

ROC Analizine Bir Seçenek: LOWESS Yöntemi

An Alternative to ROC Analysis: LOWESS

Nural Bekiroğlu¹, Hasan Batırel²

¹Marmara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, İstanbul, Türkiye

²Marmara Üniversitesi Tıp Fakültesi, Göğüs Cerrahisi Anabilim Dalı, İstanbul, Türkiye

Öz

Amaç: Klinik bir tanıya karar verme sürecinde, sürekli bir veri için ikili sınıflandırma sisteminde optimum kesim noktası araştırılırken çoğunlukla ROC (alıcı işletim karakteristiği) analizi tercih edilir. Optimum kesim noktası, duyarlılık ve seçicilik arasındaki ilişkinin grafiksel bir gösterimi olan ROC eğrisinden tespit edilir. Ancak ROC eğrisindeki kesim noktasının çok net belirlenmediği durumlarda, ROC eğrilerinde düzleştirilmenin önemini vurgulanmıştır. Bu çalışmanın amacı, ROC eğrilerine Kernel düzleştirme yönteminin uygulanması yerine bir seçenek olarak LOWESS (yerel ağırlıklı saçılım grafiğini düzleştirme) yönteminin kullanılmasını önermek ve bir uygulama ile göstermektir.

Yöntemler: ROC yöntemi yerine önerilen LOWESS yöntemi, karmaşık ilişkileri bir regresyon denklemiyle ifade etmek yerine daha çok görsel bir ifade için tercih edilir. Çalışmamızda, ROC eğrisi ve LOWESS yöntemiyle elde edilen eğrilerde net ve doğru kesim noktalarının bulunması bir uygulama ile irdelenmiş ve yöntemler tartışılmıştır. Uygulamada ise, akciğer rezeksiyonu sırasında aşırı sıvı uygulamasının akciğer yaralanması için bir risk faktörü olduğu bilindiğinden; postoperatif pulmoner komplikasyon varlığı için anlamlı risk faktörleri olarak; intraoperatif sıvı infüzyon artış hızı, ARDS (akut respiratuar distres sendromu), AAH (akut akciğer hasarı), pnömoni, atelektazi, bronkoskopi ihtiyacı ve hava kaçağı/yokluğu kullanılmıştır.

Bulgular: Kesim noktasının, ROC yöntemine göre 5,5-6 mL/kg/h arasındaki bir noktada olabileceği düşünülürken, LOWESS yöntemine göre ise 6,125 mL/kg/h noktasında olacağı saptanmıştır.

Sonuç: Sınıflama yaparken; eğrilerin daha iyi yorumlanabilmesi ve kesim noktasının daha iyi belirlenmesi için ROC eğrisine karşı bir seçenek olarak parametrik olmayan LOWESS düzleştirme yönteminin kullanılmasını önerilmektedir.

Anahtar kelimeler: LOWESS, ROC eğrisi, düzleştirme

Abstract

Objective: Receiver operating characteristic (ROC) analysis is commonly preferred for the dichotomous classification of a continuous random variable in the process of determination of the optimum cut-off point. The optimum cut-off point can be detected using ROC curve, which is a graphic presentation of the relationship between sensitivity and specificity. In the circumstances where the optimum cut-off point cannot be determined properly using ROC curves, the importance of smoothing is emphasized. In this study, it is proposed to use as an alternative locally weighted scatterplot smoothing (LOWESS) instead of ROC curves with Kernel smoothing.

Methods: In our study, determination of the accurate and clear cut-off point obtained using the curves belonging to ROC and LOWESS techniques was discussed by means of an application. Excessive fluid administration during lung resection is a risk for pulmonary injury. The significant risk factors for the presence of postoperative pulmonary complications are the infusion rate of intraoperative fluids, acute respiratory distress syndrome, acute lung injury, pneumonia, atelectasis, need for toilet bronchoscopy, prolonged air leak, and failure to expand were used in the application.

Results: According to the ROC analysis, the cut-off point should have been between 5.5-6mL/kg/h, but according to the LOWESS method, it was determined to be 6.125mL/kg/h.

Conclusion: For the dichotomous classification, to interpret curves and to determine cut-off points perfectly, LOWESS smoothing non parametric method is strongly recommended instead of the non-parametric ROC curve.

Keywords: LOWESS, ROC curve, smoothing

GİRİŞ

Model tahmini için genelde üç temel yaklaşım vardır. Bunlar, parametrik, parametrik olmayan (nonparametric) ve yarı parametrik (semi-parametric) regresyon yaklaşımlarıdır.

