

## Örtük Markov Model: Latent GOLD 5.1'e Genel Bakış\*

### Latent Markov Model: An Overview of Latent GOLD 5.1

Duygu GÜNGÖR \*\* Selva ÜLBE\*\*\*

#### Öz

Hem eğitim hem de psikoloji alanında gelecekteki durumların geçmişe bağlı olduğu düşünüldüğünde zamana bağlı değişimin nasıl gerçekleştiği önemli bir araştırma alanıdır. Bu kapsamda örtük Markov analizi gözlenebilir değişkenler kullanılarak, kategorik örtük değişkenler hakkında ileriye yönelik tahminleme yapmak için başvurulan istatistiksel bir yöntemdir. Teknolojik gelişmelerin hızlanmasıyla örtük Markov modelleri çeşitli özgün yazılımlarla incelenebilmekte ve bu yazımlar arasından seç-tıkla özelliği ile Latent GOLD programı öne çıkmaktadır. Bu yazının amacı boylamsal kategorik verilerin analizinde kullanılan ve örtük Markov modelini test eden Latent GOLD 5.1 programını uygulamacılara tanıtmaktır. Bu kapsamda öncelikle örtük Markov modelin kısa bir açıklaması yapılmıştır. İkinci olarak Latent GOLD 5.1'in genel özelliklerinden bahsedilmiştir. Son olarak bu program yardımıyla analizlerin nasıl yapılacağına bir örnek göstermek için Monte Carlo simülasyonu kullanılarak üretilen bir veri seti üzerinden model testi gerçekleştirilmiş ve çıktıların yorumlanmasıyla ilgili bilgiler sunulmuştur.

*Anahtar Kelimeler:* Örtük Markov model, Latent GOLD 5.1, boylamsal veri

#### Abstract

In the field of both education and psychology, if future situations are thought to be related to the past, how time-based change is taking place is an important research area. In this context, latent Markov analysis is a statistical method applied to make prediction about categorical latent variables by using observed variables. With acceleration of technological developments, latent Markov models can be examined with various special software, among which Latent GOLD program with select-click feature stands out. The aim of this paper is to introduce the Latent GOLD 5.1 program which is used to analyze the longitudinal categorical data and to test the latent Markov models to users. In this context, firstly a brief explanation of latent Markov model was given. Secondly, the general features of Latent GOLD 5.1 were mentioned. Finally, to give an example of how the analyses would be carried out via this program, a model test was conducted on a data set generated from Monte Carlo simulation and the information about the interpretation of the outputs was provided.

*Key Words:* Latent Markov model, Latent GOLD 5.1, longitudinal data

#### GİRİŞ

Temellerini örtük sınıf analizinden alan örtük Markov modeller hem gözlenen hem de örtük değişkenin kategorik olduğu araştırma verilerinin boylamsal analizi için kullanılmaktadır. Psikoloji alanında özellikle madde kötüye kullanımı ile ilgili çalışmalar (Cosden, Larsen, Donahue ve Nylund-Gibson, 2015; Guo, Aveyard, Fielding ve Sutton, 2009; La Flair vd., 2013; Lanza ve Bray, 2010) ve eğitim alanı (Vermunt, Langeheine ve Bockenholt, 1999) örtük Markov modelin kullanıldığı uygulamalı alanlara örnek verilebilir. Ölçme modeli ve yapısal model olarak adlandırılan iki alt bölümden oluşan örtük Markov modellerinde üç temel parametre tahminlenmektedir: örtük statü olasılıkları, madde koşullu olasılıkları ve örtük geçiş olasılıkları (ayrıntılı bilgi için bkz: Collins ve Lanza, 2010; Lanza ve Bray, 2010; Vermunt, Tran ve Magidson, 2008). Bu parametrelerden ilk ikisi örtük sınıf analizindeki parametrelerle doğrudan ilişkilidir. Örtük sınıf analizinde örtük sınıf olasılığı olarak adlandırılan parametre, örtük Markov modellerinde sınıfın dinamik yapısını tanımlamak amacıyla genellikle örtük statü olasılığı kavramı ile adlandırılmaktadır (Collins ve Lanza, 2010).  $a=$

\* Bu çalışmanın bir kısmı TÜBİTAK 2219-Yurtdışı Doktora Sonrası Araştırma Bursu ile desteklenmiştir.

