

医学統計の応用のこつ

国立保健医療科学院 技術評価部
横山 徹爾

研究計画に関する基礎知識

- データの整理
- 検定とは？
 - 数量データの差の検定
 - 質的データの差の検定
 - 検定と区間推定
- 関連の分析
 - 2つの数量データの関連
- 多変量解析

研究目的を明確にしよう。

復習

研究計画時に行うこと

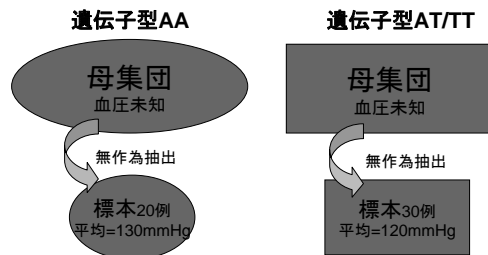
- 目的・仮説を明確に
- 最終的なまとめの図表を頭に描いておく
- 統計解析手法も決めておく
 - とはいえ、統計学の基本が分かっていると、これらを考えることができない。
- 本日の学習目標：
 - 目的(この研究で言いたいこと)に応じた基本的な統計手法を学ぶ。
 - 特に、研究デザイン、データの種類別にふさわしい手法を理解する

復習

統計学の原点：標本調査

- 調査対象(母集団)を明確にする
- そこから無作為抽出された標本を用いて推測を行う

50人の某遺伝子型を調べた。AA型の20名はAT/TT型の30名よりも収縮期血圧の平均値が10mmHg高かった。だから・・・



基本

標本数を決める

- 先行研究等を参考に、必要な標本数をあらかじめ計算する。
- 標本数の決め方にはおおむね2通りある
 - 有意差(有意な関連)を検出するために必要な標本数
 - 臨床研究ではこちらが多い
 - 例) A薬投与群では血圧が平均 4 ± 6 mmHg、B薬投与群では平均 10 ± 6 mmHg低下することが予想される。検出力80%で有意差を検出するためには、標本数は何例ずつ必要か。
 - 詳細は後述(検定を勉強してから)
 - ある精度で推定を行うために必要な標本数
 - 統計調査などではこちらが多い
 - 例) X県のメタボリックシンドローム有病率を知りたい。おおむね15%と予想される。±3%の誤差で推定するには何人調査すべきか。
 - 本日は省略。

基本

実験計画法

例) 薬剤Aの用量別効果を比較するために、5mg, 10mg, 20mgをマウス3匹ずつ計9匹に投与。

どういう順番で実験を行うか？

順番に行う??	1日目			2日目			3日目		
	朝	昼	夕	朝	昼	夕	朝	昼	夕
	5mg	5mg	5mg	10mg	10mg	10mg	20mg	20mg	20mg

- 測定値に入りうる様々な誤差
 - 個体差、籠差、温度差、慣れ、日内・日間変動
 - これら誤差が可能な限り小さくなるように計画する
- 制御できる要因は均一に
 - 温度、湿度等の実験条件は均一にする
- 制御できない要因は無作為化
 - 個体差、慣れの効果、その他未知の要因
 - 制御不可能だが、無作為化によってうち消すことができる

基本 実験計画法・基本的な考え方

例) 薬剤Aの用量別効果を比較するために、5mg, 10mg, 20mgをマウス3匹ずつ計9匹に投与。

どういふ順番で実験を行うか？

	1日目			2日目			3日目		
	朝	昼	夕	朝	昼	夕	朝	昼	夕
順番に行う	5mg	5mg	5mg	10mg	10mg	10mg	20mg	20mg	20mg
日間差の影響が入る									
完全無作為化法	10mg	20mg	10mg	5mg	5mg	20mg	5mg	10mg	20mg
少し改善したが、まだ日間差の影響が少し残る									
完備乱塊法	10mg	20mg	5mg	5mg	20mg	10mg	5mg	10mg	20mg
日間差の影響が打ち消された。日内差(朝昼夕)の影響が残っている									
ラテン方格法	5mg	10mg	20mg	10mg	20mg	5mg	20mg	5mg	10mg
日間差、日内差(朝昼夕)の影響が打ち消された									

用いる統計手法も少しずつ違う

■ 研究計画に関する基礎知識

➡ ■ データの整理

■ 検定とは？

- 数量データの差の検定

- 質的データの差の検定

- 検定と区間推定

■ 関連の分析

- 2つの数量データの関連

■ 多変量解析

いきなり複雑なことをしない。まずは簡単にデータを整理する。

復習

医学データの種類

- 計量データ: 量的に測定できる連続的な測定値
 - 連続データ (例) 身長、体重、血圧、血清総コレステロール
 - 離散データ (例) う歯の本数
- 計数データ: カテゴリー型のもの
 - 2値 (例) 性別の“男”と“女”、既往歴の“有り”と“なし”
 - カテゴリーが3つ以上
 - 順序尺度ordinal scale: 順序関係はあるが絶対量としての意味はない測定値。
 - (例) 胃癌の深達度: T1, T2, T3, T4
 - 名義尺度nominal scale: 順序関係がない分類のための変数。
 - (例) 病理分類の“腺腫”、“線維腫”、“血管腫”, etc.

