

**UKUROVA NİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

Aslı BAHADIR

KARMA REGRESYON MODELLERİ

İSTATİSTİK ANABİLİM DALI

ADANA,2010

UKUROVA NİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KARMA REGRESYON MODELLERİ

ASLI BAHADIR
YÜKSEK LİSANS TEZİ
İSTATİSTİK

Bu tez 28/01/2010 Tarihinde Aşağıdaki Jüri Üyeleri Tarafından Oybirliği/Oyçokluğu İle Kabul Edilmiştir.

İmza	İmza	İmza
Prof. Dr. Hamza EROL	Prof. Dr. Selahattin KAÇIRANLAR	Yrd. Doç. Dr. Gülsen KIRAL
DANIŞMAN	ÜYE	ÜYE

Bu tez Enstitümüz İstatistik Anabilim Dalında hazırlanmıştır.

Kod No :

Prof. Dr. İlhami YEĞİNGİL
Enstitü Müdürü

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ÖZ
YÜKSEK LİSANS TEZİ

KARMA REGRESYON MODELLERİ

ASLI BAHADIR

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI

Danışman: Prof.Dr. Hamza EROL

Yıl: 2010 Sayfa:118

Jüri : Prof.Dr. Hamza EROL

Prof.Dr. Selahattin KAÇIRANLAR

Yrd.Doç.Dr. Gülsen KIRAL

Bu çalışmada, son yıllarda regresyon analizi ve karma dağılım yöntemlerinin avantajları birleştirilerek tanımlanmış olan, karma regresyon kavramına yer verilmiştir. Homojen yapılu kitleden alınan verilerle, genellikle yanlış tahminler elde edilmektedir. Bu koşullarda, gizli sınıf regresyon analizi olarak da bilinen karma regresyon modellerinin uygulamaları savunulmuştur. Karma regresyon modelleri ve analizi ile ilgili yapılan araştırmaların incelenmesinin ardından, bu yöntemin çalışmalardaki etki ve sınırlamaları tartışılmıştır. AMOS paket programı kullanılarak yapay bir veri kümesinde karma regresyon modeli, karma regresyon analizi, dağılımlar ve hipotez testi konularına değinilmiştir. Son olarak, AMOS'ın karma regresyon analizine katkısı tartışılmış, alternatif programlar ile elde edilen sonuçların karşılaştırmaları yapılmıştır.

Anahtar Kelimeler: Karma Dağılım, Lineer Karma Regresyon Modelleri, Gizli Sınıf Regresyon, Karma Regresyon Analizi, Bayes Tahmini.

ABSTRACT

MSc. THESIS

MIXTURE REGRESSION MODELS

ASLI BAHADIR

DEPARTMENT OF STATISTICS

INSTITUTE OF NATURAL AND APPLIED SCIENCES

UNIVERSITY OF CUKUROVA

Supervisor : Prof.Dr. Hamza EROL

Year: 2010, Pages:118

Jury: Prof.Dr. Hamza EROL

Prof.Dr. Selahattin KAÇIRANLAR

Assist.Prof.Dr. Gülsen KIRAL

In this thesis, mixture regression concept which is recently defined by combining with advantages of regression analysis and mixture distribution methods, has been included. With datas taken from homogeneous structured population, generally biased estimations are obtained. In these conditions, the applications of mixture regression models, also known as latent class regression analysis, have been advocated. After the examinations of performed illustrations associated with mixture regression models and analysis, the strengths and limitations of this method are discussed in the studies. Using AMOS software program, mixture regression model, analysis, distributions and hypothesis testing have been mentioned on an artificial dataset. Finally, AMOS have been discussed about addition on mixture regression analysis and obtained results compared with alternative programs.

Keywords: Mixture Distribution, Linear Mixture Regression Models, Latent Class Regression, Mixture Regression Analysis, Bayesian Estimation.

TEŞEKKÜR

Yaptığınız iş her ne olursa olsun; emek, sabır ve azim olmadan bir hedefe ulaşmanın hiç de kolay olmadığını gösteriyor bize hayat. Fakat, bu zorlu yolda beraber yürüdüklerinizdir, sizlere bu yolları daha emin yürüten. Çünkü, bilirsiniz ki onlar yalnız sizi siz olduğunuz için severler ve desteklerini hiç eksik etmezler.

Kendimi şanslı hissediyorum. Çünkü, yürüdüğüm bu yolda varlıklarıyla bana güç veren, çok değerli insanlar vardı hayatımda...

Tüm lisans ve yüksek lisans eğitimim boyunca, benden güven ve desteklerini eksik etmeyen değerli bölüm hocalarıma ve özellikle de tez çalışmam boyunca, benle değerli zamanını paylaşan ve destekleyen sayın danışmanım Prof. Dr. Hamza EROL 'a saygı, sevgi ve teşekkürlerimi sunarım.

Bir anne, baba olmanın verdiği sorumluluktan çok daha öte, beni bugünlere getiren sevgili aileme ve benimle aynı yolda yürüyen, hayatıma renk katan değerli dostlarıma destekleri için çok teşekkür ederim.

İÇİNDEKİLER

SAYFA

ÖZ.....	I
ABSTRACT.....	II
TEŞEKKÜR.....	III
İÇİNDEKİLER.....	IV
TABLolar DİZİNİ.....	VII
ŞEKİLLER DİZİNİ.....	IX
SİMGELER VE KISALTMALAR.....	X
1. GİRİŞ.....	1
2. ÖNCEKİ ÇALIŞMALAR.....	3
3. KARMA REGRESYON KAVRAMI VE İLGİLİ ÖRNEK ÇALIŞMALAR.....	4
3.1. Karma Dağılım Modelinin Tanımlanması.....	4
3.1.1. Bir Karma Dağılım Modeli Uygulaması.....	4
3.1.1.1. Bir Karma Dağılım Modelinin Analizi.....	6
3.2. Karma Dağılım Modelinden Karma Regresyon Modeline Geçilmesi.....	8
3.2.1. Karma Regresyon Modeli.....	8
3.2.1.1. Karma Regresyon Modelleri için Varsayımlar.....	8
3.2.1.2. Model Seçiminde Çoklu Kriterler.....	9
3.3. Karma Regresyon Kavramı Üzerine Bir Örnek.....	10
3.4. Karma Regresyon Modelinin Altyapısı.....	12
3.4.1. Karma Regresyon Model Tipleri.....	12
3.4.1.1. Standart Lineer Karma Regresyon Modeli.....	13
3.4.1.2. Genişletilmiş Karma Regresyon Modeli.....	14
3.4.1.3. Rastgele Etkili Karma Regresyon Modeli.....	15
3.4.1.4. Orta Değişkene Bağlı Karma Oranlı Karma Regresyon Modeli.....	15
3.5. Rastgele Etkili Karma Regresyon ile Eğri Kümeleme.....	16
3.5.1. Model Tabanlı Eğri Kümelemenin Oluşumu.....	18
3.5.1.1. Model Tabanlı Kümeleme.....	19
3.5.1.2. Model Tabanlı Eğri Kümeleme.....	19

3.5.2. Model Tabanlı Eğri Kümelemeden Standart Regresyon	
Karmasına Geçiş	20
3.5.2.1. Eğri Gruplarının Kümelenmesi Esnasında Rastgele Etkili	
Karma Regresyonun Önemi.....	20
3.6. Karma Regresyon İle İlgili Yapılan Çalışmaların İncelenmesi.....	21
3.6.1. Pedagojik Araştırmalarda Bir Karma Regresyon Analizi Uygulaması..	21
3.6.1.1. Ölçümler	22
3.6.1.2. Analiz.....	23
3.6.1.3. Sonuçlar.....	24
3.6.2. Yenidoğan Hastanesinde Kalış Süresine Dayalı Rastgele Etkili	
Sınırlı Karma Regresyon Uygulaması.....	27
3.6.2.1. Model Yapısı.....	28
3.6.2.2. Tahmin İçin EM Algoritması.....	28
3.6.2.3. Analiz.....	29
3.6.2.4. Sonuçlar	32
4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULAMASI VE BAYES	
TAHMİN YÖNTEMİ.....	33
4.1. AMOS Paket Programı.....	33
4.1.1. AMOS'un Kurulumu ve Arayüzü.....	33
4.2. Veri Hakkında.....	35
4.2.1. İlk Veri Grubu.....	35
4.2.2. İkinci Veri Grubu.....	37
4.3. Veri Kümesi İçindeki Grup Değişkeni	38
4.4. Karma Regresyon Analizi Öncesi Adımlar.....	39
4.4.1. Veri Dosyasını Belirleme.....	41
4.4.2. Model Belirleme.....	43
4.5. Bayes Tahmini.....	46
4.5.1. Önsel Seçim.....	48
4.5.2. Bayes Analizi.....	49
4.5.3. Yakınsama Kavramı ve Değerlendirilmesi.....	53
4.5.4. Kontrol Grafikleri.....	55

4.5.5. Güven Aralığı.....	60
4.5.5.1. Güven Düzeyinin Değiştirilmesi.....	61
4.6. Karma Regresyon Analizi.....	62
4.6.1. Dağılımların İncelenmesi.....	67
4.6.2. Grupların Karşılaştırılması.....	69
4.6.3. Farklı Durumların Sınıflandırılması	70
4.7. Parametre Tahminlerini Geliştirme.....	71
4.7.1. Hipotez Testi.....	75
4.8. Tartışma ve Karşılaştırmalar.....	75
5. SONUÇ VE ÖNERİLER.....	79
KAYNAKLAR.....	81
ÖZGEÇMİŞ.....	87
EKLER.....	88

TABLolar DİZİNİ**SAYFA**

Tablo 3.1. AMOS programı ile elde edilen parametre tahminleri.....	7
Tablo 3.2. Parametre tahminleri ve sınıf büyüklükleri	25
Tablo 3.3. Gizli sınıfların orta değişkenler ile ilişkisi	26
Tablo 3.4. Normal Karma Dağılım ile Parametre Tahminleri.....	30
Tablo 3.5. Rastgele etkili iki bileşenli karma regresyon modeli sonuçları.....	31
Tablo 4.1. SPSS’de Dozaj ve Performans 1 için veri girişi.....	36
Tablo 4.2. SPSS’de Dozaj ve Performans 2 için veri girişi	37
Tablo 4.3. Grup olarak adlandırılan değişkenin SPSS’de veri girişi.....	39
Tablo 4.4. AMOS’da 1.grubun belirlenmesi.....	40
Tablo 4.5. AMOS’da 2.grubun belirlenmesi.....	40
Tablo 4.6. AMOS’da gruplar için veri dosyası belirleme.....	41
Tablo 4.7. AMOS ‘da grup değişkenin seçilmesi	42
Tablo 4.8. AMOS’da 1. ve 2. gruplar için yapılan düzenlemelerin son hali ve gruplar için atama komutunun verilmesi.....	42
Tablo 4.9. AMOS’da, veri kümesi içindeki değişkenlerin gösterildiği kutucuk	44
Tablo 4.10. AMOS’un Nesne Özellikleri kutucuğu	45
Tablo 4.11. AMOS’da analiz özelliklerinin düzenlenebildiği kutucuk.....	50
Tablo 4.12. AMOS’da Bayesian SEM penceresi	51
Tablo 4.13. AMOS’da Bayesian SEM penceresinin son hali..	55
Tablo 4.14. AMOS’da Posterior penceresi.....	56
Tablo 4.15. Age-Vocabulary parametresinin seçildiği Bayes SEM penceresi.....	56
Tablo 4.16. AMOS’ın özet tablosunda görmek istenen seçeneklerin belirlenmesi....	61
Tablo 4.17. AMOS’da % 95 güven aralıklı özet tablosu	62
Tablo 4.18. AMOS’da analiz sonucu oluşan Bayesian SEM penceresi	64
Tablo 4.19. AMOS’da 2.grup için özet tablosu	65
Tablo 4.20. AMOS ‘da Ek Tahminler penceresi	66
Tablo 4.21. AMOS yardımı ile 1.grup için ortalama bir model tahmini hesaplama..	67
Tablo 4.22. AMOS’da herbir gözlem için grup üyelik olasılıkları.....	70
Tablo 4.23. AMOS’da regresyon ağırlıklarının b olarak belirlenmesi.....	72

Tablo 4.24. AMOS’da hata varyansının ψ olarak belirlenmesi.....	73
Tablo 4.25. AMOS’da parametre kısıtlandırımlarının ardından Bayesian SEM penceresi.....	74
Tablo 4.26. SPSS’de 1.grup için tanımlayıcı istatistik tablosu	76
Tablo 4.27. SPSS’de 1.grup için katsayılar tablosu	76
Tablo 4.28. Özet tablosundan 2.grup için parametre tahminlerinin çekilmesi.....	77
Tablo 4.29. SPSS’de 2.grup için tanımlayıcı istatistik tablosu.....	77
Tablo 4.30. SPSS’de 2.grup için katsayılar tablosu.....	78
Tablo 4.31. SPSS’de veri kümesi için tanımlayıcı istatistik tablosu.....	78

ŞEKİLLER DİZİNİ

SAYFA

Şekil 3.1. Anderson verisine ait SPSS programı ile elde edilmiş saçılım grafiği.....	5
Şekil 3.2. Serbest bırakılan bit sayı ile bağlantılı hasta bitki sayısının saçılım grafiği.....	11
Şekil 3.3. Serbest bırakılan bit sayısına karşın her iki grubun sonsal üye olasılıkları.....	11
Şekil 3.4. Kuzey Atlantik üzerinden geçen kasırga yörüngeleri.....	17
Şekil 4.1. AMOS paket programı içindeki sekmeler	34
Şekil 4.2. AMOS arayüzünden genel bir görüntü	34
Şekil 4.3. SPSS paket programı ile oluşturulmuş Dozaj ve Performans 1 verisi için saçılım grafiği.....	36
Şekil 4.4. SPSS paket programı ile oluşturulmuş Dozaj ve Performans 2 için saçılım grafiği.....	38
Şekil 4.5. AMOS’da oluşturulan yol diyagramı üzerinde düzenlemeler yapılmasını sağlayan sekmeler.....	44
Şekil 4.6. AMOS’da oluşturulan yol diyagramının son hali.....	45
Şekil 4.7. AMOS’da frekans poligonu	56
Şekil 4.8. AMOS’da poligon üzerinde iki tahminin eş zamanlı göstergesi.....	58
Şekil 4.9. AMOS’da çizgi grafiği	59
Şekil 4.10. AMOS’da otokorelasyon grafiği.....	60
Şekil 4.11. AMOS’da Group number 1 için sonsal dağılım	68
Şekil 4.12. AMOS’da Group number 2 için sonsal dağılım	69
Şekil 4.13. AMOS’da hipotez testi öncesi yol diyagramı üzerinde düzenlemeler....	72
Şekil 4.14. Parametre kısıtlamalarının ardından yol diyagramı.....	73

SİMGELER VE KISALTMALAR

- AIC : Akaike Bilgi Kriteri (Akaike Information Criteria)
- AMOS : Kritik Yapıların Analizi (Analysis of Moment Structure)
- ANCOVA : Kovaryans Analizi (Analysis of Covariance)
- ARS-Math : Matematiksel Başarı Üzerine Değerlendirme Çalışmaları (Rating Studies on Math Achievement)
- BLUP : En İyi Lineer Yansız Tahmin Edici (Best Linear Unbiased Predictor)
- ECLS : Okul Öncesi Dönem Çalışması (Early Childhood Longitudinal Study)
- EM : Beklenti Maksimizasyonu (Expectation Maximization)
- EMMIX : EM Tabanlı Karma Programlama (EM Based Mixture Programme)
- ETCs : Ekstra Tropik Kasırgaların Sayısı (Number of Extra Tropical Cyclones)
- GLLM : Genelleştirilmiş Lineer Karma Model (Generalised Linear Mixed Model)
- GLLAMM: Genelleştirilmiş Lineer Gizli ve Karma Model (Generalized Linear Latent and Mixed Models)
- HME : Uzman Hiyerarşik Karma (Hierarchical Mixtures of Experts)
- LOS : Yenidoğan Hastanesinde Kalış Süresi (Length of Stay)
- MAP Estimation: Sonsal Bir Maksimum Tahmin (Maximum a Posterior Estimation)
- MCMC : Markov Zinciri Monte Carlo (Markov Chain Monte Carlo)
- PSU : Başlıca Örneklem Birimi (Prime Sampling Unit)
- REML Estimators: Rezidü Maksimum Likelihood Tahmini edicileri (Residual Maximum-Likelihood Estimators)
- SEM : Yapısal Eşitlik Modelleme (Structural Equation Modelling)
- SPSS : Sosyal Bilimler için İstatistik Paket Programı (Statistical Package for Social Sciences)
- SRS : Sosyal Değerlendirme Çalışmaları (Social Rating Studies)
- S.E : Standart Hata (Standard Error)
- S.D : Standart Sapma (Standard Deviation)
- C.S : Yakınsama İstatistiği (Convergence Statistic)

1.GİRİŞ

Regresyon analizi, istatistikte özellikle homojen yapıdaki kitleden alınan veriyi modellemek ve analiz etmek amacıyla kullanılan en temel yöntemdir. Karma dağılım modelleri ise özellikle heterojen yapıdaki kitleden alınmış veriyi modellemek ve analiz etmek amacıyla son yıllarda tercih edilen ve yaygın olarak birçok alanda başvurulan bir yöntemdir. Hem regresyon modeli, hem de karma dağılım modeli uzun yıllar önce ortaya atılan iki farklı kavramdır. Son yıllarda ise, bu iki yöntemin avantajları birleştirilerek ortaya lineer regresyon modellerinin karma modeli ya da kısaca karma regresyon modeli olarak adlandırılan yeni bir model tanımlanmıştır.

Klasik regresyon analizi, özellikle pedagojik araştırmalarda tercih edilen bir yöntemdir. Bu gibi analizlerde, örneklemdeki regresyon parametre tahminlerinin genel bir gurubunu ele alıp inceleyerek, dolaylı yoldan kitle özellikleri ile ilgili varsayımlarda bulunacak hale gelinebilir. Fakat bazı durumlarda, dolaylı yolla belirtilen bu tahminler çokta gerçekçi değildir. Örneğin, örneklem ileri seviyede matematik becerilerine sahip olan kişiler ya da düşük seviyede matematik becerilerine sahip olan kişiler şeklinde çeşitli alt kitlelerden meydana gelebilir. Böyle durumlarda, farklı alt kitlelere karşın parametre tahminleri aynıymış gibi gösterilmeye çalışıldığında, klasik regresyon modellerinde yanlış tahminler söz konusu olacaktır (Dink, 2006).

Sonuçta görülebilmektedir ki, homojen gruplarla yapılan temel varsayımlar genelde gerçekçi olmamaktadır. Bu koşullarda, gizli sınıf regresyon analizi olarak da bilinen, karma regresyon modellerinin uygulamaları savunulmaktadır.

Karma regresyonun kullanıldığı bazı akademik alanlar ve bu alanların tercih ettiği yöntemler: Ekonomi: Sistem veya regresyon değiştirme (switching regimes or regressions); Sosyal Bilimler: Gizli sınıf regresyon (latent class regression); Yapay Zeka ile Öğrenme (Machine Learning): Uzman hiyerarşik karma (HME); Sağlık Bilimleri: Rastgele etkili sınırlı karma regresyon, şeklinde sıralanabilir (Young, 2007).

Karma regresyon analizindeki amaç, veride gözlenemeyen bir kitle heterojenliğinin söz konusu olduğu durumda, örnekleme gruplar halinde bölme ve her bir grup için regresyon doğrusu oluşturmaktır.

Uygulamalar kısmında da görüleceği üzere; karma regresyon analizi, klasik regresyon analizine göre çok daha esnektir. Öyle ki, her bir gizli sınıf içerisinde regresyon parametre tahminleri değişkenlik gösterebilir ve bu gizli sınıflar içerisinde açıklanabilir.

Bu çalışmada, daha önce gerçekleştirilmiş olan araştırmalarda karma regresyon modellerinin oluşturduğu etki ve sınırlandırmalara değinilmiş, uygulamalar bölümünde ise yapay bir veri kümesi, AMOS paket programı kullanılarak karma regresyon analizi ile incelenmiştir.

Sonuçta, homojen yapılı bir veri kümesinde yapılan temel varsayımların genelde gerçekçi olmaması karma dağılım kavramına, veride gözlenemeyen kitle heterojenliği ise karma regresyon kavramının incelenmesi ve bu kavramın etkili bir şekilde kullanılabileceği paket programı arayışına yönlendirmiştir.

2.ÖNCEKİ ÇALIŞMALAR

Regresyon tabanlı karmalar üzerine daha önce yapılmış çalışmalarda, değişik yöntemlerin kullanılmasıyla iki bileşenli ve bileşenleri normal dağılımlı olan karma dağılım modelindeki parametreleri tahmin etmek amacıyla likelihood fonksiyonun tanımlanması yapılmıştır (Quandt, 1972; Quandt ve Ramsey, 1978; Hosmer, 1974). Hosmer (1974) tarafından gerçekleştirilen çalışma, Dempster ve ark. (1977) tarafından geliştirilen beklenti maksimizasyonu (EM) algoritmasının özel halidir. Ayrıca Spath (1979) grup ya da kümeye göre lineer regresyon (clusterwise linear regression) olarak adlandırılan ve olasılığa dayalı olmayan bir algoritma geliştirmiştir. Bu algoritma Hartigan ve Wong (1978) tarafından geliştirilen K-ortalamlar algoritmasına benzer. K-ortalamlar algoritması, heterojen yapıdaki veri üzerinde K farklı regresyon denkleminin çözümüne yöneliktir.

Daha sonra, Späth (1979) tarafından oluşturulan algoritma, EM algoritmasına dayalı olarak genel K-grup ya da küme durumu olarak geliştirildi (DeSarbo ve Cron 1988; Jones ve McLachlan 1992; Gaffney ve Smyth 1999). Bu algoritma daha sonra değişik durumlara ve parametrik olmayan çekirdek regresyon modellerine (non-parametric kernel regression models) genişletilmiştir. Genel rastgele etkiler içeren genelleştirilmiş lineer modellerin karması Lenk ve DeSarbo (2000), tarafından Bayesci içerikle oluşturuldu. Bu çalışmada Markov Zinciri Monte Carlo (MCMC) teknikleri kullanılarak rastgele etkiler ile genelleştirilmiş lineer modellerin karması için istatistiksel olarak Bayesci sonuç çıkarma konusu ele alınmıştır.

Jordan ve Jacobs tarafından ilk 1994 yılında başlayan ve yakın zamana kadar geliştirilen ilgili model ailesi, Uzman hiyerarşik karma (HME) modellerdir. Bu yöntem regresyonların karmalarındaki bazı özellikleri paylaşmaktadır. Bununla birlikte uzman hiyerarşik karmalar, genelleştirilmiş lineer modellerin karmasını bulunduran ağ kapıları içerir. Uzman hiyerarşik karma, matematiksel olarak regresyon karma modeline benzer. Tek fark, uzman hiyerarşik karmada denetimli bilgi kullanılmaktadır.

3. KARMA REGRESYON KAVRAMI VE İLGİLİ ÖRNEK ÇALIŞMALAR

3.1. Karma Dağılım Modelinin Tanımlanması

Karma regresyon kavramının temelini teşkil eden karma dağılım modelinin tanımlamasını bir örnek içinde açıklamaya çalışalım. Klasik bir regresyon modelinde, bütün gözlemlerin ortak kitle parametreleriyle tek bir kitleden alındığı varsayılmaktadır. Buna rağmen bir örneklem, gözlemlerin çeşitli gruplarından oluştuğunda (erkekler ve kadınlar gibi), regresyon analizi incelemeye alınabilir. Eğer bir bağımlı değişken üzerine bağımsız değişkenlerin etkileri farklıysa, bu ya grupların sabit terimlerinde ya da eğimlerinde farklılık olarak gözlemlenebilecektir. Böyle durumlarda bu gruplar farklı kitleler şeklinde dikkate alınabilir (erkek kitlesi ya da kadın kitlesi gibi). Dolayısıyla bu tip kitleler heterojen olarak dikkate alınıp, alt kitlelerinin de uygun bir şekilde özelliklerini yansıtabilmesi için farklı kitle parametrelerini gerekli kıldığı belirtilebilir. Kitle heterojenliğinin kaynağı, cinsiyet gibi gözlenen grup üyeliklerine dayalı olduğu için, çoklu grupları dikkate alarak regresyon modellerinin kullanımıyla veri analiz edilebilir. Yöntem literatüründe, baştan bu alt kitleler “grup” olarak adlandırılabilir (örn., Lubke & Muthen, 2005; Muthen, 2001). Bu cümlelerden şu sonuca varabiliyoruz ki, karma dağılım modelinin temelinde heterojenlik kavramı yer almaktadır.

Dolayısıyla, tüm kitle için düşünüldüğünde uygun olmayan bir modelin, kitle alt gruplara bölündüğünde her bir alt grup için doğruluğundan bahsedebiliyorsak, bu durumda karma modelleme uygun olacaktır (Arbuckle, 2007).

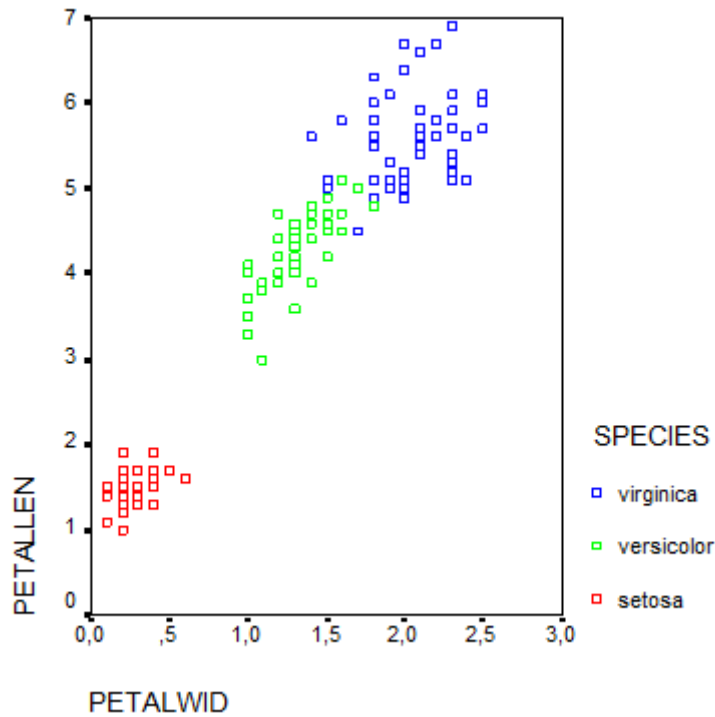
Karma modelleme, Arminger, Stein ve Wittenberg (1999), Hoshino (2001), Lee (2007), Loken (2004), Vermunt ve Magidson (2005), Zhu ve Lee (2001) tarafından yapısal eşitlik modelleme bağlamında tartışılmıştır.

3.1.1. Bir Karma Dağılım Modeli Uygulaması

AMOS paket programına kadar karma modelleme, önceden sınıflandırılmış olan durumlardan faydalanarak diğer durumları sınıflandırma şeklinde

uygulanmaktaydı. Oysa AMOS ile durumların gruplara (sınıflara) atanması çok daha pratik hale getirilmiştir. Uygulamalar kısmında AMOS paket programına ve bu program yardımı ile amaçlarımız doğrultusunda veri analizine, detaylı bir şekilde değinilecektir.

Bu anlatılanlara ışık tutması amacıyla, Anderson (1935) tarafından toparlanmış ve Fisher (1936) tarafından diskriminat analizini göstermek amaçlı kullanılmış olan veri kümesi ile çalışılmıştır. Bu orijinal verinin tamamı, Ek-1’ de paylaşılmıştır. Veri kümesi, incelenen 150 farklı bitkinin çiçekleri üzerinde yapılan dört ölçümü kapsamaktadır. İlk 50 çiçek, süsen çiçeğinin setosa türünden. Diğer 50 çiçek, süsen çiçeğinin versicolor türünden ve son 50 ise, virginica türündendir.



Şekil 3.1. Anderson verisine ait SPSS programı ile elde edilmiş saçılım grafiği

İki sayısal ölçüm olan Petal Length ve Petal Width ‘in (taç yaprağı uzunluğu ve genişliği) saçılım grafiği, bu iki ölçümün tek başına, türlere dayalı olarak çiçeklerin sınıflandırılmasında faydalı olacağını belirtmektedir. Dolayısıyla, taç yaprağın uzunluğu ve genişliği için doygun bir model kullanılabilir. Karma modellemeyi

uygularken, doymun modeller için sınırlama yoktur. Yani bir faktör analizi modeli, bir regresyon modeli veya herhangi bir model ile kullanılabilir.

Bütün setosa çiçekleri saçılım grafiğinin sol alt köşesinde toplanmıştır. Böylece, AMOS programı için Petal Length ve Petal Width'i kullanarak setosa çiçeklerini diğerlerinden ayırt etmek kolay olacaktır. Diğer taraftan, versicolor ve virginica çiçekleri grafikte üst üste gelmektedir. Bundan dolayı, saf versicolor veya virginica çiçeğinin Petal Length ve Petal Width'e dayanarak anlatılması bazen güç olabilir.

Bu örnekte, her çiçeğin türünün belirtildiği orijinal veri kümesi kullanılmayacaktır. Yerine, örnekte sadece birkaç çiçek için türlerin belirtildiği bir veri kümesi kullanılacaktır. Buradaki amaç, karma dağılım modelinde sadece belirli sınıflardan yola çıkarak AMOS 'un bu analizi gerçekleştirebileceğini ve genel anlamda karşımıza neler çıkabileceğini gösterebilmektir. Bu paket programının hiçbir sınıfın gözlenemediği durumlarda neler yapabileceği ise, dördüncü bölümde paylaşılmıştır.

Veri kümesi incelendiğinde, 10 tane setosa çiçeğinin, 10 tane versicolor ve 10 tane de virginica çiçeğinin tür bilgileri olarak ulaşılabilir olduğunu varsayalım. Bu durumda geri kalan 120 çiçeğin tür bilgileri bilinmeyecektir. AMOS veriyi analiz ederken, geri kalan çiçeklerin sınıflandırmasına yardımcı olmak amacıyla, her tür deki çiçek için 10 'ar örnek mevcut olacaktır.

3.1.1.1. Bir Karma Dağılım Modelinin Analizi

Karma dağılım modelinin analizini örnekten yola çıkarak açıklamak istersek, olağan bir üçlü-grup analizini düzenlemeye oldukça benzemektedir. Tek fark, burada önceden sınıflandırılmış olan durumlardan faydalanarak diğer durumların sınıflandırılmaya çalışılıyor olmasıdır. Bu fark ise AMOS paket programında "Assign cases to groups" onay kutusunun işaretlenmesi ile kolaylıkla gerçekleştirilebilir. Bu onay işareti, veri kümesi durumun hangi gruba ait olduğunu belirlemediğinde ya da eksik belirlediğinde, bir durumun bir gruba atanması gerektiğini anlatmaktadır.

3.KARMA REGRESYON KAVRAMI VE İLGİLİ ÖRNEK ÇALIŞMALAR

Aslı BAHADIR

Karma modellemede sadece Bayes tahmini kullanılabilir. Olağan bir üçlü-grup regresyon analizinde sahip olunabilecek parametre tahminleri, Tablo 3.1’ de görülebilmektedir.

Tablo 3.1. AMOS programı ile elde edilen parametre tahminleri

PossiblySetosa	PossiblyVersicolor	PossiblyVirginica										
	Mean	S.E.	S.D.	C.S.	Median	50% Lower bound	50% Upper bound	Skewness	Kurtosis	Min	Max	Name
Means												
PetalLength	1,462	0,000	0,026	1,000	1,462	1,444	1,479	-0,051	0,175	1,333	1,564	
PetalWidth	0,246	0,000	0,016	1,000	0,246	0,235	0,257	0,026	0,090	0,186	0,314	
Covariances												
PetalWidth<->PetalLength	0,007	0,000	0,003	1,000	0,007	0,005	0,009	0,534	0,864	-0,008	0,026	
Variances												
PetalWidth	0,013	0,000	0,003	1,000	0,012	0,011	0,014	0,918	1,693	0,006	0,032	
PetalLength	0,034	0,000	0,008	1,000	0,033	0,029	0,038	0,907	1,456	0,016	0,081	
PossiblySetosa Proportion												
Proportion	0,334	0,000	0,037	1,000	0,333	0,308	0,358	0,130	0,016	0,192	0,508	

Bu tabloda bölüm bölüm her bir grup için sonuçlar bulunmaktadır. AMOS programında bir gruptan diğerine, tablonun üst kısmında yer alan sekmeler tıklanarak geçilebilir. Bu örnekte, model parametreleri sadece ortalama, varyans ve kovaryansdan oluşmaktadır. Oldukça komplike modellerde, ayrıca regresyon ağırlıklarının ve sabitlerin olması beklenir.

Bir karma model analizinde, ayrıca kitle içerisinde her bir grubun hangi oranda tahmin edildiği gösterilebilir. Yukarıdaki tabloda, kitle içerisindeki setosa çiçeklerinin 0.334 oranında tahmin edildiği görülmektedir.

3.2. Karma Dağılım Modelinden Karma Regresyon Modeline Geçilmesi

Burada karma regresyon modelleri, sınırlı karma modellerinin genel bir çerçevesinde bir bölüm olarak açıklanmıştır (Lubke ve Muthén, 2005; Muthen, 2001; Muthen ve Muthen, 2000; Nagin ve Tremblay, 2001; Vermunt ve Magidson, 2002).

3.2.1. Karma Regresyon Modeli

Bu çalışmada gözlenemeyen kitle heterojenliğinin söz konusu olduğu durumlara özen gösterilmiştir. Diğer bir deyişle, kitle içindeki farklı grupların gizliliğine dikkat çekilmektedir (McCutcheon, 1987; Waller ve Meehl, 1998). Örneğin, öğrenciler sosyo-duygusal gelişimleri ile ilgili farklılık gösterebilir. İleri matematik becerilerine sahip çocuklar ya da düşük matematik becerilerine sahip çocuklar gibi, iki farklı türden oluşabilecek bir niteliğin mensubu da olabilirler (Ding,2006). Dolayısıyla, her iki tür için, çocukların sosyo-duygusal gelişimleri ile matematikteki başarıları arasındaki ilişkiyi tanımlayan, ayrı iki regresyon doğrusundan bahsedilebilir. Ayrıca karma regresyon modeli, klasik regresyon model ile klasik gizli sınıf modelinin bir kombinasyonu olarak gösterilebilir (Lazarsfeld ve Henry, 1968; McCutcheon, 1987).

Karma regresyon modelinin kullanımı, bağımsız değişkenlerin bir kümesiyle bağımlı değişken arasındaki ilişkinin belirlenmesi için bir araç vazifesi görmektedir. Karma regresyon analizi, klasik regresyon analizine göre çok daha esnektir. Öyle ki, her bir gizli sınıf içerisinde regresyon parametre tahminleri değişkenlik gösterebilir ve bu gizli sınıflar içerisinde açıklanabilir.

3.2.1.1. Karma Regresyon Modelleri için Varsayımlar

Gizli sınıf regresyon analizi olarak da bilinen karma regresyon modelleri (Anderson, 2004; Bouwmeester, Sijtsma ve Vermunt, 2004; Vermunt ve Magidson, 2005) bağımlı değişken ve bağımsız değişkenlerin bir kümesi arasındaki ilişki ile beraberinde veriye en iyi şekilde uyum sağlayan gizli sınıfların sayılarını tanımlamak

ve bilinen bir gizli sınıf için potansiyel açıklayıcı değişkenleri test etmek amacıyla kullanılmaktadır.

Klasik regresyon analizinde, örneklemdaki regresyon fonksiyonunun basit bir, çok değişkenli normal dağılımdan ileri geldiği varsayılır. Lineer karma regresyon modellerinde ise bundan farklı olarak, her biri bir gizli sınıfa karşılık farklı çok değişkenli normal dağılımların bir karmasının modellenmesiyle, heterojen regresyon fonksiyonları hesaba katılır. Aynı gizli sınıf içerisindeki her bir gözlem, aynı regresyon fonksiyonunu paylaşmaktadır.

Böylece karma regresyon analizinin, gözlenemeyen alt kitleler karşısında parametre farklılıklarını hesaba katarak basit kitle varsayımlarını zayıflattığı söylenebilir. Bu durum, farklı regresyon fonksiyonları etrafındaki gözlemlerin değişkenliğini gerektiren gizli sınıfların kullanılmasıyla tanımlanmıştır. Örneğin, faktörler üzerine yapılan bir çalışma, öğrencinin matematiksel başarısına etki edebilir. Yani bir araştırmacı, öğrencinin kişisel becerilerini, motivasyonunu, öğrenme metodlarını ve sınıfın büyüklüğü gibi bağımsız değişkenleri öğrencinin başarısını etkileyen faktörler olarak dahil edebilir (Ding, 2006).

Karma regresyon analizinde başlangıç noktası, öncelikle veriye en iyi uyumu sağlayan gizli sınıfların sayısını belirlemek olacaktır. Daha sonra her bir gizli sınıf içerisinde bağımlı değişkenler üzerine bağımsız değişkenlerin etkileri gözden geçirilebilir. Belirli bir gizli sınıf için sadece kişisel beceri matematiksel başarı için yeterli iken, ikinci bir sınıf için öğrenme metodu ile sınıf büyüklüğü matematiksel başarı için etkili olabilir. Yani, regresyon modelleriyle gizli sınıfların birlikte kullanımı oldukça esnektir.

