大学の講義

高校とは量的にも質的にもレベルが違う。

量的に見て：　高校の教科書は大学の講義数回分

質的に見て：　入試問題は限られた時間で解けることが前提。
経済学も経営学も実際の世界で起きている
現象を扱う。
現実に起きている問題は、解けるように作られて
いるわけではない。

問題を解く方法を覚える ⇒ 問題を立て
必要な知識を集め
解決の方法を考える

高校頭から大学頭へ進化

大学の講義

講義を一度聴いただけで理解することは困難。
自学自習の必要。

内容を100%理解する必要はない。
理解の程度も段階がある。

講義の（論理的な）流れを把握することが重要。
細部を丸暗記することは不毛。

すべての科目を同じ強度で学習する必要は無い。
究める科目と単位取得を目的にする科目を自分で決める。

この講義（統計解析）

統計的推測について、その考え方を重視
公式をマニュアルとして暗記することに大きな意味はない。

数学は数ⅡBレベル。ただし
数式の複雑さは高校教科書のレベルを超えるかもしれない。
積分はおおよその意味が理解できれば可。
数Ⅲレベルの計算技術は要求しない。
和（Σ）の記号には慣れておくこと。

基礎科目である計量経済Ⅰ（1年後期），計量経済Ⅱ（2年前期）に接続する。この講義の知識は計量経済の理解に必須。

1　統計定推測の考え方

問題 「すべてのカラスは黒い」

世界にカラスが10億羽いたとする。
10億羽全部調べる！

問題を変える 「少なくとも99.9%のカラスは黒い」

9億9900万羽のカラスを調べる！

＜統計的推測＞

1) 仮に「0.1%を超える割合で黒くないカラスが存在する」と仮定
2) 何羽か（n羽とする）のカラスを調査したとして、その時
「1)の仮定の下で黒くないカラスが見つかる確率」を考える。
3) 実際にn羽のカラスを調査する。

4) 2)で計算した確率が非常に高い（たとえば95%，99%）場合に
3)の調査で黒くないカラスが発見されなかったら不自然

→ 1)の仮定が誤り。

＜統計的推測＞

1) 仮に「0.1%を超える割合で黒くないカラスが存在する」と仮定
2) 何羽か（n羽とする）のカラスを調査したとして、その時
「1)の仮定の下で黒くないカラスが見つかる確率」を考える。
3) 実際にn羽のカラスを調査する。
4) 2)で計算した確率が非常に高い（たとえば95%，99%）場合に
3)の調査で黒くないカラスが発見されなかったら不自然

「1)の仮定の下で黒くないカラスが見つかる確率」

「全体でN羽のカラスの集団がいる。このうち0.1%は黒くないカラスである。この集団からn羽をでたらめに捕まえるとした場合、その中に黒くないカラスが含まれる確率はどれだけか。」

黒くないカラスの総数を 0.001N=m とすると、求める確率 p は

$$p = 1 - \left(\frac{N-m}{N}\right)\left(\frac{N-1-m}{N-1}\right)\cdots\left(\frac{N-n+1-m}{N-n+1}\right)$$

黒くないカラスの総数を 0.001N=m とすると、求める確率 p は

$$p = 1 - \left(\frac{N-m}{N} \right) \left(\frac{N-1-m}{N-1} \right) \cdots \left(\frac{N-n+1-m}{N-n+1} \right)$$

