

統計学 第1回 「統計学とは何だろうか？」

統計学の歴史

- ・ 統計学の歴史をずっと過去に遡っていくと、本来は、「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」であった。(ラオ)
- ・ その後「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」(マルシャス)、「国家にとって必要不可欠な人口や経済的な情報の収集」(ウォナコット)など、国家を経営するための基礎資料的な意味合いが強くなった。
- ・ 英語の statistics は、ラテン語で国家を意味する status を語源として 18 世紀半ばにドイツの哲学者アッペンウォールが作った言葉の流用。
- ・ かさばり、雑然とした生データを、解釈をやさしくしたり種々の方策決定に用いるためにまとめ上げる、グラントの生命表やケトレーの度数分布図による発展
- ・ 1834 年、英国王立統計協会設立による「統計学」成立。「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」
- ・ 現在の広い定義としては、「不確実性を考慮した論理的推論」であり、すべての自然科学や社会科学で適用される科学的分析の技術となっている。

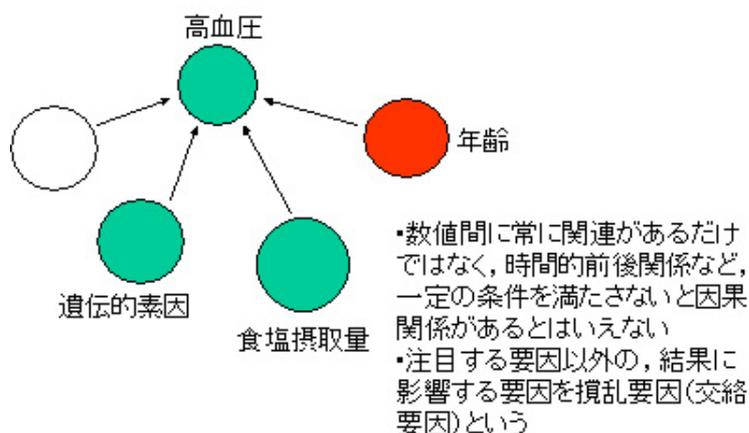
不確実性 = ランダム (乱雑さ)

- ・ 世の中のほぼすべての事象は不確実性を含んでいる
- ・ 素粒子レベルでは物理法則も不確実性を含むし(ある原子核に含まれる電子が存在する確率がゼロでない場所という意味で電子雲は決まるけれども、電子がある瞬間にどこに存在するかということ)は確率的にしかいえない)、遺伝子の発現や社会における個人の行動なども、決して決定されていない
- ・ ランダムな数字の列 = 乱数列 ~ 次の数字が予想できない、意味のないでたらめな数の集まり、例えば、国勢調査区の面積を表す数値の最初と最後を並べたものとか、20 桁の対数表の 15~19 桁目を並べたものとか、袋に入れた 500 個ずつの白ビーズと黒ビーズから、よく混ぜて 1 個ずつビーズを復元抽出(取り出しては戻す)したときの色の列など。線型合同法など、数式を使って生成される擬似乱数列というものがあり、数式がわかれば次の数字は予想できるのだが、見かけ上はでたらめな数の並びに見え、実用上十分なでたらめさをもっているため、コンピュータ上で乱数列が必要な場合は良く使われている(ただし、現在はコンピュータ上でも熱拡散の状態を測って真の乱数を得る拡張ボードが売られていて、利用可能である)。

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(幹葉表示やヒストグラムなど)
- ・ 代表値(分布の位置やばらつきを示す値)の計算
- ・ 作業仮説の明確化(ここで因果関係についての仮説を立てることが多い)
- ・ 仮説検定や区間推定を行う

因果関係と攪乱要因(例示)



統計解析の2つの原理

- ・ デザインに基づいた解析
 - ・ 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, ログランク検定など
- ・ モデルに基づいた解析
 - ・ 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など

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 がメジャーなソフトだが高価。JMP や Visual Stat などはそれほど高価ではなく、また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。そもそも統計ソフトではない。ただし少量のデータ入力には便利(大量のデータ入力にはデータベースソフトを使うか、html のフォームと cgi を組み合わせて入力環境を作ると良い)。
- ・ 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, SetupR.EXE をダウンロードして実行
 - ・ R の文法のエッセンス:(1) 代入が" $<$ "または" $_$ " (2) 範囲指定が" $:$ "でできる (3) 集合は $c()$ で定義できる (4) $q()$ で終了。help() でヘルプ。

講義を進めるためのアンケート(出席確認票の裏に記入)

- A) 数学はどれくらいできますか? 自分に当てはまる最大の番号を教えてください。
1. 数式は全然駄目
 2. 四則演算(+ - \times \div)の式変形ならOK
 3. 指数関数や累乗, べき乗, 対数が出てきてもOK
 4. 微分積分もOK
- B) この講義に期待するものは?(いくつでも)
1. 統計的なものの考え方を知りたい?
 2. 応用事例を知りたい?
 3. 統計解析手法を身につけたい?

アンケートの結果を集計すると...(その場で入力して集計)

- ・ 実は、こういうアンケートを取って結果を集計するというプロセスは、統計解析の1つの事例といえる。