

問題解決技法入門

3. Data Analysis

1. 2変数の相関 correlation (1)

クロス集計 cross tabulation

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

堀田 敬介

2変数の相関(1)

• アンケート調査

– 結果例

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

⋮

←項目(field)

属性(property)・変数(variable) 等とも言う

このデータの主キー(key)は[id]

主キーとは, レコードを特定出来る(区別できる)1つ乃至複数の項目の組合せ

[性別][年代][嗜好1][嗜好2]は主キーにできない

←レコード(record)

何のために調査するのか？

<調査の目的例>

1. 単純集計 grand total

例1)[嗜好1]で犬と猫どちらを好む人が多いか？

例2)[嗜好2]でどの飲料がよく飲まれているか？

2. 項目間の相関 correlation を知りたい

例1)[年代]と[嗜好1]には関係があるか？

例2)[性別]と[嗜好2]には関係があるか？

2変数の相関(1)

尺度によって
分析法が変わ
ることに注意

- 2変数 x, y 間の相関を調べる方法(図表と式)

<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	綠茶	珈琲	珈琲	綠茶	珈琲	紅茶	珈琲	珈琲	紅茶	質的

<2>	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

点グラフ

相関比

量的

質的

<3>	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

散布図

相関係数

2変数の関係

□ 2変数の関係1 : x (質的) $\times y$ (質的) ㊦

	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

クロス集計

性別 x と嗜好 y についてのクロス集計表

		嗜好 y			計	
		紅茶	緑茶	珈琲		
性別 x	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

クロス集計とは

- クロス集計(表) cross tabulation
 - 2つ以上の**項目間の相関関係**を知りたい時に使う集計方法のひとつ. 分割表ともよぶ

データ

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

⋮



例1)「年代」と「嗜好1」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



例2)「性別」と「嗜好2」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]

Excelの「データ」タブと「フィルタ」機能の使用方法を示すスクリーンショット。図中の①はデータ項目を選択する操作、②は「データ」タブ、③は「フィルタ」ボタンを示している。

	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 (Ctrl+Shift+L)

