

生体情報論演習

- 統計法の実践 第1回 -

2011. 6. 3.

京都大学 情報学研究科 杉山磨人

統計法とは

- 統計法を用いる目的: **仮説の検証**
 - **実験データ**だけから全体に共通する性質を考察
 - 普通, 全部の要素を実験することはできない
 - 世界中に住むすべてのマウスに対して実験可能か?
- 一般的な(生命科学での)研究手順
 1. 仮説を決める
 2. 実験をおこない, データを得る
 3. データを**グラフ**にする
 4. **仮説検定**をおこない, 結論を得る

統計法

この演習での目標

- 具体的なデータを使って統計法に慣れる
- 基本的な概念を理解する
 - 平均, 標準偏差, 標準誤差, ...
- Excelを使って様々なグラフを描く
 - ヒストグラム, 棒グラフ, ...
- 仮説検定をする
 - t 検定, ...

実験データの分布を見る

- 実験データの性質を調べる第一歩は、その分布を観察すること
- 分布を観察するためのグラフ
→ ヒストグラム
- ヒストグラムの作り方:
データを階級に分け、その度数をプロットする

実験データの例

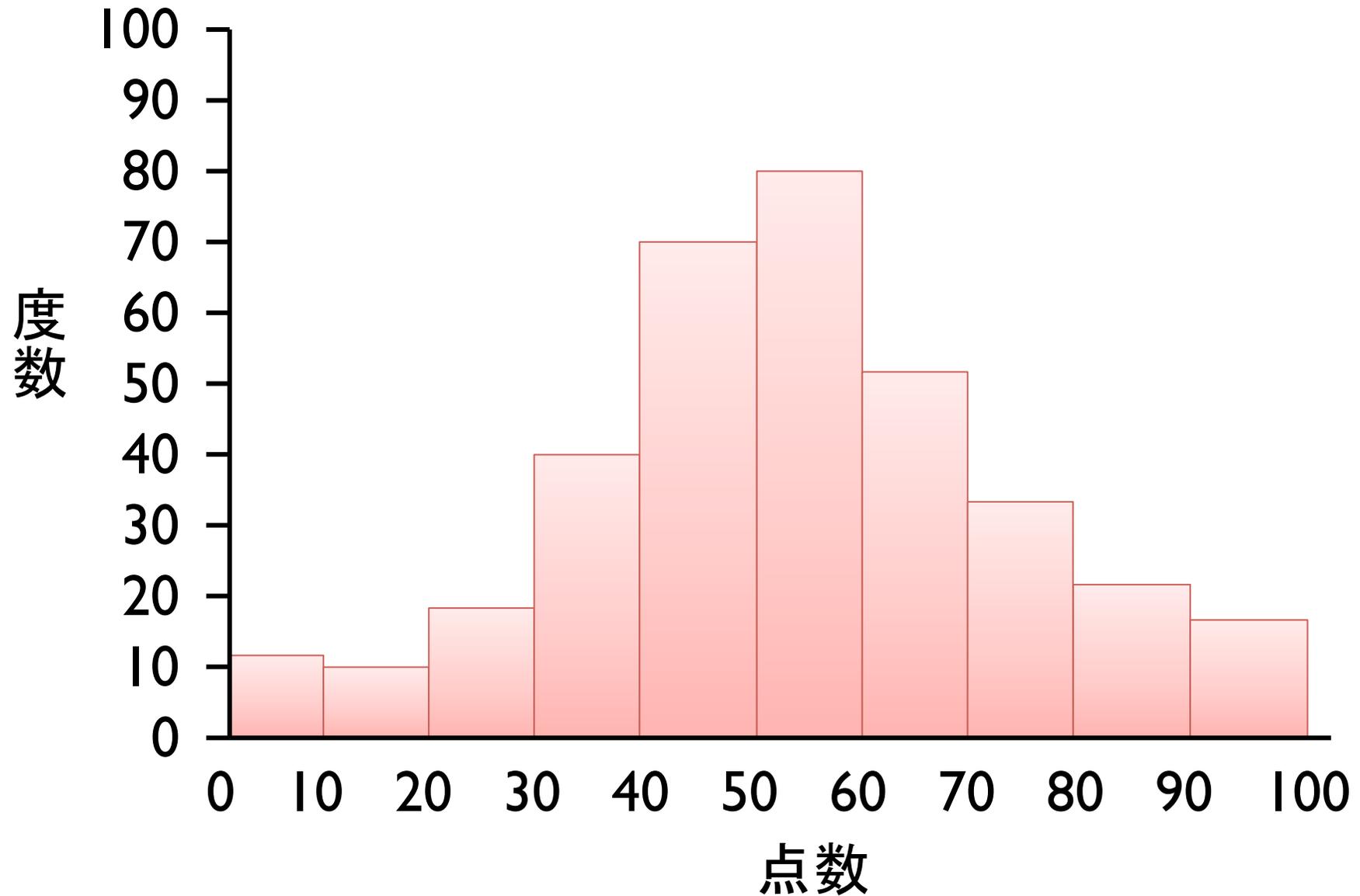
| 個体 No. | 測定値 |
|--------|-----|
| 1 | 31 |
| 2 | 87 |
| 3 | 43 |
| 4 | 25 |
| 5 | 10 |
| 6 | 31 |
| 8 | 98 |
| ⋮ | ⋮ |
| 373 | 21 |

