

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANABİLİM DALI

**SAYIM İLE ELDE EDİLEN BAĞIMLI DEĞİŞKENDE GÖZLENEMEYEN
HETEROJENLİĞİN LATENT SINIF POISSON REGRESYONU
KULLANILARAK MODELLENMESİ**

YÜKSEK LİSANS TEZİ

Nur AYCAN
Danışman: Prof Dr. Abdullah YEŞİLOVA

VAN – 2023

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANABİLİM DALI

**SAYIM İLE ELDE EDİLEN BAĞIMLI DEĞİŞKENDE GÖZLENEMEYEN
HETEROJENLİĞİN LATENT SINIF POISSON REGRESYONU
KULLANILARAK MODELLENMESİ**

YÜKSEK LİSANS TEZİ

Nur AYCAN

VAN – 2023

KABUL VE ONAY SAYFASI

Van Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Zootekni Anabilim Dalı'nda Prof. Dr. Abdullah Yeşilova danışmanlığında, Nur AYCAN tarafından sunulan “Sayım ile Elde Edilen Bağımlı Değişkende Gözlenemeyen Heterojenliğin Latent Sınıf Poisson Regresyonu Kullanılarak Modellenmesi” isimli bu çalışma Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili hükümleri gereğince/...../..... tarihinde aşağıdaki jüri tarafından oy birliği / oy çokluğu ile başarılı bulunmuş ve Yüksek lisans tezi olarak kabul edilmiştir.

Başkan: Prof Dr Abdullah Yeşilova

İmza:

Üye: Doç Dr Gazel Ser

İmza:

Üye: Doç Dr Yılmaz Kaya

İmza:

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun /.... /..... tarih ve sayılı kararı ile onaylanmıştır.

İmza

Enstitü Müdürü

ETİK BEYAN

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

İmza

Nur AYCAN



ÖZET

SAYIM İLE ELDE EDİLEN BAĞIMLI DEĞİŞKENDE GÖZLENEMEYEN HETEROJENLİĞİN LATENT SINIF POISSON REGRESYONU KULLANILARAK MODELLENMESİ

AYCAN, Nur

Yüksek Lisans Tezi, Zootečni Anabilim Dalı

Danışman: Prof. Dr. Abdullah Yeşilova

Mayıs 2023, 63 sayfa

Latent sınıf modeller, karmaşık bir veri setini kavramsal olarak yorumlanması kolay bir formatta sunmaktır. Bu nedenle son yıllarda buna yönelik sınıflandırma amaçlı istatistiksel yöntemler üzerinde durulmaktadır. Latent Poisson model yaklaşımı bu yönü ile diğer yöntemlere nazaran istatistiksel yorumlamada önemli avantajlar sağlamaktadır. Bunlar; gözlemlerin doğru sınıflandırma olasılığını sağlamak, elde edilen her alt populasyon için ayrı parametre tahminini yapmak, sınıflandırmadaki hata payını en aza indirmektir. Başka bir deyişle hem sınıflandırma yapması hem de elde edilen her bir populasyon için parametre tahmini yapabilmesidir. AIC, BIC ve entropy uyum istatistiklerine göre veri kümesinin dört alt populasyonda toplandığı tespit edilmiştir. Burada, farklı olarak faktörler, zamana bağlı değişen ve zamana bağlı değişmeyen (risk faktörleri) olmak üzere iki farklı biçimde modele dahil edilmiştir. Her bir alt populasyonun doğru sınıflandırma olasılıkları oldukça yüksek çıkmıştır. Latent Poisson regresyonu uygulamadan önce Kruskal-Wallis testi kullanılarak, akar sayımları bakımından dört çeşit ortalaması karşılaştırılmıştır. Daha sonra, veri setinin heterojen bir yapı sergilediği varsayılarak, latent Poisson regresyonu uygulanmıştır. Bu regresyon sonucunda akar sayımlarına göre dört alt populasyonlu model kabul edilmiştir. Latent Poisson regresyonun sonuçlarının daha iyi olduğu belirlenmiştir.

Anahtar kelimeler: Doğrusal olmayan regresyon, Gözlenemeyen heterojenlik, Sayım verisi

ABSTRACT

MODELING UNOBSERVABLE HETEROGENEITY USING LATENT CLASS POISSON REGRESSION IN THE DEPENDENT VARIABLE OBTAINED BY COUNTING

AYCAN, Nur

M.Sc. Thesis, Department of Animal Science

Supervisor: Prof. Dr. Abdullah YEŞİLOVA

May 2023, 63 pages

Latent class models are conceptually presenting a complex data set in a format that is easy to interpret. For this reason, statistical methods for classification purposes have been emphasized in recent years. With this aspect, the latent Poisson model approach provides significant advantages in statistical interpretation compared to other methods. These; To ensure the correct classification probability of the observations, to make a separate parameter estimation for each subpopulation obtained, and to minimize the margin of error in the classification. In other words, it makes both classification and parameter estimation for each population obtained. It has been determined that the set of tasks to AIC, BIC, and entropy compliance statistics are collected in four subpopulations. Here, factors are included in the model in two different ways as time-varying and time-invariant (risk factors). The probability of correct classification of each subpopulation was relatively high. Here, before applying the latent Poisson regression, the Kruskal-Wallis test was used to compare the mean of 4 varieties in terms of mite counts. Then, latent Poisson regression was applied, assuming the data set exhibits a heterogeneous structure. As a result of this regression, a model with four subpopulations was accepted according to mite counts. It was determined that the latent Poisson regression results were remarkable.

Keywords: Count data, Nonlinear analysis, Unobservable heterogeneity



TEŞEKKÜR

Tez çalışmamın her aşamasında bilgi ve deneyimiyle bana daima yol gösteren, desteğini hep hissettiğim ve beni daima çalışmaya teşvik eden değerli danışmanım Sayın Prof. Dr. Abdullah YEŞİLOVA başta olmak üzere eğitimim süresince sorularıma içtenlikle yanıt veren, ufkumu genişleten Sayın Doç. Dr. Gazel SER'e, yoğun iş temposu arasında desteği, motivasyonu, özveri ve sabrı için eşim Doç. Dr. Abdurrahman AYCAN, sevgili çocuklarım Belçim ve İbrahim Bengi'ye ile arkadaşlarıma sonsuz teşekkürlerimi sunarım.

2023

Nur AYCAN



İÇİNDEKİLER

Sayfa

ÖZET	i
ABSTRACT	iii
TEŞEKKÜR	v
İÇİNDEKİLER	vii
ÇİZELGELER LİSTESİ	ix
ŞEKİLLER LİSTESİ	xi
SİMGELER VE KISALTMALAR	xiii
EKLER DİZİNİ	xv
1. GİRİŞ	1
2. KAYNAK BİLDİRİŞLERİ	5
3. MATERYAL VE YÖNTEM	13
3.1 Materyal	13
3.2 Yöntem	13
3.2.1 Latent Modelin Parametrik Gösterimi	13
3.2.2 Latent Modellerde Parametre Tahmini	15
3.2.3 Karışıklı Modelde Alt Grup Sayısının Test Edilmesi	16
3.2.4 Karışıklı Modellerde En Çok Olabilirlik Yöntemi	16
3.2.5 Bilgi Matrisleri	17
3.2.6 Maksimum Olabilirlik Tahmininin Asimptotik Kovaryans Matrisi	18
3.2.7 Karışıklı Modeller için EM Algoritması	18
3.2.7.1 Doğrudan Yaklaşım	18
3.2.7.2 E Aşaması	19
3.2.7.3 M Aşaması	21
3.2.8 Model Seçiminde Uyum Ölçütleri	22
3.2.8.1 Log Algoritması	23
3.2.8.2 Akaike Bilgi Ölçütü	25
3.2.8.3 Bilgi Karmaşıklığı Kriteri	25
3.2.8.4. Bayesian Taban Uyum Ölçütleri	27

3.2.8.4.1 Bayesian Yaklaşım.....	27
3.2.8.4.2 Bayesian Bilgi Kriteri.....	27
3.2.9 Latent Sınıf Poisson Regresyon.....	28
3.2.10 Poisson ve Karışıklı Poisson Dağılımları.....	33
5. BULGULAR.....	35
6. TARTIŞMA VE SONUÇ	43
KAYNAKLAR.....	47
EKLER	51
ÖZ GEÇMİŞ.....	63



ÇİZELGELER LİSTESİ

Çizelge 3.1 Üstel fonksiyon ailesinde bazı dağılımlar	32
Çizelge 4.1 Uygun latent Poisson regresyonu için uyum ölçütleri	35
Çizelge 4.2 Bağımlı değişkenin alt populasyonlara göre dağılımı	36
Çizelge 4.3 Alt gruplara ait parametre tahminleri ve standart hataları.....	36
Çizelge 4.4 Altı alt populasyon modelde her bir alt populasyon için tahmin edilen entropy doğru sınıflandırma oranlar	37
Çizelge 4.5 Değişkenlerin alt gruplara göre ortalamaları.....	37
Çizelge 4.6 Kruskall- Wallis testi sonuçları	38
Çizelge 4.7 Dwass, Steel, Critchlow-Fligner Method çoklu karşılaştırma sonuçları.....	38
Çizelge 4.8 Latent sınıflara göre Kruskall- Wallis testi sonuçları.....	40
Çizelge 4.9 Dwass, Steel, Critchlow-Fligner Method çoklu karşılaştırma sonuçları	41

ŞEKİLLER LİSTESİ

Şekil 3.1 Karışım dağılımları konusundaki akademik yazıların yıllara göre değişimi ...	15
Şekil 3.2 Rasgele başlangıç değerleri kullanılarak EM algoritması için farklı k değerleri için üretilen parametrelere dayalı her alt grup için asimptotik (%95) elipsoidler grafiği	22
Şekil 4.1 Ortalama akar skor değerlerine göre çeşitlerin box-plot grafiği	39
Şekil 4.2 Her bir çeşidin medyan değerlerine göre dağılımı	39
Şekil 4.3 Akar sayımlarının çeşitlere göre dağılım	40
Şekil 4.4 Latent sınıflara göre, akar sayımlarının medyan değerlerine göre elde edilen dağılım	41
Şekil 4.5 Ortalama akar skorlarına göre box-plot dağılımı	42
Şekil 4.6 Çeşitlere göre akar skorlarının yoğunluk dağılımı	42



SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamalarıyla aşağıda sunulmuştur.

Simgeler

Açıklama

log

Logaritma

Kısaltmalar

Açıklama

AIC

Akaike Bilgi Ölçütü

BIC

Bayesian Bilgi Ölçütü

EM

Beklenti Maksimizasyonu

FCML

Bulanık Sınıflandırma Maksimum Olabilirlik

GLM

Genelleştirilmiş Doğrusal Modeller

GLMM

Genelleştirilmiş Karışık Doğrusal Model

LPM

Latent Poisson Model

ML

En Çok Olabilirlik Yöntemi

MPR

Karışıklı Poisson regresyonu

QL

Quasi Olabilirlik

QN

Olası Newton

DSCF

Dwass, Steel, Critchlow-Fligner Method



EKLER DİZİNİ

	Sayfa
Ek 1. Akar sayımlarının alt populasyonlara göre karışma olasılıkları	51
Ek 2. M-plus analiz sonuçları	56





1. GİRİŞ

Regresyon yöntemleri bağımlı değişkene göre doğrusal ve doğrusal olmayan olarak ikiye ayrılır. Doğrusal modellerde gözlem değerleri normal dağılım gösterir ve ölçüm ya da tartım ile elde edilebilir. Doğrusal olmayan modellerde ise normal dağılım göstermeyen sayım ile elde edilen gözlem değerleri olup Poisson dağılımı gösterir.

Poisson regresyon modeli, genelleştirilmiş doğrusal modeller içinde yer alan sürekli bir zaman aralığında meydana gelen sayım değerlerine uygulanmaktadır. Literatürde birçok tıp, sosyoloji, biyoistatistik ve aktüelya bilim alanlarında yapılan çalışmalarda bağımlı değişken kesikli yani sayıma dayalı olarak elde edilen verilerin analizinde sıklıkla Poisson Regresyon Analizi kullanılır (Mahmoud ve Alderiny, 2010, Karaca ve Olmuş, 2018). Verilerin sürekli olması durumunda doğrusal regresyon analizi kullanılabilmektedir, ancak analizlerde kullanılacak veriler her zaman sürekli halde olmayabilir. Böyle durumlarda doğrusal regresyon modeli ile yapılacak analizler yetersiz ve yanıltıcı sonuçlar vereceğinden kesikli veriler için tüm koşullar sağlandığında kullanılacak en etkin model genelleştirilmiş doğrusal modelleri esas alan Poisson regresyon modelidir. Poisson Regresyon modelleri örnek ortalaması ve örnek varyansının neredeyse eşit olduğu sayım verilerinin analiz etmek için kullanılır (Famoye ve Wang, 2004; Anonim, 2021).

Bağımlı değişkenin heterojen yapı sergilediği yani gözlemlerin birbirinde çok uzak olduğu, varyasyonun büyük olduğu durumlarda gözlenemeyen bir heterojenlik söz konusudur. Latent model ile bu heterojenliği modellemek için bağımlı değişkenleri kendi içinde popülasyonlara ayrılmaktadır. Latent modelin ilk aşamasında, gözlenemeyen heterojenliği saptayıp, homojen alt popülasyon sayısını belirlenir. İkinci aşamada ise her alt popülasyon için parametre tahmini yapılır ve herbir gözlemin ait olduğu alt popülasyon belirlenmektedir (Yeşilova ve Okut, 2003). Bu alt popülasyon aynı zamanda latent sınıf olarak adlandırılmaktadır (Dhanavanthan, 2000; Duncan vd., 2002).

Son yıllarda, özellikle büyük ve karmaşık veri setlerinin analizinde önemli ilerlemeler söz konusudur. Özellikle, latent sınıf yaklaşımı, makine öğrenme teknikleri ve yapay zeka ile ilgili çalışmalar ön plana çıkmıştır. (Wang vd., 1998; Muthén ve Muthén, 2002, Yeşilova ve Atlıhan, 2007). Latent modeller, veri setini sınıflandırmanın

yanı sıra regresyon analizini de yapmaktadır. Bağımlı değişkenin elde edilme durumuna göre olatı tüm regresyon yöntemlerinin modellenmesi mümkündür. Hem sınıflama yapması hem de regresyon analizi yapması, latent modelleme aynı zamanda karışımli model olarak bilinmektedir (Everitt ve Hand, 1981; Titterington vd., 1985; Wang vd., 1998; Chen ve Kou, 2001; Muthen ve Muthen, 2002; Kayri ve Okut, 2008)

Latent modelde, entropy doğru sınıflandırma istatistiği kullanılmaktadır. Entropy istatistiği üç aşamada kullanılmaktadır. Birincisi, veri setinin homojen dağılılacağı alt populasyon sayısının belirlemede kullanılmaktadır. En yüksek entropy olasılığına sahip model eln iyi model olarak kabul edilir. İkincisi, veri setinin homojen dağılacağı alt populasyon sayısı belirlendikten sonra her bir alt populasyon için doğru sınıflandırma olasılıklarını hesaplar. Son aşamada ise gözlemlerin hangi olasılık ile alt populasyona dahil olacağı belirler (Okut vd., 2002; Yeşilova ve Okut, 2003).

Latent Poissonu regresyon, heterojenliği elde etmede yararlı bir yaklaşım sağlar. Latent regresyon, gözlenemeyen latent sınıflar belirlemeye çalışır (Guang, 2004). Latent sınıf veya karışımli regresyon analizinin temel amacı, regresyon parametrelerinin gruptan gruba farklılık gösterdiğini ortaya koymaktır (Jeroen, 2004).

Karışık Poisson regresyon modeli olan latent Poisson (Latent Poisson Model=LPM), sayım ile elde edilen bağımlı değişkende aşırı yayılım olduğu durumlarda kullanılmaktadır. Veri seti farklı alt populasyonlardan oluşan bir heterojen yapı gösterebilir. LPM’de amaç, bir örnekteki gözlemlerin populasyondaki gözlenmemiş latent alt populasyonlara ait olabileceği varsayılır. Veri kümesinin dağılılacağı alt populasyon sayısı sonlu olduğunda sonlu latent model kullanılmaktadır. Genellikle sınırlı sayıda alt populasyona sahip olduğu varsayılarak sınırlı karışımli modeller (Finite mixture model) kullanılır. Karışımli Poisson Regresyona (Mixture Poisson Regresyon=MPR) ilişkin parametre tahminleri, EM (Expectation Maximization) ve QN (Quasi-Newton) algoritmaları kullanılmaktadır. Ayrıca Poisson dağılışı gösteren verilerde, sabit ve rasgele etkilerin bir arada olduğu karışık modeller söz konusu olduğu durumda, GLM en çok olabilirlik (Maximum Likelihood=ML) yöntemi yerine quasi-olabilirlik (Quasi Likelihood=QL) yöntemini kullanarak parametre tahmini yapmaktadır. Karışık Poisson regresyonunda, genelleştirilmiş karışık doğrusal modeller (Generalized Mixed Linear Model=GLMM) kullanılmaktadır. GLMM’de, sabit ve rasgele etkilerin her ikisi de modelde yer almaktadır.

Verileri en iyi açıklayan modelin seçiminde, Akaiki'nin bilgi ölçütü (Akaiki Information Criteria=AIC) ve Bayesian bilgi ölçütü (Bayesian Information Criteria=BIC) en çok kullanılan iki ölçüttür.





2. KAYNAK BİLDİRİŞLERİ

Wedel vd. (1993), sayım verileri için Poisson'daki heterojenliği barındıran bir yaklaşım geliştirildiğini ve geliştirilen modelin, heterojenliğin bir dağılım kesişiminden ve açıklayıcı değişkenlerin katsayılarından kaynaklandığını varsaymıştır. Ayrıca karışımın ayrık olduğunu ve sonlu bir karışım modeli formülasyonu ile sonuçlandığını varsayarak tahmin için bir EM algoritması ve model aracılığıyla sunulan kitapların müşteri satın alımlarına ilişkin verilere uygulanan algoritma, hetero regresyon modelleriyle ilgilenen bir dizi başka yaklaşımla ampirik olarak karşılaştırılmıştır.

Lin ve Tsai (2022), veriler sürekli ortak değişkenler veya birçok kategorik değişken içerdiğinde, latent sınıf inflated regresyon hem karışım bileşenlerini hem de ortalama parametreleri modellemek için bir çözüm sağladığını göstermiştir. Heterojen sayım verileri için latent bir sınıf inflated Poisson regresyon modeli önermiş olup modelin inflated verilerle iyi performans gösterdiğini gösteriyor. Önerdikleri yeni modelin inflated sayımlar için standart Poisson regresyon ve sıfır inflated Poisson regresyon modellerinden daha iyi bir uyum gösterdiğini bildirmişler.

Hynes ve Greene (2016), rekreasyonistlerin gezi tercihlerindeki heterojenliğini inceleyen bir yazılarında İrlanda'nın batı kıyısında, tarım arazilerine erişimin kısıtlandığı bir alanda plaj kullanıcılarından oluşan bir panel veri kümesine rastgele parametrelili negatif binom modeli ve latent sınıf negatif binom modeli uyguladılar. Bu yaklaşım, araştırmacının örnek verilerin yerinde toplandığını da hesaba katması gereken bir koşullu davranış seyahat maliyeti talep modelinde seçilen açıklayıcı değişkenlerin etkisine göre heterojenliği hesaba katmalarını sağladı. Analiz ayrıca bireysel tüketici artığı tahminleri geliştirdi ve tahminlerin hem rastgele parametre hem de latent sınıf özelliklerinden sistematik olarak etkilendiğini buldu.

