

風邪薬のエンドポイントは？

■ あなたなら何をエンドポイントにしますか？

- 風邪が治るとは？
 - 鼻水が出なくなる？
 - のどが痛くなる？
 - 咳が出なくなる？
 - 熱が出なくなる？



風邪をひくと、何が困りますか？

2006/11/12

31

気管支炎に対するアジスロマイシン 二重盲検比較試験 (Evans et al., 2002)

■ 主要評価項目：職場復帰までの日数

ARTICLES

Azithromycin for acute bronchitis: a randomised, double-blind, controlled trial

Author: Evans, Stuart; Hooper, Carolyn; Daniels, Laura S; Satchwell, Harlene; Charles-Davies, Yee Wang

Summary

Background The value of azithromycin for treatment of acute bronchitis is uncertain, even though the drug is commonly prescribed. We have investigated this question in a randomised, double-blind, controlled trial.

Methods Adults diagnosed with acute bronchitis, without evidence of underlying lung disease, were randomly assigned azithromycin (n=112) or vitamin C (n=108) for 5 days (total dose for each 1.5 g). All individuals were also given liquid paracetamol as and when needed with a scone. The primary outcome was improvement in health-related quality of life at 7 days; an important difference was defined as 0.5 or greater. Analysis was by intention to treat.

Findings The study was stopped by the data-monitoring and safety committee when 220 patients had been recruited. On

Introduction

Every year, more than six million US adult visit physicians for acute bronchitis, and most of them receive antibiotics.^{1,2} Many experts consider such treatment, using these main reasons: weak or conflicting experimental evidence of clinical benefit, lack of a strong biological rationale (the causative pathogens are viruses in most cases), and increasing societal concern about "antibiotic resistance."³

New randomised controlled trials of three different antibiotics for acute bronchitis have been published.⁴⁻⁶ Our demonstrated clinical benefit, "Use-ability of these studies" have been based on fundamental problems:^{7,8} different studies measured different outcomes; the reliability and validity of outcome measurements was uncertain; and some outcomes had questionable clinical significance (for example, duration of purulent sputum). Equally important, no published

2006/11/12

32

真の評価項目・代替評価項目

- 真の評価項目 (true endpoint)
 - 治療効果の最も重要かつ直接的な評価基準
 - 評価に多数の被験者数、測定が困難、長期の試験
- 代替評価項目 (surrogate endpoint)
 - 治療効果の間接的な評価基準
 - 評価に少数の被験者数、測定が簡単、短期の試験

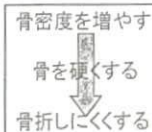
代替エンドポイントが妥当であるならば、社会的・医学的に利益が大きい

2006/11/12

33

真の評価項目・代替評価項目の事例

- 骨粗鬆症
 - 加齢とともに骨密度減少で骨の強度が低下し、骨折が起こりやすくなる疾患
 - 日本における骨粗鬆症患者数は1,000万人
 - 症状：背中や腰の痛み、身長低下、骨折など
 - 真の評価項目：骨折発生抑制効果
 - 代替評価項目：骨密度の増加



◆ 骨密度試験 (骨密度変化率 2.5%) ◆ 骨折試験 (骨折率 1.5%)

被験者数(人)	検出力(%)
75	75.9
85	82.0
95	86.6

被験者数(人)	検出力(%)
250	75.2
270	78.2
290	82.2

2006/11/12

34

良く用いられる代替評価項目

疾患	真の評価項目	代替評価項目
がん	生存期間, 無再発期間	腫瘍縮小効果
心不全	生存期間	左室駆出率
急性心筋梗塞	30日以内死亡率	再狭窄率
高脂血症	血管障害発生率	LDL値などの脂肪量
高血圧	脳卒中発生率, 死亡	血圧の減少
糖尿病	合併症発生, 死亡	血糖値 HbA1c

2006/11/12

35

症例数設計の必要性

- 臨床研究における倫理的正当性
 - 危険や不利益に晒される被験者数を最小限にする
 - 被験者が研究に参加した結果として科学的に意義のある結果をもたらす
- 症例数設計の意義
 - 帰無仮説を棄却すること
 - 棄却できない場合には薬剤効果が意味のある大きさに至らなかったということを十分な根拠をもって伝えること

2006/11/12

36

症例数設計のプロセス

- 研究において証明したい仮説に相応しいエンドポイントを決める(通常、一つが望ましい)
- このエンドポイントに期待される『臨床的に有効であると考えられる最小の効果の大きさ』を慎重に見積もる
- エンドポイントの評価に適切な統計解析法を選択、有意水準 α 、検出力(1- β)を設定
- その効果の大きさが有意水準 α 、検出力(1- β)で有意となる最小の症例数を計算

$$n = 2(Z_{1-\alpha/2} + Z_{1-\beta})^2 \times \left(\frac{\text{ばらつき(標準差)}^2}{\text{臨床的に意味のある差}} \right)^2$$

- 両側有意水準5% → $Z_{1-\alpha/2} = 1.96$, 検出力80% → $Z_{1-\beta} = 0.84$

2009/11/12

37

症例数設計の事例1

- 冠動脈アテローム性硬化の抑制を目的としたプラバスタチン多施設共同定量的血管造影試験
- エンドポイント: 平均冠動脈血管径
- 臨床的に意味のある差: プラセボ群の平均冠動脈血管径の進展率が0.2mm縮小(SD=0.34mm)、プラバスタチン投与で50%の治療効果を期待
- 両側有意水準5%, 検出力80%と設定

$$n = 2(1.96 + 0.84)^2 \times \left(\frac{0.34}{0.2 - 0.1} \right)^2 = 181.3$$

この研究仮説を検証するためには各群 181例必要

2009/11/12

38

症例数設計の事例2

- 糖尿病予防を目的とした新しい栄養指導の効果をII型糖尿病境界型と診断された男子勤労者を対象に検証するために、従来の栄養指導法を対照とした2重盲検並行群間比較試験を計画

- エンドポイント: 1年度の血糖値(負荷後120分値)
- 臨床的に意味のある差: ベースラインの血糖値は平均135mg/dl, 新しい栄養指導法は血糖値は13.5mg/dl (SD:31.5)減少する。ただし、従来の治療では2.5mg/dl減少する
- 両側有意水準5%, 検出力:80%, 脱落率15%と設定

$$n = 2(1.96 + 0.84)^2 \times \left(\frac{31.5}{13.5 - 2.5} \right)^2 = 128.5$$

脱落を考慮してこの仮説を検証するには各群 129/(1-0.15)=151.5=152例必要

2009/11/12

39

シンプソンのパラドックス

- タモキシフェン使用の有無と乳がんの再発⁸⁾

タモキシフェン	再発あり	再発なし	対象者数
使用	464	2085	2549
非使用	424	1928	2352
合計	888	4013	4901

タモキシフェン使用群再発割合: 18.2%
タモキシフェン非使用群再発割合: 18.0%
再発割合の差: 0.2%
⇒タモキシフェンの乳がん再発予防効果はみられない

- リンパ節転移の有無で層別

タモキシフェン	リンパ節転移有り		リンパ節転移無し			
	再発あり	再発なし	再発あり	再発なし		
使用	368	847	1215	96	1238	1334
非使用	253	507	760	171	1421	1592
合計	621	1354	1975	267	2659	2926

転移あり再発割合の差: -3.0%
転移あり再発割合の差: -3.5%
⇒タモキシフェンの乳がん再発予防効果あり

2009/11/12

40

交絡

- 「タモキシフェン使用グループがタモキシフェンを使用しなかった場合の再発割合」と「タモキシフェン非使用グループの再発割合」が異なる状況を交絡という
- 今回の事例では
 - リンパ節にがんが転移していると、乳がん再発の可能性が高い
 - この研究当時は、リンパ節に転移のあった患者を含む再発の可能性の高い患者に選択的にタモキシフェンが使用されていた

交絡要因であるリンパ節転移の有無を考慮した解析を行わないと、正しい結果はえられない

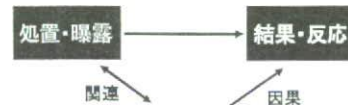
2009/11/12

41

交絡因子を取り除く方法



- ランダム化により処置と交絡因子との関連がなくなる



2009/11/12

42

交互作用

	効果		対象者数	男性		女性	
	あり	なし		効果あり	対象者数	効果あり	対象者数
実薬	90(30%)	240	300	60(40%)	150	30(20%)	150
プラセボ	45(15%)	225	300	30(20%)	150	15(10%)	150

