

問題解決技法入門

3. Data Analysis

1. Cross Tabulation

堀田 敬介

クロス集計とは

- クロス集計(表) cross tabulation

- 2つ以上の**属性間の関係**を知りたい時に使う集計方法のひとつ。分割表ともよぶ

元データ

id	性別	年齢	嗜好1	嗜好2
1	女性	34	猫	紅茶
2	女性	21	犬	紅茶
3	男性	29	猫	紅茶
4	男性	69	猫	珈琲
5	女性	38	猫	紅茶
6	男性	64	猫	紅茶
7	男性	38	犬	珈琲
8	女性	37	猫	珈琲
9	男性	16	犬	珈琲
10	女性	25	犬	珈琲
11	女性	21	犬	紅茶
12	女性	17	猫	紅茶
13	男性	20	猫	珈琲
14	男性	16	犬	珈琲
15	女性	18	犬	紅茶

⋮

加工データ

ここを加工した

id	性別	年代	嗜好1	嗜好2
1	女性	30	猫	紅茶
2	女性	20	犬	紅茶
3	男性	20	猫	紅茶
4	男性	60	猫	珈琲
5	女性	30	猫	紅茶
6	男性	60	猫	紅茶
7	男性	30	犬	珈琲
8	女性	30	猫	珈琲
9	男性	10	犬	珈琲
10	女性	20	犬	珈琲
11	女性	20	犬	紅茶
12	女性	10	猫	紅茶
13	男性	20	猫	珈琲
14	男性	10	犬	珈琲
15	女性	10	犬	紅茶

⋮

クロス集計(例1)

「年代」と「嗜好1」の人数をクロス集計

	列ラベル		
行ラベル	犬	猫	総計
10	13	7	20
20	16	16	32
30	16	23	39
40	16	25	41
50	13	16	29
60	19	17	36
70	3		3
総計	96	104	200

クロス集計(例2)

「性別」と「嗜好2」の人数をクロス集計

	列ラベル		
行ラベル	紅茶	珈琲	総計
女性	53	41	94
男性	57	49	106
総計	110	90	200

クロス集計の前に: フィルタを使おう

- 集計したいデータ項目を選択①し [データ②]-[フィルタ③]

The screenshot shows the Microsoft Excel interface. The 'Data' ribbon is highlighted with a red circle and the number '2'. The 'Filter' button in the 'Data' ribbon is also circled in red with the number '3'. The first row of the data table is circled in red with the number '1'.

	A	B	C	D	E	F	G	H	I	J	K
1											
2		id	性別	年代	嗜好1	嗜好2					
3		1	女性	30	猫	紅茶					
4		2	女性	20	犬	紅茶					
5		3	男性	20	猫	紅茶					
6		4	男性	60	猫	珈琲					
7		5	女性	30	猫	紅茶					
8		6	男性	60	猫	紅茶					

フィルター (Ctrl+Shift+L)

選択したセルにフィルターを適用します。

列見出しの矢印をクリックして、データを絞り込みます。

[詳細情報](#)

クロス集計の前に: フィルタを使おう

- フィルタをかけ, 欲しいデータだけを抽出
 - 例: 「犬」好きで「紅茶」が好きな「女性」を抽出

フィルタで選択

	A	B	C	D	E	F
1						
2		id	性別	年	嗜好	嗜好
4		2	女性	20	犬	紅茶
13		11	女性	20	犬	紅茶
17		15	女性	10	犬	紅茶
28		26	女性	40	犬	紅茶
29		27	女性	30	犬	紅茶
31		29	女性	60	犬	紅茶
32		30	女性	60	犬	紅茶
38		36	女性	50	犬	紅茶
41		39	女性	20	犬	紅茶
53		51	女性	40	犬	紅茶
54		52	女性	50	犬	紅茶
55		53	女性	40	犬	紅茶
61		59	女性	60	犬	紅茶
66		64	女性	60	犬	紅茶
74		72	女性	60	犬	紅茶
82		80	女性	20	犬	紅茶
83		81	女性	10	犬	紅茶
92		90	女性	60	犬	紅茶
106		104	女性	30	犬	紅茶
112		110	女性	40	犬	紅茶
113		111	女性	10	犬	紅茶
117		115	女性	50	犬	紅茶
118		116	女性	10	犬	紅茶
119		117	女性	50	犬	紅茶
162		160	女性	20	犬	紅茶
171		169	女性	20	犬	紅茶
186		184	女性	30	犬	紅茶
197		195	女性	10	犬	紅茶
203						

データが**選択(抽出)**されたものだけだとわかるように, 行番号が「**青色**」になっている

Excelでクロス集計

- 集計したい範囲を選択①し [挿入②]-[ピボットテーブル③]

自動保存 ● わ

ファイル ホーム② 挿入② ページレイアウト 数式 データ 校閲 表示 アドイン ヘルプ ATOK拡張ツ

③ **ピボットテーブル** おすすめ テーブル

テーブル

ピボットテーブルの作成

分析するデータを選択してください。

テーブルまたは範囲を選択(S)

テーブル/範囲(T): data1!\$B\$2:\$F\$20

外部データソースを使用(U)

接続の選択(C)...