Parametrik regresyon yaklaşımında çok fazla varsayıma ihtiyaç duyulur. Hata teriminin dağılım fonksiyonunun bilinir olduğu ve genelde normal dağıldığı ayrıca Y bağımlı değişkeni ile X açıklayıcı değişkenleri arasında doğrusal ilişki olduğu ve β parametrelerinin sonlu sayıda olduğu varsayılır. Ancak bu varsayımların yerine gelmesinin zorluğu nedeniyle parametrik yaklaşımda sonuçların güvenilirliği genelde tartışmalıdır.

Parametrik olmayan regresyon yaklaşımında ise, hiçbir varsayımda bulunulmama ile birlikte hata teriminin dağılımı ile ilgili herhangi bir varsayım yapılmaz, ayrıca açıklayıcı değişkenler arasındaki fonksiyonel yapı için bir varsayımda bulunulmaz ve β parametreler vektöründen de bahsedilmez. Ancak açıklayıcı değişken sayısının fazla olması durumunda, model tahmini zorlaştığı gibi yorumlanmasında da sorunlar yaşanır.

Yarı parametrik regresyon yöntemi ise, parametrik model varsayımlarının sağlanamaması durumunda parametrik ve parametrik olmayan yaklaşımların en iyi yönlerini alır. Hata terimi dağılımı hakkında bir varsayımda bulunulmasa bile modelden tahmin edilebileceği ve parametrik

yaklaşımında olduğu gibi Y bağımlı değişkeni ile X açıklayıcı değişkenleri arasındaki ilişkinin doğrusal olduğu varsayılır. Bu varsayımın dışında başka ek varsayımlara gereksinim duyulmadığından bazen bu modellere kısmi doğrusal model (partially linear model) de denmektedir (1).

Klinik bir tanıya karar verme sürecinde ROC (Receiver Operating Characteristic, alıcı işletim karakteristiği) analizi, lojistik regresyon vb. birçok yöntemden faydalanılır. Bu süreçte özellikle sürekli bir veri için ikili sınıflandırma sisteminde optimum kesim noktası araştırılırken çoğunlukla ROC analizi tercih edilir. Günümüzde tıp, radyoloji, psikoloji, makine öğrenme teknikleri, veri madenciliği ve benzer alanlarda oldukça fazla kullanılan ROC analizi yöntemi, parametrik olmayan bir yaklaşımdır. ROC analizi basit anlamda doğru pozitiflerin, yanlış pozitiflere olan kesri olarak da ifade edilebilir (2). Optimum kesim noktası, duyarlılık ve seçicilik arasındaki ilişkinin grafiksel bir gösterimi olan ROC eğrisinden tespit edilir. Ancak bazen ROC eğrilerinde kesim noktası çok net seçilemez. Bu tip durumlar için Lloyd, tanı/tarama sistemlerini karşılaştırmak ve ROC eğrilerini özetlemek için 'smoothing' yani düzleştirme, bir başka deyişle yumuşaklaştırma kullanılmasının önemi vurgulamıştır (3). Özellikle görsel olarak grafiklerle ifade edilebilen parametrik olmayan ROC analizi için en çok Kernel düzleştirme yöntemi tercih edilmektedir. Böylece kesim noktası daha netleşmekte ayrıca düzleştirilen ROC eğrileri altında kalan alanlar daha iyi karşılaştırılabilmektedir (4). Parametrik olmayan regresyona dayanan LOWESS (yerel ağırlıklı saçılım grafiğini düzleştirme) düzleştirme yöntemiyle elde edilen eğri ile Kernel düzleştirme yöntemi kullanılarak elde edilen ROC eğrisinin hemen hemen aynı sonuçları verdiği bilinmektedir. LOWESS düzleştiricinin bir çeşit Kernel düzleştirici olduğu söylenebilir; ancak aralarında bazı farklılıklar vardır (5).

Bu çalışmada, parametrik olmayan ROC eğrilerine Kernel düzleştirme yönteminin uygulanması yerine bir seçenek olarak yine parametrik olmayan regresyona dayanan LOWESS düzleştirme yönteminin kullanılması tartışılmış ve bir uygulama ile sunulmuştur (6).

YÖNTEMLER

Bu çalışmada yer alan 139 hasta, Ocak 2012-Eylül 2013 tarihleri arasında Marmara Üniversitesi Hastanesi'nde anatomik akciğer rezeksiyonu geçirmiş ve onamlarıyla birlikte Marmara Üniversitesi Tıp Fakültesi Etik Kurulu'na sunulmuş ve onaylanmıştır (2013-309).

