T.C.
NİĞDE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİYOLOJİ ANABİLİM DALI

REGRESYONDA YENİDEN ÖRNEKLEME YÖNTEMLERİNİN KARŞILAŞTIRMALI OLARAK İNCELENMESİ

Derviş TOPUZ

YÜKSEK LİSANS TEZİ

OCAK 2002

Fen Bilimleri Enstitüsü Müdürlüğüne:

Bu çalışma jürimiz tarafından BİYOLOJİ ANABİLİM DALI'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Başkan :Prof.Dr. Ahmet Nuri YAYINTAŞ

Üye :Doç.Dr. Tamer KAYAALP

Üye :Yrd.Doç.Dr. Suat ŞAHİNLER

Üye :Yrd.Doç.Dr. Oğuz KILIÇOĞLU

Üye :Yrd.Doç.Dr. Onur KÖKSOY

Onay:

Bu tez 13../03/2002 tarihinde, Fen Bilimleri Enstitüsü Yönetim Kurulunca belirlenmiş olan yukarıdaki jüri üyeleri tarafından uygun görülmüş ve Enstitü Yönetim Kurulunun kararıyla kabul edilmiştir.

05./04/2002

Doç.Dr. Aydın TOPÇU

Enstitü Müdürü

ÖZET

REGRESYONDA YENİDEN ÖRNEKLEME YÖNTEMLERİNİN KARŞILAŞTIRMALI OLARAK İNCELENMESİ

TOPUZ, Derviş

Niğde Üniversitesi
Fen Bilimleri Enstitüsü
Biyoloji Anabilim Dalı

Danışman : Prof. Dr. Ahmet Nuri YAYINTAŞ
Ortak Danışman : Yrd. Doç. Dr. Suat ŞAHİNLER

Ocak 2002, 68 sayfa

Regresyon analizinde yeniden örnekleme yöntemlerinin incelenmesi amacıyla bootstrap ve jackknife yöntemleri ele alınmıştır. Bu yöntemler zaman serileri, simulasyon teknikleri, tek ve çok değişkenli istatistik analizler ve regresyon analizi gibi pek çok alanlarda uygulanmaktadır.

Yeniden örnekleme yöntemleri özellikle en küçük kareler regresyon analizindeki hata değerleri ile ilgili varsayımların gerçekleşmediği durumlarda birer düzeltme yöntemi olarak da kullanılmaktadırlar.

Çalışmada, Kayseri Doğum Hastanesinden elde edilen 2000-2001 yıllarındaki 320 hastaya ait anne yaşı (X_1), gebelik süresi (X_2) ve bebeğin doğum ağırlıklarına (Y) ilişkin veriler kullanılmıştır. Önce klasik örnekleme yöntemleri ile elde edilen bu verilere ait en küçük kareler regresyon modeli ve ilgili parametre tahminleri hesaplanmış, daha sonra aynı verilere yeniden örnekleme yöntemleri uygulanarak elde edilen verilere ait en küçük kareler regresyon modeli ve ilgili parametre tahminleri tekrar tahmin edilmiş, böylece her iki şekilde elde edilen parametre tahminlerinin karşılaştırılması yapılmıştır.

Anahtar Sözcükler: Bootstrap yöntemi, Jackknife yöntemi, Yeniden örnekleme yöntemleri, Parametrik olmayan yöntemler, regresyon, Parametrik olmayan güven aralıkları

SUMMARY

COMPERATIVE INVESTIGATION OF THE RESAMPLING METHODS IN REGRESSİON ANALYSIS

TOPUZ, Derviş

University of Nigde
Graduate School of Natural and Applied Sciences
Department of Biology

Supervisor : Prof. Dr. Ahmet Nuri YAYINTAŞ
Co-Advisor : Yrd. Doç. Dr. Suat ŞAHİNLER

January 2002, 68 pages

In this study, bootstrap and jackknife methods were taken up for investigation of the resampling methods in regression analysis. These methods have a many application areas suchas the time series, simulation technigues, unique and multiple variable analysis and regression analysis etc.

In regression analysis, resampling methods are used as a correction method, in case of the error value hypotesis unrealized.

In this work, the data which were taken in kayseri maternity Hospital from 320 patient mothers, whose ages, pregnancy time, and weigth of baby are X_1, X_2 and Y, respectiveley, are used. First, the least sequare regression analysis model and its related parameters are determined from the data which were obtained with clasicall sampling method. Next, the same data are agaın used to obtain the least sequare regression model and its related parameters by aplying resampling methods. Then, these two results are compared.

Key Words: Bootstrap method, Jackknife method, resampling methods, non parametric methods, bias estimation, non parametric satety regions.

TEŞEKKÜR

Bu çalışmanın yürütülmesinde teşvik ve desteğini esirgemeyen, bundan sonraki çalışmalarımda ismini minnetle anacağım değerli hocam Prof. Dr. Ahmet Nuri YAYINTAŞ'a, ders aşamasında, konunun belirlenmesinde, istatistik analizlerin yapılması, değerlendirilmesi, tez yazımı ve düzeltilmesi aşamalarındaki çok değerli yardımlarından dolayı yardımcı danışmanım Mustafa Kemal Üniversitesi Öğretim Üyelerinden Yrd. Doç. Dr. Suat ŞAHİNLER'e, konuyla ilgili makalelerin değerlendirilmesi, tezin yazım aşamalarında ve düzeltmelerindeki değerli yardımlarından dolayı Niğde Üniversitesi Ulukışla Meslek Yüksekokulu Öğretim Görevlilerinden Duran ÖZKÖK'e ve çalışmalarım sırasındaki büyük özverilerinden dolayı eşim Özlem TOPUZ'a teşekkürlerimi borç bilirim.

İÇİNDEKİLER

ÇİZELGELER DİZİNİ

BÖLÜM I

GİRİŞ

1.1. Giriş

Bir populasyonun bilinmeyen bazı özelliklerini ortaya çıkarmak için gözlemler yapılır, veriler toplanır. Bu verilerin elde edilmesi sırasında iki yöntem söz konusu olmaktadır. Bunlardan birincisi, üzerinde araştırma yapılan populasyonu oluşturan tüm birimlerin incelenmesi ki buna tamsayım denir. Verilerin tam sayım ile elde edilmesi söz konusu olduğunda bir takım derleme hataları dışında ilgilenilen özellik ile ilgili kesin bilgiler elde edilmektedir. Populasyonun tamamı üzerinde inceleme yapmak çoğu zaman olanaksızdır. İncelenebilecek olsa bile; büyük işlemlere, zamana, personele ve bütçeye gerek vardır. Ayrıca değerlendirme uzun zaman alacağı için toplanan veriler güncelliğini yitirecek ve yararsız duruma gelecektir. Bu nedenle çabukluk, ucuzluk, yararlılık ve daha çok bilgi elde etme yönünden populasyondan alınacak örnek üzerinde çalışarak populasyona ilişkin bilgi edinilmeye çalışılır (Sümbüloğlu ve Sümbüloğlu, 1997). Bu gibi durumlarda populasyondan daha az sayıda bir alt birim seçilmekte ve bu birimdeki veriler gözlenerek incelenen değişken yada değişkenlerin istatistiklerinin elde edilmesi ile populasyon parametrelerinin tahmin edilmesi ve genellemelere gidilmesi örnekleme çalışmalarıdır.

Örneklemenin en önemli özelliği, populasyona ilişkin tutarlı, geçerli tahminlere ulaşmak için örnek hatasının minimum olmasını sağlamaktır. Örnek hatası minimuma indirgendiğinde, seçilen örnek populasyonun en iyi tahminini vermektedir. Populasyona ait parametreyi belirli bir güven aralığında kontrollü olarak tahmin etmek mümkün olmaktadır. Yapılan araştırmalarda incelenen değişken yada değişkenlerin değerleri kontrolümüz dışında olan bir çok etkenin etkisiyle değişebiliyorsa, bu değişkenlerle ilgili olarak populasyona ilişkin genel yargılara ulaşmak güçleşmektedir. Bunun yanında, doğadaki olayların değişmesi bir dereceye kadar belirli bir düzen içinde olmaktadır. Bu düzen içinde populasyondan rasgele olarak alınan örneklerin ortalama etrafinda yoğunlaşması beklenmektedir. Diğer durumda ise, ortalamadan uzaklaştıkça sapma göstermektedir. Populasyonun bilinmeyen bir parametresini tahmin etmek için populasyondan rasgele olarak birbirinden bağımsız ve eş olasılıkla örnekler alındığında, alınan bu örneklerdeki değişkenlerin dağılımı bilinmediğinden ortalama ve varyansın parametre tahmininde kullanılması tahminin geçerliliğini etkilemektedir (Püskülcü ve İkiz, 1986)

Araştırma konusu ile ilgili veriler örnekleme yöntemi ile elde edildiğinden istatistiğin önemli bir dalı olan istatistiksel çıkarım konusu gündeme gelmektedir. Çünkü, örnek verilerinden hareketle hesaplanan ortalama, oran, varyans, standart sapma, vs. gibi tahmin ediciler populasyon parametrelerinin tahmin edilmesi yada söz konusu parametre değerlerinde herhangi bir değişiklik olup olmadığının araştırılması amacıyla kullanılmaktadır.

Örnekten elde edilen bilgilerin geçerliliği, her şeyden önce, örneğin ve uygulanan istatistik yöntemlerin iyi seçilmesine bağlıdır. Örnekleme yöntemleri, bugün hemen, hemen tüm bilim dallarında, kamuoyu yoklamalarında, pazarlama araştırmalarında, uluslar arası ilişkilerde, kamu ve özel sektörün bir çok kesiminde ve diğer bir çok alanda kullanılmaktadır (Kabukçu, 1994).

Her örnekleme yönteminin olumlu ve olumsuz yönleri mevcuttur. Bu nedenle her yöntemin uygulanacağı probleme göre değerlendirilmesi gerekir. Bütün örnekleme yöntemlerinin ortak amacı; elde edilecek olan neticeyi tesir eden bazı faktörlerin etkilerini yok etmek veya bu faktörlerin etkilerinden dolayı meydana gelen örnekleme sapmalarını azaltmaktır.

Bazı örnekler bir populasyonun yalnız küçük bir parçası olduğu halde populasyonu tümüyle temsil edebilirler. Örneğin parmaktan alınan bir damla kan insan vücudundaki tüm kanın yapısını içerir. Oysa bir okuldan incelenmek için seçilen öğrencilerin o okulu tümüyle temsil etmeleri yada incelenmek için seçilen belirli sayıdaki ailelerin o bölgedeki aileleri tümüyle temsil etmeleri söz konusu değildir. Araştırılması istenilen topluluktan belli sayıda elemana sahip olan örnek veya örneklerin oluşturulması için çeşitli örnekleme yöntemleri geliştirilmiştir.

İstatistiksel tahmin etme yöntemleri, nokta ve aralık tahminleri olarak ikiye ayrılmaktadır. Populasyondan rasgele seçilen örneklerin tek bir sayısal değere göre yorumlanması nokta tahmini, parametrenin belirli bir olasılıkla minimum ve maksimum değerlerinin bulunması ise aralık tahmini olarak bilinmektedir. Tahmin etme işlemini belirleyen formülasyona tahmin edici, formülasyon gerçek verilerle işlendiğinde ise ulaşılan sayıya tahmin adı verilmektedir (Püskülcü ve İkiz, 1986)

Toplanan örneklerde tahmin hatasının minimum olması istenmektedir. Fakat parametrik tahmin etme yöntemleri örnek sayısı az olduğu zamanlarda güvenilir sonuçlar vermeyebilir ve aynı zamanda parametrik yöntemlerin varsayımı da bozulabilir. Bu durumu ortadan kaldırmak amacıyla parametrik olmayan yöntemlerden yararlanılmaktadır. Çünkü bu yöntemler aracılığı ile örnek sayısının yeterli olmadığı durumlarda belirli bir güven payı ile tahminin standart hatası minimuma indirilebilir.

Tahmin edicilerden hareketle, populasyon parametrelerinin tahmin edilmesi yada hipotez testi sürecinde tahmin edicinin elde edilmesi amacıyla;

1) En Küçük Kareler,

2) Maksimum olabilirlik,

3) Oran ve

4) Momentler

yöntemlerinden birini kullanmak mümkündür.

Ancak bu yöntemlerin hangisi kullanılırsa kullanılsın tahmin edici ile populasyon parametresi arasında tahmin hatası yada ortalama kare hata adı verilen bir fark mutlaka olmaktadır. Bu fark ise bir takım ölçme hataları dikkate alınmaz ise sistematik hatalar (sapma) ve tesadüfi hatalar (örnekleme hataları) olmak üzere iki kaynağa sahiptir.

Tahmin edicilerin populasyon parametresinden tek yönde sapma gösterdiği sistematik hatalar (sapma) birçok nedene bağlı olarak ortaya çıkmakta ve bu hataya yol açan neden ortadan kaldırılmadıkça da yok olmamaktadır. Örnekleme yöntemlerinde, genellikle tahmin edicilerden hareketle populasyon parametresi tahmin yada test edilirken kullanılan tekniğin sapma içermeyen bir teknik olduğu varsayılmaktadır. Ancak, her zaman sapmasız tahmin ediciler bulmak mümkün olmadığı gibi, bu sapmayı tahmin etmek (ölçmek) de mümkün değildir.

Tesadüfi hata adı verilen örnekleme hataları ise tahmin edici ile populasyon parametresi arasında farklı yön ve düzeylerde ortaya çıkan hatalardır. Bu fark örnek hacmi arttıkça azalmakta ve populasyon hacmine eşit olduğunda sıfıra inmektedir. Örnekleme hataları ise standart hata adı verilen bir ölçü ile ortaya konmaktadır. Standart hata tahmin edilen değerlerin dağılımının beklenen değeri ile populasyon parametresi arasındaki farkın kareli ortalaması biçiminde elde edilmektedir. Ancak standart hatanın elde edilmesi için tahmin edicilerin örnekleme dağılımının oluşturulması gerekmektedir. Örnekleme dağılımının oluşturulması ise bilindiği gibi populasyondan mümkün ve muhtemel n hacimlik tüm örneklerin çekilmesi ile mümkündür (Çömlekçi, 1988).

Kuramsal olarak olanaklı görünen bu yöntemin (populasyondan mümkün ve muhtemel n hacimlik örneklerin çekilmesi) uygulanması her zaman mümkün olmayabilir. Bu nedenle de birtakım varsayımlar (normallik) üzerine kurulu merkezi limit teoreminden hareketle, standart hatanın elde edilmesi mümkün olabilmektedir. Ancak, söz konusu teoremin normallik varsayımı her zaman geçerli olmamaktadır.

Bunun yanı sıra, teoremin uygulanması için populasyon parametre yada parametrelerinin de bilinmesi gerekmektedir. Ancak bunun mantıklı bir geçerliliği yoktur. Çünkü, populasyon parametreleri ancak tamsayım sonucu elde edilebilirler. Bu durumda ise tahmin yada test yapmak gereği kalmaz. İstatistiksel yöntemlerde amaç, zaten bilinmeyen populasyon parametresinin tahmin yada test edilmesidir.

Bu nedenle standart hatanın gerçek değerinin hesaplanması yerine örneğe ait değerlerden hareketle tahmin edilmesi söz konusu olmaktadır. Ancak bu tahmin ile standart hatanın gerçek değeri arasındaki farkın büyüklüğünü ölçmek ve standart hata tahmininin güvenilirliğini test etmek mümkün değildir.

İstatistiksel yöntemlere parametrik olmayan bir yaklaşım sağlayan yeniden örnekleme yöntemleri, örnekleme yöntemlerinin söz konusu zayıflık ve dezavantajlarını ortadan kaldırma amacı taşımaktadır.

Populasyondan elde edilen örneklerin populasyonun yeterli bir göstergesi olmasına bağlı olarak, yeniden örnekleme yöntemleri, verilen örneklerden alt örneklerin seçilmesi prensibine dayanmaktadır. Bu örneklerden hareketle sapma, varyans, güven aralıkları vb populasyona bağlı bazı değerler tahmin edilmeye çalışılır. Tahmin edilen bu değerler bir modelin parametreleri için doğruluk ölçüleridir (Efron ve Tibshirani, 1993).
Yeterli sayıda birim üzerinde araştırma yapılamadığı zaman, parametrik tahminler hatalı olmakta ve bir parametre tahmin edicisinin taşıması istenilen istatistik özellikleri yerine gelmemektedir. Bu nedenle, olasılık dağılımı bilinmeyen bir F dağılımından rasgele alınan örneklerin hatalarını minimuma indirgemek, örnekleme dağılımının oluşturulması ve güçlü güven aralıkları oluşturmak için parametrik olmayan örnekleme yöntemlerinden yararlanılmaktadır. Böylece dağılımı bilinmeyen F dağılımından alınan n sayıdaki küçük bir örnekten elde edilen istatistiklerden yararlanarak parametre tahminlerinin minimum

hatayla belirlenmesi mümkün olmaktadır. Bu yöntemlerden en yaygın olarak kullanılanları Bootstrap ve Jackknife yeniden örnekleme yöntemleridir (Efron ve Tibshirani, 1986).

Söz konusu yöntemler, daha çok tahmin konularında kullanılırken, sınırlı da olsa değişik yeniden örnekleme yöntemleri aynı konuda da kullanılabilmektedir. Bu konulardan birisi de en küçük kareler regresyon analizidir. En küçük kareler yöntemi ile doğrusal regresyon modeli uydurulması, istatistik işlemler içinde geniş kullanım alanı bulan ve sebep-sonuç ilişkisine sahip iki veya daha fazla değişken arasındaki ilişkiyi belirleyen bir analizdir (Şahinler, 1997)

En küçük kareler regresyon analizi yapılırken gözlem değerleri, değişkenler ve hata hakkında birtakım varsayımlar ileri sürülmektedir. Ancak bu varsayımlar sağlandığında yapılan hesaplamalar istatistiki olarak geçerlilik kazanır. Yoksa cebirsel olarak doğru yapılan bir işlemin, bu varsayımlar tutmadığı müddetçe istatistiki bir değeri olmaz. Çünkü; varsayımların bozulmalarının test sonuçları ve yorumlar üzerine çok önemli etkileri olabilmektedir. Varsayımların sağlanmaması uydurulan modelin populasyonu iyi temsil etmediğini ve kararlı bir şekilde genelleştirilemeyeceğini gösterir. Dolayısıyla da bu model ile yapılan ileriye yönelik tahminlerin hatalı olma ihtimali çok yüksek olur. Yeniden örnekleme yöntemleri, en küçük kareler regresyon analizinin varsayımlarına gerek duymaksızın, parametrelerin tahmin edilmesi, populasyona ait varyans değerlerinin tahmin edilmesi ve tahmin hatalarının hesaplanmasında yararlı olmaktadır (Shao, 1995).

Yeniden örnekleme yöntemlerinin bilgisayarlarla birlikte uygulanmasındaki kolaylık, yöntemlerin kısa denilecek bir süre içinde çok popüler yapmıştır. Bootstrap yöntemi işletilerek elde edilen karakteristik dağılım fonksiyon, beklenen değerler gibi ifadeler ilgili gösterimlerin üzerine (*) işareti konularak, diğerlerinden ayırt edilecektir. Yıldız işareti X veri setinin gerçek olmadığını fakat daha ziyade rasgele bir örnekleme olduğunu veya X'in yeniden örneklenmiş bir versiyonu olduğunu göstermektedir. Diğer bir değişle bootstrap örnekleri X_1^* X_2^* X_3^* ,........,X_n^* n nesneli ($X_1, X_2, X_3,..........,X_n$) örnekten yeniden seçme ile elde edilen n büyüklüğündeki rasgele bir örnektir.

Bu çalışmada Kayseri Doğum Hastanesinden elde edilen 2000-2001 yıllarındaki 320 hastaya ait anne yaşı (X_1), gebelik süresi (X_2) ve bebeğin doğum ağırlıklarına (Y) ilişkin veriler kullanılmıştır. Bu verilere ait EKK regresyon doğrusu ile yeniden örnekleme yöntemleri (Boostrap ve Jackknife yöntemleri) uygulanarak elde edilen EKK regresyon doğrusuna ait parametre tahminleri karşılaştırılmıştır.

1.2. Önceki Çalışmalar

Efron (1979) Bootstrap yöntemine ilişkin yaptığı çalışmada, istatistiğin bootstrap dağılımını elde etmede kullanılan üç farklı yöntemi tanıtmıştır. İlk yöntem, doğrudan kuramsal hesabı içeren yöntemdir. Çalışmada, kuramsal hesabı içeren yöntemin tanıtılması iki örnek ile desteklenmiştir. İkinci yöntem Monte Carlo yaklaşımı, üçüncü yöntem ise Taylor serisi açılımının kullanıldığı yöntemdir. Çalışmada ortanca tahmininin Monte Carlo yaklaşımı ile bir değerlendirmesi yapılmıştır. Araştırmacı çalışmasında bootstrap yönteminin regresyon modelinde nasıl uygulandığını da açıklamıştır. Yaptığı çalışmada örnek hacmini artırmaksızın populasyon parametresi ile tahmin edici arasındaki sapmanın azaltılacağını ve teorik olarak elde edilmesi mümkün gibi görünen ama uygulama da söz konusu olmayan tahmin edicilerin örnekleme dağılımının oluşturulabileceğini belirterek örnekleme dağılımlarının bu şekilde oluşturulmasıyla tahmin edicilerin standart hatasının daha sağlıklı olarak elde edilmesinin sağlanabileceğini bildirmiştir.

Efron ve diğ. (1981) standart sapmanın parametrik olmayan tahminlerini elde etmek için geliştirilmiş parametrik olmayan yöntemleri incelemişler, bir Monte Carlo denemesi ile standart normal dağılımlı bir populasyon için korelasyon katsayısının standart sapmasını değişik yöntemler kullanılarak elde etmişlerdir. Araştırmacılar bu çalışmalarında jackknife ve bootstrap yöntemlerinin doğrusal olmayan regresyon problemlerinde regresyon parametreleri için güvenilir tahminler sağlamanın yanı sıra, Cox regresyonunda Kaplan Meier tahmincilerinin dağılımı ve varyansını tahmin etmek amacıyla kullanıldığını bildirmişlerdir.

