

不適切な統計的推論

川田恵介

2025-07-13

1 不適切な推論

1.1 復習: 統計的推論

- 母集団の特徴について、推論(Inference)する
 - ▶ 前提 (ランダムサンプリングなど)を満たせば、結論 (信頼区間など)は概ね正しい
 - 典型的には、「ランダムサンプリング + 因果効果を識別する仮定、が満たさられるならば、平均因果効果は概ね 信頼区間 (5, 15) の範囲である」
- 労働経済学における実証研究における必須テクニック

1.2 不適切な推論

- 適切な推論ができると「錯覚してしまう」状況に注意が必要
- 本講義では、以下の典型例を紹介
 - ▶ 複数の推定目標(多重検定問題)
 - ▶ Y による分割

2 多重検定問題

2.1 探索的研究目標

- 複数の Y や D の中から、“顕著”な関係性を探したい
- ここまで議論してきた統計的推論の前提が崩れていることに注意
 - ▶ Imbens (2021), List et al. (2019)

2.2 労働経済学における研究目標例

- 複雑な現象/政策介入がなされており、大量の変数間の間での (因果的)関係性を理解したい
 - ▶ 職務内容のうち、賃金と相関している項目はあるか？

- ▶ ハローワーク内でのさまざまな介入の中で、有効なものはあるか？
- ここまでの議論は、関心のある Y と D は一つであることが前提
 - ▶ X は無数にあっても良い

2.3 多重検定問題

- Y と全く関係性がない/全ての負の関係性がある D から、正の関係性を持つものを、誤って発見する確率が 想定よりも課題になる
 - ▶ Y や D が”無限大”あれば、“必ず”発見できる
- “If you torture the data long enough, it will confess.”
 - ▶ Ronald H. Coase

2.4 悪例:

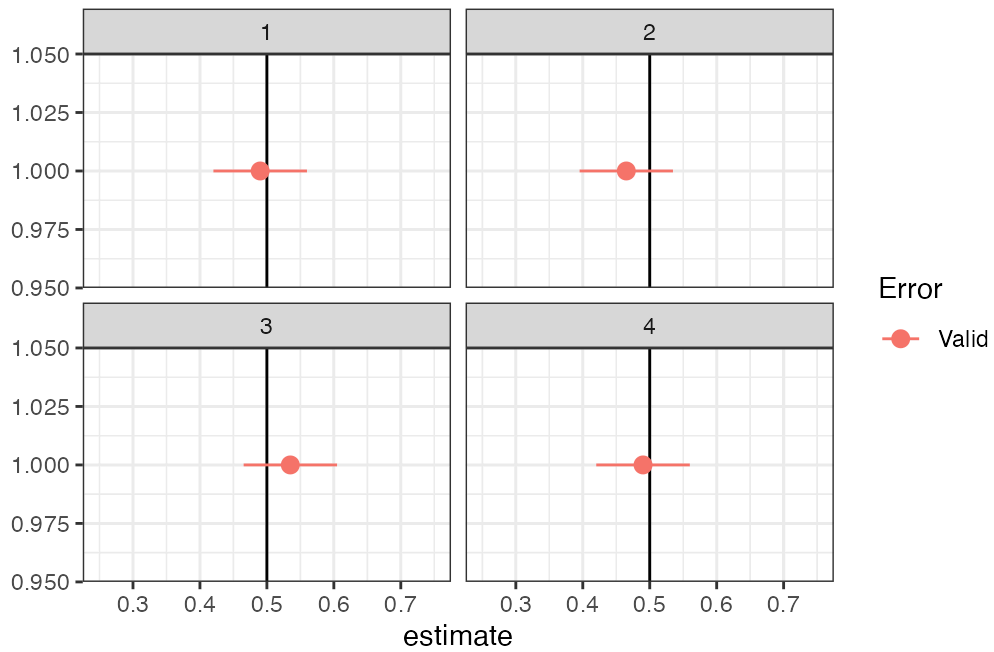
- 新聞を読む人/読まない人を比較し、どのような違いがあるのか明らかにしたい
- 「独自に収集したデータをもとに、新聞を読む人と読まない人を比較した。結果読む人の方が、“友達が多い”という結果が出た。解釈としては、、、」
 - ▶ どのくらいの特徴 ($= Y$) について推定したのか、一切記載がない
- 可能な限り事前に推定対象を絞り、推定値はすべて公表すべき