3.2.1.2. Model Seçiminde Çoklu Kriterler

Lineer karma regresyon modelleri, sonlu karma modellerin (Muthen, 2001), bir parçası olduğundan, çoklu kriterler regresyon analizi için gizli sınıfların sayısını belirlemede kullanılabilir. Çünkü farklı göstergeler, uygun bir modelin farklı yönleri ile ilgili bilgi edinilmesini sağlayacaktır. Farklı modeller arasında yapılan bu karşılaştırmalar, veri için en uygun olanının belirlenmesini sağlayacaktır. Örneğin,

likelihood oran testi (Lo, Mendell ve Rubin, 2001), farklı sayıda gizli sınıflara sahip olan karma regresyon modellerini karşılaştırmak için kullanılabilir. Anlamlı bir ki-kare değeri ($p < 0,05$), belirtilen modelin olası en küçük sınıflı bir model olduğunu gösterir. Ayrıca, final modelin seçimi Akaike (Akaike, 1973) veya Bayes bilgi kriteri (Schwarz, 1978) gibi kriterlere bağlı olabilir. Eğer, gözlemlenen kriter değerleri düşükse, bu artık uygun bir düzenlemenin göstergesidir. Bir diğer gösterge de, Entropy (Ramaswamy, DeSarbo, Reibstein ve Robinson, 1993) dir. Burada gözlemlerin model-tabanlı olasılıkları dikkate alınarak, sınıflar içerisine yerleştirilmesinde sınıflandırmanın doğruluğu belirlenmektedir. Bu olasılıklar, 0.00 ile 1.00 değerleri arasında gezinir. Yüksek değerler, en iyi sınıflandırmanın işaretidir.

Sonuçta, kitle heterojenliğini doğru olarak açıklayabilen gizli sınıfların sayısını tespit etmek zordur.

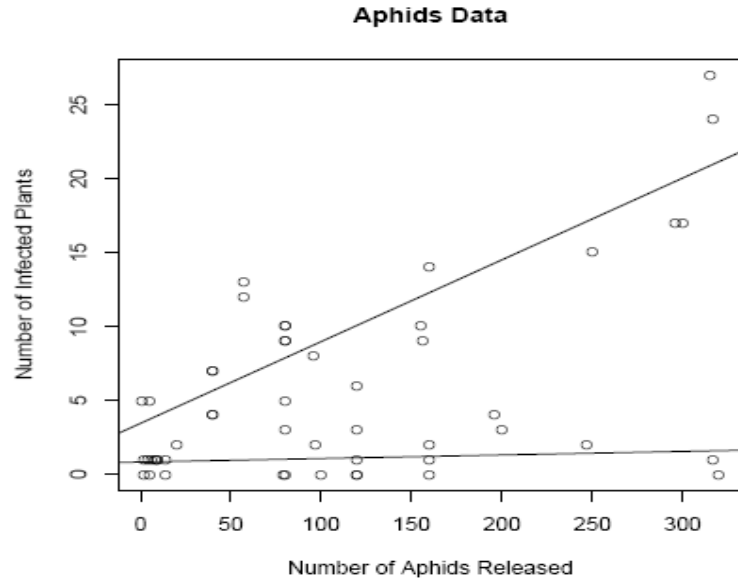
3.3. Karma Regresyon Kavramı Üzerine Bir Örnek

Şu ana kadar karma regresyon ile ilgili anlatılanlara bir örnek teşkil etmesi amacıyla, yaprak biti veri kümesini inceleyelim:

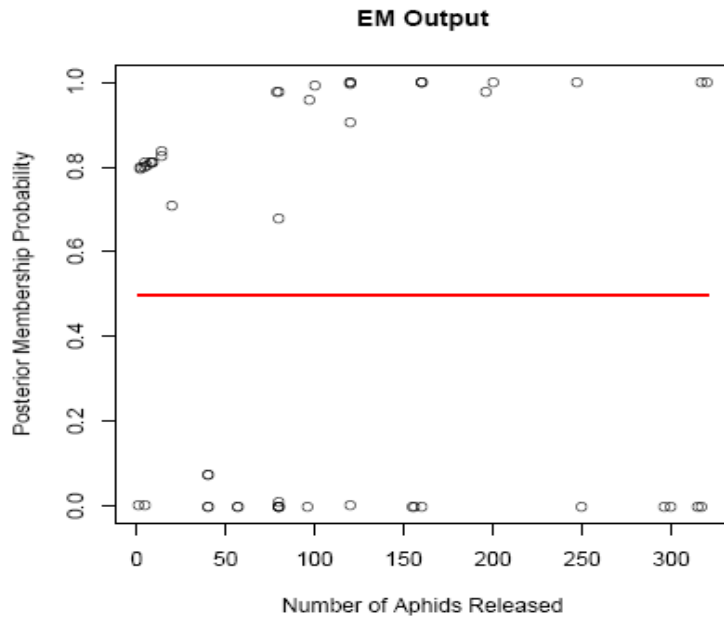
Bu veri kümesini ilk olarak Turner (2000), iki bileşenli karma regresyon ile analiz etmiştir. Deney, 81 küçük tütün bitkisi üzerine farklı zamanlarda bir miktar ham şeftali yaprak bitinin, serbest bırakılmasıyla gerçekleşmiştir. Bitkilerin her serbest bırakılışının ardından, hastalıklı bitkilerin sayısı kaydedilmiştir. Turner (2000), bazı bit gruplarının “ilk evrelerini” pas geçebileceklerini, bununda düşük seviyede virüs aktarımının bir işareti olduğunu ileri sürmüştür.

Bitki üzerine bırakılan bit sayısındaki değişime karşın hasta bitkilerdeki değişimin ilişkili oranlarının nasıl olduğunu görebilmek için kovaryanta (orta değişkene) bağlı karma oranları analize dahil etmek mümkündür.

Şekil 3.2. ve 3.3. ‘de sırası ile serbest bırakılan bit sayı ile bağlantılı hasta bitkilerin sayısının saçılım grafiği ve bit sayısına bağlı olarak her iki grubun sonsal üye olasılıklarının grafik çıktıları görülebilmektedir.



Şekil 3.2. Serbest bırakılan bit sayısı ile bağlantılı hasta bitkilerin sayısının saçılım grafiği (Young, 2007)



Şekil 3.3. Serbest bırakılan bit sayısına karşın her iki grubun sonsal üye olasılıkları (Young, 2007)

3.4. Karma Regresyon Modelinin Altyapısı

- $x_1, \dots, x_n$ orta değişkenlerinin vektörüne, doğrusal bir şekilde bağlı olan $y_1, \dots, y_n$ gözlemlenmiş değerler dikkate alınır.
- Karma Regresyon, bilinmeyen oranlarla önsel bilinen sınıfların bir sayısında ileri gelerek, (x_i, y_i) modeli için bir yöntem sağlar.
- Karma regresyonun model yapısı, şu şekilde de düşünebilir:

Faktör etkilerinin farklı regresyonlarının belirtildiği ANCOVA'da gözlemlenemeyen durumlar söz konusu iken kullanılır. Diğer bir deyişle, bir ya da daha fazla faktörün bağımlı değişken üzerindeki etkisinin incelendiği bir araştırmada, bu faktörlerin dışında bağımlı değişken üzerinde etkisi olabilecek bir ya da daha fazla değişkenin (kovaryantın) istatistiksel olarak kontrol edilemediği (gözlemlenemediği) durumlarda ANCOVA analizi yerine, karma regresyondan faydalanılabilmektedir (Young, 2007).

3.4.1. Karma Regresyon Model Tipleri

Bağımlı değişkenin sürekli olması ve lineer olma sınırlandırmaları altında, çeşitli tiplerde karma regresyon modelleri mevcuttur:

1. Standart Lineer Karma Regresyon Modeli
2. Genişletilmiş Karma Regresyon Modeli
3. Rastgele Etkili Karma Regresyon Modeli

Diğer taraftan, eğer bağımlı değişken iki değerli (dichotomous) veya sınıflamalı (nominal), iki sonuçlu (binary) veya çok terimli (multinomial) ise, lojistik karma regresyon analizi formüle edilip gösterilebilir (örn.: Orta Değişkene Bağlı Karma Oranlı Karma Regresyon Modeli).

Bu çalışmada, bağımlı değişkenin sürekli ve lineer olma sınırlamaları altındaki karma regresyon modellerine odaklanılacak ve standart lineer bir karma regresyon modeli için, dördüncü bölümde örnek bir uygulama gerçekleştirilecektir.

3.4.1.1. Standart Lineer Karma Regresyon Modeli

Lineer karma regresyon modelleri, veride kitle heterojenliğinin gözlenemediği durumlarda kullanılır (Ding 2006). Bu durumda regresyon karma modelleri, gizli ya da saklı sınıf regresyon modelleri olarak adlandırılır (Anderson 2004, Bouwmeester ve ark. 2004, Vermunt ve Magidson 2002, 2005). Gizli ya da saklı sınıf regresyon analizinde; bağımlı değişken ve gizli sınıflar içeren bağımsız değişkenlerin bir kümesi arasındaki en iyi ilişkinin modeli belirlenir. Geleneksel regresyon modelinde regresyon fonksiyonu oluşturulurken verinin çok değişkenli normal dağılıma sahip olduğu varsayımı yapılır. Lineer karma regresyon modelinde ise regresyon fonksiyonu oluşturulurken verinin herbiri bir gizli sınıfa karşılık gelen, farklı çok değişkenli normal dağılımların bir karmasına sahip olduğu varsayımı yapılır.

Standart lineer karma regresyon modelleri ile ilgili bu genel bilgilerin paylaşımının ardından model oluşumunda temel teşkil eden, karma lineer regresyonun heterojen regresyon fonksiyonu aşağıdaki eşitlikte gösterilmektedir:

$$f(y_i \setminus x_i) = \sum_{j=1}^k \lambda_j \phi(y_i \setminus x_i^T \beta_j, \sigma_j^2) \quad (3.1)$$

Bu fonksiyonda;

- $\lambda_1, \dots, \lambda_k$ karma oranlardır. $\forall_j \lambda_j \in (0,1)$ standart normal dağılımlı ve

$$\sum_{j=1}^k \lambda_j = 1 \text{ dir.}$$

- $\beta_1, \dots, \beta_k$ ve $\sigma_1^2, \dots, \sigma_k^2$ sırasıyla regresyon katsayılarının ve varyansın göstergeleridir.

- $\phi(\cdot | \mu, \tau^2)$, μ ortalama τ^2 varyansı ile normal olasılığı ifade etmektedir.

Bu bilgilere dayanarak genel bir lineer karma regresyon modeli aşağıdaki gibi formüleleştirilebilir:

$$y_{i(c)} = \beta_{0(c)} + \beta_{1(c)}x_1 + \beta_{2(c)}x_2 + \cdots \cdots + \beta_{k(c)}x_k + \varepsilon_{i(c)} \quad (3.2)$$

Bu eşitlik, altsimge c ($c = 1, 2, \dots, C$) haricinde klasik regresyon modelinin görünümüne sahiptir. Eşitlikte yer alan altsimge c , farklı gizlenmiş sınıflar etrafında parametrelerin değişkenlik gösterebileceğine işaret eder. Diğer bir ifadeyle, her bir gizli sınıf c içerisindeki gözlemler aynı parametre tahminlerine sahipken; farklı sınıflar söz konusu olduğunda, farklı parametre tahminlerine sahip olacaklardır. Eşitlikten şunu anlıyoruz ki, bağımlı değişken açıklayıcı değişkenlerin bir fonksiyonu olarak tahmin edilebilir. C -kategorili gizli sınıf değişkeni c , aynı regresyon katsayılarına sahip homojen bir altkitlenin her bir kategorideki açıklamasını içerir.

3.4.1.2. Genişletilmiş Karma Regresyon Modeli

Verinin $\varepsilon_i \sim \lambda N(0, \sigma^2) + (1 - \lambda) \text{Exp}(\alpha)$ iken, aşağıdaki şekilde modellenebileceği varsayılır:

$$y_i = x_i^T \beta + \varepsilon_i \quad (3.3)$$

Bu eşitlikte:

λ : karma oran,

$\sigma^2 > 0$: normal bileşenler için varyans,

$\alpha > 0$: üstel bileşen için bir kriterdir.

Hata teriminin üstel bileşenden ileri gelmesi sebebiyle, pozitif olasılık sözkonusu olacağından, böylesi durumda $y_i - x_i^T \beta$ rezidülerinin pozitif olması gerekliliği bizi kısıtlayacaktır.

3.4.1.3. Rastgele Etkili Karma Regresyon Modeli

- i-inci gözlemin aslında bir diziyi $((x_{i,1}, y_{i,1}), \dots, (x_{i,n_i}, y_{i,n_i}))$ belirttiği yerde $((x_1, y_1), \dots, (x_n, y_n))$ gözlemleri dikkate alınır.
- Rastgele etkili karma regresyon, her bir y_i 'nin lineer bir şekilde X_i orta değişkeni ile ilişkili olduğunu ve regresyon katsayılarının (rastgele etkilerinin) çok değişkenli normal dağılımın bir karmasını takip ettiği varsayılmaktadır.
- Heterojenliğin farklı seviyelerine karşı önlem almak için bütün regresyon doğrularını sınıflandırmada kullanılabilecek oldukça genel bir modeldir.
- Bileşenler üzerinde koşulluluk (k_i tarafından ifade edilen) sözkonusu olduğunda i-inci özne, normal regresyon modeline göre üretilmiş olan n_i uzunluğunun bir eğrisine sahiptir.

$$\begin{aligned}\varepsilon_{k_i} &\sim N_{n_i} (0, \sigma_{k_i}^2 I_{n_i}) \\ y_i &= X_i \beta_i + \varepsilon_{k_i}\end{aligned}\tag{3.4}$$

- EM algoritması, tahmin için kullanılır.

3.4.1.4. Orta Değişkene Bağlı Karma Oranlı Karma Regresyon Modeli

- Bir veya daha fazla orta değişkenin (kovaryantın) değişimi ile gerçekleşen her bir bileşen değişiminde, durumların ilgili oranlarının nasıl görüldüğü bilimsel açıdan ilgi çekici olabilir. Buna rağmen standart karma lineer regresyon modeli, orta değişkenler ile λ_j 'nin değişimini hesaba katmamaktadır.
- $\lambda_j(x_i)$ 'nin standart karma regresyon modeline dahil edilmesi ile elde edilen karma regresyon modelinin heterojen regresyon fonksiyonu aşağıda gösterildiği gibidir:

$$f(y_i \setminus x_i) = \sum_{j=1}^k \lambda_j(x_i) \phi(y_i \setminus x_i^T \beta_j, \sigma_j^2) \quad (3.5)$$

- Gizli sınıf olasılığı (karma oran), bir orta değişken z 'nin çok değişkenli lojistik regresyon fonksiyonu şeklinde değişkenlik gösterebilir:

$$P(C_i = c \mid z_i) = \frac{\exp(a_c + b_c z_i)}{\sum_{s=1}^K \exp(a_s + b_s z_i)}, \quad (3.6)$$

- Orta değişkene bağlı olan karma oranların tahmini için parametrik olmayan metodlar şart koşulmaktadır (Kernal Metodu).

3.5. Rastgele Etkili Karma Regresyon İle Eğri Kümeleme

Gaffney ve Smyth 'in 1999 yılındaki yayınlanmış çalışmalarında, odak noktaları eğri verisini modellemedir. Bunu yaparken de karma regresyon ile ilişkili olan model tabanlı eğri kümeleme algoritmasından direk olarak faydalanmışlardır.

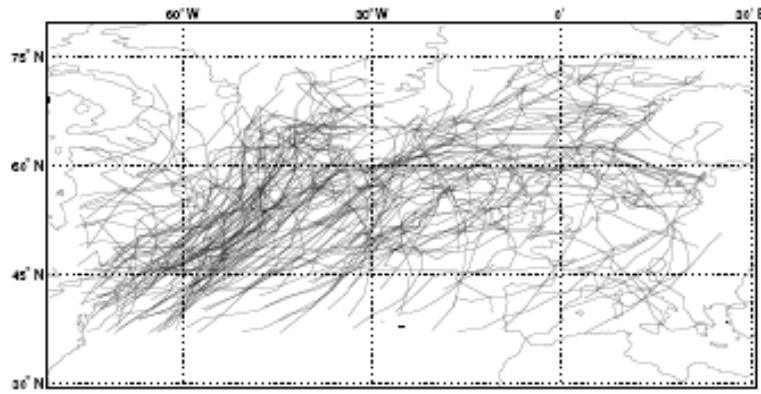
Önerilen yöntem, karma regresyonu genişletmektir. Bu da, standart regresyon karması ile lineer rastgele etkili modellerin birleşimi ile oluşturulan “rastgele etkili karma regresyon” ile sağlanmaktadır. Oluşturulan bu rastgele etkili karma regresyonda sonsal bir maksimum tahmin (MAP) kullanılarak genel bir EM algoritması geliştirilip, bu tekniğin hortum verisi (cyclone data) kümeleme probleminde nasıl uygulanabileceği gösterilmiştir.

Kümeleme; büyük veri gruplarını araştırmak ve anlamak için sıklıkla kullanılan genel bir araçtır. Birçok kümeleme algoritması, sabit boyutlu ve belirleyici özelliği olan vektörler üzerinde uygulanır. Ayrıca Gaffney ve Smyth 'in 1999 yılındaki bu çalışmalarında eğri gruplarının kümelenmesi problemiyle, nesnel veya ayrı gruplar tarafından üretilmiş uzunluk değişkenlerinin yörüngeleri bulunmaktadır.

y eğrileri, fonksiyon olarak bir bağımsız değişkene veya x değişkenlerinin bir kümesine bağlı olan zaman eksenini üzerinde ölçülmüş gözlemlerin bir dizisi iken, x zamanı belirtmektedir. Fakat genel olarak x , y olarak adlandırılan aynı aralık

üzerinde ölçülmüş değişkenlerin herhangi bir sayısı olabilir. Sabit boyutlu ve belirleyici özelliği olan vektörlerden farklı olarak y eğrileri, kayıp veri içeren gözlemleri de içinde barındırarak farklı ölçüm aralıklarında gözlemlenebilen uzunluk değişkenlerine sahip olabilir.

Çeşitli bilimsel şartlarda, bu tip veriler oldukça yaygındır. Şekil 3.4, atmosfer biliminde yörünge verisinin (trajectory data) bir kümesini göstermektedir.



Şekil 3.4. Kuzey Atlantik üzerinden geçen kasırğa yörüngeleri
(Gaffney ve Smyth, 1999)

Bu şekilde, 1980 ile 1995 yılları arasında kış aylarında (Kasım'dan-Nisan'a) Kuzey Atlantik üzerinden geçen Extra Tropik Kasırgaların sayısı (ETCs) görülebilmektedir (Gaffney, Robertson ve Smyth, 2001). Burada x eksenı boylamı, y eksenı enlemi göstermektedir. Hareketlerin yönü açık bir şekilde görülemezken, çoğu ETCs 'nin yönünün batıdan doğuya doğru olduđu görülebilmektedir. Burada eğri verisi için, zamanı dikkate alarak çok boyutludur diyebiliriz (Yani, her zaman enlem-boylam şeklinde, iki boyutlu gözlem vektörü bulunmaktadır). Atmosfer bilimcilerin araştırma soruları, kasırğa kümelerinin varlığı için kanıt buldurmayı amaçlar. Sorular, bu kümeler içerisinde düşen kar ve yağmur miktarı (yağış) veya global basınç modelleri gibi karakterize edilebilen davranış unsurlarının, bu kümelerin diğeri olağanüstü iklim koşulları ile ilişkisini tanımlayıcı cinstendir. Bu zamana kadar belirlenmiş bütün yörüngeleri göz önünde tutarak, yeni kasırgaların oldukça benzer

yörüngeleri ile ilgili tahminlerde bulunup, kasırga hareketlerinin modeli öğrenilebilir (Blender ve ark. ,1997).

Bu tip eğri verileri, standart vektör tabanlı kümeleme algoritmasıyla, sabit boyutlu ve belirleyici özellikleri olan vektörlerin bir kümesi için, veriyi azaltmak gibi bazı geçici ön işlem prosedürlerine başvurulmadan kümelenemeyebilir. Burada odaklanılan konu, regresyonun karmasıyla ya da diğer ismiyle karma regresyonla ilişkili olan model tabanlı eğri kümeleme algoritmasının bir kümesini kullanarak, eğri verisini modellemektir (De Sarbo ve Cron, 1998; Gaffney ve Smyth, 1999).

3.5.1. Model Tabanlı Eğri Kümelemenin Oluşumu

Model tabanlı eğri kümeleme, olasılığa dayalı olmayan kümeleme teknikleri ile karşılaştırıldığında yapısından kaynaklı bazı avantajlara sahiptir. Veri için üretken bir model sağlaması yanında, kayıp ve düzensiz aralıklı ölçümleri rahatlıkla elde edebilir. İki, üç veya daha yüksek boyutlu eğrilerle genelleştirilebilir. Ayrıca, zamanın bir fonksiyonu şeklinde eğri düzlüğünü açık bir şekilde modelleyebilir ve bu model örnek dışında tahmini etki bakımından değerlendirilebilir.

Önerilen yöntem, regresyon karmasını rastgele etkili regresyon karması olarak adlandırılan şekliyle geliştirmeye yöneliktir (Gaffney ve Smyth, 1999). Burada rastgele etkili regresyon karması, standart regresyon karmasıyla lineer rastgele etkili modellerin (Laird ve Ware, 1982) birleştirilmesiyle oluşmuştur. Rastgele etkili karma modeller ilk olarak Bayesci düzenlemede geçmiştir. Burada karma bileşenler genelleştirilmiş lineer modeller olarak kabul edilmektedir. (McCullagh ve Nelder, 1983). Tersine, rastgele etkili regresyon karması için, belirgin olan maksimum bir sonsal veya MAP tabanlı beklenti maksimizasyonunun (EM) geliştirmesi önerilir. Bu yaklaşım, Lenk ve Desarbo (2000) tarafından çıkarsama yapmak için kullanılan Markov Zinciri Monte Carlo (MCMC) tekniklerinin (Gelfand ve Smith, 1990) kullanımını geçersiz kılmaktadır (Gaffney ve Smyth, 1999).

3.5.1.1. Model Tabanlı Kümeleme

Model tabanlı kümeleme, olasılığa dayalı bir tekniktir. Model tabanlı kümelemede, verinin heterojen olduğu varsayılır. Bu durumda verinin, grup ya da küme tanımlı yoğunluk bileşenlerinin bir sonlu karmasıyla açıklanabileceği varsayılmaktadır (Banfield ve Raftery 1993; Fraley ve Raftery 1998).

Çok değişkenli veri kümesi X ,

$$P(x, \theta) = \sum_k \alpha_k P_k(x, \theta_k) \quad \begin{array}{l} \text{k-ıncı grup ya da küme için} \\ \theta_k \text{ parametreleriyle bileşen} \\ \text{yoğunluğu} \end{array} \quad (3.7)$$

k-ıncı grup ya da kümenin x tarafından oluşturulmasının koşulsuz olasılığı

karma yoğunluğu ile modellenir.

Standart EM tabanlı algoritmalar kullanılarak (Dempster ve ark. 1977, McLachlan ve Krishnan 1997) birçok karma modelin θ parametrelerini tahmin etmek mümkündür.

3.5.1.2. Model Tabanlı Eğri Kümeleme

Yukarıda sözü edilen karmaya benzer şekilde yapılır. Bununla birlikte $P_k(x, \theta_k)$ bileşen yoğunlukları, $P_k(y/x, \theta_k)$ koşullu bileşen yoğunlukları ile değiştirilir. Sonuçtaki karma model koşulludur. Genel olarak y , zaman ekseninde ölçülmüş gözlemlerin bir dizisi olsun. Fonksiyon olarak y , bir bağımsız değişkene veya X değişkenlerinin bir kümesine bağlıdır.

Eğri kümeleme modeli:

$$P(y/x, \theta) = \sum_k \alpha_k P_k(y/x, \theta_k) \quad (3.8)$$

şeklinde ifade edilmektedir. Bu model koşullu bileşen yoğunluklarının belirlenmesiyle tamamlanır.

3.5.2. Model Tabanlı Eğri Kümelemeden Standart Karma Regresyona Geçiş

Eğri kümeleme modelinin bileşenlerinin, regresyon bileşenleriyle değiştirilmesiyle standart karma regresyon elde edilir. Bu, K regresyon denkleminin bir karmasıyla potansiyel olarak heterojen yapıdaki veriyi modellemeye olanak sağlar.

y ve X arasındaki p-inci dereceden standart karma regresyon modeli:

$$y = X\beta_k + \varepsilon \quad \varepsilon \sim N(0, \sigma_k^2 I) \quad (3.9)$$

şeklinde ifade edilebilir.

Bu eşitlikte:

X: $n \times (p+1)$ tipinde regresyon matrisidir. X matrisinin ilk sütunu birlerden oluşur. Diğer bir ifadeyle model sabit terim içerir.

β_k : k-ıncı grup ya da küme için regresyon katsayılarıdır.

$\sigma_k^2 I$: k-ıncı kovaryans matrisidir.

Bu durum, eğri kümeleme modelindeki $P_k(y/x, \theta_k)$ koşullu bileşen yoğunluklarını, $X\beta_k$ ortalama vektörlü ve $\sigma_k^2 I$ kovaryans matrisli, $N(y/X\beta_k, \sigma_k^2 I)$ koşullu çok değişkenli normal dağılımın yoğunluklarına eşitlemeye denktir. Bu model ve modeldeki parametrelerin EM algoritmasıyla tahmini Gaffney ve Smyth (1999) tarafından çalışılmıştır.

3.5.2.1. Eğri Gruplarının Kümelenmesi Esnasında Rastgele Etkili Karma Regresyonun Önemi

Standart karma regresyon çerçevesinde, her bir kümenin θ_k parametrelili özel kümeler olarak modellenilebileceği varsayılp, eğri grupları bu kümeler üzerinde bir

karma şeklinde tanımlanabilmektedir. Bu durum, homojenliğin söz konusu olduğu alt kitleler için etkili bir değer olabilir. Oysa her bir alt kitle veya grup içerisinde, önemli bir heterojenlik söz konusu olduğunda çok daha dikkatli davranılması gereklidir.

Örneğin, K gruptan oluşan gözlemlerin bir kümesine sahip olduğumuzu varsayalım. Standart regresyon karmasının kullanımı çerçevesinde, her bir grup içindeki gözlemlerin hipotezini kurabiliriz. Bu gözlemler, uygun bir şekilde ortak grup bileşen modelini kurabilmek için yeterli derecede homojenlerdir. Yinede herhangi bir grup içerisinde önemli derecede heterojenlik söz konusu ise, bu durumda verinin daha iyi tanımlanabilmesi için, daha önce K adet olarak bildiğimiz grup sayısının artırılarak, daha fazla grubun oluşturulmasına başvurulmalıdır (Gaffney ve Smyth, 1999).

3.6. Karma Regresyon İle İlgili Yapılan Çalışmaların İncelenmesi

Karma regresyon ile ilgili temel kavramların paylaşımının ardından, pedagojik (eğitimsel) araştırmalarda ve sağlık bilimlerinin yapmış olduğu araştırmalarda, analizlerini karma regresyon yardımıyla gerçekleştirdikleri iki çalışmaya alt bölümlerde değinilecektir.

3.6.1. Pedagojik Araştırmalarda Bir Karma Regresyon Analizi Uygulaması

Cody S. Ding'in 2006 yılında yayınlanmış olan makalesinde öğrencilerin matematiksel başarılarında etkili olan faktörler, bunların birbiri ile olan ilişkileri ve bu faktörler çerçevesinde lineer karma regresyon analizinin klasik regresyon ile karşılaştırılması üzerine bir çalışma yapılmıştır.

İncelenen veri, milli eğitim istatistikleri merkezi, A.B.D eğitim departmanı tarafından sürdürülen okul öncesi çalışmasından (Early Childhood Longitudinal Study-ECLS) alınmıştır. Burada odaklanılan nokta, çocukların anaokulu ile başlayan okul öncesi deneyimleridir (Tourangeau, Nord, Lê, Pollack ve Atkins-Burnett, 2006). Ulusal bir temsilci olması amacıyla anaokulundan beşinci sınıfa kadar, çocuklardan alınan örneklem üzerinde çalışmalar gerçekleştirilmiştir. Örneklem, çeşitli ırklardan

ve dillerden bütün çocukları yansıtmaktadır. ECLS için örnekleme, ikili çerçevede 100 başlıca örnekleme birimiyle (PSU) çok aşamalı örneklemeyle dayalıdır. Bu makalede sadelik açısından sadece, beşinci sınıf öğrencilerinin 2004 yılı verileri derlenmiştir. Mevcut analiz içerisindeki örneklem, 1342 çocuktan oluşmaktadır. Bunlardan 650'si erkek, 692'si kızdır. Bu 1342 çocuktan, 797 'si beyaz, 126 'sı siyah, 230 'u İspanyol ya da Latin Amerika kökenli ve 48' i ise çok ırklıdır.

3.6.1.1.Ölçümler

Bu analizde bağımsız değişken olarak kullanılan dört tane ölçümden bahsedilmiştir. Bunlar:

Kendini ifade edebilme, Matematik üzerine anket, Kavrama becerisi (Marsh, 1990). Bu ölçüm, çocukların matematiksel yetenekleri ile ilgili kendilerini nasıl düşündüklerinin ve hissettiklerinin değerlendirmesidir. Bu ölçek takip eden konular üzerine elde edilmiştir: matematik sınıfları, matematik çalışmanın zorluğu, matematiğe karşı duyulan özel ilgi ve matematiğe karşı duyulan zevk 1 ile 4 arasında belirlenen skorlar ile ölçeklendirilmiştir. Analizde her bir katılımcının ortalama skoru kullanılmıştır.

Akademik değerlendirme, Matematiksel ölçü. Bu ölçüm, öğretmenin çocukların matematikteki akademik performanslarıyla ilgili değerlendirmeleridir. Öğretmenler, her çocuğun yeterliliğini değerlendirebilmek için takip eden konular üzerine sorular sormuşlardır: sayı kavramı, ölçümler, işlem, geometri, matematiksel stratejiler, cebirsel düşünceyle başlayabilme. Bu konular üzerine değerlendirmelerini ise 1 ile 5 arasında belirlenen skorlar ile ölçeklendirmişlerdir. Analizde her bir katılımcının ortalama skoru kullanılmıştır.

Sosyal değerlendirme, Öğrenmeye yaklaşımın ölçüsü. Bu ölçüm, öğretmenin çocukların sosyal yetenekleri ile ilgili kararlarıdır. Öğrenmeye karşı yaklaşım ölçüsü, öğrenme ortamından faydalanabilen çocukların kolaylıkla etkilenebilecek davranışlarını ölçer. Bu davranış biçimleri: çocuğun dikkati, verilen görevlerdeki sorumluluk bilinci, öğrenme gayreti, bağımsız öğrenebilmesi yani esnekliği, organize olabilme ve sınıf kurallarına uyumunu içeren altı öğeden

oluşmaktadır. Öğretmenler bu konular üzerine değerlendirmelerini 1 ile 4 arasında belirlenen skorlar ile ölçeklendirmişlerdir. Analizde her bir katılımcının ortalama skoru kullanılmıştır.

Sosyal değerlendirme ölçüsü, Oto-kontrol. Çocukların, davranışlarını kontrol edebilme yeteneklerini değerlendirebilmek amacıyla 4 öge belirlenmiştir. Bunlar: başkalarının haklarına saygı gösterebilme, öfkenin kontrol edilebilmesi, grup aktiviteler içerisinde akranlarının düşüncelerini kabul edebilme ve akranlarının baskılarına uygun bir şekilde cevap verebilme şeklinde sıralanabilir. Öğretmenler bu konular üzerine değerlendirmelerini 1 ile 4 arasında belirlenen skorlar ile ölçeklendirmişlerdir. Analizde her bir katılımcının ortalama skoru kullanılmıştır.

Yukarıdaki bütün ölçümlerde skorlar pozitif olarak kodlanmıştır. Yüksek skorlar, ileri bir kavrama becerisinin, öğretmenin çocuğun akademik başarısı üzerine olumlu bir tahminin ve sosyal yeteneğin işaretidir. Bu bağımsız değişkenler için bildirilen güvenilirlik aralığı, 0.79 ile 0.92 arasındadır (Tourangeau ve ark., 2006).

3.6.1.2. Analiz

Bu çalışmada bağımlı değişken, karma bir matematiksel yeterlilik olasılık skorudur ve bu değer, şu dokuz matematik beceri seviyesinin bir ortalaması alınarak hesaplanmıştır. Bu beceriler: sayma / sayı, orantılı ölçü, sıra / dizi, toplama / çıkarma, çarpma / bölme, sıra değeri, oran ve ölçüm, kesir ve alan / hacim şeklinde belirlenmiştir. Olasılık skorları 0.00 ile 1.00 arasında değerler alır. En büyük olasılık skoru, bütün matematik beceri seviyeleri arasındaki en yüksek başarıya işaret etmektedir.

Ek olarak, çocukların cinsiyeti ve ırkı orta değişken olarak gösterilebilir. Cinsiyet ve ırk değişkenleri, her bir gizli sınıf içerisindeki gözlemlerin sınıflandırma doğruluğunu artıracaktır. Bu araştırma içerisinde, çocukların ırkları beş kategoride belirlenmiştir: Beyaz, siyah, İspanyol ya da Latin Amerika kökenli, Asyalı (Pasifik adalı ve Kızılderili) ve çok ırklı.

Bağımlı değişkenin sürekli olması sebebiyle lineer karma regresyon modeli uygun bulunmaktadır. Bu analizde gizli kitle heterojenliğinin kaynakları

araştırılmaktadır. Bu araştırmadaki amaç, veriye en iyi şekilde uyum sağlayan gizli sınıfların sayısını belirlemektir. Böylece bağımsız değişken olarak düşünülen dört ölçüm dikkate alındığında karma regresyon modelleri, 1 gizli sınıflı modelden 4 gizli sınıflı modele kadar uygun olanı bulmak için test edilecektir. Ding bu analizde, LatentGold 4.0 programını kullanarak analizi gerçekleştirmiştir. Uygulamalar bölümünde, karma regresyon analizinin uygulanabileceği paket programlara değinilmiş. Bunlar arasından, yapay bir veri üzerinde gerçekleştirilecek analiz için AMOS 16.0 paket programı tercih edilmiştir.

Bütün bu modeller içerisinde bağımlı değişken, matematiksel yeterlilik olasılık skorlarıdır ve çocuğun cinsiyet ve ırk olarak belirlenmiş orta değişkenleriyle beraber, aynı bağımsız değişken kümesi kullanılmıştır. Bu dört model arasından, AIC bilgi kriter değeri en küçük olan model aranmaktadır. Dört modelin incelenmesinin ardından AIC bilgi kriter değerinin en küçük olması sebebiyle, üç sınıflı regresyon modelinin diğer modellere göre daha uygun olduğuna karar verilmiştir. Dikkat edilecek olursa 1 sınıflı model, diğer modellerle karşılaştırıldığında en büyük AIC'ye sahiptir. Buradan anlaşılmaktadır ki, kitle homojenliği örneklem içerisinde umulduğu gibi gerçekçi bir varsayım sağlamayacaktır. Deneysel ve anlamlı bir inceleme sonunda, üç sınıflı lineer regresyon modelinin en uygun olduğuna karar verilmiştir. Bu üç sınıflı modelde regresyon katsayıları ve hata varyansı, sınıfa bağlıdır. Bu da herhangi bir eşitlik sınırlandırması olmadan tahmin edilebilirliklerinin, esnekliğin göstergesidir.

3.6.1.3.Sonuçlar

Tablo 3.2, her 3 gizli sınıf için de regresyon katsayılarını, tahmini sınıf oranları ve ortalama matematik olasılık skorlarını içermektedir. Tablo 3.3, sınıflama profil bilgilerini göstermektedir.

Tablo 3.2'de, sınıf 1 örneklemin %57 'lik kısmını oluşturmaktadır. Matematiksel başarı, önemli ölçüde sadece öğretmenlerin matematiksel yetenek üzerine yaptıkları değerlendirmeler (ARS-Math) ile ilişkilendirilmiştir. Bu değişken matematiksel başarıda, varyansın yaklaşık %49 'luk bir kısmını açıklamaktadır. Bu

Gizli sınıf 2 ‘deki gözlemler için matematiksel başarılarının, önemli ölçüde öğretmenlerin değerlendirmeleriyle ve öğrencilerin öğrenmeye olan yaklaşımlarıyla ilişkili olduğu görülebilir. Bu iki değişken matematiksel başarıdaki varyansın yaklaşık % 63 ‘lük kısmını temsil etmektedir.

Tablo 3.2. Parametre tahminleri ve sınıf büyüklükleri (Ding, 2006)

$$^* p < 0.05$$

Tablo 3.2 ‘de standart hatalar parantez içerisinde belirtilmiştir. ^a, regresyon katsayılarının $p < 0.05$ kriterinde sıfırdan oldukça farklı olduğuna işaret eder. ^b, klasik regresyon analizinde regresyon katsayılarına işaret eder.

