

Eğitim Araştırmalarında Çok Düzeyli Meta-Analiz Modelleri: Örnek Uygulamalı Bir Rehber

Multilevel Meta-Analysis Models in Educational Research: A Practical Guide

Dilek Karaca¹, Burak Aydın² Hakan Atılğan³

¹Sorumlu Yazar, Öğretim Görevlisi, Manisa Celal Bayar Üniversitesi, dilek.karaca@cbu.edu.tr, (<https://orcid.org/0000-0001-8004-4211>)

²Doç. Dr., Ege Üniversitesi, burak.aydin@ege.edu.tr, (<https://orcid.org/0000-0003-4462-1784>)

³Prof. Dr., Ege Üniversitesi, hakan.atilgan@ege.edu.tr, (<https://orcid.org/0000-0002-5562-3446>)

Geliş Tarihi: 02.05.2024

Kabul Tarihi: 27.08.2024

ÖZ

Meta-analiz belirli alanda yapılmış çalışmalardan sistematik şekilde elde edilen nicel verileri kullanarak o alanla ilgili genel durumu ortaya koymaya çalışan istatistiksel bir yöntemdir. Geleneksel meta-analiz yöntemleri, etki büyüklükleri arasında bağımlılık olmadığı varsayımına dayanmasına rağmen özellikle sosyal bilimlerde etki büyüklüğü bağımlılığına neden olabilecek çok sayıda durum söz konusudur. Araştırmacının elde ettiği etki büyüklüklerindeki bağımlılık veri setinde kümeli bir yapı oluşturur. Geleneksel meta-analiz uygulamalarındaki etki büyüklüğü bağımlılığı sorunu ile başa çıkmak ve kümeli veri yapısını dikkate almak için önerilen yöntemlerden biri çok düzeyli meta-analitik modellerin kullanılmasıdır. Çok düzeyli modeller diğer istatistiksel çerçevelerle birleştirilebilir ve kümeli veri yapılarının daha savunulabilir şekilde çözümlenebilmesini sağlayabilir. Bu sebeple çok düzeyli modellerin sosyal bilimlerde kullanım sıklığı her geçen gün artmaktadır. Bu çalışmanın amacı örnek bir veri seti üzerinden çok düzeyli meta-analitik modellerin nasıl uygulanabileceğini göstermektir. R yazılımı kullanılarak gerçekleştirilen analizlerde *metafor* paketinin *rma.mv* fonksiyonu kullanılmıştır. Bu uygulama okuyuculara veri dosyasının düzenlenmesi, R yazılımının hazırlanması, genel etkinin hesaplanması, çalışma içi ve çalışmalar arası varyans heterojenliğinin incelenmesi ile kategorik ve sürekli değişkenlere ait moderatör analizlerinin nasıl yapılacağını adım adım anlatan bir kılavuz niteliği taşımaktadır. Çalışmada kullanılan veri dosyası ve R betiği okuyucuların kullanımı için ekler kısmında sunulmuştur.

Anahtar Kelimeler: Meta-analiz, çok düzeyli meta-analiz, çok düzeyli modeller, metafor, R yazılımı.

ABSTRACT

Meta-analysis is a statistical method that tries to reveal the general situation about that field by using quantitative data obtained systematically from studies conducted in a certain field. However, traditional meta-analysis methods assume effect size independence, which can be a significant and misleading limitation in social sciences research. The dependence in the effect sizes obtained by the researcher creates a clustered data structure, a problem that needs attention. One of the methods to address this problem is the use of multilevel meta-analytic models. Multilevel models can be combined with other statistical frameworks, providing a more defensible analysis of clustered data structures. For this reason, the frequency of multilevel models in social sciences is steadily increasing. This study demonstrates how multilevel meta-analytic models can be applied to a sample data set. In the R environment,

the `rma.mv` function of the `metafor` package was utilized for the analyses. This application is a step-by-step guide for readers on organizing the data file, preparing the R software, calculating the overall effect, examining the heterogeneity of variance within and between studies, and conducting moderator analyses with categorical and continuous variables. The data file and R script used in the study are shared as supplementary material for the readers.

Keywords: Meta-analysis, multilevel meta-analysis, multilevel models, metafor, R software.

GİRİŞ

Meta-analiz belirli bir alanda yapılmış ampirik çalışmalardan elde edilen nicel verileri kullanarak o alanla ilgili genel durumu ortaya koymaya çalışan istatistiksel bir yöntemdir. Meta-analiz terimini ilk defa kullanan Glass (1976) “analizlerin analizi” olarak tanımlamıştır. Meta-analizin amacı açıkça tanımlanmış bir konu ya da araştırma sorusu ile ilgili mevcut tüm kanıtları birleştirmek, özetlemek ve yorumlamaktır (Lipsey & Wilson, 2001). Farklı çalışmalardan elde edilen bulguları birleştirme işlemi birincil çalışmalardan elde edilen korelasyon katsayılarının özetlenmesi, p değerlerinin birleştirilmesi ve aritmetik ortalamaların ortalamasını alma şeklinde gerçekleştirilmekteyken (Fisher, 1933; Pearson, 1904) Glass, bu değerler yerine etki büyüklüğünü (*effect size*) kullanmayı önermiş ve bu öneri kabul görmüştür (Glass, 1976). Meta-analiz yapmayı mümkün kılan etki büyüklüğü değeri “bağımlı değişken” olarak ele alınır ve çalışmalar arası standartlaştırılarak elde edildiği için karşılaştırılabilir bir istatistik ortaya çıkmasını sağlar (Şen & Yıldırım, 2020). Etki büyüklükleri standartlaştırılmış bir ölçü olan Cohen’s d (Cohen, 1988), Hedges’ g (Hedges, 1982), risk oranı (*risk ratio*), olasılık oranı (*odds ratio*) ve standart ortalama farkı (*standardized mean difference*) gibi istatistiklerle ifade edilir ve birbirlerine formüller kullanılarak dönüştürülebilirler.

Meta-analiz gerçekleştirmek için izlenmesi gereken basamaklar hakkında alanyazında net bir fikir birliği bulunmamaktadır. Fakat en genel haliyle bir meta-analiz tamamlayabilmek için izlenmesi gereken adımlar: 1) araştırma konusunun/probleminin belirlenmesi; 2) ilgili birincil çalışmaların araştırılması ve toplanması; 3) birincil çalışmalardan elde edilen verilerin kodlanması; 4) elde edilen verilerin istatistiksel analizlere entegre edilmesi ve 5) analizlerden elde edilen sonuçların yorumlanması ve çıkarımlar yapılması şeklindedir. Meta-analiz kullanmak isteyen araştırmacılar meta-analiz adımları ve raporlama süreçleri ile ilgili daha detaylı bilgi için QUOROM (*The Quality of Reporting of Meta-analysis*; Moher vd., 1999), PRISMA 2020 versiyonu (*The PRISMA 2020 statement: an updated guideline for reporting systematic reviews*, Page vd., 2020) ve MOOSE (*Meta-analyses of Observational Studies*; Stroup vd., 2000) gibi başlıca kullanılan kontrol listelerini inceleyebilir ve referans alabilirler. Ayrıca eğitim ve psikoloji alanındaki çalışmaların raporlanmasında Amerikan Psikoloji Derneği yazım klavuzunda yer alan meta-analizin raporlanması ile ilgili APA-MARS (*American Psychological Association – Meta-Analyses Reporting Standards*; APA, 2008) kurallarını da inceleyebilirler. Bu çalışmada rehberlerin detaylarına daha fazla değinilmeyecek olup özellikle birincil çalışmalardan elde edilen verilerin çok düzeyli modellere entegre edilmesine odaklanılmıştır.

Meta-analiz uygulamalarında araştırmacının araştırma konusu ile ilgili yapılmış tüm çalışma verilerine ulaşmayı denemesi beklenmektedir. Fakat alanyazında genellikle istatistiksel açıdan manidar etkilerin raporlandığı, manidar etki raporlamayan çalışmaların ise alanyazında daha az yer aldığı görülmektedir. Bu durum meta-analizde yayın yanlılığı (*publication bias*) ve çekmece sorunu (*file-drawer problem*) olarak bilinen problemlere neden olmaktadır (Cheung, 2019; Cheung & Chan, 2014; Rosenthal, 1991). Yayın yanlılığı sorununu tespit etmek, çözümler sunmak için çok sayıda yöntem geliştirilmiş olmasına rağmen bu durum tamamen ortadan kaldırılamamıştır. Yayın yanlılığının çalışma sonuçlarını etkilemeyecek düzeyde olduğunu ortaya koyabilmek için huni grafiği (*funnel plot*), duyarlılık analizi (*sensitivity analyses*), kırpma ve doldurma yöntemi (*trim and fill method*, Duval & Tweedie, 2000), Begg ve Mazumdar’ın sıra

korelasyonu testi (*Begg method*, Begg & Mazumdar, 1994), güvenli N yöntemi (*fail-safe N*, Rosenthal, 1979) ve Egger regresyon testi (Egger vd., 1997) gibi prosedürlerden yararlanılmaktadır (Cheung, 2015; Hedges, 1982).

Meta-analizin temel birimi olan etki büyüklüğü, iki değişkenin arasındaki ilişkinin ya da uygulama etkisinin büyüklüğünü ifade eden bir değerdir. Meta-analizi uygulamak için, araştırmaya dahil edilen her bir çalışmanın etki büyüklüğü ile varyansı hesaplanır (Borenstein vd., 2009). Hesaplanan bu etki büyüklüğü değerleri standartlaştırılır ve ortalama etki büyüklüğü elde edilir. Çalışmalar arasındaki heterojenliğin varlığını belirlemek için *Q* testi ve *I*² istatistiklerinden faydalanılır ve heterojenlik durumu açıklamak için moderatör veya meta-regresyon analizleri yapılabilir. Analizlerde araştırma sorusuna göre sabit etki veya rasgele etkiler modellerinden biri kullanılır. Yayın yanlılığı ve duyarlılık analizleri ile meta-analiz sonuçları değerlendirilir ve raporlanır.

Meta-analiz alanyazınında üzerinde durulan önemli konulardan biri de etki büyüklükleri arasında bağımlılık olmaması varsayımdır (Assink & Wibbelink, 2016; Rosenthal, 1991). Geleneksel meta-analiz yöntemleri, etki büyüklükleri arasında bağımlılık olmadığı varsayımına dayanmaktadır. Etki büyüklükleri arasında bağımlılık olması meta-analiz sonuçlarına aşırı güven duyulması veya yanlı sonuçların ortaya konmasına neden olabilir (Van den Noortgate vd., 2013). Özellikle sosyal bilimler alanında yapılan çalışmalarda bağımlılık durumuna yol açabilecek çok sayıda durum söz konusudur. Örneğin aynı araştırmacı grubunun farklı çalışmalarının birbiriyle bağlantılı olması, aynı örneklem üzerinde aynı konu ile ilgili birden fazla çalışma yapılmış olması ya da aynı ülkede/bölgede/okulda yapılan çalışmaların birbiriyle bağlantılı olması gibi farklı senaryolar etki büyüklüğü bağımlılığına neden olabilmektedir (Cheung, 2015; Van den Noortgate vd., 2013). Bu durumlar göz önüne alındığında farklı çalışmalar arası bağımlılık durumu olabileceği gibi aynı çalışmadan elde edilen etki büyüklükleri arasında da bağımlılık olabilir. Lipsey ve Wilson (2001) meta-analizde etki büyüklüğü bağımsızlığı varsayımını karşılamak için birincil çalışma başına yalnızca bir etki büyüklüğünün dahil edilmesine vurgu yapmaktadır. Özellikle aynı çalışmadan elde edilen etki büyüklüklerinin (aynı katılımcılara, aynı ölçme araçlarına vb.) farklı çalışmalardan elde edilen etki büyüklüklerine kıyasla birbirine daha çok benzemesi -dolayısıyla bağımsız olmaması- bu vurguyu desteklemektedir (Becker, 2000; Cheung, 2015; Gooty vd., 2021; Hox vd., 2010; Van den Noortgate vd., 2013).

Alanyazında meta-analizdeki etki büyüklüklerinin bağımlılığı ile başa çıkmak için farklı yöntemler önerilmiştir (Borenstein vd., 2009; Gooty vd., 2021; Hedges & Olkin, 1988; Lipsey & Wilson, 2001; Rosenthal, 1991). Bu yollardan ilki bağımlılık problemini göz ardı etmek ve etki büyüklüklerinin bağımsız olduğu varsayımına dayalı olarak analize devam etmektir. Fakat bu yol bulguların geçerliliği hakkında sorun teşkil edeceği için önerilmez (Becker, 2000; Scammacca vd., 2014). İkinci yol birden fazla etki büyüklüğü olan çalışmalardan tek bir etki büyüklüğünün analize dahil edilmesidir. Bu durumda da araştırmacı etki büyüklüklerinden birini seçebilir ya da etki büyüklüklerinin ortalamalarını alarak analize dahil edebilir (Cheung, 2015; Marín-Martínez & Sánchez-Meca, 1999). Bağımlı etki büyüklüklerinin ortalamalarının alınması ile ilgili çeşitli yaklaşımlar geliştirilmiş olsa da (Cheung & Chan, 2004; Rosenthal & Rubin, 1986) bazı sınırlılıkları bulunmaktadır. Bu yöntemde veri kaybı fazla olduğu için hem testlerin istatistiksel gücü etkilenecek hem de meta-analizde ele alınacak araştırma sorularını sınırlayabilecektir (Cheung, 2015). Etki büyüklüklerinin ortalamasının alınması bu nedenlerden dolayı pek tavsiye edilmemektedir (Becker, 2000; Cheung, 2015; Cheung & Chan, 2014; Hedges & Pigott, 2001; Hox & de Leeuw, 2003; Scammacca vd., 2014). Üçüncü bir seçenek ise bağımlılığı istatistiksel olarak modelleme yöntemidir. Bu yöntem birincil çalışmalardaki etki büyüklüklerinin sayısını azaltmadan hepsinin dahil edilebileceği ve istatistiksel gücün çok yüksek olduğu, çok değişkenli (*multivariate*) ve çok düzeyli (*multilevel*) modellerin kullanılmasına imkân tanımaktadır (Cheung, 2015; Gooty vd., 2021; Harrer vd., 2021; Hox & de Leeuw, 2003; Scammacca vd., 2014).

