

データの分布と統計量の分布

資料：道具としての統計学(著:奥田千恵子)

範囲：第5章・データの分布

第6章・統計量の分布

目次

- 第5章
 - 2項分布
 - 正規分布
- 第6章
 - 推測統計
 - 標本平均値の分布
 - 標準偏差
 - 正規分布から導かれる分布
 - 標準正規分布
 - t分布

データをモデルに当てはめる

- 2項分布
 - 可能性が2つしかない場合の確率分布
 - コインの裏or表
 - 風邪をひいているorひいていない

2項分布・コイン投げの例

- 10個のコインを投げた時、表がx個出る確率

①x個を1組として取り出す
組み合わせの数

$${}_{10}C_x = \frac{10!}{x!(10-x)!}$$

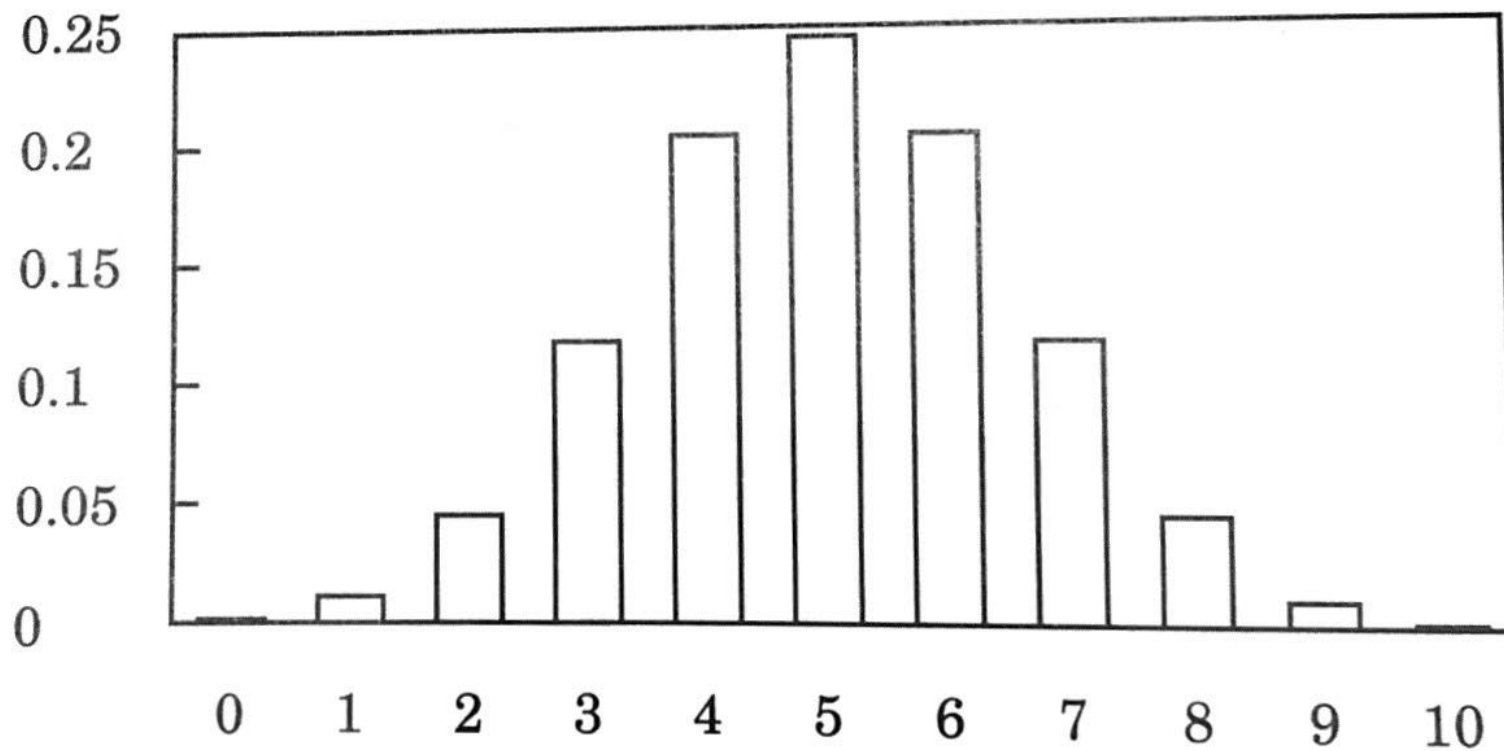
②表がx個出る確率P(X)

