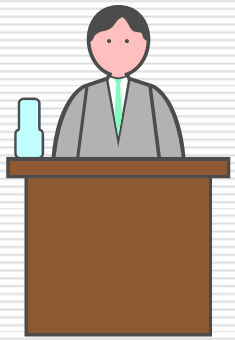
第5回　Young Investigators' Seminar
平成23年5月30日

Iwate Medical University

臨床研究デザインと医用統計の実践法入門②

岩手医科大学　消化器・肝臓内科
遠藤龍人
ryuendo@iwate-med.ac.jp



前回お話したこと

UNIVERSITY

- 臨床研究の現状と課題
  - 医学研究論文の信頼性を高めることを目指す新しい取り組み
  - STARDイニシアチブ（診断研究の報告）
- リサーチクエスチョンから臨床研究デザインへ
- 医学検査の有用性を評価する時のポイント
  - 信頼性評価：診断診察データの一致率とkappa値
  - 妥当性評価：感度、特異度、ROC曲線
- サンプルサイズの設定
  - 感度分析

研究におけるセオリーの重要性

『我流は攻において威を発するが、
守に転じて威を失う』

ラ王（北斗の拳）

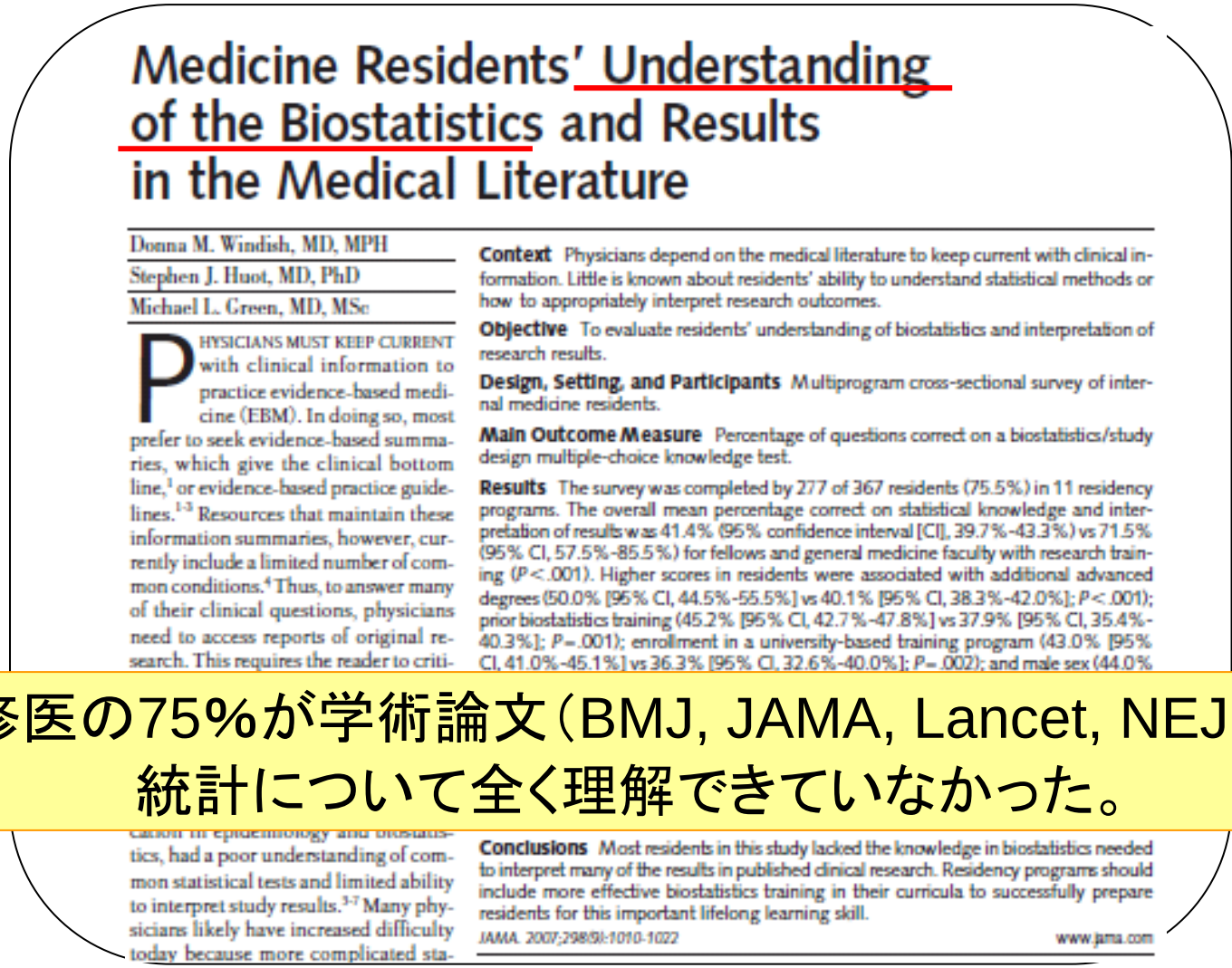
∵失敗をした時、短時間・短距離で修正可能
研究のLimitationを踏まえて討論できる

今日のアウトライン

UNIVERSITY

- 生物統計学の教育の現状
- 臨床研究デザインにおけるバイアスの制御
  - 統計学の位置づけ
- データの要約
- 統計学的検定と信頼区間
  - 平均値の差の比較方法、P値
  - パラメトリック検定とノンパラメトリック検定
- 信頼区間

レジデントに対する“生物統計学”の教育プログラムの必要性



Medicine Residents' Understanding of the Biostatistics and Results in the Medical Literature

Donna M. Windish, MD, MPH
Stephen J. Huot, MD, PhD
Michael L. Green, MD, MSc

Context Physicians depend on the medical literature to keep current with clinical information. Little is known about residents' ability to understand statistical methods or how to appropriately interpret research outcomes.

Objective To evaluate residents' understanding of biostatistics and interpretation of research results.

Design, Setting, and Participants Multiprogram cross-sectional survey of internal medicine residents.

Main Outcome Measure Percentage of questions correct on a biostatistics/study design multiple-choice knowledge test.

Results The survey was completed by 277 of 367 residents (75.5%) in 11 residency programs. The overall mean percentage correct on statistical knowledge and interpretation of results was 41.4% (95% confidence interval [CI], 39.7%-43.3%) vs 71.5% (95% CI, 57.5%-85.5%) for fellows and general medicine faculty with research training ($P<.001$). Higher scores in residents were associated with additional advanced degrees (50.0% [95% CI, 44.5%-55.5%] vs 40.1% [95% CI, 38.3%-42.0%]; $P<.001$); prior biostatistics training (45.2% [95% CI, 42.7%-47.8%] vs 37.9% [95% CI, 35.4%-40.3%]; $P=.001$); enrollment in a university-based training program (43.0% [95% CI, 41.0%-45.1%] vs 36.3% [95% CI, 32.6%-40.0%]; $P=.002$); and male sex (44.0%

Conclusions Most residents in this study lacked the knowledge in biostatistics needed to interpret many of the results in published clinical research. Residency programs should include more effective biostatistics training in their curricula to successfully prepare residents for this important lifelong learning skill.

