

**ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

**META ANALİZİNDE TEMEL İSTATİSTİKSEL METOTLAR ÜZERİNE BİR
ÇALIŞMA**

Merve KOTAN

İSTATİSTİK ANABİLİM DALI

**ANKARA
2022**

Her hakkı saklıdır

ÖZET

Yüksek Lisans Tezi

META ANALİZİNDE TEMEL İSTATİSTİKSEL METOTLAR ÜZERİNE BİR ÇALIŞMA

Merve KOTAN

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Doç. Dr. İhsan KARABULUT

Meta analizi aynı konuda farklı yer, zaman ve yığın birimleri üzerinde yapılmış birbirinden bağımsız birden çok çalışmanın sonuçlarını birleştirmeye, geçerli, güvenilir ve tutarlı bir sonuca ulaşmaya yardımcı olan yöntemler bütünüdür. Bu çalışmada meta analizinde kullanılan istatistiksel yöntemlerin kısa bir değerlendirmesi yapılmıştır. Meta analizi yönteminin açıklayıcı olması için Covid-19 pandemisi sırasında 10-17 yaş aralığında bulunan adölesan çağındaki gençlerin ekran karşısında geçirdikleri süre üzerinde cinsiyetin etkisi meta analizi yoluyla değerlendirilmiştir.

Haziran 2022, 66 sayfa

Anahtar Kelimeler: Etki büyüklüğü, etki büyüklüklerinin birleştirilmesi, özet etki heterojenlik, yayın yanlılığı

ABSTRACT

Master Thesis

A STUDY ON BASIC STATISTICAL METHODS IN META ANALYSIS

Merve KOTAN

Ankara University
Graduate School of Natural and Applied Sciences
Department of Statistics

Supervisor: Assoc. Prof. Dr. İhsan KARABULUT

Meta-analysis is a set of methods that helps to combine the results of multiple independent studies on the same subject in different places, times and mass units, and to reach a valid, reliable and consistent result. In this study, a brief evaluation of the statistical methods used in meta-analysis were made. In order to explain the meta-analysis method, the effect of gender on the time spent in front of the screen by adolescents aged 10-17 during the Covid-19 pandemic was evaluated through meta-analysis.

June 2022, 66 pages

Key Words: Effect size, combining effect sizes, summary effect, heterogeneity, publication bias

TEŞEKKÜR

Meta analizi ile ilgili yaptığım çalışmada bu konuyla tanışmamı sağlayan, bana araştırma ve öğrenme fırsatı sunan, önerileriyle beni yönlendiren ve yardımlarını esirgemeyen danışmanım Doç. Dr. İhsan KARABULUT' a (Ankara Üniversitesi İstatistik Anabilim Dalı) teşekkürlerimi sunarım.

Bu süreçte beni sabırla dinleyen, her zaman yanımda olan, motive olmam için her türlü desteği gösteren ve iyi ki varlar diyerek şükrettiğim annem Beyhan KOTAN' a, babam İsmail Hakkı KOTAN' a, kardeşim Berke KOTAN' a ve arkadaşlarıma çok teşekkür ederim.

Merve KOTAN
Ankara, Haziran 2022

İÇİNDEKİLER

TEZ ONAYI

ETİK.....	i
ÖZET.....	ii
ABSTRACT	iii
TEŞEKKÜR	iv
SİMGELER DİZİNİ	vii
ŞEKİLLER DİZİNİ	viii
ÇİZELGELER DİZİNİ	ix
1. GİRİŞ	1
2. META ANALİZİ VE KANITA DAYALI ARAŞTIRMALARDAKİ YERİ.....	3
3.ÇALIŞMALARA AİT P-DEĞERLERİNİN BİRLİKTE DEĞERLENDİRİLMESİ İÇİN KULLANILAN TESTLERİN.....	5
3.1 Tippett Metodu.....	8
3.2 Wilkinson Metodu.....	9
3.3 “p” Değerlerinin Ortalaması Metodu.....	9
3.4 Stouffers Metodu	9
3.5 Fisher Metodu	10
3.6 Logit Metodu	11
4. SAYMA YÖNTEMLERİ	12
5. ETKİ BÜYÜKLÜĞÜ.....	14
5.1 Ortalamalara Göre Etki Büyüklükleri.....	20
5.1.1 Ortalamaların farkı	20
5.1.1.1 Bağımsız gruplarla yapılan çalışmalar.....	21
5.1.1.2 Bağımlı (Eşleştirilmiş) gruplar ya da ön test- son test kullanılan çalışmalar	22
5.1.2 Standartlaştırılmış ortalamaların farkı	23
5.1.2.1 Bağımsız gruplarla yapılan çalışmalar.....	23
5.1.2.2 Bağımlı (Eşleştirilmiş) gruplar ya da ön test- son test kullanılan çalışmalar	25
5.1.3 Yanıt oranları	26

5.2 Kategorik Verilerde Etki Büyüklüğü	27
5.2.1 Risk oranı.....	27
5.2.2 Odds oranı	28
5.2.3 Risk farkı	29
5.3 Korelasyonlara Dayalı Etki Büyüklükleri	30
5.4 Etki Büyüklükleri Arasında Dönüştürme.....	31
5.4.1 Log-Odds oranından d'ye dönüştürme	31
5.4.2 d'den Log- Odds oranına dönüştürme	32
5.4.3 r'den d'ye dönüştürme	32
5.4.4 d'den r'ye dönüştürme	33
6. HETEROJENLİK.....	34
6.1 Q Testi	36
6.2 I² İndeksi.....	37
6.3 Forest Grafiği	38
7. META ANALİZİNDE İSTATİSTİKSEL MODEL SEÇİMİ.....	40
7.1 Sabit Etkili Model	41
7.2 Rasgele Etkiler Modeli.....	43
8. YAYIN YANLILIĞI.....	47
8.1 Huni Grafiği.....	49
9. UYGULAMA.....	51
10. SONUÇ.....	61
KAYNAKLAR	62
ÖZGEÇMİŞ.....	66

SİMGELER DİZİNİ

n	Örneklem büyüklüğü
k	Çalışma sayısı
μ	Yığın ortalaması
$\bar{X}$	Örneklem ortalaması
$Var(.)$	Varyans
$S^2(.)$	Varyans tahmini
$SE(.)$	Standart hata
M	Özet etki (Ağırlıklı ortalama)
$\Phi(.)$	Standart normal dağılım fonksiyonu
c_α	Kritik değer
δ	Etki büyüklüğü
θ	Gerçek (bilinmeyen) etki büyüklüğü
Y	Gerçek etki büyüklüğünün tahmin edicisi
η^2	Eta kare ilişki katsayısı
R^2	İlişki katsayısı (belirtme katsayısı)
ε	Rasgele hata terimi
τ^2	Çalışmalar arası varyans parametresi
T^2	Çalışmalar arası varyans parametresinin tahmin edicisi
w_i	Varyans tahminleri için ağırlıklandırma faktörü
S_{pooled}	Ortak-havuzlandırılmış standart sapma tahmini
Δ	Ortalamaların farkının parametresi
D	Ortalamaların farkının tahmin edicisi
d	Cohen' d istatistiği
g	Hedges g istatistiği
R	Yanıt oranı istatistiği
ρ	Korelasyon parametresi
r	Korelasyon parametresinin tahmin edicisi
π	Pi sayısı
Q	Cochran heterojenlik testi
I^2	Heterojenlik indeksi

Kısaltmalar

LL	Güven aralığı alt sınır
UL	Güven aralığı üst sınır
sd	Serbestlik derecesi
KT	Kareler toplamı
KO	Kareler ortalaması

ŞEKİLLER DİZİNİ

Şekil 5.1 2. senaryoya ait histogram grafikleri	18
Şekil 5.2 3. senaryoya ait histogram grafikleri	19
Şekil 6.1 Hedges g istatistiğinin etki büyüklüğü olarak kullanıldığı meta analizi çalışmasına ait forest grafiği	39
Şekil 8.1 Başarı değişkeni ile ilgili çalışmalara ait etki büyüklüklerini gösteren huni grafiği	50



ÇİZELGELER DİZİNİ

Çizelge 5.1 1. senaryoya ait varyans analizi	17
Çizelge 5.2 2. senaryoya ait varyans analizi	18
Çizelge 5.3 3. senaryoya ait varyans analizi	19
Çizelge 5.4 p – değeri ve etki büyüklüğü arasındaki ilişki	20
Çizelge 5.5 Cohen's d ve Hedges g etki büyüklüklerinin sınıflandırması	25
Çizelge 5.6 Sigara içme durumu ile gırtlak kanseri arasındaki ilişkiyi gösteren kontenjans tablosu	27
Çizelge 7.1 Etki büyüklüklerine ilişkin şekilsel gösterimler	41
Çizelge 9.1 Meta analizine dahil edilen çalışmaların künyesi ve özet istatistikleri	52
Çizelge 9.2 Her bir çalışma için etki büyüklükleri ve etki büyüklüklerine ait varyanslar	54
Çizelge 9.3 Sabit etkili model için gerekli hesaplamalar	58
Çizelge 9.4 Rasgele etkiler modeli için gerekli hesaplamalar	59

1. GİRİŞ

İstatistiksel disiplinlerde aynı araştırma problemi ve sorusu bağlamında araştırmaların tekrarı bilimsel araştırmaların özünde yer alır. Rasgeleliğin ve hataların varlığı bu tekrarlamaları zorunlu kılar. Aynı araştırma problemi üzerinde yapılmış çalışmaların bazılarında yokluk hipotezi reddedilirken, bazılarında yokluk hipotezi reddedilemeyebilir. Bu da konu üzerinde geniş bir sonuç yelpazesi sunar. Meta analizi aynı konu üzerinde birbirinden bağımsız olarak yapılmış, bir etkiyi tahmin eden veya test eden bilinen tüm bilimsel çalışmaların etkinin genel tahminine ulaşabilmek için uygun istatistiksel yöntemler ile birleştirilmesi için kullanılan bir yöntemdir. Aynı konu üzerinde yapılan herhangi bir çalışmanın örneklem büyüklüğü sınırlıdır ve bir çalışmanın gücü örneklem büyüklüğüne bağlıdır. Bu yüzden bilimsel bir soru meta analiz ile ele alındığında her bir çalışmanın gücü birleştirilebilir. Meta analizi, küçük ama önemli etkiler tespit edebilir ve istatistiksel gücü büyük ölçüde artırabilir (Whitlock ve Schuller 2015; Borenstein vd. 2009). Bu çalışmada meta analizinde kullanılan istatistiksel yöntemlerin kısa bir değerlendirmesi ve bir meta analizi uygulaması yapılacaktır.

Çalışmanın ikinci bölümünde meta analizi ve kanıta dayalı araştırmalardaki yeri ve meta analizde kullanılan kavramlara değinilecektir.

Çalışmanın üçüncü bölümünde nicel araştırma sentezinde kullanılan bir yöntem olan çalışmalara ait p – değerlerinin birleştirilmesinde kullanılan metotlar detaylı olarak açıklanacaktır.

Çalışmanın dördüncü bölümünde nicel araştırma sentezinde kullanılan bir diğer yöntem olan sayma yöntemleri sunularak yöntemin olumlu ve olumsuz yönlerine değinilecektir.

Çalışmanın beşinci bölümünde meta analizinde kullanılan etki büyüklüğü kavramı açıklanacak ve verinin ölçüm düzeyine bağlı olarak kullanılan etki büyüklüğü indekslerine detaylı olarak değinilecektir.

Çalışmanın altıncı bölümünde heterojenlik kavramına ve heterojenlik testlerine değinilecektir.

Çalışmanın yedinci bölümünde etki büyüklüklerinin birleştirilmesi için hangi istatistiksel modelin kullanılması gerektiği, her iki model için özet etki tahmininin ve güven aralıkları tahminlerinin nasıl yapılacağına değinilecektir.

Çalışmanın sekizinci bölümünde meta analizinde karşılaşılabilen yayın yanlılığı sorunundan, nasıl meydana geldiğinden ve tespit edilme yöntemlerinden bahsedilecektir.

Çalışmanın dokuzuncu bölümünde meta analizi yönteminin açıklayıcı olması için 10-17 yaş aralığında bulunan adölesan çağındaki gençlerin ekran karşısında geçirdikleri süre üzerinde cinsiyetin etkisi meta analizi yoluyla değerlendirilecektir.

2. META ANALİZİ VE KANITA DAYALI ARAŞTIRMALARDAKİ YERİ

Sosyal bilimler, fen bilimleri ve sağlık bilimlerinde aynı araştırma problemi ve sorusu çerçevesinde araştırmaların tekrarı bilimsel araştırmaların özünde yer alır. Gözlemlerin rasgeleliği, ölçüm hatalarının varlığı ve metodolojik zorluklar yapılan deneylerin tekrarlanmasını gerekli kılar. Seçilen örneklem yığını ne kadar iyi temsil ederse etsin yapılacak tahmin ve çıkarımlar hatalar ve yanlışlıklar içerebilir, aynı şekilde istatistiksel hipotezler ve sonuçlar ne olursa olsun iki tip hatayı (I. Tip ve II. Tip hata) içerecektir. Bilimsel çalışmalarda (matematik ve mantık disiplinlerini dışarıda tutarak) bilimsel bir teori hiçbir zaman kanıtlanamaz. Yanıltmadığı sürece olgu ve olaylar hakkındaki bilimsel açıklamalar başarılı bir şekilde kullanılmaya devam edilir. Daha başarılı bilimsel teorilerle teknoloji gelişir, daha duyarlı ölçme araçları geliştirilir ve deneyler tasarlanır. Başarısı azalan teoriler yeni teorilerle yer değiştirir, kullanım kapsamları daralır ve bazen de yanlış ve eksiklikleri olduğu düşünülür. Günümüzde bilimsel çalışmaların sayısı hızla artış göstermektedir. Bu artışla birlikte aynı araştırma problemi üzerinde yapılmış birbirinden bağımsız çalışmalarda bazen tutarsız sonuçlara ulaşılabilmektedir. Diğer yandan aynı araştırma problemi üzerinde yapılan çalışmalarda sonuçlarının birleştirilmesi yeni araştırma problemleri belirlenmesine ve sonuçların nasıl birleştirilmesi gerektiğine ilişkin soruları beraberinde getirir.

Farklı çalışmalardan elde edilen istatistiksel kanıtları (statistical evidences) birleştirmek için iki farklı yaklaşıma başvurulmuştur. Birinci yaklaşım, çalışmalara ait p – değerlerinin birlikte değerlendirilmesine dayanırken, diğeri çalışmalara ait işlem etkilerinin tahmin edilmesine dayanır. Her iki yaklaşımın varlığı 1930’ lardan beri bilinmektedir. Meta analizi, aynı ya da benzer araştırma problemleri için yapılan çalışmalara ait etki büyüklüklerinin (ilişki katsayıları veya p –değeri gibi) uygun istatistiksel yöntemler kullanılarak birleştirilmesi ve daha güvenilir, geçerli sonuç elde edilebilmesi için kullanılan istatistiksel bir yöntemdir (Hedges ve Olkin 1985).

Meta analizi bu adı almadan önce de aynı konu üzerinde birbirleriyle ilgili olduğu düşünülen bilimsel çalışma ve araştırma sonuçları belli ilkelere dayanmadan bir literatür taraması biçiminde bilimsel yazında yer almaktaydı. Bugün kullanılan meta analizi

araçlarının gelişimi de bu sıralarda devam etmekteydi. Bu çalışmaların meta analizi başlığı altında yer alması ve anılması Glass (1976)'a dayandırılır. Meta analizi genellikle dahil edilen her bir çalışmaya özet istatistikler (descriptive) biçimindeki verilere dayandığından, istatistiksel analiz sonuçlarının bir sentezidir (Hedges ve Olkin 1985). Bir meta analizi, belirli bir etkinin tahmin edildiği veya etkinin anlamlılığının test edildiği bazı ölçütlere uyan bilimsel çalışmaları derler ve genel etkinin tahminini elde etmek için bunları nicel olarak birleştirir (Whitlock ve Schuller 2015). Meta analizine dahil edilen çalışmaların sonuçlarının geçerlilik ve güvenilirlik açısından uygun olması önemlidir.



3. ÇALIŞMALARA AİT P-DEĞERLERİNİN BİRLİKTE DEĞERLENDİRİLMESİ İÇİN KULLANILAN TESTLERİN

Çalışmalara ait p-değerlerinin birlikte değerlendirilmesi için kullanılan testler, nicel araştırma sentezi için geliştirilen ilk istatistiksel yöntemlerdir. Bu yöntemlere çok amaçlı veya parametrik olmayan yöntemler denir. Bu yöntemler test istatistiklerinden gözlenen p-değerlerine bağlı olduğundan, verilerin dağılımına bağlı değildir. Çalışmalara ait p-değerlerinin birlikte değerlendirilmesi için kullanılan testlerin, verilerin dağılımına bağlı olmaması uygulamalarda büyük kolaylık sağlar. Kilit nokta, H_0 yokluk hipotezinin doğruluğu altında sürekli test istatistiklerinden gözlenen p-değerlerinin farklı istatistikler ve farklı örnekler altında düzgün dağılıma sahip olmasıdır.

