İstatistikte Güncel Konular-III

EDİTÖR: PROF. DR. ÖZLEM ALPU

EĞİTİM
yayınevi

İstatistikte Güncel Konular-III

Editör: Prof. Dr. Özlem Alpu

Yayınevi Grubu Genel Başkanı: Yusuf Ziya Aydoğan (yza@egitimyayinevi.com)
Genel Yayın Yönetmeni: Yusuf Yavuz (yusufyavuz@egitimyayinevi.com)
Sayfa Tasarımı: Kübra Konca Nam
Kapak Tasarımı: Eğitim Yayınevi Grafik Birimi

T.C. Kültür ve Turizm Bakanlığı
Yayıncı Sertifika No: 76780

E-ISBN: 978-625-385-272-6
1. Baskı, Haziran 2025

Baskı Cilt: Sayfa Basım Sanayi
Tevfik İleri Mah. Emek Cad. Polat Sok. No: 2 Pursaklar / Ankara
Matbaa Sertifika No: 77079

Kütüphane Kimlik Kartı

İstatistikte Güncel Konular-III

Editör: Prof. Dr. Özlem Alpu
III+86 s., 160x240 mm
Kaynakça var, dizin yok.
E-ISBN: 978-625-385-272-6

EĞİTİM
yayınevi

Yayınevi Türkiye Ofis: İstanbul: Eğitim Yayınevi Tic. Ltd. Şti., Atakent mah. Yasemen sok. No: 4/B, Ümraniye, İstanbul, Türkiye

Konya: Eğitim Yayınevi Tic. Ltd. Şti., Fevzi Çakmak Mah. 10721 Sok. B Blok, No: 16/B, Safakent, Karatay, Konya, Türkiye
+90 332 351 92 85, +90 533 151 50 42
bilgi@egitimyayinevi.com

Yayınevi Amerika Ofis: New York: Egitim Publishing Group, Inc.
P.O. Box 768/Armonk, New York, 10504-0768, United States of America
americaoffice@egitimyayinevi.com

Lojistik ve Sevkiyat Merkezi: Kitapmatik Lojistik ve Sevkiyat Merkezi, Fevzi Çakmak Mah. 10721 Sok. B Blok, No: 16/B, Safakent, Karatay, Konya, Türkiye
sevkiyat@egitimyayinevi.com

Kitabevi Şubesi: Eğitim Kitabevi, Şükran mah. Rampalı 121, Meram, Konya, Türkiye
+90 332 499 90 00
bilgi@egitimkitabevi.com

İnternet Satış: www.kitapmatik.com.tr
bilgi@kitapmatik.com.tr

EĞİTİM YAYINEVİ GRUBU

İÇİNDEKİLER

BİR YÖNLÜ VARYANS ANALİZİNE ALTERNATİF OLAN PARAMETRİK k-ÖRNEKLEM TESTLERİ [*]

Aslı Ceren Macunluoğlu[**], Gökhan Ocakoğlu[***]

Giriş

İstatistiksel veri analizi, araştırılmak istenen belirli bir konu ile ilgili olarak elde edilen verilerin düzeltilmesi, incelenmesi, özetlenmesi, analiz edilmesi ve analiz sonucunda elde edilen sonuçların yorumlanması için kullanılan bilimsel bir süreçtir (Moore, 2014). İstatistiksel veri analizine başlamadan önce araştırma yapılacak konunun belirlenmesi, araştırmanın amacına uygun verilerin toplanması, incelenmek istenen konuya ilişkin sorularının belirlenmesi büyük önem taşımaktadır (Montgomery, 2013). İstatistiksel veri analizinin bir diğer önemli adımı ise kullanılacak olan test yönteminin seçimidir. Parametrik testler, belirli varsayımlar altında aralıklı ya da oransal ölçekle ölçülmüş veriler üzerinde uygulanabilen testlerdir. Parametrik testlerin doğru şekilde uygulanabilmesi için gözlemlenen verilerin, belirli varsayımlara dayalı olarak seçilmiş bir anakütleden geldiği kabulüne dayanır. Bu varsayımlar arasında genellikle verilerin normal dağılıma uygunluk göstermesi ve varyansların homojen olması gibi koşullar yer alır. Dolayısıyla, parametrik testler, araştırılan problemlerin çözümünde ve elde edilen sonuçların yorumlanmasında, bu varsayımların sağlandığı durumlarda tercih edilen yöntemlerdir.

Bir yönlü varyans analizi (ANOVA), üç veya daha fazla grup ortalamasının karşılaştırılmasında kullanılan en önemli parametrik testlerden biridir (Luepsen, 2017). ANOVA, F testi temelinde çalışır ve farklı gruplara ait ortalamaların istatistiksel olarak anlamlı bir farklılık gösterip göstermediğini belirlemek için F istatistiğini kullanır. ANOVA'nın uygulanabilmesi için belirli varsayımların sağlanması gerekmektedir. Bu varsayımlar; gruplardaki gözlem değerlerinin bağımsız olması, her grubun

[*] Bu çalışma Prof. Dr. Gökhan OCAKOĞLU danışmanlığında 01.08.2019 tarihinde tamamladığımız "Bir Yönlü Varyans Analizine Alternatif Olan Parametrik Ve Parametrik Olmayan K-Örneklem Test Prosedürlerinin Performanslarının Karşılaştırılması" başlıklı yüksek lisans tezi esas alınarak hazırlanmıştır (Yüksek Lisans Tezi, Bursa Uludağ Üniversitesi Sağlık Bilimleri Enstitüsü, Bursa, Türkiye, 2019).

[**] Doktora öğrencisi, Bursa Uludağ Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı, Bursa, Türkiye cerenmacunluoglu@gmail.com, ORCID: orcid.org/0000-0002-6802-5998

[***] Prof. Dr., Bursa Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Bursa, Türkiye ocakoglu@uludag.edu.tr , ORCID: orcid.org/0000-0002-1114-6051

ait olduğu anakütlenin normal dağılıma uygunluk göstermesi ve grupların varyanslarının homojen olmasıdır. Eğer bu varsayımlardan biri veya birkaçı ihlal edilirse, ANOVA'nın sonuçları güvenilirliğini kaybedebilir ve testin Tip I hata olasılığı artabilir. Bu tür durumlarda, bir yönlü varyans analizine literatürde alternatif olarak belirtilen parametrik olmayan test yöntemleri tercih edilmelidir.

Bu bölümde, bir yönlü varyans analizine ve bir yönlü varyans analizinin parametrik alternatifi olarak literatürde yer alan testlere yer verilmiştir.

Bir yönlü varyans analizi (ANOVA testi)

Bir yönlü varyans analizi, üç ya da daha fazla grubun ortalamasının karşılaştırılmasında kullanılan en önemli parametrik testlerden biridir (Luepsen, 2017).

X_{ij}, i. gruptaki, j. örneklem olsun. Her bir X_{ij}'nin, μ_i ve σ_i^2 parametrelerine sahip normal dağılıma uygunluk gösterdiği kabul edilir. μ_i ve σ_i^2 parametreleri, Eşitlik (1) ve Eşitlik (2) yardımıyla elde edilir (Patrick, 2007).

$$\bar{X}_{i.} = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i} \quad (1)$$

$$S_i^2 = \frac{\sum(X_{ij} - \bar{X}_{..})^2}{n_i - 1} \quad (2)$$

$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$ şeklinde tek faktörlü bir model olarak tanımlansın. Burada;

Y_{ij}: i. grup, j. gözlemdeki cevap değişkeni,

μ: ortalama,

α_j: j. gözlem etkisi,

ε_{ij}: hata terimidir.

Hata terimi ε_{ij}'ler, 0 ortalamaya, σ_i^2 ortak varyansa sahip normal dağılıma uygunluk gösterir ve hata terimlerinin birbirlerinden bağımsız oldukları varsayılır.

Karşılaştırılacak gruplara ait ortalamalar arasında farklılık olup olmadığını belirlemek için kurulan hipotezler aşağıdaki gibidir;

$H_0: \mu_1 = \cdots = \mu_k = 0$ i= 1, 2, 3, …, k

H_1: en az bir $\mu_i \neq 0$

H_0 hipotezi, "en az bir ortalama farklıdır" şeklinde olan hipoteze karşı sınanır.

Bir yönlü varyans analizine ait test istatistiği, Eşitlik (5) yardımıyla hesaplanır.

$$\boldsymbol{Gruplar\ Arası\ Kareler\ Ortalaması\ (GAKO)} = \frac{\sum_{i=1}^{k} n_i(X_{i.}-\bar{X}_{..})^2}{k-1} \quad (3)$$

$$\boldsymbol{Grup\ İçi\ Kareler\ Ortalaması\ (GİKO)} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i.})^2}{N-k} \quad (4)$$

$$\boldsymbol{F = \frac{GAKO}{GİKO}} \quad (5)$$

Hesaplanan F > $F_{\alpha;\mathrm{k}-1,N-k}$ ise H_0 hipotezi reddedilir.

Welch Testi

Welch testi, üç veya daha fazla anakütle ortalamalarının karşılaştırılabilmesi için geliştirilen, bir yönlü varyans analizine alternatif olarak önerilen parametrik bir testtir (Welch, 1951). Welch testi, iki bağımsız grup arasındaki ortalamaların karşılaştırılması sırasında karşılaşılan Behren-Fisher probleminin ortadan kaldırılması için önerilen bir testtir.

Welch testi, verilerin dağılımının normal dağılıma uygunluk göstermesi durumunda bir yönlü varyans analizine alternatif bir test olmakla beraber varyansların homojenlik varsayımını ihlal etmesine karşı da sağlam bir testtir (Cribbie, Fiksenbaum, & Keselman, 2012).

Test istatistiğini hesaplayabilmek için öncelikle ağırlıklandırma katsayısı (w_i) Eşitlik (6) ile hesaplanır. Ağırlıklandırma katsayısı hesaplanırken örneklem sayısından ve örneklem varyansından yararlanılır.

$$w_i = n_i/s_i^2 \quad (6)$$

Karşılaştırılacak gruplara ait ortalamalar arasında farklılık olup olmadığını belirlemek için kurulan hipotezler aşağıdaki gibidir;

$H_0: \mu_1 = \cdots = \mu_k = 0$ i= 1, 2, 3, …, k

H_1: en az bir $\mu_i \neq 0$

H_0 hipotezi, "en az bir ortalama farklıdır" şeklinde olan hipoteze karşı sınanır.

Test istatistiği Eşitlik (1.7) ile elde edilir (Welch, 1951).

$$F_W = \frac{\sum_{i=1}^{k} w_i(\bar{x}_i - \hat{\mu})^2/(k-1)}{1+[2(k-2)/k^2-1]\sum h_i} \quad (7)$$

Eşitlik (7)'de yer alan $\hat{\mu}$, Eşitlik (8), W Eşitlik (9) ve h_i Eşitlik (10) yardımı ile hesaplanır.

$$\hat{\mu} = \sum_{i=1}^{k} w_i x_i / W \quad (8)$$

$$W = \sum_{i=1}^{k} w_i \quad (9)$$

$$h_i = (1 - w_i/W)^2/(n_i - 1) \quad (10)$$

Hesaplanan test istatistiği F_W, $F_{k\text{-}1,\, f}$ dağılımına sahiptir.

Tablo değerinin belirlenmesinde kullanılan serbestlik derecesi "*f*", Eşitlik (11) yardımı ile elde edilir.

$$f = (k^2 - 1)/(3\sum h_i) \quad (11)$$

Hesaplanan test istatistiği $F_W > F_{\alpha;k\text{-}1,\, f}$ ise H_0 hipotezi reddedilir.

Alexander-Govern Testi

Alexander-Govern testi, verilerin normal dağılım göstermesi durumunda bir yönlü varyans analizi testi yerine tercih edilebilecek parametrik bir testtir. Alexander-Govern testi, varyansların homojenliği varsayımı ihlal ihlaline karşı da sağlam bir testtir (Alexander ve Govern, 1994).

Alexander-Govern test istatistiğini hesaplayabilmek için Eşitlik (12) yardımıyla her grup için standart hata değeri hesaplanır. Sonrasında Eşitlik (13) yardımıyla her grup için ağırlıklandırma katsayısı (W_i) hesaplanır.

$$S_{\bar{X}_i} = \sqrt{\frac{\sum_{j=1}^{n_i}\left(X_j - \bar{X}_i\right)^2}{n_i(n_i-1)}} = \sqrt{\frac{S_i^2}{n}} \quad (12)$$

$$W_i = \frac{1/S_i^2}{\sum_{i=1}^{k}(1/S_i^2)} \quad (13)$$

Yine, her bir grup için hesaplanması gereken t test istatistiğinin hesaplanmasında kullanılan genel ortalamanın varyansla ağırlıklandırılmış tahmini (X^+) hesaplanır.

$$X^+ = \sum_{i=1}^{k} W_i \bar{X}_i \quad (14)$$

Eşitlik (15) yardımı ile t test istatistiği hesaplanır.

$$t_i = \frac{\bar{X}_i - X^+}{S_{\bar{X}_i}} \quad (15)$$

Eşitlik (15) yardımı ile hesaplanan değerler, normalizasyon yaklaşımı kullanılarak standart normal dağılıma (Z_i) dönüştürülür. Z_i, Eşitlik (16) yardımıyla hesaplanır (Hill, 1970).

$$Z_i = c + \frac{(c^3+3c)}{b} - \frac{(4c^7+33c^5+240c^3+855)}{(10b^2+8bc^4+1000b)} \quad (16)$$

Burada a_i, b, c ve v_i aşağıdaki Eşitlikler yardımı ile elde edilir.

$$a_i = v_i - 0.5 \quad (17)$$

$$b = 48a^2 \quad (18)$$

$$c = \left[a * ln\left(1 + \frac{t_i^2}{v_i}\right)\right]^{\frac{1}{2}} \quad (19)$$

$$v_i = n_i - 1 \quad (20)$$

Karşılaştırılacak gruplara ait ortalamalar arasında farklılık olup olmadığını belirlemek için kurulan hipotezler aşağıdaki gibidir;

$H_0: \mu_1 = \cdots = \mu_k = 0$ i= 1, 2, 3, …, k

H_1: en az bir $\mu_i \neq 0$

Test istatistiği Eşitlik (21) yardımıyla elde edilir.

$$AG = \sum_{i=1}^{k} Z_i^2 \quad (21)$$

Eğer AG > χ^2_{k-1} ise H_0 hipotezi reddedilir.

James Second-Order Testi

James Second-Order testi, bir yönlü varyans analizi testine alternatif olarak James (1951) tarafından geliştirilmiştir. James Second-Order testi, grup varyanslarının homojen olmadığı durumlarda sağlam bir testtir.

James Second-Order istatistiğini hesaplayabilmek için gruplara ait standart hata değerleri hesaplanır. Sonrasında Eşitlik (22) yardımıyla her grup için ağırlıklandırma katsayısı (a_i), hesaplanır.

$$a_i = \frac{1/S_i^2}{\sum_{i=1}^{k}(1/S_i^2)} \quad (22)$$

Genel ortalamanın varyansla ağırlıklandırılmış tahmini, $\overline{Y}$, Eşitlik (23) ile hesaplanır.

$$\bar{Y} = \sum_{i=1}^{k} a_i \bar{X}_i \quad (23)$$

t istatistiği, Eşitlik (24) ile hesaplanır.

$$t_i = \frac{\bar{X}_i - \bar{Y}}{S_i^2} \tag{24}$$

Hipotezler;

H_0: $\mu_1 = \ldots = \mu_k$ ve

H_1: $\mu_i \neq \mu_j$ (en az bir ortalama farklıdır) şeklinde kurulur.

Test istatistiği ise Eşitlik (25) yardımıyla belirlenir (Cribbie, Fiksenbaum ve Keselman, 2012).

$$J = \sum_{i=1}^{k} t_i^2 \tag{25}$$

Eğer J> CV_α ise H_0 hipotezi reddedilir. CV_α ise Eşitlik (26) ile hesaplanır (Cribbie, Fiksenbaum, & Keselman, 2012).

$$\begin{aligned} CV_\alpha = {} & C + \frac{1}{2}(3\chi_4 + \chi_2)T + \frac{1}{16}(3\chi_4 + \chi_2)^2\left(1 - \frac{J-3}{C}\right)T^2 + \frac{1}{2}(3\chi_4 + \\ & \chi_2)\Big[(8R_{23} - 10R_{22} + 4R_{21} - 6R_{12}^2 + 8R_{12}R_{11} - 4R_{11}^2) + (2R_{23} - \\ & 4R_{22} + 2R_{21} - 2R_{12}^2 + 4R_{12}R_{11} - 2R_{11}^2)(\chi_2 - 1) + \frac{1}{4}(-R_{12}^2 + 4R_{12}R_{11} - \\ & 2R_{12}R_{10} - 4R_{11}^2 + 4R_{11}R_{10} - R_{10}^2)(3\chi_4 - 2\chi_2 - 1)\Big] + (R_{23} - 3R_{22} + \\ & 3R_{21} - R_{20})(5\chi_6 + 2\chi_4 + \chi_2) + \frac{3}{16}(R_{12}^2 - 4R_{23} + 6R_{22} - 4R_{21} + \\ & R_{20})(35\chi_8 + 15\chi_6 + 9\chi_4 + 5\chi_2) + \frac{1}{16}(-R2_{22} + 4R_{21} - \\ & R_{20} + 2R_{12}R_{10} - 4R_{11}R_{10} - 4R_{11}R_{10} + R_{10}^2) \times (9\chi_8 - 3\chi_6 - \\ & 5\chi_4 - \chi_2) + \frac{1}{4}(-R_{22} + R_{11}^2)(27\chi_8 + 3\chi_6 + \chi_4 + \chi_2) + \\ & \frac{1}{4}(R_{23} - R_{12}R_{11})(45\chi_8 + 9\chi_6 + 7\chi_4 + 3\chi_2) \end{aligned} \tag{26}$$

Kritik değer hesabında yer alan C değeri, (k-1) serbestlik derecesine sahip χ^2 tablo değerine karşılık gelir. Kritik değer hesabında yer alan bir başka değer T değeri ise Eşitlik (27) yardımı ile elde edilir.

$$T = \sum_{i=1}^{k} \frac{(1-a_i)^2}{n_i - 1} \tag{27}$$

Yine kritik değer hesabında yer alan χ_{2s} değeri, Eşitlik (28) yardımı ile hesaplanır.

$$\chi_{2s} = \frac{\chi_{k-1,\alpha}^{2s}}{[(k-1)(k+1)\ldots(k+2s-3)]} \tag{28}$$

R_{st} ise Eşitlik (29) yardımı ile hesaplanır.

$$R_{st} = \sum_{i=1}^{k} \frac{a_i^t}{(n_i - 1)^2} \tag{29}$$

Brown-Forsythe Testi

Brown-Forsythe testi, bir yönlü varyans analizinin parametrik alternatifi olan önerilen bir testtir. Örneklem sayısının küçük olduğu veri setlerinde anakütle varyanslarının homojenlik varsayımını sağlamaması ancak normallik varsayımını sağlanması durumunda bir yönlü varyans analizinin performansının yetersiz kalmasından dolayı geliştirilmiş bir testtir. Brown ve Forsythe (1974) çalışmalarında, gruplardaki örneklem sayısını 4 ile 21 arasında ele almış ve küçük hacimli olarak kabul etmiştir.

Hipotezler;

$H_0: \mu_1 = \cdots = \mu_k$ ve

H_1: en az bir $\mu_i \neq 0$ (en az bir ortalama farklıdır) şeklindedir.

Test istatistiği Eşitlik (30) yardımıyla hesaplanır (Brown, & Forsythe, 1974).

$$BF = \frac{\sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2}{\sum_{i=1}^{k}(1-n_i/N)(S_i)^2} \quad (30)$$

n_i: i. grubun örneklem büyüklüğü

N: toplam örneklem büyüklüğü,

$\overline{X}_i$: i. grubun ortalaması

$\overline{X}$: genel ortalama

S_i: i. grubun örneklem varyansı

Test istatistiği dağılımının, *(k-1)* ve *f* serbestlik derecesine sahip F dağılımı olduğu varsayılır (Roth, 1983).

Serbestlik derecesinin hesaplanmasında kullanılan "C_i"değeri, Eşitlik (31) ve Eşitlik (32) yardımıyla belirlenir.

$$f = \left(\sum_{i=1}^{k} \frac{C_i^2}{n_i - 1}\right)^{-1} \quad (31)$$

$$C_i = \frac{\left(1-\frac{n_i}{N}\right){S_i}^2}{\left[\sum_{i=1}^{k}\left(1-\frac{n_i}{N}\right){S_i}^2\right]} \quad (32)$$

Hesaplanan test istatistiği BF > $F_{\alpha;k-1,\,f}$ ise H_0 hipotezi reddedilir.

SONUÇ

İstatistiksel veri analizinde, araştırılmak istenen konu ile ilgili olarak toplanan verilerin düzeltilmesi, incelenmesi, özetlenmesi, analiz edilmesi ve analiz sonucunda elde edilen sonuçların yorumlanması ne kadar önemliyse elde edilen veriyi analiz etmek için kullanılacak olan test yönteminin belirlenmesi de o kadar önemlidir. Uygun test yönteminin belirlenebilmesi için belirli varsayımların sağlanması gerekmektedir. Bu varsayımlar arasında genellikle verilerin normal dağılım göstermesi ve varyansların homojen olması gibi koşullar yer alır. Dolayısıyla, parametrik testler, araştırılan problemlerin çözümünde ve elde edilen sonuçların yorumlanmasında, bu varsayımların sağlandığı durumlarda tercih edilen yöntemlerdir. Bir yönlü varyans analizine (ANOVA) alternatif olarak literatürde belirtilen ve bu bölümde de yer verilen parametrik testlerinin kullanılması önerilmektedir.

Kaynaklar

Kitap

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah: Lawrence Erlbaum Associates.

Montgomery, D.C. (2013). *Design and Analysis of Experiments*. Eight Edition, John Wiley & Sons Inc., Hoboken.

Moore, D. S., McCabe, G. P., & Craig, B. A. (2014). *Introduction to the practice of statistics*. Eighth edition/Student edition. New York, W.H. Freeman and Company, a Macmillan Higher Education Company.

Dergi Makalesi

Alexander, R., Govern, D. (1994) A New and Simpler Approximation for ANOVA under Variance Heterogeneity. *Journal of Educational Statistics* 2: 91-101.

Bishop, T., Dudewicz, E. (1978.) Exact analysis of variance with unequal variances: test procedures and tables. *Technometrics* 20: 419-430.

Bonett, D. G., & Price, R. M. (2002). Statistical inference for a linear function of medians: Confidence intervals, hypothesis testing, and sample size requirements. *Psychological Methods*, 7(3), 370–383. https://doi.org/10.1037/1082-989X.7.3.370

Brown M., Forsythe A. (1974) The Small Sample Behavior of Some Statistics Which Test the Equality of Several Means. *Technometrics* 16: 129-132.

Buning H. (1997) Robust Analysis of Variance. Journal of Applied Statistics 24: 319-332.

Cribbie R., Fiksenbaum L., Keselman H. et al (2012) Effect of non-normality on test statistics for one-way independent groups designs. *British Journal of Mathematical and Statistical Psychology* 56–73.

Hill G. (1970) Algorithm 395. Student's t-distribution. *Communications of the ACM* 13(10): 617-619.

James G. S. (1951) The Comparison of Several Groups of Observations When the Ratios of the Population Variances are Unknown. *Biometrika*, 38 (3/4): 324-329.

James G. S. (1951) The Comparison of Several Groups of Observations When the Ratios of the Population Variances are Unknown. *Biometrika*, 38 (3/4): 324-329.

Kohr R., Games P. (1974) Robustness of the Analysis of Variance, the Welch Procedure and a Box Procedure to Heterogeneous Variances. *Journal of Experimental Education* 43:61-69.

Leech N. L., Onwuegbuzie A. (2002) A Call for Greater Use of Nonparametric Statistics. Speeches/Meeting Papers.

Luepsen H. (2017) Haiko Luepsen Comparison of nonparametric analysis of variance methods: A vote for van der Waerden. *Communications in Statistics - Simulation and Computation*, DOI: 10.1080/03610918.2017.1353613.

Moder K. (2010) Alternatives to F-Test in One Way ANOVA in case of Heterogeneity of Variances (a simulation study). *Psychological Test and Assessment Modeling* 52(4): 343-353.

Oshima T., Algina J. (1992) Type I Error Rates for James's Second-Order Test and Wilcox's Hm Test under Heteroscedasticity and Non-Normality. *British Journal of Mathematical and Statistical Psychology* 45(2): 255–263.

Peterson K. (2002) Six Modifications Of The Aligned Rank Transform Test For Interaction. *Journal of Modern Applied Statistical Methods* 100-109.

Piepho H.-P. (1996) A Monte Carlo test for variance homogeneity in linear models. *Biometrical Journal* 38:461-473.

Roth A. J. (1983) Robust Trend Tests Derived and Simulated: Analogs of the Welch and Brown-Forsythe. *Journal of the American Statistical Association* 78: 972-980.

Satterthwaite F. E. (1941) Synthesis of Variance. Psychometrika 6(5): 309-316.

Welch B. (1951) On the Comparison of Several Mean Values: An Alternative Approach. *Biometrika* 38: 330-336.

Welch B. L. (1947) The Generalization of 'Student's' Problem when Several Different Population Variances are Involved. *Biometrika* 34(1-2):28-35.

Yayımlanmamış Tez

Patrick J. (2007) *Simulations to Analyze Type I Error and Power in The Anova F Test and Nonparametric Alternatives* (Unpublished Master's thesis). University of West Florida, USA.

ETKİ BÜYÜKLÜĞÜ YÖNTEMLERİ*

Ayşegül Yabacı Tak**, İlker Ercan***

Giriş

p-değeri, iki grup arasındaki gözlemlenen farkın rastlantısal olarak ortaya çıkma olasılığını ifade eder. Eğer p-değeri belirlenen alfa düzeyinden yüksekse, gözlenen farkın örnekleme bağlı varyasyonla açıklanabileceği varsayılır. Ancak çok büyük örneklemlerde yapılan analizlerde p-değerleri çoğu zaman anlamlı sonuçlar verebilir; bu durum her zaman pratik veya klinik olarak anlamlı bir fark olduğu anlamına gelmez. Dolayısıyla, yalnızca istatistiksel anlamlılık saptanması, gerçek ve önemli bir farkın varlığını garanti etmez. p-değeri örneklem büyüklüğü ve etki büyüklüğünün her ikisinden etkilenirken, etki büyüklüğü çoğunlukla örneklem genişliğinden ayrı düşünülmektedir. Bu sebeple, geniş örneklemlerle yapılan çalışmalarda sadece p-değerinin raporlanması, bulguların doğru ve bütüncül yorumlanması açısından yetersiz kalabilir (P. Ellis, 2009; Sullivan & Feinn, 2012).

Etki büyüklüğü (Effect Size), bir müdahale/tedavinin ya da uygulamanın etkisinin nicel olarak değerlendirilmesini sağlayan pratik ve standartlaştırılmış bir ölçüttür. Farklı araştırma alanlarında kolaylıkla hesaplanabilir, yorumlanabilir ve elde edilen sonuçlara doğrudan uygulanabilir niteliktedir. Etki büyüklüğü, yalnızca istatistiksel anlamlılık düzeyine odaklanmaktan ziyade, gözlemlenen etkinin gerçek ve göreli büyüklüğünü ortaya koyarak daha derinlemesine ve anlamlı yorumlara zemin hazırlar. Bu yönüyle, araştırma bulgularının gücünü değerlendirmede ve sonuçların genellenebilirliğini yorumlamada temel bir göstergedir. Literatürde, bağlama ve veri türüne bağlı olarak farklı etki büyüklüğü ölçütleri tanımlanmıştır. Oakes (1986), etki büyüklüğünü örneklem ortalamaları arasındaki fark, birden fazla örneklem ortalamasının varyansı ya da tek bir örneklemde gözlenen ilişkinin gücü şeklinde tanımlamıştır. Cohen (1988) ise bu kavramı, farklı müdahalelerin karşılaştırıldığı çalışmalarda

** Doç. Dr., Bezmiâlem Vakıf Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, İstanbul, aysegulyabaci@gmail.com, ORCID: orcid.org/ 0000-0002-5813-3397
*** Prof. Dr., Bursa Uludağ Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Bursa, iercan@msn.com, ORCID: orcid.org/ 0000-0002-2382-290X

grup ortalamaları arasındaki farkın büyüklüğünü yansıtan bir ölçüt olarak ifade etmiştir. Kramer and Rosenthal (1999), etki büyüklüğünü yokluk hipotezinin reddedilme olasılığı olarak değerlendirirken, Thompson (2002) ise etki büyüklüğünün örneklem bulgularının sıfır hipotezinden ne ölçüde ayrıldığını ortaya koyduğunu ifade etmiştir. Nakagawa and Cuthill (2007), etki büyüklüğü terimini farklı anlamlarda ele almıştır: (i) Bir etkinin büyüklüğünü öngören istatistiksel ölçü (etki istatistiği) olarak; (ii) bu istatistiklerden hesaplanan spesifik sayısal değerler olarak; (iii) ve nihayetinde, bu değerler temelinde yapılan yorumsal değerlendirmeler olarak. Bu son yorumlar, etki büyüklüğünün biyolojik anlamı ya da klinik/pratik önemini içerebilir.

