

保健統計学

- Interaction (交互作用)
 - 新谷『今日から使える医療統計』Lesson 10
 - ロスマン『ロスマンの疫学』篠原出版 第 11 章(原著: Rothman KJ "Epidemiology: An Introduction 2nd ed." Oxford Univ. Press, Chapter 11
 - Rothman KJ, Greenland S, Lash TL "Modern Epidemiology 3rd ed." LWW, Chapter 5
- 感度・特異度: 新谷前掲書 Lesson 11, ロスマン前掲書第 13 章
 - http://minato.sip21c.org/publichealth/epidemiology.pdf の 10.4
 - Raffle A, Gray M "Screening: Evidence and Practice" Oxford Univ. Press, 2007.
- 同等性・非劣性の解析: 新谷前掲書 Lesson 12
 - 吉村功「検証的比較臨床試験の計画において考慮すべきこと」統計数理, 46(1): 81-95, 1998. <https://www.pmda.go.jp/files/000204955.pdf>
 - 佐藤俊哉「非劣性試験に潜む問題点」 <https://www.pmda.go.jp/files/000204955.pdf>
- 生存分析: 新谷前掲書 Lesson 13
 - エビデンスベーストヘルスケア特講資料第 16 章

Interaction (交互作用)

- 効果の修飾 (effect modification) と言われるが、より厳密には効果指標の修飾 (effect-measure modification)
 - 指標の選択に依存
 - 相加的効果 (additive effect) は効果を比で見るときは交互作用なしだが差で見るときは交互作用ありのこともある。例えば、新谷『今日から使える医療統計』の図 10-1A でも、効果指標が**リスク差**ならばワルファリンの有無によってアスピリンの効果は異なるので、統計的交互作用ありといえる
 - 相乗的効果 (multiplicative effect)
 - 生物学的交互作用(注目している要因が、1つの因果パイの構成要因となっている)とは区別して考えるべき
- 解析は層別解析により、層によって関連が異なることで示すか、交互作用項を説明変数に含めた多変量解析で示すのが普通
 - アウトカムが量的な変数であれば、交互作用項を含めた二元配置分散分析。交互作用があるときは、交互作用項の係数が統計学的に有意に 0 と異なる
 - アウトカムがカテゴリ変数ならロジスティック回帰分析で交互作用項を含めれば OK

二元配置分散分析で交互作用を見る例

- 以下の URL の解説は、分散分析の仕組みから丁寧に図解しておりお薦め http://www.cwk.zaq.ne.jp/fkhud708/files/R-intro/R-stat-intro_06.pdf
- 実行するコードは <http://minato.sip21c.org/QOL-ANOVA.R> に載せた
- 大雑把なプロセスとしては以下
 - データを読む(データは縦長形式で。アウトカムとなる量的変数は1つの変数として縦に積み、2つの要因は別々のカテゴリ変数として与える)
 - 2つの要因の組合せごとの平均値、標準偏差、サンプルサイズを計算する
 - 標準偏差をサンプルサイズの平方根で割って標準誤差を出す
 - 2つの要因の組合せごとの平均 ± 標準誤差(または標準偏差。明記すればどちらの場合もありうる)の折れ線グラフを描く
 - アウトカムとなる量的変数を目的変数、2つの要因とその交互作用の線形結合を説明変数とした線型モデルの当てはめを行う
 - サンプルサイズが組合せごとに異なっているときはタイプ III の平方和を使った分散分析が必要になるので、R では car パッケージの Anova() 関数で type=3 オプションを使うが、サンプルサイズがすべて同じなら通常の anova() で良い。結果を見て、交互作用項の p 値が予め決めた有意水準(通常は 0.05)より小さければ、アウトカム変数に対する2つの要因の効果には統計学的に有意な交互作用があると判定する

感度・特異度

- スクリーニングの性能を示す指標値
- 既に疾病ありとわかっている人 (a+c) 人と疾病なしとわかっている人 (b+d) 人を、何らかの検査方法で検査して、陽性という結果になったのがそれぞれ a 人、b 人であったとき、下記のクロス集計表ができる

	疾病	
	あり	なし
検査陽性	a	b
陰性	c	d
- このとき、 $a/(a+c)$ が感度 (sensitivity = 病気の人を正しく病気と判定する確率) であり、 $d/(b+d)$ が特異度 (specificity = 病気なしの人を正しく病気なしと判定する確率) である。
- 通常、感度や特異度が 0.9 もあれば性能の高い検査と言えるが、スクリーニング対象集団における有病割合が低いときは**陽性的中率**(陽性と判定されたうち本当に疾病ありである確率)が低くなることもある
 - 有病割合 0.1% の病気について、感度も特異度も 0.9 の検査で陽性と出た場合、本当にその病気である確率(陽性的中率)は? (ヒント) 有病割合 0.1% なら、 $(a+c):(b+d)=1:999$ になるはず
 - 全体で 10000 人とすると、 $a+c=10$, $b+d=9990$ 。感度が 0.9 なので $a=9$, $c=1$ 。特異度が 0.9 なので、 $b=999$, $d=8991$ 。陽性的中率は $a/(a+b)$ なので $9/1008=0.009$

