

**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

KONUMSAL MİKRO SİMÜLASYON İÇİN POPÜLASYON SENTEZLEME

YÜKSEK LİSANS TEZİ

Sevim İREY ZOBA

HAZİRAN 2017

TRABZON



KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

KONUMSAL MİKRO SİMÜLASYON İÇİN POPÜLASYON SENTEZLEME

Sevim İREY ZOBA

Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünde
"YÜKSEK LİSANS (İSTATİSTİK)
Unvanı Verilmesi İçin Kabul Edilen Tezdir.

Tezin Enstitüye Verildiği Tarih : 12 /05 /2017

Tezin Savunma Tarihi : 16 /06 /2017

Tez Danışmanı : Yrd. Doç. Dr. Uğur ŞEVİK

Trabzon 2017

**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**İstatistik ve Bilgisayar Bilimleri Anabilim Dalında
Sevim İREY ZOBA Tarafından Hazırlanan**

KONUMSAL MİKRO SİMÜLASYON İÇİN POPÜLASYON SENTEZLEME

**başlıklı bu çalışma, Enstitü Yönetim Kurulunun 16/ 05/ 2017 gün ve 1702 sayılı
kararıyla oluşturulan jüri tarafından yapılan sınavda
YÜKSEK LİSANS TEZİ
olarak kabul edilmiştir.**

Jüri Üyeleri

Başkan : Doç. Dr. Zafer KÜÇÜK

Üye : Yrd. Doç. Dr. Uğur ŞEVİK

Üye : Yrd. Doç. Dr. Fatma Gül AKGÜL



Prof. Dr. Sadettin KORKMAZ

Enstitü Müdürü

ÖNSÖZ

“Konumsal Mikro Simülasyon için Popülasyon Sentezleme” isimli bu tez Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik ve Bilgisayar Bilimleri Anabilim Dalı, Yüksek Lisans Programı’nda hazırlanmıştır.

Tez çalışma süresinde katkılarıyla beni yönlendiren benden yardımlarını, desteğini, sabrını ve bilgisini esirgemeyen değerli danışman hocam Sayın Yrd. Doç. Dr. Uğur ŞEVİK’e teşekkürü bir borç bilirim.

Ayrıca tez sürecinde R programını öğrenmemde yardımcı olan değerli hocam Sayın Yrd. Doç. Dr. Tolga BERBER’e ve bugüne kadar eğitimimde emeği geçen tüm hocalarıma teşekkürlerimi sunarım.

Son olarak, tüm hayatım boyunca her zaman beni destekleyen, her adımda arkamda duran aileme ve sevgili eşim Dr. Cuma Ali ZOBA’ya sonsuz teşekkürlerimi sunarım.

Bu tezin, bundan sonraki çalışmalara katkı sağlamasını temenni ederim.

Sevim İREY ZOBA
Trabzon 2017

TEZ ETİK BEYANNAMESİ

Yüksek Lisans Tezi olarak sunduğum “Konumsal Mikro Simülasyon İçin Popülasyon Sentezleme” başlıklı bu çalışmayı baştan sona kadar danışmanım Yrd. Doç. Dr. Uğur ŞEVİK’in sorumluluğunda tamamladığım 1, verileri/örnekleri kendim topladığımı, deneyleri/analizleri ilgili laboratuvarlarda yaptığımı/yaptırdığımı, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim.
12/05/2017

Sevim İREY ZOBA

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ.....	III
TEZ ETİK BEYANNAMESİ.....	IV
İÇİNDEKİLER.....	V
ÖZET	VII
SUMMARY	VIII
ŞEKİLLER DİZİNİ	IX
TABLolar DİZİNİ.....	X
SEMBOLLER DİZİNİ	XI
1. GENEL BİLGİLER.....	1
1.1. Giriş	1
1.2. Popülasyon Sentezi.....	4
1.2.1. Popülasyon Sentezi Yöntemleri.....	4
1.2.1.1. Adımlı Oransal Ayarlama (AOA) Tekniği	7
1.2.1.2. Adımlı Oransal Güncelleme (AOG) Tekniği	17
1.3. Mikro Simülasyon	21
1.4. Konumsal Mikro Simülasyon	22
1.4.1. Konumsal Mikro Simülasyon Uygulamaları	31
1.4.2. Konumsal Mikro Simülasyon Varsayımları	34
1.4.3. Konumsal Mikro Veri.....	35
1.5. Popülasyon Sentezlemede Alternatif Yaklaşımlar	35
1.6. Popülasyon Sentezleyiciler	45
1.7. PUMS Veri Kaynağı.....	46
2. YAPILAN ÇALIŞMALAR.....	50
2.1. R ile Konumsal Mikro Simülasyon	50
2.2. Konumsal Mikro Simülasyonda Ağırlıklandırma, Tamsayılaştırma ve Genişleme Süreçleri.....	52
2.3. Yapılan Çalışma.....	59

2.3.1.	Ağırlık Matrisinin Oluşturulması	61
2.3.2.	“ipfp” Kütüphanesi ile Konumsal Mikro Simülasyon.....	65
2.3.3.	“mipfp” Kütüphanesi ile Konumsal Mikro Simülasyon.....	67
2.3.4.	Tamsayılaştırma.....	70
2.3.5.	Genişleme	72
3.	BULGULAR VE SONUÇLAR.....	77
4.	ÖNERİLER.....	86
5.	KAYNAKLAR	88

ÖZGEÇMİŞ



Yüksek Lisans

ÖZET

KONUMSAL MİKRO SİMÜLASYON İÇİN POPÜLASYON SENTEZLEME

Sevim İREY ZOBA

Karadeniz Teknik Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik ve Bilgisayar Bilimleri Anabilim Dalı
Danışman: Yrd. Doç. Dr. Uğur ŞEVİK
2017, 93 Sayfa

Popülasyon sentezi ayrık örneklem verilerini genişleterek ilgilenilen hanehalkı ve bireysel verileri toplam nüfusa benzetmede kullanılan sentetik bir nüfus üretme işlemidir. Popülasyon sentezi konusunda literatürde en sık kullanılan yöntem Adımlı Oransal Ayarlama (AOA) tekniğidir. Bu yöntem tüm popülasyon için bireysel nüfus kayıtları üretmekte kullanılmaktadır. Mikro veri her bir birey ve hanehalkı hakkında toplanan bilgileri içeren bireysel kayıtlardır. Mikro simülasyon ise sistemdeki bireysel aktörlerin davranışlarını taklit ederek sistemin durumunun çoğaltılması ya da tahmini için kullanılan bir mekanizmadır. Konumsal mikro simülasyon ise, coğrafi alanlara atanan bireysel düzeyli verilerin oluşturulması, analizi ve modellemesidir. Popülasyon sentezi, konumsal mikro simülasyonun önemli bir safhasıdır. Konumsal mikro veri, konumsal mikro simülasyondan elde edilen faydalı çıktıdır. Konumsal mikro simülasyon için popülasyon üretimi, ülkemizde henüz pek çalışılan bir alan değildir. Herhangi bir alana ait açık kaynaklı bir mikro veri bulmak ülkemizde şimdilerde çok mümkün olmadığı için bu konu ülkemizde henüz ilgi görmemiştir. Bu tezin asıl amacı ise konumsal mikro simülasyon çalışmasına odaklanmaktadır. Bu çalışma AOA yöntemi baz alınarak R programı yardımıyla yapılmıştır. Çalışmada konumsal mikro simülasyonun nasıl oluşturulduğu iki konumlu bir bölgeden elde edilen konumsal mikro veri ile gösterilmiştir. Bu konumsal mikro simülasyon sonuçları “tamsayılaştırma” ve “genişleme” süreçlerinden geçirilerek nihai konumsal mikro simülasyon verilerinin nasıl elde edildiği gösterilmiştir.

Anahtar Kelimeler: Popülasyon Sentezi, Adımlı Oransal Ayarlama, Mikro Veri, Mikro Simülasyon, Konumsal Mikro Simülasyon, Konumsal Mikro Veri, Tamsayılaştırma, Genişleme.

Master Thesis

SUMMARY

POPULATION SYNTHESIS FOR SPATIAL MICROSIMULATION

Sevim İREY ZOBA

Karadeniz Technical University
The Graduate School of Natural and Applied Sciences
Statistical and Computer Science Graduate Program
Supervisor: Assist. Prof. Dr. Uğur ŞEVİK
2017, 93 Pages

Population synthesis is a process of producing a synthetic population that is used to simulate households and individual data of interest by expanding discrete sampling data to the total population. The most commonly used method in the literature about population synthesis is The Iterative Proportional Fitting / IPF technique. This method is used to generate individual population records for the entire population. Micro data are individual records that contain information about each individual and the household. Microsimulation is a mechanism used to reproduce or predict the state of the system by mimicking the behavior of individual actors in the system. Spatial microsimulation is the creation, analysis and modeling of individual level data assigned to geographical areas. Population synthesis is an important phase of spatial microsimulation. Spatial microdata is useful output from spatial microsimulation. Population production for spatial microsimulation is not yet a field in our country. This issue has not yet attracted attention in our country since it is not possible in our country to find open source micro data belonging to any field at present. The main purpose of this thesis focuses on spatial microsimulation study. This study was done with the help of R program based on IPF method. In this study, how the spatial microsimulation is generated is shown by the spatial microdata obtained from a two-position region. It has been shown how the final spatial microsimulation data is obtained by passing these spatial microsimulation results through "integerisation" and "expansion" processes.

Key Words: Population synthesis, The Iterative Proportional Fitting (IPF), Micro data, Microsimulation, Spatial Microsimulation, Spatial Micro Data, Integerisation, Expansion.

ŞEKİLLER DİZİNİ

Sayfa No

Şekil 1. Üç değişkene ait gözlemlerin bir listesi	5
Şekil 2. X ve Y değişkenli adımlı oransal ayarlama tekniğinin resimli gösterimi.....	6
Şekil 3. AOA tekniği örneği için gerekli veriler	12
Şekil 4. Adım 1a-Hanehalkı geliri için ayarlama.....	12
Şekil 5. Adım 1b-Hanehalkı büyüklüğü için ayarlama	13
Şekil 6. Adım 2a-Hanehalkı geliri için 2. ayarlama.....	14
Şekil 7. Adım 2b-Hanehalkı büyüklüğü için 2. ayarlama	14
Şekil 8. Adım 3a-Hanehalkı geliri için 3. ayarlama.....	15
Şekil 9. Adım 3b-Hanehalkı büyüklüğü için 3. ayarlama	16
Şekil 10. AOG tekniği örneği için gerekli veriler	19
Şekil 11. Adım 1-Hanehalkı tipi1 ve hanehalkı tipi2 için ayarlama	19
Şekil 12. Adım 1-Birey tipi1, birey tipi2 ve birey tipi3 için ayarlama & uyum ölçüsünün hesaplanması	20
Şekil 13. Adım 401-Nihai ağırlıklı toplam, nihai AOG ve AOA ağırlıkları ve uyum ölçüsü sonuçları	21
Şekil 14. Konumsal mikro veri ve daha yaygın bir şekilde kullanılan resmi veri tipleri arasındaki farklılıkları gösteren basitleştirilmiş bir akış şeması	26
Şekil 15. Stokastik kombinatoriyal optimizasyon ve AOA gibi deterministik yeniden ağırlıklandırma algoritmalarını içeren konumsal mikro veri oluşturulmasında farklı yaklaşımların şematik gösterimi.....	54
Şekil 16. Konum 1’de sentezlenen ve gerçek verilerin karşılaştırılması	78
Şekil 17. Konum 2’de sentezlenen ve gerçek verilerin karşılaştırılması	79
Şekil 18. Mikro veri, konum 1 ve konum 2’ ye ait simülasyon verilerinin yüzdelerinin karşılaştırılması	85

TABLolar DİZİNİ

	<u>Sayfa No</u>
Tablo 1. Bazı popülasyon sentezleyiciler ve özellikleri.....	46
Tablo 2. Konum 1 ve konum 2’ de bulunan bireylerin cinsiyet ve yaş bilgileri.....	59
Tablo 3. Mikro veride bulunan bireylerin cinsiyet, yaş, okuma yazma durumu ve sürücü belgesine sahip olma bilgileri	60
Tablo 4. Mikro veride bulunan bireylerin ortalama gelir bilgileri	60
Tablo 5. Mikro veride hanehalkının sahip olduğu toplam binek araç sayısı.....	61
Tablo 6. “mipfp” ile mikro verinin simülasyon sonuçları özet tablosu.....	70
Tablo 7. “mipfp” ile elde edilen sonuçlarının tamsayılaştırılmış hali.....	72
Tablo 8. Konum 1’deki bireylerin genişleme süreci sonrası cinsiyet değişkeni simülasyon sonuçları.....	74
Tablo 9. Konum 1’deki bireylerin genişleme süreci sonrası ortalama gelir değişkeni simülasyon sonuçları.....	74
Tablo 10. Konum 1’deki bireylerin genişleme süreci sonrası yaş değişkeni simülasyon sonuçları	75
Tablo 11. “mipfp” ile genişleme süreci sonrası simülasyon sonuçları.....	76
Tablo 12. Konumlardaki gerçek ve sentezlenen birey sayısı karşılaştırması.....	78
Tablo 13. Konum 1’e ait nihai konumsal mikro simülasyon sonuçları ve karşılaştırması .	79
Tablo 14. Konum 2’ye ait nihai konumsal mikro simülasyon sonuçları ve karşılaştırması	80
Tablo 15. Konum 1’e ait nihai ortalama gelir tahmini sonuçları	80
Tablo 16. Konum 2’ye ait nihai ortalama gelir tahmini sonuçları	81
Tablo 17. Konum 1 ve konum 2’ye ait toplam ortalama gelir ve kişi başı ortalama gelir tahmini	82
Tablo 18. Konum 1 ve konum 2’ye ait okuma-yazma durumu tahmini sonuçları.....	82
Tablo 19. Konum 1 ve konum 2’ye ait sürücü belgesi tahmin sonuçları.....	82
Tablo 20. Konum 1 ve konum 2’ye ait hane halklarının sahip olduğu toplam binek araç sayısı tahmini sonuçları.....	83
Tablo 21. GREGWT ile konum 1 ve konum 2’ye ait toplam ve kişi başı ortalama gelir tahmini	84

SEMBOLLER DİZİNİ

C	: Olumsallık tablosundaki hücrelerin toplam sayısı ($C=I \times J \times K$)
I, J, K	: Sırasıyla X, Y ve Z değişkenlerinin kategori sayısı
n_{ijk} ya da n	: Boyutu $I \times J \times K$ olan kaynak örneğin çok yönlü olumsallık tablosu
n	: Kaynak örneğin büyüklüğü
n_{ijk}	: n 'in tek bir hücresi
N_{ijk} ya da N	: Hedef popülasyonun çok yönlü olumsallık tablosu
N	: Hedef popülasyonun büyüklüğü
N_{ijk}	: Hedef tablo N 'de tek bir hücre
N_{i++}	: X değişkeninin her i kategorisi için toplam gözlemleri gösteren N_{ijk} 'nin bir marjinalini içeren tek yönlü tablo
N_{i++}	: N_{i++} 'de tek bir girdi
N_{ij+}	: X değişkeninin her i kategorisi ve Y değişkeninin her j kategorisi için toplam gözlemleri gösteren N 'in bir marjinalini içeren iki yönlü tablo
\widehat{N}_{ijk} ya da \widehat{N}	: Hedef çok yönlü N tablosunun AOA tahmini
\widehat{N}_{ijk}	: \widehat{N} tahmininde AOA sürecinde tek bir hücre
π_{ijk} (ya da π_{ijk})	: Sayıların yerine olasılıkların tablosu ($E[n_{ijk}] = n\pi_{ijk}$)
Π_{ijk} (ya da Π_{ijk})	: Sayıların yerine olasılıkların tablosu ($E[N_{ijk}] = N\Pi_{ijk}$)
X ya da $X(i)$: I kategorili X değişkeni
Y ya da $Y(j)$: J kategorili Y değişkeni
Z ya da $Z(k)$: K kategorili Z değişkeni

1. GENEL BİLGİLER

1.1. Giriş

Ülkeler için konumsal alanlara ilişkin istatistiksel tahmin gereksinimi sık sık karşılaşılan bir problem olmaktadır. İstatistiksel tahminlerin gerçeğe en yakın bir şekilde yapılması için eldeki verilerin uyumlu olması gerekmektedir. Bu nedenle, araştırmalarda kullanılmak üzere örneklemelerden simülasyon teknikleri kullanarak kitleye en yakın bireylerden oluşan sentetik bir popülasyon oluşturulabilmektedir.

Mikro simülasyon, sistemdeki bireysel aktörlerin davranışlarını taklit ederek dinamik ve karmaşık sistemin durumunun çoğaltılması ya da tahmini için kullanılan bir mekanizmadır. Mikro simülasyon son yıllarda sosyal konuların araştırılması ve Ar-Ge çalışmaları için oldukça ilgi duyulan bir konu olmuştur. Örneğin, ekonomistler vergi politikalarını analiz etmek için hanehalkı gelir yapısının mikro simülasyon modellerini kullanmaktadırlar. Kentsel ve bölgesel çalışan bilim adamları refah politikaları ve istihdamın etkilerini değerlendirmek; ulaştırma mühendisleri ve planlamacılar ise yönetimin talep ettiği çeşitli politikaların etkilerini analiz etmek ve faaliyet bazlı seyahat modelleri incelemek için mikro simülasyona başvurumaktadırlar. Faaliyet tabanlı seyahat talebi tahmini yapabilmek için simülasyon yapılacak alandaki hanehalkı mikro verilerine ihtiyaç duyulmaktadır. Şehirlerdeki trafik alanları ve küçük nüfuslu ilçelerin trafik talep modellemesi için sentetik popülasyon verisi üretmek oldukça zordur. Mikro alanlarda trafik talep modellerini oluşturmak ve seyahat aktivitelerinin arkasında yatan faktörleri keşfetmek için de mikro simülasyon önemli bir platform oluşturmaktadır.

Mikro simülasyon modellemesi için en uygun veri kaynağı düzenli aralıklarla ülkelerden toplanan nüfus sayımı verisidir. Fakat bazı sorunlar, mikro simülasyon için girdi verisi olarak nüfus verisi kullanımını engelleyebilmektedir. Bu sorunların ilki birçok ülkede tam nüfus sayımının olmayıp, sadece ortak kullanım olarak adlandırılan küçük bir alt örneğe ulaşılabilmesidir. İkinci sorun ise nüfus sayımının oldukça seyrek olarak yapılmasıdır. Bazı ülkelerde yapılan iki ardışık sayım arasında 10 yıl geçebilmektedir. Bu durum mikro simülasyon için kullanılan baz yılı seçimini kısıtlamaktadır. Bu tür sorunların üstesinden gelebilmek için Popülasyon Sentezleme yaklaşımı kullanılabilir (Müller ve Axhausen, 2010).

Mikro simülasyon modellerinde sistemin gelecekteki durumunu tahmin etmek için popülasyonun kararları zamanla simüle edilmektedir. İlk adım, popülasyonun tanımı olmakla birlikte bu genellikle bireyler ve hanehalkları olmaktadır. Eğer çalışılan alanın tüm nüfusunun anlık simülasyonu varsa başlangıç olarak kullanılabilir. Ancak, böyle bir veri gizlilik ve maliyet kısıtları yüzünden çoğunlukla elde edilememektedir. Bu sorun, korelasyon yapısı ve marjinal toplamlar gibi belirli ölçütlere uyan popülasyonun toplamsal olmayan bir temsilini elde etmek için farklı veri kaynaklarının birleştirilmesi ile çözülebilir. Bu işlem popülasyon sentezi olarak adlandırılabilir. Popülasyon sentezi için yapılan son çalışmalar incelendiğinde, söz konusu çalışmaların hepsinin iki ana görevi olduğu görülmektedir. İlki, geçmiş nüfus sayımı ya da anketlerden alınan başlangıç popülasyonunun şimdiki nüfusa ayarlanması; ikincisi ise, hanehalkının seçimi ve isteğe bağlı olarak hanehalklarının coğrafi alanlara atanmasıdır.

Konumsal mikro simülasyon yönteminin doğası ve insanları modelleyebilme kabiliyetiyle çeşitli alanlarda kullanılması faydalı bir yaklaşım olabilmektedir. Örneğin araştırma yapılan sahada, ikamet eden bireylerin eğitiminin geleceği ve yerel düzeyde sağlık ve ulaşım hizmetleri hakkında planlamalar yapmak için konumsal mikro veriden yararlanılabilir. Gelecek elbette belirsizdir ancak konumsal mikro simülasyon gelecek hakkında çeşitli tahminleri yapmak için senaryo temelli planlamalar sunarak geleceği tahmin ederek faydalı bilgiler sağlayabilmektedir.

Konumsal mikro simülasyon, sosyal bilimlerdeki karmaşık konuları inceleyen ve son zamanlarda çok kullanılan bir yaklaşımdır. Bu yaklaşım ulaşım, sağlık, eğitim gibi farklı alanlarda yaygın olarak kullanılmakta ve aynı şekilde bu yaklaşımı daha birçok alana da uygulamak mümkün olabilmektedir. Yaklaşımın temelinde bireylerin konumlara atanması gibi büyük saha çözümlerinde yer alan bireysel düzeydeki verilere ait yaklaşımlar yer almaktadır. Konumsal mikro veri tüm konumsal mikro simülasyonu araştırmaları için temel oluşturur.

Popülasyon sentezi konumsal mikro simülasyonda önemli bir safhadır. Genellikle konumsal mikro veri, veri gizliliği konusundaki endişeler nedeniyle kullanılmamaktadır. Genellikle sentetik konumsal mikro veri tabanları, nüfus sayım verilerinden elde edilen toplamsal veriler ile ilgilenilen popülasyonu temsil eden anketlerden elde edilen bireysel düzeydeki verilerin birleştirilmesiyle oluşturulurlar.

Konumsal mikro veri, bireysel düzeyde ve coğrafi olarak belirli süreçlerin devreye girdiği durumlarda simülasyon için faydalı olmaktadır. Konumsal mikro simülasyon çok

değişkenli modelleme ve analize de olanak vermektedir. Konumsal mikro simülasyon aynı zamanda birey tabanlı modellerle de örtüşmekte ve başlangıç koşulları için fayda sağlamaktadır. Konumsal mikro simülasyonun birçok faydası olmasına rağmen, bölgesel bilim alanlarının dışında çok az bilinmektedir. Konumsal simülasyon, kamu politikalarını bilgilendirmek için yeni alanlara uygulanacak büyük bir potansiyele sahiptir. Örneğin; sosyal fayda çalışmaları, barınma, ulaşım ve sürdürülebilir kentsel planlamada konumsal mikro simülasyon kullanılmaktadır.

Günümüzde, konumsal mikro simülasyon giderek artan bir ilgiye sahiptir. Bilgisayarlar, yazılımlar ve veri erişilebilirliğindeki gelişmeler yöntemlerin eskisinden daha fazla kullanılabilmesi anlamına da gelmektedir. Neredeyse dünyanın her yerinde bireysel düzeyde küçük idari alanların popülasyonlarını simüle etmek artık mümkün olabilmektedir.

Bu çalışmanın asıl amacı, konumsal mikro simülasyon yaklaşımının irdelenmesidir. Bu amaçla, öncelikle popülasyon sentezi ve popülasyon sentezleme yöntemleri üzerinde literatür araştırması yapılmış, daha sonra da konumsal mikro simülasyon ve konumsal mikro simülasyon oluşturma süreçleri ayrıntılı olarak incelenmiştir. Ek olarak popülasyon sentezlemede kullanılan alternatif yaklaşımlardan, popülasyon sentezi için oluşturulmuş popülasyon sentezleyici uygulamalardan ve popülasyon sentezlemede en sık kullanılan PUMS veri kaynağından da bahsedilerek literatür incelemesine yer verilmiştir. Ülkemizde popülasyon sentezi, mikro simülasyon ve konumsal mikro simülasyon konularında pek az çalışma yapıldığı görülmüştür. Bu çalışmada, konumsal mikro simülasyonun temel yöntemlerinden olan Adımlı Oransal Ayarlama (AOA) yöntemi R programı kullanılarak ele alınmıştır. R programında konumsal mikro simülasyon oluşturma süreçlerinde kullanılan AOA yöntemi baz alınarak yazılmış “ipfp” ve “mipfp” kütüphanelerine yer verilmiştir. Bu kütüphaneler konumsal mikro simülasyon uygulamalarında da kullanılmıştır. Elde edilen nihai tahminler gerçek veri ile karşılaştırılmıştır. Bununla birlikte, popülasyon sentezine alternatif bir yaklaşım olan GREGWT yönteminden bahsedilmiş ve yapılan çalışmalarda GREGWT ile sentezlenen birey verilerinden yararlanarak, konumlara ait toplam ortalama gelir ve kişi başı ortalama gelirin nasıl tahmin edildiği de gösterilmiştir. Sonuç olarak, GREGWT ve AOA (“ipfp”) yöntemlerinden konumlara ait elde edilen toplam ortalama gelir ve kişi başı ortalama gelir sonuçları karşılaştırılmıştır.

1.2. Popülasyon Sentezi

Popülasyon sentezi ayırık örneklem verilerini genişleterek ilgilenilen hanehalkı ve bireysel verileri toplam nüfusa benzetmek için kullanılan sentetik bir nüfus üretme işlemidir. Bu amaçla, tahminlerin elde edilmesinde kullanılan çeşitli yaklaşımlar geliştirilmiştir. Popülasyon sentezi konusunda literatürde en sık kullanılan yaklaşımlar Adımlı Oransal Ayarlama (AOA) ve Adımlı Oransal Güncelleme (AOG) tekniğidir. Daha önceki çalışmalarda bu yöntemler ile elde edilen sentetik veriler, nüfus sayımı verileri ile karşılaştırıldığında sentetik verilerin ne derece temsil edilebilir olduğu gösterilmiştir (Müller ve Axhausen, 2010). Bu yaklaşımlar sayesinde tüm popülasyon için bireysel nüfus kayıtları üretilebilmektedir.

1.2.1. Popülasyon Sentezi Yöntemleri

Hemen hemen tüm popülasyon sentezi süreçleri çok yönlü olumsuzluk tablolarında saklanan verilere dayanmaktadır. Bu verilerin tipini anlamaya ve açıklamaya yardım etmek için ilk olarak tutarlı bir gösterim sisteminin tanımlanması gerekmektedir.

Sayısal değerler ve olumsuzluk tablolarındaki tek hücrelerdeki değerler italik harf karakteri ile ifade edilecektir (örneğin, n ya da n_{ijk}). Çok yönlü olumsuzluk tabloları ve marjinaler, birden fazla hücrede olduğunu belirtmek için italik olmayan harf karakteri ile temsil edilecektir (örneğin, n ya da n_{ijk}). Olumsuzluk tabloları bir, iki ya da çok yönlü olabilmektedir. Kullanılan simgelerin sayısı olumsuzluk tablosunun boyutunu belirler (örneğin n_{ijk}).

Popülasyon sentezini anlamak için, rastgele olarak 3 değişken ele alınsın. X, Y ve Z değişkenlerinin eş zamanlı olarak değiştiği ve bu değişkenlerin sırasıyla I, J ve K kategorileri ile sınıflandırıldığı varsayılınsın. Bu değişkenler doğal olarak kesikli ya da sürekli olabilmektedir. Ancak sürekli değişkenler sonlu bir dizi şeklinde kesikli kategoriler halinde gruplandırılmaktadır. 'i' değişkeni X'in bir kategorisini ifade eder ve kategoriler $\{1, 2, \dots, I\}$ olarak etiketlenmiştir. Bu durum Y ve Z değişkenleri için de aynı olmaktadır. (Örneğin, bu değişkenlerin yaş, eğitim ve konum gibi birey özelliklerini temsil ettiği varsayılır.) Popülasyon sentezinde kullanılan π_{ijk} ifadesi olasılığı temsil etmektedir. π_{ijk} olasılığında rasgele bir gözlemin ilk değişkeni i, ikinci değişkeni j ve üçüncü değişkeni ise

k kategorisi ile sınıflandırılmaktadır. Mikro verinin bulunduğu tabloda $C = I \times J \times K$ tane hücre mevcuttur. Bu hücrelerden her birinin uygun kategorili gözlemlerinin sayısı n_{ijk} ile ifade edilmektedir. Bu gözlem sayıları katlıterimli π_{ijk} olasılık kitle fonksiyonuna dahil olan gözlemlerdir. Şekil 1’de iki değişkenli bir olumsallık tablosu oluşturmak için kullanılan 3 değişkene ait gözlemlerin listesi görülmektedir.

<i>Kayıt #</i>	<i>X(i)</i>	<i>Y(j)</i>	<i>Z(k)</i>
...
1663	2	2	7
1664	2	3	1
1665	2	3	5
1666	2	3	5
1667	2	3	7
1668	2	4	1
...

Şekil 1. Üç değişkene ait gözlemlerin bir listesi

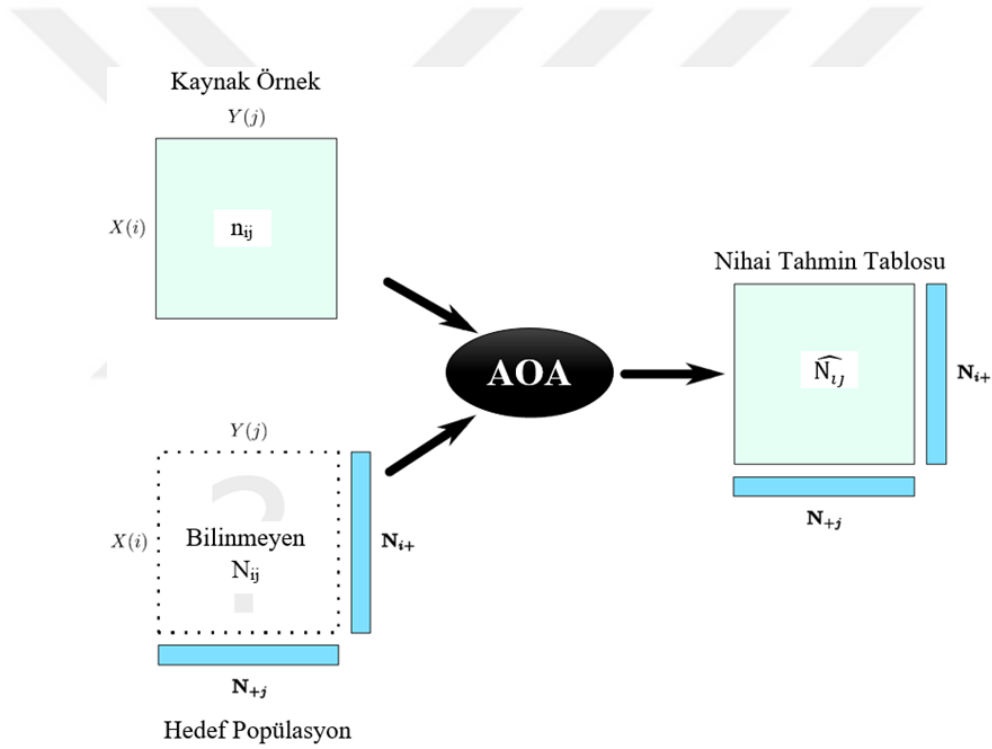
Şekil 1’deki görülen liste tabanlı tablo ve olumsallık tablosu arasındaki bağlantı çok değişkenli kategorik veriyi temsil etmektedir. Liste tabanlı tablo, her bir satırın tek bir gözlemi temsil ettiği değerlerin listesini göstermektedir. Tablodaki X, Y ve Z değişkenlerinin i, j ve k kategorilerine ayrıldığı görülmektedir. Olumsallık tablosu ise sadece X ve Y değişkenlerinin kullanıldığı gözlemlerin bir çapraz tablosudur. Tablodaki her bir n_{ij} X=i ve Y=j değerlerini aldığı gözlemlerin sayısı olmaktadır. Olumsallık tablosu, liste tabanlı temsil edilen tablodan n_{ij} gözlemlerinin belirli bir kümesine karşılık gelir.

Herhangi bir olumsallık tablosu, bir ya da daha çok boyutta toplanarak daha küçük boyutlu bir tabloya daraltılabilmektedir. Bu daraltılmış tabloya marjinal tablo adı verilmektedir. n_{i++} notasyonu, ikinci ve üçüncü değişkenleri daraltılmış sadece X değişkeninin I kategorisinin bırakıldığı marjinal tablosu için kullanılmaktadır. Notasyonda + sembolleri j ve k’nın tüm kategorileri üzerine n_{ijk} ’yı toplayarak marjinaler elde edilmektedir. Tablolanmış örneğin toplam büyüklüğü n_{+++} ifadesi ya da sadece n ile gösterilmektedir.

