



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

BOYLAMSAL VERİLERDE ÇOK DÜZEYLİ DOĞRUSAL REGRESYON VE KANTİL
REGRESYON YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Begüm KAFKAS

Doktora Tezi

Ankara, 2023

Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

Daha ileriye... En İyiyeye...



Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

BOYLAMSAL VERİLERDE ÇOK DÜZEYLİ DOĞRUSAL REGRESYON VE KANTİL
REGRESYON YÖNTEMLERİNİN KARŞILAŞTIRILMASI
COMPARISON OF MULTILEVEL LINEAR REGRESSION AND MULTILEVEL QUANTILE
REGRESSION METHODS IN LONGITUDINAL DATA

Begüm KAFKAS

Doktora Tezi

Ankara, 2023

Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

Beg¼m KAFKAS'ın hazırladıđı “Boylamsal Verilerde ok D¼zeyli Dođrusal Regresyon ve Kantil Regresyon Y¼ntemlerinin Karşılaştıırılması” bařlıklı bu alıřma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitim ¼lme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiřtir.

J¼ri Bařkanı

İmza

J¼ri Üyesi (Danıřman)

İmza

J¼ri Üyesi

İmza

J¼ri Üyesi

İmza

J¼ri Üyesi

İmza

Bu tez Hacettepe ¼niversitesi Lisans¼st¼ Eđitim, ¼đretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri ¼yeleri tarafından 20/ 06 / 2023 tarihinde uygun g¼r¼lm¼ř ve Enstit¼ Y¼netim Kurulunca / / tarihi itibarıyla kabul edilmiřtir.

Prof. Dr. İsmail Hakkı MİRİCİ
Eđitim Bilimleri Enstitüsü M¼d¼r¼

Öz

Klasik basit doğrusal regresyon analizi, gözlemlerin bağımsız ve özdeş olduğu varsayımına dayanmaktadır. Bu varsayım hiyerarşik veriler gibi durumlarda karşılanmayabilir. Hiyerarşik verilerde gözlemlerin birbirinden bağımsız olması ilkesinin ihlal ediliyor olması regresyon analizinde yanlı sonuçların ortaya çıkmasına neden olmaktadır. Bu tür sorunların giderilmesi için çok düzeyli modelleme yöntemlerine başvurulmaktadır. Ancak dağılımın çarpık olduğu durumlarda modellemelerde kullanılan yöntemlerin verimliliğinin düşük olması nedeniyle alternatif yöntem olarak kullanılan kantil regresyon modellemeleri geliştirilmiştir. Klasik regresyon koşullarının sağlanmadığı durumlarda avantajlı olan kantil regresyon modelleriyle dağılımın her noktası hakkında bilgi edinilebilmektedir. Bu araştırmada bireylere ait boylamsal üç gözlem değerini içeren farklı örneklem büyüklüğü (50; 500; 1000) ve (hata dağılımlarının çarpıklıklarına göre) farklı dağılımlara sahip veri yapılarında çok düzeyli regresyon (rasgele kesişim ve rasgele eğim modeli) ve çok düzeyli kantil regresyon analiz yöntemlerinin performanslarının karşılaştırılmıştır. Bu amaçla, boylamsal veri setlerinde çok düzeyli regresyon ve çok düzeyli kantil regresyon yönteminde farklı kantil düzeylerinde (0.10; 0.25; 0.50; 0.75; 0.90) farklı örneklem büyüklüklerinde (50; 500; 1000) model veri uyumu değerleri ile mutlak hata ve yanlılık değerleri incelenmiştir. Analizlerde R programının alt paketlerinden olan qrLMM ve nlme ile çok düzeyli kantil regresyon analizi yapılmıştır.

Anahtar sözcükler: kantil regresyon, regresyon, boylamsal veri, hiyerarşik veri, çarpıklık

Abstract

Classical simple linear regression analysis assumes that the observations are independent and identical. This assumption is not met in situations such as hierarchical data. The fact that the principle of independence of observations in hierarchical data is violated causes biased results in regression analysis. Multilevel modeling techniques are used to solve such problems. However, in cases where the distribution is skewed, quantile regression models, which are used as an alternative method, have been developed due to the low efficiency of the methods used in multilevel modeling. In cases where classical regression conditions are not met, information about each point of the distribution can be obtained with quantile regression models, which are advantageous. In this study, the performances of multilevel regression (random intercept and random slope model) and multilevel quantile regression analysis methods were compared in data structures with three longitudinal observation values belonging to individuals with different sample sizes and different distributions. For this purpose, model-data fit values, absolute error and bias values, and changes in intercept and slope coefficients at different quantile levels (0.10; 0.25; 0.50; 0.75; 0.90) and different sample sizes (50; 500; 1000) in multilevel regression and multilevel quantile regression methods in longitudinal data sets were analysed. The variation of absolute error and bias values was examined. In the analysis, multilevel quantile regression analysis were performed with qrLMM and nlme, which are sub-packages of the R program. As a result of the analyses, it was seen that multilevel quantile regression made more unbiased predictions than multilevel regression analysis in all data structures. In addition, although there was no significant difference between the intercept value obtained from multilevel regression and the intercept value obtained from the median in any case, it was observed that a significant difference occurred between the slope coefficients in many cases.

Keywords: quantile regression, regression, longitudinal data, hierarchical data, skewness

Teşekkür

Doktora tez çalışmam boyunca hızlı dönütlerle beni yönlendiren, emeğini, sabrını, vaktini ve desteğini esirmeyen, kaygılandığımda ve her umutsuzluğa kapıldığımda beni motive eden danışmanım Doç. Dr. Derya ÇOBANOĞLU AKTAN'a,

Tezimin ilerleme sürecinde görüşleriyle yönlendiren Prof. Dr. Burcu ATAR ve Prof. Dr. Hakan Yavuz ATAR'a,

Doktora eğitimim boyunca ders aldığım, yoluma ışık tutan, bilgi ve tecrübelerinden faydalandığım başta Prof. Dr. Selahattin GELBAL, Prof. Dr. Hülya KELEÇİOĞLU olmak üzere, mesleki hayatımda da desteklerini esirgemeyen sayın Hocam Prof. Dr. Cem Oktay Güzeller ve Eğitimde Ölçme ve Değerlendirme Bilim Dalı'ndaki tüm hocalarıma ayrı ayrı teşekkür ederim.

Doktora eğitimimin başından beri desteğini hiç esirgemeyen dostlarım Ar. Gör. Melike ÖZER TAYMUR'a, vaktini ve emeğini esirgemediği tez çalışmamda teknik destek sağlayan eşi Erkan TAYMUR, Muammer YÜKSEL ile süreçte katkı sunan Ar. Gör. Ömer KAMIŞ'a,

Lisansüstü eğitimim boyunca maddi-manevi destek sunan annem, babam ve kardeşim Fatih ÖZTEMÜR'e,

Son olarak sonsuz sabrı ve anlayışı ile bu stresli süreci atlatmamda yardımcı olan sevgili eşim Gökay KAFKAS ve ailemize yeni katılan sevgili miniğime sonsuz teşekkür ederim.

İçindekiler

Kabul ve Onay.....	ii
Öz.....	iii
Abstract.....	iv
Teşekkür.....	v
İçindekiler.....	vi
Tablolar Dizini.....	ix
Şekiller Dizini.....	xii
Simgeler ve Kısaltmalar Dizini.....	xiv
Bölüm 1 Giriş.....	1
Problem Durumu.....	2
Araştırmanın Amacı ve Önemi.....	10
Sınırlılıklar.....	13
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	14
Regresyon Yöntemi.....	14
Kantil Fonksiyonları.....	17
Kantil Regresyon Modeli.....	26
Çok Düzeyli Boylamsal Verilerde Analiz Yöntemleri.....	33
Verilerin Merkezileştirilmesi.....	59
Maksimum Olasılık (Maximum Likelihood-ML) Kestirim Yöntemi.....	60
Kantil Regresyon ve Çok Düzeyli Boylamsal Veri Yapısı ile İlişkisi.....	61
Çok Düzeyli Kantil Regresyon Modeli.....	63
Çok Düzeyli Modeller İle Çok Düzeyli Kantil Regresyonda Model Seçimi	
Kriterleri.....	67
İlgili Araştırmalar.....	68

Bölüm 3 Yöntem.....	85
Araştırmanın Türü	85
Veri Toplama Araçları	85
Çalışmanın Verilerinin Üretilmesi ve Simülasyon Koşulları	95
Model	98
Parametrelerin Belirlenmesi	99
Verilerin Analizi	102
Bölüm 4 Bulgular, Yorumlar ve Tartışma.....	105
Birinci Alt Probleme İlişkin Bulgular ve Yorum	105
İkinci Alt Probleme İlişkin Bulgular ve Yorum	115
Üçüncü Alt Probleme İlişkin Bulgular ve Yorum	127
Tartışma.....	145
Bölüm 5 Sonuç ve Öneriler.....	148
Birinci Alt Probleme İlişkin Sonuçlar.....	148
İkinci Alt Probleme İlişkin Sonuçlar	149
Üçüncü Alt Probleme İlişkin Sonuçlar	150
Öneriler	152
Kaynaklar	154
EK-A: Parametrelere Ait Ortalama Kesişim ve Eğim Değerleri.....	clxxviii
EK B: QRLMM Analiz Sonuçlarına Ait Örnek İterasyon Grafikleri (q10).....	clxxix
EK-C: QRLMM Analiz Sonuçlarına Ait Örnek İterasyon Grafikleri (q25)	clxxx
EK-D: QRLMM Analiz Sonuçlarına Ait Örnek İterasyon Grafikleri (q50)	clxxxı
EK-E: QRLMM Analiz Sonuçlarına Ait Örnek İterasyon Grafikleri (q75).....	clxxxii
EK-F: QRLMM Analiz Sonuçlarına Ait Örnek İterasyon Grafikleri (q90).....	clxxxiii
EK-G: Araştırma Etik Komisyon İzin Muafiyeti Formu/ Araştırma Etik Komisyonu Onay Bildirimi	clxxxiv
EK-H: Etik Beyanı.....	clxxxv

EK-I: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu.....	clxxxvi
EK-J: Thesis/Dissertation Originality Report.....	clxxxvii
EK-K: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	clxxxviii

Tablolar Dizini

Tablo 1 Çok Düzeyli Modele Ait Parametrelerin Tanımı ve Yorumlanması	59
Tablo 2 Cinsiyet Değişkenine Göre Betimsel İstatistikler	86
Tablo 3 PIAT Verisine Ait Betimsel İstatistikler	87
Tablo 4 Bireylerin Ölçüm Gruplarına Ait Betimsel İstatistikler	87
Tablo 5 Bireylerin Ölçüm Gruplarına Ait Puanların Korelasyonları	87
Tablo 6 Boş Modele Ait Analiz Sonuçları	90
Tablo 7 Rasgele Kesişim Sabit Eğim Modeli ile Rasgele Kesişim Rasgele Eğim Modellerine Göre Analiz Sonuçları	91
Tablo 8 PIAT Verisine Ait Farklı Modellere Ait Model Veri Uyumu Değerleri	91
Tablo 9 PIAT Verisinin Çok Düzeyli Kantil Regresyon Analizine Göre Farklı Kantil Düzeylerindeki Model Veri Uyumu Değerleri	92
Tablo 10 PIAT Verisinin Çok Düzeyli Kantil Regresyon Analizine Göre Farklı Kantil Düzeylerindeki Parametre Kestirim Değerleri	92
Tablo 11 Farklı Örneklem Büyüklüklerine Göre Elde Edilen Parametre ve Standart Hata Miktarları	95
Tablo 12 Çalışmanın Koşulları	97
Tablo 13 Simülasyon Koşulları	97
Tablo 14 Düzey 1 ve Düzey 2 Hata Dağılımların Farklılıklarına Göre Farklı Örneklem Büyüklüklerindeki AIC Değerleri	106
Tablo 15 Düzey 1 ve Düzey 2 Hata Dağılımların Farklılıklarına Göre Farklı Örneklem Büyüklüklerindeki BIC Değerleri	107
Tablo 16 Düzey 1 ve Düzey 2 Hata Dağılımların Farklılıklarına Göre Farklı Örneklem Büyüklüklerindeki logLike Değerleri	109
Tablo 17 Düzey 1 ve Düzey 2 Hataların Farklı Dağılımlarına Göre Mutlak Hata ve Yanlılık Değerleri	115
Tablo 18 Düzey 1 ve Düzey 2 Hata Terimlerinin Normal Dağılım Gösterdiği ($\varepsilon \sim [N, N]$)	

<i>Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değerleri</i>	128
Tablo 19 <i>Düzye 1 ve Düzye 2 Hata Terimlerinin Normal Dağılım Gösterdiği ($\epsilon \sim [N, N]$)</i>	
<i>Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri.....</i>	129
Tablo 20 <i>Düzye 1 Hata Terimlerinin Normal ve Düzye 2 Hata Terimlerinin X^2 Dağılımı Gösterdiği ($\epsilon \sim [N, X^2]$) Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değerleri.....</i>	130
Tablo 21 <i>Düzye 1 Hata Terimlerinin Normal ve Düzye 2 Hata Terimlerinin X^2 Dağılımı Gösterdiği ($\epsilon \sim [N, X^2]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri</i>	131
Tablo 22 <i>Düzye 1 Hata Terimlerinin Normal ve Düzye 2 Hata Terimlerinin Student-t Dağılım Gösterdiği ($\epsilon \sim [N, T]$) Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değerleri.....</i>	132
Tablo 23 <i>Düzye 1 Hata Terimlerinin Normal ve Düzye 2 Hata Terimlerinin Student-t Dağılım Gösterdiği ($\epsilon \sim [N, T]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri.....</i>	133
Tablo 24 <i>Düzye 1 Hata Terimlerinin X^2 ve Düzye 2 Hata Terimlerinin Normal Dağılım Gösterdiği ($\epsilon \sim [X^2, N]$) Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değerleri.....</i>	134
Tablo 25 <i>Düzye 1 Hata Terimlerinin X^2 ve Düzye 2 Hata Terimlerinin Normal Dağılım Gösterdiği ($\epsilon \sim [X^2, N]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri</i>	135
Tablo 26 <i>Düzye 1 ve Düzye 2 Hata Terimlerinin X^2 Dağılım Gösterdiği ($\epsilon \sim [X^2, X^2]$) Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değerleri.....</i>	136
Tablo 27 <i>Düzye 1 ve Düzye 2 Hata Terimlerinin X^2 Dağılım Gösterdiği ($\epsilon \sim [X^2, X^2]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri.....</i>	137

Tablo 28 <i>Düzey 1 Hata Terimlerinin X_2 ve Düzey 2 Hataların Student-t Dağılımı Gösterdiği ($\varepsilon \sim [X_2, T]$) Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değerleri.....</i>	138
Tablo 29 <i>Düzey 1 Hata Terimlerinin X_2 ve Düzey 2 Hataların Student-t Dağılımı Gösterdiği ($\varepsilon \sim [X_2, T]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri</i>	139
Tablo 30 <i>Düzey 1 Hata Terimlerinin Student-t ve Düzey 2 Hataların Normal Dağılımı Gösterdiği ($\varepsilon \sim [T, N]$) Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değerleri.....</i>	140
Tablo 31 <i>Düzey 1 Hata Terimlerinin Student-t ve Düzey 2 Hataların Normal Dağılımı Gösterdiği ($\varepsilon \sim [T, N]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri</i>	141
Tablo 32 <i>Düzey 1 Hata Terimlerinin Student-t ve Düzey 2 Hataların X_2 Dağılımı Gösterdiği ($\varepsilon \sim [T, X_2]$) Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değerleri.....</i>	142
Tablo 33 <i>Düzey 1 Hata Terimlerinin Student-t ve Düzey 2 Hataların X_2 Dağılımı Gösterdiği ($\varepsilon \sim [T, X_2]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri</i>	143
Tablo 34 <i>Düzey 1 ve Düzey 2 Hata Terimlerinin Student-t Dağılımı Gösterdiği ($\varepsilon \sim [T, T]$) Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değerleri</i>	144
Tablo 35 <i>Düzey 1 ve Düzey 2 Hata Terimlerinin Student-t Dağılımı Gösterdiği ($\varepsilon \sim [T, T]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri.....</i>	145

Şekiller Dizini

Şekil 1 <i>Kümülatif Dağılım Fonksiyonu (Gilchrist, 2000).</i>	18
Şekil 2 <i>Olasılık Yoğunluk Fonksiyonu (Gilchrist, 2000).</i>	19
Şekil 3 <i>Kantil Fonksiyonu (Gilchrist, 2000).</i>	20
Şekil 4 <i>Kümülatif Dağılım Fonksiyonu ile Kantil Fonksiyonun Karşılaştırılması (Hao & Naiman, 2007).</i>	21
Şekil 5 <i>Normal Kantil Fonksiyonu ile Sağa Çarpık Kantil Fonksiyonun Karşılaştırılması (Hao & Naiman, 2007).</i>	23
Şekil 6 <i>Minimizasyon Sorunlarını Tanımlamak İçin Kullanılan Farklı Fonksiyonlar (Hao & Naiman, 2007).</i>	25
Şekil 7 <i>Kontrol Fonksiyonun Farklı Kantil Değerlerindeki Görüntüleri (Park ve diğerleri, 2012).</i>	25
Şekil 8 <i>Tek Bir Doğrusal Regresyon Doğrusu ve 19 Ayrı Kantil Regresyon Doğrusu (Hao & Naiman, 2007).</i>	28
Şekil 9 <i>Tekrarlı Ölçümlerin Bireylerin İçine Yuvalanma Diyagramı</i>	35
Şekil 10 <i>Veri Kutusu</i>	36
Şekil 11 <i>Bir Bireye Ait Dört Adet Ölçüm Değeri</i>	43
Şekil 12 <i>Tekrarlı Ölçümlerden Elde Edilen Bireysel Eğriler (Yörüngeler)</i>	44
Şekil 13 <i>Bir örneklemden tüm bireylere ait büyüme eğrileri</i>	45
Şekil 14 <i>Zamanın Sabit ve Rasgele Etkilerinin Gösterimi (Hoffman, 2015)</i>	46
Şekil 15 <i>Farklı Regresyon Denklemlerinin Kesişim ve Eğim Değerlerinin İlişkisinin Gösterimi (Jones & Bullen, 1994).</i>	48
Şekil 16 <i>%5 Random Bireylere Ait Büyüme Noktaları</i>	88
Şekil 17 <i>%5 Random Bireylere Ait Büyüme Grafikleri</i>	89
Şekil 18 <i>%5 Random Bireylere Ait Büyüme Çizgileri</i>	89
Şekil 19 <i>Düzey 1 Hata Dağılımlarına Ait Grafikler</i>	100

Şekil 20 <i>Düzye 2 Hata Dağılımlarına Ait Grafikler</i>	100
Şekil 21 <i>X² (Ki kare) Hata Dağılımlarına Ait Grafikler.....</i>	101
Şekil 22 <i>Student T Hata Dağılımlarına Ait Grafikler.....</i>	102
Şekil 23 <i>N=50 Örnekleme Büyüklüğü İçin Model Veri Uyumu Değerleri</i>	110
Şekil 24 <i>N=500 Örnekleme Büyüklüğü İçin Model Veri Uyumu Değerleri</i>	112
Şekil 25 <i>N=1000 Örnekleme Büyüklüğü İçin Model Veri Uyumu Değerleri</i>	113
Şekil 26 <i>Düzye 1 ve Düzye 2 Hata Terimlerinin Normal Dağılım Gösterdiği ($\epsilon \sim [N, N]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri</i>	117
Şekil 27 <i>Düzye 1 Hata Terimlerinin Normal Ve Düzye 2 Hata Terimlerinin X² Dağılım Gösterdiği ($\epsilon \sim [N, kk]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri.....</i>	118
Şekil 28 <i>Düzye 1 Hata Terimlerinin Normal ve Düzye 2 Hata Terimlerinin Student's T₃ Dağılım Gösterdiği ($\epsilon \sim [N, T]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri.....</i>	120
Şekil 29 <i>Düzye 1 Hata Terimlerinin X² ve Düzye 2 Hata Terimlerinin Normal Dağılım Gösterdiği ($\epsilon \sim [kk, N]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri.....</i>	121
Şekil 30 <i>Düzye 1 ve Düzye 2 Hata Terimlerinin X² Dağılımı Gösterdiği ($\epsilon \sim [kk, kk]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri</i>	122
Şekil 31 <i>Düzye 1 Hata Terimlerinin X² ve Düzye 2 Hata Terimlerinin Student's T₃ dağılımı Gösterdiği ($\epsilon \sim [kk, T]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri</i>	123
Şekil 32 <i>Düzye 1 Hata Terimlerinin Student's T₃ ve Düzye 2 Hata Terimlerinin Normal Dağılım Gösterdiği ($\epsilon \sim [T, N]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri.....</i>	124
Şekil 33 <i>Düzye 1 Hata Terimlerinin Student's T₃ ve Düzye 2 Hata Terimlerinin X² Dağılım Gösterdiği ($\epsilon \sim [T, kk]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri.....</i>	125
Şekil 34 <i>Düzye 1 ve Düzye 2 Hata Terimlerinin Student's T₃ Dağılım Gösterdiği ($\epsilon \sim [T, T]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri</i>	126

Simgeler ve Kısaltmalar Dizini

ALD: Asimetrik Laplace Dağılımı

EKK: En Küçük Kareler Yöntemi

LSAY: Longitudinal Surveys of American Youth (Amerikan Gençliği Boylamsal Araştırmaları)

MEB: Milli Eğitim Bakanlığı

NLS: National Longitudinal Surveys of Labor Market Experience (Ulusal Boylamsal İşgücü Piyasası Deneyimi Araştırmaları)

NLSY-CYA: National Longitudinal Survey of Youth Child and Young Adult (Ulusal Boylamsal Gençlik Çocuk ve Genç Yetişkin Araştırması)

OLS: Ordinary Least Squares (Sıradan En Küçük Kareler Yöntemi)

PIAT: The Peabody Individual Achievement Test (Peabody Bireysel Başarı Testi)

PISA: The Programme for International Student Assessment (Uluslararası Öğrenci Değerlendirme Programı)

PSID: University of Michigan's Panel Study of Income Dynamics (Michigan Üniversitesi Gelir Dinamikleri Paneli Çalışması)

SLID: Canadian Survey of Labour Income Dynamics (Kanada Gelir Dinamikleri Araştırması)

TÜİK: Türkiye İstatistik Kurumu

Bölüm 1

Giriş

Zaman içindeki değişiklikleri düzenli aralıklarla izlemek, birçok araştırma alanındaki karmaşık konuları incelemek ve bu konular hakkında veri toplamak için yararlı bir yoldur (Caruana ve diğerleri, 2015). Bu değişiklikleri izlemek amacıyla yürütülen çalışmalarda, eğitimde ve istatistikte kullanılan verilerin özelliklerine göre sınıflandırıldığında karşımıza zaman serisi, yatay kesit ve boylamsal (dikey kesit) veriler çıkmaktadır. Zaman serisi verileri; değişkenlerin değerlerinin zaman birimlerine göre değişimini içeren veri yapılarıdır. Ülkelerin ortalama başarı puanlarının değerlendirildiği bir çalışmada, Türkiye'nin bugüne dek katılmış olduğu her bir döngüdeki (örneğin 2015, 2018, 2021) PISA ortalama matematik başarı puanlarının sonuçları bu verilere örnek olarak gösterilebilir. Yatay kesit veriler ise; zamanın belirli bir noktasında yani zaman sabitken, farklı birimlerden toplanan veriler olarak adlandırılmaktadır. Bu veri yapılarında "birim"; birey, öğrenci, sınıf, okul, bölge gibi birimleri ifade etmek için kullanılmaktadır. Türkiye, Almanya, Singapur gibi bazı ülkelerin belirli bir yıldaki (örneğin 2018) PISA ortalama matematik başarı puanları verileri yatay kesit veri setine örnek olarak gösterilebilir.

Boylamsal veriler ise, bireyler, sınıflar, okullar gibi birimlere ait yatay kesit gözlemlerinin, belli bir dönemde bir araya getirilmesiyle elde edilen veri yapılarıdır. Boylamsal veriler kendi içlerinde; trend/eğilim, panel ve kohort veriler olarak da üç ayrı şekilde gruplandırılabilir. Trend/eğilim çalışmalarında, farklı zaman dilimlerinde toplanan veriler söz konusudur. Verinin elde edildiği örneklem grubu her veri toplama esnasında yeniden oluşturulur; böylece zaman içindeki trend durumlar belirlenmeye çalışılır. Panel çalışmalarda veriler, önceden tanımlanmış bir grup içindeki belirli bireyler kullanılarak farklı zaman dilimlerinde toplanır. Son olarak kohort çalışmalarında ise zaman içinde değişmeyen bir evren grubundan, farklı zaman dilimlerinde farklı örneklemeler kullanılarak verilerin toplanması söz konusudur (Caruana ve diğerleri, 2015). Bu bağlamda PISA çalışmaları farklı zaman dilimlerinde ve farklı örneklem gruplarında yürütüldüğü için,

Türkiye, Almanya, Singapur gibi bazı ülkelerin bugüne dek katılmış oldukları her bir döngüdeki (örneğin; 2012, 2015 ve 2018) PISA ortalama matematik başarı puanları sonuçlarının bir araya getirilerek oluşturulan veri yapısı, trend boylamsal veri setine örnek olarak gösterilebilir. Boylamsal araştırmalarda, belirli bireyleri uzun süreler boyunca - genellikle yıllar veya on yıllar boyunca - takip etmek için sürekli veya tekrarlanan ölçümler kullanılır (Caruana ve diğerleri, 2015). Aynı bireylere ait gözlemlerin zaman içinde iki ya da daha fazla incelenmesiyle elde edilen tekrarlı ölçümler boylamsal veri olarak adlandırılmaktadır (Fitzmaurice ve diğerleri, 2004). Analiz edilen gözlemler aynıdır ve en azından gözlemlenebilen herhangi bir dönemdeki durumuyla karşılaştırılabilir niteliktedir. Veriler önceden tanımlanmış bir grup içindeki belirli bireyler için toplandığından, bir bütün olarak grup veya belirli bireyler için zaman içindeki değişimi analiz etmek için uygun istatistiksel testler kullanılabilir (Van Belle ve diğerleri, 2004). Analiz, verinin gözlemlenen dönemler içinde ya da arasında karşılaştırılmasını içermelidir. Boylamsal araştırmalarda yapılan düzenlemelerle araştırılan değişkende dönemler arasındaki farkların veya değişimlerin incelenmesi mümkün olmaktadır (Menard, 1991). Yani boylamsal bir çalışmanın amacı çoğu zaman, belirli bir süre içindeki değişiklikleri izlemek, ölçmek ve değişimi etkileyen faktörleri karakterize etmenin yanı sıra (Singer & Willett, 2003; Geraci & Bottai, 2007; Ployhart & Vandenberg, 2010; Caruana ve diğerleri, 2015) büyümenin işlevsel biçimini tanımlamaktır (Nese ve diğerleri, 2002).

Problem Durumu

Eğitim bilimleri alanında öğrenci başarısındaki değişikliklerin izlenmesi ve bu başarıya etki eden nedenlerin modellenmesi oldukça önemlidir. Teorik olarak, tüm disiplinlerdeki araştırmacılar, zaman aralıklarındaki değişikliklere göre değişkenler arasındaki doğrusal ilişkiler açısından düşünme eğiliminde olmuştur. Başka bir deyişle araştırmacılar, X'teki bir artışın Y'de bir artışa (veya azalmaya) yol açacağını ve Y'nin daha sonra bu seviyede sabit kalacağını varsaymaktadır (Kelly & McGrath, 1988). Aslında, X'teki bir artışın Y üzerindeki etkisinin devam edip etmediği, kalıcılığın süresi ve

ayrıca Y'nin zaman içinde işlevsel bir form kaymasına sahip olup olmadığı da araştırmacılar için önemli bir ilgi alanıdır (Caruana ve diğerleri, 2015). Bireylerde meydana gelen zaman içindeki değişiklikler doğrusal, doğrusal olmayan veya süreksiz olabilirler (Ployhart & Vandenberg, 2010). Bu değişiklikler, farklı hızlarda veya tempolarda meydana gelebilir; bazı değişiklikler artımlı bazıları ise katlanarak meydana gelebilir, bazı yapılar ritmik veya döngüsel kalıplarda değişebilir (Mitchell & James, 2001). Boylamsal çalışmaların eğitim bilimlerindeki yeri incelendiğinde, yapılan çalışmalarda akademik performansın zaman içindeki yörüngesi (trajectory) ve bu yörüngelerin hangi faktörlerle ilişkili olduğu (Yapar, 2014); farklı bireysel özelliklerin (Hamed, 2012), farklı öğretim uygulamaları, müfredat veya okul ikliminin (Hanushek ve diğerleri, 2011; Sakız ve diğerleri, 2021), eğitim politikalarındaki değişikliklerin/tutarlılığın (Newmann ve diğerleri, 2001) zaman içinde nasıl değişikliklere yol açtığı gibi sorular araştırmaların odağında yer almaktadır. Tüm bu olası araştırma soruları göz önünde bulundurulduğunda, boylamsal çalışmaların dinamik bir süreç olduğu da görülmektedir. Buna karşılık kesitsel analizler, belirli bir grubun belirli bir anda nasıl görüldüğüne dair bir "anlık görüntü" alması olarak da tanımlanabilen; belirli bir durumda birden fazla değişkeni analiz edebilen ancak zamanın ölçülen değişkenler üzerindeki etkisine ilişkin hiçbir bilgi sağlamayan araştırma türleridir (Caruana ve diğerleri, 2015) ve doğası gereği durağandır (Mitchell & James, 2001; Ployhart & Vandenberg, 2010; Caruana ve diğerleri, 2015). Oberfield (2014) zamansal süreçlerle ilgili araştırmalar için kesitsel verilere güvenmenin, "belirli bir andaki değişimi açıklayan faktörler, zaman içindeki değişimi açıklayan faktörlerle aynıdır" şeklindeki sorunlu bir varsayım üzerine inşa edildiğine işaret eder. Araştırmacılar ayrıca, kesitsel verilerde olumlu görünen ilişkilerin, zaman dilimleri boyunca bireysel değişiklikleri düşündüğümüzde aslında olumsuz olduğunu bulmuşlardır (Vancouver ve diğerleri, 2001). Nedensellik açısından, kesitsel verilerle çalışan araştırmacılar, zaman içinde değişkenler arasındaki değişiklikleri gözlemleyemezler ve nedensel ilişkileri doğrudan test edemezler (Popper, 1959). Bağımlı ve bağımsız değişkenlerin zamana bağlı sıralanması ve bu değişkenlerin birimler içinde tekrarlanan gözlemleri sayesinde, araştırmacılar nedensel

yönleri inceleme ve zaman içinde meydana gelen değişkenlerdeki değişiklikleri gözlemlene fırsatına sahip olurlar (Lynn, 2009; Kling ve diğerleri, 2015). Eğitim araştırmalarındaki önemli sorular, geçmişin ve ilk deneyimlerin sonuçları ne ölçüde etkilediği ve gençlerin eğitim deneyimlerinin zaman içinde nasıl değiştiği ile ilgilidir. Bu alanları incelemek için, farklı yaşlar için zengin veriler sağlayan veri setlerine ihtiyaç vardır (Carroll, 2019). Boylamsal verilerin kullanımıyla ilgili tartışmaların çoğu, tahmin ve nedensellik durumlarının tanımlanması açısından kesitsel verilere göre sahip olduğu avantajlara odaklanmaktadır. Hsiao (2007), boylamsal veri yapılarının bir dizi avantajını özetlemektedir. Boylamsal veri yapılarıyla sürdürülen analizler, değişkenlerin ölçümleri birden çok zaman periyodu boyunca tekrarlandığında mevcut olan veri miktarı hem serbestlik derecesini hem de varyasyonu artırarak tahminin etkinliğini artırır; araştırmacının birime özgü etkileri (örn. rasgele veya sabit etkiler) kontrol ederek tahminler üzerindeki gözlemlenemeyen heterojenliğin etkilerini sınırlamasına izin verir. Bu, kesitsel analizlerde ortaya çıkan değişken yanlılığı atlayan tehditler göz önüne alındığında özellikle önemlidir. Son olarak, yapı, dinamik ilişkilerin modellenmesine izin verir. Bağımsız değişkenlerin gecikmeli etkileri ve karşılıklı nedensellik dahil olmak üzere karmaşık ilişkiler, eğer veriler yeterli zaman boyunca her bir gözlem için yeterli sayıda tekrarlanan ölçüm içeriyorsa modellenebilir (Anderson & Hsiao, 1982). Boylamsal veri yapılarıyla sürdürülen analizler, değişkenlerin ölçümleri birden çok zaman periyodu boyunca tekrarlandığında mevcut olan veri miktarı hem serbestlik derecesini hem de varyasyonu artırarak tahminin etkinliğini artırır; araştırmacının birime özgü etkileri (örn. rasgele veya sabit etkiler) kontrol ederek tahminler üzerindeki gözlemlenemeyen heterojenliğin etkilerini sınırlamasına izin verir. Bu, kesitsel analizlerde ortaya çıkan bağımlı değişkeni önemli ölçüde etkileyen değişkenlerin modele dahil edilmemesi olarak da bilinen, dışlanan değişken yanlılığı gibi tehditler göz önüne alındığında özellikle önemli olmaktadır. Son olarak, yapı, dinamik ilişkilerin modellenmesine izin verir. Bağımsız değişkenlerin gecikmeli etkileri ve karşılıklı nedensellik dahil olmak üzere karmaşık ilişkiler, eğer veriler yeterli zaman boyunca her bir gözlem için yeterli sayıda tekrarlanan

ölçüm içeriyorsa modellenabilir (Anderson & Hsiao, 1982). Boylamsal çalışmalar bireyler ve/veya grupların belirli süreler boyunca gözlemlenmesi sebebiyle, zaman içindeki değişiklikleri izlemeyi mümkün kılarak (Singer & Willett, 2003) bireylerin farklı aralıklarla birden fazla tekrarlı ölçümlerle incelenmesi gelişim düzeyleri ve bu süreçler hakkında araştırmacılara bilgi vermesinin yanı sıra, bireylerin gelişimi etkileyen risk ve koruyucu faktörlerin incelenmesine de olanak tanımaktadır (Wagner ve diğerleri, 2005). Rutter (1994), risk faktörlerinin, katılımcıların bunları geriye dönük olarak hatırlamasındansa, ortaya çıktıkça ölçülen bir davranış modelini inceleyerek daha iyi kavramsallaştırılabileceğine dikkat çeker. Genel olarak boylamsal çalışmalar, eğitim sonuçlarının ve bireylere ait farklı özelliklerin risk faktörü oluşturup oluşturmadığını anlamada ve bu risk faktörlerinin belirli zaman içerisinde nasıl bir seyir izlediğini anlamak, eğitimde elde edilen çıktılarının iyileştirilmesinin, eğitim politikalarının şekillendirilmesinin yollarının belirlenmesi açısından önemlidir. Eğitim bilimleri alanında yapılması planlanan boylamsal çalışmalar, araştırmancının ilgi alanlarına bağlı olarak değişmekle birlikte birçok araştırma sorusunun yanıtı olan sonuçların elde edilmesini sağlamaktadır. Örneğin, bu çalışma türü risk faktörlerinin (bireysel, toplumsal) belirlenmesi ve bu risk faktörleri ile bireyin bilişsel gelişimi arasındaki ilişkiyi belirlemek, bireylerin doğal öğrenme süreçlerinin seyrinin takibi yardımıyla farklı sürelerdeki öğretim yöntemlerinin sonuçlarını değerlendirmek için özellikle yararlıdır (Caruana ve diğerleri, 2015). Bireylerin tekrarlanan gözlemleri ile bireysel değişkenlikler daha iyi yorumlanmakta, gözlenen değişken ile nedensellik oluşturmada yardımcı olmakta, araştırmacıların bir maruz kalma veya risk faktörü ile gözlenen durum arasındaki geçici ilişkiyi incelemesine tanıyarak, risk faktörünün gerçekten belirli bir sonuca neden olup olmadığını belirlemeyi mümkün ve böylece belirli sonuçlar için risk altında olma olasılığına sahip farklı grupların belirlenmesi ve varsa özel ihtiyaçlarının karşılanması için özel müdahalelerin gerçekleştirilmesine olanak sağlamaktadır.

Tüm bu amaçlarla yapılmış olan çalışmalar incelendiğinde son zamanlarda, her biri farklı zamanlarda gözlemlenmiş, birçok ülke ve binlerce bireye ait veriler mevcuttur. Türkiye dışındaki uzun yıllardır sürdürülen boylamsal çalışmalara bakıldığında 1960'lı yıllardan itibaren iş gücü piyasasındaki bireylere ait verileri toplayan NLS (National Longitudinal Surveys of Labor Market Experience), 1968 yılında başlayan ve 5000 değişken ile 15000'de fazla birey için toplanan PSID (University of Michigan's Panel Study of Income Dynamics), 1993 yılında başlatılan SLID (Canadian Survey of Labour Income Dynamics), 1985 yılından beri yürütülen LSAY (Longitudinal Surveys of American Youth) gibi çalışmalar bireylere ait değişkenlerin yanı sıra eğitimsel verilerini de toplamaktadır. Böylece bireylerin gelişimleri bir bütün olarak görülebilmekte ve bağımlı değişkeni açıklayıcı değişkenler hakkında detaylı bilgiler elde etmek mümkün olmaktadır. Ülkemizde ise bu şekilde hazırlanmış ve süregelen detaylı bir çalışma bulunmamakla birlikte TÜİK tarafından şehir ve bölge bazlı oluşturulan veriler araştırmalarda sıklıkla kullanılmaktadır. Türkiye'de eğitim alanında bireylerin uzun süre izlenmesine yönelik çalışmalar henüz yaygın olarak yapılmamaktadır. MEB 20. Şura Tavsiye Kararları ve 2023 Vizyon Belgesi'nde de yer alan planlar arasında erken çocukluk dönemindeki risk gruplarını belirlemek için tarama ve periyodik değerlendirmeler yapılması ve sonuçlar çerçevesinde erken müdahale programlarının hazırlanması; öğrencilerin tanımlanan yeterliklere ne derece ulaştıklarını anlamak için ulusal izleme ve değerlendirme çalışmaları geliştirilerek sürdürülmesi; tüm öğrencilerin akademik performanslarının periyodik olarak takip edilmesi gerekliliği yer almaktadır (Milli Eğitim Bakanlığı, 2021). Bu politika ve kararlar çerçevesinde MEB'in Öğrenci İzleme Çalışmaları gibi çalışmaları ile öğrenci gelişimi ve öğrenmelerinin süreç içerisindeki takibinin başlatması son yıllarda sürdürülen araştırmalar arasında yer almaktadır.

Boylamsal verilere amaçları ve avantajları dışında farklı bir açıdan bakıldığında, istatistiksel olarak, yürütülen araştırmalarda gözlenen değerlere ait dağılımın bütünlüğünün korunmasının yanı sıra, dağılımın farklı özellikler gösteren her bir alt grubun ya da

dağılıma ait farklı yüzdeler kesitlerinin de kendi içinde incelenerek doğru yorumların yapılması gerekmektedir. Örneğin Türkiye’de eğitimde fırsat ve cinsiyet eşitsizliği yaşayan ya da başarısız olan grupların varlığı çeşitli raporlamalarda dile getirilmekte olup (Feriver & Arık, 2021; TEDMEM, 2022) bu raporlar değerlendirildiğinde, çalışmalarda elde edilen veri setlerinin kendi içince incelenerek politika yapıcıların bu durumlara uygun stratejiler geliştirmesinin önemi daha net anlaşılmaktadır.

Boylamsal çalışmalar, gerçekleştirilmesi zorlu ve yürütmesi maliyetli olsa da, değerli veriler sağlarlar. Ülkemizde de öğrencilerden elde edilecek boylamsal verilerin toplanmasının zor ve iyi planlanmış, uzun süreli çalışmalar gerektirmesinden dolayı eğitim bilimleri alanında bu konuda büyük bir eksiklik olduğu söylenebilir. Böylesi değerli bilgi kaynağı sunma özelliğinde olan bu veri kümelerinin eğitim araştırmalarında nasıl kullanıldığının anlaşılması, verilerin verimli ve anlamlı bir şekilde kullanılmasını destekleyebilir ve gelecekteki araştırmalara yol gösterici nitelikte olabilir (Carroll, 2019). Boylamsal veriler kullanılarak yürütülen analizlerden bazıları; tanımlayıcı istatistikler, Büyüme Eğrisi Analizleri (Growth Curve Analysis), Tekrarlanan Ölçümler Varyans Analizi (Repeated Measures Analysis of Variance (ANOVA), Zaman Serisi Analizi (Time Series Analysis), Hayatta Kalma Analizi (Survival Analysis), Gizil Büyüme Eğrisi Modellemesi (Latent Growth Curve Modeling), Yapısal Eşitlik Modellemesi (Structural Equation Modeling) olarak gruplandırılabilir. Tanımlayıcı istatistikler çalışmadaki değişkenlerin özelliklerini özetlemek ve tanımlamak amacıyla yapılırken; zaman içindeki bireysel/grup düzeyindeki değişim yörüngelerini incelemek için Büyüme Eğrisi Analizleri; belirli bir odak noktasının etkilerin incelemesi ve zaman noktalarında ya da gruplar arasında önemli farklılıkların belirlenmesi söz konusu iken Tekrarlanan Ölçümler Varyans Analizi kullanılmaktadır. Zaman içinde değişkenlerdeki değişiklikleri tespit etmek ve tahmin etmek/kalıpları ve eğilimleri belirlemek için düzenli aralıklarla toplanan verileri analiz etmeye odaklanan Zaman Serisi Analizi (Lehr & Newton, 1978; Marston, 1988; Li ve diğerleri, 2019); bir olayın meydana gelmesine kadar geçen süreyi incelerken kullanılan

Hayatta Kalma Analizi (Plank ve diğeri, 2008; No ve diğeri, 2016) yapılan boylamsal çalışmalarda sıklıkla kullanılan tekniklerdendir. Bunların yanı sıra Gizil Büyüme Eğrisi Modellemesinin (Palardy, 2008; Phan, 2012; Joensuu ve diğeri, 2013); bireysel düzeydeki büyüme yörüngelerini modellemek ve zaman içinde gözlenen değişikliklerin altında yatan gizli yapıların (gizli faktörler/gizli sınıflar) incelenmesinde; Yapısal Eşitlik Modellemesinin ise hem gözlenen hem de gizli değişkenler dahil olmak üzere değişkenler arasındaki karmaşık ilişkileri boylamsal bir bağlamda incelenmek istenildiğinde kullanıldığı (Marsh & Yeung, 1997; Xu & Gao, 2014) görülmektedir. Boylamsal verilerde analiz seçimi, yapılan araştırmanın amacına, araştırılan hipoteze ve veri özelliklerine bağlı olarak değişmektedir. Zhu (2013), boylamsal veri analizi “sanatının” teorik ilişkiler hakkındaki varsayımlara, belirli veri yapılarına ve yanlılığın kaynağına ilişkin belirli varsayımlara dayandığına dikkat çeker. Bu çalışmalarda değinilen nokta, veri analizinde, bir modelin geçerli tahmini açısından önemli hususlarının, teorik ilişkilerinin, veri yapılarının ve araştırmacıların varsayımlarının çok önemli olduğudur (Caruana ve diğeri, 2015). Bu konuda dikkat edilmesi gereken konulardan biri de elde edilen verilerin modellenmesinde uygun yöntemin kullanılması sorunudur. Veri yapısına uygun analizlerin tercih edilmemesi hem analizlerin güvenilirliğini etkilemektedir hem de çıkan sonuçların yanlış yorumlanmasına ve yanlış kararlar verilmesine neden olmaktadır. Kwok vd. (2018)’e göre araştırmacılar, karmaşık boylamsal verileri, özellikle farklı eğitim ortamlarından gelen verileri analiz ederken, modelin eşit derecede önemli rasgele kısmını (yani, rasgele etki varyansları) göz ardı ederek genellikle yalnızca ortalama kısma (yani regresyon katsayılarına) odaklanırlar. Çoğu eğitim araştırmacısı, verilerini analiz ederken karmaşık veri yapısını dikkate almanın önemini fark etse de, analizlerinde karmaşık veri yapısını tam olarak nasıl ele alacaklarının farkında olmayabilirler ve sonuç olarak, yalnızca kısmen dikkate alabilirler.

Son olarak çok düzeyli veriler söz konusu olduğunda göz önünde bulundurulması gereken bir diğer konu da varsayımların kontrol edilmesidir. Değişkenliğin birey içi ve

bireyler arası bileşenlerine ilişkin parametrik varsayımlar yapılırsa, grup ortalamasını karakterize eden regresyon parametrelerini ve birey içi ve bireyler arası heterojenliğin büyüklüğünü karakterize eden varyans bileşenlerini tahmin etmek için maksimum olasılık yöntemleri kullanılabilir. Çok düzeyli verilerde bağımlı değişkenler için, birey içi hataların normal olarak dağıldığını ve kesişmelerin ve eğimlerin bireyler arasında normal olarak dağıldığını varsaymak uygundur (Van Belle ve diğerleri, 2004). Genellikle, klasik regresyon modellerinde, bağımlı değişken ile bağımsız değişken(ler) arasındaki ilişki koşullu ortalama bir fonksiyon tanımlanarak analizler gerçekleştirilir. Bu fonksiyon, bağımlı değişkenin ortalama değerine odaklanarak bağımsız değişken(ler) ile olan ilişkiyi açıklamaktadır. Bağımlı değişken normal dağılım göstermediği durumlarda, bağımlı değişkenin ortalama değerine odaklanmış olan bu koşullu fonksiyon bağımsız değişken(ler)le olan ilişkiyi tanımlamada yeterli olmamaktadır. Çok düzeyli regresyon analizleri, boylamsal veri analizlerinde yaygın olarak kullanılsa da verilerin karmaşık doğasını tam olarak yakalayamadığı ya da eksik/yetersiz kalabildiği bazı durumlar olmaktadır. Örneğin çok düzeyli regresyon teknikleri, değişkenler arasında doğrusal ilişkilerin olduğunu varsayar ve zaman içinde doğrusal olmayan değişiklikleri yeterince açıklayamayabildiği gibi zamanla değişen ortak değişkenlerin varlığı, gözlemlerin bağımsız olmaması gibi durumlarda da verimli olmamaktadır. Koenker ve Basset (1978) tarafından geliştirilen Kantil Regresyon yöntemi, bağımlı değişkenin normal dağılıma sahip olmadığı, uç değerlerin baskın olduğu veya bağımlı değişkenin dağılımın belli aralıkları (kantil değerleri) arasındaki farklılıkların araştırıldığı durumlarda klasik doğrusal regresyon yöntemine alternatif olarak geliştirilmiştir. Kantil regresyon, klasik doğrusal regresyon yöntemindeki gibi tek bir regresyon katsayısı vererek model kurulması yerine, incelenen farklı kantil aralıkları için farklı regresyon katsayıları vererek araştırmacılara yeni perspektifler sunmaktadır. Böylece kantil regresyon yönteminin en önemli avantajı, modellemede kullanılan bağımsız değişken(ler)in, bağımlı değişkeni yordamadaki etkisini araştırmacının istediği farklı kantiller boyunca ayrı ayrı incelenmesine olanak tanınması olarak görülmektedir.

Tüm bu durumlar eşliğinde, özellikle sosyal bilimler ve eğitim alanındaki veri yapılarının bazı durumlarda varsayımları sağlaması zor olduğu da göz önünde bulundurulduğunda, doğru modellemeler yapılabilmesi için farklı çarpıklık düzeylerine sahip karmaşık veri yapılarında farklı regresyon yöntemlerinin verimlilikleri araştırılmalıdır. Ayrıca karmaşık veri yapılarının ilgililenen belirli alt dilimlerinde, veri gruplarının ayrı ayrı değerlendirilmesi gerektiği, uygulanması gereken uygun yöntemlerin belirlenmesi ve bu yöntemlere dikkat çekilmesi, elde edilen bulguların açık, anlaşılır şekilde araştırmacılara sunulması gerekmektedir.

Bu çalışmada hem klasik regresyon varsayımlarını karşılayan uygun veri yapılarının analize dahil edilmesinin hem de varsayımları karşılamayan veri yapıları durumunda kullanılacak alternatif çok düzeyli regresyon analizlerinin tanıtılarak bu yöntemlerin verimliliğinin belirlenmesi amaçlanmaktadır. Ülkemizde eğitim alanında boylamsal çalışmalara sık yer verilmemesi ve böylece öğrenci gelişim düzeylerinin incelenmesi, takip edilmesinin önemine dikkat çekilmesi de alanda yapılacak çalışmalara katkı sağlayacaktır.

Araştırmanın Amacı ve Önemi

Eğitimde ve istatistikte bağımlı ve bağımsız değişkenlerin arasındaki ilişkileri incelenmek için kullanılan yöntemler arasında regresyon analizleri de yer almaktadır. Regresyon analizi, aralarında ilişki bulunan iki ya da daha fazla değişken arasındaki ilişkiyi belirlemek ve bu ilişkiyi kullanarak o konu ile ilgili yordamalar/kestirimler yapabilmek amacıyla kullanılan bir yöntemdir (Birkes & Dodge, 1993). Regresyon analizi ile araştırmacı "değişkenler arasında ilişki var mıdır, varsa bu ilişkinin gücü nedir?" sorularına yanıt aramaktadır. Bu araştırma soruları ile yola çıkılan çalışmalarda Türkiye'de eğitim alanında birçok çalışmanın yer aldığı görülmektedir (örn. Fan, 2001; Aşkar & Yurduğül, 2009; Acar & Öğretmen, 2012; Atar, 2010; Atar, 2014; Yapar, 2014; Yavuz ve diğerleri, 2019).

Alanyazında çok düzeyli modelleri kullanan arařtırmalar incelendiğinde bu çalışmaların son yıllarda sayılarının da giderek artmakta olduđu gör÷lmektedir. Bu durumun yanı sıra boylamsal verilerde varsayımları göz ardı edilmekte ya da varsayımların karşılanmadığı durumlar için kullanılabilir alternatif yöntemler konusundaki çalışmalar kısıtlıdır (Geraci & Bottai, 2007; Harding & Lamarche, 2009; Liu & Bottai, 2009; Farcomeni, 2012; Galvaro, 2011; Rahman & Vossmeier, 2019). Bu durumun sebepleri arasında arařtırmacıların veri yapılarını incelemekten tek düzeyli regresyon yöntemlerini tercih etmeleri ya da boylamsal veri yapılarına uygun çok düzeyli regresyon analizi yöntemleri hakkındaki bilgi eksikliği olabileceği düşün÷lmektedir. Bu sebeple bu çalışmada, çok düzeyli regresyon analizi yöntemlerinin eksik kaldığı durumlara dair bilinenleri doğrulamak ve açıklamakla beraber, alternatif olarak sunulan kantil regresyon yöntemi tanıtılmıştır. Bu çalışmada farklı koşullar altında (farklı örneklem büyüklüğü ve farklı düzey 1 /düzey 2 hata terimlerinin dağılımlarında) simüle edilmiş boylamsal veri yapılarında çok düzeyli regresyon ve çok düzeyli kantil regresyon yöntemlerinin performansları karşılaştırılması amaçlanmıştır. Bu arařtırmanın diđer bir amacı da, farklı koşullarda simüle edilmiş veri yapılarında kantil regresyon yönteminin klasik regresyon yöntemlerine göre sağladığı avantajları açıklamakla birlikte, kantil regresyonun tercih edilmediği durumları analiz etmek ve farklı koşullarda simüle edilen veri yapılarının analizleri sonucu elde edilen eğim ve kesişim parametrelerinin farklılaşmalarını incelemektir. Son olarak, klasik regresyon analizlerinin varsayımlarının ihlali durumunda hangi çarpıklık değerlerine dayanıklı olduğunu açıklamak da arařtırmanın amaçları arasında yer almaktadır.

Yapılan eğitim alanındaki arařtırmalarda çok düzeyli veri yapısına sahip boylamsal arařtırmaların sınırlı olması, ölçümlerin elde edilmesinde katılımcıların değerlendirme sayısı bakımından farklılık göstermesi ve rasgele olması nedeniyle her sene ölçüm değerlerinin elde edilememiş olması ve rasgele olmayan gözlem kayıpları nedeniyle bu çalışma, arařtırmacılara belirli koşullara sahip veri setlerinde hangi analizlerin yapılması

gerektiği konusunda yol gösterici olması açısından önemlidir. Bu doğrultuda bu çalışma ile eğitim bilimleri alanında bireylerin sosyoekonomik düzey, cinsiyet, göçmenlik durumu, ırk gibi farklı faktörlere sahip öğrencilerin puanlarının, tutumlarının, ilgilerinin vb. incelenmesinde farklılıklardan kaynaklanan durumların ait olduğu alt veri setleri (veri dilimleri) açısından incelenmesinin önemine dikkat çekmesi amacıyla önemli olduğu düşünülmektedir. Ayrıca alanyazın tarandığında boylamsal çalışmalarla daha çok sosyal bilimler, biyoistatistik, ekonometri gibi alanlarda çalışmalar yapıldığı ve eğitim bilimleri alanı ile terminoloji açısından farklı kullanımların olduğu da görülmüştür. Bu çalışmayla diğer tüm alanlarla birlikte ortak bir terminolojide ilerlenmesi de ve bu terminolojinin eğitim bilimleri alanına tanıtılması da yönüyle de araştırmanın önemleri arasında yer almaktadır. Ayrıca Türkiye’de eğitim bilimleri alanında boylamsal veri yapılarında çok düzeyli regresyon analizinde farklı koşullarda tasarlanmış veri setlerinin incelenmesi üzerine henüz çalışılmamış olması da bu çalışmanın özgünlüğünü ortaya koymaktadır.

Araştırma Problemi

Bu araştırmada bireylere ait farklı örneklem büyüklüğü ve Düzey 1 ile Düzey 2 hata terimlerinin farklı dağılımlara sahip bağımlı değişkene sahip veri yapılarında çok düzeyli regresyon (rasgele kesişim ve rasgele eğim modeli) ve çok düzeyli kantil regresyon analiz yöntemlerinin performanslarının karşılaştırılmıştır. Bu kapsam doğrultusunda aşağıdaki araştırma sorularına yanıt aranmıştır:

Alt Problemler

1. Simüle edilmiş boylamsal veri setlerinde hata düzeylerinin farklı dağılımlarına göre (Normal- Kikare- Student t(3)), her bir örneklem büyüklüğü için (50; 500; 1000), çok düzeyli klasik regresyon analizinde ve kantil regresyonun farklı kantil düzeylerinde (0.10; 0.25; 0.50; 0.75; 0.90) **model veri uyumu değerleri** (AIC-BIC-logLike) nasıl değişmektedir?

2. Simüle edilmiş boylamsal veri setlerinde, farklı örneklem büyüklüğü ve farklı Düzey 1 ile Düzey 2 hata terimlerinin dağılımlarında, çok düzeyli klasik regresyon analizi ile çok düzeyli kantil regresyon analiz yöntemlerinin **performansları** nasıldır?

3. Simüle edilmiş boylamsal veri setlerinde, farklı Düzey 1 ile Düzey 2 hata terimlerinin dağılımları ve farklı örneklem büyüklüklerinde, çok düzeyli klasik regresyon analizi ile çok düzeyli kantil regresyon analiz yöntemlerinden elde edilen ortalama kesişim ve eğim katsayıları **önemli ölçüde farklılık göstermekte** midir?

Sınırlılıklar

Matematik başarısı bilişsel ve duyuşsal olarak ayrı ayrı ele alınabilecek bir yapıdır. Simülasyon çalışmasında kullanılması planlanan parametrelerin değerleri PIAT veri setinden elde edilen kestirimler ile sınırlıdır. Bu çalışmada çok düzeyli kantil regresyon analizinde kullanılan R paketinin (qrlmm) kütüphanesinde yer alan modellemelerin değişken sayısını kısıtlı tutması nedeniyle seçilen değişken sayısı çalışmanın sınırlılıklarındandır. Ayrıca R paketinin (qrlmm) 4 tekrarlı 27 bireye ait gözlemlerin analizlerini 7.59 dakikada yapması sebebiyle örneklem sayısı ile iterasyon sayısının sınırlı tutulması çalışmamızın sınırlılıklarındandır.

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu bölümde kuramsal temelleri iki kısımda ele alınacak ve son bölümde ilgili araştırmalar sunulmaktadır. İlk kısımda regresyon yöntemi ve kantil regresyon yöntemi tanıtılacak ardından ikinci kısımda çok düzeyli veri yapılarında kullanılan klasik regresyon ve kantil regresyon yöntemlerine değinilmiştir. Son olarak, boylamsal verilerde çok düzeyli kantil regresyon modeli ile ilgili bilgiler ve bu konuda yapılmış ilgili araştırmalar sunulmuştur.

Regresyon Yöntemi

Regresyon teknikleri, bağımsız değişkenlerin bağımlı değişkenle değişen derecelerde ilişkili olduğu bir veri setine uygulanabilir. Bu analiz yöntemi, bağımlı değişken olarak ifade edilen Y ve bağımsız değişken olarak kullanılan X değişken(ler)i arasındaki ortalama ilişkinin matematiksel bir fonksiyonla açıklanması olarak da tanımlanabilir. Bu ilişkiyi açıklamak için kurulan model "regresyon modeli" olarak adlandırılır (Barnett ve Lewis, 1994; Tabachnick ve Fidell, 2007).

Regresyon teknikleri bilinen sayısal bir değişkenden bilinmeyen, aralarında ilişki olan bir başka sayısal değişkeni tahmin için uygulanırsa basit doğrusal regresyon (simple linear regression) olarak tanımlanır. İki değişken arasındaki ilişki

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (1)$$

eşitliği ile gösterilir. ε_i lerin sıfır ortalamalı, eşit varyanslı σ^2 hata terimleri, bağımsız gözlenemeyen rasgele değişkenler olduğu varsayıldığında, X değişkenine bağımsız değişken, Y değişkenine bağımlı değişken denilmekte ve β_0 , β_1 katsayıları regresyon parametreleri olarak tanımlanmaktadır (Alpar, 2013).

Diğer bir yandan basit doğrusal regresyon modelinin bağımlı değişkenini etkileyebilecek, bağımlı değişkenlerdeki değişimlerin nedeni olabilecek değişkenlerden

yararlanarak bir bağımlı değişkeni yordamak amacıyla modelleme yapıldığında ise regresyon modeli, iki değişkenli regresyon analizinin bir uzantısı olan “çoklu doğrusal regresyon” (multiple linear regression) olarak tanımlanır. Kurulan matematiksel model;

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon_i \quad i = 1, 2, \dots, n \quad (2)$$

eşitliği ile gösterilir. Tanımlanan modelde β_0 sabit parametre, $(\beta_1, \beta_2, \dots, \beta_k)$ katsayıları ise kısmi regresyon katsayılarıdır (Alpar, 2013; Tabachnick ve Fidell, 2007).

Bağımlı değişken Y_i , hata terimi ε_i 'nin doğrusal bir fonksiyondur. Basit doğrusal regresyonun temel varsayımlarından biri hataların ε_i 'ler normal dağılmasıdır (Alpar, 2013).

Regresyon analizinde değişkenler uygun kuramsal çerçevede tanımlanmalıdır. Regresyon analizi değişkenler arasındaki ilişkiyi açığa çıkarır fakat bu durum ilişkilerin nedenselliği anlamı taşımamaktadır. Regresyon denkleminin en iyi halini alması için bağımsız değişken seçimine dikkat edilmelidir. Her bir bağımsız değişkenin bağımlı değişkenle güçlü ilişkili ve diğer bağımsız değişkenlerle ilişkisiz olması önem göstermektedir. Ayrıca veri setlerinin yeterli büyüklüğe sahip olması, çoklu doğrusal bağıllık olmaması ve uç değer içermemesi de analiz öncesi kontrol edilmesi gereken durumlardandır (Tabachnick & Fidell, 2007). Uç değer, bir veri kümesinde gözlemlerin çoğunun sahip olduğu dağılıma veya modele uymayan gözlemler olarak ifade edilebilir (Barnett & Lewis, 1994). Uç değer içeren veri kümesinde varsayımların sağlanamamasından dolayı kurulan regresyon modelinden alınan sonuçlar da yanıltıcı olmaktadır (Goodal, 1983; Ryan, 1997). Uç değerlerin veri kümesinden çıkartılması regresyon denklemini tamamen veya kısmen değiştirebilmektedir. Bu nedenle büyük artık değerlere sahip olan gözlemler, regresyon analizinde oldukça etkilidirler. Böyle durumlarda uç değerlerin tespiti ve sonuçların güvenilirliği için güçlü regresyon yöntemlerini tercih etmek daha uygundur (Rousseuw & Leroy, 1987). Tüm bu nedenlerden ötürü regresyon analizi öncesinde verilerin incelenmesi oldukça önemli bir yer tutmaktadır.

Regresyon analizi bazı varsayımlara dayanmaktadır. Bu varsayımlar;

- Hataların normal dağılıyor olması,
- Hataların ortalamasının sıfır olması ($E(\varepsilon_i)=0 \quad i=1, 2, 3, \dots, n)$,
- Hata terimlerinin varyanslarının eşit olması ($\text{Var}(\varepsilon_i)=\sigma^2 \quad i=1, 2, 3, \dots, n)$,
- Hatalar arasında otokorelasyonun olmaması ($E(\varepsilon_i, \varepsilon_j)=0 \quad i \neq j$),
- Bağımsız değişkenlerin tesadüfi değişken olmamasıdır (Tabachnick & Fidell, 2007).

Regresyon analizinin amacı, denklemden yordanan Y değerlerini ölçme sonucu elde edilen Y değerlerine mümkün olduğu kadar yaklaştıran ve regresyon katsayıları olarak adlandırılan bağımsız değişkenlere ait β değerleri setine ulaşmaktır (Tabachnick & Fidell, 2007).

