

# Seçkisiz-Küme Deneylerini Planlamada Varyans Parametrelerinin Kestirilmesi ve Standartlaştırılması: Araştırmacılar için Kısa Bir Rehber

Metin BULUŞ\*

Sakine GÖÇER ŞAHİN \*\*

## Özet

Geçtiğimiz on yılı kapsayan literatür taraması, bireysel seçkisiz atamanın mümkün olmadığı durumlarda, önemli kararlar alınırken etkili kanıt üretebilecek altın standardındaki seçkisiz-küme deneylerinin (SKD) Türkiye'de eğitim ve psikoloji alanlarında yetersiz sayıda olduğunu göstermektedir. SKD sayısının azlığı sadece programın etkililiği konusunda toplam bilgiye zarar vermemekle kalmayıp, önceki bilgiler mevcut olmadığından veya en iyi ihtimalle eksik olduğundan bu tür çalışmaların etkili bir şekilde tasarlanmasını da engellemektedir. Bu çalışmada, mevcut verilerden yola çıkarak varyans parametrelerinin nasıl kestirilebileceği, nasıl standart formlara dönüştürülebileceği ve böylece bunların nasıl yeteri kadar güçlü SKD'lerin planlanmasında kullanılabileceği gösterilmiştir. Bu çalışmada belirtilen amacı gerçekleştirmek üzere halka açık yazılım kullanılmıştır. Burada istatistiksel modelleri tanıtır, parametreleri tanımlayarak, istatistiksel model ve güç formüllerinde varyans parametrelerini tahmini gösterimlerle (notasyon) ilişkilendirerek araştırmacılara adım adım rehberlik edilmeye çalışılmıştır.

**Anahtar Kelimeler:** seçkisiz-küme deneyleri, varyans kestirimi, istatistiksel güç analizi, gerekli en küçük örneklem büyüklüğü.

## GİRİŞ

Seçkisiz-küme deneyleri (SKD), deneklerin işlem koşullarına bağımsız olarak değil, bir grup olarak atandığı deneysel tasarımlardır. Eğitim araştırmalarında, son on yıl içerisinde SKD'lere yönelik ilginin arttığı görülmektedir (Spybrook, Shi, & Kelcey, 2016). Anahtar kelime olarak yalnızca "SKD" kullanıldığında, Bilim Ağı üzerine Makaleler veri tabanında (the Journals on the Web of Science database) SKD'lerle ilgili 1000'den fazla akademik makale yayınlandığı belirlenmiştir. Her ne kadar SKD'ler seçkisiz birey deneyleri kadar etkili olmasa da bir müdahalenin/programın doğası, kümelerin (birey gruplarının) işlem koşullarına atanmasını mümkün kılabilir. Bunun için birtakım sebepler bulunmaktadır. İlk olarak, küme düzeyinde bir program uygulamak daha uygun olabilir. İkinci olarak, mevcut kümeleri kullanmak, maliyeti düşürme ve uygulamanın uygunluğu açısından oldukça yararlı olabilir. Üçüncü olarak, aynı kapsamdaki organizasyonun içerisindeki bazı deneklerin uygulamadan mahrum bırakılması etik olmayabilir. Örneğin, bazı öğrencileri etkili olabileceğini düşündüğümüz bir uygulamaya tabi tutarken bazılarını dışlamak eğitimde etik olmayan bir uygulama olarak kabul edilebilir. Ayrıca SKD'ler, aynı topluluktaki bireylerin işlem koşullarına seçkisiz olarak atanmaması durumunda oluşabilecek kirlenme (contamination) riskini azaltır.

Ancak seçkisiz birey deneylerle kıyaslandığında SKD'leri tasarlamak daha zahmetlidir ve benzer bir istatistiksel güç elde etmek için daha çok katılımcıyla ihtiyaç duyulur; ayrıca öngörülen istatistiksel analizler de daha karmaşıktır (Hayes & Moulton, 2017). Kümelemeyi yok sayan istatistiksel yöntemler

\*Öğr.Gör.Dr., Adıyaman Üniversitesi, Adıyaman, TÜRKİYE, e-mail: [bulusmetin@gmail.com](mailto:bulusmetin@gmail.com), ORCID ID: [orcid.org/0000-0003-4348-6322](https://orcid.org/0000-0003-4348-6322)

\*\*Doktora sonrası araştırmacı, University of Wisconsin-Madison, WI, USA, e-mail: [sgocersahin@gmail.com](mailto:sgocersahin@gmail.com), ORCID ID: [orcid.org/0000-0002-6914-354X](https://orcid.org/0000-0002-6914-354X)

Bu makaleye atıfta bulunmak için:

Buluş, M. & Göçer Şahin, S. (2019). Estimation and standardization of variance parameters for planning cluster-randomized trials: a short guide for researchers. *Journal of Measurement and Evaluation in Education and Psychology*, 10(2), 179-201. DOI: 10.21031/epod.530642

Geliş Tarihi: 22.02.2019  
Kabul Tarihi: 13.06.2019

yanıltıcı sonuçlar verebilir çünkü bunlar, hangi kümeden geldiklerine bakmaksızın bütün deneklerin bağımsız katkı sağladığını varsaymaktadır. Eğitim ortamlarında bağımsız gözlem varsayımı genellikle bağlamsal etkilerin bir sonucu olarak ihlal edilmektedir. Örneğin aynı sınıftaki öğrencilerin aralarında iletişim/etkileşim olabileceği için bu öğrenciler/gözlemler birbirinden bağımsız olmayabilir. Benzer şekilde aynı okuldaki öğrenciler ve öğretmenler, kütüphane ya da laboratuvar gibi diğer okullardan farklılık gösteren ve benzer bağlamsal etkiye sahip olabilecek kaynakları paylaşmaktadırlar. Böyle durumlarda kümelenmeyi yok sayan yöntemleri uygulamak (örneğin en küçük kareler yöntemi), çok dar olan güven aralıklarına ve çok küçük olan p-değerlerinin üretilmesine sebep olabilir (Bland, 2004). Deneysel tasarımlar söz konusu olduğunda, daha dar güven aralıkları ve daha küçük p-değerleri aslında anlamlı bir farklılık olmayan bir durumda bile anlamlı farklılıklara işaret edebilecekleri için araştırmacıları yanlış yönlendirebilir.

İstatistiksel analiz ve örnekleme şemasına bağlı olarak kümelenmeyi ele almanın farklı yolları vardır. Bu sorunu çözenin yollarından biri, kümeye dayanıklı standart hatalara dayalı çıkarımlarda bulunmaktır (örneğin; Cameron & Miller, 2015). Eğer sonuçlar birkaç kümeden oluşan ve genelleştirilemeyen belirli bir alt grup ile ilgili ise, bir başka alternatif de istatistiksel modele işlem göstergesiyle birlikte sabit etkiler olarak kümelenme üyeliğinin dahil edilmesidir. Buna rağmen Hiyerarşik Lineer (doğrusal) Modelleri (HLM) uygulamak (Raudenbush & Bryk, 2002) eğitimde daha yaygındır. Araştırmacılar kümeye dayalı standart hataları ya da örnekleme şemasına bağlı olarak, sabit etkiler kestirim yöntemini kullansalar da varyansı, küme içi ve kümeler arası şeklinde ayrıştırmak kolay değildir, ki bu bizim çalışma boyunca varyans parametrelerini kestirmek ve standartlaştırmak için kullanacağımız bir stratejidir. Dolayısıyla eğitimlerin/programların etkinliği ile ilgili araştırmalara paralel olarak bu çalışmada HLM formülleri referans olarak alınmıştır.

Aynı şekilde, güç analizleri kestirimin standart hatasına dayandığından, benzer gruplama yapısına sahip (aynı okul ve aynı sınıftaki çocuk) çalışmalar planlanırken bağlamsal etkiler dikkate alınmalıdır. Bir araştırmacının deneysel bir çalışma öncesinde (apriori veya önsel) güç oranını tahmin edebileceği yaklaşık standart hata formüllerini türeten çeşitli çalışmalar vardır (örneğin; Bloom, 1995; Bloom, 2006; Bloom, Bos, & Lee, 1999; Dong & Maynard, 2013; Hedges & Rhoads, 2010; Konstantopoulos, 2009a, 2009b).

Dünya genelindeki birçok eğitim sisteminde ve ülkede SKD'lerin kullanımına dair artan ilgiye rağmen, geçtiğimiz on yılın literatür incelemesi, Türkiye'deki eğitim ve psikoloji araştırmalarında SKD'lerin kullanımının çok az olduğunu göstermektedir. Ayrıca mevcut çalışmalarda istatistiksel güç analizi ya yok ya da örneklemin gruplama yapısını dikkate almamaktadır. Ulakbim Tübitak Dergi Park Veri Tabanı'nda bulunan Türkçe dergilerdeki eğitim araştırmaları alanında 2018 yılında yayınlanan 174 deneysel çalışma, etkili örneklem büyüklüğüne karar vermek için herhangi bir güç analizi yapıp yapılmadığını görmek adına yazarlar tarafından incelenmiştir. Bu çalışmaların hiçbiri SKD'lerle ilgili deneysel çalışmalar olmasa da yazarların herhangi bir güç analizi bulgusuna yer vermediği gözlenmiştir. Bu çalışmalarda sonuçlar, çoğunlukla aslında var olmasına rağmen deneyin önemli bir işlem etkisini tespit edemediği küçük örneklem büyüklüğünden olumsuz etkilenmektedir.

Önsel güç analizi ile ilgili önemli bir nokta da formüllerde kullanılan varyans parametrelerinin bilinmemesidir. Güç hesaplamaları için ihtiyaç duyulan diğer parametreler ya genel olarak kabul edilen standartlardır ya da kestirime ihtiyaç duyulmaz veya geniş kapsamlı metodolojik uzmanlık gerektirmezler. Örneğin eğitim araştırmalarında standart bir uygulama, güç oranını %80'de tutmak, %5 oranında 1. tip hataya sahip olmak ve işlem etkinliğini test eden çift yönlü hipotezler kurmaktır (Dong & Maynard, 2013). Ayrıca, örneklem büyüklüğü bilgileri (örneğin, okul başına ortalama öğrenci sayısı) idari kayıtlardan elde edilebilir veya betimsel istatistikler aracılığıyla hesaplanabilir.

