経済統計分析 8
推定, 検定

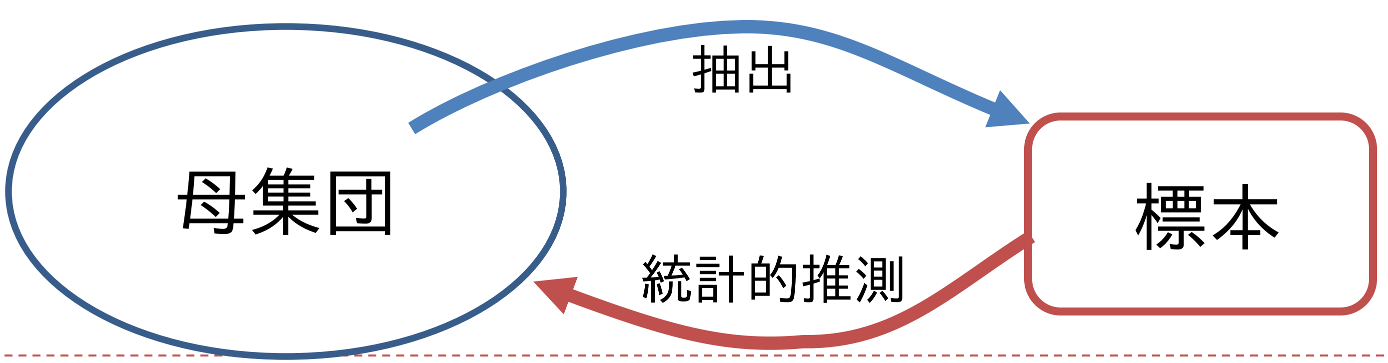
今日のおはなし.

- 統計的推測 statistical inference のひとつ
  - 推定の考え方
  - 推定量が持つべき望ましい性質
  - 平均についての点推定
  - 区間推定, 信頼区間の形成
  - 検定
  - 政策評価

- 今日のタネ
  - 吉田耕作. 2006. 直感的統計学. 日経BP.
  - 中村隆英ほか. 1984. 統計入門. 東大出版会.

なにができるようになりたいか

- ある変数が他の変数に与える効果の大きさの数量化
  - 確率論的な言葉遣いでは「同時分布の特性値の値を知りたい」
- 母集団の一部を取り出して，その部分だけから全体を推測
  - サンプル（sample）を抽出して，母数を推測
  - 標本平均については，中心極限定理が成り立つ
- 中心極限定理を使って，母平均を推測できないか?
  - 中心極限定理によれば，サンプルが十分に大きく，無作為抽出であれば，標本平均は母平均の「すぐそば」にある



推定量としての標本平均

- 標本平均の性質
  - 中心極限定理によれば, サンプルが十分に大きく, 無作為抽出であれば, かなり一般的な条件のもとで, 母集団の分布の形状によらず, 「真の平均（母平均）」の近くにある確率はかなり高い
  - ということは, 標本平均を母平均の「推定量 estimator 」としても, それほど間違った値にはならないはず
  - ただし, 標本平均は確率変数だから, 「真の平均」と一致するとは限らない
- 点推定量としての標本平均
  - 母平均の値を「推定 estimate 」するために標本平均を使うのはもっともらしい
  - 標本平均は母平均に確率収束するので, 「標本平均は母平均の一致推定量 consistent estimator だ」という.

推定論

- 推定 estimation とは
  - 標本を用いて, 母集団の未知の特性値のbest guessを計算すること
  - その計算方法のことを推定量 estimator という
  - その計算結果としての値のことを推定値 estimate という
  - 母平均だけでなく, 母分散や共分散なども推定できる

- 推定量は確率変数
  - 推定量は標本から計算できるから, 確率変数
  - 標本が変われば, 推定値は変化する
  - だから, 真のパラメタ（母集団の特性値）が得られるわけではない（ベイズ的な考え方はかなり異なる）
  - よって, 推定値が真のパラメタからどれくらい離れるか, が評価できるとうれしい.

