KAYIP GÖZLEM OLMASI DURUMUNDA KİTLE ORTALAMASI TAHMİNİ

ESTIMATION OF THE POPULATION MEAN IN THE CASE OF MISSING OBSERVATIONS

ESRA SATICI

Hacettepe Üniversitesi

Lisansüstü Eğitim – Öğretim ve Sınav Yönetmeliğinin

İSTATİSTİK Anabilim Dalı İçin Öngördüğü

DOKTORA TEZİ

olarak hazırlanmıştır.

2009

Fen Bilimleri Enstitüsü Müdürlüğü'ne,

Bu çalışma jürimiz tarafından **İSTATİSTİK ANABİLİM DALI 'nda DOKTORA TEZİ** olarak kabul edilmiştir.

Başkan	:…..................................... Prof. Dr. Öztaş AYHAN
Üye (Danışman)	:.…..................................... Doç. Dr. Cem KADILAR
Üye	:…..................................... Prof. Dr. Hülya ÇINGI
Üye	:…..................................... Doç. Dr. Meral ÇETİN
Üye	:…..................................... Yrd. Doç. Dr. Yaprak Arzu ÖZDEMİR

ONAY

Bu tez ...../...../2009 tarihinde Enstitü Yönetim Kurulunca kabul edilmiştir.

Prof.Dr Erdem YAZGAN
Fen Bilimleri Enstitüsü Müdürü

KAYIP GÖZLEM OLMASI DURUMUNDA KİTLE ORTALAMASI TAHMİNİ

Esra SATICI

Öz

Bu çalışmada, kayıp veri analiz yöntemleri ele alınmış, günümüze kadar literatürde yer alan kayıp veri olması durumunda kitle ortalaması tahmini için önerilen tahmin ediciler incelenmiştir. Bu tahmin edicilerin yan ve hata kareler ortalamaları çıkarımı yapılmıştır. İlgili tahmin ediciler birbiriyle karşılaştırılmış ve etkinlikleri incelenmiştir. Ayrıca, ard arda örnekleme yöntemi tanıtıldıktan sonra, ard arda örnekleme yönteminde kayıp veri olması durumunda kitle ortalaması tahmin edicileri de incelenmiştir.

Çalışmanın Beşinci Bölümünde, ard arda örnekleme yönteminde kayıp gözlem olması durumunda son araştırma kitle ortalaması tahmini için yeni bir tahmin edici önerilmiştir. Bu tahmin edicinin minimum hata kareler ortalaması ve optimal yenileme ilkesi kapsamında minimum eşleştirme oranı elde edilmiştir. Önerilen tahmin edici, Singh ve Priyanka (2007) tahmin edicisi ile karşılaştırılmış ve etkin olma koşulu bulunmuştur.

Sayısal örnekte, Çıngı vd. (2007) TÜBİTAK projesinde kullanılan Milli Eğitim Bakanlığı verileri kullanılmıştır. Öğretmen ve dershane sayıları yardımcı değişken, orta öğretim kurumları öğrenci seçme sınavını kazanan öğrenci sayıları ilgilenilen değişken olmak üzere, önerilen ve Singh ve Priyanka (2007) tahmin edicilerine göre kitle ortalaması tahmin edilmiş, farklı cevapsızlık oranı (W), ilişki katsayısı, ikinci araştırmada cevapsızlardan çekilen alt örneklem oranı (k) ve yeni örneklem oranı (μ) için hata kareler ortalamaları hesaplanmıştır.

Çalışmanın son bölümünde ise, sayısal örnekte elde edilen sonuçlara bağlı olarak önerilen tahmin edicinin etkinlikleri incelenmiş ve yorumlanmıştır.

ANAHTAR KELİMELER: Kayıp veri, ard arda örnekleme yöntemi, oransal tahmin edici, yan, hata kareler ortalaması, optimal yenileme ilkesi.

Danışman: Doç.Dr. Cem KADILAR, Hacettepe Üniversitesi, İstatistik Bölümü.

ESTIMATION OF THE POPULATION MEAN IN THE CASE OF MISSING OBSERVATIONS

Esra SATICI

Abstract

In this study, various kinds of estimators of the population mean in the case of missing data are investigated. Bias and mean square errors of these estimators are obtained. These estimators are compared with each other and efficient conditions are examined. Furthermore, after introducing the successive sampling method, estimators of the population mean in successive sampling are presented in the case of missing data.

In the fifth section of this study, a new estimator for the population mean of current occasion on successive sampling is proposed when there are missing data in the data set. Minimum mean square error of this proposed estimator and minimum fraction of matching based on optimal replacement policy are obtained. This estimator is compared with the estimator suggested by Singh and Priyanka (2007) and the efficient condition for the proposed estimator is found.

In numerical example, the data of Ministry of National Education compiled in a TUBITAK's project by Çıngı *et al.* (2007) are used. Considering number of teachers and private teaching institutions as the auxiliary variables, the population mean of number of successful students in the student selection examination for the secondary schools is estimated using the proposed estimator and the Singh-Priyanka estimator. The mean square errors of these estimators are also computed according to various values for the fraction of missing data (W), coefficient of correlation, fraction of subsampling (k) and fraction of matching (μ).

In the last section of the study, according to the results obtained in numerical example, the efficient of proposed estimator is examined and discussed.

KEYWORDS: Missing data, successive sampling, ratio estimator, bias, mean square error, optimal replacement policy.

Advisor: Assoc. Prof. Dr. Cem KADILAR, Hacettepe University, Department of Statistics.

TEŞEKKÜR

Tez çalışmam süresince, değerli yorumlarıyla, sabrıyla çalışmama yön veren, beraber çalışmaktan büyük keyif aldığım, kendisinden çok şey öğrendiğim, çok sevdiğim ve saydığım değerli danışmanım Doç.Dr. Cem KADILAR'a,

Değerlendirmeleriyle, eleştirileriyle ve tecrübeleriyle çalışmama büyük katkıları bulunan, tez komite üyelerim Prof.Dr. Hülya ÇINGI ve Prof.Dr. Öztaş AYHAN'a,

Doktora yapmamda çok büyük payı olan, bana büyük destek veren, hiçbir zaman hakkını ödeyemeyeceğim, çok güvendiğim değerli hocam Prof.Dr. Süleyman GÜNAY'a,

Bana, özleyeceğim bu güzel ve huzurlu çalışma ortamını sağlayan, her anlamda her zaman destek olan başta Dr.Ayten YİĞİTER, Doç.Dr. Durdu SERTKAYA KARASOY, Nilgün ÖZGÜL, Doç.Dr. Meral CANDAN ÇETİN, Yrd.Doç.Dr. Özge UÇAR, Dr. Özlem SUNAR ve Dr. Serpil AKTAŞ ALTUNAY olmak üzere tüm H.Ü. İstatistik Bölümü ailesine,

Desteklerinden ötürü, Hacettepe Üniversitesi Bilimsel Araştırmalar Birimi'ne,

Ve en önemlisi, bana her zaman güvenen, hep daha iyisini yapmam için yön gösteren, hep yanımda olan, hayatımı kolaylaştıran canım annem, babam, kardeşim, sevgili hayat arkadaşım ve bana şans getiren biricik sevimli oğlum Çağan'a çok teşekkür ederim.

İÇİNDEKİLER DİZİNİ

İÇİNDEKİLER DİZİNİ (devam ediyor)

Sayfa

İÇİNDEKİLER DİZİNİ (devam ediyor)

Sayfa

ŞEKİLLER DİZİNİ

Sayfa

ÇİZELGELER DİZİNİ

Sayfa

BİRİNCİ BÖLÜM

1. GİRİŞ

Günümüzde pek çok alanda alınması gereken önlemlerin doğruluğunun araştırılması, bu önlemlerin geliştirilmesi ve yerinde karar verilebilmesi istatistik verilerine bağlıdır. İstatistik verilerinin elde edilebilmesinde, hakkında bilgi edinilmek istenilen kitlenin tamamının gözlenmesi yerine bu kitleden uygun yöntemlerle seçilmiş örneklemlerden faydalanılabilinir. İstatistiksel çözümleme, istatistiksel çıkarsama, klasik karar işlemleri ya da tümevarımsal istatistik gibi farklı isimlerle bilinen istatistiksel yöntemler, örneklem adı verilen sınırlı sayıda birime dayanarak kitle için sonuç çıkarma olanağı verir. Planlanan bir araştırmayla ilgili veriler, haberleşme araçlarıyla ya da eğitimli anketörler aracılığıyla elde edilebilir. Veri toplama sırasında karşılaşılabilecek cevapsız bırakma ya da örneklem birimine ulaşamama durumlarında, yani kayıp veri olması durumunda klasik istatistiksel yöntemler uygulanamamaktadır.

Kayıp veri 1970'lerden bu yana tartışılan bir konu olup ve hemen hemen bütün araştırma çalışmalarının ortak problemidir. Çoğunlukla, araştırmalarda örnekleme ilişkin birimlerden elde edilen bilgiler toplanır. Ancak nedenlerin çeşitliliğinden dolayı, birimlerin küçük bir bölümü için, ya hiçbir bilgi ya da bir veya birden fazla değişken bilgisi elde edilemez. Kayıp veri çözümü için, çeşitli alternatif yaklaşımlar geliştirilmesine rağmen, birçok çalışmada kayıp veriler silinmekte ve geriye kalan eksiksiz gözlemler üzerinden klasik istatistiksel yöntemler uygulanmaktadır. Bu durum sadece önemli bilgilerin kaybına değil, aynı zamanda örneklem büyüklüğünün küçülmesine neden olduğundan tahminlerin güvenilirliğinin azalmasına, yan ve varyansın artmasına neden olmaktadır.

Kayıp veri olması durumunda kitle ortalaması tahmini ile ilgili literatürde yer alan çalışmaları, yardımcı değişken kullanma ve kullanmama, yardımcı değişkende de kayıp veri olması ya da olmaması şeklinde sınıflandırmak mümkündür. Bu sınıflandırmada yardımcı değişken bilgisi kullanılmadan kitle ortalaması tahminine ilişkin çalışmaya Gamrot (2007) örnek verilebilir. Singh ve Horn (2000), Singh ve Deo (2003) ve Rueda vd. (2005) tarafından yapılan çalışmalar ise, kayıp veri olması durumunda yardımcı değişken bilgisi kullanılarak kitle ortalaması tahmin edicisi elde etmeye yönelik çalışmalardan bazılarıdır. Fakat Singh ve Horn (2000), Singh ve Deo (2003) yardımcı değişkende kayıp veri olmadığını varsayarken, Rueda vd. (2005) yardımcı değişkende de kayıp veri olması durumunu dikkate almıştır.

Birçok araştırmada, aynı kitle ve aynı değişkenler ile ilgili çalışmalar tekrarlanabilir. Örneğin, birçok ülkede, işgücü anketleri çalışan ve çalışmayan oranını belirlemek amacıyla her ay tekrarlanır. Aynı şekilde tüketici fiyat endeksi hesaplamaları için her ay ürün fiyatları incelenmektedir. Eğer kitle birimleri, takip eden araştırmalarda değişmiyor ise, önceki araştırma verileri son araştırma kitle parametre tahminleri için kullanılabilir. Önceki araştırma verilerinden seçilen alt örneklemin, sonraki araştırmada kullanılması ard arda örnekleme yöntemi olarak adlandırılmaktadır. Önceki araştırma verilerini kullanmak maliyeti azaltma ve tahmin edicinin doğruluğunu arttırma gibi yararlar sağlarken, örneklemde aynı birimlerdeki kayıplılık durumu başarısızlığa ve tahmin edicinin etkinliğinin azalmasına neden olmaktadır.

Ard arda örnekleme yöntemi, Jessen'in tarım alanında, daha önceki araştırma bilgilerini kullanmasıyla çalışılmaya başlanmıştır (Singh, 2005). Daha sonra teori, Rao ve Graham (1964), Gupta (1979), Das (1982) ve diğer bazı araştırmacılar tarafından geliştirilmiştir. Bilindiği gibi, örnekleme teorisinde yardımcı değişken bilgisinden sıklıkla yararlanılmaktadır. Dolayısıyla ard arda örnekleme yönteminde de yardımcı değişken bilgisini kullanan çeşitli çalışmalar yapılmıştır. Sen (1971), güncel araştırmada kitle ortalaması tahmini için iki yardımcı değişken bilgisinden yararlanmıştır. Daha sonra Sen (1973), bu çalışmasını daha fazla yardımcı değişken için genişletmiştir. Singh ve Singh (2001) son araştırmada da yardımcı değişken bilgisini kullanarak ard arda örnekleme yönteminde, ikinci araştırma kitle ortalaması tahmini için yeni bir çalışma sunmuşlardır. Singh ve Deo (2003) bu çalışmayı ard arda yapılan h araştırma için genişletmiştir. Singh (2005) ard arda iki araştırmada seçilen örneklemlerden, son araştırmanın kitle ortalamasını tahmin edebilmek için her iki araştırmanın yardımcı değişken bilgisinden yararlanarak, zincirleme-oransal tahmin edici sunmuştur. Buna benzer biçimde, Singh ve Priyanka (2008) son araştırma kitle ortalaması tahmini için zincirleme fark ve regresyon tipi tahmin edicileri sunmuşlardır. Aynı zamanda Singh ve Priyanka (2007) kayıp veri olduğu durum için ard arda örnekleme yönteminden yararlanarak son araştırma kitle ortalaması için yeni bir tahmin edici sunmuşlar ve önerilen tahmin edicinin minimum hata kareler ortalamasını optimum yenileme oranına bağlı olarak elde etmişlerdir. Kayıp veri olduğu durumda, bu çalışmalara paralel, ard arda örnekleme yöntemi için kitle ortalaması tahminine ilişkin çok fazla bir çalışmaya rastlanamamaktadır.

Bu tez çalışmasının amacı, araştırmalarda karşılaşılan kayıp veri problemine dikkat çekmek ve kayıp veri olduğu durumda ard arda örnekleme yöntemi için yardımcı değişken bilgisinden faydalanarak, incelenen tahmin edicilerden daha etkin yeni bir tahmin edici sunmaktır.

Bu çalışmada, kayıp veri analiz yöntemleri ele alınmış, günümüze kadar literatürde yer alan kayıp veri olması durumunda kitle ortalaması tahmini için önerilen tahmin ediciler incelenmiştir. Ard arda örnekleme yöntemi irdelendikten sonra ard arda örnekleme yönteminde kayıp veri olması durumunda, son araştırma kitle ortalaması tahmini için yeni bir oransal tahmin edici önerilerek bu tahmin edicinin hata kareler ortalaması bulunmuş ve etkinliği sınanmıştır. Son olarak, mevcut verilerle yapılan bir uygulama sonucunda teoride elde edilen sonuçlar sayısal örnek ile desteklenmiştir.

İKİNCİ BÖLÜM

2. GENEL BİLGİLER

En sade biçimiyle kayıp veri, toplanması planlanan veri kümesi ile gerçekte elde edilen veri kümesi arasındaki fark olarak tanımlanır. Kayıp veri, alan araştırmalarında sıklıkla karşılaşılan bir problemdir. Bunun nedenleri kaynakların (anketin tamamlanması için ayrılan bütçe, çalışan sayısı, zaman) çoğu zaman anket kapsamındaki kişilerin tümüyle görüşmek için yetersiz kalması, örneklem birimlerinin bazen araştırmaya katılmaması ya da anketteki bütün soruları cevaplamayı reddetmesi, araştırmacının bütün örneklem birimleri ile iletişime geçememesi veya kazara bilgilerin kaybedilebilmesidir.

Belirtilen koşullara göre seçilen örneklemde, kayıp veriler nedeniyle örneklem büyüklüğünde bir azalma olmaktadır. Ayrıca kayıp verilerin temsil ettiği bilgi nedeniyle, örneklemin, çekildiği kitleyi artık tam anlamıyla temsil ettiği söylenemez. Örneğin, gelir durumu ile ilgili bir araştırmada, cevapsızlar cevaplayıcılara göre daha varlıklı kişiler ise ve analizler sadece cevaplayanlar üzerinden yapıldıysa, sonuçlar bozuluma uğramış demektir.

Bütün anket çalışmalarında, hem örneklem biriminde (sample unit) hem de değişkenlerinde (item) cevapsızlık olması mümkündür. Bu durumda kayıp veriyi, aşağıda verilen iki başlık altında sınıflandırabiliriz:

- Birim cevapsızlığı, bir ya da birden fazla kişi ile görüşülememesi durumudur. Görüşmenin yapılamaması, kişinin belirtilen adreste bulunamaması, görüşmeyi reddetmesi gibi nedenlere bağlı olabilir.

- Değişken cevapsızlığı ise, cevaplayıcının bir ya da birden fazla soruyu cevaplamamasından kaynaklanmaktadır. Soru ile ilgili bir yorumunun olmaması ya da ilgili soruyu cevaplamaya gönülsüz olması buna neden olabilir. Anket formunun düzgün hazırlanması, soruların titiz seçilmesi, en başta cevapsızlık oranının azaltılmasında önemli bir rol oynamaktadır.

Ayrıntılı cevapsızlık (veya cevap) tanımlaması için, cevap yapısı oluşturulur. Cevap gösterge matrisi, tam veri matrisi ile aynı biçimde ve büyüklüktedir. Genel olarak **R** ile gösterilen bu gösterge matrisinin satırları, *n* boyutlu birimleri, sütunları ise *k* boyutlu değişkenleri göstermek üzere *nxk* boyutludur ve bu matriste "0" kayıp veriyi, "1" gözlenen veriyi göstermektedir. Daha ayrıntılı bir gösterimde, her tür kayıp veri için farklı bir gösterge de kullanılabilir. Örneğin, "-1" ulaşılamamış, "-2" red eden, "-3" bilmiyor gibi. Analizler kayıp veri olmadığında daha kolay yürütülmektedir. Bu durumda veri kümesi yapısı *dikdörtgen (rectangular)* şeklinde tanımlanır, yani *nxk* boyutlu veri matrisinde kayıp veri yoktur (Longford, 2005).

Anket yapıldıktan sonra cevapsızlık ile çalışmanın ilk adımı, eksikliğin ve etkilerinin yapısının oluşturulmasıdır. Kayıp veri ile ilgili yöntemin seçimi sadece kayıp verinin boyutuna değil aynı zamanda yapısına da bağlıdır. Yani cevapsızlığın yapısı kişilerin soruları cevaplamamaya eğilimi olması veya soruların her bölümünde bir kısım sorular cevaplanırken (veya cevaplanmazken) bir kısım soruların hiç cevaplanmamasına (veya her zaman cevaplanmasına) ya da değişkenlerden biri veya birkaçının değerleri ile ilgili olup olmadığı gibi nedenlere bağlı olarak belirlenir. Bu kayıp veri mekanizmalarına bağlı olarak, uygun yöntemlere karar verilmektedir.

2.1. Kayıp Veri Mekanizmaları

Cevapsızlık mekanizması, tam veri matrisi **Y** verildiğinde **R** cevap gösterge matrisinin koşullu dağılımı, f(**R** | **Y**), biçiminde tanımlanır. Little ve Rubin (2002) bu mekanizmaları, tam rasgele kayıp: TRK (missing completely at random), rasgele kayıp: RK (missing at random) ve rasgele olmayan kayıp: ROK (missing not at random) olmak üzere üç temel kategoriye ayırmıştır.

- ***Tam Rasgele Kayıp (TRK):*** Kayıp veri yani cevaplanmama olasılığı, ilgili değişkenlerden bağımsız ise, kayıp veri TRK olarak adlandırılır. Buna örnek olarak, biyokimya laboratuarında ilgilenilen değişkene ilişkin, parametreleri ölçülmemiş bir kan örneği tüpünün kazayla kırılması ya da anket kağıdının kaybedilmesi verilebilir. Bu tip kayıplılık tam rasgeledir. Gözlenen kayıp olasılığı diğer cevaplayıcıların, örneğimize göre diğer hastaların, özellikleri ile ilgili değildir. TRK koşulu, cevap verenler ve cevap vermeyenlerin gözlenen veri dağılımlarının karşılaştırılmasıyla incelenebilir. Kayıp veri TRK ise, kayıp veri içermeyen değişkenler kümesinin, hedef kitlenin rasgele örneklemi olduğu açıktır. Bu nedenle, kayıp veri incelemelerinde birçok yöntem yansız sonuçlar vermektedir (Donders vd., 2006). TRK'da tüm **Y**, **R**'den bağımsız olacağından, mekanizma biçimi aşağıdaki gibi verilebilir:

Kesikli ise; $P\ (R = r \mid Y = y) = P(R = r)$

Sürekli ise; $f_{R|Y}(r|y) = f_R(r)$

(Little ve Rubin, 2002).

- ***Rasgele Kayıp (RK):*** Kayıp veri olasılığı araştırma konusuna bağlı ise bu kayıp veri türü RK olarak adlandırılır. Tek taraflı bir bağımsızlık söz konusudur. Yani, X ve Y gibi iki değişken alındığında, X değişkenindeki cevapsızlık olasılığı Y değişkenine bağlı iken, Y değişkenindeki cevapsızlık olasılığı X değişkenine bağlı değildir. RK mekanizmasında kayıp veri, kayıplıkları belirlenen diğer cevaplayıcı özelliklerinde koşullu rasgele olup bunlar analizler sırasında elde edilebilir. Örneğin, biyokimyada teşhis testinin tahmin değeri elde edilmek istensin. Test sonuçları bütün ilgili hastalar için biliniyor ancak hasta olmayanların rasgele örneklemi için bilinmiyor olsun. Bu durumda kayıp veri RK olacaktır: Kişinin özelliklerine bağlı gözlenen kayıp veri rasgeledir (hastalık olması veya olmaması). Genellikle kayıp veri RK olduğu durumda, gösterge (indicator) ve ortalama veri türetme gibi bütün basit yöntemler yanlı sonuçlar vermektedir. Bununla birlikte, tekli ve çoklu veri türetme gibi daha ayrıntılı yöntemler yansız sonuçlar vermektedir (Donders vd., 2006). Bir araştırmada, bireylerden birinin durumu yeterince kontrol edilemiyorsa, araştırmadan çıkartılabilmesi RK'a örnektir. RK, **Y** tam veri matrisinin sadece gözlenmiş bileşenlerine bağlı olmasından dolayı daha az bağlayıcı bir varsayıma sahiptir. Bu durum, mekanizma biçimini aşağıdaki gibi gösterilmesini sağlamaktadır:

Kesikli ise; $P\ (R = r \mid Y = y) = P(R = r \mid y_{göz})$

Sürekli ise; $f_{R|Y}(r|y) = f_{R|Y_{göz}}(r|y_{göz})$

(Little ve Rubin, 2002).

- ***Rasgele Olmayan Kayıp (ROK):*** TRK ve RK olmayan kayıplar ROK'tur. Bu tür kayıp verilerde cevaplamama olasılığı cevaplayıcının özelliklerine, yani bir değişkendeki kayıp veri olasılığı, diğer değişkenlere bağlıdır. Örneğin, kişiye sorulan gelir düzeyi ile ilgili bir soruda, gelir düzeyi ortalamanın çok üstünde olan kişiler için kayıplılık söz konusu olabilir. Burada gözlenen kayıplık tam rasgele değildir, cevaplayıcının kendi özelliklerine bağlıdır. Eğer kayıp veri ROK ise, veriden değerli bilgiler kaybedilebilir. Bu tip kayıp veri probleminin çözümü için mevcut yöntemler kullanılamaz (Donders vd., 2006).

2.2. Kayıp Veri Analiz Yöntemleri

Bu bölümde, kayıp veriler için yaygın olarak kullanılan yöntemler, avantajları ve dezavantajları ile birlikte verilecektir. Yöntemler etkinlikleri (efficiency) açısından karşılaştırılacaktır. Etkinlik, örnekleme yönteminden doğrudan etkilendiğinden, örnekleme yöntemi seçiminin önemi bir kez daha vurgulanmış olacaktır.

Daha önce söylendiği gibi, kayıp veri ile çalışmada ilk adım veri matrisinin yapılması planlanan (tam-veri) analizler için düzenlenmesidir. Veri matrisini hazırlamak için kullanılan yöntemleri, "veri indirgeme" ve "veri türetme (imputation)" başlıkları altında toplayabiliriz. Bu iki yöntem de tam-veri analizlerinin uygulanmasına olanak sağlamaktadır. Tam-veri gerektirmeyen yöntemler, tam-veri (örneklem) ve ortaya çıkan eksikliği yani plandan sapma durumunu ortak ele aldığından, oldukça karmaşık bir yapıya sahiptir. Bu nedenle çalışmalarda tam-veri matrisine ulaşılmaya çalışılır. Takip eden altbölümlerde bu yöntemlere kısaca değinilecektir.

2.2.1. Veri indirgeme

Veri indirgemede, eksik veriler silinerek geriye kalan veriler üzerinden analizler yapılır. İndirgenmiş veri kümesi, örneklem büyüklüğünün azalmasına karşın planlanan analizlerin yapılmasına olanak veren tam veri yapısındadır.

Veri indirgemenin iki sakıncası vardır. Araştırma sonuçları için önemli bilgiye sahip bir gözlem kayıp olabilir ve çok iyi bir tahmini yapılabilir durumdayken bu yöntem ile tümü silinebilir. Bu durum çeşitli cevap yapıları kapsamında oldukça etkili bir sorun olabilmektedir. Bundan daha önemlisi ise, seçilen örneklemin kitleyi en iyi temsil ettiği varsayımına aykırı bir durumun gelişmesine neden olur. Veri indirgemenin, veri tamamlamaya kıyasla başarısız sonuçlar verdiğine ilişkin simülasyon çalışması, Donders vd. (2006) tarafından verilmiştir. Veri indirgeme başlığı altında, liste veya durum bazında veri silme, çiftler bazında veri silme incelenebilir.

2.2.1.1. Liste veya durum bazında veri silme

Bu veri silme yönteminde, kayıp veriye sahip gözlem birimi analizden çıkarılmaktadır. Analiz, kayıp verinin olmadığı diğer gözlemler üzerinden devam etmektedir. Kayıp veriler için kullanılan en temel yöntemdir. Bazı çalışmalarda “ampütasyon” olarak da adlandırılmaktadır. Kayıp veri mekanizması TRK olduğunda, kayıp veri içeren gözlemler çıkarıldıktan sonra geriye kalan veri kümesi tam veri kümesinin rasgele bir alt kümesi olduğundan, tahminler yansız olacaktır. Fakat daha az veri kullanıldığı için, standart hatalar büyüyecektir. RK veri mekanizmasında ise bu durum yanlı sonuçlara neden olmaktadır.

2.2.1.2. Çiftler bazında veri silme

Bu yönteme göre, ortalama, standart sapma ve korelasyon matrisi gibi özet istatistikler, her değişken çifti için kayıp olmayan veriler kullanılarak hesaplanmaktadır. Liste veya durum bazında veri silme yöntemine göre avantajı, kullanılabilecek tüm veriyi kullanmasıdır. Kayıp veri, RK mekanizmasına uyuyorsa yanlı sonuçlar vermektedir.

2.2.2. Veri türetme (imputation)

Bu bölümde, tam-veri analizleri için veri kümesi geliştirmeye yönelik yaygın olarak kullanılan çeşitli veri türetme biçimlerine yer verilecektir. Bu yöntemler, kayıp verilerin tahmin edilerek yerine konulması esasına dayandığı için, araştırma için uygun görülen örneklem büyüklüğü azalmayarak aynı kalmaktadır. Bu durum veri indirgeme yöntemlerine göre daha küçük standart hatalara, aynı zamanda önemli bilgilerin kaybedilmemesini sağlayarak elde edilen sonuçların güvenirliğinin artmasına neden olmaktadır. Veri türetme süreçlerine anketin planlama ve geliştirme süreçlerinde karar verilir. Bazı problemler cevaplayan ile görüşme aşamasında veya sorulara çalışılırken erkenden elimine edilebilir, ancak bu bütün problemlerin çözümü için genellikle mümkün değildir (maliyet, zaman kısıtları, cevap yükü nedeniyle). Bu nedenle, hataların giderilmesi için veri türetme süreci kullanılır.

2.2.2.1. Ortalama ile veri türetme

Bu yöntemde, kayıp verilerin bulunduğu *y* değişkeninin gözlenen değerlerinin ortalaması, *y* değişkenindeki her kayıp veri için kullanılır. Sıralı değişkenler için uygulanması kolay bir yöntemdir.

Ortalama ile veri türetme ile değişken değerleri ortalamaya sabitlenmiş ve örneklem ortalaması, bu işlem sonucunda değişmemiş olur. Buna karşılık, *y* değişkeninin örneklem varyansı küçülmüş olacaktır. Ancak örneklem ortalamasıyla kayıp verileri tamamlamak ile sistematik bir hata yapılmış olur.

2.2.2.2. Başka bir değişken yardımıyla veri türetme

Başka bir değişken yardımıyla veri türetmede kayıp veri yerine, karşılık gelen ilişkili diğer değişkenin değeri kullanılmaktadır. Bazı değişkenlerin yüksek ilişkili ve değerlerinin benzer olduğu varsayılsın. Bu durumda eğer bir veri kayıp ise, o veriye karşılık gelen diğer değişkenin değeri bunun yerine kullanılabilir. Dolayısıyla, her *k* değişkeni için, yerine kullanılacak bir *h* değişkeni tanımlanır. Ancak aynı birime karşılık gelen değer, her iki değişkende de kayıp veri olursa, ikinci bir yüksek ilişkili değişkenin tanımlanması gerekir. Tanımlanan bu değişkenin değeri hem *k* hem de *h* değişkenlerindeki o birime ait eksik veriler için kullanılabilir.

Ortalama ile veri türetme ve başka değişken yardımıyla veri türetme sırasıyla dikey ve yatay veri türetme tiplerine örnektir. Dikey veri türetme, aynı sütundaki bilgilere dayalıdır, yatay veri türetme ise aynı cevaplayıcıya ait yani aynı satıra ait bilgileri kullanır. Dikey veri türetme diğer gözlenen cevaplayıcıların bir özetidir. Ortanca ve tepe değeri istatistikleri dikey veri türetme için örnek olarak gösterilebilir.

2.2.2.3. Yakın komşu ile veri türetme

Yakın komşu ile veri türetmede, cevaplayıcıya (alıcı) ait tamamlanmamış kayıt, bir başka cevaplayıcının (verici) aynı değişkenine karşılık gelen tam kayıtları tarafından tamamlanır. Kayıp değerli değişkenler $k=(k_1,\ldots,k_m)$ olsun. Kayıtlı değişkenler, $k^{(c)}$ için (*k*'nın tamamlayıcısı), *k*'daki tam kayıtlı cevaplayıcılar biçiminde tanımlıdır, alıcıya diğerlerinden daha yakındır. Bu yakın komşudur. Benzer nitelikte çeşitli yakın komşular olduğunda, verici bunlar arasından rasgele seçilebilir ve bunun *k* değişken değerleri alıcının karşılık gelen değerlerinde kullanılır. Verici tanımı için basit kural, $k^{(c)}$'de değişken değerlerinin eşleştirilmesidir. Değişkenler önemliliklerine göre sıralanırlar ve böylece ilk değişkenin eşleştirilen değeri ikinci değişkenin eşleştirilen değerinden daha önemli olacaktır. Değer eşleştirme kriterleri, eğer değerler çok farklı değilse daha esnek ve yeterlidir. Bu, özellikle sürekli değişkenler için pratiktir. Daha genel olarak, fark, gerekli cevap yapısına sahip iki değişken için tanımlanabilir ve verici alıcı ile en az farka sahip cevaplayıcı olarak seçilir. Farkın tanımında dikkatli olunmalıdır. Böylece katkıda bulunan değişkenlerin göreli önem düzeyleri uygun bir şekilde yansıtılmış olur. Vericinin tam kayıtlı olmasına gerek yoktur. Bazı değişkenler yerine koymada ya da farkın hesaplanmasında göz ardı edilebilir veya bunların katkıları çok küçük olabilir. Farklar kayıp değerler için tanımlanmalıdır.

Bu yöntemin bazı dezavantajları vardır. Bazı alıcılar, çok yakın komşulara sahip olabilirler. Buna karşılık, bazı vericiler ortak olabilirler. Bunların değerleri birçok alıcı için kullanılabilirken, tam kayıtlı diğer cevaplayıcılar, herhangi bir alıcı için verici olmayabilir (Donders vd., 2006).

