

KUÇUK ALAN TAHMİN TEKNİKLERİ VE
BİRLEŞİK SENTETİK TAHMİN EDİCİLER

YÜKSEK LİSANS TEZİ
(İSTATİSTİK)

Metin KABALCI

HAZİRAN 1993

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İÇİNDEKİLER

	<u>SAYFA NO :</u>
Ö Z E T.....	V
ABSTRACT.....	VI
TEŞEKKÜR.....	VII
G İ R İ Ş.....	1
BÖLÜM 1	
KÜÇÜK ALAN TAHMİNLERİ.....	4
1.1. Simptomatik Hesaplama Teknikleri..	7
1.2. Sentetik Tahmin Teknikleri.....	8
1.3. Regresyon Simptomatik Teknikler.....	14
1.4. Örnek Regresyon Teknikleri.....	18
1.5. Sentetik Regresyon Teknikleri.....	19
1.6. Temel Birimler Tekniği.....	21

BÖLÜM 2

BİRLEŞİK SENTETİK TAHMİN EDİCİLER.....	23
2.1. Optimum Birleşik Sentetik Tahmin Ediciler.....	23
2.2. Yığın Ortalaması Ve Varyansına Da- yanarak Yapılan Birleşik Sentetik Tahmin Ediciler.....	34
2.3. Örnek Çapının Bir Fonksiyonuyla Ağır- lıklandırılmış Birleşik Sentetik Tah- min Ediciler.....	36
2.4. Hata Kareler Ortalamasına Dayanan Birleşik Sentetik Tahmin Ediciler....	37
2.5. Birleşik Sentetik Tahmin Edicilerin Ağırlıklarının Tahmini.....	39
2.6. Birleşik Sentetik Tahmin Edicilerin Karşılaştırılması İçin, Risk Etkinlik	

Fonksiyonu.....	40
BÖLÜM 3	
UYGULAMA.....	43
3.1. C (1) Birleşik Sentetik Tahmin Edici i	
İçin Uygulama.....	46
3.2. C (2) Birleşik Sentetik Tahmin Edici i	
İçin Uygulama.....	49
3.3. C (3) Birleşik Sentetik Tahmin Edici i	
İçin Uygulama.....	51
3.4. C (4) Birleşik Sentetik Tahmin Edici i	
İçin Uygulama.....	54
3.5. Sentetik Birleşik Tahmin Edicilerin	
Karşılaştırılması.....	56
BÖLÜM 4	
SONUÇ VE TARTIŞMA.....	59

KAYNAKLAR.....

63

ÖZGEÇMİŞ

KÜÇÜK ALAN TAHMİN TEKNİKLERİ VE
BİRLEŞİK SENTETİK TAHMİN EDİCİLER

(Yüksek Lisans Tezi)

Metin KABALCI
GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
HAZİRAN 1993

Ö Z E T

Bu çalışmanın birinci bölümünde, mevcut küçük alan tahmin tekniklerinin dayandığı varsayımlar ile gerekli formülasyonlar verilerek tanıtılmaktadır. İkinci bölümde, küçük alan tahmin tekniklerinden olan birleşik sentetik tahmin edicilerin dayandığı temel varsayımlar açıklanmıştır. Ayrıca, dört ayrı birleşik sentetik tahmin edici tanıtılarak bir tanesinin teorik çıkarsaması yapılmıştır. Son bölümde ise, bir nüfus araştırması için derlenmiş verilere dört ayrı birleşik sentetik tahmin edici uygulanarak tahminler yapılmış ve bu veriler için hangi birleşik sentetik tahmin edicinin uygun olduğu gösterilmiştir.

SMALL AREA ESTIMATION AND
COMPOSITE SYNTHETIC ESTIMATORS

(M.Sc.Thesis)

Mezin KABALCI

GAZI UNIVERSITY

INSTITUTE OF SCIENCE AND TECHNOLOGY

JUNE 1993

ABSTRACT

In the first chapter of this study, small area techniques and necessary formulas have been introduced. In the second chapter the basic assumptions of composite synthetic estimators which are methods of small area techniques have been explained. Introducing four different composite synthetic estimators one of them has been discussed theoretically. In the last chapter, by using census data, four different composite synthetic estimators have been applied and the one which is better suited for census data has been determined.

TEŞEKKÜR

Yüksek lisans tez çalışmam boyunca, böyle değişik bir konuda beni yönlendiren ve yardımlarını esirgemeyen tez danışmanım Doç.Dr. Müslim EKİNİ'ye, çevirilerimde yardımlarını unutamayacağım arkadaşım Ula PAÇALI'ya, her türlü manevi desteği sağlayan sevgili eşim Hacer KABALCI ve Genel Müdürüm Dr. Mustafa Tören YÜCEL'e sonsuz saygı ve teşekkürlerimi sunarım.

G İ R İ Ő

Bazı arařtırmaların uzun zaman ve maliyet gerektirmesi ve yığının alt grupları için bilgi gereksinimi nedeniyle küçük alan tahminlerine ilgi duyulmaktadır.

Küçük alan tahmin teknikleri yanlış tahminler üretmesine rağmen, zaman ve maliyet açısından avantajlı olması, uygulanmasının kolaylığı ve büyük arařtırmalar yapmadan yığının alt grupları ile ilgili ön bilgiler vermesi nedeniyle tercih edilmektedir.

Yığının büyük alt grupları ve büyük coğrafik alanlarından sağlık, sosyal ve ekonomik analizler için kullanılabilir veriler elde edilebilir. Sağlık planlamaları ve hükümet çalışmaları için güncel verilere gereksinim duyulmaktadır. Bu nedenle önceden yapılmış arařtırma sonuçları ve değerlendirilmesi uzun zaman gerektiren çalışmalar bu tür planlamalarda kullanılamamaktadır (1).

Mevcut küçük alan tahmin teknikleri birbirinden bağımsız olarak pek çok bölüme ayrılır. Bu tür tekniklerin temel metodolojisi veri gereksinimleri ve sınırlamaları ile birlikte ele alınmaktadır ve daha fazla bilgi sunma potansiyeline göre değerlendirilmektedir.

Küçük alanlar için yapılan tahminler, uzun zamandır

bir ilgi kaynağı olmuşsa da nüfus sayımları ve idari kayıt araştırmaları dışındaki alanlarda çok fazla uygulanamamıştır. Küçük alanlar için daha zengin, daha ayrıntılı, daha güncel veriye olan talebin gitikçe artması dolayısıyla bu ilgiler daha fazla alana yayılma eğilimi göstermektedir. Özellikle son yıllarda Amerika Birleşik Devletlerinde reform planlamaları, refah çalışmaları ve uygulamaları konularında bu tür teknikler uygulanmaktadır (1).

Örnekleme ve istatistik teorikiler, uzun bir süre küçük alanlar için yapılan tahminleri görmemezlikten gelmişlerdir. Ancak, nüfus bilimciler nüfus sayımları için küçük alan tahminleri ile ilgilenerek, bu konuda değişik teknikler geliştirmişlerdir. Bu nedenle küçük alan tahminleri, özellikle nüfus araştırmaları konusunda uzmanlaşmıştır. İstatistiksel teoriler genellikle genel ortalamanın tahmini ile ilgilenmekte ise de, bireysel olaylar için küçük alan tahminleri ile ilgilenen istatistiksel teoriler de bulunmaktadır. Ancak, istatistiksel teorilerde küçük alan tahminleri üzerinde pek durulmamıştır (1).

Son yıllarda, küçük alanlar için yapılan tahmin teknikleri sayesinde, bu tür tekniklerin gelişmesi sağlanmıştır. Regresyona dayanan tekniklerden farklı olarak, bayes ve deneysel bayes tekniklerine dayanan süper yığın

(Super Population)tahmin teorisi, birleřtirici teknikler ve kategorik veri analizleri yöntemleri halen arařtırma konusu olarak güncelliğini korumaktadır. Özellikle, kategorik veri analizleri yöntemi soruna, tahmin edici oranlarının belli olduđu mantıksal ve yapısal bir yaklaşım sunduđu için, arařtırmacılara özel bir ilgi konusu olmuřtur (1).

Küçük alan tahmin tekniklerinden olan birleřik sentetik tahmin ediciler, hem örnek bilgisini, hem de yığın bilgisini kullanması nedeniyle uygulamacılar tarafından tercih edilmektedir.

B Ö L Ü M 1

KÜÇÜK ALAN TAHMİN TEKNİKLERİ

Küçük alan tahminlerine, küçük alanların birbirlerine göre farklı olduğu varsayımına dayandığı için gereksinme duyulmaktadır. Küçük alanların tipi, uygulanacak tekniğin belirlenmesinde en önemli faktörlerden biridir. Bu nedenle küçük alanların tipleri genellikle aşağıdaki gibi sıralanır (2).

- i) Planlanmış alanlar. Bu alanlar birbirlerinden bağımsız olarak seçilerek araştırma için örnek uzayı oluşturulur.
- ii) Çapraz sınıflandırılmış alanlar. Örneği ikiye ayırarak örnek uzayı oluşturulur (yaş, cinsiyet gibi).
- iii) İki uç arasında ve daha az kullanılan sınıflama da örneğin bölümlerini birbirinden ayırmayarak, ana üniteler üzerinde yoğunlaşmaya eğilimli bir sınıflama yapılır.

Ancak, küçük alanların büyüklüğü (çapı) örneğin seçiminde bir faktör olarak ortaya çıkar. Bu nedenle küçük alanların büyüklüğüne dayanan sınıfların seçimine ilişkin

ölçütler aşağıdaki gibidir.

i) Küçük alanların büyüklüğü, yığının 0,10'unu kapsıyorsa bu tür alanlar, büyük dereceli alanlardır.

ii) Küçük alanların büyüklüğü, yığının 0,10'u ile 0.01 arasında ise, bu tür alanlar orta dereceli alanlardır.

iii) Küçük alanların büyüklüğü, yığının 0,01'i ile 0,001'i arasında ise, bu tür alanlar küçük dereceli alanlardır.

Yukarıda iki farklı ölçüte göre yapılan sınıflamanın sınırı çok önemli değildir. Bu sınırlar, örneğin çapına, toplamına, değişkenlere ve istatistiğe bağlı olmakla beraber uygulamacının düşünce ve öngörülerine göre belirlenebilir. Ancak, bu tür sınırlamalar araştırmacıları bazı yanlışlıklardan kurtarır (1,2).

Örnek araştırmalarındaki olabilirlik metotları, bir takım standart tahminler üretmiş olmasına rağmen, bu tahminler belli bir eğilim taşımamaktadır. Küçük alan tahmin teknikleri veri gereksinimlerini sayımlar, idari kayıtlar veya örnek araştırmalarından karşılamaktadırlar. Farklı tekniklerin veri kaynakları aşağıdaki tabloda gösterilmiştir (1).

Tablo I. Küçük Alan Tahmin Tekniklerinin Kullandığı
Veri Kaynakları

Küçük Alan Tahmin Tekniği	Veri Kaynakları		
	Sayımlar	Kayıtlar	Örnek
Simptomatik	X	X	
Sentetik	X		X
Regresyon-Simptomatik	X	X	
Örnek-Regresyon	X	X	X
Sentetik-Regresyon	X	X	X
Temel Birimler	X	X	X

Nüfus sayımlarından elde edilen veriler, ayrıntılı olma özelliği taşımasına karşın, sonuçlarının alınması uzun zaman gerektirmesi nedeni ile güncelliğini yitirmektedir. Ayrıca, kayıtlardan elde edilen veriler, ayrıntılı ve kısa sürede sonuç verici olmasına rağmen, ilgilenilen değişkenlerin tesbiti ve ölçülmesindeki sakıncalarından dolayı kullanışlı değildir. Örnek verileri ise, ihtiyaçlara uyabilmesine rağmen, ayrıntılara cevap vermez. Yalnızca örnek verilerini kullanan Hansen, Hurwitz ve Madow tarafından öne sürülen çifte örnekleme regresyonu tekniği vardır

ve bu teknik 6rnekleme regresyonu olarak bilinen tekniğin ilk halidir (1). Ayrıca, Bayes ve Deneysel Bayes yaklaşımları 6rnek verileri uygun olduėunda k6ç6k alan tahminlerinde kullanılmaktadır.

