Çok Düzeyli Modeller: Sürekli Değişken ile İki Düzeyli Model Örneği ve R Programı ile Analizi

Burak AYDIN[1]

Geliş Tarihi: 28.05.2016 *Kabul Tarihi: 16.09.2016*

Öz

Bu çalışmada iki düzeyli regresyon modelleri teori, önemli noktalar ve uygulama olmak üzere üç başlık altında tanıtılmıştır. Araştırmanın amacı çok düzeyli modellerin kullanılabilirliğini arttırmaktır. Makalenin giriş kısmında yerli ve yabancı alanyazına değinilmiştir. Teori kısmının açık, anlaşılır ve kolay takip edilebilir olması için çaba gösterilmiştir. Çok düzeyli modellerin varsayımları, bu modellerde bağımsız değişkenlerin merkezileştirilmesi ve hipotez testi konuları ayrı başlıklar altında verilmiştir. Uygulama verisi üzerinde yapılan analizler açıkça raporlanmış anlaşılabilirliği arttırmak için grafikler ile desteklenmiştir. Yapılan bütün analizler için kullanılan R betiği okuyucuların kullanımı için ekler kısmında sunulmuştur.

Anahtar Kelimeler: İki düzeyli model ve R, Çok düzeyli regresyon, hiyerarşik model

[1] Öğr. Gör. Dr., Recep Tayyip Erdoğan Üniversitesi, Eğitim Bilimleri Bölümü, Ölçme ve Değerlendirme Anabilim Dalı, Rize, Türkiye. burak.aydin@erdogan.edu.tr

Multilevel Models: The Two Level Model With Continuous Variables and an Empirical Example Using R

Submitted by 28.05.2016 *Accepted by 16.09.2016*

Abstract

This article is crafted to provide basic theory, definition of important issues and an application of a two level regression model. It aims to enrich social scientists' methodological knowledge and equip them with appropriate theories and tools to deduce defensible inferences they draw from statistical analyses. The basic theory with Kuehl (2000), Snijders and Bosker (2012), and Swaminathan and Rogers (2008) were introduced. An illustrative example was included and the R syntax was provided (see supplementary).

***Keywords**:*Multilevel regression, Hierarchical Models, R,

Giriş

Kreft ve de Leeuw (1998) çok düzeyli modellerin tarihçesini özetlemişlerdir. Yazarlar bu modellerin ondokuzuncu yüzyılda geliştirilmeye başlandığını ve eğitim alanında yapılan araştırmalarda ilk olarak Cronbach ve Webb (1975) ve Burstein,Linn ve Capell (1978) tarafından kullanıldığını belirtmişlerdir.Çok düzeyli modellerin öneminin Robinson'un (1950) bağlamsal etkileri (contextual effects) tanımlaması ile ön plana çıktığı söylenebilir.Bir değişkenin etkisi incelendiğinde ele alınan birey ya da nesnenin bağlı olduğu grubun da etkisinde kalmasına bağlamsal etkiler denir. Ayrıca Robinson (1950) ekolojik yanılma (ecological fallacy) kavramını ortaya atarak bireylerin davranışları arasındaki ilişkileri açıklamak için bu bireylerin ait oldukları gruplara dair verileri kullanmanın doğru olmayacağını ileri sürmüştür. Diğer bir deyişle yazar gruplar arası korelasyonun bireyler arası korelasyon yerine kullanılamayacağını göstermiştir. Çok düzeyli modeller hem grup verisini hem de bireysel verileri kullanarak hem grup bazında hem de bireysel anlamda istatistiksel çıkarımlar yapılmasına olanak sağlamaktadır. Ekolojik yanılma kavamına dayanan bağlamsal etkiler yaklaşımının yanında çok düzeyli modellerin dayandığı bir diğer yaklaşım da karma etkiler modelleridir (mixed effects models) (Snijders ve Bosker, 2012). Karma etkiler modelleri ANOVA ve regresyon gibi analizlerde kullanılan bir yaklaşımdır. Bu yaklaşıma göre bazı katsayılar rastgele bazıları da sabit olarak kabul edilir. ANOVA ve regresyon gibi yaygın olarak kullanılan genel doğrusal modeller (general linear models) verideki kümelenmiş (iç içe geçmiş) yapıları hesaba katmadıkları durumlarda hatalı sonuçlar sunabilir. Genel doğrusal modeller ailesinin bir diğer üyesi olan ANCOVA yöntemi ortak değişken (covariate) kullanarak gruplardan oluşan yapıların hesaba katılmasını sağlıyor olsa da cevaplayabildiği sorular çok kısıtlıdır (Roberts, 2004). Geçmiş yıllarda alanyazında hem grup yapılarını hesaba katacak hem de rastgele etkileri tahmin edebilecek modellere ihtiyaç duyulmuştur. Bu ihtiyacı karşılamak için birçok araştırmacının çalışmalarıyla günümüzde kullandığımız çok düzeyli modellerin temelleri ortaya atılmıştır (Aitkin ve Longford, 1986; De Leeuw ve Kreft, 1986; Goldstein, 1986; Longford, 1987; Raudenbush ve Bryk, 1986). Çok düzeyli modeller ANCOVA ve regresyonun birleşimine sahip ve genel doğrusal modellerin genişletilmiş bir versiyonu olarak görülebilir. Yukarıda adı geçen birçok araştırmacı tarafından eş zamanlı olarak geliştirilen projelerle ortaya çıkan çok düzeyli modeller alanyazında farklı adlarla anılmaya başlanmıştır. Halen üzerinde anlaşılan tek bir adlandırma olmamasından dolayı çok düzeyli modelleri (Goldstein, 1995; Snijders ve Bosker, 2012) kullanacak herhangi bir teori ya da uygulama çalışması için alanyazın tarama aşamasında bu modellere verilebilecek farklı isimleri bilmek yararlı olacaktır. Bu farklı isimlerden bazıları çok düzeyli doğrusal modeller (multilevel linear models), karma modeller (mixed models), rastgele regresyon katsayıları (random regression coefficient models), hiyerarşik modeller (hierarchical models ya da nested models), kovaryans bileşenler modelleri (covariance components models) ve dengelenmemiş rastgele etkiler modelleri (unbalanced models with random effects) olarak bilinir. Türkiye'deki alanyazında karşımıza çıkan adlandırmalarda da farklılıklar göze çarpmaktadır. Türkiye'deki araştırmacıların kullandığı adlandırmalar çok düzeyli modeller, çok aşamalı modeller ve hiyerarşik doğrusal modeller olarak sayılabilir. Bu çalışmada çok düzeyli modeller adlandırılması tercih edilmiştir.

Çok düzeyli modellerin kullanılmasına zemin hazırlayan durumlar iki çatı altında toplanabilir, kümeli yapılar (nested structures) ve tekrarlanan ölçümler (repeated measures). Kümeli yapılardan kasıt bireylerin gruplar içinde yer aldığı durumlardır. Sosyal bilimler alanında yapılan bir çok çalışma gruplar içinde yer alan bireylerden edinilen verileri kullanır. Sınıflar, okullar, hastahaneler, firmalar, ilçeler bu kümelere örnek verilebilir. Bireylerin kümeler içinde yer alması toplanan verilerde bir bağımlılık (dependence) oluşturur. Aynı küme içinde yer alan bireylerden toplanan veri benzerlik gösterebilir. Bu benzerliğin göz ardı edilmesi durumunda yapılan analizler hatalı çıkarımlara yol açabilir. Bunun sebebi kümelerin dikkate alınmaması durumunda hesaplanan varyansın hatalı olmasıdır. Bu duruma bir örnek verecek olursak, lise öğrencilerinin davranış bozukluklarını azaltmak üzere hazırlanmış bir müfredatın rastgele seçilen 30 lise ve her bir liseden rastgele seçilmiş 30 öğrenciye uygulandığını düşünelim. Öğrencilerin okulların içinde kümelenmiş olduğunu göz ardı etmek, seçilen her bir öğrencinin çalışma evreninden eşit olasılıkla seçildiğini varsayar. Bu varsayım hatalıdır ve aynı okulda yer alan öğrencilerin benzer oranda davranış bozukluğu gösterebilecek oluşu örneklem hatasını artıracaktır. Kish (1965) kümeli yapıların örneklem hatası üzerine olan etkilerini teori ve örnekleriyle açıklamıştır. Cornfield (1978) ise kümelerin örneklemde yer alacak bireylere özgü karakteristiklerin (idiosyncratic characteristics) eşit dağılımına olan etkisini göstermiş ve yazar küme örnekleme yönteminin kullanıldığı çalışmalarda küme etkisini göz ardı etmenin açıkça kendi kendini aldatma (self deception) olduğunu yazmıştır. Özetle, bireylerin gruplar içinde yer aldığı çalışmalarda ve özellikle kümelerin homojen olduğu durumlarda, yapılan analizler istatistiksel düzeltmeler kullanmalıdır. Çoğu zaman, kümeleri kullanarak elde edilecek ekonomik kolaylıklar istatistiksel zorlukları çözümleme için harcanacak enerjiyi telafi eder. Tekrarlanan ölçümler de aslında yapılan ölçümlerin birey içinde kümelenmesi olarak düşünülebilir. Örneğin öğrencilerin matematik yeterliliklerinde oluşan gelişmeyi anlamak için ayda bir kez olmak üzere 4 ay boyunca gözlem yapıldığında, edinilen yeterlilik puanları muhtemelen sıfırdan farklı pozitif bir korelasyona sahip olacaktır. Çünkü çoğu zaman ilk ölçümde diğer öğrencilere göre daha yüksek puan alan bir öğrencinin, takip eden ölçümlerde yine yüksek bir puan alması olasıdır. Bir diğer deyişle, aynı öğrencinin farklı ölçümlerde aldığı notlar kendi içinde benzerdir. Verilerin analizi esnasında bu ilinti dikkate alınmalıdır ve çok düzeyli modeller bu amaç için verimli bir seçenek oluşturur. Boylamsal araştırma (longitudinal study) olarak da isimlendirilebilecek bu durum çok düzeyli modeller tarafından yapılabilecek analizlerden birisidir.

