

# 大学生のための 統計学入門

## —テーマ1. データの要約—

京都大学大学院医学研究科  
臨床統計学/臨床統計家育成コース 田中司朗



# テーマ1. データの要約

- 数値データの例
- 分布の位置の指標
- バラツキの指標
- **偏差値と標準化**



# 指標と単位の関係

- 容疑者のブーツの足跡の最大長（単位はcm）から、平均・分散・標準偏差を計算した
  - 平均: 24.66
  - 分散: 0.11
  - 標準偏差: 0.33
- これらの指標の単位は何になる？
- 答え
  - 平均: cm
  - 分散:  $\text{cm}^2$
  - 標準偏差: cm

# 指標と単位の関係

- 容疑者のブーツの足跡の最大長（単位はcm）から、平均・分散・標準偏差を計算した
  - 平均: 24.66
  - 分散: 0.11
  - 標準偏差: 0.33
- 単位がmmだったら平均, 分散, 標準偏差はどうなる?
- 答え
  - 平均:  $24.66 \times 10$
  - 分散:  $0.11 \times 10^2$
  - 標準偏差:  $0.33 \times 10$

# 偏差値

- 学力テストの得点が70点だったとしても、それだけでは成績は読み取れない
  - テストが難しければ良い成績かもしれない
  - テストが簡単なら悪い成績かもしれない
- 偏差値は**単位によらない**
  - 難易度によらず、集団の中でどれくらいに位置するのかがわかるようにした指標
  - 偏差値50は、平均点に対応
  - 偏差値50を基準に、どれくらい離れているかをみる

# 偏差値

- 平均点は、分布の位置の指標だから、学生全体の成績がどのあたりにあるか（難易度）を表している
- 一方、標準偏差は、学生間の成績のバラツキを表している
- 偏差値の計算のポイント
  - 自分の得点と平均点の差をとる（**難易度の調整**）
  - 標準偏差で割る（**バラツキの調整**）
- 難易度やバラツキが異なるテストの成績を比べるために便利

# 偏差値と順位

- 偏差値70: 学生10000人のうち, 228番目くらいの成績
- 偏差値65: 学生10000人のうち, 668番目くらいの成績
- 偏差値60: 学生10000人のうち, 1587番目くらいの成績
- 偏差値55: 学生10000人のうち, 3085番目くらいの成績
- 偏差値50: 学生10000人のうち, 5000番目くらいの成績
- 偏差値45: 学生10000人のうち, 6915番目くらいの成績
- 偏差値40: 学生10000人のうち, 8413番目くらいの成績
- 偏差値35: 学生10000人のうち, 9332番目くらいの成績
- 偏差値30: 学生10000人のうち, 9772番目くらいの成績

# 偏差値の公式

- 学生  $i$  の偏差値  $= \frac{10(x_i - m)}{s} + 50$ 
  - $x_i$ : 学生  $i$  の得点
  - $m$ : 平均
  - $s$ : 標準偏差
  - 平均50, 標準偏差10, 単位は無単位になる



# 偏差値の計算例

- 学生  $i$  の偏差値  $= \frac{10(x_i - m)}{s} + 50$
- 学力テストの得点が70点, 平均点が50点だった
- 標準偏差が10点のときと20点のときで, 成績の良し悪しがどれくらい違うのか?
  - 10点のときの偏差値  $= \frac{10(70-50)}{10} + 50 = 70$
  - 20点のときの偏差値  $= \frac{10(70-50)}{20} + 50 = 60$

# 偏差値の計算例

- 学生  $i$  の偏差値  $= \frac{10(x_i - m)}{s} + 50$
- 学生Aは, 平均点60点, 標準偏差20点のテストで, 得点が70点だった
- 学生Bは, 平均点50点, 標準偏差5点のテストで, 得点が55点だった
- どちらの成績が良いのか?
  - Aの偏差値  $= \frac{10(70-60)}{20} + 50 = 55$
  - Bの偏差値  $= \frac{10(55-50)}{5} + 50 = 60$

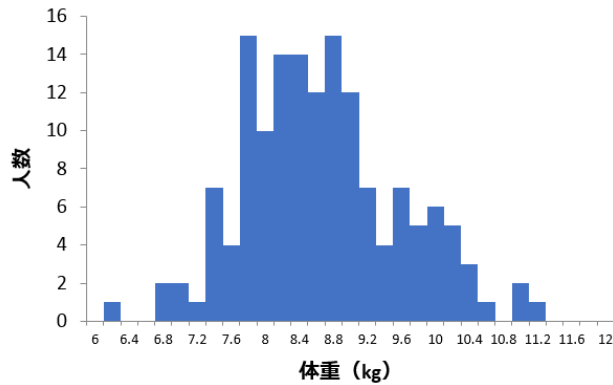
# 標準化という考え方

- 複数の測定値を比べるとき, 単位に依存するとわずらわしい
- そのようなとき**標準化**すると便利
- 標準化した値は, 個々の測定値  $x_i$  から以下のように計算される
  - 標準化した値 =  $\frac{x_i - m}{s}$ 
    - $m$ : 平均
    - $s$ : 標準偏差
- 平均ゼロ, 標準偏差1になって, 単位によらなくなる (無単位)