JAMA. 2007;298(9):1010-1022 www.jama.com

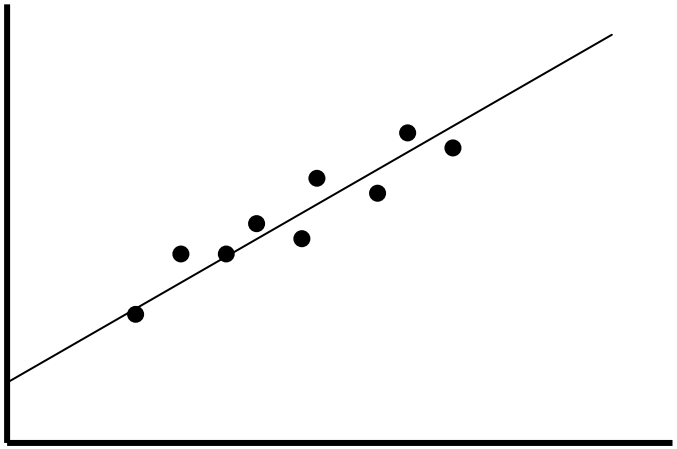
PHYSICIANS MUST KEEP CURRENT with clinical information to practice evidence-based medicine (EBM). In doing so, most prefer to seek evidence-based summaries, which give the clinical bottom line,[1] or evidence-based practice guidelines.[1-3] Resources that maintain these information summaries, however, currently include a limited number of common conditions.[4] Thus, to answer many of their clinical questions, physicians need to access reports of original research. This requires the reader to criti-

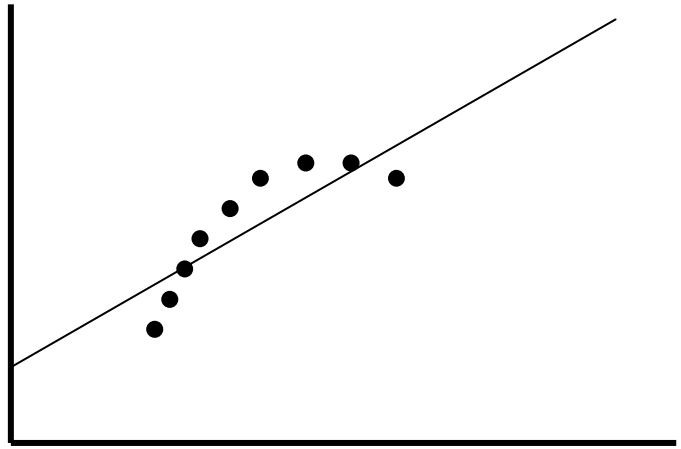
cation in epidemiology and biostatistics, had a poor understanding of common statistical tests and limited ability to interpret study results.[3-7] Many physicians likely have increased difficulty today because more complicated sta-

研修医の75%が学術論文（BMJ, JAMA, Lancet, NEJM）の統計について全く理解できていなかった。

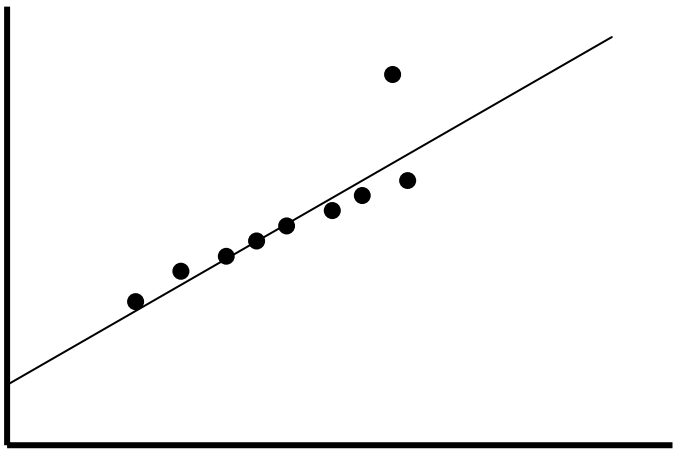
Donna MW JAMA. 2007;298:1010-22

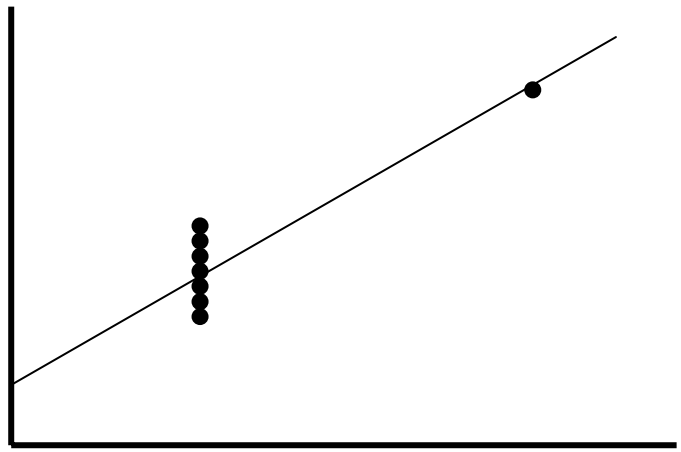
たとえば、相関係数=0.82





rやP値だけで強い結論を述べていることはないか？





少数の外れ値や変数の取りうる範囲に影響を受ける

モデルの診断（model checking）

- 相関係数

「２変数に直線的な関連があるか」をみるもの

- 変数間の影響の強さを計っているのではない

- P値は、帰無仮説「相関係数＝０」を検定している
- rやR^2だけでは弱い・・・
- 線形回帰の前提の検討

- 誤差に対する正規性
- 効果の加法性・均一性、直線的な関係

臨床研究の種類

1．記述疫学研究
（病気頻度、分布、診療パターン、自然歴）

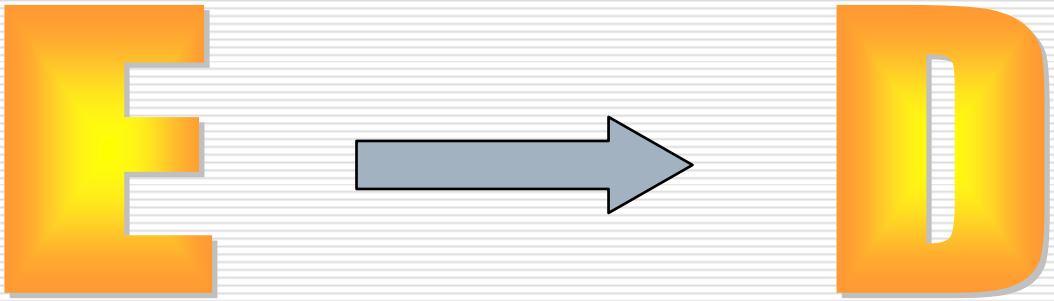
2．要因と害（or 益）との関連を分析する研究

3．治療・予防の有効性・安全性の評価

4．診断法の評価

臨床研究におけるリサーチ・クエスチョン

- 要因(治療)と 疾病(アウトカム)は関連あるか？



要因
(Exposure)

疾病
(Disease)

