

Tanı koyma amaçlı yapılan tıbbi alı malarda jackknife, bootstrap ve apraz geerlilik yöntemlerinin kullanımı

Gülhan OREK C TEMEL¹, Semra ERDO AN¹

ÖZET

Sa ık bilimlerinde veriler analiz edilirken örneklem grubu içindeki homojen olmayan dağılım yapısının olması sık karşılaşılan bir durumdur. Bu şekildeki veri yapısı parametrik testler için gerekli olan veri setinin dağılım varsayımlarını sağlayamamasına neden olacaktır. Bu gibi durumlar ile karşılaşıldığında verinin orijinal yapısı bozulmadan, aynı veri setinin yeniden örnekleme yöntemleri kullanılarak üretilip analize tabi tutulması daha etkili sonuçlar sağlayacaktır. Literatürde, Jackknife, Bootstrap ve apraz geerlilik olmak üzere üç farklı yeniden örnekleme yöntemi kullanılmaktadır. Yeniden örnekleme yöntemleri karşılaşıldığında Jackknife yönteminde verilerin normal dağılım varsayımı vardır. Ayrıca bootstrap yönteminde örneklemin seçilme eğilimi, verinin yerine koyma yöntemine göre yapılmaktadır. Yöntemler bilgisayar teknolojisini ilerlemesi ve istatistik paket programlarında bu modüllerin yer almasından dolayı giderek daha popüler hale gelmiştir. Birçok istatistik paket programının sınıflama ve modelleme amaçlı algoritmalarına bu yöntemler girmiştir. Bu nedenle uygulamada bu yöntemleri kullanmak kaçınılmazdır.

Anahtar Kelime: Yeniden Örnekleme, Jackknife, Bootstrap, apraz Geerlilik

The usage of jackknife, bootstrap and cross-validation methods in the medical studies for diagnostics

ABSTRACT

In medical sciences, the status of no homogeneous distribution within the sample group is encountered often during data analyses. That kind of data characteristic causes that the data set cannot assure distribution hypotheses that are necessary for the parametric tests. In case of the presence of these situations, processing and analyzing the same data set using resampling methods without disturbing the original status of the data offer more powerful results. In the literature, three distinct resampling methods, i.e. Jackknife, Bootstrap and cross-validation, are used. When these methods are compared, there is normal distribution assumption in Jackknife method. In Bootstrap method, the sample is selected according to data replacement method. Resampling methods have increasingly been popular through the advances in computer technology and the availability of these modules within statistical packet software. These methods are included in classification and modeling algorithms of many statistical packet programs. Therefore, it is inevitable to use these methods in practice.

Key Words: Resampling Method, Jackknife, Bootstrap, Cross Validation

G R

Örnekleme yöntemi, bilimsel bir araştırmanın vazgeçilmez amaçlarından birisidir. Örneklem büyüklüğünün yeterli büyüklükte olmaması, yeterli büyüklükte olduğu halde uygun örnekleme yönteminin kullanılmaması, alı maya alınacak olan bireylerin seçiminin doğru ve güvenilir bir şekilde yapılmaması, araştırmanın bilimsel bir temele dayanmadığının bir göstergesidir (1).

Örneklemenin en önemli özelliği, popülasyona ilişkin geerli ve tutarlı tahminlere ulaşmak için örnekleme hatasının minimum olmasını sağlamaktır. Bu koşullar sağlandığında, örneklemin popülasyonu en iyi şekilde tahmin ettiği söylenebilmektedir. Fakat parametrik tahmin etme yöntemleri, örneklem sayısı az olduğunda güvenilir sonuçlar vermemekte, aynı zamanda parametrik yöntemlerin varsayımlarını da bozabilmektedir.

Son yıllarda, verilerin parametrik yöntemlerle değerlendirilmesinin uygun olmadığı durumlarda yeniden örnekleme yöntemleri kullanılmaya başlanmıştır (2).