接続名:

このブックのデータモデルを使用する(D)

ピボットテーブル レポートを配置する場所を選択してください。

新規ワークシート(N)

既存のワークシート(E)

場所(L):

複数のテーブルを分析するかどうかを選択

このデータをデータモデルに追加する(M)

OK キャンセル

	A	B	C	D	E	F
1						
2		id	性別	年代	嗜好1	嗜好2
3		1	女性	30	猫	紅茶
4		2	女性	20	犬	紅茶
5		3	男性	20	猫	紅茶
6		4	男性	60	猫	珈琲
7		5	女性	30	猫	紅茶
8		6	男性	60	猫	紅茶
9		7	男性	30	犬	珈琲
10		8	女性	30	猫	珈琲
11		9	男性	10	犬	珈琲
12		10	女性	20	犬	珈琲
13		11	女性	20	犬	紅茶
14		12	女性	10	猫	紅茶
15		13	男性	20	猫	珈琲
16		14	男性	10	犬	珈琲
17		15	女性	10	犬	紅茶
18		16	女性	20	猫	紅茶
19		17	男性	20	犬	珈琲
20		18	女性	30	猫	珈琲

[OK④]

Excelでクロス集計

• [ピボットテーブルのフィールド]

上から下へ
drag & drop

削除・修正したい場合は
drag & drop
で戻す

- 上半分にデータの「項目(属性, フィールド)」名が並んでいる
- 下半分の「行」「列」「値」の最低3つを指定
- 「行」「列」にクロスさせたい項目を, 「値」に集計したい項目を, それぞれ該当の場所に**ドラッグ&ドロップ** →クロス集計表がシート内に完成
- 修正・編集も同様(**ドラッグ&ドロップ**)

例)左の設定でできたクロス集計表

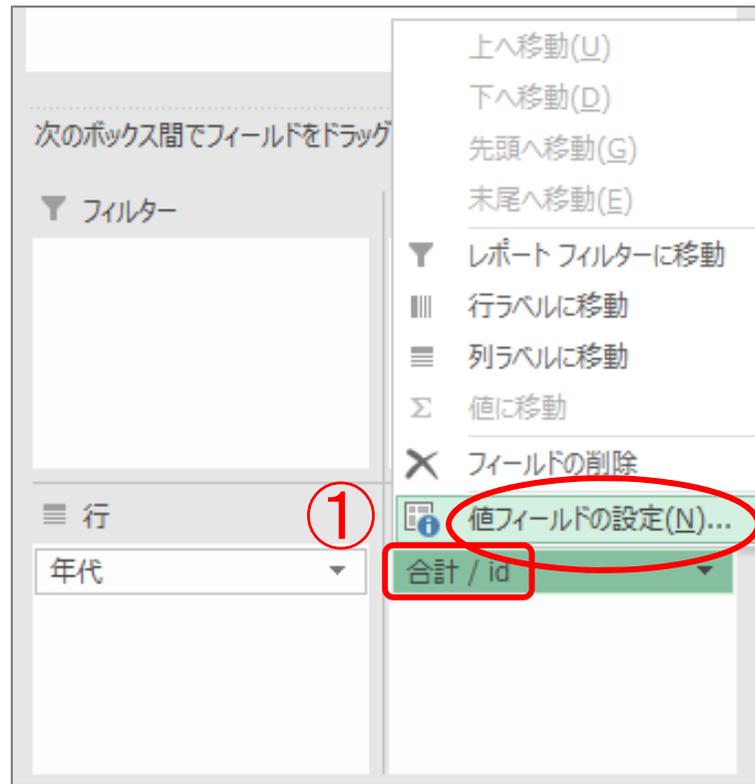
「行」=「年代」
「列」=「嗜好1」
「値」=「id」

これだと集計値がおかしい。「値」が[合計/id]なので修正する

	A	B	C	D
1				
2				
3	合計 / id	列ラベル		
4	行ラベル	犬	猫	総計
5	10	1284	727	2011
6	20	1330	1120	2450
7	30	1312	2367	3679
8	40	1780	2930	4710
9	50	1040	2065	3105
10	60	1908	1837	3745
11	70	400		400
12	総計	9054	11046	20100
13				
14				