2.2.2.4. Hot deck veri türetme

Hot deck veri türetme yakın komşu düzeniyle yakın ilişkilidir. Hot deck'de, her alıcı için verici havuzu tanımlanır ve bu havuzdan vericiler rasgele seçilir. Verici havuzu alıcıya benzer cevaplayıcıları ve $k^{(c)}$ değişkenlerinin kayıtlı değerlerini içerecek şekilde tanımlanır. Alıcılar gruplandırılır ve kümenin içinde benzer verici havuzları paylaştırılacak biçimde veri türetme tasarlanır. Ancak havuzların optimal büyüklüğünün ne olması gerektiği bu yöntemde tartışılan bir noktadır (Little ve Rubin, 2002).

Hot deck tanımı, her bir havuz üyesine, fark ölçümleri temel alınarak ağırlık atanması ve eşit olmayan seçilme olasılıklarının verilmesi biçiminde genişletilebilir.

2.2.2.5. Ağırlıklandırılmış yöntem

Anket veri kümesindeki kayıtlar, genellikle örneklem ağırlıkları ile ilişkilidir. Ağırlıklandırılmış yöntemde gözlenen değerlere verilen ağırlıklar sadece bu gözlemlerin kendilerini değil aynı zamanda kayıp verileri de temsil etmeleri esasına dayanır.

Ağırlık düzeltmesi, anket cevapsızlıklarını göstermek için bir yöntemdir. Bu yöntem değişim yapmadan tam veriden elde edilen tahmin edici özelliklerini korumayı amaçlar. Buradaki değişiklikler, ancak gerçekleştirilmiş tasarımda seçim olasılıklarının tahmini ile ilişkili olabilir. Örneğin, kadınlardan daha büyük oranda cevapsızlık yapan erkekler ile ilgili ağırlıklar artacaktır. Böylece cevaplayıcıların alt örnekleminde daha az erkek olacaktır.

Bu yöntem, grup içi alt örneklemin büyük olmasına dayanır. Aksi takdirde, çok büyük ağırlıkların ortaya çıkması gibi çeşitli anormal gelişmelere neden olabilir. Genellikle bu yöntem örneklem birimi kayıplarında kullanılır (Little ve Rubin, 2002).

2.2.2.6. Regresyon ile veri türetme

Y değişkeninin bazı kayıp değerli cevaplayıcılardan, *Z* değişkeninin ise tam değerli cevaplayıcılardan oluştuğu varsayılsın. Eğer *Y* ve *Z* ilişkili ise, bu ilişki *Y*'deki kayıp değerlerin tahmini için kullanılabilir. *Y* ve *Z*'nin aşağıdaki model kapsamında ilişkili olduğu varsayılsın:

$$Y=f(Z) + \varepsilon \ ,$$
$$Y=\beta_0 + \beta_1 Z \ .$$

Eğer *f* fonksiyonu biliniyor ise,

$$\hat{y}_j = f\left(z_j\right)$$

olur. Burada $\hat{y}_j$ kayıp verinin tahminini göstermektedir. Eğer *f* bilinmiyorsa, modelin parametreleri tahmin edilir, bununla birlikte parametrelerin hedeflenenden farklı çıkması gibi sistematik hatalar ortaya çıkabilir.

Eğer regresyonda kullanılan değişkenlerin bazılarında kayıp değerler var ise, tam kayıtların oluşturduğu alt örneklem verileri kullanılarak modelin parametreleri elde edilebilir. Normal dağılmayan *y* değişkenleri için ise genelleştirilmiş doğrusal modeller kullanılabilir. TRK mekanizması koşulları sağlandığında ve tahminler kayıp veri içermeyen diğer bağımsız değişkenlere bağlı olduğunda en küçük kareler kestirimi yöntemiyle elde edilen model parametreleri tutarlıdır (Little ve Rubin, 2002).

2.2.2.7. Veri türetme modelleri

Her bir veri türetme, doğrudan modele dayalı olarak yapılır. Her kayıp veri yerine konan değer, bütün tam kayıtları içermeyen denklem vasıtasıyla elde edilir. Model esaslı bu yöntem, kısmen kayıp veri kümesi için bilinmeyen parametrelerin tahmininde, en çok olabilirlik tahmin yöntemi gibi yöntemlerin kullanılması esasına dayanır. Avantajı esnek olması, model varsayımları altında sonuçların verilmesi ve değerlendirilmesidir (Little ve Rubin, 2002).

2.2.2.8. EM algoritması

EM algoritması son yıllarda birçok farklı alanda kullanılmaktadır. Algoritma, tam-olmayan veri problemlerinde parametre tahminlerini içeren tekrarlı bir yöntem olup en çok olabilirlik tahminlerini elde etmek için kullanılmaktadır. EM algoritması beklenen değer ve en büyükleyen olmak üzere iki adımda gerçekleşmektedir. Beklenen değer adımında kayıp verilerin, gözlenen verilere göre koşullu beklenen değerini alarak kayıp veri tahmin edilir. En büyükleme adımında, elde edilen kayıp veri tahminleri ile oluşan tam-veri kullanılarak, parametre tahminleri elde edilmektedir (Yazıcı, 2005).

2.2.2.9. Çoklu veri türetme

Çoklu veri türetme, kayıp verilerin yerine *m* tekrar sayısı olmak üzere aşağıda algoritması verilmiş simülasyonun yapıldığı bir Monte Carlo tekniğidir (Little ve Rubin, 2002).

1) Rasgele değişim içeren bir model vasıtasıyla kayıp veriler tamamlanır.
2) Bu işlem $m>1$ kez tekrarlanır. Burada *m* küçük bir değerdir (3-10 arasında). Böylelikle *m* tane tamamlanmış veri kümesi oluşturulmuş olur.
3) İstenilen standart tam veri analizleri bütün veri kümelerine uygulanır.
4) *m* defa tekrarlanmış analizlerden, *p* olasılık değerleri, güven aralıkları, varyanslar ve ortalamalar hesaplanır.

ÜÇÜNCÜ BÖLÜM

3. KAYIP VERİ OLMASI DURUMUNDA, KİTLE ORTALAMASI TAHMİNİNE İLİŞKİN LİTERATÜRDE YER ALAN ÇALIŞMALAR

Bu bölümde, örneklem birimlerine ait kayıp veri olması durumunda, klasik istatistiksel araçlar kullanılarak, kitle ortalamasının tahmini için uygun tahmin edici elde etmeye yönelik yapılmış çalışmalar incelenmiştir. Önerilen veri türetme yöntemleri ve bunlara bağlı elde edilen tahmin edicilerin özellikleri ilgili başlıklar altında verilmiştir. Literatürde verilen bazı tahmin ediciler, bu konuda temel teşkil eden, ortalama ve oransal veri türetme yöntemlerinden türetildiğinden ve bu yöntemler ile karşılaştırıldığından, öncelikle ortalama ve oransal veri türetme yöntemlerinin teorik çıkarsamaları verilmiştir.

- ***Ortalama ile veri türetme***

Ortalama ile veri türetme yöntemine göre, daha önce bahsedildiği gibi, kayıp veri yerine ilgili değişkenin kayıp veri dışındaki veriden elde edilen örneklem ortalaması konur. Cevaplı birimlerin yer aldığı küme A ve cevapsızların kümesi A^c biçiminde tanımlanmaktadır. Buna göre, her $i \in A$ için y_i değeri gözlenmektedir. A kümesinin eleman sayısı r'dir. Bununla birlikte, $i \in A^c$ birimi için y_i değeri kayıptır. Bunun için veri türetme yöntemine başvurulmuştur. Ortalama ile veri türetme yöntemine göre veriler,

$$y_i^* = \begin{cases} y_i, & i \in A \\ \bar{y}_r, & i \in A^c \end{cases}$$

biçimini almaktadır (Singh ve Horn, 2000). Buradan, kitle ortalamasının nokta tahmini ise,

$$\bar{y}_r = \frac{1}{r}\sum_{i=1}^{r} y_i \qquad (3.1)$$

olmaktadır.

Bu tahmin edici yansızdır. Hata Kareler Ortalaması (HKO) eşitliği ise aşağıda verilmiştir:

$$HKO(\bar{y}_r) = V(\bar{y}_r) \cong \left(\frac{1}{r} - \frac{1}{N}\right) S_y^2 . \tag{3.2}$$

- ***Oransal veri türetme***

Burada veri türetmenin x_i yardımcı değişkeni yardımıyla yapıldığı varsayılmaktadır. Kayıp olan y_i değeri, bx_i oransal yöntemi ile tahmin edilir. Burada regresyon modelinin katsayısı $b = \sum_{i=1}^{r} y_i \Big/ \sum_{i=1}^{r} x_i$ olarak tanımlanmıştır. Oransal veri türetmeden sonra verilerin durumu aşağıda özetlendiği gibidir:

$$y_i^* = \begin{cases} y_i, & i \in A \\ bx_i, & i \in A^c \end{cases}$$

Veri türetmenin bu biçimi oransal veri türetme yöntemi (ratio type imputation) olarak adlandırılır (Singh ve Horn, 2000). Bu yönteme göre, kitle ortalamasının nokta tahmini,

$$\bar{y}_{ORAN} = \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r} \tag{3.3}$$

biçimindedir. Burada $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{x}_r = \frac{1}{r}\sum_{i=1}^{r} x_i$ ve $\bar{y}_r = \frac{1}{r}\sum_{i=1}^{r} y_i$ göstermektedir. Bu tahmin edicinin elde edilmesi,

$$\bar{y}_{ORAN} = \frac{1}{n}\left[\sum^{r} y_i + b\sum^{n-r} x_i\right]$$

$$=\frac{1}{n}\left[\sum^{r} y_i + \frac{\sum^{r} y_i}{\sum^{r} x_i}\sum^{n-r} x_i\right] = \frac{1}{n}\left[\frac{\sum^{r} y_i \sum^{r} x_i + \sum^{r} y_i \sum^{n-r} x_i}{\sum^{r} x_i}\right]$$

$$=\frac{1}{n}\left[\frac{\sum^{r} y_i\left(\sum^{r} x_i + \sum^{n-r} x_i\right)}{\sum^{r} x_i}\right] = \frac{1}{n}\left[\frac{\sum^{r} y_i \sum^{n} x_i}{\sum^{r} x_i}\right]$$

$$=\frac{r}{r}\frac{\sum^{r} y_i}{\sum^{r} x_i}\frac{1}{n}\sum^{n} x_i$$

$$\Rightarrow \bar{y}_{ORAN} = \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r}$$

biçiminde verilebilir. Bu tahmin edicinin yan ve HKO eşitlikleri aşağıda verildiği gibidir (Singh ve Deo, 2003):

$$HKO(\bar{y}_{ORAN}) \cong \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2 + \left(\frac{1}{r}-\frac{1}{n}\right)\left(R^2 S_x^2 - 2RS_{xy}\right), \tag{3.4}$$

$$Yan(\bar{y}_{ORAN}) \cong \left(\frac{1}{r}-\frac{1}{n}\right)\bar{Y}\left(C_x^2 - C_{xy}\right). \tag{3.5}$$

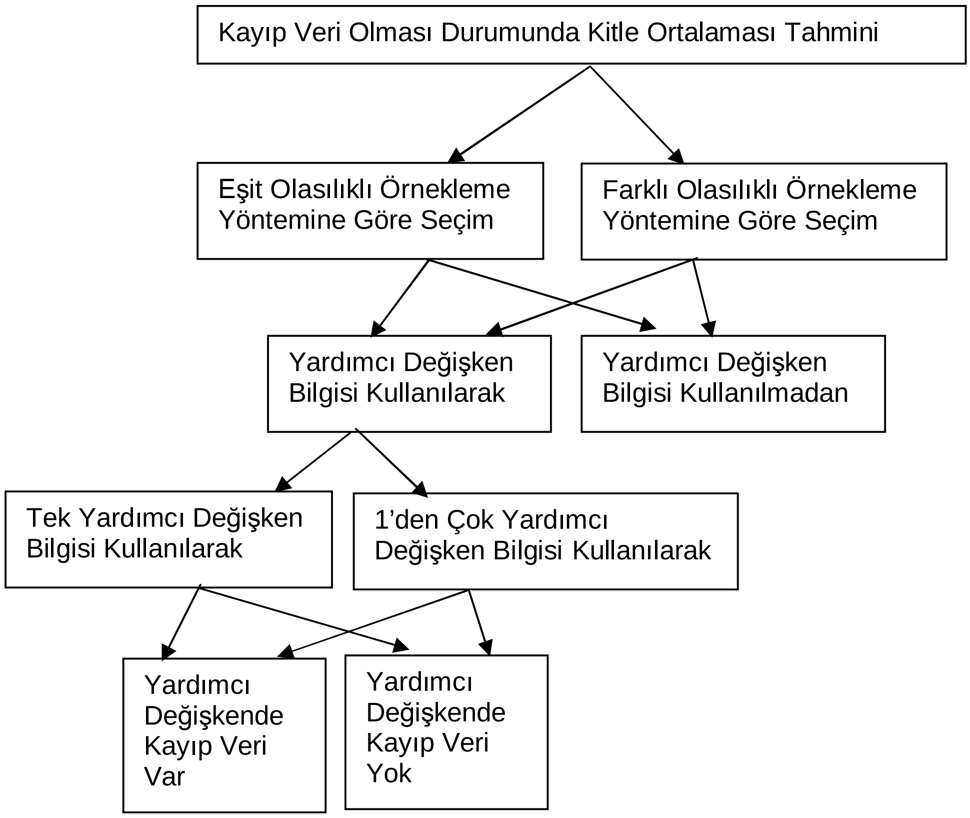
Burada S_x, S_y ilgili değişkenlerin kitle varyanslarını, S_{xy} X ve Y arasındaki kitle kovaryansını, C_x ve C_{xy} kitle değişim katsayılarını göstermektedir.

Oransal veri türetme ve ortalama ile veri türetme HKO açısından karşılaştırıldığında,

$$R = \frac{\bar{Y}}{\bar{X}} < \frac{2S_{xy}}{S_x^2} = 2B \tag{3.6}$$

koşulu sağlandığı sürece, oransal veri türetme yönteminin daha etkin olduğu sonucuna varılır.

Literatürde yer alan çalışmaları aşağıda verilen şema yardımıyla gruplandırabiliriz:



3.1. Yardımcı Değişken Bilgisi Kullanmayan Tahmin Ediciler

3.1.1 Gamrot tahmin edicisi

Cevapsızlık etkilerinin giderilmesi için, daha önce de bir kısmından bahsedildiği gibi, birçok yöntem geliştirilmiştir. Bunların en önemlilerinden biri de Hansen ve Horwitz (1946) tarafından sunulan çift örnekleme (double sampling) yöntemidir. Bu yöntem, cevapsızların oluşturduğu alt örneklemden tekrar bir alt örneklem çekimine dayanır. Bu sürecin genelleştirilmesiyle ilgili çeşitli çalışmalar sunulmuştur. Sarndal vd. (1992) tarafından, seçim yöntemine ilişkin kısıt getirmeden yapılan uygulama bunlardan biridir. İkinci safhada cevapsızlığın olmaması varsayımı altında, ikinci safhadaki veriler kullanılarak oluşturulan kitle ortalamasının tahmin edicisi yansızdır.

Gamrot (2007), yukarıda bahsedilen varsayımın gerçekleşmediği durum için, yani ikinci safhada da cevapsızlık olması durumunu kapsayan, zincirleme örnekleme yöntemine göre elde edilen ortalamanın tahmin edicisinin özelliklerini araştırmıştır. Her iki safhada da, herhangi bir örnekleme tasarımının kullanıldığı genelleştirilmiş zincirleme örnekleme yöntemi için, yan ve varyans ifadeleri elde edilmiş, tabakalı rasgele örneklemede uygulanan ayrı (separate) yönteminin kullanıldığı stokastik cevapsızlık olduğu varsayılmıştır.

Buna göre, N büyüklüğünde bir U kitlesinden, ilk sırada dahil ediliş olasılığı π_i ve ikinci sırada dahil ediliş olasılığı π_{ij} olmak üzere $p(s)$ örnekleme yöntemine göre n büyüklüğünde s rasgele örneklemi çekilmektedir. Gamrot (2007) çalışmasında, stokastik cevapsızlığın olduğu varsayılmıştır. Bu, her *i.* kitle birimi eğer örneklemde de yer alıyorsa $\rho_{i|s}$ koşullu cevap olasılığına sahip olduğu anlamına gelir. Burada *i.* ve *j.* birimler birlikte eğer örnekleme seçilmişlerse koşullu cevap olasılıkları, $\rho_{ij|s}$ ile gösterilir. s örneklemi, birbiriyle çakışmayan n_1 birimli cevaplı birimlerin oluşturduğu A ve n_2 birimli cevapsız birimlerin oluşturduğu A^c alt kümelere ayrılmıştır. Böylece cevapsızlık, bilinmeyen $q(A|s)$ cevap olasılık dağılımı tarafından ifade edilir. $q(A|s)$ ve cevap olasılıkları arasındaki ilişki aşağıdaki biçimde verilebilir (Gamrot, 2007):

$$\rho_{i|s} = \sum_{i \in A, A \subseteq s} q(A|s) \ , \tag{3.7}$$

$$\rho_{ij|s} = \sum_{i,j \in A, A \subseteq s} q(A|s) \ . \tag{3.8}$$

İkinci safhada ise n' büyüklüğünde s' alt örneklemi A^c'den herhangi başka bir örnekleme tasarımına, $p'(s'|s, A, A^c)$'e göre çekilir. Burada sırasıyla birinci ve ikinci dahil ediliş olasılıkları $\pi_{i|s,A,A^c}$ ve $\pi_{ij|s,A,A^c}$, $i, j \in A^c$ ile gösterilir. Bundan sonraki adım, alt örneklem birimlerinin toplanmasıdır. Gamrot, ikinci safhada da cevapsızlık olabileceği varsayımı altında ve ilk safhadaki gibi belirli biçimde olmayan stokastik cevapsızlık mekanizmasını ele almıştır. Sonuç olarak, her *i.* kitle birimi, eğer alt örnekleme seçilmişse, cevaplılık $\rho_{i|s,A,A^c,s'}$ koşullu olasılığına sahiptir. *i.* ve *j.* birimler birlikte s' örneklemine seçilmişlerse koşullu cevap olasılığı $\rho_{ij|s,A,A^c,s'}$ ile gösterilir .

Burada da s' alt örneklemi, cevaplıların oluşturduğu A' ve cevapsızların oluşturduğu $A^{c\prime}$ altkümelerine bölünmüş olur.

Bu safhada da cevapsızlık, bilinmeyen $q'\left(A' \middle| s, A, A^c, s'\right)$ olasılık dağılımı tarafından ifade edilir. Burada ikinci safha cevap olasılıkları aşağıdaki ifadelere göre hesaplanır:

$$\rho_{i|s,A,A^c,s'} = \sum_{i \in A', A' \subseteq s'} q'\left(A' \middle| s, A, A^c, s'\right), \qquad (3.9)$$

$$\rho_{ij|s,A,A^c,s'} = \sum_{i,j \in A', A' \subseteq s'} q'\left(A' \middle| s, A, A^c, s'\right). \qquad (3.10)$$

Burada $p(s)$, $q(A|s)$, $p'\left(s' \middle| s, A, A^c\right)$ ve $q'\left(A' \middle| s, A, A^c, s'\right)$ olasılık dağılımları olmak üzere, dört rasgele tahmin kaynağı mevcuttur. Bütün beklenen değerler aksi söylenmedikçe bu olasılık dağılımlarına göre hesaplanmıştır.

Aşağıda, kitle ortalamasının Horvitz-Thompson tahmin edicisi verilmiştir (Horvitz-Thompson tahmin edicileriyle ilgili bilgi EK 1'dedir):

$$\hat{\bar{y}}_{HT} = \frac{1}{N} \sum_{i \in s_1} \frac{y_i}{\pi_i}. \qquad (3.11)$$

Eğer ilk safhada cevapsızlık yoksa $A \equiv s$ olacaktır ve tahmin edici aşağıda verilen varyans ifadesiyle yansız elde edilecektir:

$$V_*\left(\hat{\bar{y}}_{HT}\right) = \frac{1}{N^2} \sum_{i,j \in U} y_i y_j \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right). \qquad (3.12)$$

Stokastik cevapsızlık durumunda çoğu zaman *A*, *s*'den farklıdır ve bu durumda Eşitlik (3.12)'de verilen tahmin edici yansız olmayacaktır. Her birim cevaplılığının *s*'ye bağımlı olmadığını, başka bir deyişle $\rho_{i|s} = \rho_i$, $\rho_{ij|s} = \rho_{ij}$ olduğunu varsayalım. Bu durumda Eşitlik (3.11)'de verilen tahmin edicinin yanı ve varyansı aşağıda verildiği gibi elde edilir:

$$Yan\left(\hat{\bar{y}}_{HT}\right) = -\frac{1}{N}\sum_{i\in U}\left(1-\rho_i\right)y_i \ , \tag{3.13}$$

$$V\left(\hat{\bar{y}}_{HT}\right) = \frac{1}{N^2}\sum_{i,j\in U} y_i y_j \frac{\pi_{ij}\rho_{ij} - \pi_i\pi_j\rho_i\rho_j}{\pi_i\pi_j} \tag{3.14}$$

(Nargundkar ve Joshi, 1975).

Bu sonuçlar cevapsızlık durumlarında daha iyi tahmin ediciler elde edilmesini sağlamış ve Gamrot (2007) ikinci safha verilerini kullanarak aşağıda verilen tahmin ediciyi önermiştir:

$$\hat{\bar{y}}_{\pi} = \frac{1}{N}\left(\sum_{i\in A}\frac{y_i}{\pi_i} + \sum_{i\in A'}\frac{y_i}{\pi_i\pi_{i|s,A,A^c}}\right). \tag{3.15}$$

Genellikle ikinci safhada cevapsızlığın olmadığı varsayılır, bu da $A' \equiv s'$ anlamına gelir. Bu varsayım altında, Eşitlik (3.15)'de verilen tahmin edici, ilk safhada cevap dağılımına bakılmaksızın kitle ortalaması için yansızdır ve varyansı aşağıda verildiği gibi elde edilir (Gamrot, 2007):

$$V\left(\hat{\bar{y}}_{\pi}\right) = \frac{1}{N^2}\sum_{i,j\in U} y_i y_j\left(\frac{\pi_{ij}}{\pi_i\pi_j} - 1\right) + \frac{1}{N^2}E_{pq}\left(\sum_{i,j\in s_2}\frac{y_i y_j}{\pi_i\pi_j}\left(\frac{\pi_{ij|s,A,A^c}}{\pi_{i|s,A,A^c}\pi_{j|s,A,A^c}} - 1\right)\right). \tag{3.16}$$

Burada $E_{pq}(.)$, ilk safha örnekleme yöntemini ve ilk safha cevap dağılımını göstermektedir. Beklenen değer operatörünü ortadan kaldırmak için, $p'\left(s'|s, A, A^c\right)$ ve $q\left(A|s\right)$ ifadelerine ihtiyaç vardır. Örneğin, her iki safhada da yerine koymadan basit rasgele örnekleme yönteminin kullanıldığı, alt örneklem büyüklüklerinin $n' = cn_2$, 0<c<1 ve $A' \equiv s'$ olduğu varsayılsın. Bu durumda, Eşitlik (3.15) ile verilen tahmin edici, Hansen ve Hurwitz (1946) tarafından sunulan aşağıdaki tahmin ediciye dönüşecektir:

$$\hat{\bar{y}}_{HH} = \frac{n_1}{n}\bar{y}_A + \frac{n_2}{n}\bar{y}_{s'} \ . \tag{3.17}$$

Burada, $\bar{y}_A = \frac{1}{n_1}\sum_{i\in A} y_i$ ve $\bar{y}_{s'} = \frac{1}{n'}\sum_{i\in s'} y_i$ olarak alınmaktadır.

Her iki safhada da cevapsızlığın oluşumu ile ilgili daha genel durum ele alınırsa, cevapsızlığın *s*'ye bağımlı olmadığı varsayımı altında, $\rho_{i|s} = \rho_i$, $\rho_{ij|s} = \rho_{i|s,A,A^c,s'} = \rho'_i$ ve $\rho_{ij|s,A,A^c,s'} = \rho'_{ij}$, $i, j \in U$ olmak üzere, Eşitlik (3.15)'de verilen tahmin edicinin yan ve varyansı aşağıda verilmiştir (Gamrot, 2007):

$$Yan\left(\hat{\hat{y}}_\pi\right) = -\frac{1}{N}\sum_{i\in U} y_i\left(1-\rho_i\right)\left(1-\rho'_i\right), \tag{3.18}$$

$$\begin{aligned} V\left(\hat{\hat{y}}_\pi\right) &= V_*\left(\hat{\hat{y}}_{HT}\right) + \frac{1}{N^2} E_{pq}\left(\sum_{i,j\in s_2} \frac{y_i y_j\left(\pi_{ij|s,A,A^c}\rho'_{ij} - \pi_{i|s,A,A^c}\pi_{j|s,A,A^c}\rho'_i\rho'_j\right)}{\pi_i\pi_j\pi_{i|s,A,A^c}\pi_{j|s,A,A^c}}\right) \\ &+ \frac{1}{N^2}\sum_{i,j\in U}\frac{y_i y_j\left(1-\rho'_j\right)\pi_{ij}}{\pi_i\pi_j}\left(\left(\rho_{ij}-\rho_i\right)\left(1-\rho'_i\right)-\left(1-\rho_j\right)\left(1+\rho'_i\right)\right) \\ &+ \frac{1}{N^2}\sum_{i,j\in U} y_i y_j\left(1-\rho'_j\right)\left(1-\rho_j\right)\left(1+\rho_i+\rho'_i-\rho_i\rho'_i\right). \end{aligned} \tag{3.19}$$

Görüldüğü gibi yan her iki safhada da cevap olasılıklarına bağımlıdır ancak ilk ve ikinci safhada dahil ediliş olasılıklarından bağımsız olduğundan, Eşitlik (3.15)'de verilen tahmin edicinin asimptotik yansız olduğu söylenemez. $i \in U$ için $y_i > 0$ olduğunda $\left|Yan\left(\hat{\hat{y}}_\pi\right)\right| \le \left|Yan\left(\hat{\hat{y}}_{HT}\right)\right|$ olur, yani ikinci safhada cevapsızlık varsa ikinci safha verilerinin kullanılması, Horvitz-Thompson tahmin edicisinin yanını azaltır.

İkinci safhada tam cevap için (yani $A' \equiv s'$), Eşitlik (3.19)'da verilen varyans, Eşitlik (3.10) ile verilen varyansa eşit olur. Her iki safhada tam cevaplılıkta, $V_*\left(\hat{\hat{y}}_{HT}\right)$ daha da fazla küçülür. $E_{pq}(.)$, ilk safha örnekleme yöntemine ve ilk safha cevap dağılımına göre beklenen değeri gösterir. $p'\left(s'\middle|s, A, A^c\right)$ ve $q\left(A\middle|s\right)$ ile ilgili ek varsayımlarla beklenen değer kaybolur.

İkinci safhada cevapsızlık görüldüğünde, varyans formülündeki birimlerin bilinmeyen cevap olasılıklarının varlığından dolayı, Eşitlik (3.19)'daki varyansın tahmin edicisinin elde edilmesi oldukça zordur.

Gamrot (2007) ayrıca, yukarıda tanımlanan genel sürece ait, her iki safhada da *s* örnekleminden bağımsız tek biçim cevapsızlık mekanizması ve ilk safhada deterministik cevapsızlık olması özel durumlarını da incelemiştir.

3.2. Yardımcı Değişken Bilgisi Kullanan Tahmin Ediciler

3.2.1. Tek yardımcı değişken kullanan ve yardımcı değişkende kayıp veri olmayan tahmin ediciler

3.2.1.1. Singh ve Deo tahmin edicisi

Singh ve Deo (2003) çalışmalarında, kitle ortalaması, $\overline{Y}$, tahmini için, *n* birimlik yerine koymadan *s* basit rasgele örneklemini çekmişlerdir. Çalışmalarında, cevapsızların oluşturduğu, (*n*-*r*) elemanlı A^c kümesinde, kayıp olan her y_i değeri tahmini için aşağıda gösterilen "Kuvvet Dönüşümü" diye adlandırılan veri türetme yöntemi önerilmiştir. Bu veri türetme yönteminde verilerin görüntüsü,

$$y_i^* = \begin{cases} y_i, & i \in A \\ \bar{y}_r \left[n \left(\dfrac{\bar{x}_n}{\bar{x}_r} \right)^{\alpha} - r \right] \dfrac{x_i}{\sum_{i \in A^c} x_i}, & i \in A^c \end{cases}$$

biçimindedir. Buradaki α, sonuç olarak elde edilecek olan ortalama tahmin edicinin varyansını en küçükleyen bir sabittir. Buna göre, kitle ortalamasının tahmin edicisi,

$$\bar{y}_{SD1} = \frac{1}{n} \left\{ \sum_{i=1}^{r} y_i + \sum_{i=1}^{n-r} \bar{y}_r \left[n \left(\frac{\bar{x}_n}{\bar{x}_r} \right)^{\alpha} - r \right] \frac{x_i}{\sum_{i \in A} x_i} \right\}$$

$$= r\bar{y}_r + \bar{y}_r\left[n\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\alpha} - r\right]\frac{\sum^{n-r} x_i}{\sum^{n-r} x_i} = n\bar{y}_{SD1}$$

$$= \bar{y}_r\left[r + n\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\alpha} - r\right]\frac{\sum^{n-r} x_i}{\sum^{n-r} x_i} = n\bar{y}_{SD1} \Rightarrow$$

$$\bar{y}_{SD1} = \bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\alpha} \tag{3.20}$$

biçiminde elde edilir. Burada, α =0 için $\bar{y}_{SD1} = \bar{y}_r$; α =1 için $\bar{y}_{SD1} = \bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)$ ve α = -1 için ise $\bar{y}_{SD1} = \bar{y}_r\left(\frac{\bar{x}_r}{\bar{x}_n}\right)$ biçiminde elde edilir. Bu sonuçlara göre, önerilen kuvvet dönüşümü veri türetme yönteminin, ortalama, çarpımsal ve oransal veri türetme yöntemlerini içeren genel bir tahmin edici sınıfı olduğu söylenebilir. $\bar{y}_{SD1}$ tahmin edicisinin yanı ve HKO'sı aşağıdaki biçimde elde edilebilir:

$$Yan(\bar{y}_{SD1}) \cong \bar{Y}\left(\frac{1}{r} - \frac{1}{n}\right)\left[\frac{\alpha(\alpha+1)}{2}C_x^2 - \alpha C_{xy}\right], \tag{3.21}$$

$$\alpha_{opt} = \rho\frac{C_y}{C_x} \Rightarrow HKO_{\min}(\bar{y}_{SD1}) = HKO(\bar{y}_{ORAN}) - \left(\frac{1}{r} - \frac{1}{n}\right)S_x^2\left(\frac{S_{xy}}{S_x^2} - \frac{\bar{Y}}{\bar{X}}\right)^2. \tag{3.22}$$

Eşitlik (3.21) ve (3.22)'de verilen yan ve HKO eşitliklerin çıkarımı EK 2'de verilmiştir.

Eşitlik (3.22)'de görüldüğü gibi, Singh ve Deo (2003) tarafından önerilen kuvvet dönüşümüne göre elde edilen kitle ortalaması tahmin edicisi, oransal tahmin ediciden daha etkindir.

Satıcı ve Kadılar (2009) çalışmasında, Eşitlik (3.20)'de verilen Singh ve Deo (2003) tahmin edicisine paralel olarak, X değişkenine ilişkin kitle ortalamasının bilindiği durum için aşağıdaki tahmin ediciyi önermişlerdir:

$$\bar{y}_{SK1} = \bar{y}_r \left(\frac{\bar{X}}{\bar{x}_r} \right)^{\alpha} \tag{3.23}$$

Burada $\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}$ göstermektedir. α, *HKO*($\bar{y}_{SK1}$) eşitliğini en küçük yapan sabittir.