1.1. Simptomatik Hesaplama Teknikleri.

K6ç6k alan tahmin t6kniklerinin en eskilerinden birisi simptomatik tekniklerdir. Son yıllarda, elde edilen verilere dayanan istatistiksel iliřkilerle ilgili olarak, n6fus arařtırmalarında simptomatik hesaplama teknikleri kullanılmaktadır. Amerika Birleřik Devletlerinde, en son yapılan y6resel n6fus tahmini arařtırmasında bu teknikler kullanılmıřtır. Genel n6fus sayımı ile ilgili matematiksel eřitlikler, genellikle doėum, 6l6m yada g6çleri tahmin eden eřitliklerdir.

Simptomatik hesaplama teknikleri, geniř anlamda iki sınırlaması olan y6ntemlere dayanmaktadır. İlk sınırlama olarak doėum ve 6l6m gibi verilerin iyi yapılmıř kayıtlarına dayanırken, ikinci sınırlama olarak ta simptomatik deėiřkenlerin karřılıklı iliřkili olduėu (Gelir-Gider, 6retim-T6ketim gibi) n6fus sayımlarına dayanmaktadır.

Simptomatik hesaplama tekniklerinden birisi de Bogue tarafından geliřtirilen 6nemli oranlar tekniėidir (1,3). Doėum ve 6l6m oranlarını tahmin etmekte kullanılan bu teknik,

küçük alan doğum ve ölüm oranlarının, büyük alanlardaki oranlarla aynı ölçüde değiştiği varsayımına dayanmaktadır. Bu tekniğe alternatif olarak Bogue ve Duncan tarafından birleşik yöntem geliştirilmiştir (4). Bu yöntem küçük alan nüfusunu ayrı, ayrı yaş gruplarına bölmek suretiyle oluşturulan örnek yardımı ile tahmin yapmaktadır.

Büyük şehir ve metropoller için, nüfus tahminini elde etmek için kullanılan bir başka yöntem de, hanehalkı tahmin yöntemidir. Bu yöntem küçük alanlardaki hane halkı nüfusunun tahminini yaparak, her bir hane için ortalama birey sayısını tahmin etmektedir. Bu yöntem, hanehalkı sayısının değişmesi ile nüfusun değiştiği mantığını kullanmaktadır.

Simptomatik hesaplama tekniklerinin en yenisi idari veri kayıtları yöntemidir. Amerika Nüfus İşleri Merkezi (U.S. Bureau of the Census) tarafından geliştirilen bu yöntemin temel özelliği idari kayıtların güncel olarak yenilenmesidir. Bu yöntemde verilerin sınıflandırılması yerine bireysel kayıtların kullanılmasıyla, küçük alanlar için tahmin yapılmasına imkan tanınmaktadır (5).

1.2. Sentetik Tahmin Teknikleri.

Sentetik tahmin yöntemi, yığının farklı alt gruplarından, ilgilenilen değişkenin bazı yüksek seviyelerinden

alınan örnek verilerinden tahminler yaparak, bu tahminleri her küçük alan için oransal olarak ölçümlemektedir.Örneğin, A.B.D.inde devlet tarafından yapılan işsizlik tahminleri, yaş, cinsiyet ve ırk gibi çapraz tablolanmış veriler,bölgesel işsizliği tahmin etmek için her bir bölgenin işsizlik oranına göre ölçümlendiğinde ve devletin her bir bölge için işsizlik oranları ile karşılaştırıldığında, fark küçük ise, tahminler doğru olacaktır.

Bu yaklaşıma, sentetik ismi N.C.H.S. (Ulusal Sağlık İstatistikleri Merkezi) kuruluşu tarafından verilmiştir(6). Bu kuruluş tarafından, ulusal sağlık araştırmalarına dayanan, devletin fiziksel yetersizliklerinin tahminini hesaplamak için sentetik tahminler kullanılmıştır. Bu tahmin metoduna sentetik ismi, tahminlerin araştırmalardan türetilmediği, yani dolaylı yöntemlere dayandığı için verilmiştir. Ancak bu terim günümüzde daha çok ortaya çıkan tahminlerin doğruluğunu artırmak için, benzer küçük alanlardan bilgi alınması yöntemlerine dayanmaktadır.

Son zamanlarda sentetik tahminler, küçük alan karakteristiklerini tahmin etmek için, bir çok araştırmayla bağlantılı olarak kullanılmaktadır. A.B.D. Nüfus Sayımları Bürosu tarafından yürütülen çalışmalar ve 1970 Genel Nüfus Sayımlarından elde edilen veriler, Gonzales tarafından re-

vize edilerek yayınlanmıştır (7). Sentetik tahminlerin hatalarının ölçümü, Gonzales ve Waksberg tarafından tartışılmıştır (8). Schaible, Brock ve Schnack Texas bölgesindeki işsizlik oranının sentetik tahminin hata kareler ortalamalarını karşılaştıran bir çalışma yapmışlardır(9). Purcell ve Linacre, Avusturalya Nüfus Sayım Bürosu tarafından yürütülen iki deneysel çalışmaya dayanarak gelir ve iş gücünün sentetik tahminini üreten çalışmalar yapmışlardır (10).

Sentetik tahminler, kamu sağlığı alanlarında da kullanılmaktadır. Levy, resmi ölüm kayıtlarını esas alarak, dört farklı nedenden meydana gelen ölüm olaylarının, küçük alanlardaki gelişimini izlemek üzere sentetik tahminleri kullanmıştır(11). A.B.D.inde Kısmi iş kaybı oranı Namekata, Levy ve O'Rourke tarafından, nüfus sayımından elde edilen verilerle sentetik tahmin teknikleriyle hesaplanmıştır(12).

Sentetik tahmin yöntemi şu şekilde formüle edilir; bir dizi küçük alan içinde bir X karakteristiği tahmin edilmek istenilsin, bu X karakteristiği çapraz olarak sınıflandırılır ve gruplar birbiri ile kesişmeyecek şekilde, yığının alt gruplarından alınmış ayrıntılı verilerle tahmin yapılır.

h.küçük alan için X karakteristiğinin toplamının sentetik tahmini aşağıdaki gibidir.

$$\hat{X}_h = \sum_{hg} \hat{X}_{hg} = \sum_{hg} (Y_{hg} / Y_{.g}) X'_{.g} \quad (1.1)$$

formülde;

X_{hg} : X 'in h . küçük alandaki değeri

$X'_{.g}$: X 'in örnek toplamı

Y_{hg} : Y 'nin h . küçük alandaki değeri

$Y_{.g}$: Y 'nin örnek toplamı

(1.1) ile verilen sentetik tahmin edici, g grupları için basit oransal tahmin edicinin bir türüdür ve bu tahmin edici büyük alanlar için de kullanılmaktadır. Genellikle yardımcı Y değişkeni, yığında bulunan birim sayısı olduğunda, $Y_{hg} = N_{hg}$ şeklindeki eşitlik yazılarak (1.1) ile verilen tahmin edici,

$$\hat{X}_h = \sum_{hg} (N_{hg} / N_{.g}) X'_{.g}$$

olarak yeniden formüle edilir.

h . küçük alanların toplamı (1.1)'e dayanarak Gonzales tarafından aşağıdaki gibi önerilmiştir (13).

$$\sum_h \hat{X}_h = \sum_h \sum_g (Y_{hg} / Y_{.g}) X'_{.g}$$

$$\hat{\Sigma X_h} = \Sigma_g \Sigma_h (Y_{hg} / Y_{.g}) X'_h$$

burada, $\Sigma_h (Y_{hg} / Y_{.g}) = 1$ olduğu gözönüne alınırsa,

$$= \Sigma_g X'_g = X' \quad \text{elde edilir.}$$

h . küçük alandaki, X karakteristiğinin ortalamasının tahmini aşağıdaki formülle bulunur.

$$\bar{X}_h = \Sigma_g (Y_{hg} / Y_{.g}) \bar{X}'_g \quad (1.2)$$

(1.1) ve (1.2) ile verilen tahmin edicilerin her ikisi de yardımcı değişkenin oranını belirlemekte Y_{hg} 'yi kullanmaktadır ama, (1.2) küçük alanlar içinde, alt grubların oranlarını kullanırken, (1.1) küçük alanların oranlarını kullanmaktadır.

Sentetik tahmin ediciler, iki nedenle varyansı azaltarak yanlış tahminler yapmaktadırlar. Bu nedenlerden birincisi, temel 'homojenlik' varsayımlarından sapmaların olması ve ikinci neden ise, $Y_{hg} / Y_{.g}$ oranının eski verilere dayanması sebebiyle verilerin yapısının değişmiş olmasıdır.

Sentetik tahmin edicilerin, $X_{hg} / Y_{.g} = X_{hg} / Y_{hg}$ eşitliği sağlanmadıkça yanlış tahminler yaptığını görmek için, aşağıdaki ifade yazılmaktadır.

$$E(\hat{X}_h - X_h) = E(\Sigma_g (Y_{hg} / Y_{.g}) X'_g - \Sigma_g X'_g)$$

$$E(\hat{X}_h - X_h) = \sum_g Y_{hg} \left(\frac{X_{hg}}{Y_{hg}} - \frac{X_{hg}}{Y_{hg}} \right)$$

Sentetik tahmin edicilerin değerlendirilmesi yanlışlık nedeniyle oldukça zordur. Bu nedenle hata kareler ortalamasının incelenmesi gerekir, ancak, X_{hg} 'nin gerçek değerleri hakkında bilgi eksikliğinden dolayı, hata kareler ortalamasını da tahmin etmek oldukça güçtür. Bu problem, bazı yansız örnek tahminlerinin (geniş örnek çeşitliliğine sahip ol-salar bile) oluşturulabildiği durumlarda, teorik olarak çö-zülebilir. Ancak, küçük alan varyans tahminlerinin sabit kalma olasılığı yüksektir. Mümkün bir çözüm olarak, küçük alanın hata kareler ortalamasının aritmetik ortalamasının kullanılması Gonzales ve Waksberg tarafından öneril-miştir (8). Önerilen hata kareler ortalaması aşağıdaki gi-bidir.

$$\sum_h E(\hat{X}_h - X_h)^2 / H$$

Burada H küçük alanların sayısıdır. Bu hata kareler ortalaması bazı sınırlı varsayımlara dayanmaktadır. X_{hg} 'nin yansız tahminine dayanmayan ve nüfus sayımlarından elde edilen X değişkeninin gelişiminin kullanılması değişik bir yaklaşım olarak Ghangurde ve Singh tarafından öneril-miştir (14).

Gruplanan değişkenler ile ilgilenilen değişkenler ara-

sındaki ilişki yüksek ise, sentetik tahminlerin büyük alanlar için ortalamaya yakın olması, bu tahminlerin en önemli güçlüklerindendir. Küçük alanlar için muhtelif değerlerle sentetik tahminler yetersiz kalır. Ancak, değişkenlerin gruplaması dikkatli bir şekilde yapıldığında yeterlilik sağlanır. Bu yöntemin en önemli avantajı hesaplamanın kolaylığıdır. Kamuoyu araştırmaları için, sentetik tahmin tekniği oldukça geçerlidir. Bu yaklaşım Cohen tarafından da ele alınmıştır (15).