2.5 数値例

- あるプロスポーツにおいて、コイントスで先行後攻を決めている
 - ▶ **研究目標:** 「一部のチームが不正コインを使用しており、コイントスの結果が 50:50 になっていないのではないか？」
 - 実際には不正はないとする
- 各チームのコインを回収し、実際にコイントスを繰り返してテスト

2.6 Recap: 95 % 信頼区間

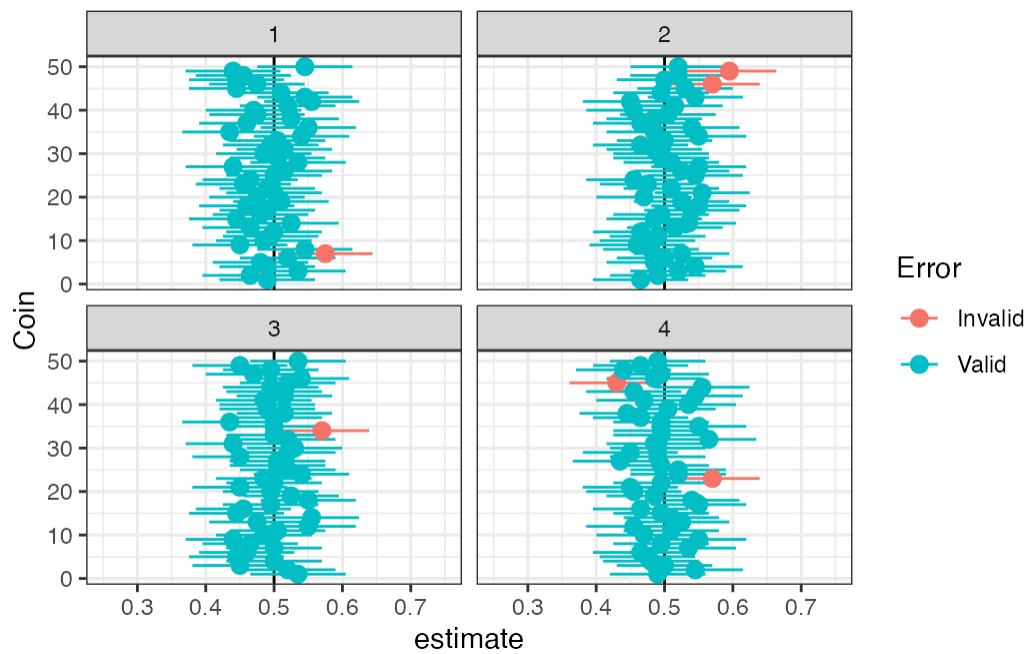
- ある一つのコインについて、4 名の研究者がコイントスを 200 回行い、表の割合、および信頼区間を計算した
 - ▶ ほとんど(例: 95%) の研究者について、信頼区間は真の値(0.5)を含む



2.7 例. 多重検定

- 50 個のコインについて、推定した

- 全員がチートコインを誤って発見した



2.8 直感

- 誤検出された”明確な”結果 = 目立ちたい研究者にとって、当たりくじ (5% の確率で引ける)
 - ▶ “どんなに当たりくじでも、無限に引けば当たる”
 - 50 回引けば、92.3% の確率で引ける
 - ▶ 統計的推論 (仮定が正しければ、結論は概ね正しい) が崩壊する
- 最低限の対策: 何回引いた(パラメタを推定した)のか、必ず明示する

