



## 2 1 組のデータの解析

### 2.1 データの特徴の記述

#### テキスト

芳賀敏郎（2011）医薬品開発のための統計解析

第1部 基礎 改訂版、サイエンティスト社、p.275



# 第1部 基礎

---

- 1. 統計の基礎 . . . . .
  - 1.1 宝くじの期待値と分散、1.2 サイコロの目の数の期待値と分散
  - 1.3 分散の加法性・中心極限定理・正規分布、1.4 統計的推測、1.5 モデル
- 2. **1組のデータの解析** **2.1 データの特徴の記述**、2.2 データのグラフ表示と外れ値
  - 2.3 対数変換と対数正規分布、2.4 平均に関する推測（母標準偏差  $\sigma$  既知）
  - 2.5 分散に関する推測、2.6 平均に関する推測（母標準偏差  $\sigma$  未知）
- 3. 2組のデータの解析 **3.1 データのグラフ化**、3.2 平均値の差の  $t$  検定、3.3 分散の違いの検定
  - 3.4 分散が異なる場合の平均値の差の比較
  - 3.5 対応のある場合の平均値の差の  $t$  検定、3.6 検出力と  $n$  の決め方
  - 3.7 ノンパラメトリック検定
- 4. 相関・回帰 . . . . .
  - 4.1 散布図、4.2 相関係数、4.3 回帰モデルとモデルの推定
  - 4.4 誤差を考慮した推定、4.5 回帰分析適用上の諸問題



## 2.1 データの特徴の記述

p.53

- (1) 中央値
- (2) 平均値
- (3) 平方和
- (4) 平均平方、標準偏差と変動係数
- (5) 間隔尺度と比例尺度
- (6) 四分位値と四分位範囲
- (7) ひずみ と とがり
- (8) 頑健性
- (9) JMPによる解析

テキストの  
該当ページ

★プレゼンテーションの  
スピーカーノートを、  
PDFの注釈に変換してあります

使用するファイル

Excel ファイル「基本改2.xls」

JMPファイル「2-演習.jmp」

サイエンティスト社のホームページ  
からダウンロード

JMP 10.0.2 の出力を表示

# はじめに

## ●本章「1組のデータの解析」の内容

連続変数： $n$ 個の観測値（連続変数）  $x_1, x_2, x_3, \dots, x_{n-1}, x_n$   
母集団から抽出されたサンプル（標本）

基本統計量：サンプルのデータの特徴を表す数値、母数を推測

代表値：中央値、平均値

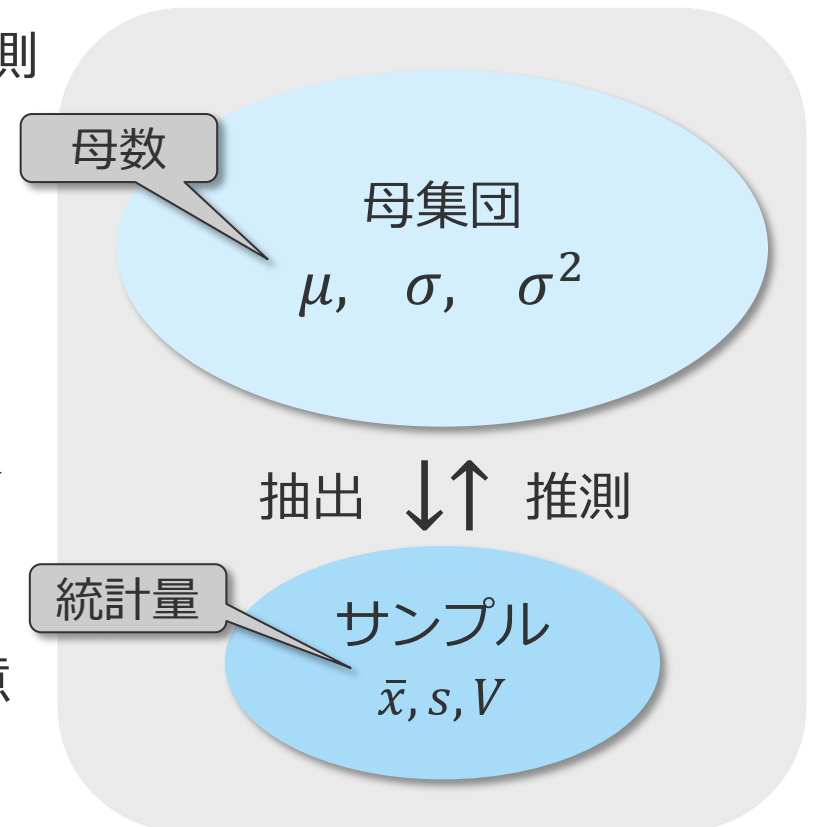
ばらつき：分散、標準偏差、四分位範囲など

### 記号

母数：母平均  $\mu$ 、母標準偏差  $\sigma$ 、母分散  $\sigma^2$

統計量：平均  $\bar{x}$ 、標準偏差  $s$ 、平均平方（分散） $V$   
平方和  $S$

注) 母平均などの「母」を省略する場合があるので注意  
母集団とサンプル（標本）を意識して区別する



# はじめに

---

## ●本節の事例

事例 1 : 8都市から1人ずつ一か所に集まる場合、どこに集まると移動距離が最小になるか  
母集団とサンプル（標本）という関係ではない  
中央値、平均値、分散、標準偏差の計算方法を理解するための単純な数値の事例





## (1) 中央値

代表値：中央値、平均値

## ●事例 1 : 8都市から1人ずつ1か所に集まる最短の集合場所

課題 8都市から1人ずつ一か所に集まる場合、全員の移動距離が最小になる地点は？

起点からの距離 東京 1、横浜 2、静岡 3、名古屋 4、京都 5、大阪 7、神戸 10、広島 18

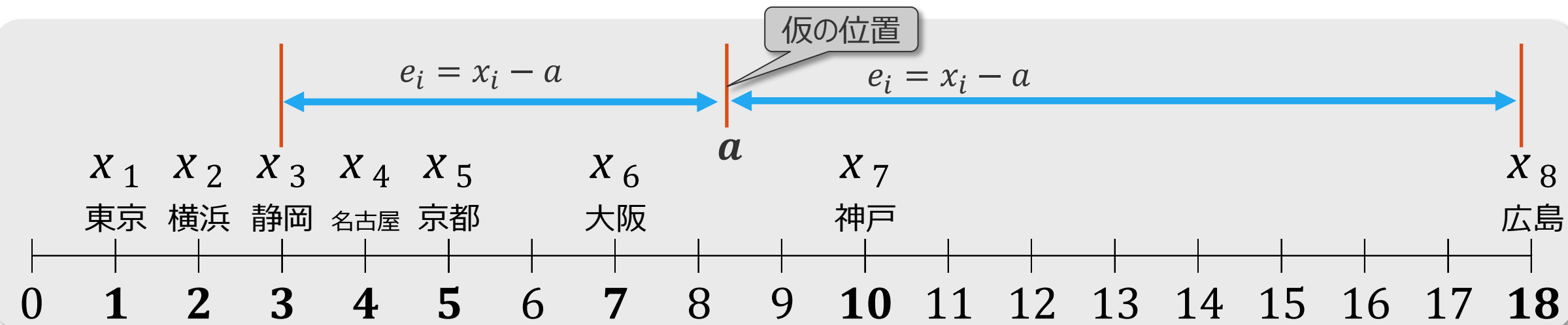
距離の表し方

単位なしの切りのいい数値

起点から各都市までの距離  $x_i$  ( $i = 1, 2, \dots, 8$ )  $x_1 = 1, x_2 = 2, \dots, x_7 = 10, x_8 = 18$

起点から集合場所までの距離  $a$  都市の間でも可、例えば 5.4

各都市と集合場所までの距離  $e_i = x_i - a$   $e_i$  が全体として最小になる  $a$  を決める



## ●事例 1 : 8都市から1人ずつ1か所に集まる最短の集合場所

起点から各都市の距離  $x_i$  . . . 観測値 (一般の場合に拡張するための表現)

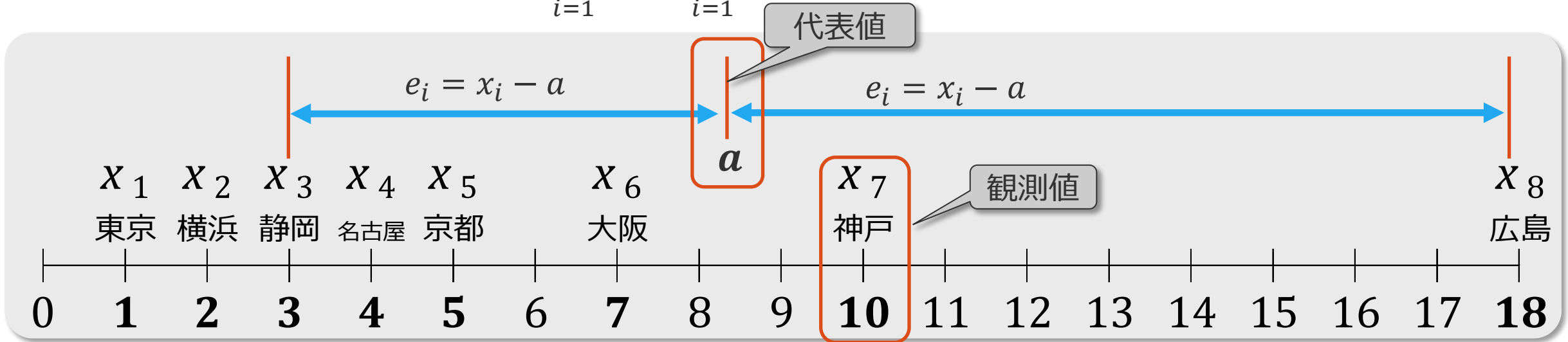
集合場所  $a$  . . . 代表値 (一般の場合に拡張するための表現)

移動距離の絶対値  $|e_i| = |x_i - a|$

移動距離の絶対値の合計  $T$

$$T = \sum_{i=1}^n |e_i| = \sum_{i=1}^n |x_i - a| \Rightarrow \min \quad (2.1.1)$$

$T$ が最小になる  $a$



## ●事例 1 : 8都市から1人ずつ1か所に集まる最短の集合場所

起点から各都市の距離  $x_i$  . . . 観測値 (一般の場合に拡張するための表現)

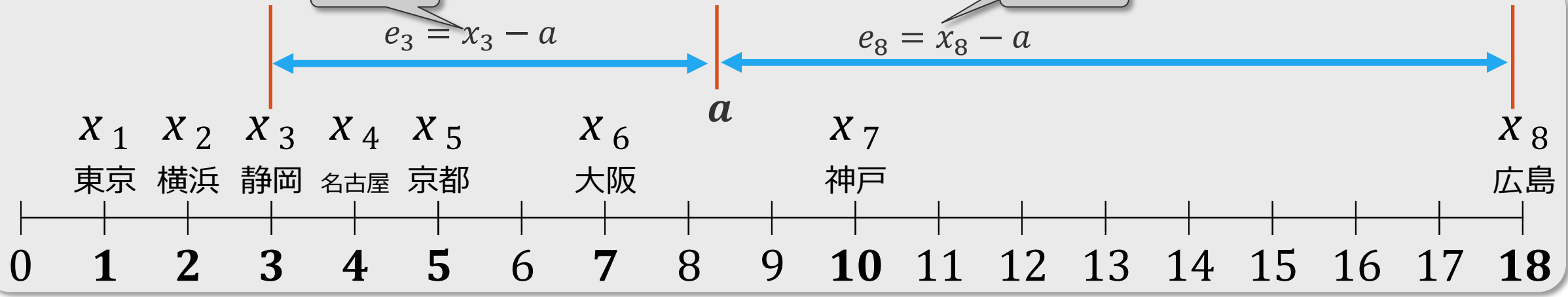
集合場所  $a$  . . . 代表値 (一般の場合に拡張するための表現)

移動距離の絶対値  $|e_i| = |x_i - a|$  単に  $e_i$  を加算しただけでは相殺されてしまう

移動距離の絶対値の合計  $T$   
 $T$  が最小になる  $a$   $T = \sum_{i=1}^n |e_i| = \sum_{i=1}^n |x_i - a| \Rightarrow \min$  (2.1.1) T が最小になる a を決める

マイナス プラス

$e_3 = x_3 - a$   $e_8 = x_8 - a$



## ● 集合場所 $a$ と移動距離の絶対値の合計 $T$

集合場所（代表値）の距離  $a$  を 0~18 まで変化させて  $T$  を計算

$a$  を 0~18 まで  
変化させる

表示 2.1.1 データおよび  $a$  と  $T, S$  との関係（改変）

各都市の位置

| 観測値         | $a$ (代表値) |    |    |     |           |           |    |    |    |   |    |    |    |    |    |    |    |    |    |
|-------------|-----------|----|----|-----|-----------|-----------|----|----|----|---|----|----|----|----|----|----|----|----|----|
|             | 東京        | 横浜 | 静岡 | 名古屋 | 京都        | 大阪        |    | 神戸 |    |   |    | 広島 |    |    |    |    |    |    |    |
|             | 0         | 1  | 2  | 3   | 4         | 5         | 6  | 7  | 8  | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 東京 $x_1$ 1  | 1         | 0  | 1  | 2   | 3         | 4         | 6  |    | 9  |   |    |    | 17 |    |    |    |    |    |    |
| 横浜 $x_2$ 2  | 2         | 1  | 0  | 1   | 2         | 3         | 5  |    | 8  |   |    |    | 16 |    |    |    |    |    |    |
| 静岡 $x_3$ 3  | 3         | 2  | 1  | 0   | 1         | 2         | 4  |    | 7  |   |    |    | 15 |    |    |    |    |    |    |
| 名古屋 $x_4$ 4 | 4         | 3  | 2  | 1   | 0         | 1         | 3  |    | 6  |   |    |    | 14 |    |    |    |    |    |    |
| 京都 $x_5$ 5  | 5         | 4  | 3  | 2   | 1         | 0         | 2  |    | 5  |   |    |    | 13 |    |    |    |    |    |    |
| 大阪 $x_6$ 7  | 7         | 6  | 5  | 4   | 3         | 2         | 0  |    | 3  |   |    |    | 11 |    |    |    |    |    |    |
| 神戸 $x_7$ 10 | 10        | 9  | 8  | 7   | 6         | 5         | 3  |    | 0  |   |    |    | 8  |    |    |    |    |    |    |
| 広島 $x_8$ 18 | 18        | 17 | 16 | 15  | 14        | 13        | 11 |    | 8  |   |    |    | 0  |    |    |    |    |    |    |
| $T$         | 50        | 42 | 36 | 32  | <b>30</b> | <b>30</b> | 34 |    | 46 |   |    |    | 94 |    |    |    |    |    |    |

$|e_i| = |x_i - a|$

$$T = \sum_{i=1}^n |e_i|$$

## ● 集合場所 $a$ と移動距離の絶対値の合計 $T$

集合場所 (代表値)  $a$  を 0~18 まで変化させて  $T$  を計算

表示 2.1.1 データおよび  $a$  と  $T, S$  との関係 (改変)

| 各都市の位置<br>観測値 | $a$ (代表値) |    |    |     |    |    |    |    |    |  |  |  |  |  |  |  |  |  |  |
|---------------|-----------|----|----|-----|----|----|----|----|----|--|--|--|--|--|--|--|--|--|--|
|               | 東京        | 横浜 | 静岡 | 名古屋 | 京都 | 大阪 | 神戸 | 広島 |    |  |  |  |  |  |  |  |  |  |  |
| 東京 $x_1$ 1    | 0         | 1  | 2  | 3   | 4  | 6  | 9  | 17 |    |  |  |  |  |  |  |  |  |  |  |
| 横浜 $x_2$ 2    | 1         | 0  | 1  | 2   | 3  | 5  | 8  | 16 |    |  |  |  |  |  |  |  |  |  |  |
| 静岡 $x_3$ 3    | 2         | 1  | 0  | 1   | 2  | 4  | 7  | 15 |    |  |  |  |  |  |  |  |  |  |  |
| 名古屋 $x_4$ 4   | 3         | 2  | 1  | 0   | 1  | 3  | 6  | 14 |    |  |  |  |  |  |  |  |  |  |  |
| 京都 $x_5$ 5    | 4         | 3  | 2  | 1   | 0  | 2  | 5  | 13 |    |  |  |  |  |  |  |  |  |  |  |
| 大阪 $x_6$ 7    | 6         | 5  | 4  | 3   | 2  | 0  | 3  | 11 |    |  |  |  |  |  |  |  |  |  |  |
| 神戸 $x_7$ 10   | 9         | 8  | 7  | 6   | 5  | 3  | 0  | 8  |    |  |  |  |  |  |  |  |  |  |  |
| 広島 $x_8$ 18   | 17        | 16 | 15 | 14  | 13 | 11 | 8  | 0  |    |  |  |  |  |  |  |  |  |  |  |
| $T$           | 50        | 42 | 36 | 32  | 30 | 30 | 34 | 46 | 94 |  |  |  |  |  |  |  |  |  |  |

仮に  $a=10$  の場合とする

$|e_1| = |x_1 - a|$   
 $= |1 - 10|$   
 $= 9$   
 東京 - 神戸

$|e_5| = |x_5 - a|$   
 $= |5 - 10|$   
 $= 5$   
 京都 - 神戸

$$T = \sum_{i=1}^n |e_i|$$

## ● 集合場所 $a$ と移動距離の絶対値の合計 $T$

集合場所 (代表値)  $a$  を 0~18 まで変化させて  $T$  を計算

表示 2.1.1 データおよび  $a$  と  $T, S$  との関係 (改変)

| 各都市の位置<br>観測値 | $a$ (代表値) |    |    |    |    |    |   |    |   |   |    |    |    |    |    |    |    |    |    |
|---------------|-----------|----|----|----|----|----|---|----|---|---|----|----|----|----|----|----|----|----|----|
|               | 0         | 1  | 2  | 3  | 4  | 5  | 6 | 7  | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 東京 $x_1$ 1    | 1         | 0  | 1  | 2  | 3  | 4  |   | 6  |   |   | 9  |    |    |    |    |    |    |    | 17 |
| 横浜 $x_2$ 2    | 2         | 1  | 0  | 1  | 2  | 3  |   | 5  |   |   | 8  |    |    |    |    |    |    |    | 16 |
| 静岡 $x_3$ 3    | 3         | 2  | 1  | 0  | 1  | 2  |   | 4  |   |   | 7  |    |    |    |    |    |    |    | 15 |
| 名古屋 $x_4$ 4   | 4         | 3  | 2  | 1  | 0  | 1  |   | 3  |   |   | 6  |    |    |    |    |    |    |    | 14 |
| 京都 $x_5$ 5    | 5         | 4  | 3  | 2  | 1  | 0  |   | 2  |   |   | 5  |    |    |    |    |    |    |    | 13 |
| 大阪 $x_6$ 7    | 7         | 6  | 5  | 4  | 3  | 2  |   | 0  |   |   | 3  |    |    |    |    |    |    |    | 11 |
| 神戸 $x_7$ 10   | 10        | 9  | 8  | 7  | 6  | 5  |   |    |   |   | 0  |    |    |    |    |    |    |    | 8  |
| 広島 $x_8$ 18   | 18        | 17 | 16 | 15 | 14 | 13 |   | 11 |   |   | 8  |    |    |    |    |    |    |    | 0  |
| $T$           | 50        | 42 | 36 | 32 | 30 | 30 |   | 34 |   |   | 46 |    |    |    |    |    |    |    | 94 |

$$T = \sum_{i=1}^n |e_i|$$

$a$

$$|e_1| = |x_1 - a| = |1 - 10| = 9$$

東京 - 神戸

$$|e_5| = |x_5 - a| = |5 - 10| = 5$$

京都 - 神戸

$T$  の最小値

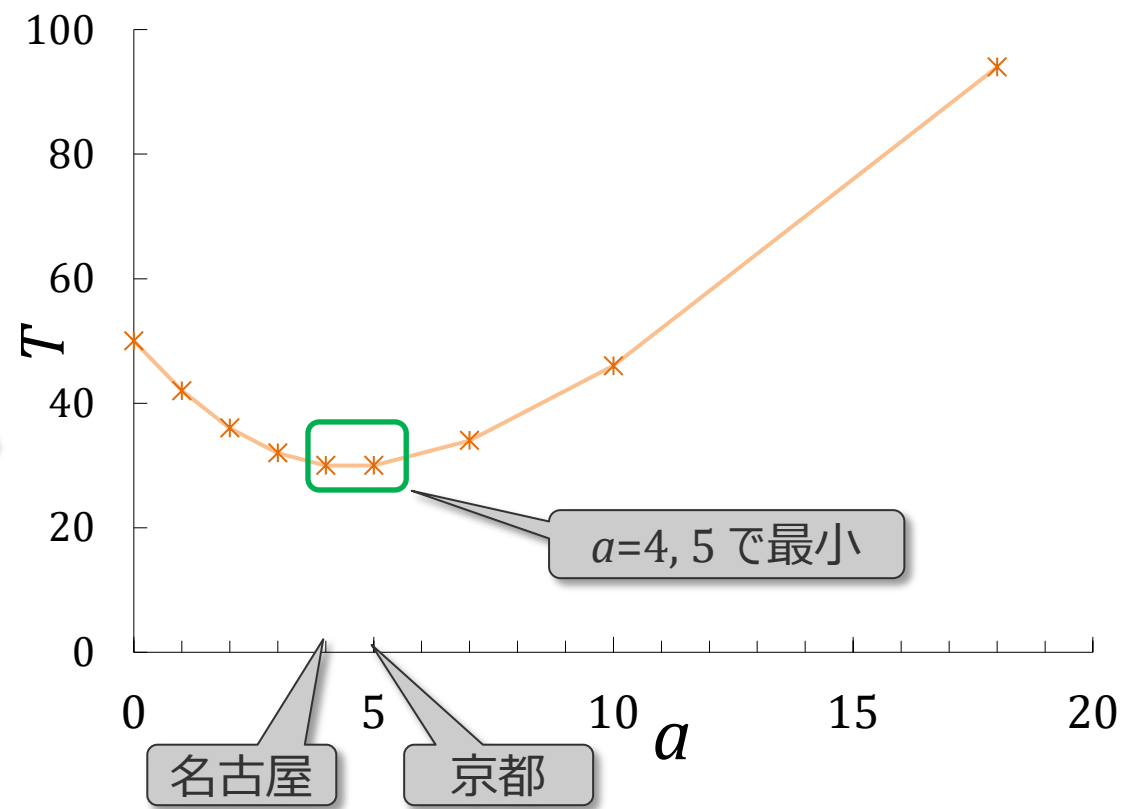
## ● 集合場所 $a$ と移動距離の絶対値の合計 $T$

集合場所 (代表値)  $a$  を 0~18 まで変化させて  $T$  を計算

表示 2.1.1 データおよび  $a$  と  $T, S$  との関係 (改変)

| $a$ | 観測値   |     | 東京 横浜 静岡 名古屋 京都 |    |    |    |           | ...       |   |
|-----|-------|-----|-----------------|----|----|----|-----------|-----------|---|
|     |       |     | 0               | 1  | 2  | 3  | 4         |           | 5 |
| 東京  | $x_1$ | 1   | 1               | 0  | 1  | 2  | 3         | 4         |   |
| 横浜  | $x_2$ | 2   | 2               | 1  | 0  | 1  | 2         | 3         |   |
| 静岡  | $x_3$ | 3   | 3               | 2  | 1  | 0  | 1         | 2         |   |
| 名古屋 | $x_4$ | 4   | 4               | 3  | 2  | 1  | 0         | 1         |   |
| 京都  | $x_5$ | 5   | 5               | 4  | 3  | 2  | 1         | 0         |   |
| 大阪  | $x_6$ | 7   | 7               | 6  | 5  | 4  | 3         | 2         |   |
| 神戸  | $x_7$ | 10  | 10              | 9  | 8  | 7  | 6         | 5         |   |
| 広島  | $x_8$ | 18  | 18              | 17 | 16 | 15 | 14        | 13        |   |
|     |       | $T$ | 50              | 42 | 36 | 32 | <b>30</b> | <b>30</b> |   |

$$T = \sum_{i=1}^n |e_i|$$

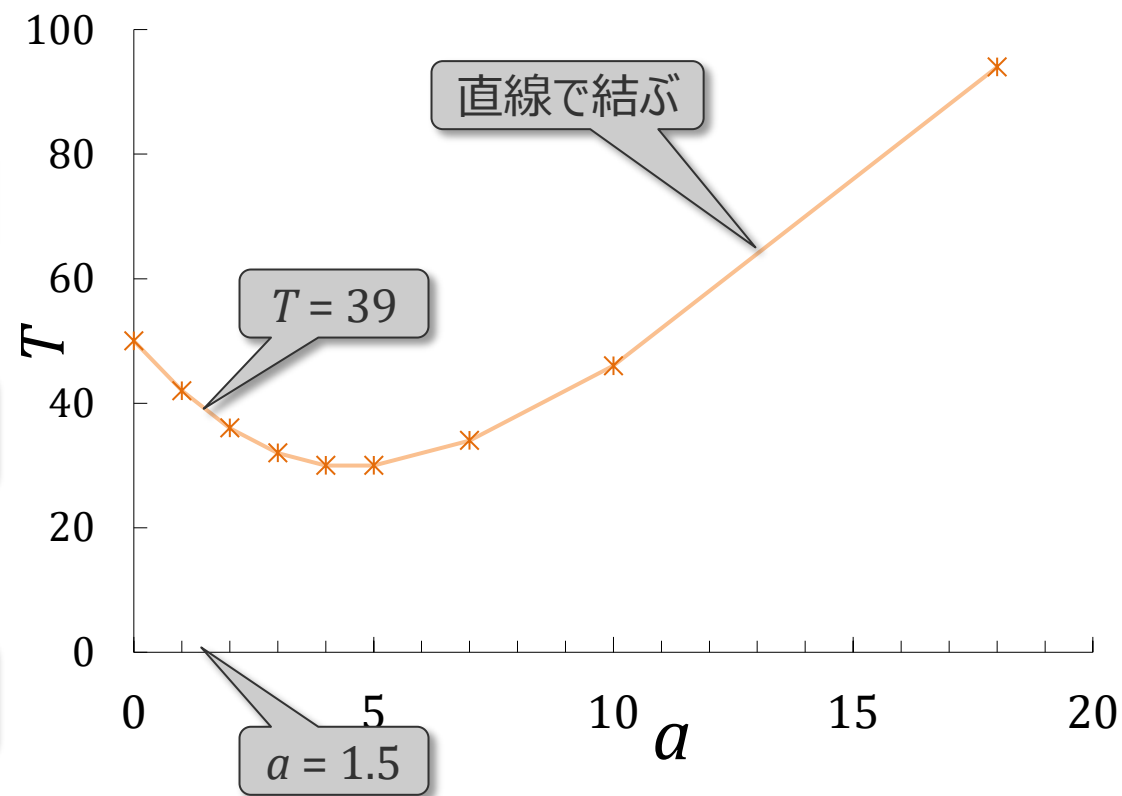


## ● 集合場所 $a$ と移動距離の絶対値の合計 $T$

集合場所 (代表値)  $a$  を 0~18 まで変化させて  $T$  を計算

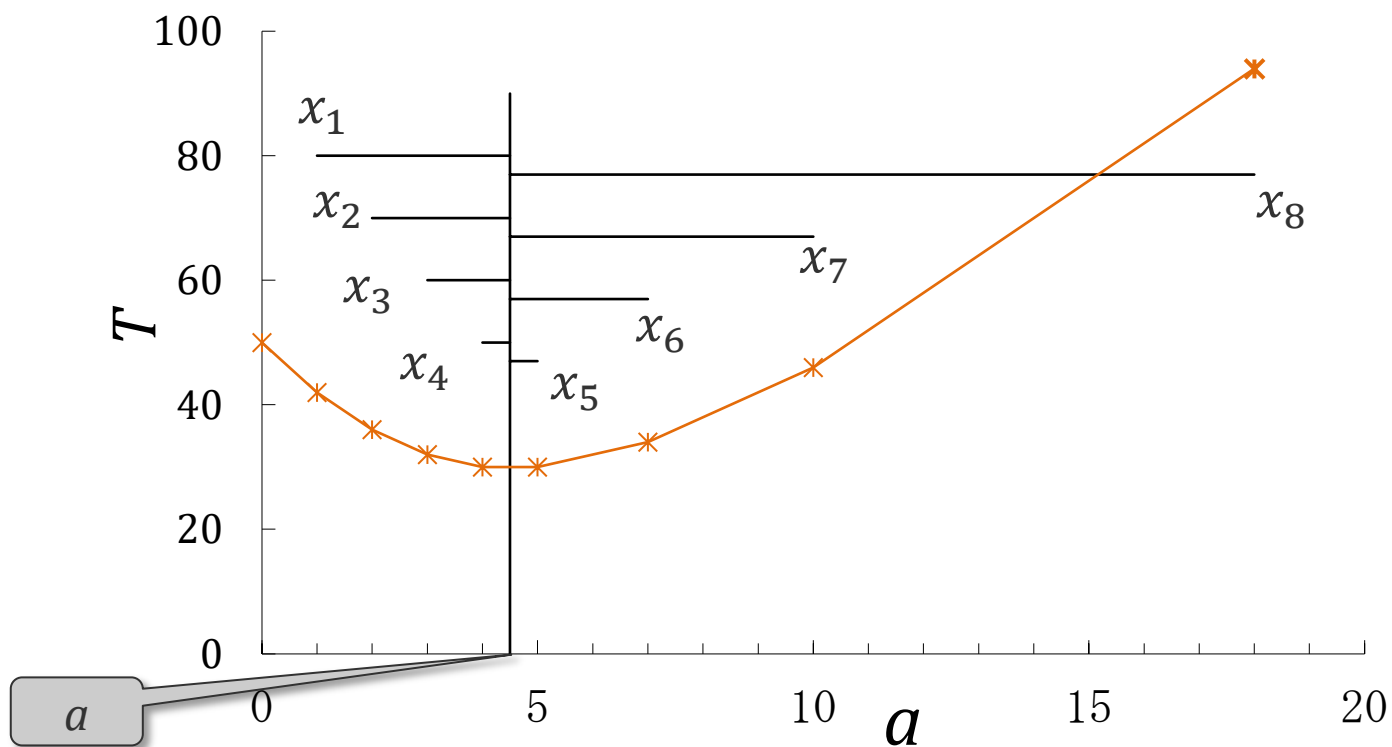
表示 2.1.1 データおよび  $a$  と  $T, S$  との関係 (改変)

| 観測値 |       |    | $a$ (代表値) |    |      |    |                               |
|-----|-------|----|-----------|----|------|----|-------------------------------|
|     |       |    | 0         | 1  | 1.5  | 2  |                               |
| 東京  | $x_1$ | 1  | 1         | 0  | 0.5  | 1  | $a = \frac{1+2}{2} = 1.5$     |
| 横浜  | $x_2$ | 2  | 2         | 1  | 0.5  | 0  |                               |
| 静岡  | $x_3$ | 3  | 3         | 2  | 1.5  | 1  | $ e_4  = \frac{3+2}{2} = 2.5$ |
| 名古屋 | $x_4$ | 4  | 4         | 3  | 2.5  | 2  |                               |
| 京都  | $x_5$ | 5  | 5         | 4  | 3.5  | 3  | $T = \frac{42+36}{2} = 39$    |
| 大阪  | $x_6$ | 7  | 7         | 6  | 5.5  | 5  |                               |
| 神戸  | $x_7$ | 10 | 10        | 9  | 8.5  | 8  |                               |
| 広島  | $x_8$ | 18 | 18        | 17 | 16.5 | 16 |                               |
| $T$ |       |    | 50        | 42 | 39   | 36 |                               |



## ● 集合場所 $a$ と移動距離の絶対値の合計 $T$

観測値を昇順に並べ、小さい方から  $i$  番目の値を  $x_i$  とする (この事例では元々の  $x_i$  と同一)



$$\begin{aligned}
 T &= \sum_{i=1}^n |e_i| \\
 &= \sum_{i=1}^n |x_i - a| \\
 &= |x_1 - a| + |x_8 - a| \\
 &\quad + |x_2 - a| + |x_7 - a| \\
 &\quad + |x_3 - a| + |x_6 - a| \\
 &\quad + |x_4 - a| + |x_5 - a|
 \end{aligned}$$

$T$  を最小にする  $a$  は?