İki bağımsız grubun ortalamaları arasındaki farkın incelenmesi genellikle, örneklemlerin normal dağılım varsayımını sağladığı durumlarda t testi ile gerçekleştirilir. Ancak normal dağılım varsayımı sağlanmazsa, parametrik olmayan bir alternatif olan Mann-Whitney U testi (U istatistiği) tercih edilir. Bununla birlikte hem t hem de U istatistikleri yalnızca istatistiksel anlamlılığı değerlendirir; gözlemlenen farkın büyüklüğü hakkında doğrudan bilgi sunmazlar. Bu nedenle, istatistiksel önemin yanında etkinin büyüklüğünü de değerlendirebilmek için etki büyüklüğü ölçütleri kullanılmalıdır.

İki bağımsız grup için dağılım varsayımları sağlandığında, grup ortalamaları arasındaki farkın standart sapmaya oranlanmasıyla hesaplanan Cohen's d istatistiği önerilmektedir (Cohen, 1962). Eğer her iki grubun örneklem büyüklükleri eşitse, bu durumda Cohen's d, iki grup ortalaması farkının birleştirilmiş (pooled) standart sapmaya bölünmesiyle hesaplanır (Cohen, 1988). Ancak örneklem genişliği küçükse (n <20), Cohen's d istatistiği sapma gösterebilir. Bu durumda, küçük örneklemler için daha uygun olan ve Cohen's d'ye benzer olmakla birlikte düzeltilmiş bir versiyonu olan Hedges' g istatistiği önerilmektedir (Hedges, 1981). Benzer şekilde, Glass's delta yöntemi ise etki büyüklüğünü hesaplamak için sadece kontrol grubunun standart sapmasını kullanmayı önerir. Bu yaklaşım, özellikle deney grubunun varyansının müdahaleden etkilenebileceği durumlarda daha güvenilir sonuçlar verir (Glass, Smith, & McGaw, 1981).

Örneklemin normal dağılım varsayımını karşılamadığı durumlarda, parametrik etki büyüklüğü ölçütleri (örneğin Cohen's d, Hedges' g) güvenilirliğini yitirebilir ve gözlemlenen etkinin gerçek büyüklüğü hakkında

yanıltıcı sonuçlar verebilir (Grissom & Kim, 2012). Bu nedenle, dağılım varsayımının sağlanmadığı durumlar için parametrik olmayan alternatif etki büyüklüğü ölçütleri geliştirilmiştir. Bu bağlamda, Cliff (1993) tarafından geliştirilen Cliff's delta, iki bağımsız grup arasında karşılaştırma yaparken bir gruptaki değerlerin diğer gruptaki değerlerden büyük olma olasılığına dayalı bir etki büyüklüğü ölçütü olarak önerilmiştir. Cliff's delta, doğrudan sıralara dayalı olarak, bir grubun diğerine üstünlüğünü nicel olarak ifade eder. Benzer şekilde, Vargha and Delaney (2000) tarafından önerilen VDA (Vargha-Delaney A) istatistiği, stokastik üstünlük kavramına dayanan parametrik olmayan bir etki büyüklüğü ölçüsüdür ve örneklemler arasında rastgele seçilen bir gözlemin diğer gruptan seçilen gözleme üstün olma olasılığını ifade eder. Ayrıca, Mann-Whitney U testiyle birlikte kullanılan etkili parametrik olmayan etki büyüklüğü ölçütlerinden biri de Glass's Rank Biserial Correlation Coefficient'tır. Bu korelasyon katsayısı, iki grup arasında sıralara dayalı ilişkiyi ölçer ve Cureton (1956), Glass (1965) ile Wendt (1972) tarafından farklı formülasyonlarla tanımlanmıştır. Bu üç yaklaşım, aynı temel kavrama dayansa da hesaplama biçimleri bakımından küçük farklılıklar içermektedir.

Etki büyüklüğünün yorumlanmasında en yaygın kullanılan sınıflandırma, etki düzeyinin "küçük", "orta" ve "büyük" şeklinde kategorize edilmesidir. Bu kategorilere ait referans değerler ilk olarak Jacob Cohen (1962) tarafından önerilmiş olup, 0.20 küçük, 0.50 orta 0.80 büyük etki büyüklüğü olarak tanımlanmıştır. Bu sınıflandırmanın temel amacı, özellikle çalışmanın planlama aşamasında gerçekleştirilen güç analizleri yoluyla gerekli minimum örneklem büyüklüğünü belirlemeye yardımcı olmaktır (Valentine & Cooper, 2008). Zamanla, bu sınıflandırma daha da detaylandırılmış ve Sawilowsky (2009) tarafından geliştirilen genişletilmiş bir referans çerçevesi literatüre kazandırılmıştır. Bu çerçevede; 0.01 çok küçük, 0.20 küçük, 0.50 orta, 0.80 büyük, 1.20 çok büyük ve 2.0 ise oldukça büyük etki büyüklüğü olarak tanımlanmıştır. Parametrik olmayan etki büyüklüğü yöntemleri açısından ise ölçüm değerleri genellikle -1 ile +1 aralığında yer almakta olup, bu değerler karşılaştırılan iki grup arasındaki yön ve büyüklük farklılıklarını simgeler. Ayrıca, Jacob Cohen (1962) diğer etki büyüklüğü ölçütlerinin (örneğin korelasyon katsayısı, eta kare vb.) Cohen's d'ye karşılık gelen yaklaşık değerlerini de referans olması açısından literatüre kazandırmıştır.

Etki Büyüklüğü

Araştırmalarda uygulanan istatistiksel analizler, genellikle iki bağımsız grup arasında anlamlı bir farkın bulunmadığını varsayan sıfır hipotezini (H_0) test etmeyi hedefler. Bu test sonucunda elde edilen anlamlılık düzeyi (p-değeri), H_0 hipotezinin reddi halinde yapılabilecek Tip I hata (yanlış pozitiflik) olasılığını ifade eder. Ancak, p-değeri yalnızca bu hata olasılığına dair bilgi sunmakta olup, sıfır hipotezinin doğruluk derecesi ya da içeriğindeki belirsizlik düzeyi hakkında kapsamlı bir değerlendirme sağlamaz. Dolayısıyla, p-değeri tek başına sıfır hipotezinin geçerliliği konusunda yeterli çıkarım yapmaya olanak tanımaz.

$$t = \frac{\overline{Y_a} - \overline{Y_b}}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}} \qquad (1)$$

$\overline{Y_a}$: grup a'nın ortalaması,

$\overline{Y_b}$: grup b'nin ortalaması,

s_a, s_b: grup a ve grup b'nin standart sapması,

n_a, n_b: grup a ve grup b'nin örneklem genişliği

Eşitlik-1 kapsamında değerlendirildiğinde, *t* istatistiğinin istatistiksel olarak anlamlılığa ulaşması yalnızca grup ortalamaları arasındaki farkın büyüklüğüne bağlı değildir. Aynı ortalama farkı için örneklem büyüklüğü arttıkça *t* istatistiğinin değeri yükselmekte ve buna bağlı olarak *p*-değeri azalmaktadır. Bu durum, örneklem büyüklüğündeki artışın yalnızca büyük farkları anlamlı hale getirmekle kalmayıp, aynı zamanda önemsiz düzeydeki küçük farkların da istatistiksel olarak anlamlı bulunmasına neden olabileceğini göstermektedir (Grissom & Kim, 2005). Öte yandan, büyük örneklem büyüklüğü yalnızca test istatistiği üzerinde etkili olmakla kalmaz; aynı zamanda popülasyonu daha iyi temsil etme kapasitesini artırır, istatistiksel analizlerin tekrarlanabilirliğini güçlendirir, testin istatistiksel gücünü (power) yükseltir ve varsayımlara yönelik olası ihlaller karşısında analizlerin daha dayanıklı (robust) olmasına katkı sağlar.

Sağlık alanında yürütülen araştırmalarda, p-değerinin kabul edilen anlamlılık düzeyinden küçük olması, bir tedavi uygulamasının istatistiksel

olarak diğerinden anlamlı düzeyde farklı olduğunu ya da tedaviyle sonuç değişkeni arasında anlamlı bir ilişki bulunduğunu gösterir. Ancak özellikle plasebo ve aktif tedavi gruplarıyla yapılan klinik çalışmalarda, bir tedavinin etkinliğinin derecesi veya etkisinin büyüklüğü yalnızca p-değeri ile değerlendirilemez. Çünkü p-değeri, farkın varlığına ilişkin bilgi sunarken, farkın büyüklüğü veya klinik anlamı hakkında doğrudan bilgi sağlamaz. Bu bağlamda, tedavinin üstünlüğü ve etkisinin gücü, etki büyüklüğü ölçütleri ile daha doğru ve anlamlı bir şekilde ortaya konulabilir.

Etki büyüklüğü, literatürde farklı şekillerde tanımlanmıştır. Jacob Cohen (1962), etki büyüklüğünü ilk olarak "sıfır hipotezinden sapmanın bir ölçüsü" olarak tanımlamış, daha sonra Jacob Cohen (1988) "farklı tedavi/müdahaleleri karşılaştıran çalışmalarda grupların ortalamaları arasındaki farkın büyüklüğü" şeklinde ifade etmiştir. Kazis, Anderson ve Meenan'a (1989) göre etki büyüklüğü, ya bir grubun zaman içindeki değişimini standart ölçülerle ifade eden bir araçtır ya da iki grup arasındaki farkı ortaya koyan nicel bir ölçüdür. Olejnik and Algina (2003), ise etki büyüklüğünü, "örneklem genişliğinden ayrı olarak popülasyonlar arasındaki farkı ya da değişkenler arası ilişkiyi ölçen standart bir istatistik" olarak tanımlamıştır. Nakagawa and Cuthill (2007), etki büyüklüğünü üç şekilde ele almıştır: (a) Etkinin büyüklüğünü ölçen bir istatistik (örneğin korelasyon katsayısı r), (b) bu istatistiklerden elde edilen sayısal değerler (örneğin r = 0.3), ve (c) bu değerlerin yorumu (örneğin r = 0.3 için "orta" etki). Bu farklı tanımlar, etki büyüklüğünün farklı amaçlarla kullanılabileceğini ve kullanım amacının netleştirilmesi gerektiğini göstermektedir. Bazı araştırmacılar ise bu tanımların sınırlı olduğunu ve literatürde "etki büyüklüğü" olarak ifade edilen çoğu ölçümün bu sınırlı tanımlarla dışarıda kaldığını savunmuştur. Henson (2006) etki büyüklüğünü hem "sıfır hipotezinden sapmanın bir ölçüsü" hem de "ilgilenilen etkinin büyüklüğünü yansıtan bir ölçü" olarak tanımlamıştır. Kelley & Preacher (2012) ise en genel anlamda etki büyüklüğünü, "bir müdahalenin, tedavinin ya da durumun etkisinin büyüklüğünü nicel olarak ifade eden bir ölçü" şeklinde özetlemiştir.

Etki büyüklüğü, yalnızca klinik ya da uygulamaya yönelik bir etkinin derecesini göstermekle kalmaz; aynı zamanda araştırma planlama sürecinde, özellikle örneklem büyüklüğünü belirlemek için yapılan güç analizlerinde ve meta-analiz çalışmalarında önemli bir rol oynar. İstatistiksel güç, yanlış bir sıfır hipotezinin (H_0) kabul edilmeme olasılığıdır. Güç arttıkça etki büyüklüğü de artacaktır. Bu nedenle, araştırmacıların araştırma öncesinde

muhtemel etki büyüklüğünü tahmin etmesi ya da çalışmanın amacına uygun minimum etki büyüklüğüne karar vermesi gerekir. Preacher ve Kelley (2011), bir etki büyüklüğü ölçütünün aşağıdaki dört maddeye sahip olması gerektiğini vurgulamıştır:

i) Etki büyüklüğü uygun bir ölçekte ifade edilmelidir. Etki büyüklüğü, kullanılan ölçüm aracı ve araştırma sorusu dikkate alınarak doğru şekilde ölçeklendirilmelidir. Anlamlı ve doğru yorum yapabilmek için yorumlanabilir bir ölçek kullanılması gerekir. Genellikle standartlaştırılmış etki büyüklüğü tercih edilir. Bu standardizasyon, araştırmacının her yeni durumda farklı bir değerlendirme ölçütü belirleme ihtiyacını ortadan kaldırır (Cohen, 1988).

ii) Etki büyüklüğü ile güven aralığı da raporlanmalıdır. Etki büyüklüğü, örneklem üzerinden hesaplanan bir tahmindir ve bu değer, gerçek (popülasyon) etkiden farklı olabilir. Bu nedenle, etki büyüklüğünün ne kadar güvenilir olduğunu göstermek için güven aralıklarıyla birlikte raporlanması gerekir.

iii) Etki büyüklüğü örneklem büyüklüğünden ayrı düşünülmelidir. Etki büyüklüğü, popülasyondaki gerçek etkinin bir tahmini olarak kabul edilir ve bu nedenle, örneklem büyüklüğünden etkilenmemelidir. Aynı etki büyüklüğü, farklı örneklem büyüklükleriyle benzer şekilde hesaplanabilmelidir.

iv) Etki büyüklüğü hesaplamaları güvenilir olmalıdır. Etki büyüklüğü tahminlerinin istatistiksel olarak iyi özellikler taşıması gerekir. Yani bu tahminlerin yansız (bias içermeyen), tutarlı (örneklem büyüdükçe gerçek değere yaklaşan) ve etkin (minimum hata ile tahmin yapan) olması beklenir.

Etki Büyüklüğünün Yönleri

Etki büyüklüğü üç temel boyutta ele alınabilir. Birinci boyut, araştırmada ilgilenilen bilgi türünü ifade eder. İkinci boyut, etki büyüklüğünün belirli istatistiksel formüller veya parametrelerle ilişkilendirilerek hesaplanmasını, yani işlevsel hale getirilmesini kapsar. Üçüncü boyut ise, hesaplanan etki büyüklüğünün aldığı nicel değerdir.

- **Etki Büyüklüğü Boyutu:** Fizikte boyutlar, genelleştirilmiş ölçüm kavramları olarak değerlendirilir (Carman, 1969; B. Ellis, 1968; Ipsen, 1960). Örneğin kütle, yoğunluk, kuvvet ve enerji, uzunluk gibi fiziksel büyüklükler birer boyuttur ve her biri farklı birimlerle ölçülebilir. Etki büyüklüğü bağlamında ise "boyut" kavramı, ölçülen bilginin soyut bir tanımı olarak ele alınır. Başka bir deyişle etki büyüklüğünün boyutu, sabit bir fiziksel birime bağlı olmayan ancak ölçülebilir bir bilgi türünü temsil eder ve araştırmada incelenen durum hakkında genel bir çerçeve sunar. Örneğin, araştırma konusu iki değişken arasındaki ilişki ise, etki büyüklüğünün boyutu bu ilişkiyi yansıtan korelasyon katsayısı, kovaryans ya da regresyon katsayısı gibi ölçülerle ifade edilebilir. Etki büyüklüğünün boyutuna örnek olarak varyans, standart sapma, dağılım aralığı (range) ve çeyrekler arası aralık (IQR) gibi istatistiksel değişkenlik ölçütleri verilebilir (Kelley & Preacher, 2012).
- **Etki Büyüklüğü Ölçüsü:** Etki büyüklüğü ölçütü ya da etki büyüklüğü indeksi, belirli bir olgunun, tedavinin veya müdahalenin etkisinin derecesini nicel olarak değerlendirmek amacıyla kullanılan istatistiksel bir göstergedir. İki grup ortalaması arasındaki farkı değerlendirmede yaygın olarak kullanılan etki büyüklüğü, Eşitlik-2'de tanımlanan standartlaştırılmış ortalama farkı ile ifade edilir.

$$Etki\ Büyüklüğü = \frac{\bar{x}_1 - \bar{x}_2}{S_{pooled}} \qquad (2)$$

Formülde, $\bar{x}_j$ (j=1,2) j.nci grup ortalamasını ve S_{pooled} grup içi varyansın yansız tahmininin karekökü̈dür. Etki büyüklüğünü değerlendirmeye yönelik bir başka örnek ise, hata kareler ortalamasının karekökü̈ne dayanan RMSEA (Root Mean Square Error of Approximation) yaklaşımıdır. RMSEA, etki büyüklüğü ölçütü olarak Eşitlik-3'te verilen şekilde tanımlanmaktadır.

$$\hat{\varepsilon} = \sqrt{max\left\{0, \frac{\hat{F}_0}{v}\right\}} \qquad (3)$$

Burada $\hat{F}_0$ popülasyona ait maksimum olabilirlik fark fonksiyonunun tahminini; v ise serbestlik derecesini ifade etmektedir. Genel olarak etki

büyüklüğü, bir olgu ya da tedavi/müdahalenin etkisini değerlendirmek amacıyla veri, istatistik veya parametrelere dayalı olarak hesaplanan nicel ve açık bir göstergedir (Kelley & Preacher, 2012).

- **Etki Büyüklüğü Değeri:** Bir tedavi ya da müdahalenin etkisini incelemek için, önce uygun bir etki büyüklüğü boyutu belirlenir, ardından buna uygun bir etki büyüklüğü ölçüsü uygulanır. Bu süreç sonunda elde edilen sayısal sonuç, etki büyüklüğü değeri olarak adlandırılır ve ilgili etkinin büyüklüğünü temsil eder. Etki büyüklüğünün boyutu, ölçüsü ve değeri çoğu zaman kısaca "etki büyüklüğü" terimiyle ifade edilir. Ancak hangi yönün kastedildiğini açıkça belirtmek, özellikle bilimsel çalışmalarda anlam karışıklığını önlemek açısından önemlidir. Bir etki büyüklüğü değerinin anlamlı bir şekilde yorumlanabilmesi için, bu değerin hangi ölçüye göre hesaplandığının açıkça belirtilmesi gerekir. Çünkü etki büyüklüğü ölçüsü, belirlenen boyutun nasıl işlediğini tanımlar. Dolayısıyla, etki büyüklüğü değeri yalnızca ölçüye dayalı olarak anlam kazanır ve bu nedenle hem ölçü hem de hesaplama yöntemi net bir şekilde ifade edilmelidir (Kelley & Preacher, 2012).

Etki Büyüklüğü Tahminlerinde Güven Aralığının Rolü

Bilimsel ve deneysel araştırmalarda etki büyüklüğünün raporlanması, popülasyondaki gerçek etki büyüklüğüne ilişkin bir nokta tahmini sunar. Ancak bu tahminin kesinliği hakkında bilgi vermek için, etki büyüklüğüyle birlikte güven aralığının (CI) da raporlanması gerekir. Neyman (1937), nokta tahmininin (T) her zaman gerçek popülasyon parametresi (θ) ile birebir örtüşmesinin mümkün olmadığını vurgulamış ve bu nedenle aralık tahminlerinin geliştirilmesinin gerekli olduğunu belirtmiştir. Buna göre güven aralığı, Eşitlik-4'teki biçimiyle ifade edilebilir.

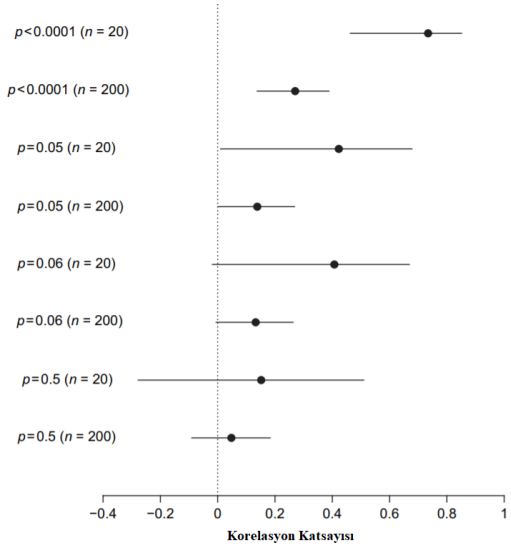
$$\underline{\theta} = T - K_1 S_T \quad ve \quad \bar{\theta} = T - K_2 S_T \qquad (4)$$

Burada, T, anakütle parametresi θ'nın örneklem üzerinden elde edilen tahminidir. S_T bu nokta tahmininin standart sapmasını ifade ederken; K_1 ve K_2, güven aralığını belirleyen sabitlerdir. Alt sınır $\underline{\theta}$ ve $\bar{\theta}$ üst sınır olmak üzere, güven aralığı popülasyon parametresinin tahmini olan θ'nın olası aralığını gösterir. Eşitlik-4'te belirtildiği üzere, standart sapma ne kadar

küçükse, yapılan tahminin doğruluğu ve güvenilirliği o kadar yüksek olacaktır.

Cox ve Snell (1981), güven aralığı (CI) ile ilgili altı temel noktaya dikkat çekmiştir: İlk olarak, güven aralıkları ile istatistiksel anlamlılık testleri yakından ilişkilidir. $(1-\alpha)\times\%100$ Güven aralığı, iki kuyruklu bir testte α anlamlılık düzeyine karşılık gelen, verilerle tutarlı olabilecek olası parametre değerlerini içerir. İkinci olarak, güven aralıkları genellikle simetrik olmalıdır.
Ancak bazı durumlarda bu simetriyi sağlamak için veri dönüşümü gerekebilir. Üçüncü, bazı analizlerde güven aralıkları yerine güven bölgeleri kullanılmalıdır. Özellikle verilerin iki ayrı değeri desteklediği durumlarda, iki parçalı bir güven bölgesi oluşturulabilir. Dördüncü, birden fazla aralık tahmin en uygun olanı seçilmelidir.
Bu seçim yöntemin hesaplamada kolaylık sağlaması, daha kesin tahminler üretmesi ve istatistiksel varsayım ihlallerine karşı duyarlı olmamasına göre yapılmalıdır. Beşinci, eğer 'nuisance' (ilgilenilmeyen ama etkili) parametreler varsa, güven aralığının kapsama olasılığı yine de belirtilen α düzeyinde olmalıdır. Altıncı ve son olarak, aynı popülasyondan tekrar tekrar örnek alınması durumunda oluşturulan güven aralıklarının büyük kısmı, gerçek popülasyon parametresini içerecektir.

Etki büyüklüğüne ait güven aralığının genişliği, yapılan tahminin ne kadar hassas olduğunu gösterir. Güven aralığını etki büyüklüğü tahminiyle birlikte değerlendirmek, yalnızca klasik anlamda istatistiksel anlamlılık hakkında değil, aynı zamanda p-değeriyle elde edilemeyen ek bilgiler hakkında da fikir verir. Bu yaklaşım, güven aralıklarının yalnızca anlamlılık testleri için değil, aynı zamanda etki büyüklüğüne dair olası değer aralığını yüksek olasılıkla yansıtması açısından da önemli olduğunu ortaya koyar.



Şekil-1. Etki büyüklüğü tahminleri (örneğin korelasyon katsayısı) ve buna karşılık gelen güven aralıkları (CI), örneklem büyüklüğüne bağlı olarak değişiklik gösterebilir. Aynı p-değerine sahip iki farklı analiz, farklı örneklem büyüklüklerine dayanıyorsa, elde edilen etki büyüklüğü tahminleri ve güven aralıkları da farklı olacaktır. Örneğin, p <0.0001 gibi oldukça anlamlı kabul edilen bir p-değeri, farklı örneklem büyüklüklerinde farklı düzeyde etki büyüklüğü ve farklı güven aralıklarıyla sonuçlanabilir (Nakagawa & Cuthill, 2007).

Şekil 1'den anlaşılacağı gibi, etki büyüklükleri ve bunlara ait güven aralıkları, p-değerlerinin tek başına sağlayamadığı bilgileri, özellikle etkinin yönü ve büyüklüğü hakkında, ortaya koyabilir. Etki büyüklüğü ve güven aralığına dayalı yorum yapmak, sonuçların daha doğru anlaşılmasını ve verilerden daha güçlü istatistiksel çıkarımlar elde edilmesini sağlar. Çoğu araştırmacı, istatistiksel hipotez testi sonucunu sadece "anlamlı" ($p < \alpha$) ya da "anlamlı değil" ($p > \alpha$) olarak değerlendirme eğilimindedir. Ancak bu yaklaşım yanıltıcıdır. Örneğin, $p = 0.06$ ile $p = 0.07$ arasındaki fark oldukça küçüktür ve etki büyüklüğü açısından önemli bir fark yaratmaz. Buna rağmen $p > 0.05$ olduğunda, genellikle hiçbir etkinin olmadığı yönünde hatalı bir sonuca varılır. Oysa bu sadece verilerin etkinin varlığını göstermek için yetersiz olduğunu ifade eder, etkinin hiç olmadığı anlamına gelmez. Bu nedenle, istatistiksel analizlerde sadece p-değeriyle sınırlı kalmak yerine, etki büyüklüğü ve etki büyüklüğüyle birlikte güven aralıklarını da raporlamak, istatistiksel olarak anlamlı olmayan sonuçların bile daha doğru ve kapsamlı şekilde yorumlanmasını sağlar (Cohen, 1992; Fisher, 1935; Nakagawa & Cuthill, 2007).

Etki Büyüklüğü Üzerinde Etkili Olan Kriterler

Etki büyüklüğü genellikle basit hesaplamalı ve yorumlanması kolay bir ölçü olsa da bazı dış etkenlere karşı duyarlıdır. Bu nedenle, etki büyüklüğü kullanılırken dikkatli olunmalı ve sonuçlar bu hassasiyetler göz önünde bulundurularak değerlendirilmelidir. Coe (2002), bu tür etkileri şu şekilde özetlemektedir:

i. Hangi Standart Sapma Kullanılmalı?

Etki büyüklüğünü hesaplarken karşılaşılan ilk problem, hangi standart sapmanın hesaplanacağıdır. Genellikle, kontrol grubu plasebo ya da müdahale edilmeyen popülasyonu temsil ettiği için, bu grubun standart sapması en uygun tahmin olarak görülür. Ancak kontrol grubunun örneklem genişliği yeterince büyük değilse, sadece bu gruba ait standart sapma doğru bir tahmin sunmayabilir. Bu nedenle, genellikle hem deney hem de kontrol gruplarının verilerini içeren birleştirilmiş (pooled) standart sapma kullanmak daha doğru bir yaklaşımdır. Pooled standart sapma, iki grubun standart sapmalarının ağırlıklı ortalamasıdır ve bu sapmaların aynı anakütle parametresinin tahminleri olduğu varsayımına dayanır. Başka bir deyişle, kontrol ve deney grupları arasındaki standart sapma farkı yalnızca örnekleme bağlı olmalı; sistematik bir fark bulunmamalıdır. Eğer iki grup arasında gerçek bir varyans farkı varsa ya da standart sapmalar farklı popülasyonlara aitse, bu durumda pooled standart sapma kullanılmamalıdır.

ii. Yanlılık Düzeltmesi

Her ne kadar etki büyüklüğünün hesaplanmasında birleştirilmiş (pooled) standart sapma kullanımı, yalnızca kontrol grubuna ait standart sapmaya göre daha isabetli bir tahmin sunsa da bu yaklaşımın küçük bir yanlılık içerdiği bilinmektedir. Genellikle bu yöntem, gerçek popülasyon etki büyüklüğünden biraz daha büyük değerler üretme eğilimindedir. Bu soruna çözüm olarak Hedges ve Olkin (1984), bu yanlılığı azaltmak amacıyla yaklaşık bir düzeltme formülü önermiştir.

$$g = \frac{\mu_1 - \mu_2}{s_{pooled}} \times \left(\frac{N-3}{N-2.25}\right) \times \sqrt{\frac{N-2}{N}} \qquad (5)$$

iii. Normal Dağılım Varsayımının İhlali

Etki büyüklüklerinin doğru şekilde yorumlanabilmesi, genellikle hem deney hem de kontrol gruplarının normal dağılıma sahip olduğu varsayımına

dayanır. Ancak bu varsayım geçerli değilse, normal dağılıma göre hesaplanmış etki büyüklükleri ile normal olmayan dağılımlara dayanan etki büyüklüklerini karşılaştırmak güçleşir. Bu durumda yapılan yorumlar yanıltıcı olabilir ve alternatif yöntemlerin kullanılması gerekebilir.

iv. Ölçüm Güvenilirliği

Etki büyüklüğünü etkileyebilecek diğer önemli faktör, kullanılan ölçüm aracının güvenilirliğidir. Klasik ölçüm kuramına göre, herhangi bir gözlem sonucu hem gerçek değeri hem de ölçüm hatasını içerir. Bu nedenle, bir örneklemde elde edilen puanlardaki varyasyon (standart sapma), sadece bireyler arasındaki gerçek farklılıkları değil, aynı zamanda ölçüm hatalarının etkisini de yansıtır. Ölçüm güvenilirliği düşükse, standart sapma olduğundan fazla görünerek etki büyüklüğünü olduğundan küçük gösterebilir ya da tam tersi, yanıltıcı sonuçlara yol açabilir.

