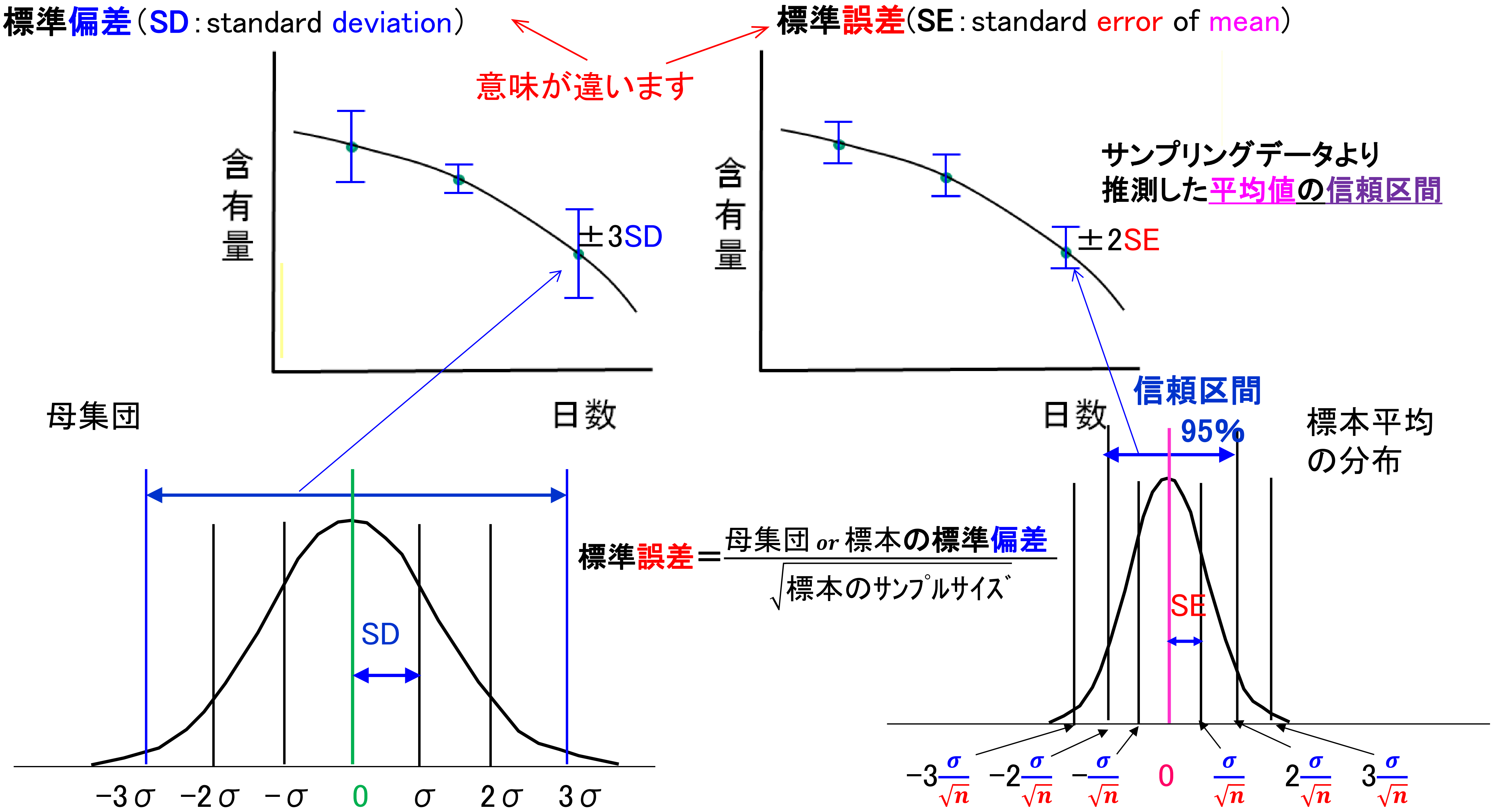
統計的推定とは？

標本の特性→母集団の特性を推定

↓

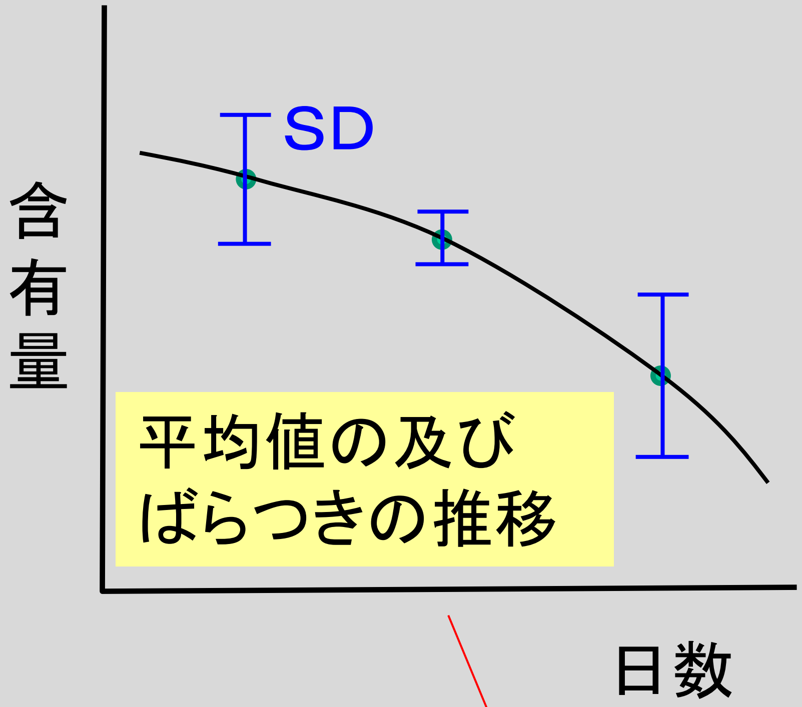
・平均値
・分散

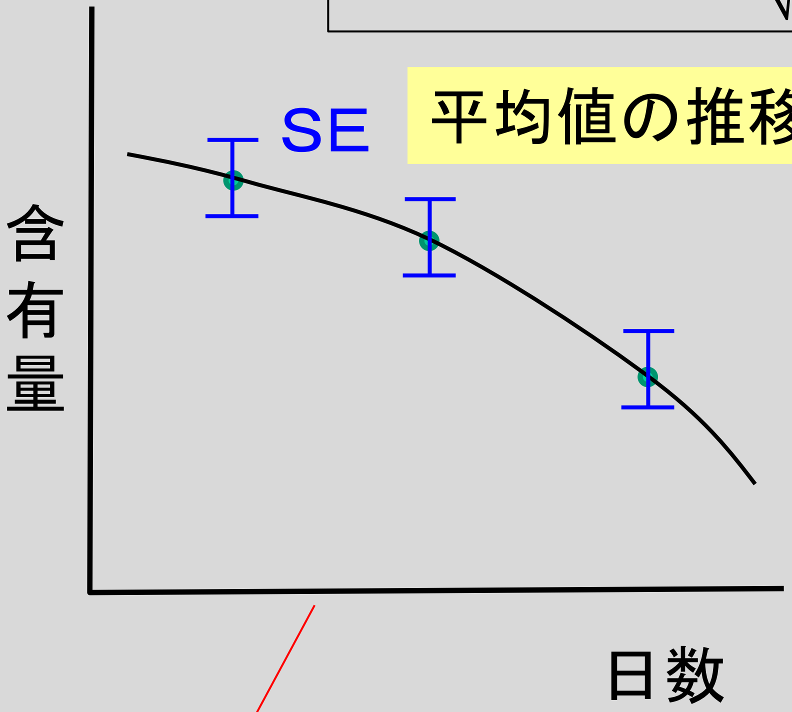


標準偏差（SD：standard deviation）
標準誤差(SE：standard error of mean)
意味が違います
含有量
±3SD
日数
含有量
サンプリングデータより
推測した平均値の信頼区間
±2SE
日数
母集団
信頼区間
95%
標本平均
の分布
標準誤差＝母集団 or 標本の標準偏差 / √標本のサンプルサイズ
SD
SE
−3σ
−2σ
−σ
0
σ
2σ
3σ
$-3\frac{\sigma}{\sqrt{n}}$
$-2\frac{\sigma}{\sqrt{n}}$
$-\frac{\sigma}{\sqrt{n}}$
0
$\frac{\sigma}{\sqrt{n}}$
$2\frac{\sigma}{\sqrt{n}}$
$3\frac{\sigma}{\sqrt{n}}$