参考: 統計学入門(東京大学出版会)

実験データの度数分布表

| 階級 | 度数 |
|----------|-----|
| 0 ~ 10 | 12 |
| 11 ~ 20 | 10 |
| 21 ~ 30 | 19 |
| 31 ~ 40 | 42 |
| 41 ~ 50 | 72 |
| 51 ~ 60 | 82 |
| 61 ~ 70 | 54 |
| 71 ~ 80 | 38 |
| 81 ~ 90 | 25 |
| 91 ~ 100 | 19 |
| 合計 | 373 |

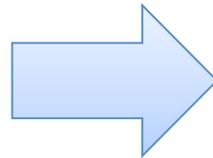
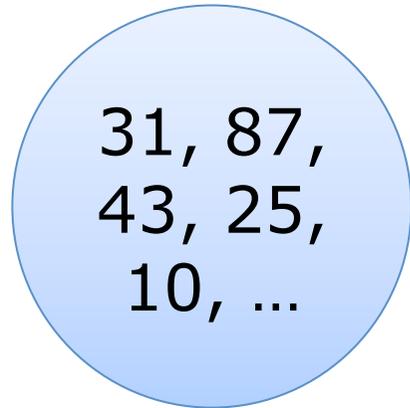
実験データの分布をヒストグラムで表す



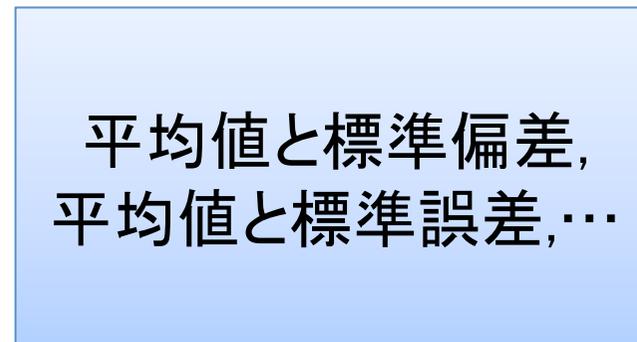
統計ではデータの分布に注目する

- データを分布として捉える
 - 分布とは、「データの散らばり方」
- 分布をいくつかの数値（パラメータ）で要約する
 - 特に、「平均値と標準偏差」、「平均値と標準誤差」の組み合わせが重要