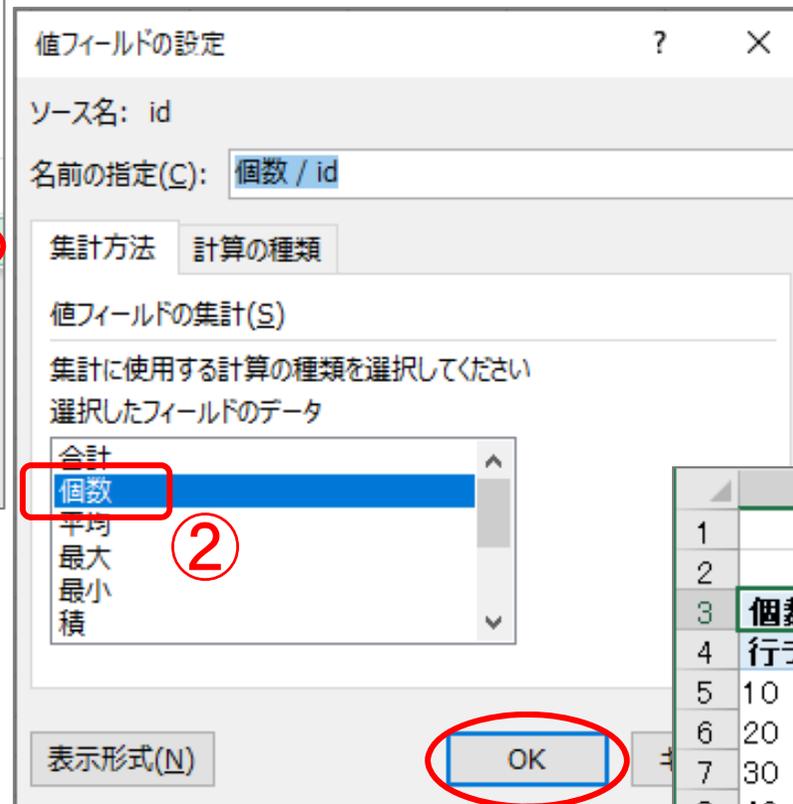
Excelでクロス集計

• [ピボットテーブルのフィールド]の修正



① [合計/id]を右クリックし、「値フィールドの設定(N)」を選択

② [個数]を選択し、[OK]クリック

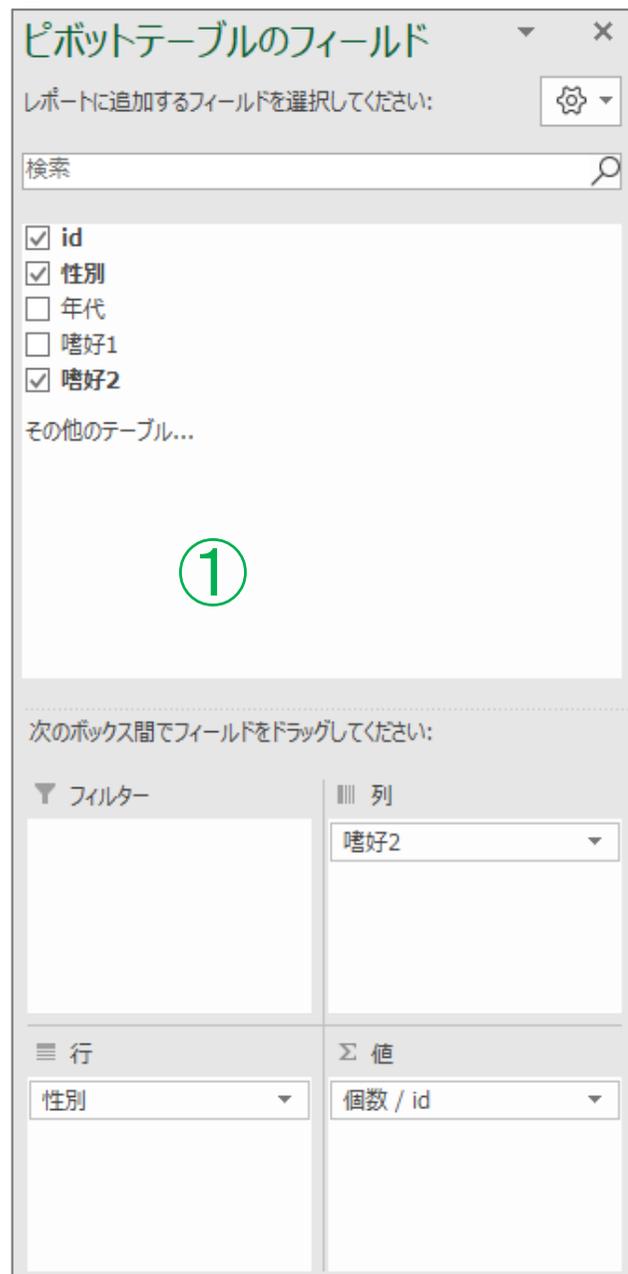


正しい(欲しい)
集計表になった

	A	B	C	D
1				
2				
3	個数 / id	列ラベル		
4	行ラベル	犬	猫	総計
5	10	13	7	20
6	20	16	16	32
7	30	16	23	39
8	40	16	25	41
9	50	13	16	29
10	60	19	17	36
11	70	3		3
12	総計	96	104	200

Excelでクロス集計

• [ピボットテーブルのフィールド]の**操作1**



① [年代]→[性別]に変更

※[年代]を下から上(①のあたり)へdrag&drop
or [年代]の☑を外す(どちらでもよい)

※[性別]を上から下([行]の中)へdrag&drop

② [嗜好1]→[嗜好2]に変更

※[嗜好1]を下から上(①のあたり)へdrag&drop
or [嗜好1]の☑を外す(どちらでもよい)

※[嗜好2]を上から下([列]の中)へdrag&drop

	A	B	C	D
1				
2				
3	個数 / id	列ラベル		
4	行ラベル	紅茶	珈琲	総計
5	女性	53	41	94
6	男性	57	49	106
7	総計	110	90	200
8				

新しい(欲しい)集計表に変わった

Excelでクロス集計

• [ピボットテーブルのフィールド]の操作2

id
 性別
 年代
 嗜好1
 嗜好2
 その他のテーブル...

次のボックス間でフィールドをドラッグしてください:

▼ フィルター	≡ 列
	嗜好1 ▼
	嗜好2 ▼
≡ 行	Σ 値
性別 ▼	個数 / id ▼



	A	B	C	D	E	F	G	H
1								
2								
3	個数 / id	列ラベル ▼						
4	行ラベル ▼	☐ 犬		犬 集計	☐ 猫		猫 集計	総計
5		☐ 紅茶	珈琲		紅茶	珈琲		
6	女性		28 18	46	25 23		48	94
7	男性		30 20	50	27 29		56	106
8	総計		58 38	96	52 52		104	200

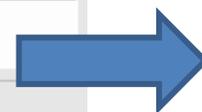
「行」=「性別」
 「列」=「嗜好1」「嗜好2」
 「値」=「id(個数)」

id
 性別
 年代
 嗜好1
 嗜好2
 その他のテーブル...