昇順に並んでいることを意識

観測値  $x_i$  と  $a$  までの距離

最小値、最大値の組合せ

最小値の1つ後、最大値の1つ前

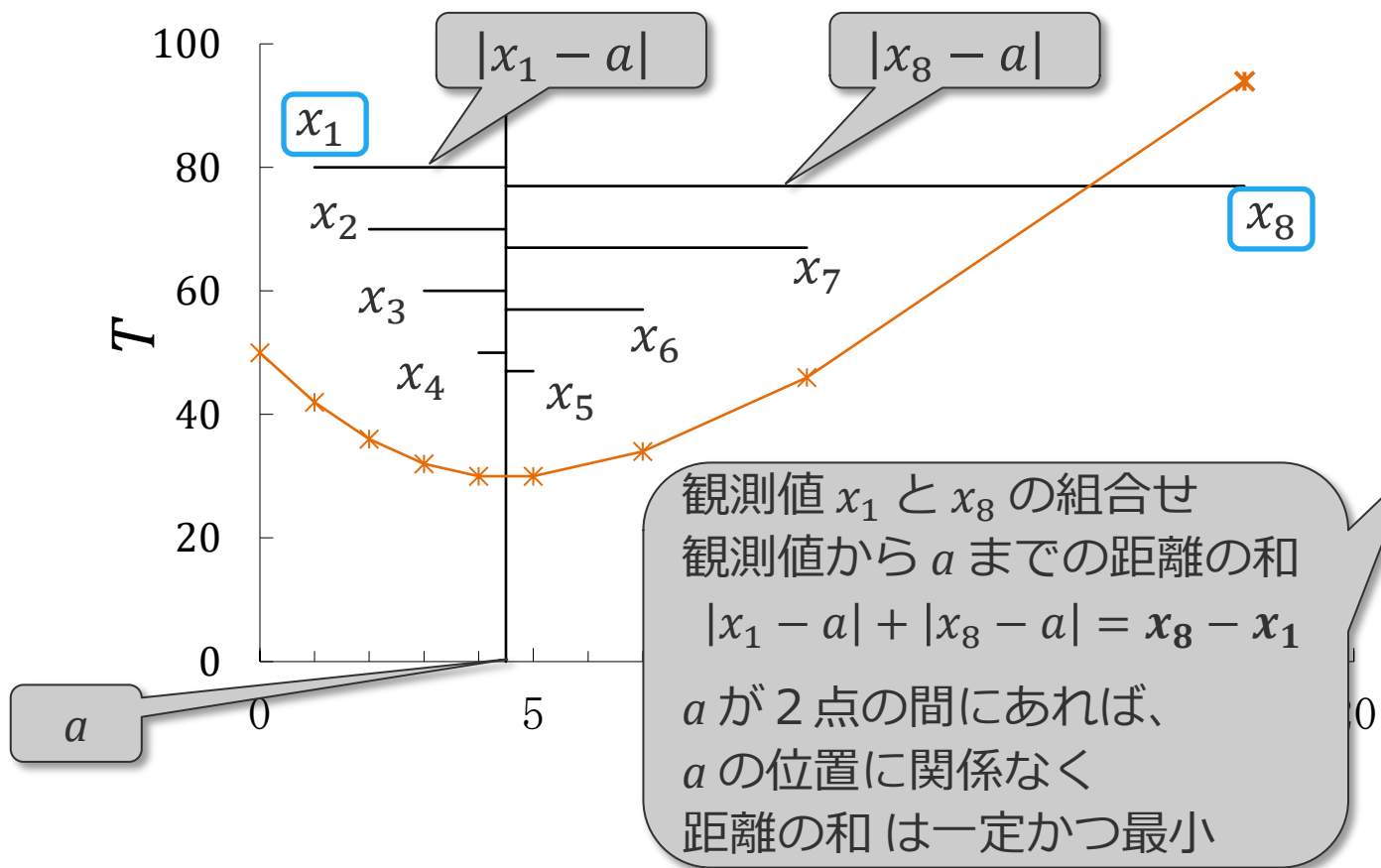
最小値の2つ後、最大値の2つ前

最小値の3つ後、最大値の3つ前

2地点ごとに組合せ

## ● 集合場所 $a$ と移動距離の絶対値の合計 $T$

観測値を昇順に並べ、小さい方から  $i$  番目の値を  $x_i$  とする（この事例では元々の  $x_i$  と同一）



$$T = \sum_{i=1}^n |e_i|$$

$$= \sum_{i=1}^n |x_i - a|$$

$$= |x_1 - a| + |x_8 - a|$$

$$+ |x_2 - a| + |x_7 - a|$$

$$+ |x_3 - a| + |x_6 - a|$$

$$+ |x_4 - a| + |x_5 - a|$$

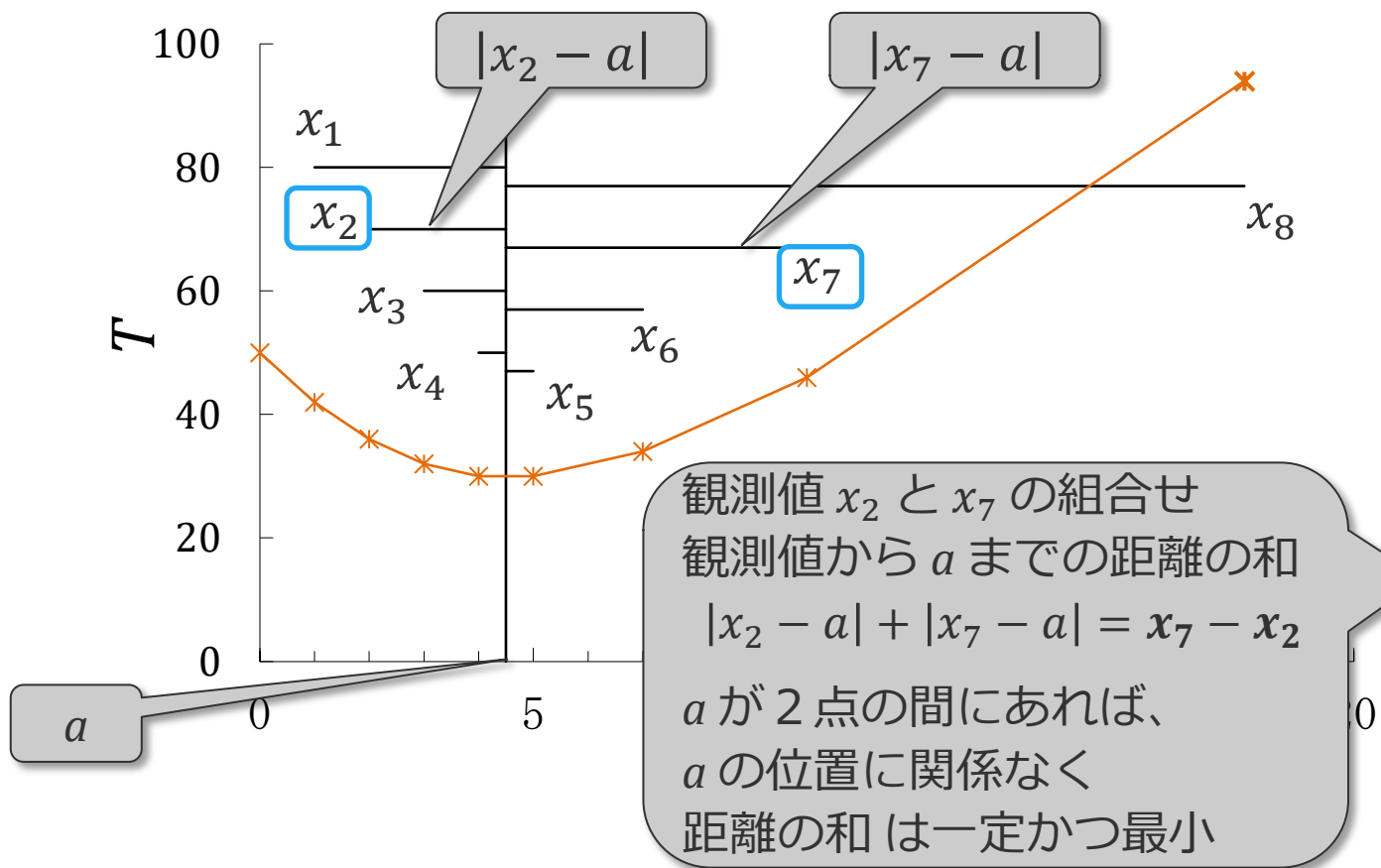
$T$  を最小にする  $a$  は？

観測値  $x_i$  と  $a$  までの距離

最小値の距離 + 最大値の距離

## ● 集合場所 $a$ と移動距離の絶対値の合計 $T$

観測値を昇順に並べ、小さい方から  $i$  番目の値を  $x_i$  とする (この事例では元々の  $x_i$  と同一)



$$T = \sum_{i=1}^n |e_i|$$

$$= \sum_{i=1}^n |x_i - a|$$

$$= |x_1 - a| + |x_8 - a|$$

$$+ |x_2 - a| + |x_7 - a|$$

$$+ |x_3 - a| + |x_6 - a|$$

$$+ |x_4 - a| + |x_5 - a|$$

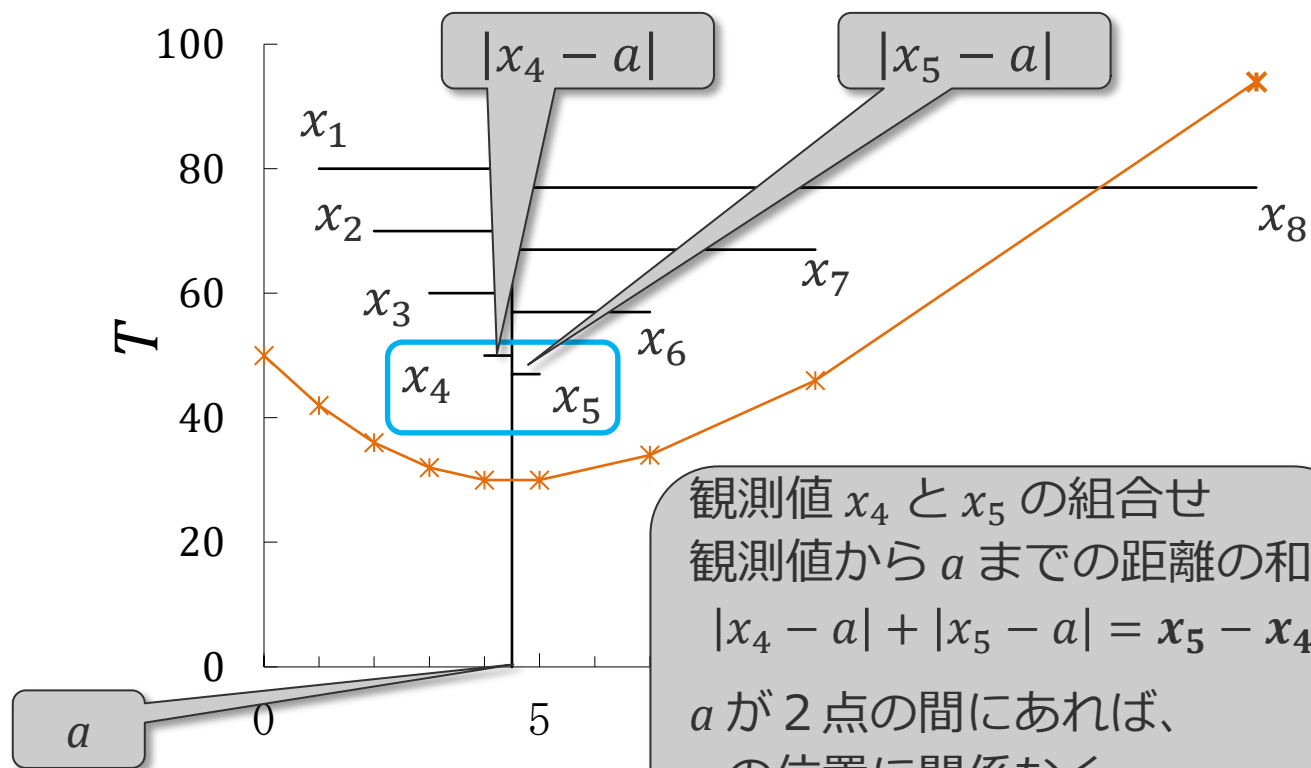
$T$  を最小にする  $a$  は?

観測値  $x_i$  と  $a$  までの距離

最小値の1つ後  
 + 最大値の1つ前

## ● 集合場所 $a$ と移動距離の絶対値の合計 $T$

観測値を昇順に並べ、小さい方から  $i$  番目の値を  $x_i$  とする (この事例では元々の  $x_i$  と同一)



観測値  $x_4$  と  $x_5$  の組合せ  
 観測値から  $a$  までの距離の和  
 $|x_4 - a| + |x_5 - a| = x_5 - x_4$   
 $a$  が 2 点の間であれば、  
 $a$  の位置に関係なく  
 距離の和は一定かつ最小

$$T = \sum_{i=1}^n |e_i|$$

$$= \sum_{i=1}^n |x_i - a|$$

$$= |x_1 - a| + |x_8 - a|$$

$$+ |x_2 - a| + |x_7 - a|$$

$$+ |x_3 - a| + |x_6 - a|$$

$$+ |x_4 - a| + |x_5 - a|$$

観測値  $x_i$  と  $a$  までの距離

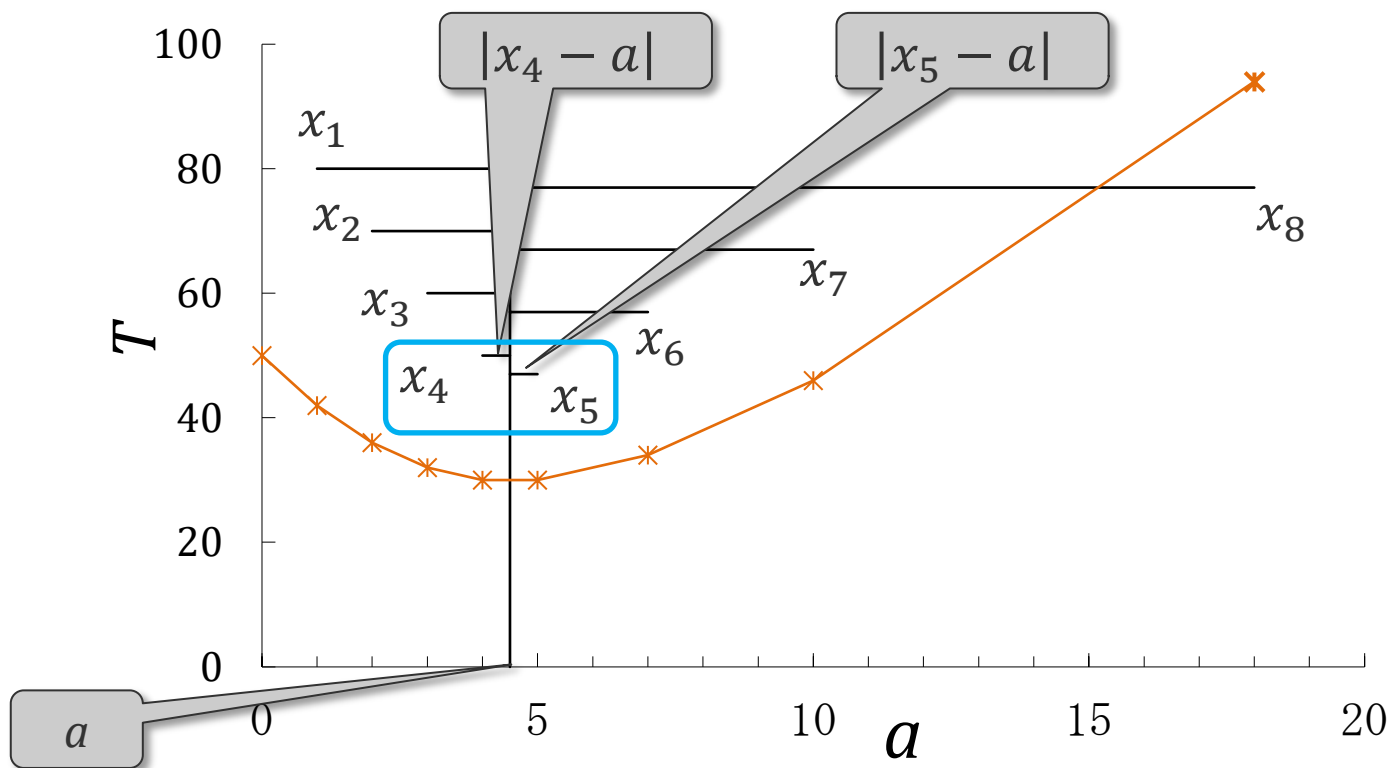
最小値の2つ後  
 + 最大値の2つ前

最小値の3つ後  
 + 最大値の3つ前

$T$  を最小にする  $a$  は?

## ● 集合場所 $a$ と移動距離の絶対値の合計 $T$

観測値を昇順に並べ、小さい方から  $i$  番目の値を  $x_i$  とする（この事例では元々の  $x_i$  と同一）

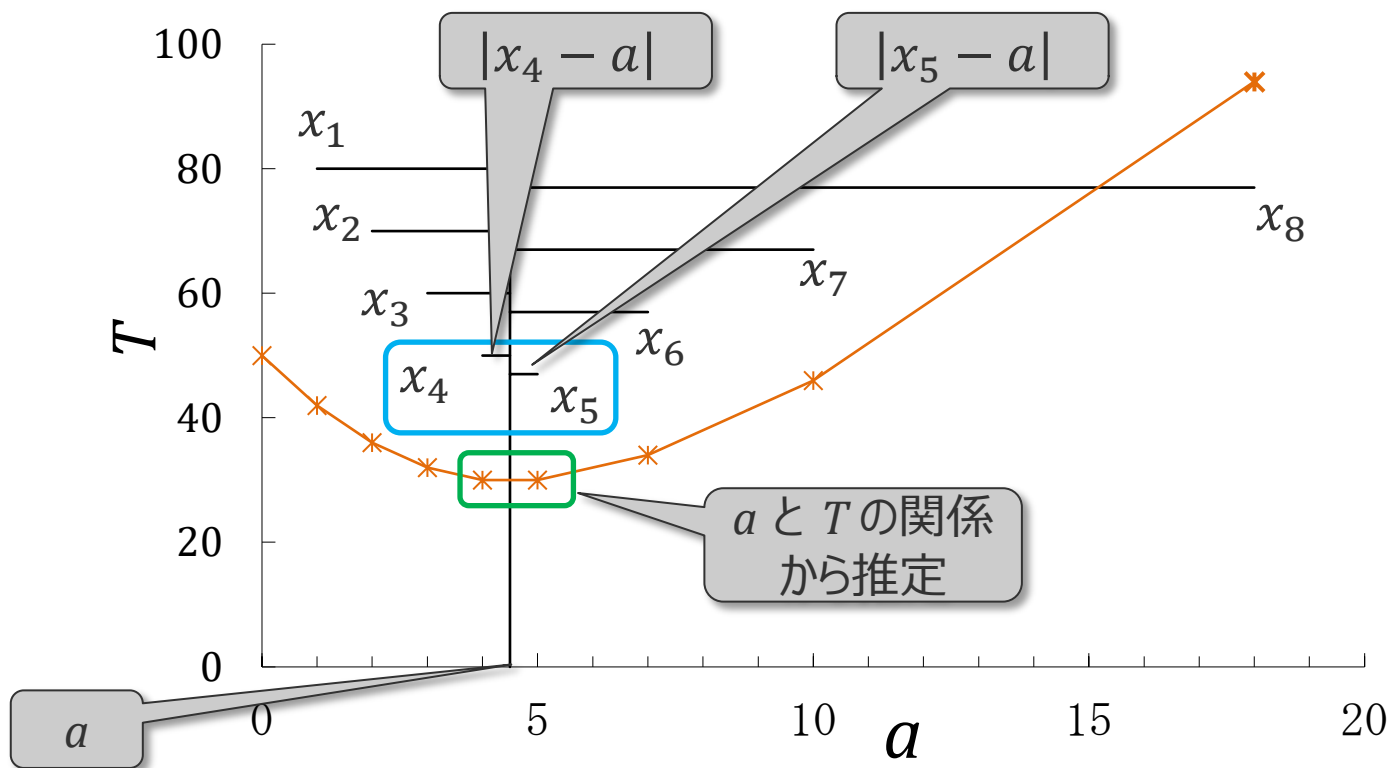


$$\begin{aligned}
 T &= \sum_{i=1}^n |e_i| \\
 &= \sum_{i=1}^n |x_i - a| \\
 &= |x_1 - a| + |x_8 - a| \\
 &\quad + |x_2 - a| + |x_7 - a| \\
 &\quad + |x_3 - a| + |x_6 - a| \\
 &\quad + |x_4 - a| + |x_5 - a|
 \end{aligned}$$

$a$  が 2 点間の間にあれば、  
距離の和は  $a$  によらず  
一定かつ最小  
最終的に  $a$  は  $x_4$  と  $x_5$  の  
間に落ち着く

## ● 集合場所 $a$ と移動距離の絶対値の合計 $T$

観測値を昇順に並べ、小さい方から  $i$  番目の値を  $x_i$  とする（この事例では元々の  $x_i$  と同一）

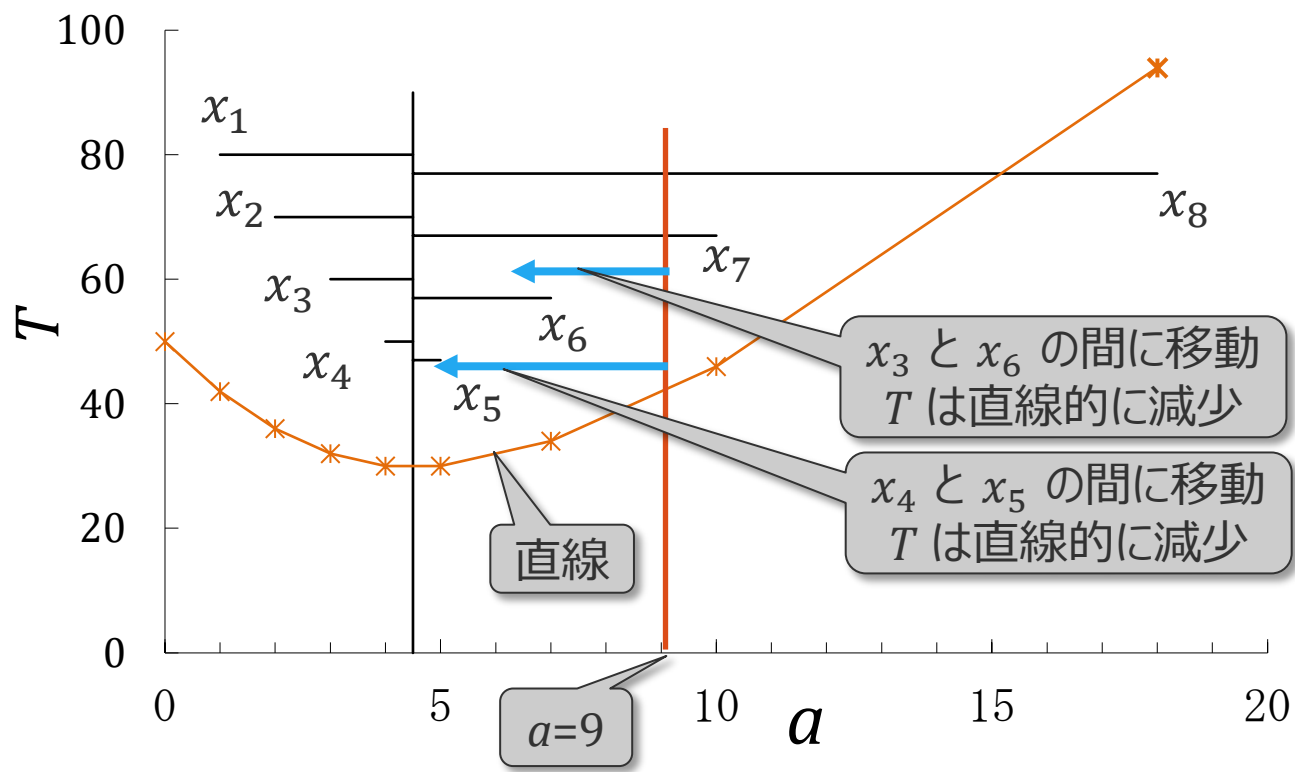


$$\begin{aligned}
 T &= \sum_{i=1}^n |e_i| \\
 &= \sum_{i=1}^n |x_i - a| \\
 &= |x_1 - a| + |x_8 - a| \\
 &\quad + |x_2 - a| + |x_7 - a| \\
 &\quad + |x_3 - a| + |x_6 - a| \\
 &\quad + |x_4 - a| + |x_5 - a|
 \end{aligned}$$