Lineer karma regresyon modeli ile klasik regresyon analizini karşılaştırmak amacıyla; klasik regresyon analizi, cinsiyet ve ırk bakımından kontrol ederken aynı bağımlı değişken ve bağımsız değişkenler ile uygulanmıştır. Sonuçlar Tablo 3.2 ‘nin son sütununda gösterildiği gibidir.

Tablo 3.3. Gizli sınıfların orta değişkenler ile ilişkisi (Ding, 2006)

	Class1	Class2	Class3
Covariates			
Gender			
Male	56.63%	38.81%	25.36%
Female	43.37%	61.19%	74.64%
Race			
White	82.21%	32.62%	0.81%
Black	3.43%	18.47%	7.96%
Hispanic	8.99%	28.87%	14.84%
Asian	3.05%	15.63%	62.32%
Multiracial	2.32%	4.40%	14.07%

Ding’in bu pedagojik araştırmasının sonucunda anlaşılmaktadır ki; yüksek matematik skorlarına sahip bir çocuk genellikle, matematikte yüksek kavrama becerisi olan, matematiksel becerileriyle ilgili öğretmeni tarafından yüksek değerlendirmelere sahip ve öğrenmeye yaklaşımının öğretmen tarafından değerlendirmesinin de yüksek olduğu bir öğrenci olacaktır.

Ayrıca, daha önce karma regresyon ile ilgili paylaşılmış olan varsayımlar düşünüldüğünde:

Her bir gizli sınıf için regresyon parametre tahminlerinin değişkenlik göstermesi ve bu gizli sınıflar içerisinde açıklanması, veriye en iyi uyum sağlayan gizli sınıf sayılarının tanımlanması ve bilinen bir gizli sınıf için potansiyel açıklayıcı değişkenleri test etme şeklinde sıralayabileceğimiz varsayımların, bu araştırma içinde dikkate alındığı görülebilmektedir.

3.6.2.Yenidoğan Hastanesinde Kalış Süresine Dayalı Rastgele Etkili Sınırlı Karma Regresyon Uygulaması

Ng, Yau ve Lee 2003 yılında yayınlanan çalışmalarında, bir iki bileşenli karma regresyon modelini, heterojenlik kavramını ve aynı zamanda gözlemler arasında ilişkinin var olduğu durumları göz önünde tutarak incelemişlerdir.

Örnek olması amacıyla metodu, yenidoğan kalış süresinin (LOS) analizinde uygulamışlardır. Bu çalışmalarında ilgili faktörlerin tanımlanmasının hastanedeki kalış sürelerini etkilediği gösterilmiştir. Gerçekleştirilen bu araştırmayla, sağlık hizmetleri planlaması ve kaynak ayrılması için önemli bilgiler elde edilebilir.

Bu makaledeki amaç, LOS sonuçlarındaki heterojenlik ile özdeki ilişkililiğin ikisine birden uyum sağlayan bir iki bileşenli karma modelinin önerilmesi üzerinedir. Başka bir açıdan bakacak olursak, bileşen dağılımları ve karma olasılıkta eş zamanlı rastgele etkilerin belirlenmesiyle oluşturulan Thompson ve ark. (1998) 'nın iki bileşenli karma regresyon modelinin genişletilmiş şekli dikkate alınabilir. Parametre tahminleri, McGilchrist (1994) ve McGilchrist ve Yau (1995) yaklaşımıyla genelleştirilmiş lineer karma modelini (GLMM) takiben elde edilmiştir.

Ng ve arkadaşlarının belirttiği problem, çeşitli belli başlı alt kitlelerden birine ait olabilen durumların, gözlemlenmiş neticelerinin heterojen bir ortamın sonuçları olmasıdır.

Rastgele etkili sınırlı karma modeller üzerine yapılan çalışma, hastanedeki kalış sürelerinin (LOS) analizi üzerine yapılmıştır. Hastaların hastanedeki kalış süreleri, kolaylıkla ulaşabilecek bir indikatördür. Kalış süreleri aynı zamanda veremin kabul edilebilir bir dayanağıdır (Leyland ve Body, 1997; Lee ve ark.,1998). İlişkili grup teşhisleri gibi; pek çok hastanenin vakaya göre yaklaşım şekli, uzun süreli kalan hastaların verem dayanaklarının, kısa süreli hastalara göre farklı olduğu varsayımı üzerinedir. Böylece, heterojen LOS gözlemleri bir iki bileşenli karma model ile modellenebilir. Bu da, bir bileşenin kısa süreli alt kitleye dahil olurken, diğer bileşenin uzun süreli alt kitleye dahil olmasıyla gerçekleşecektir (Xiao ve ark., 1999; Lee ve ark., 2001).

Ng ve arkadaşlarına göre aynı hastaneden toparlanmış gözlemlerin çoğunlukla ilişkili olması, LOS gibi hasta sonuçlarının analizinde bir problem teşkil etmektedir. Kümelenmiş verinin bağımlı olması (hastaların hastanelere yerleştirilmesi) yüzeysel topluluklar ve yanıltıcı neticelerle sonuçlanabilir. Leung 1998’deki son çalışmasında bir lineer etkili karma modelini, uç LOS gözlemlerinin keyfi olarak düzenlenmesinin ardından, bağımlılık durumunu hesaba katarak dikkate almıştır.

3.6.2.1.Model Yapısı

İki bileşenli dağılımın bir karmasında, genel anlamda herhangi bir kayba uğramadan kısa süreli ve uzun süreli yatan hasta alt grupları sırasıyla dahil edilerek değerlendirilebilir. Y_{ij} i-inci hastanedeki j-inci gözlem için LOS’i, M hastane sayısını, n_i ise i-inci hastanedeki gözlemlerin sayısını ve N, $N = \sum_{i=1}^M n_i$ şeklinde gözlemlerin toplam sayısını göstermektedir. Y_{ij} ile ilişkili olarak risk değişkenlerinin bir vektörü x_{ij} şeklinde ifade edilebilir.

Y’nin olasılık yoğunluk fonksiyonu için bir iki bileşenli karma model,

$$f(y_{ij}; x_{ij}) = \pi(x_{ij})\phi_1(y_{ij}; x_{ij}) + (1 - \pi(x_{ij}))\phi_2(y_{ij}; x_{ij}) \quad (3.10)$$

formunu alır.

Burada, $\pi(x_{ij})$ ilk bileşene veya alt kitleye ait olan hastanın olasılığını ifade etmektedir. $\phi_g(y; x)$, $g=1,2$ g-inci bileşen dağılımıdır.

3.6.2.2.Tahmin için EM algoritması

EM algoritması, eksik verilere sahip birimlerin tahmininde kullanılan maksimum benzerlik tahminlerini içeren bir yöntemdir. Bu algoritma, bir objenin hangi kümeye ait olduğunu belirlemede kesin mesafe ölçütlerini kullanmak yerine tahminsel ölçütleri kullanmayı tercih eder ve objelerin hangi kümeye yerleşeceğini, çan eğrisi şeklinde bir yapı içerisinde belirler.

Tahmin prosedürü, McGilchrist ve Yau (1995) ‘dan ileri gelmektedir. Metod en iyi lineer yansız tahmin edici (BLUP) ile başlayıp, uygun rezidü maksimum-likelihood (REML) tahmin edicilerin ifade edilmesiyle geliştirilmektedir.

Ψ bilinmeyen parametrelerinin, BLUP tahminleri EM algoritması yolu ile çözülen, $\partial L/\partial \Psi = 0$ eşitliğinin bir çözümü şeklinde ifade edilebilir (McLachlan ve Krishnan, 1997). Bir eksik veri problemi şeklindeki tahmin prosedürünü ortaya çıkarmak için, gözlenemeyen değişken vektörü z öne sürülmektedir. Her y_{ij} gözlemi için, uygun bir iki boyutlu indikatör z_{ij} değişkeni mevcuttur. Örneğin, $z_{ij} = (1,0)^T$ y_{ij} ’nin ilk bileşene ait olduğunu göstermektedir. Diğer taraftan eksik verili rastgele etkilerin tüm verinin büyük bir kısmını oluşturuyor olması EM algoritmasını önemli derecede yavaşlatacaktır (Meng ve Dyk, 1998). Ayrıca, eksik değişkenler olarak rastgele etkilere işlemler yapıldığında, E-adım içindeki log-likelihood ifadesinin analitik formu, elde etmesi güç olan yüksek boyutlu integrasyonlar içermektedir. Öyle ki, sayısal integrasyon daha fazla güvenilir veya uygulanabilir olamayacaktır.

3.6.2.3. Analiz

Doğumdan sonra, bebeğin yenidoğan hastanesinde yeterli bir süre kalması, önemli bir konudur. Bazı çalışmalarla varılan kanıya göre, yeni doğmuş bir bebeğin hastalığının tekrar nüksetme riskinin yüksek olması sebebiyle yeni doğan bir bebeğin erken taburcu edilmesi herhangi bir riske girmeden gerçekleştirilmelidir (Marbella ve ark.,1998). Yenidoğan hastanesinde kalış süresi ve bunun belirleyici faktörleri, taburcunun planlanması açısından önem arz ederken, aynı zamanda hastanenin bu konu ile ilgili kaynak ayırması açısından da faydalı olacaktır.

Ng ve arkadaşları (2003) tarafından yapılan bu çalışma, 1998 ile 1999 yılları arasında Batı Avustralya bölgesindeki 23 hastanede toplam 206 yeni doğmuş bebek verisine dayalı olarak gerçekleşmiştir. Sonuç değişkeni (yenidoğan yatış süresi), taburcu olunmasına izin verilene kadar geçen gün sayısı olarak tanımlanmıştır. Bu örneklem için, LOS (hastanede kalış süresi) aralığı 1 ile 46 gün arasında belirlenmiştir.

Yenidoğan kalış süresinin potansiyel belirleyicileri; teşhis sayısı, prosedür sayısı, doğuştan durumlar (Avustralya yerlisi / Yerli olmayan), izin şekli (isteğe bağlı / acil durum) ve cinsiyet şeklinde sıralanabilir. LOS değişkenin çarpıklığından dolayı (Leung ve ark.,1998) ham veriye bir log-dönüşümü uygulanmıştır.

Veri ilk olarak bileşen sayısı üzerinde herhangi bir kısıtlamada bulunmadan normal karma dağılım ile EMMIX yazılımı kullanılarak oluşturulmuştur. (McLachen ve Peel, 2000). Daha sonra karma bileşenlerin sayısı, Akaike (AIC) ve Bayesin (BIC) belirleme kriterlerine dayalı olarak belirlenmiştir ve likelihood-oran test istatistiğinin p-değerini belirlemek için parametrik bootstrap örnekleme yaklaşımı tanımlanmıştır. Bu seçim kriterlerinin uygulanması sonucu parametrelerin maksimum likelihood tahminleri, Tablo 3.4 ‘de görülmektedir.

Tablo 3.4. Normal Karma Dağılım ile Parametre Tahminleri (Ng ve ark., 2003)

Number of components, g	Parameter estimates	Log-likelihood	AIC	BIC	p – value ^a for g versus g+1
1	$\pi_1 = 1.000$ $\pi_2 = 2.398$ $\sigma_1^2 = 0.733$	-260.340	524.7	531.3	0.01
2	$\pi_1 = 0.640, \pi_2 = 0.360$ $\mu_1 = 1.993, \sigma_1^2 = 0.636$ $\mu_2 = 3.118, \sigma_2^2 = 0.097$	-244.306	498.6	515.3	0.40
3	$\pi_1 = 0.066, \pi_2 = 0.552, \pi_3 = 0.382$ $\mu_1 = 0.563, \sigma_1^2 = 0.227$ $\mu_2 = 2.097, \sigma_2^2 = 0.353$ $\mu_3 = 3.149, \sigma_3^2 = 0.093$	-241.380	498.8	525.4	-

^a, 99 bootstrap örnekleme yaklaşımına dayalıdır.

İki bileşenli normal karma dağılımının, basit yoğunluk fonksiyonundan daha iyi bir uyum sağladığı bulunmuştur (AIC belirleme kriteri en küçük). Fakat modele ekstra bir bileşen eklenildiğindeki durumlar henüz geliştirilmemiştir.

Tablo 3.5’de veri için, rastgele etkili bir iki-bileşenli karma regresyon modelinin oluşumundaki sonuçların özeti görülmektedir. Teşhis sayısı ve uygulanan prosedür sayısı gibi kliniksel ölçümler, yenidoğan kalış süresi üzerinde önemli etkiye sahiptir. Özellikle, teşhis sayısı kısa süreli kalış alt gruplarında kalış süresini uzatmak için önemli bir indikatördür. Diğer taraftan, uygulanan prosedür sayısındaki bir artış, grup içerisindeki kalış süresinin uzamasının yanı sıra, uzun süreli kalış alt kitlesine ait olan olasılığın artışına da sebep olacaktır. Erkek bebekler uzun süreli alt gruplarda en yüksek olasılığa sahip iken, Avustralya yerlileri daha uzun süreli kalış süresini tecrübe etmektedirler. Acil hasta kabulünde, hastanın hastanede kaldığı süresinin daha uzun bir periyotta devam etmesi muhtemel bir durumdur. Kısa süreli kalımların alt gruplarında negatif katsayı, hızlı bir şekilde karar verilerek uygun bir tedavinin veya prosedürün uygulanmasıyla az sayıda acil giriş vakalarını yansıtırken, diğer acil müdahalelerde, uzun süreli kalışlar için daha yüksek bir olasılık söz konusudur. Son olarak, kısa süreli kalan hastalar için kalış süresi içerisindeki hastane değişkenliğinin önemli oluşu, ilgili varyans bileşen tahminiyle açık olarak görülebilmektedir (Ng ve ark., 2003).

Tablo 3.5. Rastgele etkili iki bileşenli karma regresyon modeli sonuçları
(Ng ve ark., 2003)

Parameters	Reference category	Logistic part	Short-stay component	Long-stay component
Constant		3.183*(1.045)	1.499*(0.141)	2.406*(0.176)
Number of diagnosis		0.288(0.156)	0.207*(0.032)	0.008(0.016)
Number of procedures		-1.024*(0.304)	-0.081(0.070)	0.044*(0.018)
Indigenous status:		0.928(0.657)	0.481(0.118)	0.530*(0.100)
Aboriginal	Non aboriginal			
Admission Status:		-2.979*(0.926)	-0.582*(0.121)	0.018(0.110)
Emergency	Elective			
Gender:		-1.659*(0.545)	-0.130(0.113)	0.101(0.064)
Male	Female			
Variance component		0.246(0.513)	0.188*(0.081)	0.116(0.061)
σ_g			0.460	0.214

*p-value<0.01.

3.6.2.4. Sonuçlar

Ng ve arkadaşlarının bu örnek çalışmalarında, tez çalışmasında odaklanılan bağımlı değişkenin sürekli ve lineer olma sınırlandırmalarından farklı olarak logistik tipte bir yapı söz konusu olsa da, bu örneğin paylaşılmasındaki amaç; genel anlamda sağlık bilimlerinde karma regresyonun yerini ve analiz sonuçlarının nasıl yorumlanabileceğini paylaşabilmektir.

Bu çalışmada yenidoğana ait kalış süresi için iki-bileşenli karma regresyon modelinin uygulaması gösterilmektedir. Hastanede yatan yenidoğmuş bebeklerin, süreçteki önemli dereceli heterojenliği vurgulanmıştır. İki bileşenli karma modelin kullanılmasıyla metod, daha önce Leung ve ark. (1998) tarafından bahsedilen verinin keyfi düzenlemesini önlemiştir. Ayrıca Thompson ve ark. (1998) 'ın yaklaşımı, lineer açıklayıcı değişkenler içerisindeki rastgele etkiler vasıtasıyla hastaneler arasındaki değişimin eş zamanlı olarak hesaba katılması için iki bileşenli regresyon karma modelini geliştirmektedir. Yenidoğana ait kalış süresi ile ilgili faktörler, karma bileşen üyeleri ve tanımlanmış olan bütün kısa ve uzun süreli kalışların alt grupları için kalış süresinin ortalamasında belirleyicidir. Bu, taburcu planlaması ve kalış süresinin etkili yönetimi açısından kural oluşturucuları için önemli bilgiler sağlamaktadır (Ng ve ark., 2003).

Bu çalışmada GLMM yaklaşımıyla, rastgele etkili bir iki bileşenli normal karma regresyon modeli önerilmiştir. Metod, üstel aileden gelen yoğunluk bileşeniyle g-bileşenli karma model şeklinde genelleştirilebilir (Ng ve ark., 2003). Dolayısıyla, sınırlı karma GLMM'nin gelişiminde genel bir ifade elde edilecektir.

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULAMASI VE BAYES TAHMİN YÖNTEMİ

Karma regresyon analizinin gerçekleştirilebileceği çeşitli paket programları mevcuttur. Bunlardan belli başlıları: Mplus (Muthen ve Muthen, 2001), GLLAMM (Skrondal ve Rabe-Hesketh, 2004), LatentGold (Vermunt ve Magidson, 2005) ve AMOS 16.0 (Arbuckle, 2007) şeklinde sıralanabilir.

Bu bölümde, yapay bir veri kümesi ile karma regresyon analizinin uygulaması SPSS Science ürün ailesinden olan AMOS 16.0 yazılımı ile gerçekleştirilecektir.

4.1. AMOS Paket Programı

AMOS (Analysis of Moment Structure), modelleri analiz ederken faktör analizi veya regresyondan daha fazlasına ihtiyaç duyulduğunda kullanılabilecek, kullanımı kolay bir yapısal eşitlik modelleme (SEMs) aracıdır.

Özellikle gözleme dayalı verilerle çalışanlar için büyük kolaylıklar sağlayan AMOS yardımıyla; ölçme veriler, yarı deneysel veriler ve panel verileri için nedensel modeller kurulabilir. Gözlem verilerine göre oluşturabilecek “latent” değişkenleri ile modellerinizi zenginleştirebilirsiniz. “Full Information Maximum Likelihood” metodu ile daha doğru tahminler yapabilir, uygunluk faktör analizi (confirmatory factor analysis) ile oluşturulan değişkenleri kontrol edebilir, kullanımı kolay görsel rota (yol) diyagramı ile modeller kolaylıkla çizebilir, değişkenleri dikdörtgen veya oval şekillerle tanımlayabilir, aralarındaki ilişkiyi göstermek için bu değişkenleri oklarla birleştirebilirsiniz. Basit diyalog kutularıyla parametre kısıtları ekleyebilir, çalışma sonucunda oluşan parametre tahminlerine göre modelinizi kolaylıkla değiştirebilirsiniz.

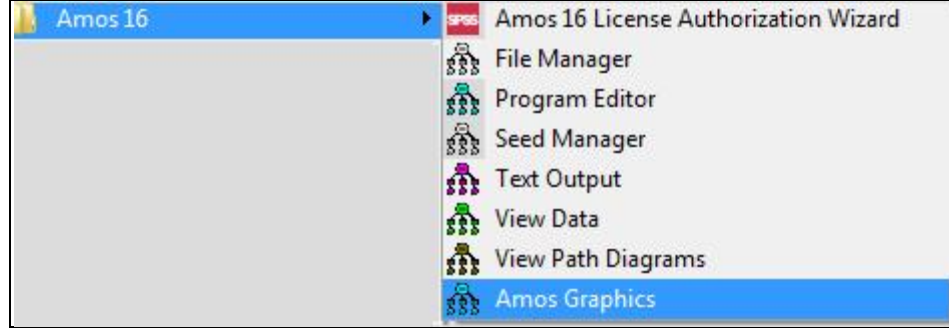
4.1.1. AMOS’un kurulumu ve arayüzü

AMOS 16.0 paket programının kurulumunun ardından (AMOS 16 License Authorization Wizard’ı kullanarak) gerçekleştirilmesi planlanan analizler, Şekil 4.1

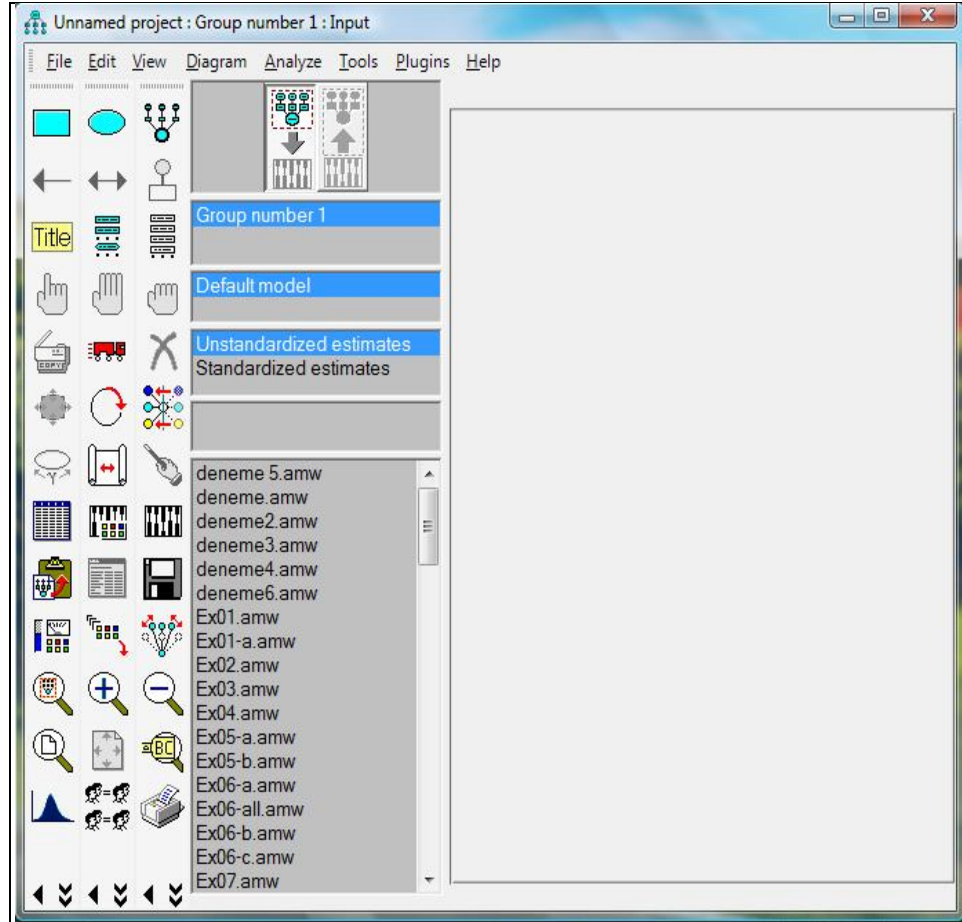
4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

de görüldüğü üzere AMOS Graphics sekmesinin seçilmesi ile karşımıza gelen, arayüz ekranı (Şekil 4.2) üzerindeki menüler yardımı ile meydana gelecektir.



Şekil 4.1. AMOS paket programı içindeki sekmeler



Şekil 4.2. AMOS arayüzünden genel bir görüntü

Şekil 4.2’ deki arayüz incelendiğinde, en üstte File, Edit,... şeklinde menüler, sol sütun boyunca analizlerin gerçekleştirilmesinde yardımcı kısayol tuşları, sağ taraftaki büyük dikdörtgen alan içerisinde yol (rota) diyagram alanı, orta bölümde ise grup sayısı ve tahmin şekilleri ile hemen altında kayıtlı dosyaların genel bir görüntüsü görülebilmektedir.

4.2. Veri Hakkında

Burada karma regresyon kavramını ve analizinin nasıl gerçekleştirildiğini açıklamak amacıyla, iki yapay veri grubu kullanılacaktır.

4.2.1. İlk Veri Grubu

İlk olarak incelenen veri kümesi, Dosage ve Performance 1 (Dozaj ve Performans 1) dir. Dosage (dozaj), bazı tedavilerin yoğunluğunu, performance (performans) ise sadece bazı performans ölçülerini belirtmektedir. Group (grup), sayısal olmayan (string) bir değişkendir ve daha sonra karma regresyon analizindeki rolü açıklanacaktır.

Toplamda $n=1000$ ’lik örneklemin ilk 50 gözlemi Tablo 4.1 ‘de görülmektedir. Bu 1000 gözlemden oluşan örneklemin bütün verileri EK-2’de paylaşılmıştır. Şekil 4.3 ‘de Dozaj ve Performans 1 verisinin saçılım grafiği görülmektedir. Grafik, örneklem içinde iki farklı insan grubunu göstermektedir. Bir grupta dozajın yükselmesi ile performansta artış gözlemlenirken, diğerinde dozajdaki artışla performansta düşüş gözlemlenmektedir. Dolayısıyla burada varılabilecek ilk kanı, veri kümesinin heterojen bir yapıya sahip olduğudur. Bütün bir örneklem için, tek bir regresyon doğrusu oluşturmaya çalışmak yanlış olurdu. Diğer taraftan, her bir grup için iki düz çizgi, veriyi iyi bir şekilde açıklamaktadır.

Dozaj ve Performans 1 veri kümesi, burada sadece karma regresyon yapısına sahip bir veri kümesini ve nasıl bir saçılım grafiği oluşturacağını göstermek amacıyla kullanılmıştır. Karma regresyon analizinin detaylı anlatımı diğer veri kümesi üzerinde gösterilecektir

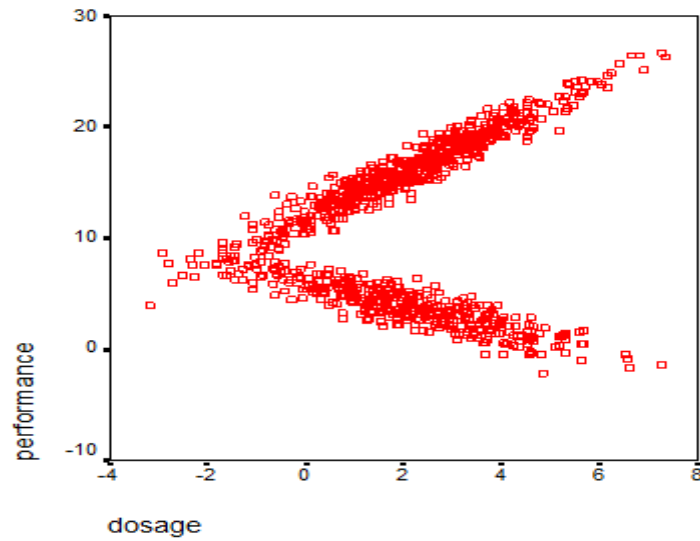
4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

Tablo 4.1. SPSS’de Dozaj ve Performans 1 için veri girişi

	dosage	performa	group		dosage	performa	group
1	5,26	-,35		26	3,97	1,37	
2	2,10	4,86		27	1,46	4,12	
3	1,93	5,20		28	5,15	1,12	
4	1,21	3,50		29	3,21	3,93	
5	1,25	2,61		30	4,00	2,14	
6	-,86	6,64		31	2,06	3,02	
7	3,85	2,31		32	3,37	2,07	
8	2,51	2,60		33	1,72	5,94	
9	1,29	4,06		34	4,36	,76	
10	,90	6,12		35	2,31	5,10	
11	2,11	4,23		36	2,49	3,93	
12	,42	6,68		37	-,11	6,47	
13	2,53	3,48		38	3,55	3,92	
14	,88	6,51		39	2,02	2,94	
15	,34	5,39		40	2,66	2,13	
16	4,88	1,90		41	1,39	4,74	
17	2,77	2,66		42	2,69	3,97	
18	1,14	4,26		43	1,86	4,97	
19	4,04	,55		44	4,50	1,12	
20	2,06	3,40		45	-1,46	6,26	
21	-,41	5,22		46	3,35	1,75	
22	-1,07	5,34		47	3,83	2,06	
23	1,54	3,63		48	-,61	4,86	
24	,73	4,86		49	,63	3,99	
25	,89	4,66		50	,74	2,69	

Graph



Şekil 4.3. SPSS paket programı ile oluşturulmuş Dozaj ve Performans 1 verisi için saçılım grafiği

4.2.2. İkinci Veri Grubu

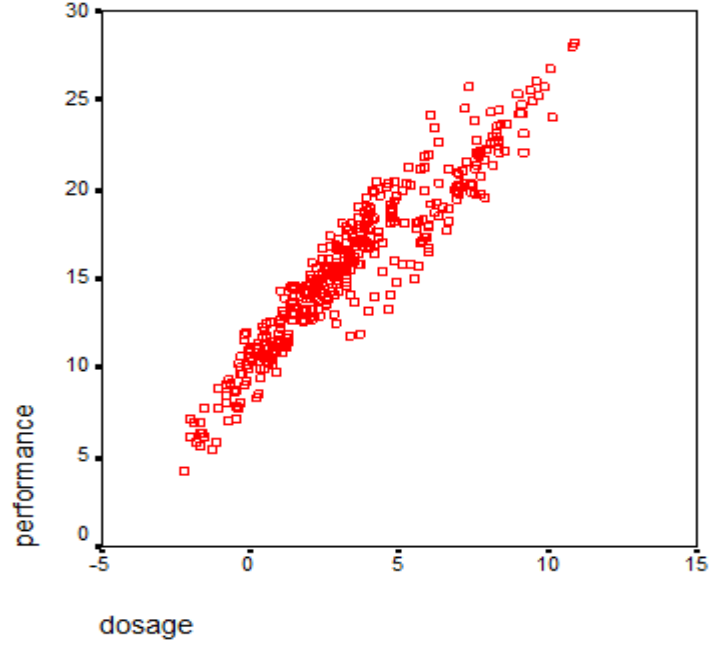
Dosage And Performance 2 (Dozaj ve Performans 2) olarak adlandırılan diğer veri kümesi ise, ilk veri kümesinde olduğu gibi dozaj, performans ve grup değişkenlerinden oluşmaktadır.

Tablo 4.2. SPSS’de Dozaj ve Performans 2 için veri girişi

	dosage	performa	group		dosage	performa	group
1	6,66	21,20		26	7,27	21,56	
2	5,66	15,70		27	7,64	22,05	
3	6,06	19,20		28	8,15	22,98	
4	9,19	23,13		29	7,48	20,24	
5	6,94	20,99		30	6,84	19,96	
6	5,16	18,04		31	5,72	16,97	
7	4,18	13,96		32	6,44	18,99	
8	8,08	22,52		33	8,21	22,69	
9	4,68	14,06		34	6,99	20,25	
10	7,86	19,55		35	4,87	14,84	
11	2,76	12,97		36	5,57	17,77	
12	3,70	11,96		37	6,68	18,22	
13	5,54	15,00		38	7,69	21,82	
14	7,06	20,14		39	10,76	28,04	
15	7,22	20,35		40	5,80	17,07	
16	9,84	25,74		41	6,20	18,66	
17	4,44	15,36		42	6,92	19,84	
18	8,63	23,68		43	8,29	23,54	
19	8,37	22,69		44	7,68	21,96	
20	6,93	19,47		45	8,44	23,64	
21	7,64	22,72		46	7,74	20,74	
22	5,66	18,18		47	5,87	17,33	
23	5,87	17,27		48	9,47	24,97	
24	5,37	15,81		49	5,99	16,66	
25	7,01	20,87		50	7,63	21,36	

Toplamda n=400 lük örneklemin ilk 50 gözlemi yukarıdaki gibidir. Bu 400 gözlemden oluşan veri kümesinin tümü EK-3’de paylaşılmıştır.

Graph



Şekil 4.4. SPSS paket programı ile oluşturulmuş Dozaj ve Performans 2 için saçılım grafiği

Şekil 4.4 ‘deki saçılım grafiğinde iki grubun varlığından bahsedebiliriz. Burada her grup kendi regresyon doğrusuna sahiptir. Her iki grup için de ayrı ayrı incelendiğinde dozajdaki bir birimlik artışa karşın performansta yaklaşık olarak iki birimlik artış söz konusudur. Bu sebeple, iki grup için regresyon doğrusunun eğimi yaklaşık olarak 2 dir. Öte yandan, iki grupta farklı sabitlere sahiptir. Herhangi bir dozaj düzeyinde bir grup içindeki performans 5 birim veya daha üzeridir. Bu veri kümesinin karma regresyon analizi, örneklemi gruplar halinde bölme ve her bir grup için ayrı bir regresyon doğrusu oluşturma girişimde bulunacaktır.

4.3. Veri Kümesi İçindeki Grup Değişkeni

Bu iki veri kümesinde de sayısal olmayan (string) bir yapıya sahip ve herhangi bir veri içermeyen, grup (group) olarak adlandırılan değişkenler

tanımlanmıştır. Değişkeni grup olarak adlandırmak önemli değildir. Herhangi bir değişken ismi de olabilirdi. Önemli olan sayısal olmayan bir değişken olmasıdır.

Analiz başlamadan önce eğer bazı durumlar gruplara atanmış ise, veri kümesinin grup sütunu içine grup isimleri koyulabilir. Örneğin, analizden önce örneklemin yüksek ve düşük performanslar gösterdiğini biliyorsak ve örneklem içindeki ilk iki kişinin yüksek performanslı ve daha sonraki üç kişinin ise düşük performanslı olduğunu biliyorsak, 4.3 numaralı veri tablosunun grup sütunu içine aşağıdaki tabloda görülebileceği şekliyle bilgileri girebiliriz:

Tablo 4.3. Grup olarak adlandırılan değişkenin SPSS’de veri girişi

	dosage	performa	group
1	6,66	21,20	high
2	5,66	15,70	high
3	6,06	19,20	low
4	9,19	23,13	low
5	6,94	20,99	low
6	5,16	18,04	
7	4,18	13,96	
8	8,08	22,52	
9	4,68	14,06	
10	7,86	19,55	
11	2,76	12,97	
12	3,70	11,96	
13	5,54	15,00	
14	7,06	20,14	
15	7,22	20,35	

Önceden sınıflandırılmış olan bu beş durum, kalan durumları sınıflandırmada yardımcı olmak amacıyla kullanılabilir. Seçilmiş farklı durumların gruplara olan ön atamaları, burada sadece bir varsayım olarak girilmiştir. Bu örnekte, hiçbir durum için gruplara ön atamada bulunmayıp, bu AMOS ‘a bırakılacaktır.

4.4. Karma Regresyon Analizi Öncesi Adımlar

Lineer karma regresyon modeli ile incelemeye alabileceğimiz kitle heterojenliğinin olası varlığında, bağımlı değişken ve bağımsız değişkenler arasında

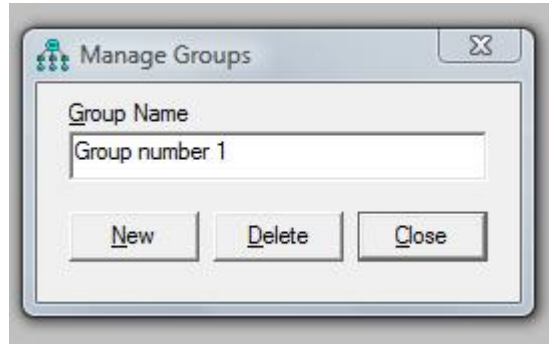
4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

nasıl bir ilişki olduğunun açıklaması Dozaj ve Performans 2 veri kümesinin analiziyle elde edilecektir. Gizli sınıf değişkeni c , bilinmeyen heterojenliğin modellenmesinde kullanılır. Orta değişken olarak ifade edebileceğimiz gözlenen grup üye değişkenleri, heterojenliğin göstergeleridir. Lineer karma regresyon analizinde ilk olarak, gizli sınıfların belirlenmiş bir sayısına ihtiyaç duyulmaktadır. Modeldeki tahmin sürecinde, model parametreleri tahmin edilir ve her bir sınıfa ait bütün gözlemlerin sonsal olasılıkları hesaplanır.

- AMOS menüsünden File $\implies$ New seçilerek, yeni bir yol diyagramının başlanır.
- Menüden Analyze $\implies$ Manage Groups seçilir.

Tablo 4.4. AMOS’da 1.grubun belirlenmesi



- İkinci bir grup oluşturmak için New i tıklarız.

Tablo 4.5. AMOS’da 2.grubun belirlenmesi



4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

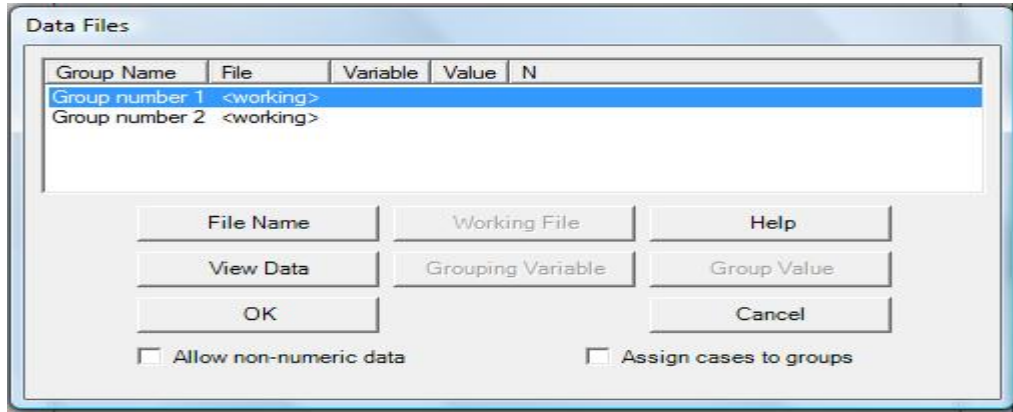
- Close u tıklarız.

Bu örnekte, bir iki-gruplu karma regresyon modeli oluşturulmaktadır.

