教養科目／**B** 自然の摂理の探求

2026 年度 統計学（集中１期）

第３回

1．4．3．その他の代表値

a．最頻値（モード）

(I) データが離散型の場合

度数分布表において度数が最も大きい階級値

(II) データが連続型の場合

度数分布表において度数が最も大きい級中央値

Note

- 中心的位置を表す.
- 代表値が複数個存在する場合もある.

１．４．３．その他の代表値

b．最小値，最大値

データの中で最小の値，最大の値をそれぞれ**最小値，最大値**

c．範囲

$$\text{範囲}\ R = \text{最大値} - \text{最小値}$$

ばらつきを表す代表値.

１．４．３．その他の代表値

d．四分位数，四分位範囲

・**第１四分位数 (だいいちしぶんいすう)**Q_1

データを大きさの順に並べ，小さい方から 25% に位置する値.

・**第３四分位数 (だいさんしぶんいすう)**Q_3

データを大きさの順に並べ， 大きい方から 25% に位置する値.

・**データの個数が 100 の場合**

第 1 四分位数は小さい方から 25 番目の値

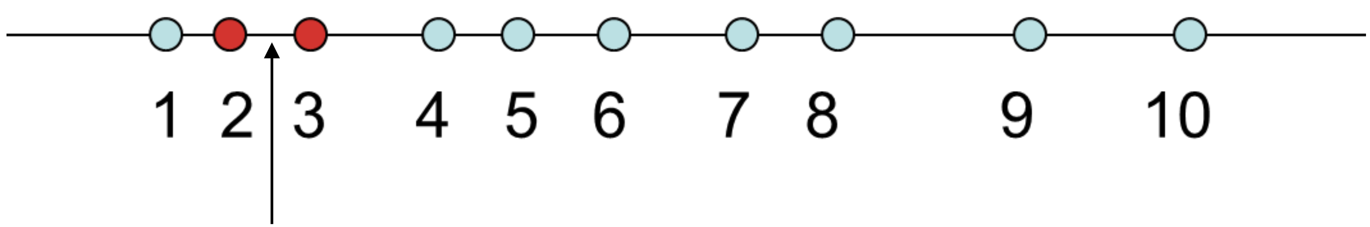
第 3 四分位数は大きい方から 25 番目の値

1. 4. 3. その他の代表値

小さい方から 25% に位置する値の候補が 2 つある場合，それらの平均を第 1 四分位数

・データの個数が 10 の場合

第 1 四分位数は小さい方から 2 番目の値と 3 番目の値の平均



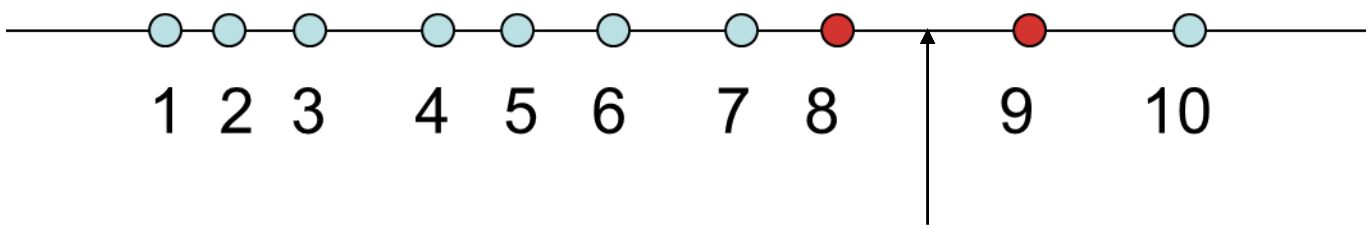
$$10 \times 0.25 = 2.5$$

1．4．3．その他の代表値

大きい方から 25% に位置する値の候補が 2 つある場合，
それらの平均を第 3 四分位数

・データの個数が 10 の場合

第 3 四分位数は大きい方から 2 番目の値と 3 番目の値の平均



$$10 \times 0.25 = 2.5$$

1．4．3．その他の代表値

・まとめ

$$
\text{第 1 四分位数 } Q_1 \quad = \quad \begin{cases} \text{小さい方から 25\% に位置する値} \\ \qquad\qquad\qquad\qquad \text{(データの個数が 4 の倍数のとき)}, \\ \dfrac{\text{小さい方から 25\% に位置する値の 2 つの候補の和}}{2} \\ \qquad\qquad\qquad\qquad \text{(データの個数が 4 の倍数でないとき)}, \end{cases}
$$

