統計計算の基礎

Y-teraya

2025 年 10 月 28 日

概要

本資料は，統計の基礎的な計算方法をまとめたものである．実験計画法を主として扱い，実務で直結する内容（Excel 函数など）を備忘録として記しておく．ほか，線形代数や微分積分も取り扱う．

また，総和 $\sum$ の記号に苦手意識を持っている人も多いので，理解して欲しい重要な式のみ $\sum$ 無しで表現した式も併記している．

目次

青文字をクリックすると，対応したページに遷移します.

留意事項

1. 色付き文字やハイライトは重要事項または強調箇所である.
2. 自身の好み（独断と偏見）で作成しているため，旧字体や座標を行列で記載している箇所がある.
3.

第 I 部
統計のデータ

1 代表値 Average

データ全体を分布中心のデータ 1 つで表したものを代表値という．主に 3 つの値のことを指し，平均値 Mean，中央値 Median，最頻値 Mode である．ただし，これらの値がデータの代表ではない可能性もあるため，扱うときには必ずデータの代表として機能しているのか確認する必要がある [3, 4].

1.1 平均値 Mean

主に算術平均のことを指す．全データを合計し，データの数で割ることで求められる．平均値を $\bar{x}$[*1]，データ数を n，各データを x_k とすると，以下である[*2].

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k \tag{1.1a}$$

$\sum$ 無しに記述すると，以下である．

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_k}{k} \tag{1.1b}$$

メリット [1] 全てのデータを考慮できる．

デメリット 外れ値に弱い．

1.2 中央値 Median

データを**大きさの順に**並べた際の，真ん中の値を指す．例えば，データの集合 A を以下の通りに設定する．

$$A = \{x_k | 1 \leq k \leq 5\} \ (x_1 < x_2 < x_3 < x_4 < x_5) \tag{1.2}$$

このとき，集合 A の中央値は x_3 である．

メリット [1] 外れ値に強い．

デメリット 全てのデータを十分に考慮に入れることができない．

[*1] μ と置くこともある．

[*2] $\sum$ の計算は，基本的に $1 \leq k \leq n$ $(k \in \boldsymbol{Z})$ の範囲内での**総和**を示す記号である．$\sum$ の下に書いてある数字から，上に書いてある数字までを**カウントアップして足したもの**である．

1.3 最頻値 Mode

データの集合内で，同じデータが存在する場合がある．その同じデータが最も出現する値を指す．例えば，データの集合 B を以下の通りに設定する．

$$B = \{x_k | 1 \le k \le 5\} \ (x_1 < x_2 = x_3 = x_4 < x_5) \tag{1.3}$$

このとき，集合 B の最頻値は x_2，x_3，x_4 である．

メリット [1] 外れ値に強い．
デメリット 1 つに決まらないことがある．また，サンプルサイズが少ないと使えない．

1.4 最大値 Maximum・最小値 Minimum

データの集合内で，一番大きい値を最大値，小さい値を最小値という．例えば，集合 A，式 (1.2) を基にして考える．このとき，最大値は x_5，最小値は x_1 である．

2 散布度 Dispersion

2.1 分散 Variance・標準偏差 Standard deviation

各データと**平均との差**の平均である．一般にデータの**バラつき**を表す．分散を σ^2[*3]，データ数を n，各データを x_k とすると，以下である．

$$\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2 \tag{2.1a}$$

$\sum$ 無しに記述すると，以下である．

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_k - \bar{x})^2}{n} \tag{2.1b}$$

標準偏差はこの 2 乗を外したもののことである．すなわち，分散 σ^2 をルート $\sqrt{\quad}$ に入れたものである．そのため，標準偏差を σ[*4]とすると以下である．

$$\sigma = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2} \tag{2.2}$$

---

*3 s^2 と置くこともある．
*4 s と置くこともある．

余談: なぜ 2 乗した誤差を考えるのか？ [2]

平均との差であれば，絶対値を用いれば良いと考えられる．絶対値で計算する平均偏差と呼ばれるものを紹介する．平均偏差は以下で求められる．

$$MAE = \frac{1}{n}\sum_{k=1}^{n}|x_k - \bar{x}| \tag{2.3}$$

ただし，これは数学的に扱いにくい．以下に 3 点挙げる．

1. 絶対値は函数として微分可能でなく，少々扱いづらい．
2. 場合分けが必要になることがある．
3. $\bar{x} \neq \arg\min_{\bar{x}} \sum_{k=1}^{n}|x_k - \bar{x}|$ である．

特に 3 つ目が問題であるが，話が難しくなるので詳しくは参考文献の記事を読んで欲しい．

2.2 範囲 Range

単に最大値と最小値の差である．例えば，集合 A，式 (1.2) を基にして考える．このとき，範囲は $x_5 - x_1$ である．

第 II 部
データのイメージ

3　スカラー Scalar とベクトル Vector

スカラーは**1 つの数**であり，ベクトルは**2 つ以上の数を束ねたもの**である．そしてベクトルの表現として**文字の上に矢印** $\vec{x}$ **を描いて**ベクトルを表現する．ただ，**非常に見づらく**なりやすいという欠点がある．そのため，大学以降ではスカラーはそのまま，ベクトルは**太字** $\boldsymbol{x}$ または **2 重文字** $\mathbb{x}$（文字に余計な線を 1 つ入れるだけ）で表現する．よく，スカラーは大きさ**だけ**持つ量で，ベクトルは大きさと**向きも**持つ量と理解している人も多い．それは何故か？

$$A = 2$$

図 1: スカラー量

$$\boldsymbol{B} = \begin{pmatrix} 2 & 1 & 3 \end{pmatrix}$$

図 2: ベクトル量

スカラーのイメージは，**1 次元**である．すなわち，x 軸だけの数直線を考えると単なる**大きさ**にすぎない．次にベクトルのイメージは，**2 次元**や **3 次元**である．x 軸だけでは，大きさしか表せなかったのに対し，数を束にすることで **2 つ目以降の数によって方向が決まる**．