Etki Büyüklüğünün Kullanımında Dikkat Edilmesi Gereken Noktalar

Etki büyüklüğü, gruplar arasındaki farkların ne kadar büyük olduğunu anlamamıza yardımcı olurken, istatistiksel anlamlılık ise bu farkların tesadüfen oluşup oluşmadığını sorgular. Etki büyüklüğü kullanımına dair bazı temel öneriler aşağıda özetlenmiştir (Coe, 2002):

i) Etki büyüklüğü, müdahalenin etkisini gösteren standartlaştırılmış ve ölçekten bağımsız bir ölçüdür. Ölçeği bilinmeyen ya da farklı olan çalışmalar arasında karşılaştırma yapmayı kolaylaştırır.

ii) Etki büyüklüğünün yorumlanması genellikle deney ve kontrol gruplarının normal dağıldığı ve benzer standart sapmalara sahip olduğu varsayımına dayanır. Bu varsayımlar altında, etki büyüklüğü deney ve kontrol grubu dağılımlarının ne ölçüde örtüştüğünü gösterir.

iii) Etki büyüklüğü ile güven aralığı kullanmak, istatistiksel anlamlılıktan ziyade etkinin büyüklüğüne ve anlamına odaklanmayı sağlar.

iv) Etki büyüklüğü ve buna ait güven aralıkları hem birincil araştırmalarda hem de meta-analizlerde mutlaka hesaplanmalı ve raporlanmalıdır.

v) Eğer örneklem dar bir ölçekte toplanmışsa, normal dağılmıyorsa, kullanılan ölçüm aracı bilinmiyorsa veya güvenilirliği düşükse, etki büyüklüğü yorumlanırken dikkatli olunmalıdır.

vi) Bu gibi durumlarda, gruplar arası doğrudan fark (ham ortalama farkı) ve onun güven aralığı kullanmak daha uygun olabilir. Özellikle sonuçlar bilinen bir ölçekte ifade ediliyorsa, dağılımlar normal değilse veya gruplar arasında standart sapmalar belirgin şekilde farklıysa bu yöntem tercih edilmelidir.

vii) Farklı sonuç ölçütlerine, farklı müdahalelere, farklı düzeylere ya da farklı popülasyonlara ait etki büyüklüklerini karşılaştırırken ya da bir araya getirirken özenli olunmalıdır.

viii) Etki, nedensellik çağrıştırır. Bu sebeple, "etki büyüklüğü" ifadesi, yalnızca nedensel bir ilişki kurulabiliyorsa ve bu gerekçelendirilebiliyorsa kullanılmalıdır.

Bağımsız İki Grup İçin Kullanılan Parametrik Etki Büyüklüğü Ölçütleri

Bağımsız iki grup karşılaştırıldığında, gruplar normal dağılım gösterip varyansları benzer olduğunda, Cohen's *d*, Glass's delta ve Hedges' *g* olmak üzere üç yaygın standartlaştırılmış etki büyüklüğü ölçütü kullanılabilir. Bu üç ölçüt aynı temel anakütle farkını tahmin etmeyi amaçlar ancak standartlaştırma sırasında kullanılan payda (standart sapma) açısından birbirinden ayrılırlar. Verilerin normal dağılım göstermemesi durumunda uygulanan doğrusal olmayan dönüşümler, bu etki büyüklüğü tahminlerinin değerlerini ve yorumlarını etkileyebilir (Kraemer & Andrews, 1982). Bu nedenle, bu ölçütler yalnızca müdahale etkisinin büyüklüğünü değil, aynı zamanda kullanılan dönüşüm yöntemini de yansıtabilir. Ancak normallik varsayımı sağlandığı sürece, Cohen's *d*, Glass's delta ve Hedges' *g*, hem güvenilir hem de uzun vadede doğru sonuçlar veren (asimptotik olarak etkin) tahminciler olarak kabul edilir (Hedges & Olkin, 1984; Peng & Chen, 2014).

1. Cohen d Etki Büyüklüğü Ölçütü

Cohen's *d*, alternatif hipotezin sıfır hipotezinden ne ölçüde ayrıldığını göstermek amacıyla Jacob Cohen (1962) tarafından geliştirilmiştir. Bu etki büyüklüğü ölçütü, iki grup arasında varyansların eşit olduğu varsayımı altında, grup ortalamaları arasındaki farkın grupların standart sapmasına oranlanmasıyla hesaplanır. Elde edilen fark, başlangıçta değişkenin ölçü biriminde ifade edilir; ancak bu değerin standartlaştırılması, farkın ilgili popülasyonun standart sapmasına bölünmesiyle sağlanır. Bu sayede, ölçüm biriminden bağımsız,

karşılaştırılabilir bir etki büyüklüğü elde edilmiş olur. Cohen's d etki büyüklüğü Eşitlik-6'daki gibi tanımlanmaktadır.

$$d = \frac{\mu_1 - \mu_2}{\sigma} \qquad (6)$$

Burada; μ_1: Birinci grubun ortalaması, μ_2: İkinci grubun ortalaması, σ: Anakütle standart sapmasını ifade etmektedir.

Standart sapma, bir veri setindeki değerlerin yayılımını yani dağılım genişliğini ölçen bir istatistiktir. Etki büyüklüğü hesaplamalarında genellikle popülasyonun standart sapmasına başvurulur. Ancak uygulamada bu değerin bilinmesi nadirdir. Bu sebeple, standart sapma genellikle deney grubundan, kontrol grubundan veya her iki grubun verilerinden elde edilen *birleştirilmiş (pooled)* bir tahminle belirlenir. Bu durumda, iki bağımsız grup arasındaki farkın büyüklüğünü hesaplamak için kullanılan formül, J. Cohen (1988) tarafından önerilen ve Eşitlik-7'de gösterilen yapıdadır.

$$d = \frac{\mu_1 - \mu_2}{ss_{pooled}} \qquad (7)$$

Burada;

$$ss_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \qquad (8)$$

n : Örneklem sayısı, s_1: Birinci grubun standart sapması, s_2: İkinci grubun standart sapmasını ifade etmektedir.

Eğer her iki grupta yer alan gözlem sayıları eşitse, gruplar arası ortalama varyasyonun daha doğru bir şekilde tahmin edilebilmesi için birleştirilmiş (pooled) standart sapma kullanılabilir. Bu durumda, grup sayısı k olmak üzere, birleştirilmiş standart sapma J. Cohen (1988) tarafından önerilen şekilde, Eşitlik-9'da gösterildiği biçimde hesaplanır.

$$ss_{pooled} = \sqrt{\frac{s_1^2 + s_2^2 + \cdots + s_k^2}{k}} \qquad (9)$$

Cohen's *d*, öngörüde bulunmak için oldukça kullanışlı bir etki büyüklüğü ölçütüdür; ancak bazı sınırlılıkları da vardır:

1. Öncelikle, Cohen *d*, popülasyon etki büyüklüğünü tam olarak yansıtmaz ve yanlı bir tahmin sunabilir (Hedges, 1981).
2. İkinci olarak, Cohen *d*'nin hesaplanmasında kullanılan ortak varyans tahmini, örneklem büyüklüğünden etkilenir. Bu da özellikle küçük örneklemlerde sapmaya yol açabilir (Keselman ve ark., 2008).
3. Üçüncü olarak, iki grubun varyansları eşit değilse, Cohen *d* ile güvenilir bir tahmin yapmak zordur. Çünkü bu durumda ortak popülasyon varyansı belirsizdir.
4. Ayrıca, Cohen *d*, örneklem büyüklüklerinin oranına bağlıdır ve bu nedenle örneklemler dengesizse sonuçlar etkilenebilir (Ruscio, 2008).
5. Bir diğer önemli sınırlama ise, Cohen *d*'nin klinik anlamda her zaman yorumlanabilir olmamasıdır. Kraemer ve Kupfer (2006), Cohen *d*'nin örneğin bir hastalığın önlenmesinde ya da tedavisinde ne kadar etkili olduğunu doğrudan yansıtamayacağını belirtmiştir. Çünkü klinik önem, sadece istatistiksel farkla değil, tedavinin hasta üzerindeki etkisiyle ilgilidir. Örneğin, çocuk felcini önleyen düşük riskli bir aşı ile yüksek riskli bir tedavi yöntemi aynı şekilde değerlendirilmemelidir.

Bu nedenle, bir müdahalenin klinik açıdan anlamlı olup olmadığını değerlendirebilmek için yalnızca Cohen *d* yeterli değildir. Spesifik tedaviler için "klinik önem eşiği" bilgisi gereklidir (Jacobson ve ark., 1984; Jacobson & Truax, 1992).

2. Glass delta Etki Büyüklüğü Ölçütü

Glass's delta, deneysel çalışmalarda bir grubun deney, diğerinin kontrol grubu olarak belirlendiği durumlar için geliştirilmiş bir etki büyüklüğü ölçütüdür ve özellikle meta-analizlerde kullanılır. Bu ölçüt, iki grup arasındaki ortalama farkın kontrol grubuna ait örneklem standart sapmasına oranlanmasıyla elde edilir. Alternatif olarak, eğer iki grup varyans açısından homojen kabul ediliyorsa, ortak (sınıf içi) standart sapma tahmini de kullanılabilir (Glass, 1976; Glass ve ark., 1981). Glass delta, bu tanımlamaya göre Eşitlik-10'da gösterilen formülle hesaplanmaktadır.

$$Glass\ delta = \frac{\mu_1 - \mu_2}{ss_2} \qquad (10)$$

Burada; μ_1: Deney grubu ortalaması, μ_2: Kontrol grubu ortalaması ve ss_2: Kontrol grubunun standart sapmasını ifade etmektedir.

Glass (1976) tarafından önerilen yöntemden bu yana, Glass delta'nın hesaplanmasında kontrol grubunun standart sapmasının kullanılması konusunda çeşitli tartışmalar gündeme gelmiştir. Bazı araştırmacılar, raporlama amacına göre yalnızca kontrol grubunun değil, deney grubunun standart sapmasının da standartlaştırıcı olarak kullanılabileceğini savunmuştur. Çünkü bu iki farklı yaklaşım, elde edilen bulgunun farklı yönlerini yansıtabilir (Peng & Chen, 2014). Glass delta, farklı araştırma tasarımlarına bağlı olarak farklı popülasyon parametrelerini tahmin etme yeteneğine sahiptir (Kraemer & Andrews, 1982). Ancak, Helena Chmura (1982), Kraemer ve Kupfer (2006), Cohen's *d* ölçütünde olduğu gibi, Glass delta'nın da müdahalenin klinik açıdan anlamlı olup olmadığını doğrudan ortaya koyamadığını vurgulamışlardır. Yani bu ölçüt, tedavinin hasta üzerindeki etkisini doğrudan yansıtan klinik yorumlar sunmakta yetersiz kalabilir.

3. Hedge g Etki Büyüklüğü Ölçütü

Hedges' *g*, etki büyüklüğünü ölçmek için kullanılan istatistiksel bir göstergedir. Yapı olarak Cohen's *d* ile büyük ölçüde benzerlik gösterir; ancak küçük örneklem büyüklüklerinde (özellikle n <20) daha doğru sonuçlar verdiği için *düzeltilmiş etki büyüklüğü* olarak da bilinir. Bu düzeltme, küçük örneklem kaynaklı yanlılığı azaltmayı amaçlar. Hedges' *g* değeri, matematiksel olarak Eşitlik-11'de gösterilen formül aracılığıyla hesaplanır.

$$g = \frac{\mu_1 - \mu_2}{MS_{within}} \qquad (11)$$

Burada; $ss_{pooled} = MS_{within}$, μ_1: Deney grubu ortalaması, μ_2: Kontrol grubu ortalamasını ifade etmektedir.

Hedges' *g*, Cohen's *d* ve Glass's delta gibi etki büyüklüğü ölçütlerindeki yanlılığı düzeltmek amacıyla geliştirilmiştir. Bu düzeltme, ilgili tahmin edicinin bir düzeltme katsayısıyla çarpılması yoluyla yapılır (Hedges, 1981). Örneklem büyüklükleri eşit olduğunda ve *g*, Cohen's *d*'den türetildiğinde,

Hedges' *g* popülasyon etki büyüklüğünün yansız, en küçük varyanslı ve en uygun tahminini sağlar (UMVUE - Uniformly Minimum Variance Unbiased Estimator). Ancak iki grup varyans bakımından eşit değilse, Hedges' *g* hâlâ Cohen's *d*'ye göre daha başarılı sonuçlar verir ve örneklem büyüklüğüne daha az duyarlıdır (Peng & Chen, 2014). Bununla birlikte, Kraemer ve Kupfer (2006), Hedges' *g*'nin de tıpkı Cohen's *d* ve Glass's delta gibi, tedavinin klinik anlamını tek başına açıklamakta yetersiz kaldığını belirtmiştir. Yani bu ölçütler, bir müdahalenin hasta üzerindeki etkisinin pratik veya klinik önemini doğrudan yansıtmaz.

Cohen's d, Glass's delta ve Hedges' g Etki Büyüklüğü Ölçümleri İçin Yorumlama Kriterleri ve Referans Aralıkları

Etki büyüklüğünün yüzdelik (persentil) değerlere dönüştürülebilmesi önemli bir özelliktir. Cohen's *d*, bu tür dönüşümlere olanak tanıyan etki büyüklüğü ölçütlerinden biridir. Cohen (1988), *d* değerlerini genel olarak şu şekilde sınıflandırmıştır: küçük (d = 0.2), orta (d = 0.5) ve büyük (d ≥ 0.8). Ancak bu sınıflandırma, ölçüm aracının güvenilirliği veya çalışma grubundaki bireylerin çeşitliliği gibi faktörleri dikkate almaz. İki grup arasındaki etki büyüklüğü, aynı zamanda bir grubun diğerine göre ortalama yüzdelik konumunu veya iki grubun dağılımlarının ne kadar örtüştüğünü gösterebilir. Örneğin, *d = 0* olduğunda, ikinci grubun ortalaması birinci grubun 50. yüzdeliğine denk gelir; bu da iki dağılımın tamamen örtüştüğü ve fark olmadığı anlamına gelir. Eğer *d = 0.8* ise, bu durumda ikinci grubun ortalaması, birinci grubun dağılımında yaklaşık 79. persentile denk gelir. Yani grup 2'den rastgele seçilen bir birey, grup 1'deki bireylerin %79'undan daha yüksek bir değere sahip olacaktır. Bu etki büyüklüğü düzeyinde iki dağılımın yaklaşık %53'ü örtüşmektedir (Sullivan & Feinn, 2012).

Tablo 1. Etki Büyüklüğünün Yorumlanması

Yorum	Etki Büyüklüğü	Yüzdelik (Percentil)	Örtüşmeme Yüzdesi
	0	50	0
Küçük	0.2	58	15
Orta	0.5	69	33
Büyük	0.8	79	47
	1	84	55
	1.5	93	71
	2	97	81

Etki büyüklüğünü yorumlamanın alternatif yollarından biri, standartlaştırılmış ortalama fark (*d*) ile korelasyon katsayısı (*r*) arasında

ilişki kurmaktır. Eğer grup üyeliği bir kukla (dummy) değişken olarak tanımlanırsa, örneğin: kontrol grubu için 0, deney grubu için 1 atanırsa, bu değişken ile sonuç değişkeni arasındaki korelasyon değeri *r* olarak hesaplanabilir. Bu şekilde elde edilen *r* değeri, standartlaştırılmış ortalama fark olan *d* ile matematiksel olarak ilişkilidir ve genellikle Eşitlik-12 kullanılarak *d* değerine dönüştürülebilir.

$$r^2 = \frac{d^2}{d^2 + 4} \qquad (12)$$

Bu formül, A ve B gruplarının eşit örneklem büyüklüklerine sahip olduğu durumlar için uygundur. Ancak örneklem büyüklükleri eşit değilse, bu dengesizlik korelasyon katsayısı (*r*) ile ilişki derecesi değerlendirilirken mutlaka dikkate alınmalıdır. Böyle durumlarda, *r* hesaplamak için daha genel bir formül kullanılması gerekir.

$$r = \frac{d}{\sqrt{d^2 + (\frac{1}{pq})}} \qquad (13)$$

Burada *p* ve *q* A ve B örneklemlerinde; *p:* A'nın oranı, *q:* B'nin oranı olarak ifade edilmektedir.

Bağımsız İki Grup İçin Parametrik Olmayan Etki Büyüklüğü Ölçütleri

Cohen's *d*, Hedges' *g* ve Glass's delta gibi etki büyüklüğü ölçütleri, normal dağılım varsayımı altında kullanılan parametrik yöntemlerdir. Ancak gerçek dünyadaki birçok durumda popülasyonlar normal dağılmayabilir ve sadece ortalama (konum) değil, aynı zamanda varyans (ölçek) ve şekil gibi farklı özellikler de önem kazanabilir. Bu tür durumlarda, normal dağılım varsayımı gerektirmeyen, yani parametrik olmayan etki büyüklüğü ölçüleri tercih edilmelidir. İki bağımsız grubun karşılaştırıldığı analizlerde kullanılabilecek başlıca parametrik olmayan etki büyüklüğü ölçütleri şunlardır: Cliff's delta, Glass Rank Biserial Korelasyon Katsayısı ve Vargha-Delaney A (VDA). Bu çalışmada, söz konusu üç yöntem detaylı olarak ele alınacaktır.

1. Cliff's Delta Etki Büyüklüğü Ölçütü

İki bağımsız grup için kullanılan parametrik olmayan etki büyüklüğü ölçütleri, grup ortalamalarına dayanmaz (Cliff, 1993). Bu tür yöntemler, verilerin değerleri yerine sıralamalarını dikkate alır (Hess & Kromrey, 2004). Cliff's delta, özellikle çarpık dağılımların bulunduğu durumlarda ya da likert tipi ölçeklerle toplanan verilerin analizinde, Cohen's *d*'ye kıyasla daha sağlam ve güvenilir bir etki büyüklüğü ölçütü olarak öne çıkar. Bu yöntemin hesaplanma şekli Eşitlik-14'te gösterilmektedir.

$$Cliff\ Delta = \frac{\#(x_1 > x_2) - \#(x_1 < x_2)}{n_1 n_2} \qquad (14)$$

$x_1\ ve\ x_2$: grup 1 ve grup 2 içindeki puanlar, $n_1\ ve\ n_2$: örneklem büyüklüğü, #: sayma sembolüdür.

Cliff's delta, iki bağımsız gruptan rastgele seçilen gözlemler arasında hangisinin daha büyük olma olasılığını tahmin eden bir ölçüttür. Başka bir deyişle, bir grubun değerlerinin diğer gruba göre ne ölçüde üstün olduğunu yansıtır. Cliff (1993), bu ölçütü iki dağılım arasındaki örtüşme düzeyini yansıtan bir gösterge olarak tanımlamıştır. Cliff's delta değeri her zaman -1 ile +1 arasında değişir. +1 ya da -1 değeri, gruplar arasında hiçbir örtüşme olmadığını; 0 değeri ise her iki grubun dağılımlarının tamamen örtüştüğünü ifade eder. Bu yöntem, gözlem değerlerinin vektör olarak ele alınmasına olanak tanır. Eşitlik-14'te tanımlandığı şekilde, iki grup arasındaki tüm olası karşılaştırmalar, matris biçiminde değerlendirilerek, cebirsel işlemlerle Cliff's delta değeri hesaplanabilir.

Grup 1 ve Grup 2, $\mathfrak{R}$ uzayındaki vektörler, $\boldsymbol{m}$ ve $\boldsymbol{n}$ boyutlar olmak üzere, Grup 1 $\in \mathfrak{R}^{\boldsymbol{m}}$ ve Grup 2 $\in \mathfrak{R}^{\boldsymbol{n}}$, $\forall_{\boldsymbol{n}} \in N$ olsun. N. Cliff (1993) tarafından önerilen ve $\boldsymbol{\delta} \in \mathfrak{R}^{\boldsymbol{mxn}}$ olarak ifade edilen matris eşitlik-15'te verilen $\boldsymbol{\delta}$: $(\mathfrak{R}^{\boldsymbol{m}}, \mathfrak{R}^{\boldsymbol{n}}) \rightarrow \mathfrak{R}^{\boldsymbol{mxn}}$ fonksiyonu ile elde edilebilir

$$\boldsymbol{\delta_{ij}} = \begin{cases} +1 \rightarrow Grup1_i > Grup2_j, \forall_i, \forall_j \\ -1 \rightarrow Grup1_i < Grup2_j, \forall_i, \forall_j \\ 0 \rightarrow Grup1_i = Grup2_j, \forall_i, \forall_j \end{cases} \qquad (15)$$

Böylece Eşitlik-15 kullanılarak $\boldsymbol{\delta_{ij}}$'deki her bir eleman için +1, -1 ve 0 olmak üzere yalnızca üç olası değere sahip *m* satır ve *n* sütun matrisi oluşturulur. Bu değerler Eşitlik-15'teki kurala göre *i*.nci satır *j*.nci sütundaki $\boldsymbol{\delta_{ij}}$ matrisine atanır. Eğer grup 1'deki *i*. değer grup 2'deki *j.* değerden büyük ise $\boldsymbol{\delta_{ij}} = +1$, grup 1'deki *i*. değer grup 2'deki *j.* değerden küçük ise $\boldsymbol{\delta_{ij}} = -1$ ve grup 1'deki *i*. değer grup 2'deki *j.* değere eşit ise $\boldsymbol{\delta_{ij}} = 0$ olur. Cliff delta değeri, satırlar veya sütunlar için sırasıyla Eşitlik-16 ve Eşitlik-17'de gösterildiği gibi elde edilebilir.

$$Satırlar\ için\ Delta = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\boldsymbol{\delta_{ij}} \qquad (16)$$

$$Sütunlar\ için\ Delta = \frac{1}{mn}\sum_{j=1}^{n}\sum_{i=1}^{m}\boldsymbol{\delta_{ij}} \qquad (17)$$

Bu hesaplama sonucunda elde edilen değer, Eşitlik-14'te gösterildiği üzere Cliff's delta etki büyüklüğü olarak tanımlanır (Cliff, 1993). Cliff delta için yorumlamada rehber olarak kabul edilen sınır değerler şunlardır: 0.11 küçük, 0.28 orta ve 0.43 büyük etki büyüklüğüne karşılık gelir (Peng & Chen, 2014).

2. Vargha & Delaney A (Vda) Etki Büyüklüğü Ölçütü

VDA (Vargha-Delaney A) ölçütü, grup 1'den rastgele seçilen bir gözlemin, grup 2'den rastgele seçilen bir gözlemden daha büyük olma olasılığına dayanır. Buna ek olarak, iki grubun eşit puanlara sahip olma durumunun yarısı da hesaba katılır. Matematiksel olarak bu olasılık şu şekilde ifade edilir: $\boldsymbol{P(X_1 > X_2) + 0.5\,P(X_1 = X_2)}$ (Vargha & Delaney, 2000). VDA, popülasyon düzeyinde yansız bir etki büyüklüğü tahminidir ve aynı zamanda Cliff's delta'nın doğrusal bir dönüşümüdür. Bu özelliğiyle, özellikle klinik anlamlılık açısından yorumlanabilirliği yüksek bir ölçüttür (Kramer & Kupfer, 2006). Ayrıca, ROC eğrisinin altında kalan alanla ilişkili olması da VDA'nın dikkat çeken bir başka yönüdür. VDA, yalnızca iki grup arasındaki farkları değerlendirmekle kalmaz; aynı zamanda birden fazla grup arasındaki stokastik benzerlik ya da farklılıkları ve ilişkili ölçümler arasında stokastik eşitliği incelemek için de kullanılabilir (Vargha & Delaney, 2000). Peng & Chen (2014) tarafından aktarıldığı üzere, VDA ölçütü için önerilen yorumlama sınırları şu şekildedir: 0.56 küçük, 0.64 orta ve 0.71 büyük etki büyüklüğüne karşılık gelir.

McGraw ve Wong (1992) tarafından geliştirilen "Ortak Dil İstatistiği" (*Common Language Statistics – CL*), etki büyüklüğünü yorumlamayı kolaylaştırmak amacıyla bir olasılık değeri olarak ifade eden bir ölçüttür. Sürekli dağılımlar söz konusu olduğunda, CL; bir gruptan rastgele seçilen bir skorun, diğer gruptan seçilen bir skordan daha yüksek olma olasılığı olarak tanımlanır.Bu ölçüt, istatistiksel sonuçların daha anlaşılır hale getirilmesini sağlar ve olasılık teorisine dayanarak Eşitlik-18'de gösterilen formülle hesaplanır.

$$\boldsymbol{CL = P(X_1 > X_2)} \qquad \textbf{(18)}$$

Burada $\boldsymbol{X_1}$ ve $\boldsymbol{X_2}$, sırasıyla birinci ve ikinci popülasyondan rastgele seçilen gözlem değerlerini temsil eder. McGraw ve Wong (1992), çalışmalarında Ortak Dil İstatistiği (CL) ile ilgili şu üç önemli noktaya değinmiştir:

1. Eğer veriler normal dağılım gösteriyor ve her iki grubun varyansları eşitse ($\boldsymbol{\sigma_1 = \sigma_2}$), bu durumda CL için doğrudan bir nokta tahmini yapılabilir.
2. CL, sadece normal dağılımlarla sınırlı değildir; normal olmayan dağılımlarda da kullanılabilir. Ayrıca, normallik ve varyans homojenliği varsayımlarının ihlali durumunda da sağlam (robust) bir ölçüttür.
3. CL, çoklu grup karşılaştırmaları, ilişkili örneklemler ve kesikli veri yapıları gibi daha karmaşık durumlara da genelleştirilmiştir.

Vargha ve Delaney (2000) tarafından geliştirilen VDA ölçütü, CL'nin bu genelleştirilmiş biçimidir ve stokastik üstünlük ölçütü olarak adlandırılır. VDA'nın hesaplanma biçimi Eşitlik-19'da sunulmuştur.

$$VDA = \frac{\#(x_1 > x_2) + 0.5\#(x_1 = x_2)}{n_1 n_2} \qquad (19)$$

$x_1\ ve\ x_2$: grup 1 ve grup 2 içindeki puanlar, $n_1\ ve\ n_2$: örneklem büyüklüğü, #: sayma sembolüdür.

VDA ölçütü için $\boldsymbol{VDA_{12} = 1 - VDA_{21}}$ eşitliği her zaman geçerlidir. Ancak bunun tersi her zaman geçerli değildir. Eğer bir değişken X, her iki popülasyonda da aynı dağılıma sahipse, bu durumda $\boldsymbol{VDA_{12} = VDA_{21} = 0.5}$ olur. Bu da her iki grubun da diğerine üstünlük sağlamadığını, yani bir

grubun diğerine göre sistematik olarak daha yüksek değerlere sahip olmadığını gösterir. Dolayısıyla bu durum, iki grubun stokastik olarak eşit olduğu, yani dağılım yapılarının birbirinden anlamlı biçimde farklı olmadığı anlamına gelir.