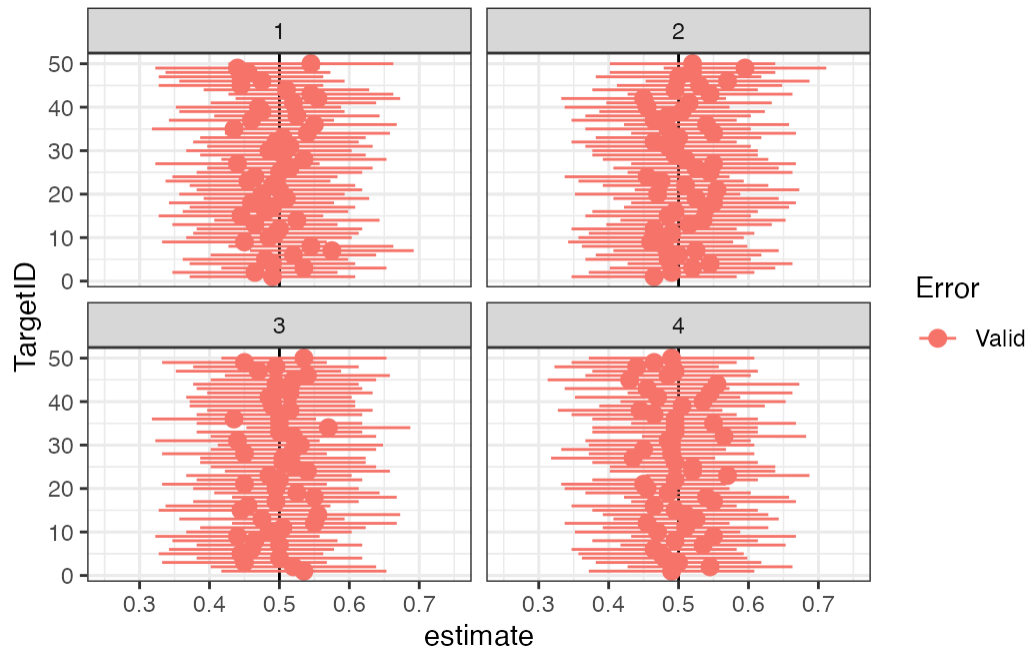
2.9 Family-wise confidence interval

- 複数の信頼区間について、一つ以上 ミスを犯す確率を一定以下(5%)に抑える
 - ▶ \Leftrightarrow 通常の信頼区間 = 一つのパラメタについてミスを犯す確率を抑える
- Study-wise confidence interval と呼ばれる
- 信頼区間を適切に拡張する必要がある
 - ▶ 広げ方について様々な提案

2.10 Bonferroni 法

- かなり保守的な修正法
- k 個のパラメタについて、一つ以上ミスを犯す確率を 5% 以下に抑えたいのであれば、 $\frac{5}{k}\%$ 信頼区間を計算
- パラメタが増えると無限に広がる
 - ▶ 他の方法については、Introduction to statistical learning 13 章 参照