Her ne kadar mevcut verilerden standartlaştırılmış varyans parametrelerini belirlemek ile ilgili güncel bir alanyazın var olsa da (örneğin; Hedberg & Hedges, 2014; Hedberg, 2016; Hedges & Hedberg, 2013; Spybrook, Westine, & Taylor, 2016; Westine, 2016; Westine, Spybrook, & Taylor, 2014; Zopluoglu, 2012), büyük çoğunluğu ABD'deki K-12 akademik çıktılara odaklanan bu sonuçlar, belirli bir coğrafi alana ya da başka konuları içeren alanlara uygulanamayabilir. Varyans parametreleri genellikle ilgili örneğe ve konuya özgü olup, bu parametreler tercihen ilgilenilen coğrafi alana mümkün olduğunca yakın ve inceleme altındaki içeriğe mümkün olduğunca benzer olan önceki

araştırmalardan veya ampirik verilerden elde edilmelidir. Böylece varyans parametrelerinin benzer türdeki daha önce yapılan araştırmalardan kestirilmesi ve standartlaştırılması, özellikle de çok az ön bilgi olan ya da hiç ön bilgi olmayan durumlarda, araştırmacılar için vazgeçilmez bir araç haline gelir.

### **Çalışmanın Amacı**

Bu çalışmanın amacı, eğitim ve psikoloji alanındaki araştırmacılara çok az ön bilgi olan ya da hiç ön bilgi olmayan durumlarda etkili SKD'ler planlama konusunda rehberlik etmektir. Spesifik olarak bu çalışma okuyuculara, HLM kullanarak mevcut verilerden varyans parametrelerini kestirme, bunları sınıf içi korelasyon katsayıları ve  $R^2$  değerleri gibi iyi bilinen varyans parametreleri bakımından standart hale getirme ve standart hale getirilmiş parametreleri istatistiksel güç ve SKD'lerin planlanması için gereken minimum örneklem büyüklüğü hesaplamalarında kullanılmasını göstermeyi amaçlamaktadır.

## **YÖNTEM**

Bu çalışmada HLM'de iki ve üç seviyeli SKD'ler için modeller ile karma model formları ele alınmış olup, parametreler Dong ve Maynard'ın (2013) çalışmasındaki gibi tanımlanmıştır. Ayrıca R ortamındaki (R Core Team, 2019) `lme4` (Bates, Maechler, Bolker, & Walker, 2015) kütüphanesi yoluyla işlem etkisinin nasıl kestirileceği ve varyans parametrelerinin nasıl elde edileceği gösterilmiştir. Son olarak da standartlaştırılmış araştırma deseni parametrelerinden bazılarını (yani sınıf içi korelasyon katsayıları ve  $R^2$  değerleri) hesaplamak için tahmini varyans parametreleri (standart hale getirilmemiş) kullanılmış ve bunlar `PowerUpR` (Bulus, Dong, Kelcey, & Spybrook, 2019) ile istatistiksel güç analizi işlemlerinde kullanılmıştır. Çoğu durumda R ortamında iki kütüphane kullanmak SKD'leri analiz etmek ve planlamak için yeterli olacaktır; ancak işlemin karmaşıklığına bağlı olarak, araştırmacılar tercih edilen herhangi bir yazılım veya platformu da kullanabilirler.

İdeal olarak bir SKD'den elde edilen sonuçlar, bağımlı değişkendeki değişkenlik, eş-değişkenlerin açıklayıcılık gücü ve pek çok istatistiksel model yoluyla elde edilebilecek işlem etkisi bakımından bilgilendirici olmalıdır. SKD'lerin hem planlaması hem de analizinde araştırmacıları bilgilendirebilen minimal olarak yeterli olan modeller, boş (null) ve tam (full) modellerdir. Boş modeller (ayrıca koşulsuz model olarak da bilinir) çıktıda (yani, bağımlı değişken) koşulsuz değişime dair bir farkındalık elde etmek için kullanılabilirken, tam model hem işlem etkisini hem de bağımlı değişkendeki koşullu değişimi tahmin etmek için kullanılabilir. İki ve üç seviyeli SKD'ler için boş ve tam modeller aşağıda tanımlanmıştır.

### **İki Seviyeli SKD'ler**

#### *Koşulsuz Değişkenliği Kestirmek için Boş Model*

Buradaki koşulsuz model, aşağıda tanımlandığı gibi, tam modelden elde edilen parametrelerle birlikte standartlaştırılmış varyans parametrelerini de hesaplamak için kullanılacak olan  $\sigma^2$  ve  $\tau^2$  varyans parametrelerini elde etmek için kullanılabilir.

HLM formülü:

$$\text{Seviye 1: } Y_{ij} = \beta_{0j} + r_{ij}$$

$$\text{Seviye 2: } \beta_{0j} = \gamma_{00} + \mu_{0j}$$

Karma model formülü:

$$Y_{ij} = \gamma_{00} + \mu_{0j} + r_{ij}$$

Burada  $r_{ij}$  ve  $\mu_{0j}$  seviye 1 ve seviye 2'den artıklar (residuals) olup, sırasıyla  $r_{ij} \sim N(0, \sigma^2)$  ve  $\mu_{0j} \sim N(0, \tau^2)$  olarak normal dağılımı takip ederler. Dolayısıyla,  $\sigma^2$  ve  $\tau^2$  sırasıyla seviye 1 ve seviye 2 birimleri arasındaki çıktıda bulunan değişkenliktir.  $Y_{ij}$ ,  $j$  kümesindeki  $i$  deneği için 1.seviye ile ilgili çıktıdır,  $\beta_{0j}$  1.seviye ara kesit (bu durumda,  $j$  kümesindeki deneklerin ortalaması),  $\gamma_{00}$  ise 2.seviye ara kesittir (bu durumda, bütün kümelerdeki bütün deneklerin ortalamasıdır – genel ortalama).

#### *İşlem Etkisi ve Koşullu Değişkenliği Kestirmek için Tam Model*

Buradaki model, aşağıda tanımlandığı gibi, koşulsuz modelden elde edilen parametrelerin yanı sıra standartlaştırılmış varyans parametrelerini de hesaplamak için kullanılan  $\sigma_{|X}^2$  ve  $\tau_{|W}^2$  varyans parametrelerini elde etmek için kullanılabilir.

HLM formülü:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \delta T_j + \gamma_{01}W_j + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

Karma model formülü:

$$Y_{ij} = \gamma_{00} + \delta T_j + \gamma_{01}W_j + \gamma_{10}X_{ij} + \mu_{0j} + r_{ij}$$

Burada  $r_{ij}$  ve  $\mu_{0j}$  seviye 1 ve seviye 2'den elde edilen artıklar olup, sırasıyla  $r_{ij} \sim N(0, \sigma_{|X}^2)$  ve  $\mu_{0j} \sim N(0, \tau_{|W}^2)$  olarak normal dağılım gösterirler. Dolayısıyla,  $\sigma_{|X}^2$  ve  $\tau_{|W}^2$  sırasıyla seviye 1 ve seviye 2 birimleri arasındaki çıktıda bulunan koşullu varyanslardır.  $Y_{ij}$ ,  $j$  kümesindeki  $i$  deneği için 1.seviye ilgili çıktı,  $X_{ij}$   $j$  kümesindeki  $i$  deneği için 1.seviye eş-değişken,  $T_j$  işlem koşulu ( $j$  kümesi işleme atandıysa 1, atanmadıysa 0) ve  $W_j$ ,  $j$  kümesi için 2.seviye eş-değişken,  $\beta_{0j}$  1.seviye ara kesit,  $\gamma_{00}$  2.seviye ara kesit,  $\delta$  işlem etkisi,  $\beta_{1j}$  veya  $\gamma_{10}$ , 1.seviye eş-değişken olan  $X_{ij}$  için regresyon ağırlığı ve  $\gamma_{01}$  2.seviye eş-değişken olan  $W_j$  için regresyon ağırlığıdır.

Koşulsuz ve tam modellerden standartlaştırılmamış varyans parametrelerine dayanarak standartlaştırılmış varyans parametreleri hesaplanabilir.  $\rho = \tau^2 / (\tau^2 + \sigma^2)$  2.seviye birimler arasındaki çıktıda bulunan varyans oranını temsil etmektedir (ayrıca literatürde sınıf içi korelasyon katsayısı olarak da ifade edilmektedir),  $R_1^2 = 1 - \sigma_{|X}^2 / \sigma^2$  1.seviye eş-değişkenler tarafından açıklanan çıktıdaki varyans oranıdır,  $R_2^2 = 1 - \tau_{|W}^2 / \tau^2$  ise 2.seviye eş-değişkenler tarafından açıklanan çıktıdaki varyans oranıdır. Ayrıca işlem etkisi  $\delta^* = \delta / \sqrt{\tau^2 + \sigma^2}$  olarak Cohen'in  $d$ 'si şeklinde standart hale getirilebilir; bundan sonra işlem etkisinden çoğunlukla etki büyüklüğü olarak bahsedilecektir.

Tam modelde, işlem etkisi ve ilişkili  $t$  istatistiği için bir tahmin elde edebilir. “Bir işlem etkisi vardır” hipotezi, boş hipotezin “işlem etkisi yoktur”una karşı test edilir. Tam modelde  $t$  istatistikleri, Tip I hata oranı verilen kritik  $t$  değeri ile karşılaştırarak ( $\alpha$ , altta yatan evrende aslında yokken işlem etkisinin saptanma olasılığı) sonuçların tesadüfün ötesinde açıklanıp açıklanmayacağı incelenebilir. Benzer şekilde,  $t$  istatistiklerini bilerek, Tip II hata oranı hakkında bir fikre sahip olunabilir ( $\beta$ , altta yatan evrende aslında bir etki varken hiçbir işlem etkisinin saptanmama olasılığı). Uygulamada, aslında altta yatan evrende bir etki olduğunda bir işlem etkisinin saptanma olasılığı ile ilgilenilir. Bu istatistiksel güç olup,  $(1 - \beta)$  olarak adlandırılır. İstatistiksel gücü hesaplamak için, deney zaten tamamlandığı için kullanışlı olmasa bile deneyden sonra  $t$  istatistikleri kullanılabilir. Bununla birlikte, daha önce birçok kez tekrar edilmiş bir örneklem büyüklüğü aracılığıyla yeterli istatistiksel güç üretmesi muhtemel olan bir deney planlayabiliriz. Bir deneyden önce istatistiksel gücü hesaplamak için önceki çalışmalardan elde edilen bazı bilgilere ihtiyaç duyulur. Bunlar, anlamlı bir işlem etkisinin ne olacağına dair bir kestirim (genellikle eğitimde 0,20 veya 0,25 olarak ayarlanır ancak alanyazında

daha büyük etki büyüklüğü varsa bu katsayı arttırılabilir) ve beklenen  $t$  istatistiklerini hesaplamak için kullanılacak standart hatadır.