4.4.1. Veri Dosyasını Belirleme

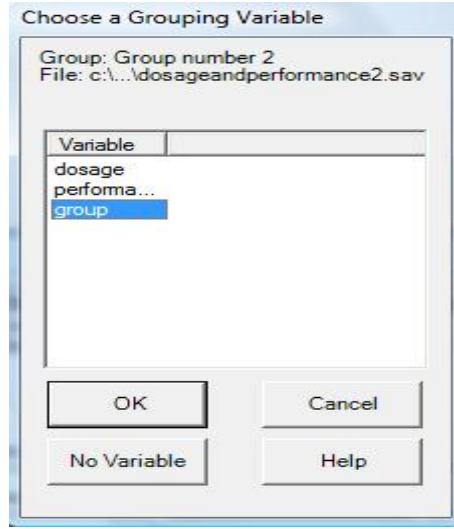
- Menüden, File $\Rightarrow$ Data Files seçilir.

Tablo 4.6. AMOS’da gruplar için veri dosyası belirleme



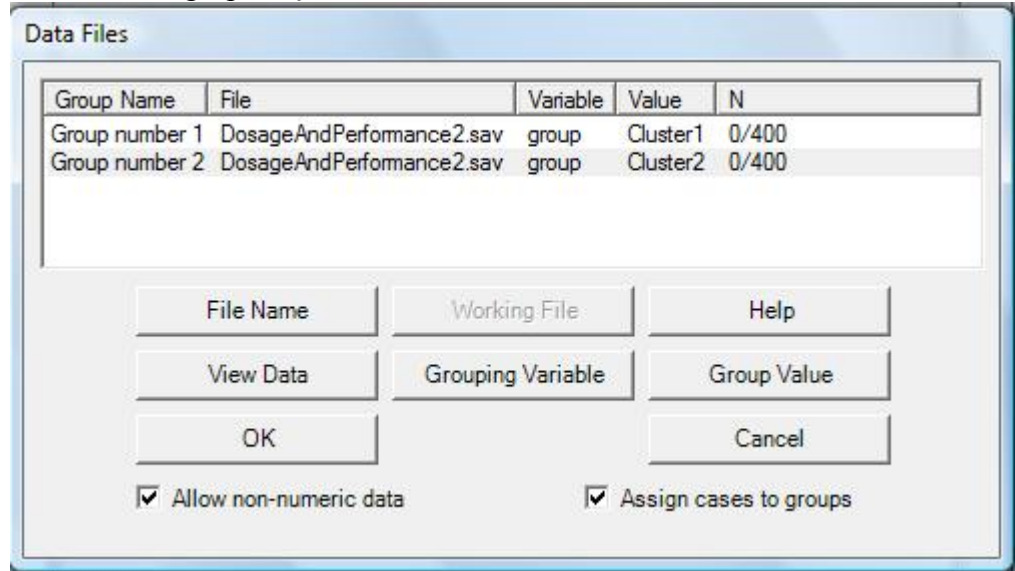
- Group number 1 satırının seçilebilmesi için tıklanması gereklidir.
- File Name tıklanıp, AMOS Examples Directory içindeki Dosage And Performance 2.sav dosyası seçilip open a basılır.
- Grouping Variable seçilir ve çıkan “Choose a Grouping Variable” diyalog kutusu (Tablo 4.7) içindeki “group” değişkeni iki kere tıklanır. Böylece programa, “group” olarak adlandırılan değişkenin bir gruptan diğerine ayırt edici olarak kullanılacağını sinyalini vermiş oluruz.

Tablo 4.7. AMOS’da grup değişkenin seçilmesi



➤ Group number 2 için de 1. grup için uygulanan adımların aynısı yani, aynı veri dosyası (Dosage And Performance 2.sav) ve aynı grup değişkeni (group) seçilerek tekrarlanır.

Tablo 4.8. AMOS’da 1. ve 2. gruplar için yapılan düzenlemelerin son hali ve gruplar için atama komutunun verilmesi



➤ “Assign cases to groups” işaretlenerek durumlar gruplara atanır.

Böylece otomatik olarak yan taraftaki, “sayısal olmayan verilerin dikkate alınması” ifadesi de işaretlenmiş olacaktır.

- Data Files diyalog kutusunu kapatmak için OK kliklenir.

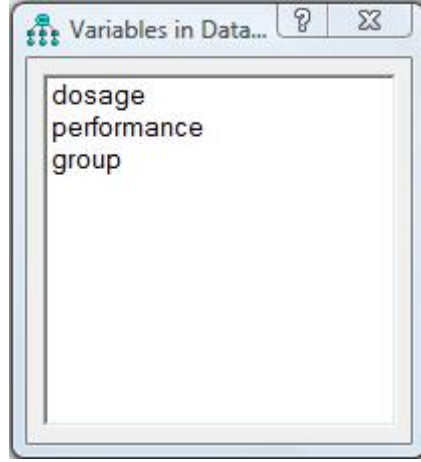
Şimdiye kadar olağan bir, çoklu grup analizi söz konusuydu, ta ki durumların gruplara atanması onay kutusu işaretlenene kadar. Böylece; bu çoklu grup analizi, karma model analizine dönüşmüş oldu. Bu onay işareti, eğer veri içinde gruplama değişkeni hala daha bir durumu bir gruba atamadı ise, AMOS’ a bir durumun bir gruba atanması gerektiğini anlatmaktadır. Dikkat edecek olursak, gruplama değişkenine bir değer belirlemek için, Group Value’yu işaretlemeye ihtiyaç yoktur. Veri dosyası gruplama değişkeni (group) için herhangi bir değer içermemektedir. Bu sebeple program, otomatik olarak grup değişkeni için değerler oluşturmuştur: Group number 1’deki durumlar için Cluster 1, Group number 2 ‘deki durumlar için Cluster 2 şeklinde.

4.4.2.Model Belirleme

Aşağıdaki adımlar takip edilerek, regresyon modeli için bir yol diyagramı çizilebilir:

- View  Variables in Dataset

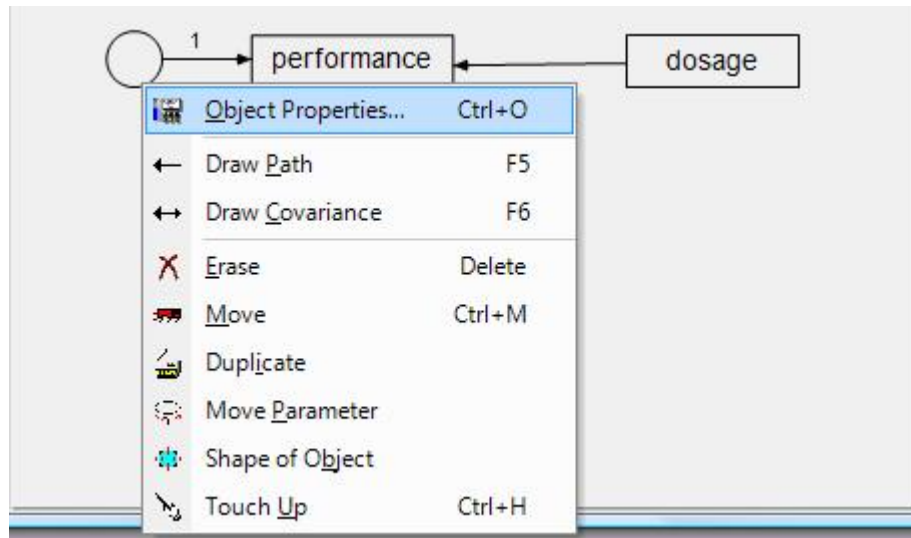
Tablo 4.9. AMOS’da, veri kümesi içindeki değişkenlerin gösterildiği kutucuk



Tablo 4.9’deki kutucuktan, bağımlı değişken performans ile bağımsız değişken dozaj, diyagramı oluşturmak istediğimiz alana sürüklenir.

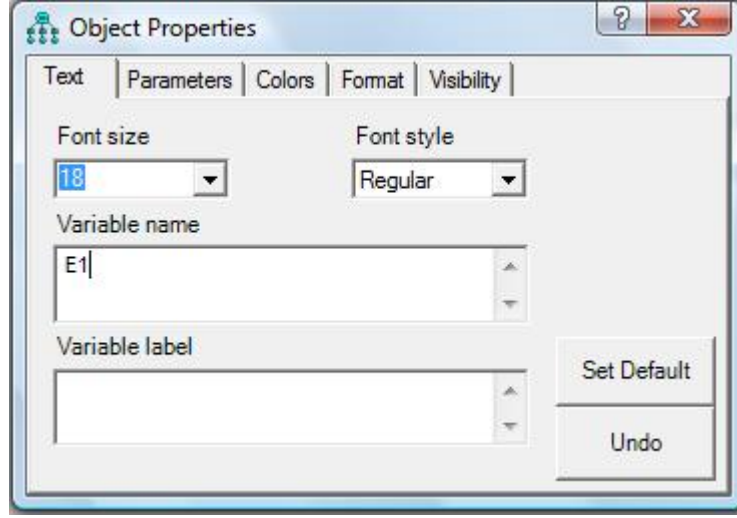
Değişkenlerin boş alana sürüklenmesi ile oluşturulan yol diyagramında, şekil 4.5’ de görüldüğü üzere ok üzerinde sağ tıklanıp “Object Properties” seçilir.

Object Properties sekmesinin seçilmesinin ardından karşımıza, bize yol diyagramı üzerinde çeşitli değişiklikler gerçekleştirebilmeyi sağlayan, Tablo 4.10’deki kutucuk gelecektir.



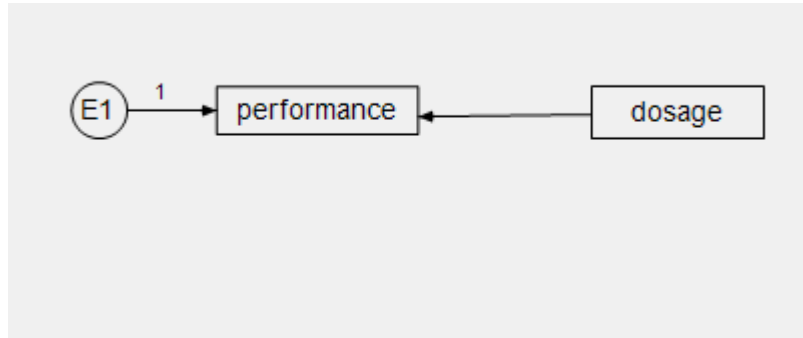
Şekil 4.5. AMOS’da oluşturulan yol diyagramı üzerinde düzenlemeler yapılmasını sağlayan sekmeler

Tablo 4.10. AMOS'un Nesne Özellikleri kutucuğu



Variable Name kısmına E1 yazılır. Böylece, dikkate alınmayan bağımsız değişkenler olabileceğinden verilerin tesadüfî değişimlerini gösteren hata terimi modele eklenir ve model içerisinde E1 olarak adlandırılır.

Yol diyagramının son hali;



Şekil 4.6. AMOS'da oluşturulan yol diyagramının son hali

Yukarıda görsel olarak ifade ettiğimiz yol diyagramını aynı zamanda, her bir grup için aşağıdaki şekliyle modelleyebiliriz:

1.grup için:

$$y_{i1} = \beta_{01} + \beta_{11}x_i + E_{i1} \quad (4.1)$$

2.grup için:

$$y_{i2} = \beta_{02} + \beta_{12}x_i + E_{i2} \quad (4.2)$$

Karma regresyon analizini gerçekleştirmek için öncelikle uygulamamız gereken adımlar yukarıda anlatılmıştır. Bayes tahmin yöntemi kullanılarak gerçekleştirilecek analize geçmeden önce, bu kavram ve AMOS paket programı ile nasıl uygulandığı üzerine ön açıklamalarda bulunulacaktır.

4.5. Bayes Tahmini

Bayesci bir bakış açısında, gerçek model parametreleri bilinmediğinden rastgele olarak dikkate alınıp, ortak olasılık dağılımı ile belirlenirler. Sadece bu dağılıma bakarak, bazı durumlarda parametrelerin farklı olduğu veya değiştiği ileri sürülmemektedir. Bundan ziyade dağılım, bilgi durumumuzu veya parametreler hakkında güncel olarak neler bilindiğini özetlemek amacıyla tasarlanmıştır.

Veri görülmeden önce, parametrelerin dağılımı önsel dağılım (prior distribution) olarak adlandırılır. Sonra veri gözlenir, veri tarafından sağlanan bulgular Bayes teorimi olarak adlandırılan bilindik bir formül tarafından önsel dağılım ile birleştirilir. Sonuçta, parametreler için güncellenmiş olan bu dağılım, sonsal dağılım (posterior distribution) şeklinde adlandırılır. Sonsal dağılım, önsel düşüncenin ve deneysel bulgunun bir kombinasyonunu yansıtmaktadır (Bolstad, 2004).

Araştırmacılar zor görselleştirmeler ve bir modelin parametreleri için ortak sonsal dağılımı yorumlamayla ilgilenmektedir. Dolayısıyla Bayes analizini uygularken, yorumlaması kolay olan sonsal dağılım özetlerine ihtiyaç vardır. Tek tek, her bir parametre için marjinal sonsal yoğunluğu grafiklendirmek, iyi bir başlama şeklidir. Genelde, marjinal sonsal dağılımlar özellikle büyük veri örneklemelerinde parametreler için, normal dağılıma benzeme eğilimindedir. Sonsal ortalama olarak adlandırılan, marjinal sonsal dağılımın ortalaması, bir parametre tahmini şeklinde bildirilebilir. Sonsal standart sapma, dağılımın standart sapması,

klasik standart hataya kesin olmamakla beraber benzeyen, kullanışlı bir ölçüdür (Arbuckle, 2007).

Bir güven aralığının benzeri, marjinal sonsal dağılımın yüzdeliğinden hesaplanabilir; 2.5 ile 97.5 yüzdelikleri arasındaki aralık, %95 lik bir Bayes güven aralığını oluşturabilir. Eğer marjinal sonsal dağılım yaklaşık olarak normal ise, %95 güven aralığı yaklaşık olarak ± 1.96 sonsal standart sapmalı, sonsal ortalamaya eşit olur. O zaman, güven aralığı parametre tahmini için normal örnekleme dağılımını varsayan olağan bir güven aralığı ile esasen aynıdır. Eğer sonsal dağılım normal değilse, aralık sonsal ortalama ile ilgili simetrik olmayacaktır. Öyleyse, Bayes versiyonu, klasik olana göre genellikle daha iyi özelliklere sahiptir.

Klasik bir güven aralığından farklı olarak Bayes güven aralığı, parametre ile ilgili bir olasılık ifadesiyle yorumlanmıştır;

$$P(a \leq \theta \leq b) = 0.95 \quad (4.3)$$

(4.3) nolu eşitliğin tam olarak anlamı, θ 'nın gerçek değerinin % 95 güvenle a ve b arasında yer aldığıdır. Marjinal sonsal bir dağılımın kuyruk alanları, hipotez testi için Bayes p değeri gibi de kullanılabilir. Eğer alanın % 96.5 i marjinal sonsal yoğunluk altında θ için a değerinin sağ tarafında kalıyor ise, $\theta > a$ alternatif hipotezine karşılık $\theta \leq a$ sıfır hipotez testi için Bayes p değeri, 0.045 dir. Bu durumda, % 96.5 güvenle alternatif hipotezin doğru olduğu söylenecektir. ($p < 0.05$)

Bayesci sonuç fikrinin geçmişi 18 inci yüzyıla kadar dayanmasına rağmen, istatistikçiler tarafından, yakın zamana kadar nadiren kullanılmıştır. Bayes metodlarının uygulanmasında yaşanan tereddüt, önsel dağılım seçiminde kişisel yorumlamadan ileri gelmektedir. Fakat Bayes analizinin nadiren kullanılmasındaki en önemli sebep, ortak sonsal dağılımların özetlenmesi için hesaba dayalı olan metodların zorluğu ve ulaşılamazlığıdır. Simülasyon tekniklerinin yeni bir sınıfı olan Markov Zinciri Monte Carlo (MCMC) kullanarak, karmaşık problemlerde dahi büyük-boyutlu ortak sonsal dağılımlardan, parametrelerin rastgele değerlerini çekmek olasıdır. Sonsal özetleri MCMC ile elde etmek, histogram grafiklerini

oluşturmak, örneklem ortalamalarını ve yüzdeliklerini hesaplamak kadar kolay olacaktır (Arbuckle, 2007).

4.5.1.Önsel Seçim

Bir önsel dağılım, bilinmeyen parametrenin nerede (yani hangi aralıkta) bulunabileceği hakkında araştırmacıların düşüncelerini sayısallaştırır. Kitle içerisinde bir değişkenin nasıl dağılmış olduğu ile ilgili bilgi, araştırmacılara ilgili parametreler için uygun önselleri seçmelerinde yardımcı olabilir (Arbuckle, 2007). Ayrıca gözlenen değişkenlerin sınırlandırılmış olması, parametreler üzerinde sınırlandırmalar koymak için bizlere yardımcı olacaktır. Örneğin, derecelmeler toplamıyla gerçekleştirilen Likert tipi bir anketten alınan değerlerin 0,1,...,10 şeklinde olduğu düşünüldüğünde, ortalamasının 0 ve 10 arasında değerler almak zorunda olduğu söylenebilir. Ayrıca elde edilebilecek maksimum varyans değeri 25 dir. Bu ögenin ortalama ve varyansı için önsel dağılımları, sınırlandırmaları uygulamak için belirleyici olabilir.

Pek çok durumda, mümkün olduğunca az bilgi paylaşan (sınırlandırmaları olan) önsel dağılımı belirlemek isteriz. Eğer parametre değerlerinin olasılıkları oldukça geniş bir aralık üzerinde yayılmışsa, önsel dağılım dağınık olarak adlandırılır. Aksi istenmediği takdirde AMOS, tek düze bir dağılım uygular.

Dağınık önsel dağılımlar, genellikle bilgilendirici olmayan olarak adlandırılrsa da, burada bu terim kullanacaktır. Fakat, hiçbir önsel dağılımın bilgilendirici olmadığını söylemek yanlış olur. Kabul edilebilir değerlerin bütün aralıkları üzerinde tek düze dağılımdan söz edemeyiz. Çünkü eğer parametre dönüştürülmüş ise, tek düze olmayı sonlandıracaktır. (Varsayalım ki, örnek için bir değişkenin varyansı 0 dan ∞ a düzgün dağılımlı olsun; o halde standart sapma düzgün dağılımlı olmayacaktır.) Bütün önsel dağılımlar az da olsa bilgi taşır. Veriden elde edilen belirtiler, önsel dağılımın etkisini azaltır. Örneklem çok küçük olmadıkça veya model ya da önsel dağılım eğer veriyle kuvvetli bir şekilde çelişki içinde değilse, Bayes analizden elde edilen cevaplar önsel bilgilerin değiştirilmesiyle çok az bir değişme eğilimi gösterecektir. AMOS, herhangi bir parametre için önsel dağılım

değişimini araştırmacılar için kolaylaştırmaktadır. Böylece bu gibi hassas kontroller daha kolay uygulanabilir (Arbuckle, 2007).

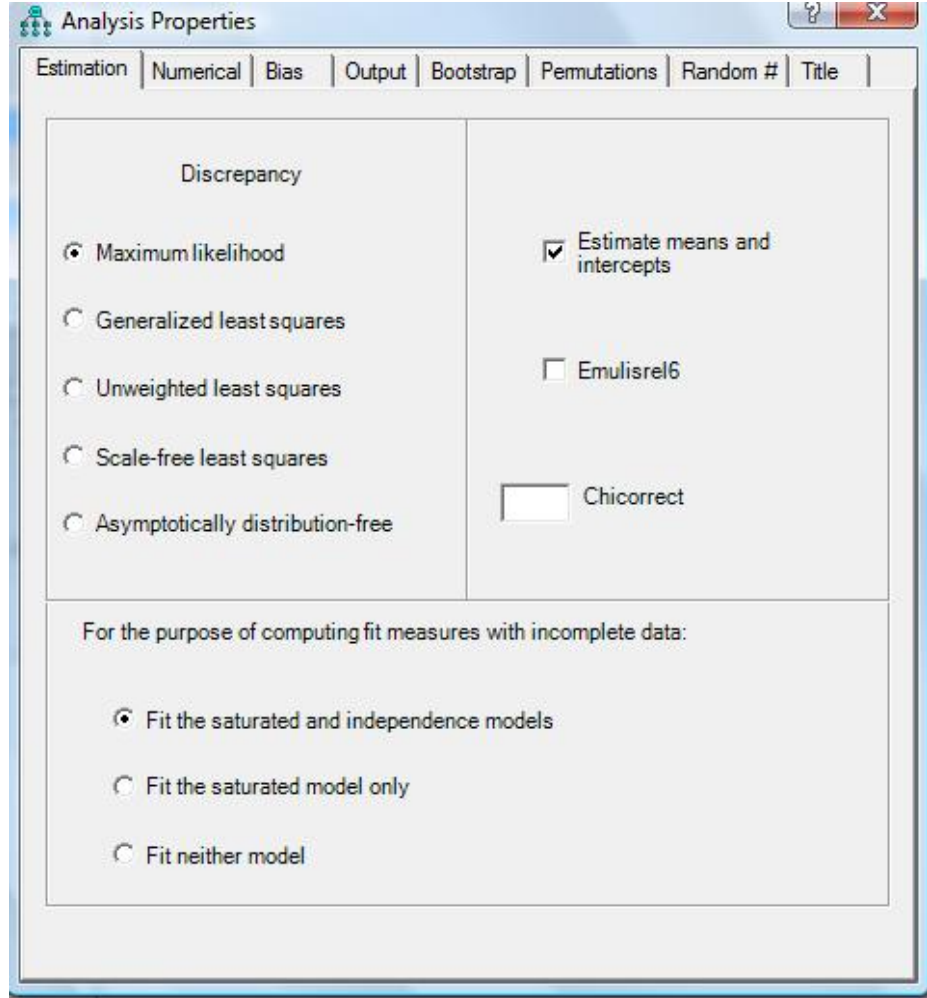
4.5.2. Bayes Analizi

Bayes analizi burada, Attig 'in 1983 yılında oluşturduğu hafıza verisinde (spatial memory data) gerçekleştirilecektir. Attig bu veri kümesinde, çeşitli sayfalardan oluşan 40 konun yer aldığı bir tanıtım kitapçığını kullanmış ve kişiler her bir konu ile ilgili üç hafıza testine tabi tutulmuştur. Bu kişiler ile bir süre eğitim çalışmaları yapıldıktan sonra hafıza testi aynı şekilde tekrar uygulanmıştır. Böylece eğitimden önce üç, eğitimden sonra üç performans ölçümü yapılmıştır. Attig ek olarak, yaş, cinsiyet ve eğitim seviyesini dikkate alarak skorları bir kelime testi üzerine kaydetmiştir. Bu veri kümesinin tamamı EK-4 de incelenebilir.

Bayes analizi belirtilmiş, ortalama ve sabit tahminlerini gerektirmektedir. AMOS 'da herhangi bir Bayes analizini gerçekleştirmeden önce, AMOS'a ilk olarak ortalamalar ve sabitleri tahmin etmesi gerekliliği anlatılır. Bu da aşağıdaki adımların takibi ile gerçekleşmektedir:

- Menüden View $\Rightarrow$ Analysis Properties seçilir.
- Estimate means and intercepts seçeneği işaretlenir.

Tablo 4.11. AMOS’da analiz özelliklerinin düzenlenebildiği kutucuk



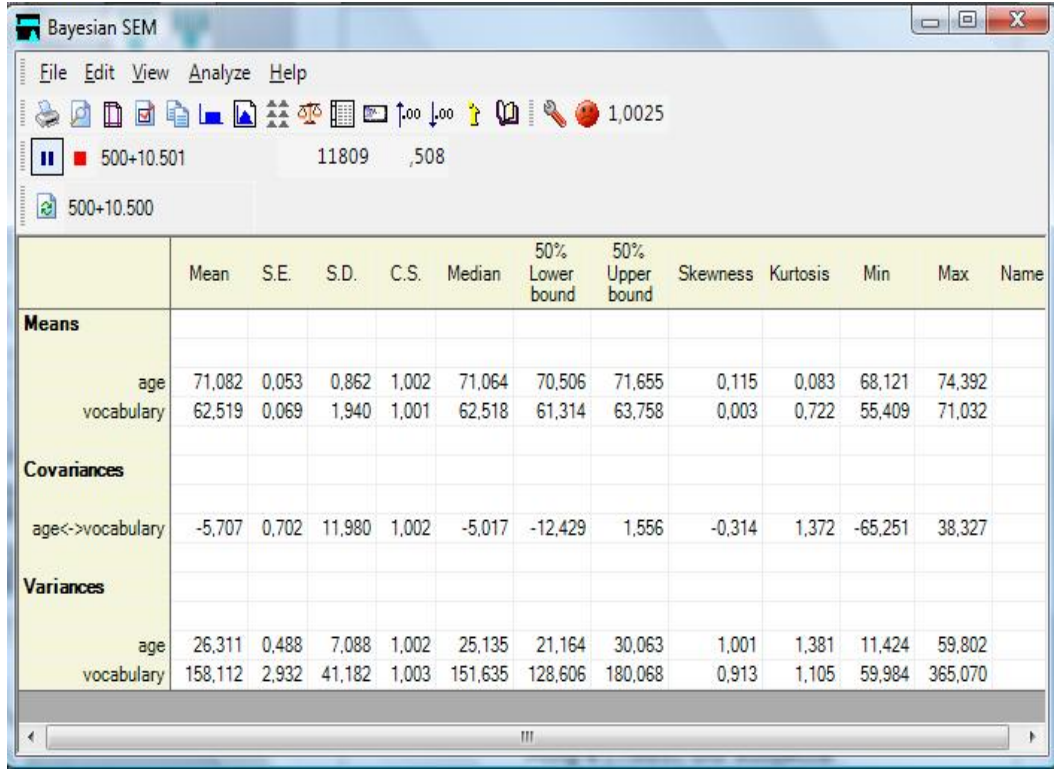
➤ Analizi uygulamak için, menüden Analyze $\Rightarrow$ Bayesian Estimation seçilir veya kısa yol tuşuyla Ctrl+B tuşlanır.

Bu adım sonrası, Bayesian SEM penceresi görünür ve MCMC algoritması hemen örneklemeler oluşturmaya başlar.

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

Tablo 4.12. AMOS'da Bayesian SEM penceresi



	Mean	S.E.	S.D.	C.S.	Median	50% Lower bound	50% Upper bound	Skewness	Kurtosis	Min	Max	Name
Means												
age	71,082	0,053	0,862	1,002	71,064	70,506	71,655	0,115	0,083	68,121	74,392	
vocabulary	62,519	0,069	1,940	1,001	62,518	61,314	63,758	0,003	0,722	55,409	71,032	
Covariances												
age<->vocabulary	-5,707	0,702	11,980	1,002	-5,017	-12,429	1,556	-0,314	1,372	-65,251	38,327	
Variances												
age	26,311	0,488	7,088	1,002	25,135	21,164	30,063	1,001	1,381	11,424	59,802	
vocabulary	158,112	2,932	41,182	1,003	151,635	128,606	180,068	0,913	1,105	59,984	365,070	

Bayesian SEM penceresi, pencerenin üst kısmında bir araç çubuğuna, altta ise sonuç özet tablosuna sahiptir.

Özet tablosunun her bir satırı, tek bir model parametresinin marjinal sonsal dağılımını açıklamaktadır. Mean olarak etiketlenen ilk sütun, sonsal dağılımın merkezi veya ortalama olan sonsal ortalamasını içerir. Örneğin burada, “age – vocabulary kovaryansı için sonsal ortalama, -5.707 dir.” şeklinde bir yorumda bulunabiliriz. Bu ilk sütun, veriye ve önsel dağılıma dayalı olarak parametrenin Bayes nokta tahmini şeklinde kullanılabilir. Büyük veri kümeleri ile sonsal ortalama, maksimum benzerlik tahminine yakın olma eğiliminde olacaktır.

AMOS tarafından kullanılan MCMC algoritmasının, yüksek boyutlu ortak sonsal dağılımlar ile parametrelerin rastgele değerlerini düzenlediğinden ve bu düzenlemeyi parametrelerin sonsal dağılımının Monte Carlo simülasyonu ile gerçekleştirdiğinden bahsedilebilir. Örneğin, Mean sütununda sunulan değer, gerçek sonsal ortalama değildir. Fakat, MCMC prosedürleri tarafından üretilmiş rastgele örneklemeler arasında, ortalama ile ifade edilen bir tahmindir. Monte Carlo

örneklemesine dayandırılan sonsal ortalama içerisindeki belirsizliğin ne kadar olduğu ile ilgili en azından genel anlamda bir fikre sahip olmak önemlidir.

S.E olarak adlandırılan ikinci sütun, standart hatanın bir tahminini sunmaktadır. S.E de, Monte- Carlo tahmini sonsal ortalamanın, gerçek sonsal ortalamadan ne kadar uzakta yer alabileceği belirtilmektedir. MCMC prosedürü daha fazla örnek üretmeye devam ettikçe, sonsal ortalamanın tahmini daha kesin hale gelir ve S.E ‘de yavaş yavaş düşer. Dikkat edecek olursak bu S.E, parametrenin bilinmeyen gerçek değerinden sonsal ortalamanın ne kadar uzakta yer aldığı bir tahmini değildir. Yani, bir parametre için %95 aralığının genişliği şeklinde $\pm 2S.E$ değerlerinden birini kullanmak ister.

Sonsal ortalama ve bilinmeyen gerçek parametre arasındaki benzer bir uzaklık S.D olarak etiketlenen üçüncü sütunda sunulmuştur. S.D ‘de ki sayılar ise maksimum benzerlik tahmini içindeki standart hataya benzer. Diğer sütunlarda ise, yakınsama istatistiği (C.S), her parametre için medyan değeri, her parametre için dağılımın en düşük ve en yüksek % 50 sınırları ve her parametre için çarpıklık, basıklık, minimum ve maksimum değerleri yer almaktadır. En düşük ve en yüksek % 50 sınırları, Bayesin %50 güven aralığına benzer olan % 50 Bayes güven kümesinin bitim noktalarıdır. Çoğumuz % 95 güven düzeyini kullanmaya alışkınızdır. Bu yüzden, güven düzeyinin %95’ e nasıl değiştirileceği ilerde paylaşılacaktır.

Analyze $\implies$ Bayesian Estimation’ ı seçtiğimizde, MCMC algoritması hemen örneklemeye başlar ve süreci durdurmak için Pause Sampling tıklanana kadar devam eder. Tablo 4.11 de 500+10501=11001 tamamlanmış örneklemin ardından, örnekleme durdurulmuştur. AMOS, analiz için muhafaza edilmiş ilk örnekleme düzenlemeden önce belleğe yazılmış 500 örnekleme üretilip, ayırmıştır. AMOS, belleğe yazılmış örneklemleri gerçek ortak sonsal dağılıma yakınsamak için, MCMC prosedürüne izin vererek devam ettirir. AMOS belleğe yazılmış örneklemleri düzenleyip ayırdıktan sonra, ortak sonsal dağılımın neye benzeyeceğini açık bir şekilde verebilmek için ek örneklemler düzenler. Tablo 4.12 de gösterilmiş olan örnekte, AMOS bu analiz örneklemlerinin 10.501’ i ile devam etmektedir. Aslında görüntülenen sonuçlar, 500 belleğe yazılmış olan ve 10.500 analiz örneklemleri içindir. Çünkü AMOS ‘da örnekleme algoritmasının kullanımı çok hızlıdır. Her

örneklemden sonra, Bayesian SEM penceresinde özet tablosu hızla sonuçlanırken, aynı zamanda bu durum analizi yavaşlatacaktır. Bu tip problemlerden uzak durmak için, her 1.000 örneklemden sonra AMOS sonuçları yeniler.

4.5.3. Yakınsama Kavramı ve Değerlendirilmesi

MCMC algoritmasının yakınsaması, maksimum benzerlik gibi rastgele olmayan bir metodun yakınsamasından oldukça farklıdır. MCMC yakınsamasını tam olarak anlayabilmek için, iki farklı tipi ayırt edebilmemiz gerekir:

Dağılım içerisindeki yakınsama olarak adlandırabileceğimiz ilk tipdeki yakınsamanın anlamı, analiz örneklemlerinin gerçekte, parametrelerin gerçek ortak sonsal dağılımlarından çekilir olmasıdır. Algoritma ilk başlangıç değerlerini yavaş yavaş unuttuğu esnada, dağılım içerisindeki yakınsama yazılmış period içerisinde yer alır. Çünkü, bu örneklem gerçekte sonsal dağılımın temsilcisi olmayabilir, iptal edilmişlerdir. Varsayılan yazılmış 500 'lük period oldukça klasiktir ve pek çok problem için gereğinden uzundur. Yazılmış period aşıldığı zaman ve AMOS, analiz örneklemelerini toplamaya başladığında, sonsal ortalama gibi özet istatistiklerinin tahmininin doğru olması için bu örneklemelerin yeterli olup olmadığı sorusu sorulabilir.

Bu soru, sonsal özetlerin yakınsaması olarak adlandırılan ikinci tipteki yakınsamayla alakalıdır. Sonsal özetlerin yakınsaması, analiz örneklemlerinin bağımsız olmadığı fakat aslında otokorelasyonlu zaman serileri oldukları gerçeği ile karmaşık hale gelmiştir. 1001 inci örneklem 1000 inci örneklemle, o da 999 uncu ile ilişkilidir, vb. Bu korelasyonlar MCMC'nin kendi yapısından kaynaklıdır ve bu korelasyonlardan dolayı, 5.500 (yada sayı ne olursa olsun) analiz örnekleminden elde edilen özet istatistikleri, bu 5.500 örneklemin bağımsız olması durumunda elde edilebileceklerden daha değişkendir. Bütün bunlara rağmen, gittikçe daha çok analiz örneklemeleri biriktirmek, sonsal özetleri yavaş yavaş stabilize hale getirecektir (Arbuckle, 2007).

AMOS, yakınsamayı kontrol edebileceğiniz çeşitli tanılayıcılar sağlar (Arbuckle, 2007). Tablo 4.11'de Bayesian SEM penceresinin araç çubuğunda yer

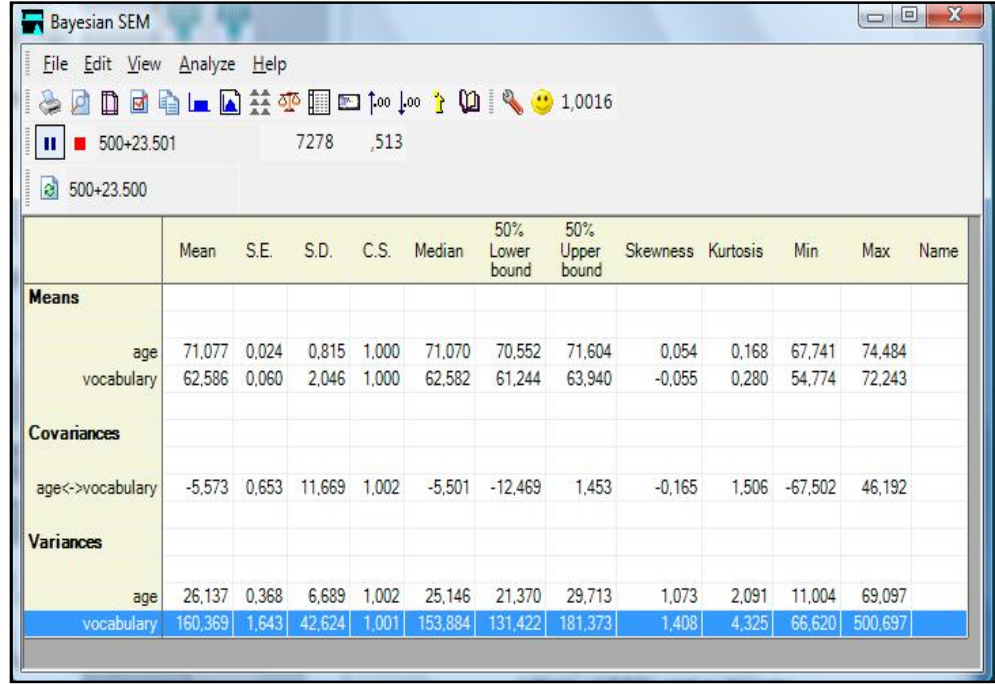
alan 1.0025 değerine dikkat edilmesi gerekir. Bu, Gelman, Carlin, Stern ve Rubin (2004) tarafından ileri sürülen, ölçüme dayalı genel bir yakınsama istatistiğidir. Sürekli, ekran yenilenir ve özet tablosu içerisinde her bir parametre için AMOS C.S ‘ yi günceller; araç çubuğu üzerindeki C.S değeri, farklı C.S değerlerinin en büyüğüdür. Eğer, C.S değerlerinin en büyüğü 1.002 değerinden küçük ise AMOS yakınsamaya sahip olmak için prosedürü değerlendirir. Bu standartlarda, 1.0025 olan maksimum C.S değeri yeterince küçük değildir. AMOS tüm C.S değerleri yeterince küçük olmadığı zaman, üzgün bir yüz ifadesi,  gösterir. Yakınsama istatistiği, analiz örnekleminin kısımları içindeki değişkenliği, bu kısımların değişkenliği ile karşılaştırır. 1.000 değeri mükemmel yakınsamayı temsil eder. Büyük değerler sonsal özetlerin daha fazla analiz örneklemi oluşturularak daha titiz bir şekilde yapılması gerektiğinin işaretidir. Pause Sampling butonuna ikinci kez tıklamak, AMOS’ a örnekleme sürecini kaldığı yerden devam ettirmesi için bir talimattır. Ayrıca, Analyze menüsünden Pause Sampling’ i seçerek veya Ctrl+E kullanılarak da örnekleme durdurmak ve yeniden başlatmak için talimat verilebilir.

