

2-1-1.データの種類

2. データリテラシー 2-1. データを読む

ご利用・改変等ご自由です。必要でしたらPPTファイルも差し上げます。ただし国のご支援で作成したものですので、もし講義等でご利用でしたら、利用実績を把握するために、右記までご一報ください (uchida@ait.kyushu-u.ac.jp) 2020.4 内田誠一@九州大学

この項目で学ぶこと

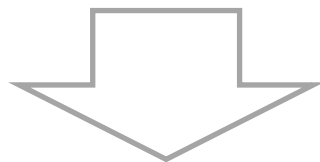
- データには「質的データ」と「量的データ」の2種類がある
 - それぞれさらに2つに分けられ、合計4種類
- 種類によって、データの見方や処理の仕方が違う

この項目を学ぶ意義

- 同じ数字に見えても、データとして意味が違うものがあることがわかる
- 自分のデータに対して、「できる処理・できない処理」があることがわかる
 - データの種類や処理に敏感になれる！

「データ」とは？ (デジタル大辞泉より引用)

- 物事の推論の基礎となる事実。また、参考となる資料・情報。「ーを集める」「確実なー」
- コンピューターで、プログラムを使った処理の対象となる記号化・**数字化**された資料。



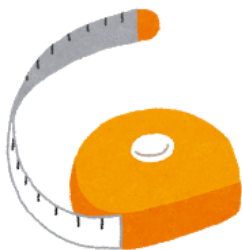
とりあえず、**データ = 数字**と考えてみよう

同じ数字データ「2」を考える

全部同じ数字 “2”

2

cm



2

年度



2

レベル

1:非常に悪い
2:悪い
3:普通
4:よい
5:非常によい

2

番系統
(バス路線)



トライ1:「どちらが大きい?」のように比べられるか?

より長い

4

より未来

4

よりよい

4

比較不能

4

2

2

2

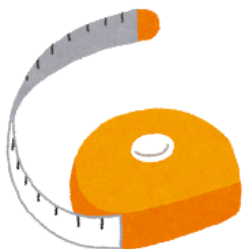
2

cm

年度

レベル

番系統
(バス路線)



- 1:非常に悪い
- 2:悪い
- 3:普通
- 4:よい
- 5:非常によい



トライ2: 足せるか？

2cmに1cmを
足したら3cm

2年度に1年を
足したら3年度

2(悪い)に
1(非常に悪い)を
足したら
3(普通)

2番(西回り)に
1番(直通)を
足したら
3番(東回り)

2

2

2

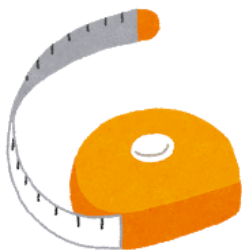
2

cm

年度

レベル

番系統
(バス路線)



1:非常に悪い
2:悪い
3:普通
4:よい
5:非常によい

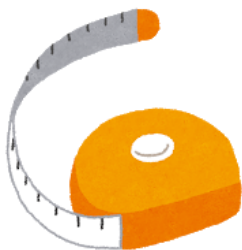


トライ3: 2倍できるか？

2cmを2倍したら4cm

2

cm



2年度を2倍したら4年度



2

年度



2(悪い)を2倍したら4(よい)



2

レベル

- 1:非常に悪い
- 2:悪い
- 3:普通
- 4:よい
- 5:非常によい

2番(西回り)を2倍したら4番(急行)



2

番系統
(バス路線)

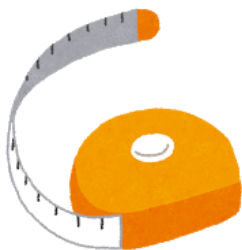


同じ数字データ「2」なのに、**できることが違う!?**
→ なぜならば**データの種類が違う**から!

全部同じ数字“2”だが、データの種類が違う

2

cm



2

年度



2

レベル

1:非常に悪い
2:悪い
3:普通
4:よい
5:非常によい

2

番系統
(バス路線)



データの種類

比率
データ

2

cm

間隔
データ

2

年度

順位
データ

2

レベル

カテゴリ
データ

2

番系統
(バス路線)



本当に「数値」

(数値でないと扱えないデータ)

みせかけの「数値」

(単に区別するための番号付け)

カテゴリデータ

- 他と区別するために形式的に数字になっているだけ
 - = 数字そのものに意味はない
- 四則演算（加減乗除）すべて意味がない
 - 「同じものか違うものか」の判断だけが可能
 - 3番バスと5番バスは違うが、3番バスと3番バスは同じ
- 例
 - 電話番号
 - 背番号
 - バスの系統番号
 - 「1:女性, 2:男性」



順位データ

- 並べるためにつけられた数字
 - 3位は5位より高順位, 1位は3位より高順位
 - ただし, 「3位と5位の差」=「1位と3位の差」とは限らない
- 四則演算（加減乗除）すべて意味がない
 - 並べることだけは可能
- 例
 - アンケート結果
 - 成績順位

1:非常に悪い
2:悪い
3:普通
4:よい
5:非常によい

間隔データ

- 差に意味があるデータ
 - $3\text{年度} - 1\text{年度} = 5\text{年度} - 3\text{年度} = 2\text{年度}$
 - 差ができるので和もできる
 - 数直線上には載せられる
- しかし、積や除算はできない
 - 乗除ができないので、割合にも意味がない。
- 例
 - (華氏・摂氏で測る)温度
 - 「温度が10度下がる」とは言うが、「温度が10%下がる」とは言わない
 - 西暦年，日付