1.3. Regresyon Simptomatik Teknikler.

Regresyon Simptomatik tekniklerin diğer tekniklerden ayırıcı özelliği simptomatik değişkenler ile ilgilenilen değişkenler arasındaki fonksiyonel ilişkinin uygunluğuna dayanmasıdır.

Çoklu regresyon modeli, Schmitt ve Crosetti tarafından oran-korelasyon metodunun yeniden düzenlenmesinden sonra küçük alan tahmini yapılmasında önemli bir araç haline gelmiştir (16). İlgilenilen değişken ile simptomatik değişkenler arasındaki benzer ilişkinin hesaplanabileceği varsayımına dayanan oran-korelasyon tekniği ilk olarak Snow tarafından önerilmiştir (17).

Bu teknik aşağıdaki adımların izlenmesiyle formüle edilir,

i) Her h. küçük alanın toplamının oranı Y_i için p
simptomatik değişkeni t=1,2 zamanları için hesap-
landıktan sonra t=3 zamanı için hesaplanır (1. ve
2. zamanlar, son iki genel sayım yıllarını, 3.zaman
ise 1.ve 2. zaman arasında hesaplanmak istenen za-
manı gösterir). Simptomatik değişken aşağıdaki gibi
hesaplanır.

$$P_{hti} = Y_{hti} / \sum_h Y_{hti}$$

Formülde,

P_{hti} : t zamanında h küçük alanındaki i.simptomatik
değişkenin oranı.

Y_{hti} : t zamanında h küçük alanındaki i.simptomatik
değişkenin değeri.

ii) Benzer olarak X değişkeni ile 1. ve 2. zamanlar
için aşağıdaki gibi hesaplanır.

$$q_{ht} = X_{ht} / \sum_h X_{ht}$$

iii) Hesaplanan P_{hti} ve q_{ht} yardımıyla aşağıdaki oranlar
hesaplanır.

$$\tau_{hai} = P_{h2i} / P_{h1i}$$

$$R_{ha} = q_{h2} / q_{h1} \quad i = 1, 2, \dots, P. \quad (1.3)$$

Simptomatik deęişkenlerin 2.ve 3. zamanları arasındaki deęişimi için,

$$\tau_{hbi} = P_{h3i} / P_{h2i} \quad (i = 1,2,...,p) \quad (1.4)$$

eşitlięi yazılır.

iv) Çoklu regresyon kullanılarak, fonksiyonel model (1.3)'de verilen oransal deęişkenler kullanılarak,

$$R_{ha} = \beta_0 + \beta_1 \tau_{ha1} + \dots + \beta_p \tau_{hap}$$

şeklinde ifade edilir.

Formülde, $\beta_0, \beta_1, \dots, \beta_p$ regrasyon katsayıları olup, iki temel periyot arasındaki deęişimler yardımıyla hesaplanır.

v) $\beta_0, \beta_1, \dots, \beta_p$ regresyon katsayıları, deneysel ilişkilere dayanır, periyotlar içi oranlardan gözlenen katsayılar, periyotlar içi oranlarla kullanılır.

(1.4) ile verilen simptomatik deęişkenler için, periyotlar arası dönemin oransal tahmin edicisi,

$$R_{hb} = \beta_0 + \beta_1 \tau_{hbi} + \dots + \beta_p \tau_{hbi}$$

olarak yazılır.

vi) R_{hb} ifadesi, küçük alan oranları ile çarpılarak aşağıdaki gibi formüle edilerek kullanılır.

$$\hat{X}_{h3} = R_{hb} \left(\frac{\sum_{h=1}^H X_{h2}}{\sum_{h=1}^H X_{h2}} \right) \sum_{h=1}^H X_{h3}$$

Elde edilen $\hat{X}_{h3}$ tahmin edicisi Regresyon Simptomatik tahmin edici olup, en önemli varsayımı, nüfus sayımı periyotları içinde bağımlı ve bağımsız değişkenler arasında gözlenen istatistiksel ilişkinin, nüfus sayımı sonrası da devam edeceğidir. Bunun yeterliliği ise, çoklu korelasyonun büyüklüğüne bağlıdır.

Küçük alan tahmin metotları diğer metotlara göre simptomatik bilgiye daha fazla gereksinim duyar. Namboodiri, teorik problemleri, oransal korelasyon metodu ile birlikte ortaya koyarak, regresyon tahminleri sonuçlarının ortalaması için bir yöntem önermiştir (18). Her bağımsız tabakanın farklı ilişkilerinin tahminlerine dayanan, oran korelasyon metodu için bir yaklaşım, Rosenberg tarafından önerilmiştir (19). Pursell, bir simptomatik değişkene, bir yapay değişken ekleyerek benzer bir çalışma yapmıştır (20).

Oran korelasyon metodu, genellikle nüfus tahminleri için kullanılmakta ise de, diğer istatistiksel tahminlerde de kullanılmaktadır. Bu metodun değişik bir yaklaşımı O'Hare tarafından fark korelasyonu adı altında sunulmuştur (21). Bu yöntemde, (1.3) ve (1.4) ile verilen zaman noktala-

rının payları arasındaki fark, burada oran olarak kullanılmıştır. Fark korelasyonu aşağıdaki gibi formüle edilir.

$$d_{hai} = p_{h2i} - p_{h1i}$$

$$d_{hbi} = p_{h3i} - p_{h2i}$$

$$D_{ha} = q_{h2} - q_{h1}$$

Bu yöntem, Çoklu regresyon eşitliğindeki r_{hai} , r_{hbi} ve R_{ha} ifadeleri yerine d_{hi} 'yi kullanmaktadır.

1.4. Örnek Regresyon Teknikleri

Örnek Regresyon metodu, her küçük alan için hesaplanan bağımsız simptomatik değişkeni kullanan regresyon modeline dayanır. Bu yöntem, modeldeki parametreleri nüfus sayımı sonrası elde edilen güncel verilerle belirler. Örnek Regresyon modeli, Oran-Korelasyon yöntemindeki gibi sağlanması zor varsayımlara dayanmadığı için daha kullanışlıdır. Hensen yöresel nüfus tahminini yapmak için 1953 yılında bu yöntemi kısmen kullanmıştır (2).

Örnek Regresyon yöntemi, Oran Korelasyon metoduna benzer olarak formüle edilir. Ancak bu yöntemde hesaplamalar örnek verileri üzerinden yapılır.

(1.4) ile verilen oranlar hesaplanarak, genel nüfus sayımı sonrası tahmin çoklu regresyon kullanılarak,

$$\hat{R}_{hb} = \beta_0 + \beta_1 \tau_{hb1} + \dots + \beta_p \tau_{hbp} \quad (1.5)$$

şeklinde formüle edilir.

Güncel tahminler yapmak için, küçük alan simptomatik değişkenlerini (1.5) eşitliğinde yerine konur. Örnek Regresyon yöntemi, klasik regresyon modeline göre en önemli avantajı, yığının ilgilenilen karakteristiklerinde meydana gelen değişikliklerden etkilenmemesidir. Fakat, örnek verilerine dayanması nedeniyle de, örnekleme hatası içereceğinden bilgi eksikliği ve yanlışlık ortaya çıkar.

Örnek Regresyon yönteminin Ericksen tarafından önerilen hata kareler ortalaması,

$$HKO(\hat{R}_{hb}) = ((n-p-1) \sigma_u^2/n) + ((p+1) \sigma_v^2/n) \quad (1.6)$$

ile ifade edilmektedir (22).

Formülde, n : örnek sayısı p : simptomatik değişken sayısı

σ_u^2 : Açıklanamayan varyans

σ_v^2 : Açıklanabilen varyans

1.5. Sentetik Regresyon Teknikleri.

Bölüm 1.2.'de anlatılan sentetik tahmin yaklaşımı, kü-

gük alanlardaki ortak deęişkenlerdeki zamana baęlı olarak meydana gelen deęişiklikleri gözönünde bulundurmaz. Bu durumu ortadan kaldırmak için Sentetik Regresyon taknięi önerilmiştir. Bu teknik regresyon ilişkisine simptomatik deęişkenler eklenerek Levy tarafından aşığıdaki gibi formüle edilmiştir (1,11).

$$X_h^* = \alpha + \beta Y_h + \epsilon_h \quad (1.7)$$

Burada,

$$X_h^* = \frac{\hat{X}_h - X_h}{\hat{X}_h} 100$$

$\hat{X}_h$: X_h 'in sentetik tahmin edicisi,

X_h : X_h 'in gerçek deęeri,

Y_h : Simptomatik deęişkenin gerçek deęeri,

α ve β : Regresyon katsayıları,

ϵ_h : Hata terimi,

$\hat{\alpha}$ ve $\hat{\beta}$ katsayılarının tahmin edicileri sırasıyla α ve β olsun, hata terimi ihmal edilerek ($\epsilon_h = 0$), X_h 'in tah-

min edicisi (1.7) eşitliği yardımıyla,

$$\hat{X}_h^{**} = \hat{X}_h \left(\frac{\hat{\alpha} + \hat{\beta} Y_h}{100} + 1 \right) \quad (1.8)$$

olarak formüle edilir.

$\hat{X}_h^{**}$ bilinmeyen X_h 'ın gerçek değeri olduğundan, regresyon katsayılarının tahmini için değişik bir yöntem uygulanmalıdır. Bu değişik yöntem için, küçük alanların toplanarak azalan kareler yöntemi ile α ve β 'nin tahmin edilmesi önerilmektedir. Bu yaklaşım çoklu regresyon yöntemi içinde uygulanabilir.

Sentetik regresyon yöntemi ile ilgili olarak Nicholls detaylı bir çalışma yaparak bu konuda bazı yaklaşımlar önermiştir (23).

1.6. Temel Birimler Tekniği.

Küçük alanları, örnek araştırmalarından elde edilecek daha küçük temel birimlere bölmeye dayanan bu teknik, Kalsbeek tarafından önerilmiştir (1,24). Temel birimler küçük alanların homojen alt gruplarıdır. Bu model, temel birimlerin ağırlıklı kombinasyonlarını küçük alan tahminleri ağırlıklarıyla birleştirilerek oluşturulur.

k temel biriminin tahmini aşağıdaki formül yardımıyla hesaplanır.

$$\bar{X}_g = \sum_{l=1}^n w_{gl} \bar{X}_{gl} \quad g = 1, 2, \dots, k \quad (1.9)$$

Formülde,

n : Temel birim sayısı

w_{gl} : Temel birim grubu

w_{gl} : Ağırlık

$\bar{X}_{gl}$: g. grupta l. temel birimin örnek ortalaması

$\bar{X}_g$: g. temel birimin ağırlıklı ortalaması

Temel birimler tekniğinin sonuç tahmin edicisi k grupları için, (1.9)'un ağırlıklı ortalaması alınarak her bir küçük alan için aşağıdaki gibi bulunur.

$$\bar{X}_h = \sum_{g=1}^k \theta_{gh} \bar{X}_g \quad (1.10)$$

θ_{gh} ilgilenilen h küçük alanı içindeki k gruplarının birleşimlerini göstermektedir.