Çalışmalara ait p-değerlerinin birlikte değerlendirilmesi için kullanılan birkaç farklı yöntem vardır. Bu yöntemlerin hangisinin kullanılması gerektiğini belirlemek için seçime rehberlik eden genel analitik sonuçlar vardır. Bu sonuçların her biri, bazı kriterlere göre optimal olan sonuçları birleştirmek ve bu test yöntemlerini ortaya çıkarmak için tasarlanmıştır. Kriterlerin hiçbirisi çalışmalara ait p-değerlerinin birlikte değerlendirilmesi için kullanılan yöntemlerin tüm özelliklerini kapsamaz ve bir kritere göre optimal olan yöntemler, başka bir kritere göre optimal olmayabilir. Bu yüzden karar vericiler bu kriterlerin hepsini dikkate alarak seçim yapamazlar, çünkü bu kayıpları arttırır.

Genel analitik sonuçlardan birincisi; çalışmalara ait p-değerlerinin birlikte değerlendirilmesi için kullanılan testlerin kabul edilebilirliğidir. Kabul edilebilir bir test bir şey için en iyidir (Burada en iyi terimi, başka bir testin eşit derecede iyi olduğu durumu içerir). Kabul edilebilir testler birbirlerine karşı üstün değildir, ancak kabul edilemeye karşı üstündür. Kabul edilebilirlik yararlı bir ilke olsa da, kabul edilebilirliğin yararlılığı pratikte başka nedenlerden dolayı sorunludur. Kabul edilemez olduğu bilinen bir aday test olduğu varsayılın. O halde bu aday testten daha iyi bir test vardır. Ancak belirli bir alternatif için daha iyi bir testin var olduğu bilinse bile bu test

bulunamayabilir. Bu nedenle, daha iyi bir testin bulunabildiği durumlar dışında kabul edilemez testler kullanılabilir.

Genel analitik sonuçlardan ikincisi; monotonluktur. Monotonluk özelliği, çalışmalara ait p-değerlerinin birlikte değerlendirilmesi için kullanılan test yöntemleri, p-değer kümesi için H_0 yokluk hipotezini reddederse aynı zamanda bileşen bazında daha küçük herhangi bir p-değer kümesi için de H_0 yokluk hipotezinin reddedilmesini gerektirir. Daha açık bir ifadeyle, $p_1, \dots, p_k$ değer kümesi H_0 yokluk hipotezinin reddedilmesine yol açarsa, $p_1^*, \dots, p_k^*$ değer kümesi çalışmalara ait p-değerlerinin birlikte değerlendirilmesi için kullanılan test yöntemi monoton olarak adlandırılır. Burada,

$$p_1^* \leq p_1, \dots, p_k^* \leq p_k \quad i = 1, \dots, k \quad (3.1)$$

dır. O halde $p_1^*, \dots, p_k^*$ anlamlılık düzeyleri kümesi de H_0 yokluk hipotezinin reddedilmesine yol açar (Hedges ve Olkin 1985).

Monotonluk ve kabul edilebilirlik ilkelerini birbirine bağlayan önemli sonuç, monoton olan test yönteminin kabul edilebilir olması ve bu nedenle bazı test durumları için optimal olmasıdır. Monoton olan bir test kabul edilebilirdir (Birnbaum 1954).

$T_1, T_2, \dots, T_k$ test istatistiklerinin her birinin dağılımı tek parametrelili ailedense, p-değerlerinin birlikte değerlendirmesi için kullanılan test dışbükeydir. Böylece, testin kabul edilebilirliğinin belirlenmesi için testin kabul bölgesinin dışbükeyliğinin belirlenmesi yeterli olacaktır. Bir bölgenin dışbükey olup olmadığı kolayca belirlenebilir, örneğin bölgedeki herhangi iki nokta için iki noktayı birleştiren doğru tümüyle bölgenin içindeyse bölge dışbükeydir.

Genel analitik sonuçlardan üçüncüsü ise etkinliktir. Etkinliğin birçok tanımı vardır. Belirli bir etkinlik ölçüsünü hesaplamak genellikle kolay değildir. Etkinlik ölçütleri; belirli bir güç düzeyini elde etmek için gereken örneklem büyüklüklerinin karşılaştırmasını, aynı örneklem büyüklüğünde güç karşılaştırmalarını ve anlamlılık

düzeylerinin karşılaştırılmalarını içerir. Test, örneklem büyük olduğunda "en anlamlı" sonuçları veriyorsa, B-optimal testi (Bahadur verimliliği) olarak adlandırılır. Bu durumda test, büyük örneklem için mümkün olan en küçük anlamlılık düzeyine sahip olacaktır. Etkinliğin bu tanımı, anlamlılık düzeyinin sıfıra yaklaşması temeline dayanmaktadır. B-optimal olan bir testin anlamlılık düzeyi, örneklem büyüklüğü arttıkça hızlı bir şekilde sıfır olma eğilimindedir. Bununla birlikte, birkaç farklı test B-optimal olabilir, bu nedenle bu kavram her zaman belirli bir testi ayırmak için yeterli olmayabilir (Hedges ve Olkin 1985).

Ortalamalar, ortalamaların farkları veya korelasyonlar gibi $\theta_1, \theta_2, \dots, \theta_k$ parametreleriyle karakterize edilen k tane bağımsız çalışma olduğu varsayalım. i. çalışmanın yokluk hipotezini test etmek için kullanılacak bir test istatistiği T_i ürettiği varsayalım ve hipotez,

$$H_{0i}: \theta_i = 0 \quad i = 1, 2, \dots, k \quad (3.3)$$

biçiminde verilsin. Test istatistiğinin büyük değerleri H_0 yokluk hipotezinin reddedilmesine yol açar. Yokluk hipotezi H_0 , etkilerin anlamlı olmadığına, yani tüm θ 'ların sıfır olduğu anlamına gelmektedir. Hipotez,

$$H_0: \theta_1 = \theta_2 = \dots = \theta_k = 0 \quad (3.4)$$

biçiminde verilir. i. çalışma için tek kuyruklu p-değeri;

$$p_i = P(T_i \geq t_{i0}) \quad (3.5)$$

biçimindedir, ve t_{i0} i. çalışmada elde edilen test istatistiğinin değerini göstermektedir. Burada $T_1, T_2, \dots, T_k$ istatistiklerinin dağılımlarının aynı olması gerekmez. H_{0i} yokluk hipotezi doğru ise p_i , $[0, 1]$ aralığında düzgün dağılıma sahiptir.

Alternatif hipotezler arasında ayırım yapmak yararlıdır. Çalışmalara ait θ_i parametreleri negatif veya pozitif değerler alabilir.

θ parametresinin, $\theta_i \geq 0$ olduğu önceden biliniyor olsun bu durumda alternatif hipotez;

$$H_1: \text{En az bir } \theta_i > 0, i = 1, 2, \dots, k \quad (3.6)$$

dir. H_1 hipotezi, varyans analizinde kullanılan F-istatistikleri ile benzerlik taşır.

θ parametresinin, $\theta_i \leq 0$ $\theta_i \geq 0$ gibi tek taraflı fakat negatif veya pozitif olduğu bilinmiyorsa alternatif hipotez;

$$H_2: \text{En az biri } \theta_i \neq 0, i = 1, 2, \dots, k \quad (3.7)$$

dir. H_2 hipotezi, farklılıkların tutarlı olduğunu ancak her iki yönde de olabileceğini belirtir. Bu alternatif hipotez, bir t-istatistiği veya bir korelasyon katsayısı ile ilgili test istatistikleriyle benzerlik taşır.

Çalışmaların hepsinin aynı işarete sahip olmadığı ve farklı etkiler gösteren çalışmalar için iki yönlü alternatif hipotez;

$$H_3: \text{En az biri } \theta_i \neq 0 \quad i = 1, 2, \dots, k \quad (3.8)$$

dir. H_3 hipotezi, büyük veya küçük değerleri reddeden ve farklı araştırmalardaki etkilerin işaretlerinin aynı olmasını gerektirmeyen istatistikleri kullanan çalışmaları birleştirirken geçerlidir (Hedges ve Olkin 1985).

3.1 Tippett Metodu

Sonuçların birlikte değerlendirilmesine ilişkin ilk test Tippett tarafından önerilmiştir. $p_1, \dots, p_k$ birbirinden bağımsız p – değerleri (sürekli test istatistiklerinden) ise, her biri H_0 hipotezi altında düzgün dağılıma sahiptir. Bu nedenle, $p_{[1]}, p_1, \dots, p_k$ 'nin minimum sıra istatistiği olsun, $p_{[1]}$ ile $1 - (1 - \alpha)^{1/k}$ α anlamlılık düzeyinde karşılaştırılır. Eğer;

$$p_{[1]} < 1 - (1 - \alpha)^{1/k} \quad (3.9)$$

ise H_0 hipotezi reddedilir. Veriler üstel ailenin bir üyesi (süreklili) ise Tippett testinin tercih edilmesi gerekir (Hedges ve Olkin 1985).

3.2 Wilkinson Metodu

Tippett'in prosedürünün bir genellemesi Wilkinson'a (1951) dayanmaktadır ve bir test istatistiği olarak sadece en küçük p – değerini değil, aynı zamanda r . en küçük p – değeri $p[r]$ ' yi kullanır (Hedges ve Olkin 1985).

Wilkinson'a bağlı olarak bu yöntem, r . en küçük p – değeri $p[r]$ küçükse, yani sabit r değeri için bir c kritik değerinden küçükse H_0 hipotezi reddedilir. H_0 hipotezi altında $p[r]$, r ve $k - r + 1$ parametreleriyle beta dağılır. Bu test için c kritik değeri aşağıdaki denklemden belirlenir (Hartung vd. 2008).

$$\alpha = \int_0^c \frac{u^{r-1}(1-u)^{k-r}}{\beta(r, k-r+1)} du \quad (3.10)$$

3.3 “ p ” Değerlerinin Ortalaması Metodu

p değerlerinin ortalamasına dayalı testte, ortalama küçük olduğunda H_0 hipotezi reddedilir. Ancak, küçük k değerleri için ortalamanın dağılımı basit değildir ve bu nedenle bu test nadir kullanılır. Öte yandan, büyük k değerleri için, p_i değerlerinin ortalamasının dağılımına merkezi limit teoremi kullanılarak normal dağılımla yaklaşımda bulunulabilir (Hartung vd. 2008).

3.4 Stouffers Metodu

Çalışmalara ait p – değerlerinin birlikte değerlendirilmesi için kullanılan bir başka yöntem Stouffer Suchman, DeVinney, Star ve Williams (1949) ve Liptak (1958) tarafından bağımsız olarak önerilen ters normal yöntemdir (Hedges ve Olkin 1985). Bu

yöntem aslında Stouffer ve arkadaşlarına ait olup tam olarak Mosteller ve Bush (1954) ile Rosenthal (1978) tarafından tanımlanmıştır. Çalışmada z test istatistiği kullanıldığında tercih edilen bir yöntemdir. Bu yöntem, her bir p_i değerinin normal dağılım kritik ya da yüzdelik değerlerine (z değerine) dönüştürülmesini sağlar. $\Phi(\cdot)$ standart normal dağılım fonksiyonunu gösterir, Z_i standart normal dağılımlı rasgele değişken olmak üzere,

$p_i = \Phi(Z_i) \sim \text{Düzgün}(0,1)$ dağılımlı olarak tanımlanır. H_0 hipotezi doğru varsayıldığında,

$$Z = \frac{Z_1 + \dots + Z_k}{\sqrt{k}} = \frac{\Phi^{-1}(p_1) + \dots + \Phi^{-1}(p_k)}{\sqrt{k}} \quad (3.11)$$

istatistiği standart normal dağılıma sahiptir. Bu yöneme ilişkin alternatif hipotezin H_3 olduğu çalışmaların işaretlerinin farklı olduğu ve en az birinin 0 dan farklı olduğu duruma uygundur. Buna göre test istatistiği,

$$Z \geq z_{1-\frac{\alpha}{2}} \text{ veya } Z \leq -z_{1-\frac{\alpha}{2}}$$

ise H_3 reddedilir.

3.5 Fisher Metodu

Ters ki-kare dönüşümünün özel bir durumu olan bu yöntem Fisher (1932) tarafından tanımlanmıştır ve meta analizinde yaygın olarak kullanılmaktadır (Hartung vd. 2008). Verilen k bağımsız çalışma ve p değerleri $p_1, \dots, p_k$ p değerlerini birleştirmek için $p_1 p_2 \dots p_k$ değerlerini kullanır. Fisher, düzgün dağılım ile ki-kare dağılımı arasında bir bağlantı kurmuştur. U rasgele değişkeni düzgün bir dağılıma sahip ise $-2 \log U$ iki serbestlik dereceli ki-kare dağılır. Sonuç olarak; H_0 hipotezi doğru ise $-2 \log p_i$ iki serbestlik dereceli ki-kare dağılır. Buradan bu denklem,

$$-2 \log(p_1 p_2 \dots p_k) = -2 \log p_1 - 2 \log p_2 - \dots - 2 \log p_k \quad (3.12)$$

biçiminde verilir. Bağımsız ki-kare değişkenlerinin toplamı da ki-kare olduğundan bu dağılım $2k$ serbestlik derecesine sahip bir ki-kare dağılımıdır.

$$P = -2 \sum_{i=1}^k \log p_i \geq c \quad (3.13)$$

ise H_0 hipotezi reddedilir.

Burada c kritik değeri $2k$ serbestlik dereceli ki-kare dağılımının üst kuyruğundan elde edilir (Hedges ve Olkin 1985).

3.6 Logit Metodu

Çalışmalara ait p – değerlerinin birlikte değerlendirilmesi için kullanılan başka bir yöntem George (1977) tarafından önerilmiştir. Her bir p – değeri bir logite, $\log [p / (1 - p)]$ ' ye dönüştürülür ve ardından logitler birleştirilerek,

$$L = \log \frac{p_1}{1-p_1} + \dots + \log \frac{p_k}{1-p_k} \quad (3.14)$$

elde edilir. L 'nin dağılımı basit değildir, ancak H_0 hipotezi doğru olduğunda George ve Mudholkar (1979), L 'nin dağılımının (sabit hariç) $5k + 4$ serbestlik dereceli Student t-dağılımı ile yaklaşık olarak tahmin edilebileceğini göstermiştir. Bu nedenle, logit istatistiğini kullanan test yöntemi,

$$L^* = |L| \sqrt{\frac{(0.3)(5k+4)}{k(5k+2)}} > c \quad (3.15)$$

biçimindedir, burada kritik değer olan c , $5k + 4$ serbestlik dereceli t-dağılımdan elde edilir. (3.15'deki 0.3 terimi, $3/\pi^2$ ile daha doğru bir şekilde verilir)

Büyük k değerleri için;

$$\sqrt{\frac{(0.3)(5k+4)}{\pi^2(5k+2)}} \cong 0.55 \quad (3.16)$$

$$L^* \cong \left[\frac{0.55}{\sqrt{k}} \right] L \quad (3.17)$$

biçimindedir (Hedges ve Olkin 1985).

4. SAYMA YÖNTEMLERİ

Sayma yöntemleri nicel araştırma sentezi için geliştirilen bir yöntemdir. Her bir çalışmadan çok az bilgi gerektirir. Örneğin; ortalamaların farkının işaretini, korelasyon katsayısının işaretini veya bir hipotez testinin istatistiksel olarak anlamlı bir sonuç verip vermediğini bilmek yeterlidir. Geleneksel sayma veya kutu-puan incelemesi, çalışmalardaki hipotez testlerinin sonuçlarına göre çalışmaları kategorilere ayırarak işlem etkilerinin varlığı hakkında çıkarımlar yapmaya çalışır. Geleneksel sayma yöntemlerinde kullanılan çıkarımlar kusurludur ve yanıltıcı olabilir. Sezgisel olarak, çalışmaların büyük bir kısmı istatistiksel olarak anlamlı olursa bu durum etkinin sıfır olmadığının kanıtı olabilir. Bunun tersine, az sayıda çalışma istatistiksel olarak anlamlı olursa, etkinin sıfır olmadığına dair kanıtlar zayıf görünecektir.

Sayma yönteminde mevcut çalışmalar; ortalamaların farkının pozitif olduğu anlamlı sonuç verenler, ortalamaların farkının negatif olduğu anlamlı sonuç verenler ve anlamlı sonuç vermeyenler olmak üzere üç kategoriye ayrılır.

Pozitif anlamlı bir sonucun olasılığı,

$$p \equiv P(\text{anlamlı sonuç} | \delta, n) = \int_{c_\alpha}^{\infty} f(t; \delta, n) d_t \quad (4.1)$$

biçimindedir, burada $f(t; \delta, n)$, etki büyüklüğü δ olan n büyüklüğündeki örnek için t istatistiğinin olasılık yoğunluk fonksiyonudur ve c_α t istatistiğinin kritik değeridir.

Her bir denemenin bir başarı (pozitif anlamlı bir sonuç) veya bir başarısızlık (anlamlı olmayan veya negatif anlamlı sonuç) olarak sonuçlandığı bir dizi deneme olduğu varsayalım. Her çalışmanın anlamlı sonuç verme olasılığı eşittir ve p ile gösterilir. Bu olasılık, her bir çalışmadaki istatistiksel testin gücüdür. Bu çalışmalarda başarıların sayısının p parametrelili binom dağılımına sahip olduğu bilinir. Sayma yöntemlerinde pozitif anlamlı sonuçlara sahip çalışmaların oranı üçte biri (veya daha genel olarak kritik değer olan C_0) aşarsa, bir etkinin yani δ 'nın sıfırdan büyük olduğuna karar verilir. Küçük örneklerde, pozitif anlamlı sonuçlar veren k tane çalışmanın önceden belirlenmiş

bir C_0 kritik deęerini ařma olasılıęını hesaplamak için binom daęılımını kullanılır (Hedges ve Olkin 1985).