Çoklu olumsallık tabloları genellikle eş zamanlı olarak kabul edilmektedir. AOA tekniğinin tipik bir uygulamasında ‘kaynak’ popülasyon örneklendirilerek, n_{ij} ’nin çok yönlü tablosu çapraz olarak sınıflandırılmaktadır. Benzer şekilde yapılandırılmış çok yönlü

N_{ij} tablosu, hedef popülasyonun bir kısmı için oluşturulmaktadır. Fakat hedef popülasyon hakkında yeterli bilgi her zaman mevcut olmayabilmektedir. Genellikle N_{i+} ve N_{+j} gibi bazı marjinal toplamalarının bir kısmı bilinmektedir. Uygulamaya bağlı olarak, hedef ve kaynak popülasyonlar farklı ya da aynı olabilir; örneğin popülasyonlar aynı olduğunda, hedef marjinaller popülasyonun %100 oranda tam bir örneğini teşkil ederken, kaynak örneklemin boyutu ise %1 ile %5 arası küçük değerler alabilmektedir.

Hedef nüfusun çok yönlü tablosundaki N_{ij} her zaman bilinmemektedir, ancak AOA tekniği \widehat{N}_{ij} 'nin tahminini bulmak için kullanılabilir. Bu tahmin n_{ij} tablosunun tekrarlanan modifikasyonları aracılığıyla elde edilmektedir. Tüm süreç ve ilgili notasyonlar Şekil 2 ve Semboller Dizininde gösterilmiştir.



Şekil 2. X ve Y değişkenli adımlı oransal ayarlama tekniğinin resimli gösterimi

Şekil 2'de kaynak tablodaki n_{ij} , bilinen hedef marjinalleri olan N_{i+} ve N_{+j} 'ye eşlenecek şekilde simüle edilerek Nihai Tahmin Tablosu'ndaki \widehat{N}_{ij} değerleri elde edilmektedir. Böylece kaynak tablo bilinmeyen hedef tablo N_{ij} 'ye benzetilmektedir. Semboller dizininde yer alan notasyonlar AOA ve çok yönlü tablolar için kullanılmaktadır. AOA yaklaşımı, bilinmeyen hedef popülasyonun çok yönlü olumsuzluk tablosunu tahmin etmek için kullanılmaktadır. AOA, bu tahmini yaparken kaynak örneğin tablosunu hedef

popülasyonun bilinen marjinallerine genişletme işlemi ile gerçekleştirmektedir. Bu notasyon tablosu X, Y, Z gibi üç değişken için olabilmekle birlikte daha fazla değişken için de kullanılabilir. Kaynak tablo n_{ij} ve hedef marjinaler N_{i+} değerleri genellikle tamsayıdır. Ancak AOA tarafından üretilen \hat{N}_{ij} tahmin değerlerini içeren hedef tablo reel değerlere sahiptir.

Popülasyon sentezi yöntemlerine geçmeden önce mikro verinin tanımının yapılması gerekmektedir. Mikro veri, her bir birey ve hanehalkı ile ilgili demografik ve coğrafi bilgileri içeren bireysel kayıtlardır. Mikro veriler hanehalkından toplanan anketlerin bilgisayar ortamında hazırlanmış kayıt türlerinden oluşmaktadır. Nüfus idaresi toplanan bu mikro verileri; çeşitli raporlar, özet dosyalar ve özel tablolar oluşturmak için özet verileri üretmek amacıyla kullanılmaktadır. Popülasyon sentezlemede literatürde en sık kullanılan veri kaynağı ABD'nin nüfus sayımından elde edilen PUMS veri kaynağıdır. PUMS verileri mikro verilerden oluşmaktadır. Çalışmanın 1.7 başlığı altında PUMS verisi hakkında daha ayrıntılı bilgi verilmiştir.

1.2.1.1. Adımlı Oransal Ayarlama (AOA) Tekniği

AOA algoritması, genellikle Deming ve Stephan tarafından geliştirildiği bilinmektedir (Deming ve Stephan, 1940). Csizs'ar'a göre, AOA algoritması 1937'de bir Alman yayını tarafından telefon endüstrisine uygulandığı için Deming ve Stephan'dan daha önce kullanılmıştır (Csizs'ar, 1998). Yöntem literatürde çalışma alanı ve içeriğine bağlı olarak birçok isimle anılmaktadır. İstatistikçiler bu yöntemi olumsuzluk tablolarına uygulamakta ve tablo standardizasyonunda kullanılmaktadırlar. Ulaştırma alanında çalışan araştırmacılar bu yöntemi seyahat planlama ve trafik modellemede kullanılmaktadırlar (Ortúzar ve Willumsen, 2002; Furness, 1965). Ekonomistler ise bu yöntemi reel ekonominin sektörel düzeyde ve bir bütün olarak incelenmesinde yararlanılan Girdi – Çıktı modelleri için kullanılmaktadırlar (Lahr ve Mesnard, 2004). AOA yöntemi ayrıca matris sıralaması ya da matris ölçeklemesi olarak da bilinmektedir (Müller ve Axhausen, 2010). Klasik AOA yöntemi etkileşim terimleri olmayan log-linear model olarak düşünülebilir (Agresti, 2002).

AOA algoritması, hedef popülasyon elde etmek için kaynak olumsuzluk tablosunu, bilinen marjinal toplamlara genişletmeye yarayan bir ayarlama yöntemi şeklinde de ifade edilebilmektedir. Deming ve Stephan başlangıçta örnekleme farklılığındaki değişimleri

açıklamak için bu yöntemi önermişlerdir (Deming ve Stephan, 1940). Marjinal toplamların tamamının bilindiğini, kaynak tablo ve hedef marjinalerin aynı popülasyon üzerinden ölçülebileceğini düşünmüşlerdir. Fakat kaynak tablo hata payı ile örnekleme süreci yardımıyla ölçülmüştür. Daha sonra AOA yöntemi daha doğru marjinal toplamları elde etmek için örnekten türetilmiş hücreleri ayarlamak için kullanılmıştır. Çapraz tablolarla tahmin yapılmak istendiğinde, örneklemeden ve diğer kaynaklardan elde edilen veri ilgili teori ile bazen uyumluluk göstermeyebilir. Bu sorunun üstesinden gelmek için Deming ve Stephan AOA yöntemine başvurmuşlardır.

Deming ve Stephan, d boyutlu olasılık tablosunu, bir boyutlu ve daha üst seviye marjinaler ($d-1$ boyuta kadar) üzerinde çalışmışlardır (Deming ve Stephan, 1940). Çalışmalarında başlangıç hücre değerlerinde ya da marjinalerde sıfır değerlerin etkisini ele almamışlardır.

Deming ve Stephan AOA algoritmasının iki kriteri karşılayan tek bir çözüm ürettiğini iddia etmişlerdir. Önerilen yöntem tamamen marjinal kısıtlamaları karşılamaktadır.

$$\sum_j \widehat{N}_{ij} = N_{i+}, \quad \sum_i \widehat{N}_{ij} = N_{j+} \quad (1)$$

AOA ağırlıklı en küçük kareler kriterini minimize eder.

$$\sum_i \sum_j \frac{(n_{ij}/n - \widehat{N}_{ij}/N)^2}{n_{ij}} \quad (2)$$

Stephan, AOA'nın en küçük kareler kriterini sadece yaklaşık olarak minimize ettiğini fark etmiştir (Stephan, 1942). Bu nedenle, en küçük kareler kriterini minimize etmek için farklı bir algoritma önermiştir. Bir diğer çalışmada, Ireland ve Kullback orijinal AOA algoritması değiştirmiş ve algoritmanın ilgi çekici özelliklere sahip olduğunu bulmuşlardır (Ireland ve Kullback, 1968). Bir diğer çalışmada, AOA yöntemi tarafından tahmin edilen \widehat{N}_{ij} 'nin göreceli entropiyi (ya da ayırma bilgi kriterini) minimize ettiği gösterilmiştir (Little ve Wu, 1991; Csisz'ar, 1998).

Entropi genellikle π_{ij} ve $\widehat{\Pi}_{ij}$ olasılıkları ile tanımlanmaktadır.

$$I(\widehat{\Pi}|\pi) = \sum_i \sum_j \widehat{\Pi}_{ij} \log \left(\frac{\widehat{\Pi}_{ij}}{\pi_{ij}} \right) \quad (3)$$

$\widehat{N}_{ij} = N\widehat{\pi}_{ij}$ ve $n_{ij} = n\pi_{ij}$ kabul edildiğinde bu formül,

$$I(\widehat{N}|n) = I(\widehat{\Pi}|\pi) = \log(n/N) + \frac{1}{N} \sum_i \sum_j \widehat{N}_{ij} \log(\widehat{N}_{ij}/n_{ij}) \quad (4)$$

$$I(\widehat{N}|n) = \frac{1}{N} (-N \log \left(\frac{N}{n} \right) + \sum_i \sum_j \widehat{N}_{ij} \log \left(\frac{\widehat{N}_{ij}}{n_{ij}} \right)) \quad (5)$$

olur. Sabit bir hedef popülasyon büyüklüğü (N) için bu eşitlik minimize edilebilmektedir.

$$\sum_i \sum_j \widehat{N}_{ij} \log(\widehat{N}_{ij}/n_{ij}) \quad (6)$$

Entropi çıktısı simetrik değildir çünkü genelde $I(\widehat{N}|n) \neq I(n|\widehat{N})$ olmaktadır. Ireland ve Kullback'ın çalışmasında yakınsama kriterinin ispatı yapılmıştır. Ayrıca çalışmalarında algoritmanın bazı türevleri ve yakınsaması hakkındaki teorik çalışmalara da yer verilmiştir (Ireland ve Kullback, 1968). Csisz'ar tarafından 1975 yılında yayımlanan makalede bu çalışma biraz daha geliştirilmiştir (Csisz'ar, 1975). Csisz'ar'ın yöntemi daha önce yayımlanan konuyla ilgili makalelerden biraz daha genel ele alınmıştır. Özellikle başlangıç hücrelerindeki sıfır değerli hücrelerin çözümü için denklem (7)'deki kural benimsemiştir:

$$\log 0 = -\infty, \log \frac{a}{0} = +\infty, 0 \cdot \pm\infty = 0 \quad (7)$$

Csisz'ar bu kuralı benimsedikten sonra, sıfır değerli hücreler için tolerans değeri ile yakınsamayı kanıtlamıştır. AOA tekniği ve geleneksel tahmini birleştiren diğer bir düzenleme ise, Oh ve Scheuren tarafından önerilmiştir (Oh ve Scheuren, 1987). Thionet AOA tekniğini, doğrusal programlama probleminin bir çözümü olarak incelemiştir (Thionet, 1961; Thionet, 1963; Thionet, 1964).

AOA yöntemi kaynak tabloya ilişkin entropiyi en aza indirirken, (1)'de belirtilen marjinal kısıtları tahmin etmek için de kullanılmaktadır. Entropinin azaltılması ve çapraz tablo oranlarının korunması AOA tekniğinin en önemli özellikleri olarak bilinmektedir. Nümerik analiz yöntemlerinden biri olan Newton yöntemi gibi, (1-7) denklem sistemlerini çözmek için de alternatif yaklaşımlar mevcuttur (Atkinson, 1989). Newton yöntemi, ikinci dereceden (karesel) daha hızlı bir yakınsama değerinin daha iyi olduğunu göstermektedir

(Atkinson, 1989). Bu yöntem aynı zamanda sisteme ilişkin varyans-kovaryans matrisini ve parametreleri de tahmin edebilmektedir. Ancak bu yöntemin hesaplamalı bellekte performansı oldukça düşüktür ve yöntem yüksek boyutlu problemlerde karşımıza çıkan büyük denklem sistemleri için de çok kullanışlı olamamaktadır. Bilgisayar biliminde bilinen asimptotik Landau “O()” gösterimi kullanıldığında (Cormen vd., 1990), C hücreli bir olumsuzluk tablosunu ayarlamak için Newton yöntemi “ $O(C^2)$ ” belleğe ihtiyaç duyarken, AOA yöntemi $O(C)$ kadar bir belleğe ihtiyaç duymaktadır (Agresti, 2002; Fienberg ve Meyer, 2004). Ek olarak, kaynak tablo ile ayarlanmış tablonun karşılaştırılmasında (6)’ncı eşitliğin minimum ayırma bilgisi, olası tek optimizasyon kriteri değildir. Little ve Wu çalışmalarında kaynak örnek ve hedef marjinalerin farklı popülasyonlardan alındığı özel bir durumu incelemişlerdir (Little ve Wu, 1991). Analizlerinde minimum ayırma bilgisi, minimum en küçük kareler, maximum log likelihood ve minimum ki-kare olmak üzere dört farklı performans kriterini karşılaştırmışlardır. Belirli problemler için, diğer optimizasyon kriterleri minimum ayırma bilgisi kriterine göre bazı avantajlar sunabilmektedir. AOA tekniği, örneklem verisinin tablo halinde ifadesini ve bu tablo üzerinde Fienberg tarafından verilen kavramlar kullanılarak birtakım ayarlamalar yapıldıktan sonra, ana kitleye ait göstergelerin tahminini içermektedir (Fienberg, 1968).

AOA yaklaşımı belirli bir kategorideki popülasyonun sayısını, ilgili marjinal toplama eşlik eden hücrelerin korelasyon yapısını ve korunan kontrol değişkenlerinin dağılımını tahmin etmektedir. Çok boyutlu olumsuzluk tablosu ya da çapraz tablo marjinalleri de hücre sıklıklarından oluşmakta ve tüm kontrol değişkenleri karşılıklı bir modda tekrarlanmaktadır. Geçerli kontrol değişkeninin her bir kategorisi için, olasılık tablosuna karşılık gelen popülasyon sayısının toplamına eşitleyebilmek için oransal olarak ayarlanmaktadır. Bu iterasyonların her biri AOA adımı olarak adlandırılmaktadır. Bu iterasyon döngüsü, marjinal toplamaların bağıl hatası, kullanıcı tarafından belirlenen eşik değerine ulaşana kadar sürmektedir. Marjinal kısıtlamaları karşılayan tüm olasılık tabloları arasında sonuç tablosu, başlangıç tablosuna en çok benzeyen tablolardan biridir.

Uygulamada, yakınsama problemleri sadece tüm satır ya da sütunlar sıfır ise ve buna karşılık gelen marjinaler sıfırdan farklı ise ortaya çıkmaktadır. Temel formülasyonda AOA, kitlenin sadece bir düzeyini tahmin etmektedir. Yani; popülasyon ya bireye dayalı nitelikleri ya da hanehalkı niteliklerini tahmin etmekte, ancak ikisini birden aynı anda tahmin edememektedir. Bazen AOA, tüm popülasyondaki bireysel nitelikleri hanehalkı

niteliklerine dönüştürmek için yeterli olabilmektedir. Bu durumda AOA, hanehalkı dağılımını tahmin etmekte kullanılabilir (Arentze vd., 2007). Bazen, kitlenin birden çok düzeyine eş zamanlı olarak ayarlanması için Adımlı Oransal Güncelleme gibi farklı algoritmalara başvurulması gerekebilir. Literatürdeki popülasyon sentezleyicilerinden biri olan PopGen kütüphanesi hariç diğer tüm popülasyon sentezleyiciler ayarlama adımı için AOA tekniğini kullanmaktadırlar. AOA Tekniğinin adımları aşağıdaki gibidir:

1. Hanehalkı kontrol değişkenleri seçilir.
2. Nüfus sayımı özet dosyalarından bu değişkenlere ilişkin marjinal dağılımlar elde edilir.
3. Bir mikro veri, örnek veri setinden (PUMS, trafik anketleri) bileşik dağılım hücre matrisleri kullanılarak oluşturulur.
4. Örnek veriyi verilen marjinal kontrol toplamlarına (tüm nüfus sayısı toplamına) benzetmek amacıyla AOA tekniği kullanarak hücre matrisi ile ifade edilen birleşik dağılım korunur ve hücre matrisi genişletilir.
5. Seçim olasılıkları mikro veri örneğindeki hanehalkları için tahmin edilir.
6. Seçim olasılıkları genişletilmiş hücre frekanslarına eşit olduğunda hanehalkları verileri elde edilir.
7. Elde edilen sentetik popülasyon, uyum iyiliği testi ile kontrol edilir ve gerekirse sentetik popülasyon yeniden hesaplanır.
8. Sonuç olarak, sentetik popülasyon sentezlenen hanehalkları içerisindeki bireylerden oluşur.

AOA tekniğinin kullanılmasına ilişkin bir örnek aşağıda incelenmiştir. Örnek için gerekli olan iki tür veri vardır. Birincisi PUMS ya da hanehalkı seyahat anketlerinden elde edilen mikro alan verisi, ikincisi ise nüfus sayımı özet dosyalarından ya da tahmini olarak elde edilen toplam nüfus verisidir. Bu örnekte amaç hanehalkı geliri ve hanehalkı büyüklüğü kategorileri için mikro bir alandan elde edilen veriler kullanılarak tüm nüfusa yönelik tahminler yapmaktır. Hanehalkı büyüklüğü 3 (1 bireyli hanehalkı, 2 bireyli hanehalkı & 3 ve daha fazla bireyli hanehalkı), hanehalkı geliri ise 2 kategoriden (düşük gelirli hanehalkı & yüksek gelirli hanehalkı) oluşmaktadır. Bu kategorilere ait mikro alan verisi (PUMS) ve tüm nüfusa ait toplam veri Şekil 3'te gösterilmiştir.

PUMS ya da Hanehalkı Seyahat Anketleri'nden elde edilen veri						
Hanehalkı Geliri İçin Ayarlama		Hanehalkı Büyüklüğü İçin Ayarlama	Hanehalkı Gelir Kategorisi		Toplam	Hanehalkı Büyüklüğü Marjinaleri
			Düşük	Yüksek		
Hanehalkı Büyüklüğü Kategorisi	1	---	3.0	1.0	4.0	30.0
	2	---	2.0	4.0	6.0	40.0
	3 ve Daha Fazlası	---	2.0	1.0	3.0	30.0
Toplam			7.0	6.0		
Hanehalkı Gelir Marjinaleri			60.0	40.0		

Nüfus Sayımı Özet Dosyaları'ndan elde edilen veri

Şekil 3. AOA tekniği örneği için gerekli veriler

- Adım 1a: Hanehalkı Geliri için ayarlama

Hanehalkı Geliri kategorisi için ayarlama 3 adımda gerçekleşir: Ayarlama katsayılarının/ağırlıklarının belirlenmesi, sıklıkların ayarlanması ve genel toplamların ayarlanmasıdır. Örneğe ait ilk hanehalkı geliri ayarlaması Şekil 4'te gösterilmiştir.

Adım 1a: Hanehalkı Geliri İçin Ayarlama		Hanehalkı Gelir Kategorisi		Toplam	Hanehalkı Büyüklüğü Marjinaleri
		Düşük	Yüksek		
1 Ayarlama Hanehalkı Geliri İçin Ayarlama		$60/7=8.57$	$40/6=6.7$		
Hanehalkı Büyüklüğü Kategorisi	1	$3.0*8.57=25.7$	$1.0*6.67=6.7$	32.4	30.0
	2	$2.0*8.57=17.1$	$4.0*6.67=26.7$	43.8	40.0
	3 ve Daha Fazlası	$2.0*8.57=17.1$	$1.0*6.67=6.7$		30.0
Toplam		60.0	40.0		
Hanehalkı Gelir Marjinaleri		60.0	40.0		

3 Toplamların Ayarlanması

Şekil 4. Adım 1a-Hanehalkı geliri için ayarlama

- Adım 1b: Hanehalkı Büyüklüğü için ayarlama

Hanehalkı Büyüklüğü kategorisi için ayarlama 3 adımda gerçekleşir: Ayarlama katsayılarının/ağırlıklarının belirlenmesi, sıklıkların ayarlanması ve genel toplamların ayarlanmasıdır. Örneğe ait ilk hanehalkı büyüklüğü ayarlaması Şekil 5’te gösterilmiştir.

Adım 1b: Hanehalkı Büyüklüğü İçin Ayarlama		Hanehalkı Gelir Kategorisi		Toplam	Hanehalkı Büyüklüğü Marjinaleri
		Düşük	Yüksek		
Hanehalkı Geliri İçin Ayarlama		Hanehalkı Büyüklüğü İçin Ayarlama	---	---	
Hanehalkı Büyüklüğü Kategorisi	1	$30.0/32.4=0.93$	$25.7*0.93=23.8$	$6.7*0.93=6.2$	30.0
	2	$40.0/43.8=0.91$	$17.1*0.91=15.7$	$26.7*0.91=24.3$	40.0
	3 ve Daha Fazlası	$30.0/23.8=1.26$	$17.1*1.26=21.6$	$6.7*1.26=8.4$	30.0
Toplam			61.1	38.9	
Hanehalkı Gelir Marjinaleri		1 Ayarlama	60.0	40.0	3 Toplamların Ayarlanması
		2 Sıklıkların Ayarlanması			

Şekil 5. Adım 1b-Hanehalkı büyüklüğü için ayarlama

Adım 1’de hem Hanehalkı Geliri hem de Hanehalkı Büyüklüğü için ayarlama yapılmıştır. Yapılan bu ayarlama sonucunda hanehalkı büyüklüğü kategorisine ait yapılan ayarlamaların nüfus sayımından elde edilen hanehalkı büyüklüğü marjinalerine eşit olduğu görülmektedir. Ancak, Hanehalkı Geliri kategorisine ait yapılan ayarlamalar, nüfus sayımından elde edilen Hanehalkı Gelir marjinalerine eşit olmadığından ayarlama işlemine devam edilir.

- Adım 2a: Hanehalkı Geliri için 2. ayarlama

Bu adımda Adım 1a’da olduğu gibi hanehalkı geliri için yeniden ağırlıklar elde edilir, sıklıklar ve toplam marjinaler ayarlanır. Bu adıma ait ayarlamalar Şekil 6’da gösterilmiştir.

Adım 2a: Hanehalkı Geliri İçin Ayarlama		Hanehalkı Gelir Kategorisi		Toplam	Hanehalkı Büyüklüğü Marjinalleri
		Düşük	Yüksek		
Hanehalkı Geliri İçin Ayarlama		Hanehalkı Büyüklüğü İçin Ayarlama	$60/61.1=0.98$	$40/38.9=1.03$	
Hanehalkı Büyüklüğü Kategorisi	1	---	$23.8*0.98=23.4$	$6.2*1.03=6.3$	29,8
	2	---	$15.7*0.98=15.4$	$24.3*1.03=25$	40,4
	3 ve Daha Fazlası	---	$21.6*0.98=21.2$	$8.4*1.03=8.6$	29,9
Toplam			60.0	40.0	
Hanehalkı Gelir Marjinalleri			60.0	40.0	

Şekil 6. Adım 2a-Hanehalkı geliri için 2. ayarlama

- Adım 2b: Hanehalkı Büyüklüğü için 2. ayarlama

Bu adımda Adım 1b'de olduğu gibi hanehalkı büyüklüğü için yeniden ağırlıklar elde edilir, sıklıklar ve toplam marjinaller ayarlanır. Bu adıma ait ayarlamalar Şekil 7'de gösterilmiştir.

Adım 2b: Hanehalkı Büyüklüğü İçin Ayarlama		Hanehalkı Gelir Kategorisi		Toplam	Hanehalkı Büyüklüğü Marjinalleri
		Düşük	Yüksek		
Hanehalkı Geliri İçin Ayarlama		Hanehalkı Büyüklüğü İçin Ayarlama	---	---	
Hanehalkı Büyüklüğü Kategorisi	1	$30.0/29,8=1.01$	$23.4*1.01=23.6$	$6.3*1.01=6.4$	30.0
	2	$40.0/40.4=0.99$	$15.4*0.99=15.2$	$25*0.99=24.8$	40.0
	3 ve Daha Fazlası	$30.0/29,9=1.00$	$21.2*1.00=21.3$	$8.6*1.00=8.7$	30.0
Toplam			60.2	39.8	
Hanehalkı Gelir Marjinalleri			60.0	40.0	

Şekil 7. Adım 2b-Hanehalkı büyüklüğü için 2. ayarlama

Adım 2'de hem hanehalkı geliri hem de hanehalkı büyüklüğü için ikinci bir ayarlama yapılmıştır. Yapılan bu ikinci ayarlama sonucunda hanehalkı büyüklüğü kategorisine ait yapılan ayarlamaların nüfus sayımından elde edilen hanehalkı büyüklüğü marjinallerine eşit olduğu görülmektedir. Ancak hanehalkı gelir kategorisine ait yapılan ayarlamalar, nüfus

sayımından elde edilen hanehalkı gelir marjinallerine eşit olmadığından ayarlama işlemine devam edilir.

- Adım 3a: Hanehalkı Geliri için 3. Ayarlama

Burada diğer adımlarda olduğu gibi hanehalkı geliri için yeniden ağırlıklar elde edilir, sıklıklar ve toplam marjinaller ayarlanır. Bu adıma ait ayarlamalar Şekil 8’de gösterilmiştir.

Adım 3a: Hanehalkı Geliri İçin Ayarlama		Hanehalkı Gelir Kategorisi		Toplam	Hanehalkı Büyüklüğü Marjinaleri
		Düşük	Yüksek		
Hanehalkı Geliri İçin Ayarlama		Hanehalkı Büyüklüğü İçin Ayarlama	$60/60.2=1.00$	$40/39.8=1.00$	
Hanehalkı Büyüklüğü Kategorisi	1	---	$23.6*1.00=23.5$	$6.4*1.00=6.4$	30.0
	2	---	$15.2*1.00=15.2$	$24.8*1.00=24.9$	40.1
	3 ve Daha Fazlası	---	$21.3*1.00=21.3$	$8.7*1.00=8.7$	30.0
Toplam			60.0	40.0	
Hanehalkı Gelir Marjinaleri			60.0	40.0	

Şekil 8. Adım 3a-Hanehalkı geliri için 3. ayarlama

- Adım 3b: Hanehalkı Büyüklüğü için 3. Ayarlama

Burada diğer adımlarda olduğu gibi hanehalkı büyüklüğü için yeniden ağırlıklar elde edilir, sıklıklar ve toplam marjinaller ayarlanır. Bu adıma ait ayarlamalar Şekil 9’da gösterilmiştir.

Adım 3b: Hanehalkı Büyüklüğü İçin Ayarlama		Hanehalkı Gelir Kategorisi		Toplam	Hanehalkı Büyüklüğü Marjinaleri	
		Düşük	Yüksek			
Hanehalkı Geliri İçin Ayarlama		Hanehalkı Büyüklüğü İçin Ayarlama	---	---		
Hanehalkı Büyüklüğü Kategorisi	1	$30.0/30.0=1.00$	$23.5*1.00=23.6$	$6.4*1.00=6.4$	30.0	30.0
	2	$40.0/40.1=1.00$	$15.2*1.00=15.2$	$24.9*1.00=24.8$	40.0	40.0
	3 ve Daha Fazlası	$30.0/30.0=1.00$	$21.3*1.00=21.3$	$8.7*1.00=8.7$	30.0	30.0
Toplam			60.0	40.0		
Hanehalkı Gelir Marjinaleri			60.0	40.0		

Hanehalkı Sıklıkları
Yakınsama Sağlandı

Şekil 9. Adım 3b-Hanehalkı büyüklüğü için 3. ayarlama

Adım 3'te hem hanehalkı geliri hem de hanehalkı büyüklüğü için üçüncü bir ayarlama yapılmıştır. Yapılan bu ayarlama sonucunda hem hanehalkı büyüklüğü kategorisine hem de hanehalkı geliri kategorisine ait yapılan ayarlamaların nüfus sayımından elde edilen ilgili kategoriye ait marjinallerine eşit olduğu görülmektedir. Bu da üçüncü adımda istenen yakınsamanın sağlandığını göstermektedir. Elde edilen hanehalkı sıklıkları Şekil 9'da gösterilmektedir. Mikro bir alandan elde edilen hanehalkı kategorilerine ait verilere AOA tekniği uygulanılarak, tüm nüfusa ait verilerin ilgili kategorilere uygun bir şekilde dağıtıldığı görülmektedir.

Özetle AOA yöntemi bir veri birleştirme tekniğidir ve ölçümleme tahmincilerinin bir türüdür. Hedef popülasyon için kaynak çok boyutlu olumsuzluk tabloları ve daha küçük boyutlu marjinal tablolardan elde ettiği bilgileri birleştirir. Yöntem kaynak tabloya göre ayırma bilgi kriterini minimize ederken marjinal tablolara mükemmel uyum sağlar.

Ayrıca bireyleri konumlara atamak için kullanılan en sık ve en iyi bilinen deterministik yöntem AOA tekniğidir. AOA yöntemi, bireyleri bölgelere atayan basit, hızlı ve yaygın kullanılan bir yöntem olarak uzun süredir kullanılmaktadır. AOA her bireyin hangi bölgeye nasıl ait olduğunu gösteren tamsayı olmayan ağırlıkları hesaplamayı içerir. Bu bir yeniden ağırlıklandırma yöntemidir. Uygulama yönteminden bağımsız olarak AOA ağırlık matrisi ile temsil edilen her birey bölge kombinasyonu için maksimum olabilirlik

değerlerinin hesaplanmasıyla bireyleri bölgelere atamak için kullanılan bir yöntemdir. Bu bölgelerdeki bireylerin en olası yapılandırılması olarak görülebilir.

1.2.1.2. Adımlı Oransal Güncelleme (AOG) Tekniği

Popülasyon sentezleyiciler genellikle standart AOA yaklaşımı kullanarak ilgilenilen değişkenlerin birleşik dağılımlarını üretmek için hanehalkı özniteliklerinin nüfus sayımına dayalı marjinal dağılımlarını kullanmaktadır. Hanehalkları bileşik dağılımla uyumlu olarak mevcut örnekten rasgele bir şekilde oluşturulmaktadır. Fakat bu geleneksel yaklaşımlar, bireysel özniteliklerin bileşik dağılımlarını ve bireysel düzeydeki özelliklerini kontrol etmemektedir. AOG yaklaşımı, AOA yaklaşımının aksine sezgisel bir yaklaşımdır. AOG, hem hanehalkı düzeyli hem de birey düzeyli ilgilenilen özellikleri, etkili bir hesaplama yöntemi ile simüle ederek sentetik popülasyonlar üretmektedir. Bu algoritma, hem hanehalkı hem de birey özelliklerini simüle ederken belirli bir tipin (birleşik dağılımdaki hücre) hanehalkları arasındaki ağırlıkları ayarlama ve yeniden belirleme işlemlerini tekrarlayarak hesaplamaktadır (Ye vd., 2009). PopGen popülasyon sentezleyicisinde temel fikir bu algoritmaya dayanmaktadır. AOG süreci sırasında çoklu hiyerarşi düzeyleri için eş zamanlı olarak kontroller sağlanmaktadır. Önerilen algoritma AOA tekniğinin seyrek liste varyantı ile birçok benzerliğe sahiptir. Hanehalkının niteliklerindeki farklılıkları açıklamak için örnek hanehalkı kullanılarak ağırlıklar yeniden elde edilmektedir. Güncelleme işlemi yakınsama sağlanıncaya kadar devam eder ve sonuç olarak, ağırlıklar hem birey düzeyli hem de hanehalkı düzeyli kısıtlara eşitlenmektedir. Ye vd. göre AOG tekniği uygulamada iyi performans sergileyerek diğer benzerlerinden ayrılmaktadır. Yazarlar algoritmanın geometrik bir açıklamasını sunarak, yakınsamanın teorik bir ispatının olmadığını belirtmişlerdir (Ye vd., 2009). AOG Tekniğinin adımları aşağıdaki gibidir:

1. Hanehalkı ve birey tipi kısıtları tahmin etmek: Bu adım için gerekli veriler, hanehalkı ve birey örnek verileri ile hanehalkı ve bireysel düzeyde marjinal dağılımlardır. Bu adımda sıfır hücre problemini ayarlamak için önsel ayarlamalar ve sıfır marjinal problemini ayarlamak için marjinal ayarlamalar yapılmaktadır. Hanehalkı ile birey tipi kısıtları hesaplamak için AOA tekniği kullanılmaktadır.
2. Hanehalklarının ağırlıklarının tahmini: Bu adım için gerekli veriler hanehalkı/birey örnek verileri ve birinci adımdan elde edilen hanehalkı/birey

kısıtlarıdır. Bu adımda hem hanehalkı hem de birey kısıtlarını karşılayan örnek hanehalkı ağırlıklarının tahmin etmek için AOG algoritması çalıştırılmaktadır.

3. Sentetik popülasyon üretmek: Bu adım için gerekli veriler, hanehalkı/birey örnek verileri ve ikinci adımda elde edilen hanehalkı ağırlıklarıdır. Bu adımda sentetik popülasyondaki farklı hanehalkı tiplerinin sıklıklarının elde etmek için yuvarlama işlemleri uygulanmaktadır. Hesaplanan ağırlıklar kullanılarak hanehalkı seçim olasılıklarının tahmin edilmektedir. Her bir hanehalkını hücre frekanslarına benzetmek için seçim olasılıklarına dayalı olarak örnek hanehalkları elde edilmektedir. En uygun sentetik popülasyon elde edilene kadar süreç tekrar etmektedir.