B Ö L Ü M 2

BİRLEŞİK SENTETİK TAHMİN EDİCİLER

Litaratürde oldukça yeni bir teknik olan ve henüz tartışma aşamasındaki birleşik sentetik tahmin ediciler, küçük alanlar için hızlı ve ekonomik bir tahmin yöntemi olarak kabul edilmektedir. Doğrudan doğruya yapılan direkt tahminler sentetik tahminlere göre daha yansız tahminler yapması nedeniyle hem örnek parametresi hem de yığın parametresi birlikte kullanılarak ağırlıklı ortalamalarının alınması fikri ortaya atılmış, tek tahmin edici için yapılan birleşik tahmin edici için de yapılmıştır. Aslında bu fikir yeni olmayıp 1936'da genetik alanında yapılan bir çalışmada kısmen uygulanmıştır (25).

2.1. Optimum Birleşik Sentetik Tahmin Ediciler.

Optimum birleşik sentetik tahmin edici, deneysel bayes tahmin edicisi tanımlanarak, bazı varsayımlarla elde edilir. Geçmiş araştırmalardan elde edilen verileri kullanarak ortalamayı tahmin eden tahmin ediciye deneysel bayes tahmin edicisi denir. Fakat, bazı koşullu dağılımlar için, geçmiş araştırmalar arasındaki periyotlar uzadığında,sonsal

ortalamaya yakınsıyan deneysel bayes tahmin edicisi bulunamaz. Ancak, buna rağmen doğrusal optimum bir θ tahmin edicisi kolayca bulunur (26).

Aşağıdaki durumlar gözönüne alınsın,

i) Parametreler $\theta_1, \theta_2, \dots, \theta_t$ olsun,

ii) Tahmin ediciler $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_t$ olsun, bu tahmin ediciler sonlu varyanslı ve yansız olsun,

Binomiyal durumda, istenen özelliğin göreceli sayısı $\bar{X}_i$ olarak seçilir, seçilen bu $\bar{X}_i$ yansız bir tahmin edicidir. Küçük alanlar için, ortalama ve/veya toplamın tahmin edildiği durumlarda, örnek çapı küçük olduğunda ve θ_i ile ilgili bilgiler elde edilebildiğinde $\bar{X}_i$ daha iyi geliştirilir.

$\bar{X}_i$ 'yi daha doğru tahmin etmek üzere yapılan deneylerden her biri, kendine ait parametresi ile aşağıdaki yapıya sahiptir (27).

i) Tahmin edilmesi gereken θ parametresi, aslında θ rasgele değişkeninin, bilinmeyen $G(\theta)$ fonksiyonunun, ilk iki sonlu momentidir.

ii) Küçük alanlar için, $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_t$ yansız tahmin edicilerin dağılımı, $F(\bar{X}_i | \theta_i)$ ($i=1,2,\dots,t$) olup,

bir deneye ait parametre θ hiç bir zaman bilinmez, ama tahmin edilebilir.

iii) Bir deney için, rasgele değişken X_i gözlenir ve bu X_i 'in dağılımı $P(X_i|\theta)$ dir.

Bu durumda X_i 'in ilk gözlenmesi θ ile gösterilirse, X_1 ile $P(X_1|\theta)$ 'den gözlenen X_1 , θ ile θ 'nin ikinci kez gözlenmesi, X_2 ile $P(X_2|\theta)$ 'den gözlenen X_2 ve genel olarak X_m ile $P(X_m|\theta)$ den gözlenen X_m elde edilir. m. tekrarda geçerli (X_1, θ) çiftinin gözlenmesiyle $(X_1, X_2, \dots, X_m)$ verilerinin tamamı elde edilmiş durumdadır (27).

$G(\theta)$ 'nin bilinmemesi nedeniyle, deneysel bayes tahmin edicisi $E(\theta|X_1, X_2, \dots, X_t)$ 'i tahmin etmede $(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_t)$ 'yi kullanır. Ancak, bazı özel dağılımlar için, başka varsayımlar yapılmadıkça $E(\theta|X_1, X_2, \dots, X_t)$ tahmin edilebilir, fakat bu tahmin tutarlı olmayabilir. Bu nedenle aşağıdaki tahmin edici üzerinde durulur (27).

$$C_i = \alpha_i \bar{X}_i + \beta_i \quad (2.1)$$

Bu doğrusal tahmin edici, bu tür tahmin ediciler sınıfı içerisinde hata kareler ortalaması ilk iki önsel momente bağlı olup, her ikisi de $(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_m)$ ile tahmin edilir.

Bu doğrusal tahmin edici, kısıt olmaksızın türetilmiş sınırsız bayes tahmin edicisi gibidir (25,28).

(2.1) ile gösterilen doğrusal tahmin edicinin α ve β parametreleri, doğrusal tahmin edicinin hata kareler ortalaması minimum yapılarak aşağıdaki gibi bulunur (27),

$$E(C_i - \theta_i)^2 = E(\alpha_i \bar{X}_i + \beta_i - \theta_i)^2$$

$$= \int_0^\infty \int_0^\infty (\alpha_i \bar{X}_i + \beta_i - \theta_i)^2 \delta F(\bar{X}_i | \theta_i) \delta G(\theta_i) \quad (2.2)$$

$E(\bar{X}_i | \theta_i) = \theta_i$ ve $\delta F(\bar{X}_i | \theta_i)$ ile $\delta G(\theta_i)$ 'nin ilk iki momentinin olduğu varsayımı yapılarak, (2.2) ifadesi minimize edildiğinde, α_i ve β_i parametreleri için,

$$\alpha_i = \frac{V(\theta_i)}{V(\theta_i) + E(V(\bar{X}_i | \theta_i))}$$

$$\beta_i = (1 - \alpha_i) \mu$$

eşitlikler elde edilir (27).

İSPAT: (2.2) ifadesini $R(G, \theta)$ ile göstererek, iki katlı

integral alınırrsa,

$$R(G, \theta) = \int_0^{\infty} \int_0^{\infty} (\alpha \bar{X}_i + \beta - \theta_i)^2 \delta F(\bar{X}_i | \theta_i) \delta G(\theta_i)$$

olur.

$R(G, \theta)$ ifadesinin $\delta F(\bar{X}_i | \theta_i)$ 'ya göre birinci integrali alınarak A ile gösterilsin.

$$\begin{aligned} A &= \int_0^{\infty} (\alpha^2 \bar{X}_i^2 + 2\alpha \beta \bar{X}_i + \beta^2 - 2\alpha \bar{X}_i \theta_i - 2\beta \theta_i + \theta_i^2) \delta F(\bar{X}_i | \theta_i) \\ A &= \int_0^{\infty} \alpha^2 \bar{X}_i^2 \delta F(\bar{X}_i | \theta_i) + \int_0^{\infty} 2\alpha \beta \bar{X}_i \delta F(\bar{X}_i | \theta_i) + \int_0^{\infty} \beta^2 \delta F(\bar{X}_i | \theta_i) \\ &\quad - \int_0^{\infty} 2\alpha \bar{X}_i \theta_i \delta F(\bar{X}_i | \theta_i) - \int_0^{\infty} 2\beta \theta_i \delta F(\bar{X}_i | \theta_i) + \int_0^{\infty} \theta_i^2 \delta F(\bar{X}_i | \theta_i) \end{aligned}$$

A ifadesindeki integraller tek tek alınırssa,

$$\int_0^{\infty} \alpha^2 \bar{X}_i^2 \delta F(\bar{X}_i | \theta_i) = \alpha^2 E(\bar{X}_i^2 | \theta_i)$$

$$\int_0^{\infty} 2\alpha \beta \bar{X}_i \delta F(\bar{X}_i | \theta_i) = 2\alpha \beta E(\bar{X}_i | \theta_i)$$

$$\int_0^{\infty} \beta^2 \delta F(\bar{X}_i | \theta_i) = \beta^2$$

$$\int_0^{\infty} 2\alpha \bar{X} \theta \delta F(\bar{X} | \theta) = 2\alpha \theta E(\bar{X} | \theta)$$

$$\int_0^{\infty} 2\beta \theta \delta F(\bar{X} | \theta) = 2\beta \theta$$

$$\int_0^{\infty} \theta^2 \delta F(\bar{X} | \theta) = \theta^2$$

Alınan integraller yerine konulduğunda A ifadesi,

$$A = \alpha^2 E(\bar{X}^2 | \theta) + 2\alpha \beta E(\bar{X} | \theta) + \beta^2 - 2\alpha \theta E(\bar{X} | \theta) - 2\beta \theta + \theta^2$$

olarak bulunur. $R(G, \theta)$ 'nin ikinci integrali alınır,

$$\begin{aligned} R(G, \theta) &= \int_0^{\infty} A \delta G(\theta) \\ &= \int_0^{\infty} [\alpha^2 E(\bar{X}^2 | \theta) + 2\alpha \beta E(\bar{X} | \theta) + \beta^2 - 2\alpha \theta E(\bar{X} | \theta) - 2\beta \theta + \theta^2] \delta G(\theta) \\ &= \int_0^{\infty} \alpha^2 E(\bar{X}^2 | \theta) \delta G(\theta) + \int_0^{\infty} 2\alpha \beta E(\bar{X} | \theta) \delta G(\theta) \\ &\quad + \int_0^{\infty} \beta^2 \delta G(\theta) - \int_0^{\infty} 2\alpha \theta E(\bar{X} | \theta) \delta G(\theta) \end{aligned}$$

$$- \int_0^{\infty} 2\beta \theta_i \delta G(\theta_i) + \int_0^{\infty} \theta_i^2 \delta G(\theta_i) \text{ olarak yazılır,}$$

$R(G, \theta)$ ifadesindeki integraller alınırrsa,

$$\int_0^{\infty} \alpha_i^2 E(\bar{X}_i^2 | \theta_i) \delta G(\theta_i) = \alpha_i^2 E(E(\bar{X}_i^2 | \theta_i))$$

$$\int_0^{\infty} 2\alpha_i \beta_i E(\bar{X}_i | \theta_i) \delta G(\theta_i) = 2\alpha_i \beta_i E(E(\bar{X}_i | \theta_i))$$

$$\int_0^{\infty} \beta_i^2 \delta G(\theta_i) = \beta_i^2$$

$$\int_0^{\infty} 2\alpha_i \theta_i E(\bar{X}_i | \theta_i) \delta G(\theta_i) = 2\alpha_i E(\theta_i E(\bar{X}_i | \theta_i))$$

$$\int_0^{\infty} 2\beta_i \theta_i \delta G(\theta_i) = 2\beta_i E(\theta_i)$$

$$\int_0^{\infty} \theta_i^2 \delta G(\theta_i) = E(\theta_i^2)$$

$R(G, \theta)$ ifadesinde alınan integral sonuçları yerine konulursa,

$$\begin{aligned} R(G, \theta) = & \alpha_i^2 E(E(\bar{X}_i^2 | \theta_i)) + 2\alpha_i \beta_i E(E(\bar{X}_i | \theta_i)) + \beta_i^2 \\ & - 2\alpha_i E(\theta_i E(\bar{X}_i | \theta_i)) - 2\beta_i E(\theta_i) + E(\theta_i^2) \end{aligned} \quad (2.3)$$

eşitliği elde edilir.

α_i ve β_i katsayılarını minimum yapmak üzere, (2.3) ifadesinin α_i ve β_i 'e göre kısmi türevleri alınır,

$$\frac{\partial R(G, \theta)}{\partial \alpha_i} = 2\alpha_i E(E(\bar{X}^2 | \theta_i)) + 2\beta_i E(E(\bar{X} | \theta_i)) - 2 E(\theta_i E(\bar{X} | \theta_i)) = 0 \quad (2.4)$$

$$\frac{\partial R(G, \theta)}{\partial \beta_i} = 2\alpha_i E(E(\bar{X} | \theta_i)) + 2\beta_i - 2 E(\theta_i) = 0 \quad (2.5)$$

α_i ve β_i değerlerinin minimum olduğunu görmek için, (2.4) ve (2.5)'in tekrar kısmi türevleri alındığında,

$$\frac{\partial^2 R(G, \theta)}{\partial \alpha_i^2} = 2 E(E(\bar{X}^2 | \theta_i)) > 0 \quad (2.6)$$

$$\frac{\partial^2 R(G, \theta)}{\partial \beta_i^2} = 2 > 0 \quad (2.7)$$

olduğu görülür.

