

SASによる診断・検査 データの統計解析

阿部研自

佐伯浩之



ROC曲線のAUC間の
検定？なにそれ！
そんなマイナーな
手法しらな～い．



ぎく！臨床試験では、
あんまり使わないん
だよね（多分...）．
論文とかないの？



うーん．このままだ
と診断薬を扱う会社
の統計屋として恥ず
かしいね．お勉強し
ておきましょうか！

放射線科のドクター
からROC曲線のAUC
の比較してって言われ
たんだけどわかる？

えー！でもあなたの
会社，造影剤とか診
断薬の開発してる
よね～！

診断検査に関する包
括的なテキストが出
版されたから，一緒
にお勉強してみよー
よ！ねー．



2 × 2分割表データの解析

マンモグラフィによる乳癌のスクリーニング診断

| Cancer Status | Test Result | | Total |
|---------------|-------------|----------|----------|
| | Positive | Negative | |
| Present | 29 (a) | 1 (b) | 30 (a+b) |
| Absent | 19 (c) | 11 (d) | 30 (c+d) |
| Total | 48 (a+c) | 12 (b+d) | 60 (N) |

$$Se = \frac{a}{a+b}$$

$$FNR = \frac{b}{a+b} = 1 - Se$$

$$LR(+) = \frac{Se}{1 - Sp}$$

$$Sp = \frac{d}{c+d}$$

$$FPR = \frac{c}{c+d} = 1 - Sp$$

$$LR(-) = \frac{1 - Se}{Sp}$$

$$OR = \frac{LR(+)}{LR(-)} = \frac{Se \times Sp}{FNR \times FPR}$$

$$Prevalence = \frac{a+b}{N} \quad Accuracy = \frac{a+d}{N}$$

こんな感じの
統計量に要約
するよね．



SAS freq output

| | リスク | 漸近標準誤差 | 95% 信頼区間 (漸近) | | 95% 信頼区間 (正確) | | 95% 信頼区間 (Wilson) | |
|-----|--------|--------|--------------------|--------|--------------------|--------|------------------------|--------|
| 感度 | 0.9667 | 0.0328 | 0.9024 | 1.0000 | 0.8278 | 0.9992 | 0.8333 | 0.9941 |
| 偽陽性 | 0.6333 | 0.0880 | 0.4609 | 0.8058 | 0.4386 | 0.8007 | 0.4551 | 0.7813 |

| | リスク | 漸近標準誤差 | 95% 信頼区間 (漸近) | | 95% 信頼区間 (正確) | | 95% 信頼区間 (Wilson) | |
|-----|--------|--------|--------------------|--------|--------------------|--------|------------------------|--------|
| 偽陰性 | 0.0333 | 0.0328 | 0.0000 | 0.0976 | 0.0008 | 0.1722 | 0.0059 | 0.1667 |
| 特異度 | 0.3667 | 0.0880 | 0.1942 | 0.5391 | 0.1993 | 0.5614 | 0.2187 | 0.5449 |

| 研究の種類 | 値 | | 95% 信頼区間 |
|-------|---------|--------|----------|
| オッズ比 | 16.7895 | 2.0006 | 140.8977 |
| 陽性尤度比 | 1.5263 | 1.1533 | 2.0201 |
| 陰性尤度比 | 0.0909 | 0.0125 | 0.6608 |



へー . FREQの
Outputと感度 ,
特異度はこんな風
に対応してるのね !

SASの信頼区間で
WaldとClopper-
Pearsonだけだから
最近話題 ? のWilson法
も作ってみたよ



リンパ腫診断のPETとSPECTの比較 (真の状態と各診断検査法の分割表)

FDG PET

| | Positive | Negative | |
|---------|----------|----------|----|
| with | 42 | 3 | 45 |
| without | 11 | 28 | 39 |
| | 53 | 31 | 84 |

Wilson score CI

Sensitivity : 0.93 (0.821 , 0.977)

Specificity : 0.72 (0.562 , 0.835)

^{67}Ga SPECT

| | Positive | Negative | |
|---------|----------|----------|----|
| with | 33 | 12 | 45 |
| without | 19 | 20 | 39 |
| | 52 | 32 | 84 |

Wilson score CI

Sensitivity : 0.73 (0.590 , 0.840)

Specificity : 0.51 (0.362 , 0.661)

