

# 情報科学 【AI・データサイエンス】

## 第7回 アンケート解析

アンケート調査の形式  
アンケート結果の集計

# はじめに

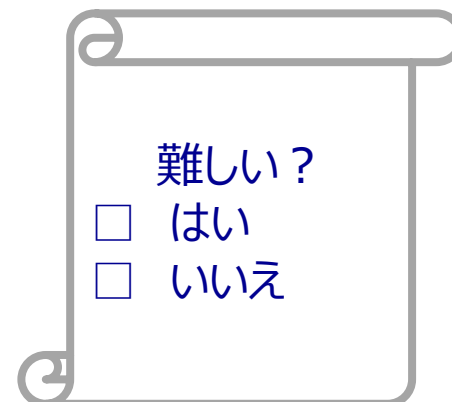
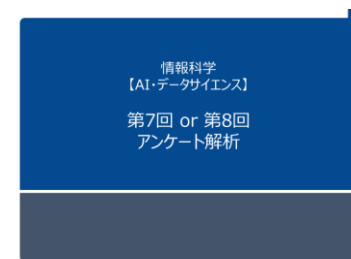
- 「アンケートって、データ？」という方
  - 立派なデータ！：人から得られる基本的なデータの一つ
- あらゆる分野で頻繁に利用されているデータ取得方法



# アンケート調査

## だれに調査すればよいかは→第4回

- 質問に対する回答をデータとして収集
  - 回答の定型化が可能



# アンケート調査の形式

# アンケート調査の回答法

- 選択回答法

- 単数回答：選択肢の中から1つだけ選択
- 複数回答：選択肢の中から該当するものを複数選択

- 自由回答法

- 数量回答
  - 順位回答
  - 数値回答
- 文字回答

# 単数回答（1/3）

## 二項選択

- 2つの回答選択肢から 1 つ選んでもらう方法
- 例

質問：あなたはパソコンを持っていますか？

1. 持っている
2. 持っていない

# 単数回答（2/3） 多項選択

- 3つ以上の回答選択肢から 1 つ選んでもらう方法
- 例：

質問：この車を選んだ理由は何ですか？  
最も重要視したものを1つ選んでください

1. スタイルのよさ
2. 室内の広さ
3. 性能の良さ
4. 燃費の良さ
5. メーカーへの信頼

# 単数回答（3/3）

## 選択肢に順序がある場合も

### ● 代表例1：範囲を回答するケース

質問：あなたのこの一年間での  
税込み収入はおいくらくらいですか？  
次の中から選択してください

1. 200万円未満
2. 200～400万円未満
3. 400～600万円未満
4. 600～800万円未満
5. 800万円以上

### ● 代表例2：レベルを回答するケース

質問：あなたは現在の生活にどの程度  
満足されていますか。その程度をお知らせください

1. 非常に満足
2. やや満足
3. どちらともいえない
4. やや不満
5. 非常に不満



# 複数回答 (1/2)

## 無制限法

- 好きなだけ選んでもらう方法
- 例：質問：あなたは日ごろのストレスを解消する方法としてどのようなことを行っていますか？（○はいくつでも）

- |             |              |
|-------------|--------------|
| 1. スポーツをする  | 8. 友人と過ごす    |
| 2. 趣味を楽しむ   | 9. 恋人と過ごす    |
| 3. 酒を飲む     | 10. 家族と団らんする |
| 4. 買い物をする   | 11. 特に何もしない  |
| 5. 旅行をする    |              |
| 6. ドリンク剤を飲む |              |
| 7. 家でごろごろする |              |

# 複数回答 (2/2)

## 制限法

- 指定した個数だけ選んでもらう方法
- 例：質問：あなたの好きな食べものを3つ選んでください  
(○は3つまで)

- |          |            |
|----------|------------|
| 1. すきやき  | 7. お寿司     |
| 2. とんかつ  | 8. おでん     |
| 3. 餃子    | 9. うなぎ     |
| 4. ハンバーグ | 10. 天ぷら    |
| 5. オムライス | 11. さしみ    |
| 6. コロッケ  | 12. カレーライス |