$$\begin{aligned} P(X) &= {}_{10}C_x \times p^x (1-p)^{10-x} \\ &= {}_{10}C_x \times \left(\frac{1}{2}\right)^{10} \end{aligned}$$

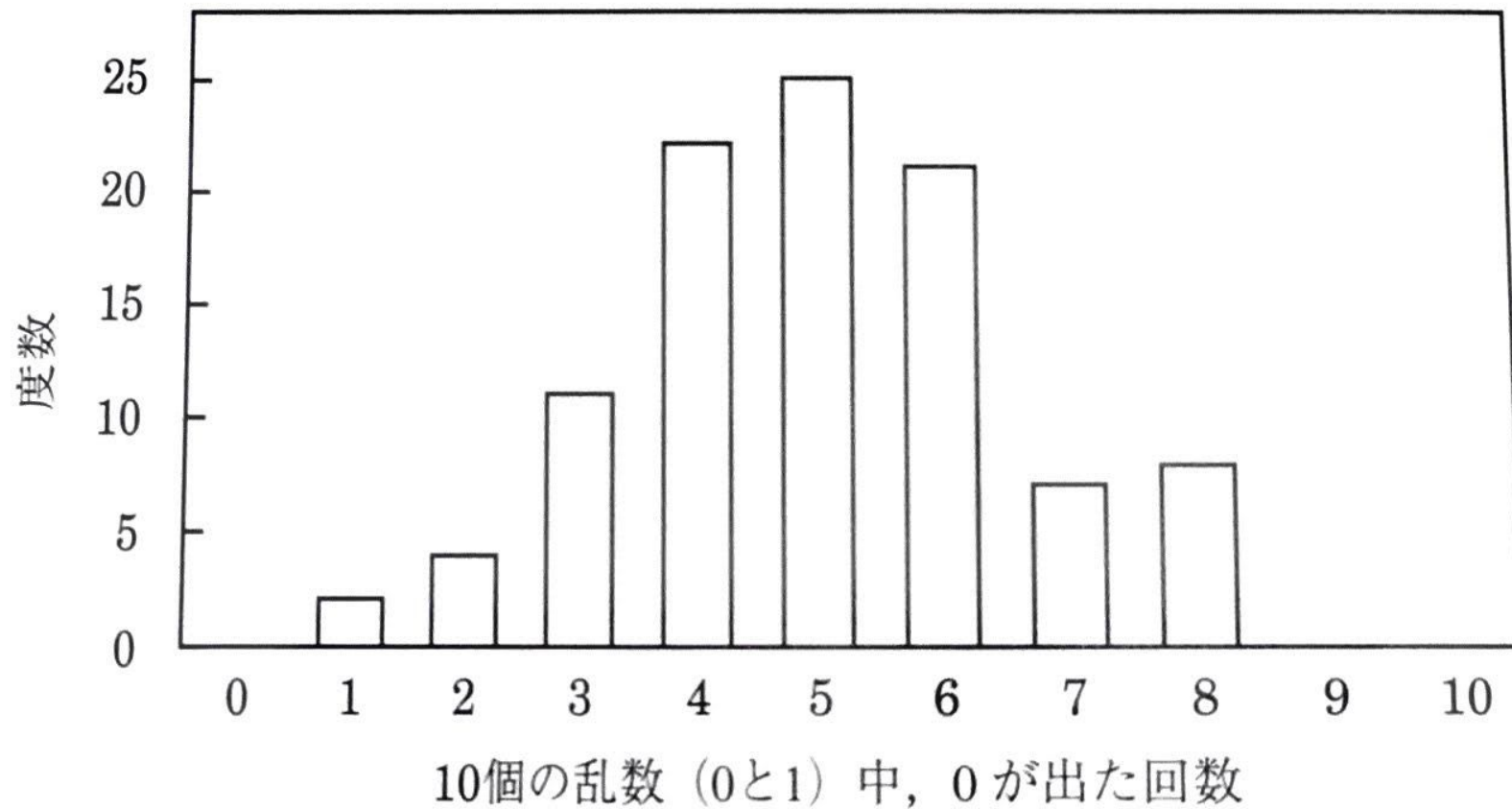
p: 表が出る確率

x	${}_{10}C_x$	P(X)
0	1	0.000977
1	10	0.009766
2	45	0.043945
3	120	0.117188
4	210	0.205078
5	252	0.246094
6	210	0.205078
7	120	0.117188
8	45	0.043945
9	10	0.009766
10	1	0.000977

