1月13日（金）

講義4　統計（1）

国立保健医療科学院
技術評価部
主任研究官　横山徹爾

学習目的

- 調査における標本抽出法の種類を説明できる。
- ランダム誤差と系統誤差について説明できる。
- サンプルサイズの決め方の基本原則がわかる。
- 測定誤差を減らす方法について説明できる。
- １日の食事調査で評価する問題点がわかる。
- 代表値と散布度の使い方がわかる。
- 標準偏差と標準誤差の違いがわかる。
- EARとRDAを統計学的観点から説明できる。

要約

医学、栄養学、保健学研究は対象とする人々を観察することから始まる。「観察する」とは、その人たちの持つ様々な「特性」を調査、記録、分析することである。疫学・統計学は、これら調査、記録、分析をいかに行うかについての「方法論」を提供する。

母集団の特性を推定するためには、標本調査を行うのが一般的であり、その際には無作為抽出を行わなければならない。標本数が小さければ偶然誤差が大きく、回収率が低ければ系統誤差（バイアス）が生じる可能性がある。標本数を決めるためには、あらかじめ必要な推定精度を定めなければならない。むやみに標本数を大きくするよりも、回収率を高める方が大切である。

測定には誤差がつきものであり、誤差の原因と種類を理解して、誤差を小さくする工夫をすべきである。例えば、１日の食事調査では日間変動による偶然誤差が大きく、個人レベル、集団レベルのいずれでも評価が困難である。多数日の調査によってこの偶然誤差を小さくすることができる。

調査データを整理する手順として、最初にデータの要約を行う。これには、代表値と散布度、パーセント点で表す等の方法がある。データの分布の代表的なものとして正規分布があり、その性質は統計学の様々な場面で利用される。食事摂取基準において、推定平均必要量、推奨量が定められた手順を理解するためには、正規分布の性質を知っておくべきである。

第1回日本栄養改善学会実践栄養学研究冬季集中セミナー
2005.1.13-15

統計（1）

国立保健医療科学院技術評価部
横山徹爾

1. 統計とは

- 医学、栄養学、保健学研究は対象とする人々を観察することから始まる。「観察する」とは、その人たちの持つ様々な「特性」を調査、記録、分析することである。疫学・統計学は、これら調査、記録、分析をいかに行うかについての「方法論」を提供する

統計にはいろいろな目的がある。目的に応じて使うことが大切

統計の目的

- 総量の把握：国勢調査人口
- 集団の構造や特性の把握：年齢3区分人口割合、死因別死亡率、栄養素摂取量
- 複数要因間に存在する法則、因果関係の探究：年齢と死亡率、塩分摂取量と血圧、血圧と脳卒中罹患率。分析疫学の多くはこれを目的とする。
- 予測：将来推計人口。
- 評価：保健活動の量的な評価。実験疫学（介入研究）の評価。

など。

常に対象集団を意識する

- 対象集団：観察の対象として設定された集団。
- 全数調査（悉皆調査）：対象集団の構成員全員の調査。国勢調査など。<u>記述統計学</u>。
- 標本調査：対象集団から（一部分）抽出した標本の調査。対象集団全体のことを母集団と呼ぶ。<u>推測統計学では標本から母集団の特性を推測する。</u>

2. 標本調査の基礎

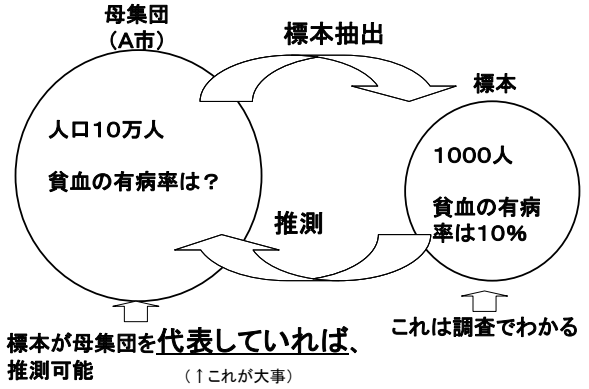
（1）標本調査と無作為抽出

無作為抽出されていない標本を用いて、ものを言うことはできない

標本調査の目的

- 母集団における
  身長、体重、栄養摂取量の分布。
  肥満、貧血、高脂血症の有病率。
  を知りたい。
- 全員調べれば（悉皆調査）わかるが、とうてい不可能である。
- そこで、一部の人たち（標本）を調べて、母集団全体の様子を推測しよう！

母集団と標本



無作為抽出

- 標本調査によって母集団の特性を推定するためには、必ず無作為抽出を行う。
- まず、母集団を確認。例えば、国民健康・栄養調査では日本人全体が母集団、県民健康・栄養調査では県民全体が母集団である。
- 標本抽出を行う場合は、母集団をいくつかの抽出単位（“個人”、“世帯”、“国勢調査区”、“市区町村”など目的に応じて決める）に分け、全ての抽出単位が選ばれる確率が等しくなるように工夫する。これを無作為抽出といい、例えば、それぞれの抽出単位に通し番号を付け、乱数によって標本を選び出せばよい。

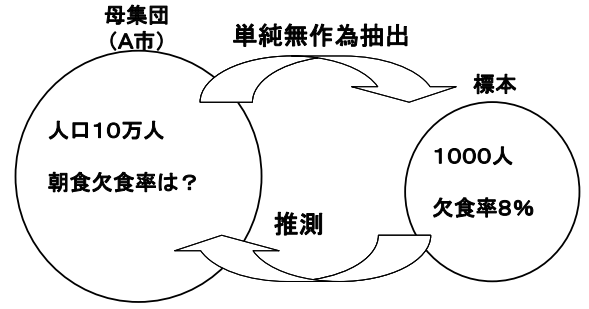
無作為抽出の方法

- 標本から母集団の特性を推測するためには、無作為抽出を行わなければならない。代表的でない（偏りのある）データから母集団の特性は推測できない！
- 単純無作為抽出法
- クラスター抽出法
- 層化クラスター抽出法
- など

単純無作為抽出法

- A市の全住民のうちから、住民基本台帳から乱数によって選んだ1000名を対象として調査を行う、というように、母集団を構成する個人を抽出単位として無作為抽出を行う方法。
- 抽出人数÷全人口を抽出率という。
- 住民アンケートなどで、広く用いられている。
- 乱数→乱数表、サイコロ、コンピュータなどを利用する。コンピュータを用いるのが現実的（エクセルのRAND()関数など）

