ロジスティックモデルとROC AUC分析を組み合わせた検査性能の評価と疫学基本モデル評価方法

古川敏仁、杉本典子
株式会社　バイオスタティスティカルリサーチ

Test Performance Evaluation in Epidemiological Basic Model Using ROC AUC with logistic regression

Toshihito Furukawa, Noriko Sugimoto
Biostatistical Research Co.,LTD.

SASユーザー総会 アカデミア/テクノロジー&ソリューション セッション 2009

要旨:

健常群、疾患群を診断する検査の性能評価のためには、両群のリスク背景因子いわゆる基本モデルを考慮したROC AUC分析が必要であり、それはロジスティック多変量解析における診断能の定量的評価を可能とする方法である。

キーワード:検査診断能　ROC　AUC　疫学　基本モデル　logistic model

検査値 X の目的

- 例: 診断
  ある閾値cをもとに疾患(Disease)と正常(Health)を区分する

  もし、X＞c ならば 疾患と判定
  もし、X≦c ならば 正常と判定

- 例:予後の予測
  ある閾値cをもとに予後良好(Survival)と不良(Death)を区分する

  もし、X＞c ならば 生存率が高いと判定
  もし、X≦c ならば 生存率が低いと判定

診断性能評価上の問題

- ある閾値 c をもとにした性能判定の限界
  感度(Sensitivity)、特異度(Specificity)、
  正確度(Accuracy)
- 多変量鑑別モデル(例:ロジスティックモデル)、有意な項目の組み合わせはわかっても、その項目の診断性能への寄与は分かりずらい
- 疫学的な問題
  そもそも、他の予後因子(背景因子)で説明される以上の臨床的な有用性がその検査には存在するか

問題解決

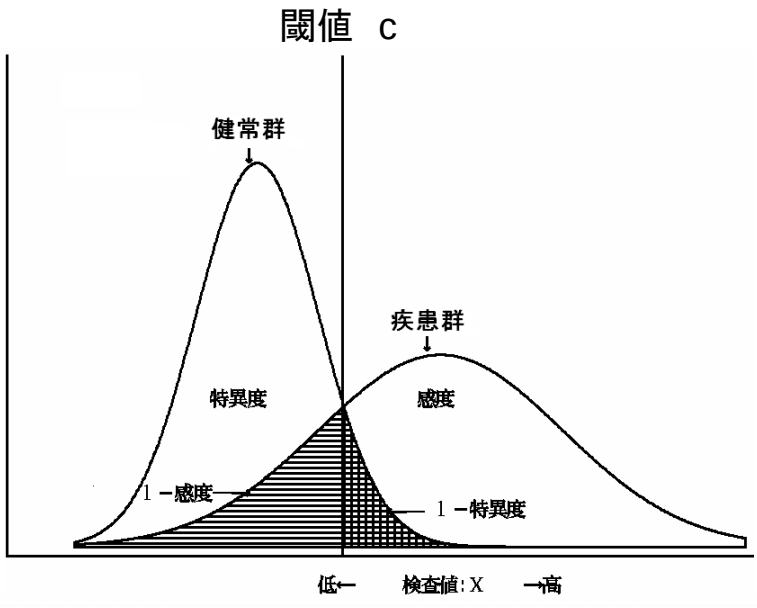
- 今回はこれらの問題をROCのAUCを用いて解決します。
- 疫学的には基本モデルの説明をします。

- 同様の問題を生存時間の予後判定や、Cox回帰を用いた場合の背景因子を考慮した予後検査診断能の評価に拡張いたします。

検査Xの性能指標の定義とROCについて

Contents

・診断検査Xの評価指標
　　感度、特異度、正確度
・ROCとAUCの説明
・ROC の分散推定
・2つの検査AUCの差の検定



検査値Xをある閾値cで診断する場合の検査性能指標の定義

- 疾患群（Disease）の例数m人、
  健常群（Health） の例数n人、
  全体で N=m+n人
- 感度（Sensitivity）
  疾患群m人中、検査値Xがcを超える人の割合

$$sens(c) = \frac{1}{m}\sum_{i=1}^{m} I(Xi > c) = \widehat{P}(Xi > c \mid Disease)$$

$$I(Xi > c): \text{陽性} = 1、\ \text{陰性} = 0$$

検査値Xをある閾値cで診断する場合の検査性能指標の定義

- 特異度（Specificity）
  健常群n人中、検査値Xがc以下の人の割合

$$spec(c) = \frac{1}{n}\sum_{j=1}^{n} I(Xj \le c) = \hat{P}(Xj \le c \mid Health)$$