基本 ポイント: 一見同じ質的データに見えても、順序尺度で量反応関係に注目する場合は、用いる統計手法が違う

復習

データを整理する

■ いきなり平均・標準偏差を計算しない!

- まず、ヒストグラム等を描いて分布を視覚的に確認

- その後、適切な要約統計量を決めて分布の特徴を表現する

■ いきなり検定しない!

- まず、図や要約統計量で比較して特徴を確認

- その後、適切な方法で検定

復習

分布型を確認

- 統計学的方法のうち、よく使うパラメトリックな方法(t検定など)では、左右対称な分布(正規分布)を前提としているものが多い。
- 従って、可能ならば、何らかの変換によって正規分布に近似させてから処理すべきである。
 - 対数変換、平方根変換、Box-Cox(べき)変換など
- 正規分布に近似できない場合、ノンパラメトリックな方法を考慮(後述)。

図2 正規分布

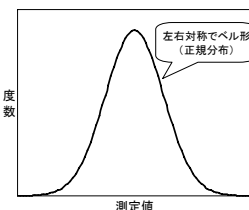
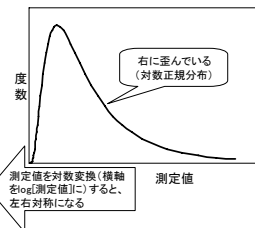
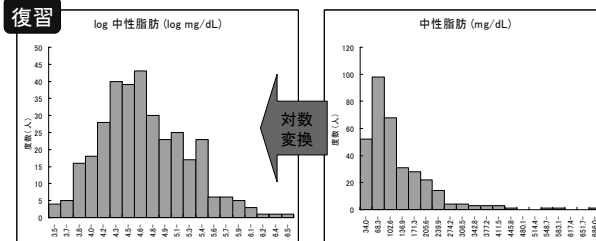


図3 対数正規分布



測定値を対数変換(横軸をlog測定値)にすると、左右対称になる

復習



■ 対数正規分布の典型例

- 細菌数、中性脂肪、AST、ALT、 γ -GTPなど

■ 正規分布の典型例

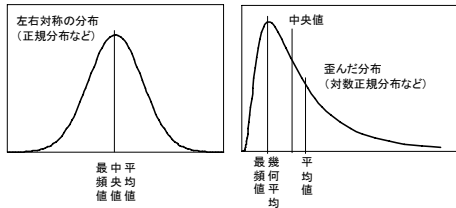
- 身長、体重など

■ 医学データは、少し右裾が長いことが多い

復習

代表値(中心位置の指標)

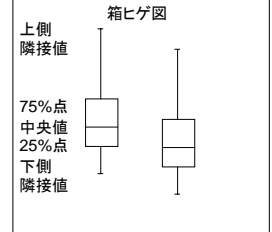
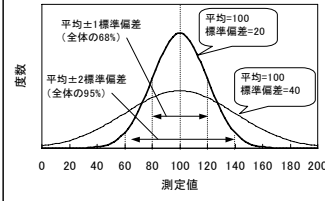
図4 分布型と代表値



- 平均値・・・左右対称な場合に有用
- 中央値・・・非対称等、歪んだ分布の場合

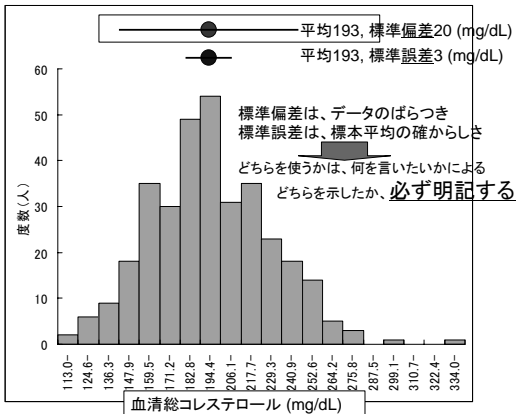
復習

図5 標準偏差はバラツキの指標



- 代表値(中心位置の指標)と散布度(バラツキの指標)として、
 - 「平均と標準偏差」
 - 「中央値と四分偏差」
 の組合せがよく用いられる。

復習 標準偏差と標準誤差を混同しない



- 研究計画に関する基礎知識
- データの整理
- ➔ 検定とは？
 - 数量データの差の検定
 - 質的データの差の検定
 - 検定と区間推定
- 関連の分析
 - 2つの数量データの関連
- 多変量解析

いろいろな検定方法があるが、基本原理はただ一つなのでそれを理解しよう。

復習

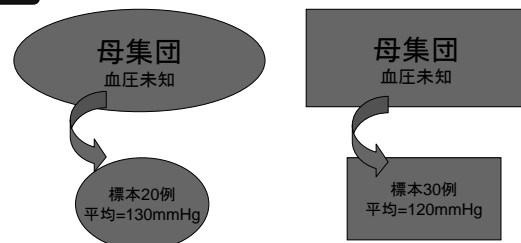
検定

- 検定とは
 - 観測された差(や関連)が偶然によるものか否かを判断する方法
- 検定の論法
 - 「真実(母集団)は差(や関連)がない」と仮定する(=帰無仮説 H_0)
 - 帰無仮説が正しい場合に、標本において観測された差(や関連)が生じる確率(P値)を計算する
 - その確率が十分に小さければ(例えば $P < 0.05$)、帰無仮説が正しい場合に偶然では起こりにくいことが起きたということなので、帰無仮説を棄却して「真実は差(や関連)がある」(=対立仮説 H_1)と判断する。(一般に、「有意差がある」という)