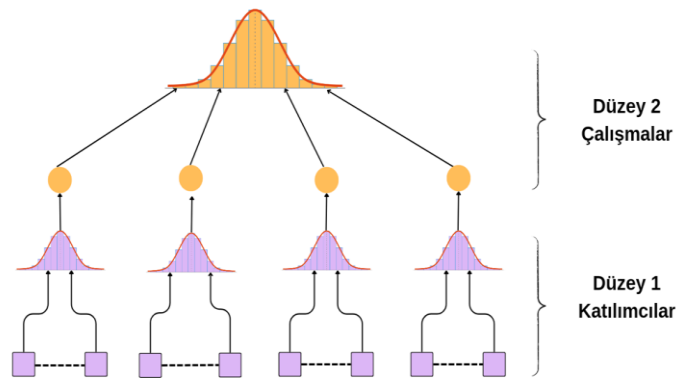
Meta-analizde bağımlı etki büyüklükleri ile başa çıkmak için çok düzeyli modellerin uygulandığı yaklaşımlar güçlü istatistiksel tekniklerin kullanımına imkân vermesi ve diğer yöntemlere göre daha avantajlı olması nedeniyle araştırmacılar tarafından son yıllarda tercih edilmeye başlanmıştır (Cheung, 2014; Gooty vd., 2021; Harrer vd., 2021; Hox & de Leeuw, 2003; Van den Noortgate vd., 2013). Bu yaklaşımlara göre geleneksel meta-analiz modelleri çok düzeyli meta-analiz modellerinin özel bir durumu olarak değerlendirilmektedir. Geleneksel meta-analiz modelindeki veriler için birinci düzeye örneklem varyansı, ikinci düzeye çalışma düzeyi varyansı (*within-study variance*) ve üçüncü düzeye de çalışmalar arası varyansın (*between-study variance*) açıklanmasına olanak tanıyarak çok düzeyli modellerle analiz edilebilir duruma gelmiştir (Assink & Wibbelink, 2016; Hox & de Leeuw, 2003).

Geleneksel meta-analiz için çok genel bir model olan rasgele etkiler (*random effects*) modelinin odak noktası, birleştirilmiş bir sonucun istatistiksel anlamlılığını belirlemek değil, farklı çalışmalarda bulunan etki büyüklüklerinin değişkenliğini analiz etmektir. Türkçe alanyazında *random* kelimesinin karşılığı olarak *rasgele*, *seçkisiz* ve *yansız* terimlerinin kullanıldığı görülebilir. Bununla birlikte Türkçe meta-analiz alanyazımında *rasgele* kelimesinin daha çok benimsendiği söylenebilir. Midway (2022) *rasgele* etkiler terimindeki *rasgele* (*random*) kelimesinin yanıltıcı olabileceğini çünkü *rasgele* etkiler modelinin özünde gerçek değil örtük bir *rasgelelik* barındırdığını; bir diğer ifadeyle, *rasgeleliğin* meta-analizi yapan kişinin alanyazında tespit ettiği etki büyüklüklerinin daha büyük bir etki büyüklüğü kümesinden *rasgele* çekilmiş etkiler olduğunu kabul etmesiyle oluştuğunu belirtmiştir. Eğer çalışmalar arası varyans büyük ve anlamlıysa çalışma sonuçları heterojen kabul edilir. Bu durum çalışmaların hepsinin aynı sonucu vermediğini, sonuçların bir dağılıma sahip olduğunu ve mevcut çalışma örnekleminin bu dağılımın ortalamasını ve varyansını tahmin etmek için kullanıldığı anlamına gelmektedir. Hox vd. (2010) çalışmalar arasındaki varyansın açıklanabilmesi için çalışma sonuçlarını etkileyen değişkenlerin yani moderatör değişkenlerin incelenmesinin gerekli olduğunu ifade etmektedir.

Meta-analizin doğası gereği çok düzeyli bir yapıya sahip olduğu söylenebilir. Geleneksel meta-analizde *rasgele* etkiler modeli incelendiğinde aslında iki hata kaynağı olduğunu ve bunlardan birinin bireysel çalışmalardan diğerinin çalışmalar arası heterojenlikten kaynaklandığı söylenebilir. Buradaki hata kaynakları iki düzeye karşılık gelecek şekilde incelenirse katılımcılar bir düzeye, çalışmalar da bir başka düzeye denk geleceklerdir (Harrer vd., 2021). Düzey 1'e katılımcıların, Düzey 2'ye çalışmaların yerleştirilmesiyle geleneksel *rasgele* etkiler modelinin çok düzeyli yapıda olduğu görselleştirilerek ifade edilebilir. Harrer vd. (2021) çalışmasında yer alan çizimlerin ilham kaynağı olduğu ve benzer bir görsel dil kullanmak amacıyla geleneksel *rasgele* etkiler modelinin çok düzeyli yapısı Şekil 1'de simgelenmiştir.

Şekil 1

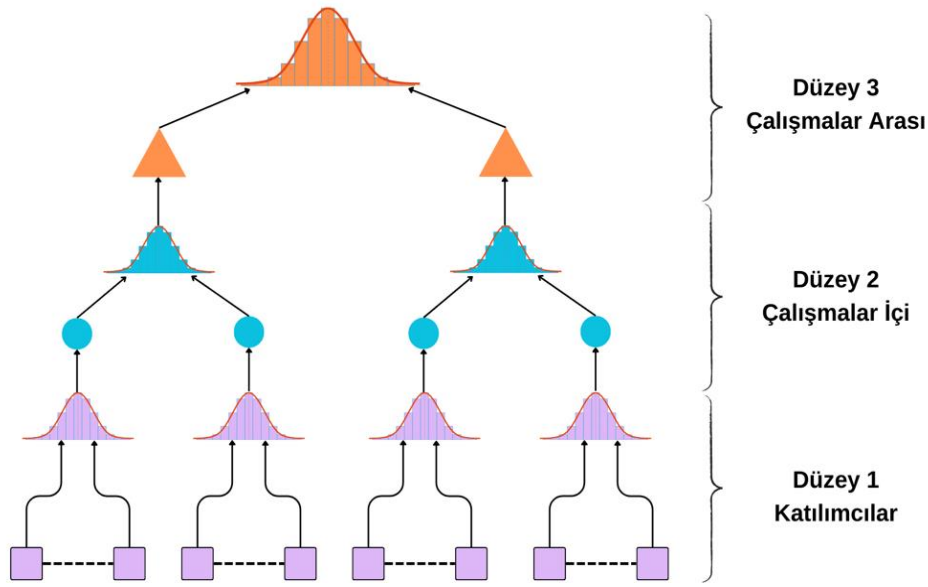
Geleneksel Rasgele Etkiler Modelinin Çok Düzeyli Yapısı



Bir meta-analiz çalışması yürütüldüğünde birinci düzeydeki veriler genellikle “havuzlanmış (*pooled*)” biçimde ulaşılan verilerdir. İkinci düzeyde de meta-analizin gerçekleştirildiği ve birleştirme işleminin yapıldığı çalışmalar bulunmaktadır. Bu açıdan bakıldığında katılımcıların çalışmalar içinde “iç içe” geçtiği kümelenmiş yapılar olduğu görülmektedir (Harrer vd., 2021). Bir adım daha ileriye giderek verilerle ilgili mevcut durumun daha derinlemesine incelenmesine imkân tanıyan ve bu yapıyı genişletmenin mümkün olduğu üç düzeyli meta-analitik modeller önerilmektedir (Cheung, 2015; Hox & de Leeuw, 2003; Van den Noortgate vd., 2013). Modellerin üç düzeyli modellere geliştirilmesiyle birlikte de etki büyüklüğü bağımlılığının dikkate alınacağı belirtilmiştir. Üç düzeyli bir meta-analiz modelinin örnek gösterimi Şekil 2’de sunulmuştur. Şekillerin çiziminde Harrer vd., (2021) kaynağından faydalanılmıştır.

Şekil 2

Üç Düzeyli Meta-Analiz Modelinin Yapısı



Üç düzeyli meta-analiz modeli incelendiğinde üç havuzlama adımı içerdiği görülmektedir. Birincil çalışmalardan elde edilen veriler birleştirilerek bir etki büyüklüğü hesaplanır (Düzyen 1). Bu etki büyüklükleri ikinci düzeyde gösterilen birkaç küme içinde kümelenir (Düzyen 2). Bu kümeler bireysel çalışmalar ya da çalışmaların alt kümeleri olabilir. Ve son olarak birleştirilmiş küme etkilerinin havuzlanması ile genel gerçek etki büyüklüğüne (μ) ulaşılır. Kavramsal olarak bu ortalama etki, sabit veya rasgele etkiler modelindeki havuzlanmış gerçek etkiye oldukça yakındır. Ancak aradaki fark; birincil çalışmalardaki bağımlı etki büyüklüklerini dikkate alarak hesaba katan bir modele dayanmasıdır (Harrer vd., 2021).

1.1 Çok Düzeyli Meta-Analiz

İki düzeyli meta-analiz modelinde birinci düzey katılımcıları, ikinci düzey çalışmaları temsil etmektedir (Hox vd., 2010). İki düzeyli meta-analiz modelinde birinci düzey varyansı doğrudan etki büyüklükleri ile hesaplanırken, ikinci düzey varyansı etki büyüklükleri tahminlerinde çalışmalar arası varyansı yansıtmaktadır. İki düzeyli hiyerarşik doğrusal modellerde ham veri kullanarak birinci ve ikinci düzey varyans ve değişken katsayıları hesaplanırken, iki düzeyli meta-analiz modellerinde birincil çalışmalardan elde edilen özet veriler (ortalama ve standart sapma) kullanılarak hesaplamalar yapılır. Bu nokta standart çok düzeyli

modellerle çok düzeyli meta-analiz modellerinin ayrıştığı noktadır. Çok düzeyli meta-analiz modellerinde ham veri yerine farklı çalışmalardan elde edilen özet istatistikler kullanıldığından ve özet istatistiklerin elde edildiği farklı çalışmalar farklı ölçeklerde olduğu için etki büyüklükleri hesaplanarak aynı ölçek üzerine yerleştirilmeye çalışılır. Birinci düzey çalışma içi ve ikinci düzey çalışmalar arası olarak ele alındığında iki düzeyli meta-analiz modeline ait eşitlikler aşağıdaki gibi sunulabilir (Hox & de Leeuw, 2003; Konstantopoulos, 2011).

Düzye 1	$y_i = \lambda_i + e_i$	
Düzye 2	$\lambda_i = \beta_0 + u_i$	Eşitlik 1
Birleştirilmiş eşitlik	$y_i = \beta_0 + u_i + e_i$	

Eşitlik 1’de $e_i \sim N(0, v_i)$, $u_i \sim N(0, \tau^2)$ olduğu ve örneklem varyansının (v_i) bilindiği varsayımları kabul edilir. Denklemdaki y_i normal dağılım gösteren etki büyüklüklerini, λ_i etki büyüklükleri ortalamasını ve v_i de varyansı ifade etmektedir. Çalışmalar arası modelde (ikinci düzeyde), popülasyon parametresi tahminleri genel bir ortalama (β_0) etrafında dağılım gösterir ve u_i çalışmalara özgü rasgele dağılımı ifade eder. Birinci ve ikinci düzey eşitliklerinin birleştirilmesiyle oluşan birleştirilmiş modele çok düzeyli model alanyazınında boş (*null/empty*) ya da koşulsuz (*unconditional*) model denir. Çalışmanın amacına göre boş modele bağımsız değişkenler (moderatör) eklenerek model genişletilebilir. Araştırmacılar araştırma özelliklerine göre bu modele bir düzey daha ekleyerek üç düzeyli meta-analiz modeli oluşturabilirler.

Üç düzeyli meta-analiz modelleri için Konstantopoulos (2011) ve Van den Noortgate vd. (2013) iki düzeyli rasgele etkiler modelinin genişletilmiş hali olan modeli önermişlerdir. Eklenen üçüncü düzey aynı çalışma içinde yer alan etki büyüklüklerinin o çalışma içerisinde kümelenmesine izin vererek bir düzey oluşturur. Birinci düzeyin katılımcılar, ikinci düzeyin çalışma kümeleri, üçüncü düzeyin de genel etki için kurulduğu söylenebilir. Üç düzeyli meta-analiz modelinde birinci düzey birincil çalışmalardan elde edilen etki büyüklükleri arası değişkenliği, ikinci düzey etki büyüklüklerinin kümeleştikleri çalışmalar arası değişkenliği ve üçüncü düzey de bağımlı etki büyüklüklerinin kümeleştigi farklı çalışmalar arasındaki değişkenliği açıklamak için kullanılır (Şen & Akbaş, 2016).

Üç düzeyli modelde, birinci düzeydeki varyans birincil çalışmalardan toplanan etki büyüklüklerinden hesaplandığı için bilinen varyans olarak adlandırılır. Bu durumda tahmin edilmeye çalışılan varyanslar ikinci ve üçüncü düzeyde olduğu için bu düzeylere ait varyanslar tahmin edilmeye çalışılır. Yani bir başka ifadeyle ikinci düzey varyansı “çalışma içi çeşitlilik (*within-study*)” ve üçüncü düzey varyansı da “çalışmalar arası çeşitlilik (*between-study*)” olarak adlandırılabilir.