Çoklu regresyonda parametre kestirimleri standartlaştırılmamış regresyon katsayıları yani β ağırlıklarıdır. Bir bağımsız değişken için β ağırlığı, diğer tüm bağımsız değişkenler sabit tutulduğunda, bağımsız değişkendeki bir birim değişiklikte bağımlı değişkendeki değişimi temsil eder. Parametre kestirimlerinin doğruluğu, bağımsız değişkenlerin hatasız ölçüldüğü varsayımı dahil çoklu doğrusal regresyon analizlerinin varsayımlarının karşılanmasına bağlıdır (Tabachnick & Fidell, 2007).

Regresyon analizinde en çok kullanılan tahmin yöntemlerden bir tanesi En Küçük Kareler (EKK) tahmin yöntemidir. Bu yöntemde parametrelerin tahmini, gözlenen (X_i, Y_i) veri çiftlerinin oluşturduğu noktalar ile örneklemin regresyon eğrisi $Y_i = \beta_0 + \beta_1 X_1$ değerlerini verdiğiinden gözlemlenen Y değerleri ile bu eğri arasındaki uzaklık hata terimlerini (ε_i) vermektedir. EKK yöntemi bu regresyon doğrusu arasındaki sapmaların (hata terimlerinin) karelerinin toplamını ($\sum \varepsilon_i^2$) en küçük yapacak şekilde katsayı tahmin değerlerini hesaplar. Burada karelerin alınmasının nedeni ise toplam alınırken artı ve eksi değerlerin birbirini götürmesini engellemektir (Hamilton, 1992). Bu yöntemde Y_i ile $E(Y_i)$

arasındaki farkların kareleri toplamı minimize edilerek çoklu doğrusal regresyonun normal denklemleri elde edilerek

$$\beta_1 = \mu(Y) - \beta_2 \mu(X_2) - \beta_3 \mu(X_3) - \dots - \beta_k \mu(X_k) \quad (3)$$

olarak yordanacaktır.

Sıradan en küçük kareler regresyonunda, artıkların normal, homoskedastik (sabit bir varyansa sahip) ve korelasyonsuz olduğu varsayımını karşılamayabilir. Bazen bu tür değişkenler sıradan en küçük kareler varsayımlarını karşılamak için dönüştürülebilir ve diğer tüm doğrusal regresyon varsayımlarının karşılandığı varsayılarak daha sonra doğrusal regresyon kullanılır. Sürekli bir bağımlı değişken için uygun bir dönüşüm veya geleneksel istatistiksel metodoloji açık olmadığında, araştırmacı bunun yerine yordayıcı değişkenler ile bağımlı değişken arasındaki ilişkiyi belirli bir ilgi miktarında (belki de dağılımın uç noktalarında) veya çoklu kantillerde (yüzdelerinde) aynı anda incelemek isteyebilir. Örneğin, geleneksel yöntemler kullanılarak çarpık bir değişken üzerindeki bir öğretim yönteminin etkisini tahmin etmek için genellikle bir log dönüşümü kullanılabilir. Gruplar arasındaki fark, ortalama öğretim yönteminin etkisini tahmin etmek için bir konum kayması kullanılarak log dönüştürülmüş değişkenin ortalamasındaki bir fark olarak uygun şekilde değerlendirilebilir.

Kantil Fonksiyonları

İstatistikte bir dağılımı betimlemek için kullanılan en yaygın ve en basit ölçekler; aritmetik ortalama ve standart sapma değerleridir. Ancak, tek başına aritmetik ortalama ve standart sapma değerlerinin kullanılması dağılım hakkında daha fazla bilgi veren önemli özelliklerin ihmal edilmesine de neden olmaktadır. Aynı zamanda birçok araştırmacı için bu ölçekler, çarpık bir dağılımın şeklini ve konumu açıklamanın en iyi yolu olarak görülmemektedir. Herhangi bir dağılımın şeklini ve konumu belirlemenin dört yolu sırasıyla; kümülatif dağılım fonksiyonu, olasılık yoğunluk fonksiyonu, kantil fonksiyonu ve kantil yoğunluk fonksiyonudur (Gilchrist, 2000; Hao & Naiman, 2007).

Kümülatif Dağılım Fonksiyonu

Bir bağımlı değişkenin simetrik olmayan şeklini açıklayabilmek için kümülatif dağılım fonksiyonundan yararlanır (Hao & Naiman, 2007). Klasik doğrusal regresyon analizinin varsayımlarından biri olan, hata terimlerinin normal dağılması ile ilişkili olan kümülatif dağılım fonksiyonu, kantil regresyon tahmincileri ile karşılaştırılmaktadır (Tareghian & Rasmussen, 2013). Kümülatif dağılım fonksiyonu $F(x)$ ile gösterilir (Hao & Naiman, 2007) ve herhangi bir değerinden küçük veya eşit olma olasılığını veren, monoton artan sürekli bir fonksiyon olarak tanımlanmıştır (Hao & Naiman, 2007; Huang, 2009; Gilchrist, 2000). Bu fonksiyon:

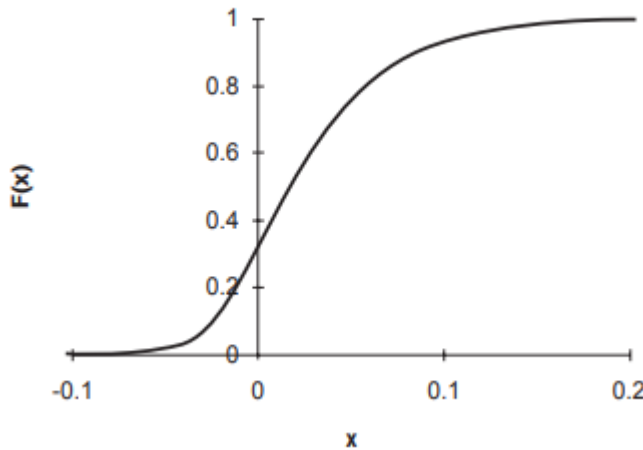
$$F(x) = \text{Olasılık}(\text{rassal değişken } X \leq x) \quad \text{yani;}$$

$$F(x) = P(\text{rassal değişken } X \leq x) \quad 0 \leq F(x) \leq 1, \quad x \in R \quad (4)$$

eşitliği ile gösterilebilir.

Şekil 1

Kümülatif Dağılım Fonksiyonu (Gilchrist, 2000).



Şekil 1 incelendiğinde, kümülatif fonksiyon değeri olan $F(x)$ 'in her zaman 0 ile 1 arasında değer aldığı ve x değerleri sağa doğru gittikçe 1'e yaklaştığı görülmektedir. Kümülatif dağılım fonksiyonu sürekli artan bir fonksiyondur ve bir dağılımın tek yönlü tanımını verir (Gilchrist, 2000; Huang, 2009).

Olasılık Yoğunluk Fonksiyonu

Bir dağılımı tanımlamanın ikinci yolu olan olasılık yoğunluk fonksiyonu, herhangi bir değişkenin alabileceği değerler ile bu değerleri alma olasılıkları arasındaki ilişkiyi göstermektedir (Saraçoğlu & Çevik, 1995). Olasılık yoğunluk fonksiyonu, kümülatif dağılım fonksiyonunun türevine eşittir (Gilchrist, 2000; Schulze, 2004). Bu fonksiyon;

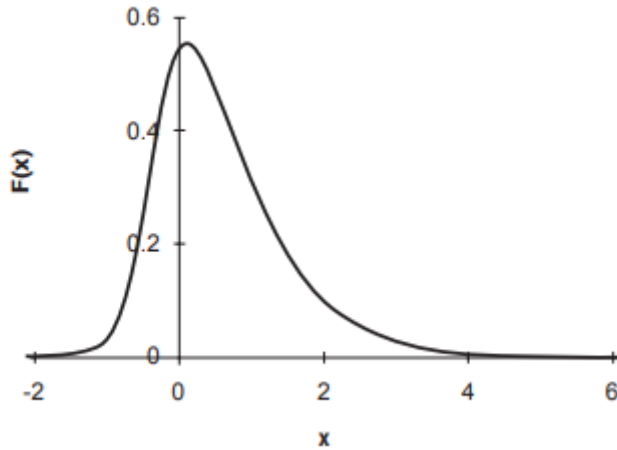
$$F(x)dx = \text{Olasılık}(x \leq \text{rassal değişken } X \leq x + dx) \text{ yani};$$

$$F'(x)dx = P(x \leq X \leq x + dx) \quad (5)$$

eşitliği ile gösterilir.

Şekil 2

Olasılık Yoğunluk Fonksiyonu (Gilchrist, 2000).



Şekil 2 ile gösterilen olasılık yoğunluk fonksiyonunda $F(x)$ eğrisinin altında kalan alan, gözlemlenen değerlerin olasılık değerlerinin toplamına yani 1'e eşittir. Ortalama değer 0 ve standart sapma değerinin 1 olması halinde bu fonksiyon standart normal dağılıma sahiptir (Rousass, 1997).

Kantil Fonksiyonu

Bir dağılımın şeklini açıklamanın üçüncü yolu ise kantil fonksiyondur (Gilchrist, 2000).

$x_p = \text{Olasılık } (X \leq x_p) = p$ olan x değeri olmak üzere, $x_p = Q(p)$ kantil fonksiyonu, X değerinin x_p 'den küçük olma olasılığını belirtir.

$$Q(p) = P(X \leq x) = p \quad (6)$$

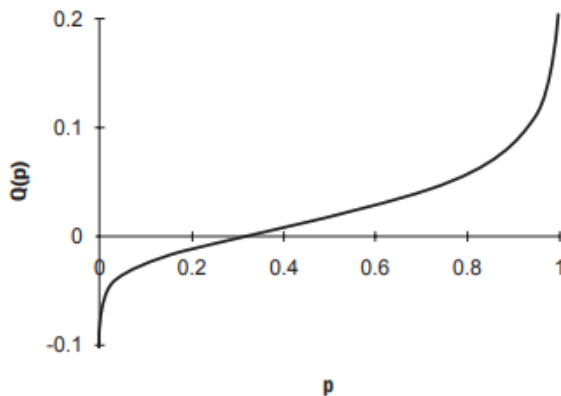
olarak gösterilir. $Q(p)$, F kümülatif dağılım fonksiyonunun kantil fonksiyonu olarak adlandırılır.

$$F(x) = P(X \leq x), \quad Q(p) = F^{-1}(p) = \inf\{x: F(x) \geq p\} \quad 0 < p < 1, \quad (7)$$

şeklinde ifade edilir. Burada $\inf\{x: F(x) \geq p\}$ gösterimi, $F(x)$ fonksiyonunun p 'den büyük veya ona eşit olduğu x değerler kümesinin alt sınırını temsil eder.

Şekil 3

Kantil Fonksiyonu (Gilchrist, 2000).



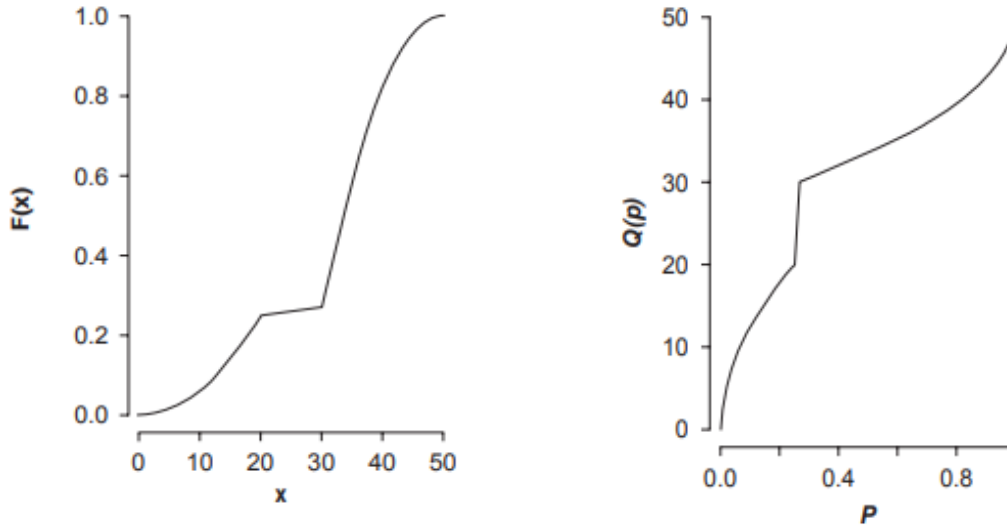
Şekil 3'e göre $Q(p)$, X 'in p . kantili olarak adlandırılır. Kantil değeri değişken dağılımında yer alan ve dağılımı kendisinden büyük ve kendisinden küçük değerler olarak ikiye bölen herhangi bir değerdir. Değerlerin % p 'si, p . Kantilden (p olasılık değerini ifade etmektedir) daha küçüktür. Kantil fonksiyonunda medyan değeri $Q(0.5)$ ile 1. ve 3. kartiller de sırasıyla $Q(0.25)$ ve $Q(0.75)$ ile ifade edilir (Koenker, 2005). Kantil değeri toplam frekansın yüzde kaçının bu değer altında olduğunu göstermektedir. Örneğin, birinci kantil değeri olan $Q(0.25)$ değeri, toplam frekansın ilk %25'lik kısmı bu değer altında olduğunu ifade etmektedir. Şekil 3 incelendiğinde monoton bir şekilde artan ve sola doğru sürekli bir

fonksiyon olduđu gör÷lmektedir (Schulze, 2004). Kantil ölçekleri aykırı deđerlerden daha az etkilenir. Bu sebeple kantil fonksiyonu, yoğun kuyruklu dađılımlarda, dađılım fonksiyonuna bir alternatif olarak da gör÷lmektedir (Soni ve diđerleri, 2012).

Küm÷latif dađılım fonksiyonu ile kantil fonksiyon birbirlerinin tersine eşittir (Parzen, 1979; Hao & Naiman, 2007) ve her ikisi de sürekli artan fonksiyonlardır. Bu durum Şekil 4'te gösterilmiştir.

Şekil 4

Küm÷latif Dađılım Fonksiyonu ile Kantil Fonksiyonun Karşılaştırılması (Hao & Naiman, 2007).



(a) Küm÷latif dađılım fonksiyonu

(b) Kantil fonksiyonu

Şekil 4'e incelendiğinde (a) panelinde küm÷latif dađılım fonksiyonu, (b) panelinde ise küm÷latif dađılım fonksiyonunun tersi olan kantil fonksiyonu yer almaktadır.

Kantil Yođunluk Fonksiyonu

Olasılık yođunluk fonksiyonu, küm÷latif dađılım fonksiyonunun türevine eşit olduđu gibi kantil yođunluk fonksiyonu da kantil fonksiyonun türevine eşittir (Gilchrist, 2000). Dađılımları tanımlamak ve modellemek için gerekli deđerleri elde etmede kullanılır.

$$q(p) = \frac{dQ(p)}{dp} \quad (8)$$

Kantil fonksiyon azalmayan bir fonksiyon olduğu için eğimi hiçbir zaman negatif olamaz. Bu sebeple $q(p)$ kantil yoğunluk fonksiyonu daima $0 \leq p < 1$ arasında yer alır (Gilchrist, 2000).

Bu anlamda kantil yoğunluk fonksiyonu, olasılık yoğunluk fonksiyonundan farklılaşmaktadır; çünkü olasılık yoğunluk fonksiyonu sonsuz tanım aralığına sahiptir. Jone (1992) tarafından kantil yoğunluk fonksiyonu ve olasılık yoğunluk fonksiyonu arasındaki ilişki şöyle tanımlanır:

$$f(x) = \frac{1}{q(Q^{-1}(x))} \text{ iken } q(p) = \frac{1}{f(F^{-1}(p))} \quad (9)$$

Kantile Dayanan Konum ve Şekil Ölçüleri.

Bir dağılımın şeklini iki temel özellik tanımlar: ölçek ve çarpıklık. Geleneksel olarak ölçek, veri noktalarının ortalamadan sapmalarının ikinci dereceden bir fonksiyonunu içeren bir dağılımın ikinci momentine dayanan standart sapma ile ölçülür. Bu ölçünün simetrik bir dağılım için yorumlanması kolaydır, ancak dağılım oldukça asimmetrik hale geldiğinde yorumu bozulma eğiliminde olup ağır kuyruklu dağılımlar için de yanıltıcıdır. Sosyal bilimler alanındaki çalışmalarda kullanılan dağılımların çoğu çarpık veya ağır kuyruklu olduğundan, ölçeklerini karakterize etmek için standart sapmayı kullanmak sorunlu hale gelir. Bir dağılımın yayılmasını standart sapmaya dayanmadan yakalamak için, yayılmayı seçilen bir p 'de kantil tabanlı ölçek ölçüsü (quantile-based scale measure-QSC) geliştirilmiştir (Hao & Naiman, 2007; Ally, 2010).

$$QSC^{(p)} = Q^{1-(p)} - Q^{(p)}, \quad p < 0.5 \quad (10)$$

(10) eşitliğinde görüldüğü üzere dağılımın orta kısmında kalan belirli bir kısımdaki yayılımı ölçmek için birbirlerinin tümleyeni olan iki kantil fonksiyon değerinin farkı alınır. Örneğin; dağılımın %90'lık orta kısmının yayılımını açıklanmak için $Q(0.95)$ değerinden $Q(0.05)$ değeri çıkarılır.

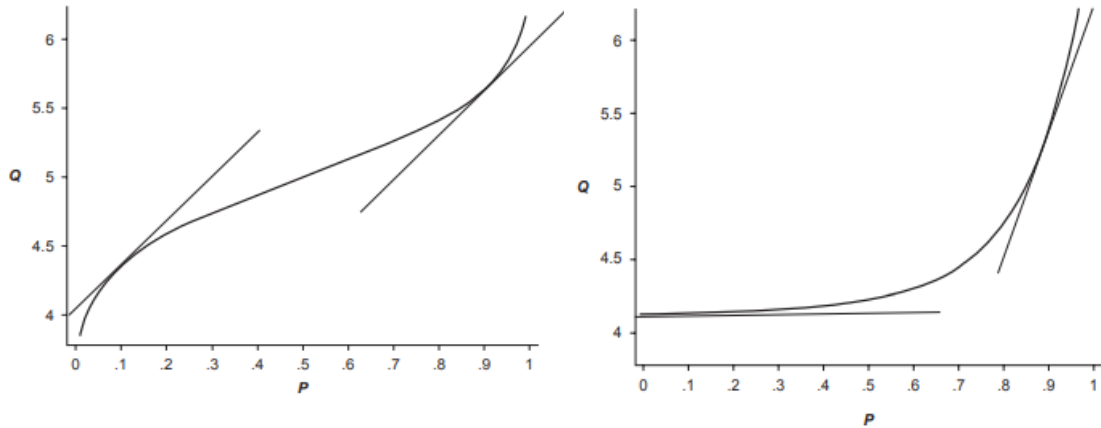
Normal dağılımlarda birbirinin simetriği olan iki kantil fonksiyonun eğimleri birbirine eşittir. Yani 0,25 ve 0,75 kantillerinin eğimleri eşit değer almaktadır. Ancak çarpıklık durumu farklı olan dağılımlarda bu değerler asimetric olmaktadır. Bu yüzden kantil tabanlı çarpıklık ölçüsü (quantile-based skewness-QSK) geliştirilmiştir (Hao & Naiman, 2007).

$$QSK = \left(\frac{Q^{(1-p)} - Q^{(0.5)}}{Q^{(0.5)} - Q^{(p)}} \right) - 1 \quad p < 0.5 \quad (11)$$

Şekil 5'in (a) paneli incelendiğinde 0,10 ve 0,90 kantillerinin eğimlerinin birbirlerine eşit olduğu gözlemlenebilir. Ancak Şekil 5'in (b) paneli incelendiğinde sağa çarpık bir dağılımda kantil fonksiyon değerleri asimetric olur.

Şekil 5

Normal Kantil Fonksiyonu ile Sağa Çarpık Kantil Fonksiyonun Karşılaştırılması (Hao & Naiman, 2007).



(a) Normal kantil fonksiyonu

(b) Sağa çarpık kantil fonksiyonu

Bir dağılımının merkezi, dağılıma ait bir noktadan ortalama kareli sapmanın yani $E[(Y - \mu)^2]$ ifadesinin minimum olduğu noktanın bulunmasıyla elde edilebilir. Benzer şekilde, n büyüklüğünde bir örneklem için ortalama kareli sapma, $\frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2$ ifadesini minimum yapan noktasının bulunması bir minimizasyon probleminin çözümleri arasında yer alır. Medyan (m) de benzer bir minimizasyon özelliğe sahiptir yani dağılımın verilen bir

noktasının, medyan değerine ortalama mutlak uzaklığı $E|Y - m|$ yardımıyla ölçülebilir. Benzer şekilde, bir örneklem ile çalışıldığında m değerinden ortalama mutlak uzaklık, $\frac{1}{n} \sum_{i=1}^n |y_i - m|$ değeri kadar olacaktır (Hao & Naiman, 2007).

Kantil regresyonun medyan regresyonun genişletilmiş hali olduğu bilindiğinden (Shulze, 2004), medyanın bu gösterimi diğer kantiller de için ifade edilerek genelleştirilebilir. Herhangi bir $p \in (0,1)$ değeri için, Y 'den verilen bir q kantili değerine uzaklık, mutlak uzaklık ile ölçülür. Ancak; Y değerinin q kantilinin sağında ya da solunda olmasına göre farklı bir ağırlıklandırma ile hesaplama yapılır. Y değerinden q kantiline olan uzaklık,

$$d_p(Y, q) = \begin{cases} (1-p)|Y - q| & Y < q \\ p|Y - q| & Y \geq q \end{cases} \quad (12)$$

şeklinde gösterilir. q değeri, p . kantile eşit olduğunda uzaklık değeri ise en küçük, p . kantil değeri, q değerine eşit olduğunda ise ortalama uzaklık en küçük olur. Bu durum

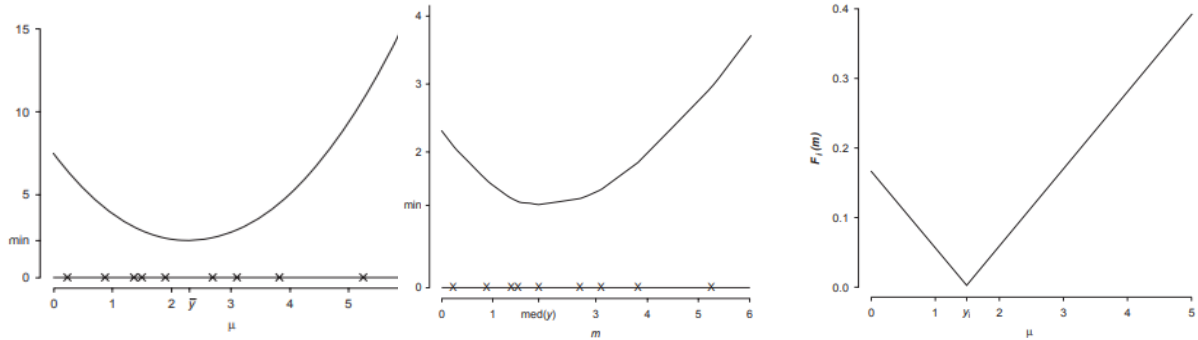
$$\frac{1}{n} \sum_{i=1}^n d_p(y_i, q) = \frac{1-p}{n} \sum_{y_i < q} |y_i - q| + \frac{p}{n} \sum_{y_i \geq q} |y_i - q| \quad (13)$$

şeklinde ifade edilir (Hao & Naiman, 2007).

Şekil 6'da (a) panelinde minimizasyon probleminin çözümleri olarak ortalama yani μ 'den ortalama kareli sapma, (b) panelinde minimizasyon probleminin çözümleri olarak medyan yani m 'den ortalama uzaklık ve (c) panelinde bir minimizasyon probleminin çözümü için medyanı tanımlamak için kullanılan V şekilli kontrol fonksiyonlar gösterilmiştir.

Şekil 6

Minimizasyon Sorunlarını Tanımlamak İçin Kullanılan Farklı Fonksiyonlar (Hao & Naiman, 2007).

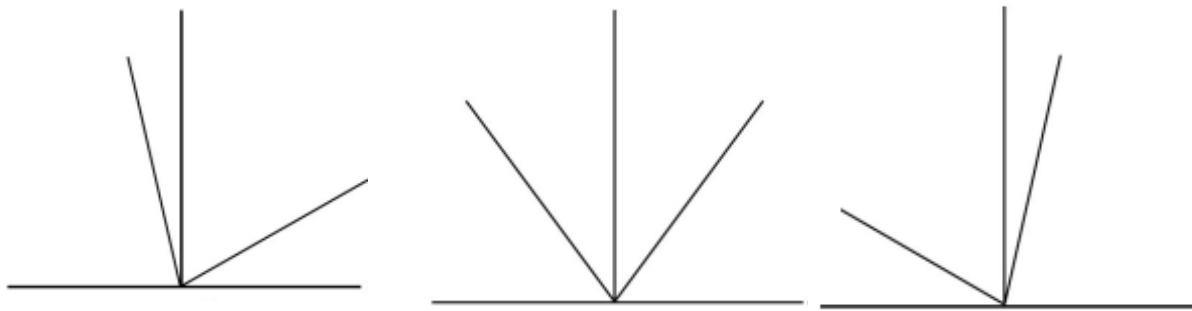


(a) μ 'den ortalama kareli sapma (b) m 'den ortalama uzaklık (c) V şekilli kontrol fonksiyonu

Kontrol fonksiyonun farklı kantillerdeki görüntüleri ise Şekil 7' de gösterilmiştir.

Şekil 7

Kontrol Fonksiyonun Farklı Kantil Değerlerindeki Görüntüleri (Park ve diğerleri, 2012).



(a) Kantil değeri=0.25 (b) Kantil değeri=0.50 (c) Kantil değeri=0.75

Kantiller bir optimizasyon probleminin çözümü ile elde edilir. Bu optimizasyon problemi, rasgele değişkenlerin önsel dağılım fonksiyonlarından yararlanarak nokta tahmini yapma problemi olarak düşünülebilir. Burada amaç, beklenen kaybı minimize edecek y_i tahminlerini bulmaktır (Koenker, 2005).

Kantil Regresyon Modeli

Kantil regresyon modeli aslında basit bir konum (location) modelidir. Kantil regresyon, dağılım grafiklerinde hem konumu hem de yayılmayı incelemek için etkili bir araçtır. Basit konum modeli

$$Y_i = \beta + \varepsilon_i \quad (14)$$

olarak ifade edildiğinde, burada yer alan Y_i , simetrik F dağılım fonksiyonuna sahip, bağımsız, özdeş dağılımlı, β medyanlı tesadüfi değişkendir. Bu ifadeyi doğrusal regresyon modeli;

$$Y_i = X_i\beta + \varepsilon_i \quad (15)$$

için genelleştirdiği durumda; X_i , bağımsız değişken vektörüdür. Ayrıca ε_i bağımsız, sıfır etrafında simetrik ve F dağılımına sahiptir. Bu modelden yola çıkılarak kantil regresyon modeli şu şekilde ifade edilir:

$$Y_i = \beta_0^{(p)} + \beta_1^{(p)}X_i + \varepsilon_i^{(p)} \quad 0 < p < 1 \quad (16)$$

Bu eşitlikte p kantil değerini ifade etmektedir. Klasik doğrusal regresyon analizinde Y 'nin koşullu ortalama değeri, herhangi verilen bir X_i değeri için;

$$E(Y_i|X_i) = \beta_0 + \beta_1 X_i \quad (17)$$

formülü ile hesaplanır. Bu eşitlikte hata terimi olan ε_i 'nin beklenen değeri sifıra eşit olmasına karşın kantil regresyon modelinde verilen herhangi bir X_i değeri için p . koşullu kantilin beklenen değeri;

$$Q^p(Y_i|X_i) = \beta_0^{(p)} + \beta_1^{(p)}X_i \quad (18)$$

eşitliği ile hesaplanır. Bu sayede p . koşullu kantil değeri, kantil parametreleri olan $\beta_0^{(p)}, \beta_1^{(p)}$ ve X_i değerine bağlı olarak değişir. Doğrusal regresyon modelinde olduğu gibi hata terimleri $\varepsilon_i^{(p)}$ 'nin beklenen değeri sifıra eşit kabul edilmektedir (Hao & Naiman, 2007).

Kantil regresyonda i . gözlem için hata değeri $\varepsilon_i^{(p)}$ değeri farklı kantillerde farklı değerlere sahip olur. Örneğin p kantil için ve herhangi bir q kantil için;

$$Q^p(Y_i|X_i) = \beta_0^{(p)} + \beta_1^{(p)}X_i + \varepsilon_i^{(p)} \quad \text{ve} \quad Q^q(Y_i|X_i) = \beta_0^{(q)} + \beta_1^{(q)}X_i + \varepsilon_i^{(q)} \quad \text{regresyon}$$

eşitliklerinde $\varepsilon_i^{(p)}$ ve $\varepsilon_i^{(q)}$ hata terimleri arasındaki fark hesaplandığında;

$$\varepsilon_i^{(p)} - \varepsilon_i^{(q)} = \beta_0^{(p)} - \beta_0^{(q)} + (\beta_1^{(p)} - \beta_1^{(q)})X_i \quad (19)$$

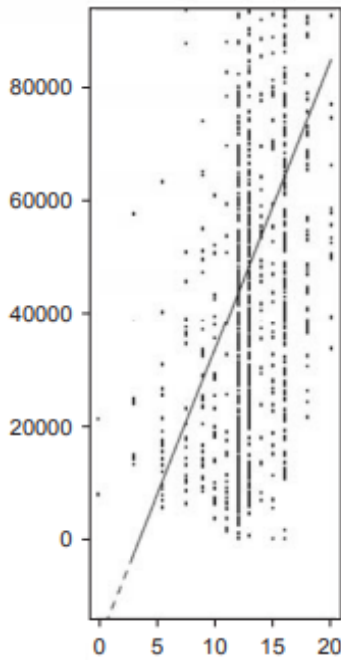
olacaktır. Bu hesaplama incelendiğinde, belirli bir i gözlem değerinin p ve q gibi iki farklı koşullu kantil regresyon denklemlerine ait hata terimleri arası fark sifıra eşit olmamaktadır. Dolayısıyla, hata terimlerinin i . gözlem için farklı kantillerde farklı değerlere sahip olacağı yorumu yapılabilir. Kısacası, hata terimleri gözlem değerlerine değil de kantillere bağlıdır. Ancak kantil regresyon yönteminde hata terimi olan $\varepsilon_i^{(p)}$ değerinin bağımsız olarak dağıldığı ve bu değer beklenen değerinin sifıra eşit olduğu varsayılır. En küçük mutlak sapmalar yönteminde olduğu gibi kantil regresyonda da hata terimleri gözlem değerlerinin $Q^{(p)}$ kantil doğrusuna dikey uzaklıkları olarak tanımlanmaktadır (Hao & Naiman, 2007).

Genel olarak kantil regresyon, artıkların mutlak değerlerinin toplamı ile hesaplanan asimetrik bir norm değerine dayanır. Buradaki asimetri, negatif ve pozitif artıkların farklı ağırlıklar alması anlamına gelmektedir. Artıkların karelerinin toplamıyla hesaplanan norm değerinin ise aynı şekilde asimetrik hale getirilebileceği hakkında sınırlı sayıda yöntem çalışması bulunmaktadır (Schnabel & Eilers, 2009). Tamamen dağılımdan bağımsız olan bu yaklaşım, bağımlı değişkenin koşullu kantili üzerindeki regresyon etkilerinin varsayıldığı (Koenker & Bassett 1978) ve modelleri büyük bir kantil kümesi için birleştirmek, daha sonra bağımlı değişkenin tam koşullu dağılımını karakterize etmeye izin veren bir fonksiyon tanımlanmasını gerektirir. Farklı kantillerde farklı sonuçların meydana gelmesi, yordanan değişkenin koşullu dağılımının farklı noktalarındaki yordayıcı değişkenlerdeki değişikliklere farklı tepki vermesi olarak açıklanabilir (Keskin, 2012).

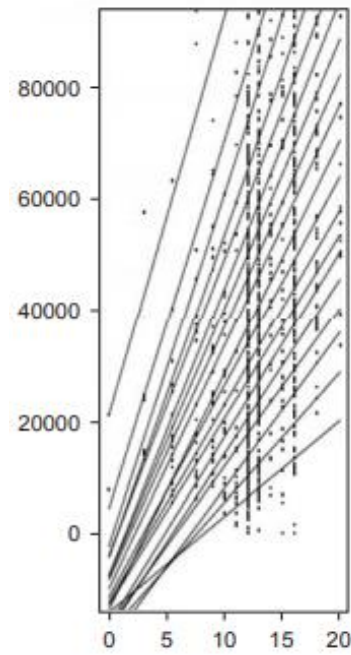
Yukarıdaki eşitliklerde de belirtildiği üzere, sadece bir eşitlik ile ifade edilen ve tek bir koşullu ortalamaya sahip olan, klasik doğrusal regresyon modelinin eşitliğinden farklı olarak koşullu regresyon modeli sayısız koşullu kantile sahiptir. Kantil regresyonun bir avantajı da veri setini istenildiği zaman dağılımın her noktası hakkında bilgi verecek şekilde belli kantillere ayırarak uç gözlemlerin etkisini ortadan kaldırmak ve bağımlı değişkeni ortalama olarak açıklamak yerine alt, üst ve orta gibi 19 farklı grup kadar açıklayabilmesidir (Van, 2013). Kantiller birbirine eşit uzaklıkta olmak zorunda olmamasına rağmen uygulamada eşit aralık kantillerin seçilmesi verilerin yorumlanmasını kolaylaştırmaktadır. Ayrıca her bir kantile bağlı olarak regresyon katsayılarını ayrı ayrı belirlenmektedir (Chen & Wei, 2005). Örneğin, eğer kantil regresyon modeli ile 19 kantil belirtilmek istenirse, 19 eşitlik için 19 ayrı regresyon katsayı verir. Doğrusal regresyon ve kantil regresyon için regresyon doğruları Şekil 8’de gösterilmiştir.

Şekil 8

Tek Bir Doğrusal Regresyon Doğrusu ve 19 Ayrı Kantil Regresyon Doğrusu (Hao & Naiman, 2007).



(a) Lineer Regresyon Doğrusu



(b) Kantil Regresyon Doğruları

Şekil 8'de (a) paneli incelendiğinde doğrusal regresyon için çizilmiş olan, tek regresyon doğrusu ortalama değişimi belirtir. Kantil regresyon için çizilen (b) paneli ise 19 kantil regresyon doğrusunu göstermektedir (Hao ve Naiman, 2007).

(4) numaralı eşitlik ile belirtilen $Y_i = \beta + \varepsilon_i$ regresyon denklemiyle ifade edilen modelde p . örnek kantili,

$$\min \frac{1}{n} \left\{ \sum_{i: y_i \geq \beta} p |y_i - \beta| + \sum_{i: y_i < \beta} (1 - p) |y_i - \beta| \right\} \quad (20)$$

ifadesinin minimizasyonu ile elde edilir. Bu ifadeyi doğrusal regresyon modeli

$$Y_i = X_i' \beta + \varepsilon_i \quad (21)$$

için genelleştirdiği durumda; X_i' , bağımsız değişken vektör olduğu durum için p . kantil regresyonu,

$$\min \frac{1}{n} \left\{ \sum_{i: y_i \geq x_i' \beta} p |y_i - x_i' \beta| + \sum_{i: y_i < x_i' \beta} (1 - p) |y_i - x_i' \beta| \right\} \quad (22)$$

ifadesinin minimizasyonu ile yordanabilir ve Y 'nin p . kantili olarak da adlandırılabilir (Saçaklı, 2005). Böylece p . kantil regresyon doğrusunun parametreleri;

$$\hat{\beta}_{(p)} = \min_{\beta} \frac{1}{n} \left\{ \sum_{i: y_i \geq x_i' \beta} p |y_i - x_i' \beta| + \sum_{i: y_i < x_i' \beta} (1 - p) |y_i - x_i' \beta| \right\} \quad (23)$$

Eşitliğin en küçüklenmesiyle elde edilmiş olur (Chiang & Li, 2012).

Kantil regresyon, veriler, sıradan en küçük kareler regresyonunun anahtarı olan hataların normalliği varsayımında başarısız olduğunda geleneksel analizi tamamlayabilen bir araçtır. Yordanan/kestirilen regresyon modelleriyle, nedenselliğin tek yönlü olmasına rağmen bağımlı değişken üzerindeki değişikliklerin ne olabileceği açıklanmaya çalışılır. Doğrusal regresyon analizinin bazı varsayımlara sahiptir ve bu varsayımların karşılanmasının mümkün olmadığı durumlarda değişkenler arasındaki ilişkinin varlığı ve gücünün yorumlanması yanlış sonuçlar vermektedir. Ayrıca araştırma sürecinde kuramsal temele dayanarak kurulan uygun regresyon modeli tanımlansa bile değişkenler arasındaki

ilişkinin nasıl daha iyi açıklanabilir sorusu araştırmacıları alternatif regresyon modellerinin gerekliliğini düşündürmüştür.

Bir alternatif olarak, Koenker ve Bassett (1978) tarafından regresyon modelinin bir genellemesi olarak tanımlanan kantil (dilim) regresyon (quantile regression/QR) bağımlı değişkenin koşullu kantillerini tahmin etmeye odaklanarak, bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi modellemek için kullanılan istatistiksel bir yöntemdir. Bağımlı değişkenin koşullu ortalamasını tahmin eden sıradan en küçük kareler (EKK) regresyonundan farklı olarak, medyan, kartiller veya istenen diğer kantil dilimleri gibi farklı kantil dilimlerinin modellenmesine izin vererek bağımlı değişkenin koşullu dağılımı üzerinden verinin bütününe daha kapsamlı bir resmi sağlar. Kantil regresyon, bağımlı değişkenin gözlenen değerleri ile yordanan kantilleri arasındaki mutlak veya karesel farklara dayanan, seçilen kantil değerinin kısıtlamalarına tabi olan bir kayıp fonksiyonunun en az değere indirilmesini içerir. Dolayısıyla kantil regresyon yönteminin belirlenen her kantil için ayrı ayrı model tahminlemesi yapıyor olması araştırmacılara araştırmanın boyutunu genişletme imkanı sunmaktadır. Böylece doğrusal regresyon modellerinde genel olarak kullanılan geleneksel en küçük kareler kestirimi (EKK) yöntemiyle yordanan tek parametre değeri yerine her bir kantil için model parametrelerini sunması ile araştırmacıya daha geniş bir perspektif sağlamaktadır. Ayrıca en küçük kareler kestirimi (EKK) yönteminin uç değerlere karşı daha hassas olması nedeniyle bu yöntem verilerin normalden sapma gösterdiği durumlarda daha az verimli hale gelmektedir. Bu durum da değişkenler arasındaki ilişkinin yanlış yorumlanmasına neden olmaktadır. Özetle (doğrusal) regresyonun amacı, ortak değişkenlerin bir fonksiyonu olarak bir yanıt değişkeninin beklenen değerlerini tahmin etmek iken; kantil regresyon sıralama istatistiklerini regresyon ayarına genelleştirerek çok daha geniş bir görüş sunar. Böylece yalnızca medyanlar değil, aynı zamanda daha düşük ve daha yüksek kantil değerlerindeki (örneğin 0.10; 0.25; 0.75; 0.90 kantil değerlerindeki) regresyon analizine ait

parametre kestirimlerinin deęerleri elde edilerek, alıřılacak yanıtın (baęımlı deęiřkenin) hem konumu hem de yayılması saęlanır (Schnabel & Eilers, 2009).

Birok alıřmanın amacı daęılımın alt veya st grubunda yer alan verilere ynelik gerekleřtirmek ve tm bu avantajlar dřnldęnde kantil regresyon, modern regresyon analizinin yanı sıra finans, tıp ve evre gibi sayısız pratik uygulamalarda cazip bir istatistik aracı haline gelmiřtir. Bu yntemin saęlık, iktisat, ekoloji vb. alanlarında sıklıkla kullanılmasının nedeni daęılımın u noktalarında yer alan verilerden elde edilecek bilgilerin olduka nemli olmasıdır. rneęin saęlık alanında HIV arařtırmasında, bireylerin daha yksek risk altında olduęu CD4 sayı daęılımının sol kuyruęunda tedavinin etkisi daha nemlidir. Daha dřk risk altındaki hastalar iin bir tedavi faydalı olmasına raęmen ancak sol kuyruktakiler iin zararlı (komplikasyonlar veya yan etkiler nedeniyle) olabilmektedir (Bruneau ve dięerleri, 2017). Benzer řekilde eęitim sonularındaki eřiřsizlięin belirleyicilerini ve mekanizmalarını arařtırmak iin gmenlik, cinsiyet vb. deęiřkenlerinin de etkileri olduka nemli olmaktadır (Javier ve dięerleri, 2011).

Kantil deęeri, deęiřkenin daęılımında yer alan ve daęılımı kendisinden byk olanlar ve kendisinden kk olanlar diye ikiye blen herhangi bir deęerdir. Bir seriye ait gzlem deęerleri kkten byęe doęru sıralandıęında serinin toplam frekansını 2, 4, 10 ve 100 eřiř paraya ayıran deęerler genel olarak kantil (eřiřblen) olarak adlandırılmaktadır. Verideki deęerlerin daęılımını iki eřiř paraya blmek iin hesaplanan deęerlere medyan, drt eřiř paraya blmek iin hesaplanan deęerlere kartil, on eřiř paraya blmek iin hesaplanan deęerlere desil, yz eřiř paraya blmek iin hesaplanan deęerlere santil adı verilmektedir (Turanlı ve Griř, 2005).

Kantil Regresyonun (KR) en belirgin zelliklerinden biri orta deęiřkenlerin yalnızca merkezdeki deęil aynı zamanda sonu daęılımı kuyruklarındaki etkisini tanımlayan dilimi zel etkileri tahmin edebilme kabiliyetidir. Kořullu ortalama regresyon yoluyla elde edilen ortalama etki gibi merkezi etkiler bir orta deęiřkenin etkisinin ilgili ekici zet istatistiklerini saęlarken, deęiřken hem merkezi hem de kuyruk dilimleri aynı ynde etkilemedike tam

dağılım etkisini tanımlamak konusunda başarısızdırlar. Ayrıca, birçok durumda ilgi dağılımının merkezi dışında orta değişkenleri etkisine odaklanmaktadır. Örneğin, bir iş eğitim programının etkinliği çalışmasında, eğitimin kazanç dağılımının alt kuyruğundaki etkisinin kamu politikası için eğitimin dağılımının ortalama üzerindeki etkisinden daha fazla ilgi görmesi çok daha olasıdır (Chernozhukov & Hansen, 2008).

Eş varyanslılık gibi ortalama regresyon gereksinimlerinin ihlal edildiği veya ilgilenilen aralığın koşullu dağılımın dış bölgelerinde olduğu durumlarda, kantil regresyon bağımlılıkları klasik yöntemlerden daha doğru bir şekilde açıklayabilir (Waldman, 2018).

EKK yöntemiyle yapılan regresyon analizinde, bağımlı değişkenin bağımsız değişken(ler) arasındaki ilişki dağılımının sadece ortalaması ile incelenirken dağılımın tamamı hakkında bilgi sağlamaz. Bunun üstesinden gelmek için alan yazında yeni regresyon yöntemleri geliştirilmiştir. Regresyon eğrisi, dağılımın farklı yüzdelik noktalarına göre ayrı ayrı hesaplanıp, klasik regresyondan daha kompleks ve detaylı bir tablo elde edilebilir (Mosteller & Tukey, 1977). Regresyon analizinin yanlı yorumlanmasına neden olabilecek bu sorunu çözmek üzere Koenker ve Bassett (1978) kantil regresyon ismini verdikleri yeni bir yöntem geliştirmiştir. Bu yöntem lineer regresyona ait bağımlı değişkenlerin dağılımının, örneklemin farklı yüzdelik dilimlerinde tahminlenmesine izin vermektedir. Başka bir ifade ile tek bir orta noktanın hesaplanması yerine dağılımdaki birçok yüzdelik dilimde orta nokta ayrı ayrı hesaplanmaktadır.

Bağımlı değişkeninin belirgin asimetrisi ve artıkların normal olmayan dağılımı nedeniyle daha sağlam tahmin yöntemlerinin varlığının araştırılmasıyla kantil regresyon hem sağlamlık özellikleri hem de tepki değişkeninin daha kapsamlı analiz olanağı nedeniyle iyi bir seçenek olarak ortaya çıkmıştır. Bu tür regresyonun temel özellikleri dağılım etkilerinin tam görünümü ile eksiksiz ortalama regresyon analizinin yapılabilmesi, uç değerler ve asimetriye karşı sağlamlılık, dağılımdan bağımsızlık, değişen varyansın doğal işlemlerine sahip olma ve monoton dönüşümlerle eş değişkenli olması olarak sıralanabilir (Piombo, 2013).

Kısacası kantil regresyon yöntemi Koenker ve Bassett (1978) tarafından önerildiğinden beri iktisat, ekonomi vb. alanlarda teorik ve uygulama alanlarında oldukça ilgi görmüş ve yaygın kullanılmaktadır. Bağımlı değişkenin dağılımının kantil dilimleri ile bağımsız değişkenler arasındaki ilişkiyi detaylı olarak ortaya koymak amacıyla uygulanan bu yöntem asimetrik mutlak artık karelerin minimizasyonuna dayandırılmaktadır ve kantil dilimleri bağımsız değişkenin dağılımı hakkında tümüyle bilgi edinilmesini sağladığından, şartlı ortalama esasına dayanan En Küçük Kareler (EKK) regresyon yöntemine önemli bir alternatif olmaktadır. Kantil regresyon yöntemi yatay kesit verilerinin yanı sıra boylamsal (dikey kesit) veri analizlerinde de kullanılabilir.

Çalışmanın devamında çok düzeyli kantil regresyon ile ilgili açıklayıcı bilgilere yer verilmeden önce çok düzeyli boylamsal verilerin analizi hakkında bilgiler sunulmaktadır.

Çok Düzeyli Boylamsal Verilerde Analiz Yöntemleri

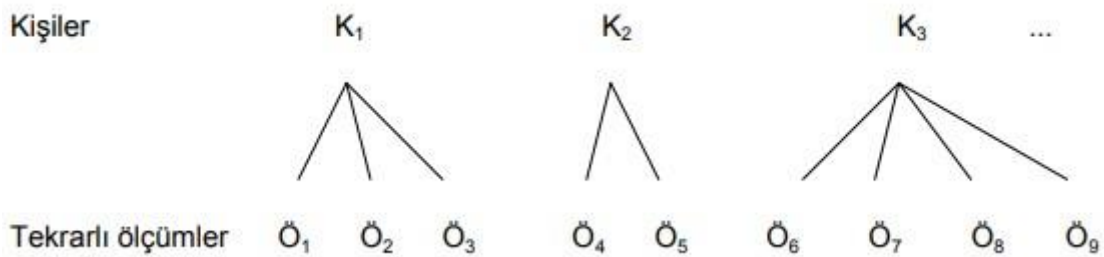
Çoklu regresyon, tipik olarak basit rasgele örneklemeden veya büyük ölçekli tarama araştırmalarında bulunan karmaşık, çok aşamalı örnekleme stratejilerinden bireysel düzeyde veriler olarak elde edilen kesitsel verileri incelemek için kullanılan popüler bir analitik seçimidir (Heck & Thomas, 2020). Klasik basit doğrusal regresyon analizi, gözlemlerin bağımsız ve özdeş olması varsayımını gerektirir. Ancak bu varsayım hiyerarşik veriler ve benzeri durumlarda karşılanmayabilir (Osborne, 2000; Raudenbush & Bryk, 2002). Anket araştırmalarında, örneğin verilerin seçkisiz (random) değil de coğrafi bölgelerden küme örnekleme ile elde edildiği durumlarda bireysel gözlemler arasındaki bağımsızlık ilkesi ihlal edilir. Genel olarak, aynı gruptan olan katılımcılar birbirlerine farklı gruplardan katılımcılara göre daha benzer (olumlu bağıntılı) olacaklardır ve bu yüzden bağımsız gözlemler varsayımı bozulacaktır. Eğer bu varsayım sağlanmazsa, regresyon katsayılarının standart hata tahminleri, genellikle daha küçük hesaplanıp Tip 1 hatanın artmasına yol açacaktır (Piombo, 2013). Tipik bir tek seviyeli regresyon modelinde, kesişmeyi tanımlayan katsayıların ve eğimlerin, örnek verilerden yordanan evrendeki sabit

değerleri temsil ettiğini varsayabiliriz (Yani, her bir yordayıcının bağımlı değişken üzerindeki etkisini açıklayan bir kesişim katsayısı ve bir eğim katsayısı olduğu kastedilmektedir). Ayrıca örneklemdaki bireylerin bağımsız olduğunu varsayabiliriz; yani, verilerin evrenin rasgele bir örneğinden seçildiği varsayımı üzerine ilerlenmektedir. Çok düzeyli modeller bağımlılık konusuna değinilmesi için uygundur ama aynı zamanda kümeler içindeki ve arasındaki ilişkileri anlamamıza ve her bir hiyerarşi seviyesindeki değişkeni analiz etmemiz ve açıklayabilmemize olanak sağlar. Aslında, çok düzeyli modelleri gruplara göre değişiklik gösteren katsayıları olan (kesişmeler ve eğriler) bir regresyon olarak görebiliriz. Boylamsal veri analizinde de klasik regresyon modellerinde olduğu gibi, bir bağımlı değişken ile bir ya da birden fazla bağımsız değişken ile hata terimi bulunmaktadır. Değişkenler hem birimlere hem de zamana göre değişim gösterebileceği için farklı indislerle modellenmektedirler. Son yıllarda iç içe geçmiş veri yapılarıyla yapılan araştırmalarda tek düzeyli analitik tekniklerin uygulanmasına ilişkin kavramsal ve metodolojik kaygılar, çok düzeyli modellemenin gelişmesine yol açmıştır (Heck & Thomas, 2020). Hiyerarşik/kümelenmiş verilerle yürütülen araştırmalardaki bu tür sorunlar, çok düzeyli modelleme (multilevel modeling) tekniklerinin gelişmesine yol açmıştır (Hox, 2010). Verilerin analizinde hiyerarşik yapıyı dikkate alan bu yöntemler literatürde farklı alanlarda farklı isimlendirmelerle anılmaktadır; çok düzeyli regresyon modelleri (multilevel regression models), hiyerarşik lineer modeller (hierarchical linear models - HLM), karmaşık ve seçkisiz etkiler modelleri (mixed and random effects models), seçkisiz katsayılar modelleri (random coefficients models) ve çok düzeyli kovaryans yapıları modelleri (multilevel covariance structure models) (Heck, 2001; Piombo, 2013). Genellikle “büyüme eğrisi (growth curve)” veya “büyüme yörüngesi (growth trajectory)” modelleri olarak adlandırılan çok düzeyli boylamsal modeller, bağımlı değişkendeki (örneğin matematik başarı puanları) zaman içindeki değişimi inceler. Literatürde farklı isimlendirmeler yer alsa da her tür araştırma sorusunu yanıtlamak, farklı bir istatistiksel yaklaşım gerektirir. Bireysel büyüme modellemesi (Rogosa ve diğerleri, 1982; Willett, 1988), çok düzeyli modelleme (Goldstein, 1995), hiyerarşik doğrusal modelleme

(Raudenbush & Bryk, 2002), rasgele katsayılı regresyon olarak bilinen yöntemleri kullanarak değişimle ilgili soruları ele alınırken kullanılan çalışmalardan olup (Hedeker ve diğerleri, 1994) ve karma modelleme (Pinheiro & Bates, 2000) hayatta kalma analizi (Cox & Oakes, 1984), olay geçmişi analizi (Allison, 1984; Tuma & Hannan, 1984), başarısızlık zamanı analizi (Kalbfleish & Prentice, 1980) olarak bilinen yöntemleri kullanarak olayın meydana gelmesiyle ilgili sorular ele alınabilmektedir. Çalışmanın diğer bölümlerinde Goldstein (1995)'in de literatürde kullandığı şekilde boylamsal verilerde “çok düzeyli modelleme” isimlendirilmesi kullanılmıştır.

Şekil 9

Tekrarlı Ölçümlerin Bireylerin İçine Yuvalanma Diyagramı



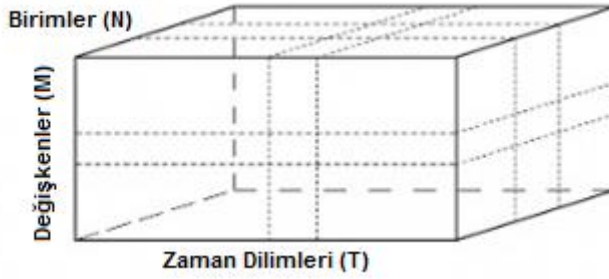
Çok düzeyli verilere bir okuldaki öğrencilerin sınıflarda örnek gösterilebilir. Boylamsal veriler de çok düzeyli iç içe geçmiş verilerdir. Boylamsal veriler iç içe geçmiş yani kümelenmiştir. Kümelenme, tek bir bireyin farklı zaman dilimlerinde tekrarlanan ölçümlerinden oluşmaktadır. Diğer bir deyişle boylamsal veriler (dikey kesit verileri) çok düzeyli verilerin özel bir halidir (Fitzmaurice ve diğerleri, 2004). Bir kümenin içerisindeki elemanlar korelasyona sahiptirler ve bu korelasyonun analizde göz ardı edilmemesi gerekmektedir. Tekrarlı ölçümlerin kişilerin içine yuvalanmış durumları Şekil 9'daki gibi olmaktadır.

Boylamsal çalışmalarda, veriler toplandıktan sonra bu veriler bir veri kutusu (data box) oluşturmaktadır. Birimlerin birinci eksen, değişkenlerin ikinci eksen ve zamanın

üçüncü eksen olarak belirlendiği bu kutu kavramı ilk defa 1946 yılında Cattell tarafından kavramsallaştırılmış ve adına da kovaryans grafiği denilmiştir (Bijleveld & Kamp, 1998).

Şekil 10

Veri Kutusu



Boylamsal verilerin birim olarak adlandırılan birçok yapının bir araya gelmesi sonucu oluştuğu ve bu yapıdaki her bir birimin kendine özgü özelliklere sahip olduğu göz ardı edilmemelidir. Birimlerin kendine özgü özelliklerini yansıtan değişken/değişkenlere “birim etki” adı verilir. Birim etki, birimden birime değişen zamana göre ise sabit olan bir değişkendir. Birim olarak bireylerden bahsediliyorsa kişilik özellikleri, yetenek, cinsiyet, ırk vb. değişkenler birim etkiye örnek olarak gösterilebilir. Boylamsal verilerde birim etkinin varlığının yanı sıra zaman etkisi de yer almaktadır. Her bir zaman diliminin kendine özgü özellikleri vardır. Bu zaman özelliklerini yansıtan değişken/değişkenler “zaman etkisi” olarak adlandırılmaktadır (Yerdelen Tatoğlu,2018).

Şekil 10’da gösterilen veri kutusundaki her hücre, belirli bir isim ve zaman noktası için bir değişkenin değerini temsil etmektedir. Bu durum, her bir değişken için zaman içindeki değişim modellerinin görselleştirilmesine ve ayrıca eksik/aykırı verilerin değerlerinin fark edilmesine/tanımlanmasına olanak sağlar. Öte yandan, ilerleyen paragraflarda detaylıca açıklanacak olan Düzey 1 ve Düzey 2 denklemleri ile değişkenler arasındaki ilişkileri zamana bağlı olarak modellemek matematiksel olarak fayda sağlamaktadır. Düzey 1 denklemleri, her birim için her değişkendeki zaman içindeki değişimi modellemek için kullanılır. Bu denklemler genellikle, değişkenin önceki değerini

ve ilgili yordayıcıları içerir. Düzey 2 denklemleri ise birimler arasındaki değişim oranındaki bireysel farklılıkları modellemek için kullanılır. Bu denklemler tipik olarak değişim oranını etkileyebilecek bireysel düzeyde yordayıcıları içerir. Özetle, Şekil 3'te tanılanan veri kutusu ile gözlemlenen veriler kontrol edilebilirken, Düzey 1 ve Düzey 2 eşitlikleri ile gözlemlenen verileri modelleme imkanımız olmaktadır. Veriler Düzey 1 eşitlikleri ile zaman içindeki birim içi değişiklikleri yakalarken, Düzey 2 eşitlikleri bu değişikliklerdeki birim içi farkları yakalayarak modellenmesine katkı sunar. Veri kutusu ve Düzey 1 ile Düzey 2 eşitlikleriyle birlikte, değişkenlerin zaman içinde birbirleriyle nasıl ilişkili olduğunu ve bu ilişkilerin farklı birimler arasında nasıl değiştiğini anlamak için bir çerçeve sağlarlar.

Araştırmacılar tarafından boylamsal veri yapısında, basit bir regresyon denklemi ile açıklanamayan araştırma sorularına yanıt aramak için model geliştirme adımları izlenmektedir. Boylamsal verilerdeki değişim süreçlerini temsil etmek için ihtiyacımız olan istatistiksel modele ait araştırma soruları da Düzey 1 ve Düzey 2'ye ait olmalıdır. Bunun için öncelikle veri yapısının düzeyleri (level of data) aşağıda kısaca tanıtılacaktır (West ve diğerleri, 2015).

Düzey 1; kümelenmiş bir veri kümesinde verinin en ayrıntılı düzeyde gözlemleri belirtir. Düzey 1, kümelenmiş veri setlerinde çalışmadaki her bir analiz birimini, boylamsal verilerde ya da tekrarlı ölçümlerde ise aynı analiz ünitesindeki tekrarlanan ölçümleri temsil eder. Sürekli bağımlı değişken her zaman verinin birinci düzeyinde ölçülür. Mümkün olan en basit hiyerarşik doğrusal model, rasgele etkilere sahip tek yönlü bir ANOVA'ya eşdeğerdir. Bu durumda, Düzey 1 modelindeki β_{1j} , tüm j değerleri için sifıra eşitlendiğinde Düzey 1 eşitliği;

$$Y_{ij} = \beta_{1j} + r_{ij} \quad (24)$$

eşitliği ile matematiksel olarak ifade edilebilir. Veri yapısına göre oluşturulan Düzey 1, bir üst düzey içine yuvalanan birimlerin kendi aralarındaki değişimin belirlendiği modeldir ve hücre içi model (within-cell model) olarak tanımlanır (Raudenbush & Bryk, 2002).

Düzyey 2; hiyerarşinin bir sonraki seviyesini temsil eder. Düzyey 2, kümedenmiş veri setlerinde gözlemlerin birim kümelerini, tekrarlı ölçüm ya da boylamsal verilerde ise analiz birimlerini temsil eder. Düzyey 1 hatası olan , r_{ij} 'nin, ortalamasının sıfır ve Düzyey 1 varyansının sabit olduđu varsayılmakta olup bu model, her bir Düzyey 1 birimin içindeki bağımlı değışkeni, yalnızca Düzyey 2 parametresi olan kesişme noktası (β_{0j}) ile yordamaktadır. Bu durumda, β_{0j} , j. birim için sadece ortalama sonuca eşittir (Yani $\beta_{0j} = \mu_{y_j}$). Rasgele etkilere sahip tek yönlü ANOVA için Düzyey 2 modeli, γ_{01} 'in sıfıra ayarlandığı durumda;

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (25)$$

eşitliği ile matematiksel olarak ifade edilebilir. Düzyey 2 hücreler arasındaki değışim bir üst düzeyde yer alan grup değışkenlerine dayandırılır ve hücreler arası model olarak (between-cell model) tanımlanır (Raudenbush & Bryk, 2002). Özetle oluşturulan modele göre Düzyey 1 alt modeli “Zaman içindeki bireysel değışim nasıl tanımlanır?” ve Düzyey 2 alt modeli “Bireyler arasındaki değışimin kaynağı nasıl tanımlanır?” sorularına yanıt aranırken kullanılmaktadır. Her iki düzeyden elde edilen bu yanıtlar birlikte ele alınarak çok düzeyli bir istatistiksel model olarak bilinen yapıyı oluşturmaktadır (Bryk & Raudenbush, 1987; Rogosa & Willett, 1985). Çalışmanın ilerleyen paragraflarında Düzyey 1 ve Düzyey 2 eşitliklerine göre çok düzeyli verilerin matematiksel modellenmesi hakkında detaylı bilgi verilmiştir.

Bireye Özgü Etkilerin İncelenmesi

Boylamsal verilerin çok düzeyli tekniklerle modellenmesinden önce araştırmacının bireye özgü etkiler hakkında kuramsal çerçeveyi oluşturması gerekmektedir. İstatistikte bireye özgü etki hakkında tesadüfi (rasgele) etki varsayımı ve sabit etki varsayımı olmak üzere iki ortak varsayım yapılabilir. Bu kavramlar farklı alanlarda farklı anlamlara gelebilmektedir. Örneğin istatistik ve biyoistatistikte, sabit etkiler gözlenen değışkenlerin yordayıcı değışkene etkileri olarak, sosyal bilimler ve ekonometride ise kümelerin tüm

ölçümleri arasındaki yoğunluğu temsil etmesi; tesadüfi etkiler ise gözlenemeyen değişkenlerin etkileri olarak yorumlanmaktadır. Çalışmamızın devamında öncelikle farklı alanlarda sabit etkiler, rasgele etkiler olarak gruplandırılan kavramların içeriğine ardından da eğitim bilimleri alanında son zamanlarda çok düzeyli boylamsal modeller ya da büyüme modelleri olarak gruplandırılan kavramların denklemlerle modellenmelerine yer verilmiştir.