望ましい推定量

- 推定量を「定義」するのは自由
  - 標本平均だけが平均の推定量ではない
  - なんらかの加重平均や切り落し平均を推定量にしてもよい
  - しかし, best guessであってほしい
- 一致性（consistency）
  - 推定量が真の値に確率収束する
  - 標本の大きさが大きくなるほど, 真の値の周りの区間に入る確率が1へ近づく
  - 例：標本平均は母平均の一致推定量
- 不偏性（unbiasedness）
  - 推定量の平均（期待値）が真の値に等しい
  - 例：標本平均は母平均の不偏推定量
  - 不偏推定量のうち, 分散が小さいものを「有効推定量」と呼ぶ

一致性と不偏性

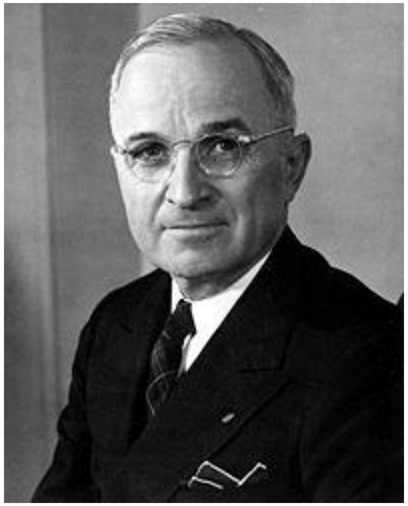
- 標本平均は一致性と不偏性を持つ.
- 一致性のほうが望ましいものの, 大きなサンプルが必要

- 一致性があっても不偏性のない推定量, というのもある

無作為抽出でしょうか?

- 推定量の一致性
  - 中心極限定理の応用として証明される
  - ここで扱った中心極限定理は, 各観測値がi.i.d.であることを前提
  - 逆に, 各観測値がi.i.d.でなければ(無作為抽出でなければ), 標本平均は母平均に確率収束するとは限らない
- Sample selection bias
  - 無作為抽出でないサンプルについては, 標本平均は母平均の一致推定量にならない(バイアス・偏りがある, という)
  - ある特性・属性を持った観測値をover-sample, over-representしているとすれば, そのような観測値に引っぱられる
  - 例:ケーキ食べ放題の店で「甘いもの好きですか?」と聞いたら?
  - 例:家計調査の消費額はあまり変動しない?

1948年のアメリカ大統領選挙



- 「番狂わせ」が起こったと言われる
  - 民主党:ハリー･S･トルーマン(現職)
  - 共和党:トマス･E･デューイ
- 選挙運動におけるデューイの優位
  - トルーマンの人気の下落, 民主党の分裂
  - デューイは慎重な選挙戦を展開
- あらゆる予測がデューイの勝利を支持
  - 選挙直後の電話調査はデューイ優位
  - 共和党よりの「シカゴ･トリビューン」は「デューイ勝利!」と報道
- 結果:トルーマンの勝利
  - オハイオ, カリフォルニア, イリノイを制した
  - 電話を持っている有権者は裕福で, デューイを支持する傾向

