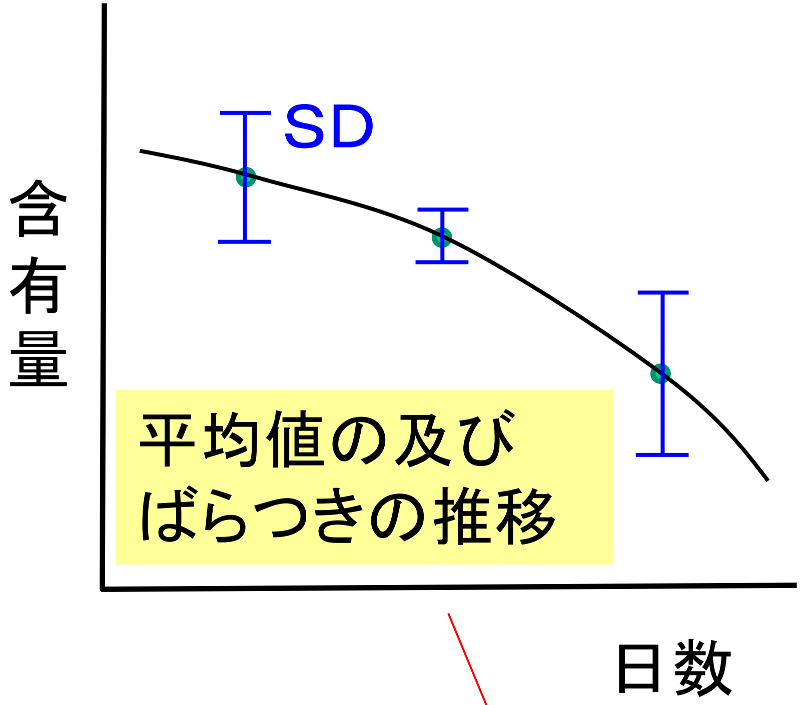
統計的推定とは？

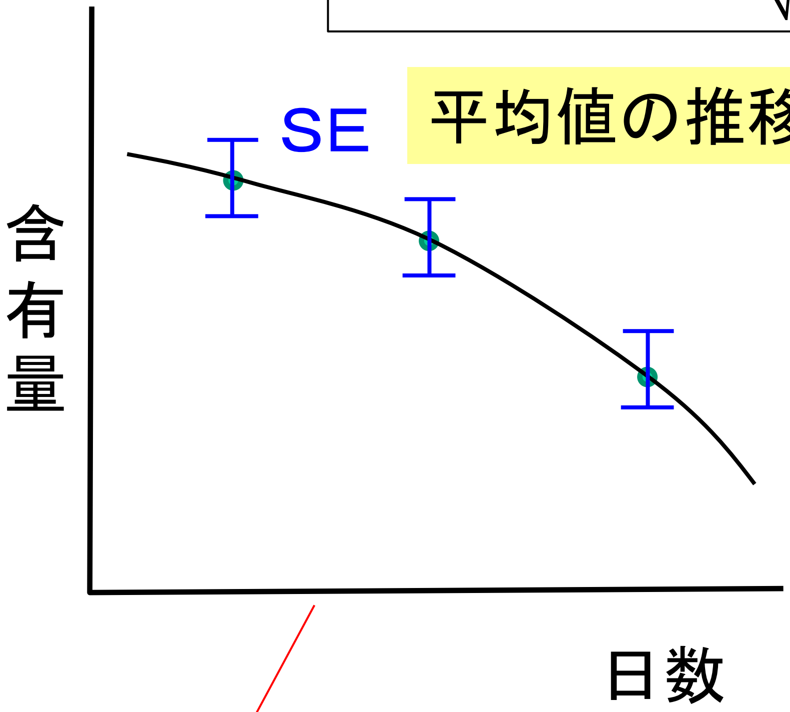
標本の特性→母集団の特性を推定

- 平均値
- 分散

標準偏差　と　標準誤差　の相違

$$\text{標準誤差} = \frac{\text{母集団 } or \text{ 標本の標準偏差}}{\sqrt{\text{標本のサンプルサイズ}}}$$





