KÜÇÜK ALAN KESTİRİM TEKNİKLERİ VE BOLU İLİ İŞSİZLİK ORANININ KESTİRİMİ ÜZERİNE BİR UYGULAMA

VOLKAN SEVİNÇ
Yüksek Lisans Tezi

İstatistik Anabilim Dalı
HAZİRAN - 1998

Volkan SEVİNÇ'in Yüksek Lisans tezi olarak hazırladığı KÜÇÜK ALAN KESTİRİM TEKNİKLERİ VE BOLU İLİ İŞSİZLİK ORANININ KESTİRİMİ ÜZERİNE BİR UYGULAMA başlıklı tez ..1..7..1998 tarihinde aşağıdaki jüri tarafından Lisansüstü Öğretim Yönetmeliğinin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

Üye (Tez Danışmanı)	: Prof. Dr. Mahir Ulusoy
Üye	: Prof. Dr. Ersoy CANKÜYER
Üye	: Doç. Dr. Embiya AĞAOĞLU

Anadolu Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 10..07..1998. tarih ve ...12/8......... Sayılı kararıyla onaylanmıştır.

ABSTRACT

Master of Science Thesis

SMALL DOMAIN ESTIMATION TECHNIQUES and
AN APPLICATION FOR ESTIMATING THE UNEMPLOYMENT RATE
IN BOLU PROVINCE

VOLKAN SEVİNÇ

Anadolu University

Graduate School of Natural and Applied Sciences

Statistics Program

Supervisor : Prof. Dr. Mahir ULUSOY

1998, Page 69

Demand for estimates of small area characteristics, is increasingly needed. Usual estimation techniques are not very efficient on small areas , so special estimation techniques have been developed for such areas. In this study, these techniques are presented with their data requirements, characteristics, and methodologies. In addition, the performances of the techniques are explained and compared with each other.In the application section Bolu province has been choosen as a small domain, and the unemployment rate in this area has been estimated with three different techniques. The unemployment rate results given by these techniques, are examined and compared with each other.

Keywords: Small Domains, Small Area Estimates, Local Area Estimates, Synthetic Estimates, Postcensal Estimates

TEŞEKKÜR

Bu tezi yazarken bana göstermiş olduğu destek, yardım ve değerli katkılarından dolayı başta Sayın Hocam Prof. Dr. Mahir Ulusoy'a, destek, anlayış ve yardımlarını benden esirgemeyen Sayın Hocam Prof. Dr. Ersoy Canküyer'e, çalışmalarım için gerekli desteği ve kolaylığı sağlayan Sayın Hocam Prof. Dr. Nuray Serter'e, tavsiye, yardım ve sağladığı kaynaklardan ötürü Sayın Hocam Prof. Dr. Öztaş Ayhan'a, gösterdiği yardım ve destekten ötürü D.İ.E.'den Sayın Hasibe Dedeş'e sonsuz teşekkürlerimi sunarım.

Ayrıca her zaman yanımda olan, beni her zaman seven ve destekleyen anneme, babama ve canım kardeşime sonsuz teşekkürlerimi ve sevgilerimi sunarım.

İÇİNDEKİLER

Sayfa

ŞEKİLLER DİZİNİ

ÇİZELGELER DİZİNİ

SİMGELER ve KISALTMALAR DİZİNİ

Y_{hg} :*h*. küçük alan ve *g*. alt grup için ilgili değişkene ait değer

$x'_{.g}$:Büyük alan içindeki g alt grubu için *x* özelliğinin örneğe dayalı kestirimi.

$\bar{x}_h$: *x* özelliği için küçük alan ortalamasının sentetik kestiricisi

$\bar{x}_{.g}$: Büyük alan içindeki g alt grubu için *x* özelliğinin ortalaması kestirimi

X : *x* özelliğinin gerçek değeri

P_{hti} : Her bir *h* küçük alanı için toplam oranlar

Y_{hti} : *t* zamanında, *h*. küçük alanında *i*. Semptomatik veri için değer

q_{ht} : her bir küçük alandaki oranlar

W_s : Ağırlıklar

X_h : Gerçek değer
Y_h : Semptomatik değişkenin değeri
α ve β : Kestirilmesi gereken regresyon katsayıları
ε_h : Rastgele hata terimleri

$\bar{Y}_0$: İlçe nüfusunda ortalama göreceli değişim

A : İkili çiftin *A* tabakası

B : İkili çiftin *B* tabakası

N : Her bir tabakadaki ilçe sayısı

P : Toplam oran

i : Değişken numarası

j : Alt alan numarası

k : Zaman aralığı

B : Regresyon katsayılarının bir px1 matrisi

u : Stokastik hata terimlerinin bir nx1 matrisi

v : Örnekleme hatalarının nx1 vektörü

σ_u^2 : Tabakalanmış birimler arası açıklanmayan varyans

σ_v^2 : Tabaka içi varyans

Y : Özellik değişkeninin gerçek değeri

$\hat{Y}$: Özellik değişkeninin kestirilmiş değeri

n_g : Temel birimlerin g. grubundaki örnek temel birim sayısı

w_{gl} : Ağırlık

$\bar{x}'_{gl}$: g. gruptaki l. temel birim için ortalamanın örnek kestirimi

$\bar{x}'_g$: Temel birimlerin g. grubu için ağırlıklı ortalama

E_ξ : Olasılık dağılımı ξ ' a göre beklenen değer

w_{hig} : Önceden belirlenmiş ağırlıklar.

$i.av.\lambda_{.j}$: $i.$ satır için $\lambda_{.j}$ 'nin ağırlıklı ortalaması

$j.av.\lambda_{i.}$: $j.$ sütun için $\lambda_{i.}$ 'nın ağırlıklı ortalaması

S_{ZZ} : Z'lerin ortalamalar için ayarlanmış, ortalama vektörel-çarpımları

S_{YZ} : Y' nin A ile birlikte, ortalamalar için ayarlanmış, ortalama vektörel çarpımları

I : Birim matris

y_{ij} : $i.$ örnek alanında $j.$örneklenmiş birim için ilgilenilen özellik

$X_{ij} = (x_{ij1} .... x_{ijk})'$: yardımcı değerlere ait bir k vektör

$\beta = (\beta_1 .... \beta_k)$: Bilinmeyen parametrelerin bir k vektörü

n_i : $i.$ küçük alanda gözlenen örneklenmiş birimlerin sayısı $(\sum n_i = n)$

$p{:}\hat{T}$: Bir örnekleme ve kestirim stratejisi

$-\lambda_{i.}, -\lambda_{.j}$: isteğe bağlı Lagrange çarpanları

M.S.E. : Hata kareleri ortalamaları

Y.K.K. : Yapı koruyan kestirim

T.O.U. : Tekrarlı oran uydurulması

H.K.O.K : Hata ortalamaları kareleri

% KGF : Yüzdelik kesin göreceli fark

NIDA : National Institute on Drug Abuse

NCHS : National Center for Health Statistics

S.K.S. : Standart semptomatik kayıt sistemleri

D.İ.E. : Devlet İstatistik Enstitüsü

H.Ü.N.E.E. : Hacettepe Üniversitesi Nüfus Etütleri Enstitüsü

M.L.E. : En büyük olabilirlik tahmini

1. GİRİŞ

Son yıllarda, bir ülke içindeki çeşitli coğrafi alanlarla ilgili istatistiksel veri talebinde önemli bir artış olmuştur. Bu çeşit veriler, yerel yönetimler ve hükümetler tarafından bölgesel politikalar oluşturmak amacıyla veya resmi olmayan kuruluşlarca, yatırım, araştırma gibi nedenlerle, güncel ve geleceğe dönük planlar yapmak amacıyla sıkça kullanılmaktadır.

Küçük alan istatistiklerinin elde edilmesinde çeşitli zorluklar vardır. Bu tür veriler nüfus sayımları veya resmi kayıtlar dışında çoğu zaman elde edilemezler. Örnek araştırmaları geniş alanlar için oldukça güncel ve güvenilir veriler sağlasa da , küçük alanlarda başarılı olamamaktadırlar. Çünkü küçük alanlardan seçilen örneklerin büyüklükleri yetersiz olmakta, hatta seçilen örnek alanlarda, örnek elemanlarının bulunmaması durumuyla karşılaşılabilmektedir. Bu nedenle küçük alan kestirimleri yapmak için özel teknikler geliştirilmiştir.

Küçük alan kestirim teknikleri gerektirdikleri veri, varsayımlar ve uygulanış şekilleri bakımından çeşitlilikler gösterirler. Bu çalışmada, önce küçük alan kavramı ve çeşitlerinin tanımı yapılacak, daha sonra küçük alan kestiriminde kullanılan belli başlı ve yaygın teknikler hakkında bilgi verilecektir.

Ekonomik ve sosyal yönden bakıldığında, işsizlik olgusu bir ülke ya da bölge için oldukça önemlidir. Gerek işsizliğin önlenmesine yönelik politikalar oluşturmakta, gerek yardım programlarının belirlenmesinde söz konusu bölgeye ait işsizlik oranı hakkında fikir sahibi olunması gerekmektedir. Çalışmanın dördüncü bölümünde yer alan uygulamada, Bolu İli bir küçük alan olarak seçilmiş ve bu küçük alandaki işsizlik oranı, elde edilebilen verilerin uygulanmasını olası kıldığı üç ayrı teknikle kestirilmiştir.

Sonuç bölümünde ise, anlatılan teknikler arasında karşılaştırmalar ve yorumlar verilmiştir. Uygulamada elde edilen bulgular ve bunların arasındaki farklar açıklanmıştır.

2. KÜÇÜK ALANLAR

Bu bölümde küçük alan kavramı açıklanacak ve tanımı yapılacaktır. Alanların çeşitli kriterlere göre sınıflandırılması verilecektir.

2.1. Küçük Alan Kavramı

Öncelikle alan kavramının bir tanımı yapılması gerekirse şöyle bir ifade kullanmak olasıdır : Alan, bir ana kütlenin herhangi bir alt bölümü ya da alt kümesi olarak tanımlanabilir [1]. Başka bir tanım ise şu şekilde verilebilir: Alan, bir ana kütlenin, örnekleme yolu ile elde edilen ve kendisi için özel bir kestirime ihtiyaç duyulan bir alt kümesidir [2]. Alanların çeşitli tipleri vardır. Bunlar :

a) Planlı alanlar :

Bu alanlarda ayrı ayrı örnekler alan için planlanmış tasarlanmış ve seçilmiştir. Bunların birleşimi, örneğin bütününü oluşturur. Bunlara örnek olarak ana bölgeler ve diğer ana birimlerden oluşturulmuş katmanlar verilebilir.

b) Çapraz sınıflar :

Çapraz sınıflar , çeşitli kriterlere göre çapraz olarak sınıflandırılmış yani tablolanmış verilerdeki, her bir birim olarak tanımlanabilir. Yaş, cinsiyet, meslek, ve eğitim sınıfları bu birimlere örnek gösterilebilirler.

c) Bölgeler :

Diğer iki tipten daha az kullanılan alan tipi de bölgelerdir. Bölgeler söz konusu alan içindeki daha küçük parçalardır. Bölgelerin seçimi, örnekleme yapılırken dikkate alınmaz. Fakat seçilen örneklerin içinde dağınık halde yer alırlar.

Alanların büyüklüğü, metot seçimini de etkiler bu nedenle alanların bir de büyüklüklerine göre sınıflandırılmalarında fayda vardır. Bu sınıflandırma da şu şekilde yapılabilir :

a) Ana alanlar :

Ana kütle büyüklüğünün yaklaşık 1/10 unu veya daha fazlasını kapsayan alanlardır. Bu alan tiplerine örnek olarak: Ana bölgeler, 10 yaş grubu çocukları veya meslek grupları gibi daha kategorik sınıflar verilebilir.

b) Dar alanlar :

Bu alanlar da ana kütle büyüklüğünün 1/10 ve 1/100 arasındaki kısmını kapsayan alanlardır. Eyalet nüfusları, herhangi bir yaş grubunun nüfusu,meslek-eğitim gibi iki yönlü sınıflandırmalarda bir hücreye düşen nüfus , ya da iş kaybı gibi tek tek küçük sınıflandırmaların oluşturduğu nüfus bu alanlara örnek olarak verilebilir.

c) Mini alanlar :

Bu alanlar ana kütle büyüklüğünün 1/100 ve 1/10000 arasındaki kısmını barındırırlar. Örneğin: ilçe nüfusları, ya da yaş-meslek-eğitim gibi üç yönlü sınıflandırmalar.

d) Seyrek alanlar :

Seyrek alanlar ise ana kütlenin 1/10000 den daha az kısmını barındırırlar. Bunlara örnek olarak da, çeşitli etnik gruplara göre sınıflandırılan sağlık servis alanları veya yerel yerleşim birimlerine göre sınıflandırılmış belirli kronik hastalıkları olan bireyler verilebilir.

Bu sınıfların sınırları kesin olarak alınmamalıdırlar. Bu sınırlar örnek ve ana kütle büyüklüğüne değişken ve istatistiklere, ölçüm ve kararlara göre değişebilirler. Fakat bunlar bize bu sınıflar arasındaki uygulama farklılıklarını göstermek açısından yararlıdırlar.

Alan kavramı, planlı alanlar kavramı ile sınırlandırılıp, bu Birleşmiş Milletler [3] planlı alanlar tanımı ile son bir kez daha tanımlanacak olunursa : Hakkında kesin bir rakamsal bilgi sağlamak için araştırma planlanan herhangi bir alt alan, bir çalışma alanı olarak adlandırılabilir.

3. KÜÇÜK ALAN KESTİRİM TEKNİKLERİNDE DOĞRUDAN ve DOLAYLI KESTİRİCİLER

Bu bölümde, küçük alanlara ait çeşitli karakteristiklerin kestiriminde kullanılan yöntemler teknikleri ve kullandıkları veri tipi ile birlikte açıklanacaktır.

Araştırma örneklemesinde amaç, bilinmeyen bir kütle karakterini kestirmektir. (kütle ortalaması, toplamı vb.). Örnekler sadece ana kütle karakterini kestirmek için değil, alt kütle ya da alanlar için kestirimler yapmak amacı ile de tasarlanırlar. Bu amaçla kullanılan kestiriciler eğer ilgilenilen değişken değerlerini sadece alan içindeki birimlerden ve zaman aralığından alıp kullanıyorsa, bunlar doğrudan kestiriciler olarak adlandırılırlar. Dolaylı kestiriciler ise, alandan ve/veya alan dışındaki zaman aralığından ve ilgilenen zamandan alınmış değişken değerlerini kullanırlar.

Doğrudan kestiriciler için çeşitli kaynaklara göre kullanılan çeşitli isimler vardır. Royall [4] bu kestiricileri, 'doğrudan', Gonzalez [5], 'yansız', Holt, Smith et al. [6], 'standart', Gonzalez [7], 'geçerli', Kalton [8], 'örnek-bazlı' olarak adlandırmışlardır. Doğrudan kestiriciler, geleneksel kestirim teknikleri ile elde edilen kestiricilerdir. Bu tür kestiriciler, sadece ilgilenilen alandan alınılan veriye dayanırlar. Bazen bu verilere nüfus sayımı veya resmi kayıtlardan elde edilen yardımcı veriler eklenebilir. Doğrudan kestiriciler, küçük alan kestirimlerinde genelde yetersiz, ve küçük örnek büyüklüğü yüzünden büyük varyanslıdırlar.

Dolaylı kestiriciler de çeşitli adlar altında anılırlar. En çok kullanılan terim, küçük alan kestiricileridir. Çünkü bu kestiricilerin bir çoğu, coğrafi alanlar için kestirimler yaparlar. Fakat küçük sözcüğü pek doğru değildir. Küçük olan alanın ya da alan içindeki kütlenin küçüklüğünden ziyade, örnek gözlem sayısının azlığı ve aranılan standart doğrudan kestiricilerin büyük varyanslı olmasıdır. Aynı şekilde alan kelimesi de tam doğru değildir; şöyle ki: Bu teknikler herhangi isteğe bağlı bir bölgeye uygulanabilir. Bu bölgenin de bir coğrafi bölge olarak tanımlanması gerekmez. Bu kestiricileri tanımlamak için, Ericksen [9], 'yerel alan', Purcell and Kish [1], 'küçük alan', Laake [10], 'alt alan', Holt, Smith et al. [6], 'küçük alt gruplar', Dalenius [11], 'dolaylı', Särndal [12], 'model-bağımlı' terimlerini kullanmışlardır. Sentetik kestiriciler terimi de bu dolaylı tip kestiricilerin tanımı için National Institute on Drug Abuse (NIDA) [13], ve özel bir dolaylı kestiricinin tanımı için National Center for Health Statistics (NCHS) [14] tarafından kullanılmıştır. Araştırma yapan kişiler, bazen dolaylı kestiricileri model bazlı kestiriciler olarak da adlandırırlar. Fakat bu isim nadir kullanılır. Bazen doğrudan kestiricileri tanımlamak için bile kullanılır. Dolaylı kestiricilerin üç tipi vardır :

a) Alan dolaylı kestirici :

Bu tür kestiriciler, ilgilenilen değişken değerlerini başka bir alandan alıp kullanırlar ama başka bir zaman aralığından almazlar.

b) Zaman dolaylı kestirici :

Bu tür kestiriciler ise, ilgilenen değişken değerlerini başka bir zaman aralığından alırlar ama, başka bir alandan alıp kullanmazlar.

c) Zaman ve alan dolaylı kestirici :

Bu üçüncü tip dolaylı kestiriciler ise, ilgilenilen değişken değerlerini hem başka bir alandan, hem de başka bir zaman aralığından sağlarlar.

Doğrudan veya dolaylı kestiricilerin yukarıda yapılan tanımlarının hiçbirinin, alan ya da zaman aralığının dışından yardımcı değişkenlerin kullanılıp kullanılmamasına göre değişmediğine dikkat edilmelidir.

Bütün dolaylı kestiriciler verilen bir analiz için eşit olarak uygun olmazlar. Örneğin, eğer analizin amacı verilen bir zaman aralığı içinde alanlar arasında karşılaştırma yapmak ise, alan dolaylı kestiricilerden ziyade, zaman dolaylı kestiricilerden birini seçmek daha doğru olur. Alan dolaylı kestiriciler, bazı model parametrelerine göre ilgilenilen değişkenlerin ortalamasının, alanlar arasında aynı olduğunu varsayan modellere dayanırlar. Fakat zaman dolaylı kestiriciler, bu değişkenlerin, aynı şekilde zaman aralıkları boyunca aynı olduğunu varsayan modellere dayanırlar. Pratikte alan ve zaman dolaylı kestiricilerin performansı, elde ne kadar bilgi olduğuna ve bu bilgiyle kullanılacak modelin gerçek ihtiyaca ne kadar cevap verebileceğine bağlıdır. Dolaylı kestirim tekniklerinin belli başlı çeşitleri şunlardır :

3.1. Semptomatik Kayıt Teknikleri

Bu teknikler, kayıt sistemleri ile eski sayım verileri arasındaki istatistiksel ilişkilere dayanır. Standart semptomatik kayıt sistemleri (S.K.S.), küçük alan kestirim tekniklerinin en eskisidir. Çeşitli kayıtları kullanan bu tür tekniklerin çoğu ,U.S. Bureau of the Census [15] tarafından önerilmiş ve test edilmiştir. Bu tekniklerin bir çoğu toplam nüfusun küçük alan kestirimleriyle ilgilenir. Bu tekniklerin çeşitleri :

3.1.1. Nüfus sayımı bileşenleri teknikleri

Bu bileşenlerin iki çeşidi vardır. İkisi de okul çocukları için göç oranının, toplam nüfus içindeki orana eşit olacağı varsayımından hareket ederler. fakat bu metotların farklılaştığı nokta ise, okul çocukları arasındaki göç oranının kestiriliş şeklidir. U.S. Bureau of the Census [16], bu metodu, genellikle sivil göç miktarını hesaplamak için kullanmıştır.

3.1.2. Gerçek oranlar tekniği

Küçük alan nüfus büyüklüklerinin kestirimi için başka bir teknik de bu tekniktir. İlk tanımı Bogue [17] tarafından verilmiştir. Bu yaklaşım, geniş alan ölüm ve doğum oranlarını ve nüfussal-oran işlemini kullanarak, gereken yerel alan nüfusu kestirimleri yapar. Bu tekniğin dayandığı varsayım, geniş alanlardaki ölüm ve doğum oranlarının küçük alanlarla aynı oranda değiştiğidir.