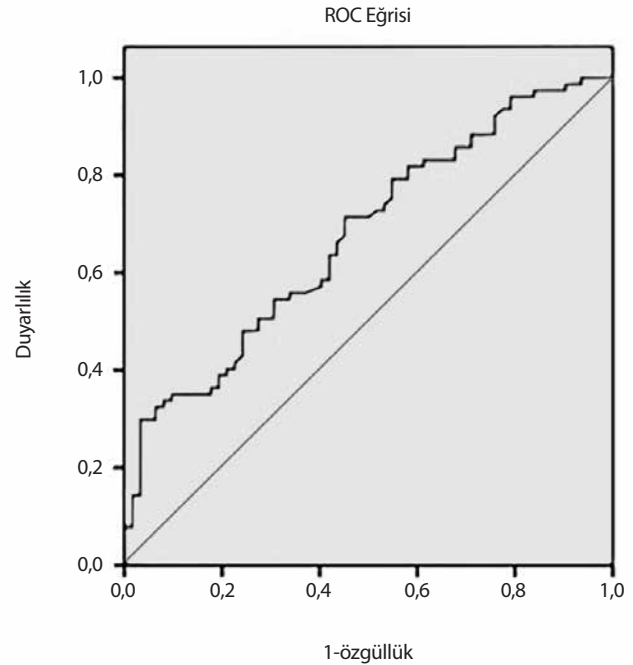
Çalışmanın analizinde, 1979 yılında Cleveland tarafından geliştirilmiş ve sunulmuş olan LOWESS yöntemi kullanılmıştır. LOWESS yerel ağırlıklandırılmış regresyonların düzleştirilmiş serpilme grafiğinin adıdır ve karmaşık ilişkileri bir regresyon denklemiyle ifade etmek yerine daha çok görsel bir ifade için tercih edilir. LOWESS yöntemi ve LOESS yöntemi (yerel polinomial regresyon) genelde aynıdır. Bu modeller, verileri parçalara ayırır ve her bir parçada ayrı regresyonlar (piecewise regression) oluşturur, sonra bu yerel ağırlıklandırılmış regresyonları birbirine düzleştirerek bağlar ve bir grafikte sunar.

İstatistiksel analizler için kullanılan paket program STATA (Texas, ABD)'dir.

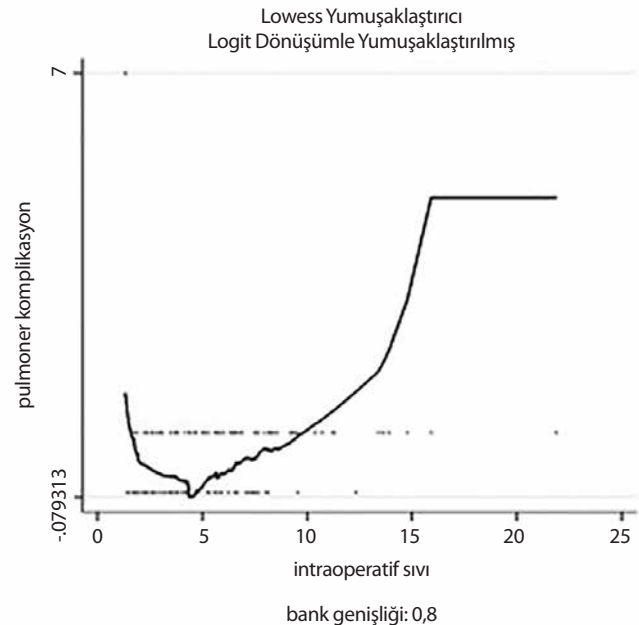
BULGULAR

Akciğer rezeksiyonu sırasında aşırı sıvı uygulamasının akciğer yaranması için bir risk faktörü olduğu bilinmektedir. Çalışmaya alınan, intraoperatif sıvı infüzyon artış hızı, ARDS (akut solunum yolu yetersizliği sendromu), ALI (akut akciğer hasarı), pnömoni, atelektazi,

bronkoskopi ihtiyacı ve hava kaçağı/yokluğu parametreleri; postoperatif pulmoner komplikasyon varlığı için anlamlı risk faktörlerini oluşturmaktadır. ROC eğrisine göre intraoperatif sıvı infüzyon artış hızı parametresi için postoperatif pulmoner komplikasyon oluşması/ oluşmamasını ayırtıracak optimum kesim noktası değerinin; duyarlılık %55 ve (1-özgüllük) %30 ile 5,5-6 mL/kg/h arasındaki bir noktada olabileceği ön görülmüştür (Şekil 1). Eğri altında kalan alan ise 0,67'dir ($p=0,001$).