describe association between exposure and outcome

要因とアウトカムの関連を調べる

バイアス(Bias:偏り)とは？

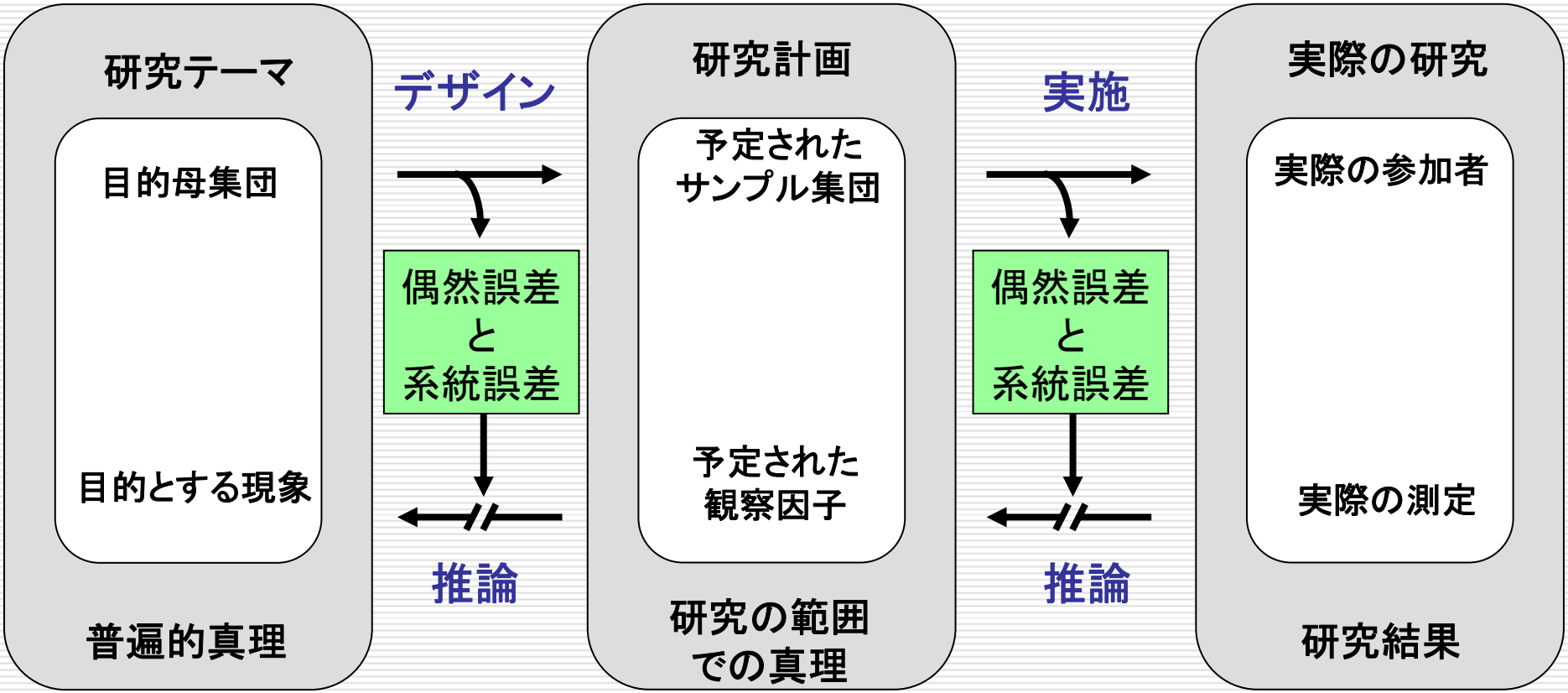
- 真の値から系統的に乖離した結果を生じさせる、あらゆる段階での推論プロセス

- 3大バイアス

1. 選択バイアス(selection bias)
2. 測定バイアス(meaurement/observation bias)
3. 交絡バイアス(交絡因子)(confounding bias/factor)

研究のダイナミズム（研究デザインが大切な理由）

Hulley S

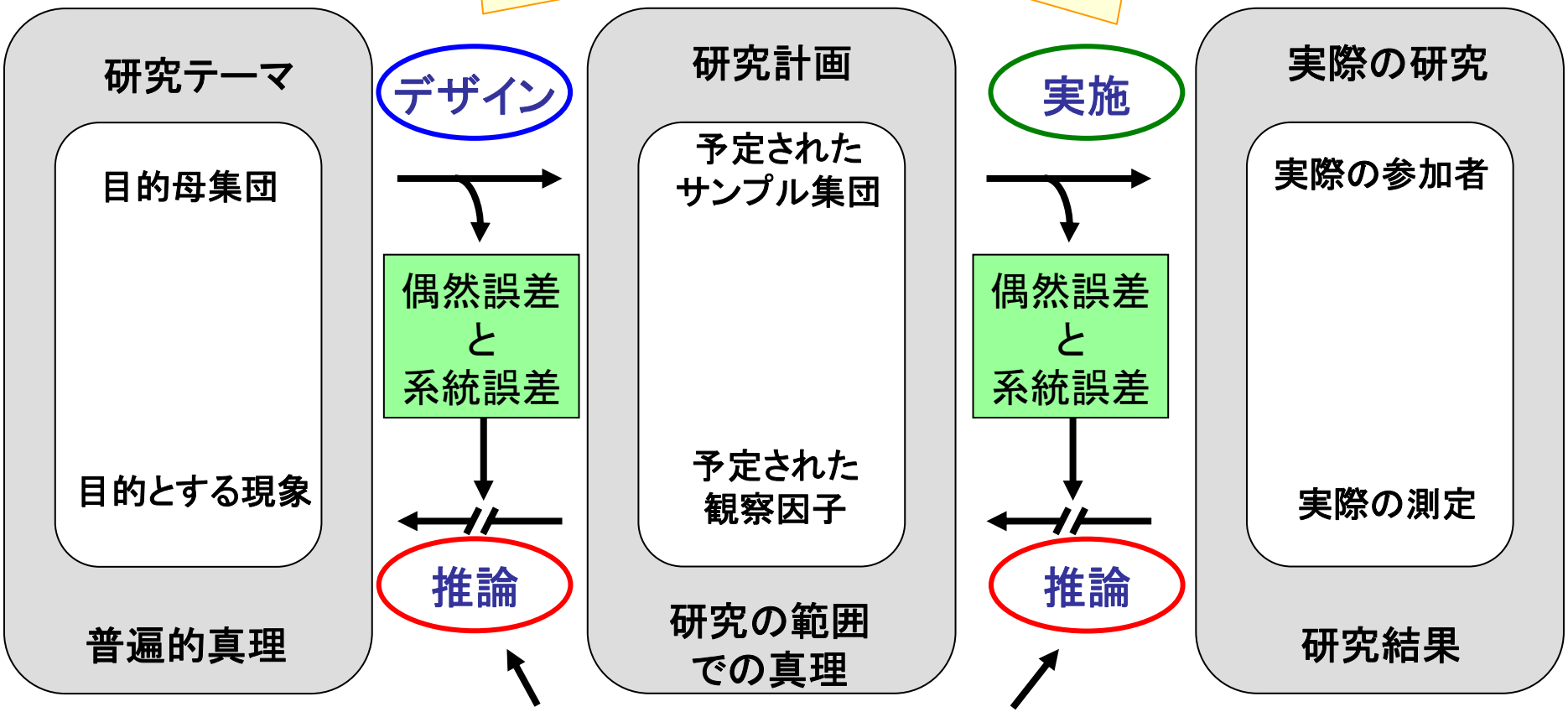


研究テーマに対する正しい解答が得られるかどうかは、研究デザインや実施の段階で、推論の妨げとなる誤差の混入をいかにうまくコントロールするかにかかっている。

研究のダイナミズム（研究デザインが大切な理由）

母集団をよく代表しているか？

正しく行われるいるか？



妥当な推測方法を用いているか？

例えば、「比較妥当性」の確保

- 選択バイアス⇒解析では除去できない
- 交絡の調整
  - <u>デザインの段階</u>
    - 限定・マッチング
    - ランダム化
  - <u>解析の段階</u>
    - 層別・サブグループ内で比較（標準化）
    - モデルによる調整（関数関係を仮定）
      - 線形回帰分析
      - ロジスティック回帰分析
      - Cox回帰分析

正当な評価のために
- Random sampling
- 同一施設の**連続症例**を対象
  - 研究期間の患者人数
  - 同意・非同意別の人数を明記

交絡要因の調整（回帰モデル）

治療効果

治療法：二次的手術

手術からの生存時間
手術からの無再発生存時間

共変量

手術時年齢	病期
組織型	Performance status
手術年	術式
残存腫瘍径	術後化学療法

共変量の効果

- 他の要因の影響を調整
- 交絡要因の調整

問題点

- 多変量解析でも交絡因子を調整しきれない

多変量解析とは

- F(転帰)=F(性)+F(年齢)+F(病型)+F(病期)+・・・

  - 複数の要因を独立してどの程度効いているかを数学モデルを使って計算するもの
    - 得られた係数 β は「要因」が1だけ増えると結果がどれほど変わるか」
  - 二値も連続量も同列に扱う
  - 設定する関数モデルに依存
    - スコア、相互作用
  - 説明変数が増えると信頼性が低下
  - 投入出来るのは既知の因子のみ

交絡要因

- 説明変数の数
- 多い：交絡の調整可能、リアルワールドに近い
  ⇔解析複雑、必要n数も多くなる
- 少ない：シンプル、解析もすっきり、必要n数少ない
  ⇔交絡、結論の弱さ

- 外してはいけない変数は入れる⇒文献検索！

  例：C型肝炎におけるIFN療法の調整要因（年齢、性、BMI）

コメント

- 多変量解析は個々の要因が独立してどれほど効いているかを知るもの
- まず単変量解析
- つづいて層別化して分析
- そして多変量解析に
  - 最適な関数モデルを設定
  - 投入する因子を厳選