$a$  が 2 点間の間にあれば、  
距離の和は  $a$  によらず  
一定かつ最小  
最終的に  $a$  は  $x_4$  と  $x_5$  の  
間に落ち着く

## ● 集合場所 $a$ と移動距離の絶対値の合計 $T$

観測値を昇順に並べ、小さい方から  $i$  番目の値を  $x_i$  とする（この事例では元々の  $x_i$  と同一）



$$\begin{aligned}
 T &= \sum_{i=1}^n |e_i| \\
 &= \sum_{i=1}^n |x_i - a| \\
 &= |x_1 - a| + |x_8 - a| \\
 &\quad + |x_2 - a| + |x_7 - a| \\
 &\quad + |x_3 - a| + |x_6 - a| \\
 &\quad + |x_4 - a| + |x_5 - a|
 \end{aligned}$$

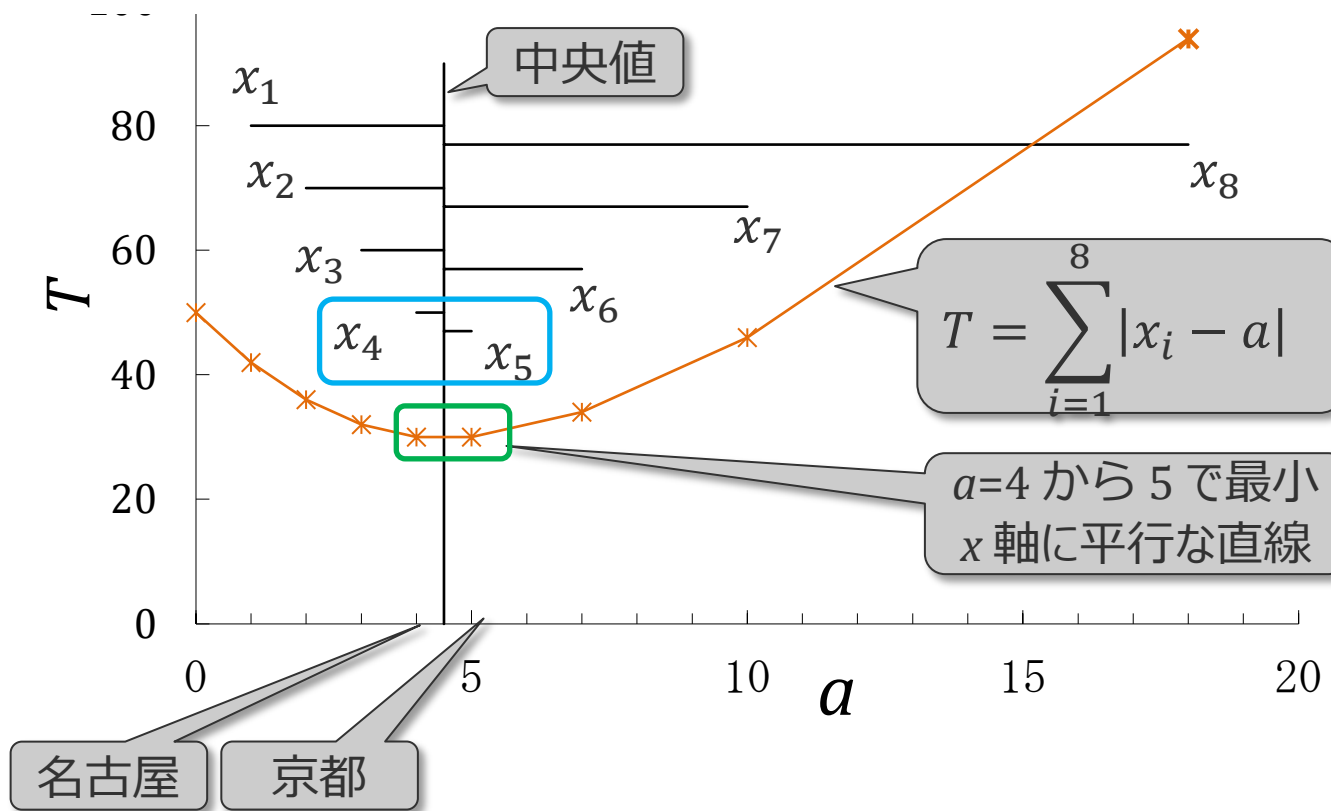
$a$  が 2 点間の間にあれば、  
距離の和は  $a$  によらず  
一定かつ最小  
最終的に  $a$  は  $x_4$  と  $x_5$  の  
間に落ち着く

（2点間の外側に  $a$  があると、観測値から  $a$  までの距離は大きくなる）

## ●中央値の意味

都市（観測値）の数が偶数：中央の2都市の midpoint（名古屋と京都の midpoint、すなわち平均）

都市（観測値）の数が奇数：中央の1都市（ $i=8$ の広島がなかった場合、 $i=4$ の名古屋）

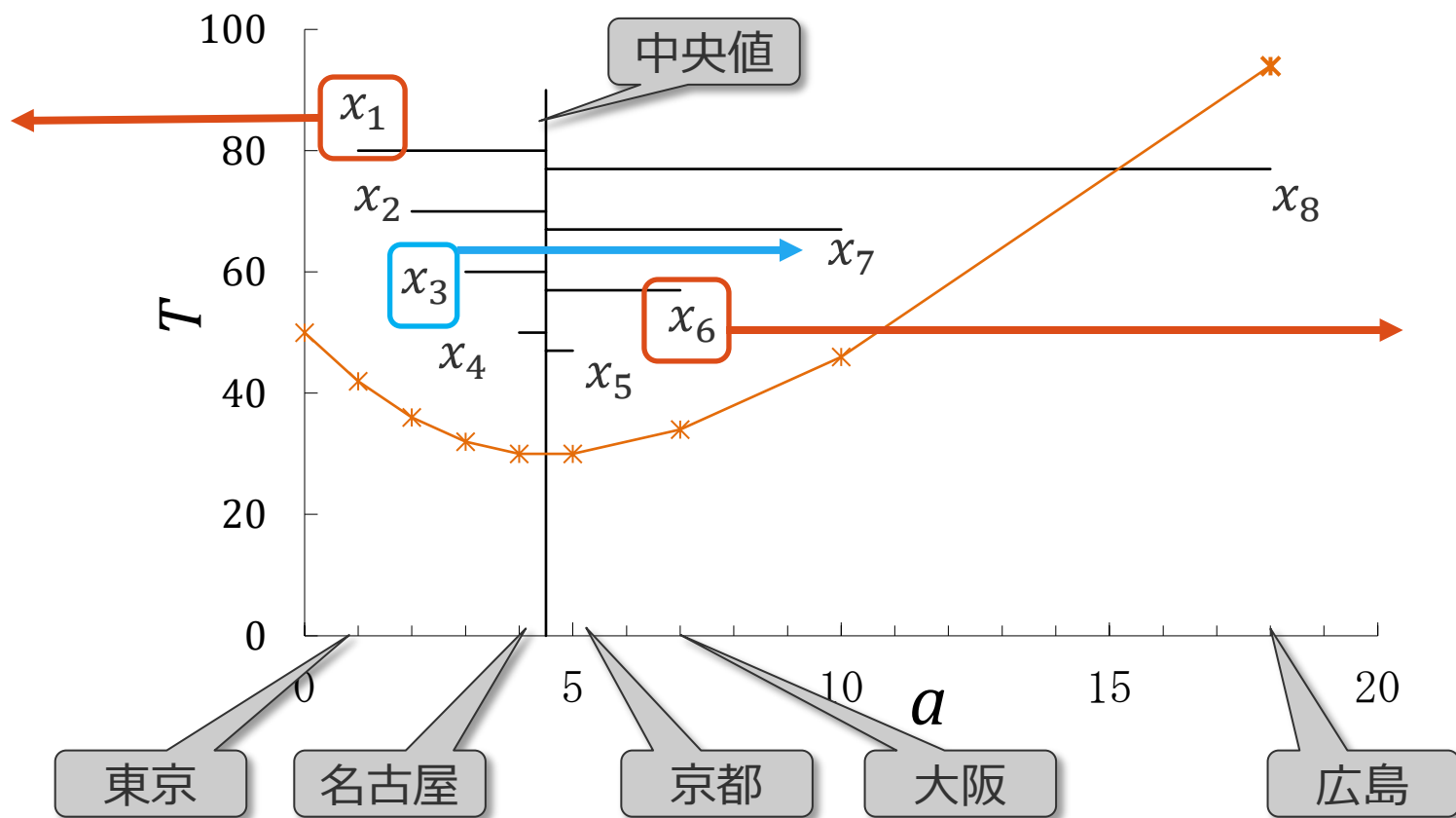


$$\begin{aligned} T &= \sum_{i=1}^n |e_i| \\ &= \sum_{i=1}^n |x_i - a| \\ &= |x_1 - a| + |x_8 - a| \\ &\quad + |x_2 - a| + |x_7 - a| \\ &\quad + |x_3 - a| + |x_6 - a| \\ &\quad + |x_4 - a| + |x_5 - a| \end{aligned}$$

$a$  が 2 点間の間にあれば、  
距離の和は  $a$  によらず  
一定かつ最小  
最終的に  $a$  は  $x_4$  と  $x_5$  の  
間に落ち着く

中央値  
2 地点の midpoint

## ●中央値の頑健性（後述）



中央値には頑健性がある (→(8))

東京の人が札幌に転居しても、  
中央値は変わらない

大阪の人が博多に転居しても、  
中央値は変わらない

名古屋より東の人が京都より西に  
転居すると、中央値は京都－大阪  
の間に移動

## ●中央値の算出と性質

中央値 (Median)  $\tilde{x}$  (エックス・チルダ(tilde)、エックス波型)

観測値  $x_i$  ( $x_1, x_2, x_3, \dots, x_{n-1}, x_n$ , 昇順に並び替える)

$n$  が奇数の場合 中央の値

$$\tilde{x} = x_{(i)} \quad i = \frac{n + 1}{2} \quad (2.1.2)$$

$n$  が偶数の場合 (中央の2つの値の平均)

$$\tilde{x} = \frac{x_{(i)} + x_{(i+1)}}{2} \quad i = \frac{n}{2}$$

$n = 7$   
 $i = (7 + 1)/2 = 4$

$n = 8$   
 $i = 8/2 = 4$   
 $i + 1 = 4 + 1 = 5$

頑健性がある

外れ値の影響を受けにくい

Excel の関数

=MEDIAN(データの集まり)

=MEDIAN(B4:B11)

=MEDIAN(B4:B8, C3, C7:C10)

2つのセルを「:」で結ぶ→データの範囲

「,」で結ぶ→データの集まり



## (2) 平均值

代表值：中央値、平均值

## ●事例 1 : 8都市から1人ずつ1か所に集まる最短の集合場所

起点から各都市の距離  $x_i$  . . . 観測値 (一般の場合に拡張するための表現)

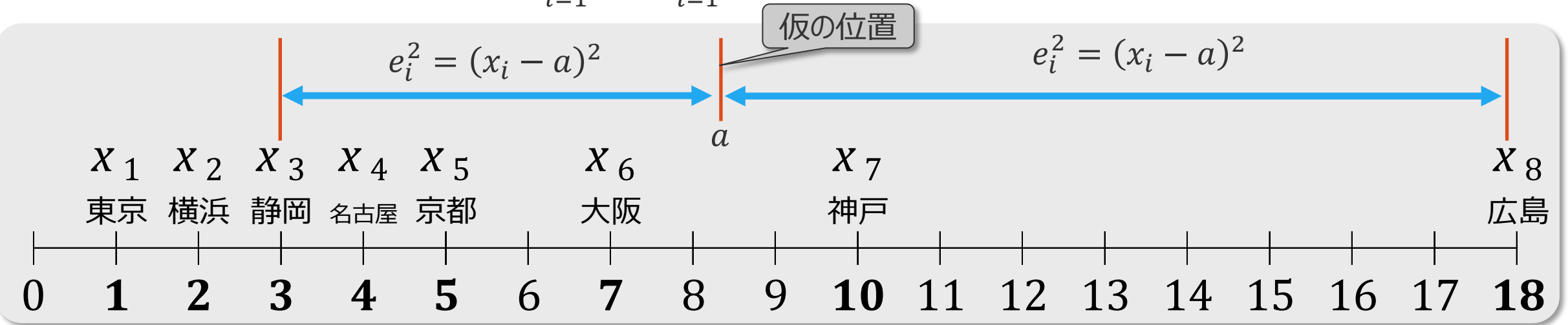
集合場所  $a$  . . . 代表値 (一般の場合に拡張するための表現)

移動距離の2乗  $e_i^2 = (x_i - a)^2$

移動距離の2乗和

$$S = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (x_i - a)^2 \Rightarrow \min \quad (2.1.3)$$

$S$ が最小になる  $a$



## ●事例 1 : 8都市から1人ずつ1か所に集まる最短の集合場所

起点から各都市の距離  $x_i$  . . . 観測値

集合場所  $a$  . . . 代表値

移動距離の2乗  $e_i^2 = (x_i - a)^2$

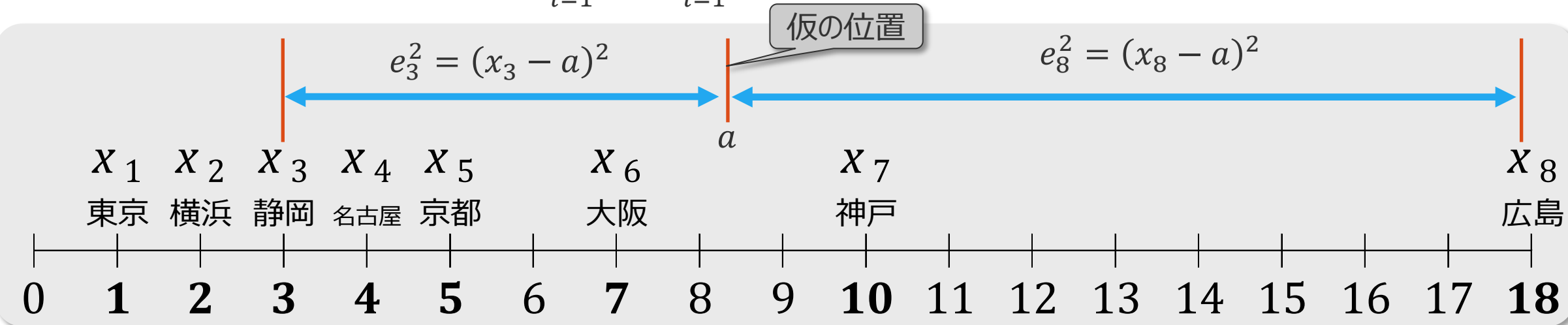
移動距離の2乗和

$S$

$S$ が最小になる  $a$

$$S = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (x_i - a)^2 \Rightarrow \min \quad (2.1.3)$$

$$T = \sum_{i=1}^n |e_i| = \sum_{i=1}^n |x_i - a| \Rightarrow \min \quad (2.1.1)$$



## ● 集合場所 $a$ と移動距離の 2 乗値の合計 $S$

集合場所 (代表値)  $a$  を 0~18 まで変化させて  $S$  を計算

$a$  を 0~18 まで  
変化させる

表示 2.1.1 データおよび  $a$  と  $T, S$  との関係

各都市の位置

| 観測値      | $a$ (平均値) |     |     |     |           |           |   |     |   |   |     |    |    |    |    |    |    |    |      |
|----------|-----------|-----|-----|-----|-----------|-----------|---|-----|---|---|-----|----|----|----|----|----|----|----|------|
|          | 0         | 1   | 2   | 3   | 4         | 5         | 6 | 7   | 8 | 9 | 10  | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18   |
| $x_1$ 1  | 1         | 0   | 1   | 2   | 3         | 4         |   | 6   |   |   | 9   |    |    |    |    |    |    |    | 17   |
| $x_2$ 2  | 2         | 1   | 0   | 1   | 2         | 3         |   | 5   |   |   | 8   |    |    |    |    |    |    |    | 16   |
| $x_3$ 3  | 3         | 2   | 1   | 0   | 1         | 2         |   | 4   |   |   | 7   |    |    |    |    |    |    |    | 15   |
| $x_4$ 4  | 4         | 3   | 2   | 1   | 0         | 1         |   | 3   |   |   | 6   |    |    |    |    |    |    |    | 14   |
| $x_5$ 5  | 5         | 4   | 3   | 2   | 1         | 0         |   | 2   |   |   | 5   |    |    |    |    |    |    |    | 13   |
| $x_6$ 7  | 7         | 6   | 5   | 4   | 3         | 2         |   | 0   |   |   | 3   |    |    |    |    |    |    |    | 11   |
| $x_7$ 10 | 10        | 9   | 8   | 7   | 6         | 5         |   | 3   |   |   | 0   |    |    |    |    |    |    |    | 8    |
| $x_8$ 18 | 18        | 17  | 16  | 15  | 14        | 13        |   | 11  |   |   | 8   |    |    |    |    |    |    |    | 0    |
| $T$      | 50        | 42  | 36  | 32  | <b>30</b> | <b>30</b> |   | 34  |   |   | 46  |    |    |    |    |    |    |    | 94   |
| $S$      | 528       | 436 | 360 | 300 | 256       | 228       |   | 220 |   |   | 328 |    |    |    |    |    |    |    | 1320 |

$|e_i| = |x_i - a|$

$$S = \sum_{i=1}^n (x_i - a)^2$$

## ● 集合場所 $a$ と移動距離の 2 乗値の合計 $S$

集合場所 (代表値)  $a$  を 0~18 まで変化させて  $S$  を計算

表示 2.1.1 データおよび  $a$  と  $T, S$  との関係

| 各都市の位置 | 観測値      | $a$ (平均値) |     |     |     |           |           |   |     |   |            |    |    |    |    |    |    |    |    |      |
|--------|----------|-----------|-----|-----|-----|-----------|-----------|---|-----|---|------------|----|----|----|----|----|----|----|----|------|
|        |          | 0         | 1   | 2   | 3   | 4         | 5         | 6 | 7   | 8 | 9          | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18   |
| 東京     | $x_1$ 1  | 1         | 0   | 1   | 2   | 3         | 4         |   | 6   |   | 9          |    |    |    |    |    |    |    |    | 17   |
|        | $x_2$ 2  | 2         | 1   | 0   | 1   | 2         | 3         |   | 5   |   | 8          |    |    |    |    |    |    |    |    | 16   |
|        | $x_3$ 3  | 3         | 2   | 1   | 0   | 1         | 2         |   | 4   |   | 7          |    |    |    |    |    |    |    |    | 15   |
|        | $x_4$ 4  | 4         | 3   | 2   | 1   | 0         | 1         |   | 3   |   | 6          |    |    |    |    |    |    |    |    | 14   |
|        | $x_5$ 5  | 5         | 4   | 3   | 2   | 1         | 0         |   | 2   |   | 5          |    |    |    |    |    |    |    |    | 13   |
|        | $x_6$ 7  | 7         | 6   | 5   | 4   | 3         | 2         |   | 0   |   | 3          |    |    |    |    |    |    |    |    | 11   |
|        | $x_7$ 10 | 10        | 9   | 8   | 7   | 6         | 5         |   | 3   |   | 0          |    |    |    |    |    |    |    |    | 8    |
|        | $x_8$ 18 | 18        | 17  | 16  | 15  | 14        | 13        |   | 11  |   | 8          |    |    |    |    |    |    |    |    | 0    |
|        | $T$      | 50        | 42  | 36  | 32  | <b>30</b> | <b>30</b> |   | 34  |   | 46         |    |    |    |    |    |    |    |    | 94   |
|        | $S$      | 528       | 436 | 360 | 300 | 256       | 228       |   | 220 |   | <b>328</b> |    |    |    |    |    |    |    |    | 1320 |

$a$  を 0~18 まで変化させる

$$|e_1| = |x_1 - a| = |1 - 10| = 9$$

$$S = 9^2 + 8^2 + 7^2 \dots \dots + 0^2 + 8^2$$

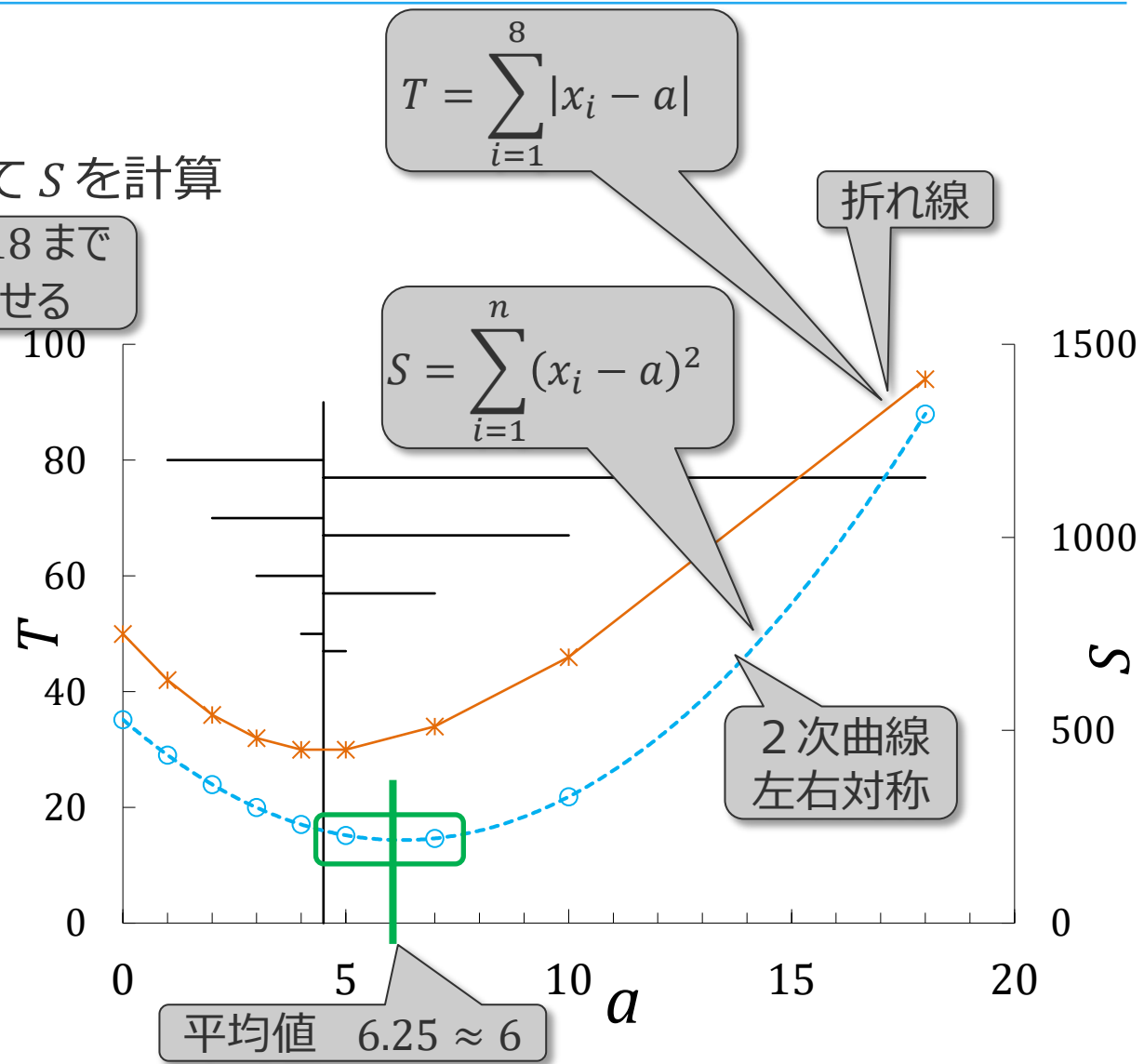
$$S = \sum_{i=1}^n (x_i - a)^2$$

## ● 集合場所 $a$ と移動距離の2乗値の合計 $S$

集合場所 (代表値)  $a$  を 0~18 まで変化させて  $S$  を計算

表示 2.1.1 データおよび  $a$  と  $T, S$  との関係

| 観測値   |    | $a$ を 0~18 まで変化させる |     |     |     |           |            |            |    |
|-------|----|--------------------|-----|-----|-----|-----------|------------|------------|----|
|       |    | 0                  | 1   | 2   | 3   | 4         | 5          | 6          | 7  |
| $x_1$ | 1  | 1                  | 0   | 1   | 2   | 3         | 4          |            | 6  |
| $x_2$ | 2  | 2                  | 1   | 0   | 1   | 2         | 3          |            | 5  |
| $x_3$ | 3  | 3                  | 2   | 1   | 0   | 1         | 2          |            | 4  |
| $x_4$ | 4  | 4                  | 3   | 2   | 1   | 0         | 1          |            | 3  |
| $x_5$ | 5  | 5                  | 4   | 3   | 2   | 1         | 0          |            | 2  |
| $x_6$ | 7  | 7                  | 6   | 5   | 4   | 3         | 2          |            | 0  |
| $x_7$ | 10 | 10                 | 9   | 8   | 7   | 6         | 5          |            | 3  |
| $x_8$ | 18 | 18                 | 17  | 16  | 15  | 14        | 13         |            | 11 |
| $T$   |    | 50                 | 42  | 36  | 32  | <b>30</b> | <b>30</b>  |            | 34 |
| $S$   |    | 528                | 436 | 360 | 300 | 256       | <b>228</b> | <b>220</b> |    |



# 平均値

影響が大

## ● 集合場所 $a$ と移動距離の 2 乗値の合計 $S$

各都市と集合場所までの距離の 2 乗値

$a = 6$  (平均値) とした場合、

$$S = \sum_{i=1}^8 (x_i - a)^2 = \sum_{i=1}^8 (x_i - 6)^2$$

$(x_i - 6)^2$  は正方形の面積に相当

$S$  は面積の総和

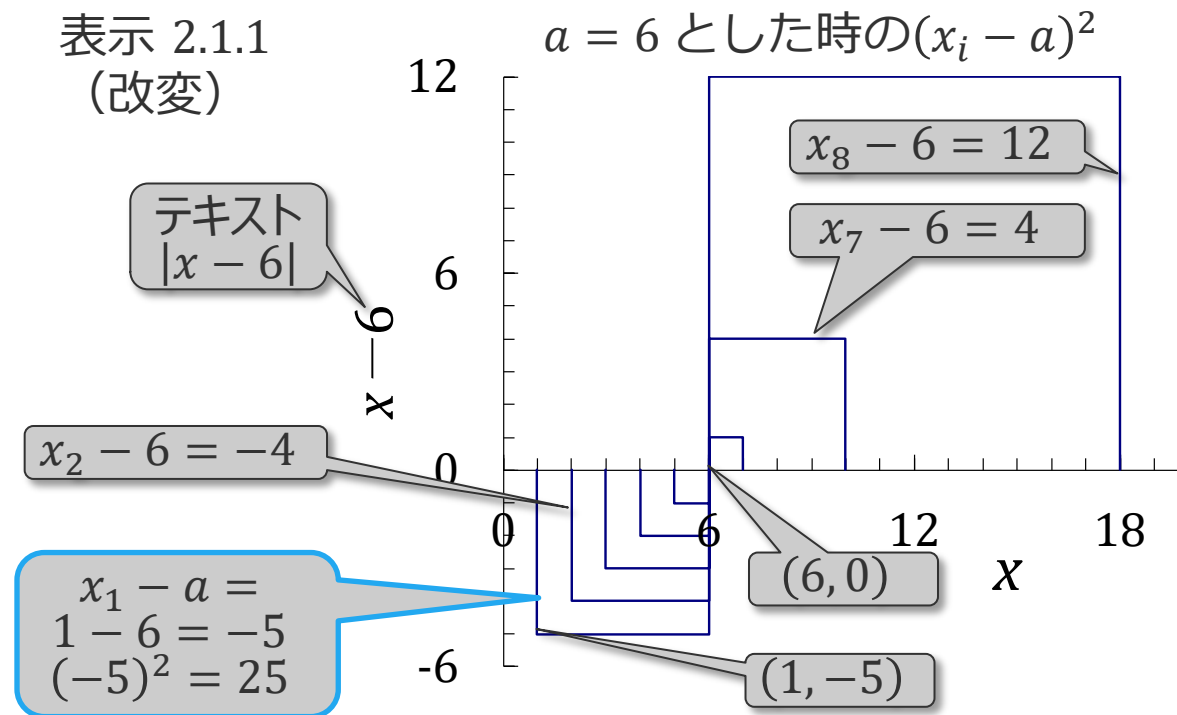
$a=6$  から大きく離れた  $x_8 = 18$  の

面積が大きく、その影響が大きい

$\alpha = 6$  とした場合の各都市  $x_i$  と集合場所  $a$  との距離

|               |    |    |    |    |    |   |    |     |
|---------------|----|----|----|----|----|---|----|-----|
| $i$           | 1  | 2  | 3  | 4  | 5  | 6 | 7  | 8   |
| $x_i$         | 1  | 2  | 3  | 4  | 5  | 7 | 10 | 18  |
| $x_i - 6$     | -5 | -4 | -3 | -2 | -1 | 1 | 4  | 12  |
| $(x_i - 6)^2$ | 25 | 16 | 9  | 4  | 1  | 1 | 16 | 144 |