市民アンケート



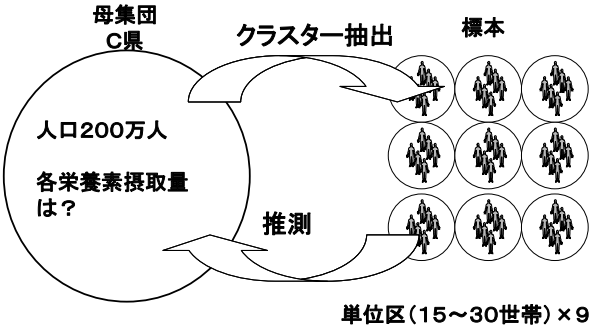
層化無作為抽出法

- B県を複数の地区＝層（保健所・二次医療圏別・市町村など）に分け、各地区内では単純無作為抽出を行うというように、対象集団をあらかじめ複数の層に分けてから無作為抽出する方法。
- 地理的な区分だけでなく、性・年齢階級などで層化することも多い。

クラスター抽出法

- B県内の国勢調査区から、乱数によって選んだ10地区の住民全員を対象として調査を行う、というように、母集団をいくつかの“集落”＝クラスター（市区町村、国勢調査区、単位区、世帯など）に分け、クラスターを抽出単位として無作為抽出を行い、選ばれたクラスター内の構成員全員を調査対象とする方法。
- 調査地域が広い場合（例えば全県レベルの調査）の訪問調査などで行われることが多い。移動の手間が小さくて済むが、同じ調査人数ならば（単純無作為抽出よりも）推定精度は低い。
- 県民健康・栄養調査は、この方法が多い。

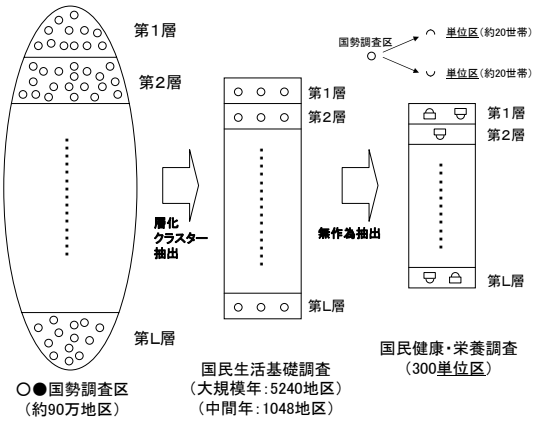
県民健康・栄養調査の標本抽出



層化クラスター抽出法

- 層化無作為抽出の抽出単位を集落（クラスター）にしたもの。
- 国民健康・栄養調査はこの方法（に近い）。
- 都道府県（と政令指定都市）別に、国勢調査区を分割した“単位区”（15～30世帯程度）を抽出単位としてクラスター抽出を行う。（実際には、国民生活基礎調査で選ばれた国勢調査区から再抽出している）

国民健康・栄養調査の標本抽出の概略



2. 標本調査の基礎

（2）標本誤差とバイアス

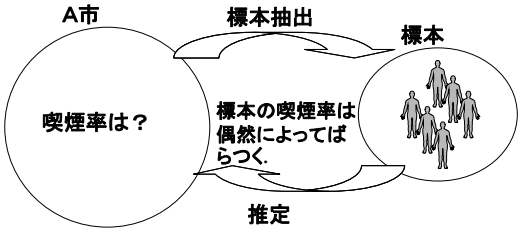
調査には誤差がつきもの。2種類の誤差を理解する。

標本調査と誤差・偏り

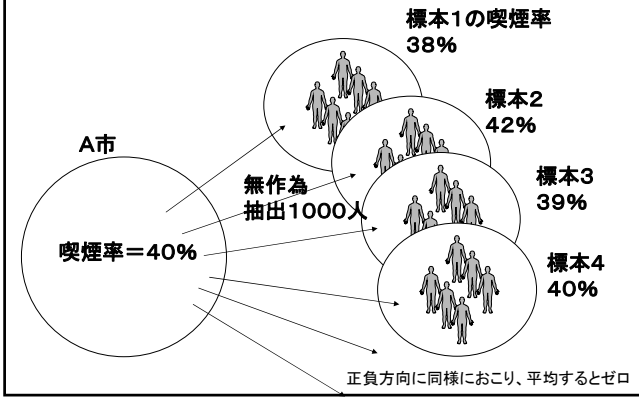
- 誤差：真の値と、観察した値とのずれ。標本調査ではつきもの。
- ランダム誤差：偶然現象によって生じたずれ。標本抽出における誤差を特に標本誤差という。統計学である程度制御可能。
- 系統的誤差：何らかの理由により、一定方向（正または負）に生じたずれ。「偏り」ともいう。統計学で制御不可能。

ランダム誤差（標本誤差）

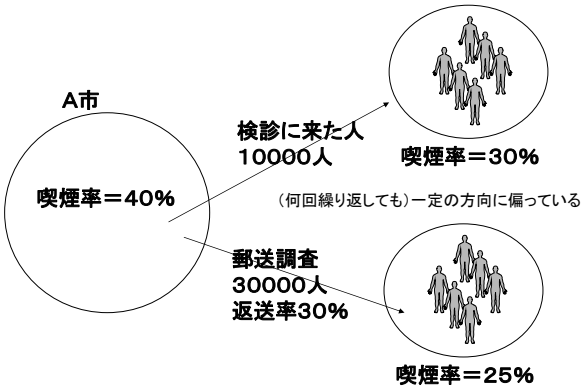
- ランダム誤差
  偶然によっておこる。標本抽出におけるランダム誤差（標本誤差）は、標本数が多いほど小さい。正負方向に同様におこり、平均するとゼロになる。



