

統計量の分布

資料: 道具としての統計学

範囲: 第6章 (P.53～P.72)

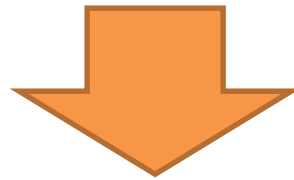
著者: 奥田 千恵子

60191008

大石 聖也

推測統計

- ・医薬研究において、1つ1つのデータは
唯一無二のもの



- ・推測統計
1回限りの調査や少数の実験データから
母集団の特性を推定

標本平均値の分布

- ・同じ大きさの標本抽出を繰り返す
→ 標本平均値 $\bar{X}_1 \sim \bar{X}_n$ は母平均値 μ の両側にばらつく

中心極限定理

- ・標本数が多い
 - 母集団がどんな分布でも、標本平均値の分布は正規分布に近似的に従う
- ・条件はゆるやか

中心極限定理

乱数による例

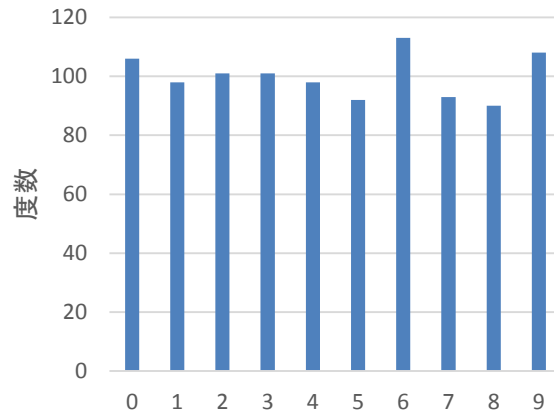


図1、1000個の乱数

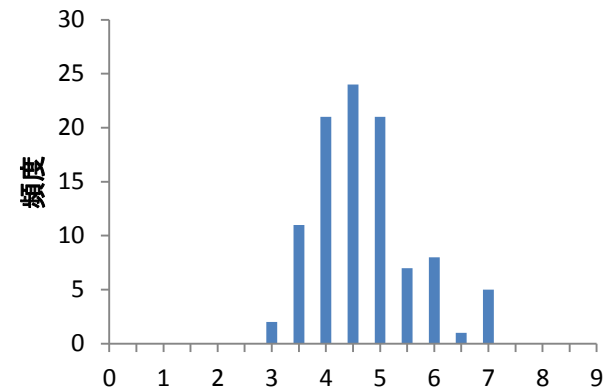


図2、100個の標本平均値

点推定

- ・母集団の分布が分かっているとき、標本から得た1つの値によってその母数を推定
- ・母平均値に関して、中心極限定理より
→ 標本平均値 = 母平均値

点推定の信頼度

	A	B	C
1	460	420	540
2	440	470	450
3	460	500	420
4	470	450	480
5	520	470	440
6		460	490
7		530	450
8		470	520
9		450	380
10		480	530
平均値	470	470	470
標準偏差	30.0	29.814	51.424

標準誤差

- ・母平均値の点推定で、標本平均値の信頼性の指標となる
- ・どのような統計量についても、それぞれの標本分布についての標準誤差を求められる

平均値の標準誤差

- ・平均値の標準誤差 $SEM = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$
(σ : 母標準偏差、 n : 標本数)

- ・ σ は未知なので、標本標準偏差SDで代用

$$SEM = \frac{SD}{\sqrt{n}}$$

平均値の標準誤差 例

$$A: SD=30.0, n=5, SEM = 30.0/\sqrt{5} = 13.416$$

$$B: SD=29.8, n=10, SEM = 29.8/\sqrt{10} = 9.428$$

$$C: SD=51.4, n=10, SEM = 51.4/10 = 16.262$$

- ・標準誤差が最も低いのはB
→ 最も信頼度が高い

比率の標準誤差

- ・比率の標準誤差

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

(n: 標本数、 $\hat{p}$: 標本比率)