Geleneksel tek düzeyli istatisik yöntemleriyle (regresyon ve ANOVA) karşılaştırıldığında çok düzeyli modeller birçok avantaj sunmaktadır. İlk olarak diyebiliriz ki çok düzeyli modeller geleneksel yöntemlerden daha etkili bir analiz imkanı sunmaktadır. Örneğin 200 farklı okuldan gelen öğrencileri içeren bir veri setindeki öğrencilerin ders çalışma saatleri ile sınav puanları arasındaki ilişkiyi incelemek isteyen bir araştırmacının farklı okullardaki ilişkiyi resmetmek için 200 farklı regresyon grafiği çizmesi gerekebilir. Bunun aksine çok düzeyli model kullanıldığında tek bir grafik üzerinden 200 farklı okula ait değişimi gösteren çizgiler çizilebilir. İkinci olarak hiyerarşik yapıya sahip verilerin analizinde çok düzeyli modellerin kullanılması istatistiksel olarak daha doğru olmaktadır. Verideki hiyerarşik yapıları göz ardı eden modeller standart hataların düşük çıkmasına ve dolayısıyla I. tür hata yapılmasına yol açmaktadır. Son olarak çok düzeyli modeller yapılabilecek istatistiksel analizler için genel bir çatı oluşturmaktadır. Çok düzeyli modelleri kullanarak ANOVA, regresyon, ANCOVA, meta-analiz ve boylamsal analizler yapabilmek mümkündür. Bunlara ek olarak hem bağımlı değişkenin hem de bağımsız değişkenin sürekli veya kategorik olduğu durumlarda analizlere katılabilmesine olanak sağlaması çok düzeyli modellerin geleneksel çok düzeyli olmayan modellere karşı daha esnek bir yapıda olduğunu göstermektedir.

Çok düzeyli (hiyerarşik) yapılara sahip verileri analiz etmek için son yıllarda birçok model geliştirilmiştir (Snijders ve Bosker, 2012). Bu yöntemlerin arasında uygulamada en çok kullanılanı çok düzeyli modeller olmuştur. Bu çok düzeyli modeller eğitim araştırmalarında da araştırmacılar tarafından çok fazla kullanılmaya başlanmıştır (Schreiber ve Griffin, 2004). Yurtdışında çok fazla tercih edilen çok düzeyli modellerin yurtdışıyla karşılaştırıldığında Türkiye'de çok daha az araştırmacı tarafından (Acar, 2013; Acar ve

Öğretmen, 2012; Atar, 2010; Atar, 2014; Can, Somer, Korkmaz, Dural ve Öğretmen, 2011; Çoker, 2009; Deniz-Başar, Özden ve Bağdatlı-Kalkan, 2013; Erol-Korkmaz, 2014; Gölbaşı-Şimşek ve Noyan, 2008; Gölbaşı-Şimşek ve Noyan, 2009; Güvendir, 2014; Kadıoğlu, ve Uzuntiryaki-Kondakcı, 2014; Karabay, Yıldırım ve Güler, 2015; Noyan ve Yıldız, 2006; Şahin, 2011; Yılmaz ve Aztekin, 2012) kullanıldığı görülmektedir. Bu çalışmalar ve kullandıkları çok düzeyli modeller hakkındaki bilgileri aşağıda kısaca sınıflandırılarak sunulmuştur.

Çoker (2009) çok düzeyli modelleri kullanarak yaptığı tez çalışmasında çok düzeyli regresyon modelleri ile çok düzeyli yapısal eşitlik modellerini karşılaştırmıştır. Bu çalışmaya benzer bir metodoloji çalışmasında da Atar (2010) TIMSS 1999 verisini kullanarak iki düzeyli hiyerarşik doğrusal model ile basit doğrusal regresyon model analizlerini karşılaştırmıştır. Çok düzeyli modellerin kullanıldığı bir diğer tez çalışmasında (Acar, 2013) 8. Sınıf öğrencilerinin Türkçe dersi başarısının bazı öğrenci ve okul özellikleri ile ilişkisi incelenmiştir.

Yılmaz ve Aztekin (2012) çalışmasında PISA 2009 verileri üzerinden Türk öğrencilerinin başarılarını etkileyen faktörler iki düzeyli hiyerarşik doğrusal model kullanılarak incelenmiştir. Buna benzer bir çalışmada da Karabay ve diğerleri (2015) iki düzeyli aşamalı doğrusal modelleri kullanarak farklı yıllardaki (2003, 2006 ve 2009) PISA verileri üzerinden öğrencilerin matematik okuryazarlığının öğrenci ve okul özellikleri ile ilişkisini incelemiştir. Uluslarası sınavlardan elde edilen Türkiye verisinin analizine bir diğer örnek de Atar'ın (2014) TIMSS 2011 sınavındaki Fen Bilgisi dersi başarısına bazı öğretmen niteliklerinin etkisini incelediği çok düzeyli çalışması verilebilir. Son olarak Acar ve Öğretmen (2012) çalışmasında 2006-PISA Fen Bilimleri Testi Türkiye verilerine göre Türk öğrencilerinin performanslarının, öğrenci ve okul düzeylerine göre farklılaşıp farklılaşmadığını iki düzeyli model ile analiz etmiştir.

Noyan ve Yıldız (2006) çalışmasında bir üniversitedeki öğretim elemanlarının öğrenciler tarafından değerlendirildiği anket sonuçlarını kullanarak iki düzeyli (bölümler ve bölümler içindeki öğrencilerden) hiyerarşik doğrusal regresyon analizi yapmışlardır. Kadıoğlu ve Uzuntiryaki-Kondakcı (2014) çalışması ise lise öğrencilerinin öğrenme stratejileri ve hedef yönelimleri arasındaki ilişkiyi iki düzeyli (öğrenci ve sınıf düzeyleri) model ile incelemiştir.

Çok düzeyli modellerin meta-analiz ve yapısal eşitlik modelleri gibi diğer istatistik yöntemleri ile birleştirilebilmesi çok düzeyli modellerin sunduğu esneklikler arasındadır

(Çoker, 2009; Şen ve Akbaş, (baskıda)). Gölbaşı-Şimşek ve Noyan (2009) çalışmasında algılanan öğretim üyesi etkinliğinin öğrenci sadakatine etkisini araştırmak için çok aşamalı yapısal eşitlik modeli kullanılmıştır. Çok düzeyli yapısal eşitlik modelini kullanan başka bir çalışma da Can ve diğerlerinin (2011) çalışmasıdır.

Genelde eğitim araştırmalarında uygulandığı gözlemlenen çok düzeyli modellere ek olarak Erol-Korkmaz'ın (2014) bilgisayar teknolojileri ve yazılım sektöründe görev yapan çalışanlardan elde edilen anket verisinin hiyerarşik doğrusal modeller ile analizini yaptığı çalışma örnek verilebilir. Eğitim alanı dışından başka bir örnek de Şahin'in (2011) uluslararası askeri bir örgütte çalışan liderlerden (üstlerden) ve bu liderlere bağlı askerlerden (astlardan) elde edilen verilerin çok düzeyli modeller ile yaptığı analizleri verilebilir. Eğitim alanı dışındaki başka bir çalışma da Gölbaşı-Şimşek ve Noyan'ın (2008) çok aşamalı doğrulayıcı faktör analizi kullanarak, Türkiye'deki ilçeler için kullanılan gelişmişlik indeksinin doğrulanması üzerine yapılan çalışması örnek verilebilir.

Yukarıda belirtilen çalışmaların hepsi kesitsel verilerin çok düzeyli modeller ile analizlerini içerirken Deniz-Başar ve diğerlerinin (2013) çalışması bebeklerin 12, 24 ve 36 aylık dönemlerdeki gelişimlerini ölçmek için kullanılan bir ölçekten elde ettikleri verinin boylamsal analizini çok düzeyli model ile gerçekleştirdiği bir çalışmadır.

Araştırmanın Amacı

Alanyazında çok düzeyli modelleri kullanan araştırmalar incelendiğinde bu çalışmaların çoğunun son beş yıl içerisinde yapıldığı ve sayısının giderek artmakta olduğu göze çarpmaktadır. Bununla birlikte çok düzeyli modeller yerine regresyon gibi geleneksel yöntemler halen çoğu araştırmacının tercihi olmaktadır. Bunun başlıca sebeplerinden birisi Türkiye'deki araştırmacıların çoğunun verinin yapısına (tek düzeyli ya da çok düzeyli) bakmaksızın toplam puanları kullanarak çok daha kolay uygulanabilen regresyon analizlerini uygulaması gelmektedir. Diğer bir sebep de çok düzeyli modellerin gelişmiş istatistik bilgisi gerektirmesi ve bu yöntemlerin Türkiye'deki istatistik kaynaklarında yeterince açıklanmamasıdır. Bu çalışmayla genelde çok düzeyli modellerin özelde de iki düzeyli modellerin Türkiye'deki araştırmacılara tanıtılması amaçlanmaktadır. Çok düzeyli modellerin daha iyi anlaşılabilmesini sağlamak için gerçek bir çok düzeyli veri seti kullanılarak bir uygulama yapılmıştır. Bu uygulama ve yapılan analizler hakkındaki detayları takip eden bölümde yer verilmiştir. Yöntem kısmının teori, varsayımlar ve hipotez testi kısımlarında yer yer Aydın (2014) tarafından verilen çerçeve takip edilmiştir.

Yöntem

İzleyen bölümlerde, kümelerin ikinci, bireylerin ise birinci düzeyi oluşturduğu iki düzeyli modeller ele alınacaktır. Snijders ve Bosker'in (2012) notasyonunu kullanırsak, toplam küme sayısı J, herhangi bir j ($j = 1,2,\ldots,J$) kümesinin toplam birey sayısı n_j ,toplam örneklem sayısı $\mathrm{M} = \sum_{\mathrm{j}=1}^{\mathrm{J}} n_j$, kümelerin içindeki bireyler $i(i = 1,2,\ldots,n_j)$, sonuç değişkeni (outcome) Y_{ij}, birinci düzey açıklayıcı değişkenler X_{ij} ve son olarak ikinci düzey açıklayıcı değişkenler Z_j ile simgelenmiştir. Bu çalışmada iki düzeyli modellerin iki farklı alt modeli tanıtılacaktır, boş model (the null ya da empty model) ve rassal kesim-eğim modeli (random intercepts ya da intercepts and slopes as outcomes). Raudenbush ve Bryk (2002), boş modelden daha karmaşık, rassal kesim-eğim modellerinden ise daha sade olan üç farklı model daha tanıtmıştır. Bu üç ara model makalenin sadeliği adına değinilmemiştir. Ayrıca, Y, X ve Z değişkenleri birer sürekli değişken olarak düşünülmüştür.

Boş Model

Boş bir model olabilecek en sade çok düzeyli modeldir. Hiç bir açıklayıcı değişken içermez;

$$Y_{ij} = \gamma_{00} + u_{0j} + \varepsilon_{ij} \tag{1}$$

Bu modelde, γ_{00} genel ortalamayı bir diğer değişle sabit etkiyi, u_{0j} j kümesi ile bağdaşık rassal etkiyi ve ε_{ij} ise j kümesinde yer alan i numaralı bireyle bağdaşık rassal etkiyi simgeler. Rassal etkiler için yapılacak bir normal dağılım varsayımı tahminleme işleminde büyük kolaylıklar sağlar. Bu yüzden rassal etkilerinin ortalamasının 0, varyanslarının küme düzeyi için τ^2 ve birey düzeyi için σ^2 olduğu varsayılır. Ayrıca rassal hataların birbirinden bağımsız olduğu varsayılır ve bu yüzden $var(Y_{ij}) = \sigma^2 + \tau^2$ dir. Boş bir model genel ortalama için bir tahmin ve bu tahmine yönelik bir güven aralığı (confidence interval) hesaplamada kullanılabilir. Fakat çoğu zaman boş bir modelin amacı küme-içi korelasyon katsayısını (KİK, intra class correlation coefficient) yani toplam varyansın ne kadarlık kısmının kümelerin etkisi ile oluştuğunu hesaplamaktır.