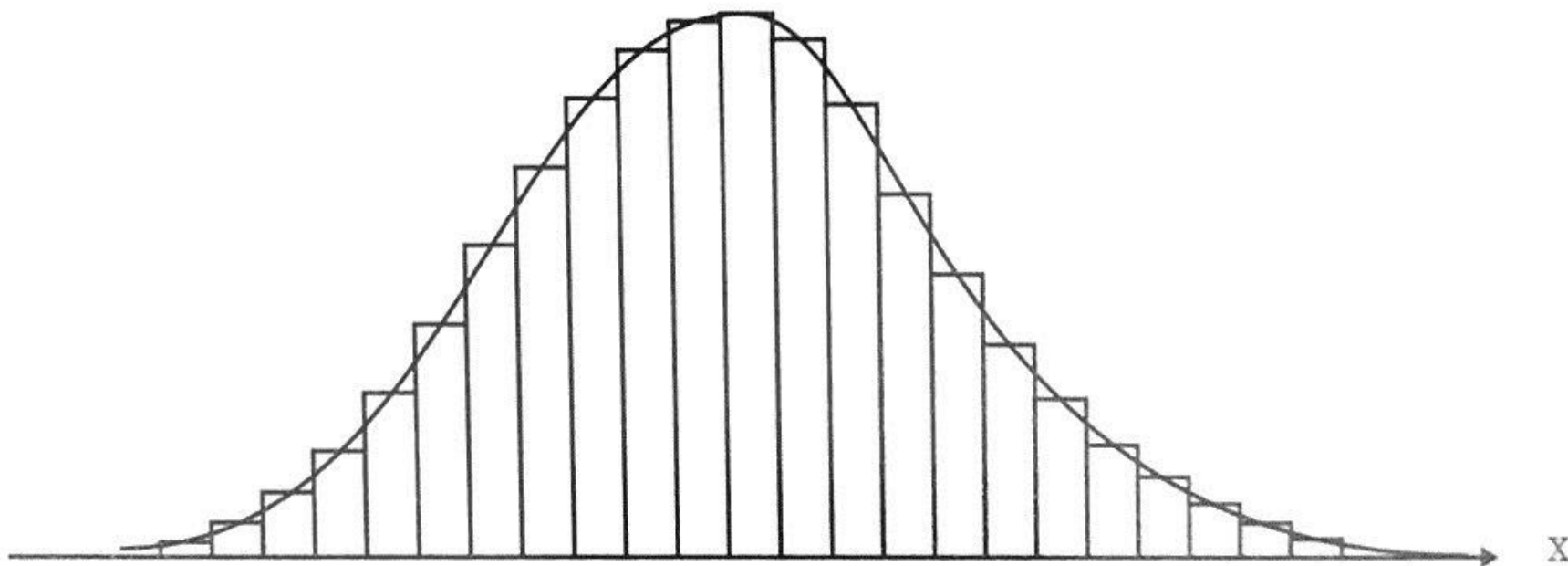
2項分布・コイン投げの例・図



2項分布・実験データとしてのサイコロ投げ

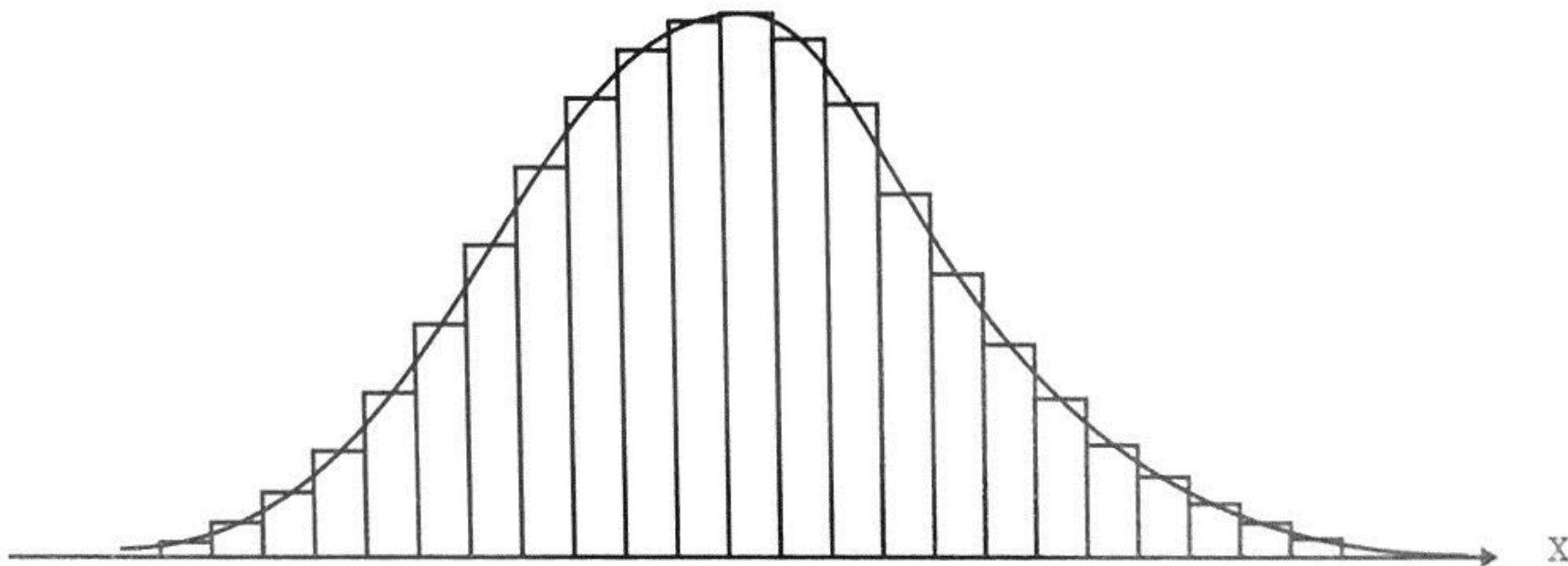


2項分布と正規分布



- コインの枚数 n を増やすと、二項分布は正規分布に近づく

正規分布



- 平均値の付近に集積するようなデータの分布を表した確率分布
 - スコットランド兵士の胸囲
 - 自然界のものはすべて、正規分布に従う(?)
- 偶発的な測定誤差を含むデータの集まりは正規分布に従う

正規分布の大切さ

- 統計解析は**母集団が正規分布に従う事を前提とするものが多い**。
(例)t検定、分散分析、直線回帰など
- **ただ、ソフトウェア工学の分野では母集団が正規分布に従わない事が多いらしいです...**
- **正規分布に従わない統計解析法をノンパラメトリックな検定法という**