3. Rank-Biserial Korelasyon Katsayısı Ölçütü

Parametrik olmayan dağılımlarda *ortak dil etki büyüklüğü* sıkça kullanılsa da, bu tür analizlerde en yaygın başvurulan ölçütlerden biri de korelasyon katsayısıdır. Mann-Whitney U testi için önerilen etki büyüklüğü ölçütü, rank-biserial korelasyon katsayısıdır. Mann-Whitney testi bağlamında bu korelasyon katsayısını hesaplamak için üç farklı yöntem geliştirilmiştir. Bu yöntemlerden biri, Cureton (1956) tarafından önerilmiştir. Cureton'un yaklaşımında, bir grubun sıralamalarının diğer gruba kıyasla daha yüksek olması temel alınır. Bu yöntem, "uygun çift" ve "uygun olmayan çift" kavramlarına dayanır. Uygun çiftlerin sayısı *P*, uygun olmayan çiftlerin sayısı ise *Q* ile gösterilir. Etki büyüklüğü, bu iki değerin farkının, alınabilecek maksimum fark değerine bölünmesiyle hesaplanır. Böylece gruplar arası farkın derecesi sıralamalara dayalı olarak ölçülmüş olur.

$$r_{rb} = \frac{(P - Q)}{P_{max}} \qquad (20)$$

Rank-biserial korelasyon katsayısına dayalı etki büyüklüğü için önerilen ikinci formül, Glass (1965) tarafından geliştirilmiştir. Bu formül, özellikle madde analizine yönelik çalışmalarda, ölçek geliştirme sürecinde kullanılmıştır. Glass'ın amacı, $\boldsymbol{r_{rb}}$ (rank-biserial korelasyon) katsayısının Pearson korelasyonu gibi, sıralamalı veriler için Spearman korelasyonunu da tahmin edebilmesini sağlamaktır. Bu doğrultuda, Glass (1965), daha önce Cureton (1956) tarafından önerilen formüle denk bir alternatif sunmuştur. Ancak bu formül, özellikle bilgi düzeyini ölçen testler için daha uygundur. Çünkü bir soruya doğru yanıt veren bireylerin, toplam test puanları açısından yanlış yanıt verenlere göre daha yüksek sıraya sahip olması durumunda, rank-biserial korelasyon değeri de yüksek çıkmaktadır. Glass'ın önerdiği bu korelasyon katsayısına dayalı etki büyüklüğü ölçütü, Eşitlik-21'de gösterilen formülle hesaplanmaktadır.

$$r_{rb} = \frac{2(\bar{Y}_1 - \bar{Y}_0)}{N} \qquad (21)$$

Burada; $\bar{Y}_1$: Teste doğru cevap veren kişilerin ortalama rankları, $\bar{Y}_0$: Teste yanlış cevap veren kişilerin ortalama rankları ve *N*: Toplam kişi sayısını ifade etmektedir.

Wendt (1972), Mann-Whitney U testi için etki büyüklüğünü hesaplamaya yönelik üçüncü bir formül önermiştir. Bu formülün geliştirilmesindeki temel amaç, araştırmacıların Mann-Whitney testi sonrası etki büyüklüğünü daha kolay bir şekilde raporlayabilmelerini sağlamaktır. Wendt'in önerdiği yöntem, doğrudan *U* istatistiği ile iki grubun örneklem büyüklüklerini kullanarak *rank-biserial korelasyon katsayısını* hesaplamaya olanak tanır. Bu basitleştirilmiş yaklaşım, Eşitlik-22'de gösterildiği şekilde ifade edilmektedir.

$$r_{rb} = \frac{1 - (2U)}{n_1 \times n_2} \qquad (22)$$

U istatistiği sıfır olduğunda, *rank-biserial korelasyon* katsayısı ($\mathbf{r_{rb}}$) en yüksek değeri olan 1'e ulaşır. Ancak U istatistiği yönsüz bir ölçü olduğundan, Wendt (1972) formülüyle hesaplanan $\mathbf{r_{rb}}$ değeri de yön bilgisi içermez ve her zaman pozitif bir değer olarak elde edilir. Bu nedenle, bu formülle hesaplanan korelasyon katsayısı sadece ilişkinin büyüklüğünü gösterir, yönünü yansıtmaz.

Kerby (2014), Mann-Whitney U testi için dördüncü bir etki büyüklüğü ölçütü önermiştir. Bu yeni yaklaşım, Cureton'un (1956) geliştirdiği formülün yarıya bölünmesiyle elde edilmiştir. Böylece Kerby, daha basit ve doğrudan yorumlanabilir bir etki büyüklüğü formülü sunmayı amaçlamıştır.

$$r_{rb} = \left(\frac{P}{P_{max}}\right) - \left(\frac{Q}{P_{max}}\right) \qquad (23)$$

Eşitlik-23'te yer alan formülde, ilk oran uygun çiftlerin oranı (*f*), ikinci oran ise uygun olmayan çiftlerin oranı (*u*) olarak tanımlanır. Bu formül, basitçe $\mathbf{r_{rb}}$= f-u şeklinde ifade edilir. Bu basit fark formülünün birkaç önemli avantajı bulunmaktadır:

1. Etki büyüklüğünü, kolay anlaşılır bir ölçü olan ortak dil etki büyüklüğü perspektifinden sunar. Bu sayede sonuçların yorumu daha sezgisel hale gelir.

2. İşaretin yönü, yorum açısından anlamlıdır. Pozitif bir rrbr_{rb}rrb değeri, verilerin beklenen yönde olduğunu; negatif bir değer ise verilerin tahmin edilenin tersine hareket ettiğini gösterir.
3. Formül oldukça kolay yorumlanabilir bir yapıdadır. Kullanıcıya etkili ve hızlı bir şekilde sonuç hakkında bilgi sunar.

Rank-biserial korelasyon katsayısı için önerilen yorumlama aralıkları ise şu şekildedir:

- $\mathbf{0.1 \leq r_{rb} < 0.30}$ → küçük etki büyüklüğü
- $\mathbf{0.30 \leq r_{rb} < 0.50}$→ orta etki büyüklüğü
- $\mathbf{r_{rb} \geq 0.50}$ → büyük etki büyüklüğü

Etki Büyüklüğü Ölçütlerinin Birbirine Dönüştürülmesi

Bağımsız gruplar, eşleştirilmiş gruplar ya da diğer farklı araştırma tasarımları kullanılsa da aynı etki büyüklüğü türü hesaplanabilir. Etki büyüklüğü kavramı tüm bu tasarımlarda aynı anlamı taşıdığı için, farklı çalışmalardan elde edilen etki büyüklüğü tahminlerinin birleştirilmesinde herhangi bir sakınca bulunmaz. McGraw ve Wong (1992) tarafından önerilen bu yaklaşım, *Ortak Dil Etki Büyüklüğü (Common Language Effect Size – CL)* olarak adlandırılmıştır. CL ölçütü, diğer etki büyüklüğü istatistiklerine kıyasla daha kolay yorumlanabilir; çünkü farkın büyüklüğünü bir olasılık değeriyle ifade eder. Daha spesifik olarak CL, deney grubundan rastgele seçilen bir bireyin, kontrol grubundan rastgele seçilen bir bireyden daha yüksek puan alma olasılığını tahmin eder. Bu olasılık, z-istatistiği kullanılarak hesaplanır ve bu hesaplama Eşitlik-24'te gösterilmiştir.

$$z = \frac{\left|\overline{Y_e} - \overline{Y_c}\right|}{\sqrt{S_e^2 + S_c^2}} \qquad (24)$$

Burada; $\overline{Y_e}$: Deney grubunun ortalaması, $\overline{Y_c}$: Kontrol grubunun ortalaması, S_e: Deney grubunun standart sapması ve S_c: Kontrol grubunun standart sapmasını ifade etmektedir.

Moore, McCabe ve Craig (2016) tarafından verilen örnekte, deney grubunun ortalaması 51.476 ve standart sapması 11.007, kontrol grubunun ortalaması ise 39.545 ve standart sapması 14.628 olarak verilmiştir. Bu verilere göre, iki grup arasındaki farkı değerlendirmek için hesaplanan *z-istatistiği* şu şekilde elde edilir:

$$z = \frac{51.476 - 39.545}{\sqrt{11.007^2 + 14.628^2}} = 0.652 \; ve \; p(z < 0.652) = 0.743$$

Bu sonuca göre, deney grubundan rastlantısal olarak seçilen bir bireyin, kontrol grubundan rastlantısal olarak seçilen bir bireyden daha yüksek bir değere sahip olma olasılığı %74,3'tür. Bu tür bir yorum, etki büyüklüğünün bir olasılık ifadesine dönüştürülmesiyle elde edilir. Aynı yaklaşım, daha genel ve karşılaştırılabilir bir yorum sağlamak amacıyla, farklı etki büyüklüğü ölçütlerinin *Cohen's d* gibi standart bir ölçüte dönüştürülmesiyle de uygulanabilir. Bu sayede, farklı çalışmalar ve yöntemler arasında etki büyüklüğü daha kolay karşılaştırılabilir hale gelir.

Cohen d değerinin Rank-Biserial Korelasyon Katsayısı (r_{rb}) değerine dönüşümü: Eşitlik-25 kullanılarak korelasyon katsayısından (r_{rb}), standartlaştırılmış bir ortalama fark (d)'a dönüşüm yapılabilir.

$$r_{rb} = \frac{d}{\sqrt{d^2 + a}} \qquad (25)$$

Burada;

$n_1 \neq n_2$ Olduğunda a bir düzeltme faktörüdür. a, n_1 ile n_2 oranına bağlıdır.

$$a = \frac{(n_1 + n_2)^2}{n_1 n_2} \qquad (26)$$

Rank-Biserial Korelasyon Katsayısı (r_{rb}) değerinin Cohen d değerine dönüşümü: Standartlaştırılmış bir ortalama fark (d)'den korelasyon katsayısı (r_{rb})'na, dönüşüm Eşitlik-27 kullanılarak yapılabilmektedir.

$$r_{rb} = \frac{2r}{\sqrt{1 - r^2}} \qquad (27)$$

Cliff Delta (δ) değerinin Cohen d değerine dönüşümü: Cliff delta (δ) etki büyüklüğü tahminini, Cohen d etki büyüklüğü tahminine, Φ^{-1} normal kümülatif dağılım fonksiyonun tersi olmak üzere Eşitlik-28 kullanılmaktadır.

$$d(\delta) = 2z\frac{-1}{\delta - 2}\ , \qquad z_p \equiv \Phi^{-1}(p) = AUC^{-1}(p) \tag{28}$$

VDA ile Cliff delta (δ) etki büyüklüğü arasında doğrusal bir ilişki olduğundan $VDA = \frac{(Cliff\ Delta+1)}{2}$ şeklinde dönüştürülmektedir.

SONUÇ

Bilimsel araştırmalarda birçok araştırmacı, elde edilen p-değerinin kabul edilen anlamlılık düzeyinden (genellikle $\alpha = 0.05$) küçük olması durumunda sonucu "istatistiksel olarak anlamlı" kabul etmektedir. Ancak p-değeri yalnızca gözlemlenen farkın ya da ilişkinin sıfır hipotezi altında şansa bağlı oluşma olasılığını ifade eder; bu nedenle, bulgunun büyüklüğü ya da pratik/klinik önemi hakkında doğrudan bilgi sunmaz.

Etki büyüklüğü (effect size) ise bu boşluğu dolduran ve müdahalenin, uygulamanın ya da ilişkinin büyüklüğüne ilişkin daha kapsamlı ve yorumlanabilir bir çerçeve sunan istatistiksel bir ölçüdür. p-değerinden farklı olarak etki büyüklüğü, çalışmanın örneklem büyüklüğünden bağımsız olarak etki düzeyine ilişkin daha doğrudan bir değerlendirme yapılmasına olanak tanır. Bu nedenle, yalnızca istatistiksel anlamlılığa değil, etkinin büyüklüğüne ve olası pratik önemine odaklanan analizlerde etki büyüklüğü kritik bir rol oynar.

Kullanılacak etki büyüklüğü ölçütü;

- Gruplar arası ilişkinin yapısına (bağımsız ya da eşleştirilmiş/ilişkili),
- Verinin ölçüm düzeyine ve dağılım özelliklerine (özellikle sürekli veriler için normal dağılım varsayımı),
- Araştırma tasarımına (örneğin iki grup mu yoksa çoklu grup karşılaştırması mı yapıldığına) bağlı olarak seçilmelidir.

Sonuç olarak, etki büyüklüğü, istatistiksel anlamlılık testlerinin tamamlayıcısı niteliğindedir ve çalışmanın bilimsel geçerliliğini, pratik anlamını ve yorum gücünü artırmak için mutlaka rapor edilmelidir.

Kaynaklar

Kitap

Carman, R. A. (1969). Numbers and units for physics: John Wiley & Sons.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences. 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cox, D. R., & Snell, E. J. (1981). Applied statistics-principles and examples (Vol. 2): CRC Press.

Ellis, B. (1968). Basic concepts of measurement. Cambridge University Press

Fisher, R. (1935). Design Of Experiments (Hafner, New York, NY).

Glass, G. V., Smith, M. L., & McGaw, B. (1981). Meta-analysis in social research: Sage Publications, Incorporated.

Grissom, R. J., & Kim, J. J. (2005). Effect sizes for research: A broad practical approach: Lawrence Erlbaum Associates Publishers.

Grissom, R. J., & Kim, J. J. (2012). Effect sizes for research: Univariate and multivariate applications: Routledge.

Ipsen, D. C. (1960). Units, dimensions, and dimensionless numbers: McGraw-Hill.

Moore, D., McCabe, G., & Craig, B. (2016). Introduction to the practice of statistics. San Francisco: W. H. In: Freeman.

Dergi Makalesi

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. Psychological bulletin, 114(3), 494.

Coe, R. (2002). It's the effect size, stupid: What effect size is and why it is important. British Educational Research Association Annual conference, https://f.hubspotusercontent30.net/hubfs/5191137/attachments/ebe/ESguide.pdf

Cohen, J. (1992). Things I have learned (so far). Paper presented at the Annual Convention of the American Psychological Association, 98th, Aug, 1990, Boston, MA, US; Presented at the aforementioned conference.

Cureton, E. E. (1956). Rank-biserial correlation. *Psychometrika*, 21(3), 287-290.

Ellis, P. (2009). Thresholds for interpreting effect sizes. Retrieved January, 13, 2014.

Glass, G. V. (1965). A ranking variable analogue of biserial correlation: Implications for short-cut item analysis. *Journal of Educational Measurement*, 2(1), 91-95.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3-8.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107-128.

Hedges, L. V., & Olkin, I. (1984). Nonparametric estimators of effect size in meta-analysis. *Psychological bulletin*, 96(3), 573.

Henson, R. K. (2006). Effect-size measures and meta-analytic thinking in counseling psychology research. *The Counseling Psychologist*, 34(5), 601-629.

Hess, M. R., & Kromrey, J. D. (2004). Robust confidence intervals for effect sizes: A comparative study of Cohen'sd and Cliff's delta under non-normality and heterogeneous variances. Paper presented at the annual meeting of the American Educational Research Association.

Kazis, L. E., Anderson, J. J., & Meenan, R. F. (1989). Effect sizes for interpreting changes in health status. *Medical care*, S178-S189.

Kramer, S. H., & Rosenthal, R. (1999). Effect sizes and significance levels in small-sample research. *Statistical strategies for small sample research*, 59-79.

Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. *Comprehensive Psychology*, 3, 11.

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, 111(2), 361.

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, 82(4), 591-605.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological methods*, 8(4), 434.

Peng, C.-Y. J., & Chen, L.-T. (2014). Beyond Cohen's d: Alternative effect size measures for between-subject designs. *The Journal of Experimental Education*, 82(1), 22-50.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychological methods*, 16(2), 93.

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 8(2), 26.

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of Graduate Medical Education*, 4(3), 279-282.

Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, 25(2), 101-132.

Valentine, J. C., & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The Study Design and Implementation Assessment Device (Study DIAD*). Psychological methods*, 13(2), 130.

Wendt, H. W. (1972). Dealing with a common problem in Social science: A simplified rank-biserial coefficient of correlation based on the U statistic. *European Journal of Social Psychology*, 2(4), 463-465.

META ANALİZİNDE YANLILIK DEĞERLENDİRME YÖNTEMLERİ

Fisun Kaşkır Kesin[**], İlker Ercan[***]

Giriş

Meta analizi, bireysel çalışmaların çelişkili sonuçlar vermesi ya da hiç sonuç vermemesi, aynı hipotezleri içeren birden fazla çalışma yapılarak ortak bir sonuca ihtiyaç duyulması, örnekleme alınacak birim sayısının az olmasına sebep olan zaman, maliyet, az görünen bir özelliğin incelenmesi gibi kısıtlardan dolayı birim sayısı az olan örneklemlerle çalışılmak zorunda kalınarak gereken güce ulaşılamaması gibi durumlarda uygulanır. Meta analizi, çalışma sonuçlarını istatistiksel olarak birleştirerek, evren parametresine ilişkin özetlenmiş tek bir etki büyüklüğü elde etmeyi sağlar.

Meta analizine dahil edilecek çalışmaların bulunduğu örneklemin, ilgili çalışmaların evrenini temsil etmesi gerekir. Bu nedenle meta analizinde doğru ve güvenilir sonuçlar elde etmenin ilk adımı olarak meta analizine dahil edilecek çalışmalar sistemli ve dikkatlice seçilmeli ve bir yanlılığın oluşmasının önüne geçilmelidir.

Meta Analizinde Yanlılık

Meta analizi ilgili tüm çalışmaların sistematik incelemeleri sonucu, matematiksel olarak bu çalışmaların sentezidir (Sterne, & Harbord, 2004). Meta analizine katılan çalışmalar ilgili literatürün yanlı bir örneklemi olduğunda, meta analizi sonucu elde edilen ortalama etkide de bu yanlılık görülür (Borenstein ve ark., 2009). Bu durum, meta analizinin ne kadar sistematik ve kapsamlı olduğuna bakılmaksızın sonuçların geçerliliğini tehdit eder (Rothstein ve ark., 2005). Bu nedenle meta analizine dahil edilecek çalışmalar mümkün olduğunca tarafsız ve eksiksiz seçilmelidir.

Meta analizinde ortaya çıkabilecek bir yanlılık durumu özellikle sağlıkla ilgili araştırmalarda hastaların zarar görebileceği ciddi sonuçlara yol açabilir. Ayrıca, bir çalışmanın yayınlanmaması biçiminde ortaya çıkan yanlılık, çalışmanın başından sonuna kullanılan tüm kaynakların boşa harcanmasına ve etik problemlere de neden olur (Rothstein ve ark., 2005).

[**] Dr., Düzce Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Mülkiyeti Koruma ve Güvenlik Bölümü Sosyal Güvenlik Programı, Düzce, kaskir.fisun@gmail.com, ORCID: orcid.org/0000-0002-6524-2997
[***] Prof. Dr., Bursa Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Bursa, iercan@msn.com, ORCID: orcid.org/0000-0002-2382-290X

Araştırmacıların yayınlanmak üzere dergiye gönderdiği bütün kaliteli çalışmalar istatistiksel olarak anlamlılığa bakılmaksızın editörlerce kabul edildiğinde, ya da sonuçların yayınlanmasındaki gecikmeler giderildiğinde yanlılık sorunu çoğunlukla engellenebilir (Whitehead, 2002). Meta analizinde meydana gelebilecek yanlılığı engelleyebilmek için önerilen stratejilerden biri de herhangi bir çalışmaya başlamadan ya da sonuçları elde etmeden önce, çalışmaların kaydedilerek, daha sonra meta analizlerinde kullanılmak üzere yansız bir örnekleme çerçevesi oluşturmaktır (Rothstein ve ark., 2005). Çalışmalara başlanmadan önce kaydedilmesiyle, çalışmaların ne zaman yapıldığı bilinebilecek, çalışma sonuçları takip edilebilecek, gerekli durumlarda sorumlu araştırmacıyla iletişim kurulabilecektir. Kayıtların farklı biçimlerde ortaya çıkabilecek yanlılıkları önleyebilme potansiyeli de vardır. Kayıtlar sayesinde devam eden çalışmalar takip edilebileceğinden gereksiz yapılacak çalışma tekrarları ve bu tekrarların neden olacağı yanlılıklar önlenmiş olur. Ayrıca, çalışma sonuçları ortaya çıkmadan önce kaydedileceğinden, sonuçların yönüne ve büyüklüğüne göre seçilerek kullanılmasıyla ortaya çıkabilecek yanlılıklar da önlenmiş olur. İdeal olarak istenen tam bir kayıt listesi olsa da şart değildir.

Meta analizi sürecinde güvenilir testlerle yanlılık saptandığında ya da böyle bir durumdan şüphe duyulduğunda, literatürde bazı öneriler yer almaktadır. Sutton ve arkadaşları (2000), ilk olarak, analize dahil edilmemiş çalışmaların araştırılarak tekrar literatür taramasına dönülmesini önermektedir. İkinci yaklaşım, meta analizine devam edilmesi ancak elde edilen bulguların yorumlanmasında dikkatli olunarak olası yanlılıkların göz önünde bulundurulmasıdır. Üçüncü seçenek ise, uygulanabilirlikleri ve geçerlilikleri zaman zaman tartışılsa da, yanlılığın etkilerini düzeltmeye yönelik istatistiksel yöntemlerin kullanılmasıdır. Bu düzeltici yöntemlerin, yalnızca bir tür duyarlılık analizi çerçevesinde değerlendirilmesi gerektiği vurgulanmakta; araştırmacıların, sonuçların farklılık gösterip göstermediğini incelemek amacıyla, hem düzeltilmiş hem de düzeltilmemiş analizleri birlikte yürütmeleri önerilmektedir (Sutton ve ark., 2000). Jin ve çalışma arkadaşları (2014) ise, yanlılık tespiti durumunda veya böyle bir durumdan şüphelenildiğinde, duyarlılık analizi amacıyla Copas seçim modelinin kullanılmasını önermektedir. Rothstein ve arkadaşları (2005), meta analizlerinde yanlılık kontrolünün mutlaka yapılması gerektiğini ve eğer yanlılık mevcutsa, eksik çalışmaların olası etkilerini değerlendirmek amacıyla duyarlılık analizlerinin uygulanmasının önemini vurgulamaktadır. Öte yandan, Wang ve Liu (2016) istatistiksel testler aracılığıyla yanlılık

saptandığında, özellikle küçük örneklem hacmine sahip çalışmaların bu yanlılığa neden olabileceğini ve bu nedenle söz konusu çalışmaların meta analizinden çıkarılmasının yerinde olacağını ifade etmişlerdir.

Meta Analizinde Yanlılığa Neden Olabilen Yanlılık Çeşitleri

Meta analizlerinde, elde edilen etki büyüklüklerinin gerçeği yansıtmamasına neden olan çeşitli yanlılık türleri literatürde tanımlanmıştır. Bu yanlılıklar arasında yayılma yanlılığı (dissemination bias), tam yayın yanlılığı (full publication bias), yayın yeri yanlılığı (place of publication bias), veri tabanı ya da indeksleme yanlılığı (database/indexing bias), medya ilgisi yanlılığı (media attention bias), ülke yanlılığı (country bias), atıf yanlılığı (citation bias), gri literatür yanlılığı (gray literature bias), zaman gecikmesi yanlılığı (time-lag bias), kesme yanlılığı (truncation bias), tekrarlama yanlılığı (duplication bias), sonuç yanlılığı (outcome bias), benzerlik yanlılığı (familiarity bias), maliyet yanlılığı (cost bias), ulaşılabilirlik yanlılığı (availability bias), yayın yanlılığı (publication bias) ve dil yanlılığı (language bias) yer almaktadır (Borenstein ve ark., 2009; Easterbrook et al., 1991; Egger et al., 1997; Ioannidis, 1998; Rothstein et al., 2005; Song et al., 2010; Stern & Simes, 1997; Toews et al., 2017).

Bu yanlılık türlerinin tamamı, meta analizi çalışmalarının dayandığı literatürün tüm araştırma evrenini temsil edememesi nedeniyle analiz sonuçlarının geçerliliğini benzer şekilde olumsuz etkiler (Rothstein ve ark., 2005). Söz konusu yanlılıklar için literatürde; "yayılma yanlılığı" (Song et al., 2010), "yayın yanlılığı" (Rothstein et al., 2005), "ulaşılabilirlik yanlılığı" ve "seçim yanlılığı" (Hunter & Schmidt, 2004) gibi farklı terimlere de yer verilmektedir. Ayrıca bu tür yanlılıklar topluca “yayın ve ilgili yanlılıklar” başlığı altında da ele alınmaktadır. Bu bölümde, anlatım birliğini sağlamak adına genel olarak "yanlılık" kavramı kullanılmıştır.

Meta analizinin ilk aşaması olan dahil edilecek çalışmaları arama süreci ile sonuçları etkileyebilecek yanlılıklar aşağıda açıklanmıştır.

Yayın yanlılığı

Yayın yanlılığı, bir çalışmanın yayınlanmasının ya da yayınlanmamasının çalışma sonuçlarının istatistiksel olarak anlamlı olup olmadığına ya da yönüne bağlı olarak değişmesidir (Rothstein ve ark., 2005). Eğer ilgili konu ile alakalı literatürde yer alan çalışmalar, tamamlanmış bütün çalışmaların evrenini temsil etmiyorsa yayın yanlılığından söz edilebilir. İstatistiksel olarak anlamlı olmayan sonuçlar, çalışılan konuya olan ilginin az olması,

örneklem hacmi sorunları, çalışmanın makaleye dönüştürülmesinin devam etmesi gibi nedenler tamamlanmış olan çalışmaların yayınlanmama nedenleri olabilir. Göreli olarak daha yüksek etki büyüklüğü raporlayan ya da istatistiksel olarak anlamlı sonuçlara ulaşan çalışmaların yayınlanma olasılığı daha fazladır. Yayınlanan çalışmaların meta analizine dahil edilme olasılıkları da yüksek olduğundan, literatürde yer alan herhangi bir yanlılığın meta analizi sonuçlarını etkilemesi olasıdır (Borenstein ve ark., 2009).

Yayılma yanlılığı

Eğer çalışmaların sonuçlarının yayılma profili, elde edilen sonuçların yönüne ya da gücüne bağlı ise yayılma yanlılığı söz konusu olur. Araştırma sonuçlarının ulaşılabilirliği ya da muhtemel kullanıcılar tarafından tanımlanması ihtimali yayılma profili olarak adlandırılır. Bir çalışmanın yayınlanma profili; yayınlandığı zaman, yer ve yöntemlerden başlayarak, erişilebilirlik düzeyine ve hatta yayınlanıp yayınlanmadığına kadar geniş bir yelpazeyi kapsar (Song ve ark., 2010).

Dil yanlılığı

Çalışmanın yayınlandığı dil çalışma sonuçlarının gücüne ve yönüne bağlı ise dil yanlılığı söz konusudur (Borenstein ve ark., 2009; Song ve ark., 2010). İngilizce dilindeki dergilerin ve veri tabanlarının yaygın kullanımı, İngilizce dilinde yayınlanan araştırmaların meta analizlerine dahil edilmesi ya da araştırmacıların bildiği dillere dayalı olarak literatür taraması yapılması gibi faktörler, dil yanlılıklarının ortaya çıkmasına neden olabilir. Bu durum, istatistiksel olarak anlamlı sonuçlar elde eden çalışmaların aşırı bir şekilde örneklenmesine yol açabilir (Rothstein ve ark., 2005). Önemli İngilizce dergilerde istatistiksel olarak anlamlı sonuçların, istatistiksel olarak anlamlı olmayan çalışmaların ise yazarların kendi dillerinde yayınlanması daha olasıdır. Meta analizine katılacak bütün çalışmaların yayınlandıkları dillerden bağımsız olarak aranması ve çalışmaya alınmaması, o çalışmanın dil yanlılığı içermemesinin en iyi yoludur.

