İstatistik Araştırma Dergisi Cilt 03 No 01 Sayfa 85-101 Nisan 2004

KARMAŞIK ÖRNEKLEME PLANLARINDA ÇEŞİTLİ VARYANS TAHMİN YÖNTEMLERİ VE UYGULAMA

Funda TERCAN* Hülya ÇINGI**

ÖZET

Bu çalışmada karmaşık örnekleme planlarında varyans tahmini için kullanılan yöntemler incelendi. İlk bölümde konuya giriş yapıldı. İkinci bölümde karmaşık örnekleme planlarında varyans tahmin yöntemlerinden olan Taylor serisi yöntemi, üçüncü bölümde rasgele grup yöntemlerinden bağımlı rasgele grup ile bağımsız rasgele grup yöntemleri, dördüncü bölümde tekrarlı yöntemlerden dengeli olarak tekrarlanan tekrarlı yöntem, jacknife yöntemi ve bootstrap yöntemi, beşinci bölümde ise genelleştirilmiş varyans fonksiyonu incelenerek kuramsal özellikleri, yöntemlerin uygulanış şekilleri, örneklem planındaki kısıtlamaları üzerinde duruldu. Altıncı bölümde açıklanan yöntemler kuramsal özellikleri, uygulanış şekilleri, sonuçlardaki doğruluk ve maliyetleri bakımından karşılaştırıldı. Yedinci bölümde ise bugüne dek karmaşık örnekleme planlarında varyans eşitlikleri incelenmemiş olan oransal, çarpımsal, regresyon tahmin edicileri için açıklanan yöntemler kullanılarak elde edilen varyans eşitlikleri verildi. Sekizinci bölümde açıklanan yöntemleri kullanarak Ankara ilinde bulunan 843 sanayi kuruluşundan tabakalı sistematik örnekleme ile elde edilen 163 birimlik örneklemden bazı özelliklerin tahminleri ve varyans tahminlerini elde etmek amacıyla yazılmış programdan elde edilen sonuçlardan bazı örnekler verildi ve elde edilen bu sonuçlar ile yöntemler karşılaştırıldı.

Anahtar Kelimeler: *Varyans tahmini, karmaşık örneklemler, yeniden (tekrarlı) örnekleme yöntemleri*

1.GİRİŞ

Son kırk yılda örnekleme araştırmalarının kuramı ve uygulamaları çok fazla genişlemiştir. Örnekleme araştırmalarının sayısı fazlalaştığı için verileri ve sonuçları analiz etmek için de birçok yönteme gerek duyulmaktadır. Bu araştırmalarda önemli olan tahmin edilmek istenen istatistiğin tahmini olduğu gibi bu tahminin ne kadar bir hataya sahip olduğudur. İşte bu hatanın ölçümü de varyansdır. Basit örnekleme planları için varyans tahmini genellikle bilinmesine karşın örnekleme planı karmaşık olduğunda bu tahminin ne olacağı birçok araştırmacı tarafından pek araştırılmamıştır. Bu konuda en çok yapılan hata örnekleme planı ne olursa olsun varyans tahmini için Basit Rasgele Örneklemenin varyans eşitliklerinin kullanılmasıdır.

* Hacettepe Üniversitesi, İstatistik Bölümü
** Hacettepe Üniversitesi, İstatistik Bölümü

Basit Rasgele Örnekleme, Tabakalı Örnekleme, Küme Örneklemesi gibi basit örnekleme planlarında varyans eşitlikleri bilinmekte ve çok kolay bir şekilde uygulanmaktadır. Ancak örnekleme planı karmaşıklaştıkça bu eşitliklerin elde edilişi güç olmakta hatta imkansızlaşmaktadır. Karmaşık örnekleme planları; tabakalama, kümeleme, rasgele örnekleme, çok aşamalı örnekleme, ikili örnekleme, eşit olmayan olasılıkla seçim gibi birçok örnekleme planlarından bazılarının bir arada yalnızca bir örnekleme planına uygulanmasıyla elde edilir. Örnekleme planı karmaşıklaştıkça tahminin varyans eşitliğinin elde edilişi de güçleşecektir. (Wolter, 1985; Sarndal,Swensson and Wretman, 1991; Toprak, 1996; Lohr, 1999)

Bu çalışmada, yukarıda bahsedilen karmaşık örnekleme planlarında varyans tahmin yöntemlerinden olan Taylor serisi yöntemi (TSY), rasgele grup yöntemi (RGY), tekrarlı örnekleme yöntemlerinden dengeli olarak tekrarlanan tekrarlı yöntemi (DTT), jacknife yöntemi (J) ve bootstrap yöntemi (B) son olarak da genelleştirilmiş varyans fonksiyonu (GVF) yöntemleri açıklanacaktır. Ayrıca bugüne dek incelenmemiş olan oransal, çarpımsal ve regresyon tahmin edicilerinin varyans eşitlikleri bu yöntemler yardımıyla elde edilecek; yöntemleri uygulamak için yazılan bilgisayar programı yardımıyla bir uygulama yapılacaktır.

2. TAYLOR SERİSİ YÖNTEMİ

Basit ve karmaşık örnekleme tasarımlarının her ikisinde de doğrusal olmayan istatistiklerin varyanslarının tahmin edilmesi kaçınılmaz ve istenilen bir durumdur. Doğrusal olmayan istatistiklerin varyansının tahmininin en kullanışlı yollarından biri tahmini doğrusal bir fonksiyona yakınlaştırmaktır. Daha sonra bu yaklaştırmadan varyans eşitliğine gidilir. İşte çok eski zamanlardan beri çok iyi bir şekilde bilinen ve uygulanan Taylor serisi yöntemi sayesinde doğrusal olmayan istatistikler doğrusal şekle dönüştürülerek varyansları elde edilir. Bu yönteme ayrıca delta yöntemi de denilmektedir.

Bu yöntemde tahmin edicinin formu, tahmindeki raslantı değişkeni sayısı ve tahmin edicinin karmaşıklığı ya da örnekleme tasarımının karmaşıklığı üzerinde kısıtlama yoktur. h gibi varyansını elde etmek istediğimiz fonksiyonun $\{X_1, X_2, ..., X_k\}$ gibi değişkenlerden oluştuğu düşünülürse Taylor serisi açılımı şöyledir;

$$h(\hat{X}_1, \hat{X}_2, ..., \hat{X}_k) \approx h(X_1, X_2, ..., X_k) + \sum_{i=1}^{k} d_j(\hat{X}_j - X_j) \quad (1)$$

Şimdiye kadar yapılan büyük araştırmalarda 1. dereceden sonraki açılımın önemsiz olduğu görülmüştür. Bu eşitlikteki d_j, h fonksiyonunun her değişkeninin beklenen değer noktası için (X=x) elde edilen kısmi türevlerden oluşan $1 \times k$ boyutlu bir vektördür.

$$d_j = \frac{\partial h(a_1, a_2, ..., a_k)}{\partial a_j} \Bigg|_{X_1, X_2, ... X_k}$$

Daha sonra varyans eşitliğine ulaşmak için $h(X_1, X_2,..., X_k)$ ifadesi sol tarafa atılarak her iki tarafın beklenen değeri alınır.

$$Var\left[h(\hat{X}_1, \hat{X}_2,..., \hat{X}_k)\right] \approx E\left[\sum_{i=1}^{k} d_j(\hat{X}_j - X_j)\right]^2 \qquad (2)$$

Dolayısıyla varyans eşitliğimizi genelleştirirsek

$$Var\left[h(\hat{X}_1, \hat{X}_2,..., \hat{X}_k)\right] \approx d\Sigma d' \qquad (3)$$

Burada d $1 \times k$ boyutlu kısmi türevlerden oluşan vektör, Σ ise $k \times k$ boyutlu varyans kovaryans matrisidir.