標準偏差　と　標準誤差　の相違

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





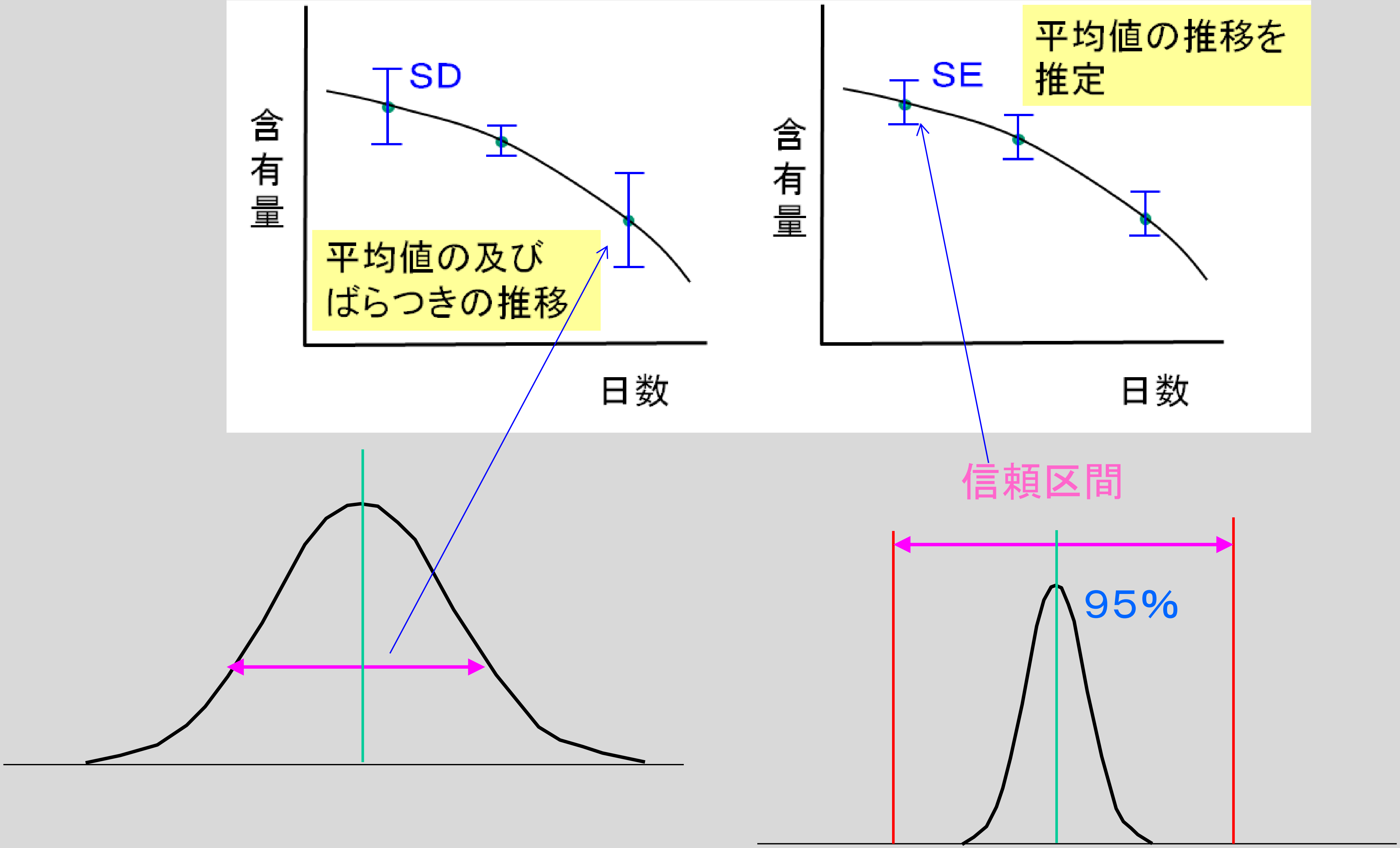
平均値の推移を推定

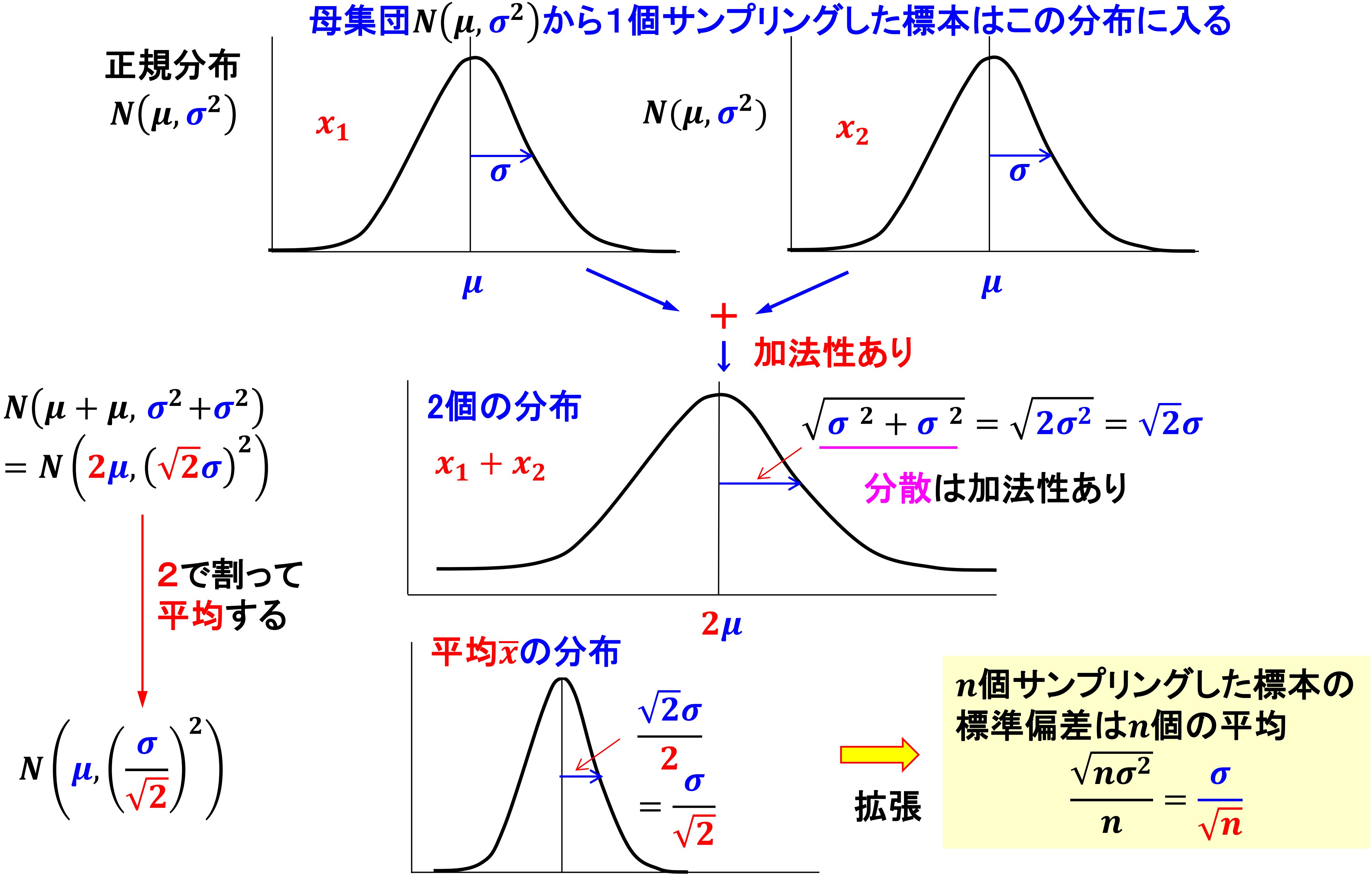
意味が違います

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

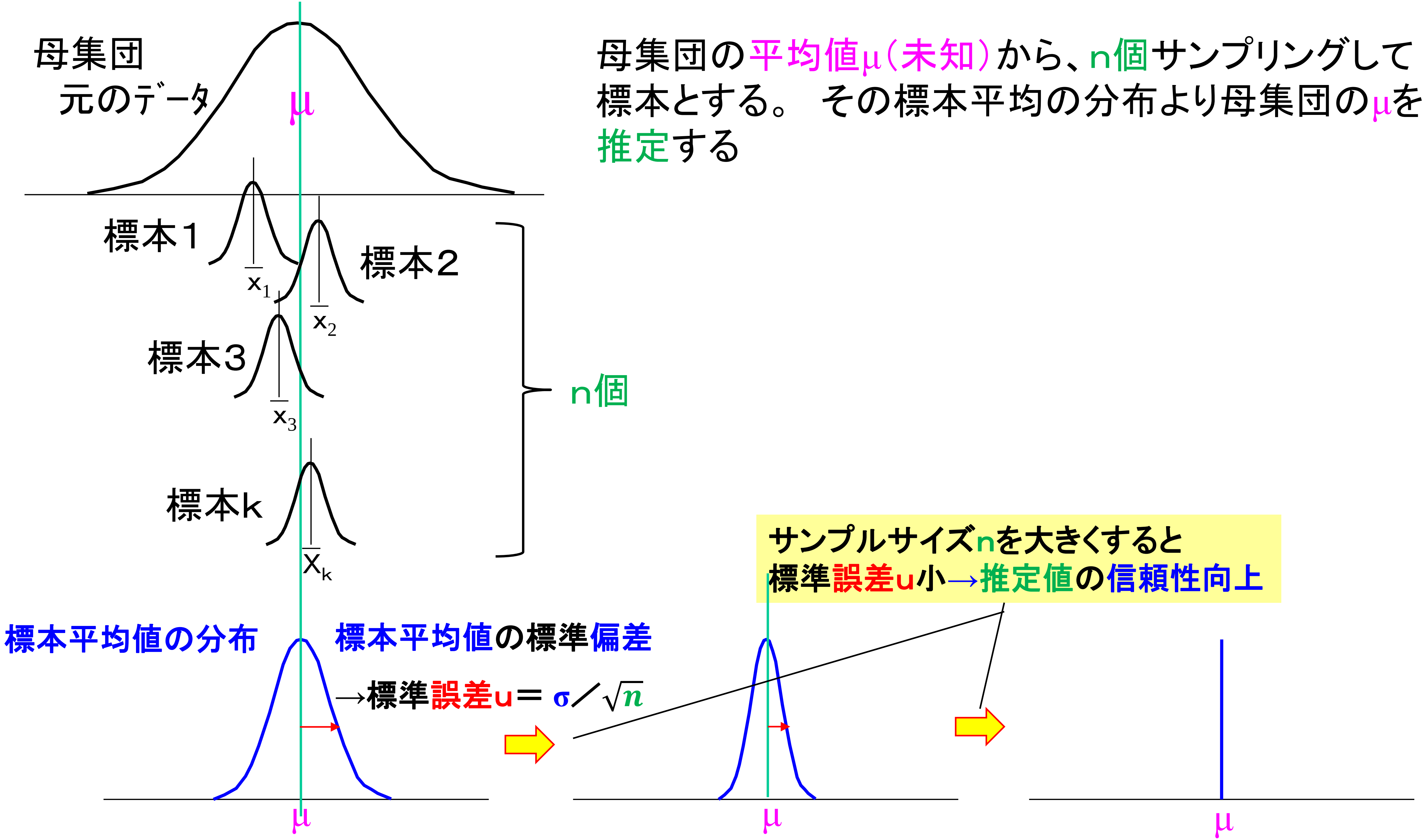
データそのもののばらつき

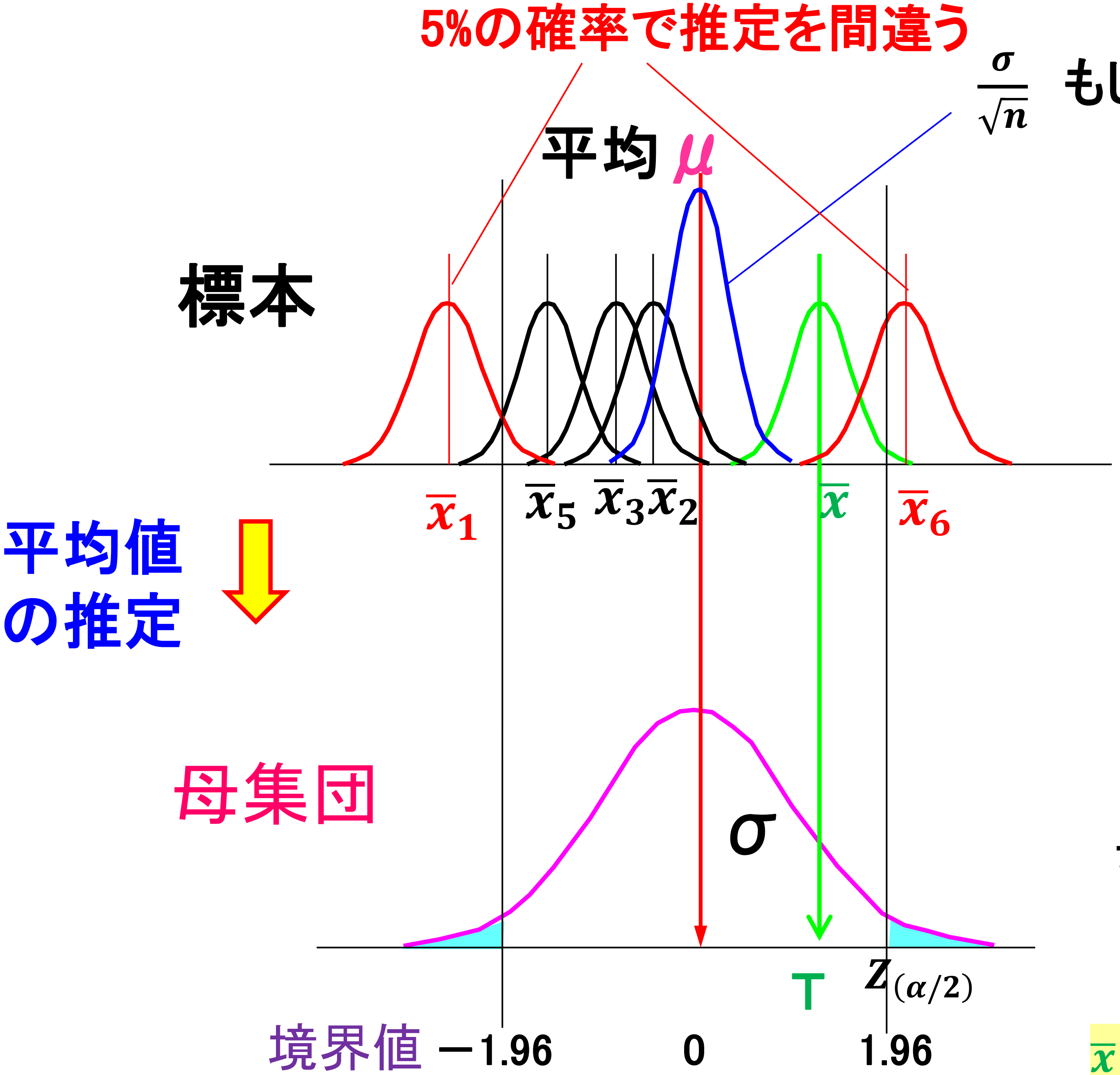
サンプリングデータより
推測した平均値の信頼区間


SD
含有量
平均値の及び
ばらつきの推移
日数
SE
平均値の推移を
推定
含有量
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
$N(\mu, \sigma^2)$
x_1
σ
μ
$N(\mu, \sigma^2)$
x_2
σ
μ
+
加法性あり
$N(\mu+\mu, \sigma^2+\sigma^2)$
$= N\left(2\mu, (\sqrt{2}\sigma)^2\right)$