\*\* Yrd. Doç. Dr., Dokuz Eylül Üniversitesi, Edebiyat Fakültesi, İzmir-Türkiye, e-posta: duygu.gungor@deu.edu.tr

\*\*\* Araş. Gör., Dokuz Eylül Üniversitesi, Edebiyat Fakültesi, İzmir-Türkiye, e-posta: selva.ulbe@deu.edu.tr

1, 2, ....., A Zaman-1'deki örtük statüler olsun. Bu durumda Zaman-1'deki örtük statü olasılığı  $\pi_a^A$  olarak gösterilebilir.  $\sum \pi_a^A = 1$  olacağından A-1 tane parametre tahminlenecektir (Collins ve Lanza, 2010). Koşullu olasılıklar ise faktör analizindeki faktör yüklerine benzemekte ve tıpkı örtük sınıf analizindeki gibi yorumlanmaktadır. Örtük sınıf analizinden farklı olarak, örtük Markov analizinde her z zamanı için farklı koşullu olasılıklar hesaplanabilmekte ve bu olasılıklar örtük statüler hakkında bilgi vermektedir. A, Zaman-1'deki örtük statüler, B, Zaman-2'deki örtük statüler, D, Y, G ise hem Zaman-1 hem de Zaman-2 için gözlenen değişkenler olmak üzere, D değişkeninin  $i=1, 2, \dots, I$ ; Y değişkeninin  $j=1, 2, \dots, J$  ve G değişkeninin  $k=1, 2, \dots, K$  tane tepki kategorisi olsun. Bu durumda  $a$  örtük statüsünde olan bir bireyin/gözlemin Zaman-1'deki koşullu olasılıkları;  $\pi_{ia}^{I/A}$ ,  $\pi_{ja}^{J/A}$ ,  $\pi_{ka}^{K/A}$ , Zaman-2'deki koşullu olasılıkları ise  $\pi_{ib}^{I/B}$ ,  $\pi_{jb}^{J/B}$ ,  $\pi_{kb}^{K/B}$  şeklinde olacaktır. Zaman-1'de A örtük statüsünde olan bir bireyin/gözlemin Zaman-2'de B örtük statüsünde olma koşullu olasılığı ise  $\pi_{ba}^{B/A}$  şeklinde gösterilir ve bu koşullu olasılıklar Zaman-1'den Zaman-2'ye örtük statüler arasındaki örtük geçiş oranını temsil eder. Z zamanında alınan ölçümlerde bir kişi ancak bir örtük statüde yer alabilmektedir ve örtük statü olasılıklarının toplamı z zamanı için 1'e eşittir. Ancak z+1. zamanda kişiler, A örtük statüsünden B örtük statüsüne geçiş yapabilir ve statülerin yaygınlığı z. ve z+1. zamanda farklılaşabilir.

Son yıllarda artış gösteren örtük Markov model çalışmaları Mplus (Muthen ve Muthen, 1998-2007) ve Latent GOLD (Vermunt ve Magidson, 2013) gibi ticari programlarla ve PROC LTA (Lanza ve Collins, 2008) ve LMest (Bartolucci ve Pandolfi, 2016) gibi özgür yazılımlarla yürütülebilmektedir. Bu yazının amacı Latent GOLD 5.1 programı kullanılarak örtük Markov analizinin yapılması ve çıktıların yorumlanması konusunda uygulayıcılara bilgi sunmaktır. Bu programla pek çok olası ve karmaşık model test edilebilmekle birlikte, bu yazıda programı anlaşılır kılmak için birinci düzey örtük Markov model üzerine odaklanılmıştır.

### Latent GOLD

Latent GOLD, Statistical Innovations Incorporated'e bağlı olarak Vermunt ve Magidson (2016) tarafından geliştirilmiş ticari bir istatistik paket programıdır. Program şu an için yalnızca Windows donanımlı bilgisayarlarda kullanılabilir. Basic, Advance ve Syntax versiyonları bulunan program, temel olarak örtük değişken modellerin parametrelerinin tahmin edilmesinde geniş alternatifler sunmaktadır. Bu üç farklı versiyon farklı şekilde ücretlendirilmektedir. Akademik kullanıcılar yıllık (345\$) ya da sınırsız (795\$) lisans seçeneklerini kullanabilirken her bir lisansın üç bilgisayar için kullanılmasına izin verilmektedir. Öğrencilere daha ekonomik çözümler sunulurken, herhangi bir akademik lisansa sahip olan kullanıcılara 2017 itibarıyla 25\$'lık sınıf kullanımı için lisanslama seçeneği de sunulmaktadır. <http://www.statisticalinnovations.com> sitesinden programa ait güncel bilgilere ulaşılabilir; programın deneme (demo) versiyonu ücretsiz olarak indirilip incelenebilir. Deneme versiyonu ile yalnızca programda bulunan örnek veriler üzerinden çalışılabilmektedir. Ayrıca aynı sitede programa ait kullanıcı kılavuzu, teknik bilgiler ve sentaks modülü kılavuzu olmak üzere üç ayrı kitap bulunmaktadır.