Tekrarlanan ölçümlerden elde edilen veriler genelde geleneksel bir analiz yöntemi olan ANOVA kullanılarak analiz edilmektedir. Ancak boylamsal verinin ANOVA ile analiz edilmesi; küresellik varsayımı, desen etkisi (hiyerarşik örnekleme) ve tüm desen ve veri seti için gereksinimler olmak üzere ortaya çıkardığı bilinen üç problem bulunmaktadır. Çok düzeyli modeller ise varyans ve kovaryans bileşenlerinin açıkça yordanmasını sağlamakta ve küresellik varsayımı gerektirmemektedir. Ayrıca çok düzeyli modelleme örnekleme hiyerarşisini hesaba katarak eksik verileri analiz etme yeteneğine de sahiptir (Quené & Van den Bergh, 2004). ANOVA literatürünün kalıcı bir fikir ayrılığı noktası, tüm etki gruplarına rasgele değişkenler olarak davranılmasını savunduğumuz için ele alınması gereken konu sabit ve rasgele etkilerin uygun kullanımınıdır. Bu etkiler için bazı araştırmacılar farklı zamanlarda farklı tanımlar yapmışlardır. Eisenhart (1947) varyans bileşenlerini tahmin etmede sabit ve tesadüfi etkileri birbirinden ayırır. Farklı alanlarda farklı kaynaklarda sabit ve tesadüfi etkilerin tanımlamaları da farklılık göstermektedir. Örneğin Kreft ve Leeue (1998) sabit etkileri birey içi, rasgele etkileri ise bireyler arası farklılık göstermek olarak tanımlamışlardır. Bir boylamsal büyüme çalışmasında rasgele kesişimin a_i ve sabit eğimin B olarak tanımlandığı modelde, farklı i bireyleri için birbirine paralel çizgilere veya $y_{it}=a_i+Bt$ denkleminin karşılıklı geleceğini belirterek sabit ve rasgele katsayıları matematiksel olarak birbirinden ayrı modellemişlerdir. Searle, Caselle & McCulloh (1992)'ye göre araştırmada etkilerin kendisi ilgi odağında ise etkiler sabit, evren ilgi odağında ise etkilere rasgele (rassal, tesadüfi) etki denir. İstatistiklerde, varyans bileşenleri modeli olarak da adlandırılan rasgele etkiler modeli, model parametrelerinin rasgele değişkenler olduğu istatistiksel bir modeldir. Analiz edilen verilerin farklılıkları bu

hierarchy ile ilgili olan farklı evrenlerin hiyerarşisinden alındığını varsayan bir tür hiyerarşik doğrusal modeldir. Ekonometri biliminde, sabit etkiler varsayılmadığında yani bireysel etkilerin varlığına izin verildiğinde hiyerarşik verilerinin analizinde rasgele etkiler modelleri kullanılır. Rasgele etkiler varsayımı geçerliyse, rasgele etkiler modeli sabit etkiler modelinden daha etkilidir. Ancak, bu varsayım geçerli değilse, rasgele etkiler modeli tutarlı değildir.

Sabit ve Rasgele Etkiler. Sabit etki, tüm gözlemlerde aynı olan bir değişkenin etkisini ölçerken, rasgele etki, her bir grup veya alt grup için farklı olan bir değişkenin etkisini ölçer. Rasgele etkiler, farklı gruplar veya alt gruplar arasındaki değişkenliği modellemeye yardımcı olur.

Boylamsal çalışmalarda ise veri modellerinde yer alan parametrelerin farklı birim ve/veya farklı zamanlarda değer almasına göre sabit etkiler ve tesadüfi etkiler varsayımları yapılmaktadır. Sabit etki varsayımı, bireye özgü etkinin bağımsız değişkenlerle ilişkili olması durumu iken rasgele (tesadüfi) etkiler varsayımı, bireysel gözlemlenemeyen heterojenliğin bağımsız değişkenlerle korelasyonlu olmamasıdır (Wooldridge, 2010). Sabit etki modeli, model parametrelerinin tamamının veya bir kısmının rasgele değişken olarak kabul edildiği tesadüfi etkiler modellerinin ve karışık modellerin aksine, model parametrelerinin sabit veya rasgele olmayan miktarlar olduğu istatistiksel bir modeldir. Yani boylamsal verilerde sabit etki modelleri, tüm zamanlarda tekrarlanan ölçümlerin yoğunluğunun, ölçülemeyen değişkenlerin kontrolüne izin vermesi üzerine dayanır (Greene, 2011). Genel olarak, veriler gözlemlenen birkaç faktöre göre gruplandırılabilir. Örneklemeden elde edilen gözlemler, her gruba için sabit veya rasgele etkiler olarak modellenir. Sabit etkiler modelinde her grup ortalaması, gruba özgü sabit bir miktardır. Aynı konu için boylamsal gözlemlerin bulunduğu boylamsal verilerde, sabit etkiler araştırmaya özgü araçları temsil eder. Boylamsal veri analizinde, sabit etkiler tahmincisi (iç tahminci olarak da bilinir) terimi, bu sabit etkileri (her birey için

bir zamanla deęişmeyen kesişme) içeren regresyon modelindeki katsayılar için bir yordayıcıyı belirtmek için kullanılır.

Rasgele faktör, incelenen düzey evreninden rasgele örneklendięi düşünölebilecek düzeylere sahip bir sınıflandırma deęişkenidir. Rasgele faktörün olası tüm seviyeleri veri setinde mevcut deęildir, ancak araştırmacının niyeti tüm düzey evreni hakkında çıkarımlarda bulunmaktadır. Hem kümelenmiş hem de tekrarlanan ölçümler/boylamsal veri kümelerinde Düzey 2 ve Düzey 3 birimlerini tanımlayan sınıflandırma deęişkenleri genellikle rasgele faktörler olarak kabul edilir. Bir analizde rasgele faktörler dikkate alınır, böylece bağımlı deęişkendeki rasgele faktörlerin düzeyleri arasındaki varyasyon deęerlendirilebilir ve veri analizinin sonuçları, rasgele faktörün düzeylerinin daha büyük bir evrenine genelleştirilebilir (West ve dięerleri, 2015).

Green ve Tukey (1960)'e göre eęer örneklem evrenin tamamını kapsıyorsa, ilgilenilen deęişken sabit, eęer örneklem evrenin ufak bir parçası ise ilgilenilen deęişken rassaldır. LaMotte (1983)'e göre, bir etkinin rasgele deęişkenin gerçekleşmiş deęeri olduęu varsayıyorsa bu etki rassal etki olarak adlandırılır. Snijder ve Bosker (1999)'a göre sabit etkiler en küçük kareler yöntemiyle (ya da daha genel olarak en çok olabilirlik yöntemiyle) ve rassal etkiler küçölme ile hesaplanır. Bu tanımlama çok düzeyli modellerin alan yazındaki standart tanımıdır. Tüm bu tanımlamalara raęmen sabit ve rasgele etkiler üzerine bir tartışmaya odaklanmanın yolu, verilerde deęişme olmadan bir dizi etkinin sabitten rasgele hale geldiğinde elde edilen çıkarımların nasıl deęiştiiğinin yorumlanmasıdır (Gelman, 2005). Rasgele etkiler modeli, sabit etkiler modelinin özel bir durumudur (Wooldridge, 2010). Daha çok bireye özgü özelliklerin etkisi olduęu ve genellikle bu özelliklerin bilinmeyen, gizli deęişkenler olduęu varsayılır (Diggle ve dięerleri, 2002; Fitzmaurice ve dięerleri, 2004; Laird & James, 1982; Gardiner ve dięerleri, 2009, Ramsey & Shaffer, 2002).

Boylamsal veri setlerinde bir deęişim modelinde zamanın bir yordayıcı olabileceğini düşönen araştırmacıların cevaplanması gereken sorular bulunmaktadır.

Öncelikle, zaman içinde ortalama bir deęişiklięin var olup olmadığına yanıt aranmalıdır. Yani baęımlı deęişkenin zaman içindeki gözlem sonuçlarını grafik olarak çizildiğinde, çizilen bu eğrilerin eğiminin varlığı (doęruların düzden farklı olup olmadığı) belirlenmelidir. Ortalamada zaman içinde herhangi bir deęişiklięi tanımlayacak olan model parametreleri, modeldeki zamanın araştırılan baęımlı deęişken için sabit etkileridir. Sabit ve rasgele şöyle bir yoruma sahiptirler: sabit etki terimi herkesin aynı etkiyi kazandığı anlamına gelir. Zamanın sabit etkileri, varyans modellerinin analizinin test etmek için yani zaman içindeki ortalama farklılıkların görülmesi amacıyla özel olarak tasarlandığı durumlardır. Dolayısıyla şimdiye kadar düşündüğümüz zamanın sabit etkileri olmasına rağmen ortalama deęişime ek olarak, yanıtlanması gereken bir dięer soru bireysel deęişim hakkında, her bireyde aynı deęişimin gözlenip gözlenmediğidir. Başka bir deyişle, zaman içindeki eğim için sabit etki her birey için açıklayıcı yeterlikte olacağı mı yoksa her bireyin zaman içindeki grafiklerindeki eğimi için bireysel rasgele etkilerinin de dâhil edilmesine ihtiyaç duyulup duyulmadığı belirlenmelidir. Rasgele etki terimi her bireyin kendi etkisini elde ettiği anlamına gelir. Ortalamalar için ortalamadaki deęişimi temsil eden zamanın sabit etkilerinin aksine, zamanın rasgele etkileri varyans için modelin bir parçası olarak deęişimdeki bireysel farklılıkları temsil eder. Yani, her bir birey için farklı bir sabit zaman eğimi tahmin etmek yerine, sabit zaman eğiminden bireysel sapmaların kişiler arasındaki varyansın yordanması daha uygun olmaktadır. Zamanın etkisi için rasgele eğimler, varyans analizinin (ANOVA) bir parçası değildir (Hoffman, 2015).

Kısacası sabit bir faktör kavramı, en yaygın olarak standart bir ANOVA veya ANCOVA modelinin ayarlanmasında kullanılır. Sabit faktör, araştırmacının araştırmaya ilgi duyan tüm seviyeleri (veya koşulları) dahil ettiği kategorik veya sınıflandırma deęişkeni olarak tanımlanmaktadır. Sabit faktörler, cinsiyet gibi niteliksel ortak deęişkenleri içerebilir ya da bölge veya katman gibi bir anket örnekleme tasarımının veya randomize bir öğrenme teknikleri araştırmasındaki öğretim yöntemi gibi bir çalışma tasarımının ima ettiği sınıflandırma deęişkenleri; veya yaş grubu gibi gözlemsel bir çalışmada sıralı

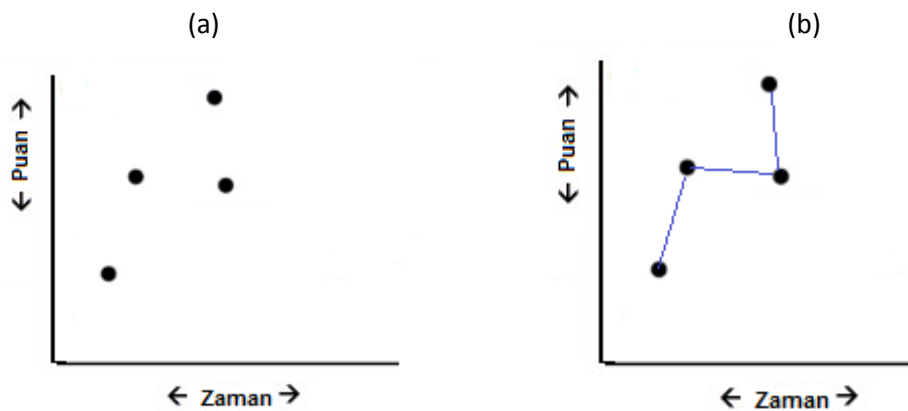
sınıflandırma değişkenleri olabilir. Sabit bir faktörün seviyeleri, belirli koşulları temsil edecek şekilde seçilir (West ve diğerleri, 2015). Özetle ANOVA'nın sınırlıkları arasında gösterilen; tekrarlanan ölçümlerdeki ortalamaların değişimlerine odaklanıyor olması ve birçok teorisyenin sadece ortalamadaki değişimden değil, bireysel değişimden de (özellikler bireyler arası değil birey içi değişim) endişe etmesinden, ANOVA/ MANOVA'nın bireylerdeki durum ve değişimin çağdaş teorilere iyi bir şekilde karşılık vermemesi vb. durumlar sebebiyle regresyon temelli modeller gelişmiştir.

Tüm bu durumlar göz önüne alındığında bu çalışmada değişimin sürekli yörüngesini elde etmek için aynı veri yapısına farklı bakış açısıyla yaklaşarak, her bireyin kendi içinde ve daha sonra bireyler karşısındaki değişimini kestirebileceğimiz veri için model inşa edilmiştir. Ancak öncelikle bir bireye ait gözlemlerden elde edilen ölçümlerin grafiğinin yorumlanmasının incelenmesinin faydalı olacağı düşünülmektedir.

Regresyon eğrileri kavramının ardından kesişim ve eğim terimlerinin daha iyi anlaşılması için, bireylerin zamana bağlı puanlarının değişimini ve zamanın etkisini inceleyen bir çalışmayı göz önüne alalım. Şekil 11'de verilen y ekseninin puan değeri (bağımlı değişken), x ekseninin zaman olduğu bir grafik verilmiştir.

Şekil 11

Bir Bireye Ait Dört Adet Ölçüm Değeri

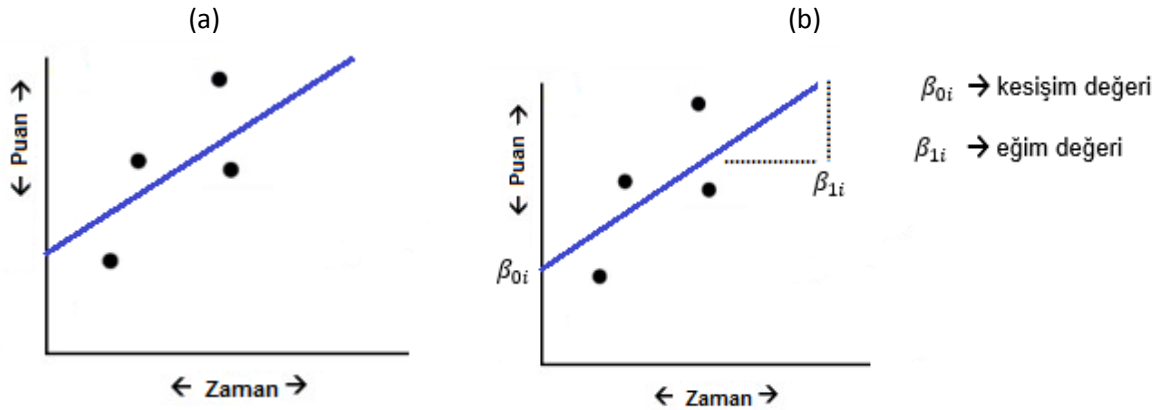


Şekil 11'de (a) paneli incelendiğinde bir bireye ait farklı zamanlardaki puan sonucunun ölçüm değeri gösterilmiştir. Bu dört puan değeri grafiğinde yapılan tekrarlı

ölçümlerde art arda zamanlardaki değişimler arasında (b) panelindeki gibi bağlantı kurulamamaktadır. Bu sebeple bireye ait tekrarlı ölçümlerden elde edilen bireysel eğriler kavramı ile ilgilenmekteyiz.

Şekil 12

Tekrarlı Ölçümlerden Elde Edilen Bireysel Eğriler (Yörüngeler)

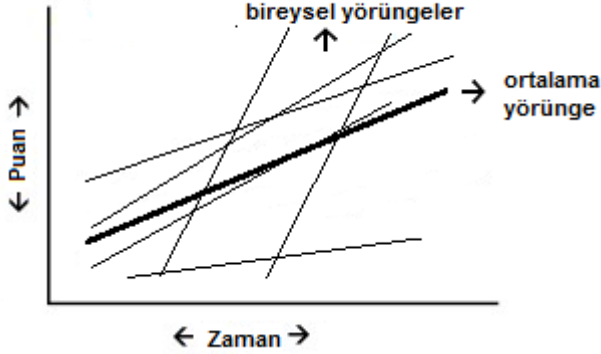


Bir bireye ait üç ya da daha fazla ölçüm değeri, değişimi hatadan daha doğru ayırabilir. Yani tekrarlı ölçümlerde en az üç ölçümün bulunması analizler de daha güvenilir sonuçlar elde etmemize yarayacaktır. Şekil 12 incelendiğinde (a) panelinde her bir birey için tekrarlanan ölçüm değerleri kullanılarak en uygun eğri kestirilmiştir. Büyüme eğrileri de denilen bireysel eğrilerin (b) panelinde gösterildiği gibi bir başlangıç noktası vardır. Yani zamanın sıfır noktasında bireyin regresyon eğrisinin y eksenindeki değeri olan β_{0i} (i bireyine ait t=0 zamanına ait) kesişim (intercept) ve regresyon eğrisindeki değişim oranı olan eğim değerinin β_{1i} değeri de eğim (slope) katsayısı olarak adlandırılır. Bir bireye ait kesişim ve eğim katsayı değerleri bireyler üzerindeki değişimi görmemize imkân vermektedir. Rasgele etkilerdeki değişim, dağılım ile tanımlanır. Genellikle harfi harfine her birey için katsayıları kestiremeyiz. Daha ziyade dağılımın parametrelerini elde ederiz ki bu da bize elimizdeki örneklemden tüm bireyler için çıkarım yapmaya izin vermektedir.

Her bireyin bireysel büyüme eğrilerinden oluşan Şekil 13'de gösterilen grafikte ise kalın çizgi ile gösterilen regresyon eğrisi örneklemin ortalaması üzerinden elde edilen örneklem eğrisidir.

Şekil 13

Bir örnekleme'deki tüm bireylere ait büyüme eğrileri

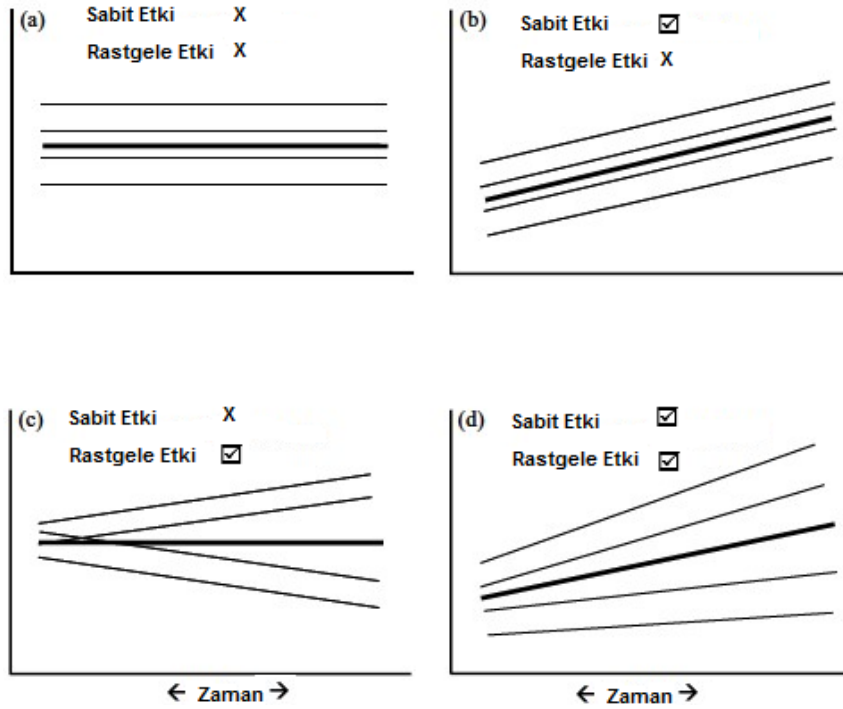


Değişimin sürekli yörüngesini elde etmek için aynı veri yapısına farklı bakış açısı ile yaklaşmalıyız. Her bireyin kendi içinde ve daha sonra bireyler karşısındaki değişimini kestirebileceğimiz veri modelleri inşa etmeliyiz. Bunun için öncelikli olarak zamanın rasgele ve sabit etkileri kavramlarının grafik gösterimleri üzerinde açıklamalarını da incelemeliyiz.

Her ne kadar terminolojik ifadeler farklı disiplin alanlarında farklılık gösterse de, zamanın rasgele etkisi ve zaman için rasgele eğim terimleri benzer anlamlar ifade eder yani; burada ana fikir, rasgele her bireyin kendi etkisini veya bir tahminleyicinin eğim değerini aldığı şeklindedir. Araştırmacıların boylamsal verilerde herhangi bir modelleme yapmadan önce zamanın sabit ve/veya rasgele etkilerinin hangi kombinasyonlarını gözlemleyebileceği, x ekseninin zaman, y ekseninin ise bağımlı (sonuç) değişkeni olduğu bir grafik üzerinde Şekil 6'daki gibi gösterilmiştir. Şekil 14'de gösterilen her paneldeki grafikteki ince çizgiler örnekleme'deki bireylerin zamana bağlı eğimlerini, kalın çizgi ise örneklemin zaman içindeki ortalama değişiminin sabit eğimini temsil etmektedir. Her panelde yer alan \square işareti önemli bir etkinin varlığını, \times işareti ise önemsiz bir etkinin varlığını gösterir (Hoffman, 2015).

Şekil 14

Zamanın Sabit ve Rasgele Etkilerinin Gösterimi (Hoffman, 2015)



İlk olarak, Şekil 14'deki sol üst panel (a), zamanın sabit ve rasgele olmayan etkilerini gösterir, yani bağımlı değişken üzerinde zamanın önemli bir etkisi yoktur. Zamanın sabit bir etkisini olmadığını ortalamada bir değişiklik olmamasından anlayabilmekteyiz. Şekil 14'deki (a) paneli incelendiğinde kalın çizgi ile gösterilen zaman içindeki ortalamanın düz olduğu görülmektedir. Paralel çizgiler arasındaki dikey mesafe kişiler arasında farklılık gösterdiği için her birey kendi rasgele kesişim noktasına sahip olsa da sabit eğim olduğu için, bireylerin zamanın etkisi için kendi rasgele eğimlerine ihtiyaçları yoktur, dolayısıyla bu durum göz önüne alınarak zamanın bağımlı değişken üzerinde rasgele bir etkisinin de olmadığı çıkarımını yapabiliriz.

İkinci olarak, Şekil 14'deki sağ üst panel (b) zamanın sabit bir etkisini gösterir, ancak zamanın rasgele bir etkisini göstermez. Zamanın sabit bir etkisi vardır, (b) paneli incelendiğinde kalın çizgi ile gösterilen zaman içindeki eğim düz değildir çünkü bağımlı değişken örneklem ortalamasında zamanla artış göstermektedir. Ancak zamanın rasgele bir etkisi yoktur, çünkü zaman içindeki ortalama artış için sabit eğim, her bireyin bireysel değişim oranını tanımlamak ve grafikte gösterimi için yeterlidir. (Hoffman, 2015).

Üçüncü olarak, Şekil 14'deki sol alt panel (c) belki de en ilginç senaryo olan; bağımlı değişken üzerinde zamanın sabit bir etkisi olmadığı, ancak rasgele bir etkisi olduğundaki durumu göstermektedir. Bağımlı değişken üzerinde zamanın sabit bir etkisi yoktur, çünkü (c) paneli de incelendiğinde ortalamanın zaman içinde bir değişimi yoktur, yani (a) panelinde olduğu gibi kalın çizgi ile gösterilen eğim düzdür. Ancak, sol alt panel (c) incelendiğinde her birey sabit eğimin ötesinde zamanın etkisi için kendi farklı eğimlerine ihtiyaç duymaktadır. Bu durumu zamanın sabit etkisi bireysel etkileri tanımlamadığından zamanın rasgele bir etkisi vardır şeklinde de yorumlayabiliriz. Sol alt panel (c) incelendiğinde bazı bireylerin zamanla artış bazılarında ise azalma göstermesi, genel olarak ortalama bir değişim örüntüsü ile sonuçlanmamaktadır. Böylece zamanın bağımlı değişken üzerinde sabit bir etkisinin olmadığı görülmektedir ancak zamanın sıfır etkisi bireylerdeki kimseyi tanımlayamayacağı için modellemelerde bilgilendirici olmamaktadır.

Son olarak, Şekil 14'deki sağ alt panel (d) zamanın bağımlı değişken üzerinde hem sabit hem de rasgele etkilerinin olduğu en yaygın modeli gösterir. Sağ alt panel (d) incelendiğinde kalın çizginin zaman içinde artış gösterdiği görülmektedir. Yani ortalama etki zamanla artmaktadır ve bu durum düz olmayan bir çizgi ile belli bir eğimde belirtilmiştir. Dolayısıyla zamanın bağımlı değişken üzerinde sabit bir etkisinin olduğu söylenebilir, fakat bu sabit etki ince çizgilerle gösterilen bireylerin eğimlerini yeterince tanımlamaz. Sağ alt panel (d) incelendiğinde bazı bireylerin ortalamadan daha hızlı bir oranda artış gösterirken, bazı bireylerin ortalamadan daha yavaş bir oranda arttığı dolayısıyla bireylerin zamanın sabit eğiminin ötesinde hala kendi bireysel farklı eğimlerine ihtiyaç duydukları görülmektedir. Dolayısıyla bu durumda zamanın bağımlı değişken üzerinde rasgele bir etkisinin de var olduğu söylenebilir.

Şekil 14 incelendiğinde ince çizgilerle gösterilen bireysel grafiklerin eğim ve kesişim noktalarının özelliklerinden de bahsetmek boyamsal modellerdeki kurulan regresyon denklemlerini anlamamız ve yorumlamamız açısından faydalı olacaktır. Şekil

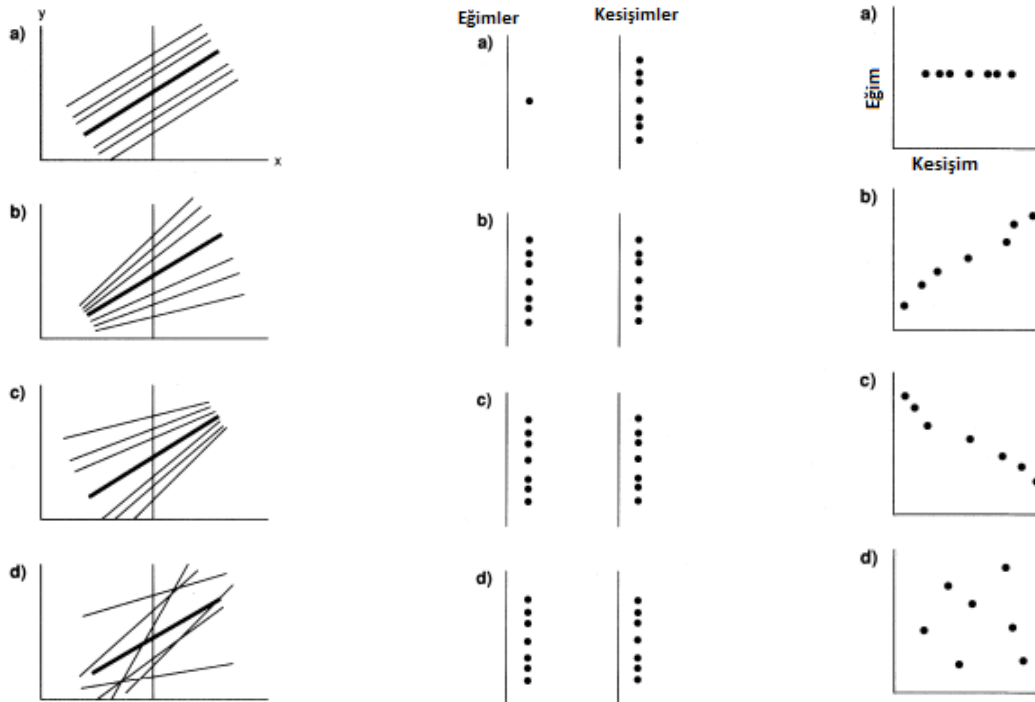
15'de x ve y eksenlerinde iki değişken arasındaki ilişkinin incelendiği regresyon denkleminin grafikleri gösterilmiştir. Bu grafiklerde ilk sütun regresyon denklemlerini, ikinci sütun kesişim ve eğim değerlerinin dağılımlarının nokta grafiklerini, üçüncü sütun ise kesişim ile eğim değerlerinin saçılım grafiklerini temsil etmektedir.

Bu grafiklerde ilk sütundaki grafikte yer alan kalın çizgiler örnekleme ait ortalama değeri ifade etmektedir. Zaman içindeki değişimin farklılığını görselleştirmek için yordayıcı değişkenlerin düzeyleri tarafından çizilen eğrilerin oluşturduğu beklenen grafiklere bir örnek olarak Şekil 15'deki ilk panel (a) incelendiğinde her bir regresyon denkleminin farklı kesişim (intercept) noktasına sahip olmasına rağmen tek bir eğim (slope) değerinin olduğu görülmektedir. Farklı başlangıç değerleriyle çizilen bu grafikte her bireyin birbirinden farklı kesişim katsayısına sahip olduğu ancak her bireydeki değişim oranının da aynı olduğu (sabit eğim) görülmektedir. Dolayısıyla bu grafiğe ait nokta grafiğinde tek eğim ancak birden fazla kesişim değeri bulunmaktadır.

Şekil 15

Farklı Regresyon Denklemlerinin Kesişim ve Eğim Değerlerinin İlişkinin Gösterimi

(Jones & Bullen, 1994)



Şekil 15'deki (b) ve (c) panelleri incelendiğinde çizilen grafikte her bireyin farklı başlangıç noktası ve farklı oranlarda eğim değerine sahip olduğu görülmektedir. Şekil 15'deki (d) paneli incelendiğinde bu grafiğe ait nokta grafiklerinin hem kesişim hem de eğim değerleri kendi içlerinde birbirinden farklı değerler aldığı görülmektedir.

Özetlemek gerekirse Şekil 15'de gösterilen zaman içindeki değişim modelleri çok basit olmasına rağmen, zamanın sabit etkilerinin amacı, zaman içindeki ortalama değişim modelini tanımlamaktır. Zamanın rasgele etkilerinin amacı, bu değişim modellerindeki bireysel farklılıkları tanımlamak ve aynı zamanda varyans ve kovaryansın zaman içinde gerektiğinde değişmesine izin vermektir (Hoffman, 2015). Bununla birlikte, aynı bireyin ölçümleri birbiriyle ilişkili olabilir. Bu nedenle, bu tür iç çok değişkenli bağımlılıklarla ilgilenen özel modellere ihtiyacımız var. Boylamsal analizlerde, zaman genellikle önemli bir model değişkenidir (Dinov, 2018). Çok düzeyli regresyon modellerinde sıklıkla amaç düzeyler arasındaki niceliksel değişimin incelenmesidir. Bu modellerde genellikle amaç bireyler arasındaki değişimi görmekten çok farklılıkları yordayıcı değişkenin bir fonksiyonu olarak açıklamaktır.

Sabit ve rasgele etkileri özetlemek için örnek vermek gerekirse; çok düzeyli modellemelerde miktar ve değişim açısından öğrenciler arasındaki rasgele farklılıklar ölçülebilir ve tahmin edilebilmektedir. Bu çerçevede çok düzeyli modelleri iki taraflı tanımlamak gerekmektedir. İlk kısım, ortalama başlangıç noktasını ve öğrenciler genelinde matematik puanlarındaki ortalama değişimi açıklamak için kullanılan sabit kesişim ve sabit regresyon eğimlerini içeren ortalama modelinden oluşur. Bu sabit etkiler, geleneksel doğrusal regresyon modelindeki etkilere benzer şekilde yorumlanır. Yani bütün öngörücüler 0'a eşit olduğunda sabit kesişim ortalama matematik puanını gösterir; aynı zamanda sürekli bir öngörücü için sabit regresyon eğimini, yordayıcıdaki bir birimlik artış başına matematik puanındaki ortalama değişimini ve kategorik bir öngörücü için sabit regresyon eğimi, iki grup arasındaki matematik puanındaki ortalama farkı temsil eder. Çok düzeyli modellemenin ikinci tarafı ise, tahmindeki rasgele hataların öğrenciler ve durumlar

arasında nasıl dağıtıldığını ve ilişkili olduğunu içeren varyans modelinden oluşur. Doğrusal bir regresyon modelinde, tek hata terimi, bir öğrencinin yordanan bağımlı değişken değerinden ve gözlemlenen sonuç değerinden sapmayı temsil eden artık değerlerdir ve yordayıcı değişkenler tarafından açıklanan veya azaltılan bu artıkların varyansıdır. Bununla birlikte, boylamsal verilerin hiyerarşik yapısı göz önüne alındığında, öğrenciler arasında rasgele etkiler olarak adlandırılan hatalar ve artıklar olarak adlandırılan öğrenci içi hatalar ile hem öğrenciler arasında hem de öğrencilerin içinde tahmin hataları meydana gelecektir. Bu hatalar ölçümlerdeki sapmaları temsil eder, bu nedenle, rasgele bir kesişme ortalama veya sabit kesişmeden öğrenciye özgü sapma olup, rasgele bir eğim ortalama veya sabit regresyon eğiminden öğrenciye özgü sapma ve bir belirli bir durumda bir öğrencinin matematik puanının kendi tahmin edilen büyüme yörüngesinden duruma özel sapması olarak temsil edilmektedir. Öğrenciler arasındaki ve öğrencinin kendi içindeki farklılıkları ölçen ve bize öğrenci ve/veya durum düzeyinde yordayıcıların araçlar için modele dahil edilip edilmeyeceğini söyleyen, rasgele etkilerin ve artıkların varyansı olarak tanımlanmaktadır (Hoffman, 2017).

Çok Düzeyli Boylamsal Verilerin Modellenmesi. Çok düzeyli boylamsal verilerin modellenmesinde farklı alanlarda farklı gruplandırmalar yapılarak analizler gerçekleştirilmektedir. Örneğin ekonomi ve biyoistatistikte panel veriler;

Klasik Model, Sabit Etkiler Modeli ve Tesadüfi Etkiler Modeli olarak üç grup altında toparlanmıştır. Her model için oluşturulan regresyon denklemleri ise tek bir formül altında birleştirilmiştir. Bu modellerin hangisinin kullanılacağına belirlenmesi için bazı durumlar söz konusudur. Yerdelen Tatoğlu (2018)'na göre, gözlenemeyen birim etkilere hata terimi gibi tesadüfi bir değişken olarak davranılıyorsa, "tesadüfi etkiler"; her bir yatay kesit gözlem için yordanan bir parametre gibi davranılıyorsa "sabit etkiler" söz konusu olmaktadır. Genel olarak, eğer yatay kesit boyut büyük bir anakütleden tesadüfi olarak çekilmişse tesadüfi etkileri, eğer daha spesifik bir veri set söz konusu ise, sabit etkileri düşünmek daha mantıklı olabilmektedir. Ayrıca tesadüfi etkiler modelinde, birim etkiler ile

yordayıcı değişkenler arasındaki korelasyonun sıfırdan farklı olduğu varsayılmakta iken, sabit etkiler modelinde bu korelasyonun sıfırdan farklı olabilmesine izin verilmektedir. Veri setlerinin test etme sürecinde ise birim etkinin varlığı için Lagrange Çarpanları, Benzerlik Oran Testi, F Testi, Score Testi gibi testler kullanılır. Birim etkinin varlığı test edildikten sonra ise Hausmann Testi ile modelin tesadüfi etkiler ya da sabit etkiler modeline uygunluğu belirlenir. Genel olarak boylamsal verilerin regresyon denklemi ise;

$$Y_i = X_i\beta + z_i b_i + \varepsilon_i \quad (26)$$

şeklinde olmaktadır. (26) numara ile gösterilen bu eşitlikte X_i : sabit etkili ortak değişkenler matrisini, β : Sabit etkileri, z_i : Tesadüfi etkili ortak değişkenler matrisini, b_i : tesadüfi etkileri ve ε_i : ise hata terimini göstermektedir. Bazı regresyon parametreleri birimden birime değişkenlik göstermektedir. Görüldüğü üzere sabit katsayı ve eğim katsayısı her birey için farklıdır, yani birimden birime değişmektedir. Verilen eşitlikteki gösterim katsayılar matrisinden oluşmaktadır. Bu eşitliğin iki düzeyli doğrusal modele genişletilmesiyle elde edilen eşitlik ise;

$$Y_{ij} = X_{ij}\beta + z_{ij} b_{ij} + \varepsilon_{ij} \quad (27)$$

şeklinde olacaktır. (27) numara ile gösterilen bu eşitlikte β : sabit etkileri, z_{ij} : düzey-2 deki rasgele etkileri, b_{ij} düzey 2 de değişen ancak düzey 1 de sabit olan etkileri göstermektedir.

Bireysel değişikliklerin birçok incelemesi iki düzeyli bir büyüme modeli olarak temsil edilebilir. Düzey 1’de, her bir bireyin zaman içindeki ardışık ölçümleri bireysel bir büyüme yörüngesi ve rasgele hata ile tanımlanır. Düzey 2’de, bireylerin grupları arasındaki yörüngelerdeki farklılıklar incelenebilir. Bireyleri (i) ve ölçüm durumlarını (t) tanımlamak için iki alt simge kullanarak ilerlenen gösterimlerde, i bireyi için t zamanında gözlemlenen durumun, Y_{ti} , sistematik bir büyüme yörüngesi ile rasgele hatanın eklenmesiyle elde edilen bir fonksiyonu olduğunu varsayılmaktadır (Heck ve diğerleri, 2010). Bu şekilde belirtilen eşitliklerin betimlediği polinom eğrileri, standart doğrusal modelleme prosedürleri

kullanılarak yordanabildiklerinden dolayı genellikle bireysel değişimi incelemek için kullanılırlar (Hox, 2010).

Yukarıda belirtilen (26) ve (27) numaralı eşitliklerdeki katsayı matrisleri yerine katsayıların kendileriyle oluşturulan ve zamana bağlı bireysel değişimlerin incelendiği bir çalışmada Düzey 1 eşitliği şu şekilde olmaktadır:

$$Y_{ti} = \pi_{0i} + \pi_{1i}a_{ti} + \varepsilon_{ti} \quad (28)$$

(26) numaralı eşitliğe göre bu model, düz bir çizginin (lineer regresyon eğrisi) her bireyin zaman içindeki gerçek değişimini yeterince temsil ettiğini tanımlar. Burada a_{ti} değişkeni zamanla değişen ilgi değişkeni olup π_{0i} katsayısı kesişimi, π_{1i} katsayısı doğrusal büyüme oranını tanımlar. Ayrıca örnek verilerde gözlemlenen doğrusallıktan herhangi bir sapmanın rasgele ölçüm hatasından (ε_{ti}) kaynaklandığını yani bireyler içindeki büyüme tahminindeki varyasyonu temsil ettiğini varsayılır. Eğim parametresi olan π_{1i} her zaman aralığında bireylerdeki değişimi temsil eder. Doğrusal bileşen, birim zaman başına değişim oranını tanımlar. Zamanla ilgili a değişkeninin doğrusal bileşeninin, veri setinde 0, 1, 2 olarak kodlanması kesişimin (π_{0i}) bireylerin gerçek başlangıç durumu (yani birinci zamandaki düzeltilmiş Y düzeyleri) olarak yorumlanmasını sağlar (Heck ve diğerleri, 2010).

Düzey 1 modelinin diğer kısmı stokastik kısım veya her bir i bireyinin t vesilesiyle ölçülmesindeki varyasyonu temsil eden kısımdır. Modelin bu kısmı, her bir bireyin gerçek değişim yörüngesinin (yani, gözlemlenen ve gerçek yörünge arasındaki fark) ölçülmesiyle ilişkili bir miktar hata (ε_{ti}) olduğunu yansıtır. Hata terimleri her ölçüm durumu için dahil edilebilir. Hatalar gözlemlenemez olması Düzey 1'deki dağılımları hakkında bazı varsayımlarda bulunmamız gerektiği anlamına gelmektedir (Singer & Willett, 2003). Yaygın olarak, her bir hatanın bağımsız ve normal dağıldığı, ortalamasının sıfır ve varyansının sabit olduğu, durumdan duruma ve kişiden kişiye basit bir artık yapısına sahip olduğu varsayılır ve;

$$\varepsilon_{ti} \sim N(0, \sigma_{\varepsilon}^2) \quad (29)$$

yani ortalaması 0 ve varyansı σ_{ε}^2 olan bir dağılım gösterdiği şeklinde ifade edilir. Ancak boylamsal verilerde bu tür basit Düzey 1 hata yapısı daha az güvenilir olabilir (Singer & Willett, 2003).

Singer ve Willett'in (2003) belirttiği gibi, Düzey 1 modeli tüm bireysel değişim yörüngelerinin (28) numaralı eşitlikte gösterildiği gibi aynı cebirsel forma sahip olduğunu varsayar, ancak her birey tam olarak aynı yörüngeye sahip değildir. Her bir kişi farklı kesişim ve eğim katsayılarına sahip olabileceğinden, Düzey 1 modeli başka bir tür rasgele katsayılar modelini temsil etmektedir. Düzey 1 modeline zamanla değişen diğer ortak değişken türlerini de (örneğin cinsiyet, sosyoekonomik durum) eklenebilmektedir. Bireysel değişim analizindeki kesişim (yani, sonucun başlangıç gibi belirli bir zamandaki seviyesi) ve büyüme oranı (yani, bağımlı değişkendeki değişme hızı) en önemli parametreler olarak karşımıza çıkmaktadır. Bireyler arasında iki birey içi katsayısının rasgele değişen olarak incelendiği durumda

$$\pi_{0i} = \beta_{00} + r_{0i} \quad (30)$$

$$\pi_{1i} = \beta_{10} + r_{1i} \quad (31)$$

Burada β katsayıları kesişim noktalarını ve r her eşitlik için ilgili artık bileşenleri temsil etmektedir. Eşitlik (30) ile (31) in eşitlik (28)'te yerine yerleştirilmesiyle, Düzey 2 yordayıcıları olmadan sabit ve rasgele bileşenleri incelemek için tek eşitlikli modele ulaşabiliriz:

$$Y_{ti} = \beta_{00} + \beta_{10}zaman_{ti} + r_{0i} + r_{1i} + \varepsilon_{ti} \quad (32)$$

Düzey 2 ile belirtilen (32) numaralı eşitlikte, öğrencilerin başlangıçtaki başarı seviyelerinde ve büyüme yörüngelerinde gözlemlenen farklılıkları açıklamak için modele bir dizi yordayıcı (örneğin cinsiyet) eklenebilir. Bu durumda bireylerin cinsiyetinin, bireysel kesişim noktalarındaki ve doğrusal büyüme oranlarındaki farklılıkları açıklayabileceği önerilecektir. Düzeyler arası etkileşimler, öğretmen etkinliği gibi Düzey 2 (bireyler arası)

değişkenlerinin Düzey 1 eğim katsayısı, yani öğrencilerin yıllık test büyüme oranları üzerindeki etkilerini içerir. Düzeyler arası etkileşimler, cinsiyet gibi Düzey 2 değişkenlerinin (yani bireyler arası değişkenlerin), Düzey 1 eğim katsayısı, yani bireylerin zaman içindeki büyüme oranları üzerindeki etkilerini içerir. (Heck vd, 2010). Bu durumda Düzey 2 eşitliği;

$$\pi_{0i} = \beta_{00} + \beta_{10}(cinsiyet)_i + r_{0i} \quad (33)$$

$$\pi_{1i} = \beta_{10} + \beta_{11}(cinsiyet)_i + r_{1i} \quad (34)$$

şeklinde formülize edilebilir. Burada burada r_{0i} ve r_{1i} sırasıyla, bireyler arasındaki kesişim ve eğim parametrelerinin ölçülmesiyle ilişkili varyasyonu temsil etmektedir. Yukarıda yer alan açıklamalarla, bu çalışmada bireysel değişimi araştırmak için iki seviyeli temel bir model sunulmuş olup sırasıyla çok düzeyli boylamsal veri modellemeleri açıklanmıştır.

Eğitim bilimlerinde daha çok bu regresyon denklemlerinin düzeylere göre değişkenler açısından incelenmesi söz konusu olmaktadır. Örneğin bir boylamsal çalışmada tekrarlı ölçümler bireylerin içine yuvalandığı için birinci düzey değişkenleri zaman değişkeni olurken ikinci düzey değişkeni bireye ait özellikler (cinsiyet, ırk vb.) olmaktadır. Farklı alanlarda farklı notasyonlarla gösterimler olmasına rağmen bu çalışmada mümkün olduğunca iki gösterim de sunularak ilerlenmiştir. Ancak kullanılacak olan R paketlerinin daha çok ekonomi ve biyoistatistikte kullanılmasından dolayı gösterimlerde matrissel notasyonlara da yer verilmiştir.

Çalışmamızın devamında, değişimle ilgili kişi içi ve kişiler arası soruları aynı anda ele almamıza nasıl izin verdiğini gösteren çok düzeyli değişim modelini tanıtılmıştır. İstatistiksel modeli yazmanın birkaç yolu olmasına rağmen, bağımlı değişkendeki değişim için çok düzeyli model, her bir kişinin zaman içinde nasıl değiştiğini açıklayan bir düzey 1 alt model ve bu değişikliklerin insanlar arasında nasıl farklılaştığını açıklayan bir düzey-2 model olan bir çift yardımcı modelin aynı anda varsayılmasıyla belirlenmiştir. (Bryk & Raudenbush, 1987; Rogosa & Willett, 1985).

Boylamsal Verilerde Çok Düzeyli Regresyon Fonksiyonunu Kestirmek İçin Kurulan Modeller

Çalışmamız için uygun modelin seçilmesi amacıyla öncelikli olarak uygun modele karar vererek ilerlemeliyiz. Araştırmacıların kullanmak isteyebilecekleri çok sayıda farklı çok düzeyli modele rağmen, çok geniş kapsamlı model hakkında; Boş Modeller, Rasgele Kesişim Modelleri, Rasgele Kesişim ve Eğim Modelleri olarak düşünmek faydalı olacaktır (Luke, 2004). Yine alanyazında bu modellemelerin aynı eşitliklerle gösterilmesine rağmen farklı şekillerde isimlendirildikleri de görülmektedir (Goldstein, 1995; Singer & Willett, 2003; Luke, 2004; Heck & Thomas, 2020).

Bu modellerde bağımlı değişkendeki değişim; zamana bakılmaksızın bireyler arasındaki değişim ve hem bireyler hem de zamana bağlı değişim olarak iki önemli şekilde bölümlere ayrılır. Bu durumların modellenerek incelenmesi ve böylece bu modelleri kullanarak bağımlı değişken üzerinde keşfetmeye değer sistematik bir değişiklik olup olmadığı ve bu varyasyonun birey içinde veya bireyler arasından kaynaklanıp kaynaklanmadığına dair yorum yapılarak uygun model üzerinde ilerlenebilmesi amacıyla modeller hakkında bilgi verilmiştir.

Model 1: Boş Model. Hiçbir yordayıcı değişkenin modelde yer almadığı ve “boş model (null model)/ büyümesiz model (no growth model)/ koşulsuz ortalama modeli” diye de anılan modeldir. Bu model bağımlı değişkendeki değişimi zaman içinde tanımlamak yerine, bağımlı değişkene ait varyasyonu basitçe tanımlar ve bölümlere ayırır. Bu model, tüm ölçümler ve bireyler arasındaki bağımlı değişkenin tüm grup ortalamasını tahmin eden sabit bir etkiye sahiptir (Singer & Willett, 2003). Bu model (35) numaralı eşitlik i birey, j zaman değişkeni olmak üzere (35) numaralı eşitlikle aşağıda gösterilmiştir (Heck & Thomas, 2020):

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij} \quad (35)$$

Boş model, Düzey 1'de, i kişisi için gerçek bireysel değişim yörüngesinin tamamen düz olduğunu ve β_{0i} yüksekliğinde bulunduğunu şart koşar. Yörünge, zamansal bir öngörücü ile ilişkili bir eğim parametresinden yoksun olduğu için eğimi yoktur/düzdür. (35) numaralı eşitlikle belirtilen Düzey 1 modelinde β_{0j} sabit parametre yani kesişim katsayısı olup ε_{ij} birinci düzey hata terimi olarak gösterilir ve bu hata terimi modeldeki rasgele parameter olarak tanımlanır.

Düzey 2 modeli ise (36) numaralı eşitlikle aşağıdaki gibi tanımlanmıştır:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (36)$$

Düzey 2 alt modelinin tek parçası, bu düz yörüngelerin yükseklik bakımından farklılık gösterebilmesine rağmen, evrendeki herkes için ortalama yüksekliklerinin γ_{00} olduğunu şart koşar. Yani yükseklikteki bireyler arası herhangi bir değişiklik, yordayıcılarla bağlantılı değildir. Özetle, her birey bir kesişim noktasına sahiptir ancak Düzey 1'e ait regresyon denklemini gösteren eşitlik (35) incelendiğinde, bu eşitlikte zamana bağlı fonksiyon olmadığı için model tarafından kestirilen puanlarda herhangi bir değişiklik bulunmamaktadır. Yani bu modelde zaman içinde bağımlı değişken üzerinde bir değişimin olmadığı kabulü vardır (Singer & Willett, 2003).

Kurulan modelin güvenilirliğini hesaplamak ve toplam varyansın kısımlarını tanımlamak için sınıf içi korelasyon katsayısı (intra-class correlation coefficient) kullanılır. Y'deki toplam varyasyon sadece birey içi ve bireyler arası varyans bileşenlerinin toplamıdır. Bu katsayı;

$$\rho = \sigma_0^2 / (\sigma_0^2 + \sigma_\varepsilon^2) \quad (37)$$

ile hesaplanır. Büyümesiz model, zaman içindeki genel performans seviyesini temsil eden bir kesişme olan bir gizli değişkene sahiptir.

Model 2: Birinci Düzey Yordayıcılı Model (Sabit Eğim Rasgele Kesişim Modeli). Klasik regresyon yönteminde araştırmacılar herhangi bir açıklayıcı değişkeni, bağımlı değişken üzerindeki varyansı açıklamaya katkısına dayalı olarak modele

ekleyerek model geliştirme adımlarını sürdürürler. Geleneksel regresyon modellerindeki kesişim, tüm yordayıcılar ($X_{1,...,q}$) sıfıra (0) ayarlandığında sonuç değişkeninin değeri olarak yorumlanır. Bu yorumlama, kukla değişkenler veya 0'ı sınırlayan ölçeklere sahip sürekli ölçümlerde olduğu gibi, tüm X'lerin anlamlı bir 0 noktasına sahip olması durumunda yeterli olmasına rağmen genellikle bu koşulun geçerli olmaması göz önünde bulundurulduğunda, kesişim değerleri araştırmacılar için çok az yorumlayıcı değere sahiptir. Çok düzeyli modellerde ise alt düzey kesişimler ve eğimler sonuç olarak ele alındığından, bu değişkenlerin gerçekçi sayısal anlamlara sahip olması kritik önem taşımaktadır. Bu nedenle, düzey 1 değişkenleri genellikle çeşitli "merkezeleme" biçimleriyle dönüştürülerek her değişken için anlamlı 0 değerleri elde edilir (Heck & Thomas, 2020). Dolayısıyla, kesişim ve eğim katsayılarının sabit olduğu klasik regresyon analizi yönteminden farklı olarak, çok düzeyli regresyon analizinde araştırmacıların, Düzey 1 katsayılarının sabit mi, Düzey 2'de rasgele değişen mi yoksa Düzey 2'de rasgele değişmeyen mi olması gerektiğini belirlemesini gerektirir. Bu ikinci tür spesifikasyonda, Düzey 1 katsayısı gruplar arasında değişebilir, ancak bu değişim daha çok tamamen grup düzeyindeki bir yordayıcının düzeylerine bağlı olup rasgele değildir (Raudenbush & Bryk, 2002).

Düzey-1 kestiricilerinin X ile gösterildiği ve i indisinin birey, j indisinin ölçüm zamanı olarak tanımlandığı bir model (38) numaralı eşitlik ile aşağıdaki gibi gösterilmiştir;

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij} \quad (38)$$

Yukarıda belirtilen (38) numaralı eşitlikte Y_{ij} değeri i bireyinin j zamanındaki bağımlı değişkeni, β_{0j} değeri j indisiyle gösterilen ölçüm tekrarı boyunca değişen kesişim parametresini, β_{1j} değeri ise her ölçüm zamanı için ortak ve sabit olan eğim parametresini temsil etmektedir. X_{ij} ile gösterilen değer, yordayıcı değişkeni ve ε_{ij} değeri ise i bireyinin j zamanındaki ölçüm değerlerinin hatasını gösterir.

(38) numaralı eşitlikte verilen modelin Düzey 2 eşitliği ise:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (39)$$

$$\beta_{1j} = \gamma_{10} \quad (40)$$

olarak gösterilmiştir. (39) numaralı eşitlikte verilen β_{0j} değeri, j zamanı boyunca kesişen birinci düzey kesişim parametresini, γ_{00} değeri zaman içindeki ortalama kesişim katsayısını, u_{0j} değeri ise j zamanındaki ölçüm değerinin hata değerini belirtmektedir.

Model 3: Birinci Düzey Yordayıcı Model (Rasgele Eğim Rasgele Kesişim Modeli). Bu model Düzey 1 eşitliğinin (38) numaralı eşitlikteki gibi temsil edildiği ancak Düzey 2 eşitliklerinin değiştiği modelle şöyle açıklanmaktadır:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (41)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (42)$$

Burada β_{0j} rasgele değişen kesişme noktasını ve β_{1j} rasgele değişen kesişme noktasını temsil eder.

Yukarıda bahsedilen modeller göz önünde bulundurulduğunda çalışmamızın ilerleyen kısımları için açıklayıcı olması açısından örneklemek ve özetlemek gerekirse;

Eşitlik (38) ile verilen eşitliklerin cinsiyet kategorik değişkenine göre düzenlenmiş hali (43) numaralı eşitlikteki gibidir:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Cinsiyet)_i + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(Cinsiyet)_i + u_{1j} \quad (43)$$

Tüm 2. düzey alt modeller gibi, (43) numaralı eşitliğin de her biri düzenli bir regresyon modelini andıran birden fazla bileşeni vardır. Bu bileşenler birlikte ele alındığında, bir bireyin büyüme yörüngesinin kesişimini (β_{0j}) ve eğimini (β_{1j}), yordayıcı değişken olan cinsiyet ile ilişkili olabilecek 2. seviye sonuçlar olarak ele alır. Her bileşenin ayrıca, bir kişinin 1. düzey parametrelerinin (β 'lar) diğerlerinden stokastik olarak farklı olmasına izin veren kendi artık değeri (u_{0j} ve u_{1j}) vardır. Eşitlik (43)'te görülmemeyle birlikte, bu düzey 2 alt modelin iki bileşeni yedi evren parametresine sahiptir. Bunlar eşitlik

(43)'te gösterilen dört regresyon parametresi (γ 'ler) ve üç artık varyans/kovaryans parametresidir. Verilerdeki değişimin incelenmesi sürecinde çok düzeyli model kurallarına uyulduğu takdirde bu parametrelerin tümü yordayabilmektedir. Eşitlik (43) göz önüne alındığında cinsiyet değerinin (1) kız/(0) erkek olduğu durum için bireylerin matematik başarılarının gözlemlendiği bir örnek çalışma için bu parametreler ve parametrelere ait tanımlar Tablo 1'de örneklendirilerek özetlenmiştir.

Tablo 1

Çok Düzeyli Modele Ait Parametrelerin Tanımı ve Yorumlanması

Sembol	Tanım/Açıklayıcı Yorum
Düzey 1 modeli için (Eşitlik 38)	
Bireysel büyüme parametreleri	β_{0j} Evrendeki her bireyin j zamanındaki gerçek değişim yörüngesinin kesişme değeri yani i bireyinin başlangıçtaki başarı puanlarının gerçek durumu
	β_{1j} Evrendeki her bireyin j zamanındaki gerçek değişim yörüngesinin eğim değeri yani matematik başarı puanlarındaki gerçek yıllık değişim oranı
Varyans bileşeni	σ_ε^2 Evrendeki her i bireyi için, tüm ölçüm durumlarında Düzey-1 kalıntı varyansı. Yani bu değer her i bireyinin varsayılan değişim yörüngesi etrafındaki gözlemlenen verilerinin net (dikey) dağılımını özetler.
Düzey 2 modeli için (Eşitlik 43)	
	γ_{00} Düzey 1'deki kesişmelerin evren ortalaması, düzey 2'nin öngörücü değeri 0 olan bireyler için β_{0j} değeridir. Yani cinsiyeti 0 olanlar için ortalama gerçek başlangıç değeridir.
	γ_{01} Düzey 1'deki kesişimindeki evren ortalama farkı, düzey 2 öngörücüsündeki 1 birimlik fark için β_{0j} değeridir. Yani kız ve erkek cinsiyetler arasındaki ortalama gerçek başlangıç değeridir.
	γ_{10} 2. düzey öngörücü değeri 0 olan bireyler için 1. düzey eğimlerin evren ortalaması, β_{1j} değeridir. Yani cinsiyet değeri 0 olanlar için ortalama yıllık gerçek değişim oranıdır.
	γ_{11} Düzey 2 yordayıcısındaki 1 birimlik fark için düzey 1 eğimindeki evren ortalama farkı, β_{1j} değeridir. Yani kız ve erkekler arasındaki ortalama yıllık gerçek değişim miktarıdır.
	σ_0^2 Evrendeki tüm bireylerde gerçek kesişme noktası olan β_{0j} 'deki düzey 2'deki artık varyansdır.
	σ_1^2 Evrendeki tüm bireylerde gerçek eğimde 2. seviye artık varyans, β_{1j} .
Sabit Etkiler	Evrendeki tüm bireylerde gerçek kesişim, β_{0j} ve gerçek eğim, β_{1j}
Varyans bileşeni	σ_{10} arasındaki düzey 2'deki artık kovaryansı değeridir.

Verilerin Merkezileştirilmesi

Doğrusal regresyon analizinde kesişim terimi β_0 , yalnızca denklemdeki tüm bağımsız değişkenler anlamlı bir sıfır noktasına sahip olduğunda anlamlı bir şekilde yorumlanabilir. Bununla birlikte, birçok sosyal bilim değişkeninin anlamlı bir sıfır noktası

yoktur, çünkü örneğin sıfır yaş, sıfır zeka, sıfır dışadönüklük veya okuldaki sıfır başarı genellikle herhangi bir bireyden ve/veya psikolojik ölçeklerden bunları ölçmesi beklenmez. Sosyal bilimlerde araştırmalarda kullanılan veri yapılarına ait yapılar, sıfır puanlarının mümkün veya anlamlı olmadığı bir şekilde tasarlanmıştır. Bu tür değişkenler için sıfırı anlamlı bir puan yapmanın bir yolu, onları ortalamaktır (Geiser, 2013). Merkezleme, her bir ham puandan bir tür anlamlı sabitin çıkarılması anlamına gelir. Genellikle bu amaçla aritmetik ortalama (ortalama) kullanılır. Ortalamadaki puanların merkezlenmesi, bir değişkenin sıfır noktasının değişkenin gözlenen ortalamasına kaymasına neden olur. Bu, anlamlı bir sıfır noktasıyla sonuçlanır (orijinal sıfır noktası anlamlı olmasa bile), çünkü merkezlemeden sonraki sıfır değeri, örneklemdaki tipik (ortalama) durumu ifade eder (Cohen ve diğerleri, 2003).

Maksimum Olasılık (Maximum Likelihood-ML) Kestirim Yöntemi

Çok düzeyli modellerdeki parametrelerin hesaplanmasında kullanılan istatistiksel teori, yürütülen analizin türüne, bağımlı değişkenin ölçüm ölçeğine ve gruplar içinde ve gruplar arasında analiz edilen verilerin özelliklerine de bağlı olduğundan, OLS regresyonu ile karşılaştırıldığında nispeten karmaşıktır. Model parametrelerini yordamak için iteratif (yinelemeli) tahmin prosedürleri gereklidir. Maksimum olasılık (ML), dengesiz veri ve rasgele kesişim ve eğim varyansları içeren çok düzeyli modeller gibi kapalı biçimli çözümleri olmayan tahmin modellerine yaygın olarak uyarlanabilen bir tahmin yaklaşımıdır. Amaç, verileri gözleme olasılığını en üst düzeye çıkaran parametre tahminlerini bulmaktır. Bu, modeldeki rasgele parametreler için, verilerin olasılığıyla çarpılan doğal logaritmanın negatifini en aza indiren bir dizi ağırlık belirleyen iteratif bir süreci içerir. Makine öğrenimi ile model tahmini, hızları ve yakınsamanın güvenilirliği açısından farklılık gösterebilen birkaç hesaplama algoritmasından birinin kullanılmasını gerektirir (Raudenbush & Bryk, 2002; McCoach ve diğerleri, 2018; Heck, 2020). Örneklem boyutu arttıkça, ML tahminlerine ait sonuçlar asimptotik olarak daha tutarlı hale gelir, evren parametrelerinin bilinmeyen gerçek değerlerine yakınsamaları artar ve asimptotik

olarak normal dağılırlar. Ayrıca standart hataları diğer yöntemlerle elde edilen değerlere göre daha küçüktür. Başka bir avantajı da, ML tahminlerinin herhangi bir fonksiyonunun aynı zamanda bir ML tahmini olmasıdır. Bu, yordanan büyüme yörüngelerinin (başlangıç durumuna ilişkin ML tahminlerinden ve değişim oranlarından oluşturulan) gerçek yörüngelerin ML tahminleri olduğu anlamına gelir. İstatistikçiler tutarlı ve verimli, iyi kurulmuş normal teoriden yararlanan ve daha karmaşık nicelikler için makul tahminler üretebilen tahminleri tercih ettikleri için ML yöntemi sıklıkla tercih edilmektedir (Judith & Singer, 2003).

