

2020/5/11 北大SPH・統計解析の基礎⑦

ロジスティック回帰



北海道大学 医学統計学
横田 勲

今回の内容

- ▶ 3水準以上をもつ分割表の解析
 - ▶ 傾向検定
 - ▶ 比例オッズモデル、一般化ロジットモデル
- ▶ 統計モデルの作り方
 - ▶ ROC解析

3×3の分割表

- ▶ 人種と乳がんステージの関係

	In situ	Local	Distant	Total
Caucasian	124	761	669	1554
African American	36	224	301	561
Asian	221	104	109	434

Chernick MR. The Essentials of Biostatistics for Physicians, Nurses, and Clinicians. 2011.

独立性のカイ二乗検定①

- ▶ p_{11} : CaucasianのIn situである確率
- ▶ p_{12} : CaucasianのLocalである確率
- ▶ p_{13} : CaucasianのDistantである確率
 - ▶ $p_{13} = 1 - p_{11} - p_{12}$
- ▶ p_{21} : African AmericanのIn situである確率
- ▶ p_{31} : AsianのIn situである確率
- ▶ $\bar{p}_{.1}$: 集団全体のIn situである確率

独立性のカイ二乗検定②

- ▶ 3群ともアウトカムの確率が同じ、という帰無仮説

$$\text{▶ } p_{11} = p_{21} = p_{31} (= \bar{p}_{.1})$$

$$\text{▶ } p_{21} - p_{11} = 0, p_{31} - p_{11} = 0$$

$$\text{▶ } p_{12} = p_{22} = p_{32} (= \bar{p}_{.2})$$

$$\text{▶ } p_{22} - p_{12} = 0, p_{32} - p_{12} = 0$$

$$\text{▶ } p_{13} = p_{23} = p_{33} (= \bar{p}_{.3})$$

- ▶ 自由度4のカイ二乗検定

数値例

- ▶ $\chi^2 = 552.10 > 9.49$
 - ▶ 自由度4の(両側)5%点 : 9.49
- ▶ アウトカムを構成する確率のいずれかがどこかの群間にて異なる

順序性のあるデータ①

7

▶ 濃度とアウトカム

- ▶ アウトカムは二値、曝露に順序性

便中Hb濃度[$\mu\text{g/mL}$]	AVN*	Non-AVN
31-100	15 (5%)	268 (95%)
100-300	2 (2%)	126 (98%)
301-500	1 (3%)	28 (97%)
>501	4 (7%)	56 (93%)

*AVN: advanced neoplasia
Kawamura T, et al. *JGH Open* 2020.

順序性のあるデータ②

8

▶ 反射のスコア

- ▶ 曝露は二値、アウトカムに順序性

反射度合い	1	2	3	4
試験薬 (n=74)	59 (79.7%)	1 (1.4%)	11 (14.9%)	3 (4.1%)
対照薬 (n=71)	31 (43.7%)	6 (8.5%)	25 (35.2%)	9 (12.7%)

Shimizu M, et al. *PLoS One* 2019: e0223947.

傾向検定

9

- ▶ 曝露カテゴリが上位であるほど、イベント発生確率が高い (低い)
 - ▶ Cochran-Armitage test for trend
- ▶ 試験群は対照群に比べ、より上位 (下位) のアウトカム水準だ
 - ▶ Mantel test for trend
 - ▶ Wilcoxonの順位和検定
- ▶ 3水準以上の群とアウトカム
 - ▶ 一般化Mantel検定

各水準を代表する点数をおく

10

- ▶ 「傾向」を考えるため

	等間隔(table)			中央順位(midrank)			
	1	2	3		3.5	12.5	30.5
1	1	2	4	4	1	2	4
2	2	4	8	14.5	2	4	8
3	3	6	12	32	3	6	12

Cochran-Armitage傾向検定①

11

▶ 濃度とアウトカム

便中Hb濃度[$\mu\text{g/mL}$]	AVN	Non-AVN
31-100	15 (5%)	268 (95%)
100-300	2 (2%)	126 (98%)
301-500	1 (3%)	28 (97%)
>501	4 (7%)	56 (93%)

- ▶ 独立性のカイ二乗検定: $\chi^2 = 3.7907$
- ▶ 傾向検定: $Z = -0.02, \chi^2 = 0.0005$

Cochran-Armitage傾向検定②

12

- ▶ H_0 : 曝露カテゴリ間で、イベント確率に直線的な関係がない
- ▶ $\pi_i = \alpha + \beta x_i$ として $\beta = 0$ の検定
 - ▶ π_i : 曝露水準 i でのイベント確率
 - ▶ x_i : 曝露水準を代表する点数
 - ▶ SASやJMPでのデフォルトは等間隔(table)点数
- ▶ 数値例 $3.7907 = 0.0005 + 3.7902$