データの集まり

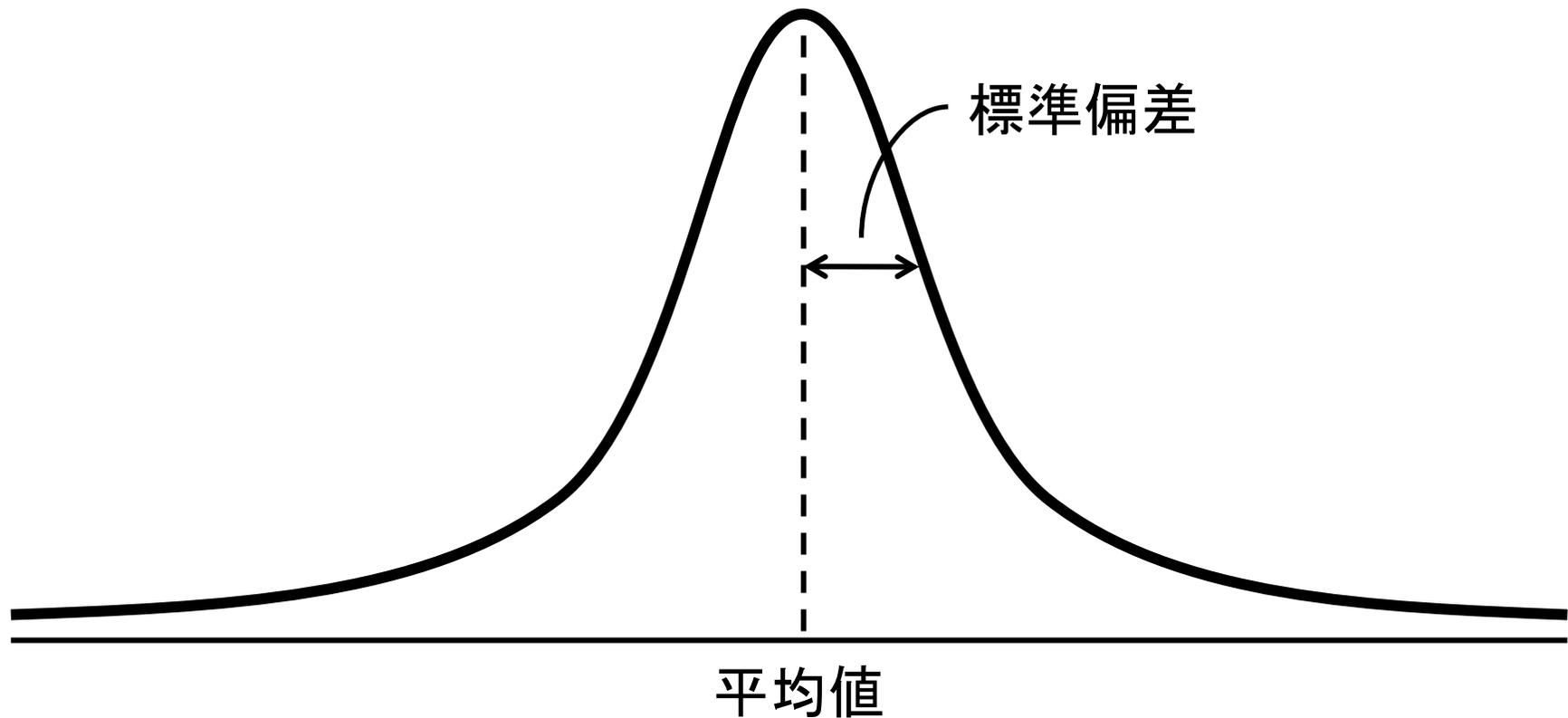


分布のパラメータ

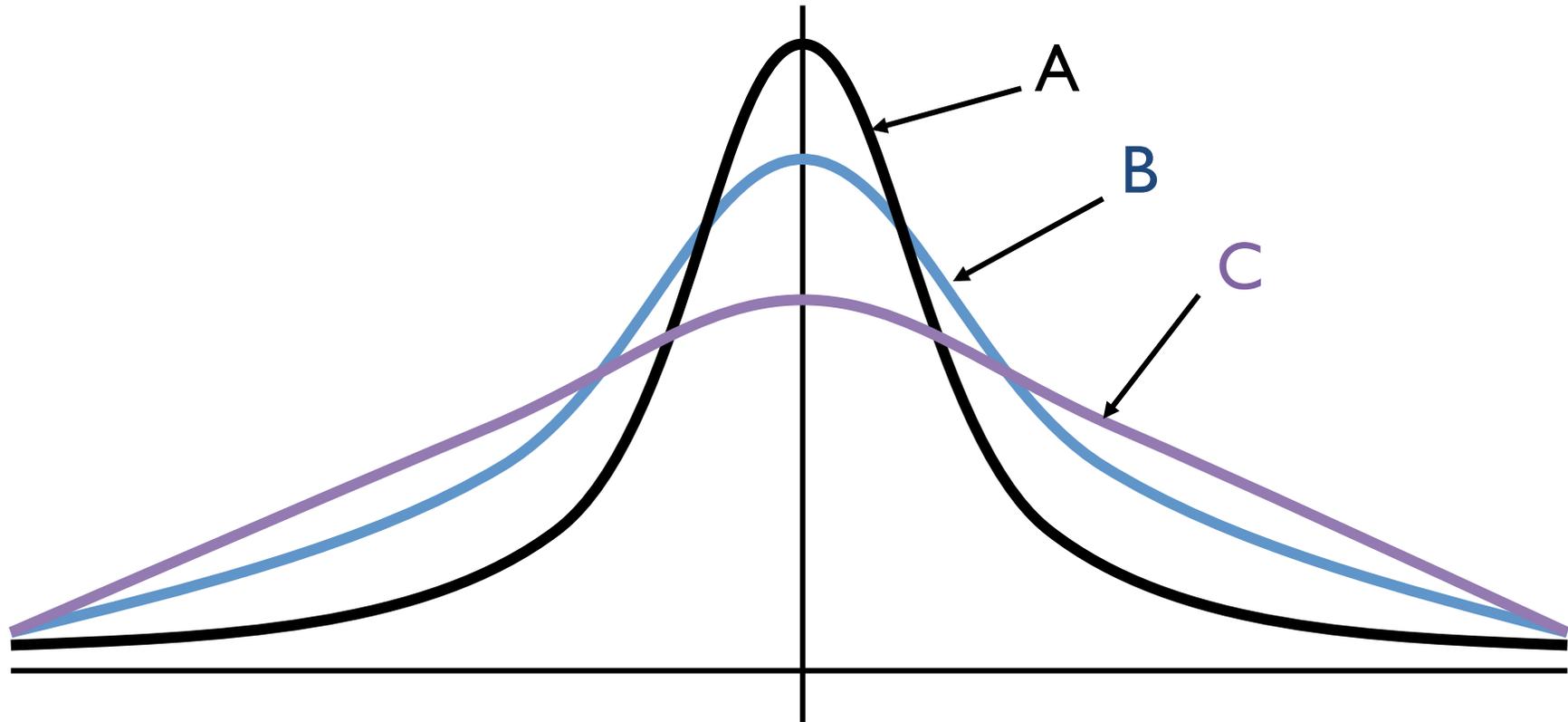


平均値と標準偏差

- **平均値**: データの分布の真ん中の位置を示す
- **標準偏差**: データの分布の広がりを示す

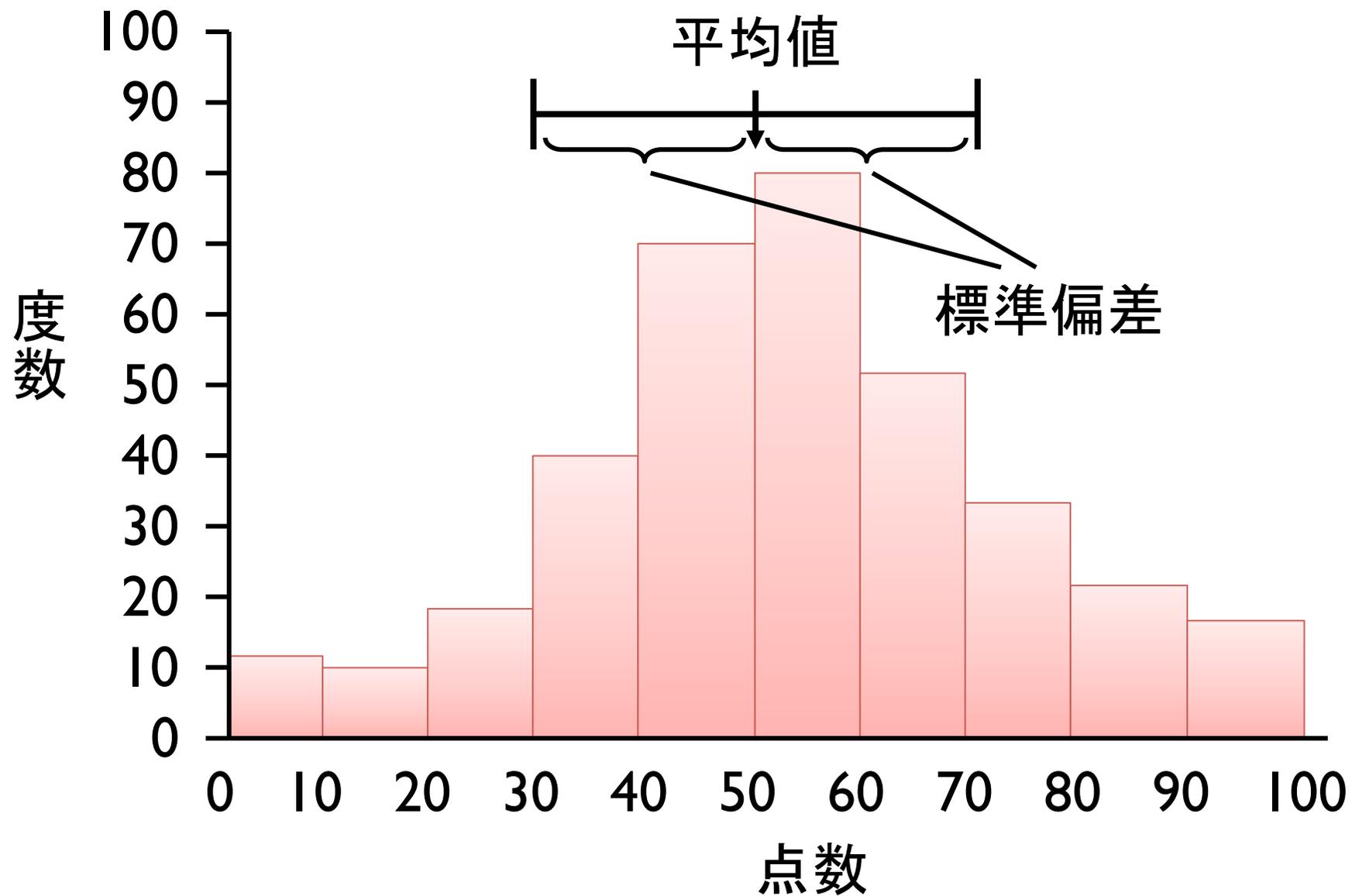