（抽出における）ランダム誤差の概念



系統的誤差（偏り）の概念



ランダム誤差（標本誤差）を減らすには

- 調査人数を増やす。

系統的誤差（偏り）を減らすには

- 無作為抽出法を用いる。
- 回収率を高める。

2. 標本調査の基礎

（3）誤差率とサンプルサイズ

どのくらいの標本サイズが必要ですか？と漠然と聞かないで下さい。

調査人数の決め方

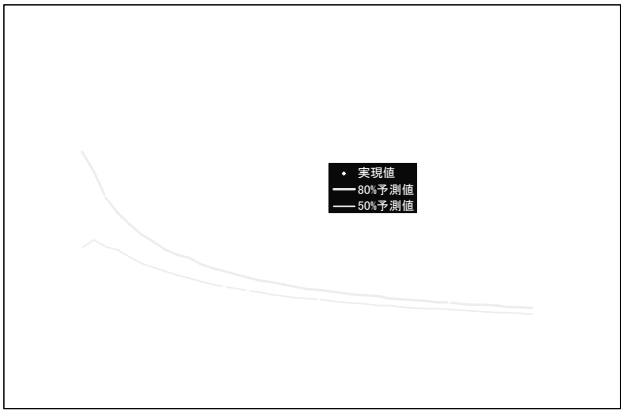
- 調査が終わって集計する段階で、
  - 「統計として有効な人数ですか？」

  という質問をしてはいけない。
- 調査人数は、あらかじめ十分な精度が得られるように計画して決めておくべきもの。
- 調査前に統計の専門家に相談する！
  - 相談のしかた
    - 20代男性の朝食欠食率を知りたい。おおむね10%程度だろうと思う。誤差率5%とすると何名調査したらよいか？
    - 市民のカルシウム摂取量の平均値を知りたい。過去の調査では平均550mg、標準偏差は290mgだった。誤差率5%とすると何名調査したらよいか？
- 検定することを前提として決めることもある。

調査人数（客体数）

- あらかじめ定めた誤差率（例えば1％）を達成するために必要な人数を調査する（ただし、実際には予算・期間等の制約を受ける）。
- 誤差率は、“標準誤差÷推定値（平均等）”と定義され、誤差率が小さいほど推定精度が高い。一般に、調査人数が多いほど誤差率は小さい。
- 必要以上に調査人数を増やすよりも、回収率を向上させるために労力を投じるべきである。

国民健康・栄養調査方式での都道府県別単位区数と誤差率との関係



「国民健康・栄養調査における各種指標の設定及び精度の向上に関する研究」（主任研究者：吉池信男）平成16年度報告書より

どちらの結果を信じる？

- 母集団の人口は50万人。朝食欠食率を知りたい。いずれも単純無作為抽出。
  （1）郵送調査5万人（返送率30％）。
  （2）郵送調査1000人（返送率95％）。

3．測定誤差

測定には誤差がつきもの。誤差を減らす方法を学びましょう。

“測ったもの（測定値）”

測定値＝真の値＋誤差

- “きちんと（正しく）測る”とは、
→可能な限り“誤差が小さくなるように”測ること。

- 誤差が大きいと、例えば、
地域間の差＜測定値の誤差・・・地域間の比較不能
経年変化＜測定値の誤差・・・経年的な比較不能

誤差を減らすには？

- 血圧、腹囲、食物摂取量など、健康・栄養調査ではこれらの値を可能な限り“誤差なく正確に”測定することが望まれるが、“誤差のない測定はあり得ない”と言われるほど、一般の人々から医学・栄養学的なデータを得ることは難しく、測定値には様々な誤差が入ってくる可能性がある。
- しかし、誤差を減らすことは可能である。そのためには、まず、誤差の種類と原因を理解し、調査を計画・実施する者は、それに基づいて可能な限り誤差が小さくなるように配慮すべき。

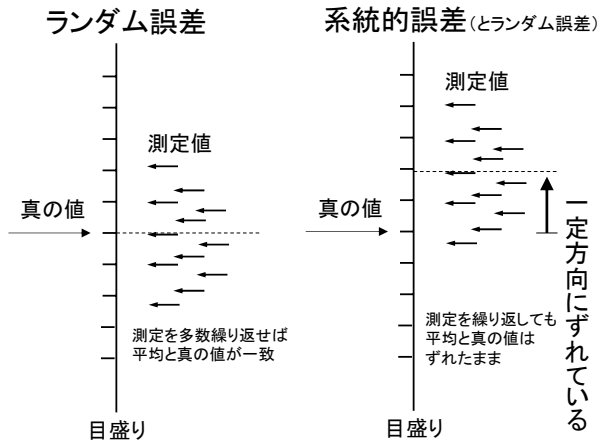
誤差とは？

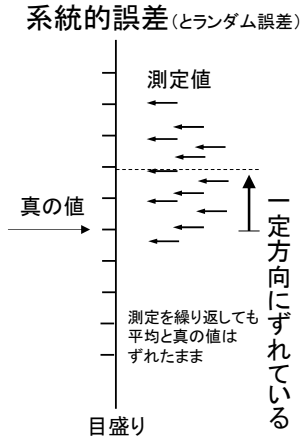
- 種類（どの段階で生じるか）によって、
  標本誤差・・・標本抽出の段階で生じる誤差（既述）
  測定誤差・・・測定の段階で生じる誤差

- 性質によって
  系統的誤差・・・測定値が真の値から特定の方向にずれている（偏り）
  ランダム誤差・・・“ずれ”が、特定の方向に偏らない（平均するとゼロ）

測定誤差

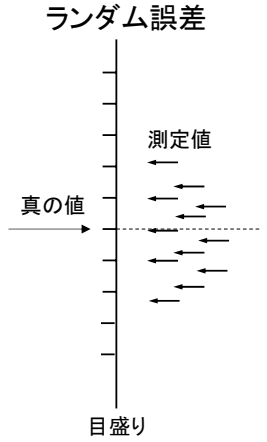
- 測定の段階で生じる誤差
- 性質によって
  系統的誤差・・・測定値が真の値から特定の方向にずれて（偏って）いる
  ランダム誤差・・・“ずれ”が、特定の方向にずれない（偏らない＝ずれの平均がゼロ）





測定の系統的誤差の例

- 被験者の測定条件
  - 食事記録調査を週末だけ行った、夏だけ行った
- 被検者の“くせ”
  - 肥満者の過少申告
- 測定者の“くせ”
  - 未熟なインタビューアによる24時間思い出し法
  - コード付けの不統一
- 測定機器の“くせ”
  - 食品成分表
  - 絶対量の評価を想定していないFFQ