### *Dengeli Örneklem Büyüklüğü ve Homojen Varyans Kapsamında Standart Hata Formülü*

Seviye 1 varyanslarının  $J$  sayıdaki seviye 2 birimlerinde eşit olduğu ve seviye 1 örneklem büyüklüklerinin dengelenmiş olduğu varsayılırsa, standartlaştırılmış varyans şu formu alır:

$$Var(\delta^*) = \frac{\rho(1 - R_2^2)}{p(1 - p)J} + \frac{(1 - \rho)(1 - R_1^2)}{p(1 - p)nJ}$$

Uygulama etkisinin standart hatası  $SE(\delta^*) = \sqrt{Var(\delta^*)}$ 'dir ve eğer  $\delta^*$  ve  $SE(\delta^*)$  biliniyorsa, hangi çıkarımların yapıldığına dayanarak  $t$  istatistikleri hesaplanabilir.  $\delta^*/SE(\delta^*)$ ,  $g$ , 2.seviyede eklenen eş-değişkenlerin sayısını temsil etmekte olup,  $J - g - 2$  serbestlik derecesi ile  $t$  dağılımını takip eder (Bloom, 2006, syf. 17; Dong & Maynard, 2013, syf. 51). Çift yönlü hipotez testi için istatistiksel güç  $(1 - \beta)$  şu şekilde hesaplanabilir:

$$1 - \beta = P(t_{df}(\lambda) > t_{df,1-\alpha/2}(0)) + P(t_{df}(\lambda) < t_{df,\alpha/2}(0))$$

Burada iki seviyeli SKD için serbestlik derecesi,  $df = J - g - 2$ 'dir;  $t_{df,\alpha/2}(0)$ ,  $df$  serbestlik derecesi ve  $\alpha/2$  olasılık ile merkezi  $t$  dağılımı ile ilişkili olan bir istatistiktir;  $t_{df}(\lambda)$  ise  $\lambda = \delta^*/SE(\delta^*)$  merkezizlik parametresi ve  $df$  serbestlik derecesi ile merkezi olmayan  $t$  dağılımı ile ilişkili olan bir istatistiktir.  $\alpha$  ve  $\beta$  ise I. Tip ve II. Tip hata oranlarıdır (bkz. Hedges & Rhoads, 2010; Moerbeek & Safarkhani, 2018). Aşağıda, varyans parametrelerinin nasıl kestirileceği ve  $Var(\delta^*)$  formülünde ihtiyaç duyulan parametrelerin nasıl hesaplanacağı gösterilmiştir.

### *İşlem Etkisi ve Varyans Bileşenlerinin Kestirimi ve Standartlaştırılması*

Önceden yüklenmediyse `lme4` ve `PowerUpR` kütüphaneleri, R ortamında `install.packages(c("lme4", "PowerUpR"))` komutu kullanılarak yüklenmelidir. Bu paketler, `library(lme4)` ve `library(PowerUpR)` komutları kullanılarak mevcut R oturumuna yüklenebilir. R'deki varyans kestirim işlemlerini göstermek için, eğitim ortamlarını dikkate alarak, 100 okulda 2000 öğrencisi olan CRT2 isimli basit iki seviyeli bir SKD verisi üretiyoruz (okul başına 20 öğrenci). Veride beş değişken bulunmaktadır: okul numarası (`schid`), bir 1.seviye çıktı değişkeni (`outcome`), bir 2.seviye işlem değişkeni (`treatment`), bir 1.seviye eş-değişken (`covx`), ve bir 2.seviye eş-değişken (`covw`). 1.seviye ve 2.seviye eş-değişkenlerin sayısı analiz sürecini çok değiştirmeyecektir. Çıktı değişkeni sürekli olup matematik, fen veya okuma puanları gibi belirli bir konu için bir başarı göstergesi olarak kabul edilebilir. İşlem; fen, teknoloji, mühendislik ve matematik (STEM) programı gibi öğrencinin başarı puanlarını arttırmayı amaçlayan herhangi bir uygulama olabilir. 1.seviye ve 2.seviye eş-değişkenler, öğrenci ön test puanı ve ortalama okul düzeyi ön test puanı olabilir. İlk olarak üretilen verinin ilk birkaç satırı aşağıda gösterilmiştir. Her bir okul kendine özgü bir okul numarasına sahiptir (`schid`). Okullar işlem koşullarına atandığından, aynı okul numaraları işlem değişkeni için aynı değerlere sahip olacaktır (`treatment`). 1.seviye (öğrenciler) ve 2.seviye (okullar) eş-değişkenler (`covx` ve `covw`), standart normal dağılımı takip eder ve çıktı (`outcome`), bu değişkenlerin rastgele hata eklenmiş bazı 1.seviye ve 2.seviyelerin doğrusal bir fonksiyonudur (Bkz. Ek A'daki veri üretim süreci). Bu noktadan sonra R komut dizileri gölgelendirilmiş kutular içerisinde. Kodların yanı sıra, yorumlar `##--` ve çıktılar `##` ile başlamaktadır.

```
head(CRT2)
##      schid treatment  outcome      COVX      COVW
## 1      1          0 -0.7145407 -0.37560287 0.2533185
## 2      1          0  0.2411899 -0.56187636 0.2533185
## 3      1          0 -0.8423327 -0.34391723 0.2533185
## 4      1          0 -0.9780591  0.09049665 0.2533185
## 5      1          0  3.2965023  1.59850877 0.2533185
## 6      1          0  1.7267023 -0.08856511 0.2533185
```

İlk olarak, sınıf içi korelasyon katsayılarını hesaplamak üzere koşulsuz model için varyans parametreleri kestirilir. Çıktı; okul ortalamaları ( $\tau_2$ ) ve öğrenciler arasındaki ( $\sigma_2$ ) değişimi gösteren iki rastgele etki için varyansı içerir. İkisinin toplamı çıktı değişkeninin varyansı ile neredeyse aynıdır. Böylelikle, sınıf içi korelasyon ( $\rho_2$ ) olarak da bilinen okullar arasındaki varyans oranı hesaplanabilir.

```
## -- install.packages(c("lme4", "PowerUpR"))
library(lme4) # for estimation
library(PowerUpR) # for power analysis

## -- null model (unconditional model)
null.model <- lmer(outcome ~ (1 | schid), data = CRT2)
print(VarCorr(null.model), comp = "Variance")

## Groups   Name          Variance
## schid    (Intercept)  1.2253
## Residual                                1.9601

## -- variance parameters
tau2 <- 1.2253
sigma2 <- 1.9601

## -- intra-class correlation coefficient
rho2 <- tau2 / (tau2 + sigma2)
round(rho2, 2)

## [1] 0.38
```

Daha sonra,  $R^2$  değerlerini hesaplamak için koşulsuz model varyans parametreleri aracılığıyla tam modele ait varyans parametreleri kestirilir. Yine çıktı, seviye 2 ve seviye 1 yordayıcıları tarafından açıklananların ötesinde, bağımlı değişkendeki koşullu değişimi (okullar ( $\tau_{2w}$ ) ve öğrenciler ( $\sigma_{2x}$ ) arasındaki) gösteren iki rastgele etkinin varyansını içerir. Okullar ve öğrenciler arasındaki varyansın bir kısmı sırasıyla 2.seviye ve 1.seviye yordayıcılar tarafından açıklandığı için, varyans bileşenlerinin boş modele göre azaldığını göz önünde bulundurun. 2.seviye ve 1.seviye için varyansta azalma oranı kullanılarak, her bir ( $r_{21}$  and  $r_{22}$ ) seviye için  $R^2$  değerleri hesaplanabilir.

```
## -- full model
full.model <- lmer(outcome ~ treatment + covx + covw + (1 | schid),
data = CRT2)
print(VarCorr(full.model), comp = "Variance")

## Groups   Name                Variance
## schid    (Intercept) 0.85332
## Residual                                0.98335

## -- variance parameters
tau2w <- 0.8533
sigma2x <- 0.9834

## -- R-squared values for level 1 and level 2
r21 <- 1 - (sigma2x / sigma2)
r22 <- 1 - (tau2w / tau2)

round(r21, 2)

## [1] 0.5

round(r22, 2)

## [1] 0.3
```

Ayrıca işlem etkisini (delta) hesaplayabilir ve bunu, Cohen'in  $d$ 'si ( $es$ ) şeklinde bulunan çıktı varyansı ile standartlaştırabiliriz. Böylelikle, etki daha önceki alanyazın ile kıyaslanabilir, gelecekteki çalışmalarla da karşılaştırılabilir ve gerekirse istatistiksel güç analizi işlemlerinde de kullanılabilir.

```
## -- treatment effect
coef(summary(full.model))["treatment",]

## Estimate Std. Error t value
## 0.9849094 0.1930537 5.1017374

delta <- 0.9849
es <- delta / sqrt(sigma2 + tau2)
round(es, 2)

## [1] 0.55
```

### *İstatistiksel Güç ve Gereken Minimum Örneklem Büyüklüğü Hesaplamaları*

İstatistiksel güç ve gereken minimum örneklem büyüklüğünü belirlemeden önce netleştirilmesi gereken birkaç nokta bulunmaktadır. Daha önce güç analizinde kullanmak üzere varyans parametreleri kestirilmiş ve bunlar standart hale getirilmişti. Ancak bu aşamada başka parametrelere de ihtiyaç duyulabilir. Bu parametrelerin çoğunluğu ya genel olarak kabul edilmiş standartlara sahiptir ya da biliniyordur (ya da metodolojik uzmanlık gerektirmeyen basit işlemlerle elde edilebilir). Eğitim araştırmalarında 0,20 veya 0,25'lik bir etki büyüklüğü ( $es$ ) için testin gücünü hesaplamak, 0,05 Tip I hata oranına ( $\alpha$ ) sahip olmak, ve iki yönlü (*two.tailed*) bir hipotezi testi etmek yaygındır. Diğer taraftan, amaç minimum gerekli örneklem büyüklüğünü bulmak olduğunda ise bunlara ek olarak, güç oranının %80 olduğu varsayılmaktadır. Dahası, okulların yarısını işlem grubuna ( $p$ ) atamak, ideal güç oranı veya ideal gereken minimum örneklem büyüklüğünü üretir ( $p = 0,50$  iken standart hata formülünün paydasındaki  $p(1 - p)$ 'nin maksimum olduğunu göz önünde bulundurun).