¹ Mersin Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilimler Anabilim Dalı

Correspondence: Dr. Semra ERDO AN e-posta: semraerdogan@gmail.com

Bununla birlikte, sa lık bilimlerinden elde edilen veriler analiz edilirken veri setinin yeterli büyüklükte olmamasından ya da veriler arasındaki ilişkilerden kaynaklı olarak (tekrar edilebilirlik özelliğinden) veri seti yetersiz kalabilmektedir. Yeniden örnekleme yöntemlerinin temeli, teorik da lımlara dayanmakta ve aynı örneklem içerisinde yeniden örnekler seçmeye ba lı olmaktadır. Bu seçim i lemi ya orijinal verinin yerini, sırasını ya da ili kisini de i tirerek verinin yerine koyulması ya da koyulmaması esasına göre yapılmaktadır. Sınıflama ve modelleme amaçlı yapılan çalı malar ileri istatistiksel yöntemlerin konusu içerisine girmektedir (3). Sınıflama amaçlı yapılan çalı malarda, modelin ba arısı e ri altında kalan alan ya da modelin diagnostik do ruluk istatistikleri ile ölçülmektedir. Modelleme amaçlı yapılan çalı malarda ise regresyon katsayısı istatistik açıdan anlamlı bir modelin yorumlanmasında önem ta ımaktadır. Son yıllarda geleneksel yöntemlere alternatif olarak geli tirilen yeni yöntemlerde regresyon katsayısı, diagnostik do ruluk oranı, e ri altında kalan alan hesaplanması elde edilirken yeniden örnekleme yakla ımı sıklıkla kullanılmaktadır (4-8).

Bu çalı manın amacı, yeniden örnekleme yöntemlerini tanıtmak ve bu yöntemleri kar ıla tırmaktır.

GENEL B LG LER

Yeniden örnekleme yöntemlerinin temeli permütasyon test yöntemine dayanmaktadır. Permütasyon test yöntemi, randomizasyon test, yeniden randomizasyon test ya da exact test olarak da adlandırılmaktadır. Aslında permütasyon test istatistiksel anlamlılık testlerinin bir tipidir. Bu test klasik istatistikte öncü olan R. A. Fisher tarafından 1930'larda ortaya atılmı tır. Fisher permütasyon testi, örnek ortalamaları arasındaki gözlenen farklılı ın olup olmadı ının belirlenmesi eklindeki deneme planında kullanılmaktadır. ki grubun ortalaması kar ıla tırıldı ında verinin sırası ve grubu, yerine koymama yakla ımı kullanılarak de i tirilip birinci, ikinci ve daha fazla permütasyonlar olu turularak bu ekilde veriyi yeniden örnekleme (9). Sonuçta her permütasyonda tesadüfî olarak istatistiksel analizlerin gerçekle tirilebilece ini göstermektedir.

Literatürde, Jackknife, Bootstrap ve çapraz geçerlilik olmak üzere üç farklı yeniden örnekleme yöntemi kullanılmaktadır.

Jackknife Yöntemi

Jackknife yeniden örnekleme yöntemi istatistiksel yanlılıkları ortadan kaldırmak amacıyla ilk olarak 1956'da Quenouille tarafından kullanılmı tır. Daha sonra hipotez testleri ve güven aralı ı olu turmak için 1958'de Tukey tarafından da genellenmi tır (10-11).

Bu prosedür genellikle veri setinde a ırı uç de erlerin varlı ında kullanılmaktadır. Jackknife yöntemi her a amada bir obje ya da bireye ait veriyi dı arıda bırakarak aynı testin tekrar edilmesidir. Bu nedenle bu yöntem birini dı arıda bırak (leave one out) yöntemi olarak da isimlendirilmektedir. Bu ekilde bu yöntem biasları azaltmada etkilidir (2).

Yöntemin temel mantı ı, veri setinde her bir gözlem de eri bir kez dı arıda bırakılarak geride kalan gözlemlerden örneklem istatistikleri hesaplanmaya dayanmaktadır. Veri setindeki her bir gözlem dı arıda bırakılarak sahte gözlem de erleri (pseduevalues, sözde de erler) elde edilir. Sahte gözlenen de erler bir çe it veri transformasyonu olarak ba ımsız ve aynı da ılıma sahipleri mi gibi dü ünülebilir.