Kantil Regresyon ve Çok Düzeyli Boylamsal Veri Yapısı ile İlişkisi

Bu çalışmanın kantil regresyon alt başlığında belirtildiği üzere, bir bağımlı değişkeninin ortak değişkenleri ile koşullu ortalaması arasındaki ilişkiyi modelleyen geleneksel sıradan en küçük kareler regresyonunun aksine, kantil regresyonun, ortak değişkenler ile bağımlı değişkeninin koşullu kantilleri arasındaki ilişkiyi modellediği ve kantil regresyonun, literatürde genellikle sıradan en küçük kareler regresyonuna çekici bir alternatif olarak kabul edilmektedir (He & Shi, 1996; He ve diğerleri, 2002; Lee, 2003; Koenker, 2005; Tian ve diğerleri, 2016). Çok düzeyli regresyon modellemeleri göz önünde bulundurulduğunda ise, sürekli veriler için tasarlanan mevcut çok düzeyli (hiyerarşik) modellemelerin çoğu, bağımlı değişkenlerin ortak değişkenlere doğrusal olarak bağlı olduğunu ve/veya hata dağılımlarının bilindiği varsayılmaktadır. Bu iki varsayım birçok uygulamada pratik olmayabilir (Chen ve diğerleri, 2014). Çok düzeyli regresyon modellemelerinin çoğu, normal olarak dağıtılmış kümeye özgü rasgele etkiler ve rasgele hatalar yoluyla kümeler arası heterojenliği kontrol ederek, bağımlı değişken üzerindeki ortak değişken etkilerini tahmin eder. Bununla birlikte, bu merkeziliğe dayalı çıkarımsal çerçeve, bağımlı değişkenin koşullu dağılımı (rasgele terimlere bağlı olarak) çarpık, çok modlu olduğunda veya atipik (farklı uçlarda yer alan) gözlemlerden etkilendiğinde genellikle yetersiz kalır. Buna karşılık, bağımlı değişkeninin tüm koşullu dağılımını ölçen koşullu kantil regresyon yöntemleri (Koenker, 2004; Koenker, 2005), bağımlı değişkenin

keyfi kantil dilimlerinde ortak deęişken etkilerinin deęerlendirilmesini saęlayabilecek şekilde geliřtirilmiřtir. Ek olarak, kantil regresyon yntemleri, hata teriminin sıfır kořullu nicelięe sahip olması dıřında, hata terimleri zerinde herhangi bir daęılım varsayımı dayatmaz (Galarza ve dięerleri, 2017).

Kantil regresyon yntemi bařlangıçta, tek deęiřkenli bir çerçeve altında geliřtirilmiř olsa da, çok dzeyli (kmelenmiř) veri yapılarının birçok alanda daha sık kullanılmasının ardından yapılan çalıřmalarla asimetrik Laplace daęılımını (ALD) kullanarak karma modelleme çerçevesine (klasik veya Bayesian) doęru geniřletilmiřtir. Kantil regresyon modelleri (Koenker & Park, 1996), normal olmayan verileri ve heterojen kořullu daęılımları modelleme esneklięi nedeniyle, boylamsal analizlerde ortalamaya dayalı regresyon modellerine nemli bir alternatif olarak kullanılmıřtır (Lee ve dięerleri, 2020). Koenker (2004)'den bu yana, boylamsal veriler iin kantil regresyon yntemleri hakkında byyen bir literatr bulunmaktadır. Koenker (2004)'e gre, boylamsal yapılara sahip verilerin analizi iin Gauss kořulları altında en kk kareler tahmincilerine neredeyse tamamen odaklanma bir zorluk olarak alınabilir: "Boylamsal veri analizine ynelik daha esnek, daha saęlam bir yaklařım Gauss rasgele etkiler çerçevesinin dıřında oluřturulabilir mi?" sorusu çerçevesinde Koenker (2004) tarafından kantil regresyonun byle bir geliřmede daha etkili bir rol oynayabileceęini vurgulamıřtır. Aynı zamanda kantil regresyonun hiyerarřik yapıdaki veri yapılarına geniřletilme ihtiyacı, olduka farklı yaklařımlara yol amıřtır. Bunlar; daęılımdan baęımsız ve olabilirlięe dayalı yntemler olarak kabaca iki gruba ayrılabilirler (Lee ve dięerleri, 2020). İlki, sabit etkileri (Koenker 2004; Lamarche 2010; Galvao ve Montes-Rojas 2010; Galvao 2011) ve aęırlıklandırılmıř yaklařımları (Lipsitz ve dięerleri, 1997; Karlsson 2008; Fu & Wang 2012) ierir. İkincisi ise, esas olarak asimetrik Laplace (AL) yoęunluęuna (Geraci & Bottai 2007; Liu & Bottai 2009; Yuan & Yin 2010; Lee & Neocleous 2010; Farcomeni 2012) veya dięer parametrik daęılımlara (Reich ve dięerleri 2010) dayanmaktadır. Bu kategoriler hibir şekilde birbirini dıřlamaz, rneęin, Koenker (2004) tarafından nerilen ceza yntemleri, Geraci ve Bottai (2007) tarafından

önerilen çift üstel rasgele etkilere sahip asimetrik Laplace regresyonu ile sıkı bir ilişkiye sahip olabilir.

Koenker ve Machado (1999) kantil regresyon ve ilgili çıkarım süreçleri için bir uyum iyiliği süreci tanıttılar ve hata terimi için asimetrik bir Laplace dağılımının (ALD) parametrik varsayımı altında olabilirlik oranı istatistiğini değerlendirdiler. Bu nedenle, bir polinom kantil regresyon fonksiyonuna dayanan parametrik bir yaklaşımı benimseyen Yu ve Moyeed (2001) dahil olmak üzere birçok yazar tarafından dikkate alınan kantil regresyon üzerine Bayesci çıkarım için olası bir yaklaşım sağlar. Boylamsal verilerin modellenmesinde, parametre tahminlerinde yanlılığı önlemek için her birey üzerinde tekrarlanan ölçümlerden kaynaklanan bağımlılık dikkate alınmalıdır (Farcomeni, 2012).

Çok düzeyli kantil regresyon yöntemi ile ilgili literatür artık kapsamlı olmasına rağmen, bu yöntem için olasılık temelli bir bakış açısıyla kesin çıkarımlar yapan herhangi bir çalışma bulunmamaktadır (Galarza ve diğerleri, 2017). Bu çalışmada, tam olasılık tabanlı uygulamanın EM algoritmasının stokastik bir versiyonunu takip ettiği (Stochastic Approximation of the EM, SAEM) sağlam bir parametrik ALD tabanlı çok düzeyli kantil regresyon yöntemi aracılığıyla yapılacak analizler hakkında bilgi verilecektir. Çalışmanın devamı özellikle kantil regresyon ve ALD arasındaki bağlantı ile EM ve SAEM algoritmalarının ana hatları olmak üzere bazı ön bilgiler sunularak ilerletilmiştir.

Çok Düzeyli Kantil Regresyon Modeli

Genel olarak rasgele etkilere sahip çok düzeyli regresyon modeli matrisel formda şöyle tanımlanır;

$$y_{ij} = x_{ij}^T \beta_p + z_{ij} b_i + \varepsilon_{ij}, \quad i = 1, \dots, n; \quad j = 1, \dots, n_i \quad (44)$$

Eşitlik (44)'e göre y_{ij} , i. bireyin j. gözlemdir; x_{ij}^T , $N \times k$ boyutunda tasarım matrisidir; β , $k \times 1$ boyutunda sabit efekt parametreleridir; z_{ij} , $q \times 1$ boyutunda tasarım

matrisidir; ve b_i , $q \times 1$ boyutunda rasgele etkilerdir. Böylece, çok düzeyli regresyon modeli için p . kantil regresyon eşitlik (45) olarak yazılır.

$$Q_p(y_{ij}|x_{ij}, b_i) = x_{ij}^T \beta_p + z_{ij} b_i \quad (45)$$

(45) numara ile verilen çok düzeyli regresyon modelinin p . kantile bağlı fonksiyonunu sırasıyla rasgele etkiler içermeyen, rasgele kesişim içeren ve hem rasgele kesişim hem de rasgele eğim içeren modeller olarak genişletildiğinde (Luo ve diğerleri, 2011) $t_j=j, i = 1, \dots, n; j = 1, \dots, m$ olmak üzere;

Model 1: Rasgele Etkiler İçermeyen Kantil Regresyon Fonksiyonu.

$$y_{ij} = x_{ij}^T \beta_p + \varepsilon_{ij} \quad \text{ve} \quad Q_{\varepsilon_{ij}}(p|x_{ij}) = 0 \quad (46)$$

şeklinde ifade edilir. Burada $x_{ij}^T = (1, t_j)$, $\beta = (\beta_0, \beta_1)^T$ olarak gösterilebilir.

Model 2: Rasgele Kesişimleri İçeren Kantil Regresyon Fonksiyonu.

$$y_{ij} = x_{ij}^T \beta_p + z_{ij}^T b_i + \varepsilon_{ij} \quad \text{ve} \quad Q_{\varepsilon_{ij}}(p|x_{ij}, b_i) = 0 \quad (47)$$

şeklinde ifade edilir. Burada $x_{ij}^T = (1, t_j)$, $\beta = (\beta_0, \beta_1)^T$, $z_{ij}^T = 1$, $b_i = b_{i0}$

olarak gösterilebilir.

Model 3: Rasgele Kesişim ve Rasgele Eğimleri İçeren Kantil Regresyon Fonksiyonu.

$$y_{ij} = x_{ij}^T \beta_p + z_{ij}^T b_i + \varepsilon_{ij} \quad \text{ve} \quad Q_{\varepsilon_{ij}}(p|x_{ij}, b_i) = 0 \quad (48)$$

şeklinde ifade edilir. Burada $x_{ij}^T = (1, t_j)$, $\beta = (\beta_0, \beta_1)^T$, $z_{ij}^T = (1, t_j)$, $b_i = (b_{i0}, b_{i1})^T$ olarak gösterilebilir.

Skolastik Yaklaşım Beklenti Maksimizasyonu (Stochastic Approach Expectation Method-SAEM) Kestirim Yöntemi.

Beklenti maksimizasyonu (EM) algoritması, fonksiyonların maksimumlarını bulmak için güçlü bir hesaplama tekniğidir. Eksik veri modellerinde maksimum olasılık, tahmin için

istatistiklerde yaygın olarak kullanılır. Ancak bazı durumlarda, beklenti adımı kapalı biçimde gerçekleştirilemediği için bu yöntem uygulanamaz. Bu sorunların üstesinden gelmek için, EM algoritmasının beklenti adımını bir stokastik yaklaşım prosedürünün bir yinelemesiyle değiştiren stokastik yaklaşım EM (SAEM) adlı yeni bir yöntem Delyon ve diğerleri (1999) tarafından yapılan çalışmada tanıtıldı. SAEM algoritmasının yakınsaması, birçok pratik duruma uygulanabilir koşullar altında kurulur. Bunlardan biri de; EM Algoritmasının Stokastik Yaklaşımının (SAEM), çok düzeyli boylamsal veriler için kantil regresyonu gerçekleştirmek için kullanılmasıdır. Asimetrik Laplace Dağılımının hiyerarşik temsilini açıklamak için, sabit etkilerin ve varyans bileşenlerinin tam maksimum olabilirlik tahminleri SAEM'e dayalı olarak türetilir (Sun ve diğerleri, 2017). Ayrıca SAEM algoritması, her yinelemede, koşullu dağılımla eksik verileri art arda simüle eder ve modelin bilinmeyen parametrelerini günceller. Böylece SAEM, her yinelemede çok sayıda kayıp veriyi simüle etme zahmetli sorununu ortadan kaldırarak daha hızlı ve verimli bir çözüm yolu sağlar (Galarza ve diğerleri, 2017). k yinelemesinde SAEM algoritması aşağıdaki gibi ilerler:

E- Adımı:

Stokastik Yaklaşım: $\hat{\theta}^{(k)}$ fonksiyonu θ 'nın k-ıncı iterasyonu olmak üzere, $Q(\theta|\hat{\theta}^{(k)})$ fonksiyonunu şu şekilde güncellenir;

$$Q(\theta|\hat{\theta}^{(k)}) \approx Q(\theta|\hat{\theta}^{(k-1)}) + \partial_k \left[\frac{1}{m} \sum_{l=1}^m \log f(y_{gözlenen}, p^{(l,k)}; \theta) - Q(\theta|\hat{\theta}^{(k-1)}) \right] \quad (49)$$

M- Adımı:

Maksimizasyon: $\hat{\theta}^{(k)}$ fonksiyonu $\hat{\theta}^{(k+1)} = \text{argmax} Q(\theta|\hat{\theta}^{(k)})$ olarak güncellenir.

Kantil Regresyon ve ALD Arasındaki İlişki.

Asimetrik Laplace dağılımı (ALD), asimetriye izin vermek için standart Laplace dağılımını genişleten sürekli bir olasılık dağılımı olup asimetrik verilerin modellenmesinin

gerekli olduğu alanlarda kullanılır (Kozubowski & Podgorski, 2000). Verilerdeki bağımlı değişkenin koşullu dağılımını modellemek ve çarpıklığı yakalamak için esnek bir çerçeve sağladığından boylamsal veriler için çok düzeyli kantil regresyon yönteminde kullanılır. Lou ve diğerleri (2011) tarafından gösterilen lineer koşullu kantil fonksiyonu;

$$Q_{y_i}(p|x_i) = x_i^T \beta_p \quad (50)$$

şeklinde tanımlanmıştır. Bu eşitlikte $Q_{y_i}(\cdot) = F^{-1}_{y_i}(\cdot)$ fonksiyonu, x_i 'ye koşullu bağımlı değişkenin kümülatif dağılım fonksiyonunun tersi olup p -inci kantil regresyonda, aşağıdaki gibi belirtilen minimizasyon problemi ortaya çıkmaktadır:

$$\min \sum_{i=1}^n \rho_p(y_i - x_i^T \beta_p) \quad (51)$$

(51) numaralı ifadede ($\rho_p(u) = u(p - 1(u \leq 0))$) kontrol ya da azaltma fonksiyonu olarak tanımlanmıştır. Eşitlik (51)'deki toplam değer minimizasyonu ile maksimum olasılık teorisi arasındaki bağlantı ALD tarafından sağlanır. Bu çarpık dağılım, Koenker ve Machado (1999) ile Yu ve Moyeed (2001) tarafından yapılan çalışmada ortaya çıkmıştır. Rasgele bir değişken Y 'nin μ , σ , τ parametreleriyle $Y \sim \text{ALD}(\mu, \sigma, \tau)$ şeklinde bir ALD olarak dağıldığını söyleyebiliriz ve buna karşılık gelen olasılık yoğunluğu

$$f(y|\mu, \sigma, p) = \frac{p(1-p)}{\sigma} \exp \left\{ -\rho_p \left(\frac{y-\mu}{\sigma} \right) \right\} \quad (52)$$

şeklinde ifade edilebilir. (52) numara ile gösterilen eşitlikte $0 < p < 1$ çarpıklık parametresidir, σ ölçek parametresidir ve $-\infty < \mu < +\infty$ konum parametresidir. $\mu_i = x_i^T \beta_p$ ve $y = (y_1, \dots, y_n)^T$ olarak kabul edilip, $y_i \sim \text{ALD}(\mu_i, \sigma, p)$ olduğunu varsayarsak, o zaman n adet bağımsız gözlemlerin olasılığı, bir orantı sabiti hariç,

$$L(\beta, \sigma; y, p) \propto \frac{1}{\sigma^n} \exp \left\{ -\sum_{i=1}^n \rho_p \left(\frac{y_i - x_i^T \beta_p}{\sigma} \right) \right\} \quad (53)$$

σ 'yı rahatsız edici/sıkıcı (nuisance parameter) bir parametre olarak kabul edersek, o zaman eşitlik (53)'deki olasılığın parametreye göre maksimizasyonu, eşitlik (51)'e eşdeğerdir.

Kantil regresyonda, genellikle farklı kantiller için eğim katsayılarını karşılaştırmak ilgi çekicidir. O zaman ALD, eğim katsayılarının farklı kantil düzeyi için farklı olabileceği durumla nasıl başa çıkabilir sorusu üzerine, ALD kullanan Bayes modelinde, x üzerindeki koşullu y 'nin farklı kantillerinin aynı eğime sahip olduğunu ima eden $y \sim \text{ALD}(\mu, \sigma, p)$ varsayımlarını uygularız. Ancak, sadece $y \sim \text{ALD}(\mu, \sigma, p)$ ise y 'nin p kantilini hesaplarız ve farklı p için aslında farklı bir model kullanırız. Dolayısıyla, eşitlik (26) doğru olduğu sürece, maksimum olasılık tahmin edicisinin (MLE) eşitlik (26)'daki gerçek β_p 'ya yakınsayacağı anlamında olasılık tutarlıdır. Bu nedenle, Bayes analizinde ALD kullanıldığında, yine de kantil fonksiyonun tutarlı bir tahminini elde edebiliriz ve farklı p için eğim katsayıları farklı olabilir.

Çok Düzeyli Modeller İle Çok Düzeyli Kantil Regresyonda Model Seçimi Kriterleri

Model seçimi, gerçeklikle ilgili farklı teorileri temsil eden iki veya daha fazla modeli karşılaştırmaktan ve bu modellerden hangisinin mevcut verileri en iyi şekilde tanımladığına karar vermektir. Sosyal ve davranış bilimlerinde model seçimine yönelik baskın yaklaşım, boş hipotez testidir (NHT). Bu yaklaşım, yıllar boyunca, pratik konulardan daha felsefi konulara kadar birçok eleştiri aldı (Raftery, 1995; Wagenmakers, 2007; Weakliam, 2004). NHT'yi model seçimi amacıyla kullanmanın en belirgin pratik sınırlamalarından bazıları şunlardır: iki modelin karşılaştırılmasıyla ilgili sınırlaması; bu iki modelin iç içe olması gerekliliği; ve boş model için kanıt bulmanın imkansızlığı.

Bu sınırlamalardan etkilenmeyen alternatif bir model seçimi yaklaşımı, An Information Criterion (AIC) geliştiren Akaike (1973) tarafından başlatıldı. AIC'nin tanıtılmasından kısa bir süre sonra, Schwarz (1978) tarafından Schwarz Bilgi Kriteri (SIC) olarak da anılan Bayes Bilgi Kriteri (BIC) tanıtıldı. BIC; AIC'ye güçlü bir benzerlik gösterse de, iki önlemin oldukça farklı kökleri vardır: AIC, bilgi teorisi çerçevesinde türetilirken, BIC, doğası gereği Bayesçi bir kavram olan sonsal model olasılığına dayanır. Bu bilgi kriterleri, modelin log olabilirliği biçimindeki uyumunu ve modelin boyutu biçimindeki model

karmaşıklığını dengeleyen, cezalandırılmış bir log olasılığı olarak kabul edilebilir. Bir bilgi kriteri için daha küçük değerler, daha uygun modeller anlamına gelir. Genel olarak, bilgi kriterlerinin hesaplanması kolaydır ve çoğu yazılım paketi yukarıda belirtilen bilgi kriterlerinden en az birini rapor eder. AIC ve BIC bilgi kriterleri ise;

$$AIC = -2LL + 2k \quad (54)$$

$$BIC = -2LL + k \ln(N) \quad (55)$$

olarak gösterilir. Burada k, parametre sayısı ve N, gözlem sayısını belirtmektedir.

Model seçimi için bilgi kriterlerinin avantajlarına ve bunların çoğu çok düzeyli modelleme (MLM) programlarına dahil edilmesine rağmen, MLM'de model seçimine ilişkin literatüre NHT hakimdir (Bryk & Raudenbush, 1992; Longford, 1995; Snijders & Bosker, 1999; Hox, 2010). Bilgi kriterleri, modellerin gerçeğin yalnızca yaklaşık değerleri olduğu fikrine dayanmaktadır, bu nedenle mesele “doğru” modeli bulmak değil, incelenen model seti verili olarak en iyi yaklaşan modeli bulmaktır (Burnham & Anderson, 2004). Elde edilen bilgi kriteri değerlerinden daha küçük değerlere sahip modeller daha iyi modeller olarak kabul edilir. Başka türlü ifade edilirse, model seçimi, model belirsizliğini hesaba katmak için kullanılabilir ve tek bir en iyi modeli aramak yerine, küçük bir makul modeller kümesi aramak için kullanılabilir (Draper, 1995).

İlgili Araştırmalar

Birçok disiplinde kullanılan kantil regresyon yönteminin temelleri Koenker ve Bassett (1978) tarafından atılmıştır. Bu çalışmada geleneksel en küçük kareler yordayıcısının, Gauss olmayan hatalara sahip doğrusal modellerde ciddi şekilde yetersiz olabileceğini vurgulayarak kantil regresyon fonksiyonunu tanımlamışlardır. Bu yeni regresyon yönteminin tanıtılmasının ardından özellikle ekonomi ve biyoistatistik alanında sıklıkla kullanımına yer verilmiştir. Bu çalışmaların bazıları kantil regresyonun gelişmesine katkı sağlayabilecek yöntem ve istatistik çalışmaları bazıları ise uygulamaya yönelik gerçek verilerle yapılmış çalışmaları içermektedir. Kantil regresyonun tanıtımı ve

matematiksel gösterimi ile ilgili kapsamlar Koenker ve Hallock (2001), Lu vd. (2003), Naiman (2007), Koenker (2005) tarafından yapılan çalışmalarda yer almaktadır. Ancak alan bazlı daha spesifik örnekler de literatürde yer almış bulunmaktadır. Örneğin kantil regresyonun ekonometri alanında yorumlanması için Machado ve Mata (2005), Portekiz'deki verilerden yararlanarak, 1986-1995 dönemi için ücret dağılımındaki değişiklikleri, bu değişikliklere katkıda bulunan çeşitli faktörlerde ayırtırmak için kantil regresyon yöntemini kullanmışlardır ve eğitim düzeylerinde gözlenen artışın daha büyük ücret eşitsizliğine kesin bir şekilde katkıda bulunduğunu raporlamışlardır.

Buchinsky (1994) tarafından, 1964 ve 1988 yılları arasında ABD ücret yapısındaki değişiklikleri açıklamak için kantil regresyon yöntemi kullanılmıştır. Çalışmada koşullu kantiller arasındaki farkla ölçülen grup içi ücret eşitsizliğindeki değişiklikleri incelenmiş; okullaşma ve deneyim getirilerinin ücret dağılımının farklı kantil değerlerinde farklılık gösterdiğini, ancak değişim modellerinin benzer olduğunu belirtmişlerdir. Ayrıca çeşitli beceri gruplarında ücret eşitsizliğinde de önemli farklılıklar bulunduğu da vurgulanmıştır.

Martins ve Pereira (2004), okullaşma ile düzeyler arası ücret eşitsizliği arasındaki bağlantıyı ele almak amacıyla kantil regresyon uygulamışlardır. Çalışmaya göre incelenen 16 gelişmiş ülkedeki verilerle yapılan analize göre, okullaşma getirilerinin ücret dağılımı üzerinden arttığını belirtmişlerdir. Yani gözlemlenemeyen özellikleri nedeniyle koşullu ücret dağılımının en üstünde yer alan bireylerde okula bağlı kazanç artışı daha yüksek olduğunu vurgulamışlardır.

Rodriguez ve diğerleri (2008) yaptıkları çalışmada kantil regresyon kullanarak Avrupa düzeyinde ücretler ve eğitim arasındaki ilişkiyi analiz etmişlerdir. Veriler, 14 Avrupa ülkesinden (Avusturya, Belçika, Danimarka, Finlandiya, Fransa, Almanya, Yunanistan, İrlanda, İtalya, Lüksemburg, Hollanda, Portekiz, İspanya ve Birleşik Krallık) oluşan bir örneklemden elde edilmiştir. Bu çalışmada ortalama ve cari ücretler için regresyonlar yürüterek Avrupa düzeyinde ücretler ve eğitim arasındaki ilişkinin homojen ve zaman içinde istikrarlı olup olmadığının araştırılmıştır. Elde edilen bulgulara göre,

eđitim ve kazanç arasındaki iliřkide ok fazla heterojenlik olduđu, ancak zaman deđiřkenliđini kontrol etmek iin mevcut cretleri ortalama cretlerle karřılařtırsak lkelere gre daha kk farklılıklar olduđunu ve ayrıca, bir niversite diplomasının orta đretim diplomasından daha deđerli olduđunu raporlamıřlardır.

Epidemiyolojide Austin ve Schull (2003) hastane nakil aralıđı ile hasta ve sistem zellikleri arasındaki iliřkiyi deđerlendirmek iin geleneksel lineer regresyonun performansını kantil regresyonun performansına karřı incelenmesinde kullanmıřlardır.

Ekolojide, Cade ve arkadaşları (2005) habitat iliřkilerinde daha yararlı belirsizlik aralıkları sađlamak iin kantil regresyonun faydasını ve performansını deđerlendirmek iin, habitatın llen yordayıcılarıyla birlikte deđerřen hayvan tepkisinin istatistiksel dađılımları zerindeki eksik bilgilerin etkilerini gzlemek iin simlasyon alıřması yapmıřlardır. Aynı alanda Cade (2011) kantil regresyon tahminlerinin tek ulu gven aralıklarının kullanımını sunmak iin eřdeđerlik testi zerinden topraktaki arsenik oranları ile bir alıřma gerekleřtirmiřtir. Kantil regresyonun heterojen etkileri nicelemek iin bir yntem sađladıđını ve bu nedenle kantiller arasında sabit sınır deđerleri olmayan denklik blgelerini dřnmenin mantıklı olacađına sonucuna ulařmıřtır.

See ve arkadaşları (2021) habitat kalitesini deđerlendirmek iin sađlam yntemler ve metrikler oluřturmak, nesli tkenmekte olan Pasifik alabalıklarının geri kazanılması iin; Muller vd. (2018) farklı evresel etmenlerin hayvan aktiviteleri zerine etkilerinin incelenmesinde; Pringle (2008) fillerin habitat yaratma alıřmalarındaki nemi iin Kenya'dan elde edilen veriler zerinde kantil regresyon yntemini kullanmıřlardır.

Scharf vd. (1998) yaptıđı alıřmada balıklar iin av uzunluđu ve avcı uzunluđu hakkında birka bađımsız veri seti kullanarak en kk kareler ve en kk mutlak deđerler modellerine dayalı regresyon tekniklerini test etmiřler ve tutarlılık iin tahmini eđimleri karřılařtırmıřlardır. alıřmanın sonularına gre, en kk kareler regresyon tekniklerinin, gzlemlerin dađılımındaki y deđerlerine ve dzensizliklere zellikle duyarlı olduđunu ve sıklıkla st ve alt sınırlar iin tutarsız eđim tahminleri rettiđini gstermiřtir.

Buna karşılık, en az mutlak değerler modellerine dayanan kantil regresyon teknikleri, üst ve alt sınır eğimlerinin tutarlı tahminlerini sağlarken, veri kümeleri içindeki uzak y değerleri ve seyreklik için sağlam görünüyordu. Ayrıca kantil regresyon kullanımı, araştırmacı tarafından fazladan keyfi karar verme ihtiyacını ortadan kaldırdığına vurgu yapmışlardır.

Harding ve Lamarche (2009), endojen değişkenlere ve bağımsız değişkenlerle ilişkili bireysel etkilere sahip panel veri modellerine bir kantil regresyon yaklaşımı sunmak adına gerçekleştirdikleri simülasyon çalışmasında, 100 ve 250 örneklem büyüklüğünde X^2 ve $N(0,1)$ dağılımlarını da kullanarak regresyon parametrelerine ait mutlak hata ve yanlılık değerlerini hesaplayarak klasik regresyon yöntemi ve kantil regresyon yöntemlerinin farklı yaklaşımlarıyla elde edilen metotlarla karşılaştırmışlardır.

Galvao (2011) tarafından gerçekleştirilen ekonomi alanındaki tek düzeyli kantil regresyon üzerine yapılan simülasyon çalışmasında hata dağılımı, normal dağılım ve Student-t(3) dağılımları kullanılarak incelenmiştir. Mutlak hata değeri ile ilgili olarak, Gauss durumunda, OLS tabanlı yordayıcıların, ilgili kantil regresyon yordayıcılarından daha iyi performans gösterdiği; buna karşılık, yeniliklerin daha ağır kuyruklu Student-t(3) dağılımı için, kantil yordayıcılarından alınan mutlak hata sonuçlarının, ilgili OLS tabanlı yordayıcılarından daha küçük olduğunu belirtmiştir.

Sobotka vd. (2011) Hindistan'dan alınan çocuklukta yetersiz beslenme verileri üzerinde bir beklenti regresyon modelinin değerlendirilmesinde kantil regresyonun güven aralıkları ile birlikte değerlendirilmesi ve semiparametrik beklenti fonksiyonlarını tanımlamaya çalışmışlardır.

Spiegel vd. (2013) boylamsal veri setinde fetal büyüme çalışmalarında kantil regresyonu kullanmıştır. 12 yıllık çalışma sırasında, İsrail'in kuzeyinden 12 ila 42 haftalık gebeliğin 17.708 gebeden bebeğe ait fetal ölçümler elde edilerek bu ölçümlere dayalı olarak büyüme çizelgeleri oluşturulmuştur.

Yu (2014), Bilgi ve iletişim teknolojisinin (BİT) ölçümü ve tahmini ve BİT'in gelişimini tahmin etmek için 1999'dan 2007'ye kadar olan dönemde BİT'in internet kullanıcıları üzerindeki etkisinin analizinde toplam 163 ekonomiyi içeren Dünya Telekomünikasyon/BİT Göstergeleri ile kantil regresyon analizi gerçekleştirmiştir.

Smith vd. (2015) doğum ağırlığı ve gebelik yaşının kantil fonksiyonlarını ozon maruziyeti ve ebeveyn yaşı, ırk ve eğitim düzeyi dahil olmak üzere çoklu belirleyicilerle ilişkilendirerek çok düzeyli kantil fonksiyon modeliyle çalışmışlardır. Doğum ağırlığı ile her gebelik haftası için ayrı ayrı öngörücüler ve gebelik yaşı ile öngörücüler arasında Teksas Halk Sağlığı Bölgelerinde ayrı ayrı ilişkiler kurarak yorumlamalarda bulunmuşlardır.

McDonald vd. (2017) bel çevresi ölçümlerinin hem yetişkin hem de ergen kadınlar arasında kardiyorespiratuar zindeliği arasındaki ilişkiyi ırksal/etnik değişkenler kapsamında incelemişlerdir. Analiz, 1999'dan 2004'e kadar olan Ulusal Sağlık ve Beslenme İnceleme Anketi verileri kullanılarak 12-49 yaş arası toplam 3977 kişiye uygulanmıştır.

Altın Yavuz ve Gündoğan Aşık (2017) bir mühendislik uygulaması üzerinde kantil regresyon yöntemini EKK yordayıcıları ile karşılaştırmıştır. Beton kırma deneyi için elde edilen sonuçlara göre, kendi verilerinde EKK yöntemi ile elde edilen modelin çıkarsama amaçlı kullanılamayacağı; 0.75 ve 0.25 kantil değerine göre kurulan regresyon denkleminin çıkarsama amaçlı kullanılabileceğini belirtmişlerdir.

Schwarz vd. (2019) doğum ağırlığının farklı kantillerde ırk/etnik gruplar arasındaki bu ilişkideki heterojenliği incelemek için çok düzeyli kantil regresyon modelleri kullanarak kantil regresyon sonuçlarının farklı kantil değerlerinde aynı değere sahip olmadıklarını bulmuşlar ve annelerin ırk/etnik gruplarına göre doğum ağırlığına ilişkin matematiksel ilişkileri açıklamışlardır. Lamichhane vd. (2020) yaptıkları çalışmada annenin hava kirliliğine maruz kalması ile doğum ağırlığı dağılımı arasında bir ilişkinin sosyoekonomik duruma (SES) göre ne ölçüde değiştiği araştırmak için, Seul'de 2007'den 2017'ye kadar tekil doğumlara ilişkin verilerle (n = 1739) kantil regresyon yöntemini kullanarak analiz

yapmışlardır. Yapılan çalışma ile Kore'de hava kirliliğinin doğum ağırlığı dağılımı üzerindeki olumsuz etkilerine dair kanıtlar sunulmuş ve hava kirliliğinin doğum ağırlığı dağılımının alt ucundaki daha güçlü olumsuz etkisi, dağılımın üst ucundan daha büyük klinik öneme sahip olabileceğini belirtmişlerdir.

Rahman & Vossmeier (2019) ortak ve bireye özgü parametrelerin yanı sıra birkaç ortak değişkenle ilişkili çok değişkenli heterojenliği hesaba katabilmesi açısından esnek bir yaklaşım olarak Yeni bir Markov zinciri Monte Carlo (MCMC) yöntemi, modele uyacak şekilde tasarlanmış ve hesaplama verimliliği bir simülasyon çalışmasında göstermişlerdir. Simülasyonda 25., 50. ve 75. dilimler için sınırlara ve birlere karşılık gelen gözlem sayısı sırasıyla (1566; 3444), (2588; 2412) ve (3536; 1464) şeklindedir. Bu yöntemi Amerika Birleşik Devletleri'nde kadınların işgücüne katılımını ve ev sahipliğini incelemek için uygulamışlardır. Önerilen yaklaşımın, ortak ve bireye özgü parametrelerin yanı sıra çeşitli ortak değişkenlerle ilişkili çok değişkenli heterojenliği hesaba katabilmesi açısından esnek olduğu belirtilmiştir. Lee vd. (2020) yaptıkları çalışmada, aralıklı kayıp veriler için gizli sınıfı birleşik kantil regresyon modeline dahil eden belirli bir çoklu atama stratejisi önererek farklı boylamsal verileri belirlemekle ilgilenilen durumlarda veri ilişkilerini daha iyi anlamlandırmak için zaman içinde farklı hastalık sonuçlarına yol açabilecek riskleri içeren veri kalıplarını incelemişlerdir. Aralıklı kayıp değerlere sahip ilaç kullanım verileri için gözlemlenmemiş heterojenlik kümesini içeren bir Bayes kantil regresyon (BQR) kullanan, gizli sınıf tabanlı çoklu değerlendirme yaklaşımı önermişler ve yapılan simülasyon çalışması sonucunda elde edilen bulgular ile, önerilen yöntemin belirli veri dağıtım senaryoları altında geleneksel gizli sınıf tabanlı çoklu değerlendirme yaklaşımı yöntemlerinden daha iyi performans gösterdiğini belirtmişlerdir.

Eğitim bilimleri alanında kantil regresyon çalışmaları incelendiğinde; Mwabu ve Schultz (1996) Güney Afrika'da ırklara göre eğitim getirileri için alternatif açıklamalar için kantil regresyonları kullanmışlardır. Çalışma sonucunda, Güney Afrika'da okula özel ücret getirisinin farklı ırklara göre değişmesinden dolayı ortaya çıkan eşitsizliğinden yola

çıkarak, ücret fonksiyona dayanarak, Afrikalılar arasında getirilerin artıkların dağılımındaki onda bir oranında artmadığını belirtmişlerdir. Bu kantil regresyona göre, özel ücret getirilerini en üst düzeye çıkarmak açısından, diğer her şey eşitken, yükseköğretim kayıtlarını beyazlardan Afrikalılara kaydırmanın etkili olacağını vurgulamışlardır.

Eide ve Showalter (1998) standartlaştırılmış testlerdeki okul kalitesi ile performans arasındaki ilişkinin, “test puanı kazanımlarının” koşullu dağılımında farklı noktalarda farklılık gösterip göstermediğini tahmin etmek için kantil regresyonları kullanmışlardır. Çalışma ile okul kalitesi girdilerinin matematik puanlarındaki değişiklikler üzerindeki etkisine ilişkin kantil regresyon tahminlerini sunmuşlardır ve matematik puanı değişikliklerinin koşullu dağılımında farklı noktalar arasında bazı önemli farklılıklar olduğunu belirtmişlerdir. Bulgular ile matematik puanı değişikliklerinin koşullu dağılımının tepesindeki performansın, uzayan bir okul yılı tarafından iyileştirildiğini, dağılımın en altında yer alan performansın, diğer değişkenler eşit tutularak, fazladan ders süresinden fayda sağlamadığını, öğrenci başına harcamaların koşullu dağılımın alt kısmı için matematik puanlarını artırabileceği raporlaştırılmıştır. Hartog vd. (2001) çalışmalarında, 1980'ler ve 1990'ların başlarında Portekiz'de eğitime geri dönüşlerin evrimini incelemek için kantil regresyondan yararlanmışlardır. Çalışma sonucunda, özellikle 1986 yılında Avrupa Birliği'ne katıldıktan sonra eğitime geri dönüşün arttığını belirtmişlerdir. Sonuçlara göre OLS ile modellemenin ücret yapısının önemli özelliklerini gözden kaçırdığını da kantil regresyon (QR) analizi ile, şartlı ücret dağılımında eğitimin etkisinin sabit olmadığını ve koşullu ücret dağılımında eğitime geri dönüş, daha yüksek dilimlerde olanlar için daha yüksek olduğu vurgulanmıştır.

Martins ve Pereira (2004) eğitim getirilerinin kantil regresyon tahminlerini, okullaşma ve ücret eşitsizliği arasındaki ilişkiyi ele almak için, 1980 ve 1995 yılları arasında on altı Avrupa ülkesinden elde edilen kanıtlara dayanarak eğitimin ücret eşitsizliği üzerindeki etkisini ele almışlardır. Çalışma ile gözlemlenemeyen özellikleri

nedeniyle koşullu ücret dağılımının en üstünde yer alan bireylerde okula bağlı kazanç artışı daha yüksek bulunmuştur.

Buchinsky (2002) çalışmasında, kadınlar için eğitimin getirilerindeki değişiklikleri analiz etmek için kantil regresyon tekniklerini kullanmıştır. Çalışma kapsamında kullanılan veriler 1968, 1973, 1979, 1986 ve 1990 yıllarına ait Mart Mevcut Nüfus Araştırmasından elde edilmiştir. Çalışma bulgularına göre eğitimin getirilerinin, genç gruplar için çok büyük oranda arttığı ancak daha yaşlı gruplar için çok az artış gösterdiği belirtilmiştir. Genel olarak eğitim getirilerinin, ilk ölçüm yıllarında düşük kantillerde daha yüksek olduğu ve ölçüm periyodunun sonunda ise daha yüksek kantillerde daha yüksek değere ulaştığı görülmüştür.

Rangvid (2003) eğitsel akran etkilerini tahmin etmek için OECD PISA örneğinin 2000 yılına ait verilerden elde edilen verileri Danimarka için kayıt verileriyle birleştirerek kantil regresyon yöntemini kullanmıştır. Kantil regresyon sonuçlarına göre, koşullu test puan dağılımının farklı noktalarında farklı akran grubu etkileri olabileceği görülmüştür. Pozitif ve anlamlı akran düzeyi etkisinin, zayıf öğrenciler için en güçlü etkide olduğu ve koşullu test puan dağılımına göre sürekli olarak azalığı vurgulanmıştır. Yani genel olarak, akran grubu etkilerinin, koşullu test puanı dağılımının alt ucunda daha güçlü olduğu, genel okuma performansının en üst düzeye çıkartılması için, öğrencilerin yeteneklerine göre gruplandırılmasından ziyade karıştırılmasının tercih edilmesi gerektiği belirtilmiştir.

Birch ve Miller (2006) öğrencilerin üniversitedeki ilk yıllarında başarıları, kişisel özellikler ve önceki okullarının özelliklerini dikkate alarak, bu faktörlerin etkilerinin okul derecesi dağılımı boyunca nasıl değiştiğini araştırmak için kantil regresyon kullanmışlardır. Bu sonuçların öğrenci seçimi ve ayrıca yetenekli lise öğrencilerini çekmek için bursların kullanılma şekli için etkileri olduğunu vurgulamışlardır.

Sakellariou (2006) Singapur'daki mesleki eğitime genel bir bakış sunmak için kantil regresyon yönteminden yararlanarak, kazanç dağılımı boyunca eğitimden sağlanan fayda modelini araştırmak için yola çıkmıştır. Eğitimin etkisinin dağılımın neresinde daha belirgin

olduğunu belirlemeye yardımcı olması için uygulanan kantil regresyon modeli ile Singapur'daki genel ve mesleki/teknik eğitim arasında, özellikle erkek-kadın farklılıklarına odaklanarak karşılaştırılmıştır. Bulgulara dayanarak, Singapur'daki mesleki eğitim sistemi özellikle orta mesleki niteliklere sahip kadınlara iyi hizmet verdiği, genel eğitim görmüş kadınlara kıyasla daha fazla kazandıkları, daha yüksek işgücü katılımına ve daha yüksek istihdam oranlarına sahip oldukları belirtilmiştir. Ayrıca özellikle, kazanç dağılımının alt dilimlerinde yer alan kadınlar (genel eğitim almış kadınlar için 10. ve 25. dilim ve mesleki eğitim almış kadınlar için 10. dilim) herhangi bir deneyim getirisinden yararlanmamakta oldukları vurgulanmıştır.

Budría ve Moro-Egido (2008) yaptıkları çalışmada, 1994-2001 dönemi için İspanya'da eğitim ve ücret eşitsizliği arasındaki bağlantıyı araştırmışlar ve kantil regresyona dayanarak, farklı nüfus gruplarının koşullu ücret dağılımlarını tanımlamıştır. Çalışmanın bulgularına göre, koşullu ücret dağılımının daha yüksek dilimlerine doğru ilerledikçe, yükseköğretimin getirilerinin oldukça arttığı belirtilmiştir.

Tushar (2012) makalesinde, ulusal düzeyde temsili bir anket kullanarak Hindistan'da eğitimin getirilerini tahmin etmek için kantil regresyonu kullanmıştır. Genel olarak toplumun dezavantajlı sosyal grupları daha düşük ücret alma eğiliminde olduğunu ve aile geçmişinin bireylerin kazançlarını etkileyen önemli bir belirleyici olduğunu belirtmiştir.

Alejo vd. (2014) Arjantin'de eğitimin dağıtım etkilerini incelemek için koşulsuz kantil regresyon yöntemlerini kullanmışlardır. Çalışmada, çoğunlukla eğitimin kazançlar üzerindeki güçlü heterojen etkilerinin etkisinden dolayı, eğitimin Arjantin'deki artan eşitsizliğe olumlu katkıda bulunduğu belirtilmiştir.

Giambona ve Porcu (2015) yaptıkları çalışmada kantil regresyon yaklaşımını kullanarak 15 yaşındaki İtalyan öğrencilerin okuma başarısını etkileyen bireysel arka plan özelliklerini incelemek için 2009 OECD-PISA anketini analiz etmişlerdir. Çalışmada elde edilen sonuçlara göre, yordayıcı değişkenlerin okuma başarısı üzerindeki önemli

etkilerinin farklı kantil dilimlerinde farklı şekilde çalıştığı görülmüştür. Bu durumda düşük ve yüksek performansa sahip öğrencilerin başarıları için farklı yollar önerildiğini vurgulamışlardır. Özellikle, bazı aile geçmişi belirleyicilerinin (ebeveyn eğitimi, evde bilgisayar bulunması ve evde ev ödevi için masa bulunması), devam edilen okul programı ve öğrencinin ikamet ettiği bölge, düşük ve yüksek performanslı okuyucular için önemli ancak farklı bir rol oynadığı belirtilmiştir.

Tighe ve Schatschneider (2016) yaptıkları araştırmada, yetişkin temel eğitim öğrencilerinde beş okuduğunu anlama düzeyinde morfolojik farkındalık ve kelime bilgisinin ortak ve benzersiz katkılarını araştırmak için kantil regresyon yöntemini kullanmışlardır. Çalışma sonucunda morfolojik farkındalığın, okuduğunu anlamının daha düşük seviyelerinde en büyük benzersiz tahmin yeteneğini sergilediği, kelime bilgisinin, okuduğunu anlamının daha yüksek düzeylerinde en iyi tahmin yeteneğine sahip olduğu ortaya çıkmıştır.

Salinas ve Salinas (2016) yaptıkları çalışmada, eğitim, iş özelemleri ve öznel iyi oluş arasındaki ilişkiyi analiz etmişlerdir. Avrupa Sosyal Araştırması'ndan elde edilen veriler kullanılarak kantil regresyon yöntemi kullanılarak mutluluğun farklı kantil noktalarında eğitim değişkenlerinin etkisini değerlendirmişlerdir. Elde edilen bulgulara göre, eğitimin öznel iyi oluş üzerindeki etkisinin dağılım boyunca değiştiği ve eğitimin etkisinin mutluluk dağılımının en üstünde daha düşük düzeyde olduğu görülmüştür. Ayrıca, eğitimin bireylerde yapmayı umdukları işe ilişkin belirli istekler oluşturduğu ve bu istekler karşılanmadığında insanların psikolojik bir maliyete maruz kaldıkları belirtilmiştir.

Costanzo ve Desimoni (2017) INVALSI büyük ölçekli değerlendirmelerinden elde edilen İtalyan ilkököl verilerini, hem kantil hem de standart regresyon yaklaşımları kullanılarak analiz ettikleri çalışmasında, matematik ve okuma puanları, öğrencilerin özelliklerine ve politikayla ilgili olmalarına göre seçilen coğrafi değişkenlere göre değişimlerini incelemişlerdir. Elde edilen sonuçlara göre, İtalya'da cinsiyet ve göçmen statüsünün rolünün, öğrencilerin performansının tüm koşullu dağılımında değiştiğini

görülmüştür. Bu bulgular ile kantil regresyon analizinin, eğitim çıktılarındaki eşitsizliğin belirleyicilerini ve mekanizmalarını keşfetmek için yararlı bir araç olduğunu vurgulanmıştır.

Hyun Sook (2017) tarafından yapılan çalışmada, Kore tarafından 2015 yılında gerçekleştirilen özel eğitim harcamaları anketinden elde edilen veriler, özel eğitim harcamalarını etkileyen faktörlerin araştırılması için kantil regresyonla analiz edilmiştir. Analiz sonuçları, öğrencinin cinsiyetinin, velinin yaşının ve okul sonrası katılım zamanının ve maliyetinin önemli olmadığını göstermiştir. Hanehalkı geliri, özel eğitim harcamalarının tüm düzeyleri (kantilleri) için aynı büyüklüğe oranla pozitif anlamlı olduğunu belirtmiştir. Widyanti (2018) çalışmasında kantil regresyonu kullanarak özellikle Endonezya'da eğitime dönüş ile ücret eşitsizliği arasındaki ilişkiyi incelemeyi amaçlamıştır. Çalışma sonucunda eğitimin farklı kantillerinde eğitime geri dönüş oranlarındaki önemli farklılıkların, aynı eğitim içindeki ücret dağılımının artması nedeniyle eğitimin artan ücret eşitsizliğine katkıda bulunduğunu belirtmişlerdir.

Çok düzeyli kantil regresyonla ilgili çalışmalar ise daha çok yöntem geliştirme ve yeni modeller inşa etme üzerine ilerlemiştir. Boyutluluk çalışmalarında Honda (2004) yerel L1 regresyonu ile değişen katsayıları tahmin ederek koşullu medyanı tahmin etmiş ve Cai ve Xu (2008) ise zaman serisi süreçleri için değişen katsayılı kantil regresyon modellerini araştırarak üç aşamalı yarı parametrik bir prosedür önermişlerdir.

He ve Shi (1996), He ve Liang (2000) ve Lee (2003) parametrik ve yarı parametrik veri örnekleri için kısmen doğrusal kantil regresyon modellerini tartışmışlardır. Karlsson (2008), doğrusal olmayan boylamsal verileri anlamlandırmak için ağırlıklı bir yaklaşım türetmiştir. Farklı bir çerçevede, yarı parametrik kısmen doğrusal modeller Wang ve diğerleri (2009) tarafından incelenmiştir. Son olarak Geraci ve Bottai (2007), kantil regresyonun amaç fonksiyonuyla yakından bağlantılı olan asimetric Laplace dağılımı (ALD) yoluyla rasgele bir kesişme ve sonucu modelleme yoluyla çıkarım yapmıştır. Bu simülasyon çalışmasında dağılımlar sırasıyla, $N(0,1)$, Student-t(3) ve X^2 dağılımı olarak ele alınmıştır. Kullanılan model biyoistatistik alanında sıklıkla kullanılan konum kaydırma

modeli (location shift model) olup 25 ve 50 kantilleri için regresyon parametrelerinin yanlılık değerleri ve verimlerini karşılaştırmışlardır. Elde edilen sonuçlara göre, önerilen yöntem simüle edilen senaryolarda genel olarak diğerinden daha iyi performans gösterdiğini belirtilmiştir.

Tian ve Chen (2006), hata dağılımının normallik varsayımını esneten kantil regresyonun hiyerarşik verilere uygulamışlardır. Geraci ve Bottai (2006) ile Geraci (2013) kantil regresyon için, "Çok Düzeyli Kantil Regresyon" adını verdikleri karma etkili yeni bir yöntem, "Doğrusal Kantil Karma Model"i tanıtmışlardır. Sürekli bağımlı değişkenler için hiyerarşik veri mevcut olduğunda rasgele etkilerin birimler arasındaki bağımlılığı dikkate alarak modele eklendiği bir koşullu kantil regresyon önermektedirler.

Zou ve Yuan (2008), tahmin amaç fonksiyonunu K adet kantil fonksiyonunun ortalaması olarak ayarlayan bileşik kantil regresyonunu (composite quantile regression) tanıtmışlardır. Zou ve Yuan (2008), kantil regresyonun en küçük kareler kestirimi ile karşılaştırıldığında küçük bir nispi verimliliğe yol açabileceğini göstermişler ve sağlam ve verimli bir tahmin oluşturmak için çoklu kantillerin bilgilerini bir araya getiren Birleşik Kantil Regresyon kestirimini önermişlerdir. Birleşik Kantil Regresyonun en küçük kareler kestiriminden çok daha verimli olabileceğini belirtmişlerdir. Birleşik kantil regresyonun hata dağılımının normal dağılımdan çok uzak noktalara saptığı durumlarda doğrusal modellerde (Zou ve Yuan, 2008) ve yarı-parametrik değişken katsayılı kısmi doğrusal modellerde (Kai ve diğerleri, 2011) güçlü ve sağlam olduğu kanıtlanmıştır.

Kai ve arkadaşları (2010) parametrik olmayan regresyon için lokal çok terimli Birleşik Kantil Regresyon üzerinde çalışmışlardır. Kai ve arkadaşları (2011) yarı-parametrik değişken katsayılı kısmi doğrusal modeller için birleşik kantil ve değişken seçimi gerçekleştirmişlerdir. Jiang ve arkadaşları (2012) doğrusal olmayan modeller için Birleşik Kantil Analizi yoluyla değişken seçimini araştırmışlardır.

Chen ve arkadaşları (2014) çok düzeyli doğrusal modeli ile birleşik kantil regresyon modellemesinin birleştirilmesiyle elde edilen ve birleşik çok düzeyli kantil regresyon modeli

adı verilen yeni bir modeli ortaya koymuşlardır. Bu yeni modelin, model çok düzeyli modellerin bütün özelliklerini paylaşan, marjinal etkiler olarak bilinen tahmini katsayı vektörü tepki değişkeninin aykırı gözlemlerine dirençli olma özelliklerine sahip olduğunu belirtmişlerdir.

Liu ve Bottai (2009) tarafından gerçekleştirilen simülasyon çalışmasında ise hata terimlerinin dağılımları sırasıyla; normal, Student-t(3) ve X^2 dağılımı olarak tasarlanmıştır. Her senaryoda, 1000 simüle edilmiş veri seti üretilerek yapılan çalışmada sabit regresyon katsayılarını; 0.25, 0.50, 0.75 kantil düzeylerinde hesaplayarak çok düzeyli regresyon modeli ile karşılaştırmak için yanlılık değerleri hesaplanmıştır. Sabit regresyon kesişimi değerinin hata teriminin normal ve Student-t(3) dağılım gösterdiği durumlarda kantil fonksiyonunun 0.50, 0.75 ve 0.25'ten daha etkili olduğu; hata terimi X^2 dağıldığında, alt çeyreklerde sabit regresyon kesişimi değerinin daha etkili olduğunu belirtmişlerdir. Hata terimi normal dağıldığında ortalama regresyonun medyan regresyondan daha iyi performans gösterdiği ve dağılımın ağır kuyrukları olduğunda daha az verimli olduğunu raporlamışlardır.

Farcomeni (2012), Gizli Markov Kantil regresyon modelini önererek, hata terimlerinin normal, Student-t(3) ve X^2 dağılımı olduğu, rasgele etkiler için ise normal, Student-t(3) ve X^2 dağılımı ve gizli Markov dağılımlarını kullandığı simülasyon çalışmasında 1000 veri seti ile (0.25, 0.5, 0.75) kantil düzeylerinde model parametrelerinin yanlılık ve standart hata değerlerini incelemiştir. Çalışma sonucunda aşırı dağılımı modellemek için gizli Markov yapısının birçok avantajı olduğu, bu modelin bilinmeyen gizli dağılımın basit ve esnek bir modellemesine izin verdiğini belirtmiştir. Ayrıca sabit parametreler için ortaya çıkan tahminlerin simülasyonlarda iyi performans gösterdiği raporlanmıştır.

Tzavidis vd. (2016) yaptıkları çalışmada model parametrelerinin standart hatalarını tahmin etmek için, Düzey 1 ve Düzey 2 hata terimlerinden dağılımını $N(0,3)$, $N(0,5)$, Student t-(3) ve Laplace dağılımlarını kullanarak iki düzeyli konum kaydırma modeli

yardımıyla, toplam 1259 gözlem ile simülasyon çalışması gerçekleştirmişlerdir. Her iki hata düzeyinin de normal dağıldığı durumda, 0.50 kantil düzeyinde lineer (doğrusal) rasgele etkiler regresyonundan sabit etkilerin yordayıcılarının, M-kantil regresyonundan karşılık gelen yordayıcılardan daha verimli olduğunu belirtmişlerdir. Kantil düzeylerinin $q=0.75$ ve $q=0.9$ olduğu durumda, M-kantil rasgele etkiler modelinin sabit etkilerinin yordayıcılarının, karşılık gelen M-kantil regresyonunun yordayıcılarından daha verimli olduğu raporlanmıştır.

Cai ve Xiao (2012) kısmen değişen katsayılara sahip dinamik modeller için kantil regresyon tahminini incelemişler ve hem parametrik hem de parametrik olmayan fonksiyonel katsayıların tahminini önermişlerdir. Ayrıca başka bir boyut küçültme modelleme yöntemini, kısmen değişen katsayı modellerini ele alarak, bazı değişkenlerde katsayılar da doğrusallığa ve diğer değişkenlerde doğrusal olmayan durumlara izin veren bir model önermişlerdir.

Kong ve ark.(2016), bağımlı değişkenin koşullu ortalama işlevini karakterize etmek için yüksek boyutlarda yeni bir işlevsel “kısmen doğrusal model sınıfı” önermiştir, önerdikleri model yalnızca koşullu dağılımın merkezi eğilimine odaklanmıştır. Wang vd. (2016) Bayesyan kantil yapısal eşitlik modellemesinde 500 ve 1000 örneklem büyüklüğü kullanarak simülasyon çalışması gerçekleştirmişlerdir.

Ülkemizde ise ekonomi ve istatistik alanlarında kantil regresyonu tanıtmaya ve yöntem geliştirme çalışmalarının (Saçaklı, 2005; Keskin, 2012; Yılmaz, 2014; Karacan, 2018; Yaman, 2019) yanı sıra uygulamaya yönelik çalışmaların ağır bastığı söylenebilir. Bu çalışmalarda yurtdışında olduğu gibi bankacılık, finans ve işgücü ile ilgili konular cinsiyet, gelir, sosyoekonomik düzey vb. çeşitli değişkenler bağlamında incelenmiştir (Tansel ve Bodur, 2012; Akyol, 2013; Çelik, 2013; Gökçe, 2013; Koşan, 2014; Kutlu, 2016; Akgün, 2017; Tokatlıoğlu, 2018; Gürsoy, 2019; Kılınç, 2020).

Çiçek (2022) yaptığı çalışmada, kantil yapısal eşitlik modellemesi çalışmasında örneklem büyüklüğü olarak 100, 500, 1000 ve 3000 seçerek hata terimlerinin farklı

dağılımlar gösterdiği durumlar için Bayesyen kantil yapısal eşitlik modellemesinin performansını değerlendirmiştir. Hata terimlerinin normallikten saptığı durumlarda Bayesyen kantil yapısal eşitlik modellemesinin Bayesyen YEM'e göre daha iyi sonuçlar verdiğini raporlamıştır.

Eğitim alanında ise sadece PISA 2015 verisi üzerinde yapılmış çalışmalar bulunmaktadır. Bu çalışmalarda Çobanoğlu Aktan (2017), Türk öğrencilerin başarı düzeyi ile ilgili faktörleri kantil regresyon yöntemiyle incelemiş ve sonuçları sıradan doğrusal regresyon ile karşılaştırmıştır. Çalışmada elde edilen bulgulara göre, ekonomik, sosyal ve kültürel statü endeksinin etkisinin farklı miktarlarda değişmekte olduğu raporlanmıştır. Bu etkinin maksimum 75. kantilde ve tüm kantillerde istatistiksel olarak anlamlı olduğu; bununla birlikte, test kaygısının düşük kantillerde istatistiksel olarak anlamlı bir etkisinin bulunmadığı görülmüştür.

Tongal (2017) gerçekleştirdiği çalışmada PISA verisinde yer alan bazı değişkenlerin (epistemolojik inançlar, ailenin en yüksek eğitim düzeyi, evdeki eğitim materyalleri, sosyal statü vb.) fen testi başarı puanına etkisini Türkiye'deki bölgeler bazında incelemiştir. Ailenin en yüksek eğitim düzeyi ve sosyal statü değişkenlerinin etkisi Akdeniz Bölgesi'nde diğer bölgelere kıyasla daha yüksek iken ebeveyn iş statüsü indeksi, ebeveyn kültür düzeyi, evdeki eğitim materyalleri ve epistemolojik inançlar değişkenlerinin diğer bölgelere kıyasla daha az etkisi olduğu belirtilmiştir. Ebeveynlerin duygusal desteği ve öğretmenin adaleti değişkenlerinin Güneydoğu Anadolu'da diğer bölgelerden ayrılarak etkisinin daha yüksek olduğu belirtilmiştir. Her üç bölgede de disiplin açısından sınıfın iklimi değişkeninin etkisinin Türkiye ortalamasından daha yüksek olduğu ve sosyal statü değişkeninin etkisinin düşük, orta ve yüksek puan alan öğrencilerde tüm bölgelerde fazlaca farklılık göstermediği belirtilmiştir. Epistemolojik inançlar değişkeninin etkisinin İstanbul Bölgesi'nde yüksek puan alan öğrencilerde orta ve düşük puan alan öğrencilere göre etkisinin belirgin şekilde az olduğu; ayrıca bu değişkenin etkisinin Akdeniz Bölgesi'nde düşük puan alan öğrencilerde orta ve yüksek puan alan öğrencilere oranla

daha az olduđu belirtilmiřtir. ebeveynlerin duygusal desteęi deęiřkeninin etkisi dūřuk, orta ve yksek puan alan oęrencilerde Tkrkiye geneli, İstanbul ve Akdeniz Bklgelerine fazlaca bir deęiřim gstermezken Gneydoęu Anadolu Bklgesi'nde orta düzeyde puan alan oęrencilerde etkisinin ok yksek olduđu vurgulanmıřtır.

Demirez (2018) yaptığı alıřmada bazı deęiřkenlerin (fen oz yeterlięi, fen etkinlikleri sayısı, haftalık fen oęrenme zamanı vb.) fen bařarı puanına etkisinde Tkrkiye, Singapur ve Almanya'yı karřılařtırmıřtır. alıřma sonucunda, fen oz yeterlięi ve oęretmen temelli fen eęitimi deęiřkenlerinin etkisinin dięer lkelere kıyasla Singapur'da daha yksek; haftalık fen Oęrenme zamanı ve okul dıřında oęrenmeye ayrılan zaman deęiřkenlerinin dięer lkelere oranla daha az etkisi olduđu belirtilmiřtir. Fen etkinlikleri sayısı, bilgi ve iletiřim teknolojileri ve okul dıřında oęrenmeye ayrılan zaman deęiřkenlerinin etkisinin dięer lkelere oranla Tkrkiye'de daha yksek olduđu; iřbirlięi ve ekip alıřmasından zevk alma ve ekonomik, kltrrel ve sosyal statü deęiřkenlerinin etkisinin Tkrkiye'deki oęrencilerde en az olduđu belirtilmiřtir. Haftalık fen oęrenme zamanı deęiřkeninin, fen bařarı puanına etkisinin Almanya'da en yksek olduđu; iřbirlięi ve ekip alıřmasına deęer verme deęiřkeninin Almanya'da dūřuk, orta ve yksek puan alan oęrencilerde keskin bir deęiřim gsterirken, Tkrkiye'de bu deęiřkenin fen bařarı puanını yordamamakta olduđu raporlanmıřtır. İřbirlięi ve ekip alıřmasından zevk alma deęiřkeninin etkisi dūřuk, orta ve yksek puan alan oęrencilerde Tkrkiye ve Almanya'da fazlaca bir deęiřim gstermezken, Singapur'da dūřuk düzeyde puan alan oęrencilerde etkisinin ok yksek olduđu belirtilmiřtir. Arařtırma temelli fen oęretim yntemi deęiřkeninin Tkrkiye'de en fazla yksek puan alan oęrencileri etkilerken Singapur'da bu deęiřkenin etkisi en fazla dūřuk düzeyde puan alan oęrencilerde etkiledięi belirtilmiřtir. Bilgi ve iletiřim teknolojileri deęiřkeninin Tkrkiye'de orta puan düzeyindeki oęrencileri en fazla etkiledięi gözlenmiřtir. Ailenin Refah düzeyi deęiřkeninin Tkrkiye'de orta puan alan oęrenciler ile Almanya'da yksek puan alan oęrenciler hari dięer tm oęrenci gruplarında bu deęiřkenin fen bařarı puanını yordamamakta olduđu raporlanmıřtır.

Uluğ (2019), PISA 2015 verisinde yer alan Türk öğrencilerinin bazı değişkenler bağlamında (matematik okuryazarlığı puanı, haftalık fen ders saati, işbirliği ve ekip çalışması, öğretmen adilliği vb.) fen okuryazarlığı ve okuma becerileri başarısına etkisini incelemiştir. Çalışmanın sonuçlarına göre, 0.9 kantil aralığında, haftalık fen ders saati ile 0.5 - 0.9 kantil aralığında haftalık okul dışı çalışma süreleri değişkeninin anlamsız olduğunu belirtmiştir. Matematik okuryazarlığı, işbirliği ve ekip çalışması, öğretmen adilliği, bilim etkinlikleri, geniş bilim konularına ilgi, okula ait olma duygusu, öğrenci tutumları motivasyon eğilimleri ve sınıf tekrarı değişkenlerinin 0.1, 0.5 ve 0.9 kantil aralıklarında anlamlı olduğu saptanmıştır.