Tablo 4.13 örnekleminin bir süreliğine yeniden başlatılmasının ve durdurulmasının ardından sonuçları göstermektedir. Bu noktada ekran, en son 23.500 üncü örnekleme güncellenmişken 23.501 analiz örnekleme sahibizdir. En büyük C.S değeri 1,0016 dır. Bu değer, 1.002 kriterinden aşağıda bir değer olduğu için uygun bir yakınsama söz konusudur. Yeterli bir yakınsama belirtildiği için, AMOS’ da mutlu bir yüz ifadesi  görülmektedir. Gelman ve ark. (2004), 1.10 veya daha küçük değerlerin pek çok analiz için uygun olduğunu ileri sürmüşlerdir. Fakat, varsayılan kriter olan 1.002 değeri daha klasiktir. Bu kriter tarafından MCMC sınırlama kararı, özet tablosunun değişimi sonladıracağı anlamına gelmemektedir. MCMC algoritması çalışmayı sürdürdüğü sürece, özet tablosu değişmeye devam edecektir. Araç çubuğu üzerinde tüm C.S değerleri 1.000‘ a yaklaşırken, ek örneklem almak daha fazla kesinlik kazanacağı anlamına gelmemektedir. Bu yüzden örnekleme durmamız uygun olacaktır.

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

Tablo 4.13. AMOS'da Bayesian SEM penceresinin son hali



	Mean	S.E.	S.D.	C.S.	Median	50% Lower bound	50% Upper bound	Skewness	Kurtosis	Min	Max	Name
Means												
age	71,077	0,024	0,815	1,000	71,070	70,552	71,604	0,054	0,168	67,741	74,484	
vocabulary	62,586	0,060	2,046	1,000	62,582	61,244	63,940	-0,055	0,280	54,774	72,243	
Covariances												
age<->vocabulary	-5,573	0,653	11,669	1,002	-5,501	-12,469	1,453	-0,165	1,506	-67,502	46,192	
Variances												
age	26,137	0,368	6,689	1,002	25,146	21,370	29,713	1,073	2,091	11,004	69,097	
vocabulary	160,369	1,643	42,624	1,001	153,884	131,422	181,373	1,408	4,325	66,620	500,697	

4.5.4. Kontrol Grafikleri

C.S değerlerine ek olarak AMOS, çeşitli grafikleri önermektedir. Bu grafikler, Bayes MCMC metodunun yakınsamasını kontrol etmede yardımcı olacaktır (Arbuckle, L. J.,2007).

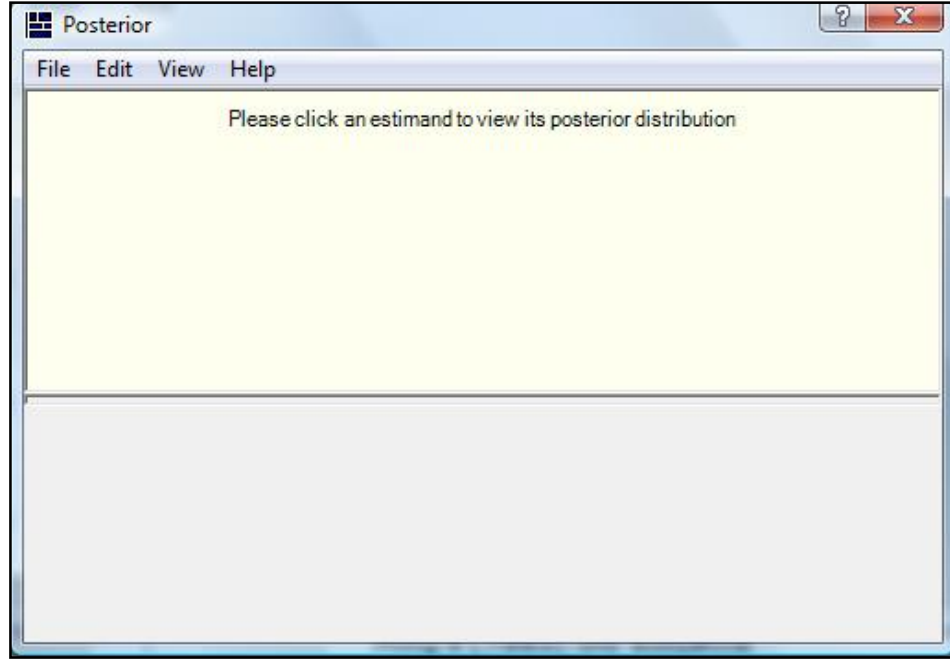
Grafikleri görüntülemek için:

- Menüden View $\Rightarrow$ Posterior seçilir ve AMOS, Posterior diyalog kutusunu görüntüler.

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

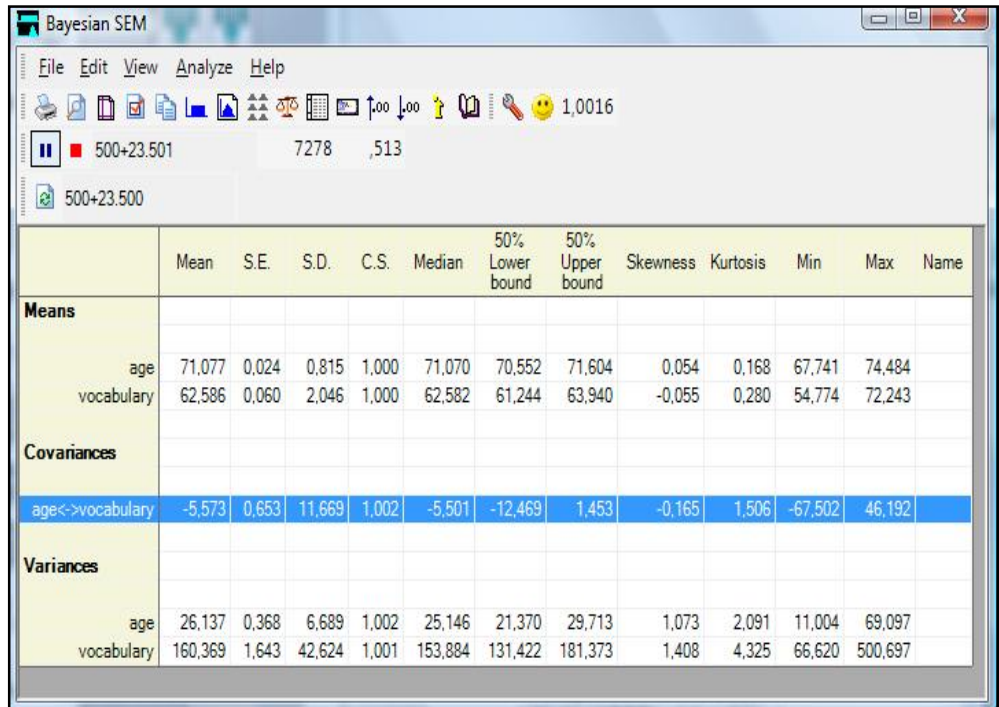
ASLI BAHADIR

Tablo 4.14. AMOS'da Posterior penceresi



➤ Tablo 4.15 'de görüldüğü üzere, Bayesian SEM penceresinden age <-> vocabulary parametresi seçilir.

Tablo 4.15. Age-Vocabulary parametresinin seçildiği Bayes SEM penceresi

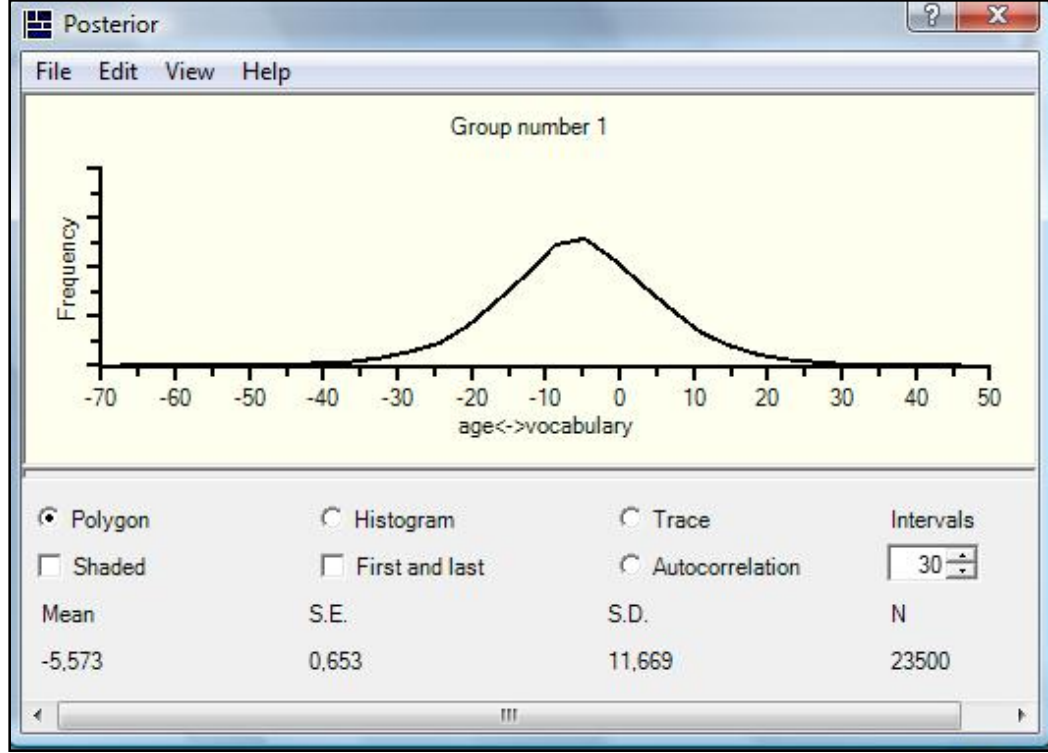


	Mean	S.E.	S.D.	C.S.	Median	50% Lower bound	50% Upper bound	Skewness	Kurtosis	Min	Max	Name
Means												
age	71,077	0,024	0,815	1,000	71,070	70,552	71,604	0,054	0,168	67,741	74,484	
vocabulary	62,586	0,060	2,046	1,000	62,582	61,244	63,940	-0,055	0,280	54,774	72,243	
Covariances												
age<->vocabulary	-5,573	0,653	11,669	1,002	-5,501	-12,469	1,453	-0,165	1,506	-67,502	46,192	
Variances												
age	26,137	0,368	6,689	1,002	25,146	21,370	29,713	1,073	2,091	11,004	69,097	
vocabulary	160,369	1,643	42,624	1,001	153,884	131,422	181,373	1,408	4,325	66,620	500,697	

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

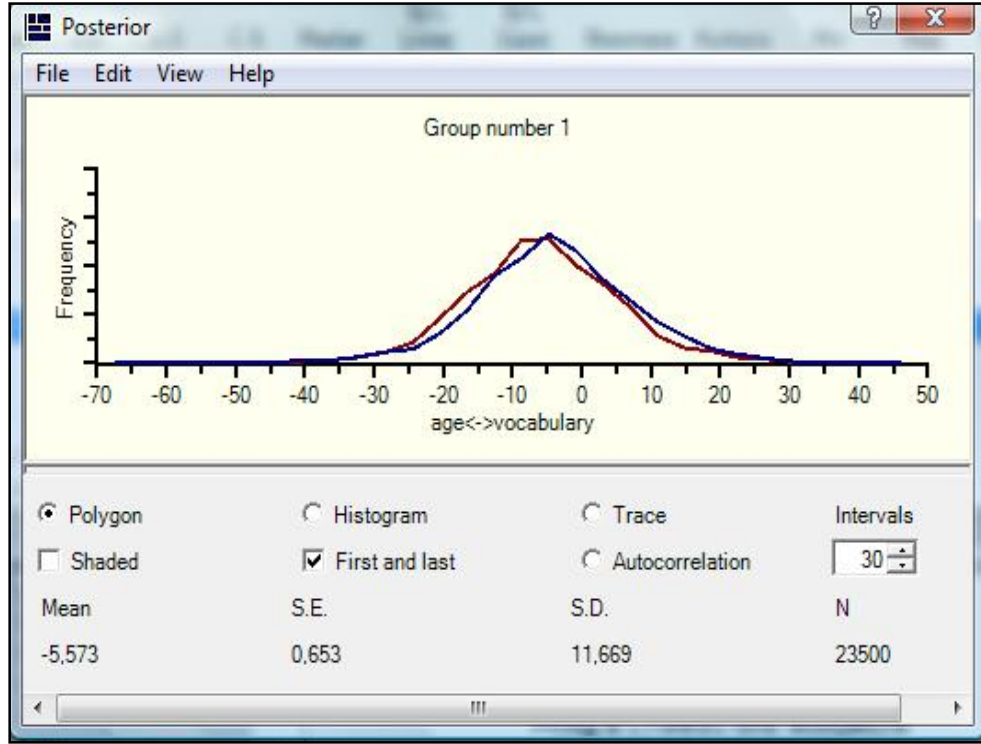
ASLI BAHADIR

Şekil 4.7’de gösterilen posterior diyalog kutusu, 23.500 örneklem üzerindeki age-vocabulary kovaryans dağılımının frekans poligonunu göstermektedir.



Şekil 4.7. AMOS’da frekans poligonu

Dağılımın iki tahminin eş zamanlı göstergeleri, AMOS sonsal dağılıma yakınsarken karar vermede kullanabilecek görsel yardımcılardır. Bunlardan biri birikimli örneklemelerden ilk üçünden elde edilen, diğeryse son üçünden elde edilendir. Aynı grafik üzerinde marjinal sonsalın iki tahminini göstermek için Posterior penceresinde “First and Last” seçilir.



Şekil 4.8. AMOS’da poligon üzerinde iki tahminin eş zamanlı göstergesi

Şekil 4.8’den anlaşılmaktadır ki, analiz örneklemelerinin ilk üç ve son üç dağılımları hemen hemen aynıdır ve burada AMOS’un başarılı bir şekilde, age-vocabulary kovaryansında sonsal dağılımın önemli özelliklerini belirttiği ileri sürülmektedir (Arbuckle, 2007).

Dikkat edecek olursak, sonsal dağılımın yaklaşık değer olarak -5 merkezli olduğu görülmektedir. Bu durum, bu parametre için ortalama değeri ile uyumludur. Görsel inceleme göstermektedir ki, standart sapma yaklaşık olarak 11 olan S.D değeri ile kabul görmektedir.

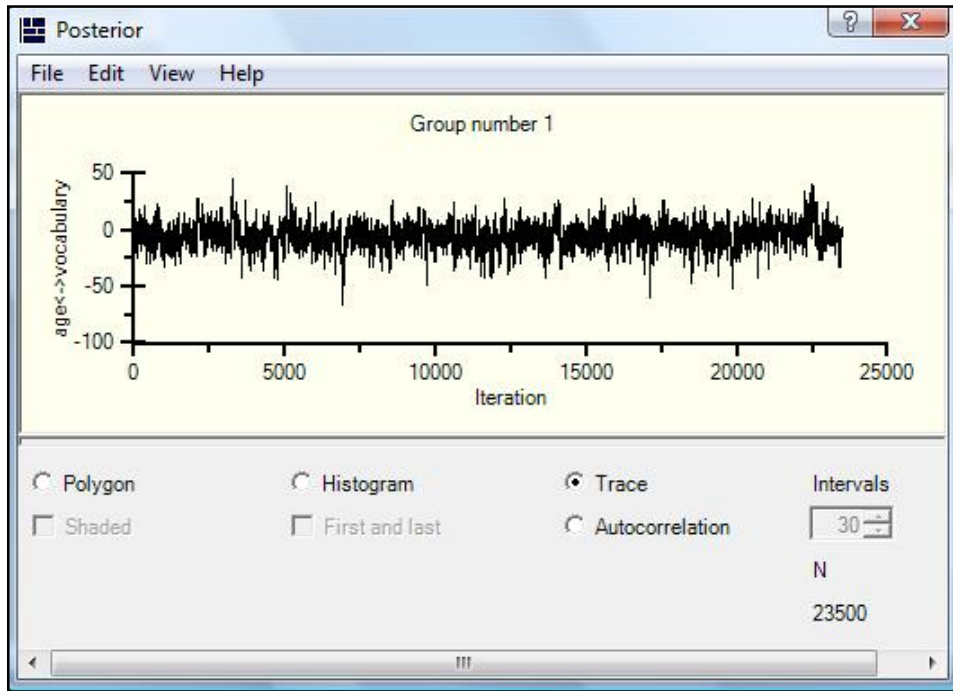
Ayrıca dikkat edecek olursak, örnek değerlerinin yarıdan fazlası sıfır değerinin sol tarafında kalmaktadır. Bu kovaryans parametresinin gerçek değerinin negatif olduğunun bir kanıtı olabilir fakat bu sonuç istatistiksel olarak öneme sahip değildir. Çünkü 0’ın sağ tarafında kalan oran hala oldukça büyüktür. Eğer 0’ın sağ tarafında kalan örnek değerlerin oranı oldukça küçük ise, örneğin; %5’den küçük ise bu durumda, kovaryans parametresinin sıfıra eşit veya küçük olması ve sıfırdan büyük olma olarak ifade edilebilen hipotezlerinden, sıfır hipotezi reddedilebilir.

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

Yakınsamayı değerlendirirken kullanabilecek bir diğer grafik ise, çizgi grafiğidir (Trace Plot). Bazen zaman-serisi grafiği olarak da adlandırılan çizgi grafiği, bir parametrenin zaman içerisindeki örnek değerlerini gösterir. Bu grafik, dağılım içerisindeki MCMC prosedürün ne kadar çabuk yakınsadığına karar vermede yardımcı olur (Arbuckle, 2007).

- Çizgi grafiğini görüntülemek için Trace seçilir.



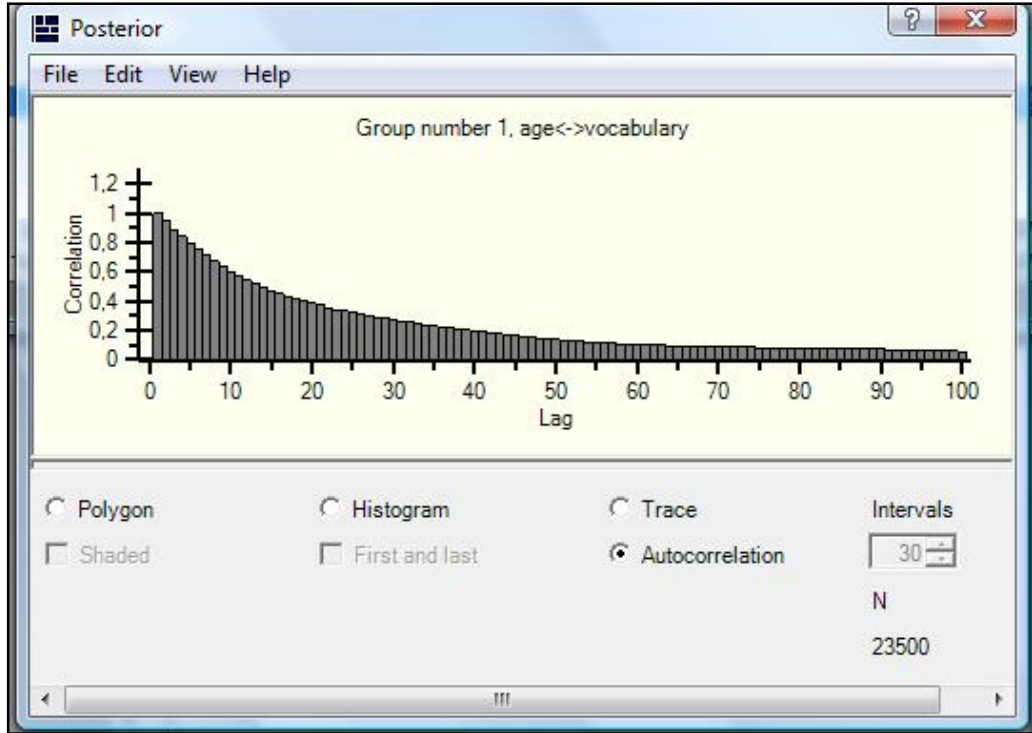
Şekil 4.9. AMOS’da çizgi grafiği

Burada gösterilen grafik oldukça idealdir. Uzun süreli olmayan hızlı çıkış ve iniş değişimlerini gösterir. Bu grafiği birkaç yatay kesite bölecek olursak, herhangi bir kesit içindeki çizgi, diğer kesitlerdeki çizgilerden çok da farklı olmayacaktır. Bu da dağılım içindeki yakınsamanın hızla gerçekleştiğinin göstergesidir (Arbuckle, 2007). Grafik içerisindeki uzun süreli inişler ve çıkışlar yavaş yakınsamaya işaret eder. (Uzun süreli eğilimler, örneklemelerin sayısına dayalı olan bu grafiğin yatay eksenine ile ilişkilidir. Daha fazla örneklem aldıkça, çizgi grafiği bir akordeon gibi sıkışır ve yavaş çıkışlar ve inişler neticede hızlı çıkış ve iniş değişimleri şeklinde

görünmeye başlar.) Hızlı çıkış ve iniş hareketleri, herhangi bir iterasyondaki örneklenmiş değerin, örneklemelerin toplam sayısı ile az ilişkili olan k değerleri için, k iterasyondan sonra ki örneklem değeri ile ilişkisizliğini anlatmaktadır.

Örneklemeler arasındaki korelasyonun ne kadar sürede bittiğini görebilmek için, otokorelasyon grafiği (autocorrelation plot) incelenebilir. Otokorelasyon grafiği, herhangi bir iterasyonda örneklenmiş değer ile $k=1,2,3,\dots$ için k iterasyondan sonraki örneklenmiş değer arasındaki tahmin edilmiş korelasyonu gösterir.

- Bu grafiği göstermek için Autocorrelation seçilir.



Şekil 4.10. AMOS’da otokorelasyon grafiği

4.5.5. Güven Aralığı

Bayesian SEM penceresindeki özet tablosunu tekrar hatırlayacak olursak, her bir tahmin için Bayesin güven aralığının en düşük ve en yüksek bitiş noktalarını görebiliyorduk. Klasik % 50 güven aralığı ile aynı olacak şekilde AMOS, Bayes

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

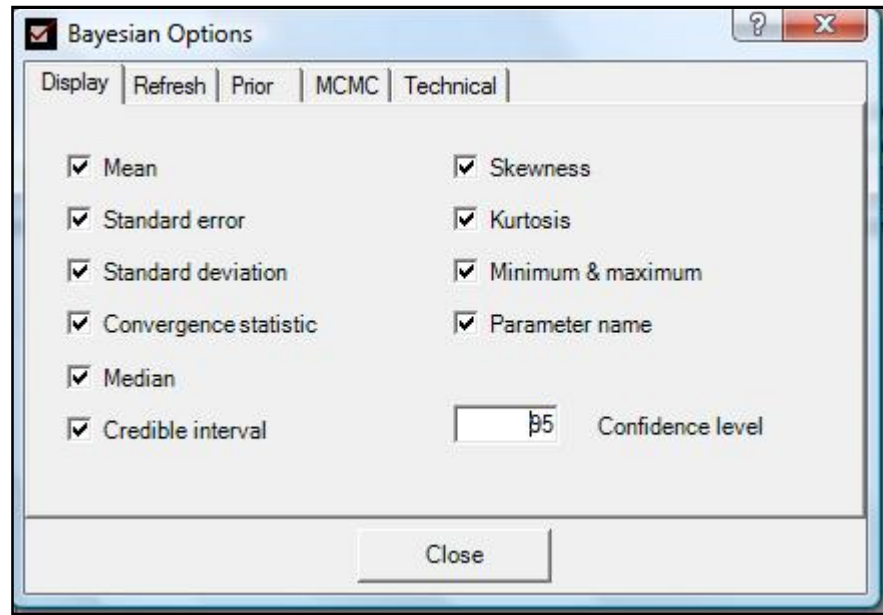
Analizi konulu başlık altındaki özet tablolarında, varsayım ile %50 aralığı göstermektedir.

Araştırmacılar, genellikle %95 güven aralıklarını raporlarlar bu sebeple; %95 içindeki sonsal olasılığa uygun olması için sınırlar değiştirilmek istenebilir.

4.5.5.1. Güven Düzeyinin Değiştirilmesi

- Options diyalog kutusunda Display sekmesi tıklanır.
- Confidence level (güven düzeyi) değerine 95 girilir.

Tablo 4.16. AMOS’ın özet tablosunda görmek istenen seçeneklerin belirlenmesi



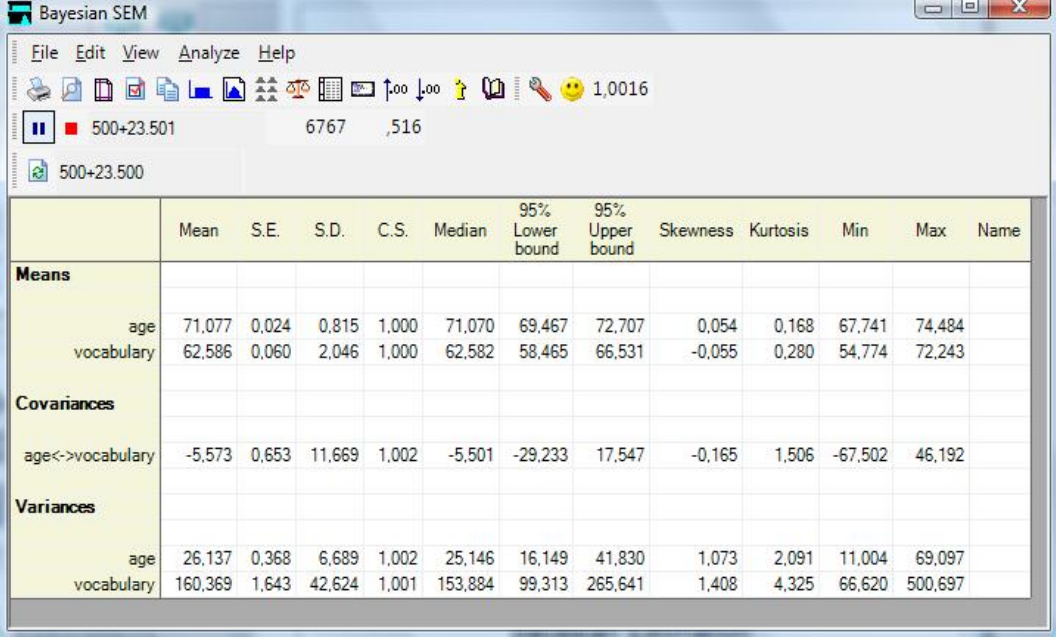
- Close butonuna basılır.

Böylece özet tablosunda sınırlar, Tablo 4.17’de görülebileceği üzere % 95 güven ile gösterilebilecektir.

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

Tablo 4.17. AMOS’da % 95 güven aralıklı özet tablosu



	Mean	S.E.	S.D.	C.S.	Median	95% Lower bound	95% Upper bound	Skewness	Kurtosis	Min	Max	Name
Means												
age	71,077	0,024	0,815	1,000	71,070	69,467	72,707	0,054	0,168	67,741	74,484	
vocabulary	62,586	0,060	2,046	1,000	62,582	58,465	66,531	-0,055	0,280	54,774	72,243	
Covariances												
age<->vocabulary	-5,573	0,653	11,669	1,002	-5,501	-29,233	17,547	-0,165	1,506	-67,502	46,192	
Variances												
age	26,137	0,368	6,689	1,002	25,146	16,149	41,830	1,073	2,091	11,004	69,097	
vocabulary	160,369	1,643	42,624	1,001	153,884	99,313	265,641	1,408	4,325	66,620	500,697	

4.6. Karma Regresyon Analizi

Karma Regresyon analizini gerçekleştirmek için başvuru olan Bayes tahmin yönteminin önemi ve bu yöntemin AMOS ile uygulanma şekli, 4.5 nolu alt bölümlerde detaylı bir şekilde incelenmiştir. Hatırlayacak olursak, analize başlamadan önce AMOS ‘da gerçekleştirilmesi gereken adımlar; grup sayısını belirleme, her bir grup için veri dosyasını belirleme ve model belirleme şeklinde sıralanmaktaydı. Bu adımlar ise, Dosage And Performance 2 (Dozaj ve Performans 2) adlı bir veri kümesi üzerinde gerçekleştirilmekteydi.

Bütün bu hatırlatmaların ardından, AMOS’da kaldığımız yerden devam edecek olursak analiz, AMOS’a ilk olarak ortalamalar ve sabitleri tahmin etmesi anlatılarak başlayacaktır.

- Menüden, View $\longrightarrow$ Analysis Properties seçilir.
- Estimate means and intercepts işaretlenir.

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

Bu adımlar ile karşımıza çıkacak kutucuk, daha önce Bayes tahmini anlatılırken gösterdiğimiz Tablo 4.11 'in aynısıdır.

- Araç çubuğu üzerindeki  kliklenir.
- Veya
- Menüden, Analyze $\Rightarrow$ Bayesian Estimation seçilir.

Menüden Bayesian Estimation' ı seçtiğimizde, MCMC algoritması hemen örneklemeye başlar ve süreci durdurmak için Pause Sampling tıklanana kadar devam eder.

Karma regresyon modelinin temelinde yatan karma modellemede sadece Bayesci tahmin kullanılabilir olmasından dolayı, AMOS ara yüzünde calculate estimates  butonu (kısayol tuşu) kullanılır durumda değildir.

Tablo 4.18 'deki Bayesian SEM penceresi açıldıktan sonra,  ifadesi  ifadesine dönene kadar beklenir. Ya da başka bir deyişle, yakınsama istatistik değerlerinin en büyüğü 1.002 değerinden küçük bir değer alana kadar beklenir. Bayesian SEM penceresi içindeki tahminlerin tablosu aşağıdakine benzer bir tablo olmalıdır.

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

Tablo 4.18. AMOS’da analiz sonucu oluşan Bayesian SEM penceresi

	Mean	S.E.	S.D.	C.S.	Median	95% Lower bound	95% Upper bound	Skewness	Kurtosis	Min	Max	Name
Regression weights												
performance<-dosage	2.083	0.000	0.070	1.000	2.083	1.945	2.217	-0.053	0.111	1.757	2.361	
Means												
dosage	7.114	0.001	0.180	1.000	7.116	6.759	7.461	-0.038	0.048	6.363	7.853	
Intercepts												
performance	5.394	0.004	0.521	1.000	5.387	4.393	6.420	0.065	0.119	3.153	7.948	
Variances												
dosage	2.896	0.005	0.458	1.000	2.890	2.128	3.929	0.632	0.739	1.557	5.519	
E1	1.027	0.002	0.174	1.000	1.007	0.738	1.418	0.707	1.040	0.478	2.090	
Group number 1 Proportion												
Proportion	0.247	0.000	0.022	1.000	0.246	0.205	0.291	0.117	0.029	0.162	0.338	

Bayesian SEM penceresi, olağan bir çoklu grup regresyon analizinde sahip olabileceğimiz bütün parametre tahminlerini özet tablosunda göstermektedir. Özet tablosunun her bir satırı, tek bir model parametresinin marjinal sonsal dağılımını açıklamaktadır. Mean olarak etiketlenen ilk sütun, sonsal dağılımın merkezi veya ortalama olan sonsal ortalamasını içerir. Örneğin, performans sabiti için sonsal ortalama 5.394 dür. Her bir grup için tahminlerin ayrı bir tablosu vardır. Bunun için sekmelerden birini tıklamamız yeterli olacaktır.

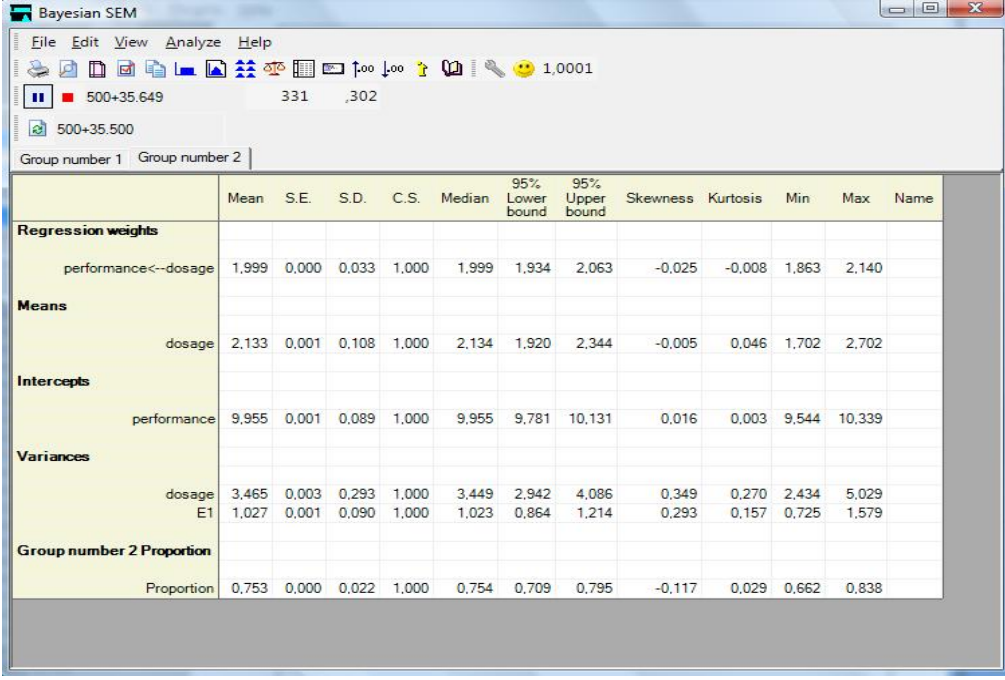
Ekran, 35.500 üncü örnekleme güncellenmişken, 35.694 analiz örnekleme sahip olunmaktadır. Tabloda Proportion olarak adlandırılan satırda, kitle içerisinde grubun hangi oranda tahmin edildiği gösterilmektedir. 1 numaralı grup için tahminleri gösteren yukarıdaki şekil, kitle içerisinde 1.grubun (Group Number 1) 0.247 oranında tahmin edildiğini göstermektedir. Diğer bir ifadeyle, 1.grubun karma oranı $P_1=0.247$ dir. (AMOS’un bizlere özet tablosunda hazır olarak sunduğu, bu karma oranın, matematiksel olarak hesaplanma şekli EK-6’da paylaşılmıştır.)

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

Analiz sonrası özet tablosu üzerinde yer alan 2.gruba da, aynı Bayesian SEM penceresinden “Group Number 2” sekmesi seçilerek ulaşılp, benzer mantıkla grup ile ilgili yorumlarda bulunabilir.

Tablo 4.19. AMOS’da 2.grup için özet tablosu



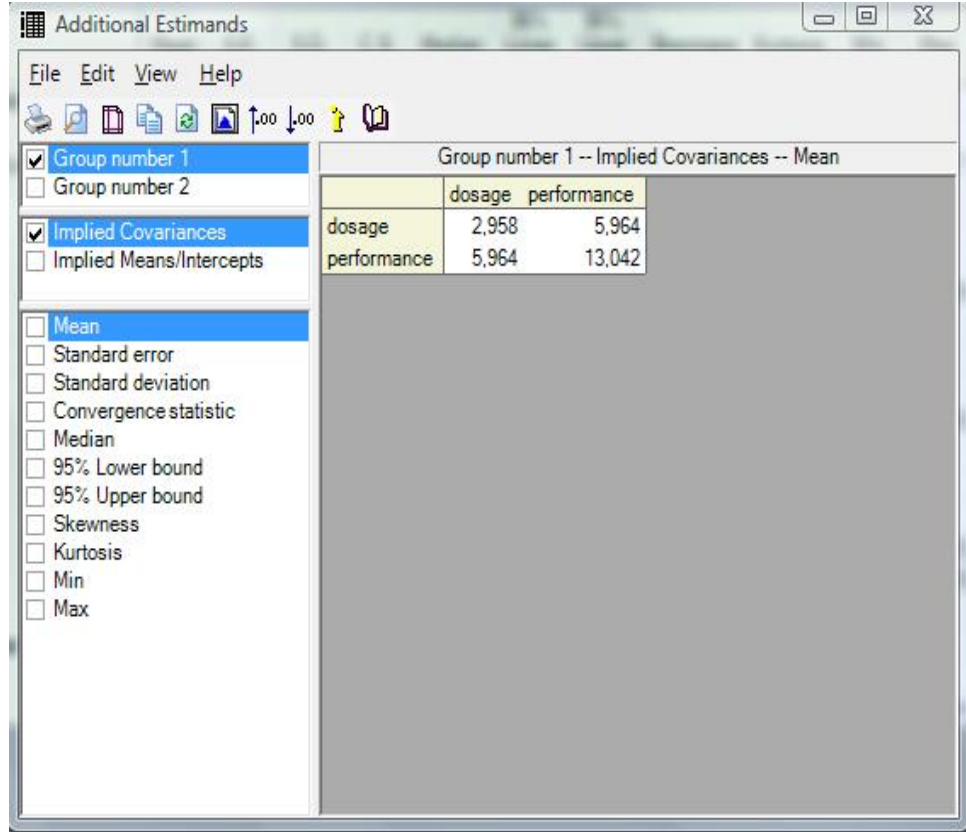
	Mean	S.E.	S.D.	C.S.	Median	95% Lower bound	95% Upper bound	Skewness	Kurtosis	Min	Max	Name
Regression weights												
performance<--dosage	1,999	0,000	0,033	1,000	1,999	1,934	2,063	-0,025	-0,008	1,863	2,140	
Means												
dosage	2,133	0,001	0,108	1,000	2,134	1,920	2,344	-0,005	0,046	1,702	2,702	
Intercepts												
performance	9,955	0,001	0,089	1,000	9,955	9,781	10,131	0,016	0,003	9,544	10,339	
Variances												
dosage	3,465	0,003	0,293	1,000	3,449	2,942	4,086	0,349	0,270	2,434	5,029	
E1	1,027	0,001	0,090	1,000	1,023	0,864	1,214	0,293	0,157	0,725	1,579	
Group number 2 Proportion												
Proportion	0,753	0,000	0,022	1,000	0,754	0,709	0,795	-0,117	0,029	0,662	0,838	

Bayesian SEM penceresinde gösterilen parametre tahminlerinden başka, bu pencerenin araç çubuğunda yer alan Ek Tahminler (Additional Estimands) butonunun tıklanması ile çeşitli niceliklerin tahminlerine de ulaşılabilir. Örneğin, Tablo 4.20’de 1.grup içindeki bütün kovaryanslar için sonsal ortalamalar gösterilmektedir.