復習

遺伝子型AA

遺伝子型BB



帰無仮説(AAとBBで母集団の血圧の平均は同じ)が正しい場合に標本平均に10mmHgの差が生じる確率は？

- t検定で1%($P=0.01$)と計算された。
- 帰無仮説が正しければめったに生じない現象がおきたといえる。従って、たぶん帰無仮説は正しくないのだろう。
- 対立仮説(AAとBBで母集団の血圧の平均は異なる)を採用。

復習 検定における2種類の判断ミス

- 検定は万能ではなく、しばしば誤った判断に陥ることがある。

		判断(検定結果)	
		差がある	差がない(あるとはいえない)
真実	差がある	○	第2種の過誤(βエラー)
	差がない	第1種の過誤(αエラー)	○

第2種の過誤が生じない確率のことを検出力(パワー)という

P値は、第1種の過誤が生じる確率。判断の基準とする確率を有意水準という。

一般に、標本数が小さいほど検出力も小さい=第2種の過誤が生じやすい
→例数設計の必要性

復習 “有意差なし”は“差がない”ことを積極的に示したわけではない!

■ 例1

- 降圧薬AとBを5匹ずつのマウスに投与した。A薬とB薬で血圧の低下幅の平均値の差は10mmHgで、有意差はなかった。
- 降圧薬AとBを20匹ずつのマウスに投与した。A薬とB薬で血圧の低下幅の平均値の差は10mmHgで、有意差があった。

■ 「差がない」ことを証明するためには、ケチって小標本にすればいい??? (そんな馬鹿な!)

■ 応用 - 「同等性の検定(後述)」を行う必要がある。あるいは、検出力を計算すると参考になる。

基本

有意差検出のための 再び、標本数の決め方

- 必要な情報
 - 想定される差の大きさD (どのくらい大きな差や強い関連が想定されるのか?)
 - 先行研究等の情報を勘案して決める。
 - データのパラッキ (=分散:連続量の場合)や保有率(割合の差、オッズ比の場合)も必要。
 - 有意水準 α と検出力 $1-\beta$ (どのくらい確実に有意差を検出したいのか?)
 - $\alpha=0.05$ 、 $1-\beta=0.8$ とすることが多い(目的等に応じて決める)
 - その意味するところ
 - 母集団で真に差Dがあるならば、80%の確率で有意差 ($P<0.05$) が得られるように標本数を決めよう。
- 例
 - 先行研究等より、A薬投与群では血圧が平均4mmHg、B薬投与群では平均10mmHg低下することが予想される ($D=6\text{mmHg}$)。また、いずれの群も低下幅の標準偏差は5mmHg程度と予想される。有意水準 $\alpha=0.05$ 、検出力 $1-\beta=0.80$ とすると、標本数は何例ずつ必要か。
- 計算方法
 - 研究デザイン、データの種類(連続量か、割合か、リスク比か、etc.)によって違うが、おおむね上記の情報が必要。成書参照。
 - 一般に、差Dが大きいほど、データのバラツキ(測定誤差を含む)が小さいほど、必要な標本数は少ない。

■ 研究計画に関する基礎知識

■ データの整理

■ 検定とは?

- 数量データの差の検定
- 質的データの差の検定
- 検定と区間推定

■ 関連の分析

- 2つの数量データの関連

■ 多変量解析

数量データの比較では、平均値などの代表値を複数の群間で比較することに興味がある

復習 パラメトリックとノンパラメトリックな方法

■ パラメトリックな検定

- 母集団の分布に特定の分布型(例えば正規分布)を仮定した検定方法
- 母集団の分布が正規分布か否かの判断
 - 経験による
 - 標本分布による
 - ヒストグラム、正規確率紙等で視覚的に判断
 - 尖度、歪度、正規性の検定は参考程度に
- 著しく正規分布ではない分布型なのに無理にパラメトリックな検定を用いると、その検定結果は信頼できない!

■ ノンパラメトリックな検定

- 母集団の分布に特定の分布型を仮定しない検定方法

復習 独立な2群の差の検定

帰無仮説 $H_0: \mu_1 = \mu_2$
対立仮説 $H_1: \mu_1 \neq \mu_2$

■ パラメトリックな検定

- Student t検定

- 正規分布、等分散

- Welch t検定

- 正規分布、不等分散

等分散の判断は正規確率紙による視覚的検討、およびF検定t検定の4倍程度の有意水準を使うことが多い(F検定: $P<0.20$ で不等分散)

■ ノンパラメトリックな検定

- Mann-Whitney U検定

- 非正規分布、(等分散)

対数変換等を行っても正規分布に近づかない場合

■ 小標本では母分布型の判断が難しいが、どうする?