Eşitlik 1’de sadece i alt indisi yer alır. Üç düzeyli modeli tanımlamak için i ve j alt indisleri kullanılabilir, i ikinci düzey birimleri, j de üçüncü düzey birimleri temsil eder. Eşitlik 2 ile üç düzeyli meta-analiz modeli tanımlanabilir (Cheung, 2014; Konstantopoulos, 2011; Van den Noortgate vd., 2013).

Birinci düzey	$y_{ij} = \lambda_{ij} + e_{ij}$	
İkinci düzey	$\lambda_{ij} = \beta_{0j} + u_{ij}$	
Üçüncü düzey	$\beta_{0j} = \gamma_{00} + u_{0j}$	Eşitlik 2
Birleştirilmiş eşitlik	$y_{ij} = \gamma_{00} + u_{0j} + u_{ij} + e_{ij}$	

Eşitlik 2’de $e_{ij} \sim N(0, v_i)$, $u_{ij} \sim N(0, \tau^2)$ olduğu ve örneklem varyansının (v_i) bilindiği varsayımları kabul edilir. Tıpkı iki düzeyli modelde olduğu gibi y_{ij} normal dağılım gösteren etki büyüklüklerini, π_{ij} etki büyüklükleri ortalamasını ve v_i varyansı ifade eder. Örneğin y_{12} ikinci çalışma içindeki birinci etki büyüklüğünü, y_{22} ikinci çalışma içindeki ikinci etki büyüklüğünü ifade eder. Bu sayede birincil çalışmalardaki birden fazla etki büyüklükleri aynı anda analize dahil edilebilir. İkinci düzeyde popülasyon parametresi tahminleri β_{0j} simgesiyle ifade edilen genel bir ortalama etrafında dağılım gösterirken, u_{ij} ikinci düzey rasgele dağılımı belirten düzey-2 varyansıdır. Üçüncü düzeyde tahminler genel bir ortalama etrafında (γ_{00}) dağılım gösterirken, u_{0j} sıfır aritmetik ortalamaya ve çalışmalar arası varyansa sahip olan düzey-3 varyansıdır (Şen & Akbaş, 2016). İkinci düzey modellerde olduğu gibi çalışma ile ilgili amaca yönelik ya da araştırmacı tarafından araştırılması istenen değişkenler varsa düzeylerdeki heterojenliği açıklamak için modellere eklenerek araştırılabilir (Cheung, 2014).

Son yıllarda gerek geleneksel meta-analiz çalışmalarının gerekse çok düzeyli meta-analiz çalışmalarının sayısı giderek artmaktadır. Fakat özellikle eğitim ve psikoloji alanında yapılan çalışmalarda hala geleneksel meta-analiz yöntemlerinin sıklıkla tercih edildiği ve bağımsızlık varsayımının ihlal edildiği görülmektedir (Şen & Akbaş, 2016). Etki büyüklüklerinin bağımlılığı konusunda geleneksel meta-analiz çalışmalarından daha iyi bir seçenek olan çok düzeyli meta-analiz çalışmaları uluslararası düzeyde giderek artan bir ilgi görmesine rağmen (Assink & Wibbelink, 2016; Cheung, 2014; Gooty vd., 2021; Hox & de Leeuw, 2003; Konstantopoulos, 2011; Marsh vd., 2009; Pastor & Lazowski, 2018; Van Den Noortgate vd., 2013; Harrer vd., 2021) Türkiye’de yapılan çalışma sayısının ise çok sınırlı olduğu görülmektedir. Ulusal Akademik Ağ ve Bilgi Merkezi (ULAKBİM) Keşif veri tabanında “*multilevel meta-analysis*” anahtar kelimeleri ile yapılan arama sonucuna göre son 10 yılda 195 çalışmaya; “*multilevel meta-analysis*” ve “*education*” kelimeleri ile yapılan taramada 36 çalışmaya ulaşılmışna rağmen “*çok düzeyli meta-analiz*” kelimeleri ile yapılan taramada Şen ve Akbaş (2016)’a ait olan bir çalışmaya ulaşılmıştır. Bu çalışmada da geleneksel meta-analiz ve çok düzeyli meta-analiz uygulamalarından elde edilen sonuçlar karşılaştırılmış olup analizin adım adım nasıl yapılacağına değinilmemiştir. Çalışmada SAS (*Statistical Analysis System*) programı kullanılarak analizler gerçekleştirilmiş olup kullanılan SAS kodları okuyucularla paylaşılmıştır. Elde edilen sonuca göre, Türkiye’de ve Türkçe alanyazında meta-analize çok düzeyli modeller uygulama yaklaşımının henüz yeterince keşfedilmediği ve yapılan çalışmanın çok sınırlı olduğu görülmektedir. Bu durum birkaç temel nedenden kaynaklanıyor olabilir. İlk olarak meta-analize çok düzeyli yaklaşım uygulamalarının oldukça karmaşık olması araştırmacıların bu yöntemi kullanmaktan kaçınmalarına neden olabilir. İkinci olarak araştırmacılar tarafından bu yöntemin pek bilinmemesi ve bu nedenden dolayı tercih edilmemesi olabilir. Meta-analize çok düzeyli modellerin nasıl uygulanacağını anlatan detaylı çalışmaların yeterli sayıda olmaması başka bir neden olabilirken diğer bir neden istatistiksel analizlerin gerçekleştirilmesi için gerekli yazılımların/programların çoğunun ücretli olması olabilir. R’nin ücretsiz ve açık kaynak kodlu bir yazılım olması, bu alanda çalışma yapmak isteyen araştırmacılar tarafından ulaşılabilirliğini kolaylaştırmaktadır. Ayrıca Pastor ve Lazowski (2018) tarafından çok düzeyli meta-analiz uygulamaları gerçekleştirilirken farklı yazılım programları karşılaştırılmıştır. Çalışma sonucuna göre R yazılımının ücretsiz olması yanında, diğer yazılımlara göre tahminleyici seçeneğinin fazla olması, diğer programlara göre daha kapsamlı ve kullanışlı çıktılar sunması gibi avantajları nedeniyle ön plana çıktığı da görülmektedir. Bu bağlamda bu çalışma meta-analize çok düzeyli modellerin uygulanmasına yönelik farkındalık oluşturmak, araştırmacılara yol haritası oluşturmak, ücretsiz ve kolay ulaşılabilir R programı ile analizin nasıl gerçekleştirileceğini göstermek ve alandaki bu eksiklikleri gidermeye yönelik önemli katkılar sunmak amacıyla gerçekleştirilmiştir.

Çalışmamız çok düzeyli meta-analiz modellerinin nasıl yapılacağını adım adım açıklaması, Türkçe yazılması, Türkiye’deki araştırmacıları ve okuyucuları hedef alması, eğitim bilimleri alanında yapılmış olması, özgün bir veri seti kullanılması ve dolayısıyla yapılmış benzer

çalışmaların bulgu ve sonuçlarından farklılık göstermesi, moderatör analiz sonuçlarının istatistiksel olarak daha güvenilir hale gelmesi, katsayıların anlamlılığının değerlendirilmesi ve güven aralıklarının hesaplanabilmesi için `clubSandwich` (Pustejovsky, 2023) paketinin kullanılması açısından farklılık göstermektedir. Bu farklılıkların çalışmamızın bilimsel değerine ve Türkçe alan yazına olumlu katkılar sunması hedeflenmektedir.

1.2. Çalışmanın Amacı

Bu çalışmanın amacı örnek bir veri seti üzerinden iki ve üç düzeyli modellerin R yazılımındaki `metafor` paketin `rma.mv` fonksiyonu kullanılarak meta-analitik modellere nasıl uygulanabileceğini göstermektir. Bu uygulama okuyucuya;

- veri dosyasının düzenlenmesi,
- R yazılımının hazırlanması,
- genel etkinin hesaplanması,
- çalışma içi ve çalışmalar arası varyansın heterojenliğinin incelenmesi,
- kategorik ve sürekli moderatör analizlerinin gerçekleştirilmesi ve
- çoklu moderatör analizi modellerinin incelenmesi adımlarını içeren bir kılavuz niteliği taşımaktadır.

YÖNTEM

2.1. Veri setinin tanıtımı ve hazırlanması

Argümantasyon Tabanlı Bilim Öğrenme (ATBÖ) yaklaşımı uygulamalarının öğrencilerin akademik başarılarına etkisini belirlemek amacıyla ATBÖ yaklaşımının kullanıldığı, yarı deneysel çalışmalarla yürütülmüş, parametrik testler kullanılmış (*t* testi, ANOVA vb.), akademik başarı ölçen Yükseköğretim Kurulu Tez Merkezi (YÖKTEZ) veri tabanında yayınlanan yüksek lisans ve doktora tez çalışmaları incelenmiştir. Bu çalışmada kullanılacak örnek veri seti için araştırmacıların belirlediği kriterler çerçevesinde yeterli veriye sahip olan 15 çalışmadan 53 etki büyüklüğü elde edilmiştir. Veri setine dahil edilen çalışmalar ve etki büyüklükleri ekte sunulmuştur.

Veri setindeki etki büyüklüklerini R ile analiz etmeden önce üç düzeyli meta-analitik modellerin analizi için veri setinin nasıl düzenlendiği önem arz etmektedir. Bu çalışmada kullanılacak veri setinin bir kısmı Şekil 3'te gösterilmektedir. Bu tabloda her bir satır bir etki büyüklüğünü temsil etmektedir. Veri setine dahil edilen her bir bağımsız birincil çalışmayı temsil etmesi için `YayınID` etiketi, çalışmalardan çıkarılan her bir etki büyüklüğü değişkenini tanımlamak için `CalismaID` etiketi kullanılmıştır. Etki büyüklüğü değerleri `yi` etiketli değişken ile tanımlanırken, `vi` etiketi etki büyüklüklerinin örnekleme varyansını belirtmek için kullanılmıştır. Bu dört sütun üç düzeyli meta-analitik modellerin analizini gerçekleştirmek için gerekli zorunlu verileri içermektedir. Çalışmayla ilgili incelenmek istenilen diğer değişkenlerden çalışmaların tez türü için `tez_yl` ve `tez_dr`; örneklem için `ortaokul`, `lise`, `lisans`; çalışmalarda kullanılan soru tiplerinden açık uçlu maddeler için `au`, çoktan seçmeli maddeler için `cs` etiketleri ve çalışmanın uygulanma süresi değişkeni için `hafta` etiketi kullanılmıştır. `phafta` etiketi içinde grup ortalaması etrafında merkezleştirilmiş etkinliklerin uygulanma süresi değişkeni yer almaktadır ve çalışmanın ilerleyen bölümünde hesaplanarak veri setine dahil edilecektir. Veri setinde yer alan tez türü, örneklemi, uygulanan soru tipleri ve çalışmaların uygulanma süresi değişkenleri moderatör analizlerinde kullanılmak üzere veri setine dahil edilmiş değişkenlerdir. Örneklem ile ilgili değişkenlerde referans grup belirlenerek ve diğer iki değişken kukla değişkenlere dönüştürülerek üç grup tanımlanmıştır. Örneğin örnekleme ortaokul öğrencileri seçilmiş ise ortaokul öğrencileri ile ilgili yapılan çalışmalar "1", diğer (lise ve lisans düzeyi) örneklem grupları "0" olarak kodlanmıştır. `tez_yl` ve `tez_dr` değişkenleri yapılan

çalışmaların yüksek lisans ve doktora tezi olmasına, `cs` ve `au` değişkenleri de çalışmalarda kullanılan soru tiplerine göre 1-0 kodlanarak veri setine dahil edilmişlerdir; kullanıcılar hangi kategoriye referans almak istiyorlarsa `tez_yl` veya `tez_dr`, `cs` veya `au` değişkenlerinden birini yordayıcı olarak kullanabilirler. Bu çalışmada `tez_dr` ve `au` değişkenleri kullanılmıştır.

Veri setindeki etki büyüklükleri ve varyansları birincil çalışmaların ortalama ve standart sapmaları kullanılarak araştırmacılar tarafından hesaplanmıştır. Hesaplama R yazılımı kullanılarak Viechtbauer (2010) tarafından geliştirilen `metafor` paketindeki `escalc` (*effect size calculator*) fonksiyonu kullanılmıştır. Fonksiyonda ortalama farklardan hesaplanan Cohen's *d* etki büyüklüğü hesaplaması için "SMD" (Hedges, 1982), örnekleme varyansı hesaplaması için "LS2" tahminleme yöntemi (Borenstein vd., 2009) kullanılmıştır. Kullanılan fonksiyon etki büyüklüğü hesaplamasında örneklem sayısı 20'nin altına düşmesi durumunda Hedges' *g* düzeltme formülünü otomatik olarak kullanmaktadır (Hedges, 1982). Etki büyüklüğü hesaplaması bu makalenin kapsamı dışında olup detaylı okuma için Viechtbauer (2010) ve Morris (2008) kaynakları incelenebilir.

2.2. Verilerin R Programına Aktarılması ve Analizlerin Yapılması

R istatistiksel analizlerin ve grafiklerin hazırlanmasını sağlayan açık kaynak kodlu ücretsiz bir yazılımdır. Zengin paket ve eklenti sistemine sahip olmasıyla kullanıcılara özelleştirilmiş analizler yapma imkânı sunmaktadır. RStudio (2023) gibi bütünleştirilmiş ortamlar ile çalışması nedeniyle de kodlama, görselleştirme ve analiz işlemlerinde büyük kolaylık sağlamaktadır. Yazılım, Windows, MacOS ve Linux gibi farklı işletim sistemlerinde sorunsuz çalışabilmektedir. R yazılımına ait en güncel sürümü kendi web sitesinden indirilerek temin edilebilir (<https://www.r-project.org>).