Gage vd. (2016), okul ikliminin öğrencilerin akademik, sosyal ve davranışsal sonuçlarını etkilediğini ve okul ikliminin iyileştirilmesi ile akademik, sosyal ve davranışsal zorlukların önlenmesi için umut verici bir yol sağladığını gösterdikleri bir çalışmada verilerin analizi yöntemi için log-lineer bir Poisson latent sınıf regresyon modeliyle sonuçlanan bir Poisson dağılımı kullandılar. Latent sınıf regresyonunun amacı, öğrenci disiplin yönlendirmelerine ilişkin olarak okul iklimi anketindeki yanıt modellerine dayalı olarak öğrencilerin sınıflarının tanımlanmasıdır. En uygun modelin

ve nihai latent sınıf sayısının belirlenmesi, Bayes Bilgi Kriterlerinin (BIC), Akaike Bilgi Kriterlerinin (AIC) minimizasyonuna, sınıflandırma hatasının minimizasyonuna ve açıklanan toplam varyansın oranına dayalı olarak yapılmıştır. Yaptıkları çalışmada her sınıf içinde, ofis disiplini yönlendirmeleri ile ölçülen öğrenci sosyal ve davranışsal performansını öngören belirli öğeleri belirlediler. Her sınıf içindeki akademik, sosyal, davranışsal ve demografik profilleri araştırıp uygulama ve araştırma için çıkarımları tartıştılar. Bu ön çalışmanın sonuçları, genel olarak, okullarda okul iklimi ve ofis disiplini yönlendirmeleri ile ilişkin algılarına dayalı olarak üç geniş öğrenci sınıfının tanımlanabileceğini, en sorunlu davranışlara sahip öğrencilerin, uygun davranışlarını pekiştirmek için öğretmenlere ihtiyaç duyduğunu doğrulamaktadır.

Guo vd. (2006), makalesinin odak noktası hem kategorik sonuç hem de yordayıcı değişkenler gizli olduğunda (yani doğrudan gözlemlenemediğinde) bu ilişkiyi oluşturmak için bir model geliştirmektir. Bu model, latent sınıf regresyon modelini, gizli tahmin ediciler üzerinde gerilemeyi içerecek şekilde genişletmekteydi. Maksimum olabilirlik tahmini kullanılarak bunu gerçekleştirmek için iki sayısal yöntem açıklandı: Monte Carlo beklenti ve maksimizasyon algoritması ve Gauss kareleme ve ardından yarı-Newton algoritması. Halk sağlığı, psikoloji ve sosyal bilimler araştırmalarında, birçok araştırma sorusu, kategorik bir sonuç değişkeni ile sürekli yordayıcı değişkenler arasındaki ilişkiyi araştırdığından bu makalede de ergen sağlığıyla ilgili bir veri örneği, modelin farklı senaryolar altındaki davranışını incelemek için simülasyon çalışması yapılmış ve gizli yeme bozukluğu risk sınıflarının gizli faktör vücut memnuniyeti tarafından tahmin edildiği gösteri için kullanılmıştır.

Harel vd. (2013), meme kanseri şiddetindeki ırksal ve etnik eşitsizlikler üzerine yapılan bir çalışmadan bir örnek kullanılarak latent sınıf regresyonu ile çoklu kategorik sonuçları analiz kullanmıştır. Açık öğelere yanıt vermemek yaygın bir komplikasyon olsa da, latent sınıf regresyonu çıkarımları maksimum olasılık, çoklu değerlendirme ve iki aşamalı çoklu değerlendirme kullanılarak değerlendirilmiştir.

Weller vd. (2020), adolesanlarda mental sağlık bulguları üzerine yaptıkları bir çalışmanın latent sınıf analizi ile çözümlemiştir. Latent sınıf analizinin genellikle belirli dışsal özellikleri paylaşan popülasyonlar içindeki niteliksel olarak farklı alt grupları belirlemek için kullanılan istatistiksel bir prosedür olduğunu ve bu analizin altında yatan varsayımın, gözlemlenmeyen gruplara (veya sınıflara) anket soruları, değerlendirme

göstergeleri veya ölçeklerdeki puan kalıplarıyla açıklanabileceğini göstermiştir. Bu uygulamanın aktif ve gelişmeye devam etmek bir araştırma alanı olduğu ve yaklaşımı daha fazla araştırmacı uygulamaya başladıkça, latent sınıf analizinin yürütülmesindeki temel hususlar hakkında ayrıntılı bilgilere ihtiyaç duyulduğunu vurgulamıştır.

Wang vd. (1998), yaptıkları çalışmada, oranlarda ortak değişkenler içeren bir Poisson karışım modelleri sınıfını incelemektedir. Bu model, özel durumlar olarak Poisson regresyonu ve bağımsız Poisson karışımlarını içermektedir. EM ve yarı-Newton algoritmalarına dayalı tahmin yöntemleri, bu tahminlerin özellikleri, bir model seçim prosedürü, kalıntı analizi ve uyum iyiliği testi tartışılmaktadır. Bir Monte Carlo çalışması, uygulama ve model seçimi konularını araştırır. Bu metodoloji, nöbet sıklığını ve Ames salmonella testi verilerini analiz etmek için kullanılır.

Rabe- Hesketh ve Skrondal (2007) makalesinde kadınların işe katılımına yönelik tutumlar için ortaya çıkarılan verilerin analizinde rastgele yanıt, latent sınıf ve sıfır şişirilmiş Poisson modellerini belirtmek için açıklanan cihazlar, sıradan genelleştirilmiş doğrusal modellerle de kullanılabileceği gösterildi.

Yamaguchi (2000), sosyoloji alanındaki bir çalışmada çok terimli logit latent sınıf regresyon modellerinin yöntemini ve uygulamasını açıklamakta olup bu model grubunun sürekli latent değişkenler için gizli değişken regresyonunun kategorik analoğu olduğunu göstermiştir. Bu modeller, grup değişkenli log-doğrusal latent sınıf modellerinin regresyon uzantıları olup sosyolojik araştırmalarda bu modellerin potansiyel yararlılığı çok yüksek olduğu düşünülmektedir. Bir uygulama, Japon kadınları arasında gizli kalmış üç toplumsal cinsiyet rolü tutumu sınıfının yordayıcılarına odaklanmaktadır. Bu sınıflar, "geleneksel toplumsal cinsiyet rolü savunucuları", "toplumsal cinsiyet eşitliği savunucuları" ve "çalışma karşıtı toplumsal cinsiyet eşitliği savunucuları" olarak saptandı. Analiz, diğerleriyle karşılaştırılan her bir sınıfın özelliklerini tanımlayarak yöntemin yararlılığını gösterdi.

Park vd. (2020), hem kategorik hem de sürekli ortak değişkenleri olan ampirik bir veri setinde önerdikleri algoritma, çok değişkenli bir normallik varsayımı gerektirmeden latent sınıf regresyonda nispeten küçük bir örneklem için kullanılabildiğini savunmaktadır. Bu çalışmada, latent sınıf regresyonu olarak da bilinen ortak değişkenli latent sınıf regresyonu için bulanık bir kümeleme yöntemi ve genelleştirilmiş yapılandırılmış bileşen analizi kullanan farklı bir tahmin yaklaşımı

önerilmektedir. Bu yeni yaklaşım, dağılım varsayımının muaf ve parametreleri tahmin etmede kararlıdır. Maksimum olasılık tahmini latent sınıf analizinde kullanılan üç adımlı yaklaşıma paralel olarak, bir bulanık kümesel genelleştirilmiş yapılandırılmış bileşen analizi algoritmasını latent sınıf regresyona genişletilerek yapılmıştır. Yeni algoritma sosyal, davranışsal ve sağlık bilimlerine daha yaygın uygulanabilir.

Tarpey ve Petkova (2010) gizli regresyon modelini göstermek için simülasyon örneği ile depresyon ve ilaç tedavisi ile plasebo tepkisi için modelleme yapılmıştır. Sonlu karışım modeli, gizli bir Bernoulli öngörücüsü ile bir regresyon modeli olarak oluşturulmuştur. Latent bir regresyon modeli, ayrık Bernoulli öngörücüsünün beta dağılımına sahip sürekli bir gizli kestiriciyle değiştirilmesine dayalı olarak önerilmiştir. Latent regresyon modeli için motivasyon, farklı latent sınıfların olmadığı, bunun yerine bireylerin sürekli bir gizli değişkene göre değiştiği uygulamalardan kaynaklanır

Bouwmeester vd. (2004), bilişsel gelişim araştırmasının amacı, gizli bilişsel süreçleri veya yapıları yaş, bilişsel davranış ve çevresel etkiler gibi açık değişkenler aracılığıyla açıklamaktır. Bu yazıda, latent sınıf regresyon modelinin kullanışlılığı, bilişsel gelişim fenomenlerini incelemek için tartışılmaktadır. Bu modeli kullanarak, gizli ve açık değişkenler arasındaki ilişkiler, bilişsel gelişim kuramı tarafından yapılan güçlü muhtemel varsayımlara ihtiyaç duymadan ampirik veriler aracılığıyla açıklanabilmekteydi. Latent sınıf regresyon modelinde, belirli bilişsel davranışla karakterize edilebilecek bir dizi sınıf ayırt edilip, bilişsel davranış üzerindeki çevresel etkiler, farklı (gelişimsel) sınıflar için değişiklik gösterebilirdi. Sonuçlar, Beş Sınıflı bir modelin verilere en iyi şekilde uyduğunu ve latent sınıfların yaş, strateji kullanımı (bilişsel davranış) ve görev özelliklerinin strateji kullanımı üzerindeki etkisi (çevresel etkiler) açısından farklılık gösterdiğini gösterdi.

Lipowetsky ve Conklin (2005), modelleri veri kümelerine göre sınıflandırarak en az bir veri segmentine ait gözlemlerin maksimum olabilirlik hedefi geliştirdi. Çözüm, tüm modellerin katsayılarını ve uyum özelliklerini tanımlayan yinelemeli olarak yeniden ağırlıklandırılmış en küçük kareler prosedürüne indirgediğini gösterdi. Bu yaklaşımın, regresyon modelleriyle birlikte, her bir sınıfa ait her bir gözlemin olasılıklarını verdiği ve bu tekniğin, karma dağılımların parametrelerini bulmak için de kullanılabilirliğini bildirdi.

Mbotwa vd. (2021) çalışmasında, latent sınıf regresyon modellemesinin öngörü keskinliğini: standart genelleştirilmiş doğrusal model (GLM); ve alt grupların/sınıfların üyeliğini içeren GLM'leri (önceki latent sınıf analizi yoluyla tanımlanır) alternatif veya ek aday öngörücüler olarak karşılaştırmayı amaçladı. 1.802 kalp yetmezliği hastasından elde edilen gerçek dünya demografik ve klinik verilerini kullanan çalışma, hayatta kalmanın tek aday öngörücüsü olarak latent sınıf analizi standart çok değişkenli GLM'lerden daha düşük olduğunu bulmuştur. Latent sınıf analizi yoluyla alt grup/sınıf üyeliğinin bu dört ortak değişkenle birlikte çok değişkenli bir GLM'de aday tahmin ediciler olarak dahil edilmesi, tahmin keskinliğinde herhangi bir gelişme göstermedi. Buna karşılık, latent sınıf regresyon modellemesi, tahmin keskinliğinde %18-22'lik bir iyileşme ile sonuçlandı ve klinik kaynakların verimli tahsisini geliştirmek için en uygun olan modelleri seçmek için tahmin keskinliğini entropiye karşı dengelemenin mümkün olacağı sonucun farklı riskini ele almak için (bu örnekte hayatta kalma) bir dizi alternatif model sağladı. Bu bulgular ile, latent sınıf regresyon modellemenin GLM'lerin öngörü keskinliğini artırabileceği ve tahminlerinin klinik faydasını artırabileceğine dair ilke kanıtı sağlandı. Bu iyileştirmeler, özellikle büyük ve rutin olarak toplanan klinik veri kümeleriyle ve ikili sayımla kullanım ve sürekli değişkenler için sonuç tahminlerini belirlerken aynı zamanda latent sınıf yapısı oluşturabilen alternatif tekniklerin kullanımı da dahil olmak üzere daha fazla dikkat ve araştırmayı garanti etmiştir.

Mufudza ve Erol (2016), Poisson karışımli regresyon modelleri aracılığıyla kalp hastalığını tahmin ve teşhis etmek için model tabanlı kümeleme tekniklerinin kullanımına odaklanmıştır. Poisson karışımli regresyon modellerinin analizi ve uygulaması burada standart ve eşlik eden değişken karışım regresyon modelleri olarak iki farklı sınıf altında ele alınmıştır. Sonuçlara göre; iki bileşenli eşlik eden değişken Poisson karışımı regresyon modelinin, düşük Bayes Bilgi Kriteri değeri nedeniyle kalp hastalığını hem standart Poisson karışımı regresyon modelinden hem de sıradan genel doğrusal Poisson regresyon modelinden daha iyi tahmin ettiğini göstermekteydi. Ayrıca, Sıfır Yoğunlaştırılmış Poisson Karışımı Regresyon modeli, hem bireyleri yüksek veya düşük risk kategorilerinde kümelediğinden hem de mevcut kümelere bileşen bazında kalp hastalığı oranını tahmin ettiğinden, tüm modeller arasında kalp tahmini için en iyi model olduğu ortaya çıktı. Poisson karışım regresyon modelini kullanarak ana riskleri

bileşen bazında belirleyerek kalp hastalığı tahmininin etkili bir şekilde yapılabileceği sonucuna varılmıştır.

Wang vd. (1998), makalesine göre sayısal veri modelleri, şirket düzeyinde patentler ile araştırma ve geliştirme harcamaları arasındaki ilişkiyi analiz etmek için kullanılır ve hem Poisson oranlarında hem de karıştırma olasılıklarında ortak değişkenlere sahip sonlu bir karma Poisson regresyon modeli kullanılarak aşırı dağılımı hesaba katmaktadır. EM ve yarı-Newton algoritmaları kullanılarak maksimum olabilirlik tahmini tartışılmıştır. Monte Carlo çalışmalarına göre, (a) cezalı olabilirlik kriterleri, model seçimi için güvenilir bir temeldir ve karıştırma dağılımı için sürekli mi yoksa sonlu desteğin mi daha uygun olduğunu belirlemek için kullanılabilir ve (b) bir karıştırma dağılımı yanlış belirtildiğinde, parametre tahminleri yansız kalır ancak varyansları şişirir.

Yang ve Lai (2005), heterojen sayım verilerinin analizinde Poisson regresyonu için bulanık sınıf modeli adı verilen yeni bir yaklaşım önermiş olup bulanık küme kavramı ve bulanık sınıflandırma maksimum olabilirlik (FCML) prosedürleri temelinde, bulanık sınıf Poisson regresyon modelleri için bir FCML algoritması oluşturmuştur. Geleneksel olarak, latent sınıf regresyon modelleri için EM algoritması kullanılmıştır. Böylece, parametreleri tahmin etmek için EM ve FCML algoritmalarının doğruluğu ve etkinliği karşılaştırılmıştır. Sonuçlar, önerilen FCML algoritmasının daha iyi doğruluk ve etkinlik sunduğunu ve heterojen sayım verileri için regresyon analizi için başka bir iyi araç olarak kullanılabileceğini göstermektedir.

Coxe vd. (2009), sayım verileri sonuç değişkeninin düşük aritmetik ortalamaya sahip bir sayı olduğu durumlarda, standart sıradan en küçük kareler regresyonu yanı sıra sonuçlar verebildiği ifade etmiş, sayım verileri için uygun analizler sağlayan regresyon modellerine bir giriş sağlanması amaçlanmıştır. Standart Poisson regresyonunu bir örnek olarak tartışıp yorumlayarak Poisson regresyonunun iki çeşidi olan aşırı dağılmış regresyon ve negatif binom regresyon, Poisson standart Poisson regresyonunun temel bir varsayımı ihlal edildiğinde daha iyi sonuçlar sağlayabilen regresyon olarak tanıtılmıştır. Ayrıca, davranışı asla göstermeyecek bir yanıtlayıcı alt grubunun örneğe dahil edildiği aşırı sıfır sorunlarını ve sıfır sayımı olan yanıtlayıcıların örnekleme planı tarafından hariç tutulduğu kesik sıfır sorunlarını tartışarak Poisson regresyon modelleri

ailesinin, sayım verilerinin iyileştirilmiş ve uygulanması kolay analizlerini sağladığını vurgulamıştır.

Özmen ve Famoye (2007), Türkiye Dalyan'da 1991-1993 yılları arasında güneşe maruz kalarak ölen *Caretta Caretta* yavrularının ölüm hızlarını modellemişlerdir. Bölge, denizden uzaklık, dönem açıklayıcı değişken olarak kullanılmıştır. Çok sayıda sıfır ve aşırı yayılım içeren zoolojik veri seti Poisson, negatif binom, genelleştirilmiş Poisson, sıfır yığılmalı Poisson ve genelleştirilmiş sıfır yığılmalı Poisson regresyon modelleri ile incelenmiştir. Seçim kriterleri sonucunda veri seti için genelleştirilmiş Poisson regresyon modelinin diğer sayım verisi modellerine göre daha iyi olduğu sonucuna varılmıştır.





3. MATERYAL VE YÖNTEM

3.1 Materyal

Çalışma Antalya ilinin Finike ilçesinde 6 farklı turuncgil bahçesinde ve 2 farklı nar bahçesinde yürütülmüştür. Bu bahçelerden portakal bahçelerinde Washington, greyfurt bahçesinde Altıntop, mandarin bahçesinde Satsuma ve nar bahçelerinde Hicaz çeşitleri üretimi yapılmaktadır. Çalışmalarda ilk gözlemlere 2019 yılında başlanmış olup aynı yöntemlerle 2019 yılında devam edilmiştir. *Ceratitis capitata*'nın populasyon yoğunluğunu takip etmek için her bir bahçeye 2 adet olmak üzere toplam 16 adet feromon izleme tuzağı yerleştirilmiştir. Feromon kapsülleri 30-45 gün ara ile değiştirilmiştir. Haftalık aralıklarla feromon tuzakları yapışan Akdeniz Meyve sineği erginleri sayılarak kaydedilmiştir. Larvaların zararından etkilenen vuruklu meyvelerin tespiti için bahçe büyüklüğüne bağlı olarak 5-10 adet ağaç 2-3 dakika boyunca gözlenmiş, vuruklu görülen meyveler not edilmiştir.