正規性の検定法

- 正規確率紙
 - グラフ用紙にプロットする
- コルモゴロフ・スミルノフの適合度試験
- リリフォースの検定
- シャピロ・ウィルクスの検定

→この検定法は「正規分布である」と言えるわけではなく
正規分布から極端に離れていないといえるだけ

正規分布のパラメータ

- データが近似的に正規分布に従う

→母集団が正規分布に従う

母集団		標本	
母平均値	μ	標本平均値	$\bar{X}$
母標準偏差	σ	標本標準偏差	SD

- μ と σ の母数で分布の性質を表すことができる
- 現実データの $\bar{X}$ SDは母数の推定値として用いられる

第6章・統計量の分布

- 推測統計

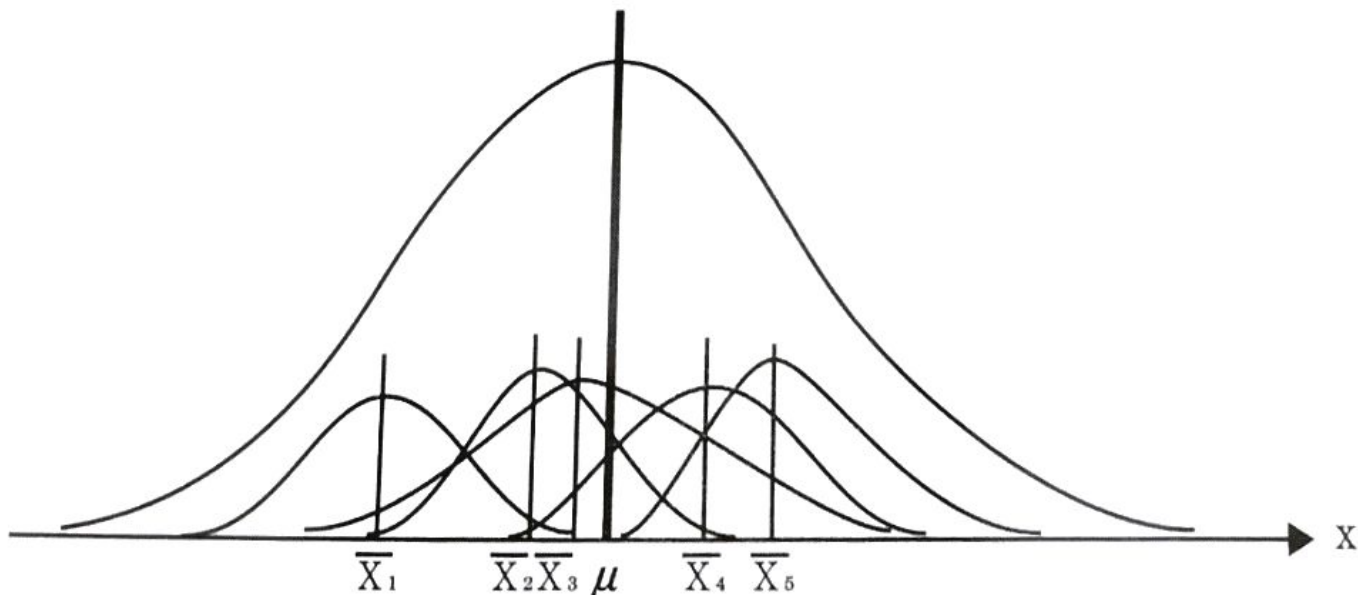
- 1回限りの調査や少数の実験データの統計量から、不確実さの度合いを数量化した上で、母集団の特性を推定する