Uyum ölçüsü (δ değeri): AOG tekniği ile ayarlanmış ağırlıklı toplam ve hedef popülasyon kısıtları arasındaki mutlak göreceli sapmaları ölçmektedir. Tüm kısıtlamaları karşılayan bu değer, uyum iyiliği ölçüsü olarak kullanılmaktadır. Bu değer aynı zamanda izleme için de kullanılabilir ve bu amaçla AOG tekniği algoritması için bir yakınsama kriteri olarak kullanılmaktadır.

$$\delta_i = \frac{|d_{i,j}w_i - c_j|}{c_j} \quad (8)$$

$d_{i,j}w_i$: AOG ile ayarlanmış ağırlıklı toplam

c_j : Hedef popülasyon kısıtı

AOG tekniğinin kullanılmasına ilişkin bir örnek aşağıda incelenmiştir. Örnek için gerekli olan hanehalkı ve birey tipine ait mikro alan verisi ve nüfus sayımı özet dosyalarından elde edilen toplam veri ve kısıtlar Şekil 10'da gösterilmiştir. Hanehalkı tipi iki, birey tipi ise üç kategoriden oluşmaktadır.

HANE HALKI KİMLİĞİ	BAŞLANGIÇ AĞIRLIKLARI	HANEHALKI TİPİ1	HANEHALKI TİPİ2	BİREY TİPİ1	BİREY TİPİ2	BİREY TİPİ3
1	1	1	0	1	1	1
2	1	1	0	1	0	1
3	1	1	0	2	1	0
4	1	0	1	1	0	2
5	1	0	1	0	2	1
6	1	0	1	1	1	0
7	1	0	1	2	1	2
8	1	0	1	1	1	0
Ağırlıklı Toplam		3.00	5.00	9.00	7.00	7.00
Kısıtlar		35.00	65.00	91.00	65.00	104.00
δ0		0.9143	0.9231	0.9011	0.8923	0.9327

Şekil 10. AOG tekniği örneği için gerekli veriler

- Adım 1: Hanehalkı Tipi1 ve Hanehalkı Tipi2 için ayarlama

Hanehalkı Tipi kategorisi için ayarlamalar Şekil 11’de gösterilmiştir. Öncelikle Hanehalkı Tipi1 için birinci ağırlıklar hesaplanır. Elde edilen bu ağırlık hücre sıklıkları ile çarpılarak yeni ağırlıklı toplam elde edilir (Ağırlıklı Toplam1). Daha sonra Hanehalkı Tipi2 için ikinci ağırlıklar hesaplanır. Elde edilen bu yeni ağırlık, hücre sıklıkları ile çarpılarak ikinci ağırlıklı toplam elde edilir (Ağırlıklı Toplam2).

Hane Halkı Kimliği	Başlangıç Ağırlıkları	Hanehalkı Tipi1	Hanehalkı Tipi2	Birey Tipi1	Birey Tipi2	Birey Tipi3	Ağırlıklar1	Ağırlıklar2
1	1	1	0	1	1	1	11,67	11,67
2	1	1	0	1	0	1	11,67	11,67
3	1	1	0	2	1	0	11,67	11,67
4	1	0	1	1	0	2	1.00	13.00
5	1	0	1	0	2	1	1.00	13.00
6	1	0	1	1	1	0	1.00	13.00
7	1	0	1	2	1	2	1.00	13.00
8	1	0	1	1	1	0	1.00	13.00
Ağırlıklı Toplam0		3.00	5.00	9.00	7.00	7.00		
Kısıtlar		35.00	65.00	91.00	65.00	104.00	35/3=11.67	65/5=13.00
δ0		0.9143	0.9231	0.9011	0.8923	0.9327		
Ağırlıklı Toplam1		35.00	5.00	51.67	28.33	28.33		
Ağırlıklı Toplam2		35.00	65.00	111.67	88.33	88.33		

= Birey Tipi1 * Ağırlıklar1
= Birey Tipi3 * Ağırlıklar2

Şekil 11. Adım 1-Hanehalkı tipi1 ve hanehalkı tipi2 için ayarlama

- Adım 1: Birey Tipi1, Birey Tipi2 ve Birey Tipi3 için ayarlama & Uyum Ölçüsünün hesaplanması

Bu adımda Birey Tipi kategorisi için ayarlamalar yapılmış ve bu ayarlamalardan sonra uyum iyiliği hesaplanmıştır. Bu adımda yapılan hesaplamalar ve ölçümler Şekil 12'de gösterilmiştir. Birey Tipi1 için üçüncü ağırlıklar hesaplanmış ve elde edilen bu ağırlıklar hücre sıklıkları ile çarpılarak Ağırlıklı Toplam3 elde edilmiştir. Daha sonra aynı işlemler Birey Tipi2 için yapılarak Ağırlıklı Toplam4 ve Birey Tipi3 için yapılarak Ağırlıklı Toplam5 elde edilmiştir. Birey Tipi kısıtlarına ait ayarlamalar yapıldıktan sonra yakınsamanın sağlanıp sağlanmadığını kontrol etmek amacıyla uyum ölçüsü değerleri hesaplanmıştır. Uyum ölçüsü değerlerinin sıfır ya da sıfıra çok yakın olması yakınsamanın sağlandığını göstermektedir. Ancak birinci adım sonucunda hesaplanan uyum ölçüsü değerlerinin ($\delta 1$) sıfırdan farklı olduğu Şekil 12'de görülmektedir.

Hane Halkı Kimliği	Başlangıç Ağırlıkları	Hanehalkı Tipi1	Hanehalkı Tipi2	Birey Tipi1	Birey Tipi2	Birey Tipi3	Ağırl.1	Ağırl.2	Ağırl.3	Ağırl.4	Ağırl.5
1	1	1	0	1	1	1	11,67	11,67	$11.67*0.81=9.51$	$9.51*0.85=8.05$	12,37
2	1	1	0	1	0	1	11,67	11,67	$11.67*0.81=9.51$	9,51	14.61
3	1	1	0	2	1	0	11,67	11,67	$11.67*0.81=9.51$	$9.51*0.85=8.05$	8,05
4	1	0	1	1	0	2	1.00	13.00	$13.00*0.81=10.59$	10,59	16.28
5	1	0	1	0	2	1	1.00	13.00	13.00	$13.00*0.85=11.00$	16.91
6	1	0	1	1	1	0	1.00	13.00	$13.00*0.81=10.59$	$10.59*0.85=8.97$	8,97
7	1	0	1	2	1	2	1.00	13.00	$13.00*0.81=10.59$	$10.59*0.85=8.97$	13.78
8	1	0	1	1	1	0	1.00	13.00	$13.00*0.81=10.59$	$10.59*0.85=8.97$	8,97
Ağırlıklı Toplam0		3.00	5.00	9.00	7.00	7.00					
Kısıtlar		35.00	65.00	91.00	65.00	104.00	$35/3=11.67$	$65/5=13.00$	$91/111.67=0.81$	$65/76.80=0.85$	$104/67.68=1.54$
$\delta 0$		0.9143	0.9231	0.9011	0.8923	0.9327					
Ağırlıklı Toplam1		35.00	5.00	51.67	28.33	28.33					
Ağırlıklı Toplam2		35.00	65.00	111.67	88.33	88.33					
Ağırlıklı Toplam3		28.52	55.38	91.00	76.80	74.39					
Ağırlıklı Toplam4		25.60	48.50	80.11	65.00	67.68					
Ağırlıklı Toplam5		35.02	64.90	104.84	85.94	104.00					
$\delta 1$		$(35.02-35)/35=0.0006$	$(64.90-65)/65=0.0015$	$(104.84-91)/91=0.1521$	$(85.94-65)/65=0.3222$	$(104-104)/104=0$					

Şekil 12. Adım 1-Birey tipi1, birey tipi2 ve birey tipi3 için ayarlama & uyum ölçüsünün hesaplanması

Hanehalkı Tipi ve Birey Tipi kategorileri için ayarlamalar yapılmış ve uyum ölçüsü değerlerinin sıfıra yakınması 401. Adımda gerçekleşmiştir. 401. Adıma ait Nihai Ağırlıklı Toplam, Nihai AOG, Nihai AOA Ağırlıkları ve Uyum Ölçüsü değerleri Şekil 13'te gösterilmiştir.

Hanehalkı Kimliği	Başlangıç Ağırlıkları	Hanehalkı Tipi1	Hanehalkı Tipi2	Birey Tipi1	Birey Tipi2	Birey Tipi3	Nihai AOG Ağırlıkları	AOA Ağırlıkları
1	1	1	0	1	1	1	1,36	11,67
2	1	1	0	1	0	1	25,66	11,67
3	1	1	0	2	1	0	7,98	11,67
4	1	0	1	1	0	2	27,79	13,00
5	1	0	1	0	2	1	18,45	13,00
6	1	0	1	1	1	0	8,64	13,00
7	1	0	1	2	1	2	1,47	13,00
8	1	0	1	1	1	0	8,64	13,00
Ağırlıklı Toplam0		3.00	5.00	9.00	7.00	7.00		
Kısıtlar		35.00	65.00	91.00	65.00	104.00		
δ0		0.9143	0.9231	0.9011	0.8923	0.9327		
Nihai Ağırlıklı Toplam		35.00	65.00	91.00	65.00	104.00		
δ401(AOG)		0.0000	0.0000	0.0000	0.0000	0.0000		

Şekil 13. Adım 401-Nihai ağırlıklı toplam, nihai AOG ve AOA ağırlıkları ve uyum ölçüsü sonuçları

Şekil 13'ten Nihai Ağırlıklı Toplamının başlangıçta elde edilen nüfus verisi kısıtlarına eşit olduğu görülmektedir. Mikro alandan elde edilen hanehalkı ve birey tipi kategorilerine ait verilere AOG tekniği uygulanılarak, tüm nüfusa ait verilerin ilgili kategorilere uygun bir şekilde dağıtıldığı görülmektedir.

Özetle, hanehalkları ağırlıklarını tahmin etmek için AOG tekniği kullanılmaktadır. Bu algoritma hem hanehalkı hem de birey dağılımlarını sentezlemek için örnek hanehalkı dağılımlarını tahmin eder. Sezgisel yorumuyla basit, pratik ve sayısal olarak izlenebilir bir yöntemdir.

1.3. Mikro Simülasyon

Mikro simülasyon ilk defa Guy Orcutt tarafından kullanılmıştır (Orcutt, 1957). Mikro simülasyon, belirli bir çevrede yaşayan bireyler, hane halkları ve firmalar gibi

aktörlerin durumlarını ve davranışlarını simüle etmek için kullanılan bir “metodoloji” olarak tanımlanmaktadır. Mikro simülasyon büyük ölçekli hanehalkları yaratmak için ya da bireysel düzeyli veri setleri yaratmak için kullanılan bir yöntem olarak bilinmektedir. Bu yöntem, veri üretiminin ardından nüfus yapılarındaki ya da devlet politikalarındaki değişikliklerin etkilerini incelemek için “... olursa ne olur” analizlerinde kullanılmaktadır. Mikro simülasyon, Orcutt ve çağdaşlarının 1961 yaptığı çalışmadan sonra sosyal bilimlerde sıklıkla kullanılan yöntemlerden biri haline gelmiştir (Orcutt vd., 1961). Birkaç yazar Orcutt’un çalışmasını genişletmiştir (Sutherland, 1995; Orcutt ve Glazer, 1980; Zaidi vd., 2009). Örneğin devlet politikaları uygulamadan önce, politikaların etkilerini incelemek için sürece bu metodolojiyi dahil edilmişlerdir. Mikro simülasyonun temel önceliği topluluk davranışının daha gerçekçi bir resminin, bireysel davranışa bakarak ve dikkate alınan sistemdeki birimler arasındaki etkileşimleri modelleyerek sağlanabildiğidir. Mikro simülasyon modelleme metodolojileri, sosyal ve ekonomik politikaların değerlendirilmesinde, vergi-fayda analizinde ve diğer kamu politikaları alanlarında uzun süredir kullanılmaktadır. Vergi değişikliklerinin milli gelir üzerindeki etkisi ya da sağlık politikası değişikliklerinin popülasyon üzerindeki etkisi gibi modellemelerin çoğunda mikro simülasyon ulusal ölçekte uygulanmıştır. Coğrafyacılara tarafından uygulanan mikro simülasyon, bu alana konumsal bir boyut kazandırmıştır. Son birkaç yılda literatürde uygulanan konumsal mikro simülasyon yöntemlerinin sayısı oldukça fazladır.

1.4. Konumsal Mikro Simülasyon

Konumsal mikro simülasyon, bireysel ve coğrafi düzeyde eş zamanlı meydana gelen olayları modellemede kullanılan bir yöntemdir. Örneğin modern bir şehrin işleyişi insan etkileşimlerinin karmaşık bir ağını içermektedir. Böyle bir karmaşıklığı simüle etmek imkânsız görülebilmektedir. Ancak bu karmaşıklığı, konumsal mikro simülasyon onu oluşturan parçalara bölerek (ayrık coğrafi alanlara ve popülasyonun alt örneklerine gibi), günlük kullanılan bir dizüstü bilgisayarla şehir sisteminin kilit yönlerini modellemek için kullanılabilir.

Konumsal mikro simülasyonun, yönteminin doğası gereğiyle ve insanları modelleyebilme kabiliyetiyle çeşitli alanlarda kullanılması faydalı olmaktadır. Örneğin, çalışılan sahada ikamet eden bireylerin eğitiminin geleceği hakkında ve yerel düzeyde sağlık ve ulaşım hizmetleri hakkında planlamalar yapmak için konumsal mikro veriden

yararlanmak faydalı olabilmektedir. Gelecek elbette belirsizdir ancak konumsal mikro simülasyon gelecek hakkında çeşitli tahminleri yapmak için senaryo temelli planlamalar sunarak geleceği tahmin etmede fayda sağlayabilmektedir.

Konumsal mikro simülasyon genellikle akademik çalışmalarda kullanılan bir konu olarak görülmektedir. Bazı modelleme girişimlerinin (Hagerstrand, 1952; Wilson ve Pownall, 1976) konumsal mikro simülasyon olduğu düşünülürken, konumsal mikro simülasyonun ilk modelinin Clarke vd. (Clarke vd., 1984) tarafından geliştirilen sağlık planlaması modeli olduğu bilinmektedir. “HIPS (Health Information and Planning System)” olarak bilinen bu model, İngiliz sağlık merkezi yöneticileri için geliştirilmiştir. Bu model her konum için toplamsal veriden bir başlangıç popülasyonu oluşturmaktadır. Bu başlangıç popülasyonun demografik özellikleri ise her yıl yenilenmektedir. Clarke bu çalışmadan sonra konumsal mikro simülasyon ile ilgili birçok makalede ve kitapta da yer almıştır (Clarke ve Wilson, 1985; Clarke ve Holm, 1987; Birkin ve Clarke, 1988; Birkin ve Clarke, 1989). Birkin ve Clarke AOA yöntemini kullanarak “Sentez” olarak adlandırılan konumsal mikro simülasyon yöntemini geliştirmiştir (Birkin ve Clarke, 1988). Clarke, bütün bu önemli çalışmalarından dolayı konumsal mikro simülasyonun öncüsü olarak bilinmektedir. Açık kaynaklı yazılımların varlığı ve veri girişi kolaylığı gibi bilgisayar bilimlerindeki son gelişmeler ile birlikte, konumsal mikro simülasyon günümüzde hiç olmadığı kadar daha geniş bir soru yelpazesine cevap vermektedir. Fakat konumsal mikro simülasyon ve bununla ilişkili yöntemler hala emekleme aşamasında olduğu için, konunun henüz çalışılmamış birçok araştırma alanı da bulunmaktadır. Konumsal mikro simülasyon son zamanlarda çok çalışılan bir konu olmasına rağmen, konu hala birçok insana uzak gelmektedir.

Şirketlerin, hanehalklarının ve hükümetlerin çok fazla olduğu bir dünya farz edilsin, kamu yararına çalışan araştırmacıların ve karar mercilerinin gelecek hakkında farklı senaryoları bulmak için konumsal bir mikro veriyi modellemek ve onları test etmek için araçları olduğunu düşünölsün. Bu durumda insanlar ellerindeki bu sonuçlara dayalı olarak gelecekte daha bilinçli karar verebilirlerdi. Bu ütöpik hayal dünyasında iklim değışikliği, eşitsizlik ve hastalık gibi problemler çözülebilirdi. Bu problemler aslında günümüz dünyasına ait sorunlardır ve konumsal mikro simülasyon bu sorunlara bir çözüm olabilmektedir. Konumsal mikro simülasyon karmaşık problemlere yeni yaklaşımlar sağlayabilmektedir ve sonuç olarak daha iyi karar vermeye olanak sağlayan bir yöntem olabilmektedir. Bu yöntem mevcut bilgiler üzerine ışık tutarak, ideolojik önyargılardan

uzak bir şekilde karar verme süreçlerinde değişiklikler yapmaya fayda sağlayabilmektedir. Araştırmacılar genellikle eksik ya da erişilemeyen veriler ile çalışmaktadır. Mevcut veri setleri, karmaşık süreçleri anlamak için gerekli konuma ya da zamana ait verilerini içermemektedir. Halka açık bulunan veri setleri, örneğin gelir gibi konumsal mikro simülasyon için gerekli kilit verilerden yoksundur. Veri kullanımını ve erişimini engelleyen katı koşullar vatandaşların gizliliğini korumaya amaçlamaktadır. Ancak bu durum demokratik karar vermeye de engel olabilmektedir.

Yüksek kaliteli veriler sağlandığı zaman da bu verilerden yararlanarak sonuçlar üretmek de çok zor olabilmektedir. Ancak, bilgi güçtür ve bu yüzden konumsal mikro simülasyon gibi yöntemler günümüzde çok talep görmektedir.

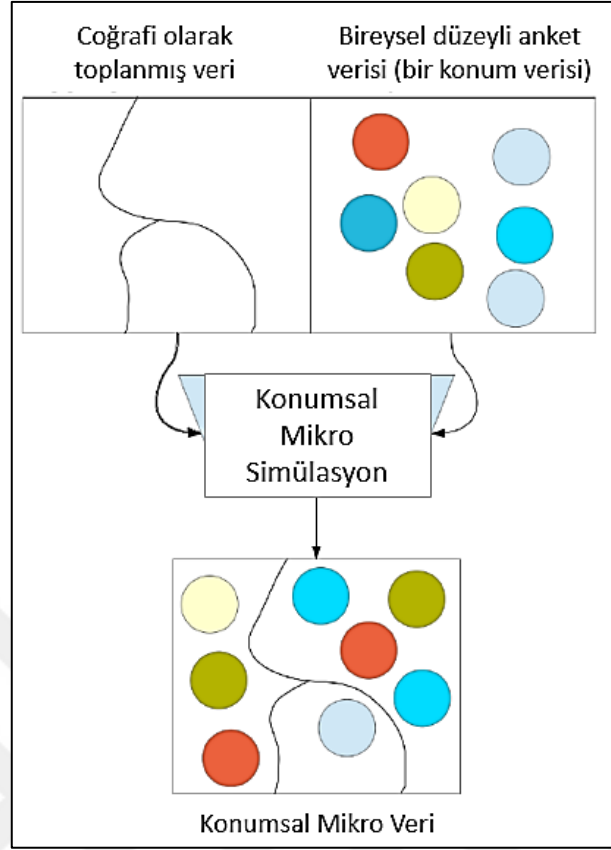
Konumsal mikro simülasyon sosyal bilimlerdeki karmaşık konuları incelemektedir ve son zamanlarda çok kullanılan bir yaklaşım olarak karşımıza çıkmaktadır. Bu yaklaşım ulaşım, sağlık, eğitim gibi farklı alanlarda yaygın olarak kullanılmaktadır ve aynı şekilde bu yaklaşımı daha birçok alana da uygulamak mümkündür. Yaklaşımın temelinde, büyük saha çözümlerinde yer alan bireysel düzeydeki verilerden yararlanarak bireylerin konumlara atanması gibi süreçler yer almaktadır. Konumsal mikro veri, tüm konumsal mikro simülasyon araştırmaları için temel oluşturmaktadır.

Popülasyon sentezi, konumsal mikro simülasyonda önemli bir safha olarak bilinmektedir. Genellikle konumsal mikro veri, veri gizliliği konusundaki endişeler nedeniyle kullanılmamaktadır. Genellikle sentetik konumsal mikro veri tabanları, nüfus sayımı verilerinden elde edilen toplamsal veriler ile ilgilenilen popülasyonu temsil eden anketlerden elde edilen bireysel düzeyli mikro verilerin birleştirilmesiyle oluşturulmaktadır. Elde edilen konumsal mikro veri, bireysel düzeyde ve coğrafi olarak belirli süreçlerin devreye girdiği durumlarda faydalı olmaktadır. Konumsal mikro simülasyon, çok değişkenli modellemeye ve analize de olanak sağlamaktadır. Konumsal mikro simülasyon aynı zamanda birey tabanlı modellemeyle de örtüşmektedir. Konumsal mikro simülasyondan elde edilen konumsal mikro veri, birey tabanlı modelleme için başlangıç verisi olarak kullanılmaktadır. Günümüz dünya sonunlarına çok fazla faydası olmasına rağmen konumsal mikro simülasyon, bölgesel bilim alanlarının dışında çok az bilinmektedir. Aslında konumsal simülasyon, kamu politikalarını bilgilendirmek amacıyla yeni alanlara uygulanabilecek büyük bir potansiyele sahiptir. Sosyal fayda çalışmaları; konumsal mikro simülasyon kullanılarak barınmada, ulaşım ve sürdürülebilir kentsel planlamada halihazırda zaten kullanılmaktadır.

Günümüzde konumsal mikro simülasyon konusuna ilgi giderek artmaktadır. Bilgisayarlardaki, yazılımlardaki ve veri erişilebilirliğindeki gelişmeler sayesinde; konumsal mikro simülasyon yöntemi eskisinden daha fazla kullanılabilir duruma gelmiştir. Neredeyse dünyanın her yerinde bireysel düzeyli küçük alanların popülasyonlarını simüle etmek artık konumsal mikro simülasyonla mümkün olmaktadır.

Yine de konumsal mikro simülasyonun anlamı, birçok kişi için belirsizlik yaratmaktadır. Konumsal mikro simülasyon terimini çevreleyen belirsizlik; yöntemin doğal olarak karmaşık olmasından, yöntemin çok değişkenli çalışmasından ve kısmen de araştırmacıların kendilerinden kaynaklanan nedenlerden dolayı ortaya çıkmaktadır. Akademik literatürdeki konumsal mikro simülasyon çalışmalarının bazıları, konumsal mikro simülasyon teriminin anlamı konusunda net olamamıştır ve bu nedenle konumsal mikro simülasyonun anlamı hakkında çok fazla tutarsızlıklar bulunmaktadır.

Konumsal mikro simülasyon, bireysel ve coğrafi olarak toplanmış veri kümelerini birleştirerek, bireyleri konumlara atayan, konumsal mikro veri üretmeyi sağlayan bir yöntem olarak tanımlanabilir (Şekil 14). Konumsal mikro simülasyon, bu tanımı ile aşağı yukarı popülasyon sentezi ile eş anlamlı olmaktadır. Konumsal mikro simülasyon, konumsal mikro veriye dayalı olarak çok seviyeli olayları anlamaya yarayan bir yaklaşım olarak görülmektedir.



Şekil 14. Konumsal mikro veri ve daha yaygın bir şekilde kullanılan resmi veri tipleri arasındaki farklılıkları gösteren basitleştirilmiş bir akış şeması

Hem coğrafi olarak toplanmış verilerin, hem de coğrafi olmayan bireysel düzeyli verilerin belirgin dezavantajları vardır. Konumsal mikro simülasyon bu dezavantajların üstesinden gelmeye, araştırma uygulamaları için faydalı yeni çıktılar yaratmaya ve daha karmaşık modeller için de girdiler yaratmaya çalışmaktadır. Şekil 14'teki şematik diyagram geleneksel resmi veriler ile popülasyon sentezi sırasında üretilen konumsal mikro veriyi karşılaştırmaktadır. Şekil 14'ü özetlemek gerekirse; burada coğrafi olarak elde edilen veriler “kısıtlar”, bireysel düzeyde toplanan veriler ise “mikro veri” olarak adlandırılmaktadır. Kısıtlar ile mikro veriler birleştirilerek konumsal mikro veri elde edilmektedir.

Konumsal mikro simülasyon teriminin en genel anlamı, coğrafi alanlara atanan bireysel düzeyli verilerin oluşturulması, bu verilerin analizi ve modellemesidir.

Konumsal mikro simülasyon, ulaşım sistemleri ve konut piyasaları gibi coğrafi alan üzerinde meydana gelen karmaşık olguların analizi için çok uygun bir yaklaşım

olmaktadır. Çünkü bu olgular sentetik verilerin oluşumunu içerdiğinden, bu yöntem mevcut verilerin kısıtlı olduğu durumlar için uygun olmaktadır.

Konumsal mikro simülasyon, bireysel düzeyli verilerin avantajları ile coğrafi verilerin coğrafi avantajlarını birleştirmenin bir yolu olarak görülmektedir. Eğer doğru kullanılırsa konumsal mikro simülasyon yöntemi, bu verileri tek bir analizle birleştirerek her iki veriden de yarar sağlamak için kullanılabilir. Bu yüzden konumsal mikro simülasyonun, bireysel düzeyli ve coğrafi düzeyli verilerin olduğu ancak bu verileri birleştirmenin bir yolu olmadığı durumlarda kullanılması faydalı olmaktadır.

Konumsal mikro simülasyonun tipik kullanım örnekleri; demografideki değişikliklerin ve nüfus artışının konumsal dağılımının modellenmesi, yüksek düzeyli ulaşım modellenmesi, bireylerin davranışsal değişiminin senaryo temelli modellenmesi olarak verilir. Buna ek olarak konumsal mikro simülasyonun tipik bir kullanımı da, elde edilen konumsal mikro veri çıktısının bireysel düzeyli modellemelere girdi olarak kullanılmasıdır. Günümüzde büyük bireysel ve alan düzeyli veri kümelerinin kullanılabilir duruma geldiği ve binlerce kişilik konumlarda büyük ölçekli modeller oluşturmak için bilgisayarların yeterince güçlü olduğu düşünüldüğünde, konumsal mikro simülasyon hala emekleme döneminde bir alan olduğu görülmektedir. Konumsal mikro simülasyon yönteminin önemli faydalarının olabileceği ancak yöntemin henüz hiç kullanılmadığı birçok uygulama alanı da bulunmaktadır.

Konumsal mikro simülasyon terimi ile ilgili bazı tanımlar aşağıda verilmiştir:

1. Konumsal mikro simülasyonun; sadece bireyler, gruplar ve zaman periyodları arasında değil, aynı zamanda coğrafi alanlar üzerinde bir şeylerin nasıl değiştiği ile ilgisi vardır. Yani daha geniş alana ait mikro simülasyondan konumsal mikro simülasyonu ayıran şey budur.
2. Konumsal mikro simülasyon, mikro sözcüğünün ifade ettiği gibi bireysel düzeyli konuları araştırır.
3. Konumsal mikro simülasyon, simülasyon sözcüğünün ifade ettiği gibi modelleme amacıyla hayali verilerin oluşturulmasını içerir.

SimCity, oyuncunun kendi kentsel altyapısını oluşturduğu ve sanal vatandaşları üzerinde etki gösterdiği popüler bir bilgisayar oyunu serisidir. SimCity benzetmesi, konumsal mikro simülasyonu anlamaya yardımcı olmaktadır. Fakat uygulamada SimCity'in temel amaçları (eğlence, eğitim, kâr), konumsal mikro simülasyonunkinden oldukça farklıdır. Fakat bazı açılardan SimCity ile konumsal mikro simülasyonu

karşılaştırmak uygundur: SimCity coğrafi alanlara dağıtacağı sanal bireyler yaratır ve konumsal mikro simülasyonun yaptığı gibi model denemelerinden oluşan bir çerçeve sunar. SimCity, kentsel planlama öğretimi için (Gaber, 2007) ve kentsel sistemlerin karmaşık bilgisayar simülasyonlarının nasıl olduğunu göstermek için kullanılabilir. SimCity konseptinin birkaç açık kaynaklı versiyonu da artık mevcuttur (örneğin LinCity-NG, Micropolis ve Simutrans).

Konumsal mikro simülasyon terimi, genellikle popülasyon sentezini içeren ortalama bir süreci tanımlamak için kullanılmaktadır. Popülasyon sentezi, coğrafi bölgelere atanan bireysel düzeyli verileri üretmek için kullanılmaktadır. Popülasyon sentezi, konumsal mikro simülasyon yaklaşımının önemli bir bileşeni olarak bilinmektedir. Popülasyon sentezinin konumsal mikro simülasyondaki amacı, toplamsal düzeyli kısıtlara olabildiğince benzeyen her konum için gerçekçi veriler oluşturmaktır. Popülasyon sentezi, genellikle coğrafi ve bireysel düzeyli veriler arasında paylaşılan değişkenlere bağlı olarak anket veri setindeki bireylerin bölgelere yerleştirilmesi sürecini içerir. Toplamsal düzeyli verilerde mevcut olmayan ancak mikro veride mevcut olan hedef değişkenler olduğu durumlarda ise, popülasyon sentezi süreci toplamsal düzeyli veride mevcut olmayan ek hedef değişkene ait bilgiyi simüle etmek için kullanılabilir. Bu bağlamda; popülasyon sentezi küçük alan tahmininin uzun zamandır var olan alanının bir parçası olarak görülebilmektedir (Rao ve Molina, 2015).

Konumsal mikro simülasyon, genel anlamda mikro simülasyonun bir alt kümesi olarak görülmektedir. Konumsal mikro simülasyonu genel mikro simülasyon sürecinden ayıran belirleyici özelliği, onun ağırlıklı olarak coğrafi terimlerle tanımlanan çevresinden kaynaklanmaktadır. Konumsal mikro simülasyonda bireyler karakteristik özelliklerini ve davranışsal çıkarımlarını etkileyen küçük konum parçalarına yerleştirilmektedirler. Orcutt bu geniş bakış açısı ile, daha önce konumsal mikro simülasyon terimini kullanmamasına rağmen neden Orcutt'un sıklıkla bu alanın kurucularından biri olarak görüldüğünü açıklamaya yardımcı olmaktadır (Orcutt, 1957).

Pek çok yeni ve nadiren kullanılan kelimelerde olduğu gibi konumsal mikro simülasyon terimi hakkında karışıklıklar mevcuttur ve bu terimin anlamı çalışılan alana ve çalışan bireylere göre değişebilmektedir. Örneğin, bir ekonomist için konumsal mikro simülasyonun anlamı, farklı konumlardaki bireylerin fiyattaki veya politikadaki değişikliklere nasıl tepki verdiği gibi zamansal bir süreci modellemesi olabilmektedir. Bir ulaşım planlamacısı için bu terim ulaşım ağında araçların tam hareketini simüle etme

anlamına gelebilmektedir. Konu hakkında hiçbir şey bilmeyen biri için bu terim anlamsız bir söz gibi de gelebilmektedir. Dolayısıyla konumsal mikro simülasyonun ne olduğunun yanı sıra ne olmadığını da ele alınması gerekmektedir. Fakat her ne olursa olsun bu terim coğrafi bölgelere bir tür tahmin yöntemi ile gruplandırılan bireysel düzeyli verilerin oluşturulmasını içermektedir.

Konumsal mikro simülasyon terimi hakkında çok fazla karmaşa yaşanmaktadır. Bu kısmen konumsal mikro simülasyon ve diğer yöntemler arasında yakın ilişki olmasından kaynaklanmaktadır, kısmen de konumsal mikro simülasyonun insanlar tarafından olduğundan daha karmaşık olarak düşünülmesinden kaynaklanmaktadır. Konumsal mikro simülasyon ve onun ne olduğu hakkında akılcı tanımlar olmasının yanı sıra bu terimin ne olmadığını tanımlamak da faydalı olmaktadır.

- Konumsal mikro simülasyon küçük alan tahmini değildir:

Küçük alan tahmini, küçük bir alana ait toplam sayıları tahmin etmekten ibarettir. Örneğin bir küçük alanda gelecekteki bir yıl için bölgenin toplam nüfusu, bu yolla tahmin edebilmektedir. Fakat burada elde edilen tahminlerde, her bir birey hakkında bilgilere sahip olunamamaktadır, elde edilen tahminler alan istatistikleri ile kısıtlı olmaktadır. Öte yandan konumsal mikro simülasyon gerçekten mikro düzey odaklı çalışmaktadır. Bu yüzden konumsal mikro simülasyondan elde edilen tahminler birey düzeyli popülasyon tahmini olarak görülmektedir. Konumsal mikro simülasyon sayesinde küçük alan tahminine karşılık gelen toplam nüfus sayımı verileri ve kendi yerel yönetim teşkilatı olan şehir veya bölgeye ait (örneğin belediye) makro seviyeli bilgiler elde edebilmektedir.