同一の患者で実施した2つの診断方法の感度や特異度を比較するには、どうしたらいいでしょう？



リンパ腫診断のPETとSPECTの比較 (疾患有無別の各診断検査法の対応のある比較)

Sensitivity (With Disease)

| | ⁶⁷ Ga SPECT | | |
|----------|------------------------|----------|----|
| FDG PET | Positive | Negative | |
| Positive | 32 | 10 | 42 |
| Negative | 1 | 2 | 3 |
| | 33 | 12 | 45 |

McNemar test : P=0.007

Wald CI : 0.200 (0.068, 0.332)

Wilson score CI : 0.200 (0.056, 0.344)

Specificity (Without Disease)

| | ⁶⁷ Ga SPECT | | |
|----------|------------------------|----------|----|
| FDG PET | Positive | Negative | |
| Positive | 7 | 4 | 11 |
| Negative | 12 | 16 | 28 |
| | 19 | 20 | 39 |

McNemar test : P=0.046

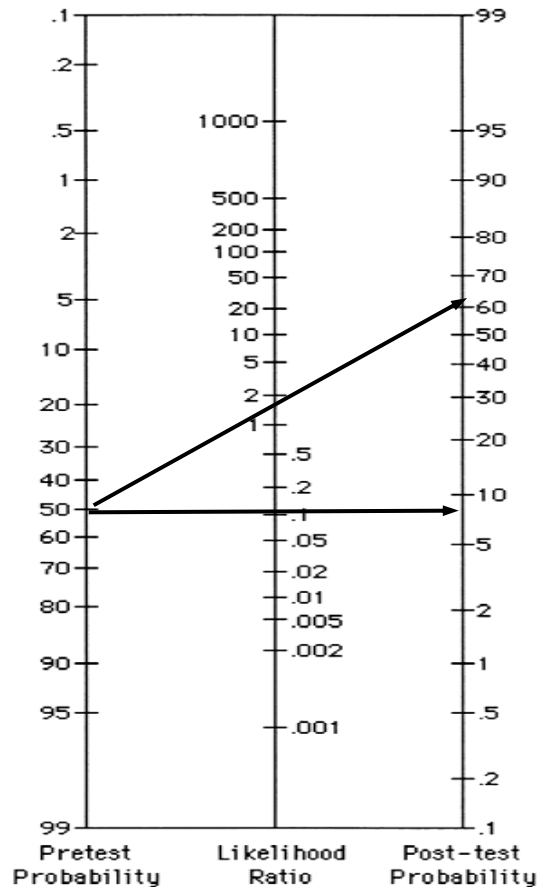
Wald CI : -0.205 (-0.396, -0.015)

Wilson score CI : -0.205 (-0.383, -0.004)

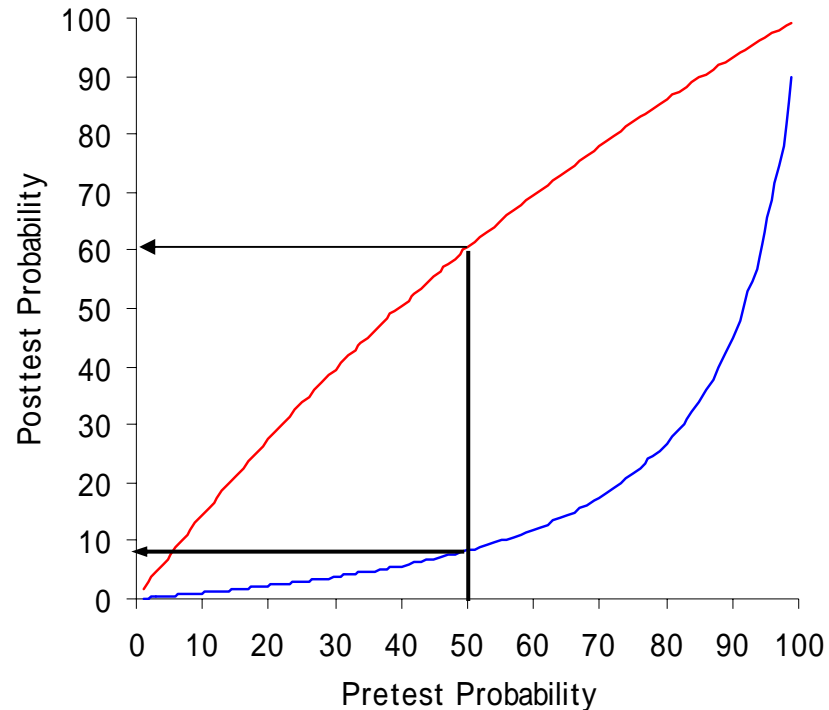


疾患有無別に，McNemar検定をすれば感度と特異度の比較になるんだね！対応のある比率の差の信頼区間は，SASではサポートしてないから，WaldとNewcombe-Wilson(CC)を作ってみたよ！