Yapılan simülasyon çalışmaları incelendiğinde (Geraci & Bottai, 2007; Liu & Bottai, 2009; Farcomeni, 2012; Tzavidis ve diğerleri, 2016), daha çok regresyon parametresinin kestirimlerine odaklanıldığı ve tüm kantil düzeylerinde bir incelemeye gidilmediği görülmektedir. Bu kapsamda bu çalışmada tüm kantil düzeylerinde model veri uyumu ve bağımlı değişkenin kestirimine ait mutlak hata ve yanlılık değerlerinin incelenmemiş olması literatürdeki boşluğu doldurmak amacıyla çalışmamıza yön vermiştir.

Bölüm 3

Yöntem

Bu bölümde araştırmanın türü, simülasyon koşulları, simülasyon koşullarına bağlı olarak verilerin üretilmesi ve verilerin analizi ile ilgili açıklamalı bilgiler sunulmuştur.

Araştırmanın Türü

Bu araştırmada durumları betimlemenin ötesinde değişkenlerin birbirleriyle ne şekilde ilişkili olduklarını incelendiğinden ilişkisel araştırma yöntemi kullanılmıştır (Büyüköztürk ve diğerleri, 2014). Bu çalışmada analize dahil edilecek veri setleri bilgisayar programı kullanılarak türetildiğinden araştırmamız bir Monte Carlo simülasyon çalışmasıdır (Gentle, 1985).

Veri Toplama Araçları

Çalışmada kullanılan veri setlerinin oluşturulmasında PIAT (Peabody Individual Achievement Tests-Peabody Bireysel Başarı Testleri) veri setinin analiz edilmesiyle elde edilen katsayılar yaklaşık olarak kullanılmıştır. Bu amaçla kullanılan boylamsal veri, çocuklara tekrar tekrar matematik testlerinin uygulandığı, Amerika Birleşik Devletleri Çalışma İstatistikleri Bürosu (Bureau of Labor Statistics) tarafından toplanan ve araştırmacıların kullanımına açık olan ulusal boylamsal anket NLSY-CYA'dan (Ulusal Boylamsal Gençlik Çocuk ve Genç Yetişkin Araştırması, İnsan Kaynakları Araştırma Merkezi, 2004) elde edilmiştir. NLSY-CYA'nın yaptığı araştırmaların verileri açık kaynaklı olup (<https://www.nlsinfo.org/investigator/pages/search.jsp>) erişim sınırı bulunmamaktadır.

PIAT, bireyin matematik dahil farklı alanlardaki akademik başarısını değerlendirmek için kullanılan standartlaştırılmış bir test dizisi olup, yüksek test-tekrar test güvenilirliği ve eşzamanlı geçerliliği ile beş yaş ve üstü çocuklar için çeşitli akademik başarıların ölçütü olarak kabul edilmektedir. Bu test, bir çocuğun temel eğitimde öğretildiği şekilde matematikteki başarısını ölçen ve artan zorlukta 84 adet çoktan seçmeli madde

bulunarak uygulanmaktadır. Tekrarlı gözlemlerde bağımlı değişken olarak seçilen ve Grimm vd. (2010) tarafından analiz edilen Peabody Bireysel Başarı Testinden (Dunn & Markwardt, 1970) gelen matematik puanlarının ölçülmesi ve raporlanması bakımından katılımcılar değerlendirme sayısına göre farklılık gösterir. Bireyler ikinci sınıftan sekizinci sınıfa kadar 2 yıl arayla değerlendirilmiştir ve bazı bireylerin bazı gözlemlerinde eksik değerler bulunmaktadır. Araştırmacıların erişimine açık bu veri setinde, 1986 yılından 2014 yılına kadar her iki yılda bir uygulanmış toplam 12 adet PIAT matematik alt testi puanları bulunmaktadır. Bu çalışmada her bireylerin 6, 8, 10 ve 12. yaşlarına ait eksik gözlem içermeyen alt veri seti kullanılmıştır. Bunun için veri setinde doğum tarihi verilen bireylerin belirtilen yaşlardaki ölçüm değerleri kullanılarak alt veri seti oluşturulmuştur. Bunun için veri setinde doğum tarihi verilen bireylerin belirtilen yaşlardaki ölçüm değerleri kullanılarak alt veri seti oluşturulmuştur. Bu veri setine ait betimsel istatistikler Tablo 2'de gösterilmiştir.

Tablo 2

Cinsiyet Değişkenine Göre Betimsel İstatistikler

	Kişi Sayısı	%	Ortalama	Standart Sapma	İlk Ölçüm Değeri	Standart Sapma
Kız	1006	48.80%	35.88	16.92	15.32	6.20
Erkek	1057	51.20%	35.64	15.996	15.67	5.69
Toplam	2063	100%	35.75	16.452	15.50	5.94

Tablo 2 incelendiğinde toplam 2063 bireyin kız ve erkek oranlarının sırasıyla %48.8 ve %51.20 olduğu görülmektedir. Yani yaklaşık olarak kız ve erkek cinsiyetlerin dağılımları birbirine eşittir. Ayrıca kız ve erkek öğrencilerin ortalama puanlarının birbirine çok yakın değerlerde olduğu (35.88 ve 35.64) görülmektedir. 2063 bireye ait 8252 adet ölçüm değerinin betimsel istatistik değerleri incelendiğinde; ölçüm değerlerinin ortalamasının toplamda 35.75 ve standart sapmasının 16.452 olduğu görülmüştür. Ayrıca sadece ilk ölçüm değerleri incelendiğinde ilk ölçüm değerlerinde kız ve erkeklerin birbirine yakın ortalama değere sahip olduğu (15.32 ve 15.67) ve genel ilk ölçüm değer

ortalamasının ise 15.50, standart sapmasının ise 5.94 olduğu görülmektedir. Verilerimizin taramasının ardından elde edilen diğer tanımlayıcı istatistikler Tablo 3'te gösterilmiştir.

Tablo 3

PIAT Verisine Ait Betimsel İstatistikler

Kişi Sayısı	Ortalama	Standart Sapma	Medyan	En Küçük Ölçüm Değeri	En Yüksek Ölçüm Değeri	Çarpıklık	Basıklık
2063	35.75	16.452	38	0	84	0.010	0.052

Tablo 3 incelendiğinde verilerimizin çarpıklık ve basıklık değerlerinin sırasıyla 0.010 ile 0.052 olduğu görülmektedir. En yüksek ölçüm değeri 84 ve en düşük ölçüm değeri 0'dır. Bireylerin her ölçüm grubuna ait betimsel istatistikler ise Tablo 4'te verilmiştir.

Tablo 4

Bireylerin Ölçüm Gruplarına Ait Betimsel İstatistikler

Ölçüm Yaşı	Ölçüm Sayısı	Ortalama	Sd	Medyan	Min.	Max.	Ranj	Çarpıklık	Basıklık
6	2063	15.5	5.95	14	0	52	52	1.25	3.01
8	2063	31.47	9.67	31	3	66	66	0.09	-0.39
10	2063	44.40	9.71	45	0	84	84	-0.38	1.35
12	2063	51.64	10.19	51	4	84	80	-0.13	0.84

Tablo 4 incelendiğinde bireylerin ölçüm yaşı arttıkça gözlemlenen ortalama puanlarında artış görülmektedir. Bireylerin ölçüm gruplarına ait puanların birbiri ile korelasyon durumu ise Tablo 5'te gösterilmiştir.

Tablo 5

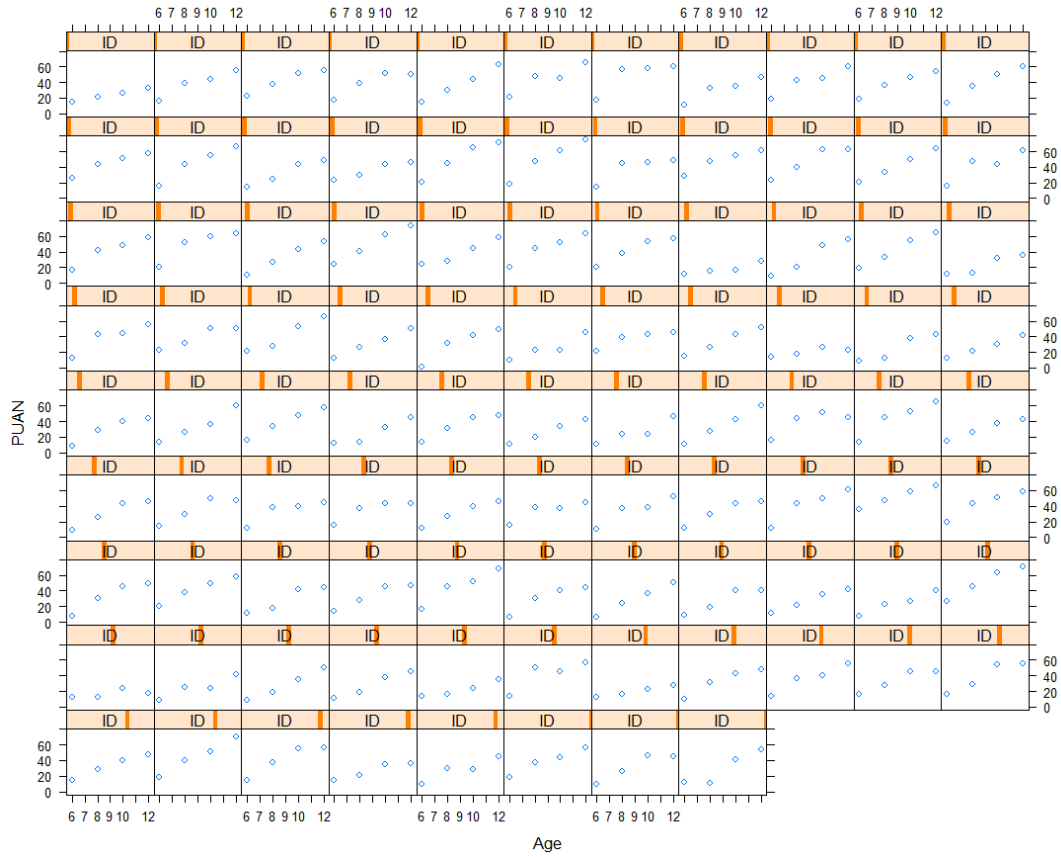
Bireylerin Ölçüm Gruplarına Ait Puanların Korelasyonları

	6	8	10	12
6	1.00	0.544	0.493	0.482
8	0.544	1.00	0.651	0.627
10	0.493	0.651	1.00	0.704
12	0.482	0.627	0.704	1.00

Tablo 5 incelendiğinde art arda iki ölçüm arasında korelasyon değeri daha yüksekken, yıllar arası değer arttıkça korelasyon değerinin azalmakta olduğu görülmektedir. Tüm betimsel istatistik değerleri incelendiğinde puanların gitgide artış eğiliminde olduğu ve bir trend gösterdiği yorumu yapılabilir. Verilerin görselleştirilmesi adına R programında random seçilen yaklaşık % 5'lik örnekleme ait bireysel puanlarının grafiği Şekil 16'da gösterilmiştir.

Şekil 16

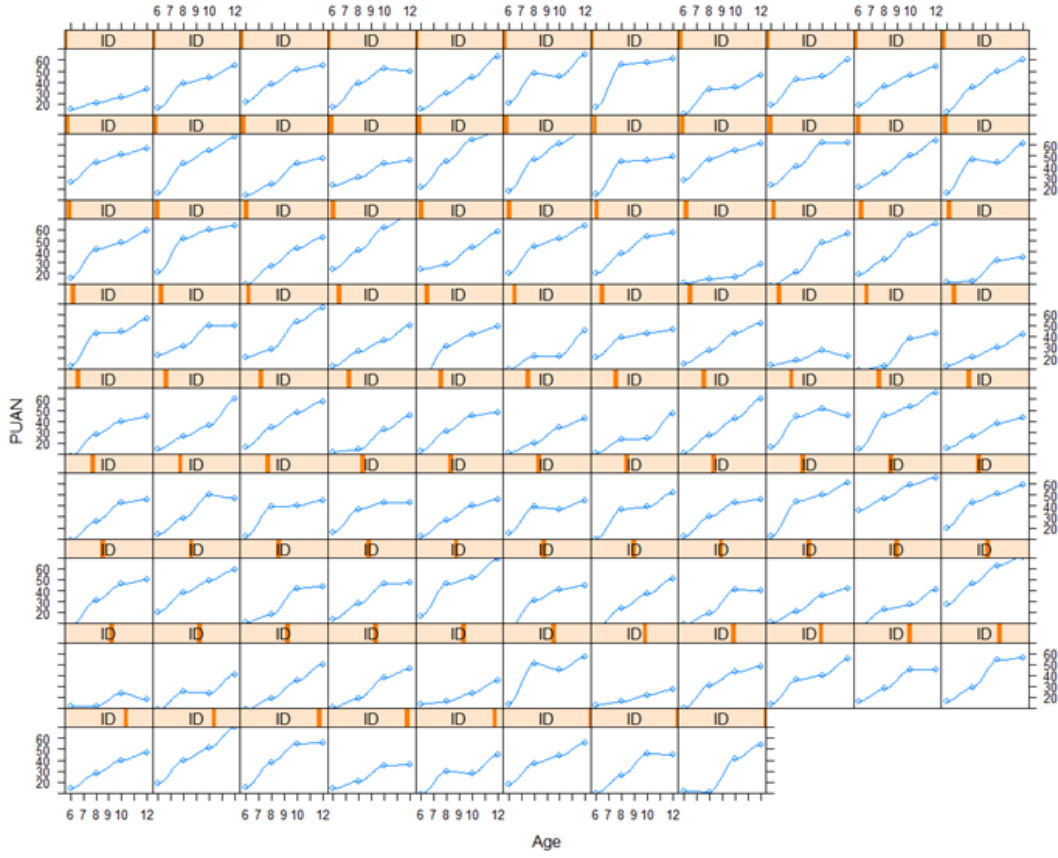
%5 Random Bireylere Ait Büyüme Noktaları



Şekil 16 incelendiğinde bireylerin farklı ölçüm yıllarındaki ölçüm değerlerinin trend gösterdiği görülmektedir. Bu noktaların lineer doğru parçaları ile simgelenen regresyon doğruları ile görselleştirilerek grafikleştirilmesi ise Şekil 17'de gösterilmiştir. Şekil 17 incelendiğinde yaklaşık % 5 random örnekleme seçilen bireylerin puanlarının genellikle artış gösterme eğiliminde olduğunun yanı sıra rasgele kesişim ve rasgele eğim ile modellenebilecek veri yapısına sahip olduğu yorumu da yapılabilir.

Şekil 17

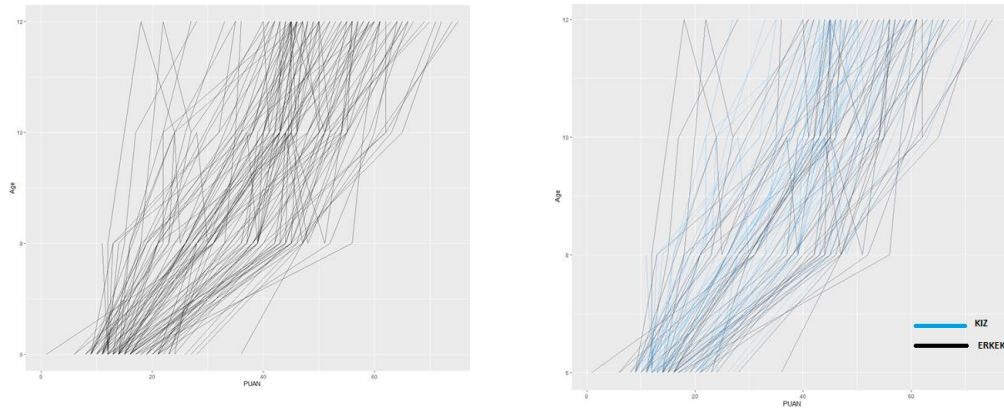
%5 Random Bireylere Ait Büyüme Grafikleri



Benzer şekilde yaklaşık % 5 random örnekleme seçilen bireylerin büyüme çizgileri Şekil 18'de (a) panelinde, cinsiyete göre büyüme çizgileri ise (b) panelinde gösterilmiştir.

Şekil 18

%5 Random Bireylere Ait Büyüme Çizgileri



(a)

(b)

Şekil 18'de (a) paneli incelendiğinde bireylerin büyüme çizgilerinin rasgele kesişim noktasına ve rasgele eğim değerine sahip olduğu, benzer şekilde; (b) paneli incelendiğinde cinsiyet değişkenine göre de aynı rasgele kesişim ve rasgele eğim durumunun geçerli olduğu görülmektedir.

PIAT veri setinde model veri uyumunun görülmesi amacıyla boş model, rasgele kesişim sabit eğim modeli ve rasgele eğim rasgele kesişim modelleri ayrı ayrı kullanılmıştır. Boş model sonrası elde edilen analiz sonuçları ise Tablo 6'da gösterilmiştir.

Tablo 6

Boş Modele Ait Analiz Sonuçları

	Kestirilen Değer	Standart Hata	t değeri
Sabit Etkiler			
Kesişim	17.55	0.189	92.51
Zaman	6.06	0.031	191.84
Rasgele Etkiler			
	Varyans	Standart Sapma	
Bireyden Kaynaklı	45.37	6.736	
Hatalardan Kaynaklı	41.26	6.424	

Tablo 6 incelendiğinde model özetinin rasgele etkiler bölümü veri setinin varyans kaynağı hakkında bilgi vermektedir. Tablo 6'ya göre bireylerden kaynaklanan varyans 45.37 iken, artıklardan kaynaklanan varyans değeri 41.26 olduğu görülmektedir. Sınıf içi korelasyon katsayısı (ICC) hesaplandığında; $ICC = 45 / (45 + 41) = 0.52$ değeri elde edilmektedir. Yani açıklanan varyansın % 52'si bireylerdeki değişimlerden (bireyler arası farklılıktan) kaynaklanmaktadır. Bir başka deyişle bireylerin ölçüm değerleri olan PIAT matematik puanlarının değerinin varyansının %52'si bireyler arası farklılıklar ile açıklanabilir. Bu durumda veri yapısının boş modelden ziyade daha karmaşık bir model ile açıklanabileceği varsayımı ile rasgele kesişim sabit eğim modeli ile rasgele kesişim rasgele eğim modelleri arka arkaya uygulanmıştır. Bu modeller ile elde edilen analiz sonuçları Tablo 7'de toplu olarak gösterilmiştir.

Tablo 7

Rasgele Kesişim Sabit Eğim Modeli Ile Rasgele Kesişim Rasgele Eğim Modellerine Göre Analiz Sonuçları

	Kestirilen Değer	Standart Hata	t değeri	AIC	BIC	logLike
Sabit Etkiler						
Rasgele Kesişim Sabit Eğim Modeli						
Kesişim	17.55	0.189	92.513	57606.78	57634.86	-28541.84
Zaman	6.066	0.031	191.84			
Rasgele Kesişim Rasgele Eğim Modeli						
Kesişim	17.17	0.471	36.448	57099.3	57148.43	-28542.65
Zaman	6.066	0.034	175.799			

Tablo 6 ve Tablo 7'den elde edilen sonuçların karşılaştırılması ve en uygun modelin karar verilmesi amacıyla bu üç model arasında ANOVA testi uygulanmıştır. ANOVA testinin sonuçları ise Tablo 8'de gösterilmiştir.

Tablo 8

PIAT Verisine Ait Farklı Modellere Ait Model Veri Uyumu Değerleri

Model	AIC	BIC	LogLike
Boş Model	696943.350	69664.400	-34818.680
Rasgele Kesişim Sabit Eğim Modeli	57606.780	57634.860	-28541.840
Rasgele Kesişim Rasgele Eğim Modeli	57099.300	57148.430	-28542.650

Tablo 8 incelendiğinde 2063 bireyden elde edilen 8252 gözleme ait farklı modellerin model veri uyum indeksleri karşılaştırıldığında rasgele kesişim ve rasgele eğim modelinin bu veriye en iyi uyum sağladığı söylenebilir. Bu çalışmanın simülasyon kısmında da buradan hareketle bu model kullanılmıştır. Kullanılan PIAT verisinin R paketlerinden qrlmm paketi ile yapılan çok düzeyli kantil regresyon analizine göre model veri uyumu değerleri ise Tablo 9'da gösterilmiştir. Tablo 9 incelendiğinde, PIAT verisinin farklı kantil düzeylerinde en iyi model veri uyumunu 0.25 ve 0.50 kantil düzeylerinde gösterdiği görülmüştür. Ayrıca bu değerlerin Tablo 8'de verilen PIAT verisinin çok düzeyli regresyon analizi ile elde edilen rasgele kesişim ve rasgele eğim modelindeki model veri uyumu değerleri ile benzerlik gösterdiği söylenebilir.

Tablo 9

PIAT Verisinin Çok Düzeyli Kantil Regresyon Analizine Göre Farklı Kantil Düzeylerindeki Model Veri Uyumu Değerleri

Kantil Düzeyi	AIC	BIC	Log-like
q10	59489.16	59538.29	-29737.58
q25	58415.75	58464.88	-29200.88
q50	58471.20	58520.32	-29228.60
q75	59403.43	59452.56	-29694.72
q90	62183.41	62232.54	-31084.71

Çok düzeyli kantil regresyon analizinde, ortalama yerine bağımlı değişkenin kantilleri tahmin edilir. Böylece analiz, PIAT veri setinde bağımlı değişkenin dağılımında farklı noktalarda bağımsız değişkenler ile bağımlı değişken arasındaki ilişki hakkında bilgi sağlayarak ilişkinin gücünü göstermektedir. Ayrıca rasgele etkiler, qrlmm paketi ile her bir kuantil için ayrı ayrı tahmin edilmekte ve bu etkiler, dağılımın farklı noktalarında rasgele etkilerin bağımsız değişkenler ile bağımlı değişken arasındaki ilişkiyi nasıl etkilediği hakkında bilgi vermektedir. Bağımsız değişkenlerin tahmin edilen katsayıları, tahmin edilen her kantil düzeyinde bağımsız değişkenler ile bağımlı değişken arasındaki ilişkinin gücünü gösterir. PIAT verisinin çok düzeyli kantil regresyon analizine göre elde edilen parametre kestirim değerleri ise Tablo 10'da verilmiştir.

Tablo 10

PIAT Verisinin Çok Düzeyli Kantil Regresyon Analizine Göre Farklı Kantil Düzeylerindeki Parametre Kestirim Değerleri

Kantil Düzeyi	Kesişim	Kesişime Ait Standart Hata	Zaman	Zamana Ait Standart Hata	Cinsiyet	Cinsiyete Ait Standart Hata
q10	9.213	0.181	5.362	0.031	0.302	0.111
q25	11.670	0.198	5.827	0.035	0.483	0.119
q50	15.903	0.231	6.438	0.045	0.155	0.135
q75	22.361	0.242	6.631	0.051	-0.463	0.136
q90	29.328	0.238	6.531	0.051	-1.261	0.134

Tablo 10 incelendiğinde, PIAT verisinin her bir kantil düzeyinde elde edilen regresyon denklemlerine ait katsayılar ve standart hata değerleri görülmektedir. Buna göre, her kantil düzeyinden elde edilen kesişim parametresi, zaman parametresi ve cinsiyet parametresi ile ayrı ayrı modellemeler yapılabilmektedir. Çok düzeyli kantil regresyon analizi ile tahmin edilen katsayılar, değişkenlerin matematik başarı puanı arasındaki ilişkinin, tahmin edilen kantil dilimlerinin her birindeki gücünü göstermektedir. Bu kapsamda her bir kantil düzeyi için elde edilen çok düzeyli regresyon denklemleri i birey t ise zaman birimini göstermek üzere;

$$Y_{it} = (\gamma_{00} + \gamma_{01}(Cinsiyet)_i + u_{0t}) + (\gamma_{10} + \gamma_{11}(Cinsiyet)_i + u_{1t})X_{it} + \varepsilon_{it} \quad (56)$$

eşitliği üzerinden ifade edilmek istenildiğinde;

0.10 kantil düzeyi için;

$$Y_{it} = (9.213 + 0.302(Cinsiyet)_i + u_{0t}) + (5.362 + 0.302(Cinsiyet)_i + u_{1t})X_{it} + \varepsilon_{it} \quad (57)$$

0.25 kantil düzeyi için;

$$Y_{it} = (11.670 + 0.483(Cinsiyet)_i + u_{0t}) + (5.827 + 0.483(Cinsiyet)_i + u_{1t})X_{it} + \varepsilon_{it} \quad (58)$$

0.50 kantil düzeyi için;

$$Y_{it} = (15.903 + 0.155(Cinsiyet)_i + u_{0t}) + (6.438 + 0.155(Cinsiyet)_i + u_{1t})X_{it} + \varepsilon_{it} \quad (59)$$

0.75 kantil düzeyi için;

$$Y_{it} = (22.361 - 0.463(Cinsiyet)_i + u_{0t}) + (6.631 - 0.463(Cinsiyet)_i + u_{1t})X_{it} + \varepsilon_{it} \quad (60)$$

0.90 kantil düzeyi için;

$$Y_{it} = (29.328 - 1.261(Cinsiyet)_i + u_{0t}) + (6.531 - 1.261(Cinsiyet)_i + u_{1t})X_{it} + \varepsilon_{it} \quad (61)$$

Eşitlikleriyle ifade edilebilir. Yani her kantil düzeyinden elde edilen bilgi miktarları ve bu bilgiye bağlı olarak kurulan matematiksel modeller kendi içinde ayrı ayrı değerlendirilmelidir.

Özetle, bu çalışma kapsamında bireylerin başarı puanı olarak Peabody Bireysel Başarı Testleri (PIAT) matematik puanları bağımlı değişken, cinsiyet özellikleri ise bağımsız değişken olarak seçilmiştir. Seçilen PIAT alt veri setinde yer alan toplam 2063 bireyin kız ve erkek oranlarının sırasıyla %48.8 ve %51.20 olduğu ortalama puanlarının birbirine çok yakın değerlerde olduğu (35.88 ve 35.64) görülmüştür. Veri setinin rasgele %5'lik örnekleme ile seçilen bireylerin ölçümlerinin görselleştirilmesi ile bireylerin farklı ölçüm yıllarındaki ölçüm değerlerinin trend gösterdiği yani verinin rasgele eğim ve rasgele kesişim modeline uygun şekilde modellenilebilecek bir yapıya sahip olduğu ve yapılan analizlerden elde edilen bulgular ile de farklı modellerin model veri uyum indeksleri karşılaştırıldığında rasgele kesişim ve rasgele eğim modelinin bu veriye en iyi uyum sağladığı görülmüştür. Tüm bu durumlar göz önünde bulundurulduğunda, simülasyon koşullarımızın ve modelimizin çerçevesi PIAT verilerinin analizlerinden elde edilen değerlerle belirlenmiştir. Son olarak simülasyon denkleminiz ile elde edilecek veri setlerinin oluşturulmasında belirlenecek örneklem büyüklüğü için PIAT veri setinin hem tamamına hem de farklı örneklem büyüklük değerlerini içeren alt örneklemlemleriyle (N=50, N=500, N=1000) ayrı ayrı rasgele kesişim ve rasgele eğim modeline göre analizler yapılmıştır. Zaman ve cinsiyet değişkeninin dahil edilerek oluşturulan modellerin analizleri sonucu elde edilen parametre katsayılarına ait kestirimler ve standart hata değerleri Tablo 11'de verilmiştir. Örneklem büyüklüğü arttıkça parametre kestirim değerlerine ait standart hata miktarının git gide azalma eğiliminde olduğu görülmüş olup örneklem büyüklüğünün parametrelerin en az hata ile kestiriminde etkili olduğu düşünülerek verilerimiz farklı örneklem büyüklükleri için oluşturulmuştur.

Tablo 11*Farklı Örneklem Büyüklüklerine Göre Elde Edilen Parametre ve Standart Hata Miktarları*

Örneklem Büyüklüğü	Kesişim Parametresi	Kesişim Parametresine Ait Standart Hata	Zaman Parametresi	Zaman Parametresine Ait Standart Hata	Cinsiyet Parametresi	Cinsiyet Parametresine Ait Standart Hata
N=50	19.963	3.241	5.884	0.233	-2.727	2.025
N=500	17.049	0.950	6.040	0.070	0.378	0.597
N=1000	17.890	0.659	6.100	0.048	-0.209	0.410
PIAT Verisi	17.172	0.471	6.067	0.035	0.252	0.295

Tablo 11 incelendiğinde örneklem büyüklüğü arttıkça verinin tamamına ait parametre kestirim değerlerine yaklaşıldığı ve bu parametrelere ait standart hata miktarının git gide azalma eğiliminde olduğu görülmektedir. Dolayısıyla örneklem büyüklüğünün parametrelerin en az hata ile kestiriminde etkili olduğu yorumu yapılabilir. Bu kapsamda bu çalışmada türetilcek farklı örneklem büyüklükleri sırasıyla 50, 500 ve 1000 olarak belirlenmiştir.

Çalışmanın Verilerinin Üretilmesi ve Simülasyon Koşulları

Çalışmada kullanılacak veri setleri R programında türetilmiştir. Çok düzeyli boylamsal veri setlerinin oluşturulmasında bağımlı değişkeni etkileyen cinsiyet kategorik değişkeni seçilmiştir. Cinsiyet değişkeni kendi içinde Kadın ve Erkek diye kodlanmış olup dağılım sayıları PIAT veri analizinden elde edilen yaklaşık değerler göz önünde bulundurularak yarı yarıya dağılım sağlanmıştır. Literatüre göre, Monte Carlo yönteminde “rasgele değişkenlerin dağılımlarının özellikleri simüle edilmiş rasgele sayılar kullanılarak araştırılır” (Gentle, 1985). Bu doğrultuda bu çalışmada, türetilen veriler eğitim bilimleri alanında çok düzeyli denklem ile modellenen veri yapıları dikkate alınarak, aynı bireylerin üç gözlemleri üzerinden ve büyüme eğrisi modeli göstermeleri için her bir gözlemin bir

önceki gözlemden fazla olması şartı, zamana bağlı değişkenin denkleme eklenmesiyle oluşturulmuştur.

Simülasyon denkleminin oluşturulması için PIAT analizinden elde edilen katsayılar çok düzeyli regresyon denkleminin uygun hale dönüştürülmüştür. Yapılan analizler sonucu elde edilen katsayılar i bireyleri, t ise zamanı göstermek üzere;

$$Y_{it} = \beta_{0t} + \beta_{1t}X_{it} + \varepsilon_{it} \quad (38)$$

eşitliğinin rasgele eğim ve rasgele kesişim modeline göre;

$$\beta_{0t} = \gamma_{00} + \gamma_{01}(Cinsiyet)_i + u_{0t}$$

$$\beta_{1t} = \gamma_{10} + \gamma_{11}(Cinsiyet)_i + u_{1t} \quad (43)$$

alt eşitlikleri baz alınarak birleştirilmiş model üzerinden ilerlendiğinde;

$$Y_{it} = (\gamma_{00} + \gamma_{01}(Cinsiyet)_i + u_{0t}) + (\gamma_{10} + \gamma_{11}(Cinsiyet)_i + u_{1t})X_{it} + \varepsilon_{it} \quad (56)$$

$$Y_{it} = (17.17 + 0.252(Cinsiyet)_i + u_{0t}) + (6.067 + 0.252(Cinsiyet)_i + u_{1t})X_{it} + \varepsilon_{it} \quad (62)$$

Eşitliği üzerinden üç gözlemlenmiş boylamsal veriler, çok düzeyli regresyon analizi yöntemine göre elde edilmiştir. Çalışmada kullanılacak veri setlerinden çok düzeyli veri yapısı ise R programında farklı simülasyon paketlerine ihtiyaç duyulmadan “data.table” ve “Matrix” paketleri yardımıyla türetilmiştir. Denkleme ait katsayılar PIAT veri setine ait değerlerden elde edilmiştir. Bu çalışmada bir ders başarısı üzerine bireylerin gözlemlerine dair veriler içerdiği düşünülerek tasarlandığı için başlangıç değeri 17.17 olarak; cinsiyet değişkenine ait katsayımız ise eğim ve kesişim parametrelerinde aynı ve 0.252 olarak atanmıştır. Düzey 1 ve Düzey 2’ye ait hatalar (u_{0t} , u_{1t} ve ε_{it}) farklı dağılım yapıları kullanılarak denkleme dahil edilmiştir (Bu dağılımlar model alt başlığında detaylı açıklanmıştır). Tablo 11’de görüldüğü üzere örneklem büyüklüğünün etkili olduğu düşünülerek çalışmamızın koşulları oluşturulurken farklı örneklem büyüklüklerine göre

(50, 500, 1000) türetilmiştir. Bu çalışmada farklı faktörlere göre simüle edilmiş veriler için çok düzeyli boylamsal veri yapısını temsil eden rasgele kesişim ve rasgele eğim modelini belirten bir denklem üzerinden ilerlenmiştir. Çalışma çok düzeyli boylamsal veri yapısının denklemine ait Düzey 1 ve Düzey 2 hata terimlerinin farklı dağılımları ve örneklem büyüklüğü koşullarına göre tasarlanmıştır. Bu faktörlere göre oluşturulan veri setlerine ait simülasyon koşulları Tablo 12’de gösterilmiştir.

Tablo 12

Çalışmanın Koşulları

Düzey 1 Hata Teriminin	
Dağılımları	$N(0,3)$; X^2 (df=3) ; Student’s T3 (df=3)
Düzey 2 Hata Teriminin	
Dağılımları	$N(0,5)$; X^2 (df=3) ; Student’s T3 (df=3)
Örneklem Büyüklüğü (N)	50 ; 500 ; 1000
Yöntem	Çok düzeyli regresyon yöntemi ; Çok düzeyli kantil regresyon yöntemi

Tablo 12’de belirlenen simülasyon faktörleri doğrultusunda oluşturulan veri setlerinin koşulları ise Tablo 13’te verilmiştir.

Tablo 13

Simülasyon Koşulları

<u>Koşul Sayısı</u>	<u>Örneklem Büyüklüğü</u>	<u>Düzey 1 Hatalarının Dağılımı</u>	<u>Düzey 2 Hatalarının Dağılımı</u>
Koşul 1		$N(0,3)$	$N(0,5)$
Koşul 2		$N(0,3)$	X^2 (df=3)
Koşul 3		$N(0,3)$	Student’s T3 (df=3)
Koşul 4		X^2 (df=3)	$N(0,5)$
Koşul 5	N=50	X^2 (df=3)	X^2 (df=3)
Koşul 6		X^2 (df=3)	Student’s T3 (df=3)
Koşul 7		Student’s T3 (df=3)	$N(0,5)$
Koşul 8		Student’s T3 (df=3)	X^2 (df=3)
Koşul 9		Student’s T3 (df=3)	Student’s T3 (df=3)

Tablo 13. (devamı)

Koşul 10		N(0,3)	N(0,5)
Koşul 11		N(0,3)	X^2 (df=3)
Koşul 12		N(0,3)	Student's T3 (df=3)
Koşul 13		X^2 (df=3)	N(0,5)
Koşul 14	N=500	X^2 (df=3)	X^2 (df=3)
Koşul 15		X^2 (df=3)	Student's T3 (df=3)
Koşul 16		Student's T3 (df=3)	N(0,5)
Koşul 17		Student's T3 (df=3)	X^2 (df=3)
Koşul 18		Student's T3 (df=3)	Student's T3 (df=3)
Koşul 19		N(0,3)	N(0,5)
Koşul 20		N(0,3)	X^2 (df=3)
Koşul 21		N(0,3)	Student's T3 (df=3)
Koşul 22		X^2 (df=3)	N(0,5)
Koşul 23	N=1000	X^2 (df=3)	X^2 (df=3)
Koşul 24		X^2 (df=3)	Student's T3 (df=3)
Koşul 25		Student's T3 (df=3)	N(0,5)
Koşul 26		Student's T3 (df=3)	X^2 (df=3)
Koşul 27		Student's T3 (df=3)	Student's T3 (df=3)

Model

Bu çalışma bireylerin zamanla puanlarında artış olacağı yani bir büyüme gösterdikleri düşünülen bir testin simülasyonu şeklinde kurgulanmıştır. Bu çalışmada, belirtilen çerçeve göz önünde bulundurulmuş ve çok düzeyli denkleme ait parametreler belirlenmiş ve bu parametrelere bağlı olarak veriler türetilmiştir. Çok düzeyli kantil regresyon modeli analizinde katsayıların belirtilmesine izin verir. Rasgele kesişmeler tekdüze (değiştirilebilir) bir korelasyon yapısını ifade ederken, rasgele eğimler korelasyon yapısının yaşa bağlı olmasına izin verir ve bu, tekrarlanan ölçüm verileri için daha gerçekçi bir yapı sunabilir. Her ne kadar teorik olarak mümkün olsa da, çok düzeyli kantil regresyon yöntemi analizinde rasgele etkiler modellerinde rasgele eğimlere izin vermek karmaşıktır ve model uyumunda yakınsama sorunlarına yol açabilir (Tzavidis ve diğerleri, 2016). Bu çalışmada, rasgele eğim ve rasgele kesişim modelinin paket programlarında yakınsama sorunu ile karşılaşılmaması için optimizasyon kodları kullanılmış olup simülasyonda olarak ise rasgele kesişim ve rasgele eğim modeline göre modelleme ve analizler gerçekleştirilmiştir.

Parametrelerin Belirlenmesi

Parametrelerin seçiminde literatürde kantil regresyon ve çok düzeyli kantil regresyon ile yapılan simülasyon çalışmaları incelendiğinde belirlenen dağılımlar ve katsayılar göz önünde bulundurulmuştur. Geraci ve Bottai (2007) tarafından gerçekleştirilen simülasyon çalışmasında dağılımlar sırasıyla, $N(0,1)$, Student-t(3) ve X^2 (df=3) dağılımı olarak ele alınmıştır. Liu ve Bottai (2009) tarafından gerçekleştirilen simülasyon çalışmasında ise hata terimlerinin dağılımları sırasıyla; normal, Student-t(3) ve X^2 (df=3) dağılımı olarak tasarlanmıştır. Galvao (2011) tarafından gerçekleştirilen ekonomi alanındaki tek düzeyli kantil regresyon üzerine yapılan simülasyon çalışmasında hata dağılımı, normal dağılım ve Student-t(3) dağılımları kullanılarak incelenmiştir. Tzavidis vd. (2016) yaptıkları çalışmada düzey 1 hata terimlerinin dağılımını $N(0,5)$ ve düzey 2 hata terimlerinin dağılımını ise $N(0,3)$ olarak belirleyen bir çalışma gerçekleştirmişlerdir. Benzer şekilde aynı çalışmanın bir diğer simülasyon koşulu ise Düzey 1 ve Düzey 2 hata terimlerinin dağılımlarının Student t-3 dağılımına sahip olması şeklinde planlanmıştır. Yapılan simülasyon çalışmaları incelendiğinde özellikle sağlık alanında yapılan çalışmalarda (Geraci & Bottai, 2007; Liu & Bottai, 2009; Tzavidis ve diğerleri, 2016) konum kaydırma modeli ile simüle edilen veri yapılarının incelendiği görülmüştür. Literatürde yapılan çalışmalar da göz önünde bulundurulduğunda eğitim alanındaki veri yapılarının modellenmesine uygun şekilde kullanılacak rasgele eğitim ve rasgele kesişim modeli kullanılarak simülasyon verileri üretilmiştir. Çalışmamız kapsamında, Düzey 1 hata dağılımları için $N(0,5)$ / Student's T3 (df=3) / X^2 (df=3) ve Düzey 2 hata terimlerinin dağılımları içinse $N(0,3)$ / Student's T3 (df=3) / X^2 (df=3) olarak belirlenmiştir.

Normal Dağılım

Düzey 1 ve Düzey 2 hata terimlerinin dağılımı normal dağılımın ortalama ve standart sapma değerleri üzerinden $N(\text{ortalama}, \text{standart sapma})$ olarak ifade edilmiştir. Düzey 1 hata terimlerinin dağılımı için $N(0,5)$ ve Düzey 2 hata terimlerinin dağılımı için ise

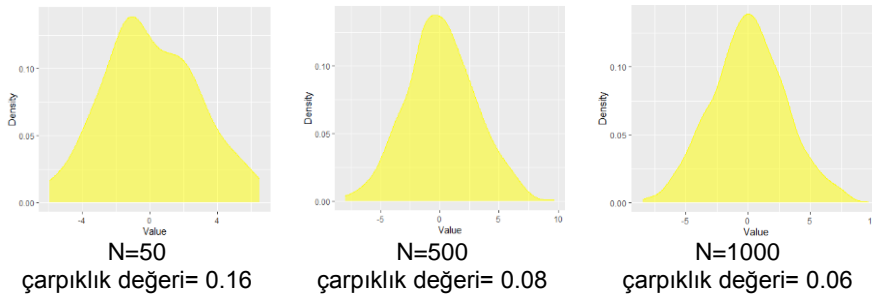
$N(0,3)$ parametreleri belirlenmiştir. Düzey 1 hata terimlerinin dağılımların oluşturulmasına yönelik analiz kodları aşağıdaki gibidir:

```
Nd_3 <- rnorm(n, mean = 0, sd = sqrt(9))
```

Simülasyon koşullarında kullanılan Düzey 1 hata dağılımlarına ait grafikler ve çarpıklık değerleri, örneklem büyüklüğüne göre Şekil 19'da gösterilmiştir.

Şekil 19

Düzey 1 Hata Dağılımlarına Ait Grafikler



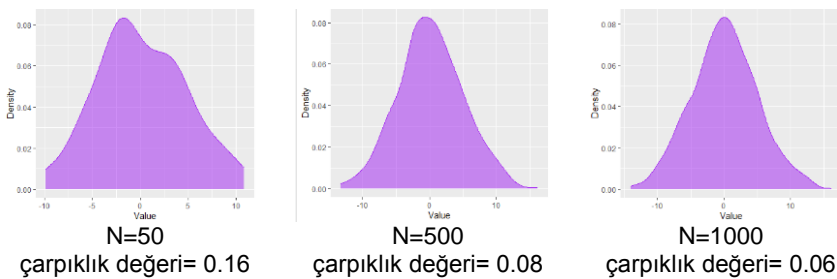
Düzey 2 hata terimlerinin dağılımların oluşturulmasına yönelik analiz kodları aşağıdaki gibidir:

```
Nd_5 <- rnorm(n, mean = 0, sd = sqrt(25))
```

Simülasyon koşullarında kullanılan Düzey 2 hata dağılımlarına ait grafikler ve çarpıklık değerleri, $N=50$, $N=500$ ve $N=1000$ örneklem büyüklüğüne göre Şekil 20'de gösterilmiştir.

Şekil 20

Düzey 2 Hata Dağılımlarına Ait Grafikler



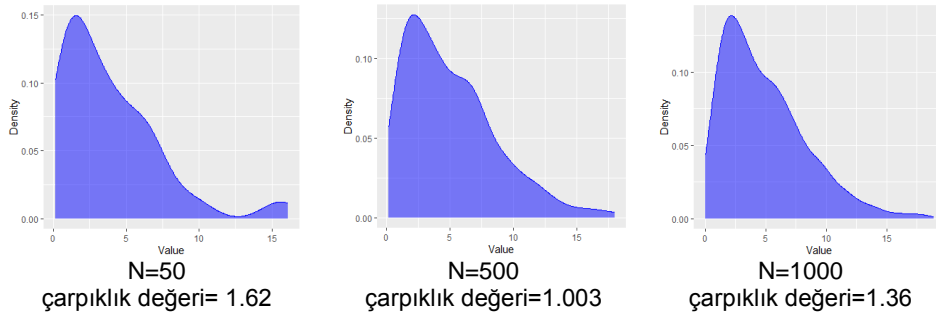
χ^2 (Ki kare) Dağılımı

Aşırı dağılımlı çarpık verilerin temsil edilmesinde kullanılan Kikare Dağılımı istatistikte yaygın olarak kullanılan sürekli bir olasılık dağılımıdır. Düzey 1 ve Düzey 2 hata terimlerinin normal dağılımdan farklı bir yapı göstermesi amacıyla serbestlik derecesi (df=3) olarak kullanılarak oluşturulan bu dağılıma ait analiz kodu aşağıdaki gibidir:

```
kikare <- rchisq(n, df = 3, ncp = 2)
```

Şekil 21

χ^2 (Ki kare) Hata Dağılımlarına Ait Grafikler



Simülasyon koşullarında kullanılan hata dağılımlarına ait grafikler ve çarpıklık değerleri, örneklem büyüklüğüne göre Şekil 21’de gösterilmiştir.

Student T Dağılımı

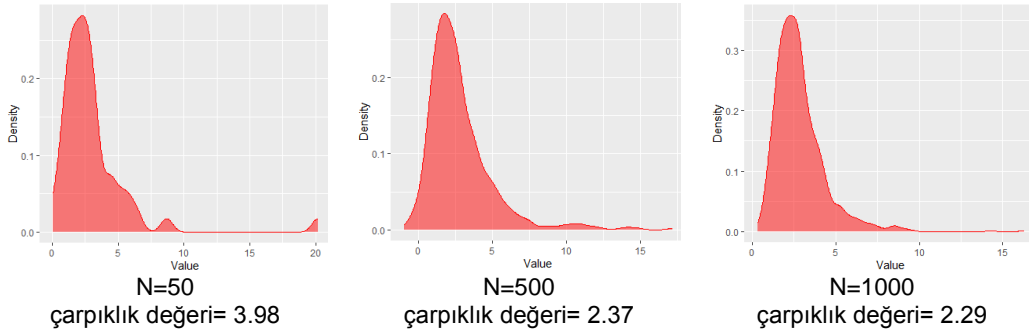
Koşullu dağılımların çarpıklığının modellenmesinde (dağılımın şekli, ölçeği ve yeri belirlenerek) kullanılan bu dağılım Düzey 1 ve Düzey 2 hata terimlerinin normal dağılımdan farklı bir yapı göstermesi amacıyla serbestlik derecesi (df=3) olarak kullanılarak oluşturulan bu dağılıma ait analiz kodu aşağıdaki gibidir:

```
St3 <- rt(n, df = 3, ncp = 2)
```

Simülasyon koşullarında kullanılan hata dağılımlarına ait grafikler ve çarpıklık değerleri, örneklem büyüklüğüne göre Şekil 22’de gösterilmiştir.

Şekil 22

Student T Hata Dağılımlarına Ait Grafikler



Verilerin Analizi

Monte Carlo çalışmalarında yordanan parametrelerin kesinliğinin genellikle yinleme sayısı ile artması beklenir. Bunun nedeni, çalıştırılan tekrar (replikasyon) sayısı arttıkça örneklem boyutunun artması ve bunun da daha doğru tahminleri sağlamasıdır. Bir Monte Carlo çalışmasında araştırmacılar, belirli bir istatistiksel modelin veya hipotezin davranışını simüle eden veri kümeleri oluşturmak için rasgele örnekleme kullanırlar ve ilgili parametreler daha sonra simüle edilmiş her bir veri setinden yordanan sonuçlar bilinen değerlerle karşılaştırılır. Yalnızca birkaç yinleme kullanılırsa, yordanan parametreler, kesin olmayan sonuçlara yol açabilen önemli örnekleme hatasına maruz kalabilir. Ancak yinleme sayısı arttıkça, tüm simülasyonlardaki tahminlerin ortalaması, parametrelerin gerçek değerine yaklaşmalıdır ve bu da daha kesin tahminler sağlar (Robert & Caselle, 2004; Spanier ve diğerleri, 2010; Feinberg & Rubright, 2016). Diğer kantil regresyon simülasyon çalışmalarında kullanıldığı gibi (Geraci & Bottai, 2007; Liu & Bottai, 2009; Tzavidis ve diğerleri, 2016), bu çalışmada da replikasyon sayısının 100 olarak belirlenmesinin yeterli olacağı düşünülmektedir. Analiz edilen tüm koşullar için 100 replikasyon sonucu elde edilmiş simülasyon veri setindeki gerçek gözlenen değer ile analiz sonucunda kestirilen değerler karşılaştırılarak yanlılık (BIAS) ve normalleştirilmiş hata kareleri ortalamasının karekökü (Normalized Root Mean Square Error - NRMSE) değerleri hesaplatılmıştır.

$$Yanlılık = \frac{\sum_{n=1}^N (\hat{Y}_i - Y_i)}{N} \quad (63)$$

$$NRMSE = \left(\sqrt{\frac{\sum_{n=1}^N (\hat{Y}_i - Y_i)^2}{N}} \right) \cdot \frac{1}{\bar{Y}} \quad (64)$$

olarak gösterilir. Burada;

\hat{Y}_i : Bağımlı değişkene ait kestirilen değer

Y_i : Bağımlı değişkene ait gözlenen (gerçek) değer

\bar{Y} : Bağımlı değişkenin ortalama değeri

N : Örneklem büyüklüğü olarak gösterilmektedir.

Son olarak üçüncü araştırma problemimiz kapsamında, çalışmada kullanılan denklem yardımıyla türetilen 100 replikasyonlu veri setlerinin, çok düzeyli boylamsal regresyon analizi ve çok düzeyli kantil regresyon analizinin farklı kantil düzeylerinden elde edilen ortalama kesişim ve eğim katsayılarının karşılaştırılması amacıyla Wald testi uygulanmıştır. Wald değeri, iki katsayı arasındaki farkın önemini değerlendirmek için kullanılan istatistiksel bir test istatistiğidir. Kantil regresyon bağlamında Wald değeri, farklı kantil seviyelerinde elde edilen katsayıların birbirinden önemli ölçüde farklı olup olmadığını belirlemesinde kullanılmıştır. Her bir analiz yöntemi ile elde edilen katsayılar hem medyan ($Q(50)$) arasında hem de her kantil düzeyinin $(1 - Q(p))$ düzeyinden elde edilen katsayılar karşılaştırılmıştır. Bu karşılaştırmalarda kullanılan, aynı ortak değişkene karşılık gelen ancak farklı kantil düzeyleri boyunca H_0 hipotez testinin gerçekleştirilebilmesi için (65) numaralı denklemlerle ifade edilen Wald istatistiğine ait eşitlik;

$$\text{Wald istatistiği} = \frac{(\hat{\beta}_j^{(p)} - \hat{\beta}_j^{(q)})^2}{\hat{\sigma}_{\hat{\beta}_j^{(p)} - \hat{\beta}_j^{(q)}}^2} \quad (65)$$

olarak gösterilir. Bu eşitlikte p ve q farklı kantil düzeyleri olmak üzere;

$\hat{\beta}_j^{(p)}$: p . kantil düzeyinde kestirilen regresyon katsayısı

$\hat{\beta}_j^{(q)}$: q. kantil düzeyinde kestirilen regresyon katsayısı

$\hat{\sigma}^2$: $\hat{\beta}_j^{(p)}$ ve $\hat{\beta}_j^{(q)}$ farkının tahmini varyansdır (Hao & Naiman, 2007).

Sıfır hipotezi altında, Wald istatistiği bir serbestlik dereceli yaklaşık bir X^2 dağılımına sahiptir (Hao & Naiman, 2007). Bu çalışma kapsamında Wald istatistiği ile elde edilen değer, $p < 0.05$ anlamlılık düzeyinde X^2 tablo değeri ile karşılaştırılmıştır. Hesaplanan değer tablodaki değerden büyük olması halinde yokluk hipotezi reddedilerek ilgili parametrelerin anlamlı olduğu yorumu yapılmıştır.

Bölüm 4

Bulgular, Yorumlar ve Tartışma

Bu bölümde simülasyon çalışması ile elde edilen veri setlerinin analiz sonuçlarına değinilerek, araştırmanın alt problemlerine ilişkin bulgular ve yorumlar sunulmaktadır. Araştırmanın amacı kapsamında simüle edilen veri setleri, “nlme” ve “qrlmm” paketlerinde analiz edilerek bulgular üç alt problemde sunulmuştur. Gerçekleştirilen analizlerle elde edilen model veri uyumu değerleri, yanlılık, mutlak hata değerleri ve kesişim ile eğim katsayılarının farklılaşmasına ilişkin bulgular alt problemlerdeki tablo ve grafiklerle sunulmuştur.

Birinci Alt Probleme İlişkin Bulgular ve Yorum

Simüle edilmiş boylamsal veri setlerinde hata düzeylerinin farklı dağılımlarına göre (Normal- Kikare- Student t(3)), her bir örneklem büyüklüğü için (50; 500; 1000), çok düzeyli klasik regresyon analizinde ve kantil regresyonun farklı kantil düzeylerinde (0.10; 0.25; 0.50; 0.75; 0.90) **model veri uyumu değerleri** (AIC-BIC-logLike) nasıl değişmektedir?

Birinci alt probleme ilişkin yapılan analizler sonucunda, Düzey 1 ve Düzey 2 hata dağılımların farklılıklarına göre; farklı örneklem (50; 500; 1000) ve farklı kantil düzeylerinde (0.10; 0.25; 0.50; 0.75; 0.90) elde edilen AIC değerleri Tablo 14’te verilmiştir. Tablo 14 incelendiğinde, Düzey 1 ve Düzey 2 hata dağılımların farklılıklarına göre farklı örneklem büyüklüklerindeki çok düzeyli regresyon ve çok düzeyli kantil regresyon analiz sonuçlarından elde edilen AIC değerleri görülmektedir. Tablo 14’ e göre N=50 için AIC değerinin 799.637 ile 1036.083 arasında; N=500 için 7876.182 ile 10573.429 ve N=1000 için 13747.067 ile 20667.104 arasında değişmektedir. Örneklem büyüklüğü arttıkça AIC değerlerinin arttığı, her iki hata düzeyinde de normal dağılım olduğu durumda çok düzeyli regresyon analizi yöntemi ile kantil regresyondan elde edilen medyana ait AIC değerlerinin birbirine yakın olduğu görülmüştür. AIC değerinin çok düzeyli regresyon analizi için en

düşük olduğu dağılım ise her iki hata düzeyinin de çok çarpık olduğu durumda ([T,T]) gözlenmiştir.

Tablo 14

Düzye 1 ve Düzye 2 Hata Dağılımların Farklılıklarına Göre Farklı Örneklem Büyüklüklerindeki AIC Değerleri

		AIC Değerleri					
Düzye 1 ve Düzye 2 Hatalar	Örneklem	Çok düzeyli regresyon analizi yöntemi	q50	q10	q25	q75	q90
[N,N]	50	963.567	992.978	1035.125	1006.695	1006.854	1036.084
	500	9624.872	9870.668	10318.37	10011.22	10016.26	10322.74
	1000	19267.67	19768.08	20654.91	20053.45	20059.8	20667.1
[N,kk]	50	903.501	924.922	938.631	919.177	957.836	1002.908
	500	8994.157	9157.478	9294.311	9103.906	9519.293	9986.085
	1000	18025.71	18359	18626.48	18230.14	19064.1	20008.22
[N,T]	50	963.567	992.978	1035.125	1006.695	1006.854	1036.084
	500	8374.338	8260.524	8590.484	8259.208	8652.12	9245.439
	1000	15042.04	15263.34	16337.98	15486.1	15842.89	16862.81
[kk,N]	50	891.378	905.564	897.383	891.569	948.078	997.001
	500	9753.133	10041.25	10414.32	10142.57	10236.5	10573.43
	1000	19479.96	20050.06	20799.56	20248.99	20450.86	21125.57
[kk,kk]	50	822.016	853.413	877.136	856.275	887.375	943.688
	500	9143.156	9397.316	9425.007	9304.636	9762.602	10259.88
	1000	18301.76	18817.46	18871.16	18619.73	19547.97	20534
[kk,T]	50	976.083	1010.002	1043.278	1018.511	1029.883	1063.712
	500	8474.829	8533.461	8746.4	8521.93	8917.726	9571.99
	1000	15366.15	15953.6	16857.22	16185.22	16554.44	17833.89
[T,N]	50	951.935	975.863	1014.811	988.059	994.269	1023.92
	500	9578.921	9746.683	10153.99	9864.803	9944.343	10292.05
	1000	18852.63	19215.62	20062.67	19487.07	19545.47	20159.41
[T,kk]	50	891.378	905.564	897.383	891.569	948.078	997.001
	500	8964.484	9012.563	8937.415	8855.594	9461.467	9991.161
	1000	17482.44	17567.59	17417.72	17260.84	18449.6	19456.17
[T,T]	50	799.637	799.614	806.119	789.51	845.153	899.772
	500	8210.368	7986.14	8057.955	7876.183	8458.045	9038.349
	1000	13951.57	13762.64	14291.56	13747.07	14457.46	15329.85

Ayrıca dağılımın normal olduğu durumlarda ([N,N]) kantil düzeylerinden elde edilen AIC değerleri birbirine çok yakın hesaplanmıştır. Düzye 2 dağılımının çok çarpık olduğu

durumda ([N,T]) küçük örnekleme AIC değerleri dağılımın normal olduğu durumdaki ([N,N]) AIC değerleri ile aynı hesaplanmıştır.

Bu çalışmada kullanılan veri yapılarından elde edilen BIC değerleri Tablo 15' te verilmiştir.

Tablo 15

Düzye 1 ve Düzye 2 Hata Dağılımların Farklılıklarına Göre Farklı Örnekleme

Büyüküklerindeki BIC Değerleri

		BIC Değerleri					
Düzye 1 ve Düzye 2 Hatalar	Örnekleme	Çok düzeyli regresyon analizi yöntemi	q50	q10	q25	q75	q90
[N,N]	50	984.5	1014.053	1056.199	1027.77	1027.929	1057.158
	500	9662.05	9907.861	10355.57	10048.41	10053.45	10359.93
	1000	19309.7	19810.12	20696.96	20095.49	20101.85	20709.15
[N,kk]	50	924.434	945.997	959.705	940.251	978.911	1023.982
	500	9031.336	9194.67	9331.503	9141.098	9556.486	10023.28
	1000	18067.75	18401.04	18668.53	18272.18	19106.14	20050.26
[N,T]	50	984.5	1014.053	1056.199	1027.77	1027.929	1057.158
	500	8411.516	8297.717	8627.677	8296.401	8689.313	9282.632
	1000	15084.08	15305.38	16380.02	15528.14	15884.93	16904.86
[kk,N]	50	912.311	926.639	918.457	912.644	969.152	1018.076
	500	9790.311	10078.44	10451.51	10179.76	10273.69	10610.62
	1000	19522	20092.11	20841.61	20291.03	20492.9	21167.61
[kk,kk]	50	842.949	874.487	898.211	877.349	908.449	964.763
	500	9180.334	9434.509	9462.2	9341.828	9799.795	10297.08
	1000	18343.8	18859.51	18913.21	18661.77	19590.01	20576.05
[kk,T]	50	997.016	1031.076	1064.353	1039.585	1050.958	1084.787
	500	8512.008	8570.654	8783.592	8559.123	8954.919	9609.182
	1000	15408.19	15995.64	16899.27	16227.27	16596.48	17875.93
[T,N]	50	972.868	996.938	1035.886	1009.134	1015.343	1044.994
	500	9616.099	9783.875	10191.18	9901.995	9981.535	10329.24
	1000	18894.66	19257.66	20104.72	19529.12	19587.52	20201.45
[T,kk]	50	912.311	926.639	918.457	912.644	969.152	1018.076
	500	9001.663	9049.756	8974.608	8892.786	9498.659	10028.35
	1000	17524.48	17609.63	17459.77	17302.89	18491.64	19498.21
[T,T]	50	820.57	820.689	827.194	810.585	866.228	920.846
	500	8247.546	8023.332	8095.148	7913.375	8495.237	9075.541
	1000	13993.6	13804.68	14333.61	13789.11	14499.5	15371.89

Tablo 15'te verilen BIC deęerleri incelendięinde, $N=50$ için 820.570 ile 1084.786 arasında; $N=500$ için 7913.375 ile 10610.622 ve $N=1000$ için 15084.077 ile 21167.612 arasında deęiřtięi görölmektedir. BIC deęerleri karřılařtırıldıęında, AIC deęerleri le benzer řekilde çok düzeyli regresyon analizi için tüm örnekleme büyüklüklerinde, her iki düzey hataların çok çarpık olması durumunda ([T,T]) en küçük deęerler elde edilmiřtir. Ayrıca bu daęılımda tüm çok düzeyli regresyondan elde edilen BIC deęerleri ile tüm kantil düzeylerinden elde edilen BIC deęerlerinin birbirlerine çok yakın hesaplandıęı görölmektedir. Özellikle $N=1000$ örnekleme büyüklüęünde, her iki düzey hata daęılımının normal olması durumundaki ([N,N]) kantillerden elde edilen BIC deęerleri arasındaki farklılařma ile her iki düzeyin çok çarpık olması durumunda ([T,T]) kantillerden elde edilen BIC deęerleri arasındaki fark göz önünde bulundurulduęunda, çok çarpık daęılımların daha iyi model veri uyumu gösterdięi ve ayrıca tüm kantil düzeylerinde bu uyumu sürdürdüęü söylenebilir. Daęılımın çarpıklıkları normalden çok çarpıęa doęru arttıęı (sırasıyla [N,N]; [kk,kk] ve [T,T] incelendięinde), tüm örnekleme büyüklüklerinde ve tüm kantil düzeylerinde, BIC deęerleri azalma eğilimi göstermiřtir. Dolayısıyla her iki hata düzeyi daęılımında da çarpıklık arttıęı model veri uyumunun iyileřiđi görölmüřtür.

Bu çalıřmada kullanılan veri yapılarından elde edilen log-Like deęerleri Tablo 16'da verilmiřtir. Tablo 16'da yer alan loglike deęerleri incelendięinde, $N=50$ örnekleme büyüklüęü için -524.856 ile -392.818 arasında; $N=500$ için - 5279.714 ile - 3931.091 ve $N=1000$ için ise -10555.783 ile -6866.533 arasında deęiřtięi görölmektedir.

Maksimum olabilirlik sürecinin bir parçası olarak hesaplanan logLike deęerleri incelendięinde, AIC ve BIC deęerlerinin yorumları ile benzer yorumlar elde edilebilmektedir. En iyi model veri uyumu tüm örneklemlerde her iki düzey hata daęılımlarının da çok çarpık olduęu [T,T] daęılımında gözlenmiřtir. Daęılımın çarpıklıęı arttıęı loglike deęerlerinin arttıęı dolayısıyla daha iyi model veri uyumu sağladıęı yorumu yapılabilir.