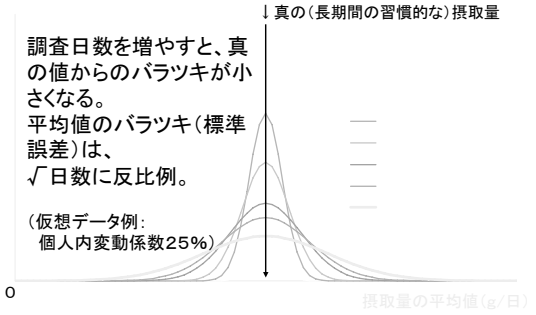


測定のランダム誤差の例

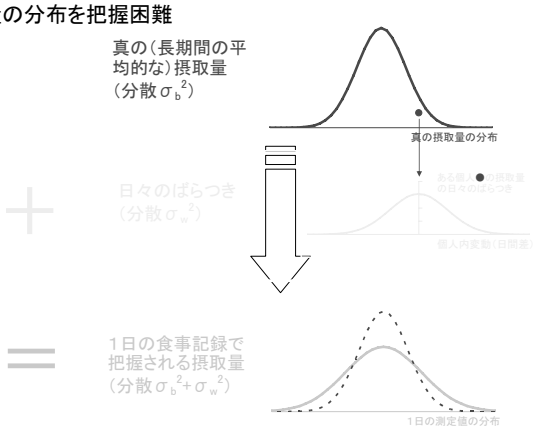
- 日間変動、日内変動
  - 被測定者のその日、その時々の体調の微妙な変化によって、測定時の血圧がたまたま（全くの偶然に）いつもより高め、あるいは低めだった。
  - 調査日は、いつもと比べてたまたま野菜摂取量が少なかった。
- 複数回測定すれば、その平均が真の値に近づく。
  - 同じ人に血圧を2回ずつ測定してその平均値をとる。
  - 複数日の食事記録調査を行いその平均値をとる。

食事記録による測定値は何を見ているのか？

- 測定値
  ＝真の（長期間の習慣的な）摂取量
  ＋日間変動（←調査日数を増やして平均をとると減少）（＋その他の誤差）



1日の食事記録では、個人の摂取量、および集団レベルでの摂取量の分布を把握困難



4. データの整理

いきなり複雑なことをしてはいけない。まずは基本から。

調査データの種類

- 質的データ:
  - 2値 （例）性別の“男”と“女”、欠食の“有”と“無”
  - 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）肥満度分類の “低体重”, “普通体重”, “肥満”
    - 名義尺度nominal scale：順序関係がない分類のための変数。
      - （例）職業の“管理職”, “事務職”, “技術職”
- 数量データ：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、栄養素摂取量、血清総コレステロール
  - 離散データ （例）う歯の本数

質的データの整理

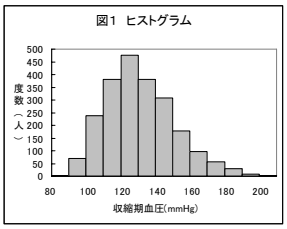
- 準備
  - 単純に頻度を記述。欠損値の数、変な値が紛れ込んでいないかなどをチェック。
- 単純集計
  - 地区・性・年齢階級別頻度を整理する。
- クロス集計
  - 二つの要因の関連を調べるために行う（交絡変数に注意）。
    - 例：喫煙状況と高血圧の頻度
  - 検定
    - 二つの要因が関係しているように見えるが、これは偶然だろうか？という問いに答えるために検定する。χ^2検定など（後述）。

		高血圧		
		あり	なし	計
喫煙状況	既喫煙（やめた）	20	10	30
	現在喫煙	15	35	50
	非喫煙	5	15	20
	計	40	60	100

χ^2=12.85（自由度=2），P<0.005

量的データの整理

- 準備
  - ヒストグラムを描き、分布の形を確認する。欠損値の数、変な値が紛れ込んでいないかなどをチェック。
  - カテゴリー型に変換する場合（BMIを肥満の有無で3群に分けるなど）は、以後、質的データの整理と同様。
- 単純集計
  - 地区・性・年齢階級別の要約統計量（代表値等）で分布の特徴を記述。
    - 正規分布の場合・・・平均と標準偏差
    - 歪んだ分布の場合・・・中央値と25，75パーセント点など
- 群間比較
  - 二つの要因の関連を調べるために行う（交絡変数に注意）。
    - 例：A地区とB地区でどちらが食塩摂取量が多いか？
      - 平均値と標準偏差を群間で比較する。
      - 検定：二つの地区間で食塩摂取量の平均値が異なるように見えるが、これは偶然だろうか？という問いに答えるために検定する（後述）。



- 階級数は、√標本数＋1前後を目安にすると形が分かりやすい。
- 分布の形を確認する。
  - 左右対称か？→右裾が長い場合、対数変換を考慮
  - 外れ値はないか？→標本として適切か検討
  - 分布の中心位置はどのあたりか？→代表値（平均、中央値など）
  - 分布のばらつき具合は？→散布度（標準偏差、四分偏差など）

分布型を確認

- 統計学的方法の多くは、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、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ると都合がよいこともある。（対数変換など）

図2　正規分布

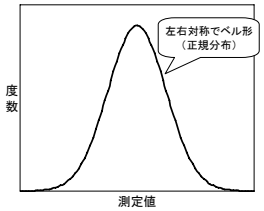
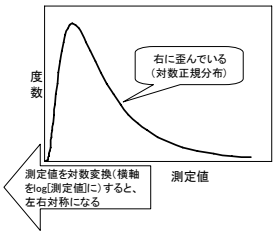
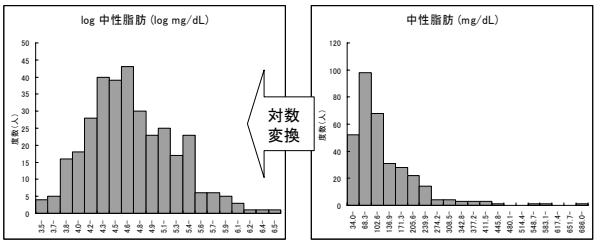


図3　対数正規分布

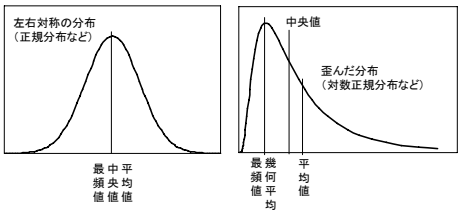




- 対数正規分布の典型例
  - 中性脂肪、ある種のビタミン類摂取量、など
- 正規分布の典型例
  - 身長、体重、エネルギー摂取量など
- 医学データは、少し右裾が長いことが多い