陽性・陰性予測値 (PPV, NPV) : FaganのノモグラムとBayes曲線



例：検査前確率（有病率）50%の患者さんが
感度96%，特異度37%の診断検査を受診して
陽性なら60%，陰性なら9%の検査後確率となる

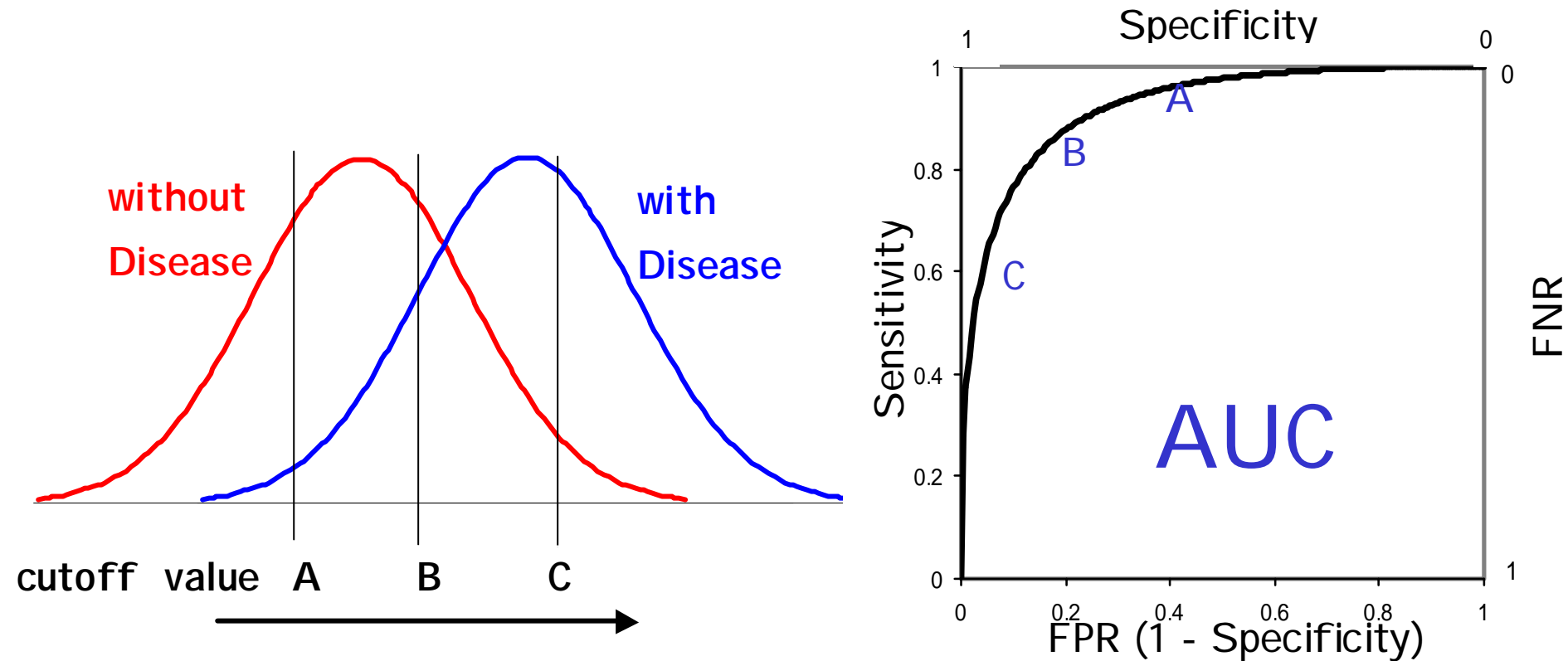


感度と特異度で検査
の性能は分るけど，
実際の医療現場では
予測値が重要だよね～

そうだよね～．でも
検査前確率の設定が
難しいよね．Bayes
曲線はlogisticプロシ
ジャで推定だね！



ROC曲線とAUCの推定(1)

