

確率・統計 前期 第6回

1次元のデータの度数分布と代表値

稲積 泰宏 (いなづみ やすひろ)

度数分布に関する用語

- **変量**：観察や測定によって得られる数値（例：身長，点数など）
- **階級**：データの区間（例：0～9，10～19など）
- **階級幅**：各階級の幅（例：0～9の階級幅は 10）
- **階級値**：各階級の中央の値（例：0～9 の階級値は 4.5）
- **度数**：各階級に属するデータの個数
- **累積度数**：ある階級までの度数の合計
- **相対度数**：各階級の度数を全体の個数で割った値
- **累積相対度数**：ある階級までの相対度数の合計

度数分布表とは？

- 変量の値をいくつかの階級に分け，それぞれに含まれるデータの個数（**度数**）を整理した表
- 階級の中央の値は**階級値**，階級の幅は**階級幅**と呼ばれる
- 相対度数や累積度数などを加えることで，分布の特徴がさらに分かりやすくなる

度数分布表の例（テストの点数）

階級（点数）	度数	相対度数	累積度数	累積相対度数
0～9	1	0.025	1	0.025
10～19	2	0.050	3	0.075
20～29	3	0.075	6	0.150
30～39	4	0.100	10	0.250
40～49	6	0.150	16	0.400
50～59	8	0.200	24	0.600
60～69	5	0.125	29	0.725
70～79	3	0.075	32	0.800
80～89	2	0.050	34	0.850
90～100	1	0.025	35	0.875

問1

度数分布のグラフ化

- **ヒストグラム**：横軸に階級値，縦軸に度数をとった棒グラフ
- 各棒の高さがその階級のデータの多さを示す
- **度数折れ線**：横軸に階級値，縦軸に度数をとって折れ線グラフにしたもの
- データの増減や分布の形を視覚的にとらえることができる

問2

代表値とは？

- データの**中心的な傾向**を示す値
- 主に次の3つがある：
 - **平均値**
 - **中央値**
 - **最頻値**

平均値

- 全データの合計をデータの個数で割った値

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

中央値

- データを小さい順に並べたときの中央の値
 - データ数が奇数：中央の値
 - データ数が偶数：中央2つの平均
- 例：(2, 4, 7) → 中央値は 4
- 例：(2, 4, 6, 8) → 中央値は $(4+6)/2 = 5$

最頻値

- 最も多く出現する値
- 度数分布表で度数が最大の階級の**階級値**を取る
- 例：(1, 2, 2, 3, 4) → 最頻値は 2

代表値の特徴

代表値	特徴
平均値	全体のバランスをとる値
中央値	外れ値の影響を受けにくい
最頻値	最もよく現れる値

平均の性質

変量 x, y の間に $y = ax + b$ の関係があるとき、それぞれの平均を \bar{x}, \bar{y} とすると

$$\bar{y} = a\bar{x} + b$$

例題1)

次の数値は、ある工場で生産された電球の寿命（単位 時間）である。
このデータから電球の寿命 x の平均を求めよ。

1485 1511 1499 1484 1483 1490 1489 1513 1487 1502
1486 1514 1501 1507 1484 1509 1475 1516 1503 1474

まとめ

- 度数分布はデータを分類・集約して整理する方法
- 代表値（平均・中央値・最頻値）はデータの中心傾向を示す
- ヒストグラムなどを活用すると、分布の様子が視覚的に分かる
- 質問、感想、わからなかったことなどをチャット欄に書いてください。
- 会議が閉じている場合は、チャットを送ってください。
- **次回までの課題：** Basic 72-77