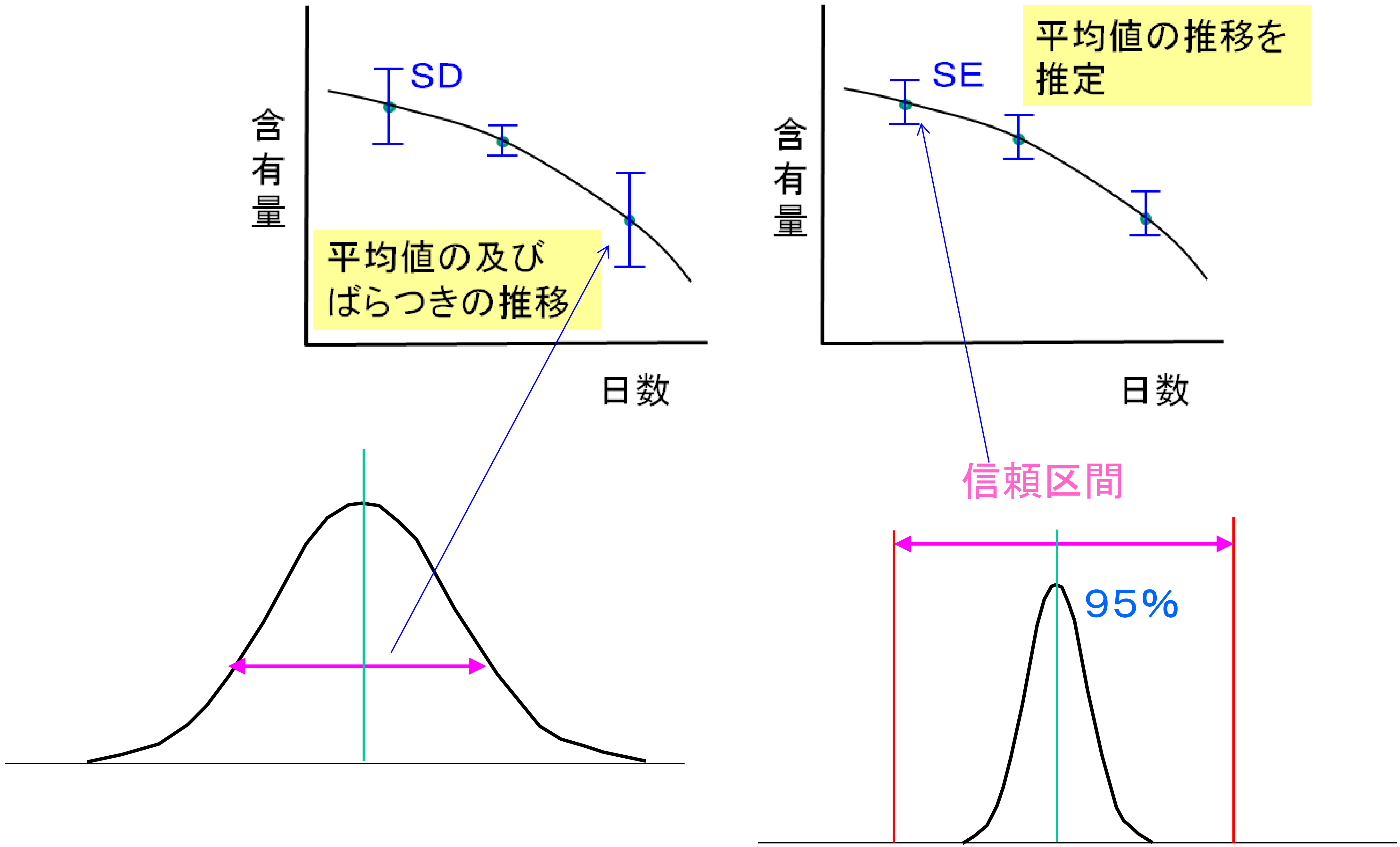
平均値の推移を推定

意味が違います

標準偏差（SD：standard deviation）　標準誤差(SE：standard error of mean)

データそのもののばらつき

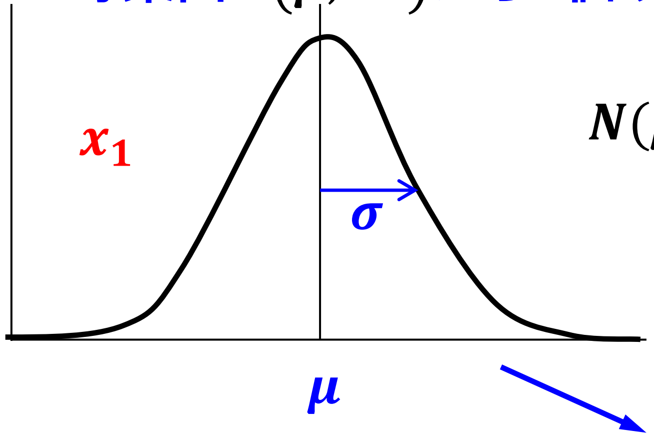
サンプリングデータより
推測した平均値の信頼区間


含有量
SD
平均値の及び
ばらつきの推移
日数
含有量
SE
平均値の推移を
推定
日数
信頼区間
95%

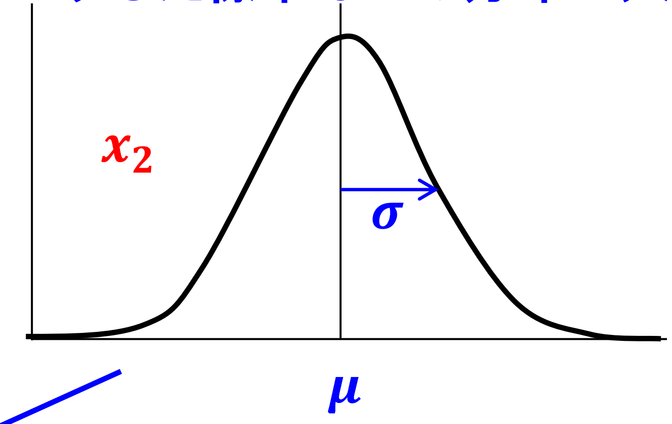
母集団$N(\mu, \sigma^2)$から1個サンプリングした標本はこの分布に入る