Örneğin, sadece bir tane sekizinci sınıf şubesi olan okulların her birinden 10 tane sekizinci sınıf öğrencisinin ve toplamda 30 okulun rastgele seçildiği bir çalışma için öğrencilerin matematik yeteneklerinin ölçüldüğünü ve her bir öğrenciye bir matematik puanı verildiğini varsayalım. Bu 300 öğrencinin ortalama puanı boş bir modeldeki sabit etkiye tekabül eder (γ_{00}). Okul düzeyindeki puanlar o okuldaki 10 öğrencinin ortalaması olarak hesaplandığında, her okul muhtemelen genel ortalamadan daha farklı bir ortalamaya sahip olacaktır. Bu farklılık ikinci düzeyde bir varyans oluşturur. Aynı okul içerisindeki öğrencilerin aldığı farklı puanlar ise birinci düzeyde bir varyans oluşturur.

Matematiksel işlemler açısından daha sade olduğu için boş modelin teorik altyapısı olarak varyans analizi ayrışımı (ANOVA decomposition) kullanılmıştır. Her kümenin eşit sayıda birey içerdiği durumda $n_1 = n_2 = \ldots = n_j = n$, ANOVA ayrışımını kullanırsak (Kuehl, 2000);

Grup içi kareler toplamı (Sum of Squares Within) $SSW = \sum_{j=1}^{J} \sum_{i=1}^{n} (Y_{ij} - Y_{.j})^2$

Grup içi ortalama kareler (Mean Squares for Within) $MSW = \frac{SSW}{J(n-1)}$

Gruplar arası kareler toplamı (SS Between) $SSB = \sum_{j=1}^{J} \sum_{i=1}^{n} (Y_{.j} - Y_{..})^2$

Gruplar arası ortalama kareler (MSfor Between) $MSB = \frac{SSB}{J-1}$

Beklenen grup içi ortalama kareler $E(MSW) = MSW = \sigma^2$

Beklenen gruplar arası ortalama kareler $E(MSB) = \sigma^2 + n\tau^2$

Bu dengeli (balanced) tasarıda, σ^2 in tahmini , $\hat{\sigma}^2$, MSW ve τ^2 tahmini

$\hat{\tau}^2 (MSB - MSW)/n$ dir.

Genel ortalama,

$$\hat{\gamma_{00}} = \frac{1}{J} \sum_{j=1}^{J} Y_{.j} \tag{2}$$

küme-içi korelasyon katsayısı,

$$\text{KİK} = \hat{\rho} = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}^2} \tag{3}$$

formülleri ile hesaplanabilir. Bu tahminler için, belirlenecek bir I. tür hata oranı (alpha) ile güven aralıkları hesaplanabilir. $100(1-\alpha)\%$ güven aralığı;

$$\frac{SSW}{\chi^2_{\alpha/2,J(n-1)}} < \sigma^2 < \frac{SSW}{\chi^2_{(1-\alpha/2),J(n-1)}} \tag{4}$$

$$\frac{SSB(1-(F_{\alpha/2,J-1,J(n-1)}/(MSB/MSW))}{n\chi^2_{\alpha/2,(J-1)}} < \tau^2 < \frac{SSB(1-(F_{1-\alpha/2,J-1,J(n-1)}/(MSB/MSW))}{n\chi^2_{1-\alpha/2,(J-1)}} \quad (5)$$

$$var(\hat{\gamma}_{00}) = \frac{\sigma^2 + n\tau^2}{nJ} \Rightarrow \hat{\gamma}_{00} - \sqrt{\frac{MSB}{nJ}}|t_{\alpha/2,J-1}| < \gamma_{00} < \hat{\gamma}_{00} + \sqrt{\frac{MSB}{nJ}}|t_{\alpha/2,J-1}| \quad (6)$$

$$\frac{(MSB/MSW) - F_{\alpha/2,J-1,J(n-1)}}{(MSB/MSW) + (n-1)F_{\alpha/2,J-1,J(n-1)}} < \rho < \frac{(MSB/MSW) - F_{1-\alpha/2,J-1,J(n-1)}}{(MSB/MSW) + (n-1)F_{1-\alpha/2,J-1,J(n-1)}} \quad (7)$$

Snijders ve Bosker (2012) Eşitlik 3 ile hesaplanan KİK değerini şu şekilde yorumlar: KİK, kümelerden oluşan bir evrenden rastgele seçilmiş bir kümeden rastgele seçilmiş iki bireyin benzerliğidir (homogeneity). Eşitlik 3 için dengeli bir tasarıda standart hata Fisher (1958) tarafından Eşitlik 8'de verilmiştir;

$$SE(\hat{\rho}) = (1-\rho)(1+(n-1)\rho)\sqrt{\frac{2}{n(n-1)(J-1)}} \quad (8)$$

KİK değerinin popülasyon bazında sıfıra eşit olduğu boş hipotezi ($\rho = 0$) Eşitlik 9 ile sınanabilir;

$$F = \frac{MSB}{MSW} \quad (9)$$

Bu F istatistiği $J-1$ ve $J(n-1)$ serbestlik derecesine sahip bir F dağılımını takip eder. Snijders ve Bosker (2012) bu boş hipotezin reddedildiği durumda istatiksel çıkarımların (statistical inference) tek düzeyli modellere nazaran çok düzeyli modeller kullanıldığında daha geçerli olduğunu söyler. Hesaplanan KİK, tasarı etkisini (design effect) tahmin etmek için de kullanılabilir;

$$TE = 1 + (n-1)\rho \quad (10)$$

Bu istatistik, basit rassal örneklem (kümelerin olmadığı bir evrenden elde edilen örneklem) ile kıyaslandığında, örneklem varyansının (sample variance) ne oranda kümeler tarafından etkilendiğini ortaya koyar. Örneğin tasarı etkisinin 2, örneklemin ise rastgele seçilmiş 30 küme ve her kümeden rastgele seçilmiş 10 bireyden oluşan 300 kişi olduğu bir durumda, aritmetik ortalamanın standart hatası, kümelerden oluşmayan bir evrenden rastgele seçilmiş 150 (300/2) kişilik bir örneklem için hesaplanan aritmetik ortalamanın standart hatasına eşittir.

Rassal Kesim-Eğim Modeli

Yukarıda kısaca değinilen boş model hiçbir açıklayıcı değişken içermez. Rassal kesim-eğim modelleri ise her iki düzeyde açıklayıcılar ve bu açıklayıcıların rassal etkilerine modelde yer verebilir. Birinci düzeydeki açıklayıcı sayısının p, ikinci düzeydeki açıklayıcı sayısının q olduğu iki düzeyli bir model[2] Eşitlik 11 ile verilmiştir.

$$Y_{ij} = \gamma_{00} + \gamma_{01}Z_{1j} + \cdots + \gamma_{0q}Z_{qj} + \gamma_{10}(X_{1ij}) + \cdots + \gamma_{p0}(X_{pij}) + u_{.j} + \varepsilon_{ij} \tag{11}$$

Eşitlik 11 çoğu kaynakta anlaşılabilirliği arttırmak için iki parça halinde verilir;

$$\begin{aligned} &\text{Düzey 1: } Y_{ij} = \beta_{0j} + \beta_{1j}(X_{1ij}) + \ldots + \beta_{pj}(X_{pij}) + \varepsilon_{ij} \\ &\text{Düzey 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}Z_{1j} + \ldots + \gamma_{0q}Z_{qj} + u_{0j} \\ &\qquad \beta_{1} = \gamma_{10} + u_{1j} \\ &\qquad \vdots \\ &\qquad \beta_{p} = \gamma_{p0} + u_{pj} \end{aligned} \tag{12}$$

Rassal eğim-kesim modellerinin teorik altyapısı matrisler kullanarak daha kolay anlaşılabilir. Swaminathan ve Rogers (2008) notasyonunu kullanırsak;

$$Y = X\beta + e \tag{13}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}_{M1}, X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{M1} & \cdots & x_{Mp} \end{bmatrix}_{(M\times(p+1))}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{((p+1)\times M)}, e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{bmatrix}_{(M\times 1)} \tag{14}$$

$$\beta = Z\gamma + u \tag{15}$$

[2] Düzeyler arası etkileşimin (crosslevel interaction) olmadığı varsayılmıştır.

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \beta_p \end{bmatrix}_{((p+1)\times 1)}, Z = \begin{bmatrix} Z_0 \\ Z_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ Z_p \end{bmatrix}_{((p+1)\times(1+q+p))}, \gamma = \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \vdots \\ \gamma_{0q} \\ \gamma_{10} \\ \gamma_{11} \\ \vdots \\ \gamma_{pq} \end{bmatrix}_{((1+q+p)\times 1)}, u = \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ u_0 \end{bmatrix}_{((p+1)\times 1)} \quad (16)$$

Z matrisinin her bir satırı açıklayıcı değişken değerlerini içeren bir vektördür. Gözlemler baz alındığında, regresyon eşitliklerini her bir küme için ayrı ayrı yazarsak;

$$Y_j = X_j\beta_j + e_j \quad (17)$$

$$e_j \sim N(0, \sigma^2 I) \quad (18)$$

Bu eşitlikte Y_j, j kümesinde yer alan bağımlı değişken değerlerini içeren bir $n_j \times 1$ vektördür, X_j açıklayıcı değişkenleri içeren bir $n_j \times (p + 1)$ matrisdir, β_j katsayıları içeren bir $(p + 1) \times 1$ vektördür, I, $n_j \times n_j$ birim matris, e_j normal dağılımlı rassal hataları içeren bir $n_j \times 1$ vektördür. Bu rassal hataların ortalamasının bir sıfır vektörü ve varyans-kovaryans matrisinin bütün diyagonal elemanlarının σ^2 geri kalan elemanlarının ise sıfır olduğu varsayılır.

$$\beta_j = Z_j\gamma + u_j \quad (19)$$

$$u_j \sim N(0, T) \quad (20)$$

Eşitlik 19 ile verilen Z_j açıklayıcı değişkenleri içeren bir $((p + 1) \times (1 + q + p))$ matrisdir, γ sabit etkileri içeren bir $((1 + q + p) \times 1)$ vektördür, u_j ikinci düzey rassal hataları içeren bir $(p + q) \times 1$ vektördür. İkinci düzey rassal hataların sıfır ortalama ve bir $((p + 1) \times (p + 1))$ varyans-kovaryans matrisi ile normal dağıldığı varsayılır. Bu varyans-kovaryans matrisi T ile simgelenmiştir. T matrisinin diyagonal elemanları sabit etkilerin varyansını ve diyagonal olmayan elemanlarıda kovaryansları içerir.