Latent Gold 5.0'dan daha önceki versiyonlarda sentaks modülü ile yapılabilen Markov analizleri yeni sürümle birlikte Latent GOLD 5.0 ve Latent GOLD 5.1 Advanced modülü ile de yürütülebilmektedir. Bu gelişmeyle birlikte modeller, seç-tıkla şeklinde kullanıcı dostu olarak da yürütülebilmektedir.

### Veri Girişi

Markov analizlerinde özellikle veri girişinin doğru şekilde yapılması çok önemlidir. Klasik analizlerde boylamsal verilerin girişi her bir katılımcı için satır bazında yapılmakta iken örtük Markov analizlerinde ise verilerin sütun bazlı girilmesi gerekmektedir. Bu noktada boylamsal çalışmalarda fazlaca bulunan kayıp verilerin veri setinde mutlaka tanıtılması gerekmektedir. Veri seti

hazırlanırken her gözleme ve zaman dilimine ait değerler iki sütunda tanımlanmalıdır. Tablo 1'de üç gözlenen iki kategorili değişkenin bulunduğu ve dört zaman diliminde ölçüm alınmış örnek bir veri seti sunulmuş, olası kayıp değerler (99) gösterilmiştir.

Tablo 1. Örnek Veri Seti

Denek No	Zaman	Madde_1	Madde_2	Madde_3
1	1	0	1	1
1	2	0	0	1
1	3	99	99	99
1	4	99	0	1
2	1	1	1	1
2	2	1	1	1
2	3	1	1	1
2	4	0	0	1
.	.	.	.	.
.	.	.	.	.
n	4	.	.	.

Program, veri setini okurken her gözlem ve her zaman dilimi için verinin girilmiş olduğunu varsayarak modeli tahminlemektedir. Bu nedenle veri girişi uygun şekilde yapılmazsa sonuçlar yorumlanabilir olmayacaktır ya da yanlış şekilde karşımıza çıkacaktır (Vermunt ve Magidson, 2013). Veri setinde denek numaraları ile ölçüm alınan zaman tutarsız şekilde girilirse uygulayıcılar, boylamsal profilde ölçüm alınan zaman sayısından daha fazla zamanın tanımlı olması gibi bir sorunla karşılaşabilirler.

### Simülasyon Verisinin Özellikleri

Bu yazının temel amacı, Latent GOLD programı ile örtük Markov modellerinin nasıl yürütülebileceğine ilişkin bir örnekle birlikte; kullanıcılara, program hakkında genel bir bilgi sunmaktır. Latent GOLD programının üstün özelliklerinden biri de çoklu simülasyon verilerini hızlı bir şekilde üretmeye ve tek bir komut ile aynı anda çok sayıda veri setini analiz etmeye olanak sağlamasıdır. Her ne kadar simülasyon çalışmalarının nasıl yapıldığı bu yazının kapsamında olmasa da kullanılan örnek veri, Latent GOLD programında Monte Carlo simülasyon yöntemi kullanılarak üretilmiştir. Veri seti oluşturulurken daha önce yapılmış çalışmalar (Collins ve Wugalter, 1992; Gudicha, Schmittmann, ve Vermunt, 2015; Vermunt, 2010) temel alınarak iki ayrı zaman dilimi, iki kategorili beş madde, 200 örneklem büyüklüğü kullanılarak 1000 replikasyon sonucu elde edilmiş simülasyon verileri için;