ROC 曲線

- 既に病気か病気でないかの確定診断がついている人について量的な測定値がある状況
- 検査データの最大値から最小値までカットオフ値を順に変えて陽性・陰性を判定し、横軸に 1 - 特異度、縦軸に感度をとってプロットして結んだ曲線
- 2つの目的
 - 最適カットオフ値の決定(測定値のいくつ以上、あるいはいくつ以下を陽性と判定すべきか?)
 - 感度も特異度もなるべく大きい方がいい = ROC 曲線のグラフが最も左上隅に近づく点を与えるカットオフが最適(右下隅から最も遠い点という計算方法もあるが)
 - 複数の検査方法の性能比較
 - 曲線下面積 (AUC) が大きいほどいい検査
 - 左上隅に近いところを通る曲線ほど高性能
- (裏面参照)

同等性・非劣性の解析 (佐藤俊哉先生の資料参照)

- 通常の臨床試験は優性試験
 - 新薬は従来薬(あるいはプラセボ)より有意に優れた効果を示さねばならない
 - 「差がない」が帰無仮説
- しかしジェネリック医薬品の試験は、先発医薬品と同等でも良い。劣っていなければ良い。先発医薬品よりも安く同等な性能を提供できれば十分。このような状況でやるのが同等性試験や非劣性試験
 - 検定ではなく、予め同等マージンや非劣性マージンを決めておいて、信頼区間を推定する
 - 同等性試験では信頼区間が同等マージン内に収まれば同等とみなす
 - 非劣性試験では信頼区間の下限が非劣性マージンを下回らなければ劣っていないとみなす
 - 通常、非劣性試験の方がサンプルサイズは小さくて済む

生存分析(生存時間解析)

- 臨床試験などで、死亡や副作用発生の有無だけを情報にすると 1 個体について生存/死亡とか、副作用有り/無しのような 2 値情報 1 つしか得られない。
- 死亡や副作用という事象が発生するまでの時間を使えば、作用が強ければ短い待ち時間で発生するし、作用が弱ければ長い時間がかかることが示唆され、2 値情報よりも豊富な情報が得られる → 生存分析 (survival analysis = 生存時間解析)
- すべての個体に事象が発生するまで待てることは稀で、通常は事象が発生していない個体が残っているうちに観察期間が終了してしまう = 右側打ち切り (right censoring)
 - 右側打ち切り個体には、おそらく観察を打ち切った時点よりも後で事象が発生するはずなので、欠損にしてしまうと、事象発生までの待ち時間を過小評価することになる
 - 生存時間解析では、こうした打ち切りデータを正しく扱うための工夫がなされている

生存時間解析の典型的な手法

- (ノンパラメトリックな方法)
 - 打ち切りを考慮した生存時間(イベント発生までの時間)の中央値と 95% 信頼区間の推定 → カプラン・マイヤ法
 - 2 群の生存時間の差の検定 → ログランク検定
- (セミパラメトリックな方法)
 - すべての要因が「なし」である基準人の死亡リスク(イベントが起こるリスク。通常、瞬間のリスクと考えればよい「ハザード」で示される)に対して、要因となる変数が「あり」となることでリスクが何倍になるかの推定(他のリスクの影響も考慮し、そのリスクの有無によって対数ハザード曲線が平行移動すると考える、「比例ハザードモデル」を当てはめる) → コックス回帰
- (パラメトリックな方法)
 - イベントが起こるまでの時間が何らかの既知の分布(ワイブル分布など)に当てはまっていると考え、分布を当てはめてパラメータ推定 → 加速モデル

検査性能の評価としての感度, 特異度, ROC

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): その検査で陽性が陰性かと、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-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-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

	疾病	健康
陽性	a	b
陰性	c	d

- 熱のある患者は血中の原虫感染強度が高い→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」

- 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 = $b/(b+d) = 1 - \text{特異度}$
- 陽性反応の中心 (positive predictive value) = $a/(a+b)$
- 陰性反応の中心 (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 = $(d/(b+d))/(c/(a+c)) = \text{特異度}/(1 - \text{感度})$



```

> epi.testsi(.Table, conf.level = 0.95)
      Disease positive Disease negative Total
Test positive      16          156      172
Test negative      23          159      182
Total              39          315      354

Point estimates and 95 % CIs:
-----
Apparent prevalence      0.055 (0.027, 0.099)
True prevalence         0.126 (0.082, 0.184)
Sensitivity              0.204 (0.122, 0.328)
Specificity             0.981 (0.946, 0.996)
Positive predictive value 0.7 (0.348, 0.933)
Negative predictive value 0.807 (0.853, 0.946)
Diagnostic accuracy     0.896 (0.842, 0.936)
Likelihood ratio of a positive test 16.13 (4.685, 58.008)
Likelihood ratio of a negative test 0.709 (0.545, 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、うつ病でない患者(または健康なボランティア)の両方を対象にしないといけない。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

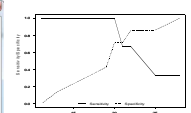
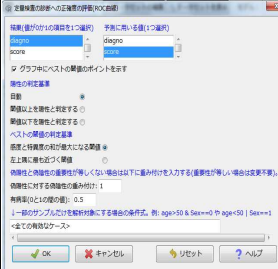
*この質問紙得点におけるうつの診断基準が、「18点以上を陽性と判定する」とすると、診断のためのクロス集計表は以下。感度は1(3/3)で、特異度は約0.43(3/7)となる。

	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。

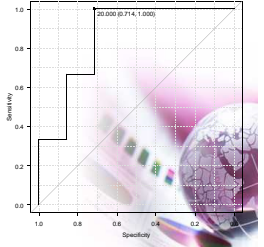
例2. うつ得点の基準値設定のためのROC分析

- 「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



「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0.6044-1(DeLong法)

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		



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comp two、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-value = 0.9218 → 5%水準で有意差なし
 AUC of roc1 AUC of roc2
 0.8928571 0.9017857