Efron ve Gong (1983) yaptıkları çalışmada, standart sapma tahmininde kullanılan parametrik olmayan tahmin yöntemleri tanıtmışlar; bootstrap yönteminin standart sapma tahmininde kullanışını adım adım anlatarak populasyon dağılımının normal ve negatif üstel olduğu durumlar için ortalamanın standart sapmasının bootstrap ve jackknife tahminleri ile elde edildiğini bildirmişlerdir.

Wu (1985) yaptığı çalışmada eşit varyanslılık durumunda $\hat{\beta}$'nin bootstrap varyans tahmini sapmasız olurken, heterojen varyanslılık durumunda sapmalı ve tutarsız olduğundan, bu yöntemin heterojen varyans durumunda tahmin yapmaya uygun olmadığını bildirmiştir.

Stine (1985) yaptığı çalışmada $Y = X_i\beta_i + \varepsilon$ doğrusal regresyon modelinde, hata değerlerinin dağılımını oluşturmak için, $Sapma_\varepsilon = e_i - \bar{e}$ değerine $\frac{1}{n}$ olasılığı vererek, deneysel dağılımın tahmin edilebildiğini göstermiştir.

Wu (1986) yaptığı çalışmada bootstrap yönteminin uygulamasında hata değerlerinin normal dağılıma sahip olduğu varsayımlarının gerçekleşmediği ve $\hat{\beta}$'nın standart hatasının sapmalı sonuç verdiği durumlarda, $Sapma_\varepsilon = (e_i - \bar{e})$'nin yerine, $\frac{e_i}{(1 - kn^{-1})^{1/2}}$ normalleştirilmiş hata terimlerinden ε_i^*'ler seçilerek, bootstrap Y^* değerlerinin elde edileceğini bildirmiştir.

Efron ve Tibshirani (1986) yaptıkları çalışmada, standart sapmanın bootstrap tahminini ele almışlar; parametre tahmininde hangi tahmin edicinin kullanılması gerektiğini; tahmin edicinin parametre için ne kadar doğru bilgi taşıdığı sorularının önemli olduğunu ifade ederek bootstrapın ikinci soruya yanıt vermek için geliştirilmiş bir yöntem olduğunu bildirmişlerdir. Araştırmacılar çalışmalarında çeşitli bootstrap güven aralığı yöntemlerine yer vererek bootstrap tekrarlama sayısının ne kadar geniş tutulması gerektiğine de değinmişler ve jackknife yöntemi ile bootstrap yöntemi arasındaki ilişkileri ele almışlardır.

Diciccio ve Tibshirani (1987) yaptıkları çalışmada, çeşitli bootstrap güven aralıklarını ele almışlar, varyans ve korelasyon katsayısı için ilgili güven aralığı yöntemleri kullanılarak güven aralıkları elde etmişlerdir.

Hall (1986) bootstrap güven aralıklarının kuramsal olarak karşılaştırıldığı bir çalışma yapmış, çalışmasında kuramsal bir bakış açısıyla bootstrap güven aralıklarını karşılaştırmıştır.

Oudewicz (1988) yaptığı çalışmada elde edilen tahmin edicilerden hareketle $1-2\alpha$ güven düzeyindeki jackknife güven aralıklarını oluşturarak, örnek hacminin 30 ve üzerinde olması durumlarında normal dağılım koşullarından hareketle t yerine z tablo değerlerinin kullanılacağını bildirmiştir.

Stine (1990) bootstrap yöntemi için gerekli tüm kopyalamaların bilgisayar teknolojisindeki gelişmelere paralel olarak gerçekleştirilebileceğini bildirmiştir.

Efron (1990) yaptığı çalışmada, bootstrap tekrarlama sayısını azaltma amacına yönelik olarak daha etkili bootstrap hesaplamaları yapmış bootstrap tekrarlarının büyüklüğü konusunda çeşitli görüşleri bildirerek 50 ile 200 arasında bir sayı olmasının yeterli olduğunu belirtmiştir. Tek örnek durumu için bootstrap algoritmasını çok açık bir ifadeyle göstermiştir. Bootstrap sapma tahmini konusunu, hem bilinen şekliyle hem de daha az bootstrap tekrarlamasını gerektiren farklı bir şekliyle ele almıştır. İlgilenilen istatistiğin, ana örnek kullanılarak elde edilen değeri S^0 ile, bootstrap örnekleri kullanılarak elde edilen değeri S^* ile, B tane bootstrap örneğinden elde edilen S^* değerlerinin ortalamasını da $\overline{S}^*$ ile göstermiş ve bootstrap yan (ikincil) tahmininin $\overline{Yan_B} = \overline{S}^* - S^0$ ifadesine eşit olduğunu bildirmiştir. Bu noktada, $\overline{Yan_B}$'den daha etkili olan, yani daha az bootstrap tekrarlamasını gerektiren $\widehat{Yan_B}$ yan (ikincil) tahmini ele almıştır. $P = (P_1, P_2, P_3, ..., P_n)$ yeniden örnekleme olasılık vektörü olmak üzere $\overline{P} = \sum_{i=1}^{B} P^b / B$, her bir öğesi B tane yeniden örneklem vektörünün öğelerinin ortalamaları olan n boyutlu bir vektörü göstermek üzere $\overline{Yan_B}$' den daha etkili olan yan tahmini $\widehat{Yan_B} = \overline{S} - S(\overline{P})$ eşitliği ile vermiştir. Buna dayanak olarak ta Taylor serisi açılımının bootstrapta kullanımını göstererek bu yaklaşıma ilişkin kuramsal bilgi vermiştir.

Money (1993) jackknife yönteminde parametrik varsayımlar kullanmak yerine, örneğe ait verilerin değişkenliğinin açıklanması yoluyla istatistiğin değişkenliğini açıklamaya yönelik yöntem olduğunu belirterek, yöntem sürecinin aşamalarını açıklamıştır. Ayrıca, bootstrap yönteminin örnekleme yöntemlerinden farklı bir yöntem olduğunu belirterek yöntemin örnekleme dağılımını oluşturmak için güçlü varsayımlar ve analitik formüller yerine büyük sayıda tekrarlanmış hesaplamalar içerdiğini, bootstrap yönteminin, örnekleme yöntemlerinin yetersizliklerini ortadan kaldırmak, tahmin edicinin örnekleme dağılımının oluşturulmasını sağlamak hem de sapmanın azaltılması amacına yönelik parametrik olmayan bir yöntem olarak geliştirildiğini belirtmiştir.

Efron (1994) bootstrap yöntemi ile örnek ortalamalarının örnek dağılımlarının standart sapmasını tahmin etmek için gerekli sürecin aşamalarını ve formüllerini göstermiştir. Ayrıca yeniden örnekleme yöntemlerinde, olasılıkların ve yeniden örnekleme olasılık vektörünün önemini belirterek, jackknife standart sapma değerlerinin nasıl tahmin edileceğini $\hat{\sigma}_{jack} = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{(-j)} - \bar{\hat{\theta}}_{(-j)})^2}$ formülü ile belirtmiştir.

Shao (1995) en küçük kareler tahmin edicileri üzerinde herhangi bir üstünlüğe sahip olmamasına rağmen, regresyon analizinde bootstrap ve jackknife tahmin edicilerinin varsayımlardan olan sapmalardan etkilenmediklerini, bootstrap ve jackknife yöntemlerinin güçlüğünün, sabit varyans varsayımının varsayılıp varsayılmadığına bağlı olduğunu, yaptığı çalışmada gözlem çiftlerinden B tane örnek oluşturarak, her bir alt örnek için, β'nın bootstrap tahmin edicisini $(\hat{\beta}^*)$ elde ederek, bootstrap tahmin edicilerinin beklenen değerinin, yaklaşık olarak, β'nın en küçük kareler tahmin edicisine eşit olduğunu göstermiştir.

Shao (1996) çalışmasında regresyon analizinde yöntemin kullanılması sırasında yeniden örnekleme sürecinin, analiz öncesinde gözlemlerin yeniden örneklenmesi ya da analiz sonrasında hataların yeniden örneklenmesi biçiminde olacağını bildirmiştir.

Jeremy (1996) yeniden örnekleme yöntemlerinin tanımlı ve bağımsız veriler için kullanılabildiği gibi zaman serilerinde de uygulanabileceğini bildirmiştir.

Strawderman ve Wells (1997) birikimli tehlike (hazard) ve yaşam fonksiyonları için bootstrap güven sınırlarının geçerliliği üzerine çalışmışlar, rasgele olarak azalan veri için bootstrap yönteminin yararlı olduğunu belirterek, çeşitli ilaç türleri için yaşam olasılıklarına ilişkin bootstrap güven aralıklarını elde etmişlerdir.

Zeng ve Davidan (1993) çalışmalarında, bağışıklık sisteminde bootstrap güven aralıklarının kullanımı üzerinde durmuşlar, konu ile ilgili olarak birde simülasyon çalışmasına yer vermişlerdir.

Fox (1997) yaptığı çalışmada bootstrap yönteminin, hata terimlerinden tekrarlı örnekler seçerek, tahmin değerlerine $(\hat{Y}_i)$ eklendiği için hata terimlerinin normal dağılıma sahip olduğunu varsaymış, örneğe ait değerlerin deneysel dağılımını oluşturarak, örnekleme yöntemlerindeki bazı sakıncaların ortadan kalkmasını ve bootstraplama sürecinde izlenecek adımların belirlenmesi gerektiğini bildirmiştir.

Yeniden örnekleme yöntemleri ile ilgili bugüne dek yapılmış çalışmalar sürekli dağılımların parametrelerine ilişkin yeniden örnekleme yöntemlerinin nokta ve aralık tahminlerini içermektedir. Bu çalışmada yeniden örnekleme yöntemlerinin özellikle bootstrap ve jackknife yöntemlerinin, parametre tahmini ve güven aralıklarının oluşturulması için çoklu regresyon modellerinin analizinde yöntemlerin kullanılarak karşılaştırılması ele alınmıştır.

BÖLÜM II

ÖRNEKLEME YÖNTEMLERİ

Yeniden örnekleme yöntemlerinin daha iyi anlaşılması için bu yöntemlere temel oluşturan klasik örnekleme yöntemlerinin bazı özelliklerinin bilinmesinde fayda olacağı düşüncesinden yola çıkılarak aşağıdaki kriterler ele alınmıştır.

2.1. Verilerin Elde Edilmesi

İstatistik yöntem biliminde veriler tamsayım yapılarak elde edildiğinde populasyon hakkında kesin bilgiler elde etmek söz konusu olsa da,

a) Maliyet

b) Zaman

c) Eleman

d) Populasyonun belirsiz olması nedeniyle tüm birimlere ulaşmanın mümkün olmaması

e) Bazı durumlarda tamsayım yapılması halinde birimlerin yok olma tehlikesinin bulunması gibi çeşitli nedenlerle bu her zaman mümkün olmamaktadır (İşçil, 1977).

Bu durumda örnekleme yapma gereği doğmaktadır. Örnekleme "populasyon hakkında genel yargılara varmak amacıyla bir populasyondan o populasyonu temsil edebilecek daha az sayıda bir topluluğu seçme işlemidir". Örnek ise, "birbirine bağlı olmayan ve her biri aynı olasılık kurallarına bağlı n sayıdaki birimden oluşan topluluk " biçiminde tanımlanabilir.

Verilerin örnekleme yöntemi ile elde edilmesi söz konusu olduğunda populasyondan çekilen örneğin;

a) Temsil gücü yüksek olmalı,

b) Tesadüfen seçilmiş olmalı ve

c) Yeterli büyüklükte olmalı koşullarının yanı sıra başarılı bir örnekleme yapabilmek için ;

a) Örnek, çekilecek populasyon hakkında özel bilgiler bulundurmalıdır.

b) Seçme işlemi ilgilenilen özellik ve değişkenlerden bağımsız olmalıdır.

c) Örnekteki birimler birbirinden bağımsız olmalıdır.

d) Örneğe alınacak verilerin tümüne istisnasız aynı koşullar uygulanmalıdır.

e) Örnek herhangi bir ön yargıya yer vermeksizin ve sistematik bir farklılık yaratmayacak biçimde seçilmelidir.

f) Verilerin elde edildiği alanlar ile diğer alanlar arasında önemli farklılıklar bulunmamalıdır.
g) Alan çalışmaları ile ilgili olarak israfa yol açmamalıdır (Yoğurtçugil, 1976).

Araştırılması istenilen populasyondan belli sayıda elemana sahip olan örnek veya örneklerin oluşturulması için çeşitli örnekleme yöntemleri geliştirilmiştir. Bu yöntemlerden genellikle aşağıdaki örnekleme yöntemleri kullanılmaktadır (Bağırkan, 1980);

1) Basit tesadüfi örnekleme yöntemi
2) Sistematik örnekleme yöntemi
3) Tabakalı örnekleme yöntemi
4) Küme örnekleme yöntemi
5) Monografi örnekleme yöntemi
6) Kararlı örnekleme yöntemi
7) Kota örneklemesi yöntemi
8) Latin kare örnekleme yöntemi

2.2. Klasik Örnekleme Yöntemlerinin Avantajları

Örnekleme yöntemlerinin avantajlarını şu şekilde sıralaya biliriz (Çil, 2000)

a) Zamandan tasarruf
b) Maliyetten tasarruf
c) Daha az elaman ile sonuca varmak
d) Populasyondaki bazı birimlere ulaşmanın imkansız olması
e) Bazı araştırmalarda örneklemenin tek seçenek olması

2.3. İstatistiksel Yöntemlerin Amacı ve Klasik örnekleme Dağılımları

İstatistiksel Yöntemlerin temel amacı; Bilinmeyen belirli bir $f(x;\theta)$ dağılımından n birimlik bir tesadüfi örnek seçilerek $\hat{\theta}$ değerini hesaplamak ve bunun da θ populasyon değeri (parametresinin) tahmin edicisi olarak kabul etmektir. $\hat{\theta}$ tahmin edicisi reel sayılar kümesinde bir noktadır, bunun için yapılan tahmin nokta tahmini olarak isimlendirilir. Tahmin edici, tahmin amacıyla kuramsal örnek değerlerini kullanacak bir fonksiyon ve bu fonksiyonun aldığı değere de tahmin denir.

Örnekleme dağılımları bir populasyondan çekilebilecek mümkün ve muhtemel tüm örneklerden yararlanılarak oluşturulan teorik dağılımlardır ve hipotez kontrolleri için gereklidir. N sayıdaki bir populasyondan iadeli (yerine koyarak) ve iadesiz (yerine koymadan) olarak seçilecek n sayıdaki mümkün ve muhtemel tüm örneklerin sayısı, iadesiz seçimlerde $C\binom{N}{n} = \frac{N!}{n!(N-n)!}$ kombinasyonu şeklinde olurken, iadeli seçimde bu sayı N^n 'e ulaşmaktadır (İşçil, 1975).

Bu dağılımlar aritmetik ortalamanın örnekleme dağılımı, standart sapma, değişim, oran ve ortanca için de oluşturulabilir. Ortalamaların örnekleme dağılımı bir populasyondan seçilebilecek tüm olası örneklerin ortalamalarının oluşturduğu bir dağılım olup ortalamaların her biri populasyon ortalamasının bir tahmin edicisidir. Bu dağılımın parametreleri $\mu_{\bar{X}}$ *ve* $\sigma_{\bar{X}}$ ' dir.

Örnekleme dağılımlarının önemi; tahmin değerleri ile populasyon parametreleri arasındaki ilişkilerin kurulması ve kavranmasına yardımcı kuramsal araçlar olmalarından kaynaklanmaktadır. Çünkü tahmin değerlerinden hareketle populasyon parametreleri hakkında genelleme yapmanın yolu örnekleme dağılımlarından geçmektedir. Ancak, bu tür örnekleme dağılımlarının oluşturulması oldukça zor olduğundan uygulamada bu örnekleme dağılımları oluşturulmaz (Gürsakal, 1998). Gerçek hayatta populasyondan tek bir örnek çekilmekte ve bu örnekten hareketle populasyon hakkında genelleme yapılmaktadır. İstatistiksel yöntemlerde örnekleme dağılımları oluşturulmaksızın genelleme yapmanın yolu merkezi limit teoreminden geçmektedir.

Merkezi limit teoremi gereğince, ortalaması μ ve varyansı σ^2 olan herhangi bir populasyondan iadeli olarak seçilen n hacimlik örneklerin ortalaması, ortalaması μ ve varyansı $\frac{\sigma^2}{n}$ olan normal dağılım göstermektedir. Bu nedenle de tahmin edicilere ilişkin örnekleme hatalarını ortaya koymak amacıyla örnekleme dağılımını oluşturmak gibi son derece zor, neredeyse olanaksız bir çabaya girişmek yerine, bunu populasyon varyansından hareketle ortaya koymanın mümkün olduğu ileri sürülmektedir (Serper, 1996).

2.4. Parametre Tahmin Edicilerinin Özellikleri

İyi bir nokta tahmin edicisinin sahip olması gereken veya arzu edilen bazı özellikleri aşağıda verilmektedir:

1) Sapmasızlık (sistematik hatasızlık, eğilimsizlik)

2) En küçük varyans

3) Etkinlik

4) Ortalama kare hata

5) Yeterlilik,

6) Tutarlılıktır.

2.4.1. Sapmasızlık

Bir populasyondan çekilen çok sayıdaki örnek için elde edilen parametre tahmin değerleri birbirinden aynı zamanda populasyon parametrelerinden farklıdır. Bir başka değişle, nokta tahmin edicisinin populasyon parametresine eşit olması hemen hemen olanaksızdır ve mutlaka bir tahmin hatası söz konusu olacaktır. Ancak, tahmin edicinin sapmasızlığı ile anlatılmak istenen tahmin hatalarının cebirsel toplamlarının sıfıra eşit olmasıdır. Bu durum nokta tahmin edicisinin sapmasızlığını anlatmaktadır.

θ parametresinin tahmin edicisi $\hat{\theta}$ olsun. Eğer bir tahmin edicinin beklenen değeri, populasyon parametresine eşit $(E(\hat{\theta}) = \theta)$ ise $\hat{\theta}$, θ 'nın sapmasız (tarafsız) bir tahmin edicisidir denir (Çakır, 2000).

2.4.2. En küçük varyans

Bir tahmin edici, başka bir tahmin yöntemiyle bulunmuş sapmalı veya sapmasız başka bir tahmin ediciye göre, en küçük varyansa sahipse, böyle tahmin edicilere iyi bir tahmin edici denilebilir. Bu durum aşağıdaki gibi ifade edilebilir (Tarı, 1999).

$$E\left[\hat{\theta} - E(\hat{\theta})\right]^2 < E\left[\tilde{\theta} - E(\tilde{\theta})\right]^2 \qquad (1)$$

veya daha kısa olarak, $Var(\hat{\theta}) < Var(\tilde{\theta})$

Burada $\tilde{\theta}$, populasyonun gerçek parametresi θ 'nin sapmalı veya sapmasız herhangi bir başka tahmin edicisidir.

2.4.3. Etkinlik

Bir populasyon parametresinin tahmini için aynı anda birden fazla sapmasız ve tutarlı tahmin ediciler belirlenmiş olabilir. Bu tahmin ediciler içinde hangisinin seçileceği sorusunun cevabı iyi bir tahmin edicide bulunması gereken "etkinlik" özelliği ile açıklanmaktadır(İşyar, 1999).

Sapmasız ve tutarlı tahmin edicilerden daha küçük varyansa sahip olan $\hat{\theta}$ tahmin edicisi etkin olarak nitelendirilir.

2.4.4. Ortalama kare hata (OKH)

Sapmasızlık ve en küçük varyans özelliklerinin bir arada bulunmadığı durumlarda seçim, ortalama kare hata kriterine bakılarak yapılabilir. Bu kriter, sapmasızlık ve en küçük varyans özelliklerinin bir bileşimidir. Bir tahmin edici en küçük ortalama kare hataya sahipse, en küçük ortalama kare hata tahmin edici olur. Ortalama kare hata, tahmin edicinin, populasyon parametresi(θ) ile olan farklarının karelerinin beklenen değeri olarak ifade edilir.

$$OKH(\hat{\theta}) = E(\hat{\theta}-\theta)^2 \tag{2}$$

Ortalama kare hatası(OKH), sapmalı(yanlı) tahmin ediciyi sapmasız tahmin edici ile karşılaştırmada faydalıdır; çünkü ortalama kare hata;

$$E\left[(\hat{\theta}-\theta)^2\right] = E\left[(\hat{\theta}-E(\hat{\theta}))-(\theta-E(\hat{\theta}))\right]^2$$

$$=E\left\{\left[\hat{\theta}-E(\hat{\theta})\right]^2 - 2\left[\hat{\theta}-E(\hat{\theta})\right]\left[\theta-E(\hat{\theta})\right]+\left[\theta-E(\hat{\theta})\right]^2\right\}$$

elde edilir. Burada ortadaki terim sıfır olduğundan OKH($\hat{\theta}$)

$$E\left[(\hat{\theta}-\theta)^2\right] = E\left[\hat{\theta}-E(\hat{\theta})\right]^2 + \left[\theta-E(\hat{\theta})\right]^2$$

olur. Burada, sağ taraftaki ilk terim, $\hat{\theta}$ tahmin edicisinin varyansıdır. İkinci terim ise $\hat{\theta}$ nın dağılımının beklenen değeri yani ortalaması ile gerçek populasyon parametresi θ arasındaki farktır, yani bir tahmin edici olarak $\hat{\theta}$'in sapmasıdır. Sapma b($\hat{\theta}$) ile gösterilirse,

$$b(\hat{\theta}) = \theta - E(\hat{\theta}) \tag{3}$$

şeklinde tanımlanır. Böylece OKH($\hat{\theta}$);

$$OKH(\hat{\theta}) = Var(\hat{\theta}) + \left[b(\hat{\theta})\right]^2 \tag{4}$$

olur (Korum, 1971). Ortalama kare hata, karesi negatif olamayan iki terimin toplamıdır ve varyans ve sapmaya bağlı olarak değişecektir. Dolayısıyla tahmin edici sapmasız olduğunda ortalama kare hata yalnızca tahmin edicinin varyansına eşit olacaktır. Ancak uygulamada her zaman sapmasız tahmin ediciler bulmak mümkün olmamaktadır. Bu durumda sapmalı tahmin ediciler arasından sapması en küçük olanın seçilmesi önerilmektedir.