$$
\text{第 3 四分位数 } Q_3 \quad = \quad \begin{cases} \text{大きい方から 25\% に位置する値} \\ \qquad\qquad\qquad\qquad \text{(データの個数が 4 の倍数のとき)}, \\ \dfrac{\text{大きい方から 25\% に位置する値の 2 つの候補の和}}{2} \\ \qquad\qquad\qquad\qquad \text{(データの個数が 4 の倍数でないとき)}, \end{cases}
$$

$$
\text{四分位範囲 } R_Q \quad = \quad \text{第 3 四分位数 } Q_3 - \text{第 1 四分位数 } Q_1.
$$

1. 4. 3. その他の代表値

例 1.5 のその他の代表値は以下のとおりです.

- 最頻値

表 1.7 (p.8) から度数がもっとも大きい階級は 13.0 ～ 16.5 である.
この階級の級中央値は 14.75 です.

$$\text{最頻値} = 14.75$$

- 最小値，最大値（表 1.6 (p.7) を参照)

$$\underset{(\text{長野})}{\text{最小値} = 9}, \quad \underset{(\text{大阪})}{\text{最大値} = 30}.$$

- 範囲

最大値は 30, 最小値は 9 なので

$$\text{範囲} = 30 - 9 = 21.$$

9 11 11 11 12 12 12 12 13 13 13 14 14 14 14 14 15 15 15 15
16 16 16 16 16 17 17 17 17 17 18 18 18 19 19 19 19 19 20 21
21 21 21 23 23 23 30