3.2 Yöntem

3.2.1 Latent Modelin Parametrik Gösterimi

Bir çok uygulamada, alt grup yoğunluklarının bazı parametrik ailelere ait oldukları belirtilmiştir. Bu durumda, bilinmeyen parametrelerin karışımındaki i ninci alt grup yoğunluğu için kabul edilen formdaki vektörü q_i olmak üzere, $f_i(y_j)$ alt grup yoğunlukları, $f_i(y_j; q_i)$ şeklinde belirtilir. $f(y_j)$ karışım yoğunluğu Eşitlik (3.1) deki gibi ifade edilebilir.

$$f(y_j; \gamma) = \prod_{i=1}^g p_i f_i(y_j; q_i) \quad (3.1)$$

Eşitlik (3.1) 'deki γ vektörü, karışımındaki bilinmeyen tüm parametreleri içerir ve Eşitlik (3.2) 'deki gibi gösterilir.

$$\psi = (\pi_1, \pi_2, \dots, \pi_{g-1}, \xi^T)^T \quad (3.2)$$

Eşitlik (3.2)' deki X vektörü, $\theta_1, \theta_2, \dots, \theta_g$ vektörleri içindeki tüm parametreleri içerir. $\mathcal{Y}$ için belirlenmiş parametre uzayı $\mathcal{W}$ ile gösterilmektedir. ρ_i karışma oranlarının toplamı 1 e eşit olacağı için bunlardan biri gereksizdir. Bu yüzden, Eşitlik (3.2)' de g ninci karışım oranını gösterecek olan ρ_g ifadesi keyfi olarak ihmal edilmiştir.

Parametrik bir karışımı tanımlamak için yukarıdaki eşitlikler kullanıldığında, Kanji (1985)'nin uçak inişlerinde karşılaşılan rüzgâr makaslarının dağılımlarının modellenmesinde önerdiği gibi ortak bir m ortalamaya sahip Laplace bileşeni ve tek değişkenli normal bileşenin karışımı ele alınmaktadır. Bu modelde rüzgâr makaslarının Y_j karışım yoğunluğu;

$$\mathcal{Y} = (p_1, x^T)^T \text{ olduğunda ve } x = (ms^2, k)^T \text{ iken} \quad (3.3)$$

$$f(y_j; \mathcal{Y}) = \rho_1 f(y_j; ms^2) + \rho_2 (2k)^{-1} \exp(-|y_j - m|/k)$$

şeklinde gösterilmektedir. Çoğu zaman alt grup yoğunlukları, aynı parametrik aileye ait olarak belirtilir. $\mathcal{Q}$ için parametre uzayı Q şeklinde ifade edildiğinde,

$$\{f(y_j, q) : q \in Q\} \quad (3.4)$$

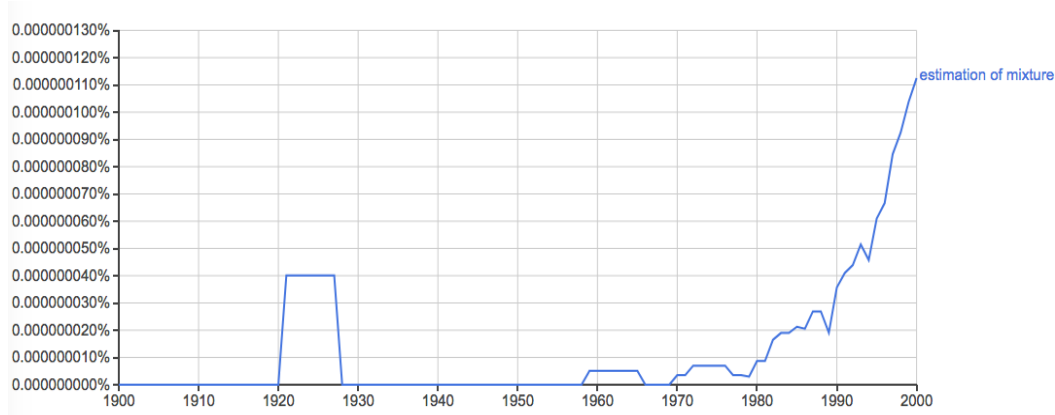
olarak gösterilen parametrik ailenin genel üyesi, $f(x; q)$ şeklinde ifade edilir ve $f(y_j; \mathcal{Y})$ karışım yoğunluğu;

$$f(y_j; \mathcal{Y}) = \sum_{i=1}^g \rho_i f(y_j; q_i) \quad (3.5)$$

şeklinde gösterilir. Genel bir gösterim olarak f sembolü, yoğunluk fonksiyonunu belirtmek amacıyla kullanılmaktadır. Örneğin $f(y_i; \mathcal{Y})$ ifadesi karışım yoğunluğunu göstermek için kullanılmıştır. $f(y_i; q_i)$ ifadesi ise Eşitlik (3.3) ailesinin altındaki i ninci alt grup yoğunluğunu göstermektedir.

3.2.2 Latent Modellerde Parametre Tahmini

Karışımli dağılımları tahmin edebilmek için çeşitli yaklaşımlar ortaya konmuştur. Bu yaklaşımlar, grafiksel metotları, momentler, en yakın mesafe, en çok olabilirlik (Maximum Likelihood - ML) ve bayesian gibi metotları içermektedir. Karışımlar için tahmin metodolojileri üzerine bu kadar geniş bir literatür olmasının ana nedeni olarak, parametre tahmini için açık formüllerin tipik olarak mevcut olmamasıdır (Titterington, 1997). Örneğin, karışma oranları ve alt grup ortalama ve varyans/kovaryansları için en çok olabilirlik tahmini (MLE-Maximum Likelihood Estimation), normal karışımlar söz konusu olduğunda kapalı bir formda yazılıp gösterilemez. Bu olabilirlik tahminleri iteratif olarak hesaplanır. Pearson (1894)'ün çalışmasında, iki tek değişkenli normal dağılım karışımının parametre tahmini yapılırken momentler metodu kullanılmıştır. Son yıllarda artan bir şekilde (Şekil 3.1), normal karışımların parametre tahminleri üzerine çeşitli çalışmalar gerçekleştirilmiştir (Lindsay ve Başak, 1993; Furman ve Lindsay, 1994; Ramos-López vd., 2018; Orellana vd., 2018).



Şekil 3.1 Karışım dağılımları konusundaki akademik yazıların yıllara göre değişimi

3.2.3 Karışımli Modelde Alt Grup Sayısının Test Edilmesi

Karışımli modellerin bazı uygulamalarında, karışımın içinde bulunan g alt grup sayısı hakkında yeterli bir ön bilginin bulunmadığı durumlarla karşılaşmaktadır. Bu duruma örnek olarak, alt grupların, normal dağılım gösteren harici gruplara karşılık gelmesi verilebilir. Bununla birlikte, birçok durumda, alt grup sayısı, alt grup yoğunluklarındaki parametrelerle birlikte verilerden çıkarılmalıdır. Örneğin, McLaren vd. (1991), iki bileşenli log normal karışım yoğunluğu kullanarak, önceden anemi hastalığına yakalanmış kişilerin akyuvar hücre hacimleri üzerindeki bazı dağılımları açıklamaya çalışmışlardır. Sağlıklı kişilerdeki akyuvar hacim dağılımları sadece bir log normal dağılım ile modellenebiliyorken, anemisi tam olarak iyileşmemiş hastalarda akyuvar hacimleri demir tedavisi bitiminde iki log normal dağılımın karışımı şeklinde görülmüştür. McLaren vd. (1991) çalışmasında, istatistiksel testler yapılarak, demir tedavisine ne zaman son verilebileceği konusunda hekimlere erken teşhis imkânı sunulabileceği sonucuna ulaşılmıştır.

3.2.4 Karışımli Modellerde En Çok Olabilirlik Yöntemi

EM algoritması geliştirilinceye kadar geçen sürede, en çok olabilirlik, karışımli dağılımların modellenmesinde sıkça kullanılmıştır.

3.2.4.1 Karışımli modellerde en çok olabilirlik

En çok olabilirlik yaklaşımı ile d - boyutundaki Y parametre vektörünün $f(y_j, Y)$ dağılımına uyduğu varsayıldığında, Y_j şansa bağlı vektörü çalışmadaki j inci olguyu göstermek şartıyla, $\hat{Y}$ tahmini, düzenli duruma uygun çözüm olarak Eşitlik (3.6)' da gösterilen en çok olabilirliği sağlar.

$$\begin{aligned} \nabla L(Y) / \nabla(Y) = 0 \quad \text{veya} \quad \nabla \log L(Y) / \nabla(Y) = 0 \\ L(Y) = \prod_{j=1}^n f(y_j; Y) \end{aligned} \quad (3.6)$$

$L(Y)$, Y şeklindeki bağımsız olduğu varsayılan $y_1, \dots, y_n$ değişkenlerin en çok olabilirlik fonksiyonudur.

Daha kısa olarak ifade edilecek olunursa, en çok olabilirlik tahmin işleminin amacı, her bir n için tahmin değerlerini saptamaktır (yukarıda $\hat{Y}$ şeklinde gösterilmiştir). Böylelikle en çok olabilirlik tahmini, en çok olabilirlik eşitliğinin kök dizisini tutarlı ve asimptotik olarak etkili bir şekilde sağlar. Böyle bir dizi, uygun düzenlilik koşulları sağlandığında bilinir. Olasılığın 1 e tamamlanma eğiliminde olması ile parametre uzayı içinde bu kökler yerel maksimuma yakınsar. Bu köklerin tutarlı sıralaması aslında benzersizdir.

3.2.5 Bilgi Matrisleri

Gözlenen değerler $y = (y_1^T, \dots, y_n^T)^T$ olmak üzere log olabilirlik fonksiyonunun $S(y; Y)$ gradyan vektörü (istatistik skoru)

$$S(y; Y) = \nabla \log L(Y) / \nabla(Y) \quad (3.7)$$

şeklinde gösterilir. Eşitlik (3.7)' de gösterilen gradyan matrisi kullanılarak, Y parametre vektörü hakkındaki $I(Y)$ beklenen bilgi matrisi

$$I(Y) = E_Y \{ S(Y; Y) S^T(Y; Y) \}, \quad (3.8)$$

şeklinde ifade edilebilir. Gözlenen bilgi matrisi diye de tanımlanan Log olabilirlik fonksiyonunun Hessian matrisi

$$I(Y; y) = - \nabla^2 \log L(Y) / \nabla Y \nabla Y^T \quad (3.9)$$

şeklinde gösterilir. $I(Y; y)$ Eşitlik (3.9)' da gösterildiği gibi olduğunda, düzenli şartlar altında $I(Y)$ beklenen bilgi matrisi

$$I(Y) = E_Y \{ I(Y; y) \} \quad (3.10)$$

şeklinde ifade edilir.

3.2.6 Maksimum Olabilirlik Tahmininin Asimptotik Kovaryans Matrisi

Maksimum olabilirlik tahmininin asimptotik kovaryans matrisi, beklenen bilgi matrisi $I(Y)$ nin tersine eşittir ve $I(\hat{Y})$ ye yakınsar. $(A)_{rs}$ standart gösterim olarak, A matrisinin r inci satırı, s inci sütununu gösterdiği göz önünde bulundurularak, $\hat{Y}_r = (\hat{Y})_r$ nin standart hatası

$$SE(\hat{Y}_r) \gg \left(I^{-1}(\hat{Y}) \right)_{rr}^{1/2} \quad (r=1, \dots, d) \quad (3.11)$$

şeklinde ifade edilir.

3.2.7 Karışımli Modeller için EM Algoritması

3.2.7.1 Doğrudan yaklaşım

X vektörü, önsel olarak bilinen $q_1, \dots, q_g$ içindeki tüm parametreleri içermek kaydıyla, karışımli modeldeki bilinmeyen tüm parametreleri kapsayan Y vektörü;

$$Y = (p_1, \dots, p_{g-1}, x^T)^T \quad (3.12)$$

şeklinde gösterilir.

Gözlenen rastgele örneklem; $y = (y_1^T, \dots, y_n^T)^T$ şeklinde olup, üzerinde maksimum olabilirlik için EM algoritması uygulanacak karışımli modelin parametrik gösterimi;

$$f(y_j; Y) = \prod_{i=1}^g p_i f_i(y_j, q_i) \quad (3.13)$$

şeklinde ifade edilir.

Y vektörü için log olabilirlik, gözlenen veri yardımıyla

$$\log L(Y) = \sum_{j=1}^n \log f(y_j; Y) = \sum_{j=1}^n \log \left\{ \sum_{i=1}^g p_i f_i(y_j, q_i) \right\} \quad (3.14)$$

şeklinde oluşturulabilir.

Olabilirlik eşitliğinin,

$$\nabla \log L(Y) / \nabla Y = 0 \quad (3.15)$$

şeklinde ifade edildiği düşünüldüğünde, Eşitlik (3.2) deki ifadenin Y vektörüne göre türevinin alınıp sıfıra eşitlenmesi gerekmektedir. EM algoritması bu çözümü iteratif metotlarla çözmeyi önermektedir.

3.2.7.2 E- aşaması

$Z_1, \dots, Z_n$ ilişkili alt grup etiket vektörü mevcut olmadığından, $y = (y_1^T, \dots, y_n^T)^T$ gözlenen veri vektörü tam olmayan bir görünümüdür. Burada, her bir y_j 'in, Eşitlik (3.13) de gösterilen karışımli modelin alt gruplarından biri olan i 'nci alt gruba dahil olup olmasına göre $z_j = (z_j)_i$ 1 veya 0 değerini alarak, g - boyutlu vektör olan Z_j vektörü oluşacaktır ($i = 1, \dots, g; j = 1, \dots, n$). Böylelikle

$$z = (z_1^T, \dots, z_n^T)^T \quad (3.16)$$

olmak kaydıyla, tam veri (*complete-data*)

$$y_c = (y^T, z^T)^T \quad (3.17)$$

şeklinde ifade edilir. $Z_1, \dots, Z_n$ alt grup etiket vektörleri, $Z_1, \dots, Z_n$ rastgele vektörlerinin gerçekleşen değerleri olarak alınmıştır. Tam veri için log olabilirlik;

$$\log L_c(Y) = \sum_{i=1}^g \sum_{j=1}^g z_{ij} \{ \log p_i + \log f_i(y_j; q_i) \} \quad (3.18)$$

şeklinde ifade edilebilir. EM algoritması, Z_{ij} değişkenini eksik veri (*missing data*) olarak ele almaktadır. EM algoritması, E aşaması (*Expectation: Beklenti*) ve M aşaması (*Maximization: Maksimizasyon*) olarak iki adımdan oluşmaktadır. Probleme, gözlenen y verisi üzerinde $\log L_c(Y)$ tam veri log olabilirliğin koşullu beklenti değeri olarak hesaplanan Z_j şeklindeki gözlemlenemeyen verilerin eklenmesi, E aşaması ile sağlanmaktadır. Y değişkeninin başlangıç değeri $Y^{(0)}$ şeklinde gösterilmektedir. EM algoritmasının ilk iterasyonunda, yani E aşamasında, verilmiş olan y gözlenen değerleri için, Y yerine $Y^{(0)}$ kullanarak, $\log L_c(Y)$ 'nin koşullu beklentisi;

$$Q(Y; Y^{(0)}) = E_{Y^{(0)}} \{ \log L_c(Y) | y \} \quad (3.19)$$

şeklinde hesaplanır. Beklenti operatörü olarak gösterilen E 'nin altındisi olan $Y^{(0)}$, bu beklentinin Y için $Y^{(0)}$ kullanarak hesaplanacağını ifade etmektedir. $k+1$ inci iterasyondaki E aşaması için hesaplanması gereken $Q(Y; Y^{(k)})$ ifadesinde, $Y^{(k)}$ değeri, Y 'nin k 'inci EM iterasyonu sonucundaki değerini gösterir. Gözlemlenemeyen Z_{ij} ile bağlantılı $\log L_c(Y)$ 'nin $k+1$ inci EM iterasyonu sonucundaki koşullu beklenti değeri Z_{ij}

$$E_{Y^{(k)}}(Z_{ij} | y) = pr_{Y^{(k)}} \{ Z_{ij} = 1 | y \} = t_i(y_j; Y^{(k)}) \quad (3.20)$$

şeklinde ifade edilir. $i=1, \dots, g$ ve $j=1, \dots, n$ olmak üzere, örneklemin j 'inci üyesini gösteren y_j gözlenen değerinin i ninci alt gruba ait olma sonsal(*posterior*) olasılığı;

$$t_i(y_j; Y^{(k)}) = \frac{p_i^{(k)} f_i(y_j; q_i^{(k)})}{f(y_j; Y^{(k)})} = \frac{p_i^{(k)} f_i(y_j; q_i^{(k)})}{\prod_{h=1}^g p_h^{(k)} f_h(y_j; q_h^{(k)})} \quad (3.21)$$

şeklinde gösterilir. Verilen bir y vektörü için, Eşitlik (3.20) i, Eşitlik (3.18) de kullanarak, koşullu beklenti fonksiyonu;

$$Q(Y; Y^{(k)}) = \prod_{i=1}^g \prod_{j=1}^n t_i(y_j; Y^{(k)}) \{ \log p_i + \log f_i(y_j; q_i) \} \quad (3.22)$$

şeklinde hesaplanır.

3.2.7.3 M-aşaması

M aşaması, W parametre uzayındaki Y vektörüne göre, $Q(Y; Y^{(k)})$ 'nin $k+1$ inci iterasyonu sonucunda global maksimumlarını bulmayı içermektedir. Sonlu karışım modelleri için, p_i karışım oranlarının güncellenmiş $p_i^{(k+1)}$ tahminleri, alt gruplardaki bilinmeyen parametreleri içeren X parametre vektörünün güncellenmiş X_i tahminlerinden bağımsız olarak hesaplanmaktadır.

Eğer, Z_{ij} gözlemlenebiliyorsa, p_i nin en çok olabilirlik tahmini

$$\hat{p}_i = \frac{\sum_{j=1}^n Z_{ij}}{n} \quad (i=1, \dots, g). \quad (3.23)$$

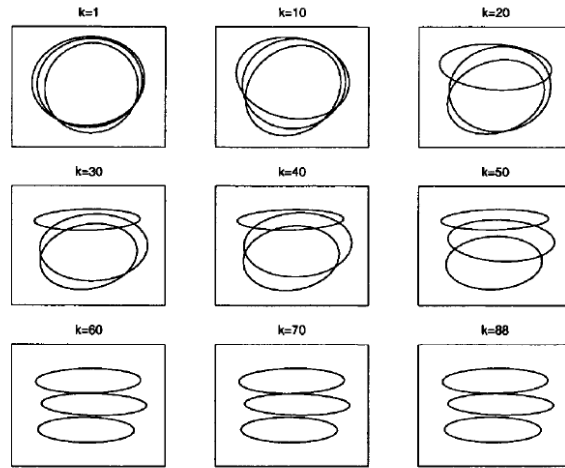
şeklinde ifade edilebilir. E- aşamasındaki her bir Z_{ij} 'nin, o anki koşullu beklentisi olan $t_i(y_j; Y^{(k)})$ ile yer değiştirmesi gibi, Eşitlik (3.27) de verilen Z_{ij} yerine $t_i(y_j; Y^{(k)})$ konularak, p_i 'nin güncel tahmini;

$$\rho_i^{(k+1)} = \frac{1}{n} \sum_{j=1}^n t_i(y_j; Y^{(k)}) \quad (i=1, \dots, g). \quad (3.24)$$

şeklinde hesaplanır.

3.2.8 Model Seçiminde Uyum Ölçütleri

Karışıklı modellerde, her karışma bileşeni, latent bir alt grup değişkeni gibi düşünülebileceğinden uygun alt sayısının seçimi ayrı bir problem olarak ele alınmaktadır. Alt grup sayısını belirlerken, tahmini model parametrelerinin değerleri alınır. Bu amaçla, en çok olabilirlik tahminlerini hesaplamak için EM algoritması kullanılır. Farklı alt grup sayısı gösterimler Şekil 3.2’de verilmiştir.