分散・標準偏差の推定

- 標本分散，標本標準偏差が不偏・一致推定量

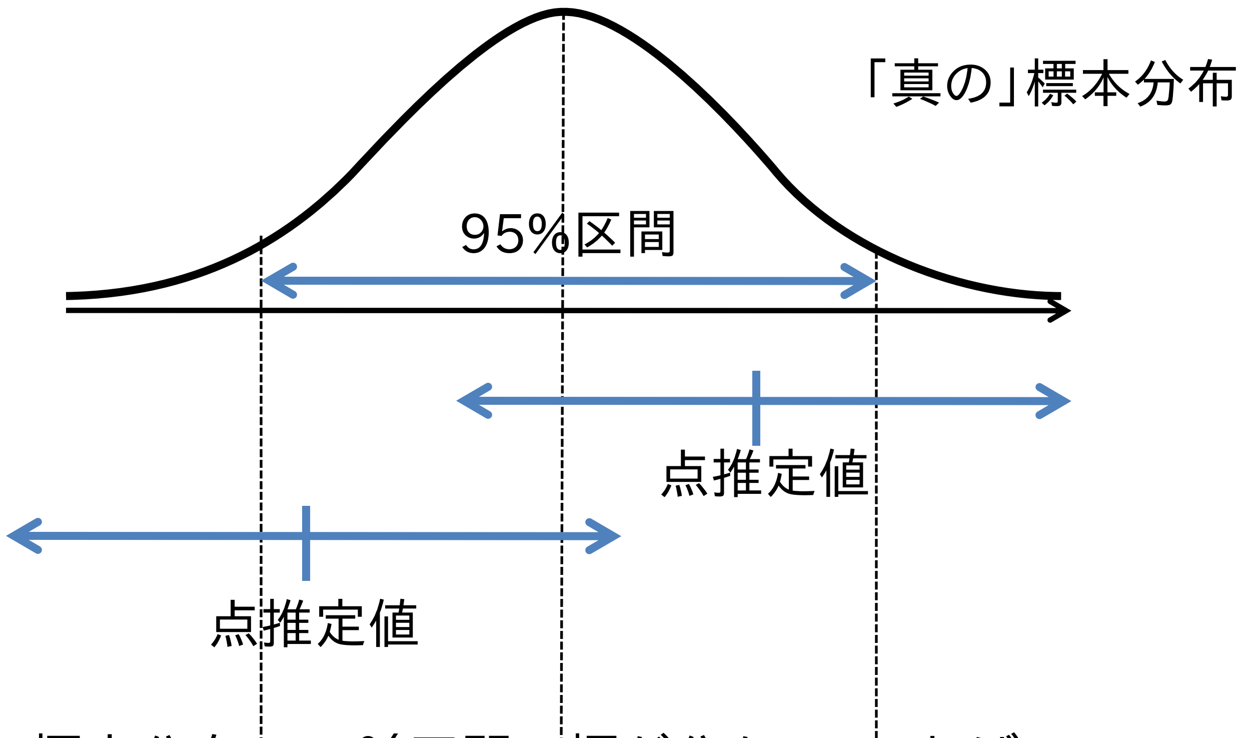
$$s_Y^2 = \frac{\left(Y_1 - \bar{Y}\right)^2 + \ldots + \left(Y_n - \bar{Y}\right)^2}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2$$

  - 証明は省略
- 自由度修正
  - 記述統計のときと分母が異なる
  - 一致性の観点からはあまり問題ではない
  - 「平均」というパラメタをすでに1つ推定しているのでその修正
  - 標本平均は母平均よりもデータに「近い」ので，その違いを修正

区間推定／信頼区間の形成

- 例：平均の話.
- 標本平均はサンプルによって異なる
  - 標本平均は確率変数ですから.
  - 母平均とぴたっと一致することはありえない
  - では,「点」で推定するよりも「この範囲にある」と示すのもよい?
- 区間推定
  - 特性値のbest guessを1つの値ではなく, 区間や範囲で表すこと
  - ただし,「この区間に必ず入る」とは言えないし, 言えたとしても役に立たない(「コインの表が出る確率は0から1のあいだにあります」).
- 95%信頼区間
  - 「95%の確率でこの範囲に母平均がある」
  - 母平均がこの範囲にあれば, 手元にある標本平均が観察される確率が95%

信頼区間の形成



- 「真の」標本分布の95%区間の幅が分かっていれば.
  - 点推定値を中心に同じ幅を取れば, 真の平均はそのあいだに入る
- 「真の」標本分布の95%区間を知るには, 標本分布の標準偏差が必要
  - 標本標準偏差で代理させる：不偏・一致推定量だから.