- Konumsal mikro simülasyon (tamamen) birey tabanlı bir modelleme değildir:

Konumsal mikro simülasyon bireylerin oluşturulmasını ve analizini içermektedir ve bu bireylerin bölgelere ve hanelere yerleştirilmesini sağlamaktadır. Fakat bu, bireyler arasında mutlaka bir etkileşimin olduğu anlamına gelmemektedir. Bunun için birey tabanlı modelleme gerekli olmaktadır. Yöntem “simülasyon” kelimesini içerdiği için, bireylerin zamanla, mekanla, birbirleriyle ve çevreyle etkileşimde bulunduğu bireysel davranışların ayrıntılı modellemesini içerdiği varsayılmaktadır. Konumsal mikro simülasyon genellikle sonuçların geniş ölçüde tahmin edilebileceği, modellemesi daha üstten aşağıya bir yaklaşım olarak bilinmektedir. Buna karşılık birey tabanlı modelleme ise aşağıdan yukarıya doğru bir yaklaşım olarak bilinmektedir (Batty, 2005).

Konumsal mikro simülasyonun nerede bittiği ve birey tabanlı modellemenin nerede başladığı belirten açıkça tanımlanmış bir sınır bulunmamaktadır ve bu iki yaklaşımın

birbiri ile yakından bağlantılı olduğu bilinmektedir. Konumsal mikro simülasyonun parçası olarak üretilen sentetik popülasyonlar, birey tabanlı modelleme için mükemmel bir başlangıç noktası oluşturabilmektedir. Bu nedenle birey tabanlı modelleme konumsal mikro simülasyonun bir uzantısı olarak görülebilmektedir. Konumsal mikro simülasyon bireyleri üretip ve onların özelliklerine göre alanlara çeşitli “... olursa ne olur” senaryoları ile yerleştirirken, birey tabanlı modelleme bireylerin zamanla, mekanla, birbirleriyle ve çevreleriyle etkileşime girmelerine izin vererek konumsal ve zamansal çözümler sunmaktadır. Bu yüzden birey tabanlı modelleme, her birey için daha yüksek hesaplama gereksinimine sahip olmaktadır. Konumsal mikro simülasyon modelleri de milyonlarca insanı kapsadığı için büyük hesaplama gereksinimi duymaktadır.

Konumsal mikro simülasyon, ileri düzeyli bireysel tabanlı modellemenin bir alt kümesi olarak görülebilmektedir. Konumsal mikro simülasyon modellerinin sonuçları yalnızca zaman içindeki belirli anlık durumlara uygulanırken ve bireyler tek bir alana sabitlenmektedir. Bu yüzden birey tabanlı modelleme, konumsal mikro simülasyon modellerine daha karmaşık etkileşimler sağlayarak konumsal mikro simülasyon modellerine ek değerler katabilmektedir.

Özetle; birey tabanlı modelleme ve konumsal mikro simülasyon coğrafi alan ve zaman üzerine çalışan, bireysel düzeyli süreçlerin analizini yapan, birbiri ile ilişkili, örtüşen ve tamamlayıcı yaklaşımlar olarak tanımlanabilmektedirler. Ana kavramsal fark; konumsal mikro simülasyon çalışmasının hesaplama gücü ve model karmaşıklığına ilişkin nedenlerden dolayı konum ve zaman aralıklarının birey tabanlı modellemeye göre daha büyük olmasıdır.

- Konumsal mikro simülasyon gerçek veriden, yeni veri üretmez:

Konumsal mikro simülasyonla görünüşte yeni bireyler oluşturulmakta ve bireyler bölgelere yerleştirilmektedir. Dünyada yeni bilgilerin bir şekilde yaratıldığını düşünmek cazip olmaktadır. Ancak durum böyle değildir. Yeni bireyler farklı bir düzen içerisinde ve farklı kombinasyonlarda olsa da bu sentezlenen yeni bireyler, bireysel düzeyli verideki bireylerin basit bir tekrarıdır başka bir şey değildir. Bu yüzden veri kümesinin çeşitliliği arttırılmaz da, veri kümesinin toplamsal düzeyde özellikleri basit bir şekilde değiştirilmektedir. Konumsal mikro simülasyon, sürece dahil edilen tüm verileri göz önünde bulundurarak eksiksiz bir veri oluşturmaktadır. Konumsal mikro simülasyon, gerçek nüfusa ait benzetim yapabilmek için tüm bilgileri bir araya getirmenin yoludur

ancak yöntem, gerçek veriden yeni veri üretmemektedir. Başka bir deyişle elde edilen veri bilinen bireylerin sadece simülasyonudur.

- Konumsal mikro simülasyon çoğunlukla konumsal değildir:

Konumsal mikro simülasyonun en şaşırtıcı özelliği yöntemin tamamen konumsal olmamasıdır. Yöntemin bu adı kullanmasının tek sebebi, uygulayıcıların coğrafi alanları temsil eden idari bölgeleri “değişken” olarak kullanmasıdır. Ancak bir evdeki yaş grubu verileri ve yatak odası sayısı gibi veriler de konumsal mikro simülasyon için değişken olabilmektedir. Benzer şekilde coğrafi konum da değişken olarak kullanılabilir. Tabii ki, konumsal mikro simülasyondan elde edilen konumsal mikro veriler, haritalar ve analizler konumsal olmaktadır. Sadece konumsal mikro veriyi üretmek için kullanılan yöntem, konumsal bir özelliğe sahip değildir.

Tam olarak belirtmek gerekirse konumsal mikro simülasyonun doğasında konumsallık bulunmamaktadır. Bölgelerin coğrafi koordinatları, bireylerin hangi konuma atandığı, ev ve iş yerleri gibi konumsal nitelikler onlar üretildikten sonra konumsal mikro verilere kolayca eklenebilmektedir. Burada kritik olan değişken olarak coğrafi değişkenlerin kullanımınıdır ve konumsal mikro simülasyonu diğer mikro simülasyon türlerinden ayıran özelliğin bu olduğu bilinmektedir.

Konumsal mikro simülasyonun yaygın kullanımı basitçe, var olmayan verilerin model tahminlerini oluşturmaktır. Girilen mikro veri ve kısıtlar, simüle edilmiş kısıtların gerçeğe uygun olmasını sağlamaktadır. Ortaya çıkan sentetik konumsal mikro veri, yerel düzeyde eksik verileri tahmin etmek için son derece kullanışlı olmaktadır. Eğer çıktıda yer alan hedef değişkenler kısıtlarda mevcut değilse, bu değişkenlerin tahminleri konumsal mikro veriden elde edilen veriden çıkartılabilmektedir.

1.4.1. Konumsal Mikro Simülasyon Uygulamaları

Konumsal mikro simülasyon çok çeşitli uygulama alanlarına sahiptir. Yaklaşımın kullanıldığı birçok alan bulunmaktadır. Başlıca uygulama alanlarının ise sağlık, ekonomik politika değerlendirmesi ve ulaşım olduğu bilinmektedir.

Sağlık verilerinin konumsal mikro simülasyon ile modellenmesinin iki nedeni bulunmaktadır. Birincisi birçok kronik hastalık; biyolojik, fizyolojik, çevresel, sosyal ve ekonomik faktörler dahil olmak üzere çok yönlü etkileşimlerin sonucu ortaya çıkmaktadır.

Bu faktörlerin tümünün nasıl değiştiğini anlamak ilgilenilen hastalıklardaki değişimleri anlamayı kolaylaştırmaktadır. İkincisi hastalık verileri, faydalı politika analizleri yapabilmek için gerekli konumsal ölçeğe her zaman sahip olmamaktadır. Konumsal mikro simülasyonun sağlık uygulamalarına ait klasik bir örnek olarak İngiltere'nin Leeds şehrinde küçük alan düzeyinde sigara içme oranını tahmin eden bir çalışma verilmektedir (Tomintz vd., 2008). Bu çalışmada sigara içme konumsal mikro simülasyonda hedef değişken olmaktadır. Bu birçok bireysel düzeyli araştırmada belirtilmiştir ancak sigara içme oranlarının bir yerden diğer bir yere nasıl değiştiği hakkında şaşırtıcı derecede az bilgi bulunmaktadır. Sigara içme oranlarına dayanarak sigara bıraktırma ünitelerinin nereye konumlandırılacağını belirlemek zor olmaktadır. Bu nedenle, sentetik konumsal mikro veriler, bireylerin sigarayı bırakmaları için açılacak yeni kliniklerin nereye konumlanacağını belirlemek için kullanılabilir. Alternatif olarak konumsal mikro veriler bir tütün firması tarafından, yeni bir şube açılmak istendiğinde nereye yatırım yapacaklarına karar vermek için kullanılabilir. Araştırmacılar çalışmalarında sigarayı bırakma kliniklerinin optimal yerlerde bulunmadıklarını göstermişler ve ayrıca halk sağlığı kampanyalarının maliyet etkinliğini potansiyel olarak geliştiren yeni kliniklerin optimal konumlarını da belirtmişlerdir (Tomintz vd., 2008).

Simobesity sağlık alanında konumsal mikro simülasyonun kullanıldığı başka bir araştırma projesidir ve konumsal mikro simülasyon bu çalışmada demografik kısıt değişkenlerine bağlı olarak yerel düzeyde obezite yaygınlığını tahmin etmek için kullanılmıştır (Edwards ve Clarke, 2013). Bu çalışmadan elde edilen sonuçlara göre 'obezojenik' olarak etiketlenen araç bağımlı kentsel çevreler, hareketsiz yaşam biçimleri ve bunun sonucu olarak hastalıklar üzerine yeni kanıtlar ortaya çıkmıştır. Bu bağlamda bir konumsal mikro simülasyon modelinde sosyo-demografik ve konumsal coğrafi değişkenlerin birleştirilmesi için büyük bir potansiyel bulunmaktadır. Tomintz, Clarke ve Rigby tarafından tanımlanan kurallardan yararlanarak, böyle bir modelin çıktıları halk sağlığı girişimlerinin yararlarını en üst düzeye çıkarmakta ve fiziksel hareketsizliğin üstesinden gelmek için yerel müdahaleleri hedef almaya yardımcı olabilmektedir (Tomintz vd., 2008).

Ekonomik politika değerlendirmesinden kaynaklanan etkilerin sosyal-demografik dağılımı mikro simülasyonun en yaygın uygulamalarından (bu alanda yapılan analiz genellikle konumsal olmamasına rağmen) biri olarak görülmektedir. Politika değişikliğinin farklı gelir ve sosyo-demografik gruplar üzerindeki etkisinin araştırıldığı 'sosyal etki

değerlendirmesi' uygulanmış mikro simülasyon arařtırmalarının klasik bir örneđi olarak bilinmektedir. Sıklıkla bu simülasyonlar hükümet tarafından üstlenilmekte ve bu simülasyonların etkilerdeki konumsal deđişkenlikten ziyade popülasyondaki genel kaymalara odaklanmaktadır. Avrupa birliđi tarafından finanse edilen EUROMOD projesi ve aynı isimli yazılım paketi bu tip çalıřmaların en büyüđü olarak görölmektedir. EUROMOD yazılımı, birçok ölkede hükümet analistleri ve arařtırma ajansları tarafından politika reformlarının dađılımsal etkilerini tahmin etmek için kullanılmaktadır. Arařtırma sonuçları, mikro simülasyona dayalı bu modelleme çalıřmasının ulusal düzeyli politikaları ilgilendiren önemli yeni deliller sunabildiđini göstermiřtir (Avram vd., 2012).

Konumsal mikro simülasyon, EUROMOD ve diđer ekonomik mikro simülasyon modellerinin kullandıđı tekniklere çok benzeyen teknikler kullanmaktadır (Sutherland ve Figari, 2013). Ancak, ekonomik politika deđerlendirmesi için yapılan mikro simülasyon arařtırmalarının çođu konum üzerindeki etkileri incelememektedir.

Ulařım modellemesi, analizlerinin temelinde bireysel verileri kullanan bir alan olarak karřımıza çıkmaktadır. Ulařım modellemelerinden biri olan MATSim gibi büyük ölçekli modeller; bireysel özellikleri seyahat alışkanlıkları belirlemek, aktivite üretmek ve rota planlaması yapmak amaçlarıyla konumsal mikro verileri kullanmaktadır. Aynı amaçlarla konumsal mikro veri üreten konumsal mikro simülasyon modelleri, ulařım literatüründe olmasına rađmen bu süreç ulařım modellemesinde her zaman popülasyon sentezi olarak anılmaktadır (Müller ve Axhausen, 2010).

Genellikle ulařım modellemesinde sentetik popülasyon üretme sürecine daha az önem verilmektedir. Çünkü odak noktası bireylerin özelliklerinden ziyade bireylerin hareketi olmaktadır. Konumsal mikro simülasyonu, ulařım modelleriyle birleřtiren potansiyel çalıřmalar da mevcuttur. Ulařım modelleme çalıřmalarında konumsal mikro simülasyon potansiyelinin büyük ve geliřmiř bir örneđi Barthelemy ve Toint tarafından yapılan çalıřmada ele alınmıřtır (Barthelemy ve Toint, 2015). Bu çalıřmada, Belçika'daki ulařım sisteminin tamamını temsil eden on milyon kiři cođrafi olarak konumlandırılmıř ve sentetik nüfusa seyahat davranıřları da atanarak stokastik bir model oluřturulmuřtur. Barthelemy ve Toint bu çalıřmada konumsal mikro simülasyondan elde ettikleri sentetik mikro veriyi birey tabanlı modelleme yaklařımı ile birleřtirilerek, çok büyük bir ulařım sistemini ayrıntılı bir řekilde karakterize etmiřlerdir (Barthelemy ve Toint, 2015).

1.4.2. Konumsal Mikro Simülasyon Varsayımları

Herhangi bir simülasyon tekniğinde olduğu gibi konumsal mikro simülasyon da varsayımlara dayanmaktadır. Bu varsayımların bazılarının tüm durumlarda sağlanması pek mümkün olmamaktadır. Varsayımların geçerli olmadığı durumlar konumsal mikro simülasyon yapılmasına engel teşkil etmemektedir. Henri Poincaré'nin söylediği gibi "Bu gibi durumlarda hiç öngörü yapamamaktansa kesinlik olmaksızın olsa bile öngörü yapmak daha iyidir." (Barthelemy, 2014). Bununla birlikte, konumsal mikro simülasyonun kullanıcılarının ve araştırmacının tüketicilerinin; konumsal mikro simülasyonun sonuçlarının gerçek olmadığını ancak belli bir alandaki popülasyonun en iyi tahmini olduğunu anlamaları önemlidir. Varsayımların anlaşılmadığı durumlarda konumsal mikro simülasyon hatalı sonuçlar doğurabilmektedir. Bu nedenle konumsal mikro simülasyonu kullanan araştırmacıların görevi, sonuçların dayandığı varsayımları açık bir şekilde belirtmek ve bu varsayımların uygulamada ne dereceye kadar sağlanması gerektiğini ifade etmektir. Kabaca ifade etmek gerekirse tüm konumsal mikro simülasyon modellerinin altında yatan 4 ana varsayım bulunmaktadır:

1. Bireysel düzeyli mikro veri çalışma alanını temsil etmelidir.
2. Hedef değişken, kısıt değişkenlerine ve onların etkileşimlerine bir bakıma bağlı olmalıdır, konum ve zaman ise nispeten sabittir.
3. Kısıt değişkenleri arasındaki ilişkiler konumsal olarak bağımlı değildir.
4. Girdi mikro veri seti ve kısıtlar yeterince büyük ve çalışma bölgesinde yer almalı, bireylerin ve konumların çeşitliliğini arttırmak için yeterince ayrıntılı olmalıdır.

Açıkçası, gerçek dünya karmaşıktır ve birçok süreç konumsal olarak bağımlıdır. Bu da ikinci ve üçüncü varsayımları geçersiz kılmaktadır. Değişkenler arasındaki ilişkilerin konumsal olarak bağımsız olmasının ne ölçüde kabul edilebileceği çoğu kez bilinmemektedir. Fakat çoklu değişkenler arasındaki ilişkilerin konumsal bağımlılığını kontrol etmenin yolları mevcuttur. Burada yer alan varsayımların, girdi verilerine referansla konumsal mikro simülasyon araştırmasının başında ele alınması gerekmektedir.

1.4.3. Konumsal Mikro Veri

Konumsal mikro veri diye adlandırılan veri, konumsal mikro simülasyondan elde edilen faydalı çıktı olarak tanımlanmaktadır. Bu veri, girdi mikro verisinde olduğu gibi her bir satırın birey değişkenini içermesinin yanında bireyin coğrafi konumu ile alakalı ek bir değişken de içermektedir. İdeal bir konumsal mikro veri seti, konumsal olmayan bireysel düzeyli girdi popülasyonundaki mevcut bilgiyle sahip olunan her konuma ait toplamsal kısıtları temsil eden bireylerin simülasyon verisinden oluşmaktadır.

1.5. Popülasyon Sentezlemede Alternatif Yaklaşımlar

Konumsal mikro veri üretiminde AOA ve AOG yöntemleri kullanılabilir yöntemlerden sadece ikisidir. Araştırmacılar modelleri için rahat kullanabilecekleri ve modellerine en iyi uyan bir yöntem seçme eğilimindedirler. İdeal yöntemin oluşturulması genellikle zaman alıcı olmaktadır. Çoğu araştırmacı ayrıntılarla uğraşmak yerine yöntemleri doğrudan gerçek dünya verilerine uygulamaya odaklanmaktadır. Çok büyük veri kümeleri olması durumunda çok büyük veri kümeleri ile çalışmak için optimize edilen belirli yöntemler kullanılmadığında, konumsal mikro simülasyon sonuçlarını elde etmek mümkün olmayabilmektedir. Her şeyden önce hangi yöntemlerin hangi uygulamalar için en iyisi olduğu hakkında herhangi bir genel fikir birliği bulunmamaktadır. Bu yüzden her bir uygulamada hangi yöntemin en uygun olduğunu belirlemek için yöntemler denenmektedir.

AOA yöntemine ilginç bir alternatif GREGWT algoritmasıdır. GREGWT geliştirilmiş regresyon ağırlıklandırma işlemine dayanan bir yöntem olarak bilinmektedir. İlk olarak Avustralya İstatistik Bürosu'nun İstatistik Birimi tarafından SAS dilinde uygulanmıştır. Algoritma bir dizi başlangıç ağırlığını geliştirilmiş bir regresyon ağırlıklandırma yöntemi yeniden ağırlıklandırmaktadır. Ortaya çıkan ağırlıklar bir araya getirildiğinde her konum için seçilen bireyler kısıt değişkenlerine uyarlanmaktadır. AOA gibi, GREGWT nin ağırlıkları da tamsayı değildir. Bu, GREGWT algoritması yardımıyla bireysel düzeyli mikro veri elde etmek için, bir çeşit tamsayılaştırma algoritmalarına ihtiyaç duyulacağı anlamına gelmektedir. Bu algoritmanın basitleştirilmiş bir versiyonu da

Rahman tarafından ele alınmıştır (Rahman, 2009). Algoritmanın ayrıntılı bir şekilde incelendiği çalışma ise Tanton vd. tarafından yapılan çalışmadır (Tanton vd., 2011).

Genel olarak optimizasyon problemi, amaç fonksiyonu olarak adlandırılan bir fonksiyondan oluşmaktadır. Bu fonksiyonun sonucu maksimum ya da minimum olmalıdır. Bu fonksiyon olası girdilerin tüm alanı için tanımlanmayabilir. Bu fonksiyonun tanımlandığı alan çözüm uzayı olarak adlandırılmaktadır. Ayrıca optimizasyon problemleri fonksiyonların argümanlarının alabileceği değerler üzerindeki sınırlamalarla kısıtsız ya da kısıtlı olabilmektedir (Boyd ve Vandenberghe, 2004). Kısıtlar varsa, çözüm uzayı yalnızca amaç fonksiyonunun görüntüsünün bir parçasını içermektedir. Amaç fonksiyonu ve kısıtların her ikisi de çözüm uzayını tanımlamak için gerekli olmaktadır. Bu çerçevede, popülasyon sentezi kısıtlı bir optimizasyon problemi olarak görülebilmektedir. x , n boyutlu bir vektör ve bu vektörün değerleri (x_1, x_2, \dots, x_n) olsun. Bu durumda amaç fonksiyonunun değeri $f_0(x)$ x 'e bağlı olmaktadır. x 'in olası değerleri m uzunluğunda bir parametre vektörü sayesinde $(par_1, par_2, \dots, par_m)$ tanımlanmaktadır. Burada m kısıtların sayısıdır. Bu tür bir problem şu şekilde ifade edilebilmektedir:

$$\begin{cases} \min f_0(x_1, x_2, \dots, x_n) \\ f_i(x) \geq par_i, i=1, \dots, m \end{cases} \quad (9)$$

Bunu popülasyon sentezi problemine uygulamak için, par_i parametreleri 0'ı ve $f_i(x)=x$ 'i temsil eder, çünkü tüm hücrelerin pozitif olması gerekmektedir. Minimize edilecek olan $f_0(x)$ toplam kısıt değişkeni ve gerçek ağırlık matrisi arasındaki uzaklıktır. x , $f_0(x)$ minimize etmek için hesaplanan ağırlıkları temsil etmektedir. Konumsal mikro simülasyon durumunda tüm ağırlıkların pozitif veya sıfır olmalısı gibi nispeten daha basit kısıtlar bulunmaktadır.

$$\{a_{ij} \in \mathbb{R}^+ \cup \{0\} \forall i,j\} \quad (10)$$

Konumsal mikro simülasyonu bir optimizasyon problemi gibi görmek, kısıtlı optimizasyon tekniklerini kullanarak çözümlerin bulunmasına olanak sağlamaktadır. Bu yöntem yeniden ağırlıklandırma işlemini gerçekleştirilmede herhangi bir optimizasyon algoritmasını kullanmaktadır.

Temel R fonksiyonu “optim”, amaç fonksiyonlarını sayısal olarak çözmek için genel amaçlı bir optimizasyon çerçevesi sağlamaktadır. Konumsal mikro simülasyonun amaç fonksiyonuna bağlı olarak bireysel düzeyli veri kümesini yeniden ağırlıklandırmak için herhangi bir genel optimizasyon algoritması kullanılabilir. Farklı ağırlıklandırma yöntemleri farklı durumlarda için uygun olabilmektedir ve bu yöntemler içerisinde herhangi bir üstün yöntem bulunmamaktadır. Ancak konumsal mikro simülasyon için hangi yöntemin daha uygun olduğunu test etmek, bazı algoritmaların diğerlerinden daha verimli olduğunu açıkça ortaya koymaktadır.

Kombinatorial optimizasyon, optimizasyon problemlerini ayrı bir seçenek yelpazesi ile çözmek için kullanılmaktadır. Kombinatorial optimizasyon ile konumsal mikro simülasyonu ele alan bir çalışma, Clarke ve Williamson tarafından su talebini tahmin etmek için geliştirilen modelde kullanılmıştır (Clarke vd., 1997; Williamson vd., 1998). Bu modelde Williamson tarafından “kombinatorial optimizasyon” adı verilen bir yöntem geliştirilmiş ve kullanılmıştır. Her bölgedeki bireylere ağırlık atamak yerine, kombinatorial optimizasyon bölgeyi doldurmak için uygun aday listesini belirtmekte ve daha sonra yeni bireyleri değiş tokuş etmektedir. Minimize edilecek amaç fonksiyonu, değiş tokuşun yararlı olup olmadığına karar vermek için kullanılmaktadır. Birçok mevcut kombinatorial optimizasyon yöntemi bulunmaktadır. Bunlardan hangisinin uygun olduğu, aday kombinasyonunun nasıl seçildiğine ve değerlendirmeden sonra ne olacağının nasıl belirleneceğine bağlı olarak değişmektedir. Kombinatorial optimizasyon bireyleri bölgelere atamak için kullanılan AOA yöntemine iyi bir alternatif olarak görülmektedir. Bu yöntem olasılıksaldır ve sonuçlar tamsayı ağırlıklarıyla sonuçlanmaktadır. Kombinatorial optimizasyon girdi mikro veri setlerinin çok büyük olduğu uygulamalar için de daha uygun olabilmektedir. Anket veri kümesinin boyutu arttıkça deterministik AOA algoritmasını kullanmanın hız avantajları azalmaktadır. AOA tamsayı olmayan ağırlıklar üretir ve bu yüzden bu ağırlıkları nihai bireysel düzeyli popülasyona dönüştürmek için tamsayılaştırma çözümleri kullanılmaktadır. Bu nedenle de, AOA'nın her iterasyonunda tüm ağırlık matrisinin hesaplanması gerekmektedir. Kombinatorial optimizasyon ise her iterasyonda sadece aday birey önerileri sunmaktadır. Eğer amaç fonksiyonunun hesaplanması uzun zaman alıyorsa, kombinatorial optimizasyon hesaplama açısından yorucu olabilmektedir. Çünkü bir ya da birkaç birey değiş tokuş işleminden sonra önerilen her yeni popülasyon sürecinde uyum iyiliği yeniden değerlendirilmektedir.

Genetik algoritmalar da alternatif yöntemlere dahildir ve örneğin endüstri gibi bazı alanlarda popüler olarak kullanılmaktadır. Bu tür bir algoritma, amaç fonksiyonu birkaç yerel minimuma sahip olduğunda ve genel olanı bulunmak istendiği zaman çok etkili olabilmektedir (Hermes ve Poulsen, 2012). Konumsal mikro simülasyon için tasarlanmış kombinatoriyal optimizasyon şimdilik R dışında programlarda uygulanabilmektedir. Harland'ın Java tabanlı bir esnek modelleme çerçevesi kullanarak konuyu ele alan pratik bir uygulama çalışması bulunmaktadır (Harland, 2013).

“simPop” kütüphanesi, sentetik mikro veri üretmek ve bu verileri modellemek için alternatif yöntemler sunmaktadır. Kütüphanenin kullanışlı bir özelliği örnek veri ekinin olmasıdır. 'AB-SILC' veri tabanından (AB gelirleri ve yaşam koşulları istatistikleri) ve gelişmekte olan dünyadaki veri setlerinden veriler bu alternatif yöntemleri test etmek için eklenmiştir. Bu kütüphanesi popülasyon sentezi yapmak için modelleme tabanlı yöntemler, ayarlama ve kombinatoriyal optimizasyon algoritmaları da dahil olmak üzere çeşitli alternatif yöntemler sunmaktadır (Meindl vd., 2015).

“Kentsel veri bilimi araçları” bir popülasyon sentezi bileşenini içeren şehirleri modellemek için diğer yaklaşımlara göre daha da iddialı bir yaklaşım sunmaktadır. Bu yaklaşım sadece popülasyon sentezi değil, görselleştirme ve modelleme de sağlamaktadır. Python'da yazılmıştır (The Urban Data Science Toolkit -UDST).

AOA ve AOG yöntemlerinin yaygın uygulamaları göz önüne alındığında, daha karmaşık yapıları popülasyonları sentezlemek için AOA ve AOG yöntemlerine dayalı olarak yöntemlerin çeşitli uzantıları ve mutasyonları da geliştirilmiştir. Heterojenlik eksikliği ve mikro verilerin kısıtları ile başa çıkmak için araştırmacılar istatistiksel öğrenme metotlarından yararlanarak yeni yöntemler geliştirmişlerdir. Bu yöntemlerin temel prensibi, olası koşullu modellere dayanarak değişkenleri sıralı olarak güncellemektir.

- Bayes Ağları Yaklaşımı:

Bayes ağları bir grafik model kullanarak bir dizi değişken arasındaki olasılıksal ilişkileri (nedensellik veya bağımlılık) kodlamaktadır. Bayes ağlarının grafiksel gösteriminin sağladığı yüksek etkinlik ve avantajlar göz önüne alındığında, bu veri odaklı yaklaşım sınırlı bir mikro veriye sahip popülasyon sisteminin temel yapısını belirleyebilmektedir. Bu anlamda Bayes ağları modelleri popülasyon sistemlerinin yapısını öğrenmek için güçlü araçlar olarak kullanılmaktadır. Bayes ağları yaklaşımı özellikle mevcut mikro veriler kısıtlı iken ilgilenilen özneliklerin sayısı çok fazla olduğu durumda kullanılmaktadır.

Bir dizi rasgele deęişken arasında nedensellięi ve baęımlılıęı belirlemek için alternatif bir modelleme paradigması olarak Bayes aęları, karmaşık ilişkileri basit bir grafiksel model halinde özetlemek için veri odaklı bir çerçeveye sağlamaktadır. Bayes aęları karmaşık baęımlılık yapılarını kısa ve öz bir yapıya dönüştürmektedir (Pearl, 2000; Koller ve Friedman, 2009)

Bayes aęları modeli olasılıksal çıkarımda ve mantık problemlerinde yoğun şekilde kullanılmaktadır. Özel bir durum olarak, anket verilerinden elde edilen bilgileri analiz etmek ve yorumlamak için de uygulanmaktadır. Örneęin Sebastiani ve Ramoni Britanya'daki Genel Hanehalkı Anketleri verilerini analiz etmek ve yorumlamak için Bayes aęları modelini kullanmışlardır (Sebastiani ve Ramoni, 2001).

Ulaşım araştırmalarında Bayes aęları modeli yaklaşımı kullanan birçok çalışma mevcuttur. Bunlardan bazıları Janssens vd. (Janssens vd., 2006) tarafından yapılan birey tabanlı aktivite simülasyonu ve tahmini, Zhang ve Taylor (Zhang ve Taylor, 2006) tarafından yapılan kaza-olay modellemesi ve Castillo vd. (Castillo vd., 2008) tarafından yapılan trafik akış tahmini çalışmalarıdır.

Bayes aęları yöntemi yapay gözlemler üretmek için temel dağılımın olası bir simülasyonu olarak ta kullanılabilir. Trafik simülasyonu seyahat davranış analizi konusunda Bayes aęları yaklaşımı ilk kez Xie ve Waller tarafından uygulanmıştır (Xie ve Waller, 2010). Xie ve Waller San Francisco'daki hanehalkı seyahat anketlerini bireylerin davranış modellemesini ölçmek için Bayes aęları modelini kullanmışlardır.

Bayes aęları, bir dizi deęişkenler arasındaki olasılıksal ilişkileri etkili bir şekilde kodlayan birleşik olasılık dağılımının grafiksel bir temsilidir. Daha önce geliştirilen olasılıksal yaklaşımlara benzer şekilde Lijun Sun ve Alexander Erath popülasyon sentezi problemini birleşik olasılık dağılımının bir çıkarımı olarak düşünmüşlerdir (Sun ve Erath, 2015). Bu anlamda Sun ve Erath, Bayes aęları modelinin popülasyon sisteminin yapısının özünü temsil eden ve simülasyon yapmaya olanak sağlayan ve aynı zamanda gizlilięi koruyan etkili bir araç olduğunu öne sürmüşlerdir. Sun ve Erath çalışmalarında önerdikleri Bayes aęları yaklaşımının birleşik dağılımın altında yatan özellikleri karakterize etmede güçlü olduğunu ve verilerin aşırı uyum gösterme problemini olabildiğince azalttığını göstermişlerdir (Sun ve Erath, 2015).

Sun ve Erath mikro yapıda karmaşık olasılık dağılımını kodlamak için etkili bir araç olarak ve veriden bireyler hakkında bilgi almanın basit bir yolu olarak Bayes aęlarını kullanmayı tercih etmişlerdir.

Nedensel ilişkileri ve olasılıksal işlemleri birleştiren bir model olarak Bayes ağları bileşik dağılımın tahminini basitleştiren alternatif bir yöntem olmuştur. Bu yöntem, popülasyon sistemlerindeki farklı öznelilikler arasındaki önemli ilişkileri tanımlayarak birleşik dağılımı etkili bir biçimde simüle etmekte ve yeniden üretmektedir.

Özetle Bayes ağları ilgilenilen değişkenlerin olasılık dağılımlarını etkili bir şekilde kodlayan bir grafik yöntemidir (Heckerman, 1998; Pearl, 2000; Koller ve Friedman, 2009).

- Monte Carlo Teknikleri Yaklaşımı:

Markov zinciri Monte Carlo yöntemleri bilgisayar tabanlı simülasyon teknikleridir. Bu teknikler çok karmaşık stokastik modellerden/süreçlerden bağımlı rasgele olaylar dizisini simüle etmek için kullanılabilir (Hastings, 1970). Bu yöntemler birçok uzaysal ölçekte farklı veri kaynaklarını kullanımını açısından esneklik sağlamaktadır. Önsel bilgi sistematik bir şekilde elde edilmektedir. Varsayımlar tutarlı bir şekilde uygulanmaktadır. Yöntemler hesaplamalıdır ve bellek açısından güçlüdür. Markov zinciri Monte Carlo yöntemleri fizik, görüntü işleme vb. olmak üzere birçok alanda yoğun şekilde kullanılmaktadır.