表示 2.1.1  
(改変)



## ●最小 2 乗法による平均値の導出

$S$  を最小にする  $a$  を解析的に求める

$$S = \sum_{i=1}^n (x_i - a)^2 \quad (2.1.3)$$

$$\frac{dS}{da} = 0$$

$S$  を  $a$  で微分して 0 と置く  
(接線の傾きが 0)

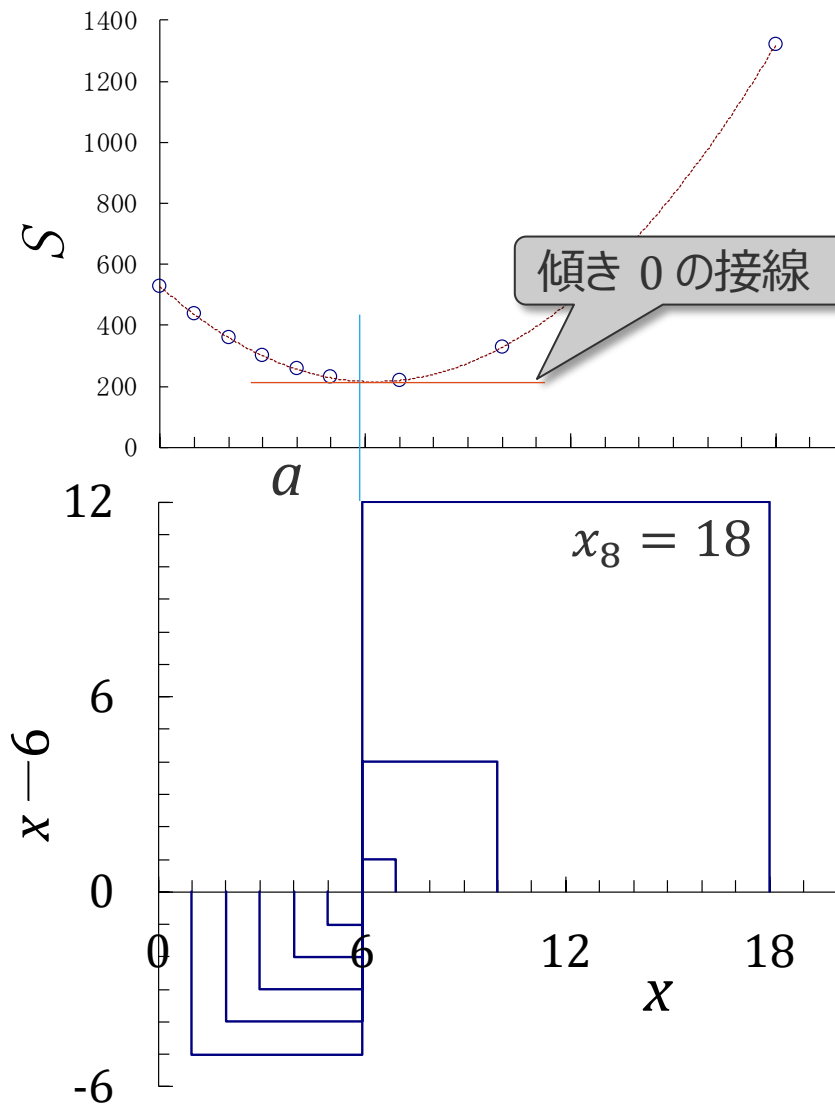
$$\sum_{i=1}^n e_i = \sum_{i=1}^n (x_i - a) = 0 \quad (2.1.4)$$

$$a = \frac{\sum x_i}{n} = \bar{x}. \quad (2.1.5)$$

平均値

$e_i$  は残差、この合計は 0 になる

微分の詳細は p.56, 微分を用いない解法は p.115 を参照



# 平均値

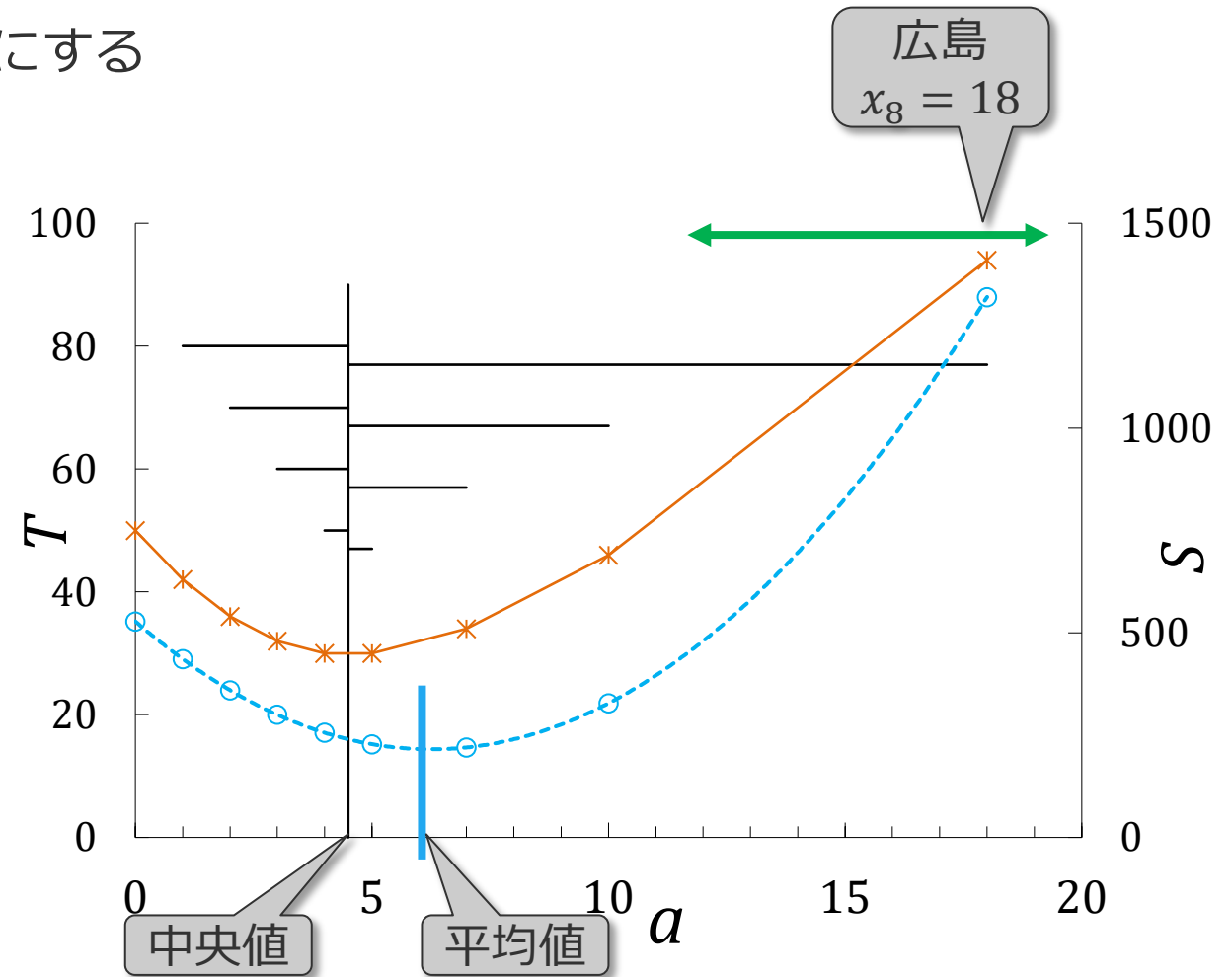
## ●中央値と平均値の比較（事例 1）

中央値：各人の移動距離（直線）の和を最小にする

$$T = \sum_{i=1}^n |e_i| = \sum_{i=1}^n |x_i - a| \Rightarrow \min \quad (2.1.1)$$

平均値：各人の移動距離の2乗（面積）の和を最小にする  
(平均からの距離が遠い値の影響が大きい)

$$S = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (x_i - a)^2 \Rightarrow \min \quad (2.1.3)$$





## ●平均値

$\bar{x}$  (エックス・バー)

$\bar{x}$ . テキストでは  $i$  について平均したことを示すため添え字ドットを付す

$$T = x_1 + x_2 + \dots + x_n = \sum_{i=1}^n x_i \qquad \bar{x} = \frac{T}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

平均値は最小 2 乗法から導かれる  
外れ値の影響が大きい

Excel による計算

=AVERAGE(データの集まり)

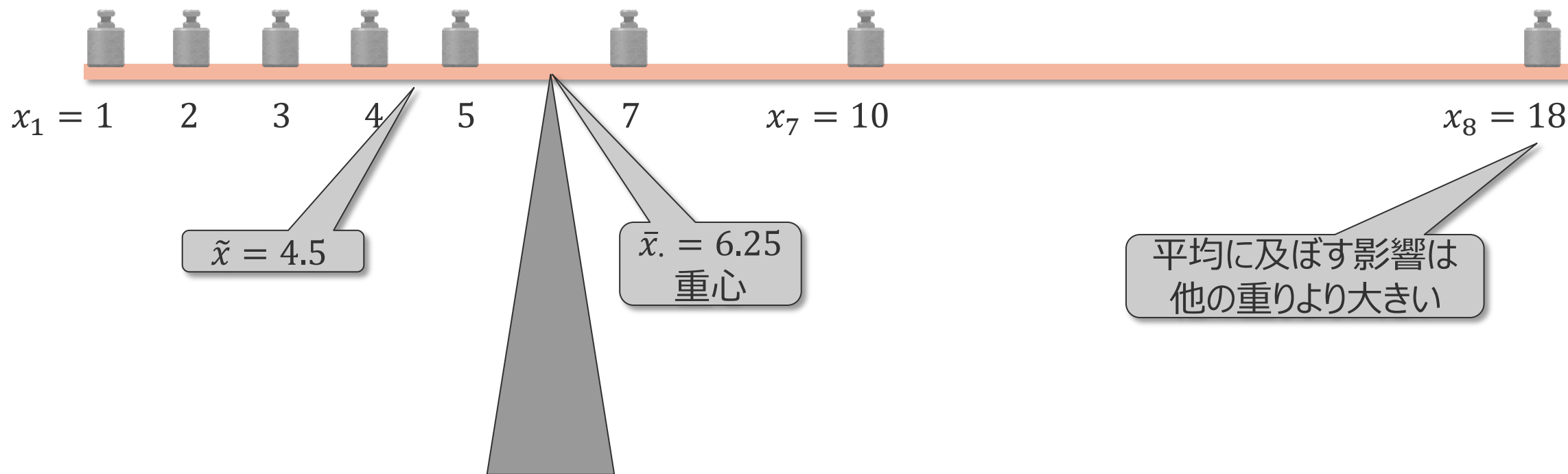
= AVERAGE(B4:B11)

= AVERAGE(B4:B8, C3, C7:C10)

## ●平均値は重心（事例 1）

棒状の板（極く軽量）の上に、同じ重量の重りを 8 個、8 都市（事例 1）の位置  $x_i$  に乗せる  
平均値  $\bar{x} = 6.25$  の位置を支点にすると釣り合う（=重心に相当）

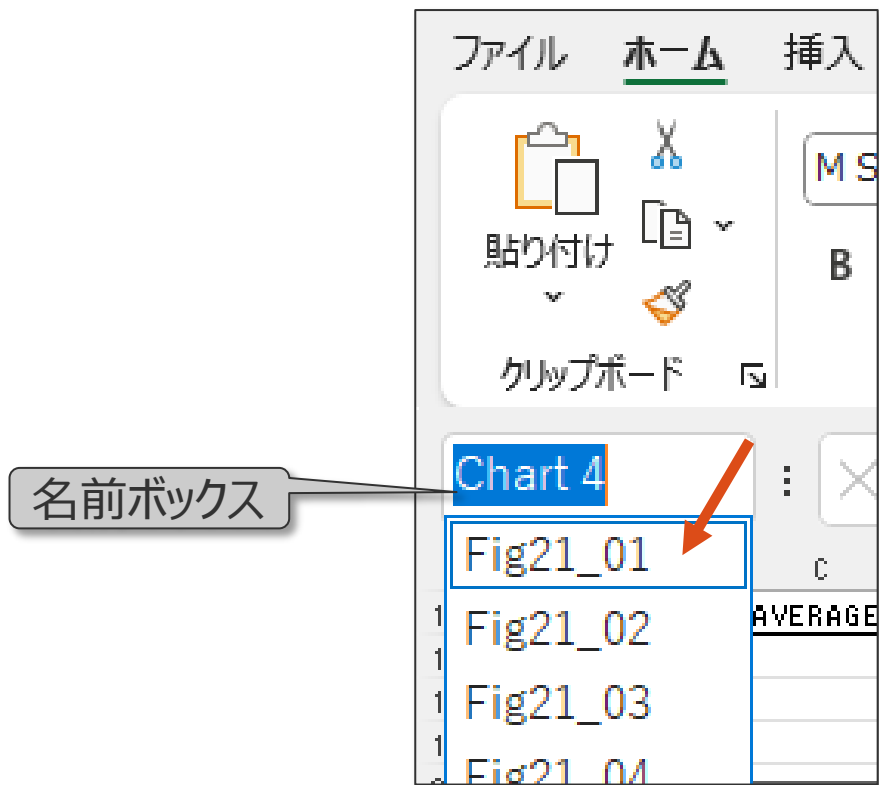
中央値  $\tilde{x} = 4.5$  の位置を支点にしても釣り合わない（支点の左右にある重りが同数）



## ●Excel 関数による計算

Excel ファイル「基礎改2.xls」

名前ボックスから「表示2.1.1」 (Fig21\_01) を選択



表示 2.1.1 データおよび $a$ と $T, S$ との関係

|    | A   | B    | C          |
|----|-----|------|------------|
| 3  | i   | x    | Excel関数    |
| 4  | 1   | 1    |            |
| 5  | 2   | 2    |            |
| 6  | 3   | 3    |            |
| 7  | 4   | 4    |            |
| 8  | 5   | 5    |            |
| 9  | 6   | 7    |            |
| 10 | 7   | 10   |            |
| 11 | 8   | 18   |            |
| 12 | 中央値 | 4.5  | =MEDIAN()  |
| 13 | n   | 8    | =COUNT()   |
| 14 | 合計  | 50   | =SUM()     |
| 15 | 平均  | 6.25 | =合計/n      |
| 16 | 平均  | 6.25 | =AVERAGE() |

## ●Excel 関数による計算

データ：事例 1（8都市から1人ずつ集まる最短の集合場合）

表示 2.1.1 データおよび  $a$  と  $T, S$  との関係

Excel 関数

引数

- =MEDIAN(B4:B11) 中央値
- =COUNT(B4:B11) 数値の数 データ数  $n$
- =SUM(B4:B11) 合計
- =B14/B13 平均 (合計 /  $n$ )
- =AVERAGE(B4:B11) 平均

セル範囲  
B4:B11

|    | A   | B    | C          |
|----|-----|------|------------|
| 3  | i   | x    | Excel関数    |
| 4  | 1   | 1    |            |
| 5  | 2   | 2    |            |
| 6  | 3   | 3    |            |
| 7  | 4   | 4    |            |
| 8  | 5   | 5    |            |
| 9  | 6   | 7    |            |
| 10 | 7   | 10   |            |
| 11 | 8   | 18   |            |
| 12 | 中央値 | 4.5  | =MEDIAN()  |
| 13 | n   | 8    | =COUNT()   |
| 14 | 合計  | 50   | =SUM()     |
| 15 | 平均  | 6.25 | =合計 / n    |
| 16 | 平均  | 6.25 | =AVERAGE() |

平均値に  
強く影響

### 中央値と平均値の比較

中央値 4.5 に比べ平均値は 6.25 と大きい → 18 が強く影響  
 誤差が正規分布に従うとき、平均値が最良の代表値  
 正規分布から大きく外れる場合は中央値が代表値として適す  
 → 頑健性 (p.65)

## ●Excel 関数による計算

Excel ファイル「基礎改2.xls」、名前ボックスから「表示2.1.2」 (Fig21\_02) を選択

表示 2.1.2 平方和、平均平方、標準偏差の計算

これ以降、  
平均、平方和、標準偏差  
変動係数の説明に使用

|    | A    | B     | C     | D      | E                | F                | G            |
|----|------|-------|-------|--------|------------------|------------------|--------------|
| 3  | i    | x     | e     | e^2    | Excel 関数         |                  |              |
| 4  | 1    | 1     | -5.25 | 27.56  |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2     | -4.25 | 18.06  |                  |                  |              |
| 6  | 3    | 3     | -3.25 | 10.56  |                  |                  |              |
| 7  | 4    | 4     | -2.25 | 5.06   |                  |                  |              |
| 8  | 5    | 5     | -1.25 | 1.56   |                  |                  |              |
| 9  | 6    | 7     | 0.75  | 0.56   |                  |                  |              |
| 10 | 7    | 10    | 3.75  | 14.06  |                  |                  |              |
| 11 | 8    | 18    | 11.75 | 138.06 |                  |                  |              |
| 12 | 平均   | 6.25  | 0     |        | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  | 215.5 | 215.5 | 215.5  | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    | 8     |       |        | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  | 7     |       |        | =B14-1           |                  |              |
| 16 | 平均平方 | 30.79 | 30.79 |        | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 | 5.55  | 5.55  |        | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 | 0.89  |       |        | =B17/B12         |                  |              |

## ●Excel 関数による計算

データ  $x_i$  とその平均値  $\bar{x}$  を算出

表示 2.1.2 平方和、平均平方、標準偏差の計算

セル範囲  
B4:B11

|    | A    | B     | C     | D      | E                | F                | G            |
|----|------|-------|-------|--------|------------------|------------------|--------------|
| 3  | i    | x     | e     | e^2    | Excel 関数         |                  |              |
| 4  | 1    | 1     | -5.25 | 27.56  |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2     | -4.25 | 18.06  |                  |                  |              |
| 6  | 3    | 3     | -3.25 | 10.56  |                  |                  |              |
| 7  | 4    | 4     | -2.25 | 5.06   |                  |                  |              |
| 8  | 5    | 5     | -1.25 | 1.56   |                  |                  |              |
| 9  | 6    | 7     | 0.75  | 0.56   |                  |                  |              |
| 10 | 7    | 10    | 3.75  | 14.06  |                  |                  |              |
| 11 | 8    | 18    | 11.75 | 158.06 |                  |                  |              |
| 12 | 平均   | 6.25  | 0     |        | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  | 215.5 | 215.5 | 215.5  | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    | 8     |       |        | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  | 7     |       |        | =B14-1           |                  |              |
| 16 | 平均平方 | 30.79 | 30.79 |        | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 | 5.55  | 5.55  |        | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 | 0.89  |       |        | =B17/B12         |                  |              |

$$\bar{x} = \frac{T}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$



## (3) 平方和

ばつらきの指標：平方和、平均平方、標準偏差、変動係数

## ●事例 1 : 8都市から1人ずつ集まる最短の集合場合

観測値のばらつき  $e_i$  の2乗和を最小にする  $a$  : 平均値

$$S = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (x_i - a)^2 \Rightarrow \min \quad (2.1.3)$$

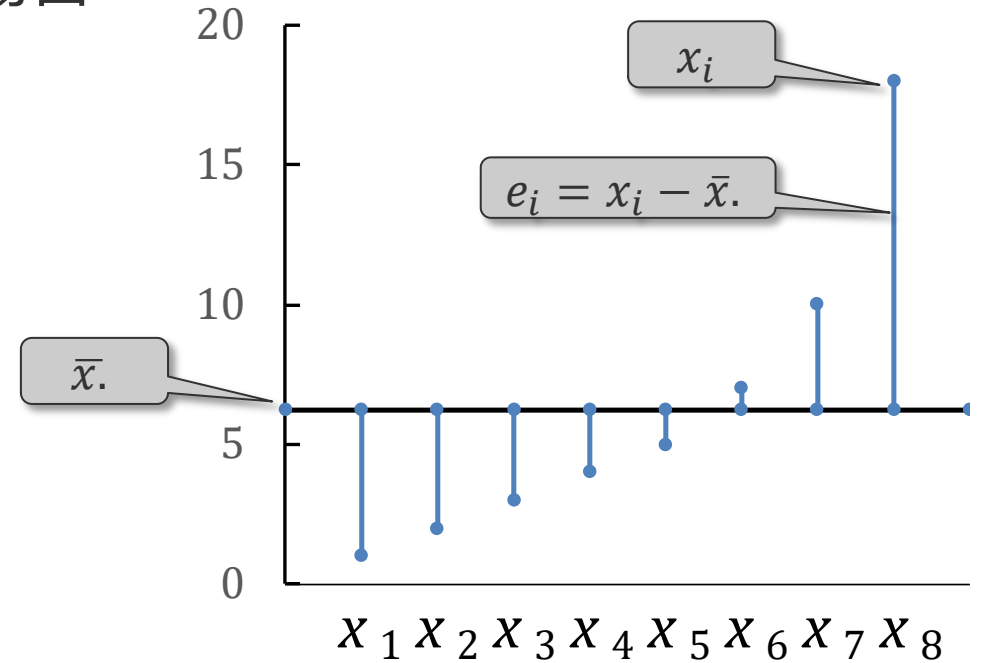
平均値を求めた式 (2.1.3) の  $a$  に平均  $\bar{x}$  を代入

$$S = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n e_i^2 \quad (2.1.6)$$

$e_i = x_i - \bar{x}$  残差 (観測値とサンプルの平均値との距離、Residual Sum of Squares)

$S$  残差平方和 (平方和) . . . ばらつきの総量を表す

注) 偏差 (観測値と母平均との距離)  $x_i - \mu$



## ●Excel による計算

データ  $x_i$ 、残差  $e_i$ 、残差の2乗値  $e_i^2$  と、その平均値を算出

表示 2.1.2 平方和、平均平方、標準偏差の計算

|    | A    | B     | C     | D              | E                | F                | G            |
|----|------|-------|-------|----------------|------------------|------------------|--------------|
|    | i    | x     | e     | e <sup>2</sup> |                  | Excel 関数         |              |
| 4  | 1    | 1     | -5.25 | 27.56          |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2     | -4.25 | 18.06          |                  |                  |              |
| 6  | 3    | 3     | -3.25 | 10.56          |                  |                  |              |
| 7  | 4    | 4     | -2.25 | 5.06           |                  |                  |              |
| 8  | 5    | 5     | -1.25 | 1.56           |                  |                  |              |
| 9  | 6    | 7     | 0.75  | 0.56           |                  |                  |              |
| 10 | 7    | 10    | 3.75  | 14.06          |                  |                  |              |
| 11 | 8    | 18    | 11.75 | 138.06         |                  |                  |              |
| 12 | 平均   | 6.25  | 0     |                | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  | 215.5 | 215.5 | 215.5          | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    | 8     |       |                | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  | 7     |       |                | =B14-1           |                  |              |
| 16 | 平均平方 | 30.79 | 30.79 |                | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 | 5.55  | 5.55  |                | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 | 0.89  |       |                | =B17/B12         |                  |              |

セル範囲  
B4:B11

$$e_i = x_i - \bar{x}$$

$$e_i^2$$

## ●Excel による計算

データ  $x_i$ 、残差  $e_i$ 、残差の2乗値  $e_i^2$  と、その平均値を算出

表示 2.1.2 平方和、平均平方、標準偏差の計算

|    | A    | B     | C     | D      | E                | F                | G            |
|----|------|-------|-------|--------|------------------|------------------|--------------|
|    | i    | x     | e     | e^2    | Excel 関数         |                  |              |
| 4  | 1    | 1     | -5.25 | 27.56  |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2     | -4.25 | 18.06  |                  |                  |              |
| 6  | 3    | 3     | -3.25 | 10.56  |                  |                  |              |
| 7  | 4    | 4     | -2.25 | 5.06   |                  |                  |              |
| 8  | 5    | 5     | -1.25 | 1.56   |                  |                  |              |
| 9  | 6    | 7     | 0.75  | 0.56   |                  |                  |              |
| 10 | 7    | 10    | 3.75  | 14.06  |                  |                  |              |
| 11 | 8    | 18    | 11.75 | 138.06 |                  |                  |              |
| 12 | 平均   | 6.25  | 0     |        | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  | 215.5 | 215.5 | 215.5  | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    | 8     |       |        | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  | 7     |       |        | =B14-1           |                  |              |
| 16 | 平均平方 | 30.79 | 30.79 |        | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 | 5.55  | 5.55  |        | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 | 0.89  |       |        | =B17/B12         |                  |              |

セル範囲  
B4:B11

$$e_i = x_i - \bar{x}$$

$$\sum_{i=1}^n e_i = \sum_{i=1}^n (x_i - \bar{x}) = 0 \quad (2.1.4)$$

合計が 0 と同等

## ●Excel による計算

データ  $x_i$ 、残差  $e_i$ 、残差の2乗値  $e_i^2$  と、その平均値を算出

表示 2.1.2 平方和、平均平方、標準偏差の計算

同一の平方和を計算

=DEVSQ(B4:B11)

=SUMSQ(C4:C11)

=SUM(D4:D11)

関数の混同に注意

$$S = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$= \sum_{i=1}^n e_i^2 \quad (2.1.6)$$

|    | A    | B     | C     | D      | E                | F                | G            |
|----|------|-------|-------|--------|------------------|------------------|--------------|
| 3  | i    | x     | e     | e^2    |                  | Excel 関数         |              |
| 4  | 1    | 1     | -5.25 | 27.56  |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2     | -4.25 | 18.06  |                  |                  |              |
| 6  | 3    | 3     | -3.25 | 10.56  |                  |                  |              |
| 7  | 4    | 4     | -2.25 | 5.06   |                  |                  |              |
| 8  | 5    | 5     | -1.25 | 1.56   |                  |                  |              |
| 9  | 6    | 7     | 0.75  | 0.56   |                  |                  |              |
| 10 | 7    | 10    | 3.75  | 14.06  |                  |                  |              |
| 11 | 8    | 18    | 11.75 | 138.06 |                  |                  |              |
| 12 | 平均   | 6.25  | 0     |        | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  | 215.5 | 215.5 | 215.5  | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    | 8     |       |        | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  | 7     |       |        | =B14-1           |                  |              |
| 16 | 平均平方 | 30.79 | 30.79 |        | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 | 5.55  | 5.55  |        | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 | 0.89  |       |        | =B17/B12         |                  |              |

$e_i = x_i - \bar{x}$

$e_i^2$



## (4) 平均平方、標準偏差と変動係数

ばつらきの指標：平方和、平均平方、標準偏差、変動係数



- 平均平方 (Mean square、分散 (Variance))