- 正確度（Accuracy）
  検査を受けたN人が、疾患群は陽性、健常群は陰性と正しく診断された割合

$$acc(c) = \frac{1}{N}\left(\sum_{i=1}^{m} I(Xi > c) + \sum_{j=1}^{n} I(Xj \le c)\right) = \hat{P}c(\text{正しく診断})$$

検査性能指標の問題
例：疾患A　検査X

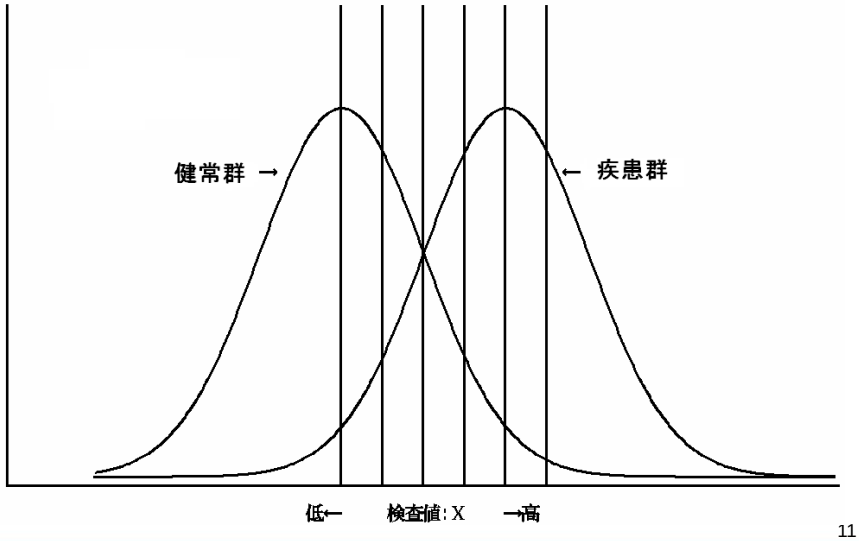
	疾患 200例		健常人 1000例		全体 1200例	
論文基準値	陽性	陽性率	陰性	陰性率	正診	正診率
18以下	151	75.5%	633	63.3%	784	65.3%
40以下	109	54.5%	851	85.1%	960	80.0%
100以下	73	36.5%	960	96.0%	1033	86.1%
250以下	25	12.5%	993	99.3%	1018	84.8%

- カットオフの設定により検査性能値は異なる
  =感度と特異度はトレードオフの関係
- 疾患群と健常群の比により正診率は異なる

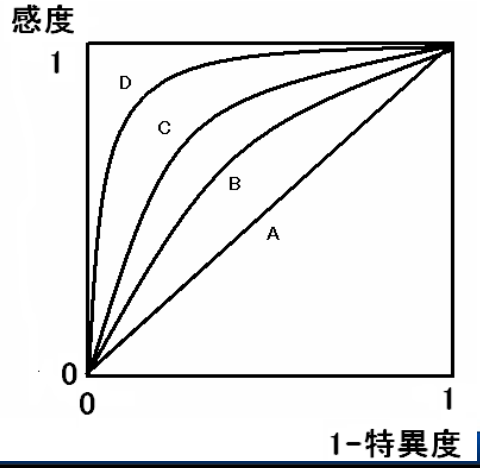
ROC曲線

- カットオフを連続的に変化
- 縦軸：感度　　横軸：1－特異度
- 曲線が左上角に近いほど検査性能が高い
- 曲線が対角線上＝診断能力はない

・カットオフを連続的に変化



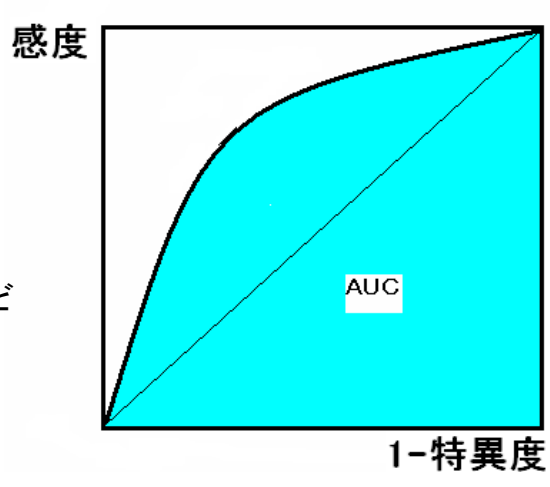
- 縦軸：感度　横軸：1－特異度
- 曲線が左上角に近いほど検査性能が高い
- 曲線が対角線上＝診断能力はない