Sapmasız fakat büyük varyanslı ile sapmalı fakat küçük varyanslı tahmin ediciler arasında tercih yapılırken, OKH' sı en küçük olan tahmin edici seçilir. Ancak modelin amacı bağımlı değişken için yapılan tahminlere kesinlik kazandırmak ise, en düşük varyanslı fakat sapmalı tahmin edici tercih edilebilir.

2.4.5. Yeterlilik

Yeterlilik özelliği, tahmin edicinin, tahmin edilecek parametre hakkında örnekte mevcut bulunan bütün bilgiyi kullanmasıdır. Bu anlamda, örnekteki tüm bilgiyi kullanan tahmin edici yeterli tahmin edici olmaktadır. Bu doğrultuda örneğin, aritmetik ortalama, n hacimlik örnekteki tüm bilgiyi kullandığı için yeterli tahmin edici özelliğine sahip olurken mod veya medyan gibi diğer tahmin ediciler örnekteki tüm bilgiyi kullanmadıkları için, populasyon ortalamasının yeterli bir tahmin edicisi değillerdir (Çakır, 2000).

2.4.6. Tutarlılık (kararlılık) özelliği

Eğer bir $\hat{\theta}$ tahmin edicisi sapmasızsa ve örnek büyüklüğü sonsuza giderken varyansı sıfıra yaklaşırsa bu tahmin edici tutarlıdır. Bu özellik de matematiksel olarak,

$$\lim_{n\to\infty} P(\left|\hat{\theta}-\theta\right| \leq a)=1 \quad (5)$$

$$\lim_{n\to\infty} Var(\hat{\theta})=0 \quad (6)$$

ise, "nokta tahmin edicisi tutarlıdır" şeklinde ifade edilir. Diğer bir ifadeyle büyüklüğü, arttıkça hem sapma ve hem de varyans azalır ve n birim sayısı sonsuza yaklaşırken varyans sıfir olursa böyle bir tahmin edici tutarlıdır denir.

2.5. Örnekleme Hataları

Daha öncede belirtildiği gibi, istatistiğin en önemli problemlerinden biri örnek değerlerinden hareketle bilinmeyen populasyon parametrelerinin tahmin edilmesidir. Söz konusu populasyon parametrelerinin tahmininde, N hacimlik populasyon birimleri arasından seçilen sadece n hacimlik bir örnekten yararlanıldığı için, tahmin değerlerinin populasyon parametresine eşit çıkması hemen hemen mümkün olamamaktadır.

Tahmin edici ($\hat{\theta}$) ile populasyon parametreleri (θ) arasındaki bu fark, tahmin hatası olarak adlandırılmaktadır. Verilerin sınıflanması ve onlara ilişkin tabloların düzenlenmesi aşamasında ortaya çıkan "adi hatalar" bir yana bırakılırsa örneklemede iki önemli hata ile karşılaşılmaktadır (Serper ve Aytaç, 1999);

1)Tesadüfi hatalar

2)Sistematik hatalar

dır. Böylece örnekleme sonucunda oluşan toplam hata,

$$Toplam\ hata = \sqrt{(Sistematik\ hata)^2 + (Tesadüfi\ hata)^2} \quad (7)$$

olur.

Toplam hatanın küçülmesi; Sistematik veya tesadüfi hatanın küçülmesi ile yada ikisinin birden küçültülmesiyle mümkündür.

2.5.1. Tesadüfi hatalar

Bir tahmin sistematik hata (sapma) içermese bile mutlaka bir örnekleme hatası içerecektir. Örnekleme hatasını kontrol etmek ve belli olasılıklarla belli sınırlar içinde tutmak mümkündür. Bu amaçla örnekleme hatalarının ortalamasının hesaplanması gerekmektedir. Ancak, bu hatalar iki yönlü ve denkleşen türden olduklarından, bunların aritmetik ortalaması daima sıfır çıkmakta ve bu nedenle kareli ortalama kullanılma gereği doğmaktadır. Elde edilen değer ise "tahmin değerlerinin standart hatası" olarak adlandırılmaktadır (Serper, 1996). Bir tahmin edicinin standart hatasının küçük çıkması istenir. Çünkü, standart hatanın küçük çıkması durumunda ;

a) Tekrarlandığı takdirde, örnekten örneğe elde edilecek tahmin değerleri de küçük çıkacaktır.

b) Örnekten bulunan değer ile aynı koşul ve kurallarda uygulanacak tam sayım sonucu elde edilecek değer arasındaki fark da küçük çıkacaktır.

c) Sonuçlar örnek değişkenliğine bağlı koşullardan daha az etkilenecektir (Yoğurtçugil, 1976).

Bir populasyondan çok sayıda n hacimlik örnek seçtiğimiz ve bunların mesela aritmetik ortalamalarını hesapladığımız zaman ortalamaların birbirlerinden ve populasyon ortalamasından farklı olduğunu görmekteyiz. Bu farklar veya hatalara "Tesadüfi hata" (örnek hatası) adı verilir.

Tesadüfi hatalar, her iki yönde etkili olduklarından, birbirlerinin etkilerini yok edebilirler. Bu özelliğin bir sonucu olarak, bir tahmin edici populasyon parametresinden, ya da aynı populasyondan çekilmiş iki ayrı tahmin edici birbirinden farklı olsa bile, örnekleme bölünmesinin ortalaması populasyon parametresine eşit olur. Tesadüfi hatalar örnek hacmi büyüdükçe büyük sayılar kanunu gereğince küçülürler (Aytaç, 1999).

Eğer tahmin edici örnek ortalaması ise $\hat{\theta}_n = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ olacaktır. Bu ortalamanın varyansı ise $Var(\overline{X}) = \frac{1}{n} Var(X_i)$ olacaktır. Buradaki $Var(X_i)$ ise uygulamada örnek varyansı tarafından tahmin edilmektedir. Bu nedenle de standart hatanın gerçek değerine ulaşmak yerine ancak onun tahmini elde edilmiş olmaktadır.

2.5.2. Sistematik hatalar (sapma)

Daha öncede belirtildiği gibi iyi bir nokta tahmini sapmasız olmalıdır. N hacimli bir populasyondan n hacimlik tüm mümkün ve muhtemel örnekler çekildiğinde , bu örneklerin tümü için hesaplanan değerlerin kuramsal dağılımının beklenen değeri populasyon parametresine eşit ise tahmin sapmasız sistematik hata içermeyen olacaktır (Yoğurtçugil, 1976). Aksi halde örneklemenin bir sistematik hata içerdiği söylenecektir.

Örnek hacminden bağımsız olarak ortaya çıkan sistematik hata, tahmin edicinin populasyon parametresinden tek yönde sapma göstermesi biçiminde ortaya çıkmakta ve bu tür bir hata durumunda, o populasyondan elde edilecek tahmin ediciler populasyon parametresinden ya her zaman büyük ya da her zaman küçük olmaktadır. Populasyondan çekilen birimlerin tesadüfi seçimle elde edilmeleri söz konusu olduğundan sistematik hatanın bulunmaması gerekirse de bu tür hatalar (Serper ve Aytaç, 1999);

a) Örneğin seçimi sırasında kullanılan tekniğin yanlış olmasından,

b) Soruların yanlış anlamaya yol açacak biçimde sorulmuş olmasından,

c) Örnekteki bazı birimler ile ilgili bilgi sağlanamaması nedeniyle, kayıp verilerin ortaya çıkmasından

d) Populasyonun iyi tanımlanamaması nedeniyle, bazı birimlerin gözlenememesi yada bazılarının birden fazla gözleme tabi tutulmasından,

e) Tahmin edicilerden hareketle populasyon parametresinin tahmin edilmesi sırasında kullanılan tekniğin yanlış olmasından kaynaklanabilir.

Sistematik hatanın ortadan kaldırılmasının tek yolu, hataya neden olan sorunun saptanarak giderilmesidir. Bu tür hataların örnek hacmi ile ilgisi bulunmadığından örnek hacminin arttırılması sistematik hatayı azaltmayacaktır.

2.6. Klasik Örnekleme Yöntemlerinin Zayıf Yönleri

Yukarıdaki koşullara sahip klasik örnekleme yöntemlerinin zaman içinde birtakım zayıflık ve dezavantajları olduğu ortaya çıkmıştır. Bu zayıflıklar aşağıdaki biçimde özetlene bilir (Sümbülloğlu ve Sümbülloğlu, 1987)

a) Örnekleme seçimini bilinçsiz olarak yapma, yanlış örnekleme yöntemi uygulama, denekleri seçerken hata yapma,

b) Örneğin yeterli sayıda bireyi içermemesi (örnek büyüklüğünün yetersiz olması)

c) Bilerek yada bilmeyerek örnek seçiminde taraf tutma,

d) Veri toplamada hata yapma, eksik, yanlış yada yararsız veri toplama, verilerin denetimini gereği gibi yapamama, analizde hatalar yapma,

e) Deneklerden yanıt almada başarısızlığa uğrama,

f) Bulunamayan bir deneğin yerine hemen başka birini yedek denek olarak kullanma.

Klasik örnekleme yöntemlerinin yukarıda sayılan zayıflıkları giderebilmek için yeniden örnekleme yöntemlerinden yararlanılmaktadır. Bu yöntemler bölüm ikide verilmiştir.

BÖLÜM III

MATERYAL ve YÖNTEM

3.1. Materyal

Bu çalışmada Kayseri Doğum Hastanesinden elde edilen 2000-2001 yıllarındaki 320 hastaya ait anne yaşı (X_1), gebelik süresi (X_2) ve bebeğin doğum ağırlıklarına (Y) ilişkin veriler kullanılmıştır. Bu verilere ait EKK regresyon doğrusu ile yeniden örnekleme yöntemleri (Boostrap ve Jackknife yöntemleri) uygulanarak elde edilen EKK regresyon doğrusuna ait parametre tahminleri karşılaştırılmıştır.

Bu verilerin analizi için EXCEL 2000, S-PLUS 2000 for WINDOWS ve SPSS for WINDOWS versiyon 10.01 paket programları kullanılmıştır.

3.2. Yöntem

Bu bölümde önce yeniden örnekleme yöntemleri genel olarak ele alınacak daha sonra her bir yöntemin özellikleri ayrı ayrı irdelenecektir.

Bütün bilim dallarında yapılan araştırmaların amacı elde edilen örneklerden yararlanarak üzerinde çalışılan populasyonun bilinmeyen parametreleri hakkında sonuç çıkarmaktır. Sonuç çıkarma iki şekilde yapılabilir:

1) Populasyonu temsil edecek parametrelerin tahmini ve,
2) Populasyon parametreleri ile ilgili hipotezlerin test edilmesi şeklindedir.

Hem yeniden örnekleme yöntemleri hem de örnekleme yöntemleri yukarıdaki amaçlara sahiptir. Ancak, iki yöntem arasındaki temel fark örnekleme dağılımının nasıl elde edildiği konusundadır.

Herhangi bir tahmin ediciye ait örnekleme dağılımının şekli örnek genişliğine, örneklerin çekildiği populasyonun şekline ve parametrelerine bağlıdır. Bir tahmin edicinin ($\hat{\theta}$) örnekleme dağılımı, n hacimlik tüm örneklerin tahmin edicilerinin dağılımı demektir. Genellikle istatistiksel yöntemler tahmin edicilerin örnekleme dağılımının belli koşullarda normal dağılıma sahip olduğunu gösteren merkezi limit teoremine dayanır.

Örnekleme yöntemlerine parametrik olmayan bir yaklaşım sağlayan yeniden örnekleme yöntemlerinin ne olduğunu ve örnekleme yöntemlerinden farkını anlamak için örnekleme dağılımları ve bunların özellikleriyle ilgili düşüncelere temel oluşturan tesadüfi gözlemler kavramından hareket etmekte fayda vardır.

Örnekleme dağılımı; bir populasyondan seçilen örneklerden elde edilen parametre tahmin değerlerinden oluşturulmuş dağılım olarak tanımlanır.

Bir populasyondan alınan rasgele örneklerinin her birisi için tahmin değerleri hesaplandığında örnek dağılımları ortaya çıkar. Mesela; her bir örneğin ortalaması hesaplanmışsa elde edilen $\bar{x}_i$ dağılımı ortalamaların örnek dağılımıdır. Aynı şekilde, her örnek için p oranları hesaplandığında oranların örnek dağılımı elde edilir.

Bir örnekleme dağılımı, bilinen veya varsayım gerektirmeyen bir populasyon dağılımına dayanılarak meydana getirilir. Diğer taraftan aynı populasyona dayanılarak, her biri belirli bir örnek büyüklüğü için hesaplanmış tahmin edicinin sonsuz sayıda örnekleme dağılımı oluşturulabilir. Ayrıca aynı populasyondan iki veya daha çok sayıda farklı tahmin edici için örnekleme dağılımı oluşturulabilir.

Örnekleme yöntemlerinin zayıflıklarını ve dezavantajlarını ortadan kaldırmaya yönelik örneklemede alt örnek kullanımı düşüncesi populasyon hakkında gerçekçi olmayan varsayımlar yapmak yerine eldeki örnekten populasyonun özellikleri(sapma ve standart hata tahminleri) hakkında daha iyi sonuçlar elde etme amacına yöneliktir. Yeniden örnekleme yöntemleri yeni yöntemler değildir. Uzun bir geçmişe sahiptir ve Parametrik olmayan en önemli yeniden örnekleme yöntemlerinden olan bootstrap ve jackknife yöntemleri, tahmin edicinin varyansını birtakım model ve karmaşık formüllere gerek olmaksızın elde edilmesini sağlar.

Bootstrap, bu tür tahminler yapmak amacıyla kullanılan duyarlı hesaplamalar içeren parametrik olmayan bir yöntemdir ve diğer örnekleme yöntemlerinden farklıdır. Bu yöntem, tahmin edicilerin örnekleme dağılımını oluşturmak için güçlü varsayımlar ve analitik formüller yerine çok sayıda tekrarlanmış hesaplamalar içerir. Bootstrap, bir tahmin edicinin örnekleme dağılımını oluşturmak amacıyla verilerin çok sayıda yeniden örneklenmesi amacı taşır (Money, 1993). Örnekleme yöntemleri örnekleme dağılımının oluşturulması ile ilgili olarak tahmin edicilerin $(\hat{\theta})$ dağılım biçimi konusunda bazı varsayımlar kullanırken, bootstrap yöntemi örnek verilerini populasyon verileri gibi düşündüğünden varsayıma gerek duymaz. Böylece bu yöntemler, araştırmacının çözümlenemez ve savunulamaz varsayımların bulunduğu durumlarda tahmin yapmasına izin verir. Bununla birlikte bootstrap, kendi başına bir istatistik değildir. Daha doğrusu populasyon parametreleri hakkında daha doğru tahmin yapmak amacıyla kullanılır. Bootstrap, örnek ve populasyon arasındaki benzerliğe güvenmektedir.

Yeniden örnekleme yöntemleri orijinal örneğe herhangi bir bilgi eklemez (Şengün, 1999). Bundan dolayı bootstrap gibi yeniden örnekleme yöntemleri örnek bilgisini yönlendiren yöntemler olarak adlandırılmaktadır.

Yeniden örnekleme yöntemlerinde, olasılıkların ve yeniden örnekleme olasılık vektörünün önemi büyüktür. X in x_i değerini alması olasılığı $P(X=x_i)=f(x_i)$ ile gösterilirse o halde $P(x_i)=P^*(X=x_i)$ dir. Böylece X'in alabileceği tüm değerler için $f(x_0)+f(x_1)+....+f(x_n)=1$ dir.

X, sonlu sayıdaki $x_1, x_2, ..., x_n$ değerlerini alabilen kesikli tesadüfi değişken ve buna karşılık gelen olasılıklar,

$$P(x_i)=P^*(X=x_i)=p^*(x_i)=p_i^*, \qquad i=1,2,3,.....n \tag{8}$$

olsun. Bu takdirde aşağıdaki koşulları sağlayan P(x) fonksiyonuna X'in olasılık fonksiyonu denir.

Burada,

1) $p_i^* \geq 0$ (9)

2) $\sum_{i=1}^{n} p_i^* = 1$ (10)

dir. Buradaki, p_i^* : n hacimlik bir örnekte $x_1, x_2, ..., x_n$ gözlemleri için yeniden örnekleme olasılık vektörü olup $p_i^* = (p_1^*, p_2^*, p_3^*, ..., p_n^*)$ dır.

3.2.1. Bootstrap Yöntemi

Bootstrap yöntemi 1979 yılında Efron tarafindan bir yeniden örnekleme yöntemi olarak ileri sürülmüş ve bootstrap tahmin edicilerini kullanarak bir tahmin edicinin dağılımını tahmin etmek amacıyla geliştirilmiştir (Bülbül ve Altaş, 1998).

Yöntem, bir başka yeniden örnekleme yöntemi olan jackknife yöntemine alternatif olarak ve söz konusu yöntemden daha kolay uygulanabilir ve çok daha güvenilir olduğu belirtilerek ileri sürülmüştür (Efron, 1979). Yöntemin önemi, gözlenen örnek verilerinden hareketle, tahminin standart hatasını minimuma indirerek populasyon parametrelerinin tahmin edildiği istatistiksel yöntemler aşamasında ortaya çıkmaktadır.

Literatüre bakıldığı zaman birkaç farklı bootstrap'a rastlanılabilir. Bunlar, basit bootstrap, çift bootstrap, ağırlıklı bootstrap, tekrarlamalı bootstrap, doğal (wild) bootstrap, ardışık bootstrap, ve daha bir çoğu sayılabilmektedir (Şengün, 1999). Bu çalışmada basit bootstrap örnekleme yöntemi açıklanmıştır.

Örneğe ait verilerin yeniden örneklenmesi mantığına dayanan bu yöntemin örnekleme dağılımının bulunuşunda parametrik ve parametrik olmayan yöntemler kullanılabilmektedir. Parametrik olmayan yöntemle bootstrap dağılımının tespitinde, bilinmeyen dağılım yerine istatistiksel fonksiyonun tanımına göre gözlemlerin deneysel dağılımı konularak işe başlanır. Ardından Monte Carlo simülasyonu ile tekrar tekrar değer türetilerek yaklaştırma yapılır. Parametrik yöntemde ise başlangıçta deneysel dağılım yerine örneğin elde edildiği populasyonun dağılımı konusunda bilgi varsa, örneğin normal dağıldığı biliniyorsa bu durumda "Parametrik Bootstrap'' söz konusu olur (Atalay ve İnal, 1999).

Populasyon parametresinin tahmin edicisi olan $\hat{\theta}$ 'nın örnekleme dağılımının oluşturulmasının amacı, söz konusu populasyon parametresinin tahmin edilmesi ya da test edilmesidir. Ancak, teorik olarak mümkün olan bu yöntemin uygulanabilirliği konusunda kuşkular bulunmaktadır. Tahmin edicinin örnekleme dağılımını oluşturmak imkansız olmasa da son derece güç ve zaman alıcı bir iştir. Ancak, tahmin edicinin deneysel örnekleme dağılımını oluşturmak amacıyla ortaya atılan bootstrap yöntemi bu sakıncayı ortadan kaldırmaktadır. Bu mantık doğrultusunda gerekli bootstrap algoritması, aşağıdaki biçimde tanımlanabilir (Fox, 1997):

1) Populasyondan n hacimlik bir örneğin elde edilmesi;

populasyon $\Rightarrow$ örnek

bu örnek kullanılarak populasyon parametresinin tahmin edicisinin hesaplanması

2) Elde edilen bu örnek populasyon ile ilgili başka hiçbir bilgi olmadığından bu populasyonun tek en iyi tahmin edicisi kabul edilir. Bu nedenle bu örnek populasyon gibi kabul edilerek her defasında iadeli seçimle her bir gözlemin örneğe girme olasılığını 1/n alarak n hacimlik bir örneğin yeniden elde edilmesi ve bu sürecin B kez tekrarlanması;

$$\text{Örnek} \Rightarrow \begin{cases} \Rightarrow & \text{1. Bootstrap örneği} \\ \Rightarrow & \text{2. Bootstrap örneği} \\ & . \\ & . \\ \Rightarrow & \text{B. Bootstrap örneği} \end{cases}$$

3) Her bootstrap örneği için ilgilenilen tahmin edicinin hesaplanması,

4) B sayıda örnekten hareketle bu tahmin edicilerin örnekleme dağılımının elde edilmesi.

5) Elde edilen bu dağılımdan, dağılımla ilgili ortalama, standart sapma ve standart hata gibi önemli tahmin ediciler ile parametre tahmin değerlerinin elde edilmesi,

6) Sonuçta bu tahminler kullanılarak populasyon hakkında yorumların yapılması.

Yukarıdaki algoritma bootstrap yönteminin mantığını genel olarak açıklayan bir algoritmadır. Yeniden örnekleme sayısı olan B, uygulamaya bağlıdır. Aslında n hacimlik bir örnekten teorik olarak n^n sayıda bootstrap örneği oluşturmak mümkünse de bu hem gereksizdir hem de zaman kaybına neden olmaktadır (Stine, 1990). Ancak, bootstrap yöntemi için gerekli tüm bootstrap örnekleri bilgisayar teknolojisindeki gelişmeye paralel olarak gerçekleştirilebilmektedir. Bununla birlikte ortalama, standart hata vb. her tahmin edici için farklı büyüklükte bootstrap örnekleri oluşturulabilmektedir (Leger, 1992).