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
2μ
2で割って
平均する
$N\left(\mu, \left(\frac{\sigma}{\sqrt{2}}\right)^2\right)$
平均$\bar{x}$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
母集団
元のデータ
μ
標本1
$\bar{x}_1$
標本2
$\bar{x}_2$
標本3
$\bar{x}_3$
標本k
$\bar{X}_k$
n個
標本平均値の分布
標本平均値の標準偏差
→標準誤差u＝ $\sigma／\sqrt{n}$
サンプルサイズnを大きくすると
標準誤差u小→推定値の信頼性向上
μ
μ
μ



標本平均の標準偏差

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

n： サンプル数

標本平均 $\bar{x}$ を基準化する

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

$\alpha=0.05$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

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

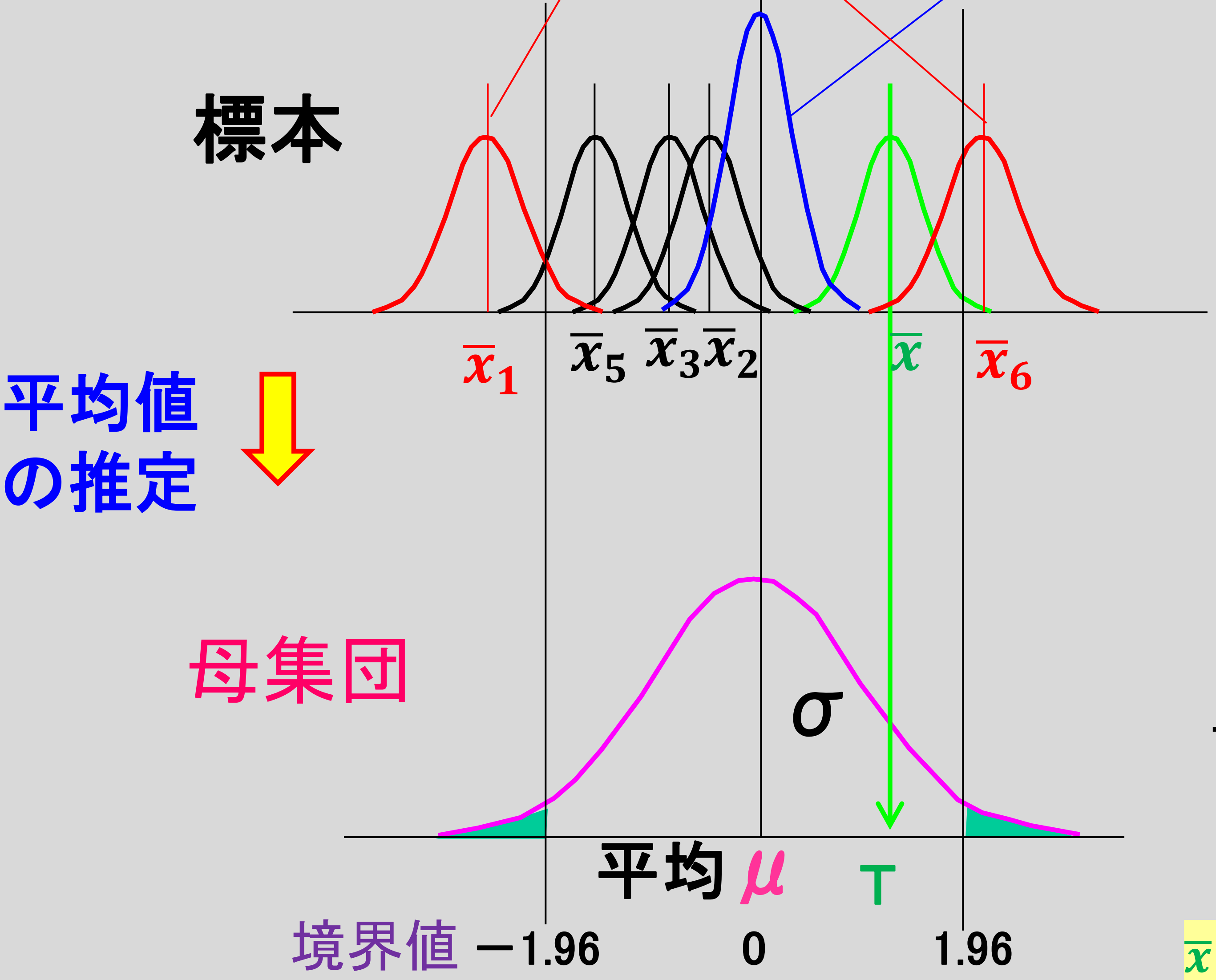
分布の推定
・性能評価
・信頼性推定

標本平均の標準偏差

5%の確率で推定を間違う

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

n： サンプル数



標本平均 $\bar{x}$ を基準化する

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

α＝0.05の時

Z(α/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、

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

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

間違えても良い確率

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

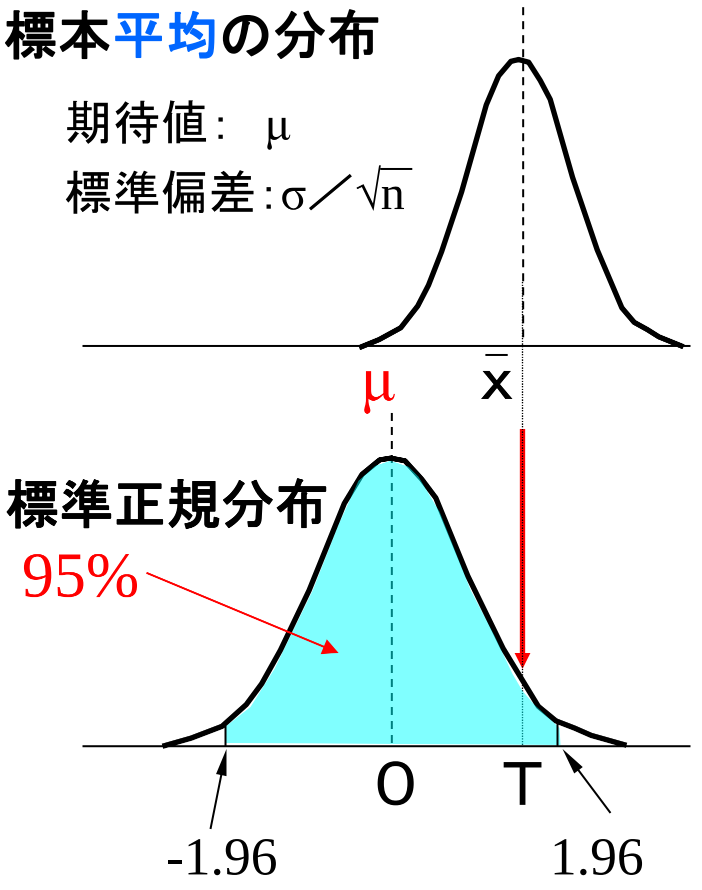
$$T=\frac{\bar{x}-\text{期待値}}{\text{標準偏差}}=\frac{\bar{x}-\mu}{\sigma／\sqrt{n}}$$