X , başarı sayısını göstermek üzere,

$$\begin{aligned} X &\sim \text{Binom}(k, p) \\ P(X > C_0) &= P\left\{\frac{X}{k} > C_0\right\} \\ &= \sum_{i=[C_0+1]}^k \binom{k}{i} p^i (1-p)^{k-i} \end{aligned} \quad (4.2)$$

dir. C_0 kritik deęeri 0 ile 1 arasında deęer alır.

Etki büyüklüęü;

$$\delta = \frac{\mu_E - \mu_C}{\sigma} \quad (4.3)$$

olarak tanımlansın. Burada μ_E ve μ_C sırasıyla deney ve kontrol grupları için yığın ortalamaları ve σ ortak standart sapmadır. Etki büyüklüęü küçük ($\delta < 0.5$) olduęunda, standart sayım yöntemi, etkileri tespit etmede başarısız olur. Dahil edilen çalışmaların sayısı arttıkça durum deęiřmez. Örneęin, $\delta < 0.3$ olduęunda, çalışma sayısı 10'dan 50'ye çıkarılırsa standart sayımın bir etkiyi tespit etme olasılıęı azalır. Sayma yöntemi kanıt miktarı kadar yanlış karar verme eğilimindedir. Sezgisel düzeyde, her çalışmadaki hipotez testinin yanlış bir karar verme olasılıęı vardır. Bu durumda etki sıfır deęildir ve yanlış karar, gerçek (sıfır olmayan) bir etkiyi tespit etmede başarısız olur. Aslında, çalışmaların sayısı arttıkça, önemli sonuçlar veren çalışmaların oranı, yaklaşık olarak testin gücü kadardır. İstatistiksel testlerin ortalama gücü p , kesme kriteri C_0 'dan küçükse, sayımın doęru kararı verme olasılıęı, çalışmaların sayısı arttıkça sıfır olma eğilimindedir. Bu nedenle, $p < C_0$ ise, bir karar prosedürü olarak sayımın gücü, çalışma sayısı büyüdükçe sıfır olma eğilimindedir (Hedges ve Olkin 1980). Sayma yöntemleri, istatistiksel olarak anlamlı olan çalışmaların sayısının istatistiksel olarak anlamlı olmayan çalışmaların sayısı ile karşılaştırılması işlemidir. Oy sayımı, anlamsız bir p – deęerini, etkinin olmaması şeklinde deęerlendirir. Küçük, orta ve büyük etki büyüklükleri, düşük istatistiksel güç nedeniyle anlamlı olmayan bir p – deęeri verebilir. Bu nedenle oy sayımı hiçbir zaman geçerli bir yaklaşım deęildir (Borenstein vd. 2009).

5. ETKİ BÜYÜKLÜĞÜ

Etki büyüklüğü; işlem etkisinin büyüklüğünü, iki değişken arasındaki ilişkinin gücünü veya iki grup arasındaki farkı yansıtan bir indekstir. Etki büyüklüğü terimi farklı alanlarda farklı şekillerde kullanılabilir. Örneğin, tıpta etki büyüklüğü genellikle “tedavi etkisi” olarak kullanılır ve odds oranlarına, risk oranlarına veya risk farklılıklarına dayanır. Sosyal bilimlerde ise olduğu gibi “etki büyüklüğü” terimi kullanılır ve etki büyüklüğü genellikle standartlaştırılmış ortalamaların farkları veya korelasyonları ifade eder.

Meta analizine dahil edilecek çalışmalar belirlenirken üç temel özelliğe dikkat edilmelidir. Birincisi, farklı çalışmalardan ve benzer yığınlar üzerinde çalışılarak elde edilen etki büyüklüklerinin, aynı şeyi (karakteristik veya değişkeni) ölçüyor olmaları, yani; birbirleriyle karşılaştırılabilir olmaları gerekir. İkincisi, etki büyüklüğü tahminlerinin yayımlanmış araştırma raporlarında bulunması ve hesaplanabilir olmasıdır. Çalışmalara ait ilgilenilen etki büyüklüğü hesaplamalarında kullanılacak istatistiklerin bulunması ve verilerin yeniden analizine ihtiyaç duyulmaması gerekir. Üçüncüsü ise konu olan etki büyüklüğünün istatistiksel analize elverişli olması beklenir. İstatistiklere ait örneklem dağılımları kolayca elde edilebilmelidir. Örneğin; varyans tahminlerinin yapılabilmesi, güven aralıklarının hesaplanabilmesi ancak örneklem dağılımı hakkında bilgi varsa olanaklıdır. Bu üç özelliğe ek olarak, ele alınan etki büyüklüğü anlamlandırılabilir veya yorumlanabilir olmalıdır.

Dikkate alınacak etki büyüklüğü genellikle çalışmanın türüne veya çalışmada yer alan değişkenlerin ölçüm düzeylerine bağlıdır. Meta analizine dahil edilen çalışmalarda verilen istatistikler, iki gruptaki ortalamalara ve standart sapmalara dayalıysa uygun etki büyüklüğü genellikle ortalamaların farkı, standartlaştırılmış ortalamaların farkı veya yanıt oranı (response ratio) olacaktır. Bir varyans analizi çalışmasında etki büyüklüğü olarak η^2 (eta kare) seçilirken, bir doğrusal regresyon çalışmasında etki büyüklüğü olarak R^2 kullanılır (Özsoy ve Özsoy 2013). İstatistikler, kategorik bir sonuca dayalıysa uygun etki büyüklüğü genellikle risk oranı, odds oranı veya risk farkı olacaktır.

İstatistikler, iki değişken arasında bir ilişkiye dayalıysa uygun etki büyüklüğü ilişki katsayısı olacaktır.

Meta analizinde her bir çalışmada ele alınan etki büyüklükleri kullanılarak bir özet etki büyüklüğü (summary effect) hesaplanır. Etki büyüklüğüne ilişkin duyarlılığı (precision) daha yüksek (varyansı daha küçük) olan çalışmalar daha fazla bilgi taşır ve bu çalışmalara meta analizinde daha fazla ağırlık verilir (Borenstein vd. 2009).

Genellikle gerçek etki büyüklüğü bir parametre olarak “ θ ” ile tahmini ise “ Y ” ile gösterilir.

Etki büyüklüğünün istatistiksel çözümlemelerdeki yeri ve önemi hakkında bir düşünceye sahip olmak için Tarlow (2016)’un öğretici çalışmasından yararlanılacaktır. Tek yönlü (faktörlü) varyans analizi (ANOVA) için yapılan çalışmada, Tarlow (2016) hipotezlerin testinde kullanılan p – değeri ve η^2 etki büyüklükleri arasındaki ilişkiyi ele almaktadır. Bu makalede açıklayıcı değişkenin 3 düzeye sahip olduğu sabit (durağan) etkili ANOVA için bilgisayarla veri üreterek 3 farklı senaryo üzerinde değerlendirmeler yapılmıştır. Ele alınan model $\varepsilon_{ij} \sim N(0, \sigma^2)$ b.b.a.d. $\sum_{i=1}^3 \alpha_i = 0$ olmak üzere

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad i = 1, 2, 3; \quad j = 1, 2, \dots, 8; \quad n_1 = n_2 = n_3$$

dir. Burada,

Y_{ij} : i. denemedeki j. gözlem değeri

α_i : i. denemenin etkisi

μ : Genel ortalama

ε_{ij} : Rasgele hata terimleri

ifade edilmektedir.

Bir ANOVA tablosunda p -değeri ve etki büyüklüğü η^2 modelin anlamlılık değerlendirilmesinde kullanılır. Modelin etki büyüklüğü η^2 temelde tüm örnek ortalamaları arasındaki farkların büyüklüğünün bir ölçüsüdür ve η^2 'nin büyük olması düzeylere ait örnek ortalamalarının birbirinden çok farklı olduğunu, düzey etkilerinin önemli ölçüde farklılaştığını gösterir. Diğer taraftan p – değeri hipotez testinde istatistiksel kanıtın büyüklüğüne ilişkin bir niceliktir.

η^2 bir ilişki katsayısı olup ölçme düzeyi nominal bir değişken ile ölçme düzeyi en az eşit aralıklı olan iki değişken arasındaki ilişkinin ölçüsünü belirler, $0 \leq \eta^2 \leq 1$ arasında değer alır, η^2 değeri 1'e yaklaştıkça ilişkinin güçlendiği anlaşılır.

η^2 'nin tahmin edicisi:

$$\hat{\eta}^2 = \frac{\sum_{j=1}^J n_j (\bar{x}_{.j} - \bar{x}_{..})^2}{\sum_{j=1}^J \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{..})^2} \quad (5.1)$$

biçimindedir. Burada,

$\bar{x}_{..}$: Örneklem genel ortalamasını,

$\bar{x}_{.j}$: j.düzeye ait örnek ortalamasını,

ifade eder. η^2 ilişki katsayısı bağımlı değişkendeki toplam varyansın ne oranda bağımsız değişkenle açıklanabildiğini gösterir (Öngel 1980).

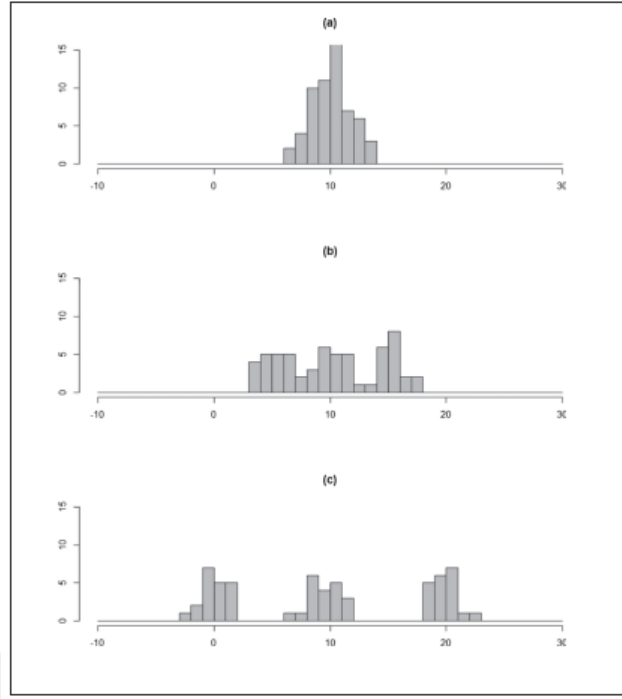
Tarlow'un makalesindeki ilk senaryo Amerika'da SAT Matematik sınavına tabi tutulmuş öğrencilerin test puanları ile ilgilidir. Öğrencilerin 3 farklı hazırlık kursundan birine katıldığı varsayılmış ve 3 farklı (Kurs A, Kurs B ve Kurs C) hazırlık kursunun her birinden rasgele ve birbirinden bağımsız olarak 8'er öğrenci seçilmiştir. Bu kursların öğrencilerin aldıkları test puanları üzerinde ne kadar etkili olduğu, 3 farklı hazırlık kursundaki öğrencilerin ortalama başarı puanları arasında anlamlı bir farklılık olup olmadığı araştırılmıştır. Her 3 hazırlık kursundaki öğrencilerin test puanlarının normal dağılımlı ve aynı yığın varyansına (σ^2) sahip olduğu varsayılmıştır. Bu senaryo için varyans analizi aşağıda verilmiştir.

Çizelge 5.1 1. senaryoya ait varyans analizi

Kaynak	<i>KT</i>	<i>sd</i>	<i>KO</i>	<i>F</i>	<i>p –değeri</i>	$\hat{\eta}^2$
Gruplar arası	26352	2	13176	1.80	0.19	0.15
Grup içi	153497	21	7309			
Toplam	179849	23	7820			

Burada $p - \text{değeri} = 0.19$ ($p - \text{değeri} > 0.05$) olduğundan 3 farklı hazırlık kursundan alınan test puanları arasında anlamlı bir fark olmadığı sonucu çıkarılmıştır. Etki büyüklüğü ($\eta^2 = 0.15$) öğrenci test puanları ile hazırlık kursları arasında anlamlı (güçlü) bir ilişki olmadığını ifade eder.

2. senaryoda öğrencilerin üç derse ilişkin memnuniyet puanları verilmiştir. Öğrencilerin üç derse ilişkin memnuniyet puanlarını arasında anlamlı bir fark olup olmadığı, öğrencilerin bazı dersleri diğerlerinden farklı olarak değerlendirip değerlendirmedikleri araştırılmıştır. Örnek a’da gruplar arasındaki memnuniyet ortalamaları birbirlerine çok yakın olacak şekilde, örnek b ve örnek c’de gruplar arasındaki memnuniyet ortalamaları birbirlerinden gittikçe farklılaşacak şekilde veri üretilmiştir. Her bir derste öğrenci sayısı eşit ve $n = 20$ dir. Ortalamaların farklılıklarının vurgulanabilmesi için standart sapma her bir grup için 1.3 olacak şekilde yapay olarak oluşturulmuştur. Etki büyüklüğü, derslere göre memnuniyet puanının ne kadar farklı olduğunu açıklayacaktır.



Şekil 5.1 2. senaryoya ait histogram grafikleri (Tarlow 2016)

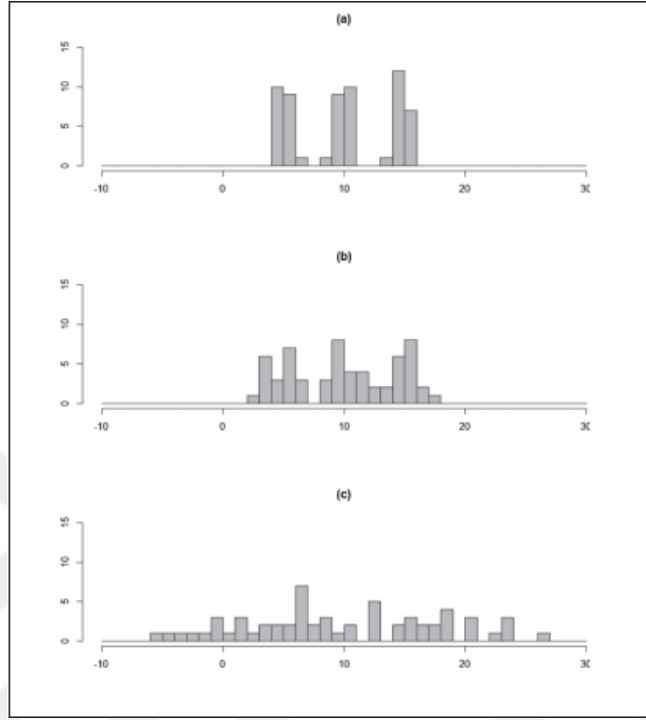
Çizelge 5.2 2. senaryoya ait varyans analizi (Tarlow 2016)

	Kaynak	KT	sd	KO	F	p –değeri	$\hat{\eta}^2$
a	Gruplar arası	73.7	2	36.9	24.31	<0.001	0.46
	Grup içi	86.5	57	1.5			
	Toplam	160.3	59	2.7			
b	Gruplar arası	1010	2	505	341.2	<0.001	0.92
	Grup içi	84.4	57	1.5			
	Toplam	1094.4	59	18.5			
c	Gruplar arası	3954	2	1976.8	1520	<0.001	0.98
	Grup içi	74	57	1.3			
	Toplam	4028	59	68.3			

Şekil 5.1'e bakıldığında a'da ortalamalar birbirine çok yakındır ve ortalamaların farkı küçüktür, bu nedenle etki de küçüktür. b ve c'de ise ortalamalar birbirlerinden çok farklı ve ortalamalar arasındaki fark fazla olduğu için etkinin de büyümesi beklenir.

3. senaryoda üç farklı öğrenci grubunun memnuniyet puanları verilmiştir. Üç farklı öğrenci grubunun memnuniyet puanları arasında anlamlı bir fark olup olmadığı

araştırılmıştır. Her bir öğrenci grubu eşit örnek büyüklüğüne (n=20) sahiptir. Örneklem ortalamaları 5, 10, 15’dir ve aynı varyansa sahiptirler.



Şekil 5.2 3. senaryoya ait histogram grafikleri (Tarlow 2016)

Çizelge 5.3 3.senaryoya ait varyans analizi (Tarlow 2016)

	Kaynak	<i>KT</i>	<i>sd</i>	<i>KO</i>	<i>F</i>	<i>p – değeri</i>	$\hat{\eta}^2$
a	Gruplar arası	984.8	2	492.4	2017	<0.001	0.99
	Grup içi	13.9	57	0.2			
	Toplam	998.7	59	16.9			
b	Gruplar arası	1058.9	2	529.4	361.4	<0.001	0.93
	Grup içi	83.5	57	1.5			
	Toplam	1142.4	59	19.4			
c	Gruplar arası	637	2	318.4	5.6	0.006	0.16
	Grup içi	3261	57	57.2			
	Toplam	3898	59	66.1			

Şekil 5.3'e bakıldığında, a'da ve b'de ortalamalar arasındaki fark büyüktür ve bu nedenle etki büyüktür. c'de ise ortalamalar birbirine çok yakındır ve ortalamaların farkı küçüktür dolayısıyla etki de küçük olacaktır.

Çizelge 5.4 p – değeri ve etki büyüklüğü arasındaki ilişki (Tarlow 2016)

	Küçük p – değeri	Büyük p – değeri
Küçük η^2	En az iki yığın ortalamasının farklı olduğunu ancak bu farklılığın çok fazla olmadığını gösterir.	Yığın ortalamalarında herhangi bir farklılık olmadığını gösterir.
Büyük η^2	En az iki yığın ortalamasının farklı olduğunu ve bu farklılığın fazla olduğunu gösterir.	Örnek ortalamalarının birbirinden çok farklı olmasına rağmen, yığın ortalamalarında herhangi bir farklılık olmadığını gösterir.