正規分布

$N(\mu, \sigma^2)$



$N(\mu, \sigma^2)$



$+$

↓ 加法性あり

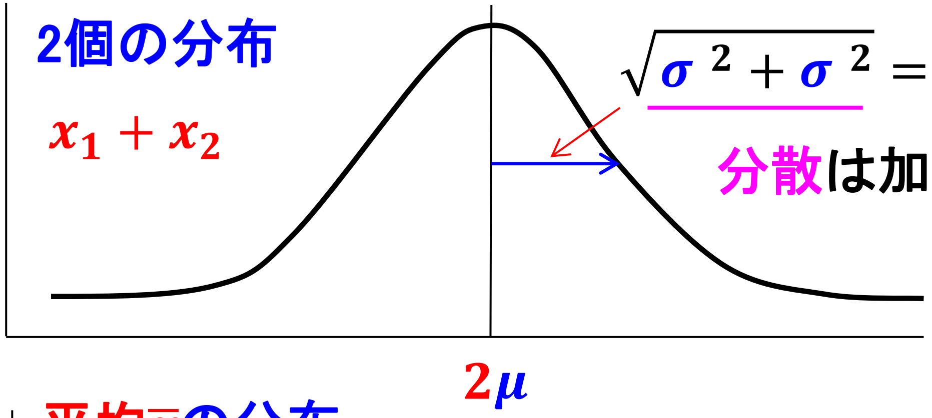
$$N(\mu+\mu,\ \sigma^2+\sigma^2)$$
$$= N\left(2\mu, (\sqrt{2}\sigma)^2\right)$$

2で割って平均する

$$N\left(\mu, \left(\frac{\sigma}{\sqrt{2}}\right)^2\right)$$

2個の分布

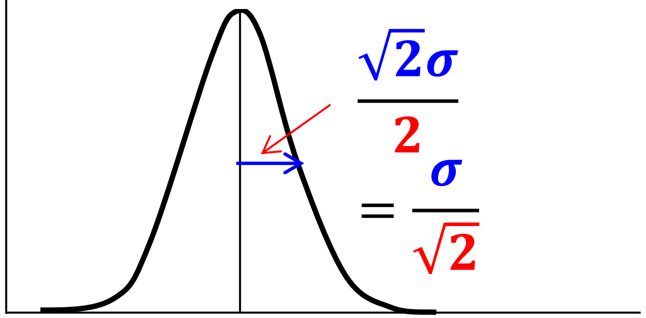
$x_1 + x_2$



$$\sqrt{\sigma^2+\sigma^2} = \sqrt{2\sigma^2} = \sqrt{2}\sigma$$

分散は加法性あり

平均$\overline{x}$の分布



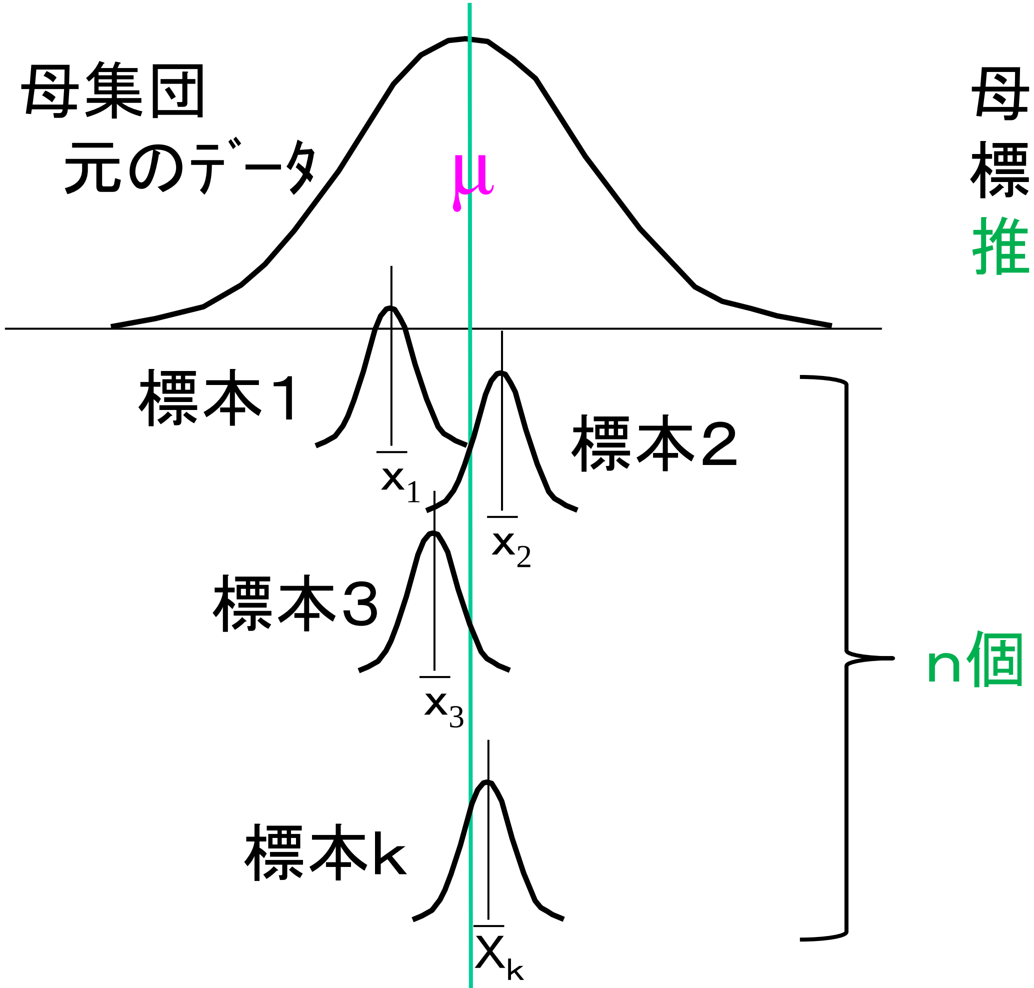
$$\frac{\sqrt{2}\sigma}{2} = \frac{\sigma}{\sqrt{2}}$$