- ノンパラメトリックな検定を行うと、検出力が著しく落ちることがある
- パラメトリックな検定を行うと、検定結果が信頼できない恐れ
 - 十分な標本サイズになるように事前に計画しましょう

基本

「独立な」2群とは？

40代男性における飲酒と血圧の関係

	飲酒者(n=100)		非飲酒者(n=100)		P値
	平均	SD	平均	SD	
収縮期血圧	130	20	125	18	?

- 飲酒者も非飲酒者も無作為抽出標本
 - 「飲酒者でたまたま高血圧者が多めに選ばれた」
 - 「すると、非飲酒者でもたまたま高血圧者が多めに選ばれやすい、なんてことはない！」
- 一方の群がどのように選ばれようとも、他方の群には全く影響しない→「独立な」2群 ... 対応のないデータ

「独立でない」2群とは？

40代男性における右腕と左腕の血圧測定値の違い

	右腕(n=100)		左腕(n=100)		P値
	平均	SD	平均	SD	
収縮期血圧	130	20	125	18	?

- 同じ人の右腕と左腕なので、似た値をとりやすい。
- 右腕が高ければ、左腕も高い。一方の値が他方の値に関係している→「独立でない」2群 ... 対応のあるデータ

復習 例題1: どのような検定を行うか

40歳代男性における飲酒習慣と血圧等との関連

	飲酒者(n=100)		非飲酒者(n=80)		P値
	平均	SD	平均	SD	
収縮期血圧 (mmHg)	130	20	125	18	?
中性脂肪 (mg/dL)	140	110	100	80	?
γ-GTP (IU/L)	70	80	30	40	?

- 収縮期血圧
 - ほぼ正規分布、等分散→Student t検定
- 中性脂肪
 - 対数変換したところ正規分布、等分散→Student t検定
- γ-GTP
 - 変換しても強く歪んだ分布→Mann-Whitney U検定

復習 例題2: どのような検定を行うか

40歳代男性における飲酒習慣と血圧等との関連

	飲酒者(n=100)		非飲酒者(n=80)		やめた(n=20)	
	平均	SD	平均	SD	平均	SD
収縮期血圧 (mmHg)	130	20	125	18	135	22

- t検定(またはU検定)を全ての組合せについて繰り返す??? (ダメ!)
- 1回の検定につき、第1種の過誤が5%の確率で生じる。
- 3回検定を繰り返すと、 $1-(1-0.05)^3=14\%$ の確率で第1種の過誤が生じる。
- 有意水準5%といいながら、実は14%の確率で3つの検定のうち1つ以上で第1種の過誤を生じる!
- 検定の多重性の問題

復習 独立な3群の差の検定 帰無仮説 $H_0: \mu_1 = \mu_2 = \mu_3$
対立仮説 $H_1: H_0$ ではない

■ パラメトリックな検定

- 一元配置分散分析
 - 正規分布、等分散

等分散の判断は正規確率紙による視覚的検討、およびBartlett検定など
歪んだ分布は対数変換を行うと、正規分布に近づくだけでなく、等分散に近づくこともある

■ ノンパラメトリックな検定

- Kruskal-Wallis検定
 - 非正規分布、不等分散

対数変換等を行っても正規分布に近づかない、不等分散の場合

- 3群のどこかに差があるということが示される。
- どこどこの組合せに差があるかは、まだ分からない!

復習 例題3: どのような検定を行うか

40歳代男性におけるALDH2遺伝子型と血圧との関連

	ALDH2遺伝子型						P値
	1/1 (n=240)		1/2 (n=140)		2/2 (n=20)		
	平均	SD	平均	SD	平均	SD	
収縮期血圧 (mmHg)	130	20	125	18	110	17	?
中性脂肪 (mg/dL)	140	110	110	80	100	77	?
γ-GTP (IU/L)	70	80	40	40	35	30	?

- 収縮期血圧
 - ほぼ正規分布、等分散→一元配置分散分析
- 中性脂肪
 - 対数変換したところ正規分布、等分散→一元配置分散分析
- γ-GTP
 - 変換しても強く歪んだ分布→Kruskal-Wallis検定

復習 独立な3群以上の差の検定と対比較

■ 分散分析

- 帰無仮説 $H_0: \mu_1 = \mu_2 = \mu_3$
- 対立仮説 $H_1: \mu_1 = \mu_2 = \mu_3$ ではない
 - つまり、どれとどれに差があるかは分からない。そこで...

■ 対比較

- どこどこの組合せに差があるか、興味のある組合せに対して検定を繰り返す(対比較)。ただし、検定を繰り返しても第1種の過誤の生じる確率が5%を超えないように工夫する。
 - 全ての組合せに興味がある場合: Tukey法 (Tukey-Kramer法)
 - 一つの対照群と残りの群を比較する場合: Dunnett法
 - 後で一部のカテゴリーを併合していろいろな比較を行う(線型比較): Scheffe法
 - 特定の組合せに興味があり、検定回数を事前に決めておく場合: Bonferroni法 (Holm法)
 - 3群の場合のみ、分散分析で有意になったという前提で3回のt検定を繰り返す方法: Fisher LSD法

復習 例題2' : どのような検定を行うか

40歳代男性における飲酒習慣と血圧等との関連

	飲酒者 (n=100)		非飲酒者 (n=80)		やめた (n=20)	
	平均	SD	平均	SD	平均	SD
収縮期血圧 (mmHg)	130	20	125	18	135	22

- まず、一元配置分散分析を行う
- どこどこに差があるかを確認するために、
 - 3通り全て興味がある場合・・・Tukey-Kramer法
 - 非飲酒者との2群の比較に興味がある場合・・・Dunnett法
 - 上記2法では分散分析を省略することもある
 - 注意: 結果を見てから差のありそうな比較だけをしちゃダメ!