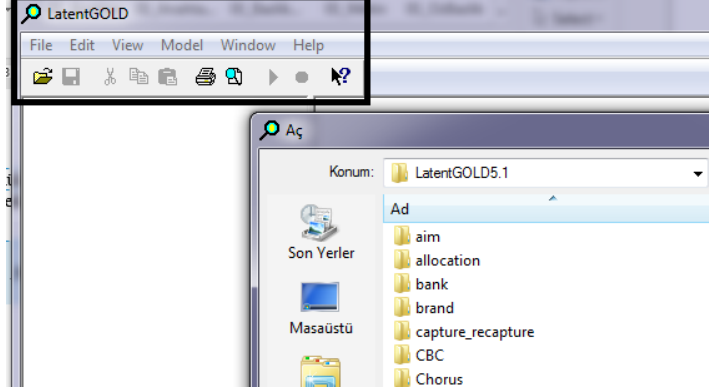
- eşit örtük statülerine sahip (.50, .50) iki örtük statü tanımlanmıştır.
- İki kategorili beş maddenin koşullu olasılıkları birinci örtük statüde olanlar için .90, .10 ve ikinci örtük statüde olanlar için .10, .90 olarak değişimlenmiştir.
- örtük geçiş matrisi ( $A^t$ ) aşağıdaki gibidir;

$$A^{(t)} = \begin{bmatrix} .80 & .20 \\ .10 & .90 \end{bmatrix}$$

Örneğin bu veri seti, eğitim araştırmalarında sınıfın iki farklı yetenek düzeyinde (birinci örtük statü düşük yetenek düzeyi, ikinci örtük statü yüksek yetenek düzeyi gibi) olduğu durumlarda iyi bir eğitim sonucu yüksek yetenek düzeyindeki grup yeteneklerini korurken, diğer grubun da aynı yetenek düzeyine ulaşmayı başardığı durumlarda karşımıza çıkabilir.

### Verinin Programa Tanıtılması

Latent GOLD programı .sav, .dat, .csv, .txt, .rsp ve .cho uzantılı veri setlerini okuyabilmektedir. Veri setinin kayıtlı olduğu klasör, programın dosya aç kısmı tıklanarak Şekil 1’de sunulduğu şekilde açılmaktadır.



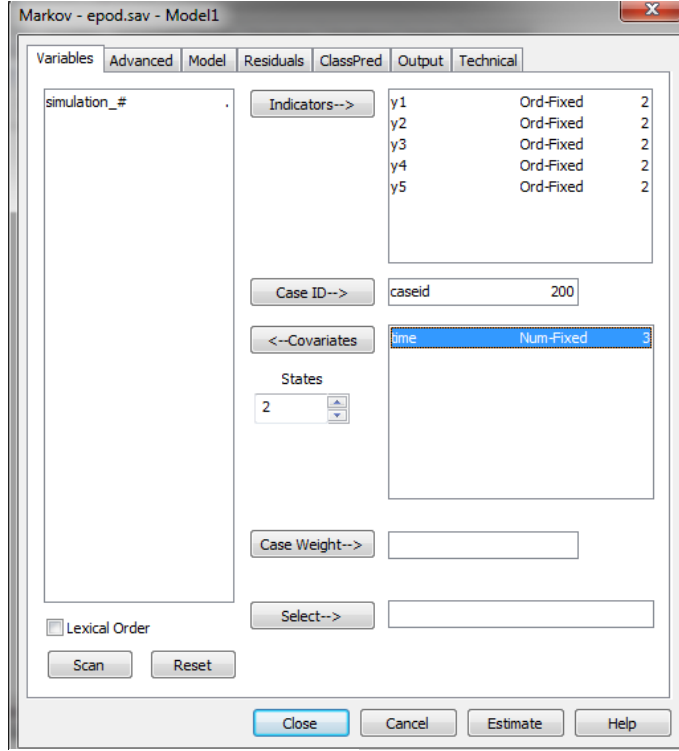
Şekil 1. Veri Setinin Tanıtılması

Programda çeşitli modeller test edilecekse uygun klasörlerin açılması önerilmektedir. Çünkü analiz sonuçları, veri setinin olduğu klasöre otomatik olarak kaydedilmektedir. Aynı klasörde aynı isimle analiz ya da çıktı dosyası kaydedilirse son işlem önceki işlemin üzerine yazılacak ve bu durum bulgu kayıplarına yol açacaktır.