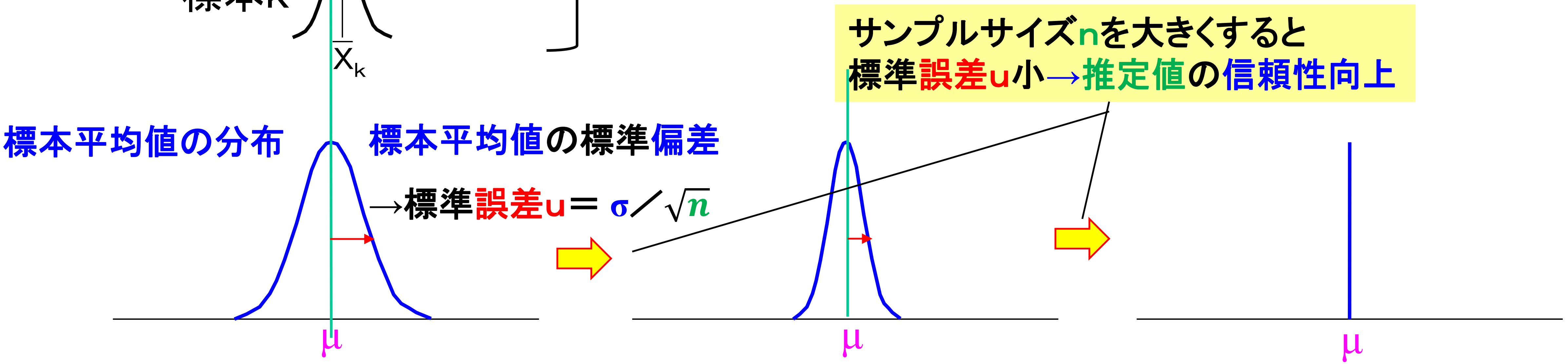
拡張

n個サンプリングした標本の標準偏差はn個の平均

$$\frac{\sqrt{n\sigma^2}}{n} = \frac{\sigma}{\sqrt{n}}$$

母集団の平均値μ(未知)から、n個サンプリングして標本とする。 その標本平均の分布より母集団のμを推定する





分布の推定
- 性能評価
- 信頼性推定

5%の確率で推定を間違う

標本平均の標準偏差

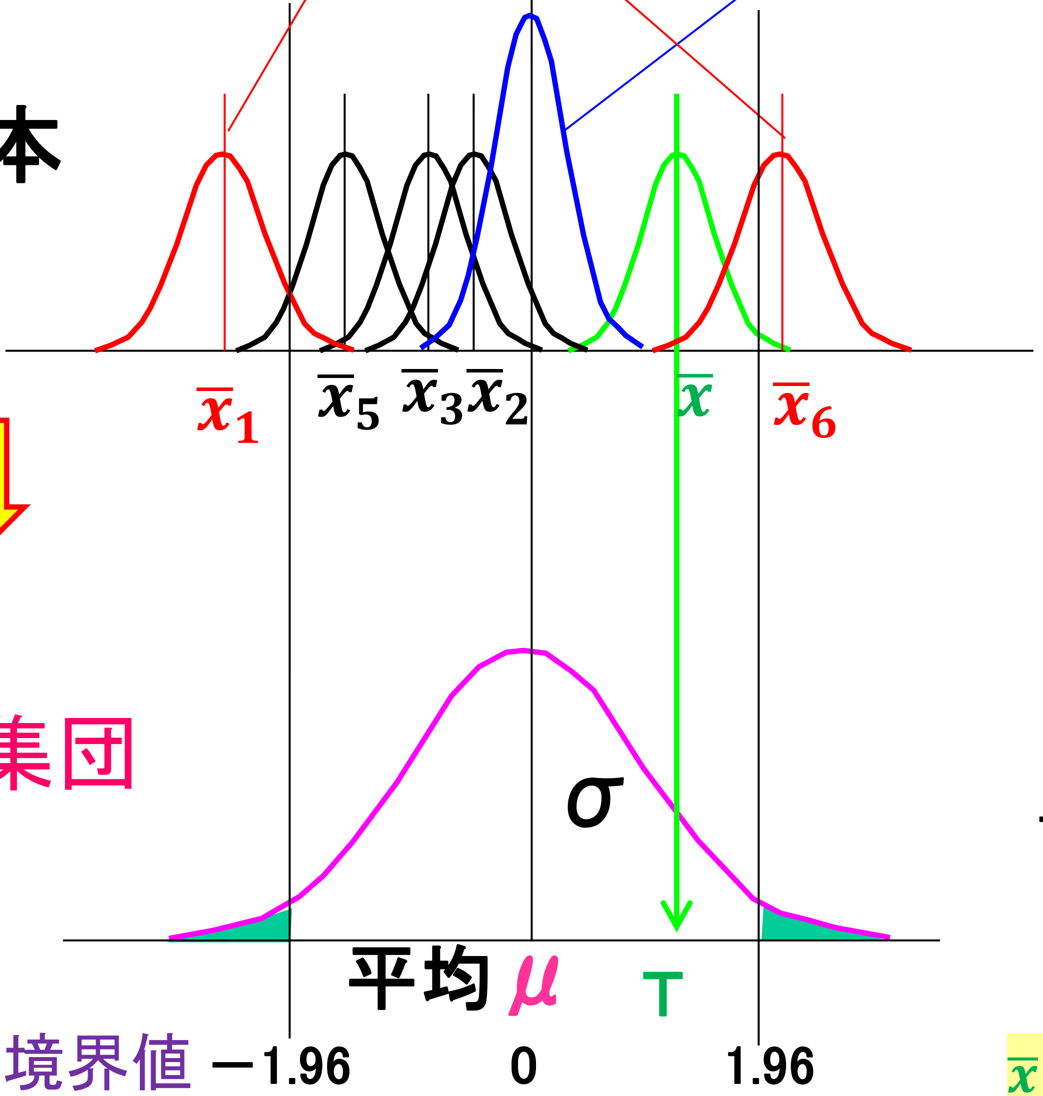
$\dfrac{\sigma}{\sqrt{n}}$ もしくは $\dfrac{s}{\sqrt{n-1}}$