次のボックス間でフィールドをドラッグしてください:

▼ フィルター	≡ 列
	嗜好1 ▼
≡ 行	Σ 値
性別 ▼	個数 / id ▼
年代 ▼	

「行」=「性別」「年代」
 「列」=「嗜好1」
 「値」=「id(個数)」

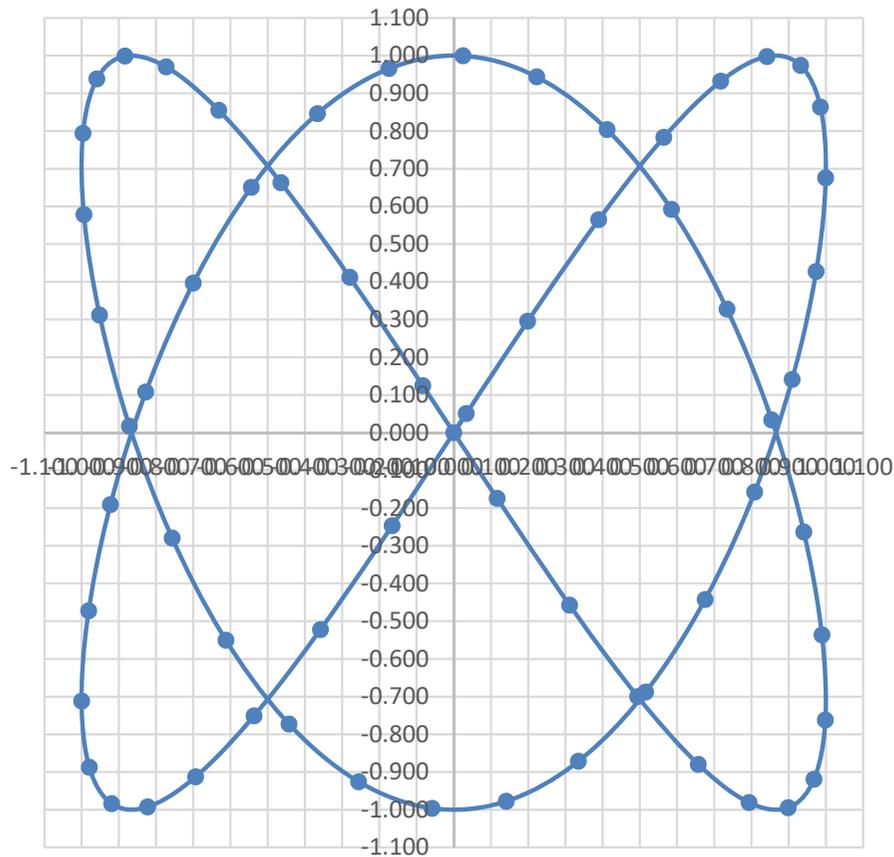


	A	B	C	D
1				
2				
3	個数 / id	列ラベル ▼		
4	行ラベル ▼	☐ 犬	猫	総計
5	☐ 女性		46 48	94
6	10		7 6	13
7	20		9 6	15
8	30		4 10	14
9	40		10 12	22
10	50		6 9	15
11	60		10 5	15
12	☐ 男性		50 56	106
13	10		6 1	7
14	20		7 10	17
15	30		12 13	25
16	40		6 13	19
17	50		7 7	14
18	60		9 12	21
19	70		3	3
20	総計		96 104	200

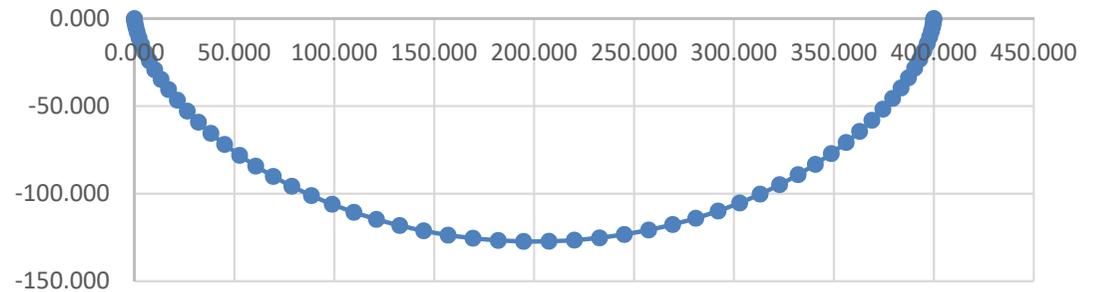
• 媒介変数表記とExcel散布図(平滑線)による関数の描画

- リサージュ曲線
- サイクロイド曲線
- 2次曲線(円・楕円・双曲線)

