

2010.7.6 現代日本論演習 (田中重人)

第10講「分散分析」

1. 相関比の意味
2. エフェクト・サイズと相関比
3. 推測統計の基礎
4. 区間推定

【相関比の意味】

ある個体の値を x , 全体平均を M , 層別平均を m とすると、全体平均との差（偏差）は

$$x - M = (x - m) + (m - M)$$

次のデータの平均値と SD は？

{1, 1, 2, 2, 3, 5, 4, 5, 4, 3}

これをふたつの層に分割すると：

{1, 1, 2, 2}

{3, 5, 4, 5, 4, 3}

★ 【モデルとデータの乖離】

連関係数も相関比も、モデルとデータの乖離を表した値と解釈できる

- 特定の仮定（モデル）の下で予測される値
（期待度数・仮想 SD）を求める
- 実際のデータの値と比較する
- 0～1 の範囲の係数になるように調整する

多くの統計手法がこのタイプに属する

【ES と η の関係】

$$ES^2 = \frac{\eta^2}{1 - \eta^2} \times \frac{N^2}{n_1 n_2}$$

特に、2層の大きさが同じ ($n_1 = n_2$) なら、

$$ES^2 = \frac{4\eta^2}{1 - \eta^2}$$

層の大きさがちがえば、ES はこれより大きくなる

※ このように ES と η は互いに変換できる。

→ 両方示すのは冗長

【注意事項】

層別の平均値を分析する場合、
各層の人数は一定以上必要

(最低 20 人?)

→カテゴリ統合が必要になることがある

【記述統計と推測統計】

記述統計 (descriptive statistics)

= データ (**ケース**) の特徴を
数値や図表にまとめる

推測統計 (inferential statistics)

= 確率的な**誤差**を考慮して、
母集団の特徴を推測する

(教科書 pp. 3–5)

【無作為抽出】

random sampling

母集団から計画標本を選ぶ際に、

すべての個体の抽出確率が等しくなる

ように抽出する

→ 「**等確率標本**」 (probability sample)

袋のなかに色つきの玉が 60 個入っている:

赤色: 30 個

青色: 30 個

玉を n 個取り出したとき、その色は……?

→ 全世界から n 人を無作為抽出したとき、そのなかに ○○ の人は何%ふくまれるか?

【区間推定】

interval estimation

「答えは たぶん この範囲内にある」



信頼率 (confidence level) を適当に設定して

信頼区間 (confidence interval) を求める

全世界から 400 人を無作為抽出:

うどん が好き: 240 人

そば が好き: 160 人

うどんが好きな人の比率は?

$$0.6 \pm 1.96 \times \sqrt{(0.6 \times 0.4 / 400)}$$

答: ~ % (95%信頼区間)

【比率の区間推定】

標本の規模がじゅうぶん大きく ($n > 30$)、

比率があまり偏っていない ($0.1 < m < 0.9$) とき、

95%信頼区間は

$$m \pm 1.96 \times \sqrt{\frac{m(1-m)}{n}}$$

標準誤差

(standard error)

【平均値の区間推定】

母集団における平均値についても、同様の計算ができる。
ただし、正規分布を仮定：

$$\underbrace{\bar{m}}_{\text{標本平均}} \pm \underbrace{1.96}_{\substack{\downarrow \\ t \text{ 臨界値}}} \times \underbrace{\frac{SD}{\sqrt{n}}}_{\text{標準誤差}}$$

※ 「 t 臨界値」は n によって変化するが、 $n > 200$ で 1.96 に収束 (教科書 p. 281)。