Efron 1979, yukarıda söz konusu olan sürecin devreye girmesiyle, örnek hacmini arttırmaksızın populasyon parametresi ile tahmin edici arasındaki sapmanın azaltılacağını ve teorik olarak elde edilmesi mümkün gibi görünen ama uygulama da söz konusu olamayan tahmin edicilerin örnekleme dağılımının oluşturulabileceğini ileri sürmektedir. Örnekleme dağılımının bu şekilde oluşturulması tahmin edicinin standart hatasının daha sağlıklı olarak elde edilmesini sağlayacaktır.

3.2.1.1. Bootstrap dağılımının sapma tahmini

Eğer bir tahmin edici sapmalıysa ve bu sapma bilinmiyorsa, populasyon parametresi ile ilgili kesin tahminde bulunmak mümkün olmaz. Bazı tahmin ediciler sapmasızdır. Örneğin, örnek ortalaması, populasyon ortalamasının sapmasız bir tahmin edicisidir.

Ancak, sapmasızlık varsayımının sağlanamadığı birçok durum vardır. Çoğu parametrelerin tahmin edicilerinin sapmalı olduğu bilinmektedir. İki örnek ortalamasının oranı buna örnektir (Money, 1993). Bu tür sapmalı tahmin edicilerle çalışmak bilimsel birtakım sakıncalar taşımakla birlikte örnekleme yöntemleri bu konuya açıklık getirmek ve bu sorunu çözmek konusunda yetersiz kalmaktadır.

Bootstrap yöntemi, örnekleme yönteminin bu yetersizliklerini ortadan kaldırmak amacıyla hem tahmin edicilerin örnekleme dağılımının oluşturulması hem de sapmanın azaltılması amacına yönelik parametrik olmayan bir yöntemdir. Bootstrap sapma tahmini, Eşitlik (3)' e benzer şekilde $\hat{\theta}$ ve bootstrap örnekleme dağılımının beklenen değeri arasındaki farktır:

$$b^*(\hat{\theta}) = \hat{\theta} - E(\hat{\theta}^*)$$

$$b^*(\hat{\theta}) = \hat{\theta} - \bar{\hat{\theta}}_B^* \qquad (11)$$

biçiminde elde edilmektedir. Burada,

$\hat{\theta}$: Parametre tahmin edicisini,

$b^*(\hat{\theta})$: Bootstrap sapma tahminini,

$E(\hat{\theta}^*)$: Bootstrap tahmin edicisinin beklenen değerini,

$\bar{\hat{\theta}}_B^*$: Bootstrap tahmin edicilerinin aritmetik ortalaması olup,

$$\bar{\hat{\theta}}_B^* = \frac{\sum_{b=1}^{B} \hat{\theta}_b^*}{B} \qquad (12)$$

dir. Burada da,

$\hat{\theta}_b^*$: b' inci bootstrap tahmin edicisini

B: Bootstrap tekrar(kopya) sayısını göstermektedir (Birich, 1995).

Eşitlik (11)'de de açıkça görüldüğü gibi buradaki sapma tahmini, nokta tahmin edicisi ile bootstrap nokta tahmin edicisi arasındaki farktır.

Efron $b^*(\hat{\theta})$ değerinin, tahmin edici ile populasyon değeri arasındaki farktan daha küçük olduğunu ve bu tür bir yaklaşımla hem sapmanın tahmin edilmesinin hem de bunun azaltılmasının olanaklı olduğunu belirtmiştir (Efron, 1990).

3.2.1.2. Bootstrap dağılımının standart hata tahmini

Populasyon normal bir dağılım gösteriyorsa, standart sapması da belli ise, örneğin büyüklüğü ne olursa olsun, örnek aritmetik ortalamasının olasılık dağılımı normal dağılımdır. Populasyon normal dağılım göstermemekle birlikte örnek yeterince büyükse, merkezi limit teoremi gereğince örnek ortalamasının dağılımı yine normal olacaktır (Aytaç, 1991). Uygulamada populasyon standart sapmasının bilinmesi (tamsayım dışında) olanağı yoktur. Bu nedenle, örneğin standart hatası,

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} \qquad (13)$$

formülü ile hesaplanır. Burada:

s_x : Örneğin standart sapma değeri

$s_{\bar{x}}$: Örnek ortalamasının Standart hatası

n: Örnek hacmini göstermektedir.

Bu değere, ortalamaya ait örnekleme dağılımının örnekten tahmin edilen standart sapması veya kısaca ortalamanın standart hatası denir. Örneklerin çekildiği populasyon sağa veya sola çarpık bir dağılım gösteriyorsa ortalamaya ait örnek dağılımının şekli tesadüfen çekilen örneklerin genişliğine bağlıdır. Normal dağılım dışındaki dağılımlardan çekilen örneklerin genişliği arttıkça ortalamaya ait örnek dağılımı normal dağılıma yaklaşır. Ancak tahmin edilen standart hatanın gerçek değerinin ne ölçüde yansıttığı konusunda kesin bir ölçü bulunmamaktadır. Bu düşünceden yola çıkan bootstrap yöntemi, tahmin edicilerin dağılımının standart sapmasının oluşturulacak bootstrap dağılımından hareketle tahmin edilmesinin daha sağlıklı olacağını savunmaktadır (Efron, 1981)

Bootstrap tekrar sayısının büyüklüğü çalışılan veri üzerindeki deneyime bağlıdır. Bootstrap tekrar sayısının genellikle $\hat{\theta}$ ' nın standart sapmasını tahmin etmek için 50-200 olması, θ ' nın güven aralıklarının tahmini içinde en azından 1000 olması önerilmektedir. Diğer taraftan, 10 ile 80 arasındaki örnek büyüklüğünde, tekrar sayısının 100 olması gerektiğine ilişkin görüşlerde vardır (Efron, 1990).

Standart hatanın bootstrap tahmini, istatistiğin bootstrap tekrarlarının standart hatasıdır ve,

$$s_B^* = \sqrt{\frac{\sum_{b=1}^{B} (\hat{\theta}_b^* - \bar{\hat{\theta}}_B^*)^2}{B-1}} \qquad (14)$$

eşitliği ile hesaplanır. Buradaki terimler Eşitlik 11 ve 12'deki terimler olup,

s_B^* : Standart hatanın Bootstrap tahmin değeri

B: Bootstrap tekrar sayısını göstermektedir.

3.2.1.3. Bootstrap güven aralıkları

Populasyon parametresi ile ilgili güven aralıklarının oluşturulması amacıyla tahmin edicinin örnekleme dağılımının tahmin edilmesi gerekmektedir. Örnekleme yönteminde $\hat{F}(\theta)$ 'nın normal ya da t dağılımına sahip olduğu varsayımı yapılmaktadır. Bu varsayım doğrultusunda populasyon parametresinin 1-α anlamlılık düzeyinde güven sınırları oluşturulmaktadır. Populasyon varyansı biliniyorken, θ için önerilen güven aralığı (n>30);

$$p\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \theta < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \qquad (15)$$

eşitliği ile elde edilir. Bu aralık, $\hat{\theta}$' nın örnekleme dağılımının şekli konusundaki varsayım geçerli olduğunda % 100(1-α) anlamlılık düzeyinde populasyon parametresini içerecektir. Ancak, söz konusu varsayım her zaman doğru olmayabilir. Bu durumda örnekleme yöntemleri geçerliliğini yitirecektir (Money, 1993).

Bootstrap güven aralıkları, örnekleme yöntemlerindeki yukarıda sözü edilen zayıf yönleri gidermek amacı taşımaktadır(Hall, 1986).

Bootstrap güven aralıkları;

a) Normal Yaklaşım Yöntemi

b) Yüzdelik Yöntemi

olmak üzere iki yöntemle elde edilir.

a) Normal Yaklaşım Yöntemi

Normal yaklaşım yöntemi güven aralıkları elde etmede kullanılan parametrik yöntemle büyük benzerlik gösteren bir yöntemdir. Bu yöntem, bir istatistik için normallik varsayımının gerçekleştiği ancak standart hatayı elde etmekle ilgili analitik formüllerin bulunmadığı durumda kullanılan bir yöntemdir.

Normal yaklaşım yöntemi ile bootstrap güven aralığı,

$$p\left(\hat{\theta}^* - z_{\alpha/2}s_B^* < \theta < \hat{\theta}^* + z_{\alpha/2}s_B^*\right) = 1-\alpha \qquad (16)$$

eşitliği ile elde edilir.

Bootstrap, parametrik olmayan bir yöntem olarak tasarlanmıştır ve normal yaklaşım güven aralıkları güçlü bir normallik varsayıma dayanır. Bu varsayımın sağlanamadığı durumlarda, normal yaklaşım yöntemiyle elde edilen güven aralıkları, parametrik yöntemle elde edilen güven aralıklarından daha iyi sonuçlar vermezler. Bu şekilde elde edilen aralıkların örnekleme yöntemine ait aralıklar ile çok büyük benzerlikleri vardır. Söz konusu aralık örnek değerlerinin normal dağıldığı durumlarda oldukça az (B=50-200) sayıda yeniden örnekleme sayısı gerektirdiğinden oldukça yararlıdır (Money, 1993).

Ancak güçlü parametrik varsayımların geçerliliğini yitirdiği durumlarda ne örnekleme yöntemi ile elde edilen güven aralıklarının ne de standart aralıkların oluşturulması söz konusu olmayacaktır. Bu durumda aşağıdaki parametrik olmayan bootstrap güven aralığının oluşturulması gerekmektedir.

b) Yüzdelik Yöntemi

Yüzdelik yönteminde temel yaklaşım oldukça basittir. Bu yaklaşımda $\hat{\theta}$ için elde edilen bootstrap kopyalarından oluşan örnekleme dağılımı kullanılarak 1-α güven düzeyinde dağılımın alt ve üst $\alpha/2$ yüzdeleri güven sınırlarını oluşturmaktadır.Bir başka değişle, alt sınır $\alpha/2$,üst sınır ise 1- $\alpha/2$ lik yüzdelere karşılık gelmektedir. Örneğin , populasyon parametresi için% 95 güven düzeyinde bir aralık oluşturulmak istendiğinde alt sınır bootstrap kümülatif (yığılmalı) örnekleme dağılımının % 2.5, üst sınır ise % 97.5 'una karşılık gelecektir (Atalay ve İnal, 1999).

Yüzdelik yöntemde örnekleme dağılımının parametrelerini tahmin etmek için karmaşık formüllere gerek yoktur. Yalnızca, bootstrap örnekleme dağılımının oluşturulması yeterlidir. Bu yönüyle güven aralıklarının oluşturulması oldukça kolaydır.
Ancak iki zayıf noktası bulunmaktadır;

1) Küçük örnekler de performansı düşüktür,

2) Yüzdelik yöntemde bootstrap örnekleme dağılımının simetrik olduğu varsayımı bulunmaktadır. Bu varsayımın gerçekleşmemesi durumunda sonuçlar geçerliliğini yitirmektedir.

3.2.2. Jackknife Yöntemi

Stine (1990)' a göre Jackknife Yöntemi ilk kez 1949 yılında Quenouilli tarafından ortaya atılmış, daha sonra 1958 yılında Tukey tarafından güven aralığı yaklaşımıyla geliştirilmiştir. Yöntem, populasyon parametrelerinin tahmin edilmesinde örnek hatasının minimuma indirilmesinde, tahmin edicinin sapmasının hesaplanmasında kullanılmasına ek olarak güçlü güven aralıkları oluşturulması amacına yönelik geliştirilmiştir.

Bootsrap gibi jackknife da, parametrik varsayımlar kullanmak yerine, örnek değişkenliğinin açıklanması yoluyla elde edilen tahmin edicilerin güvenilirliğini artırmaya yöneliktir. Ancak, jackknife örneğin değişkenliğini farklı bir yolla açıklamaya çalışan bir yöntemdir.

Bootsrap'ın sloganı "iadeli örnek" jackknife'ın sloganı ise "birini dışarıda bırak" tır. Bu şekilde yöntem aşağıdaki şekilde uygulanır.

Bilinmeyen bir F(x) olasılık dağılımından bağımsız ve genişliği n olan $X = (x_1, x_2, \ldots, x_n)$, örnek veri setinden, sırasıyla her defasında örnekteki verilerden biri dışarıda bırakılarak geriye kalan n-1 tane gözlemi olan n tane jackknife örneği oluşturulur.

$\hat{\theta}$ tahmin edicisinin oluşan jackknife örneklerinde aldığı değerler,

$\hat{\theta}_{(-j)} = \hat{\theta}(X_{(j)}) = \hat{\theta}(X_1, X_2, X_3, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n)$, $j = 1,2,3,\ldots,n$ elde edilir. Bu değerler kullanılarak

$$\bar{\hat{\theta}}_{(-j)} = \frac{\sum_{j=1}^{n} \hat{\theta}_{(-j)}}{n} \qquad (17)$$

değeri hesaplanır. Burada;

$\hat{\theta}_{(-j)}$: j'nci gözlem çıkarıldıktan sonra geriye kalan n-1 gözlemden hesaplanan tahmin değerini(j'nci jackknife tahmin değeri),

$\bar{\hat{\theta}}_{(-j)}$: n tane jackknife tahmin değerlerinin ortalamasını göstermektedir.

3.2.2.1. Jackknife sapma tahmini

Daha önce de belirtildiği gibi jackknife yönteminin ilk ortaya atılma amacı tahmin edicinin sapmasının tahmin edilmesidir.

Tahmin edicinin $(\hat{\theta})$ sapmasının jackknife tahmini Quenouille tarafından,

$$b(\hat{\theta}_{(-j)})_{jack} = (n-1)\,(\bar{\hat{\theta}}_{(-j)} - \hat{\theta}) \text{ ve}$$

$$= (n-1)\left\{\bar{\hat{\theta}}_{(-j)} - \hat{\theta}(x_1, x_2, \ldots, x_n)\right\} \tag{18}$$

ile elde edilir. Burada;

$b(\hat{\theta}_{(-j)})_{jack}$:j' nci jackknife tahmin edicisinin sapma değeri

$\bar{\hat{\theta}}_{(-j)}$: 17 nolu eşitlikten elde dilen değeri,

$\hat{\theta}$: ana örnekten hesaplanan ilgili parametre tahmin edicisini,

n: örnek hacmini,

n-1: Jackknife örnek hacmini göstermektedir (Efron, 1982). Eşitlik (17) deki $\bar{\hat{\theta}}_{(-j)}$ kullanılarak θ' nın sapma azaltan jackknife tahmin edicisi $\tilde{\theta}_{JACKK}$, Quenouille tarafından,

$$\tilde{\theta}_{JACKK} = \hat{\theta} - b(\hat{\theta}_{(-j)})_{jack}$$

şeklinde tanımlanmıştır. Burada $b(\hat{\theta}_{(-j)})_{jack}$ nin eşitlik (18) deki değeri yerine yazılırsa,

$$\tilde{\theta}_{JACKK} = \hat{\theta}(x_1, x_2, \ldots, x_n) - (n-1)\left\{\bar{\hat{\theta}}_{(-j)} - \hat{\theta}(x_1, x_2, \ldots, x_n)\right\}$$

$$\tilde{\theta}_{JACKK} = n\hat{\theta} - (n-1)\bar{\hat{\theta}}_{(-j)} \tag{19}$$

eşitliği elde edilir. Buradaki terimler daha önceki eşitliklerde verildiği şekildedir.

Örneğin θ parametresi F(x) dağılımının populasyon ortalaması ve θ nın bir tahmin edicisi örnek ortalaması da,

$$\hat{\theta}(X_1, X_2, X_3, \ldots, X_n) = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

şeklinde verilen örnek ortalaması olsun. Bu durumda sapmanın Jackknife tahmini Eşitlik 18'den,

$$b(\hat{\theta}_{(-j)})_{jack} = (n-1)\,(\bar{\hat{\theta}}_{(-j)} - \hat{\theta})$$

$$= (n-1)\left(\frac{\sum_{J=1}^{n} \hat{\theta}_{(-j)}}{n} - \hat{\theta}\right)$$

$$= (n-1)\left(\frac{\sum_{i=1}^{n} \bar{x}_j}{n} - \bar{x}\right)$$

$$=(n-1)\left(\frac{\sum_{j=1}^{n}\left(\sum_{\substack{j=1 \\ i\neq j}}^{n} x_j / n-1\right)}{n} - \bar{x}\right)$$

$$= (n-1)\left(\frac{\sum_{j=1}^{n} x_j}{n} - \bar{x}\right) \tag{20}$$

$$= 0$$

şeklinde hesaplanır.

3.2.2.2. Jackknife standart hata tahmini

Bilindiği gibi $\hat{\theta}$ tahmin edicisinin standart sapması varyansın kareköküdür. Varyans ise şu şekilde hesaplanır;

$$Var = E\left[\hat{\theta} - E\left(\hat{\theta}\right)\right]^2$$

Tukey(1958) varyansın tahmin edicisinin;

$$s^2 = \frac{n-1}{n}\sum_{j=1}^{n}\left[\hat{\theta}_{(-j)} - \bar{\hat{\theta}}_{(-j)}\right]^2 \tag{21}$$

şeklinde elde edildiğini belirtmektedir. Burada;

$\hat{\theta}_{(-j)}$: j'nci gözlem çıkarıldıktan sonra geriye kalan n-1 gözlemden hesaplanan tahmin değerini(j'nci jackknife tahmin değeri),

$\bar{\hat{\theta}}_{(-j)}$: 17 nolu eşitlik ile hesaplanan değeri göstermektedir.

Buradan hareketle Jackknife standart sapması ise;

$S_{JACKK} = \sqrt{S^2}$ olmaktadır.

Bu durum için jackknife tahmin edicisi;

$$s_{JACKK} = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{(-j)} - \bar{\hat{\theta}}_{(-J)})^2} \qquad (22)$$

olarak tahmin edilmektedir.(Efron 1982, Efron and Gong 1983).

Burada hemen şunu belirtmekte fayda vardır. $\hat{\theta}$ herhangi bir tahmin edici olmak üzere bu tahmin edicinin standart sapmasının (varyansının) Jackknife tahminini (22) den bulmak her zaman mümkün olmaktadır.

Jackknife yöntemi modeller ile ilgili varsayımlara daha az bağlı olan ve örnekleme yöntemindekine benzer teorik formüller gerektirmeyen bir yöntemdir. Ancak, bir tahmin edicinin n defa hesaplanmasını gerektirmektedir. Bu hesaplamanın yapılması oldukça zaman alıcı olmakla birlikte, bilgisayarların gelişimi bugün bunu olanaklı kılmıştır. Bugün jackknife istatistiksel analizlerde kullanılan oldukça yararlı ve popüler bir yöntem haline gelmiştir.

3.2.2.3. Jackknife güven aralıkları

Jackknife güven aralıkları daha öncede belirtildiği gibi (Stine, 1990), 1958 yılında ilk kez Tukey tarafından geliştirilmiştir. Daha önceki bölümlerde jackknife tahmin edicisinin ve bu tahmin edicinin dağılımının standart sapma tahminlerinin nasıl ele edildikleri üzerinde durulmuştu. Elde edilen bu tahmin edicilerden hareketle % 100(1-2α) güven düzeyindeki jackknife güven aralıkları;

$$P(\hat{\theta}_{(-j)} - t_{n-1,\alpha}\ s_{JACKK} < \theta < \hat{\theta}_{(-j)} + t_{n-1,\alpha}\ s_{JACKK}) = 1 - 2\alpha \qquad (23)$$

biçiminde elde edilmektedir. Burada;

$t_{n-1,\alpha}$: α düzeyinde ve n-1 serbestlik derecesinde t tablo değeri,

$\hat{\theta}_{(-j)}$: jackknife tahmin edicisi

s_{JACKK} : Jackknife standart hata tahminidir.

Örnek hacminin 30 ve üzerinde olması durumunda ise parametrik çıkarımda olduğu gibi normal dağılım koşullarından hareketle t yerine z tablo değerleri kullanılmaktadır(Edward ve Oudewicz, 1988).

3.2.3. Yeniden örnekleme yöntemlerinin regresyon analizinde kullanımı

Bootstrap ve jackknife yöntemleri, örnek ortalaması, sapma, standart hatanın hesaplanması, güven aralıklarının oluşturulmasında uygulanabildiği gibi, doğrusal regresyonda, parametrik olmayan regresyonda ve çok değişkenli modellerde de uygulama alanına sahiptir.

Yöntemlerin kullanılması son yıllarda hızla gelişmiş ve bilgisayarlarla birlikte uygulanması ile kısa bir süre içinde çok kullanılan birer yöntem haline gelmişlerdir.

Bu bölümde öncelikle En küçük Kareler regresyon modeli kısaca anlatılıp, daha sonra yeniden örnekleme yöntemlerinin (bootstrap ve jackknife yöntemleri) en küçük kareler regresyon analizinde nasıl uygulandığı, uygulamada karşılaşılan güçlükler ve çözümü ile ilgili öneriler üzerinde durulmuştur.

3.2.3.1. Klasik en küçük kareler regresyon analiz yöntemi

Herhangi bir problemin incelenmesinde birden fazla sayıda önemli açıklayıcı değişkenin bulunduğu bir ilişki söz konusu olabilir. Bu ilişki doğrusal bir fonksiyonla ifade edilebilir (Ertek,1978). k tane bağımsız değişkenin bulunduğu en küçük kareler yöntemiyle uydurulan çoklu doğrusal regresyon modeli genelde;

$$Y = X\beta + \varepsilon \qquad (24)$$

eşitliği ile tanımlanır. Burada;

Y: (nx1) boyutlu gözlem değerleri vektörü, i = 1,2,3,....,n bağımlı (açıklanan) değişken,

X: nxp bilinen sabit değerler matrisini,

β: px1 boyutlu bilinmeyen parametreler vektörünü,

ε : (nx1) boyutlu, ortalaması sıfır, varyansı σ^2 (sabit) olan bağımsız şans değişkenleri vektörüdür.

Burada ise;

n: gözlem sayısını,

p: parametre sayısını (β_0 dahil)

k: bağımsız değişken sayısını göstermektedir.