(2.6) ve (2.7) ifadeleri gereği, (2.4) ve (2.5)'den elde edilecek α_i ve β_i değerleri minimum olacaktır.

(2.4) ve (2.5) ifadelerinden α_i ve β_i 'yi elde etmek

için aşağıdaki düzenleme yapılırsa,

$$\alpha_i E(E(\bar{X}_i^2 | \theta_i)) + \beta_i E(E(\bar{X}_i | \theta_i)) = E(\theta_i E(\bar{X}_i | \theta_i)) \quad (2.8)$$

$$\alpha_i E(E(\bar{X}_i | \theta_i)) + \beta_i = E(\theta_i) \quad (2.9)$$

elde edilir.

Başlangıçta verilen $E(\bar{X}_i | \theta_i) = \theta_i$ koşulu (2.8) ve (2.9) eşitliklerinde yerine konursa,

$$\alpha_i E(E(\bar{X}_i^2 | \theta_i)) + \beta_i E(\theta_i) = E(\theta_i^2) \quad (2.10)$$

$$\alpha_i E(\theta_i) + \beta_i = E(\theta_i) \quad (2.11)$$

elde edilir.

(2.10) ifadesinden α_i , (2.11) ifadesinden β_i çekilirse,

$$\beta_i = (1 - \alpha_i) E(\theta_i) \quad (2.12)$$

$$\alpha_i = \frac{E(\theta_i^2) - \beta_i E(\theta_i)}{E(E(\bar{X}_i^2 | \theta_i))} \quad (2.13)$$

olur.

(2.12) eşitliğinden β_i , (2.13)'teki β_i yerine konursa α_i için,

$$\begin{aligned}
\alpha_i &= \frac{E(\theta_i^2) - E(\theta_i)^2}{E(E(\bar{X}_i^2 | \theta_i) + E(\theta_i)^2)} \\
&= \frac{V(\theta_i)}{V(\theta_i) + E(V(\bar{X}_i | \theta_i))} \quad (2.14)
\end{aligned}$$

eşitliği elde edilir.

Binomiyal durumda, θ ile ilgili olarak aşağıdaki eşitlikler geçerlidir,

$$V(\theta) = \sigma^2$$

$$E(\theta) = \mu$$

$$E(V(\bar{X}_i | \theta)) = \mu(1 - \mu) / n_i - V(\theta) / n_i$$

Bu eşitlikler (2.12) ve (2.14)'de yerine konulursa, α_i ve β_i için,

$$\alpha_i = \frac{\sigma^2}{(1 - 1/n_i)\sigma^2 + \mu(1 - \mu)/n_i} \quad (2.15)$$

$$\beta_i = (1 - \alpha_i) \mu \quad (2.16)$$

eşitliklerini yazmak mümkündür.

(2.2) ile verilen hata kareler ortalamasını minimum yapan α_i ve β_i değerleri sırasıyla (2.15) ve (2.16)'daki gibi bulunarak ispat tamamlanır.

(2.15) ve (2.16) ifadelerine dayanarak, (2.1)'de verilen tahmin edici,

$$C_i = \alpha_i \bar{X}_i + (1 - \alpha_i) \mu \quad (2.17)$$

şeklinde formüle edilir.

(2.15) ile verilen ifadede, n_i i. küçük alandan elde edilen gözlem sayısıdır. (2.17) ifadesi incelendiğinde örnek ortalamasının α_i ile, yığın ortalamasının ise $(1 - \alpha_i)$ ile ağırlıklandırıldığı görülmektedir.

(2.17) eşitliğinde $\alpha_i = W_i$ yazılarak, (2.1) ile verilen tahmin edici (gösterimlerde kolaylık sağlama için 1'inci tahmin edici olarak gösterilmiştir.) için,

$$C_i(1) = W_i \bar{X}_i + (1 - W_i) \mu \quad (2.18)$$

eşitliği yazılır.

(2.18) ile verilen tahmin edici, doğrusal birleşik tahmin edicidir ve birleşik sentetik tahmin edici olarak adlandırılır. (2.18) ile verilen sentetik birleşik tahmin edici Copas ve Griffin tarafından öne sürülmüştür (26,27).

2.2. Yığın Ortalaması ve Varyansına Dayanarak Yapılan Birleşik Sentetik Tahmin Ediciler.

(2.18) ile verilen $C_i(1)$ gibi birleşik sentetik tahmin ediciler önsel ortalamaya doğru yaklaşırlar. Bu nedenle değişkenlik, tahmin edilen değerler arasındaki değişim, $\theta_1, \theta_2, \dots, \theta_T$ 'nin gerçek değerleri arasındaki farktan daha küçük olacaktır. Bunu görmek için, aşağıdaki çıkarıma izinlenirse,

$$\begin{aligned}
 E \left(\frac{1}{T} \sum_{i=1}^T (C_i(1) - \mu)^2 \right) &= E \left(\frac{1}{T} \sum_{i=1}^T (W_i \bar{X}_i + (1 - W_i) \mu - \mu)^2 \right) \\
 &= E \left(\frac{1}{T} \sum_{i=1}^T (W_i \bar{X}_i + \mu - \mu W_i - \mu)^2 \right) \\
 &= E \left(\frac{1}{T} \sum_{i=1}^T W_i^2 (\bar{X}_i - \mu)^2 \right) \\
 &= \sigma^2 \frac{1}{T} \sum_{i=1}^T \bar{W}_i^2 \leq \sigma^2 \quad (2.19)
 \end{aligned}$$

olduğu görülür.

Küçük alan tahminleri ile ilgili olarak, kullanıcılar σ^2 ve küçük alanlar arasındaki değişimle ilgili olarak bazı önsel bilgilere sahiptir. Bazı durumlarda (2.18) ile verilen birleşik sentetik tahmin ediciler ile ilgili varsayımlar sağlanmayabilir yada istatistiksel analizciler için bu

varsayımlar yeterli olmayabilir. Bu nedenle, (2.1)'de tanımlanan doğrusal tahmin edicinin α_i ve β_i katsayılarını (2.2) deki gibi tahmin etmek yerine aşağıdaki gibi belirlenir.

$$E(C_i(1)) = \mu \quad (2,20)$$

(2.20) eşitliğini elde etmek üzere, çıkarsama izlenirse,

$$\begin{aligned} E(C_i(1)) &= E(W_i \bar{X}_i + (1 - W_i)\mu) = E(W_i \bar{X}_i + \mu - W_i \mu) \\ &= E(W_i \bar{X}_i) + E(\mu) - E(W_i \mu) \\ &= W_i E(\bar{X}_i) + E(\mu) - W_i E(\mu) \\ &= W_i \mu + \mu - W_i \mu = \mu \end{aligned}$$

olduğu görülür. Ayrıca,

$$E(C_i(1) - \mu)^2 = \sigma^2 \text{ 'dir.} \quad (2.21)$$

(2.20) ve (2.21) ifadelerine dayanarak α_i ve β_i için,

$$\alpha_i = W_i^{\frac{1}{2}}$$

$$\beta_i = (1 - W_i^{\frac{1}{2}})$$

eşitlikleri elde edilir.

Bu durumda belirlenen ağırlıklar (2.1) doğrusal tahmin edicisinde yerine konularak yeni bir birleşik sentetik tahmin edici,

$$C_i(2) = W_i \bar{X}_i + (1 - W_i) \mu \quad (2.22)$$

şeklinde verilmiştir.

$C_i(2)$ birleşik sentetik tahmin edicisi $C_i(1)$ birleşik sentetik tahmin edicisine göre örnek ortalaması $\bar{X}_i$ 'ye daha büyük ağırlık vermektedir. Eğer önsel $G(\theta)$ fonksiyonu normal dağılıma sahipse, $C_i(2)$ birleşik sentetik tahmin edicisi de θ gibi aynı dağılıma sahiptir (29).

2.3. Örnek Çapının Bir Fonksiyonu ile Ağırlıklandırılmış Birleşik Sentetik Tahmin Ediciler.

Küçük alan tahminleri ile ilgili olarak, basit bir birleşik sentetik tahmin edici W_i gibi bir ağırlık seçilerek oluşturulur. Bu yaklaşımdan hareketle, literatürde bir çok birleşik sentetik tahmin edici önerilmiştir.

Schaible tarafından öne sürülen, örnek çapının bir fonksiyonu ile ağırlıklandırılmış birleşik sentetik tahmin ediciye ilişkin W_i ağırlığı,

$$W_i = \frac{1}{(1 + n_i^{-\frac{1}{2}})}$$

olarak tanımlanmaktadır (30).

Burada n_i i. küçük alanın örnek çapıdır. Verilen W_i ağırlığı kullanılarak yeni bir birleşik sentetik tahmin edici,

$$C_i(3) = \frac{1}{(1 + n_i^{-\frac{1}{2}})} \bar{X}_i + (1 - \frac{1}{(1 + n_i^{-\frac{1}{2}})}) \mu \quad (2.23)$$

şaklinde tanımlanmaktadır (30).

2.4. Hata Kareler Ortalamasına Dayanan Birleşik Sentetik Tahmin Ediciler.

Her küçük alan için θ_i 'yi veren hata kareler ortalaması aşağıdaki gibi tanımlanır.

$$E(C_i - \theta_i | \theta_i)^2 = W_i^2 \theta_i (1 - \theta_i)/n_i + (1 - W_i)^2 (\theta_i - \mu)^2 \quad (2.24)$$

(2.24) ile verilen hata kareler ortalamasını minimum yapan W_i ağırlığı aşağıdaki gibidir,

$$W_i = \frac{(\theta_i - \mu)^2}{\theta_i (1 - \theta_i)/n_i + (\theta_i - \mu)^2} \quad (2.25)$$

İSPAT : (2.24) ifadesini minimum yapan (2.25)'deki W_i ağırlığını bulmak üzere, (2.24)'ün W_i ye göre kısmi türevi alınarak 0'a eşitlenirse,

$$\frac{\partial E((C_i - \theta_i)^2 | \theta_i)}{\partial W_i} = 2 W_i \theta_i (1 - \theta_i) / n - 2 (1 - W_i)(\theta_i - \mu)^2 = 0 \quad (2.26)$$

olur ve (2.26)'daki birinci kısmi türevin eksterm değerinin minimum veya maksimum olduğunu tesbit etmek için, W_i ye göre ikinci kısmi türev alınır,

$$\frac{\partial^2 E((C_i - \theta_i)^2 | \theta_i)}{\partial W_i^2} = 2 \theta_i (1 - \theta_i) / n + 2(\theta_i - \mu)^2 \geq 0 \quad (2.27)$$

olduğu görülür.

(2.27) ifadesinin bir sonucu olarak, bulunacak W_i minimumdur. Bu durumda (2.26) eşitliğinden W_i çekildiğinde (2.25) ile verilen ağırlık,

$$W_i \equiv \frac{(\theta_i - \mu)^2}{\theta_i (1 - \theta_i) / n + (\theta_i - \mu)^2}$$

olarak elde edilir.