Bu çalışmadaki örnekte, okul başına 20 öğrencinin ( $n$ ) olduğu ve toplamda 100 okul ( $J$ ) bulunduğu bilinmektedir. Bu örneğe ilişkin istatistiksel güç şu şekilde hesaplanabilir:

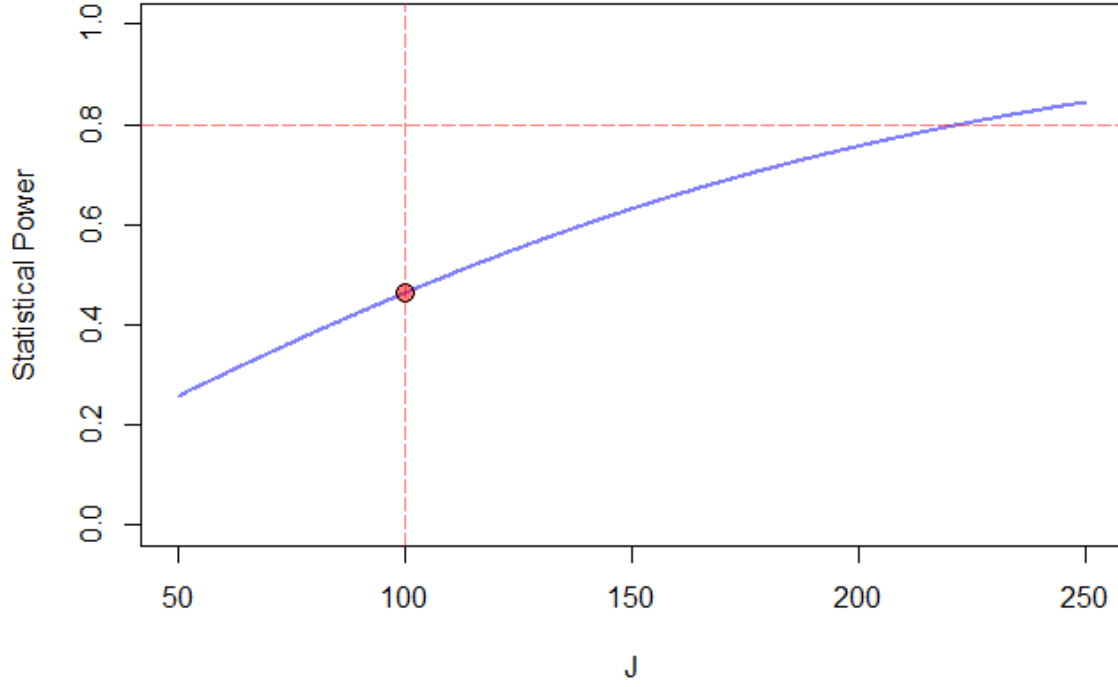
```
## -- power analysis
design <- power.cra2r2(es = .20, alpha = .05, two.tailed = TRUE,
                      rho2 = .38, r21 = .50, g2 = 1, r22 = .30,
                      p = .50, n = 20, J = 100)

##
## Statistical power:
## -----
## 0.463
## -----
## Degrees of freedom: 97
## Standardized standard error: 0.106
## Type I error rate: 0.05
## Type II error rate: 0.537
## Two-tailed test: TRUE
```

Burada, daha önce tanımlanan parametrelere ek olarak,  $g_2$  2.seviyede eklenen eş-değişkenlerin sayısıdır. Verilerden elde edilen parametreler, %46,3 oranında bir güç üretir; bu da bu deneyi daha çok sayıda tekrarladığımız zaman, aslında altta yatan evrende işlem etkisi varsa, tekrarların %46,3'ünde istatistiksel olarak anlamlı bir işlem etkisi tespit edeceğimiz anlamına gelir. Güç analizi literatüründe bu, tavsiye edilen %80'lik referans güç oranının altındadır. Diğer bir deyişle, bu durum, bir uygulamanın etkili olup olmadığına karar vermek için yazı tura atmaktan daha kötüdür. Şekil 1, referans güç oranından ne kadar uzakta olduğumuzu göstermektedir. Bu grafiğe göre 200 ile 250 okul arasında bir sayıdan oluşan bir örneklemin, %80 güç oranıyla sonuç üretme kabiliyetine sahip olduğu görülmektedir.

```
plot(design, ypar = "power", locate = TRUE, xlim = c(50, 250))
```





Şekil 1. İki Seviyeli SKD Örneği için Okul Sayısının Bir Fonksiyonu Olarak İstatistiksel Güç

%80 güç oranı ile 0,20'lik bir etki büyüklüğünü belirlemek için ihtiyaç duyulan okulların kesin sayısı, PowerUpR (aşağıda komutları verilmiştir) ya da PowerUp!'da (Şekil 2) şu şekilde bulunabilir:

```
# -- minimum required sample size
mrss.cra2r2(power = .80, es = .20, alpha = .05, two.tailed = TRUE,
            rho2 = .38, r21 = .50, g2 = 1, r22 = .30,
            p = .50, n = 20)

## J = 223
```

Model 3.1: Sample Size Calculator for 2-Level Cluster Random Assignment Design (CRA2_2_)— Treatment at Level 2		
Assumptions		Comments
MRES = MDES	0.20	MRES = MDES
Alpha Level ( $\alpha$ )	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Power ( $1-\beta$ )	0.80	Statistical power (1-probability of a Type II error)
Rho (ICC)	0.38	Proportion of variance in outcome that is between clusters
n (Average Cluster Size)	20	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
Sample Retention Rate: Level 2 units	100%	Proportion of Level 2 units retained in analysis sample
Sample Retention Rate: Level 1 units	100%	Proportion of Level 1 units retained in analysis sample
P	0.500	Proportion of sample randomized to treatment: $J_T / (J_T + J_C)$
$R_1^2$	0.500	Proportion of variance in Level 1 outcome explained by Level 1 covariates
$R_2^2$	0.300	Proportion of variance in Level 2 outcome explained by Level 2 covariates
$g^*$	1	Number of Level 2 covariates
Priori-M (Multiplier)	2.81	Computed from Priori-T1 and Priori-T2
M (Multiplier)	2.81	Automatically computed
J (Sample Size [Clusters #])	223	Number of clusters needed for given MRES

**RUN**

Note: The parameters in yellow cells need to be specified. Then click "RUN" to calculate sample size.

### Şekil 2. İki Seviyeli SKD Örneği için Gereken Minimum Okul Sayısı

Buna göre okul başına düşen ortalama öğrenci sayısı ( $n = 20$ ), sınıf içi korelasyon katsayısı ( $\rho = 0,38$ ), 1.seviyedeki eş-değişkenlerin açıklayıcılık gücü ( $R_1^2 = 0,50$ ) ve 2.seviyedeki eş-değişkenlerin açıklayıcılık gücü ( $R_2^2 = 0,30$ ) bakımından elimizde olana benzer bir örneklem ile işlem etkisinin çift yönlü hipotez testinde %80'lik bir testin gücü ve %5'lik I.Tip hata oranı ile 0,20'lik bir etki büyüklüğünü tespit etmek için en az 223 okula ihtiyacımız vardır.

### Eş-değişkenlerin Açıklayıcılık Gücü

Araştırmacılar, iki seviyeli bir SKD uygulamasından önce güç oranını artırmak için örneklem büyüklüğünü kontrol edebilir. Ancak bazı durumlarda daha çok birim örneklemek mümkün değildir ya da çok yüksek maliyetlere sebep olabilir. Bu durumda bir seviye için eş-değişkenlerin açıklayıcılık gücü, daha fazla bilgi toplanarak artırılabilir ve bu da güç oranını iyileştirir. Burada doğal olarak akla gelen soru 1.seviye birimlere dair mi yoksa 2.seviye birimlere dair mi daha fazla bilgi toplanması gerektiğidir. Bu soruyu cevaplamak için  $R_1^2$  veya  $R_2^2$ 'deki değişikliklerin ne ölçüde işlem etkisi için varyansta değişikliğe sebep olduğunu,  $Var(\delta^*)$ 'nin  $R_1^2$  veya  $R_2^2$ 'ye göre birinci türevini alarak belirleyebiliriz. Buna göre  $Var(\delta^*)$ 'daki değişiklikler,  $R_1^2$  veya  $R_2^2$ 'deki değişikliklerin aksi yönünde meydana gelmektedir (negatif işaretleri göz önünde bulundurunuz). Bu da  $R_1^2$  veya  $R_2^2$  arttırıldığında,  $Var(\delta^*)$ 'ın azalacağı ve böylece güç oranının gelişeceği anlamına gelmektedir.

$$\frac{\partial Var(\delta^*)}{\partial R_2^2} = -\frac{\rho}{p(1-p)J}$$

$$\frac{\partial \text{Var}(\delta^*)}{\partial R_1^2} = -\frac{(1-\rho)}{p(1-p)nj}$$

Kısıtlı kaynaklardan dolayı araştırmacılar,  $\text{Var}(\delta^*)$ 'ı görece daha fazla azaltan bir seviyede bilgi toplamayı uygun görebilir. Bu durumda  $R_2^2$ 'yi arttırmak,  $R_1^2$ 'i aynı oranda arttırdığımızda meydana gelecek azalmaya kıyasla, varyansı  $(\rho n)/(1-\rho)$  kez daha fazla azaltır (iki türevin oranından elde edilmiştir). Dolayısıyla, eş-değişkenlerin açıklayıcılık gücünü 2.seviyede arttırmaya odaklanmak daha etkili bir stratejidir.

Örneğin, iki seviyeli SKD örneği için  $R_2^2$ 'yi 0,40'tan 0,50'ye yükseltmek (0,10'luk bir artış), varyansı 0,01126'dan 0,00974'e düşürür (0,00152'lik bir düşüş) ve bu da karşılığında güç oranını %46,3'ten %51,9'a yükseltir. Ancak  $R_1^2$ 'yi 0,30'dan 0,40'a yükseltmek (0,10'luk bir artış), sınıra yakın bir şekilde varyansı 0,01126'dan 0,011136'ya düşürür (0,000124'lük bir düşüş) ve bu da karşılığında güç oranını %46,3'ten %46,7'ye yükseltir. Varyans azalmalarının oranı  $(\rho n)/(1-\rho)$  formülü kullanıldığında elde edilecek orandır yani 12,26'dır. Bu da  $R_2^2$ 'yi 0,10 arttırmanın,  $R_1^2$ 'yi aynı oranda arttırdığımızda meydana gelecek azalmaya kıyasla, varyansı 12,26 kez daha fazla azaltacağı anlamına gelmektedir.

### **Üç Seviyeli SKD'ler**

#### *Koşulsuz Değişkenliği Kestirmek için Boş Model*

Koşulsuz model, aşağıda tanımlandığı gibi, tam modelden elde edilen parametrelerin yanı sıra standartlaştırılmış varyans parametrelerini hesaplamak için kullanılacak olan  $\sigma^2$ ,  $\tau_2^2$  ve  $\tau_3^2$  varyans parametrelerini elde etmek için kullanılabilir.