# 自由回答法

- 数量回答

- 質問に対して**数値**で自由に答えてもらう方法

- 例：

質問：あなたのこの一年の税込み年収はおいくらですか？  
(                      万円)

- 文字回答

- 質問に対して**言葉**で自由に答えてもらう方法

- 例：

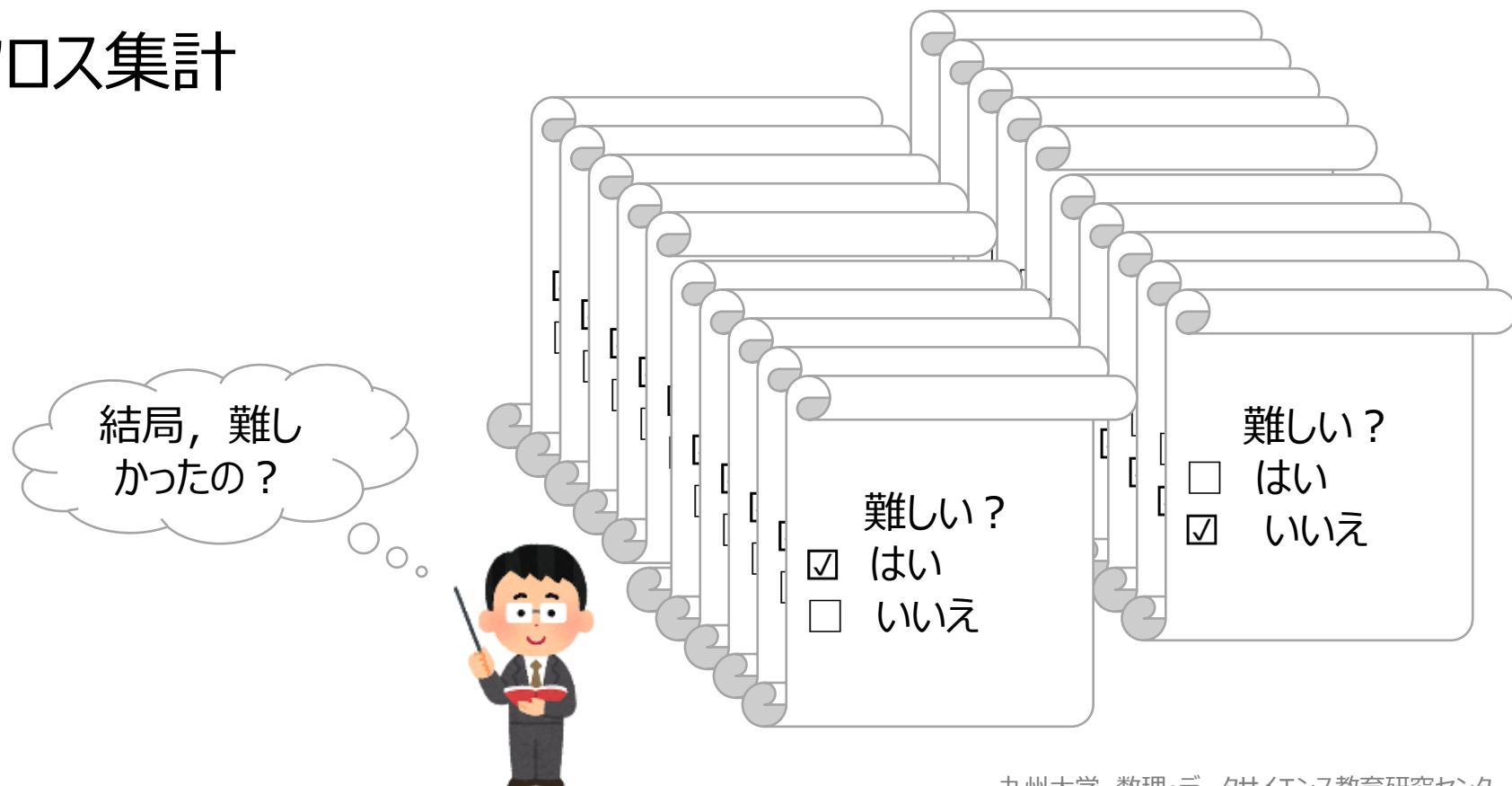
質問：あなたが今年最も活躍したと思われるタレントは誰ですか？  
(                      )