Şekil 1. ROC yöntemiyle elde edilen eğri



Şekil 2. LOWESS düzleştirme yöntemiyle elden edilen eğri

Grafik yöntemine dayanan ve parametrik olmayan regresyon yöntemi olan LOWESS düzleştiriciye göre analiz yapıldığında ise daha hassas ve doğru bir sonuç olan 6,125 noktasının önemli ayrıştırıcı noktaya sahip olduğu görülmektedir (Şekil 2).

TARTIŞMA

Alıcı işletim karakteristiği analizinde yarı parametrik yaklaşımın, parametrik ve parametrik olmayan yaklaşımlara göre daha iyi sonuçlar verdiği gösterilmiştir (7).

Bu çalışmada, her ne kadar parametrik olmayan ROC eğrilerine Kernel düzleştirme yönteminin uygulanması yerine, yine parametrik olmayan LOWESS düzleştirme yönteminin kullanılması önerilse de, Bauer (9) doğrusal olmayan ilişkiler için gerçekleştirilen LOWESS regresyona seçenек olarak yarı parametrik modelleme olan SEMM'in (Yapısal Eşitlik Karışım Modeli) kullanılmasını önermiştir. Söz konusu düzleştirme yönteminin geçmişte özellikle LOWESS parametrik olmayan regresyona dayandığından LOWESS düzleştirme yöntemi bu nedenle büyük önem taşımaktadır (8-10). LOWESS yönteminin SEMM'e karşı en zayıf yönü verilerin bir regresyon denklemi şeklinde karakterize edilememesidir. Çünkü LOWESS yöntemi daha çok eğrileri düzleştirme algoritmaları üzerine yönelik olup görsel olma özelliği ile öne çıkar. Bu da, parametrik olmayan yöntemin en önemli zayıflığı sayılmaktadır. Çünkü LOWESS tamamen parametrik olmayan bir yöntem iken, SEMM ise yarı parametrik bir yöntemdir. SEMM ile doğrusal olmayan etkileri az da olsa dağılım varsayımları karşılığında test etmek söz konusudur ancak böyle bir durum LOWESS yönteminde gerçekleşemez. SEMM yönteminin en önemli dezavantajı çok teknik bir yöntem olması ve bilgisayar programlarında kullanımının zorluğudur. Bu nedenlerden dolayı, LOWESS yöntemi SEMM yöntemine karşı iyi bir seçenек olmaktadır.

Diğer yandan, LOWESS düzleştirme yöntemiyle elde edilen eğri ile Kernel düzleştirme yöntemi kullanılarak elde edilen ROC eğrisi arasındaki en önemli fark, LOWESS yönteminin veri setinde bulunan uç noktaların olumsuz etkilerini bertaraf etme özelliğine sahip olmasıdır. Bu özellik nedeniyle LOWESS yöntemi Kernel düzleştirme yöntemine göre daha çok tercih edilir.

Yerel ağırlıklı saçılım grafiğini düzleştirme modelinde öncelikle yerel polinomial regresyonlar oluşturulur ve sonra birbirlerine bağlanır ancak yerel eğriler birbirlerine bağlanırken düzleştirilerek bağlama yapılır. Gerçekleşen bu eğrileri düzleştiren bağlama tekniği ise, parametrik olmayan modellerin en önemli özelliğini oluşturur (11). LOWESS yönteminin bir çeşit regresyon grafiği yöntemi olması özelliğiyle, STATA, SAS/INSIGHT ve R-CODE gibi bir takım yazılım ve istatistiksel paket programlarında yer almakta ve ayrıca kolayca kullanılmaktadır (7, 8, 12).

ROC eğrisindeki kesim noktasının daha gerçeğe yakın, doğru ve net olarak belirlenmesinde güçlük çekildiğinde, yarı parametrik bir yöntem olan SEMM'in çok teknik bir yöntem olması ve bilgisayar programlarında kullanımının zorluğu nedeniyle ve ayrıca KERNEL düzleştirme yönteminin de veri setindeki uç noktalardan etkilenmesi sonucu iyi bir eğri vermemesi nedeniyle, eğrinin daha iyi yorumlanabilmesine seçenек olarak parametrik olmayan LOWESS düzleştirme yönteminin kullanılması önerilmektedir.

