

支持率調査

- 25%が支持すると答えた！
1000人中250人？
40人中10人??
4人中1人???
- 同じ25%でも、信頼度は異なる

推定値の信頼度

- 0.25を中心とした区間
 - (0.25-C, 0.25+C)で示す
- 区間は確率95%で真の支持率を含む
 - 95%信頼区間と呼ぶ
- 区間の幅が短いほど信頼度は高い
- 区間の幅の決定方法は

平均値の信頼区間

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$
 $\bar{X} = \frac{X_1 + \dots + X_n}{n} \sim N(\mu, \sigma^2/n)$

区間 $\left(\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right)$
 は確率0.95で母平均 μ を含む
 μ の95%信頼区間とよぶ

2項確率の信頼区間

- $T \sim B(n, p)$, p は成功率
- n が大ならば $B(n, p) \approx N(np, np(1-p))$
- 近似的に $T \sim N(np, np(1-p))$
成功率 $\hat{p} = \frac{T}{n} \sim N\left(p, \frac{p(1-p)}{n}\right)$
- p の95%信頼区間は

$$\left(\hat{p} - 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$
 - n の中の p は未知なので、推定値の \hat{p} で置き換える

支持率調査(続)

- 40人中10人
観察比率 $\hat{p} = 10/40 = 0.25$
- $$1.96 \sqrt{\frac{0.25(1-0.25)}{40}} = 0.134$$
- $0.25 - 0.134 = 0.116, 0.25 + 0.134 = 0.384$
- 95%信頼区間は (0.116, 0.384)

n	下限	上限
40	0.116	0.384
60	0.140	0.360
80	0.155	0.345
100	0.165	0.335
200	0.190	0.310
300	0.201	0.299
400	0.208	0.292
500	0.212	0.288

CONFIDENCE関数

- CONFIDENCE(α , 標準偏差, 標本数)
- α
 - 信頼度を計算するために使用する有意水準を指定します。信頼度は $100 \times (1-\alpha)\%$ で計算されます。つまり、 $\alpha = 0.05$ であるとき、信頼度は95%になります。
- 標準偏差
 - データ範囲に対する母集団の標準偏差を指定します。これは、既知であると仮定されます。
- 標本数
 - 標本数を指定します。

CONFIDENCE関数の使用例

- 郊外に住む会社員50人を標本として、通勤時間を調査
 - 片道の平均時間が30分
 - 母集団の標準偏差は2.5
- $\alpha = 0.05$ とすると
 - CONFIDENCE(0.05, 2.5, 50) は 0.69291
- 95%信頼区間
 - 30 ± 0.69291 およそ [29.3, 30.7]
- 母集団平均 μ がこの範囲にある場合
 - 30よりも μ から離れた標本平均が得られる確率は、0.05より大きい
- 母集団平均 μ がこの範囲にない場合
 - 30よりも μ から離れた標本平均が得られる確率は、0.05より小さい