

İki Düzeyli HLM Modelleri

Dr. Sedat ŞEN

Bu dersin videosuna <https://ayeum.com/ders/hiyerarsik-lineer-modellemehlm-egitimi> adresinden ulaşabilirsiniz.

Bu Sunumda...

- Geleneksel regresyon analizi ve varsayımları
- İki düzeyli HLM modeli temel kavramları
- Parametre türleri
- İstatistiksel kestirim
- İki düzeyli HLM model türleri
- Merkezileştirme
- Model uyumu ve model karşılaştırması

Regresyon Tekrar

- Eğer elimizde tek bir sınıfa ait öğrenciler olsaydı bu öğrencilerden elde edilen veriler üzerinden analiz yapmak isteseydik tercihimiz regresyon olacaktı.

$$Y_i = \beta_0 + \beta_1 X_i + r_i$$

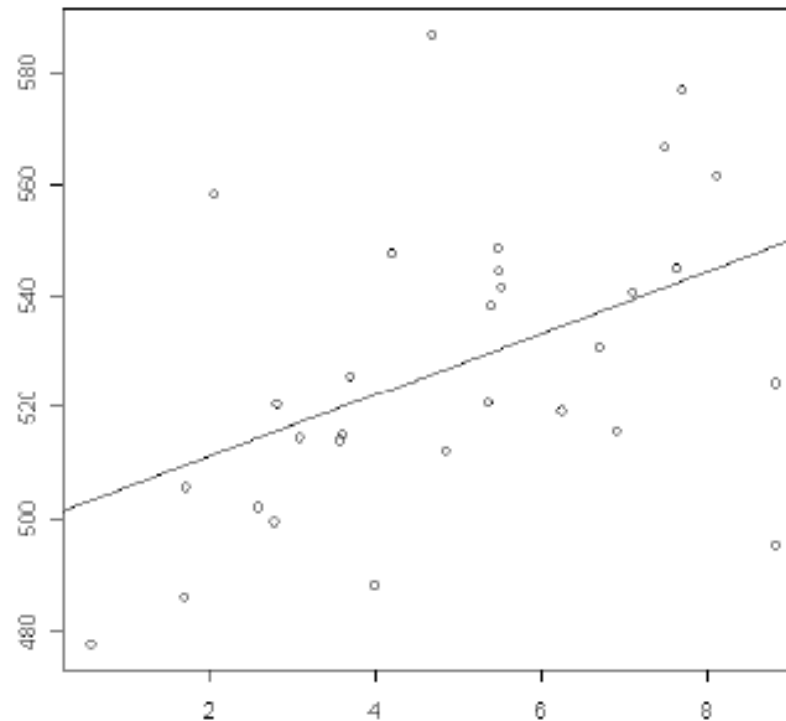
Regresyon Tekrar

$$Y_i = \beta_0 + \beta_1 X_i + r_i$$

- Y_i : Matematik başarısı (**sürekli değişken**)
- X_i : Sosyo-ekonomik-statü (SES)
- β_0 : KESEN: SES değeri sıfır olan öğrenci için yordanan puan
- β_1 : EĞİM: SES değerindeki bir birim artışa karşılık gelen matematik başarısı artışı
- r_i : ARTIK DEĞER(HATA): Yordanan ve gerçek değer arasındaki fark

Regresyon Grafiği

- Sabit kesen sabit eğim



Regresyon Varsayımları

- Tüm değişkenlerin hatasız ölçüldüğü varsayılır
- Hataların beklenen değeri (ortalaması) sıfırdır
- Regresyon, tüm artıkların sabit bir varyansa sahip bir popülasyondan çekildiğini varsayar (homoskedasticity)
- Normal dağılıma sahip olmalı
- Otokorelasyon yok (**farklı durumlar için hata terimleri arasında ilişki yoktur**)
- Öğrenciler arası hatalar bağımsızdır yani korelasyonsuzdur.

Regresyondan HLM'e doğru

- Elimizde tek bir okul ya da sınıf olduğunda regresyonun ürettiği tek bir eşitlik iş görecektir.
- Eğer birden fazla sınıf ya da okuldan öğrenci var ise bu öğrenciler arası benzerlik yani **bağımlılık** olacağından her bir okul için farklı regresyon eşitliği çizmek gerekecek.
- Her bir okul için j alt imi kullanacağız.
- Şimdi elimizde öğrenciler için i harfi okullar için j harfi kullanılan bir denklem olacak.