# 調査結果の集計

# 調査集計

アンケート調査の個人情報を集団情報に変換する手段

- 単純集計
- クロス集計



# 数量データとカテゴリデータ

## データの一般的な4分類 (1/2)

### ● 数量データ

- データ間の大小関係を比較したり、演算を行ったときに意味のある数値となるデータ
- 例：年収データ

#### ● 量的データ

##### ● 比率データ

- 積や除算ができる。和や差もできる。Ex. 体重、年収、長さ

「温度が2倍」「温度70%減」とは言わない

##### ● 間隔データ

- 積や除算に意味がない。ただし和や差はできる。Ex. (華氏・摂氏で測る)温度、西暦年

#### ● 質的データ

##### ● 順位データ

- 四則演算（加減乗除）すべて意味がない。ただし並べることができる。
- Ex. アンケート結果（5:非常によい, 4:よい, 3:ふつう, 2:わるい, 1:非常に悪い）。成績順

「非常によい〜ふつう=わるい」とはならない

##### ● カテゴリデータ

- 形式的に数値になっているだけ。
- Ex. 「1:女性, 2:男性」, 電話番号, 背番号, バスの系統番号

九州大学 数理・データサイエンス教育研究センター

### ● カテゴリデータ

- 形式的に数値になっているだけ
- データ間の大小比較や演算をしても無意味
- 例：「1:カレー, 2:うどん, 3:親子丼, 4:ジロー風スパ」

※ジロー風スパは九大学食伝説のメニューです

九州大学 数理・データサイエンス教育研究センター

# カテゴリデータの単純集計

- 選択肢に回答した人数を数えること
  - 度数分布表ができる

1:カレー	2:うどん	3:親子丼	4:ジロー風スパ
19	28	6	153

- 数えた人数の回答全体に占める割合(%)にしてもよい

1:カレー	2:うどん	3:親子丼	4:ジロー風スパ
9.2	13.6	2.9	74.3

# 数量データの単純集計

- 平均値, 中央値, 標準偏差などの**基本統計量**を計算
  - 基本統計量とは? → データの基本的な特性を表す値
  - 基本統計量の例
    - 分布全体をひとつの数で表す代表値
      - 平均値, 中央値, 最頻値 (後述)
    - データのばらつきを表す値
      - 最小・最大値, 分散・標準偏差
- など



# 平均値とは： データの合計をデータの個数で割ったもの

- データは平均を中心として分布している
- しかし「はずれ値」に弱いという問題も

クラス1		クラス2	
氏名	年齢	氏名	年齢
A	19	M	18
B	18	N	19
C	19	O	22
D	19	P	18
E	20	Q	18
F	19	R	19
G	18	S	21
H	19	T	19
I	20	U	19
J	19	V	18
K	19	W	66
L	19	X	19
平均値	19		23

一人だけ妙に高年齢  
(はずれ値)

はずれ値の影響を受けて  
しまった平均  
(23歳以上は一人だけなのに...)

Excelの関数：  
AVERAGE(数値1, 数値2, )

- アンケートでも, こうした「はずれ値」が頻繁に発生
  - 特に「不真面目な回答者」ははずれ値になりやすい

# 中央値とは：

データに順位をつけた時，ちょうど真ん中の順位にくる値

- 平均値と違い，はずれ値の影響を受けづらい

クラス1		クラス2	
氏名	年齢	氏名	年齢
A	19	M	18
B	18	N	19
C	19	O	22
D	19	P	18
E	20	Q	18
F	19	R	19
G	18	S	21
H	19	T	19
I	20	U	19
J	19	V	18
K	19	W	66
L	19	X	19
平均値	19		23
中央値	19		19

例：

18
18
18
18
19
19
19
19
19
21
22
66

個数が偶数のときは  
これらの平均が中央値

並び替え

Excelの関数：  
MEDIAN(数値1, 数値2, )

