

演習7 統計の基礎

担当: 加藤由紀

0 連絡事項: 演習実施日より前にしなければいけないこと

- 事前準備 1. 演習7のウェブページを読み, 必要な資料をダウンロードする.
演習7のページ <http://www.damp.tottori-u.ac.jp/~kato/KBJ/> (♡)
- 事前準備 2. 演習実施日の前の週に予習レポートを提出する. 予習レポートの内容・締切日時・提出方法等については, ウェブページ(♡)を見よ.
- 事前準備 3. 演習当日には次の物が必要である. これらを準備する.
 - 教科書・参考書: 次の本を教科書として使う.
薩摩順吉, 確率・統計, 岩波書店. (1年次講義「確率統計学」の教科書).
本演習では, 主に教科書 §6-1~6-4 の内容を扱う. 教科書を持っていない人は, 統計学の本を持参すること.
 - 次の計算をするための道具 (数表, 電卓, PC など).
 - * 平方根の値を求める (例えば $\sqrt{12.34}$).
 - * $\gamma = 0.90, 0.95, 0.99$ に対して, $\int_{-z}^z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\zeta^2\right) d\zeta = \gamma$ を満たす z の値を求める. 教科書を持っている人は, 巻末附表2の数表を利用できる.
 - * 10-20個程度のデータについて, 平均と分散を求める. 紙と鉛筆でも計算できるが, 楽をしたい人はPCを使ってもかまわない.
 - A4のレポート用紙.

1 概要

統計学は理工系分野のみならず社会の広い分野で利用されている. 本演習では, その基礎を学ぶため, 推定と検定の初歩的問題に取り組む. 推定・検定の目的は, 大雑把に言うと母集団の情報を予想することだ (図 1-3 参照). 予想には当たりはずれがある. 統計学は, まったくのあてずっぽうで予想するよりも当たりの確率が高くなるような予想方法を教えてくれる.

演習の時間には, §3の例題と似たような問題を解く. 問題は演習当日に配布される. 時間の余った人にはより発展的な問題が与えられる.

2 成績評価方法

原則として, 予習レポート10点 + 演習レポート90点とする.

3 例題

演習当日には以下の例題と似たような問題を解く。例題の解答解説はウェブページ(♡)にある。演習当日までに解説をよく読み、予習しておくこと。予習の便宜のため、教科書に関連情報が載っている箇所を以下の[かっこ内]に示す。

ある工場から 100 万個の部品を買った。納品書によると、部品の長さの平均値は 100 mm である。納入された部品が納品書の通りかどうか調べたい。

納入された部品の長さを知りたいければ、部品の長さを測ればよい。しかし 100 万個の部品の長さを測るなんてことはやりたくない。そこで、100 万個の中から 7 個だけ取り出して長さを調べた。その結果は次の通りであった。

抜き取り検査の結果 (単位は mm)

111, 98, 99, 102, 107, 104, 97

このデータから、納入された部品の長さについて何が言えるか考えよう。

納入された部品と検査した部品の関係について

問題設定を統計学の言葉で書きなおしておく。

100 万個の部品の長さ ($Y_1, Y_2, \dots, Y_{1000000}$) が母集団である。母集団の分布を特徴づける量 (母集団の平均値や分散など) を母数と呼ぶ。7 個の部品の長さ (X_1, X_2, \dots, X_7) は、母集団から抽出した標本である。標本を調べることによって母集団の情報を得ようというのが、統計学の目的だ (図 1)。 [§5-1 母集団と標本]

母数は 100 万個のデータから決まる値だ。それを 7 個のデータから決めるなんてことは、そもそも不可能だ。それでもなんとかして母数がいくらになるかをあててやろうという訳である。

なお、この例題では問題を簡単にしたいので、母集団は正規分布に従うものと仮定する。

例題 A 標本の統計量を計算する [§5-2 標本の整理, 例 3, 問題 5-2 の 2]
検査した 7 個の部品の長さについて、平均 \bar{X} , 分散 S^2 , 標準偏差 S を求めよ。

例題 B1 点推定 [§6-1 点推定, 例題 6.1]