5.1 Ortalamalara Göre Etki Büyüklükleri

Analizde yer alan çalışmalardaki istatistikler, ortalamaları ve standart sapmaları verdiğinde ortalamalara göre etki büyüklükleri kullanılır. Ortalamalara göre etki büyüklükleri; genellikle ortalamaların farkı, standartlaştırılmış ortalamaların farkı veya yanıt oranı olacaktır. Aşağıda “D” örneklem ortalamaları arasındaki farkın tahminini, “d” ise standartlaştırılmış örneklem ortalamaları arasındaki farkın tahminini göstermek için kullanılmıştır. Bir meta analizi, bağımsız gruplar arasında farkların yer aldığı çalışmaları ve eşleştirilmiş gruplar arasındaki farkların yer aldığı çalışmaları içerebilir. Etki büyüklüğü d veya Hedges g 'si çalışma tasarımından bağımsız olarak aynı anlama sahiptir. Bu nedenle, uygun formül kullanılarak her bir çalışmadan etki büyüklüğü ve varyans hesaplanarak tüm çalışmalar aynı analize dahil edilebilir (Borenstein vd. 2009).

5.1.1 Ortalamaların farkı

Ortalamaların farkı, meta analizine dahil edilecek tüm çalışmaların aynı ölçeği (örneğin sosyal bilimlerde kullanılan Beck Depresyon Ölçeği) kullanması durumunda aynı ölçüm

birimi kullanılmış olduğundan ölçüm birimi etkisi arındırılmadan etki büyüklüğü olarak kullanılabilir.

5.1.1.1 Bağımsız gruplarla yapılan çalışmalar

İki bağımsız grup ortalamalarının karşılaştırılmak istendiği varsayalım. İki grubun (örneğin tedavi ve kontrol grupları) yığın ortalamaları μ_1 ve μ_2 olmak üzere yığın ortalamaları farkı,

$$\Delta = \mu_1 - \mu_2 \quad (5.2)$$

dir.

Bu bağımsız gruplardan alınan n_1 ve n_2 çaplı rasgele örneklemlemlerle yapılan bir çalışmada, $\bar{X}_1$ ve $\bar{X}_2$ iki gruba ait örneklem ortalamaları olmak üzere Δ 'nın tahmin edicisi,

$$D = \bar{X}_1 - \bar{X}_2 \quad (5.3)$$

dir.

Eğer iki yığının standart sapmalarının aynı olduğu varsayılırsa ($\sigma_1 = \sigma_2 = \sigma$),

D 'nin varyansının tahmin edicisi,

$$S^2_D = \frac{n_1 + n_2}{n_1 n_2} S^2_{pooled} \quad (5.4)$$

dir, burada S_1 ve S_2 iki gruba ait örneklem standart sapmaları olmak üzere S^2_{pooled} iki gruba ait ortak (veya havuzlandırılmış) varyans tahmin edicisi,

$$S^2_{pooled} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad (5.5)$$

dir.

Eğer iki yığının standart sapmalarının aynı olmadığı varsayılırsa ($\sigma_1 \neq \sigma_2$),

D 'nin varyansının tahmin edicisi,

$$S^2_D = \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \quad (5.6)$$

biçimindedir.

5.1.1.2 Bağımlı (Eşleştirilmiş) gruplar ya da ön test- son test kullanılan çalışmalar

Eşleştirme, bir çalışmadaki deneklerin veya gözlemlerin bir ölçüte göre eşleştirildiği (örneğin, kardeşler veya hastalığın aynı evresindeki hastalar gibi) ve her bir eşin iki üyesinin daha sonra farklı gruplara atanması işlemidir. Böyle bir çalışmada gözlem birimi çifttir ve bu tasarımın avantajı, her bir çiftin kendi kontrolü altında olması, hata terimini azaltması ve istatistiksel gücü artırmasıdır. Etkinin büyüklüğü, çiftler (örneğin kardeşler) arasındaki korelasyona bağlıdır ve daha yüksek bir korelasyon, daha düşük bir varyans (artan duyarlılık) sağlar. D 'nin örnek tahmini sadece örnek ortalamalarının farkıdır.

Ortalamaların farkı $\bar{X}_{fark}$ ve bu farklılıkların standart sapması S_{fark} ile gösterilir.

Ortalamaların farkının tahmin edicisi,

$$D = \bar{X}_{fark}$$

$$\bar{X}_{fark} = \bar{X}_1 - \bar{X}_2 \quad (5.7)$$

dir.

D 'nin varyansının tahmin edicisi,

$$S^2_D = \frac{S^2_{fark}}{n} \quad (5.8)$$

dir. Burada n çiftlerin sayısıdır.

S_{fark} ,

$$S_{fark} = \sqrt{S_1^2 + S_2^2 - 2 \times r \times S_1 \times S_2}$$

$$S_{fark} = \sqrt{2 \times S_{pooled}^2 (1 - r)} \quad (5.9)$$

biçimindedir, burada r eşleşen çiftler (örneğin kardeşler) arasındaki korelasyondur.

Eşleşen tasarımlar için formüller, ön test- son test kullanılan çalışmalar için de geçerlidir. Ön test- son test kullanılan çalışmalardaki ortalamalar, eşleşen gruptaki ortalamalara karşılık gelir. Ön test-son test kullanılan çalışmalarda n , denek sayısını ve r , ön test- son test arasındaki korelasyonu ifade eder (Borenstein vd. 2009).

5.1.2 Standartlaştırılmış ortalamaların farkı

Ortalamaların farkı, meta analizine dahil edilen çalışmalar aynı ölçekleri kullanmış olduğunda yararlı bir indekstir. Meta analizine dahil edilecek çalışmalar farklı ölçekler kullanıyorsa, ölçüm ölçeği çalışmadan çalışmaya farklılık gösterecektir. Bu yüzden çalışmaların karşılaştırılabilir olması için standartlaştırılmış ortalamaların farkı kullanılmalıdır. Standartlaştırılmış ortalamaların farkı (d veya g), tüm etki büyüklüklerini ortak bir metriğe dönüştürür ve böylece farklı sonuç ölçütlerinin analize dahil edilmesini sağlar. Standartlaştırılmış ortalamaların farkı her çalışmadaki ortalamaların farkının o çalışmanın standart sapmasına bölünmesiyle elde edilir (Borenstein vd. 2009). Standartlaştırılmış ortalamaların farkı, iki farklı ölçeğe ya da ölçüm birimine dayanan çalışmaların karşılaştırılabilir olmasını sağlar (Hedges ve Olkin 1985).

Birbirinden bağımsız iki farklı grup ortalamalarının karşılaştırılmak istendiği varsayalım. Eğer iki grubun yığın standart sapması eşitse $\sigma_1 = \sigma_2 = \sigma$ standartlaştırılmış ortalamaların fark parametresi,

$$\delta = \frac{\mu_1 - \mu_2}{\sigma} \quad (5.10)$$

biçimindedir.

5.1.2.1 Bağımsız gruplarla yapılan çalışmalar

Birbirinden bağımsız iki farklı grubun standartlaştırılmış ortalamaların farkının tahmin edicisi,

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{pooled}} \quad (5.11)$$

dir, burada $\bar{X}_1$ ve $\bar{X}_2$ grupların örnek ortalamaları, S_{pooled} ise iki gruba ait ortak-havuzlandırılmış standart sapmadır ve

$$S_{pooled} = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \quad (5.12)$$

biçiminde ifade edilir. Burada n_1 ve n_2 iki gruba ait örneklem büyüklüğü, S_1 ve S_2 iki gruba ait örneklem standart sapmasıdır. Standartlaştırılmış ortalamaların farkına ait örnek tahmini, genellikle Cohen's d olarak adlandırılır.

Standartlaştırılmış ortalamaların fark parametresi " δ " ile bu parametrenin tahmin edicisi ise " d " ile gösterilecektir.

d 'nin varyansının tahmin edicisi,

$$S_d^2 = \frac{n_1+n_2}{n_1n_2} + \frac{d^2}{2(n_1+n_2)} \quad (5.13)$$

dir.

" d " tahmini örneklem büyüklüğü küçük olduğunda yanlılığa sahip olabilir. Bu yanlılık basit bir düzeltme yapılarak kaldırılabilir. Bu dönüşümü yapabilmek için J. Hedges (1981) olarak adlandırılan bir düzeltme faktörü kullanılır.

Düzeltilme faktörü (J),

$$J = 1 - \frac{3}{4sd-1} \quad (5.14)$$

olup, $sd = n_1 + n_2 - 2$ 'dir.

Hedges g tahmini,

$$g = J \times d \quad (5.15)$$

ile ifade edilir.

Hedges g istatistiğinin varyansının tahmin edicisi,

$$S^2_g = J^2 \times S^2_d \quad (5.16)$$

biçimindedir. Düzeltme faktörü (J) her zaman 1'den küçüktür ve bu nedenle “ g ” mutlak değerinde her zaman d 'den küçük olacaktır dolayısıyla g 'nin varyansı her zaman d 'nin varyansından küçük olacaktır. Bununla birlikte, sd çok küçük olmadığı sürece (örneğin 10'dan küçük) yani fark önemsiz olmadığı sürece J 1'e çok yakın bir değer olacaktır (Hedges 1981).

Çizelge 5.5 Cohen's d ve Hedges g etki büyüklüklerinin sınıflandırması (Cohen 1988)

Etki Büyüklüğünün Yorumu	Etki Büyüklüğü
Küçük	0.2
Orta	0.5
Büyük	0.8

5.1.2.2 Bağımlı (Eşleştirilmiş) gruplar ya da ön test- son test kullanılan çalışmalar

Eşleşen grupları veya ön test- son test kullanan çalışmalardan standartlaştırılmış ortalamaların farkı (δ) tahmin edilebilir.

δ 'nin tahmin edicisi,

$$d = \frac{\bar{X}_{fark}}{S_{grupiçi}} = \frac{\bar{X}_1 - \bar{X}_2}{S_{grupiçi}} \quad (5.17)$$

biçimindedir. Eşleşen gruplarda grup içi standart sapmanın tahmin edicisi,

$$S_{grupiçi} = \frac{S_{fark}}{\sqrt{2(1-r)}} \quad (5.18)$$

şeklinde hesaplanır. Burada r , gözlem çiftleri arasındaki korelasyondur (örneğin, ön test-son test arasındaki korelasyon).

d 'nin varyansının tahmin edicisi,

$$S^2_d = \left(\frac{1}{n} + \frac{d^2}{2n} \right) 2(1-r) \quad (5.19)$$

biçimindedir.

Grup içindeki standart sapmayı, farkın standart sapmasından hesaplayabilmek için ön test-son test arasındaki korelasyon gerekir bu yüzden korelasyonun bilindiği veya yüksek hassasiyetle tahmin edilebileceği varsayılır (Borenstein vd. 2009).

5.1.3 Yanıt oranları

Sonucun fiziksel bir ölçekte (uzunluk, alan veya kütle gibi) ölçüldüğü ve sıfır olma ihtimalinin düşük olduğu araştırma alanlarında, iki gruptaki ortalamaların oranı etki büyüklüğü indeksi olabilir. Deneysel ekolojide bu etki büyüklüğü indeksi yanıt oranı olarak adlandırılır (Hedges vd. 1999). Yanıt oranı, test puanları veya tutum ölçekleri gibi sonuç ölçen çalışmalar (örneğin sosyal bilim çalışması) için anlamlı değildir çünkü mutlak sıfır noktaları yoktur. Yanıt oranı yalnızca ölçüm düzeyi oran ölçeği kullanıldığında anlamlı olur. Yanıt oranı (R) ekolojide sıklıkla kullanılır.

Yanıt oranları için hesaplamalar bir log-ölçeğinde gerçekleştirilir. Yanıt oranının logaritması alınır ve meta analizindeki tüm adımları gerçekleştirmek için kullanılır. Ancak o zaman sonuçlar orijinal metriğe geri dönüştürülebilir.

Yanıt oranı,

$$R = \frac{\bar{X}_1}{\bar{X}_2} \quad (5.20)$$

biçimindedir, burada $\bar{X}_1$ ve $\bar{X}_2$ grup ortalamalarıdır.

Yanıt oranının logaritması ve yanıt oranının varyansının tahmin edicisi,

$$\ln R = \ln(R) = \ln\left(\frac{\bar{X}_1}{\bar{X}_2}\right) = \ln(\bar{X}_1) - \ln(\bar{X}_2) \quad (5.21)$$

$$S^2_{\ln R} = S^2_{pooled} \left(\frac{1}{n_1(\bar{X}_1)^2} + \frac{1}{n_2(\bar{X}_2)^2} \right) \quad (5.22)$$

biçimindedir, burada S_{pooled} iki gruba ait ortak-havuzlandırılmış standart sapmadır.

Yanıt oranı için bir varyans hesaplanmaz. Bunun yerine, özet etki, güven aralıkları vb. hesaplamaların yapılabilmesi için logaritması alınmış yanıt oranı ve bunun varyansı kullanılır. Daha sonra, *LL* ve *UL*'nin sırasıyla alt ve üst sınırları temsil ettiği yerler kullanılarak bu değerlerin her biri tekrar yanıt oranlarına dönüştürülebilir.

5.2 Kategorik Verilerde Etki Büyüklüğü

Kategorik verilerde etki büyüklüğü risk oranı, odds oranı veya risk farkı ile hesaplanır. Risk oranı ve odds oranı göreceli ölçülerdir. Bu yüzden farklılıklara karşı risk farkına göre daha duyarsız olabilirler. Risk oranı veya odds oranı ile çalışmak için tüm değerler logaritmik değerlere dönüştürülür ve analizleri gerçekleştirilir. Ardından sonuçları rapor edebilmek için tekrar oran değerlerine dönüştürülür. Risk farkı için böyle bir dönüşüme gerek yoktur.

Risk oranı, Odds oranı ve Risk farkı terimlerinin daha iyi anlaşılabilmesi ve nasıl hesaplandığının gösterilmesi için aşağıdaki çizelgeden yararlanılacaktır.

Çizelge 5.6 Sigara içme durumu ile gırtlak kanseri arasındaki ilişkiyi gösteren kontenjans tablosu

	Gırtlak Kanseri		Toplam
	Hastalığa yakalananlar	Hastalığa yakalanmayanlar	
Sigara içen	A	B	A+B
Sigara içmeyen	C	D	C+D
Toplam	A+C	B+D	

5.2.1 Risk oranı

Risk oranı basitçe iki riskin oranıdır. Örneğin, sigara içenlerin gırtlak kanserine yakalanma olasılığının, sigara içmeyenlerin gırtlak kanserine yakalanma olasılığına

bölünmesidir. Risk oranları için hesaplamalar logartima ölçeğinde yapılır. Log-risk oranı ve log-risk oranının standart hatası hesaplanır ve meta analizindeki tüm adımları gerçekleştirmek için bu değerler kullanılır. Her grupta bir olayın riski (Gırtlak kanserine yakalanma) hesaplanabilir (Sigara içen ve içmeyene göre).

Risk oranı (risk ratio),

$$Risk\ Oranı = \frac{A/(A+B)}{C/(C+D)} \quad (5.23)$$

biçimindedir.

Log-risk oranı, Log-risk oranının varyansı ve Log-risk oranının standart hatası,

$$LogRiskOranı = \ln(RiskOranı) \quad (5.24)$$

$$Var(LogRiskOranı) = \frac{1}{A} - \frac{1}{(A+B)} + \frac{1}{C} - \frac{1}{(C+D)} \quad (5.25)$$

$$SE(LogRiskOranı) = \sqrt{Var(LogRiskOranı)} \quad (5.26)$$

dır.

Orijinal metrikte risk oranı için bir varyans hesaplamak yerine, risk oranının logaritması alınarak varyans hesaplanır. Daha sonra bu değerler risk oranlarına dönüştürülür.

$$RiskOranı = \exp(LogRiskOranı) \quad (5.27)$$

$$LL_{RiskOranı} = \exp(LL_{LogRiskOranı}) \quad (5.28)$$

$$UL_{RiskOranı} = \exp(UL_{LogRiskOranı}) \quad (5.29)$$

burada LL ve UL , sırasıyla güven aralığının alt ve üst sınırlarını temsil eder (Borenstein vd. 2009).

5.2.2 Odds oranı

Olasılık oranı olarak da kullanılır. Odds oranı; hastalık ve etki arasındaki ilişkinin ölçüsüdür. Sigara içen kişilerin; gırtlak kanserine yakalanma olasılığının, yakalanmama olasılığına oranı ile sigara içmeyen kişilerin; gırtlak kanserine yakalanma olasılığının,

yakalanmama olasılığına bölünmesi ile elde edilir. Odds oranları için hesaplamalar bir logaritma ölçeğinde yapılır. Log- odds oranı ve log- odds oranının standart hatası hesaplanır.

Odds oranı,

$$OddsOranı = \frac{A \times D}{B \times C} \quad (5.30)$$

biçimindedir.

Log-Odds oranı, Log-Odds oranının varyansı ve Log-Odds oranının standart hatası,

$$LogOddsOranı = \ln(OddsOranı) \quad (5.31)$$

$$Var(LogOddsOranı) = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D} \quad (5.32)$$

$$SE(LogOddsOranı) = \sqrt{Var(LogOddsOranı)} \quad (5.33)$$

biçimindedir.

$$OddsOranı = \exp(LogOddsOranı) \quad (5.34)$$

$$LL_{OddsOranı} = \exp(LL_{LogOddsOranı}) \quad (5.35)$$

$$UL_{OddsOranı} = \exp(UL_{LogOddsOranı}) \quad (5.36)$$

burada *LL* ve *UL* sırasıyla güven aralığının alt ve üst sınırlarını temsil eder (Borenstein vd. 2009).