Sonuç yanlılığı

Sonuç yanlılığı, araştırmacının birincil bulguları seçerek raporlaması, bulguların yönü ve istatistiksel olarak anlamlı olup olmamasına göre de diğer bulguları raporlamamasıdır. Yayınlanacak makalelerde dergi politikaları gereği bütün sonuçları verecek yer ve imkân olmadığında sonuçların mecburi olarak seçilmesi gerekir. Bu şekilde, en dikkat çekici, en

olumlu ve en önemli bulguların sunulması sonucu yanlılık ortaya çıkabilir (Rothstein ve ark., 2005; Song ve ark., 2010).

Tekrarlama yanlılığı

Benzer makalelerin ya da aynı verilerden elde edilen makalelerin birden fazla dergide yayınlanması tekrarlayan yayın olarak tanımlanır (Song ve ark., 2010). Tekrarlama yanlılığı ise çalışmaların birden fazla defa yayınlanması sonucu birden fazla defa meta analizine katılmasıyla ortaya çıkar (Rothstein ve ark, 2005). Tekrarlayan yayınlar, bir çalışmanın bir bölümü, hipotezleri, yöntemleri, sonuçları, tartışması gibi kısımlarını paylaşarak, uzun süreli çalışma sonuçlarını sunmak ya da başka bir kitleye göre uygun formatta hazırlanmış sonuçları sunmak amacıyla ortaya çıkabilir (Song ve ark., 2010). Tekrarlayan bu yayınlar yazarlar değiştiği, birim sayısı farklı verildiği ya da deneme tasarımı farklı tanımlandığı için fark edilemeyerek meta analizine dahil edilir ve yanlılık meydana gelir. İstatistiksel olarak anlamlı sonuçlar elde eden çalışmaların, anlamlı olmayanlara kıyasla daha fazla tekrar yayınlanma ve yüksek etki faktörlü dergilerde yer alma olasılığı bulunmaktadır. Bu durum, meta analizlerinde etki büyüklüğünün abartılı bir şekilde tahmin edilmesine yol açabilir (Easterbrook ve ark., 1991).

Tekrarlanan yayınlar "gizli" ve "açık" olarak sınıflandırılabilmektedir. Bir çalışmada, orijinal çalışmaların uygun bir şekilde referansı ile verilerin tekrar analiz edilerek yayınlanması açık tekrarlayan yayın, aynı verilerin farklı yer veya zamanlarda orijinal yayınlara gerekli atıflar yapılmadan yayınlanması gizli tekrarlayan yayın olarak tanımlanmaktadır (Song ve ark., 2010). Tekrarlayan yayın yanlılığından kaçınmak için çalışmaları meta analizine bir defa dahil etmek önemlidir. Bunun için çalışmalarda kullanılan verilerin kaynakları açıkça verilmeli, atıfları yapılmalı, tekrarlanarak yayınlanan çalışmalar arasında çapraz referanslar olmalıdır (Rothstein ve ark., 2005).

Atıf yanlılığı

Araştırma bulgularının nitelik ve yönüne göre atıflanması ya da atıflanmaması nedeni ile atıf yanlılığı ortaya çıkabilir. Meta analizine dahil edilen yayınların referansları incelenerek, dahil edilecek diğer yayınlar belirlenebilmektedir. Fakat referanslarından faydalanılan yayın ilgilileri ikna etmek ve yazarın kendi fikrini desteklemek amacıyla atıfta bulunmuşsa meta

analizi için yanlı bir çalışma çerçevesi sunmuş olur ve atıf yanlılığı ortaya çıkar (Rothstein ve ark., 2005).

Gri literatür yanlılığı

Üçüncü Uluslararası Gri literatür Konferansı (Third International Conference on Grey Literature) tarafından gri literatür "devlet, akademi, iş dünyası ve sanayi tarafından basılı ya da elektronik olarak yayınlanan, fakat ticari yayıncılar tarafından kontrol edilmeyen" tanımı yapılmıştır (Rothstein ve ark., 2005; Song ve ark., 2010). Gri literatür, konferans özetleri, araştırma ve teknik raporlar, kitap bölümleri, internet siteleri, seminer bildirileri, veri arşivleri, tezler, politika raporları, ödevler, kişisel yazışmalar, devam eden araştırma veri tabanları gibi çeşitli kaynakları içermektedir.

Gri literatür yanlılığı ise kayda değer çalışmaların hakemli dergilerde yayınlanma ihtimali fazlayken, önem derecesi daha düşük olan çalışma bulgularının, hakemli dergilerin dışında kalan yayınlarda yayınlanması nedeniyle, hakemli dergilerde bildirilen sonuçların gri literatürde verilenlerden düzenli olarak farklı olmasından kaynaklanır. Hakemli dergilerde yayınlanan araştırmalar, genellikle daha olumlu sonuçlar elde etme ve daha büyük etki büyüklükleri bildirme eğilimindedir. Bu bağlamda, gri literatürün meta analizlerine dâhil edilmesi, genel etki büyüklüğünün daha küçük bir değere indirgenmesine neden olabilir (Song ve ark., 2010). Ayrıca, sistematik yanlılıkları en aza indirgemek adına gri literatürün meta analizi süreçlerine dâhil edilmesi önemli bir strateji olarak öne çıkmaktadır.

Ulaşılabilirlik yanlılığı

Meta analizine dahil edilen çalışmaların erişimi kolay çalışmalardan seçilmesi sonucu ortaya çıkması muhtemel yanlılık ulaşılabilirlik yanlılığıdır (Rothstein ve ark., 2005). Çalışmaya seçilen veriler sadece kolaylıkla ulaşılabilen belirli çalışma türlerinden elde edildiğinde ulaşılabilirlik yanlılığı ortaya çıkar.

Maliyet yanlılığı

Meta analizine dahil edilen çalışmaların maliyeti az olan ya da ücretsiz çalışmalardan seçilmesi sonucu meydana gelebilecek yanlılıktır (Rothstein ve ark., 2005). Maliyeti az olan ya da ücretsiz çalışmalar ilgili literatürün temsilcisi olmayabilir.

Benzerlik yanlılığı

Meta analizine sadece belirli bir alandan seçilen çalışmaların dahil edilmesi, benzerlik yanlılığının ortaya çıkmasına yol açabilir (Rothstein ve ark., 2005).

Gecikme yanlılığı

Bir çalışmanın yayınlanma hızının sonuçların gücü ve yönüne göre değişiyorsa yani, bir çalışmanın tamamlanması ve yayınlanması için geçen süre çalışma sonuçlarından etkileniyorsa gecikme yanlılığı söz konusudur (Ioannidis, 1998; Song ve ark., 2010)

Yazarların çalışmalarına ait dikkat çeken bulguları yayınlaması konusunda, editörlerin de daha çok haber değeri taşıyan bulguları yayınlama konusunda daha hevesli olması gibi nedenlerle istatistiksel olarak anlamlı sonuçları olan çalışmalar, anlamsız olanlara göre daha erken yayınlanma eğilimi gösterirler (Ioannidis, 1998; Song ve ark., 2010; Toews ve ark., 2017). Çalışmanın yayın zamanı ile etki büyüklüğü arasında negatif bir ilişki söz konusudur (Jennions, & Moller, 2002). Zaman içinde yapılan çalışmalarda bildirilen azalan etki büyüklüğünün, başka bir değişle meta analizinde etki büyüklüğünün zamansal eğilimlerinin gecikme yanlılığının bir sonucu olması muhtemeldir (Song ve ark., 2010).

Kesme yanlılığı

Çalışmaların yayınlanacağı bilimsel dergilerde yer alan kelime sınırı çalışmaların bütün bulgularının raporlanabilmesi için sınırlayıcı olabilir ve eksik raporlamaya neden olur. Kitap, rapor, tez gibi yayınlarda sınırlama olmaması ya da daha az olmasından dolayı çalışmalara ait bütün bulguların raporlanması ihtimali daha çoktur (Toews ve ark., 2017).

Bütün yayın yanlılığı

Meta analizlere dahil etmek için özet çalışmalar önemli bir kaynaktır. Araştırmacılar tarafından, çalışmayı derginin kabul etmeyeceği ya da çalışmanın iyi olmadığı düşüncesi gibi düşünceler nedeniyle bilimsel toplantılarda özet olarak sunulan bu çalışmalar, daha sonra tam olarak yayınlanmayabilmektedir. Özet olarak sunulan çalışmaların tamamının yayınlanması bulguların gücü ya da yönüne göre değişmesiyle bütün yayın yanlılığı söz konusu olmaktadır (Song ve ark., 2010).

Yayın yeri yanlılığı

Olumlu sonuçlar bildiren araştırmalar, genellikle daha yüksek etki faktörüne sahip ve daha geniş kitlelere ulaşan dergilerde yayınlanma eğilimindedir. Bu durum, kimi zaman dergilerin belirli konulara öncelik veren yayın politikalarından da kaynaklanabilir (Song ve ark., 2010). Çalışmanın sonuçlarının yönü ya da etkisinin büyüklüğü ile yayınlandığı dergi arasında bir ilişki olduğunda ise bu durum "yayın yeri yanlılığı" olarak tanımlanır (Rothstein ve ark., 2005).

Veri tabanı yanlılığı- indeksleme yanlılığı

Araştırmalar genellikle, öncelikle çevrimiçi kaynaklardan yapılan bibliyografik taramalarla başlar. Ancak MEDLINE gibi veri tabanları, ilgili alandaki tüm çalışmaları kapsamayabilir. Eğer yalnızca indekslenmiş çalışmalarla sınırlı kalınıyor ve bu çalışmalar, indekslenmemiş olanlardan sistematik olarak farklılık gösteriyorsa, bu durum "veri tabanı yanlılığı" olarak adlandırılır (Song ve ark., 2010).

Medya dikkat yanlılığı

Bilimsel ve tıbbi gelişmelerle ilgili bilgilerin büyük bir kısmı genel halk tarafından popüler medya aracılığıyla edinilmektedir. Medyanın bu gelişmeleri sunma biçimi, toplumun konuya dair algısını önemli ölçüde şekillendirebilir. Eğer çarpıcı ya da dikkat çekici bulgular içeren araştırmalar, gazete, radyo veya televizyon gibi medya organlarında diğer çalışmalara göre daha fazla yer buluyorsa, bu durum "medya dikkat yanlılığı" olarak tanımlanır (Song ve ark., 2010).

Ülke yanlılığı

Aynı konu üzerine farklı ülkelerde yürütülen araştırmalarda elde edilen sonuçların değişkenlik göstermesi, sıklıkla "ülke yanlılığı" ile ilişkilendirilir. Özellikle gelişmekte olan ülkelerden gelen çalışmalarda, alışılmadık derecede yüksek oranda olumlu sonuçlar bildirilmesi bu yanlılığa örnek teşkil edebilir (Song ve ark., 2010).

Yanlılıktan Kaçınmak İçin Yapılması Gerekenler

Doğru ve geçerli bir meta analizi yapabilmek için ilk gereken, analize katılacak çalışmaların belirli bir sisteme uygun ve dikkatlice seçilmesidir. Analize dahil edilecek çalışmaların örnekleminin ise evreni temsil edebilmesi için bütün kaynakların araştırılması ve hangi çalışmaların meta analizine seçileceği önem arz eder. Fakat meta analizine dahil edilecek

çalışmalar nerede araştırılırsa araştırılsın, meta analizi çalışmasını okuyacak olanların sonuçların geçerliliğini değerlendirebilmeleri için, meta analizi yazarlarının açık bir şekilde hangi kaynakların incelendiğini ve araştırıldığını belirtmeleri, hangi stratejilerin kullanıldığını, hangi çalışmaların dahil edildiğini belirtmeleri ve belgelemeleri de önemlidir. Eğer araştırmalar sınırlı bir kapsamda gerçekleştirildiyse, meta analizi sonuçları da yanlı olacaktır anlamına gelebilir (Rothstein ve ark., 2005). Bu yanlılığının önüne geçebilmek adına aşağıda verilen kaynaklarında aranması önemlidir.

Merkezi kayıt defterlerinin aranması

Meta analizi konusu ile ilgili elektronik bibliyografik veri tabanlarının araştırılması gerekir. Meta analizlerinde yanlılık oluşturma riskini azaltmak için "Cochrane Kontrollü Araştırmalar Merkezi Kayıt Defteri" (CENTRAL) gibi merkezi kayıt defterleri oluşturulmuştur. CENTRAL, bu açıdan en kapsamlı kayıt kaynaklarından biri olup, MEDLINE, EMBASE gibi elektronik veri tabanlarında yer almayan randomize çalışma raporlarını, İngilizce dışındaki dillerde yayınlanan alıntıları, konferans bildirilerinde yer alan referansları ve ulaşılması güç diğer kaynaklardan alınan alıntıları da içermektedir (Rothstein ve ark., 2005).

Elektronik veri tabanlarında ve el ile arama

MEDLINE, EMBASE ve ERIC gibi elektronik bibliyografik veri tabanları, meta analizine katılacak çalışmaların belirlenmesinde önemli bir referans kaynağı sunmaktadır. Ancak yalnızca bu veri tabanlarıyla sınırlı kalmak araştırmada yanlılığa yol açabilir, çünkü her çalışma bu platformlarda yer almamaktadır. Bazı araştırmalar yalnızca özet biçiminde yayınlanmış ya da veri tabanları tarafından düzenli olarak indekslenmeyen dergi eklerinde bulunmuş olabilir. Ayrıca, bazı çalışmalar veri tabanı indeksleme sistemlerinin günümüzdeki kadar gelişmiş olmadığı dönemlerde yayınlandığından, bu raporların da gözden kaçma ihtimali vardır (Rothstein ve ark., 2005).

Konferans bildirilerinin aranması

Meta analizine dahil edilecek çalışmaların belirlenmesi için önemli kaynaklardan biri de konferans bildirileridir. Bu bildirilerde yer alan özetler çoğunlukla elektronik veri tabanlarında yer almadıklarından ya da zaman kısıtı, sonuçların önemli olmadığı düşüncesi, dergilerin çalışmaları kabul etmeyecekleri düşüncesi gibi nedenlerle tam olarak yayınlanmadıklarından,

ulaşmak için sadece konferans bildirilerini taramak gerekir (Rothstein ve ark., 2005).

Araştırmacılarla iletişim kurma

Araştırmacılar yayınlanmış çalışmalardan farklı sonuçlara sahip olması muhtemel bazı çalışmalarını yayınlamazlar. Buna bağlı olarak da yanlılığı azaltmak için bu çalışmaları belirlemek ve değerlendirme sürecine almak gerekir. Alanında uzman kişiler veya ilgili sektörle irtibata geçmek yayınlanmamış ya da henüz bitirilmemiş çalışmaları belirlemenin bir yoludur. Şirketlerin bazıları ise, araştırmaları kaydederek devam eden ya da tamamlanan çalışmaların kayıtlarını açık hale getirerek erişime sunarlar (Rothstein ve ark., 2005).

Araştırma kayıtlarının aranması

Yayınlanmayacak ya da devam eden çalışmaları belirlemeye imkân sağlayan araştırma kayıtları önemlidir çünkü, bir meta analizi güncellenmek istendiğinde buradaki çalışmalar da değerlendirilebilir. Ayrıca devam eden çalışmaların kaydı ile gecikme yanlılığının ihtimali de değerlendirilebilir (Rothstein ve ark., 2005).

İnternette arama

İnternet, tamamlanan ya da süregelen araştırmalar ya da yayınlanmamış araştırmalar için önemli bir kaynak olmakla birlikte arama motorlarının karmaşık aramaya izin vermemesi ya da çok sayıda web sitesi belirleyebilmesi gibi nedenlerle oldukça zahmetlidir (Rothstein ve ark., 2005).

Farklı bilimler arasındaki aramadaki farklılıkları hesaba katmak

Meta analizlerinde çalışmaları belirleme sürecinde, farklı disiplinlerin kendine özgü yanlılık riskleri olabilir. Örneğin, tıp alanındaki araştırmalar genellikle hakemli dergilerde yayınlanır ve büyük bibliyografik veri tabanlarında kolaylıkla erişilebilir. Buna karşın, sosyal bilimler literatürü daha çeşitli bir yapı sergiler; çalışmaların indekslendiği veri tabanlarının kapsamı ve kalitesi değişkenlik gösterebilir ve çoğu zaman indeksleme terimlerinin eksikliği söz konusu olabilir (Rothstein ve ark., 2005).

Literatürde Geliştirilen Meta Analizinde Yanlılık Oluşturma Yaklaşımları

Kale ve Nirpharake (2017), meta analizinde yanlılık oluşturmak için etki büyüklüğünün olumsuz tarafında yer alan çalışmaların %20'sini çıkararak yeni bir küme oluşturmuşlardır.

Peters ve arkadaşları (2006), çalışmalarında funnel grafiği asimetrisini iki farklı biçimde simüle etmişlerdir. Birinci yöntemde, bir çalışmanın p-değeri arttıkça meta analizine dahil edilme olasılığı azaldığı varsayılmış ve bu nedenle yanlılık, etki büyüklüğü ile ilgili p-değerine dayalı olarak modellenmiştir. İkinci yöntemde ise, daha büyük odds ratio (OR) değerlerine sahip çalışmaların istatistiksel anlamlılığa ulaşma olasılığı daha yüksek olduğu için, bu tür çalışmalarda daha az yanlılık olacağı varsayılmış ve yanlılık doğrudan etki büyüklüğüne göre uyarılmıştır.

Duval ve Tweedie (2000), 35 çalışmadan oluşan ve anakütle ortalama etkisi sıfır (0) olan tek bir rassal etkiler meta analizini simüle etmişlerdir. Yanlılık, bu meta analizinde de funnel grafiğinden 5 çalışmanın çıkarılması yoluyla uygulanmıştır.

Sterne ve arkadaşları (2000) ise 5, 10, 20 ve 30 çalışmadan oluşan "tipik" meta analizi örnekleri üretmiş ve bunlar üzerinde simülasyonlar gerçekleştirmiştir. Amaçları, farklı derecelerdeki yanlılık durumlarında yanlılık olmayan, orta derecede ve şiddetli yanlılık koşulları ağırlıklı regresyon yöntemi ile sıra korelasyon testinin duyarlılığını (yanlılığı tespit etme gücünü) ve yanlış pozitif sonuç üretme olasılıklarını değerlendirmektir. Yanlılığın derecesi, eşitlik (1) de verilen etki büyüklüğü (log OR) ile onun standart hatası (SE) arasında doğrusal bir ilişki varsayımıyla tanımlanmıştır. Bu doğrultuda, simülasyonlar yanlılık katsayıları 0 (yanlılık yok), -0.5 (orta şiddette yanlılık) ve -1 (şiddetli yanlılık) olacak şekilde gerçekleştirilmiştir.

log(OR)= Gerçek etki büyüklüğü (log OR) + (yanlılık katsayısı x etki büyüklüğünün standart hatası (log OR)). (1)

Meta Analizinde Yanlılık Belirlemede Kullanılan Yöntemler

Funnel plot grafikleri, meta analizinde yanlılık araştırması için görsel bir araçtır fakat görsel olmasından dolayı araştırmacıdan araştırmacıya farklı biçimlerde yorumlanabilmektedir. Funnel plot grafiklerinin yanlılığını öznel bir şekilde değerlendirme imkânı sunduğu için, bu değerlendirmeyi daha objektif hale getirebilmek amacıyla Begg, Egger, Schwarzer, Harbord,

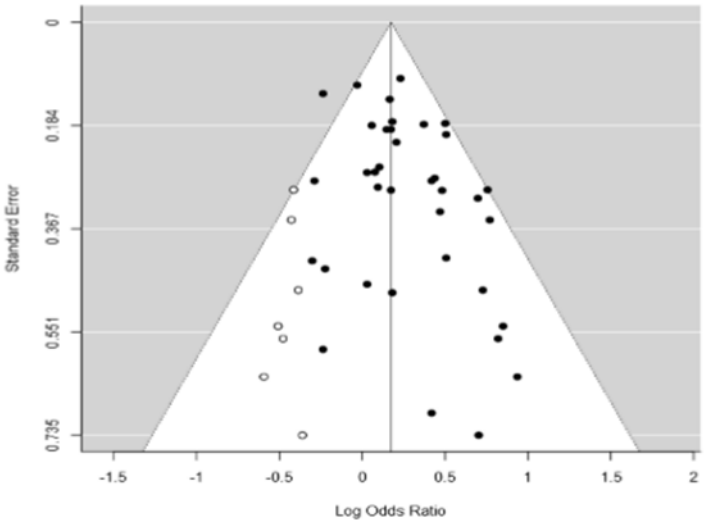
Macaskill, Peters, Thompson, Arcsin-Begg, Arcsin-Egger ve Arcsin-Thompson gibi istatistiksel testler geliştirilmiştir. Bu testler, yanlılığın objektif bir biçimde değerlendirilmesini sağlar.

Funnel Plot Grafiği

Light ve Pillemer (1984) tarafından geliştirilen funnel plot grafiği, meta analizine dahil edilen çalışmaların büyüklükleri ile tahmin edilen etki büyüklüklerinin dağılımını görselleştiren bir tekniktir. Bu yöntem, çalışmaların örneklem büyüklüklerine göre etki büyüklüklerinin ilişkisini inceleyerek, araştırmadaki olası yanlılıkları tespit etmeye yardımcı olur (Rothstein ve ark., 2005).

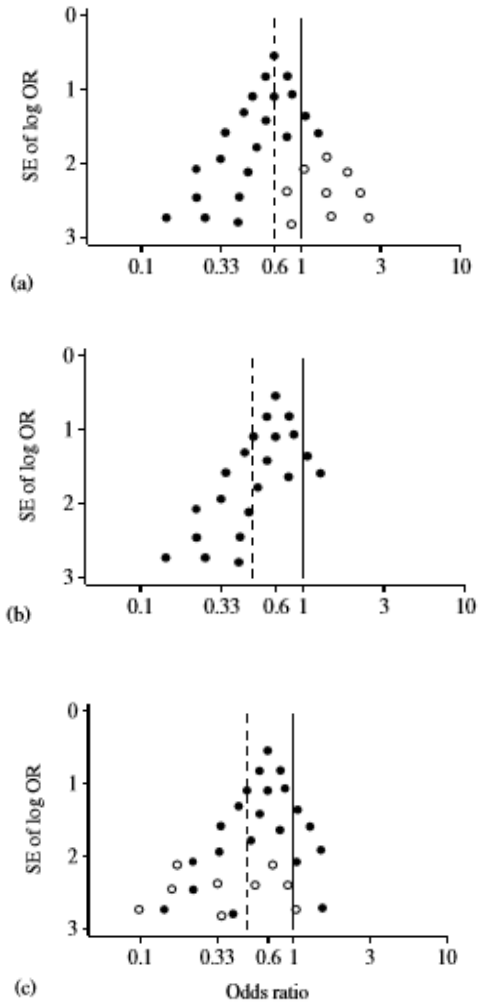
Çalışmaların örneklem hacmi az olduğunda sonuçlarının istatistiksel olarak anlamlı kabul edilmesi için gerekli etki büyüklüğü artar. Ayrıca daha büyük örneklem hacimli çalışmaların sonuçları istatistiksel olarak anlamsız olsa bile daha çok zaman ve para yatırımı yapılmasından dolayı yöntemsel kalitesi yüksek olur ve yayınlanma ihtimali artar. Böylelikle, çalışmanın etki büyüklüğü ile örneklem hacmi arasındaki ilişki meta analizindeki yanlılığın ortaya çıkmasına neden olabilir (Sterne, Gavaghan, & Egger, 2000). Yani, yatay eksende etki büyüklüğü, dikey eksende ise standart hata, standart hatanın tersi, örneklem hacmi, varyans, varyansın tersi ya da çalışma büyüklüğüne dayalı ağırlıkların bulunduğu dağılım grafikleriyle, yani funnel plot grafiklerle meta analizindeki yanlılıklar gösterilebilir. Dikey eksende meta analizine alınan çalışmaların örneklem hacimleri ölçeği yer aldığında, örneklem hacimli az olan çalışmalardan elde edilen sonuçlar grafiğin alt tarafına yayılır, örneklem hacmi çok olan çalışma sonuçları arasında ise dağılım daralır. Yani, meta analizine alınan çalışmaların örneklem hacmi büyüdükçe, etki büyüklüğünün tahminindeki duyarlılık (precision) artar (Egger ve ark., 1997; Sterne ve ark., 2000).

Şekil 1'e göre meta analizinde yer alan her çalışmaya karşılık bir içi dolu nokta gelmekte, dikey düz çizgi etki büyüklüğünü, eğik noktalı çizgiler ise etki büyüklüğünün %95 güven aralıklarını göstermektedir. Şekil 1 incelendiğinde bazı çalışmaları temsil eden noktaların %95 güven aralıklarının dışında konumlandığı ve genel olarak ise asimetrik bir görünümde dağıldığı görülmektedir. Funnel plot grafiğindeki asimetri meta analizinde yer alan yanlılık biçiminde, güven aralıkları dışında yer alan her bir çalışmayı temsil eden noktalar ise çalışmaların heterojen olduğu biçiminde yorumlanabilir.



Şekil 1. Funnel Plot Grafiği (Duval, 2005)

Yanlılığın olmadığı bir funnel plot grafiği, simetrik ters bir huniye benzer (Şekil 2(a)), grafik çarpık ve asimetrik ise yanlılık vardır. Örnek olarak, istatistiksel olarak anlamlı etkiler göstermeyen örneklem hacmi daha az olan çalışmalar yayınlanmadığı için bir yanlılık söz konusu ise grafik, şekil (2(b))'de olduğu gibi grafiğin sağ alt tarafında boşluk olacak ve asimetrik görünecektir. Asimetri ne kadar belirginse, yanlılık da o kadar yüksek kabul edilebilir (Egger ve ark., 1997; Rothstein ve ark., 2005).



Şekil 1. Varsayımsal funnel plot grafiklerinin örnekleri: (a) yanlılık bulunmadığında simetrik funnel plot grafiği; (b) yanlılık mevcut olduğunda asimetrik funnel plot grafiği; (c) küçük örneklem hacmine sahip çalışmalarda, düşük metodolojik kalite nedeniyle yanlılık oluştuğunda ortaya çıkan asimetrik funnel plot grafiği (Rothstein ve ark., 2005).

Funnel Plot Grafiği Asimetrisinin Değerlendirmesinde Kullanılan Yöntemler

Funnel plot grafiğindeki asimetriyi incelemek için farklı istatistiksel testler önerilmiştir. Bu bölümde Begg, Egger, Schwarzer, Harbord, Peters, Macaskill, Thompson ve Arcsin yöntemlerine yer verilmiştir. Söz konusu testler, iki ana grupta sınıflandırılabilir: İlki, Begg ve Mazumdar'ın (1994) çalışmasından esinlenerek sıra korelasyonuna dayanan parametrik olmayan

testlerdir; ikincisi ise Egger testinin çeşitli modifikasyonlarını içeren regresyon tabanlı testlerdir. Bu testler aracılığıyla funnel plot grafiği asimetrisinin saptanmasında, anlamlılık düzeyinin 0,10 olarak belirlenmesi önerilmektedir (Egger ve ark., 1997; Sterne ve ark., 2000).

Aşağıda sunulan bazı yöntemlerde, ikili verilere ait tanımlayıcı istatistiklerin hesaplanmasında Tablo 1'de yer alan kontenjans tablosu temel alınacaktır.

Tablo 1. İkili veri analizlerinde etki büyüklüğünü hesaplamak amacıyla kullanılan iki yönlü frekans tablosu

		Sonuç		Toplam
		+	-	
Etken	Hasta(Tedavi) (H+)	a	b	(a+b=e)
	Kontrol (H-)	c	d	(c+d=f)
	Toplam	(a+c=g)	(b+d=h)	n (Toplam)

Begg'in Yöntemi

Begg ve Mazumdar (1994) tarafından geliştirilen yöntemde, düzeltilmiş sıra korelasyonu yaklaşımıyla etki büyüklüğü tahminleri ile bu tahminlere ait örnekleme varyansları arasındaki olası ilişki analiz edilmiştir. Meta analizlerinde yanlılık söz konusu olduğunda, daha küçük örneklem büyüklüklerine sahip çalışmaların daha büyük etki büyüklükleri göstermesi beklenir. Bu durumun temelinde, küçük örneklem büyüklüğüne sahip çalışmaların daha yüksek varyans tahminleri üretmesi yatmaktadır; bu da etki büyüklüğü ile varyans arasında pozitif yönlü bir ilişki oluşmasına neden olmaktadır (Egger ve ark., 1997).