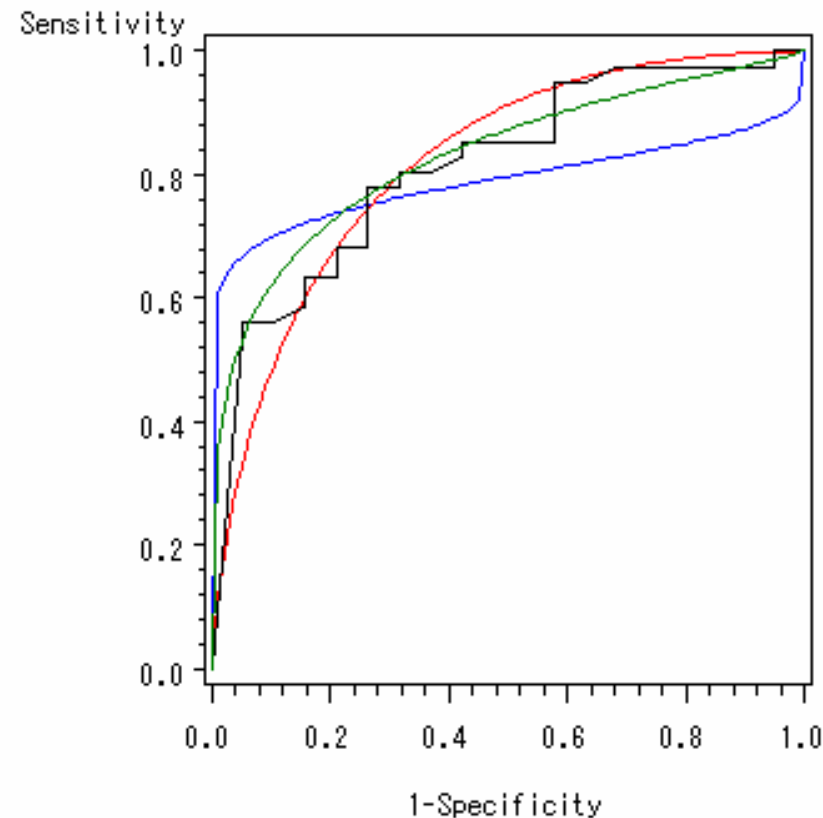


Cutoff値を連続的に動かして検査の性能を把握するんだね。
曲線下面積 (AUC) : 0.5 ~ 1.0でROC曲線を要約して
診断検査の無能, 有効を解釈することもできるね!



ROC曲線とAUCの推定(2)

CK-BBアイソザイムによる頭部外傷の予後予測



| method | | AUC | SE(AUC) |
|-----------|-------------------------|--------|---------|
| Empirical | DeLong et al. | 0.8286 | 0.0525 |
| | Bootstrap | | 0.0516 |
| Binormal | ROCKIT (MLE) | 0.8313 | 0.0519 |
| | direct(raw-data) | 0.7903 | 0.0525 |
| | direct(log-transformed) | 0.8194 | 0.0581 |