Tablo 16*Düzyey 1 ve Düzyey 2 Hata Dağılımların Farklılıklarına Göre Farklı Örneklem**Büyüküklerindeki logLike Değerleri*

Düzyey 1 ve Düzyey 2 Hatalar	Örneklem	Çok düzyeyli regresyon analizi yöntemi	Log-Like Değerleri				
			q50	q10	q25	q75	q90
[N,N]	50	-474.784	-489.489	-510.562	-496.348	-496.427	-511.042
	500	-4805.44	-4928.33	-5152.19	-4998.61	-5001.13	-5154.37
	1000	-9626.83	-9877.04	-10320.5	-10019.7	-10022.9	-10326.6
[N,kk]	50	-444.75	-455.461	-462.315	-452.588	-471.918	-494.454
	500	-4490.08	-4571.74	-4640.16	-4544.95	-4752.65	-4986.04
	1000	-9005.86	-9172.5	-9306.24	-9108.07	-9525.05	-9997.11
[N,T]	50	-474.784	-489.489	-510.562	-496.348	-496.427	-511.042
	500	-4180.17	-4123.26	-4288.24	-4122.6	-4319.06	-4615.72
	1000	-7514.02	-7624.67	-8161.99	-7736.05	-7914.44	-8424.41
[kk,N]	50	-438.689	-445.782	-441.691	-438.785	-467.039	-491.501
	500	-4869.57	-5013.63	-5200.16	-5064.29	-5111.25	-5279.72
	1000	-9732.98	-10018	-10392.8	-10117.5	-10218.4	-10555.8
[kk,kk]	50	-404.008	-419.706	-431.568	-421.137	-436.687	-464.844
	500	-4564.58	-4691.66	-4705.5	-4645.32	-4874.3	-5122.94
	1000	-9143.88	-9401.73	-9428.58	-9302.86	-9766.98	-10260
[kk,T]	50	-481.041	-498.001	-514.639	-502.256	-507.942	-524.856
	500	-4230.42	-4259.73	-4366.2	-4253.97	-4451.86	-4779
	1000	-7676.08	-7969.8	-8421.61	-8085.61	-8270.22	-8909.94
[T,N]	50	-468.968	-480.932	-500.406	-487.03	-490.134	-504.96
	500	-4782.46	-4866.34	-5069.99	-4925.4	-4965.17	-5139.03
	1000	-9419.31	-9600.81	-10024.3	-9736.54	-9765.74	-10072.7
[T,kk]	50	-438.689	-445.782	-441.691	-438.785	-467.039	-491.501
	500	-4475.24	-4499.28	-4461.71	-4420.8	-4723.73	-4988.58
	1000	-8734.22	-8776.79	-8701.86	-8623.42	-9217.8	-9721.08
[T,T]	50	-392.819	-392.807	-396.06	-387.755	-415.577	-442.886
	500	-4098.18	-3986.07	-4021.98	-3931.09	-4222.02	-4512.17
	1000	-6968.78	-6874.32	-7138.78	-6866.53	-7221.73	-7657.92

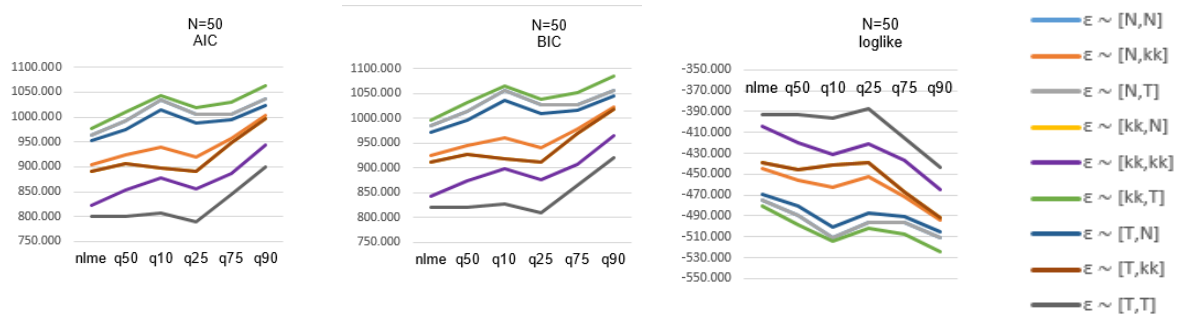
Düzyey 1 dağılımının az çarpık olduđu durumda (X^2 dağılımı) düzyey 2 hata dağılımlarının normal olduđu durumdan az çarpık olduđu duruma geçerken, yani çarpıklığının arttığı durumda loglike değerinin azaldığı görülmüştür. Ancak aynı dağılımdan düzyey 2 hataların çok çarpıklaştığı duruma geçerken loglike değerinin artış

gösterdiği görülmüştür. Dolayısıyla, Düzey 2 hataların normal ve çok çarpık olduğu dağılımlar (sırasıyla $[kk, N]$ ve $[kk, T]$), Düzey 1 hata dağılımlarının az çarpık olduğu durumlara göre ($[kk, kk]$) daha iyi model veri uyumu göstermektedir yorumu yapılabilir. AIC, BIC değerleriyle benzer şekilde tüm dağılımlar ve tüm örneklem büyüklüklerinde q90 kantil düzeyi diğerlerine göre daha zayıf model veri uyumu göstermiştir. Tablo 14, Tablo 15 ve Tablo 16, ilerleyen kısımlarda görselleştirilen şekillerle birlikte değerlendirilmiştir.

Araştırma problemimiz kapsamında ayrıca farklı örneklem büyüklüklerine göre model veri uyumu değerlerine ilişkin analiz sonuçları ayrı ayrı incelenmiştir. Buna göre $N=50$ örneklem büyüklüğünde Düzey 1 ve Düzey 2 hatalarının farklı dağılımlarına göre üretilen verilerin analizlerine ait model veri uyumu bulgularına ait grafikler Şekil 23'te verilmiştir.

Şekil 23

N=50 Örneklem Büyüklüğü İçin Model Veri Uyumu Değerleri



* $[N, N]$ ve $[N, T]$ ile $[T, kk]$ ve $[kk, N]$ hata dağılımlarına ait AIC, BIC ve loglike değerleri eşit hesaplandığı için Şekil 20'de 7 grafik değeri görülmektedir.

**Aşağıdaki şekillerle ilgili açıklamalarda harflerin parantez içindeki sırası Düzey 1 ve Düzey 2 hata terimlerinin dağılımlarını göstermek üzere, N, T ve kk harflerine göre:

$[N, N]$: Normal Dağılımı

$[N, T]$: Düzey 2 hata terimleri dağılımının çok çarpık olduğu durumlar

$[kk, N]$: Düzey 1 hata terimlerinin az çarpık, Düzey 2 hata terimlerinin normal dağılım olduğu durumlar

$[T, T]$: Düzey 1 ve 2 hata terimleri dağılımının çok çarpık olduğu durumlar şeklinde ifade edilmiştir.

Tablo 14, Tablo 15 ile Tablo 16 ve Şekil 23 birlikte incelendiğinde, $N=50$ örneklem büyüklüğünde tüm kantil düzeylerinde (q10, q25, q50, q75, q90) en iyi model veri uyumunu Düzey 1 ve 2 hata terimleri dağılımının çok çarpık olduğu $[T, T]$ hata dağılımı ile

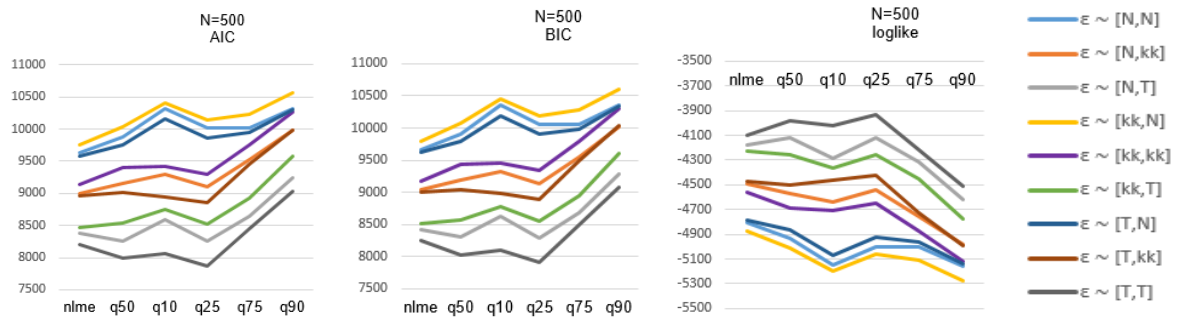
q25 kantil düzeyinde gösterdiği, model veri uyumu diğerlerine kıyasla en az olduğu modelin ise Düzey 1 hata terimlerinin az çarpık, Düzey 2 hata terimlerinin çok çarpık olduğu [kk,T] ile q90 kantil düzeyinde gösterdiği görülmektedir. Buna göre, çarpıklığı en fazla olan dağılımın en iyi model veri uyumu gösterdiği yorumu yapılabilir. Ayrıca Düzey 1 hataların normal dağılım gösterdiği (N(0,3) dağılımı) koşullar kendi içinde değerlendirilirse, en iyi model veri uyumunun Düzey 2 hata terimlerinin az çarpık olduğu durumda ([N,kk] dağılımı) ve q25 kantil düzeyinde olduğu görülmüştür. Düzey 2 hata terimlerinin normal ve çok çarpık olduğu durumlarda (sırasıyla [N,N] ve [N,T] dağılımları) hata dağılımlarına sahip veri setlerinin model veri uyum değerlerinin aynı ve az çarpık olan duruma göre ([N,kk] dağılımı) daha zayıf olduğu görülmektedir. Düzey 1 hataların az çarpık olduğu (X^2 dağılımı) koşullar kendi içinde değerlendirilirse, en iyi model veri uyumunun Düzey 2 hataların az çarpık olduğu ([kk,kk] dağılımı) durumda ve q50 kantil düzeyinde, en zayıf model veri uyumunun Düzey 2 hata dağılımının çok çarpık olduğu ([kk,T] dağılımı) durumda ve q90 kantil düzeyinde gösterdiği görülmektedir. Düzey 1 hataların çok çarpık olduğu (Student's t(3) dağılımı) koşullar kendi içinde değerlendirilirse, en iyi model veri uyumunun Düzey 2 hataların çok çarpık olduğu ([T,T] dağılımı) durumda q25 kantil düzeyinde, en zayıf model veri uyumunun Düzey 2 hataların normal olduğu ([T,N] dağılımı) durumda q90 kantil düzeyinde gösterdiği görülmektedir. Son olarak Düzey 2 hataların normal (N(0,5) dağılımı) olduğu koşullar kendi içinde değerlendirilirse, en iyi model veri uyumunun Düzey 1 hataların az çarpık ([kk,N] dağılımı) olduğu durumda ve q25 kantil düzeyinde, en zayıf model veri uyumunun ise Düzey 1 hataların normal ([N,N] dağılımı) olduğu durumda ve q90 kantil düzeyinde gösterdiği görülmektedir. Düzey 2 hataların az çarpık (X^2 dağılımı) olduğu koşullar kendi içinde değerlendirilirse, en iyi model veri uyumunun Düzey 1 hataların az çarpık ([kk,kk] dağılımı) olduğu durumda ve q50 kantil düzeyinde, en zayıf model veri uyumunun Düzey 1 hataların normal olduğu ([N,kk] dağılımı) durumda ve q90 kantil düzeyinde gösterdiği görülmüştür. Düzey 2 hataların çok çarpık olduğu (Student's t(3) dağılımı) koşullar kendi içinde değerlendirilirse, en iyi model veri uyumunun Düzey 1 hataların çok çarpık olduğu ([T,T] dağılımı) durumda

ve q25 kantil düzeyinde, en zayıf model veri uyumunun Düzey 1 hataların normal ([N,T] dağılımı) olduğu durumda ve q90 kantil düzeyinde gösterdiği görülmektedir.

N=500 örneklem büyüklüğü için farklı koşullara göre üretilen verilerin analizlerine ait model veri uyumu bulgularının grafikleri Şekil 24'te verilmiştir.

Şekil 24

N=500 Örneklem Büyüklüğü İçin Model Veri Uyumu Değerleri



Tablo 14, Tablo 15 ile Tablo 16 ve Şekil 24 birlikte incelendiğinde N=500 örneklem büyüklüğünde tüm kantil düzeylerinde (q10, q25, q50, q75, q90) en iyi model veri uyumu Düzey 1 ve Düzey 2 hata dağılımlarının çok çarpık ([T,T] dağılımı) ve q25 kantil düzeyinde olduğu, model veri uyumunun diğerlerine kıyasla en az olduğu model ise q90 kantil düzeyinde Düzey 1 hata dağılımının az çarpık, Düzey 2 hata dağılımının normal dağılım gösterdiği ([kk,N] dağılımı) durumda elde edilmiştir.

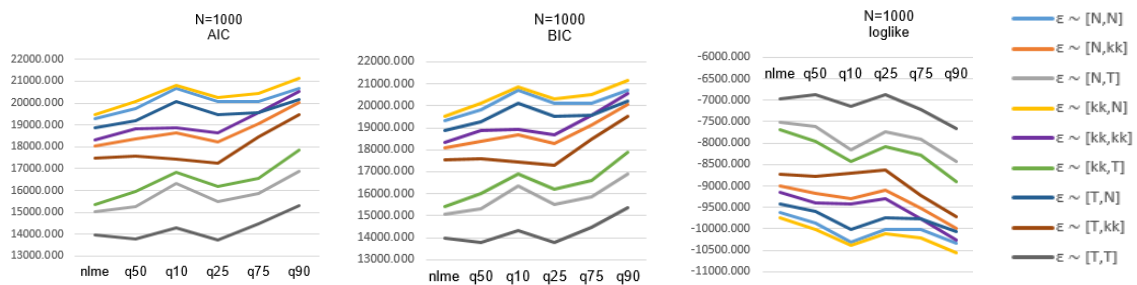
N=500 örneklem büyüklüğünde Düzey 1 hataları kendi içinde değerlendirilirse; Düzey 1 hatalarının normal ($N(0,3)$ dağılımı) olduğu durumda, en iyi model veri uyumunun Düzey 2 hataların çok çarpık olduğu ([N,T] dağılımı) durumda ve q50 kantil düzeyinde gösterdiği görülmüştür. En zayıf model veri uyumunun Düzey 2 hata dağılımının normal ([N,N] dağılımı) ve q90 kantil düzeyinde olduğu görülmüştür. Düzey 1 hatalarının az çarpık (X^2 dağılımı) olduğu durumda, en iyi model veri uyumunun Düzey 2 hata dağılımının çok çarpık ([kk,T] dağılımı) olduğu durumda ve q25 kantil düzeyinde, en zayıf model veri uyumunun Düzey 2 hata dağılımının normal ([kk,N] dağılımı) ve q90 kantil düzeyinde

olduğu durumda gösterdiği görülmüştür. Düzey 1 hatalarının çok çarpık (Student's T3 dağılımı) olduğu durumda, en iyi model veri uyumunun Düzey 2 hataların çok çarpık ([T, T] dağılımı) ve q25 kantil düzeyinde, en zayıf model veri uyumunun yine aynı dağılımda q90 kantil düzeyinde olduğu görülmüştür.

N=500 örneklem büyüklüğünde Düzey 2 hataları kendi içinde değerlendirilirse; Düzey 2 hatalarının normal (N(0,5) dağılımı) olduğu durumda, en iyi model veri uyumunu Düzey 1 hata dağılımlarının çok çarpık olduğu ([T, N] dağılımı) ve q50 kantil düzeyinde, en zayıf model veri uyumunun ise Düzey 1 hata dağılımının az çarpık olduğu ([kk, N] dağılımı) ve q90 kantil düzeyinde gösterdiği görülmüştür. Düzey 2 hatalarının çok çarpık (Student's T3 dağılımı) olduğu durumda, en iyi model veri uyumunun Düzey 1 hata dağılımlarının çok çarpık ([T, T] dağılımı) ve q25 kantil düzeyinde, en zayıf model veri uyumunun ise Düzey 1 hata dağılımlarının az çarpık ([kk, T] dağılımı) ve q90 kantil düzeyinde olduğu görülmüştür. Düzey 2 hatalarının az çarpık (X^2 dağılımı) olduğu durumda, en iyi model veri uyumunun Düzey 1 hata dağılımlarının çok çarpık ([T, kk] dağılımı) ve q25 kantil düzeyinde, en zayıf model veri uyumunun ise Düzey 1 hata dağılımlarının az çarpık ([kk, kk] dağılımı) ve q90 kantil düzeyinde gösterdiği görülmüştür.

Şekil 25

N=1000 Örneklem Büyüklüğü İçin Model Veri Uyumu Değerleri



Tablo 14, Tablo 15 ile Tablo 16 ve Şekil 25 birlikte incelendiğinde, N=1000 örneklem büyüklüğünde tüm kantil düzeylerinde (q10, q25, q50, q75, q90) en iyi model veri uyumunu Düzey 1 ve Düzey 2 hataların çok çarpık ([T, T] dağılımı) ve q25 kantil

düzeyinde göstermiş olup, model veri uyumu diğerlerine kıyasla en az olduğu modelin ise ile q90 kantil düzeyinde Düzey 1 hataların az, Düzey 2 hataların çok çarpık olduğu ([kk,T] dağılımı) durumda elde edilmiştir.

N=1000 örneklem büyüklüğünde Düzey 1 hataları kendi içinde değerlendirilirse; Düzey 1 hatalarının normal (N(0,3) dağılımı) olduğu durumda, en iyi model veri uyumu Düzey 2 hata dağılımının çok çarpık olduğu ([N,T] dağılımı) ve q50 kantil düzeyinde görülmüştür. En zayıf model veri uyumu Düzey 2 hata dağılımlarının normal olduğu ([N,N] dağılımı) ve q90 kantil düzeyinde görülmüştür. Düzey 1 hatalarının az çarpık (X^2 dağılımı) olduğu durumda, en iyi model veri uyumu Düzey 2 hatalarının çok çarpık olduğu ([kk,T] dağılımı) ve q50 kantil düzeyinde, en zayıf model veri uyumunun Düzey 2 hata dağılımlarının normal olduğu ([kk,N] dağılımı) ve q90 kantil düzeyinde görülmüştür. Düzey 1 hatalarının çok çarpık (Student's T3 dağılımı) olduğu durumda, en iyi model veri uyumu Düzey 2 hatalarının çok çarpık olduğu ([T,T] dağılımı) ve q25 kantil düzeyinde, en zayıf model veri uyumu Düzey 1 hataların normal olduğu ([N,T] dağılımı) ve q90 kantil düzeyinde görülmüştür. N=1000 örneklem büyüklüğünde Düzey 2 hataları kendi içinde değerlendirilirse; Düzey 2 hatalarının normal (N(0,5) dağılımı) olduğu durumda, en iyi model veri uyumu Düzey 1 hata dağılımlarının çok çarpık olduğu ([T,N] dağılımı) ve q50 kantil düzeyinde, en zayıf model veri uyumu ise Düzey 1 hata dağılımının az çarpık olduğu ([kk,N] dağılımı) ve q90 kantil düzeyinde görülmüştür. Düzey 2 hatalarının az çarpık (X^2 dağılımı) olduğu durumda, en iyi model veri uyumu Düzey 1 hata dağılımlarının çok çarpık olduğu ([T,N] dağılımı) ve q25 kantil düzeyinde, en zayıf model veri uyumu ise Düzey 1 hata dağılımının az çarpık olduğu ([kk,N] dağılımı) ve q90 kantil düzeyinde görülmüştür. Düzey 2 hatalarının çok çarpık (Student's T3 dağılımı) olduğu durumda, en iyi model veri uyumu Düzey 1 hata dağılımlarının çok çarpık olduğu ([T,T] dağılımı) ve q25 kantil düzeyinde, en zayıf model veri uyumu Düzey 1 hata dağılımlarının az çarpık olduğu ([kk,T] dağılımı) ve q90 kantil düzeyinde görülmüştür.

İkinci Alt Probleme İlişkin Bulgular ve Yorum

Simüle edilmiş boylamsal veri setlerinde, farklı örneklem büyüklüğü ve farklı Düzey 1 ile Düzey 2 hata terimlerinin dağılımlarında, çok düzeyli klasik regresyon analizi ile çok düzeyli kantil regresyon analiz yöntemlerinin performansları nasıldır?

İkinci alt probleme ilişkin yapılan analizler sonucunda, farklı örneklem büyüklüğü ve farklı Düzey 1 ile Düzey 2 hata terimlerinin dağılımlarına göre elde edilen mutlak hata ve yanlılık değerleri Tablo 17’de verilmiştir. Tablo 17 incelendiğinde, elde edilen değerlere göre mutlak hata değerinin 0.026 ila 0.088 arasında değiştiği; yanlılık değerlerinin ise -7.377 ila 8.636 arasında değiştiği görülmektedir. Tablo 17’deki bilgiler aşağıdaki görselleri verilen grafiklerle birlikte açıklanarak ele alınmıştır.

Tablo 17

Düzey 1 ve Düzey 2 Hataların Farklı Dağılımlarına Göre Mutlak Hata ve Yanlılık Değerleri

Hata Dağılımları	Örnekleme	Mutlak Hata						Yanlılık					
		nlme	q50	q10	q25	q75	q90	nlme	q50	q10	q25	q75	q90
[N,N]	50	0.052	0.053	0.079	0.061	0.064	0.086	4.357	0.295	-6.841	-3.480	4.183	7.859
	500	0.053	0.053	0.081	0.061	0.065	0.086	4.687	0.399	-7.121	-3.613	4.390	7.990
	1000	0.053	0.054	0.081	0.062	0.066	0.088	1.283	0.492	-7.152	-3.590	4.558	8.123
[N,kk]	50	0.033	0.034	0.048	0.039	0.040	0.053	2.946	-0.042	-5.230	-2.994	3.313	6.361
	500	0.034	0.034	0.049	0.039	0.041	0.056	7.473	0.070	-5.479	-2.935	3.476	6.933
	1000	0.034	0.034	0.049	0.039	0.041	0.056	5.706	0.094	-5.465	-2.957	3.469	6.937
[N,T]	50	0.052	0.053	0.079	0.061	0.064	0.086	4.357	0.295	-6.841	-3.480	4.183	7.859
	500	0.034	0.034	0.047	0.038	0.040	0.053	3.248	0.282	-4.521	-2.253	2.904	5.478
	1000	0.030	0.030	0.043	0.033	0.036	0.048	6.024	0.366	-4.174	-2.027	2.818	5.090

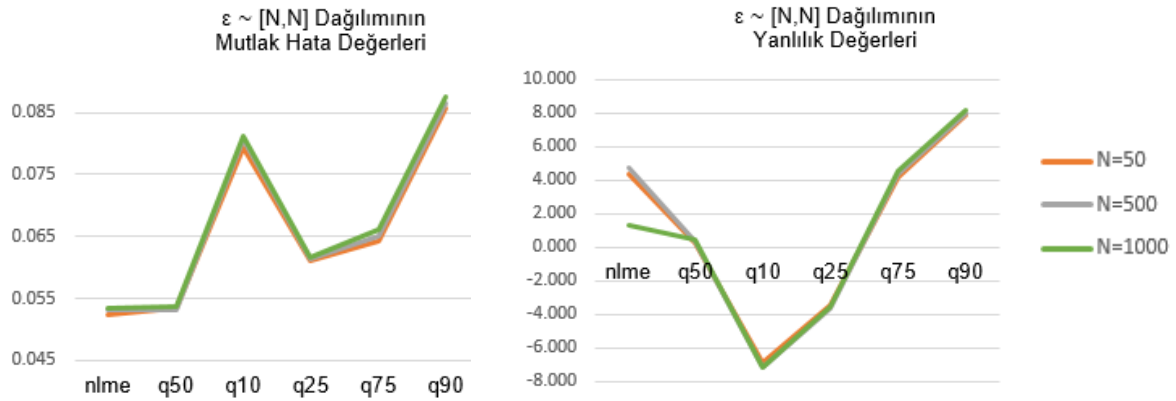
Tablo 17 (devamı)

[kk,N]	50	0.028	0.028	0.038	0.032	0.033	0.045	3.196	-0.167	-4.650	-2.792	2.978	6.249
	500	0.043	0.043	0.064	0.049	0.052	0.070	1.194	0.280	-7.360	-3.852	4.594	8.636
	1000	0.042	0.043	0.064	0.049	0.052	0.070	2.468	0.307	-7.377	-3.847	4.583	8.601
[kk,kk]	50	0.027	0.028	0.037	0.031	0.034	0.048	4.256	0.063	-4.377	-2.584	3.380	6.775
	500	0.029	0.029	0.041	0.034	0.034	0.047	2.283	-0.282	-5.827	-3.480	3.503	7.428
	1000	0.029	0.029	0.041	0.034	0.034	0.048	3.983	-0.205	-5.826	-3.419	3.604	7.578
[kk,T]	50	0.042	0.043	0.062	0.049	0.052	0.069	1.010	0.233	-6.917	-3.646	4.625	8.237
	500	0.028	0.029	0.039	0.032	0.033	0.046	2.074	-0.145	-4.711	-2.721	2.975	6.310
	1000	0.026	0.026	0.036	0.030	0.031	0.042	4.685	-0.102	-4.391	-2.509	2.842	5.966
[T,N]	50	0.042	0.044	0.065	0.050	0.052	0.068	2.571	0.292	-6.577	-3.469	4.027	7.231
	500	0.044	0.044	0.066	0.051	0.054	0.071	1.607	0.332	-6.694	-3.423	4.186	7.649
	1000	0.041	0.041	0.062	0.047	0.051	0.068	5.407	0.448	-6.408	-3.240	4.140	7.450
[T,kk]	50	0.028	0.028	0.038	0.032	0.033	0.045	3.196	-0.167	-4.650	-2.792	2.978	6.249
	500	0.029	0.029	0.040	0.033	0.034	0.047	3.117	-0.223	-4.833	-2.827	3.024	6.525
	1000	0.026	0.026	0.036	0.030	0.031	0.044	3.817	-0.165	-4.493	-2.626	2.899	6.203
[T,T]	50	0.026	0.026	0.035	0.029	0.031	0.042	6.849	-0.068	-3.597	-2.044	2.411	5.025
	500	0.027	0.027	0.036	0.030	0.031	0.041	1.085	-0.054	-3.652	-2.019	2.227	4.855
	1000	0.020	0.020	0.027	0.022	0.024	0.032	1.204	0.213	-2.872	-1.475	2.029	3.880

Farklı Düzey 1 ve Düzey 2 hata terimlerinin dağılımlarının farklı örneklem büyüklüklerine göre oluşturulan veri setlerindeki mutlak hata ve yanlılık değerleri ayrı ayrı incelenmiştir. Düzey 1 ve Düzey 2 hata terimlerinin normal dağılım gösterdiği ([N,N] dağılımı) veri yapıları için yapılan analizler sonucu elde edilen mutlak hata ve yanlılık bulgularına ait grafikler Şekil 26'da verilmiştir.

Şekil 26

Düzyey 1 ve Düzyey 2 Hata Terimlerinin Normal Dağılım Gösterdiği ($\epsilon \sim [N,N]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri



*N=500 ve N=1000 için mutlak hata değerleri birbirine çok yakın değerler olarak hesaplandığı için mutlak hata değerleri grafiğinde 2 adet grafik bulunmaktadır.

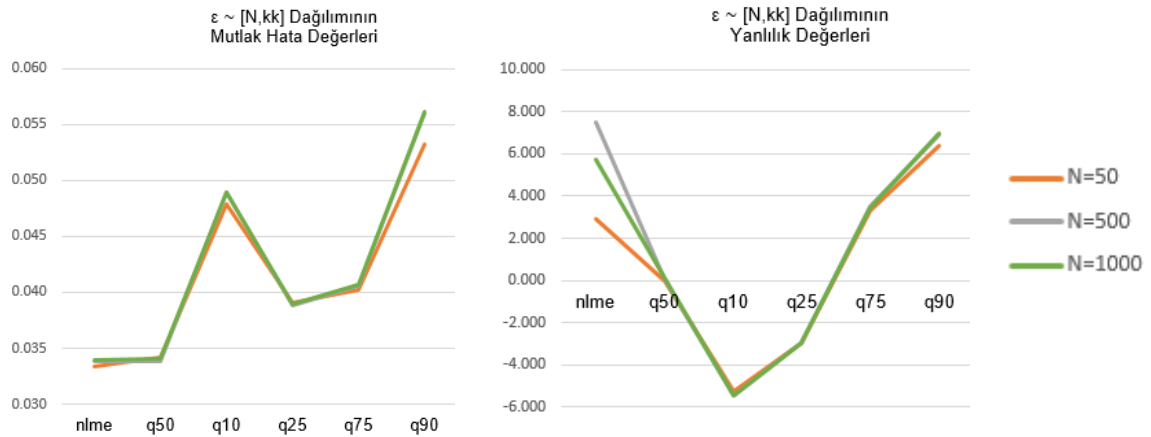
Çok düzeyli kantil regresyon modelinde, dikkate alınması gereken birden fazla kantil dilimi vardır ve her bir kantil düzeyi için elde edilen mutlak hata değeri, özellikle o kantildeki ortalama tahmin hatasını gösterir. Tablo 17 ile Şekil 26 birlikte incelendiğinde, genel olarak, N=50 örneklem büyüklüğünde farklı kantiller için mutlak hata değerleri 0.053 ile 0.086 arasında değişirken yanlılık değerleri ise -6.841 ile 7.859 arasında olduğu görülmektedir. N=500 örneklem büyüklüğünde, farklı kantiller için mutlak hata değerleri 0.053 ile 0.086 arasında değişirken yanlılık değerleri -7.121 ile 7.990 arasındadır. N=1000 örneklem büyüklüğünde, farklı kantiller için mutlak hata değerleri 0.053 ile 0.088 arasında değişirken yanlılık değerleri -8.123 ile 7.152 arasındadır. Tüm örneklem büyüklükleri için, çok düzeyli regresyon analizi (nlme) ve q50'nin mutlak hata değerlerinin birbirine çok yakın/aynı olduğu; yanlılık değerinde ise çok düzeyli regresyon yönteminde örneklem büyüklüğünde 500'den 100'e çıkarılmasının yanlılık değerini düşürdüğü görülmüştür. Çok düzeyli kantil regresyon bağlamında yanlılık değeri, modelin tahminlerinin belirli bir kantildeki gerçek değerlerden sistematik sapmasına atıfta bulunur. Modelin gerçek değerleri sürekli olarak fazla tahmin etme veya hafife alma eğiliminin yönü ve büyüklüğü

hakkında fikir verir. Her bir kantil düzeyi için elde edilen yanlılık değerinin pozitif olması durumunda, tahminlerin gerçek değerlerden daha yüksek olma eğiliminde olduğu anlamına gelirken, negatif yanlılık değeri, daha düşük olma eğiliminde oldukları anlamına gelir. Buna göre her iki hata düzeyinin de normal olduğu ($[N,N]$ dağılımı) durumda q50 kantil düzeyi daima çok düzeyli regresyon analizine göre daha yansız kestirimler yapmıştır. Ayrıca farklı kantil düzeyleri kendi içlerinde karşılaştırıldığında, q10, q25, q75 ve q90 kantillerinden elde edilen mutlak hata değerinin örneklem büyüklüğüne göre değişim göstermediği görülmüştür. Tüm örneklem büyüklüklerinde en yüksek mutlak hata daima q90 kantil düzeyinde elde edilmiştir.

Düzyer 1 hata terimlerinin normal ve Düzyer 2 hata terimlerinin az çarpık olduğu (X^2 dağılım) durumlardaki veri yapıları için ($[N,kk]$ dağılımı) yapılan analizler sonucu elde edilen mutlak yanlılık ve hata değerlerine ilişkin bulgulara ait grafikler Şekil 27'de verilmiştir.

Şekil 27

Düzyer 1 Hata Terimlerinin Normal Ve Düzyer 2 Hata Terimlerinin X^2 Dağılım Gösterdiği ($\epsilon \sim [N,kk]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri



*N=500 ve N=1000 için mutlak hata değerleri aynı değerler olarak hesaplandığı için mutlak hata değerleri grafiğinde 2 adet grafik bulunmaktadır.

Tablo 17 ve Şekil 27 birlikte incelendiğinde, tüm örneklem büyüklükleri için, medyandan ve q25 kantil düzeyinden elde edilen mutlak hata değerinin örneklem

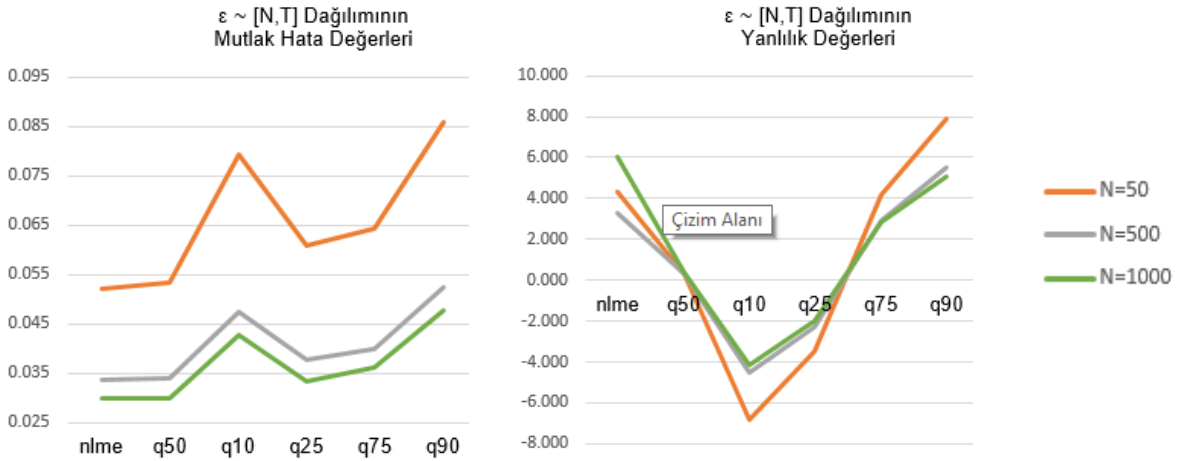
büyükliğünden bağımsız olarak aynı değer olarak hesaplanmıştır. Yanlılık değerleri incelendiğinde ise, çok düzeyli regresyon analizi ile elde edilen kestirimlerin q50 kantil düzeyinden elde edilen kestirime göre daha yüksek değerlerde olduğu görülmektedir. Dolayısıyla medyandan elde edilen kestirimler, çok düzeyli regresyon analizinden elde edilen kestirime göre daha yansızdır yorumu yapılabilir. Düzey 1 hatalar normal dağılım gösterirken, Düzey 2 hataların çarpıklığının artması durumunda, yani sırasıyla [N,N] dağılımı ile [N,kk] dağılımı karşılaştırıldığında mutlak hataların hem çok düzeyli regresyon analizinde hem de tüm kantil düzeylerinde ayrı ayrı olmak üzere daha düşük hesaplandığı görülmektedir. Sonuç olarak, Düzey 1 hata terimlerinin normal ve Düzey 2 hata terimlerinin az çarpık olduğu (X^2 dağılımı) veri yapıları için ([N,kk] dağılımı) örneklem büyüklüğünün artmasının mutlak hata değerlerinin hesaplanmasında etkili olmadığı ancak yanlılık değerlerinin hesaplanmasında farklılıklar oluşturduğu yorumu yapılabilir.

Düzey 1 hata terimlerinin normal ve Düzey 2 hata terimlerinin çok çarpık olduğu (Student's T3 dağılımı) veri yapıları için ([N,T] dağılımı) yapılan analizler sonucu elde edilen mutlak yanlılık ve hata değerlerine ilişkin bulgulara ait grafikler Şekil 28'de verilmiştir. Tablo 17 ve Şekil 28 birlikte incelendiğinde, örneklem büyüklüğündeki artışın hem çok düzeyli regresyon analizinde hem de tüm kantil düzeylerinde ayrı ayrı mutlak hata değerini azalttığı görülmektedir. Ayrıca çok düzeyli regresyon analizinden elde edilen mutlak hata değerleri ile medyandan elde edilen mutlak hata değeri birbirine çok yakın olmasına rağmen yanlılık değerleri incelendiğinde, medyan değeri olan q50 kantilinin daha az yanlı kestirimlerde bulunduğu söylenebilir.

Tablo 17'te verilen kantil düzeylerine ait yanlılık değerleri incelendiğinde birbirine simetrik kantil düzeylerinin (q90 ile q10; q25 ile q75) zıt işaretli yanlılık değerleri ile hesaplama yaptığı görülmektedir.

Şekil 28

Düzye 1 Hata Terimlerinin Normal ve Düzye 2 Hata Terimlerinin Student's T3 Dağılım Gösterdiği ($\epsilon \sim [N, T]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri



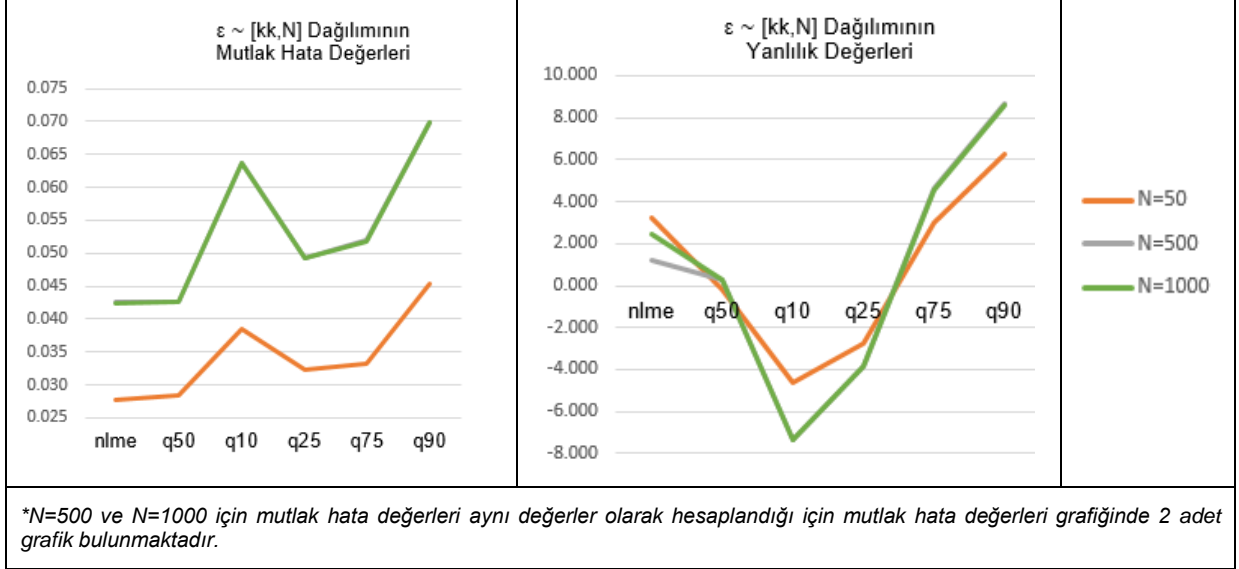
Son olarak mutlak hata ve yanlılık değerleri her iki hata düzeyinin dağılımlarının normal olduğu durumlarla karşılaştırıldığında benzer sonuçlar vermiştir. Yani verilerin normal $[N, N]$ ve düzey 2 hata terimlerinin çok çarpık olduğu $[N, T]$ dağılımlarda benzer sonuçlar elde edilebilmektedir. Şekil 26, Şekil 27 ve Şekil 28 birlikte incelendiğinde, düzey 1 hataların normal olduğu koşullarda ise en az mutlak hata ve yanlılık değeri ise düzey 2 hataların az çarpık olduğu durumlarda $[N, kk]$ elde edildiği görülmektedir.

Düzye 1 hata terimlerinin az çarpık (X^2 dağılımı) ve Düzye 2 hata terimlerinin normal dağılım gösterdiği veri yapıları için ($[kk, N]$ dağılımı) yapılan analizler sonucu elde edilen mutlak yanlılık ve hata değerlerine ilişkin bulgulara ait grafikler Şekil 29'da verilmiştir.

Tablo 17 ve Şekil 29 birlikte incelendiğinde, en düşük mutlak hata değerinin küçük örneklem olan $N=50$ 'de elde edildiği görülmüştür. $N=500$ ve $N=1000$ örneklem büyüklükleri benzer mutlak hata değerleri vermiştir. Şekil 29 incelendiğinde, küçük örneklemden büyük örnekleme geçişte mutlak hata değerleri artmıştır.

Şekil 29

Düzyey 1 Hata Terimlerinin X^2 ve Düzyey 2 Hata Terimlerinin Normal Dağılım Gösterdiği
($\varepsilon \sim [kk, N]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri



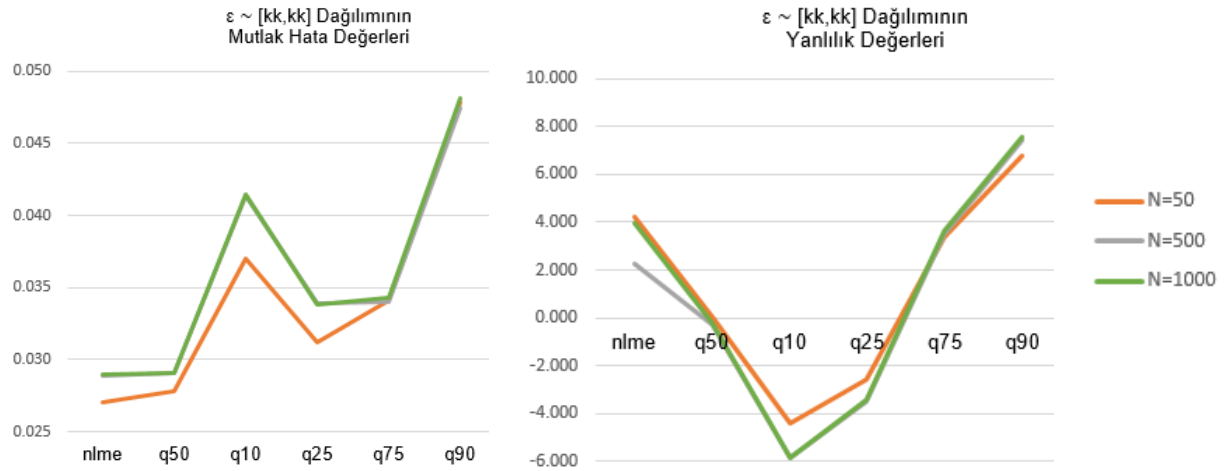
Şekil 29 incelendiğinde, çok düzeyli regresyon analizinden elde edilen mutlak hata değerleri ile medyandan elde edilen mutlak hata değeri birbirine çok yakın olmasına rağmen yanlılık değerleri incelendiğinde, medyan değeri olan q50 kantilinin daha az yanlı kestirimlerde bulunduğu söylenebilir. Diğer dağılımlarda olduğu gibi q90 kantil düzeyinde yanlılık değerleri en yüksek hesaplanmıştır.

Düzyey 1 ve Düzyey 2 hata terimlerinin az çarpık (X^2 dağılımı) olduğu veri yapıları için ([kk,kk] dağılımı) yapılan analizler sonucu elde edilen mutlak hata ve yanlılık değerlerine ilişkin bulgulara ait grafikler Şekil 30'da verilmiştir.

Tablo 17 ve Şekil 30 birlikte incelendiğinde, örneklem büyüklüğündeki artışın hem çok düzeyli regresyon analizinde hem de tüm kantil düzeylerinden elde edilen mutlak hata değerlerinin hesaplanmasında benzer sonuçlar verdiği görülmektedir.

Şekil 30

Düzyey 1 ve Düzyey 2 Hata Terimlerinin X^2 Dağılımı Gösterdiği ($\varepsilon \sim [kk, kk]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri



*N=500 ve N=1000 için mutlak hata değerleri aynı değerler olarak hesaplandığı için mutlak hata değerleri grafiğinde 2 adet grafik bulunmaktadır.

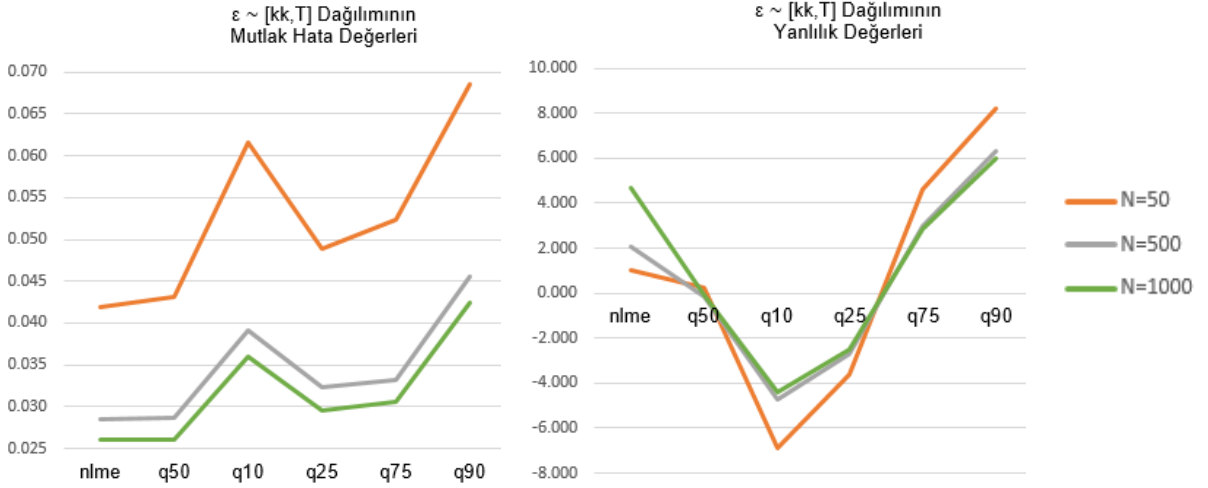
Tüm örneklem büyüklükleri için, çok düzeyli regresyon analizi ve medyan değeri olan q50 kantilinden elde edilen mutlak hata değerlerinin benzer olduğu görülürken yanlılık değerleri incelendiğinde, q50 kantil regresyondan daha az yanlı sonuçlar elde edilmiştir. Şekil 30'da yanlılık değerleri incelendiğinde, birbirinin simetriği olan kantil düzeylerinin (q10 ile q90 ve q25 ile q75) zıt işaretli hesaplamalar yaptığı ve q90 kantil düzeyinden elde edilen yanlılık değerinin diğer kantillere göre daha yüksek olduğu görülmektedir. Son olarak Düzyey 1 hata terimlerinin dağılımları az çarpıkken, düzyey 2 hata terimlerinin çarpıklıklarına göre $[kk, N]$ ve $[kk, kk]$ dağılımları karşılaştırıldığında, çarpıklıktaki artışın sonucunda büyük örneklerde daha az hata değeri ile hesaplamalar yapıldığı yorumu yapılabilir.

Düzyey 1 hata terimlerinin az çarpık (X^2 dağılımı) ve Düzyey 2 hata terimlerinin çok çarpık (Student's T3 dağılımı) olduğu veri yapıları için ($[kk, T]$ dağılımı) yapılan analizler sonucu elde edilen mutlak yanlılık ve hata değerlerine ilişkin bulgulara ait grafikler Şekil 31'de verilmiştir.

Şekil 31

Düzyey 1 Hata Terimlerinin X^2 ve Düzyey 2 Hata Terimlerinin Student's T3 dağılımı

Gösterdiği ($\varepsilon \sim [kk, T]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri



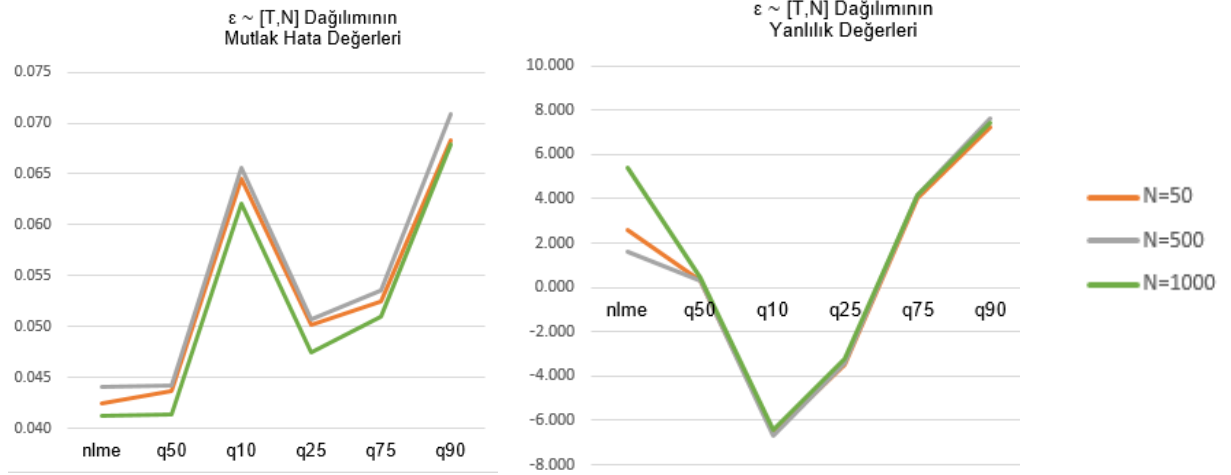
Tablo 17 ve Şekil 31 birlikte incelendiğinde, örneklem büyüklüğündeki artışın hem çok düzeyli regresyon analizinde hem de tüm kantil düzeylerinden elde edilen mutlak hata değerlerini düşürdüğü görülmüştür. Çok düzeyli regresyon analizinden elde edilen mutlak hata değeri örneklem büyüklüğüne bağlı olarak düşmesine rağmen yanlılık değeri artış göstermektedir. Dolayısıyla bu dağılımda çok düzeyli regresyon analizinin küçük örneklemelerde daha yansız kestirimler yaptığı söylenebilir. Aksi şekilde tüm kantil düzeylerinden elde edilen yanlılık değerleri ise örneklem büyüklüğü arttıkça azalma eğilimi göstermiştir. Yani bu dağılımın kantil düzeylerindeki hesaplamalarda daha yansız kestirimler yapılmasını sağladığı yorumu yapılabilir. Son olarak düzey 1 hataların çok çarpık olduğu durumlarda, düzey 2 hataların çarpıklıkları arttıkça küçük örneklemde mutlak hata değerinin arttığı ancak büyük örneklemelerde mutlak hata değerinin azaldığı görülmektedir.

Düzyey 1 hata terimlerinin çok çarpık (Student's T3 dağılımı) ve Düzyey 2 hata terimlerinin normal dağılım gösterdiği veri yapıları için ($[T, N]$ dağılımı) yapılan analizler

sonucu elde edilen mutlak yanlışlık ve hata değerlerine ilişkin bulgulara ait grafikler Şekil 32’te verilmiştir.

Şekil 32

Düzyer 1 Hata Terimlerinin Student’s T3 ve Düzyer 2 Hata Terimlerinin Normal Dağılım Gösterdiği ($\epsilon \sim [T,N]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Değerleri



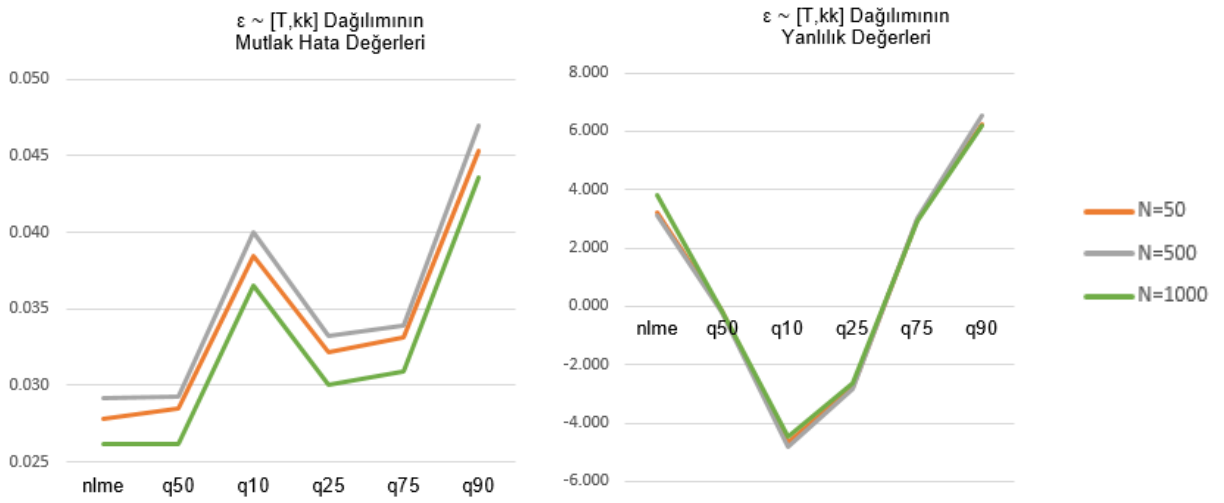
Tablo 17 ve Şekil 32 birlikte incelendiğinde, tüm örneklem büyüklüklerinde hem çok düzeyli regresyon analizi hem de farklı kantil düzeylerinden elde edilen mutlak hata değerlerinin benzer olduğu görülmektedir. Dolayısıyla düzey 1 hata terimlerinin çok çarpık olduğu düzey 2 hata terimlerinin normal dağılım gösterdiği durumlarda mutlak hata değerinin hesaplanmasında örneklem büyüklüğünün etkisi bulunmamaktadır yorumu yapılabilir. Yanlılık değerleri incelendiğinde ise medyandan elde edilen kestirimler daima çok düzeyli regresyon yöntemine göre daha az yanlış sonuçlar vermektedir. Farklı kantil düzeylerinde hesaplanan yanlışlık değerleri birbirine çok yakın olmakla birlikte diğer tüm dağılımlarda olduğu gibi birbirinin simetrisi olan kantil düzeyleri (q10 ile q90 ve q25 ile q75) daima zıt işaretli yanlışlık göstermiştir. Son olarak yine diğer tüm dağılımlarda olduğu gibi mutlak hata ve yanlışlık değerinin en yüksek olduğu kantil düzeyinin q90 kantili olduğu görülmüştür.

Düzyey 1 hata terimlerinin çok çarpık (Student's T3 dağılımı) ve Düzyey 2 hata terimlerinin az çarpık (X^2 dağılımı) olduđu veri yapıları için ([T,kk] dağılımı) yapılan analizler sonucu elde edilen mutlak yanlışlık ve hata deđerlerine ilişkin bulgulara ait grafikler Şekil 33'te verilmiştir.

Şekil 33

Düzyey 1 Hata Terimlerinin Student's T3 ve Düzyey 2 Hata Terimlerinin X^2 Dağılım

Gösterdiği ($\epsilon \sim [T,kk]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Deđerleri

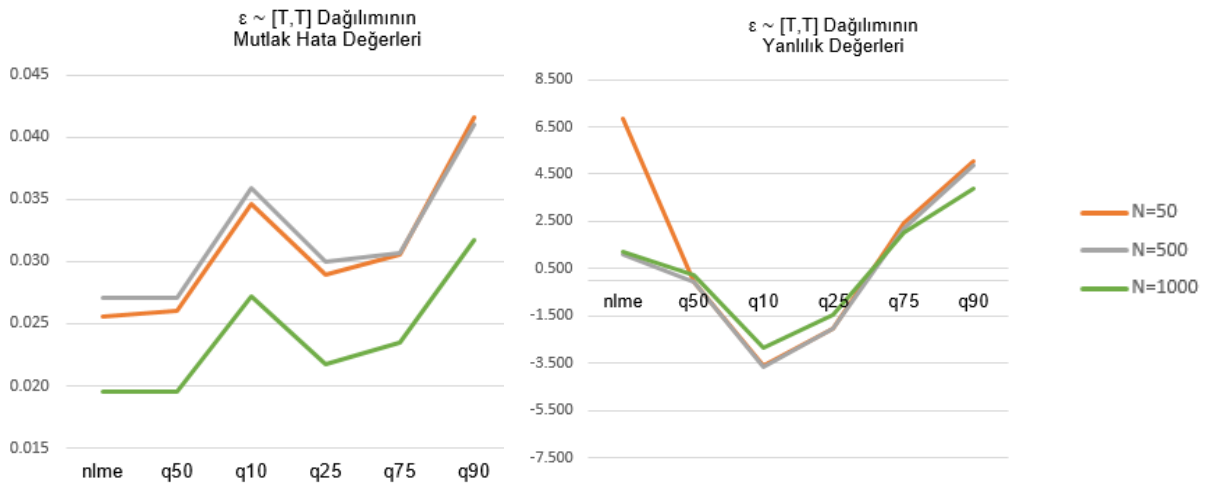


Tablo 17 ve Şekil 33 birlikte incelendiğinde, tüm örneklem büyüklüklerinde çok düzeyli regresyon ile medyandan elde edilen mutlak hata deđerlerinin aynı olduđu ancak yanlışlık deđerleri incelendiğinde q50 kantil düzeyinde daha az yanlış hesaplamaların yapıldığı görülmüştür. Tüm kantil düzeylerinden elde edilen mutlak hatalar da örneklem büyüklüğüne göre birbirine çok yakın deđerlerde hesaplanmıştır. En yüksek mutlak hata ve yanlışlık deđerleri ise q90 kantilinden elde edilmiştir. Son olarak düzyey 1 hataların dağılımı çok çarpıkken, düzyey 2 hataların çarpıklığı normalden az çarpığa doğru karşılaştırılırsa ([T,N] ve [T,kk]), düzyey 2 hata terimlerinin çarpıklığı arttıkça mutlak hata deđerinin hesaplanmasında daha etkili sonuçlar alındığı yorumu yapılabilir.

Düzye 1 ve Düzye 2 hata terimlerinin çok çarpık olduđu (Student's T3 dağılımı) veri yapıları için ([T,T] dağılımı) yapılan analizler sonucu elde edilen mutlak hata ve yanlılık deđerlerinin bulgularına ait grafikler Şekil 34'te verilmiştir.

Şekil 34

Düzye 1 ve Düzye 2 Hata Terimlerinin Student's T3 Dağılım Gösterdiği ($\epsilon \sim [T, T]$) Veri Yapıları İçin Mutlak Hata ve Yanlılık Deđerleri



Tablo 17 ve Şekil 34 birlikte incelendiğinde, tüm örneklem büyüklükleri için, çok düzeyli regresyon yönteminden elde edilen mutlak hata deđerleri ile medyandan elde edilen mutlak hata deđerinin aynı olduđu görülmektedir. Ancak yanlılık deđerleri incelendiğinde tüm örneklem büyüklüklerinde daima medyandan elde edilen hesaplamaların çok düzeyli regresyon yöntemine göre daha yansız sonuçlar verdiđi görülmektedir. Dolayısıyla örneklem büyüklüğündeki artışın bu dağılımda mutlak hata deđerinin hesaplanmasında çok etkisinin bulunmadığı ancak yanlılık hesaplamasında etkisinin görüldüğü yorumu yapılabilir. Son olarak en yüksek mutlak hata ve yanlılık deđerlerinin q90 kantilinden elde edildiđi ve birbirinin simetriđi olan kantil düzeylerinden (q10 ile q90 ve q25 ile q75) elde edilen yanlılık deđerlerinin ise birbirine zıt işaretili olduđu görülmüştür.

Üçüncü Alt Probleme İlişkin Bulgular ve Yorum

Simüle edilmiş boylamsal veri setlerinde, farklı Düzey 1 ile Düzey 2 hata terimlerinin dağılımları ve farklı örneklem büyüklüklerinde, çok düzeyli klasik regresyon analizi ile çok düzeyli kantil regresyon analiz yöntemlerinden elde edilen ortalama kesişim ve eğim katsayıları önemli ölçüde farklılık göstermekte midir?

Üçüncü alt probleme ilişkin yapılan analizler öncesinde kesişim ve eğim parametre değerleri incelenmiştir. Elde edilen kesişim/eğim katsayıları EK'te verilmiştir. Elde edilen değerler göz önünde bulundurulduğunda, dağılımın çarpık olması durumunun katsayıların farklılaşmasını yakalayıp yakalayamadığı incelenmiştir. Dağılımlar normal iken simülasyonda üretilen katsayı (17.17) ile çok düzeyli regresyon analizinden kestirilen kesişim katsayısının benzer değerler aldığı ancak dağılımların çarpıklığı arttıkça bu değerlerin yükseldiği görülmüştür. Küçük örneklem olan N=50 örneklem büyüklüğünde, büyük örneklemelere göre daima kesişim katsayısı değeri daha düşük elde edilmiştir. Çok düzeyli regresyon analizi için kestirilen en yüksek kesişim katsayısı (28.044) ise her iki düzey hatalarının da az çarpık olduğu [kk,kk] dağılımında elde edildiği görülmüştür. Farklı kantil düzeylerinden elde edilen kesişim katsayıları incelendiğinde ise en yüksek kesişim katsayısının daima q90 kantilinden elde edildiği görülmüştür. Her iki düzey hata terimlerinin dağılımı normal olduğunda, q10 ve q25 kantillerinden elde edilen kesişim katsayılarının tüm örneklem büyüklüklerinde, simülasyonda türetilen kesişim katsayısından daha düşük olduğu görülmüştür. Benzer şekilde kestirilen eğim değerleri incelendiğinde çok düzeyli regresyon analizi ve tüm kantil düzeylerinden elde edilen eğim katsayılarının simülasyonda türetilen eğim katsayısı ile benzer sonuçlar verdiği görülmektedir. Ayrıca tüm dağılımlarda çok düzeyli regresyon analizi ile medyandan elde edilen eğim katsayı değerleri birbirine çok yakın hesaplanmıştır. Kesişim katsayısı ile benzer şekilde, tüm dağılımlarda ve tüm örneklem büyüklüklerinde en büyük eğim katsayısı daima q90 kantil düzeyinden elde edilmiştir.