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

Tablo 4.20. AMOS ‘da Ek Tahminler penceresi



Group number 1 -- Implied Covariances -- Mean		
	dosage	performance
dosage	2,958	5,964
performance	5,964	13,042

Analiz sonrası, özet tablosundan faydalanarak gruplar için ayrı ayrı sonsal dağılımlı ortalama bir model tahmini hesaplanabilir.

Özet tablosunda (Tablo 4.21) işaretlenen parametre tahminlerinin yardımıyla, 1.grup için sonsal dağılımlı ortalama bir model tahmininin hesaplaması aşağıdaki şekilde gerçekleştirilebilir:

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

Tablo 4.21. AMOS yardımı ile 1.grup için ortalama bir model tahmini hesaplama

	Mean	S.E.	S.D.	C.S.	Median	95% Lower bound	95% Upper bound	Skewness	Kurtosis	Min	Max	Name
Regression weights												
performance<--dosage	2.083	0.000	0.070	1.000	2.083	1.945	2.217	-0.053	0.111	1.767	2.361	
Means												
dosage	7.114	0.001	0.180	1.000	7.116	6.759	7.461	-0.038	0.048	6.363	7.853	
Intercepts												
performance	5.394	0.004	0.521	1.000	5.387	4.393	6.420	0.065	0.119	3.153	7.948	
Variances												
dosage	2.896	0.005	0.458	1.000	2.850	2.128	3.929	0.632	0.739	1.557	5.519	
E1	1.027	0.002	0.174	1.000	1.007	0.738	1.418	0.707	1.040	0.478	2.090	
Group number 1 Proportion												
Proportion	0.247	0.000	0.000	0.000	0.000	0.000	0.000	0.291	0.117	0.029	0.162	0.338

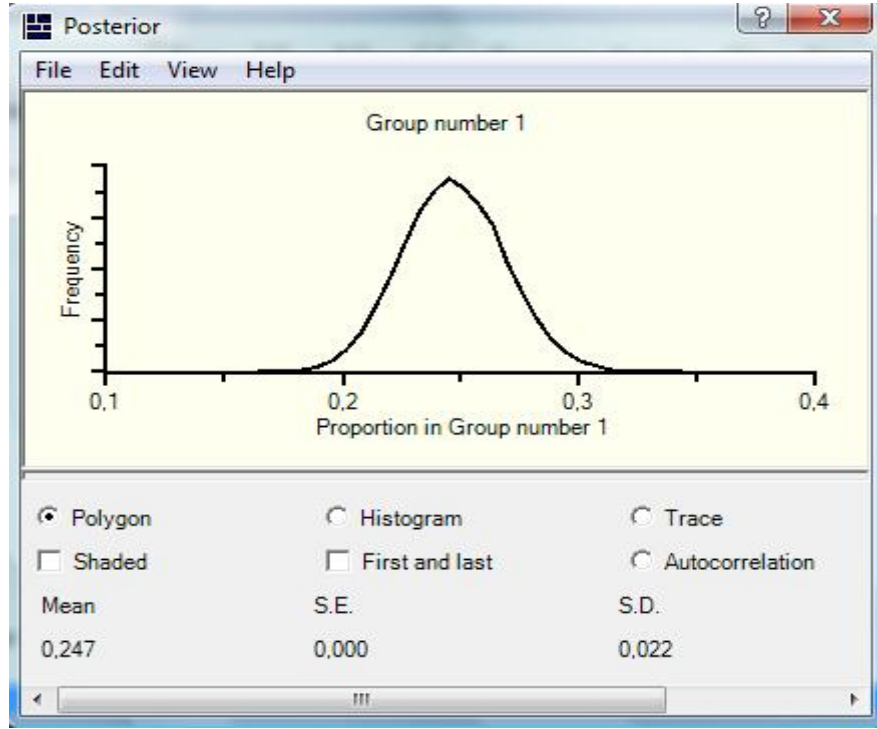
$$\begin{aligned}
 \hat{y}_1 &= \hat{\beta}_{01} + \hat{\beta}_{11}\bar{x} \\
 &= 5.394 + 2.083(7.114) \\
 \hat{y}_1 &= 20.212
 \end{aligned}
 \tag{4.4}$$

Bu eşitlik; 1.gruba ait ortalama dozaj değerine karşılık, tahmini ortalama performans sonucunu vermektedir.

4.6.1. Dağılımların İncelenmesi

Analiz sonucunda elde edilen parametre tahminlerinin ve kitle içinde her bir grubun tahmin edilme oranlarının dağılımı, AMOS ile incelenebilir.

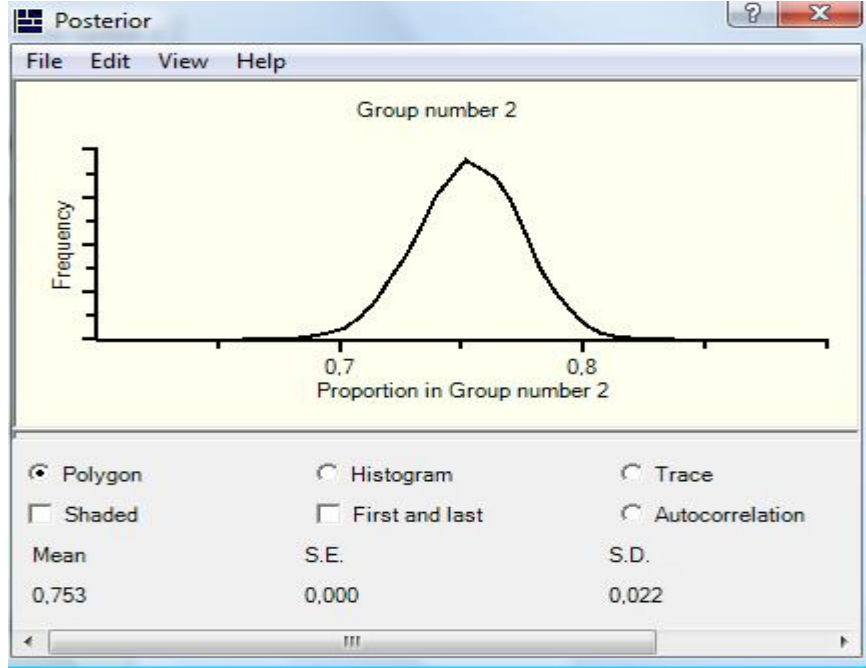
Örneğin, 1.grubun kitle oranının tahmin edilmiş sonsal dağılımını görebilmek için, Tablo 4.19'un özet tablosunda yer alan oran satırı (proportion) üzerinde sağ tıklanır ve açılan menüde Show Posterior seçilir.



Şekil 4.11. AMOS'da 1.grup için sonsal dağılım

Posterior penceresinde gördüğümüz frekans poligonunda, kitle içerisinde 1.grubun (1 numaralı grubun) tahmin edilme oranının pratikte 0.15 ile 0.35 arasında bir yerde olacağı görülebilmektedir.

Benzer adımlar ile 2.grup için de sonsal dağılım incelenebilir. Şekil 4.11'deki poligonda 2.grubun (2 numaralı grubun) tahmin edilme oranının pratikte 0.65 ile 0.85 arasında bir yerde olacağı yorumu yapılabilir.



Şekil 4.12. AMOS’da 2.grup için sonsal dağılım

4.6.2. Grupların Karşılaştırılması

Her iki grup için de özet tabloları (Tablo 4.18-4.19) incelenip, karşılaştırıldığında göze çarpan farklılıklardan bahsedilebilir.

Örneğin, 1 numaralı grup içindeki tahmini regresyon ağırlığı ve sabiti, 2 numaralı grup içindeki tahminler ile karşılaştıralım: 1.grupta regresyon ağırlığının tahmini 2.083 iken, sabitin tahmini 5.394 dür. 2.grupta regresyon ağırlığının tahmini (1.999), 1.grup içindeki ile yaklaşık olarak aynıdır. Oysa, 2.gruptaki sabit tahmini (9.955), 1.gruptakinden oldukça yüksektir.

Her iki grup için de göze çarpan bu tahminlerin karşılaştırılmalarının ardından anlaşılıyor ki; 2 numaralı grup yüksek, 1 numaralı grup ise düşük performanslı bir gruptur.

4.6.3. Farklı Durumların Sınıflandırılması

Her bir durum için, grup üyelik olasılıkları elde edilebilir:

- Posterior Predictive  butonu tıklanır.

Veya

- Menüden View  Posterior Predictive seçilir.

Tablo 4.22. AMOS’da herbir gözlem için grup üyelik olasılıkları

	dosage	performance	P(Cluster1)	P(Cluster2)
1	6,66331007602931	21,2031284997218	0,88	0,12
2	5,66102565586862	15,6987674948943	1,00	0,00
3	6,05921131399695	19,2048361186385	0,98	0,02
4	9,18741643404798	23,1294140957387	1,00	0,00
5	6,94140644349163	20,9890517035103	1,00	0,00
6	5,16054293700744	18,0430561625521	0,56	0,44
7	4,17650111795259	13,9604306100494	1,00	0,00
8	8,07913550523847	22,5151975204992	1,00	0,00
9	4,67521627594292	14,0616719804701	1,00	0,00
10	7,85787684574796	19,546086381681	1,00	0,00
11	2,7627255307244	12,9679285391251	0,08	0,92
12	3,6995932072631	11,958726583912	1,00	0,00
13	5,53950501114771	15,0037101134325	1,00	0,00
14	7,06075641893955	20,1350290402946	1,00	0,00
15	7,2170160618839	20,3485096288145	1,00	0,00
16	9,8401696468436	25,7369478114549	1,00	0,00
17	4,44159268082361	15,3602709609774	0,98	0,02
18	8,63318871213828	23,677847019763	1,00	0,00
19	8,37492785176125	22,6912370106325	1,00	0,00
20	6,92518967262505	19,4741881742043	1,00	0,00
21	7,63502745118693	22,7196293582141	0,99	0,01
22	5,66055867006606	18,1796428517578	0,99	0,01
23	5,87157585646565	17,2726584025363	1,00	0,00
24	5,3691316167084	15,8140178732247	1,00	0,00
25	7,01435564647898	20,8650988907426	1,00	0,00
26	7,26916462531103	21,5553808958383	1,00	0,00
27	7,63868845163894	22,0462181215458	1,00	0,00
28	8,14525079484013	22,9844514410935	1,00	0,00
29	7,48371460791481	20,2398105563852	1,00	0,00
30	6,83913761484707	19,9633239360852	1,00	0,00
31	5,72284811383529	16,9670943271575	1,00	0,00
32	6,43915130064085	18,9858458316289	1,00	0,00
33	8,20789759063962	22,6851673869118	1,00	0,00

The Posterior Predictive Distributions penceresi her bir durum için, o durumun 1.grup (Cluster 1) veya 2.grup (Cluster 2) da bulunma olasılığını göstermektedir. Yukarıdaki tabloda, 400 gözlemden yalnızca 33'ü gösterilmektedir. Bütün durumlar için grup üyelik olasılıkları EK-5 'de paylaşılmıştır.

1.durumun, 1.grupta bulunma olasılığı 0.88 olarak tahmin edilmiştir. 2.grupta bulunma olasılığı ise 0.12 dir. Hatırlayacak olursak, 1.grup yaklaşık olarak 5.394 sabitine sahip idi. 2. grup ise, yaklaşık olarak 9.954 sabitine sahipti. Bu sebeple, 1.grup (1 numaralı grup) düşük performanslı bir gruptur. Böylece, örneklem içindeki ilk kişi %88 olasılıkla düşük performans grubunda, %12 olasılıkla da yüksek performans grubunda yer almaktadır.

4.7. Parametre Tahminlerini Geliştirme

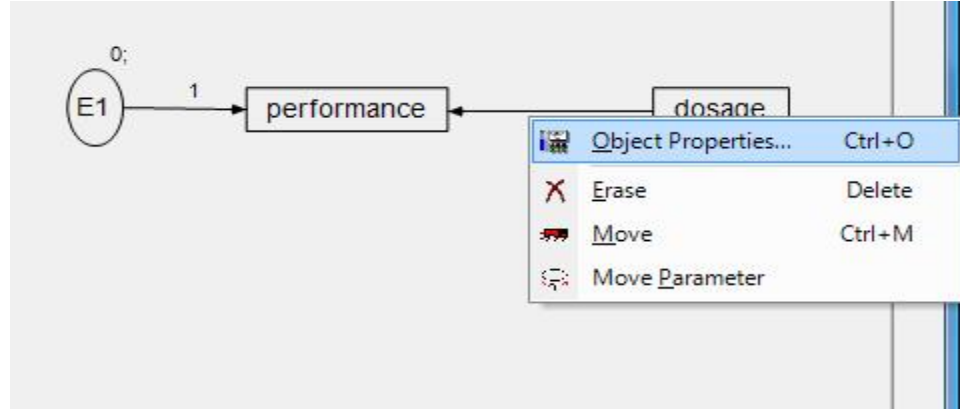
Parametre tahminleri geliştirilebilir. Ayrıca AMOS'un kümeleri oluşturma yeterliliğini de geliştirebilirsiniz (Arbuckle, 2007). Bu da, tahmin edilmesi gerekli parametrelerin sayısı azaltılarak yapılabilir. İki grup için regresyon doğrularının eğiminin (regresyon ağırlığının) yaklaşık olarak aynı olduğu görülebilmisti (Şekil 4.4). Aynı zamanda, her bir regresyon doğrusunun değişkenliğinin iki grup içinde yaklaşık olarak aynı olduğu görülmektedir. Karma model analizi içine, her iki grup için de eğimleri ve hata varyansları aynı olan hipotezleri dahil etmek olasıdır. Böylece, tahmin edilmiş olan farklı parametrelerin sayıları indirgenmiş olunur.

Bahsedilenlerin uygulaması aşağıdaki şekilde gösterilebilir:

- Yol diagramında, dozaj ve performans değişkenlerini bağlayan tek yönlü ok üzerinde sağ tıklanıp açılan pencerede Object Properties seçilir (Şekil 4.12).
- Object Properties kutusunda, parameters sekmesi seçilip, Regression weight kısmına b yazılır (Tablo 4.23).

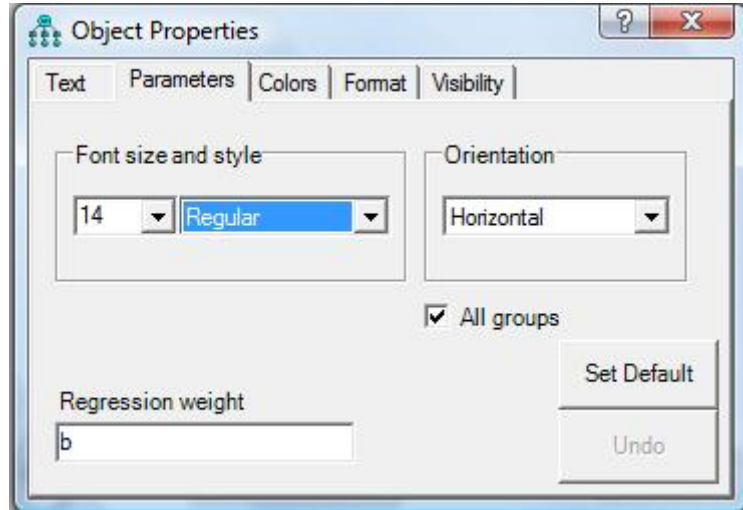
4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR



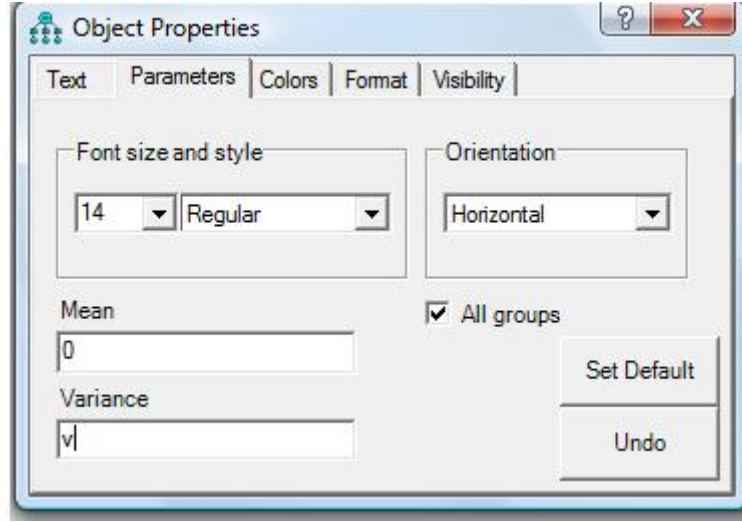
Şekil 4.13. AMOS’da hipotez testi öncesi yol diyagramı üzerinde düzenlemeler

Tablo 4.23. AMOS’da regresyon ağırlıklarının b olarak belirlenmesi

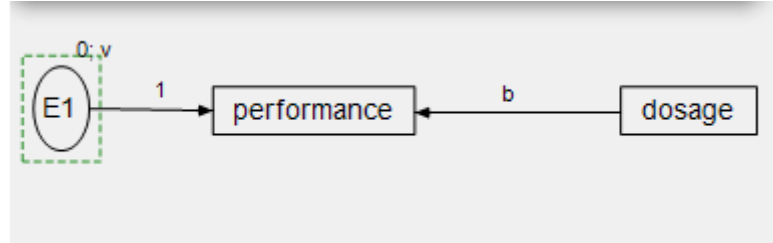


- Object Properties diyalog kutusu açık iken, yol diyagramı içindeki E1 üzerine tıklanır.
- Object Properties diyalog kutusunda variance text box kısmına parametre ismi v girilir (Tablo 4.24).

Tablo 4.24. AMOS’da hata varyansının v olarak belirlenmesi



Bu parametre kısıtlamalarının ardından yol diyagramının son hali Şekil 4.13’deki gibi olacaktır:



Şekil 4.14. Parametre kısıtlamalarının ardından yol diyagramı

İki model içinde eğim ve hata varyansının aynı olma sınırlandırmalarının ardından, Bayes butonu tıklanarak karma model analizi tekrar edilebilir. Böylece, Bayesian SEM penceresinin son hali Tablo 4.25’deki gibi olacaktır.

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

Tablo 4.25. AMOS’da parametre kısıtlandırmalarının ardından Bayesian SEM penceresi

	Mean	S.E.	S.D.	C.S.	Median	95% Lower bound	95% Upper bound	Skewness	Kurtosis	Min	Max	Name
Regression weights												
performance--dosage	2.017	0.000	0.029	1.000	2.017	1.961	2.073	-0.006	0.003	1.904	2.137	b
Means												
dosage	7.092	0.001	0.181	1.000	7.094	6.733	7.450	+0.016	0.115	6.351	7.829	
Intercepts												
performance	5.876	0.001	0.228	1.000	5.878	5.427	6.322	-0.004	0.010	4.980	6.900	
Variances												
E1	1.012	0.000	0.075	1.000	1.008	0.874	1.170	0.288	0.185	0.762	1.383	v
dosage	2.956	0.004	0.465	1.000	2.910	2.181	3.968	0.635	0.758	1.656	5.879	
Group number 1 Proportion												
Proportion	0.249	0.000	0.022	1.000	0.249	0.207	0.293	0.117	0.047	0.166	0.344	

Parametre kısıtlandırmalarının ardından, tabloda işaretlenen sonsal parametre tahminlerinin yardımıyla, 1.grup için sonsal dağılımlı ortalama bir model tahmini hesaplanabilir:

$$\begin{aligned}
 \hat{y}_1 &= \hat{\beta}_{01} + \hat{\beta}_{11}\bar{x} \\
 &= 5.876 + (2.017 \times 7.092) \\
 \hat{y}_1 &= 20.180
 \end{aligned}
 \tag{4.5}$$

Sonuç olarak, kısıtlamaların ardından 1.grub’un ortalama dozaj değerine karşılık, tahmini ortalama performans değeri 21.180 dir. Benzer şekilde 2.grup içinde hesaplamalar yapılabilir.

4.7.1. Hipotez Testi

Belirtilen kısıtlamaların ardından şu hipotezler yazılabilir:

Ho: Her iki grubun hata varyansları arasında fark yoktur.

Ha: Her iki grubun hata varyansları arasında fark vardır.

Ho: Her iki grubun eğimleri arasında fark yoktur.

Ha: Her iki grubun eğimleri arasında fark vardır.

Bayesian SEM penceresinin araç çubuğunda yer alan Fit Measures  simgesi tıklandığında hesaplanan sonsal tahmin edici p değeri (posterior predictive p value) 0.57 olarak hesaplanmaktadır. Bu durumda, $0.57 > 0.05$ olduğundan Ho hipotezleri reddedilemez. Yani, her iki grup içinde eğimler ve hata varyansları arasında fark olmadığı sonucuna varılabilir.

4.8. Tartışma ve Karşılaştırmalar

Bu bölüm boyunca, yapay bir veri kümesinde karma regresyon analizinin AMOS paket programı kullanılarak nasıl gerçekleştirilebileceği üzerine odaklanılmıştır.

Bütün bu incelemelerin ardından anlaşılmaktadır ki, AMOS'un bu analiz gerçekleştirilirken ki en etkin rolü, gözlenemeyen kitle heterojenliğinde hangi durumun (gözlemin), hangi gruba ait olduğunu belirleyebilmesidir. Bu belirlemenin ardından, AMOS ile hesapladığımız $\hat{y}_1$, $\hat{y}_2$ değerleri SPSS paket programı ile her bir grup için basit doğrusal regresyon analizi uygulanarak da, yaklaşık olarak elde edilebilir.

Örneğin AMOS'da Tablo 4.21 i kullanılarak, $\hat{y}_1 = 20.212$ tahmini ortalama performans değerini (Eşitlik 4.4) elde etmiştik. Bu sonucun, SPSS paket programı kullanarak elde edilen aşağıdaki tablolar yardımıyla da, yaklaşık olarak elde edilebileceğini görebiliriz:

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

Tablo 4.26. SPSS’de 1.grup için tanımlayıcı istatistik tablosu

Descriptive Statistics			
	Mean	Std. Deviation	N
performance	20,2599	3,57425	97
dosage	7,1359	1,63600	97

Tablo 4.27. SPSS’de 1.grup için katsayılar tablosu

Coefficients ^a								
		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	5,280	,455		11,615	,000	4,378	6,182
	dosage	2,099	,062	,961	33,800	,000	1,976	2,223

a. Dependent Variable: performance

Uygun değerlerin model içerisine yerleştirilmesiyle;

$$\begin{aligned}\hat{y}_1 &= 5.280 + (2.099 \times 7.1359) \\ &= 20,259\end{aligned}\quad (4.5)$$

elde edilir.

4.5 numaralı eşitlikte tahmini ortalama performans değeri hesaplanmaktadır.

Böylece, 1. grup için her iki paket programı ile elde edilen tahmin değerlerinin yaklaşık olarak aynı olduğu söylenebilir.

2.grup için de, aynı mantığın takip edilmesi ile karşılaştırmalarda bulunabilir. Öyle ki, Tablo 4.28’de daire içine alınan parametre tahminlerinin model içine yerleştirilmesi ile AMOS ‘da tahmini ortalama performans değeri hesaplanabilir:

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

Tablo 4.28. Özet tablosundan 2.grup için parametre tahminlerinin çekilmesi

Group number 1	Group number 2												
		Mean	S.E.	S.D.	C.S.	Median	95% Lower bound	95% Upper bound	Skewness	Kurtosis	Min	Max	Name
Regression weights													
	performance<-dosage	1,999	0,000	0,033	1,000	1,999	1,934	2,063	-0,025	-0,008	1,863	2,140	
Means													
	dosage	2,133	0,001	0,108	1,000	2,134	1,920	2,344	-0,005	0,046	1,702	2,702	
Intercepts													
	performance	9,955	0,001	0,089	1,000	9,955	9,781	10,131	0,016	0,003	9,544	10,339	
Variances													
	dosage	3,465	0,003	0,293	1,000	3,449	2,942	4,086	0,349	0,270	2,434	5,029	
	E1	1,027	0,001	0,090	1,000	1,023	0,864	1,214	0,293	0,157	0,725	1,579	
Group number 2 Proportion													
	Proportion	0,753	0,000	0,022	1,000	0,754	0,709	0,795	-0,117	0,029	0,662	0,838	

$$\begin{aligned}
 \hat{Y}_2 &= \hat{\beta}_{02} + \hat{\beta}_{12}\bar{x} \\
 &= 9.955 + (1.999 \times 2.133) \\
 &= 14.218
 \end{aligned}
 \tag{4.6}$$

elde edilir.

Böylece, 2.grup için ortalama dozaj değerine karşılık, tahmini ortalama performans değerinin 14.218 olduğu söylenebilir.

Bu sonucun yaklaşık bir değeri, SPSS paket programı kullanılarak elde edilen aşağıdaki tablolar yardımıyla da, 2. grup için elde edilebilir.

Tablo 4.29. SPSS’de 2.grup için tanımlayıcı istatistik tablosu

Descriptive Statistics			
	Mean	Std. Deviation	N
performance	14,1956	3,81125	303
dosage	2,1213	1,83355	303

4. AMOS İLE KARMA REGRESYON ANALİZİ UYGULMASI VE BAYES TAHMİN YÖNTEMİ

ASLI BAHADIR

Tablo 4.30. SPSS’de 2.grup için katsayılar tablosu

Coefficients ^a								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	9,942	,089		112,311	,000	9,768	10,116
	dosage	2,005	,032	,965	63,464	,000	1,943	2,067

a. Dependent Variable: performance

Uygun değerlerin model içerisine yerleştirilmesiyle;

$$\begin{aligned}
 \hat{y}_2 &= \hat{\beta}_{02} + \hat{\beta}_{12}\bar{x} \\
 &= 9.942 + (2.005 \times 2.1213) \\
 &= 14.195
 \end{aligned}
 \tag{4.7}$$

elde edilmektedir.

Görülebildiği üzere; her iki paket programıyla elde edilen değerler 2. grup için de, yaklaşık olarak aynıdır.

Sonuç olarak; her iki grubu da içinde barındıran karma regresyon modelinin, AMOS ile elde edilen ortalama tahmini performans değeri;

$$\begin{aligned}
 \hat{y} &= P_1\hat{y}_1 + P_2\hat{y}_2 \\
 &= (0.247 \times 20.212) + (0.753 \times 14.218) \\
 &= 15.6985
 \end{aligned}
 \tag{4.8}$$

dir.

SPSS ile elde edilen karma regresyon modelinin ortalama tahmini performans değeri de aşağıdaki tabloda görüldüğü üzere yaklaşık olarak aynı hesaplanmıştır.

Tablo 4.31. SPSS’de veri kümesi için tanımlayıcı istatistik tablosu

Descriptive Statistics			
	Mean	Std. Deviation	N
performance	15,6662	4,56515	400
dosage	3,3374	2,79631	400

5. SONUÇLAR VE ÖNERİLER

Bu çalışmada, veride gözlenemeyen bir kitle heterojenliğinin söz konusu olduğu durumlarda, karma regresyon modellerinin analizi ve uygulamaları üzerine odaklanılmıştır.

Karma regresyon modelleri, kitle heterojenliğini incelemek için bir araçtır. Karma regresyon modellerinin analizi ise, örneklemi gruplar halinde bölme ve her bir grup için regresyon doğrusu oluşturmayı amaç edinen bir analiz şeklidir.

Klasik regresyon analizi ile karma regresyon analizinin karşılaştırıldığı araştırmalar incelenmiş; klasik regresyon analizinin tek bir eşitliğin bütün gözlemler için uygun olabileceğini varsayarken, karma regresyon analizinde bir örneklem içindeki gözlemlerin farklı alt kitlelerin bir üyesi olabileceğinin öngörüldüğü sonuçlar ile karşılaşmıştır. Sonuçta bu çalışmalarda ortak bir noktada buluşmuş ve homojen gruplarla yapılan temel varsayımların çok da gerçekçi olmadığı kanısına varılmıştır.

Kitle heterojenliği ile ilgili yapılan araştırmalarda, karma regresyon modelinin çoklu grup modellemekten farklı bir analitik teknik olduğu vurgulanmış, basit bir regresyon fonksiyon formu içerisine çoklu gizli sınıfların yerleştirilebileceğine işaret edilmiştir. Karma regresyon analizinde önerilen, gizli sınıflar arasında farklı regresyon fonksiyonlarını tanımlamaktır. Bu gizli sınıflar farklı yorumlara sahiptir ve örneklem içerisinde heterojenliğin çeşitliliğini tanımlamak için kullanılabilir.

Karma regresyon kavramı, karma regresyon kullanılarak gerçekleştirilen çeşitli araştırmalar ve karma regresyon analizinin bir takım sınırlamaları gibi temel kavramların örnekler ile açıklanmasının ardından, uygulama kısmında bütün bu kavramların somut bir şekilde incelenebilmesi amacıyla detaylı bir uygulama yapılmıştır. Bunun için yapay bir veri grubunda AMOS paket programı kullanılarak karma regresyonun model yapısının görünümü, analiz şekli, dağılımlar ve hipotez testleri, model tahminin hesaplanması ve karma oranın hesaplanması şeklinde sıralayabileceğimiz alt başlık konuları detaylı bir şekilde incelenmiştir.

Sonuç olarak; bu uygulama ile AMOS paket programının analizdeki en etkin rolünün, gözlenemeyen kitle heterojenliğinde hangi durumun (gözlemin), hangi

gruba ya da sınıfa ait olduğunu belirleyebilmesi olduğu sonucuna varılmıştır. Bu belirlemenin ardından, AMOS ile hesaplanan model tahminlerinin SPSS paket programı ile her bir grup için basit doğrusal regresyon analizi uygulanarak da yaklaşık olarak elde edilebileceği görülmüştür.

Burada, bağımlı değişkenin sürekli olması ve lineer olma sınırlandırmaları altında lineer karma regresyon model tipleri incelenmiştir. Bu sınırlamalar dışında kalan, diğer karma regresyon çeşitleri ve model tipleri incelenebilir.

Bu tez çalışmasının, karma regresyon kavramı üzerine bir temel teşkil etmesi amacıyla; bir kısım incelenen örneklerde ve özellikle gerçekleştirilen uygulamada basit ve tek değişkenli karma regresyon üzerine odaklanılmıştır. Bu çalışma, çoklu veya çok düzeyli karma regresyon uygulamaları ile geliştirilebilir. Şimdiye kadar, en iyi modelin seçimi için hangi model tespit göstergelerinin kullanabileceği konusunda görüş birliğine varılamamıştır. Dolayısıyla, model seçimi ile ilgili belirsizlik hala sürmektedir. Bu nedenle bir alt başlık olarak incelenen model tespit göstergeleri üzerine daha detaylı çalışılabilir.

KAYNAKLAR

- AKAIKE, H., 1973. Information Theory as an Extension of the Maximum Likelihood Principle. In B. N. Petrov & F. Csaki (Eds.), Second International Symposium on Information Theory (pp. 267-281). Akademiai Kiado, Budapest.
- ANDERSON, E., 1935. The Irises of the Gaspé Peninsula. Bulletin of the American Iris Society, 59: 2-5.
- ANDERSON, E.B., 2004. Latent Regression Analysis Based On The Rating Scale Model. Psychology Science, 46(2): 209-226.
- ARBUCLE, J. L., 2007. AMOS 16.0 User's Guide. AMOS Development Corporation, United States of America, 631p.
- ARMINGER, G., STEIN P., and WITTENBERG J., 1999. Mixtures of Conditional Mean- and Covariance Structure Models. Psychometrika, 64(4):475-494.
- ATTIG, M.S., 1983. The Processing of Spatial Information by Adults. Presented at the Annual Meeting of The Gerontological Society, San Francisco.
- BANFIELD, J.D. and RAFTERY, A.E., 1993. Model-Based Gaussian and Non-Gaussian Clustering. Biomet-rics, 49: 803-821
- BLENDER, R., FRAEDRICH, K. and LUNKEIT, F., 1997. Identification of Cyclone-Track Regimes in the North Atlantic. Quart J. Royal Meteor. Soc., 123: 727-741.
- BOLSTAD, W. M. 2004. Introduction to Bayesian Statistics. Hoboken, NJ: John Wiley and Sons.
- BOUWMEESTER, S., SIJTSMA, K., and VERMUNT, J. K., 2004. Latent Class Regression Analysis For Describing Cognitive Developmental Phenomena: An Application To Transitive Reasoning. European Journal of Developmental Psychology, 1(1):67-86.
- DEMPSTER, A. P., LAIRD, N.M., and RUBIN, D.B., 1977. Maximum Likelihood For Incomplete Data Via The EM Algorithm. J. R. Stat. Soc., B, 39:1-38.
- DESARBO, W.S., and CRON, W.L., 1988. A Maximum Likelihood Methodology For Clusterwise Linear Regression. Journal of Classification, 5(1):249-282.

- DING, C.S., 2006. Using Regression Mixture Analysis In Educational Research. Practical Assessment Research & Evaluation, Vol. 11, No. 11. 1-11. <http://pareonline.net/pdf/v11n11.pdf>
- FRALEY, C., and RAFTERY, A.E., 1998. How Many Clusters? Which Clustering Methods? Answers via Model-Based Cluster Analysis. The Computer Journal, 41(8): 578-588
- FISHER, R. A., 1936. The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics, 7: 179-188.
- GAFNEY, S., ROBERTSON, A., and SMYTH, P., 2001. Clustering of Extra-Tropical Cyclone Trajectories Using Mixtures of Regression Models. Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Fourth Workshop on Mining Scientific Datasets.
- GAFFNEY, S. J., and SMYTH, P., 1999. Trajectory Clustering Using Mixtures Of Regression Models. In Proceedings Of The ACM 1999 Conference On Knowledge Discovery And Data Mining, S. Chaudhuri And D. Madigan (eds.), New York, NY: ACM, 63-72.
- GELFAND, A.E., and SMITH, A.F.M., 1990. Sampling Based Approaches to Calculating Marginal Densities. Journal of the American Statistical Association, 85: 398-409.
- GELMAN, A., CARLIN, J. B., STERN, H. S., and RUBIN, D. B., 2004. Bayesian Data Analysis. 2nd ed. Boca Raton: Chapman and Hall/CRC.
- HARTIGAN, J. A., and WONG, M. A., 1978. Algorithm AS 136: A K-Means Algorithm. Appl. Stat., 28:100-108.
- HOSHINO, T., 2001. Bayesian Inference for Finite Mixtures in Confirmatory Factor Analysis. Behaviormetrika, 28(1): 37-63.
- HOSMER, D., 1974. Maximum Likelihood Estimates Of The Parameters Of A Mixture Of Two Regression Lines. Communications In Statistics, 3(10): 995-1006.
- JONES, P. N., and MCLACHEN, G. J., 1992. Fitting finite mixture models in a Regression Context. Austral. J. Statist., 34(2): 233-240

- JORDAN, M. I., and JACOBS, R. A., 1994. Hierarchical Mixtures Of Experts And The EM Algorithm. *Neural Computation*, 6: 181-214.
- LAIRD, N.M., and WARE, J.H., 1982. Random Effects Models for Longitudinal Data. *Biometrics*, 38: 963-974.
- LAZARSFELD, P. F., and HENRY, N. W., 1968. *Latent Structure Analysis*. Boston: Houghton Mifflin Company.
- LEE, A.H., NG, A.S.K., YAU, K.K.W., 2001. Determinants of Maternity Length of Stay: A Gamma Mixture Risk-Adjusted Model. *Health Care Manage.Sci.*, 4: 249-255.
- LEE, A.H., XIAO, J., VEMURI, S.R., ZHAO., 1998. A Discordancy Test Approach to Identify Outliers of Length of Hospital stay. *Statist.Med.*, 17: 2199-2206.
- LEE, S.Y., 2007. *Structural Equation Modeling: A Bayesian Approach*. Chichester, UK: John Wiley and Sons.
- LENK, P.J., and DESARBO, W.S., 2000. Bayesian Inference For Finite Mixtures Of Generalized Linear Models With Random Effects. *Psychometrika*, 65(1): 93-119.
- LEUNK, K.M., ELASHOFF, R.M., HASAN, M.M., LEGORRETA, A.P., 1998. Hospital and Patient-Related Characteristics Determining Maternity Length of Stay: A Hierarchical Linear Model Approach. *Ammer. J. Public Health* 88: 377-381.
- LEYLAND, A.H., BODDY, F.A., 1997. Measuring Performance In Hospital Care: Length of Stay In Gynaecology. *European J. Pulic Health* 7: 136-143.
- LO, Y., MENDELL, N., and Rubin, D. B., 2001. Testing the Number of Components in a Normal Mixture. *Biometrika*, 88: 767-778.
- LOKEN, E., 2004. Using Latent Class Analysis to Model Temperament Types. *Multivariate Behavioral Research*, 39(4): 625–652.
- LUBKE, G. H., and MUTHEN, B., 2005. Investigating Population Heterogeneity with Factor Mixture Models. *Psychological Methods*, 10(1): 21-39.
- MARBELLA, A.M, CHETTY, V.K., LAYDE, P.M., 1998. Neonatal Hospital Lengths of Stay, Readmissions, and charges. *Pediatrics*, 101: 32-36.