5.2.3 Risk farkı

Risk farkı, iki risk arasındaki farktır. Risk oranları ve olasılık oranlarının aksine, risk farkları hesaplamaları log ölçeğinde yapılmaz. Her grupta bir olayın riski (örneğin gırtlak kanserine yakalanma) hesaplanabilir (örneğin, sigara içen ve sigara içmeyene göre). Bu risklerdeki fark daha sonra bir etki büyüklüğü (risk farkı) olarak kullanılır.

Risk farkı,

$$RiskFarkı = \left(\frac{A}{A+B} \right) - \left(\frac{C}{C+D} \right) \quad (5.37)$$

biçimindedir.

Risk farkının varyansı,

$$Var(RiskFarkı) = \frac{AB}{(A+B)^3} + \frac{CD}{(C+D)^3} \quad (5.38)$$

Risk farkının standart hatası,

$$SE(RiskFarkı) = \sqrt{Var(RiskFarkı)} \quad (5.39)$$

biçimindedir (Borenstein vd. 2009).

5.3 Korelasyonlara Dayalı Etki Büyüklükleri

Yayımlanan istatistiklerinde iki sürekli değişken arasındaki korelasyon katsayısını rapor eden çalışmalar için, korelasyon katsayısı etki büyüklüğü indeksi olarak kullanılabilir. Fisher z dönüşümü kullanılarak korelasyonlar dönüştürülür ve bu indeksi kullanarak analizi yapılır. Ardından, özet değerler rapor için tekrar korelasyonlara dönüştürülür.

Korelasyon, θ gibi orijinal ölçeklerdeki farklı metrikleri hesaba katmak için standartlaştırılmış sezgisel bir ölçüdür. Korelasyon katsayısının yığın parametresi “ ρ ” ile parametrenin tahmini ise “ r ” ile gösterilir.

r 'nin varyansının tahmin edicisi,

$$S^2_r = \frac{(1-r^2)^2}{n-1} \quad (5.40)$$

biçimindedir ve n örneklem büyüklüğüdür. Varyans korelasyona bağlı olduğundan, doğrudan korelasyon katsayısı üzerinde işlem yapılmaz. Bunun yerine, korelasyon katsayısı Fisher z ölçeğine dönüştürülür ve tüm analizler dönüştürülmüş değerler kullanılarak gerçekleştirilir. Özet etki ve güven aralığı gibi sonuçlar daha sonra rapor için tekrar korelasyonlara dönüştürülür.

r 'den Fisher's z 'ye dönüşüm şu şekilde verilir,

$$z = 0.5 \times \ln \left(\frac{1+r}{1-r} \right) \quad (5.41)$$

z 'nin varyansının tahmin edicisi,

$$S^2_z = \frac{1}{n-3} \quad (5.42)$$

z 'nin standart hata tahmini,

$$SE(z) = \sqrt{S^2_z} \quad (5.43)$$

biçimindedir.

Fisher's z dönüşümü ile çalışırken, korelasyon için varyans kullanılmaz. Bunun yerine, dönüştürülmüş değerin varyansı kullanılır. Daha sonra bu değerlerin her biri tekrar korelasyon birimlerine dönüştürülür (Borenstein vd. 2009).

$$r = \frac{e^{2z}-1}{e^{2z}+1} \quad (5.44)$$

5.4 Etki Büyüklükleri Arasında Dönüştürme

Meta analizine dahil edilecek çalışmalar farklı ölçüm düzeyine sahip olabilir, bu da çalışmalarda farklı etki büyüklüğü indekslerinin kullanılabileceğini ifade eder. Örneğin; çalışmalardan bazıları standartlaştırılmış ortalamaların farkı, bazıları odds oranı, bazıları ise korelasyon katsayısını etki büyüklüğü indeksi olarak kullanabilir. Buradaki önemli nokta karşılaştırılabilir olduğuna karar verilen çalışmalardan elde edilen özet etkiyi hesaplamamanın mantıklı olmasıdır. Farklı indeksler olduğu için öncelikle bunların ortak bir metriğe dönüştürülmesi gerekir (Borenstein vd. 2009).

5.4.1 Log-Odds oranından d 'ye dönüştürme

Bir log-odds oranı (Log-OddsRatio) standartlaştırılmış ortalamaların farkı d 'ye dönüştürülebilir.

Bu dönüşüm,

$$d = \text{LogOddsOranu} \times \frac{\sqrt{3}}{\pi} \quad (5.45)$$

biçimindedir, burada π matematiksel sabittir (yaklaşık 3.14159).

d 'nin varyansı,

$$\text{Var}(d) = \text{Var}(\text{LogOddsOranu}) \times \frac{3}{\pi^2} \quad (5.46)$$

biçimindedir, burada $\text{Var}(\text{LogOddsOranu})$ log- odds oranının varyansıdır (Borenstein vd. 2009).

5.4.2 d 'den Log- Odds oranına dönüştürme

Standartlaştırılmış ortalamaların farkı d , log- odds oranına (Log-OddsRatio) dönüştürülebilir.

Bu dönüşüm,

$$\text{LogOddsOranu} = d \frac{\pi}{\sqrt{3}} \quad (5.47)$$

Biçimindedir, burada π matematiksel sabittir (yaklaşık 3.14159).

Varyansı,

$$\text{Var}(\text{LogOddsOranu}) = \text{Var}(d) \frac{\pi^2}{3} \quad (5.48)$$

biçimindedir. Bu dönüşümü kullanmak için sürekli verilerin lojistik dağılıma sahip olduğu varsayılır.

5.4.3 r 'den d 'ye dönüştürme

Korelasyonu (r) standartlaştırılmış bir ortalamaların farkına (d) dönüştürme,

Bu dönüşüm,

$$d = \frac{2r}{\sqrt{1-r^2}} \quad (5.49)$$

biçimindedir.

Bu şekilde hesaplanan (r 'den dönüştürülen) d 'nin varyans tahmin edicisi,

$$S^2_d = \frac{4S^2_r}{(1-r^2)^3} \quad (5.50)$$

biçimindedir.

5.4.4 d 'den r 'ye dönüştürme

Standartlaştırılmış bir ortalamaların farkı (d) korelasyona (r) dönüştürme,

$$r = \frac{d}{d^2+a} \quad (5.51)$$

biçimindedir, burada a , $n_1 \neq n_2$ olduğu durumlar için bir düzeltme faktörüdür,

$$a = \frac{(n_1+n_2)^2}{n_1n_2} \quad (5.52)$$

Düzeltilme faktörü (a), bu sayıların mutlak değerlerinden ziyade n_1 'in n_2 'ye oranına bağlıdır. Bu nedenle, eğer n_1 ve n_2 tam olarak bilinmiyorsa, $n_1 = n_2$ kullanılır, bu da “ a ” değerinin 4’e eşit olmasına neden olur.

Bu şekilde hesaplanan (d 'den dönüştürülmüş) r 'nin varyans tahmin edicisi,

$$S^2_r = \frac{a^2 S^2_d}{(d^2+a)^3} \quad (5.53)$$

biçimindedir.

Bu dönüşümü uygularken, tedavi ve kontrol gruplarını oluşturmak için sürekli bir değişkenin ikiye ayrıldığı varsayılır. Fisher's z ve d arasında dönüşüm yapıldığında, yalnızca bağımsız değişken hakkında varsayımlarda bulunulur. Log-odds oranı ve d arasında dönüşüm yapıldığında yalnızca bağımlı değişken hakkında varsayımlarda bulunulur. Bu nedenle, iki varsayım kümesi birbirinden bağımsızdır ve birinin diğerinin geçerliliği için herhangi bir çıkarımı yoktur. Bu nedenle, Fisher's z 'den d 'ye kadar her iki varsayım da uygulanabilir ve bu log-odds oranına ve bunun tersine çevrilebilir (Borenstein vd. 2009).

6. HETEROJENLİK

Çalışmalar; ele alınan model, çalışmanın yürütülmesindeki farklılıklar, katılımcılar, gözlem birimleri, örneklem büyüklükleri vb. etkenler nedeniyle değişkenlik gösterebilirler. Heterojenlik klinik, metodolojik ve istatistiksel olarak üç kategoriye ayrılır. Katılımcıların yaşı, cinsiyeti, uygulanan tedavi yöntemleri gibi farklılıklara klinik heterojenlik denir. Çalışmanın kalitesi, çalışmanın süresi gibi deneme düzeni farklılıklarına metodolojik heterojenlik denir. Gerçek etkilerin çalışmalar arasında gösterdiği farklılıklara ise istatistiksel heterojenlik denir (Huedo-Medina vd. 2006). İstatistiksel heterojenlik, değerlendirilen gerçek etkiler çalışmalar arasında farklılık gösterdiğinde ortaya çıkar ve çalışmaların sonuçları arasındaki farklılık beklenenin üzerindeyse saptanabilir. İstatistiksel heterojenliği ele almak, birçok sistematik incelemenin en zahmetli yönlerinden biridir(Higgins ve Thompson 2002).

Meta analizinde kullanılan çalışmalardaki istatistiksel heterojenliği açıklayan iki değişkenlik kaynağı olabilir. Bunlardan biri, çalışma içi değişkenlik olarak da adlandırılan örnekleme hatasından kaynaklanan değişkenliktir. Örnekleme hatasından kaynaklanan değişkenlik meta analizinde her zaman mevcuttur, çünkü her çalışma farklı örneklem kullanır. Heterojenliğin diğer kaynağı ise, her bir çalışmadan tahmin edilen etki büyüklükleri arasında değişkenlik olduğunda, meta analizinde ortaya çıkabilen çalışmalar arası değişkenliktir. Çalışmalar arası değişkenlik, örnek özellikleri, işlem farklılıkları vb. gibi çalışmalar arasında farklılık gösteren birçok özelliğin etkisinden kaynaklanmaktadır (Hunter ve Schmidt 2000).

Heterojenliğin varlığı çalışmayı yapacak kişinin uygulamaya karar vereceği istatistiksel modeli etkileyeceğinden sonuçların birleştirilmesinden önce heterojenliği değerlendirmek çok önemlidir. Bu nedenle, çalışmaların sonuçları yalnızca örnekleme hatasına göre değişkenlik gösterdiğinde ortak etki elde edebilmek için sabit etkili model kullanılır. Çalışma sonuçları örnekleme hatasından daha fazla değişkenlik içerdiğinde hem çalışmalar içi hem de çalışmalar arası varyansı hesaba katmak için rasgele etkiler modeli kullanılır.

Bir meta analizinde gerçek heterojenliği (istatistiksel heterojenliği) ölçebilmek için çalışmalar arası varyansı tahmin etmek gerekir. Rasgele etkiler modeli kullanıldığında, çalışmalar arası varyans, meta analizinin her bir çalışma için tahmin edilen gerçek etki büyüklüklerinin ne kadar farklı olduğunu gösterir. Çalışmalar arası varyansın yaygın olarak kullanılan tahmin edicisi, DerSimonian ve Laird (1986) tarafından önerilen momentler yöntemine dayanan tahmin edicidir.

Çalışmalar arası varyans parametresi τ^2 ile, tahmin edicisi ise T^2 ile gösterilir.

T^2 ,

$$T^2 = \begin{cases} \frac{Q-(k-1)}{c} & Q > k - 1 \\ 0 & Q \leq k - 1 \end{cases} \quad (6.1)$$

$$c = \sum_{i=1}^k W_i - \frac{\sum_{i=1}^k W_i^2}{\sum_{i=1}^k W_i} \quad (6.2)$$

biçimindedir. Burada w_i , sabit etkili modelini ($w_i = \frac{1}{\text{Var}(Y_i)}$) varsayan i. çalışma için ağırlıklandırma faktörüdür, k çalışma sayısıdır ve Q , Cochran (1954) tarafından önerilen heterojenlik testidir.

Bir meta analizinde gerçek heterojenliğin (istatistiksel heterojenliğin) olup olmadığının belirlenebilmesi için, Cochran (1954) tarafından önerilen Q testi kullanılır. Q , $(k - 1)$ serbestlik derecesi ile ki-kare dağılır. Burada k çalışma sayısıdır. H_0 yokluk hipotezi “Etki büyüklükleri homojendir” şeklinde kurulur. H_0 yokluk hipotezinin reddedilmesi genellikle meta analizi uygulayacak kişinin sabit etkili modeli kullanmasına yol açar çünkü tahmin edilen etki büyüklüklerinin yalnızca örnekleme hatasıyla farklılaştığı varsayılır. Buna karşılık, homojenlik varsayımını reddetmek hem çalışmalar içi hem de çalışmalar arası değişkenliği içeren bir rasgele etkiler modelinin uygulanmasına yol açabilir (Huedo-Medina vd. 2006).

6.1 Q Testi

Meta analizine dahil edilen çalışmaların etkileri arasında heterojenliğin olup olmadığını belirlemek için kullanılır.

Q istatistiği,

$$Q = \sum_{i=1}^k W_i Y_i^2 - \frac{(\sum_{i=1}^k W_i Y_i)^2}{\sum_{i=1}^k W_i} \quad (6.3)$$

biçiminde tanımlanır, burada w_i , sabit etkili modeli ($w_i = \frac{1}{\text{var}(Y_i)}$) varsayan i . çalışma için ağırlıklandırma faktörüdür.

Heterojenlik testi için hipotezler,

$$H_0: \delta_1 = \delta_2 = \dots = \delta_k$$

$H_1: \text{En az biri farklıdır.}$

şeklinde kurulur. Çalışma içi varyansların, (σ^2 'nin) bilindiği varsayılırsa, Q istatistiği yokluk hipotezi altında $(k - 1)$ serbestlik dereceli ki-kare dağılır. Bu nedenle, belirli bir anlamlılık düzeyi (α) için kritik noktadan daha büyük Q değerleri, yokluk hipotezinin reddedilmesine ve çalışmalar arasında istatistiksel olarak anlamlı bir değişkenlik olduğu sonucuna varılmasını sağlar.

Q istatistiği, bir toplam şeklindedir ve meta analizine dahil edilen çalışma sayısına bağlıdır. Meta analizi az sayıda çalışma içerdiğinde çalışmalar arasındaki gerçek heterojenliği saptama gücü zayıftır. Bu nedenle, az sayıda çalışma içeren meta analizi için Q testi anlamlı olmayan (H_0 hipotezi reddedilemez) bir sonuç verdiğinde ve çalışmalar arasında gerçek bir heterojenlik olduğunda çalışmayı yapan kişinin hatalı bir şekilde sabit etkili modeli kullanmasına neden olabilir. Bu yüzden anlamlılık düzeyi (α) olarak 0.05 yerine 0.10 kullanılması önerilmiştir.

Q testi yalnızca heterojenliğin varlığı hakkında bilgi verir, heterojenliğin kapsamı ve boyutu hakkında bilgi vermez. Bir meta analizinde kullanılan etki büyüklüğü indeksleri verinin ölçüm düzeyine bağlı olduğundan, Q testi ile farklı etki büyüklüğü indeksleri (standartlaştırılmış ortalamaların farkları, korelasyon katsayıları, olasılık oranları) kullanan çalışmalardan tahmin edilen iki değeri karşılaştırmak mümkün değildir. Heterojenlik testi çalışma sayısından etkilenmemeli ve farklı etki büyüklüğü indeksleri uygulayan çalışmalar arasında karşılaştırılabilir olması için ölçeğe bağımsız olmalıdır. Q testinin eksikliklerini gidermek için Higgins ve Thompson meta analizinde heterojenliği değerlendirmede H^2 , R^2 ve I^2 indekslerini önermiştir. Üç indekste birbiriyle ilişkilidir. Burada yalnızca I^2 indeksine yer verilmiştir (Huedo-Medina vd. 2006).

6.2 I^2 İndeksi

I^2 indeksi, Q değerini serbestlik derecesiyle ($sd = k - 1$) karşılaştırarak heterojenliğin derecesini ölçer.

I^2 indeksi,

$$I^2 = \begin{cases} \frac{Q - (k-1)}{Q} \times 100\% & Q > (k - 1) \\ 0 & Q \leq (k - 1) \end{cases} \quad (6.4)$$

olarak tanımlanır. I^2 indeksi, heterojenlikten yani çalışmalar arası değişkenlikten dolayı her bir çalışmanın etki büyüklüğündeki toplam değişkenliğin yüzdesi olarak tanımlanabilir. Örneğin, $I^2 = 0$ olan bir meta analizi çalışması, tüm değişkenliğin yalnızca çalışmalar içindeki örnekleme hatasından kaynaklandığı, $I^2 = 50$ olan bir meta analizi çalışmasında, etki büyüklükleri arasındaki toplam değişkenliğin yarısının çalışmalar arasındaki değişkenlikten kaynaklandığı anlamına gelmektedir. Yaklaşık %25 ($I^2 = 25$), %50 ($I^2 = 50$) ve % 75 ($I^2 = 75$) yüzdeleri sırasıyla düşük, orta ve yüksek heterojenlik anlamına gelmektedir (Higgins vd. 2003).

I^2 indeksi ile çalışmalar arası varyans, τ^2 , doğrudan ilişkilidir: τ^2 ne kadar yüksekse, I^2 indeksi de o kadar yüksek olur.