Sonuç olarak Taylor serisi yöntemi uzun yıllardır kullanıldığı için teorisi iyi gelişmiş bir yöntemdir. Örnekleme planı üzerinde hiçbir kısıtlaması yoktur. Daha sonraki bölümlerde anlatılacak olan dengeli olarak tekrarlanan tekrarlı yöntem gibi her tabakasında 2 tane ön örneklem birimi içerme şeklinde bir kısıtlaması yoktur. Bütün bu avantajlarına rağmen bu yöntemin dezavantajları da mevcuttur. Hesaplamalar karmaşık olabilir. Ayrıca bazı istatistiklerin kısmi türevlerinin elde edilişi de zor olabilir. Son olarak bu yöntemin uygulanabilmesi için örneklemimizin yeterince büyük olması gerekir. Ayrıca çarpık bir dağılıma sahip kitlelerde doğrusallaştırma yöntemi doğru sonuçlar vermemektedir.(Woddruff, 1971; Kish and Frankel ,1974; Woddruff and Causey, 1976; Wolter, 1985; Lohr, 1999)

3. RASGELE GRUP YÖNTEMİ (RANDOM GROUP METHOD)

Burada bağımsız ve bağımlı olmak üzere iki tane rasgele grup yöntemi vardır.

3.1. Bağımsız Rasgele Grup Yöntemi (Independent Random Group Method)

Bu yöntemin bir diğer adı araştırma tasarımını tekrarlama (replicating the survey design)'dır. Bu yöntemle ortalama, toplam gibi doğrusal, iki değişkenin birbirine oranı, regresyon katsayısı ve korelasyon katsayısı gibi doğrusal olmayan istatistiklerin varyansları elde edilebilir. Yöntemin uygulanışı kısaca şöyledir.

1. S gibi bir kitleden örnekleme planına göre S_1 gibi bir örneklem seçilir. Buradaki örnekleme planı üzerinde hiçbir kısıtlama yoktur.
2. Birinci örneklemin seçimini takiben S_1 örneklemi yerine konulur ve aynı örnekleme planı ile S_2 örneklemi seçilir.
3. Bu süreç k seçilen örneklem sayısını göstermek üzere $k \geq 2$ olacak şekilde $S_1, S_2,..., S_k$ elde edilinceye kadar uygulanır.

Burada θ tahmin edilmek istenen parametre , $\hat{\theta}_\alpha, \alpha$. tekrardan elde edilen tahmin değeri olmak üzere ortak tahmin ediciyi elde etmek için tüm tekrarlardan elde edilen tahminlerin ortalaması alınır. Buna göre;

$$\hat{\bar{\theta}} = \frac{\sum_{\alpha=1}^{k} \hat{\theta}_{\alpha}}{k} \tag{4}$$

Buna göre varyans eşitliğimiz;

$$\hat{V}_1(\hat{\bar{\theta}}) = \frac{\sum_{\alpha=1}^{k} (\hat{\theta}_{\alpha} - \hat{\bar{\theta}})^2}{k(k-1)} \tag{5}$$

Bağımsız rasgele grup yaklaşımı altında, k tane örneklemin bağımsız seçildiği ve $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$' nın bağımsız ve aynı dağılımlı raslantı değişkenleri olduğu düşünüldüğünde $\hat{V}_1(\hat{\bar{\theta}})$ yansızdır. Burada $\hat{\bar{\theta}}$ her tekrardan elde edilen sonucun ortalaması $\hat{\theta}$ ise toplu örneklemden elde edilen tahmin değeridir. Doğrusal bir tahmin edici için $\hat{\bar{\theta}} = \hat{\theta}$' dir. Ancak doğrusal olmayan bir tahmin edici için $\hat{\bar{\theta}} \neq \hat{\theta}$' dir. Buna göre doğrusal olmayan bir tahmin edici için (5) 'deki varyans eşitliğine seçenek olarak aşağıdaki varyans eşitliği de kullanılabilir;

$$\hat{V}_2(\hat{\bar{\theta}}) = \frac{\sum_{\alpha=1}^{k} (\hat{\theta}_{\alpha} - \hat{\theta})^2}{k(k-1)} \tag{6}$$

Doğrusal bir tahmin edici de $\hat{\bar{\theta}} = \hat{\theta}$ olduğu için $\hat{V}_1(\hat{\bar{\theta}}) = \hat{V}_2(\hat{\bar{\theta}})$olacaktır. Bu iki varyans eşitliği arasındaki ilişkiye bakarsak;

$$\sum_{\alpha=1}^{k} (\hat{\theta}_{\alpha} - \hat{\theta})^2 = \sum_{\alpha=}^{k} (\hat{\theta}_{\alpha} - \hat{\bar{\theta}})^2 + k(\hat{\bar{\theta}} - \hat{\theta})^2 \tag{7}$$

(7) eşitliğine göre $\hat{V}_1(\hat{\bar{\theta}}) \leq \hat{V}_2(\hat{\bar{\theta}})$'dir. $\hat{V}_2(\hat{\bar{\theta}})$, $\hat{V}_1(\hat{\bar{\theta}})$'dan daha büyük olmasına rağmen genellikle tercih edilen varyans eşitliğidir. Birçok karmaşık örnekleme planında $(\hat{\bar{\theta}} - \hat{\theta})^2$ değeri fazla önemli olmayabilir. Dolayısıyla $\hat{V}_2(\hat{\bar{\theta}})$ ile $\hat{V}_1(\hat{\bar{\theta}})$ değerleri birbirine çok yakındır.

Bu yöntemle varyans hesabı oldukça kolaydır. İstatistiğin karmaşıklığı ya da örnekleme planının karmaşıklığı bu hesabı fazla etkilemez. Varyans tahmini için özel yazılımlara ihtiyaç yoktur. Uygun bir varyans tahmini için bağımız rasgele grup sayısının fazla olması gerekir. Ancak uygulama da bu grup sayısı fazla olmamaktadır. Mahalanobis en az 4, Deming ise en az 10 rasgele grubun kullanılması gerektiğini önermiştir.(Wolter, 1985; Sarndal,Swensson and Wretman, 1991; Verma, 1992; Lohr, 1999)

3.2. Bağımlı Rasgele Grup Yöntemi (Dependent Random Group Method)

Bu yöntemin bir diğer adı da örneklemi rasgele gruplara bölme (dividing the sample into random groups)'dir. Bu yöntemde kitleden seçilen geniş ana örneklemimiz alt rasgele gruplara bölünür. Bölünen her bir rasgele grubun seçim planının ana örneklemin seçim planıyla aynı olmasına dikkat edilir. Rasgele gruplar yerine konulmadan oluşturulduğu için tahminler az da olsa bir yana sahiptir.

Burada da bağımsız rasgele grup yönteminde olduğu gibi θ tahmin edilmek istenen parametre , $\hat{\theta}_{\alpha}$,α. rasgele gruptan elde edilen tahmin değeri olmak üzere ortak tahmin ediciyi elde etmek için tüm gruplardan elde edilen tahminlerin ortalaması alınarak elde edilir. Varyans eşitlikleri içinse yine bağımlı rasgele grup yönteminde kullanılan varyans eşitlikleri kullanılır. Burada da doğrusal olmayan istatistikler için iki seçenek varyans eşitliği kullanılabilir.Burada da doğrusal istatistikler için her iki varyans eşitliği birbirine eşittir.

