

統計学

- 場所：F204（1月15日を除く）
- 時間：16:10-17:40
- 担当：中澤 港（看護学部）
- 連絡先：内線 1453 <minato@ypu.jp>
- Web-URL: <http://phi.ypu.jp/stat.html>
- テキスト
中澤 港「Rによる統計解析の基礎」ピアソン・エデュケーション，2003年10月15日
- 講義の目標
統計学の考え方の基礎を理解すると同時に，実際に多くのデータを集めて分析する技術の初歩を解説し，身に付けてもらうこと

講義計画

- 10月2日 統計学とは何だろうか？
- 10月9日 統計的な考え方の基礎＝確率と確率分布
- ☆10月16日と23日は講義なし
- 10月30日 データの尺度・データの図示
- 11月6日 データを1つの値にまとめる
- 11月13日 比率に関する推定と検定
- 11月20日 カテゴリ変数2つの分析（1）
- 11月27日 カテゴリ変数2つの分析（2）
- 12月4日 平均値に関する推定と検定
- 12月11日 2群の差に関するノンパラメトリックな検定
- 12月18日 多群間の差を調べる～一元配置分散分析と多重比較
- 1月8日 相関と回帰
- 1月15日 時系列データと間隔データの扱い方（教室注意！）
- 1月22日 一般化線型モデル入門
- 1月29日 高度な分析法についての概説
- 2月5日 実際の分析に際して
- 2月12日 試験（持ち込み可）

統計学の歴史

- 無文字時代は、「家畜や他の財産の帳簿をつけるために原始人が木につけた刻み目」
- マルシャス「ある国およびそこに生きている生命の状態や発展についての、最も完全で、最も根拠のある知識」
- 国家としての人口統計，産業統計
- 英国王立統計協会「人間に関係することがらで、数量で表現することが可能で、一般的な法則を導き出すのに十分なだけ積み重ねられたもの」(1834年)
- 20世紀以降、「不確実性」を無視できなくなった諸科学の礎。 Salsburg, D. “The lady tasting tea”

