

現代日本論演習／比較現代日本論研究演習 I 「統計分析の基礎」

第11講 推測統計の基礎と区間推定

田中重人 (東北大学文学部准教授)

[テーマ] 推測統計の基礎

1 前回課題について

男性: 1, 2, 3, 3, 4, 4 → 平均 2.833

女性: 2, 3, 4, 4, 5 → 平均 3.600

全体の平均: 3.182

グループ別平均値を当てはめた「仮想」データの平方和は、つぎのようになる。下線部に注意。

$$\text{グループ間平方和} = 6(\underline{2.833 - 3.182})^2 + 5(\underline{3.6 - 3.182})^2 = 1.603 \quad (1)$$

これを $n (=11)$ で割って平方根をとると標準偏差が得られる。

$$\text{仮想 SD} = \sqrt{\frac{1.603}{11}} = 0.382 \quad (2)$$

$$\eta = \frac{\text{仮想 SD}}{\text{実際の SD}} = \frac{0.382}{1.113} = 0.343 \quad (3)$$

ただし、SPSS では平方和を $n - 1 (=10)$ で割って「標準偏差」を求めているので、注意。度数がある程度大きくなれば (およそ $n > 200$ の場合)、このことによる違いは気にしなくてよい。

2 復習

- 記述統計と推測統計 (教科書 pp. 3-5)
- 母集団と標本 (第3講資料)
- 無作為抽出と確率標本

3 統計的推測のふたつの方法

- 袋のなかに色つきの玉がたくさん入っている。ここから8個取り出したところ、すべて赤であった。
- 全世界から8人を無作為抽出して麺類の好みをきいたところ、全員が「うどんが好き」と答えた。
- 2010年FIFAワールドカップの際、ある水族館で、国旗を貼った餌箱のどちらからタコが餌を食べるかで勝敗を占った。全8回について、先に食べたほうのチームが勝利 <[http://ja.wikipedia.org/wiki/パウロ_\(タコ\)](http://ja.wikipedia.org/wiki/パウロ_(タコ))>

このような情報 (= 標本統計量) から、母集団における統計量 (= 母比率) を推測する

区間推定: ある統計量の母集団における値について、確率的な推測によって範囲を求める → 母比率はたぶん ○○ から ×× の範囲にある

統計的検定: ある統計量の母集団における値について何らかの「帰無仮説」(null hypothesis) を設け、それが棄却できるかを判定する → 母比率が0.5だと考えてよいか?

統計的検定のほうが計算が簡単であるため、よくつかわれている。区間推定を論文等で目にする機会はあまりないが、きちんと理解するにはまず区間推定の考え方をおさえるのがよい。

4 母比率の区間推定

4.1 区間推定の原理

- (1) 「信頼率」を決めておく (たとえば95%)
- (2) $(1 - \text{信頼率})$ の確率を両極の事象に設定する (高いほう、低いほうからそれぞれ2.5%ずつを除く)
- (3) 母集団における値がいくつであれば、この両極端を除いた区間に測定値が入るかを計算する。統計量の性質に応じて、これを計算するための式と数表があるので、それを利用する。

このようにして求めた、母集団においてありうる値の集合が「信頼区間」である。通常、最初に決めた「信頼率」を明示して、「95%信頼区間」などのようにいう。

4.2 母比率の区間推定

標本の規模 n がじゅうぶん大きく ($n > 30$)、比率 m があまり偏っていない ($0.1 < m < 0.9$) とき、母比率の95%信頼区間は次の式で求められる:

$$m \pm 1.96 \sqrt{\frac{m(1-m)}{n}} \quad (4)$$

4.3 例題

全世界から400人を無作為抽出して麺類の好みを訊いたところ、「うどんが好き」と答えた人が240人であった。このとき、母集団(全世界の人々)におけるうどん好きの比率の95%信頼区間を求めよ(欠損値はないものとする)。