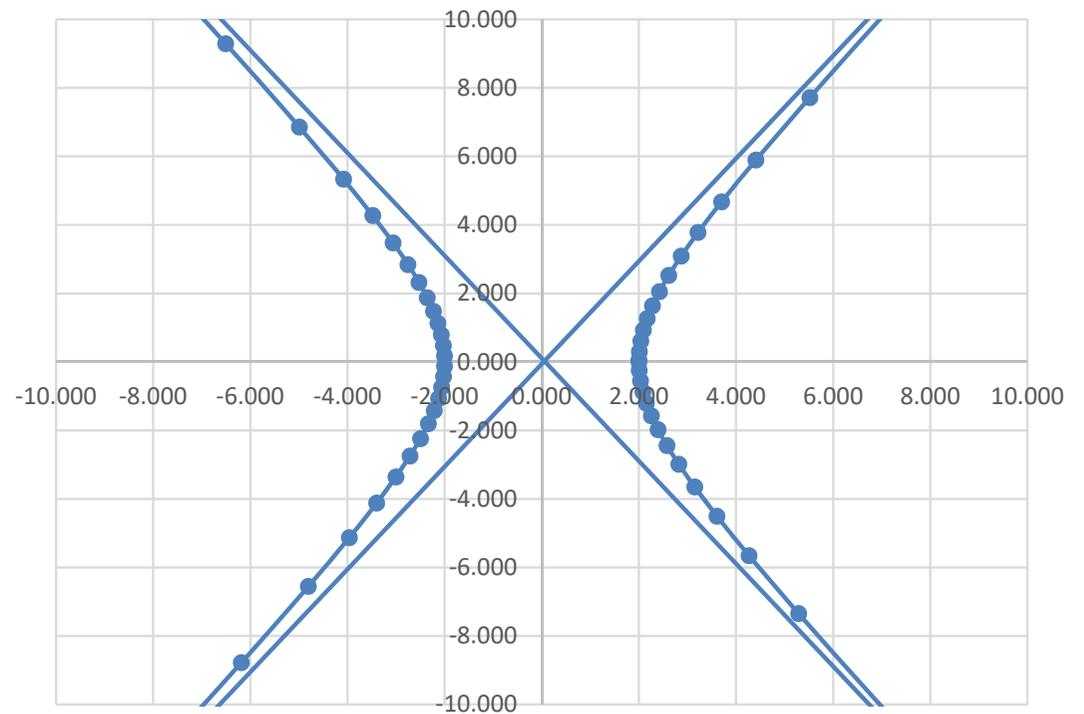
リサージュ曲線



サイクロイド曲線



2次曲線



【参考】

2変数間の分析法

尺度によって
分析法が変わる
ことに注意

- 2変数 x, y の相関関係を調べる方法(図表と式)

例1

	A	B	C	D	E	F	G	H	I	J	尺度
性別 x	男	男	女	男	男	男	女	女	男	女	質的
嗜好 y	紅茶	緑茶	珈琲	珈琲	緑茶	珈琲	紅茶	珈琲	珈琲	紅茶	質的



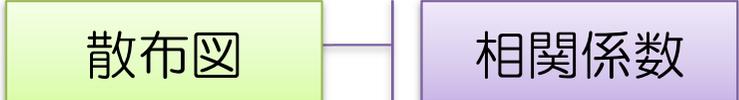
例2

	A	B	C	D	E	F	G	H	I	J	尺度
飲量 x	15	32	16	30	50	12	14	24	18	19	量的
嗜好 y	紅茶	緑茶	珈琲	珈琲	緑茶	珈琲	紅茶	珈琲	珈琲	紅茶	質的



例3

	A	B	C	D	E	F	G	H	I	J	尺度
身長 x	176	170	163	173	170	171	165	170	176	156	量的
体重 y	61	73	54	65	67	62	51	57	77	43	量的



2変数の関係

□ 2変数の関係1 : x (質的) $\times y$ (質的) 図

	A	B	C	D	E	F	G	H	I	J	
性別 x	男	男	女	男	男	男	女	女	男	女	質的
嗜好 y	紅茶	緑茶	珈琲	珈琲	緑茶	珈琲	紅茶	珈琲	珈琲	紅茶	質的

クロス集計

	紅茶	緑茶	珈琲	計	
男	1	2	3	6	} 周辺度数
女	2	0	2	4	
計	3	2	5	10	← 総度数

} 周辺度数

2変数の関係

□ 2変数の関係1 : x (質的) $\times y$ (質的)式

	紅茶	緑茶	珈琲	計	連関係数		紅茶	緑茶	珈琲	計
男	1	2	3	6	クロス集計 から 理論度数 を 求める	男	1.8	1.2	3.0	6
女	2	0	2	4		女	1.2	0.8	2.0	4
計	3	2	5	10		計	3	2	5	10

$$1.8 = \frac{3 \cdot 6}{10}$$

$$2.0 = \frac{5 \cdot 4}{10}$$

□ クラメルの連関係数 *Cramer's coefficient of association*

$$V = \sqrt{\frac{\chi^2}{n \cdot m}}$$

$$(0 \leq V \leq 1)$$

$$\chi^2 = \frac{(1-1.8)^2}{1.8} + \frac{(2-1.2)^2}{1.2} + \dots + \frac{(0-0.8)^2}{0.8} + \frac{(2-2.0)^2}{2.0}$$