復習 例題5: どのような検定を行うか

試験薬A投与前後のネズミ20匹の体重(g)の変化

投与前		投与後		後-前		P値
平均	SD	平均	SD	平均	SD	
200	30	180	28	-20	10	?

- 対応のあるt検定(パラメトリック)
- Wilcoxon符号付き順位検定(ノンパラメトリック)
 - 理由
 - 投与前と投与後の体重が独立ではないから。
 - 対応のないt検定やU検定は、独立な2群の比較

応用

例題6: 独立でない(対応のある) 3群以上の差の検定

マウスガードのフィット感(VAS得点)

マウスガードA	マウスガードB	マウスガードC	P値
中央値 (25-75%点)	中央値 (25-75%点)	中央値 (25-75%点)	
7.5 (5.0-8.8)	6.2 (4.0-7.5)	8.0 (5.0-9.0)	?

方法: 30名の被験者に、各人3種類全てを順に装着してもらい、VASにより1~10点で評価。装着順序はA-B-C、A-C-B、B-C-A、B-A-C、C-A-B、C-B-A各群5名ずつ無作為割付。

- Friedmanの順位検定(ノンパラメトリック)
- 2元配置分散分析[個体×マウスガード](パラメトリック)
 - 理由
 - 同一被験者が3種類全てを評価しているから、A、B、Cの得点は独立ではないから。
 - 1元配置分散分析やKruskal-Wallis検定は独立な多群の比較
 - 対比較は、2群比較の繰り返しをBonferroni調整するなど。

応用

例題7: 同一人物から複数標本を得た場合の検定(個々のデータが独立でない)

50歳代男性における喫煙習慣と歯周ポケットの深さ(mm)

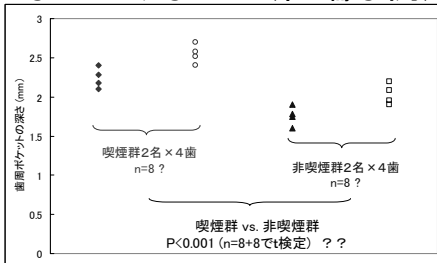
喫煙 (80名)			非喫煙 (60名)			P値
N	平均	SD	N	平均	SD	
320	2.5	0.4	240	1.8	0.3	?

一人につき4本ずつ調べたので、Nは人数の4倍。

- Student t検定? Mann-Whitney U検定?
 - N=320と240のまま検定しちゃダメ!
 - 理由: 同一人物の4本の測定値は独立でない(類似しやすい)から
 - 歯周ポケットの深い人が1人いると、いっぺんに4個の大きなデータが入る。
 - 対応のない検定は、独立な2群の、全て独立な標本の比較
- 個人毎に4本の平均値を計算して1人1つの値にしてからt検定
 - 特殊な方法として、GEEという方法もある。

応用

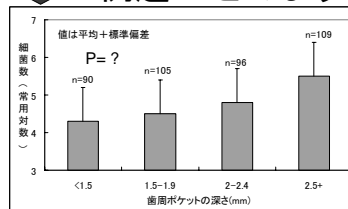
なぜいけないか(極端な例)



- 同一人物のデータはよく似ているとすると(この例は極端にしてある)、常識的に考えても、上の図はたまたま喫煙群の2人はポケットが深めの人たちだったと解釈すべきだろう。
- ところが、nが増えたので高度に有意??
- このデータは、母集団から全ての標本を独立に無作為抽出(t検定の前提)するのではなく、同じ人から似たような値を4回も取っている。
- 4本の平均を個人の値とするのが無難だろう。

応用

例題8: どのような検定を行うか



方法:
人数100名×4歯=計400歯
歯周ポケットの深さと細菌数との関係を調べた

- N=400として、1元配置分散分析? 回帰分析?
 - N=400のまま検定しちゃダメ!
 - 理由: 同一人物の4本の測定値は独立でない(類似しやすい)から
 - 歯周ポケットが深く細菌が多い人が1人いると、いっぺんに4個の似たデータが入る。
 - そのまま検定すると過度に(間違っって)有意になりやすい可能性あり
- GEE(一般化推定方程式)
 - 同一個体内での相関(類似性)を補正できる。
 - SASのProc GENMODなど。

■ 研究計画に関する基礎知識

■ データの整理

■ 検定とは？

- 数量データの差の検定
- 質的データの差の検定
- 検定と区間推定

■ 関連の分析

- 2つの数量データの関連

■ 多変量解析

質的データの比較では、割合の差を複数群間で比較することに興味がある

復習 例題9: どのような検定を行うか

		高血圧		計
		あり	なし	
果 物 頻 度 取	週2日以下	(55%) 22	18	40
	週3~5日	(40%) 8	12	20
	週6日以上	(25%) 10	30	40
計		40	60	100