納入された部品の長さを推定せよ。推定の方法は複数あるが、本演習では「不偏推定量」を使うことにする。

- (a) 納入された 100 万個の部品の長さについて、平均 μ を推定せよ。
- (b) 納入された 100 万個の部品の長さについて、分散 σ^2 を推定せよ。

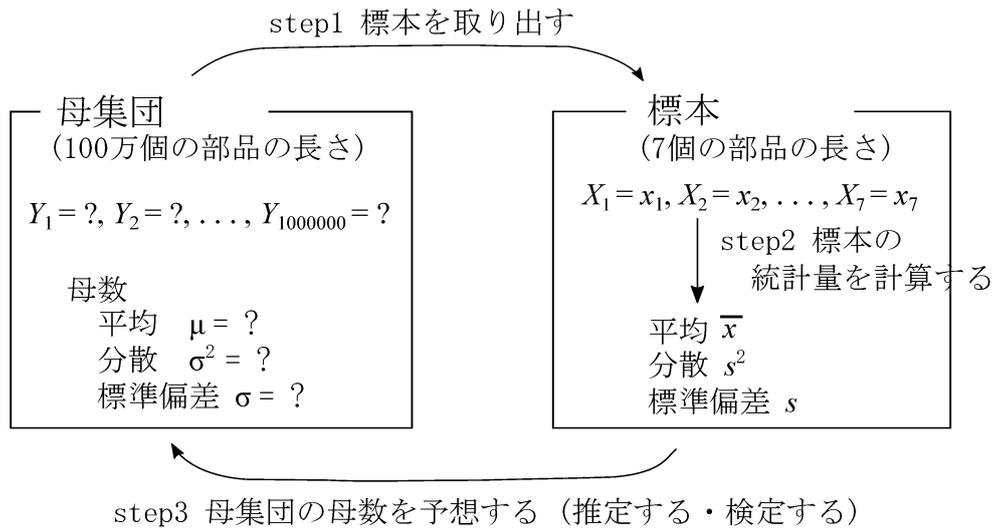


図 1: 例題において、母集団の情報(母数)を予想する際の手順.

例題 B2 区間推定

[§6-2 区間推定. 例題 6.3]

納入された 100 万個の部品の長さについて、平均 μ は未知だが、分散だけは $\sigma^2 = 9.0\text{mm}^2$ であることがわかっているものとする。このとき、平均 μ の信頼区間を推定せよ。

例題 C 検定

[§6-3 仮説と検定. §6-4 母数の検定, 例題 6.7]

納品書には部品の長さが平均 100mm と書いてあるが、抜き取り検査の結果は平均 $\bar{X} = 102.6\text{mm}$ だった。このとき、100 万個の部品の長さの平均値は納品書の通りと考えてよいだろうか？あるいは納品書の記述とは違うと考えるべきだろうか？この問いに答えるために、次の仮説を検定せよ。

仮説 H_0 : 納入された部品の平均長さは $\mu = 100\text{mm}$ である。

対立仮説は $H_1: \mu \neq 100\text{mm}$ とする。仮説 H_0 が棄却されれば「納品書の記述と違う」と言える。簡単のために、母集団の分散は既知とする (例題 B2 と同様に $\sigma^2 = 9.0\text{mm}^2$)。

4 おまけの情報: 推定・検定の応用例

推定・検定の応用例を図 2,3 に示す。

図 2 は物理実験の例である。1 年次の物理学実験演習で、測定値の不確かさを求めた経験があることと思う。あれは、例題 B2 の区間推定とだいたい同じことをやっている。実験で得られる測定値には誤差が含まれていて、個々の測定値にばらつきがある。そこで、無限個の測定値という仮想的な母集団を考えて、この母集団の平均 μ が求める値 X であるとみなす。 X の値を、有限個の測定値 (これが標本である) から推定する。