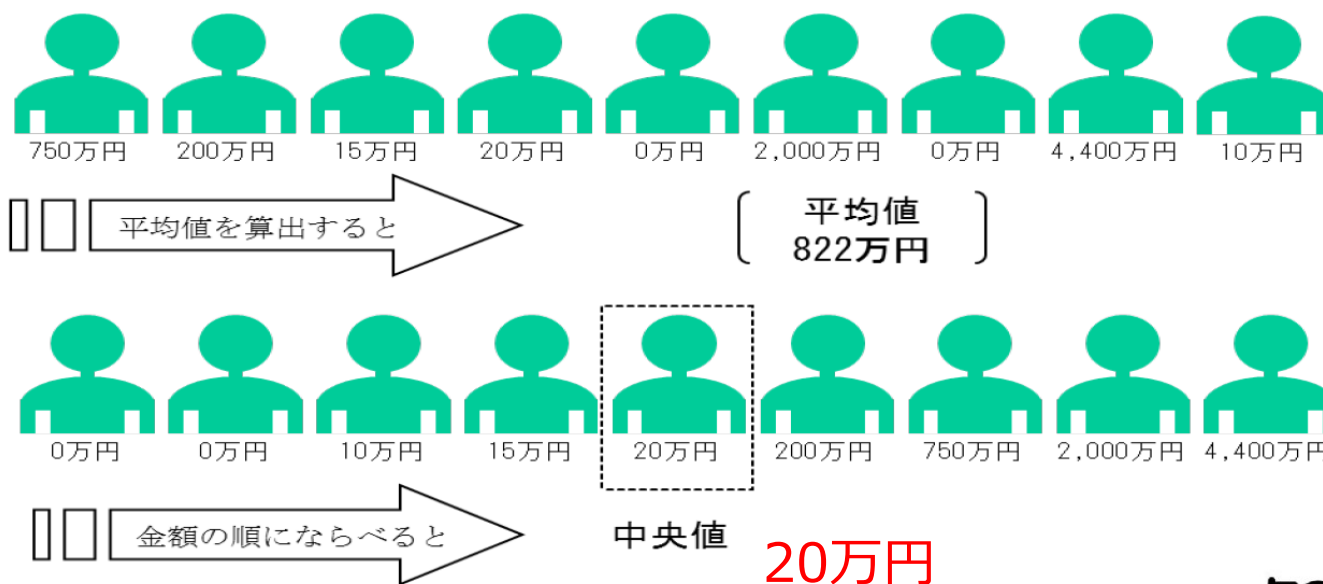
# 平均値と中央値は、時に結構違う： 単身世帯の金融資産保有額

対象：全国2,500世帯（20歳以上70歳未満で、単身世帯を構成する者）

平均値：822万円

実際は…

少数の高額保有者が  
平均を引き上げる！



出典：「家計の金融行動に関する世論調査」[単身世帯調査]（2016年）

# 分散・標準偏差とは

データのばらつきをあらわす指標

- 分散

各データの平均からのずれの2乗を合計して  
データ数で割ったもの

Excelの関数：

VAR.P(数値1, 数値2, )

- 標準偏差

分散の平方根

Excelの関数：

STDEV.P(数値1, 数値2, )

クラス1		クラス2	
氏名	年齢	氏名	年齢
A	19	M	18
B	18	N	19
C	19	O	22
D	19	P	18
E	20	Q	18
F	19	R	19
G	18	S	21
H	19	T	19
I	20	U	19
J	19	V	18
K	19	W	66
L	19	X	19
平均値			23
中央値			19
最頻値			19
最小値			18
最大値			66
分散			169.5
標準偏差			13.01922

# もう一度：度数分布表

- データ値の頻度を計算したもの
  - どんな値が，何回観測されたか
- カテゴリデータ，整数データの場合
  - 男性が何人，女性が何人
  - 18歳が何人，19歳が何人，...
- 値を区間に分ける場合
  - 0～4歳が何人，5～9歳が何人，...