Birden Fazla Okul Durumu (HLM)

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

- Y_{ij} : Matematik başarısı
- X_{ij} : Sosyo-ekonomik-statü (SES)
- β_{0j} : j okuluna ait KESEN: SES değeri sıfır olan okuldaki bir öğrenci için yordanan başarı puan
- β_{1j} : j okuluna ait EĞİM: j okulundaki bir öğrencinin SES değerindeki bir birim artışa karşılık gelen matematik başarısı artışı
- r_{ij} : öğrenci/okul için ARTIK DEĞER(HATA): Yordanan ve gerçek değer arasındaki fark

Birden Fazla Okul Durumu (HLM)

- Yani, eğer rasgele okulları seçip her okul için regresyon çizgisini hesaplarsak.
- Kesen rasgele bir değişken olarak düşünebiliriz.
- Eğimi rasgele bir değişken olarak düşünebiliriz.
- Hem kesen hem de eğitim, okulların diğer özellikleri tarafından tahmin edilebilir. Yer (kırsal, kentsel), boyut vb...
- HLM hem β_{0j} için hem de β_{1j} için regresyon modelleri oluşturur. Bu kısım Düzey-2 regresyon olarak adlandırılır.

İkinci Düzey Regresyon

- Her β_{0j} ve β_{1j} için regresyon modelleri oluşturulur

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$$

- W_j okul düzeyi bağımsız değişken
- γ_{00} kesenlerin keseni yani $W_j=0$ olduğunda β_{0j} 'nin alacağı değerdir.
- γ_{01} kesenlerin eğimi yani W_j 1 birim arttığında β_{0j} 'nin ne kadar değişeceğini gösterir.
- γ_{10} eğimlerin keseni yani $W_j=0$ olduğunda β_{1j} 'nin alacağı değerdir.
- γ_{11} eğimlerin eğimi yani W_j 1 birim arttığında β_{1j} 'nin ne kadar değişeceğini gösterir.

Artık Değerler Üzerine Varsayımlar

- u_{0j} ve u_{1j} , ikinci düzeye özgü artık değerlerdir.
- Normal regresyon çizgisinde olduğu gibi, okullar arasında şu varsayımlarda bulunacağız:
- Hem u_{0j} hem de u_{1j} normal olarak dağılır.
- Hem u_{0j} , hem de u_{1j} , 0 ortalamaya sahiptir.
- τ_{00} u_{0j} 'nin varyansıdır.
- τ_{11} ise u_{1j} 'nin varyansıdır.
- Ve u_{0j} ve u_{1j} 'in τ_{10} ile bir ilişkisi var.
- Spesifik olarak u Çok Değişkenli Normal Dağılıma sahiptir.

Bir HLM Modeli (2 Düzeyli)

Level-1

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij}) + r_{ij}$$

Level-2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$$

Düzey-2 modelinde Düzey-1'e ait regresyon katsayıları bağımlı değişken olarak kullanılır.

İki Düzeyli Model

- Düzey-2 modelinin HLM'ye özgü olan ve normal regresyon denkleminde ayıran iki yeni terim (u_{0j} ve u_{1j}) kullanması dikkat çekicidir.
- Regresyonda olduğu gibi hatalar artık bağımsız değil.
- Bu iki terim Düzey-1 ve Düzey-2 arasında bağımlılık olduğunu gösterir.

İki Düzeyli HLM Modelleri

- **Koşulsuz Model:** Null model (one-way random effects ANOVA model)-random intercept model
- **Ortalamalar Bağımlı Olarak Modeli:** Predicting mean school achievement (Means-as-Outcomes regression)
- **ANCOVA Modeli:** One-way ANCOVA with random effects
- **Rasgele Katsayı Modeli:** Random coefficients model
- **Kesenler ve eğimler Ortalama Olarak Modeli:** Intercepts and Slopes as outcomes

İki Düzeyli HLM Parametreleri

- İki düzeyli hiyerarşik modeller, üç tip parametre tahminini içerir. İlk parametre tipi sabit etkilerdir ve bunlar gruplar arasında farklılık göstermez. γ_{00} , γ_{01} , γ_{10} , γ_{11}
- İkinci parametre tipi, gruplar arasında değişiklik göstermesine izin verilen rasgele Düzey-1 katsayılarıdır (β_{0j} ve β_{1j}).
- Üçüncü parametre tahmin tipi, aşağıdakileri içeren varyans-kovaryans bileşenlerine ilişkindir: (1) Düzey-2 hata terimleri arasındaki kovaryans (τ_{01}); (2) Düzey-1 hata terimindeki varyans (σ^2) ve (3) 2. Düzey hata terimindeki varyans (τ_{00} ve τ_{11}).

HLM'de Varyans Bileşenleri

$$\text{Var}(r_{ij}) = \sigma^2; E(r_{ij}) = 0$$

$$\text{Var}(u_0) = \tau_{00}; E(u_{0j}) = 0$$

$$\text{Var}(u_1) = \tau_{11}; E(u_{1j}) = 0$$

$$\text{Var} \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} = T = \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}$$

$$\text{Cov}(r_{ij}, u_j) = 0$$

Sabit (Fixed) vs Rasgele (Random)