Varsayımlar:

a) Doğrusallık Varsayımı: Bu varsayım $Y = x\beta + \varepsilon$ modelinin tanımında gereklidir ve her bir gözlenen y_i değeri X' in i nci satırının (x_i) doğrusal bir fonksiyonudur. Yani,

$$y_i = x_i\beta_i + \varepsilon_i \quad i=1,2,......,n \tag{25}$$

b) Hesaplana bilme varsayımı: β' nın tek tahmin edicisini bula bilmek için $(X'X)^{-1}$ 'in hesaplana bilmesi veya rank(X)= p olması gerekir. Burada p: parametre sayısıdır.

c) Dağılımla ilgili varsayımlar: En küçük karelere bağlı istatistik analizlerde;

1) X' in hatasız ölçüldüğü,

2) ε_i'lerin x_i (i=1,2,...,n) değerlerine bağlı olmadığı,

3) $\varepsilon_i \approx N(0,\sigma^2 I)$

olduğu varsayılır.

d) Genel Varsayım: Tüm gözlemler aynı oranda güvenilir ve en küçük kareler sonuçlarını belirlemede eşit role sahip olmalıdır.

Klasik En küçük kareler yönteminde(EKKY) iyi tahmin yapabilmek için bir takım varsayımların gerçekleşmesi gerekmektedir. Eğer yukarıda verilen varsayımlar sağlanırsa en küçük kareler yöntemiyle elde edilen standart tahmin sonuçları ve önemli bazı özellikleri aşağıdaki şekilde özetlenebilir(Şahinler, 1997).

β'nın en küçük kareler tahmin edicisi ,

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{26}$$

formülü ile elde edilir.

1) px1 boyutlu $\hat{\beta}$ vektörü şu özelliklere sahiptir;

a) $\hat{\beta}$, β'nın sapmasız tahmin edicisidir, yani

$$E(\hat{\beta}) = \beta \tag{27}$$

b) $\hat{\beta}$, β için en iyi doğrusal sapmasız tahmin edicidir, yani doğrusal sapmasız tahmin ediciler arasında en küçük varyansa sahiptir. $\hat{\beta}$ 'nın varyansı,

$$Var(\hat{\beta}) = \sigma^2(X'X)^{-1} \tag{28}$$

$\hat{\beta}$, ortalaması μ (px1 vektör) ve varyansı $\sum$(pxp matris) olan p boyutlu çok değişkenli normal dağılış gösterir, yani;

$$\hat{\beta} \approx N_p\left[\beta, \sigma^2(X'X)^{-1}\right]$$

2) nx1 boyutlu tahmin değerleri vektörü,

$$\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = PY \tag{29}$$

ile elde edilir. Burada;

$$P = X(X'X)^{-1}X' \tag{30}$$

olup simetrik ve kendileyen bir matristir. $\hat{Y}$ şu özelliklere sahiptir.

a) $E(\hat{Y}) = X\beta$ (31)

b) $Var(\hat{Y}) = \sigma^2 P$ (32)

c) $\hat{Y} \approx N_n(X\beta, \sigma^2 P)$ (33)

Burada, p: eşitlik(30)' da verilen matristir.

3). nx1 boyutlu hata değerleri vektörü,

$$e = Y - \hat{Y} = Y - PY = (1-P)Y \tag{34}$$

eşitliği ile elde edilir ve şu özelliklere sahiptir;

a) E(e)=0 (35)

b) $Var(e) = \sigma^2(1-P)$ (36)

c) $e \approx N_n\left[0, \sigma^2(1-P)\right]$ (37)

d) $\dfrac{e'e}{\sigma^2} \approx \chi^2_{(n-p)}$ (38)

olup burada,

$\chi^2_{(n-p)}$: n-p serbestlik dereceli χ^2 dağılımını,

e'e: hata kareler toplamını,

göstermektedir.

4). σ^2'nin sapmasız tahmin edicisi,

$$s^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{n-p} = \frac{\sum_{i=1}^{n} e_i^2}{n-p} = \frac{e'e}{n-p} \tag{39}$$

ile elde edilir.

$Var(\hat{\beta}) = s^2(X'X)^{-1}$ ise, $(X'X)^{-1}$ matrisi A ile gösterilirse,

$$\mathrm{Var}(\hat{\beta}) = s^2(X'X)^{-1} = s^2 A_{ii} = s^2 \begin{bmatrix} a_{11} & a_{12}\ldots\ldots a_{1k} \\ a_{21} & a_{22}\ldots\ldots a_{2k} \\ . & \\ . & \\ a_{k1} & a_{k1}\ldots\ldots a_{kk} \end{bmatrix} \qquad (40)$$

olur. Burada A_{ii}; $(X'X)^{-1}$ matrisinin köşegen elemanları, s^2 ise Eşitlik (39) ile elde edilen değerdir. Böylece $\hat{\beta}$ 'nin standart hatası,

$$Se(\hat{\beta}) = s\sqrt{(X'X)^{-1}_{ii}}$$

$$Se(\hat{\beta}) = s\sqrt{A_{ii}} \qquad (41)$$

ile elde edilir(Bülbül ve Altaş, 1998).

Yukarıda da bahsedildiği gibi klasik En küçük kareler yönteminde iyi tahmin yapabilmek için bir takım varsayımların gerçekleşmesi gerekmektedir. Sabit varyans varsayımının gerçekleşmediği durumlarda, yapılan tahmin güvenilir olmamaktadır. Böyle durumlarda, en küçük karelerin varsayımlarına gerek duymaksızın parametrelerin tahmin edilmesi, parametrelerin varyans değerlerinin elde edilmesine ve tahmin hatalarının hesaplanmasına olanak sağlayan yeniden örnekleme yöntemleri kullanılabilmektedir (Efron, 1990).

Bu yöntemlerden en çok kullanılan Bootstrap ve Jackknife yöntemleri ile en küçük kareler regresyon analizinin nasıl yapıldığı aşağıda verilmiştir.

3.2.3.2. En küçük kareler regresyon analizinde bootstrap yeniden örnekleme yöntemi

En küçük kareler regresyon analizinin temelleri, hata teriminin (ε) analizine dayanır. En küçük kareler regresyon analizinde hata terimlerinin tekrarlanması ile uygulanan bu yöntem, 1979 yılında Bradley Efron tarafından ileri sürülmüş ve klasik en küçük kareler yönteminden daha etkili parametre tahminleri elde etmek amacı ile geliştirilmiştir ve hata teriminin yeniden örneklenmesi olarak bilinir.

β nın bir bootstrap tahmin edicisini elde etmek için Bölüm 3.2.1. de verilen algoritma aşağıdaki şekilde izlenir:

1) Populasyondan şansa bağlı olarak n sayıda bir örnek seçilir.
2) Seçilen bu örneğe ait EKK regresyon doğrusu oluşturulur.

3) Bu modelden (34) nolu eşitlik yardımıyla e_i değerleri hesaplanır.

4) Elde edilen e_i değerlerine 1/n olasılığı verilerek her biri n hacminde B tane bootstrap hata alt örnekleri oluşturulur. Böylece deneysel dağılım fonksiyonu $(\hat{F}_\varepsilon(x))$,

$$\hat{F}_\varepsilon(x) = \left\{ \hat{e}_i^2 \leq x \right\} / n \tag{42}$$

şeklinde elde edilir (Shao,1995).

5) Oluşan bu deneysel dağılım fonksiyonundan (12) nolu formül kullanılarak bootstrap hata değerlerinin ortalaması,

$$\bar{\hat{\varepsilon}}_i^* = \frac{\sum_{b=1}^{B} \hat{\varepsilon}_{bi}}{B} \tag{43}$$

şeklinde hesaplanır. Burada,

$\bar{\hat{\varepsilon}}_i^*$: i'nci bootstrap hata tahmin edicisi

$\hat{\varepsilon}_{bi}$: b'nci bootstrap örneğine ait i'nci hata tahmin edicisi

6) Elde edilen $\bar{\hat{\varepsilon}}_i^*$ değerleri 2. adımda oluşturulan modeldeki e_i'ler yerine konarak

$$Y_i^* = \hat{\beta} X + \bar{\hat{\varepsilon}}_i^* \tag{44}$$

şeklinde bootstrap Y^* değerleri hesaplanır.

7). Y^* ve X'den hareketle β 'nın bootstrap tahmin edicisi, en küçük kareler yöntemi ile,

$$\hat{\beta}^* = (X' X)^{-1} X' Y^* \tag{45}$$

şeklinde hesaplanır (Liu, 1988). Elde edilen bu tahmin edici sapmasız olup

$E(\hat{\beta}^*) = (X'X)^{-1} X' E(Y^*) = \hat{\beta}$ olmaktadır.

Bootstrap yönteminin bu uygulamasında hata terimlerinden tekrarlı örnekler seçilerek, tahmin değerlerine $(\hat{Y}_i)$ eklendiği için hata terimlerinin normal dağılıma sahip olduğu varsayılmaktadır (Fox, 1997). Bu varsayım gerçekleşmiyorsa, $\hat{\beta}^*$ 'nın standart hatası sapmalı sonuç verir. Bu nedenle, $Sapma_\varepsilon = (e_i - \bar{e})$'nin yerine, $\dfrac{e_i}{(1 - pn^{-1})^{1/2}}$ normalleştirilmiş hata terimlerinden ε_i^*'ler seçilerek, bootstrap Y değerleri elde edilir(Wu, 1986).

Katsayıların standart hatalarını hesaplamak için kullanılan 40 nolu eşitlikteki s^2 değeri ,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\overset{\wedge *}{\varepsilon}_i - \overset{\wedge *}{\varepsilon}_{ort})' \, (\overset{\wedge *}{\varepsilon}_i - \overset{\wedge *}{\varepsilon}_{ort}) \qquad (46)$$

şeklinde hesaplanır. Burada da,

$\overset{\wedge *}{\varepsilon}_i$: Eşitlik 43'ten elde edilen değeri

$\overset{\wedge *}{\varepsilon}_{ort}$: Bootstrap hata tahmin edicilerinin aritmetik ortalaması olup, $\overset{\wedge *}{\varepsilon}_{ort} = \frac{\sum_{i=1}^{n} \overset{\wedge *}{\varepsilon}_i}{n}$ şeklinde hesaplanır. Böylece 41 nolu eşitlikten elde edilen matrisin köşegen değerleri katsayılara ait standart hataları verir.

Bootstrap yönteminde, hata terimlerinin dağılımına ilişkin varsayım yapılmamasına rağmen, Y^*'lar doğrusal modele göre oluşturularak, modelin fonksiyonel biçiminin doğrusal olduğu varsayılır. Bunun yanında hata terimlerinin normalliği varsayımından hareket edildiği için gerçek hatalar sabit varyansa sahip değilse, bu özellik alt örnek hata terimleri nede yansımayacaktır (Fox, 1997).

3.2.3.3. En küçük kareler regresyon analizinde jackknife yeniden örnekleme yöntemi

Jackknife yeniden örnekleme yöntemi ile en küçük kareler regresyon parametrelerinin tahmin edilmesinde aşağıdaki algoritma izlenir.

1) Populasyondan şansa bağlı olarak n hacimlik bir örnek seçilir.

2) Örnek veri setinden her defasında sıra ile bir gözlem dışarıda bırakılarak her birisi n-1 gözlemden oluşan n tane jackknife alt örnekleri oluşturulur.

3) Oluşturulan bu n tane jackknife alt örneklerine ait n tane en küçük kareler regresyon denklemi ile ilgili parametre tahmin edicileri hesaplanır.

4) Regresyon denklemindeki her bir değişkene ait katsayılar kullanılarak jackknife tahmin edicisi,

$$\hat{\beta}_{JACKJ} = \frac{\sum_{i=1}^{n} \hat{\beta}_{(ij)}}{n} \qquad (47)$$

eşitliği yardımıyla j nci parametre tahmin edicisinin n tane örnekteki değerlerinin aritmetik ortalaması olarak hesaplanır. Burada ,

$\hat{\beta}_{(ij)}$: i'nci jackknife örneğinden hesaplanan j' nci tahmin edici

$\hat{\beta}_{JACKJ}$: j'nci jackknife tahmin edicisi

n: Örnek genişliğini göstermektedir."

5) Eşitlik (47) ile hesaplanan ortalama değerler β'nın jackknife tahmin edicicisi olarak kullanılır.

Var($\hat{\beta}$)' nın jackknife tahmin edicisi (Efron, 1982).

$$Var\ (\hat{\beta}_{JACKJ}) = \frac{n-1}{n}\sum_{i=1}^{n}\left[\hat{\beta}_{(ij)} - \hat{\beta}_{(JACKJ)}\right]'\left[\hat{\beta}_{(ij)} - \hat{\beta}_{(JACKJ)}\right] \tag{48}$$

ile elde edilir. Bu matrisin köşegen elemanları Var($\hat{\beta}$)'ı verir.

3.2.4. Parametrelerin güven aralıkları tahminleri

Doğrusal regresyon denkleminde hata terimlerinin ortalaması sıfır ve sabit varyans (σ^2) ile bağımsız tesadüfi değişken oldukları ve bu tesadüfi değişkenlerin normal dağılıma uygun oldukları varsayılmıştır. $Y_1, Y_2, ..., Y_n$ tesadüfi değişkenlerinin de normal dağılımlı olduğu varsayıldığında, $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, ..., \hat{\beta}_k$ tahmin edicilerinin her biri bir normal değişken olur.

3.2.4.1. Normal yaklaşım Yöntemi

Bölüm 3.2.4.1. deki güven aralıklarına benzer şekilde $\hat{\beta}^*$ nın Normal yaklaşım yöntemi ile güven aralığı için, $\hat{\beta}^*$ 'nın standart hatası, eşitlik 41 den elde edilerek .

$$\left(\hat{\beta}^* - t_{n-p,\alpha/2} * S_e(\hat{\beta}^*) < \beta < \hat{\beta}^* + t_{n-p,\alpha/2} * S_e(\hat{\beta}^*)\right) = 1-\alpha \tag{49}$$

eşitliği ile elde edilir (Saama, 1996). Burada:

$\hat{\beta}^*$: i'nci bootstrap tahmin edicisi

$S_e(\hat{\beta}^*)$: Bootstrap standart hata değeri

$t_{n-p,\square/2}$: n-p serbestlik dereceli, $\alpha/2$ önem seviyesindeki t cetvel değeridir.

3.2.4.2. Yüzdelik yöntemi

Yüzdelik yönteminde temel yaklaşım oldukça basittir. Bu yaklaşımda $\hat{\beta}$ için elde edilen bootstrap kopyalarından oluşan örnekleme dağılımı kullanılarak 1-α güven düzeyinde dağılımın alt sınırı α/2 ,üst sınır ise 1- α/2 lik yüzdelere karşılık gelmektedir .

Bootstrap için yüzdelik yöntemi ile parametreler için güven aralığı,

$$(\hat{\beta}^*_A < \beta < \hat{\beta}^*_U) \quad (50)$$

şeklinde elde edilir. Burada;

$\hat{\beta}^*_A$: Alt sınır tekrar sayısına giren tahmin edicilerin aritmetik ortalaması,

$\hat{\beta}^*_U$: Üst sınır tekrar sayısına giren tahmin edicilerin aritmetik ortalaması olup tüm alt örneklere ait parametreler tahmin edilip küçükten büyüğe doğru sıralandıktan sonra,

Alt sınır tekrar sayısı = B* α/2 = B_A

Üst sınır tekrar sayısı = B* (1- α /2) = B_U

Şeklinde hesaplanır.

3.2.4.3. Jackknife Güven Aralığı

Parametre tahminlerinin jackknife güven aralıkları Bölüm 3.2.2.3'deki güven aralıklarına benzer biçimde aşağıdaki şekilde oluşturulur. ise bunun için $\hat{\beta}_{JACKJ}$ 'nin standart hatası eşitlik 41 den hesaplanarak yerine konulur ise aşağıdaki güven aralığı elde edilir. ,

$$(\hat{\beta}_{JACKJ} - t_{n-p,\alpha/2} * S_e(\hat{\beta}_{JACKJ}) < \beta < \hat{\beta}_{JACKJ} + t_{n-p,\alpha/2} * S_e(\hat{\beta}_{JACKJ})) \quad (51)$$

şeklinde elde edilir. biçiminde elde edilmektedir. Burada;

$t_{n-p,\alpha/2}$: α önem seviyesinde ve n-p serbestlik derecesinde t cetvel değeri,

$\hat{\beta}_{JACKJ}$: Eşitlik 47 ile elde edilen jackknife tahmin edicisi

$S_e(\hat{\beta}_{JACKJ})$: Eşitlik 48 ile elde edilen jackknife standart hatasını göstermektedir.

Örnek hacminin 30 ve üzerinde olması durumunda ise parametrik çıkarımda olduğu gibi normal dağılım koşullarından hareketle t yerine z tablo değerleri kullanılmaktadır(Edward, 1988).

BÖLÜM IV

ARAŞTIRMA BULGULARI

Bu bölümde materyal kısmında sözü edilen (EK 1'de) veriler populasyona ait veriler kabul edilmiş ve En küçük kareler yöntemi ile regresyon modeli oluşturularak parametreler hesaplanmıştır. Daha sonra, hesaplanan bu parametrelerin, klasik örnekleme yöntemleri ile elde edilen tahmin ediciler ile mi yoksa yeniden örnekleme yöntemleri kullanılarak elde edilen tahmin edicilerle mi daha iyi tahmin edilebildiğini ortaya koymak amacıyla aynı verilerden şans örneklemesi ile n=100 (EK2) ve n=30 (EK3) olan örnekler alınmış ve parametreler tahmin edilmiştir. Daha sonra EK 2 ve 3'deki veriler ana örnek olarak kullanılıp yöntem kısmında sözü edilen bootstrap ve jackknife yeniden örnekleme yöntemleri ile örnekler oluşturulmuştur. Bootstrap için oluşturulan hata terimlerine ait örnekler (EK4, EK5) Elde edilen Y^* değerleri (EK6, EK7), jackknife örneklerine ait veriler (EK8, EK9), bu verilere ait en küçük kareler regresyon doğrusu ve ilgili parametreler tahmin edilmiş ve her veri grubu için sonuçlar karşılaştırılmıştır.

4.1. En Küçük Kareler Regresyon Analiz Sonuçları

Materyal kısmında (EK1'de) verilen doğum hastanesine ait değerler kullanılarak en küçük kareler regresyon analizi yapılırsa,

$$Y = -2373.8 + 11.757X_1 + 132.047X_2 \qquad (57)$$

eşitliği elde edilir. Bu eşitlik için yapılan varyans analizinde regresyon istatistik olarak önemli bulunmuştur ($p<0.01$). Eşitlik (57) de verilen regresyona ait bazı parametre değerleri Çizelge 1'de özetlenmiştir.

Çizelge 1. EK 1'deki Veriler İçin (N=320) Bazı Parametre Değerleri

Değişkenler	β	**Standart Hata**	**t**	**Güven aralıkları**
Sabit	-2373,8	1032,160	-2,337*	(-4396.8) - (-350.7)
X_1	11.757	4,223	2,745	(3.42) - (19.98)
X_2	132.047	25,843	5,054	(81.39) – (182.69)

4.2. Klasik Örnekleme Yöntemleri İle Elde Edilen Örneklerin En Küçük Kareler Regresyon Analizi Sonuçları

Klasik örnekleme yöntemleri ile seçilen n=100 hacimlik örneğe (EK 2) ait en küçük kareler regresyon modeli uydurulursa,

$$\hat{Y} = -2831.659 + 4.251X_1 + 149.327X_2 \tag{58}$$

eşitliği elde edilir. Bu eşitlik için yapılan varyans analizinde regresyon istatistik olarak önemli bulunmuştur ($p<0.01$). eşitliği elde edilir. Eşitlik (58) de verilen regresyon sonuçları Çizelge 2. özetlenmiştir.

Çizelge 2. EK 2'deki Veriler İçin (n=100) Bazı Tanımlayıcı İstatistikler

Değişkenler	$\hat{\beta}$	**Standart Hata**	**t**	**Güven aralıkları**
Sabit	-2831.659	1700.946	-1.645*	(-6209.7) – (546.46)
X_1	4.251	7.444	0.571	(-10.52) – (19.13)
X_2	149.327	42.165	3.541	(65.59) – (233.05)

Klasik örnekleme yöntemleri ile seçilen n=30 hacimlik örneğe (EK 3) ait en küçük kareler regresyon modeli uydurulursa,

$$\hat{Y} = -2691.13 + 24.71X_1 + 129.90X_2 \tag{59}$$

eşitliği elde edilir. Bu eşitlik için yapılan varyans analizinde regresyon istatistik olarak önemli bulunmuştur ($p<0.01$). Eşitlik (59) de verilen regresyon sonuçları Çizelge 3'de özetlenmiştir.

Çizelge 3. EK 3'deki Veriler İçin (n=30) Bazı Tanımlayıcı İstatistikler

Değişkenler	$\hat{\beta}$	**Standart Hata**	**t**	**Güven aralıkları**
Sabit	-2691.13	2830.065	-0.971*	(-8555.21) –(3059.35)
X_1	24.71	16.008	1.543	(-10.65) – (55.03)
X_2	129.90	71.786	1.809	(-14.19) - (280.41)

4.3. Yeniden Örnekleme Yöntemleri İle Elde Edilen Örneklerin En Küçük Kareler Regresyon Analizi Sonuçları

4.3.1. Bootstrap Yeniden Örnekleme Yöntemi İle Elde Edilen Örneğe Ait E.K.K. Regresyon Analiz Sonuçları

Bölüm 3.2.4.2.1'deki hata terimlerine dayalı bootstrap algoritması EK 1'deki verilere aşağıdaki şekilde uygulanmıştır.

1) Populasyondan (EK 1) n=100 hacimlik örnek seçilmiştir (EK 2).

2) Seçilen bu örneğe ait en küçük kareler regresyon doğrusu bulunmuştur (Eşitl 58).

3) Eşitlik (34) yardımıyla bu doğruya ait 100 tane e_i değeri hesaplanmıştır (EK 4).

4) Bu hata (e_i) terimlerine her bir değere 1/100 olasılığı verilerek iadeli olarak EXCEL paket programının ''örnekleme'' modülü yardımıyla her biri 100 hacimden oluşan 1000 tane bootstrap hata örnekleri oluşturulmuştur (EK 4).