2.11 例. Bonferoni 法



3 Yによる分割

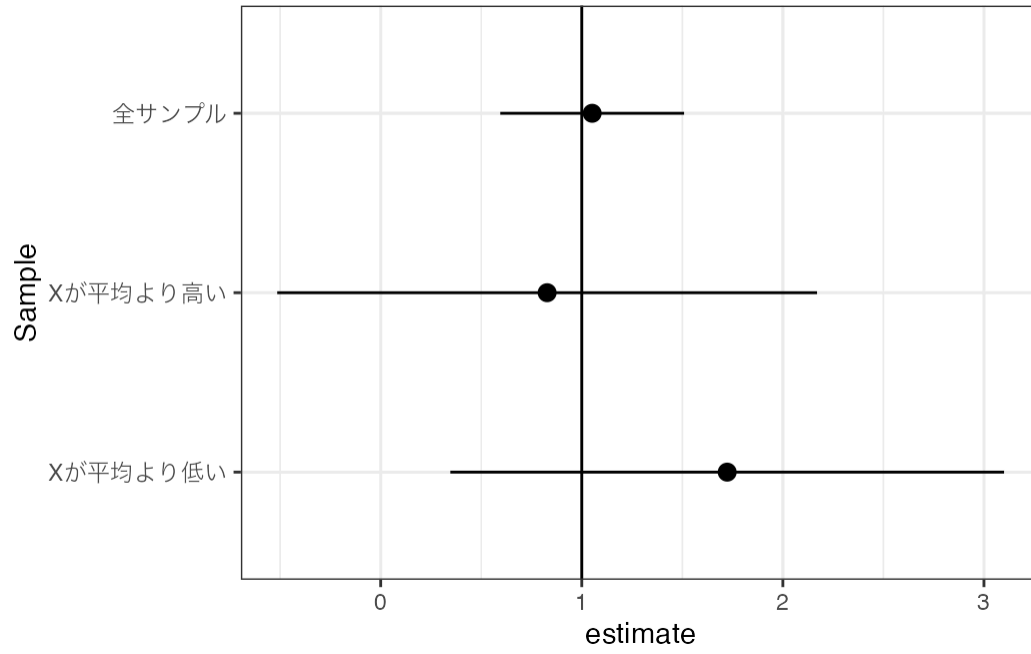
3.1 嚴重注意

- 労働経済学において、しばしば Y でサンプルを分割した、サブサンプル分析をしたくなる”誘惑”に駆られる
 - ▶ 例となる研究目標: 立場の弱い (Y が平均よりも低い) グループにおいて、介入は効果を持つのか?
 - Y が平均よりも低い/高い事例に 2 分割し、サブグループごとに $Y \sim D$ を推定する
 - 通常の統計的推論は不可能になる
 - D や X による分解は問題ない
- Abadie et al. (2018)