平均の信頼区間の形成の手続き

- 信頼度を決める
  - 90%, 95%, 99%がよく用いられる
  - 95%のばあい, 95%区間は平均から±1.96×標準偏差
- 必要な点推定量を求める
  - 標本平均, 標本分散（標本標準偏差）
- 標準誤差 standard error を求める
  - 標本分布の標準偏差の推定量のこと
  - 標本平均の標準偏差 = （母分散／サンプルサイズ）の平方根
  - 標準誤差 = （標本分散／サンプルサイズ）の平方根
= 標本標準偏差／（サンプルサイズの平方根）
- 信頼区間を求める
  - 95%のばあい, 標本平均±1.96×標準誤差

仮説検定 hypothesis testing

- 「平均はこれくらいではないか」という予想がすでにあるとする
  - 計算された標本平均が予想と一致することはない
  - たとえ予想が正しいとしても，標本平均は確率変数
  - 例：さいころの目の平均はぴったり3.5にはならない
- 「予想とあまりにかけ離れていたら」予想は間違い
  - 例：さいころの目の平均が6だとしたら，疑うほうがもっともらしい
  - 例：さいころの目の平均が3.4だとしたら？　3.1だとしたら？
  - 「あまりにかけ離れている」とは？
- 標本平均の分布を利用しよう
  - 予想が正しいとして95%区間を作ってみて，その区間に予想した値が入っていれば，確率95%で予想は正しい

例:さいころの目の平均値の仮説検定

- さいころを100回投げてみる.
  - ふつうのさいころだと仮定してみよう
- 予想が正しいときの標本平均の分布を求める
  - 真の平均は3.5だから, 標本平均の平均も3.5
  - 真の分散は約2.92
  - 標本平均の分散は 2.92/100 = 0.0292, 標準誤差は0.171
  - 95%区間は, [3.5–1.96×0.171, 3.5+1.96×0.171] = [3.165, 3.835]
- 実際に投げてみて.
  - 標本平均が3.4　→95%区間に入っている　　→予想は正しいらしい
  - 標本平均が3.1　→95%区間に入っていない　→予想は間違いらしい

  - 「確実」にはいえない.

仮説検定の用語

- 仮説検定とは:
  - 母集団の分布の特性値について仮説を立て, その真偽を統計的(確率的)に判定すること
- 仮説:
  - しばしば, 否定したい形で仮説を表す(帰無仮説 null hypothesis). 帰無仮説の逆を対立仮説(alternative)と呼ぶ
  - 例:ふつうのさいころだ　→「平均は3.5だ」
  - 「期待値が～に等しい」という定式化が多い
- 統計的な判定
  - どれほど「かけ離れて」いたら, 予想が間違いだと判定するか
  - 例:95%区間を使う　→5%の有意水準(significant level)
  - p値:帰無仮説が正しいときに, より「かけ離れた」値になる確率
  - 予想が間違い　→棄却(reject)
  - 予想は正しい　→受容(accept)

仮説検定の一般的な手続き

- 手続き
  1. 仮説を立てる.
  2. 有意水準を決める.
  3. 検定統計量(test statistics)を計算する.
  4. p値を求めて, 棄却/受容を判定する.
- 検定統計量
  - p値が計算できるためには, 帰無仮説が正しいときの分布がわかっていないといけない
  - ある手続きに従えば, 既知の分布(たとえば正規分布)に従うような確率変数があれば便利 ←ここで中心極限定理が登場
- 仮説検定のロジック
  - もし帰無仮説が正しければ, 検定統計量が既知の分布に従う
  - 計算された検定統計量の値から, 実現する確率(p値)が求まる

例:さいころの目の平均値の仮説検定

- 平均値の仮説検定
  - このときの検定統計量をt値(t-statistics)と呼ぶ

$$t\text{-stat} = \frac{\text{標本平均}-\text{仮説}}{\text{標準誤差}} = \frac{\bar{Y}-\mu_{Y,0}}{\mathrm{SE}\left(\bar{Y}\right)} = \frac{\bar{Y}-\mu_{Y,0}}{\sqrt{s_Y^2/n}} \xrightarrow{d} N(0,1)$$

- 手続き

1. 仮説を立てる:平均は3.5
2. 有意水準を決める:5%
3. 検定統計量(t値)を計算する
   - 標本平均が3.4なら, (3.4 – 3.5)/ 0.171 = -0.586 → p = 0.279
   - 標本平均が3.1なら, (3.1 – 3.5)/ 0.171 = -2.342 →p = 0.009
4. p値を求めて,棄却/受容を判定する
   - 標本平均が3.4なら受容,3.1なら棄却
   - t-統計量の5%有意水準に対応するのは<u>1.96</u>.