AOA ve AOG yöntemlerinin başlıca eksiklikleri:

- a. Mevcut veriye uyan başka çözümler varken tek bir olumsuzluk tablosuna ait ayarlama yapılması,
- b. Heterojenliği kaybetme ve doğru bir popülasyon sentezi yerine mikro verilerin sadece kopyalanması,
- c. Veri üretimi sırasında ağırlıkları belirlerken verilerin doğruluğuna aşırı güvenilmesi,
- d. Sentezlenen bireylerin özelliklerinin sayısındaki artışa göre yetersiz ölçeklenebilirliğin olmasıdır.

Geleneksel popülasyon sentezleme yöntemlerin yerine, gerçek popülasyon verilerinden sentetik popülasyon sentezlemek için Markov zinciri Monte Carlo yöntemleri de kullanılabilir. Markov zinciri Monte Carlo yöntemleri kullanılarak yukarıda bahsedilen sorunların üstesinden gelinebilmektedir. Farooq vd. tahmin yapmak için farklı seçim modellerini kullanmayı önermişlerdir ve veri üretim modeli olarak Markov zinciri Monte Carlo algoritmasını uygulamışlardır (Farooq vd., 2013).

- Koşullu Monte Carlo:

Normal Monte Carlo sentezi kısıtlı/az veri kullanılarak kolaylıkla uygulanabilmektedir ancak koşullu Monte Carlo ile sentez ise sadece biraz daha karmaşıktır. X ve Z özelliklerinin bilindiği ve Y özelliğinin ise P(X, Y, Z) bileşik olasılık dağılımını kullanarak sentezlenmesi gerektiği varsayalım. Koşullu olasılık için bu formül:

$$P(Y|X,Z) = \frac{P(X,Y,Z)}{P(X,Z)} \quad (11)$$

P(Y|X,Z) den ilgilenilen olayı bulmak için, P(X, Y, Z)' nin katkıda bulunacak hücrelerinin etkili bir şekilde bulunması mümkün olmalıdır. Bu liste tabanlı veri yapısı kullanıldığında otomatik değildir. Çünkü (i,j,k) hücresi ile ilgili satırlara rasgele erişim etkin değildir. Bu sorunla başa çıkabilmek için liste verilen özelliklere göre sıralanmaktadır. Bu yöntem ilgili satırlara ulaşmayı O (log n) asimptotik performansı ile kolaylaştırmış olur. Pritchard ve Miller ortalama performans O (Nlog n) ve performansı arttırmak için popülasyon sentezi çalışmalarını C dilinde uygulamışlardır (Pritchard ve Miller, 2012). Pritchard ve Miller AOA, Monte Carlo, Koşullu Monte Carlo yöntemleri ile popülasyon sentezleme sürecinin adımlarını şu şekilde tanımlamışlardır:

1. (a) PUMS verisi ve Özet İstatistik Dosyaları (Nüfus verileri vb.) kullanarak hanehalklarına yönelik ayarlamaların AOA tekniği kullanılarak yapılması.
(b) PUMS verisi ve Özet İstatistik Dosyaları (Nüfus verileri vb.) kullanılarak bireylere yönelik ayarlamaların AOA tekniği kullanılarak yapılması.
2. PUMS verisi ve Özet İstatistik Dosyaları (Nüfus verileri vb.) kullanılarak ailelere yönelik ayarlamaların AOA tekniği kullanılarak yapılması. (Ek olarak burada hanehalkı ve bireylerin paylaştığı özniteliklerin dağılımının ayarlanması yapılır.)
3. Hanehalkının bir listesini sentezlemek için Monte Carlo yönteminin kullanılması. (1. Adımda (a) sentezlenen hanehalkı bilgisi kullanılır.)
4. Bir ya da daha çok ailelik her hanehalkı için, hanehalkı özelliklerinden yararlanarak aile/aileleri sentezlemek. Burada Koşullu Monte Carlo kullanılır. (2. Adımda sentezlenen aile bilgisi ve 3. Adımda MC ile sentezlenen hanehalkı bilgisi biliniyor).

5. (a) Her aile için, aile özelliklerinden yararlanarak bireyleri sentezlemek. Burada Koşullu Monte Carlo kullanılır (4.adımda Koşullu MC ile sentezlenen aile bilgisi ve 1. Adımda (b) sentezlenen birey bilgisi kullanılır).

(b) Her hanehalkı için, hanehalkı özelliklerinden yararlanarak ailesi olmayan bireyleri sentezlemek. Burada Koşullu Monte Carlo kullanılır (3. Adımda MC ile sentezlenen hanehalkı bilgisi ve 1. Adımda (b) sentezlenen birey bilgisi kullanılır).

(c) Yabancı / geçici sakinler (hanehalkı ile ilgisi olmayan ve kurumsal olmayan bireyler) sentezlemek. 1. Adımda (b) sentezlenen birey bilgisi MC kullanılarak sentezleme yapılır.

AOA'nın girdileri PUMS'ın liste tabanlı verilerini, marjinal kısıtları, q gibi bir sonlandırma hata terimini ve yineleme sınırını içermektedir. Marjinal kısıtlar tam çok yönlü olasılık tabloları ya da PUMS'taki sütunlardır. Marjinal kısıtlar AOA'nın geleneksel yaklaşımında kullanılmaktadırlar. Pritchard ve Miller AOA sürecinin bir kısmında, performans sebeplerinden dolayı C dilini kullanmışlardır. Pritchard ve Miller popülasyon sentezi yaparken AOA sürecinin geri kalan kısmı için R dilinin yeterli performansı sağladığını öne sürmüşlerdir. Pritchard ve Miller çalışmalarında ayrıca yuvarlama ve küçültme işlemleri kullanmışlardır. Marjinallerin tamamını toplamsal tablolarda yuvarlama hatalarını azaltmak için kullanmışlardır (Pritchard ve Miller, 2012).

- En Yakın k-Komşu Algoritması Yaklaşımı:

Sınıflandırma yöntemlerinden birisi olan En Yakın k-Komşu Algoritması geniş bir uygulama alanı bulmuştur. Bu yöntem, sınıfları belli olan bir örnek kümesindeki gözlem değerlerinden yararlanarak, örneğe katılacak yeni bir gözlemin hangi sınıfa ait olduğunu belirlemek amacıyla kullanılmaktadır.

Popülasyon sentezi aşamasında anket verisi nüfus sayımından elde edilen verilere eşleştirilirken, ana amaç anket verisinde g bölgesindeki Y_{ig} özelliklerine sahip j . hanehalkının X_{ijg} özelliklerine sahip i . birey özellikli bireyleri için en iyi eşleşmeyi bulmaktır. Bu eşleşme işlemi biter bitmez daha sonraki aşama örnek veri kullanılarak hanehalkı düzeyindeki ekstra özellikleri bulmaktır. Bu özellikleri bulmak için en yakın k-komşu algoritması yaklaşımı Namazi-Rad vd. tarafından yılında kullanılmıştır (Namazi-Rad vd., 2017). Bu çalışmada nüfus verileri ile g bölgesinin hanehalkı ve birey düzeyli anket verilerini eşleştirmek için en yakın k-komşu yaklaşımının kullanıldığı algoritmaya yer verilmiştir. Bu algoritmanın ana amacı hanehalkı ve birey özellikleri göz önüne

alındığında anketteki en benzer hanehalkını bulmaktır. Bu algoritmada, popülasyonun her bir hanehalkı için (örneğin Y_{jg}) örnek verideki en benzer hanehalkları çoğunluk oy sınıflandırıcıları kullanılarak bulunmuştur. Daha sonra anket verilerindeki hanehalkları içerisinde en yakın k-komşu algoritmasında bulunan uygun çözümlerden bireyler için arama yapılmıştır. Ve daha sonra hedef popülasyon hanehalkında (örneğin Y_{jg}) yer alan bireylerin özellikleri karşılaştırılmıştır (Namazi-Rad vd., 2017).

- Saklı Markov Modeli:

Popülasyon sentezi teknikleri (a) ayarlama yöntemleri (AOA ve AOG), (b) kombinatoriyal optimizasyon yöntemleri olmak üzere iki ana aileye ayrılmaktadır. Son birkaç yılda, popülasyon sentezi teknikleri ailesine üçüncü bir aile olarak görülen Markov işlem tabanlı yöntemler eklenmiştir. Bu yöntemlere örnek olarak Markov zinciri Monte Carlo simülasyonu verilebilir. Saadi vd. bu mevcut yöntemlerden daha iyi bir alternatif olarak kullanılabilen Saklı Markov Modeli tabanlı yaklaşıma dayalı genişletilmiş bir yöntem kullanmışlardır (Saadi vd., 2016). Bu yaklaşım veri hazırlama ve model eğitimi bakımından büyük bir esneklik ve verimlilik ile nitelenmektedir. Saklı Markov Modeli bir majinal dağılım ve sınırsız sayıda mikro örnekte verilen bir popülasyonun yapısal şeklini yeniden elde edebilmektedir. Tahmin edilmek istenen popülasyonun sadece bir marjinal dağılımı tüm popülasyonun sentezini yapabilmek için bir sınır şartı olarak kullanılabilir. Saadi ve ark. yaptıkları çalışmada Saklı Markov Modeli yaklaşımının standart ayarlama yaklaşımları olan AOA ve AOG'nin aksine mikro verinin tüm heterojenliğini elde etmemize olanak sağladığını göstermişlerdir. Bu yöntemin kabul edilebilir bir hata oranı ile marjinal dağılımlar ve onlara karşılık gelen çok değişkenli ortak dağılımları bularak doğru sonuçları verdiğini göstermişlerdir (Saadi vd., 2016). Dahası Saklı Markov Modeli küçük örnek büyüklükleri için AOA yönteminden daha iyi performans göstermektedir. Saklı Markov Modeli daha gerçekçi popülasyon tahminleri bulmak için birden fazla veri kaynağından sağlanan bilgileri birleştirebilmektedir.

Metodolojinin genelleştirilebilirliği ve verimliliği göz önüne alındığında, Saklı Markov Modeli yaklaşımı birey tabanlı mikro simülasyon modellemesinde önemli bir etki yaratmıştır. Yöntem, popülasyon sentezleme işlemi yapmak için birden fazla veri kaynağının (örneğin, mikro-örnekler ve toplamsal marjinal dağılımlar) birleştirilmesine olanak sağlamaktadır. Saklı Markov Modeli yaklaşımı mikro veride var olan heterojenliği elde etmede daha başarılı olmuştur. Bu yaklaşım hem kesikli hem de sürekli değişkenlere uygulanabilmektedir. Sınırsız sayıdaki mikro örnekler Saklı Markov Modeli yaklaşımı

kullanarak birleştirilebilmektedirler. Saklı Markov Modeli yaklaşımı AOA ile karşılaştırıldığında AOA yönteminden daha az veri kullandığı ve daha iyi sonuçlar sağladığı gösterilmiştir (Saadi vd., 2016). Saklı Markov Modeli altta yatan iç durumların gözlemciden gizlenmesi gereken bir Markov sürecidir. Hipotezler sistemdeki durumların sayısı ile ilgilidir ve durum geçiş olasılıklarının bilindiği varsayılmaktadır. Bu yüzden Markov zincirinin her durumu her durumun ortaya çıkma olasılığını ifade eden çıkış sembolü ve başka bir duruma geçişi sağlayan olasılıklara karşılık gelen geçiş olasılıkları olmak üzere iki parametre ile karakterize edilir (Ibe, 2013).

- Yapay Sinir Ağları (YSA):

Popülasyon sentezlemede son yıllarda kullanılan yeni yaklaşımlardan birisi de “Yapay Sinir Ağları” yaklaşımıdır. Yapay sinir ağları insan beyninin çalışma sisteminin yapay olarak benzetimi çabalarının bir sonucu olarak ortaya çıkmıştır. En genel anlamda bir yapay sinir ağı insan beynindeki birçok nöronun ya da yapay olarak basit işlemcilerin birbirlerine değişik etki seviyeleri ile bağlanması sonucu oluşan karmaşık bir sistem olarak düşünülebilir. Önceleri temel tıp birimlerinde insan beynindeki nöronların matematiksel modelleme çabaları ile başlayan çalışmalar, geçtiğimiz on sene içerisinde, disipline bir şekil almıştır. Yapay sinir ağları bugün fizik, matematik, elektrik ve bilgisayar mühendisliği gibi çok farklı bilim dallarında araştırma konusu haline gelmiştir. Yapay sinir ağlarının pratik kullanımı genelde, çok farklı yapıda ve formlarda bulunabilen enformasyon verilerini hızlı bir şekilde tanımlama ve algılama üzerinedir. Aslında mühendislik uygulamalarında yapay sinir ağlarının geniş çaplı kullanımının en önemli nedeni, klasik tekniklerle çözümü zor problemler için etkin bir alternatif oluşturmasıdır. Çünkü bilgisayarlar insanın beyinsel yeteneğinin en zayıf olduğu çarpma, bölme gibi matematiksel ve algoritmik hesaplama işlemlerinde hız ve doğruluk açısından yüzlerce kat başarılı olmalarına rağmen insan beyninin öğrenme ve tanıma gibi işlevlerini hala yeteri kadar gerçekleştirilememektedir.

Mohammadian vd. 2010 yılında yaptıkları çalışmada Ulusal Hanehalkı Seyahat Anketleri (UHSA) kümeleri ve bu kümelere karşılık gelen diğer bölgenin kümeleri arasında seyahat özelliklerinin aktarıldığı için, uygulama bölgesinin sentetik popülasyonunun da kümelenmesi gerektiğini vurgulamışlardır (Mohammadian vd., 2010). Bu amaçla yaptıkları çalışmada Yapay Sinir Ağları (YSA) kullanmışlardır. Yapay Sinir Ağları modeli burada UHSA’daki aynı küme şemasını çoğaltmak için kullanılmıştır. Bu sebeple UHSA hanehalkı üç alt gruba (60%, 30% ve 10%) ayrılmıştır. Bu alt gruplar sırası

ile YSA modelini eğitmede, doğrulamada ve test etmede kullanılmıştır. Elde edilen model 2 gizli katmana sahiptir ve model, test alt kümesinde %90'dan daha fazla doğruluk üretebilmektedir. Mohammadian vd. yaptıkları çalışmada simülasyon sonuçlarının iyileştirilmesi için güncelleme yöntemi önermişlerdir. Burada Bayes teoremine dayanan Bayes güncellemesi yöntemi kullanılmışlardır. Transfer modeli gözden kaçan yerel özelliklerini hesaba katarak büyük ölçüde geliştirilebilmiştir. Bu yöntem yerel örnek verisi var olan her yerde kullanılabilir. Mohammadian vd. yaptıkları çalışmada UHSA'dan seçtikleri rasgele bir örneği yerel örnek olarak kullanmışlardır. Popülasyon sentezi modeli; uygulama bölgesinin demografik, sosyo-ekonomik ve arazi kullanımı özelliklerinin hepsini içeren tüm popülasyon verisini üretmek için kullanılmaktadır. Daha sonra transfer modeli elde edilen bu hanehalklarına uygulanmaktadır. Uygulama bölgesinin sentetik hanehalkı seyahat anketi verisini üretmek için Monte Carlo simülasyonundan yararlanılmaktadır. Monte Carlo simülasyonu veri kaynağı olarak seyahat özelliklerinin güncellenmiş halini kullanılmaktadır (Mohammadian vd., 2010).

1.6. Popülasyon Sentezleyiciler

Bireylerin seyahat aktivitelerinin zaman ve yer olarak simüle edildiği trafik talep modellemelerinde mikro simülasyon yaklaşımının ortaya çıkması, sentetik popülasyon sentezleyicilerinin gelişmesini sağlamıştır (Ye vd., 2009).

Popülasyon sentezleme amacıyla bugüne kadar birçok yaklaşım kullanılmıştır. Her sentezleyicinin farklı veri giriş yöntemi vardır ve her bir sentezleyicinin doğruluğu için farklı uyum iyiliği testleri kullanılmaktadır (Müller ve Axhausen, 2010). Tablo 1'de bazı popülasyon sentezleyicileri ve genel özellikleri yer almaktadır.

Tablo 1. Bazı popülasyon sentezleyiciler ve özellikleri

Popülasyon Sentezleyici	Araştırmacılar	Sentezleyicinin Kullanıldığı Alan	Sentezleyicinin Özelliği
PopSynWin	Auld ve ark. (2008), Auld ve ark. (2010).	Chicago, Illinois.	Otomatik kategorizasyon ve kişi düzeyindeki kısıtlamaları kontrol etmek için hanehalkı seçim olasılıklarının ayarlanması üzerinde durulmaktadır.
ILUTE	Salvini ve Miller (2005), Pritchard (2008), Miller (2009).	Toronto, Canada.	AOA sırasında büyük alanlarda çalışmak için yeni bir yaklaşım kullanılmaktadır. Olasılık tablosu, özniteliklerin kombinasyonu ile liste yapısı olarak temsil edilmektedir. Ek olarak, girdi verileri olmadan kişi/hanehalkı ilişkileri oluşturmak için bir çözüm önerilmiştir.
PopGen	Ye ve ark. (2009).	Maricopa County, Arizona.	Kişi ve hanehalkları için eş zamanlı bir ayarlama sağlayan yeni bir teknik (AOG) kullanılmaktadır.
FSUMTS	Srinivasan ve Ma (2009), Srinivasan ve ark. (2008)	Tampa Bay ve South-East Florida	İstenen hanehalkı seçimi için olasılıksal yöntem kullanmak yerine sıralamalı bir yöntem kullanılmaktadır. Ayrıca, geriye dönük çıkarımlarda bulunularak sonuçlar çeşitlendirilmiştir.
CEMDAP	Pinjari ve ark. (2006) Guo ve Bhat (2007)	Dallas-Fort Worth area, Texas	Çok yönlü olasılık tablolarını birleştirmek için bir yöntem geliştirilmiştir.
ALBATROSS	Arentze ve ark. (2007)	European region	Hanehalkı düzeyi, dağılım ön işleme adımı sırasında kişi düzeyli dağılımdan hesaplanmaktadır.

1.7. PUMS Veri Kaynağı

Popülasyon sentezlemede literatürde en sık kullanılan veri kaynağı, PUMS (Public Use Micro data Sample) veri kaynağıdır. PUMS, bir mikro veri tipidir. PUMS kütükleri ABD'deki boş konut birimlerini, ilgilenilen alandaki insanların bilgilerini ve ilgilenilen alanın %1 ya da %5'ini yansıtan kayıtları içermektedir. Ek olarak ilgilenilen alana ait insanların dörtte birine ait bilgileri de içermektedir. Bu veri dosyasında her kişi ya da konuta ait bireysel ağırlıklar bulunmaktadır. Bu ağırlıklar bireysel kayıtlara uygulandığı zaman örnek, ilgili toplam kitleye genişleyebilmektedir.

Veri kaynağının konut kaydında yer alan öğelerden bazılarına örnek olarak, yüzölçümü; tarımsal satışlar; yatak odası; apartman aidatı; kira sözleşmesi; aile gelirleri; aile, alt aile ve ilişkilerinin kayıtları; çiftlik kayıtları; yangın, tehlike, sel sigortası; kullanılan yakıt; brüt kira; kalorifer yakıtı; hanehalkı geliri; hanehalkı tipi; mutfak imkânları; dilsel izolasyon; karavan masrafları; ipotek ödemesi; ipotek durumu; sıhhi tesisler; sahip olunan çocuk sayısı ve yaşları; emlak vergileri; oda sayısı; binaların büyüklüğü; devlet kodu; telefon hizmeti; memuriyet süresi; mevcut araç bilgisi ve bina yaşı verilebilir (PUMS Technical Documentation, 2008).

Birey kaydında yer alan öğelerden bazılarına örnek olarak, ingilizce konuşma yeteneği; yaş; soy; vatandaşlık; işçi sınıfları; sakatlık durumu; kazanç; eğitimsel kazanımları; dede-nine bilgileri; İspanyol kökeni; çalışılan saatler; gelir; sanayi; evde konuşulan dil; medeni hal; işe giderken kullanılan araç türü; göç bilgisi; göç durumu; hareketlilik durumu; tecrübe kıdem süresi; askerlik yaşı; meslek; kişinin ağırlığı; kişisel bakım kısıtları; doğum yeri; iş yerinin durumu; yoksulluk durumu; ırk; akrabalık; okul kaydı ve okul türü; iş için yola çıkma zamanı; iş için yolda harcanan zaman; araç kullanım süresi; yıllık çalışılan hafta süresi; işe giriş yılı ve iş durumu verilebilir (PUMS Technical Documentation, 2008).

PUMS verisi “super-PUMA (super-Public Use Microdata Areas) ve PUMA (Public Use Microdata Areas)” olarak bilinen coğrafi birimleri içermektedir. Minimum popülasyon eşikleri PUMA ve super-PUMA, PUMS verisinin gizliliğini korumak için oluşturulmuştur. %1’lik dosyalar için, super-PUMA’lar en az 400,000 bir nüfusa sahiptir ve bir PUMA ya da %5’lik PUMS dosyalarında tanımlanan bitişik PUMA’ların bir grubundan oluşmaktadır. %5’lik dosyalar her biri en az 100,000 nüfusa sahip PUMA’ları içermektedir. %5’lik dosyalar aynı zamanda super-PUMA’ların kodlarını da göstermektedir. Her devlet ayrı ayrı tanımlanmakta ve bir ya da daha fazla PUMA ve super-PUMA verilerinden oluşabilmektedir. Büyük metropol alanlar super-PUMA ve PUMA gibi alt bölümlere bölünebilmektedir.

PUMS verileri yaygın olarak değişkenler arasındaki ilişkileri tanımlayan istatistiksel modeller geliştirmek için kullanılmaktadır. PUMS dosyaları bireyler ve hanehalkları hakkında düzenlenmemiş kayıtlar dizidir. PUMS dosyaları bireyler ve hanehalklarına ait örneğin ilişki, cinsiyet, eğitim düzeyi ve çalışma durumu gibi bireysel ve özel yanıt bilgilerini içermektedir. PUMS dosyalarında 1 yıllık, 3 yıllık ve 5 yıllık veriler yer almaktadır. 3 yıllık ve 5 yıllık PUMS dosyaları PUMS’in 1 yıllık verilerinin çok yıllı

kombinasyonlarından oluşmaktadır. Bu kombinasyonlar oluşturulurken çeşitli ağırlıklandırmalar ve enflasyon ayarlama faktörleri gibi düzeltme ayarlarından yararlanılmaktadır. PUMS dosyaları, örneğin gaziler ve üniversite öğrencileri gibi belirli nüfus gruplarının özel bilgilerini elde etmek için ya da standart tablolardan belirli veri kategorileri (örneğin resmi yoksulluk eşiği yüzde 90 ile yüzde 99 arasında gelire sahip aileler) elde etmek mümkün olmadığı zaman kullanılabilir.

PUMS dosyalarından en yaygın yararlanan kullanıcı grup, değişkenler arasındaki ilişkileri modellemekle ilgilenen akademik araştırmacılar olmaktadır. Diğer yaygın kullanıcı grup ise, hükümet ve iş dünyasında çalışan araştırmacılar olmaktadır. Bu araştırmacılar, standart tablolardan daha farklı yollarla kategorize edilen ya da birbirlerine karşı çapraz tablosu oluşturulamayan özellikler ile ilgili çalışmak için PUMS dosyalarından yararlanmaktadırlar. Standart veri tabloları kullanıcıların ilgilendiği birçok soruna cevap verirken, bazı sorular bu tablolar tarafından cevaplandırılmamaktadır. Örneğin, standart tablolar yabancı uyruklu yerlilerin eğitimleri ile yoksulluk durumlarını gösteren bir tablo sağlayamamaktadırlar. Böyle bir tablo, PUMS dosyaları tarafından oluşturulabilmektedir. PUMS dosyaları aynı zamanda bir veri kullanıcısının standart tablolardan elde edemeyeceği kategorilere de ulaşma imkânı sağlamaktadır. Örneğin, standart tabloların çoğu 65 yaş ve üzeri, 55 – 64 yaş gibi yaş aralıkları kullanmaktadır. Fakat New York eyaletinde, Yaşlanma Ofisi tarafından yürütülen programlar 60 yaş ve üstü popülasyon için tasarlanmıştır. Bu yüzden bu ofisin, programlarının herhangi bir değişikliğinin etkisini araştırmak istedikleri zaman, birincil veri kaynağı olarak PUMS dosyalarına başvurmaları gerekmektedir (White ve Mesenbourg, 2009).

PUMS verileri, bireyler ve hanehalkları hakkındaki bilgilerin ifşa edilmesinin önlenmesi için gizli mikro veri kayıtlarından oluşmaktadır. PUMS dosyalarının gizliliği kısmen; veri değiş tokuşu, seçilen değişkenlerin üst kodlaması, coğrafi nüfus eşikleri, büyük hanehalkları için yaş değişimi ve bazı kategorik değişkenler üzerindeki ayrıntıların azaltılması süreçleri yardımıyla korunmaktadır (PUMS Technical Documentation, 2008). Veri değiş tokuşu, sıklık tablolarında gizliliği korumak için tasarlanmış veri ifşası kısıtlama yöntemlerinden biridir. Veri değiş tokuş işlemi, kaynak veriyi düzenleyerek ya da örnek kayıtlarını değiştirerek yapılmaktadır. Değiş tokuş bireysel kayıtlara uygulanmaktadır ve böylece mikro verilerin gizliliği korunmuş olmaktadır. Üst kodlama, tek bir kategoriye yerleştiren dağılımın belirli bir yüzdesi üzerinde veya tamamında veri ifşası kısıtlama yöntemlerinden biridir. Coğrafi nüfus eşikleri, nüfus sayımı belirli bir seviyenin altındaki

coğrafi birimlerin hanehalkları ve bireylere ait verilerin ifşasının önlenmesinde kullanılan bir yöntemdir. Yaş değişimi (hanehalkı üyelerinin yaşlarının değişimi), gizlilik konusundaki endişeler nedeniyle büyük hanehalkları (on ya da daha fazla bireyden oluşan hanehalkları) için gerekli olmaktadır. Kategorik değişkenler üzerindeki ayrıntıların azaltılması, eğer her bir kategorideki oluşumların sayısı belirlenen minimum eşiği karşılamıyorsa kullanılan bir yöntemdir.

PUMS mikro veri dosyaları özellikle “kullanıcıların kendilerinin oluşturabileceği” özel tablolar oluşturulmasına olanak vermektedir. Kullanıcılar istenen değişken kümesiyle ilişkilendirilen çok çeşitli tablolar oluşturabilmektedirler. Mikro veri örnekleri, belirli küçük coğrafi alanların tanımlanmayı ya da küçük popülasyonlar için ayrıntılı çapraz tablolar yapmayı gerektirmeyen çalışmaları yapan kullanıcılar için faydalı olmaktadır. Mikro veri kullanıcıları sıklıkla, mevcut nüfus tablolarında gösterilmeyen nüfus değişkenleri arasındaki ilişkileri incelemek için ya da özel olarak tanımlanmış popülasyonların özelliklerini ele almak için çalışmaktadır (PUMS Technical Documentation, 2008).

2. YAPILAN ÇALIŞMALAR

2.1. R ile Konumsal Mikro Simülasyon

Araştırmanın esnekliği, etkinliği ve tekrarlanabilirliği üzerinde yazılım kurallarının büyük etkisi vardır. Neredeyse 30 yıl önce, Clarke ve Holm mikro simülasyon için kullanılan programlama dili seçiminin dikkate alınması gereken bir konu olduğunu ileri sürmüşlerdir (Clarke ve Holm, 1987). Bunun o zaman olduğu gibi şimdi de doğru olduğu görülmektedir. Yazılım konusu, konu ile ilgili makalelerde nadiren tartışılmaktadır ve kaliteli konumsal mikro simülasyon paket programları azdır. “The Flexible Modelling Framework (FMF)” çeşitli modelleme işlerini yapabilen Java’da yazılmış önemli bir istisnadır.

Birey tabanlı modelleme alanında güçlü ve farklı bir yazılım topluluğu bulunmaktadır ve bunların çoğu konumsal mikro simülasyon çalışan araştırmacılar için yararlı olabilecek kavramlar ortaya çıkarmışlardır. Örneğin Mannion ve ark. JAMSIM’i tanıtmışlardır (Mannion vd., 2012). JAMSIM, iyi bir birey tabanlı modelleme uygulaması yapmak için R dili ile yazılan, sonuçların grafiklenmesinde ve analizinde kullanılan yeni bir yazılım çerçevesidir. JAMSIM ve alandaki diğer yazılımların, (NetLogo ve MASON ABM sistemleri ve bunlara ilişkin eklentiler dahil) rakip olarak değil bu alanda geliştirilen R yazılımlarına paralel olarak gelişmekte olduğu görülmektedir.

Yazılım seçilmesine etki eden faktörler; yazılımın maliyeti, ne kadar zamandır kullanıldığı, özellikleri ve performansdır. Araştırmacılar için en önemlileri, analizin öğrenme, yazma, uyarlama ve ilişki kurma kolaylığı ve hızıdır. R, bu alanların her birinde üstünlük sağlamaktadır.

R, güçlü bir grafiksel kullanıcı arayüzüne (GUI) sahip olmasına rağmen, örneğin Microsoft Excel ve SPSS gibi istatistiksel programlarla karşılaştırıldığında düşük seviyeli bir dildir. R, verileri analiz etmede ve modellemede büyük bir esneklik sunmaktadır ve kullanıcı tanımlı işlemlerin kolaylıkla oluşturulmasına olanak vermektedir. Bu özellikler, konumsal mikro simülasyonun yapılabilmesi için yazılımda istenen nitelikleridir. Öte yandan R, C ve Python gibi genel amaçlı dillere kıyasla yüksek seviyeli bir dildir. İstatistiksel işlemleri yapmak için sıfırdan kod yazmak yerine, R genellikle önceden yazılmış kodları kullanma olanağı sunmaktadır. Örneğin x değişkeninin ortalamasını

hesaplamak için; Python'da 20 karaktere ihtiyaç varken ($\text{float}(\text{sum}(x))/\text{len}(x)$), R'de sadece 7 karakter yeterli olmaktadır ($\text{mean}(x)$). Önceden yazılmış fonksiyon kodları ile konumsal mikro simülasyon modellerinin yazılması, okunması ve sonuçların analiz edilmesi açısından kolaylık sağlaması açısından R faydalı olmaktadır. R ve Python'da ortalamanın hesaplanması örneği çok klişe bir örnek olabilmektedir. Daha geniş açıdan bakıldığında R, istatistiksel verilerle çalışmak için tasarlanmıştır. Bu yüzden R kurulduğu zaman birçok fonksiyon zaten bünyesinde vardır. Örneğin R bünyesindeki $\text{lm}()$ kodu, doğrusal regresyon modeli oluşturmak için kullanıldığında doğrudan istatistiksel analizin yapılmasına olanak sağlamaktadır. Birey tabanlı modellemede, sonuçların istatistiksel analizi, genellikle modeli çalıştırmaktan daha fazla zaman almaktadır (Thiele vd., 2014). Aynı durum konumsal mikro simülasyon modellemesi için de geçerli olmaktadır. R nin istatistiksel analiz yetenekleri sayesinde, R konumsal mikro simülasyon modellemesi için ideal bir tercih olmaktadır.

R aktif ve büyüyen bir kullanıcı topluluğuna sahiptir ve R'nin genişletilmesi kolaydır. R'nin aşırı esnekliği, diğer programlama dillerinde yazılmış kodları çalıştırmasına da olanak sağlamaktadır. R'de kullanılan geniş bir algoritma yelpazesi bulunmaktadır. Bu özellik R kodunu önemli ölçüde hızlandırmak için kullanılmıştır. Örneğin, yeni R paketi "dplyr" veri işleme görevlerinin zor işlerini yapmak için C++ kodunu kullanmaktadır. R'nin esnekliğinin, açık kaynaklı oluşunun ve güçlü kullanıcı topluluğunun sonucu olarak, R'nin yeteneklerini yeni fonksiyonlarla genişleten binlerce paket bulunmaktadır. Yapılan gelişmeler her zaman R ekosistemine eklenmektedir.