比率データ

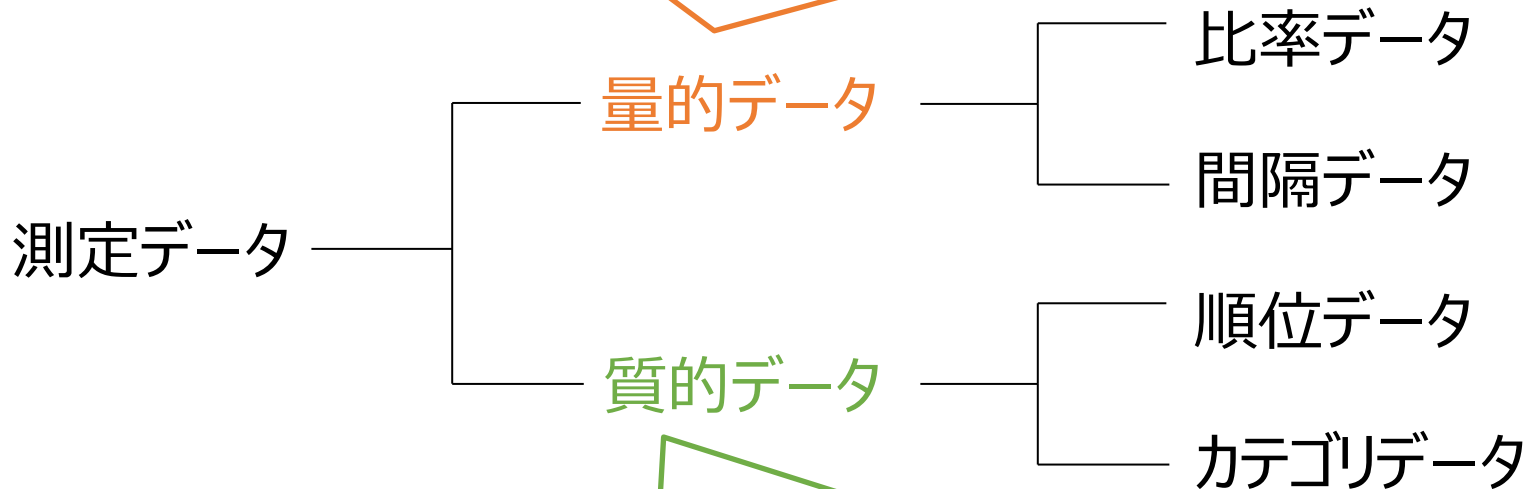
- 和・差に加え，積や除算も可能!
 - 「2倍多い」「10%減少」などの比率計算も可能

- 例
 - 体重，長さなど，多くの物理量
 - 年齢
 - 収入
 - 温度（絶対零度を0とした場合）



量的データと質的データ

数値で表すことが適切なデータ



内容を区別する目的で
数値を使っているデータ

データの種類によって使える手法が 大きく異なってくる

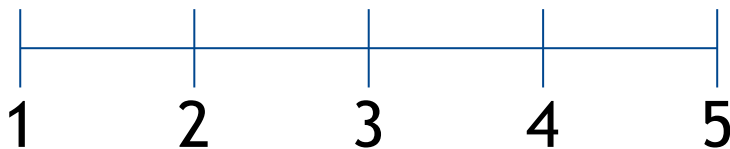
	名称	可能な演算	主な代表値	主な事例
量的 データ	比率データ	+ - × ÷	各種平均	質量, 長さ, 年齢, 時間, 金額
	間隔データ	+ -	算術平均	温度 (摂氏) , 知能指数
質的 データ	順位データ	> =	中央値, 最頻値	満足度, 選好度, 硬度
	カテゴリ データ	度数カウント	最頻値	電話番号, 性別, 血液型

例：平均 = データを全部足して，データの個数で割る
→ 質的データは平均が出せない

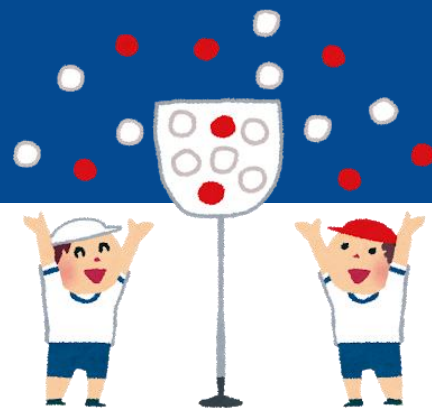
参考：アンケート結果の平均

1:非常に悪い
2:悪い
3:普通
4:よい
5:非常によい

- 5段階アンケートの結果, 「5, 4, 2, 4, 3, 1, 1, 3」
 - 平均 $23/8 \div 2.9$ だから, みんな「普通」ぐらい??
 - ↑これは間違いか?
- 厳密に言えば間違い
 - 「5と4の差」と「4と3の差」が同じとは限らないので...
- 対策
 - 数直線を準備して, 回答者に等間隔性を意識させる
 - →間隔データに近くできる→平均が計算できる



参考：個数は何データ？



- 比率データです！
- 「3番のバス」の3はカテゴリデータですが、「3番のバスが10台通った」の10は比率データ
- 「5段階アンケート」の選択肢1,2,3,4,5は順序データですが、「5と答えた人は20人」の20は比率データ

まとめ・参考資料

- データには種類がある
- 種類によって扱える計算が異なる
 - 例：バスの番号の平均値を計算しても無意味！
- 自分の扱っているデータがどの種類のデータなのか、データ分析の前に、まず理解しよう