Her kümenin eşit sayıda bireye sahip olduğu ve X matrislerinin aynı olduğu varsayıldığında, iki kademeli sıradan en küçük kareler yöntemi (two-stage ordinary least squares) Raudenbush ve Bryk (2002, s. 42-44) tarafından şu şekilde verilmiştir;

$$\hat{\beta}_j = (X_j'X_j)^{-1}X_j'Y_j \tag{21}$$

$$\hat{\gamma} = (\sum(Z_j'Z_j)^{-1}\sum Z_j'\hat{\beta}_j \tag{22}$$

$$\hat{u}_j = \hat{\beta}_j - Z_j\hat{\gamma} \tag{23}$$

$$\hat{e}_{ij} = Y_{ij} - X_{ij}Z_j\hat{\gamma} - X_{ij}\hat{u}_j \tag{24}$$

Eğer σ^2 ve τ^2 tahminleri mevcut ise iki kademeli en küçük kareler tahmin edicisinin örneklem varyansı Swaminathan ve Rogers (2008) tarafından şu şekilde verilmiştir;

$$V_j = \sigma^2(X_j'X_j)^{-1} \tag{25}$$

$$\underset{j}{\Delta} = var(u_j + e_j) = T + V_j \tag{26}$$

$$Var(\hat{\gamma}) = (\sum_{j=1}^{N} Z_j'\, Z_j)^{-1} \left[\sum_{j=1}^{N} Z_j' \underset{j}{\Delta} Z_j \right] (\sum_{j=1}^{N} Z_j'\, Z_j)^{-1} \tag{27}$$

Çok düzeyli model parametrelerinin tahmini değerleri farklı tahmin yöntemleriyle elde edilebilir (Kreft ve De Leeuw, 1998; Raudenbush ve Bryk, 2002). Sabit ve rassal etkiler eşzamanlı olarak en büyük olabilirlik (maximum likelihood) yöntemi ile ya da ayrı ayrı en küçük kareler yöntemi ile tahmin edilebilir. Bayes tahmin yöntemleri de çok düzeyli modelleri analiz etmek için kullanılabilmektedir. En büyük olabilirlik yöntemi E-M algoritması ile uygulanabilirken Bayes tahmin yöntemi de Gibbs örnekleyicisi (Gibbs Sampler) ile uygulanabilmektedir. Çok düzeyli modelleri analiz edebilmek için birçok istatistik programı kullanılabilmektedir. Bunların başında Harvey Goldstein tarafından geliştirilen MLwiN ile Bryk, Raudenbush ve Congdon (1996) tarafından geliştirilen HLM programları gelmektedir. Bu programlara ek olarak örtük değişken analizleri için geliştirilen Mplus (Muthén ve Muthén, 1998-2012) programı da çok düzeyli modellerin analizinde kullanılabilmektedir. Yukarıda belirtilen çok düzeyli istatistik programlarının yanında R, SAS ve STATA gibi genel istatistik programları da çok düzeyli modellerin analizine imkan sağlayan paket ve fonksiyonlar içermektedir. R programında çok düzeyli verileri analiz edebilmek için başlıca lmer, multilevel ve nlme paketleri, SAS programında PROC MIXED fonksiyonu ve STATA programında xtreg ve xtmixed prosedürleri kullanılabilmektedir.

Çok Düzeyli Modellerde Varsayımlar

Diğer istatistiksel modellerde olduğu gibi çok düzeyli regresyon için de modelin doğru kurulması sonuçların geçerli olması için gereklidir. Çok düzeyli bir regresyon için; önemli değişkenlerin modelin dışında bırakılması (Kim ve Frees, 2006), bağımlı bağımsız değişken arasındaki ilişkinin doğru tanımlanmaması, varyans kaynaklarının doğru tespit edilememesi ya da istatistiksel çözümleme esnasında yapılan varsayımların gerçekçi olmaması elde edilen sonuçların güvenirliğini zedeleyebilir. Bahsedilen bu varsayımların detaylı tanımları Raudenbush ve Bryk (2002, s 254-255) tarafından verilmiştir. Çok düzeyli modelde hata terimlerinin görsel olarak incelenmesi varsayım ihlalleri hakkında bilgi verebilir. Grafikler incelenerek normallik varsayımı, doğrusal ilişki varsayamı ve eş-varyans (homoscedastisity) varsayımı gözden geçirilebilir. Normal bir regresyon analizinde olduğu gibi, çok düzeyli modellerde de aykırı değerlerin (outliers) incelenmesi gerekir. Snijders ve Bosker (2012) çok düzeyli modellerin varsayımlarının gözden geçirilmesini tavsiye eder. Hox (2010) hata grafiklerinin nasıl yorumlanacağını örneklerle açıklamıştır. Çok düzeyli modellerde aykırı değerlerin ve etkili veri noktalarının saptanması için influence.ME (Rense, Manfred ve Ben, 2012) kütüphanesi R yazılımı ile kullanılabilir, benzer analizler SAS yazılımında da mevcuttur (SAS, 2004).

Raudenbush ve Bryk (2002) birinci düzey hataların normal dağılmadığı durumlarda standart hataların her iki düzeyde de yanlı olacağını, ikinci düzey hataların normal dağılmadığı durumlarda ise ikinci düzeydeki standart hataların yanlı olacağını yazmıştır. Bir ya da daha fazla varsayımın geçerliliği kuşkulu ise tahmin yöntemleri çeşitli uyarlamalarla (modification) kullanılabilir, değişikliklerden sonra bu tahmin edicilerin daha dirençli (robust) olması beklenir.

Bağımsız Değişkenlerin Merkezileştirilmesi

Merkezileştirme seçenekleri genel olarak üç kategoride incelenir; genel ortalama etrafında merkezileştirme (GOM), küme ortalaması etrafında merkezileştirme (KOM) ve herhangi bir merkezileştirilmenin yapılmaması durumu. Çok düzeyli regresyon modellerinde, bağımsız değişkenlerin, özellikle birinci düzey bağımsız değişkenlerin merkezileştirme ile yeniden tanımlanmasının istatistiksel sonuçları vardır. Tek düzeyli regresyon modellerinde kullanılacak merkezileştirme (GOM), sadece regresyon sabitinin tahminini ve yorumu değiştirecekken, çok düzeyli modellerde uygulanacak merkezileştirme regresyon sabitine ek olarak modeldeki regresyon katsayılarının büyüklüğünü ve yorumunu da değiştirebilir.

Bununla birlikte, çok düzeyli bir modelde sadece GOM kullanılması, modellerin eşitliğini değiştirmez (Kreft , de Leeuw ve Aiken, 1995). Bir diğer deyişle, GOM kullanılan modeller ve merkezileştirme kullanılmayan modellerin eşitliği matematiksel olarak ispatlanabilir (Paccagnella, 2006). Burada bahsedilen eşitlik, model-veri uyumunun ve tahmin edilen bağımlı değişken değerlerinin (predicted values) aynı olmasıdır. Bu noktada çok düzeyli modellerde GOM kullanmanın avantajı regresyon sabitinin yorumunun kolaylaşmasıdır. Buna ilave olarak, Raudenbush ve Bryk (2002), modelde düzeyler arası etkileşimin yer aldığı durumlarda, GOM kullanmanın veriler arasındaki doğrusal bağlılığı azaltabileceğini belirtmiştir.

Merkezileştime konusunun nispeten daha karmaşık olduğu durum KOM'dur. Birinci düzey bir değişkenin KOM ile yeniden tanımlanması durumunda yeni model GOM ya da merkezileştirme kullanılmayan modelden farklıdır. İki düzeyli modellerde KOM kullanımının detaylı bir tartışması Algina ve Swaminathan (2011) tarafından verilmiştir. KOM kullanmanın özellikle önemli olduğu nokta bağlamsal etkinin (context effect) mevcut olduğu durumlardır. Bağlamsal etki, birinci düzey bir değişkeninin küme-içi ve kümeler-arası etkisi birbirinden farklı ise mevcuttur.

Çok Düzeyli Modellerde Hipotez Testi Alternatifleri

Regresyon modellerinde sadece sabit etkiler olduğu için sabit katsayıların tahmini yapıldığında hipotez testi işlemi bu sabit katsayıların sıfıra eşit olduğunu (H_0: $\beta = 0$) varsayan sıfır hipotezinin testi ile gerçekleştirilir. Burada sıfır hipotezinin istatistiksel olarak sıfırdan farklı olup olmadığı t-testi ile sınanır. Bu uygulamada *t* değeri regresyon katsayısının tahmini değerinin standart hata değerine bölünmesiyle elde edilir ($t = \hat{\beta}/\widehat{SH}$). Serbestlik dercesi değeri de örneklem büyüklüğünden parametre sayısının çıkarılmasıyla elde edilerek (N-k) hesaplanır ve t-değerinin anlamlı bulunup bulunmadığını sınamak için kullanılır. Çok düzeyli modellerde parametre tahminleri hipotez etmede t-testinin yanı sıra birçok farklı yöntem kullanılmaktadır. Çok düzeyli modeller sabit parametrelerin yanında rastgele parametreler de içerdiği için sabit ve rastgele kısımlara ait parametre tahminleri için ayrı ayrı hipotez testleri uygulanabilir. Öncelikle çok düzeyli modellerdeki sabit parametre tahmin değerlerini test etmek için regresyonda da olduğu gibi *t*-testi kullanılabilir. Burada da kesenin ya da eğimin tahmini değeri standart hata değerine bölünür ve yaklaşık olarak *Z* veya *t* dağılımına sahip olan bir değer elde edilir. Bu oran Wald oranı olarak da adlandırılır (Snijders ve Bosker, 2012). Bu *Z* ya da *t* dağılımına sahip değeri test etmek için Düzey 1 ya da Düzey 2 eleman sayısı ile bağımsız değişken sayısının farkından bir çıkarılarak elde edilen (n-q-1) serbestlik derecesi değeri kullanılır. Çok düzeyli modellerde serbestlik derecesi hesaplaması farklı düzeylerin farklı örneklem büyüklüklerine sahip olması sebebiyle regresyon analizlerinde olduğu gibi kesinlik göstermemektedir. Bu serbestlik derecesi değeri her analizde farklı değer verdiği için yaklaşık serbestlik derecesi (approximate degrees of freedom) olarak da adlandırılır. Bu yaklaşık serbestlik derecesinden kaynaklanan problemi çözmek için farklı serbestlik derecesi hesaplama yöntemleri geliştirilmiştir (Kenward ve Roger, 1997; Manor ve Zucker, 2004; Satterthwaite, 1946). Çok düzeyli modellerdeki sabit etkiler üzerine çıkarım yapabilmek için geliştirilen alternatif strateji en çok olabilirlik oran testinin (likelihood ratio test) kullanılmasıdır (Manor ve Zucker, 2004). Manor ve Zucker (2004) sabit katsayıların tahmini üzerine çıkarım yapabilmek için geliştirilmiş sekiz farklı yöntemin küçük örneklemlerde karşılaştırmasını yapmıştır. Bu yöntemler en büyük olabilirlik (ML: maximum likehood) ve kısıtlandırılmış en büyük olabilirlik (REML: restricted maximum likelihood) tahmin yöntemleriyle elde edilen standart asimptotik normal teori testleri (standard asymptotic normal-theory tests), en çok olabilirlik oran testi ve bu testin değiştirilmiş (modified) verisyonunu, ve *t* ve *F* dağılımlarının düzeltilmiş testlerini içeren yöntemlerdir. Manor ve Zucker'in (2004) çalışmasına göre REML tahmini ile elde edilen Satterwaite prosedürü ve ML tahmini ile elde edilen Bartlett düzeltilmiş en çok olabilirlik oran testi uygun I. tür hata oranları ile iyi performans göstermişlerdir. Bunlara ek olarak Kenward-Rogers yöntemi de iyi performans gösteren diğer bir yöntem olarak öne çıkmıştır.