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

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



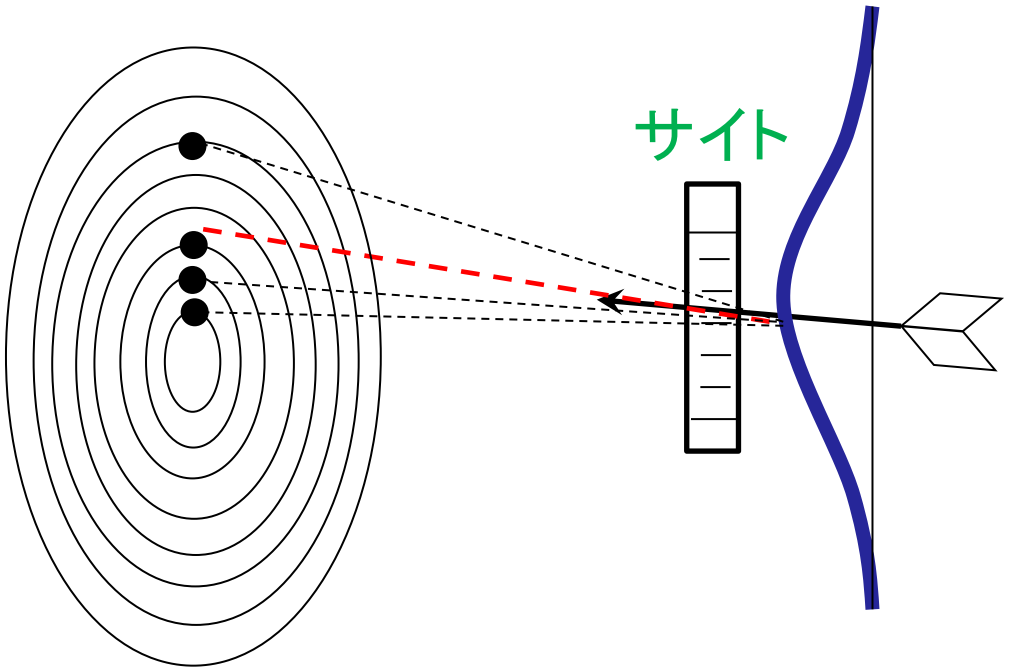
母集団の標準偏差σが未知のとき

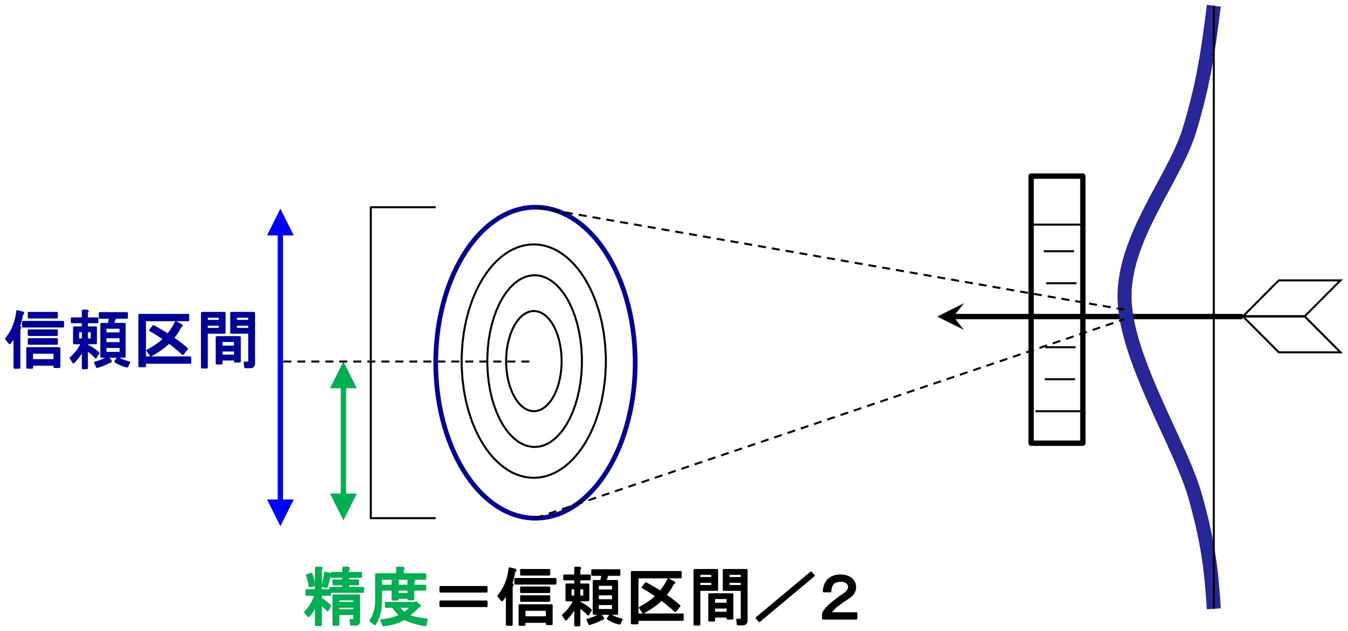
ｻﾝﾌﾟﾙｻｲｽﾞが大きければ（n≧100）、　u≒σ

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

推定のイメージ





的の中心を狙って**数本試射**して、サイトの目盛を読み取り、その**平均の目盛**を記録

⇨

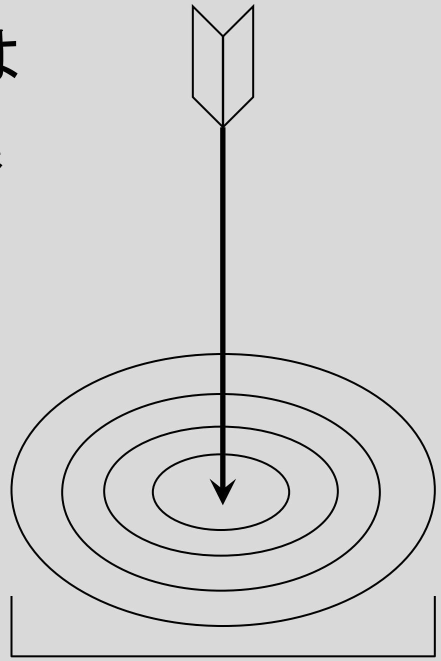
試射での読み取り平均より、的の中心を推定してサイトの目盛幅を調整する

