散らばりの尺度

標本分散の定義

<u>分散</u>：分布の広がりに関する情報

（平均はこの情報を持っていない）

<u>標本分散</u>：偏差の二乗和を自由度で割ったもの

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}$$

次式を用いると計算が楽になる場合がある

$$s^2 = \frac{1}{n-1}\left\{(x_1^2 + x_2^2 + \cdots + x_n^2) - \frac{1}{n}(x_1 + x_2 + \cdots + x_n)^2\right\}$$

例：標本分散を求める

健康な20歳の女性8名の最高血圧の測定値

106， 98， 116， 96， 100， 112， 100， 102

- 標本分散を求める

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1}$$

  - 手順1：標本平均を求める
  - 手順2：標本分散を求める

例：標本分散を求める

- データ

  106, 98, 116, 96, 100, 112, 100, 102

- 標本平均

$$\bar{x} = \frac{106 + 98 + \cdots + 102}{8} = 104$$

- 標本分散

$$s^2 = \frac{(106 - 104)^2 + (98 - 104)^2 + \cdots + (102 - 104)^2}{8 - 1} \fallingdotseq 49.7$$

簡便な推定法(標本平均)

- 標本数が大きい場合、計算が煩雑になる
  ⇒ 簡便推定法
  - 手順１ ： 度数分布表を作成する
  - 手順２ ： 度数分布表から標本平均を求める
  - 手順３ ： 度数分布表から標本分散を求める

- 例 ： 11ヶ月以上で生まれた初生児の体重の度数分布表
  - この度数分布表から、標本平均、標本分散を推定する

級 (以上) – (未満)	級代表値	度数
1.0-1.5	1.25	1
1.5-2.0	1.75	1
2.0-2.5	2.25	12
2.5-3.0	2.75	90
3.0-3.5	3.25	217
3.5-4.0	3.75	124
4.0-4.5	4.25	52
4.5-5.0	4.75	8
5.0-5.5	5.25	2
計		507

例：簡便推定法（標本分散）

- 標本平均の簡便推定法 （精度は落ちるが計算が容易）

$$
\begin{aligned}
\text{標本平均} &= \frac{\text{各階級についての (級代表値} \times \text{度数) の和}}{\text{標本数}} \\
&= \frac{1 \times 1.25 + 1 \times 1.75 + 12 \times 2.25 + \cdots + 8 \times 4.75 + 2 \times 5.25}{507} \\
&= 3.39
\end{aligned}
$$

級 （以上）－（未満）	級代表値	度数
1.0-1.5	1.25	1
1.5-2.0	1.75	1
2.0-2.5	2.25	12
2.5-3.0	2.75	90
3.0-3.5	3.25	217
3.5-4.0	3.75	124
4.0-4.5	4.25	52
4.5-5.0	4.75	8
5.0-5.5	5.25	2
計		507

例：簡便推定法

- 標本平均の簡便推定法（精度は落ちるが計算が容易）

$$
\text{標本分散} = \frac{\text{各階級についての}((\text{級代表値} - \text{標本平均}) \times \text{度数})\text{の和}}{\text{標本数}}
$$

$$
= \frac{1 \times (1.25 - 3.39)^2 + 1 \times (1.75 - 3.39)^2 + \cdots + 2 \times (5.25 - 3.39)^2}{507}
$$

$$
= 0.277
$$

級 （以上）–（未満）	級代表値	度数
1.0-1.5	1.25	1
1.5-2.0	1.75	1
2.0-2.5	2.25	12
2.5-3.0	2.75	90
3.0-3.5	3.25	217
3.5-4.0	3.75	124
4.0-4.5	4.25	52
4.5-5.0	4.75	8
5.0-5.5	5.25	2
計		507