12

ROCのAUC（Area Under the Curve）

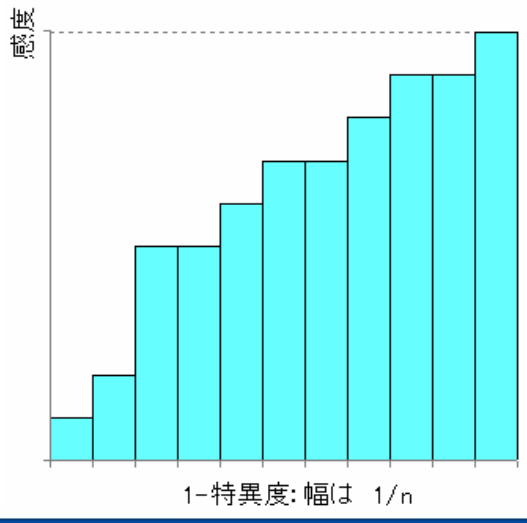
- AUC=1
  完全な検査
- AUC=0.5
  無意味な検査
- AUCは1.0に近いほど
  良い検査



13

AUCの重要な性質

- AUC　台形法
- 健常人に着目



AUC（台形法）の重要な性質：感度

- 健常人に着目：健常人n人を検査値Xjの小さい順にならべ、個々のXjをカットオフとしたときの感度sens（Xj)を台形法にて求めると

$$AUC = \sum_{j=1}^{n} \frac{1}{n} sens(Xj) \qquad (1)$$

$$= \sum_{j=1}^{n} f(Xj)sens(Xj) = E(sens)$$

となり、AUCは感度の期待値となることがわかる。

$$sens(Xj) = \{(m + R(H)j - R(S)j)/m\}$$

であることから式（1）は

$$AUC = \frac{1}{nm} \sum_{j=1}^{n} (m + R(H)j - R(S)j) \qquad (2)$$

$R(H)j$：健常人(H)n人中のjの順位

$R(H)j$ ：全例(S)n＋m人中のjの順位

AUC（台形法）の重要な性質：特異度

$$AUC = \sum_{i=1}^{m} \frac{1}{m} sens(Xi) = E(spec) \qquad (3)$$

となり、AUCは得意度の期待値でもあることがわかります。

$$spec(Xi) = \{(R(S)i - R(D)i)/n\}$$

であることから式(3)は

$$AUC = \frac{1}{nm} \sum_{i=1}^{m} (R(S)i - R(D)i) \qquad (4)$$

$R(D)i$ ：疾患群(D)m人中のiの順位

$R(S)i$ ：全例(S)n＋m人中のiの順位

AUCの分散

- AUCの分散はAUCが感度、特異度の期待値であることから経験的に以下に求めることができる。

AUCの分散

$$\mathrm{var}(AUC) = \frac{1}{m(m-1)} \sum_{i=1}^{m} (spec(Xi) - AUC)^2 + \frac{1}{n(n-1)} \sum_{j=1}^{n} (sens(Xj) - AUC)^2$$

同一症例に対し同時に測定された検査のAUC比較

- 今、検査X、検査Yが同一症例に対し同時に測定されたと仮定し、検査XのROC AUCをAUCx、検査YのAUCをAUCyとする。
- 臨床的には AUCx、AUCyの差がしばしば問題となる。
- Dif（AUC）=AUCx-AUCy

AUCの比較の検定

$$\mathrm{var}(Dif(AUC)) = \mathrm{var}(AUCx) + Var(AUCy) - 2\,\mathrm{cov}(AUCx, AUCy)$$

$$\mathrm{cov}(AUCx, AUCy) = \frac{1}{m(m-1)} \sum_{i=1}^{m} (specx(i) - AUCx)(specy(i) - AUCy) + \frac{1}{n(n-1)} \sum_{j=1}^{n} (sensx(j) - AUCx)(sensy(j) - AUCy)$$