- **Sabit Etkiler**, bir değişken veya faktörün ayırık, amaçlı olarak seçilen veya mevcut değerlerini temsil eder.
 - Sabit etkiler bağımlı değişken üzerinde sabit etki gösterir.
 - Rasgele değişkenlik sadece bireylerin (Düzey 1) etkisi olarak ortaya çıkar.
 - Sadece kullanılan belirli değerlere genellenebilir.
- **Rasgele Etkiler**, bir değişken veya faktörün daha sürekli veya rasgele örneklenmiş değerlerini temsil eder.
 - Rasgele etkiler bağımlı değişken üzerinde değişken etki gösterir
 - Değişkenlik Düzey 1 ve Düzey 2'de gerçekleşir.
 - Değişkenliği inceleyebilir ve modelleyebilir
 - Değerler popülasyonuna genelleme yapabilir

Sabit (Fixed) Etkiler

Eğer

- Gruplar benzersiz varlıklar olarak kabul edilir
- Grup değerleri araştırmacı tarafından tasarım veya manipülasyon yoluyla belirlenirse
- Küçük j (<10) varsa gücü artırmak için

Sabit etkileri kullanınız.

Rasgele (Random) Etkiler

Eğer

- Daha büyük bir popülasyondan örneklem olarak kabul edilen gruplar varsa
- Araştırmacı grup düzeyi değişkenlerinin etkilerini test etmek istiyorsa
- Araştırmacı, grup düzeyindeki farklılıkları anlamak istiyorsa
- Küçük j (<10) varsa ve kestirimi geliştirmek isteniyorsa

Rasgele etkileri kullanınız.
$$\begin{bmatrix} U_{0s} \\ U_{1s} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix} \right)$$

HLM'de İstatistiksel Kestirim

- HLM'de üç parametre türü için istatistiksel kestirim farklı kestirim yöntemleri ile yapılabilir:
- **Tam Maksimum olabilirlik** (Full Maximum Likelihood)-FML
- **Kısıtlı Maksimum olabilirlik** (Restricted Maximum Likelihood)-REML*
- **Ampirik Bayes Kestirimi**-Empirical Bayes Estimation

*RML'nin özellikle j küçük olduğunda daha iyi tahminlere yol açması beklenir.

HLM'de İstatistiksel Kestirim

- İki parametre kestirimi tekniği (**ML** ve **REML**) yalnızca sabit etkinin tahmininde ve serbestlik derecelerinin kestirimde farklılık gösterir.
- Çok büyük örneklem boyutlarında REML ve ML'den elde edilen çok düzeyli model parametre tahminleri çok benzerdir.
- Bununla birlikte, uygulamalı araştırmanın tipik olduğu düşünülen örneklem boyutlarında, REML, rastgele etkiler ve sabit etki standart hataları hakkında daha kesin tahminler vermektedir (Snijders ve Bosker, 1999).
- ML rastgele etkiler tahminleri negatif yanlıdır.

HLM'de İstatistiksel Kestirim

HLM'de İstatistiksel Kestirim

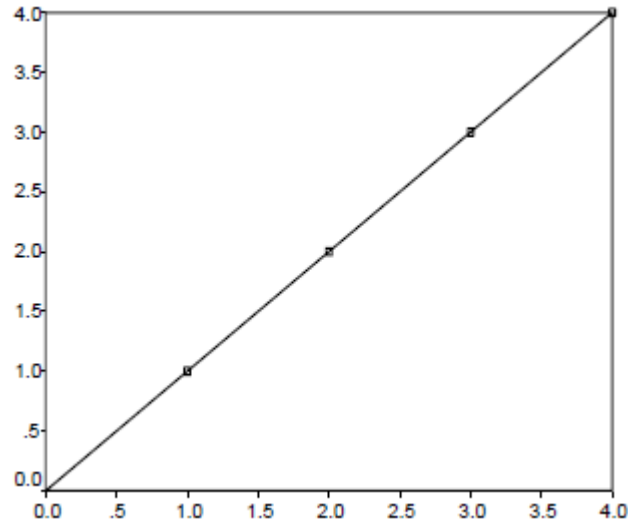
- Katsayıları ve standart hatalarının kestirimini
- Sabit etkiler ve rasgele etkilere (varyans bileşenleri) ait değerleri içerir.
- Manidarlık için z, t ya da ki-kare istatistiği hesaplanır. HLM programında t-istatistiği ve ki-kare istatistiği hesaplanır.
- İstatistiksel kestirimde dikkat edilmesi gereken diğer konular: merkezileştirme, parametre güvenirliği ve artık değerlerdir.