Şekil 3.2 Rasgele başlangıç değerleri kullanılarak EM algoritması için farklı k değerleri için üretilen parametrelere dayalı her alt grup için asimptotik (%95) elipsoidler grafiği (McLachlan ve Peel, 2000)

3.2.8.1 Log Olabilirliğin Sapması

Model seçiminde, oluşturulan (fitted) modelin gerçek verilerine ulaşılması amacıyla Kullback-Leibler (1951) yaklaşımı kullanılabilir. $f(w)$; gerçek yoğunluğu ifade ederse, $f(w)$ 'nin $f(w; \hat{Y})$ tahminine göre Kullback-Leibler değeri;

$$I\{f(w); f(w; \hat{Y})\} = \int \hat{f}(w) \log f(w) dw - \int \hat{f}(w) \log f(w; \hat{Y}) dw \quad (3.25)$$

şeklinde ifade edilir.

Bu ifade, $f(w; \hat{Y})$ 'ye göre $\hat{f}(w)$ 'nin sapmasının bir ölçüsü olarak verilmektedir. Burada amaç, Eşitlik 3.29'daki Kullback-Leibler ifadesini küçük yapmaya çalışmaktır.

Eşitlik 3.29'da verilen denklemin sağ tarafındaki ilk terim modele bağlı değildir, yalnızca ikinci terim modele ilişkin bir terim olup, F gerçek dağılımı göstermek ve

$y = \left(y_1^T, \dots, y_n^T \right)^T$ gözlenen verileri içermek şartıyla, bu ikinci terim;

$$\begin{aligned} h(y; F) &= \int \hat{f}(w) \log f(w; \hat{Y}) dw \\ &= \int \hat{f} \log f(w; \hat{Y}) dF(w) \end{aligned} \quad (3.26)$$

şeklinde genişletilebilir. $h(y; F)$ 'nin basit bir tahmini, Eşitlik 26 dalı ifadede F yerine $\hat{F}$ deneysel dağılım fonksiyonu bırakılarak;

$$\begin{aligned} h(y; \hat{F}_n) &= \frac{1}{n} \sum_{j=1}^n \log f(y_j; \hat{Y}) \\ &= \frac{1}{n} \log L(\hat{Y}) \end{aligned} \quad (3.27)$$

şeklinde ifade edilir ve her bir $y_j (j=1, \dots, n)$ gözleminde ağırlıkça $1/n$ kadar bulunur. Buradan beklenen log yoğunluğunun aşırı tahmini olarak;

$$\int \hat{f} \log f(w) dF(w) \quad (3.28)$$

elde edilir.

Genellikle $\hat{F}_n$ deneysel dağılım fonksiyonu, gerçekteki F dağılımına, $F_{\hat{Y}}$ ile ifade edilen oluşturulmuş dağılım fonksiyonunun yakınsadığından daha fazla yakınsar. Bağımsız $Y_1, \dots, Y_n$ 'nin ortak dağılım fonksiyonu olarak F kullanıldığında beklenti değeri E_F olmak şartıyla, Eşitlik (3.56) nın bir tahmin edicisi olarak $h(y; \hat{F}_n)$ 'nin sapması;

$$\begin{aligned} b(F) &= E_F \left\{ h(Y; \hat{F}_n) - h(Y; F) \right\} \\ &= E_F \left\{ \frac{1}{n} \sum_{j=1}^n \log f(y_j; \hat{Y}) - \int \log f(w; \hat{Y}) dF(w) \right\} \end{aligned} \quad (3.29)$$

olarak ifade edilir.

$b(F)$ sapma teriminin uygun tahmini yapılarak, sapması düzeltilmiş log-olabilirlik;

$$\log L(\hat{\Psi}) - b(F) \quad (3.30)$$

şeklinde ifade edilir ve bu fonksiyon, model seçiminde bilgi ölçütü olarak kullanılabilir. Burada amaç, seçilen modelde mevcut bileşen sayısını en üst düzeye çıkarmak (Eş. (3.30)) ve böylece Kullback-Leibler bilgisini (Eş.(3.25)) en aza indirmektir.

Literatürde, bu şekilde oluşturulan bilgi ölçütleri, genellikle bu farkın negatif değerinin iki katı olacak şekilde, yani;

$$-2\log L(\hat{Y}) + 2C \quad (3.31)$$

şeklinde gösterilir. Eşitlik (3.31)' deki' sağındaki ilk terim uyum eksikliğini (*lack of fit*) ölçmekte, ikinci terimdeki C ise modelin karmaşıklığını ölçen ceza (*penalty*) terimini göstermektedir. Buradaki amaç, kriteri en aza indirgeyen bir modelin seçilebilmesidir.

3.2.8.2 Akaike Bilgi Ölçütü

Akaike çalışmalarında (1973, 1974) $b(F)$ 'nin asimptotik olarak d 'ye eşit olduğunu göstermiştir; burada d , modeldeki toplam parametre sayısına eşittir. Akaike bilgi ölçütü (Akaike's Information Criterion - AIC),

(3.32)

$$-2 \log L(\hat{\Psi}) + 2d$$

değerini en aza indirebilen modeli seçer. Akaike'den sonra, AIC'nin bir karışım içindeki alt grup sayısını seçmek için kullanımı konusunda, yeni yaklaşımlar içeren çalışmalar yapılmıştır (Bozdogan ve Sclove, 1984; Sclove, 1987).

Konishi ve Kitagawa (1996), gerçek yoğunluğun $f(y_j)$, öngörülen parametrik aileye ait olmadığı ve parametre vektörünün zorunlu olarak en çok olabilirliğe göre tahmin edilmediği durumlarda, ilgili asimptotik sapmayı türetmişlerdir. Bununla birlikte bu asimptotik açılımların $b(F)$ için geçerliliği, normal asimptotik teorisi için gereken tekdüzelik (*regularity*) koşullarına bağlıdır. Bu şekliyle, Titterington vd. (1985) tarafından değinilen LRTS (Likelihood Ratio Test Statistic)'nin boş dağılımı geçerli kılınmış olur. Bu durumda, LRTS, bir karışım modelindeki alt grup sayısı üzerinde yapılan testler için işlevsiz hale gelmektedir.

Buna rağmen, AIC kriteri, karışımli modellerin düzenini değerlendirmek için sıklıkla kullanılmaktadır. Koehler ve Murphree (1988), AIC'nin düzeninin tutarsız olduğunu ve modellere uyma eğiliminde olduğunu gözlemlemiştir. Karışım bağlamında, AIC'nin doğru sayıda alt grubu, olduğundan fazla hesaplama eğiliminde olduğu bildirilmiştir. (Soromenho, 1994; Celeux ve Soromenho, 1996).

3.2.8.3 Bilgi Karmaşıklığı Kriteri

Bozdogan, AIC'nin performansını iyileştirmek için bilgi karmaşıklığı (*Informational Complexity-ICOMP*) kriterini önermiştir (Bozdogan, 1990; Bozdogan, 1993). Bu kriter, modeldeki parametre sayısı d olmak kaydıyla, Eşitlik (3.33) ile aşağıdaki gibi ifade edilebilir.

$$IOMP(g) = -2\log L(\hat{Y}) + C_1 + C_2 \quad (3.33)$$

Eşitlik (3.33)'deki C_1 ;

$$C_1 = d\log\{d^{-1}trI^{-1}(\hat{Y})\} \quad (3.34)$$

şeklinde ifade edilir ve C_2 ise;

$$C_1 = d\log\{d^{-1}trI^{-1}(\hat{Y})\} \quad (3.35)$$

şeklinde gösterilebilir. Bu kriterde, yeni bir parametre oluşturma(*reparameterization*) sırasında modelin değişken olması, sorun olarak görülen bir konudur. C_1 ve C_2 ifadelerindeki, örneklemdeki beklenti değeri olan $I(\hat{Y})$ 'yi, tek bir gözlemdeki beklenti değeri ile değiştirebiliriz. Bu işlem, $d\log n$ 'in iki terimden de çıkarılması ile sonuçlanacaktır. Bu kriteri pratikte uygulayabilmek için, beklenen bilgi matrisi olan $I(\hat{Y})$ 'nin tahmin edilmesi gerekmektedir.

Bozdoğan (1993), bu eşitliği bir karışım model içerisinde uygulamanın zor olduğunu düşünerek, sınıflandırılmış bir örneklem için (tahmini) beklenen bilgi matrisini kullanmak yerine bu matrise yaklaşma yönünde çalışmıştır. g -alt gruplu normal karışımli model için, $d = gp + \frac{1}{2}gp(p+1)$ olmak şartıyla, C_1 ve C_2 'nin tahmin edilebilmesinde, aşağıdaki iki denklem kullanılabilir.

$$C_1 = d\log\left[d^{-1}\sum_{i=1}^g\left\{p_i^{-1}tr\hat{S}_i + \frac{1}{2}tr(\hat{S}_i\hat{S}_i) + \frac{1}{2}(tr\hat{S}_i)^2 + \sum_{v=1}^p(\hat{S}_i)_{wv}^2\right\}\right]$$

ve

$$C_2 = (p+2)\sum_{i=1}^g\log|\hat{S}_i| - p\sum_{i=1}^g\log(n\hat{p}_i) + gp\log(2n) \quad (3.36)$$

Normal karışımli modeller ile gerçekleştirdikleri simülasyonlarda, Celeux ve Soromenho (1996) tarafından, alt grupların kovaryans matrislerinin çok farklı olduğu durumlarda, ICOMP'nin, karışımli model alt grup sayısını, AIC'den daha belirgin bir şekilde, olduğundan daha fazla tahmin etme (*overestimate*) eğiliminde olabileceği belirtilmiştir.

3.2.8.4 Bayesian Taban Uyum Ölçütleri

.

3.2.8.4.1 Bayesian Yaklaşım

Model seçimi için bir Bayesian çerçevesi içerisinde türetilmiş bazı kriterler mevcuttur. Fakat bu yaklaşımlar, Bayesian olmayan bir çerçevede de uygulanabilir. Bu nedenle bu yaklaşımlar, Bayesian olan veya başka bir bakış açısıyla ele alınan karışımli modellerde, alt grup sayısının seçiminde tercih edilmektedir. Bu başlık altında, yalnızca Bayesian çerçevede uygulanan bazı kriterler açıklanmıştır.

Bayesian temelli bilgi kriterleri konusunda ilk olarak, Schwarz (1978) tarafından, Bayesian bilgi kriterinin (BIC) kullanımında entegre olabilirliğe göre bir yaklaşım sunulmuştur. Bu yaklaşımın genel teorik gerekçeleri, bileşen sayısının ortaya çıkarılması için, frekanslara dayalı bir yaklaşımla, veri setini benzer düzenliliğin olduğu gruplara ayırma esasına dayanır.

3.2.8.4.2 Bayesian Bilgi Kriteri

$$\log p(y) = \log L(\hat{Y}) + \log p(\hat{Y}) - \frac{1}{2} \log |I(\hat{Y}; y)| + \frac{1}{2} d \log(2p) \quad (3.37)$$

Schwarz (1978)'in önerdiği, Eşitlik (3.37)'de $O(1)$ teriminin ihmal edilmesi ile;

$$|I(\hat{Y}; y)| = O(n^d) \quad (3.38)$$

olmak şartıyla, karışımli modelde, g alt grupların sayısı için, negatif log olabilirlik cezası(*penalty*) iki ile çarpım olacak şekilde, model seçiminde en küçük olarak elde edilecek Bayesian bilgi kriteri (BIC);

$$-2\log L(Y) + d\log n \quad (3.39)$$

şeklinde hesaplanır.

3.2.9 Latent Sınıf Poisson Regresyon

$Y_1, \dots, Y_n$ Bağımlı değişkenin farklı alt sınıflardan oluştuğu varsayılmaktadır ve i 'nci y gözleminin hangi k sınıfında olması gerektiği λ_{ik} oranı ile belirlenir. Bu durumda karışımli poisson dağılımı Eşitlik 3.40' daki gibidir.

$$f_k(y_i | \lambda_{ik}) = \frac{e^{-\lambda_{ik}} (\lambda_{ik})^{y_i}}{y_i!}, i = 1, \dots, n, k = 1, \dots, c. \quad (3.40)$$

Olabilirlik fonksiyonu şöyledir.

$$\ln f_{ik}(y_i | \lambda_{ik}) = y_i \ln(\lambda_{ik}) - \lambda_{ik} - \ln(y_i!). \quad (3.41)$$

Genelleştirilmiş doğrusal modeline göre (McCullag ve Nelder), kanonik parametresi $Q(\lambda_{ik}) = \ln(\lambda_{ik})$ şeklindedir. Poisson regression modelinde, kanonik bağlantı fonksiyonu kullanarak $x_{i1}, \dots, x_{iL}$ (bağımsız değişkenleri) aşağıdaki gibidir.

$$x_{i0} = 1. \text{ ile } \ln \lambda_{ik} = \beta_{0k} + \sum_{\ell=1}^L x_{i\ell} \beta_{\ell k} = \sum_{\ell=0}^L x_{i\ell} \beta_{\ell k},$$

Latent sınıf karışımli Poisson modelini esas alarak parametre tahmini yapmaktadır. Bu nedenle, latent sınıf Poisson regresyonu Poisson log-lineer regresyonunu esas alığında; k latent sınıf sayısına göre K ($K=1, 2, \dots, c$) ifade edilir.

$$f(y_i | \alpha, \beta) = \sum_{k=1}^c \alpha_k f_k(y_i | \lambda_{ik}), i = 1, \dots, n, \quad (3.42)$$

$\beta_k = \{\beta_{0k}, \beta_{1k}, \dots, \beta_{Lk}\}$ ile

$\beta = \{\beta_1, \dots, \beta_c\}$ ve $\ln y_{i|k} = \beta_{0k} + x_{i1}\beta_{1k} + \dots + x_{iL}\beta_{Lk}$ olduğunda

Burada $\alpha = \{\alpha_1, \dots, \alpha_c\}$ ve α_k gösterir ki, k' inci sınıfın oranı $\sum_{k=1}^c \alpha_k = 1$.

Dempster vd. (1977) EM algoritması önermiş olup Wedel vd. (1993) latent sınıf Poisson regresyon modelleri için EM kullanmışlardır. EM algoritmasını elde etmek için $\{y_1, \dots, y_n\}$ nin tamamlanmamış bir veri seti olduğunu ve latent sınıf değişkeni

$z_i = (z_{1i}, \dots, z_{ci})$ nin

$$z_{ki} = z_k(y_i) = \begin{cases} 1, & \text{eğer } y_i \in k \text{ inci sınıf} \\ 0, & \text{diğer} \end{cases} \text{ olduğunda} \quad (3.43)$$

ve onun dağılım fonksiyonu $f(z_i) = \prod_{k=1}^c \alpha_k^{z_{ki}}$ multinominal dağılım göstermektedir.

Böylece, latent sınıf poisson regresyonu Eşitlik (3.44)'daki gibi yazılabilir.

$$f(y_i | z_i) = \prod_{k=1}^c (f_k(y_i | \beta_k))^{z_{ki}} \quad (3.44)$$

$$f(y_1, \dots, y_n, z_1, \dots, z_n) = \prod_{i=1}^n f(y_i, z_i) = \prod_{i=1}^n f(y_i | z_i) f(z_i) = \prod_{i=1}^n \left(\prod_{k=1}^c (f_k(y_i | \beta_k))^{z_{ki}} \alpha_k^{z_{ki}} \right). \quad (3.45)$$

Olabilirlik fonksiyonu ise şöyledir:

$$\ln L = \sum_{i=1}^n \sum_{k=1}^c z_{ki} \ln f_k(y_i | \beta_k) + \sum_{i=1}^n \sum_{k=1}^c z_{ki} \ln \alpha_k. \quad (3.46)$$

Çünkü E –aşamasında latent sınıf Poisson regresyonu aşağıdaki gibi yazılabilir.

$$F(z_{ki} | y_i) = \frac{f(y_i, z_{ki})}{f(y_i)} = \frac{f(y_i | z_{ki}) f(z_{ki})}{\sum_{s=1}^c \alpha_s f_s(y_i | \beta_s)} = \frac{(\alpha_k f_k(y_i | \beta_k))^{z_{ki}} \left(\sum_{s=1}^c \alpha_s f_s(y_i | \beta_s) \right)^{1-z_{ki}}}{\sum_{s=1}^c \alpha_s f_s(y_i | \beta_s)} \quad (3.47)$$

$$\hat{z}_{ki} = E[z_{ki} | y_i, \alpha, \beta_k] = 1 \cdot \frac{\alpha_k f_k / (y_i | \beta_k)}{\sum_{s=1}^c \alpha_s f_s / (y_i | \beta_s)} + 0 = \frac{\alpha_k f_k / (y_i | \beta_k)}{\sum_{s=1}^c \alpha_s f_s / (y_i | \beta_s)}. \quad (3.48)$$

$$E(\ln L) = \sum_{i=1}^n \sum_{k=1}^c \hat{z}_{ki} \ln f_k(y_i | \beta_k) + \sum_{i=1}^n \sum_{k=1}^c \hat{z}_{ki} \ln \alpha_k. \quad (3.49)$$

M-aşamasında her bir popülasyon için, karışma olasılıkları $\hat{\alpha}_k$ ve parametre tahminleri sırasıyla hesaplanır.

$$\hat{\alpha}_k = \frac{\sum_{i=1}^n \hat{z}_{ki}}{n}. \quad (3.50)$$

$$L(\beta) = \sum_{i=1}^n \sum_{k=1}^c \hat{z}_{ki} \ln f_k(y_i | \beta_k) + \sum_{i=1}^n \sum_{k=1}^c \hat{z}_{ki} \ln \hat{\alpha}_k. \quad (3.51)$$

olsun. Ardından

$$\frac{\partial}{\partial \beta_{\ell k}} L(\beta) = \sum_{i=1}^n \hat{z}_{ki} \frac{\partial \ln f_k(y_i | \lambda_{\eta|k})}{\partial \lambda_{\eta|k}} \frac{\partial \lambda_{\eta|k}}{\partial \beta_{\ell k}}. \quad (3.52)$$

$$\ln f_k(y_i | \lambda_{\eta|k}) = y_i \ln(\lambda_{\eta|k}) - \lambda_{\eta|k} \ln(y_i!) \quad (3.53)$$

$$\ln \lambda_{\eta|k} = \sum_{\ell=1}^L x_{i\ell} \beta_{\ell k}, \frac{\partial L(\beta)}{\partial \lambda_{\eta|k}} = (y_i / \lambda_{\eta|k} - 1), \frac{\partial \lambda_{\eta|k}}{\partial \beta_{\ell k}} = x_{i\ell} \lambda_{\eta|k}. \quad (3.54)$$

Böylece,

$$\frac{\partial}{\partial \beta_{\ell k}} L(\beta) = \sum_{i=1}^n z_{ki} \left(\frac{y_i}{\lambda_{i|k}} - 1 \right) x_{i\ell} \lambda_{i|k} = \sum_{i=1}^n z_{ki} (y_i - \lambda_{i|k}) x_{i\ell}.$$

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1L} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nL} \end{pmatrix} \text{ ve } W_k = \begin{pmatrix} z_{k1}^0 \lambda_{1|k} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & z_{kn}^0 \lambda_{n|k} \end{pmatrix} \text{ olduğunda}$$

Fisher bilgi matrisi ise

$$-E \left[\frac{\partial^2 L(\beta)}{\partial \beta_{\ell k} \partial \beta_{\ell' k}} \right] = \sum_{i=1}^n z_{ki} \lambda_{i|k} x_{i\ell} x_{i\ell'} = \left\{ X' W_k X \right\}_{\ell \ell'} \quad (3.55)$$

Genel olarak, Newton-Raphson prosedürü $\ln \lambda_{i|k}$ i ayarlamak için ve sonra β_k tahmin etmek için kullanılır.