また、Delong[1]らは、この経験的分散に基づく下記の統計量が自由度1のχ2乗分布に従うことを示している。

$$\frac{Dif(\mathrm{AUC})}{\mathrm{var}(Dif(AUC))}$$

ROC AUCの疫学データへの応用

Contents
- ・基本モデルとは
- ・基本モデルと検査性能
- ・ロジスティック変数選択とROC AUC

基本モデルとは

- 近年の大規模データに基づく疫学研究の進展により
  疾患ごとの被験者背景要因のリスクが明確になりつつある
  この疾患ごとの被験者リスクモデルを基本モデルとここでは呼ぶ

- 例：メタボリックシンドロームと成人病基本リスク
  ウエスト周囲径が男性で85cm、女性で90cm以上かつ
  下記が2つ以上該当
  血清脂質異常（例：トリグリセリド値150mg/dL以上、または
  HDLコレステロール値40mg/dL未満）
  血圧高値（例：SBP130mmHg以上、またはDBP85mmHg以上）
  高血糖（例：空腹時血糖値110mg/dL）

検査性能評価上の問題点

- 検査性能は、疾患群の感度、健常群の特異度をもとに評価される
- もともと、健常群と疾患群では被験者背景（基本リスク）が違う可能性がある。
- 検査値が基本リスクと相関する場合、一見有効な診断検査であっても、同じ基本リスク集団では、診断能を持たない可能性がある。

22

ある検査の評価：疾患Aの診断
健常群　1000人、疾患群　200人

- Logistic　Regression
- 統計的に有意　Odds比　1.117（/10）　検査値が10高くなるとリスクは約1.1倍
- 検査値が100高くなるとOdds比3.00・・！

Odds比推定	推定			
推定	95%下限	95%上限	Wald χ2	p値
1.117	1.140	1.095	155	>0.0001

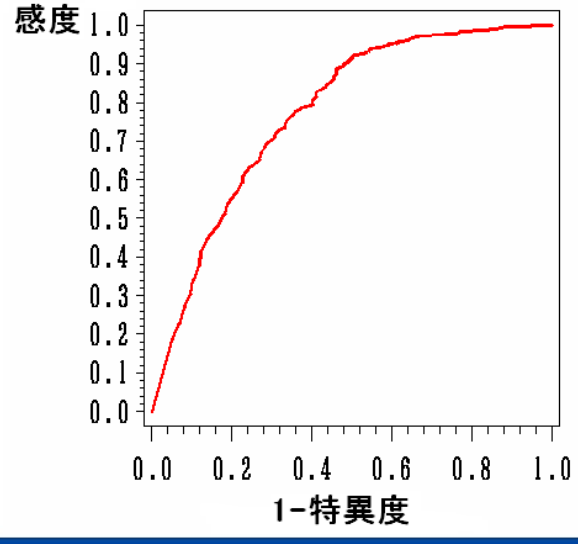
23

検査性能指標の問題
例：疾患A　検査X

論文基準値	疾患 200例 陽性	陽性率	健常人 1000例 陰性	陰性率	全体 1200例 正診	正診率
18以下	151	75.5%	633	63.3%	784	65.3%
40以下	109	54.5%	851	85.1%	960	80.0%
100以下	73	36.5%	960	96.0%	1033	86.1%
250以下	25	12.5%	993	99.3%	1018	84.8%