平方和 (S) : データ数が増えると大きくなる (ばらつきの大きさの比較につかえない)

$$S = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.1.3)$$

平均平方 (V) : データ 1 個当たりの平方和 (分散) 、ばらつきが比較できる

データ数  $n$  ではなく自由度  $\nu = (n - 1)$  で割る (p.99、p.117参照)

$$V = \frac{S}{n - 1} = \frac{S}{\nu} \quad (2.2.7)$$

二一

(注) 平方和を  $n$  で割る分散もあるが、自由度  $\nu = (n - 1)$  で割る方法で統一する  
不偏分散という名称もあるが、現在では使われなくなった

平方和の平均 (Sum of Squares の平均) の意、これが平均平方 (Mean Square) に短縮

- 平均平方 (Mean square、分散 (Variance))

平方和 (S) : データ数が増えると大きくなる (ばらつきの大きさの比較につかえない)

$$S = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.1.3)$$

平均平方は分散の概念を拡張したものの  
分散は1つの群内でのばらつきの指標  
平均平方は複数の群間での変動も含め、  
より広範囲なばらつきの指標として利用

平均平方 (V) : データ1個当たりの平方和 (分散)、ばらつきが比較できる

データ数  $n$  ではなく自由度  $\nu = (n - 1)$  で割る (p.99、p.117参照)

$$V = \frac{S}{n - 1} = \frac{S}{\nu} \quad (2.2.7)$$

統一

(注) 平方和を  $n$  で割る分散もあるが、自由度  $\nu = (n - 1)$  で割る方法で統一する  
不偏分散という名称もあるが、現在では使われなくなった

平方和の平均 (Sum of Squares の平均) の意、これが平均平方 (Mean Square) に短縮



- 標準偏差 (Standard Deviation)

標準偏差 ( $s$ ) : 平均平方 (分散) の平方根、s.d. または SD と表すこともある

$V$  の単位は観測値の 2 乗、観測値の平方根を取って単位を戻す

$$S = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$V = \frac{S}{n-1} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$s = \sqrt{V}$$

小文字

## ●自由度 (Degree of Freedom)

自由度の表記

$\nu$  (ニュー)

$\phi$  (ファイ)

$f$

平方和の自由度の意味

自由に動けるデータの個数

平方和  $S$  は  $n$  個の  $e_i$  を 2 乗して加えたもの

残差  $e_i$  の合計は必ず 0 になるので、

任意の  $(n-1)$  個の  $e_i$  を決めると、

残りの 1 個は自動的に決まる

( $e_i$  は、 $x$  と平均値との差であるから)

自由に決められる個数は  $(n-1)$  個

$$S = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n e_i^2 \quad (2.1.6)$$

| $i$                 | $x_i$                    | $e_i$           | $e_i^2$         |
|---------------------|--------------------------|-----------------|-----------------|
| 1                   | 1                        | -5.25           | 27.56           |
| 2                   | 2                        | -4.25           | 18.06           |
| 3                   | 3                        | -3.25           | 10.56           |
| 4                   | 4                        | -2.25           | 5.06            |
| 5                   | 5                        | -1.25           | 1.56            |
| 6                   | 7                        | 0.75            | 0.56            |
| 7                   | 10                       | 3.75            | 14.06           |
| 8                   | 18                       | 11.8            | 138.06          |
| 合計、<br>$n$ 、<br>平均、 | 平方和<br>自由度 $\nu$<br>平均平方 | 50<br>8<br>6.25 | 0<br>7<br>30.79 |

8 ではなく 7 で平方和を割る

# 母数と統計量

## ●母数（パラメータ）の定義と性質

母平均  $\mu$  (p.16)

$$\mu = E[x]$$

母分散  $\sigma^2$  (p.17)

$$\sigma^2 = V[x] = E[(x - E[x])^2]$$

母標準偏差  $\sigma$

$$\sigma = D[x] = \sqrt{V[x]}$$

母分散の性質 (第1章参照)

$$V[ax + b] = a^2V[x]$$

$$V[x + y] = V[x] + V[y]$$

( $x, y$  が独立、分散の加法性)

$$V[a_1x_1 + a_2x_2 + \cdots + a_nx_n]$$

$$= a_1^2V[x_1] + a_2^2V[x_2] + \cdots + a_n^2V[x_n]$$

( $x_1, x_2, \cdots, x_n$  が互いに独立、分散の加法性)

# 母数と統計量

## ●統計量の定義

(1) 母集団の中心位置を推測する統計量

平均  $\bar{x}$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

中央値  $\tilde{x}$

観測値を昇順に並べ替えて

$n$  が奇数個なら中央の観測値

$n$  が偶数個なら中央の2つの観測値の平均

(2) 母集団の広がり具合を推測する統計量

平方和  $S$

$$S = \sum_{i=1}^n (x_i - \bar{x})^2$$

平均平方 (分散)  $V$

$$V = \frac{S}{n-1} = \frac{S}{\nu}$$

自由度 ( $n-1$ )

標準偏差  $s$

$$s = \sqrt{V}$$

注)  $n$  で割る平均平方もあるが、ここでは取り上げない

# 母数と統計量

## ● サンプルの統計量による母数の推定

サンプルの平均 ( $\bar{x}$ ) は母平均 ( $\mu$ ) の偏りのない推定値

$$E[\bar{x}] = \mu$$

サンプルの平均平方 (分散、 $V$ ) は母分散 ( $\sigma^2$ ) の偏りのない推定値

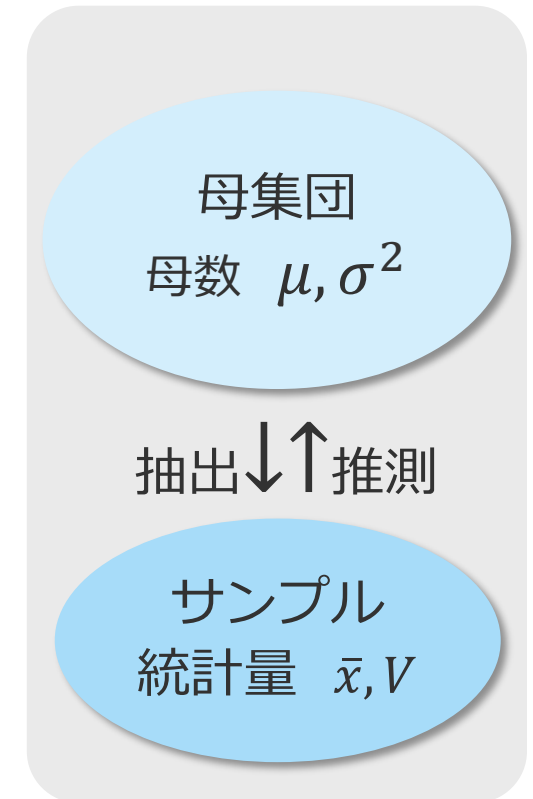
$$E[V] = \sigma^2$$

サンプルの標準偏差 ( $s$ ) の期待値は母標準偏差 ( $\sigma$ ) よりもやや小さい

$s^2$  の期待値は  $\sigma^2$  であるが、 $s$  の期待値は  $\sigma$  ではない

ただし、 $\sigma$  の推定値としても、現実の解析には問題はない

$$E[s] < \sigma \quad (E[s] \sim \sigma) \quad (\text{サンプルサイズが小さいとき注意})$$



# 変動係数

## ●変動係数の算出

変動係数

標準偏差と平均値の比可（100倍して%で表すこともある）

Coefficient of Variation、CV

$$CV = \frac{s}{\bar{x}}$$

単位に依存しない：単位がないため、異なる単位や尺度のデータを比較可

相対的なばらつきの評価：平均値の異なるデータ同士を比較する際に有用

格差の指標：経済分析における所得格差、地域格差など

間隔尺度では計算できない（摂氏温度、テストの点数、偏差値、対数変換値）

マイナスの値があるデータでは計算できない、プラスの値でも0に近い値がある場合に注意

# 変動係数

## ●変動係数の利用

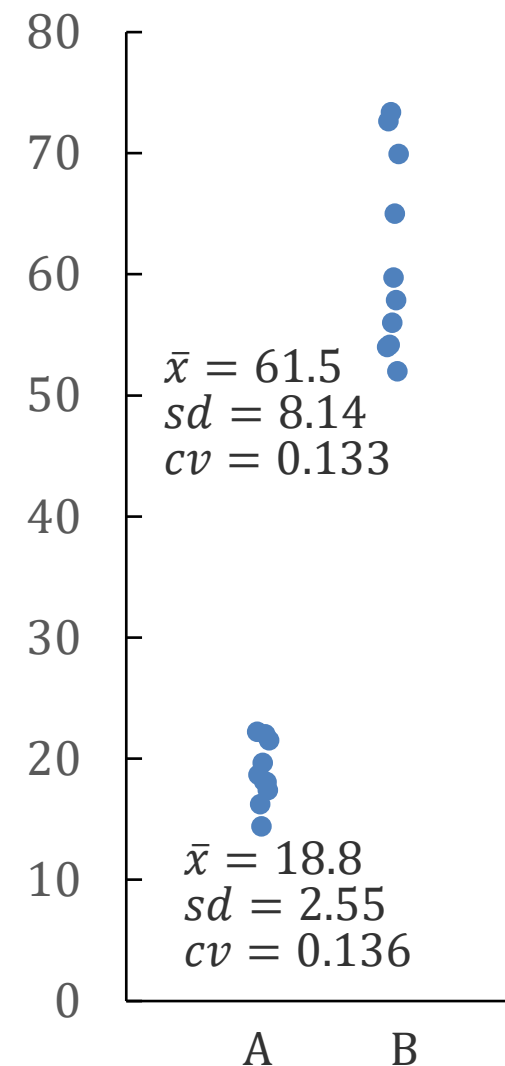
単位が異なる場合、標準偏差ではなく、変動係数で比較  
平均値が大きく異なる場合、ばらつきの比較を変動係数で行う  
(平均値が高いほどバラツキは大きい傾向がある)

事例 a : 男子高校 3 年生の身長と体重のバラツキの比較  
体重の変動係数は身長の変動係数よりも顕著に大きい

事例 b : 男子高校 3 年生と男子園児の体重のバラツキの比較  
変動係数は、両者の間で顕著な差がみられない

| 事例 a<br>学校保健統計調査<br>(文部科学省) | 男子高校 3 年生  |            |
|-----------------------------|------------|------------|
|                             | 身長<br>(cm) | 体重<br>(kg) |
| 平均                          | 170.7      | 62.5       |
| 標準偏差                        | 5.78       | 10.1       |
| 変動係数                        | 0.034      | 0.162      |

| 事例 b | 男子の体重       |            |
|------|-------------|------------|
|      | 高校生<br>(kg) | 園児<br>(kg) |
| 平均   | 62.5        | 18.9       |
| 標準偏差 | 10.1        | 2.61       |
| 変動係数 | 0.162       | 0.138      |



## ●Excel 関数による計算

表示2.1.2 に使われている Excel 関数

=AVERAGE(B4:B11)

平均

=DEVSQ(B4:B11)

残差平方和

=SUMSQ(C4:C11)

2乗和

=VAR(B4:B11)

平均平方 (分散) 新しい関数は=VAR.S

=STDEV(B4:B11)

標準偏差

新しい関数は=STDEV.S

=SQRT(C16)

平方根

=COUNT(B4:B11)

数値の個数

両者は間違いやすい

( ) 内の数値から平均値を引いた差の2乗の和

( ) 内の数値をそのまま2乗した和

=VAR.Pは  $S/n$

=STDEV.Pは  $\sqrt{S/n}$

使わない

注) 変動係数を直接求める関数はない

新しい関数を使っていく

# 平均平方、標準偏差、変動係数

## ●Excel による計算

表示 2.1.2 平方和、平均平方、標準偏差の計算

|    | A    | B     | C     | D              | E                | F                | G            |
|----|------|-------|-------|----------------|------------------|------------------|--------------|
|    | i    | x     | e     | e <sup>2</sup> | Excel 関数         |                  |              |
| 4  | 1    | 1     | -5.25 | 27.56          |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2     | -4.25 | 18.06          |                  |                  |              |
| 6  | 3    | 3     | -3.25 | 10.56          |                  |                  |              |
| 7  | 4    | 4     | -2.25 | 5.06           |                  |                  |              |
| 8  | 5    | 5     | -1.25 | 1.56           |                  |                  |              |
| 9  | 6    | 7     | 0.75  | 0.56           |                  |                  |              |
| 10 | 7    | 10    | 3.75  | 14.06          |                  |                  |              |
| 11 | 8    | 18    | 11.75 | 138.06         |                  |                  |              |
| 12 | 平均   | 6.25  | 0     |                | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  | 215.5 | 215.5 | 215.5          | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    | 8     |       |                | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  | 7     |       |                | =B14-1           |                  |              |
| 16 | 平均平方 | 30.79 | 30.79 |                | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 | 5.55  | 5.55  |                | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 | 0.89  |       |                | =B17/B12         |                  |              |

セル範囲  
B4:B11

## ●Excel による計算

セルに入力されている計算式を消去、計算式を再入力して表を再現

表示 2.1.2 平方和、平均平方、標準偏差の計算

ポイント

範囲指定の後 DEL キー

オートフィル機能

(+マークの利用)

相対参照と絶対参照

(\$マークの利用)

(F4キーの利用)

[§1.1](#) を参照

計算式の消去、再入力

|    | A    | B  | C | D   | E                | F                | G            |
|----|------|----|---|-----|------------------|------------------|--------------|
| 3  | i    | x  | e | e^2 | Excel 関数         |                  |              |
| 4  | 1    | 1  |   |     |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2  |   |     |                  |                  |              |
| 6  | 3    | 3  |   |     |                  |                  |              |
| 7  | 4    | 4  |   |     |                  |                  |              |
| 8  | 5    | 5  |   |     |                  |                  |              |
| 9  | 6    | 7  |   |     |                  |                  |              |
| 10 | 7    | 10 |   |     |                  |                  |              |
| 11 | 8    | 18 |   |     |                  |                  |              |
| 12 | 平均   |    |   |     | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  |    |   |     | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    |    |   |     | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  |    |   |     | =B14-1           |                  |              |
| 16 | 平均平方 |    |   |     | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 |    |   |     | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 |    |   |     | =B17/B12         |                  |              |

# 平均平方、標準偏差、変動係数

## ●Excel による計算

- (1) セル範囲 B4:B11 に 8 個のデータ（事例 1）を入力
- (2) セル B12 に「=AVERAGE(B4:B11)」を入力して平均値を得る

|    | A    | B  | C | D              | E                | F                | G            |
|----|------|----|---|----------------|------------------|------------------|--------------|
|    | i    | x  | e | e <sup>2</sup> | Excel 関数         |                  |              |
| 4  | 1    | 1  |   |                |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2  |   |                |                  |                  |              |
| 6  | 3    | 3  |   |                |                  |                  |              |
| 7  | 4    | 4  |   |                |                  |                  |              |
| 8  | 5    | 5  |   |                |                  |                  |              |
| 9  | 6    | 7  |   |                |                  |                  |              |
| 10 | 7    | 10 |   |                |                  |                  |              |
| 11 | 8    | 18 |   |                |                  |                  |              |
| 12 | 平均   |    |   |                | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  |    |   |                | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    |    |   |                | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  |    |   |                | =B14-1           |                  |              |
| 16 | 平均平方 |    |   |                | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 |    |   |                | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 |    |   |                | =B17/B12         |                  |              |

セル範囲  
B4:B11

# 平均平方、標準偏差、変動係数

## ●Excel による計算

(3) セル C4 に、「=B4-B\$12」を入力して残差を得る

(4) セル範囲 C5:C11 に、セルC4 の計算式をオートフィルでコピーする

(5)セル C12 に、セル B12 の  
計算式をコピー  
(合計 0 を確認)

|    | A    | B    | C | D   | E                | F                | G            |
|----|------|------|---|-----|------------------|------------------|--------------|
| 3  | i    | x    | e | e^2 |                  | Excel 関数         |              |
| 4  | 1    | 1    |   |     |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2    |   |     |                  |                  |              |
| 6  | 3    | 3    |   |     |                  |                  |              |
| 7  | 4    | 4    |   |     |                  |                  |              |
| 8  | 5    | 5    |   |     |                  |                  |              |
| 9  | 6    | 7    |   |     |                  |                  |              |
| 10 | 7    | 10   |   |     |                  |                  |              |
| 11 | 8    | 18   |   |     |                  |                  |              |
| 12 | 平均   | 6.25 |   |     | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  |      |   |     | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    |      |   |     | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  |      |   |     | =B14-1           |                  |              |
| 16 | 平均平方 |      |   |     | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 |      |   |     | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 |      |   |     | =B17/B12         |                  |              |

# 平均平方、標準偏差、変動係数

## ●Excel による計算

(6) セル D4 に「=C4^2」を入力して残差の 2 乗値を得る

(7) セル範囲 D5:D11 に、セル D4 の計算式をオートフィルでコピーする

|    | A    | B    | C     | D   | E                | F                | G            |
|----|------|------|-------|-----|------------------|------------------|--------------|
| 3  | i    | x    | e     | e^2 | Excel 関数         |                  |              |
| 4  | 1    | 1    | -5.25 |     |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2    | -4.25 |     |                  |                  |              |
| 6  | 3    | 3    | -3.25 |     |                  |                  |              |
| 7  | 4    | 4    | -2.25 |     |                  |                  |              |
| 8  | 5    | 5    | -1.25 |     |                  |                  |              |
| 9  | 6    | 7    | 0.75  |     |                  |                  |              |
| 10 | 7    | 10   | 3.75  |     |                  |                  |              |
| 11 | 8    | 18   | 11.75 |     |                  |                  |              |
| 12 | 平均   | 6.25 | 0     |     | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  |      |       |     | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    |      |       |     | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  |      |       |     | =B14-1           |                  |              |
| 16 | 平均平方 |      |       |     | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 |      |       |     | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 |      |       |     | =B17/B12         |                  |              |

# 平均平方、標準偏差、変動係数

## ●Excel による計算

(8) セル B13 に「=DEVSQ(B4:B11)」を入力して平方和を得る

(9) セル C13 に「=SUMSQ(B4:B11)」を入力して平方和を得る

(10) セル D13 に

「SUM(B4:B11)」を入力して平方和を得る

|    | A    | B    | C     | D      | E                | F                | G            |
|----|------|------|-------|--------|------------------|------------------|--------------|
| 3  | i    | x    | e     | e^2    | Excel 関数         |                  |              |
| 4  | 1    | 1    | -5.25 | 27.56  |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2    | -4.25 | 18.06  |                  |                  |              |
| 6  | 3    | 3    | -3.25 | 10.56  |                  |                  |              |
| 7  | 4    | 4    | -2.25 | 5.06   |                  |                  |              |
| 8  | 5    | 5    | -1.25 | 1.56   |                  |                  |              |
| 9  | 6    | 7    | 0.75  | 0.56   |                  |                  |              |
| 10 | 7    | 10   | 3.75  | 14.06  |                  |                  |              |
| 11 | 8    | 18   | 11.75 | 138.06 |                  |                  |              |
| 12 | 平均   | 6.25 | 0     |        | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  |      |       |        | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    |      |       |        | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  |      |       |        | =B14-1           |                  |              |
| 16 | 平均平方 |      |       |        | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 |      |       |        | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 |      |       |        | =B17/B12         |                  |              |

## ●Excel による計算

(11) セル B14 に「=COUNT(B4:B11)」を入力して  $n$  を得る

(12) セル B15 に「=B14-1」を入力して自由度を得る

|    | A    | B     | C     | D              | E                | F                | G            |
|----|------|-------|-------|----------------|------------------|------------------|--------------|
| 3  | i    | x     | e     | e <sup>2</sup> | Excel 関数         |                  |              |
| 4  | 1    | 1     | -5.25 | 27.56          |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2     | -4.25 | 18.06          |                  |                  |              |
| 6  | 3    | 3     | -3.25 | 10.56          |                  |                  |              |
| 7  | 4    | 4     | -2.25 | 5.06           |                  |                  |              |
| 8  | 5    | 5     | -1.25 | 1.56           |                  |                  |              |
| 9  | 6    | 7     | 0.75  | 0.56           |                  |                  |              |
| 10 | 7    | 10    | 3.75  | 14.06          |                  |                  |              |
| 11 | 8    | 18    | 11.75 | 138.06         |                  |                  |              |
| 12 | 平均   | 6.25  | 0     |                | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  | 215.5 | 215.5 | 215.5          | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    |       |       |                | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  |       |       |                | =B14-1           |                  |              |
| 16 | 平均平方 |       |       |                | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 |       |       |                | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 |       |       |                | =B17/B12         |                  |              |

# 平均平方、標準偏差、変動係数

## ●Excel による計算

(13) セル B16 に「=VAR(B4:B11)」を入力して平均平方を得る

(14) セル C16 に「=C13 / B15」を入力して平均平方を得る

|    | A    | B     | C     | D      | E                | F                | G            |
|----|------|-------|-------|--------|------------------|------------------|--------------|
| 3  | i    | x     | e     | e^2    | Excel 関数         |                  |              |
| 4  | 1    | 1     | -5.25 | 27.56  |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2     | -4.25 | 18.06  |                  |                  |              |
| 6  | 3    | 3     | -3.25 | 10.56  |                  |                  |              |
| 7  | 4    | 4     | -2.25 | 5.06   |                  |                  |              |
| 8  | 5    | 5     | -1.25 | 1.56   |                  |                  |              |
| 9  | 6    | 7     | 0.75  | 0.56   |                  |                  |              |
| 10 | 7    | 10    | 3.75  | 14.06  |                  |                  |              |
| 11 | 8    | 18    | 11.75 | 138.06 |                  |                  |              |
| 12 | 平均   | 6.25  | 0     |        | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  | 215.5 | 215.5 | 215.5  | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    | 8     |       |        | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  | 7     |       |        | =B14-1           |                  |              |
| 16 | 平均平方 |       |       |        | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 |       |       |        | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 |       |       |        | =B17/B12         |                  |              |

# 平均平方、標準偏差、変動係数

## ●Excelによる計算

(15) セル B17 に「=STDEV(B4:B11)」を入力して標準偏差を得る

(16) セル C17 に「=SQRT(C16)」を入力して標準偏差を得る

|    | A    | B     | C     | D      | E                | F                | G            |
|----|------|-------|-------|--------|------------------|------------------|--------------|
| 3  | i    | x     | e     | e^2    | Excel 関数         |                  |              |
| 4  | 1    | 1     | -5.25 | 27.56  |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2     | -4.25 | 18.06  |                  |                  |              |
| 6  | 3    | 3     | -3.25 | 10.56  |                  |                  |              |
| 7  | 4    | 4     | -2.25 | 5.06   |                  |                  |              |
| 8  | 5    | 5     | -1.25 | 1.56   |                  |                  |              |
| 9  | 6    | 7     | 0.75  | 0.56   |                  |                  |              |
| 10 | 7    | 10    | 3.75  | 14.06  |                  |                  |              |
| 11 | 8    | 18    | 11.75 | 138.06 |                  |                  |              |
| 12 | 平均   | 6.25  | 0     |        | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  | 215.5 | 215.5 | 215.5  | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    | 8     |       |        | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  | 7     |       |        | =B14-1           |                  |              |
| 16 | 平均平方 | 30.79 | 30.79 |        | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 |       |       |        | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 |       |       |        | =B17/B12         |                  |              |

# 平均平方、標準偏差、変動係数

## ●Excelによる計算

(17) セル B18 に「=B17/B12」を入力して変動係数を得る

|    | A    | B     | C     | D      | E                | F                | G            |
|----|------|-------|-------|--------|------------------|------------------|--------------|
| 3  | i    | x     | e     | e^2    | Excel 関数         |                  |              |
| 4  | 1    | 1     | -5.25 | 27.56  |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2     | -4.25 | 18.06  |                  |                  |              |
| 6  | 3    | 3     | -3.25 | 10.56  |                  |                  |              |
| 7  | 4    | 4     | -2.25 | 5.06   |                  |                  |              |
| 8  | 5    | 5     | -1.25 | 1.56   |                  |                  |              |
| 9  | 6    | 7     | 0.75  | 0.56   |                  |                  |              |
| 10 | 7    | 10    | 3.75  | 14.06  |                  |                  |              |
| 11 | 8    | 18    | 11.75 | 138.06 |                  |                  |              |
| 12 | 平均   | 6.25  | 0     |        | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  | 215.5 | 215.5 | 215.5  | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    | 8     |       |        | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  | 7     |       |        | =B14-1           |                  |              |
| 16 | 平均平方 | 30.79 | 30.79 |        | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 | 5.55  | 5.55  |        | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 |       |       |        | =B17/B12         |                  |              |

## ●Excel による計算

セルに入力されている計算式を消去、計算式を再入力して表を再現

表示 2.1.2 平方和、平均平方、標準偏差の計算

|    | A    | B     | C     | D      | E                | F                | G            |
|----|------|-------|-------|--------|------------------|------------------|--------------|
| 3  | i    | x     | e     | e^2    | Excel 関数         |                  |              |
| 4  | 1    | 1     | -5.25 | 27.56  |                  | =B4-B\$12        | =C4^2        |
| 5  | 2    | 2     | -4.25 | 18.06  |                  |                  |              |
| 6  | 3    | 3     | -3.25 | 10.56  |                  |                  |              |
| 7  | 4    | 4     | -2.25 | 5.06   |                  |                  |              |
| 8  | 5    | 5     | -1.25 | 1.56   |                  |                  |              |
| 9  | 6    | 7     | 0.75  | 0.56   |                  |                  |              |
| 10 | 7    | 10    | 3.75  | 14.06  |                  |                  |              |
| 11 | 8    | 18    | 11.75 | 138.06 |                  |                  |              |
| 12 | 平均   | 6.25  | 0     |        | =AVERAGE(B4:B11) | =AVERAGE(C4:C11) |              |
| 13 | 平方和  | 215.5 | 215.5 | 215.5  | =DEVSQ(B4:B11)   | =SUMSQ(C4:C11)   | =SUM(D4:D11) |
| 14 | n    | 8     |       |        | =COUNT(B4:B11)   |                  |              |
| 15 | 自由度  | 7     |       |        | =B14-1           |                  |              |
| 16 | 平均平方 | 30.79 | 30.79 |        | =VAR(B4:B11)     | =C13/B15         |              |
| 17 | 標準偏差 | 5.55  | 5.55  |        | =STDEV(B4:B11)   | =SQRT(C16)       |              |
| 18 | 変動係数 | 0.89  |       |        | =B17/B12         |                  |              |



## (5) 間隔尺度と比例尺度

変動係数に関連して説明



## ●間隔尺度と比例尺度（量的変数）

因子の影響や関係性を定量的に評価するために因子を変数と見なして解析  
連続尺度には間隔尺度と比例尺度がある

### 間隔尺度

間隔に意味を持ち、比は意味を持たない（値の大小関係と値の差に意味がある）

原点（0）の取り方に任意性がある量（原点が相対的で、絶対的な原点が存在しない）

加減算は可、乗除算は不可（比率に意味がない）

変動係数に意味がない

対数変換してはいけない（[§2.3](#)）

### 比例尺度

加減算は可、乗除算は可（比率に意味がある）

一般的な連続変数

原点に絶対的な意味がある（無という意味）

### 変数の種類（[§0.1](#)）

|      |      |      |
|------|------|------|
| 質的変数 | 名義変数 |      |
|      | 順序尺度 |      |
| 連続変数 | 連続尺度 | 間隔尺度 |
|      |      | 比例尺度 |

## ●間隔尺度の例と統計解析ソフトでの扱い

摂氏温度、華氏温度は間隔尺度

摂氏 40°C は 20°C の倍とは言えないが、差は20°C  
(絶対温度は比例尺度、0 に絶対的な意味あり)

西暦、昭和、対数変換値なども間隔尺度

JMP などの統計解析ソフトは

間隔尺度と比例尺度を判断できない

(JMP の場合、いずれも連続尺度)

解析者が判断して統計解析ソフトを利用する

平均値、分散、標準偏差は計算可

同じデータを 3 種類の単位で比較

|      | 摂氏温度   | 華氏温度   | 絶対温度  |
|------|--------|--------|-------|
|      | 38.5   | 101.3  | 312   |
|      | 23.3   | 73.9   | 296   |
|      | 46.3   | 115.3  | 319   |
|      | 36.9   | 98.4   | 310   |
| 平均   | 36.3   | 97.3   | 309   |
| 標準偏差 | 9.56   | 17.2   | 9.56  |
| 変動係数 | ×0.264 | ×0.177 | 0.031 |