$y_k = (y_{1|k}, \dots, y_{n|k})'$ olduğunda

$$y_{i|k} = \ln \lambda_{i|k} + (y_i - \lambda_{i|k}) \frac{d \ln \lambda_{i|k}}{d \lambda_{i|k}} = \ln \lambda_{i|k} + (y_i - \lambda_{i|k}) \frac{1}{\lambda_{i|k}}. \quad (3.56)$$

$$X' W_k X \hat{\beta}_k = (X' W_k y_k)^{-1} \text{ olsun.}$$

$$\text{Ardından ise } \hat{\beta}_k = (X' W_k X)^{-1} X' W_k y_k. \quad (3.57)$$

Bu haliyle, EM algoritmini aşağıdaki gibi verebiliriz. EM algoritmi;

Adım 1. $2 \leq c \leq n$ ve $\varepsilon > 0$. sabitle. Başlangıç değerleri $z_k^{(0)}$ ve $\beta_k^{(0)}$ ve $s=1$ olsun.

Adım 2. $\beta_k^{(s-1)}$ ile $\ln y_{i|k}^{(s)}$ hesapla

Adım 3. $z_k^{(s-1)}$ ile $\alpha_k^{(s)}$ hesapla

Adım 4. $z_k^{(s-1)}$ ve $\lambda_{i|k}^{(s)}$ ile $W_k^{(s)}$ hesapla

Adım 5. $\lambda_{i|k}^{(s)}$ ile $y_k^{(s)}$ hesapla

Adım 6. $\beta_k^{(s-1)}$ u $\beta_k^{(s)}$ a ve $z_k^{(s-1)}$ i $z_k^{(s)}$ e güncelle

Adım 7. Eğer $\|z_k^{(s)} - z_k^{(s-1)}\| < \varepsilon$ duruyorsa ve $s = s+1$ ise adım 2 'ye dön.

Yaygın olarak kullanılan bağlantı (link) fonksiyonları Çizelge 3.1'de özetlenmiştir.

Çizelge 3.1. Üstel fonksiyon ailesinde bazı dağılımlar

Dağılım	Gösterimi	Dağılım fonksiyonu	Bağlantı
Binomial	$B(K, \mu)$	$\binom{K}{y} \left(\frac{\mu}{K}\right)^y \left(1 - \frac{\mu}{K}\right)^{K-y}$	$\ln(\mu / (K - \mu))$
Poisson	$p(\mu)$	$\frac{e^{-\mu} \mu^y}{y!}$	$\ln(\mu)$
Negatif binomial	$NB(\mu, v)$	$\binom{v}{y} \frac{r(v+y)}{y! r(v)} \left(\frac{\mu}{v+\mu}\right)^y$	$\ln(\mu / (v + \mu))$
Multinomial	$M(\mu)$	$\prod_{k=1}^K \mu_k^{y_k}$	$\ln(\mu / (\mu_k + \mu_K))$
Normal	$N(\mu, \sigma)$	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y - \mu)^2}{2\sigma^2}\right]$	μ
Multivariate normal	$MVN(\mu, \Sigma)$	$\frac{1}{(2\pi)^{K/2} \Sigma ^{1/2}} \exp\left[-1/2(y_n - \mu)' \Sigma^{-1} (y_n - \mu)\right]$	μ
Üstel	$E(\mu)$	$\frac{1}{\mu} \exp\left[-\frac{y}{\mu}\right]$	$\frac{1}{\mu}$
Gamma	$G(\mu, v)$	$\frac{1}{\Gamma(v)} \left(\frac{y}{\mu}\right)^{v-1} \exp\left[-\frac{y}{\mu}\right]$	$\frac{1}{\mu}$
Dirichlet	$D(\mu)$	$\frac{r\left(\sum_{k=1}^K \mu_k\right) \prod_{k=1}^K y_k^{\mu_k - 1}}{\prod_{k=1}^K \Gamma(\mu_k)}$	$\ln(\mu / (\mu_k + \mu_K))$

3.2.10 Poisson ve Karışımli Poisson Dağılımları

Kesit sayım veri analizinin başlangıç noktası Poisson regresyon modelidir. Poisson dağılımı, koşullu varyans ve koşullu ortalamanın (eşit dağılım) eşitliğini varsayar.

$$f(y_i | x_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, y_i = 0, 1, 2, \dots \quad (3.58)$$

Ortalama ve varyanslar ile

$$E[y_i | x_i] = V[y_i | x_i] = \mu_i = \exp(x_i \beta) \quad (3.59)$$

Poisson log-olabilirlik fonksiyonu;

$$L(\beta) = \sum_{i=1}^n [y_i x_i \beta - \mu_i - \log(y_i!)] \quad (3.60)$$

Poisson dağılımının ima ettiği eş dağılım çok kısıtlayıcıdır.

Genellikle y_i nin koşullu varyansı $\mu_i + \tau \mu_i^2$ e eşittir ve koşullu ortalamayı aşıyor. Pozitif bulaşma ve gözlemlenmemiş heterojenlik aşırı yayılıma neden olabilir. Bu sorunları ele almak için karışık Poisson modeli parametresine gözlemlenmemiş özel bir $\square_i$ etkisi dahil eder; bu özel etki rastgele veya sabit olarak ele alınabilir. Rastgele etki durumlarında, $\square_i$ ve $\square_i$ arasındaki ilişki

$$\square_i \mu_i = \exp(x_i \beta + \varepsilon_i) = \exp(x_i \beta) \exp(\varepsilon_i) = \mu_i v_i \quad (3.61)$$

olur. Rastgele $\square_i$ etkisi olası spesifikasyon hatalarını dikkate alır. Karışık Poisson modelinin dağılımının kesin biçimi, v_i 'nin olasılık dağılımının özel seçimine bağlıdır. v_i nin yoğunluğu $g(v_i)$ olsun, sonra şartlı yoğunluğu ise

$$f(y_i | x_i) = \int \exp(-\mu_i) \mu_i^{y_i} / r(y_i + 1) g(v_i) dv_i \quad (3.62)$$

olarak bulunur.



4. BULGULAR

Bu çalışmada, akar sayımları üzerine çeşit, sıcaklık ve nem etkilerine bağlı olarak veri setinin nasıl bir heterojen yapı sergilediğini saptamak için latent Poisson regresyon uygulanmıştır. Gerekli istatistiksel analizler Mplus ve SAS yazılım programı kullanılarak yapılmıştır. Veri setinin kendi içerisinde homojen olmak üzere kaç alt popülasyona ayrıldığını belirlemek için kullanılan AIC ve BIC uyum ölçütleri ile entropy doğru sınıflandırmaya ilişkin istatistikler Çizelge 4.1’ de verilmiştir.

En yüksek AIC ve BIC değerleri 1 alt popülasyonlu modelde sırasıyla sırasıyla 2163.415 ve 2238.341; en küçük ise beş alt popülasyonlu modelde sırasıyla 2071.646 ve 2137.106 olarak elde edilmişlerdir. 1 alt popülasyonlu model aynı zamanda Poisson regresyonunu göstermektedir. Böylece, AIC ve BIC bilgi ölçütlerine göre veri setinin kendi içlerinde homojen 4 alt popülasyondan oluşmuştur. AIC ve BIC ait değerlerin, 4 alt popülasyonlu modele kadar azaldığı daha sonra ise arttığı gözlenmiştir. Bu nedenle 5 ve 6 alt popülasyonlu modellerden sonra AIC ve BIC değerlerine yer verilmemiştir. Doğru sınıflandırma olasılığını gösteren entropi değeri en yüksek 4 alt popülasyonlu modelde %92.1 olarak bulunmuştur. Bununla birlikte en düşük entropy oranı 1 alt popülasyonlu modelde %76.3 olarak elde edilmiştir.

Dört alt popülasyonlu karışımli Poisson regresyonundaki her bir alt popülasyon için elde edilen entropy doğru sınıflandırma olasılıkları Çizelge 4.1’de verilmiştir. Gözlemlerin doğru sınıflandırma olasılıklarının yüksek olduğu saptanmıştır.

Çizelge 4.1 Latent Poisson regresyonu için uyum ölçütleri

Alt gruplar	Model bilgi ölçütleri		
	AIC	BIC	Entropy (%)
Bir alt popülasyonlu model	2163.415	2238.341	76.3
İki alt popülasyonlu model	2090.711	2154.914	89.6
Üç alt popülasyonlu model	2076.128	2144.034	90.0
Dört alt popülasyonlu model	2071.646	2137.106	92.1
Beş alt popülasyonlu model	2074.811	2142.318	90.2
Altı alt popülasyonlu model	2081.635	2149.021	90.1

Latent Poisson regresyon modeli kullanılarak elde edilen 4 alt popülasyonlu modelde, her bir alt gruba giren bireylerin sayısı ve karışma olasılıkları (mixing probability) sırasıyla 142(%71.3), 12(%6.20), 26(%13.3) ve 16(%8.20) olarak elde

edilmiştir (Çizelge 4.2). En fazla bireyin alt populasyon 1’de, en az bireyin ise alt populasyon 2’de olduğu belirlenmiştir.

Çizelge 4.2 Bağımlı değişkenin alt populasyonlara göre dağılımı

Alt populasyonlar	N	Oran
1	142	0.713
2	12	0.062
3	26	0.133
4	16	0.082

Veri seti için elde edilen en iyi model 4 alt populasyonlu modelden oluştuğu için bu model ilişkin parametre tahminleri Çizelge 4.3’te verilmiştir. Birinci alt populasyonda akar sayımları üzerine çeşit etkisinin istatistiksel olarak önemli ($p<0.01$); sıcaklık ve nem değişkenlerinin etkileri ise istatistiksel olarak önemsiz oldukları saptanmıştır. İkinci alt populasyonda, akar sayımları üzerine sıcaklık ve nem bağımsız değişkenlerinin etkileri istatistiksel olarak önemli ($p<0.01$); çeşit bağımsız değişkeninin etkisi istatistiksel olarak önemsiz olduğu saptanmıştır. Üçüncü alt populasyonda, akar sayımları üzerine nemin etkisi istatistiksel olarak önemli bulunmuşken ($p<0.05$); buna sıcaklık ve nem değişkenlerinin etkileri ise istatistiksel olarak önemsiz oldukları saptanmıştır. Dördüncü alt populasyonda akar sayımları üzerine sıcaklık ve nem etkileri istatistiksel olarak önemli bulunmuşken ($p<0.001$), çeşit etkisinin istatistiksel olarak önemsiz oldukları belirlenmiştir.

Çizelge 4.3 Alt gruplara ait parametre tahminleri ve standart hataları

Alt populasyonlar	β_0	Çeşit	Sıcaklık	Nem
1	22.344(3.828)	-3.858(1.103)**	-0.043(0.027)	-0.004(0.005)
2	259.292(81.799)	-6.080(25.363)	0.299(0.149)**	-0.302(0.038)***
3	140.839(41.168)	-15.109(11.435)	0.335(0.084)***	-0.130(0.020)**
4	441.207(20.371)	-1.163 (8.901)	1.736(0.124)***	-0.409(0.016)***

Dört alt populasyonlu latent Poisson regresyonundaki her bir alt populasyon için elde edilen entropy doğru sınıflandırma olasılıkları Çizelge 4.4’te verilmiştir. Gözlemlerin doğru sınıflandırma olasılıklarının yüksek olduğu saptanmıştır. Elde edilen alt populasyon 1, alt populasyon 2, alt populasyon 3 ve alt populasyon 4 için entropy doğru sınıflandırma oranları sırasıyla %97.3, %93.6, %95.4 ve %94.3 olarak elde

edilmiştir. Her bir alt grup için elde edilen entropy doğru sınıflandırma oranları oldukça yüksek elde edilmiştir.

Çizelge 4.4. Dört alt populasyon modelde her bir alt populasyon için tahmin edilen entropy doğru sınıflandırma oranları

	Alt gruplar			
	Alt populasyon 1	Alt populasyon 2	Alt populasyon 3	Alt populasyon 4
Alt populasyon 1	0.973	0.000	0.027	0.000
Alt populasyon 2	0.000	0.936	0.064	0.000
Alt populasyon 3	0.000	0.046	0.954	0.000
Alt populasyon 4	0.000	0.000	0.057	0.943

Bu çalışmada kullanılan değişkenlerin her bir alt populasyondaki ortalama değerleri aşağıdaki Çizelge 4.5'te verilmiştir. Akar sayımları bağımlı değişkenini en yüksek ortalaması dördüncü alt populasyonda (368.500), daha sonra sırasıyla üçüncü alt populasyonda (173.200), ikinci alt populasyonda (82.067) ve birinci alt populasyonda (10.947) olarak elde edilmiştir. Nem bağımsız değişkeni alt populasyonlara dağılımı sırasıyla; 53.53, 53.51, 51.95 ve 56.55 olarak elde edilmiştir. Sıcaklık bağımsız değişkeni 4 alt populasyona dağılımı sırasıyla; 17.77, 22.66, 25.25 ve 26.55 olarak elde edilmiştir (Çizelge 4.5).

Portakal çeşidinin alt populasyonlara dağılımı; birinci alt populasyonda 29, ikinci alt populasyonda 4, üçüncü alt populasyonda 10 ve dördüncü alt populasyonda 6 olmuştur. Greyfurt çeşidinin alt populasyonlara dağılımı; birinci alt populasyonda 32, ikinci alt populasyonda 16 olmuştur. Mandalina çeşidinin alt populasyonlara dağılımı; birinci alt populasyonda 29, ikinci alt populasyonda 4, üçüncü alt populasyonda 10 ve dördüncü alt populasyonda 6 olmuştur. Nar çeşidinin alt populasyonlara dağılımı; birinci alt populasyonda 43, ikinci alt populasyonda 6 olarak elde edilmiştir.

Çizelge 4.5 Değişkenlerin alt gruplara göre ortalamaları

Alt populasyon	N	Akar ortalaması	Çeşit (Sayı, %)				Nem	Sıcaklık
			Portakal	Greyfurt	Mandalina	Nar		
1	142	10.947	29(%59.2)	32(%66.7)	29(%59.2)	43(%87.8)	53.52	17.77
2	12	82.067	4(%8.2)	16(%33.3)	4(%8.2)	6(%12.2)	53.51	22.66
3	26	173.200	10(%20.4)	-	10(%20.4)	-	51.95	25.25
4	16	368.500	6(%12.24)	-	6(%12.24)	-	56.55	26.55

Akar sayımları normal dağılım göstermediğinden dolayı hem çeşitlere göre hem de latent sınıflar oluştuktan sonra sınıflara göre akar sayımları arasındaki farklılığı saptamak için Kruskal-Wallis analizi uygulanmıştır. Böyle her iki durum arasındaki farklılıklar saptanmıştır.

Akar sayımları bakımından çeşitler arasındaki farklılığı saptamak için uygulanan Kruskal-Wallis analizine göre, çeşitler arasındaki fark istatistiksel olarak önemli bulunmuştur ($p < 0.001$).

Çeşitler arasındaki farklı belirlemek için elde edilen Wilcoxon Z testinin dağılımı, medyan testinin dağılımı ve oransal dağılım sırasıyla Şekil 4.1, Şekil 4.2, Şekil 4.3, Çizelge 4.6 ve Çizelge 4.7’de verilmiştir.

Çizelge 4.6’de portakal, mandalina, greyfurt ve nar çeşitleri için akar sayımlarının ortalama skorları sırasıyla; 120.225, 115.163, 99.427 ve 57.214 olarak elde edilmiştir.

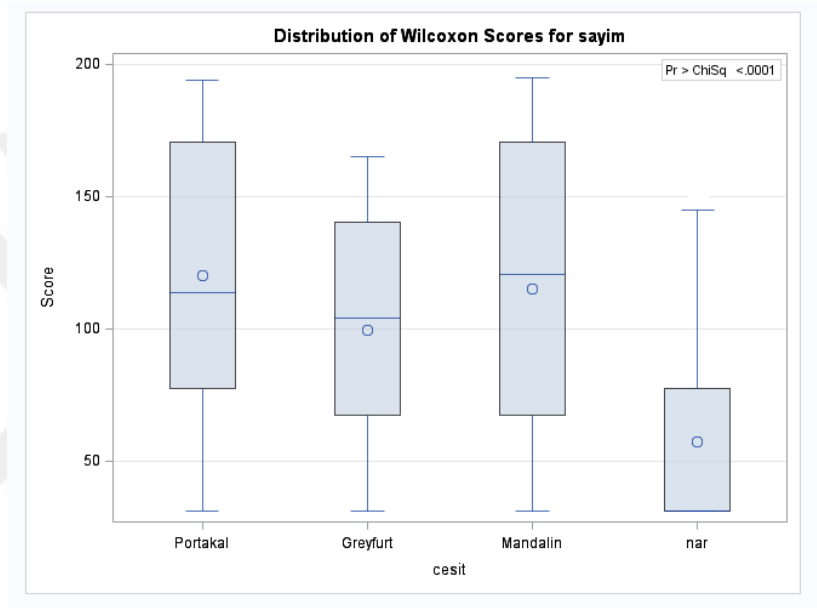
Çizelge 4.6 Kruskal- Wallis testi sonuçları

Çesit	N	Toplam Rank	Std Dev	Ortalama Rank
Portakal	49	5891.00	336.485	120.224
Greyfurt	48	4772.50	334.173	99.427
Mandalin	49	5643.00	336.485	115.163
Nar	49	2803.50	336.485	57.214

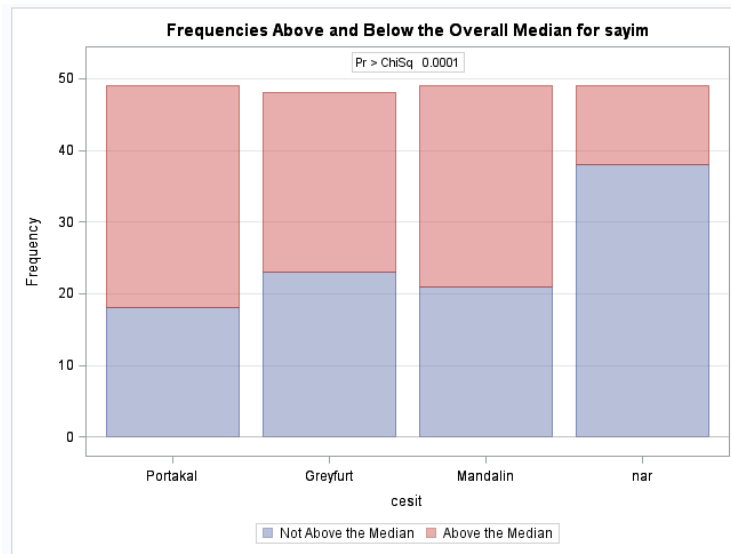
Çizelge 4.7. Dwass, Steel, Critchlow-Fligner Method (DSCF) çoklu karşılaştırma sonuçları

Karşılaştırmalar	Wilcoxon Z	DSCF Value	Pr > DSCF
Portakal-Greyfurt	2.0048	2.8353	0.1861
Portakal-Mandalin	0.4025	0.5692	0.9780
Portakal-nar	5.5615	7.8651	0.0001
Greyfurt-Mandalin	-1.5997	2.2623	0.3788
Greyfurt-nar	4.3052	6.0884	0.0001
Mandalin-nar	5.0446	7.1341	0.0001

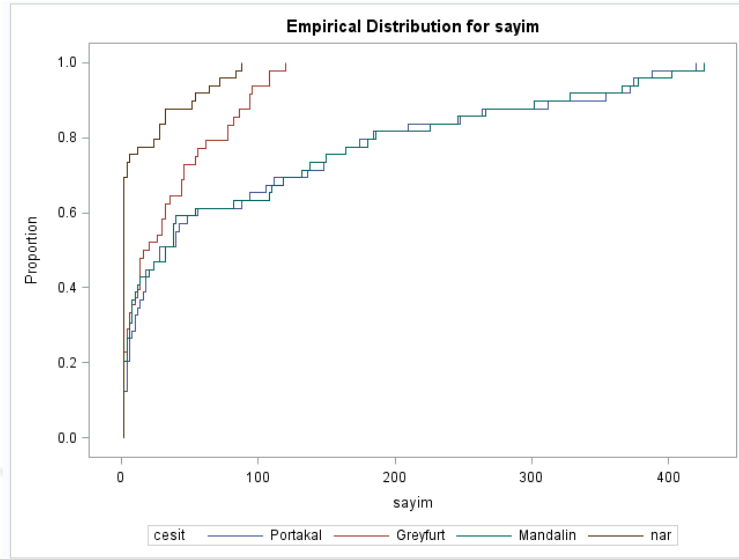
Hangi çeşitlerin birbirinden farklı olduğunu belirleme için yapılan DSFC testinin sonuçlarına bakıldığında (Çizelge 4.7), portakal ile nar; greyfurt ile nar ve mandalina ile nar arasındaki fark istatistiksel olarak önemli bulunmuşken, diğer ikili karşılaştırmalar önemsiz bulunmuştur. Bilindiği üzere Kruskal-Wallis gibi parametrik olmayan testler karşılaştırma yaparken ortalama ölçüsü olarak medyanı esas almaktadır. Her dört çeşidin akar sayımlarının medyan değerlerine göre elde edilen medyan dağılımları Şekil 4.1’de verilmiştir.