Eşitlik (3.23)'de verilen tahmin edicinin, α_{opt} için en küçük *HKO* eşitliği aşağıda verilmiştir:

$$\alpha_{opt} = \rho \frac{C_y}{C_x} \Rightarrow HKO_{\min}(\bar{y}_{SK1}) = \left(\frac{1}{r} - \frac{1}{N} \right) S_y^2 \left(1 - \rho^2 \right) \tag{3.24}$$

Bu tahmin edicinin HKO eşitliğinin elde edilişi EK 4'de verilmiştir. Eşitlik (3.23)'de verilen tahmin edici verilen Singh ve Deo (2003) tahmin edicisi ile karşılaştırılmış ve Singh ve Deo (2003) tahmin edicisinden her zaman daha etkin olduğu gösterilmiştir (Satıcı ve Kadılar, 2009).

Singh ve Deo (2003) çalışmasında, yukarıda bahsedildiği gibi, ayrıca çarpımsal veri türetme yöntemine de değinilmiştir. Çarpımsal veri türetme yönteminin kullanımı sınırlı olmasına karşın (negatif ilişki gerektirdiğinden), tıbbi ve sosyal bilimler ile ilişkili özel problemlerin çözümünde çok önemlidir. Bu bilimlerle ilişkili birçok değişken mevcuttur. Örneğin, yaşla negatif ilişkili, (a) uykuda geçirilen süre, (b) işitme eğilimi, (c) görme niteliği, (d) saç teli sayısı, (e) çalışma kapasitesi v.b. değişkenler mevcuttur. Bu değişkenleri ve buna benzer örnekleri artırmak mümkündür. İlgilenilen bu değişkenlere ait bilgiler kayıpsa, fakat insanların yaşına ulaşılabiliyorsa, çarpımsal veri türetme yönteminin kullanılması faydalı olacaktır. Önerdikleri kuvvet veri türetme yönteminde $\alpha = -1$ olarak alınırsa, çarpımsal veri türetme yöntemine ulaşılır ve veriler aşağıdaki biçimi alır:

$$y_i^* = \begin{cases} y_i, & i \in A \\ \bar{y}_r \left[\dfrac{n\bar{x}_r - r\bar{x}_n}{\bar{x}_n} \right] \dfrac{x_i}{\sum_{i \in A^c} x_i}, & i \in A^c \end{cases}$$

Bu veriler esas alındığında, kitle ortalamasının tahmin edicisi ise,

$$\bar{y}_Ç = \bar{y}_r \left(\frac{\bar{x}_r}{\bar{x}_n} \right) \qquad (3.25)$$

biçiminde elde edilir. Çarpımsal tahmin ediciye ait HKO ise aşağıda gösterildiği gibidir:

$$HKO(\bar{y}_Ç) \cong \left(\frac{1}{r} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{r} - \frac{1}{n} \right) \left[R^2 S_x^2 + 2RS_{xy} \right]. \qquad (3.26)$$

Singh ve Deo (2003) çalışmasında, çoklu yardımcı değişken bilgisini içeren yeni bir veri türetme yöntemi de aşağıda gösterildiği şekliyle verilmiştir:

$$y_i^* = \begin{cases} y_i, & i \in A \\ \bar{y}_r \left[n \prod_{j=1}^{p} \left(\dfrac{\bar{x}_{nj}}{\bar{x}_{rj}} \right)^{\alpha_j} - r \right] \dfrac{\sum_{j=1}^{p} x_{ij}}{\sum_{i \in A^c} \sum_{j=1}^{p} x_{ij}}, & i \in A^c \end{cases}$$

Burada $\prod_{j=1}^{p} x_j = x_1 x_2 \ldots x_p$ *p*-terim çarpımıdır. Bu yöntem altında, kitle ortalamasının nokta tahmin edicisi,

$$\bar{y}_{SD2} = \bar{y}_r \prod_{j=1}^{p} \left(\frac{\bar{x}_{nj}}{\bar{x}_{rj}} \right)^{\alpha_j} \qquad (3.27)$$

biçiminde elde edilir. Bu tahmin edici, Srivastava (1967) tahmin edicisinin çok değişkenli genelleştirilmiş halidir. $\bar{y}_{SD2}$ tahmin edicisinin minimum varyansı ise,

$$V(\bar{y}_{SD2}) = \left(\frac{1}{r} - \frac{1}{N}\right) S_y^2 - \left(\frac{1}{r} - \frac{1}{n}\right) S_y^2 \left(1 - R^2_{yx_1 \ldots x_p}\right) \tag{3.28}$$

biçimindedir. Burada yer alan $R^2_{yx_1 \ldots x_p}$, çoklu korelasyon katsayısını ifade etmektedir.

3.2.1.2. Singh ve Horn tahmin edicisi

Singh ve Horn (2000) çalışmalarında, ortalama ve oransal yöntemlerinden yararlanarak yeni bir veri türetme yöntemi ve buna bağlı yeni bir kitle ortalaması tahmin edicisi önermişlerdir. Bu veri türetme yönteminde sadece ilgilenilen değişkenin kayıp verileri için değil aynı zamanda gözlenen verileri için de yardımcı değişkenden faydalanılarak tahmin değerleri verilmiştir. Burada amaç, en küçük HKO'a sahip kitle ortalaması tahmin edicisine ulaşmaktır. Buna göre, cevaplıların oluşturduğu *A* ve cevapsızların oluşturduğu A^c alt kümelerindeki veriler aşağıdaki gibi olmaktadır:

$$y_i^* = \begin{cases} \dfrac{\alpha n y_i}{r} + (1-\alpha) b x_i, & i \in A \\ (1-\alpha) b x_i, & i \in A^c \end{cases}$$

Burada yer alan α değeri, veri türetmeden elde edilen tahmin edicinin HKO'ı en küçükleyen bir sabittir. Bu veri biçimine göre elde edilen ortalamaya ilişkin tahmin edici ise,

$$\bar{y}_{SH} = \alpha\, \bar{y}_r + (1-\alpha) \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r} \tag{3.29}$$

biçiminde tanımlanır. Bu tanımın gösterimi,

$$\bar{y}_{SH} = \frac{\sum^{r} \left[\dfrac{\alpha\, n y_i}{r} + (1-\alpha) b x_i \right] + \sum^{n-r} (1-\alpha) b x_i}{n}$$

$$= \frac{\alpha n\bar{y}_r + b(1-\alpha)\sum^{n} x_i}{n} = \alpha\bar{y}_r + b(1-\alpha)\bar{x}_n$$

Burada $\quad b = \dfrac{\sum^{r} y_i}{\sum^{r} x_i} \Rightarrow b = \dfrac{\sum^{r} y_i / r}{\sum^{r} x_i / r} = \dfrac{\bar{y}_r}{\bar{x}_r}$.

Buna göre,

$$\bar{y}_{SH} = \alpha\,\bar{y}_r + (1-\alpha)\bar{y}_r\,\frac{\bar{x}_n}{\bar{x}_r}$$

biçiminde elde edilir. Bu tahmin edici, Chakrabarty (1968), Vos (1980), Adhavryu ve Gupta (1983) tarafından sunulan kitle ortalaması tahmin edicileri ile benzerdir:

$$\bar{y}_{CVAG} = \alpha\,\bar{y}_n + (1-\alpha)\bar{y}_n\,\frac{\bar{X}}{\bar{x}_n}.$$

$\bar{y}_{SH}$ tahmin edicisine ait yan ve HKO eşitlikleri aşağıda verilmiştir:

$$Yan(\bar{y}_{SH}) \cong (1-\alpha)\left(\frac{1}{r}-\frac{1}{n}\right)\bar{Y}(C_x^2 - C_{xy}), \qquad (3.30)$$

$$HKO(\bar{y}_{SH}) \cong \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2 + \left(\frac{1}{r}-\frac{1}{n}\right)\left((\alpha-1)^2 R^2 S_x^2 - 2(\alpha-1)RS_{xy}\right). \qquad (3.31)$$

$HKO(\bar{y}_{SH})$ eşitliğini en küçükleyen optimal α değeri ise,

$$\alpha_{opt} = 1-\rho\frac{C_y}{C_x} \qquad (3.32)$$

olarak bulunmuştur. Bu değer, Eşitlik (3.31)'de yerine konarak, $HKO_{\min}(\bar{y}_{SH})$ aşağıdaki gibi elde edilir:

$$HKO_{\min}(\bar{y}_{SH}) \cong \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2-\left(\frac{1}{r}-\frac{1}{n}\right)\rho^2 S_y^2 . \tag{3.33}$$

Eşitlik (3.33)'ün $2\left(\frac{1}{r}-\frac{1}{n}\right)RS_{xy}$ ifadesi eklenip çıkartılarak, $HKO(\bar{y}_{ORAN})$ cinsinden yazılması sağlanabilir;

$$HKO_{\min}(\bar{y}_{SH}) = HKO(\bar{y}_{ORAN}) - \underbrace{\left(\frac{1}{r}-\frac{1}{n}\right)\bar{Y}^2 C_x^2\left(1-\rho\frac{C_y}{C_x}\right)^2}_{>0} . \tag{3.34}$$

Buradan, $HKO(\bar{y}_{SH}) < HKO(\bar{y}_{ORAN})$ olduğu söylenebilir.

$HKO(\bar{y}_{SH})$ değerini en küçük yapan α değeri incelendiğinde, kitle değerlerine bağlı olduğu görülmektedir. Ayrıca, kitle karakteristiklerine bağlı $K=\rho C_y / C_x$ parametresinin bilinmediği durumlar için, Singh ve Horn (2000) çalışmasında, α_i tahmin edicileri önerilmiştir:

i) $\hat{\alpha}_1 = 1-\dfrac{\bar{x}_r\, s_{xy}^*}{\bar{y}_r\, s_x^{*2}}$, $s_{xy}^* = (r-1)^{-1}\sum^r (y_i-\bar{y}_r)(x_i-\bar{x}_r)$ ve $s_x^{*2} = (r-1)^{-1}\sum^r (x_i-\bar{x}_r)^2$

ii) $\hat{\alpha}_2 = 1-\dfrac{\bar{x}_n\, s_{xy}^*}{\bar{y}_r\, s_x^2}$, $s_x^2 = (n-1)^{-1}\sum^n (x_i-\bar{x}_n)^2$

Sonsuz büyüklükteki kitlelerde, tahmin edicilerin hata kareler ortalamasının hemen hemen benzer sonuçlar vermesinden dolayı $\hat{\alpha}_1$ ve $\hat{\alpha}_2$'nin hangisinin kullanılacağı önemli olmamasına karşın, sonlu büyüklükteki kitlelerde $\hat{\alpha}_2$'nin daha iyi sonuçlar verdiği gözlenmiştir (Singh ve Horn, 2000).

3.2.1.3. Toutenburg, Srivastava ve Shalabh tahmin edicisi

Toutenburg vd. (2008) çalışmasında, basit rasgele örnekleme planında, yardımcı değişkene ait kitle bilgisinin bilinmediği durum da dikkate alınarak, oransal veri türetme temelli, ilgilenilen değişkene ilişkin kitle ortalaması için seçenek tahmin ediciler önerilmiştir.

İlgilenilen değişken değerleri $y_1, y_2, \ldots, y_N$ ve yardımcı değişken değerleri $x_1, x_2, \ldots, x_N$ olmak üzere N birimli sonlu kitleden, $\bar{Y}$ kitle ortalaması tahmini için, n birimli basit rasgele örnekleminin seçildiği varsayılmıştır. Cevapsızlığın rasgele olduğu düşünülerek, r tane $(y_1, x_1)\ldots, (y_r, x_r)$ tam veri ve n-r tane ilgilenilen değişkene ait kayıp veriye (KV) karşılık gelen yardımcı değişken verileri mevcuttur. Buna dayanarak, örneklemin r büyüklüklü A ile gösterilen cevap ve n-r büyüklüklü A^c ile gösterilen cevapsız alt kümelerini kapsamaktadır.

$$
\begin{array}{ll}
\underline{x_i} & \underline{y_i} \\
x_1 & \left.\begin{array}{l} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ y_r \end{array}\right\} A \\
x_2 & \\
x_3 & \\
. & \\
. & \\
. & \\
x_r & \\
x_{r+1} & \left.\begin{array}{l} KV \\ KV \\ KV \\ KV \\ KV \end{array}\right\} A^c \\
. & \\
. & \\
. & \\
x_n &
\end{array}
$$

Kayıp veriler dikkate alınmadığı zaman, bilinen kitle ortalaması, $\bar{Y}$, tahmini,

$$\bar{y}_r = \frac{1}{r}\sum_{i=1}^{r} y_i. \tag{3.35}$$

olmaktadır. Eksik veriler tamamlanıp hesaplamaya dahil edildiğinde tam-veri kümesi,

$$z_i = \begin{cases} y_i & i \in A \\ \tilde{y}_i^* & i \in A^c \end{cases}$$

biçiminde olur. Eğer kayıp olan $\tilde{y}_i$ değeri için, oransal tahmin kullanılırsa, iki seçenek söz konusudur:

1) $\tilde{y}_i^* = \bar{y}_r \left(\dfrac{\bar{X}_N}{\bar{x}_r} \right)$ (3.36)

Burada, kayıp olan ilgilenilen değişkenin verileri için, *r* büyüklüklü *s* alt örneklemine göre elde edilen oransal tahminler kullanılmıştır.

2) $\tilde{y}_i^* = \bar{y}_r \left(\dfrac{n\bar{X}_N}{r\bar{x}_r + (n-r)\bar{x}_{n-r}} \right)$ (3.37)

Eşitlik (3.37)'e dikkat edildiğinde, sadece cevap kümesinde yer alan yardımcı değişken değil aynı zamanda *n-r* büyüklüklü cevapsız kümesindeki yardımcı değişken bilgilerinin de kullanıldığı görülmektedir. Burada, $\bar{x}_r = \frac{1}{r}\sum_{i=1}^{r} x_i$, $\bar{x}_{n-r} = \frac{1}{n-r}\sum_{i=1}^{n-r} x_i$ ve $\bar{X} = \bar{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i$ göstermektedir.

Eşitlik (3.36) ve (3.37)'de verilen iki tahmin edicide de, yardımcı değişkene ait kitle bilgisinin bilindiği varsayılmaktadır. Toutenburg vd. (2008) çalışmasında, kitle bilgisinin bilinmediği durum için yukarıda verilen yöntemlere seçenek iki farklı yöntem de önermişlerdir:

3) $\tilde{y}_i^* = \bar{y}_r \left(\dfrac{x_i^c}{\bar{x}_r} \right)$, (3.38)

4) $\tilde{y}_i^* = \bar{y}_r \left(\dfrac{n x_i^c}{r\bar{x}_r + (n-r)\bar{x}_{n-r}} \right)$. (3.39)

Burada, x_i^c, cevapsızların oluşturduğu A^c alt örnekleminde, kayıp olan y_i değerine karşılık gelen yardımcı değişken değeridir. Bu veri türetme ifadelerinden $\bar{Y}$ için aşağıdaki 4 tahmin ediciye de ulaşılır:

$$\bar{y}_{t1} = \bar{y}_r \left[\frac{r\bar{x}_r + (n-r)\bar{X}_N}{n\bar{x}_r} \right] \tag{3.40}$$

Eşitlik (3.40)'da verilen tahmin ediciye,

$$\begin{aligned}
\bar{y}_{t1} &= \frac{1}{n}\left(\sum_{i \in A} y_i + \sum_{i \in A^c} \tilde{y}_i^* \right) \\
&= \frac{1}{n}\left(\sum_{i \in A} y_i + \sum_{i \in A^c} \bar{y}_r \left(\frac{\bar{X}_N}{\bar{x}_r} \right) \right) \\
&= \frac{1}{n}\left(r\bar{y}_r + (n-r)\bar{y}_r \left(\frac{\bar{X}_N}{\bar{x}_r} \right) \right) \\
&= \bar{y}_r \left[\frac{r\bar{x}_r + (n-r)\bar{X}_N}{n\bar{x}_r} \right]
\end{aligned}$$

biçiminde ulaşmak mümkündür. Aşağıdaki tahmin edicilerin de elde edilişi benzer biçimde yapılabilir.

$$\bar{y}_{t2} = \bar{y}_r \left[\frac{r^2\bar{x}_r + n(n-r)\bar{X} + r(n-r)\bar{x}_{n-r}}{rn\bar{x}_r + n(n-r)\bar{x}_{n-r}} \right], \tag{3.41}$$

$$\bar{y}_{t3} = \bar{y}_r \left[\frac{r\bar{x}_r + (n-r)\bar{x}_{n-r}}{n\bar{x}_r} \right], \tag{3.42}$$

$$\bar{y}_{t4} = \bar{y}_r \left[\frac{r^2\bar{x}_r + (2n-(n-r))(n-r)\bar{x}_{n-r}}{rn\bar{x}_r + n(n-r)\bar{x}_{n-r}} \right]. \tag{3.43}$$

Bu tahmin edicilerin yanında, bütün kayıp verilerin göz ardı edildiği, yani mevcut değerler üzerinden alınan, Eşitlik (3.35)'de verilen tahmin edici de (amputasyon yöntemine göre elde edilen) dahil edilirse, kitle ortalaması için toplam 5 tahmin ediciden bahsetmek mümkündür. Hangi tahmin edicinin hangi durumda tercih edilebileceğine, etkinlik karşılaştırması sonucu karar verilmektedir.

Yöntemlerin Karşılaştırılması:

Taylor serisi yöntemine göre elde edilen HKO'a göre, veri türetme ile elde edilen tahmin edicilerin, veri türetme kullanmadan elde edilen tahmin ediciye olan etkinliklerinin araştırılması esas alınmış ve buna göre HKO eşitlikleri karşılaştırılmıştır. Yan ve HKO eşitliklerinin incelenmesinde, gösterimlerin daha rahat verilebilmesi için aşağıdaki ifadelerden yararlanılmıştır:

$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2 , \qquad S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \overline{Y})^2 ,$$

$$\rho = \frac{1}{S_x S_y (N-1)}\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y}) , \qquad \theta = \frac{\overline{Y} S_x}{\overline{X} S_y} ,$$

Kayıplılığın rastgele olduğu durumda, aşağıdaki fonksiyonların beklenen değerleri f_k ve g_k biçiminde aşağıdaki gibi tanımlanmıştır.

$$f_k = E\left[\left(\frac{1}{r}\right)\left(\frac{n-r}{n}\right)^k\right], \qquad g_k = \frac{1}{n}E\left[\left(\frac{n-r}{n}\right)^k\right].$$

Buna göre karşılaştırmalar aşağıda verilmiştir.

- $\bar{y}_{t1}$ ve $\bar{y}_r$ karşılaştırıldığında,

$$\Delta(\bar{y}_r ; \bar{y}_{t1}) = E(\bar{y}_r - \overline{Y})^2 - E(\bar{y}_{t1} - \overline{Y})^2 = S_x S_y \left(\frac{\overline{Y}}{\overline{X}}\right)\left[2 f_1 \rho - f_2 \theta\right]$$

Taylor serisi yöntemi ile HKO eşitliği yaklaşık olarak bulunabilir;

$$h(\hat{Y}_1, \hat{Y}_2) = h(Y_1, Y_2) + \sum_{j=1}^{k} d_j\left(\hat{Y}_j - Y_j\right) + R_k$$

$\hat{Y}_1 = \bar{y}_r$, $\hat{Y}_2 = \bar{x}_r$ olmak üzere,

$$h(\hat{Y}_1, \hat{Y}_2) = \bar{y}_{t_1} = \bar{y}_r \frac{r\bar{x}_r + (n-r)\overline{X}}{n\bar{x}_r}$$

$$h(Y_1, Y_2) = \overline{Y}$$

$$d_1 = \frac{\partial h(\hat{Y}_1, \hat{Y}_2)}{\hat{Y}_1}\Big|_{Y_1, Y_2} = \frac{r\bar{x} + (n-r)\overline{X}}{n\bar{x}_r}\Big|_{\overline{Y}, \overline{X}} = 1$$

$$d_2 = \frac{\partial h(\hat{Y}_1,\hat{Y}_2)}{\hat{Y}_2}\Big|_{Y_1,Y_2} = \bar{y}_r\left[\frac{rn\bar{x}_r - n[r\bar{x} + (n-r)\bar{X}]}{n^2\bar{x}_r^2}\right]\Big|_{\bar{Y},\bar{X}}$$

$$= \bar{Y}\left[\frac{n^2\bar{X} - n(n-r)\bar{X} - n(n\bar{X} - (n-r)\bar{X} + (n-r)\bar{X}]}{n^2\bar{X}^2}\right]$$

$$= \bar{Y}\left[-\frac{n(n-r)\bar{X}}{n^2\bar{X}^2}\right] = -\frac{(n-r)}{n}\frac{\bar{Y}}{\bar{X}}$$

$$\left(\bar{y}_{t1} - \bar{Y}\right) \cong (\bar{y}_r - \bar{Y}) - \frac{(n-r)}{n}\frac{\bar{Y}}{\bar{X}}\left(\bar{x}_r - \bar{X}\right)$$

$$\left(\bar{y}_{t1} - \bar{Y}\right)^2 \cong (\bar{y}_r - \bar{Y})^2 - 2\left(\frac{n-r}{n}\right)\left(\frac{\bar{Y}}{\bar{X}}\right)(\bar{y}_r - \bar{Y})\left(\bar{x}_r - \bar{X}\right) + \left(\frac{n-r}{n}\right)^2\left(\frac{\bar{Y}}{\bar{X}}\right)^2\left(\bar{x}_r - \bar{X}\right)^2$$

Her iki tarafın beklenen değeri alındığında,

$$E\left(\bar{y}_{t1} - \bar{Y}\right)^2 \cong E(\bar{y}_r - \bar{Y})^2 - 2\left(\frac{\bar{Y}}{\bar{X}}\right)E_p\left(\frac{n-r}{n}\right)E(\bar{y}_r - \bar{Y})\left(\bar{x}_r - \bar{X}\right) + \left(\frac{\bar{Y}}{\bar{X}}\right)^2 E_p\left(\frac{n-r}{n}\right)^2 E\left(\bar{x}_r - \bar{X}\right)^2$$

eşitliğine ulaşılır. Kayıplığın rastgele olduğu daha önce belirtilmişti. Buna göre *(n-r)* bir rastlantı değişkenidir. Aynı zamanda, *(n-r)*'nin $\bar{y}_r$ ve $\bar{x}_r$'dan bağımsız olduğu düşünülerek yukarıdaki beklenen değer formuna ulaşılmıştır. Gösterim kolaylığı olması açısından verilen ifadeler kullanılarak gerekli sadeleştirmeler yapıldığında;

$$E\left(\bar{y}_{t1} - \bar{Y}\right)^2 \cong E(\bar{y}_r - \bar{Y})^2 - 2\left(\frac{\bar{Y}}{\bar{X}}\right)E_p\left(\frac{n-r}{n}\right)\frac{1}{r}S_{xy} + \left(\frac{\bar{Y}}{\bar{X}}\right)^2 E_p\left(\frac{n-r}{n}\right)^2\frac{1}{r}S_x^2,$$

$$E(\bar{y}_r - \bar{Y})^2 - E\left(\bar{y}_{t1} - \bar{Y}\right)^2 \cong 2\left(\frac{\bar{Y}}{\bar{X}}\right)f_1\, S_{xy} - \left(\frac{\bar{Y}}{\bar{X}}\right)^2 f_2\, S_x^2$$

$$= 2\left(\frac{\bar{Y}}{\bar{X}}\right)f_1\,\rho\, S_x\, S_y - \left(\frac{\bar{Y}}{\bar{X}}\right)^2 f_2\, S_x^2\,\frac{S_y}{S_y}$$

$$= 2\left(\frac{\bar{Y}}{\bar{X}}\right)f_1\,\rho\, S_x\, S_y - \left(\frac{\bar{Y}}{\bar{X}}\right)f_2\,\theta\, S_x S_y$$

$$= S_x S_y\left(\frac{\bar{Y}}{\bar{X}}\right)\left[2 f_1\rho - f_2\theta\right]$$

elde edilir. Yanın ihmal edildiği ve kitle ortalamalarının pozitif olduğu varsayımı altında, $\bar{y}_r$ ve t_1 tahmin edicilerinin varyansları karşılaştırıldığında,

$$\rho > \left(\frac{f_2}{2f_1}\right)\theta \tag{3.44}$$

koşulu altında, $\bar{y}_{t1}$ varyansının $\bar{y}_r$ varyansından küçük olduğu görülmektedir. Burada, oransal tahmin kullanmamıza neden olan *Y* ve *X*'in pozitif ilişkili olması durumu düşünüldüğünde, ρ'nun pozitif olduğu dikkate alınmalıdır.

Aynı şekilde diğer tahmin edicilere ilişkin varyanslar bulunmuş ve benzer olarak karşılaştırmalar yapılmıştır.

- $\bar{y}_{t2}$ ve $\bar{y}_r$ karşılaştırıldığında,

$$\Delta(\bar{y}_r;\bar{y}_{t2}) = E(\bar{y}_r-\bar{Y})^2 - E(\bar{y}_{t2}-\bar{Y})^2 = S_x S_y\left(\frac{\bar{Y}}{\bar{X}}\right)\left[2g_1\rho - g_2\theta\right]$$

elde edilir. Buradan

$$\rho > \left(\frac{g_2}{2g_1}\right)\theta \tag{3.43}$$

koşulu altında, $\bar{y}_{t2}$ varyansının $\bar{y}_r$ varyansından küçük olduğu görülmektedir.

- $\bar{y}_{t3}$ ve $\bar{y}_r$ karşılaştırıldığında,

$$\Delta(\bar{y}_r;\bar{y}_{t3}) = E(\bar{y}_r-\bar{Y})^2 - E(\bar{y}_{t3}-\bar{Y})^2 = S_x S_y\left(\frac{\bar{Y}}{\bar{X}}\right)\left[2\rho - \theta\right]f_1$$

elde edilir. Buradan

$$\rho > \frac{\theta}{2} \tag{3.46}$$

koşulu altında, $\bar{y}_{t3}$ varyansının $\bar{y}_r$ varyansından küçük olduğu görülmektedir. Bilindiği gibi bu koşul aynı zamanda, kayıp veri olmadığı durumda, kitle ortalaması tahmininde, oransal tahminin basit tahmine olan üstünlük koşuludur.

- $\bar{y}_{t4}$ ve $\bar{y}_r$ karşılaştırıldığında,

$$\Delta(\bar{y}_r;\bar{y}_{t4}) = E(\bar{y}_r-\bar{Y})^2 - E(\bar{y}_{t4}-\bar{Y})^2 = S_x S_y\left(\frac{\bar{Y}}{\bar{X}}\right)\left[2g_1\rho-(g_1-g_2)\theta\right]$$

elde edilir ve $\bar{y}_{t4}$ tahmin edicisi,

$$\rho > \left(\frac{g_1-g_2}{2g_1}\right)\theta \tag{3.47}$$

koşulu altında $\bar{y}_r$'dan daha etkin bir tahmin edicidir. Aynı karşılaştırmalar, $\bar{X}$'nın bilindiği ve bilinmediği durumlar için ayrı ayrı yapılabilir. Buna göre, $\bar{X}$'nın bilindiği durumda da $\bar{y}_{t1}$ ve $\bar{y}_{t2}$ tahmin edicilerini kullanabiliriz. $\bar{y}_{t1}$ tahmin edicisi sadece tam verileri kullanırken, $\bar{y}_{t2}$ tahmin edicisi eksik gözlenen tüm verileri, yani ilgilenilen değişkene ait kayıp veriye karşı gelen yardımcı değişkenin değerlerini de kullanır. Bu durumda, $\bar{y}_{t2}$ tahmin edicisinin hem daha küçük bir yana hem de aşağıdaki koşul altında daha küçük bir varyansa sahip olduğu görülmektedir:

$$\rho < \frac{1}{2}\left(\frac{f_2-g_2}{f_1-g_1}\right)\theta\ . \tag{3.48}$$

Benzer şekilde, $\bar{X}$'nın bilinmediği durumda $\bar{y}_{t3}$ ve $\bar{y}_{t4}$ gibi, tüm elde edilen verileri dikkate alan iki tahmin edici söz konusudur. Bu tahmin edicilerin ikisi de yanlıdır fakat aşağıdaki koşul altında $\bar{y}_{t3}$ daha küçük bir HKO'a sahiptir:

$$\rho > \frac{1}{2}\left(\frac{f_1-g_1+g_2}{f_1-g_1}\right)\theta\ . \tag{3.49}$$

Aynı şekilde bu tahmin edicilerin yanı da incelenebilir. Buna göre,

$$Yan(\bar{y}_{t1}) \cong E(\bar{y}_{t1}-\bar{Y}) \cong f_1(\theta-\rho)\frac{S_x S_y}{\bar{X}}, \tag{3.50}$$

$$Yan(\bar{y}_{t2}) \cong E(\bar{y}_{t2}-\bar{Y}) \cong g_1(\theta-\rho)\frac{S_x S_y}{\bar{X}}, \tag{3.51}$$

$$Yan(\bar{y}_{t3}) \cong E(\bar{y}_{t3}-\bar{Y}) \cong f_1(\theta-\rho)\frac{S_x S_y}{\bar{X}}, \tag{3.52}$$

$$Yan(\bar{y}_{t4}) \cong E(\bar{y}_{t4} - \bar{Y}) \cong g_1 \rho \frac{S_x S_y}{\bar{X}}. \tag{3.53}$$

Yan eşitliklerinin çıkarımı EK 3'te verilmiştir. Eşitlik (3.50)-(3.53)'de $\rho = \theta$ olduğu durumda $\bar{y}_{t1}$, $\bar{y}_{t2}$, $\bar{y}_{t3}$ tahmin edicilerinin yansız olduğu görülmektedir. Eğer $\theta > 1$ veya $\rho < \theta \le 1$ ise, bu tahmin ediciler pozitif yana sahiptir. Sonuç olarak, Toutenburg vd. (2008) çalışmasında önerilen tahmin edici formüllerinde, $\bar{X}$'nın bilinmesi büyük rol oynamaktadır. $\bar{y}_{t1}$ ve $\bar{y}_{t2}$ tahmin edicilerinde $\bar{X}$'nın yerine $\bar{x}_{n-r}$ kullanılarak t_3 ve t_4 tahmin edicileri elde edilmektedir. Yukarıdaki karşılaştırmalar göz önünde bulundurularak, araştırmanın kapsamı doğrultusunda uygun tahmin ediciye karar verilir.

3.2.1.4. Kadılar ve Çıngı tahmin edicileri

Kadılar ve Çıngı (2008) çalışmalarında, kayıp veri olması durumunda kitle ortalaması tahmini için, regresyon tipi oransal tahmin edici ailesi önermişlerdir. Buna göre, *N* büyüklüğünde bir kitleden basit rasgele örnekleme ile çekilen *n* büyüklüğündeki örneklemde, *r* elemanlı cevaplı birimlerin oluşturduğu alt örneklem ortalamalarını göstermek üzere ilk tahmin edici, ilgilenilen ve yardımcı değişkende elde edilen tüm bilgiyi kullanan,

$$\bar{y}_{KC1} = \frac{\bar{y}_r + b\left(\bar{X} - \bar{x}_n\right)}{\bar{x}_n}\bar{X} \tag{3.54}$$

tahmin edicisidir. İkinci tahmin edici ise, örneklemde ilgilenilen değişkende cevaplılara karşılık gelen yardımcı değişken bilgisini kullanan aşağıdaki tahmin edicidir:

$$\bar{y}_{KC2} = \frac{\bar{y}_r + b\left(\bar{X} - \bar{x}_r\right)}{\bar{x}_r}\bar{X}. \tag{3.55}$$

Yardımcı değişkene ait kitle ortalaması bilinmediği durum için ise, Kadılar ve Çıngı (2008) aşağıdaki tahmin ediciyi önermişlerdir:

$$\bar{y}_{KC3} = \frac{\bar{y}_r + b(\bar{x}_n - \bar{x}_r)}{\bar{x}_r}\bar{x}_n . \tag{3.56}$$

Burada, $b = \frac{s_{xy}}{s_x^2}$ örneklem regresyon katsayısını, s_x^2 yardımcı değişkene ait örneklem varyansını, s_{xy} yardımcı değişken ve ilgilenilen değişken arasındaki örneklem kovaryansını, $\bar{x}_r = \frac{1}{r}\sum_{i=1}^{r} x_i$, $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y}_r = \frac{1}{r}\sum_{i=1}^{r} y_i$ göstermektedir.