Çok düzeyli modellerdeki varyans bileşenlerinin tahmini katsayılarını ve standart hata değerlerini elde ettikten sonra *Z*-testi kullanarak hipotez testi uygulamak sabit katsayılarda olduğundan daha karmaşıktır. Bunun sebebi her zaman pozitif değer alan varyans tahminleridir. Varyans örnekleme dağılımlarının normal dağılım göstermemesi ya da çarpık olması normallik varsayımına dayalı t-testinin kullanılmasını zorlaştırmaktadır. Bu sebepten dolayı varyans bileşenleri için hipotez testi en çok olabilirlik oran testi kullanılarak yapılabilmektedir (Snijders ve Bosker, 2012). En çok olabilirlik oran testi daha çok birden fazla parametreyi aynı anda test etmek için ve özellikle de rastgele parametrelerin tahminlerini test etmek için tercih edilen bir yöntemdir. En çok olabilirlik oran testi biri diğerinin genişletilmiş verisyonu olan iç-içe geçmiş (nested) modellerin karşılaştırılmasında da kullanılabilir. Burada en çok olabilirlik oran testini kullanırken dikkat edilmesi gereken bir konu da hangi tahmin yönteminin (ML ya da REML) modelde kullanıldığıdır. Bu yöntem genellikle ML tahmin yöntemi kullanıldığında tercih edilir. Eğer REML tahmin yöntemi kullanarak iki model tahmin edilmiş ve en çok olabilirlik oran testi ile karşılaştırılacaksa bu iki modelin de sahip olduğu sabit etkiler aynı olmak zorundadır. Varyans bileşenlerinin tahmini değerlerinin test edilmesinde χ^2 testi de kullanılabilmektedir (Raudenbush ve Bryk, 2002). LaHauis ve Ferguson (2009) χ^2 testi ile en çok olabilirlik oran testini varyans bileşenleri için karşılaştırmış ve χ^2 testinin I. tür hata oranı açısından daha güçlü olduğunu göstermiştir. Rastgele etkileri test etmek için başka bir alternatif de permütasyon testidir (Lee ve Braun, 2012). Son olarak *F*-testinin rastgele kesen (intercept) tahmini için kullanılabileceğinden bahsetmekte yarar vardır. Eğer açıklayıcı değişkenleri kontrol ettiğimizde kesitin rastgele mi sabit mi olduğunu test etmek isteniyorsa ANCOVA yöntemi kullanılarak elde edilen *F* değeri de rastgele kesit tahmin değerinin hipotez testi için kullanılabilir (Snijders ve Bosker, 2012).

Çok düzeyli modellerde hipotez testine ek olarak model seçimi de önemli bir konudur. Eldeki veriyi analiz etmek için sabit ve rastgele olmak üzere önce kesit ve eğimin durumuna sonra da modele giren değişkenlerin durumuna karar vermek gerekmektedir. Birden fazla bağımsız değişkenin olduğu durumlarda bu değişkenler arası etkileşim (interaction) de araştırma için önemli olabilmektedir. Bu s ebeplerden dolayı öncelikle araştırmanın altında yatan teoriye bakmanın yanı sıra modele karar vermek için birçok istatistiksel yöntem kullanılmaktadır. Bu model seçme yöntemlerinin başında en çok olabilirlik oran testi ve bilgi ölçütleri (information criteria) gelmektedir. Eğer karşılaştırdığımız iki farklı model birbiri içerisinde kümelenmiş şekilde ifade edilebiliyorsa en çok olabilirlik oran değerlerinin ve serbestlik derecelerinin farkı alınır. Bu fark χ^2 dağılımı gösterdiği için serbestlik dereceleri arasındaki fark da kullanılarak hangi modelin daha iyi uyum gösterdiğine karar verilir. İkinci yöntem olan bilgi ölçütleri ise iç-içe geçmemiş modellerin karşılaştırılmasında kullanılır. Karşılaştırılan modellerden daha küçük bilgi ölçütü değerine sahip olanın daha iyi uyuma sahip olduğu sonucuna varılır.

Uygulama Verisinin Detayları

İllustrasyon amaçlı kullanılan veri 2010-2013 yılları arasında, Amerika Birleşik Devlerinde, Eğitim Bilimleri Enstitüsü tarafından desteklenmiş bir proje için toplanmıştır (Daunic ve diğ., 2012). Araştırma sorularından bir tanesi, sınıf düzeyinde verilen bir müdahelenin öğrencilerin davranışlarında olumlu bir değişime yol açıp açmayacağıdır. Proje kapsamında ilkokul 4. ve 5. sınıf öğrencileri hakkında 6 ölçme aracından 23 farklı kuramsal yapıya yönelik veri toplanmıştır. Her öğrenci ve bulunduğu sınıf için kimlik (id) bilgisi olan veri setinden, illustrasyon amaçlı 50 sınıf ve her sınıftan 12 öğrenci olmak üzere 600 kişilik bir alt küme rastgele seçilmiştir. Verilerin tamamı ile yapılmış çok düzeyli analizler için Smith ve diğerleri. (2014) incelenebilir. Kullanılan değişken, Klinik Davranış Ölçeği (KDÖ, Clinical Assessment of Behavior, Bracken ve Keith, 2004) ile oluşturulmuştur, bu puan 70 adet Likert tipi (1-5) soruya verilen cevapların toplamıdır ve 70 ila 350 arasında değişir. Bağımlı değişken müdahele sonrası toplanan KDÖ puanları (son test), bağımsız değişkenler ise müdahele öncesi toplanan KDÖ puanları (ön test) ve sınıf düzeyinde uygulanmış müdahele verisidir (treatment indicator, 1= müdahele, 0=kontrol).

Uygulama Verilerinin Analizi

Analizler R programı ile yapılmıştır, çok düzeyli modellerin analizi için nlme paketi kullanılmıştır (Pinheiro, Bates, Debroy, Sarkar ve R Core Team, 2016). Betimleyici istatistikler için psych paketi (Revelle, 2015) kullanılmıştır. İlk olarak ön-test ve son-test puanlarının dağılımları gözden geçirilmiştir ve betimsel analizler yürütülmüştür. Bu analizleri, kümeli yapıları gözardı ederek yürütülen doğrusal regresyon analizi, kümeli yapıların gerekli olup olmadığı hakkında fikir sahibi olmak için yürütülen KİK analizi takip etmiştir. Çok düzeyli regresyon modelleri olarak rassal-kesim ve rassal-kesim-eğim modelleri incelenmiştir. Uygulama verisinin analizinde merkezileştirme yöntemlerinden GOM kullanılmıştır. KOM ile yapılan bağlamsal etki analizleri ek olarak raporlanmıştır. Varsayım kontrolü için hataların dağılım grafikleri incelenmiştir.

Bulgular

Bağımlı değişken KDÖ-son test puanının dağılımı normal dağılımdan çok sapmadığı gözlenmiştir ($\bar{Y}$=151, ss=52.89, çarpıklık=0.64, basıklık=2.91). Bağımsız değişken KDÖ-ön test puanının dağılımı da normale yakındır ($\bar{X}$=146, ss=48.39, çarpıklık=0.6, basıklık=2.75). Bu değişken analizlerde genel ortalamaya göre yeniden hesaplanmıştır. Böylece bu puanın 0 olması, öğrencinin genel ortalamaya eşit bir puanının olduğunu gösterir. Diğer bağımlı değişken ikili veri olan müdahele durumudur, veri setinde yer alan 23 sınıf kontrol grubunda, 27 sınıf ise müdahale grubunda yer almıştır.

İlk olarak kümeli yapılar gözardı edilerek Eşitlik 28'de yer alan basit regresyon analizi yürütülmüştür. KDÖ-son test puanlarının, KDÖ-ön test puanları ve müdahele değişkeni ile açıklandığı regresyon eşitliği istatistiksel olarak anlamlı bulunmuştur, (F(2,597)=420.8, p<.001, düzeltilmiş-R^2=.584). Ön test puanları (b=0.84, p<.001) ve müdahele değişkeni (b=-5.83, p=.038) istatistiksel olarak anlamlı bulunmuştur. KDÖ puanlarında yüksek puanının yüksek davranış bozukluğu riski olduğu göz önüne alınınca, kümeli yapıyı göz ardı eden bir araştırmacı, müdahelenin olumlu yönde etkili olduğu sonucuna ulaşabilir.

ANOVA ayrışımı ile hesaplanan KİK değeri son test puanları için.190 bulunmuştur. Yüzde 95 güven aralığı.118-.295 olan bu KİK değeri istatiksel olarak anlamlıdır ($F(49,550)=3.82$, $p<.001$). Bir diğer deyişle, kümeli yapının bağımlı değişkenin varyansına olan etkisi göz ardı edilmemelidir. Bu KİK değeri ile tasarı etkisi 3.09 bulunur. KİK değeri ML ve REML tahminleme yöntemi ile de hesaplanabilir, örneğimiz için REML tahmini 0190, ML tahmini.186 olarak bulunmuştur. Çok düzeyli regresyon varyans tahminlerinde ML metodunun REML metoduna nazaran daha küçük tahminlerde bulunacağı Raudenbush ve Bryk (2002, s 53) tarafından not edilmiştir, fakat bu fark küme sayısı arttıkça gözardı edilebilecek kadar küçülür.