Üçüncü alt problem kapsamında her bir analiz yöntemi ile elde edilen bu kesişim/eğim katsayıları ile medyan ($Q(50)$) arasında ve her kantil düzeyinin ($1-Q(p)$) düzeyindeki kesişim/eğim katsayıları ile karşılaştırılması Wald testi kullanılarak yapılarak ayrı ayrı tablolandırılmıştır.

Düzye 1 ve Düzye 2 hataların normal ($[N,N]$ dağılımı) olduğu koşullar için elde edilen kesişim katsayılarına ait Wald istatistik değerleri Tablo 18'de verilmiştir.

Tablo 18

Düzye 1 ve Düzye 2 Hata Terimlerinin Normal Dağılım Gösterdiği ($\epsilon \sim [N,N]$) Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değerleri

Örneklem Büyüklüğü	Analiz Yöntemi	Kesişim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	($1-Q(p)$) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	17.958	0.029	
	$Q(0.10)$	11.665	12.341 (*)	61.430 (*)
	$Q(0.25)$	14.811	3.134	13.475 (*)
	$Q(0.50)$	18.254		
	$Q(0.75)$	21.798	3.242	
	$Q(0.90)$	25.411	14.260 (*)	
500	Çok düzeyli regresyon analizi	18.030	0.665	
	$Q(0.10)$	11.353	174.333 (*)	831.511 (*)
	$Q(0.25)$	14.672	44.789 (*)	185.174 (*)
	$Q(0.50)$	18.451		
	$Q(0.75)$	22.206	43.679 (*)	
	$Q(0.90)$	25.648	177.485 (*)	
1000	Çok düzeyli regresyon analizi	18.048	2.064	
	$Q(0.10)$	11.295	366.148 (*)	1724.728 (*)
	$Q(0.25)$	14.696	93.888 (*)	384.845 (*)
	$Q(0.50)$	18.575		
	$Q(0.75)$	22.378	88.687 (*)	
	$Q(0.90)$	25.777	353.113 (*)	

Düzye 1 ve Düzye 2 hataların normal olduğu koşullar için Tablo 18 incelendiğinde, tüm örneklem büyüklüklerinde çok düzeyli regresyon analizi ile elde edilen kesişim katsayıları ile medyandan elde edilen kesişim katsayısı arasında anlamlı farkın bulunmadığı görülmektedir. Tüm örneklem büyüklüklerinde, q_{10} ve q_{90} kantil düzeylerinden ayrı ayrı elde edilen kesişim katsayıları ile medyandan elde edilen kesişim katsayısı arasında anlamlı fark bulunmaktadır. Büyük örneklerde ($N=500$ ve $N=1000$)

q25 ve q75 kantil düzeylerinden ayrı ayrı elde edilen kesişim katsayıları ile medyandan elde edilen kesişim katsayısı arasında anlamlı fark bulunmaktadır. Son olarak tüm örneklem büyüklüklerinde, birbirinin simetriği olan q10 ve q90 ile q25 ve q75 kantillerinden elde edilen kesişim katsayılarının kendi aralarında anlamlı olarak farklı olduğu görülmektedir.

Düzey 1 ve Düzey 2 hataların normal ($[N,N]$ dağılımı) olduğu koşullar için elde edilen eğim katsayılarına ait Wald istatistik değerleri Tablo 19'da verilmiştir.

Tablo 19

Düzey 1 ve Düzey 2 Hata Terimlerinin Normal Dağılım Gösterdiği ($\varepsilon \sim [N,N]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri

Örneklem Büyüklüğü	Analiz Yöntemi	Eğim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	6.149	0.000	
	Q(0.10)	5.732	0.303	1.499
	Q(0.25)	5.894	0.104	0.511
	Q(0.50)	6.139		
	Q(0.75)	6.418	0.134	
	Q(0.90)	6.581	0.352	
500	Çok düzeyli regresyon analizi	6.131	0.000	
	Q(0.10)	5.731	3.304	15.295 (*)
	Q(0.25)	5.892	1.091	4.226 (*)
	Q(0.50)	6.135		
	Q(0.75)	6.351	0.866	
	Q(0.90)	6.517	2.965	
1000	Çok düzeyli regresyon analizi	6.117	0.006	
	Q(0.10)	5.726	5.769 (*)	31.233 (*)
	Q(0.25)	5.885	1.766	9.014 (*)
	Q(0.50)	6.104		
	Q(0.75)	6.361	2.426	
	Q(0.90)	6.518	6.986 (*)	

Tablo 19 incelendiğinde, küçük örneklem olan N=50 için katsayılar arasında anlamlı farklılık bulunmazken, büyük örneklem olan N=1000 için q10 ve q90 kantillerinden ayrı ayrı elde edilen eğim katsayıları ile medyandan elde edilen eğim katsayısı anlamlı olarak farklılaşmaktadır. Ayrıca N=500 ve N=1000 örneklemelerinde birbirinin simetriği olan

q10 ile q90 ve q25 ile q75 kantillerinden elde edilen eğim katsayıları da kendi aralarında anlamlı olarak farklı bulunmuştur.

Düzey 1 hataların normal ve Düzey 2 hataların az çarpık ($[N, k]$ dağılımı) olduğu koşullar için elde edilen kesişim katsayılarına ait Wald istatistik değerleri Tablo 20'de verilmiştir.

Tablo 20

Düzey 1 Hata Terimlerinin Normal ve Düzey 2 Hata Terimlerinin X^2 Dağılımı Gösterdiği ($\varepsilon \sim [N, X^2]$) Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değerleri

Örneklem Büyüklüğü	Analiz Yöntemi	Kesişim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	23.053	0.009	
	Q(0.10)	18.497	8.932 (*)	46.029 (*)
	Q(0.25)	20.373	2.804	12.795 (*)
	Q(0.50)	22.920		
	Q(0.75)	25.990	3.565	
	Q(0.90)	28.853	13.377 (*)	
500	Çok düzeyli regresyon analizi	23.006	0.000	
	Q(0.10)	18.091	134.429 (*)	718.787 (*)
	Q(0.25)	20.324	37.285 (*)	167.941 (*)
	Q(0.50)	23.006		
	Q(0.75)	26.170	45.100 (*)	
	Q(0.90)	29.527	199.079 (*)	
1000	Çok düzeyli regresyon analizi	23.034	0.012	
	Q(0.10)	18.199	262.997 (*)	1398.899 (*)
	Q(0.25)	20.357	75.808 (*)	328.161 (*)
	Q(0.50)	23.066		
	Q(0.75)	26.155	85.177 (*)	
	Q(0.90)	29.473	383.006 (*)	

Düzey 1 hataların normal ve Düzey 2 hataların az çarpık olduğu koşullar için Tablo 20 incelendiğinde, tüm örneklem büyüklüklerinde çok düzeyli regresyon analizi ile elde edilen kesişim katsayıları ile medyandan elde edilen kesişim katsayısı arasında anlamlı farkın bulunmadığı görülmektedir. Tüm örneklem büyüklüklerinde q10 ve q90 kantillerinden ayrı ayrı elde edilen ortalama kesişim katsayısının, medyan değeri ile elde edilen katsayıdan ve simetrik kantiller olan q10 ve q90 ile q25 ve q75 kantillerinden elde edilen kesişim katsayılarının kendi aralarında anlamlı olarak farklı olduğu görülmektedir.

Büyük örneklem olan N=500, N=1000 örneklem büyüklüklerinde, q25 ile q75 kantil düzeylerinden ayrı ayrı elde edilen kesişim katsayısının, medyan değeri ile elde edilen katsayıdan anlamlı olarak farklı olduğu görülmektedir.

Düzyer 1 hataların normal ve Düzyer 2 hataların az çarpık olduğu koşullar için elde edilen eğim katsayılarına ait Wald istatistik değerleri Tablo 21’de verilmiştir.

Tablo 21

Düzyer 1 Hata Terimlerinin Normal ve Düzyer 2 Hata Terimlerinin X^2 Dağılımı Gösterdiği ($\epsilon \sim [N, X^2]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri

Örneklem Büyüklüğü	Analiz Yöntemi	Eğim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	6.110	0.023	
	Q(0.10)	5.551	1.221	2.913
	Q(0.25)	5.837	0.373	1.140
	Q(0.50)	6.206		
	Q(0.75)	6.481	0.190	
	Q(0.90)	6.571	0.329	
500	Çok düzeyli regresyon analizi	6.133	0.229	
	Q(0.10)	5.559	14.160 (*)	35.387 (*)
	Q(0.25)	5.899	3.193	10.102 (*)
	Q(0.50)	6.226		
	Q(0.75)	6.483	1.761	
	Q(0.90)	6.597	3.796	
1000	Çok düzeyli regresyon analizi	6.121	0.434	
	Q(0.10)	5.515	30.742 (*)	84.018 (*)
	Q(0.25)	5.866	7.0843 (*)	23.052 (*)
	Q(0.50)	6.213		
	Q(0.75)	6.489	4.0607 (*)	
	Q(0.90)	6.646	10.253 (*)	

Tablo 21 incelendiğinde, tüm örneklem büyüklüklerinde çok düzeyli regresyon analizi ile elde edilen eğim katsayısının medyandan elde edilen eğim katsayısı ile aralarında anlamlı farkın bulunmadığı görülmektedir. Küçük örneklem olan N=50 için kantillerden elde edilen eğim katsayıları arasında anlamlı fark bulunamamıştır. Büyük örneklem olan N=500 ve N=1000 için, q10 kantil düzeyinden elde edilen eğim katsayısının medyandan elde edilen eğim katsayısı arasında; q10 ile q90 ve q25 ile q75 kantillerinin de kendi aralarında anlamlı olarak farklılaştığı görülmektedir. Ayrıca sadece

N=1000 örneklem büyüklüğünde; q90, q25 ve q75 kantillerinden ayrı ayrı elde edilen eğim katsayılarının medyandan elde edilen eğim katsayısı ile anlamlı olarak fark gösterdiği görülmektedir.

Düzyey 1 hataların normal ve Düzyey 2 hataların çok çarpık ($[N,T]$ dağılımı) olduğu koşullar için elde edilen kesişim katsayılarına ait Wald istatistik değerleri Tablo 22'de verilmiştir.

Tablo 22

Düzyey 1 Hata Terimlerinin Normal ve Düzyey 2 Hata Terimlerinin Student-t Dağılım Gösterdiği ($\varepsilon \sim [N,T]$) Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değerleri

Örneklem Büyüklüğü	Analiz Yöntemi	Kesişim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	17.958	0.029	
	Q(0.10)	11.665	12.341 (*)	61.430 (*)
	Q(0.25)	14.811	3.134	13.475 (*)
	Q(0.50)	18.254		
	Q(0.75)	21.798	3.242	
	Q(0.90)	25.411	14.260 (*)	
500	Çok düzeyli regresyon analizi	20.799	0.621	
	Q(0.10)	17.026	141.559 (*)	563.477 (*)
	Q(0.25)	18.925	38.993 (*)	151.749 (*)
	Q(0.50)	21.072		
	Q(0.75)	23.315	37.692 (*)	
	Q(0.90)	25.559	145.598 (*)	
1000	Çok düzeyli regresyon analizi	20.748	3.871 (*)	
	Q(0.10)	17.417	315.874 (*)	1204.745 (*)
	Q(0.25)	19.181	91.587 (*)	360.583 (*)
	Q(0.50)	21.160		
	Q(0.75)	23.205	92.225 (*)	
	Q(0.90)	25.107	324.898 (*)	

Düzyey 1 hataların normal ve Düzyey 2 hataların çok çarpık olduğu koşullar için Tablo 22 incelendiğinde, sadece N=1000 örneklem büyüklüğünde çok düzeyli regresyon analizi ile elde edilen kesişim katsayısı, medyandan elde edilen kesişim katsayısından anlamlı olarak farklılaştığı görülmüştür. Tüm örneklem büyüklüklerinde q10 ve q90 kantillerinden ayrı ayrı elde edilen ortalama kesişim katsayısının, medyan değeri ile elde

edilen katsayıdan anlamlı olarak farklı olduğu görülmektedir. N=500 ve N=1000 örneklem büyüklüklerinde, q25 ve q75 kantil düzeyinde elde edilen kesişim katsayıları, medyan değeri ile elde edilen kesişim katsayısından anlamlı olarak farklılaşmaktadır. Ayrıca tüm örneklem büyüklüklerinde, uç kantiller olan q10 ve q90 ile q25 ve q75 kantillerinden elde edilen kesişim katsayılarının kendi aralarında anlamlı olarak farklı olduğu görülmektedir.

Düzey 1 hataların normal ve Düzey 2 hataların çok çarpık ($[N, T]$ dağılımı) olduğu koşullar için elde edilen eğim katsayılarına ait Wald istatistik değerleri Tablo 23'te verilmiştir.

Tablo 23

Düzey 1 Hata Terimlerinin Normal ve Düzey 2 Hata Terimlerinin Student-t Dağılım

Gösterdiği ($\epsilon \sim [N, T]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri

Örneklem Büyüklüğü	Analiz Yöntemi	Eğim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	6.149	0.000	
	Q(0.10)	5.732	0.303	1.499
	Q(0.25)	5.894	0.104	0.511
	Q(0.50)	6.139		
	Q(0.75)	6.418	0.134	
	Q(0.90)	6.581	0.352	
500	Çok düzeyli regresyon analizi	6.121	0.106	
	Q(0.10)	5.385	35.096 (*)	107.681 (*)
	Q(0.25)	5.771	9.306 (*)	32.036 (*)
	Q(0.50)	6.170		
	Q(0.75)	6.545	7.145 (*)	
	Q(0.90)	6.877	23.562 (*)	
1000	Çok düzeyli regresyon analizi	6.136	0.037	
	Q(0.10)	5.349	107.567 (*)	341.918 (*)
	Q(0.25)	5.737	32.667 (*)	119.241 (*)
	Q(0.50)	6.153		
	Q(0.75)	6.567	30.287 (*)	
	Q(0.90)	6.937	93.982 (*)	

Tablo 23 incelendiğinde, küçük örneklem olan N=50 için hiçbir kantil düzeyinde eğimler arasında anlamlı fark gözlenmemiştir. Büyük örneklemelerde (N=500 ve N=1000) q10, q25, q75 ve q90 kantillerinden ayrı ayrı elde edilen eğim katsayısı ile medyandan elde edilen eğim katsayısının anlamlı olarak farklılaştığı görülmektedir. Son olarak q10 ile

q90 ve q25 ile q75 kantillerinin kendi aralarında eğim katsayılarının anlamlı fark gösterdiği görülmektedir.

Düzyey 1 hataların az çarpık ve Düzyey 2 hataların normal ($[X^2, N]$ dağılımı) olduğu koşullar için edilen keşişim katsayılarına ait Wald istatistik değeri Tablo 24'te verilmiştir.

Tablo 24

Düzyey 1 Hata Terimlerinin X^2 ve Düzyey 2 Hata Terimlerinin Normal Dağılım Gösterdiği

($\epsilon \sim [X^2, N]$) Veri Yapıları İçin Keşişim Katsayılarına Ait Hesaplanan Wald Değeri

Örneklem Büyüklüğü	Analiz Yöntemi	Keşişim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	($1-Q(p)$) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	25.857	0.034	
	$Q(0.10)$	21.892	6.884 (*)	42.191 (*)
	$Q(0.25)$	23.331	2.309	11.032 (*)
	$Q(0.50)$	25.605		
	$Q(0.75)$	28.483	3.115	
	$Q(0.90)$	31.696	13.402 (*)	
500	Çok düzeyli regresyon analizi	23.084	0.197	
	$Q(0.10)$	16.164	167.222 (*)	895.598 (*)
	$Q(0.25)$	19.459	43.584 (*)	195.744 (*)
	$Q(0.50)$	23.327		
	$Q(0.75)$	27.424	49.205 (*)	
	$Q(0.90)$	31.398	208.396 (*)	
1000	Çok düzeyli regresyon analizi	23.049	0.627	
	$Q(0.10)$	16.100	346.835 (*)	1799.188 (*)
	$Q(0.25)$	19.432	90.494 (*)	392.710 (*)
	$Q(0.50)$	23.353		
	$Q(0.75)$	27.378	96.046 (*)	
	$Q(0.90)$	31.313	409.462 (*)	

Düzyey 1 hataların az çarpık ve Düzyey 2 hataların normal olduğu koşullar için Tablo 24 incelendiğinde, tüm örneklem büyüklüklerinde çok düzeyli regresyon analizi ile elde edilen keşişim katsayıları ile medyandan elde edilen keşişim katsayısı arasında anlamlı farkın bulunmadığı görülmektedir. Tüm örneklem büyüklüklerinde q10 ve q90 kantillerinden ayrı ayrı elde edilen ortalama keşişim katsayısının, medyan değeri ile elde edilen katsayıdan anlamlı olarak farklı olduğu görülmektedir. Yine tüm örneklem büyüklüklerinde, birbirlerinin simetriği olan kantillerden q10 ile q90 ve q25 ile q75'den elde edilen keşişim katsayıları kendi aralarında anlamlı fark oluşturmuştur. Son olarak büyük

örneklemelerde (N=500 ve N=1000) q25 ve q75 kantil düzeyinden ayrı ayrı elde edilen kesişim katsayıları, medyan değeri ile elde edilen kesişim katsayısından anlamlı olarak farklılaşmaktadır.

Düzyey 1 hataların az çarpık ve Düzyey 2 hataların normal ($[X^2, N]$ dağılımı) olduğu koşullar için edilen eğim katsayılarına ait Wald istatistik değeri Tablo 25'te verilmiştir.

Tablo 25

Düzyey 1 Hata Terimlerinin X^2 ve Düzyey 2 Hata Terimlerinin Normal Dağılım Gösterdiği ($\epsilon \sim [X^2, N]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değeri

Örneklem Büyüklüğü	Analiz Yöntemi	Eğim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	6.141	0.078	
	Q(0.10)	5.493	2.519	3.369
	Q(0.25)	5.916	0.546	1.327
	Q(0.50)	6.311		
	Q(0.75)	6.556	0.171	
	Q(0.90)	6.525	0.121	
500	Çok düzeyli regresyon analizi	6.119	0.031	
	Q(0.10)	5.703	3.895 (*)	12.583 (*)
	Q(0.25)	5.901	1.132	3.826
	Q(0.50)	6.164		
	Q(0.75)	6.373	0.670	
	Q(0.90)	6.468	1.535	
1000	Çok düzeyli regresyon analizi	6.129	0.021	
	Q(0.10)	5.716	7.118 (*)	25.062 (*)
	Q(0.25)	5.914	1.918	7.833 (*)
	Q(0.50)	6.155		
	Q(0.75)	6.390	1.715	
	Q(0.90)	6.476	3.481	

Tablo 25 incelendiğinde, küçük örneklem olan N=50 için hiçbir kantil düzeyinde eğimler arasında anlamlı fark gözlenmemiştir. Büyük örneklemelerde (N=500 ve N=1000) sadece q10 kantil düzeyinden elde edilen eğim katsayısı ile medyandan elde edilen eğim katsayısı arasında ve q10 ile 90 kantillerinden elde edilen eğim katsayılarının da kendi aralarında anlamlı olarak farklılaştığı görülmektedir. Son olarak sadece N=1000 örneklem büyüklüğünde, q25 ile q75 kantillerinden elde edilen eğim katsayılarının kendi aralarında anlamlı olarak farklılaştığı görülmektedir.

Düzyey 1 ve Düzyey 2 hataların az çarpık ($[X^2, X^2]$ dağılımı) olduđu koşullar için elde edilen keşişim katsayılarına ait Wald istatistik değeri Tablo 26'da verilmiştir.

Tablo 26

Düzyey 1 ve Düzyey 2 Hata Terimlerinin X^2 Dağılım Gösterdiği ($\epsilon \sim [X^2, X^2]$) Veri Yapıları İçin Keşişim Katsayılarına Ait Hesaplanan Wald Değeri

Örneklem Büyüklüğü	Analiz Yöntemi	Keşişim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	25.862	0.006	
	Q(0.10)	22.107	9.206 (*)	63.487 (*)
	Q(0.25)	23.391	3.777	19.965 (*)
	Q(0.50)	25.767		
	Q(0.75)	28.990	6.321 (*)	
	Q(0.90)	32.474	24.916 (*)	
500	Çok düzeyli regresyon analizi	28.016	0.778	
	Q(0.10)	22.761	119.604 (*)	759.315 (*)
	Q(0.25)	24.721	38.859 (*)	189.718 (*)
	Q(0.50)	27.613		
	Q(0.75)	31.190	53.687 (*)	
	Q(0.90)	35.040	235.073 (*)	
1000	Çok düzeyli regresyon analizi	28.044	1.044	
	Q(0.10)	22.788	246.903 (*)	1592.832 (*)
	Q(0.25)	24.797	79.120 (*)	387.968 (*)
	Q(0.50)	27.713		
	Q(0.75)	31.337	110.091 (*)	
	Q(0.90)	35.268	491.872 (*)	

Düzyey 1 ve Düzyey 2 hataların az çarpık olduđu koşullar için Tablo 26 incelendiğinde, tüm örneklem büyüklüklerinde çok düzeyli regresyon analizi ile elde edilen keşişim katsayısı ile medyandan elde edilen keşişim katsayısı arasında anlamlı farkın bulunmadığı görülmektedir. Tüm örneklem büyüklüklerinde q10, q75 ve q90 kantillerinden ayrı ayrı elde edilen ortalama keşişim katsayısının, medyan değeriyle elde edilen katsayıdan anlamlı olarak farklı olduđu görülmektedir. Büyük örneklerde (N=500 ve N=1000) q25 kantil düzeyinden elde edilen keşişim katsayısı ile medyandan elde edilen keşişim katsayısı anlamlı farklılık göstermektedir. Yine tüm örneklem büyüklüklerinde, birbirlerinin simetriği olan kantillerden q10 ile q90 ve q25 ile q75'den elde edilen keşişim katsayılarının kendi aralarında anlamlı fark oluşturduđu görülmüştür.

Düzyey 1 ve Düzyey 2 hataların az çarpık ($[X^2, X^2]$ dağılımı) olduđu koşullar için elde edilen eğim katsayılarına ait Wald istatistik değeri Tablo 27’de verilmiştir.

Tablo 27

Düzyey 1 ve Düzyey 2 Hata Terimlerinin X^2 Dağılım Gösterdiği ($\epsilon \sim [X^2, X^2]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri

Örneklem Büyüklüğü	Analiz Yöntemi	Eğim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	6.079	0.070	
	Q(0.10)	5.413	2.880	4.271 (*)
	Q(0.25)	5.874	0.460	1.206
	Q(0.50)	6.208		
	Q(0.75)	6.439	0.198	
	Q(0.90)	6.511	0.296	
500	Çok düzeyli regresyon analizi	6.124	0.526	
	Q(0.10)	5.587	13.059 (*)	24.945 (*)
	Q(0.25)	5.962	2.406	6.211 (*)
	Q(0.50)	6.275		
	Q(0.75)	6.471	0.801	
	Q(0.90)	6.529	1.394	
1000	Çok düzeyli regresyon analizi	6.120	1.154	
	Q(0.10)	5.569	27.748 (*)	47.952 (*)
	Q(0.25)	5.972	4.598 (*)	11.154 (*)
	Q(0.50)	6.279		
	Q(0.75)	6.455	1.297	
	Q(0.90)	6.489	1.932	

Tablo 27 incelendiğinde, tüm örneklem büyüklüklerinde çok düzeyli regresyon analizi ile medyandan elde edilen eğim katsayısı arasında anlamlı fark bulunamazken; q10 ile q90 kantillerinden elde edilen eğim katsayılarının kendi aralarında anlamlı farklılık gösterdiği görülmektedir. Büyük örneklerde (N=500 ve N=1000) q10 kantilinden elde edilen eğim katsayısı ile medyandan elde edilen eğim katsayısı arasında ve q25 ile q75 kantillerinden elde edilen eğim katsayılarının da kendi aralarında anlamlı olarak farklılaştığı görülmektedir. Son olarak sadece N=1000 örneklem büyüklüğünde q25 kantilinden elde edilen eğim katsayısı medyandan elde edilen eğim katsayısından anlamlı olarak farklılaşmaktadır.

Düzyey 1 hataların az çarpık ve Düzyey 2 hataların çok çarpık ($[X^2, T]$ dağılımı) olduđu koşullar için elde edilen kesişim katsayılarına ait Wald istatistik değeri Tablo 28'de verilmiştir.

Tablo 28

Düzyey 1 Hata Terimlerinin X^2 ve Düzyey 2 Hataların Student-t Dağılımı Gösterdiği ($\epsilon \sim [X^2, T]$) Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değeri

Örneklem Büyüklüğü	Analiz Yöntemi	Kesişim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	23.043	0.012	
	Q(0.10)	16.692	11.974 (*)	60.698 (*)
	Q(0.25)	19.752	3.092	15.432 (*)
	Q(0.50)	23.238		
	Q(0.75)	27.425	4.374 (*)	
	Q(0.90)	30.887	15.145 (*)	
500	Çok düzeyli regresyon analizi	25.767	0.561	
	Q(0.10)	21.805	104.920 (*)	660.082 (*)
	Q(0.25)	23.307	35.141 (*)	181.535 (*)
	Q(0.50)	25.482		
	Q(0.75)	28.371	57.015 (*)	
	Q(0.90)	31.594	239.389 (*)	
1000	Çok düzeyli regresyon analizi	25.746	0.665	
	Q(0.10)	22.236	213.870 (*)	1384.778 (*)
	Q(0.25)	23.591	74.146 (*)	400.308 (*)
	Q(0.50)	25.550		
	Q(0.75)	28.196	132.289 (*)	
	Q(0.90)	31.188	545.847 (*)	

Düzyey 1 hataların az çarpık ve Düzyey 2 hataların çok çarpık olduđu koşullar için Tablo 28 incelendiğinde, tüm örneklem büyüklüklerinde çok düzeyli regresyon analizi ile elde edilen kesişim katsayısının medyandan elde edilen kesişim katsayısı ile anlamlı fark göstermediği görülmektedir. Tüm örneklem büyüklüklerinde q10, q75 ve q90 kantillerinden ayrı ayrı elde edilen ortalama kesişim katsayısının, medyan değeri ile elde edilen katsayıdan anlamlı olarak farklı olduđu görülmektedir. Yine tüm örneklem büyüklüklerinde, birbirlerinin simetriği olan kantillerden q10 ile q90 ve q25 ile q75'den elde edilen kesişim katsayılarının kendi aralarında anlamlı fark oluşturduđu görülmüştür. Büyük

örneklemelerde (N=500 ve N=1000), q25 kantil düzeyinde elde edilen kesişim katsayıları, medyan değeri ile elde edilen kesişim katsayısından anlamlı olarak farklılaşmaktadır.

Düzyey 1 hataların az çarpık ve Düzyey 2 hataların çok çarpık ($[X^2, T]$ dağılımı) olduğu koşullar için elde edilen eğim katsayılarına ait Wald istatistik değerleri Tablo 29'da verilmiştir.

Tablo 29

Düzyey 1 Hata Terimlerinin X^2 ve Düzyey 2 Hataların Student-t Dağılımı Gösterdiği ($\epsilon \sim [X^2, T]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri

Örneklem Büyüklüğü	Analiz Yöntemi	Eğim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	6.141	0.002	
	Q(0.10)	5.744	0.306	1.018
	Q(0.25)	5.882	0.131	0.460
	Q(0.50)	6.174		
	Q(0.75)	6.422	0.087	
	Q(0.90)	6.496	0.154	
500	Çok düzeyli regresyon analizi	6.115	1.311	
	Q(0.10)	5.391	38.746 (*)	59.570 (*)
	Q(0.25)	5.891	7.285 (*)	17.435 (*)
	Q(0.50)	6.298		
	Q(0.75)	6.549	2.365	
	Q(0.90)	6.644	4.023 (*)	
1000	Çok düzeyli regresyon analizi	6.128	2.988	
	Q(0.10)	5.306	128.588 (*)	201.115 (*)
	Q(0.25)	5.838	26.378 (*)	71.685 (*)
	Q(0.50)	6.283		
	Q(0.75)	6.604	12.683 (*)	
	Q(0.90)	6.743	20.541 (*)	

Tablo 29 incelendiğinde, küçük örneklem olan N=50 için farklı kantillerden elde edilen hiçbir eğim katsayısı arasında anlamlı fark bulunamamıştır. Büyük örneklemelerde (N=500 ve N=1000) q10, q25 ve q90 kantillerinden ayrı ayrı elde edilen eğim katsayılarının medyandan elde edilen eğim katsayısından anlamlı olarak farklılaştığı görülmektedir. Yine büyük örneklemelerde, q10 ile q90 ve q25 ile q75 kantillerinden elde edilen eğim değerleri de kendi aralarında anlamlı olarak farklılaşmaktadır. Son olarak

sadece $N=1000$ örneklem büyüklüğünde q_{75} kantilinden elde edilen eğim katsayısının medyandan elde edilen eğim katsayısından anlamlı olarak farklılaştığı görülmektedir.

Düzy 1 hataların çok çarpık ve Düzy 2 hataların normal ($[T, N]$ dağılımı) olduğu koşullar için kesişim katsayılarından elde edilen Wald istatistik değeri Tablo 30'da verilmiştir.

Tablo 30

Düzy 1 Hata Terimlerinin Student-t ve Düzy 2 Hataların Normal Dağılımı Gösterdiği ($\epsilon \sim [T, N]$) Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değeri

Örneklem Büyüklüğü	Analiz Yöntemi	Kesişim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	($1-Q(p)$) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	20.774	0.029	
	$Q(0.10)$	14.646	12.146 (*)	56.328 (*)
	$Q(0.25)$	17.515	3.337	13.794 (*)
	$Q(0.50)$	21.062		
	$Q(0.75)$	24.502	3.145	
	$Q(0.90)$	27.596	11.919 (*)	
500	Çok düzeyli regresyon analizi	20.796	0.394	
	$Q(0.10)$	14.600	146.589 (*)	755.669 (*)
	$Q(0.25)$	17.646	36.812 (*)	162.311 (*)
	$Q(0.50)$	21.119		
	$Q(0.75)$	24.725	39.450 (*)	
	$Q(0.90)$	28.028	160.579 (*)	
1000	Çok düzeyli regresyon analizi	20.829	1.602	
	$Q(0.10)$	14.920	287.111 (*)	1472.080 (*)
	$Q(0.25)$	17.845	73.635 (*)	313.734 (*)
	$Q(0.50)$	21.270		
	$Q(0.75)$	24.706	73.672 (*)	
	$Q(0.90)$	27.824	302.111 (*)	

Düzy 1 hataların çok çarpık ve Düzy 2 hataların normal olduğu koşullar için Tablo 30 incelendiğinde, tüm örneklem büyüklüklerinde çok düzeyli regresyon analizi ile elde edilen kesişim katsayısının medyandan elde edilen kesişim katsayısı ile anlamlı fark göstermediği görülmektedir. Tüm örneklem büyüklüklerinde q_{10} ve q_{90} kantillerinden ayrı ayrı elde edilen ortalama kesişim katsayısının, medyan değeri ile elde edilen katsayıdan anlamlı olarak farklı olduğu görülmektedir. Büyük örneklerde ($N=500$ ve $N=1000$), q_{25} ve q_{75} kantil düzeyinden ayrı ayrı elde edilen kesişim katsayıları medyan değeri ile elde

edilen kesişim katsayısından anlamlı olarak farklılaşmaktadır. Son olarak tüm örneklem büyüklüklerinde uç kantiller olan q10 ile q90 ve q25 ile q75 kantillerinden elde edilen kesişim katsayılarının kendi içlerinde anlamlı farklılık gösterdiği görülmektedir.

Düzy 1 hataların çok çarpık ve Düzy 2 hataların normal ([T,N] dağılımı) olduğu koşullar için eğim katsayılarından elde edilen Wald istatistik değeri Tablo 31’de verilmiştir.

Tablo 31

Düzy 1 Hata Terimlerinin Student-t ve Düzy 2 Hataların Normal Dağılımı Gösterdiği ($\epsilon \sim [T, N]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değeri

Örneklem Büyüklüğü	Analiz Yöntemi	Eğim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	6.107	0.000	
	Q(0.10)	5.662	0.468	1.560
	Q(0.25)	5.853	0.149	0.640
	Q(0.50)	6.118		
	Q(0.75)	6.395	0.156	
	Q(0.90)	6.483	0.272	
500	Çok düzeyli regresyon analizi	6.136	0.007	
	Q(0.10)	5.679	5.394 (*)	21.563 (*)
	Q(0.25)	5.883	1.626	6.641 (*)
	Q(0.50)	6.156		
	Q(0.75)	6.422	1.479	
	Q(0.90)	6.574	3.897 (*)	
1000	Çok düzeyli regresyon analizi	6.124	0.012	
	Q(0.10)	5.646	12.934 (*)	54.904 (*)
	Q(0.25)	5.865	3.775	14.710 (*)
	Q(0.50)	6.142		
	Q(0.75)	6.396	3.102	
	Q(0.90)	6.596	10.622 (*)	

Tablo 31 incelendiğinde, küçük örneklem olan N=50 için farklı kantillerden elde edilen hiçbir eğim katsayısı arasında anlamlı fark bulunamamıştır. Büyük örneklerde (N=500 ve N=1000) q10 ve q90 kantillerinden elde edilen eğim katsayısı ile medyandan elde edilen eğim katsayısı arasında, ayrıca simetrik kantiller olan q90 ile q10 ve q25 ile q75 kantillerinden elde edilen eğim katsayılarının da kendi aralarında anlamlı olarak farklılaştığı görülmektedir.

Düzye 1 hataların çok çarpık ve Düzye 2 hataların az çarpık ($[T, X^2]$ dağılımı) olduđu koşullar için kesişim katsayılarından elde edilen Wald istatistik değeri Tablo 32'de verilmiştir.

Tablo 32

Düzye 1 Hata Terimlerinin Student-t ve Düzye 2 Hataların X^2 Dağılımı Gösterdiği ($\epsilon \sim [T, X^2]$) Veri Yapıları İçin Kesişim Katsayılarına Ait Hesaplanan Wald Değeri

Örnekleme Büyüklüğü	Analiz Yöntemi	Kesişim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	25.857	0.034	
	Q(0.10)	21.892	6.884 (*)	42.191 (*)
	Q(0.25)	23.331	2.309	11.032 (*)
	Q(0.50)	25.605		
	Q(0.75)	28.483	3.115	
	Q(0.90)	31.696	13.402 (*)	
500	Çok düzeyli regresyon analizi	25.809	0.897	
	Q(0.10)	21.675	79.468 (*)	593.755 (*)
	Q(0.25)	23.208	24.797 (*)	132.217 (*)
	Q(0.50)	25.410		
	Q(0.75)	28.396	39.465 (*)	
	Q(0.90)	31.860	191.663 (*)	
1000	Çok düzeyli regresyon analizi	25.802	1.690	
	Q(0.10)	22.088	137.279 (*)	1117.619 (*)
	Q(0.25)	23.423	44.207 (*)	240.800 (*)
	Q(0.50)	25.445		
	Q(0.75)	28.247	72.880 (*)	
	Q(0.90)	31.466	362.036 (*)	

Düzye 1 hataların çok çarpık ve Düzye 2 hataların az çarpık olduđu koşullar için Tablo 32 incelendiğinde, tüm örnekleme büyüklüklerinde çok düzeyli regresyon analizi ile elde edilen kesişim katsayısının medyandan elde edilen kesişim katsayısı ile anlamlı fark göstermediği görülmektedir. Tüm örnekleme büyüklüklerinde q10 ve q90 kantillerinden ayrı ayrı elde edilen ortalama kesişim katsayısının, medyan değeri ile elde edilen katsayıdan anlamlı olarak farklı olduđu görülmektedir. Tüm örnekleme büyüklüklerinde simetrik kantiller olan q90 ile q10 ve q25 ile q75 kantillerinden elde edilen kesişim katsayılarının da kendi aralarında anlamlı olarak farklılaştığı görülmektedir. Büyük örneklemlerde (N=500 ve

N=1000), q25 ve q75 kantil düzeylerinden ayrı ayrı elde edilen kesişim katsayıları medyan değeri ile elde edilen kesişim katsayısından anlamlı olarak farklılaşmaktadır.

Düzyer 1 hataların çok çarpık ve Düzyer 2 hataların az çarpık ($[T, X^2]$ dağılımı) olduđu koşullar için eğim katsayılarından elde edilen Wald istatistik değeri Tablo 33'te verilmiştir.

Tablo 33

Düzyer 1 Hata Terimlerinin Student-t ve Düzyer 2 Hataların X^2 Dağılımı Gösterdiği ($\varepsilon \sim [T, X^2]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değerleri

Örneklem Büyüklüğü	Analiz Yöntemi	Eğim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	6.141	0.078	
	Q(0.10)	5.493	2.519	3.369
	Q(0.25)	5.916	0.546	1.327
	Q(0.50)	6.311		
	Q(0.75)	6.556	0.171	
	Q(0.90)	6.525	0.121	
500	Çok düzeyli regresyon analizi	6.122	0.957	
	Q(0.10)	5.412	32.844 (*)	53.747 (*)
	Q(0.25)	5.911	5.798 (*)	15.515 (*)
	Q(0.50)	6.306		
	Q(0.75)	6.575	2.226	
	Q(0.90)	6.614	2.826	
1000	Çok düzeyli regresyon analizi	6.119	2.385	
	Q(0.10)	5.336	94.740 (*)	165.149 (*)
	Q(0.25)	5.877	17.558 (*)	43.774 (*)
	Q(0.50)	6.316		
	Q(0.75)	6.588	5.635 (*)	
	Q(0.90)	6.690	10.373 (*)	

Tablo 33 incelendiğinde, küçük örneklem olan N=50 için farklı kantillerden elde edilen hiçbir eğim katsayısı arasında anlamlı fark bulunamamıştır. Büyük örneklerde (N=500 ve N=1000) q10 ve q25 kantillerinden elde edilen eğim katsayısı ile medyandan elde edilen eğim katsayısı arasında anlamlı olarak farklılık bulunmuştur. Sadece N=1000 örneklem büyüklüğünde q75 ve q90 kantillerinden ayrı ayrı elde edilen eğim katsayıları, medyandan elde edilen eğim katsayısı ile anlamlı fark göstermektedir.

Düzyey 1 ve Düzyey 2 hataların çok çarpık ($[T, T]$ dağılımı) olduđu koşullar için keşişim katsayılarından elde edilen Wald istatistik değeri Tablo 34'te verilmiştir.

Tablo 34

Düzyey 1 ve Düzyey 2 Hata Terimlerinin Student-t Dağılımı Gösterdiği ($\epsilon \sim [T, T]$) Veri Yapıları İçin Keşişim Katsayılarına Ait Hesaplanan Wald Değeri

Örneklem Büyüklüğü	Analiz Yöntemi	Keşişim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	23.555	0.028	
	Q(0.10)	20.848	5.195 (*)	35.144 (*)
	Q(0.25)	21.897	1.712	9.405 (*)
	Q(0.50)	23.380		
	Q(0.75)	25.508	2.967	
	Q(0.90)	28.006	12.987 (*)	
500	Çok düzeyli regresyon analizi	23.549	0.299	
	Q(0.10)	20.910	57.884 (*)	390.973 (*)
	Q(0.25)	21.948	17.979 (*)	89.213 (*)
	Q(0.50)	23.370		
	Q(0.75)	25.205	26.038 (*)	
	Q(0.90)	27.576	134.096 (*)	
1000	Çok düzeyli regresyon analizi	23.513	1.083	
	Q(0.10)	21.872	109.182 (*)	623.542 (*)
	Q(0.25)	22.608	36.767 (*)	167.141 (*)
	Q(0.50)	23.678		
	Q(0.75)	24.945	46.054 (*)	
	Q(0.90)	26.326	199.010 (*)	

Düzyey 1 ve Düzyey 2 hataların çok çarpık olduđu koşullar için Tablo 34 incelendiğinde, tüm örneklem büyüklüklerinde çok düzeyli regresyon analizi ile elde edilen keşişim katsayısının medyandan elde edilen keşişim katsayısı ile anlamlı fark göstermediği görülmektedir. Tüm örneklem büyüklüklerinde q10 ve q90 kantillerinden ayrı ayrı elde edilen ortalama keşişim katsayıları, medyan değeri ile elde edilen katsayıdan ve simetrik kantiller olan q25 ile q75 ve q10 ile q90 kantillerinin keşişim katsayıları ise birbirlerinden anlamlı olarak farklılaşmaktadır. Büyük örneklemeler olan N=500 ve N=1000 için q25 ve q75 kantil düzeylerinden ayrı ayrı elde edilen keşişim katsayılarının medyan değeri ile elde edilen keşişim katsayısından anlamlı olarak farklılaştığı görülmektedir.

Düzyey 1 ve Düzyey 2 hataların çok çarpık ($[T, T]$ dağılımı) olduđu koşullar için eğim katsayılarından elde edilen Wald istatistik değeri Tablo 35'te verilmiştir.

Tablo 35

Düzyey 1 ve Düzyey 2 Hata Terimlerinin Student-t Dağılımı Gösterdiği ($\epsilon \sim [T, T]$) Veri Yapıları İçin Eğim Katsayılarına Ait Hesaplanan Wald Değeri

Örneklem Büyüklüğü	Analiz Yöntemi	Eğim Katsayısı	Medyan değeri ile hesaplanan Wald değeri	(1-Q(p)) kantil düzeyinde hesaplanan Wald değeri
50	Çok düzeyli regresyon analizi	6.153	0.091	
	Q(0.10)	5.216	8.250 (*)	14.506 (*)
	Q(0.25)	5.757	1.988	6.094 (*)
	Q(0.50)	6.289		
	Q(0.75)	6.736	1.164	
	Q(0.90)	6.878	1.670	
500	Çok düzeyli regresyon analizi	6.115	1.205	
	Q(0.10)	5.116	107.232 (*)	216.080 (*)
	Q(0.25)	5.714	24.139 (*)	70.788 (*)
	Q(0.50)	6.271		
	Q(0.75)	6.720	12.893 (*)	
	Q(0.90)	6.976	27.994 (*)	
1000	Çok düzeyli regresyon analizi	6.130	1.661	
	Q(0.10)	4.932	542.190 (*)	1508.387 (*)
	Q(0.25)	5.592	135.953 (*)	456.835 (*)
	Q(0.50)	6.227		
	Q(0.75)	6.790	98.857 (*)	
	Q(0.90)	7.265	304.060 (*)	

Tablo 35 incelendiğinde, tüm örneklem büyüklüklerinde q10 kantilinden elde edilen eğim katsayısı ile medyandan elde edilen eğim katsayısı; ayrıca simetrik kantiller olan q10 ile q90 ve q25 ile q75 kantillerinden elde edilen eğim katsayıları da kendi aralarında anlamlı olarak farklılaşmaktadır. Büyük örneklerde (N=500 ve N=1000) q25, q75 ve q90 kantillerinden ayrı ayrı elde edilen eğim katsayıları ile medyandan elde edilen eğim katsayısı arasında anlamlı olarak farklılaşma olduğu görülmektedir.

Tartışma

Kantil regresyon, bağımlı değişkenin dağılımındaki herhangi bir kantil ile bağımsız değişkenler arasındaki fonksiyonel ilişkinin tahmini için geliştirilmiştir (Tareghian &

Rasmussen, 2013). Kantil regresyon, özellikle koşullu kantillerin değişkenlik gösterdiği durumlarda kullanışlıdır. Kantillere bağlı olarak regresyon katsayıları belirlenir (Chen & Wei, 2005) ve her kantil düzeyinin kendi içinde verdiği bilgi düzeyi ayrı ayrı değerlendirilir. Bazı durumlarda, araştırmacının ilgisi dağılımın merkezine değil, dağılımın farkı kantil düzeylerinde olabilir. Bu durumlarda ilgilenilen kantil düzeyi ile ilgili regresyon katsayıları, model veri uyumu değerleri ve diğer kestirim değerleri o kantil özelinde değerlendirilmelidir. Bu çalışmada her kantil düzeyine ait mutlak hata, yanlılık ve model veri uyumu indeksleri ilgilenilen kantil düzeyinde farklılaşmaktadır. Bu da Staffa vd. (2019) tarafından yapılan çalışmanın içeriğinde kantil regresyon modeli uygun olduğunda, modelleme sonuçları, bağımlı değişkenin belirli bir kantil dilimi bağlamında yorumlanmalıdır ifadesi ile eşdeğer şekilde betimlenmesi gerektiğinin kanıtı olmaktadır. Bu çalışmada, N=50 örneklem büyüklüğü için tüm kantil düzeylerinde (q10, q25, q50, q75, q90) en iyi model veri uyumunu Düzey 1 ve Düzey 2 hata terimlerinin çarpıklık düzeyi en fazla olan $\varepsilon \sim [T, T]$ dağılım ile q50 kantil düzeyinde elde edilmiştir. Farcomeni (2010) çalışmasında tüm simülasyon ortamlarında, tüm bilgi kriterlerinin q50 olduğunda çok daha iyi olduğunu belirtmiştir. Bu durum Farcomeni (2010) çalışması ile paralellik göstermektedir. Bu durumun daha uç kantillerde mevcut olan daha az bilgi miktarı ile açıklayarak kantile bağlı bir sınırlandırma teriminin kullanılmasını önermektedir. Bu çalışmada da veri yapılarının sağa çarpık olmasında dolayı q10 kantilinde sağlanan bilgi miktarı ile q90 kantilinden elde sağlanan bilgi miktarında farklılıkların olduğu görülmüştür.

Kesitsel tasarımlarda genellikle normalliğe tatmin edici bir dönüşüm bulunabilirken, boylamsal araştırmalarda bağımlı değişkene ait dağılımın şekli her seferinde değişebilir (örn. çarpıklık). Böylece, global bir dönüşüm, zamana özgü dağılımları Gauss'tan uzak hale getirebilir ve zamana özgü dönüşümler, yorumlanması oldukça zor olan parametre tahminlerine yol açabileceğini belirten Marino ve Farcomeni (2015)'nin yaptığı çalışmaya paralel olarak bu çalışmada simetrik dağılımlar için daha küçük bir verimlilik kaybı ve göreceli yanlılığın dengelendiği gözlemlenmiştir. Beklendiği gibi, en küçük kareler tahmini,

hata terimi normal olarak dağıldığında medyan regresyondan daha iyi performans gösterir, ancak dağılımın yoğun kuyrukları olduğunda oldukça verimsiz tahminler vermiştir. Düzey 1 ve Düzey 2 hata dağılımlarının normallik varsayımlarının ihlal edildiği durumlarda q50 regresyon modeli klasik çok düzeyli regresyon modeline göre daha iyi performans gösteren sonuçlar vermiştir. Bu da Geraci ve Bottai (2007) ile Tzavidis vd. (2015) çalışmalarında belirttiği gibi verilerin asimetric dağılması durumunda diğer kantillerin daha bilgilendirici olacağı yorumu ve çalışmasındaki sonuçlar ile paralellik göstermektedir.

Çalışma kapsamında Düzey 1 ve Düzey 2 hatalarının tüm çarpıklıkları için q90 kantil düzeyinde daima kesişim katsayısı medyandan elde edilen kesişim katsayısı ile farklılaşmaktadır. Büyük örneklerde ise genellikle az çarpık ve çok çarpık veri yapılarında eğim katsayısı ile medyandan elde edilen eğim katsayısı farklılaşmaktadır. Bu da eğitim alanında yapılan çalışmalarda uç kantillerin etkisini gösteren çalışmalarla (Mwabu & Schultz, 1996; Eide & Showalter, 1998; Martins & Pereira, 2004; Buchinsky, 2002) paralellik göstermektedir.

Bölüm 5

Sonuç ve Öneriler

Bu bölümde önceki bölümde verilen bulgular ve yorumlar çerçevesinde elde edilen sonuç ve önerilere yer verilmiştir.

Birinci Alt Probleme İlişkin Sonuçlar

Simüle edilmiş boylamsal veri setlerinde hata düzeylerinin farklı dağılımlarına göre (Normal- Kikare- Student t(3)), her bir örneklem büyüklüğü için (50; 100; 500; 1000), çok düzeyli klasik regresyon analizinde ve kantil regresyonun farklı kantil düzeylerinde (0.10; 0.25; 0.50; 0.75; 0.90) model veri uyumu değerlerine (AIC-BIC-logLike) ilişkin sonuçlar sıralanmıştır:

- Tüm koşullar (farklı örneklem büyüklükleri ve tüm Düzey 1 ile Düzey 2 hata terimlerinin farklı dağılımları) göz önünde bulundurulduğunda, model veri uyumunun diğerlerine kıyasla en zayıf olduğu dağılımların Düzey 1 hata dağılımlarının daima az çarpık olduğu (X^2 dağılımı) durumda gözlenmiştir. Model veri uyumunun en iyi olduğu dağılımların Düzey 1 ve Düzey 2 hata terimlerinin çok çarpık (Student T dağılımı) olduğu ([T, T] dağılımı) veri yapılarında elde edilmiştir.
- Tüm örneklem büyüklüklerinde, her iki hata düzeyinin de çarpıklığı arttıkça (sırasıyla [N, N], [kk, kk], [T, T]) model veri uyumunun iyileştiği görülmüştür.
- Hata düzeylerinin dağılımları normalken, kantillerin model veri uyumu değerleri birbirine yakın; hata düzeylerinin dağılımları çarpıklaştıkça kantillere ait model veri uyumu değerleri birbirinden farklılaşmaktadır.
- Tüm koşullarda diğer kantillere göre en zayıf model veri uyumu değerleri daima q90 kantilinden elde edilmiştir.

İkinci Alt Probleme İlişkin Sonuçlar

Simüle edilmiş boylamsal veri setlerinde, farklı örneklem büyüklüğü ve farklı Düzey 1 ile Düzey 2 hata terimlerinin dağılımlarında, çok düzeyli klasik regresyon analizi ile çok düzeyli kantil regresyon analiz yöntemlerinin performanslarının nasıl değiştiğine yönelik sonuçlar sıralanmıştır:

- Tüm örneklem büyüklükleri için en düşük mutlak hata değerleri, hem çok düzeyli regresyon analizinde hem de tüm kantil düzeylerinde daima her iki düzeyin hata dağılımlarının da çok çarpık olduğu $[T, T]$ dağılımında elde edilmiştir.
- Dağılımlarının çarpıklığı normalden değiştikçe çok düzeyli regresyon analizinden elde edilen mutlak hata değerleri ile medyandan elde edilen mutlak hata değerlerinin benzer olmalarına rağmen, q50 kantil düzeyinde daima daha yansız kestirim sonuçları elde edilmiştir.
- Düzey 1 hatalar normal dağılım gösterirken, çok düzeyli regresyon analizi ile q50 kantil düzeyi karşılaştırıldığında; en az mutlak hata Düzey 2 hataların az çarpık $([N, kk])$ olduğu durumda elde edilmiştir. Ayrıca Düzey 2 hataların normal $([N, N])$ ve çok çarpık $([N, T])$ olduğu durumlarda mutlak hata değerlerinin aynı olduğu ancak her durumda da medyandan elde edilen kestirimlerin daha yansız olduğu görülmüştür. Büyük örneklerde çarpıklık arttıkça her kantil düzeyinde elde edilen mutlak hata değeri ise azalmıştır.
- Düzey 1 hatalar az çarpık (X^2 dağılımı) olduğunda, çok düzeyli regresyon analizi ile q50 kantil düzeyi karşılaştırıldığında; en az mutlak hata Düzey 2 hataların normal $([kk, N])$ olduğu durumda elde edilmiştir. Ayrıca düzey1 hataların az çarpık olduğu koşullarda, büyük örnekler etkili olmuştur. $N=500$ ve $N=1000$ için düzey 2 hata terimlerinin çarpıklığı arttıkça tüm kantil düzeylerinde mutlak hata değeri azalmıştır.

- Düzey 1 hatalar çok çarpık (Student T dağılımı) olduğunda, çok düzeyli regresyon analizi ile q50 kantil düzeyi karşılaştırıldığında; en az mutlak hata Düzey 2 hataların çok çarpık ([T,T]) olduğu durumda elde edilmiştir. Ayrıca düzey 2 hata terimlerinin çarpıklığının artması daima tüm kantil düzeylerinde daha az mutlak hata değerinin hesaplanmasını sağlamıştır.
- Düzey 2 hatalar normal dağılım gösterirken, düzey 2 hataların çarpıklığı arttıkça büyük örneklerde (N=500 ve N=1000) daha düşük mutlak hata değerleri elde edilmiştir.
- Düzey 2 hatalar az çarpık (X^2 dağılımı) olduğunda, Düzey 1 hata terimlerinin çarpıklığı arttıkça tüm kantil düzeylerinde elde edilen mutlak hata değerlerinde azalma olduğu görülmüştür.
- Düzey 2 hatalar çok çarpık (Student T dağılımı) olduğunda, Düzey 1 hata terimlerinin çarpıklığı arttıkça tüm kantil düzeylerinde elde edilen mutlak hata değerlerinde azalma olduğu görülmüştür.
- Tüm dağılımlarda birbirinin simetriği olan kantiller (q10 ile q90 ve q25 ile q75) benzer mutlak hata değerleri üretmelerine rağmen yanlılık değerleri zıt işaretli olarak hesaplanmıştır.
- Tüm dağılımlarda q90 kantilinden elde edilen mutlak hata ve yanlılık değeri diğer kantillerden elde edilen değerlere göre en yüksek olarak hesaplanmıştır.

Üçüncü Alt Probleme İlişkin Sonuçlar

Simüle edilmiş boylamsal veri setlerinde, farklı Düzey 1 ile Düzey 2 hata terimlerinin dağılımları ve farklı örneklem büyüklüklerinde, çok düzeyli klasik regresyon analizi ile çok düzeyli kantil regresyon analiz yöntemlerinden elde edilen ortalama kesişim ve eğim katsayılarının nasıl değiştiğine yönelik sonuçlar sıralanmıştır:

- Tüm örneklem büyüklüklerinde çok düzeyli regresyon analizi ile elde edilen ortalama kesişim ve ortalama eğim katsayısı ile medyandan elde edilen ortalama kesişim ve ortalama eğim katsayıları arasında anlamlı olarak fark bulunmadığı görülmektedir.
- Düzey 1 ve düzey 2 hataların tüm dağılımlarında ve tüm örneklem büyüklüklerinde (tüm koşullarda), q10 ile q90'dan elde edilen kesişim katsayıları ile q25 ve q75'ten elde edilen kesişim katsayılarının kendi aralarında anlamlı farklılık gösterdiği görülmüştür.
- Küçük örneklem olan N=50 için tüm dağılımlardan elde edilen kesişim katsayısı ile medyandan elde edilen kesişim katsayısının farklılaşması sadece uç kantiller olan q10 ve q90 kantil düzeylerinde gerçekleşmiştir.
- Büyük örneklerde (N=500 ve N=1000) tüm kantil düzeylerinden elde edilen kesişim katsayıları ile medyandan elde edilen kesişim katsayıları, düzey 1 ve düzey 2'nin tüm hata dağılımları için anlamlı olarak farklılaşmıştır.
- Düzey 1 hataların normal dağılım gösterdiği tüm koşullarda, q90'dan elde edilen kesişim katsayısının medyandan elde edilen kesişim katsayısından anlamlı olarak farklılaştığı görülmüştür. Düzey 1 hataların normal dağılım gösterdiği büyük örneklerde (N=500 ve N=1000) q10 ile q90 ve q25 ile q75 kantillerinden elde edilen eğim katsayılarının kendi aralarında anlamlı olarak farklılaştığı görülmüştür.
- Düzey 1 hataların az çarpık olduğu (X^2 dağılımı) tüm koşullarda, q25 ve q75'ten elde edilen kesişim katsayılarının kendi aralarında anlamlı farklılık gösterdiği görülmüştür.
- Düzey 1 hataların az çarpık olduğu (X^2 dağılımı) tüm koşullarda q10 ve q90'dan ayrı ayrı elde edilen kesişim katsayılarının medyandan elde edilen kesişim katsayısından anlamlı olarak farklılaştığı görülmüştür.

- Düzey 1 hatalar az çarpıkken Düzey 2 hataların çarpıklığı arttıkça büyük örneklerde (N=500 ve N=1000) kantillerden elde edilen eğim katsayıları arasında farklılaşmanın arttığı görülmüştür. Yine büyük örneklerde q10 ile q90 kantillerinden elde edilen eğim katsayılarının kendi aralarında anlamlı olarak farklılaştığı görülmüştür.
- Düzey 1 hataların çok çarpık olduğu (Student t dağılımı) tüm koşullarda, q25 ve q75'ten elde edilen kesişim katsayılarının kendi aralarında anlamlı farklılık gösterdiği görülmüştür.
- Düzey 1 hataların çok çarpık olduğu (Student t dağılımı) tüm koşullarda q10 ve q90'dan ayrı ayrı elde edilen kesişim katsayılarının medyandan elde edilen kesişim katsayısından anlamlı olarak farklılaştığı görülmüştür.
- Düzey 1 hataların çok çarpık olduğu (Student t dağılımı) büyük örneklerde (N=500 ve N=1000) q10 ile q90 ve q25 ile q75 kantillerinden elde edilen eğim katsayılarının kendi aralarında anlamlı olarak farklılaştığı görülmüştür.

Öneriler

- Çalışma örneklem büyüklükleri ve hata dağılımlarının çarpıklık düzeyleri çeşitlendirilerek tekrarlanabilir.
- Sadece uç değerlerde (q10 ve q90) daha fazla bilgi verecek şekilde yeni dağılımlar elde edilebilir.
- Simülasyon denklemleri, kullanılan parametrelerin sabit değerlerden oluşturulmasının yerine farklı dağılımlardan oluşarak düzenlenmesi ile farklı karşılaştırmalar yapılabilir.
- Çok düzeyli kantil regresyonun farklı tahmin metodlarının değerlendirme kriterleri göz önünde bulundurularak karşılaştırmalar yapılabilir.

- Endojen ve egzogen deęişkenler bağlamında gizil deęişkenlerle çok düzeyli kantil yapısal eşitlik modellemeleri kurularak ölçüm modelleri karşılaştırılabilir.
- Kantil regresyon çalışmalarda MTK uygulamalarıyla kullanılabilir.
- Kantil regresyonda eksik veri problemine yönelik çalışmalar gerçekleştirilebilir.
- Büyük örneklemeler içeren ulusal ve uluslararası geniş ölçekli sınavlar bazında, farklı deęişkenler ele alınarak çok düzeyli kantil regresyon uygulamaları gerçekleştirilebilir. Bu örneklerde uç kantil deęerlerinden elde edilen veriler incelenebilir.
- Küçük örneklemelere sahip okullarda ve sınıf içi ölçmelerden elde edilen gözlemler ile çalışma tekrarlanabilir ve bu gözlemlerin uç kantillerde yer alan farklı özelliklere sahip öğrencilerin profilleri incelenerek eğitim politikaları oluşturulabilir.
- Öğrencilerin ilgi, tutum, beceri vb. gibi özellikleri boylamsal olarak incelenerek kantil ve çok düzeyli kantil regresyon yöntemleri kullanılabilir.

