Comparison of the Methods Used in Multi-Case Multi-Reader ROC Studies

Gülhan OREKÇİ TEMEL¹, Emine Arzu KANIK¹
¹Biyoistatistik AD, Mersin Üniversitesi Tip Fakültesi, Biyoistatistik Anabilim Dalı, Mersin

Abstract

Objective: The goals of this study are to compare the methods used in diagnostic accuracy studies in which both multi-case and multi-reader methods are employed, and to make useful suggestions for the proper use of these methods.

Material and Methods: When diagnosis tests are carried out by a single reader, they might have subjectivity based upon the reader’s personal decisions. The readers’ performance is related to their experiences, education and present states of the cases. Therefore, while the accuracy of the diagnostic tests is discussed, the reports of multi-readers should be taken into consideration within the scope of the model. There are five different methods used in multi-case and multi-reader tests of diagnostic accuracy studies.

Results: There are models dealing with the results of both multi-reader and multi diagnosis tests demonstrating the accuracy. These models were compared and presented in the tables.

Conclusion: The most common models are DBM (Dorfan Berbaum Metz) and OR (Obuchowski ve Rockett) methods.

Key Words: DBM; OR; multi reader multi case receiver operating curve

Özet


Bulgular: Hem çoklu değerlendirme hem de çoklu tanı testi sonuçlarının ele alınışı doğruluğu ortaya koyan modeller mevcuttur. Bu modeller karşılaştırılmış ve tablo halinde sunulmuştur.

Sonuç: Bu modellerden en yaygın kullanılanları DBM (Dorfan Berbaum Metz) ve OR (Obuchowski ve Rockette) metodudur.

Anahtar Kelimeler: DBM; OR; çok testli çok değerlendirme egrisi altında kalan alan
Giriş


Çok Testli Çok Değerlendirici ROC Analizi

Deneme Tasarımı

ČTÇD ROC çalışmalarda en genel deneme tasarım R değerlendircinin C vakayi K tanı testini birlikte değerlendirdiği faktöryel deneme tasarımıdır (2,3).

Kabul edelim ki, R değerlendirci, C vaka ve K tanı testi var olsun. Bu durumda her différencenin ÇÇK tane sonucu ve bütün çalısmının ise RÇKÇK tane sonucu vardır. ČTÇD analizlerinin deneme düzenleri bir tür faktöryel deneme düzenidir. Her değerlendirci bir ya da birden fazla test sonucunu değerlendirebilir. Tablo 1 de ČTÇD ROC analizi için deneme tasarımını göstergesidir.

| Tablo 1: ČTÇD Doğruluk Analizleri için Faktöryel Deneme Tasarımı |
|-----------------------|-----------------------|-----------------------|
|  | 1       | 2 | .... | K |
| Vaka (Hasta/Sağlam) | Değerlendirici | Değerlendirici | Değerlendirici | Değerlendirici |
|  | R₁ | R₂ | R₃ | R₁ | R₂ | R₃ | R₁ | R₂ | R₃ |
| 2 | Y₁₁₁ | Y₁₂₁ | Y₁₃₁ | Y₁₁₂ | Y₁₂₂ | Y₁₃₂ | Y₁₁₃ | Y₁₂₃ | Y₁₃₃ |
|  | Y₂₁₁ | Y₂₂₁ | Y₃₂₁ | Y₂₁₂ | Y₂₂₂ | Y₃₂₂ | Y₂₁₃ | Y₂₂₃ | Y₃₂₃ |
| C | Y₁₁₁ | Y₁₂₁ | Y₁₃₁ | Y₁₁₂ | Y₁₂₂ | Y₁₃₂ | Y₁₁₃ | Y₁₂₃ | Y₁₃₃ |
|  | Y₂₁₁ | Y₂₂₁ | Y₃₂₁ | Y₂₁₂ | Y₂₂₂ | Y₃₂₂ | Y₂₁₃ | Y₂₂₃ | Y₃₂₃ |
|  | Y₃₁₁ | Y₃₂₁ | Y₃₃₁ | Y₃₁₂ | Y₃₂₂ | Y₃₃₂ | Y₃₁₃ | Y₃₂₃ | Y₃₃₃ |


Çok testli çok değerlendircili çalışmalarda hasta ve değerlendirciler için rasgele etkili model kullanılır. Tanı testleri için ise etkilerin faktörler arasında değişmediğini varsayan sabit etkili model kullanılır.