Yeni bir birleşik sentetik tahmin edici (2.25) eşitli-

ğindeki w_i ağırlığı kullanılarak,

$$C_i(4) = \frac{(\theta_i - \mu)^2}{\theta_i(1 - \theta_i)/n + (\theta_i - \mu)^2} \bar{X}_i + \left(1 - \frac{(\theta_i - \mu)^2}{\theta_i(1 - \theta_i)/n + (\theta_i - \mu)^2}\right) \mu \quad (2.28)$$

şeklinde formüle edilir.

(2.28) ile verilen birleşik sentetik tahmin edici genellikle multinomial verileri kullanır (31).

2.5. Birleşik Sentetik Tahmin Edicilerin Ağırlıklı- rının Tahmini.

Bir çok uygulamada σ^2 ve μ genellikle bilinmez, ancak, $(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_T)$ ile tahmin edilir. σ^2 ve μ 'nün bilinmediği durumlarda (2.18), (2.22), ve (2.23) eşitlikleri ile verilen birleşik sentetik tahmin edicilerin w_i ağırlıklarının hesaplanmasında, σ^2 ve μ 'nün tahmini,

$$\hat{\mu} = \sum_i (n_i/n) \bar{X}_i \quad (2.29)$$

$$\hat{\sigma}^2 = \frac{\sum_i (n_i/n)(\bar{X}_i - \hat{\mu})^2 - k\hat{\mu}(1 - \hat{\mu})}{T - k} \quad (2.30)$$

eşitlikleri ile elde edilir. (2.29) ve (2.30) ifadelerinde T küçük alan sayısı olup, n ve k değerleri,

$$n = \sum_i^T n_i$$

$$k = 1/\sum_i^T n_i$$

yardımları ile bulunur.

Bazı durumlarda $\hat{\sigma}^2$ 'nin tahmin edicisi olan, $\hat{\sigma}^2$ negatif bir değer olarak ortaya çıkabilir, bu halde $\hat{\sigma}^2$ tahmin edicisi 0 olarak kabul edilir (26).

2.6. Birleşik Sentetik Tahmin Edicilerin Karşılaştırılması İçin Risk Etkinlik Fonksiyonu.

Birleşik sentetik tahmin edicilerin kullanılmasıyla yapılan küçük alan tahminlerini değerlendirirken, performansın genel ölçümü ile birlikte diğer bir çok yönünün de incelenmesi gerekir. (2.24) ile verilen, her küçük alan için θ_i 'yi veren hata kareler ortalamasının, i. küçük alan tahmininin varyansına oranı birleşik sentetik tahminlerin değer-

lendirilmesi için bir ölçüttür. i .küçük alan tahmininin varyansı,

$$v(\bar{X}_i) = w_i^2 \theta_i (1 - \theta_i) / n_i \quad (2.31)$$

eşitliği ile tanımlanır.

θ_i 'nin değeri küçükten büyüğe sıralandığında, hata kareler ortalamasının (i) 'nin fonksiyonu gibi olduğu görülür. Bunun anlamı C_i birleşik sentetik tahmin edicisi kullanılarak elde edilen kazancın, küçük alandan küçük alana değiştiğidir. Küçük bir alan için, önsel ortalamaya yakın bir θ_i değeri ile kazanç daha büyüktür.

Birleşik sentetik tahmin edicilerle yapılan küçük alan tahminlerinin duyarlılığını ölçmek ve birleşik sentetik tahmin edicilerini duyarlılık açısından birbirleriyle karşılaştırmak için (2.31) ile verilen ifade (2.24) ifadesine bölünür. Bu orana " Risk Etkinlik Fonksiyonu " denir ve,

$$R(C_i) = \frac{E \left(\frac{1}{T} \sum_{i=1}^T (\bar{X}_i - \theta_i)^2 \right)}{E \left(\frac{1}{T} \sum_{i=1}^T (C_i - \theta_i)^2 \right)} \quad (2.32)$$

şeklinde formüle edilir.

(2.32) ile verilen risk etkinlik fonksiyonu her bir birleşik sentetik tahmin edici için hesaplanarak, 1'e en yakın $R(C_i)$ 'ye ilişkin birleşik sentetik tahmin edici en uygun tahmin edici olarak seçilir.

Birleşik Sentetik Tahmin Ediciler genel olarak yanlış tahminler üretmektedir. Çok küçük örnek çapına dayanması nedeniyle uygulamada isabetli bir risk etkinlik değerine sahip olması, yanlışlık açısından önemli bir farklılık ortaya koymayacaktır.

Başka bir amaçla yapılan kalite kontrol anketinden derlenen verilere birleşik sentetik tahmin ediciler uygulandığında risk etkinlik fonksiyonu değeri bir'den büyük olarak ortaya çıkmaktadır. Bu durum yapılan anketin yapısından kaynaklanmaktadır.

B Ö L Ü M 3

U Y G U L A M A

Birinci bölümde küçük alan tahminlerinin özellikle nüfus tahminleri konusunda uzmanlaştığı, ikinci bölümde ise birleşik sentetik tahmin edicilerin küçük alan tahminlerini ekonomik ve hızlı bir şekilde yaptığı söylenmişti. Bu bağlamda, göç alan illerden elde edilen küçük alan verilerine birleşik sentetik tahmin ediciler uygulanarak, her bir küçük alan için göç oranları tahmin edilecektir. Dört ayrı birleşik sentetik tahmin edici kullanılarak yapılan bu göç tahminlerinin hangisinin daha iyi tahmin yaptığını belirlemek üzere her bir birleşik sentetik tahmin edici için risk etkinlik fonksiyonu hesaplanacaktır.

Devlet İstatistik Enstitüsü'nce nüfus sayımları ile ilgili uygulamalarda, kapsam ve elde edilen bilgilerin doğruluğu, nüfus sayımında yer alan sorulara alınan yanıtların tutarlılık derecesinin tesbiti için bir kalite kontrol çalışması yapılmıştır (*) (32).

Bu çalışmada, Türkiye'de yerleşim yerleri nüfus büyüklükleri ele alınarak tabakalara ayrılmıştır. Uygulamada eş

Kaynak : Devlet İstatistik Enstitüsü

olasılıklı yöntem kullanılarak örnek çapı bölge ve nüfus grubu içerisinde gerçekleştirilmiştir. Bu şekilde 18 il belirlenerek, bu illerden küçük çaplı örnek seçilmiş ve örneğe giren kişilere genel nüfus sayımı soruları sorulmuştur. Bu çalışmada küçük alan olarak belirlenen illerin göç oranı değişkeni ile ilgilenilecektir. Bu nedenle örneğe giren kişilere sorulan göçlerle ilgili göç sayılarına ilişkin veriler değerlendirilerek aşağıdaki Tablo 2’de verilmiştir.

Tablo 2. Göç Alan İllerdeki Örnek Çapı, Göç Sayısı ve Göç Ortalamasının Dağılımı.

İl Adı	Örnek Çapı n_i	Göç Sayısı X_i	Göç Oranı $\bar{X}_i$
Adana	730	63	0.08630
Ankara	924	196	0.21212
Artvin	170	9	0.05294
Bitlis	148	4	0.02703
Burdur	193	8	0.04145
Bursa	891	29	0.03251
Diyarbakır	368	38	0.10326
Erzurum	333	53	0.15916
Eskişehir	709	32	0.04513

(Tablo 2'nin devamı)

İl Adı	Örnek Çapı n_i	Göç Sayısı $\bar{X}_i$	Göç Oranı X_i
Gaziantep	348	30	0.08621
İstanbul	1041	178	0.17099
İzmir	1165	43	0.03691
Konya	721	86	0.11928
Niğde	356	29	0.08146
Samsun	285	43	0.15088
Tekirdağ	324	18	0.05556
Urfa	299	13	0.04348
Zonguldak	401	32	0.07980
Toplam	9407	904	

a) $\hat{\mu}$ ve $\hat{\sigma}^2$ 'nin tahminin hesaplanması,

(2.29) ile verilen formül yardımıyla, göç alan illere ilişkin yığın ortalaması,

$$\hat{\mu} = \sum_{i=1}^T (n_i / n) \bar{X}_i$$

ile hesaplandığında, $n = \sum_{i=1}^T n_i = 730+924+...401 = 9407$ ve

$$\begin{aligned}\hat{\mu} &= [(730/9407)*0.08630] + [(924/9407)*0.21212] + \dots \\ &\dots + [(401/9407)*0.07980] \\ &= 0,0961 \text{ olarak bulunur.}\end{aligned}$$

(2.30) ile verilen formül yardımıyla, göç alan illere ilişkin yığın varyansı,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^T [(n_i/n)(\bar{X}_i - \hat{\mu})^2] - \hat{\mu}(1 - \hat{\mu})/n}{T - 1/n}$$

ile hesaplandığında,

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{18-(1/9407)} \{ [(730/9407)*(0.08630 - 0.0961)^2 + \dots \\ &\dots + (401/9407)*(0.07980 - 0.0961)^2] \\ &- [0.0961*(1 - 0.0961)/9407] \} \\ &= 0,0002 \text{ olarak bulunur.}\end{aligned}$$

3.1. C (1) Birleşik Sentetik Tahmin Edici

i
İçin Uygulama.

Göç veren illerdeki, göç oranının tahmini için (2.18) ile verilen C (1) birleşik sentetik tahmin edicisi uygulanarak elde edilen göç alan illerin göç oranının tahmini ve birleşik sentetik tahmin edicilerin W_i ağırlıklarına ilişkin değerler aşağıda verilen Tablo 3'de gösterilmiştir.

Tablo 3. C (1) Birleşik Sentetik Tahmin Edicisinin
İllere Göre Göç Tahmininin Dağılımı.

İl Adı	Ağırlık W i	Göç Oranı C (1) i
Adana	0.62751	0.08995
Ankara	0.68075	0.17508
Artvin	0.28177	0.08394
Bitlis	0.25459	0.07851
Burdur	0.30815	0.07926
Bursa	0.67304	0.05330
Diyarbakır	0.45924	0.09939
Erzurum	0.43454	0.12350
Eskişehir	0.62066	0.06447
Gaziantep	0.44540	0.09169
İstanbul	0.70609	0.14898
İzmir	0.72889	0.05296
Konya	0.62461	0.11058
Niğde	0.45102	0.08950
Samsun	0.39676	0.11783
Tekirdağ	0.42782	0.07875
Urfa	0.40829	0.07462
Zonguldak	0.48063	0.08827

Tablo 3'de verilen $C(1)$ Birleşik Sentetik Tahmin Edicisi ve ağırlıkların hesaplanmasını göstermek üzere ilgili formüller Adana ili için uygulanırsa,

(2.15) ile verilen W_i ağırlığının tahmini,

$$\hat{W}_i = \frac{\hat{\sigma}^2}{(1-1/n_i)\hat{\sigma}^2 + \hat{\mu}(1-\hat{\mu})/n_i} \quad \text{ile yapılmaktadır.}$$

$\hat{\sigma}^2 = 0.0002$ ve $\hat{\mu} = 0.0961$ olarak hesaplanmıştı, Adana için örnek çapı $n_i = 730$ 'dır. Bu durumda,

$$\hat{W}_{\text{Adana}} = 0,0002 / \{ (1-1/730)0,0002 + 0,0961(1-0,0961)/730 \} \\ = 0,62751 \text{ olarak bulunur.}$$

(2.18) ile verilen birleşik sentetik tahmin edicinin tahmini,

$$\hat{C}_i(1) = \hat{W}_i \bar{X}_i + (1-\hat{W}_i) \hat{\mu}_i \quad \text{ile tanımlanmaktadır.}$$

Adana için göç oranının birleşik sentetik tahmini,

$$\hat{C}_{\text{Adana}}(1) = 0,62751 (0,08630) + (1- 0,62751) 0,0961 \\ = 0,08995 \text{ olarak bulunur.}$$

3.2. C (2) Birleşik Sentetik Tahmin Edici

i

İçin Uygulama.