HLM formülü:

$$\text{Seviye 1: } Y_{ijk} = \beta_{0jk} + r_{ijk}$$

$$\text{Seviye 2: } \beta_{0jk} = \gamma_{00k} + \mu_{0jk}$$

$$\text{Seviye 3: } \gamma_{00k} = \xi_{000} + \zeta_{00k}$$

Karma model formülü:

$$Y_{ijk} = \xi_{000} + \zeta_{00k} + \mu_{0jk} + r_{ijk}$$

Burada  $r_{ijk}$ ,  $\mu_{0jk}$ , ve  $\zeta_{00k}$ , 1.seviye, 2.seviye ve 3.seviye artıklar olup, sırasıyla  $r_{ijk} \sim N(0, \sigma^2)$ ,  $\mu_{0jk} \sim N(0, \tau_2^2)$ , ve  $\zeta_{00k} \sim N(0, \tau_3^2)$  olarak normal dağılımı takip eder. Dolayısıyla  $\sigma^2$ ,  $\tau_2^2$  ve  $\tau_3^2$  sırasıyla 1.seviye, 2.seviye ve 3.seviye birimleri arasındaki çıktıda bulunan değişkenlerdir.  $\beta_{0jk}$ , 1.seviye arakesit (bu durumda  $j$  alt kümesindeki ve  $k$  kümesindeki deneklerin ortalaması);  $\gamma_{00k}$ , 2.seviye arakesit (bu durumda  $k$  kümesindeki bütün alt kümelerdeki deneklerin ortalaması); ve  $\xi_{000}$  ise 3.seviye arakesittir (bu durumda tüm kümelerdeki bütün alt kümelerin bütün deneklerinin ortalamasıdır – genel ortalama).

#### *İşlem Etkisini ve Koşullu Değişkenliği Kestirmek için Tam Model*

Bu model, aşağıda tanımlandığı gibi, koşulsuz modelden elde edilen parametrelerin yanı sıra standartlaştırılmış varyans parametrelerini de hesaplamak için kullanılan  $\sigma_{|X}^2$ ,  $\tau_{2|W}^2$  ve  $\tau_{3|V}^2$  varyans parametrelerini elde etmek için kullanılabilir.

HLM formülü:

$$\text{Seviye 1: } Y_{ijk} = \beta_{0jk} + \beta_{1jk}X_{ijk} + r_{ijk}$$

$$\text{Seviye 2: } \beta_{0jk} = \gamma_{00k} + \gamma_{01k}W_{jk} + \mu_{0jk}$$

$$\beta_{1jk} = \gamma_{10k}$$

$$\text{Seviye 3: } \gamma_{00k} = \xi_{000} + \delta T_k + \xi_{001}V_k + \zeta_{00k}$$

$$\gamma_{01k} = \xi_{010}$$

$$\gamma_{10k} = \xi_{100}$$

Karma model formülü:

$$Y_{ijk} = \xi_{000} + \delta T_k + \xi_{001}V_k + \xi_{010}W_{jk} + \xi_{100}X_{ijk} + \zeta_{00k} + \mu_{0jk} + r_{ijk}$$

Burada  $r_{ijk}$ ,  $\mu_{0jk}$ , ve  $\zeta_{00k}$ , sırasıyla  $r_{ijk} \sim N(0, \sigma_{|X}^2)$ ,  $\mu_{0jk} \sim N(0, \tau_{2|W}^2)$ , ve  $\zeta_{00k} \sim N(0, \tau_{3|V}^2)$  olarak normal dağılımı takip eden koşullu artıklardır. Bu nedenle  $\sigma_{|X}^2$ ,  $\tau_{2|W}^2$  ve  $\tau_{3|V}^2$  sırasıyla 1.seviye, 2.seviye ve 3.seviyedeki artık varyanslardır ve tam model tarafından açıklanmazlar.  $Y_{ijk}$ ,  $k$  kümesinde bulunan  $j$  alt kümesindeki  $i$  sayıda denek için 1.seviye ilgili çıktıdır;  $X_{ijk}$ ,  $k$  kümesinde bulunan  $j$  alt kümesindeki  $i$  sayıda birey için 1.seviye eş-değişkendir;  $W_{jk}$ ,  $j$  kümesinde bulunan  $j$  alt kümesi için 2.seviye eş-değişkendir;  $T_k$ , işlem koşulu ( $k$  kümesi işleme atandıysa 1, atanmadıysa 0) ve  $V_k$  ise 3.seviye eş-değişkendir.  $\beta_{0jk}$ ,  $\gamma_{00k}$ , ve  $\xi_{000}$  sırasıyla 1.seviye, 2.seviye ve 3.seviye ara kesitlerdir.  $\delta$  işlem etkisi;  $\beta_{1jk}$  veya  $\gamma_{10k}$  veya  $\xi_{100}$  1.seviye eş-değişken olan  $X_{ijk}$  için regresyon ağırlığı;  $\gamma_{01k}$  veya  $\xi_{010}$  2.seviye eş-değişken olan  $W_{jk}$  için regresyon ağırlığı;  $W_{jk}$ , ve  $\xi_{001}$  ise 3.seviye eş-değişken olan  $V_k$  için regresyon ağırlığıdır.

İki seviyeli SKD durumuna benzer olarak, standartlaştırılmış varyans parametreleri koşulsuz ve tam modellerden elde edilen standart hale getirilmemiş varyans parametrelerine dayanarak hesaplanabilir.  $\rho_2 = \tau_2^2 / (\tau_3^2 + \tau_2^2 + \sigma^2)$  olup 2.seviye birimler arasındaki çıktıda varyans oranını;  $\rho_3 = \tau_3^2 / (\tau_3^2 + \tau_2^2 + \sigma^2)$  ise 3.seviye birimler arasındaki çıktıda varyans oranını temsil etmektedir.  $R_1^2 = 1 - \sigma_{|X}^2 / \sigma^2$  olup 1.seviye eş-değişkenler tarafından açıklanan çıktıdaki varyans oranı;  $R_2^2 = 1 - \tau_{2|W}^2 / \tau_2^2$  olup 2.seviye eş-değişkenler tarafından açıklanan çıktıdaki varyans oranını ve  $R_3^2 = 1 - \tau_{3|V}^2 / \tau_3^2$  ise 3.seviye eş-değişkenler tarafından açıklanan çıktıdaki varyans oranını temsil etmektedir. İşlem etkisi Cohen'in  $d$ 'si formunda  $\delta^* = \delta / \sqrt{\tau_3^2 + \tau_2^2 + \sigma^2}$  şeklinde standart hale getirilebilir.

#### *Dengeli Örneklem Büyüklüğü ve Homojen Varyans Kapsamında Standart Hata Formülü*

Örneklem büyüklüğünün dengeli olduğunu, diğer bir ifadeyle 2.seviye birim başına  $n$  sayıda 1.seviye birim, 3.seviye birim başına  $J$  sayıda 2.seviye birimin olduğunu varsayarsak ve ayrıca her bir 2.seviye ve 3.seviye birim içerisindeki varyansın, 2.seviyenin  $JK$  sayıdaki birimlerinde ve 3.seviyenin  $K$  sayıdaki birimlerinde aynı olduğunu varsayarsak, standart hale getirilmiş standart hata şu formu alır:

$$\text{Var}(\delta^*) = \frac{\rho_3(1 - R_3^2)}{p(1 - p)K} + \frac{\rho_2(1 - R_2^2)}{p(1 - p)JK} + \frac{(1 - \rho_2 - \rho_3)(1 - R_1^2)}{p(1 - p)nJK}$$

İki seviyeli SKD'ye benzer şekilde işlem etkisinin standart hatası  $E(\delta^*) = \sqrt{\text{Var}(\delta^*)}$ 'dir. Eğer  $\delta^*$  ve  $SE(\delta^*)$ 'yi biliyorsak, hangi çıkarımların yapıldığına dayanarak  $t$  istatistikleri ve dolayısıyla istatistiksel güç de hesaplanabilir.  $\delta^*/SE(\delta^*)$ ,  $g_3$ 'ün 3.seviyede eklenen eş-değişkenlerin sayısını

temsil etmekte olup,  $K - g_3 - 2$  serbestlik derecesi ile  $t$  dağılımını takip eder (Dong & Maynard, 2013, syf. 52). İstatistiksel güç iki seviyeli SKD'lerde olduğu gibi hesaplanabilir.

### *İşlem Etkisi ve Varyans Bileşenlerinin Kestirimi ve Standartlaştırılması*

İki seviyeli SKD durumuna benzer şekilde, eğitim ortamlarını dikkate alarak, 100 okuldaki 300 sınıfta 6000 öğrencisi olan CRT3 isimli basit bir üç seviyeli SKD verisi üretilmiştir (sınıf başına 20 öğrenci ve okul başına 3 sınıf). Veri yedi değişkeni kapsamaktadır: okul numarası (schid), sınıf numarası (clsid), bir 1.seviye çıktı değişkeni (outcome), bir 3.seviye işlem değişkeni (treatment), bir 1.seviye eş-değişken (covx), bir 2.seviye eş-değişken (covw), ve bir 3.seviye eş-değişken (covv). Üretilen verinin ilk birkaç satırı aşağıda verilmiştir. Her bir okul ve sınıf kendine özgü bir numaraya sahiptir (schid). Okullar işlem koşullarına atandığından, aynı okul ve sınıf numaraları işlem değişkeni için aynı değerlere sahip olacaktır (treatment). 1.seviye (öğrenciler), 2.seviye (sınıflar) ve 3.seviye (okullar) eş-değişkenler (covx, covw ve covv), standart normal dağılımı takip eder ve çıktı (outcome), bu değişkenlerin rastgele hata eklenmiş bazı 1.seviye, 2.seviye ve 3.seviyelerinin doğrusal bir işlevidir (Bkz. Ek A'daki veri üretim mekanizmaları).

```
head(CRT3)
##      schid clsid treatment      outcome      covx      covw      covv
## 1         1         1         0  3.0263592  0.5622673 -0.3756029  0.2533185
## 2         1         1         0  1.7124732 -0.0974125 -0.3756029  0.2533185
## 3         1         1         0  1.0353372  1.0164552 -0.3756029  0.2533185
## 4         1         1         0 -0.8436311 -1.1561674 -0.3756029  0.2533185
## 5         1         1         0  1.7452900  2.3208602 -0.3756029  0.2533185
## 6         1         1         0  0.6092003 -0.6035312 -0.3756029  0.2533185
```

İki seviyeli SKD durumuna benzer şekilde ilk olarak, sınıf içi korelasyon katsayılarını hesaplamak üzere koşulsuz model için varyans parametreleri kestirilmiştir. Çıktı, okul ortalamaları ( $\tau_{23}$ ), sınıf ortalamaları ( $\tau_{22}$ ) ve öğrenciler arasındaki ( $\sigma_2$ ) değişkenliği gösteren üç rastgele etki için varyansı içermektedir. Üçünün toplamı çıktı değişkeninin varyansı ile neredeyse aynıdır. Dolayısıyla, okullar ve sınıflar arasındaki varyans oranı hesaplanabilir.

```
## -- null model (unconditional model)
null.model <- lmer(outcome ~ (1 | schid) + (1 | clsid), data = CRT3)
print(VarCorr(null.model), comp = "Variance")

## Groups      Name          Variance
## clsid      (Intercept) 1.2593
## schid      (Intercept) 0.9969
## Residual                    1.6160

## -- variance parameters
tau23 <- 0.9969
tau22 <- 1.2593
sigma2 <- 1.6160

## -- intra-class correlation coefficients for level 2 and level 3
rho2 <- tau22 / (tau23 + tau22 + sigma2)
rho3 <- tau23 / (tau23 + tau22 + sigma2)
round(rho2, 2)

## [1] 0.33

round(rho3, 2)

## [1] 0.26
```