R akademik araştırmalarda, eğitimde ve endüstride yaygın olarak kullanılmaktadır. R sürekli olarak büyümektedir ve şu an bünyesinde 7000'den fazla paket bulunmaktadır. Örneğin R'de bulunan ipfp ve mipfp kütüphaneleri kullanıldığında, popülasyon sentezi için gerekli kodun satır sayıları ve hesaplama zamanı büyük ölçüde azaltabilmektedir. R'nin belirgin bir avantajı da herkesin bir paket yazabilmesidir. Bu aynı zamanda potansiyel bir dezavantajdır. Çünkü çok fazla paket bulunmaktadır ve yeni kullanıcılar için hangi paketlerin daha güvenilir olduğu ve hangisinin hangi durumlar için en uygun olduğunu belirlemede zorlanmaktadırlar (Hornik, 2012). Fakat bu problem R'nin açık kaynaklı olmasından dolayı hafiflemektedir. Kullanıcılar R programlamayı öğrenmek istiyorlarsa herhangi bir fonksiyonun tam olarak nasıl çalıştığını açık kaynaklar sayesinde görebilmektedirler.

R’yi birbirinden ayırık komutlar dizisi olarak değil, birbirine bağlı komutlardan oluşan bir dil olarak düşünmek faydalı olmaktadır. Burada yazılan kodların önemli bir görevi de, fikirleri bir kişiden diğer bir kişiye iletmektir. R’yi daha çok erişilebilir yapan yeni gelişme ise RStudio’dur. RStudio sayesinde her kullanıcı R’yi kolaylıkla kullanabilmektedir.

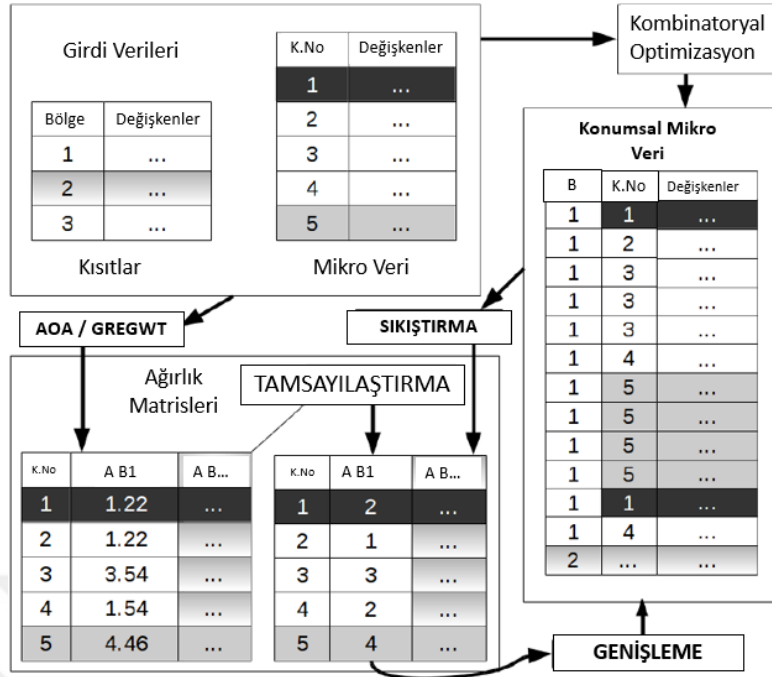
2.2. Konumsal Mikro Simülasyonda Ağırlıklandırma, Tamsayılaştırma ve Genişleme Süreçleri

Konumsal mikro simülasyonda, bireyleri konumlara atamak için çeşitli yöntemler kullanılabilir. İstatistiksel analiz yöntemlerinin çoğunda olduğu gibi deterministik ve stokastik yöntemler bulunmaktadır. AOA gibi deterministik yöntemlerin sonuçları asla değişmemektedir, rasgele ve olasılık sayıları kullanılmadığından sonuçtaki ağırlıklar her zaman aynı olmaktadır. Diğer yandan benzetim tavlama gibi stokastik yöntemler rasgele sayıları kullanmaktadırlar. Literatürde stokastik ve deterministik yaklaşımlar arasındaki uçurum genellikle yeniden ağırlıklandırma ve kombinatoriyal optimizasyon yöntemleri arasındaki ayrımla eşleştirilebilmektedir. Yeniden ağırlıklandırma yöntemleri genellikle her birey-bölge kombinasyonu için tamsayı olmayan ağırlıkları hesaplamaktadır. Kombinatoriyal optimizasyon yöntemleri ise genellikle her bölgeye mikro veri anketinden bireyleri tek tek rasgele atayarak çalışmakta ve her değişikliğin ardından uyum iyiliğini yeniden hesaplamaktadır. Eğer bir kişi söz konusu bölgeye atandıktan sonra gözlemlenen ve simüle edilen sonuçlar arasındaki uyum iyiliği artarsa birey o bölgede kalmaktadır. Eğer uyum bozulursa birey o bölgeden çıkartılmakta ve bölgeye farklı bir bireyin yerleştirilmesinin etkisi test edilmektedir.

Kesirli ağırlıkların yeniden ağırlıklandırılması ve kombinatoriyal optimizasyon algoritmaları arasındaki bu ayrım önemlidir. Kombinatoriyal optimizasyon yöntemleri her bir bireyin bir bölgeye yerleştirilmesi ile sonuçlanırken, yeniden ağırlıklandırma yöntemleri ise her bölgeye yerleştirilen bireylerin kesirleriyle sonuçlanmaktadır. Konumsal mikro veri yaratmada kombinatoriyal optimizasyon ve yeniden ağırlıklandırma yöntemleri arasındaki ayrım aslında görüldüğü kadar net değildir.

AOA gibi yeniden ağırlıklandırma algoritmaları ile üretilen kesirli ağırlıklar, tamsayı ağırlıklarına “tamsayılaştırma” süreci ile dönüştürülür (Şekil 15). “Genişleme” süreci boyunca tamsayılaştırma yöntemi ile üretilen tamsayı ağırlık matrisi, nihai konumsal mikro

veri çıktısına dönüştürülebilmektedir. Bu yüzden “tamsayılaştırma ve genişleme” süreçlerinin birleşmesi ile oluşan süreçler, ağırlık matrislerinin kombinatoriyal optimizasyon algoritmalarının doğrudan ürettiği çıktı formatına dönüştürülmesine olanak sağlamaktadır. Başka bir deyişle kesirli ağırlıklandırma popülasyon sentezi sürecinde kombinatoriyal optimizasyon yaklaşımlarına alternatif olarak kullanılabilir. Bu sürecin tersi de mümkündür. Kombinatoriyal optimizasyon algoritmalarıyla üretilen sentetik konumsal mikro veri “sıkıştırma” olarak adlandırılan bir adımla daha kompakt bir ağırlık matrisine dönüştürülebilmektedir. Bu tamsayı ağırlığı, “tamsayılaştırma” ile üretilen tamsayı matrisi ile aynı boyuta sahip olmaktadır. Tamsayılaştırma, genişletme ve sıkıştırma süreçleri popülasyon sentezinde kesikli ağırlıklandırma ve kombinatoriyal optimizasyon yaklaşımlarının aslında aynı şey olarak görülmesine izin vermektedir. Popülasyon sentezinin farklı yöntemleri arasındaki bu eşdeğerlik, kombinatoriyal optimizasyon yaklaşımlarının kesikli ağırlıklandırma yöntemlerinin özel bir durumu olarak görülebilmesi ya da tam tersi şeklinde ifade edilebilmektedir. Bu yüzden konumsal mikro veri üreten tüm deterministik ve stokastik yaklaşımlar, bireylere ağırlıkları atayan farklı yöntemler ve algoritmalar olarak görülebilmektedir. Bir konumu temsil eden bireylere yüksek ağırlık verilmektedir. (Kombinatoriyal optimizasyonda bu çoğu kez çoğaltmaya eşdeğerdir.) Bir konumda nadir bulunan kişilere ise düşük ağırlık verilmektedir. Buna ek olarak bir konumu temsil eden bireyler olmayabilir, bu durumda ise ağırlık sifıra eşit olmaktadır.



Şekil 15. Stokastik kombinatoriyal optimizasyon ve AOA gibi deterministik yeniden ağırlıklandırma algoritmalarını içeren konumsal mikro veri oluşturulmasında farklı yaklaşımların şematik gösterimi

Ağırlık kavramı, popülasyon sentezinin konumsal mikro veriyi nasıl ürettiğini anlamak için önem taşımaktadır. Bireylerin tümünün özellikleri hakkında bilgimizin olmadığı, sadece her konumun toplam nüfusunu bildiğimiz bir bölge hayal edelim. Bu durumda örnekte bulunan özelliklerin dağılımının bütün nüfusunun dağılımını temsil ettiğini varsayabiliriz. Bu senaryoda bireyler örnekten rasgele seçilecek ve rasgele bölgelere yerleştirilecek ve her bir konumdaki birey özelliklerinin dağılımı mikro verilerle asimptotik olarak aynı olacaktır.

Yeniden ağırlıklandırma yöntemleri konumdaki her bireye ağırlık katmaktan ibarettir. Bu yöntem, elimizde konumun temsili bir örneği varsa ve popülasyonun azınlığı da buna dahil edilmişse başarılı olmaktadır. Aksine, elimizde bireysel düzeyde nüfusun çoğunluğu elimizde varsa ve örneğin elimizde hala çalışan yaşlı bir birey yoksa bu tür birey nihai verimizde olamaz. Bu durumu önlemek için kullanılan öneriler, ya bu aykırı değerleri içerecek şekilde örneği değiştirmek ya da yeni bireyler yaratmak için bazı bireysel düzeyde verilerin karıştırılmasına izin veren genetik algoritma gibi farklı bir yaklaşım kullanmaktır. Bu sorunu çözmek için diğer bir kullanışlı alternatif tamsayı programlamada, sezgisel bir yöntem olan tabu arama algoritmasıdır (Glover, 1997; Glover

ve Laguna, 2013). Bu sezgisel yaklaşım, AOA tarafından üretilen başlangıç çözümünü geliştirmek için kullanılabilir. Bu yaklaşımın, AOA tarafından üretilen başlangıç çözümünü geliştirmek için kullanılabilir.

Konumsal mikro simülasyon üretilmesi süreci, her bir konumdaki her bir bireye ağırlık tahsis edilmesi sürecidir. Belirli bir birey-konum kombinasyonuna ait ağırlık ne kadar yüksekse, bireyin o konuma ait olması o kadar yüksek olmaktadır.

Ağırlık matrisinin bir özelliği, her bir sütun toplamının her konumdaki toplam nüfusa eşit olmasıdır. Matrisler bireysel düzeyli verileri coğrafi olarak toplanmış verilere bağlamaktadır. Bu ağırlık matrislerini üretmek için birçok yöntem bulunmaktadır. Bu yöntemler bazen literatürde yeniden ağırlıklandırma yöntemleri olarak adlandırılmaktadır (Tanton vd., 2011). Bunlar, AOA gibi “deterministik” ve benzetim tavlama gibi sözde rasgele sayı üreticisine dayanan “olasılıksal” yöntemlerden oluşmaktadır.

Konumsal mikro simülasyonda AOA bireyleri konumlara atamak için kullanılmaktadır. AOA yaparken hesaplama kolaylığı sağlayan yaklaşımlar bulunmaktadır. Araştırmacıya zamandan tasarruf etmesini sağlayan ve AOA yöntemini uygulamada kullanılan R kütüphaneleri bulunmaktadır (“ipfp” ve “mipfp” kütüphaneleri). Kullanıcı her iterasyonda çok fazla koda ihtiyaç duymadan bu kütüphaneleri kullanarak bireyleri konumlara atayan kesirli ağırlık matrisleri üretebilmektedir.

AOA tekniği R’deki “ipfp” kütüphanesi ile daha az kod kullanma imkânı bularak, daha hızlı çalışmaktadır. “ipfp” fonksiyonu AOA algoritmasını C dilinde çalıştırmaktadır. R’de düzinelerce kod satırı ile yazılan AOA algoritmasının iki satıra yoğunlaştırılması etkileyici olmaktadır. C dili tamsayıda başarısızdır, bu yüzden girdi kısıtları sayısal veri olmalıdır. “ipfp” kütüphanesinde kodlar C dilinde oluşturulmuş ve optimize edilmiştir. “ipfp” kütüphanesi verilerin yüklendiği ve sonuçları döndüren sihirli bir kutuya benzetilmektedir. Yakınsama, iki ardışık yineleme arasındaki farkı içeren matris normunun fonksiyonda elle ayarlanabilen tolerans değerine ulaşması anlamına gelmektedir. Varsayılan tolerans değeri çok küçük bir sayıdır (0.0000000000000000222). Yinelemelerin sayısı da “maxit (maximum iterasyon)” argümanı ile belirtilerek de ayarlanabilmektedir. R çıktısında simülasyondan önceki ve sonrası ağırlıklar arasındaki farkları içeren uzaklık yer almaktadır. Bu iki matris eşit olduğu zaman algoritma yakınsamakta ve uzaklık matrisi de sıfır olmaktadır. Eğer uzaklık tolerans değerinin altına düşerse algoritma durmaktadır. Bilgisayar gerçek sayıların sayısal yaklaşımlarını hesaplamaktadır. Örneğin sonuç 4/3 olduğunda bilgisayar bunu sonsuz sayıda ondalık olarak kaydetmemektedir, sayıyı kesmekte ve yaklaşımı hesaplamaktadır. “maxit” ’i durdurma ölçütü olarak kullanıldığı

zaman dikkatli olunması gerekmektedir. Çünkü her uygulamada, uyum seviyesine ulaşmak için gerekli argümanların sayısını tahmin etmek imkânsızdır. Bu yüzden maxit argümanı çok büyük olduğunda gereksiz yere zaman kaybı olmaktadır, küçük olduğu zaman ise sonuç yakınsamayabilmektedir. Tolerans değeri kullanmak da aynı şekilde tehlikeli olmaktadır. Elimizde tekrarlanan aynı matrisler varsa fakat uzaklık matrisi üzerindeki yaklaşımlar argümandan büyük ise algoritma devam edecektir. Bu yüzden tolerans (tol) ve maxit birlikte kullanılmalıdır. “ipfp” ile üretilen ağırlıklar kesirlidir. Birey tabanlı modelleme uygulamaları için tamsayı ağırlıklar gerekli olmaktadır. Tamsayı ağırlıklar “ipfp” adından sonra “tamsayılaştırma” süreciyle üretilmektedir. Tamsayılaştırma sürecinde iki yöntem geliştirilmiştir. İlki bireyleri ağırlıkları ile orantılı bir olasılıkla rasgele bir şekilde seçmektedir. İkincisi ise yuvarlama yöntemi olarak tanımlanmaktadır. Literatürde tamsayılaştırmada en iyi bir yöntem, Lovelace ve Ballas tarafından tanıtılan TRS (‘Truncate, Replicate, Sample’) yöntemi olarak bilinmektedir (Lovelace ve Ballas, 2013).

“mipfp” kütüphanesi, girdileri ağırlık matrisi haline dönüştürmek için kullanılan diğer bir R kütüphanesidir. “mipfp” kütüphanesi, “ipfp” kütüphanesiyle kıyaslandığında AOA algoritmasının daha genelleştirilmiş bir uygulaması olduğu söylenebilmektedir. “mipfp” kütüphanesi de popülasyon sentezi için tasarlanmıştır. “ipfp” kütüphanesi (çapraz tablolaştırılmamış) karşılıklı olarak özel kısıt tabloları içeren iki boyutlu ağırlık matrisi ile çalışma olanağı sunmaktadır. “ipfp” kütüphanesi, çapraz tablosu olmayan ve sadece marjinaleri olan kısıtları kullanan uygulamalar için faydalı olmaktadır. “mipfp” ise daha esnek, örneğin yaş/cinsiyet ve yaş/sınıf kombinasyonları gibi kısıt değişkenlerine sahip çoklu çapraz tablolarla çalışmaya olanak sunmaktadır. “mipfp” bu yüzden AOA’nın çok boyutlu bir uygulaması olarak düşünülebilmektedir. Dolayısıyla çok boyutlu olabilmektedir. Bu bakımda “mipfp”, sadece iki boyutlu durumları çözebilen “ipfp” den daha avantajlı olmaktadır. “mipfp” ana fonksiyonu “Ipfp” dir. Bu fonksiyon, bir dizi toplamsal düzeyli kısıt tablosu seçeneklerine dayanan N boyutlu bir ağırlık matrisini doldurmaktadır.

“ipfp” kütüphanesinde olduğu gibi, “mipfp” kütüphanesinde de tol ve/veya maxit argümanları belirlenip durdurma kriteri seçilebilmektedir. Maksimum iterasyon sayısına ulaşıldıktan sonra ya da yakınsama sağlandıktan sonra (hangisi önce olursa), fonksiyon algoritmanın yakınsaması hakkında diğer bilgilerin yanı sıra güncellenmiş diziyi içeren

bir listeyi döndürmektedir. Eğer hedef marjinaler tutarlı değil ise girdi verileri frekansların yerine olasılıkları dikkate alarak normalize edilmektedir.

“mipfp” kütüphanesinin diğer bir avantajı çapraz tablolanmış kısıtlamaların eklenmesine izin veriyor olmasıdır. Tamsayılaştırma ve genişleme adımları “ipfp” nin adımlarına oldukça benzemektedir. Mantık olarak aslında tamamen aynıdır. “ipfp” vektörleri, “mipfp” dizileri kullanmaktadır. Bu sebeple diziler için başka bir genişleme süreci üretilmiştir. Sonuç olarak, “mipfp” kütüphanesi üretilen dizilerin doğruluğunu değerlendirmek için maksimum olabilirlik, minimum ki-kare ve en küçük kareler yaklaşımı gibi problemleri çözmeye olanak sağlayan işlevselliklerle diğer yöntemleri de içermektedir.

Büyük veri kümelerinde çalıştırılmak üzere tasarlanmış olsa da, “mipfp” daha küçük veri kümelerinde de kullanılabilir. mipfp ve ipfp kütüphaneleri arasındaki en büyük fark, yazıldıkları dildir. mipfp R dilinde, ipfp ise C dilinde yazılmıştır. Diğer bir fark ise mipfp nin amacı çok çeşitli veri girişlerini (çapraz tablo ya da marjinal dağılımlar gibi) kabul ederek konumsal mikro simülasyon üretmek için AOA ‘yı uygulamaktır. Aksine ipfp $Ax=b$ şeklindeki bir cebir problemini çözmek için oluşturulmuştur. Bunun amacı A matrisi ve b vektörünü tanımladıktan sonra $Ax=b$ gibi bir x vektörünü bulmaktır. Bu yüzden boyutlar sabitlenir (iki boyutlu bir matris ve iki vektör). “mipfp” R dilinde yazıldığı için R kullanıcıları tarafından kolaylıkla kullanılabilir. Buna karşılık ipfp C dilinde yazıldığı için, hızlı olmasına rağmen nispeten anlaması daha zor olarak tanımlanmaktadır. “ipfp” her bir bölgede girdi mikro verilerindeki her bireye ağırlık vermektedir. “mipfp” ise bundan farklı çalışmaktadır. Kısıt değişkenleri bakımından aynı özelliklere sahip bireyler girdi hücresinde sadece bir kez görülmektedirler, sayıları başlangıç ağırlığı ile temsil edilmektedir. Bu nedenle mipfp, kısıt değişkeni kategorilerinin her bir kombinasyonundaki kişi sayısını belirlemektedir. Bu özellik mipfp’yi hesaplamada ipfp’den daha verimli yapmaktadır. “ipfp” kısıtlar ve başlangıç birey düzeli verilere ihtiyaç duyarken, “mipfp” kısıtlar ve anketten elde edilen olumsuzluk tablosuna ihtiyaç duymaktadır. “mipfp” mikro veriler mevcut olmadığı durumlarda kullanılmalıdır. “ipfp” sonucunda bireyler için ağırlıklar yaratılırken, “mipfp” sonucunda olumsuzluk tablosu için ağırlıklar üretilmektedir. “mipfp” ve “ipfp” sonuçları birbirine dönüştürülebilir. “mipfp” kütüphanesinin bir avantajı da belleğidir. “ipfp” nin bireyleri kadar satır içeren tabloya ihtiyacı vardır. Buna karşılık “mipfp” nin farklı kategorilerin sayısına karşılık gelen bir boyutlu tabloya ihtiyacı vardır. “mipfp”, “ipfp” ’e göre daha uyarlanabilir.

Tüm durumlar ele alındığında her iki kütüphane de aynı sonuçları vermektedir. Birey tabanlı modellemede en önemlisi adım, her iki yöntemle üretilen ağırlık matrislerinin konumsal mikro veriye dönüştürülebilme aşamasında tamsayılaştırma ve genişleme süreçlerine girmelerinin gerektiğidir.

Tamsayılaştırma kesirli ağırlıkları minimum bilgi kaybıyla tam sayıya dönüştürme süreci olarak tanımlanmaktadır. Ağırlıkların yuvarlanması basit bir tamsayılaştırma yöntemidir fakat yuvarlamada sonuçlar çok zayıftır. Bu işlem aynı zamanda değiştirilmeden ağırlıklandırılmış örnekleme olarak da ifade edilebilmektedir. Literatürde sık kullanılan iki tamsayılaştırma yöntemi bulunmaktadır. Bunlardan ilki ağırlıkları basit olasılıklar olarak ele almaktadır. İkincisi ise her kesirli ağırlığın en yakın altındaki ve en yakın üstündeki tam sayıdan elde edebileceği maksimum ve minimum tamsayı ağırlıklarını kısıtlamaktadır ve bu yöntem literatürde TRS (Truncate Replicate Sample) olarak bilinmektedir.

Tamsayılaştırma; sentetik konumsal mikro veriden elde edilen bireylerin sayısına karşılık gelen, kesirli sayılardan oluşan bir vektörün bir tamsayı vektörüne dönüştürülmesi süreci olarak tanımlanır. Yeni vektörün uzunluğu ağırlık vektörüne eşit olmalıdır. Yüksek ağırlıklar tamsayılaştırılmış versiyonda da yüksek olmalıdır. Bu yüzden yüksek ağırlıkların çalışmada etkin olabilmesi için düşük ağırlıklı olanlardan orantılı olarak daha sık örneklenmesi gerekmektedir. Her bölge için sentetik ve kısıt verileri arasındaki uyumu maksimize etmek için bu tamsayılaştırma sürecini sistematik olarak üstlenen bir algoritmaya ihtiyaç vardır.

Birçok mevcut tamsayılaştırma yöntemi vardır. Lovelace ve Ballas bunların 5 tanesini test etmiş ve orantılı olasılıkların olasılıksal yöntemlerinin ve TRS'nin deterministik rakiplerine göre daha iyi sonuç gösterdiğini bulmuşlardır (Lovelace ve Ballas, 2013). TRS yöntemi 3 adımdan oluşmaktadır. İlk adım tüm ağırlıkların tamsayıları kısmını tutarak ondalık kısımları kesmektir (Truncate). İkinci adım bu tamsayıları konumdaki her bireyin sayısı gibi düşünerek tekrarlamaktır (Replicate). Son adım ise bölgedeki birey sayısını en iyi şekilde örnekleme kullanmaktır (Sample). Bu aşama ondalık ağırlıklara karşılık gelen olasılıklarla bir örnekleme kullanmaktadır. TRS daha kompakt bir şekilde kısıtlanmış entegrasyon yöntemidir. Bu tanım, TRS kullanılarak üretilen konumsal mikro verilerin farklı tamsayılaştırma algoritmaları kullanılarak üretilen konumsal mikro verilerle kıyaslandığında TRS'nin toplam kısıtlara daha iyi uymasının nedenini açıklamaktadır (Lovelace ve Ballas, 2013).

Nihai konumsal mikro veriyi oluşturmak için “genişleme süreci” uygulanmaktadır. Seçilen yönteme bağlı olarak, ya her birey için bir ağırlık vektörü (örneğin “ipfp”) ya da her olası farklı kategori için bir ağırlık vektörü (örneğin “mipfp”) sonucu elde edilmektedir. Genişleme adımı bir ya da daha fazla ağırlık yapısına sahip olduğunda biraz farklılık göstermektedir. Bireyler için üretilecek ağırlık vektöründe, her bir tamsayılaştırılmış ağırlık her bir birey için gerekli tekrarların sayısına karşılık gelmektedir. Kategoriler için üretilecek ağırlık vektöründe, veri yapısı bireylerinkinden tamamen farklı olduğu için genişleme süreci de farklı olmaktadır. Bireylerin ağırlıklarında sadece bireylerin çoğaltılmasına ihtiyaç duyuluyorken, kategorilerin ağırlıklarında bireyleri yeniden oluşturmak için kategorilerin isimleri yeniden elde edilmektedir.

2.3. Yapılan Çalışma

Bu tezde yapılan konumsal mikro simülasyon çalışmasında kullanılan veri iki konumdan ve toplam 1207902 bireyden oluşmaktadır. Birinci konumda 359597 birey, ikinci konumda ise 848305 birey bulunmaktadır. Her bir bölgenin yaş ve cinsiyet bilgileri Tablo 2’de gösterilmiştir.

Tablo 2. Konum 1 ve konum 2’ de bulunan bireylerin cinsiyet ve yaş bilgileri

	Kadın	Erkek	0-49 Yaş	50 ve 50+ Yaş
Konum 1	178945	180652	288023	71574
(%)	49,76	50,24	80,10	19,90
Konum 2	429746	418559	670363	177942
(%)	50,66	49,34	79,02	20,98

Çalışmada toplam 12443 bireye ait mikro veri kullanılmıştır. Bu mikro veride birinci konumda 2956, ikinci konumda ise 9487 birey bulunmaktadır. Mikro veri bireylerin cinsiyet, yaş, ortalama gelir, okuma yazma durumu, buldukları hanhalklarında sahip oldukları toplam binek araç sayısı ve otomobil sürücü belgesine sahip olma durumları hakkında bilgiler içermektedir.

Tablo 3. Mikro veride bulunan bireylerin cinsiyet, yaş, okuma yazma durumu ve sürücü belgesine sahip olma bilgileri

Cinsiyet	Erkek	Kadın
Birey Sayısı	6439	6004
(%)	51,75	48,25
Yaş	0-49 Yaş	50 ve 50+ Yaş
Birey Sayısı	9089	3354
(%)	73,05	26,95
Okuma Yazma Durumu	Bilen	Bilmeyen
Birey Sayısı	12049	394
(%)	96,83	3,17
Sürücü Belgesi	Var	Yok
Birey Sayısı	5766	6677
(%)	46,34	53,66

Tablo 4. Mikro veride bulunan bireylerin ortalama gelir bilgileri

Ortalama Gelir	Birey Sayısı	(%)
125 TL	58	0,47
375 TL	110	0,88
625 TL	381	3,06
875 TL	2070	16,64
1125 TL	1810	14,55
1375 TL	1532	12,31
1625 TL	1373	11,03
1875 TL	1171	9,41
2125 TL	928	7,46
2375 TL	779	6,26
3000 TL	1068	8,58
4250 TL	778	6,25
6250 TL	299	2,40
8750 TL	57	0,46
10000 TL	29	0,23

Tablo 5. Mikro veride hanehalkının sahip olduğu toplam binek araç sayısı

Hanehalkının Sahip Olduğu Toplam Binek Araç Sayısı	Birey Sayısı	(%)
0	6198	49,81
1	5789	46,52
2	395	3,17
3	61	0,49

İki konumda mevcut toplam yaş ve cinsiyet bilgilerinden (Tablo 2) ve bu konumlardaki mikro verisine ulaşılan bireylerin (Tablo 3,4,5) bilgilerinden yararlanılarak bir konumsal mikro simülasyon çalışması yapılmıştır. Çalışmanın 2.1. bölümünde bahsedilen sebeplerden dolayı R programı tercih edilmiştir. Çalışmada öncelikle popülasyon sentezi sürecinin en önemli adımı olan ağırlık matrisinin oluşturulmasından bahsedilmiştir. Daha sonra “ipfp” ve “mipfp” kütüphaneleri kullanılarak AOA ile mikro simülasyon çalışması yapılmıştır. Bu mikro simülasyon sonuçları “tamsayılaştırma” ve “genişleme” süreçlerinden geçirilerek nihai konumsal mikro simülasyon sonuçlarının nasıl elde edildiği gösterilmiştir.

2.3.1. Ağırlık Matrisinin Oluşturulması

Bir bireyin bir konuma ait olması, onların o konumdaki ağırlıkları ile temsil edilmektedir. Her bir ağırlık bireyi bir konuma bağlamaktadır. Bu yüzden ağırlıkların sayısı mikro verideki bireylerin sayıları ile çarpıldığında konumdaki toplam sayıya eşit olmaktadır. Ağırlık kavramı popülasyon sentezinin konumsal mikro veriyi nasıl ürettiğini anlamak için önemlidir.

R’de AOA yöntemi kodları 2013 yılında Lovelace tarafından yazılmıştır (Lovelace ve Ballas, 2013). R’de AOA yöntemi her iterasyonda her bir kısıtın ağırlıklarını kaydederek çalışmaktadır. Her iterasyonda ağırlıklılar yeniden güncellenmektedir.

Her bireyin bulunduğu konumu temsil eden bir ağırlık matrisinin nasıl oluşturulduğu örneği bu bölümde üretilmiştir. Algoritma konum konum çalışmakta ve ağırlık matrisi de sütun sütun doldurularak çalışmaktadır. Öncelikle bir ağırlık matrisi ve her konumdaki bireylerin marjinal dağılımları oluşturulmaktadır. Satırların konumları, sütunların ise

değişkenlerin kategorilerini içerdiği bir nesne oluşturulur. Daha sonra her adımda belleği korumak için ağırlık matrisi kopyalanmaktadır.

```
mikroveri_top0 <- t(apply(veri, 1, function(x) 1 * mikroveri_top))
colnames(mikroveri_top0) <- names(veri)
agirliklar1 <- agirliklar2 <- agirliklar
```

R’de AOA iç içe for döngüleri ile yazılabilmektedir. Örneğin verideki yaş değişkenine göre ayarlama yapılması ve ağırlık matrisinin oluşturulması istendiğinde döngü aşağıdaki gibidir:

```
for(j in 1:n_konum){
  for(i in 1:n_yas){
    index <- mikroveri_kat[, i] == 1
    agirliklar1[index, j] <- agirliklar[index, j] * yas[j, i] / mikroveri_top0[j, i]
  }
  print(agirliklar1)
}
```

Sadece yaş değişkenini ele alarak “agirliklar1” ağırlık matrisi elde edilmiştir. Burada elde edilen ağırlık matrisi 12443x2 boyutludur. Yukarıdaki kod yaş kısıtlarına uyacak şekilde her ağırlığı bir katsayı ile çarparak ağırlık matrisini güncellemektedir. İlk adımda tüm ağırlıklar 1 ile başlatılmıştır ve yeni ağırlıklar katsayılara eşittir. Algoritma konum konum ilerletilir ve matrisin her bir sütunu bir bölgeye karşılık gelmektedir. Algoritma önce konum 1 daha sonra konum 2 için hesaplama yapmaktadır. Bu, matrisin neden sütun sütun doldurulduğunu açıklamaktadır. İlk konum için her bireyin ağırlıkları ağırlık matrisinin ilk sütunundadır. Ve her bireyin bireysel özellikleri “mikroveri_kat” matrisinde yer almaktadır. “mikroveri_kat” ve “agirliklar1” ’in ilk sütunu çarpıldığında bir vektör elde edilmektedir. Bu vektörün değerleri konum 1’in her bir kategorisindeki bireylerin sayısına karşılık gelmektedir. Toplamları bulmak için aşağıdaki kod yazılmıştır:

```
mikroveri_top2 <- mikroveri_top1 <- mikroveri_top0 * NA
```

oluşturulmuş ve bir for döngüsü kurulmuştur:

```
for(i in 1:n_konum){
  mikroveri_top1[i, ] <- colSums(mikroveri_kat * agirliklar1[, i])
}
```

Burada sadece yaş kısıtından yararlanılarak ağırlıklandırılmış bireysel düzeyli veri seti oluşturulmuştur. Kodun doğru çalışıp çalışmadığını kontrol edilmek istediğinde ise her bir konumdaki var olan birey sayısı ile her konumdaki simüle edilen birey sayısını karşılaştırmak gerekir.

```
rowSums(mikroveri_top1[, 1:2]) # her bir konumda simüle edilen birey sayısı
[1] 359597 848305
rowSums(veri[, 1:2]) # her bir konumda var olan gerçek birey sayısı
[1] 359597 848305
rowSums(mikroveri_top1[, 1:2]) == rowSums(veri[, 1:2])
[1] TRUE TRUE
```

Yukarıdaki doğrulama tekniğine göre elde edilen simülasyon sonucu doğrudur. Burada simüle edilen toplam birey sayısına bakıldığında sonuç aşağıdaki gibidir:

	K	E	0-49 yas	50 ve 50+ yas
[1,]	288023	71574	262829.1	96767.89
[2,]	670363	177942	620012.5	228292.47

Bu sonuca henüz tamsayılaştırma ve genişleme süreçleri uygulanmadığı için, sonuçlar tamsayı değildir. Gerçek birey sayısı ile simüle edilen bireylerin ne kadar uyumlu olduğuna bakılmak istendiğinde korelasyon sonucuna bakılmaktadır. Korelasyon sonucu -1 ile 1 arasında değer almaktadır. Sonuç simülasyon verisi ile gerçek simülasyon verisi arasındaki korelasyon sayısı 1 olduğunda, veriler arasındaki uyumun mükemmel olduğu söylenmektedir.

```

vec <- function(x) as.numeric(as.matrix(x))
cor(vec(mikroveri_top0), vec(veri))
[1] 0.6869546 # Konum 1 ayarlaması sonrası korelasyon
cor(vec(mikroveri_top1), vec(veri))
[1] 0.7724485 # Konum 1 & Konum 2 ayarlaması sonrası korelasyon

```

Sadece yaş değişkeninden yararlanılarak elde edilen simülasyon sonucu ile gerçek veri arasındaki korelasyon katsayısı 0.77 olarak bulunmuştur.

