

統計学メモ

@mathnuko

2020/02/12 ~

目 次

第 I 部	統計学	3
第 1 章	統計学の目的	4
1.1	統計学とは何のための学問か？	4
1.1.1	「データ」とは？	4
1.1.2	統計学とは？	4
1.2	統計学の体系：記述統計学と推測統計学	5
1.3	記述統計学，推測統計学の基本的な流れ	6
1.3.1	データ分析における問題設定	6
1.3.2	データ分析の流れ	7
1.4	補足 データの解釈：「定性的な解釈」と「定量的な解釈」	7
第 2 章	記述統計学	8
第 3 章	推測統計学	9
第 II 部	数理統計	10
第 4 章	正規分布の導出	11
第 III 部	日本統計学会 - 統計検定に関するまとめ	12
第 5 章	統計検定の情報の整理	13
5.1	統計検定 2 級	13
5.1.1	試験要旨	13
5.2	統計検定 1 級	14
5.2.1	試験要旨	14

はじめに

この文書は統計学のポイントをまとめるためのノートである.¹⁾hoge

¹⁾脚注サンプル

第I部
統計学

第 1 章 統計学の目的

まず、「統計学とは何のための学問か」についてまとめる。目的を理解した上で統計学の体系について説明する。

1.1 統計学とは何のための学問か？

1.1.1 「データ」とは？

統計学は「データの要約，解釈を行うための学問」であるが，そもそも「データ」とは何かを定義しておく。ざっくり言うと「データ = 事実，資料」である。

- データ … 立論・計算の基礎となる，既知のあるいは認容された事実，資料。

自然科学の分野では，しばしば「データ」，「事実」は，「現象（研究対象）を観測した際に得られる数値」，つまり「数値データ」を指すことが多い¹⁾。

また，「データ」は数値データに限らず，文字，文書なども立派な「データ」である。数値データではないデータを統計学で扱う場合には，前処理として数値データへの変換などが行われた後に利用される。

1.1.2 統計学とは？

「統計学とはどのような学問か」という問いに対するいくつかの説明を挙げておく。

- 数値データの要約，解釈を行う上での理論的根拠を提供するための学問
- （赤本²⁾の説明を要約）数値データをどのように分析し，どのような判断を下したら良いかを論ずるための学問
- 対象とする集団，現象の数値データからその性質，法則性を導き出すための学問

¹⁾数値データとしては個数，長さ，体積，重さ，特定の事柄が起こった回数，時刻，続いた時間の長さなどがある。

²⁾統計学入門（東京大学出版）

1.2 統計学の体系：記述統計学と推測統計学

「データに対する解釈を与える」という統計学の目的を達成するためにはどうすれば良いか？

ある現象の法則性を導くためには、まず、現象に関するデータを観測、取得する必要がある。ここでは、「ある現象に関するデータを観測・取得済み」の状態を考える。それらのデータから現象の法則性を導き出すためには、以下の 2 つの手順が考えられる。

- データから現象の法則性を道きび出すための方法
 - 手元のデータを丹念に調べ、その特徴、性質を調べ、現象の規則性、法則を見出す。
 - 手元のデータを用い、「論理性のある」推測により現象の規則性、法則を見出す。

上記手順はどちらも「統計学」が扱う範疇であるが、これらはそのまま統計学の分類に対応する。前者は「記述統計学」³⁾、後者は「推測統計学」⁴⁾と呼ばれる学問に分類される。ざっくり説明すると以下のようになる。