$$n = 10$$

$$m = \min\{2-1, 3-1\}$$

ピアソンの
 χ^2 統計量

(行数-1)と(列数-1)
の小さい方

2変数の関係

□ 2変数の関係1：x(質的)×y(質的)式

□ クラメルの連関係数 *Cramer's coefficient of association*

	紅	緑	珈	計
男	0	3	9	12
女	6	0	0	6
計	6	3	9	18

	紅	緑	珈	計
男	3	1	8	12
女	3	2	1	6
計	6	3	9	18

	紅	緑	珈	計
男	4	2	6	12
女	2	1	3	6
計	6	3	9	18

$$\chi^2 = \frac{(0-4)^2}{4} + \frac{(3-2)^2}{2} + \frac{(9-6)^2}{6} + \frac{(6-2)^2}{2} + \frac{(0-1)^2}{1} + \frac{(0-3)^2}{3}$$

$$= 18$$

$n = 18$
 $m = \min\{2-1, 3-1\} = 1$

$$\chi^2 = \frac{(3-4)^2}{4} + \frac{(1-2)^2}{2} + \frac{(8-6)^2}{6} + \frac{(3-2)^2}{2} + \frac{(2-1)^2}{1} + \frac{(1-3)^2}{3}$$

$$= 17/4$$

$n = 18$
 $m = \min\{2-1, 3-1\} = 1$

$$\chi^2 = \frac{(4-4)^2}{4} + \frac{(2-2)^2}{2} + \frac{(6-6)^2}{6} + \frac{(2-2)^2}{2} + \frac{(1-1)^2}{1} + \frac{(3-3)^2}{3}$$

$$= 0$$

$n = 18$
 $m = \min\{2-1, 3-1\} = 1$

→ $V = \sqrt{\frac{18}{18 \cdot 1}} = 1$ 嗜好と性別は **完全相関**

→ $V = \sqrt{\frac{17/4}{18 \cdot 1}} \approx 0.49$ 嗜好と性別は **多少相関**

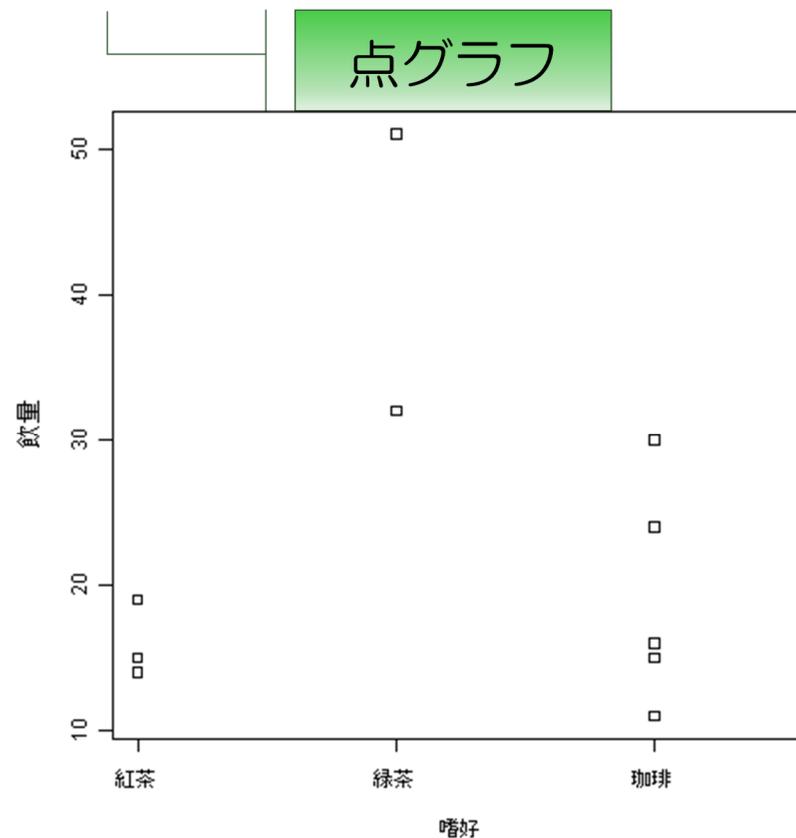
→ $V = \sqrt{\frac{0}{18 \cdot 1}} = 0$ 嗜好と性別は **無相関**

2変数の関係

□ 2変数の関係2 : x (量的) $\times y$ (質的)

	A	B	C	D	E	F	G	H	I	J
飲量 x	15	32	16	30	50	12	14	24	18	19
嗜好 y	紅茶	緑茶	珈琲	珈琲	緑茶	珈琲	紅茶	珈琲	珈琲	紅茶