# テーマ1のまとめ

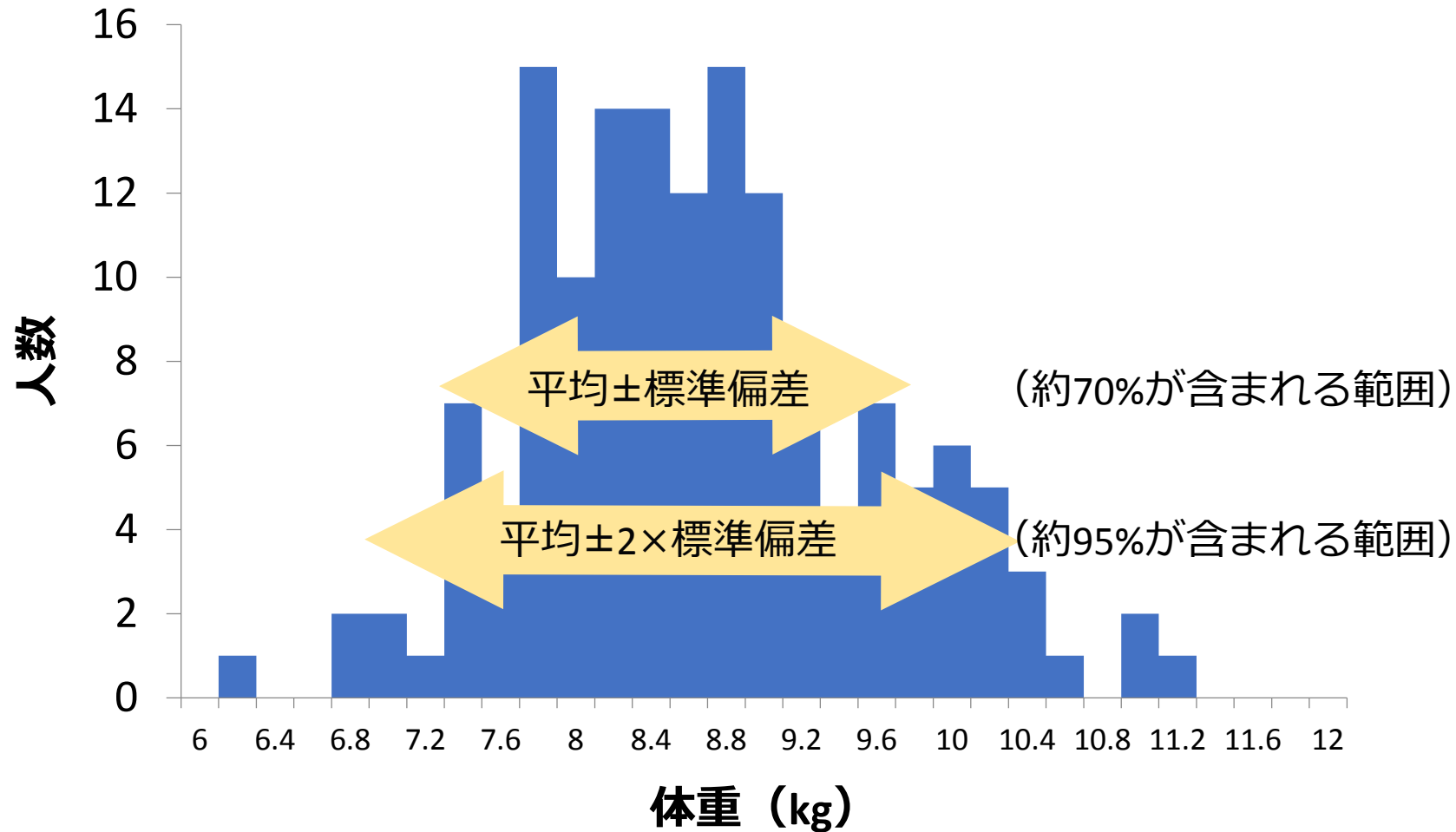
- データの散らばり具合のことを分布という

9.2kg	9.6kg	9.3kg	10.9kg	7.7kg	9.1kg	9.7kg	8.0kg	6.2kg	10.2kg
8.5kg	7.8kg	7.7kg	8.7kg	10.9kg	7.1kg	8.7kg	8.2kg	8.9kg	8.7kg
8.1kg	7.5kg	9.6kg	7.7kg	7.7kg	8.3kg	8.3kg	8.9kg	8.9kg	9.9kg
7.8kg	8.0kg	8.9kg	8.3kg	7.2kg	9.4kg	8.8kg	7.9kg	9.2kg	8.7kg
8.9kg	8.5kg	7.9kg	7.5kg	8.3kg	8.2kg	9.6kg	7.3kg	8.3kg	8.0kg
8.1kg	8.8kg	8.8kg	9.0kg	7.8kg	9.9kg	8.4kg	11.1kg	7.7kg	8.5kg
10.3kg	7.3kg	7.5kg	8.0kg	8.5kg	9.8kg	7.8kg	8.5kg	8.7kg	8.9kg
8.9kg	8.1kg	7.9kg	10.2kg	7.6kg	10.1kg	9.8kg	7.7kg	10.1kg	8.4kg
8.6kg	7.7kg	9.1kg	9.0kg	7.8kg	10.3kg	8.5kg	10.1kg	8.6kg	8.0kg
7.7kg	7.9kg	8.5kg	8.8kg	10.2kg	8.7kg	9.9kg	8.7kg	8.0kg	9.5kg
9.4kg	8.1kg	8.5kg	8.9kg	8.4kg	8.0kg	9.5kg	8.4kg	8.8kg	8.7kg
8.3kg	7.6kg	9.0kg	9.9kg	8.2kg	8.4kg	10.6kg	7.9kg	8.8kg	9.2kg