- MARSH, H. W., 1990. Self-Description Questionnaire Manual. Campbelltown N. S. W., Australia: University of Western Sydney, Macarthur.
- MCCULLAGH, P., and NELDER, J.A., 1983. Generalized Linear Models, New York: Chapman and Hall.
- MCCUTCHEON, A. L., 1987. Latent Class Analysis. Thousand Oaks, CA: Sage Publications, Inc.
- MCGILCHRIST, C.A., 1994. Estimation in Generalised Mixed Models. *J. Roy. Statist. Soc. B.*, 56: 61-69.
- MCGILCHRIST, C.A., YAU, K.K.W., 1995. The Derivation of BLUP, ML, REML Estimation Methods for Generalized Linear Mixed Models. *Commun Statist. Theory Method*, 24: 2963-2980.
- MCLACHLAN, G.J., and KRISHNAN, T., 1997. The EM Algorithm and Extensions, New York, NY: John Wiley and Sons.
- MCLACHLAN, G.J., PEEL, D., 2000. Finite Mixture Models. Wiley, New York.
- MENG, X.L., VAN DYK, D.A., 1998. Fast EM- Type Implementations for Mixed Effects Models. *J. Roy. Statist. Soc. B.*, 60: 559-578.
- MUTHEN, B. O., 2001. Second-Generation Structural Equation Modeling with a Combination of Categorical and Continuous Latent Variables: New Opportunities for Latent Class/Latent Growth Modeling. In L. M. Collins & A. Sayer (Eds.), *New Methods for the Analysis of Change* (pp. 291-322). Washington, DC: APA.
- MUTHEN, B., and ASPAROUHOV, T., 2009. Multilevel Regression Mixture Analysis. *Journal of the Royal Statistical Society*, 172(3): 639-657
- MUTHEN, B. O., and MUTHEN, L. K., 2000. Integrating Person-Centered and Variable-Centered Analyses: Growth Mixture Modeling with Latent Trajectory Classes. *Alcoholism: Clinical and Experimental Research*, 24: 882-891.
- MUTHEN, L. K., and MUTHEN, B. O., 2001. *Mplus User's Guide*. Los Angeles, CA: Muthen & Muthen.
- NAGIN, D., and TREMBLAY, R. E., 2001. Analyzing Developmental Trajectories of Distinct but Related Behaviors: A Group-Based Method. *Psychological Methods*, 6: 18-34.

- QUANDT, R.E., 1972. A New Approach to Estimating Switching Regressions. *Journal of the American Statistical Association*, 67: 306-310.
- QUANT, R.E., and RAMSEY, J.B., 1978. Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association*, 73: 730-738.
- RAMASWAMY, V., DESARBO, W., REIBSTEIN, D., and ROBINSON, W., 1993. An Empirical Pooling Approach for Estimating Marketing Mix Elasticities with PIMS Data. *Marketing Science*, 12: 103-124.
- SCHWARZ, G., 1978. Estimating the Dimension of a Model. *Annals of Statistics*, 6: 461-464.
- SPATH, H., 1979. Algorithm 39: Clusterwise Linear Regression. *Computing*, 22: 367-373.
- SKRONDAL, A., and RABE-HESKETH, S., 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca, Raton: Chapman & Hall/CRC.
- THOMPSON, T.J., SMITH, P.J., 1998. Finite Mixtures Models with Concomitant Information: Assessing Diagnostic Criteria For Diabetes. *Appl. Statist.* 47: 393-404.
- TOURANGEAU, K., NORD, C., LE, T., POLLACK, J. M., and ATKINS-BURNETT, S., 2006. Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Combined User's Manual for the ECLS-K Fifth-Grade Data Files and Electronic Codebooks (NCES 2006–032). Washington, DC: U.S. Department of Education: National Center for Education Statistics.
- TURNER, T.R., 2000. Estimating the Propagation Rate of a Viral Infection of Potato Plants via Mixtures of Regressions. *Applied Statistics*, 49: 371-384.
- VERMUNT, J. K., and MAGIDSON, J., 2002. Latent Class Cluster Analysis. In J. A. Hagenaars & L. M. Allan (Eds.), *Applied Latent Class Analysis*. Cambridge: Cambridge University Press.
- _____, 2005a. *Latent Gold 4.0 User's Guide*. Belmont, MA: Statistical Innovations Inc.

- _____, 2005b. Structural Equation Models: Mixture Models. In: Encyclopedia of Statistics in Behavioral Science, B. Everitt and D. Howell, Eds. Chichester, UK: John Wiley and Sons, 1922–1927.
- WALLER, N., and MEEHL, P.E., 1998. Multivariate Taxometric Procedures: Distinguishing Types from Continua. Thousand Oaks, CA: Sage Publications, Inc.
- XIAO, J., LEE, A.H., VERMURI, S.R., 1999. Mixture Distribution Analysis of Length of Hospital Stay for Efficient Funding. *Socio-Econom. Planning Sci.* 33: 39-59.
- YAU, K.K.W., LEE, A.H., and NG, A.S.K., 2003. Finite Mixture Regression Model with Random Effects: Application to Neonatal Hospital Length of Stay. *Computational Statistics & Data Analysis* 41: 359-366.
- YOUNG, D., 2007. Mixtures of Regressions. Rao Prize Day: Joint Work with Dr. David Hunter.
- ZHU, H. T., and LEE, S. Y., 2001. A Bayesian Analysis of Finite Mixtures in the Lisrel Model. *Psychometrika*, 66(1): 133–152.

ÖZGEÇMİŞ

1983 yılında Adana ‘da doğdu. Öğrenim hayatı Adana ‘da geçti. İlköğretimi Bilfen Koleji’nde bitirdi. Ardından Anadolu Lisesi sınavlarına girerek, hazırlık, ortaokul ve lise eğitimini, ÇEAŞ Seyhan Anadolu Lisesi’nde tamamladı. 2003 yılında Çukurova Üniversitesi Fen Edebiyat Fakültesi İstatistik bölümüne girdi ve 2007 yılında bölüm üçüncülüğü ile lisans eğitimini bitirdi. Aynı yıl içerisinde istatistik yüksek lisans eğitimine başladı. İki yıllık yüksek lisans eğitiminin ders aşamasını tamamlayıp, kısa bir dönem sağlık sektöründe çalıştıktan sonra, buradan ayrılıp tez çalışmasına odaklandı. Eğitimine doktora ile devam etmek istiyor.

EKLER

EK-1

Anderson (1935) tarafından toparlanmış veri kümesi, incelenen 150 farklı bitkinin çiçekleri üzerinde yapılan ölçümleri göstermektedir. Bu veri kümesine, <http://www.amosdevelopment.com/video/index.htm> internet adresinden ulaşılabilir.

case no	sepalen	sepalwid	petallen	petalwid	species
1	5,1	3,5	1,4	0,2	setosa
2	4,9	3	1,4	0,2	setosa
3	4,7	3,2	1,3	0,2	setosa
4	4,6	3,1	1,5	0,2	setosa
5	5	3,6	1,4	0,2	setosa
6	5,4	3,9	1,7	0,4	setosa
7	4,6	3,4	1,4	0,3	setosa
8	5	3,4	1,5	0,2	setosa
9	4,4	2,9	1,4	0,2	setosa
10	4,9	3,1	1,5	0,1	setosa
11	5,4	3,7	1,5	0,2	setosa
12	4,8	3,4	1,6	0,2	setosa
13	4,8	3	1,4	0,1	setosa
14	4,3	3	1,1	0,1	setosa
15	5,8	4	1,2	0,2	setosa
16	5,7	4,4	1,5	0,4	setosa
17	5,4	3,9	1,3	0,4	setosa
18	5,1	3,5	1,4	0,3	setosa
19	5,7	3,8	1,7	0,3	setosa
20	5,1	3,8	1,5	0,3	setosa
21	5,4	3,4	1,7	0,2	setosa
22	5,1	3,7	1,5	0,4	setosa
23	4,6	3,6	1	0,2	setosa
24	5,1	3,3	1,7	0,5	setosa
25	4,8	3,4	1,9	0,2	setosa
26	5	3	1,6	0,2	setosa
27	5	3,4	1,6	0,4	setosa
28	5,2	3,5	1,5	0,2	setosa
29	5,2	3,4	1,4	0,2	setosa
30	4,7	3,2	1,6	0,2	setosa

31	4,8	3,1	1,6	0,2	setosa
32	5,4	3,4	1,5	0,4	setosa
33	5,2	4,1	1,5	0,1	setosa
34	5,5	4,2	1,4	0,2	setosa
35	4,9	3,1	1,5	0,2	setosa
36	5	3,2	1,2	0,2	setosa
37	5,5	3,5	1,3	0,2	setosa
38	4,9	3,6	1,4	0,1	setosa
39	4,4	3	1,3	0,2	setosa
40	5,1	3,4	1,5	0,2	setosa
41	5	3,5	1,3	0,3	setosa
42	4,5	2,3	1,3	0,3	setosa
43	4,4	3,2	1,3	0,2	setosa
44	5	3,5	1,6	0,6	setosa
45	5,1	3,8	1,9	0,4	setosa
46	4,8	3	1,4	0,3	setosa
47	5,1	3,8	1,6	0,2	setosa
48	4,6	3,2	1,4	0,2	setosa
49	5,3	3,7	1,5	0,2	setosa
50	5	3,3	1,4	0,2	setosa
51	7	3,2	4,7	1,4	versicolor
52	6,4	3,2	4,5	1,5	versicolor
53	6,9	3,1	4,9	1,5	versicolor
54	5,5	2,3	4	1,3	versicolor
55	6,5	2,8	4,6	1,5	versicolor
56	5,7	2,8	4,5	1,3	versicolor
57	6,3	3,3	4,7	1,6	versicolor
58	4,9	2,4	3,3	1	versicolor
59	6,6	2,9	4,6	1,3	versicolor
60	5,2	2,7	3,9	1,4	versicolor
61	5	2	3,5	1	versicolor
62	5,9	3	4,2	1,5	versicolor
63	6	2,2	4	1	versicolor
64	6,1	2,9	4,7	1,4	versicolor
65	5,6	2,9	3,6	1,3	versicolor
66	6,7	3,1	4,4	1,4	versicolor
67	5,6	3	4,5	1,5	versicolor
68	5,8	2,7	4,1	1	versicolor
69	6,2	2,2	4,5	1,5	versicolor
70	5,6	2,5	3,9	1,1	versicolor
71	5,9	3,2	4,8	1,8	versicolor

72	6,1	2,8	4	1,3	versicolor
73	6,3	2,5	4,9	1,5	versicolor
74	6,1	2,8	4,7	1,2	versicolor
75	6,4	2,9	4,3	1,3	versicolor
76	6,6	3	4,4	1,4	versicolor
77	6,8	2,8	4,8	1,4	versicolor
78	6,7	3	5	1,7	versicolor
79	6	2,9	4,5	1,5	versicolor
80	5,7	2,6	3,5	1	versicolor
81	5,5	2,4	3,8	1,1	versicolor
82	5,5	2,4	3,7	1	versicolor
83	5,8	2,7	3,9	1,2	versicolor
84	6	2,7	5,1	1,6	versicolor
85	5,4	3	4,5	1,5	versicolor
86	6	3,4	4,5	1,6	versicolor
87	6,7	3,1	4,7	1,5	versicolor
88	6,3	2,3	4,4	1,3	versicolor
89	5,6	3	4,1	1,3	versicolor
90	5,5	2,5	4	1,3	versicolor
91	5,5	2,6	4,4	1,2	versicolor
92	6,1	3	4,6	1,4	versicolor
93	5,8	2,6	4	1,2	versicolor
94	5	2,3	3,3	1	versicolor
95	5,6	2,7	4,2	1,3	versicolor
96	5,7	3	4,2	1,2	versicolor
97	5,7	2,9	4,2	1,3	versicolor
98	6,2	2,9	4,3	1,3	versicolor
99	5,1	2,5	3	1,1	versicolor
100	5,7	2,8	4,1	1,3	versicolor
101	6,3	3,3	6	2,5	virginica
102	5,8	2,7	5,1	1,9	virginica
103	7,1	3	5,9	2,1	virginica
104	6,3	2,9	5,6	1,8	virginica
105	6,5	3	5,8	2,2	virginica
106	7,6	3	6,6	2,1	virginica
107	4,9	2,5	4,5	1,7	virginica
108	7,3	2,9	6,3	1,8	virginica
109	6,7	2,5	5,8	1,8	virginica
110	7,2	3,6	6,1	2,5	virginica
111	6,5	3,2	5,1	2	virginica
112	6,4	2,7	5,3	1,9	virginica

113	6,8	3	5,5	2,1	virginica
114	5,7	2,5	5	2	virginica
115	5,8	2,8	5,1	2,4	virginica
116	6,4	3,2	5,3	2,3	virginica
117	6,5	3	5,5	1,8	virginica
118	7,7	3,8	6,7	2,2	virginica
119	7,7	2,6	6,9	2,3	virginica
120	6	2,2	5	1,5	virginica
121	6,9	3,2	5,7	2,3	virginica
122	5,6	2,8	4,9	2	virginica
123	7,7	2,8	6,7	2	virginica
124	6,3	2,7	4,9	1,8	virginica
125	6,7	3,3	5,7	2,1	virginica
126	7,2	3,2	6	1,8	virginica
127	6,2	2,8	4,8	1,8	virginica
128	6,1	3	4,9	1,8	virginica
129	6,4	2,8	5,6	2,1	virginica
130	7,2	3	5,8	1,6	virginica
131	7,4	2,8	6,1	1,9	virginica
132	7,9	3,8	6,4	2	virginica
133	6,4	2,8	5,6	2,2	virginica
134	6,3	2,8	5,1	1,5	virginica
135	6,1	2,6	5,6	1,4	virginica
136	7,7	3	6,1	2,3	virginica
137	6,3	3,4	5,6	2,4	virginica
138	6,4	3,1	5,5	1,8	virginica
139	6	3	4,8	1,8	virginica
140	6,9	3,1	5,4	2,1	virginica
141	6,7	3,1	5,6	2,4	virginica
142	6,9	3,1	5,1	2,3	virginica
143	5,8	2,7	5,1	1,9	virginica
144	6,8	3,2	5,9	2,3	virginica
145	6,7	3,3	5,7	2,5	virginica
146	6,7	3	5,2	2,3	virginica
147	6,3	2,5	5	1,9	virginica
148	6,5	3	5,2	2	virginica
149	6,2	3,4	5,4	2,3	virginica
150	5,9	3	5,1	1,8	virginica

EK-2

Dozaj ve Performans 1 için yapay veri kümesi aşağıdaki gibidir. Bu veri kümesine, <http://www.amosdevelopment.com/video/index.htm> internet adresinden ulaşılabilir.

case no	dosage	performance	group	case no	dosage	performance	group
1	5,26	-0,35		36	2,49	3,93	
2	2,1	4,86		37	-0,11	6,47	
3	1,93	5,2		38	3,55	3,92	
4	1,21	3,5		39	2,02	2,94	
5	1,25	2,61		40	2,66	2,13	
6	-0,86	6,64		41	1,39	4,74	
7	3,85	2,31		42	2,69	3,97	
8	2,51	2,6		43	1,86	4,97	
9	1,29	4,06		44	4,5	1,12	
10	0,9	6,12		45	-1,46	6,26	
11	2,11	4,23		46	3,35	1,75	
12	0,42	6,68		47	3,83	2,06	
13	2,53	3,48		48	-0,61	4,86	
14	0,88	6,51		49	0,63	3,99	
15	0,34	5,39		50	0,74	2,69	
16	4,88	1,9		51	3,86	1,36	
17	2,77	2,66		52	-1,54	8,45	
18	1,14	4,26		53	-0,74	8,35	
19	4,04	0,55		54	1,78	3,15	
20	2,06	3,4		55	-1,26	7,2	
21	-0,41	5,22		56	1,72	3,43	
22	-1,07	5,34		57	3,56	3,86	
23	1,54	3,63		58	2,68	3,4	
24	0,73	4,86		59	1,63	4,01	
25	0,89	4,66		60	4,13	0,42	
26	3,97	1,37		61	1,84	2,81	
27	1,46	4,12		62	4,55	1,12	
28	5,15	1,12		63	3,98	2,52	
29	3,21	3,93		64	1,05	4,18	
30	4	2,14		65	0,92	4,28	
31	2,06	3,02		66	-0,43	7,81	
32	3,37	2,07		67	2,35	3,52	
33	1,72	5,94		68	-0,17	6,52	
34	4,36	0,76		69	-1,63	6,73	
35	2,31	5,1		70	4,17	2,02	

case no	dosage	performance	group	case no	dosage	performance	group
71	3,66	-0,49		106	4,29	2,07	
72	-1,61	6,64		107	3,28	2,7	
73	0,75	5,63		108	2,27	4,71	
74	2,6	3,81		109	0,45	7,2	
75	3,65	3,06		110	-0,54	7,5	
76	1,63	6,06		111	2,75	2,2	
77	4,5	0,46		112	0,11	5,15	
78	1,43	6,05		113	1,82	3,56	
79	-1,55	7,85		114	3,71	3,94	
80	-0,68	5,87		115	3,73	2,18	
81	2,68	3,72		116	-0,3	6,25	
82	3,36	4,64		117	2,22	1,68	
83	1,7	2,88		118	2,68	2,49	
84	1,38	5,2		119	0,27	6,81	
85	-0,09	4,57		120	5,2	1,13	
86	3,72	0,9		121	3,2	3,33	
87	3,75	3,25		122	2,16	4,16	
88	3,78	1,74		123	2,11	2,89	
89	1,66	3,86		124	0,11	5,85	
90	2,97	3,01		125	1,64	4,29	
91	0,27	5,49		126	1,08	5,83	
92	1,58	4,94		127	2,79	2,45	
93	-2,81	7,79		128	3,31	1,23	
94	0,97	4,97		129	1,85	2,81	
95	4,54	2,17		130	3,6	1,65	
96	-0,9	7,83		131	1,11	3,1	
97	3,06	1,4		132	1,94	5,27	
98	0,65	5,5		133	2,13	4,89	
99	1,77	4,78		134	-0,27	4,4	
100	3,88	2,09		135	1,29	4,4	
101	-0,43	7,04		136	3	2,54	
102	0,23	5,62		137	4,74	1,33	
103	2,85	2,17		138	0,62	5,87	
104	1,28	4,94		139	1,64	3,6	
105	1,99	3,35		140	-0,97	7,19	

case no	dosage	performance	group	case no	dosage	performance	group
141	1,37	3,37		176	4,34	1,08	
142	1,56	3,39		177	3,25	2,82	
143	2,12	5,24		178	3,15	3,75	
144	2,46	5,36		179	2,57	2,91	
145	1,05	5,39		180	0,22	6,85	
146	3,44	1,84		181	0,86	5,82	
147	2,18	3,91		182	3,7	3	
148	-1,52	7,13		183	4	2,56	
149	-1,01	7,6		184	2,28	6,4	
150	1,67	3,58		185	2,83	2,7	
151	1,78	4,51		186	2,93	3,83	
152	0,06	7,6		187	1,58	2,05	
153	2,98	3,26		188	2,92	2,15	
154	0,92	5,33		189	2,68	3,07	
155	4,26	1,21		190	1,17	6,83	
156	1,43	5,36		191	0,65	6,28	
157	1,74	5,88		192	2,19	4,11	
158	1,48	4,32		193	1,86	2,21	
159	0,2	5,57		194	0,47	6,75	
160	3,65	0,54		195	4,07	0,48	
161	4,58	2,85		196	0,45	5,59	
162	1,84	5,67		197	3,46	3,7	
163	0,29	4,42		198	4,78	-0,45	
164	2,72	5,1		199	-0,6	7,91	
165	2,04	4,33		200	1,88	2,25	
166	1,63	4,09		201	0,13	5,53	
167	1,28	5,95		202	0,84	5,67	
168	3,1	2,93		203	0,96	4,21	
169	1,92	5,58		204	0,88	5,15	
170	2,31	3,4		205	2,53	3,22	
171	3,38	2,17		206	0,04	5,7	
172	0,61	5,98		207	-0,46	6,24	
173	3,15	2,92		208	2,94	2,05	
174	2,93	2,93		209	5,64	1,64	
175	2,36	3,49		210	5,67	0,48	

case no	dosage	performance	group	case no	dosage	performance	group
211	3,54	2,38		246	1,51	2,51	
212	0,71	5,83		247	1,78	4,11	
213	4,54	0,87		248	3,21	4,57	
214	3,37	1,01		249	1,97	4,57	
215	-1,16	7,54		250	-0,01	5,35	
216	2,6	3,61		251	2,18	4,89	
217	1,9	3,58		252	0,97	5,65	
218	2,24	2,13		253	2,89	3,43	
219	0,77	6,57		254	3,41	1,27	
220	-0,21	6,93		255	2,42	3,45	
221	1,85	2,57		256	2,71	3,37	
222	2,07	3,47		257	4,62	2,04	
223	1,95	4,25		258	0,58	6,09	
224	3,28	3,25		259	-0,12	7,17	
225	1,02	5,67		260	2,88	2,95	
226	4,31	2,9		261	-1,28	6,29	
227	0,03	6,41		262	2,92	3,48	
228	2,44	2,84		263	2,62	5,82	
229	3,11	3,67		264	1,48	3,25	
230	1,35	4,38		265	2,94	2,92	
231	0,68	7,08		266	-0,34	6,02	
232	2,73	1,92		267	2,41	3,82	
233	4,82	-2,17		268	5,6	0,45	
234	4,42	1,21		269	3,62	0,63	
235	1,44	5,58		270	2,29	3,06	
236	1,04	4,65		271	0,44	6,79	
237	0,71	4,75		272	2,92	2,64	
238	3,58	2,56		273	2,41	1,71	
239	-0,57	6,71		274	1,12	4,33	
240	4,63	0,66		275	2,7	3,71	
241	-1,45	8,38		276	1,19	3,42	
242	-1,09	6,04		277	0,68	5,42	
243	1,67	6,33		278	3,31	3,18	
244	0,67	5,61		279	-2,31	7,62	
245	1,51	5,37		280	3,08	4,22	

case no	dosage	performance	group	case no	dosage	performance	group
281	1,86	5,18		316	4,16	1,48	
282	3,57	1,23		317	-2,93	8,72	
283	1,08	5,88		318	1,2	3,61	
284	0,9	6,4		319	1,35	4,12	
285	1,66	5,46		320	1,71	4,97	
286	0,2	4,99		321	3,71	3,14	
287	1,03	6,38		322	-0,92	6,51	
288	4,57	-0,38		323	1,59	4,7	
289	3,42	3,97		324	0,26	6,14	
290	2,05	4,04		325	2,69	3,93	
291	1,3	2,68		326	4,05	-0,41	
292	0,36	5,45		327	3,41	1,28	
293	-1,71	6,62		328	3,18	2,02	
294	3,79	3,52		329	3,85	2,02	
295	2,3	4,15		330	1,33	5,87	
296	3,65	2,68		331	1,6	5,11	
297	3,14	1,19		332	4,23	0,35	
298	3,08	4,32		333	1,84	6,11	
299	3,93	3,86		334	1,71	4,53	
300	3,28	1,47		335	1,3	5,43	
301	5,07	0,19		336	0,29	6,78	
302	2,48	3,65		337	3,63	2,67	
303	2,08	2,63		338	0,96	4,55	
304	-0,59	6,85		339	2,01	3,71	
305	1,99	4,81		340	5,27	0,92	
306	2,86	4,36		341	2,51	1,57	
307	3,83	3,18		342	1,45	3,74	
308	2,68	3,31		343	0,44	3,94	
309	0,33	5,06		344	3,11	0,8	
310	3,79	1,64		345	3,3	4,16	
311	-0,63	7,05		346	-0,6	7,33	
312	3,38	2,14		347	4,31	0,19	
313	6,54	-0,91		348	3,32	2,53	
314	2,17	4,47		349	0,55	4,24	
315	0,88	6,15		350	2,97	1,74	

case no	dosage	performance	group	case no	dosage	performance	group
351	-1,47	7,95		386	0,93	5,89	
352	3,52	1,01		387	-0,65	7,26	
353	1,69	3,53		388	-2,33	8,17	
354	4,62	1,05		389	5,26	1,41	
355	1	4,77		390	2,3	2,59	
356	4,59	-0,19		391	2,18	3,05	
357	0,74	4,89		392	6,61	-1,64	
358	5,57	1,48		393	7,26	-1,37	
359	4,06	2,79		394	4,63	1,16	
360	-2,2	8,68		395	-0,49	6,9	
361	5,16	0,33		396	-1,35	6,82	
362	0,41	6,18		397	2	3,63	
363	3,94	2,38		398	2,54	4,6	
364	1,33	4,27		399	0,93	5,57	
365	6,51	-0,44		400	1,91	3,27	
366	3,32	3,13		401	0,75	11,73	
367	-0,08	5,37		402	1,92	14,98	
368	5,62	-0,99		403	4,07	20,05	
369	4,51	1,5		404	3,29	19,36	
370	1,37	4,92		405	3,9	20,79	
371	0,79	4,86		406	0,16	12,93	
372	0,84	5,37		407	1,87	14,95	
373	4,75	1,35		408	2,94	18,38	
374	3,65	2,35		409	-0,52	12,42	
375	2,48	4,85		410	2,67	18,57	
376	4,53	-0,51		411	2,17	14,97	
377	0,28	5,32		412	1,72	15,66	
378	0,69	5,96		413	2,46	16,79	
379	0,74	3,26		414	2,61	18,21	
380	5,24	1,24		415	3,09	18,23	
381	0,37	4,98		416	0,83	15,13	
382	1,28	3,8		417	3,67	19,66	
383	3,57	1,83		418	1,36	16,4	
384	-0,18	6,08		419	0,82	12,77	
385	1,13	4,56		420	2,95	17,84	

case no	dosage	performance	group	case no	dosage	performance	group
421	1,34	14,22		456	-0,11	11,6	
422	-1,44	7,99		457	3,4	20,6	
423	2,27	18,25		458	2,99	18,85	
424	3,35	20,19		459	1,37	15,27	
425	1,85	15,68		460	0,36	12,91	
426	2,77	19		461	2,78	17,02	
427	2,77	17,83		462	3,17	18,71	
428	2,43	17,1		463	-0,83	9,32	
429	-2,73	6,06		464	-1,8	7,75	
430	4,29	21,2		465	3,25	17,9	
431	2,06	16,49		466	0,08	11,63	
432	3,17	19,68		467	2,23	16,64	
433	1,82	15,44		468	2,4	15,56	
434	0,24	12,84		469	-1,67	8,61	
435	0,59	13,87		470	0,41	13,97	
436	1,72	16,67		471	1,79	15,22	
437	1,93	16,79		472	0,43	13,22	
438	5,16	22,74		473	2,1	15,51	
439	3,56	19,74		474	4,01	20,47	
440	1,64	14,44		475	3,99	19,28	
441	0,4	13,41		476	0,01	10,79	
442	3,99	19,18		477	1,78	16,03	
443	5,33	21,6		478	3,01	16,82	
444	2,06	15,42		479	1,12	14,06	
445	-0,36	13,04		480	0,2	12,01	
446	5,69	23,61		481	2,14	13,41	
447	1,54	14,73		482	0,53	13,55	
448	4,58	20,92		483	1,74	15,39	
449	4,58	19,72		484	-1,08	9,55	
450	1,65	14,78		485	1,5	14,16	
451	1,59	16,06		486	-2,06	7,55	
452	0,13	13,67		487	1,89	16,31	
453	0,75	14,18		488	2,38	16,87	
454	4,61	19,98		489	3,53	18,44	
455	-2,27	6,57		490	0,54	13,4	

case no	dosage	performance	group	case no	dosage	performance	group
491	1,26	14,26		526	5,28	22,74	
492	3,35	19,41		527	1,45	12,84	
493	4,31	21,3		528	2,44	17,35	
494	4,25	21,74		529	3,81	20,28	
495	2,89	17,27		530	3,64	19,51	
496	3,46	18,7		531	6,81	26,49	
497	1,86	15,31		532	0,86	13,93	
498	-1,15	7,68		533	0	11,47	
499	3,19	18,94		534	0,53	13,06	
500	-1,06	10,81		535	1,05	14,66	
501	1,57	16,95		536	0,59	14,46	
502	1,29	15		537	2,96	19,9	
503	3,96	19,52		538	4,17	22,18	
504	5,01	21,4		539	1,25	16,06	
505	-1,41	9,48		540	3,86	19,14	
506	2	15,43		541	2,01	14,52	
507	2,24	16,85		542	4,55	22,42	
508	2,29	16,04		543	2,25	15,11	
509	2,94	19,07		544	6,15	23,56	
510	-0,59	9,78		545	2,34	16,76	
511	-0,78	9,7		546	2,8	16,93	
512	5,57	22,74		547	2,59	17,04	
513	0,75	13,23		548	6,65	26,46	
514	2,59	18,93		549	-0,02	11,49	
515	2,24	15,03		550	1,15	14,34	
516	2,89	18,56		551	-0,47	10,11	
517	0,3	13,58		552	2,76	17,74	
518	0,87	12,34		553	2,25	17,07	
519	1,9	15,84		554	1,61	14,54	
520	0,5	12,44		555	1,33	14,3	
521	2,44	18,41		556	3,26	19,07	
522	0,34	13,1		557	1,46	15,58	
523	3,64	19,37		558	1,03	15,32	
524	1,22	14,96		559	0,81	13,85	
525	0,79	12,97		560	4,13	20,82	

case no	dosage	performance	group	case no	dosage	performance	group
561	4,24	20,86		601	1,58	14,53	
562	0,81	13,67		602	1,14	14,65	
563	2,58	15,53		603	0,41	11,55	
564	-0,74	11,16		604	1,49	14	
565	2,98	18,86		605	1,36	15,92	
566	2,93	18,3		606	4,59	19,62	
567	2,7	17,99		607	2,51	17,53	
568	3,43	18,76		608	-0,04	12,42	
569	-1,8	7,62		609	2,92	17,28	
570	3,78	19,54		610	0,99	13,33	
571	0,57	13,73		611	1,38	15,51	
572	3,76	18,7		612	2,58	16,27	
573	3,04	18,04		613	4,14	20,8	
574	4,74	22,17		614	1,67	15,32	
575	2,03	14,46		615	0,74	14,07	
576	2,7	18,22		616	2,79	17,48	
577	-0,99	9,85		617	-0,63	13,84	
578	2,64	16,84		618	0,69	13,38	
579	4,49	21,9		619	4,29	19,61	
580	4,54	20,65		620	0,71	13,14	
581	2,6	16,54		621	1,72	15,97	
582	0,92	13,43		622	5,44	21,92	
583	0,59	10,69		623	3,38	19	
584	2,83	15,79		624	0,1	11,01	
585	2,8	17,3		625	0,5	11,35	
586	1,98	16,98		626	2,72	18,72	
587	3,28	17,21		627	3,79	20,51	
588	1,06	13,56		628	0,52	12,32	
589	3,49	17,91		629	5,34	23,8	
590	0,66	12,53		630	1,59	16,53	
591	5,61	23,04		631	0,68	11,69	
592	-0,75	10,22		632	-0,43	10,79	
593	3,08	18,46		633	2,77	15,39	
594	3,44	19,27		634	1,02	14,28	
595	0,97	12,92		635	2,26	15,46	
596	-0,43	9,51		636	2,7	15,21	
597	5,48	23,08		637	3,66	19,53	
598	2,59	16,37		638	2,76	17,17	
599	2,21	16,9		639	2,5	16,41	
600	4,49	18,72		640	1,72	17,27	

case no	dosage	performance	group	case no	dosage	performance	group
641	0,82	13,31		681	2,42	17,06	
642	4,09	20,73		682	0,75	12,57	
643	0,7	14,94		683	2,14	16,84	
644	1,99	15,06		684	1,47	13,92	
645	2,28	16,97		685	1,4	14,48	
646	3,42	18,85		686	-3,18	3,88	
647	0,51	13,33		687	3	16,43	
648	5,47	24,03		688	5,15	19,7	
649	2,03	16,63		689	1,69	16,04	
650	3,13	19,07		690	-1,06	11,16	
651	4,03	19,26		691	5,61	24,21	
652	1,74	15,35		692	3,71	19,23	
653	6,13	24,56		693	2,93	17,76	
654	-0,32	11,65		694	1,69	14,44	
655	3,16	18,03		695	0,87	13,1	
656	3,73	21,69		696	1,14	13,55	
657	1,26	14,33		697	0,47	11,79	
658	6,4	25,66		698	3,13	17,84	
659	1,9	15,09		699	2,36	15,61	
660	0,78	13,13		700	1,15	13,78	
661	2,43	16,12		701	1,24	14,26	
662	1,61	14,19		702	1,87	14,9	
663	2,4	19,39		703	0,97	14,14	
664	1,84	16,84		704	0,9	14,88	
665	3,05	17,87		705	4,33	21,05	
666	4,46	20,16		706	-0,66	9,87	
667	3,31	18,25		707	3,51	18,35	
668	2,93	17,74		708	1,59	14,29	
669	4,8	21,99		709	2,3	17,23	
670	2,9	17,01		710	0,81	14,83	
671	3,43	17,82		711	-0,84	11,41	
672	3,53	17,32		712	3,32	18,17	
673	3,41	19,21		713	0,06	10,49	
674	-2,5	6,67		714	1,6	14,3	
675	1,7	13,63		715	4,22	21,34	
676	0,83	13,89		716	1,14	12,83	
677	0,03	10,81		717	3,35	18,26	
678	1,64	14,35		718	4,79	20,76	
679	1,38	14,88		719	3,75	19,06	
680	2,56	17,45		720	2,7	17,08	

case no	dosage	performance	group	case no	dosage	performance	group
721	0,08	12,04		761	1,25	14,2	
722	1,69	15,79		762	2,11	15,98	
723	1,48	16,09		763	-0,18	11,78	
724	2,61	15,75		764	1,92	14,63	
725	2,41	16,5		765	4,49	21,27	
726	1,11	16,03		766	2,79	17	
727	0,28	12,92		767	2,27	16,3	
728	4,15	20,24		768	2,99	17,84	
729	-0,27	13,71		769	3,57	18,3	
730	0,37	11,85		770	0,37	12,13	
731	4,09	19,15		771	7,34	26,25	
732	-0,89	8,97		772	1,24	15,46	
733	1,18	14,56		773	3,28	18,2	
734	6,89	25,06		774	-0,18	11,46	
735	0,01	11,08		775	-0,05	11,8	
736	-0,09	11,48		776	-0,48	10,61	
737	4,34	20,01		777	3,05	19,65	
738	2,06	16,96		778	6,24	24,85	
739	3,31	19,11		779	1,76	14,26	
740	1,3	14,15		780	0,48	12,86	
741	4,02	20,48		781	4,55	19,5	
742	4,27	21,51		782	3,03	17,78	
743	1,62	15,26		783	2,53	15,32	
744	1,03	13,59		784	1,84	15,49	
745	1,05	14,67		785	0,56	12,44	
746	0,53	13,02		786	4,62	21,22	
747	-1,6	8,17		787	1,53	15,63	
748	1,78	17,3		788	2,71	19,41	
749	0,88	13,59		789	0,07	13,72	
750	2,31	16,65		790	1,09	13,52	
751	1,92	15,71		791	2,56	16,27	
752	2,77	18,46		792	2,37	15,13	
753	-0,69	9,26		793	3,92	19,83	
754	4,5	19,12		794	0,38	12,37	
755	2,64	17,34		795	2,29	16,15	
756	-0,19	10,89		796	5,96	24	
757	0,5	13,67		797	3,69	19,23	
758	0,57	13,92		798	4,16	20,24	
759	2,53	16,19		799	3,61	18,28	
760	2,87	17		800	3,19	18,41	

case no	dosage	performance	group	case no	dosage	performance	group
801	5,28	21,43		841	2,58	18,56	
802	1,26	14,46		842	3,68	19,13	
803	5,65	23,11		843	2	14,38	
804	3,13	17,98		844	2,65	16,16	
805	1,09	15,22		845	5,66	24,17	
806	1,52	15,8		846	4,18	20,47	
807	3,26	16,43		847	-0,32	11,29	
808	-0,92	8,77		848	2,07	18,3	
809	3,83	20,95		849	0,14	14,68	
810	1,15	13,52		850	3,22	18,83	
811	0,99	12,68		851	3,13	18,73	
812	2,75	17,09		852	2,37	17,1	
813	0,5	13,84		853	3,81	18,82	
814	2,48	16,07		854	0,99	12,7	
815	2,14	16,6		855	0,93	13,71	
816	4,51	22,13		856	3,51	17,43	
817	2,4	18,11		857	1,6	15,3	
818	3,39	20,49		858	3,66	20,22	
819	1,7	13,03		859	2,58	17,47	
820	1,42	14,99		860	4,04	20,1	
821	2,55	17,58		861	1,66	14,28	
822	-1,44	9,14		862	2,29	15,09	
823	-1,67	8,1		863	-0,09	11,32	
824	2,32	16,76		864	0,97	13,63	
825	3,49	17,69		865	1,85	15,19	
826	3,49	18,22		866	1,96	15,79	
827	1,93	15,12		867	3,36	18,41	
828	2,13	13,95		868	3,37	18,75	
829	3,61	19,47		869	0,91	13,17	
830	1,52	15,71		870	-0,06	12,35	
831	0,59	11,66		871	2,88	17,58	
832	-1,7	8,89		872	0,3	12,63	
833	1	13,84		873	3,14	19,13	
834	4,36	20,31		874	3,08	20,12	
835	3,22	19,63		875	1,3	15,53	
836	1,27	12,57		876	1,36	13,03	
837	0,43	12,9		877	2,91	17,21	
838	2,75	17,94		878	2,15	16,49	
839	-0,47	10,34		879	-0,54	11,87	
840	3,55	19,41		880	3,55	18,36	

case no	dosage	performance	group	case no	dosage	performance	group
881	1,33	13,71		921	-0,13	13,25	
882	-0,61	10,75		922	1,53	16,05	
883	0,54	10,65		923	1,95	15,38	
884	3,77	18,91		924	3,93	19,72	
885	1,28	14,7		925	2,04	15,81	
886	3,15	17,05		926	2	15,71	
887	2,32	16,81		927	1,05	15,33	
888	4,04	21,85		928	2,76	18,45	
889	4,24	20,32		929	1,53	15,75	
890	0,52	15,74		930	3,52	17,99	
891	-0,05	10,39		931	1,33	16,62	
892	3,07	19,29		932	3,53	17,36	
893	3,9	20,58		933	3,58	21	
894	0,8	15,07		934	2,37	16,64	
895	2,63	16,67		935	5,82	24,1	
896	1,65	15,1		936	1,53	14,85	
897	4,9	22,02		937	2,12	15,92	
898	3,25	18,93		938	3,7	19,62	
899	1,59	15,63		939	3,54	18,81	
900	1,71	14,68		940	2,95	16,61	
901	1,73	14,23		941	-0,13	11,63	
902	1,13	15,81		942	3,26	17,54	
903	1,23	13,71		943	0,92	14,5	
904	0,43	15,47		944	-1,04	8,23	
905	2,01	15,05		945	3,27	16,53	
906	1,12	14,47		946	1,43	15,16	
907	2,19	16,08		947	2,06	15,85	
908	3,15	18,66		948	2,69	16,28	
909	-0,3	11,47		949	1,72	14,92	
910	2,44	18,18		950	-1,24	11,95	
911	1,04	13,33		951	3,85	19,36	
912	-0,04	11,58		952	3,01	17,8	
913	-0,24	10,36		953	1,48	15,45	
914	5,22	22,36		954	2,49	17,44	
915	-0,44	10,55		955	0,4	13,13	
916	3,89	21,28		956	2,76	17,16	
917	0,72	12,4		957	-0,73	10,18	
918	3,26	18,58		958	-0,21	12,61	
919	1,32	14,2		959	2,64	18,04	
920	3,26	18,69		960	1,18	14,06	

case no	dosage	performance	group
961	6	23,73	
962	1,43	13,33	
963	3,96	20,49	
964	-0,54	9	
965	3,27	17,58	
966	3,03	18,95	
967	3,46	17,74	
968	3,93	19,47	
969	3,11	17,92	
970	2,69	17,01	
971	0,55	12,29	
972	5,32	23,96	
973	1,21	13,32	
974	3,07	17,99	
975	2,78	17,8	
976	1,5	15,66	
977	2,11	14,94	
978	0,94	15,37	
979	1,48	14,73	
980	3,67	19,96	
981	0,98	12,51	
982	2,68	17,44	
983	0,37	12,06	
984	-1,71	9,67	
985	3,49	18,88	
986	-0,63	10,82	
987	1,66	15,34	
988	2,93	18,18	
989	0,32	13,18	
990	1,34	15,02	
991	1,6	15,29	
992	1,35	14,32	
993	-0,52	9,04	
994	1,69	14,59	
995	2,49	15,16	
996	1,49	15,94	
997	2,93	19,39	
998	3,54	19,8	
999	1,42	13,67	
1000	7,28	26,62	