Excelの関数：

COUNTIF(範囲, 数値)

「範囲」中に「数値」が何回出てくるか求めることができる

クラス1		クラス2	
氏名	年齢	氏名	年齢
A	19	M	18
B	18	N	19
C	19	O	22
D	19	P	18
E	20	Q	18
F	19	R	19
G	18	S	21
H	19	T	19
I	20	U	19
J	19	V	18
K	19	W	66
L	19	X	19
年齢	度数	年齢	度数
18	2	18	4
19	8	19	5
20	2	20	0
21	0	21	1
22	0	22	1
66	0	66	1

# クロス集計

# クロス集計とは

- 2つの質問項目をクロスして表を作成することにより、相互の関係を明らかに

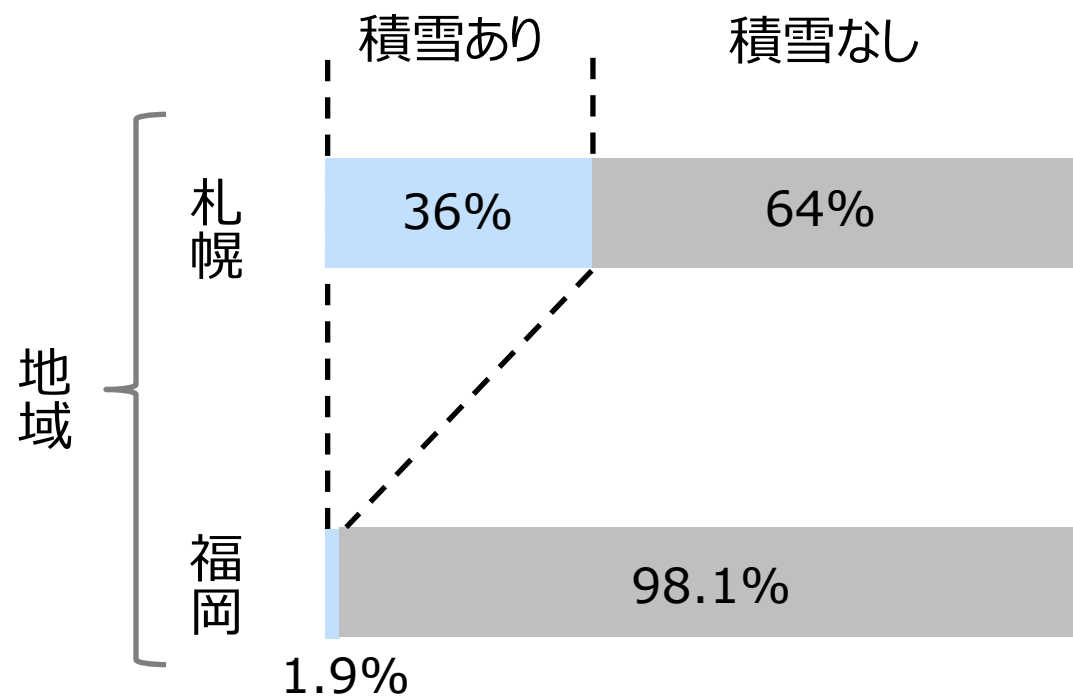
	自宅から通学	一人暮らしで通学
朝ごはん食べる	83	15
朝ごはん食べない	29	47

※数値はいい加減

- 回答が「カテゴリデータ or 数量データ」で3つに分類
  - 「カテゴリ/カテゴリ」クロス集計
  - 「カテゴリ/数量」クロス集計
  - 「数量/数量」クロス集計