標本分散の分布

例：標本分散の分布

日本人の18歳の女性の身長：正規分布

- 無作為に5人を抽出し、平均と分散を推定
  - 1回目　151.1, 160.9, 154.3, 161.2, 152.5

    標本平均：156.0,　標本分散：22.55
  - 2回目　161.8, 157.1, 159.3, 148.4, 165.9

    標本平均：158.5,　標本分散：42.57

⋮

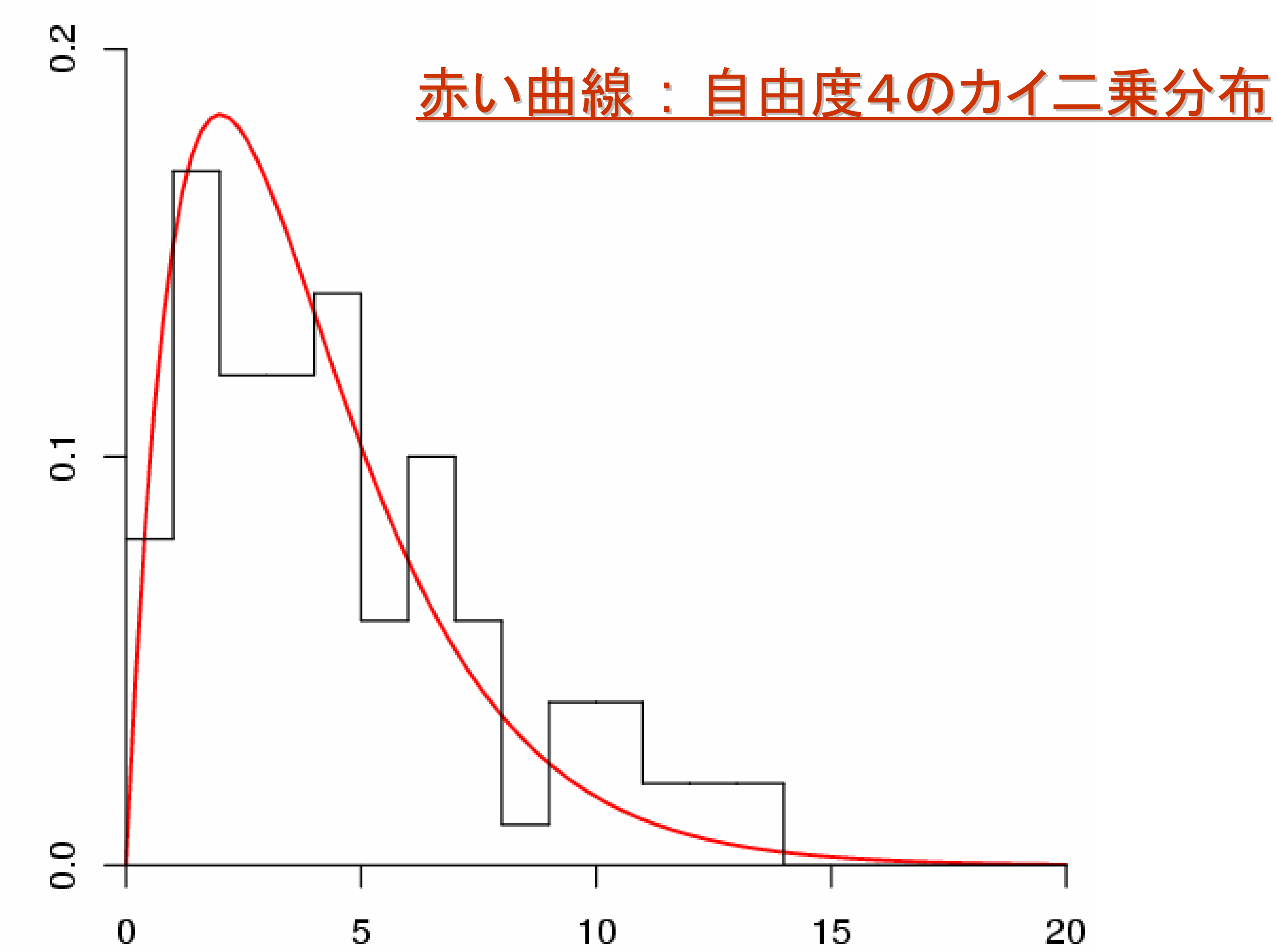
標本分散はどのような分布にしたがうか？

標本分散の分布

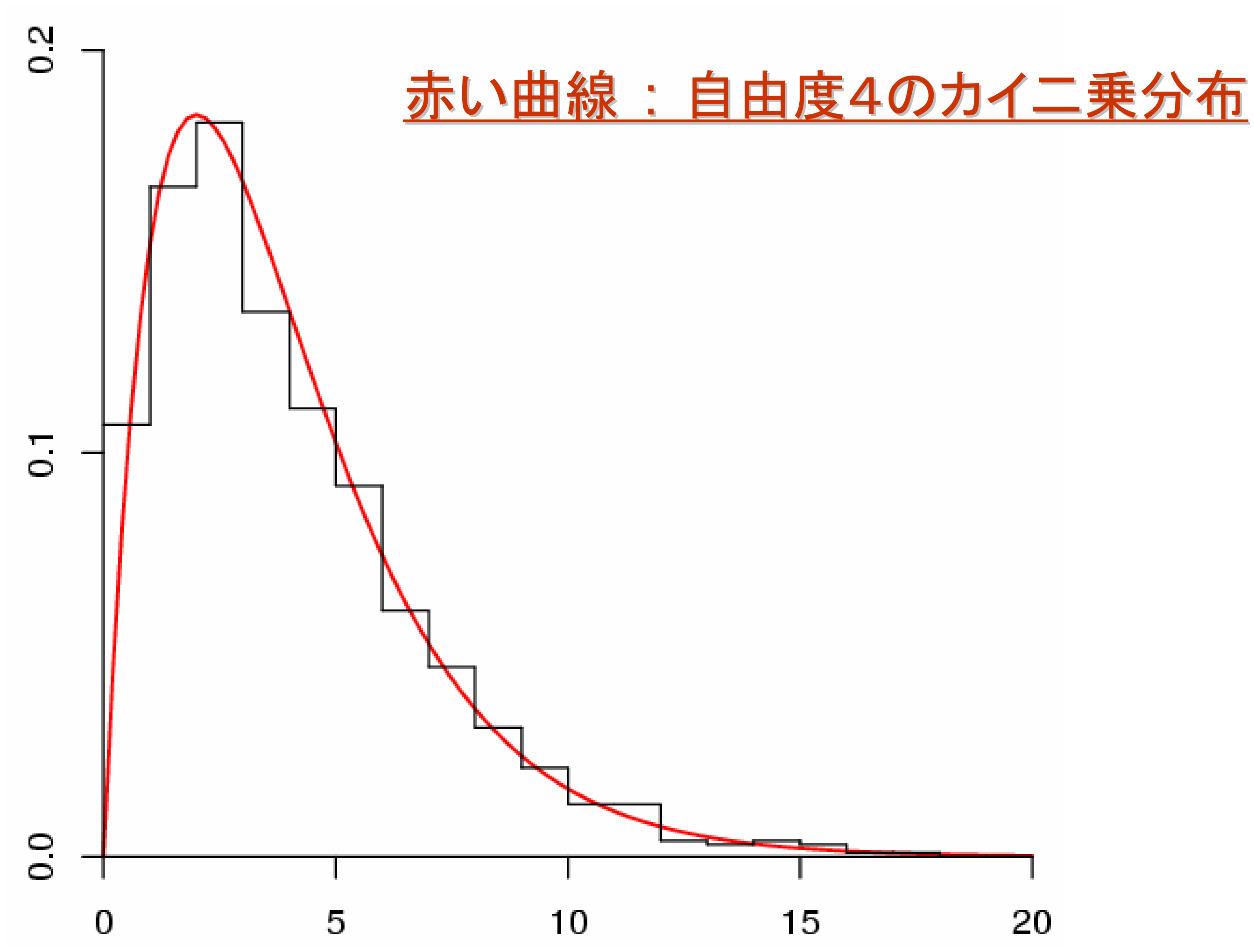
- 次の統計量が，自由度n-1のカイ二乗分布にしたがうことが知られている
  - $$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2}$$

