

# 検査性能の評価

## ROC, カッパ統計量, Bland-Altmanプロット

- 新しい検査方法を開発する際に必要なこと:
  - 感度と特異度を計算
    - データ(元々カテゴリデータの場合): その検査で陽性か陰性かと, 信頼性が確立した標準的な方法(gold standard)で検査して真にその病気かどうか
      - 感度 = 真にその病気の人を測った時に検査陽性となる割合
      - 特異度 = その病気でない人を測った時に検査陰性となる割合
    - データ(元々連続量の場合): 検査値, gold standardで真にその病気かどうか
      - ROC (Receiver Operating Characteristic)分析: 検査で陽性/陰性を判別する閾値を段階的に変え, 感度が1に,  $(1 - \text{特異度})$ が0に最も近い結果を与える値を最適閾値として求める。通常,  $(1 - \text{特異度})$ を横軸, 感度を縦軸にとって閾値を変えて得られる値を曲線で結ぶ(ROC曲線と呼ばれる)
  - 複数の検査方法をROC分析で比較する
    - ROC曲線の下での面積(AUC)が最も大きい方法が, 最も性能が良い方法と判定される。
    - 実際の方法決定には, 性能以外にも, コストや実施のしやすさなども考慮される



# 例1. 原虫感染強度が弱いときのマラリア迅速診断キットの性能評価

- マラリアには何種類もRDT(迅速診断キット)がある。元々、マラリア患者の他の熱病患者と区別するために開発された
- 熱のある患者は血中の原虫感染強度が強い→特異度が高いことが重要。感度は中程度でいい
- 原虫感染強度が弱いときの積極的疫学調査(症状がない一般住民を対象とした検査)にも有用
- ソロモン諸島での三日熱マラリアについてPan-R malariaを使った検査結果は、以下のように得られる  
「統計解析」「検査の正確度の評価」「定性検査の診断への正確度の評価」

	疾病	健康
陽性	a	b
陰性	c	d

- 感度 (sensitivity) =  $a/(a+c)$  ※ positive in disease と覚えるといい。
- 偽陰性率 =  $c/(a+c) = 1 - \text{感度}$
- 特異度 (specificity) =  $d/(b+d)$  ※ negative in health と覚えるといい。
- 偽陽性率 =  $b/(b+d) = 1 - \text{特異度}$
- 陽性反応的中度 (positive predictive value) =  $a/(a+b)$
- 陰性反応的中度 (negative predictive value) =  $d/(c+d)$
- 陽性尤度比 =  $(a/(a+c))/(b/(b+d)) = \text{感度}/(1 - \text{特異度})$
- 陰性尤度比 =  $(d/(b+d))/(c/(a+c)) = \text{特異度}/(1 - \text{感度})$

## 出カウインドウ

```
> epi.tests(.Table, conf.level = 0.95)
              Disease positive Disease negative Total
Test positive          7             3         10
Test negative         16            156        172
Total                  23            159        182

Point estimates and 95 % CIs:
-----
Apparent prevalence          0.055 (0.027, 0.099)
True prevalence              0.126 (0.082, 0.184)
Sensitivity                   0.304 (0.132, 0.529)
Specificity                   0.981 (0.946, 0.996)
Positive predictive value     0.7 (0.348, 0.933)
Negative predictive value     0.907 (0.853, 0.946)
Diagnostic accuracy           0.896 (0.842, 0.936)
Likelihood ratio of a positive test 16.13 (4.485, 58.008)
Likelihood ratio of a negative test 0.709 (0.541, 0.93)
```



## 例2. 診断のために数値の基準値を決定

- 質問紙に基づいたうつ得点により, うつ病のスクリーニングを行う
- 必要なこと: 臨床診断でうつ病とわかっている患者と, うつ病でないとわかっている患者(または健康なボランティア)の両方を対象にしないではいけない。対象者全員について, 同じ質問紙によって得た得点を用いる(下表で2行目が得点, 3行目が臨床診断)

1	2	3	4	5	6	7	8	9	10
20	13	19	21	22	28	11	25	16	19
うつ	非うつ	非うつ	非うつ	うつ	うつ	非うつ	非うつ	非うつ	非うつ

\* この質問紙得点におけるうつの診断基準が, 「18点以上を陽性と判定する」とすると, 診断のためのクロス集計表は以下。感度は1 (3/3) で, 特異度は約0.43 (3/7) となる。

	うつ	非うつ
陽性	3	4
陰性	0	3

\* 基準値を変えることにより, 感度と特異度がどちらも高くなるような点を探索することができる。



# 例2. うつ得点の基準値設定のためのROC分析

- 「ファイル」「新しいデータセットを作成する(直接入力)」からデータを入力 →
- 「統計解析」[検査の正確度の評価]「定量検査の診断への正確度の評価(ROC曲線)」から変数を選んで[OK]をクリック