Kabul edelim ki parametresini tahmin etmek istiyoruz. i birey sayısını göstermek üzere $i=1,2,...,n$ olan bir veri setinin de genel ortalaması olsun. Jackknife yönteminin her a amasında bir obje ya da bireyin gözlem de eri silindi inden j. veri silindi inde örneklem ortalaması Denklem-1'deki gibi hesaplanmaktadır (2, 10-11).

$$\bar{\theta}_j = \frac{1}{n-1} \sum_{i \neq j}^n x_i \quad (1)$$

j verinin gerçek gözlem de erine kar ılık bu yöntemde sahte gözlem de eri hesaplanaca ından bu sahte gözlem de eri Denklem 2'deki gibi hesaplanmaktadır.

$$x_j = n\bar{\theta} - (n-1)\bar{\theta}_{-j} \quad (2)$$

Her gözlem de eri için bu ekilde sözde de erler hesaplanır. Bu çerçevede parametresi için istatistikleri tahmin etmek istedi imizde sözde de erler Denklem 3'deki gibi hesaplanmaktadır.

$$\hat{\theta}_j^* = n\hat{\theta} - (n-1)\hat{\theta}_j \quad (3)$$

Sözde de erlerin ortalaması Denklem 4'deki, varyansı da Denklem 5'deki gibi hesaplanmaktadır.

$$\hat{\theta}^* = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i^* \quad (4)$$

$$\text{Var}(\hat{\theta}^*) = \frac{\sum_{j=1}^n (\hat{\theta}_j^* - \hat{\theta}^*)^2}{(n-1)} \quad (5)$$

$\hat{\theta}^*$ 'nin örnekleme hatası ise sözde de erlerin örnekleminin varyansından elde edilece inden Denklem 6'daki gibi hesaplanmaktadır.

$$\text{Var}(\hat{\theta}^*) = \frac{\text{Var}(\hat{\theta}_j^*)}{n} = \frac{\sum_{j=1}^n (\hat{\theta}_j^* - \hat{\theta}^*)^2}{n(n-1)} \quad (6)$$

Benzer bir yaklaşımla güven düzeyi için sözde de erlerin güven aralığı Denklem 7'de verilmektedir (2, 10-11).

$$\hat{\theta}^* \mp t_{\alpha/2, n-1} \sqrt{\frac{\sum_{j=1}^n (\hat{\theta}_j^* - \hat{\theta}^*)^2}{n(n-1)}} \quad (7)$$

Bootstrap Yöntemi

Yöntem ilk olarak 1979 yılında Bradley Efron tarafından Jackknife yöntemine alternatif olarak ortaya atılmıştır. 1993 yılında da Efron ve Tibshirani tarafından geliştirilmiştir. Bootstrap metodu standart sapma, güven aralığı gibi istatistiklerde ve parametrik olmayan tahminleme problemlerinde kullanılan basit ve güvenilir bir yöntemdir. Yöntemin temel mantığı herhangi büyüklükteki mevcut veri setindeki gözlemlerin ansa bağılı olarak yer değiştirilerek yeniden örnekleme ile yeni veri setleri oluşturulmasına dayanmaktadır (12-16). Böylece mevcut veri setinden mümkün olabildiğince fazla miktarda bilgi alınabilmektedir. Yöntemde ilk olarak orijinal örneklemden yerine koyma yöntemi ile örnek seçilerek bir bootstrap örnekleme oluşturulur. Ardından bu şekilde çok sayıda bootstrap örneklemleri oluşturulur. Bu

örneklemeler ile tahmin edicinin güven aralığı ve testin istatistik anlamlılığı hesaplanır.

İstatistikte çoğu zaman incelenen parametrelerin normal dağılım varsayımına dayanılarak tahmin edilebilmesi söz konusu olurken bu yöntemde veri setinden yadeli örneklem alınarak parametreler dağılım varsayımından bağımsız olarak da tahmin edilebilmektedir. Eksik ve/veya küçük veri setleri dahi istatistiksel çıkarımla yapmak için kullanılabilir (17).