n： サンプル数

標本

平均値の推定

母集団



標本平均 $\bar{x}$ を基準化する

$$T=\frac{\bar{x}-\text{期待値}}{\text{標準誤差}}=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \quad \text{or} \quad \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$$

$\alpha=0.05$の時

$Z(\alpha/2)=1.96$

$$-1.96<T=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}<1.96$$

変形して平均値 μ の推定式は

$$\bar{x}-1.96(\sigma/\sqrt{n})<\mu<\bar{x}+1.96(\sigma/\sqrt{n})$$

母集団の標準偏差σ不明のとき

$$-1.96<T=\frac{\bar{x}-\mu}{s/\sqrt{n-1}}<1.96$$

変形して平均値 μ の推定式は

$$\bar{x}-1.96(s/\sqrt{n-1})<\mu<\bar{x}+1.96(s/\sqrt{n-1})$$

サンプルサイズの算出式

母集団の平均値 μ は、

n個の標本の平均値　$\bar{x}$

母集団の標準偏差　σ

信頼度の境界値　$Z_{(\alpha/2)}$　で推定できる

n ≧100の時

Z推定

$$\bar{x} - \boxed{Z_{(\alpha/2)}\left(\frac{\sigma}{\sqrt{n}}\right)} < \mu < \bar{x} + \boxed{Z_{(\alpha/2)}\left(\frac{\sigma}{\sqrt{n}}\right)}$$

許容誤差 E とすると

$$E = \boxed{Z_{(\alpha/2)}\left(\frac{\sigma}{\sqrt{n}}\right)}$$

サンプルサイズ n は、

$$\boxed{n = \left(\frac{Z_{(\alpha/2)}\sigma}{E}\right)^2}$$

間違えても
良い確率

両側 α	$Z_{(\alpha/2)}$
1％	2.57
5％	1.96
10％	1.64

サンプルサイズの算出式

母集団の平均値 μ は、

n個の標本の平均値　$\bar{x}$

標本の標準偏差　s ←母集団の標準偏差 σ が不明な場合

信頼度の境界値　$t_{(\alpha/2,n-1)}$　で推定できる

n <100の時

t推定

$$\bar{x} - \boxed{t_{(\alpha/2,n-1)}\left(\frac{s}{\sqrt{n-1}}\right)} < \mu < \bar{x} + \boxed{t_{(\alpha/2,n-1)}\left(\frac{s}{\sqrt{n-1}}\right)}$$

許容誤差 E とすると

$$E = \boxed{t_{(\alpha/2,n-1)}\left(\frac{s}{\sqrt{n-1}}\right)}$$

サンプルサイズ n は、

$$\boxed{n = \left(\frac{t_{(\alpha/2,n-1)}s}{E}\right)^2 + 1}$$

間違えても良い確率

α	$n-1$	$t_{(\alpha/2,n-1)}$
5%	1	12.70
	2	4.30
	4	2.78
	9	2.26
	19	2.09

統計的推定:

母集団
ある市の中学1年男子12,000人
平均体重μは不明　←統計的推定

標本	
サンプル数	n＝100人
平均値	$\bar{x}$＝44.1kg
標準偏差	u＝ 3.9kg

標本平均 $\bar{x}$ を基準化する

$$T = \frac{\bar{x} - \text{期待値}}{\text{標準偏差}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