平均値の検定

- H_0:「平均値が〜に等しい」に対する検定
- 検定統計量はt-統計量
  - 母分散は一般に未知なので, 標準誤差を用いる

- 「t分布を使う」といわれることも?
  - t-統計量は標準正規分布に漸近的に従う（中心極限定理）
  - サンプルサイズが小さいときには, 近似が十分ではないことも.
  - 「母集団の分布が正規分布のとき」, t-統計量は t-分布に従う
  - 母集団分布は正規分布とは限らないし, サンプルは大きくしよう!

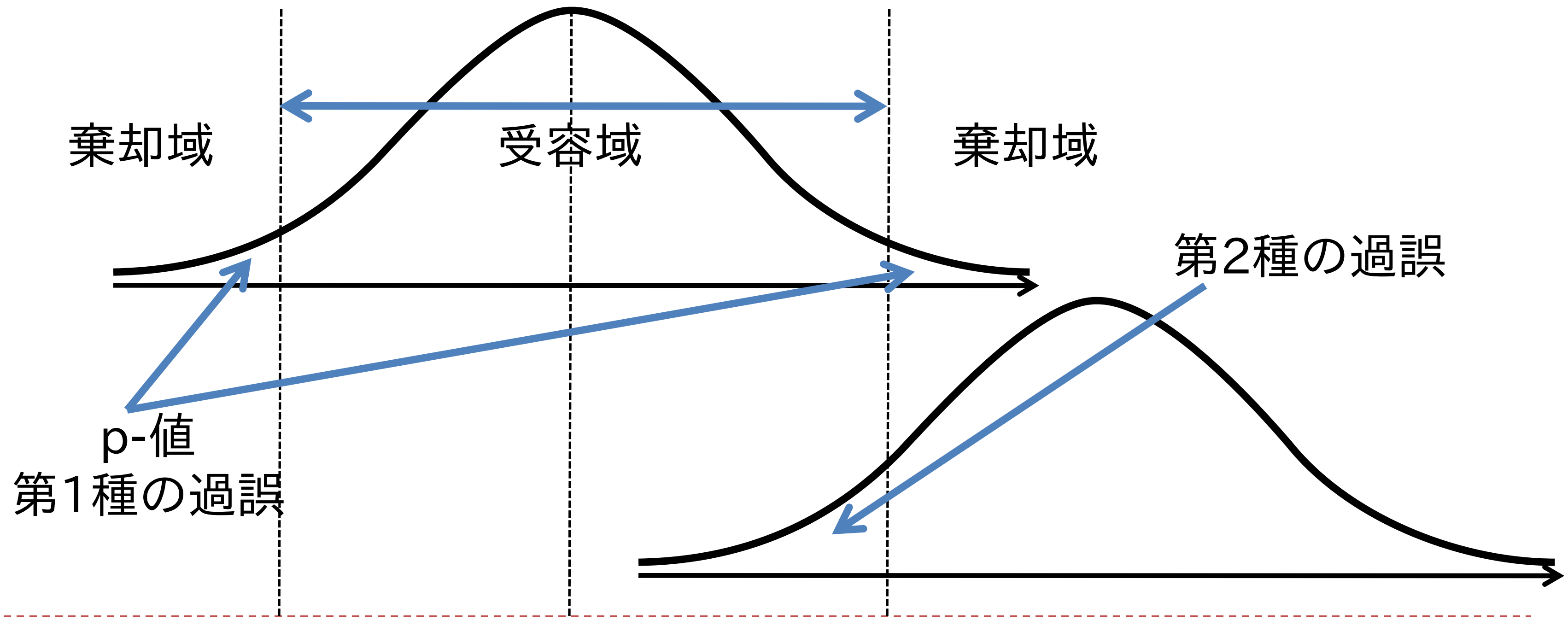
検定における過誤

- 統計的検定は確率的
  - 例：さいころを100回振って標本平均が3.1以下になることも0.9%の確率でありうる
  - ひょっとすると，「滅多に起きないこと」が，いま起きているのかも？
- 2種類の過誤
  - 第1種の過誤（type-I error）：仮説が正しいのに棄却する過誤
  - 第2種の過誤（type-II error）：仮説が偽なのに受容する過誤
- 検定の正しさ
  - Size：仮説が真のときに受容する確率
  - Power：仮説が偽のときに棄却する確率

	受容 accept	棄却 reject
真 true	OK	Type-I error
偽 false	Type-II error	OK

検定における過誤

- 2種類の過誤
  - サンプルが所与であれば, 第1種の過誤と第2種の過誤をともに減らすことはできず, トレードオフの関係にある
  - 「受容」も, 帰無仮説が真だ, と強く宣言しているわけではない
  - p-値 = 第1種の過誤を犯す確率



平均値の差の検定

- 2つの異なる母集団の平均の差の値を検定
- 帰無仮説は, H_0: $\mu_m - \mu_w = d_0$
  - 対立仮説はH_1: $\mu_m - \mu_w \neq d_0$
- 母平均の差の推定量は, 標本平均の差
  - それぞれ独立に母平均を平均とする正規分布に従う（←中心極限定理）
- 検定統計量（t-統計量）
  - 標本平均は独立に正規分布に従う
  - その差の分散は, 標本平均の分散の和

$$t\text{-stat} = \frac{\text{標本平均}-\text{仮説}}{\text{標準誤差}} = \frac{\bar{Y}_m - \bar{Y}_w - d_0}{\mathrm{SE}\left(\bar{Y}_m - \bar{Y}_w\right)} = \frac{\bar{Y}_m - \bar{Y}_w - d_0}{\sqrt{\dfrac{s_m^2}{n_m} + \dfrac{s_w^2}{n_w}}} \xrightarrow{d} N(0,1)$$

例：男女間賃金格差の検定