図3は、選挙の結果を予想する例である。例えば衆議院議員選挙の投票日の夜、まだ開票が終わっていないのに、テレビ放送局は当選確実の速報を出し、選挙事務所ではバンザイー!とかやっている。開票が半分も終わっていないのになぜ当選確実と言えるのか不思議に思ったことはないだろうか? あれは、大雑把に言うと、一部の人の投票結果(これが標本)に基づいて、全員の投票結果(これが母集団)を予想している。

統計学が使われる場面は他にもいろいろある。各自で調べてみるとよいだろう。

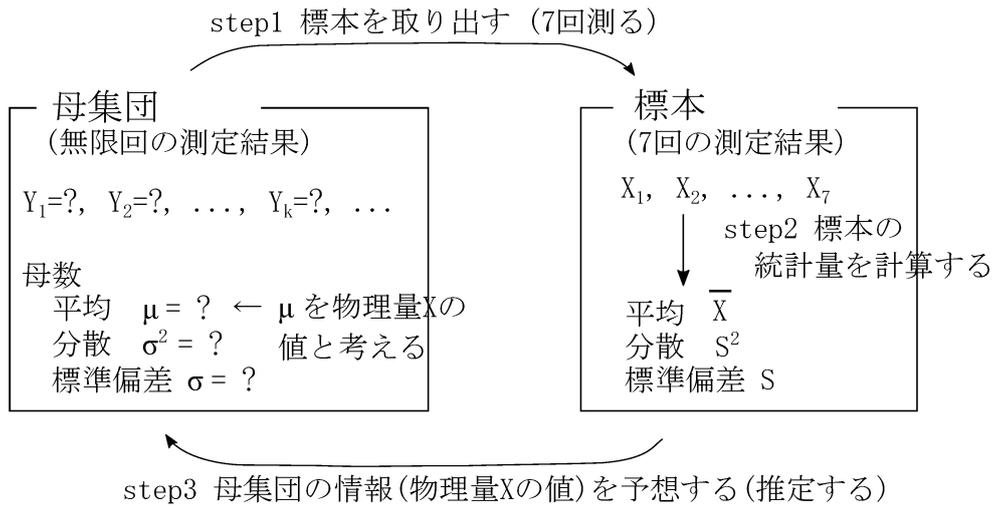


図 2: 統計学の応用例 1. 実験で物理量 X の値を求める.

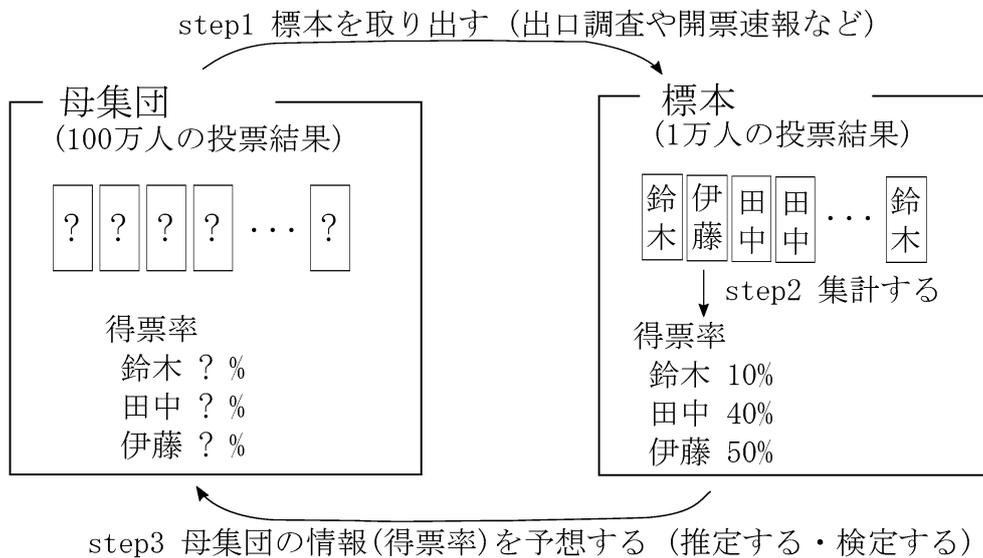


図 3: 統計学の応用例 2. 議員を選ぶ選挙で当選確実の判定をする.