■ 効果の割合の差

- 全体: 30%-15%=15%
- 男性: 40%-20%=20%
- 女性: 20%-10%=10%

■ 男女が完全にバランスしているのでバイアスではない

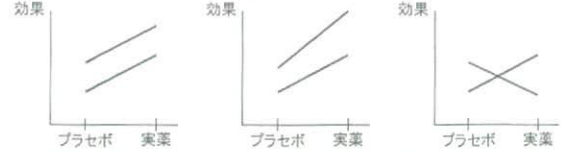
ある水準で効果の大きさが異なることを交互作用という

2009/11/12

43

交互作用の種類

■ 交互作用なし ■ 量的交互作用 ■ 質的交互作用



■ 交互作用なし: 男性における治療効果と女性における治療効果の大きさと方向がほぼ同じ

- 量的交互作用: 性別で治療効果の大きさは異なるが、方向は異なる
- 質的交互作用: 性別で治療効果の方向が異なる

2009/11/12

44

交絡と交互作用の解釈

■ 両者はよく間違われるが全く性質が異なるもの

■ 交絡がみられた場合

- 重要な予後因子が群間で不均衡である
- 研究計画に問題があった
- 交絡なし⇒研究計画の妥当性の保証

■ 交互作用がある場合

- 治療効果が、サブグループ間で異なる
- 薬剤の適応の限定
- 交互作用なし⇒一般化可能性のある効果

交絡は積極的に除去すべきもの
交互作用は治療効果が異なることを報告すべきもの

2009/11/12

45

脳代謝改善薬

脳疾患4薬 効かず

厚生省「承認取り消し」
痴呆症に多用 売上総額800億円

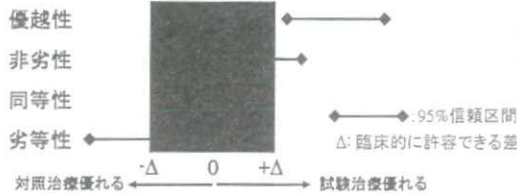
大蔵省再評価を要求

プラセボ対象の再試験実施したところ、有効性を検証できず

承認当時は、ホバテを対象薬として同等性試験が実施された

46

優越性試験・非劣性試験



- 優越性(superiority): 『臨床的に無意味な差が臨床的に有意』となる減少をさげ、 Δ 以上の差があることを示す
- 非劣性(non-inferiority): Δ 以上劣っていないことを示す
- 同等性(equivalence): Δ 以内の差であれば同等と示す

2009/11/12

47

非劣性試験の意義

■ 試験治療には対照治療にないメリットがあって、多少有効性が劣っても患者のコンプライアンス向上などの利点があるものを評価できる

■ 非劣性マージン(Δ)の導入

- 決定的に決めることはできないため、臨床家と統計家が相談して決める臨床的に許容できる差
 - 抗菌薬では10%が良く使われている
 - 実薬とプラセボの効果の差の2分の1ないし、3分の1

■ 被験者数の比較

試験治療の効果	対照治療の効果	優越性試験	非劣性試験	
60%	60%	∞	364	$\Delta=10\%$
62%	60%	9368	250	両側有意水準5%
65%	60%	1484	166	検出力:80%

2009/11/12

48

クロスオーバー試験(cross-over)



引用: 東京大学 大橋博雄先生 臨床試験登録システム シンポジウム (2005年2月2日) 講演スライドより
2008/11/12 49

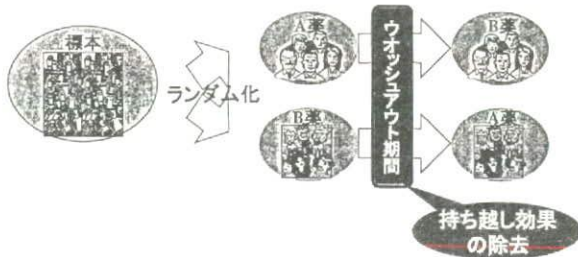
クロスオーバー 試験

- 同一対象に時期を変えて異なる介入を実施, 各治療法についての個体内の評価を得る
- 応用場面
 - 治療法の効果発現が速やかで, かつ中止後は患者が元の状態にすぐに戻るもの
 - 軽い慢性的な疾患など
 - 対象とする疾患が, 短期間に改善が期待できるもの
 - 軽い喘息, 皮膚疾患の貼付剤など
 - 被験者の病状が安定しているもの
 - 安定した糖尿病, 睡眠薬など
- 長所: 並行群間比較試験に比べ症例数が少なく済む
- 短所: 持ち越し効果が存在すると結果にバイアス

2008/11/12

58

クロスオーバーデザイン



2008/11/12

51

中間解析について¹⁰

- 中間解析とは
 - 試験が正式に完了する前に行われる有効性または安全性に関してキーを開けて試験治療群間の比較をする解析
 - 中間解析の目的
 - 試験を早期に中止すること
 - 研究中の試験治療の優越性が疑いなく立証された場合
 - 適切な試験治療の差を示す見込みのないことが判明した場合
 - 許容できない有害作用が明らかになった場合
 - 中間解析の長所・短所
 - 長所: 被験者を科学的・倫理的に保護可能
 - 短所: 情報量の不足, 試験の質が低下する可能性あり
- 長期試験や生命に関わる試験では良く実施される**

2008/11/12

52

中間解析を実施するための要件

- 合理的な臨床試験なのに中間解析が必要
 - 長期試験, 重篤な副作用, 悪い忍容性, ばらつき全く不明, 有効性が劣る群
- 用いる手法は妥当で, 確立したものである
 - 第一種の過誤の適切な制御
- 判断が独立かつ適切に行われる保証がある
 - 独立データモニタリング委員会
 - 委員は, 研究関係者や利害関係者であってはダメ
- 試験に偏りが生じない情報秘匿の保証がある
 - 守秘義務, 情報伝達の制約, 標準手順の事前明記, 任意監査が可能

2008/11/12

53

最近の流行?

- アダプティブデザイン
 - 試験の効率化を図るために, 中間解析の結果を利用して, 試験登録基準, 治療薬剤の割付け, 症例数などを適応的に変更するデザイン
 - 症例数の再設定
 - 被験者割付け比率の変更
 - 治療群の選別的継続
 - 被験者母集団の変更
 - 優越性試験と非劣性試験の変更
- ベイズ流のアプローチ
 - がんのPhase IにおけるCRM (Continual Reassessment Method)
- クラスターランダム化
 - 参加者を例えば施設ごとにランダム化

2008/11/12

54

今日のまとめ

- 臨床研究の3つの目標を達成するために研究計画及び解析計画の工夫をする
 - 精度を高くする
 - 症例数設計, データの評価方法
 - 比較可能性をあげる
 - ランダム化や盲検化を適用
 - 一般化可能性を保証する
 - 適格基準・除外基準の設定, 交互作用の評価, サブグループ解析, ITT解析

新たなエビデンスを千葉大から世界へ発信し、
良い医療を確[®]していきましょう

2009/11/12

55

引用文献1

- [1]厚生省薬務局新医薬品課, 抗悪性腫瘍薬の臨床評価方法に関するガイドライン, 2005.
- [2]名郷直樹. EBM実践ワークブック. 南江堂, 1999.
- [3] David Salsburg. The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century, W. H. Freeman, NY, 2001.
- [4] Cochran, WG. & Cox, G.M. Experimental Designs, 2nd ed., NY: John Wiley & Sons, 1957.
- [5] Medical Research Council. Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment of pulmonary tuberculosis. British Medical Journal 1948; 2: 769-83.
- [6] 佐久間昭. 佐久間昭の世界-日本の薬効評価と生物統計の歩み 1996. サイエントリスト社

2009/11/12

56

引用文献2

- [7] The Coronary Drug Project. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. N Engl J Med 1980; 303: 1038-41.
- [8] 佐藤俊哉. 交絡 事実と反実の比較. 科学 2008; 78: 427-429
- [9] 村岡了一・小出俊彦. 臨床試験のための中間解析. サイエントリスト社. 2004.
- [10] 大橋靖雄. 臨床試験のデザイン. Surgery Frontier 2003; 10: 422-426
- [11] 椿広計, 佐藤俊哉, 藤田利治. これからの臨床試験—医薬品の科学的評価 - 原理と方法. 朝倉書店. 1999.

2009/11/12

57

お薦め!

- これからの臨床試験
医薬品の科学的評価
—原理と方法
- 椿広計, 藤田利治,
佐藤俊哉
- 朝倉書店
- 1991年



2009/11/12

58

*Thank you for your
kind attention!*

2009/11/12

59

生物統計学入門

第3回 生物統計方法論I

千葉大学医学部附属病院
臨床試験部
Harvard School of Public Health,
Department of Biostatistics

佐藤 泰憲

今日のお話

- 1 推定
 - 1 点推定
 - 1 区間推定
 - 信頼区間
- 1 検定
 - 1 帰無仮説と対立仮説
 - 1 2種類の過誤
 - 1 有意水準とP値
 - 1 統計学的有意と医学的有意
 - 1 パラメトリックとノンパラメトリック
 - 1 検定の多重性の問題

2

統計的推測

対象母集団

統計的推測

標本

10人中7人に有効

推定

●の有効率は何%くらい?