Verilerin R programına aktarılması, analizde kullanılacak paket programların yüklenmesi ve aktif hale getirilmesi gerekmektedir. Çok düzeyli meta-analiz için Viechtbauer (2010) tarafından geliştirilen `metafor` paket programı ve fonksiyonları kullanılacaktır. Paketin yüklenmesi, aktif hale getirilmesi ve veri setinin programa aktarılması için aşağıdaki kodlar kullanılabilir.

```
#metafor paket yükleme ve aktif hale getirme
install.packages("metafor")
library(metafor)
```

Kurulan paket içeriğini görmek için `?metafor` ya da `help(metafor)` komutları yazılarak ve Run butonuna basılarak detaylı incelenebilir. Verilerin doğru aktarılıp aktarılmadığını kontrol edebilmek için `head` fonksiyonu kullanılarak parantez içine dosya adı ve satır sayısı (örneğin ilk 10 satır için 10) yazılarak Run tuşuna basılır. Aşağıdaki kod çalıştırıldığında görülen ekran Şekil 3'te gösterilmiştir. Ayrıca üzerinde istatistiksel analizlerin yapılacağı veri dosyası RStudio ara yüzünün sağ tarafında bulunan sekmeler kullanılarak da programa aktarılabilir.

```
#veriyi programa aktarma
dataATBO <- read.csv("dataATBO.csv")
head(dataATBO, 10)
```

Şekil 3

Veri Dosyasının Görünümü

YayınID	CalismaID	tez_yl	tez_dr	cs	au	ortaokul	lise	lisans	hafta	yi	vi
1	1	1	0	1	0	1	0	0	6	0.537	0.051
1	2	1	0	0	1	1	0	0	6	0.815	0.053
2	3	0	1	1	1	0	1	0	10	1.567	0.043
2	4	0	1	0	1	0	1	0	10	0.626	0.034
3	5	1	0	1	0	0	0	1	13	0.643	0.066
3	6	1	0	1	0	0	0	1	13	0.826	0.069
4	7	0	1	1	0	0	0	1	8	1.023	0.055
4	8	0	1	0	1	0	0	1	8	1.495	0.063
5	9	1	0	1	0	1	0	0	4	0.540	0.063
5	10	1	0	0	1	1	0	0	4	1.385	0.075

Veri setindeki değişkenlerden kategorik değişkenleri tanımlamak için `lapply` fonksiyonu, incelemek ve kontrol etmek için `str` fonksiyonu kullanılabilir. Burada kullanılan veri seti için 3-9 arası tüm değişkenler kategorik olduğundan fonksiyonda tanımlama yapabilmek için kullanılacak kod aşağıdaki gibi olur. Kodlar çalıştırıldığında görülen ekran Şekil 4'te sunulmuştur.

```
#kategorik değişkenleri tanımla
dataATBO[,3:9] <- lapply(dataATBO[,3:9], as.factor)
#veri yapısını inceleme
str(dataATBO)
```

Şekil 4

Veri Setindeki Değişkenler

```
'data.frame': 53 obs. of 12 variables:
 $ YayınID : int 1 1 2 2 3 3 4 4 5 5 ...
 $ CalismaID: int 1 2 3 4 5 6 7 8 9 10 ...
 $ tez_yl : Factor w/ 2 levels "0","1": 2 2 1 1 2 2 1 1 2 2 ...
 $ tez_dr : Factor w/ 2 levels "0","1": 1 1 2 2 1 1 2 2 1 1 ...
 $ cs : Factor w/ 2 levels "0","1": 2 1 2 1 2 2 2 1 2 1 ...
 $ au : Factor w/ 2 levels "0","1": 1 2 2 2 1 1 1 2 1 2 ...
 $ ortaokul : Factor w/ 2 levels "0","1": 2 2 1 1 1 1 1 1 2 2 ...
 $ lise : Factor w/ 2 levels "0","1": 1 1 2 2 1 1 1 1 1 1 ...
 $ lisans : Factor w/ 2 levels "0","1": 1 1 1 1 2 2 2 2 1 1 ...
 $ hafta : num 6 6 10 10 13 13 8 8 4 4 ...
 $ yi : num 0.538 0.815 1.567 0.626 0.644 ...
 $ vi : num 0.0515 0.0539 0.0431 0.0345 0.0665 ...
```

Çok düzeyli meta-analizin verdiği önemli esnekliklerden biri de analize dahil edilecek birincil çalışmaların her birinden birden fazla etki büyüklüğünün analize dahil edilebilmesine imkân tanınmasıdır. Veri setinde hangi birincil çalışmadan kaç tane etki büyüklüğünün analize dahil edildiğini görmek için aşağıdaki kod kullanılabilir.

```
#frekans tablosu
table(dataATBO$YayınID)
```

Şekil 5

Çalışmalara Göre Etki Büyüklükleri Sayısı

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	2	2	2	2	6	4	8	2	4	2	6	6	3	2

Şekil 5'te 15 birincil çalışmadan elde edilen etki büyüklüklerinin çalışma başına 2 ile 8 arasında değiştiği görülmektedir. aggregate fonksiyonunu kullanarak veri setindeki değişkenlere ait betimsel istatistikler incelenebilir. Etki büyüklüklerinin hem birincil çalışma (makale/tez) bazında hem de genel olarak betimsel istatistiklerini incelemek için aşağıdaki kodlar kullanılabilir ve elde edilen sonuçlar Şekil 6'da sunulmuştur.

```
#birincil çalışma (makale/tez) bazında betimsel istatistikler
round(aggregate(yi ~ YayinID , data=dataATBO, FUN=function(x) c(mean=mean(x), sd=sd(x),
min=min(x), max=max(x))), 3)
#etki büyüklüklerine ait genel betimsel istatistikler
round(c(summary(dataATBO$yi), SD=sd(dataATBO$yi)),2)
```

Şekil 6

Birincil Çalışma Bazında Etki Büyüklükleri İstatistikleri

YayinID	yi.mean	yi.sd	yi.min	yi.max
1	0.676	0.196	0.538	0.815
2	1.097	0.665	0.626	1.567
3	0.735	0.129	0.644	0.826
4	1.260	0.334	1.024	1.496
5	0.963	0.598	0.541	1.386
6	0.871	0.903	-0.111	1.893
7	0.405	0.315	0.140	0.787
8	0.332	0.521	-0.340	1.141
9	0.760	0.121	0.674	0.846
10	1.132	1.124	0.022	2.258

Birincil çalışmaların uygulanma süresi sürekli değişken olarak veri setine dahil edilmiştir. Fakat analiz sonucunun daha kolay yorumlanabilmesi ve değişkene ait anlamlı bir sıfır oluşturmak için genel ortalama etrafında merkezleştirilmiş (*grand mean centering*–GOM) değerleri kullanılarak analizler gerçekleştirilmiştir. Çalışmamız özelinde hafta değişkeni için merkezleştirme yapılması zorunlu olmamakla birlikte yapılacak bir başka çalışmada kullanılacak sürekli bir değişken için merkezleştirmenin nasıl yapıldığı örneklendirilmek amaçlanmaktadır. Örneğin Assink ve Wibbelink (2016)'in yapmış oldukları çalışmada, çalışmanın yapıldığı yıl değişkeni için merkezleştirme yapılmıştır. GOM, değişkenin her bir değerinden değişkene ait genel ortalama değerinin çıkarılması ile elde edilir. Merkezleştirme yapabilmek için kullanılacak kod aşağıda sunulmuştur.

```
#Hafta değişkenini genel ortalama etrafında merkezleştirme
dataATBO$phafta<-dataATBO$hafta-mean(dataATBO$hafta, na.rm=TRUE)

#Hafta ve merkezleştirilmiş hafta değişkenlerini inceleme
haftadegiskenleri <- dataATBO[,c("hafta", "phafta")]
print(haftadegiskenleri)
```

Kodda kullanılan argümanlardan `dataATBO$phafta`, veri setinde `phafta` adında yeni bir değişkenin tanımlanacağını belirtmek için kullanılmaktadır. GOM yapabilmek için hafta değişkenine ait her bir değerden, hafta değişkenine ait genel ortalamanın çıkarılması komutu yazılmıştır. Veri seti içinden sadece hafta ve merkezleştirilmiş hafta değişkenleri incelenebilir. Bunun için kodda veri setinden hangi değişkenlerin seçileceği ve bu seçime bir isim tanımlaması (`haftadegiskenleri`) yapılabilir. Kod çalıştırıldığında oluşan yeni değişken ve değerlerinden bir kısmı Şekil 7’de sunulmuştur.

Şekil 7

Hafta Değişkenlerine Ait Veriler

hafta	phafta
6.0	-3.132
6.0	-3.132
10.0	0.867
10.0	0.867
13.0	3.867
13.0	3.867
8.0	-1.132
8.0	-1.132
4.0	-5.132
4.0	-5.132

Yapılan bu düzenlemelerden sonra veri seti çok düzeyli meta-analizi gerçekleştirmek için hazır hale getirilmiş olur.

2.3. İki Düzeyli Boş Model

ATBÖ yaklaşımı uygulamalarının öğrencilerin akademik başarılarına etkisini araştırmak için öncelikle yalnızca genel etkiyi temsil eden iki düzeyli boş/koşulsuz model (*null/empty/unconditional*) işe koşulmuştur. Boş model en basit modeldir ve herhangi bir yordayıcı içermediğinden yordayıcısız model olarak da bilinir. Bu iki düzeyli rasgele etkiler modeli kümeli yapının ihlal edildiği modeldir ve dolayısıyla da geleneksel meta-analiz ile sonuçların birbirine eşdeğer çıkması beklenmektedir (Şen & Akbaş, 2016). Fakat her ne kadar birbirine yakın sonuçlar verse de çok düzeyli modellerin avantajlarının daha fazla olmasından dolayı geleneksel meta-analiz yerine kullanılmaları tavsiye edilmektedir (Hox & de Leeuw, 2003).

İki düzeyli boş model için `metafor` paketinin `rma.mv` fonksiyonu kullanılacaktır. Kod yazılıp Run tuşuna basılarak çalıştırıldığında oluşan sonuç Şekil 8’de gösterilmiştir.

```
#İki-düzeyli boş model (diğer ifadeyle bilinen rasgele etkiler modeli, kümeli yapının ihlal edildiği durum)
bosmodel2d <- rma(yi, vi, data=dataATBO)
bosmodel2d
confint(bosmodel2d, digits=2)
```

Fonksiyondaki argümanlar; `bosmodel2d`: `rma.mv` fonksiyonun sonuçlarının saklanacağı nesneye verilen isim; `yi`: çalışma verisindeki etki büyüklüklerini içeren değişkenin; `vi`: çalışma verisindeki varyansları içeren değişkenin adıdır. `Confint` fonksiyonu ise belirtilen modelin parametreleri için güven aralıklarının hesaplanmasını sağlarken, `digits=2` komutu ile ondalık basamak sayısının iki olması belirtilmiştir.

Şekil 8

İki Düzeyli Boş Model Sonucu

```
Random-Effects Model (k = 53; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.4152 (SE = 0.0995)
tau (square root of estimated tau^2 value)      : 0.6443
I^2 (total heterogeneity / total variability)   : 83.67%
H^2 (total variability / sampling variability): 6.12

Test for Heterogeneity:
Q(df = 52) = 267.5242, p-val < .0001

Model Results:
estimate      se      zval      pval      ci.lb      ci.ub
0.7303  0.0980  7.4511  <.0001  0.5382  0.9224  ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Şekil 8’de sunulan sonuçların detayları incelendiğinde veri kümesinin 53 etki büyüklüğü ($k=53$) içerdiğini ve modeldeki parametreleri tahmin etmek için kısıtlı en çok olabilirlik (*REML-Restricted Maximum Likelihood*) tahmin yönteminin kullanıldığı görülmektedir. *rma.mv* fonksiyonunda tahminleme yöntemi formülde özellikle belirtilmediğinde varsayılan olarak REML kullanılmaktadır. Bununla birlikte farklı bir tahminleme yöntemi kullanılmak istenirse kodda belirtilebilir. Örneğin en çok olabilirlik kullanılmak istenirse formüle “method=ML” eklemesi yapılabilir.

Heterojenlik testi sonuçları p değerinin .001’den küçük olduğunu, veri setindeki tüm etki büyüklükleri arasında önemli farklılıklar olduğunu ifade etmektedir ($Q(df(52))=267.52$). Tahmini toplam heterojenlik miktarı için τ^2 değerleri, toplam heterojenlik ile toplam değişkenlik oranı için I^2 değerleri, toplam değişkenlik ile örnekleme değişkenliği arasındaki oran için H^2 değerleri incelenebilir. Tahmini toplam heterojenlik .099 standart hata ile .415 olarak elde edilmiştir. Çözümleme sonuçları arasındaki heterojenliğin ($I^2=\%83.67$, $H^2=6.12$) oldukça yüksek olduğunu göstermektedir. Heterojenliğin tanımlanması, ölçülmesi ve yorumlanması hakkında daha detaylı bilgi için Borenstein vd. (2009) ile Higgins ve Thompson (2002) kaynakları incelenebilir.

Genel etki sonucu için model sonuçları bölümündeki değerler incelenebilir. *Estimate* genel etki büyüklüğü değerini, *se* standart hatayı, *zval* z değerini, *pval* p değerini, *ci.lb* ve *ci.ub* güven aralığının alt ve üst sınır değerlerini ifade etmektedir. Çözümleme sonucuna göre ATBÖ yaklaşımı uygulamaları ile öğrencilerin akademik başarıları arasındaki genel ilişkinin .098 standart hata ile .730 (Cohen’s d değerlerine göre büyük etki) olduğu ve bu etkinin istatistiksel olarak anlamlı olduğu ifade edilebilir. Güven aralığı .538 ile .922 arasındadır. Araştırmacılar *predict(bosmodel2d)* kodunu kullanarak tahmin aralıklarını da elde edebilirler.

2.4. Üç Düzeyli Boş Model

ATBÖ yaklaşımı uygulamalarının öğrencilerin akademik başarılarına etkisini araştırmak için genel etkiyi temsil eden üç düzeyli boş model işe koşulmuştur. Üç düzeyli boş modelin iki düzeyli boş modelden farkı kümeli yapının dikkate alınmasıdır. İki düzeyli modelin genişletilmiş hali olan üç düzeyli modelde, eklenen üçüncü düzey aynı çalışma içinde yer alan etki büyüklüklerinin o çalışma içinde kümelenmesine olanak tanır. Birinci düzey birincil çalışmalardan elde edilen etki büyüklükleri arasındaki heterojenliği, ikinci düzey çalışmalar arasındaki heterojenliği ve üçüncü düzey de bağımlı etki büyüklüklerinin kümelendiği farklı çalışmalar arasındaki heterojenliği açıklar. Örneğin öğrenci başarısının incelendiği bir meta-analiz çalışmasında ilk düzey her çalışma içerisindeki öğrenciler arasındaki farklılıklara odaklanırken ikinci düzey çalışmanın yürütüldüğü okullar arasındaki farklılıklara odaklanır. Üçüncü düzey ise okulların bulunduğu şehir ya da ülkelere odaklanabilir.

Metafor paketinin `rma.mv` fonksiyonundaki argümanlar; `bosmodel3d`: `rma.mv` fonksiyonun sonuçlarının saklanacağı nesneye verilen isim; `random`: rasgele etkiler modeli; `list(~1|CalismaID,~1|YayinID)`: meta-analitik modelin üç düzeyli yapısını tanımlamak için gerekli argümandır. Bu argümandaki `CalismaID` değişkeni birincil çalışmalar içindeki etki büyüklükleri arasındaki varyansın dağıtıldığı üç düzeyli modelin ikinci düzeyini tanımlamak, `YayinID` değişkeni de çalışmalar arasındaki varyansın dağıtıldığı üç düzeyli modelin üçüncü düzeyini tanımlamak için kullanılmıştır. `tdist=T`: test istatistiklerinin ve güven aralıklarının t dağılımına dayalı olması gerektiğini belirtmektedir. Aşağıdaki kod yazılıp Run tuşuna basılarak çalıştırıldığında oluşan sonuç Şekil 9’da gösterilmiştir.