HLM'de İstatistiksel Kestirim

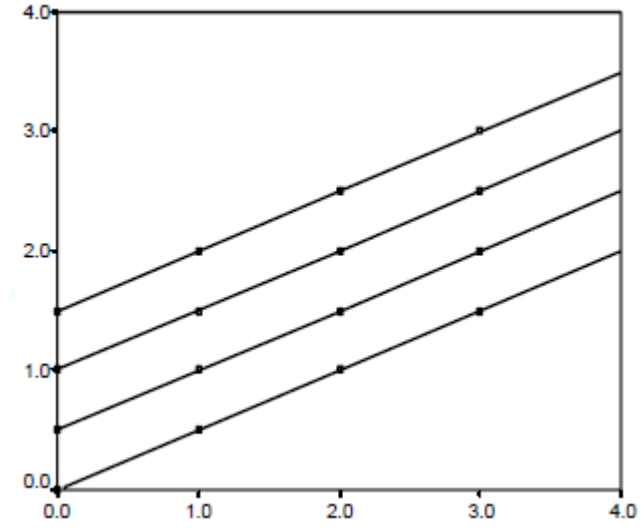
- HLM'de İstatistiksel Kestirim bazı algoritmalar vasıtasıyla yinelemeli analizler yapılarak gerçekleştirilir. HLM'de kullanılan algoritmalar:
- Expectation-Maximization (EM)
- Fisher scoring
- Iterative Generalized Least Squares (IGLS)
- Restricted IGLS (RIGLS)

Sabit (Fixed) vs Rasgele (Random)

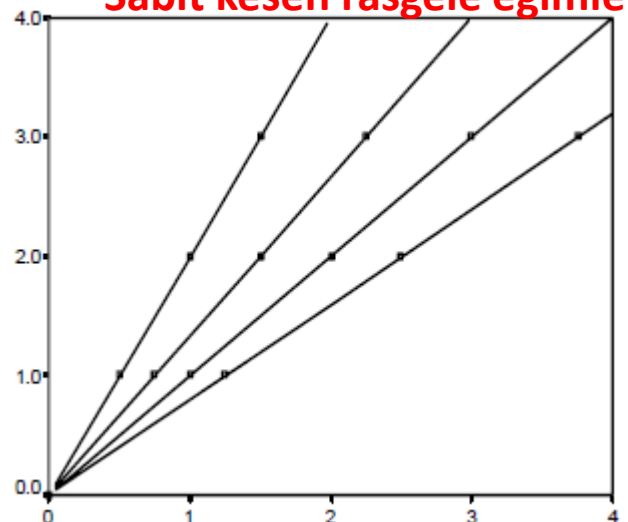
Sabit kesen sabit eğimler



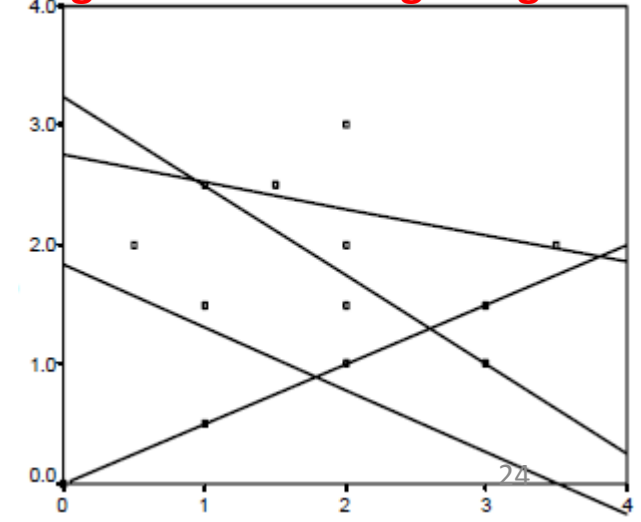
Rasgele kesenler sabit eğim



Sabit kesen rasgele eğimler



Rasgele kesenler rasgele eğimler



ANOVA Model

- Modellerin en basitidir. Koşulsuz (**unconditional**) model ya da **null model** olarak da adlandırılır.
- Toplam varyansı Düzey-1 ve Düzey-2 varyansları olmak üzere ikiye ayırır.

Düzey 1

$$Y_{ij} = \beta_{0j} + r_{ij}$$

Düzey 2

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

Birleştirilmiş eşitlik:

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

Bireysel farklılıklar-
Düzey 1 artık değeri

Genel ortalama

Dr. Sedat ŞEN

Grup sapması-Düzey 2 artık değeri

ANOVA Model

- Koşulsuz modelde varyans aşağıdaki gibi iki parçaya ayrıştırılır:

$$\text{Var}(Y_{ij}) = \text{Var}(u_{0j}) + \text{Var}(r_{ij})$$

$$= \tau_{00} + \sigma^2$$

Gruplar arası varyans

Gruplar içi varyans

ANOVA Model

- Bağımlı değişken varyansı ($\tau_{00} + \sigma^2$) toplamına eşittir . τ_{00} gruplar arası değişkenliği σ^2 ise gruplar içi değişkenliği temsil eder.
- **INTRACLASST CORRELATION** (ICC): Bu model sayesinde, Düzey 2'ye atfedilebilecek varyans oranını da hesaplayabiliriz (yani sınıf içi korelasyonu).
- ICC değeri(ρ) Düzey 2 varyansının toplam varyansa bölünmesiyle hesaplanır. 0 ile 1 arasında değişir. 1'e yaklaştıkça ikinci düzeyde varyasyonun çok olduğuna işaret eder. **Bu değer HLM ile OLS regresyonu arasında seçim yapmamıza yardımcı olur.**

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

Koşullu Modeller

- Mutlaka HLM analizine koşulsuz model ile başlamalı ve ICC değerini hesaplamalıyız. Daha sonra bu modele Düzey 1 ve Düzey 2’de ölçülen bağımsız değişkenler eklenebilir. Bu modeller koşullu (conditional) modeller olarak adlandırılır.
- Modele eklenen değişkenler de hem sabit (fixed) hem de rasgele (random) olabilir.
- Eklenen parametreye göre koşullu modeller **ANCOVA, Ortalamalar bağımlı, Rasgele katsayı, kesenler ve eğimler bağımlı** şeklinde adlandırılır.