統計の役割 ～2つ

- データを要約する
  - 代表値、データの散らばり具合など
- 検定・推定する
  - 平均値の差の検定、など
  - 有意水準（P値）、信頼区間

医学研究における統計手法の利用（探索的？/検証的？）

- 探索的（exploratory）データ解析
  - データに語らしむ：分布の形、外れ値
  - グラフを用いる：ヒストグラム、散布図
  - （外れに値に注意して）適切なデータの要約を行う
  - 手がかり（仮説）をデータから見つけ出す

- 検証的（confirmatory）データ解析
  - 証拠（仮説）の強さを評価：仮説検証
  - 統計的推測：検定と推定（主に統計的検定）
  - 試験計画・解析計画（事前宣言）

複数の手法の存在と選択

- 非常に多くの統計手法が存在
  - t-検定、χ^2検定、〇〇回帰・・・
- 統計ソフトによる実行
  - 現在では容易入手できる
- どの手法を選択するか？
  - 手法の選択の指針が不明確
- 良い教科書、生物統計家の不足
  - 数理もの（睡眠導入・・・）
  - ハウツーもの（時に応用きかず・・・）

統計手法の選択(1)

論文のTable 1.に相当

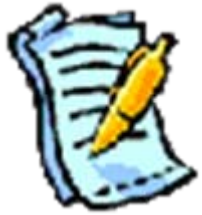
- 変数の数と解析方法
- 1変数
  - 記述統計量(平均、標準偏差、Max、Min)
- 2変数以上の関連
  - 1つの原因と1つの結果:
    - 単純な群間比較(t-検定、χ二乗検定)
  - 複数の原因と1つの結果:
    - 層別解析、回帰モデル
- (多変数相互)
  - (多変量解析諸種法・・・統計家に相談ください)

統計手法の選択(2)

- データ(変数)の型
  - 名義変数
  - 2値変数
  - 順序変数
  - 計数データ
  - 連続データ(生存時間データ)

- 目的および証明したい仮説
  - 群間比較、複数群、用量反応

変数による統計手法の分類（参考）



変数の型	2値	連続	生存期間
解析目的			
分布の記述	頻度集計 分割表	ヒストグラム 平均、S.D.、相関	Kaplan-Meier法
単純な群比較	χ^2検定 リスク推定	t 検定 平均値の差	log-rank検定 率の推定
回帰モデル	Logistic回帰	重回帰	Cox回帰
層別解析	標準化 Mantel-Haenszel法	分散分析	層別log-rank検定

データ解析の前に行うべきこと

- 最初から複雑なモデルで解析する？
  - 重回帰、ロジスティク回帰、Cox回帰・・・

- データ解析の準備体操
  - 欠測値の有無
  - 外れ値の評価
    - ヒストグラム
    - 箱ひげ図
    - 正規確率プロット
    - Kolmogorov-Smirnov検定
  - グラフ化、要約統計量の計算

データの要約（連続値）

- データ（変数）の種類にあわせて適切なグラフ
  - ヒストグラム、箱ヒゲ図、幹葉プロット

- データ解析の要約
  - 平均、標準偏差（S.D.）、Min-Max
  - 標準誤差（S.E.＝S.D./√n）
  - 平均± 95％信頼区間：データの分布と平均値の分布

・信頼区間については、後ほど・・・

平均、分散、標準偏差

データ：$x_1, x_2, x_3, \cdots x_n$

$$\text{平均} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\text{分散} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$$\text{標準偏差} = \sqrt{\text{分散}}$$

$$\text{標準誤差} = \frac{\text{標準偏差}}{\sqrt{n}}$$

標準偏差と標準誤差

- 標準偏差（Standard Deviation：S.D.）
  - データのばらつきを示す指標のひとつ
- 標準誤差（Standard Error：S.E.）
  - データではなく、推定値のばらつきを示す指標のひと
  - **推定値の標準偏差**

標準偏差と標準誤差のイメージ ～データの分布と推定値の分布

- 例：患者100名の血圧を測定
  - 血圧の分布
  - ヒストグラム、ボックスプロット
  - データの平均値、標準偏差、・・・・

- 「患者100名の血圧を測定する」研究を100回繰り返す
  - 血圧の平均値の分布
  - 平均値（推定値）の平均値、平均値の標準偏差

標準誤差

変数の種類　Types of variables

要約の仕方や持ち込む統計手法が異なる

質的データ		名義尺度 2値尺度 順序尺度
量的データ	比例性	間隔尺度 比尺度
	連続性	連続量 離散量

変数の種類　Types of variables

- 名義変数 nominal variable
  - 例：血液型{A, B, O, AB型}、{賛成、反対、未回答}など
  - 2つ以上のカテゴリー。順序なし
- 2値変数 binary valiable
  - 例：性別{男、女}、転帰{生存、死亡}
  - Alb値{＜3.5 g/dl、≧ 3.5 g/dl }
  - 2つのカテゴリー
- 順序変数 ordinal variable
  - {改善、不変、悪化}、がんの病期（ステージ）など
  - 何らかの自然な順序

変数の種類　Types of variables

- 連続変数 continuous variable
  - 間隔尺度と比尺度

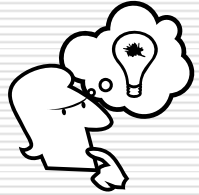
- 間隔尺度 interval scale
  - 距離尺度とも。比尺度と異なり、数値の差に意味がある（比例関係はない）
  - 例：温度：10℃から15℃になったときに、50％の温度上昇があったとは言わない。10℃→15℃、100℃→105℃とも5℃の温度上昇

変数の種類　Types of variables

- 比尺度 ratio scale
  - 比例尺度とも。**差とともに数値の比にも意味がある**。比が定義できる→絶対零点を持つ
  - 例：身長、体重など：**ゼロや負の値をとらない**ので比尺度
  - 例：同じ体重10 kg増加でも・・・