Eşitlik 29 boş modeli, Eşitlik 30 rassal-kesim modelini, Eşitlik 31 ise rassal-kesim-eğim modelini gösterir. Tablo 1 bu modeller ile yapılan tahminleri, standart hatalarını ve p-değerlerini göstermektedir. Bu tablodan görülebileceği gibi, ikinci düzey bağımsız değişken olan müdahelenin etkisi, tek düzeyli modelin aksine, çok düzeyli modellerde istatistiksel olarak anlamlı bulunmamıştır. Bir diğer deyişle I. tür hata yapılmamıştır.

$$Y_i = \beta_0 + \beta_1(X_i - \overline{X}_i) + \beta_2(T_i) + \varepsilon_i \tag{28}$$

$$\begin{aligned} &\text{Düzey 1;}\quad Y_{ij} = \beta_{0j} + \varepsilon_{ij} \\ &\text{Düzey 2;}\quad \beta_{0j} = \gamma_{00} + u_{0j} \end{aligned} \tag{29}$$

$$\begin{aligned} &\text{Düzey 1;}\quad Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{\overline{X}}_{ij}) + \varepsilon_{ij} \\ &\text{Düzey 2;}\quad \beta_{0j} = \gamma_{00} + \gamma_{01}T_j + u_{0j} \\ &\qquad\qquad\quad \beta_1 = \gamma_{10} \end{aligned} \tag{30}$$

$$\begin{aligned} &\text{Düzey 1;}\quad Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{\overline{X}}_{ij}) + \varepsilon_{ij} \\ &\text{Düzey 2;}\quad \beta_{0j} = \gamma_{00} + \gamma_{01}T_j + u_{0j} \\ &\qquad\qquad\quad \beta_{1j} = \gamma_{10} + u_{1j} \end{aligned} \tag{31}$$

Boş model, rassal kesim modeli ve rassal kesim-eğim modellerinde tahminleyici olarak kısıtlandırılmış en çok olabilirlik (REML) yöntemi kullanılmıştır. Modelde yer alan regresyon katsayılarının istatistiksel anlamlılığı için t-testi prosedürü kullanılmıştır (serbestlik derecesi hesaplamaları için Pinheiro ve Bates, 2000, s. 90-91). Boş model ile hesaplanan γ_{00}bağımsız değişkenin ortalamasıdır, 150.91 olarak hesaplanan bu değer istatistiksel olarak sıfırdan farklıdır (t_{550}=39.66, p<.001). Eşitlik 6 ile γ_{00} için hesaplanan %95 güven aralığı 143.26 ile 158.55 arasındadır.

Tablo 1

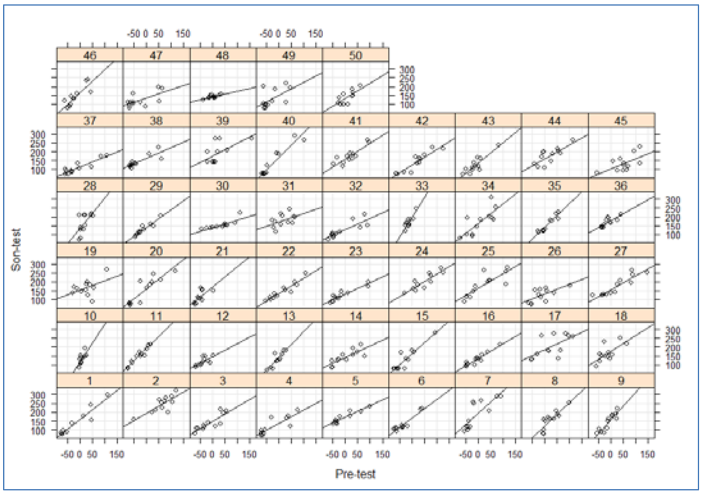
Model sonuçları

Model	γ_{00}	γ_{10}	γ_{01}	σ^2	τ^2	$\tau^2_{eğim}$
Tek Düzey Regresyon	154.06(2.06)*	0.84(0.03)*	-5.83(2.81)*	1165		
Boş model	150.91(3.81)*			2273	535	
Rassal kesim	154.13(4.07)*	0.85(0.03)*	-5.97(5.54)	868	309	
Rassal-kesim-eğim	153.39(3.80)*	0.87(0.04)*	-3.17(4.99)	806	288	0.03

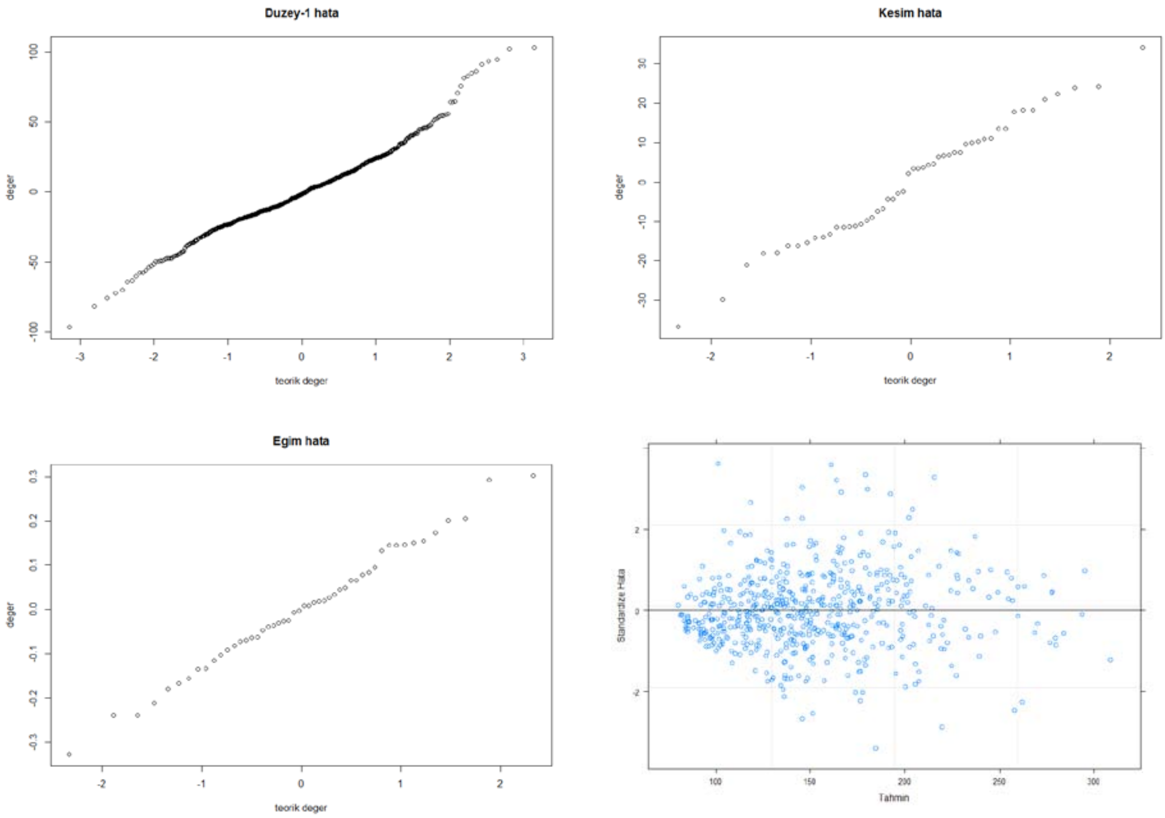
Not. * p<.05; γ_{00}: Sabit, γ_{10}: Ön test, γ_{01}:Müdahele

Rassal kesim modeli ile hesaplanan γ_{00}, ön teste göre düzeltimiş, kontrol grubu üyelerinin genel ortalamasıdır, 154.13 olarak hesaplanan bu değer istatistiksel olarak sıfırdan farklıdır (t_{549}=37.85, p<.001). Genel ortalama etrafında merkezileştirilmiş ön test için hesaplanan γ_{10}değeri 0.85 bulunmuştur ve bu değer istatistiksel olarak sıfırdan farklıdır (t_{549}=30.60, p<.001). Bu değer, ön test puanlarında görülen her bir artışın tahmin edilen son test puanlarında 0.85 puan artışla ilişkilidir denilebilir.

Müdahele verisinin katsayısı γ_{01}=-5.97 olarak hesaplanmıştır ve bu değerin sıfıra eşit olduğu sıfır hipotezi reddedilemez (t_{48}=-1.08, p=0.29) şeklinde çıkmıştır. Rassal kesim-eğim modeli için tahminler benzerdir, γ_{00}=153.39 (t_{549}=40.38, p<.001), γ_{10}=0.86 (t_{549}=22.28, p<.001) ve γ_{01}=-3.17 (t_{48}=-0.63, p=0.53). Rassal kesim modelinden farklı olarak modele eklenen rassal eğim hata teriminin varyansı olabilirlik oran testine (deviance test) göre anlamlı bulunmuştur, bu teste göre hesaplanan fark 20.42, 2 serbest dereceli ki-kare dağılımına göre anlamlı (p<.001) bulunmuştur. Eğimlerin farklı olduğu Şekil-1 de görülebilir, bu grafik R paketi lattice (Sarkar, 2008) ile çizilmiştir. Rassal kesim-eğim modeli için hataların normal dağıldığı varsayımının Şekil-2 de yer alan grafikler incelenerek güçlü bir şekilde ihlal edilmediği sonucuna varılmıştır. Ayrıca, model ile tahmin edilen değerlerin standardize edilmiş birinci düzey hatalara göre çizilen saçılım grafiğinde eş-varyanslılığın güçlü bir şekilde ihlal edilmediği görülmüştür. Varsayım ihlalleri durumunda araştırmacılar modeli gözden geçirmeyi, dönüşüm (transformation) yapmayı ya da değiştirilmiş (modified) daha dirençli tahminleyiciler kullanmayı deneyebilirler.[3]



Şekil 1. Her sınıf için regresyon grafiği, ilk 23 sınıf kontrol grubu



Şekil 2. Rassal-kesim-eğim modeli hatalarının dağılım grafiği.

[3] Varsayımlar hakkındaki kabullerimizi kontrol etmek amacıyla analizler Mplus yazılımının varsayım ihlallerine karşı dirençli en büyük olabilirlik (MLR) tahminleyicisi ile tekrarlanmıştır ve sonuçlar kayda değer şekilde değişmemiştir.

Klinik davranış ölçeği verileri üzerinde yapılan analizde, GOM kullanılmaması durumunda regresyon sabiti ön test puanı 0 olan öğrencilerin ortalamasını verecektir. Fakat bu ölçekten gelen puanlar 70 ila 350 arasında değer aldığı için, regresyon sabiti merkezileştirme yapılmadığı durumda gerçekçi bir yoruma sahip değildir.

