比較現代日本論研究演習/現代日本論演習 (田中重人)

第 2 回「測定値と誤差」(2004.10.21)

1. 「真の値」と測定値
2. 誤差の種類と対策
3. 標本抽出のプロセス
4. 中心極限定理
5. 平均値の区間推定

【「真の値」と測定値】

<u>測定値</u> ＝ <u>真の値</u> ＋ <u>誤差</u>

記述　　　　　　　　推測

【誤差（error）の種類】

- 測定上の誤差

  計器の故障・測定精度の問題
  回答者の間違い・虚偽の回答
  調査員の間違い・不正
  調査票の不備
  入力ミス

- 対象者の選択に起因する誤差

【誤差への対策：科学的原則論】

誤差はゼロにはならない。

→追試を通じた再現性のチェック

しかし実際には追試はめったに行われない
・研究資源の問題
・時間の問題

【現実的な対策】

誤差の発生原因と
その大きさについて推定・公表

→ 追試をおこなう人の助けになる

→ 追試がなくても誤差について見当がつく

【統計学があつかえる誤差】

- 発生メカニズムが既知
- 誤差の範囲が確率的に決まる

無作為標本抽出にともなう
「**標本誤差**」がその典型である

【標本抽出の４段階モデル】

ユニバース (universe)

母集団 (population)

計画標本 (designed sample)

有効標本 (valid sample / case)

★ 伝統的な推測統計学では 4 段階にわけずに、2 段階で考えるのがふつう：

母集団=Universe + population

標本 = (designed/valid) sample

【無作為抽出】

random sampling

母集団から計画標本を選ぶ際に、

<u>すべての個体の抽出確率が等しくなる</u>

ように抽出する

➔「**等確率標本**」(probability sample)

※完全な無作為抽出はむずかしいので、実際には、
系統抽出、多段抽出、層化抽出などの技法が使われる

【標本誤差の推定】

「標本誤差」(sampling error)

＝無作為抽出による誤差

★　方向性をもたない

★　確率的に決まる

★　**標本数が大きいほど誤差の範囲が小さい**

➔「統計的推測」によって範囲を推定できる

【中心極限定理】

central limit theorem

★　等確率標本の平均値は、母集団の平均値より高くなったり低くなったりする。

★　しかし平均的にみれば母集団の平均値に一致すると期待できる (点推定)