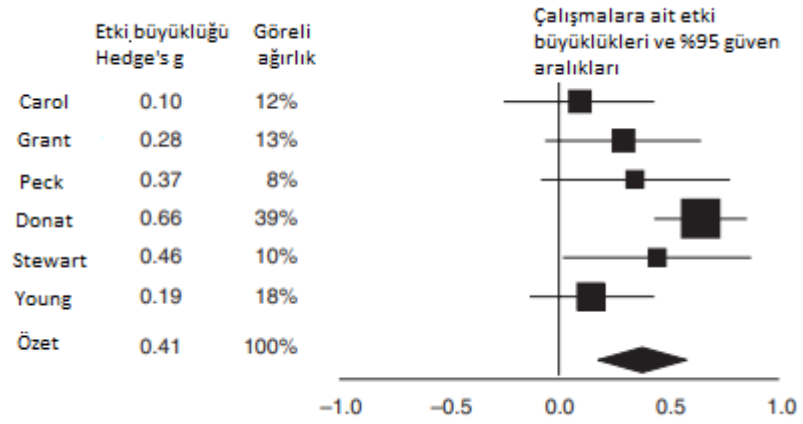
$$I^2 = \frac{c\tau^2}{Q} \quad (6.5)$$

Farklı sayıda çalışma ve farklı etki büyüklüğü indeksleri ile meta analizlerden elde edilen I^2 indeksleri doğrudan karşılaştırılabilir.

Higgins ve Thompson (2002)'a göre, I^2 indeksinin güven aralığı 0 değerini içeriyorsa, çalışmayı yapan kişi her bir çalışmanın homojen olduğunu varsayabilir. Güven aralığı 0 değerini içermiyorsa, gerçek heterojenliğin varlığına dair kanıt olabilir. I^2 indeksi yalnızca meta analizindeki heterojenliği değil, aynı zamanda bu heterojenliğin derecesini de değerlendirdiğinden, meta analizinde çalışmalar arasında gerçek heterojenliğin olup olmadığını değerlendirmede Q testinden daha fazla önerilen bir yöntem olmalıdır (Huedo-Medina vd. 2006).

6.3 Forest Grafiği

Heterojenliğin belirlenebilmesi için heterojenlik ölçütlerine ek olarak bazı grafiksel metotlardan da yararlanılır. Forest grafiği meta analizi diyagramı olarak adlandırılır ve meta analizinde sıklıkla kullanılır. Grafikte her bir çalışma için gözlenen etki büyüklüğü kare, özet etki büyüklüğü ise karo ile temsil edilir. Her bir çalışma için karenin boyutu çalışmanın varyansı ile ters orantılıdır. Kare ne kadar büyük olursa çalışma ağırlığı da o kadar büyüktür. Her çalışmaya ait etki büyüklüğü bir güven aralığına sahiptir ve karenin ortasından geçen çizgi %95 güven aralığını temsil etmektedir. Düşey olarak geçen çizgi etkisizlik çizgisidir (no effect line). Etkisizlik çizgisi, logaritma kullanılarak hesaplanan etki büyüklüklerinde 1, logaritma kullanılmadan hesaplanan etki büyüklüklerinde 0 olarak kabul edilir (Bakıoğlu ve Özcan 2016).



Şekil 6.1 Hedges g istatistiğinin etki büyüklüğü olarak kullanıldığı meta analizi çalışmasına ait forest grafiği (Borenstein vd. 2009)

Forest grafiğine bakıldığında etki büyüklüğü olarak Hedges g kullanıldığı için etkisizlik çizgisi 0 olarak kabul edilmiştir. En büyük etki büyüklüğüne sahip olan çalışma Donat'a, en geniş güven aralığına sahip olan çalışma Peck' e aittir. Donat'ın ve Young' un çalışmalarına ait etki büyüklüklerini temsil eden kareler diğerlerine göre daha büyük olduğundan çalışma ağırlıkları daha fazladır.

7. META ANALİZİNDE İSTATİSTİKSEL MODEL SEÇİMİ

Meta analizinde analize dahil edilecek çalışmalar belirlendikten sonra, sonuçların birleştirilebilmesi için bu çalışmalara uygun istatistiksel model ve yöntemlerin seçilmesi gerekir. Meta analizi genellikle sabit etkili modele veya rasgele etkiler modeline dayanır. Sabit etkili model altında, tüm çalışmaların gerçek bir etki büyüklüğüne (sabit etki terimi) sahip olduğu ve gözlenen etkilerdeki tüm farklılıkların örneklem hatasından kaynaklandığı varsayılır. Rasgele etkiler modeli altında, gerçek etkinin çalışmadan çalışmaya değişebileceği varsayılır. Örneğin; katılımcıların yaşı, eğitim durumu, boy uzunluğu vb. değişkenler nedeniyle çalışmalarda etki büyüklüğü daha yüksek (veya daha düşük) olabilir. Çalışmalar, katılımcılar ve müdahalelerin uygulanmasında farklılık gösterebileceğinden farklı çalışmalar için farklı etki büyüklükleri olabilir.

Sabit etki, gözlem birimlerinin işleme tabi tutulan etki ya da etkilerin araştırmacı tarafından planlı olarak önceden belirlendiği, yani etkinin rasgele belirlenmemiş olduğunu ifade eder. Örneğin; sigara markalarına göre nikotin miktarı konusunda araştırma yapan bir araştırmacı üzerinde gözlem yapacağı sigaraları markalarını özel olarak merak edip ona göre belirlediyse marka etkisi veya markalar sabit etki olarak değerlendirilir. Rasgele etki, gözlem birimlerinin işleme tabi tutulan etki ya da etkilerin araştırmacı tarafından rasgele belirlenmiş olduğunu ifade eder. Örneğin; sigara markalarına göre nikotin miktarı konusunda araştırma yapan bir araştırmacı gözlem yapacağı sigara markalarını piyasada bulunan sigara markaları arasından rasgele olarak belirlemiş ve bunlar üzerinden gözlem yapıyorsa marka etkisi rasgele etki olarak değerlendirilir ve model buna göre tanımlanır.

Bir çalışmanın gerçek etki büyüklüğü, temel alınan etki büyüklüğüdür ve çalışmanın sonsuz büyüklükte bir örneklem büyüklüğüne sahip olması durumunda (ve dolayısıyla örneklem hatası olmaması) gözlenen etki büyüklüğüdür. Gerçek etki ile gözlenen etki arasında farkı göstermek için farklı semboller kullanılır. Her bir çalışma için gerçek etki

daire, gözlenen etki kare ile temsil edilir. Birleştirilmiş çalışmalarda; gerçek etki üçgen, gözlenen etki ise karo ile temsil edilir.

Çizelge 7.1 Etki büyüklüklerine ilişkin şekilsel gösterimler

	Gerçek etki	Gözlenen etki
Her bir çalışma için		
Birleştirilmiş çalışmalarda		

Gerçek(bilinmeyen) etki büyüklüğü teta (θ) ile gösterilir. Gerçek etki büyüklüğüne ilişkin tahmin ise $\hat{\theta}$ ile gösterilir.

7.1 Sabit Etkili Model

Sabit etkili model altında meta analizine dahil edilen tüm çalışmaların gerçek bir etki büyüklüğüne sahip olduğu varsayılır. Anlamlı bir şekilde birbirinden farklı olmayan gerçek etki büyüklüğüne sahip çalışmalar homojen olarak adlandırılır ve etki büyüklüklerinin birleştirilmesi için sabit etkili model kullanılır. Yani, etki büyüklüğünü etkileyebilecek tüm faktörler tüm çalışmalar için aynıdır. Sabit etkili modelde özet etki, gerçek etki büyüklüğüne ilişkin tahmindir ve tüm çalışmalara ait ortak etki büyüklüğünün tahminidir. H_0 yokluk hipotezi ortak etkinin sıfır (fark için) veya bir (oran için) olmasıdır. Tüm çalışmalar ortak bir gerçek etkiye sahip olduğu için, gözlenen etki büyüklüğü sadece her çalışmada bulunan rasgele hata nedeniyle çalışmadan çalışmaya değişir. Eğer bir çalışmanın örneklem büyüklüğü sonsuz olursa, örnekleme hatası sıfır olur ve her bir çalışma için gözlenen etki gerçek etki ile aynı olur. Fakat her çalışmada örneklem büyüklüğü sonsuz değildir, bu nedenle örnekleme hatası mevcuttur ve çalışmalarda gözlenen etki ile gerçek etki aynı değildir.

Sabit etkili modeli tahmin edebilmek için 3 farklı yöntem kullanılmaktadır, bunlar Ters Varyans, Peto ve Mantel-Haenszel yöntemi olarak adlandırılmıştır (Bakıoğlu ve Göktaş 2018). Burada ters varyans yöntemi kullanılarak sabit etkili model tahmini yapılacaktır.

$$Y_i = \theta + \varepsilon_i \quad (7.1)$$

Gerçek etki büyüklüğünün en kesin tahminini elde edebilmek ve her çalışmaya atanan ağırlığın hesaplanabilmesi için çalışma varyansının tersi olan ağırlıklı bir ortalama hesaplanır. Bu yöntem özet etkinin daha duyarlı tahminini veren çalışmalara daha fazla ağırlık vermek amacıyla kullanılır.

i. çalışma için çalışma içi varyans,

$$Var(Y_i) = \frac{\sigma^2}{n} \quad (7.2)$$

ve her bir çalışmaya atanan ağırlık,

$$W_i = \frac{1}{Var(Y_i)} \quad (7.3)$$

dır. Her çalışmaya bir ağırlık atandığında, özet etki tüm çalışmalarda ağırlıklı ortalama ile verilir.

Özet etki (Ağırlıklı ortalama),

$$M = \frac{\sum_{i=1}^k W_i Y_i}{\sum_{i=1}^k W_i} \quad (7.4)$$

biçimindedir. Özet etkinin varyansı;

$$Var(M) = \frac{1}{\sum_{i=1}^k W_i} \quad (7.5)$$

Özet etkinin tahmini standart hatası;

$$SE(M) = \sqrt{Var(M)} \quad (7.6)$$

Özet etki için %95 güven düzeyinde alt ve üst sınırlar;

$$LL_M = M - 1.96 \times SE(M) \quad (7.7)$$

$$UL_M = M + 1.96 \times SE(M) \quad (7.8)$$

Özet etkinin sıfır olduğu yokluk hipotezini test etmek için z değeri;

$$Z = \frac{M}{SE(M)} \quad (7.9)$$

Tek kuyruklu bir test için p – değeri;

$$p = 1 - \Phi(\pm|z|) \quad (7.10)$$

fark beklenen yönde ise '+', aksi takdirde '-' seçilir. İki kuyruklu bir test için p – değeri;

$$p = 2[1 - (\Phi(|z|))] \quad (7.11)$$

dir. Burada $\Phi(z)$ standart normal dağılım fonksiyonudur (Borenstein vd. 2009).

7.2 Rasgele Etkiler Modeli

Çalışmalar; katılımcılar, müdahaleler ve işlemlerden dolayı farklılık gösterebileceğinden farklı etki büyüklüklerine sahip olabilir. Rasgele etkiler modeli altında, her çalışmada bulunan çalışma içi rasgele hata ve çalışmalar arası varyans olmak üzere iki varyasyon kaynağı vardır. Çalışmalar arasındaki ve çalışma içerisindeki değişimi değerlendirmek için rasgele etkiler modeli kullanılarak meta analizi yapılmalıdır. Rasgele etkiler modelinde, genellikle gerçek etkilerin normal dağıldığı varsayılır. Rasgele etkiler modelinde özet etki; tek bir gerçek etkinin tahmini değil, etkilerin dağılımının ortalamasının tahminidir. H_0 yokluk hipotezi ortalama etki büyüklüğünün sıfır olmasıdır. Meta analizinde kullanılacak çalışmalar yayımlanmış literatürden toplandıysa rasgele etkiler modelinin kullanılması gerekir(Borenstein vd. 2010).

Herhangi bir çalışma için gözlenen etki Y_i , genel ortalama μ ile gösterilir. Çalışmanın gerçek etkisinin genel ortalamadan sapması ve çalışmanın gözlenen etkisinin çalışmanın gerçek etkisinden sapması ile verilir.

$$[(\theta - \mu) = \zeta_i]$$

$$[(Y_i - \theta) = \varepsilon_i]$$

Yani;

$$Y_i = \mu + \zeta_i + \varepsilon_i \quad (7.12)$$

dir. Bu nedenle, herhangi bir çalışmada gözlenen Y_i etkisinin μ 'den ne kadar sapacağını tahmin etmek için hem ζ_i 'nin varyansını hem de ε_i 'nin varyansının dikkate alınması gerekir. μ 'den θ 'ya (gerçek etkiye) olan mesafe, çalışmalar arasındaki gerçek etkilerin standart sapmasına bağlıdır, buna τ (veya varyansı için τ^2) denir.

Meta analizinde, gözlenen etkiler tahmin edilmeye başlanır ve yığın etkisi tahmin edilmeye çalışılır. Aslında amaç, genel ortalamayı tahmin etmek için Y_i rasgele etkiler modelini kullanmaktır. Çalışmaların örneklem büyüklükleri farklıysa, daha büyük çalışmalardan elde edilen tahminler, daha küçük çalışmalardan elde edilen tahminlerden daha duyarlı olacaktır. Bu durumda daha duyarlı tahminlere daha fazla ağırlık vermek mantıklıdır. Genel ortalamanın en kesin tahminini elde edebilmek ve her çalışmaya atanan ağırlığın hesaplanabilmesi için çalışmanın varyansının tersi olan ağırlıklı bir ortalama hesaplanır. Yani, daha küçük varyansa (daha fazla kesinlik) sahip olan bir çalışmanın ağırlığı daha fazla olacaktır (Hedges ve Vevea 1998). Rasgele etkiler modeli altında bir çalışmanın varyansını hesaplamak için çalışma içi varyans ve çalışmalar arası varyansın (τ^2 'nin) bilinmesi gerekir. Çalışmalar arası varyans bileşeni hesaba katıldığı için çalışmaların homojenliği hakkında yorum yapılabilir.

τ^2 'yi tahmin etmek için, momentler yöntemi (veya DerSimonian ve Laird) kullanılır.

$$T^2 = \frac{Q - (k-1)}{c} \quad (7.13)$$

$$Q = \sum_{i=1}^k W_i Y_i^2 - \frac{(\sum_{i=1}^k W_i Y_i)^2}{\sum_{i=1}^k W_i} \quad (7.14)$$

$$c = \sum_{i=1}^k W_i - \frac{\sum_{i=1}^k W_i^2}{\sum_{i=1}^k W_i} \quad (7.15)$$

dir. Rasgele etkiler modelinde ortalama etki büyüklüğünü tahmin etmek için her çalışma, varyansının tersi ile ağırlıklandırılacaktır. Sabit etkili model ile arasındaki farklılık, varyansın çalışmalar arası varyansın tahminini (T^2) de içermesidir. Varyans,

alışmalar arası varyans bileşenini içerdği için rasgele etkiler modeli için güven aralıkları daha geniş olacaktır. Rasgele etkiler modelini temsil etmesi için bir yıldız işareti (*) eklenmiştir.

i.alışma için alışma içi varyans,

$$Var(Y_i) = \frac{\sigma^2}{n} \quad (7.16)$$

Her bir alışmaya atanan ağırlık;

$$W_i^* = \frac{1}{Var(Y_i^*)} \quad (7.17)$$

alışmalar arası varyans, T^2 'dir. Yani;

$$Var(Y_i^*) = Var(Y_i) + T^2 \quad (7.18)$$

Özet etki (Ağırlıklı ortalama),

$$M^* = \frac{\sum_{i=1}^k W_i^* Y_i}{\sum_{i=1}^k W_i^*} \quad (7.19)$$

Özet etkinin varyansı;

$$Var(M^*) = \frac{1}{\sum_{i=1}^k W_i^*} \quad (7.20)$$

Özet etkinin tahmini standart hatası;

$$SE(M^*) = \sqrt{Var(M^*)} \quad (7.21)$$

Özet etki için %95 güvenle alt ve üst sınırlar;

$$LL_{M^*} = M - 1.96 \times SE(M^*) \quad (7.22)$$

$$UL_{M^*} = M + 1.96 \times SE(M^*) \quad (7.23)$$

Özet etkinin sıfır olduğu yokluk hipotezini test etmek için z değeri;

$$z^* = \frac{M^*}{SE(M^*)} \quad (7.24)$$

Tek kuyruklu bir test için p – değeri;

$$p^* = 1 - \Phi(\pm|z^*|) \quad (7.25)$$

fark beklenen yönde ise '+' veya aksi takdirde '-' seçilir.

İki kuyruklu bir test için p – değeri;

$$p^* = 2[1 - (\Phi(|z^*|))] \quad (7.26)$$

burada $\Phi(z^*)$ standart normal kümülatif dağılımdır.

8. YAYIN YANLILIĞI

Saygı duyulan, sevilen ve etki faktörü (Q_1) yüksek bir dergide yayımlanan makaleler okunduğunda okuyucu yazılanlara inanma eğiliminde olur. Etki faktörü derginin başarısını gösterir. Derginin başarısı yayınların başarısı ile doğrudan ilişkilidir. Böyle dergilerde yayımlanan makaleler, muhtemelen alanında en iyi bilim insanları, uzman hakemler tarafından incelenmiş ve makalelere yayımlanma kararı verilmiştir. Makalelere bakıldığında yazarlar bir sonuç çıkarabilmek için istatistiksel analizleri kullanmış ve bunu doğru şekilde yapmışlardır. Fakat, özellikle etki faktörü yüksek ve yaygın olarak okunan dergilerde yayımlanan makalelerin, yapılan tüm çalışmalardan rasgele olarak seçilmediği ortaya çıkar. Bunun iki sebebi vardır. Birincisi; “nitelikli” yayınların fazla olmaması ve dergide yayımlanabilecek alanın sınırlı olmasıdır. İkincisi; ilgi çekici makalelerin sıkıcı olanlardan daha çok tercih edilip daha sık yayımlanmasıdır. Bu şekilde bakıldığında yanlış bir şey yok gibi görülse de bu yanlılık ciddi sorunlara yol açabilir (Whitlock ve Schluter 2015).