変数の種類 ([§0.1](#))

質的変数 名義変数  
順序尺度

連続変数 連続尺度 間隔尺度  
比例尺度

統計解析ソフトは  
区別できない



## (6) 四分位値と四分位範囲

平均値－標準偏差  
中央値－四分位範囲

## ●四分位値（四分位数、四分位点）

### 四分位値

$n$  個のデータを昇順に並べる

全データ数を 4 等分する区切りの値が四分位値

下側四分位値（第 1 四分位値） $Q_1$  25% の位置

中央値（第 2 四分位値） $Q_2$  50% の位置

上側四分位値（第 3 四分位値） $Q_3$  75% の位置

四分位値の求め方には何種類もの方法がある

データ数によって区切になる位置に数値がない

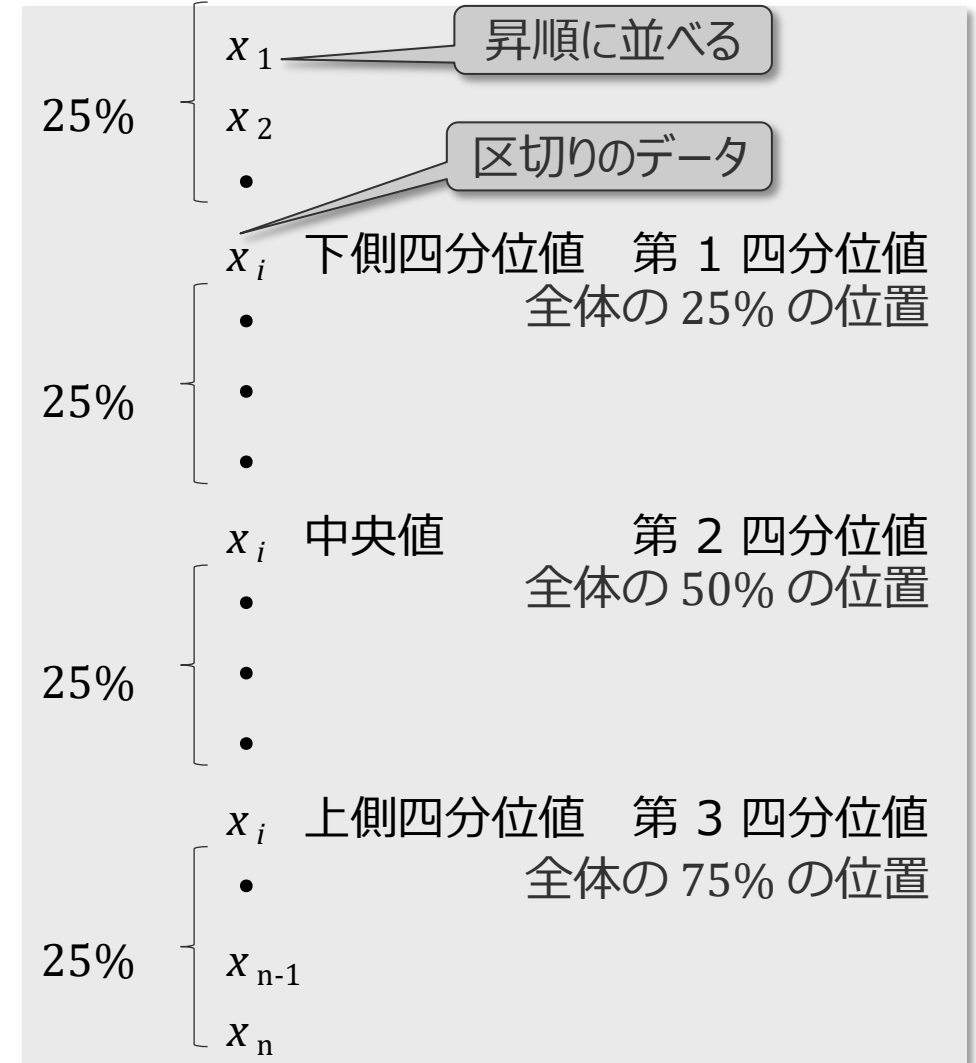
$n$  が奇数か偶数か、区切り位置を含めるか否か

統計ソフトによって採用している方法が異なる

箱ひげ図に反映（箱ひげ図も異なってくる）

データ数が多ければ違いを気にすることはない

### 四分位値のイメージ





## ●四分位値の計算

Excel の QUARTILE 関数：以下の 2 つの関数に分かれた

QUARTILE.INC：0～100% の間 (0 と 100 を含む) の百分率に基づいて配列の四分位数を返す  
QUARTILE 関数と同じ機能を持つ、今後は QUARTILE.INC を使う

QUARTILE.EXC：0～100% の間 (0 と 100 を除く) の百分率に基づいて配列の四分位数を返す  
最小値と最大値の外側に 1 つずつ空の値を補って QUARTILE.INC を実行  
QUARTILE.EXC (配列,0) と QUARTILE.EXC (配列,4) の戻り値はない

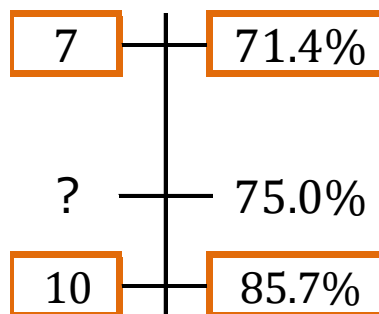
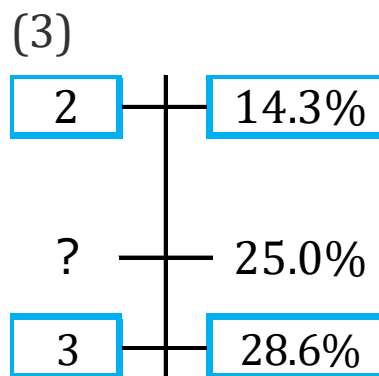
|        |                      |                               |
|--------|----------------------|-------------------------------|
| 最小値    | =QUARTILE.INC(配列, 0) | =QUARTILE.EXC(配列, 0) (戻り値はない) |
| 下側四分位値 | =QUARTILE.INC(配列, 1) | =QUARTILE.EXC(配列, 1)          |
| 中央値    | =QUARTILE.INC(配列, 2) | =QUARTILE.EXC(配列, 2)          |
| 上側四分位値 | =QUARTILE.INC(配列, 3) | =QUARTILE.EXC(配列, 3)          |
| 最大値    | =QUARTILE.INC(配列, 4) | =QUARTILE.EXC(配列, 4) (戻り値はない) |

## ●四分位値の計算

Excel の QUARTILE.INC 関数 (事例 1、 $n=8$  偶数の場合)

| (1) |      | (2)    |           |
|-----|------|--------|-----------|
| $x$ |      | $x$ 順位 | 率 (%)     |
| 2   |      | 1 1    | 0.0 最小値   |
| 7   | →    | 2 2    | 14.3 ←25% |
| 1   |      | 3 3    | 28.6      |
| 4   | 昇順に  | 4 4    | 42.9      |
| 18  | 並び替え | 5 5    | 57.1      |
| 5   |      | 7 6    | 71.4 ←75% |
| 10  |      | 10 7   | 85.7      |
| 3   |      | 18 8   | 100.0 最大値 |

$$\text{率} = (\text{順位} - 1) / (n - 1)$$



(4) 下側四分位値

$$2 + (3 - 2) \times \frac{25 - 14.3}{28.6 - 14.3} = 2.75$$

(5) 中央値

$$\frac{5 + 4}{2} = 4.50$$

(6) 上側四分位値

$$7 + (10 - 7) \times \frac{75 - 71.4}{85.7 - 71.4} = 7.75$$

順位に並び替えて数値の間を内分  
 全体の 25% の位置が下側四分位値  
 全体の 75% の位置が上側四分位値

# 四分位値

## ●四分位値の計算

Excel の QUARTILE.INC 関数 (単純な事例、サンプルサイズが小さすぎる)

$$\text{率} = (\text{順位} - 1) / (n - 1)$$

n=4 偶数

| x  | 順位 | 率(%) |
|----|----|------|
| 10 | 1  | 0    |
| 11 | 2  | 33   |
| 12 | 3  | 66.6 |
| 13 | 4  | 100  |

n=5 奇数

| x  | 順位 | 率(%) |
|----|----|------|
| 10 | 1  | 0    |
| 11 | 2  | 25   |
| 12 | 3  | 50   |
| 13 | 4  | 75   |
| 14 | 5  | 100  |

n=6 偶数

| x  | 順位 | 率(%) |
|----|----|------|
| 10 | 1  | 0    |
| 11 | 2  | 20   |
| 12 | 3  | 40   |
| 13 | 4  | 60   |
| 14 | 5  | 80   |
| 15 | 6  | 100  |

データ範囲

|        |                        |
|--------|------------------------|
| 最小値    | 10.0                   |
| 下側四分位値 | 10.8 = 0.25*10+0.75*11 |
| 中央値    | 11.5 = (11+12)/2       |
| 上側四分位値 | 12.3 = 0.75*12+0.25*13 |
| 最大値    | 13.0                   |

|                    |
|--------------------|
| 10.0               |
| 11.0 10,11,12 の中央値 |
| 12.0 全体の中央値        |
| 13.0 12,13,14 の中央値 |
| 14.0               |

|                                   |
|-----------------------------------|
| 10.0                              |
| 11.3 = 11+(12-11)*(25-20)/(40-20) |
| 12.5 = (12+13)/2                  |
| 13.8 = 13+(14-13)*(75-60)/(80-60) |
| 15.0                              |

下側四分位値 = 10+(11-10)\*(25-0)/(33-0)  
 上側四分位値 = 12+(13-12)\*(75-66.6)/(100-66.6)

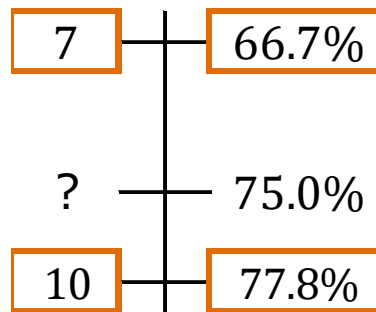
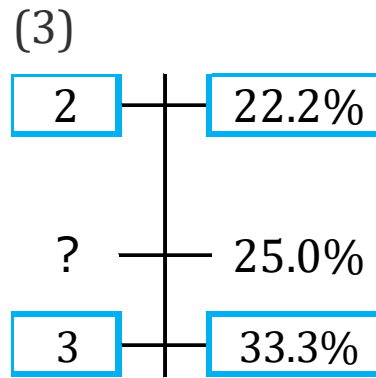
= QUARTILE.INC(データ範囲, k)  
 k=0~4

## ●四分位値の計算

Excel の QUARTILE.EXC 関数 (事例 1、 $n=8$  偶数の場合)

| (1) | (2)           |
|-----|---------------|
| $x$ | $x$ 順位 率 (%)  |
|     | 追加 1 0.0      |
| 2   | 1 2 11.1      |
| 7   | 2 3 22.2 ←25% |
| 1   | 昇順に 3 4 33.3  |
| 4   | 並び替え 4 5 44.4 |
| 18  | 5 6 55.6      |
| 5   | 7 7 66.7 ←75% |
| 10  | 10 8 77.8     |
| 3   | 18 9 88.9     |
|     | 追加 10 100.0   |

$$\text{率} = (\text{順位} - 1) / (n - 1)$$



(4) 下側四分位値

$$2 + (3 - 2) \times \frac{25 - 22.2}{33.3 - 22.2} = 2.25$$

(5) 中央値

$$\frac{5 + 4}{2} = 4.50$$

(6) 上側四分位値

$$7 + (10 - 7) \times \frac{75 - 66.7}{77.8 - 66.7} = 9.25$$

順位に並び替えて数値の間を内分  
 全体の 25% の位置が下側四分位値  
 全体の 75% の位置が上側四分位値

## ●四分位値の計算

Excel の QUARTILE.INC 関数 : 小さい方から  $(n - 1)p + 1$  番目の値 (一般化)

$n$  : データ数  
 $p$  : 0.25, 0.75

下側四分位値 ( $p=0.25$ ,  $n=8$ )

$$(n - 1)p + 1 = (8 - 1) \times 0.25 + 1 = 2.75 \text{ 番目} \quad 2 + (3 - 2) \times 0.75 = 2.75$$

2 と 3 の間を  
0.75:0.25 に内分

上側四分位値 ( $p=0.75$ ,  $n=8$ )

$$(n - 1)p + 1 = (8 - 1) \times 0.75 + 1 = 6.25 \text{ 番目} \quad 7 + (10 - 7) \times 0.25 = 7.75$$

7 と 10 の間を  
0.25:0.75 に内分

JMP : 小さい方から  $(n + 1)p$  番目の値 (一般化)

下側四分位値 ( $p=0.25$ ,  $n=8$ )

$$(n + 1)p = (8 + 1) \times 0.25 = 2.25 \text{ 番目} \quad 2 + (3 - 2) \times 0.25 = 2.25$$

上側四分位値 ( $p=0.75$ ,  $n=8$ )

$$(n + 1)p = (8 + 1) \times 0.75 = 6.75 \text{ 番目} \quad 7 + (10 - 7) \times 0.75 = 9.25$$

事例 1

| $x$ | 順位 |
|-----|----|
| 1   | 1  |
| 2   | 2  |
| 3   | 3  |
| 4   | 4  |
| 5   | 5  |
| 7   | 6  |
| 10  | 7  |
| 18  | 8  |

## ●四分位値の計算

$n$  : データ数  
 $p$  : 0.25, 0.75

Excel の QUARTILE.INC 関数 : 小さい方から  $(n - 1)p + 1$  番目の値 (一般化)

下側四分位値 ( $p=0.25$ ,  $n=8$ )

$$(n - 1)p + 1 = (8 - 1) \times 0.25 + 1 = 2.75 \text{ 番目} \quad 2 + (3 - 2) \times 0.75 = 2.75$$

上側四分位値 ( $p=0.75$ ,  $n=8$ )

$$(n - 1)p + 1 = (8 - 1) \times 0.75 + 1 = 6.25 \text{ 番目} \quad 7 + (10 - 7) \times 0.25 = 7.75$$

JMP : 小さい方から  $(n + 1)p$  番目の値 (一般化)

下側四分位値 ( $p=0.25$ ,  $n=8$ )

$$(n + 1)p = (8 + 1) \times 0.25 = 2.25 \text{ 番目} \quad 2 + (3 - 2) \times 0.25 = 2.25$$

上側四分位値 ( $p=0.75$ ,  $n=8$ )

$$(n + 1)p = (8 + 1) \times 0.75 = 6.75 \text{ 番目} \quad 7 + (10 - 7) \times 0.75 = 9.25$$

事例 1

| $x$ | 順位 |
|-----|----|
| 1   | 1  |
| 2   | 2  |
| 3   | 3  |
| 4   | 4  |
| 5   | 5  |
| 7   | 6  |
| 10  | 7  |
| 18  | 8  |

## ●四分位値の計算

Excel の QUARTILE.INC 関数：小さい方から  $(n - 1)p + 1$  番目の値（一般化）

下側四分位値 ( $p=0.25$ 、 $n=8$ )

$$(n - 1)p + 1 = (8 - 1) \times 0.25 + 1 = 2.75 \text{ 番目} \quad 2 + (3 - 2) \times 0.75 = 2.75$$

$$2 \times 0.25 + 3 \times 0.75$$

上側四分位値 ( $p=0.75$ 、 $n=8$ )

$$(n - 1)p + 1 = (8 - 1) \times 0.75 + 1 = 6.25 \text{ 番目} \quad 7 + (10 - 7) \times 0.25 = 7.75$$

$$7 \times 0.75 + 10 \times 0.25$$

JMP：小さい方から  $(n + 1)p$  番目の値（一般化）

下側四分位値 ( $p=0.25$ 、 $n=8$ )

$$(n + 1)p = (8 + 1) \times 0.25 = 2.25 \text{ 番目} \quad 2 + (3 - 2) \times 0.25 = 2.25$$

$$2 \times 0.75 + 3 \times 0.25$$

上側四分位値 ( $p=0.75$ 、 $n=8$ )

$$(n + 1)p = (8 + 1) \times 0.75 = 6.75 \text{ 番目} \quad 7 + (10 - 7) \times 0.75 = 9.25$$

$$7 \times 0.25 + 10 \times 0.75$$



## ●四分位値の計算

高校の教科書における定義（補足）

データの個数が偶数個の場合

中央値は中央に位置する 2 つの値の平均値

中央値で 2 つのグループに分け、

それぞれのグループの中央値を下側四分位値と上側四分位値にする

データの個数が奇数個の場合

中央値は中央に位置する値

中央値を除外して 2 つのグループに分け、

それぞれのグループの中央値を下側四分位値と上側四分位値にする

統計ソフト R には 9 種類の四分位値を出力する関数がある

## ●四分位範囲と標準偏差の近似推定値

### 四分位範囲

四分位範囲 = 上側四分位値 - 下側四分位値  
ほぼ半分のデータが含まれる

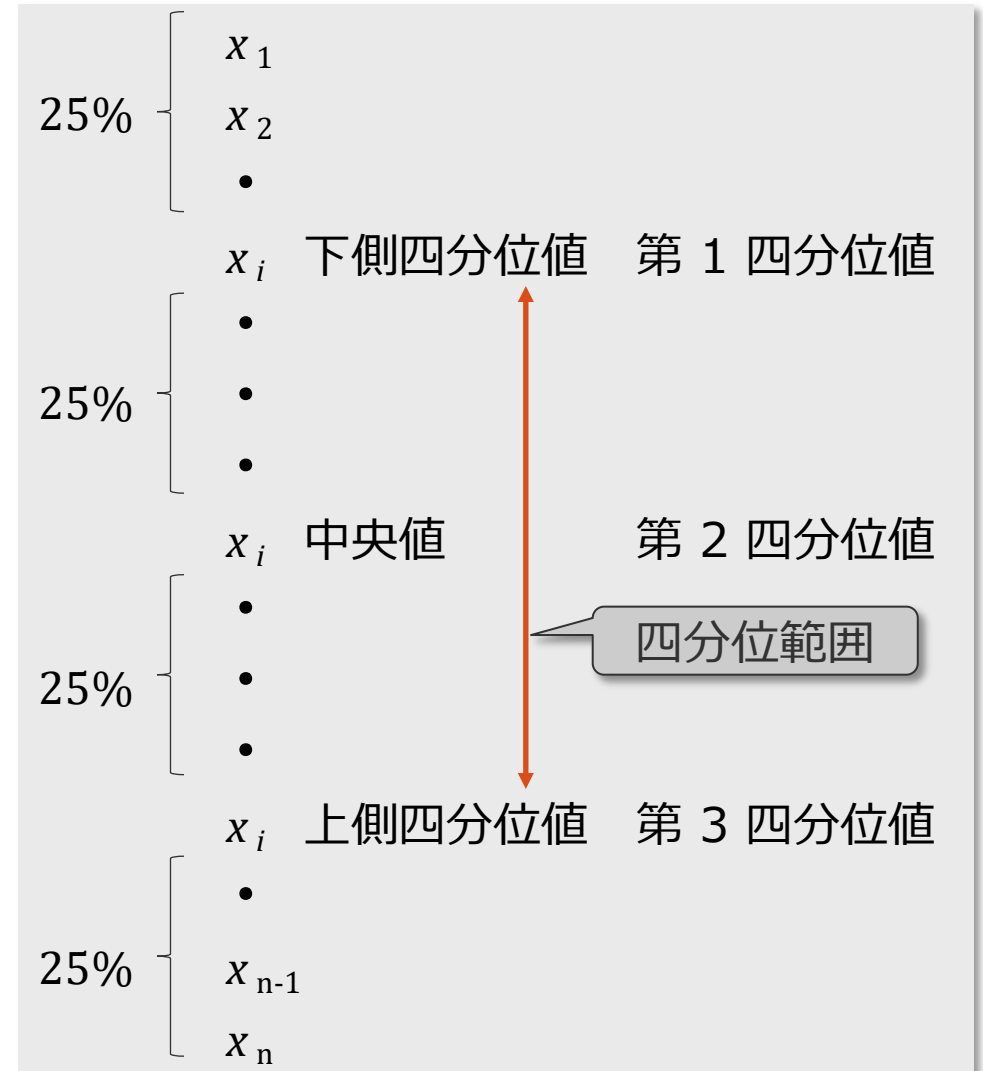
### 標準偏差の近似推定値 $s^*$

データが正規分布に従うとき,  
データの50%が含まれる範囲は  $\mu \pm 0.675\sigma$   
(§1.3)

$$\text{四分位範囲} \approx 0.675\sigma \times 2 = 1.35\sigma$$

$$s^* = \frac{\text{四分位範囲}}{1.35}$$

外れ値の影響を受けにくい (頑健性がある)



## ●四分位範囲と標準偏差の近似推定値

### 四分位範囲

四分位範囲 = 上側四分位値 - 下側四分位値  
ほぼ半分のデータが含まれる

### 標準偏差の近似推定値 $s^*$

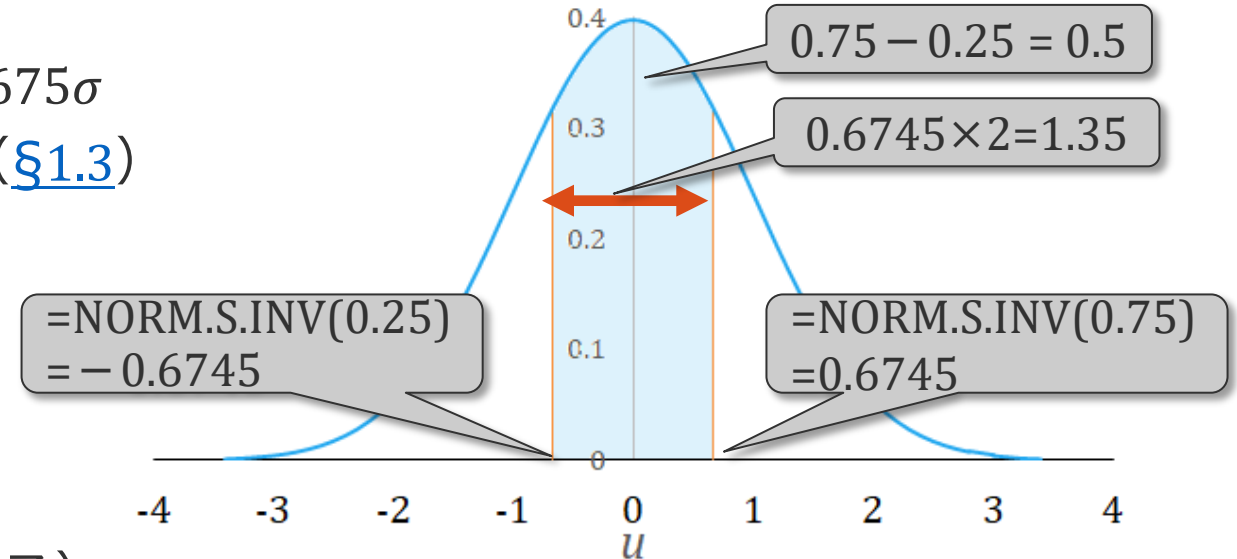
データが正規分布に従うとき,  
データの50%が含まれる範囲は  $\mu \pm 0.675\sigma$   
(§1.3)

四分位範囲  $\approx 0.675\sigma \times 2 = 1.35\sigma$

$$s^* = \frac{\text{四分位範囲}}{1.35}$$

外れ値の影響を受けにくい (頑健性がある)

標準正規分布で50%が  
含まれる範囲は  $-0.657 < u < 0.657$   
(演習 1.3.5 p.30)



## ●Excel による計算

Excel ファイル「基礎改2.xls」

名前ボックスから「表示2.1.3」 (Fig21\_03) を選択

表示 2.1.3 四分位値、ひずみ、とがりの計算

|    | I          | J     | K     | L                       |
|----|------------|-------|-------|-------------------------|
| 3  | i          | x1    | x2    | Excel 関数                |
| 4  | 1          | 1     | 1     |                         |
| 5  | 2          | 2     | 3     |                         |
| 6  | 3          | 3     | 5     |                         |
| 7  | 4          | 4     | 8     |                         |
| 8  | 5          | 5     | 9     |                         |
| 9  | 6          | 7     | 11    |                         |
| 10 | 7          | 10    | 13    |                         |
| 11 | 8          | 18    | 18    |                         |
| 12 | n          | 8     | 8     | =COUNT(J4:J11)          |
| 13 | 最小値        | 1.00  | 1.00  | =QUARTILE(J\$4:J\$11,0) |
| 14 | 下側四分位値     | 2.75  | 4.50  | =QUARTILE(J\$4:J\$11,1) |
| 15 | 中央値        | 4.50  | 8.50  | =QUARTILE(J\$4:J\$11,2) |
| 16 | 上側四分位値     | 7.75  | 11.50 | =QUARTILE(J\$4:J\$11,3) |
| 17 | 最大値        | 18.00 | 18.00 | =QUARTILE(J\$4:J\$11,4) |
| 18 | 四分位範囲      | 5.00  | 7.00  | =J16-J14                |
| 19 | 四分位範囲/1.35 | 3.70  | 5.19  | =J18/1.35               |
| 20 | ひずみ        | 1.57  | 0.37  | =SKEW(J4:J11)           |
| 21 | とがり        | 2.51  | -0.28 | =KURT(J4:J11)           |
| 22 | 標準偏差       | 5.55  | 5.55  | =STDEV(J4:J11)          |

## ●Excel による計算

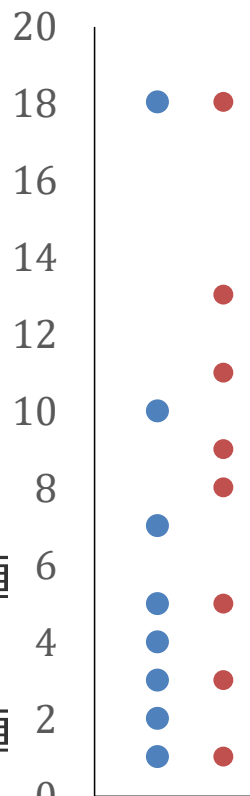
Excel関数

- =QUARTILE( J4:J11, 0) : 最小値
- =QUARTILE( J4:J11, 1) : 下側四分位値
- =QUARTILE( J4:J11, 2) : 中央値
- =QUARTILE( J4:J11, 3) : 上側四分位値
- =QUARTILE( J4:J11, 4) : 最大値



- =QUARTILE.INC( J4:J11, 0) : 最小値
- =QUARTILE.INC( J4:J11, 1) : 下側四分位値
- =QUARTILE.INC( J4:J11, 2) : 中央値
- =QUARTILE.INC( J4:J11, 3) : 上側四分位値
- =QUARTILE.INC( J4:J11, 4) : 最大値

2種類のデータ



表示 2.1.3 四分位値、ひずみ、とがりの計算

|    | I          | J     | K     | L                       |
|----|------------|-------|-------|-------------------------|
| 2  |            | i     | x1    | x2                      |
| 4  |            | 1     | 1     |                         |
| 5  |            | 2     | 2     |                         |
| 6  |            | 3     | 3     |                         |
| 7  |            | 4     | 4     |                         |
| 8  |            | 5     | 5     |                         |
| 9  |            | 6     | 7     |                         |
| 10 |            | 7     | 10    |                         |
| 11 |            | 8     | 18    |                         |
| 12 | n          | 8     | 8     | =COUNT(J4:J11)          |
| 13 | 最小値        | 1.00  | 1.00  | =QUARTILE(J\$4:J\$11,0) |
| 14 | 下側四分位値     | 2.75  | 4.50  | =QUARTILE(J\$4:J\$11,1) |
| 15 | 中央値        | 4.50  | 8.50  | =QUARTILE(J\$4:J\$11,2) |
| 16 | 上側四分位値     | 7.75  | 11.50 | =QUARTILE(J\$4:J\$11,3) |
| 17 | 最大値        | 18.00 | 18.00 | =QUARTILE(J\$4:J\$11,4) |
| 18 | 四分位範囲      | 5.00  | 7.00  | =J16-J14                |
| 19 | 四分位範囲/1.35 | 3.70  | 5.19  | =J18/1.35               |
| 20 | ひずみ        | 1.57  | 0.37  | =SKEW(J4:J11)           |
| 21 | とがり        | 2.51  | -0.28 | =KURT(J4:J11)           |
| 22 | 標準偏差       | 5.55  | 5.55  | =STDEV(J4:J11)          |