- シミュレーション
  - 平均0，分散25の正規分布にしたがう乱数を5つ生成し，上のカイ二乗統計量を求める
  - この操作を繰り返し行い，ヒストグラムを描く
    - 繰り返し回数：100，1000，10000

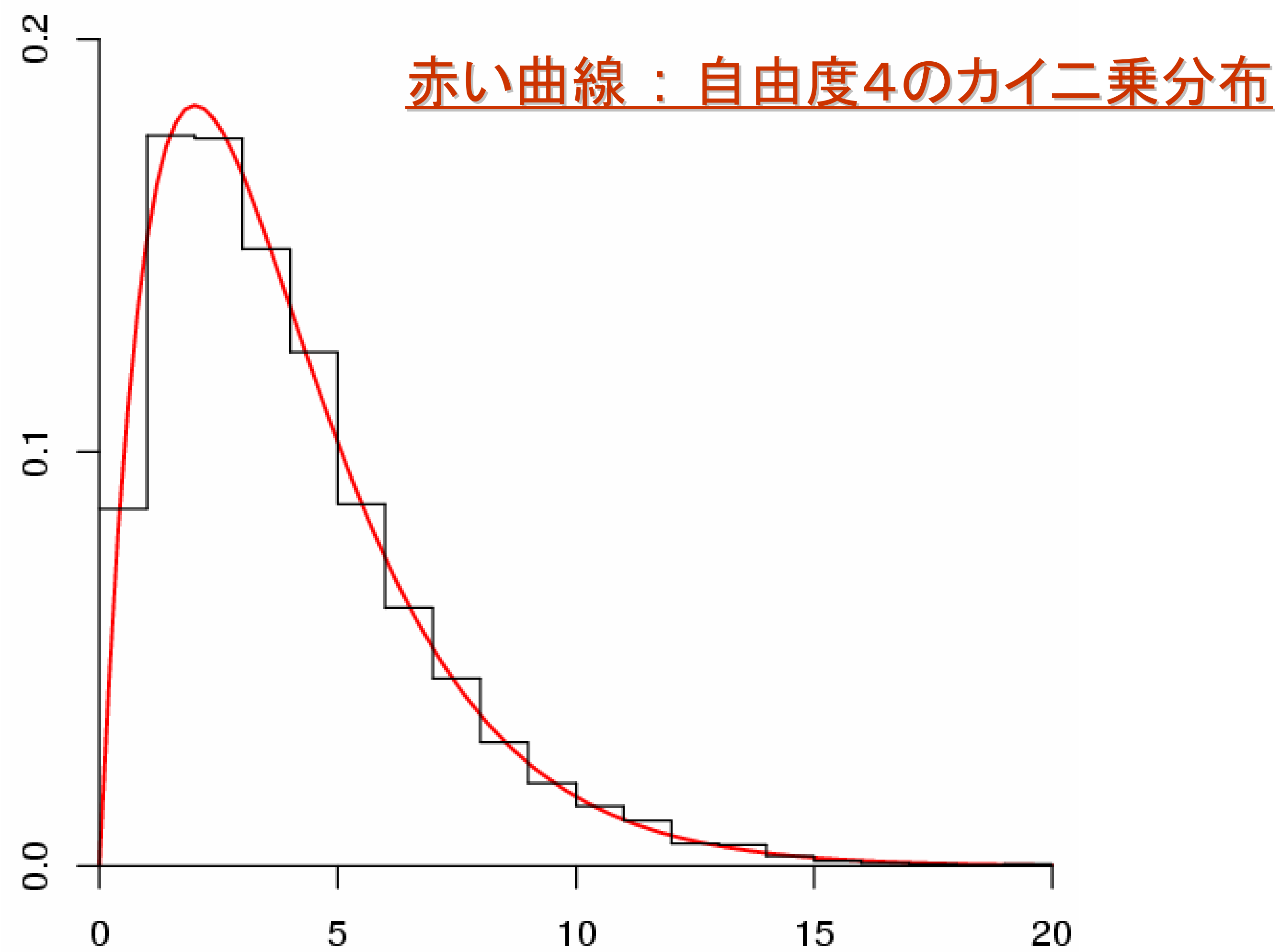
シミュレーション：100回の繰り返し



シミュレーション：1000回の繰り返し

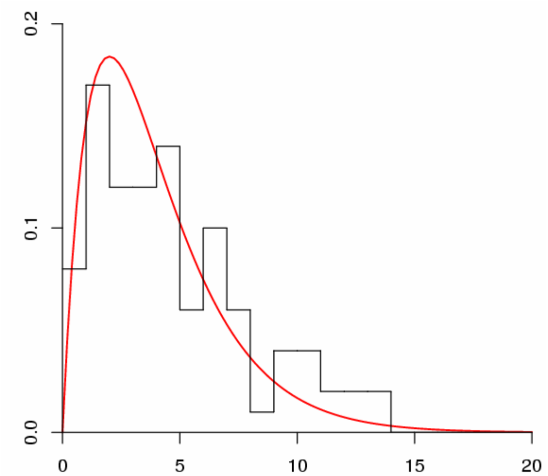


シミュレーション：10000回の繰り返し

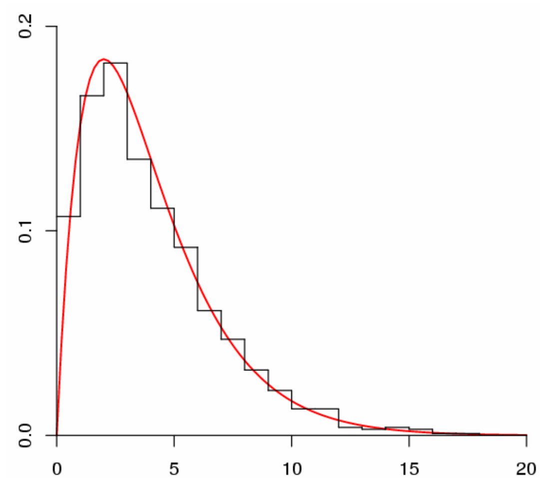


シミュレーション：標本分散の分布