■ χ^2 検定、Fisherの正確な検定

- 果物摂取頻度と高血圧有病率との関連の有無

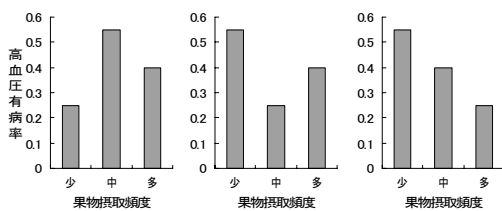
応用

■ 拡張Mantel検定、ロジスティック回帰のトレンド検定

- 果物摂取頻度と高血圧有病率との順序的な関連の有無

応用

①量-反応関係なし ②量-反応関係なし ③量-反応関係あり



	①の場合	②の場合	③の場合
独立性の χ^2 検定	$p=0.024$	$p=0.024$	$p=0.024$
傾向性の検定 (拡張Mantel検定)	$p=0.10$	$p=0.10$	$p=0.0064$

量-反応関係を積極的に示すためには、拡張Mantel検定の方がよい。
ロジスティック回帰を用いてもよい。

復習 例題10: どのような検定を行うか

100人の聴力低下所見の有無

		右耳	
		所見あり	なし
左 耳	所見あり	5	13
	なし	17	65

■ McNemar検定

- 理由

- 右耳と左耳のデータが独立ではないから。
 - 同一人物なので右耳と左耳の聴力が似ている傾向あり
- 独立性の χ^2 検定は、独立な2群の比較

応用

- 3カテゴリー以上の場合(所見+, ±, -など)には、
 - Bowkerの対称性検定 (Bowker's Test of Symmetry)

応用

例題11: どのような検定を行うか

50歳代男性における喫煙習慣と歯周ポケット $\geq 4\text{mm}$ の割合

喫煙 (80名)		非喫煙 (60名)		P値
N	%	N	%	
320	33	240	25	?

一人につき4本ずつ調べたので、Nは人数の4倍。

■ χ^2 検定？

- N=320と240のまま検定しちゃダメ！
- 理由: 先ほどと同じく、同一人物の4本の測定値は独立でないから
- χ^2 検定は、独立な複数群の、全て独立な標本の比較

■ 個人毎に $\geq 4\text{mm}$ の本数(0~4)をその人の1つの値にしてからMann-Whitney U検定など。

- 特殊な方法として、GEEという方法もある。

■ 研究計画

■ データの整理

■ 検定とは？

- 数量データの差の検定
- 質的データの差の検定
- 検定と区間推定

■ 関連の分析

- 2つの数量データの関連

検定よりも区間推定が望ましいことも多い

復習

検定と区間推定

この検定結果をどう解釈するか？

40歳代男性における飲酒習慣と血圧との関連

	飲酒者		非飲酒者		P値
	平均	SD	平均	SD	
収縮期血圧 (mmHg)	130	20	125	18	0.02

- 有意差あり
 - 母平均の差が5mmHgということを示したわけではない。「差なし」を否定しただけ。
- 区間推定
 - 母平均の差(95%信頼区間)=5.0(1.2-8.8)のように示した方が有用かも。
 - 95%信頼区間が0を含んでいなければ、5%水準で有意差ありを意味する。

基本

検定と区間推定

- 検定は、「差がある」ということを言えるが、どの程度の大きさの差があるのか？という問には答えられない。
- 区間推定は、どの程度の大きさの差があるのかを幅をもって示すことができる。検定と同じ意味合いも持つ。
 - 例) 平均値の差(95%CI) 5.2 (0.9, 9.5) 有意
 - 0を含むか? 5.2 (-0.8, 11.2) 有意でない
 - オッズ比(95%CI) 2.7 (1.2, 6.1) 有意
 - 1を含むか? 2.7 (0.7, 10.0) 有意でない
- 差の大きさに興味がある場合は、検定よりも推定。

応用

再び、

“有意差なし”は“差がない”ことを積極的に示したわけではない！

例題 12. そこで、同等性の検定、非劣性の検定

口腔内装置による閉塞型無呼吸症候群の治療効果

最大中間呼吸速度の改善度

	人数	平均	標準誤差	90%信頼区間
A: 75%前方位	50	10.0	0.32	
B: 50%前方位	50	9.8	0.32	
平均の差		-0.2	0.45	-0.95~+0.55

有意差なし
= 差がない

- 「通常はAを用いるが、Bでもいいのでは？(Bも劣っていない)」ということを知りたい。
- Bの平均改善度がAの±10%(±1)以内ならば、同等とみなそう。
 - 信頼区間が±1に収まっているので同等。
- Bの平均改善度がAより10%(-1)以上劣っていなければ、非劣性とみなそう。
 - 信頼区間の下限が-1以上なので非劣性(劣っていない)。
- どこまで許容するかは、臨床的意義によって決める。
- 有意水準 $\alpha = 5\%$ の場合は90% (=1-2 α) 信頼区間で判断する。