Tam model için çıktı, yine, seviye 3, seviye 2 ve seviye 1 yordayıcıları tarafından açıklananların ötesinde, bağımlı değişkendeki koşullu değişimi (okullar ( $\tau_{23v}$ ), sınıflar ( $\tau_{22w}$ ) ve öğrenciler ( $\sigma_{2x}$ ) arasındaki gösteren üç rastgele etkinin varyansını içerir. Okullar arasındaki, sınıflar arasındaki ve öğrenciler arasındaki değişkenliğin bazıları sırasıyla 3.seviye, 2.seviye ve 1.seviye değişkenler tarafından açıklandığı için, 3.seviye, 2.seviye ve 1.seviye için varyansta azaltma oranı kullanılarak her biri için ( $r_{23}$ ,  $r_{22}$  ve  $r_{21}$ )  $R^2$  değerleri hesaplanabilir.

```
## -- full model
full.model <- lmer(outcome~ treatment + covx + covw + covv +
                  (1 | schid) + (1 | clsid), data = CRT3)
print(VarCorr(full.model), comp = "Variance")

## Groups   Name                Variance
## clsid    (Intercept)         1.06824
## schid    (Intercept)         0.71853
## Residual                            1.00901

## -- variance parameters
tau23v <- 0.7185
tau22w <- 1.0682
sigma2x <- 1.0090

## -- R-squared values for level 1, level 2 and level 3
r21 <- 1 - (sigma2x / sigma2)
r22 <- 1 - (tau22w / tau22)
r23 <- 1 - (tau23v / tau23)
round(r21, 2)

## [1] 0.38

round(r22, 2)

## [1] 0.15

round(r23, 2)

## [1] 0.28

## -- treatment effect
coef(summary(full.model))["treatment",]

## Estimate Std. Error    t value
## 0.9323254 0.2124156 4.3891572

delta <- 0.9323
es <- delta / sqrt(sigma2 + tau22 + tau22)
round(es, 2)

## [1] 0.46
```

### *İstatistiksel Güç ve Gereken Minimum Örneklem Büyüklüğü Hesaplamaları*

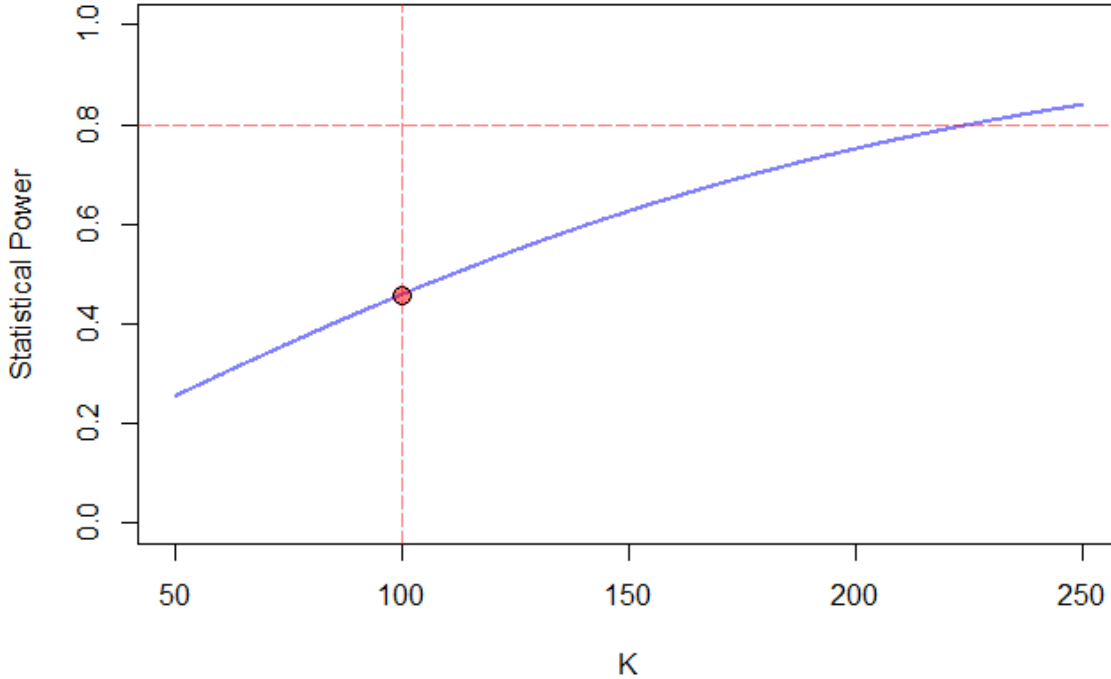
Güç analizi için varsayılan parametreler, iki seviyeli SKD için gerekli olan parametreler ile aynıdır. İki seviyeli SKD durumundan farklı olarak, sınıf başına 20 öğrenci ( $n$ ), okul başına 3 sınıf ( $J$ ) ve toplamda 100 okul ( $K$ ) bulunmaktadır. Bu durum için istatistiksel güç şu şekilde hesaplanabilir:

```
## -- power analysis
design <- power.cra3r3(es = .20, alpha = .05, two.tailed = TRUE,
                    rho2 = .33, rho3 = .26,
                    r21 = .38, r22 = .15, g3 = 1, r23 = .28,
                    p = .50, n = 20, J = 3, K = 100)

##
## Statistical power:
## -----
## 0.458
## -----
## Degrees of freedom: 97
## Standardized standard error: 0.107
## Type I error rate: 0.05
## Type II error rate: 0.542
## Two-tailed test: TRUE
```

Burada yukarıda hesaplanan parametrelere ek olarak,  $g_3$  3.seviyede eklenen eş-değişkenlerin sayısıdır. Verilerden elde edilen parametreler %45,8 oranında bir güç üretir; bu da bu deneyi daha çok sayıda tekrarladığımız zaman, aslında altta yatan evrende işlem etkisi varsa, tekrarların %45,8'i kadar istatistiksel olarak anlamlı bir işlem etkisi tespit edeceğimiz anlamına gelir. Şekil 3, referans güç oranından ne kadar uzakta olduğumuzu göstermektedir. Buna göre 200 ile 250 okul arasında bir sayıdan oluşan bir örneklemin, %80 güç oranıyla etkili bir sonuç üretme kabiliyetine sahip olduğu görülmektedir.

```
plot(design, ypar = "power", locate = TRUE, xlim = c(50, 250))
```



Şekil 3. Üç Seviyeli SKD Örneği için Okul Sayısının Bir Fonksiyonu Olarak İstatistiksel Güç



%80 güç oranı ile 0,20'lik bir etki büyüklüğünü belirlemek için ihtiyaç duyulan minimum okul sayısını hesaplamak için PowerUpR (komutları aşağıda verilmiştir) ya da PowerUp! (Şekil 4) şu şekilde kullanılabilir:

```
# -- minimum required sample size
mrss.cra3r3(power = .80, es = .20, alpha = .05, two.tailed = TRUE,
            rho2 = .33, rho3 = .26,
            r21 = .38, r22 = .15, g3 = 1, r23 = .28,
            p = .50, n = 20, J = 3)

## K = 226
```

Model 3.2: Sample Size Calculator for 3-Level Cluster Random Assignment Designs (CRA3_3r)— Treatment at Level 3		
Assumptions		Comments
MRES = MDES	0.20	Minimum Relevant Effect Size = Minimum Detectable Effect Size
Alpha Level ( $\alpha$ )	0.05	Probability of Type I error
Two-tailed or One-tailed Test?	2	
Power (1- $\beta$ )	0.80	Statistical power (1 - probability of Type II error)
Rho <sub>3</sub> (ICC <sub>3</sub> )	0.26	Proportion of variance in outcome between Level 3 units: $V_3/(V_1+V_2+V_3)$
Rho <sub>2</sub> (ICC <sub>2</sub> )	0.33	Proportion of variance between Level 2 units: $V_2/(V_1 + V_2 + V_3)$
P	0.50	Proportion of Level-3 units randomized to treatment
R <sub>1</sub> <sup>2</sup>	0.38	Proportion of variance in Level 1 outcome explained by the Level 1 covariates
R <sub>2</sub> <sup>2</sup>	0.15	Proportion of variance in Level 2 outcome explained by the Level 2 covariates
R <sub>3</sub> <sup>2</sup>	0.28	Proportion of variance in Level 3 outcome explained by the Level 3 covariates
g <sub>3</sub> *	1	Number of Level 3 covariates
n (Average Sample Size for Level 1)	20	Mean number of Level 1 units per Level 2 unit (harmonic mean recommended)
J (Average Sample Size for Level 2)	3	Mean number of Level 2 units per Level 3 unit (harmonic mean recommended)
Priori-J (Sample Size [Clusters #])	226	
Priori-T <sub>1</sub> (for desired precision)	1.97	Computed from given alpha Level, two-tailed or one-tailed test
Priori-T <sub>2</sub> (for desired precision)	0.84	Computed from given power Level
Priori-M (Multiplier)	2.81	Computed from Priori-T <sub>1</sub> and Priori-T <sub>2</sub>
M (Multiplier)	2.81	Automatically computed
K (Sample Size [# of Level 3 units])	226	Number of Level 3 clusters needed for given MDES.

**RUN**

Note: The parameters in yellow cells need to be specified. Then click "RUN" to calculate sample size.

Şekil 4. Üç Seviyeli SKD Örneği için Gereken Minimum Okul Sayısı

Bu sonuçlar, sınıf başına düşen ortalama öğrenci sayısı ( $n = 20$ ), okul başına düşen ortalama sınıf sayısı ( $J = 3$ ), sınıf içi korelasyon katsayıları ( $\rho_2 = 0,33$  and  $\rho_3 = 0,26$ ), 1.seviyedeki ( $R_1^2 = 0,38$ ), 2.seviyedeki ( $R_2^2 = 0,15$ ) ve 3.seviyedeki ( $R_3^2 = 0,28$ ) eş-değişkenlerin açıklayıcılık gücü bakımından elimizdekine benzer bir örneklem ile gerçekleştirilen güç analizi sonucu, işlem etkisinin çift yönlü hipotez testinde %80'lik bir testin gücü ve %5'lik I.Tip hata oranı ile 0,20'lik bir etki büyüklüğünü tespit etmek için en az 226 okula ihtiyacımız olduğunu önermektedir.