Sıra korelasyon testinin geçerliliği sağlanmadan önce, etki büyüklüklerinin standartlaştırılması gerekmektedir. Bu işlem, varyansları dengelemek amacıyla yapılır (Begg & Mazumdar, 1994). θ_i^*, i numaralı çalışmadan elde edilen etki büyüklüğü tahmini (θ_i , yani log odds oranı) ile bu tahmine ait varyans v_i kullanılarak hesaplanan, standartlaştırılmış etki büyüklüğünü göstermektedir. Standartlaştırılmış etki büyüklükleri, eşitlik (2)'de gösterildiği şekilde ifade edilir.

$$\theta_i^* = (\theta_i - \bar{\theta}) / \sqrt{v_i^*} \qquad (2)$$

Eşitlik (2) kullanılarak standartlaştırılmış etki büyüklükleri elde edilir ve bu sayede varyanslar arasında dengeleme sağlanır. Bu kapsamda, ağırlıklı ortalama etki büyüklüğü $\bar{\theta}$, $\bar{\theta} = (\sum \theta_i v_i^{-1})/\sum v_i^{-1}$ formülüyle tanımlanır. Her bir çalışmaya ait düzeltilmiş varyans ise $v_i^* = v_i - (\sum v_j^{-1})^{-1}$ şeklindedir ve bu değer, $(\theta_i - \bar{\theta})$ farkının varyansını ifade eder (Rothstein ve ark., 2005).

Begg testi, θ_i^* ile v_i^* arasındaki Kendall sıra korelasyonuna dayanır; yani iki değerin sıralarının karşılaştırılmasına odaklanır. Örnek olarak, θ_i^*'nin en büyük değeri 1. sırada, sonraki en büyük değeri 2. sıradadır. k çalışma için toplamda k(k-1)/2 çalışma çifti bulunur. i ve j çalışmalarına ait, (θ_i^*, v_i^*) ve (θ_j^*, v_j^*) çiftlerinin sıralamaları aynı yönde farklılık gösteriyorsa bu çiftler uyumlu kabul edilir. Başka bir ifadeyle, i numaralı çalışmanın hem θ_i^* hem de v_i^* sıralamaları, j numaralı çalışmanınkilerden ya tamamen daha düşük ya da tamamen daha yüksekse, bu bir uyum durumu olarak kabul edilir. Ancak sıralamalardan biri daha yüksekken diğeri daha düşükse, bu durum sıralar arasında uyumsuzluk olarak değerlendirilir. x uyumlu sıralar, y ise uyumsuz sıralar olduğunda normalleştirilmiş test istatistiği z eşitlik (3) de verildiği gibi hesaplanır (Rothstein ve ark., 2005; Jin, Zhou, & He, 2014).

$$z = \frac{x-y}{\sqrt{k(k-1)(2k+5)/18}} \quad (3)$$

Eşitlik (3)'te tanımlanan z test istatistiği, meta analizinde yanlılık bulunmadığını varsayan sıfır hipotezi altında, asimptotik olarak standart normal dağılıma uymaktadır. α anlamlılık düzeyinde sıfır hipotezi $|z| = z_{1-\alpha/2}$ değeriyle reddedildiğinde meta analizinde yanlılık olmadığını gösterir.

Varyansının $\{v_i\}$ sıralaması $\{s_i\}$, standart hatalarının sıralamalarıyla örtüştüğünden, Begg testi, standart hatalar ile standart etki büyüklükleri arasındaki sıra korelasyonu incelemeye eşdeğerdir (Rothstein ve ark., 2005). Funnel plot grafiğindeki çarpıklığın, $\{v_i\}$ ve $\{\theta_i\}$ arasında korelasyondan kaynaklanacağı varsayımıyla bu test prosedürü Begg ve Mazumdar (1994) ile Jin ve arkadaşları (2014) tarafından önerilmiştir. Buna göre, sıfır hipotezi altında meta analizinde yanlılık bulunmadığı kabul edildiğinde, meta analizine katılan çalışmaların etki büyüklükleri ile kesinlik (θ_i ve v_i) arasında herhangi bir ilişki bulunmamaktadır.

Egger'ın Yöntemi

Egger testi, funnel plot grafiği asimetrisini lineer regresyon yöntemi ile değerlendiren bir yöntemdir.

$$E[z_i] = \beta_0 + \beta_1 prec_i \quad (4)$$

Eşitlik (4)'te $E[z_i]$, θ_i'nin (i. çalışmadan elde edilen etki büyüklüğü tahmini) değerinin standart hata (s_i) ile bölünmesiyle elde edilen, standartlaştırılmış etki büyüklüğüdür ve bu formül (z_i=θ_i/s_i) şeklinde ifade edilir. $prec_i$ = $1/s_i$,β_0 sabit bir terim ve β_1 eğim katsayısıdır (Rothstein ve ark., 2005).

Egger testi, meta analizinde yanlılığın varlığını test ederken, sıfır hipotezini yanlılık olmadığı yönünde kurar. Bu hipotez, asimptotik olarak k-2 serbestlik derecesine sahip Student t dağılımına dayanarak değerlendirilir ve β_0 sabitinin sıfır olmadığı kabul edilerek test edilir. Meta-analizlerinde yanlılık olmadığına dair sıfır hipotezi, α anlamlılık düzeyine göre $\left|\hat{\beta}_0 / s_i(\hat{\beta}_0)\right| = |t| > t_{k-2;1-\alpha/2}$ koşulunun sağlanması durumunda reddedilir. Bu prosedür, yanlılık söz konusu olduğunda doğrusal ilişkinin geçerli olduğu varsayımına dayanmaktadır.

Eğer β_0 = 0 koşulu sağlanıyorsa, funnel plot grafiği simetrik hale gelir ve regresyon doğrusu orijinden geçer. β_1 (eğim) parametresi, etki büyüklüğünün yanı sıra, etki yönünü de gösterir. Eğer funnel plot grafiğinde asimetri mevcut ise regresyon çizgisi orijinden geçmez ve β_0 bir asimetriyi ölçen bir gösterge sağlar. Sıfırdan sapmanın büyüklüğü ölçüsünde ise asimetri belirginleşir (Egger ve ark., 1997; Rothstein ve ark., 2005).

Meta analizi çalışmalarında sabitin istatistiksel anlamlılığı çoğunlukla α=0,10 anlamlılık seviyesinde test edilir (Tang, & Liu, 2000). Sabitin negatif bir değer aldığı durumlar, daha küçük örneklem büyüklüğüne sahip çalışmaların, daha büyük örneklemli çalışmalara kıyasla daha güçlü etkiler gösterdiğini ortaya koymaktadır (Egger et al., 1997).

Egger ve arkadaşları (1997), Egger testinin alternatif bir versiyonunu, etki tahminlerinin varyansının tersine, yani $1/v$ 'ye dayalı ağırlıklandırma kullanarak geliştirmiştir (Jin et al., 2014; Rothstein et al., 2005). Bu yaklaşımda, z'nin "prec" üzerindeki ağırlıksız regresyonu, her çalışmadaki tahmin edilen etki büyüklüğünün varyansının ($1/v_i$) tersine orantılı ağırlıklarla hesaplanan standart hatalar (s) üzerinde yapılan etki büyüklüğünün (θ) ağırlıklı regresyonuyla eşdeğer olup, şu şekilde ifade

edilir: ($E[\theta_i] = \beta_1 + \beta_0 s_i$). Bu iki regresyon arasındaki tek fark, β_0 ve β_1 katsayılarının rollerinin yer değiştirmiş olmasıdır; ilk regresyondaki sabit, ikinci regresyondaki eğim katsayısına karşılık gelir. Bu bağlamda, funnel plot grafiği asimetrisi üzerine yapılan regresyon testi, etki büyüklüğünün standart hata ile karşılaştırıldığı funnel plot grafiğindeki doğrusal eğilim testine karşılık gelmektedir (Rothstein et al., 2005).

Schwarzer'ın Yöntemi

Seyrek ikili sonuçlara sahip meta analizlerinde yanlılık tespiti için Schwarzer ve arkadaşları (2007), Begg testine dayalı bir yöntem önermiştir. Bu yöntemde, Begg testinde kullanılan standart etki büyüklüğü ve varyans yerine sırasıyla standartlaştırılmış hücre sayıları ve yeni tanımlanan varyans kullanılmıştır. Buna ek olarak, gözlenen hücre sayısının (a_i) olasılığı, eşitlik (5)'te belirtilen merkezi olmayan hipergeometrik dağılımı izlediği varsayılmaktadır.

$$P(a_i;\theta_i) = \frac{\binom{a_i+b_i}{a_i}\binom{c_i+d_i}{c_i}\theta_i^{a_i}}{\sum_{x\in U_i}\binom{a_i+b_i}{x}\binom{c_i+d_i}{a_i+c_i-x}\theta_i^{x}} \quad (5)$$

Sabit marjinal toplam U_i göz önünde bulundurulduğunda, a_i değişkeninin alabileceği olası değerlerin aralığı tanımlanır. Belirli bir θ_i değeri için, merkezi olmayan hipergeometrik dağılımın beklenen değeri ve varyansı aşağıda belirtilmiştir.

$$E(a_i;\theta_i)=\sum_{x\in U_i} x\, P(x;\theta_i) \quad (6)$$

$$Var(a_i;\theta_i)=\sum_{x\in U_i}(x - E(a_i;\theta_i))^2\, P(x;\theta_i) \quad (7)$$

Schwarzer'ın yönteminde, bilinmeyen OR (θ_i) değerlerini tahmin etmek Mantel–Haenszel OR ($\hat{\theta}_i^{MH}$) değerleri kullanılır. Koşullu maksimum olabilirlik tahmin edicisinin özellikleri dikkate alındığında, odds oranı için elde edilen $\hat{\theta}_i^{MH}$ değerinin asimptotik varyansı, $1/Var\left(a_i;\hat{\theta}_i^{MH}\right)$ formülüyle yaklaşık olarak hesaplanabilir. Schwarzer ve arkadaşları (2007) funnel plot grafiğindeki asimetriyi Kendall τ ile iki şekilde test etmişlerdir. Birinci yaklaşımda, asimptotik varyans ($\hat{v}_i$) standartlaştırılmış etki büyüklüğünün hesaplamasına dâhil edilerek, bu değer $1 / Var\left(a_i;\hat{\theta}_i^{MH}\right)$ ile değiştirilir. İkinci yaklaşım ise, standartlaştırılmış hücre sayısı ile $1 / Var\left(a_i;\hat{\theta}_i^{MH}\right)$ arasındaki korelasyonun istatistiksel olarak test edilmesini içerir. Bu kapsamda standartlaştırılmış hücre sayısı şu şekilde tanımlanmaktadır:

$s(a_i; \hat{\theta}_i^{MH}) = (a_i - E(a_i; \hat{\theta}_i^{MH}))/(Var(a_i; \hat{\theta}_i^{MH})^{1/2})$ (Jin ve ark., 2014; Schwarzer, Antes, & Schumacher, 2007).

Arcsin Farkına Dayanan Testler (Rucker'in Testleri)

Arcsin dönüşümüne dayanan ve binom türü rastgele değişkenler için stabilize edilmiş varyansı temel alan bu modifiye yöntem, Rücker ve ark. (2008) tarafından geliştirilmiştir. Söz konusu yöntemde, etki büyüklüğü olarak arcsin farkı (Δ_i) esas alınarak Arcsin regresyon testi oluşturulmaktadır. Gözlenen risk farkının Arcsin dönüşümü ve yaklaşık varyansı sırasıyla şu şekilde tanımlanmaktadır: $\Delta_i = \arcsin(a_i/(e_i))^{1/2} - \arcsin(b_i/(f_i))^{1/2}$ ve $\Gamma_i = 1/4e_i + 1/4(f_i)$ Bu yöntemde, orijinal θ_i^* ve $\hat{v}_i$ değerleri yerine Δ_i ve Γ_i kullanılarak Begg, Egger ve Thompson testlerinin özel bir biçimi elde edilmektedir (Jin et al., 2014). Bu yaklaşımın dikkat çekici yönü, varyans tahminlerinin yalnızca örneklem büyüklüklerine (satır toplamlarına) dayanması ve etki büyüklüğü tahminlerinden bağımsız olmasıdır.

Macaskill'in Yöntemi

Macaskill ve arkadaşları (2001) tarafından geliştirilen bu yöntem, tahmini etki büyüklüğünü ($\hat{\theta}_i$) bağımlı, çalışma büyüklüğünü (n_i) ise bağımsız değişken olarak ele alan ağırlıklı bir regresyon yaklaşımıdır. Söz konusu modelde, her bir gözlem; ya tahminin ters varyansı (FIV) ya da ilgili çalışmanın birleşik varyansının tersine (FPV) göre ağırlıklandırılarak analiz edilmiştir. Ağırlıklandırma katsayısı genellikle $1/\hat{v}_i$ veya $[1/(a_i + c_i) + 1/(b_i + d_i)]^{-1}$ şeklindedir. Bu koşullarda regresyon modeli $\hat{\theta}_i = \beta_0 + \beta_1 n_i + \varepsilon_i$ biçiminde ifade edilir. Meta analizinde yanlılık söz konusu değilse regresyon eğimi olan β_1 sıfıra eşittir. Ancak β_1 katsayısının sıfırdan anlamlı şekilde farklı olması, etki büyüklüğü ile örneklem hacmi arasında sistematik bir ilişki bulunduğunu ve bunun potansiyel bir yanlılık göstergesi olabileceğini düşündürmektedir (Jin et al., 2014; Macaskill, Walter & Irwig, 2001).

Harbord'ın Yöntemi

Egger testinin bir türevi olan bu yöntem, Harbord ve arkadaşları (2006) tarafından geliştirilmiştir. Yöntem, her çalışma için etki büyüklüğü tahmini olarak etkin skor ($Z_i = a_i - (a_i + c_i) \times e_i/n_i$) ve buna karşılık gelen varyans skoru ($V_i = (a_i + c_i) \times (b_i + d_i) \times e_i \times f_i/[n_i^2 \times (n_i - 1)]$) hesaplamalarına dayanır. Egger ve arkadaşları (1997) tarafından önerilen

klasik regresyon testinde etki büyüklükleri ile bunlara ait standart hatalar kullanılırken, modifiye edilmiş regresyon testinde bu değerlerin tahminleri ve yaklaşık standart hataları esas alınmaktadır. Bu testte, V ağırlığı ile Z/V ve $1 / \sqrt{V}$ kullanılarak kurulan ağırlıklı regresyon modeli şu şekilde tanımlanır: $Z_i/V_i=\beta_0 + \beta_1 1/\sqrt{V_i} + \varepsilon_i$'dir. Burada V_i ağırlığı temsil ederken, hata terimi $\varepsilon_i \sim N\left(0, \frac{\hat{v}_i}{V_i} \times \phi\right)$ şeklinde tanımlanmıştır. Regresyon katsayısı için β_1 = 0 olan sıfır hipotezi, herhangi bir yanlılığın bulunmadığını ifade eder ve bu hipotez iki yönlü t testi ile değerlendirilir (Harbord, Harris & Sterne, 2009; Jin et al., 2014). Bu yöntemin önemli bir avantajı, varyansın yalnızca marjinal toplamlar üzerinden tahmin edilebilmesine olanak tanımasıdır (Schwarzer ve ark., 2015).

Peters'in Yöntemi

Peters ve arkadaşları (2006) tarafından geliştirilen bu yöntem, Macaskill'in funnel plot regresyon yönteminin (FPV) örneklem büyüklüğüne dayalı bir modifikasyonudur. Yöntem, toplam örneklem hacminin tersinin bağımsız değişken olarak kullanıldığı ağırlıklı bir regresyon modeline dayanır. Ağırlık $[1/(a_i + c_i) + 1/(b_i + d_i)]^{-1}$ biçiminde tanımlanırken, hata terimi $\varepsilon_i \sim N(0, se_i^2 \times \phi)$ şeklindedir. Ağırlıklı regresyon analizi kapsamında oluşturulan model, $\hat{\theta}_i = \beta_0 + \beta_1 1/(a_i + b_i + c_i + d_i) + \varepsilon_i$ biçimindedir. Bu modelde test edilen sıfır hipotezi, β_1 = 0 durumunu yani yanlılığın bulunmadığını belirtir. Eğim katsayısının istatistiksel anlamlılığı, iki taraflı t testi aracılığıyla analiz edilmektedir (Harbord ve ark., 2009; Jin ve ark., 2014).

Thompson'ın Yöntemi

Thompson ve Sharp (1999), çalışmalar arası heterojenliği dikkate alan bir regresyon tabanlı test önermiştir. Bu yöntem, Egger testinin bir uzantısı olup, meta analizlerindeki toplanabilir varyans bileşenini moment tahminiyle belirler ve standart hata üzerine kurulu etki büyüklüğü regresyonunu kullanır. Regresyon modeli aşağıdaki şekilde ifade edilir:

$$\theta_i = \beta_0 + \beta_1 \times s_i + \varepsilon_i \quad , \quad \varepsilon_i \sim N(0, s_i^2 + \tau^2) \qquad (8)$$

Burada, τ^2 çalışmalar arası varyansı, s_i her bir çalışmanın standart hatası ve $w_i = 1/(s_i^2 + \tau^2)$ regresyon ağırlığıdır. β_0 ve β_1 katsayıları, ağırlıklı en küçük kareler yöntemiyle tahmin edilir. Test istatistiği, serbestlik derecesi 2

olan bir t dağılımını takip eder. $\beta_1 = 0$ şeklindeki sıfır hipotezi, çalışmalarda yayın yanlılığının bulunmadığını ifade eder.

SONUÇ

Meta analizinde en önemli süreçlerden bir tanesi yanlılık araştırmasıdır. Meta analizinde saptanan olası yanlılıkların, elde edilen sonuçlar üzerinde anlamlı bir etki yaratmadığı bilimsel olarak ortaya konmalıdır. Meta analizinde yanlılık değerlendirmesinde funnel plot grafikleri öznel bir değerlendirme sunarken, objektif değerlendirme yanlılık belirlemeye yönelik istatistiksel testlerle yapılmaktadır. Meta analizinin yayın yanlılığının değerlendirilmeden sunulması düşünülemez. Meta analizinde yanlılık bulunması durumunda da uygun istatistiksel yaklaşımlar uygulanır.

Kaynaklar

Kitap

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-Analysis with R*. Springer International Publishing, Switzerland.

Whitehead, A. (2002). *Meta-Analysis of Controlled Clinical Trials*. Wiley Publication, England.

Dergi Makalesi

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. Biometrics, 50(4), 1088-1101. Erişim adresi: https://www.jstor.org/stable/2533446 Ozan, C. ve Köse, E. (2014). Eğitim programları ve öğretim alanındaki araştırma eğilimleri. *Sakarya University Journal of Education, 4(*1), 116-136.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta Analysis*. United Kingdom, John Wiley & Sons, Ltd.

Duval, S. J. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.). *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 127–144). Chichester, England: Wiley.

Easterbrook, P. J., Berlin, J. A., Gopalan, R., & Matthews, D.R. (1991). Publication bias in clinical research. *Lancet*. 337, 867–72. doi: 10.1016/0140-6736(91)90201-y

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315, 629–634. doi: 10.1136/bmj.315.7109.629

Harbord, R. M., Egger, M., & Sterne, J. A. C. (2006). A modified test for smallstudy effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine*, 25, 3443–3457. doi: 10.1002/sim.2380

Harbord, R. M., Harris, R. J., & Sterne, J. A. C. (2009). Updated tests for small-study effects in meta-analyses. *The Stata Journal*, 9(2), 197–210. doi: 10.1177/1536867X0900900202

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage Publications, Inc.

Ioannidis, J. P. A. (1998). Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA*, 279(4), 281-286. doi: 10.1001/jama.279.4.281

Jennions, M. D., & Moller, A. P. (2002). *Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution*. Proceedings of the Royal Society of London Series B: Biological Sciences, 269(1486),43–48. doi: 10.1098/rspb.2001.1832

Jin, Z. C., Zhou, X. H., & He, J. (2014). Statistical methods for dealing with publication bias in meta-analysis. *Statistics in Medicine* 34(2), 343–360. doi: 10.1002/sim.6342

Macaskill, P., Walter, S.D., & Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine*, 20(4),641–654. doi: 10.1002/sim.698

Peters, J. I., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of Two Methods to Detect Publication Bias in Meta-analysis. *JAMA*, 295(6), 676-680. doi: 10.1001/jama.295.6.676

Rothstein, H. R., Sutton, A.J., & Borenstein, M. (2005). *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. John Wiley & Sons, Ltd.

Schwarzer, G., Antes, G., & Schumacher, M. (2007). A Test for Publication Bias in Meta-Analysis With Sparse Binary Data. *Statistics in Medicine*, 26, 721-733, doi: 10.1002/sim.2588.

Song, F., Parekh, S., Hooper, L., Loke, Y. K., Ryder, J., Sutton, A.J., …Harvey, I. (2010). Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess*, 14(8), doi: 10.3310/hta14080.

Stern, J. M., & Simes, R. J. (1997). Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ*, 315, 640-645. doi: 10.1136/bmj.315.7109.640

Sterne, J. A. C., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literatüre. *Journal of Clinical Epidemiology*, 53(11), 1119-1129. doi: 10.1016/s0895-4356(00)00242-0

Sterne, J. A. C., & Harbord, R. M. (2004). Funnel plots in meta-analysis. *The Stata Journal*, 4 (2), 127–141. Erişim adresi: http://www.stata-journal.com/sjpdf.html?articlenum=st0061

Sutton, A. J., Duval, S. J., Tweedie, R. L., Abrams, K. R., & Jones, D. R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *BMJ*, 320(7249), 1574-1577. doi: 10.1136/bmj.320.7249.1574

Tang, J. L., & Liu, J. L. Y. (2000). Misleading funnel plot for detection of bias in meta-analysis. *Journal of Clinical Epidemiology*, 53(5), 477-484. doi: 10.1016/s0895-4356(99)00204-8

Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, 18(20), 2693-2708. doi: 10.1002/(sici)1097-0258(19991030)18:20<2693::aid-sim235>3.0.co;2-v

Toews, I., Booth, A., Berg, R. C., Lewin, S., Glenton, C., Munthe-Kaas, H. M., …, Meerpohl, J. J. (2017). Further exploration of dissemination bias in qualitative research requiredto facilitate assessment within qualitative evidence syntheses. *Journal of Clinical Epidemiology*. 88, 133-139. doi: 10.1016/j.jclinepi.2017.04.010

Wang, K. S., & Liu, X. (2016). Statistical methods in the meta-analysis of prevalence of human diseases. *Journal of Biostatistics and Epidemiology*. 2(1), 20-24. Erişim adresi: https://jbe.tums.ac.ir/index.php/jbe/article/view/46

META BULANIK FONKSİYONLAR*

Nihat Tak**

Giriş

Son yıllarda veri, dünyanın en değerli kaynaklarından biri haline gelmiştir. Gelişen teknolojiler, artan dijitalleşme ve büyüyen veri setleri, veri biliminin önemini her zamankinden daha fazla artırmıştır. Veri bilimi, bu devasa verileri anlamlandırarak, içgörü, öngörü elde etme ve karar verme anlamında günümüzde büyük bir önem arz etmektedir. Birçok disiplini bir araya getiren veri bilimi, temel olarak istatistik ve programlama bilgilerinin harmanlamasından oluşur denilebilir. Veri bilimi, karmaşık problemlere çözümler sunarken, aynı zamanda geleceği tahmin etme ve inovasyonu destekleme gücüne sahiptir. Örneğin, sağlık sektöründe hastalıkların erken teşhisinden, finans alanında risk yönetimine, e-ticarette kişiselleştirilmiş önerilere kadar geniş bir uygulama yelpazesi vardır.

Verinin büyümesiyle birlikte veri biliminde kullanılan yöntemlerin çeşitliliği de hızlıca artmaktadır. Örneğin, zaman serisi verisini modelleyebilmek ve öngörü elde edebilmek için literatürde karşımıza yüzlerce hatta binlerce farklı öngörü yöntemi çıkmaktadır. Veya bir kesit verisini modelleyebilmek ve tahminlerde bulunabilmek için literatürde yüzlerce farklı regresyon yaklaşımı ile karşılaşmak mümkündür. Bu örnekler, diğer tip veriler içinde genellenebilmektedir. Modellemek istediğimiz probleme ilişkin onlarca, yüzlerce hatta binlerce farklı yöntem ile karşılaşmak mümkündür. Bu yöntemlerden hangisi veya hangilerinin veri setine uygun olduğunu anlamak çok zor veya mümkün değildir. Çünkü bu yöntemlerden bazıları bazı veri setlerine iyi uyum sağlarken bazı veri setleri için iyi tahminler/öngörüler elde edememektedir. Örneğin, elimizdeki veri seti klasik lineer regresyon yöntemindeki en küçük kareler (EKK) tahmincilerinin varsayımlarını sağlıyorsa o zaman EKK tahmincilerinden daha iyi bir tahminci aramak vakit kaybı olacaktır. Çünkü EKK, varsayımlar sağlandığında en iyi model kurma garantisini vermektedir. Ancak, gerçek hayat verilerinde bu varsayımların sağlanması çok mümkün olmamaktadır.

** Doçent Dr., Marmara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, İstanbul, nihattak@gmail.com, ORCID: https://orcid.org/0000-0001-8796-5101

Bundan dolayı literatürde varsayımlardan bağımsız yüzlerce farklı yöntem bulmak mümkündür. Örneğin, Yapay Sinir Ağları (YSA) varsayımlardan bağımsız güvenilir olmayan tahminciler olarak karşımıza çıkmaktadır. Güvenilir olmayan tahminciler olarak adlandırılmasının nedeni ise parametre tahminlerini rastgele yapmasından kaynaklanmaktadır. Yani, başlangıç olarak rastgele değerler seçilir ve bu değerlerin iteratif olarak güncellenerek amaç fonksiyonunu (hata kareler toplamı) minimum yapacak parametre değerlerini araması mantığını dayanmaktadır. Dolayısıyla YSA ile model kurulup tahminler elde edilirse, her defasında farklı tahminler elde edilmektedir. Alternatif ve esnek hesaplama mantığına dayalı birçok yöntem içerisinde rastgelelik barındırdığından, bu alternatif yöntemler kesin çözüm veremezler. Bundan dolayı, bu yöntemler bazı veri setleri için iyi performans sergilerken, bazı veri setleri için çok kötü performans sergileyebilirler. Bu kitap bölümünde, girdisi farklı yöntemlerin çıktıları olan Meta Bulanık Fonksiyonlar tanıtılacak ve farklı alanlardaki kullanımlarından bahsedilip, öngörü üzerine uygulamaları verilecektir.

Meta bulanık fonksiyonlar (MFF) yöntemi Tak (2018) tarafından önerilmiştir. Yöntemin çalışma prensibi meta analizinin temel mantığını içerisinde barındırmaktadır. Meta-analizi Glass (1973) tarafından önerilen bir yöntemdir. Glass çalışmasında, psikoterapi üzerine yapılan 375 farklı çalışmanın sonuçlarını birleştirerek çıkarımlar elde etmiştir. Dolayısıyla, meta-analiz daha çok tıp alanında yaygın olarak kullanılan ve aynı amaca yönelik farklı çalışmaların sonuçlarından istatistiksel olarak yorum yapmayı amaçlayan bir yaklaşımdır. MFF yönteminin esin kaynağı meta-analizidir. MFF yönteminde meta-analizinden farklı olarak aynı amaca yönelik farklı çalışmaların sonuçları birleştirmeye değil de aynı amaca yönelik önerilmiş farklı metotlara ait sonuçların tek bir veri seti için birleştirilmesi amaçlanmıştır.

Bu çalışmanın geri kalanında kısaca MFF ile ilgili literatür çalışması, yöntemin çalışma prensibi, veri uygulaması ve sonuçların tartışılması yapılacaktır.

Literatür Taraması

MFF bulanık kümeleme yaklaşımı ile meta-analiz mantığının birleşimini ifade eden, özellikle belirsizlik içeren karmaşık sistemlerde etkin çözümler sağlayan güncel bir yöntemdir.