- 記述統計

- 研究対象の集合としての性質を様々な統計量を用いて、整理・記述する

標本平均値の分布

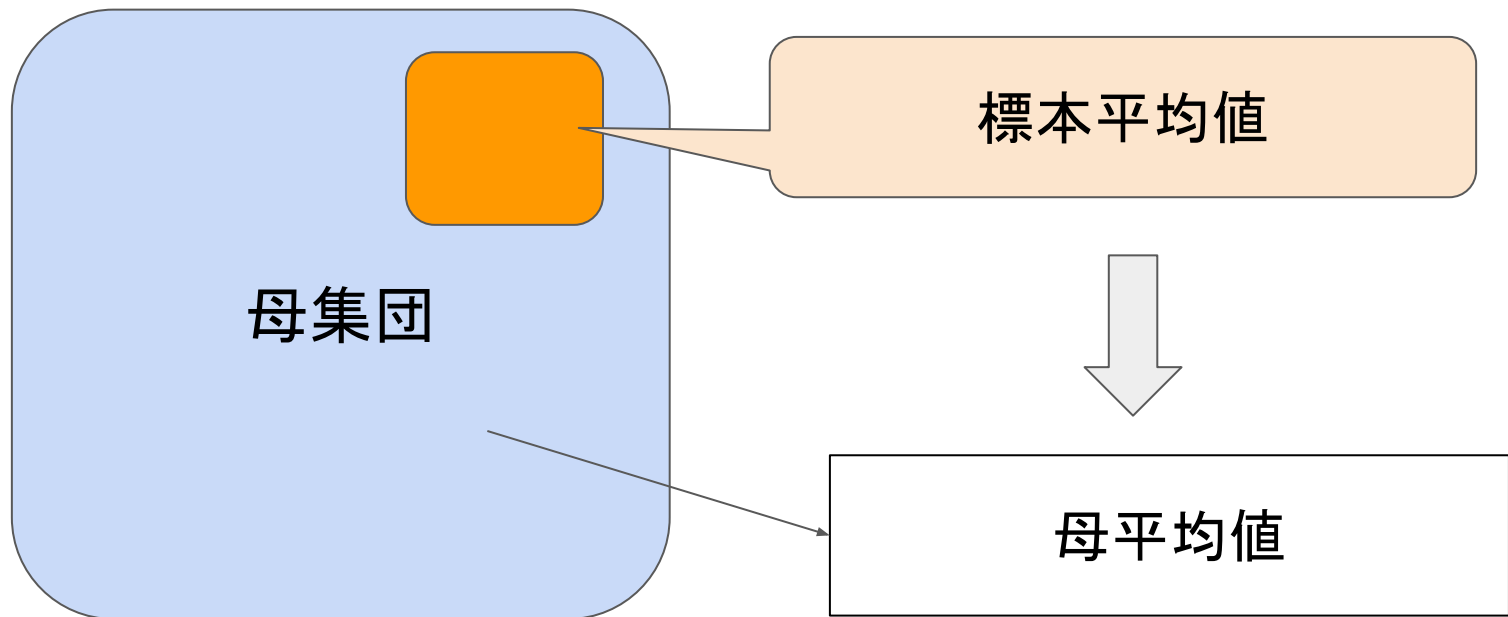
- 中心極限定理
 - 標本数が大きければ、母集団の分布型にかかわらず、標本平均の分布は正規分布に従う