Şekil 4.1 Ortalama akar skor değerlerine göre çeşitlerin box-plot grafiği



Şekil 4.2 Her bir çeşidin medyan değerlerine göre dağılımı



Şekil 4.3 Akar sayımlarının çeşitlere göre dağılımı

Akar sayımlarına göre latent sınıflar arasındaki farklılığı saptamak için uygulanan Kruskal-Wallis analizine göre, çeşitler arasındaki fark istatistiksel olarak önemli bulunmuştur ($p < 0.001$).

Latent sınıflar arasındaki farklı belirlemek için elde edilen Wilcoxon Z testinin dağılımı, medyan testinin dağılımı ve oransal dağılım sırasıyla Şekil 4.4, Şekil 4.5, şekil 4.6, Çizelge 4.8 ve Çizelge 4.9’da verilmiştir. Çizelge 4.8’de latent 1, 2, 3 ve 4 sınıfları için akar sayımlarının ortalama skorları sırasıyla; 67.263, 148.883, 171.175 ve 189.500 olarak elde edilmiştir.

Çizelge 4.8 Alt populasyonlara göre Kruskal- Wallis testi sonuçları

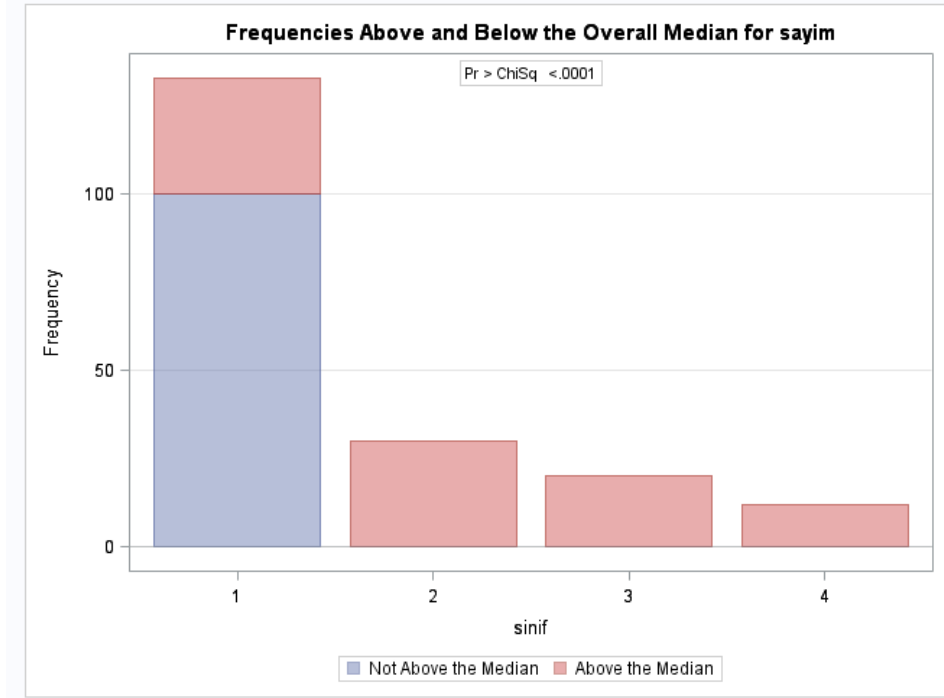
Alt populasyon	N	Toplam	Std Dev	Ortalama
1	142	8946.00	361.254	67.263
2	12	4466.50	279.894	148.883
3	26	3423.50	235.356	171.175
4	16	2274.00	186.426	189.500

Hangi latent sınıfların birbirinden farklı olduğunu belirleme için yapılan DSFC testinin sonuçlarına bakıldığında (Çizelge 4.9), elde edilen dört laten sınıfın ortalama akar skorlarının birbirinden farklı olduğu sağtanmıştır.

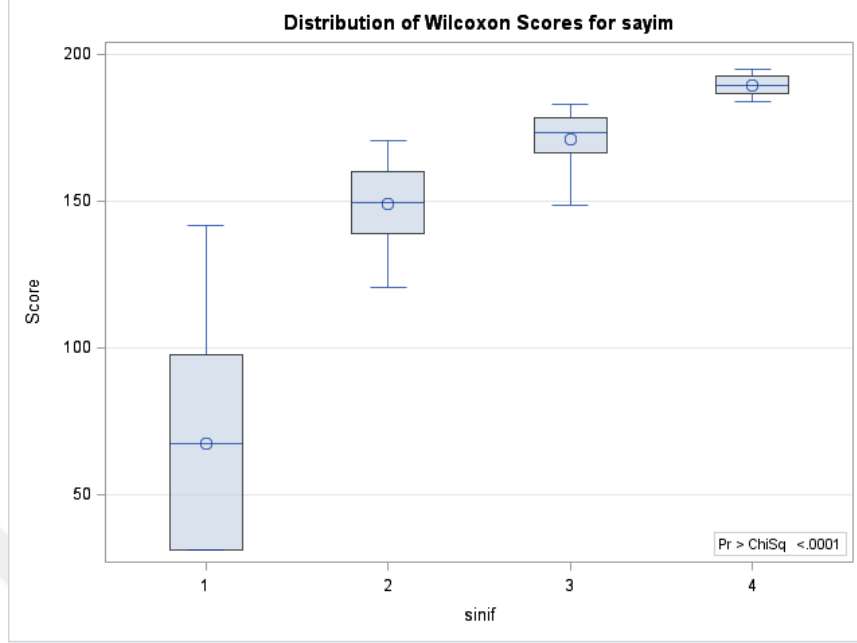
Çizelge 4.9 Dwass, Steel, Critchlow-Fligner Method çoklu karşılaştırma sonuçları

Alt popülasyon karşılaştırmaları	Wilcoxon Z istatistiği	DSCF istatistiği	Pr > DSCF
1-2	-8.6257	12.1985	<.0001
1-3	-7.4413	10.5235	<.0001
1-4	-5.9556	8.4225	<.0001
2-3	-5.0221	7.1023	<.0001
2-4	-5.0143	7.0912	<.0001
3-4	-4.6714	6.6064	<.0001

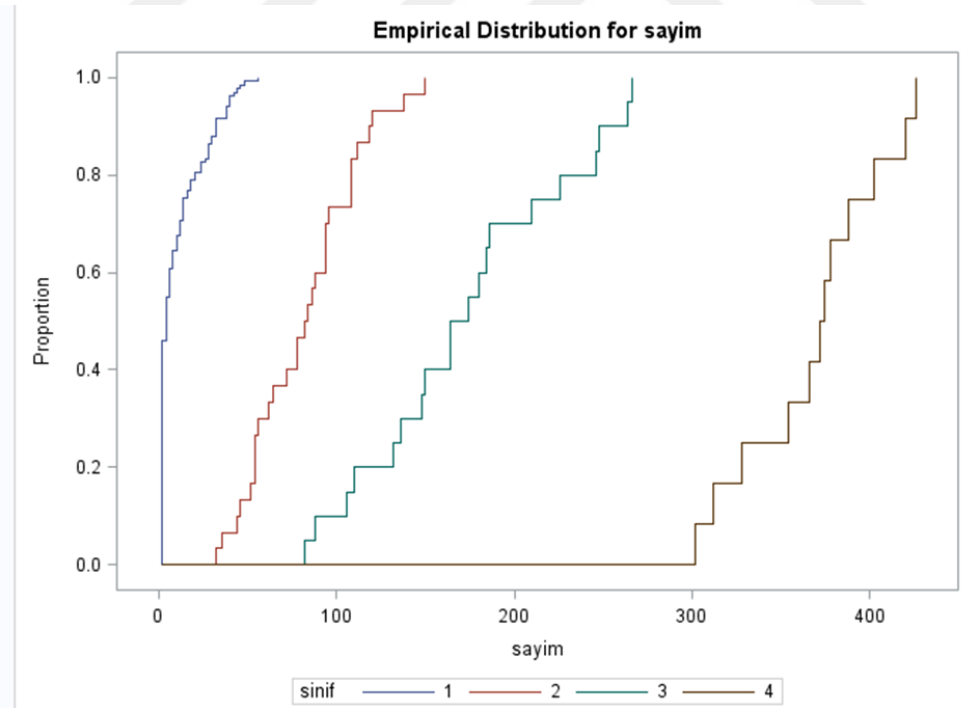
Elde edilen latent sınıflara göre, akar sayımlarının medyan değerlerine göre elde edilen dağılım Şekil 4.4’de; ortalama akar skorlarına göre box-plot dağılımı Şekil 4.5’de ve çeşitlere göre akar skorlarının yoğunluk dağılımı Şekil 4.6’da verilmiştir.



Şekil 4.4 Alt popülasyonlara göre, akar sayımlarının medyan değerlerine göre elde edilen dağılım



Şekil 4.5 Ortalama akar skorlarına göre box-plot dağılımı



Şekil 4.6 Alt populasyonlara göre akar skorlarının yoğunluk dağılımı

5. TARTIŞMA VE SONUÇ

Latent modeller verilerin elde edilme durumlarına göre farklı isimler almaktadır. Bağımlı değişkenin sürekli olması durumunda latent Gaussian regresyon; bağımlı değişkenin kesikli olması durumunda latent Poisson regresyon; bağımlı değişkenin binom dağılımı göstermesi durumunda latent binary lojistik regresyon ve bağımlı değişkenin çoklu binom dağılımı göstermesi durumunda ise latent multinominal regresyon olarak adlandırılmaktadır. Bu yönü ile latent modeller çok geniş bir istatistiksel alana sahiptir. Bunun yanı sıra kullanılan AIC, BIC, entropy model uyum ölçütleri de uygun modelin belirlenmesinde oldukça önemlidir (Wang vd., 1996; Yeşilova vd., 2006; 1998; Abut, 2019)

Çizelge 4.1’de verilen 1 alt populasyonlu modeldeki AIC ve BIC değerleri sırasıyla 2071.647 ve 2238.341 olarak elde edilmiştir. Bu model aynı zamanda Poisson regresyonunu göstermektedir. Hem AIC hem de BIC değerleri diğer alt populasyonlu modellere nazaran daha yüksek çıkmıştır. Bunun nedeninin akar sayımlarındaki varyasyonun büyük olmasıdır. Akar sayımlarının Poisson ortalaması 23.42 iken, Poisson varyansı 1251.284 olarak elde edilmiştir. Varyansın ortalmadan yüksek çıkması aşırı yayılıma neden olmuştur. Bu nedenle bir alt populasyonlu modelin uyum ölçütleri diğer alt populasyonlu modellere göre daha yüksek olmuştur. Buna paralel olarak entropy doğru sınıflandırma istatistiği de %76.3 olarak elde edilmiştir. Bu çizelge 4.1’deki en küçük entropy değeridir. Yapılan benzer çalışmalarda, aşırı yayılımlı modellerde AIC ve BIC değerleri çok yüksek, entropy istatistiği ise daha küçük olmaktadır (Wang vd., 1996; Wang ve Putterman, 1998)

Bu çalışmada farklı alt populasyon modelleri arasında, AIC, BIC, Log-L ve Entropy uyum ölçütleri kullanılarak dört alt populasyonlu modelin uygun olduğu tespit edilmiştir. Latent Poisson modellerin kullanıldığı durumlarda, en uygun modeli belirlemede söz konusu ölçütlerin birlikte değerlendirilmesi gerekmektedir. Entropy değerinin düşmeye başladığı noktada sınıflandırmaya son verilmelidir (Muthen ve Muthen, 2002; Roeder vd., 1999). Bundan dolayı dört alt populyasyondan sonraki model parametre tahminleri verilmemiştir. Dört alt populasyonlu modelde, AIC ve BIC bir alt populyasyona göre daha büyükken, log-L ölçütü daha küçük bulunmuştur. Entropy ölçütü dört alt populasyonlu modelde en büyük değerine ulaşmıştır. Bu durum verilerin

ne kadar doğru sınıflandırıldığını gösterir. Bundan dolayı, veri kümesinin dört alt popülasyonda toplandığı tespit edilmiştir. Burada, farklı olarak faktörler, zamana bağlı değişen ve zamana bağlı değişmeyen (risk faktörleri) olmak üzere iki farklı biçimde modele dahil edilmiştir. Bunlardan birinci grup, zamanla değişmeyen çeşit grubudur. Genel olarak cinsiyet, ırk, gibi demografik karakterler bu grubu temsil eder ve literatürde bu grup kovaryans risk faktörü olarak bilinir. Bu çalışmada çeşit faktörü olarak modele alınmıştır. Risk faktörler akar sayımlarından koşullu bağımsız olduğu varsayılmıştır. Başka bir ifade ile alt popülasyonların bilinmesi durumunda ($Y|C=k$) risk faktörü ile akar sayımlarını bağımsız kabul edilmektedir. Yani risk faktörü modele alınarak çok tepeli logit yardımı verilerin hangi olasılıkla hangi alt popülasyona girdiği belirlenmiştir. Bu değerlendirme şekli Roeder vd. (1999) yaklaşımı ile aynı paralelliktedir. İkinci alt popülasyon kovaryans grubu ise zamanla değişen etkilerden oluşmaktadır. Bu çalışmada sıcaklık ve nem bu grubu temsil etmektedirler. Bu kısım kovaryans grubu $\log(\lambda) = X'\beta$ modeli için kullanılmıştır. Risk faktörünün modelde olmaması durumunda BIC ve AIC ölçütleri daha büyük bulunmuştur. Yani modelin uyumu daha kötü olmuştur. Bu bulgular Roeder vd. (1999) tarafından da ortaya konulmuştur.

En iyi model olarak seçilen 4 alt popülasyonlu modelde, her bir alt popülasyonun doğru sınıflandırma olasılıklarına bakıldığında, birinci alt popülasyonda gözlemlerin %97.3'ü doğru sınıflandırmışken, %2.7'si üçüncü alt popülasyona; ikinci alt popülasyonda gözlemlerin %93.6'sı doğru sınıflandırmışken, %6.4'ü üçüncü alt popülasyona; üçüncü alt popülasyonda gözlemlerin %95.4'ü doğru sınıflandırmışken, %4.6'sı ikinci alt popülasyona ve dördüncü alt popülasyonda gözlemlerin %94.3'ü doğru sınıflandırmışken, %5.7'si üçüncü alt popülasyona dahil oldukları saptanmıştır. Doğru sınıflandırma olasılığının çok yüksek olması elde edilen latent sınıflarının ne derece doğru olduğunu göstermektedir (Muthen ve Muthen, 2002; Guang, 2004; Li vd., 2011).

Bu çalışmada, akar sayımları çok geniş bir aralıkta değiştiğinden dolayı Poisson varyansı Poisson ortalamasından oldukça büyük çıkmıştır. Bu gibi durumlarda dönüşüm uygulamak yetersiz kalabilmektedir. Hem sayım verisi olması hem de dönüşümün yetersiz kalması parametrik olmayan testleri gerekmektedir. Burada latent Poisson regresyonu uygulamadan önce Kruskal-Wallis testi kullanılarak, akar sayımları

bakımından 4 çeşit ortalaması karşılaştırılmıştır. Daha sonra, veri setinin heterojen bir yapı sergilediği varsayılarak, latent Poisson regresyonu uygulanmıştır. Bu regresyon sonucunda akar sayımlarına göre 4 alt populasyonlu model kabul edilmiştir. Sonuç olarak, her iki durumda yapılan istatistiksel analizler karşılaştırıldığında;

1. Akar sayımları bakımından her bir çeşit kendi içinde heterojen bir yapı gösterirken; latent Poisson regresyonda her bir latent sınıf kendi içinde homojenlik göstermiştir. Zaten latent sınıf regresyonunda amaç, kendi içinde homojen, aralarında ise heterojen yapı gösteren sınıfların elde edilmesidir.
2. Çeşitler DFSC testine tabi tutulduğunda, diğer çeşitler ile nar çeşidi arasındaki farklılıklar önemli bulunmuşken, latent Poisson regresyonunu sonucu en iyi model olarak kabul edilen dört alt sınıflı modelde, tüm sınıfların etkileri birbirinden farklılık göstermiştir. Bu, latent Poisson regresyonunun sonuçlarının daha keskin ve net olduğunu göstermektedir (Muthen ve Muthen, 2002; Yeşilova vd., 2008; Ng ve McLachlan, 2014)



KAYNAKLAR

- Abut S. (2019). *Normal dağılım gösteren verilerin analizinde Gaussian karışımli regresyonun kullanılması*. Doktora tezi, Van Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü, Van, Türkiye.
- Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 60(2), 255-265.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Anonim. (2021). Poisson Regression / Regression of Counts: Definition. Erişim tarihi: 4 Ocak 2023. Erişim adresi: <https://www.statisticshowto.com/poisson-regression/>
- Bouwmeester, S., Sijtsma, K., 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.
- Bozdogan, H., Sclove, S. L. (1984). Multi-sample cluster analysis using Akaike's information criterion. *Annals of the Institute of Statistical Mathematics*, 36(1), 163-180.
- Bozdogan, H. (1990). On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. *Communications in Statistics-Theory and Methods*, 19(1), 221-278.
- Bozdogan, H. (1993). Choosing the number of component clusters in the mixture-model using a new informational complexity criterion of the inverse-Fisher information matrix. In **Information and Classification: Concepts, Methods and Applications Proceedings of the 16th Annual Conference of the "Gesellschaft für Klassifikation eV" University of Dortmund, Berlin Heidelberg**.
- Celeux, G., Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13, 195-212.
- Coxe, S., West, S. G., Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of Personality Assessment*, 91(2), 121-136.
- Chen, Z., Kou, L. (2001). A Note on the Estimation of the Multinomial Logit Model with Random Effects. *The American Statistician*, 55(2), 89-95.
- Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1-22.
- Dhanavanthan, P. (2000). Estimation of the Parameters of Compound Intervened Poisson Distribution. *Biometrical Journal*, 42(3), 315-320.
- Duncan, T. E., Susan, S. C., Strycker, L. A., Okut, H. (2002). Chapter Addendum to an Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Applications. Growth Mixture Modeling of Adolescent Alcohol Use Data. **Oregon Research Institute**, Oregon.
- Everitt, B. S., Hand, D. J. (1981). Finite mixture distributions. **London: Chapman and Hall**.
- Famoye F., Wang W. W. (2004). Censored generalized Poisson regression model *Computational Statistics & Data Analysis*; 46, 547-560.
- Furman, W. D., Lindsay, B. G. (1994). Testing for the number of components in a mixture of normal distributions using moment estimators. *Computational Statistics & Data Analysis*, 17(5), 473-492.