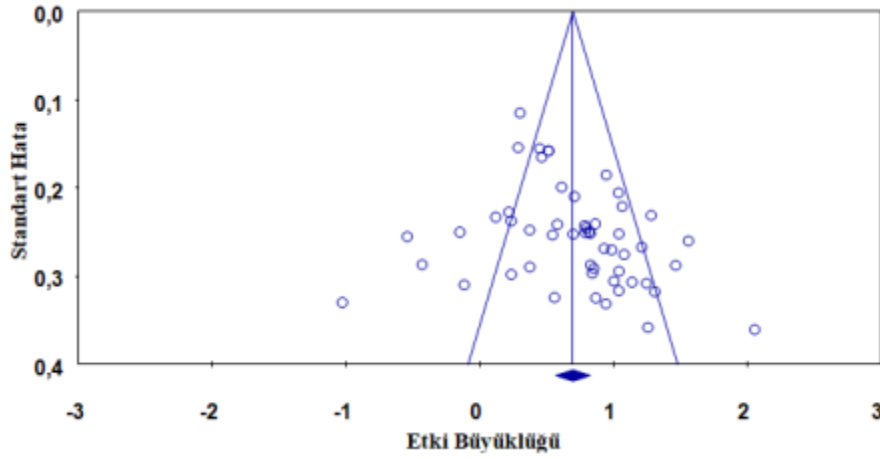
Makalede araştırmaya ait olan yokluk hipotezi reddedilemiyorsa editörler ve yazarlar tarafından makalenin ilgi çekici olmadığı düşünülür. Bu nedenle yokluk hipotezini reddeden makalelerin yayımlanma olasılığı, reddedemeyenlere göre daha fazladır. Ayrıca bir deneyin büyük etkilerini açıklayan makaleler, küçük etkiler gösterenlerden daha ilginçtir. Bu nedenle yayımlanmış makaleler, yayımlanmamış çalışmalara göre daha büyük etkiler gösterme eğilimindedir. Araştırmacılar, “anlamsız sonuçları” yayımlamakta ($p > 0.05$) çekimser davranırlar. Anlamlı sonuç ($p < 0.05$) veren çalışmaların yaygın olarak okunan dergilere girme olasılığı daha yüksektir. Sonuçta 0.05'ten düşük bir p değeri elde eden bir çalışmanın yayımlanma olasılığının, 0.10'dan büyük bir p değeri elde eden çalışmalardan yaklaşık 2.5 kat daha yüksektir (Easterbrook vd. 1991; Dickersin vd. 1992). Düşük veya anlamsız etkiler bulan niteliksiz çalışmaların, benzer şekilde düşük veya anlamsız etkilere sahip nitelikli çalışmalara göre yayımlanmamış olma olasılığı daha yüksektir. Nitelikli bir çalışma yapmış olan bilim insanları, çalışmalarının karşılığını almak için çalışmanın sonucu ne olursa olsun yayımlamaya karardır. Diğer yandan, niteliksiz bir çalışma yürüten ve anlamsız sonuçlar alan bir araştırmacının, çalışmanın yetersiz kaldığını varsayması ve

yayımlanmamış bırakması oldukça olasıdır (Whitlock ve Schluter 2015). Yani yalnızca belirli bir standarda ulaşan çalışmaları seçici olarak yayımlayarak bulguları yanlı hale getirirken, olmayanları bastırmaya yayın yanlılığı denir (Kulinskaya vd. 2008). Yayımlanan bilimsel çalışmaların önemli bir kısmı yanlı bir seçimin sonucu olabilir. Yapılan tüm çalışmalardan hesaplanan etki ile yayımlanan çalışmalardan hesaplanan ortalama etki arasındaki farka yayın yanlılığı denir. Yanlılığın ne kadar olduğu ve yayımlanmayan araştırmalarda hangi sonuçların elde edildiğini bulabilmek kolay değildir, çünkü yayın yanlılığını ölçmek zordur (Whitlock ve Schluter 2015).

Yayın yanlılığı, çeşitli bilimsel alanlarda büyük endişe uyandıran bir konudur. Genellikle yayımlanmış verilere dayanan meta analizi ve karar analizinin yanı sıra sistematik incelemelerin geçerliliğini ve genelleştirilmesini etkileyebilecek ciddi bir sorundur. Yayın yanlılığının varlığını belirlemek ve hatta düzeltmek için çeşitli yöntemler önerilmiş olsa da, hepsinin kendi sınırlamaları vardır. Bu nedenle, en iyi seçenek, ya yapılan her araştırmanın varlığını kayıt altına alarak ya da sonuçları ne olursa olsun tüm çalışmaları yayımlayarak, ilk etapta ortaya çıkmasını önlemek olabilir. Bununla birlikte, zaman ve mekân kısıtlamaları bu ideallere ulaşılmasını zorlaştırabilir. Yayın yanlılığı sorununun üstesinden gelinene kadar tüm hakemler ve okuyucular, yanlı bir deneysel sonuç örneği ile karşılaşabileceklerinin farkında olmalı ve sonuçlarının gücünü buna göre değerlendirmelidirler (Thornton ve Lee 2000). Bunların haricinde yayın yanlılığını belirleme yöntemleri seçim modelleri ve huni grafiğine dayalı yöntemler olarak ikiye ayrılır. Seçim modelleri, genel etki büyüklüğü tahminini bulmak için ağırlık fonksiyonlarından yararlanır ve genellikle yayın yanlılığının etkisini değerlendirmek için duyarlılık analizleri olarak kullanılır. Huni grafiğine dayalı yöntemler, bir huni grafiğinin görsel olarak incelemesini, regresyon ve sıralama testleri ile parametrik olmayan kırp ve doldur yöntemini içerir. Bu yaklaşımlar uygulamalarda yaygın olarak kullanılmasına rağmen, yayın yanlılığını ölçmek için alınan önlemler literatürde az kullanılmaktadır. Bu tür ölçümler meta analizin bir özelliği olarak kullanılabilir ve farklı meta analizler arasında yayın yanlılıklarının karşılaştırılmasına izin verirler (Lin ve Chu 2018).

8.1 Huni Grafiđi

Huni grafiđi, yayın yanlılıđını tespit etme işlevi görür ve meta analizine dahil edilen çalışmaların grafiksel olarak görselleştirilmesini sağlar (Dowdy vd. 2021). Huni grafiđi, yayımlanmış çalışmalarda tespit edilen etki büyüklüğünün ve bunlara karşılık gelen örneklem büyüklüğü, varyans veya standart hatanın grafiđidir. X eksenini üzerinde etki büyüklüğü, Y eksenini üzerinde örneklem büyüklüğü, varyans veya standart hata kullanılarak çizilir. Y ekseninde standart hata, örneklem büyüklüğü veya varyansa göre daha sık kullanılır. Standart hata, meta analizine dahil edilen çalışmaları ağırlıklandırmak için kullanılan bir istatistiktir. Bu, asimetriyi tanımlamayı kolaylaştırabilir (Borenstein vd. 2009). Daha yüksek hassasiyete sahip çalışmalar daha düşük standart hataya sahipken, daha az hassasiyete sahip çalışmalar daha yüksek standart hataya sahiptir (Dowdy vd. 2021). Örneklem büyüklüğü fazla olan çalışmalar grafiđin üst kısmına doğru görünür ve genelde ortalama etki büyüklüğü etrafında kümelenir. Örneklem büyüklüğü küçük çalışmalar ise grafiđin alt kısmına doğru görünür ve (daha küçük çalışmaların etki büyüklüklerinde daha fazla örnekleme hatası değişkenliği olduğundan) geniş bir değer aralığına yayılma eğilimindedir. Bu desen bir huniyi andırır, bu yüzden grafik bu isimle adlandırılır (Light ve Pillemer 1984). Çalışmaların etki büyüklükleri, huni grafiđi kullanılarak yayın yanlılıđını değerlendirmek için çizilirse ve huni şekli oluşturacak şekilde simetrik olarak dağıtılırsa, yayın yanlılıđından şüphe edilmez (Light vd. 1994). Çalışmalar asimetrik olarak dağıtılırsa, yani standart hataların yüksek ve etki büyüklüklerinin küçük olduğu çalışmalar dahil edilmemişse, bu yayın yanlılıđının kanıtını gösterir (Sterne ve Egger, 2001).



Şekil 8.1 Başarı değişkeni ile ilgili çalışmalara ait etki büyüklüklerini gösteren huni grafiği (Sarı ve Ören 2020)

Başarı değişkeni ile ilgili çalışmalara ait etki büyüklüklerini gösteren huni grafiğine bakıldığında her bir çalışmanın etki büyüklüğü ve buna karşılık gelen standart hata genellikle huni çizgilerinin içinde olduğu için simetriye yakın bir dağılım gösterdiği söylenebilir. Bu çalışmada düşük yayın yanlılığı mevcut olabilir (Sarı ve Ören 2020).

9. UYGULAMA

Bu uygulamada, 10-17 yaş aralığında bulunan adölesan çağındaki gençlerin ekran karşısında geçirdikleri süre (screen time) üzerinde cinsiyetin etkisi meta analizi yoluyla değerlendirilecektir. Metinde “ekran karşısında geçirilen süre” ifadesi ekran süresi olarak kullanılacaktır. Covid-19 pandemisi sırasında 10-17 yaş aralığındaki adölesan çağındaki gençlerin ekran sürelerine ilişkin değişik ülkelerde yapılmış çalışmalar incelenmiştir. Bu konuda pandemi döneminde yapılmış pek çok çalışma olmakla birlikte verilen yaş ve yığın sınırlaması kullanılarak Web of Science veri tabanından aynı özellik üzerinde (cinsiyete göre ekran süresi) gözlem yapılan, etki büyüklüğü hesaplamalarında kullanılabilecek istatistikler bulunan ve analize elverişli olan 5 çalışma meta analizine dahil edilmiştir. Wunsch vd. (2021) çalışmasında fiziksel aktivite ve ekran süresinin Covid-19 öncesi ve Covid-19 pandemisi sırasında farklılık gösterip göstermediğini araştırmıştır. Nagata vd. (2022) çalışmasında Covid-19 pandemisi sırasında ebeveyn ve gençlerde sosyodemografik özelliklerin ekran süresi üzerinde bir etkisi olup olmadığını araştırmıştır. Marckhoff vd. (2022) çalışmasında Covid-19 pandemisine bağlı kısıtlamaların Alman gençlerinde fiziksel aktivite, ekran süresi ve zihinsel sağlık üzerindeki etkisini araştırmıştır. Xiao vd. (2021) çalışmasında Covid-19 pandemisi sırasında Çinli gençler arasındaki fiziksel aktivite ve ekran süresinin duygudurum bozukluğu ve ebeveynlerle olan çatışmalarla ilişkisini araştırmıştır. Nagata vd. (2022) çalışmasında gençlerin Covid-19 pandemisi sırasında ekran kullanımını sosyo-demografik kategorilere göre değerlendirmiş ve ekran kullanımıyla ilişkili zihinsel sağlık ve dayanıklılık faktörlerini araştırmıştır. Bu çalışmada örnek olarak kullanılacak veriler bu çalışmalardan derlenmiştir. Konu üzerinde eğitim ve gelişmeyi içerecek uzman değerlendirmesi yapılmayacak olup, konunun herkes tarafından anlaşılır olması nedeniyle seçilmiştir. Çalışmaların künyesi ve gereksinim duyulacak istatistiksel veriler aşağıdaki tabloda verilmiştir.

Çizelge 9.1 Meta analizine dahil edilen çalışmaların künyesi ve özet istatistikleri

Çalışmaya ait yayın başlığı	Yazar	Ülke	Yayın yılı	Yayınlandığı dergi	Cinsiyet					
					Erkek			Kız		
					n	Ortalama	Standart Sapma	n	Ortalama	Standart Sapma
Physical Activity, Screen Time, and Mood Disturbance Among Chinese Adolescents During COVID-19	Xiao vd.	Çin	2021	JOURNAL OF PSYCHOSOCIAL NURSING	862	2.38	2.41	818	2.56	2.52
COVID-19: Effects of Pandemic Related Restrictions on Physical Activity, Screen Time, and Mental Well-being in German adolescents	Marckhoff vd.	Almanya	2022	Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie	411	8.04	4.18	627	7.68	3.76
Screen Time Use Among US Adolescents During the COVID-19 Pandemic: Findings From the Adolescent Brain Cognitive Development (ABCD) Study	Nagata vd.	Amerika Birleşik Devletleri	2022	JAMA Pediatrics	2668	8.18	5.92	2744	7.23	5.52
Parent-Adolescent Discrepancies in Adolescent Recreational Screen Time Reporting During the Coronavirus Disease 2019 Pandemic	Nagata vd.	Amerika Birleşik Devletleri	2022	Academic Pediatrics	2619	3.99	3.37	2716	3.76	3.21
The Impact of COVID-19 on the Interrelation of Physical Activity, Screen Time and Health-Related Quality of Life in Children and Adolescents in Germany: Results of the Motorik-Modul Study	Wunsch vd.	Almanya	2021	Multidisipliner Digital Publishing Institute	344	4.99	2.51	402	4.17	2.33

Her bir çalışmada verilen betimleyici istatistikler saat/gün ölçüm değerleri kullanılarak hesaplanmıştır. Çoğu sosyal bilim çalışmalarında kullanılan yapılara (construct) ilişkin ölçek farklılıkları bu çalışmaların değerlendirilmesinde etkili olmayacaktır; karşılaştırmalara elverişlidir. Örneklem ortalamaları ve standart sapmalar, etki büyüklüğü indeksi olarak standartlaştırılmış ortalamaların farkı (Cohen's d) hesaplamalarında kullanılacaktır.

1. ve 2. çalışmalar için işlemler detaylı olarak gösterilecektir.

Her bir çalışmanın etki büyüklüğünü hesaplamak için S_{pooled} (örnekleme ilişkin ortak-havuzlandırılmış standart sapma) hesaplanacaktır.

1. Çalışma için,

Ortak-havuzlandırılmış standart sapma,

$$S_{pooled_1} = \sqrt{\frac{(862 - 1) \times 2.41^2 + (818 - 1) \times 2.52^2}{(862 + 818 - 2)}}$$

$$S_{pooled_1} = 2.46$$

olarak hesaplanır.

Etki büyüklüğü (Cohen's d),

$$d_1 = \frac{2.38 - 2.56}{2.46}$$

$$d_1 = -0.073$$

Etki büyüklüğünün varyans tahmini,

$$S^2_{d_1} = \frac{862 + 818}{862 \times 818} + \frac{(-0.073)^2}{2(862 + 818)}$$

$$S^2_{d_1} = 0.00238$$

olarak hesaplanır.

2. Çalışma için,

Ortak-havuzlandırılmış standart sapma,

$$S_{pooled_2} = \sqrt{\frac{(411 - 1) \times 4.18^2 + (627 - 1) \times 3.76^2}{(411 + 627 - 2)}}$$

$$S_{pooled_2} = 3.93$$

olarak hesaplanır.

Etki büyüklüğü (Cohen's d),

$$d_2 = \frac{8.04 - 7.68}{3.93}$$

$$d_2 = 0.091$$

Etki büyüklüğünün varyans tahmini,

$$S^2_{d_2} = \frac{411 + 627}{411 \times 627} + \frac{(0.091)^2}{2(411 + 627)}$$

$$S^2_{d_2} = 0.00403$$

olarak hesaplanır.

Bu şekilde 5 çalışma için işlemler tekrarlanır. Her bir çalışma için etki büyüklükleri ve etki büyüklüklerine ait varyanslar çizelge 9.2'de verilmiştir.

Çizelge 9.2 Her bir çalışma için etki büyüklükleri ve etki büyüklüklerine ait varyanslar

Çalışma	Etki büyüklüğü Y_i	Çalışma içi varyans(Varyans) $S^2_{Y_i}$
Xiao vd.	-0.073	0.00238
Marckhoff vd.	0.091	0.00403
Nagata vd.	0.960	0.00083
Nagata vd.	0.069	0.00075
Wunsch vd.	0.340	0.0547

Etki büyüklükleri hesaplandıktan sonra, etki büyüklüklerini birleştirmek için hangi modelin kullanılması gerektiğinin belirlenmesi gerekir. Model seçimi yapılabilmesi için heterojenlik testlerinden yararlanılır. Heterojenlik testinin sonucu dikkate alınmaksızın her iki model için işlemlerin nasıl yapıldığı gösterilecektir.

Sabit etkili model için,

Her çalışma için atanan ağırlıklar,

$$W_1 = \frac{1}{S^2_{Y_i}}$$

$$W_1 = \frac{1}{0.00238} = 420.17$$

$$W_2 = \frac{1}{0.00403} = 248.14$$

olarak hesaplanır. Özet etki,

$$M = \frac{\sum_{i=1}^5 W_i Y_i}{\sum_{i=1}^5 W_i}$$

$$W_1 Y_1 = 420.17 \times (-0.073) = -30.67$$

$$W_2 Y_2 = 248.14 \times 0.091 = 22.58$$

$$\sum_{i=1}^5 W_i Y_i = 1302.7$$

$$\sum_{i=1}^5 W_i = 3389.27$$

$$M = \frac{1302.7}{3389.27}$$

$$M = 0.384$$

olarak hesaplanır. Özet etki Cohen (1988)'in etki büyüklüklerini sınıflamasına göre küçük olarak nitelendirilebilir.