\swarrow 独立性のカイ二乗検定 (自由度3)
 \swarrow CA傾向検定 (自由度1)
 \swarrow 直線的ではない分 (自由度2)

Mantel傾向検定①

13

▶ 反射のスコア

反射度合い	1	2	3	4
試験薬 (n=74)	59 (79.7%)	1 (1.4%)	11 (14.9%)	3 (4.1%)
対照薬 (n=71)	31 (43.7%)	6 (8.5%)	25 (35.2%)	9 (12.7%)

Shimizu M, et al. *PLoS One* 2019: e0223947.▶ $\chi^2 = 16.87, Z = 4.11 > 1.96$

▶ 片側2.5%水準で有意差あり

Mantel傾向検定②

14

▶ 曝露とアウトカムとの関連が直線的であることを言いたい

▶ Mantel-Haenszelのカイ二乗検定

▶ $\chi^2 = (n-1)r^2$ (自由度は1)▶ r : 例えば等間隔(table)点数を与えて求めたPearsonの相関係数▶ n : 全体の人数

Wilcoxonの順位和検定を利用

15

▶ 分割表の各水準に与える点数を中央順位(midrank)とした場合の傾向検定

Non-model-basedな検定

16

▶ ここまで登場した傾向検定は水準を代表する点数だけ指定して直線的に増加するかを検討した

▶ 効果の指標で評価できない

▶ ロジスティック回帰を応用してオッズ比で表現してみよう

アウトカムが3水準以上

17

▶ アウトカムに順序性なし

▶ 一般化ロジットモデル
(ベースラインカテゴリロジットモデル)

▶ アウトカムに順序性あり

▶ 比例オッズモデル (累積ロジットモデル)

▶ 隣接カテゴリロジットモデル

▶ 逐次ロジットモデル

例: 4項アウトカム

18

▶ 水準 j ($j = 1, 2, 3, 4$)となる確率を p_j ▶ $p_1 + p_2 + p_3 + p_4 = 1$

一般化ロジットモデル①

19

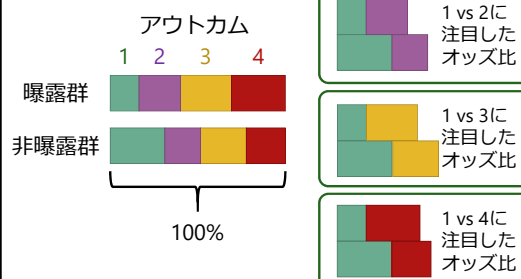
▶ ベースライン水準に対するオッズ

▶ 例えば、第1水準

$$\begin{cases} \log \frac{p_2}{p_1} = \alpha_2 + \beta_2 x \\ \log \frac{p_3}{p_1} = \alpha_3 + \beta_3 x \\ \log \frac{p_4}{p_1} = \alpha_4 + \beta_4 x \end{cases}$$

一般化ロジットモデル②

20



比例オッズモデル①

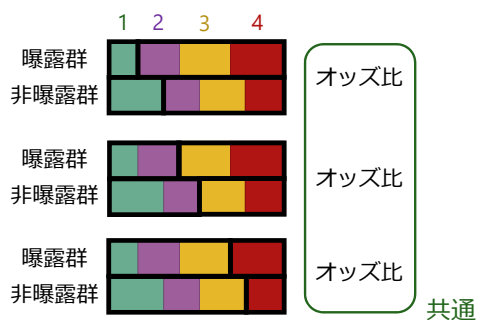
21

▶ ある水準以上/未満のオッズをモデル化

$$\begin{cases} \log \frac{p_2 + p_3 + p_4}{p_1} = \alpha_2 + \beta x \\ \log \frac{p_3 + p_4}{p_1 + p_2} = \alpha_3 + \beta x \\ \log \frac{p_4}{p_1 + p_2 + p_3} = \alpha_4 + \beta x \quad \text{共通} \end{cases}$$

比例オッズモデル②

22



隣接カテゴリオッズモデル

23

▶ ある水準と隣の水準のオッズ

$$\begin{cases} \log \frac{p_2}{p_1} = \alpha_2 + \beta x \\ \log \frac{p_3}{p_2} = \alpha_3 + \beta x \\ \log \frac{p_4}{p_3} = \alpha_4 + \beta x \quad \text{共通} \end{cases}$$