Bu yöntemle de varyans hesabı oldukça kolaydır. Yine aynı şekilde istatistiğin karmaşıklığı ya da örnekleme planının karmaşıklığı bu hesabı fazla etkilemez. Burada da varyans tahmini için özel yazılımlara ihtiyaç yoktur. Ancak bazı örnekleme planlarında rasgele grupları ana örneklemin planında oluşturmak zor olabilir. Grup sayısının fazla olabilmesi için de örneklemimizin geniş olmasına ihtiyaç duyulur. (Wolter, 1985; Sarndal,Swensson and Wretman, 1991; Verma, 1992; Lohr, 1999)

4. TEKRARLI YÖNTEMLER (RESAMPLING METHODS)

Bu yöntemler dengeli olarak tekrarlanan tekrarlı yöntem (balanced repeated replication method), jacknife yöntemi (jacknife method) ve bootstrap yöntem (bootstrap method) ' idir.

4.1. Dengeli Olarak Tekrarlanan Tekrarlı Yöntem (Balanced Repeated Replication Method)

Bu yöntemin temelinde her tabakasında iki tane ön örneklem birimi (ÖÖB) ya da birim bulunan ya da seçilen planlar için kullanılması yer almaktadır. Bu yöntemin uygulandığı planlar çok fazla tabakaya sahiptir. Yöntemin uygulanışının temelinde ise her tabakada yer alan iki ön örneklem biriminden (ÖÖB) birinin seçilerek yarı örneklemlerin oluşturulması yatar. Her tabakasında iki tane ön örneklem birimi içeren planlarda bağımlı rasgele grup yöntemi kullanılarak her tabakadan bir gruba bir ön örneklem birimi alınarak iki tane rasgele grup oluşturularak tahminler yapılabilir. Ancak bu yöntemle rasgele grup sayımız iki çok az olduğu için tahminlerimize yan gelecektir. İşte böyle bir durumda önerilen yöntem ise dengeli olarak tekrarlanan tekrarlı yöntemdir.

Her bir tabakada yer alan birimler ya da ön örneklem birimlerini y_{h1} ve y_{h2} olarak adlandıralım. Her bir tabakadaki bu iki birimden birisi seçilerek yarı örneklemler oluşturulur. Burada L tabaka sayısı olmak üzere 2^L tane mümkün olası yarı örneklemimiz vardır. Bu iki birimden hangisinin yarı örneklemimize seçileceği McCarthy (1966) tarafından bir denge kuralı şeklinde açıklanmıştır. Bu kurala göre yarı örneklemlerimizi oluşturmak için bir δ vektörü tanımlanır. Bu vektör tanımı şöyledir;

$$\delta_h^{(\alpha)} = \begin{cases} 1 & \text{eğer h. tabakadaki 1. birim } \alpha\text{. yarı örneklemde ise} \\ -1 & \text{eğer h. tabakadaki 2. birim } \alpha\text{. yarı örneklemde ise} \end{cases}$$

Burada da α. yarı örneklemdeki tahmin değerimiz, bu örneklemde yer alan birimlerin ağırlıkları ile çarpılarak toplamın alınmasıyla elde edilir. δ vektörüne göre tahmin eşitliğimiz;

$$\bar{y}_{st,\alpha} = \sum_{h=1}^{L} W_h(\delta_h^{(\alpha)} y_{h1} - \delta_h^{(\alpha)} y_{h2}) \tag{8}$$

olur. Yarı örneklemlerin k kez tekrar edildiğini, yani her tabakadaki iki birimden birini seçerek oluşturulan yarı örneklemlerin k tane olduğunu düşünelim. Böylece $\delta_h^{(\alpha)}$ vektörü her bir tekrar için elde edileceğinden matris halini alacaktır. Bu matris dik bir matristir. İşte sözü edilen denge kuralına göre $\sum_{h\neq h'} \delta_h^{(\alpha)}\delta_{h'}^{(\alpha)} = 0$ olmalıdır. Yani bu matristeki herhangi iki kolonunun çarpımlar toplamı sıfır olmalıdır. Buna " tam denge" kuralı denir.

Burada tabaka sayısı L çok büyük olduğunda 2^L olası yarı örneklem sayısı da çok fazla olacaktır. Dolayısıyla bu durumda çok karışık işlemler yer alacak ve hesaplama maliyeti çok fazla olacaktır. Bunun için tekrar sayısı olan yarı örneklem sayısına dikkat edilmesi gerekir. Yarı örneklemiz 2^L tane olan olası yarı örneklem setinden seçilmelidir. Yarı örneklem sayısı olan k, L tabaka sayısından fazla olup 4' ün katı şeklinde olmalıdır. Dolayısıyla k, $L+1 \leq k \leq L+4$ olacak şekilde seçilebilir. Örneğin 47 tane tabakamız varsa 48 tekrar yeterli olacaktır. 48'den fazla tekrar sayısı hesaplamaları karıştıracak ve maliyeti artıracaktır.

Buradaki varyans eşitliğimiz ise;

$$\hat{V}_{DTT}(\bar{y}_{st}) = \frac{1}{k}\left[\sum_{\alpha=1}^{k}(\bar{y}_{st,\alpha} - \bar{y}_{st})^2\right] \tag{9}$$

Doğrusal olmayan istatistiklerin varyansının tahmininde (9) eşitliğine alternatif varyans eşitlikleri yer almaktadır.

$\hat{\theta}_\alpha$: α. yarı örneklemden elde edilen tahmin,

$\hat{\theta}_\alpha^c$: α. yarı örneklemin tamamlayanından yani bu yarı örneklemde yer almayan birimlerden elde edilen tahmin olmak üzere varyans eşitliklerimiz;

$$\hat{V}_{DTT}^c(\hat{\theta}) = \frac{1}{k}\sum_{\alpha=1}^{k}(\hat{\theta}_\alpha^c - \hat{\theta})^2 \tag{10}$$

$$\hat{\bar{V}}_{DTT}(\hat{\theta}) = \frac{1}{2}\left[V_{DTT}(\hat{\theta}) + V_{DTT}^c(\hat{\theta})\right] \tag{11}$$

$$\hat{V}_{DTT}^{+}(\hat{\theta}) = \frac{1}{4k}\sum_{\alpha=1}^{k}(\hat{\theta}_\alpha - \hat{\theta}_\alpha^c)^2 \tag{12}$$

Bu yöntemin her tabakasında iki birim veya iki ön örnekleme birimi olduğunda uygulanabildiğini söylemiştik. Ama her tabakasında iki birim bulunmayan planlarda iki yönteme başvurulur. İlki her tabakadaki birim sayısı m_h bir tamsayı olacak şekilde $n_h = 2m_h$ olacak biçimde iki gruba bölünür. Bu grupların ortalaması alınarak iki tane birim oluşturulur. Daha sonra bu iki birime daha önce anlatılan biçimde denge kuralı kullanılarak yarı örneklemler oluşturulur. Yani yöntem bu iki birim üzerine uygulanır ve tahminlere ulaşılır. İkincisi ise h. tabakada her birinin genişliği iki olan m_h tane yapay tabakaya bölünür. $L = \sum_{h=1}^{L} m_h$ tane yapay tabaka elde edilir. δ vektörü tanımına göre birim değil tabakalar seçilerek yarı örneklemler oluşturulur.