標準正規分布表において95%の確率になるのは$-1.96 < T < 1.96$

よって$-1.96 < (\bar{x} - \mu) / \sigma / \sqrt{n} < 1.96$

$$\bar{x} - 1.96\,\sigma / \sqrt{n} < \mu < \bar{x} + 1.96\,\sigma / \sqrt{n}$$

母集団の標準偏差σが未知のとき

ｻﾝﾌﾟﾙｻｲｽﾞが大きければ(n≧100)、 u≒σ

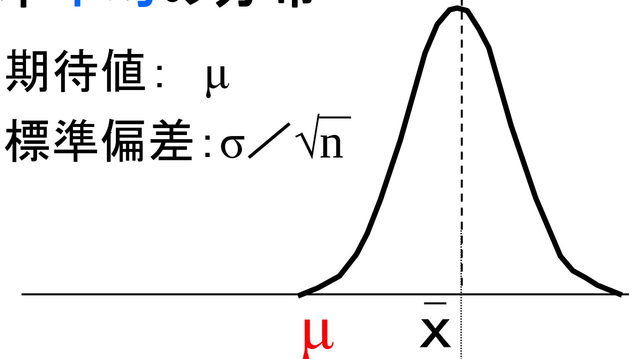
標本の数値を代入して、母集団の平均体重は

$43.3 < \mu < 44.9$ kg と推定される

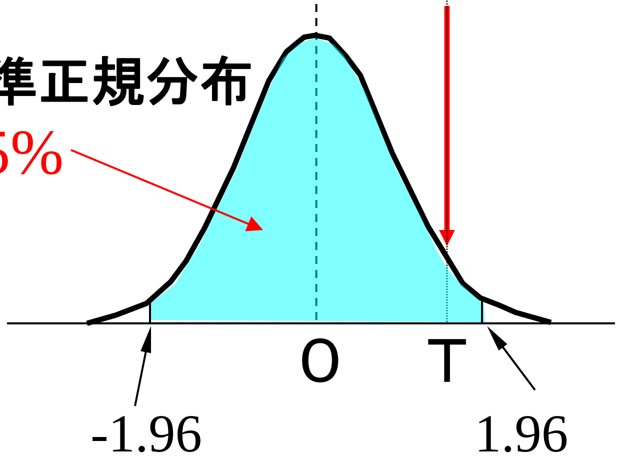
標本平均の分布

期待値: μ

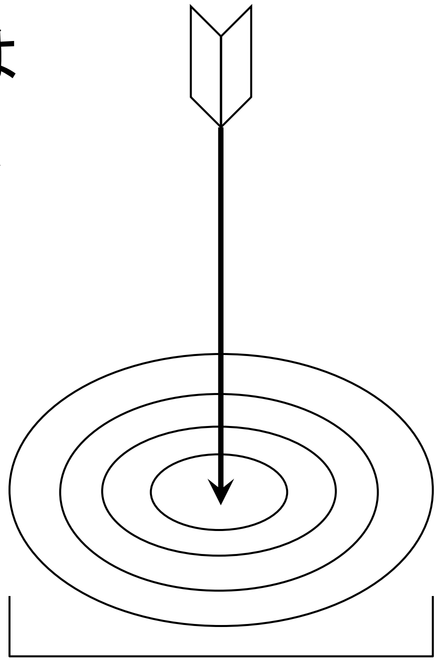
標準偏差: $\sigma / \sqrt{n}$



標準正規分布



母集団の平均体重μは
信頼度95%の確率で
この範囲内にある



この場合の有意水準は

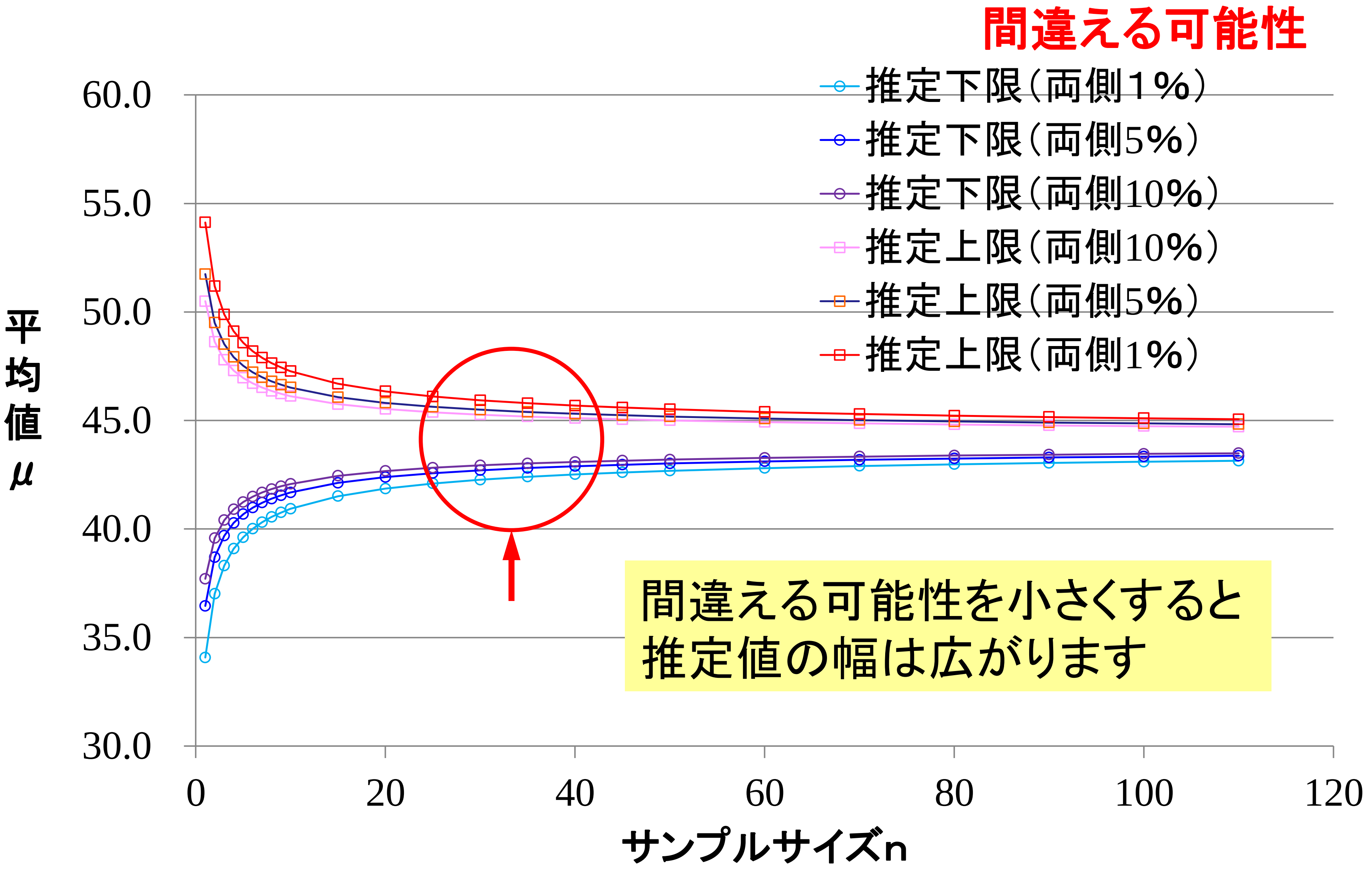
$$\frac{100-95}{100} = 0.05$$

43.3 ＜μ＜ 44.9 kg

精度＝信頼区間／2

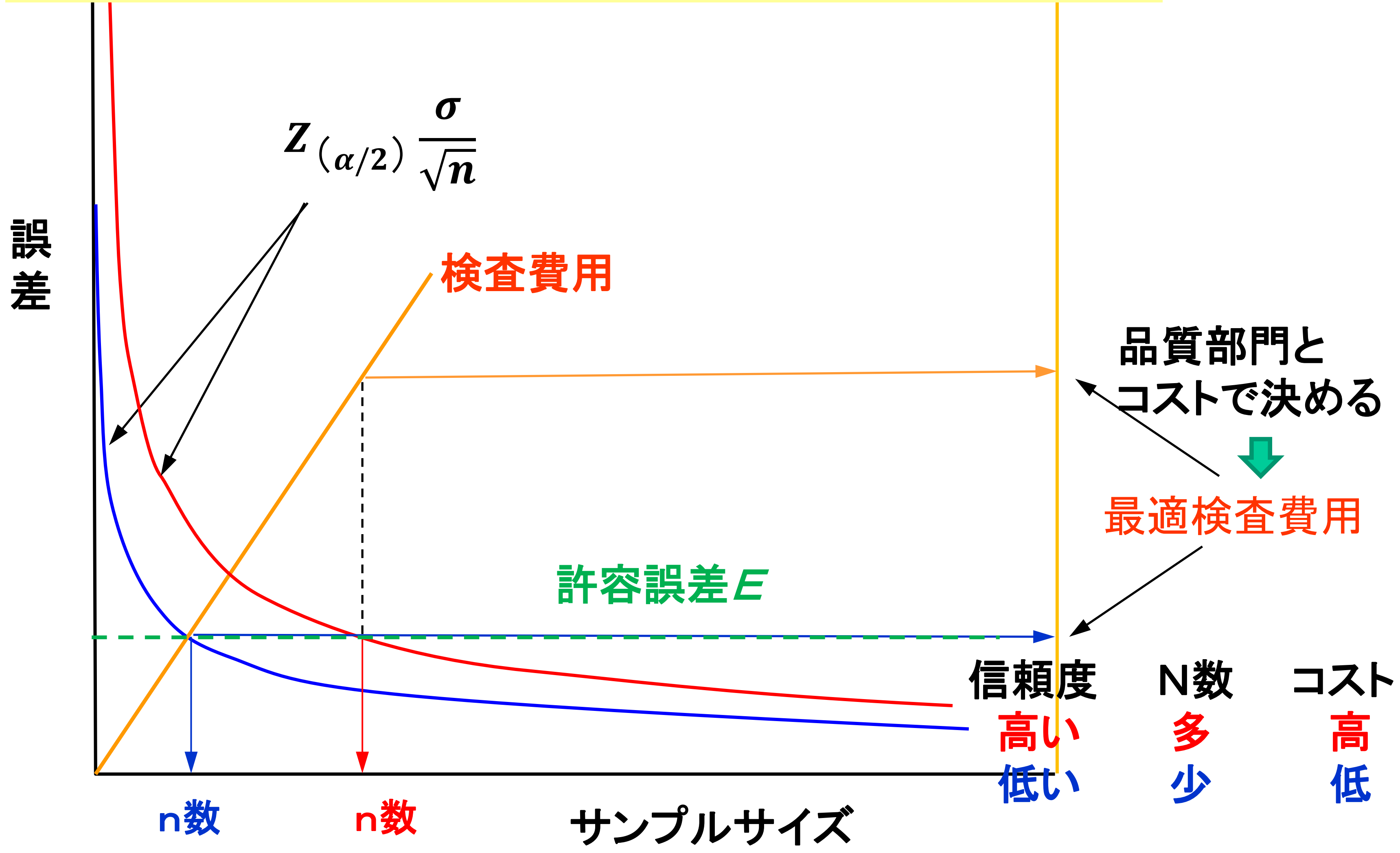
信頼区間

実験データから平均値や分散(バラつき)を推定する方法 → 統計的推定