5) Oluşturulan bu örneklerin (43) nolu eşitlik yardımıyla bootstrap hata tahmin edicileri hesaplanmıştır (EK 4).

6) Elde edilen bootstrap hata tahmin edicileri Eşitlik 58'de yerlerine konarak bootstrap Y değerleri (Y_i*) hesaplanmıştır (EK 5).

7) Oluşan yeni gözlem noktaları kullanılarak en küçük kareler yöntemi ile parametreler tahmin edilerek;

$$\hat{Y}^{*} = -2957.26 + 4.07X_1 + 152.59X_2 \qquad (60)$$

şeklinde bootstrap regresyon katsayılarına ait tahminler hesaplanmıştır. Eşitlik (60) da verilen regresyon modeli ile ilgili olarak sırasıyla Eşitlik 45, 46, 11, 49 ve 50 kullanılarak parametre tahmin edicileri, tahmin edicilere ait sapma, standart hatalar, normal yaklaşım ve yüzdelik yöntemleri ile hesaplanan güven aralık değerleri Çizelge 4'de verilmiştir.

Çizelge 4. EK 2'deki Verilere Ait (n=100) Hata Teriminin Yeniden Örneklemesine Dayanan Bazı Bootstrap Tanımlayıcı İstatistikleri.

Değişkenler	$\hat{\beta}^{*}_{ort}$	**Standart Hata**	**Sapma**	**t**	**Güven aralıkları**	
					Normal yaklaşım Yöntemi(%95)	**YüzdelikYöntemi(% 2,5-%97.5)**
Sabit	-2957.26	1718.89	-125.6	-1.72*	(-6326.28)-(411.76)	-6397.2)-(215.8)
X_1	4.07	6.495	-0.181	0.62	(-8.66)-(16.8)	(-8.8)-(16.33)
X_2	152.591	41.214	3.266	3.53	71.82-233.36	(72.9) -(240.9)

Bootstrap yönteminin bölüm 3.2.1. de bahsedilen iddialarının örnek hacminin küçülmesi ile de ne ölçüde gerçekleşebildiğini görebilmek amacıyla örnek hacmini azaltarak (n=30) bu verilerden (EK 3) yine bölüm 3.2.3.2 deki hata terimlerine dayanan bootstrap algoritmasının aşamaları izlenerek,

1) Populasyondan (EK 1) n=30 hacimlik örnek seçilmiştir (EK 3).

2) Seçilen bu örneğe ait en küçük kareler regresyon doğrusu bulunmuştur (Eşitik 59).

3) Eşitlik (34) yardımıyla bu doğruya ait 30 tane e_i değeri hesaplanmıştır (EK 6).

4) Bu hata (e_i) terimlerine her bir değere 1/30 olasılığı verilerek iadeli olarak EXCEL paket programının ''örnekleme'' modülü yardımıyla her biri 30 hacimden oluşan 1000 tane bootstrap hata örnekleri oluşturulmuştur (EK6).

5) Oluşturulan bu örneklerin (43) nolu eşitlik yardımıyla bootstrap hata tahmin edicileri hesaplanmıştır (EK6).

6) Elde edilen bootstrap hata tahmin edicileri Eşitlik 59'de yerlerine konarak bootstrap Y değerleri (Y_i*) hesaplanmıştır(EK 7).

7) Oluşan yeni gözlem noktaları kullanılarak en küçük kareler yöntemi ile parametreler tahmin edilerek;

$$\hat{Y}^{*} = -2909.42 + 24.38X_1 + 135.67X_2 \qquad (61)$$

şeklinde bootstrap regresyon katsayılarına ait tahminler hesaplanmıştır. Eşitlik (61) de verilen regresyon modeli ile ilgili olarak sırasıyla Eşitlik 45, 46, 11, 49 ve 50 kullanılarak parametre tahmin edicileri, tahmin edicilere ait sapma, standart hatalar, normal yaklaşım ve yüzdelik yöntemleri ile hesaplanan güven aralık değerleri Çizelge 5'de verilmiştir.

Çizelge 5. EK 3'deki Verilere Ait (n=30) Hata Teriminin Yeniden Örneklemesine Dayanan Bazı Bootstrap Tanımlayıcı İstatistikleri.

Değişkenler	$\hat{\beta}^{*}_{ort}$	**Standart Hata**	**Sapma**	**t**	**Güven aralıkları**	
					Normal yaklaşım Yöntemi(%95)	**YüzdelikYöntemi(%2,5-%97.5)**
Sabit	-2909.4	3583.38	-218.2	-0.811*	(-9932.84) - (411)	-10372.1)-(3714.8)
X_1	24.38	15.22	-0.336	1.60	(-5.45) - (54.21)	(-5.5)-(51.7)
X_2	135.67	89.23	5.77	1.52	(-39.22) - (310.56)	(-29.7)-(324.2)

4.3.2. Jackknife Yeniden Örnekleme Yöntemi İle Elde Edilen Örneğe Ait En Küçük Kareler Regresyon Analizi Sonuçları

Bölüm 3.2.3.3'deki jackknife yeniden örnekleme yöntemi algoritması EK 1'deki verilere aşağıdaki şekilde uygulanmıştır.

1) Populasyondan şansa bağlı olarak 100 hacimlik bir örnek veri seti oluşturul muş tur (EK 2)

2) 100 hacimlik Örnek veri setinden her defasında sıra ile bir gözlem grubu dışarıda bırakılarak her birisi 99 gözlemden oluşan 100 tane jackknife alt örnekleri oluşturulmuştur (EK 7)

3) Oluşturulan bu 100 tane jackknife alt örneklerine ait,

$$\hat{Y}_{(-1)} = -2799.73 + 4.63X_1 + 148.11X_2$$

$$\hat{Y}_{(-2)} = -2847.10 + 4.33X_i + 149.65X_2$$

$$\cdot \qquad \cdot \qquad \cdot$$

$$\cdot \qquad \cdot \qquad \cdot$$

$$\hat{Y}_{(-99)} = -2832.46 + 4.27X_i + 149.33X_2$$

$$\hat{Y}_{(-100)} = -2824.75 + 2.98X_i + 150.12X_2.$$

şeklinde 100 tane en küçük kareler regresyon denklemi oluşturularak parametre tahmincileri hesaplanır.

4) Eşitlik (47) yardımıyla yukarıdaki denklemlerde verilen katsayılar kullanılarak jackknife tahmin edicileri hesaplanır ve modelde yerine yazılırsa,

$$\hat{Y} = -2832.820 + 4.251X_1 + 149.356X_2 \qquad (62)$$

eşitliği elde edilir.

Eşitlik (62)'da verilen regresyon modeli ile ilgili olarak sırasıyla Eşitlik 47, 48,11 ve 51 kullanılarak parametre tahmin edicileri, tahmin edicilere ait sapma, standart hata ve güven aralıkları hesaplanmış ve sonuçları Çizelge 6'de verilmiştir.

Çizelge 6. EK 2'deki Verilere Ait (n=100) Jackknife Yeniden Örneklemesine Dayanan Bazı Jackknife Tanımlayıcı İstatistikleri.

Değişkenler	$\hat{\beta}_{(j)}$	**Standart Hata**	**Sapma**	**t**	**Güven aralıkları**
Sabit	-2832.820	1771.698	-114.904	-1.598	(-6340.78) - (675.14)
X_1	4.251	6.593	0.003	0.644	(-8.79) - (17.305)
X_2	149.356	44.763	2.886	3.336	(60.73) - (237.979

Örnek hacminin küçülmesi ile Jackknife tahmin edicilerinin populasyonu ne ölçüde tahmin ettiğini görebilmek amacıyla örnek hacmini azaltarak bölüm 3.2.3.3'deki jackknife yeniden örnekleme yöntemi algoritması EK 1'deki verilere aşağıdaki şekilde uygulanmıştır.

1) Populasyondan şansa bağlı olarak 30 hacimlik bir örnek veri seti oluşturulmuştur (EK 2).

2) 30 hacimlik Örnek veri setinden her defasında sıra ile bir gözlem grubu dışarıda bırakılarak her birisi 29 gözlemden oluşan 30 tane jackknife alt örnekleri oluşturulmuştur (EK 8)

3) Oluşturulan bu 30 tane jackknife alt örneklerine ait ,

$$\hat{Y}_{(-1)} = -2806.26 + 24.46X_1 + 133.26X_2$$

$$\hat{Y}_{(-2)} = -2882.67 + 16.47X_1 + 140.55X_2$$

. . . .

. . . .

. . . .

$$\hat{Y}_{(-30)} = -1938.32 + 28.34X_1 + 109.32X_2$$

şeklinde 30 tane en küçük kareler regresyon denklemi oluşturularak parametre tahmincileri hesaplanır.

4) Eşitlik (47) yardımıyla yukarıdaki denklemlerde verilen katsayılar kullanılarak jackknife tahmin edicileri hesaplanır ve modelde yerine yazılırsa,

$$\hat{Y} = -2694.4 + 24.7X_1 + 130.0X_2 \qquad (63)$$

eşitliği elde edilir.

Eşitlik 63'de verilen regresyon modeli ile ilgili olarak sırasıyla Eşitlik 47, 48, 11 ve 51 kullanılarak parametre tahmin edicileri, tahmin edicilere ait sapma, standart hata ve güven aralıkları hesaplanmış ve sonuçları Çizelge 7'de verilmiştir.

Çizelge 7. EK 3'deki Verilere Ait (n=30) Jackknife Yeniden Örneklemesine Dayanan Bazı Jackknife Tanımlayıcı İstatistikleri

Değişkenler	$\hat{\beta}_{(J)}$	**Standart Hata**	**Sapma**	**t**	**Güven aralıkları**
Sabit	-2694.4	3736.71	-94.6556	-0.721	(-10431.88) – (4915.92)
X_1	24.7	16.36	-0.4258	1.509	(-9.266) – (53.6)
X_2	130.0	92.98	2.7348	1.398	(-57.56) – (324.02)

5. TARTIŞMA ve SONUÇLAR

Bu çalışmada yöntem kısmında verilen, en küçük kareler regresyon analizinde kullanılmak üzere iki yöntem vasıtasıyla(klasik örnekleme, yeniden örnekleme(Bootstrap, Jacknife)) örnekler elde edilmiş ve bunlardan hangisi ile daha etkili parametre tahminleri yapıldığı ortaya konulmaya çalışılmış. Bu amaçla araştırma bulguları bölümünde elde edilen sonuçlar toplu olarak Çizelge 8'de verilmiştir. Karşılaştırmalar öncelikle ikili karşılaştırmalar olarak ele alınmış ve aşağıdaki sonuçlar elde edilmiştir

1) Çizelge 8'deki populasyon ve örnek verilerine ait sonuçlar incelendiğinde örneğin 320 hastaya ait X_1 değişkenine ait regresyon katsayısı 11.757 iken 30 hacimlik örnek için X_1 katsayısı 22.19 olarak bulunmuştur. Bu durumda populasyona ait katsayı ile örneğe ait katsayı arasındaki fark(sapma) (11.757 - 22.19) -10.433 olurken n=30 için standart hata değeri 16.00 olarak hesaplanmıştır. n= 100 için ise X_1 değişkenine ait regresyon katsayısı 4.251 olarak bulunmuştur. Bu durumda populasyona ait katsayı ile 100 hacimlik örneğe ait katsayı arasındaki fark (11.757 - 4.251) 7.506 olurken standart hata değeri 7.444 olmaktadır. Sabit ve X_2 değişkenine ait regresyon katsayıları için ise sapmalar örnek büyüdükçe büyümüş ancak katsayılara ait standart hatalar ve güven sınırları daha küçük hesaplanmıştır.

Bu sonuçlara göre genel olarak örnek genişliği arttıkça parametre tahminleri daha az hatalı diğer bir ifade ile daha isabetli olduğu görülmüştür. Bu sonuç, "örnek genişliği populasyon genişliğine yaklaştıkça daha isabetli tahminler yapılır" varsayımı ile uyum içerisindedir.

2) Çizelge 8'deki klasik örnek sonuçları (n=100) ile yeniden örnekleme yöntemlerinden bootstrap yönteminin sonuçları (n=100) incelendiğinde, katsayı tahminleri arasında hemen hemen hiç fark olmadığı görülmüştür ($\hat{\beta}_0$=-2831.659, $\hat{\beta}_1$=4.251 ve $\hat{\beta}_2$=149.328 iken bu değerler bootstrap yönteminde $\hat{\beta}_0^*$= -2957.26, $\hat{\beta}_1^*$=4.07 ve $\hat{\beta}_2^*$=152.59 olarak hesaplanmıştır). Ancak bootstrap yönteminde tahmin edicilerin standart hataları ($S_e(\hat{\beta}_0^*)$=1718.89, $S_e(\hat{\beta}_1^*)$=6.495 ve $S_e(\hat{\beta}_2^*)$=41.214) genellikle klasik örnekten elde edilen tahmin edicilerin standart hatalarından($S_e(\hat{\beta}_0)$=1700.94, $S_e(\hat{\beta}_1)$=7.444 ve

$S_e(\hat{\beta}_2)$=42.165) daha küçük bulunmuştur. Bu durumda n=100 için iki örnekleme yönteminde aynı tahmin değerleri bulunurken bootstrap yöntemi ile daha küçük standart hata tahminleri elde edilmiştir.

Çizelge 8. İncelenen Örnek Gruplarına Ait Bazı Tanımlayıcı Değerler

			Değişkenler		
			Sabit	**X_1**	**X_2**
Populasyon (N=320)		β	-2373,8	11.757	132.047
		St.hata	1032,16	4,22	25,84
		t	-2,337	2,745	5,054
Klasik Örnek Değerleri	**n=100**	$\hat{\beta}$	-2831.659	4.251	149.327
		Sapma	457.859	7.506	-17.28
		St. hata	1700.94	7.444	42.165
		t	-1.645	0.622	3.512
		Güven A.	(-209.7) – (546.42)	(-10.5) – (19.0)	(65.59)-(233.0)
	n=30	$\hat{\beta}$	-2747.93	22.19	133.11
		Sapma	374.13	-10.43	-1.063
		St.hata	2830.065	16.008	71.786
		t	-0.971	1.704	1.811
		Güven A.	(-8555.2) –(3059.3	(-10.65) –(55.0)	(-14.19)-(280.4)
Bootstrap Değerleri	**n=100**	$\hat{\beta}^*_{ort}$	-2957.26	4.07	152.591
		Sapma	-125.601	-0.1812	3.266
		St. Hata	1718.89	6.495	41.214
		t	-1.724	0.626	3.531
		Güven A.	(-6326.28)-(411.7)	(-8.66)-(16.8)	(71.82)-(233.3)
	n=30	$\hat{\beta}^*_{ort}$	-2909.42	24.38	135.67
		Sapma	-218.288	-0.3345	5.77
		St.H	3583.38	15.22	89.23
		t	-0.811	1.60	1.52
		Güven A	(-9932.8)-(411.4)	(-5.45)-(54.21)	(-39.22)-(310.5)
Jackknife Değerleri	**n=100**	$\hat{\beta}_{JACKJ}$	-2832.820	4.251	149.356
		Sapma	-114.904	0.003	2.886
		St. hata	1771.698	6.593	44.763
		t	-1.598	0.6447	3.3368
		Güven A.	(-6340.78)-(675.1)	(-8.79)-(17.30)	(60.73)-(237.9)
	n=30	$\hat{\beta}_{JACKJ}$	-2694.4	24.7	130.0
		Sapma	-94.655	-0.4258	2.734
		St. hata	3736.71	16.36	92.98
		t	-0.721	1.509	1.398
		Güven A.	(-1043.8)-(4915.9)	(-9.266)-(53.6)	(-57.56)-(324.0)

n=30 için sonuçlar incelendiğinde ise katsayıların tahmin değerleri yine yakın değerler almasına rağmen n=100 için olan sonuçların aksine bootstrap yöntemi ($\hat{\beta}_0^*$=2909.42, ($S_e(\hat{\beta}_0^*)$=3583.38, $\hat{\beta}_1^*$=24.38, $S_e(\hat{\beta}_1^*)$=15.22 ve $\hat{\beta}_2^*$=135.67 $S_e(\hat{\beta}_2^*)$=89.23) klasik yöntemden daha büyük standart hatalı tahminler vermiştir. Bu durumda bootstrap yönteminin her zaman, her değişken için daha küçük standart hatalı tahminler vermediği görülmektedir. Benzer sonuçlar Fox (1997)'da da görülmektedir. Ancak genellikle bootstrap yönteminin klasik yöntemlere göre daha küçük standart hatalı tahminler verdiği belirtilmektedir(Efron 1979).

3) Çizelge 8'deki klasik örnek sonuçları (n=100) ile yeniden örnekleme yöntemlerinden jackknife yönteminin sonuçları (n=100) incelendiğinde, katsayı tahminleri arasında hemen hemen hiç fark olmadığı görülmüştür ($\hat{\beta}_0$=-2831.659, $\hat{\beta}_1$=4.251 ve $\hat{\beta}_2$=149.328 iken bu değerler jackknife yönteminde $\hat{\beta}_{JACK0}$= -2832.820, $\hat{\beta}_{JACKK1}$=4.251 ve $\hat{\beta}_{JACKK2}$=149.356 olarak hesaplanmıştır). Ancak jackknife yönteminde tahmin edicilerin standart hataları ($S_e(\hat{\beta}_{JACKK0})$=1771.698, $S_e(\hat{\beta}_{JACKK1})$=6.593 ve $S_e(\hat{\beta}_{JACKK2})$=44.76) genellikle klasik örnekten elde edilen tahmin edicilerin standart hatalarından($S_e(\hat{\beta}_0)$=1700.94, $S_e(\hat{\beta}_1)$=7.444 ve $S_e(\hat{\beta}_2)$=42.165) daha büyük ($\hat{\beta}_{JACKK1}$ hariç) bulunmuştur. Bu durumda n=100 için iki örnekleme yönteminde aynı tahmin değerleri bulunurken jackknife yöntemi ile daha büyük standart hata tahminleri ($\hat{\beta}_{JACKK1}$ hariç) elde edilmiştir.

n=30 için sonuçlar incelendiğinde ise katsayıların tahmin değerleri yine yakın değerler almasına rağmen n=100 için olan sonuçların aksine jackknife yönteminin ($\hat{\beta}_{JACKK1}$ hariç) ($\hat{\beta}_{JACK0}$=-2694.4, $S_e(\hat{\beta}_{JACKK0})$ =3736.71, $\hat{\beta}_{JACKK1}$=22.17, $S_e(\hat{\beta}_{JACKK1})$=16.36 ve $\hat{\beta}_{JACKK2}$=130.0 $S_e(\hat{\beta}_{JACKK2})$=92.98). klasik yöntemden daha büyük standart hatalı tahminler vermiştir. Bu durumda jackknife yönteminden elde edilen sonuçlar klasik örnekleme yöntemlerine benzer olmasına rağmen değişkenin yapısına bağlı olarak genelde Quennouilli(1949), Tukey(1958) ve Efron (1982)'in bildirdiğinin tersine daha büyük hata ile tahminler vermiştir. Benzer sonuçlar Quennouilli(1949) ve Tukey(1958) de de görülmektedir. Ancak genellikle jackknife yönteminin klasik yöntemlere göre daha küçük standart hatalı tahminler verdiği belirtilmektedir(Efron1982).

4) Çizelge 8'deki bootstrap yöntemi ile elde edilen sonuçlar (n=100) ile jackknife yöntemini ile elde edilen sonuçlar (n=100) incelendiğinde, katsayı tahminleri arasında hemen hemen hiç fark olmadığı görülmüştür ($\hat{\beta}_0^*$= -2957.26, $\hat{\beta}_1^*$=4.07 ve $\hat{\beta}_2^*$=152.591 olarak hesaplanmıştır iken bu değerler jackknife yönteminde $\hat{\beta}_{JACK0}$= -2832.820, $\hat{\beta}_{JACKK1}$=4.251 ve $\hat{\beta}_{JACKK2}$=149.356 olarak hesaplanmıştır). Ancak jackknife yönteminde tahmin edicilerin standart hataları ($S_e(\hat{\beta}_{JACKK0})$=1771.698, $S_e(\hat{\beta}_{JACKK1})$=6.593 ve $S_e(\hat{\beta}_{JACKK2})$=44.76) genellikle bootstrap örneğinden elde edilen tahmin edicilerin standart hatalarından ($S_e(\hat{\beta}_0)$=1718.89, $S_e(\hat{\beta}_1)$=6.495 ve $S_e(\hat{\beta}_2)$=43.214) daha büyük bulunmuştur. Bu durumda n=100 için iki yeniden örnekleme yönteminde aynı tahmin değerleri bulunurken jackknife yöntemi ile daha büyük standart hata tahminleri elde edilmiştir.

n=30 için sonuçlar incelendiğinde ise katsayıların tahmin değerleri yine yakın değerler almasına rağmen n=100 için olan sonuçların aksine jackknife yönteminin ($\hat{\beta}_{JACKK0}$=-2694.4, $S_e(\hat{\beta}_{JACKK0})$ =3736.71, $\hat{\beta}_1^*$=24.7, $S_e(\hat{\beta}_{JACKK1})$=16.36 ve $\hat{\beta}_{JACKK2}$=130.0 $S_e(\hat{\beta}_{JACKK2})$=92.98) bootstrap yönteminden daha büyük standart hatalı tahminler vermiştir. Bu durumda jackknife yönteminin de bootstrap yöntemi gibi parametrik varsayımlar kullanmak yerine verilerdeki değişkenliği kullanmaya yönelik yöntemler olmasına rağmen, bootstrap yönteminin jackknife yönteminden daha az hatalı parametre tahminleri yapabildiği görülmektedir. Benzer sonuçlar Efron(1979)'da da bildirilmiştir. Bu sonuçlara göre,

1)Her üç örnekleme yönteminde de aynı örnek genişliğinde benzer sonuçlar tahmin edilmiştir.

2) Klasik örnekleme yönteminde örnek hacminin büyütülmesi ile daha az hatalı parametre tahminleri hesaplanmıştır.