DBM Metodu Teorisi

Çok Testli Çok Değerlendirici ROC

(1) \[ F_{DBM} = \frac{MS(T \times R)_{\text{pseudo}} + MS(T \times C)_{\text{pseudo}}}{MS(T \times R \times C)_{\text{pseudo}}} \]

MS(T): Tam etkinisenin kareler toplamı
MS(T\times R): Tam-değerlendirici interaksiyon etkinisinin kareler toplamı
MS(T\times R\times C): Tam-bir interaksiyon etkinisinin kareler toplamı

Karar değerleri \( F_{\text{değer}} = (F_{\text{değer}}, \alpha) \) ve tablo değerinden büyükse tanır testlerin diagnostic doğrulukları arasındaki farklılık anlamıdır. F istatistikine karşılık gelen kritik tablo değerlerinin hesaplaması eşitlik (4) ve eşitlik (5)'de verilmiştir.

\[ df = (t-1) \]

\[ df_{\text{sanat}} = \frac{(MS(T \times R)_{\text{pseudo}} + MS(T \times C)_{\text{pseudo}} - MS(T \times R \times C)_{\text{pseudo}})^2}{(t-1)(t-2)-
\]

(5) ÇTCD ROC analizlerinde performans değeri eğri altında kalan alan (AUC) 4,7-11.

Obuchowski ve Rockette Metodunun
(Düşeltilmiş F Metodu) Teorisi

Obuchowski ve Rockette tarafından OR metodları kullanılarak ve düzeltilmiş F testi olarak da ifade edilen matematiksel bir model önerilmiştir. Modelde değerlendirilere sabit etkili, tanı testleri ve vakalar rasgele etkili kullanılarak tahmin yapılmıştır. Buradaki doğruluk değerleri, DBM metodundaki gibi sözdeler değerler kullanılarak değil, gerçek değerler kullanılarak yapılmıştır. DBM metoduna açık vakaların çoklu değerlendirmeleri birinin performansını çoklu tekrarlamalar izin verir. Faktor düzeltilmiş F testi faktörlerin deneme taramasını yapan bir replikasyonda hesaplanır ve Denklem 6'daki gibi yazılabilir.

\[ \delta_{\text{testi}} = \mu + \tau + R + \epsilon \]
\( R \) ve \((\varepsilon R)\) ortalamalar sıfır varyansları \( \sigma_{R}^{2} \) ve \( \sigma_{\varepsilon R}^{2} \) olan ortak bağımsız normal dağılım gösterdikleri varsayın. Estistik 2.6 daki \( \varepsilon_{ij} \) de ortalamalar sıfır ve varyansı sabit \( \sigma_{\varepsilon}^{2} \) normal dağılımı varsayılır.

Ayrıca \( \varepsilon_{ij}; R \) ve \((\varepsilon R)\) da bağımsız olduğu varsayılır. Faktat her tani için her değerlendiricinin aynı vakayı okumasi sonucunda, \( \varepsilon_{ij} \)'nin ortak bağımsız olduğu varsayımın kabul etmeyeceğiz. Bunun yerine testler ve değerlendiriciler arasındaki hataları kovaryansları oluşturular ve Denklem 7'deki gibi bu kovaryansları ifade edilir.

\[
\text{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = \begin{bmatrix}
\text{Cov}1 & i \neq i' & j = j' \\
\text{Cov}2 & i = i' & j = j' \\
\text{Cov}3 & i \neq i' & j \neq j'
\end{bmatrix}
\]  \hspace{1cm} (7)

Obuchowski ve Rockette (15,16) bu kovaryansları Denklem 8'deki gibi sıralamadıkları,

\[
\text{Cov} \geq \text{Cov}2 \geq \text{Cov}3 = 0
\]  \hspace{1cm} (8)

Bu kovaryans değerlerinden: Denklem 9 farklı tanı testleri-aynı değerlendirici için oluşturulan \( \text{Cov}_{1} \) 'n'in, Denklem 10 aynı tanı testi-aynı değerlendirme için oluşturulan \( \text{Cov}_{2} \) 'n'in, Denklem 11'de farklı tanı testi-farklı değerlendirme için oluşturulan \( \text{Cov}_{3} \) değerinin hesaplanması göstergemektedir.

\[
\text{Cov}1 = \text{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = \text{Cov}(\hat{\beta}_{ij}, \hat{\beta}_{i'j'}) / R \text{R}
\]  \hspace{1cm} (9)

\[
\text{Cov}2 = \text{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = \text{Cov}(\hat{\beta}_{ij}, \hat{\beta}_{i'j'}) / R \text{R}
\]  \hspace{1cm} (10)

\[
\text{Cov}3 = \text{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = \text{Cov}(\hat{\beta}_{ij}, \hat{\beta}_{i'j'}) / R \text{R}
\]  \hspace{1cm} (11)

