

**【手がかり用スペース】**

前回の復習

4 種類の尺度について

名義尺度  
順序尺度  
間隔尺度  
比例尺度

尺度の値

世の中の値は 4 種類のそれぞれの値のどれかに含まれる  
次元が高い尺度は次元が低い尺度に変換可能（逆は不可能）  
高い次元の尺度の方が高度な分析を行うことができる

練習問題の解答

問 1：名義尺度の例を 2 つ以上記しなさい

問 2：比例尺度の例を 2 つ以上記しなさい

問 3：次の数字は、名義尺度、順序尺度、間隔尺度、比例尺度のうち、  
どれに属するか書きなさい。

カレンダーの日付  
洋服のサイズ  
教室番号  
モース硬度  
絶対温度

問 4：Wikipedia の「尺度水準」のページを読み、それぞれの尺度に  
おいて利用される代表値についてまとめなさい。

今後も練習問題は SNS の日記で解答のこと

本日の内容

ばらつき具合の視覚化

集団は画一的ではない（ばらつきがある）  
特に人間はばらつきが大きい  
まずは視覚化して全体をとらえる

**【要点書き出し用スペース】**

**【手がかり用スペース】**

データの特徴を把握する

データの特徴の捉え方

一般的にデータを取得した時点では全体像は把握しづらい

サンプルデータ：50 人の小学生のテスト結果

54, 48, 44, 48, 74, 46, 92, 65, 42, 18

59, 54, 48, 52, 78, 52, 28, 74, 46, 22

49, 58, 53, 56, 84, 55, 34, 76, 50, 24

54, 64, 56, 62, 38, 61, 36, 80, 55, 31

58, 54, 64, 69, 44, 65, 42, 88, 60, 36

表としてまとめるメリット

表や図にすることは直感的な理解のために重要

データの入力ミスなどを発見しやすい

「度数分布表」と呼ばれる表を作成

サンプルデータに基づいて値を求めてみよう

a) 最小値：

b) 最大値：

c) 範囲：

d) 範囲を 10 で割った値：

級間の幅の決定

級 (Class) とはデータが含まれる等間隔の区間

分かりやすく、データを数えやすいもの

10-20 程度の級数を目安とする

値も分かりやすい方が良い

度数分布表作成のまとめ

級間 : 等幅にする

真の限界：級間の真の上限と下限（一つ下の位まで書く）

中心点：級間の中心の値が「中心点」または「階級値」と呼ぶ

度数 : 頑張って数える

累積度数：小さい方から数え上げて行く。最大値は N の値と等しい

**【要点書き出し用スペース】**

【手がかり用スペース】

表 1. 50 人の小学生のテスト結果による度数分布表

級間	真の限界	中心点	度数	相対度数	累積度数

N=50

ヒストグラムの作成

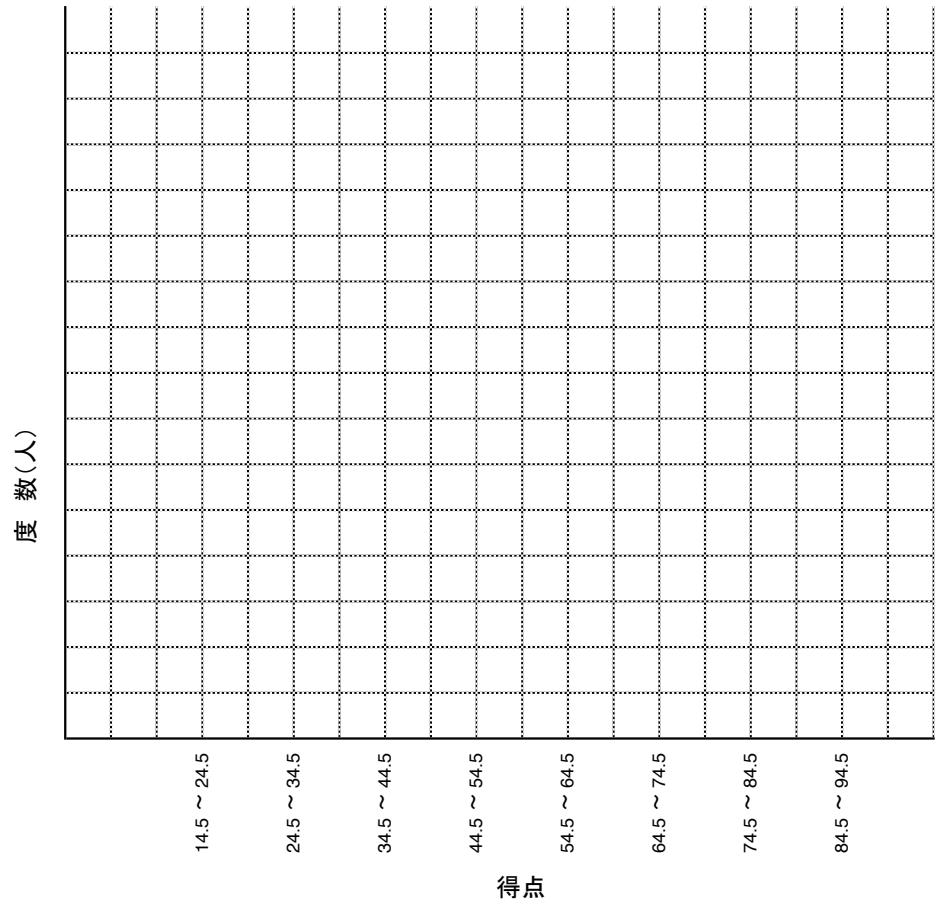


図 1. 50 人の小学生のテスト結果によるヒストグラム

【要点書き出し用スペース】

【手がかり用スペース】

ヒストグラムの作成

- 度数分布表を元に発展させる
- 棒グラフ状にデータを表記する
- 量的データは連続していることを示すために棒と棒の間に隙間を空けない
- 真の限界を基準として描く