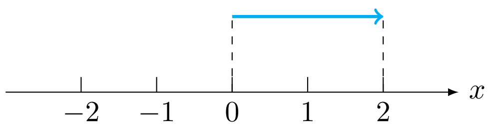

図 3: スカラーのイメージ

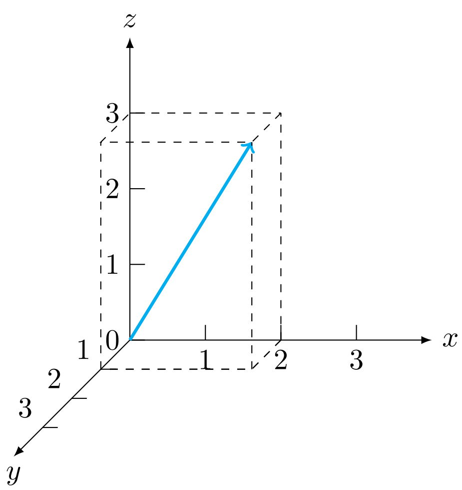

図 4: ベクトルのイメージ

簡単に説明すると，スカラーは **1 つの**数であり，ベクトルは**数と数の組**[*5]である．ベクトルは方向を表すことから，基本的に**矢印**で表現することが多い．そのとき大きさは，矢印の**長さ**で表す．

[*5] 厳密には，**線形空間の元**であるが，話が難しくなるので本資料では触れないことにする．線形空間の元として考えると，多項式もベクトル，函数もベクトル，微分方程式の解もベクトルとして捉えられる．これらのように，この世界にはベクトルでありふれている！

また，それぞれの量のことを表現するとき，それぞれに名前を付けて**スカラー量**，**ベクトル量**という．

3.1 ベクトルの加減法

スカラーはそのまま加減乗除できるが，ベクトルはそう上手くいかない．ベクトルの乗法には**内積**と**外積**の 2 種類あり，除法はできない．分かりやすい加減法から説明する．

言葉で説明するなら，矢印の終点ともう一方の矢印の始点を合わせ，1 つの折れ線矢印と見なして始点と終点を線で結ぶ．減法の場合は，矢印を逆にしてから足す．矢印が逆のベクトルを**逆ベクトル**という．

ほかに始点同士を揃えて平行四辺形にする方法もあり，運動方程式を解く際の分力を求めるときに最適である．

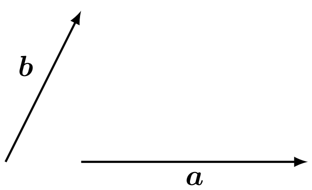

図 5: $\boldsymbol{a}$ と $\boldsymbol{b}$

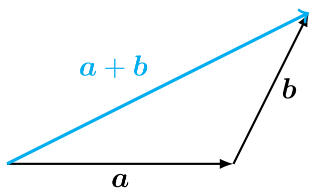

図 6: 三角形を作る方法

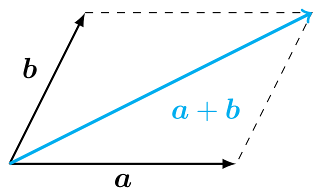

図 7: 平行四辺形を作る方法

3.2 単位ベクトル

ユークリッド空間（実数を n 個並べた全体の集合）において，3 つの直交座標をそれぞれ x 軸，y 軸，z 軸とする．そのなかで**大きさを「1」に仕立てた**ベクトルを単位ベクトルという．（単位〇〇は基本的に，〇〇の大きさを「1」に仕立てたもののことである．）

また，x 軸と平行な単位ベクトルを $\boldsymbol{i}$，y 軸と平行な単位ベクトルを $\boldsymbol{j}$，z 軸と平行な単位ベクトルを $\boldsymbol{k}$ とする．

3.3 ベクトルの成分表示

単位ベクトルと係数倍を用いて，一般にベクトルを次のような式で表せる．

$$\boldsymbol{a} = A\boldsymbol{i} + B\boldsymbol{j} + C\boldsymbol{k} \tag{3.1}$$

また，係数を座標のように表して，

$$\boldsymbol{a} = (A \quad B \quad C) \tag{3.2}$$

とも表せる．

参考文献

[1] 高校数学の美しい物語. 平均値, 中央値, 最頻値の求め方といくつかの例. https://manabitimes.jp/math/985, 2023. アクセス日: 2025-10-28.

[2] 数学の景色. データの分散・標準偏差の定義・具体例・性質まとめ. https://mathlandscape.com/variance, 2023. アクセス日: 2025-10-28.

[3] 内田誠一. 2-1-2. データの分布と代表値. https://mi.u-tokyo.ac.jp/consortium/pdf/2-1-2.pdf, 2020. アクセス日: 2025-10-28.

[4] 内田誠一. 2-1-3. 代表値の性質の違い. https://mi.u-tokyo.ac.jp/consortium/pdf/2-1-3.pdf, 2020. アクセス日: 2025-10-28.

索引