3) Büyük örnekler üzerinde inceleme yapmak çoğu zaman büyük işlemlere, zaman, eleman, maliyet gibi nedenlerden dolayı tahmin edilen parametre değerleri, güncelliğini yitirecek ve yararsız duruma gelecektir. Bu nedenle çabukluk ucuzluk, yararlılık ve daha çok bilgi elde etme yönünden Örnek genişliğini artırarak yapılan tahmin ile daha küçük örnekten yeniden örnekleme yapılarak elde edilen tahmin arasında önemsenmeyecek kadar küçük sapmalar olmaktadır. Bu durumda büyük örnek alıp tahmin yapılacağına, küçük örnekler alıp yeniden örnekleme yöntemlerini uygulamak suretiyle benzer sonuca ulaşılabilmektedir. Ancak büyük örnek yerine küçük örnekler alıp yeniden örnekleme yaparak parametre değerlerini tahmin etmek her zaman iyi sonuca götürmeyecektir. Aynı örnek genişliğine sahip örnekleme yöntemleri içinde klasik örnekleme yerine yeniden örnekleme yapmanın genellikle daha iyi sonuçlar verdiği ortaya çıkmıştır.

Yeniden örnekleme yöntemlerinin her ikisinde de ana örnek genişliğinin arttırılması daha sağlıklı sonuçlar vermemiştir.

Bu sonuçlara göre yeniden örnekleme yöntemlerinin benzer yönleri ve farklı yönleri aşağıdaki gibi çıkarılmıştır.

a) Bir populasyondan çekilen örneğin yeniden kullanılması yöntemi bootstrap yöntemi kadar yeni değildir. Bilindiği gibi bu süreç Jackknife yönteminde de kullanılmıştır.

b) Her iki yöntem de güvenilir varyans tahminlerinin kolay elde edilmesini amaçlamaktadır.

c) Sabit varyans varsayımının sağlanamadığı bir çok durumda yapılan çalışmalar bilimsel açıdan sakıncalar taşımaktadır. Bu gibi durumlarda yeniden örnekleme yöntemlerinin en küçük kareler regresyon analizinde bir düzeltme yöntemi olarak kullanılmaktadır. Böylece bu sakıncaları ortadan kaldırmaktadırlar.

Yöntemlerin Farklı yönleri ise;

1) Jackknife yönteminde verilerin örneklemesinin mantığı birini dışarıda bıraktırmaktır. Bootstrap yönteminde ise yerine koyarak yeniden örnekleme yapmaktır

2) Jackknife yöntemi daha az algoritma ve hesaplama içerirken bootstrap yöntemi daha fazla algoritma ve hesaplama içermektedir.

3) Jackknife yönteminin, bootstrapın doğrusal bir yaklaşımı olarak görülmesine rağmen

temelde bootstrap metodunun jackknife'dan daha geniş bir uygulama alanı vardır.

4) Jackknife yöntemi modeller ile ilgili varsayımlara daha az bağlı olan ve klasik örnekleme yöntemlerdekine benzer teorik formüller gerektirmeyen bir yöntemdir

5) Bootstrap tekrarlarının birbirinden bağımsız olmasına karşılık Jackknife tekrarları birbiri ile güçlü bir ilişki içinde bulunmaktadır. Örneğin; iki Jackknife ortalaması mutlaka n-2 ortak gözleme sahipken, Bootstrap tekrarları birbirinden bağımsızdır ve iki bootstrap örneğinde hiç ortak değer olmaya bilir.

6) Jackknife örneklerinin büyüklüğü n-1 iken, Bootstrap'ınki n dir.

7) Bootstrap ve Jackknife tekrarlarının elde edilme biçimi ve süreci de farklıdır. Jackknife yönteminde elde edilen tekrar sayısı n iken, Bootstrap yönteminde bu sayı n^n. kadar olabilir.

8) Bootstrap yönteminde amaç bilinmeyen bir dağılımın tahmin edilmesi amaçlanırken jackknife yönteminde herhangi bir dağımın tahmin edilmesi amacı yoktur.

Bu konuda yapılan çalışmalar ve yukarıdaki sonuçlar göz önüne alındığında uygulanan yöntemlerle ilgili aşağıdaki öneriler yapılabilir.

1) Herhangi bir çalışmada özellikle tahmin edicilerle ilgili varsayımların tutmadığı durumlarda yeniden örnekleme yöntemlerinin uygulanması genellikle daha iyi sonuçlar verecektir.

2) Yeniden örnekleme yöntemleri içinde bootstrap yöntemi, jackknife yöntemine göre daha güvenilir sonuçlar verdiğinden tercih edilmelidir ancak yöntemin her zaman güvenilir sonuçlar verdiği de söylenemez. Sonuçların güvenilirliği verilerin yapısına seçilen örneğin populasyonu iyi temsil edip etmemesine ve dağılım fonksiyonunun populasyon dağılımını iyi yansıtmasına bağlı olarak değişmektedir.

3) Bootstrap tekrar sayısının büyüklüğü çalışılan veri üzerindeki deneyime bağlı olarak tahmin edicinin standart sapmasını tahmin etmek için 50-200 arasında olması, populasyona ait güven aralıklarının oluşturulması için 1000 olması diğer taraftan 10 ile 80 arasındaki örnek büyüklüğü için tekrar sayısının 100 olması yeterli olacaktır.

Sonuç olarak bootstrap yöntemi diğer yöntemlere göre, tahmin hatalarının daha az olması, standart sapmaların daha küçük olması nedeniyle oldukça kullanışlı bir yöntemdir. Üstelik daha güçlü ve güvenilir sonuçlar elde etmek için örnek hacmini arttırmak suretiyle daha çok zaman ve maliyete katlanmak gerekirken, bootstrap tekrarları sayesinde daha güvenilir sonuçlara ulaşılmaktadır. Ancak bootstrap yönteminin her zaman güvenilir sonuçlar vereceğini ummak yanlış olur. Yöntemin başarısı elde edilen verilerin yapısına ve deneysel dağılım fonksiyonunu populasyonun dağılımını çok iyi yansıtmasına bağlı olarak değişecektir.

KAYNAKLAR

Atalay, U., İnal, C.,1999. Geometrik ve İki terimli Dağılımların Para metrelerinin Bootstrap Tahminleri. İstatistik Konferansı, 26-27 Ekim 1998, Gazi Üniversitesi, İstatistik Bölümü Devlet istatistik Enstitüsü Matbaası, Ankara, s.37-67.

Aytaç, M., 1991.Uygulamalı Parametrik Olmayan İstatistik Testleri Uludağ Üniversitesi Basımevi, Bursa.

Aytaç, M., 1999. Matematiksel İstatistik. Genişletilmiş 2. Baskı, Ezgi Kitabevi Yayınları, Bursa, s.508.

Bağırkan, Ş., 1980. İstatistiksel Analize Giriş. Er Matbaası, İstanbul, s.1-63.

Birich, J.B., 1995. Exploratory and Robust Data Analysis Using MINITAB, Virgini Tech, Blachsburg,VA.s.3-12-1,3-13-4.

Bülbül, Ş., Altaş, D.,1998. Bootstrap Yönteminin Model Seçiminde Kullanılması . III. Ulusal Ekonometri ve İstatistik Sempozyumu, 29-30 Mayıs 1997,Uludağ Üniversitesi Basımevi, Bursa, s.1037-1049.

Çakır F., 2000. 1. Sosyal bilimlerde istatistik Baskı Alfa Basım Yayım Dağıtım Ltd. Şti, İstanbul, s.450

Çil, B., 2000. İstatistik. Gözden Geçirilmiş 2. Baskı, Detay Yayıncılık, Ankara; s.369.

Çömlekçi, N., 1998. Temel İstatistik İlke ve Teknikleri. Gözden Geçirilmiş 3. Baskı, Bilim Teknik Yayınevi, İstanbul, s.500

Dicicio, T., Tibshirani, R., 1987. Bootstrap Confidence Intervals and Bootstrap Approximations. JASSA, Vol. 82, No.397.s.163-170.

Edward. J.O., 1988. Modern Mathemamatical Statistics; John Wiley and Sons inc, Newyork, s.747.

Efron, B., 1979. Bootstrap Methods: Anoher Look at the Jackknife. Ann. of Stat., Vol. 7, No. 1, s.1-26.

Efron, B., 1981. Censored Data and Bootstrap. JASA, Vol.76, No.374. s.312-319.

Efron, B., 1982. The Jackknife, The Bootstrap and Other Resampling Plans. Soc. for Ind. and App. Math., 6th Edi., Philadelpia. s37.

Erfon, B., Gong, G., 1983. Aleisurely Look at The Bootstrap, The Jackknife and Cross-Validation " The American Statistician, February, JASA, Vol. 37, No.1, s36-48.

Erfon, B., and Tıbshıranı, R, 1986. Bootstrap Methods for Standart Errors, Confidence Intervals, and Other Measures of Statistical Accuracy, Statistical Science, 1, 54-77.

Efron, B.,Tibshirani, R., 1993. And Introduction to the Bootstrap Chapman and Hall, New-York. s.45-393.

Efron B., 1990. More.Efficient Bootstrap Computations. JASA, Vol. 85, No. 409.

Ertek, T., 1978. Ekonometriye Giriş. Orta Doğu Teknik Üniversitesi, Ankara, 136s.

Excel-2000. Microsoft Corporatino, Vıa Technologies, Inc. http:// www.microsoft.com

Fox, J., 1997. Applied Regession Analysis, Linear Models and Related Methods. London, s.494-520.

Gürsakal, N., 1998. Bilgisayar Uygulamalı İstatistik II. Marmara Kitabevi Yayınları, Bursa, s.8-9.

Hall, P., 1986. On The Bootstrap and Confidence Intervals. The Ann. of Stat., Vol.14, No.4, s.1431-1452.

Jeremy, B., 1996. Recent Developments in Bootstrapping Times Series, Federal reserve Board, İnternational Finance Discussion Paper Series, September 25, s.2-34

Kabukçu, M. A., 1994.Sağlık Sosyal ve Fen Bilimlerinde Uygulamalı İstatistik Selçuk Üniversitesi, Ziraat Fakültesi, Tarım Ekonomisi Bölümü, Konya, s.270.

Korum, U., K., 1971. Matematiksel İstatistiğe Giriş. Ankara Üni. Basımevi, Ankara, s.467.

İşçil, N., 1975, İstatistiksel Kalite kontrolü Anakara İktisadi ve Ticari ilimler Akademisi Yayını,88,36-55.

İşçil, N., 1977. Örnekleme Yöntemleri Kalite matbaası Ankara, s.297-336.

İşyar, Y., 1999. Ekonometrik Modeller Uludağ Üniversitesi İkt.ve Bil.Fak. Vipaş A.Ş. 2. Baskı.-No: 17 Bursa. s. 695.

Leger, C., Politis, D. N., Romano, J. P. 1992. Bootstrap Technology and Applications. Technometrics, Vol. 34, No. 4.s.378.

Liu, Y. R., 1988 Bootstrap Procedures Under Some Non-I:I.D. Models. Ann. of Stat., Vol.16, No. 4, 1696-1708.s.

Mooney C. Z., Duval, R. D. 1993. Bostatrapping: A Nonparametric Approach to Statictical Inference. Sage Uni. Papers.

Oudewicz E. J. 1988. Modern Mathematical Statistics. John Wiley and Johns Inc., Newyork.

Püskülcü, H., F. İkiz, 1986. İstatistiğe Giriş. Ege Üniversitesi matbaası, Bornova-İzmir.

Saama, P. M., 1996. The permutatıon Bootstrap Resamplıng Dıstrıbutıons of The Regressıon Parameter. http // saama.ans.msu.edu/ permutation. Bootstrap regresion. htm.

Serper, Ö., Aytaç, M., 1999. Örnekleme. Filiz Kitabevi, İstanbul, s.15-35.

Serper, Ö.,1996. Uygulamalı İstatistik I. Genişletilmiş 3. Baskı, Filiz Kitabevi, İstanbul.s12

Serper, Ö., 1996. Uygulamalı İstatistik II. Genişletilmiş 3.Baskı, Filiz Kitabevi, İstanbul,57

Shao, J., Tu, D. 1995 The Jackknife and Bootstrap. Stat. Help Index, Version 5.0, New York.

Shao, J. 1996. Bootstrap Model Selection. JASA, Vol. 91, No. 434.s.655-665.

Stine, R., 1985. Bootstrap Prediction Intervals for Regression. Jour. of the Amer. Stat. Assoç., Vol. 80, No. 392, 1027p.

Stine, R., 1990. Modern Methods of Data Analysis. Edit. by John Fox, Scotland Sage Pub.s.325-373.

Strawderman, R. L., Wells, M. T., 1997, Accurate Bootstrap Confidence Limits for the Cumulative Hazart and Survivor Functions Under Random Censoring, Journal of the American Statistical Association, 92, 1356-1374.

S-Plus-2000 for Windws, 1999. http:/www.mathsoft.com/s-plus, Data Analysis Products Division, Mathsoft., Inc. Seattle, Washington.

SPSS – For Windows.1999. SPSS for Windows, Release 10.01 Stardart Version, SPSS Inc., USA.

Sümbülloğlu, K., Sümbülloğlu, V. 1997. Biyoistatistik .7. Baskı, Hatiboğlu Yayınevi, Ankara, s.269.

Şahinler, S, 1997. Regresyon Analizinde Etkili Gözlemlerin Belirlenmesinde Kullanılan İstatistiklerin Karşılaştırmalı olarak İncelenmesi , Doktora Tezi, Adana, s.162.

Şengün, M., 1999. Yeniden Örnekleme Metoduna Nonparametrik Yaklaşım .IV Ekonometri ve İstatistik Sempozyumu, 14-16 Mayıs 1999, Yaylım matbacılık, Antalya, s.1027-1035.

Tarı, R., 1999. Ekonometri. Alfa Basım Yayım Dağıtım San. ve Tic. Ltd. Şti., Yayın No: 609, İstanbul, s.403.

Yoğurtçugil, K., 1976. Örnekleme. Sermet Matbaası, İstanbul.

Zeng, Q., Davidan, M., 1997, Bootstrap-Adjusted Calibration Confidence Intervals for Immunoassay, Journal of the American Statistical Association, 92, 278-290.

Wu, C. F. J.,1986 Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. Ann. Of Stat., Vol. 14pp, No. 4, 1261-1295.

EK-1. Populasyona (N=320'lik) Ait Veri Seti

Sıra No	Y	X_1	X_2
1	2500	22	39
2	4000	29	40
3	2500	20	38
4	3780	25	40
5	3650	20	40
6	3150	19	40
7	3170	21	40
8	2940	31	40
9	3720	19	40
10	3090	26	40
11	3200	21	40
12	3180	25	40
13	3210	28	39
14	2950	22	39
15	3210	17	39
.	.	.	.
.	.	.	.
.	.	.	.
305	2980	32	40
306	3400	29	40
307	3000	31	37
308	3510	21	40
309	2640	20	39
310	4150	27	40
311	2760	23	40
312	3500	24	40
313	3600	30	40
314	2620	23	40
315	2950	30	40
316	3190	38	40
317	2970	27	40
318	3560	26	40
319	3360	25	38
320	3010	21	40

EK-2. n=100 Hacimlik Örnek Veri Seti

Sıra No	Y	X_1	X_2	Sıra No	Y	X_1	X_2	Sıra No	Y	X_1	X_2
1	3910	24	40	**35**	2860	32	40	**69**	3710	21	40
2	3100	18	39	**36**	3470	18	41	**70**	3040	35	40
3	2740	28	40	**37**	2760	26	37	**71**	3700	35	40
4	3000	27	39	**38**	3670	33	40	**72**	3750	27	42
5	4225	23	39	**39**	2990	35	40	**73**	3810	26	41
6	3040	17	40	**40**	3550	28	40	**74**	3180	32	41
7	1890	30	38	**41**	3460	35	40	**75**	3000	23	39
8	3710	21	40	**42**	3110	40	40	**76**	3910	30	40
9	3040	35	40	**43**	3280	18	40	**77**	3200	23	40
10	3700	35	40	**44**	4000	26	40	**78**	3390	28	40
11	3750	27	42	**45**	2780	25	40	**79**	3320	19	41
12	3810	26	41	**46**	3160	30	40	**80**	3060	35	40
13	3180	32	41	**47**	3290	27	40	**81**	3500	22	40
14	3090	26	40	**48**	2410	17	40	**82**	3210	22	40
15	3200	21	40	**49**	3400	32	39	**83**	4700	20	40
16	3180	25	40	**50**	2850	33	39	**84**	3280	24	39
17	3210	28	39	**51**	2980	32	40	**85**	2880	22	40
18	2950	22	39	**52**	3400	29	40	**86**	2900	25	39
19	3210	17	39	**53**	3000	31	37	**87**	3650	24	42
20	3390	38	40	**54**	3510	21	40	**88**	3280	28	39
21	3690	24	40	**55**	2640	20	39	**89**	3500	22	41
22	3180	22	39	**56**	4150	27	40	**90**	3000	21	41
23	3030	21	40	**57**	2760	23	40	**91**	3130	19	40
24	2910	20	40	**58**	3500	24	40	**92**	3250	25	40
25	3000	18	40	**59**	2650	33	39	**93**	3050	18	40
26	2300	30	38	**60**	2950	23	40	**94**	3180	20	40
27	2490	22	39	**61**	3400	32	39	**95**	3380	30	41
28	3180	20	40	**62**	2850	33	39	**96**	2370	25	40
29	2780	21	40	**63**	2980	32	40	**97**	3700	32	37
30	3570	31	42	**64**	3400	29	40	**98**	2980	29	42
31	2880	18	40	**65**	3000	31	37	**99**	3250	22	40
32	3000	19	37	**66**	3510	21	40	**100**	2650	19	40
33	3410	18	40	**67**	2640	20	39				
34	2970	33	39	**68**	4150	27	40				

EK-3. n=30 Hacimlik Veri Seti

Sıra No	Y	X_1	X_2
1	3030	32	40
2	2140	17	40
3	3660	19	40
4	3180	22	39
5	3000	18	40
6	3000	25	42
7	3020	18	39
8	2810	22	42
9	3600	38	39
10	3310	25	40
11	2140	21	38
12	3510	20	38
13	3630	29	41
14	3040	27	40
15	3110	20	40
16	2140	21	38
17	3510	20	38
18	3750	27	42
19	3000	33	40
20	3010	21	39
21	3150	25	40
22	3160	30	40
23	3290	27	40
24	2880	22	40
25	2250	18	39
26	2800	25	40
27	3630	29	41
28	3960	25	41
29	3000	27	39
30	2760	35	38

EK-4. n=100 için Bootstrap Hata Örnekleri

Sıra No	Hata Terimleri	Sıra No	1/100	1	2	3	4	5	6	7	.	.	97	98	99	100
1	666,56	1	0,01	-300,2	195,16	-289,18	-380,01	115,74	-877,69	31,4	.	.	446,56	446,56	-162,37	192,07
2	31,4	2	0,01	-337,93	265,07	-250,2	-46,43	-92,18	-106,86	409,8	.	.	174,78	-380,01	271,88	174,78
3	-520,44	3	0,01	-482,37	-482,37	-282,37	408,74	-289,18	893,81	129,56	.	.	-1080,2	-250,2	-467,6	-595,6
4	-106,86	4	0,01	-595,61	-24,93	446,56	-108,94	169,8	-46,43	409,8	.	.	225,8	169,8	-161,94	174,78
5	1135,14	5	0,01	289,56	-217.9	195,16	-250,2	185,89	87,04	174,78	.	.	893,81	98,89	265,07	748,0
6	-173,68	6	0,01	641,06	-200,68	-297,45	195,16	15,07	-230,2	-161,94	.	.	256,56	168,89	-354,93	-108,9
7	-1080,3	7	0,01	94,39	-67,69	-246,77	666,56	408,74	-89,86	-51,5	.	.	-354,93	-450,68	-877,69	98,89
8	479,32	8	0,01	-51,5	-167,93	135,31	-217,93	-201,46	641,06	-106,86	.	.	-246,77	-217,93	-246,77	-198,3
9	-250,2	9	0,01	641,06	479,32	-67,69	107,91	409,8	-89,86	-1080,2	.	.	168,8	1135,1	-250,2	-200,6
10	409,8	10	0,01	641,06	-246,77	408,74	-289,18	31,4	-108,94	-135,61	.	.	289,56	479,32	195,16	135,31
11	195,16	11	0,01	-217,93	409,8	1473,57	98,89	-92,18	-108,94	641,06	.	.	195,16	279,32	479,32	185,89
12	408,74	12	0,01	-217,93	666,56	289,56	-38,27	135,31	289,56	168,89	.	.	-51,5	409,8	-417,45	-38,27
13	-246,77	13	0,01	265,07	1135,14	-520,44	-106,86	-479,18	271,88	-43,96	.	.	388,3	-46,43	-217,93	169,8
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
100	98.89	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
		995		-173,68	-46,43	-43,96	-161,94	-46,43	-67,69	-217,93	.	.	-43,96	145,65	-162,37	-108,94
		996		-39,18	-572,18	-51,5	-246,77	408,74	-246,77	-354,93	.	.	271,88	-670,29	-289,18	-479,18
		997		-51,5	-520,44	-1080,2	-24,93	98,89	479,3	-289,1	.	.	-583,3	893,8	-877,6	-108,9
		998		-877,69	-230,2	135,31	192,07	-46,43	-173,68	-316,43	.	.	479,32	-24,93	115,74	87,04
		999		-161,94	-46,43	748,06	135,31	-1,85	870,53	31,4	.	.	-250,2	479,32	98,89	265,07
		1000		-246,77	-300,2	279,32	145,65	-300,2	33,81	-337,93		.	-89,86	94,39	271,88	-250,2
Hata Ortalamaları				**-9,13**	**8,69**	**-1,30**	**-2,85**	**-14,96**	**-3,41**	**1,00**	.	.	**0,42**	**25,23**	**13,96**	**9,10**