```
#üç düzeyli boş model (kümeli yapı dikkate alındığında)
bosmodel3d<- rma.mv(yi, vi, random = list(~ 1 | CalismaID, ~ 1 | YayinID), data=dataATBO, tdist=T)
summary(bosmodel3d, digits=3)
```

Şekil 9

Üç Düzeyli Boş Model Sonucu

```
Multivariate Meta-Analysis Model (k = 53; method: REML)
  logLik  Deviance      AIC      BIC      AICc
  -58.474  116.948   122.948   128.802   123.448

Variance Components:
      estim  sqrt  nlvls  fixed  factor
sigma^2.1  0.389  0.624    53    no  CalismaID
sigma^2.2  0.026  0.162    15    no  YayinID

Test for Heterogeneity:
Q(df = 52) = 267.524, p-val < .001

Model Results:
estimate  se  tval  df  pval  ci.lb  ci.ub  ***
  0.747  0.106  7.040  52  <.001  0.534  0.960  ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Şekil 9’da sunulan sonuçların detayları incelendiğinde 53 etki büyüklüğünün incelendiği ve tahminleme yönteminin REML kullanıldığı görülmektedir. Loglik, Deviance, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) ve Akaike’s Information Corrected Criterion (AICc) meta-analitik model için uyum iyiliği indeksleridir ve modelin veriye ne kadar iyi uyduğu hakkında bilgi edinilmesini sağlar. Çok düzeyli meta-analiz modellerinde ilgili indeksler için daha detaylı bilgi için Cheung (2015) kaynağı incelenebilir. Varyans bileşenleri çalışmalar içinde (2. düzey) ve çalışmalar arasında (3. düzey) dağıtılan varyans değerlerini göstermektedir. .389 çalışmalar içindeki etki büyüklükleri arasındaki varyans için tahmini değeri, .026 da çalışmalar arasındaki etki büyüklükleri arasındaki varyans için tahmini değerleridir. `nlvls` ve `factor` sütunlarındaki değerler de veri setindeki 15 çalışmadan 53 etki büyüklüğü elde edildiğini ifade etmektedir.

Heterojenlik testi sonuçları p değerinin .001’den küçük olduğunu, veri setindeki tüm etki büyüklükleri arasında önemli farklılıklar olduğunu ifade etmektedir. Fakat veri setindeki tüm etki büyüklükleri arasındaki varyansla değil, çalışma içi (2. düzey) ve çalışmalar arası (3. düzey) varyansla ilgilendiğimiz için bu sonuç şu aşamada çok bilgilendirici değildir. Örnekteki analiz sonucuna göre ATBÖ yaklaşımı uygulamaları ile öğrencilerin akademik başarıları arasındaki genel ilişkinin .106 standart hata ile .747 (Cohen’s d değerlerine göre büyük etki) olduğu ifade edilebilir. Bu genel etki anlamlıdır ($t(52)=7.040$, $p<.001$) ve güven aralığı .534 ile .960 arasındadır.

2.5. Etki Büyüklüklerindeki Heterojenliğin Öneminin Belirlenmesi

Çalışma içi varyansın (2. düzey) ve çalışmalar arası (3. düzey) varyansın anlamlı olup olmadığını belirlemek için iki ayrı *log-likelihood-ratio* testi yapılabilir. Bu testlerin tek yönlü (*one-tail*) yapılması tavsiye edilir çünkü varyans bileşenleri sıfırdan sadece pozitif yönde sapma gösterecektir (Assink & Wibbelink, 2016; Viechtbauer, 2010). R çıktısında p değerleri varsayılan olarak iki yönlü (*two-tail*) olarak hesaplandığı için p değerlerinin de ikiye bölünmesi gerekecektir.

2.5.1. Çalışmalar İçi Varyans Heterojenliği

Üç düzeyli boş modelin uygulandığı analiz sonucunda ikinci düzey için dağıtılan varyansın .389 olduğu görülmüştü. Bu varyans bileşenine ait anlamlılığı test etmek için tek yönlü *log-likelihood-ratio* testi uygulanabilir. Bu sayede meta-analitik modelde çalışma içi varyansın modele dahil edilip edilmemesi boş ve alternatif hipotezler ($H_0: \sigma^2(\text{level } 2)=0$, $H_1: \sigma^2(\text{level } 2)>0$) kurularak belirlenebilir. Test sonuçları boş hipotezi reddetmeye destek sağlarsa üç düzeyli boş modelin uyumunun yeni kurulan modelin uyumundan istatistiksel olarak daha iyi olduğunu ifade eder. Yani çalışmalar içindeki etki büyüklükleri arasında önemli farklılıklar olduğu yorumu yapılabilir (Assink & Wibbelink, 2016). Bunun için ikinci düzey varyansının sabitlendiği, üçüncü düzeyin varyansının serbestçe tahmin edildiği yeni bir model kurulur (Assink & Wibbelink, 2016). Bunun için kurulacak model ve anlamlılık testi için aşağıdaki kod kullanılabilir.

```
#Çalışmalar İçi Varyans Heterojenliği(düzye-2)
model3d_var2 <- rma.mv(yi, vi, random = list(~ 1 | CalismaID, ~ 1 | YayinID),sigma2=c(0,NA),
tdist=TRUE, data=dataATBO)
summary(model3d_var2, digits=3)
#anova testi
anova(bosmodel3d,model3d_var2)
```

Koddaki argümanlardan `model3d_var2`, üç düzeyli modelde ikinci düzey varyans heterojenliği belirlemek için kurulan fonksiyonun saklanacağı nesnenin adıdır. `sigma2=c(0,NA)` argümanı ile ilk parametrenin yani çalışma içi varyansın sıfıra sabitlendiğini, ikinci parametrenin yani çalışmalar arası varyansın serbestçe tahmin edileceği belirtilmektedir. `anova` fonksiyonu da oluşturulan iki model arasındaki model uyumunun karşılaştırılması için kullanılmaktadır. Kod çalıştırıldığında elde edilen sonuçlar Şekil 10'da sunulmaktadır.

Şekil 10

Çalışmalar İçi Varyans Heterojenliği ve ANOVA Testi Sonuçları

```
Multivariate Meta-Analysis Model (k = 53; method: REML)
logLik  Deviance    AIC      BIC     AICc
-102.424  204.847   208.847  212.750  209.092

Variance Components:
      estim  sqrt  nlvls  fixed  factor
sigma^2.1  0.000  0.000   53    yes  CalismaID
sigma^2.2  0.068  0.260   15    no   YayinID

Test for Heterogeneity:
Q(df = 52) = 267.524, p-val < .001

Model Results:
estimate  se  tval  df  pval  ci.lb  ci.ub
0.750  0.079  9.462  52  <.001  0.591  0.909 ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> #anova testi
      df    AIC    BIC    AICc  logLik  LRT  pval    QE
Full   3 122.9478 128.8015 123.4478 -58.4739
Reduced 2 208.8474 212.7499 209.0923 -102.4237 87.8996 <.0001 267.5242
```

Şekil 10'da yeni kurulan modelin ve ANOVA testinin sonucu görülmektedir. Full yazan satır boş üç düzeyli boş modeli temsil ederken, Reduced satırı model3d_var2 nesnesinde saklanan yeni modeli, *df* serbestlik derecesini, *LRT* olabilirlik oranı testini (likelihood-ratio), *pval* test istatistiğinin iki taraflı *p* değerini ifade etmektedir. Örnek sonuçları incelendiğinde boş modelin uyumu, kurulan yeni model uyumundan önemli ölçüde daha iyi olduğundan, çalışma içi varyansın önemli olduğu yorumu yapılabilir. Diğer bir ifadeyle çalışma içindeki etki büyüklükleri arasında önemli farklılıklar olduğu söylenebilir. İki yönlü *p* değerinin çok küçük olduğu ($p < .001$) ve ikiye bölünmesinin de bu sonucu değiştirmeyeceği görülmektedir.

2.5.2. Çalışmalar Arası Varyans Heterojenliği

Çalışmalar arası varyansın anlamlılığını belirlemek için yine benzer bir yol izlenecektir. Boş modelin uygulandığı modelde üçüncü düzey için dağıtılan varyans .026 olarak elde edilmişti. Yeni kurulacak modelde bu defa ikinci düzey varyans tahmini serbest bırakılacak ve üçüncü düzey varyans tahmini sifira eşitlenerek sabitlenecektir. Böylece meta-analitik modelde üçüncü düzeydeki varyansın sifira eşit olduğu ve olmadığı boş ve alternatif hipotezler ($H_0: \sigma^2(\text{level3})=0$, $H_1: \sigma^2(\text{level3}) > 0$) test edilecek ve çalışmalar arası varyansın istatistiksel olarak anlamlı olup olmadığı incelenecektir. Yapılan test sonucu boş hipotezi reddetmeye destek sağlarsa yani boş modelin uyumu yeni kurulan model uyumundan istatistiksel olarak daha iyi çıkarsa çalışmalar arasında önemli farklılıklar olduğu yorumu yapılabilir (Assink & Wibbelink, 2016). Bunun için kurulacak model ve anlamlılık testi için aşağıdaki kod kullanılabilir.

```
#Çalışmalar Arası Varyans Heterojenliği (düzey-3)
model3d_var3 <- rma.mv(yi, vi, random = list(~ 1 | CalismaID, ~ 1 | YayinID),sigma2=c(NA,0),
tdist=TRUE, data=dataATBO)
summary(model3d_var3, digits=3)
#anova testi
anova(bosmodel3d,model3d_var3)
```

Koddaki argümanlardan model3d_var3, üç düzeyli modelde üçüncü düzey varyans heterojenliği için kurulan model fonksiyonun saklanacağı nesnenin adıdır. $\text{sigma2} = c(NA, 0)$ argümanı ile ilk parametrenin yani çalışma içi varyansın serbest bırakıldığı, ikinci parametrenin yani çalışmalar arası varyansın sifira sabitlendiği belirtilmektedir. *anova* fonksiyonu da oluşturulan iki model arasındaki model uyumları karşılaştırılması için kullanılmaktadır. Kod çalıştırıldığında elde edilen sonuçlar Şekil 11'de sunulmaktadır.

Şekil 11

Çalışmalar Arası Varyans Heterojenliği ve ANOVA Testi Sonuçları

Multivariate Meta-Analysis Model (k = 53; method: REML)										
	logLik	Deviance	AIC	BIC	AICc					
	-58.656	117.312	121.312	125.214	121.556					
Variance Components:										
	estim	sqrt	nlvls	fixed	factor					
sigma^2.1	0.415	0.644	53	no	CalismaID					
sigma^2.2	0.000	0.000	15	yes	YayinID					
Test for Heterogeneity:										
Q(df = 52)	= 267.524, p-val < .001									
Model Results:										
	estimate	se	tval	df	pval	ci.lb	ci.ub			
	0.730	0.098	7.451	52	<.001	0.534	0.927	***		
Signif. codes:	0	****	0.001	***	0.01	**	0.05	.'	0.1	' ' 1
> #anova testi										
	df	AIC	BIC	AICc	logLik	LRT	pval	QE		
Full	3	122.9478	128.8015	123.4478	-58.4739			267.5242		
Reduced	2	121.3116	125.2141	121.5565	-58.6558	0.3638	0.5464	267.5242		

Analiz sonucuna göre yeni kurulan modelin uyumu boş modelin uyumundan daha iyi olduğu için çalışmalar arası varyansın önemli olmadığı yorumu yapılabilir. Bir başka ifadeyle çalışmalar arası önemli bir değişkenlik söz konusu değildir. İki yönlü p değeri .546 olarak raporlanmıştır.