# 「カテゴリ/カテゴリ」クロス集計

## ●「地域」と「積雪有無」のクロス集計



クロス集計表

	積雪あり	積雪なし
札幌	36	64
福岡	1.9	98.1

1961年から1990年までの平均値をもとに作成。  
資料：文部省国立天文台編「理科年表」



# 「カテゴリ/数量」クロス集計

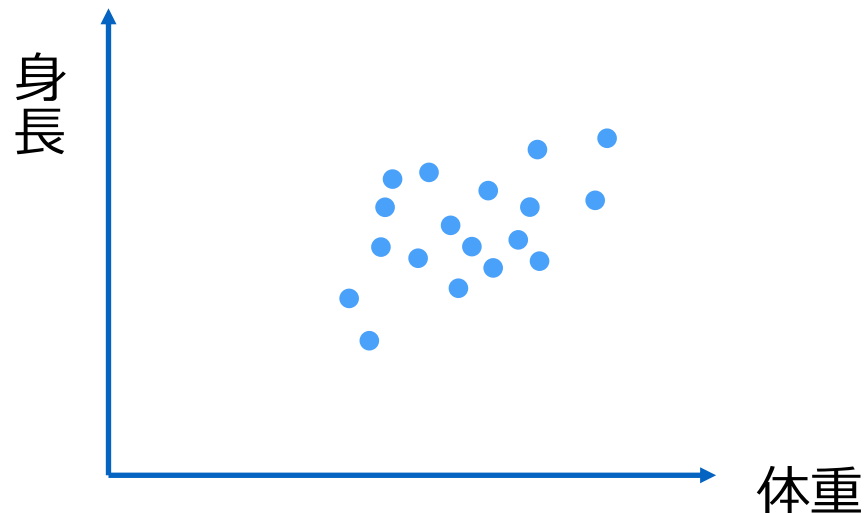
- 「果物」と「出荷量 (t)」のクロス集計



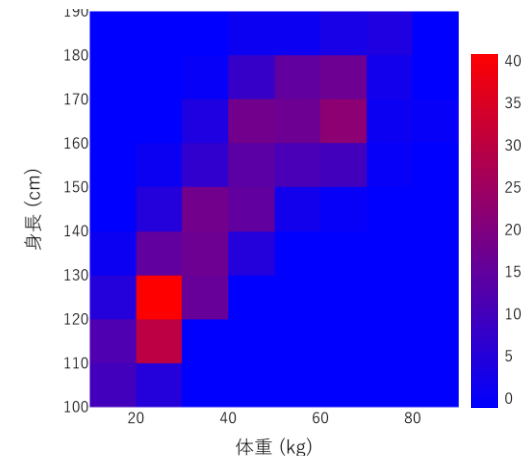
出典: 政府 作物統計調査

# 「数量/数量」クロス集計

- 散布図が便利

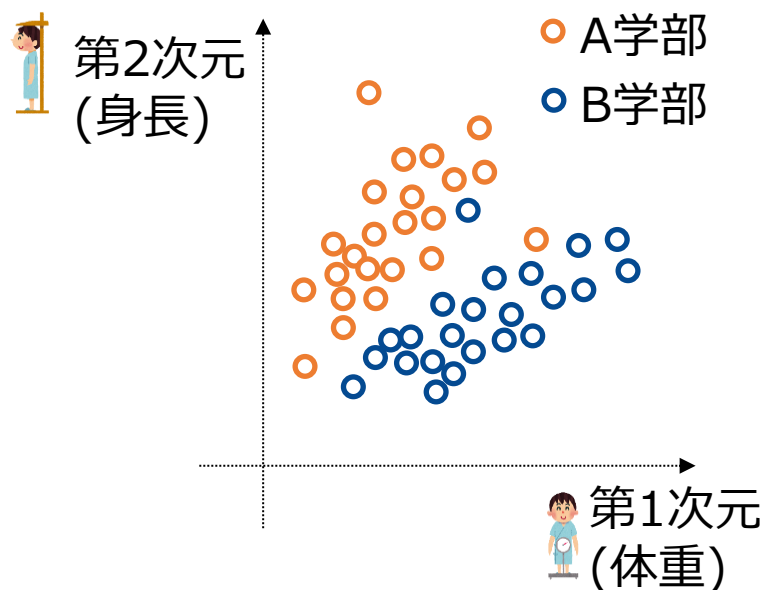


- 例えば体重を10kg, 身長を10cmごとに区切って集計すれば, 2次元ヒストグラム(ヒートマップ)になる
  - これもクロス集計表の一種と見なせる

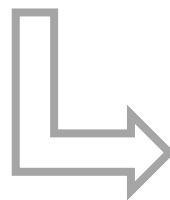


# アンケート調査からデータ分析へ

- 得られたデータを分析し，知見を与える



- A学部とB学部のデータの広がり方はどのくらい違うのか？
- 身長が増加と体重の増加に相関はある？



定量的な評価が必要⇒**数学**の出番

- ◆ 統計（平均，分散，相関...）
- ◆ ベクトル（類似度，距離計算）
- ◆ 行列（固有値分解）

# まとめ

- アンケート調査におけるデータ
  - アンケート調査の回答方法
  - アンケートデータの分類
- 調査集計
  - 単純集計
  - クロス集計