Kadılar ve Çıngı (2008) tahmin edicilerinin yan ve HKO eşitlikleri sırasıyla aşağıda verildiği gibidir:

$$Yan(\bar{y}_{KC1}) \cong \left(\frac{1}{n} - \frac{1}{N}\right)\bar{Y}C_x^2 , \tag{3.57}$$

$$Yan(\bar{y}_{KC2}) \cong \left(\frac{1}{r} - \frac{1}{N}\right)\bar{Y}C_x^2 , \tag{3.58}$$

$$Yan(\bar{y}_{KC3}) \cong \left(\frac{1}{r} - \frac{1}{n}\right)\bar{Y}\rho C_x C_y , \tag{3.59}$$

$$HKO(\bar{y}_{KC1}) \cong \left(\frac{1}{r} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{n} - \frac{1}{N}\right)S_x^2\left(R^2 - B^2\right), \tag{3.60}$$

$$HKO(\bar{y}_{KC2}) \cong \left(\frac{1}{r} - \frac{1}{N}\right)\left(S_y^2 - BS_{xy} + R^2 S_x^2\right), \tag{3.61}$$

$$HKO(\bar{y}_{KC3}) \cong \left(\frac{1}{r} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{r} - \frac{1}{n}\right)\left[(R+B)^2 S_x^2 - 2(R+B)S_{xy}\right]. \tag{3.62}$$

Kadılar ve Çıngı (2008) çalışmalarında, önerdikleri tahmin edicileri $\bar{y}_{ORAN} = \frac{\bar{y}_r}{\bar{x}_r}\bar{x}_n$ klasik oransal tahmin edici ile ve Eşitlik (3.20)'de verilen Singh ve Deo (2003) ve Eşitlik (3.29)'da verilen Singh ve Horn (2000) tahmin edicileriyle karşılaştırmışlar ve belli koşullar altında önerdikleri tahmin edicilerin her zaman daha etkin olduğunu göstermişlerdir (Ayrıntılar için bkz. Kadılar ve Çıngı, 2008).

3.2.1.5. Perri ve Diana tahmin edicileri

Perri ve Diana (2009) çalışmalarında kitle ortalaması tahmini için, Eşitlik (3.54)-Eşitlik (3.56) ile verilen Kadılar ve Çıngı tahmin edicilerine paralel üç tane yeni regresyon tahmin edici önermişlerdir:

$$\bar{y}_{DP1} = \bar{y}_r + b\left(\bar{X} - \bar{x}_n\right), \tag{3.63}$$

$$\bar{y}_{DP2} = \bar{y}_r + b\left(\bar{X} - \bar{x}_r\right), \tag{3.64}$$

$$\bar{y}_{DP3} = \bar{y}_r + b\left(\bar{x}_n - \bar{x}_r\right). \tag{3.65}$$

Bu tahmin ediciler, üç farklı regresyon tipi veri türetme yöntemine bağlıdır. Veri türetmeden sonra verilerin görüntüsü sırasıyla aşağıdaki gibidir (Perri ve Diana, 2009):

$$y^*_{i,DP1} = \begin{cases} \dfrac{ny_i}{r} + b\left(\bar{X} - x_i\right), & i \in A \\ b\left(\bar{X} - x_i\right), & i \in A^c \end{cases} \tag{3.66}$$

$$y^*_{i,DP2} = \begin{cases} \dfrac{ny_i}{r} - b\dfrac{nx_i}{r}, & i \in A \\ b\dfrac{n\bar{X}}{n-r}, & i \in A^c \end{cases} \tag{3.67}$$

$$y^*_{i,DP3} = \begin{cases} \dfrac{ny_i}{r} - b\dfrac{nx_i}{r}, & i \in A \\ b\dfrac{n\bar{x}_n}{n-r}, & i \in A^c \end{cases} \tag{3.68}$$

Diana ve Perri tahmin edicilerinin HKO eşitlikleri ise aşağıdaki gibidir:

$$HKO(\bar{y}_{DP1}) \cong S_y^2\left[\left(\frac{1}{n}-\frac{1}{N}\right)\left(1-\rho^2\right)+\left(\frac{1}{r}-\frac{1}{n}\right)\right], \tag{3.69}$$

$$HKO(\bar{y}_{DP2}) \cong \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2\left(1-\rho^2\right), \tag{3.70}$$

$$HKO(\bar{y}_{DP3}) \cong S_y^2\left[\left(\frac{1}{n}-\frac{1}{N}\right)+\left(\frac{1}{r}-\frac{1}{n}\right)\left(1-\rho^2\right)\right]. \tag{3.71}$$

Bu tahmin edicilerin HKO eşitliklerinin elde edilişi EK5'de verilmiştir.

Diana ve Perri tahmin edicileri kendi içinde karşılaştırıldığında, en etkin tahmin edici Eşitlik (3.64) ile verilen $\bar{y}_{DP2}$ tahmin edicisidir. Eşitlik (3.63)-Eşitlik (3.65) ile verilen tahmin ediciler, bunlara paralel Eşitlik (3.54)-Eşitlik (3.56)'da verilen Kadılar ve Çıngı tahmin edicileriyle karşılaştırıldığında, regresyon tipi tahmin edici olmalarından ötürü Diana ve Perri tahmin edicilerinin her zaman daha etkin olduğu görülmektedir (Diana ve Perri, 2009).

3.2.2. Yardımcı değişkende kayıp veri olan tahmin ediciler

Birçok araştırmacı tarafından, bundan önceki bölümlerde de verildiği gibi, bazı verilerin kayıp olduğu ve yardımcı değişkenin kitle ortalamasının bilindiği durumda, yerine konmadan basit rastgele örneklemede dolaylı tahmin ediciler üzerine araştırmalar yapılmıştır. Bu bölümde ise, hem ilgilenilen değişkende hem de yardımcı değişkende kayıp veri olması durumu ile ilgili çalışmalar incelenecektir.

3.2.2.1. Rueda, Gonzalez, Arcos, Roman, Martinez ve Munoz tahmin edicisi

Rueda vd. (2005) çalışmalarında, hem ilgilenilen değişkende hem de yardımcı değişkende bilgi kaybı olması durumunda, kitle ortalaması tahmini için kullanılabilecek yeni bir tahmin edici önermişlerdir. *N* birimli bir kitleden, *n* büyüklüklü *s* basit rasgele örnekleminin çekildiği bu örneklemde sadece (*n*-*p*-*q*) birimden tam gözlemin elde edilebildiği varsayılmaktadır. Buna ek olarak, örneklemin *p* biriminde *X* değişkenine ait gözlemler elde edilebilir fakat *Y* değişkeninde buna karşılık gelen veriler (:KV) kayıptır. Aynı şekilde, örneklemde, *Y* değişkenine ait *q* gözlem bilgisi mevcut fakat *X* değişkeninde bunlara karşılık değerler kayıptır. Burada, *p* ve *q*'nun $0<p,\ q<n/2$ koşullarını sağlayan tamsayı oldukları varsayılmaktadır.

Yukarıda bahsedilen *s* örneklem birimlerinin görüntüsü, aşağıda gösterildiği gibidir:

$s_1=\{i \in s/\ x_i,\ y_i$ elde edilebilir$\}$,
$s_2=\{i \in s/\ x_i$ elde edilebilir, y_i kayıp$\}$,
$s_3=\{i \in s/\ y_i$ elde edilebilir, x_i kayıp$\}$.

$$\begin{array}{ll}
\underline{x_i} & \underline{y_i}
\end{array}$$
$$\left.\begin{array}{ll}
x_1 & KV \\
x_2 & KV \\
x_3 & KV \\
. & . \\
. & . \\
. & . \\
x_p & KV
\end{array}\right\} p \text{ birimli } s_2 \text{ alt kümesi}$$
$$\left.\begin{array}{ll}
KV & y_{p+1} \\
. & . \\
. & . \\
. & . \\
KV & y_{p+q}
\end{array}\right\} q \text{ birimli } s_3 \text{ alt kümesi}$$
$$\left.\begin{array}{ll}
x_{p+q+1} & y_{p+q+1} \\
x_{p+q+2} & y_{p+q+2} \\
. & . \\
. & . \\
. & . \\
x_n & y_n
\end{array}\right\} n\text{-}p\text{-}q \text{ birimli } s_1 \text{ alt kümesi}$$

Rueda vd.(2005), kitle ortalamasının, örneklem birimlerinin ve örneklem dışında kalan birimlerin ortalamasının ağırlıklandırılmış toplamı olduğunu dikkate alarak, *B*'nin en küçük varyanslı doğrusal yansız, *b*, regresyon tahmin edicisine bağlı aşağıdaki tahmin ediciyi sunmuşlardır:

$$T^{*} = k_1\bar{y}_{s1} + k_2\bar{y}_{s3} + k_3 b\,\bar{x}_{s2}\,. \tag{3.72}$$

Burada, $k_1 = \dfrac{n-p-q(N-p)}{N(n-p)}$, $k_2 = \dfrac{q(N-p)}{N(n-p)}$ ve $k_3 = \dfrac{p}{N}$ göstermektedir. $\bar{y}_{s1}, \bar{y}_{s3}, \bar{x}_{s2}$ alt indislerde gösterilen alt örneklem parçalarına ait örneklem ortalamalarıdır.

Önerilen tahmin edicinin Yan ve HKO eşitlikleri, *s* örnekleminin basit rasgele örnekleme planına göre çekildiği varsayımıyla aşağıda verilmiştir:

$$Yan(T^{*}) \cong k_3(B\bar{X} - \bar{Y}), \tag{3.73}$$

$$HKO(T^{*}) \cong S_y^2\left[\ k_1^2 a + k_2^2\alpha + 2k_1k_2c\ \right] + B^2k_3^2S_x^2 d + 2k_3BS_{xy}\left[\ k_1e + k_2 f\ \right]. \tag{3.74}$$

Bu tahmin edicinin Yan ve HKO eşitliklerinin elde edilişi EK 6'da verilmiştir. Eşitlik (3.73) ve (3.74)'deki *a*, α, *c*, *d*, *e*, *f* katsayıları EK 6'da tanımlandığı gibi alınmaktadır.

Eşitlik (3.73)'den de görüldüğü gibi, yanın sıfır olabilmesi için aşağıdaki koşulun sağlanması gereklidir:

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

$$\frac{S_{xy}}{S_x^2}\bar{X} = \bar{Y} \rightarrow \frac{\rho S_x S_y}{S_x^2}\bar{X} = \bar{Y}$$

$$\frac{\rho S_y}{S_x/\bar{X}} = \bar{Y} \rightarrow \frac{\rho S_y}{\bar{Y}} = C_x$$

Sonuç olarak, yanın sıfır olması aşağıdaki koşula dönüşmüştür:

$$\rho = \frac{C_x}{C_y} \Rightarrow Yan(T^{*}) = 0\,.$$

Satıcı ve Kadılar (2009) çalışmasında, Eşitlik (3.72)'de verilen Rueda vd. (2005) tahmin edicisi, yardımcı değişkende kayıp veri olmadığı varsayımı altında düzenlenerek, aşağıdaki ikinci tahmin edici önerilmiştir:

$$\bar{y}_{SK2} = l_1\bar{y}_{s1} + l_2 b\bar{x}_{s2}\,. \tag{3.75}$$

Burada, yardımcı değişkende kayıp veri olmadığı durumda örneklem iki parçada incelenmektedir. İlgilenilen değişkeninin oluşturduğu gözlemlenen 1. örneklem parçası, *r* elemanlı, gözlenmeyen 2. örneklem parçası *n-r* elemanlıdır. $\bar{y}_{s1}$, 1. örneklem parçasındaki ilgilenilen değişkeninin ortalamasını, $\bar{x}_{s2}$ ise 2. örneklem parçasındaki yardımcı değişkeninin ortalamasını, $l_1 = \frac{N-n+r}{N}$ ve $l_2 = \frac{n-r}{N}$ göstermektedir.

Eşitlik (3.75)'de verilen tahmin edicinin, *HKO* eşitliği aşağıda verilmiştir:

$$HKO(\bar{y}_{SK2}) \cong l_1^2\mu S_y^2 + l_2^2\lambda B^2 S_x^2 + 2l_1 l_2 \tau B S_{xy}\,. \tag{3.76}$$

Burada S_{xy}, ilgilenilen değişken ile yardımcı değişken arasındaki kitle kovaryansını göstermekte ve μ, λ, τ katsayıları da aşağıdaki gibi tanımlanmaktadır:

$$\mu = \frac{1}{r} - \frac{1}{N}\,,$$

$$\lambda = \frac{1}{n-r} - \frac{1}{N}\,,$$

$$\tau = \begin{cases} \mu, & n - \mathrm{r} \le n/2 \\ \lambda, & n - \mathrm{r} > n/2 \end{cases}$$

Eşitlik (3.76)'da verilen HKO eşitliğini elde edebilmek için,

$V(\bar{y}_{s1}) = \left(\frac{1}{r} - \frac{1}{N}\right) S_y^2$, $V(\bar{x}_{s2}) = \left(\frac{1}{n-r} - \frac{1}{N}\right) S_x^2$ ve

$$Cov(\bar{y}_{s1}, \bar{x}_{s2}) = \begin{cases} \left(\frac{1}{r} - \frac{1}{N}\right) S_{xy} & , n\text{ - r} \le n/2 \\ \left(\frac{1}{n-r} - \frac{1}{N}\right) S_{xy} & , n\text{ - r} > n/2 \end{cases}$$

olmak üzere Eşitlik (3.75)'de verilen tahmin edicinin varyansı alındığında,

$$HKO(\bar{y}_{SK2}) \cong l_1^2 V\left(\bar{y}_{s_1}\right) + l_2^2 B^2 V\left(\bar{x}_{s2}\right) + 2 l_1 l_2 B Cov\left(\bar{y}_{s1}, \bar{x}_{s2}\right)$$

yazılabilir. Burada varyans ve kovaryans ifadeleri yerine yazıldığında Eşitlik (3.76)'da verilen *HKO* eşitliği elde edilir.

Eşitlik (3.75)'de verilen tahmin edici, yardımcı değişkende kayıp veri olmayan Eşitlik (3.20)'de verilen Singh ve Deo (2003) tahmin edicisi ile karşılaştırılmış ve *n-r*<*n*/2 varsayımı altında aşağıda verilen koşulda, Singh ve Deo (2003) tahmin edicisinden daha etkin olduğu görülmüştür (Satıcı ve Kadılar, 2009):

$$\rho^2 < \frac{n(N-r)[2N-(n-r)]}{nr(n-r) + N\left[N^2 + n(2N-2n+r)\right]}. \tag{3.77}$$

Karşılaştırmaya ilişkin ayrıntı EK 7'de verilmiştir.

3.2.2.2. Gonzalez, Rueda ve Arcos tahmin edicisi

Gonzalez vd. (2008) kitle ortalaması tahmini için, birden çok yardımcı değişken bilgisinden yararlanmış ve bu yardımcı değişkenlerde ayrı zamanlarda meydana gelen kayıplılık durumunu ele almışlardır. *X* ve Z'nin yardımcı değişken olduğu ve bunlara ilişkin $\bar{X}$ ve $\bar{Z}$ kitle ortalamalarının bilindiği varsayılmaktadır. Buna göre, *Y* ilgilenilen değişkene ilişkin $\bar{Y}=\sum_{i=1}^{N} y_i / N$ ortalamasının tahmini için çekilen s örnekleminde, üç değişken bilgisi gözlenir, (y_i, x_i, z_i), i=1,2,…,n.

Gonzalez vd.(2008), kayıp veri olması durumunda kitle ortalaması tahmini için, Rao (1994) tahmin edicisine alternatif yeni bir tahmin edici sunmuşlardır. Bütün verilerin elde edildiği ideal koşullar altında, Rao tarafından sunulan tahmin edici aşağıda verilmiştir:

$$\bar{y}_{Rao}=\bar{y}_{HT}+\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\sigma}}'\begin{pmatrix}\bar{X}-\bar{x}_{HT}\\ \bar{Z}-\bar{z}_{HT}\end{pmatrix}.$$

Burada, $\bar{x}_{HT}$, $\bar{y}_{HT}$ ve $\bar{z}_{HT}$ Horvitz-Thompson tahmin edicileri, $\hat{\boldsymbol{\Sigma}}$ kovaryans matrisinin ve $\hat{\boldsymbol{\sigma}}$ varyans vektörünün yansız tahmin edicileridir. Gonzalez vd. (2008) çalışmasında, kayıp veri yapısının ve örneklem görüntüsünün aşağıdaki biçimde olduğu varsayılmıştır:

$y_i : y_1 \dots y_{n-p-q-k}$	$y_{n-p-q-k+1} \dots y_{n-q-k}$	$y_{n-q-k+1} \dots y_{n-k}$	$KV \dots . KV$
$x_i : x_1 \dots x_{n-p-q-k}$	$x_{n-p-q-k+1} \dots x_{n-q-k}$	$KV \dots . KV$	$x_{n-k+1} \dots x_n$
$z_i : z_1 \dots z_{n-p-q-k}$	$KV \dots . KV$	$z_{n-q-k+1} \dots z_{n-k}$	$z_{n-k+1} \dots z_n$

$s_1=\{i \in s / x_i, y_i, z_i$ elde edilebilir$\}$,

$s_2=\{i \in s / x_i, y_i$ elde edilebilir, z_i kayıp$\}$,

$s_3=\{i \in s / y_i, z_i$ elde edilebilir, x_i kayıp$\}$,

$s_4=\{i \in s / x_i, z_i$ elde edilebilir, y_i kayıp$\}$.

Burada, sadece gözlenen tam veri kümesi üzerinden çalışma yapılması durumunda, örneklem büyüklüğü *n-p-q-k* birim olmakta, dolayısıyla örneklem hatası artmaktadır. Gonzalez vd., bütün gözlenen verileri dikkate alarak Rao (1994) tahmin edicisine alternatif aşağıdaki tahmin ediciyi sunmuşlardır:

$$\bar{y}_{GVF}=\bar{y}_{GV}+\mathbf{B}'\begin{pmatrix}\bar{X}-\bar{x}_{GV}\\ \bar{Z}-\bar{z}_{GV}\end{pmatrix}. \tag{3.78}$$

Burada $\bar{y}_{GV}, \bar{x}_{GV}, \bar{z}_{GV}$ ilgili değişkenlere karşılık gelen Horvitz-Thompson tahmin edicileri,

$$\bar{y}_{GV} = \sum_{i \in s_1 \cup s_2 \cup s_3} \frac{y_i}{\pi_i}, \quad \bar{x}_{GV} = \sum_{i \in s_1 \cup s_2 \cup s_4} \frac{x_i}{\pi_i} \quad \text{ve} \quad \bar{z}_{GV} = \sum_{i \in s_1 \cup s_3 \cup s_4} \frac{z_i}{\pi_i}$$

biçimindedir. Eşitlik (3.78)'deki tahmin edicinin varyansını en küçükleyen **B** optimal değeri $\mathbf{B}_{opt} = \mathbf{\Sigma}^{-1}\boldsymbol{\sigma}$ biçiminde bulunmaktadır. Burada,

$$\mathbf{\Sigma} = \begin{pmatrix} V(\bar{x}_{GV}) & Cov(\bar{x}_{GV}, \bar{z}_{GV}) \\ Cov(\bar{x}_{GV}, \bar{z}_{GV}) & V(\bar{z}_{GV}) \end{pmatrix} \text{ ve } \boldsymbol{\sigma} = \begin{pmatrix} Cov(\bar{y}_{GV}, \bar{x}_{GV}) \\ Cov(\bar{y}_{GV}, \bar{z}_{GV}) \end{pmatrix}$$

olmaktadır. Tahmin ediciler, genellikle bilinmeyen s_1, s_2, s_3, s_4'e bağlı ortalama tahmin ediciler arasındaki varyans–kovaryans değerlerine bağlıdır. Örneklem bilgileri kullanılarak, Horvitz-Thompson tahmin edicilerinin varyans–kovaryans tahmin edicilerinden, $\hat{\mathbf{\Sigma}}$ ve $\hat{\boldsymbol{\sigma}}$ hesaplanabilir.

$\mathbf{\Sigma}$ ve $\boldsymbol{\sigma}$ yerine tahmin edicileri, sırasıyla $\hat{\mathbf{\Sigma}}$ ve $\hat{\boldsymbol{\sigma}}$ kullanılarak Eşitlik (3.78)'deki formül aşağıdaki gibi alınabilir:

$$\bar{y}_{GVFopt} = \bar{y}_{GV} + \hat{\mathbf{\Sigma}}^{-1}\hat{\boldsymbol{\sigma}}^{-1} \begin{pmatrix} \bar{X} - \bar{x}_{GV} \\ \bar{Z} - \bar{z}_{GV} \end{pmatrix}. \tag{3.79}$$

Eğer basit rastgele örneklemeye göre örneklem çekimi yapılırsa, olasılıklar eşit olacağından $\mathbf{\Sigma}$ ve $\boldsymbol{\sigma}$ ifadeleri sırasıyla aşağıdaki biçimde elde edilir:

$$Cov(\bar{x}_{GV}, \bar{z}_{GV}) = \frac{S_x^2}{n-q}\left(1 - \frac{n-q}{N}\right), \quad Var(\bar{z}_{GV}) = \frac{S_z^2}{n-p}\left(1 - \frac{n-p}{N}\right),$$

$$\begin{aligned} Cov(\bar{x}_{GV}, \bar{z}_{GV}) = {} & \frac{(n-p-q)^2}{(n-q)(n-p)} Cov(\bar{x}_{s_1 \cup s_4}, \bar{z}_{s_1 \cup s_4}) + \frac{p(n-p-q)}{(n-q)(n-p)} Cov(\bar{x}_{s_{24}}, \bar{z}_{s_1 \cup s_4}) \\ & + \frac{q(n-p-q)}{(n-q)(n-p)} Cov(\bar{x}_{s_1 \cup s_4}, \bar{z}_{s_3}) + \frac{pq}{(n-q)(n-p)} Cov(\bar{x}_{s_2}, \bar{z}_{s_3}) \end{aligned}$$

$$Cov(\bar{y}_{GV},\bar{x}_{GV}) = \frac{(n-q-k)^2}{(n-k)(n-q)} Cov(\bar{y}_{s_1\cup s_2},\bar{x}_{s_1\cup s_2}) + \frac{k(n-q-k)}{(n-k)(n-q)} Cov(\bar{y}_{s_1\cup s_2},\bar{x}_{s_4})$$
$$+ \frac{q(n-q-k)}{(n-k)(n-q)} Cov(\bar{y}_{s_3},\bar{x}_{s_1\cup s_2}) + \frac{qk}{(n-k)(n-q)} Cov(\bar{y}_{s_3},\bar{x}_{s_4})$$

ve

$$Cov(\bar{y}_{GV},\bar{z}_{GV}) = \frac{(n-p-k)^2}{(n-k)(n-p)} Cov(\bar{y}_{s_1\cup s_3},\bar{z}_{s_1\cup s_3}) + \frac{k(n-p-k)}{(n-k)(n-p)} Cov(\bar{y}_{s_1\cup s_3},\bar{z}_{s_4})$$
$$+ \frac{p(n-p-k)}{(n-k)(n-p)} Cov(\bar{y}_{s_2},\bar{z}_{s_1\cup s_3}) + \frac{pk}{(n-k)(n-p)} Cov(\bar{y}_{s_2},\bar{z}_{s_4})$$

(Gonzalez vd, 2008). Burada verilen s_1, s_2, s_3, s_4'e bağlı Horvitz-Thompson tahmin edicileri arasındaki varyans ve kovaryansları, örneklemden tahmin edilebilir.

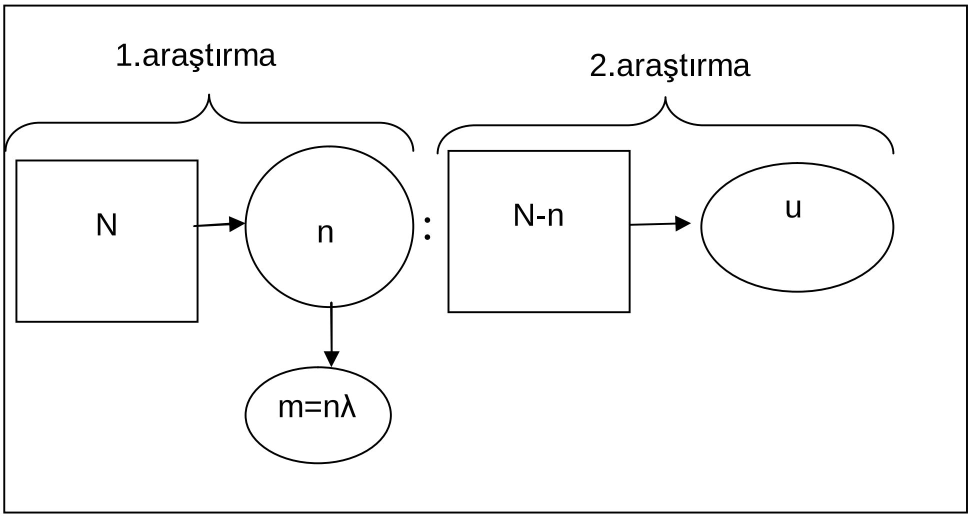
Gonzalez vd., simulasyon çalışması ile, cevapsızlık olması durumunda, önerdikleri tahmin edicinin Rao (1994) tahmin edicisinden daha etkin olduğunu göstermişlerdir.

DÖRDÜNCÜ BÖLÜM

4. ARD ARDA ÖRNEKLEME YÖNTEMİ

Birçok araştırmada, aynı kitle ve aynı ilgilenilen değişken ile ilgili araştırmalar tekrarlanmaktadır. Bu tür araştırmalarda, daha önce yapılmış araştırma bilgilerini kullanan ard arda örnekleme yöntemini kullanmak avantaj sağlamaktadır. Bu yöntemde önemli olan, daha önceki araştırmadan ne kadarlık bir örneklemin güncel araştırmaya dahil edileceği ve aynı zamanda son araştırmada, dahil edilen örneklemden başka ne kadarlık bir örneklemin yeniden inceleneceğidir. Literatürde buna “optimum yenileme ilkesi” (optimum replacement policy) denilmektedir.

$U=(u_1, u_2,\ldots,u_N)$ N elemanlı sonlu kitleden, iki araştırma için örneklem çekilmektedir. Y ilgilenilen değişken, bunun dışındaki değişkenler ise yardımcı değişkenlerdir (X,Z,..). İlk araştırmada, yerine konmadan basit rasgele örnekleme ile n birimlik örneklem çekilir. Buradan rasgele $m=n\lambda$ kadar bir kısım ikinci araştırmada kullanılmak üzere ayrılır (match: eşleştirilen kısım). İkinci araştırmada ise, kalan N-n birimden $u=(n-m)=n\mu$ birimlik yeni örneklem yerine koymadan basit rasgele örnekleme ile çekilir. Sonuç olarak, ilk araştırmada ayrılan kısım da dahil edilince, ikinci araştırmada örneklem büyüklüğü n olur. Burada λ, eşleştirme oranı (fraction of matched) ve μ, ikinci araştırmada yeni örneklem oranı (fraction of fresh samples at the second_current occasion) olarak adlandırılır. Bu örneklem görüntüsü Şekil (4.1)’de görülmektedir.



Şekil 4.1. Ard arda örnekleme yöntemi görüntüsü

4.1. Kitle Ortalaması Tahmini İçin Yapılmış Çalışmalar

4.1.1. Singh tahmin edicisi

Singh (2005) $\overline{Y}$ kitle ortalaması tahmini için, iki tane bağımsız tahmin edici önermiştir. Bunlardan birincisi ikinci araştırmada yeni çekilen *u*=*n*μ örneklemini temel alan

$$T_{1s} = \frac{\bar{y}_u}{\bar{z}_u}\overline{Z} \qquad (4.1)$$

oransal tahmin edicisidir. İkinci tahmin edici ise, iki araştırmanın ortak *m*=*n*λ birimlik örneklem parçasını esas alan aşağıdaki zincirleme oransal tahmin edicidir:

$$T_{2s} = \frac{\bar{y}_m}{\bar{x}_m}\frac{\bar{x}_n}{\bar{z}_n}\overline{Z}\ . \qquad (4.2)$$

Burada *Y*, ilgilenilen değişken, *X* ve *Z* yardımcı değişkenlerdir. Z yardımcı değişkeninin kitle ortalamasının bilindiği varsayılmaktadır. $\bar{x}_n, \bar{z}_n, \bar{y}_m, \bar{x}_m, \bar{y}_u, \bar{z}_u$ ise, alt indislerde gösterilen örneklem büyüklüklerinde, değişkenlerin örneklem ortalamalarıdır.

Sonuç olarak bu iki tahmin edicinin doğrusal kombinasyonu ise, ard arda örnekleme yönteminde $\overline{Y}$ için önerilen aşağıdaki tahmin edici

$$T_s = \phi T_{1s} + (1-\phi)T_{2s} \qquad (4.3)$$

olmaktadır. Burada ϕ, T_s tahmin edicisinin hata kareler ortalamasını minimum yapan bilinmeyen bir sabittir.

Yan ve Hata Kareler Ortalaması:

Singh (2005), kitlenin yeterince büyük olduğu (N→∞) varsayımı altında aşağıdaki eşitlikleri elde etmişlerdir:

$$Yan(T_s) = \phi Yan(T_{1s}) + (1-\phi) Yan(T_{2s}) \tag{4.4}$$

ve

$$HKO(T_s) = \phi^2 HKO(T_{1s}) + (1-\phi)^2 HKO(T_{2s}) \, . \tag{4.5}$$

Burada,

$$Yan(T_{1s}) = \frac{\bar{Y}}{u}\left(C_z^2 - \rho_{yz} C_y C_z\right), \tag{4.6}$$

$$Yan(T_{2s}) = \bar{Y}\left[\left(\frac{1}{m} - \frac{1}{n}\right)\left(C_x^2 - \rho_{yx} C_y C_x\right) + \frac{1}{n}\left(C_z^2 - \rho_{yz} C_y C_z\right)\right], \tag{4.7}$$

$$HKO(T_{1s}) = \frac{\bar{Y}^2}{u}\left(C_y^2 + C_z^2 - 2\rho_{yz} C_y C_z\right) \tag{4.8}$$

ve

$$HKO(T_{2s}) = \bar{Y}^2\left[\frac{C_y^2}{m} + \left(\frac{1}{m} - \frac{1}{n}\right)\left(C_x^2 - 2\rho_{yx} C_y C_x\right) + \frac{1}{n}\left(C_z^2 - 2\rho_{yz} C_y C_z\right)\right] \tag{4.9}$$

göstermektedir (Singh 2005).

Minimum HKO(T_S):

Eşitlik (4.5)'te verilen HKO eşitliği bilinmeyen ϕ sabitine bağlıdır. ϕ'e göre minimum HKO eşitliğine, optimum ϕ değerinin elde edilerek Eşitlik (4.5)'te yerine konulmasıyla ulaşılır. Buradan, Eşitlik (4.5), ϕ'e göre türevi alınıp sıfıra eşitlendiğinde,

$$\phi_{opt} = \frac{HKO(T_{2s})}{HKO(T_{1s}) + HKO(T_{2s})} \tag{4.10}$$

olarak bulunur. Eşitlik (4.10), Eşitlik (4.5)'te yerine koyulduğunda

$$HKO(T_s)_{\min} = \frac{HKO(T_{1s})HKO(T_{2s})}{HKO(T_{1s}) + HKO(T_{2s})}. \tag{4.11}$$

eşitliği elde edilir.

Formülleri basitleştirmek için $A = \bar{Y}^2 C_y^2$, $B = \bar{Y}^2\left(C_x^2 - 2\rho_{yx}C_yC_x\right)$ ve $C = \bar{Y}^2\left(C_z^2 - 2\rho_{yz}C_yC_z\right)$ gösterimleri kullanılmıştır. Buna göre HKO eşitlikleri,

$$HKO(T_{1s}) = \frac{A+C}{u} = \frac{A+C}{n\mu} \tag{4.12}$$

ve

$$HKO(T_{2s}) = \frac{A}{m} + \left(\frac{1}{m} - \frac{1}{n}\right)B + \frac{C}{n} = \frac{A}{n(1-\mu)} + \frac{B\mu}{n(1-\mu)} + \frac{C}{n} \tag{4.13}$$

biçiminde elde edilir. Buradan,

$$HKO(T_s)_{\min.} = \frac{(A+C)^2 + (A+C)(B-C)\mu)}{n\left[(A+C) + (B-C)\mu^2\right]} = \frac{\alpha_1^2 + \alpha_1\alpha_2\mu}{n[\alpha_1 + \alpha_2\mu^2]} \tag{4.14}$$

olur. Burada, $\alpha_1 = A + C$ ve $\alpha_2 = B - C$'i göstermektedir.