逐次オッズモデル

24

▶ ある水準とそれ以上/未満のオッズ

$$\begin{cases} \log \frac{p_2 + p_3 + p_4}{p_1} = \alpha_2 + \beta x \\ \log \frac{p_3 + p_4}{p_2} = \alpha_3 + \beta x \\ \log \frac{p_4}{p_3} = \alpha_4 + \beta x \quad \text{共通} \end{cases}$$

例：一般化ロジットモデル

25

反射度合い	1	2	3	4
試験薬 (n=74)	59 (80%)	1 (1%)	11 (15%)	3 (4%)
対照薬 (n=71)	31 (44%)	6 (9%)	25 (35%)	9 (13%)

▶ 基準カテゴリは「反射度合い=1」

	Est.	S.E.	アウトカム	オッズ比	(95%CI)
α_2	-1.64	0.45	反射度合い2	0.09	(0.01-0.76)
α_3	-0.22	0.27	反射度合い3	0.23	(0.10-0.53)
α_4	-1.24	0.38	反射度合い4	0.18	(0.04-0.69)
β_2	-2.44	1.10			
β_3	-1.46	0.42			
β_4	-1.74	0.70			

例：比例オッズモデル

26

反射度合い	1	2	3	4
試験薬 (n=74)	59 (80%)	1 (1%)	11 (15%)	3 (4%)
対照薬 (n=71)	31 (44%)	6 (9%)	25 (35%)	9 (13%)

▶ 高い度合いに関するオッズ

	Est.	S.E.	オッズ比	(95%CI)
α_2	0.19	0.23	0.21	(0.11-0.44)
α_3	-0.05	0.23		
α_4	-1.86	0.32		
β	-1.53	0.37		

医学研究での目的

27

- ▶ “X” は “疾病Y” と 関連 がある
 - ▶ X：健康状態マーカーや
疾病Yを引き起こす疾患など



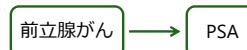
- ▶ “X” は “疾病Y” の 原因 となる
- ▶ “X” は “疾病Y” を 予測 する

より目的を明確に

前立腺がんとPSA

29

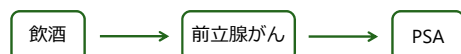
- ▶ 前立腺がんの発見・病勢と強い関連
 - ▶ スクリーニングにも用いられる
- ▶ がんの細胞壁が壊れやすいため、
がんのvolumeに応じてPSAが血液中に漏出



前立腺がんのリスク因子を検討

30

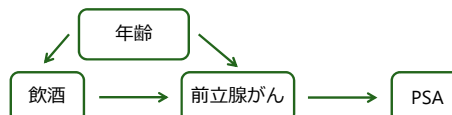
- ▶ 明らかなリスク因子は、年齢、家族歴
- ▶ 他にもリスク因子はあるに違いない！
 - ▶ 例えば、飲酒の影響を調べてみる
 - ▶ 因果関係を知りたい



交絡の影響を解析で除去

31

- ▶ 以下の条件を満たすことで、
飲酒と前立腺がんの関係を歪めてしまう
 - ▶ 年齢が高いほど前立腺がんは増える
 - ▶ 年齢と飲酒には関係がある
 - ▶ 飲酒をすれば年齢が増えるわけではない

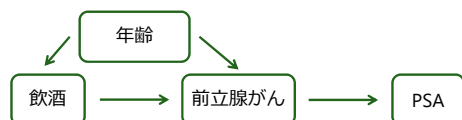


前立腺がんの予測をしたい

32

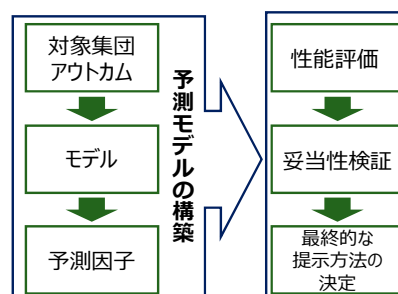
▶ 前立腺がん発生を精度良く当てたい

- ▶ どのような因子を用いてもよい
 - ▶ 年齢のようなリスク因子
 - ▶ 前立腺がんの“結果”であるPSA



予測モデル構築の流れ

34



DLBCLの新規予後予測モデル

36

- ▶ びまん性大細胞型B細胞リンパ腫
- ▶ 全生存予後を予測したい
- ▶ 臨床で簡単に利用できるスコアを作りたい
 - ▶ 年齢、血清LDH、Ann Arborステージ、ECOG-Performance Status、血清CRP、低アルブミン血症、節外（骨髄、骨、皮膚、肺／胸膜）病変
 - ▶ 変数選択により、予測に用いる因子を決定