MFF kavramının teorik temelleri ilk kez Tak (2018) tarafından ortaya konulmuştur. Bu çalışmada geri beslemeli tip-1 bulanık fonksiyonların (RFF) MFF çerçevesinde uygulanabilirliği araştırılmıştır. Tak (2020), farklı süreç yetenek ve kur kriz indeks tanımlarını meta bulanık fonksiyonları kullanarak birleştirip daha güvenilir sonuçlar elde ederek literatüre pratik açıdan önemli bir katkı sağlamıştır. Tahmin kombinasyonları üzerine yapılan sonraki çalışmalar, MFF yöntemlerinin performansını kanıtlamıştır. Örneğin, meta olasılıksal bulanık fonksiyonlarla gerçekleştirilen tahmin kombinasyonları, klasik yöntemlere kıyasla önemli doğruluk artışları göstermiştir (Tak, 2021a). İleri beslemeli sinir ağlarının tahminleme kapasitesini artırmak için tek gizli katmanlı sinir ağlarında MFF tabanlı yaklaşımlar kullanılmış ve başlangıç ağırlıklarına duyarlılığın azaltılmasıyla daha tutarlı sonuçlar elde edilmiştir (Tak, 2021b). MFF'nin pratik alandaki başarısına göç tahmini uygulamaları da örnek olarak gösterilebilir. Türkiye'nin deniz sınırlarında göçmen hareketlerinin tahmin edilmesinde, farklı tahmin yöntemlerinin entegrasyonu sayesinde daha güvenilir sonuçlar alınmıştır (Cevik ve ark., 2023). Etki büyüklüğü analizlerinde MFF tabanlı yöntemlerin kullanımıyla daha güvenilir ve kapsamlı sonuçlar sağlanmıştır. Bu yöntem, farklı etki büyüklüğü ölçümlerini entegre ederek daha güçlü analizler sunmaktadır (Yabacı Tak & Ercan, 2023). Son yıllarda adaptif tahmin kombinasyonları ve sezgisel bulanık fonksiyonların kullanıldığı çalışmalar da yapılmıştır. Bu yöntemler, tahmin hatalarını azaltma ve adaptif esnekliği artırma açısından önemli avantajlar sağlamıştır (Tak ve ark., 2021c). Ayrıca, döviz krizlerinin belirleyicilerinin analizi ve erken uyarı sistemlerinin tasarımında da MFF yöntemleri başarıyla uygulanmıştır (Gök & Tak, 2023; Tak & Gök, 2022).

Literatür incelemesi, MFF yöntemlerinin çeşitli veri yapılarında yüksek uyarlanabilirlik ve tahmin doğruluğu sağladığını ortaya koymaktadır. Gelecekte yapılacak çalışmaların bu yöntemlerin geliştirilmesine ve daha geniş alanlarda kullanımına odaklanacağı öngörülmektedir.

Meta Bulanık Fonksiyonlar

Meta Bulanık Fonksiyonlar (MFF), ilk olarak Tak (2018) tarafından önerilmiş olup, özellikle zaman serisi tahmin problemlerinde tek bir tahmin yönteminin performansını arttırmak için geliştirilmiştir. İlgili çalışmada MFF'nin temel dayanağı, belirli bir amaca yönelik geliştirilen her bir yöntemin, kısmi de olsa model tahminine katkı sağladığı varsayımıdır. Literatürde, birçok tahminleme yöntemi benzer argümanları farklı

yaklaşımlar çerçevesinde ele alarak geliştirilmiştir. Bu doğrultuda, MFF yönteminin temel varsayımı, "aynı amaca yönelik farklı yöntemlerin, verilen bir veri seti hakkında belirli düzeyde bilgi içerdiği" şeklinde ifade edilebilir. Bu bağlamda, çalışmada aynı amaca hizmet eden farklı yöntemler, verilen veri seti üzerindeki performansları doğrultusunda kümelenmektedir. Örneğin, belirli bir veri seti üzerinde yüksek başarı gösteren yöntemler, belirli bir kümede toplanırken, görece düşük performans sergileyen yöntemler farklı bir kümede yer alması beklenmektedir. Böylece, küme sayısı kadar farklı yöntem kombinasyonlarını içeren fonksiyonlar elde edilmektedir. Elde edilen fonksiyonlar arasından nihai tahmin fonksiyonunun belirlenmesi ise değerlendirme kriterleri doğrultusunda gerçekleştirilmektedir.

MFF yönteminin amacı, farklı tahmin yöntemlerinin sonuçlarını tek başına kullanmak yerine, bulanık c-ortalamalar (Fuzzy C-Means - FCM) algoritması kullanarak kümeler oluşturmak ve bu kümeler aracılığıyla yöntemlerin sonuçlarını birleştirerek daha güçlü ve doğru tahminler elde etmektir.

MFF'nin işleyişi 5 aşamadan oluşur denilebilir. Bu aşamalar;

I. Verilen veri seti için yöntemlerin belirlenmesi: Tahmin yapılacak probleme uygun mevcut yöntemler belirlenir. Veriler eğitim ve test setleri olarak ayrılır.

II. Eğitim ve tahminlerin elde edilmesi: Seçilen her yöntem eğitim verisiyle eğitilir ve eğitilen yöntemler ile test seti kullanılarak tahminler oluşturulur. Bu tahminler bir girdi matrisi oluşturur.

III. Yöntemlerin kümelenmesi: Oluşturulan tahmin matrisindeki yöntemler, sonuçlarına göre FCM algoritmasıyla kümelere ayrılır. Burada yöntemlerin hangi kümeye daha yakın olduğuna dair bulanık üyelik değerleri hesaplanır. Her küme, aslında yöntemlerin doğruluğuna göre oluşturulmuş birer "fonksiyon" olarak düşünülebilir

IV. MFF'lerin oluşturulması: Elde edilen üyelik değerleri, yöntemlerin küme içindeki ağırlıklarını belirler. Bu ağırlıklara göre her küme için ayrı bir Meta Bulanık Fonksiyon oluşturulur.

V. En iyi MFF'nin seçimi: Elde edilen fonksiyonlar arasından, değerlendirme kriterlerine (örneğin RMSE ve MAPE) göre en düşük hatayı veren fonksiyon en iyi MFF olarak seçilir.

Bu yöntemin aşamaları ve uygulama adımları, aşağıda ayrıntılı olarak açıklanmaktadır. Ayrıca yöntemin sözde kodu aşağıda verilmiştir.

Algorithm 1 MFF ile GWO-RFF Yöntemine Ait Sahte Kod

```
Eğitim ve test veri setlerinin uzunluğunu belirt
R-T1FF'ler için AR ve MA gecikme sayısını ve maksimum küme sayısını belirle
for i < c do
    for j < p do
        for k < q do
            Eğitim veri seti ile R-T1FF'leri eğit
            if RMSE < eşik değeri then
                Tahminleri Z matrisine kaydet
            end if
        end for
    end for
end for
m adet modelin çıktısını Z matrisi olarak MFFS için giriş verisi olarak kullan
MFFS için küme sayısını ve bulanık indeks parametresini başlat
while α < m do
    while i < maksimum fonksiyon/küme sayısı (c) do
        FCM ile her tahmin modelinin ilgili fonksiyon/kümedeki üyelik derecelerini belirle
        MFF'leri Denklemler ? kullanarak elde et
        MFF'lerin değerlendirme kriterlerini hesapla
        j = j + 1
    end while
    α = α + 0.1
end while
En iyi değerlendirme kriterine sahip MFF fonksiyonunu döndür
En iyi fonksiyonu kullanarak tahminleri hesapla
```

Algoritma

Adım 1. Kullanılacak olan öngörü yöntemleri ve parametre seçimleri yapılır; test ve eğitim verilerinin uzunlukları, küme sayısı, alpha-cut ve bulanık indeks değerleri belirlenir.

Adım 2. Öngörü yöntemlerinden elde edilen sonuçlar girdi matrisi olarak kullanılır ve kümeleme işleminin daha sağlıklı bir şekilde yürütülebilmesi için girdi matrisi standartlaştırılır.

$$X = [X_{ij}], i = 1,2,\dots,m\,;\; j = 1,2,\dots,ntest1 \tag{1}$$

$$Z = [Z_{ij}], i = 1,2,\dots,m\,;\; j = 1,2,\dots,ntest1 \tag{2}$$

$$Z_{ij} = \frac{X_{ij} - ort(X_{.j})}{ssapma(X_{.j})} \tag{3}$$

Adım 3. Eğitim verisi girdi matrisi olarak kullanılır ve girdi matrisi bulanık-c kümeleme yöntemi kullanılarak kümelenir. Elde edilen yöntemlere ait üyelik dereceleri, fonksiyonlar halinde ifade edilen yöntemlerin katsayıları olarak kullanılır.

$$MFFF_i(x) = \sum_{j=1}^{m} w_{ij}x_j \ , i = 1,2,..,c; x = (x_1, x_2, .., x_m) \quad (4)$$

$$w_{ij} = \frac{\mu_{ij}}{\sum_j \mu_{ij}} \quad (5)$$

burada $\boldsymbol{x_j, j}$. yönteme ait öngörüleri belirtirken, $\boldsymbol{w_{ij}, i}$. fonksiyona ait $\boldsymbol{j}$. yöntemin ağırlık katsayısını ifade etmektedir.

Adım 4. Eğitim verisi kullanılarak elde edilen ağırlık katsayıları kullanılarak, eğitim verisi için öngörüler aşağıdaki formüller yardımıyla elde edilir.

$$F_{ij} = MFFF_i(X_{.j}) \quad (6)$$

$$X_{.j} = \left[X_{1j}\ X_{2j} \ldots X_{mj}\right]^T, j = 1, 2, \ldots, ntest1 \quad (7)$$

Adım 5. Meta bulanık fonksiyonun seçimi, yani en iyi fonksiyonun bir değerlendirme kriterine göre belirlenmesi için 4. adım da elde edilen öngörüler RMSE değerlendirme kriteri kullanılarak hesaplanır. En küçük RMSE değerine sahip fonksiyon, nihai fonksiyon yani Meta Bulanık Fonksiyon (MFF) olarak seçilir.

Adım 6. Meta bulanık fonksiyon kullanılarak eğitim verisi için öngörüler aşağıdaki formüller yardımıyla elde edilir.

$$\hat{F}_j = MFFF_{best}\left(Y_{.j}\right), \ j = 1,2,\ldots ntest2 \quad (8)$$

$$Y = \left[Y_{ij}\right], i = 1,2, \ldots, m; j = 1,2, \ldots ntest2 \quad (9)$$

$$Y_j = \left[Y_{1j}\ Y_{2j} \ldots\ Y_{mj}\right]^T, \ j = 1, 2, \ldots, ntest2 \quad (10)$$

Uygulama

Uygulama olarak farklı metotlar yerine, tek bir yöntemden elde edilen farklı modeller birleştirilerek öngörü performansı arttırılmaya çalışılacaktır. Öngörü yöntemi olarak seçilen yöntem bozkurt optimizasyon temelli Geri Beslemeli 1.Tip Bulanık Fonksiyon (GWO-RFFs)' dur.

Uygulama verisi olarak İstanbul borsası (Bist100) 2009 ve 2012 yılları arası ilk 6 aylık kapanış değerleri seçilmiştir. AR için gecikme sayısı 1 ile 5 arasında, MA için gecikme sayısı 1 ile 2 arasında ve küme sayısı 2 ile 5 arasında değiştirilerek GWO-RFF yönteminden elde edilen 9 farklı modelin

son 15 gözlem için elde edilen öngörü değerleri girdi matrisi olarak kullanılmıştır. Eğitim veri setinin uzunluğu 8 olarak seçilmiş yani son 7 gözlem için öngörüler önerilen yöntem ile elde edilmiştir.

2009 Yılına Ait Veri seti

MFF yönteminde eğitimi için kullanılan 2009 yılına ait veri seti Tablo 1'de ve test veri seti ve buna ait RMSE değerleri Tablo 2 de verilmiştir.

Tablo 1 Eğitim veri seti; 2009 yılı için 9 farklı modelden elde edilen öngörüler

Gözlem	a	b	c	d	e	f	g	h
1	32843	32842.61	33061.31	33178.62	33356.24	33141.94	33134.36	33640.57
2	32806	32805.72	33159.03	33102.1	32907.74	32324.43	32840.27	33640.57
3	32203	32203.06	32531.04	32618.48	32452.1	31989.47	32633.73	33640.57
4	33043	33043.28	33338.92	33314.4	32699.27	32932.85	33150.3	33640.57
5	32829	32828.9	33031.01	33192.82	32940.28	33480.57	32977.84	33640.57
6	33095	33094.92	33356.06	33350.41	33084.76	33603.53	33440.59	33640.57
7	33485	33484.52	33651.2	33790.48	33347.12	33671.1	33856.18	33640.57
8	33666	33666.08	33794.65	33960.67	33661.69	33851.14	33797.19	32756.86

Tablo 2 Test veri seti; 2009 yılı için 9 farklı modelin öngörüleri ve RMSE değerleri

Gözlem	a	b	c	d	e	f	g	h
9	35140	35139.7	34926.15	35353.47	34559.35	34676.64	35000.92	32756.86
10	34721	34720.6	34547.22	35065.66	34999.72	35140.3	34907.47	32756.86
11	35015	35014.5	34887.24	35247.61	35042.18	35396.82	35255.89	32756.86
12	35408	35407.7	35108.24	35720.93	35257.93	35186.02	35096.48	32756.86
13	34861	34860.6	34727.41	35196.88	35186.01	34850.95	34857.23	32756.86
14	35169	35168.7	35002.28	35444.51	35103.95	34719.4	34657.6	32756.86
15	35021	35021	34844.72	35369.76	35103.37	34799.38	34952.59	32756.86
RMSE	344.91	344.963	340.929	460.1925	218.7582	225.8351	239.1376	2280.444
MAPE	0.0087	0.00868	0.008383	0.010318	0.004647	0.005424	0.006124	0.064814

MFF yöntemi, eğitim verisini girdi matrisi olarak kullanmış ve küme sayısını 2 ile 5 arasında, bulanık indeks değeri 1.5 ile 2.5 arasında 0.1 artış hızıyla değiştirerek en iyi MFF' yi aramıştır. MFF yönteminde, en iyi RMSE sonuçlarına, küme sayısı 5 ve bulanık indeks değeri 1.9 olduğunda

ulaşılmıştır. Bu durumda, Tablo 3'ten anlaşılacağı üzere en iyi öngörü sonuçları MFF_5'ten elde edilmiş ve böylece test verisi için kullanılacak olan fonksiyon MFF_5 olarak belirlenmiştir. Dolayısıyla gelecekteki öngörüler MFF_5 (Denklem 11) kullanılarak bulunacaktır.

Tablo 3 Eğitim veri seti için modellerden elde edilen ağırlıklar ve fonksiyonların RMSE değerleri

Model	MFF_1	MFF_2	MFF_3	MFF_4	MFF_5
a	0	0	0	0.322	0
b	0	0	0	0.322	0
c	0.309	0	0	0	0
d	0.31	0	0	0	0
e	0.091	0	0	0.275	0.208
f	0	0	0	0	0.695
g	0.291	0	0	0.08	0.097
h	0	1	0	0	0
RMSE	600.393	1105.708	2119.891	648.6833	**598.6337**
MAPE	0.01322	0.025324	0.057472	0.014624	**0.012033**

$$MFF_{best} = MFF_5 = 0.208 * e + 0.695 * f + 0.097 * g \quad (11)$$

Tablo 4' ten anlaşılacağı üzere test verisi için en iyi öngörü performansı da MFF_5'ten elde edilmiştir. Tablo 2 ve Tablo 4 incelendiğinde, en iyi öngörü sonuçlarına önerilen yöntemin 5. Fonksiyon kullanılarak ulaşıldığı görülmektedir.

Tablo 4 Önerilen yöntem için elde edilen öngörüler ve RMSE değerleri

Gözlem	MFF_1	MFF_2	MFF_3	MFF_4	MFF_5
9	35046.7	32756.9	31584.37	34968.9	34683.81
10	34853.7	32756.9	31484.08	34812.49	35088.48
11	35120	32756.9	32366.14	35041.54	35309.49
12	35308.1	32756.9	31546.65	35341.72	35192.25
13	34952.3	32756.9	33258.56	34950.06	34921.12
14	35048.3	32756.9	32908.01	35110.2	34793.23
15	35062.2	32756.9	32731.13	35038.23	34877.38
RMSE	258.05	2280.44	2823.868	273.8549	**190.9795**
MAPE	0.0062	0.06481	0.078817	0.006692	**0.004636**

2010 Yılına Ait Veri seti

MFF yönteminde eğitimi için kullanılan 2010 yılına ait veri seti Tablo 5'de bu set kullanılarak elde edilen model ağırlıkları ve karşılık gelen değerlendirme kriteri değerleri Tablo 6 da verilmiştir.

Tablo 5. 2010 yılı için eğitim veri seti

Model	1	2	3	4	5	6	7	8
a	52687	56448	56462	57976	57930	55748	56071	56978
b	52687	56448	56462	57976	57930	55748	56071	56978
c	52671	56424	56438	57902	57859	55718	56045	56949
d	51771	56908	56109	57735	57827	54773	55595	56677
e	54619	55443	55794	56291	56252	56070	55923	56545
f	56148	56266	55006	56694	56421	56587	58432	55572
g	55211	56391	55621	56566	57232	56706	57599	57132
h	52522	57061	57061	57817	57817	54791	54791	57061

MFF yöntemi, eğitim verisini girdi matrisi olarak kullanmış ve küme sayısını 2 ile 5 arasında, bulanık indeks değeri 1.5 ile 2.5 arasında 0.1 artış hızıyla değiştirerek en iyi MFF' yi aramıştır. MFF yönteminde, en iyi RMSE sonuçlarına, küme sayısı 3 ve bulanık indeks değeri 1.7 olduğunda

ulaşılmıştır. Bu durumda, Tablo 6'dan anlaşılacağı üzere en iyi öngörü sonuçları MFF_2'ten elde edilmiş ve böylece test verisi için kullanılacak olan fonksiyon MFF_2 olarak belirlenmiştir.

Tablo 6. 2010 yılı için modellere ait elde edilen ağırlıklar

Model	MFF_1	MFF_2	MFF_3
a	0.333	0	0
b	0.333	0	0
c	0.333	0	0
d	0	0	0.492
e	0	0.333	0
f	0	0.346	0
g	0	0.321	0
h	0	0	0.508
RMSE	1885.868	**1348.111**	2112.667
MAPE	0.025088	**0.019926**	0.031

Dolayısıyla test veri seti için öngörüler (Tablo 7) MFF_2 (Denklem 12) kullanılarak bulunacaktır.

$$MFF_{best} = MFF_2 = 0.333 * e + 0.346 * f + 0.321 \quad (12)$$

Tablo 7. 2010 yılı için önerilen yöntemden son 7 gözlem için elde edilen öngörüler

Gözlem	MFF_1	MFF_2	MFF_3
9	54432.63	54539.98	53408.48
10	54093.77	54537.92	53529.43
11	54541.13	53762.23	53867.36
12	52263.02	52552.81	52115.75
13	54085.72	53928.37	53819.5
14	54480.81	54046.16	53852.64
15	55220.8	54722.97	55002.88
RMSE	1216.588	**978.7122**	1244.967
MAPE	0.018391	**0.014846**	0.021177

Tablo 7' den anlaşılacağı üzere test verisi için de en iyi öngörü performansı MFF_2'ten elde edilmiştir. Tablo 8 incelendiğinde, MFF_2 modellerin bireysel performansından da daha iyi öngörü sonuçlarına ulaştığı görülmektedir.

Tablo 8. 2010 yılı için test veri seti ve modellere ait RMSE ve MAPE değerleri

Model	9	10	11	12	13	14	15	RMSE
a	54450	54112	54558	52257	54104	54498	55234	1221.1
b	54450	54112	54558	52257	54103	54498	55234	1221.1
c	54398	54058	54507	52275	54050	54446	55195	1208.2
d	53543	53789	54475	51697	54378	54445	55221	1360.8
e	55432	54496	54630	53635	53584	53961	54361	1198.2
f	52985	54419	53167	51463	54183	54071	55613	1312
g	55291	54710	53506	52605	54010	54108	54140	984.53
h	53278	53278	53278	52522	53278	53278	54791	1273.9
MFF_{best}	54539.98	54537.92	53762.23	52552.81	53928.37	54046.16	54722.97	**979**

2011 Yılına Ait Veri Seti

MFF yönteminde eğitimi için kullanılan 2011 yılına ait veri seti Tablo 9'da bu set kullanılarak elde edilen model ağırlıkları ve karşılık gelen değerlendirme kriteri değerleri Tablo 10' da verilmiştir.

Tablo 9. 2011 yılı için eğitim veri seti

Model	1	2	3	4	5	6	7	8
a	67956	67260	65643	66535	64585	65418	65385	63733
b	67956	67260	65643	66535	64585	65418	65385	63733
c	67987	67387	65828	66423	64644	65291	65106	63647
d	67990	67104	65662	66567	64012	65549	64752	63735
e	67256	66768	65619	65591	64776	64607	64866	63912
f	67130	66592	65452	65160	64524	64096	64413	63560
g	67561	66531	65801	65710	63977	64722	63475	63406
h	67627	67627	65998	65998	64368	65998	65998	64368

MFF yöntemi, eğitim verisini girdi matrisi olarak kullanmış ve küme sayısını 2 ile 5 arasında, bulanık indeks değeri 1.5 ile 2.5 arasında 0.1 artış hızıyla değiştirerek en iyi MFF'yi aramıştır. MFF yönteminde, en iyi RMSE sonuçlarına, küme sayısı 4 ve bulanık indeks değeri 2 olduğunda ulaşılmıştır. Bu durumda, Tablo 10'dan anlaşılacağı üzere en iyi öngörü sonuçları MFF_1'ten elde edilmiş ve böylece test verisi için kullanılacak olan fonksiyon MFF_1 olarak belirlenmiştir.

Tablo 10. 2011 yılı için modellere ait elde edilen ağırlıklar

Model	MFF_1	MFF_2	MFF_3	MFF_4
a	0	0	0	0.281
b	0	0	0	0.281
c	0	0	0	0.241
d	0.062	0.158	0.116	0.166
e	0.435	0	0	0.031
f	0.503	0	0	0
g	0	0.842	0	0
h	0	0	0.884	0
RMSE	**824.773**	838.531	1293.095	1146.721
MAPE	**0.01159**	0.01059	0.017649	0.015037

Dolayısıyla test veri seti içim öngörüler (Tablo 11), MFF_1 (Denklem 13) kullanılarak hesaplanacaktır.

$$MFF_{best} = MFF_1 = 0.062 * d + 0.435 * e + 0.503 * f \quad (13)$$

Tablo 11. 2011 yılı için önerilen yöntemden son 7 gözlem için elde edilen öngörüler

Gözlem	MFF_1	MFF_2	MFF_3	MFF_4
9	63521.58	62881.27	62826.99	63402.91
10	63029.67	62651.74	62782.92	63324.1
11	63495.86	63432.99	64405.51	64652.88
12	62942.39	63415.65	64274.68	63591.38
13	63492.22	63202.83	64353.76	63900.56
14	62726.81	62454.24	62694.33	62540.31
15	62628.96	61437.5	61186.95	61778.8
RMSE	**925.1679**	1072.114	1357.744	1043.284
MAPE	**0.012465**	0.013825	0.019215	0.01461

Tablo 11' den anlaşılacağı üzere test verisi için de en iyi öngörü performansı MFF_2'ten elde edilmiştir. Tablo 12 incelendiğinde, MFF_1 modellerin bireysel performansından da daha iyi öngörü sonuçlarına ulaştığı görülmektedir.

Tablo 12. 2011 yılı için önerilen yöntemden son 7 gözlem için elde edilen öngörüler

Model	9	10	11	12	13	14	15	RMSE
a	63299	63210	64561	63609	63755	62407	61492	1057.6
b	63299	63210	64561	63609	63755	62408	61492	1057.8
c	63615	63782	64969	63570	64074	62966	62465	1065.4
d	63501	63122	64690	63566	64246	62361	61782	1139.7
e	63121	62926	63662	63571	63339	62618	61649	986.08
f	62598	62356	63080	63263	62956	62324	61212	1202.1
g	63525	63012	63273	62826	63351	62795	62787	935.9
h	62738	62738	64368	64368	64368	62738	61109	1396.4
$\boldsymbol{MFF_{best}}$	63522	63030	63496	62942	63492	62727	62629	**925**

2010 – 2012 yılları arası Bist100 veri setleri göz önünde bulundurulduğunda, MFF birleştirilen modellerden daha iyi öngörü

performansı gösterdiği ve GWO-RFFs yönteminin öngörü performansını arttırdığı gözlemlenmiştir.

SONUÇ

Bu çalışma kapsamında, GWO-RFFs öngörü modelinin farklı parametre spesifikasyonlarından elde edilen öngörülerin birleştirilmesi yoluyla oluşturulan Meta Bulanık Fonksiyonlar (MFF) yöntemi detaylı olarak ele alınmış ve zaman serisi verileri üzerinde uygulamaları gerçekleştirilmiştir. MFF yaklaşımı, klasik tek model tabanlı tahmin yöntemlerinin ötesine geçerek, çeşitli model çıktılarının bulanık kümeleme algoritmalarıyla analiz edilmesini ve performansı yüksek öngörü fonksiyonlarının elde edilmesini hedeflemektedir.

2009, 2010 ve 2011 yıllarına ait BIST100 veri setleri üzerinde gerçekleştirilen deneysel analizler, MFF yaklaşımının bireysel tahmin yöntemlerinden daha üstün sonuçlar verdiğini ortaya koymuştur. Özellikle RMSE ve MAPE gibi değerlendirme kriterlerine göre yapılan karşılaştırmalarda, MFF ile elde edilen tahminlerin hata oranları sistematik biçimde daha düşük bulunmuştur. Bu durum, MFF'nin veri yapısına duyarlı şekilde optimize edilebilmesi ve veri setine en uygun alt model gruplarını seçebilmesiyle açıklanabilir.

Uygulama bulguları, kümelenen yöntemlerin üyelik derecelerine göre ağırlıklandırılması sayesinde daha dengeli ve genellenebilir öngörülerin elde edilebildiğini göstermiştir. Ayrıca, her yıl için farklı sayıda küme ve bulanık indeks parametresi kombinasyonları test edilerek, en düşük hata oranına sahip fonksiyonlar seçilmiş ve bu esnek yapı yöntemin güçlü yönlerinden biri olarak öne çıkmıştır.

Sonuç olarak, MFF yöntemi sadece model çeşitliliğini avantaja çeviren bir yaklaşım olmakla kalmayıp, aynı zamanda belirsizliğin hâkim olduğu tahmin problemleri için etkili ve güvenilir bir alternatif sunmaktadır. Gelecekte yapılacak çalışmalarla bu yöntemin farklı veri türlerine (örneğin, panel veriler, sınıflandırma problemleri, yüksek boyutlu veri setleri) uygulanması ve farklı meta-sezgisel algoritmalarla entegrasyonu potansiyel araştırma alanları arasında yer almaktadır. Ayrıca, etki büyüklüğü tahminleri, erken uyarı sistemleri ve karma skor indekslerinin oluşturulması gibi alanlarda MFF'nin uygulanabilirliği de önemli bir araştırma perspektifi sunmaktadır.