Kaynaklar

- Acar, T. ve Öğretmen, T. (2012). Çok düzeyli istatistiksel yöntemler ile 2006 PISA Fen Bilimleri performansının incelenmesi. *Eğitim ve Bilim*, 37(163), 121-137.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. B. N. Petrov ve F. Caski (Ed.), *Proceedings of the second international symposium on information theory*. Akademiai Kaido.
- Akgün, Ç. (2017). *Türkiye'de eğitimin ücretler üzerindeki etkisi: Dilim regresyon uygulaması* (Yüksek lisans tezi), Dokuz Eylül Üniversitesi, İzmir.
- Akyol, T. (2013). *Kantil regresyon modeli yardımıyla ülkelerin insani gelişmişlik indeksi üzerinde etkili olan faktörlerin incelenmesi* (Yüksek lisans tezi), Atatürk Üniversitesi, Erzurum.
- Alejo, J., Gabrielli, M.F. & Sosa-Escudero, W. (2014). The distributive effects of education: An unconditional quantile regression approach. *Revista de Análisis Económico – Economic Analysis Review*, 29(1), 53-76. <http://doi.org/10.4067/S0718-88702014000100003>
- Allison, P.D. (1984) *Event History Analysis*. Sage Publications, Beverly Hills. <https://doi.org/10.4135/9781412984195>
- Ally, A. (2010). *Quantile-based Methods for Prediction, Risk Measurement and Inference*. Brunel University.
- Alpar, R. (2013). *Basit doğrusal regresyon analizi: Spor, sağlık ve eğitim bilimlerinden örneklerle uygulamalı istatistik ve geçerlik-güvenirlilik*. Detay Yayıncılık.
- Altın Yavuz, A. & Gündoğan Aşık, E. (2017). Quantile regression. *International Journal of Engineering Research and Development*, 9(2), 137-146. <https://doi.org/10.29137/umagd.352530>

- Anderson, T. W. & Hsiao, C., (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1), 47-82.
- Aşkar, P. ve Yurdugül, H. (2009). Örtük büyüme modellerinin eğitim araştırmalarında kullanımı. *İlköğretim Online*, 8(2), 534-555.
- Atar, B. (2010). Basit doğrusal regresyon analizi ile hiyerarşik doğrusal modeller analizinin karşılaştırılması. *Eğitim ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 78-84.
- Atar, H. Y. (2014). Öğretmen niteliklerinin TIMSS 2011 fen başarısına çok düzeyli etkileri. *Eğitim ve Bilim Geniş Ölçekli Test Uygulamaları*, (39),172.
- Austin, P. C. & Schull, M. J. (2003). Quantile regression: A statistical tool for out-of-hospital research. *Acad. Emerg. Med*, 10, 789–797. <https://doi.org/10.1111/j.1553-2712.2003.tb00075.x>
- Barnett, V. & Lewis, T. (1994). *Outliers in statistical data* (3. Baskı). John Wiley Sons. Academic Publishers.
- Birch, E. R. & Miller, P. W. (2006). Student outcomes at university in Australia: A quantile regression approach. *Australian Economic Papers*, 45, 1-17. <https://doi.org/10.1111/j.1467-8454.2006.00274.x>
- Birkes, D. & Dodge, Y. (1993). *Alternative methods of regression*. John Wiley Sons.
- Bijleveld, C. C.J.H., Van der Kamp, L. J. T., Mooijaart, A., Van der Kloot, W. A., Van der Leeden, R., & Van der Burg, E. (1998). *Longitudinal data analysis: Designs, models and methods*. Sage Publications Ltd.
- Bruneau, L., Billaud, E., Raffi, F. & Hanf, M. (2017). Factors associated with the level of CD4 cell counts at HIV diagnosis in a French cohort: a quantile regression approach. *International Journal of STD & AIDS*, 28(4), 397-403. <https://doi.org/10.1177/0956462416650980>

- Buchinsky, M. (2002). Quantile regression with sample selection: Estimating women's return to education in the U.S. B. Fitzenberger, R. Koenker ve J. A. F. Machado (Ed.), *Economic applications of quantile regression. Studies in empirical economics*. Physica, Heidelberg. https://doi.org/10.1007/978-3-662-11592-3_4
- Buchinsky, M. (1994). Changes in the US wage structure 1963-1987: Application of quantile regression. *Econometrica*, 62(2), 405-458. <https://doi.org/10.2307/2951618>
- Budría, S. & Moro-Egido, A. I. (2008). Education, educational mismatch, and wage inequality: Evidence for Spain. *Economics of Education Review*, 27(3), 332-34. <https://doi.org/10.1016/j.econedurev.2006.10.005>
- Burnham, K. P. & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33, 261–304. <https://doi.org/10.1177/0049124104268644>
- Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2014). *Bilimsel araştırma yöntemleri* (16. Baskı). Pegem Akademi Yayınları.
- Cade, B. S. (2011). Estimating equivalence with quantile regression. *Ecological Applications*, 21, 281-289. <https://doi.org/10.1890/09-2060.1>
- Cade, B. S., Noon, B. R. & Flather, C. H. (2005). Quantile regression reveals hidden bias and uncertainty in habitat models. *Ecology*, 86, 786–800. <http://doi.org/10.1890/04-0785>
- Cai, Z. & Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics*, 167(2), 413-425. <https://doi.org/10.1016/j.jeconom.2011.09.025>
- Cai, Z. & Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, 103(484), 1595-1608. <http://doi.org/10.1198/016214508000000977>

- Carroll, M. (2019, September). *The use of longitudinal survey data in education research* [Conference presentation abstract]. The Conference of the British Educational Research Association. University of Manchester, Manchester, UK.
- Caruana, E. J., Roman, M., Hernández-Sánchez, J. & Solli, P. (2015). Longitudinal studies. *J Thorac Dis*, 7(11), E537-40. <https://doi.org/10.3978/j.issn.2072-1439.2015.10.63>
- Chen, C. & Wei, Y. (2005). Computational issues for quantile regression. *Special Issue on Quantile Regression and Related Methods*, 67(2), 399-417. <http://doi.org/10.2307/25053439>
- Chen, Y., Tian, M., Yu K. & Pan, J. (2014). Composite hierarchical linear quantile regression. *Acta Mathematicae Applicatae Sinica, English Series*, 30(1), 49–64. <http://doi.org/10.1007/s10255-014-0267-1>
- Chernozhukov, V. & Hansen, C. (2008). Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, 142, 379–398. <https://doi.org/10.1016/j.jeconom.2007.06.005>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.)*. Lawrence Erlbaum Associates Publishers.
- Costanzo, A. & Desimoni, M. (2017). Beyond the mean estimate: A quantile regression analysis of inequalities in educational outcomes using INVALSI survey data. *Large-scale Assess Educ*, 5, 14. <https://doi.org/10.1186/s40536-017-0048-4>
- Cox, D.R. & Oakes, D. (1994). *Analysis of Survival Data*. Chapman & Hall/Crc.
- Çelik, O. (2013). *Türkiye’de kamu ve özel sektörde ücret farklılıklarını belirleyen etmenler: Kantil Regresyon Yaklaşımı* (Yüksek lisans tezi), Celal Bayar Üniversitesi, Manisa.
- Çiçek, Z. (2022). *Bayesyen Kantil yapısal eşitlik modellemesi* (Doktora tezi). Uludağ Üniversitesi, Bursa.

- Çobanoğlu Aktan, D. (2017). *Examination of PISA 2015 results for Turkish students with quantile regression* [Conference presentation abstract]. European Conference on Educational Research, Kopenhag, Denmark.
- Delyon, B., Lavielle, M. & Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist*, 27(1), 94-128. <https://doi.org/10.1214/aos/1018031103>
- Demirez, G. (2018). *Bazı değişkenlerin fen başarı puanına etkisi: PISA 2015 Türkiye, Singapur ve Almanya örneği*. (Yüksek lisans tezi), Akdeniz Üniversitesi, Antalya.
- Díez, J., Gatt, S. & Racionero, S. (2011). Placing Immigrant and Minority Family and Community Members at the School's Centre: the role of community participation. *European Journal of Education*, 46: 184-196. <https://doi.org/10.1111/j.1465-3435.2011.01474.x>
- Diggle, P., Diggle, P.J. & Heagerty, P., (2002). *Analysis of Longitudinal Data*. Oxford University Press, Oxford.
- Dinov, I. D. (2018). *Data Science and Predictive Analytics: Biomedical and Health Applications Using R*. Springer.
- Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society, Series B*, 57, 45–97. <http://doi.org/10.2307/2346087>
- Dunn, L. M., & Markwardt, F. C. (1970). *Peabody individual achievement test*. Circle Pines, Mont.: American Guidance Service.
- Eide, E. & Showalter, M. H. (1998). The effect of school quality on student performance: A quantile regression approach. *Economics Letters*, 58(3), 345-350. [https://doi.org/10.1016/S0165-1765\(97\)00286-3](https://doi.org/10.1016/S0165-1765(97)00286-3)
- Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics*, 3, 1–21. <https://doi.org/10.2307/3001534>

- Fan, X. (2001). Parental involvement and students' academic achievement: A growth modeling analysis. *The Journal of Experimental Education*, 70(1), 27-61.
<https://doi.org/10.1080/00220970109599497>
- Farcomeni, A. (2012). Quantile regression for longitudinal data based on latent Markov subject-specific parameters. *Stat. Comput*, 22, 141–152.
<https://doi.org/10.1007/s11222-010-9213-0>
- Feinberg, R.A. & Rubright, J.D. (2016). Conducting Simulation Studies in Psychometrics. *Educational Measurement: Issues and Practice*, 35: 36-49.
<https://doi.org/10.1111/emip.12111>
- Feriver, Ş. ve Arık, B. M. (2021). *Eğitim izleme raporu 2021: Eğitimin içeriği*. Eğitim Reformu Girişimi. <https://www.egitimreformugirisimi.org/egitim-izleme-raporu-2021-egitimin-icerigi/>
- Fitzmaurice, G. M., Laird, N. M. & Ware, J. H. (2004). *Applied longitudinal analysis*, Wiley series in probability and statistics. John Wiley & Sons, inc. Publication.
- Fu, L. & Wang, Y. G. (2012). Quantile regression for longitudinal data with a working correlation model. *Computational Statistics & Data Analysis*, 56, 2526–2538.
<https://doi.org/10.1016/j.csda.2012.02.005>
- Galarza M., Bandyopadhyay, D. & Lachos, V. (2017). Quantile regression in linear mixed models: a stochastic approximation EM approach. *Stat Interface*, 10(3):471-482.
<https://doi.org/10.4310%2FSII.2017.v10.n3.a10>
- Galvao, A. F. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics.*, 164(1), 142–157. doi:
<https://doi.org/10.1016/j.jeconom.2011.02.016>
- Galvao, A. F. & Montes-Rojas, G. V. (2010). Penalized quantile regression for dynamic panel data. *J. Stat. Plan. Inference*, 140, 3476–3497.
<http://doi.org/10.1016/j.jspi.2010.05.008>

- Gardiner, J.C., Luo, Z. & Roman, L.A. (2009). Fixed effects, random effects and GEE: What are the differences?. *Statistics in Medicine*, 28: 221-239.
<https://doi.org/10.1002/sim.3478>
- Geiser, C. (2013). *Data analysis with Mplus*. Guilford Press.
- Gelman, A. (2005). Analysis of variance—why it is more important than ever. *The Annals of Statistics*, 33(1), 1–53.
- Gentle, J. E. (1985). Monte Carlo methods. S. Kotz ve N. L. Johnson (Ed.), *The encyclopedia of statistical sciences*, p:612–617. Wiley.
- Geraci, M. (2013). Estimation of regression quantiles in complex surveys with data missing at random: An application to birthweight determinants. *Statistical Methods in Medical Research*, 22(4). <http://doi.org/10.1177/0962280213484401>
- Geraci, M. & Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics*, 8(1), 140-154. DOI: <https://doi.org/10.1093/biostatistics/kxj039>
- Geraci, M. & Bottai, M. (2013). Linear quantile mixed models. *Statistics and Computing*, 24, 461–479. <https://doi.org/10.1007/s11222-013-9381-9>
- Giambona, F. & Porcu, M. (2015). Student background determinants of reading achievement in Italy. A quantile regression analysis. *International Journal of Educational Development*, 44, 95-107.
<https://doi.org/10.1016/j.ijedudev.2015.07.005>
- Gilchrist, W. G. (2000). *Statistical modelling with quantile functions*. Chapman & Hall/CRC.
- Goldstein, H. (1995). Hierarchical Data Modeling in the Social Sciences. *Journal of Educational and Behavioral Statistics*, 20(2), 201–204.
<https://doi.org/10.3102/10769986020002201>

- Goodal, C. (1983). *Examining residuals: Understanding robust and exploratory data analysis*. John Wiley Sons.
- Gökçe, A. (2013). *Türk bankacılık sisteminde döviz kuru değişmelerinin bankaların mali tablolarına etkileri* (Yüksek lisans tezi), Pamukkale Üniversitesi, Denizli.
- Green, B. F., Jr., & Tukey, J. W. (1960). Complex analyses of variance: General problems. *Psychometrika*, 25, 127–152. [https:// doi/10.1007/BF02288577](https://doi/10.1007/BF02288577)
- Grimm, K. J., Steele, J. S., Mashburn, A. J., Burchinal, M., & Pianta, R. C. (2010). Early behavioral associations of achievement trajectories. *Developmental Psychology*, 46(5), 976–983. [https:// doi/10.1037/a0018878](https://doi/10.1037/a0018878)
- Gürsoy, İ. (2019). *Likidite riskinin panel kantil modeller ile incelenmesi* (Yüksek lisans tezi), Marmara Üniversitesi, İstanbul.
- Hamed, H. M. (2012). *The relationship between self-efficacy beliefs and social emotional competence in at-risk girls* (Doctoral dissertation thesis). The State University of New Jersey, New Jersey.
- Hamilton, L. C. (1992). *Regression with graphics: A second course in applied statistics* (First Edition).
- Hanushek, E. A., Link, S. & Wößmann, L. (2011). *Does school autonomy make sense everywhere? Panel estimates from PISA*. Discussion Paper series, Forschungsinstitut zur Zukunft der Arbeit. <http://nbn-resolving.de>
- Hao, L. & Naiman, D. Q. (2007). *Quantile regression*. Sage Publication.
- Harding, M. ve Lamarche, C. (2009). A quantile regression approach for estimating panel data models using instrumental variables. *Economics Letters*, 104, 133–135. <https://doi.org/10.1016/j.econlet.2009.04.025>
- Hartog, J., Pereira, P. T. & Vieira, J. A. C. (2001) Changing returns to education in Portugal during the 1980s and early 1990s: OLS and quantile regression

- estimators. *Applied Economics*, 33(8), 1021-1037. <https://doi.org/10.1080/00036840122679>
- He, X. ve Liang, H. (2000). Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statistica Sinica*, 129-140.
- He, X. & Shi, P. (1996). Bivariate tensor-product B-splines in a partly linear model. *Journal of Multivariate Analysis*, 58(2), 162-181. <https://doi.org/10.1006/jmva.1996.0045>
- He, X., Zhu, Z. & Fung, W. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika*, 89, 579–590. <http://doi.org/10.1093/biomet/89.3.579>
- Heck, R. H. (2001). *Multilevel modeling with sem: New developments and techniques in structural equation modeling*. Lawrence Erlbaum Associates.
- Heck, R.H., Thomas, S.L., & Tabata, L.N. (2010). *Multilevel and Longitudinal Modeling with IBM SPSS* (1st ed.). Routledge. <https://doi.org/10.4324/9780203855263>
- Heck, R. H. & Thomas, S.,L. (2020). *An Introduction to Multilevel Modeling Techniques: MLM and SEM Approaches (Quantitative Methodology)*. Routledge.
- Hedeker, D., Flay, B. R., & Petraitis, J. (1996). Estimating individual influences of behavioral intentions: An application of random-effects modeling to the theory of reasoned action. *Journal of Consulting and Clinical Psychology*, 64(1), 109–120. <https://doi.org/10.1037/0022-006X.64.1.109>
- Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge/Taylor & Francis Group.
- Honda, T. (2004). Quantile regression in varying coefficient models. *Journal of Statistical Planning and Inference*, 121(1), 113-125. [http://doi.org/10.1016/S0378-3758\(03\)00110-1](http://doi.org/10.1016/S0378-3758(03)00110-1)
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. Routledge.

- Hsiao, C. (2007). Panel data analysis—advantages and challenges. *TEST* 16, 1–22. <https://doi.org/10.1007/s11749-007-0046-x>.
- Hyun Sook, O. (2017). Bayesian quantile regression analysis of private education expenses for high school students in Korea, *Journal of the Korean Data and Information Science Society*, 28(6), 1457-1469. <http://doi.org/10.7465/jkdi.2017.28.6.1457>
- Jiang, X. J., Jiang, J. C. & Song, X. Y. (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statist. Sinica*, 22, 1479-1506. <http://doi.org/10.5705/ss.2010.203>
- Joensuu, S., Viljamaa, A., Varamäki, E. & Tornikoski, E. (2013). Development of entrepreneurial intention in higher education and the effect of gender – a latent growth curve analysis. *Education Training*, Vol. 55, No. 8/9, pp. 781-803. <https://doi.org/10.1108/ET-06-2013-0084>
- Kai, B., Li, R. & Zou, H. (2010). Local composite quantile regression smoothing: An efficient and safe alternative to local polynomial regression. *J. Royal Stat. Soc. Series B (Statistical Methodology)*, 72, 49–69. <https://doi.org/10.1111/j.1467-9868.2009.00725.x>
- Kai, B., Li, R. & Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Annal. Stat*, 39, 305–332. <http://doi.org/10.1214/10-AOS842>
- Kalbfleisch, J. D. & Prentice, R.L. (1980). *The statistical analysis of failure time data*. John Wiley & Sons, Inc., New York. <https://doi.org/10.2307/3315078>
- Karacan, S. (2018). *Cezalı kantil regresyon modelinin yarı-parametrik bağıllılaşımli etkiler yöntemiyle tahmini ve bir uygulama* (Doktora tezi), Dumlupınar Üniversitesi, Kütahya.

- Karlsson, A. (2008). Nonlinear quantile regression estimation of longitudinal data. *Commun. Stat., Simul. Comput*, 37, 114–131. <https://doi.org/10.1080/03610910701723963>
- Kelly, J. R., & McGrath, J. E. (1988). *On time and method*. Sage Publications, Inc.
- Keskin, B. (2012). *Sağlam bir çıkarsama yöntemi: Kantil regresyon* (Yüksek lisans tezi), Muğla Sıtkı Koçman Üniversitesi, Muğla.
- Kılınç, N. (2020). *İşgücü piyasasında toplumsal cinsiyet eşitsizliği ve gelir dağılımı* (Doktora tezi), Hacettepe Üniversitesi, Ankara.
- Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1), 74–89. <https://doi.org/10.1016/j.jmva.2004.05.006>
- Koenker, R. (2005). *Quantile regression*. Cambridge University Press.
- Koenker, R. & Basset, G. S. (1978). Regression quantiles. *Econometrica*, 46, 33–50. <https://doi.org/10.2307/1913643>
- Koenker, R. & Hallock, K. (2001). Quantile regression. *Journal of Economic Perspectives*, 143–156. <http://doi.org/10.1257/jep.15.4.143>
- Koenker, R. & Machado, J. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94(448), 1296–1310. <https://doi.org/10.1080/01621459.1999.10473882>
- Koenker, R. & Park, J. B. (1996). An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, vol. 71, issue 1-2, 265-283.
- Kong, D., Xue, K., Yao, F. & Zhang, H. H. (2016). Partially functional linear regression in high dimensions. *Biometrika*, 103(1), 147-159. <https://doi.org/10.1093/biomet/asv062>

- Koşan, İ. (2014). *OECD ülkelerinde dış ticaret hadlerini etkileyen değişkenlerin panel kantil regresyon modelleri ile incelenmesi* (Yüksek lisans tezi), Marmara Üniversitesi, İstanbul.
- Kozubowski, T. & Podgorski, K. (2000). Asymmetric laplace distributions. *The Mathematical Scientist*, 25, 37-46. https://doi.org/10.1007/978-1-4612-0173-1_3
- Kreft, I., & de Leeuw, J. (1998). *Introducing multivlevel modeling*. Sage Publications, Inc. <https://doi.org/10.4135/9781849209366>
- Kutlu, G. (2016). *Determinants of household savings behavior in Turkey* (Yüksek lisans tezi), İstanbul Teknik Üniversitesi, İstanbul.
- Kwok, O. M., Lai, M. H., Tong, F., Lara-Alecio, R., Irby, B., Yoon, M. & Yeh, Y. C. (2018). Analyzing complex longitudinal data in educational research: A demonstration with project english language and literacy acquisition (ella) data using xxM. *Front Psychol.* 5(9), 790. <https://doi.org/10.3389/fpsyg.2018.00790>
- Laird, N. M., & Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, 38(4), 963–974. <https://doi.org/10.2307/2529876>
- Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometrics.*, 157, 396–498. <http://doi.org/10.1016/j.jeconom.2010.03.042>
- Lamichhane, D. K., Lee, S. Y., Ahn, K., Kim, K. W., Shin, Y. H., Suh, D. I. & Kim, H. C. (2020). Quantile regression analysis of the socioeconomic inequalities in air pollution and birth weight. *Environment International*, 142, 105875. <https://doi.org/10.1016/j.envint.2020.105875>
- LaMotte, L. R. (1983). *Fixed-random-and mixed-effects models*, Encyclopedia of Statistical Sciences, Wiley, New York, 137–141.

- Lee, S. (2003). Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric Theory*, 19, 1-31. <https://doi.org/10.1017/S0266466603191013>
- Lee, D. & Neocleous, T. (2010). Bayesian quantile regression for count data with application to environmental epidemiology. *Journal of the Royal Statistical Society Series C Applied Statistics*, 59, 905–920. <http://doi.org/10.1111/j.1467-9876.2010.00725.x>
- Lee, M., Rahbar, M. H., Gensler, L. S., Brown, M., Weisman, M. & Reveille, J. D. (2020). A latent class based imputation method under Bayesian quantile regression framework using asymmetric laplace distribution for longitudinal medication usage data with intermittent missing values. *Journal of Biopharmaceutical Statistics*, 30(1), 160-177. <https://doi.org/10.1080/10543406.2019.1684306>
- Lehr, D.K. & Newton, J.M. (1978). Time Series and Cross-Sectional Investigations of The Demand For Higher Education, *Economic Inquiry*, 16, 411-422. <https://doi.org/10.1111/j.1465-7295.1978.tb00511.x>
- Liu, Y. & Bottai, M. (2009). Mixed-Effects Models for Conditional Quantiles with Longitudinal Data. *The International Journal of Biostatistics*, 5(1). <https://doi.org/10.2202/1557-4679.1186>
- Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G. & Zhao, L. P. (1997). Quantile regression methods for longitudinal data with drop-outs: Application to CD4 cell counts of patients infected with the human immunodeficiency virus. *J. R. Stat. Soc., Ser. C, Appl. Stat*, 46, 463–476. <https://doi.org/10.1111/1467-9876.00084>
- Longford, N. T. (1995). *Random coefficients models*. Oxford University Press.
- Luke, D. A. (2004). *Multilevel modeling*. SAGE Publications, Inc.
- Lynn, P. (2009). *Methodology of longitudinal surveys*. John Wiley & Sons, Ltd.

- Machado, J. A. F. & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *J. Appl. Econom*, 20, 445–465. <https://doi.org/10.1002/jae.788>
- Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology*, 89(1), 41–54. <https://doi.org/10.1037/0022-0663.89.1.41>
- Marston, D. (1988). The Effectiveness Of Special Education: A Time Series Analysis Of Reading Performance In Regular And Special Education Settings. *The Journal of Special Education*, 21(4), 13–26. <https://doi.org/10.1177/002246698802100405>
- Martins, P. S. & Pereira, P. T. (2004). Does education reduce wage inequality? Quantile regression evidence from 16 countries. *Labour Economics*, 11(3), 355-371. <https://doi.org/10.1016/j.labeco.2003.05.003>
- McCoach, B. D., Rifken, G. G., Newton, S. D., Li, X., Kooker, J., Yomtov, D., Gambino, A. J. & Bellara, A. (2018). Does the package matter? A comparison of five common multilevel modeling software packages. *Journal of Educational and Behavioral Statistics*, 43(5), 594-627. <http://doi.org/10.3102/1076998618776348>
- McDonald, S., Ortaglia, A., Supino, C., Kacka, M., Clenin, M. & Bottai, M. (2017). Fitness adjusted racial disparities in central adiposity among women in the USA using quantile regression. *Obes Sci Pract*, 3(2), 153-161. <https://doi.org/10.1002/osp4.110>
- Menard, S. (1991). *Longitudinal research*. Sage Publications, Inc.
- Milli Eğitim Bakanlığı. (2021). 20. MEB Şura Tavsiye Kararları. https://www.meb.gov.tr/meb_iys_dosyalar/2022_02/20144118_Yuzyuze_Egitime_Donuste_180_Gun-sayfalar-104-113_1.pdf

- Mitchell, T. R. & James, L. R. (2001). Building Better Theory: Time and the Specification of When Things Happen. *The Academy of Management Review*, 26(4), 530-547. <https://doi.org/10.2307/3560240>
- Mosteller, F. & Tukey, J. W. (1977). *Data analysis and regression: A second course in statistics*. Addison-Wesley.
- Muller, B. J., Cade, B. S. & Schwarzkopf, L. (2018). Effects of environmental variables on invasive amphibian activity: Using model selection on quantiles for counts. *Ecosphere*, 9(1), e02067. <https://doi.org/10.1002/ecs2.2067>
- Mwabu, G. & Schultz, T. (1996). Education returns across quantiles of the wage function: Alternative explanations for returns to education by race in South Africa. *The American Economic Review*, 86(2), 335-339.
- Nese, J., Lai, C. & Anderson, D. (2002). *A Primer on Longitudinal Data Analysis In Education*. University of Oregon.
- Newmann, F. M., Smith, B., Allensworth, E. & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis*, 23(4), 297-321. <https://doi.org/10.3102/01623737023004297>
- No, F., Taniguchi, K. & Hirakawa, Y. (2016). School dropout at the basic education level in rural Cambodia: Identifying its causes through longitudinal survival analysis, *International Journal of Educational Development*, Volume 49, pages 215-224, ISSN 0738-0593, <https://doi.org/10.1016/j.ijedudev.2016.03.001>.
- Oberfield, Z. (2014). Public Management in Time: A Longitudinal Examination of the Full Range of Leadership Theory. *Journal of Public Administration Research and Theory*. 24, 407-429. <http://doi.org/10.1093/jopart/mus060>
- Osborn, J. W., (2000). Prediction in Multiple Regression, *Practical Assessment, Research, and Evaluation*, 7(2).

- Palardy, G.J. (2008). Differential school effects among low, middle, and high social class composition schools: a multiple group, multilevel latent growth curve analysis, *School Effectiveness and School Improvement*, 19:1, 21-49, <https://doi.org/10.1080/09243450801936845>
- Parzen, E. (2004). Quantile probability and statistical data modeling. *Statistical Science*, 19(4), 652-662. <https://doi.org/10.1214/088342304000000387>
- Phan, H.P. (2012). The Development of English and Mathematics Self-Efficacy: A Latent Growth Curve Analysis, *The Journal of Educational Research*, 105:3, 196-209, <https://doi.org/10.1080/00220671.2011.552132>
- Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer Verlag.
- Piombo, S. (2013). *Multilevel analysis in household surveys: An application to health condition data*. Alma Mater Studiorum, Universita Di Bologna.
- Plank, S. B., DeLuca, S., & Estacion, A. (2008). High School Dropout and the Role of Career and Technical Education: A Survival Analysis of Surviving High School. *Sociology of Education*, 81(4), 345–370. <https://doi.org/10.1177/003804070808100402>
- Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: The theory, design, and analysis of change. *Journal of Management*, 36(1), 94–120. <https://doi.org/10.1177/0149206309352110>
- Popper, K. R. (1959). *The logic of scientific discovery*. Basic Books.
- Pringle, R. M. (2008). Elephants as agents of habitat creation for small vertebrates at the patch scale. *Ecology*, 89, 26-33. <https://doi.org/10.1890/07-0776.1>
- Quené, H. & Van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, 43(1), 103-121. <http://doi.org/10.1016/j.specom.2004.02.004>

- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111-163. <https://doi.org/10.2307/271063>
- Rahman, M. A. & Vossmeier, A. (2019). Estimation and applications of quantile regression for binary longitudinal data. *Topics in Identification, limited dependent variables, partial observability, experimentation, and flexible modeling: Part B advances in econometrics*, 40, 157-191. Emerald Publishing Limited. <https://doi.org/10.1108/S0731-90532019000040B009>
- Rangvid, B. (2003). *Educational peer effects*. Quantile Regression Evidence from Denmark with PISA 2000 Data.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models. Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Reich, B. J., Bondell, H. D. & Wang, H. J. (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics*, 11, 337–352. <http://doi.org/10.1093/biostatistics/kxp049>
- Robert, C.P. & Casella, G., (2004). *Monte Carlo Statistical Methods*. Springer, Berlin. <http://doi.org/10.1007/978-1-4757-4145-2>
- Rodriguez, J. P., Barros, C. P. & Vieira, J. A. C. (2008). What a quantile approach can tell us about returns to education in Europe. *Education Economics*, 16(4), 391-410. <https://doi.org/10.1080/09645290801977033>
- Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, 92(3), 726–748. <https://doi.org/10.1037/0033-2909.92.3.726>
- Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika*, 50(2), 203–228. <https://doi.org/10.1007/BF02294247>
- Rousass, G. G. (1997). *A course in mathematical statistics*. Academic Press.

- Rousseeuw, P. & Leroy, A. (1987). *Robust regression and outlier detection*. John Wiley Sons.
- Rutter, M. (1994). Beyond longitudinal data: Causes, consequences, changes, and continuity. *Journal of Consulting and Clinical Psychology*, 62(5), 928–940. <https://doi/10.1037/0022-006X.62.5.928>
- Ryan, T. P. (1997). *Modern regression methods*. John Wiley Sons.
- Saçaklı, İ. (2005). *Kantil regresyon ve alternatif regresyon modelleri ile karşılaştırılması* (Yüksek lisans tezi). Marmara Üniversitesi, İstanbul.
- Sakellariou, C. (2006). Benefits of general vs vocational/technical education in Singapore using quantile regressions. *International Journal of Manpower*, 27(4), 358-376. <https://doi.org/10.1108/01437720610679214>
- Salinas-Jiménez, M. D. M., Artés, J. & Salinas-Jiménez, J. (2016), Educational mismatch and job aspirations: A subjective wellbeing analysis using quantile regression. *International Journal of Manpower*, 37(1), 115-134. <https://doi.org/10.1108/IJM-12-2014-0266>
- Saraçoğlu, B. ve Çevik, F. (1995). *Matematiksel İstatistik*. Gazi Büro Kitabevi.
- Scharf, F. S., Juanes, F. & Sutherland, M. (1998). Inferring ecological relationships from the edges of scatter diagrams: Comparison of regression techniques. *Ecology*, 79, 448-460. [https://doi.org/10.1890/0012-9658\(1998\)079\[0448:IERFTE\]2.0.CO;2](https://doi.org/10.1890/0012-9658(1998)079[0448:IERFTE]2.0.CO;2)
- Schnabel, S. & Eilers, P. H. C. (2009). *Optimal expectile smoothing, Computational Statistics & Data Analysis*. Elsevier.
- Schulze, N. (2004). *Applied quantile regression: Microeconomic, financial, and environmental analyses* (Doctoral dissertation thesis). Eberhard Karls Universität Tübingen, Germany.

- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461-464.
- Schwarz, L., Bruckner, T., Ilango, S. D., Sheridan, P., Basu, R. & Benmarhnia, T. A. (2019). Quantile regression approach to examine fine particles, term low birth weight, and racial/ethnic disparities. *Environ Epidemiol*, 3(4), e060. <https://doi.org/10.1097/EE9.000000000000060>
- Searle, S.R., Casella, G. & McCulloch, C.E., (1992). *Variance Components*. Wiley, New York. <http://doi.org/10.1002/9780470316856>
- See, K. E., Ackerman, M. W., Carmichael, R. A., Hoffmann, S. L. & Beasley, C. (2021). Estimating carrying capacity for juvenile salmon using quantile random forest models. *Ecosphere*, 12(3): e03404. <https://doi.org/10.1002/ecs2.3404>
- Singer, J. D., & Willett J.B (2003). Examining the Multilevel Model's Error Covariance Structure, *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, Oxford Academic, New York. <https://doi.org/10.1093/acprof:oso/9780195152968.003.0007>
- Smith, L. B., Reich, B. J., Herring, A. H., Langlois, P. H. & Fuentes, M. (2015). Multilevel quantile function modeling with application to birth outcomes. *Biometrics*, 71(2), 508-19. <https://doi.org/10.1111/biom.12294>
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks: Sage Publications.
- Sobotka, T., Skirbekk, V. & Philipov, D. (2011), Economic Recession and Fertility in the Developed World. *Population and Development Review*, 37: 267-306. <https://doi.org/10.1111/j.1728-4457.2011.00411.x>
- Soni, P., Dewan, I. & Kanchan, J. (2012). Nonparametric estimation of quantile density function. *Computational Statistics & Data Analysis*, 56(12), 3876-3886. <http://doi.org/10.1016/j.csda.2012.04.014>

- Spanier, J., Azmy, Y. & Sartori, E. (2010). *Monte Carlo methods*.
https://doi.org/10.1007/978-90-481-3411-3_3
- Spiegel, D. , E., Weiner, E., Yarom, I., Doveh, E., Friedman, P., Cohen, A. & Shalev, E., (2013). Establishment of Fetal Biometric Charts Using Quantile Regression Analysis. *Journal of Ultrasound in Medicine*, 32: 23-33.
<https://doi.org/10.7863/jum.2013.32.1.23>
- Staffa, S. J., Kohane D. S. & Zurakowski, D. (2019). Quantile regression and its applications: A primer for anesthesiologists. *Anesth Analg*, 128(4), 820-830.
<https://doi.org/10.1213/ANE.0000000000004017>
- Sun, Y., Gao, H. & Li, F. (2017). Using linear mixed-effects models with quantile regression to simulate the crown profile of planted pinus sylvestris var. Mongolica Trees. *Forests*, 8(11), 446.
<https://doi.org/10.3390/f8110446>
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics* (5. Baskı). Pearson.
- Tansel, A. & Bodur, F. B. (2012). Wage inequality and returns to education in Turkey: A quantile regression analysis. *Review of Development Economics*, 16, 107-121. <https://doi.org/10.1111/j.1467-9361.2011.00655.x>
- Tareghian, R. & Rasmussen, P. (2013). Analysis of arctic and antarctic sea ice extent using quantile regression. *International Journal of Climatology*, 33(5), 1079-1086.
<https://doi.org/10.1002/joc.3491>
- TEDMEM. (2022). 2021 *Eğitim Değerlendirme Raporu*. <https://tedmem.org/yayin/2021-egitim-degerlendirme-raporu>
- Tian, M. & Chen, G., (2006). Hierarchical linear regression models for conditional quantiles. *Science in China Series A: Mathematics*, 1800–1815.
<https://doi.org/10.1007/s11425-006-2023-3>

- Tian, Y., Zhu, Q. & Tian, M. (2016). Estimation of linear composite quantile regression using EM algorithm. *Statistics and Probability Letters*, 183-191. <http://doi.org/10.1016/j.spl.2016.05.019>
- Tighe, E. L. & Schatschneider, C. (2016). A quantile regression approach to understanding the relations among morphological awareness, vocabulary, and reading comprehension in adult basic education students. *Journal of Learning Disabilities*, 49(4), 424-436. <https://doi.org/10.1177/0022219414556771>
- Tokatlıođlu, Y. (2018). *Dilim regresyon yöntemi ile Türkiye'de eğitimin ücret getirisi ve cinsiyete göre ücret ayrımcılıđının analizi* (Yüksek lisans tezi), Gazi Üniversitesi, Ankara.
- Tongal, A. (2017). *PISA 2015 Türkiye verilerine göre bazı deđişkenlerin fen testi başarı puanına etkisinin kantil regresyon yöntemi ile incelenmesi*. (Yüksek lisans tezi), Akdeniz Üniversitesi, Antalya.
- Tuma, N. & Hannan, M., (1984). *Social Dynamics: Models and Methods*. Academic Press.
- Turanlı, M. ve Güriş, S. (2005). *Temel istatistik* (2. Baskı). Der Yayınları.
- Tushar, A. (2012). *Returns to education in India: Some recent evidence*. <http://www.igidr.ac.in/pdf/publication/WP-2011-017.pdf>
- Tzavidis, N., Salvati, N., Schmid, T., Flouri, E. & Midouhas, E. (2016). Longitudinal analysis of the strengths and difficulties questionnaire scores of the Millennium Cohort Study children in England using M-quantile random-effects regression. *Journal of the Royal Statistical Society Series A: Statistics in Society, Volume 179*(2), Pages 427–452. <https://doi.org/10.1111/rssa.12126>
- Uluđ, S. (2019). *PISA 2015 Türkiye uygulamasında bazı öğrenci deđişkenlerinin fen okuryazarlıđı ve okuma becerileri başarısına etkisinin incelenmesi* (Yüksek lisans tezi), Akdeniz Üniversitesi, Antalya.

- Van Belle, G., Fisher, L., Heagerty P. J. & Lumley, T. (2004). *Biostatistics: A methodology for the health sciences. Longitudinal data analysis*. John Wiley and Sons.
- Vancouver, J., Thompson, C. & Williams, A. (2001). The Changing Signs in the Relationships among Self-efficacy, Personal Goals, and Performance. *The Journal of applied psychology*, 86(4). 605-20. <http://doi.org/10.1037/0021-9010.86.4.605>
- Van, M. H. (2013). *Türkiye'de gelir farklılığının kantil regresyon modeli ile incelenmesi: 2002-2010 yılları karşılaştırması* (Yüksek lisans tezi). Dokuz Eylül Üniversitesi, İzmir.
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p-values. *Psychonomic Bulletin and Review*, 14, 779–804. <https://doi.org/10.3758/BF03194105>
- Wagner, M., Kutash, K., Duchnowski, A. J. & Epstein, M. H. (2005). The special education elementary longitudinal study and the national longitudinal transition study: Study designs and implications for children and youth with emotional disturbance. *Journal of Emotional and Behavioral Disorders*, 13(1), 25-41. <http://doi.org/10.1177/10634266050130010301>
- Waldmann E. Quantile regression: A short story on how and why. *Statistical Modelling*. 2018;18(3-4):203-218. <https://doi.org/10.1177/1471082X18759142>
- Wang, Y., Feng, X. N. & Song, X. Y. (2016). Bayesian quantile structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(2), 246-258. <http://doi.org/10.1080/10705511.2015.1033057>
- Weaklim, D. L. (2004). Introduction to the special issue on model selection. *Sociological Methods and Research*, 33, 167-187. <https://doi.org/10.1177/0049124104268642>
- West, B., Welch, K. & Galecki, A. (2015). *Linear mixed models, a practical guide using statistical software* (2th Edition). CRC Press.

- Widyanti, R.D. (2018). Wage Inequality and Return to Education in Indonesia: Quantile Regression Analysis. *Signifikan: Jurnal Ilmu Ekonomi*. Vol. 7 (1), 27 – 44. <http://doi.org/10.15408/sjie.v7i1.6071>
- Willett, T. (1988). A Cross-Linguistic Survey of the Grammaticization of Evidentiality. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, 12(1), 51 – 97. <https://doi.org/10.1075/sl.12.1.04wil>
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.
- Xu, H. & Gao, Y., (2014). The development of English learning motivation and learners' identities: A structural equation modeling analysis of longitudinal data from Chinese universities, *System*, Volume 47, pages 102-115, ISSN 0346-251X, <https://doi.org/10.1016/j.system.2014.09.020>
- Yaman, N. (2019). *Kantil regresyonda yanlı tahmin edicilerin performanslarının incelenmesi* (Yüksek lisans tezi), Necmettin Erbakan Üniversitesi, Konya.
- Yavuz, E., Tan, Ş. & Atar, H. Y. (2019). Effects of students and school variables on SBS achievements and growth in mathematic. *Journal of Measurement and Evaluation in Education and Psychology*, 10(1), 96-116. <http://doi.org/10.21031/epod.493297>
- Yılmaz, H. İ. (2014). *Türkiye'de gelir dağılımının kentsel ve kırsal alanda tüketim harcamalarına etkisi: Ekonometrik bir inceleme* (Yüksek lisans tezi), Kahramanmaraş Sütçü İmam Üniversitesi, Kahramanmaraş.
- Yu, T. (2014). A quantile regression forecasting model for ICT development. *Management Decision*, 52(7), 1263-1272. <https://doi.org/10.1108/MD-11-2012-0793>
- Yu, K. & Moyeed, R. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54(4), 437-447. [http://doi.org/10.1016/S0167-7152\(01\)00124-9](http://doi.org/10.1016/S0167-7152(01)00124-9)

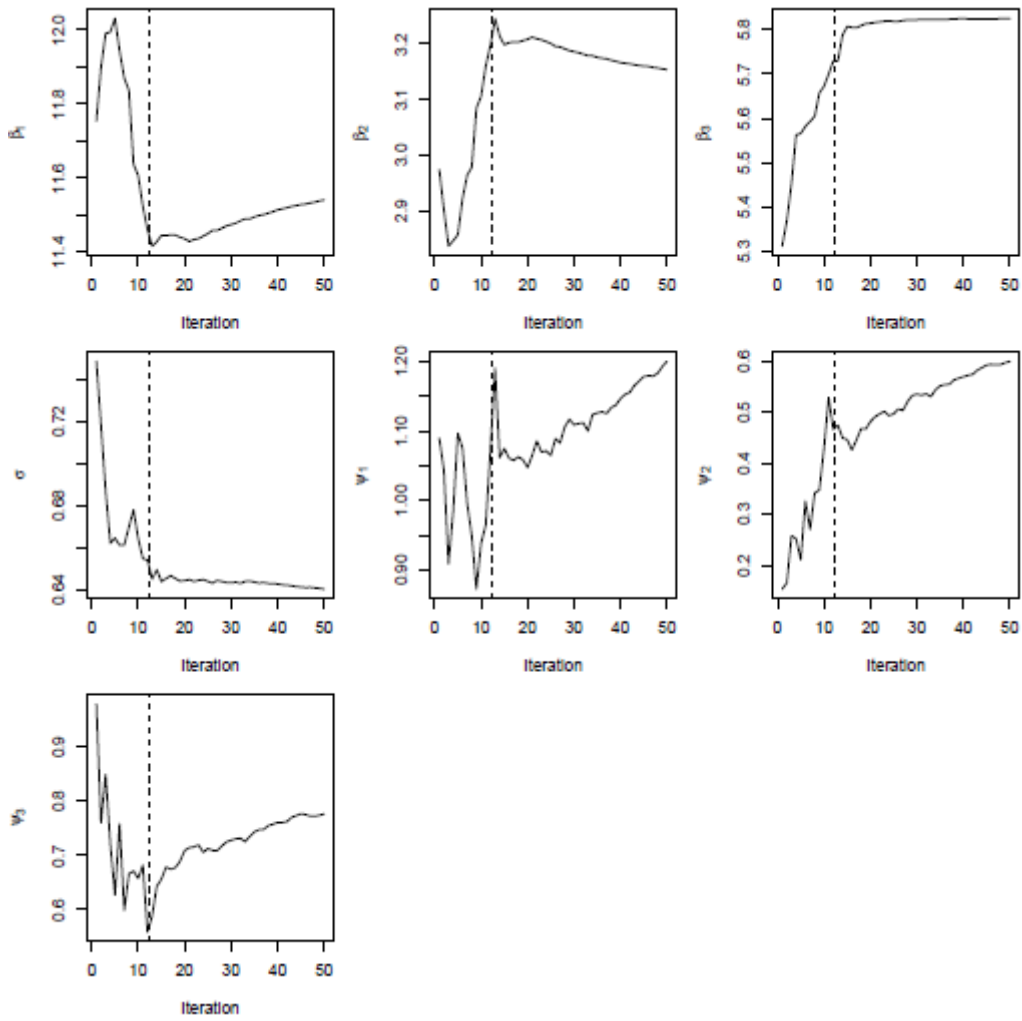
Yuan, Y. & Yin, G., (2010). Bayesian Quantile Regression for Longitudinal Studies with Nonignorable Missing Data. *Biometrics*, 66: 105-114.
<https://doi.org/10.1111/j.1541-0420.2009.01269.x>

Zou, H. & Yuan, M. (2008). Composite quantile regression and the oracle model selection. *Theory. Ann. Statist.*, 36, 1108-1126. <http://doi.org/10.1214/07-AOS507>

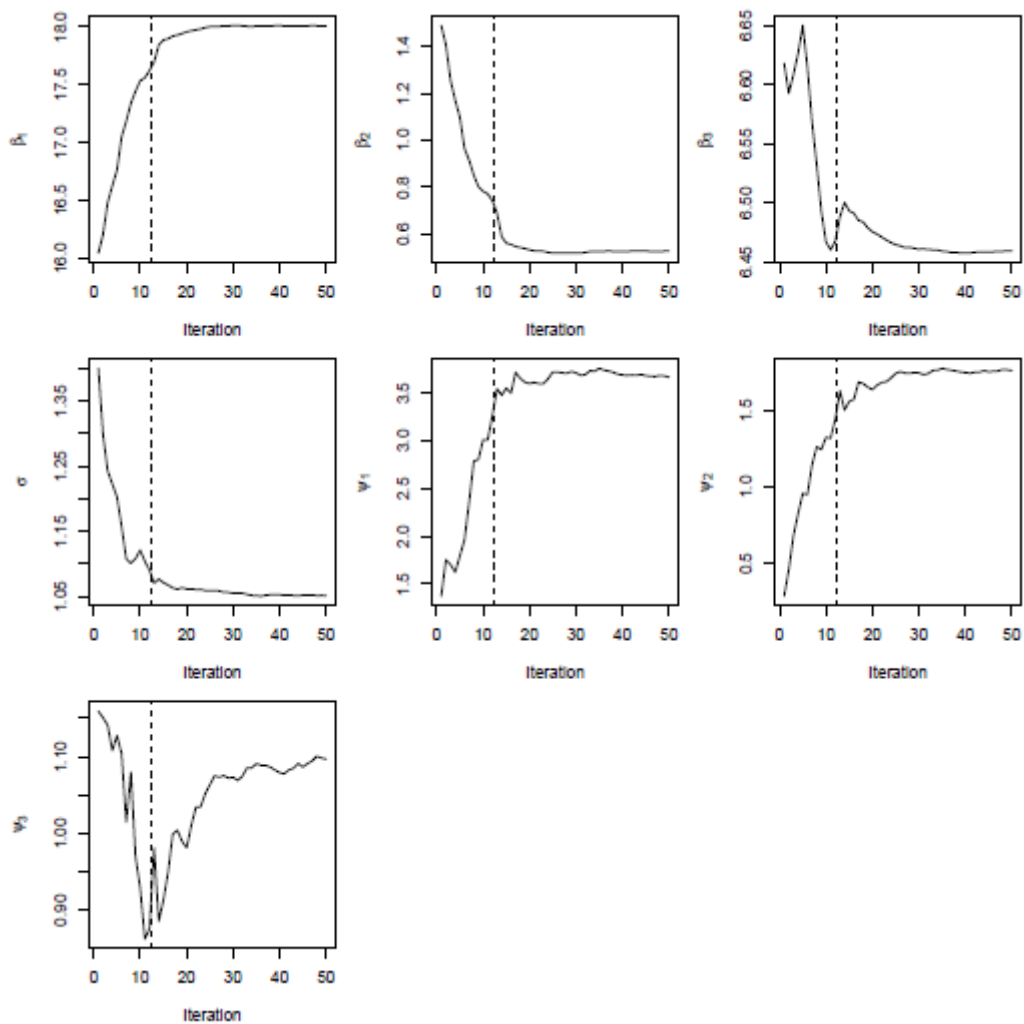
EK-A: Parametrelere Ait Ortalama Kesişim ve Eğim Değerleri

Hata Dağılımı	Örneklem	nlme	q50	q10	q25	q75	q90	nlme	q50	q10	q25	q75	q90
		kesişim	kesişim	kesişim	kesişim	kesişim	kesişim	eğim	eğim	eğim	eğim	eğim	eğim
[N,N]	50	17,958	18,254	11,665	14,811	21,798	25,411	6,149	6,139	5,732	5,894	6,418	6,581
	500	18,030	18,451	11,353	14,672	22,206	25,648	6,131	6,135	5,731	5,892	6,351	6,517
	1000	18,048	18,575	11,295	14,696	22,378	25,777	6,117	6,104	5,726	5,885	6,361	6,518
[N,kk]	50	23,053	22,920	18,497	20,373	25,990	28,853	6,110	6,206	5,551	5,837	6,481	6,571
	500	23,006	23,006	18,091	20,324	26,170	29,527	6,133	6,226	5,559	5,899	6,483	6,597
	1000	23,034	23,066	18,199	20,357	26,155	29,473	6,121	6,213	5,515	5,866	6,489	6,646
[N,T]	50	17,958	18,254	11,665	14,811	21,798	25,411	6,149	6,139	5,732	5,894	6,418	6,581
	500	20,799	21,072	17,026	18,925	23,315	25,559	6,121	6,170	5,385	5,771	6,545	6,877
	1000	20,748	21,160	17,417	19,181	23,205	25,107	6,136	6,153	5,349	5,737	6,567	6,937
[kk,N]	50	25,857	25,605	21,892	23,331	28,483	31,696	6,141	6,311	5,493	5,916	6,556	6,525
	500	23,084	23,327	16,164	19,459	27,424	31,398	6,119	6,164	5,703	5,901	6,373	6,468
	1000	23,049	23,353	16,100	19,432	27,378	31,313	6,129	6,155	5,716	5,914	6,390	6,476
[kk,kk]	50	25,862	25,767	22,107	23,391	28,990	32,474	6,079	6,208	5,413	5,874	6,439	6,511
	500	28,016	27,613	22,761	24,721	31,190	35,040	6,124	6,275	5,587	5,962	6,471	6,529
	1000	28,044	27,713	22,788	24,797	31,337	35,268	6,120	6,279	5,569	5,972	6,455	6,489
[kk,T]	50	23,043	23,238	16,692	19,752	27,425	30,887	6,141	6,174	5,744	5,882	6,422	6,496
	500	25,767	25,482	21,805	23,307	28,371	31,594	6,115	6,298	5,391	5,891	6,549	6,644
	1000	25,746	25,550	22,236	23,591	28,196	31,188	6,128	6,283	5,306	5,838	6,604	6,743
[T,N]	50	20,774	21,062	14,646	17,515	24,502	27,596	6,107	6,118	5,662	5,853	6,395	6,483
	500	20,796	21,119	14,600	17,646	24,725	28,028	6,136	6,156	5,679	5,883	6,422	6,574
	1000	20,829	21,270	14,920	17,845	24,706	27,824	6,124	6,142	5,646	5,865	6,396	6,596
[T,kk]	50	25,857	25,605	21,892	23,331	28,483	31,696	6,141	6,311	5,493	5,916	6,556	6,525
	500	25,809	25,410	21,675	23,208	28,396	31,860	6,122	6,306	5,412	5,911	6,575	6,614
	1000	25,802	25,445	22,088	23,423	28,247	31,466	6,119	6,316	5,336	5,877	6,588	6,690
[T,T]	50	23,555	23,380	20,848	21,897	25,508	28,006	6,153	6,289	5,216	5,757	6,736	6,878
	500	23,549	23,370	20,910	21,948	25,205	27,576	6,115	6,271	5,116	5,714	6,720	6,976
	1000	23,513	23,678	21,872	22,608	24,945	26,326	6,130	6,227	4,932	5,592	6,790	7,265

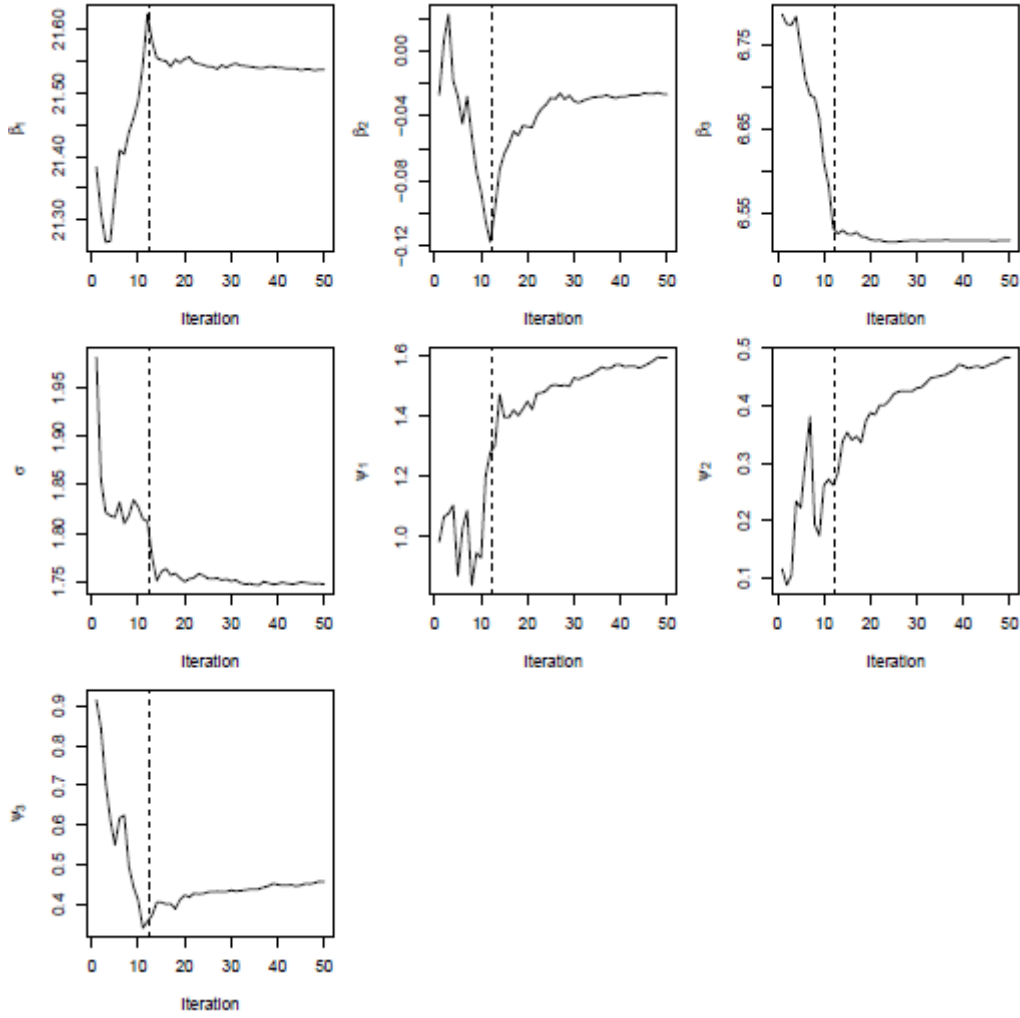
EK B: QRLMM Analiz Sonuçlarına Ait Örnek İterasyon Grafikleri (q10)



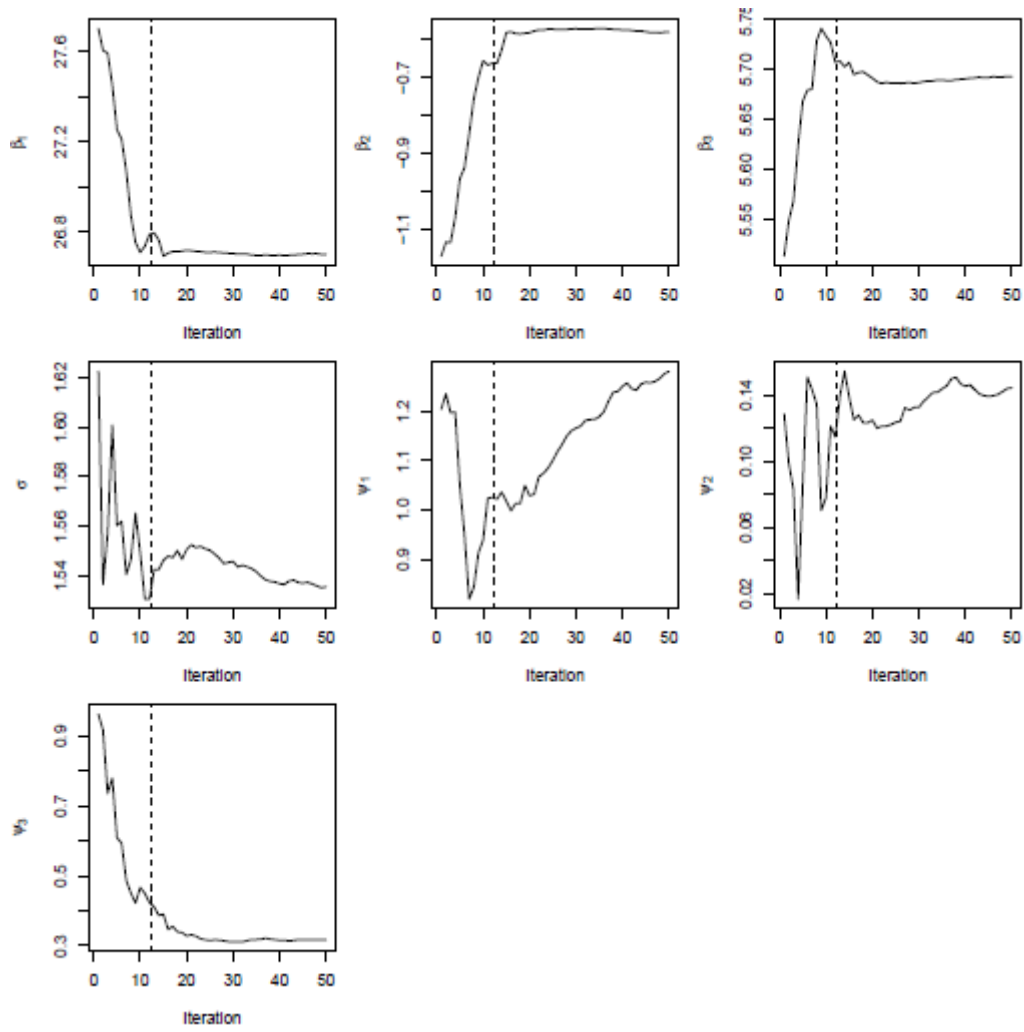
EK-C: QRLMM Analiz Sonuçlarına Ait Örnek İterasyon Grafikleri (q25)



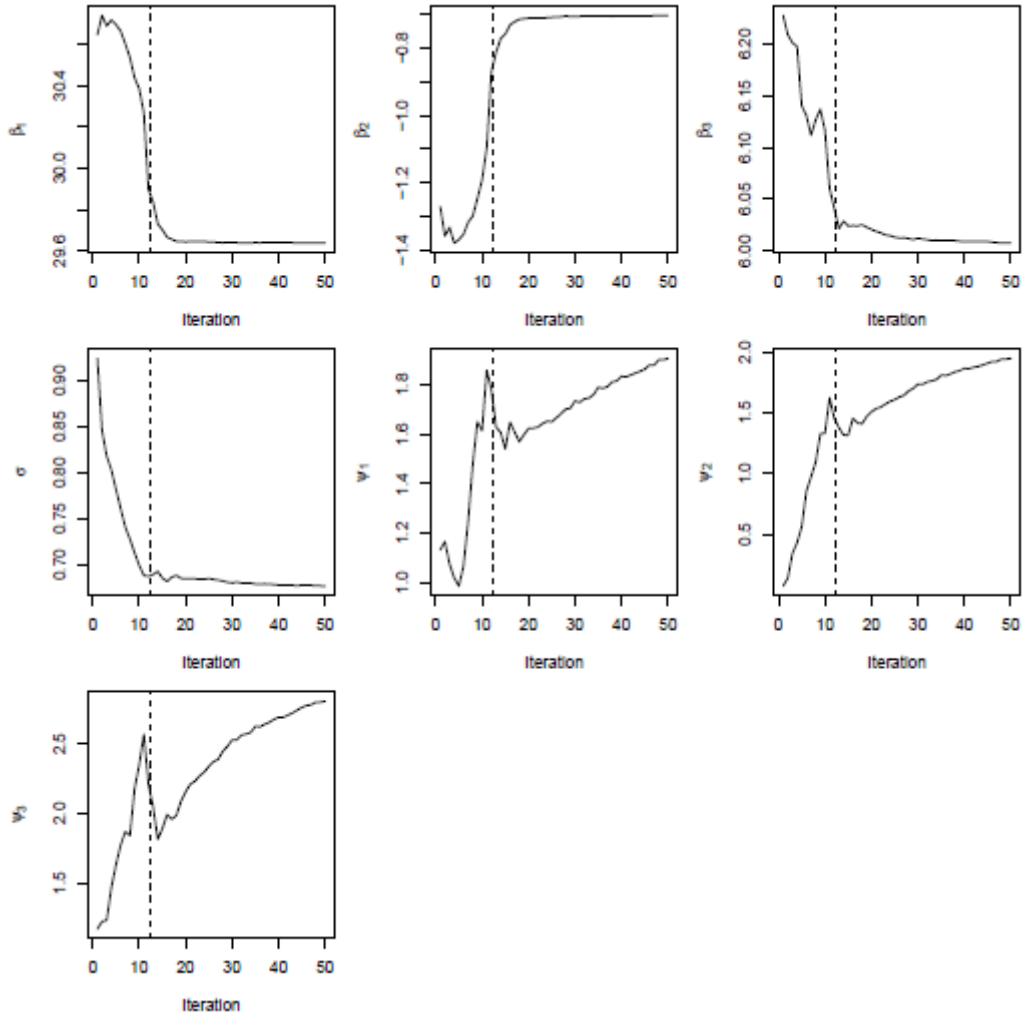
EK-D: QRLMM Analiz Sonuçlarına Ait Örnek İterasyon Grafikleri (q50)



EK-E: QRLMM Analiz Sonuçlarına Ait Örnek İterasyon Grafikleri (q75)



EK-F: QRLMM Analiz Sonuçlarına Ait Örnek İterasyon Grafikleri (q90)



EK-G: Arařtırma Etik Komisyon İzin Muafiyeti Formu/ Arařtırma Etik Komisyonu

Onay Bildirimi



T.C.
HACETTEPE ÜNİVERSİTESİ
Rektörlük



Sayı : 35853172-100
Konu : Begüm ÖZTEMÜR Hk.

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 11.04.2019 tarihli ve 51944218-300/00000548974 sayılı yazınız.

Enstitünüz Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı Doktora programı öğrencilerinden Begüm ÖZTEMÜR'in Dr. Öğr. Üyesi Derya ÇOBANOĞLU AKTAN danışmanlığında yürüttüğü "Çoklu Doğrusal Regresyon Yöntemiyle Kantil Regresyon Yönteminin Karşılaştırılması" başlıklı tez çalışması Üniversitemiz Senatosu Etik Komisyonununun 14 Mayıs 2019 tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini saygılarımla rica ederim.

e-izmalıdır
Prof. Dr. Rahime Meral NOHUTCU
Rektör Yardımcısı

EK-H: Etik Beyanı

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- * tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- * görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- * başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- * atıfta bulunduğum eserlerin bütününe kaynak olarak gösterdiğimi,
- * kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- * bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

...../...../.....

(İmza)

Begüm KAFKAS

EK-I: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu

11 / 08 / 2023

HACETTEPE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı : Boylamsal Verilerde Çok Düzeyli Doğrusal Regresyon Ve Kantil Regresyon Yöntemlerinin Karşılaştırılması

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
11/08/2023	205	297159	20/06 /2023	% 10	2144351386

Uygulanan filtreler:

- Kaynaklar hariç
- Alıntılar dâhil
- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esaslarını inceledim ve 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 eder, gereğini saygılarımla arz ederim.

Ad Soyadı: Begüm KAFKAS

Öğrenci No.: N13246985

Ana Bilim Dalı: Eğitim Bilimleri Ana Bilim Dalı

İmza

Programı: Eğitimde Ölçme ve Değerlendirme

Statüsü: Y.Lisans Doktora Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

(Doç. Dr. Derya ÇOBANOĞLU AKTAN)

EK-J: Thesis/Dissertation Originality Report

11 / 08 / 2023

HACETTEPE UNIVERSITY
Graduate School of Educational Sciences
To The Department of Educational Sciences

Thesis Title: Comparison of Multilevel Linear Regression and Multilevel Quantile Regression Methods in Longitudinal Data

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
11/08/2023	205	297159	20/06 /2023	% 10	2144351386

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Name Lastname: Begüm KAFKAS

Student No.: N13246985

Department: Educational Sciences

Program: Educational Measurement and Evaluation

Status: Masters Ph.D. Integrated Ph.D.

Signature

ADVISOR APPROVAL

APPROVED
(Assoc. Prof. Dr. Derya ÇOBANOĞLU AKTAN)

EK-K: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikrî mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge**" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü/Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren ... ay ertelenmiştir. ⁽²⁾
- Tezimle ilgili gizlilik kararı verilmiştir. ⁽³⁾

..... / /

(imza)

Öğrencinin Adı SOYADI

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanın önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir
*Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