ROC曲線の推定は色々な方法があるよ！

ノンパラメトリック法 : logisticプロシジャ

Binormal法 (最尤推定法) : Metz's ROCKIT

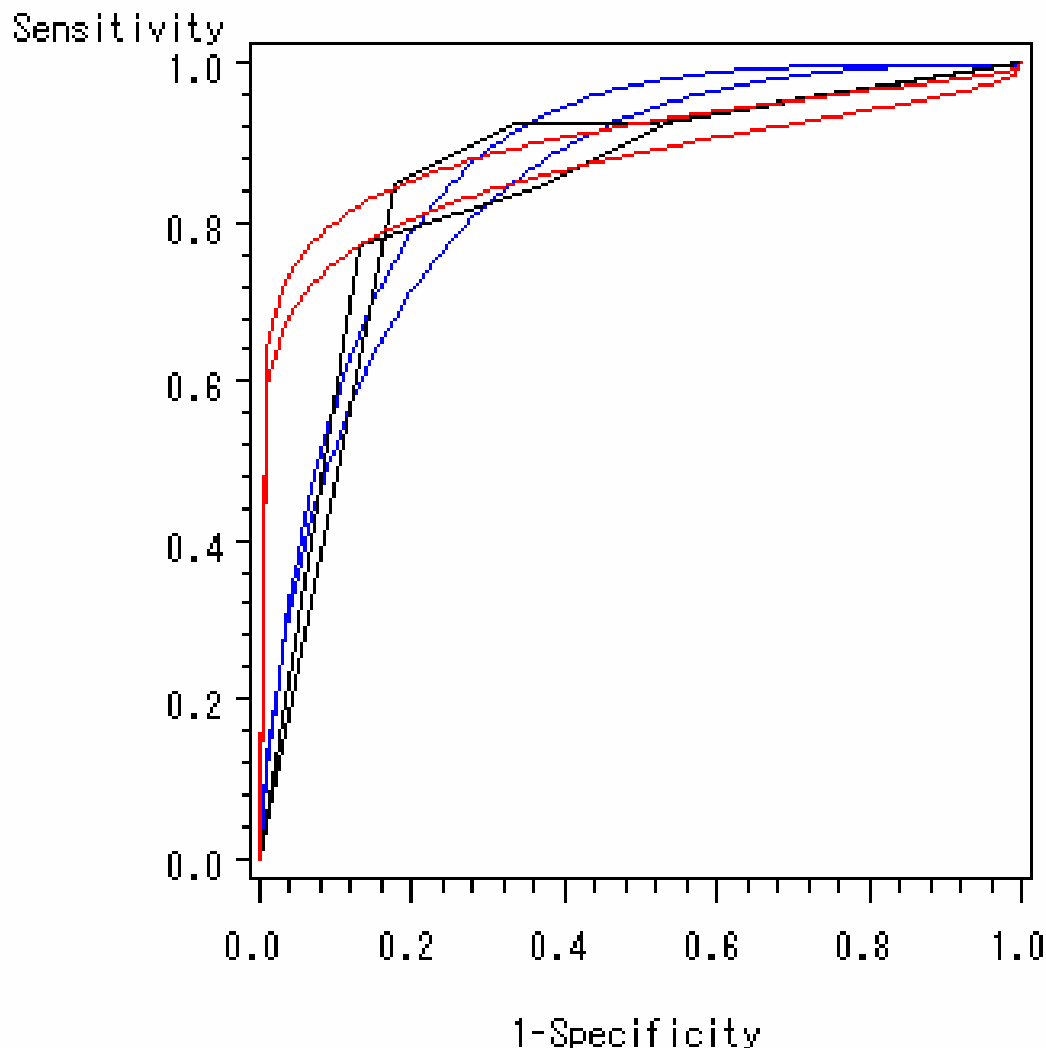
Binormal法 (直接推定法) : SAS sample code
etc.



診断検査法間のAUCの比較

マンモグラフィ (Plain and Digitized Film)