- Gage, N. A., Larson, A., Sugai, G., Chafouleas, S. M. (2016). Student perceptions of school climate as predictors of office discipline referrals. *American Educational Research Journal*, 53(3), 492-515.
- Guang Sung H. (2004). *Gaussian mixture regression and classification*, Doctoral thesis. Rice University, Houston, Texas.
- Guo, J., Wall, M., Amemiya, Y. (2006). Latent class regression on latent factors. *Biostatistics*, 7(1), 145-163.
- Harel, O., Chung, H., Miglioretti, D. (2013). Latent class regression: Inference and estimation with two- stage multiple imputation. *Biometrical Journal*, 55(4), 541-553.
- Hynes, S., Greene, W. (2016). Preference heterogeneity in contingent behaviour travel cost models with on- site samples: A Random Parameter vs. a Latent Class Approach. *Journal of Agricultural Economics*, 67(2), 348-367.
- Jeroen, K. (2004). Latent Class Regression Analysis. **Workshop at the 24th Biennial Conference of SMABS**. Jena University.
- Karaca, A. G., Olmuş, H. (2018). Sıfır Değer Aralıklı Verilerin Analizinde Sıfır Değer Aralıklı Regresyon Modellerinin İncelenmesi. *Trakya Üniversitesi Sosyal Bilimler Dergisi*, 20(2), 105-118.
- Kayri, M., Okut, H. (2008). Scrutinization of risk analysis related to success of the private skill exam by using mixture logistic regression model. *H. U. Journal of Education*, 35, 227-239.
- Konishi, S., Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika*, 83(4), 875-890.
- Koehler, A. B., Murphree, E. S. (1988). A comparison of the Akaike and Schwarz criteria for selecting model order. *Journal of the royal statistical society: Series C (Applied Statistics)*, 37(2), 187-195.
- Kullback, S., Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22, 79-86. <http://dx.doi.org/10.1214/aoms/1177729694>
- Li, S., Li, J., Yan, B., Liu, Y., Li, L. (2011). The Monte Carlo simulation of CT based on flat panel detector. **4th International Congress on Image and Signal Processing, CISP**, Shangai, China.
- Lin, T. H., Tsai, M. H. (2022). Solving unobserved heterogeneity with latent class inflated Poisson regression model. *Journal of Applied Statistics*, 49(11), 2953-2963.
- Lindsay, B. G., Basak, P. (1993). Multivariate normal mixtures: a fast consistent method of moments. *Journal of the American Statistical Association*, 88(422), 468-476.
- Lipovetsky, S., Conklin, W. M. (2005). Latent class regression model in IRLS approach. *Mathematical and Computer Modelling*, 42(3-4), 301-312.
- Mahmoud M. M., Alderiny M. M. (2010). On Estimating Parameters of Censored Generalized Poisson Regression Model. *Applied Mathematical Sciences*, 4(13), 623-635.
- Mbotwa, J. L., Kamps, M. D., Baxter, P. D., Ellison, G. T., Gilthorpe, M. S. (2021). Latent class regression improves the predictive acuity and clinical utility of survival prognostication amongst chronic heart failure patients. *Plos One*, 16(5), e0243674.
- McLaren, G. D., Nathanson, M. H., Jacobs, A., Trevett, D., Thomson, W. (1991). Regulation of intestinal iron absorption and mucosal iron kinetics in hereditary

- hemochromatosis. *The Journal of Laboratory and Clinical Medicine*, 117(5), 390-401.
- Mufudza, C., Erol, H. (2016). Poisson mixture regression models for heart disease prediction. *Computational and Mathematical Methods in Medicine*, 2016, 4083089.
- Muthen, L. K., Muthen, B. (2002). *Mplus: User's Guide*. Los Angeles: CA, Muthen Muthen.
- Ng, S. K., McLachlan, G. J. (2014). Mixture of regression models with latent variables and sparse coefficient parameters. *Compstat*, 223-231.
- Pearson, K. (1894). Contributions to the Mathematical Theory of Evolution. *Philosophical Transactions of the Royal Society of London*, 185, 71-110. <http://www.jstor.org/stable/90667>
- Ramos-López, D., Masegosa, A. R., Salmerón, A., Rumí, R., Langseth, H., Nielsen, T. D., Madsen, A. L. (2018). Scalable importance sampling estimation of Gaussian mixture posteriors in Bayesian networks. *International Journal of Approximate Reasoning*, 100, 115-134.
- Okut, H., Duncan, E. T., Duncan, C. S., Strycker, A. L. (2002). Latent variable mixture modelling: analyzing mixture and the structural portion of model. **Joint Statistical Meetings (JSM)**, New York City.
- Orellana, R., Carvajal, R., Agüero, J. C. (2018). Maximum likelihood infinite mixture distribution estimation utilizing finite Gaussian mixtures. *IFAC-PapersOnLine*, 51(15), 706-711.
- Özmen, İ., Famoye, F. (2007). Count regression models with an application to zoological data containing structural zeros, *Journal of Data Science*, 5, 491- 502.
- Park, S., Kim, S., Ryoo, J. H. (2020). Latent class regression utilizing fuzzy clusterwise generalized structured component analysis. *Mathematics*, 8(11), 2076.
- Rabe-Hesketh, S., Skrondal, A. (2007). Multilevel and latent variable modeling with composite links and exploded likelihoods. *Psychometrika*, 72(2), 123.
- Roeder, K., Lynch, K. G., Nagin, D. S. (1999). Modeling uncertainty in latent class membership: A case study in criminology. *Journal of the American Statistical Association*, 94(447), 766-776.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461-464.
- Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 333-343.
- Soromenho, G. (1994). Comparing approaches for testing the number of components in a finite mixture model. *Computational Statistics*, 9(1), 65-78.
- Tarpey, T., Petkova, E. (2010). Latent regression analysis. *Statistical Modelling*, 10(2), 133-158.
- Titterton, D. M., Smith, A. F. M., Markov, U. E. (1985). Statistical Analyses of Finite Mixture Distributions. Chichester, U.K.: **John Willey and Sons**.
- Titterton, D. M. (1997). Introduction to Gelfand and Smith 1990 Sampling-based approaches to calculating marginal densities. *Breakthroughs in Statistics*, 519-550.
- Wang, P., Puterman, M. L., Cockburn, I., Le, N. (1996). Mixed Poisson regression models with covariate dependent rates. *Biometrics*, 381-400.

- Wang, P., Cockburn, L. M., Puterman, M. L. (1998). Analysis of patent data—a mixed-Poisson-regression-model approach. *Journal of Business & Economic Statistics*, 16(1), 27-41.
- Wang, P., Puterman, M. L. (1998). Mixed logistic regression models. *Journal of Agricultural, Biological, and Environmental Statistics*, 175-200.
- Wedel, M., DeSarbo, W. S., Bult, J. R., Ramaswamy, V. (1993). A latent class Poisson regression model for heterogeneous count data. *Journal of Applied Econometrics*, 8(4), 397-411.
- Weller, B. E., Bowen, N. K., Faubert, S. J. (2020). Latent class analysis: a guide to best practice. *Journal of Black Psychology*, 46(4), 287-311.
- Yamaguchi, K. (2000). Multinomial logit latent-class regression models: An analysis of the predictors of gender-role attitudes among Japanese women. *American Journal of Sociology*, 105(6), 1702-1740.
- Yang, M. S., Lai, C. Y. (2005). Mixture Poisson regression models for heterogeneous count data based on latent and fuzzy class analysis. *Soft Computing*, 9(7), 519-524.
- Yeşilova, A., Okut, H. (2003). *Biyolojik Çalışmalardan Elde Edilen Kategorik Verilere Karışık Poisson Regresyon Analizinin Uygulanması*, Doktora tezi. Van Yüzüncü Yıl Üniversitesi Fen bilimleri Enstitüsü, Van, Türkiye.
- Yeşilova, A., Yılmaz, A., Kaki, B. (2006). Norduz erkek kuzularının bazı kesikli üreme davranış özelliklerinin analizinde doğrusal olmayan regresyon modellerin kullanılması. *Yüzüncü Yıl Üniversitesi Ziraat Fakültesi Tarım Bilim Dergisi*, 16(2), 87-92.
- Yeşilova, A., Yılmaz, A., Kaki B. (2007). Norduz erkek kuzularının bazı kesikli üremedavranış öz elliklerinin analizinde doğrusal olmayan regresyon modellerin kullanılması. *Yüzüncü Yıl Üniversitesi Ziraat Fakültesi Tarım Bilim Dergisi*, 16(2), 87-92.
- Yeşilova, A., Okut H., Kaki B. (2008). Kategorik veriler için karışımli poisson ve karışımli lojistik regresyon yöntemlerin teorik özelliklerinin incelenmesi. *Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 13(1), 19-23.

EKLER

Ek 1. Akar sayımlarının alt populasyonlara göre karışma olasılıkları

çeşitler	Akar sayımı	nem	sıcaklık	Alt populasyon1	Alt populasyon2	Alt populasyon3	Alt populasyon4	latent sınıf
Portakal	10	80.04	12.09	0,768	0,218	0,014	0	1
Portakal	8	64.07	12.04	0,743	0,255	0,001	0	1
Portakal	4	64.01	14.06	0,88	0,12	0	0	1
Portakal	2	57.03	13.05	0,812	0,188	0	0	1
Portakal	16	66.08	13.02	0,714	0,285	0,002	0	1
Portakal	2	64.04	16.08	0,979	0,021	0	0	1
Portakal	12	65.04	12.09	0,734	0,265	0,001	0	1
Portakal	28	76.01	14.02	0,614	0,367	0,019	0	1
Portakal	4	61.03	15.04	0,917	0,083	0	0	1
Portakal	2	48.08	19.01	0,996	0,004	0	0	1
Portakal	4	00.03	14.09	0,931	0,069	0	0	1
Portakal	6	61.06	7.05	0,983	0,017	0	0	1
Portakal	6	00.03	19.01	0,999	0,001	0	0	1
Portakal	18	66.08	19.04	0,986	0,014	0	0	1
Portakal	2	40.08	23.04	1	0	0	0	1
Portakal	4	35.03	27.04	1	0	0	0	1
Portakal	10	00.03	21.02	1	0	0	0	1
Portakal	40	46.07	27.02	1	0	0	0	1
Portakal	40	65.01	26.06	1	0	0	0	1
Portakal	48	63.03	23.08	0,908	0,092	0	0	1
Portakal	42	65.09	26.09	1	0	0	0	1
Portakal	28	00.02	15.02	1	0	0	0	1
Portakal	24	64.03	26.02	1	0	0	0	1
Portakal	18	58.05	26.01	1	0	0	0	1
Portakal	18	40.07	1.01	1	0	0	0	1
Portakal	112	00.03	28.01	0	1	0	0	2
Portakal	94	68.07	28.04	0,001	0,999	0	0	2
Portakal	150	60.03	28.01	0	0,997	0,003	0	2
Portakal	210	67.02	31.12	0	0	1	0	3
Portakal	248	60.04	29.07	0	0	1	0	3
Portakal	264	33.09	31.05	0	0	1	0	3
Portakal	312	66.02	28.03	0	0	0	1	4
Portakal	372	61.05	29.01	0	0	0	1	4
Portakal	420	54.08	26.09	0	0	0	1	4
Portakal	388	39.09	28.08	0	0	0	1	4
Portakal	374	54.08	25.06	0	0	0	1	4
Portakal	354	65.02	23.03	0	0	0	1	4
Portakal	184	54.03	22.08	0	0	1	0	3
Portakal	174	43.01	22.06	0	0	1	0	3

Portakal	164	51.05	22.08	0	0	1	0	3
Portakal	148	48.06	17.01	0	0	1	0	3
Portakal	136	48.02	18.03	0	0	1	0	3
Portakal	106	40.08	16.07	0	0	1	0	3
Portakal	88	75.06	18.01	0	0,105	0,895	0	3
Portakal	56	83.06	17.02	0,125	0,77	0,106	0	2
Portakal	14	85.05	13.07	0,788	0,194	0,018	0	1
Portakal	6	43.06	11.03	0,741	0,259	0	0	1
Portakal	2	79.09	12.03	0,78	0,208	0,011	0	1
Portakal	2	68.08	13.05	0,839	0,161	0	0	1
Greyfurt	4	80.04	12.09	0,752	0,243	0,005	0	1
Greyfurt	2	64.07	12.04	0,738	0,262	0	0	1
Greyfurt	2	64.01	14.06	0,818	0,182	0	0	1
Greyfurt	2	57.03	13.05	0,753	0,247	0	0	1
Greyfurt	14	66.08	13.02	0,71	0,289	0,001	0	1
Greyfurt	4	64.04	16.08	0,929	0,071	0	0	1
Greyfurt	2	65.04	12.09	0,746	0,254	0	0	1
Greyfurt	4	76.01	14.02	0,809	0,191	0	0	1
Greyfurt	2	75.08	12.09	0,755	0,243	0,001	0	1
Greyfurt	2	00.03	14.09	0,876	0,124	0	0	1
Greyfurt	2	61.06	7.05	0,966	0,034	0	0	1
Greyfurt	8	00.03	19.01	0,991	0,009	0	0	1
Greyfurt	14	66.08	19.04	0,972	0,028	0	0	1
Greyfurt	12	40.08	23.04	1	0	0	0	1
Greyfurt	2	35.03	27.04	1	0	0	0	1
Greyfurt	2	00.03	21.02	1	0	0	0	1
Greyfurt	10	46.07	27.02	1	0	0	0	1
Greyfurt	44	65.01	26.06	0,989	0,011	0	0	1
Greyfurt	32	63.03	23.08	0,99	0,01	0	0	1
Greyfurt	30	65.09	26.09	1	0	0	0	1
Greyfurt	30	00.02	15.02	1	0	0	0	1
Greyfurt	20	64.03	26.02	1	0	0	0	1
Greyfurt	14	58.05	26.01	1	0	0	0	1
Greyfurt	16	40.07	1.01	1	0	0	0	1
Greyfurt	46	00.03	28.01	0,999	0,001	0	0	1
Greyfurt	56	68.07	28.04	0,988	0,012	0	0	1
Greyfurt	82	60.03	28.01	0,004	0,996	0	0	2
Greyfurt	78	67.02	31.12	0,096	0,904	0	0	2
Greyfurt	86	60.04	29.07	0,005	0,995	0	0	2
Greyfurt	96	33.09	31.05	0	1	0	0	2
Greyfurt	120	66.02	28.03	0	1	0	0	2
Greyfurt	108	61.05	29.01	0	1	0	0	2
Greyfurt	108	54.08	26.09	0	1	0	0	2
Greyfurt	94	39.09	28.08	0	1	0	0	2

Greyfurt	78	54.08	25.06	0,002	0,998	0	0	2
Greyfurt	94	65.02	23.03	0	1	0	0	2
Greyfurt	62	54.03	22.08	0,025	0,975	0	0	2
Greyfurt	54	43.01	22.06	0,062	0,938	0	0	2
Greyfurt	46	51.05	22.08	0,407	0,593	0	0	2
Greyfurt	44	48.06	17.01	0,169	0,831	0	0	2
Greyfurt	36	48.02	18.03	0,318	0,682	0	0	2
Greyfurt	32	40.08	16.07	0,347	0,653	0	0	2
Greyfurt	26	75.06	18.01	0,787	0,213	0	0	1
Greyfurt	14	83.06	17.02	0,922	0,078	0	0	1
Greyfurt	6	85.05	13.07	0,786	0,209	0,005	0	1
Greyfurt	6	43.06	11.03	0,793	0,207	0	0	1
Greyfurt	2	79.09	12.03	0,734	0,257	0,008	0	1
Greyfurt	2	68.08	13.05	0,771	0,229	0	0	1
Mandalin	2	80.04	12.09	0,733	0,265	0,003	0	1
Mandalin	2	64.07	12.04	0,753	0,247	0	0	1
Mandalin	2	64.01	14.06	0,756	0,244	0	0	1
Mandalin	4	57.03	13.05	0,735	0,265	0	0	1
Mandalin	12	66.08	13.02	0,74	0,259	0,001	0	1
Mandalin	8	64.04	16.08	0,816	0,184	0	0	1
Mandalin	8	65.04	12.09	0,75	0,25	0	0	1
Mandalin	24	76.01	14.02	0,66	0,329	0,011	0	1
Mandalin	2	75.08	12.09	0,735	0,264	0,001	0	1
Mandalin	2	78.09	16.03	0,879	0,121	0	0	1
Mandalin	4	61.03	15.04	0,769	0,231	0	0	1
Mandalin	10	48.08	19.01	0,877	0,123	0	0	1
Mandalin	2	00.03	14.09	0,801	0,199	0	0	1
Mandalin	2	66.08	19.04	0,981	0,019	0	0	1
Mandalin	6	40.08	23.04	1	0	0	0	1
Mandalin	2	35.03	27.04	1	0	0	0	1
Mandalin	4	00.03	21.02	0,997	0,003	0	0	1
Mandalin	14	46.07	27.02	1	0	0	0	1
Mandalin	6	65.01	26.06	1	0	0	0	1
Mandalin	40	63.03	23.08	0,77	0,23	0	0	1
Mandalin	38	65.09	26.09	0,995	0,005	0	0	1
Mandalin	32	00.02	15.02	0,999	0,001	0	0	1
Mandalin	38	64.03	26.02	0,987	0,013	0	0	1
Mandalin	38	58.05	26.01	0,974	0,026	0	0	1
Mandalin	32	40.07	1.01	1	0	0	0	1
Mandalin	108	00.03	28.01	0	1	0	0	2
Mandalin	118	68.07	28.04	0	1	0	0	2
Mandalin	138	60.03	28.01	0	1	0	0	2
Mandalin	226	67.02	31.12	0	0	1	0	3
Mandalin	246	60.04	29.07	0	0	1	0	3