予測性能指標による評価

37

- ▶ 予後の悪い対象者を特定するための予測モデルがどれだけ有用かを知りたい
- ▶ 他の予測モデルと比較したい
- ▶ 予測モデルを構築する上で、overfittingを避けたい
 - ▶ ノイズまでモデルをあてはめてしまい、将来の対象者への予測性能が悪くなること

ところで「予測性能がよい」とはどういうこと？

モデルのよさ、精度の測り方

38

- ▶ モデルのあてはまり
 - ▶ 決定係数 R^2
 - ▶ 尤度とAIC
- ▶ 予測精度、予測結果のよさ
 - ▶ 平均二乗誤差、Brierスコア
 - ▶ ROC曲線、c 統計量

決定係数 R^2

39

$$R^2 = 1 - \frac{\text{残差平方和}}{\text{全体の平方和}} = \frac{\text{モデルで説明した平方和}}{\text{全体の平方和}}$$

- ▶ データの持つ全ばらつきのうち、モデルで説明した割合
 - ▶ 単回帰の場合、相関係数の2乗に一致

Akaike's Information Criterion; AIC

40

- ▶ $-2 \log L(\boldsymbol{\beta}; x) + 2K$
 - ▶ $\log L(\boldsymbol{\beta}; x)$: 対数尤度関数
 - ▶ パラメータ $\boldsymbol{\beta}$ の確からしさ、あてはまりのよさ
 - ▶ K : $\boldsymbol{\beta}$ のパラメータ数
- ▶ AICが最小となるモデルがよいモデル
 - ▶ パラメータを増やすことへのペナルティを与えた指標
 - ▶ 自由度調整済み決定係数も同様

古典的な変数選択法

41

- ▶ 基準に至るまで以下の操作を繰り返す
- ▶ 変数増加法
 - ▶ 変数候補から最もp値の小さなものを加える
- ▶ 変数減少法
 - ▶ 変数候補をすべて含めたモデルから最もp値の大きな変数を除く
- ▶ ステップワイズ法
 - ▶ 変数候補から最もp値の小さなものを加え、モデルから最もp値の大きな変数を除く

他の変数選択法

42

- ▶ p値の代わりに用いる基準
 - ▶ AIC
 - ▶ 平均二乗誤差、Brierスコア
 - ▶ c-index
 - ▶
- ▶ 総当たり法
 - ▶ 変数の組合せ全パターン調べる

平均二乗誤差 Mean Squared Error

43

- ▶ $\frac{1}{n} (\hat{y}_i - y_i)^2$
 - ▶ 予測値と実測値の差を評価
 - ▶ 平方根をとって、Root MSE; RMSE

Brier スコア

44

- ▶ イベント有無と生存確率のズレ
 - ▶ 生存時間アウトカムの場合、ある時点 t でのイベント有無と確率のズレ
- ▶ Brierスコア
 - ▶ $\{I(y = 1) - \hat{y}\}^2$
 - ▶ $\{I(T > t) - \hat{S}(t|\mathbf{X})\}^2$
 - ▶ $I(\cdot)$: かつこ内が真のときに1、それ以外は0
 - ▶ $\hat{S}(\cdot)$: 生存関数の予測値

平均Brierスコアの数値例

45

- ▶ 2人死亡、2人生存という仮想例

無情報モデル

ID	生存/死亡	予測確率	Brierスコア
1	死亡	0.5	$(1 - 0.5)^2 = 0.25$
2	死亡	0.5	$(1 - 0.5)^2 = 0.25$
3	生存	0.5	$(0 - 0.5)^2 = 0.25$
4	生存	0.5	$(0 - 0.5)^2 = 0.25$

平均Brierスコア
0.25

予測モデル

ID	生存/死亡	予測確率	Brierスコア
1	死亡	0.9	$(1 - 0.9)^2 = 0.01$
2	死亡	0.6	$(1 - 0.6)^2 = 0.16$
3	生存	0.3	$(0 - 0.3)^2 = 0.09$
4	生存	0.2	$(0 - 0.2)^2 = 0.04$

平均Brierスコア
0.075

相対Brierスコア減少 46

- ▶ 期待Brierスコアのとりうる範囲は0から0.25
 - ▶ しかも0に近いほど「予測性能がよい」
 - ▶ 集団全体の生存確率によって、上限が変化
- ▶ 無情報モデルに対する、予測モデルでの期待Brierスコアを小さくした割合
 - ▶ 0から1をとり、1に近いほど「予測性能がよい」

$$\frac{\text{Brier}_{\text{無情報モデル}} - \text{Brier}_{\text{予測モデル}}}{\text{Brier}_{\text{無情報モデル}}}$$