Ağırlık matrisi oluşturulurken mikro veride yer alan değişkenlerden ne kadar yararlanılırsa simülasyon sonucu o kadar doğru sonuç verecektir. Nitekim “mikroveri_top1” sonucuna bakıldığında sonucun gerçek verideki yaş değişkenine daha uyumlu olduğu, cinsiyet değişkeni hesaba katılmadığı için cinsiyet değişkenine göre bakıldığında ise o kadar uyumlu olmadığı açıkça görülebilmektedir. Mikro veride yer alan cinsiyet değişkeni de ele alınıp “agirliklar2” ağırlık matrisi oluşturulmuştur.

```

for(j in 1:n_konum){
for(i in 1:n_cinsiyet + n_yas){
index <- mikroveri_kat[, i] == 1
agirliklar2[index, j] <- agirliklar1[index, j] * veri[j, i] /
mikroveri_top1[j, i]
}
}

```

Burada elde edilen ağırlık matrisi “agirliklar2” de 12443x2 boyutludur. Daha sonra yukarıda olduğu gibi toplam hesaplanmıştır.

```

for(i in 1:n_konum){
mikroveri_top2[i, ] <- colSums(mikroveri_kat * agirliklar2[, i])
}

```

Cinsiyet değişkenine göre ayarlama yapıp ağırlık matrisi oluşturulduğunda simülasyon sonucu aşağıdaki gibidir:

```
mikroveri_top2
      K      E      0-49 yas      50 ve 50+ yas
[1,] 288051.7 71545.26 288023      71574
[2,] 670422.7 177882.28 670363      177942
```

Yeni oluşturulan simülasyon sonucunun gerçek veri ile uyumuna bakılmak istediğinde korelasyon değeri 0,81 olarak bulunmuştur.

```
cor(vec(mikroveri_top2),vec(veri))
[1] 0.8114438
```

Cinsiyet değişkeni de ele alındığında korelasyon değerinin arttığı görülmektedir ve gerçek veri ile uyumun giderek arttığı söylenebilir. Bu bölümde ağırlık matrisinin R de nasıl oluşturulduğu gösterilmiştir. AOA yöntemi tek bir nihai ağırlık matrisine ulaşmadan önce çoklu iterasyon gerektirmektedir. AOA yöntemini kullanmada kolaylık sağlayan R de kullanılan hazır kütüphaneler bulunmaktadır: “ipfp” ve “mipfp”. Çalışmanın 2.3.2 ve 2.3.3 bölümlerinde bu paketler kullanılarak konumsal mikro simülasyonun nasıl yapıldığı gösterilmiştir.

2.3.2. “ipfp” Kütüphanesi ile Konumsal Mikro Simülasyon

AOA yöntemi R’de “ipfp” kütüphanesi ile daha hızlı çalışma imkânı bulmuştur ve bu kütüphane yardımıyla daha az kod kullanılmaktadır. R’de AOA yöntemini “ipfp” fonksiyonu kullanarak çalıştırmak için, “ipfp” kütüphanesinin yüklenmesi gerekir.

```
library(ipfp)
veri <- apply(veri, 2, as.numeric) # girdilerin sayısal veriye dönüştürülmesi
ipfp(veri[1,], t(mikroveri_kat), x0 = rep(1, n_mikroveri)) # “ipfp” fonksiyonunun çalıştırılması
```

Ağırlık matrisinin oluşturulması çalışmanın 2.3.1 bölümünde ele alınmış ve AOA ile mikro simülasyon yapabilmek için bir satırdan fazla kod yazılmıştı. “ipfp” burda yazılan

kod satırını iki satıra yoğunlaştırmıştır. İlk satır girdi veri setinin sayısal veri setine dönüştürülmesidir. Çünkü “ipfp” fonksiyonu tamsayı verilerde başarısızdır, bu yüzden veri önce sayısal yapılmıştır. İkinci satırdaki kod ise doğrudan fonksiyonun çalıştırılması kodudur. Burada direkt “ipfp” fonksiyonu yardımı ile ağırlıklar elde edilmektedir. Yukarıdaki kodda olduğu gibi kaç yinelemenin olmasını belirtilmediği durumlarda maksimum iterasyon sayısı 1000 alınmaktadır ve yakınsama sağlanana kadar süreç devam etmektedir. Maksimum iterasyon sayısı kullanıcı tarafından belirlenebilmektedir (örneğin maxit=20 gibi). “maxit” argümanı maksimum iterasyon sayısının özelleştirilmesini sağlamaktadır.

```
ipfp(veri[1,], t(mikroveri_kat), rep(1, n_mikroveri), maxit = 20, v = T)
```

Burada yakınsamanın anlamı ise iki ardışık yineleme arasındaki farkı içeren matrisin normunun tolerans değerine ulaşmasıdır. Tolerans değeri de fonksiyonda kullanıcı tarafından ayarlanabilmektedir (örneğin tol= 0.0000001 gibi). Tolerans değeri kullanıcı tarafından girilmediğinde bu değer 2.220446e-16 olarak kabul edilmektedir.

Gerçek matris ile simülasyon matrisi eşit olduğunda algoritma yakınsamaktadır ve bu iki matris arasındaki uzaklık sıfır olmaktadır. Eğer bu iki matris arasındaki fark tolerans değerinin altında ise algoritma durmaktadır.

```
mikroveri_katt <- t(mikroveri_kat)
```

```
x0 <- rep(1, n_mikroveri) # ipfp' de ağırlık tahminlerinin başlangıç noktası
```

İki konum için “ipfp” ile ağırlık matrisi oluşturma süreci iç içe for döngüsü ile yazılabilir.

```
agirliklar_maxit_2 <- agirliklar
```

```
for(i in 1:ncol(agirliklar)){
```

```
+ agirliklar_maxit_2[,i] <- ipfp(veri[i,], mikroveri_katt, x0, maxit = 2)
```

```
+ }
```

Daha kısa bir şekilde kod yazmak istediğinde iç içe for döngüsü yerine apply fonksiyonu da kullanılabilir.

```
> agirliklar <- apply(veri, MARGIN = 1, FUN =
+function(x) ipfp(x, mikroveri_katt, x0, maxit = 20))
```

Böyle yazıldığında “ipfp” fonksiyonu yukarıdaki for döngüsünde olduğu gibi her bölgeye sırayla uygulanmaktadır.

Burada “ipfp” fonksiyonu yardımıyla elde edilen AOA ağırlıklarının mantıklı olup olmadığını kontrol etmek gerekir.

```
mikroveri_top <- t(apply(agirliklar, 2, function(x) colSums(x * mikroveri_kat)))
colnames(mikroveri_top) <- colnames(veri)
```

```
mikroveri_top
```

	K	E	0-49 yas	50 ve 50+ yas
[1,]	178945	180652	288023	71574
[2,]	429746	418559	670363	177942

```
veri
```

	K	E	0-49 yas	50 ve 50+ yas
[1,]	178945	180652	288023	71574
[2,]	429746	418559	670363	177942

```
cor(vec(mikroveri_top),vec(veri))
```

```
[1] 1
```

“ipfp” fonksiyonu yardımıyla elde edilen AOA sonuçlarının, mevcut girdi verisi kısıtlarıyla birebir örtüştüğü görülmektedir.

2.3.3. “mipfp” Kütüphanesi ile Konumsal Mikro Simülasyon

“mipfp” kütüphanesi, “ipfp” kütüphanesinin daha geliştirilmiş bir uygulamasıdır ve popülasyon sentezi için tasarlanmıştır. “ipfp” çapraz tablo olmayan kısıt tablolarından yararlanarak iki boyutlu ağırlık matrisi üretmektedir. “mipfp” daha esnektir, kısıt değişkenlerinin çoklu çapraz tablolarıyla çalışmaya olanak sağlamaktadır. “mipfp” bu

yüzden AOA yönteminin çok boyutlu bir uygulamasıdır. Bu anlamda “mipfp”, sadece iki boyutlu durumlarda çözüm sunabilen “ipfp” ’e göre daha ileri düzeylidir.

“mipfp” ’in ana fonksiyonu Ipfp()’dir. Bu paketle çalışabilmek için “mipfp” kütüphanesinin yüklenmesi gerekir (library(mipfp)).

“mipfp” kütüphanesinin nasıl çalıştığını göstermek için mevcut verilerden tahmin yapmak istensin. “mipfp” çok boyutlu çapraz tablolarla çalışmaya olanak sağladığı için, 12443 bireye ait cinsiyet, yaş, okuma yazma durumu ve sürücü belgesi değişkenleri kullanılarak hedef değişken oluşturulmuştur.

```
cinsiyet2 <- c(K=6004, E=6439) # Kadın=K, Erkek=E
yas2 <- c(G=9089, Y=3354) # 0-49 Yaş=G(Genç), 50 ve 50+ Yaş=Y(Yaşlı)
okuma <- c(B=12049, BM=394) # Okuma-Yazma Bilen =B, Okuma-Yazma
Bilmeyen=BM
surucu <- c(S= 5766, SD= 6677) # Sürücü = S , Sürücü Değil=SD
target <- list(cinsiyet2, yas2, okuma, surucu) #Hedef değişken
descript <- list(1,2,3,4)
target
  K      E
6004  6439
  G      Y
9089  3354
  B      BM
12049  394
  S      SD
5766   6677
names <- list(names(cinsiyet2),names(yas2),names(okuma), names(surucu))
weight_init <- array(1, c(2,2,2,2), dimnames = names)
weight_init[, c("G"), c("B"), c("S")] <- 0
```

Tüm kısıt değişkenleri yukarıdaki gibi kodlamıştır ve bu matrisin boyutu 2x2x2x2’dir. Bu aşamadan sonra “mipfp” çalıştırılır.

Ipfp fonksiyonu durdurma kriteri olarak tolerans değeri (tol) ve/ veya maksimum iterasyon sayısı (maxit) tanımlanmasına da izin vermektedir. İterasyon sayısı maksimum

iterasyon sayısına ulaştığında ya da yakınsama gerçekleştiğinde (hangisi önce olursa olsun), fonksiyon güncellenmiş diziyi içeren bir listeyi döndürmektedir.

Aynı zamanda, “mipfp” başlangıçta kısıt değişkenlerinin marjinaleri tutarlı olmadığı durumlarda, girdi verilerinin sıklıkları olasılıklarla normalize edilerek fonksiyonun kullanılabilmesine olanak sağlamaktadır.

Bu çalışmada mikroveri değişkenlerinin marjinalerinde tutarsızlık olmadığı için çalışma frekanslar üzerinden gerçekleştirilmiştir.

```
result <- Ipfp(weight_init, descript, target, iter = 50, print = TRUE, tol = 1e-10)
```

“mipfp” kullanılarak Ipfp fonksiyonu çalıştırıldığında yakınsama maksimum iterasyon sayısından önce gerçekleşmiş ve sonuca 46. iterasyondan sonra ulaşılmıştır.

```
Margins consistency checked!
... ITER 1
  stoping criterion: 2731.969
... ITER 2
  stoping criterion: 506.3403
.
.
.
... ITER 45
  stoping criterion: 1.118678e-10
... ITER 46
  stoping criterion: 5.820766e-11
Convergence reached after 46 iterations!
sum(result$x.hat)
12443
```

Sonuç nihai ağırlık matrisini içermektedir ve kişilerin toplam sayısı “sum(result\$x.hat)” ile kontrol edilmiştir. “result\$x.hat” ile mipfp kütüphanesi ile elde edilen sonuçlar görülmektedir.

Tablo 6. “mipfp” ile mikro verinin simülasyon sonuçları özet tablosu

		G	Y
B & S	K	0	2.497,246
	E	0	2.678,176
BM & S	K	284,97	4,62E-06
	E	305,6122	4,95425E-06
B & SD	K	3.221,788	5,22E-05
	E	3455,212	5,60E-05
BM & SD	K	5,96E-06	9,66E-20
	E	6,39E-06	1,04E-19

“mipfp” ile elde edilen 12443 bireye ait mikro simülasyon sonuçları Tablo 6’da görüldüğü gibidir. Bu tabloya bakıldığında;

- Okuma yazma bilen, sürücü belgesi olan, yaşlı ve erkek birey sayısı 2678, okuma yazma bilen, sürücü belgesi olan, yaşlı ve kadın birey sayısı 2497’dir.
- Okuma yazma bilmeyen, sürücü belgesi olan, genç ve kadın birey sayısı 285, okuma yazma bilmeyen, sürücü belgesi olan, genç ve erkek birey sayısı 306’dır.
- Okuma yazma bilen, sürücü belgesi olmayan, genç ve kadın birey sayısı 3222, okuma yazma bilen, sürücü belgesi olmayan, genç ve erkek birey sayısı 3455’tir.
- Okuma yazma bilmeyen ve sürücü belgesi olmayan birey bulunmamaktadır.

2.3.4. Tamsayılaştırma

“ipfp” ve “mipfp” kütüphaneleri ile elde edilen sonuçlara bakıldığında sonuçların kesirli olduğu görülmektedir. Bu sonuçlar nihai birey tablosu olarak kullanılamamaktadır. Tamsayılaştırma bu kesirli ağırlıkları minimum bilgi kaybıyla tamsayıya dönüştüren yöntemlere verilen addır. Bu çalışmada tamsayılaştırma yöntemi olarak daha önce bahsedilen TRS yöntemi tercih edilmiştir.

```
> int_trs <- function(x){
  + # genelleme amacıyla kullanılır, x bir vektördür.
  + # Bu fonksiyonun bir matris ile de çalışmasına izin verir.
  + xv <- as.vector(x) # trs'nin matrisler üzerinde çalışmasına izin verir.
```

```

+ xint <- floor(xv) # ağırlığın tamsayı kısmıdır.
+ r <- xv - xint # ağırlığın ondalık kısmıdır.
+ def <- round(sum(r)) # popülasyon açığı.
+ topup <- sample(length(x), size = def, prob = r)
+ xint[topup] <- xint[topup] + 1
+ dim(xint) <- dim(x)
+ dimnames(xint) <- dimnames(x)
+ xint
+ }

```

- “ipfp” ile Tamsayılaştırma:

Çalışmada yukarıda “ipfp” bölümünde elde “agirliklar” matrisinin konum 1 ağırlıklarını tamsayılaştırmak istediğimizde R kodu aşağıdaki gibidir:

```
> int_agirlik1 <- int_trs(agirliklar[,1])
```

Burada tamsayılaştırılan ağırlık matrisi “int_agirliklar1” 12443x1 boyutlu bir matristir. Bu matrisin tamsayılaştırılmadan önceki ilk 6 ağırlık değerine ve tamsayılaştırmadan sonraki ilk altı değerine bakmak için head fonksiyonunu kullanılırsa:

```
head(agirliklar[,1]) # “agirliklar” matrisinin tamsayılaştırılma sürecinden önceki ilk 6 ağırlık değeri
```

```
[1] 20.72376 32.67206 30.77094 22.00413 32.67206 30.77094
```

```
head(int_agirlik1) # “agirliklar” matrisinin tamsayılaştırılma sürecinden sonraki ilk 6 ağırlık değeri
```

```
[1] 21 33 30 22 32 30
```

- “mipfp” ile Tamsayılaştırma:

Çalışmada yukarıda “mipfp” bölümünde elde edilen “result\$x.hat” çapraz tablo matrisini (Tablo 6) tamsayılaştırmak istediğimizde R kodu aşağıdaki gibidir:

```
mipfp_int <- int_trs(result$x.hat)
mipfp_int
```

“mipfp_int” matrisi tamsayılaştırılmış ağırlıkları içermektedir. Tablo 7’de gösterilmektedir.

Tablo 7. “mipfp” ile elde edilen sonuçlarının tamsayılaştırılmış hali

		G	Y
B & S	K	0	2497
	E	0	2679
BM & S	K	285	0
	E	305	0
B & SD	K	3221	0
	E	3456	0
BM & SD	K	0	0
	E	0	0

Tamsayılaştırma işlemi ağırlığa yakın tamsayıları seçmektedir. Bu adım marjinal toplamlarda bazı hatalara neden olabilmektedir. Maksimum hata, kategori başına her zaman 1’dir. Bu hata toplam nüfus bazında düşünüldüğünde her zaman çok küçük bir hata olarak kalmaktadır.

Nihai mikro veriyi üretmek için son adım genişleme adıdır.

2.3.5. Genişleme

Konumsal mikro simülasyonda son adım genişleme sürecidir. Seçilen yönteme bağlı olarak, ya elimizde birey başına ağırlıkların olduğu bir vektör (“ipfp”) ya da farklı olası kategori başına bir ağırlık matrisi (“mipfp”) sonucu olacaktır.

- “ipfp” ile Genişleme:

Birey başına ağırlıkların genişleme durumunda (“ipfp”), her bir tamsayılaştırılmış ağırlık her birey için gerekli tekrarların sayısına karşılık gelmektedir. Bu amaçla ilk olarak örnek verilerin kimliklerine ait vektörü oluşturan fonksiyon yazılır:

```
int_expand_vector <- function(x){
+ index <- 1:length(x)
+ rep(index, round(x))
+ }
```

Bu fonksiyon bir vektörü döndürür. Bu vektörün her elementi nihai konumsal mikro veride bir bireyi temsil eder ve her hücre ilgili örnek bireyin kimlik bilgilerini içermektedir. Örneğin tamsayılaştırma bölümünde “agirlıklar” matrisinin konum 1 ağırlıklarının tamsayılaştırılmış ağırlık matrisi “int_agirlıklar1” ’in genişleme sürecinin R kodu aşağıdaki gibidir:

```
exp_indices <- int_expand_vector(int_agirlıklar1) # bireyleri ağırlıklara göre genişletir.
```

Nihai veri, basit bir şekilde var olan verilerin tekrar edilmesiyle bulunabilmektedir. Konum 1 için nihai konumsal mikro simülasyon verisine bakmak için aşağıdaki kod kullanılır:

```
mikroveri_orig[exp_indices,]
```

Konum 1’de 359597 birey bulunmaktadır. Yukarıda kod çalıştırıldığında mikro veriden, nihai konumsal mikro simülasyon verisindeki 359597 bireye ait cinsiyet, yaş ve gelir verileri elde edilmiş olur.

Konum 1 için elde edilen nihai mikro simülasyon cinsiyet değişkeni açısından incelenmek istenirse,

```
mikroveri_orig[exp_indices,]$cinsiyet=as.factor(mikroveri_orig[exp_indices,]$cinsiyet)
table(mikroveri_orig[exp_indices,]$cinsiyet)
```

Tablo 8. Konum 1'deki bireylerin genişleme süreci sonrası cinsiyet değişkeni simülasyon sonuçları

Cinsiyet	Erkek	Kadın
Birey Sayısı	180546	179051
(%)	50,21	49,79

Sonuçta 180546 erkek ve 179051 kadın bireyin simüle edildiği görülmektedir. Bu bireylerin ortalama gelir değişkenlerinin durumu incelenmek istenirse R kodu aşağıdaki gibidir:

```
mikroveri_orig[exp_indices,]$gelir=as.factor(mikroveri_orig[exp_indices,]$gelir)
table(mikroveri_orig[exp_indices,]$gelir)
```

Tablo 9. Konum 1'deki bireylerin genişleme süreci sonrası ortalama gelir değişkeni simülasyon sonuçları

Ortalama Gelir	Birey Sayısı	(%)
125 TL	1643	0,46
375 TL	3334	0,93
625 TL	10915	3,04
875 TL	59805	16,63
1125 TL	52635	14,64
1375 TL	44384	12,34
1625 TL	39520	10,99
1875 TL	33615	9,35
2125 TL	26842	7,46
2375 TL	22702	6,31
3000 TL	30910	8,60
4250 TL	22281	6,20
6250 TL	8522	2,37
8750 TL	1643	0,46
10000 TL	846	0,24

Konum 1'de yer alan bireylerin ortalama gelir bilgileri Tablo 9'da gösterilmektedir. Bu bireylerin yaş dağılımlarına bakılmak istediğinde öncelikle yaş değişkenini kategorize etmek gerekir, daha sonra da dağılıma bakılır:

```

brks <- c(0, 49, 110)
cut(mikroveri_orig[exp_indices,]$yas,breaks=brks)
labs <- c("0-49 yas","50 ve 50+ yas")
cut(mikroveri_orig[exp_indices,]$yas, breaks = brks, labels=labs)
mikroveri_orig[exp_indices,]$yas<-
cut(mikroveri_orig[exp_indices,]$yas,breaks=brks,labels = labs)
levels(mikroveri_orig[exp_indices,]$yas)
mikroveri_orig[exp_indices,]$yas=as.factor(mikroveri_orig[exp_indices,]$yas)
table(mikroveri_orig[exp_indices,]$yas)

```

Tablo 10. Konum 1’deki bireylerin genişleme süreci sonrası yaş değişkeni simülasyon sonuçları

Yaş	0-49 Yaş	50+ Yaş
Birey Sayısı	291518	68079
(%)	81,07	18,93

Burada genişleme adımından sonra konum 1 için nihai konumsal mikro simülasyon sonuçlarının nasıl elde edildiğini gösterilmiştir.

- “mipfp” ile Genişleme:

Kategori başına bir ağırlık matrisinin genişleme sürecinde (“mipfp”), birey yaratmak için kategorilerin isimleri geri alınmaktadır. Burada süreç, diziyi bir veri çerçevesine çeviren bir fonksiyonla başlamaktadır. Bu fonksiyon tamsayılaştırılmış matrisin veri çerçevesini kaydeder.

```

int_expand_array <- function(x){
  count_data <- as.data.frame.table(x) # nihai popülasyonun kategori indekslerini
  ayarlar.
  indices <- rep(1:nrow(count_data), count_data$Freq) # nihai bireyleri oluşturur.
  ind_data <- count_data[indices,]
  ind_data
}

```

Yukarıdaki kod çalıştırıldığında sonuç bir veri çerçevesi olmaktadır. Satırlar bireyleri oluşturur. Sütunlar ise değişken kategorilerdir. Genişleme adımı ise aşağıdaki gibi yazılır:

```
ind_mipfp <- int_expand_array(mipfp_int)
Genişleme sürecinden sonra simülasyon sonuçlarına bakılmak istendiğinde,
ind_mipfp$Var1 =as.factor(ind_mipfp$Var1) # Var1= Cinsiyet
table(ind_mipfp$Var1)
ind_mipfp$Var2 =as.factor(ind_mipfp$Var2) # Var2= Yaş
table(ind_mipfp$Var2)
ind_mipfp$Var3 =as.factor(ind_mipfp$Var3) # Var3= Okuma - Yazma Durumu
table(ind_mipfp$Var3)
ind_mipfp$Var4 =as.factor(ind_mipfp$Var4) # Var 4 = Sürücü Belgesi
table(ind_mipfp$Var4)
```

Burada “mipfp” ile genişleme süreci sonrası nihai konumsal mikro simülasyon sonuçlarının nasıl elde edildiğini gösterilmiştir. Buradan elde edilen sonuçlar Tablo 11’de gösterilmiştir.

Tablo 11. “mipfp” ile genişleme süreci sonrası simülasyon sonuçları

Cinsiyet	Erkek	Kadın
Birey Sayısı	6440	6003
(%)	51,76	48,24
Yaş	0-49 Yaş	50+ Yaş
Birey Sayısı	7267	5176
(%)	58,40	41,60
Okuma-Yazma Durumu	Biliyor	Bilmiyor
Birey Sayısı	11853	590
(%)	95,26	4,74
Sürücü Belgesi	Var	Yok
Birey Sayısı	5766	6677
(%)	46,34	53,66

3.BULGULAR VE SONUÇLAR

Çalışmanın ikinci bölümünde mikro veri ile konumsal mikro simülasyonun R de nasıl uygulandığı gösterilmiştir. Bu bölümde, her iki konum için “ipfp” kütüphanesi ile AOA yöntemi uygulanarak nihai konumsal mikro veri sonuçları elde edilmiş ve sonuçlar verilmiştir. Bu bölümde ayrıca gerçek veri ile sentezlenen veri kıyaslanmıştır.

```
ints_df <- NULL
set.seed(42)
for(i in 1:nrow(veri)){
  ints <- int_expand_vector(int_trs(agirliklar[, i])) # tamsayılaştırma ve genişleme
  ints_df <- rbind(ints_df, data.frame(mikroveri_orig[ints,], konum = i))
}
```

Burada “ints_df” nihai konumsal mikro veri setini temsil etmektedir ve iki konumdan oluşan ve toplam 1207902 bireye ait sonucu içermektedir.

```
nrow(ints_df)
[1] 1207902
```

İstenen bir konuma ait popülasyon sentezi sonuçlarına incelenmek istendiğinde,

```
ints_df[ints_df$konum == 1, ]# Konum 1’e ait popülasyon sentezi sonuçlarını verir
ints_df[ints_df$konum == 2, ]# Konum 2’e ait popülasyon sentezi sonuçlarını verir
```

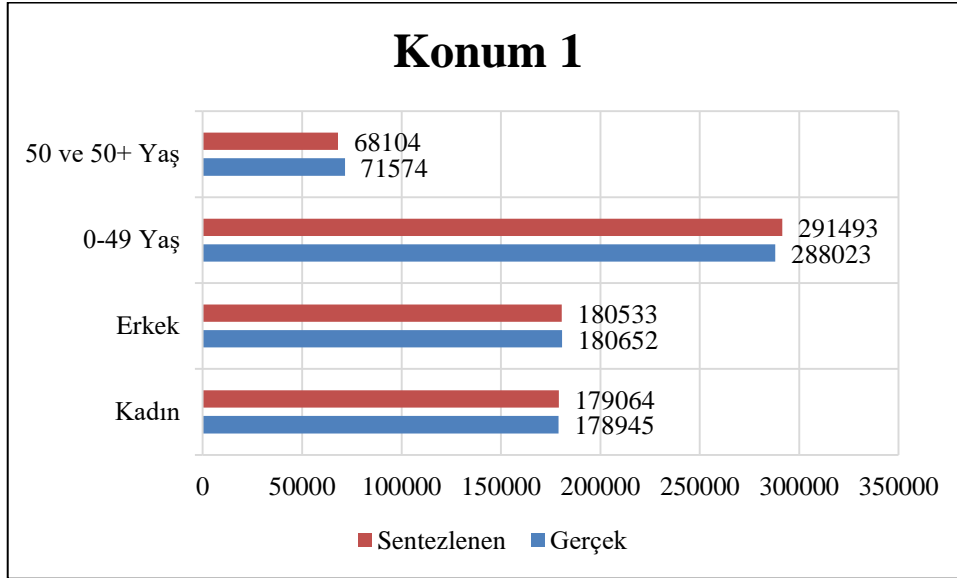
kodları yazılarak sonuçlar yorumlanabilir.

“ipfp” kullanılarak AOA yöntemi ile her iki konuma ait gerçek ve sentezlenen birey sayısı Tablo 12’de gösterilmiştir. Bu tabloya göre, yapılan konumsal mikro simülasyon çalışmasının mikro veriden popülasyon sentezlemede mükemmel başarı sağladığı görülmektedir.

Tablo 12. Konumlardaki gerçek ve sentezlenen birey sayısı karşılaştırması

	Gerçek	Sentezlenen	Fark (%)
Konum 1	359597	359597	0
Konum 2	848305	848305	0

Konum 1'deki 2956 kişilik mikro veriden konum 1 popülasyonuna (359597) ait mikro simülasyon sonuçları elde edilmiştir. Şekil 16'da konum 1'e ait gerçek ve simülasyon verileri grafik ile gösterilmiştir. Cinsiyet değişkenine bakıldığında gerçek veride toplam 178945 kadın ve 180652 erkek birey varken, mikro veriden simülasyon sonucunda elde edilen veriye bakıldığında 179064 kadın ve 180533 erkek bireyin sentezlendiği görülmektedir. Yaş değişkenine bakıldığında gerçekte 50 yaş altı 288023, 50 yaş ve üstü 71574 birey bulunuyorken, sentezlenen veriye bakıldığında 50 yaş altı 291493, 50 yaş ve üstü 68104 birey sentezlendiği görülmektedir. Konum 1'e ait nihai konumsal mikro simülasyon sonuçları özet bilgileri Tablo 13'te görülmektedir. Cinsiyet değişkenine bakıldığında gerçek veri ve simülasyon verisi arasında farkın yüzde 0,03, yaş değişkenine bakıldığında ise farkın yüzde 0,96 olduğu görülmektedir.



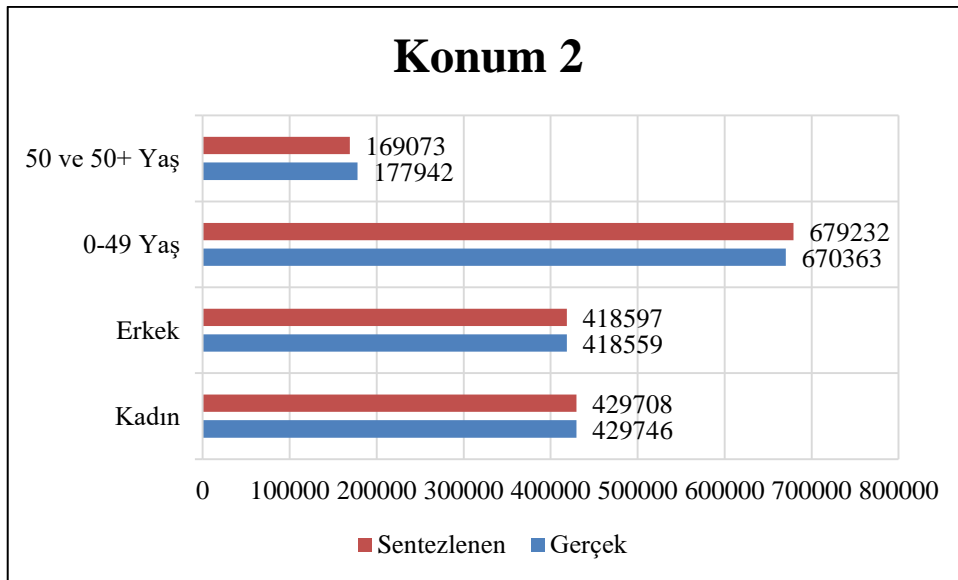
Şekil 16. Konum 1'de sentezlenen ve gerçek verilerin karşılaştırılması

Tablo 13. Konum 1'e ait nihai konumsal mikro simülasyon sonuçları ve karşılaştırması

Konum 1	Gerçek	Gerçek (%)	Sentezlenen	Sentezlenen (%)	Fark (%)
Kadın	178945	49,76	179064	49,80	0,03
Erkek	180652	50,24	180533	50,20	-0,03
0-49 Yaş	288023	80,10	291493	81,06	0,96
50 ve 50+ Yaş	71574	19,90	68104	18,94	-0,96

Konum 2'deki 9487 kişilik veriden konum 2 popülasyonuna (848305) ait mikro simülasyon sonuçları elde edilmiştir. Şekil 17'de konum 2'ye ait gerçek ve simülasyon verileri grafik ile gösterilmiştir. Cinsiyet değişkenine bakıldığında gerçek veride toplam 429746 kadın ve 418559 erkek birey varken, mikro veriden simülasyon sonucunda elde edilen veriye bakıldığında 429708 kadın ve 418597 erkek bireyin sentezlendiği görülmektedir.

Yaş değişkenine bakıldığında gerçekte 50 yaş altı 670363, 50 yaş ve üstü 177942 birey bulunuyorken, sentezlenen veriye bakıldığında 50 yaş altı 679232, 50 yaş ve üstü 169073 birey sentezlendiği görülmektedir. Konum 2'ye ait nihai konumsal mikro simülasyon sonuçları özet bilgileri Tablo 14'te görülmektedir. Cinsiyet değişkenine bakıldığında gerçek veri ve simülasyon verisi arasında farkın çok çok küçük olduğunu (yüzde -0,00448), yaş değişkenine bakıldığında ise farkın yüzde 1,05 olduğu görülmektedir.