検定

●の有効率は何%だ
●の有効率は何%でない

どっちが正しい?

- 1 推定: 母集団における薬剤Aの有効率の値を知りたい
- 1 検定: 母集団における薬剤Aの有効率が想定した値といえるか否かという二者択一の結論

3

点推定とは

- 1 薬剤Aを10人の患者に投与した結果

有効	無効	合計
7人	3人	10人

- 1 母集団における薬剤Aの有効率は70%であると推定する方法を点推定(point estimation)という
- 1 点推定により得られた推定値を点推定値という
- 1 母集団における薬剤Aの有効率をたったひとつの値として推定しており、ばらつきまで考慮した手法ではない
- 1 標本が無作為に選ばれている場合、標本から得た平均値や比率は、母集団の母平均値と母比率の点推定値となることが知られている

4

区間推定

対象母集団

10人の患者

点推定値 X_1

10人の患者

点推定値 X_2

...

10人の患者

点推定値 X_{n-1}

10人の患者

点推定値 X_n

- 1 母集団における薬剤Aの有効率の点推定値はばらつくため、ばらつきを考慮して母集団での有効率がほぼ確実に含まれるような区間を推定する方法を区間推定(interval estimation)という
- 1 区間推定により推定された値を信頼区間(confidence interval)という

5

点推定値の分布

- 1 母集団における薬剤Aの有効率の確率分布は理論上、正規分布に従う
- 1 点推定値のバラツキを評価するために標準誤差を利用する

6

(復習) 標準偏差と標準誤差

- 1 標準偏差(Standard Deviation: SD)
 - 1 データのばらつきを示す指標の一つ
- 1 標準誤差(Standard Error: SE)
 - 1 データではなく推定値のばらつきを示す指標の一つ
 - 1 平均値に限らず、推定値の標準偏差を『標準誤差』という

$$SE = \frac{SD}{\sqrt{n}}$$

- 標準誤差はサンプル数(n)に依存する
- 標準誤差は必ず標準偏差より小さくなる

2009/11/17

7

有効率の点推定値とその標準誤差

- 1 n人の患者に治療してx人に有効であれば

1 有効率の点推定値: $\hat{P} = \frac{x}{n}$

1 有効率の点推定値の標準誤差: $SE = \sqrt{\frac{x \cdot n(1-x \cdot n)}{n}}$

- 1 事例)薬剤Aにおいて10人中7人が有効

1 有効率の点推定値は7/10=0.7

1 有効率の点推定値の標準誤差は $\sqrt{\frac{0.7(1-0.7)}{10}} = 0.1449$

2009/11/17

8

信頼区間の推定



- 1 点推定値を中心として、その両側に『ある定数(K)×標準誤差』といった幅をもった区間として推定される

1 $(C_L, C_U) = (\text{点推定値} - K \times \text{標準誤差}, \text{点推定値} + K \times \text{標準誤差})$

- 1 Kの値の決め方

1 信頼区間の上限と下限を決めるためのもの

1 点推定値の分布と信頼係数により決まる

1 例)分布が正規分布、信頼係数が95%⇒K=1.96

- 1 事例での95%信頼区間

$(C_L, C_U) = (0.7 - 1.96 \times 0.1449, 0.7 + 1.96 \times 0.1449)$

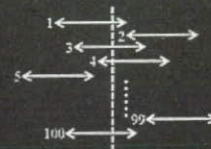
$= (0.416, 0.984)$

2009/11/17

9

信頼係数の意味

- 1 薬剤Aを10人の患者に投与することを100回繰り返すと、100個の95%信頼区間が得られる



- 1 95%信頼区間とは、100個の信頼区間のうち95個の信頼区間が、母集団における薬剤Aの有効率を含んでいると期待される区間

2009/11/17

10

被験者数による信頼区間の影響

- 1 被験者数を50人にした場合

有効	無効	合計
35人	15人	50人

1 点推定値: 35/50=0.70

1 点推定値の標準誤差: $\sqrt{\frac{0.7(1-0.7)}{50}} = 0.065$

1 95%信頼区間: $(0.7 - 1.96 \times 0.065, 0.7 + 1.96 \times 0.065)$

被験者数	標準誤差	95%信頼区間
10人	0.1449	0.416 - 0.984
50人	0.0648	0.573 - 0.827
1000人	0.0144	0.671 - 0.728

点推定値に収束

2009/11/17

11

ここまでのまとめ

- 1 点推定とは、母数(母平均値など)をデータに基づいて1つの数値を用いて推定する方法

1 推定した値を点推定値

- 1 区間推定とは、ある統計学的な確からしさで区間を設定し、その範囲内に母数(母平均値など)が存在することを示す方法

1 推定した区間を信頼区間

1 95%信頼区間とは、100回の同一の研究を行い、同一の計算方法を用いれば、95回はこの信頼区間の中に母平均値が含まれることを意味する



2009/11/17

統計的仮説検定とは

(Statistical hypothesis testing)

- 1 ある疾患に対して既存の治療薬の有効率が50%であったとする
- 1 この疾患の患者10人に薬剤Aを投与したところ下記のような結果が得られた

有効	無効	合計
7人	3人	10人

- 1 母集団における薬剤Aの有効率が、従来の治療薬の有効率50%とは違うといえるか否かを検討したい

薬剤の効果の有無などの
二者択一の判定を行う方法

2025.11.17

13

帰無仮説と対立仮説

- 1 検定をおこなうために、帰無仮説(null hypothesis)と対立仮説(alternative hypothesis)を設定する

- 1 自分が主張したいこと⇒対立仮説
- 1 自分の主張と反対のこと⇒帰無仮説

- 1 先ほどの例

- 1 帰無仮説(H_0): 母集団における薬剤Aの有効率は50%である(既存治療と差がない)
- 1 対立仮説(H_1): 母集団における薬剤Aの有効率は50%ではない(既存治療と差がある)

2025.11.17

14

2種類の過誤

(第一種の過誤と第二種の過誤)

- 1 『帰無仮説が真であるか』、『対立仮説が真であるか』を検定という手法を用いてデータから判断したい
- 1 その判断には、2種類の誤りの可能性が存在
- 1 第一種の過誤: 帰無仮説が正しいにもかかわらずそれを棄却する誤り⇒ α エラー
- 1 第二種の過誤: 対立仮説が正しいにもかかわらず対立仮説を採択しないという誤り⇒ β エラー

検定の判定	神のみが知る治療効果の真実	
	差がない	差がある
有意差なし	正しい($1-\alpha$) specificity	第二種の過誤 false-negative
有意差あり	第一種の過誤 false-positive	正しい($1-\beta$) sensitivity(power)

注) 有意差:意味のある差

2025.11.17

15

2種類の過誤の例

- 1 事例: 結婚

- 1 真に結婚すべき相手と結婚できれば正
- 1 本当は結婚すべきでない相手を断ることができれば正
- 1 あせって結婚すべきでない相手と結婚したら
⇒第一種の過誤(α エラー: α ateの α)
- 1 優柔不断で結婚すべき相手を見逃してしまったら
⇒第二種の過誤(β エラー: β onyariの β)

あまい α 大 β 小 ← 意思決定 → 厳しい α 小 β 大

2025.11.17

16

2種類の過誤確率の決め方

- 1 第一種の過誤確率と第二種の過誤確率は両方同時に小さくすることはできない
⇒第一種の過誤に重点を置いて、第一種の過誤確率が有意水準以下(たとえば5%)になるようにして、第二種の過誤確率が小さいものを探す
- 1 たとえば、新薬の臨床試験において、本当に効果がない薬が誤って世に出る可能性を避けるために第一種の過誤確率を小さく設定する
 - 1 有意水準を5%に設定すれば、効かない薬が偶然で世に出るのを20回に1回に抑える

2025.11.17

17

有意水準とは

- 1 第一種の過誤をある値以下に抑えるようにするため、事前に設定する確率の値のことを有意水準という
- 1 通常は有意水準を5%ないし1%に設定
- 1 有意水準を5%にする根拠
 - 1 統計学的な根拠はなく、慣習的に用いられている
 - R.A. Fisherが活躍した農事試験では、大学を卒業してから20年の現役生活を送る。現役時代に1回くらい誤りをおかしても許されるということで $1/20=0.05$ になったというあてにならない話もある。

2025.11.17

18

P値とは

- 1 P値のPはProbability⇒確率
- 1 実際には効果がない(差がない)にもかかわらず、偶然によって効果あり(差がある)と判断する確率
 - 1 P値大: 偶然でも起こりそう
 - 1 P値小: 偶然では起こりそうにない
 - ⇒ 偶然を超えた意味のある(有意)差
- 1 P値は小さければご利益のあるもの
- 1 どれくらい小さいと良いか?
 - 有意水準以下、すなわち5%あるいは1%以下