Göç alan illerdeki, göç oranının tahmini için (2.22) ile verilen C (2) birleşik sentetik tahmin edicisi uygulanarak elde edilen göç alan illerin göç oranının tahmini ve birleşik sentetik tahmin edicilerin W_i ağırlıklarına ilişkin değerler aşağıda verilen Tablo 4'de gösterilmiştir.

Tablo 4. C (2) Birleşik Sentetik Tahmin Edicisinin
İllere Göre Göç Tahmininin Dağılımı.

İl Adı	Ağırlık W_i	Göç Oranı C (2) i
Adana	0.79216	0.08834
Ankara	0.82508	0.19183
Artvin	0.53082	0.07319
Bitlis	0.50457	0.06125
Burdur	0.55511	0.06576
Bursa	0.82039	0.04393
Diyarbakır	0.67767	0.10095
Erzurum	0.65920	0.13767
Eskişehir	0.78782	0.05595

(Tablo 4'ün devamı)

İl Adı	Ağırlık w_i	Göç Oranı $C(2)_i$
Gaziantep	0.66738	0.08950
İstanbul	0.84029	0.15903
İzmir	0.85375	0.04557
Konya	0.79032	0.11442
Niğde	0.67158	0.08627
Samsun	0.62989	0.13060
Tekirdağ	0.65408	0.06958
Urfa	0.63898	0.06248
Zonguldak	0.69327	0.08480

Tablo 4'de verilen $C(2)_i$ Birleşik Sentetik Tahmin Edicisi ve ağırlıkların hesaplanmasını göstermek üzere ilgili formüller Adana ili için uygulanırsa,

(2.22) ile verilen Birleşik Sentetik Tahmin Edicinin w_i ağırlığının tahmini,

$$\hat{w}_i = \left[\frac{\hat{\sigma}^2}{(1-1/n_i)\hat{\sigma}^2 + \mu(1-\mu)/n_i} \right]^{\frac{1}{2}} \text{ ile yapılmaktadır.}$$

$\hat{\sigma}^2 = 0.0002$ ve $\hat{\mu} = 0.0961$ olarak hesaplanmışdı, Adana için örnek çapı $n_i = 730$ 'dır. Bu durumda,

$$W_{\text{Adana}} = [0,0002 / \{ (1 - 1/730)0,0002 + 0,0961(1 - 0,0961)/730 \}]^{\frac{1}{2}}$$

= 0,79216 olarak bulunur.

(2.18) ile verilen birleşik sentetik tahmin edicinin tahmini,

$$\hat{C}_i(2) = (\hat{W}_i) \bar{X}_i + (1 - \hat{W}_i) \hat{\mu}_i \text{ ile tanımlanmaktadır.}$$

Adana için göç oranının birleşik sentetik tahmini,

$$\hat{C}_{\text{Adana}}(2) = 0,79216 (0,08630) + (1 - 0,79216) 0,0961$$

= 0,08834 olarak bulunur.

3.3. $C_i(3)$ Birleşik Sentetik Tahmin Edici

İçin Uygulama.

Göç alan illerdeki, göç oranının tahmini için (2.23) ile verilen $C_i(3)$ birleşik sentetik tahmin edicisi uygulanarak elde edilen göç alan illerin göç oranının tahmini ve birleşik sentetik tahmin edicilerin W_i ağırlıklarına ilişkin değerler aşağıda verilen Tablo 5'de gösterilmiştir.

Tablo 5. C (3) Birleşik Sentetik Tahmin Edicisinin
i
İllere Göre Göç Tahmininin Dağılımı.

İl Adı	Ağırlık W i	Göç Oranı C (3) i
Adana	0.96431	0.08834
Ankara	0.96815	0.19183
Artvin	0.92877	0.07319
Bitlis	0.92404	0.06125
Burdur	0.93285	0.06576
Bursa	0.96760	0.04393
Diyarbakır	0.95045	0.10095
Erzurum	0.94805	0.13767
Eskişehir	0.96380	0.05595
Gaziantep	0.94912	0.08950
İstanbul	0.96994	0.15903
İzmir	0.97154	0.04557
Konya	0.96410	0.11442
Niğde	0.94967	0.08627
Samsun	0.94408	0.13060
Tekirdağ	0.94737	0.06958
Urfa	0.94533	0.06248
Zonguldak	0.95244	0.08480

Tablo 5'de verilen $C(3)$ Birleşik Sentetik Tahmin Edicisi ve ağırlıkların hesaplanmasını göstermek üzere ilgili formüller Adana ili için uygulanırsa,

(2.23) ile verilen Birleşik Sentetik Tahmin Edicinin W ağırlığının tahmini,
 i

$$W_i = \frac{1}{(1+n_i)^{-\frac{1}{2}}} \text{ ile tanımlanmaktadır.}$$

Adana için,

örnek sayısı $n_i = 730$ 'dır. Bu durumda,

$$W_{\text{Adana}} = 1 / [1 + (730)^{-\frac{1}{2}}]$$

= 0,96431 olarak bulunur.

(2.23) ile verilen birleşik sentetik tahmin edici,

$$C_i(3) = (W_i) \bar{X}_i + (1-W_i) \hat{\mu} \text{ ile tanımlanmaktadır.}$$

Adana için göç oranının birleşik sentetik tahmini,

$$C_{\text{Adana}}(3) = 0,96431 (0,08630) + (1 - 0,96431) 0,0961$$

= 0,08665 olarak bulunur.

3.4. C (4) Birleşik Sentetik Tahmin Edici i

için Uygulama.

Göç alan illerdeki, göç oranının tahmini için (2.28) ile verilen C (4) birleşik sentetik tahmin edicisi uygulanarak elde edilen göç alan illerin göç oranının tahmini ve birleşik sentetik tahmin edicilerin W_i ağırlıklarına ilişkin değerler aşağıda verilen Tablo 6'de gösterilmiştir.

Tablo 6: C (4) Birleşik Sentetik Tahmin Edicisinin
i illere Göre Göç Tahmininin Dağılımı.

İl Adı	Ağırlık W_i	Göç Oranı C (4) i
Adana	0.99886	0.08621
Ankara	0.99612	0.21133
Artvin	0.99938	0.05291
Bitlis	0.99969	0.02702
Burdur	0.99951	0.04144
Bursa	0.99960	0.03250
Diyarbakır	0.99871	0.10314
Erzurum	0.99795	0.15885
Eskişehir	0.99944	0.04511

(Tablo 6'nın devamı)

İl Adı	Ağırlık W_i	Göç Oranı $C(4)_i$
Gaziantep	0.99894	0.08613
İstanbul	0.99702	0.17051
İzmir	0.99953	0.03690
Konya	0.99834	0.11910
Niğde	0.99900	0.08139
Samsun	0.99810	0.15061
Tekirdağ	0.99934	0.05553
Urfa	0.99949	0.04346
Zonguldak	0.99902	0.07973

Tablo 6'da verilen $C(4)_i$ Birleşik Sentetik Tahmin Edicisi ve ağırlıkların hesaplanmasını göstermek üzere ilgili formüller Adana ili için uygulanırsa,

(2.25) ile verilen Birleşik Sentetik Tahmin Edicinin W_i ağırlığının tahmini,

$$W_i = \frac{(\hat{\theta}_i - \mu)^2}{\hat{\theta}_i (1 - \hat{\theta}_i)/n + (\hat{\theta}_i - \mu)^2} \quad \text{ile tanımlanmaktadır.}$$

Adana için,

örnek çapı $n = 730$ ve $\mu_i = 0,0961$ 'dir. Bu durumda,

$$W_{\text{Adana}} = \frac{(0,00670 - 0,0961)^2}{0,00670(1 - 0,00670)/730 + (0,00670 - 0,0961)^2}$$
$$= 0,99886 \text{ olarak bulunur.}$$

(2.23) ile verilen birleşik sentetik tahmin edici,

$$C_i(3) = (W_i) \bar{X}_i + (1 - W_i) \hat{\mu}_i \text{ ile tanımlanmaktadır.}$$

Adana için göç oranının birleşik sentetik tahmini,

$$C_{\text{Adana}}(3) = 0,99886 (0,08630) + (1 - 0,99886) 0,0961$$
$$= 0,08621 \text{ olarak bulunur.}$$

3.5. Birleşik Sentetik Tahmin edicilerin Karşılaştırılması.

Bölüm 2.6.'da birleşik sentetik tahmin edicilerin birbirleriyle karşılaştırılması için risk etkinlik fonksiyonuyla hesaplanan değerlerin 1'e en yakın olanının tercih edilmesi gerektiği belirtilmişti. Bu nedenle, dört ayrı birleşik sentetik tahmin ediciden bu veriler üzerinde hangisinin daha iyi tahmin yaptığını belirlemek üzere ile

verilen risk etkinlik fonksiyonu hesaplanarak aşağıdaki tabloda verilmiştir.

Tablo 7. Birleşik Sentetik Tahmin Edicilerin Risk Etkinlik Fonksiyonu Değerleri

Birleşik Sentetik Tahmin Edicinin Türü	Risk Etkinlik Fonksiyonu Değeri
$C_i(1)$	1.08922
$C_i(2)$	1.04366
$C_i(3)$	1.09466
$C_i(4)$	1.10202

Risk etkinlik fonksiyonu değeri,

$$R(C_i) = \frac{\sum_{i=1}^T (\bar{X}_i - \theta_i)^2}{\sum_{i=1}^T (C_i - \theta_i)^2} \text{ ile hesaplanmaktadır.}$$

Tablo 7'de verilen $R(C_i)$ Risk etkinlik değerleri sırasıyla,

$$R(C(1)) = \frac{(0,08630-0,00670)^2 + \dots + (0,07980-0,00340)^2}{(0,08995-0,00670)^2 + \dots + (0,08827-0,00340)^2}$$

$$R(C(1)) = \frac{0,16524}{0,15170} = 1,08922$$

$$R(C(2)) = \frac{(0,08630-0,00670)^2 + \dots + (0,07980-0,00340)^2}{(0,08834-0,00670)^2 + \dots + (0,08480-0,00340)^2}$$

$$= \frac{0,16524}{0,15833} = 1,04366$$

$$R(C(3)) = \frac{(0,08630-0,00670)^2 + \dots + (0,07980-0,00340)^2}{(0,08665-0,00670)^2 + \dots + (0,08058-0,00340)^2}$$

$$= \frac{0,16524}{0,15095} = 1,09466$$

$$R(C(4)) = \frac{(0,08630-0,00670)^2 + \dots + (0,07980-0,00340)^2}{(0,09149-0,00670)^2 + \dots + (0,08645-0,00340)^2}$$

$$= \frac{0,16524}{0,14994} = 1,10202 \text{ olarak bulunur.}$$

Tablo 7'de görüldüğü gibi, 1'en yakın risk etkinlik fonksiyonu değerine sahip birleşik sentetik tahmin edici $C(2)$ 'dir. Bu nedenle, göç tahminini bu veriler için en uygun tahmin eden birleşik sentetik tahmin edici $C(2)$ olarak kabul edilir.

B Ö L Ü M 4

SONUÇ VE TARTIŞMA

Yığının farklı özellik gösteren alt grublarına ait karakteristiklerin tahminine duyulan gereksinim nedeniyle literatürde bir çok küçük alan tahmin teknikleri önerilmiştir. Bu çalışmanın birinci bölümünde bazı küçük alan tahmin teknikleri açıklanarak, bunların dayandığı varsayımlara ve formülasyonlara yer verilmiştir.