Burada düşünülmesi gereken bir diğer husus da tekrar sayısının çok fazla olduğu durumlarda maliyeti ve hesaplamaları azaltmak için kısmi dengeye başvurulmasıdır. Burada L tabakamız G gruba bölünür ve ilk G gruptaki tabakaya tam denge kuralı uygulanır.

Dengeli olarak tekrarlanan tekrarlı yöntem bu gruptaki jacknife ve bootstrap gibi diğer yöntemlere göre daha az hesaplama gerektirmektedir. Ve tabaka sayısı fazla olduğunda çok fazla tekrar yerine kısmi denge tasarımı uygulanarak daha az tekrarla varyans elde edilebilir. Her tabakasında iki ön örneklem birimi bulunduran planlara uygulanması negatif yönüdür. Bu negatif özelliğinin yanı sıra dengeli tekrarları oluştururken oluşan teknik karmaşıklıklar ve farklı örnekleme planlardaki esneksizliği de negatif özellikleri olarak sayılabilir. Yarı örneklem sayısı (tekrar) olan k' nın çok büyük olması hesaplamaları zorlaştıracak ve maliyeti artıracaktır. Bunun için tabaka sayısına göre tekrar sayısına karar verilmelidir. Eğer tabakalarımız çok fazla ise mantıki yönden bazı tabakaları birleştirebiliriz. (Kish and Frankel, 1970; Lee, 1972; Rao and Krewski, 1981; Rao and Wu, 1985; Wolter, 1985; Sarndal, Swensson and Wretman, 1991; Raj, 1998; Lohr, 1999)

4.2. Jacknife Yöntemi (Jacknife Method)

Çok kullanışlı bir yöntem olan jacknife yöntemi ile ana örneklemin birçok alt örneklemindeki tahminleri ve bu tahminler içindeki değişikliklerden yararlanarak ana örneklemin tahminleri elde edilir. Yöntemin uygulanışı ise şöyledir; S gibi örneklemden n birimin seçildiğini düşünelim. Bu yöntemde n=mk , k grup sayısını, m her bir gruptaki birim sayısını göstermek üzere n birimimizin her birinin genişliği m olacak şekilde k gruba bölünür. Her tekrarda bir grubun silinip diğer gruplar üzerinden işlemlerin yapıldığı bir yöntemdir.

$\hat{\theta}$ tahmin edilmek istenen tahmindir ve θ formunda olup tüm örneklemden elde edilsin. $\hat{\theta}_{(\alpha)}$ ise α. grup silindikten sonra m(k-1) birimden elde edilen ve yine θ formunda bir tahmin edici olsun.

Pseudo değeri ismi verilen tahmin edici tanımlayalım ;

$$\hat{\theta}_{\alpha} = k\hat{\theta} - (k-1)\hat{\theta}_{(\alpha)} \tag{13}$$

Pseudo değerlerinin ortalaması alınırsa;

$$\hat{\bar{\theta}} = \frac{1}{k}\sum_{\alpha=1}^{k}\hat{\theta}_{\alpha} \tag{14}$$

Buradan varyans eşitliklerine geçilirse;

$$\hat{V}_{JK1}(\hat{\theta}) = \frac{1}{k(k-1)}\sum_{\alpha=1}^{k}(\hat{\theta}_\alpha - \hat{\bar{\theta}})^2 \qquad (15)$$

$$\hat{V}_{JK2}(\hat{\theta}) = \frac{1}{k(k-1)}\sum_{\alpha=1}^{k}(\hat{\theta}_\alpha - \hat{\theta})^2 \qquad (16)$$

Varyans tahminimizde $\hat{\theta}_\alpha$ yerine $\hat{\theta}_{(\alpha)}$ yani α. grup silindikten sonra m(k-1) birim üzerinden elde edilen tahmini kullanalım. Burada $\hat{\theta}_{(.)}$ değeri tüm tekrarlardan elde edilen $\hat{\theta}_{(\alpha)}$ tahminlerinin ortalaması olmak üzere;

$$\hat{V}_{JK3}(\hat{\theta}) = \frac{k-1}{k}\sum_{\alpha=1}^{k}(\hat{\theta}_{(\alpha)} - \hat{\theta}_{(.)})^2 \qquad (17)$$

$\hat{V}_{JK1}(\hat{\theta})$ ve $\hat{V}_{JK2}(\hat{\theta})$ arasındaki ilişkiye bakalım;

$$\hat{V}_{JK2}(\hat{\theta}) = = \hat{V}_{JK1}(\hat{\theta}) + \frac{k}{k(k-1)}(\hat{\bar{\theta}} - \hat{\theta})^2 \qquad (18)$$

Eşitliğin sağ tarafındaki son değer daima pozitiftir. Dolayısıyla $\hat{V}_{JK2}(\hat{\theta}) \geq \hat{V}_{JK1}(\hat{\theta})$ olacaktır. Ancak bu eşitsizlik doğrusal olmayan istatistikler için geçerlidir. Doğrusal olan istatistiklerde $\hat{\bar{\theta}} = \hat{\theta}$ olduğundan $\hat{V}_{JK2}(\hat{\theta}) = \hat{V}_{JK1}(\hat{\theta})$ olacaktır.

Burada k grup sayısının belirlerken hem hesaplama maliyetini hem de sonuçların doğruluğunu birarada düşünerek uygun grup sayısına karar verilir.

Bu yöntemi tabakalı örneklemeye uygularken her tekrarda kaç birim sileceğimize karar vermemiz gerekir. Bunun için iki durum söz konusudur. Ya her tekrarda bir tabakadan bir birim silinir dolayısıyla her tekrarda bir birim silinmiş olur. Ya da her tabakadaki n_h birimin her birinde m_h birim bulunan rasgele gruplara ayrılır ve her tekrarda bir rasgele grup silinir dolayısıyla her tekrarda m_h birim silinmiş olur. $\hat{\theta}_{(hi)}$ h. tabakada i. birim silindikten sonra veya i. m_h birim silindikten elde edilen tahmin olsun. Buna göre pseudo değerimiz;

$$\hat{\theta}_{hi} = n_h\hat{\theta} - (n_h - 1)\hat{\theta}_{(hi)} \text{ ya da} \qquad (19)$$

$$\hat{\theta}_{hi} = (LW_h + 1)\hat{\theta} - LW_h\hat{\theta}_{(hi)} \qquad (20)$$

Buradaki ağırlık değişkenleri ise örneklemenin yerine konularak ya da konulmadan yapılmasına göre değişmektedir. Şöyle ki;

$$W_h = (n_h - 1)(1 - \frac{n_h}{N_h}) \text{ Yerine konulmadan yapılan örnekleme için} \qquad (21)$$

$$W_h = (n_h - 1) \text{ Yerine konularak yapılan örnekleme için} \qquad (22)$$

Burada değişik ortalama tahminleri yer almaktadır. Bunlar;

$$\hat{\theta}_{(h.)} = \sum_{i=1}^{n_h}\hat{\theta}_{(hi)} / n_h \qquad (23)$$

$$\hat{\theta}_{(...)} = \sum_{h=1}^{L}\sum_{i=1}^{n_h}\hat{\theta}_{(hi)} / n \quad (24)$$

$$\bar{\theta}_{(...)} = \sum_{h=1}^{L}\hat{\theta}_{(h.)} / L \quad (25)$$

Bu ortalamalar kullanılarak varyans eşitliklerine geçilirse;

$$\hat{V}_{JK}^{1}(\hat{\theta}) = \sum_{h=1}^{L}\frac{W_h}{n_h}\sum_{i=1}^{n_h}(\hat{\theta}_{(hi)} - \hat{\theta}_{(h.)})^2 \quad (26)$$

$$\hat{V}_{JK}^{2}(\hat{\theta}) = \sum_{h=1}^{L}\frac{W_h}{n_h}\sum_{i=1}^{n_h}(\hat{\theta}_{(hi)} - \hat{\theta}_{(...)})^2 \quad (27)$$

$$\hat{V}_{JK}^{3}(\hat{\theta}) = \sum_{h=1}^{L}\frac{W_h}{n_h}\sum_{i=1}^{n_h}(\hat{\theta}_{(hi)} - \bar{\theta}_{(...)})^2 \quad (28)$$

$$\hat{V}_{JK}^{4}(\hat{\theta}) = \sum_{h=1}^{L}\frac{W_h}{n_h}\sum_{i=1}^{n_h}(\hat{\theta}_{(hi)} - \hat{\theta})^2 \quad (29)$$