設定した帰無仮説が正しいという前提のもとでデータから実際に得られた結果が起こる確率

2025.11.17

19

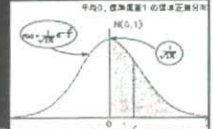
P値の計算の仕方

- 1 帰無仮説に基づいた確率分布において、データとしての検定統計量の値以上の値が得られる確率を計算する
- 1 割合の差を検出したい場合
 - 1 割合の差($p_1 - p_2$)は帰無仮説のもとで標準正規分布に従う
 - 1 割合の差の検定はZ検定(すなわち検定統計量はZ統計量)

$$Z = \frac{p_1 - p_2}{SE}, \quad SE = \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) p_0(1 - p_0)}$$

- 1 標準正規分布の確率密度関数を積分

$$P\text{-value} = \int_z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$



2025.11.17

20

仮説検定の原理

- 1 数学の「背理法」の考えの応用
 - 1 「意義」のある命題の「否定」を仮説として設定し、それをデータで「否定」する二重否定の論理で命題を証明
- 1 背理法とは?
 - ⇒ ある主張(命題)を真とすると、そこから不条理な結果が生じることを示して、この主張を否定(棄却)する論法
 - 1 命題: 犬はしろい
 - この命題を証明するために白い犬をいくら集めてきても証明できない
 - しかし、黒い犬(白くない犬)を少なくとも1匹連れてくれば命題を否定できる



2025.11.17

21

検定の原理と結果の解釈



- 1 帰無仮説が棄却された場合
 - ⇒ 対立仮説を積極的に支持⇒積極的な結論づけ
- 1 帰無仮説が棄却されなかった場合
 - ⇒ 帰無仮説を積極的に支持せず⇒結論は保留

2025.11.17

22

検定の手順

- 1 帰無仮説と対立仮説を設定する
- 1 有意水準を決める
- 1 検定手法を決め、検定統計量を計算し、データの差が偶然によって生じる確率(P値)を計算
- 1 P値があらかじめ決めた有意水準より小さければ、帰無仮説が間違っていたとして、有意差ありと結論づける

2025.11.17

23

検定の事例1: いかさまコイン

- 1 このコインが表が出やすいのではと疑いがかけられた
- 1 帰無仮説と対立仮説を設定する
 - 1 帰無仮説: 表と裏が出る確率は等しい
 - 1 対立仮説: 表と裏の出る確率は等しくない
- 2) 有意水準を決める
 - 1 5%と設定
- 3) 実験を行い、データをとる⇒表が5回続けて出たとする
- 4) 帰無仮説の下で、実験結果が得られる確率(P値)を計算
 - 1 表が5回出る確率(0.5)⁵=0.03125
- 5) 有意水準5%以下なので帰無仮説を棄却
 - ⇒ このコインはイカサマだと結論づける



2025.11.17

24

検定の事例2: 既存治療との比較

1 既存治療の効果が50%であるが、ある患者10人に薬剤Aを投与したら7人に有効であった

1) 帰無仮説と対立仮説を設定する

- 1 帰無仮説: 母集団における薬剤Aの有効率は50%である
- 1 対立仮説: 母集団における薬剤Aの有効率は50%ではない

2) 有意水準を決める

- 1 5%と設定

3) 帰無仮説の下で、このような結果が得られるP値を計算

- 1 P値=0.2059

5) 有意水準5%より大きいので帰無仮説を棄却できない
⇒このデータから結論づけることができない

2008/11/17

25

両側検定と片側検定

1 薬剤Aの有効率は50%とは違うか否かという検討

- 1 有効率が50%よりも大きいか、あるいは小さいか大小関係については考慮していない

- 1 有効率が50%とは違うか否かを検討する場合は両側検定

- 1 有効率が50%より大きい(小さい)を検討する場合は片側検定

1 イカサマコインの例(表が5回連続出現⇒イカサマ)

- 1 裏が5回連続して出る場合もあり、表裏の両方が出現する確率は $0.03125 + 0.03125 = 0.0625$ ⇒ 両側検定のP値

1 多くの場合、両側検定のP値は片側検定のP値の2倍

1 何も断らずに検定というと、両側検定を意味する

1 片側で有意差ありとなったが、両側で有意差が得られない微妙な場合もあり、事後的に向きを変更してはダメ!

2008/11/17

26

検定における被験者数の影響

1 既存治療の効果が50%であるが、ある患者n人に薬剤Aを投与したらx人に有効であった

$$Z = \frac{p_1 - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.7 - 0.5}{\sqrt{0.5(1-0.5)/n}}$$

$$P\text{-value} = P(x \leq -z, z \geq x)$$

被験者数(n)	有効(x)	標準誤差	Z	P値
10人	7人	0.158	1.265	0.205
20人	14人	0.111	1.801	0.071
30人	21人	0.091	2.197	0.028
40人	28人	0.079	2.531	0.011
50人	35人	0.070	2.828	0.0048

←Z=1.96の時 P=0.05

被験者数を増やすと標準誤差が小さくなり
その結果としてP値も小さくなる

2008/11/17

27

統計学的有意と医学的有意

1 例: 標準薬を対照とした降圧薬の試験

Type	拡張期血圧の平均値の差	被験者数	P値	判定
A	-20 mmHg	100	0.012	問題なくpositive
B	-5 mmHg	100	0.531	問題なくnegative
C	-20 mmHg	30	0.170	nが小さすぎた
D	-5 mmHg	1000	0.048	nが大きすぎた

2008/11/17

28

統計学的有意と医学的有意

1 統計学的有意差は、『偶然を超えた差がある』ことを意味するが、必ずしも医学的に意義のある差を反映しない場合がある

1 症例数を増やせば、解析の感度が増して、医学的に意味のない小さな差であっても、偶然ではないと判定してしまう

1 統計学的有意症(Significantosis) by 佐久間昭

『統計学的有意≠臨床的有意』ということ踏まえた上で、臨床的な立場から検定の結果を正しく意味づけることが重要

2008/11/17

29

検定結果の表記の注意点

1 習慣として、*: P<0.05 **: P<0.01表記されることが良くある

1 P値の値そのものを示す方がよい。

- 1 P=0.0090とP=0.00010では解釈が異なる

- 1 P値が示されていれば、事後的に有意水準を変更することが可能

1 P値を表記するときは有効数字2桁(多くても3桁)

1 検定の種類と仮説の方向を明記

2008/11/17

30

パラメトリックとノンパラメトリック

1 パラメトリック検定

- 1 母集団が、正規性(正規分布などの特定の分布に従うこと)や等分散性を仮定する方法

1 ノンパラメトリック検定

- 1 母集団が特定の分布に従うことを前提にしない方法
- 1 符号、順位など情報を使って、検定統計量を求める

2023.11.17

31

パラメトリックとノンパラメトリックの特徴

	パラメトリック	ノンパラメトリック
分布形の仮定	正規分布	必要なし
等分散性	仮定	仮定
第1種の過誤	α	常に α 以下
正規分布のとき	◎	○
外れ値が存在	×	○
変数変換	変	不変
$n < 6$	△	×

2023.11.17

32

ノンパラメトリック法の注意点

- 1 正規性からのズレや外れ値などに対してノンパラメトリック検定法は頑健である⇒頼りすぎるのは危険
- 1 パラメトリック法を適用できる条件がそろっているにもかかわらずノンパラメトリック法を用いると、検出力が低下するという問題が生ずる
 - 1 パラメトリック法の t 検定の効率が 1 として、ノンパラメトリック法の中央値検定の効率は 0.63、Wilcoxon 検定の効率は 0.95
 - 1 標本の大きさが n であるとき、 t 検定と同じ検出力を期待するならば、中央値検定では $n/0.63 = 1.59n$ の標本が必要
- 1 外れ値の原因を追究しないまま、手法に頼りすぎると実験全体の信頼性を欠く恐れがあるため、各手法の特性を知って適切な方法を選ぶ必要がある

2023.11.17

33

検定の繰り返しに

- 1 臨床研究において複数の評価項目や多種多様な臨床検査項目を測定することがある
- 1 そのひとつひとつに対して、治療効果があるかないかを検定をする場合がよくある
- 1 例) ラットを用いた毒性試験

	対照群	低用量群	中用量群	高用量群
平均値	926.0	911.9	891.5	893.0
標準偏差	25.7	20.1	39.8	35.4
N	10	10	10	10