量的
質的



2変数の関係

□ 2変数の関係2 : x (量的) $\times y$ (質的)式

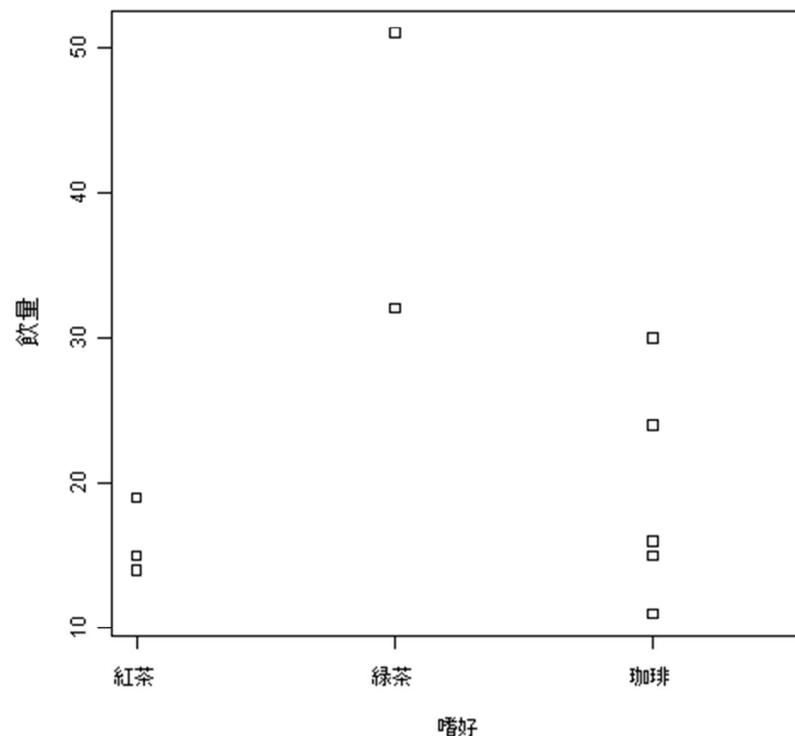
	A	B	C	D	E	F	G	H	I	J	
飲量 x	15	32	16	30	50	12	14	24	18	19	量的
嗜好 y	紅茶	綠茶	珈琲	珈琲	綠茶	珈琲	紅茶	珈琲	珈琲	紅茶	質的

相関比

□ 相関比 *correlation ratio*

$$\eta^2 = \frac{S_T}{S_B + S_T}$$

$$(0 \leq \eta^2 \leq 1)$$



2変数の関係

□ 2変数の関係2 : x (量的) $\times y$ (質的)式

□ 相関比 *correlation ratio*

$$\eta^2 = \frac{S_T}{S_B + S_T} \quad (0 \leq \eta^2 \leq 1)$$

$$\eta^2 = \frac{840}{376 + 840} \approx 0.691$$

	紅茶	緑茶	珈琲	
	14	32	12	
	15	50	16	
	19		18	
			24	
			30	
個数	3	2	5	全平均
平均	16	41	20	23
偏差平方	49	324	9	840 = S_T

$$49 = (16 - 23)^2$$

$$324 = (41 - 23)^2$$

$$9 = (20 - 23)^2$$

$$S_T = \underline{840} = 49 \times 3 + 324 \times 2 + 9 \times 5$$

級間変動

= 級平均と全平均との偏差平方の加重和

偏差平方	4	81	64	
	1	81	16	
	9		4	
			16	
			100	
計	14	162	200	376 = S_B

級間変動

$$14 = (14 - 16)^2 + (15 - 16)^2 + (19 - 16)^2$$

$$162 = (32 - 41)^2 + (50 - 41)^2$$

$$200 = (12 - 20)^2 + (16 - 20)^2 + \dots + (30 - 20)^2$$

$$S_B = \underline{376} = 14 + 162 + 200$$

級内変動

= 級内データと級平均との偏差平方の和

級内変動

2変数の関係

□ 2変数の関係2 : x (量的) $\times y$ (質的)式

□ 相関比 *correlation ratio*

$$\eta^2 = \frac{840}{0 + 840} = 1$$

$$\eta^2 = \frac{840}{376 + 840} \approx 0.691$$

$$\eta^2 = \frac{0}{314 + 0} = 0$$

嗜好と飲量は**完全相関**

	紅茶	緑茶	珈琲	
	16	41	20	
	16	41	20	
	16		20	
			20	
			20	
個数	3	2	5	全平均
平均	16	41	20	23
偏差平方和	49	324	9	840
				級間変動
偏差平方和	0	0	0	
	0	0	0	
	0		0	
			0	
			0	合計
計	0	0	0	0
				級内変動

嗜好と飲量は**多少相関**

	紅茶	緑茶	珈琲	
	14	32	12	
	15	50	16	
	19		18	
			24	
			30	
個数	3	2	5	全平均
平均	16	41	20	23
偏差平方和	49	324	9	840
				級間変動
偏差平方和	4	81	64	
	1	81	16	
	9		4	
			16	
			100	合計
計	14	162	200	376
				級内変動

嗜好と飲量は**無相関**

	紅茶	緑茶	珈琲	
	19	15	15	
	21	31	20	
	29		25	
			25	
			30	
個数	3	2	5	全平均
平均	23	23	23	23
偏差平方和	0	0	0	0
				級間変動
偏差平方和	16	64	64	
	4	64	9	
	36		4	
			4	
			49	合計
計	56	128	130	314
				級内変動