Doğrusal istatistikler için bu varyans eşitliklerinin hepsi birbirine eşittir. Eğer tekrar sayısı 100 ya da 200' den az ise bootstrap yöntemine göre hesaplamaları daha kolaydır. Tabaka genişliği ikiden fazla olan örnekleme planlarına da uygulanabildiği için dengeli olarak tekrarlanan tekrarlı yöntemin uygulanmadığı planlarda da uygulanabilir. Bu yöntemin negatif özelliği ise örneğin ortanca gibi "smooth" düzgün olmayan istatistiklerde başarısızdır. (Miller, 1974; Krewski and Rao, 1981; Rao and Wu, 1985; Wolter, 1985; Hedayat and Sinha, 1991; Sarndal, Swensson and Wretman, 1991; Verma, 1992; Efron and Tibshirani, 1993; Raj, 1998; Lohr, 1999)

4.3. Bootstrap Yöntemi (Bootstrap)

Bootstrap yönteminin ilk çıkış yılları diğer yöntemler kadar eski değildir. ilk kez Efron tarafından 1979 yılında bilgisayar tabanlı bir yöntem olarak tanıtılmış ve tahmin edicinin standart hata, güven aralığı ve yan tahmininde kullanılmıştır. Bootstrap, varyans tahmini için hiçbir teorik hesaplama gerektirmez. İlk çıkış noktası çok yakın olduğundan dolayı bu yöntem henüz tüm ayrıntıları ile incelenmemiş ve büyük örnekleme araştırmalarına uygulanmamıştır. Bootstrap yönteminin temelinde orijinal örneklemden yerine konularak elde edilen her bağımsız yeniden örneklemlerden tahminler elde edilmesi yatar. Yerine konmadan ve bağımsızlığın olmadığı durumlarda ne olacağı hala cevaplanmamış sorulardandır.

Burada yöntemin uygulanışını anlatırsak;

1. U kitlemizin dağılımı F olsun. Bu kitleden n birimli S örneklemi seçilir.
2. $\hat{F}$ dağılımlı n birimli S örneklemi kitleymiş gibi düşünülerek yine n birimli ve yerine konularak örneklemler seçilir.
3. 2. şıktaki durum B tekrar sayısı olmak üzere B kez tekrarlanır.

Burada n birimli S örnekleminden yine n birimli örneklemler yerine konularak seçilerek Bootstrap tekrarları oluşturulur. Burada seçim yerine konularak yapıldığı için seçilen bir birim tekrar seçilebilir veya bazı birimler hiç seçilmeyebilir. Dolayısıyla bu değişkenlikten yararlanılarak varyans elde edilir.

θ tahmin edilmek istenen tahmin, $\hat{\theta}$ tüm örneklemden elde edilen tahmin, $\hat{\theta}^{*}(b)$ ise b. bootstrap tekrarından (b=1,2,...,B) elde edilen tahmin olsun. Tüm tekrarlardan elde edilen tahminlerin ortalaması;

$$\hat{\theta}^{*}(.)=\sum_{b=1}^{B}\hat{\theta}^{*}(b)/B \qquad (30)$$

Varyans eşitliğimiz ise;

$$\hat{V}_B(\hat{\theta}^{*})=\frac{1}{B-1}\sum_{b=1}^{B}\left[\hat{\theta}^{*}(b)-\hat{\theta}^{*}(.)\right]^2 \qquad (31)$$

Bu varyans tahmin edicisi $B \to \infty$ iken yani bootstrap tekrar sayısı sonsuza yaklaştıkça bootstrap varyans tahmini de kitlenin varyans tahminine yaklaşmaktadır. Yani bootstrap varyans tahmini çok fazla bir yana sahip değildir.

Bootstrap yönteminin tabakalı örneklemeye uygulanışında bazı değişiklikler vardır. Efron(1979) tabakalı örneklemede karmaşıklığı önlemek için her tabakadan n_h birim yerine n_h-1 birim seçilmesi gerektiğini önermiştir.

Yan miktarı en az olan bir varyans tahminini elde etmek için gerek duyulan B tekrar sayısı kullanılan bilgisayar alt yapısına ve tahmin edilmek istenen parametreye bağlı olarak değişecektir. Buna göre Efron (1979) varyans tahmini için en az 100 güven aralığı için ise en az 1000 tekrara ihtiyaç olduğunu söylemiştir.

Bootstrap yöntemi varyans tahmininin yanında güven aralıklarının tahmininde de kullanılmaktadır. Bu yöntem kullanılarak üç çeşit güven aralığı elde etme yöntemi vardır. Bootstrap-t güven aralığı ile her bootstrap tekrarı için bootstrap-t değerleri elde edilerek t tablosu gibi bir tablo oluşturulur. Ve bu tablo değerleri kullanılarak güven aralıkları elde edilir. Diğer güven aralıkları yöntemi olan yüzdelik güven aralığı ve BC_α güven aralığı yöntemleri ile her tekrardan elde edilen $\hat{\theta}^{*}(b)$ değerleri sıralanarak değişik hesaplama yöntemleri ile güven aralıkları elde edilir.

Bootstap yöntemi varyans tahmini ve güven aralıkları için kullanılan bir yöntemdir. Bootstrap örneklemleri kitledeki birimlerin dağılımına bağlıdır. Yerine konmadan örneklem için*ne olacağı daha bulunamamıştır. Bootstrap yöntemi diğer tekrarlı yöntemlere göre çok daha yeni bir yöntemdir. Dolayısıyla çok fazla büyük araştırmalarda kullanılmamaktadır. Ancak fazla teorik hesaplamalar gerektirmediğinden ve güçlü bir bilgisayar alt yapısı gerektirmesinden dolayı araştırmacıların ilgisini çekmeye başlamıştır. Bootstrap yönteminin dengeli olarak tekrarlanan tekrarlı yönteme göre en önemli avantajı tabaka genişliğinin jacknife yönteminde olduğu gibi her keyfi durum için kolayca uygulanabilir olmasıdır. Bootstrap yöntemi jacknife yönteminin tersine düzgün ve düzgün olmayan istatistikler için de kullanılabilirdir. Ayrıca bootstrap yöntemi güven aralıklarını elde etmek için direk olarak kullanılabilir. Bütün bu üstünlüklerine rağmen bootstrap yöntemi dengeli olarak tekrarlanan tekrarlı yöntem ve jacknife yöntemine göre çok fazla hesap gerektirmektedir. Güçlü bir bilgisayar alt yapısına ihtiyacı vardır. Dengeli olarak tekrarlanan tekrarlı yöntem ve jacknife yönteminin tersine her türlü karmaşık örnekleme planındaki teorik çalışmaları yapılmamıştır. (Efron, 1979; Chao, 1985; Rao and Wu, 1988; Hedayat and Sinha, 1991; Sarndal, Swensson and Wretman, 1991;Efron and Tibshirani, 1993; Rao, 1997; Chernick, 1999; Lohr, 1999)

