

数理・データサイエンス教育強化拠点コンソーシアム
関東・首都圏ブロックワークショップ
2021 年 6 月 24 日

模擬講義 1 「データサイエンス基礎」

寒野 善博

数理・情報教育研究センター
東京大学

模擬講義では、応用基礎レベル教材 1.4 節 (p.66-69)

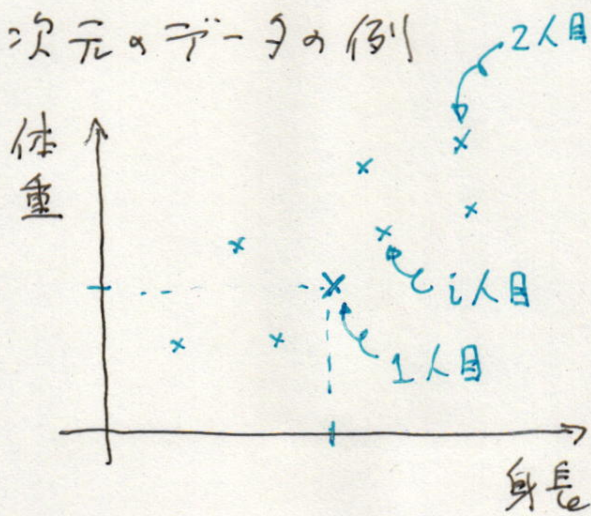
http://www.mi.u-tokyo.ac.jp/pdf/1-4_data_analysis.pdf

について書画カメラを用いて、説明した。

以下は、書画カメラに書かれたノートである。

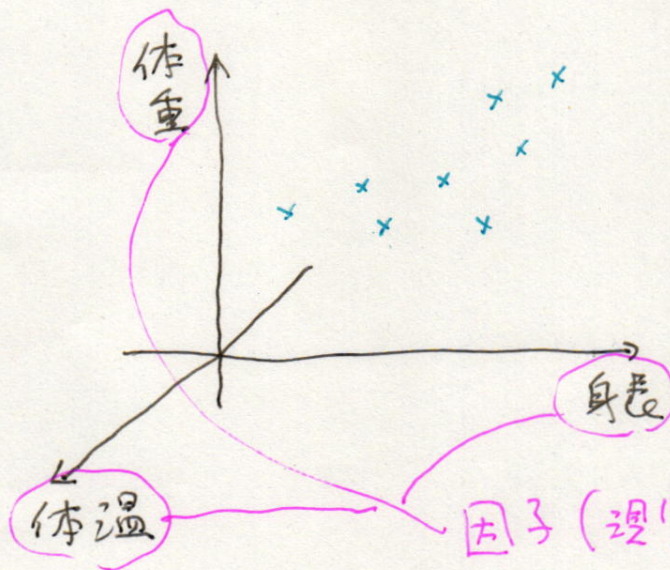
データの次元

2次元データの例



m 人分のデータ
||
(2次元)平面内の
 m 個の点の集合

3次元の場合は



3次元空間内の
 m 個の点の集合

因子(変数したもの)が3つ

p. 66 の例: 7 因子 z , $m = 100$ 人

	身長	体重		血糖値
1 人目	○	○		○
2 "	⋮	⋮		⋮
⋮				
i "	○	○		○
⋮	⋮	⋮		⋮
100 "	○	○		○

← 7次元
空間の
中の点

7 個ある

↓
列の数 = データの次元

図示は
できない。

7×100 人分

↓

2×100 人分

データの低次元への変換

簡単な例

①

3次元のデータ

	因子 1	2	3
1人目	250	71	69
2	106	158	162
3	315	243	239
4	423	325	330

↑ ↑
どちらか一方だけ
よさそう → 2次元

2次元

②

	1	2
1	250	97
2	106	101
3	315	102
4	423	98

↑
たゞ2もよさそう
→ 1次元