Mandalin	266	33.09	31.05	0	0	1	0	3
Mandalin	302	66.02	28.03	0	0	0	1	4
Mandalin	402	61.05	29.01	0	0	0	1	4
Mandalin	426	54.08	26.09	0	0	0	1	4
Mandalin	378	39.09	28.08	0	0	0	1	4
Mandalin	366	54.08	25.06	0	0	0	1	4
Mandalin	328	65.02	23.03	0	0	0	1	4
Mandalin	186	54.03	22.08	0	0	1	0	3
Mandalin	180	43.01	22.06	0	0	1	0	3
Mandalin	164	51.05	22.08	0	0	1	0	3
Mandalin	150	48.06	17.01	0	0	1	0	3
Mandalin	132	48.02	18.03	0	0	1	0	3
Mandalin	110	40.08	16.07	0	0	1	0	3
Mandalin	82	75.06	18.01	0	0,023	0,977	0	3
Mandalin	54	83.06	17.02	0,127	0,703	0,17	0	2
Mandalin	20	85.05	13.07	0,683	0,279	0,037	0	1
Mandalin	8	43.06	11.03	0,879	0,121	0	0	1
Mandalin	2	79.09	12.03	0,733	0,26	0,007	0	1
Mandalin	2	68.08	13.05	0,738	0,262	0	0	1
nar	2	80.04	12.09	0,755	0,243	0,002	0	1
nar	2	64.07	12.04	0,81	0,19	0	0	1
nar	2	64.01	14.06	0,735	0,265	0	0	1
nar	2	57.03	13.05	0,766	0,234	0	0	1
nar	2	66.08	13.02	0,766	0,234	0	0	1
nar	2	64.04	16.08	0,795	0,205	0	0	1
nar	2	65.04	12.09	0,782	0,218	0	0	1
nar	2	76.01	14.02	0,737	0,263	0	0	1
nar	2	75.08	12.09	0,763	0,236	0,001	0	1
nar	2	78.09	16.03	0,804	0,196	0	0	1
nar	2	61.03	15.04	0,737	0,263	0	0	1
nar	2	48.08	19.01	0,879	0,121	0	0	1
nar	2	00.03	14.09	0,748	0,252	0	0	1
nar	2	61.06	7.05	0,837	0,163	0	0	1
nar	2	00.03	19.01	0,952	0,048	0	0	1
nar	2	66.08	19.04	0,944	0,056	0	0	1
nar	2	40.08	23.04	0,999	0,001	0	0	1
nar	2	35.03	27.04	1	0	0	0	1
nar	2	00.03	21.02	0,99	0,01	0	0	1
nar	2	46.07	27.02	1	0	0	0	1
nar	2	65.01	26.06	1	0	0	0	1
nar	2	63.03	23.08	0,999	0,001	0	0	1
nar	2	65.09	26.09	1	0	0	0	1
nar	2	00.02	15.02	1	0	0	0	1
nar	4	64.03	26.02	1	0	0	0	1

nar	4	58.05	26.01	1	0	0	0	1
nar	6	40.07	1.01	1	0	0	0	1
nar	24	00.03	28.01	1	0	0	0	1
nar	28	68.07	28.04	1	0	0	0	1
nar	32	60.03	28.01	0,998	0,002	0	0	1
nar	12	67.02	31.12	1	0	0	0	1
nar	32	60.04	29.07	1	0	0	0	1
nar	28	33.09	31.05	1	0	0	0	1
nar	64	66.02	28.03	0,1	0,9	0	0	2
nar	72	61.05	29.01	0,013	0,987	0	0	2
nar	84	54.08	26.09	0	1	0	0	2
nar	88	39.09	28.08	0	1	0	0	2
nar	54	54.08	25.06	0,082	0,918	0	0	2
nar	52	65.02	23.03	0,081	0,919	0	0	2
nar	2	54.03	22.08	0,996	0,004	0	0	1
nar	2	43.01	22.06	0,991	0,009	0	0	1
nar	2	51.05	22.08	0,996	0,004	0	0	1
nar	2	48.06	17.01	0,761	0,239	0	0	1
nar	2	48.02	18.03	0,833	0,167	0	0	1
nar	2	40.08	16.07	0,728	0,272	0	0	1
nar	2	75.06	18.01	0,895	0,105	0	0	1
nar	2	83.06	17.02	0,868	0,132	0	0	1
nar	2	85.05	13.07	0,735	0,264	0,002	0	1
nar	2	43.06	11.03	0,919	0,081	0	0	1

Ek 2. Mplus çıktıları

Mplus VERSION 7.4
MUTHEN & MUTHEN
05/02/2023 1:39 PM
INPUT INSTRUCTIONS
INPUT READING TERMINATED NORMALLY
SUMMARY OF ANALYSIS
Number of groups 1
Number of observations 195
Number of dependent variables 1
Number of independent variables 3
Number of continuous latent variables 0
Number of categorical latent variables 1
Observed dependent variables
Continuous
SAYIM
Observed independent variables
CESIT NEM TEMP
Categorical latent variables
C
Estimator MLR
Information matrix OBSERVED
Optimization Specifications for the Quasi-Newton Algorithm for
Continuous Outcomes
Maximum number of iterations 2000
Convergence criterion 0.100D-05
Optimization Specifications for the EM Algorithm
Maximum number of iterations 1000
Convergence criteria
Loglikelihood change 0.100D-06
Relative loglikelihood change 0.100D-05
Derivative 0.100D-05
Optimization Specifications for the M step of the EM Algorithm for
Categorical Latent variables
Number of M step iterations 2
M step convergence criterion 0.100D-02
Basis for M step termination ITERATION
Optimization Specifications for the M step of the EM Algorithm for
Censored, Binary or Ordered Categorical (Ordinal), Unordered
Categorical (Nominal) and Count Outcomes
Number of M step iterations 2
M step convergence criterion 0.100D-02
Basis for M step termination ITERATION
Maximum value for logit thresholds 15
Minimum value for logit thresholds -15
Minimum expected cell size for chi-square 0.100D-01
Optimization algorithm EMA
Integration Specifications
Type STANDARD
Number of integration points 15
Dimensions of numerical integration 0
Adaptive quadrature ON
Cholesky OFF
Input data file(s)
D:\nur.dat
Input data format FREE
SAMPLE STATISTICS

SAMPLE STATISTICS

Means				
	SAYIM	CESIT	NEM	TEMP
1	60.533	2.503	60.911	21.359
Covariances				
	SAYIM	CESIT	NEM	TEMP
SAYIM	9535.182			
CESIT	-23.971	1.255		
NEM	-272.161	0.017	168.377	
TEMP	266.510	0.007	-33.603	39.036
Correlations				
	SAYIM	CESIT	NEM	TEMP
SAYIM	1.000			
CESIT	-0.219	1.000		
NEM	-0.215	0.001	1.000	
TEMP	0.437	0.001	-0.414	1.000

UNIVARIATE SAMPLE STATISTICS

UNIVARIATE HIGHER-ORDER MOMENT DESCRIPTIVE STATISTICS

Variable/ Sample Size	Mean/ Variance	Skewness/ Kurtosis	Minimum/ Maximum	% with Min/Max	Percentiles		
					20%/60%	40%/80%	Median
SAYIM	60.533	2.196	2.000	31.28%	2.000	6.000	14.000
195.000	9535.182	4.172	426.000	0.51%	30.000	94.000	
CESIT	2.503	-0.006	1.000	25.13%	1.000	2.000	3.000
195.000	1.255	-1.365	4.000	25.13%	3.000	4.000	
NEM	60.911	-0.229	33.900	2.05%	48.200	60.300	64.100
195.000	168.377	-0.631	85.500	2.05%	65.100	69.000	
TEMP	21.359	-0.043	11.300	2.05%	14.200	18.300	22.600
195.000	39.036	-1.442	32.100	2.05%	24.220	28.000	

THE MODEL ESTIMATION TERMINATED NORMALLY

MODEL FIT INFORMATION

Number of Free Parameters	20
Loglikelihood	
H0 Value	-965.121
H0 Scaling Correction Factor	1.6199
for MLR	
Information Criteria	
Akaike (AIC)	1970.243
Bayesian (BIC)	2035.703
Sample-Size Adjusted BIC	1972.346
(n* = (n + 2) / 24)	

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES BASED ON THE ESTIMATED MODEL

Latent Classes		
1	119.90644	0.61490
2	42.78871	0.21943
3	20.30486	0.10413
4	12.00000	0.06154

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON ESTIMATED POSTERIOR PROBABILITIES

Latent Classes		
1	119.90646	0.61490
2	42.78870	0.21943
3	20.30484	0.10413
4	12.00000	0.06154

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

Class Counts and Proportions

Latent Classes		
1	133	0.68205
2	30	0.15385
3	20	0.10256
4	12	0.06154

CLASSIFICATION QUALITY

Entropy 0.840

Average Latent Class Probabilities for Most Likely Latent Class Membership (Row)
by Latent Class (Column)

	1	2	3	4
1	0.887	0.112	0.001	0.000
2	0.066	0.925	0.009	0.000
3	0.000	0.006	0.994	0.000
4	0.000	0.000	0.000	1.000

Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)

	1	2	3	4
1	0.984	0.016	0.000	0.000
2	0.348	0.649	0.003	0.000
3	0.008	0.014	0.979	0.000
4	0.000	0.000	0.000	1.000

Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)

	1	2	3	4
1	13.799	9.705	0.954	0.000
2	12.761	13.383	8.004	0.000
3	8.937	9.527	13.794	0.000
4	-13.816	-13.816	-13.816	0.000

MODEL RESULTS

	Estimate	S.E.	Two-Tailed Est./S.E.	P-Value
Latent Class 1				
SAYIM	ON			
CESIT	-3.580	0.833	-4.300	0.000
NEM	0.276	0.098	2.812	0.005
TEMP	1.468	0.208	7.050	0.000
Intercepts				
SAYIM	-25.567	9.520	-2.686	0.007
Residual Variances				
SAYIM	227.295	37.474	6.065	0.000

Latent Class 2

SAYIM	ON				
CESIT		-8.430	3.267	-2.580	0.010
NEM		0.319	0.254	1.255	0.210
TEMP		6.189	0.478	12.959	0.000
Intercepts					
SAYIM		-76.645	22.736	-3.371	0.001
Residual Variances					
SAYIM		227.295	37.474	6.065	0.000

Latent Class 3

SAYIM	ON				
CESIT		0.859	3.615	0.238	0.812
NEM		-1.106	0.536	-2.062	0.039
TEMP		9.922	1.019	9.738	0.000
Intercepts					
SAYIM		1.955	24.649	0.079	0.937
Residual Variances					
SAYIM		227.295	37.474	6.065	0.000

Latent Class 4

SAYIM	ON				
CESIT		-1.500	9.176	-0.163	0.870
NEM		-2.289	1.072	-2.134	0.033
TEMP		-0.165	3.989	-0.041	0.967
Intercepts					
SAYIM		506.555	155.516	3.257	0.001
Residual Variances					
SAYIM		227.295	37.474	6.065	0.000

Categorical Latent Variables

Means					
C#1		2.302	0.307	7.487	0.000
C#2		1.271	0.348	3.650	0.000
C#3		0.526	0.369	1.425	0.154

QUALITY OF NUMERICAL RESULTS

Condition Number for the Information Matrix 0.121E-06
(ratio of smallest to largest eigenvalue)

CONFIDENCE INTERVALS OF MODEL RESULTS

	Lower .5%	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%	Upper .5%
--	-----------	------------	----------	----------	----------	------------	-----------

Latent Class 1

SAYIM	ON						
CESIT		-5.725	-5.212	-4.950	-3.580	-2.210	-1.435
NEM		0.023	0.083	0.114	0.276	0.437	0.528
TEMP		0.931	1.060	1.125	1.468	1.810	2.004
Intercepts							
SAYIM		-50.088	-44.225	-41.227	-25.567	-9.907	-6.908
Residual Variances							
SAYIM		130.771	153.847	165.651	227.295	288.939	300.743

Latent Class 2

SAYIM	ON						
CESIT		-16.845	-14.834	-13.805	-8.430	-3.056	-2.027
NEM		-0.336	-0.179	-0.099	0.319	0.737	0.817
TEMP		4.959	5.253	5.404	6.189	6.975	7.126
Intercepts							
SAYIM		-135.208	-121.207	-114.045	-76.645	-39.244	-32.082
Residual Variances							
SAYIM		130.771	153.847	165.651	227.295	288.939	300.743

Latent Class 3

SAYIM ON

CESIT	-8.451	-6.225	-5.087	0.859	6.805	7.944	10.170
NEM	-2.487	-2.157	-1.988	-1.106	-0.223	-0.054	0.276
TEMP	7.298	7.925	8.246	9.922	11.598	11.919	12.547

Intercepts

SAYIM	-61.536	-46.357	-38.593	1.955	42.503	50.267	65.446
-------	---------	---------	---------	-------	--------	--------	--------

Residual Variances

SAYIM	130.771	153.847	165.651	227.295	288.939	300.743	323.820
-------	---------	---------	---------	---------	---------	---------	---------

Latent Class 4

SAYIM ON

CESIT	-25.136	-19.485	-16.594	-1.500	13.595	16.485	22.136
NEM	-5.051	-4.391	-4.053	-2.289	-0.524	-0.186	0.474
TEMP	-10.440	-7.984	-6.727	-0.165	6.397	7.654	10.110

Intercepts

SAYIM	105.977	201.743	250.731	506.555	762.378	811.366	907.132
-------	---------	---------	---------	---------	---------	---------	---------

Residual Variances

SAYIM	130.771	153.847	165.651	227.295	288.939	300.743	323.820
-------	---------	---------	---------	---------	---------	---------	---------

Categorical Latent Variables

Means

C#1	1.510	1.699	1.796	2.302	2.808	2.904	3.094
C#2	0.374	0.589	0.698	1.271	1.844	1.954	2.169
C#3	-0.425	-0.197	-0.081	0.526	1.133	1.249	1.477

RESIDUAL OUTPUT

ESTIMATED MODEL AND RESIDUALS (OBSERVED - ESTIMATED) FOR CLASS 1

Model Estimated Means

SAYIM

1 12.039

Residuals for Means

SAYIM

1 0.000

Model Estimated Covariances

SAYIM

SAYIM 304.589

Residuals for Covariances

SAYIM

SAYIM -107.523

ESTIMATED MODEL AND RESIDUALS (OBSERVED - ESTIMATED) FOR CLASS 2

Model Estimated Means

SAYIM

1 57.675

Residuals for Means

SAYIM

1 0.000

Model Estimated Covariances

SAYIM

SAYIM	1869.813
Residuals for Covariances	
SAYIM	

SAYIM	77.291
ESTIMATED MODEL AND RESIDUALS (OBSERVED - ESTIMATED) FOR CLASS 3	
Model Estimated Means	
SAYIM	

1	170.924
Residuals for Means	
SAYIM	

1	0.000
Model Estimated Covariances	
SAYIM	

SAYIM	3313.328
Residuals for Covariances	
SAYIM	

SAYIM	9.278
ESTIMATED MODEL AND RESIDUALS (OBSERVED - ESTIMATED) FOR CLASS 4	
Model Estimated Means	
SAYIM	

1	368.500
Residuals for Means	
SAYIM	

1	0.000
Model Estimated Covariances	
SAYIM	

SAYIM	638.998
Residuals for Covariances	
SAYIM	

SAYIM	783.085	
+00	0.000E+00	0.000E+00



ÖZ GEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : Nur AYCAN

Eğitim Bilgileri

Lisans : Tıp
Üniversite : İstanbul Üniversitesi
Fakülte : Cerrahpaşa Tıp Fakültesi
Bölüm : Tıp
Mezuniyet Yılı : 2005

Yüksek Lisans :
Üniversite : Van Yüzüncü Yıl Üniversitesi
Enstitü : Fen Bilimleri Enstitüsü
Anabilim Dalı : Zootečni
Mezuniyet Yılı : 2023

Akademik Yayınlar:

- Aycan, A., Aktas, O., Guzey, F., Tufan, A., İşler, C., Aycan, N., et al. (2016). Rapidly Progressive Spontaneous Spinal Epidural Abscess. *Case Reports in Infectious Diseases*, 2016, 7958291.
- Aycan, A., Yener, U., Aycan, N., Gönüllü, E., Dursun, R., Gönüllü, H. (2015). Neurosurgical Injuries Caused by the 2011 Van Earthquake: The Experience at the Van Regional Training and Research Hospital. *The Journal of Emergency Medicine*, 49(4), 464-70.
- Ozkaya, E., Cambaz, N., Kolsuz, L. D., Aycan, N., Calis, S., Samanci, N. (2011). Vaccination coverage and risk factors for incomplete vaccination in children with recurrent wheeze. *Allergologia et Immunopathologia*, 39(4), 222-227.
- Ozkaya, E., Eker, H. H., Aycan, N., Samanci, N. (2010). Impact of maternal anxiety level on the childhood vaccination coverage. *European Journal of Pediatrics*, 169(11), 1397-1401.
- Taparlı, O. E., Shahi, P. K., Cagatay, N. S., Aycan, N., Ozaydin, B., Yapici, S., et al. (2022). Selectively compromised inner retina function following hypoxic-ischemic encephalopathy in mice: A noninvasive measure of severity of the injury. *Neurochemistry international*, 105471.
- Taş, A., Aycan, N., Aycan, A., Zengin, İ., Akyol, M. E. (2022). Povidone-iodine antisepsis in lumbar disc surgery. *Eastern Journal of Medicine*, 27(4), 549-551.
- Zafer, D., Aycan, N., Ozaydin, B., Kemanli, P., Ferrazzano, P., Levine J. E., et al. (2019). Sex differences in Hippocampal Memory and Learning following neonatal brain injury: is there a role for estrogen receptor-alpha? *Neuroendocrinology*, 109(3), 249-256.



VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
LİSANSÜSTÜ TEZ ORJİNALLİK RAPORU

Tarih/...../.....

Tez Başlığı: SAYIM İLE ELDE EDİLEN BAĞIMLI DEĞİŞKENDE
GÖZLENEMEYEN HETEROJENLİĞİN LATENT SINIF POISSON
REGRESYONU KULLANILARAK MODELLENMESİ

Yukarıda başlığı belirtilen tez çalışmamın, kapak sayfası, giriş, ana bölümler ve sonuç bölümlerinden oluşan toplam (.....) sayfalık kısmına ilişkin, .../...../..... tarihinde şahsım/tez danışmanım tarafından adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre tezimin benzerlik oranı %..... (.....) dır.

Uygulanan filtreler aşağıda verilmiştir:

- Kabul ve onay sayfası hariç,
- Teşekkür hariç,
- İçindekiler hariç,
- Simge ve kısaltmalar hariç,
- Gereç ve yöntemler hariç,
- Kaynakça hariç,
- Alıntılar hariç,
- Tezden çıkan yayınlar hariç,
- 7 kelimedenden daha az örtüşme içeren metin kısımları hariç (Limit match size to 7 words)

Van Yüzüncü Yıl Üniversitesi Lisansüstü Tez Orijinallik Raporu Alınması ve Kullanılmasına İlişkin Yönergeyi inceledim ve bu yönergede belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini bilgilerinize arz ederim.

Tarih ve İmza

Adı Soyadı: Nur AYCAN.....

Öğrenci No:

Anabilim Dalı: ...Zootečni.....

Programı: Zootečni.....

Statüsü: (x) Yüksek lisans () Doktora

DANIŞMAN
UYGUNDUR

ENSTİTÜ ONAYI
UYGUNDUR