ANCOVA Modeli

- ANOVA modeline bir bağımsız değişken eklemek, rasgele etkileşimli bir **ANCOVA** modeliyle sonuçlanır:

$$Y_{ij} = \beta_{0j} + \beta_1(X_1) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_1 = \gamma_{10}$$

- Rasgele etkili bir kesen değeri içerir.
- Sabit etkili yani bütün İkinci düzey elemanları arası sabit olan bir Düzey 1 değişkeni (X) içerir.

Ortalamalar Bağımlı (Means as Outcomes) Olarak

- Koşullu model oluşturmanın başka yolu da ANOVA modeline bir Düzey 2 bağımsız değişkeni eklemektir, bu da **means as outcomes** yani ortalamaların bağımlı değişken olduğu bir model ile sonuçlanır:

$$Y_{ij} = \beta_{0j} + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

- Düzey-2 değişkeni içerir ve düzey-1 kesenini düzey-2 değişkeninin bir fonksiyonu olarak modeller.

Açıklanan Varyans (Variance Explained)

- Her düzeyde açıklanan varyanslar aşağıdaki gibi hesaplanır:

R^2 at level 1 =

$$1 - (\sigma^2_{\text{cond}} + \tau_{\text{cond}}) / (\sigma^2_{\text{uncond}} + \tau_{\text{uncond}})$$

R^2 at level 2 =

$$1 - [(\sigma^2_{\text{cond}} / n_h) + \tau_{\text{cond}}] / [(\sigma^2_{\text{uncond}} / n_h) + \tau_{\text{uncond}}]$$

Where n_h = the harmonic mean of n for the level 2 units

$$(k / [1/n_1 + 1/n_2 + \dots 1/n_k])$$

Dr. Sedat ŞEN

Rasgele Katsayı Modeli

- Önceki modellerden ANCOVA modelinde eğim sabit etkiye sahipti. Şimdi eğim için rasgele etki eklediğimizde rasgele katsayı (**random coefficient**) modelini elde ederiz.
- Ek bir parametre, eğimlerin rasgele değişmesine neden olur.
- Hem kesen hem de eğimler artık gruptan gruba değişir.
-

$$Y_{ij} = \beta_{0j} + \beta_1(X_1) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

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

Kesen ve Eğimler Bağımlı Olarak

- Hem Düzey-1 hem de Düzey-2 değişkenini modele ekleyip hem kesen hem de eğim değerlerini rasgele etkilere sahip olacak şekilde eklediğimizde aşağıdaki modeli elde ederiz. Bu model **Intercept/Slopes as Outcomes** olarak adlandırılır.

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{.j}) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$$

Önceki modele benzer olarak, en az bir değişkenin bir fonksiyonu olarak kesen ve eğim değişkenliğini modelliyoruz.

Parametrelerin Testi

- γ_{00} ve γ_{10} parametreleri, Düzey-1 için kesen ve eğim katsayılarıdır ve sınıflar arasındaki ortalamalar alınarak hesaplanır. HLM, **t-testi istatistiği**ni kullanarak bu parametrelerin sıfırdan manidar düzeyde farklı olup olmadığını test eder.
- Kesen ve eğim varyanslarının (τ_{00} ve τ_{11}) sıfırdan manidar düzeyde farklı olup olmadığı da **χ^2 istatistiği ile test edilir**. SAS gibi programlar Wald-Z istatistiği kullanır.

Parametrelerin Testi

- Bir dizi katsayının aynı anda testi: Araştırmacılar, bir dizi yordayıcı değişkenin etkilerinin manidarlığını aynı anda test etmekle ilgilenebilirler.
- Bu, modelin, ilgilenilen değişken setiyle ve onsuz aynı veri setine yerleştirilmesi ve ardından bir olasılık oranı (LR) testi yapılması ile yapılabilir.
- **Likelihood ratio statistisc** bir ki-kare istatistiğidir ve karşılaştırılan 2 modelin deviance (sapma) değerleri arası farkın ve parametre sayıları arası farkın hesaplanmasıyla uygulanır.