Tekniğin uygulaması şöyledir :

a) Bir yıla ait nüfus sayımı verileri temel alınarak, küçük alan ve ana alan için kaba ölüm ve doğum oranları hesaplanır.

b) Şu oranlar bulunur :

$$\text{Ölüm oranı} = \frac{\text{Küçük alan ölüm oranı}}{\text{Ana alan ölüm oranı}} \qquad (3\text{-}1)$$

$$\text{Doğum oranı} = \frac{\text{Küçük alan doğum oranı}}{\text{Ana alan doğum oranı}} \qquad (3\text{-}2)$$

c) Ana alandaki gerçek oranların eğilimine bağlı olarak, yakın geçmişte kaba ölüm ve doğum oranlarının nasıl geliştiği araştırılır. Eğer herhangi bir alt alan için zamana bağlı olarak kesin bir oran değişikliği gösterisi varsa, kestirim yılı için düzeltilmiş oranları elde etmek amacıyla, temel yılın oranlarına uygulamak için bir düzeltme faktörleri seti oluşturulur. Eğer düzeltme faktörleri bulunamazsa, ana alan için gerçek oran temel yıl için düzeyi aştığında, "e" kısmında bir sabit oranla çarpmak yerine bir sabit fark ekleme metodu kullanılmalıdır.

d) Nüfus kestirimi bulunmak istenen yıl için, ana alan içinde kaba ölüm ve doğum oranları hesaplanır. (Ana alan, kendisi için daha önce düzenli kestirimler yapılan bir eyalet veya diğer bir birim olarak seçilmelidir.)

e) Kaba ölüm ve doğum oranlarını şu formül uygulanır :

Alt alan için bir kestirilmiş oran = Ana alanda istenen yıldaki kaba oran x bilinen bir faktör. (3-3)

Bilinen bir faktör "c" de tanımlanmıştır.

f) Formül kestirilen kaba doğum ve ölüm oranlarına uygulanır :

Güncel nüfusun doğru kestirimi

$$= \frac{\text{Her bir alt alan için kestirilmiş kaba oran}}{\text{Alt alanda yıl içinde kaydedilmiş doğum sayısı}} \qquad (3\text{-}4)$$

g) Tek bir ana kütle kestirimi oluşturmak için, iki kestirimin ortalaması alınır.

3.1.3. Karma metot

Bu metot, Bogue and Duncan [18] tarafından geliştirilmiştir ve aynı zamanda gerçek oranlar tekniğine bir alternatiftir. Bu metot da yerel alan nüfusunu ayrı yaş gruplarına böler ve bu grupların her biri için ayrı ayrı nüfus kestirimleri bulur. Bunları bulmak için, her bir gruba en uygun teknik ve veriyi kullanır. Bu tek tek elde edilen kestirimler toplanıp, sonuç olarak bir yerel alan kestirimi elde edilir.

3.1.4. Konut birimi metodu

Bu metot, yerel alandaki konut sayısının güncel kestirimini kullanır ve her konuttaki ortalama insan sayısını hesaba katar. Dayandığı varsayım ise, konut sayısındaki değişimin, nüfustaki değişimi yansıtacağıdır. U.S. Bureau of the Census [19] bu metodu tartışmış ve düzenleyerek bir karma metot oluşturmuştur. Rives [20], konut birimi metodunu daha da geliştirmiştir.

3.1.5. Resmi veri kayıtları metodu

Bu metot, semptomatik kayıt tekniklerinin en yenisidir. U.S. Bureau of the Census [21] tarafından kullanılmıştır. Metot, federal gelir paylaşımı programı için gereken, bütün gelir paylaşım alanlarındaki nüfusun güncellenmesi için geliştirilmiştir. Temelde nüfus sayımı bileşenleri ile aynıdır. Ayrıldıkları nokta, göç sayısının gelir servisi tarafından saptanmasıdır. Burada veri sınıflarından çok kişisel kayıtların kullanılması, çok küçük alanlar için kestirim yapılmasına izin verir. Fakat federal kullanımla sınırlıdırlar. Bu metodun detaylı bir tanımı Starsinic [22] tarafından verilmiştir.

3.2. Sentetik Kestiriciler

Sentetik kestiricilerin bulunmasında kullanılan varsayım, küçük bir alanın kendisini kapsayan büyük alana benzer olduğudur. Sentetik kestiriciler, nüfus sayımı ve resmi kayıtlardan yardımcı veri alarak da bulunabilirler. Eğer sentetik kestiriciler için gerekli söz konusu varsayım ilgilenilen alan için geçerli değilse, o zaman sentetik kestiriciler oldukça yanlı ve tutarsız olabilirler. Fakat genelde varyansları düşüktür. Bütün dolaylı kestiriciler gibi, sentetik kestiriciler de alan dolaylı, zaman dolaylı,ya da zaman ve alan dolaylı olabilir. Sentetik kestiricileri isminin kullanılmasının nedeni, bu kestiricilerin direkt olarak araştırma sonuçlarından elde edilmemesidir. Fakat şu anda bu isim, kestirimlerin güvenirliğini arttırmak için öbür benzer alanlardan veri ödünç alan metot anlamında kullanılmaktadır.

Küçük alan h içindeki x özelliği için sentetik kestirim :

$$\hat{x}_h = \sum_g \hat{x}_{hg} = \sum_g (Y_{hg} / Y_{.g}) x'_{.g} \qquad (3\text{-}5)$$

Y_{hg} : h. küçük alan ve g. alt grup için ilgili değişkene ait değer.

$x'_{.g}$: Büyük alan içindeki g alt grubu için x özelliğinin örneğe dayalı kestirimi.

Burada noktalı indisler, o indisin tümünün toplanmış halini göstermektedir.

Genellikle ilgili değişken Y, insan sayısı toplamı olarak alınır. Bu nedenle

$$Y_{hg} = N_{hg} \qquad (3\text{-}6)$$

alınır ve sentetik kestirici şu hale gelir :

$$\hat{x}_h = \sum_g (N_{hg} / N_{.g}) x'_{.g} \qquad (3\text{-}7)$$

Sentetik kestirici, oranlamalar yolu ile bulunan kestiricilerin, g sayıdaki bir grup için genişletilmiş hali olarak görülebilir. Buna ek olarak ayrı ayrı küçük alanlar üzerinde toplandığında, büyük alan kestirimlerine uyum sağlar :

$$\sum_h \hat{x}_h = \sum_h \sum_g (Y_{hg} / Y_{.g}) x'_{.g} = \sum_g \sum_h (Y_{hg} / Y_{.g}) x'_{.g} = \sum_g x'_{.g} = x'_{..} \qquad (3\text{-}8)$$

Sentetik kestirici (3-5) ilk olarak Purcell and Linacre [23] daha sonra Ghangurde and Singh [24] tarafından önerilmiştir. Fakat bu kestirici, National Center for Health Statistics (NCHS) [14] tarafından önerilen ve Gonzales [5] tarafından bulunan kestirici ile aynı değildir. Söz konusu kestiricinin kalıbı şudur :

$$\bar{x}_h = \sum_g (Y_{hg} / Y_h) \bar{x}_{.g} \qquad (3\text{-}9)$$

$\bar{x}_h$: x özelliği için küçük alan ortalamasının sentetik kestiricisi

$\bar{x}_{.g}$: Büyük alan içindeki g alt grubu için x özelliğinin ortalaması kestirimi

Genel olarak (3-9), (3-5)'in yaptığı gibi, uygun yansız geniş alan kestirimini oransal olarak eklemez. Her iki kestirici de Y_{hg} 'yi kullanır. Fakat (3-9) bunu küçük alan içinde alt gruplar için oran olarak kullanırken, (3-5) bu değeri alt gruplar içindeki küçük alanların oranı için kullanır.

Sentetik kestiriciler varyansı azaltırlar, fakat iki yönden yanlıdırlar. Birincisi ölçümlerin homojenliği varsayımından daima sapmalar olacaktır. İkincisi ise $Y_{hg} / Y_{.g}$ ağırlıkları genellikle geçmiş veriye aittir. Ve bu zaman sürecinde nüfusun yapısı değişmiş olabilir. (3-5) kestiricisinin yanlılık ifadesine bakacak olursak :

$$E[\hat{x}-X_h]=E\left[\sum_g (Y_{hg}/Y_{.g}=)x'_{.g}-\sum_g X_{hg}\right]$$

$$=\sum_g Y_{hg}(X_{.g}/Y_{.g}-X_{hg}/Y_{hg}) \qquad (3\text{-}10)$$

X: x özelliğinin gerçek değeri

Bu nedenle,

$$X_{.g}/Y_{.g}=X_{hg}/Y_{hg} \qquad (3\text{-}11)$$

olmadıkça, sentetik kestirici, X_h için yanlıdır. Bu durum da genelde pek olmaz. Dolayısıyla her zaman yanlılık olasılığı kuvvetlidir.

Sentetik kestiricilerin oluşturulması, yanlılık tehlikesi nedeni ile karmaşıktır. Bu nedenle hata kareleri ortalamalarının kestiriminde dikkat etmek gerekir. Ayrıca bunların kestirimi kolay değildir. Çünkü hata terimlerini kestirmek için gereken gerçek X_{hg} değerleri bilinmemektedir. Bu problem küçük alan düzeyi x'_{hg}'de bazı yansız örnek kestirimleri oluşturarak aşılabilir. Bu örneklerin büyük varyansları olması da kabul edilebilir. Fakat küçük alan varyans kestirimleri sağlam değildir. Bu problemin olası bir çözümü olarak Gonzalez and Waksberg [25] ilgilenen küçük alana ait ortalama hata karelerinin, ortalamalarının hesaplanmasını önerdiler. Önerdikleri hata kareleri ortalamalarının ortalamaları şu şekilde bulunmaktadır :

$$\sum_h E(\hat{x}_h-X_h)^2/H \qquad (3\text{-}12)$$

H: Küçük alan sayısı

x'_{hg}' nin kestiriminin yanlı olması tehlikesini içermeyen daha farklı bir yaklaşım ise, ilgilenilen x değerinin nüfus sayımında toplandığını varsayarak, hesaplamada nüfus sayımı verilerinin kullanılmasıdır.

3.3. Regresyon-Semptomatik İşlemler

Bu işlemlerin ortak noktaları, semptomatik veriler ve ilgilenilen veriler arasında bir fonksiyonel ilişkiye, yani bir en küçük kareler regresyon ilişkisine dayanmalarıdır. Böyle bir ilişki kurulabildiği taktirde, daha önce uydurulmuş model üzerinde en son bilgi kullanılarak küçük alan kestirimleri yapılabilir.

3.3.1. Oran-korelasyon metodu

Schmitt and Crosetti'nin [26] oran-korelasyon metodunu sunmaları ve düzeltmelerinden bu yana, çoklu regresyon, küçük alan kestirimi yapmada önemli bir rol oynamıştır. Oran korelasyon tekniği, ilk olarak Snow [27] tarafından önerilmiştir. Metot basitçe şu şekilde uygulanır :

a) Her bir p tane Yi semptomatik veri için, t=1,2 temel zaman aralığı (nüfus sayımı yılları) ve kestirim yılı t=3 baz alınarak her bir h küçük alanı için toplam oranlar hesaplanır:

$$P_{hti} = Y_{hti} \,/\, \sum\nolimits_h Y_{hti} \tag{3-13}$$

P_{hti} : Her bir h küçük alanı için toplam oranlar

Y_{hti} : t zamanında, h. küçük alanında i. Semptomatik veri için değer

Bu Y semptomatik değişkenlerine örnek olarak, doğum-ölüm kayıtları, istihdam kayıtları,seçmen kayıtları, okul kayıtları verilebilir.

b) Benzer olarak, ilgilenilen x değişkeni için her bir alt alandaki oranlar hesaplanır. Bu hesaplama t_1 ve t_2 zamanları için yapılır :

$$q_{ht} = x_{ht} \,/\, \sum\nolimits_h x_{ht} \tag{3-14}$$

q_{ht} : her bir küçük alandaki oranlar

c) Daha sonra iki oran seti hesaplanır :

Birinci set, 1. ve 2. zaman arasındaki değişiklikleri yansıtan oranlar setidir. Bu oranlar :

$$r_{hai} = P_{h2i} \,/\, P_{h1i} \qquad R_{ha} / = q_{h2} \,/\, q_{h1} \qquad i = 1,2,\ldots\ldots\ldots,p \tag{3-15}$$

İkinci oranlar seti semptomatik veriler arasında 2. ve 3. zaman aralığındaki değişikliği göstermek için kurulur. İkinci oranlar :

$$r_{hbi} = P_{h3i} \,/\, P_{h2i} \qquad i = 1,2,\ldots\ldots\ldots,p \tag{3-16}$$

d) Çoklu regresyon kullanılarak (3-15)'de verilen oran değişkenleriyle model uydurulur :

$$R_{ha} = B_0' + B_1' r_{ha1} + \ldots\ldots\ldots + B_p' r_{hap} \tag{3-17}$$

$B_0', B_1', \ldots\ldots\ldots, B_p'$: İki zaman aralığı arasındaki değişimlerden kestirilen regresyon katsayıları

e) Nüfus sayım yılı oranları arasındaki deneysel ilişkilere dayanan $B_0', B_1', \ldots\ldots\ldots, B_p'$ katsayıları, daha sonra nüfus sayım yılı dışındaki nüfus oranlarını kestirmek için, (3-16) 'da semptomatik değişkenler için verilen nüfus sayım yılı sonrası oranlarıyla kullanılırlar. Kestirilen oranlar şu şekilde verilebilir :

$$R'_{hb} = B'_0 + B'_1 r_{hb1} + \dots\dots + B'_p r_{hbp} \quad (3\text{-}18)$$

f) İlgilenilen değişken değeri ve bunun bütün küçük alanlar üzerinde toplanmış güncel değeri için kestirilen R'_{hb} oranları, bunların gerçek küçük alan oranlarıyla çarpılmasıyla gerçek sayılara dönüştürülebilirler :

$$\hat{x}_{h3} = R'_{hb}(x_{h2} / \sum_h x_{h2}) \sum_h x_{h3} \quad (3\text{-}19)$$

Bu metot için gereken semptomatik bilgi, diğerleri için gereken semptomatik bilgiden oldukça fazladır.

3.3.2. Teorik bir başlangıç noktasından oran-korelasyon metodu

Bu metot, Namboodiri [28] tarafından geliştirilmiştir. Metodun uygulanması kısaca şu şekildedir :

Öncelikle elde bir (y, x) örnek çifti vardır. İkinci örneğe (y', x') adı verilsin. Böylece birinci örnekle karışmayacaktır. $\bar{y}$ ve $\bar{x}$ sırasıyla y ve x' in ortalamaları olsun.

$$b = \sum (y - \bar{y})(x - \bar{x}) / \sum (x - \bar{x})^2 \quad (3\text{-}20)$$

y' nin x üzerine doğrusal regresyonunun eğimi olsun. Aynı şekilde (y', x')'nün ortalamaları, sırasıyla $\bar{y}'$ ve $\bar{x}'$ olsun ve y' nin de aynı şekilde x' üzerine regresyonunun eğimi b' olacaktır.

İki örneğin de aynı ana kütleden rastgele seçildiği varsayılsın. Bu koşullar altında $(y' - \bar{y}')$ değerlerinin kestiriminde şu formüller kullanılabilir :

$$(y' - \bar{y}') \text{'nün kestirimi} = (\bar{y} / \bar{x})(x' - \bar{x}') \quad (3\text{-}21)$$

$$(y' - \bar{y}') \text{'nün kestirimi} = b(x' - \bar{x}') \quad (3\text{-}22)$$

Bu problem daha genel olarak çok boyutlu duruma taşınabilir. Şöyle ki :

$(y_1, x_1, x_2, \dots\dots\dots, x_m)$ şeklinde çoklu uzaydan bir örnek vardır. İstenen şey bu örnekteki bilgiyi kullanarak, başka örneklerdeki y değerini, verilen $(x_1, x_2, \dots\dots\dots, x_m)$ değerlerini kullanarak bulmaktır.

İkinci örneğe aynı şekilde $(y', x'_1, x'_2, \dots\dots\dots, x'_m)$ adını verilsin. Birincisi de $(y, x_1, x_2, \dots\dots\dots, x_m)$ olsun. $\bar{y}, \bar{x}_1, \dots\dots\dots \bar{x}_m$, $y, x_1, \dots\dots\dots, x_m$'in sırasıyla ortalamaları olsun. Aynı şekilde $\bar{y}', x'_1, \dots\dots\dots, \bar{x}'_m$ de ikinci örneğin uygun ortalamaları olsun. ($\bar{y}'$'nün değeri bilinmeyebilir). b_s birinci örnekte y' nin x_s üzerine doğrusal regresyonun eğimi olsun. B_s de aynı örnekte y' nin $x_1, x_2, ..., x_s$ üzerine çoklu regresyonunda x_s'nin katsayısı olsun. Şimdi eğer iki örneğin de aynı kütleden rastgele çekildiği varsayılırsa, iki boyutlu olan daha önceki kestirimler, çok boyutlu düzeye taşınabilir. (3-21) ve (3-22)'ye benzer şekilde :

$$(y' - \bar{y}')\text{'nün kestirimi} = (\bar{y} / \bar{x}_s)(x_s' - \bar{x}_s'), \qquad s = 1,2,........,m \qquad (3\text{-}23)$$

$$(y' - \bar{y}')\text{'nün kestirimi} = (y' - \bar{y}') = b_s(x_s' - \bar{x}_s'), \qquad s = 1,2,........,m \qquad (3\text{-}24)$$

İki boyutlu durumdan farklı olarak, yukarıdaki formüllere ek olarak şu formül uygulanabilir :

$$(y' - \bar{y}')\text{'nün kestirimi} = \sum_{s=1}^{m} B_s(x_s' - \bar{x}_s') \qquad (3\text{-}25)$$

Bu formülde çoklu regresyon katsayıları kullanılmaktadır. (3-23)' deki kestirimler kullanılarak, iki tane karma kestirim oluşturulabilir :

$$A_I = (1/m)\sum_{s=I}^{m} (\bar{y} / \bar{x}_s)(x_s' - \bar{x}_s') \qquad (3\text{-}26)$$

ve

$$D_I = \sum_{s=I}^{m} W_s(\bar{y} / \bar{x}_s)(x_s' - \bar{x}_s') \qquad (3\text{-}27)$$

Burada,

$$\sum_{s=1}^{m} W_s = 1 \qquad (3\text{-}28)$$

W_s : Ağırlıklar

W_s'lerin kestiriminin bulunması gerekir. Aynı şekilde $(y' - \bar{y}')$'nün m kestiriminden (3-24) tarafından sağlanan şu iki karma kestirim hesaplanabilir :

$$A_{II} = (1/m)\sum_{s=1}^{m} b_s(x_s' - \bar{x}_s') \qquad (3\text{-}29)$$

$$D_{II} = \sum_{s=1}^{m} W_s^* \, b_s(x_s' - \bar{x}_s') \qquad (3\text{-}30)$$

Burada

$$\sum_{s=1}^{m} W_s^* = 1 \tag{3-31}$$

koşulu vardır. W_s^* ağırlıklarının gerçek değerleri belirlenmelidir. W_s ve W_s^* (s =1,2,........,m) ağırlıklarını bulmak için,

$$\sum\left[(y-\bar{y})-\sum_{s=1}^{m} W_s(\bar{y}/\bar{x}_s)(x_s-\bar{x}_s)\right]^2 \tag{3-32}$$

ifadesinde W_s' ler, bu ifadeyi minimum yapacak şekilde bulunurlar. Burada W_s' ler için koşul :

$$\sum W_s = 1 \tag{3-33}$$

olmasıdır. Aynı şekilde W_s^*'ler de

$$\sum W_s^* = 1 \tag{3-34}$$

koşulu altında

$$\sum\left[(y-\bar{y})-\sum_{s=1}^{m} W_s^* b_s(x_s-\bar{x}_s)\right] \tag{3-35}$$

ifadesini minimize edecek şekilde bulunurlar. W_s ve W_s^* değerlerinin bulunuşu EK-1 'de verilmiştir [28].

y' değerlerinin kestirimi, bu değerlerin ortalamalarından sapmalarını alan bir yaklaşımla bulunabilir. Bu sapma değerleri kestirildiğinde, y' değerleri kolayca elde edilebilirler. Bu yaklaşım şu formülle açıklanabilir :

y' 'nün kestirimi = $(y'-\bar{y}')$ 'nün kestirimi + $\bar{y}'$ 'nün kestirimi (3-36)

$\bar{y}'$ 'nün olası bir kestirimi $\bar{y}$ 'dur. $\bar{y}$ orijinal örnekteki y değerlerinin ortalamasıdır. y' 'nün regresyon katsayısı ve x'_1,........,x'_m değerlerine bağlı olarak kestirimi ise :

y' 'nün kestirimi = $B_0 + \sum B_s x'_s$ 'dir. (3-37)

Çoklu regresyon teorisinden ,

$$B_0 = \bar{y} - \sum B_s \bar{x}_s \tag{3-38}$$

olduğu bilinmektedir. Bu durumda (3-37) şöyle yazılabilir :

$$y' \text{'nün kestirimi} = \bar{y} - \sum B_s \bar{x}_s + \sum B_s x'_s \tag{3-39}$$

Bu da şu ifadeyle eşit olacaktır :

$$y' \text{'nün kestirimi} = \bar{y} - \sum B_s(\bar{x}_s - \bar{x}'_s) + \sum B_s(x'_s - \bar{x}'_s) \tag{3-40}$$

3.3.3. Tabakalama yolu ile güncel kütle kestirimleri geliştirilmesi

Rosenberg [29] oran-korelasyon metodunun bir genişlemesi olarak bu yaklaşımı sunmuştur. Bu yaklaşımın dayandığı nokta, aynı tabaka içinde farklı ilişkilere dayanan kestirimlerdir. Buradaki hipotez, sosyal ve ekonomik çeşitlilikler gösteren bir coğrafi alan içinde, örneğin tarımda, istihdam yüzdesi gibi, bazı demografik özelliklere göre tabakalanmış ilçelere, regresyon analizi uygulayarak, güncel kestirimler yapılabileceğini ileri sürer. Çünkü tabakalama çok daha fazla güvenilir güncel kütle kestirimleri yapmaya olanak verir [30], [31].

