

# 問題の発見・分析 ～ 統計的分析とデータの表現 ～

情報の科学 第20回授業

04情報の蓄積と管理

対応データ 20exp20.xls

# 今日のテーマ

- 数値で示す
- 数値の「意味」を知る
- わかりやすく示す

# データの種類

ここに  
注目！

データの種類	尺度	意味	単位	例
質的データ (定性的)	名義尺度	区別しかできない	ない	職業区分、電話番号
	順序尺度	大小比較ができる	ない	優良可の区分、震度
量的データ (定量的)	間隔尺度	差が意味を持つ	ある	気温、偏差値
	比率尺度	比が意味を持つ	ある	長さ、重さ

# 統計的分析

- 定量的なデータを、数値(統計量とも言う)を用いて分かりやすく示す
  - 代表的や特徴的な数値を用いる
    - 平均値、中央値、最頻値、最大値、最小値 など
  - 散らばり具合を示す
    - 分散、標準偏差、範囲、四分位偏差 など
  - 偏り具合を示す
    - 尖度(せんど)、歪度(わいど) など
  - 2変数の関係を分かりやすく示す
    - 相関、相関係数、回帰直線
  - 違いを見極める
    - 統計的仮説検定の考え方

# 同じ平均値でも、集団の性質が違う

- 大きい人と小さい人との差が大きいようだ
  - データの「偏り」を客観的に表す必要性
  - 偏りを数値化する必要性

例：A組 最大177.1 最小153.2 範囲23.9

B組 最大180.3 最小149.7 範囲30.6

# データの偏りを表す数字

分散:

- ・それぞれのデータの平均値との差をとり、
- ・その差を二乗し、平均をとったもの

標準偏差:

- ・分散の正の平方根

備考:

偏差値・標準偏差をもとに、平均が50になるように数値化したもの

# 分散と標準偏差

(数学でも学習します)

	得点	平均との差	平均との差の2乗
A	67	13	169
B	55	1	1
C	42	-12	144
D	57	3	9
E	49	-5	25
平均	54	0	分散 → 69.6

標準偏差(分散の正の平方根) → 8.342661446

# 基本統計量

☆他にも、いろいろな「傾向」を数値で表せる。

中央値（メジアン）

最頻値（モード）

四分位数

標準偏差

四分位偏差

分散

尖度（せんど：ヒストグラムの「とがり具合」）

歪度（わいど：ヒストグラムの「左右対称性」）

範囲（レンジ）

最小

最大

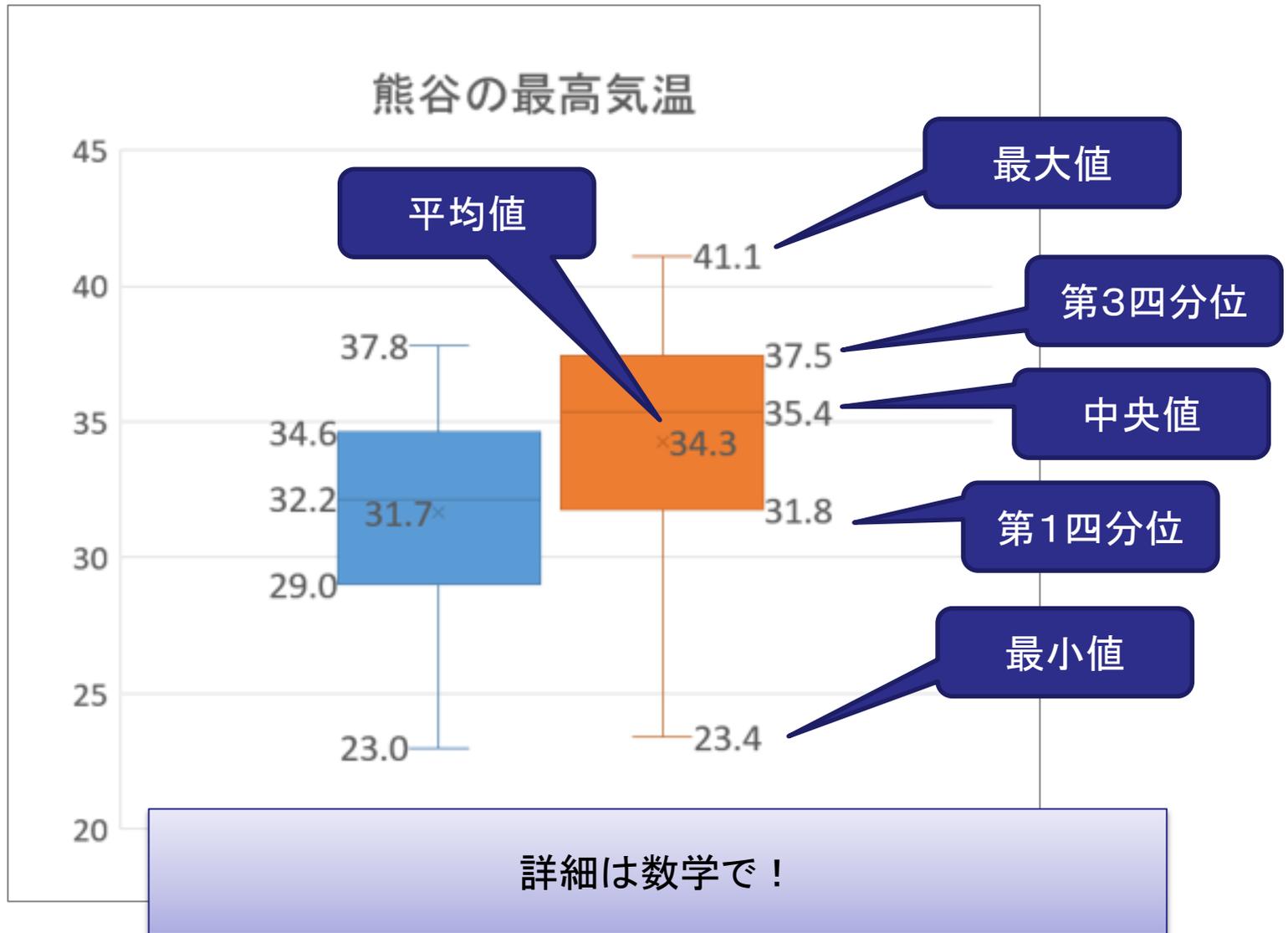
合計

標本数

# データのありかを意識

- 「気象庁 最高気温 データ」で検索

# 箱ひげ図(データを4分割する)



# 演習1

- 2017年と2018年の熊谷の最高気温について、箱ひげ図を描いてみよう。