<u>Optimum Yenileme İlkesi:</u>

İlk araştırmada seçilen örneklemin ne kadarının ayrılacağına dolayısıyla ikinci araştırmada ne kadarlık bir örneklemin çekileceğine karar verebilmek için, μ'nün tahmin edilmesi gerekir. Eşitlik (4.14)'de verilen *HKO*(T_s)$_{min.}$ ifadesini, μ'e göre minimize ederek μ'nün optimal değeri bulunabilir. Bunun içinde *HKO*(T_s)$_{min.}$ ifadesi μ'e göre türevi alınıp sıfıra eşitlendiğinde, μ'nün optimal değeri aşağıdaki gibi elde edilir:

$$\hat{\mu} = \frac{-\alpha_1 \mp \sqrt{\alpha_1^2 + \alpha_1\alpha_2}}{\alpha_2} = \frac{-\left(1-\rho_{yz}\right) \mp \sqrt{\left(1-\rho_{yz}\right)\left(1-\rho_{yx}\right)}}{\rho_{yz} - \rho_{yx}} = \mu_0 . \tag{4.15}$$

$\rho_{yz} \neq \rho_{yx}$ olduğu durumda $\hat{\mu}$ için iki değer tanımlıdır. Bu değerlerin seçiminde $0 \leq \hat{\mu} \leq 1$ eşitsizliği göz önünde bulundurulmadır. İki değer de bu sınırlar içindeyse, en optimal yani HKO değerini en küçükleyen değer, $\hat{\mu}$ değeri olarak alınabilir. $\hat{\mu}$, Eşitlik (4.14)'de yerine yazıldığında

$$HKO(T_s)_{\min.} = \frac{\alpha_1^2 + \alpha_1\alpha_2\mu_0}{n\left[\alpha_1 + \alpha_2\mu_0^2\right]} \tag{4.16}$$

eşitliği elde edilir.

Singh (2005) çalışmasında, önerdiği T_s tahmin edicisini, önceki araştırma bilgisini kullanmayan, $\bar{y}_n$, ve Sukhatme vd. (1984) tahmin edicisi, $\hat{\bar{Y}} = \bar{y}_m + \beta_{yx}\left(\bar{x}_n - \bar{x}_m\right)$, ile karşılaştırmıştır. Farklı ρ_{yx} , ρ_{yz} ve μ değerleri için, göreli etkinlik değerleri hesaplanmış ve ilişki miktarı arttıkça T_s'nin etkinliğinin arttığı görülmüştür.

4.1.2. Singh ve Priyanka tahmin edicisi

Singh ve Priyanka (2008), çalışmalarında, ard arda örnekleme yönteminde son (ikinci) araştırma kitle ortalaması, $\bar{Y}$, için regresyon katsayısına bağlı iki fark tahmin edicisinin doğrusal kombinasyonuna dayalı yeni bir tahmin edici sunmuşlardır. Daha sonra regresyon parametreleri yerine tahmin edicilerini kullanarak çalışmayı genişletmişlerdir.

Buna göre, $u=n\mu$ örneklem büyüklüğü (ikinci araştırmada yeniden çekilen) için ilk tahmin edici,

$$T_{1u} = \bar{y}_u + B_{yz}\left(\bar{Z} - \bar{z}_u\right) \tag{4.17}$$

ve her iki araştırmada da ortak kullanılan $m(=n\lambda)$ birimlik örneklem büyüklüğüne bağlı ikinci zincirleme fark tahmin edicisi ise,

$$T_{2m} = \bar{y}_m^* + B_{yx}\left(\bar{x}_n^* - \bar{x}_m^*\right) \tag{4.18}$$

şeklindedir. Burada,

$$\bar{y}_m^* = \bar{y}_m + B_{yz}\left(\bar{Z} - \bar{z}_m\right)$$
$$\bar{x}_n^* = \bar{x}_n + B_{xz}\left(\bar{Z} - \bar{z}_n\right)$$
$$\bar{x}_m^* = \bar{x}_m + B_{xz}\left(\bar{Z} - \bar{z}_m\right)$$

biçiminde tanımlanmıştır. B_{yx}, B_{xz}, B_{yz} alt indislerde gösterilen değişkenler arasında bilinen kitle regresyon katsayılarıdır.

T_{1u} ve T_{2m} tahmin edicilerinin doğrusal kombinasyonu ile, $\bar{Y}$ için Singh ve Priyanka tarafından önerilen tahmin ediciye ulaşılır:

$$\hat{T} = \varphi T_{1u} + (1-\varphi) T_{2m} . \tag{4.19}$$

Burada φ bilinmeyen bir sabittir.

Bilinmeyen regresyon parametreleri yerine tahmin edicileri kullanıldığında, $\hat{T}$ tahmin edicisi aşağıdaki biçime dönüşür:

$$\hat{T}^* = \Psi T_{1u}^* + (1-\Psi) T_{2m}^* . \tag{4.20}$$

Burada Ψ bilinmeyen bir sabit,

$$T_{1u}^* = \bar{y}_u + b_{yz}(u)\left(\bar{Z} - \bar{z}_u\right) \tag{4.21}$$

ve $\bar{y}_m^{**} = \bar{y}_m + b_{yz}(m)\left(\bar{Z} - \bar{z}_m\right)$; $\bar{x}_n^{**} = \bar{x}_n + b_{xz}(n)\left(\bar{Z} - \bar{z}_n\right)$ ve $\bar{x}_m^{**} = \bar{x}_m + b_{xz}(m)\left(\bar{Z} - \bar{z}_m\right)$ olmak üzere

$$T_{2m}^* = \bar{y}_m^{**} + b_{yx}(m)\left(\bar{x}_n^{**} - \bar{x}_m^{**}\right) \tag{4.22}$$

şeklindedir. Burada $b_{yz}(u), b_{yx}(m), b_{yz}(m), b_{xz}(n), b_{xz}(m)$, parantez içinde gösterilen örneklem büyüklüklerinde ilgili değişkenler arasındaki örneklem regresyon katsayılarıdır.

Yan ve Hata Kareler Ortalaması:

T_{1u} ve T_{2m}, $\overline{Y}$ için yansız tahmin edicilerdir. $\hat{T}$, T_{1u} ve T_{2m}'in konveks doğrusal kombinasyonu olduğundan, $\overline{Y}$ için yansız bir tahmin edicidir.

$\hat{T}$'nın HKO'su,

$$HKO(\hat{T}) = V(\hat{T}) = \varphi^2 V(T_{1u}) + (1-\varphi)^2 V(T_{2m}) \tag{4.23}$$

biçimindedir. Burada,

$$V(T_{1u}) = \frac{1}{u}\left(1-\rho_{yz}^2\right)S_y^2, \tag{4.24}$$

$$V(T_{2m}) = \left[\frac{1}{m}\left(1-\rho_{yz}^2\right) + \left(\frac{1}{m}-\frac{1}{n}\right)\left(2\rho_{yz}^2\rho_{yx} - \rho_{yx}^2\left(1+\rho_{yz}^2\right)\right)\right]S_y^2 . \tag{4.25}$$

$\hat{T}$ tahmin edicisinin varyans çıkarsamasında, kitlenin yeterince büyük olduğu kabul edilmiş (N→∞), dolayısıyla düzeltme terimi ihmal edilmiştir. Aynı zamanda $\rho_{xz} = \rho_{yz}$ olduğu varsayılmıştır. Eşitlik (4.24) ve (4.25)'te verilen T_{1u} ve T_{2m} tahmin edicilerinin varyans eşitliklerinin çıkarımı EK 8'de gösterilmiştir.

Minimum $V(\hat{T})$:

Bilinmeyen φ sabitine bağlı $\hat{T}$'nın minimum $V(\hat{T})$ eşitliğine, φ_{opt} değeri yerine yazılarak ulaşılır. φ_{opt} değeri, Eşitlik (4.10)'dakine benzer biçimde elde edilir:

$$\varphi_{opt.} = \frac{V(T_{2m})}{V(T_{1u}) + V(T_{2m})} \quad . \tag{4.26}$$

Yerine yazıldığında $V(\hat{T})_{\min.}$ ifadesi,

$$V(\hat{T})_{\min.} = \frac{V(T_{1u})V(T_{2m})}{V(T_{1u}) + V(T_{2m})} \tag{4.27}$$

olur. $V(T_{1u})$ ve $V(T_{2m})$ ifadeleri yerine yazıldığında,

$$V(\hat{T})_{\min.} = \frac{A\left[A + \mu C\right]}{\left[A + \mu^2 C\right]} \frac{S_y^2}{n} \tag{4.28}$$

biçiminde elde edilir. Burada, $A = 1 - \rho_{yz}^2$, $C = 2\rho_{yz}^2 \rho_{yx} - \rho_{yx}^2 \left(1 + \rho_{yz}^2\right)$ ve $\mu = u/n$ ikinci araştırmada yeniden çekilen örneklem oranıdır.

Eğer $\mu = 0$ ise, yani tam eşleştirme yapılıyor ise, *n*=*m* durumu için

$$V(\hat{T})_{\min.} = A\frac{S_y^2}{n} \tag{4.29}$$

olur. Eğer $\mu = 1$ ise, yani hiç eşleştirme yapılmıyor ise, *m*=0 durumu için

$$V(\hat{T})_{\min.} = A\frac{S_y^2}{n} \tag{4.30}$$

olur. Görüldüğü gibi $V(\hat{T})_{\min.}$ aynı değerleri almaktadır. Bu durumda, daha küçük varyanslı tahminler için μ için en iyi seçim yapılmalıdır.

<u>Optimum Yenileme İlkesi:</u>

$V(\hat{T})_{\min.}$ ifadesi μ'e göre türevi alınıp sıfıra eşitlendiğinde,

$$C\mu^2 + 2A\mu - A = 0 \tag{4.31}$$

eşitliğine ulaşılır. Eşitlik (4.31) çözüldüğünde, μ çözüm değerleri aşağıdaki gibi elde edilir:

$$\hat{\mu} = \frac{-A \mp \sqrt{A(A+C)}}{C}. \tag{4.32}$$

$\hat{\mu}$'nün gerçel değerler alabilmesi için $A(A+C) \geq 0$ olması gerektiği açıktır. $A = 1 - \rho_{yz}^2$ olduğundan, $0 \leq A \leq 1$ dir. Dolayısıyla, gerçel değer için, $A + C \geq 0$ olmalıdır. Optimal $\hat{\mu}$'ı, μ_0 şeklinde gösterirsek, bu durumda $V(\hat{T})_{\min.^*}$ ifadesi,

$$V(\hat{T})_{\min.^*} = \frac{A\left[A + \mu_0 C\right]}{\left[A + \mu_0^2 C\right]} \frac{S_y^2}{n} \tag{4.33}$$

biçiminde elde edilir.

Singh ve Priyanka (2008) çalışmasında, $\hat{T}$ tahmin edicisi, eşleştirme yapılmadığı durumda kitle ortalaması tahmin edicisi $\bar{y}_n$ ve farklı ρ_{yz} ve ρ_{yx} değerleri için ikinci araştırmada yardımcı değişken kullanılmadığı, $\hat{\hat{Y}} = \varphi^* \bar{y}_u + (1 - \varphi^*) \bar{y}_m'$ tahmin edicisi ile karşılaştırılmıştır (burada, $\bar{y}_m' = \bar{y}_m + B_{yx}(\bar{x}_n - \bar{x}_m)$ olmaktadır). $\bar{y}_n$ ve $\hat{\hat{Y}}$ tahmin edicileri yansız tahmin edicilerdir, buna göre bu tahmin edicilerin N→∞ için varyansları aşağıda sırasıyla verilmiştir (Singh ve Priyanka, 2008):

$$V(\bar{y}_n) = \frac{S_y^2}{n}, \tag{4.34}$$

$$V(\hat{\hat{Y}})_{\min.} = \left[1 + \sqrt{1 - \rho_{yx}^2}\right] \frac{S_y^2}{2n}. \tag{4.35}$$

$E_1 = \dfrac{V(\bar{y}_n)}{V(\hat{T})^*_{\min.}} * 100$ ve $E_2 = \dfrac{V(\hat{\hat{Y}})}{V(\hat{T})^*_{\min.}} * 100$ göreli etkinlik ifadeleri olmak üzere, farklı ρ_{yz}, ρ_{yx} ve μ_0 değerleri için E_1 ve E_2 değerleri elde edilmiş ve

i) ρ_{yx} sabitken, μ_0 azalırken ve ρ_{yz} arttıkça E_1 ve E_2'nin arttığı,

ii) ρ_{yz} sabitken, μ_0 ve ρ_{yx} arttıkça E_1'in arttığı, E_2'nin ise başlangıç değerine göre azalış gösterdiği görülmüştür.

Sonuç olarak, yardımcı değişken kullanmanın tahmin etkinliğini arttırdığı aynı zamanda yüksek ilişkili yardımcı değişkenlerinde ikinci araştırmada daha az örneklem çekilmesine katkıda bulunduğu gösterilmiştir.

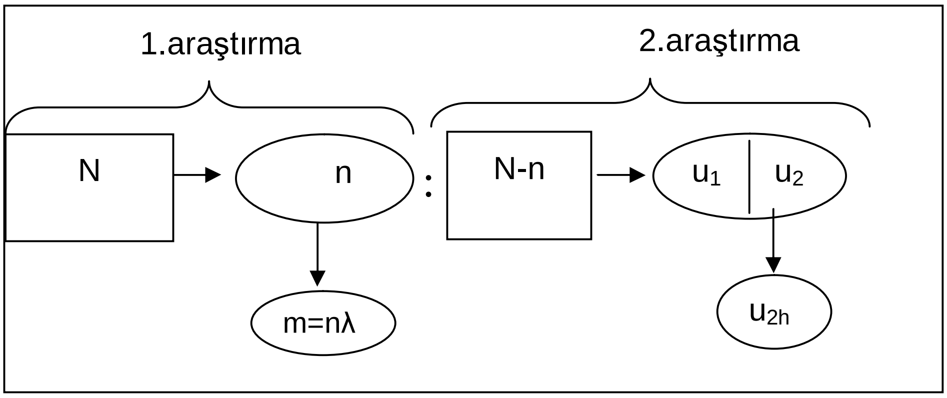
4.2. Kayıp Veri

Bu bölümde, ard arda örnekleme yönteminde, ikinci araştırmada çekilen güncel örneklemde kayıp veri olması durumunda, ikinci araştırma kitle ortalaması tahminine ilişkin teorik yapı incelenecektir.

4.2.1. Singh ve Priyanka tahmin edicisi

Ard arda örnekleme yönteminde kayıp veri olması durumunda kitle ortalaması için yapılmış sınırlı sayıda çalışma vardır. Singh ve Priyanka (2007) çalışmaları, ard arda örnekleme yönteminde, ikinci araştırmada kayıp veri ile karşılaşıldığı durumda, zincirleme örnekleme yönteminde olduğu gibi, tekrar bir alt örneklem çekme prensibine dayanmaktadır.

İlk araştırmada, N birimlik sonlu kitleden, n birimlik basit rastgele örneklemi çekilir. Buradan, $m=n\lambda$ birimlik alt örneklem ikinci araştırmada eşleştirilmek için ayrılır. Cevapsızlığın ikinci araştırmada ortaya çıktığı varsayılmaktadır, dolayısıyla kitle cevaplı N_1 ve cevapsız N_2 olmak üzere ikiye ayrılmış gibi düşünülebilir. Son (ikinci) araştırmada, kalan $N-n$ birimlik kitleden $u=n\mu=n-m$ birimlik basit rastgele örneklemi çekilir. Sonuç olarak, ikinci araştırmada incelenen örneklem büyüklüğü n birim olmaktadır. Burada da, λ ve μ, sırasıyla eşleştirme ve ikinci araştırmada yeni çekilen örneklem oranlarını göstermektedir ($\lambda + \mu = 1$). İkinci araştırmada, cevapsızlık olduğu varsayımı altında eşleştirilmeyen u birimlik örneklemin, u_1 cevaplıların, u_2 cevapsızların olmak üzere iki parçaya ayrıldığı düşünülmektedir. Buna göre, cevapsızların oluşturduğu u_2 alt kümesinden, tekrar u_{2h} alt örneklemi çekilmektedir ve bu alt örneklem birimlerinde cevapsızlık olmadığı varsayılmaktadır.



Şekil 4.2. Kayıp veri olması durumunda ard arda örnekleme yöntemi görüntüsü

Singh ve Priyanka (2007) çalışmasında, *Y* ilgilenilen değişken, X yardımcı değişkendir. $\bar{x}_n, \bar{y}_m, \bar{x}_m, \bar{y}_{u1}, \bar{y}_{u2h}$, alt indislerde gösterilen örneklem büyüklüklerinde değişkenlerin örneklem ortalamalarıdır. ρ_{yx}, *X* ve *Y* arasında ki ilişki katsayısını, S_x^2, S_y^2, gösterilen değişkenlerin kitle varyanslarını ve $W = \frac{N_2}{N}$ kitle cevapsızlık oranını göstermektedir.

Singh ve Priyanka, ikinci araştırmada kitle ortalaması, $\bar{Y}$, tahmini için iki bağımsız tahmin edici sunmuşlardır. Sonuçta, ard arda örnekleme yönteminde $\bar{Y}$ tahmini için, bu tahmin edicilerin doğrusal kombinasyonundan nihai tahmin ediciye ulaşılmıştır. Bu tahmin edicilerden, *u*=*n*μ örneklem büyüklüğüne dayalı birinci tahmin edici aşağıdaki gibidir:

$$T_1 = \bar{y}_u^* = \frac{u_1 \bar{y}_{u1} + u_2 \bar{y}_{u2h}}{u}. \tag{4.36}$$

İkinci tahmin edici ise, her iki araştırmada da ortak kullanılan *m* birimlik örnekleme dayanmaktadır:

$$T_2 = \bar{y}_m + b_{yx}\left(\bar{x}_n - \bar{x}_m\right). \tag{4.37}$$

Burada b_{yx}, bilinen kitle regresyon katsayısına eşit bir sabit olduğu varsayılmaktadır. T_1 ve T_2 tahmin edicilerinin doğrusal kombinasyonu ile, *T* tahmin edicisi aşağıdaki gibi elde edilir:

$$T = \theta T_1 + (1-\theta)T_2 \,. \tag{4.38}$$

Burada θ bilinmeyen bir sabittir.

<u>**Teorem 4.1.**</u> **:** *T* tahmin edicisi $\overline{Y}$ için yansız bir tahmin edicidir.

<u>**Tanıt :**</u> $E(T_1) = E\left(\bar{y}_u^*\right) = E\left[E\left(\bar{y}_u^*\right) \middle| u_1, u_2\right] = E\left[E\left(\dfrac{u_1\bar{y}_{u1} + u_2\bar{y}_{u2h}}{u}\right) \middle| u_1, u_2\right]$

$$= \frac{1}{u}E\left[u_1 E\left(\bar{y}_{u1}\middle|u_1,u_2\right) + u_2 E\left(\bar{y}_{u2h}\middle|u_1,u_2\right)\right] = \frac{1}{u}E\left[u_1\bar{y}_{u1} + u_2\bar{y}_{u2h}\right]$$

$$= E\left[\bar{y}_u\right] = \overline{Y} \,.$$

T_1, $\overline{Y}$ için yansız bir tahmin edicidir. T_2 fark tahmin edicisi de $\overline{Y}$ için yansızdır. Dolayısıyla T_1 ve T_2'nin konveks doğrusal kombinasyonu olan *T*'de $\overline{Y}$ için yansızdır.

T'nin varyansı,

$$V(T) = \theta^2 V(T_1) + (1-\theta)^2 V(T_2) \,. \tag{4.39}$$

Burada $k = \dfrac{u_2}{u_{2h}}$ olmak üzere

$$V(T_1) = \left[\left(\frac{1}{u} - \frac{1}{N}\right) + \frac{(k-1)N_2}{Nu}\right]S_y^2 \;, \tag{4.40}$$

$$V(T_2) = \left[\left(\frac{1}{m} - \frac{1}{n}\right)\left(1-\rho_{yx}^2\right) + \left(\frac{1}{n} - \frac{1}{N}\right)\right]S_y^2 \tag{4.41}$$

biçimindedir (Singh ve Priyanka, 2007).

T tahmin edicisinin varyansında örneklemler birbirinden bağımsız olduğu için kovaryans terimi ihmal edilmiştir. Dolayısıyla,

$$V(T) = E(T - \bar{Y})^2 = \theta^2 V(T_1) + (1-\theta)^2 V(T_2)$$

olur.

$$V(T_1) = V(\bar{y}_u^*) = V\left[E\left(\bar{y}_u^*\right) | \, u_1, u_2\right] + E\left[V\left(\bar{y}_u^*\right) | \, u_1, u_2\right]$$

$$= V\left(\bar{y}_u\right) + E\left[\frac{u_2}{u^2}(k-1)S_{y2}^2\right] = \left(\frac{1}{u} - \frac{1}{N}\right)S_y^2 + \frac{(f-1)}{u} S_{y2}^2 \frac{N_2}{N}$$

S_{y2}^2, son araştırmada cevapsızların kitle varyansıdır ve S_y^2'e eşit olduğu varsayımı altında

$$V(T_1) = \left[\left(\frac{1}{u} - \frac{1}{N}\right) + \frac{(k-1)N_2}{uN}\right]S_y^2$$

biçiminde elde edilir. Benzer olarak,

$$V(T_2) = E\left(T_2 - \bar{Y}\right)^2 = E\left[\bar{y}_m + b_{yx}\left(\bar{x}_n - \bar{x}_m\right) - \bar{Y}\right]^2$$

$$= \left[\left(\frac{1}{m} - \frac{1}{n}\right)\left(1 - \rho_{yx}^2\right) + \left(\frac{1}{n} - \frac{1}{N}\right)\right]S_y^2$$

biçiminde elde edilir.

Bilinmeyen θ sabitine bağlı *V*(*T*)'nin minimum eşitliğine, θ_{opt} değeri yerine yazılarak ulaşılır:

$$\theta_{opt.} = \frac{V(T_2)}{V(T_1) + V(T_2)}, \qquad (4.42)$$

$$V(T)_{\min.} = \frac{V(T_1)V(T_2)}{V(T_1)+V(T_2)}. \tag{4.43}$$

Eşitlik (4.43) başka bir gösterimle,

$$A_1 = \left(\frac{\rho_{yx}^2}{N} - \frac{n}{N^2}\right),\ A_2 = \left(\frac{n}{N^2} - \frac{A}{n}\rho_{yx}^2 + \frac{A-1}{N}\right),\ A_3 = A\left(\frac{1}{n} - \frac{1}{N}\right),\ A_4 = \left(\frac{2n}{N} - \rho_{yx}^2\right),$$

$$A_5 = \left(1 - A - \frac{2n}{N}\right),\ A = 1 + (k-1)W,\ W = \frac{N_2}{N} \text{ ve } \mu = \frac{u}{n},$$

sembolleri kullanılarak

$$V(T)_{\min.} = \frac{A_1\mu^2 + A_2\mu + A_3}{A_4\mu^2 + A_5\mu + A} \tag{4.44}$$

biçiminde ifade edilebilir.

<u>Optimum Yenileme İlkesi:</u>

$\mu_{opt.} = \hat{\mu}$ değerini elde edebilmek için, Eşitlik (4.44)'te verilen $V(T)_{\min.}$ ifadesinin μ'e göre birinci dereceden türevi alınarak sıfıra eşitlendiğinde,

$$\hat{\mu} = \frac{-B_2 \pm \sqrt{B_2^2 - 4B_1B_3}}{2B_1} = \mu_0 \tag{4.45}$$

ifadesine ulaşılır. Burada $B_1 = A_1A_5 - A_2A_4$, $B_2 = 2AA_1 - 2A_3A_4$, $B_3 = AA_2 - A_3A_5$ göstermektedir.

$\hat{\mu}$ değeri, Eşitlik (4.44)'de verilen $V(T)_{\min.}$ ifadesinde yerine yazılarak,

$$V(T)_{\min.^*} = \frac{A_1\mu_0^2 + A_2\mu_0 + A_3}{A_4\mu_0^2 + A_5\mu_0 + A} \tag{4.46}$$

eşitliğine ulaşılır.

Singh ve Priyanka (2007) çalışmalarında, önerdikleri tahmin ediciyi ard arda örneklemede cevapsızlığın etkisini görebilmek için, aynı koşullarda fakat cevapsızlığı dikkate almayan T^*,

$$T^* = \theta^* T_1^* + \left(1-\theta^*\right)T_2 \tag{4.47}$$

tahmin edicisi ile karşılaştırmışlardır. Burada, $T_1^* = \bar{y}_u$ ve T_2 Eşitlik (4.37)'de verildiği gibidir, θ^* ise bilinmeyen bir sabittir. T^*'nın minimum varyansı,

$$V(T^*)_{\min.^*} = \frac{C_1\mu_1^2 + C_2\mu_1 + C_3}{C_4\mu_1^2 + C_5\mu_1 + 1} S_y^2 \tag{4.48}$$

olur. Burada $\mu_1 = \frac{-D_2 \pm \sqrt{D_2^2 - 4D_1D_3}}{2D_1}$ (μ'nün optimal değeri), $C_1 = \frac{\rho_{yx}^2}{N} - \frac{n}{N^2}$, $C_2 = \frac{n}{N^2} - \frac{\rho_{yx}^2}{n}$, $C_3 = \frac{1}{n} - \frac{1}{N}$, $C_4 = \frac{2n}{N} - \rho_{yx}^2$, $C_5 = -\frac{2n}{N}$, $D_1 = C_1C_5 - C_2C_4$, $D_2 = 2C_1 - 2C_3C_4$, $D_3 = C_2 - C_3C_5$ olmaktadır.

Rasgele seçilen, μ_0, μ_1, *W*, *k* ve ρ_{yx} değerleri için, ρ_{yx} arttıkça, Singh ve Priyanka (2007) tahmin edicisinin Eşitlik (4.47)'de verilen T^* tahmin edicisine göre etkinliğinin arttığı, sabit ρ_{yx} ve *W* değerlerinde *k* ve μ arttıkça etkinliğin arttığı, bununla birlikte sabit *k* ve μ değerleri için, ρ_{yx} ve *W* değerleri arttıkça etkinliğin azaldığı gözlenmiştir.

Sonuç olarak, ard arda örnekleme yönteminde son araştırmada karşılaşılan cevapsızlığın, tahmin duyarlılığını azalttığı ve tahminlerde cevapsızlığın dikkate alınması gerektiği gösterilmiştir.

BEŞİNCİ BÖLÜM

5. ÖNERİLEN KİTLE ORTALAMASI TAHMİN EDİCİSİ

Örneklemede kitle karakteristiklerinin tahmini oldukça önemlidir. Tahmin edicilerin etkinliğinin incelenmesinde, maliyet, zaman, emek vb. koşulların yanında en önemli ölçüt HKO'dır. Literatürde, kitle ortalamasına ilişkin çeşitli tahmin ediciler mevcuttur. Daha düşük HKO'na sahip yani daha etkin tahmin edici elde etmeye yönelik çalışmalar devam etmektedir. Cevapsızlık, daha önceki konularda da bahsedildiği gibi, tahminin güvenilirliğini etkileyen bir durumdur. Bu nedenle, çalışmaların, cevapsızlığı kapsayacak şekilde genişletilmesi daha doğru sonuçlara ulaşılması konusunda fayda sağlamaktadır. Dördüncü bölümde bahsedildiği gibi, tekrarlayan araştırma konularında, ard arda örnekleme yöntemini kullanmak hem maliyet, hem zaman açısından fayda sağlarken kitle ortalaması tahmin edicisinin HKO'nın düşmesine de yardımcı olmaktadır. Bu nedenle bu çalışmada, cevapsızlık olması durumunda ard arda örnekleme yönteminde, daha önce yapılmış kitle ortalaması tahmin edicileri incelenerek yeni bir oransal tahmin edici önerilmiştir.

Buna göre, N birimden n birimlik basit rastgele örneklemi çekilir ve $m=n\lambda$ birimlik örneklem parçası ikinci araştırmada kullanılmak üzere ayrılır. İkinci araştırmada ise, kalan N-n birimden $u=n\mu$ birimlik basit rastgele örneklemi çekilir. Yeni çekilen bu örneklemde, ilgilenilen değişkende kayıp veri ile karşılaşıldığı varsayımı altında, u örnekleminin, cevaplı birimlerin oluşturduğu u_1 ve cevapsız birimlerin oluşturduğu u_2 örneklem parçalarına ayrıldığı düşünülmektedir. Cevapsız birimlerin oluşturduğu u_2 örneklem parçasından, u_{2h} alt örneklemi çekilmektedir ve bu örneklem birimlerine tekrar ulaşılmakta ve cevapsızlık olmadığı varsayılmaktadır. Burada Y ilgilenilen değişken, X herhangi ilişkili ve Z kitle ortalaması bilinen yardımcı değişkendir. Semboller, Singh ve Priyanka (2007) tahmin edicisinde olduğu gibi tanımlanmaktadır.

Bu durumda, son (ikinci) araştırmanın kitle ortalaması tahmini için, Singh (2005) tahmin edicisine paralel olarak yeni bir oransal tahmin edici önerilmiştir. Buna göre ilk olarak u örneklemine dayalı birinci tahmin edici aşağıdaki gibidir:

$$T_{öneri1} = \frac{\bar{y}_u^*}{\bar{z}_u^*}\bar{Z} \; . \tag{5.1}$$

Burada, $\bar{y}_u^* = \dfrac{u_1\bar{y}_{u1} + u_2\bar{y}_{u2h}}{u}$, $\bar{z}_u^* = \dfrac{u_1\bar{z}_{u1} + u_2\bar{z}_{u2h}}{u}$ göstermektedir.

İlk araştırmada ayrılan *m* birimlik örnekleme dayalı ikinci tahmin edici ise, Singh ve Priyanka (2007) tahmin edicisinde Eşitlik (4.37) ile tanımlanan, T_2 tahmin edicisidir. $T_{öneri1}$ ve T_2 tahmin edicilerinin doğrusal kombinasyonu ile $\bar{Y}$ tahminine ulaşılır,

$$T_{öneri} = \gamma T_{öneri1} + (1-\gamma)T_2\,. \tag{5.2}$$

Teorem 5.1. : $T_{öneri}$ tahmin edicisinin HKO eşitliği,

$$HKO(T_{öneri}) = \gamma^2 HKO(T_{öneri1}) + (1-\gamma)^2 V(T_2) \tag{5.3}$$

biçiminde olmaktadır. Burada, $R_x = \dfrac{\bar{Y}}{\bar{X}}$, $R_z = \dfrac{\bar{Y}}{\bar{Z}}$ olmak üzere,

$$HKO(T_{öneri1}) = \left[\left(\frac{1}{u} - \frac{1}{N}\right) + \frac{W(k-1)}{u}\right]\left[S_y^2 - 2R_z\rho_{yz}S_yS_z + R_z^2S_z^2\right] \tag{5.4}$$

ve

$$V(T_2) = \left[\left(\frac{1}{m} - \frac{1}{n}\right)\left(1-\rho_{yx}^2\right) + \left(\frac{1}{n} - \frac{1}{N}\right)\right]S_y^2$$

eşitliklerine sahiptir.