EK-5. n =30 için Bootstrap Hata Örnekleri

Sıra No	Hata Terimleri	1/30	Sıra No	1	2	3	4	.	.	20	21	22	23	24	25	26	27	28	29	30
1	-265,5	0,003	1	770,8	261,5	318,3	278,7	.	.	-623,9	770,8	-623,9	707,6	-132,0	278,7	187,5	-349,8	-569,7	-349,8	-498,2
2	-784,8	0,003	2	-42,1	-784,8	-86,1	-569,7	.	.	200,3	-265,5	707,6	261,5	-498,2	-168,4	27,5	-569,7	-86,1	685,7	261,5
3	685,7	0,003	3	118,1	-320,2	-382,3	278,7	.	.	118,1	-349,8	707,6	707,6	770,8	-623,9	27,5	111,0	50,5	-784,8	-168,4
4	261,5	0,003	4	27,5	-86,1	318,3	50,5	.	.	-498,2	286,1	-569,7	-168,4	27,5	118,1	-42,1	-320,2	-86,1	-349,8	770,8
5	50,5	0,003	5	-322,5	770,8	278,7	-132,0	.	.	707,6	318,3	-86,1	278,7	261,5	278,7	770,8	685,7	-168,4	-168,4	116,2
6	-382,3	0,003	6	116,2	-498,2	118,1	-86,1	.	.	-322,5	27,5	707,6	-132,0	-382,3	-784,8	-265,5	770,8	-320,2	111,0	318,3
7	200,3	0,003	7	-320,2	707,6	770,8	261,5	.	.	-623,9	318,3	318,3	111,0	261,5	318,3	770,8	-168,4	200,3	-86,1	318,3
8	-498,2	0,003	8	770,8	-569,7	-42,1	685,7	.	.	-382,3	-322,5	-42,1	-498,2	278,7	770,8	-349,8	-86,1	-42,1	318,3	-42,1
9	286,1	0,003	9	-498,2	-382,3	278,7	-569,7	.	.	-168,4	-86,1	50,5	-569,7	707,6	200,3	770,8	685,7	707,6	-322,5	318,3
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
30	706.6.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
			997	-784,8	-349,8	261,5	286,1	.	.	-320,2	-168,4	286,1	-382,3	318,3	-623,9	-382,3	-322,5	-42,1	118,1	111,0
			998	-382,3	50,5	-784,8	-168,4	.	.	286,1	187,5	-382,3	-322,5	116,2	27,5	-623,9	187,5	286,1	-498,2	685,7
			999	318,3	-322,5	770,8	-382,3	.	.	-498,2	-168,4	770,8	278,7	-569,7	-265,5	-265,5	111,0	318,3	278,7	278,7
			1000	278,7	111,0	278,7	-168,4	.	.	-623,9	-623,9	-784,8	318,3	286,1	187,5	-569,7	-623,9	-168,4	-320,2	118,1
Hata ortalamaları				**-47,56**	**130,2**	**147,5**	**-5.8**	**.**	**.**	**-203.2**	**-6.2**	**125.3**	**55.6**	**70.9**	**-43.7**	**18.9**	**47.3**	**98.2**	**-54.2**	**117.8**

EK-6. n=100 İçin Bootstrap Y* Değerleri

Sıra No	Y*	β_0	X_1	β_1	X_2	β_2	e_i
1	3247,681	-2831,66	24	4,251	40	149,327	4,236
2	3083,032	-2831,66	18	4,251	39	149,327	14,42
3	3268,009	-2831,66	28	4,251	40	149,327	7,56
4	3109,231	-2831,66	27	4,251	39	149,327	2,36
5	3101,517	-2831,66	23	4,251	39	149,327	11,65
6	3226,608	-2831,66	17	4,251	40	149,327	12,92
7	2999,097	-2831,66	30	4,251	38	149,327	28,8
8	3242,712	-2831,66	21	4,251	40	149,327	12,02
9	3279,356	-2831,66	35	4,251	40	149,327	-10,85
10	3273,076	-2831,66	35	4,251	40	149,327	-17,13
11	3543,972	-2831,66	27	4,251	42	149,327	-10,88
12	3416,014	-2831,66	26	4,251	41	149,327	14,74
13	3418,29	-2831,66	32	4,251	41	149,327	-8,49
14	3235,117	-2831,66	26	4,251	40	149,327	-16,83
15	3228,422	-2831,66	21	4,251	40	149,327	-2,27
16	3286,726	-2831,66	25	4,251	40	149,327	39,03
17	3128,012	-2831,66	28	4,251	39	149,327	16,89
18	3100,866	-2831,66	22	4,251	39	149,327	15,25
19	3074,611	-2831,66	17	4,251	39	149,327	10,25
20	3330,679	-2831,66	38	4,251	40	149,327	27,72
•	.	.	.	.	.	.	.
•	.	.	.	.	.	.	.
•	.	.	.	.	.	.	.
•	.	.	.	.	.	.	.
•	.	.	.	.	.	.	.
•	.	.	.	.	.	.	.
•	.	.	.	.	.	.	.
•	.	.	.	.	.	.	.
•	.	.	.	.	.	.	.
•	.	.	.	.	.	.	.
96	3257,836	-2831,66	25	4,251	40	149,327	10,14
97	2802,822	-2831,66	32	4,251	37	149,327	-26,65
98	3554,984	-2831,66	29	4,251	42	149,327	-8,37
99	3240,593	-2831,66	22	4,251	40	149,327	5,65
100	3224,74	-2831,66	19	4,251	40	149,327	2,55

EK-7. n=30 İçin Bootstrap Y* Değerleri

Sıra No	Y*	β_0	X_1	β_1	X_2	β_2	e_i
1	3269,19	-2691,13	32	24,71	40	129,9	-26,4
2	2926,16	-2691,13	17	24,71	40	129,9	1,22
3	2965,69	-2691,13	19	24,71	40	129,9	-8,67
4	2886,89	-2691,13	22	24,71	39	129,9	-31,7
5	2976,73	-2691,13	18	24,71	40	129,9	27,08
6	3361,91	-2691,13	25	24,71	42	129,9	-20,51
7	2831,56	-2691,13	18	24,71	39	129,9	11,81
8	3321,7	-2691,13	22	24,71	42	129,9	13,41
9	3288,38	-2691,13	38	24,71	39	129,9	-25,57
10	3131,31	-2691,13	25	24,71	40	129,9	8,69
11	2770,8	-2691,13	21	24,71	38	129,9	6,82
12	2727,89	-2691,13	20	24,71	38	129,9	-11,38
13	3357,45	-2691,13	29	24,71	41	129,9	6,09
14	3175,63	-2691,13	27	24,71	40	129,9	3,59
15	2988,61	-2691,13	20	24,71	40	129,9	-10,46
16	2752,99	-2691,13	21	24,71	38	129,9	-10,99
17	2749,69	-2691,13	20	24,71	38	129,9	10,42
18	3422,45	-2691,13	27	24,71	42	129,9	-9,39
19	3328,25	-2691,13	33	24,71	40	129,9	7,95
20	2873,21	-2691,13	21	24,71	39	129,9	-20,67
21	3154,7	-2691,13	25	24,71	40	129,9	32,08
22	3251,52	-2691,13	30	24,71	40	129,9	5,35
23	3187,77	-2691,13	27	24,71	40	129,9	15,73
24	3019,96	-2691,13	22	24,71	40	129,9	-28,53
25	2838,61	-2691,13	18	24,71	39	129,9	18,86
26	3118,85	-2691,13	25	24,71	40	129,9	-3,77
27	3337,28	-2691,13	29	24,71	41	129,9	-14,08
28	3242,5	-2691,13	25	24,71	41	129,9	-10,02
29	3032,55	-2691,13	27	24,71	39	129,9	-9,59
30	3113,92	-2691,13	35	24,71	38	129,9	4

EK-8. Jackknife (n=100)Yeniden Örnek Veri Seti

Sıra No	Ana Örnek			Jackknife Örnekleri												
				Sıra No	1			Sıra No	2			.	Sıra No	100		
	Y	X_1	X_2		Y	X_1	X_2		Y	X_1	X_2			Y	X_1	X_2
1	3910	24	40	**2**	3100	18	39	**1**	3910	24	40	.	**1**	3910	24	40
2	3100	18	39	**3**	2740	28	40	**3**	2740	28	40	.	**2**	3100	18	39
3	2740	28	40	**4**	3000	27	39	**4**	3000	27	39	.	**3**	2740	28	40
4	3000	27	39	**5**	4225	23	39	**5**	4225	23	39	.	**4**	3000	27	39
5	4225	23	39	**6**	3040	17	40	**6**	3040	17	40	.	**5**	4225	23	39
6	3040	17	40	**7**	1890	30	38	**7**	1890	30	38	.	**6**	3040	17	40
7	1890	30	38	**8**	3710	21	40	**8**	3710	21	40	.	**7**	1890	30	38
8	3710	21	40	**9**	3040	35	40	**9**	3040	35	40	.	**8**	3710	21	40
9	3040	35	40	**10**	3700	35	40	**10**	3700	35	40	.	**9**	3040	35	40
10	3700	35	40	**11**	3750	27	42	**11**	3750	27	42	.	**10**	3700	35	40
11	3750	27	42	**12**	3810	26	41	**12**	3810	26	41	.	**11**	3750	27	42
12	3810	26	41	**13**	3180	32	41	**13**	3180	32	41	.	**12**	3810	26	41
13	3180	32	41	**14**	3090	26	40	**14**	3090	26	40	.	**13**	3180	32	41
14	3090	26	40	**15**	3200	21	40	**15**	3200	21	40	.	**14**	3090	26	40
15	3200	21	40	**16**	3180	25	40	**16**	3180	25	40	.	**15**	3200	21	40
16	3180	25	40	**17**	3210	28	39	**17**	3210	28	39	.	**16**	3180	25	40
17	3210	28	39	**18**	2950	22	39	**18**	2950	22	39	.	**17**	3210	28	39
18	2950	22	39	**19**	3210	17	39	**19**	3210	17	39	.	**18**	2950	22	39
19	3210	17	39	**20**	3390	38	40	**20**	3390	38	40	.	**19**	3210	17	39
20	3390	38	40	**21**	3690	24	40	**21**	3690	24	40	.	**20**	3390	38	40
21	3690	24	40	**22**	3180	22	39	**22**	3180	22	39	.	**21**	3690	24	40
22	3180	22	39	**23**	3030	21	40	**23**	3030	21	40	.	**22**	3180	22	39
23	3030	21	40	**24**	2910	20	40	**24**	2910	20	40	.	**23**	3030	21	40
24	2910	20	40	**25**	3000	18	40	**25**	3000	18	40	.	**24**	2910	20	40
25	3000	18	40	**26**	2300	30	38	**26**	2300	30	38	.	**25**	3000	18	40
26	2300	30	38	**27**	2490	22	39	**27**	2490	22	39	.	**26**	2300	30	38
27	2490	22	39	**28**	3180	20	40	**28**	3180	20	40	.	**27**	2490	22	39
28	3180	20	40	**29**	2780	21	40	**29**	2780	21	40	.	**28**	3180	20	40
29	2780	21	40	**30**	3570	31	42	**30**	3570	31	42	.	**29**	2780	21	40
30	3570	31	42	**31**	2880	18	40	**31**	2880	18	40	.	**30**	3570	31	42
31	2880	18	40	**32**	3000	19	37	**32**	3000	19	37	.	**31**	2880	18	40
32	3000	19	37	**33**	3410	18	40	**33**	3410	18	40	.	**32**	3000	19	37
33	3410	18	40	**34**	2970	33	39	**34**	2970	33	39	.	**33**	3410	18	40
34	2970	33	39	**35**	2860	32	40	**35**	2860	32	40	.	**34**	2970	33	39
35	2860	32	40	**36**	3470	18	41	**36**	3470	18	41	.	**35**	2860	32	40
36	3470	18	41	**37**	2760	26	37	**37**	2760	26	37	.	**36**	3470	18	41
37	2760	26	37	**38**	3670	33	40	**38**	3670	33	40	.	**37**	2760	26	37
38	3670	33	40	**39**	2990	35	40	**39**	2990	35	40	.	**38**	3670	33	40

EK-8. devam

39	2990	35	40	**40**	3550	28	40	**40**	3550	28	40	.	**39**	2990	35	40
40	3550	28	40	**41**	3460	35	40	**41**	3460	35	40	.	**40**	3550	28	40
41	3460	35	40	**42**	3110	40	40	**42**	3110	40	40	.	**41**	3460	35	40
42	3110	40	40	**43**	3280	18	40	**43**	3280	18	40	.	**42**	3110	40	40
43	3280	18	40	**44**	4000	26	40	**44**	4000	26	40	.	**43**	3280	18	40
44	4000	26	40	**45**	2780	25	40	**45**	2780	25	40	.	**44**	4000	26	40
45	2780	25	40	**46**	3160	30	40	**46**	3160	30	40	.	**45**	2780	25	40
46	3160	30	40	**47**	3290	27	40	**47**	3290	27	40	.	**46**	3160	30	40
47	3290	27	40	**48**	2410	17	40	**48**	2410	17	40	.	**47**	3290	27	40
48	2410	17	40	**49**	3400	32	39	**49**	3400	32	39	.	**48**	2410	17	40
49	3400	32	39	**50**	2850	33	39	**50**	2850	33	39	.	**49**	3400	32	39
50	2850	33	39	**51**	2980	32	40	**51**	2980	32	40	.	**50**	2850	33	39
51	2980	32	40	**52**	3400	29	40	**52**	3400	29	40	.	**51**	2980	32	40
52	3400	29	40	**53**	3000	31	37	**53**	3000	31	37	.	**52**	3400	29	40
53	3000	31	37	**54**	3510	21	40	**54**	3510	21	40	.	**53**	3000	31	37
54	3510	21	40	**55**	2640	20	39	**55**	2640	20	39	.	**54**	3510	21	40
55	2640	20	39	**56**	4150	27	40	**56**	4150	27	40	.	**55**	2640	20	39
56	4150	27	40	**57**	2760	23	40	**57**	2760	23	40	.	**56**	4150	27	40
57	2760	23	40	**58**	3500	24	40	**58**	3500	24	40	.	**57**	2760	23	40
58	3500	24	40	**59**	2650	33	39	**59**	2650	33	39	.	**58**	3500	24	40
59	2650	33	39	**60**	2950	23	40	**60**	2950	23	40	.	**59**	2650	33	39
60	2950	23	40	**61**	3400	32	39	**61**	3400	32	39	.	**60**	2950	23	40
61	3400	32	39	**62**	2850	33	39	**62**	2850	33	39	.	**61**	3400	32	39
62	2850	33	39	**63**	2980	32	40	**63**	2980	32	40	.	**62**	2850	33	39
63	2980	32	40	**64**	3400	29	40	**64**	3400	29	40	.	**63**	2980	32	40
64	3400	29	40	**65**	3000	31	37	**65**	3000	31	37	.	**64**	3400	29	40
65	3000	31	37	**66**	3510	21	40	**66**	3510	21	40	.	**65**	3000	31	37
66	3510	21	40	**67**	2640	20	39	**67**	2640	20	39	.	**66**	3510	21	40
67	2640	20	39	**68**	4150	27	40	**68**	4150	27	40	.	**67**	2640	20	39
68	4150	27	40	**69**	3710	21	40	**69**	3710	21	40	.	**68**	4150	27	40
69	3710	21	40	**70**	3040	35	40	**70**	3040	35	40	.	**69**	3710	21	40
70	3040	35	40	**71**	3700	35	40	**71**	3700	35	40	.	**70**	3040	35	40
71	3700	35	40	**72**	3750	27	42	**72**	3750	27	42	.	**71**	3700	35	40
72	3750	27	42	**73**	3810	26	41	**73**	3810	26	41	.	**72**	3750	27	42
73	3810	26	41	**74**	3180	32	41	**74**	3180	32	41	.	**73**	3810	26	41
74	3180	32	41	**75**	3000	23	39	**75**	3000	23	39	.	**74**	3180	32	41
75	3000	23	39	**76**	3910	30	40	**76**	3910	30	40	.	**75**	3000	23	39
76	3910	30	40	**77**	3200	23	40	**77**	3200	23	40	.	**76**	3910	30	40
77	3200	23	40	**78**	3390	28	40	**78**	3390	28	40	.	**77**	3200	23	40
78	3390	28	40	**79**	3320	19	41	**79**	3320	19	41	.	**78**	3390	28	40
79	3320	19	41	**80**	3060	35	40	**80**	3060	35	40	.	**79**	3320	19	41
80	3060	35	40	**81**	3500	22	40	**81**	3500	22	40	.	**80**	3060	35	40
81	3500	22	40	**82**	3210	22	40	**82**	3210	22	40	.	**81**	3500	22	40
82	3210	22	40	**83**	4700	20	40	**83**	4700	20	40	.	**82**	3210	22	40
83	4700	20	40	**84**	3280	24	39	**84**	3280	24	39	.	**83**	4700	20	40
84	3280	24	39	**85**	2880	22	40	**85**	2880	22	40	.	**84**	3280	24	39

EK-8. devam

85	2880	22	40	**86**	2900	25	39	**86**	2900	25	39	.	**85**	2880	22	40
86	2900	25	39	**87**	3650	24	42	**87**	3650	24	42	.	**86**	2900	25	39
87	3650	24	42	**88**	3280	28	39	**88**	3280	28	39	.	**87**	3650	24	42
88	3280	28	39	**89**	3500	22	41	**89**	3500	22	41	.	**88**	3280	28	39
89	3500	22	41	**90**	3000	21	41	**90**	3000	21	41	.	**89**	3500	22	41
90	3000	21	41	**91**	3130	19	40	**91**	3130	19	40	.	**90**	3000	21	41
91	3130	19	40	**92**	3250	25	40	**92**	3250	25	40	.	**91**	3130	19	40
92	3250	25	40	**93**	3050	18	40	**93**	3050	18	40	.	**92**	3250	25	40
93	3050	18	40	**94**	3180	20	40	**94**	3180	20	40	.	**93**	3050	18	40
94	3180	20	40	**95**	3380	30	41	**95**	3380	30	41	.	**94**	3180	20	40
95	3380	30	41	**96**	2370	25	40	**96**	2370	25	40	.	**95**	3380	30	41
96	2370	25	40	**97**	3700	32	37	**97**	3700	32	37	.	**96**	2370	25	40
97	3700	32	37	**98**	2980	29	42	**98**	2980	29	42	.	**97**	3700	32	37
98	2980	29	42	**99**	3250	22	40	**99**	3250	22	40	.	**98**	2980	29	42
99	3250	22	40	**100**	2650	19	40	**100**	2650	19	40	.	**99**	3250	22	40
100	2650	19	40													

EK-9. Jackknife (n=30) Yeniden Örnek Veri Seti

Sıra No	Ana Örnek			Jackknife Örnekleri												
					1				2			.		30		
	Y	X_1	X_2	Sıra No	Y	X_1	X_2	Sıra No	Y	X_1	X_2		Sıra No	Y	X_1	X_2
1	3030	32	40	**2**	2140	17	40	**1**	3030	32	40	.	**1**	3030	32	40
2	2140	17	40	**3**	3660	19	40	**3**	3660	19	40	.	**2**	2140	17	40
3	3660	19	40	**4**	3180	22	39	**4**	3180	22	39	.	**3**	3660	19	40
4	3180	22	39	**5**	3000	18	40	**5**	3000	18	40	.	**4**	3180	22	39
5	3000	18	40	**6**	3000	25	42	**6**	3000	25	42	.	**5**	3000	18	40
6	3000	25	42	**7**	3020	18	39	**7**	3020	18	39	.	**6**	3000	25	42
7	3020	18	39	**8**	2810	22	42	**8**	2810	22	42	.	**7**	3020	18	39
8	2810	22	42	**9**	3600	38	39	**9**	3600	38	39	.	**8**	2810	22	42
9	3600	38	39	**10**	3310	25	40	**10**	3310	25	40	.	**9**	3600	38	39
10	3310	25	40	**11**	2140	21	38	**11**	2140	21	38	.	**10**	3310	25	40
11	2140	21	38	**12**	3510	20	38	**12**	3510	20	38	.	**11**	2140	21	38
12	3510	20	38	**13**	3630	29	41	**13**	3630	29	41	.	**12**	3510	20	38
13	3630	29	41	**14**	3040	27	40	**14**	3040	27	40	.	**13**	3630	29	41
14	3040	27	40	**15**	3110	20	40	**15**	3110	20	40	.	**14**	3040	27	40
15	3110	20	40	**16**	2140	21	38	**16**	2140	21	38	.	**15**	3110	20	40
16	2140	21	38	**17**	3510	20	38	**17**	3510	20	38	.	**16**	2140	21	38
17	3510	20	38	**18**	3750	27	42	**18**	3750	27	42	.	**17**	3510	20	38
18	3750	27	42	**19**	3000	33	40	**19**	3000	33	40	.	**18**	3750	27	42
19	3000	33	40	**20**	3010	21	39	**20**	3010	21	39	.	**19**	3000	33	40
20	3010	21	39	**21**	3150	25	40	**21**	3150	25	40	.	**20**	3010	21	39
21	3150	25	40	**22**	3160	30	40	**22**	3160	30	40	.	**21**	3150	25	40
22	3160	30	40	**23**	3290	27	40	**23**	3290	27	40	.	**22**	3160	30	40
23	3290	27	40	**24**	2880	22	40	**24**	2880	22	40	.	**23**	3290	27	40
24	2880	22	40	**25**	2250	18	39	**25**	2250	18	39	.	**24**	2880	22	40
25	2250	18	39	**26**	2800	25	40	**26**	2800	25	40	.	**25**	2250	18	39
26	2800	25	40	**27**	3630	29	41	**27**	3630	29	41	.	**26**	2800	25	40
27	3630	29	41	**28**	3960	25	41	**28**	3960	25	41	.	**27**	3630	29	41
28	3960	25	41	**29**	3000	27	39	**29**	3000	27	39	.	**28**	3960	25	41
29	3000	27	39	**30**	2760	35	38	**30**	2760	35	38	.	**29**	3000	27	39
30	2760	35	38													