- 感度、特異度はこんな感じ
- 良い検査なのか、それとも・・・・

検査XのROC曲線　AUC＝0.775



集団の基本リスクを考えると ロジスティック多変量解析-Odds推定

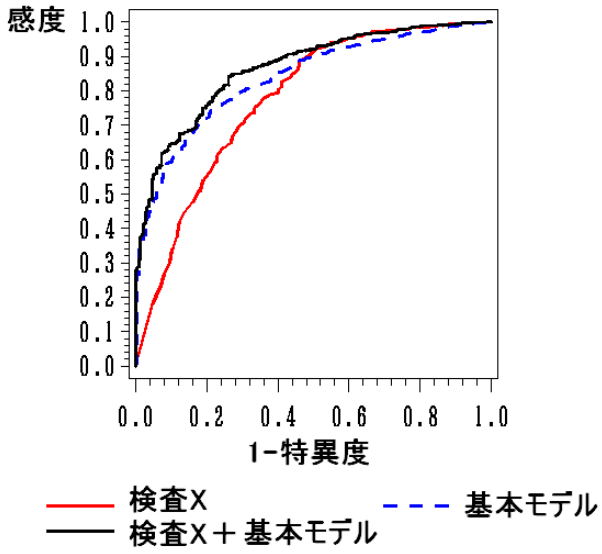
疾患Aの基本リスク

リスク因子	推定	95%下限	95%上限
年齢	0.915	0.902	0.929
性別	1.146	0.731	1.798
喫煙	1.978	1.261	3.103
高血圧	2.128	1.504	3.013
糖尿病	2.480	1.754	3.507
高コレステロール血漿	1.013	0.738	1.390

基本リスクを検査Rとして考える

- 多変量ロジスティック基本モデル
- Logit=Intercept+b1*年齢+b2*性別+b3*喫煙+b4*高血圧+b5*糖尿病+b6*高コレステロール血漿
- R=exp(logit)/(1+exp(logit))

ROC曲線の比較



検査Xの本当の性能？

- 検査XのAUCは0.775であった。
- しかし、患者集団の基本リスクによる診断でもAUCは0.850もあることがわかる
- 基本モデルに検査Xを加えたときのAUCは0.855で基本モデルより、わずかに0.005大きいだけであった。

		Confidence Intervals		基本モデルとの差
	AUC	Lower	Upper	
検査X	0.775	0.738	0.810	−0.075
基本モデル	0.850	0.813	0.861	−
基本モデル+検査X	0.855	0.844	0.888	0.005

検査診断能としての変数選択
ロジスティックモデルでは、直接的にどの程度診断能が向上したのかは分からない
Wald χ2のp値では、例数が多いと有益な情報は得られない。

AUC

AUCの差の検定	差の推定	95% Confidence Intervals		χ 2 p値
基本モデル+検査X−基本モデル	0.005	0.003	0.007	0.0007

ロジスティックモデル

	Odds比 推定 推定	95%下限	95%上限	Waldχ 2 p値
検査X	0.201	0.144	0.279	<.0001
年齢	0.945	0.930	0.960	<.0001
性別	1.190	0.750	1.888	0.460
喫煙	1.842	1.159	2.927	0.010
高血圧	2.146	1.497	3.079	<.0001
糖尿病	2.662	1.850	3.830	<.0001
高コレステロール血漿	1.132	0.814	1.575	0.461

ROC AUCの疫学データへの応用 結論(1)

- 検査の性能を評価する場合、特定の感度、特異度に影響されないROC（AUC）の評価は重要である。
- AUCは検査の感度、特異度、有病率50%時の正確度の期待値なので、検査性能の理解しやすい指標である。
- 特定の診断情報に検査Xの追加情報が臨床的に意味があるかを判断する場合、ロジスティックモデルでは、統計的に追加変数が有意かどうかは判定できても、どの程度診断能が向上したのかは分からない。
  基本モデルと基本モデル+検査XのAUCの差の評価が重要である。
- このAUC比較機能はVer9.2より、標準的にlogisticプロシジャに採用される。

ROC AUCの疫学データへの応用 結論(2)

- 検査の性能を評価する場合、健常群、疾患群間で、集団間の疾患に対してリスク要因となる背景因子が違うことを考慮しなければならない。
- リスク要因と検査値が相関する場合、検査診断性能が正しく評価されない場合がある。
  上記を確認するためには、リスク要因のみによる診断能とリスク要因+検査時の診断能をAUCで比較する必要がある。
- 疫学研究が進展するにつれ、従来有用とされていた検査が、実はリスク要因との単なる交絡を反映する事象であることが示される可能性がある。統計担当者は充分そのことを理解する必要がある。

参考文献

- [1] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach.Biometrics. 1988;44:837-845.
- [2] Li Lu,Chenwei Liu. Using the Time Dependent ROC Curve to Build Better Survival Model in SAS.NESUG 2006

時間依存性ROC　AUCとCox回帰
次回予告

Contents

- ・時間依存性ROCの定義と臨床的意味
- ・時間依存性ROC の注意点
- ・多変量リスクのROC評価-Cox回帰
- ・PGxにおける遺伝子発現群の最大リスクの評価