Küçük alan tahminleri, yanlış tahminler vermesi ve teorik çarışının tam oluşmaması dezavantajına karşın, maliyet ve zaman açısından sağladığı avantajlar nedeniyle kullanım alanı bulmaktadır. Bu tahmin tekniklerinin veri kaynaklarının geniş olması, (Örneğin, resmi kayıtlar, küçük çaplı örnek araştırmaları) değerlendirme aşaması uzun zaman gerektiren araştırmalar ile ilgili ön bilgileri elde etme kolaylığı sağlaması nedeniyle de tercih edilmektedir.

Küçük alan tahmin teknikleri uzun bir süre, İstatistik ve Örneklem teorisyenlerince kabul görmemesi nedeniyle bu tür tekniklerin teorisi hızlı bir ilerleme gösterememiştir. Nüfus Bilimciler için küçük alan tahmin teknikleri kullanışlı ve ihtiyaçlarına cevap verici nitelikte olması nedeniyle, bu tür teknikler nüfus konusunda uzmanlaşmıştır.

Bu çalışmanın ikinci bölümünde, yığın bilgisi ile örnek bilgisini birlikte kullanan ve bir küçük alan tahmin tekniği olan birleşik sentetik tahmin ediciler detaylıca açıklanmıştır. Birleşik sentetik tahmin ediciler, örnek parametresiyle yığın parametresinin ağırlıklı ortalamasını alarak küçük alan tahminleri yapması nedeniyle bu konuya mantıklı ve yapısal çözümler getirmektedir. Yine bu bölümde dört ayrı birleşik sentetik tahmin edici açıklanarak, bunlardan birinin teorik yapısı detaylıca incelenmiştir.

Birleşik sentetik tahmin ediciler, üçüncü bölümde bir nüfus araştırması için derlenen verilere uygulanmıştır. Devlet İstatistik Enstitüsü'nce yapılan nüfus sayımı kalite kontrol anketinden derlenen veriler, çok büyük bir yığından oldukça küçük örnek çekilmesi ve ilgilenilen değişkenin binomiyal bir yapıya sahip olması nedeniyle, tipik bir "küçük alan" özelliği taşımaktadır. Ayrıca, küçük alan olarak kabul ettiğimiz her bir ildeki göç oranının heterojen olması bu verilerin uygunluğunu göstermektedir. Ancak, bu kalite kontrol çalışmasında illerin göç oranı genel olarak büyük değerlere sahip olduğundan, bu durum birleşik sentetik tahminlere de yansımıştır.

Üçüncü bölümde yapılan uygulamada, dört ayrı birleşik sentetik tahmin edicinin hesaplanmasında kullanılan üçüncü

tahmin edicinin W_i ağırlığı, diğer tahmin edicilerin ağırlıklarına nazaran örnek ortalaması $\bar{X}$ 'ya daha fazla ağırlık verdiği görülmektedir.

Küçük alan olarak belirlenen illere ilişkin örnek ortalaması $\bar{X}_i$ büyükten küçüğe sıralandığında, birleşik sentetik tahmin edici değerleri de aynı düzeyde sıralanmaktadır. Örnek ortalamasına göre göç alma oranı en yüksek olan iller 0,21212 ile Ankara ve 0,15916 ile İstanbul olurken, bu iller için 1. birleşik sentetik tahmin değerleri 0,17508 ve 0,14898, 2. birleşik sentetik tahmin değerleri 0,19183 ve 0,15903, 3. birleşik sentetik tahmin değerleri 0,20843 ve 0,16874 ve 4. birleşik sentetik tahmin değerleri 0,21058 ve 0,16921 olarak hesaplanmıştır.

Tablo 3, 4, 5 ve 6'da verilen birleşik sentetik tahmin değerlerinin uygunluğunun kriteri olarak önerilen risk etkinlik fonksiyonu, $\bar{X}$ ortalamaya ilişkin hata kareler ortalamasının birleşik sentetik tahmin ediciye ilişkin hata kareler ortalamasına oranlanması ile hesaplanmaktadır. Elde edilen bu oran bir anlamda birleşik sentetik tahmin edicilerin duyarlılığını örnek ortalaması $\bar{X}$ 'ya yakınlığı ile ölçmektedir. Bu ölçüte göre 1,04366 olarak hesaplanan 2. birleşik sentetik tahmin edici bu veriler için en uygun tahmin edici olarak belirlenmiştir. Ancak, birleşik sente-

tik tahmin ediciler genel olarak yanlış tahminler yapması ve çok küçük örnek çapına dayanması nedeniyle istenen risk etkinlik değerine sahip olması, yanlışlık bakımından önemli bir kazanç sağlamayacaktır.

Bu çalışmada anlatılan teknikler uygulanırken, avantaj ve dezavantajlar gözönüne alınarak istenen amaca uygunluğu tartışılmalıdır.

Önümüzdeki yıllarda, küçük alan tahmin tekniklerinin ve birleşik sentetik tahmin edicilerin istatistiksel teorisine kazandırılacak yeni çalışmalarla, bu tür teknikler daha da gelişerek geniş bir uygulama alanı bulabilecektir.

K A Y N A K L A R

1. Purcell, N.J. Kish, L. "Estimation for small domains."
Biometrics, 35, 365-384, 1979
2. Purcell, N.J. Kish, L. "Postcensal estimates for local areas (or domains)" International Statistical Review
48, 3-18, 1980
3. Bogue, D.J. "A technique for making extensive population estimates." Journal of the American Statistical Association 45, 149-163, 1950
4. Bogue, D.J. and Duncan, B.D. A composite method for estimating postcensal population of small areas by age, sex and colour. National Office Of Vital Statistics, Vital Statistics-Special Reports 47, No.6, 1959
5. U.S. Bureau of the Census. Population estimates and projection. Current Population Reports, Series P-25 No.580, U.S. Government Printing Office, Washington, D.C. 1975
6. National Center For Health Statistics. Synthetic State Estimates of Disability, P.H.S. Publication No.1759 U.S. Government Printing Office, Washington, D.C. 1975

7. Gonzales, M.E. and Hoza, C. "Small area estimation with application to unemployment and housing estimates"
Journal of the American Statistical Association 73,
7-15, 1978
8. Gonzales M.E. and Waksberg, J. Estimation of the error of synthetic estimates. Paper presented at the first meeting of the International Association of Survey Statisticians, Vienna, Austria, 18-25, August 1973
9. Marton, A. " Synthetic estimates for small areas: Problems and results of a simulation experiment"
Statistical Journal of the United Nations ECE 4,
71-80, 1986
10. Purcell, N.J. and Linacre, S. Techniques for the estimation of small area characteristics. Paper Presented at the 3rd Australian Statistical Conference, Melbourne, Australia, 18-20 August, 1976
11. Levy, P.S. The use of mortality data in evaluating synthetic estimates. 1971 Proceedings of the Social Statistics Section, American Statistical Association, 328-331 ,1971

12. Namekata, T. Levy, P.S. and O'Rourke, T.W. Synthetic estimates of work loss disability for each state and the District of Columbia. Public Health Reports 90, 532-538, 1975
13. Gonzalez, M.E. Use and evaluation of synthetic estimates. 1973 Proceedings of the Social Statistics Section American Statistical Association, 33-36, 1973
14. Ghangurde, P.D. and Singh, M.P. "Synthetic estimation in periodic household surveys." Journal of Survey Methodology, Statistics Canada 3, 152-181, 1977
15. Cohen, S.B. A modified approach to small area estimation. Unpublished Ph.D. thesis, University of North Carolina, Chapel Hill, North Carolina, 1978
16. Schmitt, R. C. and Crosetti, A. E. "Accuracy of the ratio-correlation method of estimating postcensal population." Land Economics 30, 279-280, 1954
17. Snow, E.C. "The application of the method of multiple correlation to the estimation of postcensal populations." Journal of the Royal Statistical Society 74, 575-620, 1911

18. Namboodiri, N.K. "On the ratio-correlation and related methods of subnational population estimation." Demography 9, 443-453, 1972
19. Rosenberg, H. "Improving current population estimates through stratification." Land Economics 44, 331-338, 1976
20. Pursell, D.E. "Improving population estimates with the use of dummy variables." Demography 7, 87-91, 1970
21. O'Hare, W. "Report on a multiple regression method for making population estimates." Demography 13, 369-379, 1976
22. Ericksen, E.P. "A regression method for estimating population changes of local areas." Journal of the American Statistical Association 69, 867-875, 1974
23. Nicholls, A. A regression approach to small area estimation. Australian Bureau of Statistics, Canberra, Australia, March 1977
24. Kalsbeek, W.D. A method for obtaining local postcensal estimates for several types of variables. Unpublished Ph. D. thesis, University of Michigan, Ann Arbor, Michigan, 1973

25. Rutherford, J.R. and Krutchkoff, R.G. " The empirical bayes approach : estimating the prior distribution " Biometrika 54, 326-328, 1967
26. Copas, J. " Empirical bayes methods and the repeated use of a standard. " Biometrika 59, 349-360, 1972
27. Griffin, B.S. and Krutchkoff, R.G. " Optimal linear estimators: An empirical bayes version with application to the binomial distribution. " Biometrika 58, 195-201, 1971
28. Manton, K.G. Woodbury, M.A. Stallard, E. et al. A composite estimation model for producing stabilized health rate estimates for small areas using sample surveys: Experience from health surveys in Ethiopia, India and Indonesia. Small Area Statistics: An International Symposium, wiley, New York, 1986
29. Heldal, J. Swensen, A.R. and Thomsen, I. " Census statistics through combined use of surveys and registers. " Statistical Journal of the United Nations ECE 5, 1-9, 1987
30. Schaible, W.L. Choosing weights for composite estimates for small area statistics. To appear in 1978

Proceedings of the Social Statistics Section, American
Statistical Association, 1978

31. Bishop, Y.M.M. Fienberg, S.E. and Holland, P.W. Dis-
crete Multivariate Analysis: Theory and Practice, MIT
Press, Cambridge, 1975
32. Devlet İstatistik Enstitüsü, Genel Nüfus Sayımı için
Kalite Kontrol Anketi, (Yayınlanmamış Araştırma) Ankara
1992

Ö Z G E Ç M İ Ş

1962 yılında Kahramanmaraş'ta doğdu. İlk, Orta ve Lise eğitimini doğum yerinde tamamladıktan sonra, Hacettepe Üniversitesi Fen Fakültesi İstatistik Bölümüne girerek buradan mezun oldu. 1989-1990 Güz döneminde Gazi Üniversitesi Fen Bilimleri Enstitüsü İstatistik Ana Bilim Dalında Yüksek Lisans programına kayıt oldu. 1990 yılında ESOMAR (European Society For Opinion And Marketing Research) tarafından düzenlenen "Veri Hazırlama ve Değerlendirme" (Data Analysis and Interpretation) konulu seminer ile Türkiye ve Orta Doğu Amme İdaresi Enstitüsünce düzenlenen "Araştırma Teknikleri" konulu seminerlere katılarak sertifika aldı. Çalışma hayatına 1981 yılında Ankara Köy Hizmetleri Müdürlüğünde başladı, daha sonra Ankara Özel İdare Müdürlüğünde 8 yıl çalıştı. Halen Adalet Bakanlığı Adli Sicil ve İstatistik Genel Müdürlüğünde İstatistikçi olarak görev yapmaktadır. "Ceza Adaletinde Verimlilik" konulu bilimsel araştırması Verimlilik Dergisi'nde ve bazı hukuk dergilerinde yayınlandı. 1992 yılında Evlendi.