Kaynaklar

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3–8. https://doi.org/10.3102/0013189X005010003

Tak, N. (2018). Meta fuzzy functions: Application of recurrent type-1 fuzzy functions. *Applied Soft Computing*, 73, 1–13. https://doi.org/10.1016/j.asoc.2018.07.056

Tak, N. (2020). Meta fuzzy index functions. *Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics*, 69(1), 654–667

Tak, N. (2021). Forecast combination with meta possibilistic fuzzy functions. *Information Sciences*, 560, 168–182. https://doi.org/10.1016/j.ins.2021.01.024

Tak, N. (2021). Meta fuzzy functions based feed-forward neural networks with a single hidden layer for forecasting. *Journal of Statistical Computation and Simulation*, 91(14), 2800–2816. https://doi.org/10.1080/00949655.2020.1806782

Cevik, F. C., Gever, B., Tak, N., & Khaniyev, T. (2023). Forecast combination approach with meta-fuzzy functions for forecasting the number of immigrants within the maritime line security project in Turkey. *Soft Computing*, 27(5), 2509–2535. https://doi.org/10.1007/s00500-022-07800-7

Yabacı Tak, A., & Ercan, İ. (2023). Ensemble of effect size methods based on meta fuzzy functions. *Engineering Applications of Artificial Intelligence*, 119, 105804. https://doi.org/10.1016/j.engappai.2022.105804

Tak, N., Egrioglu, E., Bas, E., & Yolcu, U. (2021). An adaptive forecast combination approach based on meta intuitionistic fuzzy functions. *Journal of Intelligent & Fuzzy Systems*, 40(5), 9567–9581. https://doi.org/10.3233/JIFS-202021

Gök, A., & Tak, N. (2023). Dating currency crisis and assessing the determinants based on meta fuzzy index functions. *Computational Economics*, 61(3), 1225–1250. https://doi.org/10.1007/s10614-022-10243-9

Tak, N., & Gök, A. (2022). Dating currency crises and designing early warning systems: Meta-possibilistic fuzzy index functions. *International Journal of Finance & Economics*, 27(3), 3773–3790. https://doi.org/10.1002/ijfe.2350

Tak, N., Evren, A. A., Tez, M., & Egrioglu, E. (2018). Recurrent type-1 fuzzy functions approach for time series forecasting. *Applied Intelligence*, 48(1), 68–77. https://doi.org/10.1007/s10489-017-0958-4

ÇİFT MAKİNE ÖĞRENİMİ VE UYGULAMALARI

Sevgi Abdalla[**]

Giriş

Son yirmi yılda, yapay zekâ kavramının hayatımıza girmesi beraberinde verilerin analizinde devrim yaratan makine öğrenimi disiplinini doğurmuştur. Günümüzde ise makine öğrenimi sağlık, finans ve hatta edebiyat gibi ana alanlarla birlikte diğer tüm sektörlerde kullanılarak bilgiye ulaşma konusunda tatmin edici ve şaşırtıcı çıktılar üretmeye devam etmektedir.

Çeşitli makine öğrenimi yaklaşımları arasında, 2018 yılında Chernozhukov ve arkadaşları tarafından yüksek boyutlu verilerde nedensellik çıkarımını geliştirmek için tasarlanmış *çift makine öğrenimi* (ÇMÖ) literatüre tanıtılmıştır. Söz konusu yaklaşım nedensellik, veri heterojenliği ve tahmin ile ilgili karmaşık sorunları ele alma potansiyeli nedeniyle özellikle dikkat çekmektedir.

Bu bölümde, çift/yanlılıktan arındırılmış makine öğrenimi yaklaşımının teorik alt yapısı, önemi, nedensellik çıkarımını iyileştirme, yanlılığı azaltma, karmaşık veri kümelerini ele alma ve çeşitli uygulamalarda daha güvenilir tahminler sunma yeteneği üzerinde durulmuş ve vurgulanmıştır.

Çift makine öğrenimi, genellikle geleneksel istatistiksel modellerin zorlandığı yüksek boyutlu verilerle başa çıkmak için tasarlanmış iki adımlı bir süreçtir. Makine öğrenimi algoritmaları—Random Forest (Rassal Orman), Lasso ve Boosting gibi—verileri önişlemden geçirmek için kullanılır. Böylece, nedensel etkinin doğru ve tutarlı bir şekilde tahmin edilmesini sağlar. Bu uyarlanabilirlik, ÇMÖ'yü genellikle sıkı varsayımlar gerektiren ve büyük veri kümelerinde etkinlik kaybına sebep olan geleneksel tekniklerden ayırır (Athey & Imbens, 2019). Sonuçta yaklaşım, yanlılığı azaltmak ve nedensel ilişkilerin tahminlerini iyileştirmek için makine öğrenimi tekniklerini nedensel çıkarımla birleştirir.

[**] Dr. Öğr. Üyesi, Eskişehir Osmangazi Üniversitesi, Fen Fakültesi, İstatistik, Eskişehir, sayhan@ogu.edu.tr, ORCID: orcid.org /0000-0003-4177-5868

Çift Makine Öğreniminin Önemi ve Avantajları

Modern dünyada, verinin bol ve genellikle yüksek boyutlu (örneğin; çok sayıda değişken içeren büyük veri setleri) olduğu bir ortamda, geleneksel ekonometrik yöntemler, veri karmaşıklığıyla baş etme ve etkin sonuç çıkarımı için veri işleme konusunda zorlanmaktadır. Geleneksel istatistiksel yöntemler, tahmin edici değişkenlerin sayısının nispeten az olduğunu veya değişkenler arasındaki ilişkiler için belirli varsayımların (doğrusallık gibi) geçerli olduğunu kabul eder. Pratikte ise bu varsayımlar nadiren karşılanır. Bununla birlikte, söz konusu algoritmalar doğal olarak nedensel akıl yürütmeyi içermedikleri gibi tipik olarak nedensel etkileri doğrudan tahmin etmek için tasarlanmamışlardır.

Öte yandan, geleneksel yöntemlerle makine öğrenimi yöntemlerinin güçlerini birleştirerek sağlanan esneklik, ÇMÖ'nun karmaşık veri yapılarını aşırı uyumdan sakınarak yüksek boyutlu ortamlarda oldukça iyi sonuçlar üretmesine imkân verir (Bach ve ark., 2021). Ayrıca değişkenler arasındaki karmaşık, doğrusal olmayan ilişkileri modelleyebilme becerisi sunar (Kennedy & Kolesar, 2021).

Söz konusu yaklaşımı avantajları bakımından dört farklı boyutta ele almak uygun olacaktır.

A1. Confounding/Karıştırıcı Değişkenler İçin Kontrolün Sağlanması

Nedensellik analizinde karıştırma, üçüncü bir değişkenin hem denemeyi hem de sonucu etkileyerek, deneme etkisinin tahminini yanıltabilir bir durumuna karşılık gelir (Hernán ve Robins, 2020).

ÇMÖ, karıştırıcı değişkenleri modellemek ve etkilerini ortadan kaldırmak için makine öğrenimini kullanarak bu görevini etkili bir şekilde gerçekleştirmektedir. Böylece ÇMÖ, karıştırıcıların varlığında bile nedensel etkiyi daha doğru ve tarafsız bir şekilde tahmin edebilme yeteneğine sahiptir.

A2. Tahminde Doğruluk

Deneme etkilerini tahmin etmede kullanılan geleneksel istatistik yöntemleri, genellikle yüksek boyutlu veri durumunda ve analize ortak değişken/değişkenlerin dahil edilmesi durumunda da kesin olmayan tahminlere yol açmaktadır.

A3. Esneklik ve Model Bağımsızlığı

ÇMÖ'nün en önemli avantajlarından biri modelden bağımsız doğasıdır. Verilerin işlevselliği hakkında belirli varsayımlara (örneğin, doğrusallık) dayanan klasik yöntemlerin aksine, ÇMÖ, değişkenler arasında önceden tanımlanmış ilişkileri gerektirmez. Bu esneklik, araştırmacıların ÇMÖ'yü çeşitli ortamlarda, temel varsayımlarla ilgilenmeden uygulayabilmelerine olanak tanır, bu da onu daha karmaşık gerçek dünya problemlerinin çözümüne daha uyumlu hale getirir.

A4. Deneme Etkisi Tahminindeki Yanlılığın Azaltılması

Gözlemsel verilerden nedensel etkileri tahmin etmedeki büyük bir zorluk, ihmal edilen değişkenler veya ölçüm hatalarından dolayı ortaya çıkan yanlılıktır. ÇMÖ çerçevesi içinde kullanılan makine öğrenimi teknikleri, potansiyel karıştırıcıların büyük veri yığınlarını daha iyi yönetmek için bu tür yanlılığı en aza indirebilir.

Yaklaşımın avantajlarına rağmen ÇMÖ uygulanmasında bazı zorluklar da söz konusudur. Bu zorluklardan biri, makine öğrenimi algoritmalarının verinin eğitimindeki hesaplama karmaşıklığıdır. Değişkenleri seçme, model uydurma ve nedensel etkileri tahmin etme süreci, özellikle büyük veri kümelerinde önemli hesaplama kaynakları gerektirmektedir. Ayrıca ÇMÖ, büyük ölçüde kullanılan makine öğrenimi algoritmalarının kalitesine bağlıdır. Doğru olmayan model seçimi veya yanlış tanımlama, yanlı sonuçlara yol açabilmektedir.

Ek olarak, ÇMÖ karıştırıcı değişkenlerin kontrolüne yardımcı olurken, ölçülmeyen veya analize dahil edilmeyen gözlemlenememiş faktörler sorununu tam olarak çözemediği için elde edilen sonuçlar yanlı olabilmektedir. Dolayısıyla, yardımcı değişkenler veya duyarlılık analizi gibi ek stratejilerin kullanılması gerekebilir.

Literatür İncelemesi: Çift Makine Öğrenimi ve Uygulamaları

Makine öğreniminin gücünü ekonometrik tekniklerin titizliğiyle birleştirmek için geliştirilen yaklaşım, genellikle çalışma konusunun deneysel olmayan verilerden nedensel ilişkiler çıkarmak olduğu ekonomi, epidemiyoloji ve sosyal bilimler gibi alanlarda önemlidir. İzleyen kısımda, söz konusu yaklaşımın gelişimini, temel katkılarını ve farklı alanlardaki uygulamalarına ilişkin literatür özetlenmektedir.

ÇMÖ ile ilgili en temel çalışma Chernozhukov ve arkadaşları (2018) tarafından önerilmiş olup, tahminlerin tutarlılığını ve asimptotik normalliğini sağlarken nedensel çıkarım bağlamında makine öğrenimini kullanmak için genel bir çerçeve sunmuştur. Özellikle gözlem sayısına göre çok sayıda kestirici içeren ortamlarda esnekliği ve verimliliği nedeniyle ilgi çekmektedir. Sonuç olarak, yöntemin örneklem büyüklüğü arttıkça, tahminlerin gerçek nedensel etkiye yakınsadığını ve normal bir dağılım izlediğini göstermişlerdir.

Farbmacher ve arkadaşları (2020) çalışmalarında, yüksek boyutlu ortamlarda gözlemlenen karıştırıcıları kontrol etmek için nedensellik analizi ile ÇMÖ'yi birleştirerek, dolaylı ve doğrudan etkiler için sağlam tahminciler sağlamaktadır. Bu yöntemin etkinliği, sağlık sigortası kapsamı etkilerine yapılan deneysel bir uygulama ile gösterilmiştir. Knaus (2020) tarafından yazılmış bir makalede, bir programın değerlendirilmesinde çift makine öğrenmesi yöntemleri incelenmiş ve genişletilmiştir. Zhang ve arkadaşları (2021) tarafından yapılan araştırmada, yağış ürünlerini ve uydu bazlı ölçümleri birleştirmek için yenilikçi bir çift makine öğrenimi yaklaşımı önerilmektedir. Söz konusu yaklaşım yağış olaylarını tespit etmede diğer yöntemlere göre üstün performans sergilemektedir. Çalışmada karıştırıcı değişkenlerinin düzeltilmesine odaklanılmış ve bireyselleştirilmiş etki tahminlerini istikrara kavuşturmak için normalize edilmiş DR-öğrenici önerilmiştir. Bir diğer çalışma ise Fingerhut ve arkadaşları (2022) tarafından, türetilmiş ve gerçek veriler üzerinde sayısal deneyler yoluyla tahmin yanlılığını azaltmayı ve deneysel performansı iyileştirmeyi amaçlayan ÇMÖ çerçevesi içinde derin sinir ağları için koordineli bir öğrenme algoritması araştırılmıştır. Felderer ve arkadaşlarının (2023) yaptıkları çalışmada, anket istatistiklerine, ilgili parametrelerin yaklaşık olarak yanlı olmayan tahminlerini veren çift makine öğrenimi yaklaşımını tanıtmışlardır ve bu yaklaşımın yüksek boyutlu panel ortamında anketlere yanıt verilmemesi durumunda analiz etmek için nasıl kullanılabileceğini göstermişlerdir. Cohrs ve arkadaşları (2024), hibrit modellerde nedensel etkileri tahmin etmek için ÇMÖ kullanan yeni bir yaklaşım sunmuş, özellikle karbondioksit akışlarına uygulanarak, nedensellik parametrelerini tahmin etmede derin öğrenme yaklaşımlarının üstünlüğünü göstermişlerdir. Son olarak, başka bir makalede, Ahrens ve arkadaşları (2024) yapısal parametreleri tahmin etmek için ÇMÖ ile yığınlama (staking) yöntemlerini entegre edip bu yaklaşımın bilinmeyen fonksiyonel formlara karşı daha dayanıklı olduğunu

göstermişlerdir. Ayrıca çalışmada, Stata ve R'de yazılım uygulamaları sunulmuştur.

Çift Makine Öğrenimi

Bu bölümde, ÇMÖ yaklaşımının temel kavramları, matematiksel yapısı, uygulama adımları ve geleneksel istatistiksel yöntemlere göre sunduğu avantajları açıklanmıştır.

a. Problemin Tanımı ve Amaçlar

ÇMÖ'nün matematiksel alt yapısını anlamak için öncelikle istatistiksel öğrenme, nedensel çıkarım ve yüksek boyutlu ortamlarda düzenlileştirmedeki temel kavramların sıkı bir şekilde kavranması gerekir.

Nedensel çıkarım bağlamında amaç tipik olarak bir deneme değişkeni D 'nin, bir sonuç değişkeni Y üzerindeki nedensel etkisini tahmin ederken, karıştırıcı olabilecek yüksek boyutlu ortak değişkenleri (X) kontrol etmektir. Birçok gerçek yaşam probleminde ortak değişkenlerin sayısı oldukça fazla olabilir ve sıradan en küçük kareler (Ordinary Least Squares) veya araç değişkenler (Instrumental Variables) gibi geleneksel ekonometrik yöntemler aşırı uyum veya eş doğrusallık gibi sorunlar nedeniyle etkinlik kaybına hatta uygulanamaz hale gelebilir. Bu tür yüksek boyutlu ortamlardaki zorluk, verilere aşırı uyumdan koruyarak $Y,\ D\ ve\ X$ arasındaki karmaşık ilişkileri hesaba katmaktır. ÇMÖ, ilgilenilen asıl parametre dışında hesaba katılmak zorunda olunan (**nuisance**) parametreleri tahmin etmek için makine öğrenimi algoritmalarını kullanarak bu zorluğun üstesinden gelmektedir.

b. Çift Makine Öğreniminin Anahtar Kavramları

Nedensellik Çıkarımı ve Tahmin

ÇMÖ'in nedensellik çıkarımındaki rolünü anlamak için, ele aldığı temel problemle başlamak uygun olacaktır. Bu problem; *gözlemsel çalışmalarda deneme etkilerini tahmin etmektir*.

Birçok gerçek dünya probleminde, bir deneme veya müdahalenin (örneğin, ilaç, politika) bir sonuç üzerindeki etkisini (örneğin, bedensel iyileşme, satışlar) anlamak istediğimizi varsayalım. Bu, Ortalama Deneme Etkisi (ATE) veya Denenen üzerindeki Ortalama Deneme Etkisi (ATT) gibi nedensellik parametrelerini tahmin etmeyi gerektirir. Bir deneme olayı D, bir sonuç değişkeni Y ve bir ortak değişken vektörü X düşünelim. Nedensellik

çıkarımında, genellikle deneme D'nin sonuç Y üzerindeki etkisini tahmin etmek isteriz, bu da şu şekilde ifade edilebilir:

$$ATE = E[Y(1) - Y(0)] \qquad (1)$$

Söz konusu eşitlikte, Y, $Y(1)$ deneme altındaki potansiyel sonucu ve $Y(0)$ kontrol altındaki (deneme yok) potansiyel sonucu temsil eder.

Rubin (1974) tarafından geliştirilen sonuçların yapısı, nedensellik çıkarımında merkezi bir öneme sahiptir. Ancak, burada birkaç zorlukla karşılaşılmaktadır. Bu zorluklardan ikisi; karıştırıcı değişkenler (confounding variables) ve yüksek boyutlu veriyle çalışmaktır.

Karıştırıcı Değişkenler: Ortak değişkenler X, hem D hem de Y sonuç değişkenini etkileyebilmektedir. Bu da söz konusu değişkenlerin düzgün bir şekilde kontrol edilmediği durumlarda nedensellik etkilerinin tahmininde yanlılığa yol açabilmektedir.

Yüksek Boyutlu Veri durumunda, ortak değişken sayısı X'in fazla olması, tahmin sürecini karmaşıklaştırır ve bu karmaşıklık deneme etkilerini güvenilir bir şekilde tahmin etmeyi zorlaştırır.

ÇMÖ'nün amacı, bu zorlukları özellikle yüksek boyutlu veriyle çalışıldığı durumunda ele alarak nedensellik parametreleri (örneğin; ATE) için tutarlı bir tahminci sağlamaktır.

Nuisance Parametreler ve Tahmini

Nedensellik çıkarımında ilgilenilen sonuç tipik olarak hem deneme hem de çeşitli ortak değişkenlere bağlıdır. Bununla birlikte, gerçek veri oluşturma süreci genellikle bilinmediğinden, yaklaşık olarak tahmin etmek için bir modele güvenmemiz gerekir.

X ortak değişkenlerinin varlığında, sonuç Y ile deneme D değişkenleri arasındaki ilişki, bu ortak değişkenlerin etkileri tarafından karıştırılır. Bu karıştırıcı değişkenleri kontrol etmek için iki nuisance parametrenin doğru bir şekilde tahmin edilmesi esas rol oynamaktadır. Bu parametreler; (1) sonuç değişkeninin koşullu beklenen değeri ve (2) eğilim skoru (propensity score)'dur.

* ***Sonuç değişkeni Y'nin koşullu beklenen değeri.*** X ortak değişkenleri verildiğinde Y 'nin koşullu beklentisi,

$$m(X) = E[Y|X] \qquad (2)$$

* ***Eğilim Skoru (Propensity Score).*** X ortak değişkenleri verildiğinde D 'nin koşullu beklentisi,

$$g(X) = E[D|X] \qquad (3)$$

olarak gösterilir.

Lasso regresyonu (Tibshirani, 1996), Random Forest (Rassal Orman) (Breiman, 2001) ve Gradyan Artırma Makineleri (Friedman, 2001) gibi makine öğrenimi algoritmaları, büyük veri kümelerini işleme, doğrusal olmayan ilişkileri yakalama ve otomatik olarak önemli özellikleri seçme yeteneklerinden dolayı yaygın olarak kullanılmaktadır. *Amaç, tahmin doğruluğunu korurken yüksek boyutlu verilere aşırı uyum sağlamadan bu parametreleri tahmin etmektir.*

Artıklaştırma (Residualization)

Bir değişkenin, başka bir değişkenin etkisinden arınmış kısmını bulma veya bir değişkenin, başka değişkenler tarafından açıklanan kısmını ayırıp, geriye kalan "açıklanamayan" kısmı (residual) elde etme işlemine residualization ismi verilir.

İlgilenilen parametre dışındaki (nuisance) parametreler tahmin edildikten, sonraki adım, sonuç ve deneme değişkenlerini artıklaştırmaktır. Bu kısım, (Y) ve (D)'deki (X) tarafından açıklanmayan değişimi ayırt etmek için gözlemlenen değişkenlerden tahmin edilen nuisance bileşenlerini çıkarmayı içermektedir.

Bu süreç, ortak değişkenlerin etkisini hem sonuçtan hem de denemeden etkili bir şekilde kaldırarak (D) ve (Y) arasındaki nedensel ilişkinin daha iyi tanımlanmasını sağlar. Aşağıda verilen eşitlik 4 ve 5'te söz konusu ilişkiler matematiksel olarak ifade edilmiştir.

$$\tilde{Y} = Y - \hat{m}(X) \qquad (4)$$

$$\tilde{D} = D - \hat{g}(X) \qquad (5)$$

Parametre Tahmini

Sonuç ve deneme değişkenleri artıklaştırıldıktan sonra, ilgilendiğimiz nedensellik parametresi (ATE), artıklaştırılmış veriler kullanılarak tahmin edilir. En yaygın yaklaşım, artıklaştırılmış değişkenlerin çarpımının örneklem ortalamasını hesaplamaktır ve formülü eşitlik 6'da gösterilmiştir.

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Y_i \widetilde{D}_i \qquad (6)$$

Bu tahmin edici, nuisance parametre tahminlerinin tutarlı olduğu varsayıldığında, ATE için yansızdır ve tutarlıdır denir.

c. Çift Makine Öğreniminin Asimptotik Özellikleri

ÇMÖ'nün temel avantajlarından biri, asimptotikliği garanti etmesidir. Bu özelliği, tahmin edicinin örneklem hacmi büyüdükçe gerçek nedensellik parametresine yakınsadığını garanti etmesi anlamına gelir. ÇMÖ algoritmasının tahmin edicilerinin asimptotik özellikleri aşağıda özetlenmiştir.

Tutarlılık: Belirli düzenlilik koşulları altında (örneğin, makine öğrenimi modellerinin doğru spesifikasyonu ve nuisance fonksiyonlarının sınırlı olması), ÇMÖ tahmincisi gerçek nedensel etki için tutarlıdır. Yani, n örneklem boyutu arttıkça ($n \to \infty$), tahminci ($\hat{\theta}$) gerçek nedensel parametre (θ) 'ya yakınsamaktadır.

Asimptotik Normallik: ÇMÖ tahmin edicisi, düzenli koşullar altında asimptotik normal dağılımı takip eder. Bu, nedensel parametre için hipotez testi ve güven aralıkları oluşturma gibi çıkarımlara olanak tanır:

$$\sqrt{n}\left(\hat{\theta} - \theta\right)\backslash x \to \mathcal{N}(0, \sigma^2) \qquad (7)$$

Burada σ^2, söz konusu tahmincinin asimptotik varyansıdır. Bu sonuç, nuisance parametrelerini tahmin etmek için kullanılan makine öğrenimi modelleri parametrik olmasa bile geçerlidir.

Çift Makine Öğreniminin Dezavantajları

Avantajlarına rağmen, Çift Makine Öğrenimi'nin uygulanmasında birkaç hususa dikkat edilmesi gerekir.

Makine Öğrenimi Algoritması Seçimi: ÇMÖ'nün performansı, nuisance parametrelerini tahmin etmek için makine öğrenimi modelinin seçimine

bağlıdır. Random Forest (Rassal Orman) veya Sinir Ağları gibi esnek modeller karmaşık ilişkileri yakalayabilirken, doğru bir şekilde ayarlanmadıklarında veriyi aşırı uyumdan kurtaramamaktadır. Böyle durumlarda ise belirli ayarlarda düzenleme teknikleri (örneğin LASSO) veya daha basit modeller tercih edilebilir.

Boyut: ÇMÖ yüksek boyutlu ortak değişkenleri ele almak için tasarlanmış olsa da, problemin boyutu yine de zorluklar çıkarabilmektedir. Ortak değişken sayısı p, örneklem boyutuna (n) göre çok büyükse, makine öğrenimi modelleri de güvenilir tahminler üretemeyebilir. Bu durumlarda boyut azaltma teknikleri (örneğin, Temel Bileşenler Analizi) veya özellik seçimi gerekli olabilir.

Hesaplama Karmaşıklığı: Makine öğrenimi modellerini eğitmek, özellikle büyük veri setleri ile hesaplama açısından maliyetli olabilir. ÇMÖ'yi verimli bir şekilde uygulamak, paralel işlemeyi veya özel yazılım platformlarının kullanımını gerektirebilir.

Çift Makine Öğrenimi Algoritmasının Örnek Bir Problem Üzerinde Gösterimi

Örnek Problem. Çalışanlara verilen kurum içi eğitimin, verimliliği arttırıcı etkileri üzerine bir araştırma yapıldığı varyasılsın. Söz konusu problemin çözümüne ilişkin ÇMÖ algoritması aşağıdaki gibi uygulanmaktadır.

Girdi:

- S= {X, D, Y} veri setini tanımlasın
- X: Ortak değişkenler (örneğin, yaş, iş tecrübesi, öğrenim düzeyi)
- D: Deneme değişkeni (örneğin, kursa katılım durumu)
- Y: Sonuç değişkeni (örneğin, verimlilik)

Çıktı:

- Etki tahmini: Üzerinde durduğumuz denemenin sonucunu nasıl etkilediği

Adımlar:

1. S veri setini, eğitim ve test veri seti olarak ayır
2. Model Eğitimi:

Tahmin/Performans Modeli (C1):

- Girdi: X Ortak değişkenleri ve D'yi kullanarak Y'yi tahmin et
- Model: Öğrenme algoritmalarından birini kullan (örneğin, linear regression, random forest vs.)
- Çıktı: Y'nin tahmin değerleri ve artık değerlerini hesapla

Eğilim (Propensity) Skoru Modeli (C2):

- Girdi: X ortak değişkenlerini kullanarak D'yi tahmin et
- Model: Öğrenme algoritmalarından biriyle deneme olasılığını tahmin et
- Çıktı: D'nin tahmin değerleri ve artık değerlerini hesapla

3. Artık Değerlerin Hesaplaması:

Y_artık= Y - C1(Y | X, D)

D_artık = T - C2(T | X)

4. İkinci Aşama Modeli:

- Regresyon Modeli: Y_artık ~ D_artık Y artık değerleriyle D_artık modellerini tahmin eden bir model oluştur.
- Bu regresyon modeli, tahmin edilmek istenen nedensel etkiyi vermektedir.

5. Sonuçların Yorumlanması:

- Regresyon katsayısını incele ve deneme etkisini yorumla.

6. Geçerlilik Kontrolleri:

- Modelin sağlamlığını test eden geçerlilik analizlerini gerçekleştir.

SONUÇ

ÇMÖ, makine öğrenimini nedensel çıkarımla birleştirerek yüksek boyutlu verilerde nedensel etkileri tahmin etmek için güçlü bir araç olarak ortaya çıkmıştır. Birçok karıştırıcı değişkenin varlığında bile nedensel ilişkilerin tahmin edilmesine izin veren ve geleneksel ekonometrik tekniklerin ortaya çıkardığı birçok zorluğu çözen bir yaklaşım sunmaktadır. ÇMÖ, Chernozhukov ve diğerleri (2018) tarafından tanıtılmasından bu yana, heterojen deneme etkilerini, parametrik olmayan modelleri ve çapraz doğrulama tekniklerini entegre ederek tahmin iyileştirme amacına hizmet eden ve ekonomiden sağlık ve pazarlamaya kadar çeşitli alanlarda kullanılan önemli bir yaklaşım olmaktadır. Hesaplama gücü arttıkça ve makine öğrenimi algoritmaları gelişmeye devam ettikçe, ölçeklenebilirliğini, esnekliğini ve sağlamlığını geliştirmeye devam eden çalışmalarla ÇMÖ yaklaşımının uzun yıllar etkin bir yaklaşım olarak kullanılacağı öngörülmektedir.

Kaynaklar

Dergi Makalesi

Ahrens, A., Hansen, C., Schaffer, M., & Wiemann, T. (2024). Model averaging and double machine learning. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4691169

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Cohrs, K., Varando, G., Carvalhais, N., Reichstein, M., & Camps-Valls, G. (2024). Causal hybrid modeling with double machine learning - Applications in carbon flux modeling. *Machine Learning: Science and Technology*. https://doi.org/10.1088/2632-2153/ad5a60

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, 21(1), C1–C68. https://doi.org/10.1111/ectj.12097

Farbmacher, H., Huber, M., Langen, H., & Spindler, M. (2020). Causal mediation analysis with double machine learning. *The Econometrics Journal*. https://doi.org/10.1093/ectj/utac003

Felderer, B., Kueck, J., & Spindler, M. (2023). Using double machine learning to understand nonresponse in the recruitment of a mixed-mode online panel. *Social Science Computer Review, 41*(2), 461-481. https://doi.org/10.1177/08944393221095194

Fingerhut, N., Sesia, M., & Romano, Y. (2022). Coordinated double machine learning. arXiv, 6499-6513. https://doi.org/10.48550/arXiv.2206.00885

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232.

Knaus, M. (2020). Double machine learning based program evaluation under unconfoundedness. *Econometrics: Mathematical Methods & Programming eJournal*. https://doi.org/10.1093/ectj/utac015

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267–288.

Zhang, L., Li, X., Zheng, D., Zhang, K., Q., Zhao, Y., & Ge, Y. (2021). Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach. *Journal of Hydrology, 594*, 125969. https://doi.org/10.1016/J.JHYDROL.2021.125969

Kitap

Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC.