

**ÇOKLU ÇERÇEVE YÖNTEMİ VE GİRİŞİMCİ BİLGİ SİSTEMİ
VERİLERİ ÜZERİNE BİR UYGULAMA**

Mehmet Oğuzhan ÜSTÜN

**YÜKSEK LİSANS TEZİ
İSTATİSTİK ANABİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

EKİM 2014

Mehmet Oğuzhan ÜSTÜN tarafından hazırlanan “ÇOKLU ÇERÇEVE YÖNTEMİ VE GİRİŞİMCİ BİLGİ SİSTEMİ VERİLERİ ÜZERİNE BİR UYGULAMA” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi İstatistik Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Yrd. Doç. Dr. Esen GÜRBÜZSEL

İstatistik Anabilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

.....

Başkan: Doç. Dr. Mehtap AKÇİL OK

İstatistik ve Bilgisayar Bilimleri Anabilim Dalı, Başkent Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

.....

Üye: Doç. Dr. Yaprak Arzu ÖZDEMİR

İstatistik Anabilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

.....

Tez Savunma Tarihi: 31/10/2014

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Şeref SAĞIROĞLU

Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Mehmet Oğuzhan ÜSTÜN

31/10/2014

ÇOKLU ÇERÇEVE YÖNTEMİ VE GİRİŞİMCİ BİLGİ SİSTEMİ VERİLERİ ÜZERİNE BİR UYGULAMA

(Yüksek Lisans Tezi)

Mehmet Oğuzhan ÜSTÜN

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Ekim 2014

ÖZET

Yapılan çalışmalarda ilgili karakteristikler açısından istenilen veriyi elde etmek için birden fazla liste kullanma zorunluluğu doğabilir. Bu durumda, bir listede yer alan örnekleme birimi, başka bir listede de yer alma ihtimali vardır. Böyle bir durum söz konusu olduğunda iki listeden çekilecek örneklerden elde edilecek tahminler, yüksek olasılıkla hatalı sonuçlar verebilir. Bu problemin üstesinden gelebilmek için çoklu çerçeve yöntemi kullanılmaktadır. Bu çalışmada, çoklu çerçeve yönteminin duyarlılığını ortaya çıkarmak amacıyla Girişimci Bilgi Sistemi verileri kullanılarak, iki farklı devlet kurumuna ait il verileri analiz edilmiştir. Bu analiz sonucu o ilde istihdam edilen toplam çalışan sayısı, yığın ve alan büyüklüklerinin bilindiği ve bilinmediği durumlarda, farklı tahmin edicilere göre tahmin edilmiştir. Yığın büyüklükleri bilindiği durumda Hartley'in tahmin edicisi, yığın büyüklüğü bilinip, alan büyüklüğü bilinmediği zaman Fuller ve Burmeister'in tahmin edicisi daha iyi sonuç vermiştir.

Bilim Kodu : 205.1.066
Anahtar Kelimeler : Çoklu Çerçeve, Girişimci Bilgi Sistemi, Çalışan Sayısı
Sayfa Adedi : 53
Tez Yöneticisi : Yrd. Doç. Dr. Esen GÜRBÜZSEL

MULTIPLE FRAME METHOD AND AN APPLICATION ABOUT ENTREPRENEUR
INFORMATION SYSTEM

(M. Sc. Thesis)

Mehmet Oğuzhan ÜSTÜN

GAZİ UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

October 2014

ABSTRACT

In the analysis, a necessity of using more than one list might emerge to acquire the needed data in terms of related characteristics. In this case, there might be a possibility that the sampling unit in one list can also exist in another list as well. Under this circumstances, estimations from samples got from both lists most likely to give misleading results. In order to overcome this problem, multi frame method is used. In this study, in order to reveal the sensitivity of multi frame method , provincia data of two different state institutions are analyzed by using the data of Entrepreneur Information System. As a result of this analysis, total number of employees working in that province is estimated according to different estimators in cases that either population and domain size are known or not. Hartley's estimator gave better results when population are known as Fuller and Burmeister's estimator gave better results when population are known but domain size are not.

Science Code : 205.1.066
Key Words : Multiple Frame, Entrepreneur Information System, Number of
Workers
Page Number : 53
Supervisor : Assistant Professor Esen GÜRBÜZSEL

TEŞEKKÜR

Çalışmalarım boyunca değerli yardım ve katkılarıyla beni yönlendiren hocam Yrd. Doç. Dr. Esen GÜRBÜZSEL'e manevi destekleriyle beni hiçbir zaman yalnız bırakmayan babam Güntekin ÜSTÜN, annem Ayşe ÜSTÜN, ablalarım Ayşe Feyza CEYLANLI ve Emine Banu PARLATAN, arkadaşlarım Pınar DAĞ ve Caner ÇİMEN'e, değerli iş arkadaşlarım Ayşe YAMAK ve Emre ODABAŞI'na teşekkürü bir borç bilirim.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	x
SİMGELER VE KISALTMALAR.....	xi
1. GİRİŞ.....	1
2. ÇOKLU ÇERÇEVE ARAŞTIRMASI.....	7
3. İKİLİ ÇERÇEVEDE TESADÜFİ ÖRNEKLEME.....	19
3.1. N_A , N_B ve N_{ab} 'nin Bilindiği Durumda Yığın Toplamı Tahmin Edicileri	22
3.1.1. Hartley tahmin edicisi	22
3.1.2. Lund tahmin Edicisi.....	24
3.2. N_A ve N_B 'nin Bilindiği, N_{ab} 'nin Bilinmediği Durumda Yığın Toplamı Tahmin Edicileri	25
3.2.1. Hartley tahmin edicisi	26
3.2.2. Lund tahmin edicisi.....	27
3.2.3. Fuller ve Burmeister tahmin edicisi	28
3.2.4. Tek çerçeve (Single frame) tahmin edicisi	30
3.2.5. Skinner'in sıralama oranı tahmin edicisi	30
4. GİRİŞİMCİ BİLGİ SİSTEMİ	33
5. İKİLİ ÇERÇEVE İÇİN UYGULAMA	37
5.1. N_A , N_B ve N_{ab} 'nin Bilindiği Durum İçin Toplam Çalışan Sayısı Tahminleri.....	38

Sayfa

5.2. N_A ve N_B bilindiđi, N_{ab} 'nin Bilinmediđi Durum İin Toplam alıřan Sayısı Tahminleri	41
6.SONU VE NERİLER.....	49
KAYNAKLAR	51
ZGEMİř	53

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Tekil tanımlama süreci	8
Çizelge 2.2. Örnek için tekil tanımlama süreci	8
Çizelge 2.3. A, B ve C çerçevesine sahip yığın	12
Çizelge 2.4. A çerçevesine ait yığından seçilen bağımsız basit tesadüfi örnek	13
Çizelge 2.5. B çerçevesine ait yığından seçilen bağımsız tesadüfi örnek.....	13
Çizelge 2.6. C çerçevesine ait yığından seçilene bağımsız basit tesadüfi örnek	14
Çizelge 2.7. s_A , s_B ve s_C örneklerindeki tekil birimler	14
Çizelge 2.8. s_A , s_B ve s_C örneklerindeki tekrar eden birimler.....	15
Çizelge 2.9. s_A , s_B ve s_C örneklerindeki tekil ve tekrar eden birimlerin sayısı.....	15
Çizelge 3.1. İki çerçeve düzeni ve tahminleri için notasyon gösterimi	19
Çizelge 3.2. İki alan arasında herhangi bir çerçeve için yığın birimlerinin bölünmesi .	20
Çizelge 3.3. İki alan arasında herhangi bir çerçeve için seçilen örnek birimlerinin bölünmesi	20
Çizelge 5.1. N_A , N_B ve N_{ab} bilindiğinde toplam çalışan sayısı, varyansı ve değişim ve değişim katsayısı	48
Çizelge 5.2. N_A ve N_B bilindiğinde, N_{ab} bilinmediğinde toplam çalışan sayısı, varyansı ve değişim katsayısı	48

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 1.1. Örtüşen A ve B Çerçeveleri ve İki Tanım Kümesi.....	3
Şekil 1.2. Örtüşen Üç Çerçeve ve Dört Tanım Kümesi.....	3
Şekil 1.3. Örtüşen A ve B Çerçeveleri ve Üç Tanım Kümesi.....	4
Şekil 1.4. Örtüşen A, B ve C Çerçeveleri ve Yedi Tanım Kümesi.....	4

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamalarıyla birlikte aşağıda sunulmuştur;

Simgeler	Açıklama
C	Maliyet
c_i	i. örneğin seçim maliyeti
f_h	h. tabaka örneklem oranı
M	Tekil birimlerin toplam sayısı
M_i	i. çerçevedeki tekil birimlerin sayısı
M_{ij}	Hem i. hem de j. çerçevelerde yer alan birim sayısı
$M_{ij.k}$	i. ve j. çerçevelerinde yer alıp, k. çerçevede yer almayan birimlerin sayısı
$M_{ij.kl}$	i. ve j. çerçevelerde yer alıp, k. ve l. çerçevede yer almayan birimlerin sayısı
m_i	Örnek için i. tekil birimlerin sayısı
m_{ij}	Örnek için i. ve j. örnekte yer alan birimlerin sayısı
$m_{ij.k}$	Hem i. hem de j. örnekte yer alıp, k. örnekte yer almayan birimlerin sayısı
$m_{ij.kl}$	Hem i. hem de j. örnekte yer alıp, k. ve l. örnekte yer almayan birimlerin sayısı
$\bar{m}_i$	i. tekil örneklem oranı
N	Yığın çapı
N_i	i. yığın çapı
n_i	i. örnek çapı
n'_{ij}	Hem i. hem j. örnekte yer alıp, i. örnekten çekilen birimlerin sayısı
n''_{ij}	Hem i. hem j. örnekte yer alıp, j. örnekten çekilen birimlerin sayısı
s_h	h. alt sınıfın varyansı

Simgeler	Açıklama
W_h	h. sınıfın ağırlığı
x_{hi}	h. alt sınıfın i. birimi
Y_i	i. yığın toplamı
y_i	i. örnek toplamı
$\bar{Y}_i$	i. yığın ortalaması
$\bar{y}_i$	i. örnek ortalaması
x_{hi}	h. alt sınıfın i. birimi
$\bar{y}_w$	w. alt sınıf örneklem ortalaması
$\bar{Y}$	Yığın ortalaması
$\hat{Y}$	Yığın Toplamının Tahmini
Y	Yığın Toplamı
$\hat{Y}_c^{(F)}$	c. Yığın toplamı
σ^2	Yığın varyansı
σ_i^2	i. yığın varyansı
α	Tekrar oranı
β	Tekrar oranı
λ	Düzeltilme terimi
δ_i	i. çerçeveye ait olup olmama durumu

Kısaltmalar	Açıklama
DK	Değişim Katsayısı
FB	Fuller-Burmeister
GBS	Girişimci Bilgi Sistemi
H	Hartley
KOSGEB	Küçük ve Orta Ölçekli İşletmeleri Geliştirme ve Destekleme İdaresi Başkanlığı
L	Lund
NACE Rev.2	Avrupa Topluluğunda Ekonomik Faaliyetlerin İstatistikî Sınıflaması
PMLE	Psuedo- En Çok Olabilirlik Tahmin Edicisi

Kısaltmalar**Açıklama****RR**

Skinner'in Sıralama Oranı

SF

Tek Çerçeve

SGK

Sosyal Güvenlik Kurumu

TÜBİTAK

Türkiye Bilimsel ve Teknolojik Araştırma Kurumu

TÜİK

Türkiye İstatistik Kurumu

Var

Varyans

1. GİRİŞ

Yığın birimlerinin oluşturduğu liste örnekleme çerçevesi olarak adlandırılmaktadır. Örnekleme yapmak için, yığın birimlerinin listesi mevcut veya hazırlanabilir şekilde olmalıdır. Örnekleme çerçevesi ya da liste, örnekleme sürecinin etrafına inşa edilen temel taşlardan biridir. Çerçeve olmaksızın yapılan araştırmalar yanlış tahminlere neden olurlar. Örnek vermek gerekirse çerçeve; telefon rehberi, seçmen listesi, hastane kayıtları vb. olabilir. Çerçevenin güncellemesi zor ve pahalı bir süreç gerektirmektedir. Birçok güncelleme yöntemleri var olmasına rağmen, uygulamaya geçildiğinde kimi zaman güçlükler ortaya çıkmaktadır. Bu güçlükler güvenilirlik kaybı, maliyet ve zaman kısıtları olarak karşımıza çıkmaktadır.

Çerçeve eğer hedef yığının tüm birimlerini kapsarsa mükemmeldir. Diğer bir deyişle ilgilenilen yığına ait örnekleme çerçevesi ve seçilmiş örnekleme birimleri birebir eşleşmeye sahip olmalıdır. Ancak pratikte böyle mükemmel çerçevelere ulaşmak mümkün olmadığı için yığının temsil etme gücü yüksek olan çerçeveler kullanılmaktadır (Dalçık, 2010).

Kish (1965) tarafından belirlenmiş dört temel çerçeve sorunu vardır;

1. Kayıp birimler¹
2. Birden fazla birimin tek birim gibi görünmesi ²
3. Yabancı birimler veya boşluğun olması³
4. Mükerrer listeleme⁴

Kish (1965), bu sorunları azaltmak için üç genel yol belirlemiştir;

- a) Problemi düzeltmek çok maliyetli olacaksa o problem göz ardı edilebilir.
- b) Yığın yeniden tanımlanarak çerçeve tekrardan oluşturulabilir.

¹ Kayıp birimler boş bırakılan, cevap alınamayan değerlerdir.

²Çeşitli hatalardan dolayı farklı birimlerin, aynı şekilde görünmesi durumudur. Bir işletmeye ait vergi kimlik numarasının diğer işletme sahiplerinin herhangi bir kodlama hatasından dolayı, çerçevede birden fazla görünmesi bu duruma örnek olarak verilebilir.

³İlgili birimin o çerçeveye ait olmaması gerektiği halde görünmesi durumudur.

⁴Bir birimin, çerçevede birden fazla görünmesi durumudur.

c) Yığına ait liste tamamen düzeltilebilir.

Kish (1965), bu dört temel çerçeve probleminin her biri için farklı çözüm yolları önermektedir;

1. Kayıp birimlerin olması durumunda,

- Kayıp birimler ayrı bir tabaka olarak alınıp farklı çözüm yolları yaratılabilir.
- Birleştirme yöntemi yapılarak, bu listede yer almayan birim, örneğe seçilen bir birime bağlı olarak tanımlanabildiği durumda, örneğe seçilmiş gibi alınıp örneğe seçilen birimin olasılığı kullanılabilir.

2. Birim kümelerinin tek listeye bağlı olarak birlikte görünmesi durumunda,

- Tüm birimler kapsama dahil edilerek yeni liste oluşturulabilir.
- Birimlerden biri rasgele olarak seçilir ve ona göre ağırlıklandırma yapılabilir.
- Daha büyük bir örnekleme listesi yapıldıktan sonra eşit olasılıklı seçme yöntemiyle bu birimlerden seçim yapılabilir.

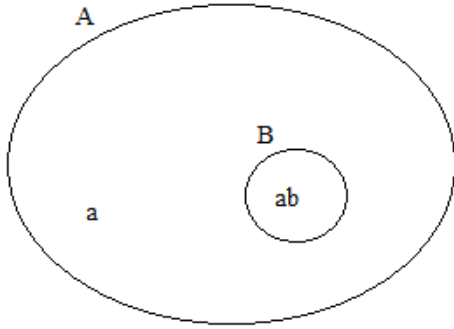
3. Yabancı birimlerin veya boşluğun olması durumunda, örnek seçimi yapıldıktan sonra, bu birimler örnek dışına atılabilir.

4. Mükerrer listeleme durumunda,

- Sorun ağırlıklandırma aşamasında çözülebilir.
- Liste kontrolü yapıp tekrar eden birimler çıkartılabilir.
- Her bir birim için tek listenin mükemmel bir şekilde oluşturulması örnek seçimi sırasında belirlenebilir.

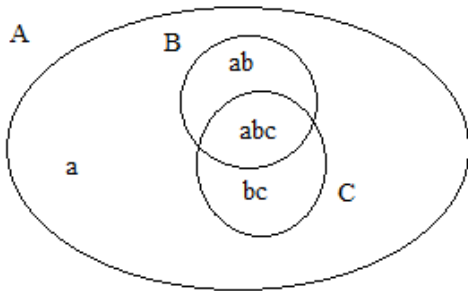
Çoklu çerçeve araştırmaları, genellikle örneklemenin maliyetini düşürmek için kullanılır. Örneğin; hanehalkı adreslerinin listesi gibi bir çerçeve mevcut olduğu zaman, örnek çekimi daha az maliyetli olan telefon rehberi gibi diğer çerçeveden alınan verilerle listeyi desteklemek daha verimli olur.

Şekil 1.1’de örnek çekimi daha maliyetli olan, tüm yığını gösteren A çerçevesi ile örnek çekimi az maliyetli olan ancak tüm yığını kapsamayan B çerçevesi gösterilmektedir. Şekil 1.1’ de görüldüğü üzere B çerçevesinin tüm birimleri A çerçevesinde de görülmektedir. Bu iki çerçevenin örtüştüğü anlamına gelmektedir ve A çerçevesinin bu yığında sahip olmadığı birim yoktur.



Şekil 1.1. Örtüşen A ve B Çerçeveleri ve İki Tanım Kümesi (Kaynak: Dalçık (2010))

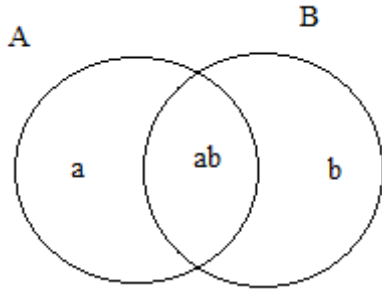
Şekil 1.2, A çerçevesinin eksiksiz yani yığın birimlerinin hepsini kapsadığı ancak B ve C çerçevelerinin eksik olduğu, örtüşen üç çerçeveyi göstermektedir. Hartley (1974) çalışmasında A çerçevesi olarak Teksas bölgesindeki çiftlik çalışmaları listesi ve B çerçevesi olarak da arazi çiftçileri listesindeki çiftçileri almıştır. Bu çalışmada B çerçevesi %100 kapsamaya sahiptir yani ancak A çerçevesi %70-%90’ını kapsamaktadır. Yani A çerçevesi ile B çerçevesi %70-%90 oranında birbirleriyle örtüşmektedir.



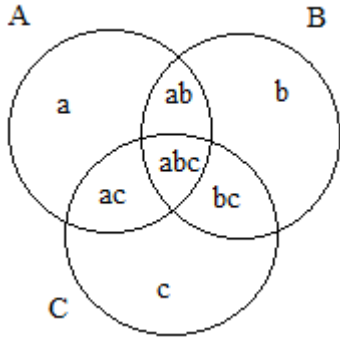
Şekil 1.2. Örtüşen Üç Çerçeve ve Dört Tanım Kümesi (Kaynak: Dalçık (2010))

Kapsamama, örnekleme çerçevelerinin en yaygın sorunlarından biridir. Araştırma tahminlerindeki kapsama hatasının etkilerini azaltmak için, çeşitli çerçeveler birleştirilebilir. Çoklu çerçeve tahmin edicileri, çoklu çerçeve araştırmalarındaki durumlar için kullanılmak üzere geliştirilmişlerdir. Bazı durumlarda örnekleme çerçeveleri

örtüşebilir. Şekil 1.3 her bir çerçevenin eksik olduğu örtüşen iki çerçeveyi ve Şekil 1.4, tüm çerçevelerin eksik olduğu örtüşen 3 çerçeveyi göstermektedir. Kalton ve Anderson (1986) çalışmalarında uyuşturucu tedavi merkezlerinden ve bağımsız hastanelerden seçtikleri örnekler doğrultusunda AIDS’li hastalar üzerine bir çoklu çerçeve araştırması yapmışlardır. Iachan ve Dennis (1993), çerçeve olarak evsiz barınaklarını, aşevlerini ve sokaklardaki belirli bölgeler olarak belirlemiş ve sokakta yaşayan insanlar üzerine bir çoklu çerçeve araştırması yapmışlardır. Buna göre evsiz insanların tamamının, evsiz barınaklarını ziyaret etmedikleri bilindiğinden daha fazla çerçeve kullanmak yanlılığı azaltmaktadır.



Şekil 1.3. Örtüşen A ve B Çerçeveleri ve Üç Tanım Kümesi (Kaynak: Dalçık (2010))



Şekil 1.4. Örtüşen A, B ve C Çerçeveleri ve Yedi Tanım Kümesi (Kaynak: Dalçık (2010))

Bu çalışmada tek bir çerçeveye bağlı kalınmaksızın birden fazla çerçeve kullanıldığında tahmin yapılması tartışılacaktır. Çünkü bazı durumlarda listelerdeki birimlerin bir kısmı ya da tamamı aynı olabilir ve listelerin birleştirilmesi mümkün olmayabilir. Örneğin, Ankara’da ikamet eden internet bankacılığı kullanıcılarına ilişkin bir araştırma yapıldığı zaman, çerçeveyi oluşturmak için her bir bankadan alınan internet kullanıcılarına ilişkin listelerin birleştirilmesi mümkün olmayabilir. Birden fazla listede aynı ad ve soyada sahip bireylerin yer alması mümkündür. Bu kişilere ait adres bilgilerine bakılarak bu durum

tespit edilebilir ancak bu durumun tespit edilmesi oldukça zordur. Böyle durumlarda çoklu çerçeve kullanımı iyi bir yaklaşım olabilir.

Çoklu Çerçeve yöntemi tam ve güncel bir çerçeve olmadığı durumlarda kullanılan en etkili yöntemlerden biridir. 1950 ve sonrasında bu konu oldukça fazla çalışılmıştır. Hartley (1962 ve 1974) çoklu çerçeve kullanımının tek çerçeve üzerinden tahmin üretme konusunda daha az maliyetli olduğunu göstermiştir. Fuller ve Burmeister (1972) ise Hartley'e göre daha küçük varyansa sahip tahmin edici geliştirmiştir. Bankier (1986) yaptığı çalışmada çoklu çerçevede tahmin üretme sürecinde Horvitz-Thompson tahmin edicisini kullanmıştır. Skinner (1991) çoklu çerçeve kullanımında Sıralı Oran Tahmin Edicisi (Raking Ratio Estimator) ile tahminler üretmiştir. Rao ve Skinner (1996) çalışmasında çerçeve büyüklüklerinin bilinme durumuna göre tahminlerin farklılık gösterebileceğini ifade etmiştir. Aynı çalışmada Pseudo-En Çok Olabilirlik Tahmin Edicisini (PMLE) kullanmış ve ikili çerçevede karmaşık örnekleme durumlarını uygulamıştır. Lohr ve Rao (2006) da PMLE'yi genelleştirerek ikiden daha fazla tahmin edici üzerinde kullanılabileceğini göstermiştir.

Çalışmanın birinci bölümünde temel çerçeve sorunlarına değinilerek bu sorunların hangi yollarla azaltılabileceğine yer verilmiştir. Ayrıca örtüşen çerçeve tanımı yapıp bununla ilgili örnekler verilmiştir. İkinci bölümde çoklu çerçeve araştırmalarına değinilerek, hem yığın için hem de örnek için tekil tanımlama süreci açıklanmıştır. Üç yığından elde edilen örneklerle küçük bir ön çalışma yapılmıştır. Üçüncü bölümde ikili çerçeveye ait çeşitli formülasyonlar, yığın ve alan büyüklüklerinin bilinip bilinmediği durumlara göre verilmiştir. Dördüncü bölümde Girişimi Bilgi Sistemi tanımlanarak hangi veri çeşitlerinin olduğu gösterilmiştir. Beşinci bölümde Girişimci Bilgi Sistemi'nden çekilen verilerle, ilgili ile ait toplam çalışan sayısı tahmin edilmiştir. Farklı tahmin edicilere ait değişim katsayıları hesaplanarak, hangi tahmin edicinin kullanılmasının daha doğru olduğu üzerinde durulmuştur. Altıncı bölümde ise sonuç ve öneriler kısmına yer verilmiştir.

2. ÇOKLU ÇERÇEVE ARAŞTIRMALARI

Çoklu çerçeve araştırmaları, çerçeveler birleştirilmeden parametre tahminleri elde etmek için örnekleri birleştiren çalışmaların bütünü olarak ifade edilebilir (Hartley, 1974).

“Çoklu Çerçeve Araştırmaları”nın kullanıldığı durumlar şu şekilde açıklanabilir; Örneklemeyle dayalı araştırma yöntemlerinde, yığının tüm birimlerini kapsayan bir çerçevenin kullanılması çok pahalı olabilir. Bunun yerine daha az maliyetli örnekleme yöntemleri kullanılarak çerçeveler oluşturulabilir. Ancak bu çerçeveler, genel olarak sadece bilinmeyen veya yığının sadece bir bölümünü kapsadığı yaklaşık olarak bilinen çerçevelerdir. Çoklu çerçeve kullanımında amaç, birkaç çerçevenin örtüşmesinin ölçüsü hakkında bir ön bilgiye gereksinim duymaksızın bu örtüşmenin kullanılabilmesini mümkün kılan genel bir yöntemi açıklamaktır (Hartley, 1962).

Buradaki varsayım ise, yığın birimlerinin tekrarı nedeniyle çerçeveler arasında oluşan yığın birimleri arasındaki kesişimin, rasgele örnek sonucu elde edilen örnek birimleri arasındaki kesişimle temsil edilebileceğidir (Şimşek, 2002).

Çoklu çerçeve kullanıldığı zaman, hedef yığının bazı birimleri birden fazla çerçeveye dahil olabilir. Bu durum “eşleşmiş çerçeve” ya da “örtüşen çerçeve” olarak literatürde yer almaktadır (Kish, 1965).

Bazı yığın birimlerinin birden çok çerçevede yer alması sorunundan dolayı, Kalton ve Anderson (1986) iki temel yaklaşım önermektedir;

- Örtüşmenin ortadan kaldırılması için çerçeveyi tekrar tanımlamak.
- Örtüşmeyi analiz esnasında telafi etmek.

Hedef yığın, tekrarları içermeyecek şekilde yeniden tanımlanırsa veya tekrarın ihmal edilecek kadar küçük bir oranda olduğu biliniyorsa bu tekrarlar göz ardı edilebilir. Eğer maliyetler çok yüksek değilse bu tekrarlar listelerden çıkartılabilir. Kish (1965) tekrarların ortadan kaldırılmasıyla ilgili bir süreç önermektedir. Bu süreç aşağıdaki Çizelge 2.1’de dört çerçevede gösterilmiştir.

Çizelge 2.1. Tekil tanımlama süreci

Çerçeve	1	2	3	4
Tekil Birimlerin Sayısı	M_1	M_2	M_3	M_4
Tekrar Eden Birimlerin Sayısı	0	M_{12}	$M_{23}+M_{13,2}$	$M_{34}+M_{24,3}+M_{14,23}$
Yığın Birimlerinin Sayısı	N_1	N_2	N_3	N_4

Birinci çerçeveden, M_1 tekil birimler olmak üzere N_1 listesinin tüm birimleri kabul edilir. N_2 listeli ikinci çerçeveden, ilk çerçevede bulunan M_{12} listesi çıkartılır ve geriye kalan ikinci çerçevedeki tekil M_2 birimleri yer alır. N_3 listeli üçüncü çerçeveden, ikinci çerçevede bulunan M_{23} listesi ve ikinci çerçevede bulunmamasına rağmen birinci çerçevede bulunan $M_{13,2}$ listesi çıkartılır ve geriye M_3 tekil birimleri kalır. N_4 listeli dördüncü çerçeveden ise üçüncü çerçevede bulunan M_{34} listesi, ikinci çerçevede bulunan ancak üçüncü çerçevede bulunmayan $M_{24,3}$ listesi ve birinci çerçevede bulunan, ikinci ve üçüncü çerçevede bulunmayan $M_{14,23}$ listesi çıkartılır ve geriye M_4 tekil birimleri kalır.

Tekil tanımlama süreci aynı zamanda bir örnek içinde uygulanabilir. Seçilen dört örnek için süreç aşağıdaki gibidir.

Çizelge 2.2. Örnek için tekil tanımlama süreci

Örnek	1	2	3	4
Tekil Birimlerin Sayısı	m_1	m_2	m_3	m_4
Tekrar Eden Birimlerin Sayısı	0	m_{12}	$m_{23}+m_{13,2}$	$m_{34}+m_{24,3}+m_{14,23}$
Örnek Birimlerinin Sayısı	n_1	n_2	n_3	n_4

$$\sum m_h = m_1 + m_2 + m_3 + \dots \quad (2.1)$$

tekil listeler olmak üzere, ilk $\bar{m}_1 = m_1 / n_1$ tabakası hariç alt sınıfların tabakalı örnekleme şekline gösterilebilir.

Tabakalar içinde n_h elemanlı basit tesadüfî örnekler, m_h elemanlı sınıf üyelerinin basit tesadüfî örnekleri ise sonuçlanır. Böylece $W_h^* = M_h / M$ ağırlıklarının sabit olduğu bilindiğinde varyans için kullanılan genel tabakalı tesadüfî denklemi ortalama içinde kullanılabilir (Kish, 1965).

$$\bar{y}_w = \sum_h^{M_h} \left(\frac{M_h}{M} \right) \sum_i^{m_h} y_{hi} \quad (2.2)$$

ve varyansı da

$$\text{var}(\bar{y}_w) = \sum W_h^{*2} (1 - f_h) (s_h^2 / m_h) \quad (2.3)$$

şeklinde gösterilir.

Genellikle M_h değeri bilinmemektedir. Fakat h . tabakadan seçilen n_h elemanlı örneklerin içerisinde yer alan, alt sınıf üyelerinin eleman sayısı olan, tesadüfî değişken m_h bilinmektedir. Alt sınıf üyelerinin oranı $\bar{m}_h = m_h / n_h$, M_{hi} değişkeninin binomial ortalaması;

$$\begin{aligned} m_{hi} &= n_{hi} = 1, & \text{alt sınıfın } m_h \text{ üyesi için} \\ m_{hi} &= 0, & \text{üye olmayan } (n_h - m_h) \text{ için} \end{aligned}$$

şeklindedir.

Alt sınıflar için hesaplamalar ve çıkarsamalar X_{hi} değişkeni için benzer bir Y_{hi} değişkeni içermektedir.

$y_{hi} = x_{hi}$, alt sınıfın m_h üyesi için

$y_{hi} = 0$, üye olmayan ($n_h - m_h$) için

şeklindedir.

Tabaka toplamı,

$$y_h = \sum_i^{n_h} y_{hi} = \sum_i^{m_h} y_{hi} \quad (2.4)$$

şeklindedir.

Tabaka ortalaması,

$$\bar{y}_h = \frac{y_h}{\bar{m}_h n_h} = \frac{y_h}{m_h} = \frac{\sum_i^{m_h} y_{hi}}{\sum_i^{m_h} m_{hi}} \quad (2.5)$$

şeklindedir.

Yığındaki M alt sınıf üyeleri için alt sınıf toplamı $Y = \sum_h^H \sum_i^{M_h} Y_{hi}$, alt sınıf ortalaması ise

$\bar{Y} = Y / M$ şeklinde tahmin edilebilir. Ağırlıklandırma yöntemi ile alt sınıf toplamı,

$$Y_w = \sum y_h \frac{N_h}{n_h} = \sum F_h y_h \quad (2.6)$$

alt sınıf örnek ortalaması ise,

$$\bar{y}_w = \frac{\sum y_h \frac{N_h}{n_h}}{\sum m_h \frac{N_h}{n_h}} = \frac{\sum F_h y_h}{\sum F_h m_h} = \sum w_h \bar{y}_h \quad (2.7)$$

$$w_h = \frac{F_h m_h}{\sum F_h m_h} \text{ ve } F_h = N_h / n_h$$

şeklinde tahmin edilebilir.

Kish (1965) tarafından önerilen alt sınıf toplamı Y_w için varyans,

$$\begin{aligned} \text{var}(Y_w) &= \sum (1 - f_h) F_h^2 \frac{n_h}{n_h - 1} \left[\sum_i^{m_h} y_{hi}^2 - \frac{y_h^2}{n_h} \right] \\ \text{var}(Y_w) &= \sum (1 - f_h) \frac{n_h^2}{n_h - 1} \bar{m}_h \left[v_h^2 + (1 - m_h) \bar{y}_h^2 \right] \end{aligned} \quad (2.8)$$

$$v_h^2 = \frac{1}{m_h} \sum_i^{m_h} (y_{hi} - \bar{y}_h)^2 \quad (2.9)$$

şeklinde hesaplanmaktadır. Kish (1965) tarafından gösterilen çoklu çerçeve araştırmasına ait uygulama Çizelge 2.3'te çerçeve sayısının üç olması durumunda incelenmiştir. Her bir çerçevedeki sanal numaralar tek bir firmaya aittir ve o firmalara ait çalışan sayısı bilgileri de Çizelge 2.3'te verilmiştir. Çerçeveler arası kesişim söz konusudur.

Çizelge 2.3. A, B ve C çerçevesine sahip yığın

A Çerçevesi			B Çerçevesi			C Çerçevesi		
Birim No	Sanal Numara	Çalışan Sayısı	Birim No	Sanal Numara	Çalışan Sayısı	Birim No	Sanal Numara	Çalışan Sayısı
1	992499970	45	1	992496835	24	1	992498980	80
2	992499880	65	2	992496865	33	2	992499355	22
3	992499835	50	3	992499580	105	3	992496730	9
4	992499790	65	4	992499835	50	4	992496535	8
5	992499775	17	5	992499970	45	5	992496370	18
6	992499670	32	6	992498335	45	6	992496205	74
7	992499625	102	7	992497900	27	7	992499880	65
8	992499610	31	8	992498650	38	8	992499610	31
9	992499580	105	9	992499775	17	9	992499205	55
10	992499355	22	10	992496730	9	10	992499055	10
11	992499310	30	11	992496535	8	11	992497105	51
12	992499205	55	12	992496370	18	12	992462425	22
13	992499130	20	13	992496205	74	13	992462365	43
14	992499085	91	14	992498980	80	14	992462275	56
15	992499055	10	15	992499355	22	15	992461930	66
16	992498980	80	16	992496220	15	16	992461885	40
17	992498890	19	17	992463445	24	17	992461840	50
18	992498845	37	18	992463430	40	18	992461780	24
19	992498650	38	19	992463355	31	19	992461765	49
20	992498485	15	20	992463280	23	20	992461630	62
21	992498335	45	21	992463085	5	21	992461615	4
22	992498215	22	22	992462965	40			
23	992498185	30	23	992462800	15			
24	992498035	17	24	992462455	32			
25	992497960	38						
26	992497900	27						
27	992497720	23						
28	992497480	98						
29	992497195	17						
30	992497105	51						

Çizelge 2.3’de verilen yığına ait yapılan çoklu çerçeve araştırmasına ilişkin uygulamada, Çizelge 2.4, Çizelge 2.5 ve Çizelge 2.6’da verilen örnekler kullanılmıştır. A, B ve C çerçevelerinden sırasıyla, birbirinden bağımsız basit tesadüfi (random) örnekler seçilmiştir. Bu örneklerin büyüklükleri $n_A=10$, $n_B=8$ ve $n_C=7$ ’ dir.

Çizelge 2.4. A çerçevesine ait yığından seçilen bağımsız basit tesadüfi örnek

Örnek A			
Örn. No. (j)	Birim No. (i)	Sanal Numara	Çalışan Sayısı (y _i)
1	16	992498980	80
2	10	992499355	22
3	17	992499775	17
4	9	992499580	105
5	3	992499835	50
6	1	992499970	45
7	21	992498335	45
8	26	992497900	27
9	22	992498215	22
10	12	992499205	55

Çizelge 2.5. B çerçevesine ait yığından seçilen bağımsız tesadüfi örnek

Örnek B			
Örn. No. (j)	Birim No. (i)	Sanal Numara	Çalışan Sayısı (y _i)
1	14	992498980	80
2	15	992499355	22
3	1	992496835	24
4	2	992496865	33
5	16	992496220	15
6	22	992462965	40
7	11	992496535	8
8	13	992496205	74

Çizelge 2.6. C çerçevesine ait yığından seçilen bağımsız basit tesadüfi örnek

Örnek C			
Örn. No. (j)	Birim No. (i)	Sanal Numara	Çalışan Sayısı (yi)
1	12	992462425	22
2	17	992461840	50
3	20	992461630	62
4	9	992499205	55
5	4	992496535	8
6	6	992496205	74
7	14	992462275	56

Örnek temelinde tekil belirleme süreci yapılmıştır. Tekil birimler ve tekrarlar sırasıyla Çizelge 2.7 ve Çizelge 2.8’ de gösterilmiştir.

Çizelge 2.7. s_A , s_B ve s_C örneklerindeki tekil birimler

A örneğindeki tekil birimler				B örneğindeki tekil birimler			
Örn. No. (j)	Birim No. (i)	Sanal Numara	Çalışan Sayısı (yi)	Örn. No. (j)	Birim No. (i)	Sanal Numara	Çalışan Sayısı (yi)
1	16	992498980	80	3	1	992496835	24
2	10	992499355	22	4	2	992496865	33
3	17	992499775	17	5	16	992496220	15
4	9	992499580	105	6	22	992462965	40
5	3	992499835	50	7	11	992496535	8
6	1	992499970	45	8	13	992496205	74
7	21	992498335	45	C örneğindeki tekil birimler			
8	26	992497900	27	Örn. No. (j)	Birim No. (i)	Sanal Numara	Çalışan Sayısı (yi)
9	22	992498215	22	1	12	992462425	22
10	12	992499205	55	2	17	992461840	50
				3	20	992461630	62
				7	14	992462275	56

Çizelge 2.8. s_A , s_B ve s_C örneklerindeki tekrar eden birimler

Hem A hem de B örneklerinde bulunan birimler			
Örn. No. (j)	Yığın No. (i)	Sanal Numara	Çalışan Sayısı (y _i)
1	14	992498980	80
2	15	992499355	22
Hem B hem de C örneklerinde bulunan birimler			
Örn. No. (j)	Yığın No. (i)	Sanal Numara	Çalışan Sayısı (y _i)
5	4	992496535	8
6	6	992496205	74
Hem C hem de A örneklerinde bulunan ancak B örneğinde bulunmayan birimler			
Örn. No. (j)	Yığın No. (i)	Sanal Numara	Çalışan Sayısı (y _i)
4	9	992499205	55

Örnek listelerinden elde edilen tekil birimler ve tekrarlar Çizelge 2.9’ da gösterilmiştir.

Çizelge 2.9. s_A , s_B ve s_C örneklerindeki tekil ve tekrar eden birimlerin sayısı

Örneklemler	A	B	C
Tekil Birimlerin Sayısı	10	6	4
Tekrar Eden Birimlerin Sayısı	0	2	2+1=3
Örneklemdaki Birimlerin Sayısı	10	8	7
$\bar{m}_h$	$\bar{m}_A=10/10=1$	$\bar{m}_B=6/8=0,75$	$\bar{m}_C= 4/7=0,571$
y_h	468	194	190

$$\hat{Y}_w = \sum y_h \frac{N_h}{n_h} = \sum F_h y_h = \frac{30}{10} 468 + \frac{24}{8} 194 + \frac{21}{3} 190 = 2556$$

$n_A=10$, $n_B=8$ ve $n_C=7$ olmak üzere toplam çalışan sayısının tahmini 2556 çıkmıştır.

Tek bir liste oluştururken, farklı çerçevelerdeki listelerin sahip olduğu birimlerin eşleştirilmesi gerekmektedir. Bu eşleştirme süreci uzun ve hataya açık bir süreçtir. Harf veya sayı hatasından dolayı yanlış yazılan isimler, adresler, numaralar vb. eşleşmenin hatalı olmasına neden olmaktadır. Bu problemlerden dolayı ortaya çıkan sorunlar dikkate

alınmalıdır. Yapılan çoklu çerçeve araştırmalarında bu durum ayrıntılı bir şekilde incelenmiştir.

Örnek araştırmalarda geçen çoklu çerçeveler Hartley (1962 ve 1974), Fuller ve Burmeister (1972), Bankier (1986), Skinner (1991), Skinner ve Rao (1996) ve Lohr ve Rao (2000) tarafından incelenmiştir. Alan ve çerçeve büyüklüğünün bilindiği veya bilinmediği durumlar için farklı tahmin ediciler önerilmiştir. Genellikle alan büyüklüğü bilinmemektedir. Fakat güncelliğini yitirmiş listelerin kullanımında alan büyüklüğünün yaklaşık olarak bilindiği durumlarda, ilgilenilen birimin hangi çerçeveye ait olduğunu belirlemek çok zorlu bir süreçtir.

Eşleşen yani örtüşen çerçeveler tanımı ilk kez Hartley (1962 ve 1974) tarafından ele alınmıştır. Genellikle çoklu çerçeve araştırmaları ikili çerçeve durumu için ele alınmıştır. Çoklu çerçeve araştırmalarında Lohr ve Rao (2006) tarafından oluşturulan bilgilere de ilerideki bölümlerde değinilecektir.

İlgilenilen yığına ait çerçeveler $F = \{1, 2, \dots, P\}$ şeklinde tanımlanmaktadır. Çerçevelerinin birleşimi hedef yığını yüzde yüz şekilde kapsadığı varsayımı altında, eğer P tane çerçeve varsa $2^P - 1$ tane de alan var demektir. İlgilendiğimiz çerçevelerin örnek yapıları farklı olabilir, ancak aşağıdaki tanımlamalar yapılmalıdır;

- Yığın içerisindeki her birim, en az bir çerçevenin birimi olmalıdır.
- Örnekteki her birimin, diğer çerçevede yer alıp almadığı bilinmemektedir.

$C \subseteq F$ için C olarak tanımlanan küme, $D_C = (\bigcap_{i \in C} A_i) \cap (\bigcap_{j \in C} A_j^k)$ 'dir. Örneğin $P=5$ olduğu durum için D_C olarak gösterilen 31 tane alan yada alt küme bulunmaktadır. Bu alanlar $C = \{(1), (2), (3), (4), (5), (12), (13), (14), (15), (23), (24), (25), (34), (35), (45), (123), (124), (125), (134), (135), (145), (234), (235), (245), (345), (1234), (1235), (1245), (1345), (2345), (12345)\}$ şeklinde ifade edilebilir.

A_F çerçevesinde yığın birimlerinin sayısı $N^{(F)}$ genellikle bilinirken, tanımlanan C kümesinde ki yığın birimlerinin sayısı N_C bilinmemektedir. Bu durumda yığın toplamı Y ,

$$Y_C = \sum_{i=1}^N \delta_i(C) y_i \quad (2.10)$$

$$\delta_i(L) = \begin{cases} 1, & i \in D_L \\ 0, & \text{diğer} \end{cases} \quad (2.11)$$

şeklinde yazılabilir.

Tüm i 'ler için $y_i=1$ olduğu durum Y_C için özel bir durumdur.

$$N_C = \sum_{i=1}^N \delta_i(C) \quad (2.12)$$

S_F , A_F 'nin olasılıksal örnekleme olsun. Çoklu çerçeve araştırmasında Y_C , C tane bağımsız tahmin ediciye sahiptir.

$$\hat{Y}_C^{(F)} = \sum_{i \in S_F} W_i^{(F)} \delta_i(C) y_i \quad (2.13)$$

Benzer şekilde NL tanım kümesinin de tahmin edicisi,

$$N_C^{(F)} = \sum_{i \in S_F} W_i^{(F)} \delta_i(C) y_i \quad (2.14)$$

Literatürde Mecatti (2007) tarafından yapılan çalışmada tanımlanan $W_i^{(F)}$ ağırlıklandırılmasının ayarlanmasını dikkate alan yığın toplamının tahmini için üç ana yaklaşım bulunmaktadır:

- Hartley (1962), Lund (1968) ve Fuller ve Burmeister (1972) tarafından en küçük varyans sağlayan ancak oldukça karmaşık yöntemleri içeren Optimal Tahmin Edicisi önerilmiştir.
- Bankier (1986), Kalton ve Anderson (1986) ve Skinner (1991) tarafından sabit ağırlıklandırma kullanarak yansız tahminleri garanti eden Tek Bazlı Tahmin Edici önerilmiştir. Tek Bazlı Tahmin Edici, Optimal Tahmin Edici'den daha etkindir.

- Skinner ve Rao (1996) ve Lohr ve Rao (2000) tarafından Optimal Tahmin Edici'den genişletilmiş uygulaması olan Pseudo En Çok Olabilirlik Tahmin Edicisi'ni önermiştir. Tek Bazlı Tahmin Edici'ye göre daha etkin bir tahmin edicidir (Aykuteli, 2011).

3. İKİLİ ÇERÇEVEDE BASİT TESADÜFİ ÖRNEKLEME

Hartley (1962 ve 1974), Fuller ve Burmeister (1972), Lund (1968), Bankier (1986), Kalton ve Anderson (1986), Skinner (1991), Skinner ve Rao (1996), Lohr ve Rao (2000 ve 2006) ve Mecatti (2005 ve 2007) tarafından çoklu çerçeveye ilişkin birçok araştırma yapılmıştır. Bu çalışmalarda alan ve çerçeve büyüklüğünün bilinip bilinmediği varsayımları altında bir çok tahmin edici geliştirilmiştir. Çizelge 3.1’de Hartley (1962) tarafından yapılan çalışmalarda kullanılan notasyonlar gösterilmiştir.

Çizelge 3.1. İki çerçeve düzeni ve tahminleri için notasyon gösterimi

	Çerçeve		Alan		
	A	B	a	b	ab
Yığındaki Birim Sayısı	N_A	N_B	N_a	N_b	N_{ab}
Örneğe Seçilen Birim Sayısı	n_A	n_B	n_a	n_b	n'_{ab}, n''_{ab}
Y için Yığın Toplamı	Y_A	Y_B	Y_a	Y_b	Y_{ab}
Y için Yığın Ortalaması	$\bar{Y}_A$	$\bar{Y}_B$	$\bar{Y}_a$	$\bar{Y}_b$	$\bar{Y}_{ab}$
Y için Örnek Toplamı	y_A	y_B	y_a	y_b	y'_{ab}, y''_{ab}
Y için Örnek Ortalaması	$\bar{y}_A$	$\bar{y}_B$	$\bar{y}_a$	$\bar{y}_b$	$\bar{y}'_{ab}, \bar{y}''_{ab}$
Örnek Birimlerinin Seçiminin Maliyeti	c_A	c_B			

Lund (1968) tarafından yapılan çalışmada, her iki çerçeve için yığın birimleri ve örnek birimleri Çizelge 3.2 ve Çizelge 3.3’de gösterilmiştir.

Çizelge 3.2. İki alan arasında herhangi bir çerçeve için yığın birimlerinin bölünmesi (Kaynak: Lund (1968))

N Büyüklüğüne Sahip Yığın		
N _A büyüklüğüne sahip A Çerçevesi		
N_a	N_{ab}	N_b
$\sigma^2_a, \bar{Y}_a$	$\sigma^2_{ab}, \bar{Y}_{ab}$	$\sigma^2_b, \bar{Y}_b$
	N _B büyüklüğüne sahip B Çerçevesi	

Çizelge 3.3. İki alan arasında herhangi bir çerçeve için seçilen örnek birimlerinin bölünmesi (Kaynak: Lund (1968))

A Çerçevesi Örneklemesi: n _A		
n_a	n'_{ab}	
$\bar{y}_a$	$\bar{y}'_{ab}$	
	B Çerçevesi Örneklemesi: n _B	
	n''_{ab}	n_b
	$\bar{y}''_{ab}$	$\bar{y}_b$

Eşleşen veya örtüşen A ve B çerçevelerinin birleşiminden oluşan U yığını;

$$U = A \cup B, \quad a = A \cap B^c$$

$$b = A^c \cup B, \quad ab = A \cap B$$

c kümenin bütünleyenidir.

U, A, B, a, b ve ab kümelerinin büyüklüğü sırasıyla N, N_A, N_B, N_a, N_b ve N_{ab} ile gösterilir ve

$$N = N_a + N_b + N_{ab} = N_a + N_B = N_A + N_b \quad (3.1)$$

$$N_A = N_a + N_{ab} \quad (3.2)$$

$$N_B = N_b + N_{ab} \quad (3.3)$$

şeklinde. A ve B çerçevelerinden sırasıyla bağımsız basit tesadüfî örnekler s_A ve s_B seçilmişlerdir ve

$$\begin{aligned} s_a &= s_A \cap a & s'_{ab} &= s_A \cap ab \\ s_b &= s_B \cap b & s''_{ab} &= s_B \cap ab \end{aligned}$$

şeklinde yazılabilir.

$s_A, s_B, s_a, s_b, s'_{ab}, s''_{ab}$ kümelerinin büyüklüğü sırasıyla, $n_A, n_B, n_a, n_b, n'_{ab}, n''_{ab}$ ile gösterilir.

En önemli nokta seçilen örnek birimlerinin hangi alana düşeceğidir. Basit tesadüfî örnekler s_A ve s_B , A ve B çerçevelerinden sırasıyla yerine konulmadan çekilmektedir.

Çoklu çerçeve araştırmalarında genel amaç, yığın toplamı olan Y 'nin tahmin edilmesidir. İkili çerçevede yığın toplamı olan Y ,

$$Y = Y_a + Y_b + Y_{ab}, \quad (3.4)$$

şeklinde yazılabilir. Y_a , a alanındaki yığın birimlerinin toplamı, Y_b , b alanındaki yığın birimlerinin toplamı, Y_{ab} ise ab alanındaki yani kesişen yığın birimlerinin toplamıdır. s_A ve s_B örnekleri sırasıyla A ve B çerçevelerinden bağımsız olarak çekilen örneklerdir. Burada belirtilen tahmin problemi, Y 'nin yansız tahmin edicisi $\hat{Y}$ 'yi ve Y 'nin varyansının tahmin edicisini oluşturmaktır. Bu durum N_A , N_B ve N_{AB} 'nin bilinmesi durumuna göre değişmektedir. Skinner ve Rao (1996) tarafından yapılan çalışmaya göre;

- 1) N_A , N_B ve N_{AB} 'nin bilinmesi,
- 2) N_A ve N_B 'nin bilinmesi fakat N_{AB} 'nin bilinmemesi,
- 3) N_A , N_B ve N_{AB} 'nin bilinmemesi

olmak üzere üç farklı durum söz konusudur. Bu çalışmada 1.ve 2. durum ele alınacaktır.

Yığın toplamının tahmini için;

- a) $n_A, n_B, n_a, n_b, n'_{ab}, n''_{ab}$
- b) $y_A, y_B, y_a, y_b, y'_{ab}, y''_{ab}$
- c) N_A ve N_B

değerlerinin bilindiği varsayılın.

Son olarak Y_A, Y_B, Y_a, Y_b ve Y_{ab} sırasıyla A, B, a, b ve ab alanları içerisinde yer alan i . birimlerine denk gelen y_i değerlerinin toplamını gösterir ve bu toplamın değerleri,

$$\begin{aligned}\bar{Y}_A &= Y_A / N_A, & \bar{Y}_B &= Y_B / N_B, & \bar{Y}_a &= Y_a / N_a, & \bar{Y}_b &= Y_b / N_b, \\ \bar{y}_A &= y_A / n_A, & \bar{y}_B &= y_B / n_B, & \bar{y}_a &= y_a / n_a, & \bar{y}_b &= y_b / n_b, \\ \bar{y}_{ab} &= y_{ab} / n_{ab}, & \bar{y}'_{ab} &= y'_{ab} / n'_{ab}, & \bar{y}''_{ab} &= y''_{ab} / n''_{ab} \\ y_{ab} &= y'_{ab} + y''_{ab} \text{ ve } n_{ab} = n'_{ab} + n''_{ab}\end{aligned}$$

şeklinde yazılır.

3.1. N_A, N_B ve N_{ab} 'nin Bilindiği Durumda Yığın Toplamı Tahmin Edicileri

Hartley (1962) ve Lund (1968) tarafından N_A, N_B ve N_{ab} 'nin bilindiği durumlar için ikili çerçeve tahmin edicileri önerilmektedir.

3.1.1. Hartley tahmin edicisi

Hartley (1962) yığın toplamı için yansız tahmin ediciyi;

$$\hat{Y}_H = N_a \bar{y}_a + N_{ab} (p \bar{y}'_{ab} + q \bar{y}''_{ab}) + N_b \bar{y}_b \quad (3.5)$$

şeklinde ifade etmiştir.

Hartley'in tahmin edicisinin varyansı,

$$V(\hat{Y}_H) = \frac{N_A^2}{n_A} \{ \sigma_a^2 (1 - \alpha) + p^2 \sigma_{ab}^2 \alpha \} + \frac{N_B^2}{n_B} \{ \sigma_b^2 (1 - \beta) + q^2 \sigma_{ab}^2 \beta \} \quad (3.6)$$

$$\alpha = N_{ab}/N_A, \quad \beta = N_{ab}/N_B$$

şeklinde yazılabilir. σ_a^2 , σ_b^2 ve σ_{ab}^2 terimleri alan kapsamında yığın varyanslarını temsil etmektedir.

Maliyetin doğrusal bir fonksiyon olduğu varsayımı altında, toplam maliyet;

$$C = n_A c_A + n_B c_B \quad (3.7)$$

şeklinde açıklanabilir.

c_A ve c_B , A ve B çerçevelerindeki birimlerinin sıralanmasının maliyetlerini tanımlamaktadır. İki çerçeve arasında örnekleme dağıtımının optimizasyonu ve p , n_A ve n_B 'nin bir fonksiyonu olan $Var(\hat{Y}_H)$ en küçüklemesinden elde edilen en uygun p değeri maliyet fonksiyonu kısıtlamasına maruz kalmaktadır (Aykuteli, 2011).

p için optimal değer,

$$\frac{c_A p^2}{c_B q^2} = \frac{\sigma_a^2(1-a) + p^2 \sigma_{ab}^2 a}{\sigma_b^2(1-\beta) + q^2 \sigma_{ab}^2 \beta} \quad (3.8)$$

olacaktır.

Optimal örnek oranı ise,

C: maliyet, olmak üzere

$$\frac{n_A}{N_A} = C \{(\sigma_a^2 (1 - a) + q^2 \sigma_{ab}^2 a)/c_A\}^{1/2} \quad (3.9)$$

$$\frac{n_B}{N_B} = C \{(\sigma_b^2 (1 - \beta) + q^2 \sigma_{ab}^2 \beta)/c_B\}^{1/2} \quad (3.10)$$

şeklinde elde edilebilir.

3.1.2. Lund tahmin edicisi

Hartley'in p değerini Lund (1968),

$$p_{0H} = \frac{\alpha n_A}{\alpha n_A + \beta n_B} \quad (3.11)$$

şeklinde tanımlamıştır.

Lund, Hartley tarafından tanımlanan uygulama süreci, çerçeveler arasında n_A ve n_B birimlerinin tesadüfi olarak bölünmesini kabul etmediğini düşünmektedir. Lund, p 'nin n'_{ab} ve n''_{ab} 'nin bir fonksiyonu olduğu durumda başarılı bir kazancın olup olmadığını incelemiştir (Aykuteli, 2011).

$\hat{Y}_H$ 'nin (3.12)'de verilen varyansı,

$$Var(\hat{Y}) = E[Var(\hat{Y} | n'_{ab}, n''_{ab})] + Var[E(\hat{Y} | n'_{ab}, n''_{ab})] \quad (3.12)$$

$$E(\hat{Y} | n'_{ab}, n''_{ab}) = Y \text{ ve } Var[E(\hat{Y} | n'_{ab}, n''_{ab})] = 0$$

sonlu yığın düzeltmesi ihmal edildiğinde,

$$Var(\hat{Y} | n'_{ab}, n''_{ab}) = \frac{N_a^2}{n_a} \sigma_a^2 + p^2 \frac{N_{ab}^2}{n'_{ab}} \sigma_{ab}^2 + (1-p)^2 \frac{N_{ab}^2}{n''_{ab}} \sigma_{ab}^2 + \frac{N_b^2}{n_b} \sigma_b^2 \quad (3.13)$$

şeklinde yazılabilir. (3.13)'deki varyansı en küçük yapan p değeri,

$$\frac{\partial Var(\hat{Y} | n'_{ab}, n''_{ab})}{\partial p} = \frac{Var(\bar{y}''_{ab})}{Var(\bar{y}'_{ab}) + Var(\bar{y}''_{ab})}$$

$$p_o = \frac{n'_{ab}}{n'_{ab} + n''_{ab}} \quad q = 1 - p \quad (3.14)$$

şeklinde ifade edilebilir.

(3.13) eşitliğiyle (3.14) eşitliği yer değiştirdiğinde beklenen değer elde edilir.

(3.14)'deki p değeri ile tahmin edicinin varyansı bulunabilir.

Yığın toplamının tahmin edicisi ve yaklaşık varyansı,

$$\hat{Y}_L = N_a \bar{y}_a + N_{ab} \bar{y}_{ab} + N_b \bar{y}_b \quad (3.16)$$

$$\bar{y}_{ab} = \frac{n'_{ab}\bar{y}'_{ab} + n''_{ab}\bar{y}''_{ab}}{n'_{ab} + n''_{ab}} \quad (3.17)$$

$$V(\hat{Y}_L) = \frac{N_A^2}{n_A} \sigma_a^2 (1 - \alpha) + \frac{N_A N_B \alpha \beta}{\alpha n_A + \beta n_B} \sigma_{ab}^2 + \frac{N_B^2}{n_B} \sigma_b^2 (1 - \beta) \quad (3.18)$$

şeklinde hesaplanır.

3.2. N_A ve N_B 'nin Bilindiği, N_{ab} 'nin Bilinmediği Durumda Yığın Toplamı Tahmin Edicileri

Literatürde, Hartley (1962,1974), Lund (1968), Fuller ve Burmeister (1972), Bankier (1986), Skinner (1991), Skinner ve Rao (1996) ve Mecatti (2005) tarafından önerilen farklı tahmin ediciler bulunmaktadır. Bu durum, A ve B çerçevelerine ilişkin listenin uzunluğunun bilinmesine karşın eşleşen büyüklüğün tanımlanmasındaki güvenilirliğin uygulamada eşleşme zorlukları nedeniyle mümkün olmamasına neden olabilmektedir.

A ve B çerçevelerinin olduğu ve sırasıyla N_A ve N_B birim sayılarını içerdiklerini varsayalım. Böylece,

$$N_A = N_a + N_{ab} \text{ ve } N_B = N_b + N_{ab}$$

olduğu ve yığındaki birimlerinin sayısının toplamı N ,

$$N = N_a + N_b + N_{ab} = N_a + N_B = N_A + N_b$$

şeklinde ifade edilebilir.

Bu bilgiler ışığında N_a , N_b ve N_{ab} 'nin tahminleri hesaplanmalıdır. N_A ve N_B 'nin bilindiği durumda sadece N_{ab} 'deki birimlerinin sayısını tahmin etmek gerekmektedir. N_{ab} tahmin edildikten sonra N_a ve N_b 'nin tahminleri,

$$\hat{N}_a = N_A - \hat{N}_{ab} \text{ ve } \hat{N}_b = N_B - \hat{N}_{ab}$$

eşitlikleri ile kolayca bulunabilir. Yığındaki birimlerin sayısının tahmini,

$$\hat{N} = N_A + N_B - \hat{N}_{ab}$$

eşitliği ile tahmin edilebilir.

3.2.1. Hartley tahmin edicisi

Hartley (1962) tarafından yapılan çalışmada N_a , N_b ve N_{ab} örneklem büyüklükleri bilinmediği durumda tabakalı örnekleme eşitliklerinin iki çerçevedeki (tabakadaki) değişkenlere uygulamayı önermiştir. Hartley tahmin ediciyi,

$$\hat{Y}_H = f_A^{-1}(y_a + p y'_{ab}) + f_B^{-1}(y_b + q y''_{ab}) \quad (3.19)$$

şeklinde elde etmiştir.

$$f_A^{-1} = \frac{N_A}{n_A}, f_B^{-1} = \frac{N_B}{n_B}$$

olarak gösterilmiştir.

Tahmin edicinin varyansı,

$$\begin{aligned} V(\hat{Y}_H) = & \frac{N_A^2}{n_A} \{ \sigma_a^2 (1 - \alpha) + p^2 \sigma_{ab}^2 \alpha + \alpha (1 - \alpha) (\bar{Y}_a - p \bar{Y}_{ab})^2 \} \\ & + \frac{N_B^2}{n_B} \{ \sigma_b^2 (1 - \beta) + q^2 \sigma_{ab}^2 \beta + \beta (1 - \beta) (\bar{Y}_b - q \bar{Y}_{ab})^2 \} \end{aligned} \quad (3.20)$$

eşitliği ile bulunur.

Optimal dağıtım ise,

$$\frac{n_A^2}{N_A^2} = C \{ (\sigma_a^2 (1 - a) + p^2 \sigma_{ab}^2 a) + a(1 - a)(\bar{Y}_a - p\bar{Y}_{ab})^2 \} / c_A \quad (3.21)$$

$$\frac{n_B^2}{N_B^2} = C \{ (\sigma_b^2 (1 - \beta) + q^2 \sigma_{ab}^2 \beta) + \beta(1 - \beta)(\bar{Y}_b - p\bar{Y}_{ab})^2 \} / c_B \quad (3.22)$$

eşitlikleri ile bulunabilir.

3.2.2. Lund tahmin edicisi

Lund (1968) yaptığı çalışmada N_a , N_b ve N_{ab} 'nin bilinmediği durumlar için tahmin ediciler önermiştir. Lund, bilinmeyen tanım kümesinin büyüklüğünün tahminini kullanmıştır. N_a ve N_b 'nin yansız tahmin edicileri sırasıyla, $N_A (n_a / n_A)$ ve $N_B (n_b / n_B)$ olarak tanımlanmaktadır. $N_A (n'_a / n_A)$ ve $N_B (n''_b / n_B)$ eşitlikleri ise N_{ab} için yansız tahmin edicilerdir.

N_{ab} 'nin iki tahmini için tanımlanmamış ağırlıkları olan p ve $(1 - p)$ 'nin kullanımı ve (3.1.2.5)'deki tahmin edicide yer değiştirilirse yığın toplamının yansız tahmin edicisi,

$$\hat{Y}_L = f_A^{-1} y_a + (p f_A^{-1} n'_{ab} + q f_B^{-1} n''_{ab}) \bar{y}_{ab} + f_B^{-1} y_b \quad (3.23)$$

$$\bar{y}_{ab} = \frac{n'_{ab} \bar{y}'_{ab} + n''_{ab} \bar{y}''_{ab}}{n'_{ab} + n''_{ab}}$$

olarak ifade edilir. Lund tahmin edicisinin varyansı ise,

$$V(\hat{Y}_L) = \frac{N_A^2}{n_A} \sigma_a^2 (1 - \alpha) + \frac{N_A N_B \alpha \beta}{\alpha n_A + \beta n_B} \sigma_{ab}^2 + \frac{N_B^2}{n_B} \sigma_b^2 (1 - \beta) + \frac{N_A^2 (1 - \alpha) \alpha}{n_A} (\bar{Y}_a - p \bar{Y}_{ab})^2 + \frac{N_B^2 (1 - \beta) \beta}{n_B} (\bar{Y}_b - q \bar{Y}_{ab})^2 \quad (3.24)$$

şeklinde ifade edilir. Eşitlik (3.2.2.2)'de yer alan son iki terim N_a , N_b ve N_{ab} 'nin bilinmediği durumda varyansı artıran terim olarak gösterilmektedir.

Lund p 'nin optimal değerini ve örneklem verisinden elde edilen optimal p 'nin tahmin edicisini,

$$\hat{p}_L = \frac{\frac{N_A n_a}{n_A^2} \bar{y}_a + \frac{N_B n_b}{n_B^2} (\bar{y}_{ab} - \bar{y}_b)}{\left(\frac{N_A n_a}{n_A^2} + \frac{N_B n_b}{n_B^2} \right) \bar{y}_{ab}} \quad (3.25)$$

şeklinde göstermiştir. Lund (1968) tarafından yapılan çalışmada Hartley'in ağırlık değişkeni p 'nin tekrarlı birimlerinden oluşan $(N_{ab}, \bar{Y}_{ab})$ üzerinden toplanan iki bireysel tahmin edici ile hesaplanması gösterilmektedir. p 'nin optimal değeri, σ_a^2 , σ_b^2 ve $\sigma_{a,b}^2$ parametrelerini içerdiğinden Lund tahmin edicisi, Hartley tahmin edicisinden daha basittir. Skinner (1991) çalışmasında p 'nin her değeri için $Var(\hat{Y}_L) \leq Var(\hat{Y}_H)$ olduğunu göstermiştir.

3.2.3. Fuller ve Burmeister tahmin edicisi

Fuller ve Burmeister tahmin ediciyi,

$$\hat{Y}_{FB} = (N_A - \hat{N}_{ab,FB}) \bar{y}_a + \hat{N}_{ab,FB} \bar{y}_{ab} + (N_B - \hat{N}_{ab,FB}) \bar{y}_b \quad (3.26)$$

şeklinde göstermiştir ve

$$\bar{y}_{ab} = p_{FB} \bar{y}'_{ab} + (1 - p_{FB}) \bar{y}''_{ab} \quad (3.27)$$

'dir.

A çerçevesinden seçilmiş, ab alanında yer alan birimlerin ortalaması;

$$\bar{y}'_{ab} = \frac{1}{n'_{ab}} \sum_{i=1}^{n'_{ab}} y_i$$

B çerçevesinden seçilmiş, ab alanında yer alan birimlerin ortalaması

$$\bar{y}''_{ab} = \frac{1}{n''_{ab}} \sum_{i=1}^{n''_{ab}} y_i$$

$$p_{FB} = \frac{n'_{ab}(1-f_B)}{n'_{ab}(1-f_B) + n''_{ab}(1-f_A)} \quad (3.28)$$

şeklinde gösterilmektedir ve bu değer varyansı minimize etmektedir.

$$[n_A g_B + n_B g_A] \hat{N}_{ab,FB}^2 - [n_A N_B g_B + n_B N_A g_A + n'_{ab} N_A g_B + n''_{ab} N_B g_A] \hat{N}_{ab,FB} + [n'_{ab} g_B + n''_{ab} g_A] N_A N_B = 0$$

$$g_A = \frac{N_A - n_A}{N_A - 1}, \quad g_B = \frac{N_B - n_b}{N_B - 1}$$

şeklinde verilmiştir.

Yukarıda verilen kuadratik form üzerinden N_{ab} tahmin edilebilir.

Bu tahmin edicinin varyansı ise;

$$\begin{aligned} Var(\hat{Y}_{FB}) = & N_a(f_A^{-1} - 1)S_a^2 + [(1 - f_B)f_A + (1 - f_A)f_B]^{-1} \\ & (1 - f_A)(1 - f_B)N_{ab}S_{ab}^2 + N_b(f_B^{-1} - 1)S_b^2 \\ & + (\bar{Y}_{ab} - \bar{Y}_a - \bar{Y}_b)^2 \frac{N_{ab}N_aN_bg_Ag_B}{n_A N_b g_B + n_B N_a g_A} \end{aligned} \quad (3.29)$$

şeklinde gösterilmektedir.

Alan varyansları ise;

$$S_a^2 = \frac{1}{N_a - 1} \sum_{i=1}^{N_a} (y_{ai} - \bar{Y}_a)^2 \quad (3.30)$$

$$S_{ab}^2 = \frac{1}{N_{ab} - 1} \sum_{i=1}^{N_{ab}} (y_{abi} - \bar{Y}_{ab})^2 \quad (3.31)$$

$$S_b^2 = \frac{1}{N_b - 1} \sum_{i=1}^{N_b} (y_{bi} - \bar{Y}_b)^2 \quad (3.32)$$

şeklinde gösterilmektedir.

3.2.4. Tek çerçeve (Single Frame) tahmin edicisi

Bankier (1986), aşağıda verilen formu kullanarak Horwitz-Thompson tahmin edicisine dayanarak bir tahmin edici önermiştir. Birden fazla örneklemden seçilen birimlerin tanımlanabileceği varsayılmıştır. Fakat, Kalton ve Anderson (1986) ve Skinner (1991)'in tahmin edicisi her iki örnekte bulunan birimlerin tanımlanmasını gerektirmemektedir.

İkili çerçeve için Tek Çerçeve tahmin edicisi aşağıdaki gibi verilmiştir;

$$\hat{Y}_{SF} = f_A^{-1}y_a + f_B^{-1}y_b + (f_A + f_B)^{-1}y_{ab} \quad (3.33)$$

Skinner (1991), $\hat{Y}_{SF}$ 'nin $\hat{Y}_H$ ya da $\hat{Y}_L$ 'nin özel bir hali olduğunu göstermiştir.

$$p = f_A (f_A + f_B)^{-1} \quad \text{ve} \quad \text{Var}(\hat{Y}_{SF}) \geq \text{Var}(\hat{Y}_{FB}).$$

$$\hat{Y}_S = f_A^{-1}[y_a + py'_{ab}] + f_B^{-1}[y_b + qy''_{ab}] \quad (3.34)$$

3.2.5. Skinner'in sıralama oranı tahmin edicisi

Skinner' in sıralama oranı tahmin edicisi için yığın toplamı,

$$\hat{Y}_{RR}^{(\infty)} = (N_A - \hat{N}_{ab,RR}) \bar{y}_a + \hat{N}_{ab,RR} \bar{y}_{ab} + (N_B - \hat{N}_{ab,RR}) \bar{y}_b \quad (3.35)$$

şeklinde gösterilmiştir.

$$n_{ab} \hat{N}_{ab,RR}^2 - [n_{ab}(N_A + N_B) + (f_A^{-1} + f_B^{-1})n_a n_b] \hat{N}_{ab,RR} + n_{ab} N_A N_B = 0$$

Bu kuadratik form üzerinden N_{ab} tahmin edilebilir.

Bu tahmin edicinin varyansı ise,

$$n \rightarrow \infty$$

iken

$$\begin{aligned}
Var(\hat{Y}_{RR}^{(\infty)}) &\sim f_A^{-1}N_a S_a^2 + f_B^{-1}N_b S_b^2 + (f_A + f_B)^{-1}N_{ab} S_{ab}^2 \\
&\quad + (\bar{Y}_{ab} - \bar{Y}_a - \bar{Y}_b)^2 \frac{N_{ab}N_aN_b}{n_A N_b + n_B N_a} (1 + \lambda^2)
\end{aligned} \tag{3.36}$$

ifadesine yaklaşmaktadır.

$$\lambda^2 = \frac{N_a N_b N_{ab}^2 (f_A^2 N_A - f_B^2 N_B)^2}{N_A N_B f_A f_B (f_A + f_B)^2 (N_{ab}^2 - N_A N_B)^2} \tag{3.37}$$

şeklinde verilmiştir.

4. GİRİŞİMCİ BİLGİ SİSTEMİ

Girişimci Bilgi Sistemi (GBS); farklı kamu kurum ve kuruluşların idari kayıtlarındaki girişim ve işyerlerine ait ekonomik faaliyetlere ilişkin verilerin, ortak standartlar çerçevesinde bir veritabanında toplanması ve bu verilerin entegrasyonu projesidir.

GBS'nin amaçları:

- Makro ve mikro düzeyde, kamu kesimindeki politika yapıcılar ile karar alıcılara, ekonomi politikaların tasarlanması, uygulanması ve etkinlik ölçümü için doğru ve güvenilir bilgilerin sunulacağı bir ortam sağlamak,
- Ekonomi analizleri çalışmalarında kullanılmak üzere, üniversitelere, enstitülere, sivil toplum kuruluşlarına, özel araştırma kuruluşları ile bireysel araştırmacılara doğru ve güvenilir bilgiler sunmak,
- Ülkemizde yatırım yapmak isteyen yerli ve yabancı girişimcilere, yapacakları yatırımların tespiti konusunda bölgesel ve sektörel bazda bilgiler sunmak,
- İlgili paydaşlarla bir geri bildirim ve koordinasyon mekanizması kurularak, kamu kurum ve kuruluşlarının idari kayıtlarındaki ekonomik faaliyet verilerinin ortak standartlarda tutulması konusuna dikkat çekmek ve bu konularda bir bilinç ve farkındalık oluşturmak,

şeklinde özetlenebilir.

Girişimci Bilgi Sistemi, Bilgi Toplumu Stratejisi (2006-2010)'nde yer almaktadır. Aynı zamanda Projeye, 2011–2013 Dönemi Orta Vadeli Program'ının Gelişme Eksenleri bölümünün Rekabet Gücünün Artırılması başlığı altında değinilmektedir. GBS'ye Bilim, Sanayi ve Teknoloji Bakanlığı Sanayi Sicil veritabanında yer alan verilerin yanı sıra;

- Gümrük ve Ticaret Bakanlığı
- Gelir İdaresi Başkanlığı
- Sosyal Güvenlik Kurumu
- TÜBİTAK
- Türk Patent Enstitüsü
- KOSGEB

- TÜİK

verileri entegre edilmiştir.

GBS ile, ülkemizde yerleşik girişimlerin;

- Mali Yapıları (Net Satış, Aktif, Kar, Borçluluk vb.),
- Sektörler ve İller Arası Ticareti,
- İstihdam (Yaş, Cinsiyet, Ücret Miktarı, Nitelik),
- İthalat ve İhracat Değerleri (Ürün, Ülke ve İl Bazında),
- Fikri Mülkiyet Hakları (Patent, Faydalı Model, Marka),
- KOSGEB ve TÜBİTAK Destekleri (Kredi ve Diğer Destekler),
- Sanayi İşletmelerinin Kurulu Kapasiteleri (Ürün ve İl Bazında),
- Sanayi İşletmelerinin Fiili Üretimleri (Ürün ve İl Bazında),
- Girişim ve İşyeri Sayıları

gibi yaklaşık 500 adet veri alanını Ölçek, Sektör, Bölge ve Teknoloji Düzeyine göre raporlanabilmektedir.

GBS, kurumlar vergisi mükellefleri ile yalnızca ticari kazanç elde eden gelir vergisi mükelleflerinin farklı kamu kurum ve kuruluşların idari kayıtlarındaki ekonomik faaliyet verilerini içermektedir. Ticari kazanç, her türlü ticari ve sınai faaliyetlerden doğan kazançlardır. Ticari faaliyette temel olan faaliyetin sermaye ve emek içeren bir organizasyona dayanması ve süreklilik arz etmesidir. GBS'de, 5510 sayılı Sosyal Sigortalar ve Genel Sağlık Sigortası Kanununun 4/a maddesine göre zorunlu sigortalı işçi çalıştıran işyerleri ile 4/a zorunlu sigortalı çalışanlar dikkate alınmıştır. 6948 Sayılı Sanayi Sicili Kanunun kapsamında Bilim, Sanayi ve Teknoloji Bakanlığı tarafından kayıt altına alınan sanayi işletmelerinin “yıllık işletme cetveli” vasıtasıyla her yıl Bakanlığa bildirilen çalışan sayıları da sisteme entegre edilmiştir. Paydaş kurum ve kuruluşların idari kayıtlarında yer almakla beraber, ticari faaliyette bulunmayan girişim/işyerlerine ait veriler GBS kapsamı dışındadır. (Zorunlu sigortalı işçi çalıştıran apartman yönetimleri, patent başvurusunda bulunan mucitler vb.) Girişim ve işyerlerinin ekonomik faaliyetleri, Avrupa Topluluğunda Ekonomik Faaliyetlerin İstatistik Sınıflaması NACE Rev. 2' ye göre sınıflandırılmıştır.

Giriřim, birinci derecede karar alma özerklięini kullanarak, mal ve hizmet üreten bir organizasyon biçimidir. Giriřim, bir veya birden fazla yerde, bir veya birden fazla faaliyet gerçekleřtirebilir. Bir giriřim yasal birime ya da yasal birimlerin birleřimine karřılık gelmektedir. SGK Çalışan Sayısı, referans dönemi son ayında adlarına bildirimde bulunulan 4/a zorunlu sigortalı çalışanların sayısıdır.

GBS’de yer alan veriler, 5429 Sayılı Türkiye İstatistik Kanunu ve ilgili dięer mevzuat çerçevesinde paylaşılmaktadır.

5. İKİLİ ÇERÇEVE İÇİN UYGULAMA

Bu çalışmanın amacı, Türkiye’de imalat sanayiinde faaliyet gösteren işletmelerin bilgileri kullanılarak, iki çerçeve için örnek çekimi ile sanayi işletmelerinde çalışan kişi sayısını tahmin etmektir.

Yapılan çalışmada *A* çerçevesi olarak Türkiye’nin seçilmiş bir ilinde, 2012 yılında imalat sanayiinde faaliyet gösteren işletmelerin Sosyal Güvenlik Kurumu’nda yer alan çalışan kayıtları, *B* çerçevesi olarak yine aynı ilde ve 2012 yılında imalat sanayide faaliyet gösteren işletmelerin Bilim, Sanayi ve Teknoloji Bakanlığı’nın sanayi sicil kayıtlarında yer alan çalışan kayıtları kullanılmıştır.

Türkiye’den seçilmiş olan söz konusu ilin verilerine, iki ayrı çerçevenin birimleri için sanal numaralar atanmıştır. Bu atama sonucu 1394 tane sanayi işletmesinin kesiştiği görülmüştür. Eksik olan birimler çerçevelerden çıkartılmış, eşleşmeyen veriler içinse yeni veriler üretilmiştir. Bunun sonucunda toplam çalışan sayısının değiştirilmemesine özen gösterilmiştir.

Bu iki çerçeveye ait hesaplamalar aşağıda verilmiştir;

$$\begin{array}{llllll}
 N_A = 4695 & Y_A = 94699 & & & & \\
 N_B = 1467 & Y_B = 80798 & & & & \\
 N_a = 3301 & Y_a = 15081 & \bar{Y}_a = 4,568 & S_a^2 = 292,469 & \sigma_a^2 = 292,381 & \\
 N_{ab} = 1394 & Y_{ab} = 79618 & \bar{Y}_{ab} = 57,114 & S_{ab}^2 = 32534,887 & \sigma_{ab}^2 = 32558,243 & \\
 N_b = 73 & Y_b = 1180 & \bar{Y}_b = 16,164 & S_b^2 = 2485,694 & \sigma_b^2 = 2451,644 &
 \end{array}$$

$$N_a = N_A - N_{ab} = 4695 - 1394 = 3301$$

$$N_b = N_B - N_{ab} = 1467 - 1394 = 73$$

$$N = N_A + N_B - N_{ab} = N_a + N_b + N_{ab}$$

$$N = 4695 + 1467 - 1394 = 3301 + 73 + 1394$$

$$N = 4768$$

$$Y = Y_A + Y_B - Y_{ab} = Y_a + Y_b + Y_{ab}$$

$$Y = 94699 + 80798 - 79618 = 15081 + 1180 + 79618$$

$$Y = 95879$$

İkili çerçeve tahmin edicileri ile toplam tahminleri yapılmadan önce iki yığından basit tesadüfî örnekleme yöntemi ile örnek çekimi yapılmıştır. Buna göre örnek hacimleri, örneklerin toplam, ortalama değerleri ve örneklem oranları şöyledir;

$$n_A = 1174 \quad y_A = 22795$$

$$n_B = 367 \quad y_B = 18443$$

$$n_a = 880 \quad y_a = 4892$$

$$n_b = 46 \quad y_b = 658$$

$$\bar{y}_a = \frac{4892}{880} = 5,559 \quad \bar{y}_b = \frac{658}{46} = 14,304$$

$$n'_{ab} = 294 \quad y'_{ab} = 17903$$

$$n''_{ab} = 321 \quad y''_{ab} = 17785$$

$$\bar{y}'_{ab} = \frac{y'_{ab}}{n'_{ab}} = \frac{17903}{294} = 60,894 \quad \bar{y}''_{ab} = \frac{y''_{ab}}{n''_{ab}} = \frac{17785}{321} = 55,404$$

$$n_{ab} = n'_{ab} + n''_{ab} = 294 + 321 = 615$$

$$y_{ab} = y'_{ab} + y''_{ab} = 17903 + 17785 = 35688$$

$$\bar{y}_{ab} = \frac{y_{ab}}{n_{ab}} = \frac{35688}{615} = 58,029$$

$$f_A = \frac{n_A}{N_A} = \frac{1174}{4695} = 0,25 \quad f_B = \frac{n_B}{N_B} = \frac{367}{1467} = 0,25$$

5.1. N_A , N_B ve N_{ab} 'nin Bilindiği Durum İçin Toplam Çalışan Sayısı Tahminleri

N_A , N_B ve N_{ab} 'nin bilindiği zaman toplam çalışan için Hartley'in yansız tahmin edicisi,

$$\hat{Y}_H = N_a \bar{y}_a + N_{ab} (p \bar{y}'_{ab} + q \bar{y}''_{ab}) + N_b \bar{y}_b$$

şeklinde gösterilmiştir.

Lund (1968) Hartley'in kuadratik p ve q değerini aşağıdaki biçimde göstermiştir. Buna göre örnekten hesaplandığı gibi;

$$p_H = \frac{\alpha n_A}{\alpha n_A + \beta n_B} = \frac{0,296 * 1167}{0,296 * 1167 + 0,95 * 367} = 0,5 \quad q_H = 0,5$$

değerleri bulunmuştur. Burada;

$$\alpha = N_{ab}/N_A = 1394/4695 = 0,296, \quad \beta = N_{ab}/N_B = 1394/1467 = 0,95$$

eşitliği kullanılarak elde edilir.

Hartley'e göre toplam çalışan sayısının tahmini;

$$\hat{Y}_H = 3301 * 5,559 + 1394 * (0,5 * 60,894 + 0,5 * 55,404) + 73 * 14,304$$

$$\hat{Y}_H = 100454,157 \cong 100454$$

şeklinde bulunur.

Bu tahmin edicinin varyansı ise;

$$V(\hat{Y}_H) = \frac{N_A^2}{n_A} \{ \sigma_a^2 (1 - \alpha) + p^2 \sigma_{ab}^2 \alpha \} + \frac{N_B^2}{n_B} \{ \sigma_b^2 (1 - \beta) + q^2 \sigma_{ab}^2 \beta \}$$

$$V(\hat{Y}_H) = \frac{4695^2}{1174} \{ 292,381 * (1 - 0,296) + 0,5^2 * 32534,89 * 0,296 \}$$

$$+ \frac{1467^2}{367} \{ 2451,644 * (1 - 0,95) + 0,5^2 * 32534,89 * 0,05 \}$$

$$V(\hat{Y}_H) = 95099696,32$$

şeklinde bulunur.

Değişim katsayısı ise aşağıdaki şekilde hesaplanır:

$$DK_H = \frac{\sqrt{V(\hat{Y}_H)}}{\hat{Y}_H} = \frac{\sqrt{95099696,32}}{100454} = 0,097$$

olarak bulunur.

N_A , N_B ve N_{ab} 'nin bilindiği zaman toplam çalışan için Lund'un yansız tahmin edicisi;

$$\hat{Y}_L = N_a \bar{y}_a + N_{ab} \bar{y}_{ab} + N_b \bar{y}_b$$

$$\hat{Y}_L = 3301 * 5,559 + 1394 * 58,029 + 73 * 14,30$$

$$\hat{Y}_L = 100286,877 = 100287$$

olarak bulunur.

Bu tahmin edicinin varyansı ise;

$$V(\hat{Y}_L) = \frac{N_A^2}{n_A} \sigma_a^2 (1 - \alpha) + \frac{N_A N_B \alpha \beta}{\alpha n_A + \beta n_B} \sigma_{ab}^2 + \frac{N_B^2}{n_B} \sigma_b^2 (1 - \beta)$$

$$V(\hat{Y}_L) = \frac{4695^2}{1174} * 292,381 * (1 - 0,296) + \frac{4965 * 1467 * 0,296 * 0,95}{0,296 * 1174 + 0,95 * 367} * 32534,89$$

$$+ \frac{1467^2}{367} * 2451,644 * (1 - 0,95)$$

$$V(\hat{Y}_L) = 95099532,91$$

olur.

Değişim katsayısı ise aşağıdaki şekilde hesaplanır:

$$DK_L = \frac{\sqrt{V(\hat{Y}_L)}}{\hat{Y}_L} = \frac{\sqrt{95099532,91}}{100287} = 0,0972$$

olarak bulunur.

N_A , N_B ve N_{ab} 'nin bilindiği zaman, bu iki tahmin edicinin değişim katsayılarına baktığımızda, $DK_H < DK_L$ 'dir. Bu durumda $\hat{Y}_H$ tahmin edicinin değişim katsayısı $\hat{Y}_L$ tahmin edicinin değişim katsayısından daha küçük olduğundan $\hat{Y}_H$ tahmin edicinin kullanılması daha iyi bir yaklaşımdır denilebilir.

5.2. N_A ve N_B 'nin Bilindiği N_{ab} 'nin Bilinmediği Durum İçin Toplam Çalışan Sayısı Tahminleri

N_A ve N_B 'nin bilindiği, N_{ab} 'nin bilinmediği yığın toplamı için Hartley'in yansız tahmin edicisi;

$$\hat{Y}_H = f_A^{-1}(y_a + p y'_{ab}) + f_B^{-1}(y_b + q y''_{ab})$$

$$\hat{Y}_H = 4 * (4892 + 0,5 * 17903) + 4 * (658 + 0,5 * 17785)$$

$$\hat{Y}_H = 93588$$

olarak bulunur.

Bu tahmin edicinin varyansı ise;

$$V(\hat{Y}_H) = \frac{N_A^2}{n_A} \{ \sigma_a^2 (1 - \alpha) + p^2 \sigma_{ab}^2 \alpha + \alpha (1 - \alpha) (\bar{Y}_a - p \bar{Y}_{ab})^2 \} \\ + \frac{N_B^2}{n_B} \{ \sigma_b^2 (1 - \beta) + q^2 \sigma_{ab}^2 \beta + \beta (1 - \beta) (\bar{Y}_b - q \bar{Y}_{ab})^2 \}$$

$$V(\hat{Y}_H) = \frac{4695^2}{1174} \{ \{ 292,381 * (1 - 0,296) + 0,5^2 * 32534,89 * 0,296 \} \\ + \{ 0,296 (1 - 0,296) (4,568 - 0,5 * 57,114)^2 \} \} \\ + \frac{1467^2}{367} \{ \{ 2451,644 * (1 - 0,95) + 0,5^2 * 32534,89 * 0,95 \} \\ + \{ 0,95 (1 - 0,95) (16,164 - 0,5 * 57,114)^2 \} \}$$

$$V(\hat{Y}_H) = 97394059,9$$

olarak bulunur.

Değişim katsayısı ise aşağıdaki şekilde hesaplanır:

$$DK_H = \frac{\sqrt{V(\hat{Y}_H)}}{\hat{Y}_H} = \frac{\sqrt{97394059,9}}{93588} = 0,105$$

olarak bulunur.

Yığın toplamı için Lund'un yansız tahmin edicisi;

$$\hat{Y}_L = f_A^{-1} y_a + (p f_A^{-1} n'_{ab} + q f_B^{-1} n''_{ab}) \bar{y}_{ab} + f_B^{-1}$$

ile hesaplanmaktadır.

Buradaki optimal p değeri aşağıdaki formülasyonla hesaplanmaktadır:

$$\hat{p}_L = \frac{\frac{N_A n_a}{n_A^2} \bar{y}_a + \frac{N_B n_b}{n_B^2} (\bar{y}_{ab} - \bar{y}_b)}{\left(\frac{N_A n_a}{n_A^2} + \frac{N_B n_b}{n_B^2} \right) \bar{y}_{ab}}$$

$$\hat{p}_L = \frac{\frac{4695 * 880}{1174^2} * 5,559 + \frac{1467 * 46}{367^2} * (58,029 - 14,304)}{\left(\frac{4695 * 880}{1174^2} + \frac{1467 * 46}{367^2} \right) * 58,029} = 0,189$$

$$\hat{q}_L = 1 - 0,189 = 0,811$$

olarak bulunur.

Buna göre Lund'a göre toplam çalışan sayısının tahmini;

$$\hat{Y}_L = 4 * 4892 + (0,189 * 4 * 294 + 0,811 * 4 * 321) * 58,029 + 4 * 658$$

$$\hat{Y}_L = 95524,748 \cong 95525$$

olarak bulunur.

Bu tahmin edicinin varyansı ise;

$$V(\hat{Y}_L) = \frac{4695^2}{1174} * 292,381 * (1 - 0,296) + \frac{4965 * 1467 * 0,296 * 0,95}{0,296 * 1174 + 0,95 * 367} * 32534,89$$

$$+ \frac{1467^2}{367} * 2451,644 * (1 - 0,95)$$

$$+ \frac{4695^2 * (1 - 0,296) * 0,296}{1174} * (4,568 - 0,189 * 57,117)^2$$

$$+ \frac{1174^2 * (1 - 0,95) * 0,95}{367} * (16,164 - 0,811 * 57,117)^2$$

$$V(\hat{Y}_L) = 95504550,94$$

olarak bulunur.

Değişim katsayısı ise aşağıdaki şekilde hesaplanır:

$$DK_L = \frac{\sqrt{V(\hat{Y}_L)}}{\hat{Y}_L} = \frac{\sqrt{95504550,94}}{95525} = 0,102$$

olarak bulunur.

Fuller ve Burmeister' a göre yığın toplamın tahmin edicisi;

$$\hat{Y}_{FB} = (N_A - \hat{N}_{ab,FB}) \bar{y}_a + \hat{N}_{ab,FB} \bar{y}_{ab} + (N_B - \hat{N}_{ab,FB}) \bar{y}_b$$

ile hesaplanır ve

$$p_{FB} = \frac{n'_{ab}(1 - f_B)}{n'_{ab}(1 - f_B) + n''_{ab}(1 - f_A)}$$

değeri;

$$\bar{y}_{ab} = p_{FB} \bar{y}'_{ab} + (1 - p_{FB}) \bar{y}''_{ab}$$

ifadesinde yerine yazılır.

Bu tahmin edici hesaplanmadan önce aşağıda verilen eşitlikten N_{ab} 'nin tahmin edilmesi gerekmektedir.

$$g_A = \frac{N_A - n_A}{N_A - 1} = \frac{4695 - 1174}{4695 - 1} = 0,75 \quad g_B = \frac{N_B - n_b}{N_B - 1} = \frac{1467 - 367}{1467 - 1} = 0,75$$

olmak üzere;

$$[n_A g_B + n_B g_A] \hat{N}_{ab,FB}^2 - [n_A N_B g_B + n_B N_A g_A + n'_{ab} N_A g_B + n''_{ab} N_B g_A] \hat{N}_{ab,FB} + [n'_{ab} g_B + n''_{ab} g_A] N_A N_B = 0$$

$$[1174 * 0,75 + 367 * 0,75] \hat{N}_{ab,FB}^2 \\ - [1174 * 1467 * 0,75 + 367 * 4695 * 0,75 + 294 * 4695 * 0,75 + 321 * 1467 * 0,75]$$

$$\hat{N}_{ab,FB} + [294 * 0,75 + 321 * 0,75] 4695 * 1467 = 0 \\ 1155,75 \hat{N}_{ab,FB}^2 - 3972420 \hat{N}_{ab,FB} + 3176889356 = 0$$

$$\hat{N}_{ab,FB}^1 = 1266,18 \cong 1266$$

$$\hat{N}_{ab,FB}^2 = 2170,91 \cong 2171$$

değerleri bulunur. İkinci değer seçilemez çünkü bu değer örneğe aldığımız toplam birim sayısından daha büyük bir sayıdır. Yapılan işlemler sonucu N_A tahmini,

$$\hat{N}_{a,FB} = N_A - \hat{N}_{ab,FB} = 4695 - 1266 = 3429$$

N_B tahmini,

$$\hat{N}_{b,FB} = N_A - \hat{N}_{ab,FB} = 1467 - 1266 = 201$$

N' 'nin tahmini ise,

$$\hat{N}_{FB} = N_A + N_B - \hat{N}_{ab,FB} = 4695 + 1467 - 1266 = 4896$$

olarak bulunur.

$$p_{FB} = \frac{294 * (1 - 0,25)}{294(1 - 0,25) + 321(1 - 0,25)} = 0,478$$

$$\bar{y}_{ab} = 0,478 * 60,894 + 0,528 * 55,404$$

$$\bar{y}_{ab} = 58,36$$

Fuller ve Burmeister'a göre toplam çalışan sayısı;

$$\hat{Y}_{FB} = (4695 - 1266) * 5,559 + 1266 * 58,36 + (1467 - 1266) * 14,304$$

$$\hat{Y}_{FB} = 95820,675 = 95821$$

olarak bulunur.

Bu tahmin edicinin varyansı ise;

$$\begin{aligned}
 Var(\hat{Y}_{FB}) &= N_a(f_A^{-1} - 1)S_a^2 + [(1 - f_B)f_A + (1 - f_A)f_B]^{-1} \\
 &\quad (1 - f_A)(1 - f_B)N_{ab}S_{ab}^2 + N_b(f_B^{-1} - 1)S_b^2 \\
 &\quad + (\bar{Y}_{ab} - \bar{Y}_a - \bar{Y}_b)^2 \frac{N_{ab}N_aN_bg_Ag_B}{n_A N_b g_B + n_B N_a g_A} \\
 Var(\hat{Y}_{FB}) &= 3301 * (4 - 1) * 292,469 + [(1 - 0,25) * 0,25 + (1 - 0,25) * 0,25]^{-1} \\
 &\quad (1 - 0,25) * (1 - 0,25) * 1394 * 32558,243 + 73 * (4 - 1) * 2485,694 \\
 &\quad + (57,114 - 4,568 - 16,164)^2 * \frac{1394*3301*73*0,75*0,75}{1174*73*0,75+367*3301*0,75} \\
 Var(\hat{Y}_{FB}) &= 71777054,02
 \end{aligned}$$

olarak bulunur.

Değişim katsayısı ise aşağıdaki şekilde hesaplanır.

$$DK_{FB} = \frac{\sqrt{V(\hat{Y}_{FB})}}{\hat{Y}_{FB}} = \frac{\sqrt{71777054,2}}{95821} = 0,088$$

olarak bulunur.

Tek Çerçeve (SF) tahmin edicisine göre yığın toplamı;

$$y_{ab} = y'_{ab} + y''_{ab} = 17903 + 17785 = 3568$$

$$\hat{Y}_{SF} = f_A^{-1}y_a + f_B^{-1}y_b + (f_A + f_B)^{-1}y_{ab}$$

$$\hat{Y}_{SF} = 4 * 4892 + 4 * 658 + 2 * 35688$$

$$\hat{Y}_{SF} = 93576$$

olarak bulunur.

N_A ve N_B 'nin bilindiği N_{ab} 'nin bilinmediği durumda, değişim katsayılarına ait sıralama $DK_{FB} < DK_L < DK_H$ şeklindedir.

Skinner(1991)' e göre $\hat{Y}_{SF}$, $\hat{Y}_H$ ya da $\hat{Y}_L$ ' nin özel bir hali olduğu için $Var(\hat{Y}_{FB}) \leq Var(\hat{Y}_{SF})$ 'dir. Bundan dolayı da $DK_{FB} < DK_{SF}$ 'dir.

Skinner'in Sıralama Oranı tahmin edicisine göre yığın toplamı;

$$\hat{Y}_{RR}^{(\infty)} = (N_A - \hat{N}_{ab,RR}) \bar{y}_a + \hat{N}_{ab,RR} \bar{y}_{ab} + (N_B - \hat{N}_{ab,RR}) \bar{y}_b$$

ile hesaplanır.

Bu tahmin edici hesaplanmadan önce aşağıda verilen eşitlikten N_{ab} 'nin tahmin edilmesi gerekmektedir.

$$\begin{aligned} n_{ab} \hat{N}_{ab,RR}^2 - [n_{ab}(N_A + N_B) + (f_A^{-1} + f_B^{-1})n_a n_b] \hat{N}_{ab,RR} + n_{ab} N_A N_B &= 0 \\ 615 \hat{N}_{ab,RR}^2 - [615 * (4695 + 1467) + (4 + 4) * 880 * 46] \hat{N}_{ab,RR} \\ + 615 * 4695 * 1467 &= 0 \\ 615 \hat{N}_{ab,RR}^2 - 4113470 \hat{N}_{ab,RR} + 4235852475 &= 0 \end{aligned}$$

$$\hat{N}_{ab,RR}^1 = 1271,44 \cong 1271$$

$$\hat{N}_{ab,RR}^2 = 5417,12 \cong 5417$$

değerleri bulunur. İkinci değer seçilemez çünkü bu değer örneğe aldığımız toplam birim sayısından daha büyük bir sayıdır.

Yapılan işlemler sonucu Skinner'a göre çalışan sayısının toplamı;

$$\hat{P}_{RR}^{(\infty)} = (4695 - 1271) * 5,559 + 1271 * 58,029 + (1467 - 1271) * 14,304$$

$$\hat{P}_{RR}^{(\infty)} = 95592,459 \cong 95592$$

olarak bulunur.

Bu tahmin edicinin varyansı ise;

$$Var(\hat{Y}_{RR}^{(\infty)}) \sim f_A^{-1} N_a S_a^2 + f_B^{-1} N_b S_b^2 + (f_A + f_B)^{-1} N_{ab} S_{ab}^2 \\ + (\bar{Y}_{ab} - \bar{Y}_a - \bar{Y}_b)^2 \frac{N_{ab} N_a N_b}{n_A N_b + n_B N_a} (1 + \lambda^2)$$

formülü ile hesaplanır ve

$$\lambda^2 = \frac{N_a N_b N_{ab}^2 (f_A^2 N_A - f_B^2 N_B)^2}{N_A N_B f_A f_B (f_A + f_B)^2 (N_{ab}^2 - N_A N_B)^2}$$

'dir.

Buna göre;

$$\lambda^2 = \frac{3301 * 73 * 1394^2 (0,25^2 * 4695 - 0,25^2 * 1467)^2}{4695 * 1467 * 0,25 * 0,25 (0,25 + 0,25) * (1394^2 - 4695 * 1467)^2}$$

$$\lambda^2 = 0,003622$$

ve

$$Var(\hat{Y}_{RR}^{(\infty)}) = 4 * 3301 * 292,469 + 4 * 73 * 2485,694 \\ + (0,25 + 0,25)^{-1} 1394 * 32534,89 \\ + (57,114 - 4,568 - 16,164)^2 * \frac{1394 * 3301 * 73}{1174 * 73 + 367 * 3301} * (1 + 0,06^2)$$

$$Var(\hat{Y}_{RR}^{(\infty)}) = 95638864,522$$

olarak bulunur.

Değişim katsayısı ise aşağıdaki şekilde hesaplanır:

$$DK_{RR} = \frac{\sqrt{V(\hat{Y}_{RR})}}{\hat{Y}_{RR}} = \frac{\sqrt{95638864,52}}{95592} = 0,102$$

olarak bulunur.

Çalışan sayısı toplamı, varyansı ve değişim katsayısı her iki durum için Çizelge 5.1 ve Çizelge 5.2’de gösterilmiştir.

Çizelge 5.1. N_A , N_B ve N_{ab} bilindiğinde toplam çalışan sayısı, varyansı ve değişim katsayısı

Hartley’in Tahmin Edicisi		
$\hat{Y}_H=100454$	$Var(\hat{Y}_H)=95099696,32$	$DK_H=0,097$
Lund’un Tahmin Edicisi		
$\hat{Y}_L=100287$	$Var(\hat{Y}_L)=95099532,91$	$DK_L=0,0972$

Çizelge 5.1’ de görüldüğü üzere Hartley’in tahmin edicisi daha küçük değişim katsayısına sahiptir. Bu yüzden N_A , N_B ve N_{ab} bilindiğinde Hartley’ in tahmin edicisini kullanmak daha iyi bir çözüm olarak düşünülebilir.

Çizelge 5.2. N_A ve N_B bilindiğinde, N_{ab} bilinmediğinde toplam çalışan sayısı, varyansı ve değişim katsayısı

Hartley’in Tahmin Edicisi		
$\hat{Y}_H=93588$	$Var(\hat{Y}_H)=97394059,9$	$DK_H=0,105$
Lund’un Tahmin Edicisi		
$\hat{Y}_L=95525$	$Var(\hat{Y}_L)=95504550,94$	$DK_L=0,102$
Fuller ve Burmeister’in Tahmin Edicisi		
$\hat{Y}_{FB}=95821$	$Var(\hat{Y}_{FB})=71777054,2$	$DK_{FB}=0,088$
Tek Çerçeve (Single Frame) Tahmin Edicisi		
$\hat{Y}_{SF}=93576$	$Var(\hat{Y}_{SF}) > Var(\hat{Y}_{FB})$	$DK_{SF} > DK_{FB}$
Skinner’in Sıralama Oran Tahmin Edicisi		
$\hat{Y}_{RR}=95592$	$Var(\hat{Y}_{RR})=95638864,52$	$DK_{RR}=0,102$

* $\hat{Y}_{SF}$, $\hat{Y}_H$ ya da $\hat{Y}_L$ ’ nin özel bir hali olduğu için $Var(\hat{Y}_{FB}) \leq Var(\hat{Y}_{SF})$ ’ dir (Skinner, 1991).

Çizelge 5.2’ de görüldüğü üzere N_A ve N_B bilindiğinde, N_{ab} bilinmediğinde Fuller ve Burmeister tahmin edicisini kullanmak daha iyi bir çözümdür. Çünkü en küçük değişim katsayısına sahip olan tahmin edici Fuller ve Burmeister’in tahmin edicisidir.

6. SONUÇ VE ÖNERİLER

Yapılacak olan her türlü örnekleme dayalı araştırmalar için çerçevenin önemi büyüktür. Çerçeve olmaksızın yapılan araştırmalar hatalı ve yanlış olur. Aynı konu üzerinde birden fazla çerçeve varsa, bu çerçeveleri birleştirmek çok fazla maliyetli olabilir ve mümkün olmayabilir. Bu çerçeveleri birleştirmek yerine çoklu çerçeve yöntemi kullanılabilir.

Çoklu çerçeve örnekleme, özellikle iki veya ikiden fazla çerçevesi bulunan konu veya alanlarda, bu çerçevelerin birleştirilmesinin mümkün olmadığı durumlar veya bu çerçevelerin birleştirilmesinden dolayı maliyetin fazla olduğu durumlarda, güvenilirliği arttırmak amacıyla kullanılan yöntemlerden biridir. Çoklu çerçeve örnekleme, birkaç çerçevenin ortak kapsamı hakkında herhangi bir ön bilgi olmaksızın, bu çerçevelerin kullanılabilmesine olanak sağlar.

Çoklu çerçeve araştırmalarının etkinliği daha da geliştirebilir, ama bunu yaparken dikkatli olmak gerekmektedir. Kullanılan çerçevelerin sayısı fazla olmamalıdır, eğer bu sayı artarsa her bir alan için çektiğimiz örneklem boyutları da küçük olacaktır. Çerçeve sayısı arttıkça zaman ona bağlı olarak alan sayısı da artacaktır. O zaman da farklı çerçevelerde yer alan birimleri eşleştirmek daha uzun sürecektir ve bazı veriler kapsam dışı kalacaktır.

Literatürdeki makaleler incelendiğinde, farklı tahmin edicilerin çeşitli varsayımlar altında önerildiği görülmektedir. Bu varsayımlar, ilgilendiğimiz yığındaki her birimin çerçevelerden en az birine ait olması ve örneklenen her bir birimin, diğer çerçevelerin bir veya birden fazlasına ait olup olmamasıdır. Kesişen çerçevelerde tahmin yapmak için kullanılan bütün yöntemler bu tezde incelenmiştir. Hartley (1962 ve 1974), Lund (1968), Fuller ve Burmeister (1972), alan büyüklüğü bilinmediği zaman toplam yığın için en uygun tahmin edicileri önermişlerdir. Fuller ve Burmeister tahmin edicisi karşılaştırma yapılan tahmin ediciler arasında literatürde de belirtildiği gibi en iyi çözümü vermiştir. Ayrıca Dalçık (2010)'ta yaptığı simülasyon çalışmasında, çektiği örneğe bağlı olarak Fuller ve Burmeister tahmin edicisinin en uygun sonucu verdiğini göstermiştir.

Bu çalışmanın amacı Bilim, Sanayi ve Teknoloji Bakanlığı ile Sosyal Güvenlik Kurumu çalışan sayısı verilerinin hiçbir veri kaybı olmadan, çerçeveler birleştirilmeden, iki liste halinde toplam çalışan sayısını tahmin etmektir. Çoklu çerçeve örnekleme kullanılarak yapılan analiz sonucunda değişim katsayısı küçük olan tahmin edici, en iyi toplam çalışan sayısı tahminini vermiştir.

Tezde, seçilen ile ait Girişimci Bilgi Sistemi çalışan sayısı verileri kullanılarak çoklu çerçeve örnekleme yapılmıştır. İki kurum verileri arasında yapılan karşılaştırma sonucunda, bazı verilerin kesişim halinde olduğu görülmüştür. Bu kesişim belirlenirken firmaların vergi kimlik numaraları baz alınmıştır. Numaralar gizli veri olduğu için sonuçlar toplulaştırılmış bir şekilde tezde sunulmuştur. Eldeki firma verilerinden çok fazla kayıp olmadan ve bu iki çerçeve tek liste haline getirilmeden, iki ayrı liste şeklinde analiz yapılmıştır. Kesişen veriler listelerden çıkartılmamıştır. Zaten çoklu çerçeve yönteminin en önemli özelliği de kesişen verilerin listelerden çıkartılmayarak analizin yapılabilmesidir.

Yapılan analiz sonucunda N_A , N_B ve N_{ab} bilindiğinde ve N_A , N_B bilindiğinde, N_{ab} bilinmediğinde çekilen örneklerden toplam çalışan sayısı bu çalışan sayısına ait varyans ve değişim katsayısı hesaplanmıştır. Birinci durumda Hartley'in tahmin edicisi, değişim katsayısı için daha küçük dolayısıyla daha iyi bir sonuç verirken, ikinci durumda ise, literatürdeki çalışmalarda olduğu gibi, Fuller ve Burmeister tahmin edicisi daha küçük ve dolayısıyla daha iyi bir sonuç vermiştir.

Ülkemizde bir sanayi işletmesinin birçok kurumda verisi bulunmaktadır. Farklı amaçlar doğrultusunda kullanılmasına rağmen bazı zamanlarda bu durum sıkıntıya yol açmaktadır. Bilimsel teoriler bu verilerin hangilerinin daha sağlıklı olduğunun araştırılmasında önemli katkı sağlamaktadır. Ancak verilerin tek bir çatı altında toplanması hem kayıt dışılığın önlenmesi noktasında merkezi idareyi güçlendirecektir hem de firmaları aynı konuda birden fazla bildirim yapma zorunluluğundan kaynaklanan iş yükünden kurtaracaktır.

KAYNAKLAR

- Ayikuteli, T. (2011), *Çoklu Çerçeve Kullanımı*, Uzmanlık Tezi, Türkiye İstatistik Kurumu, Ankara, 3-27.
- Bankier, M.D. (1986), Estimators Based on Several Stratified Samples with Applications to Multiple Frame Surveys, *Journal of the American Statistical Association*, (81), 1074-1079.
- Dalçık, A. (2010), *Multiple Frame Sampling Theory and Applicaitons*. Y. Lisans tezi, ODTÜ Fen Bilimleri Enstitüsü
- Fuller, W.A., and Burmeister, L.F. (1972), Estimators for Sample Selected From Two Overlapping Frames, *Proceedings of the Social Statistics Section*, American Statistical Association, pp. 245-249.
- Hartley, H.O. (1962), Multiple Frame Surveys, *Proceedings of the Social Statistics Section*, American Statistical Association, pp. 203-206.
- Hartley, H.O. (1974), Multiple Frame Methodology and Selected Applications, *Sankhya*, Series C, (36), 99-118.
- Iachan, R., and Dennis, M.L. (1993), A multiple frame approach to sampling the homeless and transient population. *Journal of Official Statistics*, 9(4), 747-764.
- Kalton, G., and, Anderson, D.W. (1986), Sampling Rare Population, *Journal of the Royal Statistical Society*, Series A, (149), 65-82.
- Kish, L. (1965), *Survey Sampling*, New York: John Wiley and Sons.
- Lund, R.E. (1968), Estimators in Multiple Frame Surveys, *Proceedings of the Social Statistics Section*, American Statistical Association, pp. 282-288
- Lohr, S.L., and Rao, J.N.K. (2000), Inference in Dual Frame Surveys, *Journal of the American Statistical Association*, (95), 271-280.
- Lohr, S.L., and Rao, J.N.K. (2006), Estimation in Multiple Frame Surveys, *Journal of the American Statistical Association*, (101), 1019-1030.
- Rao, J.N.K., and Skinner, C.J.(1996), Estimation in Dual Frame Surveys With Complex Designs, *Proceedings of the Survey Methods Section*, Statistical Society of Canada, pp 63-68.
- Şimşek, S. (2002), *Çoklu Çerçeve Yöntemiyle İşyerleri Örneklemesi ve Tahmin Edicileri*, Uzmanlık Tezi, Devlet İstatistik Enstitüsü, Ankara, 10-23.
- Skinner, C.J. (1991), On the efficiency of Raking Ratio Estimation for Multiple Frame Surveys, *Journal of the American Statistical Association*, (86), 779-784.

Skinner , C.J., and Rao, J.N.K. (1996), Estimation in Dual Frame Surveys with Complex Design, *Journal of the American Statistical Association*, (91), 349-356.

İnternet: Bilim, Sanayi ve Teknoloji Bakanlığı. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fgbs.sanayi.gov.tr%2F&date=2014-12-30>. Son Erişim Tarihi: 15.10.2014

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : ÜSTÜN, Mehmet Oğuzhan
Uyruğu : T.C.
Doğum Tarihi ve Yeri : 27.01.1989 Ankara
Medeni Hali : Bekar
Telefon : (0536) 236 8924
e-mail : mehmetoguzhanustun@gmail.com



Eğitim Derecesi

Okul/Program

Mezuniyet Yılı

Yüksek Lisans

Gazi Üniversitesi/ İstatistik Bölümü

2014

Lisans

Gazi Üniversitesi/ İstatistik Bölümü

2010

Lise

Batıkent Y.D.A Lisesi

2006

İş Deneyimi, Yıl

Çalıştığı Yer

Görev

2013- Devam ediyor

Bilim, Sanayi ve Teknoloji
Bakanlığı

ST Uzm.Yrd.

Yabancı Dil

İngilizce



GAZİ GELECEKTİR..