感度と特異度 47

至適基準	評価値	
	陽性	陰性
陽性	a	b
陰性	c	d

- ▶ 感度： $\frac{a}{a+b}$
 - ▶ 本当に陽性であるものを陽性といえたか
- ▶ 特異度： $\frac{d}{c+d}$
 - ▶ 本当に陰性であるものを陰性といえたか

陽性的中度、陰性的中度との違い 48

至適基準	評価値	
	陽性	陰性
陽性	a	b
陰性	c	d

- ▶ 陽性的中度： $\frac{a}{a+c}$ 、陰性的中度： $\frac{d}{b+d}$
 - ▶ 評価した結果が本当はどうであったか？
 - ▶ 真の陽性、陰性者の分布によって変わってしまう指標

カットオフ値をもって評価 49

至適基準で陰性 (本当に陰性) 至適基準で陽性 (本当に陽性)

カットオフ値

陰性と判断 陽性と判断

特異度 感度

感度と特異度はトレードオフ 50

至適基準で陰性 (本当に陰性) 至適基準で陽性 (本当に陽性)

カットオフ値

- ▶ 感度を上げれば、特異度は下がる
- ▶ 特異度を上げれば、感度は下がる

ROC曲線 51

- ▶ 縦軸に感度、横軸に偽陽性率 (1-特異度)
- ▶ カットオフ値を全範囲で動かした場合の感度と偽陽性率をプロット

感度

偽陽性率

AUC (Area Under the Curve)

ロジスティック回帰とROC曲線 52

- ▶ 至適基準の判定をアウトカムとした回帰
- ▶ カットオフ値は予測イベント確率

$$\hat{y}_i = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots)}$$

- ▶ カットオフ値は $\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots$ でも全く同じROC曲線

(ROC-)AUC 53

- ▶ ROC曲線の要約指標
 - ▶ 判別能力を表す指標として解釈
- ▶ AUC自体はモデルに依らずに計算される
 - ▶ AUC=0.5であれば、no discriminative ability
 - ▶ AUCが1に近づくほど、判別能力がよい
 - ▶ 絶対的な解釈は困難
- ▶ c (concordance) indexとも呼ばれる

c-index 54

$$\frac{\sum_{i=1}^n \sum_{j=1}^m \{I(\hat{p}_i > \hat{p}_j) + 0.5 \times I(\hat{p}_i = \hat{p}_j)\}}{nm}$$

- ▶ \hat{p}_i : 実際にイベント発生した*i*のイベント確率の推定値
- ▶ \hat{p}_j : 実際にイベント発生しなかった*j*のイベント確率の推定値
- ▶ イベント有無と予測変数の
 - ▶ 大小関係が一致していれば1点
 - ▶ 大小関係が一致していなければ0点
 - ▶ 値が等しければ0.5点 (引き分け)

c-indexの数値例① 55

- ▶ 死亡例の予測確率 : A:0.9, B:0.7, C:0.4
- ▶ 生存例の予測確率 : D:0.6, E:0.3, F:0.1

▶ 総当たり表

	生存例			
	0.6	0.3	0.1	
死亡例	0.4	×	○	c-indexは 8/9=0.89
0.7	○	○	○	
0.9	○	○	○	

○ : 一致 × : 不一致

c-indexの数値例② 56

- ▶ 死亡例の予測確率 : A:0.9, B:0.7, C:0.4
- ▶ 生存例の予測確率 : D:0.6, E:0.3, F:0.1

c-indexの数値例③ 57

- ▶ 死亡例の予測確率 : A:0.9, B:0.7, C:0.4
- ▶ 生存例の予測確率 : D:0.6, E:0.3, F:0.1

DLBCL予測モデル研究

58

	PFS		OS	
	c-index	RBSR	c-index	RBSR
R-IPI	0.668	0.122	0.642	0.135
NCCN-IPI	0.749	0.172	0.736	0.251
提案スコア(4段階)	0.703	0.183	0.740	0.305
元の0-5点スコア	0.711	0.215	0.754	0.356

RBSR : 相対Brierスコア減少

- ▶ 提案スコアが従来スコアより概ね性能がよいことを示した

まとめ

59

- ▶ 3水準以上の分割表
 - ▶ 順序性の検討に傾向検定、比例オッズモデル
- ▶ 関連から因果、予測へ
 - ▶ 予測モデルの構築のあと、妥当性の検証
 - ▶ 決定係数
 - ▶ 平均二乗誤差、Brierスコア
 - ▶ 感度、特異度、ROC曲線、c-index