	Score	Diagno	var3
1	20	1. Dep	
2	13	0. Norm	
3	19	0. Norm	
4	21	0. Norm	
5	22	1. Dep	
6	28	1. Dep	
7	11	0. Norm	
8	25	0. Norm	
9	16	0. Norm	
10	19	0. Norm	
11			

定量検査の診断への正確度の評価(ROC曲線)

結果(値が0か1の項目を1つ選択) 予測に用いる値(1つ選択)

diagno score

グラフ中にベストの閾値のポイントを示す

陽性の判定基準

自動

閾値以上を陽性と判定する

閾値以下を陽性と判定する

ベストの閾値の判定基準

感度と特異度の和が最大になる閾値

左上隅に最も近づく閾値

偽陽性と偽陰性の重要性が等しくない場合は以下に重み付けを入力する(重要性が等しい場合は変更不要)。

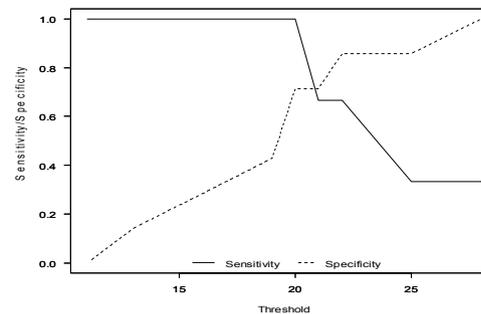
偽陽性に対する偽陰性の重み付け: 1

有病率(0と1の間の値): 0.5

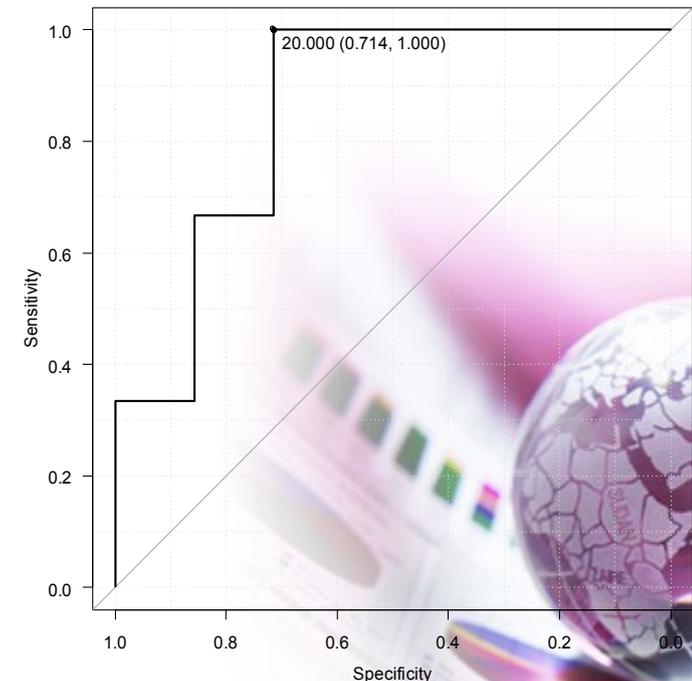
↓一部のサンプルだけを解析対象にする場合の条件式。例: age>50 & Sex==0 や age<50 | Sex==1

<全ての有効なケース>

OK キャンセル リセット ヘルプ



「20以上がうつ」という基準で最適な感度(1.0)と特異度(0.714)が得られた。曲線下面積(AUC)は0.8571, 95% CIが0.6044-1 (DeLong法)

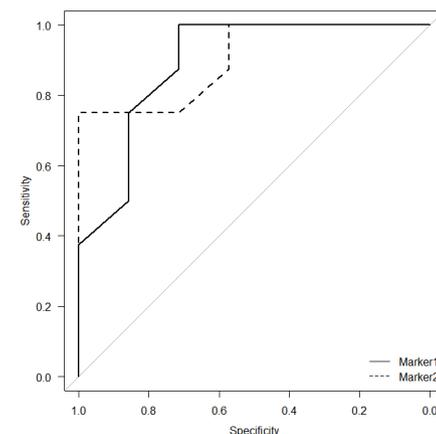


# 例3. 複数の方法をROC分析で比較

- 同じモノを評価するための2つの異なった検査の結果は異なりうる。ROC分析の結果として計算されるAUCにより、性能を比較可能。

PID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
病理診断	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
Marker1	2.2	1.8	2.4	2.2	1.7	2.5	2.9	2.3	1.8	1.1	1.3	1.4	2.3	1.0	0.8
Marker2	3.5	2.8	3.9	3.4	1.8	3.0	3.1	2.0	2.1	0.9	2.7	0.9	2.0	0.5	0.4