平均値は同じだが，分布が異なる場合

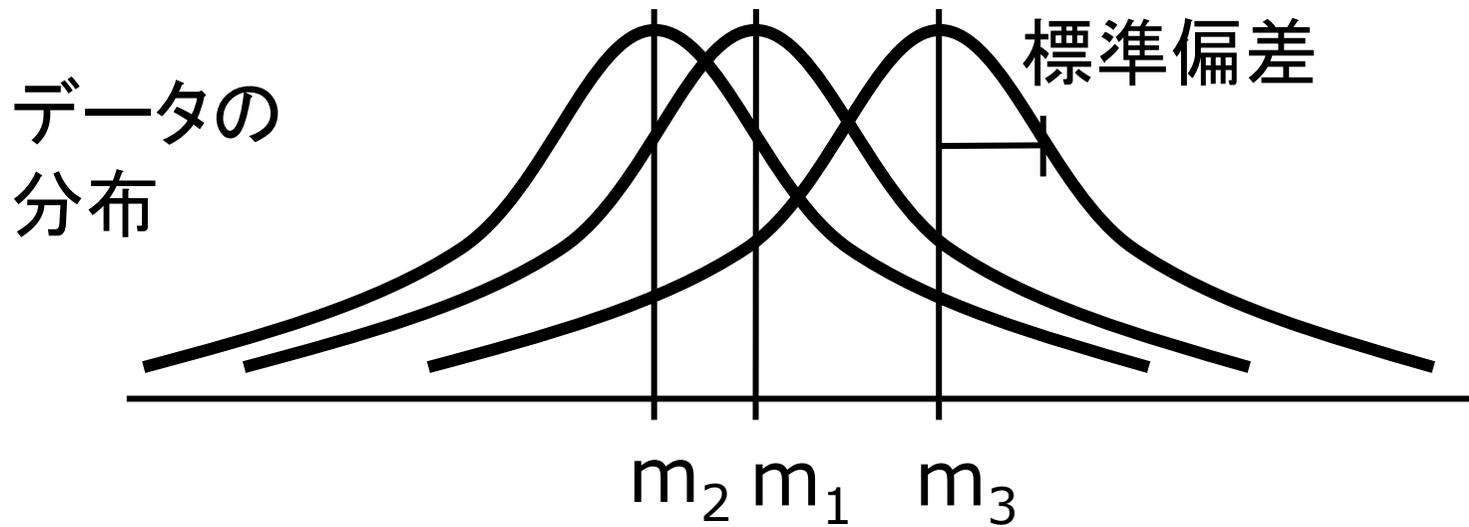


- A, B, Cの分布は平均値が等しい
 - しかし，分布は明らかに異なる
- **標準偏差**で違いを表すことができる

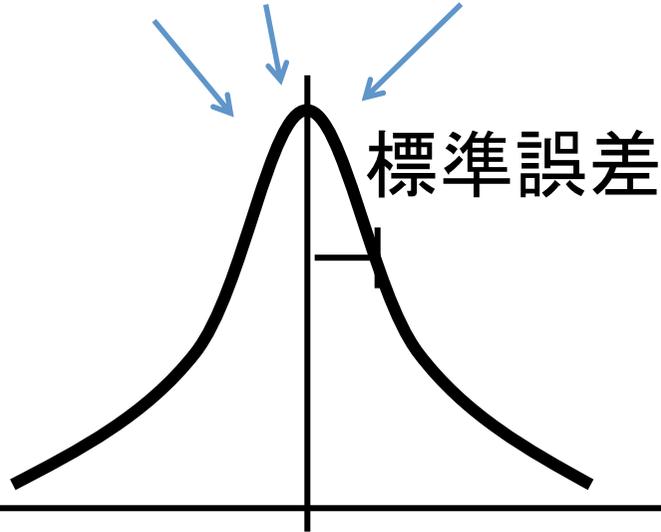
ヒストグラム上での平均値と標準偏差



データの分布と平均値の分布



データの平均値の分布



標準誤差: データの平均値の分布の標準偏差

課題

- Excelファイルにある各データ(Data1, Data2, Data3)について以下の2つをする
 1. ヒストグラムを描く
 2. 平均値, 標準偏差, 標準誤差を求める
- (なんでも良いので)3つのデータそれぞれの特徴を述べる
- 余力があれば・・・
 - データの数 n が変わると, 標準偏差と標準誤差がどのように変わるか説明する

課題

- Excelで計算, グラフ作成を行う
- 解答をPowerPointにまとめる
 - PowerPoint ファイルで提出
 - ファイル名を自分の名前と今日の日付にする
 - 例: 杉山0603.ppt
- 締め切り: 6月9日(木) 18:00
- 課題の提出先:
mahito+seitai(at)iip.ist.i.kyoto-u.ac.jp