- 統計学
 - 記述統計学 … 手元のデータの説明
 - 推測統計学 … 手元のデータから推測⁵⁾

記述統計学と推測統計学の関係が非常に分かりやすく描かれているのが以下の図。後に出てくる統計学の用語が挙げられているが、ここでは雰囲気を読む。

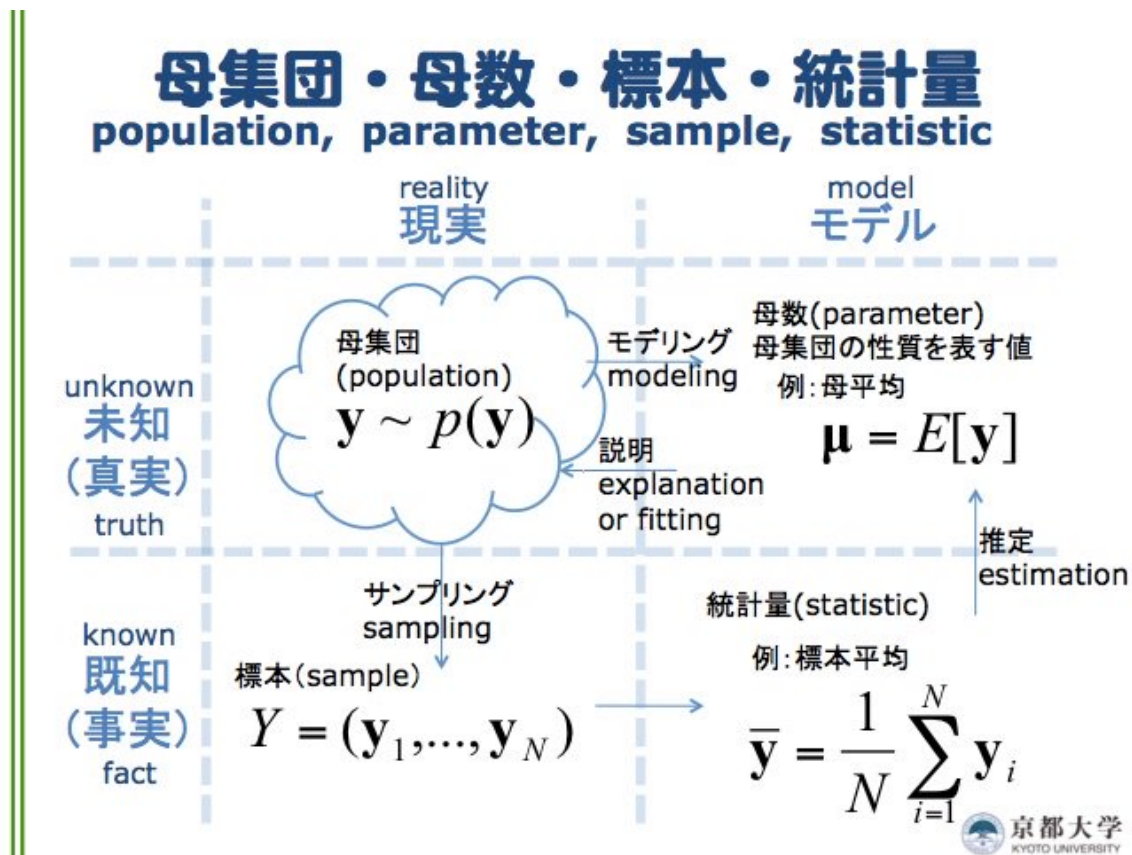


図 1.1: 統計学の体系 (hoge より引用)

³⁾記述統計学, descriptive statistics

⁴⁾推測統計学, inferential statistics, inductive statistics

⁵⁾正確には「手元のデータから『母集団の性質』の推測」

1.3 記述統計学，推測統計学の基本的な流れ

記述統計学，推測統計学の手法等を用いてデータ分析を行う際の基本的な流れを説明する。

1.3.1 データ分析における問題設定

まず，データ分析の際に必ず行わなければならないのが「問題設定」である。「○○という現象を理解したい」など，データ分析を行った結果何を得たいかという目的が必ず設定されているべきである。

また，統計学はあくまでも「データに対する解釈を与える」ための手段の 1 つである，ということは認識しておくべきである。手段が目的となってはならない。つまり，「統計学の○○という手法を使って△△という現象を理解したい」，「機械学習の△△という手法を使って何かできないか」ではなく，「△△という現象を理解したい → その現象から得たデータに適した統計学（もしくは機械学習）の○○という手法を用いる」という順番であるべきである⁶⁾。

まとめると，データ分析を行う際は，闇雲に統計学の手法を用いるのではなく，

1. 「どんな現象を理解したいか」という問題設定
2. 必要なデータの取得
3. データの解釈，データから推測（必要であれば統計学の手法を用いる）

という手順を踏むべきである。「問題設定 → データに適した統計学の手法の選択」という流れを意識したい。問題設定には例えば以下のようなものが考えられる。

- 問題設定の例

- － ある現象のデータからその傾向を見たい（記述統計学）
- － ○○という仮説を検証したい（記述統計学，推測統計学）
- － 手元のデータから未来のデータを予測したい（推測統計学）

⁶⁾現実のデータ分析でこのような綺麗な流れがあるかは分からないが，少なくとも問題設定は必ず為されるべきである

1.3.2 データ分析の流れ

前節を踏まえ、記述統計学、推測統計学の手法等を用いてデータ分析を行う際の基本的な流れをざっくりまとめてみる。両者の最終的なアウトプットの違いに着目すると良い。

記述統計学の流れ

1. 問題設定、必要なデータを決める
2. データの取得
3. データの要約
4. データの解釈

推測統計学の流れ

1. 問題設定、必要なデータを決める
2. データの取得
3. データから母集団の性質に当てはまる数理モデルを構築⁷⁾
4. 数理モデルを使った推定・推測・推論（データを生成する未知の真の確率分布）
5. 数理モデルを使った推定・推測・推論の妥当性を考察・検証

- 記述統計学のアウトプット：対象となる現象，集団の性質，特徴の説明，解釈
- 推測統計学のアウトプット：対象となる現象，集団の性質，特徴の予測．例えば，「予測を行うための数理モデル」がアウトプットになる．この場合，「数理モデルの構築 → 検証」まで行って初めて意味を成す．

データ分析においては，上記の流れは上から下に綺麗に流れていくものではなく，上記の流れを回すである．必要に応じて再度データ収集を行ったり，数理モデルの構築 → 検証のサイクルを回したり，といった

1.4 補足 データの解釈：「定性的な解釈」と「定量的な解釈」

統計学，または統計学だけでなくもっと広い意味で「データに対する解釈」を行う際に，解釈の結果として何らか（統計学で使われる指標など）の数値に落とし込むことが当たり前のように感じられるが，必ずしも数値に落とし込むことだけが「データに対する解釈」ではない⁸⁾．データに対する特徴・性質を「言葉」で表すこともできる．

⁸⁾ 現実には統計学が使われる場面（データ分析など）では，アウトプットとして何らかの数値（指標）に落とし込むことが求められるのがほとんどだとは思いますが...

第 2 章 記述統計学

第 3 章 推測統計学

第II部

数理統計

第 4 章 正規分布の導出

第III部

日本統計学会 - 統計検定に関するまとめ

第 5 章 統計検定の情報の整理

5.1 統計検定 2 級

5.1.1 試験要旨

- 公式 HP : 統計検定 2 級
- 出題範囲

5.2 統計検定 1 級

5.2.1 試験要旨

- 公式 HP : 統計検定 1 級
- 出題範囲