Tanıt : $T_{öneri1}$ ve T_2 tahmin edicileri bağımsız örneklemlere dayanmaktadır, dolayısıyla $HKO(T_{öneri})$ ifadesinde kovaryans terimi sıfırdır. Buna göre,

$$HKO(T_{öneri}) = \gamma^2 HKO(T_{öneri1}) + (1-\gamma)^2 HKO(T_2)$$

olur. $HKO(T_{öneri1})$, birinci dereceden Taylor yaklaşımı kullanılarak aşağıdaki gibi elde edilir:

$$h(Y_1, Y_2) = h\left(\bar{y}_u^*, \bar{z}_u^*\right) = T_{öneri1}$$

$$\frac{\partial T_{öneri1}}{\partial \bar{y}_u^*}\bigg|_{\bar{Y},\bar{Z}} = 1 \quad , \quad \frac{\partial T_{öneri1}}{\partial \bar{z}_u^*}\bigg|_{\bar{Y},\bar{Z}} = -\frac{\bar{Y}}{\bar{Z}} = -R_z$$

$$T_{öneri1} - \bar{Y} = (\bar{y}_u^* - \bar{Y}) - \frac{\bar{Y}}{\bar{Z}}\left(\bar{z}_u^* - \bar{Z}\right)$$

$$\left(T_{öneri1} - \bar{Y}\right)^2 = (\bar{y}_u^* - \bar{Y})^2 + \frac{\bar{Y}^2}{\bar{Z}^2}\left(\bar{z}_u^* - \bar{Z}\right)^2 - 2\frac{\bar{Y}}{\bar{Z}}(\bar{y}_u^* - \bar{Y})\left(\bar{z}_u^* - \bar{Z}\right)$$

(her iki tarafın beklenen değerini alırsak)

$$HKO(T_{öneri1}) = V(\bar{y}_u^*) + \frac{\bar{Y}^2}{\bar{Z}^2}V\left(\bar{z}_u^*\right) - 2\frac{\bar{Y}}{\bar{Z}}Cov(\bar{y}_u^*, \bar{z}_u^*)$$

$$\bar{y}_u^* = \frac{u_1}{u}\bar{y}_{u1} + \frac{u_2}{u}\bar{y}_{u2h} - \frac{u_2}{u}\bar{y}_{u2} + \frac{u_2}{u}\bar{y}_{u2}$$

biçiminde yazılabilir. Buradan,

$$\bar{y}_u^* = \frac{u_1}{u}\bar{y}_{u1} + \frac{u_2}{u}\bar{y}_{u2} + \frac{u_2}{u}\left(\bar{y}_{u2h} - \bar{y}_{u2}\right)$$

$$= \bar{y}_u + \frac{u_2}{u}\left(\bar{y}_{u2h} - \bar{y}_{u2}\right)$$

olur. Buna göre,

$$V\left(\bar{y}_u^*\right) = V\left(\bar{y}_u\right) + \frac{u_2^2}{u^2}V\left(\bar{y}_{u2h} - \bar{y}_{u2}\right) + 2\frac{u_2}{u}Cov\left[\bar{y}_u, \left(\bar{y}_{u2h} - \bar{y}_{u2}\right)\right]$$

olur. $Cov\left[\bar{y}_u, \left(\bar{y}_{u2h} - \bar{y}_{u2}\right)\right] = 0$ olduğuna göre (Cochran, 1977),

$$V\left(\bar{y}_u^*\right) = V\left(\bar{y}_u\right) + \frac{u_2^2}{u^2}V\left(\bar{y}_{u2h} - \bar{y}_{u2}\right) \tag{5.5}$$

biçiminde elde edilir.

$$V(\bar{y}_{u2h} - \bar{y}_{u2}) = V(\bar{y}_{u2h}) + V(\bar{y}_{u2}) - 2Cov(\bar{y}_{u2h}, \bar{y}_{u2})$$

$Cov(\bar{y}_{u2h}, \bar{y}_{u2}) = V(\bar{y}_{u2})$ olduğuna göre (Cochran, 1977),

$$V(\bar{y}_{u2h} - \bar{y}_{u2}) = V(\bar{y}_{u2h}) - V(\bar{y}_{u2})$$

olur.

Buna göre,

$$V(\bar{y}_{u2h} - \bar{y}_{u2}) = \left(\frac{1}{u_{2h}} - \frac{1}{N_2}\right)S_{y2}^2 - \left(\frac{1}{u_2} - \frac{1}{N_2}\right)S_{y2}^2$$

$$= \left(\frac{1}{u_{2h}} - \frac{1}{u_2}\right)S_{y2}^2$$

olur. Bu ifade, Eşitlik (5.5)'de yerine yazıldığında,

$$V(\bar{y}_u^*) = \left(\frac{1}{u} - \frac{1}{N}\right)S_y^2 + \frac{u_2^2}{u^2}\left(\frac{1}{u_{2h}} - \frac{1}{u_2}\right)S_{y2}^2$$

$$= \left(\frac{1}{u} - \frac{1}{N}\right)S_y^2 + \frac{u_2}{u^2}\left(\frac{u_2 - u_{2h}}{u_{2h}}\right)S_{y2}^2$$

$w = \dfrac{u_2}{u}$, $k = \dfrac{u_2}{u_{2h}}$ olduğuna göre,

$$V(\bar{y}_u^*) = \left(\frac{1}{u} - \frac{1}{N}\right)S_y^2 + \frac{w}{u}(k-1)S_{y2}^2$$

biçiminde elde edilir. Örneklem cevapsızlık oranı, *w*, sabit olmadığında yerine beklenen değeri, E(*w*)=*W* kullanılarak,

$$V\left(\bar{y}_u^*\right)=\left(\frac{1}{u}-\frac{1}{N}\right)S_y^2+\frac{W}{u}(k-1)S_{y2}^2$$

olarak bulunur. Aynı şekilde,

$$V(\bar{z}_u^*)=\left(\frac{1}{u}-\frac{1}{N}\right)S_z^2+\frac{W(k-1)}{u}S_{z2}^2$$

ve

$$Cov(\bar{y}_u^*,\bar{z}_u^*)=Cov\left[\left(\bar{y}_u+\frac{u_2}{u}(\bar{y}_{u2h}-\bar{y}_{u2})\right),\left(\bar{z}_u+\frac{u_2}{u}(\bar{z}_{u2h}-\bar{z}_{u2})\right)\right]$$

$$=Cov(\bar{y}_u,\bar{z}_u)+\frac{u_2}{u}Cov(\bar{y}_u,(\bar{z}_{u2h}-\bar{z}_{u2}))+\frac{u_2}{u}Cov(\bar{z}_u,(\bar{y}_{u2h}-\bar{y}_{u2}))$$
$$+\frac{u_2^2}{u^2}Cov((\bar{y}_{u2h}-\bar{y}_{u2}),(\bar{z}_{u2h}-\bar{z}_{u2}))$$

$Cov(\bar{y}_u,(\bar{z}_{u2h}-\bar{z}_{u2}))=0$ ve $Cov(\bar{z}_u,(\bar{y}_{u2h}-\bar{y}_{u2}))=0$ olduğuna göre (Cochran, 1977),

$$Cov(\bar{y}_u^*,\bar{z}_u^*)=Cov(\bar{y}_u,\bar{z}_u)+\frac{u_2^2}{u^2}Cov((\bar{y}_{u2h}-\bar{y}_{u2}),(\bar{z}_{u2h}-\bar{z}_{u2}))$$

$$=\left(\frac{1}{u}-\frac{1}{N}\right)S_{yz}^2+\frac{u_2^2}{u^2}\left(\frac{1}{u_{2h}}-\frac{1}{u_2}\right)S_{yz(2)}^2$$

$$=\left(\frac{1}{u}-\frac{1}{N}\right)S_{yz}+\frac{w(k-1)}{u}S_{yz(2)}$$

$E(w)=W$ olduğuna göre,

$$Cov(\bar{y}_u^*,\bar{z}_u^*)=\left(\frac{1}{u}-\frac{1}{N}\right)S_{yz}+\frac{W(k-1)}{u}S_{yz(2)}$$

olur. Singh ve Priyanka (2007) çalışmasında olduğu gibi, $S_{y2}^2 = S_y^2$, $S_{z2}^2 = S_z^2$ ve $S_{yz(2)} = S_{yz}$ olarak kabul edildiğinde,

$$HKO(T_{öneri1}) = \left[\left(\frac{1}{u} - \frac{1}{N}\right) + \frac{W(k-1)}{u}\right]\left[S_y^2 + \frac{\overline{Y}^2}{\overline{Z}^2} S_z^2 - 2\frac{\overline{Y}}{\overline{Z}} S_{yz}\right]$$

($\rho_{yz} = \dfrac{S_{yz}}{S_y S_z}$ olduğuna göre)

$$HKO(T_{öneri1}) = \left[\left(\frac{1}{u} - \frac{1}{N}\right) + \frac{W(k-1)}{u}\right]\left[S_y^2 - 2R_z \rho_{yz} S_y S_z + R_z^2 S_z^2\right]$$

olarak bulunur.

<u>Teorem 5.2.</u> $T_{öneri}$ tahmin edicisinin yan miktarı (:YM) eşitliği,

$$YM(T_{öneri}) \cong \gamma \frac{1}{\overline{Z}}\left(\frac{W(k-1)+1}{u} - \frac{1}{N}\right)\left(R_z S_z^2 - S_{yz}\right)$$

biçiminde olmaktadır.

Tanıt : T_2 regresyon tahminidir dolayısıyla $\overline{Y}$ için yansızdır (Singh ve Priyanka, 2007). Dolayısıyla,

$$YM(T_{öneri}) \cong \gamma \, YM(T_{öneri1})$$

olmaktadır. Burada $YM(T_{öneri1})$ ise, birinci dereceden Taylor yaklaşımı kullanılarak aşağıdaki gibi elde edilir:

$$YM(T_{öneri1}) \cong \frac{1}{2} E\left[-\frac{2}{\overline{Z}}\left(\bar{z}_u^* - \overline{Z}\right)\left(\bar{y}_u^* - \overline{Z}\right) + \frac{\overline{Y}}{\overline{Z}^2}\left(\bar{z}_u^* - \overline{Z}\right)\left(\bar{z}_u^* - \overline{Z}\right)\right]$$

$$= \frac{1}{2}\left[-\frac{2}{\overline{Z}} Cov\left(\bar{z}_u^*, \bar{y}_u^*\right) + \frac{\overline{Y}}{\overline{Z}^2} V\left(\bar{z}_u^*\right)\right]$$

Burada, $V(\bar{z}_u^*) = \left(\frac{1}{u} - \frac{1}{N}\right) S_z^2 + \frac{W(k-1)}{u} S_{z2}^2$ ve $Cov(\bar{y}_u^*, \bar{z}_u^*) = \left(\frac{1}{u} - \frac{1}{N}\right) S_{yz} + \frac{W(k-1)}{u} S_{yz(2)}$ olduğuna göre ve Singh ve Priyanka (2007) çalışmasında olduğu gibi, $S_{z2}^2 = S_z^2$ ve $S_{yz(2)} = S_{yz}$ olarak kabul edildiğinde,

$$YM(T_{öneri1}) \cong \frac{1}{\bar{Z}} \left(\frac{W(k-1)+1}{u} - \frac{1}{N} \right) \left(R_z S_z^2 - S_{yz} \right)$$

olarak bulunur.

$HKO(T_{öneri1})$ ve $V(T_2)$ ifadeleri Eşitlik (5.3)'de yerine yazıldığında $HKO(T_{öneri})$ eşitliğine ulaşılır. $HKO(T_{öneri})$ ifadesinin γ'e göre minimum değeri için, $\gamma_{opt.}$ değerinin yerine yazılması gerekmektedir. Buna göre, $\frac{\partial HKO(T_{öneri})}{\partial \gamma} = 0$ eşitliğinden elde edilen,

$$\gamma_{opt.} = \frac{V(T_2)}{HKO(T_{öneri1}) + V(T_2)} \tag{5.6}$$

ifadesi Eşitlik (5.3)'te yerine yazıldığında,

$$HKO(T_{öneri})_{\min.} = \frac{HKO(T_{öneri1}) V(T_2)}{HKO(T_{öneri1}) + V(T_2)} \tag{5.7}$$

eşitliği elde edilir. Gösterimi basitleştirmek için,

$A_1 = S_y^2 - 2R_z \rho_{yz} S_y S_z + R_z^2 S_z^2$,

$A_2 = 1 + W(k-1)$,

$A_3 = \frac{1}{n} - \frac{1}{N}$

olmak üzere,

$K_1 = n^2 A_1 A_3 - nA_1(1-\rho_{yx}^2)$,

$K_2 = NA_1 A_2 (1-\rho_{yx}^2) - nA_1 A_3 (NA_2 + n)$,

$K_3 = NnA_1 A_2 A_3$,

$K_4 = n^2 A_1 + Nn(1-\rho_{yx}^2) S_y^2 - Nn^2 A_3 S_y^2$,

$K_5 = Nn^2 A_3 S_y^2 - NnA_1 A_2 - n^2 A_1$,

$K_6 = NnA_1 A_2$

ifadeleri kullanılarak, $HKO(T_{öneri})_{\min.}$ aşağıdaki gibi elde edilir:

$$HKO(T_{öneri})_{\min.} = \frac{K_1\mu^2 + K_2\mu + K_3}{K_4\mu^2 + K_5\mu + K_6} S_y^2 . \tag{5.8}$$

<u>**Teorem 5.3. :**</u> Optimum yenileme ilkesi kapsamında, $\mu_{opt.} = \hat{\mu}$ değeri ve γ ve μ'e göre minimum $HKO(T_{öneri})$ ifadesi aşağıdaki gibi elde edilmektedir:

$$\hat{\mu} = \frac{-\Delta_2 \pm \sqrt{\Delta_2^2 - 4\Delta_1\Delta_3}}{2\Delta_1} = \mu_0 , \tag{5.9}$$

$$HKO(T_{öneri})_{\min.^*} = \frac{K_1\mu_0^2 + K_2\mu_0 + K_3}{K_4\mu_0^2 + K_5\mu_0 + K_6} S_y^2 . \tag{5.10}$$

Burada $\Delta_1 = K_1K_5 - K_2K_4$, $\Delta_2 = 2K_1K_6 - 2K_3K_4$ ve $\Delta_3 = K_2K_6 - K_3K_5$ göstermektedir.

Tanıt : $\mu_{opt.} = \hat{\mu}$ değerine, $HKO(T_{öneri})_{\min.}$ ifadesinin μ'e göre birinci dereceden türevi alınıp sıfıra eşitlenerek ulaşılabilir. $\hat{\mu} = \mu_0$ değeri, Eşitlik (5.8)'de verilen $HKO(T_{öneri})_{\min.}$ ifadesinde yerine yazılarak, $HKO(T_{öneri})_{\min.^*}$ Eşitlik (5.10)'daki gibi elde edilir.

Teorem 5.4. : Eşitlik (5.2)'de önerilen tahmin edicinin aşağıda verilen koşulda, Eşitlik (4.38)'de verilen Singh ve Priyanka (2007) tahmin edicisinden daha etkindir:

$$\frac{1}{2}\frac{C_z}{C_y} < \rho_{yz} \,. \qquad (5.11)$$

Tanıt : $HKO(T_{öneri})_{\min.} < V(T)_{\min.}$ eşitsizliğinde ifadeler yerine yazıldığında,

$$\frac{HKO(T_{öneri1})V(T_2)}{HKO(T_{öneri1})+V(T_2)} < \frac{V(T_1)V(T_2)}{V(T_1)+V(T_2)}\,,$$

$$HKO(T_{öneri1})[V(T_1)+V(T_2)] < V(T_1)[HKO(T_{öneri1})+V(T_2)]\,,$$

$$HKO(T_{öneri1})V(T_2) < V(T_1)V(T_2)\,,$$

$$HKO(T_{öneri1}) < V(T_1)$$

elde edilir. Eşitlik (5.4) ve Eşitlik (4.40) yerine yazılarak,

$$\left[\left(\frac{1}{u}-\frac{1}{N}\right)+\frac{W(k-1)}{u}\right]\left[S_y^2-2R_z\rho_{yz}S_yS_z+R_z^2S_z^2\right] < \left[\left(\frac{1}{u}-\frac{1}{N}\right)+\frac{(k-1)W}{u}\right]S_y^2\,,$$

$$S_y^2-2R_z\rho_{yz}S_yS_z+R_z^2S_z^2 < S_y^2\,,$$

$$R_z^2S_z^2 < 2R_z\rho_{yz}S_yS_z\,,$$

$R_zS_z < 2\rho_{yz}S_z$, (Ortalamaların pozitif olduğu varsayımı altında)

$$\frac{1}{2}R_z\frac{S_z}{S_y} < \rho_{yz}$$

olur ve buradan,

$$\frac{1}{2}\frac{C_z}{C_y} < \rho_{yz}$$

koşulu elde edilir.

ALTINCI BÖLÜM

6. SAYISAL ÖRNEK

Sayısal örnekte, Çıngı vd. (2007) TÜBİTAK projesinde kullanılan Milli Eğitim Bakanlığı verileri ele alınmıştır. Türkiye'de ilk ve ortaöğretim olanakları, coğrafik bölge, il ve ilçe bazında farklılıklar göstermektedir. Proje kapsamında bu farklılıklar Türkiye'deki 923 ilçede ölçülen çeşitli değişkenlere göre, 2005-2006 ve 2006-2007 öğretim yılında incelenmiş ve ilçeler gelişmişlik düzeylerine göre sınıflandırılmıştır. Bu çalışmada, orta düzeyde gelişmiş olan 261 ilçe için, ilköğretimdeki öğretmen sayısı (*X*) ve ilköğretime takviye dershane sayısı (*Z*) yardımcı değişken olarak alınarak Ortaöğretim Kurumları Öğrenci Seçme Sınavını (OKS) kazanan öğrenci sayıları (*Y*) tahmin edilecektir.

Uygulamada amaç, kuramsal olarak anlatılan Singh ve Priyanka (2007) tahmin edicisine ve önerilen tahmin ediciye göre kitle ortalaması tahminini yapmak ve bu tahmin edicilerin HKO'larını hesaplayarak etkinliklerini karşılaştırmaktır. Bahsedilen değişkenlerin 2005-2006 ve 2006-2007 öğretim yıllarına ait kitle değerleri Çizelge (6.1)'de sırasıyla verilmiştir.

Çizelge (6.1) 2005-2006 ve 2006-2007 Öğretim Yılları İçin İlköğretimdeki Öğretmen Sayısı (*X*), İlköğretime Takviye Dershane Sayısı (*Z*), OKS'yi Kazanan Öğrenci Sayısı (*Y*) Değişkenlerine Ait Kitle Bilgileri

N=261		
$\bar{Y}_{2005-2006} = 222{,}58$	$\bar{X}_{2005-2006} = 306{,}45$	$\bar{Z}_{2005-2006} = 2{,}07$
$S_y^2 = 172386{,}4$	$S_x^2 = 290706{,}7$	$S_z^2 = 17{,}5$
$C_y = 1{,}87$	$C_x = 1{,}76$	$C_z = 2{,}02$
$\rho_{yx} = 0{,}970$	$\rho_{yz} = 0{,}94$	$\rho_{xz} = 0{,}93$
$\bar{Y}_{2006-2007} = 179{,}77$	$\bar{X}_{2006-2007} = 312{,}33$	$\bar{Z}_{2006-2007} = 2{,}05$

Kitle birimi, Türkiye'de orta gelişmişlik düzeyindeki ilçelerin her biridir. Bu durumda, 261 ilçeden yerine konmadan basit rasgele örnekleme ile bir örneklem çekilmiştir. Örneklem büyüklüğünün tahmininde,

$$n_0 = \frac{t^2 S_y^2}{d^2} \Rightarrow n = \frac{n_0}{1 + \frac{n_0}{N}}, \qquad (6.1)$$

eşitliği kullanılmıştır (Çıngı,1994). Burada *d*, $\bar{Y}$ tahmini için hoşgörü miktarıdır. Hoşgörü miktarı 70 olarak alındığında, $n_0 \cong 140$ ve buradan $n \cong 90$ olarak elde edilmiştir. Buna göre, 261 ilçeden, 90 ilçe basit rasgele örnekleme ile seçilmiştir. Seçilen örnekleme ait değerler Çizelge (6.2)'de verilmiştir.

Çizelge (6.2) 2005-2006 Öğretim Yılı İçin İlköğretimdeki Öğretmen Sayısı (*X*), İlköğretime Takviye Dershane Sayısı (*Z*), OKS'yi Kazanan Öğrenci Sayısı (*Y*) Değişkenlerine Ait Örneklem Bilgileri

n=90		
$\bar{y}_n = 236{,}37$	$\bar{x}_n = 333{,}42$	$\bar{z}_n = 2{,}16$
$s^2_{y(n)} = 192906{,}3$	$s^2_{x(n)} = 313864{,}3$	$s^2_{(z)} = 15{,}44$

Singh ve Priyanka (2007) tahmin edicisi ve önerilen tahmin edici için HKO, farklı kitle cevapsızlık oranı (*W*), ikinci araştırmada cevapsızlardan çekilen alt örneklem oranı (*k*) ve yeni örneklem oranı (μ) değerleri için hesaplanmıştır. Ayrıca, bu hesaplamalarda HKO'ların, ilişki miktarına göre değişimini de inceleyebilmek için, kitleye ait $\rho_{yz} = 0{,}94$ ve $\rho_{yx} = 0{,}97$ değerleri yanında, daha düşük ve daha yüksek ilişki miktarlarında HKO hesaplanarak Çizelge (6.3) ve Çizelge (6.4)'te verilmiştir.

Çizelge (6.3) $\rho_{yx} = 0{,}97$ için farklı *W*, *k*, μ ve ρ_{yz} değerlerinde Singh ve Priyanka (2007) tahmin edicisi varyans (V(*T*)), önerilen tahmin edicinin HKO (HKO(*T*)$_{öneri}$) ve etkinlik değerleri

k		1,500									2,000					
μ		0,100			0,300			0,500			0,100			0,300		
W	ρ_{yz}	HKO(T)$_{öneri}$	V(T)	Etkinlik	HKO(T)$_{öneri}$	V(T)	Etkinlik	HKO(T)$_{öneri}$	V(T)	Etkinlik	HKO(T)$_{öneri}$	V(T)	Etkinlik	HKO(T)$_{öneri}$	V(T)	Etkinlik
0,2	0,935	892,23	1193,383	1,338	545,970	1081,817	1,981	379,240	987,801	2,605[*]	915,480	1199,400	1,310	576,520	1098,855	1,906
	0,850	1067,690	1193,383	1,118	805,94	1081,817	1,342	633,36	987,801	1,560	1082,330	1199,400	1,108	835,000	1098,855	1,316
	0,985	491,230	1193,383	2,429	209,810	1081,817	5,156	126,720	987,801	7,795[**]	518,480	1199,400	2,313	227,190	1098,855	4,837
0,4	0,935	915,480	1199,400	1,310	576,520	1098,860	1,906	407,940	1015,270	2,489	954,300	1208,913	1,267	630,790	1126,120	1,785
	0,850	1082,33	1199,400	1,108	835,000	1098,860	1,316	668,270	1015,270	1,519	1105,990	1208,913	1,093	884,010	1126,120	1,274
	0,985	518,48	1199,400	2,313	227,190	1098,860	4,837	138,990	1015,270	7,305	567,610	1208,913	2,130	260,350	1126,120	4,325
0,6	0,935	936,020	1204,514	1,287	604,710	1113,461	1,841	435,010	1039,032	2,389	985,410	1216,094	1,234	677,510	1146,970	1,693
	0,850	1094,970	1204,514	1,100	860,860	1113,461	1,293	700,030	1039,032	1,484	1124,300	1216,094	1,082	923,730	1146,970	1,242
	0,985	543,870	1204,514	2,215	244,030	1113,461	4,563	151,020	1039,032	6,880	610,690	1216,094	1,991	291,530	1146,970	3,934
0,8	0,935	954,300	1208,910	1,267	630,790	1126,120	1,785	460,600	1059,800	2,301	1010,890	1221,710	1,209	718,170	1163,440	1,620
	0,850	1105,990	1208,910	1,093	884,010	1126,120	1,274	729,030	1059,800	1,454	1138,870	1221,710	1,073	956,580	1163,440	1,216
	0,985	567,610	1208,910	2,130	260,350	1126,120	4,325	162,820	1059,800	6,509	648,770	1221,710	1,883	320,900	1163,440	3,626

Çizelge (6.3) devam ediyor

k		2,000			2,500								
μ		0,500			0,100			0,300			0,500		
W	ρ_{yz}	HKO(T)$_{öneri}$	V(T)	Etkinlik	HKO(T)$_{öneri}$	V(T)	Etkinlik	HKO(T)$_{öneri}$	V(T)	Etkinlik	HKO(T)$_{öneri}$	V(T)	Etkinlik
0,2	0,935	407,940	1015,266	2,489	936,020	1204,514	1,287	604,710	1113,461	1,841	435,010	1039,032	2,389
	0,850	668,27	1015,266	1,519	1094,970	1204,514	1,100	860,86	1113,461	1,293	700,03	1039,032	1,484
	0,985	138,990	1015,266	7,305	543,870	1204,514	2,215	244,030	1113,461	4,563	151,020	1039,032	6,880
0,4	0,935	460,600	1059,798	2,301	985,410	1216,094	1,234	677,510	1146,972	1,693	507,790	1094,350	2,155
	0,850	729,030	1059,798	1,454	1124,300	1216,094	1,082	923,730	1146,972	1,242	780,080	1094,350	1,403
	0,985	162,820	1059,798	6,509	610,690	1216,094	1,991	291,530	1146,972	3,934	185,750	1094,350	5,892
0,6	0,935	507,790	1094,350	2,155	1021,980	1224,080	1,198	736,580	1170,430	1,589	570,040	1133,750	1,989
	0,850	780,080	1094,350	1,403	1145,100	1224,080	1,069	970,960	1170,430	1,205	843,000	1133,750	1,345
	0,985	185,750	1094,350	5,892	666,200	1224,080	1,837	334,950	1170,430	3,494	218,550	1133,750	5,188
0,8	0,935	550,320	1121,940	2,039	1050,160	1229,910	1,171	785,460	1187,780	1,512	623,890	1163,230	1,864
	0,850	823,580	1121,940	1,362	1160,630	1229,910	1,060***	1007,750	1187,780	1,179	893,760	1163,230	1,302
	0,985	207,820	1121,940	5,399	713,060	1229,910	1,725	374,800	1187,780	3,169	249,580	1163,230	4,661

Çizelge (6.4) $\rho_{yx} = 0{,}8$ için farklı *W*, *k*, μ ve ρ_{yz} değerlerinde Singh ve Priyanka (2007) tahmin edicisi varyans (V(*T*)), önerilen tahmin edicinin HKO ($\text{HKO}(T)_{\text{öneri}}$) ve etkinlik değerleri

k		1,500									2,000					
μ		0,100			0,300			0,500			0,100			0,300		
W	ρ_{yz}	$\text{HKO(T)}_{\text{öneri}}$	V(T)	Etkinlik	$\text{HKO(T)}_{\text{öneri}}$	V(T)	Etkinlik	$\text{HKO(T)}_{\text{öneri}}$	V(T)	Etkinlik	$\text{HKO(T)}_{\text{öneri}}$	V(T)	Etkinlik	$\text{HKO(T)}_{\text{öneri}}$	V(T)	Etkinlik
0,2	0,935	923,49	1249,980	1,354	585,010	1246,660	2,131	413,190	1256,750	3,042*	948,430	1256,590	1,325	620,230	1269,340	2,047
	0,850	1112,770	1249,980	1,123	894	1246,660	1,394	734,06	1256,750	1,712	1128,680	1256,590	1,113	929,910	1269,340	1,365
	0,985	500,560	1249,980	2,497	215,330	1246,660	5,790	130,290	1256,750	9,646**	528,880	1256,590	2,376	233,680	1269,340	5,432
0,4	0,935	948,430	1256,590	1,325	620,230	1269,340	2,047	447,490	1301,550	2,909	990,150	1267,030	1,280	683,480	1305,860	1,911
	0,850	1128,68	1256,590	1,113	929,910	1269,340	1,365	781,400	1301,550	1,666	1154,440	1267,030	1,098	991,090	1305,860	1,318
	0,985	528,88	1256,590	2,376	233,680	1269,340	5,432	143,310	1301,550	9,082	580,100	1267,030	2,184	268,910	1305,860	4,856
0,6	0,935	970,490	1262,200	1,301	652,970	1288,870	1,974	480,280	1340,870	2,792	1023,680	1274,920	1,245	738,680	1333,980	1,806
	0,850	1142,430	1262,200	1,105	962,090	1288,870	1,340	825,170	1340,870	1,625	1174,400	1274,920	1,086	1041,300	1333,980	1,281
	0,985	555,330	1262,200	2,273	251,530	1288,870	5,124	156,130	1340,870	8,588	625,170	1274,920	2,039	302,300	1333,980	4,413
0,8	0,935	990,150	1267,030	1,280	683,480	1305,860	1,911	511,660	1375,660	2,689	1051,220	1281,090	1,219	787,270	1356,300	1,723
	0,850	1154,440	1267,030	1,098	991,090	1305,860	1,318	865,770	1375,660	1,589	1190,310	1281,090	1,076	1083,230	1356,300	1,252
	0,985	580,100	1267,030	2,184	268,910	1305,860	4,856	168,780	1375,660	8,151	665,140	1281,090	1,926	334,000	1356,300	4,061

Çizelge (6.4) devam ediyor

k		2,000			2,500								
μ		0,500			0,100			0,300			0,500		
W	ρ_{yz}	HKO(T)$_{öneri}$	V(T)	Etkinlik	HKO(T)$_{öneri}$	V(T)	Etkinlik	HKO(T)$_{öneri}$	V(T)	Etkinlik	HKO(T)$_{öneri}$	V(T)	Etkinlik
0,2	0,935	447,490	1301,550	2,909	970,490	1262,200	1,301	652,970	1288,870	1,974	480,280	1340,870	2,792
	0,850	781,4	1301,550	1,666	1142,430	1262,200	1,105	962,090	1288,870	1,340	825,17	1340,870	1,625
	0,985	143,310	1301,550	9,082	555,330	1262,200	2,273	251,530	1288,870	5,124	156,130	1340,870	8,588
0,4	0,935	511,660	1375,660	2,689	1023,680	1274,920	1,245	738,680	1333,980	1,806	570,560	1434,440	2,514
	0,850	865,770	1375,660	1,589	1174,400	1274,920	1,086	1041,300	1333,980	1,281	938,730	1434,440	1,528
	0,985	168,780	1375,660	8,151	625,170	1274,920	2,039	302,300	1333,980	4,413	193,540	1434,440	7,412
0,6	0,935	570,560	1434,440	2,514	1063,210	1283,700	1,207	809,450	1365,820	1,687	650,360	1502,900	2,311
	0,850	938,730	1434,440	1,528	1197,110	1283,700	1,072	1101,710	1365,820	1,240	1031,360	1502,900	1,457
	0,985	193,540	1434,440	7,412	683,480	1283,700	1,878	349,250	1365,820	3,911	229,410	1502,900	6,551
0,8	0,935	624,810	1482,220	2,372	1093,750	1290,120	1,180	868,870	1389,500	1,599	721,400	1555,150	2,156
	0,850	1002,440	1482,220	1,479	1214,090	1290,120	1,063***	1149,310	1389,500	1,209	1108,370	1555,150	1,403
	0,985	217,620	1482,220	6,811	732,890	1290,120	1,760	392,800	1389,500	3,537	263,840	1555,150	5,894

* : Bilinen kitle ilişki miktarına göre (ρ_{yz} =0,935) en büyük etkinlik miktarı

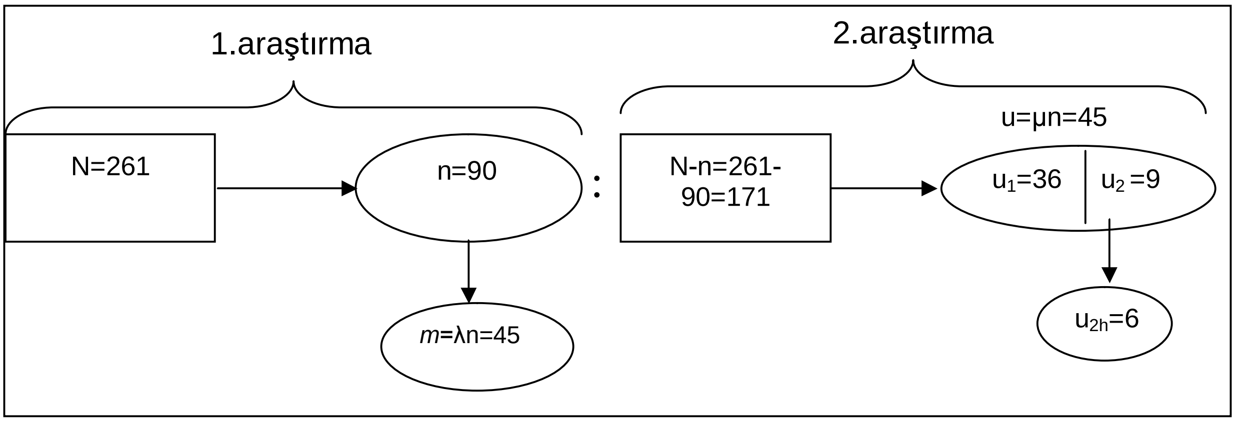
** : Tüm incelenen değerler içinde en büyük etkinlik miktarı

*** : Tüm incelenen değerler içinde en düşük etkinlik miktarı

Çizelge (6.3)-Çizelge (6.4) incelendiğinde, en büyük etkinlik miktarını veren *W*, *k*, μ değerleri sırasıyla 0,2; 1,5 ve 0,5'dir. Bu değerler, *W* ve *k*'nın incelenen en küçük, μ'nün ise incelenen en büyük değerleridir. Buradan, bu sayısal örnek için, kitledeki cevapsızlık oranı, $W = \frac{N_2}{N}$'nin yani kitlede cevapsızlık oranının küçük, ikinci araştırmada cevapsızlardan çekilecek alt kümenin ise büyük olmasının, tahmin edicinin etkinliğini artırdığı söylenebilir. Aynı zamanda, bu sayısal örnek için μ oranı büyüdükçe dolayısıyla λ oranı küçüldükçe yani birinci araştırmada eşleştirmek için ayrılan örneklem parçası küçülüp ikinci araştırmada çekilen güncel örneklem büyüklüğü arttıkça etkinliğin arttığı görülmektedir. Singh ve Priyanka (2007) tahmin edicisi, ρ_{yx}'e bağlı olduğu için, ρ_{yz}'deki değişim varyansı etkilememektedir. ρ_{yz} arttıkça, önerilen tahmin edicinin etkinliği artmaktadır. ρ_{yx}, kitle değeri 0,97 gibi yüksek bir değer olduğu için, ρ_{yx}'deki azalmanın etkinliğe etkisini görebilmek için, $\rho_{yx} = 0{,}8$ alınarak HKO değerleri incelenmiştir. Sonuçta, $\rho_{yx} = 0{,}8$ için, Çizelge (6.4) incelendiğinde, en büyük etkinliği veren *W*, *k* ve μ değerleri değişmezken, önerilen tahmin edicinin Singh ve Priyanka (2007) tahmin edicisine göre etkinlik miktarının daha da arttığı görülmektedir. Buna göre, bu sayısal örnek için, ρ_{yx} azaldıkça tahmin edicilerin HKO'ları arasındaki farkın açıldığı söylenebilir.