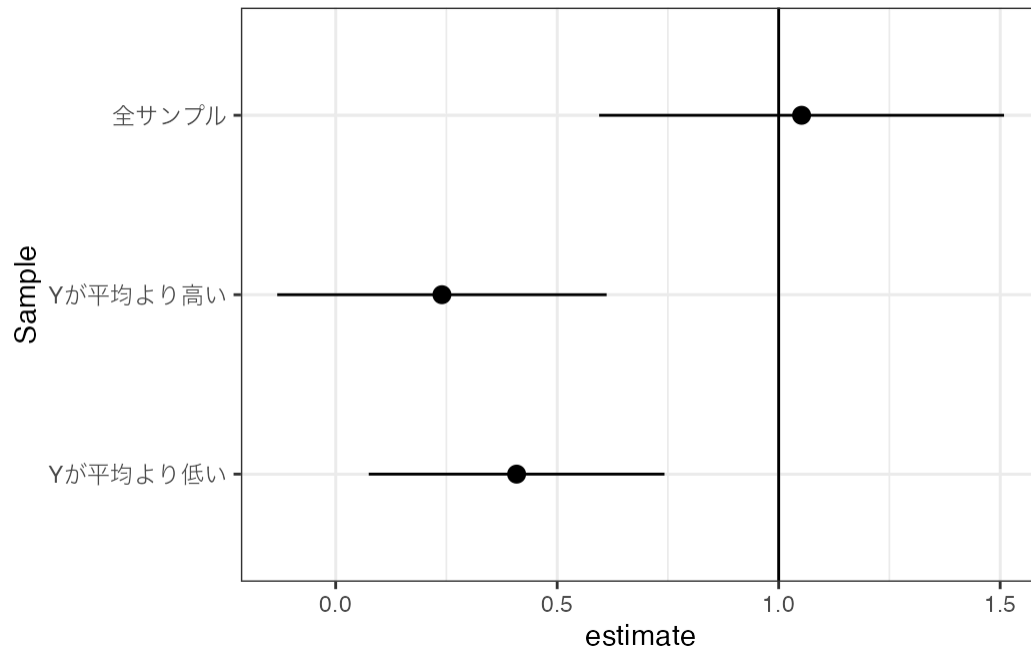
3.2 数値例

- $Y = D + u$
 - ▶ u は $-20, 20$ までの一様分布

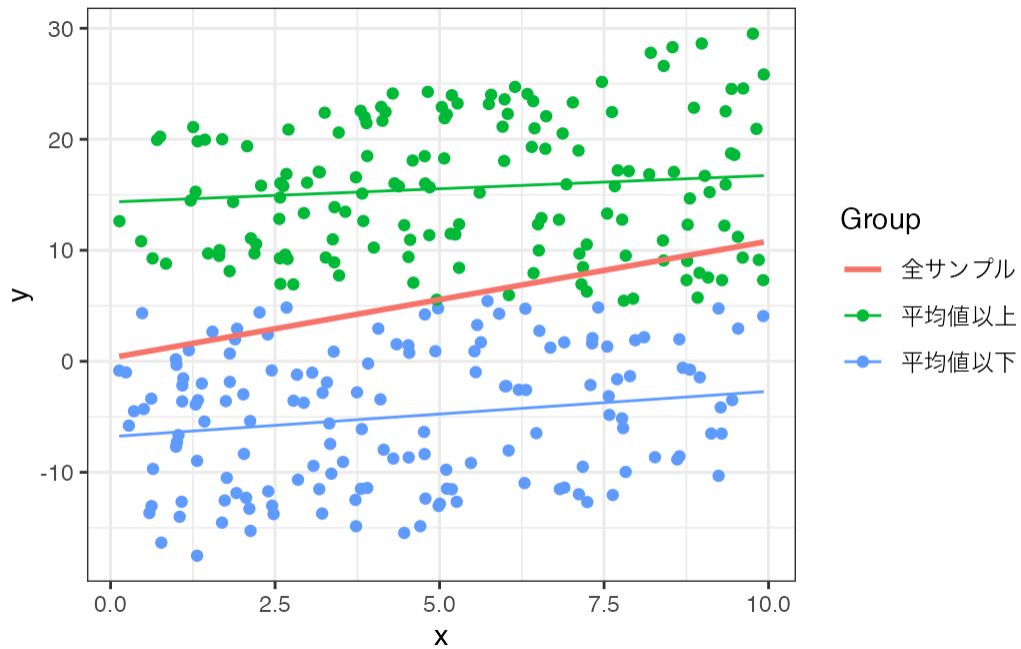
3.3 例. X による分割



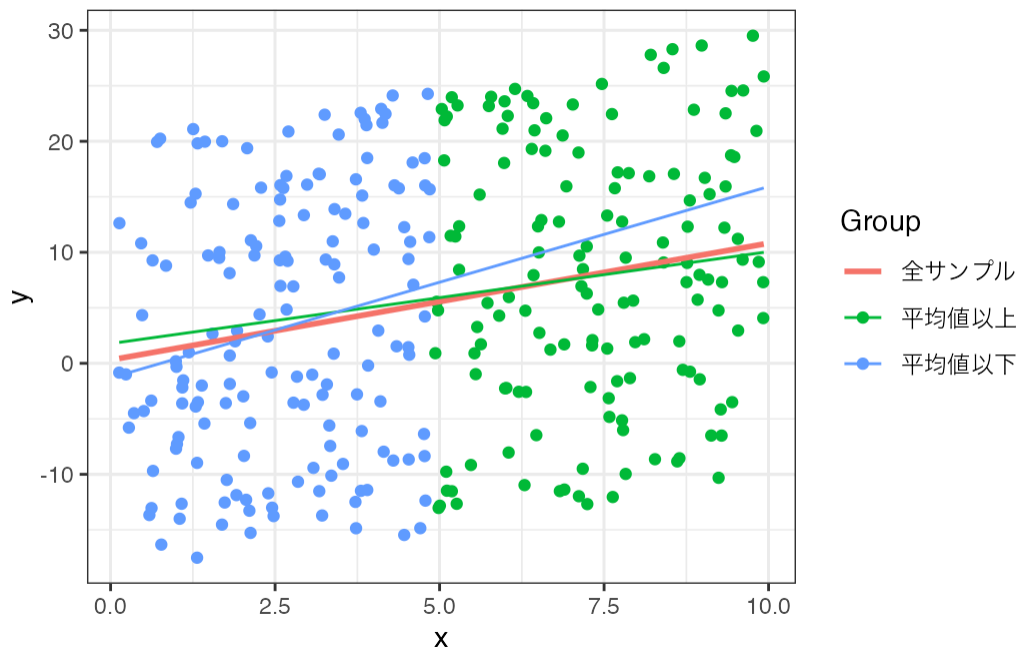
3.4 例. Y による分割



3.5 例. Y による分割



3.6 例. X による分割



3.7 Reference

Bibliography

- Abadie, A., Chingos, M. M., & West, M. R. (2018). Endogenous stratification in randomized experiments. *Review of Economics and Statistics*, 100(4), 567–580.
- Imbens, G. W. (2021). Statistical significance, p-values, and the reporting of uncertainty. *Journal of Economic Perspectives*, 35(3), 157–174.
- List, J. A., Shaikh, A. M., & Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, 22, 773–793.