Merkezileştirme (Centering)

- Regresyonda çoğunlukla kesenden çok eğimle ilgileniriz. Ama kesen için de bazı durumları dikkate almalıyız.
- Regresyon keseninin anlamı, açıklayıcı değişkenlerin değerleri ve ölçekleriyle belirlenir. Sürekli bir değişken (seviye 1 veya seviye 2'de) anlamlı bir 0 değerine sahip değilse, merkezileştirme önerilir.
- Bir regresyon modelinin düzey-1 keseninin 5 olduğunu varsayalım. Regresyon çizgisinin Y eksenini 5'te geçtiğini veya bağımlı değişkenin beklenen değerinin, modeldeki tüm açıklayıcı değişkenlerin 0 değeri olduğunda 5 olacağını söyleyebiliriz.
- Bazı açıklayıcı değişkenlerin gözlenen verilerde gerçekten 0 değeri olmaması durumunda, bu anlamlı olmayabilir.

Merkezeleştirme

- Örneğin, yetişkin deneklerin yaşı bir regresyon modeline dahil edildiğinde, 0 yaşına karşılık gelen bağımlı değişkenin değeri anlamsızdır.
- Keseni anlamlı hale getirmek için, merkezlenmiş olarak yeniden ölçeklenmiş veya dönüştürülmüş bir yaş ölçüsü kullanabiliriz; örneğin, örneklemdaki tüm katılımcılar için örneklemin ortalama yaşının gözlemlenen yaştan çıkarılması.
- Örneklemin ortalama yaşı 30 ve bireyin yaşı 40 ise, o zaman bireyin ortalama yaşının değeri $40 - 30 = 10$ olacaktır; bu, ortalama yaştan sapmanın bir ölçüsüdür.

Merkezileştirme

- Merkezileştirme özellikle etkileşime giren ana etki değişkenlerinin yorumlanmasını anlamlı hale getirir.
- Merkezileştirme HLM’de sadece anlamlı yorumlar yapılmasını sağlamaz aynı zamanda kestirimi hızlandırır ve yakınsama problemlerini azaltır.

Merkezeleştirme

- Düzey-1 ve düzey-2 yordayıcı değişkenlerini ayrı ayrı incelerken, merkezleme, regresyon katsayılarını değiştirmez, ancak kesen değerini değiştirir.
- Düzey-2 belirleyici değişkeni merkezlendiğinde, düzey-2 keseni bağımlı değişkeninin genel ortalamasına eşittir. Düzey-2 belirleyici değişkeni merkezlenmediğinde, düzey-2 keseni, düzey-2 yordayıcı değişkenleri sıfıra eşit olduğunda, bağımlı değişkeninin ortalama puanına eşittir.

Merkezileştirme

- Orijinal eşitlik

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

- Merkezileştirilmiş eşitlik

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{..}) + r_{ij}$$

Merkezileştirme temel olarak bir bağımsız değişkenden herhangi bir değerin çıkarılması yoluyla gerçekleştirilir.

Merkezileştirme

- Merkezileştirme yaparken X bağımsız değişkeninden çıkaracağımız sayıyı belirlerken karşımıza 4 durum çıkar:
- Orijinal X Metriği (yani merkezileştirme yapılmaz) kullanarak ya da
- i) **Genel ortalama etrafında** (Around the Grand Mean)
- ii) **Grup ortalamaları etrafında** (Around the Group Mean)
- iii) **Teorik bir değer etrafında** (Theoretical Location) merkezleme yaparak analizleri gerçekleştirmek.
- **Genel kural: eğer düzeyler arası etkileşimler tahmin ediliyorsa genel ortalama etrafında merkezleme yapılır.**

Orijinal Metrik

- Regresyon veya HLM’de merkezileştirme yapmadan orijinal metriği kullanabileceğimiz durumlar olabilir
- 0, yordayıcının orijinal ölçeğinde anlamlı bir nokta olduğunda merkezileştirme yapmamak mantıklı olabilir.
- Örneğin, 0 ila 14 gün arasında değişen eğitim süresi durumunda
- 0'ın plaseboyu temsil ettiği veya tedavi olmadığı bir ilacın dozunu temsil ettiği durumlarda orijinal metrik kullanılabilir
- **Fakat 200 ila 800 arasında değişen sınav puanları varsa orijinal metriği kullanmak mantıklı olmaz ve yorumlanamaz.**

Genel Ortalama Etrafında Merkezleme

- **Grand Mean Centering**
- Değişkenlerin genel ortalama etrafında merkezlenmesi demektir.
- Bağımsız değişken değerinden genel ortalama (grand mean) değerinin çıkarılması ile oluşturulur.
- Yordayıcılar (X 'ler) genel ortalamanın (M) sapması olarak ifade edilir: $(X_{ij} - M)$
- Genel ortalama etrafında merkezleme, tüm x_{ij} 'ler üzerindeki değerleri olan bir birey için y_{ij} 'nin beklenen değerine eşit bir kesen sağlar; örneklemdaki tüm bireylerde ortalama x_{ij} değerlerine eşittir.
- Hem Düzey-1 hem de Düzey-2 değişkenleri için kullanılabilir.