### Örtük Markov Modelin Tahminlemesi

Veri seti tanıtıldıktan sonra veri setinin adı -bu örnekte epod.sav- ve altında yer alan Model 1 yazısı veri setinin tanıtıldığı pencerede çıkacaktır. Model tanıtımı için iki alternatif bulunmaktadır. Birinci alternatifte Model 1 yazısının üzerine gelerek sağa tıkladığında yürütebilecek analizlerin adı çıkacaktır. İkincisinde ise pencerenin üstünde bulunan Model yazısı tıkladığında yine aynı analiz seçenekleri karşımıza çıkmaktadır. Burada “Markov” seçilirse gösterge değişkenlerinin, gözlemlerin tanımlanabileceği Caseid ve kovariate değişkenlerin sürüklenerek taşınabileceği pencere açılmaktadır. “Indicators” kısmına gösterge değişkenler sürüklenir. Gösterge değişkenler bu örnekteki gibi iki kategorili değilse ölçek düzeyi doğru şekilde tanıtılmalıdır. Ancak iki kategorili değişkenler için sınıflayıcı ve sıralayıcı seçimi parametre tahminlerini etkilememektedir. “Case ID” kısmına her gözleme ait etiket değeri, “Covariates” kısmına da olası ortak değişkenler (covariate) sürüklenebilir (Şekil 2). Bu değişkenler cinsiyet, grup, yaş gibi zamana göre sabit kalan değişkenler (time constant) olabileceği gibi iş durumu, madde kullanımı, başarı gibi her zaman diliminde değişebilecek değişkenler de olabilir (time varying covariates). Boylamsal çalışmalarda en önemli ortak değişken zamanın kendisidir. Zaman değişkeni bu kısma atılırsa geçiş olasılıkları her zaman dilimi için ayrı ayrı tahminlenirken (zaman-heterojen model) zaman değişkeni ortak değişken olarak kullanılmazsa geçiş olasılıkları her zaman diliminde eşit olacak şekilde hesaplanmaktadır (zaman-homojen model).

Model tanımlanırken belirlenmesi gereken en önemli noktalardan biri statü sayısıdır. Uygulamada her zaman dilimi için örtük statü sayısı ayrı ayrı inceleyebilir ya da araştırmacılar önsel olarak bu bilgiye sahip olabilirler. Bu örnekte önsel olarak statü sayısının iki olduğu bilindiği için “States” iki olarak belirlenmiştir. Ancak statü sayısı bilinmiyorsa farklı sayılardaki modeller iç içe test edilip veriye en iyi uyum sağlayan model seçilebilir.

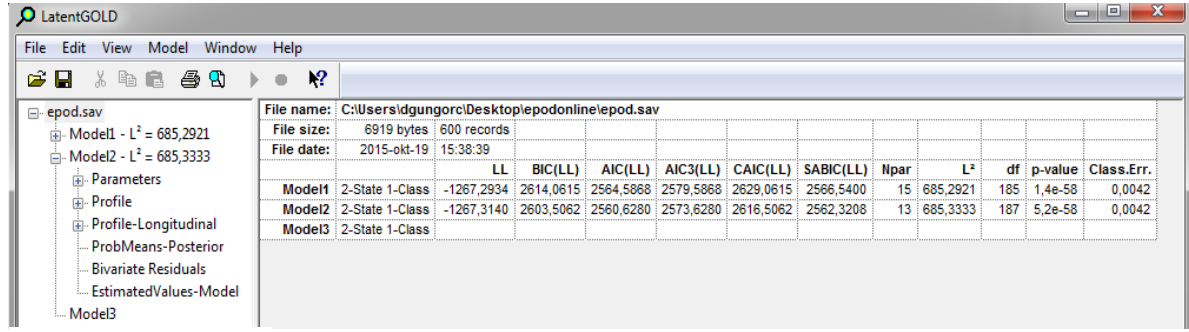


Şekil 2. Örtük Markov Model

Estimate sekmesine tıklandığında “WARNING: Time varying covariates have been removed from Model for Initial States: time” uyarısı çıkacaktır. Bu ilk zaman dilimi için zaman değişkeninin etkisinin tahminlenmeyeceği anlamına gelmektedir.