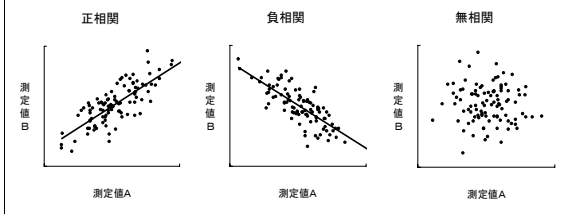
- 研究計画に関する基礎知識
- データの整理
- 検定とは？
 - 数量データの差の検定
 - 質的データの差の検定
 - 検定と区間推定
- 関連の分析
 - 2つの数量データの関連
- 多変量解析

相関分析と回帰分析がよく使われる

復習

相関と回帰

図7 正相関と負相関

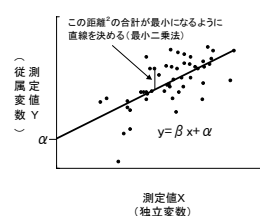


- 相関係数
 - -1~+1の値をとり、2変数の直線的な関連の強さを表す。
 - 検定も行う(帰無仮説: 母相関係数=0)

復習

相関と回帰

図8 回帰直線



- 回帰直線
 - 2つの連続量の関係を、 $y = \beta x + \alpha$ の形の1次式で表したものを。
- 回帰係数 β
 - 相関係数と違い、単位があるので、様々な値をとる。独立変数が1増加した時の、従属変数の増加量の期待値を表す。
- 相関係数の検定(帰無仮説: 母相関係数=0)
- 回帰係数の検定(帰無仮説: 母回帰係数=0)
- 両者の結果は一致する。

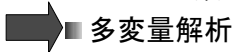
応用 例題4. 独立な3群以上のトレンド検定

降圧薬A投与量と収縮期血圧の低下幅

	降圧薬A投与量						P for trend
	5mg		10mg		20mg		
	平均	SE	平均	SE	平均	SE	
収縮期血圧 (mmHg)低下幅	-3.1	1.2	-5.6	1.3	-8.3	1.2	?

- 分散分析
 - 帰無仮説 $H_0: \mu_1 = \mu_2 = \mu_3$
 - 対立仮説 $H_1: \mu_1 = \mu_2 = \mu_3$ ではない
 - つまり、どこかとどこかに差がある、ということがいえる(でこぼこしていても可)。
- トレンド検定
 - 対立仮説 $H_1: \mu_1 < \mu_2 < \mu_3$ (または逆)
 - つまり、量反的に増加する(または減少する)、ということがいえる。
 - 計算の仕方
 - 各個体の収縮期血圧低下幅を目的変数、カテゴリーの値(5, 10, 20)を説明変数として回帰分析を行う。回帰係数のP値(H_0 : 母回帰係数=0)がP for trend。
 - カテゴリーが幅を持っている場合(1日あたり歩数をカテゴリー化した場合など)は、その平均値(中央値)をカテゴリーの値として用いるのが一般的。

- 研究計画に関する基礎知識
- データの整理
- 検定とは？
 - 数量データの差の検定
 - 質的データの差の検定
 - 検定と区間推定
- 関連の分析
 - 2つの数量データの関連

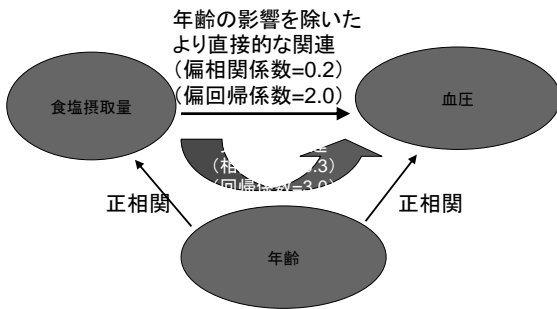


偏相関、重回帰分析が基本

応用

偏相関と重回帰

- 他の要因の影響を補正したうえで、2変数間の直線的な関連を表す方法。



応用

重回帰分析

- 注目している連続量Yと、複数の要因 X_1, X_2, \dots, X_n との関係を1次式の形で表したものを

$$Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \text{切片} + \text{誤差}$$
- $\beta_1 \sim \beta_n$ を偏回帰係数という。
- Yは正規分布(正確には誤差が正規分布)
- 重回帰分析では、偏回帰係数と切片を最小二乗法で推定して解釈する。
- どの程度よく説明できているかを表す指標として、決定係数 R^2 を参考にする。
- 多重ロジスティック回帰ではYが疾病有無のlogit、多変量Cox回帰ではYがハザードの形になっている。
 - 従って、解釈のしかたは似ている。まずは重回帰から。

応用

(単)回帰分析と重回帰分析の解釈の違い
(多重ロジスティック回帰、多変量Cox回帰も同じ)

	収縮期血圧			重回帰分析		
	(単)回帰分析			重回帰分析		
	回帰係数	標準誤差	P値	偏回帰係数	標準誤差	P値
飲酒量(合)	4.0	0.5	<0.001	4.1	0.5	<0.001
喫煙量(箱)	2.0	0.9	0.02	0.5	0.8	0.90

- (単)回帰分析の解釈
 - 飲酒量が1合多いと、血圧は4mmHg高いが、これに含まれる喫煙の影響はわからない。
 - 喫煙量が1箱多いと、血圧は2mmHg高いが、これに含まれる飲酒の影響はわからない。
- 重回帰分析の解釈
 - 喫煙の影響を除いても(調整しても)、飲酒量が1合多いと、血圧は4.1mmHg高い。
 - 飲酒の影響を除くと(調整すると)、喫煙量と血圧の関係は明らかでない。