Eşitlik (5.11)'deki koşul bu kitle değerleri için incelendiğinde, $\frac{1}{2}\frac{C_z}{C_y} = 0{,}54 < \rho_{yz} = 0{,}94$ eşitsizliği sağlanmaktadır. Bu sayısal örnekte, her durumda, önerilen tahmin edicinin daha etkin olması beklenmektedir. Çizelge (6.3) ve Çizelge (6.4) bunu desteklemektedir.

En yüksek etkinlik oranını veren *W*, *k*, μ değerlerine göre tahmin değerleri hesabı için öncelikle, 261 birimlik kitleden ilk araştırma için *n*=90 birimlik örneklem çekilir. μ=0,5 olarak alındığından, $1 - \mu = \lambda = 0{,}5$ olur. Buna göre, *n* birimlik örneklemden $m = n\lambda = 90 * 0{,}5 = 45$ birimlik alt örneklem ikinci araştırmada kullanılmak üzere ayrılır. Eşleştirmek için ayrılan bu alt örnekleme ait bilgiler Çizelge (6.5)'de verilmiştir. Ard arda örnekleme yönteminin uygulanabilmesi için, daha öncede bahsedildiği gibi, araştırmalarda ilgilenilen kitle birimlerinin değişmemesi gerekmektedir. Birinci araştırmanın kapsadığı ilçeler ile ikinci araştırmanın kapsadığı ilçeler değişmediği için burada ard arda örnekleme yöntemi uygulanmıştır. Buna göre, birinci araştırmada 261 ilçeden çekilen 90 birimlik örneklem çıkarılarak, ikinci araştırmada geriye kalan 171 ilçeden, $u = n - m = 90 - 45 = 45$ birimlik örneklem rastgele seçilmiştir. Kitle cevapsızlık oranı, W=0,2 olduğu için, örneklem cevapsızlık oranı da 0,2 alınmıştır. Buna göre, ikinci araştırmada seçilen u=45 birimlik güncel örneklem, rasgele u_1=36 cevaplı birim, u_2=9 cevapsız birim olmak üzere iki alt örnekleme ayrılmıştır. Cevapsızların oluşturduğu u_2 alt örnekleminden, $k = u_2 / u_{2h} = 1{,}5$ olarak alındığından, 6 birimlik u_{2h} alt örneklemi rastgele seçilmiştir. Buna göre örneklem görüntüsü, Şekil (6.1)'de, u_1 ve u_{2h} örneklem bilgileri ise Çizelge (6.6)'da verilmiştir.



Şekil 6.1. Sayısal örnek verileri için ard arda örnekleme yöntemi görüntüsü

Çizelge (6.5) 2005-2006 Öğretim Yılı İçin İlköğretimdeki Öğretmen Sayısı (*X*), İlköğretime Takviye Dershane Sayısı (*Z*), OKS'yi Kazanan Öğrenci Sayısı (*Y*) Değişkenlerine Ait "*m*" Alt Örneklem Bilgileri

m=45		
$\bar{y}_m = 288{,}73$	$\bar{x}_m = 393{,}11$	$\bar{z}_m = 2{,}60$
$s^2_{y(m)} = 295439{,}7$	$s^2_{x(m)} = 437542{,}8$	$s^2_{z(m)} = 22{,}38$
$b_{yx} = 0{,}8$		

Çizelge (6.6) 2006-2007 Öğretim Yılı İçin İlköğretimdeki Öğretmen Sayısı (*X*), İlköğretime Takviye Dershane Sayısı (*Z*), OKS'yi Kazanan Öğrenci Sayısı (*Y*) Değişkenlerine Ait "u_1" Örneklem Parçası ve "u_{2h}" alt Örneklem Bilgileri

u_1=36 , u_{2h}=6		
$\bar{y}_{u1} = 117{,}19$	$\bar{x}_{u1} = 211{,}75$	$\bar{z}_{u1} = 1{,}22$
$\bar{y}_{u2h} = 136{,}33$	$\bar{x}_{u2h} = 185{,}83$	$\bar{z}_{u2h} = 1{,}67$

Bu bilgiler doğrultusunda, ikinci araştırma kitle ortalaması tahmini için, önerilen tahmin edici ($T_{öneri}$) ve Singh ve Priyanka (2007) tahmin edicisi (T) için optimal katsayıları, sırasıyla Eşitlik (5.6) ve Eşitlik (4.42) kullanılarak γ_{opt} ve φ_{opt} değerleri elde edilmiş ve tahmin değerleri aşağıdaki gibi hesaplanmıştır:

$$\left.\begin{aligned}
\gamma_{opt} &= \frac{V(T_2)}{HKO(T_{öneri1}) + V(T_2)} = \frac{1368{,}12}{524{,}69 + 1368{,}12} = 0{,}72 \\
T_{öneri1} &= \frac{\bar{y}_u^*}{\bar{z}_u^*}\bar{Z} = \frac{121{,}02}{1{,}31} 2{,}07 = 191{,}23 \\
T_2 &= \bar{y}_m + b_{yx}(\bar{x}_n - \bar{x}_m) = 288{,}73 + 0{,}8(333{,}42 - 393{,}11) = 240{,}8
\end{aligned}\right\} T_{öneri} = 204{,}97$$

$$\left.\begin{aligned}
\varphi_{opt} &= \frac{V(T_2)}{V(T_1) + V(T_2)} = \frac{1368{,}12}{524{,}69 + 1368{,}12} = 0{,}28 \\
T_1 &= \bar{y}_u^* = 121{,}019 \\
T_2 &= 240{,}799
\end{aligned}\right\} T = 207{,}5$$

2006-2007 öğretim yılında ilçe başına OKS'yi kazanan ortalama öğrenci sayısı 179,77 iken, bu sayı önerilen tahmin ediciye göre 204,97, Singh ve Priyanka (2007) tahmin edicisine göre 207,5 olarak tahmin edilmiştir.

YEDİNCİ BÖLÜM

7. SONUÇ VE TARTIŞMA

Örneklem araştırmalarında cevapsızlık sıklıkla karşılaşılan bir problemdir. Cevapsızlık olması durumunda analiz yöntemleri farklılık göstermektedir. Kitle ortalaması, araştırmalarda en çok araştırılan konulardan biridir. Kitle karakteristikleri tahmini için, ulaşılabilir her bilginin kullanımı tahmin duyarlılığını arttırmaktadır. Bu nedenle, cevapsızlığı göz önünde bulundurarak çalışmaları sürdürmek gerekmektedir. Ayrıca tekrarlayan araştırmalarda, önceki araştırma bilgilerini kullanan ard arda örnekleme yönteminin, zaman, maliyet ve tahmin duyarlılığı bakımından avantaj sağladığı bilinmektedir. Ard arda örnekleme yönteminde cevapsızlık durumu, son zamanlarda çalışılmaya başlanmış bir konudur. Kitle ortalaması tahmini için sınırlı sayıda çalışma vardır. İlgilenilen değişken ve yardımcı değişken arasındaki korelasyon bilgisine göre, oransal tahmin edicilerin kullanılması da tahmin duyarlılığını artırmaktadır. Bu bilgiler ışığında, bu çalışmada cevapsızlık olması durumunda, ard arda örnekleme yönteminde yeni bir oransal tahmin edici önerilmiş ve Singh ve Priyanka (2007) tahmin edicisi ile karşılaştırılmıştır.

Önerilen tahmin edici $\frac{1}{2}\frac{C_z}{C_y} < \rho_{yz}$ koşulu altında Singh ve Priyanka (2007) tahmin edicisinden daha etkindir. Altıncı bölümde verilen sayısal örnek araştırma verileri bu koşulu sağlamaktadır. İlgilenilen bütün kitle cevapsızlık oranı (*W*), ikinci araştırmada cevapsızlardan çekilen alt örneklem oranı (*k*), yeni örnekleme oranı (μ) ve ilişki miktarları ρ_{yz} ve ρ_{yx} değerlerinde, önerilen tahmin edicinin daha etkin olduğu görülmektedir. ρ_{yz} arttıkça, önerilen tahmin edicinin etkinliği artmaktadır. Bu sayısal örnek için en büyük etkinlik oranını veren *W*, *k*, μ değerleri son araştırmada en fazla bilgiyi sağlayan sırasıyla 0,2; 1,5 ve 0,5 olarak elde edilmiştir.

Bu çalışmada, *Z* yardımcı değişkenin kitle ortalamasının bilindiği varsayılmıştır. Yardımcı değişkenin kitle ortalamasının bilinmediği durumda oransal tahmin edici geliştirilebilir. Yardımcı değişkende de kayıp veri olması ayrıca incelenebilir. Bunun yanında, ard arda örnekleme yönteminde örneklemler yerine konmadan basit rastgele örnekleme yöntemine göre seçilmektedir. Literatürde tabakalı rastgele örnekleme yöntemine göre ard arda örnekleme yönteminde yapılmış çalışmalara rastlanmamıştır. Tahmin ediciler tabakalı rastgele örnekleme yöntemine uyarlanabilir. Ayrıca cevapsızlık olması durumunda ard arda örnekleme yönteminde kitle ortalaması tahmini için regresyon tahmin ediciler geliştirilebilir.

Cevapsızlık olması durumunda ard arda örnekleme yönteminde cevapsızlardan tekrar bir alt örneklem çekilmesi yapısal olarak iki safhalı örneklemi çağrıştırmaktadır. Bu konuda elde edilmiş etkin tahmin ediciler, ard arda örnekleme yöntemine uyarlanarak çalışma genişletilebilir.

KAYNAKLAR

[1] Adhavryu, D., Gupta, P.C., 1983, On some alternative sampling strategies using auxiliary information, *Metrika*, 30, 217-226.

[2] Chakrabarty, R.P., 1968, *Contribution to the Theory of Ratio-type Estimators*, Ph.D. Thesis, Texas A & M University.

[3] Cochran, W. G., 1977, *Sampling Techniques*, John Wiley&Sons, New York.

[4] Çıngı, H., 1994, *Örnekleme Kuramı*, H.Ü. Fen Fakültesi Basımevi, Beytepe.

[5] Çıngı, H., Kadılar, C., Koçberber, G., 2007, *Türkiye Genelinde İlk ve Orta Öğretim Olanaklarının İncelenmesi ve Belirlenen Aksaklıklara Çözüm Önerilerinin Getirilmesi*, TÜBİTAK, SOBAG106K077.

[6] Das, A.K., 1982, Estimation of population ratio on two occasions, *Journal of the Indian Society of Agricultural Statistics*, 34, 1-9.

[7] Donders, A.R.T., Heijden, G.J.M.G., Stijnen, T., Moons, K.G.M., 2006, Review: A gentle introduction to imputation of missing values, *Journal of Clinical Epidemiology*, 59, 1087-1091.

[8] Gamrot, W., 2007, Mean value estimation using two-phase samples with missing data in both phases, *Acta Applicandae Mathematics, vol.96, 215-220.*

[9] Gonzalez, S., Rueda, M., Arcos, A., 2008, An improved estimator to analyse missing data, *Statistical Papers*, vol. 49(4), 791-796.

[10] Gupta, P.C., 1979, Sampling on two successive occassions, *Journal of Statistical Research*, 13, 7-16.

[11] Hansen, M.H., Hurwitz, W.N., 1946, The problem of nonresponse in sample surveys, *American Statistical Association*, 41, 517-529.

[12] Kadılar, C., Çıngı, H., 2008, Estimators for the population mean in the case of missing data, *Communications in Statistics Theory and Statistics*: 37:14, 2226-2236.

[13] Little, R.J.A. , Rubin, D.B., 2002, *Statistical Analysis With Missing Data*, Wiley-Interscience, New Jersey, USA, Second Edition.

[14] Longford, N. T., 2005, *Missing Data and Small-Area Estimation*, Springer, New York.

[15] Nargundkar, M.S., Joshi, G.B., 1975, *Non-response in Sample Surveys*, 40th Session of the ISI-Warsaw, 626-628.

[16] Perri, P.F., Diana, G., 2009, Improved estimators of population mean for missing data, *Communications in Statistics Theory and Statistics:* basım aşamasında.

[17] Rao, J.N.K., Graham, J.E., 1964, Rotation design for sampling on repeated occassions, *Journal American Statistical Association*, 59, 492-509.

[18] Rao, J.N.K., 1994, Estimating totals and distribution functions using auxiliary information at the estimation stage, *Journal of Official Statistics*, 10(2),153-165.

[19] Rueda, M.M., Gonzalez, S., Arcos, A., Roman, Y., Martinez, M.D., Munoz, J.F. (2005), *Estimation of the population mean using auxiliary information when some observations are missing*, In Proceedings of XIth International Symposium on Applied Stochastic Models and Data Analysis (ASMDA 2005), Pages 1493-1498, Brest, France, 17-20 May 2005.

[20] Sarndal, C.E., Swenson, B., Wretman, J., 1992, *Model Assisted Survey Sampling*, Springer, New York.

[21] Satıcı, E., Kadılar, C., 2009, Kayıp gözlem olduğunda kitle ortalamasının tahmini, *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi*, basım aşamasında.

[22] Sen, A.R., 1971, Successive sampling with two auxiliary variables, *Sankhya*, 33, 371-378.

[23] Sen, A.R., 1973, Theory and application of sampling on repeated occasions with several auxiliary variables, *Biometrics*, 29, 381-385.

[24] Singh, G.N., Singh V.K., 2001, On the use of auxiliary information in successive sampling, *Journal of the Indian Society Agricultural Statistics*, 54(1), 1-12.

[25] Singh, S., Deo, B., 2003, Imputation by power transformation, *Statistical Papers*, 44 (4), 555-579.

[26] Singh, G.N., 2005, On the use of chain-type ratio estimator in successive sampling, *Statistics in Transition*, 7(1), 21-26.

[27] Singh, G.N., Priyanka, K., 2007, Effect of non-response on current occasion in search of good rotation patterns on successive occasions, *Statistics in Transition*, 8 (2), 273-292.

[28] Singh, G.N., Priyanka, K., 2008, Search of good rotation patterns to improve the precision of estimates at current occasion, *Communications in Statistics: Theory and Methods*, 37, 337-348.

[29] Singh, S., Horn, S., 2000, Compromised imputation in survey sampling, *Metrika*, 51, 267-276.

[30] Srivastava, S.K., 1967, An estimator using auxiliary information in sample surveys, *Calcutta Statistical Association Bulletin*, 16, 121-132.

[31] Sukhatme, P.V., Sukhatme, B. V., Sukhatme, S., Ashok, C., 1984, *Sampling Theory of Surveys with Applications*, Iowa State University Press, Iowa, *USA, 3rd Revised Edition*.

[32] Toutenburg, H., Srivastava, V., K., Shalabh, 2008, Amputation versus imputation of missing values through ratio method in sample surveys, *Statistical Papers*, vol. 49(2), 237-247.

[33] Vos, J.W.E., 1980, Mixing of direct, ratio and product method estimators, *Statistica Neerlandica*, 34, 209-213.

[34] Yazıcı, F., 2005, *EM Algoritması ve Uzantıları*, Yayınlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

EK 1: HORVITZ-THOMPSON TAHMİN EDİCİLERİ

1,2,…N şeklinde numaralandırılmış, N tane birimden oluşan sonlu bir kitleden n büyüklüğünde s örnekleminin çekildiği varsayılsın. y_i, y ilgilenilen değişkeninin, kitledeki *i.* birim değerini göstersin. Örneklem yöntemine bağlı yaklaşımda, y_i'ler bilinmeyen sabitlerdir. Kitle toplamı tahmini için Z_i rastlantı değişkeni aşağıdaki gibi tanımlansın;

$$Z_i = \begin{cases} 1 & , i. \text{ birim örnekleme seçiliyorsa} \\ 0 & , \text{diğer durumlarda} \end{cases}$$

Basit rasgele örneklemede birimlerin örnekleme seçilme olasılıkları eşittir. Fakat uygulamada her zaman, birimlere eşit örnekleme seçilme olasılığı verilmeyebilir. Birimlere farklı seçilme olasılıkları veren örnekleme yöntemine göre, kitle parametresi için yansız bir tahmin edici Horvitz-Thompson (1946) tahmin edicisidir. Horvitz-Thompson tahmin edicisi,

$$\hat{Y} = \sum_{i=1}^{N} Z_i \frac{y_i}{\pi_i} = \sum_{i \in s} \frac{y_i}{\pi_i}$$

olarak tanımlanır. Bu tahmin edicinin dağılımı birim numaralarına, yani, Z_i rastlantı değişkenine bağımlı bir yapıdadır. Buradan,

$$E(\hat{Y}) = E\left(\sum_{i=1}^{N} Z_i \frac{y_i}{\pi_i} \right) = \sum_{i=1}^{N} \pi_i \frac{y_i}{\pi_i} = Y$$

biçiminde yansız olduğu gösterilebilmektedir. Horvitz-Thompson tahmin edicisi formülü incelendiğinde, seçilme olasılıkları düşük olan birimlerin katkısının, yüksek olasılıklı birimlerin katkısından daha fazla olduğu açıkça görülmektedir.

EK 2: $\bar{y}_{SD1}$ TAHMİN EDİCİSİNİN YAN VE HKO EŞİTLİKLERİNİN ELDE EDİLİŞİ

$$\bar{y}_{SD1} = \bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r} \right)^{\alpha}$$

$$\varepsilon = \frac{\bar{y}_r - \bar{Y}}{\bar{Y}} \ , \ \delta = \frac{\bar{x}_r - \bar{X}}{\bar{X}} \ , \ \eta = \frac{\bar{x}_n - \bar{X}}{\bar{X}}$$

biçiminde alındığında,

$$\bar{y}_{SD1} = \bar{Y}(1+\varepsilon)(1+\eta)^{\alpha}(1+\delta)^{-\alpha}$$

olur. Binom açılımları yerine koyulduğunda ve 3. dereceden terimler ihmal edildiğinde tahmin edici,

$$\bar{y}_{SD1} = \bar{Y}(1+\varepsilon)(1+\alpha\eta + \frac{\alpha(\alpha-1)}{2}\eta^2)(1-\alpha\delta + \frac{\alpha(\alpha+1)}{2}\delta^2)$$

$$\begin{aligned}
&= \bar{Y}\Big(1-\alpha\delta + \frac{\alpha(\alpha+1)}{2}\delta^2 + \alpha\eta - \alpha^2\eta\delta + \frac{\alpha^2(\alpha+1)}{2}\eta\delta^2 + \frac{\alpha(\alpha+1)}{2}\eta^2 - \frac{\alpha^2(\alpha-1)}{2}\delta\eta^2 \\
&\quad + \frac{\alpha^2(\alpha-1)(\alpha+1)}{4}\eta^2\delta^2 + \varepsilon - \alpha\delta\varepsilon + \frac{\alpha(\alpha+1)}{2}\delta^2\varepsilon + \alpha\eta\varepsilon - \alpha^2\eta\delta\varepsilon + \frac{\alpha^2(\alpha+1)}{2}\eta\delta^2\varepsilon \\
&\quad + \frac{\alpha(\alpha-1)}{2}\eta^2\varepsilon - \frac{\alpha^2(\alpha-1)}{2}\delta\eta^2\varepsilon + \frac{\alpha^2(\alpha-1)(\alpha+1)}{4}\eta^2\delta^2\varepsilon \ \Big)
\end{aligned}$$

$$\bar{y}_{SD1} \cong \bar{Y}\Big(\ 1-\alpha\delta + \frac{\alpha(\alpha+1)}{2}\delta^2 + \alpha\eta - \alpha^2\eta\delta + \frac{\alpha(\alpha-1)}{2}\eta^2 + \varepsilon - \alpha\delta\varepsilon + \alpha\eta\varepsilon \ \Big)$$

$$\bar{y}_{SD1} - \bar{Y} \cong \bar{Y}\Big(\ \varepsilon - \alpha\delta + \alpha\eta - \alpha\delta\varepsilon + \alpha\eta\varepsilon - \alpha^2\eta\delta + \frac{\alpha(\alpha+1)}{2}\delta^2 + \frac{\alpha(\alpha-1)}{2}\eta^2 \ \Big)$$

Her iki tarafın beklenen değeri alınıp gerekli sadeleştirmeler yapıldığında yan miktarına ulaşılır,

$$Yan(\bar{y}_{SD1}) \cong \bar{Y}\left(\frac{1}{r} - \frac{1}{n}\right)\left[\frac{\alpha(\alpha+1)}{2}C_x^2 - \alpha C_{xy}\right]$$

$\bar{y}_{SD1} - \bar{Y}$ ifadesinin karesinin beklenen değeri alındıktan sonra ve gerekli sadeleştirmeler yapıldığında, $HKO(\bar{y}_{SD1})$ ifadesine ulaşılır. $HKO(\bar{y}_{SD1})$ eşitliğinin, α'ya göre türevi alınıp sıfıra eşitlenerek optimal α değeri elde edilir ve bu değer $HKO(\bar{y}_{SD1})$ eşitliğinde yerinde konularak $HKO_{\min}(\bar{y}_{SD1})$ eşitliğine ulaşılır. Bu eşitlikler aşağıda verilmiştir:

$$\alpha_{opt} = \rho\frac{C_y}{C_x} \Rightarrow \quad HKO_{\min}(\bar{y}_{SD1}) = HKO(\bar{y}_{ORAN}) - \left(\frac{1}{r} - \frac{1}{n}\right)S_x^2\left(\frac{S_{xy}}{S_x^2} - \frac{\bar{Y}}{\bar{X}}\right)^2.$$

EK 3: $\bar{y}_{t1}$ TAHMİN EDİCİSİNİN YAN EŞİTLİĞİNİN ELDE EDİLİŞİ

Taylor Yöntemi özelliğinden,

$$R_k(\hat{Y}_k, \alpha) \cong \sum_{j=1}^{k}\sum_{i=1}^{k}\frac{1}{2}\underbrace{\frac{\partial^2 (Y_1,\ldots,Y_k)}{\partial Y_j \partial Y_i}}_{d_{ij}}(\hat{Y}_j - Y_j)(\hat{Y}_i - Y_i)$$

$$R_2 \cong \sum_{j=1}^{2}\sum_{i=1}^{2}\frac{1}{2}\frac{\partial^2 (Y_1, Y_2)}{\partial Y_j \partial Y_i}(\hat{Y}_j - Y_j)(\hat{Y}_i - Y_i)$$

$$\cong \frac{1}{2}\Big[\, d_{11}\left(\hat{Y}_1 - Y_1\right)\left(\hat{Y}_1 - Y_1\right) + d_{21}\left(\hat{Y}_1 - Y_1\right)\left(\hat{Y}_2 - Y_2\right) + d_{12}\left(\hat{Y}_2 - Y_2\right)\left(\hat{Y}_1 - Y_1\right)$$
$$+ d_{22}\left(\hat{Y}_2 - Y_2\right)\left(\hat{Y}_2 - Y_2\right)\,\Big]$$

$$\hat{Y}_1 = \bar{y}_r \quad , \quad \hat{Y}_2 = \bar{x}_r \Rightarrow$$

$$d_{11} = \frac{\partial}{\partial \bar{y}_r}\frac{r\bar{x}_r + (n-r)\bar{X}}{n\bar{x}_r} = 0$$

$$d_{12} = d_{21} = \frac{\partial}{\partial \bar{x}}\frac{r\bar{x}_r + (n-r)\bar{X}}{n\bar{x}_r} = \frac{rn\bar{x}_r - n\left[r\bar{x}_r + (n-r)\bar{X}\right]}{n^2 \bar{x}_r^2}$$

$$= \frac{n^2\bar{x}_r - n(n-r)\bar{x}_r - n(n\bar{x}_r - (n-r)\bar{x}_r + (n-r)\bar{X})}{n^2\bar{x}_r^2}\Big|_{\bar{X},\bar{Y}} = -\frac{n-r}{n}\frac{1}{\bar{X}}$$

$$d_{22} = \frac{\partial}{\partial \bar{x}_r}\bar{y}_r\left[\frac{rn\bar{x}_r - n\left[r\bar{x}_r + (n-r)\bar{X}\right]}{n^2\bar{x}_r^2}\right] = \frac{\partial}{\partial \bar{x}_r}\bar{y}_r\left[-\frac{p\bar{X}}{n\bar{x}_r^2}\right]$$

$$= \bar{y}_r\left[\frac{2(n-r)\bar{X}}{n\bar{x}_r^3}\right]\Big\|_{\bar{X},\bar{Y}}$$

$$= \frac{2\bar{Y}}{\bar{X}^2}\frac{n-r}{n}$$

$$(\bar{y}_{t1} - \bar{Y}) \cong \frac{1}{2}\left[\frac{2\bar{Y}}{\bar{X}^2}\frac{n-r}{n}(\bar{x}_r - \bar{X})^2 - \frac{2(n-r)}{n}\frac{1}{\bar{X}}(\bar{x}_r - \bar{X})(\bar{y}_r - \bar{Y})\right]$$

$$E(\bar{y}_{t1} - \bar{Y}) \cong -\frac{1}{\bar{X}} S_{xy} E_p \frac{1}{r} \frac{(n-r)}{n} + \frac{\bar{Y}}{\bar{X}^2} S_x^2 E_p \frac{1}{r} \frac{(n-r)}{n}$$

$$= -\frac{1}{\bar{X}} S_{xy} f_1 + \frac{\bar{Y}}{\bar{X}^2} S_x^2 f_1$$

$$= f_1 S_x^2 \frac{\bar{Y}}{\bar{X}^2} - f_1 \frac{1}{\bar{X}} \rho S_x S_y$$

$$= f_1 \frac{S_x S_y}{\bar{X}} \theta - f_1 \frac{1}{\bar{X}} \rho S_x S_y$$

$$= \frac{S_x S_y}{\bar{X}} f_1 (\theta - \rho)$$

EK 4: $\bar{y}_{SK1}$ TAHMİN EDİCİSİNİN HKO EŞİTLİĞİNİN ELDE EDİLİŞİ

Taylor serisi açılımı

$$h\left(\hat{Y}_1,\hat{Y}_2,\ldots,\hat{Y}_p\right)\cong h\left(Y_1,Y_2,\ldots,Y_p\right)+\sum_{i=1}^{p} d_j\left(\hat{Y}_j-Y_j\right)+R\left(\hat{Y},Y\right)$$

biçiminde tanımlanmaktadır. Burada p, parametre sayısıdır ve $R\left(\hat{Y},Y\right)$ kalanı göstermekte olup birinci dereceden Taylor serisi yaklaşımında ihmal edilir. $\bar{y}_{SK1}$ tahmin edicisi için p=2 olmaktadır ve

$$d_j=\frac{\partial h(t)}{\partial \hat{Y}_j}\Big|_{t=T}\ ,\ h(t)=h\left(\bar{y}_r,\bar{x}_r\right)=\bar{y}_{SK1}\ ,\ \hat{Y}_1=\bar{y}_r\ ,\ \hat{Y}_2=\bar{x}_r\ .$$

Buna göre,

$$d_1=\frac{\partial h(t)}{\partial \hat{Y}_1}\Big|_{\bar{X},\bar{Y}}=\left(\frac{\bar{X}}{\bar{x}_r}\right)^{\alpha}\Big|_{\bar{X},\bar{Y}}=1$$

$$d_2=\frac{\partial h(t)}{\partial \hat{Y}_2}\Big|_{\bar{X},\bar{Y}}=\bar{y}_r\bar{X}^{\alpha}\frac{-\alpha}{\bar{x}_r^{\alpha+1}}\Big|_{\bar{X},\bar{Y}}=-\alpha\frac{\bar{Y}}{\bar{X}}$$

olarak bulunur. Değerler yerine yazıldığında,

$$\bar{y}_{SK1}\cong\bar{Y}+\left(\bar{y}_r-\bar{Y}\right)-\alpha\frac{\bar{Y}}{\bar{X}}\left(\bar{x}_r-\bar{X}\right)\ ,$$

$$\bar{y}_{SK1}-\bar{Y}\cong\left(\bar{y}_r-\bar{Y}\right)-\alpha\frac{\bar{Y}}{\bar{X}}\left(\bar{x}_r-\bar{X}\right)$$

olur ve her iki tarafın karesi alınarak beklenen değer alındığında,

$$E\left(\bar{y}_{SK1}-\bar{Y}\right)^2\cong V\left(\bar{y}_r\right)+\alpha^2\frac{\bar{Y}^2}{\bar{X}^2}V\left(\bar{x}_r\right)-2\alpha\frac{\bar{Y}}{\bar{X}}Cov\left(\bar{x}_r,\bar{y}_r\right)$$

biçiminde elde edilir. Burada,

$$V\left(\bar{y}_r\right)=\left(\frac{1}{r}-\frac{1}{N}\right)S_y^2\ ,\ V\left(\bar{x}_r\right)=\left(\frac{1}{r}-\frac{1}{N}\right)S_x^2\ \text{ve}\ Cov\left(\bar{x}_r,\bar{y}_r\right)=\left(\frac{1}{r}-\frac{1}{N}\right)S_{xy}$$

eşitlikleri yerine yazıldığında,

$$HKO\left(\bar{y}_{SK1}\right)\cong\left(\frac{1}{r}-\frac{1}{N}\right)S_y^2+\alpha^2\frac{\bar{Y}^2}{\bar{X}^2}\left(\frac{1}{r}-\frac{1}{N}\right)S_x^2-2\alpha\frac{\bar{Y}}{\bar{X}}\left(\frac{1}{r}-\frac{1}{N}\right)S_{xy}$$

elde edilir.