### Model Seçimi

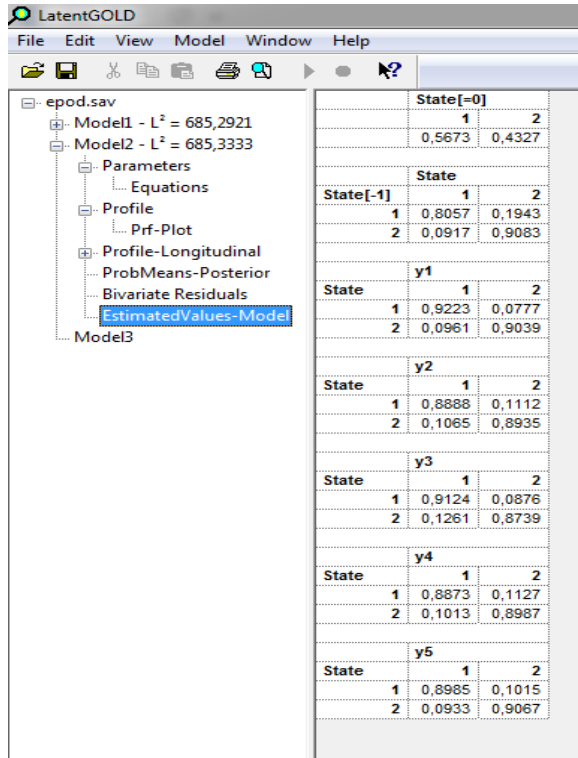
Örtük Markov modellerinde model seçim sürecinde, veriye uyum sağlayan en sade (parsimony) yani en az parametre tahminlenen model aranmaktadır. Bu koşulları sağlayan modeli bulmak için Pearson ki-kare ( $\chi^2$ ), en çok olabilirlik ki-kare oranı (likelihood ratio chi square) ( $L^2$ ) gibi istatistikler ve Akaike bilgi kriteri (Akaike Information Criteria) (AIC), Bayesian bilgi kriteri (Bayesian Information Criteri) (BIC) gibi bilgi kriterleri kullanılabilir (McCutcheon, 2002). Bilgi kriterleri kullanılarak yapılan model seçim süreci diğer modeller arasından en iyisini seçmeye yöneliktir ve hangi bilgi kriteri kullanılırsa kullanılsın kriterin aldığı daha düşük değer daha iyi model anlamına gelmektedir. Alternatif modeller farklı örtük statü sayısına sahip modeller olabileceği gibi örnekte sunulduğu şekilde parametrelere kısıt konulmuş çeşitli modellerden birini seçmeye de yönelik olabilir. Örnekte Model-1 geçiş olasılıklarının her zaman dilimi için ayrı ayrı tahminlendiği zaman-heterojen model, Model-2 ise zaman-homojen modeldir. Şekil 3'te görülebileceği gibi tüm bilgi kriterleri Model 2'de daha düşük değeri almıştır. Bu durumda Model 2'nin veriye daha iyi uyum sağladığını söylemek mümkündür. Bu seçim işlemi yapıldıktan sonra veriye en iyi uyum sağlayan modele ait parametre değerleri raporlanıp yorumlanabilir.



Şekil 3. Model Seçimi

### Örtük Markov Model Parametre Değerlerinin Yorumlanması

Seçilen modele ait parametre tahminleri, “Profile” ve “EstimatedValues-Model” sekmeleri tıklanarak incelenebilir. Şekil 4’te “EstimatedValues-Model” sekmesi tıklandığında açılan pencere gösterilmektedir.



Şekil 4. Parametre Tahminleri

State [=0] başlığının altında sunulan değerler, ölçüm alınan son zaman dilimindeki örtük statü olasılıklarını göstermektedir. Buradaki değerlere göre bir bireyin/gözlemin araştırmanın sonunda örtük statü-1’de olma olasılığının .57 ve örtük statü-2’de olma olasılığının .43 olduğu söylenebilir. Bir alttaki satırlarda sunulan olasılık değerleri ise örtük geçiş olasılıklarıdır (Seçilen model zaman-homojen model olduğu için t ve t-1 zamanlar arasındaki geçişin ölçüm alınan zaman sayısından bağımsız olarak aynı olduğu not edilmelidir). Bu matris şu şekilde yorumlanabilir; birinci zamanda örtük statü-1’de olan bireyin/gözlemin ikinci zamanda da örtük statü-1’de olma olasılığı .80’dir. Bu durumda örtük statü-1’den örtük statü-2’ye geçiş olasılığı da basitçe .20 olarak söylenebilir. Örtük statü-2’de gözlemlerin aynı statüde kalma olasılığı ise .90 ve örtük statü-1’e geçiş olasılıkları .10

olarak hesaplanmıştır. Son olarak sunulan olasılıklar ise madde tepki olasılıklarıdır ve faktör yükleri gibi yorumlanıp örtük statüler hakkında bilgi sunarlar.