Özet etkinin varyans tahmini ve standart hatası,

$$S^2_M = \frac{1}{3389.27} = 0.00029$$

$$SE_M = \sqrt{0.00029} = 0.017$$

olarak hesaplanır.

Özet etkinin alt ve üst güven sınırları,

$$LL_M = 0.384 - 1.96 \times 0.017 = 0.350$$

$$UL_M = 0.384 + 1.96 \times 0.017 = 0.417$$

olarak hesaplanır.

Z hesap değeri,

$$Z = \frac{0.384}{0.017} = 22.588$$

olarak hesaplanır.

Sabit etkili model kullanıldığında, özet etki %95 güven düzeyinde 0.35 ile 0.42 aralığında değer alır. Z hesap değeri 22.588'dir.

Rasgele etkiler modeli için,

Rasgele etkiler modelini temsil etmesi için yıldız işareti (*) eklenmiştir.

Rasgele etkiler modelinde özet etki çalışmalar arası varyans (τ^2) tahminini de içerir.

T^2 , τ^2 'nin momentler tahmin edicisidir.

Çalışmalar arası varyans tahmini (T^2),

$$T^2 = \frac{Q - (k - 1)}{c}$$

biçimindedir. Buradan Q istatistiği,

$$Q = \sum_{i=1}^k W_i Y_i^2 - \frac{(\sum_{i=1}^k W_i Y_i)^2}{\sum_{i=1}^k W_i}$$

$$W_1 Y_1^2 = 2.24$$

$$W_2 Y_2^2 = 2.06$$

$$\sum_{i=1}^k W_i Y_i^2 = 1770.59$$

$$Q = 1770.59 - \frac{1302.7^2}{3389.27}$$

$$Q = 1269.88$$

olarak hesaplanır.

Serbestlik derecesi,

$$sd = k - 1 = 4$$

dir. c istatistiği,

$$c = \sum_{i=1}^k W_i - \frac{\sum_{i=1}^k W_i^2}{\sum_{i=1}^k W_i}$$

$$c = 3389.27 - \frac{3500895.90}{3389.27}$$

$$c = 2356.34$$

olarak hesaplanır. Buradan çalışmalar arası varyans,

$$T^2 = \frac{1269.88 - 4}{2356.34}$$

$$T^2 = 0.54$$

dır. Her çalışma için atanan ağırlıklar,

$$Var(Y_i^*) = Var(Y_i) + T^2$$

$$W_i^* = \frac{1}{S^2_{Y_i^*}} = \frac{1}{S^2_{Y_i} + T^2}$$

$$W_1^* = \frac{1}{0.00238 + 0.54} = 1.844$$

$$W_2^* = \frac{1}{0.00403 + 0.54} = 1.838$$

dır. Özet etki,

$$M^* = \frac{\sum_{i=1}^5 W_i^* Y_i}{\sum_{i=1}^5 W_i^*}$$

$$W_1^* Y_1 = 1.844 \times (-0.073) = -0.14$$

$$W_2^* Y_2 = 1.838 \times 0.091 = 0.17$$

$$\sum_{i=1}^5 W_i^* Y_i = 2.56$$

$$\sum_{i=1}^5 W_i^* = 9.21$$

$$M^* = \frac{2.56}{9.21}$$

$$M^* = 0.28$$

olarak hesaplanır. Özet etki Cohen (1988)'in etki büyüklüklerini sınıflamasına göre küçük olarak nitelendirilebilir.

Özet etkinin varyansı ve standart hatası,

$$S^2_{M^*} = \frac{1}{9.21} = 0.11$$

$$SE_{M^*} = \sqrt{0.11} = 0.33$$

dır.

Özet etkinin alt ve üst güven sınırları,

$$LL_{M^*} = 0.28 - 1.96 \times 0.33 = -0.37$$

$$UL_{M^*} = 0.28 + 1.96 \times 0.33 = 0.93$$

dır.

Z hesap değeri,

$$Z^* = \frac{0.28}{0.33} = 0.85$$

dir.

Rasgele etkiler modeli kullanıldığında, özet etki %95 güven düzeyinde -0.37 ile 0.93 aralığında değer alır. Z değeri 0.85'dir.

Çizelge 9.3 Sabit etkili model için gerekli hesaplamalar

Çalışma	Etki büyüklüğü	Çalışma içi varyans (Varyans)	Ağırlıklandırma faktörü	Hesaplanan değerler		
	Y			WY	WY^2	W^2
Xiao vd.	-0.073	0.00238	420.17	-30.67	2.24	176542.82
Marckhoff vd.	0.091	0.00403	248.14	22.58	2.06	61573.46
Nagata vd.	0.960	0.00083	1204.82	1156.63	1110.36	1451591.23
Nagata vd.	0.069	0.00075	1333.33	92.00	634.80	1777768.90
Wunsch vd.	0.340	0.0547	182.81	62.16	21.13	33419.50
Toplam			3389.27	1302.7	1770.59	3500895.90

Çizelge 9.4 Rasgele etkiler modeli için gerekli hesaplamalar

Çalışma	Etki büyüklüğü	Varyans	Ağırlıklandırma faktörü	Hesaplanan değer
	Y	$S^2_Y + T^2$	W^*	W^*Y
Xiao vd.	-0.073	0.542	1.844	-0.14
Marckhoff vd.	0.091	0.544	1.838	0.17
Nagata vd.	0.960	0.541	1.849	1.78
Nagata vd.	0.069	0.541	1.849	0.13
Wunsch vd.	0.340	0.545	1.833	0.62
Toplam			9.21	2.56

Heterojenlik testi (Cochran Q),

$$H_0: \delta_1 = \delta_2 = \dots = \delta_5$$

$$H_1: \text{En az biri farklıdır}$$

$$Q = \sum_{i=1}^k W_i Y_i^2 - \frac{(\sum_{i=1}^k W_i Y_i)^2}{\sum_{i=1}^k W_i}$$

$$Q = 1770.59 - \frac{1302.7^2}{3389.27}$$

$$Q = 1269.88$$

Q istatistiği 1269.88 olarak hesaplanmıştır.

$\alpha = 0.10$ anlamlılık düzeyinde kritik değer,

$$X^2_{k-1, \alpha/2} = X^2_{5-1, 0.05}$$

$$X^2_{5-1, 0.05} = 9.49$$

olarak hesaplanır. $Q = 1269.88 > X^2_{5-1, 0.05} = 9.49$ olduğundan H_0 yokluk hipotezi reddedilir. Etki büyüklükleri homojen değildir.

Heterojenlik ölçütü I^2 ,

$$I^2 = \frac{Q - (k - 1)}{Q} \times 100\%$$

$$I^2 = \frac{1269.88 - (5 - 1)}{1269.88} \times 100\%$$

$$I^2 = \%99$$

olarak hesaplanır. Çalışmalar arası çok yüksek heterojenlik mevcuttur. Heterojenlik testi sonuçlarına göre etki büyüklüklerinin birleştirilmesi için rasgele etkiler modeli kullanılması uygun olacaktır.

Sonuç olarak; Covid-19 pandemisi sırasında, 10-17 yaş aralığında bulunan adölesan çağındaki gençlerin ekran karşısında geçirdikleri süre üzerinde cinsiyetin etkisi küçüktür ve anlamlı değildir.



10. SONUÇ

Meta analizi yöntemine genellikle her bilim dalında ihtiyaç duyulmaktadır. Meta analizi, aynı konu üzerinde birbirinden bağımsız olarak yapılan çalışmaların uygun yöntemler kullanılarak birleştirilmesi ve analiz edilmesi ile tek bir sonuç çıkarmak için kullanılan istatistiksel bir yöntemdir. Aynı konuda yapılmış çalışmaların birleştirilmesi ile daha etkin, yansız, güvenilir, doğru ve tutarlı sonuçlar amaçlanır. Meta analizi, uygulanması kolay bir yöntem değildir. Doğru sonuçların elde edebilmesi için, dahil edilecek çalışmaların dikkatli ve yansız olarak seçilmesi, çalışmaları birleştirmek için uygun istatistiksel modelin kullanılması ve elde edilen sonuçların doğru yorumlanması gerekir. Meta analizinde genellikle, sabit etkili model veya rasgele etkiler modeli kullanılmaktadır. Sabit etkili modelde heterojenlik sadece örnekleme hatasından kaynaklanırken, rasgele etkiler modelinde heterojenlik örnekleme hatası ve etki büyüklüklerinin değişkenliğinden kaynaklanır. Uygun istatistiksel model seçiminde göz önüne alınan heterojenliğin belirlenmesinde kullanılan Q istatistiği çalışma sayısı fazla olduğunda yanıltıcı sonuçlar verebilir. Bu nedenle heterojenliğin belirlenmesinde Q istatistiği ile birlikte heterojenlik ölçüleri de dikkate alınmalıdır. Uygulamada birbirinden bağımsız 5 farklı çalışma için etki büyüklükleri Cohen's d ile hesaplanmıştır. İstatistiksel model seçiminde sabit etkili ve rasgele etkiler modelinin nasıl uygulanacağı yöntemlerin açıklayıcı olması için gösterilmiştir. Heterojenlik testi dikkate alınarak model seçimi yapıldığında Cochran Q testi sonucuna göre 5 çalışma heterojen olarak değerlendirilmiştir ve çalışmalar birleştirilirken rasgele etkiler modeli kullanılması uygundur. Rasgele etkiler modeli kullanılarak elde edilen özet etki 0.28'dir ve bu Cohen (1988) etki büyüklüğü sınıflamasına göre düşük etki düzeyi olarak değerlendirilir. Covid-19 pandemisi sırasında adölesan çağındaki gençlerin ekran karşısında geçirdikleri süre ile cinsiyet arasındaki etki küçük ve anlamlı değildir.

KAYNAKLAR

- Bahadur, R. R. 1967. Rates of convergence of estimates and test statistics. *The Annals of Mathematical Statistics*, 38(2), 303-324.
- Bakioğlu, A., & Göktaş, E. 2018. An educational policy making method: Meta analysis. *Medeniyet Eğitim Araştırmaları Dergisi*, 1(2), 35-54.
- Bakioğlu, A., & Özcan, Ş. 2016. Meta analizi. Nobel Akademik Yayıncılık, 272, İstanbul.
- Birnbaum, A. 1954. Combining independent tests of significance. *Journal of the American Statistical Association*, 49(267), 559-574.
- Borenstein, M., Hedges, L. & Rothstein, H. 2009. *Introduction to Meta Analysis*. John Wiley & Sons, Inc., 421, United Kingdom.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. 2010. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research synthesis methods*, 1(2), 97-111.
- Brockwell, S., & Gordon, I. R. 2001. A comparison of statistical methods for meta-analysis. *Statistics in medicine*, 20(6), 825-840.
- Cochran, W. G. 1954. The combination of estimates from different experiments. *Biometrics*, 10(1), 101-129.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, 567, New York.
- DerSimonian, R., & Laird, N. 1986. Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3), 177-188.
- Dickersin, K., Min, Y. I., & Meinert, C. L. 1992. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *Jama*, 267(3), 374-378.
- Dowdy, A., Hantula, D. A., Travers, J. C., & Tincani, M. 2021. Meta-analytic methods to detect publication bias in behavior science research. *Perspectives on Behavior Science*, 45(1), 1-16.
- Easterbrook, P. J., Gopalan, R., Berlin, J. A., & Matthews, D. R. 1991. Publication bias in clinical research. *The Lancet*, 337(8746), 867-872.
- George, E. O. 1977. *Combining Independent One-Sided And Two-Sided Statistical Tests--Some Theory And Applications*. Doctoral Dissertaion, University Of Rochester, New York.

- Glass, G. V. 1976. Primary, secondary, and meta-analysis of research. *Educational researcher*, 5(10), 3-8.
- Hartung, J., Knapp, G., Sinha, B. K., & Sinha, B. K. 2008. *Statistical meta-analysis with applications*. John Wiley & Sons, Inc., 247, New Jersey.
- Hedges, L. V. & Olkin, I. 1980. Vote-counting methods in research synthesis. *Psychological Bulletin*, 88(2), 359-369.
- Hedges, L. V. & Olkin, I. 1985. *Statistical methods for meta-analysis*. CA: Academic Press, 369, San Diego.
- Hedges, L. V. 1981. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107-128.
- Hedges, L. V., & Vevea, J. L. 1998. Fixed-and random-effects models in meta-analysis. *Psychological methods*, 3(4), 486-504.
- Hedges, L. V., Gurevitch, J., & Curtis, P. S. 1999. The meta- analysis of response ratios in experimental ecology. *Ecology*, 80(4), 1150-1156.
- Higgins, J. P., & Thompson, S. G. 2002. Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, 21(11), 1539-1558.
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. 2003. Measuring inconsistency in meta-analyses. *British Medical Journal*, 327(7414), 557-560.
- Huedo-Medina, T. B., Sánchez-Meca, J., Marin-Martinez, F., & Botella, J. 2006. Assessing heterogeneity in meta-analysis: Q statistic or I^2 index?. *Psychological methods*, 11(2), 193-206.
- Hunter, J. E., & Schmidt, F. L. 2000. Fixed effects vs. random effects meta- analysis models: Implications for cumulative research knowledge. *International Journal of selection and assessment*, 8(4), 275-292.
- Kulinskaya, E., Morgenthaler, S., & Staudte, R. G. 2008. *Meta analysis: a guide to calibrating and combining statistical evidence*. John Wiley & Sons, Inc., 260, England.
- Light, R. J., & Pillemer, D. B. 1984. *Summing up: the science of reviewing research*. Harvard University Press, 191, Cambridge.
- Light, R. J., Singer, J. D., & Willett, J. B. 1994. The visual presentation and interpretation of meta-analyses. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis*, 439-453. Russell Sage Foundation, New York.
- Lin, L., & Chu, H. 2018. Quantifying publication bias in meta-analysis. *Biometrics*, 74(3), 785-794.

- Lipták, T. 1958. On the combination of independent tests. *Magyar Tud Akad Mat Kutato Int Kozl*, 3, 171-197.
- Marckhoff, M., Siebald, M., Timmesfeld, N., Janben, M., Romer, G., & Föcker, M. 2022. COVID-19: Effects of Pandemic Related Restrictions on Physical Activity, Screen Time, and Mental Well-being in German adolescents. *Zeitschrift für Kinder-und Jugendpsychiatrie und Psychotherapie*, 1-14.
- Mosteller, F., & Bush, R. R. 1954. *Selected quantitative techniques*. Addison-Wesley, Cambridge.
- Mudholkar, G. S., & George, E. O. 1979. The logit method for combining probabilities. In *Symposium on optimizing methods in statistics*. Academic Press, New York.
- Nagata, J. M., Cortez, C. A., Cattle, C. J., Ganson, K. T., Iyer, P., Bibbins-Domingo, K., & Baker, F. C. 2022. Screen time use among US adolescents during the COVID-19 pandemic: findings from the Adolescent Brain Cognitive Development (ABCD) study. *JAMA pediatrics*, 176(1), 94-96.
- Nagata, J. M., Cortez, C. A., Iyer, P., Ganson, K. T., Chu, J., & Conroy, A. A. 2022. Parent-Adolescent Discrepancies in Adolescent Recreational Screen Time Reporting During the Coronavirus Disease 2019 Pandemic. *Academic pediatrics*, 22(3), 413-421.
- Öngel, E. 1980. *Araştırmacılar İçin Kimi İstatistiksel Teknikler*. AITIA Yayınları, 483, Ankara.
- Özsoy, S., & Özsoy, G. 2013. Effect Size Reporting in Educational Research. *Ilkogretim Online*, 12(2), 334-346.
- Sarı, K., & Ören Şaşmaz, F. 2020. Araştırmaya dayalı öğrenme stratejisinin öğrencilerin akademik başarılarına etkisi: Bir meta analizi çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 35(3), 540-555.
- Sterne, J. A., & Egger, M. 2001. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of clinical epidemiology*, 54(10), 1046-1055.
- Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., & Williams Jr, R. M. 1949. *The american soldier: Adjustment during army life*. Princeton University Press, 599, Princeton.
- Tarlow, K. R. 2016. Teaching principles of inference with ANOVA. *Teaching Statistics*, 38(1), 16-21.
- Thornton, A., & Lee, P. 2000. Publication bias in meta-analysis: its causes and consequences. *Journal of clinical epidemiology*, 53(2), 207-216.

Whitlock, M., & Schluter, D. 2015. The analysis of biological data. W. H. Freeman and Company, 818, New York.

Wilkinson, B. 1951. A statistical consideration in psychological research. Psychological bulletin, 48(2), 156.

Wunsch, K., Nigg, C., Niessner, C., Schmidt, S. C., Oriwol, D., Hanssen-Doose, A., & Woll, A. 2021. The impact of COVID-19 on the interrelation of physical activity, screen time and health-related quality of life in children and adolescents in Germany: results of the Motorik-Modul Study. Children, 8(2), 98.

Xiao, S., Yan, Z., & Zhao, L. 2021. Physical activity, screen time, and mood disturbance among chinese adolescents during COVID-19. Journal of Psychosocial Nursing and Mental Health Services, 59(4), 14-20.