Oran-korelasyon metodunun başarısı, nüfus sayımında ortalama hata ile ölçülüp görüleceği gibi, büyük nüfuslu ilçeler için büyük, küçük nüfuslular için düşüktür. Buna ek olarak regresyon doğrularının geçerliliğinin ölçümü için birkaç kriter vardır. Bunlar nüfus değişiminin kestiricileri olarak : Havuzlu-korelasyon katsayısı (pooled correlation coefficient) ve havuzlu göreceli regresyon (pooled relative regression error) hatasıdır. Bu kriterler, sırasıyla herbir ikili çift için hesaplanan şu kestiricilerle bulunur :

$$R_{0.1245}(havuzlu) = \sqrt{\frac{N_A R^2{}_{0.1245(A)} + N_B R^2{}_{0.1245(B)}}{N_A + N_B}} \tag{3-41}$$

$$RS_{0.1245}(havuzlu) = \frac{N_A \sqrt{\dfrac{S^2{}_{0.1245(A)}}{\bar{Y}^2{}_{0(A)}}} + N_B \sqrt{\dfrac{S^2{}_{0.1245(B)}}{\bar{Y}^2{}_{0(B)}}}}{N_A + N_B} \tag{3-42}$$

$S^2{}_{0.1245}$: Kestirimin varyansı

$\bar{Y}_0$: İlçe nüfusunda ortalama göreceli değişim

A : İkili çiftin A tabakası

B : İkili çiftin B tabakası

N : Her bir tabakadaki ilçe sayısı

Bu aşamadan sonra cevaplanması gereken bir kaç soru olabilir. Örneğin, tabakaların sayısı kaç olmalıdır? Tabakaların minimum büyüklüğü ne kadar olmalıdır? Ya da tabakalama için en etkili kriter nedir? Bazen tabaka için ayrı bir regresyon kestirimi kullanmak uygundur. Eğer regresyonların doğrusal olduğundan emin olunuyorsa, ve eğer B_h'ler (h tabakası için regresyon katsayısı) bütün tabakalarda aynıysa, karma kestirici tercih edilmelidir. Eğer regresyon doğrusal ise, (böylece yanlılık tehlikesi küçülür) ama B_h'ler tabakadan tabakaya değişiyorsa, ayrı kestirim bulma yoluna gidilmelidir [32].

3.3.4. Fark-korelasyon metodu

Bu metot O'Hare [33] tarafından sunulmuştur. Bu metodun farklı tarafı, zaman içinde değişiklikleri yansıtmada kullanılan değişkenlerin oluşturulmasındadır. Buna karşılık oran-korelasyon ve fark-korelasyon metodunun izlediği yöntem oldukça benzerdir. Her iki metot da her alt alan için bakılan, geçmiş zaman içindeki iki farklı noktadaki nüfus ölçümlerini ve semptomatik veriyi buna ek olarak kestirim yılındaki semptomatik veri ölçümlerini kullanır. Semptomatik veri okul kayıtları, araba kayıtları, ölüm-doğum kayıtları gibi nüfustaki değişimleri yansıtabilecek değişkenler olabilir.

Her bir gözlem zamanındaki alt alan için, her bir değişkenin gözlem değeri, o tarihteki alan toplamı oranına dönüştürülür. Bu oranlar şu şekilde belirtilirler :

$$P_{ijk} \qquad (3\text{-}43)$$

P : Toplam oran

i : Değişken numarası

j : Alt alan numarası

k : Zaman aralığı

$k=1$: 1. zaman

$k=2$: 2. zaman

$k=3$: 3. zaman

P_{ijk} ile gösterilen oranlar, hem oran-korelasyon metodu hem de fark-korelasyon metodunda kullanılır. Fakat bir sonraki adım, iki metot arasındaki farkı gösterir. Bu da, zamandaki değişikliği yansıtan değişkenin oluşturulmasında ortaya çıkar.Fark korelasyon metodu, bir zaman içindeki iki noktaya ait oranları kullanarak oranlama yapmaktan ziyade, bu oranların farkını alarak çalışır.

1. ve 2. zamanlardaki değişimleri yansıtmada kullanılan farkların birinci kümesi şu olsun :

$$D_{ijm} = P_{ij2} - P_{ij1} \tag{3-44}$$

P_{ijk} : Toplam oran

D : Fark

i : Değişken numarası

j : Alt alan numarası

m : Bu farkların 1. ve 2. zaman arasındaki değişiklikleri yansıttığını gösterir.

2. ve 3. zamanlar arasındaki semptomatik göstergelerdeki değişiklikleri yansıtmak için kullanılan farkların ikinci kümesi de oluşturulabilir. 3. zamanın kestirim tarihi olduğu hatırlanmalıdır. Bu tarihte alt alanın nüfusu bilinmemektedir. Bu ikinci küme de şudur :

$$D_{ijm'} = P_{ij3} - P_{ij2} \tag{3-45}$$

P_{ijk} : Toplam oran

D : Farklar

i : Bütün semptomatik göstergelerin değişken numarası

j : Alt alan numarası

m' : Bu farkların 2. ve 3. zaman arasındaki değişiklikleri gösterdiğini belirtir.

Bundan sonraki adım, oranların farklarla yer değiştirilmesi dışında oran-korelasyon metodundaki adımla oldukça benzerdir. 1. ve 2. zaman arasındaki değişiklikleri yansıtan farkları değişken olarak kullanarak, bir çoklu regresyon analizi oluşturulduğunda, alt alandaki nüfus oranlarındaki farkları kestirmek için, 2. ve 3. zaman arasındaki değişimleri yansıtan semptomatik göstergelerdeki farklar ile, bu analizden çıkarılan katsayı vektörü kullanılır. 3. zaman nüfus oranı eksi 2. zaman nüfus oranı farkı kestirildiğinde, bunlar gerçek nüfus oranlarına dönüştürülebilirler. Toplam alan nüfusunun alt alan 2'deki gerçek oranı, kestirilen farka eklenir, toplam 3. zamandaki toplam alan nüfusu ile çarpılır. Kestirilen alt alan nüfusları, toplamlarının toplam alan nüfusuna eşit olması için hafifçe ayarlanabilir.

3.4. Örnek-Regresyon Metodu

Örnek-regresyon metodu, seçilmiş semptomatik gösterge verilerinden oluşturulmuş regresyon eşitliğine dayanır. Bu değişkenler, her alan için bağımsız değişkenler olarak hesaplanırlar ama, güncel örnek verisi, bağımlı değişken olarak ele alınır. Bu metot, modeldeki parametreleri belirlemek için, ağırlıklı olarak eldeki veriye dayanır. Kullanılan veri, oldukça güncel nüfus sayımına dayanır. Ve oluşturulan eşitlik, örneğin, oran-korelasyon metodunda yapıldığı gibi, mantıksal varsayımlarla sınırlandırılmaz. Hansen et al. [34] tarafından önerilen bir çalışmada, yerel kestirimler geliştirmek için bir regresyon ilişkisi oluştururken, çift örnekleme kullanılmıştır. Fakat semptomatik veri ve çok değişkenli yaklaşım kullanılmamıştır. Daha sonra Woodruff [35] bu yaklaşımı daha da geliştirmiştir. Bu yaklaşım bugün genel olarak Ericksen [36] ile anılır. Çünkü Ericksen [36] bu metodu çok değişkenli olarak ele almıştır.

Metodun ele alınışı, oran-korelasyon metoduna benzer. Tabakalanmış n örnek birimindeki bir özellik için önceden kestirimlerin bulunduğu varsayılsın. Aynı şekilde, N tane yerel birimin hepsi için p tane semptomatik göstergenin değeri de biliniyor olsun. Bu durumda şu terimler matris şeklinde ifade edilir :

Y : Bir nx1 vektördür. Özellik değişkenidir. Gerçek fakat gözlenmemiş değerler kümesidir.

X : Bir n x (p+1) matristir. Semptomatik göstergeler kümesidir. Bu göstergeler teorik olarak şu eşitlik tarafından Y ile ilişkilendirilebilirler :

$$Y = XB + u \qquad (3\text{-}46)$$

B : Regresyon katsayılarının bir px1 matrisi.

u : Stokastik hata terimlerinin bir nx1 matrisi

Bu formülün matrislerle ifade ediliş şekli şöyledir :

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} \ldots\ldots & X_{p1} \\ 1 & X_{12} & X_{22} \ldots\ldots & X_{p2} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & X_{1n} & X_{2n} \ldots\ldots & X_{pn} \end{bmatrix} \text{x} \begin{bmatrix} B_0 \\ B_1 \\ . \\ . \\ . \\ B_p \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_n \end{bmatrix} \qquad (3\text{-}47)$$

Doğrusallık varsayımı verilmiş olsun, bu durumda eğer Y değerleri gözlenmiş ise, B'yi sıradan en küçük karelerle kestirmek olasıdır. Gözlem seti şu şekilde yazılabilir :

$$Y_o = Y + v \qquad (3\text{-}48)$$

v : Örnekleme hatalarının nx1 vektörü

Gözlem değerlerini veren model daha ileri bir ifadeyle şu olacaktır :

$$Y_o = XB + u + v \qquad (3\text{-}49)$$

Burada

$$E\,(u) = 0 \tag{3-50}$$

$$E\,(uu') = \sigma_u^{\,2}(I_n), \qquad (I_n = nxn \text{ birim matris}) \tag{3-51}$$

ve

$$X\text{'in rankının } (p+1) < \text{n} \tag{3-52}$$

olduğu varsayılırsa, regresyon katsayıları matrisi B'nin kestirimini veren eşitlik :

$$\hat{B} = (X'X)^{-1} X'Y \tag{3-53}$$

olacaktır. Örneğin j birimi için, özellik değişkeninin güncel kestirimi :

$$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 X_{1j} + \hat{B}_2 X_{2j} + \ldots\ldots\ldots + \hat{B}_p X_{pj} \tag{3-54}$$

olur. Kestirimlerin hata kareleri ortalamaları ise

$$MSE = \frac{E(Y-\hat{Y})^2}{n} = \frac{(n-p-1)}{n}\sigma_u^{\,2} + \frac{(p+1)}{n}\sigma_v^{\,2} \tag{3-55}$$

$\sigma_u^{\,2}$: Tabakalanmış birimler arası açıklanmayan varyans

$\sigma_v^{\,2}$: Tabaka içi varyans

Y : Özellik değişkeninin gerçek değeri

Ŷ : Özellik değişkeninin kestirilmiş değeri

olacaktır. Hata kareleri ortalamalarının (MSE)[1] bulunuşu EK-2'de verilmiştir [36].

3.5. Sentetik-Regresyon İşlemleri

Bölüm 3.2'de tartışılan sentetik yaklaşım, değişimlere duyarlı olduğu ve bunları hesaba kattığı sürece işlerlik kazanır. Bu da, sentetik yaklaşımın bir çeşit zayıf noktasıdır. Örneğin, bir küçük alanda yeni bir fabrika açılması veya yeni bir endüstrinin kurulması, bu bölgedeki iş gücünün karekteristiğini, oldukça köklü bir biçimde etkiler. Aynı şekilde bir dinlenme merkezinin açılması, alandaki yaşlı nüfus oranını, dolayısıyla sağlık hizmetleri talebini arttıracaktır. Bu değişimler, küçük alanda meydana geldikleri için, büyük alanın içerisinde yansımayabilirler. Bu da duyarlık eksikliğine yol açar. Bu eksikliği gidermenin bir yolu, bir regresyon ilişkisi yardımı ile ek değişkenler kullanmaktır. Burada kullanılabilecek göstergeler: Doğumlar, ölümler, okul kayıtları, işsizlik kayıtları olabilir.

---

[1] MSE : Mean square error

3.5.1. Regresyon ayarlı sentetik metot

Bu metot, Levy [37] tarafından önerilmiştir. Metot, sentetik kestirim ile birlikte yerel düzeydeki semptomatik bilgiyi kullanır. Kullandığı model :

$$x_h^* = \alpha + \beta Y_h + \varepsilon_h \qquad (3\text{-}56)$$

$$x_h^* = \{(X_h - \hat{x}_h)/\hat{x}_h\}100 \qquad (3\text{-}57)$$

$\hat{x}_h$: Sentetik kestirim

X_h : Gerçek değer

Y_h : Semptomatik değişkenin değeri

α ve β : Kestirilmesi gereken regresyon katsayıları

ε_h : Rastgele hata terimleri

modelidir. α 'nın kestiricisi $\hat{\alpha}$, ve β 'nın kestiricisi $\hat{\beta}$ bulunursa, ve hata terimi ε_h çıkartılırsa, x_h 'nin bir kestiricisi olan $\hat{x}_h^*$ (3-56)' dan doğrudan elde edilebilir:

$$\hat{x}_h^* = \hat{x}_h = \left[(\hat{\alpha} + \hat{\beta} Y_h)/100 + 1\right] \qquad (3\text{-}58)$$

x_h^*, bilinmeyen gerçek değer X_h 'nin bir fonksiyonu olduğundan, doğrusal katsayıları bulmak için farklı bir metot kullanılmalıdır. Özetle böyle bir metot, küçük alanları tabaka oluşturacak biçimde düzenledikten sonra, en küçük kareler yoluyla α ve β 'nın kestirim değerlerini bulmalıdır. Bu tabakalar, oluşturulurken, X_h 'nin mantıklı ve yansız kestirimlerini verebilecek şekilde oluşturulmalıdırlar.

3.5.2. Karma sentetik-regresyon metodu

Bu Metot ise, Nicholls [38] tarafından ilk olarak geliştirilmiştir. Temelde örnek-regresyon metodu ile aynı olmakla birlikte, sentetik kestirimi ek bir bağımsız değişken olarak eklemesi yönüyle ayrılır. Semptomatik değişkenlerin uygun seçilmesiyle, teknik sentetik kestirimler ve standart örnek-regresyon kestirimlerinde gelişmeye yol açar.

Gonzalez and Hoza [39] bu metotla, küçük alandaki işsizlik miktarının kestirimini konu alan bir çalışma yapmışlardır. Bu tekniğin ana hatları şu şekildedir :

Teknik, sentetik kestirimi, yukarıda da belirtildiği gibi, regresyon eşitliğinde bağımsız değişkenlerden herhangi biri gibi kullanmaktadır. Bağımlı değişken olarak herbir ana örnek ünitelerindeki işsizliğin güncel nüfus sayımı ile elde edilmiş kestirimleri gözönüne alınmıştır. Yerel varyansı azaltmak için ek bağımsız değişkenler de kullanılmıştır. Regresyon hataları kareleri ortalaması (MSE) ise şu eşitlikle hesaplanmıştır:

$$\text{MSE} = \text{E}\left[\frac{(Y_o - \hat{Y})'(Y_o - \hat{Y})}{n}\right] - \frac{(n-2p-2)\sigma_w^{\,2}}{n} \qquad (3\text{-}59)$$

Y_o : Y'nin gözlenen değerleri

n : Gözlem sayısı

p : Bağımsız değişken sayısı

$\sigma_w^{\,2}$: Ana örnek ünitesi içindeki hata

$\sigma_w^{\,2}$ 'nin kestirimi Ericksen'e [40] dayanmaktadır.

Sentetik kestiricilerin yanlı olabilmesinden dolayı, bunların değişkenlik ve yanlılıklarını yansıtabilecek bir ölçü ya da kritere ihtiyaç vardır. Bu çeşit bir ölçüm olarak, hata kareleri ortalaması kökü (H.K.O.K.) alınabilir.

Gonzalez and Hoza'nın [39] araştırmasında, en geniş alan, bir coğrafi bölgedir. Alt alan olarak iller alınmıştır. Göreceli H.K.O.K.'nun ölçümü, yüksek ve çok düşük işsizlik miktarlarında artar. Göreceli H.K.O.K. şu şekilde hesaplanmaktadır :

$$\text{Göreceli H.K.O.K.} = \frac{\text{H.K.O.K.}}{\text{İşsizlik aralığının orta noktası}} \qquad (3\text{-}60)$$

Hata kareleri ortalamalarını azaltmak için regresyondan bazı gözlemler çıkartılabilir. Bu gözlemler bu araştırmada güncel nüfus sayımı ve regresyon kestirimleri arasındaki farktan belirlenebilirler.

3.6. Temel Birim Metodu

Ana kütle büyüklüğünden ziyade, değişkenlerle kolayca kullanılabilecek bir metot arayışları içinde, Kalsbeek [41], küçük alanları küçük birimlere ayırmaya dayanan temel birim metodunu önermiştir. Bu alanlar, küçük birimlere ayrıldıktan sonra, k tane grup olacak şekilde sınıflandırılırlar. Bu gruplar için kestirimler, bir araştırma örneklemesinden elde edilebilir. Küçük alan kestirimleri ise, bu temel kestirimlerin ağırlıklı kombinasyonlarından hesaplanabilir. Metot, sentetik kestirim metodununki kadar olmamakla birlikte, sadece yerel coğrafi bölgeler olarak tanımlanmış küçük alanlarda sağladığı pratiklik nedeni ile, daha kısıtlı bir metottur.

Metodun uygulanışı aşamasında araştırılacak alan, coğrafi birimlere bölünür. Bu birimlere, temel birim adı verilmektedir. Bu temel birimler, bloklar, ilçeler, sayım bölgeleri ya da diğer küçük coğrafi birimler olabilir. Bu birimlerin bir örneği için, örnekleme araştırması verisi gerekir. Semptomatik bilgi ve ilgilenilen özelliğe ait bilgi baz alınarak, uygun bir tabakalama işlemi ile veriler, k homojen grup halinde gruplanırlar. İlgilenilen alan da aynı şekilde ayrı ayrı birimlere ayrılır. Eldeki semptomatik bilgi kullanılarak, her birim, bu k tane grubun herbirine girecek şekilde sınıflandırılır.

Kestirim işlemi ise şu şekilde yapılır: k tane temel birim grubunun herbiri için kestirim, her grubu oluşturan örnek temel birimleri için kestirimlerin ağırlıklı ortalaması alınarak bulunur :

$$\bar{x}'_g = \sum_{l=1}^{n_g} w_{gl}\bar{x}'_{gl} \qquad g = 1,\ldots\ldots,k \qquad (3\text{-}61)$$

n_g : Temel birimlerin g. grubundaki örnek temel birim sayısı

w_{gl} : Ağırlık

$\bar{x}'_{gl}$: g. gruptaki l. temel birim için ortalamanın örnek kestirimi

$\bar{x}'_g$: Temel birimlerin g. grubu için ağırlıklı ortalama

Sonuç kestirim $\bar{x}'_h$ ise, k grubun ortalamalarının ağırlıklı ortalaması alınarak bulunur :

$$\bar{x}'_h = \sum_{g=1}^{k} \theta_{gh}\bar{x}'_g \qquad (3\text{-}62)$$

Burada θ_{gh} ağırlıkları, yerel alan h'deki k grubun bileşimini temsil eder

Bu metodun örnek-regresyon metoduna göre avantajı ilgilenilen özellik değişkenine ve semptomatik değişkenlere ait hiçbir fonksiyonel kalıp varsayımı gerektirmemesidir. Ancak metodun zayıf yanı örnek temel birimleri için elde kestirimler bulunmasını gerektirmesidir.

3.7. Karma Kestiriciler

İlgilenilen küçük alan için örnek kestirimlerine sahip olunduğunda, bu kestirimleri sentetik kestirimlerle birleştirmek yararlı olabilir. Sentetik kestiricinin düşük varyansı ile yanlılığının, ve doğrudan kestiricinin yüksek varyansı ile yansızlığının birleştirilmesinden, hem varyansı hem de yanlılığı kabul edilebilir değerde olan daha sağlıklı kestirimler yapılabilir. Bu konudaki bir yaklaşım, Schaible [42] tarafından verilmiştir :

$$\hat{T}_{(com),d^*,t^*} = w_{d^*t^*}.\hat{T}_1 + (1 - w_{d^*t^*})\hat{T}_2 \qquad (3\text{-}63)$$

$w_{d^*t^*}$: 1 ve 0 arasında bir ağırlık

$\hat{T}_1$ ve $\hat{T}_2$: Bileşen kestiriciler

Küçük alan kestirimi uygulamalarında, genellikle bir bileşen kestirici doğrudan, diğeri ise alan ya da zaman dolaylı olarak seçilir. Bileşen kestiricilerden birinin doğrudan olması isteniyorsa, ilgilenilen alandan en az bir gözlemin elde olması gerekecektir. Sentetik ve dolaylı regresyon kestiricileri, ilgilenilen alandan hiçbir gözlem değeri elde yoksa bile kullanılabilirler. $w_{d^{\bullet}t^{\bullet}}$ ağırlıklarının tanımı için çok çeşitli yaklaşımlar vardır.

Oran-korelasyon ve örnek-regresyon metotları birleştirilerek de karma kestirimler elde edilebilir. Bu metotlar birbirinden çok farklı metotlar değildirler. İkisinin arasındaki farklar, eldeki veriye ait farklı varsayımlar kullanmalarından kaynaklanır. Örnek-regresyon metodu regresyon katsayılarını kestirmek için güncel örnek bilgisini kullanırken, oran-korelasyon metodu daha kesin, fakat eski tarihli katsayıları kullanır.

3.8. James-Stein ve Bayes Kestirimleri

James-Stein işleminin küçük alan kestirimlerinde kullanılabilme olasılığı uzun süre ilgi çekmiştir. James-Stein kestiricisinin hesaplanması özetle şu şekildedir :

Herbiri eşit v varyansına ve farklı u_h ortalamalarına sahip olan H tane farklı küçük alan için x'_h örnek kestirimlerinin elde bulunduğu varsayılır. Daha önce bulunmuş p_h kestirimleri verildiğinde, u_h'nin uygun bir kestiricisi şu şekilde yazılabilir:

$$\tilde{x}_h = cx'_h + (1-c)p_h \quad , \quad 0 \le c \le 1 \qquad (3\text{-}64)$$

Önceki kestirim p_h'ler ana ortalamayı ya da daha geniş alanların ortalamalarını içerebilirler. Fakat aynı zamanda sentetik kestirimler, örnek-regresyon kestirimleri gibi kestirimler de olabilirler. Sabit bir c değeri için $\tilde{x}_h$'nin hata kareleri ortalaması şu şekilde bulunur :

$$R(u_h, \tilde{x}_h) = \sum_h E_{u_h}(u_h - \tilde{x}_h)^2 \qquad (3\text{-}65)$$

Bu ifade,

$$c = w/(w+v) \qquad (3\text{-}66)$$

$$w = \sum_h (p_h - u_h)^2 / H \qquad (3\text{-}67)$$

olacak şekilde ele alındığında minimize edilmiş olur. c'nin değeri (3-64)'de yerine konduğunda sonuç:

$$Min\ R(u_h, \tilde{x}_h) = cHv \qquad (3\text{-}68)$$

olacaktır. Aynı zamanda

$$R(u_h, x'_h) = Hv \tag{3-69}$$

olduğu da görülebilir. Sonuç olarak:

$$Min\ R(u_h, \tilde{x}_h) \leq R(u_h, x'_h) \tag{3-70}$$

olur. Bu sonuç baz alınarak James-Stein kestiricisinin,

$$\hat{c} = 1 - (H-2)v / s \tag{3-71}$$

$$s = \sum_h (x'_h - p_h)^2 \tag{3-72}$$

olmak üzere $\hat{c}$ ile kestirilen, c değerini içeren (3-64) olduğu söylenebilir. $H \geq 3$ değerleri için James-Stein kestiricisi bütün u_h 'ler için, x'_h 'nin sahip olduğundan daha az hata kareleri ortalamasına sahiptir.

3.8.1. James-Stein kestiricisinin Bayes yöntemi ile elde edilmesi

Efran and Morris [43], $k \geq 3$ normal ortalama için James-Stein kestiricisinin maksimum olabilirlik kestiricisinden daha iyi sonuç verdiğini göstermek amacıyla Bayes kestiricisinden yararlanan bir çalışma yapmışlardır.

Buna göre James-Stein kestiricisinin Bayes yöntemi ile elde edilişi şu şekildedir:

Elde k tane, $k \geq 3$ olmak üzere $\theta_1, \theta_2, \ldots\ldots\ldots \theta_k$ parametreleri olduğu varsayılsın. Her biri için bağımsız normal varyans gözlenmiş olsun :

$$x_i \setminus \theta_i \sim n(\theta_i, 1) \tag{3-73}$$

x_i : n tane bağımsız gözlem $Yij \sim n(\theta_i, \sigma^2)$' nin ortalaması

Bu durumda :

$$\mathrm{x_i} \sim n(\theta_i, \sigma^2 / n) \tag{3-74}$$

olacaktır.

Uygun bir ölçek dönüşümü yapılırsa σ^2 / n, daha uygun olan 1 değerine dönüştürülebilir. Fakat bu durumda σ^2'nin biliniyor olması gerekli ise de, sonraki adımlardaki sonuçlar bundan etkilenmeyeceklerdir. $\theta = (\theta_1, \theta_2, \ldots\ldots\ldots \theta_k)$ vektörü kestirilmek istenmektedir. Bunun için hata kareleri kullanılacaktır.

$$L(\theta, \delta) = \frac{1}{k}\|\delta - \theta\|^2 = \frac{1}{k}\sum_{i=1}^{k}(\delta_i - \theta_i)^2 \tag{3-75}$$

Çünkü amaç,

$$\delta(x) = (\delta_1(x), (\delta_2(x), \ldots\ldots\ldots, \delta_k(x)) \tag{3-76}$$

$\delta_i(x)$: θ_i 'nin kestiricisi olmak üzere, (3-76) kestiricisinin gücünü ölçmektir.

θ_i'nin kestiriminin $x=(x_1,x_2,........,x_k)$ gözlem vektörüne göre değişebilmesine izin verilebilir. x_i'nin sadece θ_i ile ilgili gibi görünmesi birşeyi değiştirmez.

$$\delta_i^0(x) = x_i \tag{3-77}$$

maksimum olabilirlik kestiricisi, bütün θ değerleri için

$$R(\theta,\delta^0) \equiv E_\theta L(\theta,\delta^0(x)) = 1 \tag{3-78}$$

risk fonksiyonuna sahiptir. James and Stein [44],

$$\delta_i^1(x) = \left(1 - \frac{k-2}{S}\right)x_i \qquad (\mathrm{S} = \|\mathrm{x}\|^2) \tag{3-79}$$

kestiricisinin, bütün θ değerleri için,

$$R(\theta,\delta^1) < 1 \tag{3-80}$$

(3-80) risk fonksiyonuna sahip olduğunu göstermişlerdir. Bu durum da James-Stein kestiricisinin, en büyük olabilirlik kestiricisinden daha iyi sonuç verdiğini kanıtlamaktadır. Kestiricinin formülü şu şekilde genelleştirilebilir :

$$S_i^t = \left(1 - \frac{(k-2)^t}{S}\right)x_i \tag{3-81}$$

James and Stein [44] aynı zamanda, yukarıdaki genelleştirilmiş formüldeki herhangi bir t değeri için δ^t'nin, en büyük olabilirlik kestiricisine üstün geleceğini de kanıtlamıştır. Sonuç olarak t için en iyi değer, James-Stein'ın [44] seçeneği olan t=1 değeridir.

3.9. Kestirim Yaklaşımı

Kestirim yaklaşımında, ilgilenilen değişken ve semptomatik değişkenler arasında bir süper ana alan olasılık modeli ilişkisi varsayımı vardır. Bunların yardımı ile optimal alt alan kestiricileri elde edilir. Küçük alan özelliklerinin kestiricileri, örneğin, ilgilenilen değişken ve semptomatik veriler arasında bir doğrusal regresyon ilişkisi varsayılarak bulunur. Royall [45], çalışmasında, en iyi kestiricinin, örnek birimlerinden gözlenmiş değerler ile, gözlenemeyen ana alan birimlerine ait olan bir kestiricinin, bir doğrusal kombinasyonu olarak açıklanabileceğini belirtmiştir. Çalışması özetle 3.9.1 bölümünde verilmiştir.

3.9.1. Kesin doğrusal regresyon modelleri doğrultusunda, sınırlı alan üstünde örnekleme teorisi

İlgilenilen alan 1,........,N şeklinde etiketlenmiş N tane tanımlanabilir birimlerden oluşmuştur. i birimi için x_i ve y_i olmak üzere iki değer vardır. x_i bilinmekte fakat y_i sabit olmakla beraber bilinmemektedir. N tane birimden bir örnek çekilecek ve örnek ünite ile ilgili y değerleri gözlenecektir. Amaç şu toplamı kestirmektir :

$$T = \sum_{i=1}^{N} y_i \tag{3-82}$$

Problem ise, bir örnekleme planı ve kestirici bulmaktan ibarettir. Bu kestirici, gözlenmeyen y değerlerinin bir fonksiyonu olmamak koşuluyla, herhangi bir gerçek değerli istatistik olabilir. Örnek seçiminde 1,........,N tane üniteden s tanesi, S tane ünitenin üzerinde tanımlanan bir olasılık fonksiyonuna göre seçilir. Örnekleme planı olarak adlandırılan P fonksiyonu $x_1,........,x_n$ değerlerine bağlı olabilir ama bilinmeyen $y_1,........,y_n$ değerlerine bağlı olamaz. Örnek büyüklüğünün beklenen değeri sabit ve n olarak adlandırılacaktır, yani eğer $n(s)$, s içinde farklı ünitelerin sayısı ise, $E\{n(s)\} = n$ olacaktır.

Ana kütlenin bir parçası, $x_1,........,x_n$'lerin tümüyle pozitif olduğu serpilme diyagramlarıyla karakterize edilebilir. N tane (x_i, y_i) noktası, (0,0) noktasından geçen bir doğru etrafında yoğunlaşır ve noktaların doğru etrafında serpilmesi x arttıkça artar. Bu tür kütleler için bir çok örnekleme planı ve kestiriciler önerilmiştir. Bunlardan ikisi:

i) Birim i içindeki π_i olasılığının x_i'ye orantılı olduğu bir örnekleme planı kullanılır ve T, Horvitz-Thompson [46] kestiricisi ile kestirilir :

$$\hat{T}_{HT} = \frac{1}{n}\sum{}_s (y_i / x_i)\sum_{i=1}^{N} x_i \tag{3-83}$$

ii) Basit rasgele bir örnek alınır:

$$p(s) = \begin{cases} 1/\binom{N}{n} & , \quad \text{n(s) = n} \\ 0 & , \quad \text{d.d.} \end{cases} \tag{3-84}$$

ve T, oran kestiricisi ile kestirilir.

$$\hat{T} = \left(\sum{}_s y_i / \sum{}_s x_i\right)\sum_{i=1}^{N} x_i \tag{3-85}$$

x_i genellikle birimin büyüklüğünün bir ölçüsüdür. Örneğin birimler çiftlikler olsa, x_i, i. çiftliğin yüzölçümü olacaktır. π_i'nin x_i'ye orantılı olduğu örnekleme planı da olasılığın, yüzölçümüyle orantılı olduğu bir örnekleme planı olacaktır. $\pi_1+........+\pi_N$ beklenen örnek büyüklüğü olduğundan ve hiçbir π_i, 1'den büyük olamayacağından bir P_{pps}, yani örnek büyüklüğü beklenen değeri n olan $\hat{T}_{HT}$ stratejisi sadece $nx_i \le x_1+........+x_N$ (i=1,........,N) olduğunda uygulanabilecektir. Buna karşılık değişik olasılıkların eklenmesiyle elde edilebilecek bir çok örnekleme planı vardır ve P_{pps} bunlardan herhangi birisini temsil eder.

$\hat{T}_{HT}$, (3-83) ile tanımlanan istatistiği belirtmektedir. Bu istatistik, olasılığın büyüklüğe oranlandığı bir örnekleme planı ile kullanıldığı zaman, Horvitz-Thompson [46] kestiricisi adını alır.

Bir örnekleme planı ve kestirici seçmek için şu model göz önünde bulundurulabilir : $y_1,........,y_N$ sayıları, bağımsız rastgele değişken $Y_1,........,Y_N$ değerlerinin gerçek, ve bilinen değerleri olsun. Y_i' lerin ortalaması βx_i ve varyansı $\sigma^2 v(x_i)$ olsun. v fonksiyonunun x>0 için $v(x)$>0 olacağı bilinmektedir ama β ve σ^2 sabitleri bilinmemektedir. ξ, $Y_1,........,Y_n$' in bileşik olasılığını temsil etsin. Bir örnekleme ve kestirim stratejisi $p{:}\hat{T}$ olarak verilebilir .Bu stratejinin de , diğer bir strateji $p'{:}\hat{T}'$' dan daha iyi olduğu söylenebilir. Bunun için koşul ilk stratejinin daha küçük hata kareleri ortalamasına (MSE) sahip olmasıdır. Yani :

$$MSE(p{:}\hat{T}) < MSE(p'{:}\hat{T}') \qquad (3\text{-}86)$$

olmasıdır. $MSE(p{:}\hat{T})$, şu şekilde tanımlanabilir :

$$MSE(p{:}\hat{T}) = E_\xi\left\{\sum_s p(s)(\hat{T}-T)^2\right\} \qquad (3\text{-}87)$$

$\sum_s p(s)(\hat{T}-T)^2$ = Örnekleme hata kareleri ortalaması

E_ξ = Olasılık dağılımı ξ ' a göre beklenen değer.

$Y_1,........,Y_N$'in dağılımına göre, daha küçük ortalama örnek hataları karesine sahip olan strateji, her zaman daha iyi bir stratejidir.

3.10. Kategorik Veri Analizi Yaklaşımı ve Yapı Koruyan Kestirim (Y.K.K.)

Küçük alan kestirimlerinin bir çoğu, frekanslı veriler şeklinde gösterilen değişkenlerle ilgilenir. Bu kestiriciler de, bir kategorik veri analizi çerçevesinde oluşturulabilirler. Kategorik veri analizi bir çok amaç için kullanılmıştır. Aynı zamanda bu analiz, frekans içermeyen özelliklerin kestiriminde kullanılmak üzere uygulanabilmektedir (kişi başına düşen milli gelir gibi). Bu nedenle bu metot küçük alan kestirimlerinde de kullanılabilmektedir. Bu yaklaşım, kestirilecek özellikle ilgili değişkenler üzerinde geniş bilgi ve esneklik sağlar. Örneğin; yaş, cinsiyet ve ırk, yerel alanlardaki işsizlik oranlarıyla oldukça ilişkili etmenlerdir. Bu tür verilere ilişkili değişkenler adı verilir. Bu veriler, ana alanı etkili bir şekilde çapraz sınıflara ayırırlar ve kestirim işleminde önemli rol oynarlar.

3.10.1. Yapı koruyan kestirim (Y.K.K.) metodu.

Metot, Purcell and Kish [47], tarafından ortaya atılmıştır. Metodun uygulamasında, öncelikle bir ortaklık yapısına ihtiyaç vardır. Bu, arzu edilen küçük alan içindeki daha önceki bir zamandaki kestirim için, ve ilişkili değişkenler arasındaki ilişkiyi oluşturmak için gereklidir. İkinci olarak bir dağıtım yapısına ihtiyaç vardır. Bu yapı da, küçük alanları içine alan bir büyük alan içinde, araştırılan değişkenler ve ilişkili değişkenler arasındaki güncel ilişkiyi oluşturmak içindir.

Ortaklık yapısını oluşturan bilgi üç boyutlu kontenjans tablolarıyla oluşturulur. N_{hig} hücreli, N kontenjans tablosunda $h = 1,........,H$, $i = 1,........I$, $g = 1,........,G$ olmak üzere,

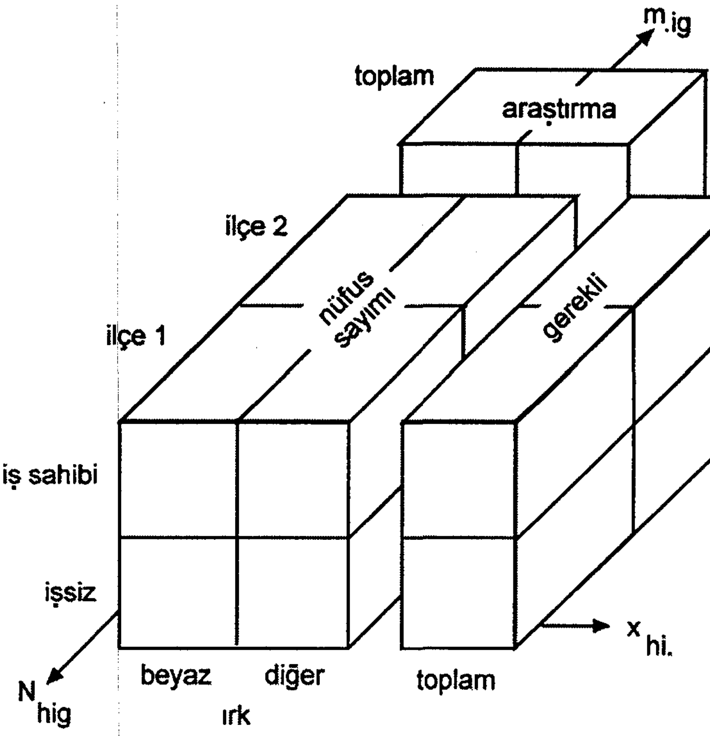
h : h. küçük alan,

i : Kestirim değişkeninin i. kategorisi ,

g : Birleşmiş değişken boyutunun g. kategorisini ,

temsil eder. Dağıtım yapısı (güncel bilgi), aslında ortaklık yapısı N' in, güncellenmiş kısımlarını yansıtır. Bu sınırları oluşturan hücre değerlerinin oluşturduğu küme, m olarak adlandırılacaktır. Örneğin: dağıtım yapısı kestirim değişkeni i ise ve ortak değişkenler ile tanımlanmış ise, m, hücre değerleri $m_{.ig}$ ile temsil edilir. Noktalı indis, h'nin bütün yerel alan değerleri için toplamını göstermektedir.

İki veri yapısı N ve m verildiğinde öncelikle ortaklık yapısı N'in güncel kestirimlerine ihtiyaç vardır. Bu kestirimler x ile gösterileceklerdir ve hücre değerleri x_{hig} olacaktır. İlgilenilen şey ise küçük alanlar için x'in kısımlarındadır. Yani, ortak değişken boyutu g üzerinde toplanarak bulunan, $x_{hi.}$ hücre değerlerinin içindedir. Bir grafik gösterimi yapmak amacıyla, iki ilçe (yerel alan) içindeki iş sahibi ve işsiz insan sayılarını kestirme problemi incelenebilir. Bu problemde, ırk faktörünün iş durumu ile oldukça ilgili olduğu varsayılırsa, bu faktör, tek ortak ikili değişken olarak kullanılabilir. Eldeki verinin yapısı Şekil 3.1'de gösterilmiştir .



Şekil 3.1. Küçük alan kestirimi için, basit bir veri yapısı örneği [47]

Ortaklık yapısı *N*, $2^3 = 8$ hücrede, bir önceki nüfus sayımından elde edilen veriyi temsil eder. Irka göre iş durumu kısmı tarafından verilen dağıtım yapısı *m* ise, güncel araştırmadan elde edilir. Irka göre iş durumu ilçeler için dört hücrede toplanmıştır.

Daha sonraki işlem, elverdiği kadar ortaklık yapısı N_{hig}'deki değişkenler arasındaki ilişkiyi korurken, dağıtım yapısındaki güncel veriye uydurmak üzere ortaklık yapısını ayarlamaktır. Sonuçta oluşan ayarlanmış tablo x_{hig}, geçicidir. Bu tablo *g* değişkeni üzerinde toplanır, ve gereken kestirici $x_{hi.}$'lar bulunur. İhtiyaç duyulan kestiriciler, sağ kısımdaki dört hücre tarafından temsil edilirler (iş durumu - ilçe kısmı) (Bkz. Şekil 3.1).

Kestiricilerin bulunması işleminde çeşitli adımlar vardır:

a) $N = \{N_{hig}\}$ ve $m = \{m_{.ig}\}$ olduğu durum.

Bu durum en önemli durumdur. Burada ortaklık yapısındaki bilgi (*h, i, g*) tamdır. Fakat küçük alan boyutu *h* dağıtım yapısındaki güncel örnek verisinde yer almamaktadır. *x*'in kestiriminde birkaç yol izlenebilir. Hepsi de aynı kestirimi vereceğinden, en basiti olan ağırlıklı en küçük kareler yaklaşımı, uygun bir yaklaşım olabilir. Standart Lagrange analizi uygulanarak şu fonksiyon minimize edilirse :

$$F(x_{hig}) = \sum_{hig} w_{hig}^{-1}(x_{hig} - N_{hig})^2 - \lambda\left(\sum_{h} x_{hig} - m_{.ig}\right) \quad (3\text{-}88)$$

w_{hig} : Önceden belirlenmiş ağırlıklar.

Ve eğer

$$w_{hig} = N_{hig} \quad (3\text{-}89)$$

olacak şekilde w_{hig} değeri seçilirse, kestirici,

$$x_{hig} = \frac{N_{hig}}{N_{.ig}} m_{.ig} \quad (3\text{-}90)$$

haline gelecektir. Her küçük alan *h*'de, ilgilenilen *i* özelliği için sonuç kestirimi :

$$x_{hi} = \sum_{g} \frac{N_{hig}}{N_{.ig}} m_{.ig} \quad (3\text{-}91)$$

olacaktır. Aynı kestiriciyi en büyük olabilirlik ya da diskriminant bilgisi kriteri minimizasyonu yöntemi ile de elde etmek olasıdır. Bu kestiriciler, parçalı oran kestiricileri genel sınıfının bir üyesidirler.

b) $N = \{N_{hig}\}$ ve $m = (\{m_{.ig}\}, \{m_{h..}\})$ olduğu durum,

(a)' daki bilgiye ek olarak, bu durumda yerel alan *h*'lerin toplamı için güncel kestirimler bulunmaktadır. Bu ek bilgi, kestirimlerin oldukça iyi elde edilmesini sağlar. Küçük alanların boyutları hakkındaki bilgi, araştırmalardan ya da kayıtlardan alınabilir. Bu bilgi tüm alanlar ya da onların bir örneği için olabilir. Bu durumda (a) kısmında belirtilen kriterlerden herhangi biri kullanılarak işlem yapılabilir. Fakat sonuçta elde edilecek olan Lagrange ifadeleri, çok fazla kolay işlenir ifadeler değildirler, yani, analitik bir çözümlerini bulmak zordur.