7.9kg	8.4kg	7.2kg	9.9kg	8.7kg	8.9kg	9.9kg	7.4kg	7.8kg	9.5kg
9.1kg	7.9kg	7.4kg	8.7kg	9.2kg	9.5kg	8.2kg	7.0kg	8.4kg	8.2kg
7.9kg	8.1kg	6.6kg	8.5kg	7.3kg	6.7kg	9.6kg	9.7kg	8.3kg	7.0kg



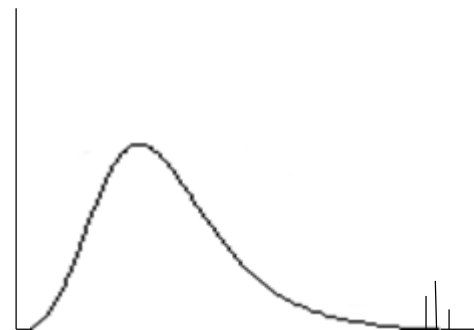
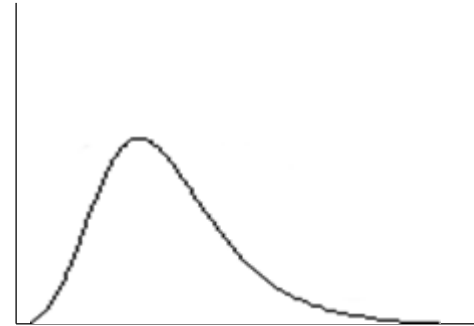
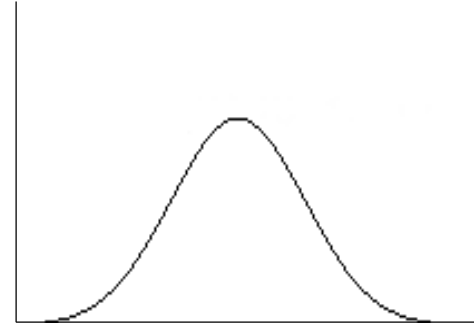
- 位置の指標
  - 平均値, 中央値, 最頻値
- バラツキの指標
  - 分散, 標準偏差, パーセント点, 四分位偏差

# テーマ1のまとめ



# テーマ1のまとめ

- データの分布が左右対称のとき
  - 平均値, 中央値, 最頻値は同じ値
  - どれを使っても大差はない
  - 平均値と標準偏差を使うのが便利
- 分布が右に裾を引いているとき
  - 平均値 > 中央値 > 最頻値
    - 値が大きい方からアルファベット順
    - $\text{mean} > \text{median} > \text{mode}$
- 外れ値があるとき
  - 最頻値は外れ値の影響を受けない
  - 中央値は影響が小さい
  - 平均値は外れ値に引っ張られる



# さらに学びたい人に

- 教科書

- 高校生からの統計入門. 加藤久和著. 筑摩書房2016
- 統計学入門. 東京大学教養学部統計学教室編. 東京大学出版会1991

- 引用文献

- 平均順位偏差値—データに負けるな! 吉村功著. 岩波ジュニア新書1984
- Stigler. Ann Stat 1981;9(3):465-74
- 総務省統計局 (<http://www.stat.go.jp>)