例

サンプルサイズは以下の2つの要素で決まる
①信頼度
②コスト



$$n=\left(\frac{Z_{(\alpha/2)}\sigma}{E}\right)^2$$

社員26人の会社で、1日当り平均吸うタバコの本数を推定することにしました。25人のデータは調査でき、平均喫煙本数$\bar{x}$＝7本、標準偏差u＝4本でした。26名の平均値を推定すると？

n＜100なので、t推定。 信頼度95%とすると

$\bar{x}-t(n-1,\alpha/2)(\sigma/\sqrt{n}) < \mu < \bar{x}+t(n-1,\alpha/2)(\sigma/\sqrt{n})$ の公式に

$t(25-1,0.05/2)=2.064$ 　$u=\sigma/\sqrt{n}=4/\sqrt{25}=4/5=0.8$ を代入して

$7-2.064\times0.8 < \mu < 7+2.064\times0.8$ 　　$5.3 < \mu < 8.7$

26人の内の25人と100，000人の内の25人の調査では信頼区間が異なる

$$\bar{x}-t(n-1,\alpha/2)(\sigma/\sqrt{n})\sqrt{\frac{N-n}{N-1}} < \mu < \bar{x}+t(n-1,\alpha/2)(\sigma/\sqrt{n})\sqrt{\frac{N-n}{N-1}}$$

$$\sqrt{\frac{N-n}{N-1}}=\sqrt{\frac{26-25}{26-1}}=0.2$$