    50 →60 kg（ 20%増増加）

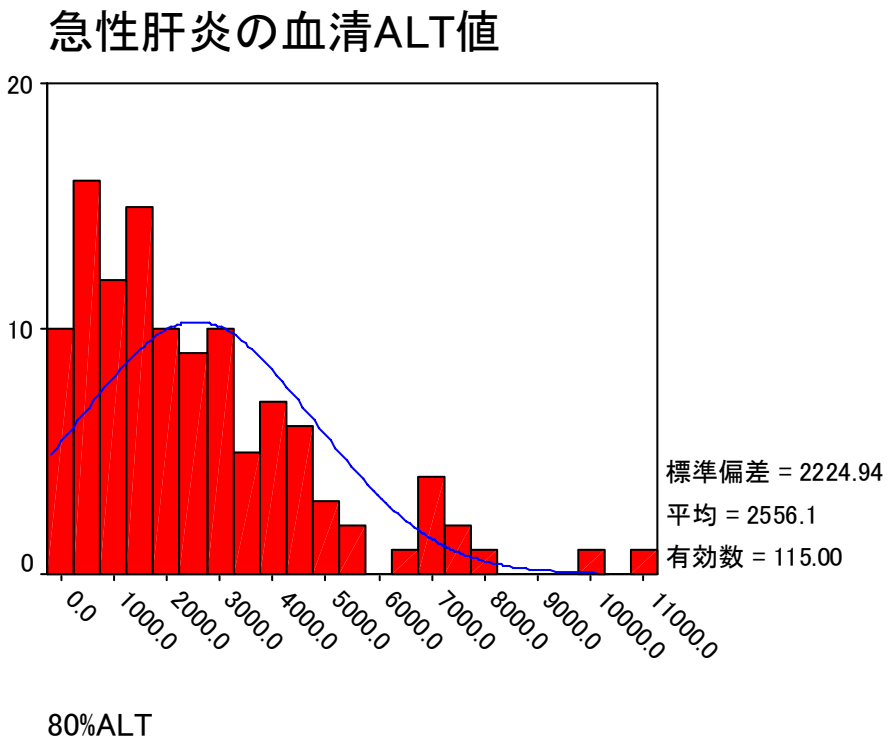
    100 →110 kg（10%増加）



ピットフォール：何でも平均値を出す（1）

- 急性肝炎患者の血清ALTの平均値と標準偏差を算出した・・・。
  - ALT：2556±2225

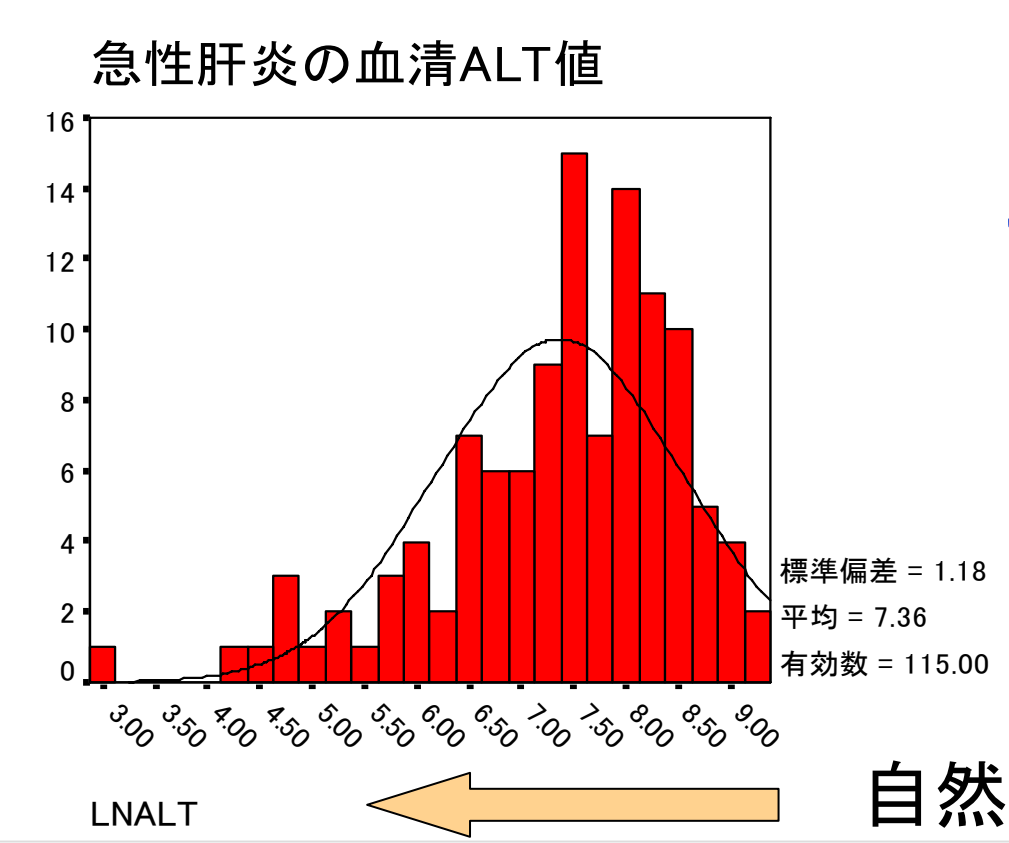
これでいい？



平均値と標準偏差

- 正規分布において成り立つ概念
- ヒストグラムを書いてみて正規分布パターンに近ければそのまま
- 関数変換により正規化できるものは正規化してから処理
- 正規化できないデータは中央値とパーセンタイル値（または分布範囲）で表現する
  - 1894 [825-3571]；25-75 パーセンタイル
  - 1894 (22-1193)；最小値～最大値

ALT値を自然対数に変換してみる



正規分布と言える？

自然対数に変換

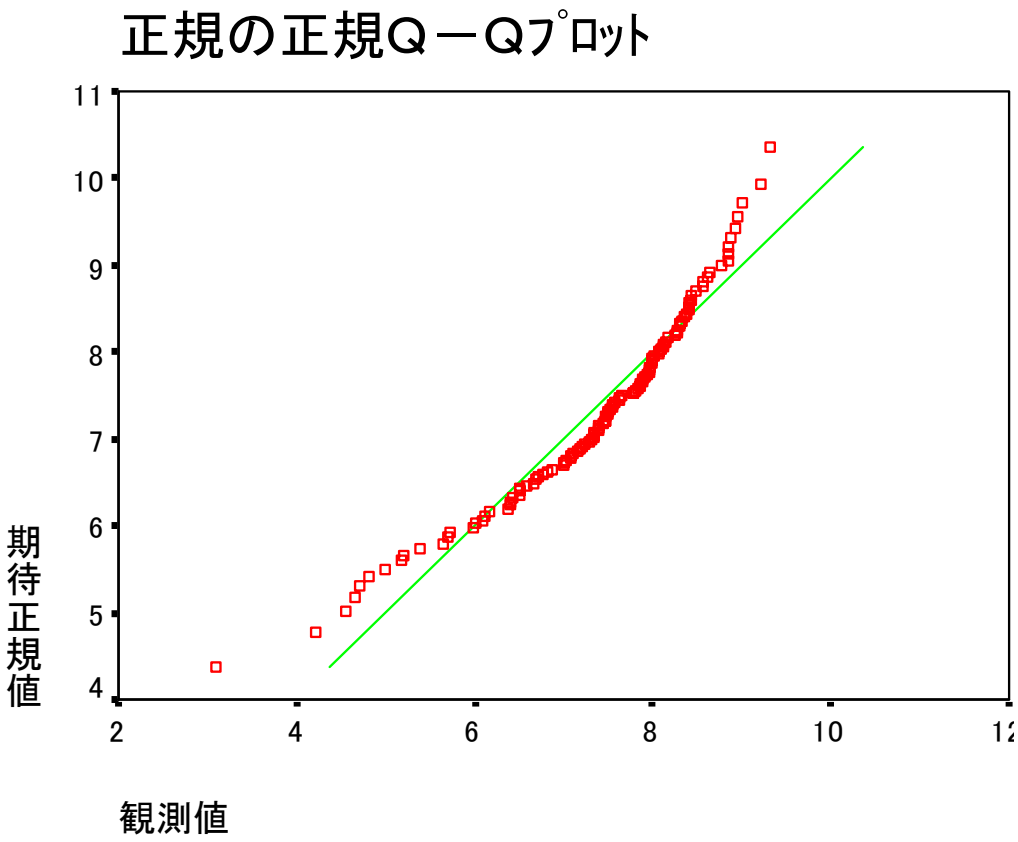
正規性の検証方法

- グラフ
  - ヒストグラムを書いてみる
- 正規Q−Qプロット
- 正規性検定（あまり使われない）
  - Skewness-Kurosis テスト（n＞2000のとき）
  - Shapiro-Wilks テスト（7≦n ≦ 2000のとき）
  - Kolmogorov-Smirnov検定

帰無仮説は、「変数の分布が正規分布と違わない」

有意確率＜0.05の場合には、正規分布と異なる

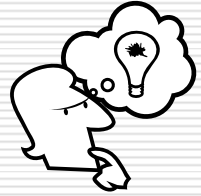
正規Q–Qプロット



正規 distribution
parameters estimated:
location=7.3602125
scale=1.1759897

- 確率プロット

  データの累積度数分布が、母集団の理論累積度数分布に一致しているかグラフで確かめる方法
- 母集団が正規分布のとき「正規プロット」と言う
- 正規直線上に点が並べば正規分布に一致