悪性・良性の確信度の比較



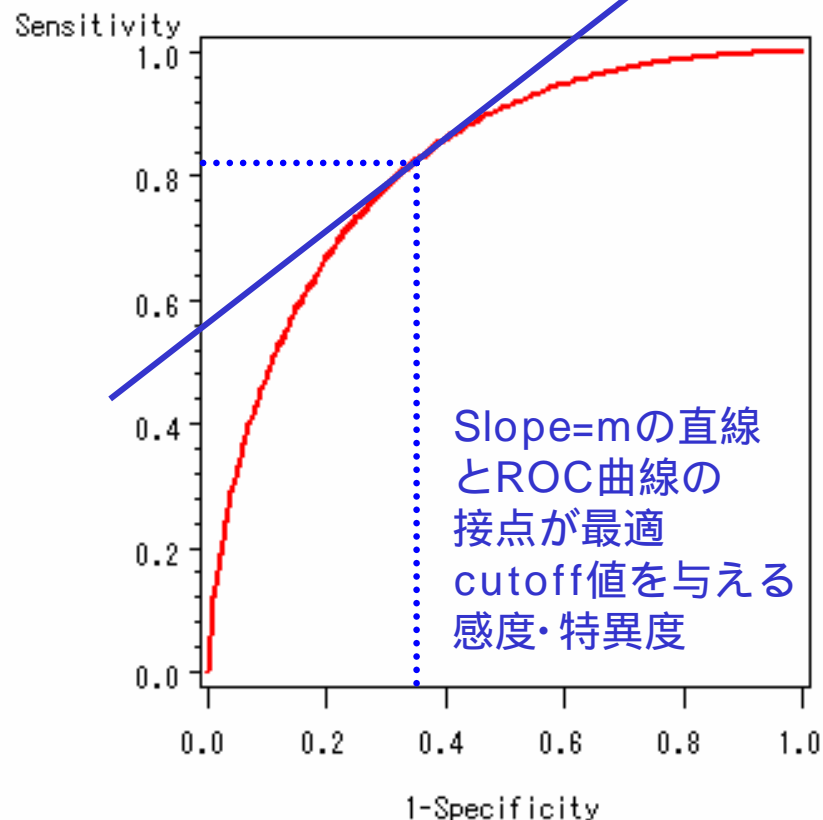
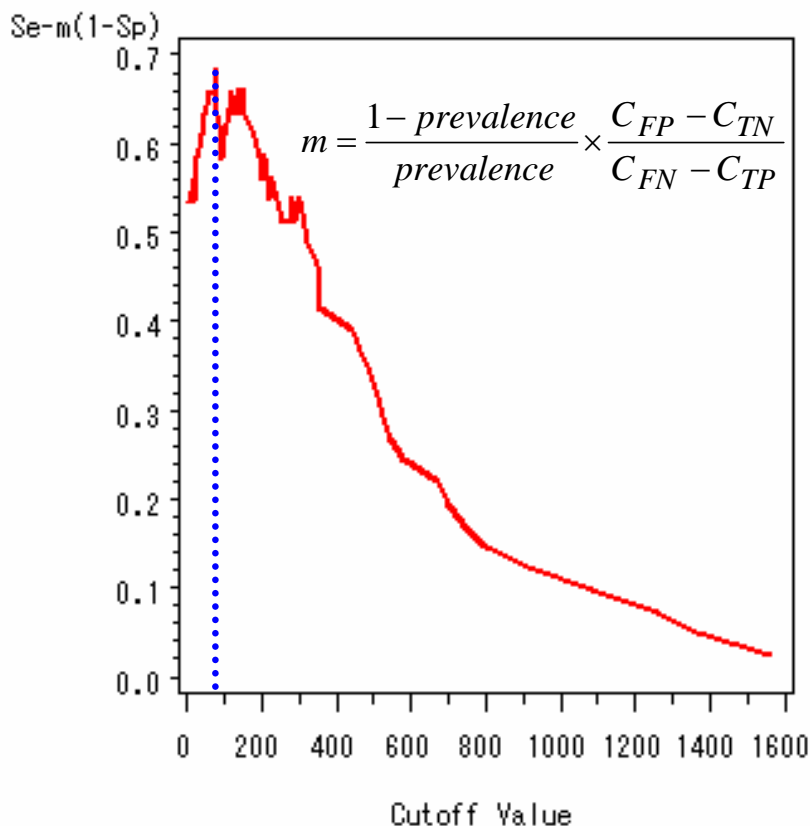
| | | AUC | | p-value |
|-----------|---------------|-------|---------|---------|
| | method | Plain | Digital | |
| Empirical | Delong et al. | 0.835 | 0.847 | P=0.783 |
| Binormal | ROCKIT(MLE) | 0.869 | 0.905 | P=0.351 |
| | direct | 0.843 | 0.875 | P=0.416 |

5段階評価の順序カテゴリ応答でROC曲線を描画してみたよ。

手法ごとに曲線もP値も結構違うんだね。



最適cutoff値の設定



Cutoff値って感度と
特異度の和が最大
になる値でいいの？

有病率と判別コスト
を設定して決めるべき
なんだけど、実際には
難しいよね～。



クラスターデータにおける診断検査間の比較

$$\hat{Se}_i = \frac{\sum_{j=1}^I N_{ij} \hat{Se}_{ij}}{\sum_{j=1}^I N_{ij}} \quad \hat{Var}(\hat{Se}_i) = \frac{1}{I_i(I_i - 1)} \sum_{j=1}^I \left[\frac{N_{ij}}{\bar{N}_i} (\hat{Se}_{ij} - \hat{Se}_i) \right]^2 \quad i = 1, 2$$