Klinik davranış ölçeği için Eşitlik 32'de verilen model koşularak ön test puanlarının bağlamsal etkisi olup olmadığı test edilebilir. Bağlamsal etkinin istatistiksel anlamlılığı KOM ile de test edilebilir. Bu durumda model Eşitlik 33'te olduğu gibi inşaa edilmelidir.

$$\begin{aligned}
&\text{Düzey 1;}\quad Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij}) + \varepsilon_{ij} \\
&\text{Düzey 2;}\quad \beta_{0j} = \gamma_{00} + \gamma_{C} X_{j} + u_{0j} \\
&\qquad\qquad \beta_{1} = \gamma_{10}
\end{aligned} \tag{32}$$

$$\begin{aligned}
&\text{Düzey 1;}\quad Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{j}) + \varepsilon_{ij} \\
&\text{Düzey 2;}\quad \beta_{0j} = \gamma_{00} + \gamma_{B} X_{j} + u_{0j} \\
&\qquad\qquad \beta_{1} = \gamma_{10}
\end{aligned} \tag{33}$$

Eşitlik 32'de ön test puanları merkezileştirilmemiş ve ikinci düzeye her sınıfın ön test ortalaması eklenmiştir. Eşitlik 33'te ön test puanları küme ortalaması etrafında merkezileşirilmiş ve ikinci düzeye her sınıfın ön-test ortalaması eklenmiştir Sonuçlar Tablo 2'de verilmiştir. Klinik davranış ölçeği için yürütülen analizde bağlamsal etkinin istatistiksel olarak anlamlı bulunmadığı görülmüştür (γ_C=-0.10, SH=0.12). Bağlamsal etki Eşitlik 33 ile oluşturulan modelden gelen γ_B ve γ_{01}'in farkına eşittir (0.76-0.86). Ön test için bir bağlamsal etkinin olmaması ve düzeyler arası bir etkileşimin modelde yer almadığı göz önünde bulundurulduğunda, Klinik Davranış Ölçeği Tablo 1'de raporlanan çok düzeyli analizlerin geçerli olduğu kabul edilebilir.

Tablo 2

Bağlamsal etki analizleri

Model	γ_{00}	γ_{01}	γ_C	γ_B	σ^2	τ^2
Eşitlik 32	39.39(16.99)*	0.86(0.03)*	-0.10(0.12)		868	312
Eşitlik 33	39.39(16.99)*	0.86(0.03)*		0.76(0.12)*	868	312

Not: * *p*<.05

Sonuçlar ve Tartışma

Kümeli yapıya sahip olan verilerin analizleri yürütülürken küme etkisinin gözardı edilmesi hatalı çıkarımlara yol açabilmektedir. Yapılabilecek hatalı çıkarımlardan biri, uygulama örneğinde gösterildiği gibi, kümeli yapıların dikkate alınarak yapılan daha geçerli analizlerde istatistiksel olarak anlamlı bulunmayan bir müdahele etkisinin, kümeli yapıların gözardı edildiği doğrusal regresyon modelinde istatistiksel olarak anlamlı bulunmasıdır. Kısaca I. tür hata yapmaktır. Çok düzeyli modeller, varyansın daha doğru bir şekilde modellenmesi, anlaşılması ve yorumlanması açısından da önemlidir. Bu makalenin amacı çok düzeyli modellerin gerekliliğini, basit teorik altyapısını, basit bir uygulamasını göstermek ve çok düzeyli modellerde önemli olan bazı noktaların kısaca özetlenmesidir. Bu konular çok düzeyli modellerin doğru bir şekilde uygulanabilmesinde kritik önem taşımaktadır. Bu konuların önemine rağmen yurtiçinde yapılan çok düzeyli model çalışmalarında bazılarının göz ardı edildiği gözlenmiştir. Çok düzeyli modellerin yorumlanmasını kolaylaştırmak adına uygulanan merkezileştirme uygulamasından birkaç çalışma (Acar, 2013; Atar, 2010; Atar, 2014; Çoker, 2009; Güvendir, 2014) dışında neredeyse hiç bahsedilmemiştir. Diğer çıkarımsal istatistik yöntemlerinde olduğu gibi çok düzeyli modeller belirli bazı varsayımlara sahiptirler. Bu varsayımların sağlanmaması bulguların geçerliğini olumsuz yönde etkilemektedir. Alanyazında bu konuda da yapılan çok düzeyli modellerde eksiklikler olduğu gözlenmiştir. Alanyazın incelendiğinde dikkati çeken başka bir konu da uluslarası büyük ölçekli sınav verisini kullanan çalışmaların örneklem özelliklerinden olan ağırlıklandırma meselesini göz ardı etmeleri olmuştur. Bu ve benzeri büyük çapta yürütülen çalışmalar çoğu zaman örneklemin nasıl seçildiği konusunda oldukça detaylı ve dikkatli bir ağırlıklandırma prosedürü takip ederler. Bu çalışmalardan edinilen veriler üzerine yapılan analizlerin genellenebilirliği bu ağırlıklandırmaların analizlerde ne kadar doğru kullanıldığı ile doğru orantılıdır.

Araştırmacılar yürütülen analizler için kullanılan R betiğini ek dosyada bulabilirler. Bu çalışma kapsamında değinilmemiş fakat çok düzeyli modellerde oldukça önemli olan konulardan bazıları, kayıp veri teknikleri, kusursuz olmayan kümeli yapılar (imperfect hierarchies), ağırlıklandırma, sürekli olmayan bağımlı değişkenler, boylamsal yapılar ve model inşaası prosedürleridir. Kümeli yapılarda kayıp veri problemi tespit eden araştırmacılar Black, Harel ve McCoach (2011), Graham (2012), Longford (2008), Yang, Kim ve Zhu (2013) kaynaklarını incelemek isteyebilirler. R kullanıcıları, *pan* (Zhao ve Schafer, 2013) paketini işe yarar bulabilirler. Bireylerin birden fazla kümeye dahil olabildiği durumlarla karşılaşan araştırmacılar Beretvas (2011) tarafından kaleme alınan kitap bölümünden faydalanabilirler. Bu çalışmada çok düzeyli modellerin hepsi için geçerli olan birçok konuya vurgu yapılmasının yanında sadece iki düzeyli kesitsel modelin uygulaması verilmesi bu çalışmanın sınırlılıkları arasında sayılabilir. Çok düzeyli model alanyazınında farklı veri durumlarında kullanılan birçok çok düzeyli model türleri bulunmaktadır. Bu modeller arasında iki düzeyli modelin genişletilmiş versiyonu olan üç düzeyli modeller, kategorik bağımlı değişkenin analizine imkan veren genelleştirilmiş (generalized) çok düzeyli modeller, boylamsal verilerin analizine imkan veren boylamsal çok düzeyli modeller verilebilir. Bu model türlerinin yanı sıra çok düzeyli modellerin diğer model türleri ile birleştirilmesinden oluşan yeni model türleri bulunmaktadır. Bu modellere çok düzeyli yapısal eşitlik modelleri, çok düzeyli madde tepki kuramı modelleri ve çok düzeyli meta-analiz modelleri örnek verilebilir. Yurtiçindeki alanyazına bakıldığında çok düzeyli yapısal eşitlik modellerinin diğer modellerden daha fazla çalışıldığı görülmektedir (Can ve diğ., 2011; Çoker, 2009; Gölbaşı-Şimşek ve Noyan, 2009). Bu çalışmada olduğu gibi yukarıda belirtilen çok düzeyli model türlerinin detaylıca anlatıldığı çalışmaların yapılmasına Türkçe alanyazında ihtiyaç vardır.

Kaynakça

Acar, M. (2013). *Öğrenci başarılarının belirlenmesi sınavında Türkçe dersi başarısının öğrenci ve okul özellikleri ile ilişkisinin hiyerarşik lineer model ile analizi*. (Doktora tezi). Ankara Üniversitesi, Ölçme ve Değerlendirme Anabilim Dalı, Ankara. http://tez2.yok.gov.tr/ adresinden edinilmiştir.

Acar, T. ve Öğretmen, T. (2012). Çok düzeyli istatistiksel yöntemler ile 2006 PISA fen bilimleri performansının incelenmesi. *Eğitim ve Bilim, 37*(163), 178-189.

Algina, J., ve Swaminathan, H. (2011). Centering in two-level nesteddesigns. J. Hox ve J. K. Roberts (Eds.), *Handbook of advancedmultilevel analysis* (s. 285–312). New York, NY: Taylor and Francis.

Aitkin, M., ve Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society. Series A (General)*, 1-43.

Atar, B. (2010). Basit doğrusal regresyon analizi ile hiyerarşik doğrusal modeller analizinin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 1*(2), 78-84.

Atar, H. Y. (2014). Öğretmen niteliklerinin TIMSS 2011 fen başarısına çok düzeyli etkileri. *Eğitim ve Bilim, 39*(172), 121-137.

Aydın, B. (2014). *Statistical power in cluster randomized trials: An evaluation of observed and latent mean covariate* (Doctoral dissertation). Retrieved from http://uf.catalog.fcla.edu/uf.jsp?st=UF033650895

Beretvas, N., S. (2011). Cross-clasified and multiple-membership models. J. Hox ve J. K. Roberts (Eds.), *Handbook of advancedmultilevel analysis* (s. 313–335). NY: Taylor and Francis.

Black, A. C., Harel, O., ve Betsy McCoach, D. (2011). Missing data techniques for multilevel data: implications of model misspecification. *Journal of Applied Statistics, 38*(9), 1845–1865. doi:10.1080/02664763.2010.529882

Bracken, B. A., ve Keith, L. K. (2004). *Clinical assessment of behavior*. Lutz, FL: Psychological Assessment Resources.

Bryk, A. S., Raudenbush, S. W., ve Congdon, R. T. (1996). *HLM: Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs*. SSI Scientific Software International.

Burstein. L., Linn, R., ve Capell, F. (1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journal of Educational Statisrics 3,* 347-183.

Can, S., Somer, O., Korkmaz, M., Dural, S., ve Öğretmen, T. (2011). Çok düzeyli yapısal eşitlik modelleri. *Türk Psikoloji Dergisi, 26*(67), 14-21.

Çoker, E. (2009). *Çok-Düzeyli regresyon modelleri ile çok düzeyli yapısal eşitlik modellerinin uygulamalı karşılaştırılması*. (Doktora tezi). Mimar Sinan Güzel Sanatlar Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul. http://tez2.yok.gov.tr/ adresinden edinilmiştir.

Cornfield, J. (1978). Randomization by Group: A formal analysis. *American Journal of Epidemiology, 108*(2),100-102.

Cronbach, L.J. ve Webb, N. (1975) Between-class and within-class effects in a reported aptitude × treatment interaction: reanalysis of a study by GL Anderson. *J. Educ. Psychol.67*:717–724.

Daunic, A. P., Naranjo, A. H., Smith, S. W., Garvan, C. W., Barber, B. R., Becker, M. K., ve Li, W. (2012). Reducing developmental risk for emotional/behavioral problems: A randomized controlled trial examining the tools for getting along curriculum. *Journal of School Psychology*, *50*(2), 149-166. doi:10.1016/j.jsp.2011.09.003

De Leeuw, J., ve Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational and Behavioral Statistics*, *11*(1), 57-85.

Deniz-Başar, Ö., Özden, Ü. H., ve Bağdatlı-Kalkan, S. (2013). Boylamsal verilerde çok düzeyli analizler: Dil gelişimine ilişkin bir uygulama. *Ekonometri ve İstatistik, 19*, 23-37.

Erol-Korkmaz, H. T. (2014). Çalışanların günlük duygu durumu ve dretim karşıtı davranışları arasındaki ilişki: Genel örgütsel adalet algısının düzenleyici rolü. *Türk Psikoloji Yazıları, 17*(33), 77-87.