- 四分位数，四分位範囲

$47 \times 0.25 = 11.7\cdots$ なので，第 1 四分位数は小さい方から 11 番目の値 13 と 12 番目の値 14 の平均，

$$\text{第 1 四分位数} = \frac{13+14}{2} = 13.5.$$

第 3 四分位数は大きい方から 11 番目の値 19 と 12 番目の値 19 の平均，

$$\text{第 3 四分位数} = \frac{19+19}{2} = 19.$$

四分位範囲は

$$\text{四分位範囲} = 19 - 13.5 = 5.5.$$

1．4．3．その他の代表値

偏差値（代表値ではない）

・偏差値はデータの値を比較しやすいように基準化した値である．

$$偏差値 = 50 + 10 \times \frac{データの値 - 標本平均}{標本標準偏差}$$

n 個のデータを

$$x_1, \quad x_2, \quad \cdots, \quad x_n$$

とし，x_i の偏差値 z_i は

$$z_i = 50 + 10 \times \frac{x_i - \bar{x}}{s_x}$$

となる．

・偏差値の性質

・偏差値の標本平均 $= 50$

・偏差値の標本標準偏差 $= 10$

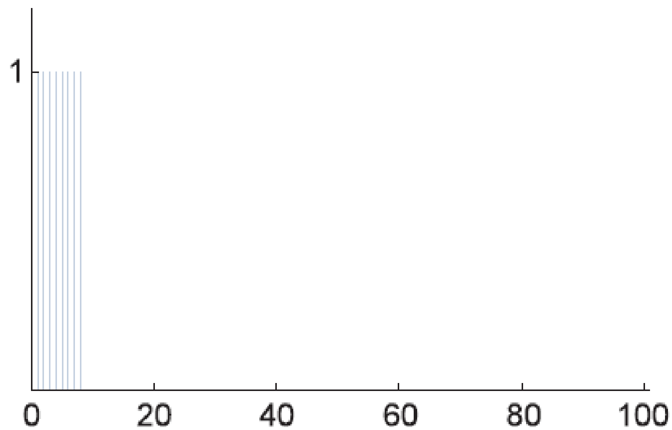
1．4．4．はずれ値

・2 組のデータ

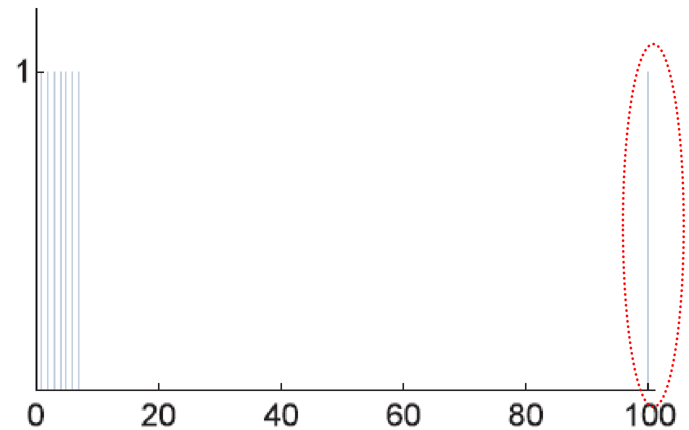
$A : 1, 2, 3, 4, 5, 6, 7, 8$ $\quad$ $B : 1, 2, 3, 4, 5, 6, 7, 100$

A と B の違いはデータの最後の値が 8 か 100 かだけ.

・A と B のヒストグラム



(i) はずれ値がない



(ii) はずれ値がある

B での値 100 は他の値より極端に大きいことがわかる.

1．4．4. はずれ値

・前スライドの例のように他のデータの値と比べて極端に大きい，または小さいデータの値を**はずれ値**という.

・ここで “極端に” という言葉が曖昧である.

比較的理解しやすい目安として，

$$\text{第 1 四分位数} - \text{四分位範囲} \times 1.5$$

より小さい，または

$$\text{第 3 四分位数} + \text{四分位範囲} \times 1.5$$

より大きいデータの値を**はずれ値**とする基準がよく使われる.

1. 4. 4. はずれ値

・例 1.5 ではずれ値があるかどうか考えてみる.

表 1.6　10 万人あたりの結核罹患者数（2010 年，厚生労働省）

北海道	12	栃 木	13	石 川	16	滋 賀	15	岡 山	15	佐 賀	21
青 森	14	群 馬	11	福 井	14	京 都	19	広 島	16	長 崎	23
岩 手	12	埼 玉	16	山 梨	15	大 阪	30	山 口	16	熊 本	17
宮 城	11	千 葉	17	長 野	9	兵 庫	21	徳 島	18	大 分	19
秋 田	14	東 京	23	岐 阜	20	奈 良	17	香 川	15	宮 崎	13
山 形	11	神奈川	17	静 岡	17	和歌山	21	愛 媛	19	鹿児島	21
福 島	12	新 潟	12	愛 知	23	鳥 取	14	高 知	18	沖 縄	19
茨 城	14	富 山	13	三 重	16	島 根	18	福 岡	19		

$9, 11, 11, 11, 12, 12, 12, 12, 13, 13, 13, 14$ ・・・　・・・

・・・　・・・ $19, 19, 19, 20, 21, 21, 21, 21, 23, 23, 23, 30$

$$\text{第 1 四分位数} - \text{四分位範囲} \times 1.5 = 13.5 - 5.5 \times 1.5 = 5.25,$$

$$\text{第 3 四分位数} + \text{四分位範囲} \times 1.5 = 19 + 5.5 \times 1.5 = 27.25$$

なので，大阪の結核罹患者数 30 ははずれ値.