	平均値の差	P値
対照群-低用量群	-14.1	0.1895
対照群-中用量群	-34.5	0.0333
対照群-高用量群	-33.0	0.0281

2023.11.17

34

検定の繰り返しの問題

- 1 検定を有意水準5%で行うと、本当に効果がないときに間違っって効果ありと判定する確率は5%である
- 1 100個の臨床検査項目について検定を100回行くと、5項目くらいは処理効果がないのに間違っって効果ありと判定することになる

2023.11.17

35

検定の繰り返しの問題

- 1 一つの研究で3回の検定を実施した場合
 - 1 1回の比較あたりの有意水準を5%としても、実際には3回の比較を行っているため、3回の比較全体で、偶然に有意差が出る確率は5%よりはるかに大きくなる
 - 1 3回の比較全体での第1種の過誤確率

$$FWER = 1 - (1 - 0.05)^3 = 0.142$$

ひとつの研究において、複数回の検定を同時に行うことによって、偶然に有意になる確率が大きくなる現象を検定の多重性という

2023.11.17

36

検定の多重性の問題が生じる状況

- 1 複数のエンドポイント
- 1 研究全体で有意でないときに、有意となるサブグループを探す
 - 1 男性・うお座・B型に効果あり?
- 1 繰り返し観測して、有意となったところを報告
- 1 さまざまな解析をおこなって、有意となったものを報告
- 1 多数の群があるとき、2群比較を多数行う

2009.11.17

37

多重性の問題への対処

- 1 検定の回数を減らす
 - 1 検証目的の研究では、必要もないのに検定を行うことを極力避ける
- 1 検定に優先順位をつける
 - 1 主要評価項目と副次評価項目
- 1 分散分析を先に行って、有意水準をコントロール
 - 1 分散分析を行って有意差があれば、多重性を考慮しない検定で群間比較
- 1 多重比較法を用いる
 - 1 ひとつの研究において複数の比較を同時に行うとき、比較の組み全体として、どれかひとつを誤って有意となる確率を有意水準以下に抑える方法

2009.11.17

38

よく用いられる多重比較法

1 Bonferroni法

- 1 m 個の帰無仮説からなる仮説族全体の有意水準を α にするために、個々の有意水準を α/m とする
 - 100,000 遺伝子を有意水準0.05で検定した場合、 P 値が 5×10^{-7} より小さければ疾患感受性遺伝子と判定



$$P(E_1 \cup E_2 \cup E_3) \leq P(E_1) + P(E_2) + P(E_3)$$

1 Holm法(Bonferroni法の改良法)

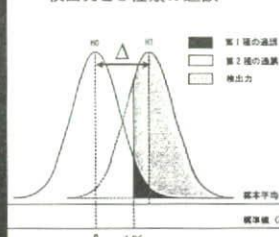
- 1 帰無仮説族全体の有意水準を α にするため、 m 個の帰無仮説の個々の有意確率を計算して小さい順に、 i 番目を有意水準 $\alpha/(m-i+1)$ で棄却か保留か判断、1つでも保留になったら、それ以後は全部保留

2009.11.17

39

症例数設計

検出力と2種類の過誤



1 症例数設計に必要なパラメータ

- 1 検定の有意水準 α
 - 通常は5% $\Rightarrow Z_{\alpha} = 1.96$
- 1 検出力(1- β)
 - 通常は80% $\Rightarrow Z_{\beta} = 0.84$
- 1 臨床的に意味のある差 Δ
- 1 ばらつきの大きさSD

$$N = 2(Z_{\alpha/2} + Z_{\beta})^2 SD^2 / \Delta^2$$

2009.11.17

40

プロトコルや論文への症例数設定根拠の記載

- 1 症例数の設定根拠
- 1 症例数設計に必要な4つのパラメータの値を記載
- 1 解析に用いる統計手法の記載
- 1 高血圧の治療薬の症例数設定根拠記載例
 - 1 『先行研究の結果より、

そこで $\Delta=15$, $SD=15$ と見積もり、 t 検定において80%以上の検出力が期待できる症例数設計を行った。ただし、両側有意水準を5%とする。

以上より、 $N=2(1.96+0.84)^2 \times 15^2/15^2=15.68$ となり、1群あたりの必要な被験者数を16例と見積もった。』

2009.11.17

41

推定と検定の関係

- 1 既存治療の効果が50%であるが、ある患者 n 人に薬剤Aを投与したら x 人に有効であった
 - 1 母集団における薬剤Aの効果を95%信頼区間により推定
 - 1 『母集団における薬剤Aの有効率は50%である』という帰無仮説を設定し、有意水準5%で検定

被験者数(n)	有効(x)	推定の結果 95%信頼区間	検定の結果 P値
10人	7人	0.416 - 0.984	0.205
20人	14人	0.499 - 0.900	0.071
30人	21人	0.536 - 0.863	0.028
40人	28人	0.557 - 0.842	0.011
50人	35人	0.573 - 0.827	0.0048

有効率の95%信頼区間が帰無仮説で設定している

50%を含んでいる場合は、検定結果も有意でない

2009.11.17

42

今日のまとめ

- 1 母集団について統計的推測を行う場合は、推定と検定の二つの方法がある
- 1 推定
 - 1 点推定:母数(母平均など)を1つの値として推定
 - 1 区間推定:母数(母平均など)を含む領域として推定
 - 1 95%信頼区間を示すことが有効
- 1 検定
 - 1 帰無仮説と対立仮説
 - 1 第1種の過誤と第2種の過誤
 - 1 有意水準とP値
 - 1 統計学的有意⇔医学的有意
 - 1 検定を複数回実施する場合⇒多重性の問題が生じる

2008.11/17

43

参考文献

- [1] 佐久間昭, 酒井弘憲, 佐藤泰憲. 医薬統計Q&A 2007. 金原出版
- [2] 椿美智子, 椿広計(訳). 医学研究のための統計的方法. 2001. サイエンティスト社
- [3] 永田靖. サンプルサイズの決め方. 朝倉書店. 2003
- [4] 永田靖, 吉田道弘. 統計的多重比較法の基礎. 1997. サイエンティスト社
- [5] 浜田知久馬. 学会・論文発表のための統計学. 1999. 真興交易医学出版部
- [6] 東京大学教養学部統計学教室. 自然科学の統計学. 1992. 東京大学出版会
- [7] 吉村功. 数理統計学. 1969. 培風館.
- [8] 吉村功. 毒性・薬効データの統計解析. 1987. サイエンティスト社

2008.11/12

44

生物統計学入門

第4回 生物統計方法論II

千葉大学医学部附属病院
臨床試験部
Harvard School of Public Health,
Department of Biostatistics

佐藤 泰憲

今日のお話

- 1 回帰分析
 - 1 回帰分析の流れ
 - 1 誤差の概念と4つの条件
 - 1 残差プロット
- 1 ロジスティック回帰
 - 1 ロジスティック回帰とは
 - 1 回帰係数とオッズ比の関係
- 1 生存時間解析
 - 1 生存時間データとは
 - 1 Kaplan-Meier plotとは
 - 1 ログランク検定・一般化Wilcoxon検定の違い
 - 1 Cox回帰とは

2

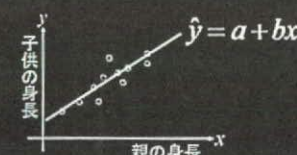
回帰(regression)とは

- 1 生物統計学者であり、「優生学」の創始者F.Galton (1822-1911)の着想
- 1 Galtonの主張
 - 1 背の高い親を持つ子の身長は親ほどには高くない
 - 1 逆に、背の低い親を持つ子の身長は親ほどには低くない
 - 1 集団平均からのずれの著しい親の子は親よりずれが小さくなる傾向がある
 - 1 第2世代は平均の方向にもどってゆく現象を回帰とよんだ

3

回帰分析(regression analysis)とは

- 1 変数間の“関係”を解析する統計的手法
 - 1 子どもの身長と親の身長の関係は？



- 1 データに基づいてaやbを推定
 - yを従属変数、目的変数といい
 - xを独立変数、説明変数という

4

回帰分析の応用例

- 1 ビール会社のビール製造計画
 - 1 応答変数を数ヶ月先の生ビール生産量として独立変数を各年の生産量・各月の生産量・天気予報(気温)などから生産量を予測し、製造計画を立てる
- 1 クレジットカードに加入する際の評価
 - 1 独立変数を年収・勤続年数・年齢・持家などから応答変数を信用評価として、加入者に支払い能力があるかどうかを評価(推定)する

5

回帰分析の流れ

```

    graph TD
      A[仮説の設定] --> B[モデルの設定]
      B --> C[パラメータの推定]
      C --> D[モデルの妥当性の検討]
      D -- 妥当でない --> E[モデルの更新]
      E --> B
      D -- 妥当 --> F[モデルに基づく推測]
  
