

統計学入門

第9講,第10講

発表者: 辻 裕真

第9講

1つのデータから母集団を推理する
— 仮説検定の考え方

統計的推定

部分から全体を推理すること

何かデータを目にしたとき,その背後に膨大なデータがあり,その中の1つを観測している

いくつかのデータを観測し,その背後に広がる膨大な全データについて観測したい

もっともらしい母集団を推定する

例題

N枚のコインを投げ,表が10枚出たという結果のみを知っているとき,投げた枚数Nを次のように想定するのが妥当かどうか.

- ①16枚 ②36枚

どうやって母数 N を推定するか

表が出る確率から考えると「 $N=20$ 」
「 $N=21$ 」や「 $N=19$ 」でもおかしくはない

N が20からどこまで離れても妥当と考えていいのかがポイント

どうやって判断するか

仮説検定

「N=16」と仮定し,妥当な仮説であるか検討

N=16のときの95パーセント予言的中区間
表の枚数のデータは近似的に

$$\text{平均値 } \mu = \frac{16}{2} = 8 \quad \text{S.D.} = \frac{\sqrt{16}}{2} = 2$$

の正規分布だと考えられるので

仮説検定

95パーセント予言的中区間は

$$-1.96 \leq \frac{x-8}{2} \leq +1.96$$

$$4.08 \leq x \leq 11.92$$

「n=16」のとき表の枚数は

「4.08枚以上11.92枚以下」と予言できる
表の枚数が10というのは想定範囲内

「N=36」の場合

$$\text{平均値 } \mu = \frac{36}{2} = 18 \quad \text{S.D.} = \frac{\sqrt{36}}{2} = 3$$

95パーセント予言的中区間は

$$-1.96 \leq \frac{x-18}{3} \leq +1.96$$

$$12.12 \leq x \leq 23.88$$

「12.12枚以上23.88枚以下」と予言
表の枚数が10というのは想定外

仮説に対する考え方

1. 母集団に対する仮説は正しく、5%の確率でしか起こらない出来事が起きてしまった.
2. 母集団に関する仮説が正しくない

統計学では2の考え方を採用

仮説検定

「 $N=16$ 」は妥当な仮説として採用する

「 $N=36$ 」は妥当ではないので、仮説を棄却する

例題の回答

16枚は想定するが、36枚は想定しない

第10講

95パーセント当たる信頼区間を探し出す
— 区間推定

予言的中区間を推定に逆用

9講での仮説の評価法をすべての母数おの
おのに対して実行すれば、「可能性として残
すべき母数の集まり」が分かる

この母数の集まりを「可能な母集団の母数
として推定される区間」とし、「95パーセント信
頼区間」と呼び、母数をこのような区間で推
定することを「区間推定」と呼ぶ

例題

N枚のコインを投げる実験を行い表が10枚出たという結果のみを知っているとき,投げた枚数Nとして考えられるのは何枚から何枚までであるか.

コインの枚数についての仮説検定

$$\text{平均値 } \mu = \frac{N}{2} \quad \text{S.D. } \sigma = \frac{\sqrt{N}}{2}$$

$$Z = \frac{10 - \mu}{\sigma} \text{ として計算し, } -1.96 \leq z \leq +1.96$$

を満たすNは採択され, 満たさないNは棄却される

これを全てのNについて実行すれば母集団のNとして妥当な数値が得られる

コインの枚数についての仮説検定

$N = 12$ のとき $z = 2.31$

$N = 13$ のとき $z = 1.91$

(中略)

$N = 30$ のとき $z = -1.83$

$N = 31$ のとき $z = -1.98$

コインの枚数についての仮説検定

「Nについての区間推定」の結果は
「 $13 \leq N \leq 30$ 」となり、「Nの95パーセント信頼
区間」と呼ぶ

平均に対する区間推定

母集団が正規分布と分かっている、S.D.は分かっているが平均値が分かっていないとき、観測されたデータから平均値を区間推定したいとき

例題

精度のあまりよくない温度計で液体の温度を測った. 計測されたデータは, 実際の温度 μ を平均とし, S.D. 5°C の正規分布に従う. 計測した温度が 20°C であったとき, 実際の温度を95%信頼区間で区間推定せよ.

平均値の区間推定

$z = \frac{20 - \mu}{5}$ として計算し, $-1.96 \leq z \leq +1.96$

を満たしていればよい

計算すると,

$$10.2 \leq \mu \leq 29.8$$

となり, 実際の温度 μ の95パーセント信頼区間は 10.2°C 以上 29.8°C 以下となる

まとめ

95パーセント予言的中区間の式

$$-1.96 \leq \frac{x - \mu}{\sigma} \leq +1.96$$

x : 観測されたデータ μ : 平均値 σ : 標準偏差

を用いて母集団の区間や平均値の区間を
推定することができる