### Eş-Değişkenlerin Açıklayıcılık Gücü

İki seviyeli SKD'ye benzer şekilde aynı sebeplerden dolayı, üçüncü seviyede daha fazla eş-değişkenler dahil etme yoluyla eş-değişkenlerin açıklayıcılık gücünü arttırmanın daha etkili olacağını akılda tutmakta fayda var. Eğer  $Var(\delta^*)$ 'ın  $R_1^2$ ,  $R_2^2$ , veya  $R_3^2$ 'e göre birinci türevini alırsak,  $Var(\delta^*)$ 'daki değişiklik  $R_1^2$ ,  $R_2^2$ , veya  $R_3^2$ 'deki değişikliklerin aksi yönünde meydana gelmektedir. Bu,  $R_1^2$ ,  $R_2^2$ , veya  $R_3^2$ 'den herhangi biri için açıklayıcılık gücünde meydana gelen artışın,  $Var(\delta^*)$ 'ı düşüreceği anlamına gelmektedir; bu da güç oranını iyileştirmektedir.

$$\frac{\partial Var(\delta^*)}{\partial R_3^2} = -\frac{\rho_3}{p(1-p)K}$$
$$\frac{\partial Var(\delta^*)}{\partial R_2^2} = -\frac{\rho_2}{p(1-p)JK}$$
$$\frac{\partial Var(\delta^*)}{\partial R_1^2} = -\frac{(1-\rho_2-\rho_3)}{p(1-p)nJK}$$

Benzer şekilde,  $R_3^2$ 'ü arttırmak,  $R_2^2$ 'yi aynı oranda arttırdığımızda meydana gelecek azalmaya kıyasla varyansı  $(\rho_3J)/\rho_2$  kez daha fazla azaltmakta ve  $R_1^2$ 'i arttırarak meydana gelecek azalmaya kıyasla ise  $(\rho_3nJ)/(1-\rho_2-\rho_3)$  kez daha fazla azaltmaktadır. Dolayısıyla eş-değişkenlerin açıklayıcılık gücünü 3.seviyede arttırmaya odaklanmak daha etkili bir stratejidir.

Örneğin, üç seviyeli SKD örneği için  $R_3^2$ 'ü 0,28'den 0,38'e yükseltmek (0,10'luk bir artış), varyansı 0,011398'den 0,010357'e düşürür (0,00104'lik bir düşüş) ve bu da karşılığında güç oranını %45,8'den %49,4'e yükseltir. Benzer şekilde  $R_2^2$ 'yi 0,15'ten 0,25'e yükseltmek (0,10'luk bir artış), varyansı 0,011398'ten 0,010957'ye düşürür ve bu da karşılığında güç oranını %45,8'den %47,3'e yükseltir. Varyans azalmalarının oranı  $(\rho_3J)/\rho_2$  formülü kullanıldığında elde edilecek oran olup sonuç 2,36'dır. Bu da  $R_3^2$ 'ü 0,10 arttırmanın,  $R_2^2$ 'yi aynı oranda arttırdığımızda meydana gelecek varyans düşüşüne kıyasla, varyansı 2,89 kez daha fazla azalttığı anlamına gelmektedir. Ancak  $R_1^2$ 'i 0,48'den 0,58'e yükseltmek (0,10'luk artış), sınıra yakın bir şekilde varyansı 0,011398'den 0,011370'e düşürür ve bu da karşılığında güç oranını %45,8'den %45,9'a yükseltir. Varyans azalmalarının oranı  $(\rho_3nJ)/(1-\rho_2-\rho_3)$  formülü kullanıldığında elde edilecek oran olup sonuç 38'dir. Bu da  $R_3^2$ 'ü 0,10 arttırmanın,  $R_1^2$ 'i aynı oranda arttırdığımızda meydana gelecek varyans düşüşüne kıyasla 38 kez daha fazla azalttığı anlamına gelmektedir.

### SONUÇ ve TARTIŞMA

Bu bilgilendirici çalışmada, iki ve üç seviyeli SKD'lerin nasıl analiz edileceği ve planlanacağı gösterilmiştir. Daha sonra istatistiksel güç analizi işlemlerinde kullanılmak üzere varyans bileşenleri kestirilmiş ve istatistiksel modeller sunulmuştur. Güç analizi programlarının çoğu standartlaştırılmış varyans parametrelerinin belirtilmesini gerektirmektedir. Ayrıca varyans parametrelerinin sınıf içi korelasyon katsayıları ve  $R^2$  değerlerine göre nasıl standartlaştırılacağı da gösterilmiştir. Bu rehber, araştırmacıların iki ve üç seviyeli SKD'leri daha hassas bir şekilde planlama çabalarında yardımcı

olacak ve böylece değerlendiricilere, paydaşlara ve politika yapıcılara güvenilir sonuçlar sağlayacaktır.

İki seviyeli ve üç seviyeli SKD'ler için istatistiksel güç hesaplamaları, girdi olarak standartlaştırılmış parametrelere izin veren herhangi bir yazılım programında yapılabilir (örneğin; Optimal Design Plus, PowerUpR ve PowerUp!). PowerUp! ve PowerUpR'da yapılan örneklem büyüklüğünün en az kaç olması gerektiğine (MRSS) ilişkin hesaplamalardan elde edilen sonuçlar, iki seviyeli SKD'ler için sadece bir parametre değiştirilmesiyle elde edilen dokuz adet birbirinden çok az farklı olan desenlerde (Tablo B1'de D1-D9) kıyaslanabilir. Aynı işlem, üç seviyeli SKD'ler için de tekrarlanabilir (Tablo B2'de D1-D12). Bu sonuçlar, her iki yazılımda yapılan MRSS hesaplamalarının oldukça benzer olduğuna ve iki farklı platformdaki sonuçların yuvarlama farkının bir sonucu olarak nadiren bir birim farklılık gösterdiğine işaret etmektedir.

Bu çalışmada eş-değişkenlerin açıklayıcılığının güç oranı üzerindeki etkisine dair yorumda bulunulmuş, daha yüksek seviyeli birimlerle ilgili daha fazla bilgi toplamanın ve bunları eş-değişkenler olarak istatistiksel modellere dahil etmenin güç oranını önemli ölçüde arttırdığı gösterilmiştir. Bireysel düzeyde eklenen eş-değişkenler ise güç oranını sınıra yakın olarak geliştirmektedir. Dolayısıyla, daha fazla küme örnekleme konusunda maddi ve uygulama konusunda bazı zorluklar varsa alternatif olarak, eş-değişkenlerin açıklayıcılık gücünü geliştirmeye odaklanılabilir.

Bazı kümeler ve içindeki bireyler, bir uygulamaya başından sonuna kadar katılmayı reddedebilir veya uygulamayı yarıda bırakabilirler; bu da güç oranının azalmasına neden olabilir. Uygulamaya katılmama ve uygulamayı bırakma oranları da önceden yapılan araştırmadan elde edilebilir ve böylelikle gereken en küçük örneklem büyüklüğü hesaplamaları buna uygun şekilde ayarlanabilir. Mevcut verileri analiz ederken veya sonuçları bildirirken, uygulamaya katılmama ve uygulamayı bırakma oranlarını belgelemek ayrıca araştırmacıların SKD'leri daha net bir şekilde tasarlamalarına yardımcı olacaktır. Akılda bulundurulması gereken bir diğer nokta ise örneğin eğitim bağlamında, örneklem büyüklüğünde düzeltmeler yapmak adına uygulamaya katılmama ve uygulamayı bırakma durumları için fazladan okullar örnekleme alınabilirken, aynı okullardan fazladan öğrenci seçilememektedir.

Bu rehberin bazı sınırlılıkları bulunmaktadır. Her ne kadar SKD'ler için varyans parametrelerinin nasıl kestirileceği gösterilmiş olsa da bir araştırmacıyı meşgul edecek, uygulamaya dair başka sorunlar da olabilir. Örneğin, kayıp veri olabilir, doğrusallığın uç değeri ya da tahmininin tutmaması gibi sorunlar bulunabilir. Eğer araştırmacılar genelleştirilebilir büyük ölçekli SKD'ler planlamak istiyor ve benzer büyük ölçekli veri setlerine erişebiliyorlarsa, ağırlıklandırma yapmaya ihtiyaç duyulabilirler. Bu tür konular, daha kapsamlı bir şekilde ele alınmayı gerektirmekte olup, bu sorunlar bu rehberin kapsamı dışında tutulmuştur.

## KAYNAKLAR

- Bland J. M. (2004). Cluster randomized trials in the medical literature: two bibliometric surveys. *BMC Medical Research Methodology*, 4(21). DOI: <https://doi.org/10.1186/1471-2288-4-21>
- Bloom, H. S. (1995). Minimum detectable effects a simple way to report the statistical power of experimental designs. *Evaluation Review*, 19(5), 547-556. DOI: <https://doi.org/10.1177/0193841X9501900504>
- Bloom, H. S. (2006). The core analytics of randomized experiments for social research. MDRC Working Papers on Research Methodology. New York, NY: MDRC. Retrieved from DOI: [https://www.mdrc.org/sites/default/files/full\\_533.pdf](https://www.mdrc.org/sites/default/files/full_533.pdf).
- Bloom, H. S., Bos, J. M., & Lee, S. W. (1999). Using cluster random assignment to measure program impacts statistical implications for the evaluation of education programs. *Evaluation Review*, 23(4), 445-469. DOI: <https://doi.org/10.1177%2F0193841X9902300405>