Literatürde, örnekleme dağılımının varsayımına göre parametrik ve parametrik olmayan bootstrap yöntemleri, örneklerin alınması ekline veya tahmin edicilerin hesaplanma yöntemlerine göre basit, çift, ayrıntılı, tekrarlamalı, do al, ardıcık ve bunun gibi pek çok bootstrap yöntemi yer almaktadır (13). Bu çalışmada standart sapma ve güven aralığı tahminlerine ait formülasyonlar, parametrik olmayan bootstrap yöntemlerine göre verilmektedir.

Populasyon parametresinin tahmin edicisinin hesaplanmasında, populasyondan n hacimlik bir örneklem alınır ve bu örneklem populasyonun en iyi tahmin edicisi olarak kabul edilir. Bu nedenle, bu örneklem populasyon gibi kabul edilerek her defasında yadeli seçimle her bir gözlemin örnekleme girme ansı 1/n alınarak n hacimli bootstrap örneklem veri setleri oluşturulmaktadır. Bu işlem istenildiği kadar tekrarlanarak B adet bootstrap gözlemleri veri seti oluşturulur (13).

Kabul edelim ki B Bootstrap seçim sayısı olmak üzere ($i = 1, 2, \dots, B$) Bootstrap tahminleri elde edilsin ($\hat{\theta}_i, i = 1, B$). Bu tahminler daha sonra ortalama ve varyansın hesaplanmasında kullanılmaktadır (12-16).

Her bootstrap örneklem grubunun tahmin ortalaması olmak üzere bütün tahminlerin ortalaması Denklem 8, standart sapması ise

$$\hat{\theta}_B = \frac{1}{B} \sum_{i=1}^B \hat{\theta}_i \quad (8)$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^B (\hat{\theta}_i - \hat{\theta}_B)^2 \quad (9)$$

Denklem 9'daki gibi hesaplanmaktadır (12-16). Güven aralığı ise güven düzeyi için Denklem

$$\hat{\theta} \mp Z_{\alpha/2} S = \hat{\theta} \mp Z_{\alpha/2} \sqrt{\frac{1}{n-1} \sum_{i=1}^B (\hat{\theta}_i - \hat{\theta}_B)^2} \quad (10)$$

10'da verilmektedir.

Manly 1997'de yaptı 1 simülasyon çalışması standart sapma hesaplaması yapabilmek için 100-200 arası bootstrap örnekler oluşturmak yeterli oldu unu göstermektedir (18).

Çapraz Geçerlilik Yöntemi

1948'de Kurtz tarafından basit çapraz geçerlilik (cross validation) yöntemi önerilmiştir (19). Kurtz'un basit çapraz geçerlilik tekniğini Mosier 1951'de geliştirerek çift çapraz geçerlilik olarak geliştirmiştir (20). Daha sonra da Krus ve Fuller tarafından 1982'de çoklu çapraz geçerlilik teknikleri geliştirilmiştir (21).

Modelin doğruluğunun test edildiği çapraz geçerlilik yönteminde veri başlangıçta test ve eğitim verisi olarak tesadüfi olarak ayrılır. Eğitim verisi modelin kurulumu (eğitimi) amacıyla kullanılır. Test verisi model kurulumunda kullanılmaz modelin doğruluğu bu yeni veri seti üzerinde test edilir (22).

Basit geçerlilikte; veri setinin %5-%33'lük bir kısmı test verisi olarak ayrılır ve modelin öğrenme amacıyla bu kısım kullanılmaz. Geriye kalan kısım üzerinde model kurulur ve gerçek değerler ile tahmin değerleri karşılaştırılarak modelin doğruluğu hesaplanır.

Çiftli çapraz geçerlilik ve çoklu çapraz geçerlilik veri setinin sınırlı sayıda olması durumunda kullanılır. Modelleme amacıyla bütün veri setinin kullanımına bu yöntemler olanak sağlar. Veri seti tesadüfi olarak iki eğitme parçaya ayrılır. Bu parçalardan biri eğitim verisi diğer test verisi olarak ikiye ayrılır. Bu şekilde bir doğruluk hesaplaması yapılır daha sonra test verisi ile eğitim verisi yer değiştirilerek bir doğruluk hesaplaması yapılır ve doğruluk değerlerinin ortalaması alınarak modelin doğruluk oranı hesaplanır. Bu algoritma çift çapraz geçerlilik olarak tanımlanır.