信頼区間は、95%のように設定
5%は大きく的から外れるということ

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

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

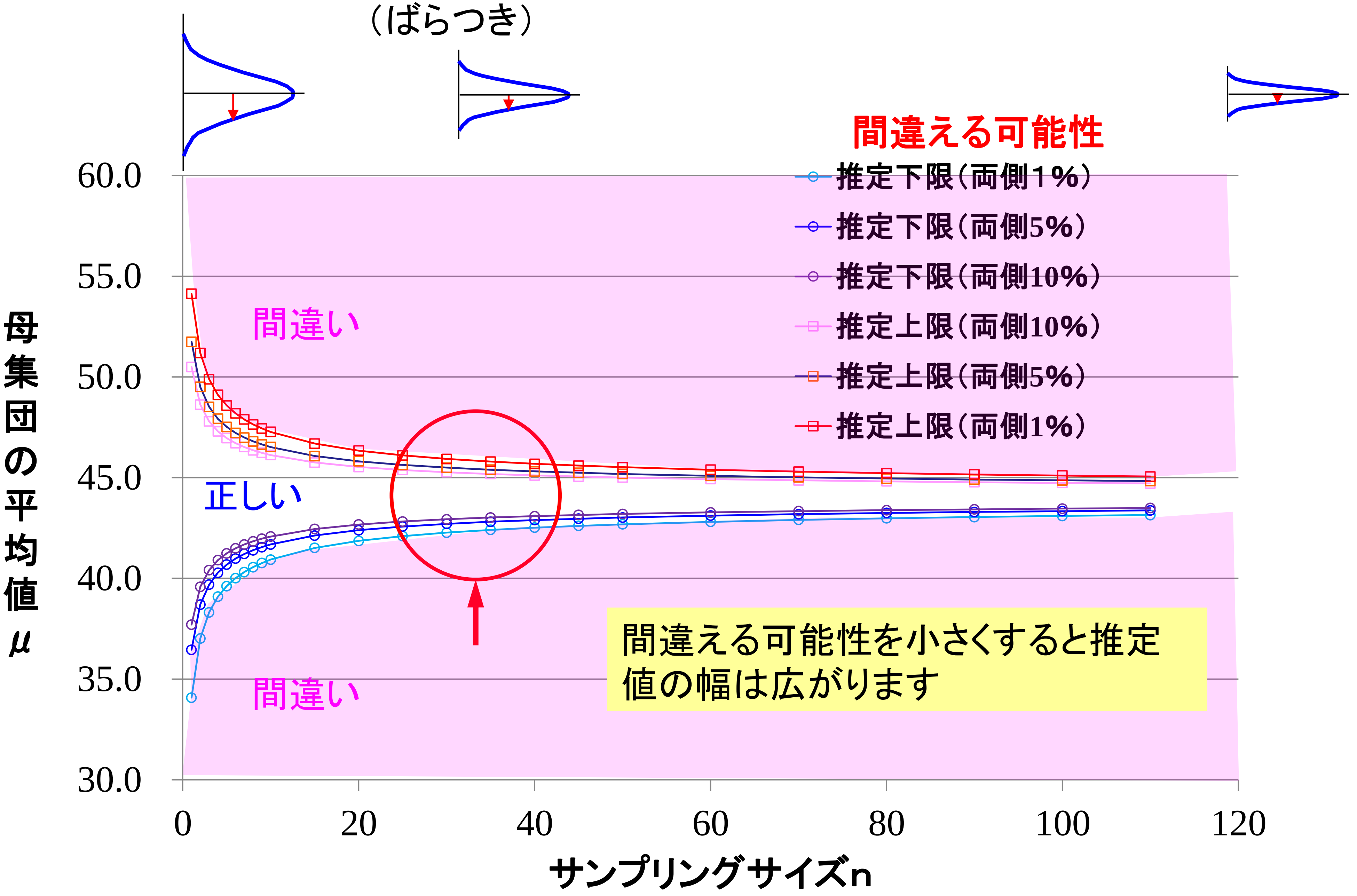


$43.3 < \mu < 44.9$ kg

精度＝信頼区間／2

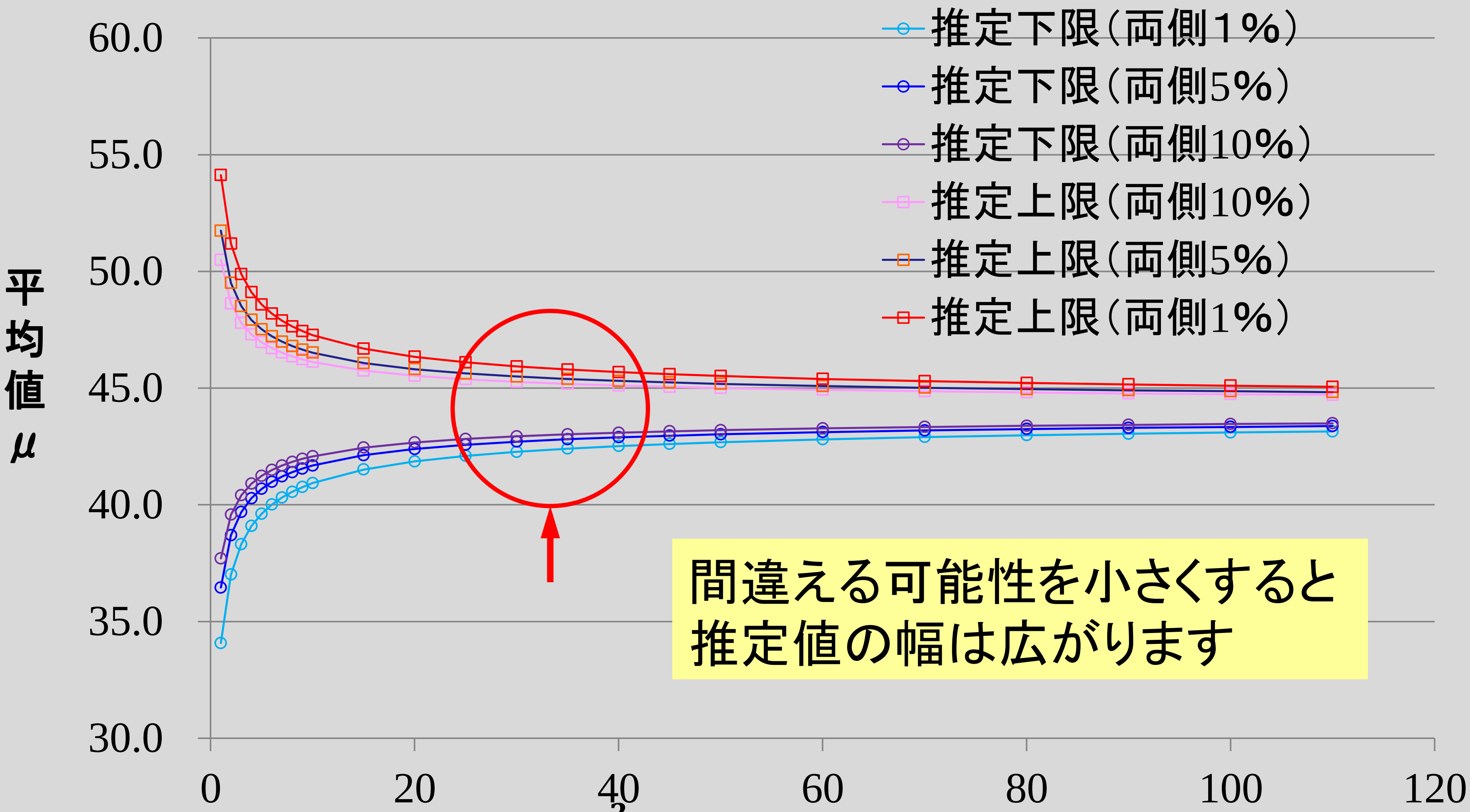
信頼区間

実験データから母集団の平均値や分散を推定する方法 → 統計的推定
（標本）　　　　　　　　　　　　（ばらつき）



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

間違える可能性

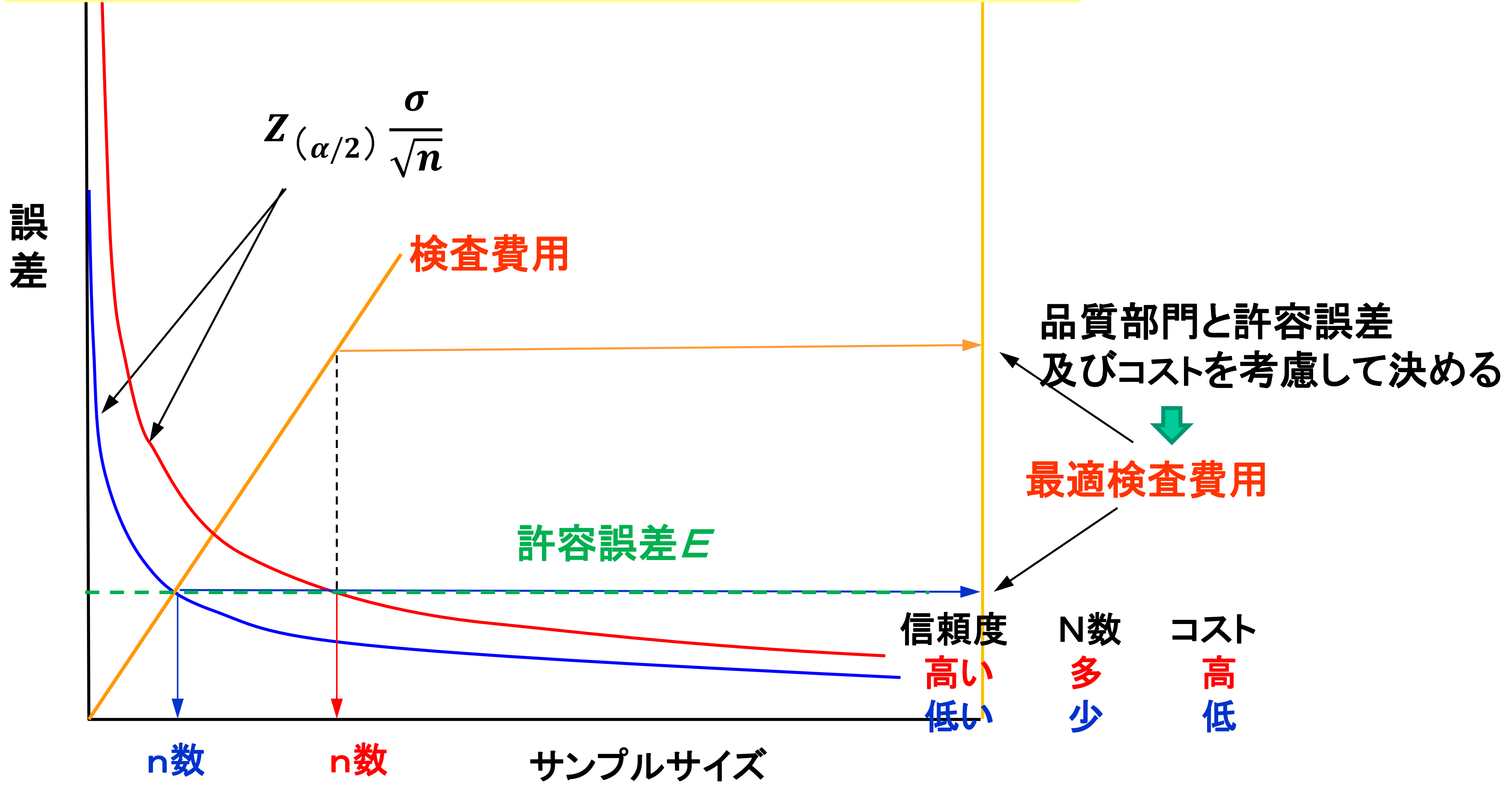