```

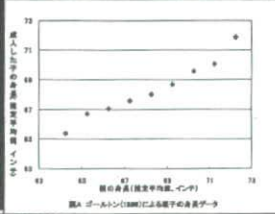
6

回帰分析の手順1・2

1 仮説の設定: 父親の身長と子どもの身長は関連がある(線型的な関係)

↳ Galtonのデータを使用

1 モデルの設定



$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

x_i : 説明変数

y_i : 目的変数

ε_i : 誤差

β_0, β_1 : パラメータ

$i = 1, \dots, n$

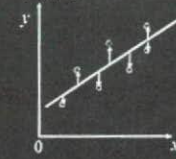
2025/11/19

7

回帰分析の手順3

1 パラメータの推定

↳ β_0, β_1 の推定法⇒最小2乗法を利用



$$Q^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$$

Q^2 が最小となる b_0 と b_1 を求める

⇒ b_0 と b_1 について偏微分方程式を解く

$$\frac{\partial Q^2}{\partial b_0} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) = 0$$

$$\frac{\partial Q^2}{\partial b_1} = -2 \sum_{i=1}^n x_i (y_i - b_0 - b_1 x_i) = 0$$

↳ Galtonのデータにおいて

$$- b_1 = 25.28, b_1 = 0.627 \Rightarrow y = 25.28 + 0.627x$$

2025/11/19

8

回帰分析の手順4

1 モデルの妥当性の検討

↳ モデルの妥当性の指標: 寄与率(R^2)

$$R^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

- 回帰モデルにより説明できたYの変動の割合
- 寄与率の取り得る範囲は0~1.0
- Rは重相関係数と呼ばれ、予測値と実測値の相関の度合

↳ Galtonの身長データにおける寄与率

$$\Rightarrow R^2 = 0.934$$

2025/11/19

9

回帰分析の手順5

モデルの更新

モデルの妥当性が認められないとき、再度モデルを設定

モデルに基づく推測

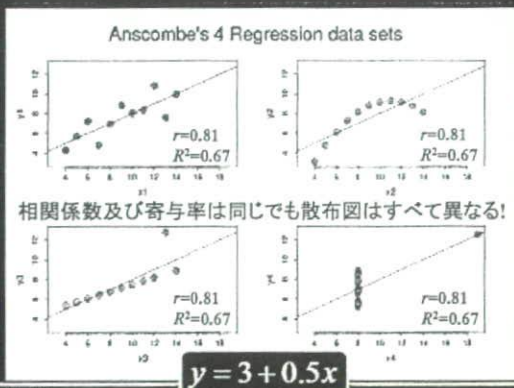
モデルの妥当性が認められたとき、推測が必要

1 疑問: そもそも R^2 の値のみでモデルの当てはまりを評価して良いか??

2025/11/19

10

【再掲】 Anscombeの相関図 (Anscombe 1973)



2025/11/19

11

Anscombeの教訓

- 1 統計量の数値のみによる判断だけで、モデルの当てはまり具合を評価することの危険性を示唆
- 1 グラフによる視覚的表現の重要性を強調!!

⇒ 統計量はモデルの誤差項の仮定が満たされる場合に限り有効

グラフにより残差を調べることが重要

2025/11/19

12

残差プロット

残差とは、実測値と予測値の差

$$e_i = y_i - \hat{y}_i$$

x_i	y_i	予測値	残差
10	8.04	8	0.04
8	6.95	7	-0.05
13	7.58	9.5	-1.92
9	8.81	7.5	1.31
11	8.33	8.5	-0.17
14	9.96	10	-0.04
6	7.24	6	1.24
4	4.26	5	-0.74
12	10.84	9	1.84
7	4.82	6.5	-1.68
5	5.68	5.5	0.18

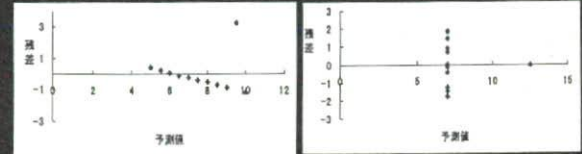
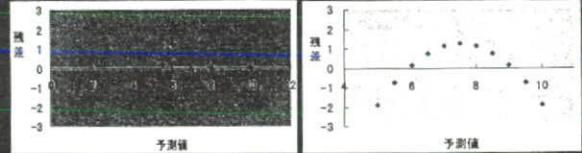
$$\hat{y}_i = 3 + 0.5x_i$$

残差プロット: 残差と説明変数、予測値または時間変数などの散布図

2008/11/19

13

残差プロットの例(Anscombeのデータ)



2008/11/19

14

残差プロットから分かること

モデルに含まれる誤差の仮定をチェック

線型モデルに基づく解析では誤差に4つの条件を必要とする

- 独立性
- 不偏性
- 等分散性
- 正規性(残差プロットから評価できない)

はずれ値の検出

モデルの妥当性を評価
適当に散らばっていればモデルは妥当

2008/11/19

15

誤差の4条件

誤差の概念

- 偶然性、不確定性に支配されている
- 数値ではあるがその値を知ることはできない

誤差の4条件

- 条件1: 独立性 \Rightarrow 誤差は独立
- 条件2: 不偏性 \Rightarrow 誤差の期待値が 0
 $E\{U_i\} = 0; i = 1, 2, \dots, n$
- 条件3: 等分散性 \Rightarrow 誤差の大きさは同じ
 $V\{U_i\} = \sigma^2; i = 1, 2, \dots, n$
- 条件4: 正規性 \Rightarrow 誤差の分布は正規分布

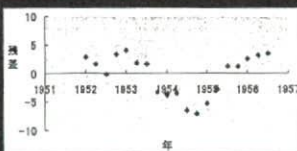
『誤差について、独立性、不偏性、等分散性、正規性、という条件を仮定した線型モデル』と『応答変数の分布は正規分布を仮定』は同値(必要十分条件)

2008/11/19

16

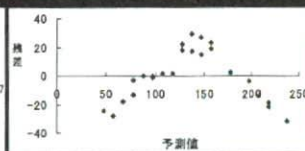
モデルに含まれる誤差の仮定のチェック

独立性



モデルの誤差項に相関がある

不偏性



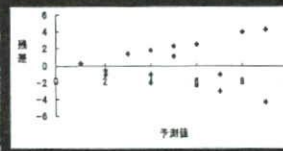
横軸の大きさによって偏りがある

2008/11/19

17

モデルに含まれる誤差の仮定のチェック

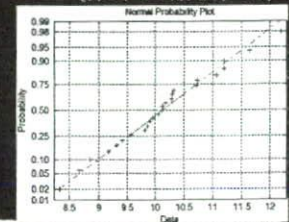
等分散性



横軸の大きさによってばらつきが大きくなる

正規性

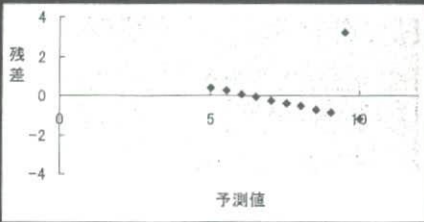
残差プロットでは評価できない
 \Rightarrow Q-Qプロットの利用
(第1回の講義参照)



2008/11/19

18

残差プロットによる外れ値の検出



1 少数個だけ飛び離れている

2156年五輪の100m最速は女性？

Oxfordのグループが過去100年のオリンピックのデータを分析し、予測したもの (Tatem et al., Nature 2004) **brief communications**

Momentous sprint at the 2156 Olympics?

Women sprinters are closing the gap on men and may one day overtake them.

The 2004 Olympic women's 100-metre sprint champion, Yelena Isinbayeva, is assured of fame and fortune, but we show here that — if current trends continue — it is the winner of the event in the 2156 Olympics whose name will be etched in sporting history forever, because this may be the first occasion on which the race is won in a faster time than the men's event.

The Athens Olympic Games could be viewed as another giant experiment in human athletic achievement. Are women narrowing the gap with men, or falling further behind? Some argue that the gains made by women in running events between the 1970s and the 1990s are decreasing as the women's achievements plateau. Others contend that there is no evidence that athletes, male or female, are reaching the limits of their potential.

In a limited test, we plot the winning times of the men's and women's Olympic 100-metre sprints over the past 100 years (ref. 3), for data sets, see sup-

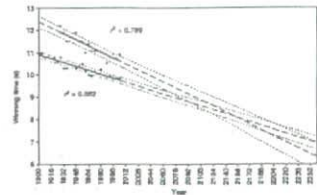


Figure 1 The winning Olympic 100-metre sprint times for men (left curve) and women (right curve) with superimposed best-fit regression lines. The regression lines are shown for men and women, respectively, and the dashed curves (dotted lines) are based on the available data at the time of writing. The projection that the women will overtake the men is based on the regression lines.