Her verinin bir kez eğitim bir kez de test verisi olarak kullanabileceği geçerlilik yöntemlerinin genellenmiş hali olan çoklu çapraz geçerlilik yöntemidir. Bu yöntemde ise veri n adet eğitme grubuna ayrılır. Bir grup test olarak geriye kalan (n-1) grupta eğitim verisi olarak kullanılır. Ve bu durum n kez tekrar eder. Yine doğruluk değerlerinin ortalaması alınarak modelin doğruluk oranı belirlenir (23).

SONUÇ VE ÖNERİLER

Üzerinde çalışılan örneklem grubu içindeki homojen olmayan dağılım yapısının olması ile veri setinin dağılım varsayımlarını sağlamaması, veri setindeki örneklem hacminin küçük olduğu gibi kısıtlılıklar uygulama amacıyla sıklıkla karşılaşılan durumlardır. Verinin orijinal yapısı bozulmadan, aynı veri setinin yeniden örnekleme yöntemleri kullanılarak üretilmesi veri setinin başlı başına ekilde kullanılabilmesine olanak sağlayacaktır. Yöntemler karşılaştırıldığında Jackknife yöntemi verilerin normal dağılım varsayımını gerektirirken, Çapraz Geçerlilik ve Bootstrap yönteminde normal dağılım varsayımı gerektirmez. Ayrıca Bootstrap yönteminde, Jackknife ve Çapraz Geçerlilik yöntemlerinden farklı olarak örneklemin seçilme ekli, yerine koyma yöntemine göre yapılır. Aynı obje ya da bireye seçilme şansını tanıyan yerine koyma yöntemi ile örnekleme yapma, yerine koymadan yapılan örnekleme seçme biçimine göre daha doğru sonuçlar vermektedir (10,11,17,21).

Jackknife yönteminde tekrarların sayısı örnek genelliği (n-1 kez) ile ilgilidir. Fakat Çapraz Geçerlilik ve Bootstrap yöntemlerinde örnek genelliğinin bir kısıtlama yoktur. Dolayısıyla, orijinal veri setinin birçok kez çoğaltılabiliyor olması bu iki yöntemin avantajıdır (10,11,17,21).

Sınıflama bilimlerinde sınıflama amaçlı yapılan çalışmalar, daha çok belirli bir hastalığın tanısının konulmasında etkili olan parametrelerin, risk faktörlerinin istatistiksel olarak modellenerek belirlenmesi şeklinde yapılmaktadır. Bu tür sınıflama metodları ileri istatistik teknikler kullanmayı gerektirir. Bilgisayar teknolojisinin ilerlemesi, istatistik paket programlarının sınıflama ve modelleme amaçlı kullandıkları algoritmaların da yeniden örnekleme yöntemlerinden herhangi birini kullanıyor olması bu yöntemlerin giderek daha popüler hale gelmesine neden olmuştur (24-26).

Yeniden örnekleme yöntemlerini kullanan sınıflama yöntemleri de literatürde yer almaktadır. Bunlardan en yaygın kullanılanları CART, MARS, Boosting Tree ve Random Forest yöntemleridir (26-29). Bu yöntemlerden CART ve MARS yöntemleri algoritmalarında Çapraz

Geçerlilik yöntemini kullanırken, Boosting Tree ve Random Forest yöntemleri ise Bootstrap örnekleme yöntemlerini kullanmaktadır. Ayrıca sınıflama amaçlı olarak veri altında kalan alan hesaplamaları da yapılmaktadır ki burada yeniden örnekleme yöntemlerinden Jackknife yeniden örnekleme yöntemini kullanan MRMC ROC analizi karımıza çıkmaktadır (6).

Sonuç olarak sağlık alanında yapılan çalışmalarda da yeniden örnekleme yöntemleri ile veri setini yeniden inceledikçe sınıflama başarılarını artıran ileri istatistik yöntemlerinin kullanılması kaçınılmazdır.