1. 4. 4. はずれ値

- 2組のデータ

A : $1, 2, 3, 4, 5, 6, 7, 8$ B : $1, 2, 3, 4, 5, 6, 7, 100$

- 標本平均と中央値に関して，はずれ値の影響を見る．
  - A は標本平均も中央値も 4.5.
  - B は標本平均が 16, 中央値は 4.5.

標本平均ははずれ値の影響を受けやすいのに対し，中央値ははずれ値の影響を受けない．

- 標本分散, 範囲, 四分位範囲に関して，はずれ値の影響があるかを見る．
  - A の標本分散は 5.25, 範囲は $8 - 1 = 7$, 四分位範囲は $7 - 2 = 5$.
  - B の標本分散は 1011.5, 範囲は $100 - 1 = 99$, 四分位範囲は $7 - 2 = 5$

標本分散や範囲ははずれ値の影響を受けやすいのに対し，四分位範囲ははずれ値の影響を受けない．

1. 4. 5. 箱ひげ図

・箱ひげ図とは中央値，第１四分位数，第３四分位数，四分位範囲，はずれ値を図で表すことにより，データがもつ特徴をヒストグラムよりも簡潔に表す方法.

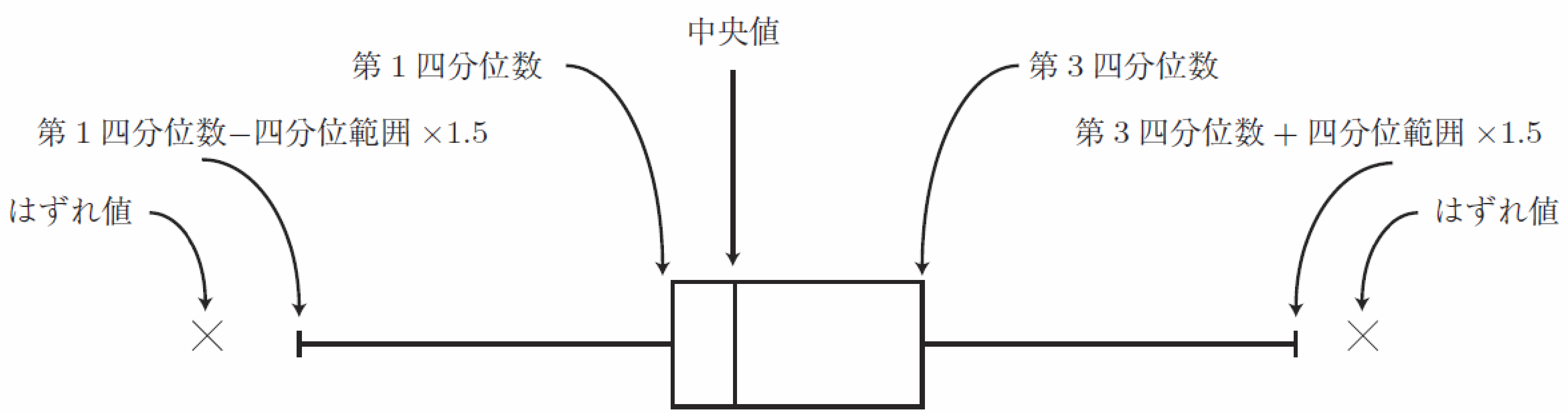


図 1.12　箱ひげ図のイメージ

1．4．5. 箱ひげ図

・ヒストグラムと箱ひげ図の関係

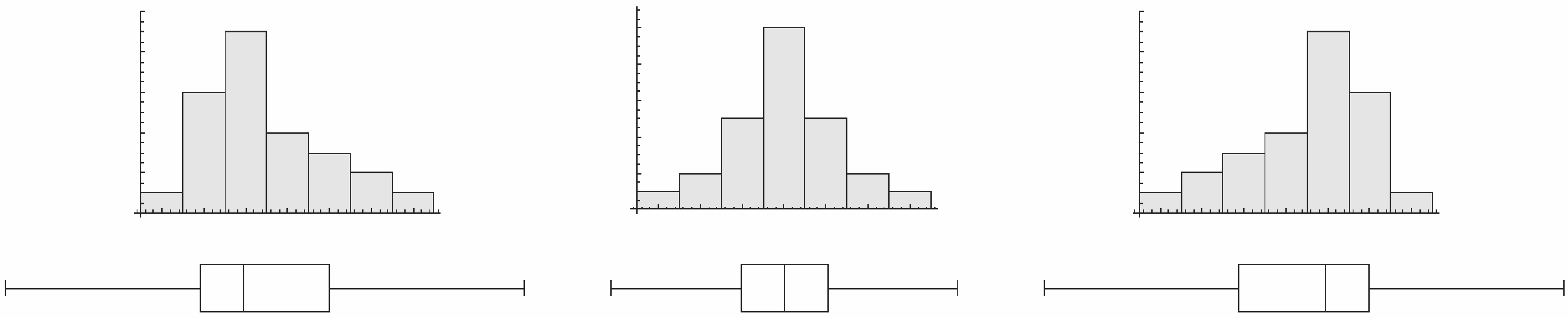


図 **1.13**　箱ひげ図とヒストグラムの関係

・ヘリコニア・ビハイ

・ヘリコニア・カリバエア（赤）

・ヘリコニア・カリバエア（黄）

1．4．5．箱ひげ図

・次のデータは熱帯アフリカ原産のヘリコニアの 3 種類の花の長さ (mm) を測った結果である.

・ヘリコニア・ビハイ（データの個数：16）

$16 = 8 \times 2$

46.34, 46.44, 46.64, 46.67, 46.75, 46.81, 46.94, 47.12, 47.12, 47.43, 48.07, 48.15, 48.34, 48.36, 50.12, 50.26