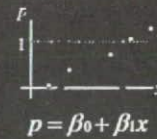
ここまでのまとめ

- 1 応答変数と説明変数の相関関係を表現したり、予測式を作る場合に回帰分析は役立つ
- 1 寄与率は、妥当性の評価指標であるが、誤差の4条件(独立性, 不偏性, 等分散性, 正規性)が成り立たない場合に適用すると誤った解釈を導く可能性がある
- 1 残差プロットを利用してデータのバラツキを視覚化することが有効

回帰分析の拡張

- 1 ここまでの話は、目的変数(y)が正規分布に従うことを前提としていた
- 1 目的変数が正規分布に従わない場合は??
- 1 事例: 毒物の量 x に対する死亡率 p のモデル

毒物の量 x_i	死亡数 y_i	合計 n_i
101	0	10
136	2	10
183	5	10
247	8	10
333	9	10
450	10	10



x が大きくなると p は [0, 1] を超える
等分散性もみだされない

ロジスティック回帰分析

- 1 応答変数(y)が二値変数の時に用いる回帰分析
- 1 二値変数とは
 - 1 二つのカテゴリーからなるカテゴリカルデータ(0/1データ)
 - 薬剤の有効・無効
 - 性別(男・女)
 - 2者択一のアンケート調査(Yes, No)
- 1 ロジスティック回帰の適用例
 - 1 予測: 大腸がんの予後予測
 - 1 推定: 大腸がんの予後に影響を与えるものを調べる



ロジスティックモデルとは

- 1 ロジット(オッズの対数を取ったもの)変換

$$\text{logit}[P(Y=1)] = \log \left[\frac{P(Y=1)}{1-P(Y=1)} \right]$$

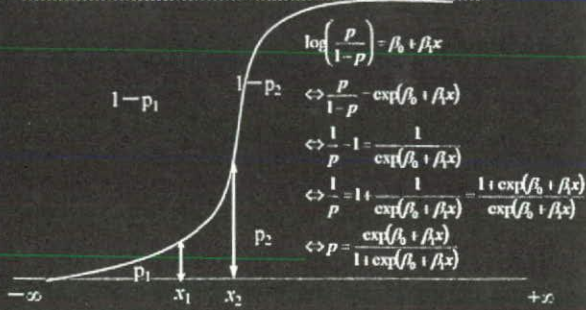
- 1 ロジットにすることで応答変数の範囲が $[-\infty, +\infty]$

- 1 ロジスティックモデル

$$\log \left[\frac{P(Y=1)}{1-P(Y=1)} \right] = \beta_0 + \beta_1 x$$

ロジスティック曲線

イベント発現確率 p



2008/11/19

25

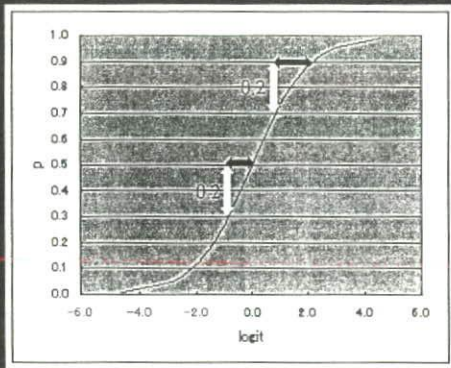
ロジットを用いる利点

- 1 範囲が $[-\infty, +\infty]$ となる
- 1 確率に対する重み付けが実質科学的
 - 1 治療効果を10%あげる場合、30%を40%にするのと70%を80%にするのでは大変さが異なる
 - ⇒ロジットではそれが考慮できる
- 1 偏回帰係数(切片や傾き)の解釈がしやすい
 - ⇒傾きがオッズ比

2008/11/19

26

確率に対する重み



2008/11/19

27

偏回帰係数の解釈

- 1 説明変数が一つで実薬とプラセボを比較する場合

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$$

1 $x=0$: プラセボ $x=1$: 新薬とする

$$x=0 \text{ のとき } \log\left(\frac{p_0}{1-p_0}\right) = \beta_0; \quad x=1 \text{ のとき } \log\left(\frac{p_1}{1-p_1}\right) = \beta_0 + \beta_1$$

$$\log\left(\frac{p_1}{1-p_1}\right) - \log\left(\frac{p_0}{1-p_0}\right) = \beta_1$$

← 対数の性質より $(\log X - \log Y) = \log(X/Y)$

$$\log\left(\frac{p_1}{1-p_1}\right) - \log\left(\frac{p_0}{1-p_0}\right) = \beta_1$$

$$\left(\frac{p_1}{1-p_1}\right) / \left(\frac{p_0}{1-p_0}\right) = \exp(\beta_1)$$

オッズ比

2008/11/19

28

【再掲】リスク比とオッズ比

1 コホート研究

1 男女50人ずつの集団を一定期間観察してイベントの発生を観察

性別	イベント		合計
	あり	なし	
男性	n_{11}	n_{12}	50
女性	n_{21}	n_{22}	50
合計	$n_{.1}$	$n_{.2}$	100

1 評価指標: リスク比

$$RR = \frac{n_{11}/(n_{11}+n_{12})}{n_{21}/(n_{21}+n_{22})}$$

1 ケースコントロール研究

1 肺がん患者50人と健常人50人について過去の喫煙歴を調査

喫煙歴	肺がん	健常人	合計
あり	n_{11}	n_{12}	$n_{1.}$
なし	n_{21}	n_{22}	$n_{2.}$
合計	50	50	100

1 評価指標: オッズ比

$$OR = \frac{n_{11}/n_{21}}{n_{12}/n_{22}}$$

発症が稀な場合
リスク比とオッズ比

2008/11/19

29

事例: 喫煙と脳卒中の関連

- 1 喫煙と脳卒中中の関連を調べた疫学研究結果

喫煙	脳卒中なし	脳卒中あり
あり	64	98
なし	211	159
合計	275	257

脳卒中なし群の喫煙割合: 23.3%
脳卒中なし群の喫煙割合: 38.1%
オッズ比: $(0.38 / (1-0.38)) / (0.23 / (1-0.23)) = 2.05$

- 1 ロジスティックモデルの設定

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \times \text{smoking}$$

- 1 解析結果(SASの出力)

	推定値	標準誤差	カイ2乗	P値	Exp(推定値)
β_0	-0.283	0.105	7.25	0.0071	
β_1	0.709	0.192	13.6	0.002	2.03

喫煙は脳卒中中のリスクを約2倍にする

2008/11/19

30

ここまでのまとめ

- 1 2値データに対して線形モデルを仮定した回帰分析を行うことはできるが妥当でない
- 1 予測値が0~1に収まらない
- 1 破綻する可能性がある
- 1 応答変数が2値データのときはロジスティック回帰分析が有効
- 1 確率の重みづけを考慮可能
- 1 回帰係数がオッズ比として解釈可能

2025/11/19

31

プラバスタチンの大規模臨床試験

- 1 高脂血症の患者さんに対してコレステロール減少効果を持つプラバスタチンの心疾患による死亡減少効果の評価

The New England Journal of Medicine
Copyright, 1995, by the Massachusetts Medical Society
 Volume 333 SEPTEMBER 16, 1995 Number 39

PREVENTION OF CORONARY HEART DISEASE WITH PRAVASTATIN IN MEN WITH HYPERCHOLESTEROLEMIA

John A. SPECTOR, M.D., Steven M. COHEN, M.D., Leo FROH, PH.D., Christopher G. FRANK, M.D., A. Ross LEVINE, M.D., Peter W. MAYER, PH.D., James H. MCKENNEY, M.D., and Christopher J. TORPACON, D.Sc., for the Heart Protection Project Group*

Abstract. Background: Lowering the blood cholesterol level may reduce the risk of coronary heart disease. This double-blind study was designed to determine whether the administration of pravastatin to men with hypercholesterolemia and no history of myocardial infarction or death from coronary heart disease would reduce the risk of death from coronary heart disease. Methods: A total of 20,325 men, aged 40 to 64 years of age, with a mean LDL cholesterol level of 277.22 mg per deciliter (7.18 mmol per liter) were randomized to 40 mg each evening of pravastatin (10 mg each evening of placebo). The average follow-up period was 5.0 years. End points were all-cause mortality, cardiovascular mortality, and the combined death, nonfatal myocardial infarction, and stroke. Results: Pravastatin lowered plasma cholesterol levels by 25 percent and low-density lipoprotein cholesterol levels by 35 percent, whereas there was no change with placebo. There were 248 deaths among men in the pravastatin group and 262 among men in the placebo group. The relative risk of death from coronary heart disease was 26 percent lower in the pravastatin group (95 percent confidence interval, 15 to 40 percent; P<0.001). Conclusions: Treatment with pravastatin significantly reduced the occurrence of myocardial infarction and death from coronary heart disease without adversely affecting the risk of death from noncardiovascular causes in men with moderate hypercholesterolemia and no history of myocardial infarction. (N Engl J Med 1995;333:1361-71.)