## SONUÇ

Zamana bağlı değişim hem eğitimciler hem de psikologlar için önemli bir araştırma alanıdır. Değişimin incelenmesinde araştırma sorusuna yetenek, bilgi düzeyi, klinik tanı gibi örtük değişkenler dahil olduğunda ve veri seti kategorik olduğunda tek alternatif örtük Markov model kullanmaktır. Teknolojik gelişmelerle birlikte hız kazanan örtük Markov çalışmalarında Latent GOLD 5.0 ve 5.1 Advanced modül programları seç-tıkla özelliği ile öne çıkmaktadır. Diğer programların sadece sentaks ile çalıştırılabildiği göz önünde bulundurulduğunda Latent GOLD'un daha kullanıcı dostu olduğunu söylemek mümkündür. Ancak araştırmacılar daha karmaşık modelleri test etmek isterse bu programlarda da sentaks yazmaları gerekmektedir. Latent GOLD büyük veri setlerini hızlı bir şekilde analiz edebilmektedir. Ancak benzetim çalışmaları yapan araştırmacıların örneğin 1000 tekar sayısı ve iç içe on model test etmek gibi bir araştırma desenleri varsa, analiz süresi parametre tahmin gücüne de bağlı olarak on ile kırk dakika arasında değişebilmektedir. Programın bir diğer avantajı kullanıcı kılavuzlarının pek çok konuda bilgiler içermesidir. Latent GOLD programın en büyük sınırlılığı ise şu an için yalnızca Windows işletim sistemi ile çalışılabilir olmasıdır.

## KAYNAKÇA

- Bartolucci, F., & Pandolfi, S. (2016). LMest package, available at <https://cran.rproject.org/web/packages/LMest/index.html>
- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. New Jersey: John Wiley ve Sons.
- Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research*, 27(1), 131-157.
- Cosden, M., Larsen, J., Donahue, M. T., & Nylund-Gibson, K. (2015). Trauma symptoms for men and women in substance abuse treatment: A latent transition analysis. *Journal of Substance Abuse Treatment*, 50, 18-25.
- Gudicha, D., Schmittmann, V., & Vermunt, J. K. (2015). Power computation for likelihood ratio tests for the transition parameters in latent Markov models. *Structural Equation Modeling: A multidisciplinary Journal*, 23, 234-245. DOI: 10.1080/10705511.2015.1014040
- Guo, B., Aveyard, P., Fielding, A., & Sutton, S. (2009). Using latent class and latent transition analysis to examine the transtheoretical model staging algorithm and sequential stage transition in adolescent smoking. *Substance Use & Misuse*, 44(14), 2028-2042.
- La Flair, L. N., Reboussin, B. A., Storr, C. L., Letourneau, E., Green, K. M., Mojtabai, R., . . . Crum, R. M. (2013). Childhood abuse and neglect and transitions in stages of alcohol involvement among women: A latent transition analysis approach. *Drug and Alcohol Dependence*, 132: 491-498.
- Lanza, S. T., & Bray, B. C. (2010). Transitions in drug use among high-risk women: An application of latent class and latent transition analysis. *Advances and Applications in Statistical Sciences*, 3(2), 203-235.
- Lanza, S. T., & Collins, L. M. (2008). A new SAS procedure for latent transition analysis: Transition in dating and sexual behavior. *Developmental Psychology*, 44, 446-456.
- McCutcheon, A. L. (2002). Basic concepts and procedures in single-and multiple-group latent class analysis In J. A. Hagenaars, & A. L. McCutcheon (Ed.), *Applied latent class analysis* (pp. 56-89). New York: Cambridge University Press.
- Muthen, B., & Muthen, L. (2007). *Mplus user's guide fifth edition*. Los Angeles: Muthen ve Muthen.
- Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18, 450-469.
- Vermunt, J. K., Langeheine, R., & Bockenholt, U. (1999). Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, 24, 179-207.
- Vermunt, J. K., & Magidson, J. (2013). *LG-Syntax user's guide: Manual for Latent GOLD 5.0 Syntax Module*. Belmont, MA: Statistical Innovations Inc.
- Vermunt, J. K., & Magidson, J. (2016). *Technical guide for latent GOLD 5.1: Basic, advanced, and syntax*. Belmont, MA: Statistical Innovations Inc.