Grup Ortalamaları Etrafında Merkezleme

- Centering around group means
- Bağımsız değişkenin grup ortalamalarına göre merkezlenmesini içerir.
- Her gruptaki bireylerin değerinden kendi grup ortalamasının değeri çıkarılır. $(X_{ij} - \bar{X})$
- Bireysel puanlar grup ortalamalarına göre yorumlanır.
- Grup ortalaması etrafında merkezleme Düzey-1 değişkenleri için önerilir.

Grup vs Genel Ortalama Etrafında

- Çoğu durumda, düzey-1 belirleyici değişkeni, düzey-1 ve düzey-2 belirleyici değişkenlerinin etkilerini bağımsız olarak incelemek ve kesin noktalarının daha doğru tahminlerini elde etmek için grup merkezli bir değişken olarak girilir.
- Tahmini değişkenleri ayrı ayrı analiz etmekle ilgilenmezsek, birinci düzey merkezlenmiş değişkeni seçeriz (örneğin, bir değişkeni diğer değişken için kontrol ederken test eden bir ANCOVA analizi genel ortalama etrafında merkezlemeyi gerektirir).
- **LEVEL-1 DEĞİŞKENİ GRUP ORTALAMASI ETRAFINDA**
- **LEVEL-1 ve LEVEL-2 DEĞİŞKENİ GENEL ORTALAMA ETRAFINDA MERKEZLENEREK EKLENEBİLİR.**

Teorik Bir Değer Etrafında Merkezeleme

- Diğer bir merkezileştirme yöntemi teoriye dayanan anlamlı bir nokta seçmemize izin verir.
- Çalışmamızda topladığımız bağımsız değişkenler ilgili bir eşik/norm değeri olduğunu biliyorsak.
- Bu değeri bağımsız değişken değerlerinden çıkartarak, grubun o noktayla ilişkili olduğu yeri ayarlayabiliriz.

Güvenirlilik Hesaplaması

- Son olarak, Okul j 'lerin ortalamasının kestiriminin güvenirliğini aşağıdaki formülle hesaplayabiliriz.

$$\hat{\lambda}_j = \textit{reliability}(\bar{Y}_{.j}) = \frac{\hat{\tau}_{00}}{(\hat{\tau}_{00} + (\sigma^2 / n_j))}$$

Bağımlı Değişkende Ne Kadar Varyasyon Bulunmaktadır?

- Bu soruyu cevaplayabilmek için boş model (empty model) oluşturulur. ANOVA modelde kesenin sabit olduğu modeldir.
- Yordayıcı/bağımsız değişken eklenmez
- Rasgele etki eklenmez, sabit etkili kesen değeri ile oluşturulan model ile bu soru cevaplanır
- Kesen = ortalama bağımlı değişken puanına eşittir
- Hata = bağımlı değişkenin varyansıdır

Bağımlı değişkende ikinci düzey değişkenine atfedilebilecek bir varyasyon var mı?

- Bu soruyu cevaplayabilmek için boş modelde kesen değeri de rasgele etkiler olarak tahmin edilir yani ANOVA modeli oluşturulur.
- Rasgele kesen varyansı sıfırdan büyük çıkarsa, «Bağımlı değişkende ikinci düzey değişkenine atfedilebilecek bir varyasyon var mı?» sorusunun cevabı evet olur.
- Hata varyansının yanında kesene ait varyans tahmin edilir.

Hangi Model?

- Buraya kadar bir çok model kuruldu. Başka modeller de kurulabilir.
- Bu modellerden hangisi ile yola devam edeceğimize model uyumuna bakarak ve model karşılaştırması yaparak karar vereceğiz
- Model uyumu ve karşılaştırması için **deviance** değeri, serbestlik derecesi, bilgi kriteri gibi indeksler kullanılır.

Model Uyumu ve Model Karşılaştırma

- -2loglikelihood (**-2LL**) değeri (Deviance/sapma değeri olarak da bilinir) bir HLM modelinin veriye ne kadar iyi uyum gösterdiğine işaret eder.
- Düşük -2LL değerleri daha iyi uyuma işaret eder.
- Deviance değerinin yanında **-2LL** değeri kullanarak hesaplanan **AIC** ve **BIC** gibi **Bilgi Kriteri İndeksleri** de bir modelin ne kadar iyi uyum gösterdiğine işaret eder. Aynı şekilde düşük değerler daha iyi uyuma işaret eder.

Model Uyum İndeksleri

- HLM'de model karşılaştırmasında olasılık oranı ya da AIC/BIC değerleri kullanılır.