KAYNAKLAR

- Sümbülo lu V, Sümbülo lu K. Klinik ve Saha Ara tırmalarında Örnekleme Yöntemleri ve Örnekleme Büyüklü ü. Ankara: Alp Ofset Matbaacılık Ltd. ti., 2005:1-15.
- Yu CH. Resampling methods: concepts, applications, and justification. Practical Assessment, Research & Evaluation 2003; 8(19). Retrieved September 22, 2011 from <http://PAREonline.net/getvn.asp?v=8&n=19>. This paper has been viewed 53,213 times since 9/29/2003.)
- Tinsley HEA, Brown SD, (editors). Handbook of Applied Multivariate Statistics and Mathematical Modelling. USA: Academic Press, 2000.
- Friedman JH. Multivariate Adaptive Regression Splines. Ann Stat 1991; 19(1):1-141.
- Schapiro RE. The Strength of weak learnability. Mach Learn 1990;5:197-227.
- Dorfman DD, Berbaum KS, Length RV. Multireader multicase Receiver Operating Characteristic Methodology: a Bootstrap Analysis. Acad Radiol 1995; 2(7):626-633.
- Hillis SL, Obuchowski NA, Scharz KM et al. A Comparison of the Dorfman-Berbaum-Metz and Obuchowski-Rockette Methods for Receiver Operating Characteristic (ROC) Data. Stat Med 2005; 24(10): 1579-1607.
- Breiman L, Friedman JH, Olshen RA, Stone CJ, (editors). Classification and regression trees. New York: Chapman&Hall,1993.
- Fisher RA, (editors). The Design of Experiment. New York: Hafner Press. 7th Ed. 1935-1966.
- Quenouille MH. Notes on bias in estimation. Biometrika 1956; 43: 353-360.
- Tukey JW. Bias and confidence in not-quite large samples (abstract). Ann. Math. Statist 1958; 29(2): 614.
- Efron B. Bootstrap methods: Another look at the jackknife. The Annals of Statistics 1979; 7(1): 1-26.
- Efron B, Tibshirani R, (editors). An Introduction to the bootstrap. New York: Chapman&Hall, 1993.
- Efron B. Censored data and Bootstrap. JASA 1981; 76 (374): 312-9.
- Peterson I. pick a sample. Science News 1991; 140: 56-57.
- Hamajima N, Yuasa H, Matsuo K. Methods for statistical inferences. Biotherapy 1999; 13(6): 739-44.
- Wehrens R, Putter H, Buydens LMC. The bootstrap: A tutorial. Chemometr Intell Lab 2000; 54: 35-52.
- Manly BFJ, (editors). Randomization, Bootstrap and Monte Carlo Methods in Biology. 2nd Ed. London: Chapman & Hall, 1997.
- Kurtz AK. A Research test of Rorschach test. Pers Psychol 1948; 1: 41-53.
- Mosier CI. Problems and designs of cross-validation. Educ Psychol Meas 1951;11: 5-11.
- Krus DJ, Fuller EA. Computer-assisted multicross-validation in regression analysis. Educ Psychol Meas 1982; 42: 187-193.
- Bishop CM, (editors). Neural Networks for Pattern Recognition. 1st Ed. Oxford: Clarendon Press, 1995.
- Hastie T, Tibshirani R, Friedman J, (editors). The Elements of Statistical Learning Data Mining. Inference and Prediction. Canada Springer; 2003.
- Efron B. How biased is the apparent error rate of a prediction rule ?. J Am Stat Assoc 1983; 78:316-331.
- Picard RP, Cook RD. Cross-validation of regression models. J Am Stat Assoc 1984; 79:575-583.
- Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. 1st Ed. Chapman&Hall/Crc, 1984;18-58.
- Friedman JH. Multivariate adaptive regression splines (with discussion). Ann Stat 1991; 19 (March):1-141.
- Death G. Boosted trees for ecological modelling and prediction. Ecological Society of America 2007; 88(1), 243-251.
- Skurichina M, Duin R. Bagging, boosting and the random subspace method for linear classifiers. Pattern Analysis & Applications 2002;5:121-135.