★　標本サイズが大きいほど、母集団の平均とのずれが小さくなる

【平均値の信頼区間】

※「母集団では正規分布」の仮定が必要

★　標本の平均値が母集団平均値からはずれる確率は正規分布にしたがう

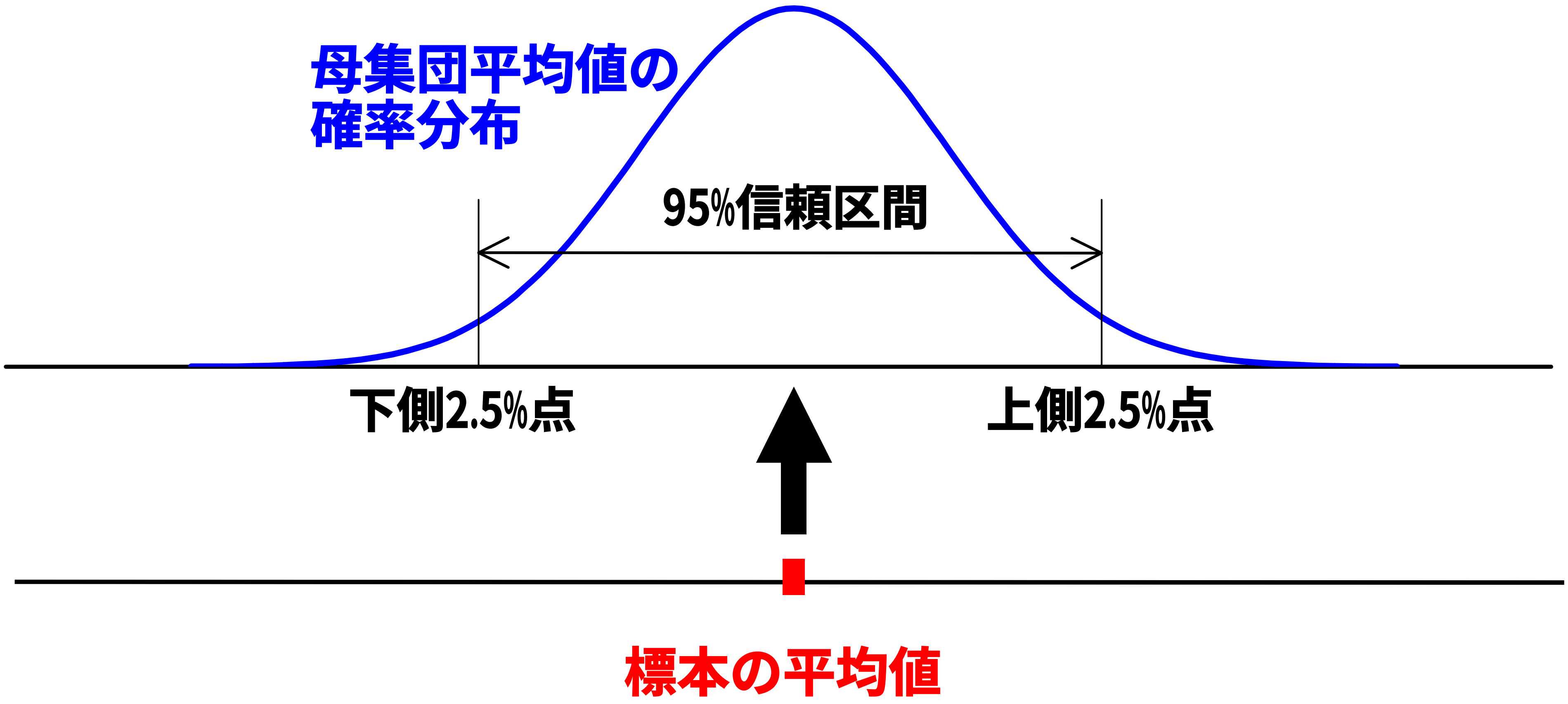
→ 標本の平均値と SD から、母集団の平均値の確率密度 (t 分布) が逆算できる

★　母集団の平均値の確率密度分布の両端を αだけ切り落としてえられる区間を (1－α) の「信頼区間」 (confidence interval) という。

★　このとき、αを「危険率」(critical rate) (1－α) を「信頼率」(confidence rate) という。

※　この値は自由に決めていいのだが、通常はα＝5%として、95%信頼区間を求める。

信頼区間のもとめかた



【無限母集団の仮定】

母集団がある程度大きければ、統計的推測のうえでは、母集団は無限大とみなしてよい。

厳密にいうと、$\dfrac{N-n}{N-1} \approx 1$ の場合

➔ 無限大の母集団から n 個の標本を無作為に選んだ場合について考える

★　無限母集団からの標本の場合の
平均値の信頼区間のおおよその値：

$$\underline{m} \pm t\text{臨界値} \times \frac{\text{SD}}{\sqrt{n}}$$

標本平均

標準誤差

※ t 臨界値は、自由度 (n－1) の t 分布にしたがって求める。n が 200 を超えると、ほぼ 1.96 になる (95%信頼率の場合)

【SPSS コマンド】

「分析」→「記述統計」→「探索的」

◎ 「従属変数」を指定
◎ パネル左下の「統計」だけをチェック

※ 信頼率を変更するには「統計」を選択
※ 「因子」を指定すると層別に分析できる

【課題】

適当な変数について、平均値の９５％信頼区間を求める。

印刷して提出