代表値（中心位置の指標）

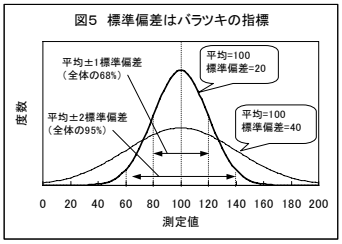
図4　分布型と代表値



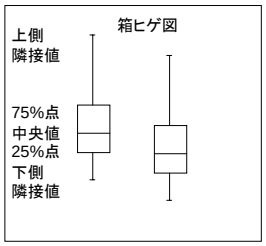
- 平均値・・・左右対称な場合に有用
- 中央値・・・非対称等、歪んだ分布の場合

散布度（バラツキの指標）

図5　標準偏差はバラツキの指標



箱ヒゲ図



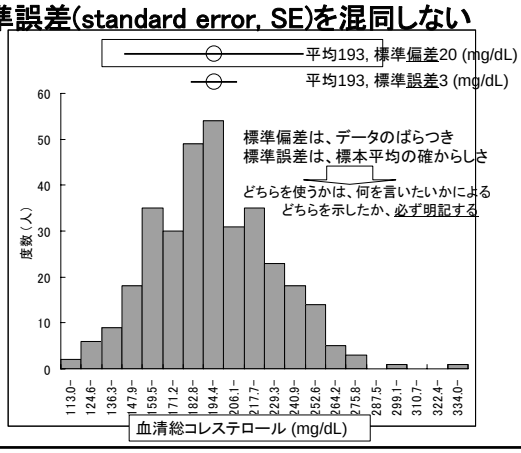
- 「平均と標準偏差」　（標準偏差^{2}を分散という）
- 「中央値と四分偏差（25%点, 75%点）」
  の組合せがよく用いられる。

パーセント点を活用しよう

◆第68表◆　動物性たんぱく質摂取量の分布（性・年齢階級別）

	年齢（歳）	食事摂取基準 EAR	RDA	UL	調査人数	平均値	標準偏差	パーセンタイル 1	5	10	25	50(中央値)	75	90	95	99
男	1～2	—	—	—	122	23.9	11.9	3.4	6.0	10.5	14.4	22.4	32.7	41.4	43.6	50.6
	3～5	—	—	—	197	28.8	12.9	5.4	9.7	13.3	19.8	26.9	37.3	44.1	51.9	73.2
	6～8	—	—	—	226	35.7	14.7	9.8	15.2	19.2	25.4	33.8	43.3	54.3	62.1	89.7
	9～11	—	—	—	212	44.3	17.9	13.8	20.1	25.6	33.3	41.5	52.8	63.5	82.2	99.1
	12～14	—	—	—	230	51.9	23.6	9.1	22.0	27.5	36.1	49.0	63.3	76.6	90.6	147.3
	15～17	—	—	—	231	55.2	25.1	15.5	23.0	28.5	38.9	52.1	65.6	85.9	103.4	142.3
	18～29	—	—	—	714	45.8	22.4	4.6	16.4	21.4	29.9	42.7	57.9	72.8	85.8	114.2
	30～49	—	—	—	1,500	45.3	21.9	8.5	17.0	20.8	30.3	42.4	56.4	72.4	84.2	111.2
	50～69	—	—	—	1,734	47.5	24.1	6.5	15.9	22.3	32.1	44.0	56.8	74.4	92.6	123.2
	70以上	—	—	—	686	36.4	19.7	2.2	9.4	14.4	23.1	32.8	46.9	62.0	73.5	98.7
	総　数	—	—	—	5,852	44.2	22.7	5.6	14.6	19.4	29.0	40.9	55.7	70.9	84.6	116.4
女	1～2	—	—	—	138	21.1	11.0	1.8	5.9	9.0	14.3	20.4	26.1	33.2	37.9	57.2
	3～5	—	—	—	199	29.0	11.8	9.8	12.8	15.3	20.1	28.2	34.6	43.4	48.6	74.5
	6～8	—	—	—	215	34.3	12.7	13.1	16.5	19.8	25.7	33.1	41.3	49.5	53.4	73.1
	9～11	—	—	—	199	38.5	14.6	12.9	18.0	22.1	27.7	36.3	47.3	57.6	64.6	93.4
	12～14	—	—	—	230	42.5	16.9	10.0	20.7	22.8	32.6	39.7	51.3	62.6	72.6	86.4
	15～17	—	—	—	209	42.9	18.3	8.0	16.6	20.9	29.0	42.1	54.2	67.4	73.4	87.7
	18～29	—	—	—	764	36.5	18.8	3.1	11.2	14.9	23.2	35.1	46.0	59.0	67.3	92.3
	30～49	—	—	—	1,635	35.5	17.4	3.5	11.0	15.7	23.2	33.5	44.6	56.6	66.8	86.2
	50～69	—	—	—	1,945	38.6	18.8	4.3	12.0	17.5	25.7	36.3	49.0	62.0	72.2	96.2
	70以上	—	—	—	967	32.1	16.7	0.1	7.5	12.2	20.9	30.5	41.9	54.9	63.6	78.2
	妊婦	—	—	—	48	35.2	16.9	8.3	10.9	16.5	21.1	33.1	43.1	56.7	68.0	88.3
	授乳婦	—	—	—	80	38.4	17.1	11.4	16.7	19.4	26.5	34.8	49.1	61.8	65.3	118.1
	小　計	—	—	—	128	37.2	17.1	9.6	15.6	17.7	25.2	34.5	45.7	61.5	66.2	88.3
	総　数	—	—	—	6,501	36.1	17.8	3.5	11.2	15.4	23.7	33.9	45.8	58.4	67.7	87.7
	総数＋小計	—	—	—	6,629	36.1	17.8	3.6	11.2	15.5	23.7	33.9	45.8	58.5	67.7	88.3
全国		—	—	—	12,481	39.9	20.6	4.5	12.5	17.3	25.9	36.9	50.5	65.2	76.0	106.5

