

統計計算の基礎

Y-teraya

2025 年 10 月 29 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 関数など）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

第Ⅰ部	初歩的な計算	1
第Ⅱ部	統計のデータ	2
1	代表値 Average	2
1.1	平均値 Mean	2
1.2	中央値 Median	3
1.3	最頻値 Mode	4
1.4	最大値 Maximum・最小値 Minimum	4
2	散布度 Dispersion	4
2.1	分散 Variance・標準偏差 Standard deviation	4
2.2	範囲 Range	5
第Ⅲ部	データのイメージ	6
3	スカラー Scalar とベクトル Vector	6
3.1	ベクトルの加減法	7
3.2	単位ベクトル	7

3.3 ベクトルの成分表示	7
-------------------------	---

[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. 本資料の著作権は、[CC BY-NC-SA 4.0](#) を適応する.

第 I 部

初歩的な計算

第 II 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}。

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k = \frac{x_1 + x_2 + \cdots + x_n}{n} \quad (1.1)$$

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。

エクセル =AVERAGE(Cell₁: Cell₂)

^{*1} μ と置くこともある。

^{*2} \sum の計算は、基本的に $1 \leq k \leq n$ ($k \in \mathbb{Z}$) の範囲内での総和を示す記号である。 \sum の下を書いてある数字から、上を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。

余談: 平均にもいろんな種類がある!

上記では、算術平均（相加平均）を取り扱った．一般に平均と言われればこの算術平均だと思えばよい．ただ，他にも場合によっては使える平均があるので紹介する．

1.1.1 幾何平均（相乗平均）

算術平均は足し算においての平均であったが，これは掛け算においての平均である．平均値を G ，データ数を n ，各データを x_k とすると，以下である^a．

$$G = \left(\prod_{k=1}^n x_k \right)^{\frac{1}{n}} = \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n} \quad (1.2)$$

主に比率や成長率，増加率など**比率や割合で変化する数の平均**を算出したいときに用いる．

1.1.2 調和平均

調和平均は，一度逆数にしたデータの算術平均のことである．平均値を H ，データ数を n ，各データを x_k とすると，以下である．

$$H^{-1} = \frac{1}{n} \left(\sum_{k=1}^n \frac{1}{x_k} \right) = \frac{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}{n} \quad (1.3)$$

主に平均速度や並列接続時の抵抗，直列接続時のコンデンサなど**比率で表示される数の平均**を算出したいときに用いる．

^a \prod の計算は \sum の足し算を**掛け算に変えた**ものである，基本的に $1 \leq k \leq n$ ($k \in \mathbf{Z}$) の範囲内での総積を示す記号である． \prod の下を書いてある数字から，上を書いてある数字までを**カウントアップして掛け**たものである．

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．

$$A = \{x_k | 1 \leq k \leq 5\} \ (x_1 < x_2 < x_3 < x_4 < x_5) \quad (1.4)$$

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.

エクセル =MEDIAN(Cell₁: Cell₂)

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。

$$B = \{x_k | 1 \leq k \leq 5\} \quad (x_1 < x_2 = x_3 < x_4 < x_5) \quad (1.5)$$

このとき、集合 B の最頻値は x_2, x_3 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.

エクセル =MODE.SNGL(Cell₁: Cell₂)

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。

エクセル (最大値) =MAX(Cell₁: Cell₂)

エクセル (最小値) =MIN(Cell₁: Cell₂)

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。

$$\sigma^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n} \quad (2.1)$$

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。

^{*3} s^2 と置くこともある。

^{*4} s と置くこともある。

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

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。

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

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 III 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 x** または **2 重文字 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

$$B = (2 \quad 1 \quad 3)$$

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、数を束ねることで**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

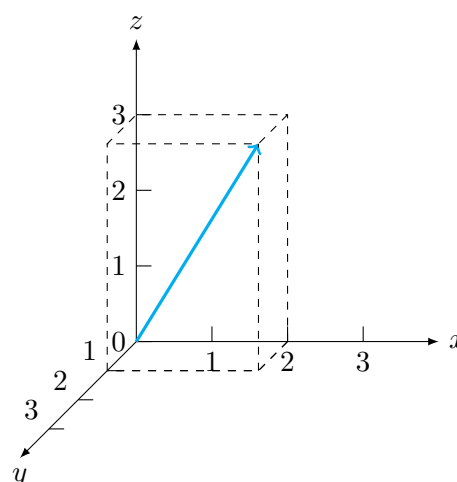


図 4: ベクトルのイメージ

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、関数もベクトル、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、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。

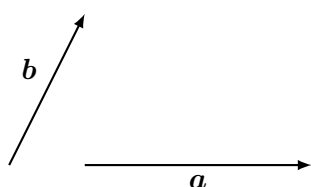


図 5: a と b

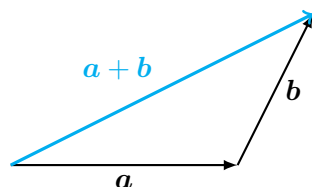


図 6: 三角形を作る方法

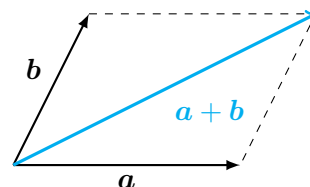


図 7: 平行四辺形を作る方法

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 \bigcirc は基本的に、 $\bigcirc\bigcirc$ 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 i 、 y 軸と平行な単位ベクトルを j 、 z 軸と平行な単位ベクトルを 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。

$$a = Ai + Bj + Ck \quad (3.1)$$

また、係数を座標のように表して、

$$a = (A \ B \ C) \quad (3.2)$$

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

Average, [2](#)

Dispersion, [4](#)

Maximum, [4](#)

Mean, [2](#)

Median, [2](#), [3](#)

Minimum, [4](#)

Mode, [2](#), [4](#)

Range, [5](#)

Standard deviation, [4](#)

Variance, [4](#)

中央値, [2](#), [3](#)

代表値, [2](#)

分散, [4](#)

外れ値, [2](#), [4](#)

平均値, [2](#), [3](#)

散布度, [4](#)

最大値, [4](#), [5](#)

最小値, [4](#), [5](#)

最頻値, [2](#), [4](#)

標準偏差, [4](#)

算術平均, [2](#)

範囲, [5](#)