(a) kısmında görüldüğü gibi, dağıtım yapısını tanımlayan sadece bir tane marjinal küme varsa, bir parçalı oran kestiricisi olan, basit bir çözüm bulunur. Burada önerilen yaklaşım, hücre frekanslarını başarılı bir şekilde ayarlayarak bu gerçekten yararlanmaktadır. Bu işlem şu şekildedir :

Birinci adımda, başlangıç değerleri, ortaklık yapısını belirleyen hücre değerlerine eşitlenir. Yani:

$$x_{hig}^{(0)} = N_{hig} \quad (3\text{-}92)$$

Bu hücre değerleri daha sonra, marjinal sınırlayıcıların birinci kümesine göre ayarlanırlar. Bu, küme dağıtım yapısı:

$$\sum_h x_{hig} = m_{.ig} \tag{3-93}$$

tarafından belirlenir. Bu işlemden sonra , yine aynı hücre değerleri, ikinci sınırlayıcılar kümesine göre ayarlanırlar. Bu küme de :

$$\sum_{ig} x_{hig} = m_{h..} \tag{3-94}$$

şeklinde belirtilir. Bu iki adım içeren tekrar zincirinde k. Tekrarda :

$$_1x_{hig}^{(k)} = \frac{x_{hig}^{(k-1)}}{x_{\cdot ig}^{(k-1)}} m_{\cdot ig} \quad \text{ve} \quad x_{hig}^{(k)} = \frac{_1x_{hig}^{(k)}}{_1x_{h..}^{(k)}} m_{h..} \tag{3-95}$$

elde edilir.

Burada $_1x_{hig}^{(k)}$'ler, marjinal sınırlayıcılar $\{m_{.ig}\}$'leri k. tekrarda ayarlamaktan doğan kestiricilerdir. Elde edilen $x_{hig}^{(k)}$ kestiricileri, bir sonraki zincire konmak üzere, girdi olarak kullanılırlar. Adımlama işlemi, tatminkar bir sonuç (yakınsayan sonuç) bulununcaya kadar sürdürülür. Bu kestiriciler, iki-adım taramalı oran kestiriciler genel sınıfının bir üyesidirler.

c) $N = \{N_{hig}\}$ ve $m = (\{m_{.ig}\}, \{m_{h.g}\})$ durumu.

Bu durumda ise, güncel bilgi bulunmaktadır. Çözüm ise (b)'deki çıkarımlardan doğrudan elde edilebilir. Tek fark, k. tekrarda ikinci adımdaki ayarlamanın şu şekilde verilmesidir :

$$x_{hig}^{(k)} = \frac{_1x_{hig}^{(k)}}{_1x_{h.g}^{(k)}} m_{h.g} \tag{3-96}$$

Çözüm, ortak değişkenler üzerinde toplandığında, yine iki-adım parçalı taramalı oran kestirimi elde edilir.

d) $N = \{N_{h.g}\}$ ve $m = \{m_{.ig}\}$ durumu.

Bu durum, tam ortaklık durumu $N = \{N_{hig}\}$'nin, kestirim yaparken kullanılacak verinin daha önceki nüfus sayımlarında toplanmamasından kaynaklanan yokluğunu sergileyen, ve sık sık karşılaşılan bir durumdur. Örneğin, doğum kontrol hapı kullanılması veya kullanılmaması gibi. Bu durumda veri, bir tamamlanmamış ortaklık durumu sergiler. Bu durum da, ortak verilerin önceki nüfus sayımlarına göre küçük alanlarla çapraz sınıflandırılmasıyla gösterilir. Sonuç olarak ortaklık yapısı üzerine elde olan bilgi $N = \{N_{h.g}\}$ ile tanımlanır. Bu yapı i kestirim verisi kategorisinden yoksundur. Bu nedenle bir geçici ortaklık yapısı tanımlamak üzere bir model kullanılmalıdır. En basit ve sık kullanılan işlem *i* kategorileri arasında oransallık olduğunu varsaymak ve,

$$x_{hi} = \sum_{g} (N_{h.g} / N_{.g}) m_{.ig} \qquad (3\text{-}97)$$

kestiricisini kullanmaktır. Bu kestirici parçalı oran kestiricisinin özel bir durumudur.

e) $N = \{N_{h.g}\}$ ve $m = (\{m_{.ig}\}, \{m_{h..}\})$ durumu

f) $N = \{N_{h.g}\}$ ve $m = (\{m_{.ig}\}, \{m_{h.g}\})$ durumu

bu durumlarda kullanılacak kestiriciler, (b) ve (c) durumlarında kullanılanlan kestiriciler bulunurken, işlem daha da ilerletilerek bulunurlar. Farkları ise, sabit ortaklık yapısının (d)'de olduğu gibi kullanılmasıdır. Bu nedenle tek fark, başlangıç değerlerinin,

$$x_{hig}^{(0)} = N_{h.g} \qquad i = 1, \ldots\ldots, I \qquad (3\text{-}98)$$

şeklinde belirlenmesidir.

Yukarıda belirtilen kestiricilerden her biri, Purcell [48] tarafından gösterildiği gibi, log-doğrusal şekilde ifade edilebilirler. Bu gerçek, dağıtım yapısı (güncel veriler) tarafından tekrar belirtilenler dışında, ortaklık yapısına (geçmişe ait veriler) giren değişkenler arasındaki bütün ilişkilerin, güncel kestirimler için de geçerli olduğunu göstermek için kullanılabilir.

3.10.2. Farklı Y.K.K.'lerin karşılaştırılması

Y.K.K.'lerin varyansları, büyük oranda dağıtım yapısı tarafından belirtilen marjinal sınırlara göre değişir. Bunun nedeni, ortaklık yapısının geçmiş zamana ait verilerden oluşturulmasıdır. Bu nedenle bu kestiricilerin varyansı, bir çok uygulamada daha az kontrol edilebilen hatalara göre küçük olacaktır. Aynı zamanda, farklı Y.K.K.'lerin etkinlikleri üzerinde oldukça fazla etkileri olacaktır.

Y.K.K. kestirimleri yanlı kestirimler olmaya yatkındırlar. Yanlılık, açık bir şekilde hem dağıtım yapısından gelen marjinal sınırlar, hem de kullanılan ortaklık yapısının şeklinden ileri gelir. Bu nedenle, yukarıda bulunan farklı Y.K.K. kestiricilerinin, oldukça farklı büyüklükte hata terimlerine sahip olmaları beklenebilir. Y.K.K. kestirimlerinin yanlılıkları, yüzdelik kesin göreceli fark *(% KGF)* ile ölçülebilir. Bu ölçüm şu şekilde verilebilir :

$$\%KGF = \frac{|x_{hi.} - X_{hi.}|}{X_{hi.}} x100 \qquad (3\text{-}99)$$

$x_{hi.}$: Y.K.K. kestirimi

$X_{hi.}$: Gerçek değer.

Eğer dağılımların yatıklığı gözlenirse, farklı Y.K.K. kestirimlerinin yüzdelik kesin göreceli farklarının medyanları veya ortalamaları da kullanılabilir. Farklı Y.K.K. kestirimlerinin yapı koruma özellikleri, Çizelge 3.1' de verilmiştir.

Çizelge 3.1. Farklı Y.K.K. kestirimlerinin yapı koruma özellikleri [47]

(y : yeni, n: nüfus sayımı, 0 : sıfır)

	Etkiler						
	Tek-yönlü			İki-yönlü			Üç-yönlü
Kestirici	h	i	g	hi	hg	ig	hig
(a)	n	y	y	n	n	y	n
(b)	y	y	y	n	n	y	n
(c)	y	y	y	n	y	y	n
(d)	n	y	y	0	n	y	0
(e)	y	y	y	0	n	y	0
(f)	y	y	y	0	y	y	0

3.11. Tekrarlı Oran Uydurulması (T.O.U.) ile Kategorik Veri Analizi

Tekrarlı oran uydurulması (T.O.U.) (tekrarlı ölçüm, marjinal düzey veya marjinal ölçüm) çözümü, ağırlıklı en küçük kareler çözümünün sadece yaklaşık bir değerini hesaplar. Fakat aynı zamanda çoklu normal dağılımın olabilirlik fonksiyonunu maksimize ederken, diskriminant bilgisini minimize eder. Elde edilen T.O.U. kestirimleri, dağıtım yapısı dışında ortaklık yapısı tarafından belirtilen bütün etkileşimleri korur. Sentetik kestirimler, bazı kesin koşullar altında, T.O.U. çözümünün birinci yaklaşık değerine eşit ve optimaldirler [49].

3.11.1. İki boyutlu problem

Örnek tabloları, tüm ana kütlenin çapraz tablolanmasından elde edilmiş olan sonuçların kestiriminde kullanılacaktır. Bu durumun özeti, ana kütle ve örnek için Çizelge 3.2' de gösterilmektedir.

Çizelge 3.2. Örnek ve Ana kütlenin tablolanması [49]

Ana kütle

$j = ...$

	1	2 ....S	
$i = 1$	N_{11}	$N_{12}........N_{1s}$	$N_{1.}$
$i = 2$	N_{21}	$N_{22}........N_{2s}$	$N_{2.}$
		N_{ij}	$N_{i.}$
$i = r$	N_{r1}	$N_{r2}.........N_{rs}$	$N_{r.}$
	$N_{.1}$	$N_{.2}.........N_{.s}$	N

-N_{ij} bilinmiyor

-Marjinal toplamlar $N_{.j}$ ve

$N_{i.}$ biliniyor

-N biliniyor

Örnek

$j = ...$

	1	2.........S	
	n_{11}	$n_{12}.........n_{1s}$	$n_{1.}$
	n_{21}	$n_{22}....... n_{2s}$	$n_{2.}$
		N_{ij}	$n_{i.}$
	n_{r1}	$n_{r2}.........n_{rs}$	$n_{r.}$
	$n_{.1}$	$n_{.2}..........n_{.s}$	n

-n_{ij} biliniyor

-Marjinal toplamlar $n_{.j}$ ve

$n_{i.}$ biliniyor

-n biliniyor

Ana kütleden rastgele seçilen n tane bireyi kapsayan örnek için, problemin hedefi, ana kütle içindeki bilinmeyen N_{ij} frekanslarını kestirmektir. Bu, hesaplanmış m_{ij} frekanslarını bularak, ve bunları ana kütle örnekleme oranı N/n yoluyla genişleterek yapılır.

En küçük kareler çözümü için aşağıdaki ifadeyi minimize eden m_{ij} değerleri aranır:

$$S = \sum\left(m_{ij} - n_{ij}\right)^2 / n_{ij} \qquad (3\text{-}100)$$

m_{ij}'ler aşağıdaki durumların birisiyle ilgili olabilir.

<u>Durum I : Marjinal toplamların bir kümesi biliniyorsa</u>

$N_{1.},.........,N_{r.}$'lerin bilindiği varsayılır. Bu durumda :

$$\sum_j m_{ij} = m_{i.} \qquad i= 1, 2,........,r \qquad (3\text{-}101)$$

gereklidir. Ayarlanmış m_{ij} üzerinde r tane eşitlik, r tane koşul oluşturur.

<u>Durum II : Marjinal toplamların her iki kümesi biliniyorsa</u>

Burada ayarlanmış hücre frekansları yalnızca (3-101) koşulunu değil aynı zamanda

$$\sum_i m_{ij} = m_{.j} \qquad j = 1, 2, \ldots\ldots, s\text{-}1 \tag{3-102}$$

koşulunu da sağlamalıdır. Bu durumda $r+s$-1 tane koşul bulunmaktadır. Her iki durumda da

$$m_{i.} = N_{i.}\, n / N \tag{3-103}$$

$$m_{ij} = N_{ij} n / N \tag{3-104}$$

olur. $m_{i.}$ ve $m_{.j}$ küçültülmüş marjinal toplamlar olarak adlandırılırlar. Yani $N_{i.}$ ve N_{ij} gerçek örnekleme oranı N/n ile bölünürler. $m_{i.}$ ve m_{ij} bağımsız değildirler.

<u>İki boyutlu durum I'in çözümü :</u>

m_{ij}' nin ayarlanmış değerlerinin bulunduğu varsayılırsa :

1) Küçük değişim δm_{ij} 'ler alınır ve (3-100) ve (3-101) diferansiyel denklemlerinden şu sonuç çıkarılır :

$$\begin{aligned} S &= \sum (m_{ij} - n_{ij})^2 / n_{ij} \\ \delta S &= 0 \end{aligned} \tag{3-105}$$

Buradan şu eşitlikler çıkarılır :

$$\sum_j m_{ij} = m_{i.} \tag{3-106}$$

$$\sum_j \delta m_{ij} = 0 \tag{3-107}$$

Bu da r tane eşitliğe yol açar.

2) İsteğe bağlı Lagrange çarpanı $-\lambda_{i.}$ kullanılarak şu fonksiyon oluşturulur:

$$f - \lambda_{i.} g = \sum (m_{ij} - n_{ij})^2 / n_{ij} - \lambda_{i.} \sum_j m_{ij} \tag{3-108}$$

Burada $f = S$ ve $g = \sum_j m_{ij.}$ 'dır. Daha sonra şu eşitlik elde edilir :

$$\frac{\delta f}{\delta m_{ij}} - \lambda_{i.} \frac{\delta g}{\delta m_{ij}} = O \Rightarrow m_{ij} = n_{ij}(1+\lambda_{i.}) \qquad (3\text{-}109)$$

3) Ayarlanmış frekans m_{ij}'ler, $\lambda_{i.}$'lar bulunduğu anda hesaplanabilirler. Bu hesaplamayı yapabilmek için (3-109) eşitliğinin sağ tarafı kullanılarak (3-102) eşitliği tekrar m_{ij} için yazılır ve şu ifade elde edilir :

$$m_{i.} = n_{i.}(1+\lambda_{i.}) \qquad (3\text{-}110)$$

$\lambda_{i.}$ şimdi elde edilmiş ve bilinmektedir. $m_{i.}$ ve $n_{i.}$ de bilindiğinden, (3-109) eşitliğinden,

$$m_{ij} = n_{ij}\,(m_{i.} / n_{i.}) \qquad (3\text{-}111)$$

$$\left(m_{i.} = N_{i.} \frac{n}{N}\right)$$

eşitliğini çıkarmak olasıdır.

4) Minimize edilmiş kareler toplamı doğrudan, ya da:

$$S = \sum_i (m_{i.} - n_{i.})^2 / n_{i.} \qquad (3\text{-}112)$$

olduğunu görerek, satır toplamlarından elde edilebilir. $i.$ satır için $(m_{i.} - n_{i.})^2/n_{i.}$ terimi, ayrı bir şekilde düşünülebilir ve s-1 serbestlik dereceli χ^2 olarak kullanılabilir. Ya da (3-112) eşitliğinde verildiği gibi bütün satırlar minimize edilmiş S içinde toplanabilir ve $r(s$-1) serbestlik dereceli χ^2 olarak kullanılabilirler.

<u>İki boyutlu durum II'nin çözümü :</u>

1) (3-107) eşitliğine ek olarak, (3-102) eşitliğinin diferansiyel çözümü yapılarak elde edilen aşağıdaki koşul da bulunmaktadır :

$$\sum_i \delta m_{ij} = 0 \qquad j = 1, 2, \ldots\ldots, S\text{-}1 \qquad (3\text{-}113)$$

2) (3-107)' nin $-\lambda_{i.}$ ile çarpılması ve (3-113)'ün $-\lambda_{.j}$ ile çarpılmasından sonra, (3-105), (3-107) ve (3-113) eşitliklerinin toplanmasıyla halen isteğe bağlı Lagrange çarpanları $-\lambda_{i.}, -\lambda_{.j}$ kullanılarak :

$$f - \lambda_{i.} g - \lambda_{.j} h = \sum (m_{ij} - n_{ij})^2 / n_{ij} - \lambda_{i.} \sum_j m_{ij} - \lambda_{.j} \sum_i m_{ij} \qquad (3\text{-}114)$$

$$f = S,\ g = \sum_j m_{ij} \text{ ve } h = \sum_i m_{ij} \tag{3-115}$$

elde edilebilir. Daha sonra :

$$\frac{\delta f}{\delta m_{ij}} - \lambda_{i.} \frac{\delta g}{\delta m_{ij}} - \lambda_{.j} \frac{\delta h}{\delta m_{ij}} = 0 \Rightarrow m_{ij} = n_{ij}\left(1 + \lambda_{i.} + \lambda_{.j}\right) \tag{3-116}$$

olacaktır. Bu durumda ayarlama değeri satırlara oranlı değildir ama her hücreyi barındırır.

3) (3-116) eşitliğindeki Lagrange çarpanlarını hesaplamak için iki terim alt alta toplanabilir ve $r+s$-1 tane normal eşitlik elde edilir. Bunları çözmek zor olsa da, Deming and Stephan [49] makalelerinde bunun çözümünü vermişlerdir. Bu zorluk nedeni ile daha kolay bir işlem 3.11.1 bölümünde verilmiştir.

3.11.2. Basitleştirilmiş bir tekrarlı oranlar işlemi

İki boyutlu durum II, örneğin, şu eşitlikle alınır:

$$\lambda_{i.} = \left(1/n_{i.}\right)\left\{m_{i.} - \sum_j n_{ij}\lambda_{ij}\right\} - 1 \tag{3-117}$$

$\left(1/n_{i.}\right)\sum_j n_{ij}\lambda_{.j}$, i. satır için $\lambda_{.j}$'nin ağırlıklı ortalaması olarak düşünülebilir. (3-117) eşitliğinden (3-116) eşitliğine aktarma yapılacak olursa, (3-116) ayarlaması

$$m_{ij} = n_{ij}\left(m_{i.}/n_{i.} + \lambda_{.j} - i.av.\lambda_{.j}\right) \tag{3-118}$$

$i.av.\lambda_{.j}$: i. satır için $\lambda_{.j}$ 'nin ağırlıklı ortalaması

haline gelir. Benzer şekilde :

$$m_{ij} = n_{ij}\left(m_{.j}/n_{.j} + \lambda_{i.} - j.av.\lambda_{i.}\right) \tag{3-119}$$

$j.av.\lambda_{i.}$: j. sütun için $\lambda_{i.}$ 'nın ağırlıklı ortalaması

yazılabilir.

Hem (3-118) hem de (3-119) eşitliğinden, ayarlama (3-116)'nın neden sütunlar veya satırlar ile orantılı olmadığı kolayca anlaşılabilir. Bunun nedeni, herhangi bir hücre için $\lambda_{.j}$ değerinin, o satır için ortalama $\lambda_{.j}$ değerine eşit olma zorunluluğunun bulunmamasıdır. Aynı şey $\lambda_{i.}$ için de geçerlidir. Herhangi bir hücre için $\lambda_{i.}$ değeri, o sütun için ortalama $\lambda_{i.}$ değerine eşit olmak zorunda değildir.

Eğer basit orantılı ayarlama *i*. satırda yatay olarak yapılırsa :

$$m'_{ij} = n_{ij}(m_{i.} / n_{i.}) \tag{3-120}$$

yatay koşul (3-101)'ler doğrulanacaktır. Yani $m'_i = m_{i.}$ olarak bulunacak, $m'_{.j} = m_{.j}$ eşitliğinin her zaman tutmadığı görülecektir. Bunun nedeni, (3-120) eşitliğinin sadece kısmi bir ayarlama olan her bir m'_{ij}'yi etkilemesidir. (3-118) eşitliğinde görüldüğü gibi m'_{ij}'ler hataya sahiptirler. Bu hata, *j*. sütun için $\lambda_{.j}$ değeri ile *i*. satır için bütün $\lambda_{.j}$ değerlerinin ortalamasının eşitsizliğinden kaynaklanır.

Bu hata, işlemi çevirerek ve bu m'_{ij}'leri orantılı bir dikey ayarlamaya sokarak azaltılabilir.

$$m''_{ij} = m'_{ij}\left(m_{.j} / m'_{.j}\right) \tag{3-121}$$

Bu eşitlik, (3-119) eşitliğinin bir uygulaması olarak görülebilir. (3-119) eşitliğinde herhangi $\lambda_{i.}$ ve *j*. sütun için ortalama $\lambda_{i.}$ değeri arasındaki eşitsizlik gözardı edilmiştir. Dikey koşullar tatmin edici bulunmayabilir, ama yatay olanların da hepsi tatmin edici olmayabilir. Çünkü bazı satır toplamları bozulmuş olabilir. (3-120) eşitliği tarafından başlatılan döngü tekrarlanmalı, ve işlem, tablo kendi kendini üretir hale gelene kadar, ve bütün yatay ve dikey koşulları sağlayana kadar sürdürülmelidir.

3.12. Varyans Bileşenleri Modelleri Yolu ile Küçük Alan Kestirimi

Küçük alan kestiriminin birçok uygulamalarında, regresyon kestirimi bir temel oluşturmuştur. Bu regresyon modelinin bir vektör olarak ifadesi şu şekilde verilebilir:

$$Y = Xb + a \tag{3-122}$$

Bu durumda regresyon kestirimi:

$$\hat{y} = X\hat{b} \ , \tag{3-123}$$

Y'nin küçük alan kestirimlerini elde etmek için kullanılabilecek bir strateji olacaktır. *b*'nin kestirilmesi, (3-123) kestiricisinin bütün performansını etkiler. a_i'lerin stokastik dağıldığı varsayılacak olursa, bu durum bir küçük alan kestiricisi oluşturmada oldukça yardımcı olacaktır.

Daha genel bir ifade ile, eğer a_i'lerin tekil olmayan kovaryans A'ya sahip olduğu düşünülürse, genelleştirilmiş en küçük kareler ile b'nin kestirimi şu şekilde olur :

$$\hat{b} = \left(X'A^{-1}X\right)^{-1} X'A^{-1}y \tag{3-124}$$

Küçük alan kestirimi problemlerinde, Y_i'ler doğrudan gözlenmezler, fakat, örnek kestirimleri olan Y_i'ler, bazı veya bütün i'ler için biliniyor olabilir. Fay [50], Y üzerindeki örnekleme hatasının etkisinin, şu rastgele vektör ile gösterilebileceğini düşünmüştür :

$$d = Y - y \tag{3-125}$$

Bu durumda $Y = y + d$ olacaktır ve örnek kestirimi modeli,

$$Y = Xb + a + d \tag{3-126}$$

olarak değişecektir. a ve d'nin bağımsız olduğu varsayılmaktadır. Yani

$a_i \sim$ Çoklu normal (0,A)

$d_i \sim$ Normal (0,D)

Bu durumda b'nin en iyi doğrusal yansız kestirimleri :

$$\hat{b} = \left\{X'(A+D)^{-1}X\right\}^{-1} X'(A+D)^{-1}y \tag{3-127}$$

ile verilebilecektir.

Küçük alanlara ait y değerleri, sabit etkiler b'lere ve rastgele etkiler a'lara göre değişirler. d ve a'nın dağılımı üzerinde ortalaması alındığında, Harville'e [51] göre en iyi doğrusal yansız kestirim :

$$\hat{Y} = X\hat{b} + A(D+A)^{-1}(Y + X\hat{b}) \tag{3-128}$$

şekline dönüşecektir.

Fay [50], bazı bağımsız değişkenlerin, örnekleme hatasına bağlı olduğu bir durum ileri sürmüştür. k tane bağımsız Z değişkeni üzerine Y^* değişkeninin doğrusal regresyonu şu şekilde verilir:

$$Y^* = m_1 + Zb^* + e \tag{3-129}$$

1 : Bir sütun vektör

m : Regresyonda sabit terim

e : Hata terimi

$Y = Xb + a$

şeklindeki regresyon modelini, Z, bağımsız kabul edilmiş olarak, X matrisi ile yeniden ifade etmek olasıdır :

$$X = \begin{vmatrix} 1 & 0 & 0.....0 \\ 0 & 1 & 0.....0 \\ 0 & 0 & 1.....0 \\ \cdot & \cdot & \cdots\cdots \\ 0 & 0 & 0.....1 \\ 1 & 0 & 0.....0 \\ 0 & 1 & 0.....0 \\ \cdot & \cdot & \cdots\cdots \end{vmatrix} \qquad (3\text{-}130)$$

Hem Y^* hem de Z, Y'nin genişletilmiş vektöründe kullanılabilir:

$$Y' = (Y_1^*, Z_{11}, Z_{12}, \ldots\ldots, Z_{1k}, Y_2^*, Y_{21}, \ldots) \qquad (3\text{-}131)$$

Bundan sonra model şu hale gelir:

$$Y^* - Zb^* = m_1 + e \qquad (3\text{-}132)$$

Bağımsız değişkenleri, bağımlı değişken Y'nin kısmına aktarmak, problemin niteliğini değiştirecektir. Katsayı vektörü b^*, Y ve Z'nin ortalamalarını göstermektedir.

Eğer,

$S_{ZZ} =$ Z'lerin ortalamalar için ayarlanmış, ortalama vektörel-çarpımları

$S_{YZ} =$ Y' nin A ile birlikte, ortalamalar için ayarlanmış, ortalama vektörel çarpımları

ise, doğrusal regresyon için standart varsayımlar :

$$A = \begin{vmatrix} A^{*} & 0 & 0\ldots.. \\ 0 & A & 0\ldots.. \\ 0 & 0 & A^{*}\ldots. \\ \ldots & \ldots & \ldots \end{vmatrix} \qquad D = \begin{vmatrix} D^{*} & 0 & 0\ldots.. \\ 0 & D^{*} & 0\ldots.. \\ 0 & 0 & D^{*}\ldots. \\ \ldots & \ldots & \ldots \end{vmatrix} \tag{3-133}$$

$$A^{*} = \begin{vmatrix} S'_{YZ}S_{ZZ}^{-1}S_{YZ} & S'_{YZ} \\ & \\ S_{YZ} & S_{ZZ} \end{vmatrix} \qquad D^{*} = \begin{vmatrix} S_{e} & 0 \\ & \\ 0 & 0 \end{vmatrix} \tag{3-134}$$

şekline dönüştürülebilirler. S_e , Y'nin örnekleme hatasının bir sayısal değerini belirtir. A ve D gibi, (3-128)' deki $A(D+A)^{-1}$ matrisi köşegenseldir.

$$A(D+A)^{-1} = \begin{vmatrix} T^{*} & 0 & 0\ldots.. \\ 0 & T^{*} & 0\ldots.. \\ 0 & 0 & T^{*}\ldots.. \\ \ldots & \ldots & \ldots \end{vmatrix} \tag{3-135}$$

ve T^{*} kestiricisi

$$T^{*} = \begin{vmatrix} 0 & S'_{YZ}S_{ZZ}^{-1} \\ & \\ 0 & I \end{vmatrix} \tag{3-136}$$

I : Birim matris

olmak üzere verilmektedir. (3-128) eşitliği Z'nin her bir bileşenini, doğrudan gözlenmiş değeriyle kestirir. Ancak, gözlenmiş Z_i değerine dayanan y_i'nin regresyon kestirimini, en iyi doğrusal yansız kestirim olarak alır. Çünkü $S_{ZZ}^{-1}S_{YZ}$, b^{*}'ın sıradan en küçük kareler kestirimidir [50].

3.13. Üç Tane Genel Karışık Doğrusal Model ve Küçük Alan Ortalamalarının Hesaplanması

Genel doğrusal karışık model (general linear mixed model) :

$$y = X\beta + Zv + e \tag{3-137}$$

şeklinde verilebilir.

Y: örnek gözlemlerinin vektörü

X ve Z : Bilinen matrisler

v ve e, 0 ortalama, ve sırasıyla G ve R kovaryans matrisleri ile bağımsız dağılmışlardır ve varyans bileşenleri adında bazı θ parametrelerine dayanmaktadırlar. Henderson [52], bilinen bir θ değeri için, $\mu = l'\beta + m'v$ 'nin en iyi doğrusal yansız kestiricisinin :

$$t(\theta, y) = l'\tilde{\beta} + m'GZ'V^{-1}(y - X\tilde{\beta}) \tag{3-138}$$

$V = R + ZGZ' =$ y'nin varyans-kovaryans matrisi

$\tilde{\beta} = (X'V^{-1}X)^{-1}(X'V^{-1}Y) = \beta$ 'nın genellenmiş en küçük kareler kestirimi

olacak şekilde verilebileceğini göstermiştir.

3.13.1. İç içe-hata regresyon modeli

Battese, Harter et al. [53], Iowa'da ilçeler (küçük alanlar) için ürün dönüm ortalamasını hesaplamak üzerinde çalışırken, iç içe-hata regresyon modelini önermişlerdir. İç içe-hata regresyon modeli, genel karışık doğrusal modelin özel bir durumudur. Belirtilen çalışmada, araştırma ile elde edilen verilerin yanısıra, uydu verilerinden de yararlanılmıştır. Önerilen model :

$$y_{ij} = X'_{ij}\beta + v_i + e_{ij} \qquad i=1,........,t \quad , \quad j=1,........,n_i \tag{3-139}$$

y_{ij} : i. örnek alanında j.örneklenmiş birim için ilgilenilen özellik

$X_{ij} = (x_{ij1}........x_{ijk})'$: yardımcı değerlere ait bir k vektör

$\beta = (\beta_1........\beta_k)$: Bilinmeyen parametrelerin bir k vektörü

n_i : i. küçük alanda gözlenen örneklenmiş birimlerin sayısı ($\sum n_i = n$)

modelidir. Rastgele hatalar: v_i'lerin, bağımsız olduğu varsayılır. Dağılımları $N(0, \sigma_v^2)$ 'dir. Bunlar dağılımları $N(0, \sigma_e^2)$ olan e_{ij}'lerden de bağımsızdırlar.

i. alan için ortalama şu şekilde yazılabilir :

$$\mu_i = \overline{X}'_i\beta + v \tag{3-140}$$

Burada, i. alan içindeki birimlerin sayısı olan N_i'nin geniş olduğu, ve i. alan için $\overline{X}_i$ 'nin, x_{ij}'nin bilinen ortalamalarının vektörü olduğu varsayılmaktadır. μ_i, sabit etkiler β 'ların ve rastgele etkiler v_i'lerin gerçek değerlerinin doğrusal bir kombinasyonudur ve i. alan için v_i verildiğinde y_{ij}'nin koşullu ortalaması olarak yorumlanabilir.

İç-içe hata regresyon modelinde, μ_i için en iyi doğrusal yansız kestirici aranacak olursa :

$$V = diag\,(V_1 \ldots\ldots\ldots V_t) \tag{3-141}$$

$$V_i = \sigma_e^2\, I_{n_i} + \sigma_v^2 1_{n_i} 1'_{n_i} \tag{3-142}$$

ve bundan dolayı

$$V^{-1} = diag\,(V_1^{-1} \ldots\ldots\ldots V_t^{-1}) \tag{3-143}$$

$$V_i^{-1} = (\sigma_e^2)^{-1} I_{n_i} - \gamma_i n_i^{-1} (\sigma_e^2)^{-1} 1_{n_i} 1'_{n_i} \tag{3-144}$$

ve

$$\gamma_i = \sigma_v^2 (\sigma_v^2 + \sigma_e^2 n_i^{-1})^{-1} \tag{3-145}$$

olur.

matris çevriminde standart bir sonuç kullanılmıştır [54].

$$(A + uv')^{-1} = A^{-1} - A^{-1}uv'A^{-1} / (1 + v'A^{-1}u) \tag{3-146}$$

Sonuç olarak, $I = \overline{X}_i$, ve i. pozisyonda 1 olacak şekilde $m = (0 \ldots\ldots\ldots 010 \ldots\ldots\ldots 0)'$ alınırsa ve $(\sigma_v^2 / \sigma_e^2)(1 - \gamma_i) = \gamma_i / n_i$ olduğu not edilirse, (3-138)'den μ_i 'nin en iyi doğrusal yansız kestirimi :

$$t_i(\sigma^2, y) = \overline{X}_i'\widetilde{\beta} + \gamma_i(\overline{y}_i - \overline{x}_i'\widetilde{\beta}\ ) \tag{3-147}$$

$$\sigma^2 = (\sigma_v^2 \sigma_e^2)'$$

$\overline{x}_i$:x_{ij}'nin, i. alan için örnek ortalaması

şeklinde bulunur.

3.13.2. Rastgele-regresyon katsayısı modeli

Dempster, Rubin et al. [55] tarafından, rastgele β ile daha genel bir model önerilmiştir. Modelin değişken x, ve orijinden geçen regresyon doğrusundan oluşan özel bir durumu şu şekilde verilebilir:

$$y_{ij} = \beta_i x_{ij} + e_{ij} = \beta x_{ij} + v_i x_{ij} + e_{ij} \qquad j = 1,........,n_i, \quad i = 1,........,t \qquad (3\text{-}148)$$

$$\beta_i = \beta + v_i$$

v_i ve e_i (3-139) modelinde olduğu gibidir.

i. alan için ortalama $\mu_i = \overline{X}_i\beta + \overline{X}_i v_i$, sabit etki β ve rastgele etki v_i'nin gerçek değerinin doğrusal bir kombinasyonudur. μ_i'nin en iyi doğrusal yansız kestirimi:

$$t_i(\sigma^2, y) = \overline{X}_i\widetilde{\beta} + \widetilde{\gamma}\overline{X}_i(\widetilde{\beta}_i - \widetilde{\beta}) \qquad (3\text{-}149)$$

$$\sigma^2 = (\sigma_v^2\sigma_e^2)',$$

$$\widetilde{\gamma}_i = \sigma_v^2\left(\sigma_v^2 + \sigma_e^2 / \sum\nolimits_j x_{ij}^2\right)^{-1}$$

$$\widetilde{\beta}_i = \sum\nolimits_j x_{ij}y_{ij} / \sum\nolimits_j x_{ij}^2$$

ve

$$\widetilde{\beta} = \sum \widetilde{\gamma}_i\widetilde{\beta}_i / \sum \widetilde{\gamma}_i \qquad (3\text{-}150)$$

şeklinde verilebilir.

3.13.3. Fay-Herriot modeli

Nüfusu 1000'den az olan küçük alanlar için kişi başına düşen geliri hesaplama çalışmasında, Fay and Herriot [56], μ_i'ye bağlı değişkenler X_i = $(x_{i1}........x_{ik})$'nın bir *k* vektörünün, her bir *i* alanı için bilindiğini varsayarlar. Buna ek olarak μ_i'ler bağımsız ve $N(X_i'\beta, A)$ dağılımlı ve β bilinmeyen parametrelerin bir *k* vektörüdür. Bir sonraki varsayım ise $\mu = (\mu_1........\mu_t)'$ verildiğinde, örnek ortalaması vektörü $\bar{y} = (\bar{y}_1,........\bar{y}_t)' = col_{1\le i\le t}(\bar{y}_i)$'nin, $N(\mu, D)$, D = *diag* ($D_1........D_t$) (D_i köşegensel elemanları bilinmekte) dağılımlı olduğudur. Model, bir doğrusal model olarak ifade edilebilir:

$$\bar{y}_i = \mu_i + e_i \quad \text{ve} \quad \mu_i = X_i'\beta + v \quad i = 1,........t \qquad (3\text{-}151)$$

$e = (e_1........e_t)'$ ve $v = (v_1........v_t)'$ bağımsız ve sırasıyla *N (0,D)* ve *N(0,AI)* dağılımlıdırlar. İç içe - hata regresyon modelinden farklı olarak, birim düzeyinde yardımcı bilgiye Fay-Herriot modelinde gerek duyulmaz.

Fay-Herriot modeli için μ_i'nin en iyi doğrusal yansız kestirimi şu şekilde elde edilebilir:

$$t_i(A, \bar{y}) = X_i'\widetilde{\beta} + \left(A / (A + D_i)\right)(\bar{y}_i - X_i'\widetilde{\beta}) \qquad (3\text{-}152)$$

$$\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}\bar{y}$$

$$V = diag\,(A+D_1 .... A+D_t) \quad \text{ve} \quad X = col_{1 \le i \le t}(X_i')$$

Normallik varsayımı altında (3-152) kestiricisi aynı zamanda bir Bayes kestiricisidir [56].

4. BİR UYGULAMA ÇALIŞMASI

Bu bölümde küçük alan kestirim teknikleri kullanılarak, elde edilebilen veriye göre, Bolu ili işsizlik oranının kestirimi üzerine uygulamalar yapılacaktır. Elde edilebilen veriler, bütün tekniklerin uygulanmasına imkan vermemektedir. Bu yüzden uygulama çalışması, veriye ve konuya uygun tekniklerin uygulanmasıyla sınırlandırılmıştır.

4.1. Bir Gerçek Oranlar Tekniği Uygulaması

Yerel alanlardaki işsizlik ve nüfus oranının geniş alanlarla aynı oranda değiştiği varsayılırsa, Bolu İli küçük alanındaki işsizlik oranını, Bolu ilini kapsayan daha geniş bir alandaki nüfus oranından kestirmek olasıdır. Bolu ilini kapsayan daha geniş bir alan seçiminde, Devlet Planlama Teşkilatı'nın (D.P.T.) 1996 yılında yayınlamış olduğu İllerin Sosyo Ekonomik Gelişmişlik Sıralaması Araştırmasına [57] göre Bolu İli'nin de içinde bulunduğu, üçüncü derece gelişmiş iller ile Hacettepe Üniversitesi Nüfus Etütleri Enstitüsü'nün (H.Ü.N.E.E.) 1994 yılında yayınlanmış olduğu Türkiye Nüfus ve Sağlık Araştırmasında [58], Bolu İli'nin de yer aldığı Türkiye orta bölge illerinden, her iki gruba da giren iller seçilmiştir. Böylece Bolu iline sosyo ekonomik ve coğrafi bakımdan benzerlik gösteren illerden oluşmuş daha geniş bir bölge elde edilmiştir. Bu iller :

1- Afyon

2- Amasya

3- Bilecik

4- Bolu

5- Kırşehir

6- Konya

7- Kütahya

8- Nevşehir

9- Niğde

10-Uşak

illeridir. Bu illere ait nüfus verileri, Devlet İstatistik Enstitüsü'nün (D.İ.E.) 1985 yılı genel nüfus sayımı [59] ve 1990 yılı genel nüfus sayımı [60] sonuçlarından derlenmiştir. Seçilen illerin 1985 yılı erkek nüfusu, işsiz erkek sayısı, işsizlik oranı ve 1990 yılına ait erkek nüfusu ile erkek işsiz sayıları Çizelge 4.1' de verilmiştir.

Çizelge 4.1. Seçilen on ile ait 1985 yılı erkek nüfusu, erkek işsiz sayısı, erkek işsizlik oranı ve 1990 yılına ait erkek nüfusu ile erkek işsiz sayıları

	Erkek Nüfusu	Erkek İşsiz Sayısı	Erkek İşsizlik Oranı	Erkek Nüfusu	Erkek İşsiz Sayısı
Şehirler	1985	1985	1985	1990	1990
1- Afyon	333328	8551	0.026	368950	11599
2- Amasya	181925	5768	0.032	179895	5818
3- Bilecik	83868	1886	0.023	90978	2144
4- Bolu	250806	4738	0.019	268870	6599
5- Kırşehir	125704	4775	0.038	124639	5243
6- Konya	878563	31298	0.036	873886	33086
7- Kütahya	273694	4857	0.018	289866	7524
8- Nevşehir	136300	2671	0.020	143812	3486
9- Niğde	272207	6282	0.023	148163	3690
10-Uşak	134066	3875	0.029	144224	4367
Toplam	2670461	74701	0.260	2633283	83556

Buna göre:

$$\text{1985 yılına ait bölge erkek işsizlik oranı} = \frac{74701}{2670461} = 0.028$$

$$\text{1990 yılına ait bölge erkek işsizlik oranı} = \frac{83556}{2633283} = 0.032$$

$$\text{1990 yılı Bolu İli erkek işsizlik oranı kestirimi} = \text{1990 yılı bölge erkek işsizlik oranı} \times \frac{\text{1985 yılı Bolu İli erkek işsizlik oranı}}{\text{1985 yılı bölge erkek işsizlik oranı}}$$

$$= 0.032 \times \frac{0.019}{0.028}$$

$$= 0.022$$

Aynı kestirimi kadınlar için de yapmak olasıdır. Seçilen on ile ait 1985 yılı kadın nüfusu, işsiz kadın sayısı, kadın işsizlik oranı ve 1990 yılına ait kadın nüfusu ile kadın işsiz sayıları, Çizelge 4.2'de verilmiştir.

Çizelge 4.2. Seçilen on ile ait 1985 yılı kadın nüfusu, kadın işsiz sayısı, kadın işsizlik oranı ve 1990 yılına ait kadın nüfusu ile kadın işsiz sayıları

	Kadın Nüfusu	Kadın İşsiz Sayısı	Kadın İşsizlik Oranı	Kadın Nüfusu	Kadın İşsiz Sayısı
Şehirler	1985	1985	1985	1990	1990
1- Afyon	333650	1464	0.004	370273	2913
2- Amasya	176364	1535	0.009	177296	1990
3- Bilecik	77041	551	0.007	84548	459
4- Bolu	253972	1185	0.005	267999	1932
5- Kırşehir	134452	1283	0.010	132223	1389
6- Konya	890487	5868	0.007	876417	4962
7- Kütahya	269690	611	0.002	288154	1231
8- Nevşehir	141829	441	0.003	145697	692
9- Niğde	288179	872	0.003	157698	1108
10- Uşak	137195	987	0.007	146059	1332
Toplam	2702859	14788	0.057	2646364	18008

Buna göre :

$$\begin{array}{l}\text{1985 yılına ait bölge}\\ \text{kadın işsizlik oranı}\end{array} = \frac{14788}{2702859} = 0.005$$

$$\begin{array}{l}\text{1990 yılına ait bölge}\\ \text{kadın işsizlik oranı}\end{array} = \frac{18008}{2646364} = 0.007$$

$$\begin{array}{l}\text{1990 yılı Bolu İli kadın}\\ \text{işsizlik oranı kestirimi}\end{array} = \begin{array}{l}\text{1990 yılı bölge}\\ \text{kadın işsizlik oranı}\end{array} \times \frac{\begin{array}{l}\text{1985 yılı Bolu İli kadın}\\ \text{işsizlik oranı}\end{array}}{\begin{array}{l}\text{1985 yılı bölge kadın}\\ \text{işsizlik oranı}\end{array}}$$

$$= 0.007 \times \frac{0.005}{0.005}$$

$$= 0.007$$

olarak bulunabilir.

4.2. Bir Sentetik Kestirim Uygulaması

Küçük alanların kendi aralarında homojen olduğu kabul edilirse, herhangi bir x özelliğinin toplam oranını, sentetik kestirim yolu ile elde etmek olasıdır. x özelliği işsizlik olarak ele alınacak olursa, bir küçük alan olarak Bolu İli'nin işsizlik oranını, alt grup olarak ele alınan, rastgele örnekleme ile seçilmiş ilçelerinin işsizlik oranından kestirmek olasıdır. Buna göre Bolu İli içinde yer alan, merkez ilçe dahil 16 ilçeden dokuz ilçe, rastgele örnekleme ile seçilmiştir. Bu ilçeler :

1- Akçakoca

2- Düzce

3- Gölyaka

4- Göynük

5- Kıbrısçık

6- Merkez

7- Mengen

8- Mudurnu

9- Seben

İşsizlik oranını, ilçeleri merkez ve kırsal kesim olarak ayırarak, aynı zamanda cinsiyete dayalı olarak da hesaplamak olasıdır. Uygulamada kullanılan veriler, Devlet İstatistik Enstitüsü (D.İ.E.) tarafından yapılan 1990 genel nüfus sayımında, Bolu İli için yayınlanmış verilerden [61] derlenmiştir.

Seçilmiş ilçelerin merkez ve kırsal kesim nüfusları ile işsizlik oranları, kadın ve erkek nüfusa dayalı olarak sırasıyla Çizelge 4.3 ve Çizelge 4.4'de görülmektedir.

Çizelge 4.3. İlçelerin erkek nüfusa dayalı merkez ve kırsal kesim nüfusları ile işsizlik oranları

	İşsizlik Oranı (Erkek) (1990)	İlçe Merkezleri Nüfusu (Erkek) (1990)	Kırsal Kesim Nüfusu (Erkek) (1990)
İlçeler	$x'_{.g}$	N_{1g}	N_{2g}
1- Akçakoca	0.045	6817	9720
2- Düzce	0.024	31684	46642
3- Gölyaka	0.024	2143	7630
4- Göynük	0.008	1904	7766
5- Kıbrısçık	0.011	850	2261
6- Merkez	0.021	32981	25809
7- Mengen	0.006	2327	7040
8- Mudurnu	0.005	2686	10558
9- Seben	0.024	2305	3205
Toplam	0.168	83697	120631

Çizelge 4.4. İlçelerin kadın nüfusa dayalı merkez ve kırsal kesim nüfusları ile işsizlik oranları

	İşsizlik Oranı (Kadın) (1990)	İlçe Merkezleri Nüfusu (Kadın) (1990)	Kırsal Kesim Nüfusu (Kadın) (1990)
İlçeler	$x'_{.g}$	N_{1g}	N_{2g}
1- Akçakoca	0.013	6765	9537
2- Düzce	0.006	30194	47806
3- Gölyaka	0.003	1788	8214
4- Göynük	0.003	1908	8498
5- Kıbrısçık	0.011	782	2823
6- Merkez	0.010	27808	26998
7- Mengen	0.006	1971	7644
8- Mudurnu	0.003	2551	11358
9- Seben	0.005	1860	3650
Toplam	0.06	75627	126528

Küçük alan içindeki işsizlik oranı x için kestirim :

$$\hat{x}_h = \sum_g \hat{x}_{hg} = \sum_g \left(\frac{N_{hg}}{N_{.g}} \right) x'_{.g} \tag{4-1}$$

$\hat{x}_h$: küçük alan h'deki toplam işsizlik oranının sentetik kestirimi

h : Küçük alanlar : Bolu İli ilçe merkezleri ($h = 1$) ve kırsal kesim ($h = 2$)

g : İlçeler (alt gruplar)

$N_{.g}$: İlçe Merkezleri ve kırsal kesim nüfusu toplamı

$x'_{.g}$: İşsizlik oranı

Formül, ilçe merkezleri ve kırsal kesim için, cinsiyete göre ayrı ayrı uygulanacak olursa :

I. 1990 yılına ait Bolu İli, ilçe merkezleri erkek işsizlik oranı kestirimi :

İl bazında erkek işsizlik oranı ve ilçe merkezleri erkek nüfusu gözönüne alınırsa (Bkz. Çizelge 4.3) :

$$\hat{x}_1 = \sum_{g=1}^{9} \hat{x}_{1g} = \sum_{g=1}^{9} \left(\frac{N_{1g}}{N_{.g}} \right) x'_{.g}$$
$$= \frac{1918.508}{204328} = 0.009$$

II. 1990 yılına ait Bolu İli, kırsal kesim erkek işsizlik oranı kestirimi :

İl bazında erkek işsizlik oranı ve kırsal kesim erkek nüfusu gözönüne alınırsa (Bkz Çizelge 4.3) :

$$\hat{x}_2 = \sum_{g=1}^{9} \hat{x}_{2g} = \sum_{g=1}^{9} \left(\frac{N_{2g}}{N_{.g}} \right) x'_{.g}$$
$$= \frac{2540.866}{204328} = 0.012$$

II. 1990 yılına ait Bolu İli, ilçe merkezleri kadın işsizlik oranı kestirimi :

İl bazında kadın işsizlik oranı ve ilçe merkezleri kadın nüfusu gözönüne alınırsa (Bkz. Çizelge 4.4) :

$$\hat{x}_1 = \sum_{g=1}^{9} \hat{x}_{1g} = \sum_{g=1}^{9} \left(\frac{N_{1g}}{N_{.g}} \right) x'_{.g}$$

$$= \frac{595.658}{202155} = 0.003$$

IV. 1990 yılına ait Bolu İli, kırsal kesim kadın işsizlik oranı kestirimi :

İl bazında kadın işsizlik oranı ve kırsal kesim kadın nüfusu gözönüne alınırsa (Bkz. Çizelge 4.4) :

$$\hat{x}_2 = \sum_{g=1}^{9} \hat{x}_{2hg} = \sum_{g=1}^{9} \left(\frac{N_{2g}}{N_{.g}} \right) x'_{.g}$$

$$= \frac{860.174}{202155} = 0.004$$

olarak hesaplanabilir.

4.3. Bir Tekrarlı Oran Uydurulması Uygulaması

Bu bölümde, bölüm 3.11 'deki, II. durumda verilen normal eşitliklerden λ'ların elde edilmesi güçlüğü nedeni ile, bir basitleştirilmiş tekrarlı oranlar işlemi örneği verilecektir.

Bolu İli küçük alan olarak ele alındığında, yaş gruplarına göre çalışan ve işsiz kişi sayısını, küçük alandaki rastgele seçilmiş bir alt alandan (örnek) yardımıyla kestirmek olasıdır. Bu örnek alt alan, Bolu İli'nde yer alan bir ilçe olarak rastgele seçilebilir. Bu ilçe Düzce İlçesi olarak seçilmiştir.

Metodun uygulama aşamasında Bolu İli'ne ait yaş guruplarına göre çalışan ve işsiz kişilerin toplam sayılarının bilindiği farz edilmektedir. Buna karşın, yaş gruplarına göre çalışan ve işsiz kişilerin, ayrı ayrı sayıları bilinmemektedir. Bu hücre değerleri bilinmiyor varsayılmalarına karşın, Bolu İli ana kütle tablosunda, kestirimlerle karşılaştırılabilmeleri açısından verilmiştir. Sözkonusu ana kütle hücre değerlerinin kestirimi, örnek alt alan olarak seçilen Düzce İlçesi'ndeki, yaş gruplarına göre çalışan ve çalışmayan kişilerin ayrı ayrı sayılarından oluşturulmuş tablo yardımıyla yapılacaktır.

Uygulamada kullanılan veri, Devlet İstatistik Enstitüsü'nün (DİE) 1990 yılında yapmış olduğu, genel nüfus sayımında, Bolu İli için yayınlanan verilerden [61] elde edilmiştir. Bolu İli (ana kütle) 1990 yılı için yaş gruplarına göre çalışan ve işsiz kişi sayısının esas (kestirimle elde edilmeye çalışılacak) tablosu Çizelge 4.5'de görülmektedir. Örnek alt alan olarak seçilen Düzce İlçesi'nin, 1990 yılı için yaş gruplarına göre çalışan ve işsiz sayısının başlangıç tablosu ise, Çizelge 4.6'da verilmiştir.

Çizelge 4.5. Bolu ili (ana kütle) 1990 yılı, yaş gruplarına göre çalışan ve işsiz sayıları (esas tablo)

		Yaş Grupları					
		12 - 14	15 - 19	20 - 24	25 - 29	29 -	$N_{i.}$
İş Durumu	Çalışan	14573	35237	31169	34741	163426	279146
	İşsiz	517	2744	2156	1096	2009	8522
	$N_{.j}$	15090	37981	33325	35837	165435	287668

Düzce için aynı tabloyu oluştururken,

$$m_{i.} = N_{i.} n / N \qquad (4\text{-}2)$$

ve

$$m_{.j} = N_{.j} n / N \qquad (4\text{-}3)$$

değerleri hesaplanır ve tabloya eklenir.

Çizelge 4.6. Düzce İlçesi (örnek) 1990 yılı, yaş gruplarına göre çalışan ve işsiz sayıları (başlangıç tablosu)

		Yaş Grupları						
		12 - 14	15 - 19	20 - 24	25 - 29	29 -	$n_{i.}$	$m_{i.}$
İş Durumu	Çalışan	409	1987	2047	2903	10506	17852	19540
	İşsiz	101	657	560	353	700	2371	597
	$n_{.j}$	510	2644	2607	3256	11206	20223	
	$m_{.j}$	1056	2659	2333	2509	11580		20137

İlk olarak *i.* satırda basit oransal ayarlamayı yapmak gerekir. Tablodaki hücreler için aşağıdaki dönüşüm yapılır :

$$m'_{ij} = n_{ij}(m_{i.} / n_{i.}) \qquad (4\text{-}4)$$

yeni oluşan tablo, Çizelge 4.7'de verilmiştir.

Çizelge 4.7. Birinci tekrar yaş grupları tablosu

		Yaş Grupları						
		12 - 14	15 - 19	20 - 24	25 - 29	29 -	$m'_{i.}$	$m_{i.}$
İş	Çalışan	448	2176	2241	3179	11504	19548	19540
Durumu	İşsiz	25	166	141	89	176	597	597
	$m'_{.j}$	473	2342	2382	3268	11680	20145	
	$m_{.j}$	1056	2659	2333	2509	11580		20137

Bu tablo incelenirse,

$$m'_{i.} = m_{i.} \qquad (4\text{-}5)$$

ve

$$m'_{.j} \neq m_{.j} \qquad (4\text{-}6)$$

olduğu görülür. Bu durumda yatay koşullar sağlanmış, dikey koşullar sağlanamamıştır. Bir tekrar daha yapmak gerekmektedir. Bunun nedeni birinci tekrarın sadece kısmi bir ayarlama olmasıdır. Bu eksiklik, işlemi sürdürüp, m'_{ij} 'lerin dikey olarak da ayarlanması ile giderilebilir. Şu eşitlikler hesaplanır :

$$m''_{ij} = m'_{ij}(m_{.j} / m'_{.j}) \qquad (4\text{-}7)$$

ve yeni tablo oluşturulur. Oluşan yeni tablo Çizelge 4.8'de verilmiştir.

Çizelge 4.8. İkinci tekrar yaş grupları tablosu

		Yaş Grupları						
		12 - 14	15 - 19	20 - 24	25 - 29	29 -	$m''_{i.}$	$m_{i.}$
İş	Çalışan	1000	2470	2194	2441	11400	19505	19540
Durumu	İşsiz	56	188	138	68	174	624	597
	$m''_{.j}$	1056	2658	2332	2509	11574	20129	
	$m_{.j}$	1056	2659	2333	2509	11580		20137

Koşullara tekrar bakıldığında,

$$m''_{i.} \neq m_{i.} \qquad (4\text{-}8)$$

ve

$$m''_{.j} = m_{.j} \tag{4-9}$$

olduğunu görülür. Bu durumda ise, yatay koşullar sağlanamamış, dikey koşullar sağlanmıştır. Tekrarlar, koşullar sağlanıncaya ve tablo kendini tekrar edinceye kadar sürdürülmelidir.

$$\tilde{m}_{ij} = m''_{ij}(m_{i.} / m''_{i.}) \tag{4-10}$$

eşitliği ile bir tekrar daha yapılmalıdır. Elde edilen yeni tablo Çizelge 4.9'da verilmiştir.

Çizelge 4.9. Üçüncü tekrar yaş grupları tablosu (sonuç tablosu)

		Yaş Grupları						
		12 - 14	15 - 19	20 - 24	25 - 29	29 -	$\tilde{m}_{i.}$	$m_{i.}$
İş	Çalışan	1002	2475	2198	2446	11423	19544	19540
Durumu	İşsiz	54	180	132	65	167	598	597
	$\tilde{m}_{.j}$	1056	2655	2330	2511	11579	20131	
	$m_{.j}$	1056	2659	2333	2509	11580		20137

En sonuncu tabloda, hesaplamalardaki sayıların yuvarlanmasından doğan küçük farklar dikkate alınmazsa, tablonun,

$$\tilde{m}_{i.} = m_{i.} \tag{4-11}$$

ve

$$\tilde{m}_{.j} = m_{.j} \tag{4-12}$$

yatay ve dikey koşullarını sağladığı görülür. Bu durumda tablo kendini tekrar eder hale gelmiştir. Bu tablo yardımıyla esas ana kütle tablosundaki hücre değerleri N_{ij}'leri kestirmek olasıdır. Bu kestirim m_{ij}'lerin ters örnekleme oranı $\frac{N}{n}$ ile genişletilmesiyle yapılabilir.

$$N_{ij} = m_{ij}\frac{N}{n} \tag{4-13}$$

eşitliğini kullanarak, ana kütle (Bolu ili) N_{ij} değerlerini kestirerek, kestirim tablosu oluşturulabilir. Bu tablo Çizelge 4.10'da görülmektedir.

Çizelge 4.10. Bolu ili 1990 yılı yaş gruplarına göre çalışan ve işsiz sayıları (kestirilmiş tablo)

		Yaş Grupları					
		12 - 14	15 - 19	20 - 24	25 - 29	29 -	$N_{i.}$
İş	Çalışan	14253	35207	31267	34794	162492	278013
Durumu	İşsiz	768	2561	1878	925	2376	8508
	$N_{.j}$	15021	37768	33145	35719	164868	286521

Son tablodaki kestirilmiş hücre değerlerinin, esas tablo hücre değerlerine oldukça yakın sonuç verdiği görülmektedir. Aradaki küçük farklar, hesaplamalardaki yuvarlamalardan ileri gelmektedir. Kestirilen işsizlik rakamlarından, Bolu İli içindeki işsizlerin, yaşlara göre, toplam işgücü içindeki kaba oranları bulunabilir. Bu oranlar, Çizelge 4.11'de verilmiştir.

Çizelge 4.11. Bolu ili 1990 yılı yaşlara göre kestirilen işsizlik oranları

	Yaş Grupları				
	12 - 14	15 - 19	20 - 24	25 - 29	29 -
İşsizlik Oranı	0.05	0.068	0.057	0.026	0.014

5. SONUÇ ve YORUMLAR

Küçük alan kestirimleri problemini çözmek için çeşitli ve birbirine alternatif teknikler bulunmakla beraber, hangi durumda hangi tekniğin uygun olduğu konusunda kesin bir ayırım yoktur. Buna rağmen, söz konusu problemin konusuna, yapısına ve veri ihtiyacına göre en iyi ya da en uygun bir teknik veya tekniklerden söz edilebilir.

Küçük alan kestiricileri, genel olarak doğrudan ve dolaylı olarak iki sınıfta toplanmaktadırlar. Doğrudan kestiriciler, çoğunlukla yetersiz ve küçük örnek büyüklüğü yüzünden büyük varyanslı olduklarından, kullanışlı değildirler. Yaygın olarak kullanılan kestiriciler, dolaylı kestiricilerdir. Küçük alan kestiricileri denildiğinde kastedilen, çoğunlukla dolaylı kestiriciler olmaktadır.

Küçük alan kestiricilerinin özelliklerine genel olarak bakılacak olursa, semptomatik kayıt tekniklerinin en eski teknikler olduğu, buna karşın etkinliklerinin çok fazla olmadığı görülmektedir. Bu tekniklerin en büyük eksiklikleri, güncel kayıtlardan elde edilen verilere bağımlı olmalarıdır. Bu tür veriler her zaman kolayca elde edilememekte veya güncelliklerini yitirmiş olmaktadırlar.

Regresyon - semptomatik ve örnek - regresyon metotları, yıllık gelir kestirimi gibi sürekli değişkenler içeren çalışmalara daha çok uyum göstermektedirler. Bu metotlardan örnek- regresyon metodu, daha büyük potansiyel ve doğruluğa sahiptir. Elde güncel örnek bulunmadığı hallerde, oran ya da fark-korelasyon yaklaşımları, ilgilenilen değişkene ait veriler önceki araştırma veya nüfus sayımlarında elde edilmiş olduğu sürece uygulanabilmektedirler.

Temel birim metodu doğrusal durumlarda iyi performans göstermesine karşın, örnek-regresyon metodu bu gibi durumlarda biraz daha başarılıdır, dolayısıyla, kullanılması daha uygun olur. Bu iki metot doğrusal olmayan durumlarda karşılaştırıldıklarında, örnek - regresyon metodu kullanıcıları doğrusal bir regresyon formülü kullanmaya zorlamamaktadır. Bazı durumlarda doğrusal olmayan regresyon eşitlikleri kullanmak gerekebilir. Bu durumda örnek - regresyon metodu yararlı bir araç olmaktadır. Ayrıca, temel birim metodunun bir dezavantajı vardır. Bu da uygulanmasının daha zor olmasıdır. Çünkü metodu uygulamak için küçük alanları daha küçük temel birimlere bölmek gerekmektedir. Bu nedenlerden dolayı, örnek - regresyon metodu daha geniş imkanlar sunmaktadır.

Sentetik metot, frekanslı veriler için en popüler ve en kolay yaklaşım gibi görünmektedir. Elde hiç bir kesin model varsayımı yokken bile, sentetik metot ilgilenilen değişken, ve alt grupları tanımlamada kullanılan değişkenler arasındaki ilişkinin sağlamlığını varsaymakta ve işlem yapabilmektedir. Buna karşın bu varsayım geçerli değilse, yanlılık bir sorun haline gelmektedir.

Sentetik ve örnek - regresyon metotlarının karşılaştırılmasında, ortak nokta fazla görünmemektedir. Bunun nedeni iki metodun farklı veri ihtiyaçları olmasıdır. Bunun yerine iki metodun kombinasyonu üzerinde durulabilir. Böylece iki metodun da kuvvetli yanlarından yararlanılabilir. Bu çeşit bir kombinasyon, karma sentetik-regresyon metodunu oluşturmuştur. Bu metot, incelenilen metotların tümünden daha geniş potansiyele sahip görünüyorsa da, ilgilenilen değişkene ait örnek verisinin elde olmasını gerektirmektedir.

İstatistiksel güvenirliğe sahip bulunan örnek kestirimlerinin bulunmadığı durumlarda, James - Stein kestirimleri yararlı bir çözümdürler. James - Stein işlemi, bazı koşullarda, verilerin değiştirilmesinden çok, eldeki verinin maksimum düzeyde kullanılması için etkin bir araçtır.

Kategorik veri analizi yaklaşımı, kestirilecek özellikle ilgili olan değişkenlerin kullanımında geniş bilgi ve esneklik sağlamaktadır. Bu yaklaşımı kullanan Y.K.K. (yapı koruyan kestirim) metodu, güncel bilgilerle bir dağıtım yapısı, ve önceki bilgi ve ilişkileri koruyan, bir ortaklık yapısı ile çalışmaktadır. Y.K.K. kestiricileri yanlılık sorunu çıkarabilirler. Diğer bir kategorik veri analizi yaklaşımı olan T.O.U. (tekrarlı oran uydurulması), kategorize edilmiş verilerin tablolanması ve örnek tablolarında yapılan bir takım ayarlamalar ile ana kütle değerlerinin kestirilmesi üzerine geliştirilmiştir.

Üç genel karışık doğrusal model için, iki aşamalı kestiricilerin, bilinen regresyon - sentetik veya regresyon kestiricilerinin etkinliğini arttırdığını söylemek olasıdır. Buna ek olarak, hata kareleri ortalamaları kestirimi, iki aşamalı kestirici ile ilgili belirsizliğe güvenilir bir ölçüm getirmektedir.

Sonuç olarak, verilerin kalitesi ve elde edilebilirliği, teknik seçimini oldukça etkilemektedir. Hiçbir teknik tek başına küçük alanlar için en iyi teknik olmadığından, tekniklerin güçlü yanlarından yararlanabilmek için, bunların kombinasyonu yoluna gitmek iyi bir yaklaşımdır. Açıklanan tekniklerin veri gereksinimleri, varsayımları ve sonuçlarının karşılaştırılması, Çizelge 5.1'de verilmiştir.