$$\hat{Cov}(\hat{Se}_1, \hat{Se}_2) = \frac{1}{I(I - 1)} \sum_{j=1}^I \left(\frac{N_j}{\bar{N}} \right)^2 (\hat{Se}_{1j} - \bar{Se})(\hat{Se}_{2j} - \bar{Se})$$

| Specificity PET | SD_PET | L_95%CI PET | U_95%CI PET | Specificity SPECT | SD_SPECT | L_95%CI SPECT | U_95%CI SPECT |
|--------------------|---------------|----------------|-----------------|----------------------|----------|------------------|------------------|
| 0.78431 | 0.069603 | 0.64789 | 0.92073 | 0.92157 | 0.035474 | 0.85204 | 0.99110 |
| Covariance | Difference SD | diff | L_95%CI Diff | U_95%CI Diff | z value | P value | |
| -.000086256 | -0.13725 | 0.079218 | -0.29252 | 0.018010 | -1.73262 | 0.083164 | |

1症例に複数個の検査対象がある場合は、個体内相関を考慮した分散が必要なんだよね

これで基本的な診断検査の解析は大丈夫だね
これからも楽しくお勉強しよー！！



研究の成果物：SAS sample code

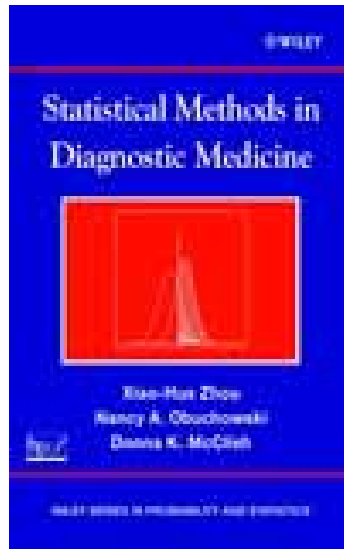
| No. | 内容 | Procedure |
|-----|---|------------------------------|
| 1 | 2 × 2 表における各種統計量とその信頼区間 | FREQ |
| 2 | Wilson's score 法による信頼区間 | Data step |
| 3 | Youden's index とその信頼区間 | Data step |
| 4 | Bayes の定理に基づく事後確率の推定と Bayes 曲線の描画 | LOGISTIC, GPLOT |
| 5 | Empirical ROC の推定 | GPLOT, LOGISTIC, NPAR1WAY |
| 6 | bootstrap 法による AUC の分散推定 | LOGISTIC, MULTTEST |
| 7 | ROCKIT 用データ変換 | Data step |
| 8 | Binormal ROC の推定（直接推定） | Data step |
| 9 | 平均損失を最小にする cutoff 値の選択 | LOGISTIC, GPLOT |
| 10 | Wald , Newcombe-Wilson 法による信頼区間 （対応のあるデータ） | FREQ , Data step |
| 11 | クラスターデータの解析 | MEANS |

Delong et al.(1988)の Empirical ROC 比較マクロの紹介

SAS社のHPに公開していただく予定です。
必要に応じ適宜加工してご活用いただければ幸いです！



使用したテキストとお役立ちWeb site



Statistical Methods in Diagnostic Medicine

Zhou, McClish, Obuchowski



The Statistical Evaluation of Medical Tests for Classification and Prediction

MS Pepe



ROC曲線とcutoffの関係をスクロールバーで体感

<http://www.anaesthetist.com/mnm/stats/roc/>

Faganのノモグラムで簡単に事後確率が求められます

<http://araw.mede.uic.edu/cgi-alansz/testcalc.pl>

FDAとEMAのガイダンスがダウンロードできます

<http://www.fda.gov/cder/guidance/index.htm>

<http://www.emea.eu.int/pdfs/human/ewp/111998en.pdf>



論文集の訂正

論文集に誤記がありましたので訂正させていただきます。

5.1 ノンパラメトリック法 (Empirical ROC approach)

誤 : Delong et al.(1988)の指数分布に基づく近似を利用した

正 : Delong et al.(1988)の分布に依らない方法を利用した

表5.1

誤 : タイトル 表2.2と表2.3のAUCと分散 $\text{Var}(\text{AUC})$

正 : タイトル 表2.2と表2.3のAUCと標準誤差 $\text{SE}(\text{AUC})$

参考文献

誤 : 5) Newcombe RG. 873-890 .

正 : 5) Newcombe RG. 857-872 .

お手数ですが、ご修正ください。