SONUÇ

Elde edilen kesim noktası özellikle klinik olarak değerlendirildiğinde de, grafik yöntemine dayanan ve parametrik olmayan regresyon yöntemi olan

LOWESS düzleştirici yönteminin, yine parametrik olmayan ve sıkça tercih edilen ROC yöntemine karşı daha iyi sonuçlar verdiği görülmüştür.

Etik Komite Onayı: Bu çalışma için etik komite onayı Marmara Üniversitesi Tıp Fakültesi Etik Kurulu'ndan alınmıştır (Karar No: 2013-309).

Hasta Onamı: Yazılı hasta onamı bu çalışmaya katılan hastadan alınmıştır.

Hakem Değerlendirmesi: Dış Bağımsız.

Yazar Katkıları: Fikir - N.B., H.B.; Tasarım - N.B.; Denetleme - N.B., H.B.; Kaynaklar - N.B., H.B.; Malzemeler - N.B., H.B.; Veri Toplanması ve/veya işleme - N.B., H.B., M.S.; Analiz ve/veya Yorum - N.B.; Literatür taraması - N.B., H.B.; Yazılı Yazan - N.B.; Eleştirel İnceleme - N.B., H.B.

Çıkar Çatışması: Yazarlar çıkar çatışması bildirmemişlerdir.

Finansal Destek: Yazarlar bu çalışma için finansal destek almadıklarını beyan etmişlerdir.

Ethics Committee Approval: Ethics committee approval was received for this study from the ethics committee of Marmara University School of Medicine (Decision No: 2013-309).

Informed Consent: Written informed consent was obtained from patient who participated in this study.

Peer-review: Externally peer-reviewed.

Author contributions: Concept - N.B., H.B.; Design - N.B.; Supervision - N.B., H.B.; Resource - N.B., H.B.; Materials - N.B., H.B.; Data Collection &/or Processing - N.B., H.B., M.S.; Analysis &/or Interpretation - N.B.; Literature Search - N.B., H.B.; Writing - N.B.; Critical Reviews - N.B., H.B.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study has received no financial support.

KAYNAKLAR

1. Akkuş Ö, Demir S, Karasoy D. İki düzeyli bağımlı değişken modelinin yarı parametrik tahmini. İstatistikçiler Dergisi 2008; 135-43.
2. Swets JA. Signal detection theory and ROC analysis in psychology and diagnostics: collected papers; Lawrence Erlbaum Associates, Publishers 1996 Mahwah, New Jersey.
3. Lloyd CJ. Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. J Am Stat Assoc 1998; 93: 1356-64. [CrossRef]
4. Zhou XH, Harezlak J. Comparison of bandwidth selection methods for kernel smoothing of ROC curves. Stat Med 2002; 21: 2045-55. [CrossRef]
5. Irizarry RA. Kernel Methods, Chapter 6: 96-106.
6. Tang LL, Zhou XH. A semiparametric separation curve approach for comparing correlated ROC data from multiple markers. J Comput Graph Stat 2012; 21: 662-76. [CrossRef]
7. Colak E, Mutlu F, Bal C, Oner S, Ozdamar K, Gok B, et al. Comparison of semiparametric, parametric, and nonparametric roc analysis for continuous diagnostic tests using a simulation study and acute coronary syndrome data. Comput Math Methods Med 2012; 1-7. [CrossRef]
8. Pek J, Sterba SK, Kok BE, Bauer DJ. Estimating and visualizing nonlinear relations among latent variables: a semiparametric approach. Multivariate Behav Res 2009; 44: 407-36. [CrossRef]
9. Bauer DJ. Structural equation modeling: a semiparametric approach to modeling nonlinear relations among latent variables, 12(4). Lawrence Erlbaum Associates, Inc. 2005: 513-35.
10. Bauer DJ, Baldasaro RE, Gottfredson NC. Diagnostic procedures for detecting nonlinear relationships between latent variables. Struct Equ Modeling 2012; 19: 157-77. [CrossRef]
11. Afifi CA, Clark VA, May S. Computer Aided Multivariate Analysis. Chapman&Hall/CRC. 2004: 111-113.
12. Jacoby WG. Loess: a nonparametric, graphical tool for depicting relationships between variables. Electoral Studies 2000; 19: 577-613. [CrossRef]