## ●Excel による計算

四分位範囲/1.35 → 標準偏差の近似推定値

標準偏差と四分位範囲/1.35

x1 : 標準偏差 > 四分位範囲/1.35

5.55    3.70

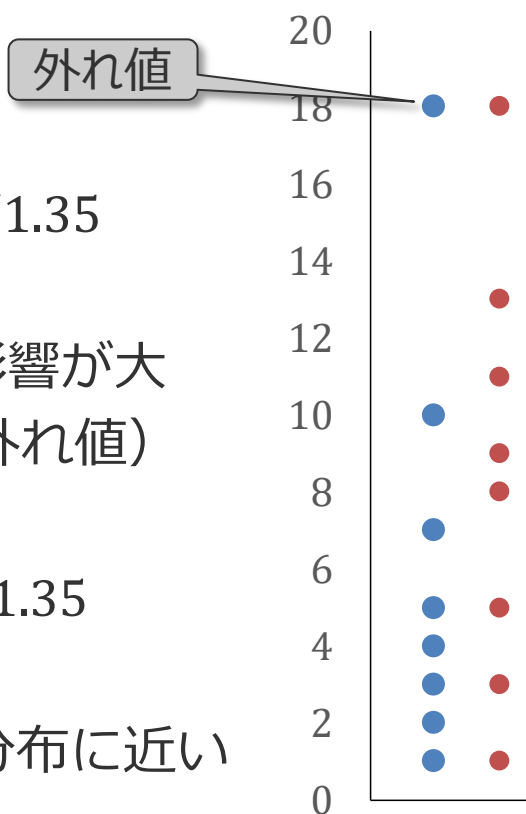
標準偏差は  $x_8 = 18$  の影響が大

5.55 は過大評価 (外れ値)

x2 : 標準偏差 = 四分位範囲/1.35

5.55    5.19

外れた値がない、正規分布に近い



表示 2.1.3 四分位値、ひずみ、とがりの計算

|    | I          | J     | K     | L                       |
|----|------------|-------|-------|-------------------------|
| 3  | i          | x1    | x2    | Excel 関数                |
| 4  | 1          | 1     | 1     |                         |
| 5  | 2          | 2     | 3     |                         |
| 6  | 3          | 3     | 5     |                         |
| 7  | 4          | 4     | 8     |                         |
| 8  | 5          | 5     | 9     |                         |
| 9  | 6          | 7     | 11    |                         |
| 10 | 7          | 10    | 13    |                         |
| 11 | 8          | 18    | 18    |                         |
| 12 | n          | 8     | 8     | =COUNT(J4:J11)          |
| 13 | 最小値        | 1.00  | 1.00  | =QUARTILE(J\$4:J\$11,0) |
| 14 | 下側四分位値     | 2.75  | 4.50  | =QUARTILE(J\$4:J\$11,1) |
| 15 | 中央値        | 4.50  | 8.50  | =QUARTILE(J\$4:J\$11,2) |
| 16 | 上側四分位値     | 7.75  | 11.50 | =QUARTILE(J\$4:J\$11,3) |
| 17 | 最大値        | 18.00 | 18.00 | =QUARTILE(J\$4:J\$11,4) |
| 18 | 四分位範囲      | 5.00  | 7.00  | =J16-J14                |
| 19 | 四分位範囲/1.35 | 3.70  | 5.19  | =J18/1.35               |
| 20 | ひずみ        | 1.57  | 0.37  | =SKEW(J4:J11)           |
| 21 | とがり        | 2.51  | -0.28 | =KURT(J4:J11)           |
| 22 | 標準偏差       | 5.55  | 5.55  | =STDEV(J4:J11)          |



## (7) ひずみ と とがり

歪度と尖度



# ひずみ（歪度） と とがり（尖度）

## ●ひずみととがりの利用目的

正規分布からの外れの度合いを表す

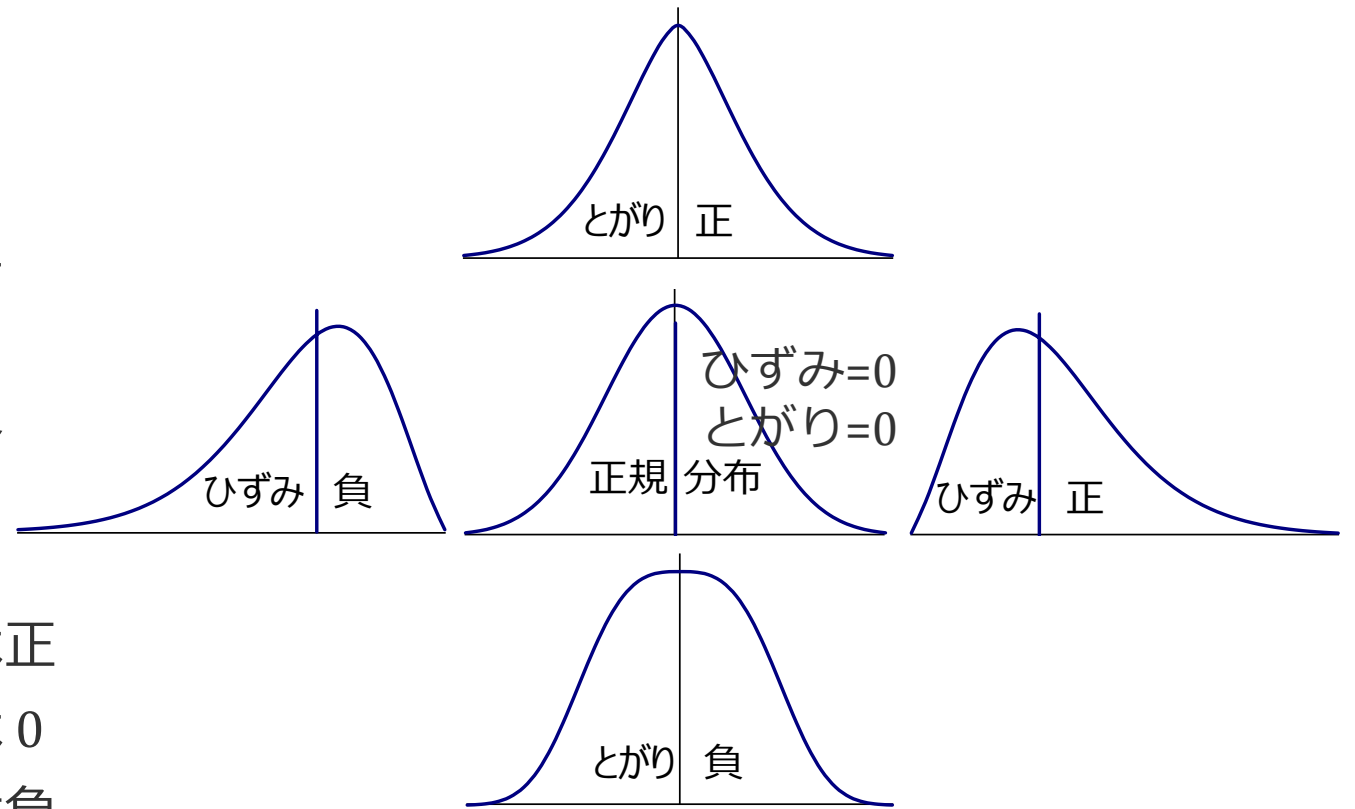
表示 2.1.4 ひずみ と とがり

### ひずみ（歪度）

- 大きい方に裾を引く → ひずみは正
- 分布が左右に対称 → ひずみは0
- 小さい方に裾を引く → ひずみは負

### とがり（尖度）

- 中央が尖って裾が長い → とがりは正
- 分布が正規分布 → とがりは0
- 中央が丸まって裾が短い → とがりは負





# ひずみ（歪度） と とがり（尖度）

- ひずみ（歪度、 $b_1$ ）、とがり（尖度、 $b_2$ ）の算出

計算式

$$u_i = \frac{x_i - \bar{x}}{s} \quad b_1 = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n u_i^3 \approx \frac{1}{n} \sum_{i=1}^n u_i^3 \quad (2.1.8)$$

$$b_2 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n u_i^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \approx \frac{1}{n} \sum_{i=1}^n u_i^4 - 3 \quad (2.1.9)$$

Excel 関数

ひずみ：=SKEW(データの集まり)

とがり：=KURT(データの集まり)

見方

母集団が正規分布でも、サンプルのひずみ、とがりは0からかなり振れる

正規分布から外れているか否かは $\pm 1.5$ を目安（ $n$ が数十程度以下のとき）

この範囲を超えた場合、分布が全体として正規分布から外れている、少数個の外れ値がある

⇒ ヒストグラムや箱ひげ図で確認（後で説明）

# ひずみ（歪度） と とがり（尖度）

## ●ひずみ（歪度）、とがり（尖度）

x1 : 外れ値がある

→ ひずみ、とがりとも 1.5 以上

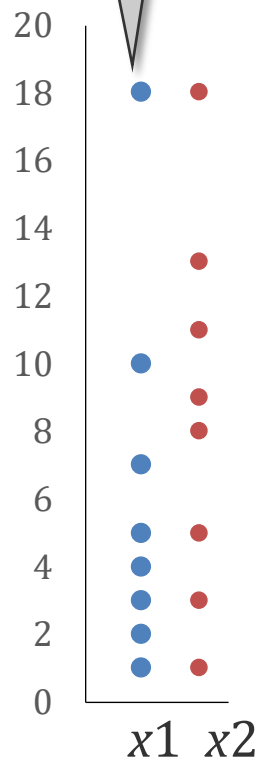
x2 : 外れ値はない、全体にばらつく

→ ひずみ、とがりとも 0 に近い

表示 2.1.3

|    | I          | J     | K     | L                       |
|----|------------|-------|-------|-------------------------|
| 3  | i          | x1    | x2    | Excel 関数                |
| 4  | 1          | 1     | 1     |                         |
| 5  | 2          | 2     | 3     |                         |
| 6  | 3          | 3     | 5     |                         |
| 7  | 4          | 4     | 8     |                         |
| 8  | 5          | 5     | 9     |                         |
| 9  | 6          | 7     | 11    |                         |
| 10 | 7          | 10    | 13    |                         |
| 11 | 8          | 18    | 18    |                         |
| 12 | n          | 8     | 8     | =COUNT(J4:J11)          |
| 13 | 最小値        | 1.00  | 1.00  | =QUARTILE(J\$4:J\$11,0) |
| 14 | 下側四分位値     | 2.75  | 4.50  | =QUARTILE(J\$4:J\$11,1) |
| 15 | 中央値        | 4.50  | 8.50  | =QUARTILE(J\$4:J\$11,2) |
| 16 | 上側四分位値     | 7.75  | 11.50 | =QUARTILE(J\$4:J\$11,3) |
| 17 | 最大値        | 18.00 | 18.00 | =QUARTILE(J\$4:J\$11,4) |
| 18 | 四分位範囲      | 5.00  | 7.00  | =J16-J14                |
| 19 | 四分位範囲/1.35 | 3.70  | 5.19  | =J18/1.35               |
| 20 | ひずみ        | 1.57  | 0.37  | =SKEW(J4:J11)           |
| 21 | とがり        | 2.51  | -0.28 | =KURT(J4:J11)           |
| 22 | 標準偏差       | 5.55  | 5.55  | =STDEV(J4:J11)          |

外れ値



# ひずみ（歪度） ととがり（尖度）

## ●外れ値の影響

(演習 2.1.1)

表示2.1.3 のJ列 x1 のデータから、最後の18を除いて再計算

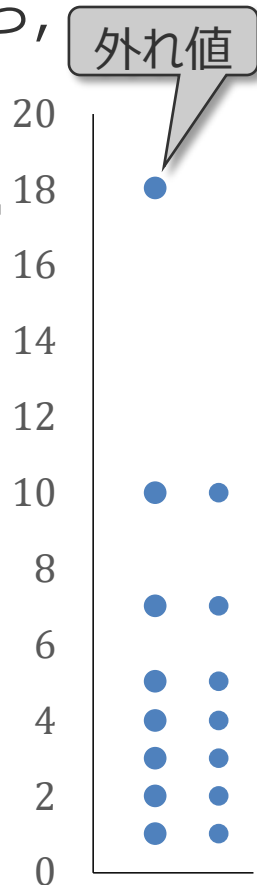
ひずみ、とがりは1以下に変化  
中央部の尖り具合よりも、裾の長さが影響する

標準偏差の変化が大きい

5.55 → 3.10

四分位範囲の変化は小さく、頑健性がある

3.70 → 2.59



表示2.1.3

|    | I          | J     |
|----|------------|-------|
| 3  | i          | x1    |
| 4  | 1          | 1     |
| ⋮  | ⋮          | ⋮     |
| 10 | 7          | 10    |
| 11 | 8          | 18    |
| 12 | n          | 8     |
| 13 | 最小値        | 1.00  |
| 14 | 下側四分位値     | 2.75  |
| 15 | 中央値        | 4.50  |
| 16 | 上側四分位値     | 7.75  |
| 17 | 最大値        | 18.00 |
| 18 | 四分位範囲      | 5.00  |
| 19 | 四分位範囲/1.35 | 3.70  |
| 20 | ひずみ        | 1.57  |
| 21 | とがり        | 2.51  |
| 22 | 標準偏差       | 5.55  |

数値の前に「\*」を添付  
データを除く

|    | I          | J     |
|----|------------|-------|
| 3  | i          | x1    |
| 4  | 1          | 1     |
| ⋮  | ⋮          | ⋮     |
| 10 | 7          | 10    |
| 11 | 8          | *18   |
| 12 | n          | 7     |
| 13 | 最小値        | 1.00  |
| 14 | 下側四分位値     | 2.50  |
| 15 | 中央値        | 4.00  |
| 16 | 上側四分位値     | 6.00  |
| 17 | 最大値        | 10.00 |
| 18 | 四分位範囲      | 3.50  |
| 19 | 四分位範囲/1.35 | 2.59  |
| 20 | ひずみ        | 0.84  |
| 21 | とがり        | 0.23  |
| 22 | 標準偏差       | 3.10  |



## (8) 頑健性

## ●頑健性（ロバストネス Robustness）

### 統計量と頑健性

平均値、標準偏差よりも、中央値、四分位値、四分位範囲／1.35の方が頑健性が高い

### 平均値

データが正規分布に従うとき、母平均  $\mu$  の推定値は平均値が最良（数理統計学的に証明）  
正規分布に従わない場合、外れ値・異質な値が混入する場合、それに対して平均値は敏感

1つの観測値  $x$  が  $d$  増加すると平均値は  $d/n$  増加  
⇒  $d$  が大きいと平均値への影響は大

### 中央値

中央値は頑健性がある  
中央値はデータの情報を十分に利用していないので、  
正規性のあるデータでは効率が悪い

|     | x1        | +d<br>→   | x2        |
|-----|-----------|-----------|-----------|
|     | 10        |           | 10        |
|     | 20        |           | 20        |
|     | 30        |           | 30        |
|     | 40        |           | 40        |
|     | <b>50</b> | <b>40</b> | <b>90</b> |
| 平均値 | 30        |           | 38        |
| 中央値 | 30        |           | 30        |

$$\begin{aligned} & 30 + \frac{d}{n} \\ &= 30 + \frac{40}{5} \\ &= 38 \end{aligned}$$

## ● トリム平均 (Trimmed Mean)

刈り込み平均値、調整平均ともいう

データを昇順または降順に並べ替えた後、

上位と下位から一定の個数または一定の割合のデータを除外、残ったデータの平均を計算

使用例

オリンピックのフィギュアスケートの採点では、最高点と最低点を除外した平均値を使用  
給与水準の比較や品質管理などでも利用、外れ値の対応策

Excel 関数

= TRIMMEAN (データ範囲, 刈り込む割合)

刈り込む割合は両側の合計割合、0.20 とすると上位と下位からそれぞれ 10% が除かれる

端数は切り捨て、刈り込む割合  $\times n/2$  が 0.9 の場合は除外数 0、1.9 の場合は除外数 1 個ずつ

10% トリム平均 (上位、下位でそれぞれ 10% 除外)、25% トリム平均 (中央平均)

トリム (trim) : 切り込む、切り取る  
トリミング (trimming) も同じような意味

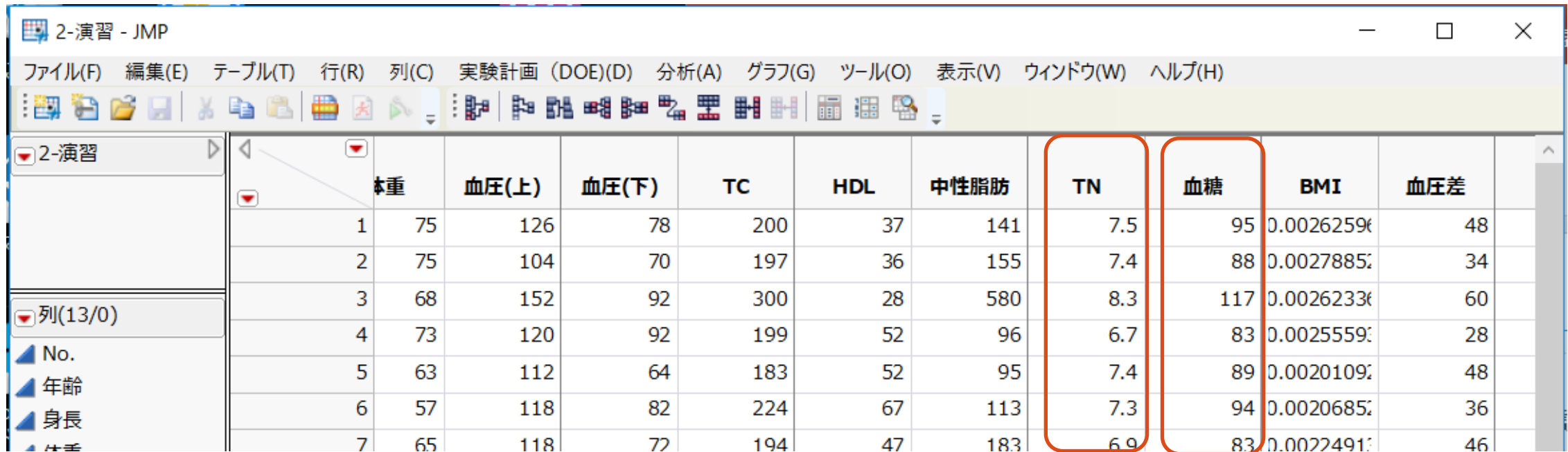


## (9) JMPによる解析

## ●データ

JMP ファイル「2-演習.jmp」の読み込み

40人の男子のNo.、年齢、身長、体重、血圧、血液の生化学検査値などが記録  
「NT（総窒素）」と「血糖」の基本統計量を求める



The screenshot shows the JMP software interface with a data table. The table has columns for '体重' (Weight), '血圧(上)' (Systolic Blood Pressure), '血圧(下)' (Diastolic Blood Pressure), 'TC' (Total Cholesterol), 'HDL' (High-Density Lipoprotein), '中性脂肪' (Triglycerides), 'TN' (Total Nitrogen), '血糖' (Blood Glucose), 'BMI', and '血圧差' (Blood Pressure Difference). The 'TN' and '血糖' columns are highlighted with red boxes. The data rows are numbered 1 through 7.

|   | 体重 | 血圧(上) | 血圧(下) | TC  | HDL | 中性脂肪 | TN  | 血糖  | BMI        | 血圧差 |
|---|----|-------|-------|-----|-----|------|-----|-----|------------|-----|
| 1 | 75 | 126   | 78    | 200 | 37  | 141  | 7.5 | 95  | 0.00262596 | 48  |
| 2 | 75 | 104   | 70    | 197 | 36  | 155  | 7.4 | 88  | 0.00278852 | 34  |
| 3 | 68 | 152   | 92    | 300 | 28  | 580  | 8.3 | 117 | 0.00262336 | 60  |
| 4 | 73 | 120   | 92    | 199 | 52  | 96   | 6.7 | 83  | 0.00255596 | 28  |
| 5 | 63 | 112   | 64    | 183 | 52  | 95   | 7.4 | 89  | 0.00201092 | 48  |
| 6 | 57 | 118   | 82    | 224 | 67  | 113  | 7.3 | 94  | 0.00206852 | 36  |
| 7 | 65 | 118   | 77    | 194 | 47  | 183  | 6.9 | 83  | 0.00274912 | 46  |

# JMPによる解析 [一変量の分布]

## ●データ「TN」「血糖」の基本統計量

表示 2.1.5 JMPのデータ表と変数の指定画面

The screenshot shows the JMP interface. At the top, the 'Analyze' menu is open, with '一変量の分布' (Univariate) selected. Below the menu, a data table is visible with columns 'TN', '血糖', and 'BI'. The data rows are:

|   | TN  | 血糖  | BI    |
|---|-----|-----|-------|
| 1 | 7.5 | 95  | 0.002 |
| 2 | 7.4 | 88  | 0.002 |
| 3 | 8.3 | 117 | 0.002 |
| 4 | 6.7 | 83  | 0.002 |
| 5 | 7.4 | 89  | 0.002 |

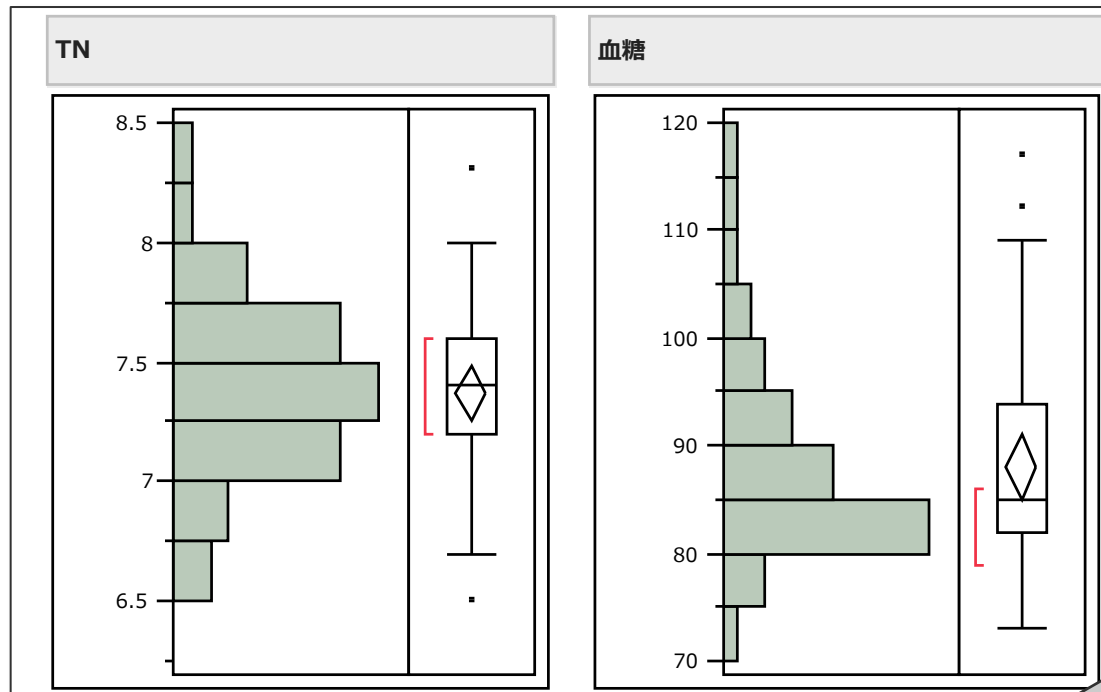
「TN」「血糖」のデータを確認

ブルーの三角は連続尺度

[分析] > [一変量の分析] > [列の選択] : TNと血糖を選択  
[選択した列に役割を割り当てる] > [Y]

The screenshot shows the 'Univariate' dialog box. The '列の選択' (Select Columns) list contains 'TN' and '血糖', both with blue triangles next to them. A callout points to these triangles with the text 'ブルーの三角は連続尺度'. The 'Y, 列' (Y, Column) field contains 'TN' and '血糖'. The 'アクション' (Action) buttons include 'OK', 'キャンセル', '削除', '前回の設定', and 'ヘルプ'. A callout points to the 'OK' button.

