Document Preview

Copyright information

Copyright Ortadogu Reklam Tanitim Yayincilik Turizm Egitim Insaat Sanayi ve Ticaret A.S. 2023

Access to the complete full text

This is a short preview of the document. Your library or institution may give you access to the complete full text for this document in ProQuest.

Full Text | Scholarly Journal

A Novel Variable Selection Procedure for Binary Logistic Regression Using Akaike Information Criteria Testing: An Example in Breast Cancer Prediction: Methodological Study (Validity Study)

Alternate title: Akaike Bilgi Kriterleri Testi Kullanılarak İkili Lojistik Regresyon için Yeni Bir Deg˘işken Seçim Prosedürü: Meme Kanseri Tahmininde Bir Örnek: Metodolojik Çalışma (Geçerlilik Çalışması)

Olaniran, Oyebayo Ridwan; Olaniran, Saidat Fehintola. Türkiye Klinikleri Biyoistatistik**; Balgat** Vol. 15, Iss. 2, (2023): 91-101. DOI:10.5336/biostatic.2023-97597

Türkiye Klinikleri Biyoistatistik Dergisi
Turkiye Klinikleri Journal of Biostatistics
Turkiye Klinikleri J Biostat. 2023;15(2

ORIGINAL RESEARCH ORİJİNAL ARAŞTIRMA

DOI: 10.5336/biostatic.2023

A Novel Variable Selection Procedure for Binary Logistic Regression Using Akaike Information Criteria Testing: An Example in Breast Cancer Prediction: Methodological Study (Validity Study)

Akaike Bilgi Kriterleri Testi Kullanılarak İkili Lojistik Regresyon iç Yeni Bir Değişken Seçim Prosedürü: Meme Kanseri Tahmininde Bir Örnek: Metodolojik Çalışma (Geçerlilik Çalışması)

Oyebayo Ridwan OLANIRAN[a], Saidat Fehintola OLANIRAN[b]

[a]Department of Statistics, Faculty of Physical Sciences, University of Ilorin, Ilorin, PMB 1515, Kwara State, Nigeria
[b]Department of Statistics and Mathematical Sciences, Faculty of Pure and Applied Science, Kwara State University, PMB 1530 Malete Kwara State, Nigeria

ABSTRACT Objective: Breast cancer is a leading cause of cancer-related death among women worldwide, with approximately 2.3 million new cases and 685,000 deaths reported in 2020 alone. One critical step in developing effective classification and prediction models is variable selection, which involves identifying a subset of relevant variables from a larger set of potential predictors. Accurate variable selection is crucial for building interpretable and robust models that are not overfit to noise, leading to improved model performance and generalization ability. In this paper, we proposed an alternative objective approach for comparing two Akaike Information Criterions (AIC) that originated from two competing models, such that the magnitude of the difference is subjected to the statistical test of significance. **Material and Methods:** We developed a new backward elimination variable selection procedure similar in spirit to the existing "stepAIC" within the environment of R statistical software. We used both simulated and Wisconsin breast cancer diagnostic datasets to compare the proposed method's variable selection and predictive performances with "stepAIC" and LASSO. **Results:** The simulation showed that the proposed AIC procedure achieved higher variable selection sensitivity, specificity and accuracy when compared to stepAIC and LASSO. Also, the proposed AIC method's prediction results are relatively comparable with stepAIC and LASSO at various simulated data dimensions. Similar supremacy results were observed with the breast cancer dataset used. **Conclusion:** The AIC-based variable selection approach proposed is a promising method that integrates AIC with statistical testing for improved variable selection in breast cancer classification and prediction.

Keywords: Breast cancer; Akaike Information Criteria; variable selection; backward selection; LASSO

ÖZET Amaç: Göğüs kanseri, yalnızca 2020 yılında bil yaklaşık 2,3 milyon yeni vaka ve 685.000 ölüm ile dünya kadınlar arasında kanser ilişkili ölümlerin başında gelen seb den biridir. Etkili sınıflandırma ve tahmin modelleri geliş kritik bir adım, daha geniş bir potansiyel öngörücü setinde değişken alt seti tanımlamayı içeren değişken seçimdir. Do ğişken seçimi, gürültüye fazla uyum sağlamayan, yorumlan sağlam modeller oluşturmada çok önemlidir. Bu durum model performansı ve generalizasyon becerisi sağlar. Bu m 2 rakip modelden oluşan 2 Akaike Bilgi Kriterleri'ni Information Criterions (AIC)] karşılaştırdığımız alternatif bir yaklaşım sunduk, öyle ki farkın büyüklüğü istatistiksel lık testine tabi tutulmuştur. **Gereç ve Yöntemler:** R istatist lımı ortamında bulunan "stepAIC"ye benzer yeni bir geriy eleme değişken seçme prosedürü geliştirdik. Sunulan meto ğişken seçimi ile "stepAIC" ve LASSO ile tahmini performa karşılaştırmak için simüle edilmiş, Wisconsin meme kans veri setlerini kullandık. **Bulgular:** Simülasyon, sunulan AI dürünün stepAIC ve LASSO'ya kıyasla yüksek değişke hassasiyeti, spesifitesi ve doğruluğu kazandığını göstermişti ca, sunulan AIC yönteminin tahmin sonuçları, simüle edile veri boyutlarında stepAIC ve LASSO ile görece karşılaştı dir. Kullanılan meme kanseri veri setinde de benzer üstünlül ları gözlemlenmiştir. **Sonuç:** AIC temelli değişken seçim mı, meme kanseri sınıflandırması ve tahmininde AIC'yi değişken seçimi için istatistiksel testlere entegre eden, um bir metottur.

Anahtar Kelimeler: Meme kanseri; Akaike Bilgi Kriterleri değişken seçimi; geriye dönük seçim;