標準偏差(standard deviation, SD)と標準誤差(standard error, SE)を混同しない



標準偏差と変動係数の関係

■ 変動係数＝標準偏差÷平均値

または、

■ 標準偏差＝平均値×変動係数

つまり、標準偏差が平均値の何倍かを表す。

標準誤差と誤差率の関係

■ 誤差率＝標準誤差÷平均値（推定値）

または、

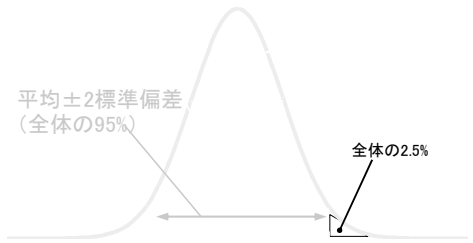
■ 標準誤差＝平均値×誤差率

つまり、標準誤差が平均値の何倍かを表す。

5. 統計学と食事摂取基準の原理

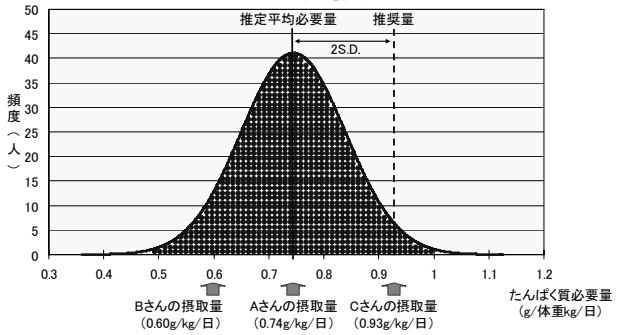
食事摂取基準では確率論が重視されています。

正規分布の性質



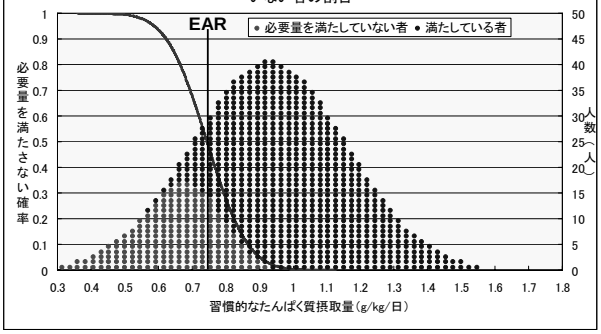
問1） 「平均」、「標準偏差」、「正規分布」という用語を使って、推定平均必要量と推奨量の関係を説明しなさい

図1. 成人1000人のたんぱく質必要量
（●が1個人を表す）

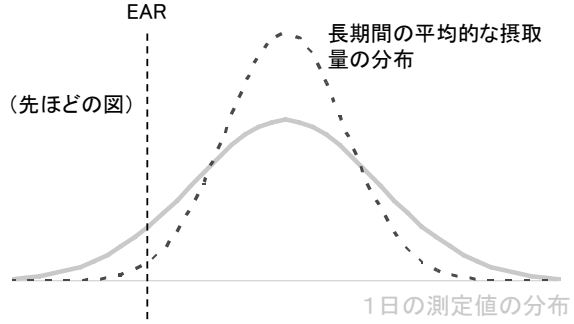


問2） 集団において、習慣的摂取量がEAR以下の者の割合は、不足者の割合と、（経験的にたまたま）ほぼ一致する。1日の食事記録調査によってこのような評価を行うと、どうなるか？

図3. 成人1000人の集団において、たんぱく質摂取量が必要量を満たしていない者の割合



EAR以下の人の割合は、1日調査ではこんなに過大評価



1月14日（土）

講義5　統計（2）

国立保健医療科学院
技術評価部
主任研究官　横山徹爾

学習目的

- 検定の概念、帰無仮説、対立仮説について説明できる。
- 検定における第1種・第2種の過誤、有意水準、検出力について説明できる。
- t検定、分散分析、Wilcoxon順位和検定、Wilcoxon符号付順位検定、χ2検定、Fisherの正確な検定、拡張Mantel検定を、どのような分析で使うか説明できる。
- 多重検定がなぜいけないのか説明できる。
- 相関と回帰の違いがわかる。
- 残差法による総エネルギー調整の原理を説明できる。
- 重回帰分析、共分散分析、ロジスティック回帰、Cox比例ハザードモデルの結果の解釈ができる。

要約

研究計画の段階では、最終的な統計解析方法と図表のまとめ方まで考えたうえで、集めるデータの種類と標本数を決める。収集されたデータは、いきなり複雑な解析をせずに、データの分布型を視覚的に確認し、要約統計量として何が適切かを検討する。

観測された差や関連が、偶然で説明できるか否かを判断するのが検定である。その際、２種類の判断ミス（第１種、第２種の過誤）が生じうる。いわゆる“有意差なし”は、“差がない”ではない。検定方法は、データの分布型等の性質によって、パラメトリック検定とノンパラメトリック検定の２種類に大別される。検定をむやみに多数回繰り返すと（多重検定）、第１種の過誤が生じやすい。例えば、多群間の比較で単純にｔ検定を繰り返してはいけない。区間推定は検定と同様の意味を持ちつつ、より有益な情報が得られることがあるので使い分ける。

関連の有無を調べる場合には、連続変数では相関分析、回帰分析等、カテゴリー変数ではχ２検定等を用いる。栄養素摂取量と疾病リスクの関係を分析する際には、総エネルギー摂取量で調整するのが一般的。

重回帰分析、共分散分析、多重ロジスティック回帰、Ｃｏｘ比例ハザードモデル等の多変量解析では、変数の選択と解釈の仕方が重要。

参考図書

丹後俊郎著：統計学のセンス・医学統計学シリーズNo.1. 朝倉書店、1998.
古川俊之監修・丹後俊郎著：新版医学への統計学, 朝倉書店、1993.

第1回日本栄養改善学会実践栄養学研究冬季集中セミナー
2005.1.13-15

統計（2）

国立保健医療科学院技術評価部
横山徹爾

1．研究計画

後で考えるのでは遅すぎる

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ないと、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。
  - 論文を読むために必要な統計手法を学ぶ。