- 標本平均値の平均値は母平均値と一致

標準誤差

- 点推定
 - 一つの標本平均値から母平均値を推定する
 - 母集団の分布の形はわかっている



標準誤差・推定の信頼度

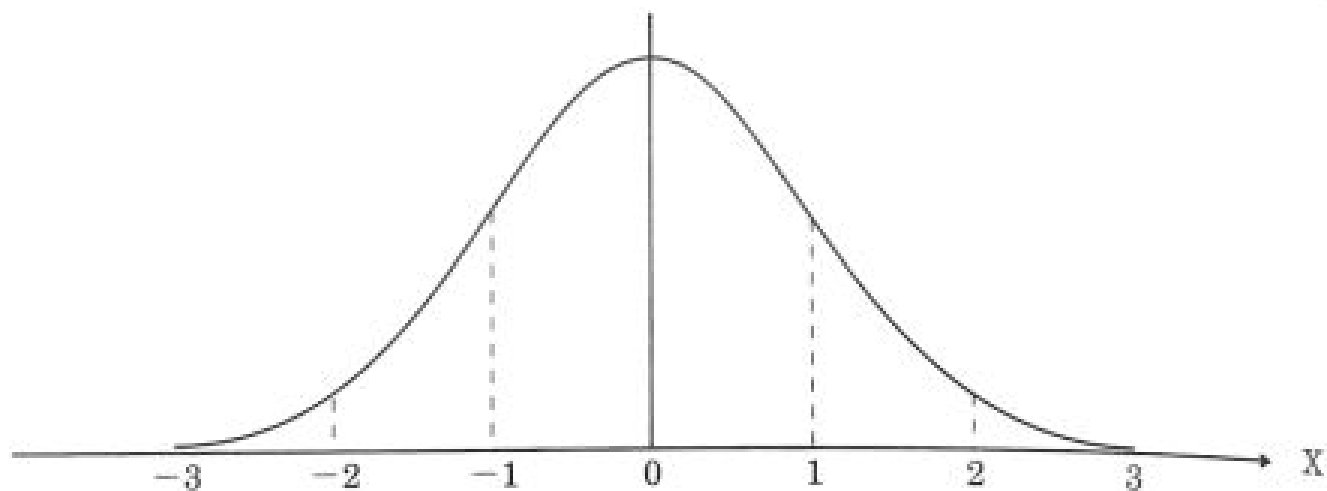
- 標準偏差(SD)で判断

Bが一番
信頼度が高い

	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($\bar{X}$)	470	470	470
標準偏差(SD)	30.0	<u>29.814</u>	51.424

正規分布から導かれる分布・標準正規分布

- 標準化
 - 母平均値や母標準偏差が異なるデータを一定の基準の揃えるように変換すること
- 標準化したデータは標準正規分布に従う



正規分布から導かれる分布・t分布

- 小標本の場合に、実験データのずれを解決するためのもの
- ガンマ関数を用いた複雑な式で表現される

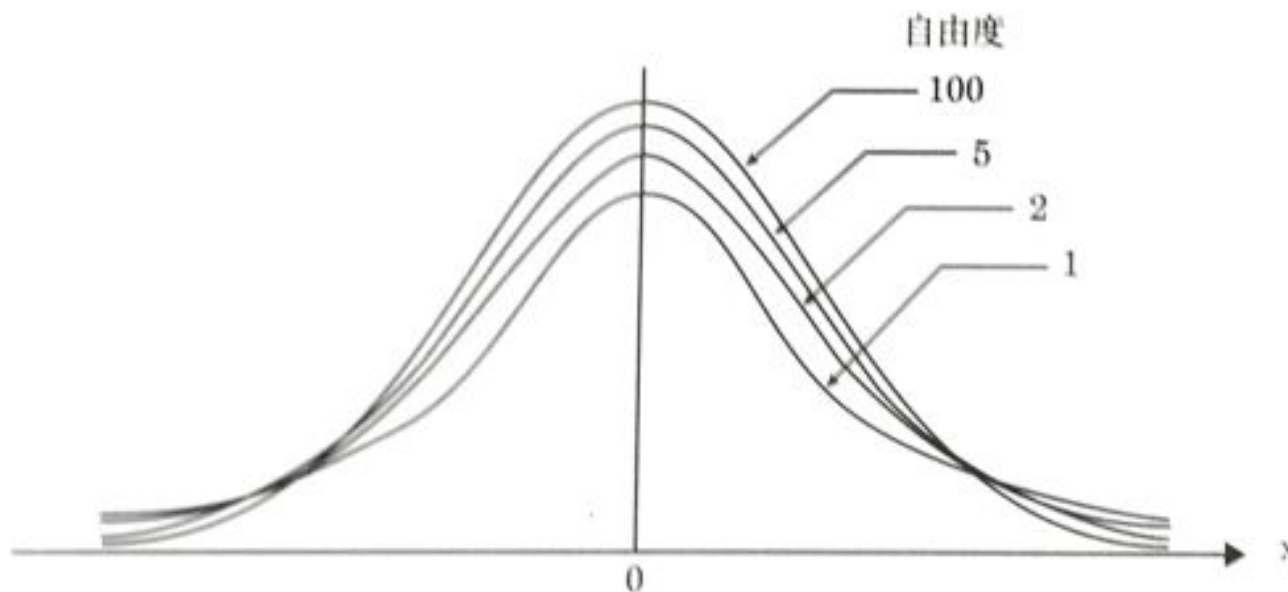


図 6-5

正規分布から導かれる分布

- 母集団と差があるかないかを判定出来る