・ヘリコニア・カリバエア（赤）（データの個数：23）

$23 \times 0.25 = 5.75$
$23 = 12 \times 2 - 1$

37.4, 37.78, 37.87, 37.97, 38.01, 38.07, 38.1, 38.2, 38.23, 38.79, 38.87, 39.16, 39.63, 39.78, 40.57, 40.66, 41.47, 41.69, 41.9, 41.93, 42.01, 42.18, 43.09

・ヘリコニア・カリバエア（黄）（データの個数：15）

$15 \times 0.25 = 3.75$
$15 = 8 \times 2 - 1$

34.57, 34.63, 35.17, 35.45, 35.68, 36.03, 36.03, 36.11, 36.52, 36.66, 36.78, 36.82, 37.02, 37.1, 38.13

1．4．5．箱ひげ図

・ヘリコニア・ビハイ (データの個数：**16**)

第1四分位数 : 46.67,　中央値 : 47.12,　第3四分位数 : 48.34

四分位範囲:$48.34 - 46.67 = 1.67$

第1四分位数 $-$ 四分位範囲 $\times 1.5 = 44.165 \fallingdotseq 44.17$,

第3四分位数 $+$ 四分位範囲 $\times 1.5 = 50.845 \fallingdotseq 50.85$

・ヘリコニア・カリバエア（赤） (データの個数：**23**)

第1四分位数 : 38.04,　中央値 : 39.16,　第3四分位数 : 41.795

四分位範囲:$41.795 - 38.04 = 3.755$

第1四分位数 $-$ 四分位範囲 $\times 1.5 = 32.4075 \fallingdotseq 32.41$,

第3四分位数 $+$ 四分位範囲 $\times 1.5 = 47.4275 \fallingdotseq 47.43$

・ヘリコニア・カリバエア（黄） (データの個数：**15**)