Şekil 17. Konum 2'de sentezlenen ve gerçek verilerin karşılaştırılması

Tablo 14. Konum 2'ye ait nihai konumsal mikro simülasyon sonuçları ve karşılaştırması

Konum 2	Gerçek	Gerçek (%)	Sentezlenen	Sentezlenen (%)	Fark (%)
Kadın	429746	50,66	429708	50,65	0,00
Erkek	418559	49,34	418597	49,35	0,00
0-49 Yaş	670363	79,02	679232	80,07	1,05
50 ve 50+ Yaş	177942	20,98	169073	19,93	-1,05

Konum 1 mikro verisinde 2956 bireye ait ortalama gelir bilgisi bulunmaktadır. Bu bilgiden yararlanarak konum 1'deki tüm bireylerin (359597) ortalama gelirleri tahmin edilmiştir. Bu tahmin sonuçlara ait özet bilgiler Tablo 15'te görülmektedir. Bu tabloya bakıldığında konum 1 popülasyonunun yüzde 16,63'ünün ortalama gelirinin 875 TL, yüzde 14,64'ünün ortalama gelirinin 1125 TL ve yüzde 12,34'ünün ortalama gelirinin 1375 TL olduğunu tahmin edilmiştir.

Tablo 15. Konum 1'e ait nihai ortalama gelir tahmini sonuçları

Konum 1		
Ortalama Gelir	Sentezlenen	(%)
125 TL	1641	0,46
375 TL	3337	0,93
625 TL	10917	3,04
875 TL	59807	16,63
1125 TL	52632	14,64
1375 TL	44368	12,34
1625 TL	39496	10,98
1875 TL	33649	9,36
2125 TL	26896	7,48
2375 TL	22695	6,31
3000 TL	30866	8,58
4250 TL	22281	6,20
6250 TL	8528	2,37
8750 TL	1638	0,46
10000 TL	846	0,24

Konum 2 mikro verisinde 9487 bireye ait ortalama gelir bilgisi bulunmaktadır. Bu bilgiden yararlanarak konum 2'deki tüm bireylerin (848305) ortalama gelirleri tahmin edilmiştir. Bu tahmin sonuçlara ait özet bilgiler Tablo 16'da görülmektedir. Bu tabloya bakıldığında konum 2 popülasyonunun %16,64'ünün ortalama gelirinin 875 TL, yüzde 14,62'sinin ortalama gelirinin 1125 TL ve yüzde 12,34'ünün ortalama gelirinin 1375 TL olduğunu tahmin edilmiştir.

Tablo 16. Konum 2'ye ait nihai ortalama gelir tahmini sonuçları

Konum 2		
Ortalama Gelir	Sentezlenen	(%)
125 TL	3872	0,46
375 TL	7821	0,92
625 TL	25765	3,04
875 TL	141161	16,64
1125 TL	124045	14,62
1375 TL	104660	12,34
1625 TL	93279	11,00
1875 TL	79328	9,35
2125 TL	63412	7,48
2375 TL	53483	6,30
3000 TL	72833	8,59
4250 TL	52677	6,21
6250 TL	20103	2,37
8750 TL	3871	0,46
10000 TL	1995	0,24

Hem konum 1'de hem de konum 2'de popülasyonun en büyük çoğunluğunun ortalama geliri 875 TL olarak tahmin edilmiştir.

Konumlara ait toplam ortalama gelir ve kişi başı ortalama gelir tahminleri Tablo 17'de yer almaktadır. Bu tabloya göre konumlarda yer alan bir bireyin ortalama geliri 1898 TL olarak tahmin edilebilir.

Tablo 17. Konum 1 ve konum 2'ye ait toplam ortalama gelir ve kişi başı ortalama gelir tahmini

Konum	Toplam Ortalama Gelir	Kişi Başı Ortalama Gelir
1	682540000	1898,069
2	1610426250	1898,405

Konumlara ait okuma yazma durumu tahmin sonuçları Tablo'18 de görülmektedir. Bu tabloya göre her iki konumda da okuma yazma bilenlerin popülasyonun yüzde 97'sini oluşturduğu görülmektedir.

Tablo 18. Konum 1 ve konum 2'ye ait okuma-yazma durumu tahmini sonuçları

Konum	Okuma-Yazma Bilen (Sentezlenen)	(%)	Okuma-Yazma Bilmeyen (Sentezlenen)	(%)
1	349442	97,18	10155	2,82
2	823633	97,09	24672	2,91

Konumlara ait bireylerin otomobil sürücü belgesi sahip olma durumu tahminlerine ilişkin sonuçlar Tablo 19'da görülmektedir. Bu tabloya göre her iki konumda da popülasyonun yaklaşık yüzde 45'inin sürücü belgesine sahip olduğu tahmin edilmiştir.

Tablo 19. Konum 1 ve konum 2'ye ait sürücü belgesi tahmin sonuçları

Konum	Sürücü Belgesi Olan	(%)	Sürücü Belgesi Olmayan	(%)
1	164467	45,74	195130	54,26
2	385491	45,44	462814	54,56

Konumlardaki hane halklarının sahip olduğu toplam binek araç sayısı tahmini sonuçları Tablo 20'de gösterilmiştir. Her iki konumda da bulunan bireylerin yaklaşık yarısının hanehalkına ait aracı olmadığı tahmin ediliyorken, yaklaşık yüzde 47'sinin ise hane halkına ait 1 adet binek aracı olduğu söylenebilmektedir.

Tablo 20. Konum 1 ve konum 2'ye ait hane halklarının sahip olduğu toplam binek araç sayısı tahmini sonuçları

Hanehalkının Sahip Olduğu Toplam Binek Araç Sayısı	0	1	2	3
Konum 1	178991	167463	11426	1717
(%)	49,78	46,57	3,18	0,48
Konum 2	422177	395092	26954	4082
(%)	49,77	46,57	3,18	0,48

Konumlara ait nihai simülasyon sonuçlarına ait değişken yüzdeleri ile mikro verideki değişkenlerin yüzdelerinin karşılaştırılması Şekil 18'de gösterilmiştir. Buna göre mikro verideki değişkenlerin yüzdeleriyle hem konum 1 hem de konum 2'deki değişken yüzdeleri arasındaki farkın çok küçük olduğu görülmektedir. Mikro veriden yararlanarak yapılan konumsal simülasyon sonuçlarının ne kadar iyi sonuç verdiği bu tablodan anlaşılabilir.

GREGWT geliştirilmiş regresyon ağırlıklandırma işlemine dayanan bir yöntemdir. Çalışma verisinde GREGWT yöntemiyle konumlara ait ortalama toplam gelir ve kişi başına ortalama tahmini yapmak istensin. Öncelikle GREGWT kütüphanesinin yüklenmesi gerekir. R kodu aşağıdaki gibi yazılır:

```
library('GREGWT')
yas <- read.csv2("yas.csv",header=T)
cinsiyet <-read.csv2("cinsiyet.csv",header=T)
mikroveri <- read.csv2("mikroveri.csv",header = T)
mikroveri$yas <- cut(mikroveri$yas,breaks = c(0, 49,
Inf),labels=c("X0.49.Yas","X50.ve.50..Yas"))
mikroveri$w <- vector(mode = "numeric",length=dim(mikroveri)[1])+1
data_in <- prepareData(cbind(cinsiyet,yas),mikroveri,census_area_id = F, breaks =
c(3))
fweights <- NULL
Result <- as.data.frame(matrix(NA, ncol=3, nrow=dim(yas)[1]))
names(Result) <- c("konum", "toplam.gelir", "kisibasına.gelir")
for(area in seq(dim(yas)[1])){
+ gregwt = GREGWT(data_in = data_in, area_code = area)
```

```

+ fw <- gregwt$final_weights
+ fweights <- c(fweights, fw)
+ sum.income <- sum(fw * mikroveri$gelir)
+ cap.income <- sum(fw * mikroveri$gelir / sum(fw))
+ Result[area,] <- c(area, sum.income, cap.income)
+ }

```

GREGWT ile elde edilen konumsal simülasyon sonucu aşağıda yer almaktadır:

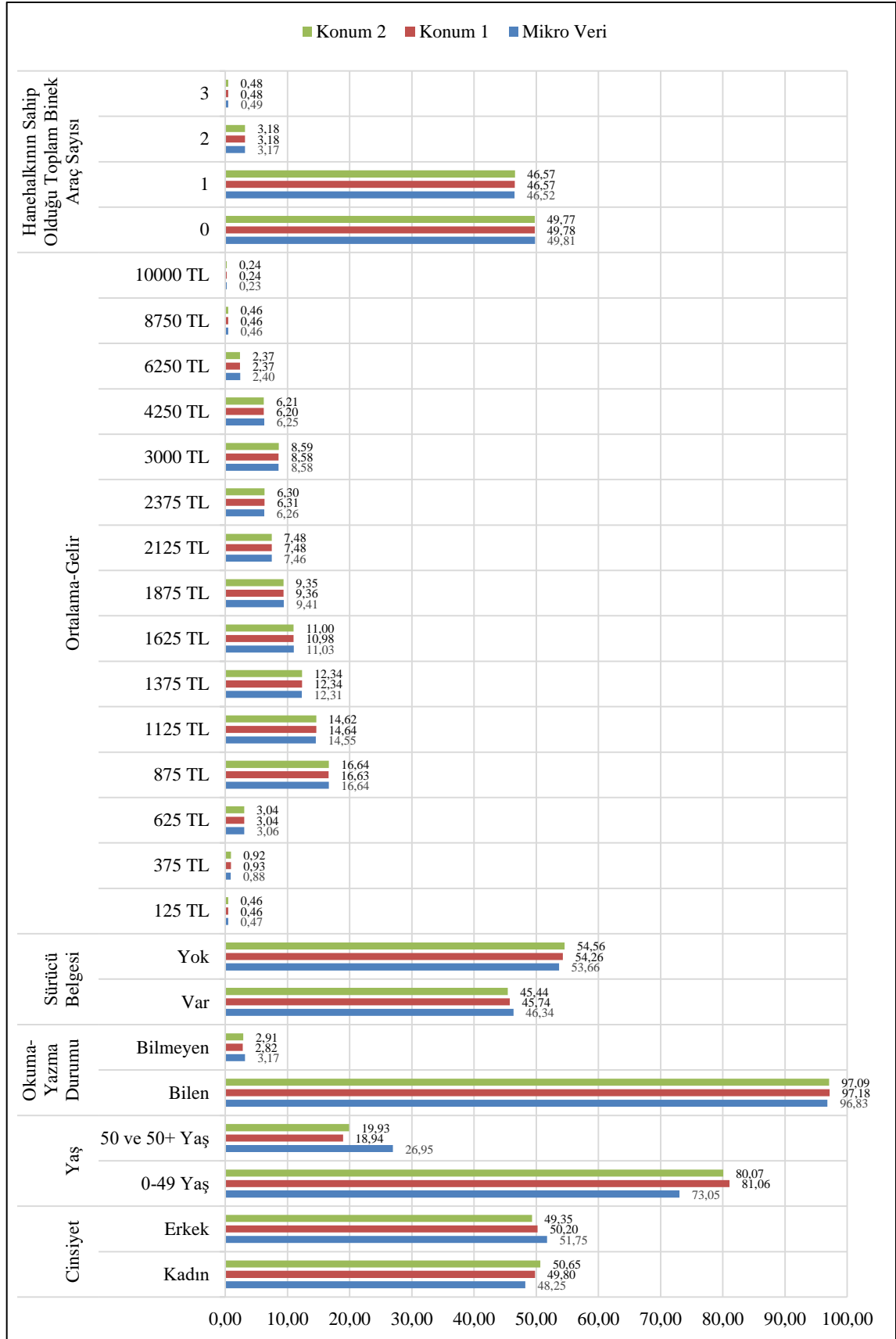
Result

konum	toplam.gelir	kisibasına.gelir
1	700403631	1899.235
2	1628416129	1899.050

Tablo 21. GREGWT ile konum 1 ve konum 2'ye ait toplam ve kişi başı ortalama gelir tahmini

Konum	Toplam Ortalama Gelir	Kişi Başı Ortalama Gelir
1	700403631	1899,235
2	1628416129	1899,050

GREGWT yöntemi kullanılarak konumlara ait toplam ortalama gelir ve kişi başı ortalama gelir tahminleri Tablo 21'de yer almaktadır. Bu tabloya göre konumlarda yer alan bir bireyin ortalama geliri 1899 TL olarak tahmin edilebilir. “ipfp” ile AOA yöntemi kullanılarak toplam ortalama gelir ve kişi başı ortalama gelir tahminlerinin verildiği tablo ile kıyaslandığında (Tablo 17), konumlarda yer alan bir bireyin ortalama geliri de 1898 TL olarak tahmin edilmişti. GREGWT ve AOA yöntemlerinden elde edilen kişi başı ortalama gelir tahminlerinin yaklaşık aynı olduğu görülmektedir. GREGWT yöntemi ile konum 1'e ait toplam ortalama gelir 700403631 TL, AOA yöntemi ile 682540000 TL olarak tahmin edilmiştir. Konum 2'ye ait toplam ortalama gelir tahminleri karşılaştırıldığında ise toplam ortalama gelirin GREGWT ile 1628416129 TL ve AOA ile 1610426250 TL olduğu görülmektedir.



Şekil 18. Mikro veri, konum 1 ve konum 2' ye ait simülasyon verilerinin yüzdelerinin karşılaştırılması

4. ÖNERİLER

İnsanođlu, ilk çağlardan bu yana geleceđe karşı ilgi duymuřtur. Uygarlık ilerledikçe yařamın karmařası artmıř ve artık gelecek hakkında belli bilgilere ulařmak meraktan ziyade ihtiyaç haline gelmiřtir. Ülkeler, bölgeler veya řehirlerde konumsal alanlara ait analiz yapma gereksinimine son yıllarda çok ihtiyaç duyulmaktadır. Günümüzde kamu ve özel kuruluşlar, iřletmeler ya da řirketler hatta bireyler birçok alanda kullanmak üzere konumsal mikro simülasyon çalışmalarını yapabilmektedirler ve buradan elde ettikleri sonuçlarla gelecekte beklenen kořullara göre bugünden önlemlerini alabilmektedirler.

Günümüzde birey tabanlı mikro simülasyon modelleri, mikro simülasyonun en üst düzeyini temsil etmektedir. Bu modeller toplu bir sistemin gelecek durumunu zaman içerisinde birkaç bireysel aktörün davranıřını taklit ederek tahmin etmektedir. Trafik talep modellemesinde, bu sistem genellikle trafik modellerinin mekânsal düzenlemesidir ve aktörler genellikle bireyler, aileler ve hanehalkları olmaktadır. Bireysel düzeyde davranıřlar analiz edildiđinde, kiřiler arasındaki ilgi çekici bađlantıları gözlemek ve modellemek mümkün olmaktadır. Örneđin, bir ailenin üyeleri bütünüyle bađımsız bir şekilde hareket edemezler. Aynı gelir kaynaklarını paylařırlar. Ulařımda birbirlerinin programlarına uygun seyahat kalıplarını ayarlamak için ya da tüm aile üyelerinin ihtiyaçlarına dayalı olarak hane mülkiyeti hakkında kararlar almak için tek bir araçta seyahat etmeyi tercih edebilirler. Ancak birey tabanlı yapıda, hem bireysel davranıřı hem de aile düzeyinde davranıřı modellemek için aile bireyleri arasındaki iliřkilerin bilinmesi gerekmektedir.

Konumsal mikro simülasyon için popülasyon üretimi, ülkemizde henüz çalışılan bir alan deđildir. Herhangi bir alana ait açık kaynaklı bir mikro veri bulmak ülkemizde řimdilerde çok mümkün olmadığı için bu konu ülkemizde henüz ilgi görmemiřtir.

Bu çalışmaya TÜBİTAK TEYDEB tarafından desteklenen 2130125 nolu "řehir İçi ve řehirlerarası Trafik řebekelerinin Bilgisayar Ortamında Modellenmesi, Simülasyonu ve Analizini Gerçekleřtirecek Yerli Bir Yazılımın Üretilmesi" projesi kapsamında gerçeđe yakın bir trafik simülasyonu oluřturmak için gerçek mikro veriden sentetik bir popülasyon üretici oluřturulması amacıyla yola çıkılmıřtır. Tezde popülasyon sentezi, mikro simülasyon ve konumsal mikro simülasyon hakkında yapılan çalışmalar incelenmiř ve

konumsal mikro simülasyonun nasıl oluşturulduğu iki konumlu bir bölgeden elde edilen konumsal mikro veri ile gösterilmiştir.

Bu tez çalışması göz önüne alındığında, konumsal mikro simülasyon konusu yeterli derecede ele alınarak, bir konumsal mikro simülasyon üretme süreci oluşturulmuştur. Ülkemize bu konuya değer katma açısından bu tezin faydalı olacağı düşünülmektedir. Bu tez ileri aşamalarda projelendirilerek ülkemiz için mikro veriler elde edildiğinde konumsal mikro simülasyon yapılması ve buradan elde edilen simülasyon verileri ile bireylerinin aktivitelerinin üretilmesi ve rotalarının planlanması gibi simülasyon çalışmalarının yapılması planlanmaktadır.



5. KAYNAKLAR

- Agresti, A., 2002. *Categorical Data Analysis*, Second Edition, John Wiley & Sons, New York, 721 s.
- Arentze, T., Timmermans, H. ve Hofman, F., 2007. Population Synthesis for Microsimulating Travel Behavior, Transportation Research Record, 11, 85–91.
- Atkinson, K. E., 1989. *An Introduction to Numerical Analysis*, John Wiley & Sons, 683 s.
- Avram, S. vd., 2012. The Distributional Effects of Fiscal Consolidation In 9 EU Countries, Social Situation Observatory Research Note 01.
- Barthelemy, J., 2014. A Parallelized Micro-Simulation Platform for Population and Mobility Behaviour-Application to Belgium, PhD thesis, University de Namur, Belgium.
- Barthelemy, J. ve Toint. P., 2015. A Stochastic and Flexible Activity Based Model for Large Population-Application to Belgium, Journal of Artificial Societies and Social Simulation, 18, 3, 15.
- Batty, M., 2005. *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*, The MIT Press, 565 s.
- Birkin, M. ve Clarke, M., 1988. Synthesis – A Synthetic Spatial Information System for Urban and Regional Analysis: Methods and Examples. Environment and Planning A, 20, 12, 1645–1671.
- Birkin, M. ve Clarke, M., 1989. The Generation of Individual and Household Incomes at The Small Area Level Using Synthesis. Regional Studies: The Journal of the Regional Studies Association, 23, 6, 535–548.
- Boyd, S. ve Vandenberghe, L., 2004. *Convex Optimization*, Cambridge University Press, Cambridge, New York, 725 s.
- Casati, D., Müller, K., Fourie, P.J., Erath, A. ve Axhausen, K.W., 2015. Synthetic Population Generation by Combining a Hierarchical, Simulation-Based Approach with Reweighting by Generalized Raking, Transportation Research Board 94th Annual Meeting, Washington, D.C.
- Castillo, E., Menéndez, J.M. ve Sánchez-Cambronero, S., 2008. Predicting Traffic Flow Using Bayesian Networks, Transport. Res. B: Meth., 42, 5, 482–509.
- Clarke, G.P., Kashti, A., McDonald, A. ve Williamson, P., 1997. Estimating Small Area Demand for Water: A New Methodology, Water and Environment Journal, 11, 3, 186-192.

- Clarke, M., Forte, P., Spowage, M. ve Wilson, A.G., 1984. A Strategic Planning Simulation Model of A District Health Service System: The In-Patient Component And Results. Third International Conference on System Science in Health Care, W. Eimeren, R. Engelbrecht ve C.D. Flagle, Berlin, Heidelberg, Springer, 949-954.
- Clarke, M. ve Wilson, A., 1985. The Dynamics of Urban Spatial Structure: The Progress of A Research Programme, Transactions of the Institute of British Geographers, 10, 4, 427–451.
- Clarke, M. ve Holm, E., 1987. Microsimulation Methods in Spatial Analysis and Planning, Geografiska Annaler Series B. Human Geography, 69, 145–164.
- Cormen, T.H., Leiserson, C.E. ve Rivest, R.L., 1990. Introduction to Algorithms. 1st edition. MIT Press, Cambridge, MA, 1028 s.
- Csisz'ar, I., 1975. I-Divergence Geometry of Probability Distributions and Minimization Problems, Annals of Probability, 3, 1, 146–159.
- Csisz'ar, I., 1998. Information Theoretic Methods in Probability and Statistics. IEEE Information Theory Society Newsletter, 1-20.
- Deming, W.E. ve Stephan, F.F., 1940. On A Least Squares Adjustment of A Sampled Frequency Table When The Expected Marginal Totals are Known, Annals of Mathematical Statistics, 11, 427-444.
- Edwards, K. ve Clarke, G., 2013. SimObesity: Combinatorial Optimisation (Deterministic) Model, Spatial Microsimulation: A Reference Guide for Users, Springer, Netherlands, 69–85.
- Farooq, B., Bierlaire, M, Hurtubia, R. ve Flötteröd, G., 2013. Simulation Based Population Synthesis, Transportation Research Part B: Methodological, 58, 243-263.
- Fienberg, S.E., 1968. The Geometry of An R X C Contingency Table, The Annals of Mathematical Statistics, 39, 1186–1190.
- Fienberg, S.E. ve Meyer, M.M. 2004. Iterative Proportional Fitting, 2nd edition, Encyclopedia of Statistical Sciences, John Wiley, New York.
- Furness, K.P., 1965. Time Function Iteration, Traffic Engineering and Control, 7, 7, 458-460.
- Gaber, J., 2007. Simulating Planning: SimCity as a Pedagogical Tool, Journal of Planning Education and Research, 27, 113–121.
- Glover, F., 1997. Tabu Search and Adaptive Memory Programming—Advances, Applications and Challenges, Springer, US, 75 s.
- Glover, F. ve Laguna, M., 2013. Tabu Search, Springer, New York, 3261-3362.

- Hagerstrand, T., 1952. The Propagation of Innovation Waves, Lund Studies in Geography, Ser. B, Human Geography; 4, 20. London: Royal University of Lund, Dept. of Geography.
- Harland, K., 2013. Microsimulation Model User Guide: Flexible Modelling Framework, National Centre for Research Methods Working Paper, 1-51.
- Hastings, W.K., 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications, Biometrika, 57, 1, 97-109.
- Heckerman, D., 1998. A Tutorial on Learning With Bayesian Networks, MIT Press, Cambridge, 301–354.
- Hermes, K. ve M. Poulsen., 2012. A Review of Current Methods to Generate Synthetic Spatial Microdata Using Reweighting and Future Directions, Computers, Environment and Urban Systems, 36, 4, 281-290.
- Hornik, K. 2012. Are There Too Many R Packages?, Austrian Journal of Statistics 41, 1, 59-66.
- Ibe, O.C. 2013. 14 - Hidden Markov Models. Markov Processes for Stochastic Modeling, Second Edition, Elsevier, Oxford, 417-451.
- Ireland, C.T. ve Kullback, S., 1968. Contingency Tables with Given Marginals, Biometrika, 55, 1, 179-188.
- Janssens, D., Wets, G., Brijs, T., Vanhoof, K., Arentze, T. ve Timmermans, H., 2006. Integrating Bayesian Networks and Decision Trees in A Sequential Rule-Based Transportation Model, Eur. J. Oper. Res., 175, 1, 16–34.
- Knudsen, D. C. ve Fotheringham. A. S., 1986. Matrix Comparison, Goodness-of-Fit and Spatial Interaction Modelling, International Regional Science Review, 10, 2, 127-147.
- Koller, D. ve Friedman, N., 2009. Probabilistic Graphical Models: Principles and Techniques, MIT press, Cambridge, 1213 s.
- Lahr, M. L. ve Mesnard, L. De., 2004. Biproportional Techniques in Input-Output Analysis: Table Updating and Structural Analysis, Economic Systems Research, 16, 2, 115-134.
- Little, R.J.A. ve Wu, M.M., 1991. Models for Contingency Tables with Known Marginals When Target and Sampled Populations Differ, Journal of the American Statistical Association, 86, 413, 87-95.
- Lovelace, R. ve D. Ballas., 2013. "Truncate, Replicate, Sample": A Method for Creating Integer Weights for Spatial Microsimulation, Computers, Environment and Urban Systems, 41, 1-11.

- Mannion, O., Lay-Yee, R., Wrapson, W. ve Davis, P., 2012. JAMSIM: A Microsimulation Modelling Policy Tool, Journal of Artificial Societies and Social Simulation, 15, 1, 8.
- Meindl, B., Templ, M., Alfons, A. ve Kowarik, A., 2015. simPop: Simulation of Synthetic Populations for Survey Data Considering Auxiliary Information, R package version 0.2, 9.
- Miller, E. J., Kriger, D. S. ve Hunt, J. D., 1999. Integrated Urban Models for Simulation of Transit and Land Use Policies: Guidelines for Implementation and Use, Transportation Research Board National Research Council, National Academy Press, Washington, D.C., 1-40.
- Mohammadian, A.K., Javanmardi, M. ve Zhang, Y., 2010. Synthetic Household Travel Survey Data Simulation, Transportation Research Part C: Emerging Technologies, 18, 6, 869–878.
- Müller, K. ve Axhausen, K.W., 2010. Population Synthesis for Microsimulation: State of The Art, Annual Meeting of The Transportation Research Board, IVT Working Paper, Swiss Federal Institute of Technology, ETH Zurich, 1-16.
- Namazi-Rad, M.R., Tanton, R., Steel, D., Mokhtarian, P. ve Das, S., 2017. An Unconstrained Statistical Matching Algorithm for Combining Individual and Household Level Geo-Specific Census and Survey Data, Computers, Environment and Urban Systems, 63, 3-14.
- Oh, H. L. ve Scheuren, F., 1987. Modified Raking Ratio Estimation, Survey Methodology, 13, 2, 209-219.
- Orcutt, G., Greenberger, M., Korbel, J. ve Rivlin, A., 1961. Microanalysis of Socioeconomic Systems: A Simulation Study, Harper and Row, New York.
- Orcutt, G. ve Glazer, A., 1980. Microanalytic Modelling and Simulation. Simulation Models: Methods and Applications. Stockholm, Industrial Institute for Economic and Social Research.
- Orcutt, G.H.G., 1957. A New Type of Socio-Economic System, The Review of Economics and Statistics, 116–123.
- Ortúzar, J.D. ve Willumsen, L.G., 2002. Modelling Transport, John Wiley & Sons, Chichester, UK.
- Pearl, J., 2000. Causality: Models, Reasoning, and Inference, Cambridge University Press, Cambridge, UK.
- Pritchard, D. R. ve Miller, E. J., 2012. Advances In Population Synthesis: Fitting Many Attributes Per Agent and Fitting to Household and Person Margins Simultaneously, Transportation 39, 3, 685–704.
- PUMS Technical Documentation., 2008. Public Use Microdata Sample - 2000 Census of Population and Housing. U.S.

- Rahman, A., 2009. Small Area Estimation Through Spatial Microsimulation Models. 2nd International Microsimulation Association Conference, Natsem University of Canberra, Canada, 1-49.
- Rao, J.N.K. ve I. Molina. 2015. Small Area Estimation, John Wiley & Sons, Canada.
- Saadi, I., Mustafa, A., Teller, J., Farooq, B. ve Cools, M. 2016. Hidden Markov Model-Based Population Synthesis, Transportation Research Part B: Methodological, 90, 1-21.
- Sebastiani, P. ve Ramoni, M., 2001. On the Use of Bayesian Networks to Analyze Survey Data, Res. Off. Stat., 4, 1, 53-64.
- Sevinç, V., 2008. Küçük Alan Tahminlerinde Adımlı Oransal Ayarlama Tekniği ile Kategorik Veri Analizi, Ekonomik ve Sosyal Araştırmalar Dergisi 4, 2, 27-43.
- Stephan, F.F., 1942. Iterative Methods of Adjusting Sample Frequency Tables When Expected Margins Are Known, The Annals of Mathematical Statistics, 13, 2, 166-178.
- Sun, L. ve Erath, A., 2015. A Bayesian Network Approach for Population Synthesis, Transportation Research Part C, 49-62.
- Sutherland, H. 1995. Static Microsimulation Models in Europe: A Survey, Cambridge Working Papers in Economics 9523, Cambridge: Faculty of Economics, University of Cambridge.
- Sutherland, H. ve Figari, F., 2013. EUROMOD: The European Union Tax-Benefit Microsimulation Model, International Journal of Microsimulation, 6, 4-26.
- Tanton, R., Vidyattama Y., Nepal, B. ve McNamara J., 2011. Small Area Estimation Using a Reweighting Algorithm, Journal of the Royal Statistical Society Series A, 174, 931-951.
- Thiele, J. C., Kurth, W. ve V. Grimm., 2014. Facilitating Parameter Estimation and Sensitivity Analysis of Agent-Based Models: A Cookbook Using Netlogo and R, Journal of Artificial Societies and Social Simulation, 17, 3, 11.
- Thionet, P., 1964. Note Sur Le Remplissage D'un Tableau À Double Entrée, J. Soc. Stat. Paris, 105, 228-247.
- Thionet, P., 1963. Sur Certaines Variantes Des Projections Du Tableau Changes Interindustriels, Bull. Inst. Internat. Stat., 40, 431-446.
- Thionet, P., 1961. Sur Le Remplissage D'un Tableau À Double Entrée, J. Soc. Stat. Paris, 102, 331-345.
- Tomintz, M.N.M., Clarke, G.P. ve Rigby, J.E.J., 2008. The Geography of Smoking in Leeds: Estimating Individual Smoking Rates and The Implications for The Location of Stop Smoking Services, Area 40, 341-353.

- Transportation Research Board. 2007. Metropolitan Travel Forecasting: Current Practice, Special Report 288, Washington, D.C.
- Upton, G. J. G. ve Fingleton, B. 1989. Spatial Data Analysis by Example, John Wiley & Sons Ltd., Chichester.
- White, K. ve Mesenbourg, T.L., 2009. A Compass for Understanding and Using American Community Survey Data - What PUMS Data Users Need to Know?, United States: U.S. Census Bureau.
- Williamson, P., Birkin, M. ve Rees, P.H., 1998. The Estimation of Population Microdata by Using Data from Small Area Statistics and Samples of Anonymised Records, Environment and Planning A, 30, 5, 785-816.
- Wilson, A.G. ve Pownall, C.E., 1976. A New Representation of the Urban System for Modelling and for the Study of Micro-Level Interdependence, Area, 8, 246-254.
- Xie, C. ve Waller, S.T., 2010. Estimation and Application of A Bayesian Network Model For Discrete Travel Choice Analysis, Transp. Lett.: Int. J. Transport. Res., 2, 125–144.
- Ye, X., Konduri, K., Pendyala, R. M., Sana, B. ve Waddell, P., 2009. A Methodology to Match Distributions of Both Household and Person Attributes in The Generation of Synthetic Populations, 88 th Annual Meeting of the Transportation Research Board, 11-15.
- Zaidi, A., Harding, A. ve Williamson, P., 2009. New Frontiers in Microsimulation Modelling: Introduction. New frontiers in Microsimulation Modelling, 31-50. Vienna, Ashgate.
- Zhang, K. ve Taylor, M.A.P., 2006. Effective Arterial Road Incident Detection: A Bayesian Network Based Algorithm, Transport. Res. C: Emerg. Technol., 14, 6, 403–417.

ÖZGEÇMİŞ

Sevim İREY ZOBBA, 26 Şubat 1990 tarihinde Konya Ereğli'de doğdu. İlköğrenimini Öğretmen Abdurrahim İlköğretim Okulu'nda, ortaöğrenimini ise Konya Ereğli Anadolu Lisesi'nde tamamladı. 2009 yılında Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümüne yerleşti. 2012 yılında T.C. Kalkınma Bakanlığı, Türkiye Cumhuriyet Merkez Bankası ve Sermaye Piyasası Kurulu'nda staj eğitimleri aldı. 2013 yılında Hacettepe Üniversitesi İstatistik Bölümünden Yüksek Şeref Öğrencisi olarak mezun oldu. Aynı yıl Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim dalında tezli yüksek lisans programına başladı. 2014 yılında TOBB TEPAV / Türkiye Ekonomi Politikaları Araştırma Vakfı'nda bir yıl araştırmacı ünvanıyla çalıştı. Şubat 2015 tarihinde Karadeniz Teknik Üniversitesi İstatistik ve Bilgisayar Bilimleri Bölümü'ne Araştırma Görevlisi olarak atandı. Mart 2015 tarihinde Hacettepe Üniversitesi Yabancı Diller Yüksekokulunda altı aylık ÖYP kapsamında gittiği İngilizce dil eğitimi kursunu tamamladı. Eylül 2015'te Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik ve Bilgisayar Bilimleri Anabilim dalında tezli yüksek lisans programına geçiş yaptı. Şubat 2015 tarihinde Karadeniz Teknik Üniversitesi başladığı görevine halen devam etmektedir. Ulusal ve uluslararası birden fazla sempozyum, kongre ve konferansa katıldı. Tez konusu ile alakalı uluslararası konferansta yayınlanmış bildirisi bulunmaktadır.