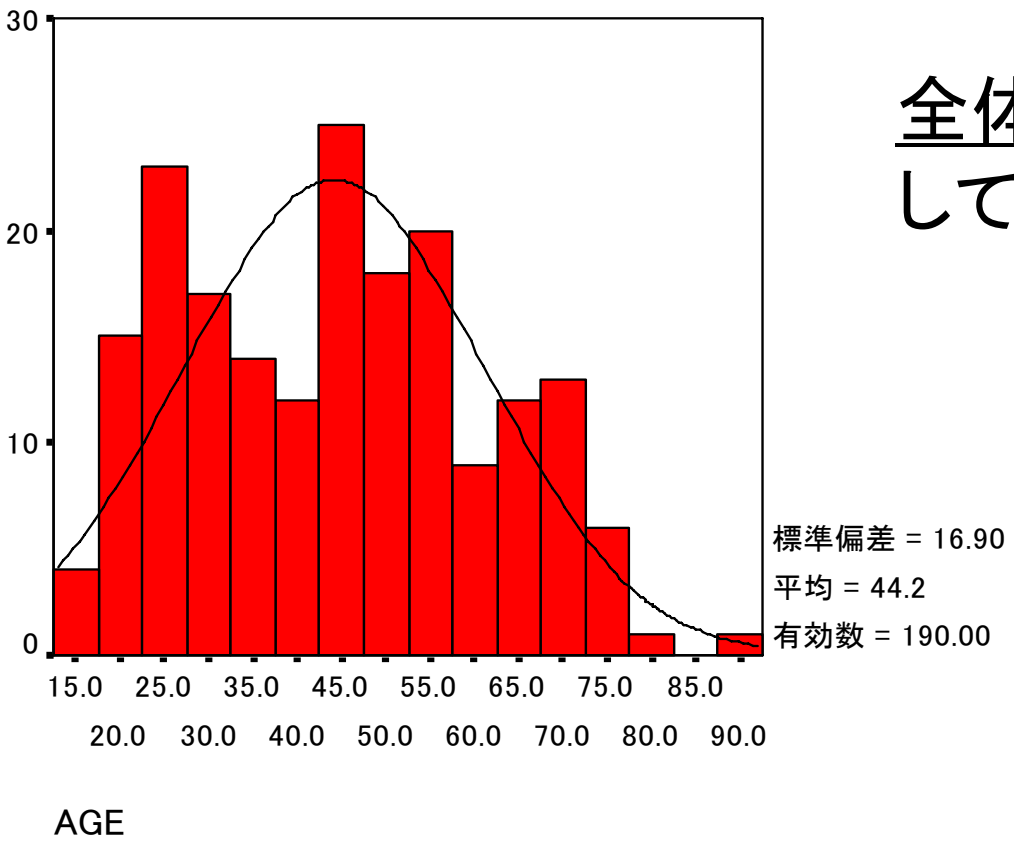


ピットフォール：何でも平均値を出す（2）

- テーマ「高齢者の○○の臨床的特徴」
- 50歳以上の「高齢群」と50歳未満の「非高齢群」に分けて比較する
- 「患者背景」で、各群の年齢を記述する
  - 高齢群（73）　　：61.7±　8.6歳
  - 非高齢群（ 117 ）　：33.2 ±10.2歳

これでいい？

ヒストグラムを書いてみる



全体としては正規分布しているようだ

だから正しい？

カテゴリーに分類した場合

- 全体としては正規分布していても各カテゴリー内では正規分布しない
  - カテゴリーごとにその測定値の平均値を算出してはいけない
- 「分類の指標」と「評価の指標」は分けるべき
  - カテゴリーごとの平均値は求めず、範囲で示す
    - 高齢群（73）　　：50～80歳
    - 非高齢群（ 117 ）：16～49歳

２つの判断の誤り（“あわて者”と”ぼんやり者”）

「ない傾向をあるとする誤り」＝αエラー（P値、有意水準）
「ある傾向をないとする誤り」＝βエラー

研究者の結論 ＼ 真の世界	差なし	差あり
差なし（帰無仮説を受理）	正しい判断	誤った判断 βエラー （第二種の過誤）
差あり（帰無仮説を棄却）	誤った判断 αエラー （第一種の過誤）	正しい判断 検出力（1-β）

有意差検定

論理学の背理法

- 前提
  - 地球は丸くない
- 論理展開
  - 水・地平線の先まで見えるはず
- 矛盾（反例）
  - 実際には見えない
- 結論：前提は誤り
  - 地球は丸い

統計学の検定

- 帰無仮説（H0）
  - 群間差なし
- データに適用
  - p値の計算
- 起こりにくい結果
  - p＜0.05・・・有意水準
- 結論：帰無仮説を棄却
  - 群間差あり

２つの仮説

- 帰無仮説：H0（証明したくない方の仮説）

  否定することで証明したい仮説を証明

  - A群とB群でがん発生率に差がない

- 対立仮説：H1（研究者が証明したい仮説）

  - がん発生率に群間差あり

P値？

- 群間差が0のとき、p値＝1.000
- 群間差が大きくなるとp値は小さくなる
- 例数が大きくなると小さくなる
- 群間差は同じでも例数が増えればp値 ↓

$$\text{P値} = \text{関数}\left[\frac{\text{群間差}}{\text{標準誤差（S.E.）}}\right]$$

$$S.E. = \frac{S.D.}{\sqrt{n}}$$

「有意差なし」は「群間差なし」？

- 有意水準
  - 統計的有意かどうかを判断する基準
  - 通常は、0.05（あくまでも慣例的な数値）

- 帰無仮説を棄却できないとき（p＞0.05）
  - 「差が等しい」ことを保証するものではない
    （「 群間差なし」でもない）
  - →「群間差ありとは言えなかった」
  - →「有意な差を認めなかった」
    （判定は保留・・・）

症例数と有意差

- 各群の平均値：100と98、標準偏差100
  - 1群　10例　　p=0.660
  - 1群　50例　　p=0.320
  - 1群 100例　　p=0.159
  - 1群 500例　　p=0.002
  - 1群1000例　　p＜0.001

単に、例数が少なくて有意差を示せなかっただけ

★RCTのP値は十分に小さくはならない
∵ RCTでは先行研究をもとに有意に出る最小限の症例に設定

平均値の差の比較方法

- モデルに基づく方法
  - t-検定
  - 3群以上は分散分析：“どこかに差がある”
- 順位に基づく方法
  - Wilcoxon順位和検定（ Mann-Whitney U検定）
  - この2つは同じ検定：p値も同値

中村好一 やさしい統計学（2006）、石村貞夫 すぐわかる統計用語（1997）

Studentの t 分布

THE PROBABLE ERROR OF A MEAN

BY STUDENT

「平均値の誤差の確率分布」

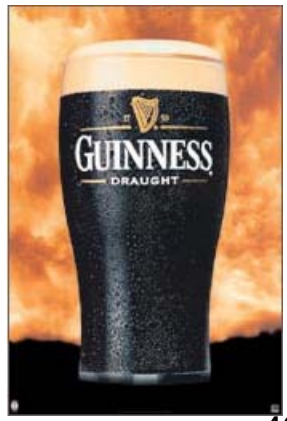
Introduction

Any experiment may he regarded as forming an individual of a "population" of experiments which might he performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a greater number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

If the number of experiments be very large, we may have precise information as to the value of the mean, but if our sample be small, we have two sources of uncertainty: (1) owing to the "error of random sampling" the mean of our series of experiments deviates more or less widely from the mean of the population, and (2) the sample is not sufficiently large to determine what is the law of distribution of individuals. It is usual, however, to assume a normal distribution, because, in a very large number of cases, this gives an approximation so close that a small sample will give no real information as to the manner in which the population deviates from normality: since some law of distribution must he assumed it is better to work with a curve whose area and ordinates are tabled, and whose properties are well known. This assumption is accordingly made in the present paper, so that its conclusions are not strictly applicable to populations known not to be normally distributed; yet it appears probable that the deviation from normality must be very extreme to load to serious error. We are concerned here solely with the first of these two sources of uncertainty.