第1四分位数 : 35.31,　中央値 : 36.11,　第3四分位数 : 36.92

四分位範囲:$36.92 - 35.31 = 1.61$

第1四分位数 $-$ 四分位範囲 $\times 1.5 = 32.895 \fallingdotseq 32.90$,

第3四分位数 $+$ 四分位範囲 $\times 1.5 = 39.335 \fallingdotseq 39.34$

1. 4. 5. 箱ひげ図

３つの箱ひげ図 ((i) ヘリコニア・ビハイ, (ii) ヘリコニア・カリバエア (赤), (iii) ヘリコニア・カリバエア (黄)) を並べて描いた図

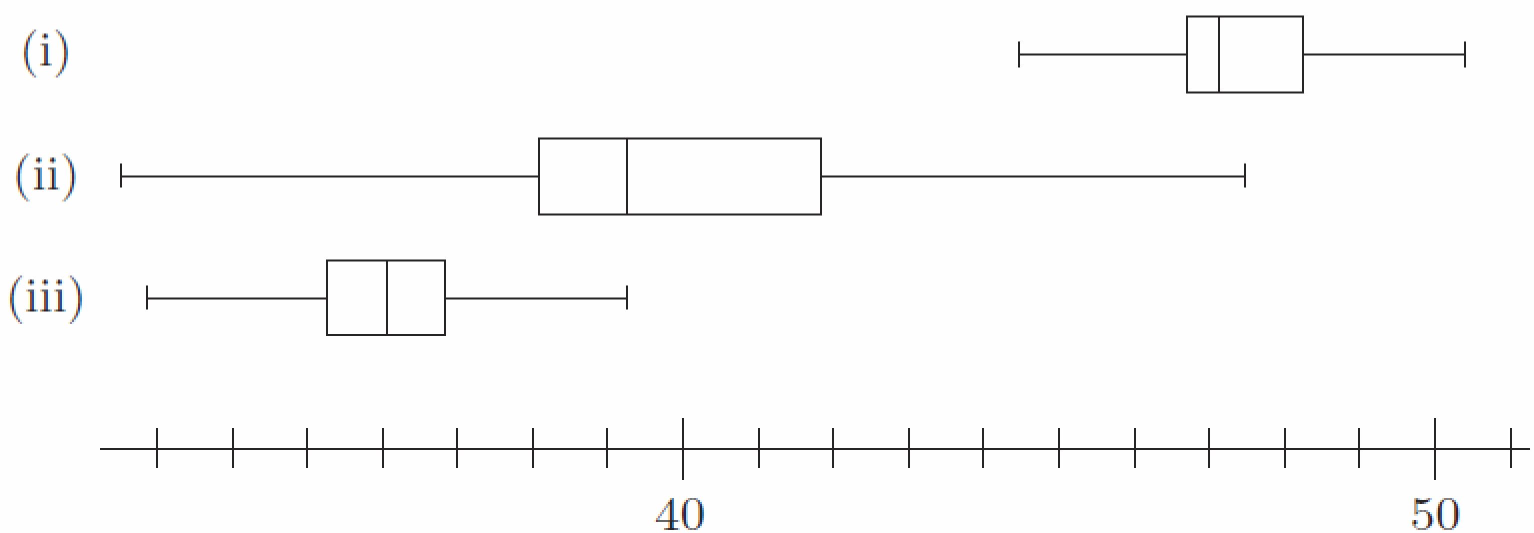


図 1.14　例 1.9 の箱ひげ図

