確率・統計　後期　第6回

母比率の区間推定

稲積　泰宏（いなづみ　やすひろ）

今日の内容

- 母比率とは
- 標本比率の性質
- 母比率の区間推定
- 正規近似の条件
- 信頼区間の幅と必要標本サイズ

母比率とは

定義

母集団において、ある特性を持つ個体の割合

記号：p

例

- 製品の不良品率
- 選挙での支持率
- 病気の有病率
- インターネット調査での賛成率

質的データと母比率

質的データ（二値データ）

- 例：合格/不合格、賛成/反対、良品/不良品

> 質的データを 0 と 1 で表現することで、母比率を扱う

ベルヌーイ試行と二項分布

ベルヌーイ試行

成功確率 p の試行を考える

$$X = \begin{cases} 1 & \text{（成功）} \\ 0 & \text{（失敗）} \end{cases}$$

n 回の独立試行

成功回数 Y は二項分布 $B(n, p)$ に従う

標本比率

定義

標本における成功の割合

$$\hat{p} = \frac{Y}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

ここで、X_i は i 番目の試行の結果（0 or 1）

> $\hat{p}$ は母比率 p の推定量

標本比率の期待値と分散

期待値

$$E[\hat{p}] = p$$

→ 不偏推定量

分散

$$V[\hat{p}] = \frac{p(1-p)}{n}$$

標本比率の分布

大標本の場合

n が十分大きいとき、中心極限定理により

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

標準化すると

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

問5）

正規近似の条件

経験則

以下の条件を満たすとき、正規近似が妥当

$$np \geq 5 \quad \text{かつ} \quad n(1-p) \geq 5$$

実用的な目安

標本サイズ $n \geq 30$ かつ p が極端でない（0.1〜0.9程度）

母比率の区間推定の問題点

標準化した統計量

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

には未知の母比率 p が含まれる

解決方法

分母の母比率 p を標本比率 $\hat{p}$ で置き換える

信頼区間の構成

信頼係数 $1-\alpha$ の信頼区間を求める

標準正規分布の上側 $\alpha/2$ 点を $z_{\alpha/2}$ とすると

$$P\left(-z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \leq z_{\alpha/2}\right) \approx 1-\alpha$$

不等式を p について解くと

$$P\left(\hat{p}-z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p}+z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \approx 1-\alpha$$

信頼区間の公式

母比率 p の信頼係数 $1-\alpha$ の信頼区間

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

ポイント

- 標準正規分布の上側確率点を使用
- 母比率 p を標本比率 $\hat{p}$ で推定
- 大標本が前提

よく使われる信頼係数と z 値

信頼係数	α	$z_{\alpha/2}$
95%	0.05	1.960
99%	0.01	2.576

95% 信頼区間では $z_{0.025} = 1.960$ を使用

例題3）

問6）

信頼区間の幅

信頼区間の幅

$$2 \times z_{\alpha/2} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

幅を狭くする方法

1. 標本サイズ n を大きくする
2. 信頼係数を下げる

> $\hat{p}(1-\hat{p})$ は $\hat{p} = 0.5$ のとき最大（0.25）
>
> → p が 0.5 に近いほど信頼区間は広くなる

例題4）

問7）

まとめ

母比率の区間推定

- 質的データ（0,1）を扱う
- 大標本では正規近似が利用可能
- 標準化には標本比率 $\hat{p}$ を使用

信頼区間の公式

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

次回までに問題集191から193を解いておくこと
レポートあります

感想を会議のチャット欄へ