$HKO(\bar{y}_{SK1})$ ifadesinde $\bar{Y}^2$ ortak parantezine alınırsa,

$$HKO(\bar{y}_{SK1}) \cong \bar{Y}^2 \left\{ \left(\frac{1}{r} - \frac{1}{N} \right) \frac{S_y^2}{\bar{Y}^2} + \alpha^2 \left(\frac{1}{r} - \frac{1}{N} \right) \frac{S_x^2}{\bar{X}^2} - 2\alpha \left(\frac{1}{r} - \frac{1}{N} \right) \frac{S_{xy}}{\bar{Y}\bar{X}} \right\}$$

$$= \bar{Y}^2 \left\{ \left(\frac{1}{r} - \frac{1}{N} \right) C_y^2 + \alpha^2 \left(\frac{1}{r} - \frac{1}{N} \right) C_x^2 - 2\alpha \left(\frac{1}{r} - \frac{1}{N} \right) \rho C_x C_y \right\}$$

biçiminde elde edilir. Burada $C_x = \frac{S_x}{\bar{X}}$, $C_y = \frac{S_y}{\bar{Y}}$ ve $\rho = \frac{S_{xy}}{S_x S_y}$ göstermektedir.

Optimal α değeri için,

$$\frac{\partial HKO(\bar{y}_{SK1})}{\partial \alpha} = \bar{Y}^2 \left\{ 2\alpha \left(\frac{1}{r} - \frac{1}{N} \right) C_x^2 - 2 \left(\frac{1}{r} - \frac{1}{N} \right) \rho C_x C_y \right\} = 0$$

eşitliği çözüldüğünde,

$$\alpha_{\text{opt}} = \rho \frac{\mathrm{C_y}}{\mathrm{C_x}}$$

olarak bulunur. Bu ifade $HKO(\bar{y}_{SK1})$'de yerine yazılırsa,

$$HKO_{\min}(\bar{y}_{SK1}) \cong \left(\frac{1}{r} - \frac{1}{N} \right) S_y^2 + \rho^2 \frac{C_y^2}{C_x^2} \frac{\bar{Y}^2}{\bar{X}^2} \left(\frac{1}{r} - \frac{1}{N} \right) S_x^2 - 2\rho \frac{C_y}{C_x} \frac{\bar{Y}}{\bar{X}} \left(\frac{1}{r} - \frac{1}{N} \right) S_{xy}$$

elde edilir. Gerekli sadeleştirmeler yapıldığında

$$HKO_{\min}(\bar{y}_{SK1}) = \left(\frac{1}{r} - \frac{1}{N} \right) S_y^2 \left(1 - \rho^2 \right)$$

biçiminde elde edilir.

EK 5: $\bar{y}_{DP1}$, $\bar{y}_{DP2}$, $\bar{y}_{DP3}$ TAHMİN EDİCİLERİNİN HKO EŞİTLİĞİNİN ELDE EDİLİŞİ

Taylor serisi açılımı

$$h\left(\hat{Y}_1,\hat{Y}_2,...,\hat{Y}_p\right)\cong h\left(Y_1,Y_2,...,Y_p\right)+\sum_{i=1}^{p} d_j\left(\hat{Y}_j-Y_j\right)+R\left(\hat{Y},Y\right)$$

biçiminde tanımlanmaktadır. Burada *p*, parametre sayısıdır ve $R\left(\hat{Y},Y\right)$ kalanı göstermekte olup birinci dereceden Taylor serisi yaklaşımında ihmal edilir. $\bar{y}_{DP1}$ tahmin edicisi için *p*=2 olmaktadır ve

$$d_j=\frac{\partial h(t)}{\partial \hat{Y}_j}\Big|_{\bar{Y},\bar{X}} \qquad h(t)=h(\bar{y}_r,\bar{x}_n)$$

Buna göre,

$$d_1=\frac{\partial h(t)}{\partial \hat{Y}_1}\Big|_{\bar{X},\bar{Y}}=1$$

$$d_2=\frac{\partial h(t)}{\partial \hat{Y}_2}\Big|_{\bar{X},\bar{Y}}=-b$$

olarak bulunur.

$$\bar{y}_{DP1}\cong\bar{Y}+\left(\bar{y}_r-\bar{Y}\right)-b\left(\bar{x}_n-\bar{X}\right)$$

$$\bar{y}_{DP1}-\bar{Y}\cong\left(\bar{y}_r-\bar{Y}\right)-b\left(\bar{x}_n-\bar{X}\right)$$

her iki tarafın karesi alınarak beklenen değeri alındığında,

$$HKO\left(\bar{y}_{DP1}\right)\cong V\left(\bar{y}_r\right)-2BCov\left(\bar{y}_r,\bar{x}_n\right)+B^2V\left(\bar{x}_n\right)$$

$$V\left(\bar{y}_r\right)=\left(\frac{1}{r}-\frac{1}{N}\right)S_y^2 \quad , \quad Cov\left(\bar{y}_r,\bar{x}_n\right)=\left(\frac{1}{n}-\frac{1}{N}\right)S_{xy} \quad , \quad V\left(\bar{x}_n\right)=\left(\frac{1}{n}-\frac{1}{N}\right)S_x^2$$

$$\begin{aligned}HKO\left(\bar{y}_{DP1}\right)&\cong\left(\frac{1}{r}-\frac{1}{N}\right)S_y^2-2\frac{S_{xy}}{S_x^2}\left(\frac{1}{n}-\frac{1}{N}\right)S_{xy}+\frac{S_{xy}^2}{S_x^4}\left(\frac{1}{n}-\frac{1}{N}\right)S_x^2\\&=S_y^2\left[\left(\frac{1}{r}-\frac{1}{N}\right)-2\left(\frac{1}{n}-\frac{1}{N}\right)\rho^2+\left(\frac{1}{n}-\frac{1}{N}\right)\rho^2\right]\\&=S_y^2\left[\left(\frac{1}{r}-\frac{1}{N}\right)-\left(\frac{1}{n}-\frac{1}{N}\right)\rho^2\right]\end{aligned}$$

$\mp\frac{1}{n}$ eklendiğinde,

$$= S_y^2\left[\left(\frac{1}{n}-\frac{1}{N}\right)\left(1-\rho^2\right)+\left(\frac{1}{r}-\frac{1}{N}\right)\right]$$

olur.

Benzer şekilde, $HKO\left(\bar{y}_{DP2}\right)$ için,

$$d_j = \frac{\partial h(t)}{\partial \hat{Y}_j}\Big|_{\bar{Y},\bar{X}} \qquad h(t) = h(\bar{y}_r, \bar{x}_r)$$

Buna göre,

$$d_1 = \frac{\partial h(t)}{\partial \hat{Y}_1}\Big|_{\bar{X},\bar{Y}} = 1$$

$$d_2 = \frac{\partial h(t)}{\partial \hat{Y}_2}\Big|_{\bar{X},\bar{Y}} = -b$$

olarak bulunur.

$$\bar{y}_{DP2} \cong \bar{Y} + \left(\bar{y}_r - \bar{Y}\right) - b\left(\bar{x}_r - \bar{X}\right)$$

$$\bar{y}_{DP2} - \bar{Y} \cong \left(\bar{y}_r - \bar{Y}\right) - b\left(\bar{x}_r - \bar{X}\right)$$

her iki tarafın karesi alınarak beklenen değeri alındığında,

$$HKO\left(\bar{y}_{DP1}\right) \cong V\left(\bar{y}_r\right) - 2BCov\left(\bar{y}_r, \bar{x}_r\right) + B^2 V\left(\bar{x}_r\right)$$

$$V\left(\bar{y}_r\right) = \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2 \quad , \quad Cov\left(\bar{y}_r, \bar{x}_r\right) = \left(\frac{1}{r}-\frac{1}{N}\right)S_{xy} \quad , \quad V\left(\bar{x}_r\right) = \left(\frac{1}{r}-\frac{1}{N}\right)S_x^2$$

$$HKO\left(\bar{y}_{DP2}\right) \cong \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2 - 2\frac{S_{xy}}{S_x^2}\left(\frac{1}{r}-\frac{1}{N}\right)S_{xy} + \frac{S_{xy}^2}{S_x^4}\left(\frac{1}{r}-\frac{1}{N}\right)S_x^2$$

$$= \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2\left(1-\rho^2\right)$$

olur.

$HKO\left(\bar{y}_{DP3}\right)$ için,

$$d_j = \frac{\partial h(t)}{\partial \hat{Y}_j}\Big|_{\bar{Y},\bar{X}} \qquad h(t) = h(\bar{y}_r, \bar{x}_n, \bar{x}_r)$$

Buna göre,

$$d_1 = \frac{\partial h(t)}{\partial \hat{Y}_1}\Big|_{\bar{X},\bar{Y}} = 1$$

$$d_2 = \frac{\partial h(t)}{\partial \hat{Y}_2}\Big|_{\overline{X},\overline{Y}} = b$$

$$d_3 = \frac{\partial h(t)}{\partial \hat{Y}_3}\Big|_{\overline{X},\overline{Y}} = -b$$

olarak bulunur.

$$\overline{y}_{DP3} \cong \overline{Y} + \left(\overline{y}_r - \overline{Y}\right) + b\left(\overline{x}_n - \overline{X}\right) - b\left(\overline{x}_r - \overline{X}\right)$$

$$\overline{y}_{DP3} - \overline{Y} \cong \left(\overline{y}_r - \overline{Y}\right) + b\left(\overline{x}_n - \overline{X}\right) - b\left(\overline{x}_r - \overline{X}\right)$$

Her iki tarafın karesi alınarak beklenen değeri alındığında,

$$HKO\left(\overline{y}_{DP3}\right) \cong V\left(\overline{y}_r\right) + B^2 V\left(\overline{x}_n\right) + B^2 V\left(\overline{x}_r\right) + 2BCov\left(\overline{y}_r, \overline{x}_n\right) - 2B^2 Cov\left(\overline{x}_r, \overline{x}_n\right) - 2BCov\left(\overline{x}_r, \overline{y}_r\right)$$

$$= \left(\frac{1}{r} - \frac{1}{N}\right) S_y^2 + \frac{S_{xy}^2}{S_x^4}\left(\frac{1}{n} - \frac{1}{N}\right) S_x^2 + \frac{S_{xy}^2}{S_x^4}\left(\frac{1}{r} - \frac{1}{N}\right) S_x^2 + 2\frac{S_{xy}}{S_x^2}\left(\frac{1}{n} - \frac{1}{N}\right) S_{xy}$$
$$- 2\frac{S_{xy}^2}{S_x^4}\left(\frac{1}{n} - \frac{1}{N}\right) S_x^2 - 2\frac{S_{xy}}{S_x^2}\left(\frac{1}{r} - \frac{1}{N}\right) S_{xy}$$

$$= S_y^2\left[\left(\frac{1}{r} - \frac{1}{N}\right) - \rho^2\left(\frac{1}{r} - \frac{1}{N}\right)\right]$$

$\mp \frac{1}{n}$ eklendiğinde,

$$= S_y^2\left[\left(\frac{1}{n} - \frac{1}{N}\right) + \left(\frac{1}{r} - \frac{1}{N}\right)(1 - \rho^2)\right]$$

olarak bulunur.

EK 6 : *T** TAHMİN EDİCİSİNİN YAN VE HKO EŞİTLİKLERİNİN ELDE EDİLİŞİ

- T^* tahmin edicisinin Yan eşitliğinin elde edilmesi:

$$Yan(T^*) \cong E\left(T^* - T\right) \cong E\left(k_1\overline{y}_{s1} + k_2\overline{y}_{s3} + k_3 b\overline{x}_{s2} - \overline{Y}\right)$$

$$= k_1\overline{Y} + k_2\overline{Y} + k_3 b\overline{X} - \overline{Y}$$

$$= \overline{Y}(k_1 + k_2 - 1) + k_3 b\overline{X}$$

$$k_1 + k_2 - 1 = \frac{n - p - q(N - p)}{N(n - p)} + \frac{q(N - p)}{N(n - p)} - 1$$

$$= \frac{(N - p)(n - p - q + q)}{N(n - p)} - 1$$

$$= \frac{(N - p)(n - p)}{N(n - p)} - 1$$

$$= \frac{N - p - N}{N} = -\frac{p}{N}$$

$$= -k_3$$

Yan eşitliği yerine yazıldığında;

$$Yan\left(T^*\right) \cong -k_3\overline{Y} + k_3 B\overline{X}$$

$$= k_3\left(B\overline{X} - \overline{Y}\right)$$

biçiminde elde edilir.

- T^* tahmin edicisinin HKO eşitliğinin elde edilmesi:

$$T^* = k_1\overline{y}_{s1} + k_2\overline{y}_{s3} + k_3 b\overline{x}_{s2}$$

$$h\left(Y_1, Y_2, Y_3\right) = h\left(\overline{y}_{s_1}, \overline{y}_{s_3}, \overline{x}_{s_2}\right) = T$$

$$T^* - T \Rightarrow \frac{\partial h(Y_1, Y_2, Y_3)}{\partial Y_1}\Big|_{\bar{X},\bar{Y}} = k_1$$

$$\frac{\partial h(Y_1, Y_2, Y_3)}{\partial Y_2}\Big|_{\bar{X},\bar{Y}} = k_2, \qquad \frac{\partial h(Y_1, Y_2, Y_3)}{\partial Y_3}\Big|_{\bar{X},\bar{Y}} = k_3 b$$

$$T^* - T = k_1\left(\bar{y}_{s_1} - \bar{Y}\right) + k_2\left(\bar{y}_{s_3} - \bar{Y}\right) + k_3 b\left(\bar{x}_{s_2} - \bar{X}\right)$$

Her iki tarafın karesi alınırsa,

$$\left(T^* - T\right)^2 = k_1^2\left(\bar{y}_{s_1} - \bar{Y}\right)^2 + k_2^2\left(\bar{y}_{s_3} - \bar{Y}\right)^2 + k_3^2 b^2\left(\bar{x}_{s_2} - \bar{X}\right)^2$$

$$+ 2k_1k_2\left(\bar{y}_{s_1} - \bar{Y}\right)\left(\bar{y}_{s_3} - \bar{Y}\right) + 2k_1k_3b\left(\bar{y}_{s_1} - \bar{Y}\right)\left(\bar{x}_{s_2} - \bar{X}\right) + 2k_2k_3b\left(\bar{y}_{s_3} - \bar{Y}\right)\left(\bar{x}_{s_2} - \bar{X}\right)$$

Her iki tarafın beklenen değeri alınarak yaklaşık HKO’ya ulaşılabilir;

$$E\left(T^* - T\right)^2 \cong k_1^2 E\left(\bar{y}_{s_1} - \bar{Y}\right)^2 + k_2^2 E\left(\bar{y}_{s_3} - \bar{Y}\right)^2 + k_3^2 b^2 E\left(\bar{x}_{s_2} - \bar{X}\right)^2 + 2k_1k_2E\left(\bar{y}_{s_1} - \bar{Y}\right)\left(\bar{y}_{s_3} - \bar{Y}\right)$$

$$+ 2k_1k_3bE\left(\bar{y}_{s_1} - \bar{Y}\right)\left(\bar{x}_{s_2} - \bar{X}\right) + 2k_2k_3bE\left(\bar{y}_{s_3} - \bar{Y}\right)\left(\bar{x}_{s_2} - \bar{X}\right)$$

$$HKO\left(T^*\right) \cong k_1^2 V\left(\bar{y}_{s_1}\right) + k_2^2 V\left(\bar{y}_{s_3}\right) + k_3^2 b^2 V\left(\bar{x}_{s_2}\right) + 2k_1k_2Cov\left(\bar{y}_{s_1}, \bar{y}_{s_3}\right) + 2k_1k_3bCov\left(\bar{y}_{s_1}, \bar{x}_{s_2}\right)$$

$$+ 2k_2k_3bCov\left(\bar{y}_{s_3}, \bar{x}_{s_2}\right) \qquad \text{(E1)}$$

Eşitlik (E1)’de yer alan kovaryans ve varyans fonksiyonlarının eşitlikleri aşağıda verilmiştir:

$$Cov\,(\bar{x}_n, \bar{y}_n) = \frac{1-f}{n} S_{yx} = \left(\frac{1}{n} - \frac{1}{N}\right) S_{yx}$$

$$V(\bar{y}_{s_1}) = \left(\frac{1}{n-p-q} - \frac{1}{N}\right) S_y^2 = aS_y^2$$

$$V(\bar{y}_{s_3}) = \left(\frac{1}{q} - \frac{1}{N}\right) S_y^2 = \alpha S_y^2$$

$$V(\bar{x}_{s_2}) = \left(\frac{1}{p} - \frac{1}{N}\right) S_x^2 = d S_x^2$$

$$Cov\left(\bar{y}_{s_1}, \bar{y}_{s_3}\right) = c S_y^2 \rightarrow c = \begin{cases} \dfrac{1}{n-p-q} - \dfrac{1}{N}, & q \leq n-p-q \\ \\ \dfrac{1}{q} - \dfrac{1}{N}, & q > n-p-q \end{cases}$$

$$Cov\left(\bar{y}_{s_1}, \bar{x}_{s_2}\right) = e\, S_{xy} \rightarrow e = \begin{cases} \dfrac{1}{n-p-q} - \dfrac{1}{N}, & p \leq n-p-q \\ \\ \dfrac{1}{p} - \dfrac{1}{N}, & p > n-p-q \end{cases}$$

$$Cov\left(\bar{y}_{s_3}, \bar{x}_{s_2}\right) = f\, S_{xy} \rightarrow f = \begin{cases} \dfrac{1}{p} - \dfrac{1}{N}, & p \geq \mathrm{q} \\ \\ \dfrac{1}{q} - \dfrac{1}{N}, & p < \mathrm{q} \end{cases}$$

olmaktadır. Bu değerler Eşitlik (E1)'de yerine yazılırsa;

$$\begin{aligned} HKO(T^*) &\cong k_1^2 a S_y^2 + k_2^2 \alpha S_y^2 + k_3^2 B^2 d S_x^2 + 2k_1 k_2 c S_y^2 + 2k_1 k_3 B e S_{xy} + 2k_2 k_3 B f S_{xy} \\ &= S_y^2 \left(k_1^2 a + k_2^2 \alpha + 2k_1 k_2 c\right) + B^2 k_3^2 d S_x^2 + 2k_3 B S_{xy} \left(e k_1 + f k_2\right) \end{aligned}$$

biçiminde elde edilir.

EK 7: $HKO(\bar{y}_{SK2})$ VE $HKO(\bar{y}_{SD})$ EŞİTLİKLERİNİN KARŞILAŞTIRILMASI

$HKO(\bar{y}_{SK2}) < HKO(\bar{y}_{SD})$ 'de ilgili eşitlikler yerine yazılarak,

$$\frac{(N-(n-r))^2}{N^2}\left(\frac{1}{r}-\frac{1}{N}\right)S_y^2+\frac{(n-r)^2}{N^2}\left(\frac{1}{n-r}-\frac{1}{N}\right)\frac{S_{xy}^2}{S_x^2}+2\frac{N-(n-r)}{N}\frac{n-r}{N}\left(\frac{1}{r}-\frac{1}{N}\right)\frac{S_{xy}^2}{S_x^2}$$

$$<\left(\frac{1}{r}-\frac{1}{N}\right)S_y^2-\left(\frac{1}{r}-\frac{1}{n}\right)\frac{S_{xy}^2}{S_x^2},$$

$$\frac{S_{xy}^2}{S_x^2}\left[\frac{1}{r}-\frac{1}{n}+\frac{p^2}{N^2}\left(\frac{1}{n-r}-\frac{1}{N}\right)+2\frac{(n-r)(N-(n-r))}{N^2}\left(\frac{1}{r}-\frac{1}{N}\right)\right]$$

$$<S_y^2\left(\frac{1}{r}-\frac{1}{N}\right)\left[1-\frac{\left(N-(n-r)\right)^2}{N^2}\right],$$

$$\rho^2\left[\frac{1}{r}-\frac{1}{n}+\frac{(n-r)^2}{N^2}\left(\frac{1}{n-r}-\frac{1}{N}\right)+2\frac{(n-r)(N-(n-r))}{N^2}\left(\frac{1}{r}-\frac{1}{N}\right)\right]$$

$$<\left(\frac{1}{r}-\frac{1}{N}\right)\left[1-\frac{\left(N-(n-r)\right)^2}{N^2}\right]$$

elde edilir. Daha sonra gerekli sadeleştirmeler yapıldıktan sonra,

$$\rho^2\left[N^3+2N^2n-Nn^2-Nn(n-r)+n(n-r)r\right]<n(N-r)\left(2N-(n-r)\right)$$

ve buradan

$$\rho^2<\frac{n(N-r)\left[2N-(n-r)\right]}{n(n-r)r+N\left[N^2+n(2N-2n+r)\right]}$$

biçiminde bulunur.

EK 8 : T_{1u} ve T_{2m} TAHMİN EDİCİSİNİN VARYANS EŞİTLİKLERİNİN ELDE EDİLİŞİ

T_{1u} ve T_{2m} yansız tahmin edicilerdir ve *u* ve *m* büyüklüklü iki bağımsız örnekleme dayanmaktadır, dolayısıyla kovaryans terimi ihmal edilerek aşağıdaki eşitliğe ulaşmak mümkündür:

$$V(\hat{T}) = \varphi^2 V(T_{1u}) + (1-\varphi)^2 V(T_{2m}) .$$

$V(T_{1u})$ ve $V(T_{2m})$ ifadeleri aşağıdaki biçimde elde edilmektedir:

$$\begin{aligned}
V(T_{1u}) &= E\left[\bar{y}_u + B_{yz}\left(\bar{Z} - \bar{z}_u\right) - \bar{Y}\right]^2 \\
&= E\left[(\bar{y}_u - \bar{Y}) + B_{yz}\left(\bar{Z} - \bar{z}_u\right)\right]^2 \\
&= E\left[(\bar{y}_u - \bar{Y})^2 + B_{yz}^2\left(\bar{Z} - \bar{z}_u\right)^2 + 2B_{yz}(\bar{y}_u - \bar{Y})\left(\bar{Z} - \bar{z}_u\right)\right] \\
&= V(\bar{y}_u) + B_{yz}^2 V(\bar{z}_u) - 2B_{yz} Cov(\bar{y}_u, \bar{z}_u) \\
&= \frac{1}{u} S_y^2 + B_{yz}^2 \frac{1}{u} S_z^2 - 2B_{yz} \frac{1}{u} S_{yz} \\
&= \frac{1}{u} S_y^2 - \frac{1}{u} \frac{S_{yz}^2}{S_z^4} S_z^2 - 2\frac{1}{u} \frac{S_{yz}^2}{S_z^2} \\
&= \frac{1}{u}\left(1 - \frac{S_{yz}^2}{S_z^2 S_y^2}\right) S_y^2
\end{aligned}$$

$$\Rightarrow V(T_{1u}) = \frac{1}{u}\left(1 - \rho_{yz}^2\right) S_y^2$$

olarak bulunur.

$$T_{2m} = \bar{y}_m^* + \beta_{yx}\left(\bar{x}_n^* - \bar{x}_m^*\right)$$

$$\left(T_{2m} - \bar{Y}\right) = \left(\bar{y}_m^* - \bar{Y}\right) + B_{yx}\left(\bar{x}_n^* - \bar{X}\right) - B_{yx}\left(\bar{x}_m^* - \bar{X}\right)$$

her iki tarafın karesi alınarak beklenen değeri alındığında,

$$E\left(T_{2m} - \bar{Y}\right)^2 = V\left(\bar{y}_m^*\right) + B_{yx}^2 V\left(\bar{x}_n^*\right) + B_{yx}^2 V\left(\bar{x}_m^*\right) + 2B_{yx} Cov\left(\bar{y}_m^*, \bar{x}_n^*\right) - 2B_{yx} Cov\left(\bar{y}_m^*, \bar{x}_m^*\right) - 2B_{yx}^2 Cov\left(\bar{x}_n^*, \bar{x}_m^*\right)$$

$V(\bar{y}_m^*)$, $V(\bar{x}_n^*)$ ve $V(\bar{x}_m^*)$, $V(T_{1u})$'a benzer biçimde aşağıdaki gibi elde edilir:

$$V(\bar{y}_m^*) = \frac{1}{m}\left(1-\rho_{yz}^2\right)S_y^2$$

$$V(\bar{x}_n^*) = \frac{1}{n}\left(1-\rho_{xz}^2\right)S_x^2$$

$$V(\bar{x}_m^*) = \frac{1}{m}\left(1-\rho_{xz}^2\right)S_x^2.$$

$$Cov(\bar{y}_m^*, \bar{x}_m^*) = E\left(\bar{y}_m^* - \bar{Y}\right)\left(\bar{x}_m^* - \bar{X}\right)$$

$$= E\left(\bar{y}_m + B_{yz}\left(\bar{Z} - \bar{z}_m\right) - \bar{Y}\right)\left(\bar{x}_m + B_{xz}\left(\bar{Z} - \bar{z}_m\right) - \bar{X}\right)$$

$$= E\left(\left(\bar{y}_m - \bar{Y}\right) + B_{yz}\left(\bar{Z} - \bar{z}_m\right)\right)\left(\left(\bar{x}_m - \bar{X}\right) + B_{xz}\left(\bar{Z} - \bar{z}_m\right)\right)$$

$$= E\left[\left(\bar{y}_m - \bar{Y}\right)\left(\bar{x}_m - \bar{X}\right) + B_{xz}\left(\bar{y}_m - \bar{Y}\right)\underbrace{\left(\bar{Z} - \bar{z}_m\right)}_{-\left(\bar{z}_m - \bar{Z}\right)} + B_{yz}\left(\bar{Z} - \bar{z}_m\right)\left(\bar{x}_m - \bar{X}\right) + B_{yz}B_{xz}\left(\bar{Z} - \bar{z}_m\right)^2\right]$$

$$= Cov\left(\bar{y}_m, \bar{x}_m\right) - B_{xz}Cov\left(\bar{y}_m, \bar{z}_m\right) - B_{yz}Cov\left(\bar{x}_m, \bar{z}_m\right) + B_{yz}B_{xz}V\left(\bar{z}_m\right)$$

$$\Rightarrow Cov(\bar{y}_m^*, \bar{x}_m^*) = \frac{1}{m}S_{xy} - \frac{1}{m}B_{xz}S_{yz} - \frac{1}{m}B_{yz}S_{xz} + \frac{1}{m}B_{yz}B_{xz}S_z^2$$

$$Cov(\bar{y}_m^*, \bar{x}_n^*) = E\left(\bar{y}_m^* - \bar{Y}\right)\left(\bar{x}_n^* - \bar{X}\right)$$

$$= E\left(\bar{y}_m + B_{yz}\left(\bar{Z} - \bar{z}_m\right) - \bar{Y}\right)\left(\bar{x}_n + B_{xz}\left(\bar{Z} - \bar{z}_n\right) - \bar{X}\right)$$

$$= E\left(\left(\bar{y}_m - \bar{Y}\right) + B_{yz}\left(\bar{Z} - \bar{z}_m\right)\right)\left(\left(\bar{x}_n - \bar{X}\right) + B_{xz}\left(\bar{Z} - \bar{z}_n\right)\right)$$

$$= E\left[\left(\bar{y}_m - \bar{Y}\right)\left(\bar{x}_n - \bar{X}\right) + B_{xz}\left(\bar{y}_m - \bar{Y}\right)\left(\bar{Z} - \bar{z}_n\right) + B_{yz}\left(\bar{Z} - \bar{z}_m\right)\left(\bar{x}_n - \bar{X}\right) + B_{yz}B_{xz}\left(\bar{Z} - \bar{z}_m\right)\left(\bar{Z} - \bar{z}_n\right)\right]$$

$$= Cov\left(\bar{y}_m, \bar{x}_n\right) - B_{xz}Cov\left(\bar{y}_m, \bar{z}_n\right) - B_{yz}Cov\left(\bar{x}_n, \bar{z}_m\right) + B_{yz}B_{xz}Cov\left(\bar{z}_n, \bar{z}_m\right)$$

$$\Rightarrow Cov(\bar{y}_m^*, \bar{x}_n^*) = \frac{1}{n}S_{xy} - \frac{1}{n}B_{xz}S_{yz} - \frac{1}{n}B_{yz}S_{xz} + \frac{1}{n}B_{yz}B_{xz}S_z^2$$

$$Cov\left(\bar{x}_n^*, \bar{x}_m^*\right) = E\left[\left(\bar{x}_n^* - \bar{X}\right)\left(\bar{x}_m^* - \bar{X}\right)\right]$$

$$= E\left[\left(\left(\bar{x}_n - \bar{X}\right) + B_{xz}\left(\bar{Z} - \bar{z}_n\right)\right)\left(\left(\bar{x}_m - \bar{X}\right) + B_{xz}\left(\bar{Z} - \bar{z}_m\right)\right)\right]$$

$$= E\left[\left(\bar{x}_n - \bar{X}\right)\left(\bar{x}_m - \bar{X}\right) + B_{xz}\left(\bar{x}_n - \bar{X}\right)\left(\bar{Z} - \bar{z}_m\right) + B_{xz}\left(\bar{Z} - \bar{z}_n\right)\left(\bar{x}_m - \bar{X}\right) + B_{xz}^2\left(\bar{Z} - \bar{z}_n\right)\left(\bar{Z} - \bar{z}_m\right)\right]$$

$$= Cov(\bar{x}_n, \bar{x}_m) - B_{xz}Cov(\bar{x}_n, \bar{z}_m) - B_{xz}Cov(\bar{x}_m, \bar{z}_n) + B_{xz}^2 Cov(\bar{z}_n, \bar{z}_m)$$

$$\Rightarrow Cov(\bar{x}_n^*, \bar{x}_m^*) = \frac{1}{n}S_x^2 - 2\frac{1}{n}B_{xz}S_{xz} + \frac{1}{n}B_{xz}^2 S_z^2$$

Varyans ve kovaryans eşitlikleri yerine yazıldığında,

$$\begin{aligned} V(T_{2m}) = {} & \frac{1}{m}(1-\rho_{yz}^2)S_y^2 + B_{yx}^2\frac{1}{n}(1-\rho_{xz}^2)S_x^2 + B_{yx}^2\frac{1}{m}(1-\rho_{xz}^2)S_x^2 \\ & + 2B_{yx}\left[\frac{1}{n}S_{xy} - \frac{1}{n}B_{xz}S_{yz} - \frac{1}{n}B_{yz}S_{xz} + \frac{1}{n}B_{yz}B_{xz}S_z^2\right] \\ & - 2B_{yx}\left[\frac{1}{m}S_{xy} - \frac{1}{m}B_{xz}S_{yz} - \frac{1}{m}B_{yz}S_{xz} + \frac{1}{m}B_{yz}B_{xz}S_z^2\right], \\ & - 2B_{yx}^2\left[\frac{1}{n}S_x^2 - 2\frac{1}{n}B_{xz}S_{xz} + \frac{1}{n}B_{xz}^2S_z^2\right] \end{aligned}$$

$B_{yx} = \dfrac{S_{xy}}{S_x^2}$, $B_{yz} = \dfrac{S_{yz}}{S_z^2}$ ve $B_{xz} = \dfrac{S_{xz}}{S_z^2}$ ifadeleri yerine yazılarak gerekli sadeleştirmeler yapıldığında,

$$V(T_{2m}) = \frac{1}{m}(1-\rho_{yz}^2)S_y^2 + \left(\frac{1}{m}+\frac{1}{n}\right)(1-\rho_{xz}^2)\frac{S_{xy}^2}{S_x^2} - 2S_y^2(\rho_{xy}^2 - \rho_{yz}^2\rho_{xy})\left(\frac{1}{\mathrm{m}}-\frac{1}{n}\right) - 2S_y^2\rho_{xy}^2(1-\rho_{yz}^2)\left(\frac{1}{\mathrm{n}}\right)$$

olur. S_y^2 ortak parantezine alınırsa,

$$V(T_{2m}) = \left[\frac{1}{m}(1-\rho_{yz}^2) + \left(\frac{1}{m}+\frac{1}{n}\right)(1-\rho_{xz}^2)\rho_{xy}^2 - 2(\rho_{xy}^2 - \rho_{yz}^2\rho_{xy})\left(\frac{1}{\mathrm{m}}-\frac{1}{n}\right) - 2\rho_{xy}^2(1-\rho_{yz}^2)\left(\frac{1}{\mathrm{n}}\right)\right]S_y^2,$$

elde edilir. $\rho_{xz} = \rho_{yz}$ varsayımı altında gerekli sadeleştirmelerden sonra,

$$V(T_{2m}) = \left[\frac{1}{m}(1-\rho_{yz}^2) + \left(\frac{1}{m}-\frac{1}{n}\right)(2\rho_{yz}^2\rho_{yx} - \rho_{yx}^2(1+\rho_{yz}^2))\right]S_y^2$$

eşitliğine ulaşılır.