5. GENELLEŞTİRİLMİŞ VARYANS FONKSİYONU (GENERALIZED VARIANCE FUNCTION)

Genelleştirilmiş varyans fonksiyonu (Generalized variance function) varyans tahmini için, tahmin ediciler ile onların varyansı ya da göreli varyansı arasında kurulan bir matematiksel model olarak tanımlanır. Bu modelin parametreleri geçmişteki verilerden ya da araştırmanın küçük bir kısmından tahmin edilir. Genelleştirilmiş varyans fonksiyonu (GVF) genellikle çok geniş çapta ve çok sık aralıklarla yapılan ve çok fazla istatistiği tahmin edilmek istenen dolayısıyla çok fazla raporlar oluşturulan araştırmalarda kullanılır.

Genelleştirilmiş varyans fonksiyonu şu üç aşamada uygulanır;

1. Adım: Araştırmamızda tahminini elde etmek istediğimiz değişkenleri demografik özelliklerine göre gruplandırılır. Genelleştirilmiş varyans fonksiyonunun başarısı bu istatistikleri ne kadar başarılı bir şekilde gruplandırdığımıza bağlıdır.

2. Adım: Araştırmanın küçük bir kısmından ya da geçmişte elde edilen verilerden tahmin ediciler ve onların varyansları ya da göreli varyansları elde edilir. Bu tahminleri elde etmek için tekrarlı örnekleme yöntemleri, doğrusallaştırma ya da diğer yöntemler kullanılır. $\hat{X}$ tahmin edilmek istenen tahmin edici olmak üzere bu tahmin edicinin varyansı ;

$$\sigma^2 = Var(\hat{X}) \tag{32}$$

Göreli varyansımız ise tahmin edicimizin varyansının tahminin kitle değerinin karesine bölümünden elde edilir;

$$V^2 = \frac{Var(\hat{X})}{X^2} \tag{33}$$

3. Adım: Bu son aşamada V^2 ile $\hat{X}$ arasında matematiksel bir model tanımlanır. Bu modelin bilinmeyen parametreleri genellikle en küçük kareler yöntemi kullanılarak tahmin edilir. Şimdiye kadar yapılan büyük araştırmalarda $V^2 = \alpha + \frac{\beta}{X}$ modeli kullanılmıştır. Buradaki V^2 uygulanan modele göre elde edilen varyansdır. Yani V^2 bağımlı değişken, $1/X$ de bağımsız değişken gibi düşünülerek α ve β katsayıları tahmin edilir. Burada daha öncede söylendiği gibi göreli varyans olan V^2 daha önce anlatılan varyans tahmini yöntemleri ile daha önceki verilerden ya da araştırmanın küçük bir kısmından elde edilen verilerden daha sonra da α ve β katsayıları en küçük kareler yöntemiyle tahmin edilir. Bunun gibi daha değişik modeller de düşünülebilir. Bunlar;

$$V^2 = \alpha + \frac{\beta}{X} + \frac{\gamma}{X^2} \tag{34}$$
$$V^2 = (\alpha + \beta X)^{-1} \tag{35}$$
$$V^2 = (\alpha + \beta X + \gamma X^2)^{-1} \tag{36}$$
$$Log(V^2) = \alpha + Log(X) \tag{37}$$

Bu modeller gibi daha birçok model mevcuttur. Ancak şimdiye kadar yapılan araştırma sonuçlarına ve araştırmacıların tecrübelerine göre $V^2 = \alpha + \frac{\beta}{X}$ en uygun ve en çok kullanılan bir model olup verdiği sonuçlar bakımından da başarılı bir modeldir. Diğer modeller için de fazla bir teorik çıkarsama ve çalışma yapılmamıştır.

Bu yöntemde α ve β ya da modeldeki parametrelerin tahmini için en küçük kareler yöntemi kullanılmaktadır. Daha sonra toplamın varyansını elde etmek için modelimiz göreli varyansa uygun olarak elde edildiği için modelimizi X^2 terimi ile çarpmamız gerekir. Buna göre toplam tahmini için varyans modelimiz;

$$V(\hat{X}) = \alpha \hat{X}^2 + \beta \hat{X} \tag{38}$$

Oran için elde edilen varyans modeli ise;

$$V(\hat{p}) = \frac{b}{\hat{X}} \hat{p}(1 - \hat{p}) \tag{39}$$

Kish ve diğer arkadaşları varyansları elde etmek için düzen etkisini (design effect) kullanmışlardır. Düzen etkisi örnekleme tasarımından elde edilen varyansın Basit Rasgele Örnekleme varyansına oranıdır. Buna göre;

$$Deff = \frac{\text{Örnekelmeden elde edilen } \sigma^2}{BRÖ\sigma^2} \tag{40}$$

Göreli varyans düzen etkisine bağlı olarak yazılabilir.

$$V^2 = \frac{NDeff}{Xn} - \frac{Deff}{n} \tag{41}$$

Genelleştirilmiş varyans fonksiyonu (GVF) genellikle çok aşamalı örnekleme planlarında, hanehalkı araştırmalarında kullanılır. Genellikle aylık ya da yıllık olarak yapılan geniş araştırmalarda da kullanılır. Bu araştırmalarda birçok tahmin edici ve bu tahmin edicilerin varyansı elde edilmek istenir ve bunlar tablolanır. Bu ise raporu sadece sayı ve tablo yığınına dönüştürür. Bunun gibi sadece sayı ve tablolardan oluşan sayfalar dolusu rapor okuyucunun fazla ilgisini çekmez. Bu raporların oluşturulması da fazla zaman ve emek kaybına neden olur. İşte Genelleştirilmiş varyans fonksiyonu sayesinde tahmin edilmek istenen tahminler gruplandığı ve bu gruplarda yer alan istatistiklerin tahminleri ve varyanslarının tahminleri çalışma öncesinde araştırmacıya sunulup, eldeki verilere uygulanan model ile varyansların elde edilmesiyle çok sık yapılan geniş araştırmalarda bu gibi olumsuzluklardan uzaklaştırılmış olur. Zamandan ve emekten kazanç sağlanmış olur. Eskiden yapılan araştırmalardan elde edilmiş verilerle şimdiki ve gelecekteki araştırmanın hesaplamaları yapılabilir. Bu yöntemin dezavatajı ise V^2 ve $\hat{X}$ arasındaki modelden tam emin olamayışımızdır. Ancak genellikle $V^2 = \alpha + \frac{\beta}{X}$ modeli kullanılmaktadır. Ayrıca konu içerisinde yazdığımız diğer modeller üzerinde fazla bir teorik çalışma yapılmamış ve büyük çaptaki araştırmalarda kullanılmamıştır. (Wolter, 1985; Vaillant, 1987; Lohr, 1999)