William Sealy Gosset
(1876～1937)



Biometrica. 1908;6:1-25

t 検定
モデルに基づく方法

- 前提
  - サンプルサイズがそれぞれn, mの2つのグループのデータは互いに独立に正規分布に従う

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{(1/n + 1/m)\ \hat{\sigma}^2}}$$

ただし、σ^2はバラツキの指標（分散）

⇒ t 分布に従う。そのt分布を使ってp値を求める

Student Biometrica. The probable error of a mean. 1908;6:1-25

Wilcoxon 検定

- 順位に置き換え
  - 治療A

    25.5　29　35.0　22　14 - - - - - - - - -
  - 治療B

    18.5　8　25.5　3　30 - - - - - - - - -

$$Z = \frac{\text{平均順位（A：15.8）ー平均順位（B：20.4）}}{\text{バラツキ（平均順位の差）}}$$

⇒ Z が正規分布に従うことからp値（=0.19）を求める

カイ二乗検定

	低下　大	低下　小	計	低下割合
A群	13	5	18	13/18 （72.2%）
B群	7	10	17	7/17 （41.2%）
	20	15		

$$\chi^2 = \frac{\text{低下割合（A）ー低下割合（B）}}{\text{バラツキ（低下割合の差）}}$$

⇒ χ^2 がカイ二乗分布に従うことからp値（=0.064）を求める

- なるべくFisher直接確率計算法を選択する
- 特に、標本数が少なく、期待値が5以下の場合

パラメトリック検定とノンパラメトリック検定

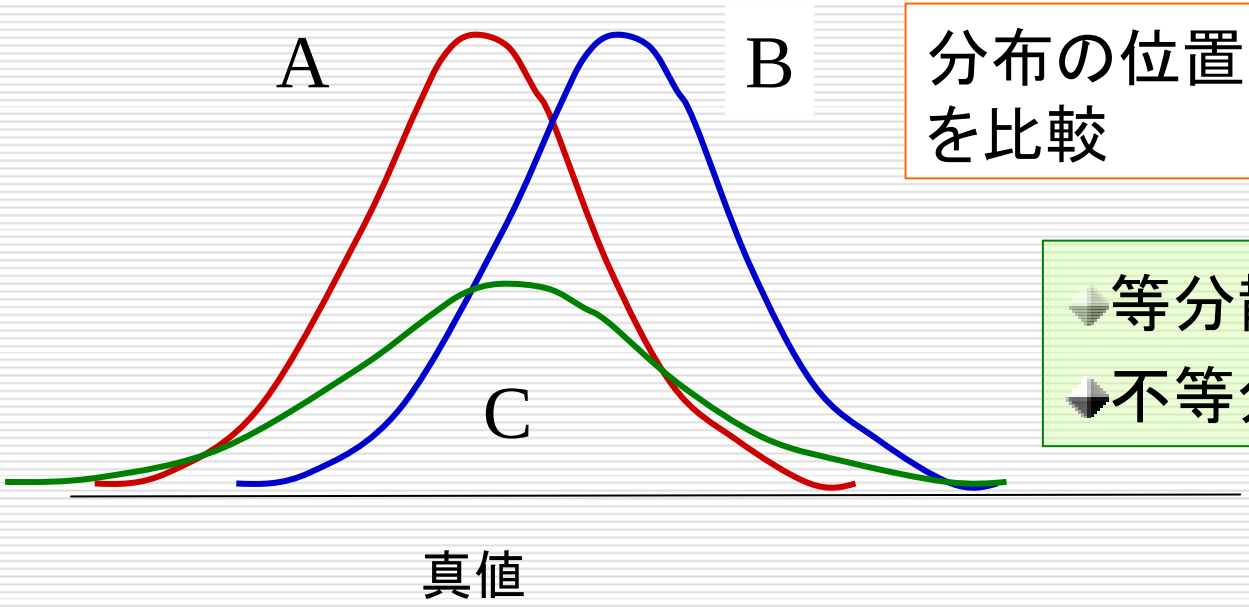
- パラメトリック検定
  - 母集団の分布に関する前提に基づく
  - t-検定：母集団は正規分布であることを前提

- ノンパラメトリック検定
  - 分布によらない検定（distribution-free test）
  - 正規母集団という前提はないため、どのような状況でも安心して使える
  - データを順位に置き換えて検定する
  - Mann-Whitney U検定、Wilcoxon順位和検定

中村好一 やさしい統計学（2006）、石村貞夫 すぐわかる統計用語（1997）

分布に関する前提

- 平均値の差を見ることが意味のある場面
  分布が左右対称（ほぼ正規分布）
  （平均は異なるが、バラツキ度合いは同じ）



分布の位置の違いを比較

- 等分散の検定：F検定
- 不等分散⇒Welchの方法

2標本の検定

- t-検定の条件（3つの条件を満たすのは厳しい・・・）
  - データが間隔尺度
  - 両群とも正規分布（逆数や対数変換をする方法もあり）
  - 両群とも分散が等しい（ばらつき具合が同じ）
    （・・・厳密に考える必要はないとの意見もある）

- ノンパラメトリック検定を行うべき場合とは？
  - 間隔尺度でないとき（スコアなどの順序尺度のとき）
  - 明らかに正規分布でないとき
  - データの分散が群によって一様でないとき
  - 分布の端で測定値が途切れているとき。測定感度以下のデータがあるとき

パラメトリックかノンパラメトリックか？

- 原則にこだわって“事前に宣言して”使い分ける
- 疑わしい場合は、ノンパラメトリック検定が無難
- 小標本の場合
  - 極端な場合を除き、事実上t検定は正当化される
  - 母集団がどういう分布か？手元のデータ以外のデータも参照
- 大標本の場合
  - 厳密にはt検定（3つの条件）を適用できる場合は殆どない・・・
  - パラメトリックでもノンパラメトリックでも結果に大きな差はない
  - 厳密に検定の妥当性をチェックする必要はないという意見も

パラメトリック検定とノンパラメトリック検

	パラメトリック検定	ノンパラメトリック検定
独立2標本の比較	対応のない t 検定	Mann-Whitney U検定 （Wilcoxon順位和検定）
関連2標本の比較	対応のある t 検定	Wilcoxon符号付順位検定
独立多標本の比較	一元配置分散分析	Kruskal-Wallis検定
関連多標本の比較	反復測定分散分析	Friedman検定
2変量の相関	Pearson相関係数	Spearman順位和相関係数

中村好一 やさしい統計学入門（2009）より）改変引用

検定よりも信頼区間！

- 仮説検定
  - 仮説が観察されたデータと矛盾しないかどうかを調べる
  - 効果の大きさは不明、二者択一の情報のみ

- 信頼区間
  - データと矛盾しない真の差（δ）の範囲
  - 差の程度（効果の大きさ）とその精度

95%信頼区間

- 両側5％水準の検定で有意にならないδ（真の値）を集めてくる
- 信頼区間の「95％」を信頼係数と呼ぶ（1―αレベル）

95％信頼区間とはなにか？※誤解が多い

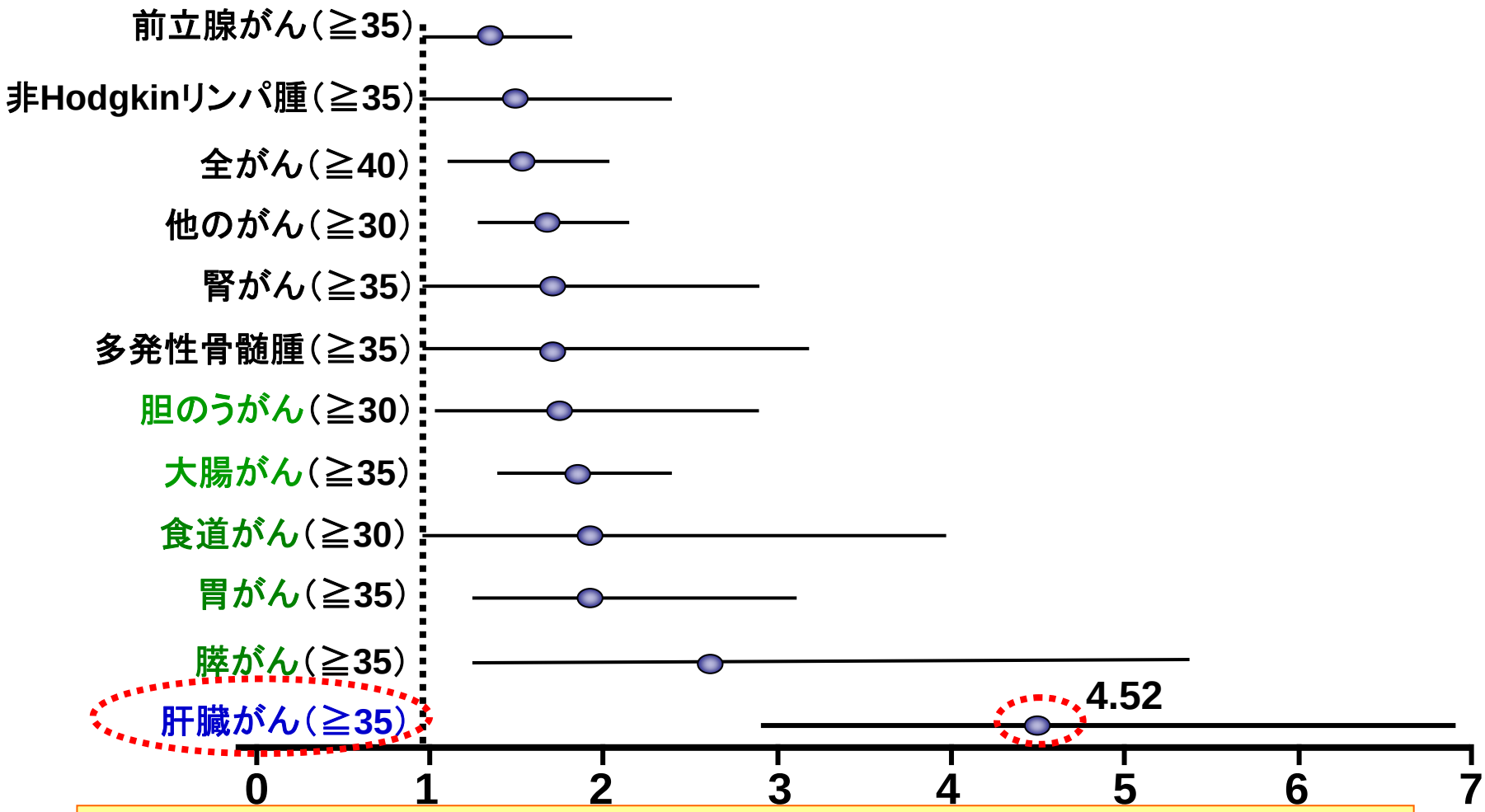
- 同じ規模の研究を繰り返して実施
- 信頼区間を求める
- 何回も繰り返して、そのうち平均して95％は真の値を含むような区間