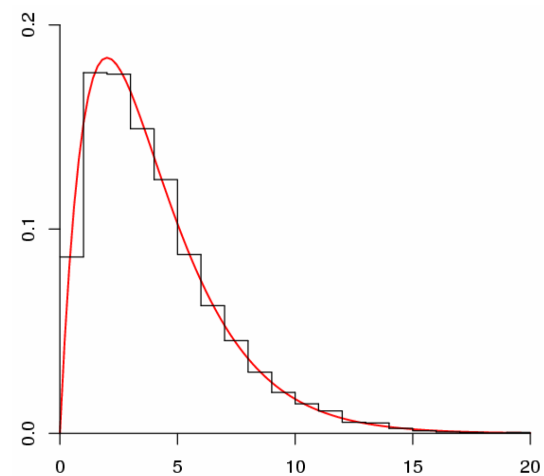
100回



1000回

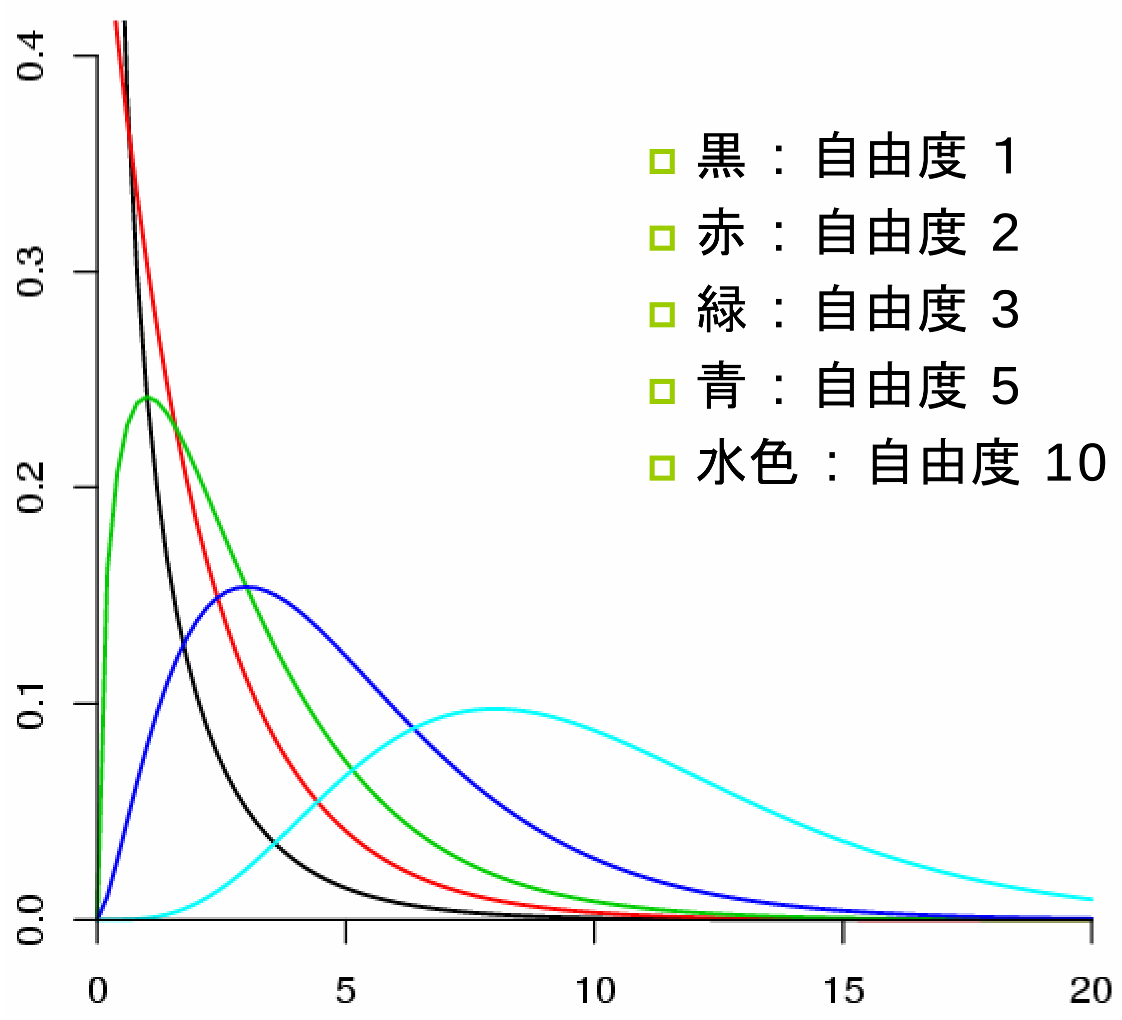


10000回



自由度 n-1 のカイ二乗分布に近づく

カイ二乗分布の密度曲線



- 黒：自由度 1
- 赤：自由度 2
- 緑：自由度 3
- 青：自由度 5
- 水色：自由度 10

変動係数

変動係数

尺度の異なる値の変動の大きさの比較

- 例1：18歳の女性の身長
  標本平均　157cm　標準偏差　5.0cm
- 例2：18歳の女性の体重
  標本平均　52kg　　標準偏差　6.1kg

測定単位が異なるものは、そのままでは比較できない

- 測定値の安定性について，変動係数を用いて比較できる

$$\text{変動係数} = \frac{\text{標準偏差}}{\text{標本平均}} \times 100$$

変動係数による比較

- 変動係数による比較

$$変動係数 = \frac{標準偏差}{標本平均} \times 100$$

- 測定値1：18歳の女性の身長
  標本平均　157cm 標準偏差　5.0cm
  変動係数　3.2%

- 測定値2：18歳の女性の体重
  標本平均　52kg　標準偏差　6.1kg
  変動係数　11.7%

体重は身長と比べて不安定な指標と考えられる