度数多角形の作成

- 度数分布表を元にした図形
- 「度数多角形」または「度数ポリゴン」
- 多角形の形で描く
- ヒストグラムの棒の頂点（中心点）を直線で結ぶ
- 左右に度数 0 の項目を想定する

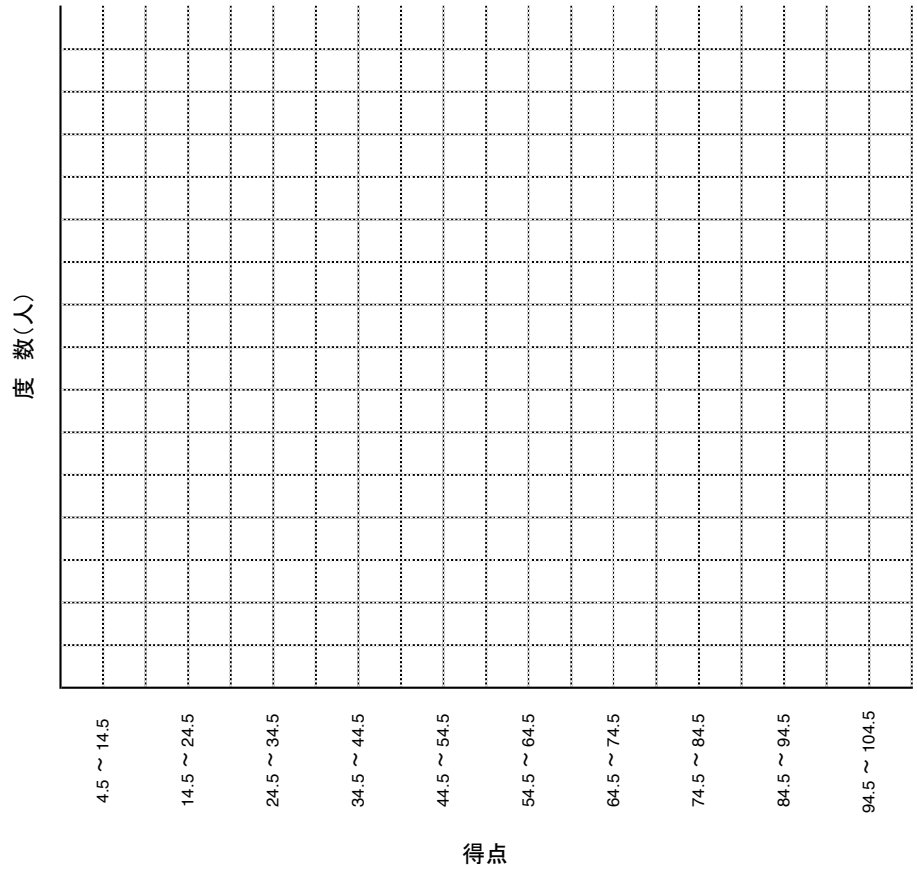


図 2. 50 人の小学生のテスト結果による度数多角形

【要点書き出し用スペース】

【手がかり用スペース】

データの視覚化

- データを見やすくまとめる
- 表や図にすると直感的な理解が得られる
- 調査や実験を行った後は必ず行うべき作業
- 視覚化のメリット
  - 全体の様子をとらえるのには良い
- 視覚化のデメリット
  - 他人に伝えるには困難

データを表現するためには他の方法も必要になる  
数値化

代表値

- 今回のサンプルデータの特徴を 1 つの数値で表現するには？
- 分布の特徴を 1 つの値で表すための数値

代表値の種類

- 平均値
- 中央値
- 最頻値

平均値

- 記号で表記する場合は  $M$  または  $\bar{X}$  で表現される
- 算出方法は結果を全て合計して人数で割ったもの

中央値 (メディアン)

- 記号で表現する場合は  $Me$  または  $Md$  で表現される
- データを小さい順に並べて中央に来た値
- データ数が偶数の場合は中央二つの値の平均値

最頻値 (モード)

- 記号で表現する場合は  $Mo$  で表現される
- 最も度数が多い級間の中心点の値

3 種類全て「分布を代表する値」足り得る

【要点書き出し用スペース】

【手がかり用スペース】

代表値の性質の比較

平均値

- 分布のすべての得点を使って求める
- 平均値からのすべての得点の差を合計すると 0 になる
- 極端な値に左右されやすい
- 数学的に操作しやすい

中央値

- 分布の特定の値に左右されづらい
- 歪曲した分布でも利用できる
- 平均値ほどには分布の性質が反映されない

最頻値

- 直感的にわかりやすい
- 極端な値には左右されづらい
- 数値処理には向かない
- 質的データに利用されることが多い

分布と 3 つの代表値の特徴

左右対称の分布

- 平均値と中央値と最頻値は一致

左側に頂点があり右に長く伸びた分布

- 正に歪曲した分布・L 型分布
- 平均値 > 中央値

右側に頂点があり左に長く伸びた分布

- 負に歪曲した分布・J 型分布
- 平均値 < 中央値

分布の歪み

正に歪んだ分布

負に歪んだ分布

来週もういちど説明します

【要点書き出し用スペース】