Buradaki kovaryans \( \text{Cov}(/ R \text{R}) \); şart değerlendirici ve tanı-değerlendirici etkisinin kovaryansıdır. Eğer hatalar bağımsız ise bir tannın etkisi test etmek için \( (H_{0} : \tau_{1} = \tau_{2} = ... = \tau_{i}) \) Denklem 12'de kullanılan F istatistiği kullanılır.

\[
F = \frac{\sum_{i=1}^{I} \hat{\beta}_{ij} - \hat{\beta}_{j} \cdot (t-1)}{\sum_{i=1}^{I} \hat{\beta}_{ij} - \hat{\beta}_{ij} - \hat{\beta}_{j} + \hat{\beta}_{j} \cdot (t-1)(t-r-1)} = \frac{\text{MS}(T)_{\hat{\beta}_{ij} \text{varial}}}{{\text{MS}(T \ast R)_{\hat{\beta}_{ij} \text{varial}} + r(\text{cov}2 - \text{cov}3)}(14)}
\]

Kullanmasından F değeri \( (F_{k, v_{k}} ; t-1, a) \) tablo değerinden büyükse tanı testlerinin diagnostik doğrulukları arasında farklılık anlamılır. F istatistiğine karşılık gelen kritik tablo değerlerinin hesaplanması Denklem 15 ve 16'da verilmiştir (7,9,12-14).

\[
df_{1} = (t-1)
\]  \hspace{1cm} (15)

\[
df_{2} = (t-1)^{2}(t-1)
\]  \hspace{1cm} (16)

**Çok Değişkenli Wilcoxon-Mann-Whitney Yöntemi**


**ROC Deneme Tasarımını İçin Hiyerarşik Sıralı Regresyon Metodu**


**Varyans Unsurlarının Bootstrap Tahmini**


Modelde veriler bootstrap yöntemi öncelikle metodunun kullanılacağı elde edilir ve varyans analizindeki varyasyon kaynaklarını bu yöntelerle hesaplanır.

Çok testli çok değerlendircili ROC analizlerinde kullanılan beş yöntem model, değerlendirmelerin modele ektsi, kullanılan modelin herhangi bir kovaryatında iyi verip vermemesi, modele kullanılan doğruluk ölçümü ve bir bilgisayar yazılımlarının olup olmadığını yordamı Tablo 2'de karşılaştırılmıştır (20-21).

<table>
<thead>
<tr>
<th>özellikler</th>
<th>DBM</th>
<th>OR</th>
<th>Çok değişkenli WMW</th>
<th>BWC</th>
<th>HROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Kısımı</td>
<td>Sözu Değerler</td>
<td>Doğrulüğün özet ölçümu</td>
<td>Her hasta için güven skoru</td>
<td>Doğrulüğün özet ölçümu</td>
<td>Her hasta için gizli değişken</td>
</tr>
<tr>
<td>Doğruluk ölçümü</td>
<td>DIAGNOSTİK istatistiklerden herhangi biri kullanılır.</td>
<td>DIAGNOSTİK istatistiklerden herhangi biri kullanılır.</td>
<td>Etiği altında kalan alan kullanılır.</td>
<td>DIAGNOSTİK istatistiklerden herhangi biri kullanılır.</td>
<td></td>
</tr>
<tr>
<td>Yazarım</td>
<td>LABMRMRC (22)</td>
<td>DBB MRMC 2.0 (23)</td>
<td>Bilgisayar yazılımı yoktur.</td>
<td>Bilgisayar yazılımı yoktur</td>
<td>Bilgisayar yazılımı yoktur</td>
</tr>
</tbody>
</table>

Sonuç

Çok testli çok değerlendircili diagnostik doğruluk çalışmalarda kullanılabilecek beş farklı yönteme rastlanmıştır. Bu metotlardan DBM metodu veri setini sözu değerlere kullanarak yeniden örneklemeye yapmaya, modele kovaryatları dahil edebilmesi, diagnostik performans istatistiklerinden herhangi birini kullanabilmesi açısından diğer metotlara üstünlük sağlamaktadır. DBM ve OR metotlarını uygulamada kullanılan birden fazla yazılımla internetten ücretsiz erişim ekmek mümkündür. Çok değişkenli WMW, BWC, HROC metotlarının herhangi bir bilgisayar yazılımlarının olmasası bu yöntemlerin dezavantajıdır. Bu üç yöntemin bilgisayar yazılımlarının oluşturulması önerilmektedir.

Kaynaklar