$$n = \left(\frac{Z_{(\alpha/2)}\sigma}{E}\right)^2$$

サンプルサイズn

例

サンプル数は以下の2つの要素で決まる
①信頼度
②コスト



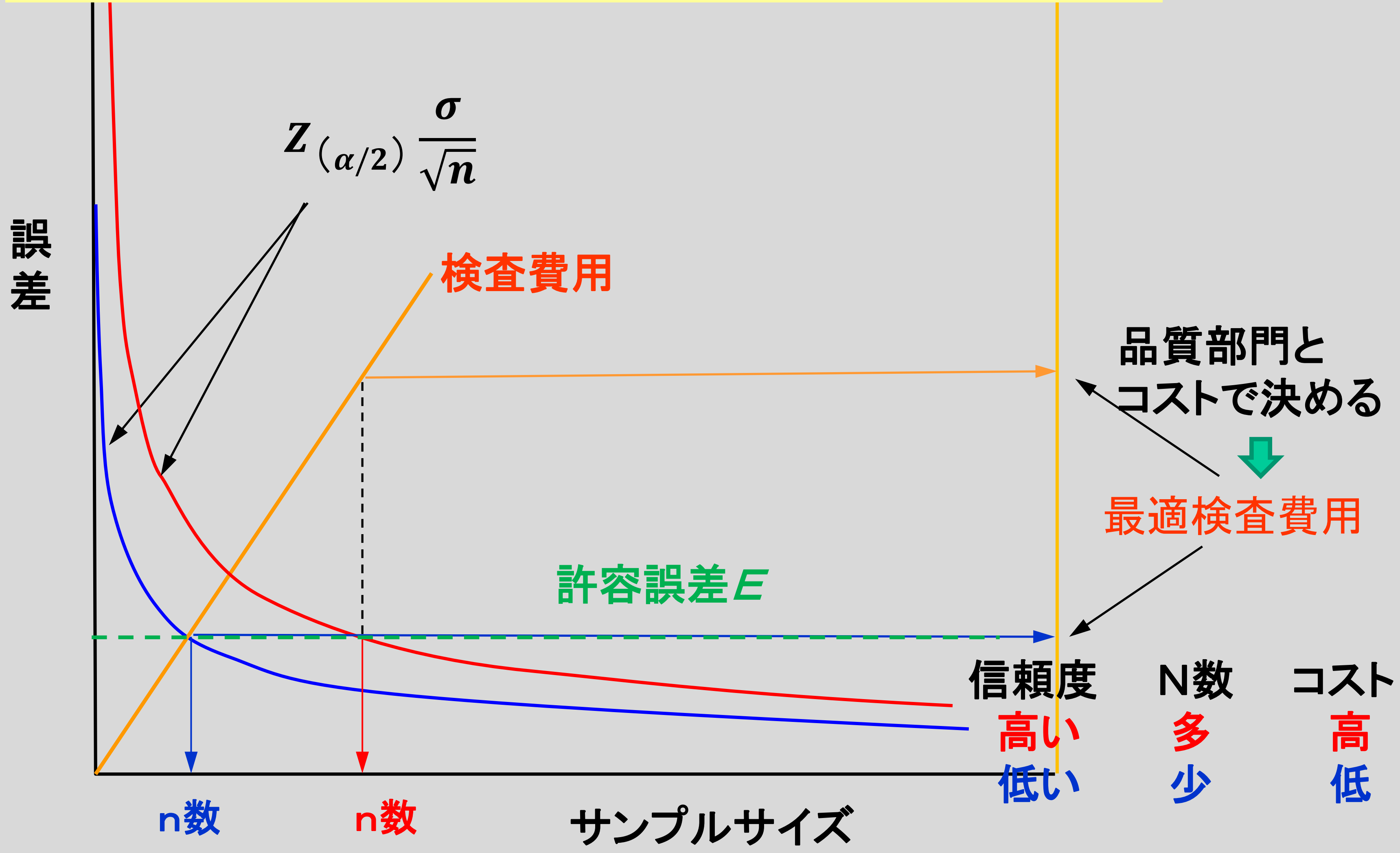
$$n=\left(\frac{Z_{(\alpha/2)}\sigma}{E}\right)^2$$

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,α/2)(σ／√n)< μ< $\bar{x}$+t(n-1,α/2)(σ／√n)の公式に

t(25-1,0.05/2)＝2.064 u＝σ／√n＝4／√25＝4／5＝0.8を代入して

7－2.064×0.8＜ μ＜ 7＋2.064×0.8 5.3 ＜ μ＜ 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