- Chi-square $-2 * \ell$
- Akaike Information Criteria

$$AIC = -2 * \ell + 2K$$

- Bayesian Informaiton Criteria

$$BIC = -2 * \ell + K * Ln(N)$$

Model Uyumu ve Model Karşılaştırma

- Aynı verilere uyması için farklı modeller kullanıldığında, modeller yuvalanmış veya yuvalanmamış olabilir.
- İç içe modeller için, bir model diğerinin alt modelidir, olasılık oranı (LR) testi model karşılaştırması için kullanılabilir. İki model arasındaki sapma istatistiklerinde veya -2LL arasındaki fark LR testi için kullanılır.

$$\begin{aligned} D_A - D_0 &= (-2LL_A) - (-2LL_0) \\ &= -2 \ln \left(\frac{L_A}{L_0} \right) \end{aligned}$$

Model Uyumu ve **Model Karşılaştırma**

- Yuvalanmayan modelleri karşılaştırırken LR testi çalışmaz.
- AIC, AICC ve BIC gibi bilgi kriterleri istatistikleri, iç içe geçmiş modellerin yanı sıra gerçek olmayan modellerin karşılaştırılmasında da uygundur.
- AIC, AICC ve BIC değerlerinin sıfıra yakın bir model olması tercih edilir. Ne kadar düşük olursa modelin veriye o kadar iyi uyum gösterdiği söylenebilir.
- SAS gibi HLM programlarında rapor edilir fakat HLM programında sunulmaz.

Model Uyumu ve Model Karşılaştırma

- ML ve REML tahmin edicilerin farklı çok düzeyli model karşılaştırmaları için uygun şekilde kullanılması gerekir.
- Modeller aynı rasgele etkilere sahipken, ancak farklı sabit etkilere varsa, model karşılaştırması için kullanılan model uyumu istatistikleri ML yöntemi altında tahmin edilmelidir.
- REML yöntemi, aynı sabit etkilere sahip, ancak farklı rasgele etkilere sahip modeller için kullanılmalıdır.
- Genel olarak konuşursak, ML model karşılaştırmaları için daha esnektir. Modeller hem farklı sabit etkilere, hem de farklı rasgele etkilere sahipse ve rasgele etkilere iç içe geçmişse ML önerilir. Aynı ilkeler AIC, AICC ve BIC gibi bilgi ölçütleri için de geçerlidir.
- **Ancak, en iyi model ML ile belirlendiğinde, “nihai” modeli tahmin etmek ve REML sonuçlarını rapor etmek için REML yöntemini kullanmalıyız.**

Model Karşılaştırması Örnek

	Deviance	Parametre sayısı
Model1	2,202.5	2
Model2	2,147.9	4

- Tablodaki değerlere göre Model2 mi Model1 mi seçilmeli?
 - Test statistic: $2,202.5 - 2,147.9 = 57.6$
 - Degrees of freedom = 2 (new parameters γ_{10} and γ_{01})
 - P-value: < 0.0001
- Karar: Model2 tercih edilmeli ($p < 0.05$).

Model Kurma Süreci

- Model oluşturma süreci çok temel bir modelle (koşulsuz modelle) başlar ve daha sonra her düzey için yordayıcılar ekleyerek devam eder.
- Her biri farklı bir soruya cevap vermeye çalışan bir dizi model oluştururuz.
- Süreç sonunda kullandığımız nihai modeli değerlendirip değişkenlerimizin niteliğine ilişkin çıkarımlar yaparız.

Model Kurma Süreci

1. Koşulsuz modeli test ediniz

2. Düzey 1 yordayıcılarını ekleyiniz

- gruplar arası varyans olup olmadığına karar veriniz
- eğer yoksa parametreleri sabit yapın
- manidar olmayan yordayıcıları çıkarıp çıkarmamaya karar verin
- Deviance testi yapın ve R-kare hesaplayın

3. Düzey 2 yordayıcılarını ekleyiniz

- Manidarlığı değerlendirin
- Deviance testi yapın ve R-kare hesaplayın

Model Kurma Süreci

4. Düzey-1 rasgele eğimlerinin testi: Eklediğimiz Düzey-1 değişkenlerinin eğimlerinin rasgele etkiye sahip olup olmadığını test etmeliyiz ona göre belirleme yapmalıyız. Burada yapacağımız yanlış belirleme sabit etkilere ait değerleri de olumsuz yönde etkileyecektir.

5. Çapraz düzey etkileşimlerinin eklenmesi: Bazı düzey-1 eğim katsayılarının rasgele olduğu tespit edilirse, makro modellerde grup düzeyindeki açıklayıcı değişkenlerin bir fonksiyonu olarak ele alınarak grup arasındaki grup varyasyonlarının açıklanması gerekir.

Model Kurma Süreci

- İdeal olarak, bahsedilen beş adımı takip ederek, tasarlanan çalışmamız için “nihai” bir model hakkında bazı fikirler ortaya çıkacaktır.
- Daha iyi bir model uyumu (örneğin, daha küçük sapma, AIC ve BIC), mutlaka bir modelin tatmin edici olduğu anlamına gelmez.
- Model seçimi hem istatistik hem de teori odaklı bir süreçtir.
- Model geliştirmenin amacı, yalnızca verilere uygun değil aynı zamanda yorumlanabilir sonuçlar da sağlayan sağlam temelli bir model bulmaktır.

Final Modelin Parametreleri

- Model karşılaştırmaları ışığında karar verdiğimiz final modele ait sabit ve rasgele etkiler parametreleri REML kestiricisi ile kestirilir ve uygun şekilde raporlanır ve sırasıyla yorumlanır.