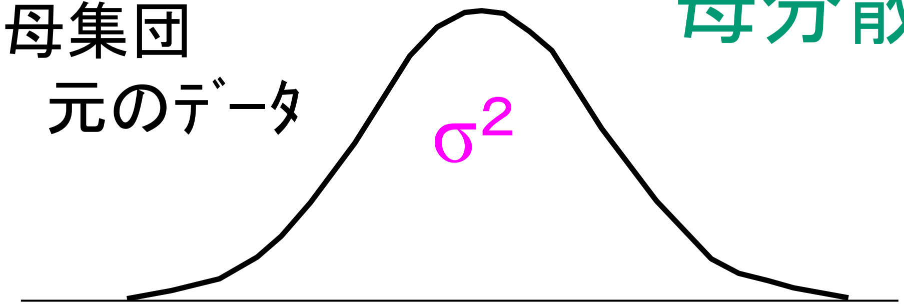
$7-2.064\times0.8\times0.2 < \mu < 7+2.064\times0.8\times0.2$

$6.7 < \mu < 7.3$

サイズが小さい有限母集団の際は補正項を乗ずる

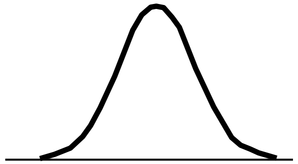
母分散の推定

母集団
元のﾃﾞｰﾀ



標本としての10人の体重を計測

67.0	64.7	52.3	59.0	54.6
59.0	58.2	52.2	52.6	48.1

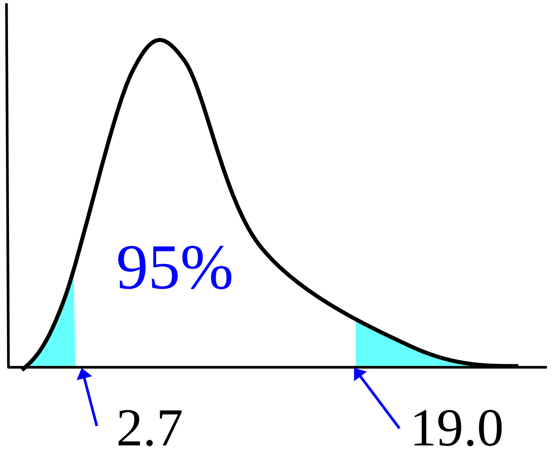


標本
サンプル数10

$$\bar{x}=\frac{67.0+64.7+\cdot\ \cdot\ \cdot+52.6+48.1}{10}=56.8$$

$$s^2=\frac{(67.0-\bar{x})^2+(64.7-\bar{x})^2+\cdot\ \cdot\ \cdot+(48.1-\bar{x})^2}{10-1}=35.4$$

Z=(n－1)s²／σ²は
自由度f=n－1のχ²分布にしたがう



Z値

$$2.7\le\boxed{\frac{(10-1)s^2}{\sigma^2}}\le19.0$$

$$\frac{9s^2}{19.0}\le\sigma^2\le\frac{9s^2}{2.7}$$

$s^2=35.4$　を代入して

$$16.7\le\sigma^2\le118.0$$