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

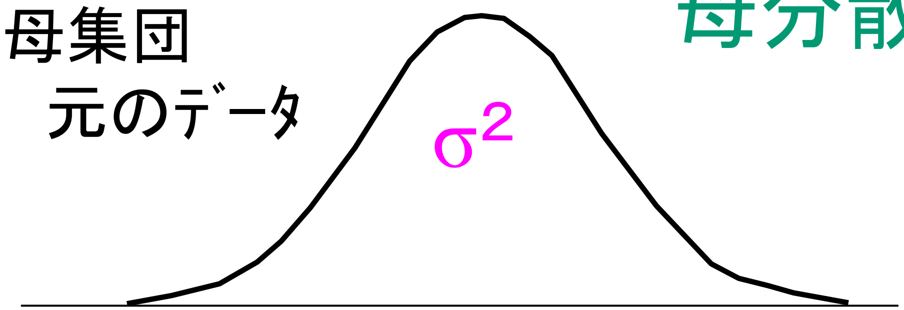
$$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{26-25}{26-1}} = 0.2$$

7－2.064×0.8×0.2＜ μ＜ 7＋2.064×0.8×0.2

6.7 ＜ μ＜ 7.3

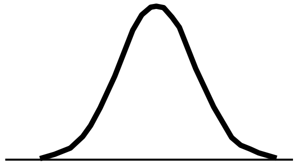
サイズが小さい有限母集団の際は補正項を乗ずる

母分散の推定



標本として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

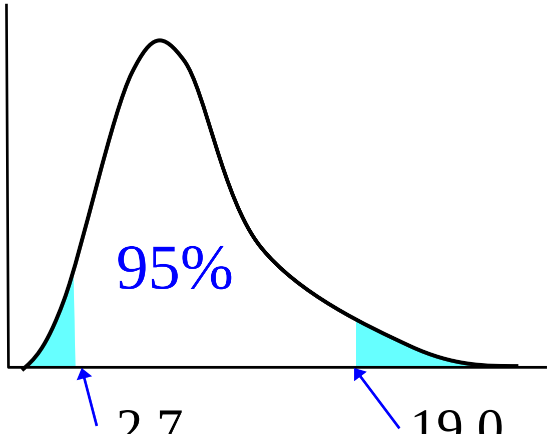


標本
サンプル数10

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

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

Z=(n－1)s²／σ²は
自由度f=n－1のχ²分布にしたがう



Z値

$$2.7 \le \boxed{\frac{(10-1)s^2}{\sigma^2}} \le 19.0$$

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

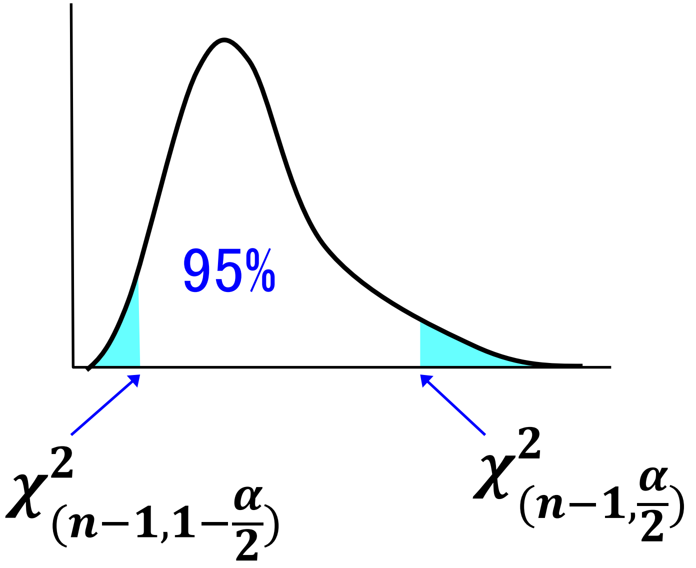
$s^2 = 35.4$ を代入して

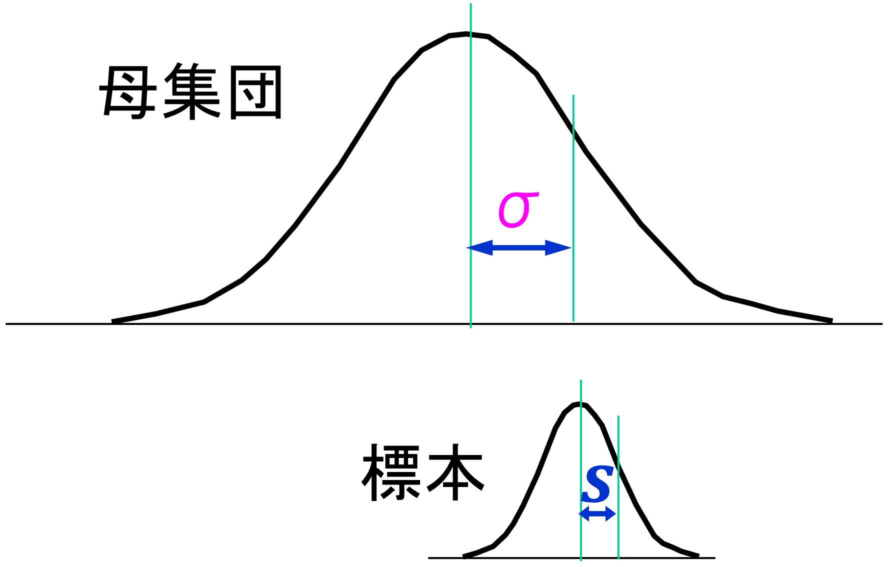
$$16.7 \le \sigma^2 \le 118.0$$

母分散の推定とサンプルサイズの関係

$\boxed{Z=(n-1)s^2/\sigma^2}$は
自由度f=n－1のχ^2分布にしたがう

α: 有意水準
n: サンプルサイズ
s: 標本の標準偏差
σ: 母集団の標準偏差
$\chi^2_{(n-1,\circ)}$: χ^2分布での境界値





$$\chi^2_{(n-1,1-\frac{\alpha}{2})} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{(n-1,\frac{\alpha}{2})}$$

$$\frac{(n-1)s^2}{\chi^2_{(n-1,\frac{\alpha}{2})}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{(n-1,1-\frac{\alpha}{2})}}$$

母集団分散の推定幅
$\Delta\sigma^2$は、

$$\Delta\sigma^2 = \frac{(n-1)s^2}{\chi^2_{(n-1,1-\frac{\alpha}{2})}} - \frac{(n-1)s^2}{\chi^2_{(n-1,\frac{\alpha}{2})}}$$

$$= (n-1)s^2\left(\frac{1}{\chi^2_{(n-1,1-\frac{\alpha}{2})}} - \frac{1}{\chi^2_{(n-1,\frac{\alpha}{2})}}\right)$$

$$\frac{\Delta\sigma^2}{s^2} = (n-1)\left(\frac{1}{\chi^2_{(n-1,1-\frac{\alpha}{2})}} - \frac{1}{\chi^2_{(n-1,\frac{\alpha}{2})}}\right)$$

$$\frac{\Delta\sigma^2}{s^2} = (n-1)\left(\frac{1}{\chi^2_{(n-1,1-\frac{\alpha}{2})}} - \frac{1}{\chi^2_{(n-1,\frac{\alpha}{2})}}\right)$$

n	$\chi^2_{(n-1,0.975)}$	$\chi^2_{(n-1,0.025)}$	$\Delta\sigma^2/s^2$
2	0.00	5.02	1018.06
3	0.05	7.38	39.23
5	0.48	11.14	7.90
10	2.70	19.02	2.86
15	5.63	26.12	1.95
20	8.91	32.85	1.55
25	12.40	39.36	1.33
30	16.05	45.72	1.17
40	23.65	58.12	0.98
50	31.55	70.22	0.86
60	39.66	82.12	0.77
70	47.92	93.86	0.70
80	56.31	105.47	0.65
90	64.79	116.99	0.61
100	73.36	128.42	0.58
110	82.00	139.78	0.55
120	90.70	151.08	0.52