

# 多変量解析 〈初歩の初歩〉に入門(1)補遺

## データコーディングの形式

なかもとけいこ  
kenakamoto@nifty.com

## 測定されるデータには水準がある

- 復習
  - 名義尺度
  - 順序尺度
  - 間隔尺度
  - 比率尺度
- 比率尺度、間隔尺度は基本的に普通の数字としてデータファイル(エクスセル)に入力してよい。
  - 足したり引いたり(二つの値の差を)かけたり割ったりできるから。
- 順序尺度と名義尺度はそうはいかない。
  - 足したり引いたりかけたり割ったりできないから。
  - 特に名義尺度の入力方法(データコーディングスキーム)には工夫が必要。

## まず見極めること

- その尺度は、排他的カテゴリーになっているのか(すなわち択一式か)、非排他的カテゴリーになっているのか。

## 架空のデータ例(択一式)

- 果物、野菜の中からそれぞれ好きなものを一つ選ぶ。

ひとりめ

<input type="checkbox"/>	りんご
<input checked="" type="checkbox"/>	みかん
<input type="checkbox"/>	イチゴ

<input type="checkbox"/>	なすび
<input type="checkbox"/>	トマト
<input checked="" type="checkbox"/>	ズッキーニ

ふたりめ

<input type="checkbox"/>	りんご
<input type="checkbox"/>	みかん
<input checked="" type="checkbox"/>	イチゴ

<input type="checkbox"/>	なすび
<input checked="" type="checkbox"/>	トマト
<input type="checkbox"/>	ズッキーニ

さんにんめ...

## コーディング方法(1)

りんご	→	1	被験者		①	②
みかん	→	2	1	ひとりめ	2	6
イチゴ	→	3	2	ふたりめ	3	5
なすび	→	4	3	さんにんめ	.	.
トマト	→	5	4	...	.	.
ズッキーニ	→	6				

## コーディング方法(2)

変数 <果物>

りんご	→	1
みかん	→	2
イチゴ	→	3

変数 <野菜>

なすび	→	1
トマト	→	2
ズッキーニ	→	3

被験者	果物	野菜	
1	ひとりめ	2	3
2	ふたりめ	3	2
3	さんにんめ	.	.
4	...	.	.



## たとえば,

- ケースの類似性を定義するには,

被験者	果物	野菜
ひとりめ	2	3
ふたりめ	3	2
さんりんめ	...	...
...	...	...

単に差の絶対値をとったり、ベクトル表現で距離をとると誤分からなくなる。

被験者	果物			果物		
	りんご	みかん	イチゴ	なすび	トマト	ズッキーニ
ひとりめ	0	1	0	0	0	1
ふたりめ	0	0	1	0	1	0
さんりんめ	...	...	...	...	...	...
...	...	...	...	...	...	...

差分をとっても割と自然に(非)類似性を定義できる

被験者	果物			果物		
	りんご	みかん	イチゴ	なすび	トマト	ズッキーニ
	Y	N	Y	N	Y	N
ひとりめ	0	1	1	0	1	1
ふたりめ	1	0	0	1	1	0
さんりんめ	...	...	...	...	...	...
...	...	...	...	...	...	...

## ということで,

- コーディングデータをどういう風に入力するかは、後の作業効率に影響します。
- おらかじめ矢を見越しての入力をおすすめします。