TABLE 3.1 Hourly Earnings in the United States of Working College Graduates, Aged 25–34: Selected Statistics from the Current Population Survey, in 1998 Dollars

	Men			Women			Difference, Men vs. Women		
Year	$\overline{Y}_m$	s_m	n_m	$\overline{Y}_w$	s_w	n_w	$\overline{Y}_m - \overline{Y}_w$	$SE(\overline{Y}_m - \overline{Y}_w)$	95% Confidence Interval for d
1992	17.57	7.50	1591	15.22	5.97	1371	2.35**	0.25	1.87–2.84
1994	16.93	7.39	1598	15.01	6.41	1358	1.92**	0.25	1.42–2.42
1996	16.88	7.29	1374	14.42	6.07	1235	2.46**	0.26	1.94–2.97
1998	17.94	7.86	1393	15.49	6.80	1210	2.45**	0.29	1.89–3.02

These estimates are computed using data on all full-time workers aged 25–34 from the CPS for the indicated years. The difference is significantly different from zero at the *5% or **1% significance level.

- 差がゼロ, と帰無仮説を設定すれば, 平均の差の検定を使える
- いずれの年でも, 平均が等しいという帰無仮説は5%有意水準で棄却

因果効果の推定／検定

- 実験や準実験, 自然実験
  - 実験群（treatment group）と対照群（control group）を設定
  - 「政策」が無作為割り当てなら理想的
- 政策の効果 treatment effect
  - 政策が0-1で表現できるとする
  - その効果は, E[*Y*| *X*=1] - E[*Y*| *X*=0]
  - 政策に効果がない, という帰無仮説の検定に「差の検定」が使える

  - 実際には「無作為割り当て」はむずかしい
  - 「無作為割り当て」になっているところだけ選ぶ？
  - その他のいろいろな方法：propensity score matchingなど
  - 例：交付団体 vs 不交付団体
  - 例：中絶の容認と犯罪率