

確率と統計

統計学：

データから情報を取り出す方法論

- ・ 一変量解析：1種類のデータの性質を調べる
- ・ 二変量解析：2種類のデータの関係を調べる
- ・ 多変量解析：3種類以上のデータの関係を調べる

確率と統計

統計学：

データから情報を取り出す方法論

- ・ 記述統計：特定の集団全体のデータを記述する
- ・ 推測統計：特定の集団の一部についてのデータから全体を推測する

確率と統計

統計学：

データから情報を取り出す方法論

- ・ 量的変数：数値
- ・ 質的変数：数値以外

§1 代表値

データの

- ・ 平均値(算術平均・相加平均)
- ・ 幾何平均
- ・ 調和平均
- ・ 中央値 (メディアン)
- ・ 最頻値 (モード)

を計算する技術を学ぶ

以下で、

データの数: N

i 番目のデータの値: x_i

代表値：集団のデータを端的に象徴する値

次は太郎、次郎の模試の得点である。

どちらが成績がよいか。

太郎 330 372 314 325 349

次郎 303 397 278 340

平均：

$$\frac{1}{5}(330+372+314+325+349)=338.0$$

$$\frac{1}{4}(303+397+278+340)=329.5$$

いろいろな平均

算術平均 $\frac{A+B}{2}$ $\bar{x} = \frac{1}{N}(x_1 + x_2 + \dots + x_N)$

$$= \frac{1}{N} \sum_{i=1}^N x_i = \sum_i \frac{x_i}{N}$$

幾何平均 \sqrt{AB}

$$m_g = (x_1 x_2 \dots x_N)^{\frac{1}{N}} = \left(\prod_{i=1}^N x_i \right)^{\frac{1}{N}}$$

調和平均 $\frac{1}{\frac{1}{2} \left(\frac{1}{A} + \frac{1}{B} \right)}$ $m_h = \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{x_i} \right)^{-1}$

次郎 303 397 278 340

算術平均 $\frac{1}{4}(303+397+278+340) = 329.5$

幾何平均 $(303 \times 397 \times 278 \times 340)^{\frac{1}{4}} = 326.5$

調和平均 $\frac{1}{\frac{1}{4} \left(\frac{1}{303} + \frac{1}{397} + \frac{1}{278} + \frac{1}{340} \right)} = 323.7$

算術平均

p.21 例題3.1

次は1月間の販売台数である。平均値は？

5	2	5
3	6	4
4	0	6
7	10	7
6	7	3

データの個数を N 、 i 番目のデータを x_i と書く。

算術平均は、

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{75}{15} = 5.0$$

次郎 303 397 278 340

$$\text{算術平均} \quad \frac{1}{4}(303+397+278+340) = 329.5$$

$$\text{幾何平均} \quad (303 \times 397 \times 278 \times 340)^{\frac{1}{4}} = 326.5$$

$$\text{調和平均} \quad \frac{1}{\frac{1}{4} \left(\frac{1}{303} + \frac{1}{397} + \frac{1}{278} + \frac{1}{340} \right)} = 323.7$$

幾何平均

p.23 例題3.2

次は運賃の値上げ率である。平均値は？

	運賃(円)	値上げ率
-	70	
第1回改定	90	1.286
第2回改定	120	1.333
第3回改定	140	1.167

$$70 \times 1.286 \times 1.333 \times 1.167 = 140.000$$

$$\text{幾何平均} : (1.286 \times 1.333 \times 1.167)^{\frac{1}{3}} = 1.260$$

$$70 \times 1.260 \times 1.260 \times 1.260 = 140.000$$

$$\text{算術平均} : \frac{1}{3}(1.286 + 1.333 + 1.167) = 1.262$$

$$70 \times 1.262 \times 1.262 \times 1.262 = 140.662$$

次郎 303 397 278 340

$$\text{幾何平均} \quad m_g = (303 \times 397 \times 278 \times 340)^{\frac{1}{4}} = 326.5$$

$$= (3.03 \times 3.97 \times 2.78 \times 3.40)^{\frac{1}{4}} \times 100$$

$$\log_a m_g = \frac{1}{4} (\log_a 303 + \log_a 397 + \log_a 278 + \log_a 340)$$

$$\Rightarrow m_g = a^{\log_a m_g} = a^{\frac{1}{N} \sum_{i=1}^N \log_a x_i}$$

調和平均

p.25 例題3.3

次は時速の変化である。平均値は？

	時速(km/h)	所要時間(h)
最初の60km	30	2.0
中間の60km	15	4.0
最後の60km	20	3.0
	合計	9.0

調和平均：
$$\frac{1}{\frac{1}{3}\left(\frac{1}{30} + \frac{1}{15} + \frac{1}{20}\right)} [\text{km/h}] = 20.0 [\text{km/h}]$$

所要時間の合計×平均時速：
 $9.0 [\text{h}] \times 20.0 [\text{km/h}] = 180 [\text{km}]$

調和平均

p.25 例題3.3

次は時速の変化である。平均値は？

	時速(km/h)	所要時間(h)
最初の60km	30	2.0
中間の60km	15	4.0
最後の60km	20	3.0
	合計	9.0

算術平均： $\frac{1}{3}(30+15+20) [\text{km/h}] = 21.7 [\text{km/h}]$

所要時間の合計×平均時速：
 $9.0 [\text{h}] \times 21.7 [\text{km/h}] = 195 [\text{km}]$

中央値(メディアン)

太郎 330 372 314 325 349
次郎 303 397 278 340

大きさ順に並べる

太郎 314 325 **330** 349 372
次郎 278 **303** **340** 397

2つの平均 321.5

中央値(メディアン)

p.27 例題3.4

次は学生の欠席日数である。中央値は？

4	2
2	3
5	3
3	4
5	4
4	5
6	5
5	5
3	6
40	40

中央値は4.5日

最頻値(モード)

4
2
5
3
5
4
6
5
3
40

欠席日数	度数
2	1
3	2
4	2
5	3
6	1
40	1
合計	10

最頻値 は5日

課題

次のデータの算術平均、幾何平均、調和平均、メディアンを求めよ。

50 55 60 65
40 50 70 80

1から10までの自然数の算術平均、幾何平均、調和平均、メディアンを求めよ。

まとめ

算術平均 $\bar{x} = \frac{1}{N}(x_1 + x_2 + \dots + x_N) = \frac{1}{N} \sum_{i=1}^N x_i$

幾何平均 $m_g = (x_1 \dots x_N)^{\frac{1}{N}} = \left(\prod_{i=1}^N x_i \right)^{\frac{1}{N}} = a^{\frac{1}{N} \sum_{i=1}^N \log_a x_i}$

調和平均 $m_h = \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{x_i} \right)^{-1}$

中央値(メディアン) 大きさ順に並べた中央の値

最頻値(モード) 最も頻繁に表れる値