2.5.3. Varyansların Düzeylere Dağılım Oranı

Çalışmalar içi ve çalışmalar arası varyansın istatistiksel anlamlılığını test edebilmenin yanında toplam varyansın meta-analitik modellerdeki düzeylere dağılım oranını da görmek mümkündür. Bunun için Cheung (2014) formülleri kullanılabilir. Bu formüllerin R sözdizimine çevrilmesi Assink ve Wibbelink (2016) tarafından şu şekilde gerçekleştirilmiştir.

```
#varyansların düzeylerde dağılım oranı
n <- length(dataATBO$vi)
list.inverse.variances <- 1 / (dataATBO$vi)
sum.inverse.variances <- sum(list.inverse.variances)
squared.sum.inverse.variances <- (sum.inverse.variances) ^ 2
list.inverse.variances.square <- 1 / (dataATBO$vi^2)
sum.inverse.variances.square <- sum(list.inverse.variances.square)
numerator <- (n - 1) * sum.inverse.variances
denominator <- squared.sum.inverse.variances - sum.inverse.variances.square
tahmini.ornekleme.varyansi <- numerator / denominator
I2_1 <- (tahmini.ornekleme.varyansi) / (bosmodel3d$sigma2[1]+ bosmodel3d$sigma2[2] + tahmini.ornekleme.varyansi)
I2_2 <- (bosmodel3d$sigma2[1]) / (bosmodel3d$sigma2[1] + bosmodel3d$sigma2[2] + tahmini.ornekleme.varyansi)
I2_3 <- (bosmodel3d$sigma2[2]) / (bosmodel3d$sigma2[1] + bosmodel3d$sigma2[2] + tahmini.ornekleme.varyansi)
amountvariancelevel1 <- I2_1 * 100
amountvariancelevel2 <- I2_2 * 100
amountvariancelevel3 <- I2_3 * 100
amountvariancelevel1
amountvariancelevel2
amountvariancelevel3
```

Kod detayları incelendiğinde veri setindeki etki büyüklükleri sayısı belirlenerek n nesnesine saklanır. Çalışmaların varyanslarının tersi ($1/vi$) alınarak `list.inverse.variances` nesnesine, bu değerlerin hepsi toplanarak `sum.inverse.variances` nesnesine tanımlanır. Toplanan bu değerlerin karesi alınarak `squared.sum.inverce.variances` nesnesine ve varyansların karelerinin tersleri alınarak ($1/vi^2$) `list.inverse.vaiances.square` adlı listeye saklanırlar. Bu listede yer alan ters varyans kareleri toplanır ve `sum.inverse.variances.square` nesnesine atanırlar. Etki büyüklükleri sayısının bir eksiği ($n-1$) ile çalışmaların varyanslarının terslerinin toplamı çarpılarak `numerator` değişkeni tanımlanır. Varyansların tersinin toplamının karesinden (`squared.sum.inverce.variances`), ters varyans kareleri toplamı (`sum.inverse.variances.square`) değerleri çıkarılır ve `denominator` değişkeni olarak atanır. Tahmini örnekleme varyansı (`tahmini.ornekleme.varyansi`) bu değişkenlerinin birbirine bölünmesiyle elde edilir. Birinci, ikinci ve üçüncü düzeylerdeki varyansların düzeyini belirlemek için tahmini örnekleme varyansı değeri ile üç düzeyli boş modelin birinci ve ikinci düzeydeki `sigma2` değerleri kullanılır. Formülde pay ve paydaya yerleştirilerek elde edilen değerlerin 100 ile çarpılmasıyla düzeylerdeki varyans yüzdelik oranları elde edilir.

Kod çalıştırıldığında elde edilen sonuç Şekil 12’de sunulmuştur. Toplam varyansın %16.33’ü düzey 1’deki varyansa (örnekleme/katılımcı varyansı), %78.38’i düzey 2’deki varyansa (çalışma içi varyans) ve %5.28’i düzey 3’teki varyansa (çalışmalar arası varyans) dağılmaktadır.

Şekil 12

Varyansların Düzeylerde Dağılıma Oranı Sonuçları

```
> amountvariancelevel1  
[1] 16.333  
> amountvariancelevel2  
[1] 78.383  
> amountvariancelevel3  
[1] 5.283
```

Yapılan analizler sonucunda ikinci düzeydeki varyansların toplam varyansa oranı istatistiksel olarak anlamlı olmasına rağmen üçüncü düzeydeki varyansların toplam varyansa oranı sıfırdan anlamlı derecede farklı bulunmamıştır. Yani bir başka ifadeyle, toplam varyansa oranla, çalışmalar içindeki etki büyüklükleri arasında önemli farklılıklar söz konusu olmasına rağmen çalışmalar arasındaki etki büyüklükleri arasında önemli farklılıklar bulunmamıştır. Bu durumun nedenleri bu çalışma özelinde düşünüldüğünde analize dahil edilen çalışma sayısının az olmasından kaynaklanıyor olabilir. Çalışma amacımıza istinaden veri setine dahil edilen birincil çalışma sayısının sınırlı tutulması araştırmacılar tarafından alınmış bir karardır. Ama söz konusu durum bir başka çalışmada karşılaşıldığı takdirde ek analizlerin yapılması önerilebilir.

Analize dahil edilen birincil çalışmaların etki büyüklüklerinin betimsel istatistikleri incelendiğinde, özellikle 6., 8., 10., 12., 13. ve 14. çalışmaların en küçük ve en büyük etki büyüklüğü değerleri arasında önemli farklılıklar olduğu görülmektedir. Bu duruma çalışmaların örneklem heterojenliği ve büyüklükleri, uygulanan metodolojik yöntem farklılıkları ve kullanılan ölçme araçları özellikleri ile türleri gibi nedenler yol açmış olabilir. Ayrıca diğer taraftan birincil çalışma sayıları ya da etki büyüklüğü sayıları az olduğunda istatistiksel güçten kaynaklı aslında anlamlı olan sonucun anlamlı çıkmaması durumu söz konusu olabilir. Bu durumda Hunter ve Schmidt (2004) tarafından önerilen %75 kuralı dikkate alınabilir. Öneriye göre toplam varyansın %75’inden daha azı örneklem varyansına (düzey 1) atfedilmesi durumunda heterojenliğin önemli olduğu ve moderatör analizlerinin incelenmesi gerektiği görüşü hakimdir. Çalışmalar içi ve çalışmalar arasında beklenenden daha fazla değişkenlik söz konusu olduğu durumlarda varyansları açıklayabilmek için moderatör analizleri yapılabilir ve varyansa neden olan durumlar derinlemesine incelenebilir.

2.6. Moderatör Analizleri

İkinci ve üçüncü düzeydeki varyansların kaynağını açıklayabilmek için moderatör analizleri yapılabilir. Bu çalışmada tek değişkenli ve çok değişkenli moderatör analizleri için örnekler sunulacaktır. Örnek uygulamalar için veri setinde yer alan iki ve üç kategorili değişkenler ile sürekli değişkenler kullanılacaktır.

2.6.1. İki Kategorili (*binary/dichotomous*) Veriler ile Moderatör Analizi

Meta-analitik modelin parametrelerini tahmin etmek için REML tahmin yöntemi kullanıldığı için, potansiyel moderatör değişkenler içeren bir modelin uyumunu, potansiyel moderatör bir değişken içermeyen bir modelin uyumu ile karşılaştırmak (ör. *log-likelihood-ratio* testi yapmak) mümkün değildir (Hox & de Leeuw, 2003). Bunun yerine modele dahil edilen bir veya daha fazla değişkenin moderatör etkisinin anlamlı olup olmadığını belirlemek için omnibus testi yapılabilir (Assink & Wibbelink, 2016). Omnibus testinde sıfır hipotezi tüm regresyon

katsayılarının sıfıra eşit olduğunu ($H_0=\beta_1=\beta_2=\beta_3=\dots=0$), alternatif hipotez ise bu regresyon katsayılarından en az birinin sıfırdan farklı olduğunu ifade etmektedir.

İki kategorili tek değişkenli moderatör analizi için incelenecek değişken birincil çalışmaların tez türü değişkenidir. Yani birincil çalışmaların tez türü değişkeninin önemli bir moderatör değişken olup olmadığı sorusuna cevap aranacaktır. Veri setinde kukla değişkenlere dönüştürülmüş `tez_yl` ve `tez_dr` değişkenleri kullanılarak oluşturulacak kod aşağıdaki gibidir.

```
#iki kategorili moderatör analizi (tezlerin yl ya da dr tezi olması)
moderator_yl <- rma.mv(yi, vi, mods = ~ tez_dr, random = list(~ 1 | CalismaID, ~ 1 | YayinID),
tdist=TRUE, data=dataATBO)
summary(moderator_yl, digits=3)

library(clubSandwich)
?coef_test
coefs = coef_test(moderator_yl, cluster = dataATBO$YayinID, vcov = "CR2")
coefs
```

Kurulan model `moderator_yl` olarak isimlendirilmiştir. Çünkü referans kategorisi olarak yüksek lisans tezi değişkeni belirlenmiştir. Referans kategorisi; basit regresyon analizine benzer şekilde bir kategori referans olarak belirlenir ve diğer kategoriler ile karşılaştırılması yapılır. `mods` argümanı `rma.mv` fonksiyonunda moderatör analizinde değişkenini tanımlamak için kullanılan argümandır. Moderatör analiz sonuçlarını istatistiksel olarak daha güvenilir hale getirmek, katsayıların anlamlılığını değerlendirmek ve güven aralıklarını hesaplamak için `clubSandwich` (Pustejovsky, 2023) paketi kullanılmıştır. Bu işlem analiz sonuçlarının daha geçerli olmasını sağlayabilir. Öncelikle `clubSandwich` paketi yüklenmeli ve aktif hale getirilmelidir. `Coefs`: istatistiksel analiz sonuçlarının saklanacağı nesnenin adı, `vcov`: varyans tahminleyicisinin belirtilmesi için kullanılan argümandır. Bu çalışmada paketin detaylarında önerilen yanlılığı azaltılmış doğrusallaştırma ayarlaması olan `CR2` kullanılmıştır. Kod çalıştırıldığında gelen sonuç ekranı Şekil 13'te verilmiştir.

Şekil 13

İki Kategorili Moderatör Analizi Sonucu

```
Multivariate Meta-Analysis Model (k = 53; method: REML)
  logLik Deviance   AIC     BIC   AICc
-56.688  113.377  121.377  129.104  122.246

Variance Components:
      estim  sqrt  nlvls  fixed  factor
sigma^2.1  0.404  0.635   53    no  CalismaID
sigma^2.2  0.000  0.000   15    no  YayinID

Test for Residual Heterogeneity:
QE(df = 51) = 261.614, p-val < .001

Test of Moderators (coefficient 2):
F(df1 = 1, df2 = 51) = 2.792, p-val = 0.101
Model Results:
      estimate      se    tval  df    pval  ci.lb  ci.ub  ***
intrcpt      0.936  0.157   5.964  51  <.001  0.621  1.251  ***
tez_dr1     -0.333  0.200  -1.671  51  0.101 -0.734  0.067

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> coefs
  Coef. Estimate      SE t-stat d.f. (Satt) p-val (Satt) Sig.
intrcpt      0.936  0.106   8.84  51.3  <0.001  ***
tez_dr1     -0.333  0.173  -1.92  51.7  0.0839
```

Tek deęişkenli iki kategorili moderatör analizi artık heterojenlik testi (*Test for Residual Heterogeneity*) sonucuna göre çalışmaların doktora tezi olmasının çalışmalar içindeki etki büyüklükleri arasındaki varyansı açıklama durumuna bakılmıştır. Analiz sonucuna göre etki büyüklükleri arasında açıklanamayan önemli bir varyans miktarı kaldığı (QE(df=51)=261.614, $p<.001$) görülmektedir. Sonuç olarak genel etkinin dahil edilen birincil çalışmaların doktora tezi olma durumundan etkilenmediği söylenebilir ($F(1,51) = 2.792, p=.101$).

Model sonuçlarından referans kategorisinin yani çalışmaların yüksek lisans tezi olma durumunun genel ortalama etkisinin .936 olduğunu ve bu ortalama etkinin sıfırdan önemli ölçüde farklı olduğunu göstermektedir ($t(5.13)=8.84, p<.001$). Çalışmaların doktora tezi olma deęişkeninin regresyon katsayısının anlamlılığını test etmek için kullanılan t istatistięi anlamlı deęildir ($t(9.77)=-1.92, p=.084$) ve bu sonuç omnibus testi sonucu ile uyumludur.

Yapılan analiz sonucuna göre ATBÖ uygulamalı etkinliklerin akademik başarı ile arasındaki genel ilişkinin birincil çalışmaların yüksek lisans ya da doktora tezi olup olmama durumundan etkilenmedięi sonucuna varılmıştır. Araştırmacılar kendi çalışma özelliklerine göre farklı deęişken veya deęişkenlerin moderatör etkisini incelemek için analize dahil edebilirler.

2.6.2. Üç Kategorili Verilerin Moderatör Analizi

Üç kategorili moderatör analizi için ATBÖ uygulamalı yaklaşımların uygulandıęı çalışmalar ile akademik başarı arasındaki genel ilişkinin örnekleme göre deęişip deęişmedięi incelenmiştir. Bunun için çalışmalarda kullanılan örneklem gruplarından ortaokul, lise ve lisans öğrencilerinin tanımlandığı deęişkenler kullanılmıştır. Deęişkenler veri setine kukla deęişkene dönüştürülerek dahil edilmişlerdir. Üç kategorili verilerin moderatör analizi iki kategorili verilerin analizine benzer şekilde yapılmaktadır. Bir referans grubu belirlenecek ve dięer iki grup analize aynı anda dahil edilecektir. Analizi gerçekleştirmek için gerekli kod aşağıda sunulmuştur.

```
#üç kategorili moderatör analizi (örneklem grubu)
orneklemgrubu <- rma.mv(yi, vi, mods = ~ ortaokul + lise, random = list(~1 | CalismaID, ~ 1 |
YayinID), tdist=TRUE, data=dataATBO)
summary(orneklemgrubu, digits=3)

coefs = coef_test(orneklemgrubu, cluster = dataATBO$YayinID, vcov = "CR2")
coefs
```

Kullanılan R kodunda iki kategorili moderatör analizinden farklı olarak `mods` argümanına iki deęişken “+” işareti kullanılarak eklenmektedir. Lisans öğrencilerinin referans grup olarak belirlendięi modele ilişkin R kodu çalıştırıldığında gelen sonuç ekranı Şekil 14’te verilmiştir.