2025/11/19

32

プラバスタチンの大規模臨床試験 方法: 統計解析

Statistical Analysis

All data were analyzed according to the intention-to-treat principle. The results of the two fasting lipoprotein profiles obtained during visits 2 and 3 were averaged to produce base-line values. The LDL cholesterol results were analyzed according to both the treatment actually received and the intention-to-treat principle. The analysis based on actual treatment used only the measured lipid levels in subjects who had attended the previous scheduled visit and who had been seen at work trial medication at that visit. For the intention-to-treat analysis, all randomized levels were included, without reference to the subject's degree of compliance at previous visits. In addition, in cases in which the lipid value was available for a scheduled visit and no medication had been found at the previous visit, the subject's base-line level was used. For each end-point category, the lengths of time to a first event were compared with use of the log-rank test, and the relative reduction in risk resulting from pravastatin treatment, with 95 percent confidence intervals, was calculated with the Cox proportional hazards model.¹⁷ In addition, Kaplan-Meier survival curves¹⁸ were used to estimate the absolute risk of each event at five years for each treatment group. When a silent myocardial infarction was detected on the basis of serial comparison of ECGs, the event was considered to have occurred midway between the first diagnostic ECG and the previous ECG. Two-sided P values were used throughout.

For the primary end point, an analysis was performed for predefined subgroups¹⁹ characterized at base line according to age (<55 years or ≥55 years), smoking status (nonsmoker or nonsmoker of cigarettes, cigars, or pipes), and whether at least two of the following risk factors were present: smoking, hypertension, a history of chest pain of intermittent character (as indicated by positive responses on the Rose questionnaire), diabetes, and a silent ECG abnormality associated with coronary heart disease (Minnesota code 4-2, 4-3, 5-2, or 5-3).

log-rank test
Cox proportional hazard model
Kaplan-Meier

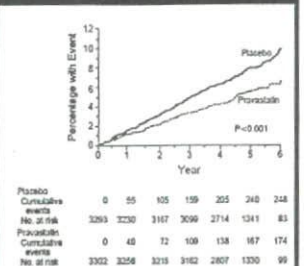
2025/11/19

33

プラバスタチンの大規模臨床試験 解析結果

Table 2. End Points of the Study*

Treatment	Pravastatin (n=10,162)	Placebo (n=10,163)	P Value	Relative Risk (95% CI)
All-cause mortality (n=248/10,162 vs 262/10,163)				
Stroke	104 (1.0%)	110 (1.1%)	0.85	0.97 (0.79-1.20)
Myocardial infarction	104 (1.0%)	110 (1.1%)	0.85	0.97 (0.79-1.20)
Death from other causes	140 (1.4%)	142 (1.4%)	0.92	1.00 (0.82-1.22)
Cardiovascular mortality (n=248/10,162 vs 262/10,163)				
Death from coronary heart disease	104 (1.0%)	142 (1.4%)	0.001	0.70 (0.55-0.90)
Death from stroke	104 (1.0%)	110 (1.1%)	0.85	0.97 (0.79-1.20)
Death from other causes	140 (1.4%)	142 (1.4%)	0.92	1.00 (0.82-1.22)
Other events				
Stroke	104 (1.0%)	110 (1.1%)	0.85	0.97 (0.79-1.20)
Myocardial infarction	104 (1.0%)	110 (1.1%)	0.85	0.97 (0.79-1.20)
Death from other causes	140 (1.4%)	142 (1.4%)	0.92	1.00 (0.82-1.22)
Death from stroke	104 (1.0%)	110 (1.1%)	0.85	0.97 (0.79-1.20)
Death from other causes	140 (1.4%)	142 (1.4%)	0.92	1.00 (0.82-1.22)



2025/11/19

34

生存時間(survival time)とは

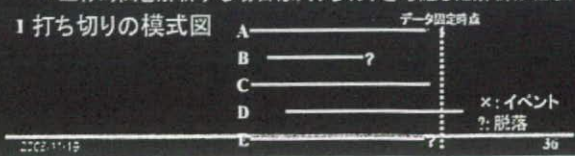
- 1 興味あるイベント(事象)が発生するまで、追跡することで得られるイベント発生までの時間
- 1 臨床研究におけるイベントの例
 - 1 プラバスタチンの試験においては、非致死性の心筋梗塞、心筋梗塞が原因の死亡などが発生するまでの時間
 - 1 アジスロマイシンの研究では、気管支炎が治り職場復帰するまでの時間
 - 1 肺癌の増悪・再発までの時間
 - 1 ICUにおいて人工呼吸器の離脱するまでの時間
- 1 生存時間を測定する研究において、脱落や転院などにより打ち切り(censoring)が生じることがよくある

2025/11/19

35

打ち切り(censoring)とは

- 1 ある時点までイベントが生じていなかったことは分かるが、それ以降のいつにイベントが発生したか正確に分からない
- 1 データ固定時点においてイベントが起きていない
- 1 生存時間を測定する研究では避けられない
 - 1 被験者全体がイベントが発生するまで追跡すると時間とコストがかかり研究の効率が下がる
 - 1 生存時間を解析する場合は、打ち切りを考慮した解析が必要



2025/11/19

36

なぜ生存時間解析が必要？

- 1 打ち切りを考慮した解析をするため生存時間解析という手法が生まれた
- 1 生存期間が長い患者ほど打ち切りを受けやすく、打ち切りを無視して解析すると結果にバイアスが生じる
- 1 時間を対象とするため、データは正の値であり、右に裾の長い分布になる(正規分布でない)

2009/11/19

37

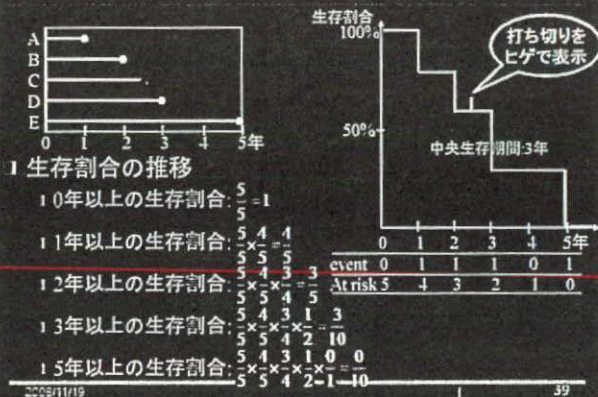
打ち切りを考慮した解析 (生存時間解析の手法)

- 1 生存時間分布の記述
 - 1 Kaplan-Meier法
- 1 生存時間の検定
 - 1 ログランク検定, 一般化ウイルクソン検定
- 1 ハザード比の推定
 - 1 Cox比例ハザードモデル

2009/11/19

38

Kaplan-Meier法



2009/11/19

39

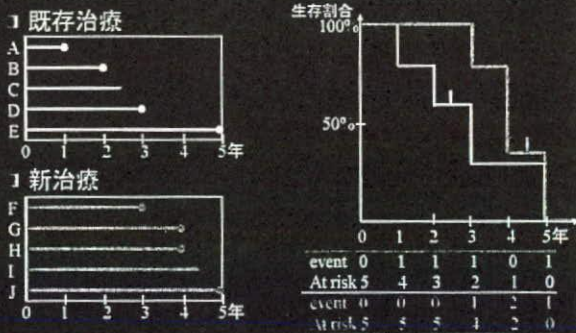
Kaplan-Meier plotの見かた

- 1 時間とともに階段の落ち方は $1/N$
 - 1 例) $0.05 \rightarrow n=20$
- 1 あまり急激に変化する場合は信頼できない
- 1 時点の後ろの方は信頼性が乏しい
 - 1 信頼区間の幅が広がる
- 1 打ち切り症例の数は？
- 1 打ち切り症例は特定の群に偏ってないか？

2009/11/19

40

2群間での生存曲線の比較



2009/11/19

41

2群間の生存曲線の比較(検定)

- 1 既存治療と新治療の生存曲線に差があるかどうか知りたい
- 1 帰無仮説: 2群の生存曲線に差がない『 $H_0: S_A = S_B$ 』
- 1 対立仮説: 2群の生存曲線に差がある『 $H_1: S_A > S_B$ 』
- 1 生存曲線の比較はノンパラ検定が有効
 - 1 生存時間は非負であり、右に裾をひいた分布
 - 1 順位に基づく検定(ログランク検定, 一般化Wilcoxon検定など)が良く使われる
- 1 検定統計量

$$-2 \ln \frac{\prod_{t=1}^k \frac{d_t}{n_t}}{\prod_{t=1}^k \frac{d_{1t}}{n_{1t}} \frac{d_{2t}}{n_{2t}}}$$

- d_t : 時点 t における観測死亡数

- d_{jt} : 時点 t における観測死亡数

2009/11/19

42