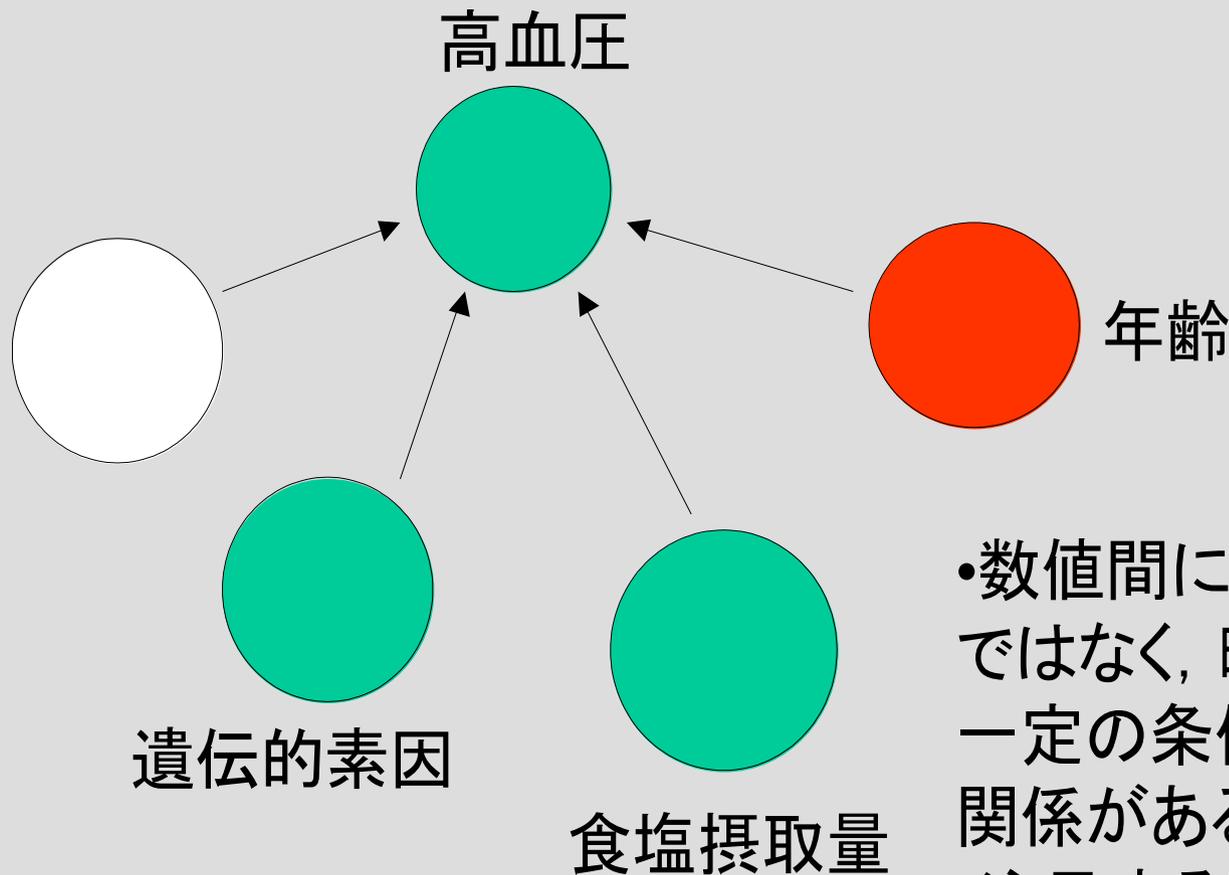
不確実性とランダム

- 世の中のほぼすべての現象は不確実性を含んでいる
- 不確実性を数学的に扱うための方法＝確率の導入～乱数列の利用
- 擬似乱数列のいろいろ
 - 乗算法
 - 線型合同法
 - M 系列乱数
 - Lag のあるフィボナッチ数列
 - メルセンヌツイスター

統計解析の手順

- 目的を明確にする
- 生データをとる
- データ化（エディティング，コーディング，データ入力）
- データの図示（幹葉表示，ヒストグラム等）
- 分布の位置やばらつきの計算
- 作業仮説の明確化
- 仮説検定や区間推定を行う（攪乱要因に配慮し，その影響を制御する必要がある）
- 因果関係についての推論を行う（先行研究の知見なども総合する必要がある）

因果関係と攪乱要因



- 数値間に常に関連があるだけでなく、時間的前後関係など、一定の条件を満たさないと因果関係があるとはいえない
- 注目する要因以外の、結果に影響する要因を攪乱要因(交絡要因)という

統計解析の2つの方針

- デザインに基づく解析
 - 推論の土俵をデータの取り方におく
 - データが、ランダムにデータを取った場合に得られるパタンの1つであると考え、その確率を計算
 - 攪乱要因は層別で制御
 - (例) 並べ替え検定, ログランク検定
- モデルに基づく解析
 - 推論の土俵をモデルにおく
 - 結果の分布を記述するために確率分布を仮定し、その未知パラメータをデータから推定する
 - 攪乱要因は説明変数としてモデルに入れる
 - (例) t検定, 比例ハザードモデル

統計解析の道具

- 実際の統計解析は，コンピュータのソフトを使って行われるのが普通。
 - SAS とか SPSS がメジャーだが高価。 JMP や Statistica などはそれほど高価ではなく，かなりマウスで操作できるので簡便。
 - Excel
 - 覚えておくといろいろ便利だが，ブラックボックスだし，それがないと何も出来ないのでは困る
 - R
 - R-project (<http://www.r-project.org/>) という国際共同研究で開発されているフリーソフト
 - R のインストール： Windows 版では rw1071.EXE をダウンロードして実行
 - コマンドを打つのが基本だが， Rcmdr というメニューもある。