2．基本的なデータ解析

（1）データの整理

いきなり複雑なことをしてはいけない！

データを整理する

（昨日の学習範囲）

- 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

2．基本的なデータ解析

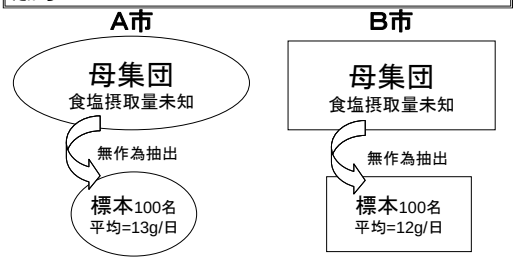
（2）検定とは

多種多様の検定方法があるが、その基本原理はたった一つ

統計学の原点：標本調査

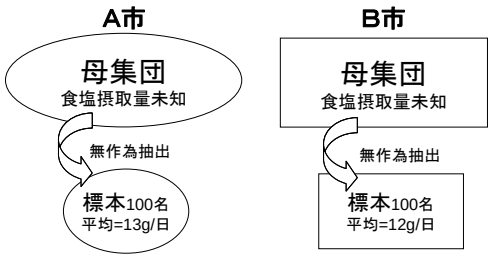
- 調査対象（母集団）を明確にする
- そこから無作為抽出された標本を用いて推測を行う

A市とB市から無作為抽出した標本100人ずつの食事調査をした。
A市の100名はB市の100名よりも食塩摂取量の平均値が1g少なかった。
だから・・・



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）



帰無仮説（A市とB市で母集団の食塩摂取量の平均は同じ）が正しい場合に標本平均に1g/日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（A市とB市で食塩摂取量の平均は異なる）を採用。

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

有意水準（P値）は、第1種の過誤が生じる確率

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

“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
  - A市とB市で　10人ずつ食事調査をした。両市の食塩摂取量の平均値の差は1g/日で、有意差はなかった。
  - A市とB市で100人ずつ食事調査をした。両市の食塩摂取量の平均値の差は1g/日で、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！）
  - 「同等性の検定」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。
  - 十分なサンプルサイズがないと、「何も言えない」

2. 基本的なデータ解析

（3）数量データの差の検定

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

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
  - 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

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

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

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)		
	平均	SD	平均	SD	P値
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

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？？？
  - 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！

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。
  - どことどこの組合せに差があるかは、まだ分からない！

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						
	1/1 (n=240)		1/2 (n=140)		2/2 (n=20)		
	平均	SD	平均	SD	平均	SD	P値
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

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
  - どこと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：Dunnet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

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
- どこと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
  - 非飲酒者と他の2群の比較に興味がある場合・・・Dunnet法
  - 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！

例題4：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		
平均	SD	平均	SD	平均	SD	P値
200	30	180	28	−20	10	?

- 対応のあるt検定
  - 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検定は、独立な2群の比較

2. 基本的なデータ解析

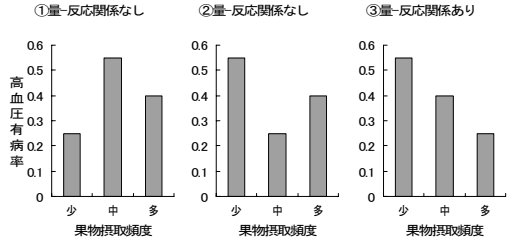
（4）質的データの差の検定

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

例題6：どのような検定を行うか

		高血圧		
		あり	なし	計
果物摂取頻度	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
- 拡張Mantel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



	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

例題7：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
    - 右耳と左耳が独立ではないから。
    - 独立性のχ^2検定は、独立な2群の比較

2. 基本的なデータ解析

（5）区間推定

検定よりも区間推定が望ましいこともある

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		
	平均	SD	平均	SD	P値
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。

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もつ。
- 差の大きさに興味がある場合は、検定よりも推定。

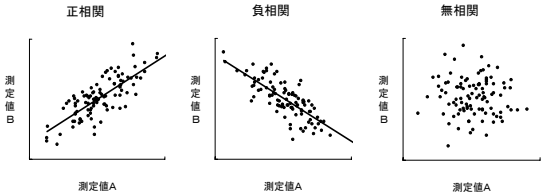
95%信頼区間は一般に、
点推定値±2×標準誤差

3. 相関と回帰

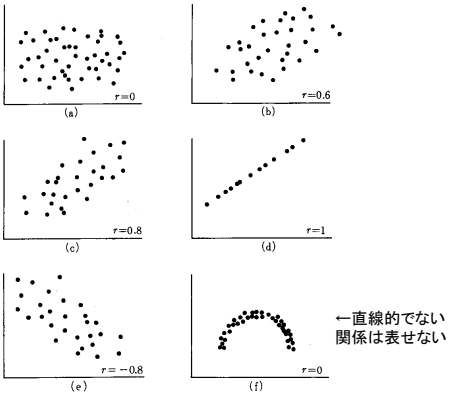
2つの数量データの関連を表す方法

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）



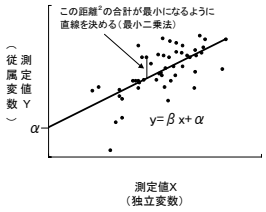
←直線的でない
関係は表せない

図 3　種々の散布図とそれに対する r の値

P.G.ホーエル著、浅井晃他訳　初等統計学第4判

相関と回帰

図8　回帰直線

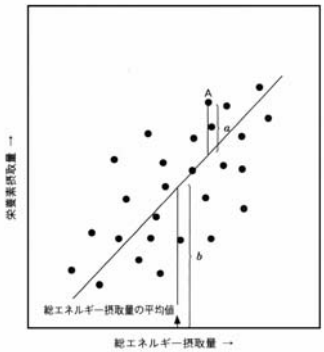


- 回帰直線
  - 2つの連続量の関係を、$y=\beta x+\alpha$ の形の1次式で表したもの。
- 回帰係数 β
  - 相関係数と違い、単位があるので、様々な値をとる。独立変数が1増加した時の、従属変数の増加量の期待値を表す。

- 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。

回帰分析を応用した、残差法によるエネルギー調整

総エネルギー摂取量の栄養素摂取量に及ぼす影響　図1



（Willett W. Nutritional Epidemiology）

- 総エネルギー摂取量が多い者は、
  - 体格が大きく、身体活動が多く、代謝効率が低い
  - ほとんどの栄養素摂取量が多い（左図）
- 栄養素摂取量と疾病リスクの関係を調べる際に、総エネルギー摂取量が交絡因子になる
  - 総エネルギー調整が必要
  - 調整栄養素摂取量
    $= b + a$

4. 交絡調整・多変量解析

注目している要因以外の影響を補正する

粗死亡率

	A地区（高齢型人口構成）			B地区（若年型人口構成）		
年齢階級	人口	死亡数	死亡率	人口	死亡数	死亡率
0～14歳	1000	10	1.0%	4000	80	2.0%
15～64歳	2000	30	1.5%	2000	40	2.0%
65歳以上	4000	160	4.0%	1000	50	5.0%
全年齢計	7000	200	2.9%	7000	170	2.4%