EK-3

Dozaj ve Performans 2 için yapay veri kümesi aşağıdaki gibidir. Bu veri kümesine, <http://www.amosdevelopment.com/video/index.htm> internet adresinden ulaşılabilir.

case no	dosage	performance	group	case no	dosage	performance	group
1	6,66	21,2		46	7,74	20,74	
2	5,66	15,7		47	5,87	17,33	
3	6,06	19,2		48	9,47	24,97	
4	9,19	23,13		49	5,99	16,66	
5	6,94	20,99		50	7,63	21,36	
6	5,16	18,04		51	5,78	18,24	
7	4,18	13,96		52	7,42	20,34	
8	8,08	22,52		53	8,37	22,77	
9	4,68	14,06		54	2,84	12,57	
10	7,86	19,55		55	8,53	22,16	
11	2,76	12,97		56	6,25	19,31	
12	3,7	11,96		57	6,01	16,46	
13	5,54	15		58	5,94	17,31	
14	7,06	20,14		59	5,57	18,05	
15	7,22	20,35		60	10,11	24,05	
16	9,84	25,74		61	7,51	19,8	
17	4,44	15,36		62	8,98	25,33	
18	8,63	23,68		63	10,89	28,23	
19	8,37	22,69		64	8,31	23,17	
20	6,93	19,47		65	7,72	22,19	
21	7,64	22,72		66	10,07	26,71	
22	5,66	18,18		67	7,01	19,75	
23	5,87	17,27		68	3,31	11,77	
24	5,37	15,81		69	9,13	24,71	
25	7,01	20,87		70	6,62	17,73	
26	7,27	21,56		71	5,96	17,96	
27	7,64	22,05		72	6,35	18,48	
28	8,15	22,98		73	9,65	25,26	
29	7,48	20,24		74	6,65	18,83	
30	6,84	19,96		75	9,17	22,07	
31	5,72	16,97		76	5,98	17,83	
32	6,44	18,99		77	4,84	16,01	
33	8,21	22,69		78	8,07	24,37	
34	6,99	20,25		79	6,35	20,33	
35	4,87	14,84		80	3,37	14,07	
36	5,57	17,77		81	3,44	13,67	
37	6,68	18,22		82	7,12	21,04	
38	7,69	21,82		83	8,39	22	
39	10,76	28,04		84	7,95	22,26	
40	5,8	17,07		85	5,11	15,76	
41	6,2	18,66		86	7,57	21,13	
42	6,92	19,84		87	9,34	25,53	
43	8,29	23,54		88	8,36	24,4	
44	7,68	21,96		89	7,91	21,65	
45	8,44	23,64		90	3,95	13,22	

case no	dosage	performance	group	case no	dosage	performance	group
91	8,13	21,31		141	-2,03	7,13	
92	9,6	26,05		142	2,1	12,89	
93	4,65	13,3		143	-0,48	7,89	
94	6,92	20,12		144	3,43	18,03	
95	7,57	23,84		145	3,69	17,23	
96	9,01	24,2		146	3,84	18,63	
97	7,38	19,89		147	0,78	10,43	
98	9,11	24,19		148	1	12,18	
99	8,31	22,96		149	6,3	22,7	
100	7,72	19,78		150	2,87	15,44	
101	1,02	11,47		151	3,18	15,24	
102	2,28	14,92		152	2,09	14,68	
103	2,4	16,09		153	-0,4	7,79	
104	2,06	12,76		154	0,66	10,99	
105	5,82	21,85		155	2,45	15,65	
106	2,52	13,57		156	2,95	16,77	
107	0,53	12,52		157	4,15	16,83	
108	0,49	10,67		158	-0,73	7,07	
109	2,41	16,09		159	-2,04	6,2	
110	-0,02	9,96		160	4,54	20,13	
111	-0,35	8,04		161	0,36	9,45	
112	-0,12	10,16		162	-0,23	9,6	
113	4,82	19,37		163	2,58	15,7	
114	5,23	20,34		164	2,66	16,79	
115	3,68	17,97		165	4,27	17,31	
116	1,87	13,94		166	1,54	14,14	
117	3,46	17,13		167	-0,42	10,22	
118	-0,22	8,99		168	1,86	13,29	
119	0,22	11,05		169	0,26	8,56	
120	2,25	13,78		170	1,1	13,19	
121	1,04	12,75		171	2,12	14,54	
122	2,59	15,2		172	1,43	12,71	
123	1,44	13,38		173	3,31	15,53	
124	1,22	11,17		174	4,72	19,18	
125	2,79	15,13		175	0,93	9,71	
126	2,21	14,28		176	0,55	11,78	
127	2,22	14,42		177	0,03	10,93	
128	2,07	13,31		178	4,54	20,21	
129	3,83	17,87		179	3,94	18,56	
130	3,63	15,81		180	0,57	11	
131	-1,68	6,33		181	1,99	12,84	
132	1,77	12,7		182	5,11	20	
133	3,58	17,98		183	0,45	10,6	
134	-0,47	7,11		184	2,86	15,24	
135	2,42	14,84		185	0,77	10,86	
136	4,67	18,12		186	7,35	25,7	
137	-1,56	7,74		187	4,76	18,41	
138	0,35	10,61		188	-1,71	5,68	
139	0,78	10,12		189	3,79	17,09	
140	3,15	16,43		190	2,88	15,48	

case no	dosage	performance	group	case no	dosage	performance	group
191	-1,28	5,45		241	3,15	15,38	
192	1,86	14,59		242	-0,68	9,09	
193	2,28	13,82		243	-0,31	9,65	
194	1,99	14,82		244	3,28	17,69	
195	0,82	11,13		245	1,74	14,54	
196	-0,81	9		246	2,01	15,25	
197	2,08	14,41		247	1,19	13,9	
198	0,49	12,14		248	2,22	14,3	
199	3,6	16,39		249	0,2	8,31	
200	1,2	12,97		250	0,09	10,29	
201	0,36	9,98		251	3,19	15,69	
202	3,48	16,01		252	2,07	15,92	
203	7,19	24,52		253	4,71	18,64	
204	1,17	11,26		254	3,06	15,46	
205	1,39	13,61		255	-1,06	8,89	
206	5,82	20,01		256	1,23	11,21	
207	0,5	10,53		257	3,03	15,57	
208	-0,75	9,31		258	-0,2	11,5	
209	4,67	18,51		259	6,17	23,42	
210	-0,48	8,75		260	1,34	12,74	
211	3,63	17,74		261	3,25	16,17	
212	3,19	16,57		262	1,4	13,12	
213	3,74	17,15		263	2,59	13,78	
214	0,29	11,43		264	2,4	14,23	
215	0,61	10,28		265	2,4	16,69	
216	1,41	13,74		266	-0,77	8,02	
217	4,74	18,86		267	4,83	20,46	
218	1,98	13,86		268	6,03	24,13	
219	-0,07	11,07		269	1,12	11,09	
220	2,88	14,69		270	-2,22	4,27	
221	2,7	15,31		271	1,01	11,64	
222	0,06	10,4		272	4,43	17	
223	3,9	16,96		273	0,55	10,31	
224	3,98	18,77		274	3,63	18,33	
225	2,89	16,71		275	2,91	15,7	
226	4,1	19,96		276	5,29	21,24	
227	1,28	11,87		277	3,51	16,82	
228	3,07	18,11		278	4,04	16,8	
229	2,72	14,31		279	-1,16	5,82	
230	5,86	21,29		280	2,12	14	
231	3,16	16,6		281	3,7	17,59	
232	2,85	15,87		282	2,85	16,87	
233	-0,17	11,88		283	4,07	18,26	
234	0,45	11,75		284	3,94	18,28	
235	0,67	12,57		285	-0,8	8,39	
236	4,81	18,21		286	-1,65	6,93	
237	2,21	14,89		287	-0,12	9,24	
238	3,99	18,99		288	5,38	20,31	
239	-0,35	9,99		289	4,15	18,36	
240	2,26	15,18		290	4,33	19,72	

case no	Dosg.	Perfm.	Grp.	case no	Dsg.	Perfm.	Grp.	case no	Dsg.	Perfm.	Grp.
291	-1,04	7,8		341	-0,13	12,02		391	4,92	19,69	
292	1,05	12,51		342	4,85	19,32		392	4,16	16,38	
293	2,18	13,92		343	4,16	19,9		393	3,44	16,15	
294	2,74	14,49		344	1,39	13,6		394	0,64	10,69	
295	-0,01	11,23		345	5,7	21,18		395	1,17	11,28	
296	2,68	17,35		346	-1,6	6,36		396	2,92	16,57	
297	0,48	9,97		347	5,99	21,99		397	0,28	10,86	
298	1,29	11,41		348	0,84	10,96		398	3,36	15,54	
299	1,72	13,2		349	4,25	20,48		399	3,31	16,1	
300	1,74	12,68		350	3,88	18,34		400	1,4	14,46	
301	2,08	14,98		351	4,02	18,86					
302	1,19	11,28		352	1,08	10,78					
303	2,54	15,59		353	2,24	13,28					
304	0,88	10,91		354	0,62	10,27					
305	3,23	17,37		355	2,58	15,49					
306	3,67	15,79		356	4,18	16,9					
307	1,22	11,75		357	2,04	13,31					
308	2,96	17,22		358	2,44	15,63					
309	3,53	16,96		359	3,36	16,03					
310	-1,57	6,14		360	4,65	20,39					
311	3,97	17,68		361	2,38	15,3					
312	2,27	14,75		362	1,42	14,57					
313	0,79	11,34		363	0,36	10,75					
314	2,33	14,44		364	-1,87	6,96					
315	1,42	12,98		365	0,72	11,56					
316	3,4	17,25		366	3,56	16,83					
317	3,91	19,59		367	1,04	14,28					
318	2,83	14,09		368	1,09	12,98					
319	3,91	18,98		369	4,02	17,81					
320	2,21	14,64		370	-0,49	8,66					
321	3,07	16,61		371	4,3	17,54					
322	2,54	13,51		372	2,73	15,38					
323	3,96	16,7		373	1,52	13,41					
324	2,96	14,94		374	3,4	15,89					
325	2,34	15,01		375	2,35	12,91					
326	0,71	10,5		376	0,44	12,31					
327	0,31	11,2		377	3,28	16,63					
328	1,62	14,02		378	3,58	18,93					
329	2,22	14,82		379	3,19	14,64					
330	2,58	16,11		380	1,53	14,47					
331	0,02	10,5		381	0,96	11,96					
332	1,94	14,62		382	0,94	11,26					
333	3,17	14,81		383	-1,8	5,84					
334	2,92	15,36		384	0,74	11,21					
335	-0,31	10,67		385	3,28	16,58					
336	1,16	11,26		386	3,65	17,76					
337	3,18	17,87		387	1,8	12,65					
338	1,44	14,37		388	2,03	14,14					
339	2,22	15,58		389	2,93	14,29					
340	1,48	12,99		390	2,3	14,25					

EK-4

Attig 'in 1983 yılında oluşturduğu hafıza verisi aşağıdaki gibidir. Bu veriye, Amos paket programı yüklenmesinin ardın, C:\Program Files\ Amos 16.0\ Examples patikasının takibi ile ulaşılmıştır.

subject	age	v_short	vocab	education	sex	recall 1	recall 2	cued 1	cued 2	place 1	place 2
1	65	12	72	16	1	5	11	5	11	30	32
2	68	14	77	18	0	12	16	14	16	35	32
3	64	14	74	17	1	11	11	10	11	33	30
4	77	13	74	16	0	3	3	3	4	31	26
5	72	6	46	12	0	8	9	11	9	27	34
6	75	14	77	12	1	10	9	10	10	28	24
7	69	7	49	12	0	8	7	10	8	29	30
8	74	12	61	12	0	7	6	8	9	30	28
9	66	11	72	12	0	8	12	9	13	28	39
10	77	11	64	12	0	8	11	10	13	32	35
11	72	11	76	20	0	11	13	12	14	28	34
12	69	13	74	13	0	10	14	12	14	33	27
13	68	12	76	16	1	6	3	7	3	24	27
14	68	13	69	12	0	8	10	11	10	30	35
15	74	6	46	4	0	11	7	10	8	32	32
16	68	8	51	12	0	5	9	7	10	29	32
17	70	9	51	12	0	8	9	9	10	27	25
18	69	14	76	20	1	13	13	13	13	29	30
19	71	9	63	12	0	10	13	10	13	23	30
20	79	10	54	8	0	11	6	13	9	32	30
21	74	10	67	16	1	8	3	7	5	27	25
22	70	13	70	12	0	7	8	9	10	29	29
23	71	4	50	14	0	4	7	3	10	27	34
24	72	12	72	20	1	10	12	12	11	31	32
25	81	8	58	10	0	5	2	9	2	27	28
26	69	6	47	12	0	8	6	13	6	29	29
27	68	8	55	14	0	8	6	9	6	28	24
28	68	9	56	12	1	8	11	8	11	27	25
29	64	11	60	12	0	12	11	13	13	29	34
30	70	8	50	12	0	11	7	13	6	24	31
31	77	8	52	12	0	12	11	8	12	28	25
32	73	12	66	11	0	11	6	12	7	32	32
33	76	7	56	8	0	9	5	9	7	23	24
34	74	9	54	8	0	8	2	7	4	34	27
35	72	4	47	5	0	7	7	9	8	30	24
36	67	0	40	6	0	7	4	7	2	34	22
37	83	12	77	12	0	10	9	10	8	33	32
38	70	14	80	14	0	8	13	8	13	32	31
39	64	14	78	13	0	12	13	12	14	26	31
40	63	13	66	12	0	9	10	11	11	29	31

EK-5

Dozaj ve Performans 2 veri kümesinin bütün gözlemler için AMOS paket programı kullanılarak hesaplanan grup üye olasılıkları aşağıdaki gibidir:

case no	performance	dosage	P(Cluster1)	P(Cluster2)
1	21,2031285	6,66331008	0,88	0,12
2	15,69876749	5,66102566	1	0
3	19,20483612	6,05921131	0,98	0,02
4	23,1294141	9,18741643	1	0
5	20,9890517	6,94140644	1	0
6	18,04305616	5,16054294	0,55	0,45
7	13,96043061	4,17650112	1	0
8	22,51519752	8,07913551	1	0
9	14,06167198	4,67521628	1	0
10	19,54608638	7,85787685	1	0
11	12,96792854	2,76272553	0,07	0,93
12	11,95872658	3,69959321	1	0
13	15,00371011	5,53950501	1	0
14	20,13502904	7,06075642	1	0
15	20,34850963	7,21701606	1	0
16	25,73694781	9,84016965	1	0
17	15,36027096	4,44159268	0,98	0,02
18	23,67784702	8,63318871	1	0
19	22,69123701	8,37492785	1	0
20	19,47418817	6,92518967	1	0
21	22,71962936	7,63502745	0,99	0,01
22	18,17964285	5,66055867	0,99	0,01
23	17,2726584	5,87157586	1	0
24	15,81401787	5,36913162	1	0
25	20,86509889	7,01435565	1	0
26	21,5553809	7,26916463	1	0
27	22,04621812	7,63868845	1	0
28	22,98445144	8,14525079	1	0
29	20,23981056	7,48371461	1	0
30	19,96332394	6,83913761	1	0
31	16,96709433	5,72284811	1	0
32	18,98584583	6,4391513	1	0
33	22,68516739	8,20789759	1	0
34	20,25002356	6,98892576	1	0
35	14,8357132	4,87241527	1	0
36	17,77108173	5,56859338	0,99	0,01
37	18,21569197	6,67660807	1	0
38	21,81732184	7,68515104	1	0
39	28,0374376	10,7628896	1	0
40	17,07205681	5,8020707	1	0
41	18,65947766	6,19835278	1	0
42	19,83926183	6,91636132	1	0
43	23,5404642	8,28670057	1	0
44	21,95765439	7,67969081	1	0
45	23,63920299	8,4449591	1	0
46	20,74091102	7,74354449	1	0
47	17,33331536	5,86545434	1	0
48	24,96692214	9,46816015	1	0
49	16,65560059	5,99459445	1	0
50	21,35760016	7,63080733	1	0

case no	performance	dosage	P(Cluster1)	P(Cluster2)
51	18,23753453	5,77616224	0,99	0,01
52	20,33838128	7,41631772	1	0
53	22,76866003	8,37343747	1	0
54	12,5657738	2,84127905	0,41	0,59
55	22,1558098	8,53096868	1	0
56	19,3093234	6,25446357	1	0
57	16,46041316	6,01073762	1	0
58	17,31493142	5,94042213	1	0
59	18,04674528	5,57439467	0,98	0,02
60	24,04772823	10,106291	1	0
61	19,79730838	7,51239372	1	0
62	25,33071912	8,97904828	1	0
63	28,22981354	10,8850001	1	0
64	23,16607707	8,31273504	1	0
65	22,19338093	7,72460966	1	0
66	26,70589699	10,0737469	1	0
67	19,75453839	7,01139035	1	0
68	11,77257081	3,30915273	1	0
69	24,70563597	9,12929352	1	0
70	17,72769762	6,61550565	1	0
71	17,96072897	5,96089648	1	0
72	18,47790795	6,3509418	1	0
73	25,25566245	9,64792522	1	0
74	18,82666493	6,65415272	1	0
75	22,06794037	9,17451817	1	0
76	17,83447619	5,97893792	1	0
77	16,0088857	4,84431705	0,99	0,01
78	24,36544929	8,06669967	0,94	0,06
79	20,32556489	6,34875861	0,92	0,08
80	14,06638162	3,37110477	0,24	0,76
81	13,6731876	3,44382298	0,72	0,28
82	21,04383952	7,12210031	1	0
83	22,00410477	8,38699414	1	0
84	22,25989015	7,94861216	1	0
85	15,76161005	5,11126999	1	0
86	21,12529151	7,56984011	1	0
87	25,53497856	9,34146177	1	0
88	24,4009606	8,3591885	0,99	0,01
89	21,64748894	7,91473758	1	0
90	13,22240473	3,9510702	1	0
91	21,30954567	8,12689862	1	0
92	26,05489073	9,60215607	1	0
93	13,30314789	4,65192225	1	0
94	20,11844652	6,91923731	1	0
95	23,83712159	7,56509277	0,58	0,42
96	24,20480782	9,00656718	1	0
97	19,88854588	7,38410165	1	0
98	24,19259422	9,11180087	1	0
99	22,95893412	8,31184723	1	0
100	19,77788833	7,71925889	1	0
101	11,46928755	1,02279085	0	1
102	14,92348099	2,27966518	0	1
103	16,08950753	2,39526231	0	1
104	12,75622934	2,0587308	0	1
105	21,85058999	5,8173389	0	1
106	13,57160048	2,52183848	0	1
107	12,51797552	0,53148318	0	1
108	10,67015185	0,48621623	0	1
109	16,08838189	2,41245341	0	1
110	9,958635793	-0,022373	0	1

case no	performance	dosage	P(Cluster1)	P(Cluster2)
111	8,043151815	-0,3470517	0	1
112	10,16497757	-0,1154461	0	1
113	19,36751488	4,8238406	0	1
114	20,33501096	5,22620099	0	1
115	17,96523335	3,67837765	0	1
116	13,94196299	1,86728085	0	1
117	17,12763362	3,46444656	0	1
118	8,993002598	-0,2202774	0	1
119	11,05232233	0,2200883	0	1
120	13,78234393	2,25485454	0	1
121	12,75095859	1,04299832	0	1
122	15,20276939	2,59470972	0	1
123	13,38483808	1,43849658	0	1
124	11,16703078	1,22019228	0	1
125	15,1303717	2,79441923	0	1
126	14,28435999	2,21376115	0	1
127	14,42290917	2,2186285	0	1
128	13,31315542	2,07314355	0	1
129	17,86726133	3,82613148	0	1
130	15,8098483	3,63134068	0	1
131	6,333905719	-1,6845652	0	1
132	12,70353223	1,77148886	0	1
133	17,97693157	3,58100498	0	1
134	7,111584668	-0,4651816	0	1
135	14,84203527	2,42238147	0	1
136	18,12181875	4,6697551	0,01	0,99
137	7,740990133	-1,5572788	0	1
138	10,60999846	0,34584624	0	1
139	10,12142508	0,77663488	0	1
140	16,43000919	3,15272054	0	1
141	7,131369131	-2,0342289	0	1
142	12,89139993	2,09537056	0	1
143	7,887249519	-0,4839839	0	1
144	18,03395001	3,42624108	0	1
145	17,22845236	3,69420091	0	1
146	18,6314848	3,83685924	0	1
147	10,43413758	0,77648276	0	1
148	12,17986123	1,00151297	0	1
149	22,69540012	6,303015	0	1
150	15,44428227	2,87099168	0	1
151	15,24085167	3,18364834	0	1
152	14,68385365	2,09402589	0	1
153	7,789584217	-0,4008463	0	1
154	10,99110192	0,6594106	0	1
155	15,64610017	2,45315713	0	1
156	16,76873028	2,94750396	0	1
157	16,82719323	4,1517344	0,01	0,99
158	7,065834555	-0,7289824	0	1
159	6,195856888	-2,0360787	0	1
160	20,12737294	4,53821959	0	1
161	9,449677699	0,36450828	0	1
162	9,604363476	-0,2335994	0	1
163	15,70224399	2,57750508	0	1
164	16,79477836	2,65897495	0	1
165	17,30678911	4,27370102	0,01	0,99
166	14,14156325	1,53807151	0	1
167	10,22019023	-0,4169918	0	1
168	13,28634093	1,85849547	0	1
169	8,556666886	0,25893971	0	1
170	13,18782194	1,09649637	0	1

case no	performance	dosage	P(Cluster1)	P(Cluster2)
171	14,5365028	2,11780038	0	1
172	12,70923491	1,42820659	0	1
173	15,52857136	3,30899595	0	1
174	19,17817418	4,72355719	0	1
175	9,71265477	0,92741466	0	1
176	11,78022915	0,55443418	0	1
177	10,92695493	0,02854628	0	1
178	20,21385147	4,54025237	0	1
179	18,55696664	3,94024425	0	1
180	11,00267739	0,56678258	0	1
181	12,83624704	1,98839536	0	1
182	19,99611586	5,11186379	0	1
183	10,60421622	0,44835749	0	1
184	15,24134398	2,8616715	0	1
185	10,85675615	0,77475969	0	1
186	25,70332331	7,3520501	0	1
187	18,4064039	4,76414353	0,01	0,99
188	5,680894638	-1,7075451	0	1
189	17,09467063	3,78890575	0	1
190	15,48225167	2,88104635	0	1
191	5,445550321	-1,2763698	0	1
192	14,58512409	1,86287389	0	1
193	13,82339721	2,27991013	0	1
194	14,81746949	1,99401771	0	1
195	11,13360367	0,81707599	0	1
196	8,996885903	-0,8113705	0	1
197	14,40781012	2,08140113	0	1
198	12,13773057	0,4910429	0	1
199	16,39254275	3,59898512	0	1
200	12,96866703	1,19704669	0	1
201	9,97642722	0,36464742	0	1
202	16,00687798	3,47979186	0	1
203	24,51927762	7,19396491	0,01	0,99
204	11,26217441	1,17493952	0	1
205	13,61092103	1,38707381	0	1
206	20,0079943	5,81985558	0,25	0,75
207	10,53399996	0,49698532	0	1
208	9,30530393	-0,7463701	0	1
209	18,51267739	4,66706419	0	1
210	8,74704322	-0,479564	0	1
211	17,74165875	3,63439825	0	1
212	16,57437236	3,18752262	0	1
213	17,15126348	3,74155979	0	1
214	11,43358207	0,28787723	0	1
215	10,27852562	0,60606421	0	1
216	13,74047171	1,40865014	0	1
217	18,85610457	4,74096663	0	1
218	13,86416686	1,97874226	0	1
219	11,0723381	-0,0715693	0	1
220	14,69254406	2,8787283	0	1
221	15,31174266	2,69578091	0	1
222	10,40450131	0,05945702	0	1
223	16,95594412	3,89910615	0	1
224	18,76979471	3,97547166	0	1
225	16,70904765	2,89189206	0	1
226	19,95576974	4,09622331	0	1
227	11,87300847	1,28163164	0	1
228	18,11221145	3,06816966	0	1
229	14,30967612	2,71717544	0	1
230	21,2892331	5,86289794	0	1

case no	performance	dosage	P(Cluster1)	P(Cluster2)
231	16,59621234	3,15902533	0	1
232	15,86583847	2,85089798	0	1
233	11,87508988	-0,1700721	0	1
234	11,75472364	0,45348801	0	1
235	12,57144248	0,67379573	0	1
236	18,21042106	4,80909276	0,03	0,97
237	14,89416726	2,20677663	0	1
238	18,99497832	3,98557475	0	1
239	9,988090885	-0,3456018	0	1
240	15,18191172	2,2587903	0	1
241	15,38135416	3,15047091	0	1
242	9,089998229	-0,682274	0	1
243	9,645792429	-0,3081196	0	1
244	17,68909002	3,28356488	0	1
245	14,54026993	1,74200946	0	1
246	15,24744913	2,00745917	0	1
247	13,90082289	1,19384311	0	1
248	14,30448482	2,21601524	0	1
249	8,307376608	0,20140242	0	1
250	10,28985672	0,09283402	0	1
251	15,68787838	3,19292289	0	1
252	15,91540518	2,06979203	0	1
253	18,64039303	4,71001079	0	1
254	15,45589028	3,05694154	0	1
255	8,886297727	-1,064744	0	1
256	11,20523562	1,22642794	0	1
257	15,57189915	3,02540953	0	1
258	11,50118597	-0,2029642	0	1
259	23,41969949	6,16710504	0	1
260	12,74230766	1,3400863	0	1
261	16,1707638	3,24636513	0	1
262	13,11884283	1,39594781	0	1
263	13,77790255	2,58929235	0	1
264	14,22667782	2,40251898	0	1
265	16,68727409	2,39668349	0	1
266	8,017057363	-0,7729234	0	1
267	20,45958439	4,82698005	0	1
268	24,13253834	6,03103952	0	1
269	11,09182318	1,11959821	0	1
270	4,267297139	-2,2240314	0	1
271	11,63998152	1,00936401	0	1
272	17,00101335	4,43462977	0,08	0,92
273	10,31348781	0,54952773	0	1
274	18,33291585	3,62694436	0	1
275	15,70198626	2,91056838	0	1
276	21,24319525	5,28748122	0	1
277	16,81929663	3,50811545	0	1
278	16,79742093	4,03776595	0	1
279	5,823605566	-1,1580606	0	1
280	13,99611422	2,12343709	0	1
281	17,58700828	3,69844358	0	1
282	16,86768774	2,84983301	0	1
283	18,25630335	4,07244795	0	1
284	18,27915732	3,94219111	0	1
285	8,39427921	-0,8037862	0	1
286	6,928338894	-1,6527581	0	1
287	9,243000658	-0,1190782	0	1
288	20,30946755	5,38149082	0	1
289	18,35966975	4,15427404	0	1
290	19,71540132	4,32616316	0	1

case no	performance	dosage	P(Cluster1)	P(Cluster2)
291	7,797007316	-1,0404931	0	1
292	12,51410631	1,05249423	0	1
293	13,91692844	2,17627897	0	1
294	14,48964157	2,74304818	0	1
295	11,22805041	-0,0106664	0	1
296	17,34523583	2,67584652	0	1
297	9,967782034	0,48245367	0	1
298	11,40856022	1,29398649	0	1
299	13,20351686	1,71578459	0	1
300	12,67929136	1,74067544	0	1
301	14,97576496	2,07975424	0	1
302	11,27716092	1,1875313	0	1
303	15,58647697	2,53867253	0	1
304	10,91026957	0,88460716	0	1
305	17,3689097	3,22794244	0	1
306	15,79319522	3,66735014	0,01	0,99
307	11,74811694	1,21939767	0	1
308	17,22432356	2,95889254	0	1
309	16,96227064	3,52621809	0	1
310	6,141804553	-1,5671512	0	1
311	17,67774633	3,97433797	0	1
312	14,746975	2,2736652	0	1
313	11,34147331	0,79433798	0	1
314	14,435088	2,32796991	0	1
315	12,9801137	1,41781099	0	1
316	17,24554045	3,3958935	0	1
317	19,58925903	3,91227199	0	1
318	14,0927476	2,82787906	0	1
319	18,98062882	3,90883952	0	1
320	14,64437678	2,21211446	0	1
321	16,61185826	3,07036131	0	1
322	13,51463892	2,53643425	0	1
323	16,70187812	3,95595147	0	1
324	14,93663508	2,96105939	0	1
325	15,01010648	2,33502175	0	1
326	10,49542327	0,70998146	0	1
327	11,19843897	0,30808968	0	1
328	14,02099667	1,61771328	0	1
329	14,81639334	2,22084526	0	1
330	16,10807499	2,58149152	0	1
331	10,49817401	0,01731795	0	1
332	14,61936538	1,93656651	0	1
333	14,80890048	3,17428906	0	1
334	15,35960661	2,91562256	0	1
335	10,66728944	-0,3101875	0	1
336	11,26345394	1,16147515	0	1
337	17,87453191	3,18022489	0	1
338	14,36956945	1,43875227	0	1
339	15,58383136	2,21771876	0	1
340	12,99454128	1,48105691	0	1
341	12,01675006	-0,131517	0	1
342	19,3215126	4,84829956	0	1
343	19,89982687	4,16166493	0	1
344	13,59665974	1,39307863	0	1
345	21,18434946	5,69794501	0	1
346	6,357536716	-1,6003217	0	1
347	21,99274361	5,99086467	0	1
348	10,95721867	0,83567142	0	1
349	20,4826519	4,25124993	0	1
350	18,33886419	3,87676752	0	1

case no	performance	dosage	P(Cluster1)	P(Cluster2)
351	18,8558468	4,01548918	0	1
352	10,77816499	1,08409151	0	1
353	13,28174982	2,24232414	0	1
354	10,26728416	0,62131667	0	1
355	15,49204404	2,57505913	0	1
356	16,90317254	4,17976588	0,01	0,99
357	13,30870865	2,04353322	0	1
358	15,63179437	2,43540471	0	1
359	16,02513948	3,3616357	0	1
360	20,39381813	4,65422699	0	1
361	15,30391157	2,38419685	0	1
362	14,57124521	1,42345947	0	1
363	10,74837707	0,35987486	0	1
364	6,95774937	-1,8666786	0	1
365	11,55731265	0,71663825	0	1
366	16,82620044	3,55771246	0	1
367	14,27649592	1,04145176	0	1
368	12,97900197	1,0922214	0	1
369	17,81267067	4,01649738	0	1
370	8,6620644	-0,4924933	0	1
371	17,53516448	4,30110366	0	1
372	15,37611255	2,7265681	0	1
373	13,41075845	1,52145402	0	1
374	15,888386	3,39997388	0	1
375	12,91460654	2,34570591	0	1
376	12,31481653	0,43798288	0	1
377	16,62972356	3,27919796	0	1
378	18,92540573	3,57754642	0	1
379	14,64193936	3,19337685	0,01	0,99
380	14,47358503	1,52850312	0	1
381	11,96300862	0,96077911	0	1
382	11,25798218	0,94348921	0	1
383	5,841064645	-1,8027218	0	1
384	11,21193196	0,74473895	0	1
385	16,57683736	3,28125737	0	1
386	17,75957245	3,65294855	0	1
387	12,64712744	1,80191618	0	1
388	14,14111475	2,02988517	0	1
389	14,28802177	2,92660168	0	1
390	14,24994697	2,29648253	0	1
391	19,69188794	4,92316404	0	1
392	16,38457477	4,16164536	0,07	0,93
393	16,14541106	3,44236184	0	1
394	10,68870927	0,63909287	0	1
395	11,2751731	1,16561487	0	1
396	16,57490107	2,92214981	0	1
397	10,85610043	0,27853662	0	1
398	15,54247639	3,36159594	0	1
399	16,09565031	3,30674672	0	1
400	14,4601617	1,39569593	0	1

EK-6

Karma Oran Hesabı

Dozaj ve Performans 2 verisine ait olan grupların, karma oran hesaplarında:

1.grubun karma oranını hesaplarken, AMOS'un Posterior Predictive Penceresinde yer alan toplam 400 durumun gruplara atanmış halleri arasından 97'sinin, 1.gruba ait gözlemler olduğu tespit edilebilir.

Buna göre karma oran;

$$P_1 = \frac{97}{400} = 0.2425 \text{ dir.}$$

Bu değer, yaklaşık olarak Bayesian SEM penceresinin (Tablo 4.18) proportion satırındaki değer ile aynıdır.

Aynı mantıkla;

$$P_2 = \frac{303}{400} = 0.757$$

2.grup için karma oran hesabı yapılabilir.

Aynı şekilde; 2. grup için elde edilen bu oranın da, Bayesian SEM penceresinin (Tablo 4.19) proportion satırındaki değeri ile yaklaşık olarak aynı olduğu görülebilmektedir.