Şekil 14

Üç Kategorili Moderatör Analizi Sonucu

```
Multivariate Meta-Analysis Model (k = 53; method: REML)
logLik Deviance AIC BIC AICc
-56.125 112.250 122.250 131.810 123.614

Variance Components:
      estim  sqrt  nlvls  fixed  factor
sigma^2.1 0.390 0.625   53    no  CalismaID
sigma^2.2 0.014 0.118   15    no  YayinID

Test for Residual Heterogeneity:
QE(df = 50) = 240.520, p-val < .001

Test of Moderators (coefficients 2:3):
F(df1 = 2, df2 = 50) = 1.238, p-val = 0.299

Model Results:
      estimate  se  tval  df  pval  ci.lb  ci.ub
intrcpt      0.968 0.199  4.855  50 <.001  0.568  1.369 ***
ortaokull    -0.345 0.236 -1.462  50 0.150 -0.818  0.129
lisel       -0.038 0.396 -0.097  50 0.923 -0.834  0.757
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> coefs
      Coef. Estimate  SE t-stat d.f. (Satt) p-val (Satt) Sig.
intrcpt  0.9681 0.112  8.624   3.76  0.00131 **
ortaokull -0.3445 0.184 -1.872   6.65  0.10563
lisel    -0.0383 0.201 -0.191   1.74  0.86868
```

Analiz sonucuna göre omnibus testi sonucu istatistiksel olarak anlamlı çıkmamıştır ($F(2,50)=1.238$, $p=.299$). Dolayısıyla örneklem grubunun moderatör bir etkisi olmadığı söylenebilir. Lisans öğrencileri için genel başarının ortalama etkisi .968'e eşittir ve bu etki sıfırdan önemli ölçüde sapmaktadır ($t(3.76)=8.62$, $p=.001$). Ortaokul öğrencilerinin genel ortalaması .968+(-.345)=.623'e, lise öğrencilerinin genel ortalaması ise .968+(-.038)=.93'e eşittir. Regresyon katsayıları anlamlı olmadığından bu etkiler ortalama etkiden önemli ölçüde farklı değillerdir. Araştırmacılar diğer örneklem gruplarının moderatör analiz sonuçları ile ilgilenmek isterse referans grup olarak belirleyebilir ve analizleri tekrarlayarak inceleyebilirler.

2.6.3. Sürekli Verilerin Moderatör Analizi

Tek değişkenli moderatör analizinin son örneğinde sürekli bir değişkenin analizi sunulacaktır. Bunun için birincil çalışmaların uygulanma süresinin (hafta cinsinden) genel etkiyi etkileyip etkilemediğine bakılmıştır. ATBÖ uygulamalarının uygulanma süreleri sürekli değişken olarak ele alınmış ve veri setinde genel ortalama etrafında merkezileştirilmiş değerleri hesaplanarak $phafta$ değişkeni içinde kaydedilmiştir. Analiz için gerekli kod aşağıda sunulmuştur.

```
#sürekli değişken moderatör analizi (uygulama süresi)
hafta <- rma.mv(yi, vi, mods = ~ phafta, random = list(~ 1 | CalismaID, ~1 | YayinID), tdist=TRUE,
data=dataATBO)
summary(hafta, digits=3)

coefs = coef_test(hafta, cluster = dataATBO$YayinID, vcov = "CR2")
coefs
```

İki kategorili moderatör veri analizinde uygulanan ile benzer olan kod çalıştırıldığında elde edilen analiz sonucu Şekil 15'te sunulmuştur.

Şekil 15

Sürekli Değişken Moderatör Analizi Sonucu

```
Multivariate Meta-Analysis Model (k = 53; method: REML)
logLik Deviance AIC BIC AICc
-57.702 115.404 123.404 131.131 124.273

Variance Components:
      estim  sqrt  nlvls  fixed  factor
sigma^2.1  0.402  0.634   53    no   CalismaID
sigma^2.2  0.021  0.145   15    no   YayinID

Test for Residual Heterogeneity:
QE(df = 51) = 266.167, p-val < .001

Test of Moderators (coefficient 2):
F(df1 = 1, df2 = 51) = 0.280, p-val = 0.599

Model Results:
      estimate  se  tval  df  pval  ci.lb  ci.ub
intrcpt  0.745  0.105  7.071  51  <.001  0.533  0.956 ***
phafta  -0.018  0.034  -0.529  51  0.599  -0.086  0.050
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> coefs
      Coef. Estimate  SE t-stat d.f. (Satt) p-val (Satt) Sig.
intrcpt  0.7446  0.1069  6.967  10.85  <0.001  ***
phafta  -0.0179  0.0355  -0.503  5.47  0.635
```

Analiz sonucuna göre uygulama süresi anlamlı bir moderatör değişken değildir çünkü omnibus testi istatistiksel olarak manidar değildir ($F(1,51)=.280, p=.599$). Regresyon sabiti de manidar değildir ($-0.018; t(5.47)=-.503, p=.635$). Sabit sıfırdan önemli ölçüde sapmaktadır ($t(10.85)=6.97, p<.001$), fakat bu sürekli moderatör analizinde çok önemli bir sonuç değildir. Sabit değer, ortalama uygulama süresini içeren birincil çalışmalardan elde edilen etki büyüklüklerini (yani GOM hafta değerine 0 verildiğinde) temsil etmektedir. Dolayısıyla, kategorik moderatörleri test etme prosedürlerinin aksine kesişme noktası bir referans kategorisinin ortalama etkisi olarak yorumlanmaz. Sürekli değişkenleri potansiyel moderatör analizi olarak test ederken, regresyon katsayıları ve bu değerlerin anlamlılık değerleri daha bilgilendiricidir (Assink & Wibbelink, 2016).

2.6.4. Çoklu Moderatör Analizi

Tek değişkenli moderatör analizleri yapılabileceği gibi birden fazla değişkenin bir arada incelendiği çoklu moderatör analizleri de yapılabilmektedir. Hox vd. (2010) moderatör etkilerin tek değişkenli modellerde ayrı ayrı değerlendirilmesinin yanı sıra tek bir modelde birden fazla değişkenin moderatör analizinin test edilmesinin makul bir strateji olduğunu belirtmektedir. Fakat yapılacak analizlerde değişkenlerin birbiri ile ilişkili olabileceğinin ve bu durumun çoklu bağlantılılık (*multicollinearity*) sorununa neden olabileceğinin göz önüne alınması gerekli olduğunu belirtmiştir.

Çoklu moderatör analiz örneği için çalışmalarda kullanılan soru tipi kategorik değişkeni ile uygulama süresi sürekli değişkeni (au ve phafta) seçilmiştir. Bu iki değişken moderatör değişkenler olarak aynı anda modele dahil edilmiştir. Analiz için gerekli kod aşağıdaki gibidir.

```
#Çoklu moderatör analizi
coklu_moderator <- rma.mv(yi, vi, mods = ~ phafta + au, random = list(~ 1 | CalismaID, ~ 1 |
YayinID), tdist=TRUE, data=dataATBO)
summary(coklu_moderator, digits=3)
```

```
coefs = coef_test(coklu_moderator, cluster = dataATBO$YayinID, vcov = "CR2")
coefs
```

R kodunun detayları incelendiğinde mods argümanına analize dahil edilecek değişkenlerin “+” işareti kullanılarak eklendiği görülmektedir. Kod çalıştırıldığında oluşan sonuç Şekil 16’da sunulmuştur.

Şekil 16

Çoklu Moderatör Analizi Sonucu

```
Multivariate Meta-Analysis Model (k = 53; method: REML)
logLik Deviance AIC BIC AICc
-44.913 89.826 99.826 109.386 101.190

Variance Components:
      estim  sqrt  nlvls  fixed  factor
sigma^2.1  0.180  0.424   53    no  CalismaID
sigma^2.2  0.058  0.240   15    no  YayinID

Test for Residual Heterogeneity:
QE(df = 50) = 183.318, p-val < .001

Test of Moderators (coefficients 2:3):
F(df1 = 2, df2 = 50) = 16.368, p-val < .001

Model Results:
      estimate      se  tval  df  pval  ci.lb  ci.ub
intrcpt  0.411  0.115  3.588  50 <.001  0.181  0.642 ***
phafta   0.023  0.031  0.749  50  0.457 -0.038  0.084
aul      0.894  0.156  5.721  50 <.001  0.580  1.208 ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> coefs
      Coef. Estimate      SE t-stat d.f. (Satt) p-val (Satt) Sig.
intrcpt  0.4114  0.0856  4.809  11.18 < 0.001 ***
phafta   0.0229  0.0307  0.744  6.18  0.48444
aul      0.8941  0.1892  4.725  8.75  0.00117 **
```

Analiz sonucuna göre moderatörlerden herhangi birinin regresyon katsayısı sıfırdan önemli ölçüde sapmaktadır. Çünkü omnibus testi istatistiksel olarak manidardır ($F(2,50)=16.368$, $p<.001$). Açık uçlu soruların regresyon katsayısı (.894), t testi anlamlı bir sonuç gösterdiğinden ($t(8.75)=4.73$, $p=.001$) sıfırdan önemli ölçüde sapmaktadır. Fakat çalışmaların uygulama süresi regresyon katsayısı (.023), t testi anlamlı bir sonuç göstermediğinden ($t(11.18)=4.81$, $p=.484$) sıfırdan önemli ölçüde sapmamaktadır. Bu nedenle soru tipinin açık uçlu olmasının ATBÖ uygulamalarının akademik başarı ile ilişkileri üzerinde moderatör etkiye sahip olduğu söylenebilir. Bu durumun alanyazın çalışmaları incelendiğinde ATBÖ uygulamalarının öğrenme amaçlı yazma aktiviteleri üzerine yoğunlaştığı için bireylerde yazma davranışlarının gelişimine katkı sağlamış olabilir.

Araştırmacılar çoklu moderatör analizi ile ilgili araştırmalarında çalışma, örneklem ve araştırma tasarımları ile ilgili birden fazla değişkene ait değişkeni modele dahil ederek moderatör etki analizini tekrarlayabilirler.

SONUÇ VE ÖNERİLER

Geleneksel meta-analiz uygulamalarında karşılaşılan en büyük problemlerden biri etki büyüklükleri arasındaki bağımlılık sorunudur. Bu bağımlılık sorunu ile başa çıkmak için en güçlü yöntemler arasında ise meta-analize çok düzeyli bir yaklaşım uygulamak gelmektedir. Fakat araştırmacılar tarafından çok bilinmeyen bu yöntem meta-analitik çalışmalara çok fazla

uygulanmamıştır. Çalışmada geleneksel ve çok düzeyli meta-analiz yaklaşımları arasındaki farklar açıklanmış olup ardından meta-analizin iki ve üç düzeyli modellere nasıl yansıtılacağı ifade edilmiştir. Çalışmanın amacı örnek bir veri seti üzerinden üç düzeyli rastgele etkiler modelinin R yazılımındaki `metafor` paketinin `rma.mv` fonksiyonu kullanılarak meta-analitik modellere nasıl uygulanabileceğini göstermektir. Bu amaçla argümantasyon tabanlı bilim öğrenme yaklaşımı uygulamalarının öğrencilerin akademik başarılarına etkisini belirlemek amacıyla Türkiye’de yapılmış yüksek lisans ve doktora tezleri incelenmiştir. Araştırmacıların belirlediği kriterlere uygun 15 çalışmadan 53 etki büyüklüğüne sahip özgün bir veri seti oluşturulmuş ve analizler bu örnek veri seti üzerinden gerçekleştirilmiştir. Bu uygulama okuyucuya veri dosyasının düzenlenmesi, R yazılımının hazırlanması, genel etkinin hesaplanması, çalışma içi ve çalışmalar arası varyansın heterojenliğinin incelenmesi, kategorik ve sürekli moderatör analizlerinin gerçekleştirilmesi ve çoklu moderatör analizi modellerinin incelenmesi adımlarını içeren bir kılavuz niteliği taşımaktadır. Dolayısıyla bu çalışma araştırmacılara çok düzeyli modelleri kullanarak meta-analiz uygulamalarını nasıl zenginleştirebileceklerini ve daha güçlü sonuçlar elde edebileceklerine yönelik bir farkındalık kazandırmak amacıyla gerçekleştirilmiştir.

Çok düzeyli meta-analitik modellerin analizinde kullanılacak yazılımın seçimi, araştırma sonuçlarının tekrarlanabilirliği açısından önemli bir faktördür. Yazılım seçimi tahminleyiciler, anlamlılık testleri ve elde edilen çıktılar üzerindeki etkisi nedeniyle büyük önem taşımaktadır. Pastor ve Lazowski (2018) yapmış oldukları çalışmada ikisi meta analize özgü (R’da metafor ve SPSS makrolar) ikisi çok düzeyli modellere özgü (HLM-*Hierarchical Linear Modeling* ve SAS PROD MIXED) olmak üzere dört yazılım programı performanslarını karşılaştırmışlardır. Çalışmada yazılım programları çıktıları arasındaki benzerlik ve farklılıklar detaylı şekilde incelenmiştir. Araştırmacılara esneklik sağladığı, farklı tahminleyicilerin kullanılmasına ve karşılaştırılmasına imkân vererek en iyi performans sunan programın R yazılımı metafor paketi olduğu sonucuna ulaşmışlardır. R yazılımının diğer programlara kıyasla daha esnek ve avantajlı olmasının yanında ücretsiz ve açık kaynak kodlu olması da araştırmacılar için daha ulaşılabilir olmasını sağlamaktadır.