比率の標準誤差 例

- ・薬の有効率の標準誤差を求める
被験者50人、Aは60%有効、Bは90%有効

$$A: SE_p = \sqrt{\frac{0.6(1-0.6)}{50}} = 0.069 \quad \therefore 6.9\%$$

$$B: SE_p = \sqrt{\frac{0.9(1-0.9)}{50}} = 0.042 \quad \therefore 4.2\%$$

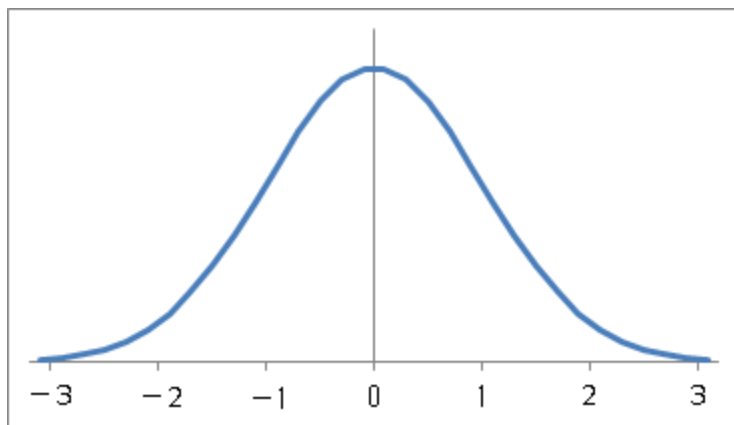
標準化

- ・さまざまな単位で測定され、母平均値などが異なるデータを一定の基準にそろえるよう変換
- ・平均値を原点にし、標準偏差が1になるように変換したもの

$$\rightarrow z\text{値} \quad Z = \frac{x - \mu}{\sigma}$$

標準正規分布

- ・正規分布 $N(\mu, \sigma^2)$ に従う母集団に属する x を標準化
→ z 値は平均値0、標準偏差1の正規分布に従う $N(0, 1^2)$



<http://www.cis.twcu.ac.jp/~konishi/stat14/stat14g.html>

標本平均値の標準化

- ・標本平均値 $\bar{X}$ を標準化

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

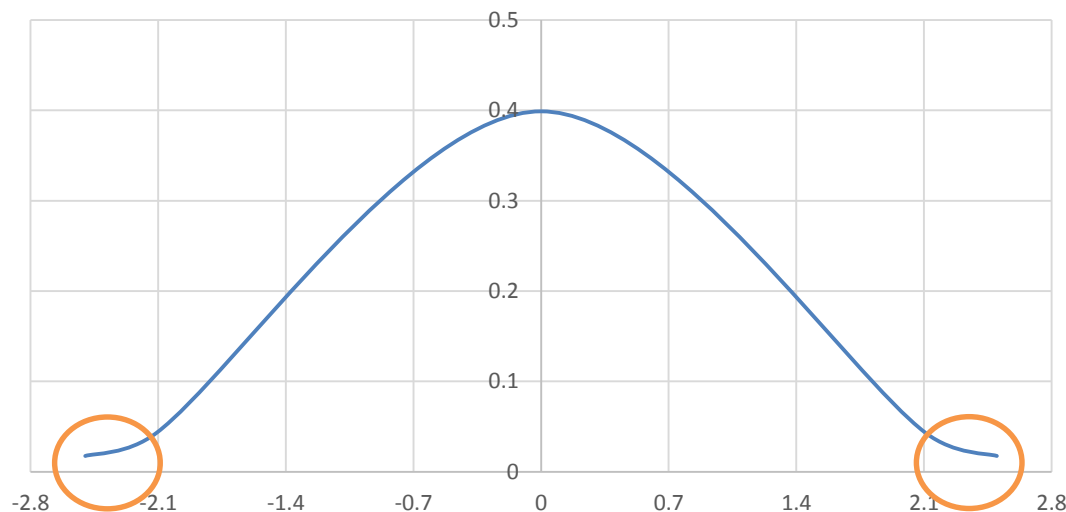
- ・例

一般男性の赤血球数 $490 \pm 30 \times 10^4 / \mu l$
身近の男性10人の平均値 $470 \times 10^4 / \mu l$

標本平均値の標準化 例

$$Z = \frac{470 - 490}{\frac{30}{\sqrt{10}}} = -2.108$$

$$P = NORMSDIT(z) \times 2 = 0.018 \times 2 \\ = 0.036 < 0.05$$



t分布

- ・母標準偏差 σ が未知のときに、標本平均値と母平均値の差の検定に用いる

$$t = \frac{\bar{X} - \mu}{\frac{SD}{\sqrt{n}}}$$

- ・t値は自由度、 $n-1$ のt分布に従う
- ・ n が100近くあるとき、t分布は標準正規分布とほぼ重なる

まとめ

- 点推定
→ 少ない調査から母集団の特性を推定する
- 標準誤差
→ 点推定の信頼度を確かめる指標
- 標準正規分布
→ 標準化した標本平均の分布