Fisher RA. (1958). *Statistical Methods for Research Workers*. 13th edition. London: Hafner Press.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, *73*, 43-56.

Goldstein, H. (1995). Hierarchical data modeling in the social sciences.*Journal of Educational and Behavioral Statistics*, *20*(2), 201-204.

Gölbaşı-Şimşek, G., ve Noyan, F. (2008). İlçelerin gelişmişlik indekslerinin oluşturulmasında çok aşamalı doğrulayıcı factor analizi yaklaşımı. *İstatikçiler Dergisi, 1*, 50-67.

Gölbaşı-Şimşek, G., ve Noyan, F. (2009). Algılanan öğretimsel etkililiğin öğrenci sadakatine etkisi: Çok aşamalı yapısal eşitlik modeli. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 36*, 109-118.

Graham, J. W. A. (2012). *Missing data: Analysis and design*. New York, NY: Springer.

Güvendir, M. A. (2014). Öğrenci başarılarının belirlenmesi sınavında öğrenci ve okul özelliklerinin türkçe başarısı ile İlişkisi. *Eğitim ve Bilim, 39*(172), 163-180.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications.* New York, NY: Routledge.

Kadıoğlu, C., ve Uzuntiryaki-Kondakcı E. (2014). Relationship between learning strategies and goal orientations: A multilevel analysis. *Eurasian Journal of Educational Research, 56,* 1-24.

Karabay, E., Yıldırım, A., ve Güler, G. (2015). Yıllara göre PISA matematik okuryazarlığının öğrenci ve okul özellikleri ile ilişkisinin aşamalı doğrusal modeller ile analizi. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi, 36*, 137-151.

Kenward, M. G., ve Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*, 983-997.

Kim, J., ve Frees, E. W. (2006). Omitted variables in multilevel models. *Psychometrika*, *71*(4), 659-690. doi:10.1007/s11336-005-1283-0

Kish, L. (1965). *Survey sampling*. New York, NY: J. Wiley.

Kreft, I. ve De Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage Publications.

Kreft, I. G., De Leeuw, J., ve Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate behavioral research*, *30*(1), 1-21.

Kuehl, R. O. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis*, 2nd edition. Pacific Grove, California: Brooks/Cole.

LaHuis, D. M., ve Ferguson, M. W. (2009). The accuracy of significance tests for slope variance components in multilevel random coefficient models. *Organizational Research Methods*, *12*(3), 418–435. doi:10.1177/1094428107308984

Lee, O. E., ve Braun, T. M. (2012). Permutation tests for random effects in linear mixed models. *Biometrics*, *68*(2), 486–93. doi:10.1111/j.1541-0420.2011.01675.x

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested effects. *Biometrika 74*, 812-27.

Longford, N. T. (2008). Missing data. *Handbook of multilevel analysis (2008)*. DE: Springer New York. doi:10.1007/978-0-387-73186-5

Manor, O., ve Zucker, D. M. (2004). Small sample inference for the fixed effects in the mixed linear model. *Computational statistics & data analysis*, *46*(4), 801-817.

Muthén, L. K., ve Muthén, B. O. (2012). Mplus. *Statistical analysis with latent variables. Version*, *7*.

Noyan, F., ve Yıldız, D. (2006). YTÜ'de öğrenci gözüyle öğretim üyesi etkinliğinin iki aşamalı modeller yardımı ile değerlendirilmesi. *Mühendislik ve Fen Bilimleri Dergisi, 1*, 34-45.

Paccagnella, O. (2006). Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *Evaluation Review*, *30(1)*, 66-85. doi:10.1177/0193841X05275649

Pinheiro J, Bates D, DebRoy S, Sarkar D ve R Core Team (2016). *nlme*: Linear and Nonlinear Mixed Effects Models. R package version 3.1-127, http://CRAN.R-project.org/package=nlme adresinden elde edildi.

Raudenbush, S.W., ve Bryk, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.

Raudenbush, S. W., ve Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Rense, N., Manfred G., ve Ben, P. (2012). Influence.ME: Tools for detecting influential data in mixed effects models. *R Journal*, *4*(2),38-47.

Revelle, W. (2015) *psych*: Procedures for personality and psychological research, Northwestern University, Evanston, Illinois, USA, http://CRAN.R-project.org/package=psych Version = 1.5.8. adresinden elde edildi.

Roberts, J. K. (2004) An introductory primer on multilevel and hierarchical linear modeling. *Learning Disabilities: A Contemporary Journal*, *2*(1), 30-38.

Robinson,W. S. (1950). Ecological correlations and the behavior of individuals. *Sociological Review*, *15*, 351-357.

Sarkar, D. (2008) *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5

SAS, S. (2004). STAT 9.1 user's guide. *SAS Institute Inc., Cary, NC*, 1291-1320.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin, 2*(6), 110-114.

Schreiber, J. B., ve Griffin, B. W. (2004). Review of multilevel modeling and multilevel studies in *The Journal of Educational Research* (1992-2002). *The Journal of Educational Research, 98*, 24-33.

Smith, S. W., Daunic, A. P., Barber, B. R., Aydın, B., Van Loan, C. L. ve Taylor, G. G. (2014). Preventing risk for significant behavior problems through a cognitive-behavioral intervention: Effects of the tools for getting along curriculum at one-year follow-up. *The Journal of Primary Prevention, 35*(5), 371-387. doi:10.1007/s10935-014-0357-0.

Snijders, T., ve Bosker, R. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Thousand Oaks, CA: Sage.

Swaminathan, H., ve Rogers, H. J. (2008). Estimation Procedures for HLM. Hox, J., ve Roberts, K., *Multilevel modeling of educational data* (s. 469-519). Charlotte, NC: IAP.

Şahin, F. (2011). Liderin kültürel zekâsının astların örgütsel vatandaşlık davranışı ile iş doyumu üzerine etkisi. *Savunma Bilimleri Dergisi, 10*(2), 80-104.

Şen, S., ve Akbaş, N. (baskıda). Çok düzeyli meta-analiz yöntemleri üzerine bir çalışma. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi.*

Yang, S., Kim, J.-K., ve Zhu, Z. (2013). Parametric fractional imputation for mixed models with nonignorable missing data. *Statistics and Its Interface, 6*(3), 339–347. doi:10.4310/SII.2013.v6.n3.a4

Yılmaz, H. B., ve Aztekin, S. (2012,Haziran). *Türkiye'deki 15 yaş grubu öğrencilerin matematik okuryazarlığı başarılarını etkileyen bazı faktörlerin okul ve öğrenci düzeyine göre incelenmesi*. X. Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresi, Niğde, Türkiye.

Zhao, J. H., ve Schafer, J. L. (2016). *pan*: Multiple imputation for multivariate panel or clustered dataR package version 1.4.

Extended Abstract

This article is crafted in Turkish to provide basic theory, definition of important issues and an application of a two level regression model. It aims to enrich social scientists' methodological knowledge and equip them with appropriate theories and tools to deduce defensible inferences they draw from statistical analyses. The data in social sciences generally have a clustered structure; individuals are nested in classrooms, schools, companies, neighborhoods and so on. Ignoring the true structure of the data when utilizing statistical procedures might yield less valid results. Literature on how to address a clustered structure is notlimited; there are several frameworks (e.g. Permutation tests, generalized estimating equations, multilevel models) and even several names exist for the same or similar framework (e.g. mixed model, hierarchical models, nested models, multilevel models). Milestones of the field might date back to the works of Robinson (1950), Kish (1965) and Cornfield (1978). Goldstein (1995) Raudenbush and Bryk (2002), Hox (2010), and Snijders and Bosker (2012) are popular multilevel introductory books at least among social scientists. Even though there are several multilevel application papers in the Turkish literature, there is a paucity of information on the theory and analytic procedures of multilevel models. In the introduction, these application papers were mentioned shortly.

The basic theory with Kuehl (2000), Snijders and Bosker (2012), and Swaminathan and Rogers (2008) were introduced in the present study. Equations 1 through 27 are related to two-level null model, random intercepts model, and random intercepts and slopes models. In the study(a) two-level regression assumptions following Raudenbush and Bryk (2002, p. 254-255), (b) centering issue mainly following Algina and Swaminathan (2011), and (c) hypotheses testing alternatives mainly summarizing the literature review by Aydin (2014) were briefly summarized. Estimator alternatives; maximum likelihood, restricted maximum likelihood, Bayesian procedures were briefly mentioned. HLM, Mplus, MLwiN, R, SAS and STATA as the software alternatives were listed.

For illustrative purposes, a subsample from the original cluster randomized trial study by Daunic et al. (2012) was taken. The subsample included 50 clusters (23 in control) and 12 elementary-schoolers in each classroom. Pretest and posttest scores were gathered using Clinical Assessment of Behavior Teacher-Rating Form (Bracken and Keith, 2004). Based on 70 different 5-point Likert type items, the scale ranged 70 to 350. The analyses were conducted with R using the nlme package (Pinheiro, Bates, Debroy, Sarkar & R Core Team, 2016) and the psych package (Revelle, 2015).

The mean value of posttest scores was 151, the standard deviation was 52.99, the skewness value was.64, and the kurtosis value was 2.91. The mean value of pretest scores was 146, the standard deviation was 48.39, the skewness value was.6, and the kurtosis value was 2.75. Intraclass correlation coefficient for the post test scores was .190 with a 95 % confidence interval of [.118, .295]. First the data set ignoring the clustered structure was analyzed and a multiple regression model was ran (Equation 28) in which the posttest scores were set as dependent variable, and grand-mean centered pretest scores and the treatment indicator (1 = treatment, 0 = control) were set as independent variables. In this multiple regression model, the treatment indicator found to be statistically significant (b = -5.83, p =.038) indicating an expected treatment effect. Then the same data set was analyzed using a random intercepts model (Equation 30), and a random intercepts and slopes model (Equation 31). The results were given in Table 1. In these multilevel models, the treatment indicator was no longer statistically significant. The distribution of the residuals was studied to examine possible assumption violations, and these graphs were presented in Figure 2. Also cluster specific regression lines in Figure 1 using the R package lattice (Sarkar, 2008) Was depicted. The final analyses aimed to detect any possible contextual effects and to illustrate group-mean centering. These models presented in Equation 32 and 33 and the results were reported in Table 2. The contextual effect of the pretest was found to be insignificant. The R syntax for all these analyses can be found in the supplementary file.

Multilevel models are important tools to analyze data with a clustered structure. This study aimed to increase social scientists’ familiarity with multilevel models by introducing the basic theory, selected issues and an application of a two level regression model. An illustrative example was included and the R syntax was provided (see supplementary). This introductory study did not include the issues related to missing data in multilevel models, so readers were referred to Black, Harel and McCoach (2011), Graham (2012), Longford (2008) and Yang, Kim and Zhu (2013). Imperfect hierarchies, categorical outcomes and weighting issues in multilevel models were not included, either. Non-Turkish readers might find the R code helpful.