- 「ファイル」「データのインポート」「ファイルまたはクリップボード、URLからテキストデータを読み込む」で、データセット名をcomptwo、データファイルの場所をインターネットのURL、フィールド区切りをタブにしてOKし、URLとして<http://minato.sip21c.org/ROC1.txt>を指定する
- データセット名としてROC1とROC2は使ってはいけない。もし使うと計算途中でデータセットが上書きされてエラーになる。
- 「統計解析」「検査の正確度の評価」「2つのROC曲線のAUCの比較」を選び、Marker1とMarker2を選んでOK
- $Z = -0.0981$ ,  $p\text{-value} = 0.9218$  → **5%水準で有意差なし**  
 AUC of roc1    AUC of roc2  
 0.8928571    0.9017857



# 2つの診断(カテゴリデータ)の一致

- $\kappa$ (カッパ)統計量は、2つのカテゴリ変数の一致度を評価するのに使える  
2つの測定結果の他、2回の繰り返し評価とか、異なる評価者による評価などにも使える
- 前の例で、「1.7を超えたら(=1.8以上)陽性」「2.8を超えたら陽性」という基準値をマーカー1と2各々の値に対して適用して診断。陽性/陰性は以下の表になる(「アクティブデータセット」「変数の操作」「変数を指定した閾値で2群に分けた新しい変数を作成する」)。

PID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Pathology	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
Marker1	1	1	1	1	0	1	1	1	1	0	0	0	1	0	0
Marker2	1	0	1	1	0	1	1	0	0	0	0	0	0	0	0

- 「統計解析」「名義変数の解析」「対応のある比率の比較(二分割表の対称性の検定, McNemar検定)」で新しい変数2つを選ぶと、クロス集計表が得られる
- 「統計解析」「検査の正確度の評価」「2つの定性検査の一致度の評価(カッパ係数)」 得られたクロス集計表の数値, 5, 4, 0, 6を入力し「OK」

```
$kappa
  est      lower      upper
1 0.5 0.08039678 0.9196032

$mcnemar
 test.statistic df    p.value
1              4    1 0.04550026
```

Kappa estimate was 0.5, which means  
“Moderate agreement”



# 同じ量の2種類の測定結果の一致

- 新しく安価あるいは迅速な測定方法を開発したとき、その測定方法が信頼できるかどうかを検討するには、同じ対象者を、従来gold standardであると考えられてきた方法で測定し、一致しているかをみる

- 検討の方法

- 対応のあるt検定→絶対値の大きさとの交互作用を検出できない
- 相関係数を求め、散布図を描き、 $x=y$ の直線も描いて、直線から系統的に点がずれていないか
- BAプロット(Bland-Altman plot) : 有名。Bland JMとAltman DGの論文がLancetに発表されて以来必須

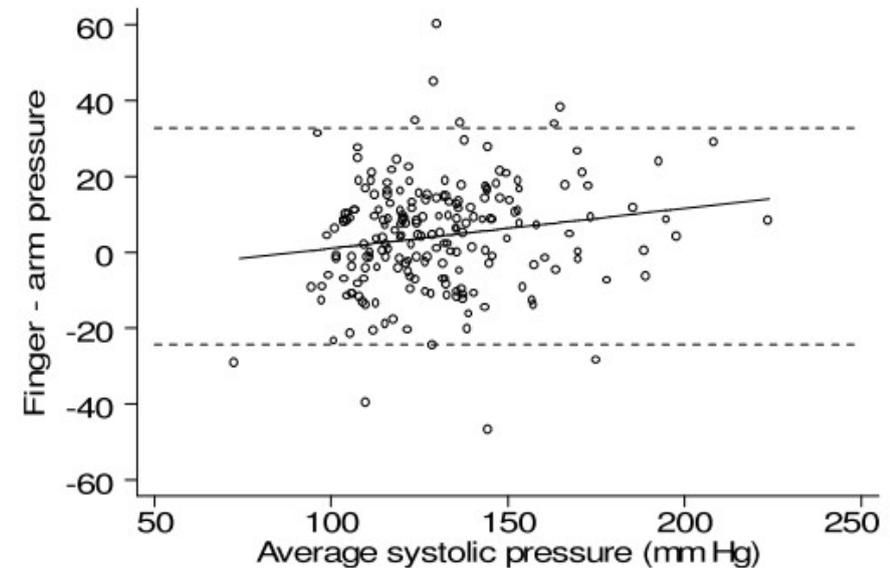


Figure 2. Difference against average of test and standard measurements, with 95% limits of agreement (broken lines) and regression line



# RでBAプロットを行う方法

- 変数の操作で、新しい変数として、検討したい2つの変数の差の変数Dと、2つの変数の平均値の変数Mを作成し、横軸にM、縦軸にDをとった散布図を描けば良い。
- MethCompパッケージのBA.plotが便利。
- ```
library(MethComp)
data(ox) # 子供61人の
# 血中酸素飽和度を血液
# ガス測定(CO)とパルス
# オキシメータ測定(pulse)
# で測定した結果
BA.plot(ox)
```