- Bulus, M., Dong, N., Kelcey, B., & Spybrook, J. (2019). PowerUpR: Power Analysis Tools for Multilevel Randomized Experiments. R package version 1.0.4. DOI: <https://CRAN.R-project.org/package=PowerUpR>
- Cameron, A. C., & Miller, d. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50, 317-372. DOI: <https://doi.org/10.3368/jhr.50.2.317>
- Dong, N., & Maynard, R. (2013). *PowerUp!*: A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasi-experimental Design Studies. *Journal of Research on Educational Effectiveness*, 6(1), 24-67. DOI: <https://doi.org/10.1080/19345747.2012.673143>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Hayes, R. J. & Moulton, L. H. (2017). *Cluster Randomized Trials* (2nd ed.). New York, NY: Chapman and Hall/CRC Press. DOI: <https://doi.org/10.4324/9781315370286>
- Hedberg, E. C. (2016). Academic and behavioral design parameters for cluster randomized trials in kindergarten: an analysis of the Early Childhood Longitudinal Study 2011 Kindergarten Cohort (ECLS-K 2011). *Evaluation Review*, 40(4), 279-313. DOI: <https://doi.org/10.1177/0193841X16655657>
- Hedberg, E. C., & Hedges, L. V. (2014). Reference values of within-district intraclass correlations of academic achievement by district characteristics: Results from a meta-analysis of district-specific values. *Evaluation Review*, 38(6), 546-582. DOI: <https://doi.org/10.1177/0193841X14554212>
- Hedges, L. V., & Hedberg, E. C. (2013). Intraclass correlations and covariate outcome correlations for planning two-and three-level cluster-randomized experiments in education. *Evaluation Review*, 37(6), 445-489. DOI: <https://doi.org/10.1177/0193841X14529126>
- Hedges, L. V., & Rhoads, C. (2010). *Statistical Power Analysis in Education Research* (NCSER 2010-3006). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from <https://files.eric.ed.gov/fulltext/ED509387.pdf>
- Konstantopoulos, S. (2009a). Using power tables to compute statistical power in multilevel experimental designs. *Practical Assessment, Research & Evaluation*, 14(10).
- Konstantopoulos, S. (2009b). Incorporating Cost in Power Analysis for Three-Level Cluster-Randomized Designs. *Evaluation Review*, 33(4), 335-357. DOI: <https://doi.org/10.1177/0193841X09337991>
- Moerbeek, M., & Safarkhani, M. (2018). The design of cluster randomized trials with random cross-classifications. *Journal of Educational and Behavioral Statistics*, 43(2), 159-181. DOI: <https://doi.org/10.3102/1076998617730303>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing [Computer software]. Vienna, Austria. Retrieved from <https://www.R-project.org>.
- Spybrook, J., Shi, R., & Kelcey, B. (2016). Progress in the past decade: an examination of the precision of cluster randomized trials funded by the U.S. Institute of Education Sciences. *International Journal of Research & Method in Education*, 39(3), 255-267. DOI: <https://doi.org/10.1080/1743727X.2016.1150454>
- Spybrook, J., Westine, C. D., & Taylor, J. A. (2016). Design parameters for impact research in science education: A multistate analysis. *AERA Open*, 2(1). DOI: <https://doi.org/10.1177/2332858415625975>
- Westine, C. D. (2016). Finding Efficiency in the Design of Large Multisite Evaluations: Estimating Variances for Science Achievement Studies. *American Journal of Evaluation*, 37(3), 311-325. DOI: <https://doi.org/10.1177/1098214015624014>

- Westine, C. D., Spybrook, J., & Taylor, J. A. (2013). An empirical investigation of variance design parameters for planning cluster-randomized trials of science achievement. *Evaluation Review*, 37(6), 490-519. DOI: <https://doi.org/10.1177/0193841X14531584>
- Zopluoglu, C. (2012). A cross-national comparison of intra-class correlation coefficient in educational achievement outcomes. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 3(1), 242-278.

**Ek A**

**Veri Üretim Süreci**

**İki Seviyeli SKD için Veri Üretim Modeli**

İki seviyeli SKD üretmek için kullanılan istatistiksel model, ana metinde tanımlanan istatistiksel modelle aynıdır. Burada yalnızca karma model formülüne yer verilmiştir.

$$Y_{ij} = \gamma_{00} + \delta T_j + \gamma_{01} W_j + \gamma_{10} X_{ij} + \mu_{0j} + r_{ij}$$

Bu formüldeki parametrelere metin içerisinde yer verilmiştir. Okul başına 20 öğrenci ( $n$ ), toplamda 100 okul ( $J$ ) şeklinde düşünülerek üretilen veride aşağıdaki parametre değerleri kullanılmıştır.

$$\begin{aligned}\gamma_{00} &= 0 \\ \delta &= 1 \\ T_j &\sim \text{BERN}(0.50) \\ \gamma_{01} &= 0.50 \\ W_j &\sim N(0,1) \\ \gamma_{10} &= 1 \\ X_{ij} &\sim N(0,1) \\ \mu_{0j} &\sim N(0,1) \\ r_{ij} &\sim N(0,1)\end{aligned}$$

```
set.seed(123) # for replication
delta <- 1
js <- 100
ns <- rep(20, js)
id <- as.factor(rep(1:js, ns))
tj <- rep(rbinom(js, 1, .50), ns)
wj <- rep(rnorm(js), ns)
uj <- rep(rnorm(js), ns)
xij <- rnorm(sum(ns))
rij <- rnorm(sum(ns))
yij <- delta * tj + 0.50 * wj + xij + uj + rij

CRT2 <- data.frame("schid" = id,
                  "treatment" = tj,
                  "outcome" = yij,
                  "covx" = xij,
                  "covw" = wj)
```

**Üç Seviyeli SKD için Veri Üretim Modeli**

Üç seviyeli SKD için karma model formülü şu şekildedir:

$$Y_{ijk} = \xi_{000} + \delta T_k + \xi_{001} V_k + \xi_{010} W_{jk} + \xi_{100} X_{ijk} + \zeta_{00k} + \mu_{0jk} + r_{ijk}$$

Burada yer alan parametreler metin içerisinde açıklanmıştır. Sınıf başına 20 öğrenci ( $n$ ), okul başına 3 sınıf ( $J$ ), toplamda 100 okul ( $K$ ) şeklinde düşünülerek üretilen veri için aşağıdaki parametre değerleri kullanılmıştır.

$$\begin{aligned}\xi_{000} &= 0 \\ \delta &= 1 \\ T_k &\sim \text{BERN}(0.50) \\ \xi_{001} &= 0.25 \\ V_k &= N(0,1) \\ \xi_{010} &= 0.50\end{aligned}$$

$$\begin{aligned}W_{jk} &\sim N(0,1) \\ \xi_{100} &= 0.75 \\ X_{ijk} &\sim N(0,1) \\ \varsigma_{00k} &\sim N(0,1) \\ \mu_{0jk} &\sim N(0,1) \\ r_{ijk} &\sim N(0,1)\end{aligned}$$

```
set.seed(123) # for replication
delta <- 1
ks <- 100
js <- rep(3, ks)
ns <- rep(20, sum(js))

id3 <- as.factor(rep(rep(1:ks, js), ns))
id2 <- as.factor(rep(rep(1:sum(js), ns)))

tk <- rep(rep(rbinom(ks, 1, .50), js), ns)
vk <- rep(rep(rnorm(ks), js), ns)
sk <- rep(rep(rnorm(ks), js), ns)
wjk <- rep(rep(rnorm(sum(js)), ns))
ujk <- rep(rep(rnorm(sum(js)), ns))
xijk <- rnorm(sum(ns))
rijk <- rnorm(sum(ns))
yijk <- delta * tk + 0.25 * vk + 0.50 * wjk + 0.75 * xijk + sk + ujk +
rijk

CRT3 <- data.frame("schid" = id3,
                  "clsid" = id2,
                  "treatment" = tk,
                  "outcome" = yijk,
                  "covx" = xijk,
                  "covw" = wjk,
                  "covv" = vk)
```

**Ek B**

**PowerUpR ve PowerUp! Karşılaştırmaları**

Tablo B1

İki seviyeli SKD'ler için Karşılaştırma

Varsayımlar	Base	D1	D2	D3	D4	D5	D6	D7	D8	D9
MRES = MDES	0.20	<b>0.40</b>	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
Alpha Level ( $\alpha$ )	0.05	0.05	<b>0.01</b>	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Two-tailed or One-tailed Test?	2	2	2	<b>1</b>	2	2	2	2	2	2
Power (1- $\beta$ )	0.80	0.80	0.80	0.80	<b>0.20</b>	0.80	0.80	0.80	0.80	0.80
Rho (ICC)	0.40	0.40	0.40	0.40	0.40	<b>0.20</b>	0.40	0.40	0.40	0.40
n (Average Cluster Size)	20	20	20	20	20	20	<b>10</b>	20	20	20
P	0.50	0.50	0.50	0.50	0.50	0.50	0.50	<b>0.30</b>	0.50	0.50
R <sub>1</sub> <sup>2</sup>	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	<b>0.20</b>	0.50
R <sub>2</sub> <sup>2</sup>	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	<b>0.50</b>
J (Sample Size [# of Level 2 units]) in <i>PowerUp!</i>	<b>234</b>	<b>60</b>	<b>348</b>	<b>184</b>	<b>41</b>	<b>128</b>	<b>246</b>	<b>239</b>	<b>241</b>	<b>171</b>
J (Sample Size [# of Level 2 units]) in <i>PowerUpR</i>	233	60	348	184	41	128	245	238	241	171

Not: g (2. Seviyede eklenen eş-değişkenlerin sayısı), dokuz desenin hepsi için 1'e sabitlenmiştir.



Tablo B2

Üç seviyeli SKD'ler için Karşılaştırma

Varsayımlar	Base	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
MRES = MDES	0.20	<b>0.40</b>	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
Alpha Level ( $\alpha$ )	0.05	0.05	<b>0.01</b>	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Two-tailed or One-tailed Test?	2	2	2	<b>1</b>	2	2	2	2	2	2	2	2	2
Power (1- $\beta$ )	0.80	0.80	0.80	0.80	<b>0.20</b>	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
Rho <sub>3</sub> (ICC <sub>3</sub> )	0.30	0.30	0.30	0.30	0.30	<b>0.15</b>	0.30	0.30	0.30	0.30	0.30	0.30	0.30
Rho <sub>2</sub> (ICC <sub>2</sub> )	0.30	0.30	0.30	0.30	0.30	0.30	<b>0.10</b>	0.30	0.30	0.30	0.30	0.30	0.30
P	0.50	0.50	0.50	0.50	0.50	0.50	0.50	<b>0.30</b>	0.50	0.50	0.50	0.50	0.50
R <sub>1</sub> <sup>2</sup>	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	<b>0.30</b>	0.50	0.50	0.50	0.50
R <sub>2</sub> <sup>2</sup>	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	<b>0.40</b>	0.50	0.50	0.50
R <sub>3</sub> <sup>2</sup>	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	<b>0.70</b>	0.50	0.50
n (Average Sample Size for Level 1)	20	20	20	20	20	20	20	20	20	20	20	<b>10</b>	20
J (Average Sample Size for Level 2)	2	2	2	2	2	2	2	2	2	2	2	2	<b>3</b>
K (Sample Size [# of Level 3 units]) in <b>PowerUp!</b>	<b>183</b>	<b>47</b>	<b>272</b>	<b>144</b>	<b>33</b>	<b>125</b>	<b>145</b>	<b>217</b>	<b>184</b>	<b>194</b>	<b>136</b>	<b>187</b>	<b>162</b>
K (Sample Size [# of Level 3 units]) in <b>PowerUpR</b>	<b>183</b>	<b>47</b>	<b>272</b>	<b>144</b>	<b>33</b>	<b>125</b>	<b>145</b>	<b>217</b>	<b>184</b>	<b>194</b>	<b>135</b>	<b>186</b>	<b>162</b>

Not:  $g_3$  (3. Seviyede eklenen eş-değişkenlerin sayısı), 12 desenin hepsi için 1'e sabitlenmiştir.