6. YÖNTEMLERİN KARŞILAŞTIRILMASI

Taylor serisi yöntemi karmaşık örnekleme planlarında varyans tahmini için kullanılan en eski dolayısıyla teorisi daha çok çalışılmış olan yöntemdir. İstenilen istatistik için varyans eşitliği elde edilebilir. Bu yöntem için birçok bilgisayar programı geliştirilmesine rağmen ilgilenilen istatistik eğer kullanılan o programda yoksa kullanıcı kendi kodunu yazmak zorunda kalabilir. Bu yöntem yeniden örnekleme yöntemleri ile karşılaştırıldığında bu yöntemlere göre karmaşık bir yapıya ve hesaplama işlemlerine gerek duymamaktadır. Ancak yeniden örnekleme yöntemleri de tüm istatistikler için kullanılabilir iken eğer istatistiğin değişkenleri için türev alınamıyorsa Taylor serisi yöntemi kullanılamaz. Rasgele grup yöntemi açıklaması, uygulanması ve hesabı oldukça kolay bir yöntemdir. Bu yöntemin olumsuz yönü ise tahminler için fazla rasgele grup sayısı kullanmasıdır. Dengeli olarak tekrarlanan tekrarlı yöntem tüm istatistikler için kullanılırken her tabakasında sadece iki tane ön örneklem birimi ya da birim bulunan planlara uygulanabilir. Jacknife tüm istatistikler ve tüm örnekleme planları için kullanılırken ortanca gibi düzgün olmayan istatistiklerde başarısızdır. Bootstrap de tüm istatistikler ve tüm örnekleme planları için kullanılabilir. Ancak bu yöntemde çok fazla tekrar gerektirir. Dengeli olarak tekrarlanan tekrarlı yöntem daha az hesaplama gerektirirken bootstrap de çok fazla tekrar gerektirdiğinden hesabı fazladır. Genelleştirilmiş varyans fonksiyonu basit ve ucuzdur ancak olumsuz özelliği uygulanan modelden tam emin olamayışımızdır. Bu yöntemler uygulanırken doğruluk, esneklik ve yönetimsel düşünceler gibi kriterlere de bakmak gerekir. (Kish and Frankel, 1974; Wolter,•1985; Toprak, 1996; Lohr. 1999)

7.ORANSAL, ÇARPIMSAL, REGRESYON TAHMİN EDİCİLERİNİN VARYANSLARI

Bilindiği gibi oransal tahmin $\hat{\bar{Y}}_O = \frac{\hat{Y}}{\hat{X}}.\bar{X}$, çarpımsal tahmin $\hat{\bar{y}}_Ç = \bar{y}\frac{\bar{x}}{\bar{X}}$ ve regresyon tahmini ise $\bar{y}_{dr} = \bar{y} + b(\bar{X} - \bar{x})$ ' dir. Şimdi bu tahmin ediciler için anlatılan yöntemler kullanılarak varyans eşitlikleri elde edilmiştir. Taylor serisi yöntemi kullanılarak oransal ve çarpımsal tahmin ediciler için elde edilen varyans eşitlikleri,

$$V(\hat{\bar{Y}}_O) = N^2 (R^2 V(\hat{X}) + V(\hat{Y}) - 2RCov(\hat{X},\hat{Y})),$$

$$V(\hat{\bar{y}}_Ç) = \frac{1}{N^4 \bar{X}^2}.(Y^2.V(\hat{X}) + X^2.V(\hat{Y}) - 2YXCov(\hat{X},\hat{Y})).$$

Rasgele grup yöntemi kullanılarak oransal ve regresyon tahmin ediciler için elde edilen varyans eşitlikleri,

$$\hat{V}_1(\hat{\bar{Y}}_{O\alpha})=\frac{\sum_{\alpha=1}^{k}(\hat{\bar{Y}}_{O\alpha}-\hat{\bar{\bar{Y}}}_O)^2}{k(k-1)},\quad \hat{V}_2(\hat{\bar{Y}}_{O\alpha})=\frac{\sum_{\alpha=1}^{k}(\hat{\bar{Y}}_{O\alpha}-\hat{\bar{Y}}_O)^2}{k(k-1)},$$

$$\hat{V}_1(\bar{y}_{dr,\alpha})=\frac{\sum_{\alpha=1}^{k}(\bar{y}_{dr,\alpha}-\bar{y}_{dr})^2}{k(k-1)}.$$

Dengeli olarak tekrarlanan tekrarlı yöntemi kullanılarak oransal ve regresyon tahmin ediciler için elde edilen varyans eşitlikleri,

$$\hat{V}_{DTT}(\bar{Y}_O)=\frac{1}{k}\sum_{\alpha=1}^{k}(\hat{\bar{Y}}_{O\alpha}-\hat{\bar{Y}}_O)^2,\quad \hat{V}^c_{DTT}(\bar{Y}_O)=\frac{1}{k}\sum_{\alpha=1}^{k}(\hat{\bar{Y}}^c_{O\alpha}-\hat{\bar{Y}}_O)^2$$

$$\hat{\bar{V}}_{DTT}(\hat{\bar{Y}}_O)=\frac{1}{2}\left[\hat{V}_{DTT}(\hat{\bar{Y}}_O)+\hat{V}^c_{DTT}(\hat{\bar{Y}}_O)\right],\quad \hat{V}^+_{DTT}(\hat{\bar{Y}}_O)=\frac{1}{4k}\sum_{\alpha=1}^{k}(\hat{\bar{Y}}_{O\alpha}-\hat{\bar{Y}}^c_O)^2,$$

$$\hat{V}_{DTT}(\bar{y}_{dr})=\frac{1}{k}\sum_{\alpha=1}^{k}(\bar{y}_{dr,\alpha}-\bar{y}_{dr})^2.$$

Jacknife yöntemi kullanılarak oransal ve regresyon tahmin ediciler için elde edilen varyans eşitlikleri,

$$\hat{V}_{JK1}(\hat{\bar{Y}}_O)=\frac{1}{k(k-1)}\sum_{\alpha=1}^{k}(\hat{\bar{Y}}_{O\alpha}-\hat{\bar{\bar{Y}}}_O)^2,\quad \hat{V}_{JK2}(\hat{\bar{Y}}_O)=\frac{1}{k(k-1)}\sum_{\alpha=1}^{k}(\hat{\bar{Y}}_{O\alpha}-\hat{\bar{Y}}_O)^2,$$

$$\hat{V}_{JK1}(\bar{y}_{dr})=\frac{1}{k(k-1)}\sum_{\alpha=1}^{k}(\bar{y}_{dr,\alpha}-\bar{\bar{y}}_{dr})^2.$$

Bootstrap yöntemi kullanılarak oransal ve regresyon tahmin ediciler için elde edilen varyans eşitlikleri,

$$\hat{V}_B(\hat{\bar{Y}}_O)=\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\bar{Y}}_{O(b)}-\hat{\bar{Y}}_{O(.)})^2,$$

$$\hat{V}_B(\bar{y}_{dr})=\frac{1}{B-1}\sum_{\alpha=1}^{k}(\bar{y}_{dr,b}-\bar{y}_{dr(.)})^2.$$

Genelleştirilmiş varyans fonksiyonu kullanılarak oransal tahmin edici için elde edilen göreli varyans eşitliğini elde etmek için önce doğrusal formda olmayan tahmin edicimiz Taylor serisi yöntemi kullanılarak doğrusallaştırıldı. Ve göreli varyanslar elde edildi. Daha sonra her iki değişken için de model denklemleri kullanılarak oransal tahminin göreli varyans denklemi elde edildi.

$$V^2(\hat{\bar{Y}}_O)=\alpha+\frac{\beta}{X}+\alpha+\frac{\beta}{Y}=2\alpha+\beta(\frac{Y+X}{XY}).$$