2変数の関係

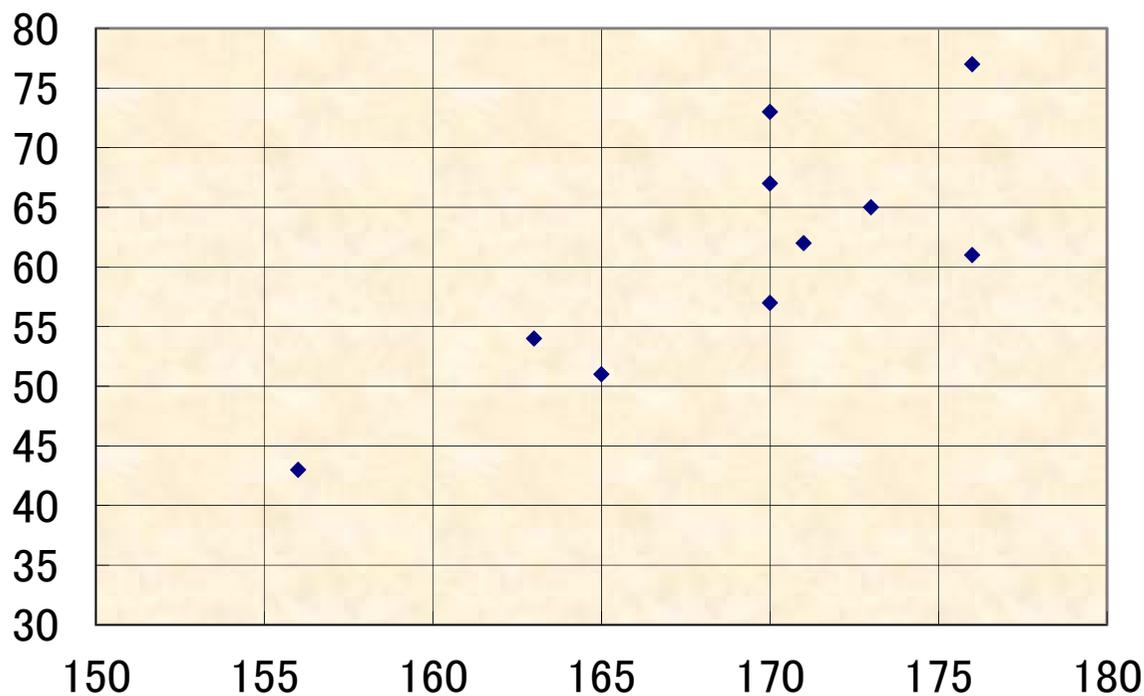
□ 2変数の関係3 : x (量的) $\times y$ (量的) 図

	A	B	C	D	E	F	G	H	I	J
身長 x	176	170	163	173	170	171	165	170	176	156
体重 y	61	73	54	65	67	62	51	57	77	43

量的

量的

散布図



2変数の関係

□ 2変数の関係3 : x (量的) $\times y$ (量的)式

	A	B	C	D	E	F	G	H	I	J	平均
身長 x	176	170	163	173	170	171	165	170	176	156	169
体重 y	61	73	54	65	67	62	51	57	77	43	61

相関係数

□ ピアソンの積率相関係数 *Pearson's product-moment correlation coefficient*

$$r_{xy} = \frac{\text{COV}_{xy}}{S_x \cdot S_y}$$

$$\approx \frac{46}{5.848 \cdot 9.706}$$

$$\approx 0.81$$

$$(-1 \leq r_{xy} \leq 1)$$

$$\left\{ \begin{array}{l} \text{COV}_{xy} = \frac{(176-169)(61-61) + \dots + (156-169)(43-61)}{10} = 46 \quad (x,y \text{の共分散}) \\ S_x = \sqrt{\frac{(176-169)^2 + \dots + (156-169)^2}{10}} \approx 5.848 \quad (x \text{の標準偏差}) \\ S_y = \sqrt{\frac{(61-61)^2 + \dots + (43-61)^2}{10}} \approx 9.706 \quad (y \text{の標準偏差}) \end{array} \right.$$

2変数の関係

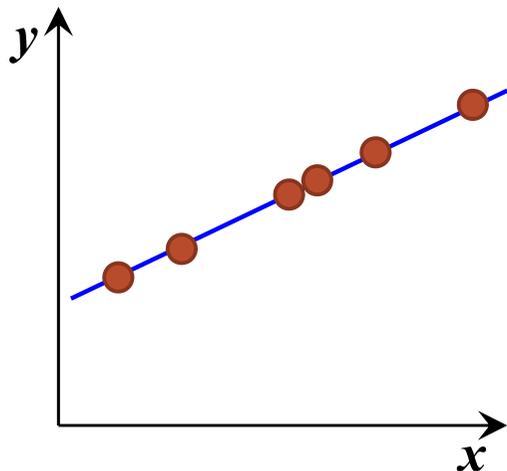
□ 2変数の関係3：x(量的)×y(量的)式

□ ピアソンの積率相関係数 *Pearson's product-moment correlation coefficient*

$$r_{xy} = \frac{\text{COV}_{xy}}{S_x \cdot S_y} \quad (-1 \leq r_{xy} \leq 1)$$

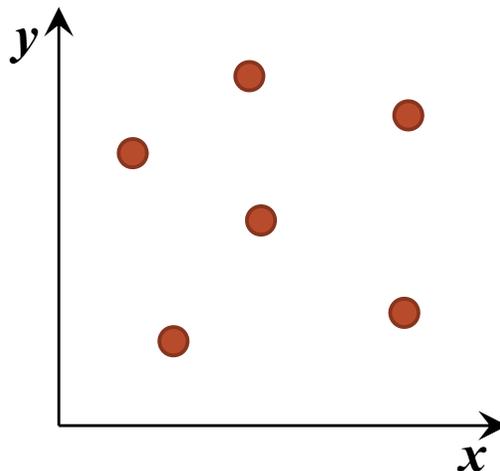
$$r_{xy} = \frac{\text{COV}_{xy}}{S_x \cdot S_y} = 1$$

身長と体重は正の相関



$$r_{xy} = \frac{\text{COV}_{xy}}{S_x \cdot S_y} = 0$$

身長と体重は無相関



$$r_{xy} = \frac{\text{COV}_{xy}}{S_x \cdot S_y} = -1$$

身長と体重は負の相関