Çizelge 5.1. Küçük alan kestirim tekniklerinin karşılaştırılması

Metot	Veri gereksinimleri	Varsayımlar	Sonuçlar
Semptomatik kayıt teknikleri	-Doğumlar, ölümler ve göç	-Mantıklı demografik ilişkiler	-Yanlıdır -Doğum ve ölümlerin iyi ve güncel kayıtlarına bağımlıdır.
Sentetik kestirim	-Örnek verisi -Frekans tipli değişkenler	-Tabakaların homojenliği	-Az varyans -Yanlılık -Yerel alan faktörlerinin gerçek etkilerinin yansıtılamaması
Oran-korelasyon	-Doğumlar, ölümler, istihdam, seçmen kütükleri, okul kayıtları -Güncel bilgi -Oranlar olarak alınan değişkenler -Sürekli değişkenler	-Nüfus sayım zamanı ve sonrası için semptomatik göstergeler ve ilgilenilen değişkenler arasındaki sabit ilişki	- Semptomatik verilerin sayısı önemlidir - Daha az güvenilir sonuçlar -Tabakalama mümkündür. -Yapısal değişiklikler yüzünden hata -Değişkenler tarafından açıklanamayan rastgele hata
Fark-korelasyon	-Farklar şeklinde alınan değişkenler -Sürekli değişkenler	-Yukarıdakiyle aynı	-Daha doğru sonuçlar -Sıfır ortalamalar nedeni ile daha güvenilir sonuçlar
Örnek-regresyon	-Nüfus sayımı sonrası verisi -Sürekli değişkenler	-Uygun fonksiyonel ilişki	-Hata kareleri ortalamasının kestirimi için tabaka içi ve dışı varyanslarının elde edilmesi
Sentetik-regresyon	-Yüksek derecede ilişkili semptomatik değişkenler -Yerel alana ait semptomatik bilgi	-Ek semptomatik değişkenlerin varolması	-Değişkenler içinde örnekleme hataları -Değişkenler tarafından açıklanamayan rastgele hatalar -Yanlılığın azalması
Temel-birim metodu	-Temel birimler için araştırma örneği	-Gruplar arası homojenlik	-Özel bir fonksiyonel şekil yoktur. -Yerel alan kestirimine uygulanabilir. -Yanlıdır -Tabakalama yapmak hataya yol açabilir. -Temel birimler için veri bulmak zordur.

Çizelge 5.1. (Devam) Küçük alan kestirim tekniklerinin karşılaştırılması

James-Stein ve Bayes kestirimleri	-Sadece örneklemeden elde edilen veri	-H tane küçük alan için eşit varyanslı ve farklı ortalamalı örnek değişkenleri	-Yardımcı veriye ihtiyaç yoktur. -c'nin seçimi zordur. -Bayes özelliklerine sahiptir. -Eğer h>3 ise normal ortalamalar, en büyük olabilirlik tahmininden baskındır
Kestirim yaklaşımı	-Bazı doğrusal regresyon altındaki verilerin oluşturduğu yardımcı bilgi	-Doğrusal regresyon -Belirli varyans varsayımı	-Güvenirlik ölçütü olarak hata kareleri ortalaması kullanılır -Performans, örnekleme planına göre değişir
Yapı koruyan kestirim	-Yerel alan kestirimi için çapraz tablolanmış veri -Frekans verisi -Niceliksel ve niteliksel veriler	-Ortaklık yapısının sağlamlığı varsayımı -Dağıtım Yapısı	-Veri nüfus sayımından elde edilebilir. -Modele göre değişir -Yanlıdır
Tekrarlı oran uydurulması	-Frekans verisi -Niceliksel ve niteliksel veriler	-Ortaklık yapısı -Dağıtım yapısı	-Esnek, doğrusal olmayan bir modele tekrarlı teknikler ile örnek verisi uydurur
Varyans bileşenleri modelleri yolu ile kestirim	-Örnekleme hatası içermeyen veriler.	-Yardımcı veride anlamlı örnekleme hatası vardır ve modelde görünmez. -Regresyon varsayımları	-Örnekleme hatası yüzünden bağımsız değişkenler ortaya çıkarlar
Genel karışık doğrusal model	-Örnek verisi -Modellere göre değişen veri	-Regresyon varsayımları -Ortalama için normallik varsayımı	-Belirsizliğin hata kareleri ortalaması yaklaşımı ile ölçülmesi

Uygulama çalışmasının kapsamı olan Bolu İli'nin işsizlik oranı kestiriminde, üç ayrı teknik kullanılmıştır. Gerçek oranlar ve sentetik kestirim teknikleri ile yapılan kestirimlerde, Bolu İli'ne ait kadın ve erkek işsizlik oranları bulunmuştur. Tekrarlı oran uydurulması tekniği ile yapılan uygulamada ise, işsizlik oranları yaş gruplarına göre kestirilmiştir. İlk iki teknikle elde edilen oranlara bakılacak olursa, 1990 yılı Bolu İli erkek işsizlik oranının, gerçek oranlar tekniği ile kestirilen 0.022 değerinin, sentetik kestirim ile kestirilen[2] 0.021 değerine oldukça yakın olduğu görülmektedir. Aynı şekilde 1990 yılı Bolu İli kadın işsizlik oranı, gerçek oranlar tekniği ile 0.007 şeklinde, sentetik kestirim tekniği ile[3] de 0.007 şeklinde, aynı değer olarak kestirilmiştir. Buradan da her iki tekniği uygulayarak aynı sonuca varıldığı

[2] Bolu İli erkekler için ilçe merkezleri ve kırsal kesim toplamı = 0.009 + 0.012 = 0.021
[3] Bolu İli kadınlar için ilçe merkezleri ve kırsal kesim toplamı = 0.003 + 0.004 = 0.007

söylenebilir. Bulunan işsizlik oranlarına bakılacak olursa : Genel olarak Bolu İli'ne ait erkek nüfus arasındaki işsizlik oranının, kadın nüfus arasındaki işsizlik oranından yaklaşık üç kat fazla olduğu görülmektedir. Aynı şekilde, kırsal kesimdeki işsizlik oranının, hem erkek hem de kadın nüfus için, ilçe merkezlerindeki işsizlik oranından daha fazla olduğu açığa çıkmaktadır. Tekrarlı oran uydurulması tekniği ile elde edilen, yaşlara göre işsizlik oranlarına bakılacak olursa : İşsizlik oranı, 12-14 yaş grubunda 0.05 iken, bir sonraki yaş grubu olan 15-19 grubunda 0.068 ile en yüksek değerine ulaşmakta, bu değerden sonra gittikçe azalarak 29 yaş ve yukarısı grubunda 0.014 ile en küçük değerini almaktadır (Bkz. Çizelge 4.11). Bundan dolayı, Bolu İli'nde 19 yaşından sonra, yaş arttıkça, işsizlik oranının azaldığı sonucuna varılabilir.

KAYNAKLAR

1. PURCELL, N.J. and KISH, L. *Estimation for Small Domains*, Biometrics, 35, June, *365-384*, 1979.

2. SÄRNDAL, C.E., *Small Area Statistics and Survey Designs*, International Scientific Conference, Volume II. Contributed Papers and Panel Discussion. Central Statistical Office, Warsaw, 1993.

3. UNITED NATIONS., *The Preparation of Sampling Survey Reports*, Statistical Papers, Series C, No:1, 1950.

4. ROYALL, R.A., *Discussion of Papers by Gonzalez and Ericksen, Proceedings of the Social Statistics Section*, American Statistical Association, *42-43*, 1973.

5. GONZALEZ, M.E., *Use and Evaluation of Synthetic Estimates, Proceedings at the Social Statistics Section*, American Statistical Association, 87, *533-540*, 1973.

6. HOLT, D., SMITH, T.M.F., and TOMBERLIN, T.J., *A Model Based Approach to Estimation for Small Subgroups of a Population*, Journal of the American Statistical Association, 74, *405-410*, 1979.

7. GONZALEZ, M.E., *Case Studies on the Use and Accuracy of Synthetic Estimates: Unemployment and Housing Applications in Synthetic Estimates for Small Areas* (National Institute on Drug Abuse, Research Monograph 24), U.S. Goverment Printing Office, Washington D.C., 1979.

8. KALTON, G., *Panel Discussion in Small Area Statistics*, John Wiley and Sons, New York, 1987.

9. ERICKSEN, E.P., *A Regression Method for Estimation Population Changes for Local Areas*, Journal of the American Statistical Association, 69, *867-875*, 1974.

10. LAAKE, P., *A Prediction Approach to Subdomain Estimation in Finite Populations*, Journal of the American Statistical Association, 74, *355-358*, 1979.

11. DALENIUS, T., *Panel Discussion in Small Area Statistics*, John Wiley and Sons, New York, 1987.

12. SÄRNDAL, C.E., *Design-Consistent versus Model-Dependent Estimation for Small Domains*, Journal of the American Statistical Association, 79, 624-631, 1984.

KAYNAKLAR (devam)

13. NATIONAL INSTITUTE ON DRUG ABUSE (NIDA)., *Synthetic Estimates for Small Areas* (Research Monograph 24), U.S. Government Printing Office, Washington D.C., 1979.

14. NATIONAL CENTER for HEALTH STATISTICS (NCHS)., *Synthetic State Estimates of Disability* (PHS Publication No. 1759), U.S. Government Printing Office, Washington D.C., 1968.

15. U.S. BUREAU OF THE CENSUS., *Population Estimates and Projection, Current Population Reports*, Series *P-25*, No.580, U.S. Goverment Printing Office, Washington D.C., 1975a.

16. U.S BUREAU OF THE CENSUS., *Population-Special Reports, Series P-47*, No.4, U.S. Government Printing Office, Washington D.C., 1947.

17. BOGUE, D.J., *A Technique for Making Extensive Population Estimates*, Journal of the American Statistical Association, 45, *149-163*, 1950.

18. 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 Report, 47, No.6, 1959.

19. U.S. BUREAU OF THE CENSUS., *Estimates of the Population of Counties and Metropolitan Areas, July 1, 1966: A Summary Report, Current Population Reports*, Series P-25, No. 427, U.S. Goverment Printing Office, Washington D.C., 1969.

20. RIVES, N.W., *A Modified Housing Unit Method for Small Area population Estimation, Proceedings of the Social Statistics Section*, American Statistical Association, *717-720*, 1976.

21. U.S. BUREAU OF THE CENSUS., *The Methods and Materials of Demography*, Third Printing, U.S. Government Printing Office, Washington D.C., 1975b.

22. STARSINIC, D.E., *Development of Population Estimates for Revenue Sharing Areas, Census Tract Papers*, Series GE 40, No.10, U.S. Bureau of the Census, U.S. Government Printing Office, Washington D.C., 1974.

23. PURCELL, N.J., and LINACRE, S., *Techniques for Estimation of Small Area Characteristics*, Paper Presented at the 3rd Australian Statistical Conference, Melbourne, Australia,18-20 August,1976.

KAYNAKLAR (devam)

24. GHANGURDE, P.D., and SINGH, M.P., *Synthetic Estimation in Periodic Household Surveys*, Journal of Survey Methodology, Statistics Canada 3, *152-181*, 1977.

25. GONZALEZ, M.E., and WAKSBERG, J., *Estimation of the Error of Synthetic Estimates*, Paper Presented at the First Meeting of the International Association of Survey Statisticions, Vienna, Austria, 18-25 August, 1973.

26. SCHMITT, R.C., and CROSETTI, A.H., *Accuracy of the Ratio-Correlation Method of Estimating Postcensal Population*, Land Economics, 30, *279-280*, 1954.

27. 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.

28. NAMBOODIRI, N.K., *On the Ratio-Correlation and Related Methods of Subnational Population Estimation*, Demography, 9, *443-453*, 1972.

29. ROSENBERG, H., *Improving Current Population Estimates Through Stratification*, Land Economics, 44, *331-338*, 1968.

30. PURSELL, D.E., *Improving Population Estimates With the Use of Dummy Variables*, Demography, 7, 87-91, 1970.

31. MARTIN, J.H., and SEROW. W.J., *Estimating Demographic Characteristic Using the Ratio-Correlation Method*, Demography, 15, *223-233*, 1978.

32. COCHRAN, W.G., *Sampling Techniques*, John Wiley and Sons Inc., New York, *204-205*, 1963.

33. O'HARE, W., *Report on a Multiple Regression Method for Making Population Estimates*, Demography, 13, *369-379*, 1976.

34. HANSEN, M.H., HURWITZ, W.N., and MADOW, W.G., *Sample Survey Methods and Theory I*, John Wiley and Sons Inc., New York, 1953.

35. WOODRUFF, R.S., *Use of a Regression Technique to Produce Area Breakdowns of the Monthly National Estimates of Retail Trade*, Journal of the American Statistical Association, 61, *496-504*, 1966.

KAYNAKLAR (devam)

36. ERICKSEN, E.P., *A Method for Combining Sample Survey Data and Symptomatic Indicators to Obtain Estimates for Local Areas*, Ph. D. thesis, University of Michigan, Ann Arbor, Michigan, 1971.

37. 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.

38. NICHOLLS, A., *A Regression Approach to Small Area Estimation*, Australian Bureau of Statistics, Canberra, Australia, March (Mimeographed), 1977.

39. GONZALEZ, 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.

40. ERICKSEN, E.P., *Population Estimation in the 1970's: The Stakes are Higher*, Institute for Survey Research, Temple University, 1975.

41. KALSBEEK, W.D., *A Method for Obtaining Local Postcensal Estimates for Several Types of Variables*, Ph. D. thesis, University of Michigan, Ann Arbor, Michigan, 1973.

42. SCHAIBLE, W.L., *Use of Small Area Estimators in U.S. Federal Programs*, Small Area Statistics and Survey Designs, International Scientific Conference, Warsaw, 1993.

43. EFRON, B., and MORRIS, C., *Stein's Estimation Rule and Its Competitors on Emprical Bayes Approach*, Journal of the American Statistical Association, 68, *117-130*, 1973.

44. JAMES, W., and STEIN, C., *Estimation with Quadratic Loss in Proceedings of the Fourth Berkeley Symposium of Mathematical Statistics and Probability*, Vol. 1, Berkeley: University of California Press, *361-379*, 1961.

45. ROYALL, RR.M., *On Finite Population Sampling Theory Under Certain Linear Regression Models*, Biometrika, 57, *377-387*, 1970.

46. HORWITZ, D.G., and THOMPSON, D.J., *A Generalization of Sampling Without Replacement from a Finite Universe*, Journal of the American Statistical Association, 47, *663-685*, 1952.

47. PURCELL, N.J., and KISH, L., *Postcensal Estimates for Local Areas (or Domains)*, International Statistical Review, 48, *3-18*, 1980.

KAYNAKLAR (devam)

48. PURCELL, N.J., *Efficient Small Domain Estimation: A Categorical Data Analysis Approach*, Ph. D. thesis, University of Michigan, Ann Arbor, Michigan, 1979.

49. DEMING, W.E., and STEPHAN, F.F., *On a Least Squares Adjustment of A Sampled Frequency Table When the Expected Marginal Totals are Known*, Annals of Mathematical Statistics, 11, *427-444*, 1940.

50. FAY, R.E., *Small Domain Estimation Through Components of Variance Models*, 46th Session of the ISI, Invited Paper 29.1, *582-394*, Tokyo, 1987.

51. HARVILLE, D.A., *Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects*, Annal of Statistics, 4, *384-395*, 1976.

52. HENDERSON, C.R., *Best Linear Unbiased Estimation and Prediction Under a Selection Model*, Biometrics, 31, *423-447*, 1975.

53. BATTESE, G.E., HARTER, R.M., and FULLER, W.A., *An Error-Component, Model for Prediction of County Crop Areas Using Survey and Satellite Data*, Journal of the American Statistical Association, 83, *28-36*, 1988.

54. RAO, C.R., *Linear Statistical Inference and Its Applications*, John Wiley, New York, 1973.

55. DEMPSTER, A.P., RUBIN, D.B., and TSUTAKAWA, R.K., *Estimation in Covariance Component Models*, Journal of the American Statistical Association, 76, *341-353*, 1981.

56. FAY, R.E., and HERRIOT, R.A., *Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data*, Journal of the American Statistical Association, 74, *269-277*, 1979.

57. DİNÇER, B., ÖZASLAN, M., SATILMIŞ, E., *İllerin Sosyo Ekonomik Gelişmişlik Sıralaması Araştırması*, Devlet Planlama Teşkilatı Yayınları, Ankara, 1996.

58. HACETTEPE ÜNİVERSİTESİ NÜFUS ETÜTLERİ ENSTİTÜSÜ (HÜNEE)., *Türkiye Nüfus ve Sağlık Araştırması*, Ankara, 1993.

59. D.İ.E., *1985 Genel Nüfus Sayımı, Nüfusun Sosyal ve Ekonomik Nitelikleri*, Devlet İstatistik Enstitüsü Matbaası, Ankara, 1987.

KAYNAKLAR (devam)

60. D.İ.E., *1990 Genel Nüfus Sayımı, Nüfusun Sosyal ve Ekonomik Nitelikleri*, Devlet İstatistik Enstitüsü Matbaası, Ankara, 1993.

61. D.İ.E., *1990 Genel Nüfus Sayımı, Nüfusun Sosyal ve Ekonomik Nitelikleri, İllere Göre Sonuçlar, Bolu*, Devlet İstatistik Enstiüsü Matbaası, 1993.

EK-1

W_s ve W_s^* DEĞERLERİNİN BULUNUŞU

W_s 'nin hangi değerleri için

$$\sum\left[(y-\bar{y})-\sum_{s=1}^{m}W_s k_s(x_s-\bar{x}_s)\right]^2 \qquad (1\text{-}1)$$

ifadesinin

$$\sum_{s=1}^{m}W_s=1 \qquad (1\text{-}2)$$

koşulu altında minimum olacağı bulmak isteniyor. (1.1)'in minimizasyonu,

$$Q=\sum_{s=1}^{m}\sum_{s=1}^{m}W_sW_tA_{st} \qquad (1\text{-}3)$$

A_{st} : $v_{00}-k_sv_{0t}-k_tv_{0t}+k_sk_tv_{st}$

v_{00} : y'nin varyansı,

v_{0S} : y'nin ve x_S 'nin varyansı

v_{0t} : y'nin ve x_i 'nin varyansı,

v_{St} : x_s'nin ve x_i ' nin kovaryansı

olmak üzere, (1.3) ifadesi matris halinde, şu şekilde yazılabilir:

WAW'

W : W_s'lerin 1 x m vektörü

W' : W'nin transpozu

A : Yukarıda tanımlanan A_{st}'lerin m x m matrisi

Eğer aynı zamanda (11........1)'ler, yani e'lerin 1 x m vektörü de belirlenirse, (1.2) koşulu altında (1.1) ifadesini minimize eden W_s'leri bulma problemi, şu eşitliği minimum olacak şekilde W_s için çözmek olarak tanımlanabilir.

$$Q^* = WAW' + \lambda(eW' - 1) \tag{1-4}$$

λ : Lagrange çarpanı

Q^*'ı minimize eden W şu eşitlikle verilir :

$$W = \frac{eA^{-1}}{eA^{-1}e'} \tag{1-5}$$

(1.5)'teki s. birim şu ifadeyle verilir :

$$W_s = \frac{A^{-1}\text{' in s. sütunundaki elemanlarının toplamı}}{A^{-1}\text{'deki bütün elemanların toplamı}}$$

EK-2

REGRESYON KESTİRİMLERİNİN HATA KARELERİ ORTALAMALARININ BULUNUŞU

Regresyon kestirimlerinin hataları :

e= $Y - X\hat{B} = (Y_o - v) - X\hat{B}$ 'dir.

$\hat{B} = (X'X)^{-1}X'Y_o$

ve

$Y_0 = XB + u + v$

olduğundan

$$e = \left[I_n - X(X'X)^{-1}X'\right]u - X(X'X)^{-1}X'v$$

olur.

$\left[I_n - X(X'X)^{-1}X'\right]$

ve

$X(X'X)^{-1}X'$

idempotent olduğundan,

$$E(e'e) = E\left\{u'\left[I_n - X(X'X)^{-1}X'\right]u\right\} + E\left[v'X(X'X)^{-1}X'v\right] - 2E\left\{u'\left[X(X'X)^{-1}X' - X(X'X)^{-}\right.\right.$$

$$= (n-p-1)\sigma_u^2 + (p+1)\sigma_v^2$$

olur. Bu durumda hata kareleri ortalamaları şu şekilde yazılabilir :

$$MSE = E(e'e)/n = E(Y-\hat{Y})'(Y-\hat{Y})/n$$

$$= \left[(n-p-1)\sigma_u^2/n\right] + \left[(p+1)\sigma_v^2/n\right]$$

Regresyon ve örnek kestirimleri arasındaki fark karelerinin ortalaması da şu şekilde elde edilebilir :

$\hat{Y} = X\hat{B}$

ve

$Y_o = XB + u + v$

olduğu bilinmektedir.

$Y_o - \hat{Y} = \left[I_n - X(X'X)^{-1}X'\right]Y_o = \left[I_n - X(X'X)^{-1}X'\right](u+v)$ 'dir.Çünkü :

$\left[I_n - X(X'X)^{-1}X'\right]XB = XB - XB = 0$'dır.

$\left[I_n - X(X'X)^{-1}X'\right]$ idempotent olduğundan,

$(Y_o - \hat{Y})'(Y_o - \hat{Y}) = (u+v)'\left[I_n - X(X'X)^{-1}X'\right](u+v)$

ve

$E(Y_0 - \hat{Y})'(Y_o - \hat{Y}) = (n-p-1)E(u+v)'(u+v)$

olur. u ve v'nin korelasyonunun 0 olduğu varsayıldığından,

$$E(Y_0 - \hat{Y})'(Y_0 - \hat{Y})/n = \left[(n-p-1)/n\right](\sigma_u^2 + \sigma_v^2)$$

olacaktır.