直接法による年齢調整死亡率

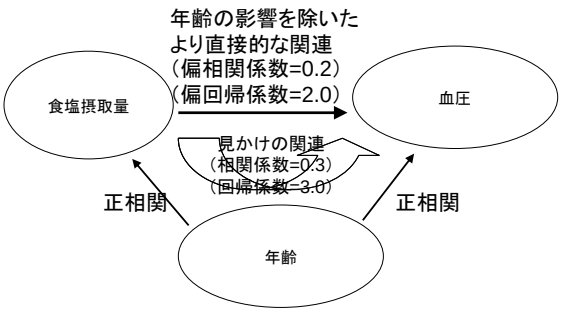
	基準集団	A地区		B地区	
年齢階級	人口	死亡率	期待死亡数	死亡率	期待死亡数
0～14歳	4000	1.0%	40	2.0%	80
15～64歳	5000	1.5%	75	2.0%	100
65歳以上	3000	4.0%	120	5.0%	150
全年齢計	12000		235		330
年齢調整死亡率（直接法）		235÷12000=1.96%		330÷12000=2.75%	

標準化死亡比（SMR）

	基準集団	A地区			B地区		
年齢階級	死亡率	人口	期待死亡数	観測死亡数	人口	期待死亡数	観測死亡数
0～14歳	1.8%	1000	18	10	4000	72	80
15～64歳	2.0%	2000	40	30	2000	40	40
65歳以上	4.2%	4000	168	160	1000	42	50
全年齢計	2.7%	7000	226	200	7000	154	170
標準化死亡比（SMR）			200÷226=0.885	(88.5)		170÷154=1.10	(110.4)

偏相関分析と重回帰分析

- 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。



重回帰分析

- 目的変数y、説明変数x_1, x_2, x_3…
  - $y = \beta_1 \times x_1 + \beta_2 \times x_2 + \beta_3 \times x_3 + \cdots +$切片$\alpha$
  - c.f. 回帰分析では$y = \beta x + \alpha$
- 解釈
  - x_2, x_3, …の影響を補正したうえでの、x_1とyとの直線的な関連の大きさがβ_1である。
  - β_2以下についても同様。
  - 全ての説明変数によって説明できるyのバラツキ（分散）の割合を、<u>決定係数</u>といい、R^2で表す。

共分散分析

- 重回帰分析の特殊型
- 複数の群間で、交絡変数の影響を調整した平均値（調整平均・最小2乗平均LSM）を計算できる。
  - 例： A市とB市の血圧の平均値を比較したい。標本調査したところ、A市の方が平均年齢が高かったので、単純に血圧の平均値を比較しても意味がない。そこで、A市とB市で年齢調整した血圧の最小2乗平均を計算して比較することとした。A市では130mmHg、B市では125mmHgであった。検定したところ$P<0.05$で有意であった。
  - 解釈： 年齢調整した血圧の最小2乗平均が、A市の方がB市よりも高いので、年齢の影響を補正してもA市の方が血圧の平均は高いと考えられる。

ロジスティック回帰

- 目的変数が病気の有り無し（食道がん有、無）、の2つの値だけをとる2値変数。
- オッズ比を計算できる。
- 回帰分析の一種

多重ロジスティック回帰

- 他の要因の影響を補正（交絡要因を調整）したうえでの、ある要因と疾病の有無との関連（調整オッズ比）を調べることができる。

一般に、点推定値（95%信頼区間）の形で表すことが多い。
例： OR = 2.7 (1.6−4.5) ・・・1を含まなければ“有意”

マッチングを行った場合は、条件付きロジスティック回帰を用いる。

多重ロジスティックモデルによる飲酒習慣と食道がんとの関連（症例対照研究）

説明変数	オッズ比※	(95%信頼区間)
飲酒量		
2合未満／日	1	基準群
2合以上／日	2.42	(1.50−3.90)
最も良く飲むアルコール飲料		
ビール	1	基準群
日本酒	1.15	(0.78−1.70)
焼酎	2.23	(1.51−3.30)
ウィスキー	3.15	(2.13−4.66)
その他	1.04	(0.70−1.54)

※年齢もモデルに含めて調整してある。

解釈例

1. 年齢と飲酒量の影響を調整しても、最もよく飲むアルコール飲料がウィスキーである者は、ビールの者に比べて食道がんリスクが高い。
2. 年齢と最もよく飲むアルコール飲料の種類の影響を調整しても、飲酒量が2合以上の者は2合未満の者より食道がんリスクが高い。

Cox比例ハザードモデルと相対危険

- A群の死亡率Pa
- B群の死亡率Pb
- A群に対するB群の死亡の相対危険
  RR=Pb／Pa
- つまり、死亡率が何倍なのかを表す。

- Cox比例ハザードモデルでは、ハザード比（相対危険の一種）を推定可能。

一般に、点推定値（95%信頼区間）の形で表すことが多い。
例： RR = 3.3 (2.0−5.5) ・・・1を含まなければ“有意”

脳梗塞の危険因子（観察期間：1977年～1992年）（コホート研究）

	男性		女性	
	相対危険	(95%信頼区間)	相対危険	(95%信頼区間)
年齢（+10歳）	3.70	(2.30−5.15)	3.91	(2.60−5.89)
平均血圧（+14mmHg）	1.70	(1.16−2.50)	−	
心房細動（有 vs 無）	4.27	(1.25−14.5)	−	
尿タンパク（＋以上 vs 未満）	2.72	(0.92−7.99)	6.85	(1.60−29.1)
眼底異常（KW分類IIa以上 vs 未満）	4.54	(1.05−19.6)	−	
喫煙（20本以上 vs 未満）	2.81	(1.28−6.16)	−	
虚血性心疾患既往歴（有 vs 無）	−		6.35	(1.46−27.7)

年齢（強制投入）、平均血圧、血清総コレステロール、心電図異常、ヘマトクリット、BMI、尿タンパク、尿糖、喫煙、飲酒（2合以上）、身体活動度、虚血性心疾患の既往歴から、STEPWISE法により変数選択を行ったCox比例ハザードモデルによる多変量解析の結果。

解釈例

男性では、年齢、血圧、心房細動の有無、尿タンパク、眼底異常の影響を補正しても、喫煙者は脳梗塞の罹患リスクが高い。