- 他の変数の影響を調整したうえで、2変数間の関連を調べるのが重回帰分析。
- 同時に用いた説明変数によって、解釈が少し変わる。

応用

重回帰分析の説明変数に関する注意
(多重ロジスティック回帰、多変量Cox回帰も同じ)

- 全く同じ意味を持つ2変数を同時に使ってはいけない。
 - 例)2回測定した血圧を、2つとも同時に説明変数に入れるのはナンセンス!
- 類似の理由で、相関が非常に強い2変数を同時に使うのは、望ましくないことが多い。
- 変数のもつ医学的な意味が変わることがあるので注意。
 - 例1)収縮期血圧SBPと拡張期血圧DBPを同時に入れると、DBPで調整したSBPって・・・“脈圧みたいなもの”?
 - 例2)身長と体重を同時に入れると、身長で調整した体重って・・・“肥満度みたいなもの”?

応用 例題13: どのような分析を行うか

歯科材料への着色の程度(値は平均±SD)

		フッ素	
		(-)	(+)
紅茶色素	(-)	1.0±0.4	2.0±0.5
	(+)	3.0±0.6	4.0±0.5

紅茶色素の
効果=2.0,
P=0.01

フッ素の効果=1.0, P=0.05

- 組合せにして4カテゴリーで一元配置分散分析?
 - 「組合せによって着色の程度が違う」ということがいえる。
 - それでフッ素の影響なのか、紅茶の影響なのか、よく分からない。
- 二元配置分散分析
 - アウトカム(着色)に及ぼす、二つの要因(フッ素、紅茶色素)の独立な影響を分析する。
 - フッ素の影響と、紅茶の影響と、分離して評価できる。

応用 例題13': 二元配置分散分析

歯科材料への着色の程度(値は平均±SD)

		フッ素	
		(-)	(+)
紅茶色素	(-)	1.0±0.4	2.0±0.5
	(+)	3.0±0.6	7.0±0.5

紅茶色素の
効果=??

フッ素の効果=??

- 二元配置分散分析
 - フッ素の有無によって紅茶色素の効果が変わる(逆も同様)。
 - そのため、紅茶色素の効果とフッ素の効果を単純には示せない。
 - 交互作用という概念が必要。

応用 例題13'': 二元配置分散分析

歯科材料への着色の程度(値は平均±SD)

		フッ素	
		(-)	(+)
紅茶色素	(-)	1.0±0.4	2.0±0.5
	(+)	3.0±0.6	7.0±0.5

紅茶色素の
主効果=2.0

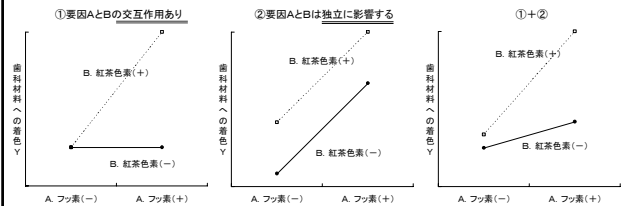
フッ素の主効果=1.0

交互作用=3.0
P=0.03

1.0+2.0+1.0=4.0のはずのところが7.0になっているので

- 二元配置分散分析(交互作用あり)
 - フッ素と紅茶色素が単独の時の効果がそれぞれの主効果。
 - 同時に組み合わせさせた時に、主効果の和にさらに上積みされる効果が交互作用。“フッ素×紅茶色素”のようにかけ算の記号で表記することが多い。
 - 交互作用がある時は、主効果だけの解釈はしない。交互作用も見て、総合的に解釈する。

応用 二元配置分散分析のキモ 交互作用を理解しよう



- ①フッ素と紅茶色素が同時に存在する時のみ、着色される。(交互作用あり)
- ②フッ素と紅茶色素は独立に、着色に影響する。(交互作用なし、主効果あり)
- ①+②両方が混ざった状態(実際はこれが多い。交互作用あり、主効果少しあり)

最後に

- 大規模な研究プロジェクトには、計画段階から疫学・生物統計学の専門家を1名、仲間に入れましょう。
- データ収集後に相談するのでは、手遅れかも。



国立保健医療科学院における 生物統計関連の教育

- 遠隔教育・生物統計学
 - いわゆるe-learning。埼玉県まで来なくても自宅等で受講できる。3ヶ月かけて教科書を1冊学習。
 - 定員30名。
 - 臨床試験に係わる臨床医向け生物統計学研修
 - 臨床試験のプロトコルを自分で作って実施しようという臨床医向け。臨床試験に特化した研修で、統計学そのものは時間をあまりかけない。
 - 専門課程・生物統計分野
 - 生物統計の本物の専門家を目指す人向け。最低1年間専念。
- いずれも昨年度実績。今年度について詳しくは：
http://www.niph.go.jp/soshiki/gijutsu/index_j.html
 ● このハンドアウトの最新版：
<http://www.niph.go.jp/soshiki/gijutsu/staffs/yokoyama/etc/hotetsu2006.pdf>