## ●グラフ表示と四分位値

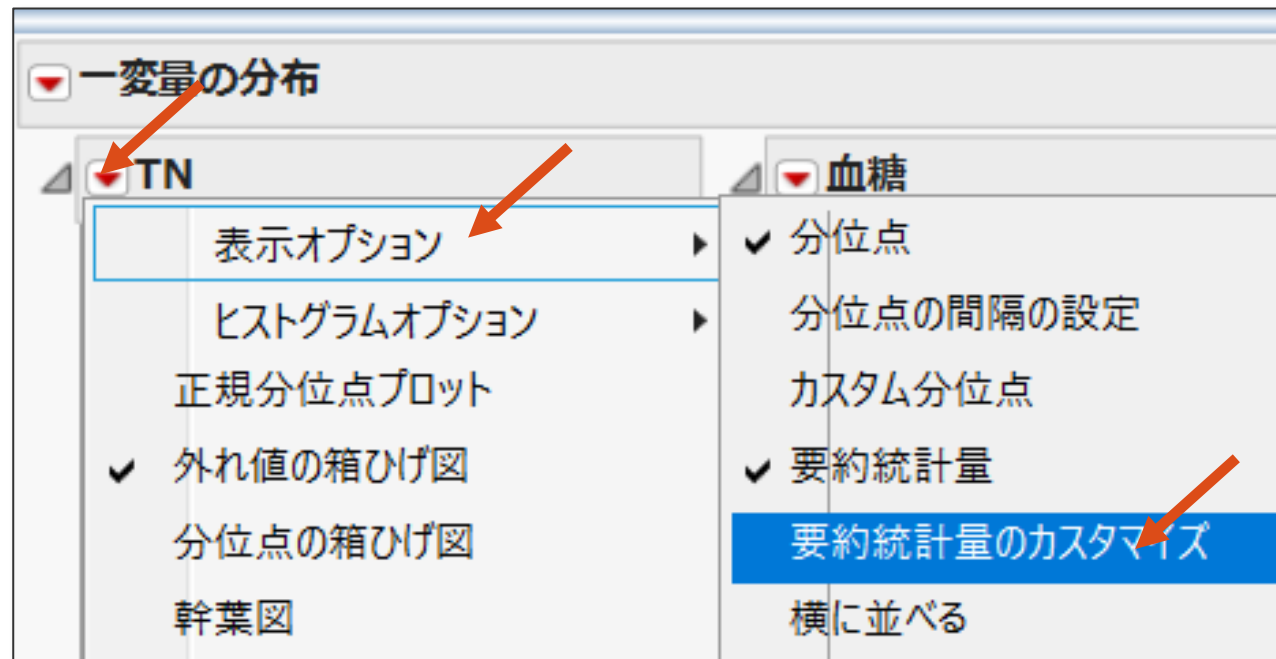


グラフは [§2.2](#) で説明

四分位値

| TN     |      |        | 血糖     |      |         |
|--------|------|--------|--------|------|---------|
| 分位点    |      |        | 分位点    |      |         |
| 100.0% | 最大値  | 8.3    | 100.0% | 最大値  | 117     |
| 99.5%  |      | 8.3    | 99.5%  |      | 117     |
| 97.5%  |      | 8.2925 | 97.5%  |      | 116.875 |
| 90.0%  |      | 7.89   | 90.0%  |      | 102.7   |
| 75.0%  | 四分位点 | 7.6    | 75.0%  | 四分位点 | 93.75   |
| 50.0%  | 中央値  | 7.4    | 50.0%  | 中央値  | 85      |
| 25.0%  | 四分位点 | 7.2    | 25.0%  | 四分位点 | 82      |
| 10.0%  |      | 6.9    | 10.0%  |      | 79.1    |
| 2.5%   |      | 6.505  | 2.5%   |      | 73.15   |
| 0.5%   |      | 6.5    | 0.5%   |      | 73      |
| 0.0%   | 最小値  | 6.5    | 0.0%   | 最小値  | 73      |

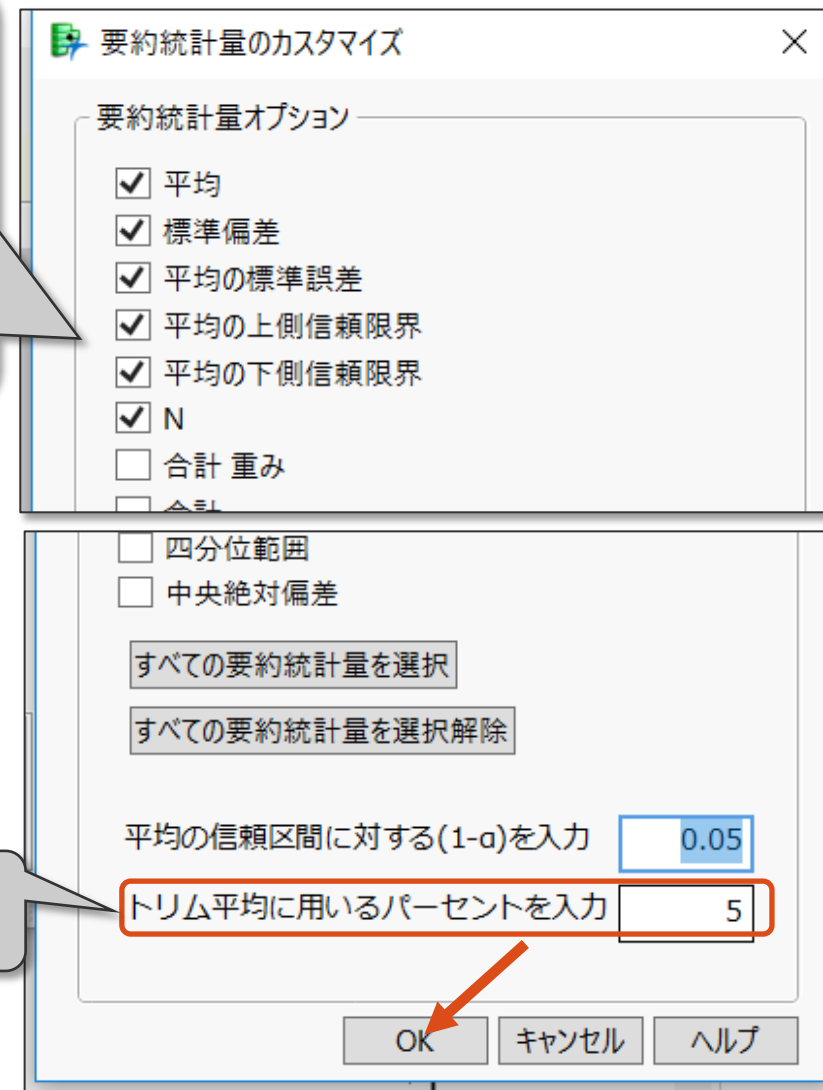
## ●要約統計量のカスタマイズ



変数名の▼> [表示オプション]> [要約統計量のカスタマイズ]

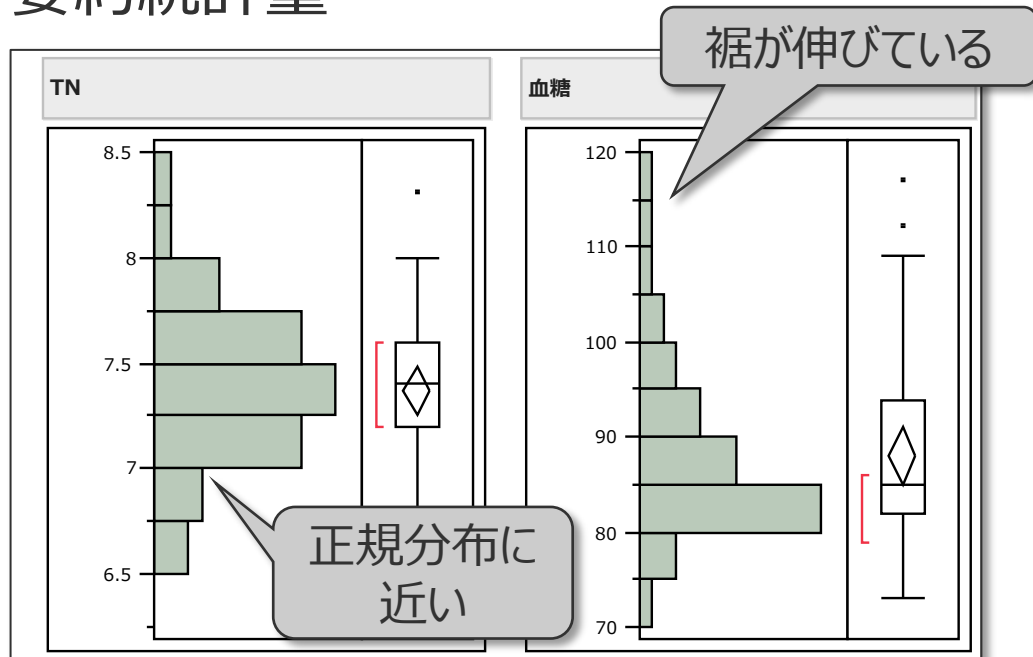
▼を押す前から Ctrl キーを押しながら作業すると表示中の全ての変数に同時指定できる

分散  
歪度  
尖度  
変動係数  
などに  
チェック  
を入れる



片側の割合  
を指定

## ●要約統計量



TNの平均 7.37 は中央値 7.4 と一致  
 血糖の平均 88.1 は中央値 85 よりも大きい  
 変動係数は、TN < 血糖 (%表示)  
 単位が異なるので、変動係数で比較

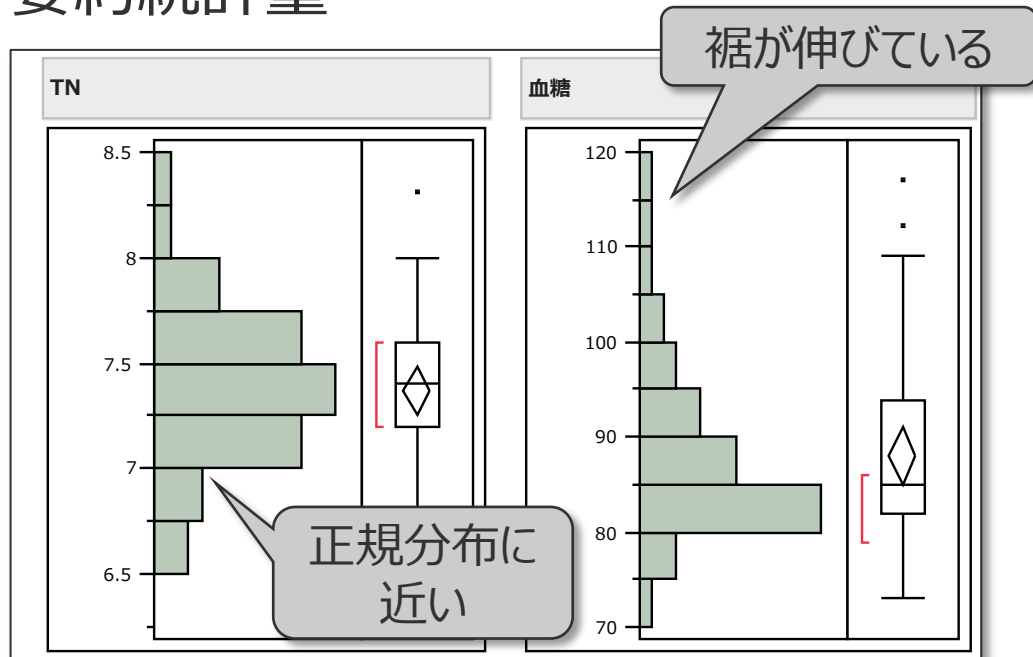
TN

| ▼ 要約統計量  |           |
|----------|-----------|
| 平均       | 7.3725    |
| 標準偏差     | 0.3573316 |
| 平均の標準誤差  | 0.0564991 |
| 平均の上側95% | 7.4867802 |
| 平均の下側95% | 7.2582198 |
| N        | 40        |
| 合計       | 294.9     |
| 分散       | 0.1276859 |
| 歪度       | 0.0950405 |
| 尖度       | 0.6325858 |
| 変動係数     | 4.8468177 |
| 中央値      | 7.4       |
| 5% トリム平均 | 7.3722222 |
| 四分位範囲    | 0.4       |

血糖

| ▼ 要約統計量  |           |
|----------|-----------|
| 平均       | 88.075    |
| 標準偏差     | 9.5283316 |
| 平均の標準誤差  | 1.5065615 |
| 平均の上側95% | 91.122308 |
| 平均の下側95% | 85.027692 |
| N        | 40        |
| 合計       | 3523      |
| 分散       | 90.789103 |
| 歪度       | 1.3815043 |
| 尖度       | 1.8324036 |
| 変動係数     | 10.818429 |
| 中央値      | 85        |
| 5% トリム平均 | 87.277778 |
| 四分位範囲    | 11.75     |

## ●要約統計量



標準偏差の近似推定値： $s^* = \text{四分位範囲} / 1.35$

TN :  $s^* = 0.4 / 1.35 = 0.296$

血糖 :  $s^* = 11.75 / 1.35 = 8.70$

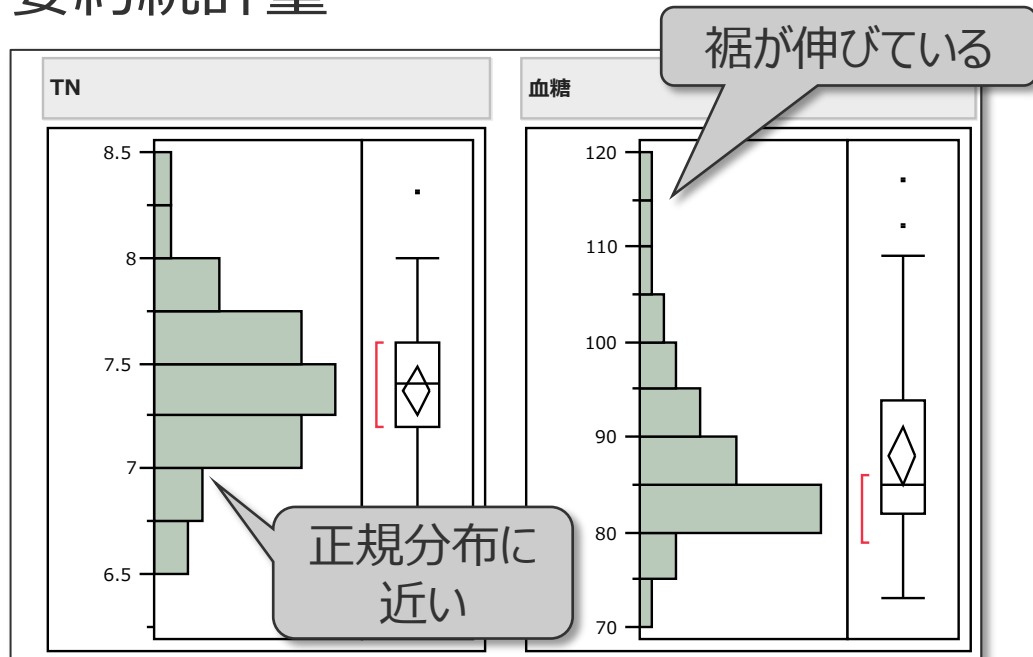
TN

| ▼ 要約統計量  |           |
|----------|-----------|
| 平均       | 7.3725    |
| 標準偏差     | 0.3573316 |
| 平均の標準誤差  | 0.0564991 |
| 平均の上側95% | 7.4867802 |
| 平均の下側95% | 7.2582198 |
| N        | 40        |
| 合計       | 294.9     |
| 分散       | 0.1276859 |
| 歪度       | 0.0950405 |
| 尖度       | 0.6325858 |
| 変動係数     | 4.8468177 |
| 中央値      | 7.4       |
| 5% トリム平均 | 7.3722222 |
| 四分位範囲    | 0.4       |

血糖

| ▼ 要約統計量  |           |
|----------|-----------|
| 平均       | 88.075    |
| 標準偏差     | 9.5283316 |
| 平均の標準誤差  | 1.5065615 |
| 平均の上側95% | 91.122308 |
| 平均の下側95% | 85.027692 |
| N        | 40        |
| 合計       | 3523      |
| 分散       | 90.789103 |
| 歪度       | 1.3815043 |
| 尖度       | 1.8324036 |
| 変動係数     | 10.818429 |
| 中央値      | 85        |
| 5% トリム平均 | 87.277778 |
| 四分位範囲    | 11.75     |

## ●要約統計量



TN

血糖

| ▼ 要約統計量  |           | ▼ 要約統計量  |           |
|----------|-----------|----------|-----------|
| 平均       | 7.3725    | 平均       | 88.075    |
| 標準偏差     | 0.3573316 | 標準偏差     | 9.5283316 |
| 平均の標準誤差  | 0.0564991 | 平均の標準誤差  | 1.5065615 |
| 平均の上側95% | 7.4867802 | 平均の上側95% | 91.122308 |
| 平均の下側95% | 7.2582198 | 平均の下側95% | 85.027692 |
| N        | 40        | N        | 40        |
| 合計       | 294.9     | 合計       | 3523      |
| 分散       | 0.1276859 | 分散       | 90.789103 |
| 歪度       | 0.0950405 | 歪度       | 1.3815043 |
| 尖度       | 0.6325858 | 尖度       | 1.8324036 |
| 変動係数     | 4.8468177 | 変動係数     | 10.818429 |
| 中央値      | 7.4       | 中央値      | 85        |
| 5% トリム平均 | 7.3722222 | 5% トリム平均 | 87.277778 |
| 四分位範囲    | 0.4       | 四分位範囲    | 11.75     |

TN は、尖度と歪度ともに目安の 1.5 以内  
 血糖は、尖度が 1.5 を超え、歪度も 1.5 に近い

# JMPによる解析 [一変量の分布]

- 四分位値のJMPとExcelによる違い  
サンプルサイズが大きい場合はほぼ一致

表示 2.1.3 Excel の  
計算結果 (改変)

| No.        | TN    | 血糖    |
|------------|-------|-------|
| 1          | 7.5   | 95    |
| 2          | 7.4   | 88    |
| ⋮          | ⋮     | ⋮     |
| ⋮          | ⋮     | ⋮     |
| ⋮          | ⋮     | ⋮     |
| 39         | 7.7   | 80    |
| 40         | 7.8   | 112   |
| 平均         | 7.37  | 88.08 |
| 標準偏差       | 0.36  | 9.53  |
| 上位四分位値     | 7.60  | 93.25 |
| 中央値        | 7.40  | 85.00 |
| 下位四分位値     | 7.20  | 82.00 |
| 四分位範囲      | 0.40  | 11.25 |
| 四分位範囲/1.35 | 0.296 | 8.333 |

| TN      |      |        | 血糖      |      |         |
|---------|------|--------|---------|------|---------|
| 分位点     |      |        | 分位点     |      |         |
| 100.0%  | 最大值  | 8.3    | 100.0%  | 最大值  | 117     |
| 99.5%   |      | 8.3    | 99.5%   |      | 117     |
| 97.5%   |      | 8.2925 | 97.5%   |      | 116.875 |
| 90.0%   |      | 7.89   | 90.0%   |      | 102.7   |
| 75.0%   | 四分位点 | 7.6    | 75.0%   | 四分位点 | 93.75   |
| 50.0%   | 中央値  | 7.4    | 50.0%   | 中央値  | 85      |
| 25.0%   | 四分位点 | 7.2    | 25.0%   | 四分位点 | 82      |
| 10.0%   |      | 6.9    | 10.0%   |      | 79.1    |
| 2.5%    |      | 6.505  | 2.5%    |      | 73.15   |
| 0.5%    |      | 6.5    | 0.5%    |      | 73      |
| 0.0%    | 最小値  | 6.5    | 0.0%    | 最小値  | 73      |
| ▼ 要約統計量 |      |        | ▼ 要約統計量 |      |         |
| 四分位範囲   |      | 0.4    | 四分位範囲   |      | 11.75   |

JMPよりやや大きめ

# JMPによる解析 [一変量の分布]

## ●データ「身長」「中性脂肪」基本統計量

JMP ファイル「2-演習.jmp」のデータの中から  
「身長」「中性脂肪」の基本統計量を得る

(演習 2.1.2 p.68)

[分析] > [一変量の分析] >

[列の選択] : 身長と中性脂肪を選択

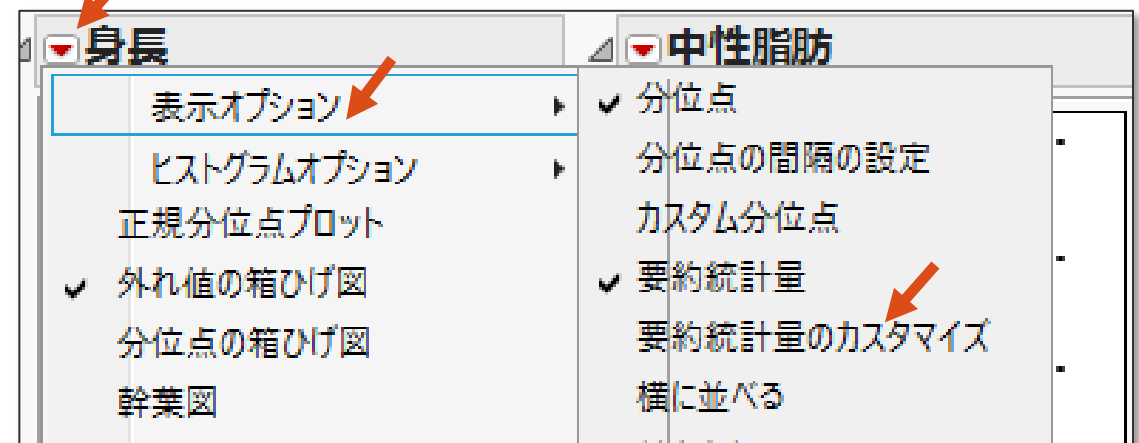
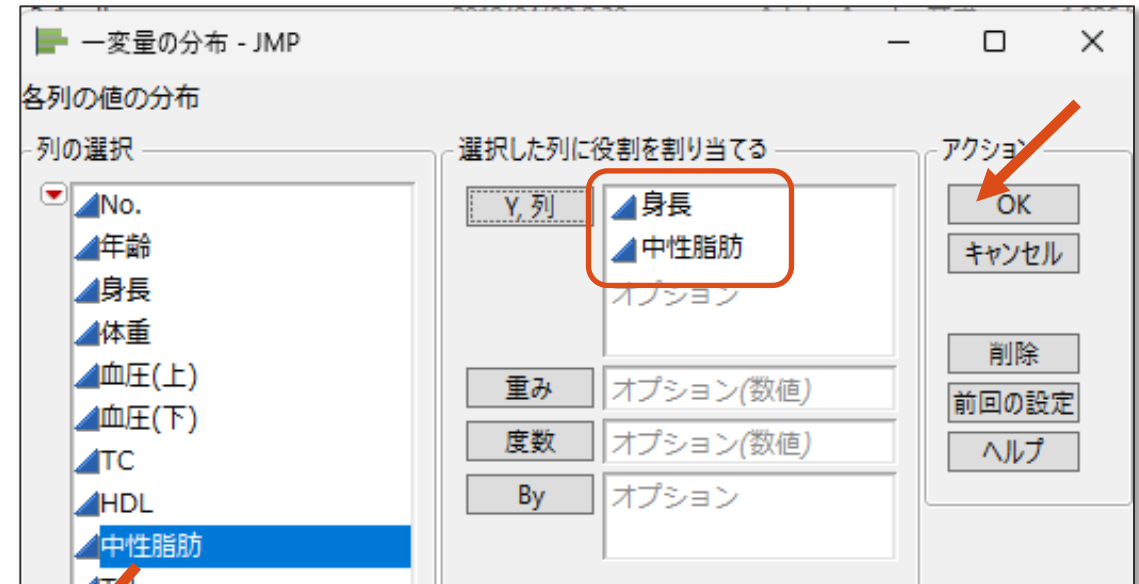
[選択した列に役割を割り当てる] > [Y]

変数名の▼>

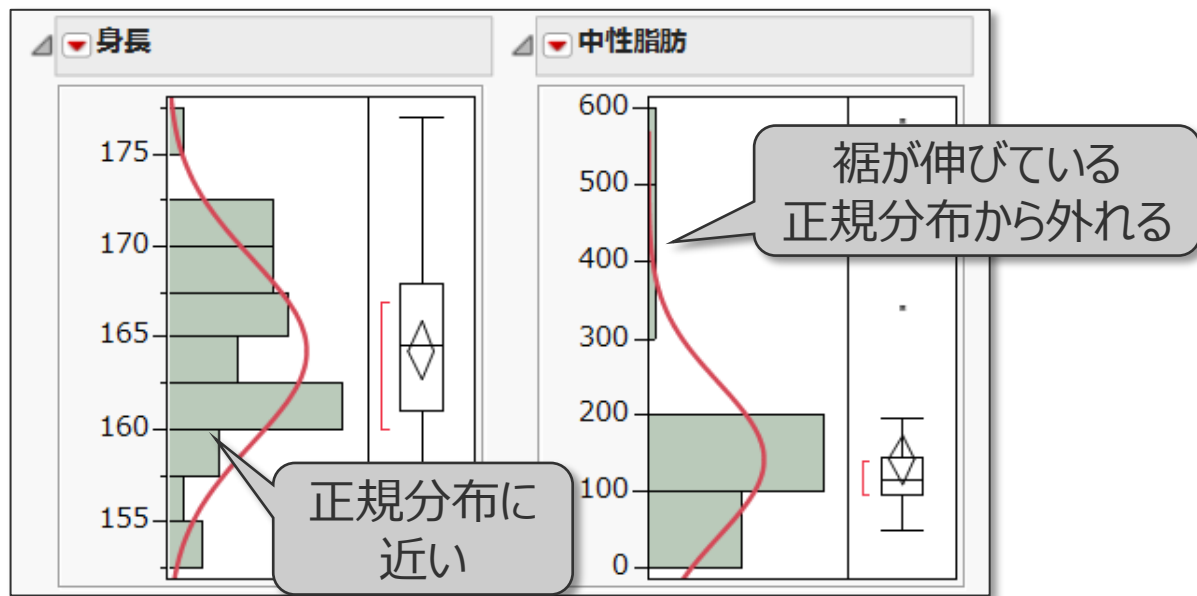
[表示オプション] >

[要約統計量のカスタマイズ]

▼を押す前から Ctrl キーを押しながら作業  
表示中の全ての変数に同時指定できる



## ●データ「身長」「中性脂肪」基本統計量



### 身長

| 分位点    |      |        |
|--------|------|--------|
| 100.0% | 最大値  | 177    |
| 99.5%  |      | 177    |
| 97.5%  |      | 176.85 |
| 90.0%  |      | 170    |
| 75.0%  | 四分位点 | 168    |
| 50.0%  | 中央値  | 164.5  |
| 25.0%  | 四分位点 | 161    |
| 10.0%  |      | 158    |
| 2.5%   |      | 154    |
| 0.5%   |      | 154    |
| 0.0%   | 最小値  | 154    |

### 要約統計量

|       |           |
|-------|-----------|
| 平均    | 164.325   |
| 標準偏差  | 4.999423  |
| 歪度    | -0.00708  |
| 尖度    | -0.063401 |
| 四分位範囲 | 7         |

### 中性脂肪

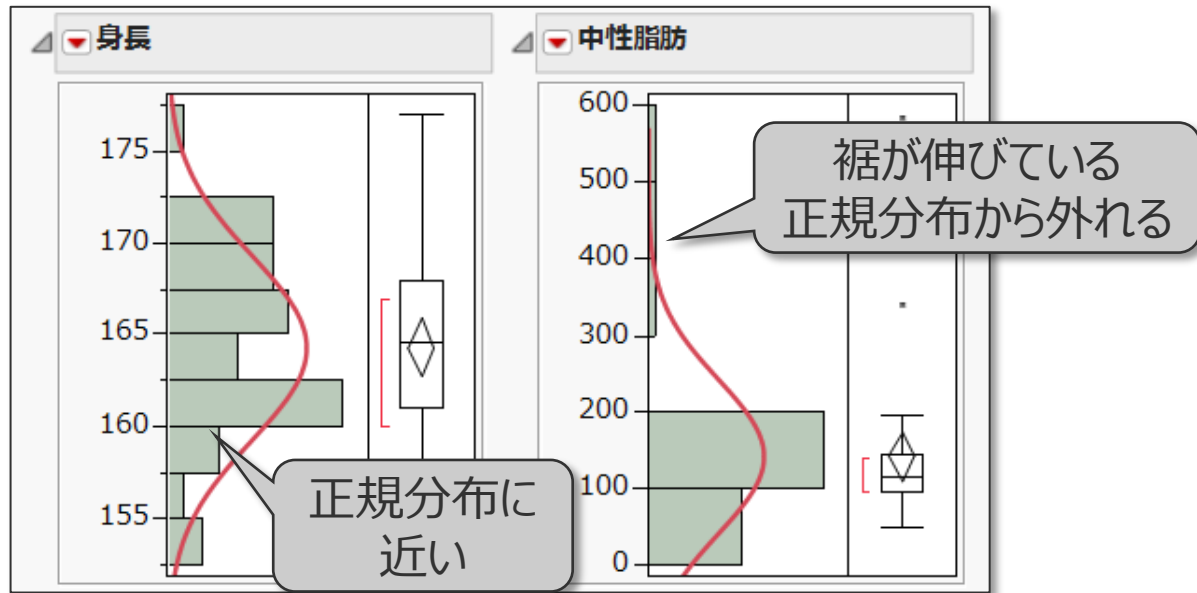
| 分位点    |      |       |
|--------|------|-------|
| 100.0% | 最大値  | 580   |
| 99.5%  |      | 580   |
| 97.5%  |      | 576.9 |
| 90.0%  |      | 194.7 |
| 75.0%  | 四分位点 | 144   |
| 50.0%  | 中央値  | 117.5 |
| 25.0%  | 四分位点 | 95.25 |
| 10.0%  |      | 74.5  |
| 2.5%   |      | 50.45 |
| 0.5%   |      | 50    |
| 0.0%   | 最小値  | 50    |

### 要約統計量

|       |           |
|-------|-----------|
| 平均    | 142.5     |
| 標準偏差  | 100.13581 |
| 歪度    | 3.1885992 |
| 尖度    | 11.026068 |
| 四分位範囲 | 48.75     |

身長は、尖度と歪度ともに目安の 1.5 以内、0 に近い  
 中性脂肪は、尖度と歪度ともに 1.5 を大きく超える

## ●データ「身長」「中性脂肪」基本統計量



| 身長       |           | 中性脂肪     |           |
|----------|-----------|----------|-----------|
| ▼ 要約統計量  |           | ▼ 要約統計量  |           |
| 平均       | 164.325   | 平均       | 142.5     |
| 標準偏差     | 4.999423  | 標準偏差     | 100.13581 |
| 平均の標準誤差  | 0.7904782 | 平均の標準誤差  | 15.832861 |
| 平均の上側95% | 165.92389 | 平均の上側95% | 174.52498 |
| 平均の下側95% | 162.72611 | 平均の下側95% | 110.47502 |
| N        | 40        | N        | 40        |
| 歪度       | -0.00708  | 歪度       | 3.1885992 |
| 尖度       | -0.063401 | 尖度       | 11.026068 |
| 変動係数     | 3.0423995 | 変動係数     | 70.270741 |
| 中央値      | 164.5     | 中央値      | 117.5     |
| 四分位範囲    | 7         | 四分位範囲    | 48.75     |

標準偏差の近似推定値： $s^* = \text{四分位範囲} / 1.35$

身長 :  $s^* = 7 / 1.35 = 5.19$  . . . . 標準偏差とほぼ一致

中性脂肪 :  $s^* = 48.75 / 1.35 = 36.11$  . . 標準偏差は大きく異なる、正規分布から外れるため

## ●データの特徴の記述

データの特徴を表す指標

中央値、四分位値、四分位範囲、四分位範囲/1.35

平均値、標準偏差、変動係数、ひずみ、とがり

計算過程で出てきた指標

残差、平方和、自由度、平均平方

これらの値の持つ意味、違い、関係を十分に理解する

Excel 関数の理解



- 作成 片瀬雅彦
- 監修 松本一彦、長谷文雄
- 作成時期 2018年10月6日
- 改訂 2018年11月18日、2024年9月3日