Vermunt, J. K., Tran, B., & Magidson, J. (2008). Latent class models in longitudinal research. In Menard, S. (Ed.), *Handbook of longitudinal research: Design, measurement, and analysis* (pp. 373-385). Burlington: Elsevier.

## EXTENDED ABSTRACT

### *Introduction*

Latent Markov models, based on their grounds of Latent Class Analyses, are used for longitudinal analysis of research data where both observed and latent data are categorical. Latent Markov models consist of two sub-parts, which are named as measurement model and structural model, and three parameters (i.e., latent status probabilities, item conditional probabilities, and latent transition probabilities) are predicted in these models. Latent status probabilities and item conditional probabilities in latent Markov models are directly associated with the parameters of Latent Class Analysis. The latent class probability parameter in Latent Class Analysis is usually named as latent status probability in Latent Markov models in order to define the dynamic structure of class. Furthermore, the conditional probabilities in Latent Markov models are interpreted as in latent class analysis. The Latent Markov model studies that have increased in recent years are conducted via number of commercial programs such as Mplus (L. Muthen ve Muthen, 1998-2007) and Latent GOLD (Vermunt ve Magidson, 2013) and via free software like PROC LTA (Lanza ve Collins, 2008) and LMest (Bartolucci and Pandolfi, 2016). The objective of this paper is to provide information about how Latent Markov Analysis is carried out and how its outcomes are interpreted by using Latent GOLD 5.1 program to researchers.

Latent GOLD is a commercial statistical package program developed by Vermunt and Magidson under Statistical Innovations Inc. Currently, the program is only available for Windows-equipped computers. With Basic, Advance and Syntax versions, the program offers a wide range of alternatives to predict the parameters of the models. Markov analysis, conducted with the syntax module in versions earlier than Latent Gold 5.0, can be now carried out by the user-friendly point-and-click interface versions, namely by Latent GOLD 5.0 and Latent GOLD 5.1 Advanced modules.

In Markov analyses, it is quite important to enter data correctly. In classical analyses, the entry of longitudinal data for each participant is done by raw by raw, but in latent Markov analysis; data must be entered column-by-column. At this point, it is absolutely necessary to enter the missing data which are likely to be found in the longitudinal studies. When the program reads the data set, it estimates the parameters of model assuming that the data has been entered for each observation and every time slice. For this reason, if the data entry is not done properly, the results cannot be interpreted or can be inaccurate (Vermunt ve Magidson, 2013). Latent GOLD program can read data sets with .sav, .dat, .csv, .txt, .rsp and .cho extensions.

While the model is being defined, one of the most crucial point that must be determined is the number of status. For each time occasion, status number can be examined separately, or the researchers may have a priori knowledge about the number of status. However, if the number of statuses is unknown, the different number of models can be tested, and then the best suited model can be selected. Also, case id and covariate variables are determined for presentation of the model. If the variable of time is considered as a covariate variable, the transition probabilities are always estimated separately for each slice (time-heterogeneous model); if not, the transition probabilities are calculated to be equal at all time occasions (time-homogeneous model).

During the model selection process, the most parsimonious model fit the data is sought in Latent Markov models. To find the best model, information criteria such as Pearson Chi-Square ( $\chi^2$ ), Likelihood Ratio Chi Square ( $L^2$ ) and Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) Information criteria can be used (McCutcheon, 2002). Model selection process using information criteria is aimed at choosing the best model among other models, and whichever information criterion is used, the lower value, which the criterion holds, points at the better model.



### **Results**

Time-dependent change is a crucial area of research for both educators and psychologists. When research question includes latent variables such as ability, knowledge level, clinical diagnosis, etc. and when the data set is categorical, the time-related change is examined by using Latent Markov Model. As compared to other programs that can only be run with syntax, Latent GOLD program with the point-click feature is more user-friendly for running Latent Markov analyses. However, researchers who want to test more complex models need to write syntax in this program. Furthermore, Latent GOLD is able to quickly analyze large dataset. Manuals of the program also provide users with useful information on a variety of topics. As for major limitation of Latent GOLD program, it can only work with Windows operating system.