8. UYGULAMA

Ankara ilinde bulunan 843 tane sanayi kuruluşundan elde edilen veriler kullanılarak sanayi kuruluşları içinde hukuki durumu anonim şirketi olanların oranı, çalışan kişi sayısı ortalaması gibi değişkenlerin tahminlerini ve varyanslarını elde edebilmek için bahsedilen yöntemleri kullanarak sonuçlara ulaşan Visual Basic dilinde bir program yazılmıştır. Elimizdeki kitleden bir örnekleme yapılmıştır. Bunun için sanayi kuruluşları faaliyet gösterdikleri alana göre tabakalanarak, her tabakadan birimler sistematik örnekleme ile seçilmiştir. Elde edilen tahmin sonuçlarından bazıları Çizelge 1' de verilmiştir;

Tahminler	Taylor Serisi Yöntemi	Rasgele Grup Yöntemi	Dengeli Olarak Tekrarlanan Tekrarlı Yöntem	Jacknife Yöntemi	Bootstrap Yöntemi
Anonim Oranı	0,373723	0,372168	0,467378	0,37372	0,375781
Çalışan Ortalaması	85,476157	68,640538	90,762159	85,47611	84,763577

Çizelge 1. Yöntemlerin uygulanması ile elde edilen tahminler

Elde edilen varyans tahmin sonuçları ise Çizelge 2' de verilmektedir;

Tahminler	Taylor Serisi Yöntemi	Rasgele Grup Yöntemi	Dengeli Olarak Tekrarlanan Tekrarlı Yöntem	Jacknife Yöntemi	Bootstrap Yöntemi
Anonim Oranı	0,000623	0,000623	0,007852	0,001897	0,002794
Çalışan Ortalaması	53,06538	53,06538	139,508982	115,05137	122,74833

Çizelge 2. Yöntemlerin uygulanması ile elde edilen varyans tahminleri

Ayrıca elde edilen sonuçlar karşılaştırılmıştır. Değişkenlerin tahminleri ve bunların varyans tahminleri için Taylor serisi ile rasgele grup yönteminden elde edilen sonuçlar birbirine çok yakındır ve en küçük olanlarıdır. Ayrıca örnekleme planımız her tabakasında iki birim içermediği için dengeli olarak tekrarlanan tekrarlı yöntemi uygulayabilmek için her tabakada iki birim bırakıldığından tüm değişkenler için en büyük sonuçları vermiştir. Jacknife ve Bootstrap yöntemleri ise Taylor serisi ve rasgele gruptan büyük fakat dengeli olarak tekrarlanan tekrarlı yöntemden küçük sonuçlar vermiştir. Dolayısıyla en iyi sonucu rasgele grup yönteminden elde ettiğimizi söyleyebiliriz.

KAYNAKLAR

CHAO, M. and LO, S. (1985), A Bootstrap Methods For Finite Population, The Indian Journal Of Statistics, 47, Series A, 399-405.

CHERNİCK, M.R. (1999), Bootstrap Methods A Practitioner's Guide: John Wiley & Sons.

COHEN, S.B. (1997), An Evulation Of Alternative PC-Based Software Packages Developed For The Analysis Of Complex Survey Data, The American Statistician , 51, No.3, 285-292.

EFRON, B. (1979), Bootstrap Methods Another Look At The Jacknife, The Annals Of Statistics, 7, No.1, 1-26.

EFRON, B. and TİBSHİRANİ, R.J.(1993), An Introduction To The Bootstrap: Chapman&Hall International Thompson Publishing.

HEDAYET, A. S. and SİNHA, B.K. (1991), Design And Inference In Finite Population Sampling, A Wiley-Interscience Publication, New York : John Wiley &Sons.

KİSH, L. and FRANKEL,, M.R.(1974), Inference From Complex Samples, Journal of Royal Statistical Society, B(36), 1-37.

Kish, L. and Frankel, M.R. (1970), Balanced Repeated Replications for Standard Errors, Journal of American Statistical Association, 65, 1071-1095.

LEE, K.H. (1972), Partially Balanced Designs For Half Sample Replication Method Of Variance Estimation, Journal of American Statisticial Association, 67, No.338, 324-334.

LEPLOWSKİ, J. and Bowles, J. (1996), Sampling Error Software For Personel Computers, The Survey Statistician, 35, 10-17.

LOHR, S.L. (1999), Sampling: Design and Analysis: Duxbury Press.

MİLLER, R.G. (1974), The Jacknife A Review, Biometrika, 61, 1-15.

RAJ, D. (1998), Sample Survey Theory, Naraso Publishing House: London.

RAO, J.N.K. (1997), Developmnets in sample survey theory: an apprasial, The Canadian Journal of Statistics, 25, 1-21.

RAO, J.N.K. and Krewski, D. (1981), Inference From Stratified Samples; Properties Of The Linearization, Jacknife And Balanced Repeated Replication Methods, The Annals Of Statistics, 9, No. 5, 1010-1019.

RAO, J.N.K. and WU, C.F.J. (1988), Resampling Inference With Complex Survey Data, Journal of American Statisticial Association, 83, No.401, 231-241.

RAO, J.N.K. and WU, C.F.J. (1985), Inference From Stratified Samples:Second-Order Analysis Of Three Methods For Nonlinear Statistics, Journal of American Statisticial Association, 80, No.391, 620-630.

SARNDAL, C.E., SWENSON B. and WRETMAN J. (1991), Model Assisted Survey Sampling, New York,: Springer Series-Verlag,.

TOPRAK, A.Ö. (1996), Karmaşık Örnekleme Planlarında Varyans Tahmini, Bilim Uzmanlığı Tezi, Orta Doğu Teknik Üniversitesi.

VAİLLANT, R. (1987), Generalized Variance Functions In Stratified Two-Stage Sampling, Journal of American Statisticial Association, 82, No.398, 499-508.

VERMA, V. (1992), Sampling Methods, Statisticial Institute for Asia and the Pacific, Tokyo.

WODDRUFF, R. S. (1971), A Simple Method For Approximating the Variance of a Comlicated Estimate, Journal of American Statisticial Association, 66, 411-414.

WOODRUFF, R.S. and Causey, B. D. (1976), Computerized method For Approximating The Variance Of A Complicated Estimate, Journal of American Statisticial Association, 71, No.354, 315-321.

WOLTER, K.M. (1985), Introduction to the Variance Estimation, Springer Series in Statistics, New York.

VARİOUS VARİANCE ESTİMATİON METHODS FOR COMPLEX SAMPLİNG SURVEY AND APPLİCATİON

ABSTRACT

In this study, variance estimation methods for complex sampling survey were examined. In the first chapter a brief introduction was made. In the following chapters, Taylor series method, dependent and independent random group methods from random group method, balanced repeated replication method, jackknife and bootstrap methods from resampling methods and lastly generalized variance function were examined, theoretical properties, methods appliance forms, restrictions of sampling were studied. .In the sixth chapter, these methods were compared with each other based on their theoretical properties, appliance forms, accuracy of results and costs. In the seventh chapter, variance equations were obtained using the methods those are described for ratio, multiplier and regression estimators whose variance equations have not been examined in complex sampling survey up to now. In the eigth chapter, some results obtained from a software application developed to obtain estimators and variance equations over a stratified systematic sampling having 163 units, obtained from 843 industrial corporations in Ankara were given. Lastly some results of estimators and variance estimators which were obtained from these methods were placed and compared with each other.

Key Words: *Variance estimation, complex survey, resampling methods*