1．4．5. 箱ひげ図

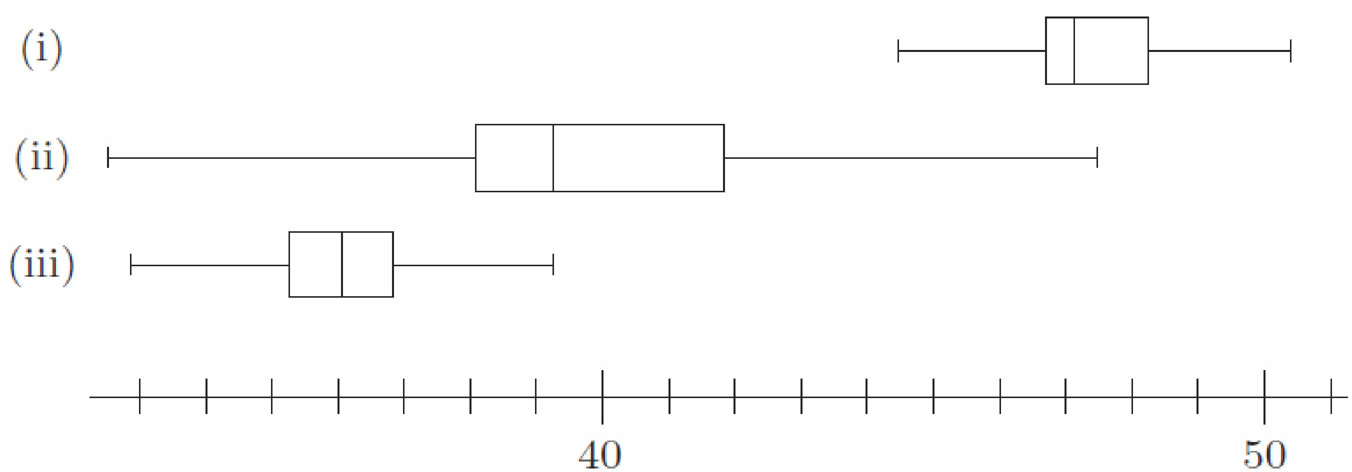


図 1.14　例 1.9 の箱ひげ図

- (i), (ii), (iii) の順にデータは全体的に小さくなっている.
- (i),(ii) は (iii) に比べて，データが小さい方に偏っている.
- (ii) は (i), (iii) に比べて，データのばらつきが大きい.
- (iii) は (i), (ii) に比べて，データの偏りが小さい.

まとめ

・いろいろな代表値

最頻値， 最小値， 最大値， 範囲， 四分位数， 四分位範囲， 偏差値, はずれ値

・箱ひげ図

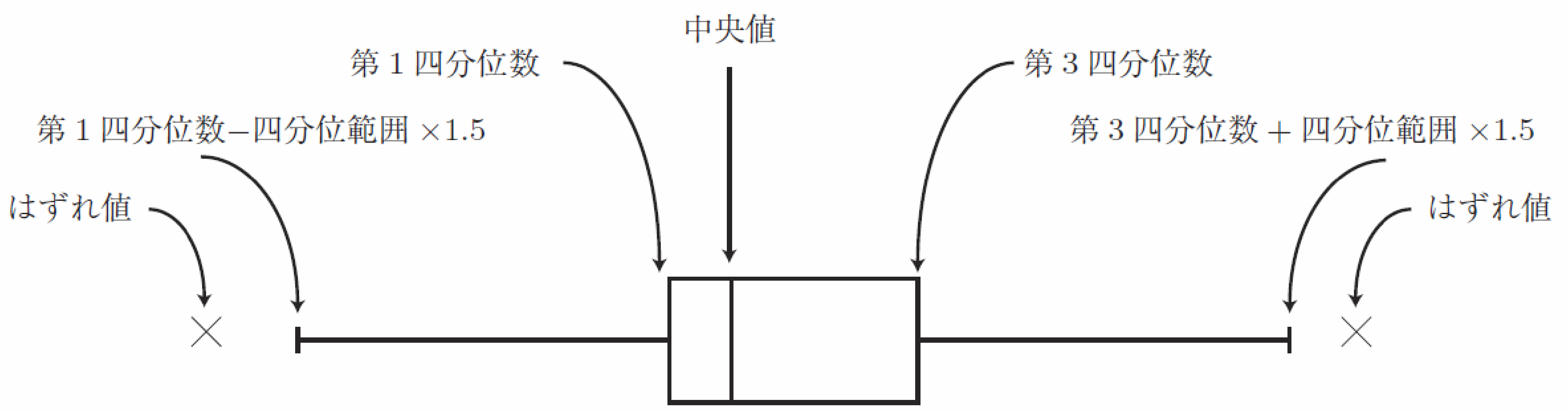


図 1.12　箱ひげ図のイメージ