Çok düzeyli meta-analiz modellerinin bir çalışmadan elde edilen birden fazla etki büyüklüğünün analize dahil edilebilmesi, çalışmalarda ortak olan değişkenlerin modele eklenerek incelenebilmesi, iki düzeyli modele ek düzey/düzeyler ekleyebilmesi gibi avantajları olması yanında en büyük avantajlarından biri de kayıp veri problemi ile daha etkili baş edebilmesi sayılabilir (Cooper vd., 2019; Hox & de Leeuw, 2003). Bu çalışmada, çok düzeyli meta-analitik modellerde eksik veya kayıp veri bulunmasının analiz sonuçları üzerindeki etkisi incelenmemiştir. Bu durum çalışmanın sınırlılıkları arasında görülmekle birlikte, gelecekteki araştırmacılar için bir öneri teşkil etmektedir. Eksik veri senaryoları içeren simülasyon çalışmaları yapılarak, bu tür verilerin analiz sonuçlarına olan etkileri incelenebilir ya da farklı yazılım programlarının eksik verilerin olması durumundaki performansları karşılaştırılabilir.

Özellikle eğitim ve psikoloji alanlarında yapılan çalışmaların genellikle geleneksel meta-analiz yöntemleriyle sınırlı olduğu göz önüne alındığında çalışma, bu alanlarda daha fazla çok düzeyli modellerin uygulanması gerekli olabileceğine vurgu yapmaktadır. Meta-analiz yöntemlerine çok düzeyli modellerin uygulanması etki büyüklükleri bağımlılığı varsayımı için kullanılacak en güçlü yöntemler arasında görülmektedir. Bu bağlamda bu geleneksel yöntemler yerine çok düzeyli modellerin meta-analiz çalışmalarına uygulanması önem arz etmektedir. Özellikle meta-analize çok düzeyli modeller uygulama yaklaşımının Türkçe alan yazında karşılaşılmamış olması, bu çalışmaların nasıl yapıldığının bilinmemesinden kaynaklanıyor olabilir. Bu çalışma meta-analiz yöntemlerinde çok düzeyli modellerin nasıl uygulanacağını adım adım açıklayarak araştırmacılara rehberlik etmeyi ve gelecek çalışmalara ışık tutarak yaygınlaşmasına katkı sağlama amacı taşımaktadır.

KAYNAKÇA

- APA. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, 63(9), 839-851. <https://doi.org/10.1037/0003-066X.63.9.839>
- Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, 12(3), 154-174. <https://doi.org/10.20982/tqmp.12.3.p154>
- Becker, B. J. (2000). Multivariate meta-analysis. *Handbook of applied multivariate statistics and mathematical modeling*, 499-525. <https://doi.org/10.1016/B978-012691360-6/50018-5>
- Begg, C. B., & Mazumdar, M. (1994). Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics*, 50(4), 1088-1101. <https://doi.org/10.2307/2533446>
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons. <http://doi.org/10.1002/9780470743386>
- Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, 19, 211-229. <https://doi.org/10.1037/a0032968>
- Cheung, M. W. L. (2015). *Meta-Analysis: A Structural Equation Modeling Approach*. John Wiley & Sons. <https://doi.org/10.1080/10705510802561295>
- Cheung, M. W. L. (2019). A Guide to Conducting a Meta-Analysis with Non-Independent Effect Sizes. *Neuropsychology Review*, 29(4), 387-396. <https://doi.org/10.1007/s11065-019-09415-6>
- Cheung, S. F., & Chan, D. K. S. (2004). Dependent Effect Sizes in Meta-Analysis: Incorporating the Degree of Interdependence. *Journal of Applied Psychology*, 89(5), 780-791. <https://doi.org/10.1037/0021-9010.89.5.780>
- Cheung, S. F., & Chan, D. K. S. (2014). Meta-analyzing dependent correlations: An SPSS macro and an R script. *Behavior Research Methods*, 46(2), 331-345. <https://doi.org/10.3758/s13428-013-0386-2>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. edition). Lawrence Erlbaum Associates, Publishers. <https://doi.org/10.4324/9780203771587>
- Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The handbook of research synthesis and meta-analysis* (3.edition). Russell Sage Foundation.
- Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455-463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *bmj*, 315(7109), 629-634. <https://doi.org/10.1136/bmj.315.7109.629>
- Fisher, R. A. (1933). Statistical Methods for Research Workers. *Nature*, 131(3307), Article 3307. <https://doi.org/10.1038/131383b0>
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, 5(10), 3-8. <https://doi.org/10.3102/0013189X005010003>

- Gooty, J., Banks, G. C., Loignon, A. C., Tonidandel, S., & Williams, C. E. (2021). Meta-Analyses as a Multi-Level Model. *Organizational Research Methods*, 24(2), 389-411. <https://doi.org/10.1177/1094428119857471>
- Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing Meta-Analysis with R: A Hands-On Guide* (1st edition). Chapman and Hall/CRC. <https://doi.org/10.1201/9781003107347>
- Hedges, L. V. (1982). Fitting Categorical Models to Effect Sizes from a Series of Experiments. *Journal of Educational Statistics*, 7(2), 119-137. <https://doi.org/10.3102/10769986007002119>
- Hedges, L. V., & Olkin, I. (1988). Statistical Methods for Meta-Analysis. *Journal of Educational Statistics*, 13(1), 75. <https://doi.org/10.2307/1164953>
- Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, 6(3), 203-217. <https://doi.org/10.1037/1082-989X.6.3.203>
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539-1558. <https://doi.org/10.1002/sim.1186>
- Hox, J. J., & De Leeuw, E. D. (2003). Multilevel models for meta-analysis. In *Multilevel modeling* (pp. 87-104). Psychology Press.
- Hox, J. J., Moerbeek, M., & Schoot, R. van de. (2010). *Multilevel Analysis: Techniques and Applications, Second Edition*. Routledge. <https://doi.org/10.4324/9780203852279>
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings, Second Edition*. SAGE.
- Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods*, 2(1), 61-76. <https://doi.org/10.1002/jrsm.35>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-Analysis*. SAGE Publications, Inc.
- Marín-Martínez, F., & Sánchez-Meca, J. (1999). Averaging Dependent Effect Sizes in Meta-Analysis: A Cautionary Note about Procedures. *The Spanish Journal of Psychology*, 2, 32-38. <https://doi.org/10.1017/S1138741600005436>
- Marsh, H. W., Bornmann, L., Mutz, R., Daniel, H.-D., & O'Mara, A. (2009). Gender Effects in the Peer Reviews of Grant Proposals: A Comprehensive Meta-Analysis Comparing Traditional and Multilevel Approaches. *Review of Educational Research*, 79(3), 1290-1326. <https://doi.org/10.3102/0034654309334143>
- Midway, S. (2022). Data analysis in R. Retrieved April, 8, 2023. https://bookdown.org/steve_midway/DAR/
- Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D. F. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. *The Lancet*, 354(9193), 1896-1900. [https://doi.org/10.1016/S0140-6736\(99\)04149-5](https://doi.org/10.1016/S0140-6736(99)04149-5)
- Morris, S. B. (2008). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods*, 11(2), 364-386. <https://doi.org/10.1177/1094428106291059>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting

systematic reviews. *Journal of clinical epidemiology*, 134, 103-112.
<https://doi.org/10.1136/bmj.n71>

Pastor, D. A., & Lazowski, R. A. (2018). On the Multilevel Nature of Meta-Analysis: A Tutorial, Comparison of Software Programs, and Discussion of Analytic Choices. *Multivariate Behavioral Research*, 53(1), 74-89. <https://doi.org/10.1080/00273171.2017.1365684>

Pearson. (1904). *Report On Certain Enteric Fever Inoculation Statistics*. 1243-1246.

Pustejovsky, J. (2023). *clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections (R package version 0.5.10)*. <https://cran.r-project.org/web/packages/clubSandwich/index.html>

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin*, 86(3), 638. <https://doi.org/10.1037/0033-2909.86.3.638>

Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412984997>

Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, 99(3), 400-406. <https://doi.org/10.1037/0033-2909.99.3.400>

RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA
URL <http://www.rstudio.com/>

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-Analysis With Complex Research Designs: Dealing With Dependence From Multiple Measures and Multiple Group Comparisons. *Review of Educational Research*, 84(3), 328-364. <https://doi.org/10.3102/0034654313500826>

Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., Moher, D., Becker, B. J., Sipe, T. A., & Thacker, S. B. (2000). Meta-analysis of observational studies in epidemiology: A proposal for reporting. *Jama*, 283(15), 2008-2012. <https://doi.org/10.1001/jama.283.15.2008>

Şen, S., & Akbaş, N. (2016). Çok Düzeyli Meta-Analiz Yöntemleri Üzerine Bir Çalışma. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 7(1). <https://doi.org/10.21031/epod.29995>

Şen, S., & Yıldırım, İ. (2020). CMA ile meta-analiz uygulamaları. *Ankara: ANI Yayıncılık*.

Van Den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, 45(2), 576-594. <https://doi.org/10.3758/s13428-012-0261-6>

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48. <https://doi.org/10.18637/jss.v036.i03>

EKLER

Makalemizde yer alan R koduna ve veri dosyasına ulaşmak için:
https://drive.google.com/drive/folders/1mGxCulaE-WllwRGCVusc6XL5yrZYu_K5?usp=sharing

EXTENDED ABSTRACT

Introduction

Meta-analysis *is a* statistical method that uses systematically collected quantitative data from a specific field in an attempt to reach a comprehensive understanding. One of the critical issues emphasized in the meta-analysis literature is the assumption of effect size independence (Assink & Wibbelink, 2016; Rosenthal, 1991). Traditional meta-analysis methods assume that there is no dependence between effect sizes. Dependence between effect sizes may lead to overconfidence in meta-analysis results or biased results (Van den Noortgate et al., 2013). Especially in studies conducted in the field of social sciences, several situations may lead to dependency. For example, different scenarios such as the connected studies of the same researcher group, more than one study on the same subject on the same sample, or the connected nature of studies conducted in the same country/region/school can create effect size dependence (Cheung, 2015; Van den Noortgate et al., 2013). Considering these situations, there may be dependence between different studies as well as between effect sizes obtained from the same study. Considering that studies in education and psychology generally utilize traditional meta-analysis methods, this study emphasizes the necessity of multilevel meta-analysis in these fields. Combining the multilevel model framework with the meta-analysis framework is one of the most flexible methods to address the effect size dependence.

Approaches in which multilevel models are applied to address the dependent effect sizes in meta-analysis have been popular in recent years because they provide flexible and robust statistical techniques and are generally more advantageous compared to other frameworks (Cheung, 2014; Gooty et al., 2021; Harrer et al., 2021; Hox & de Leeuw, 2003; Van den Noortgate et al., 2013). Based on these approaches, traditional meta-analysis models are considered a particular case of multilevel meta-analysis models. The data in the traditional meta-analysis model can be analyzed with multilevel models by allowing sample variance to be explained at the first level, within-study variance at the second level, and between-study variance at the third level (Assink & Wibbelink, 2016; Hox and de Leeuw, 2003).

Although multilevel meta-analysis studies to address the dependence of effect sizes are attracting attention at the international level (Assink & Wibbelink, 2016; Cheung, 2014; Gooty et al., 2021; Hox & de Leeuw, 2003; Konstantopoulos, 2011; Marsh et al., 2009; Van Den Noortgate et al., 2013), to our knowledge, the number of studies conducted in Turkey is very limited. The most relevant study in the Turkish literature, Şen and Akbaş (2016), the results obtained from traditional meta-analysis and multilevel meta-analysis were compared, but multilevel meta-analysis steps were not explained in detail. Based on our literature search, it is seen that the approach of applying multilevel models to meta-analysis has not yet been explored sufficiently. This may be due to several basic reasons. First of all, the complexity of the multilevel approach to meta-analysis may cause researchers to avoid using this method. Secondly, it may be that this method is not well-known by researchers, which is why it is not preferred. While another reason may be that there are not enough detailed studies explaining how to apply multilevel models to meta-analysis, another reason may be that most of the software/programs required to perform statistical analyses are proprietary. The fact that R is a free and open-source software might eliminate the last reason mentioned. Researchers who want to utilize multilevel meta-analysis can try to take advantage of the growing R community. Additionally, interested readers are referred to Pastor and Lazowski (2018) in which different software programs were compared to perform multilevel meta-analysis. The authors reported that R software stands out due to its advantages, such as being free, having more estimator options compared to other software, and providing more comprehensive and useful outputs than other programs. In this context, setting Turkish-speaking researchers as the target population, this study was carried out to raise awareness about the application of multilevel models to meta-analysis, to create a road map for researchers, to

show how to perform the analysis with the free and easily accessible R program, and to make significant contributions to eliminating these deficiencies in the field.

Method

The aim of this study, again setting the Turkish-speaking researchers as the target audience, is to show how the three-level random effects model can be applied to meta-analytic models using a sample data set - that we collected specifically for this tutorial- using the *rma.mv* function of the *metafor* package (Viechtbauer, 2010) in R software. This application provides the reader with a guide that includes the steps of organizing the data file, calculating the overall effect, examining the heterogeneity of variance within and between studies, performing categorical and continuous moderator analyses, and examining multiple moderator analysis models.

Result and Discussion

The sample data set of 53 effect sizes from 15 primary studies was collected to illustrate the multilevel meta-analysis. Before analyzing the effect sizes in the data set with R, how they should be organized for the three-level meta-analysis is explained. Reading data into the R environment, installing and activating the necessary packages, and utilizing the functions are explained step by step. Examples of two- and three-level meta-analytic models and moderator analyses are presented. This study aims to guide researchers by explaining step by step how to apply multilevel models in meta-analysis methods and to contribute to its widespread use by shedding light on future studies. The data set and the R script are shared as supplementary to allow the readers to attempt computational replication of the results. We also invite researchers to try alternative procedures given that the collected data set has potential challenges due to relatively larger within-study variance compared to the between-study variance.