✓「真の値を95％の確率で含んでいる区間」ではない

95％信頼区間＝推定値±1.96×標準誤差

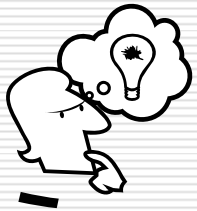
自由度（標本数－1）で決まるt分布の0.975の

肥満による発がんリスクの上昇（男性）



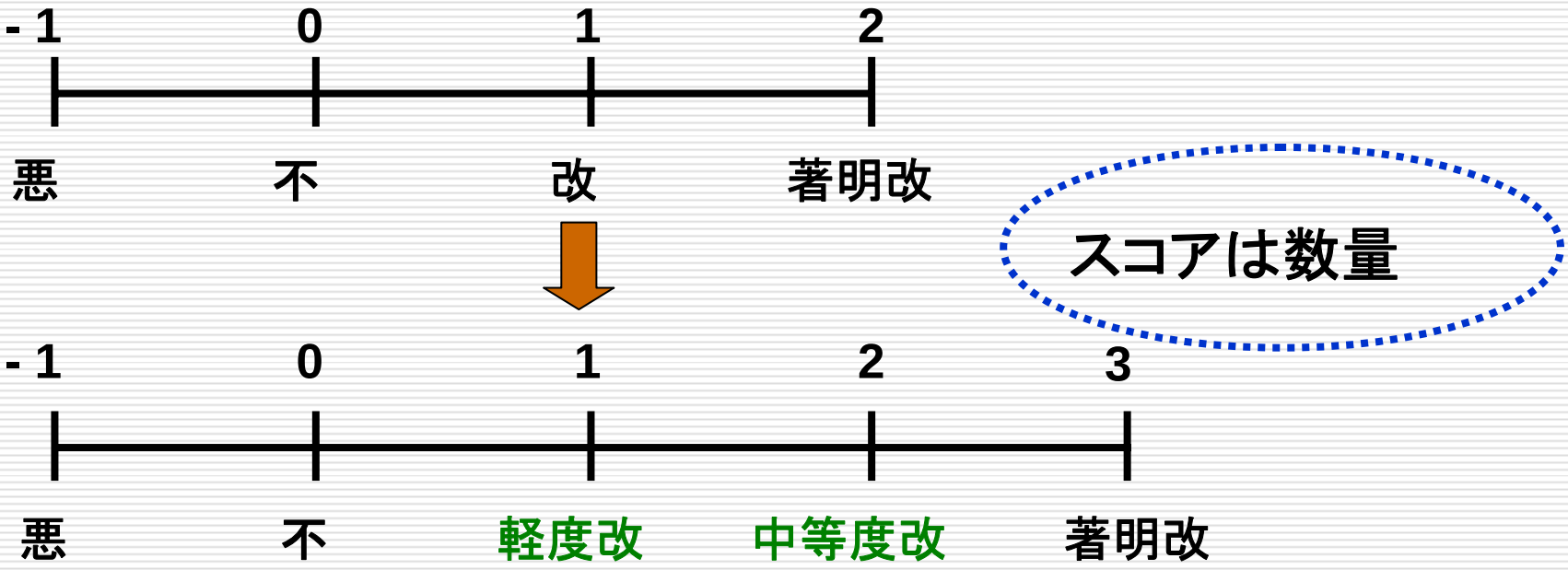
肥満男性は、肝発がんリスクが約4.5倍

Calle EE et al. N Engl J Med 2003;348:1625



ピットフォール：スコアを数量として扱う

- テーマ「○○に対する治療の効果」
- 効果をスコアにして平均値を比較した・・・



スコア化する場合の注意点

- スコアは「順序尺度」であり、間隔尺度ではない
  - 例：Glasgow coma scale, Apgar scoreなど
- 平均や標準偏差は計算できず、カテゴリー別の頻度（人数）で表現する

	結果のカテゴリー				
	-1	0	1	2	合計人数
プラセボ	6	3	1	0	10
A薬	1	0	0	9	10

順序カテゴリーデータの解析

- 頻度の差についてノンパラメトリック（非数量的）検定を行う
  - Wilcoxon-Mann-Whitney検定
  - ２×２表にしてカイ2乗検定することは間違い
    （∵順序情報を使っていない。感度が低い。）

- 数量としての裏付けがあれば、算術的に扱ってもよい
  - スコアと臨床的重要性に量ー反応関係が明らかになった場合など

- 代表値として平均を出す場合は、t検定で差を確認可能
  - 平均値自体は当然変わるが、「t値」は殆ど変わらない

Lincoln EM, et al. Medical Uses of Statistics. 1986

統計学～臨床研究者としての学習目標

- 解析方法の意味を理解する
- 解析結果を良く吟味する
- 結果の解釈を十分に行う
- 解析結果から言えること、言えないことをきっちり分けて文章にする
- （人に理解できるように平易に説明できる）

参考文献・図書

- 高橋 敬. 初学者のための統計解析. 岩手医誌 2001;54:61-5
- 津谷喜一郎・折笠秀樹 監訳. 医学統計学の活用: サイエンティスト社; 1995
- 浜田知久馬. 学会・論文発表のための統計学：統計パッケージを誤用しないために. 真興交易医書出版部; 1999
- 臨床試験のための統計的原則（ICH-E9）
  http://www.nihs.go.jp/dig/ich/ichindex.htm
- Harvey M. *Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking 2nd ed*. Oxford University Press, 2010

（津崎晃一訳：**数学いらずの医科統計学 第2版** メディカル・サイエンス・インターナショナル; 2011 ）

配布資

謝辞

- 京都大学　大学院医学研究科　社会健康医学専攻
  - 健康情報学：中山 健夫
  - 予防医療学：川村 孝、安藤昌彦
  - 医療疫学：福原 俊一、森田 智視、東 尚弘
  - 医療統計学：佐藤 俊哉、大森 崇
  - 薬剤疫学：松井 茂之

- 同　臨床研究者養成者コース（MCR）1期生
  - 石見拓
  - 片多史明
  - 川口武彦
  - 北村和也
  - 西内辰也
  - 能城　毅
  - 杉岡隆
  - 白井貴子
  - 松澤重行
  - 松田二三子