p が十分高い確率（ここでは0.95）以上になるように n を選択

$n \geq 2995$ ならば $p \geq 0.95$

よって2995羽のカラスを調べて全部黒かったら
「少なくとも99.9%のカラスは黒い」
という命題は、確率95%（有意水準5%）の確からしさで証明される。

2995羽はカラス全体10億羽の 0.0003%

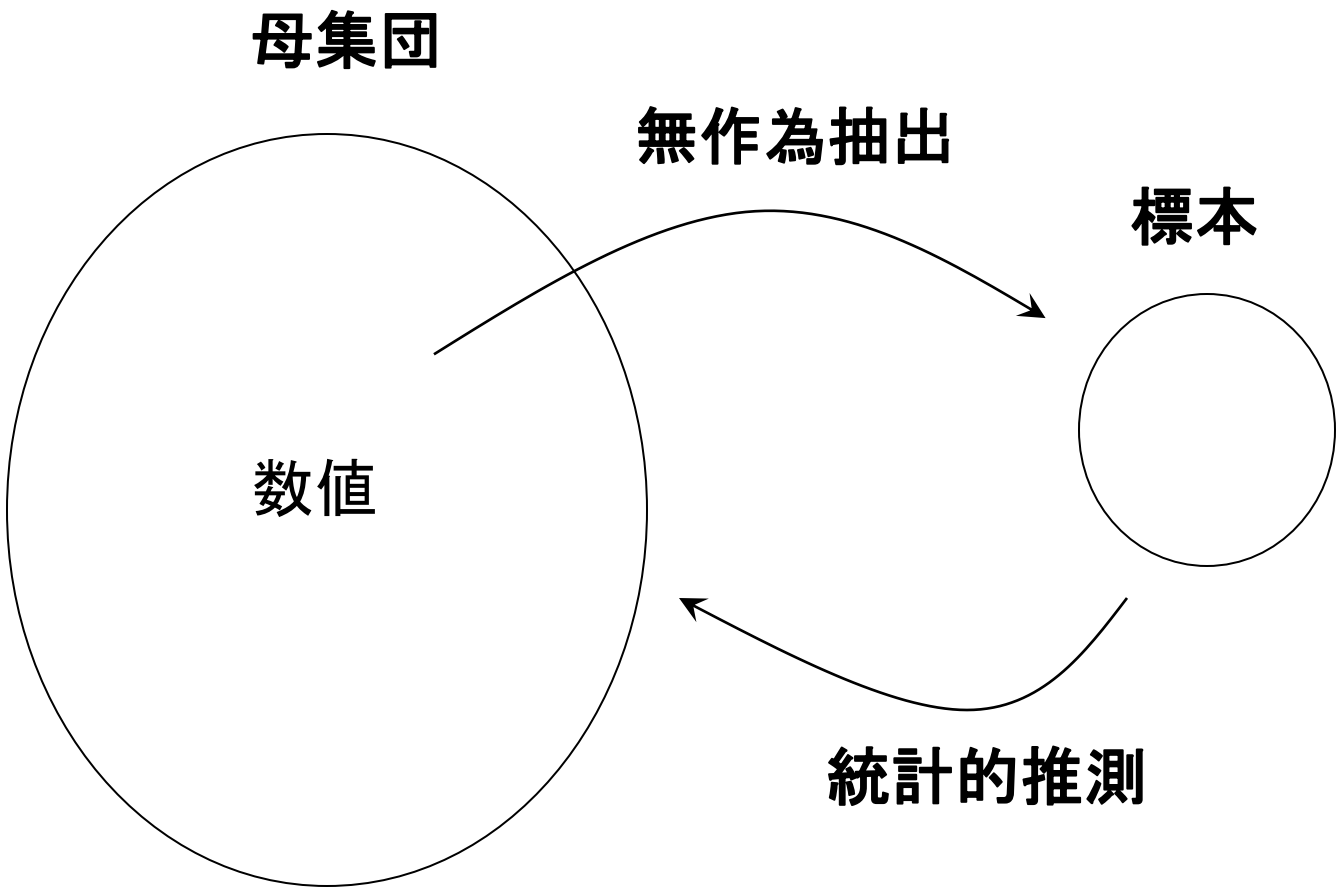
全体の 0.0003% を調べただけで意味のある結論が出せるというところに、統計的推測のパワーがある。

そのパワーの源泉は2995羽がカラス全体を代表する標本になっている、というところにある。

教科書：　木下宗七編 「入門統計学（新版）」
有斐閣　有斐閣ブックス

参考書：　森棟公夫 「統計学入門」 新世社　新経済学ライブラリ
宮川公男 「基本統計学（新版）」 有斐閣

2　母集団と標本



分析者が関心を持つ対象全体の集合が母集団
標本は母集団の部分集合

無作為抽出（random sampling）

標本は母集団の性質を正しく反映しなければならない。

そのために無作為抽出（でたらめに取り出すこと）を行う。

でたらめに取り出すことは難しい。

- 家計調査
- 1932　アメリカ大統領選挙の電話調査

RDD法（random digit dialing）
インターネット調査

1936年　アメリカ大統領選挙

この選挙は「世論調査」が行われたことで有名である。
これは「リテラリー・ダイジェスト」誌が読者および潜在的読者一千万人に郵送し、200万通を回収した質問表に基づくものだった。
最近5回の選挙での勝者を正確に予測した「リテラリー・ダイジェスト」誌は、その10月31日号で、ランドンが選挙人表370票を獲得して勝利すると宣言した。
この誤りの原因は不適切なサンプリングによるものと考えられている。
「リテラリー・ダイジェスト」誌の読者は共和党支持者が多かった。

同じ年、科学的な調査を始めた広告社役員ジョージ・ギャラップが、5,000人の無作為抽出に基づく調査でルーズベルトの当選を予測した。ギャラップの正確な予測は、ジャーナリストや実際の政治家にとって、世論調査を重要な要素にすることになった。

出典: フリー百科事典『ウィキペディア（Wikipedia）』
(http://ja.wikipedia.org/wiki/1936年アメリカ合衆国大統領選挙)

問題のある標本

1. 抽出に偏りがある。
2. 対応する母集団が不明。
3. サイズが過小。
4. 回答率が低い。

社会調査： 参考文献

田村秀 「データの罠　世論はこうしてつくられる」 集英社新書

谷岡一郎 「『社会調査』のウソ　リサーチ・リテラシーのすすめ」
文春新書

標本: 度数分布表

階級値	度数	相対度数
d_1	f_1	$r_1=f_1/n$
d_2	f_2	$r_2=f_2/n$
d_3	f_3	$r_3=f_3/n$
⋮	⋮	⋮
d_K	f_K	$r_K=f_K/n$
計	n	1

階級値は階級の中央値

標本： ヒストグラム

面積が相対度数
を表す。