選択したセルにフィルタを適用します。

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

- ※ 論理積 (かつ, AND) でフィルタがかかる

フィルタで選択

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

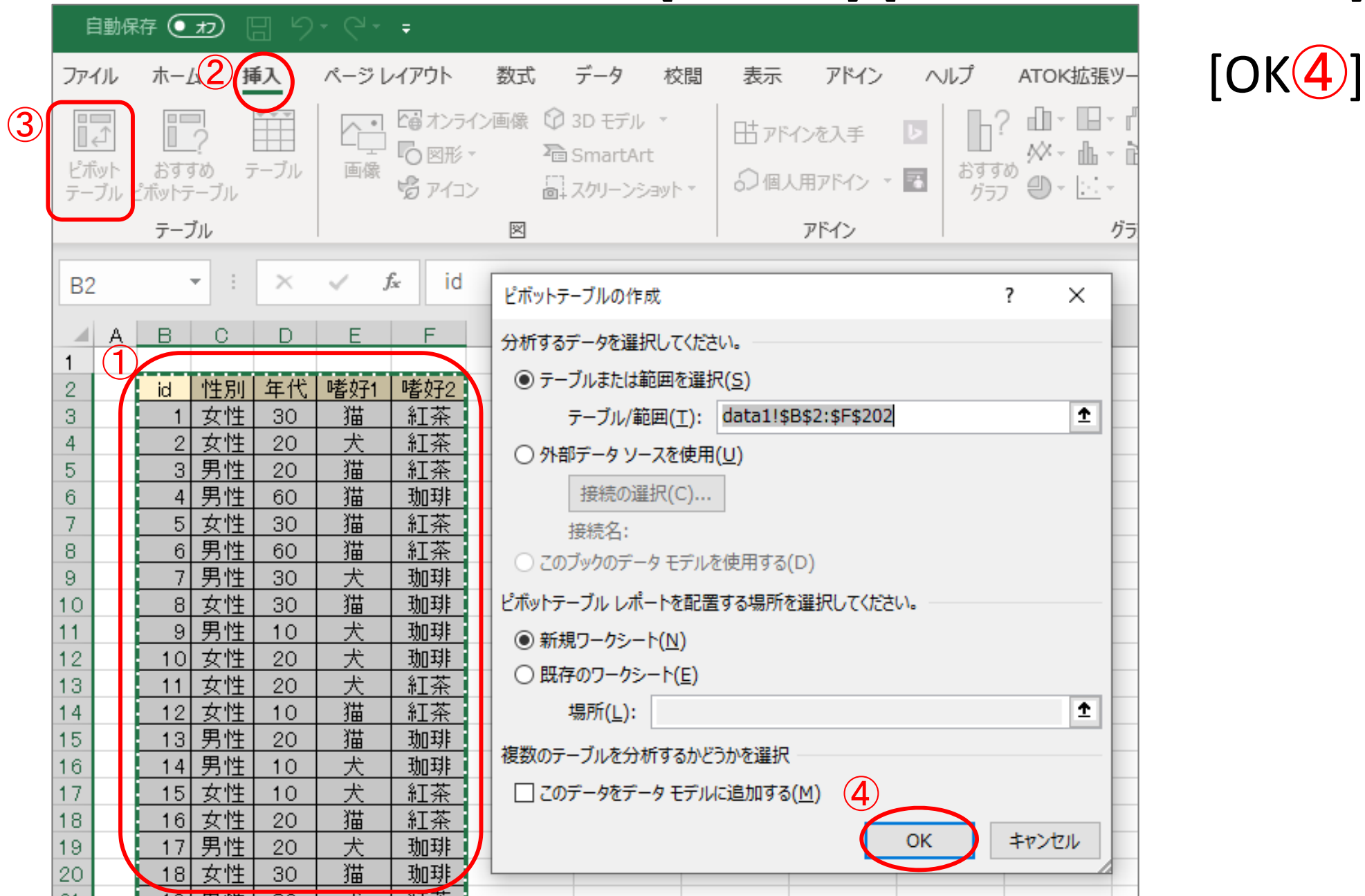
< 論理演算の例 >

- ✓ 論理積 AND/intersection/logical conjunction/logical product
- ✓ 論理和 OR/union/logical disjunction/logical add/logical sum
- ✓ 論理否定 not
- ✓ 排他的論理和 XOR/exclusive or

	A	B	C	D	E	F
1						
2		id	性別	年齢	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						

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]



Excelでクロス集計

• [ピボットテーブルのフィールド]



- 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

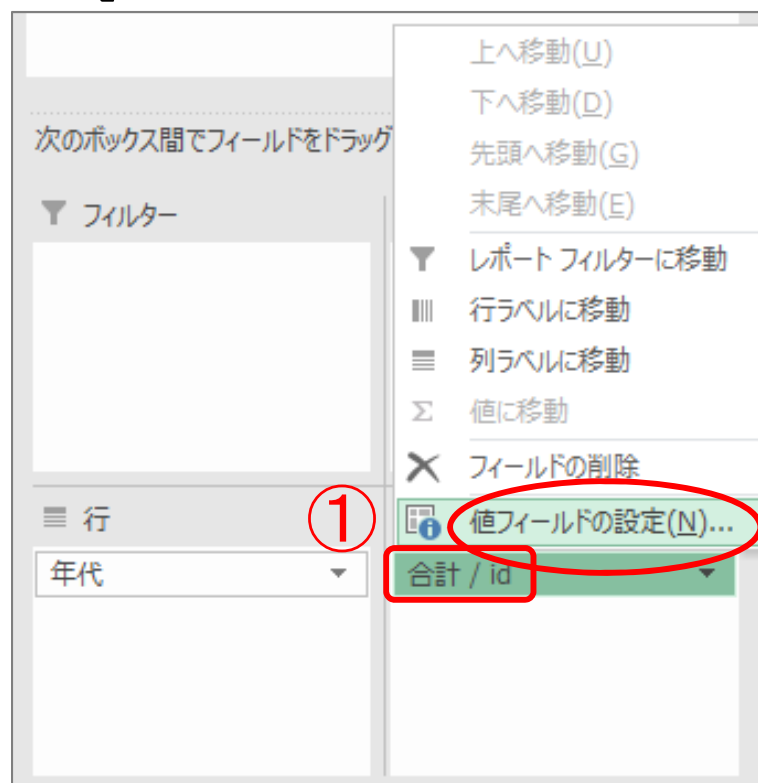
「行」=「年代」
「列」=「嗜好1」
「値」=「id」

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				

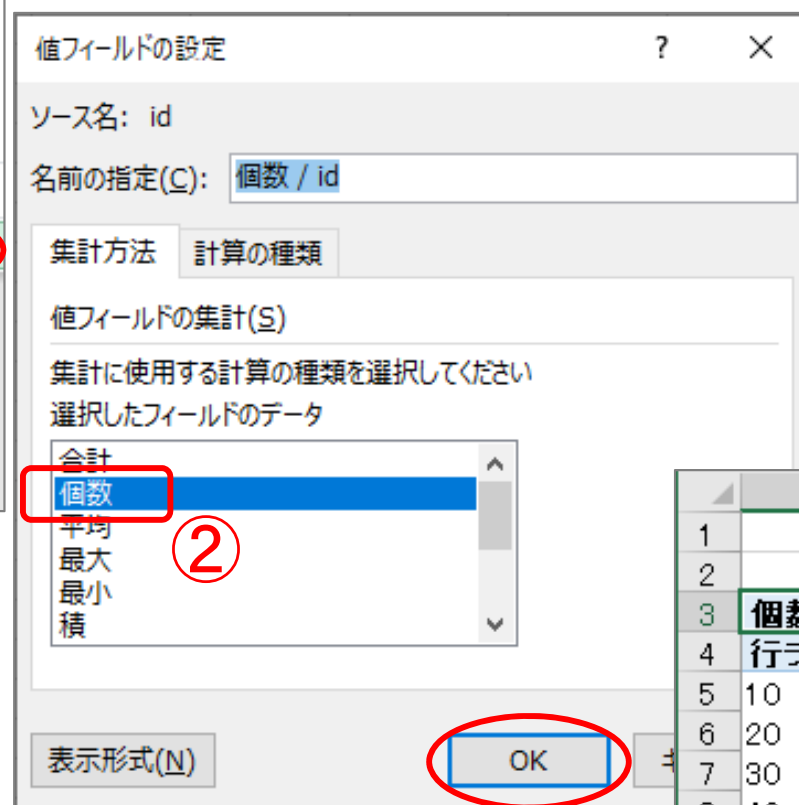
Excelでクロス集計

- [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
6	20		16	16
7	30		16	23
8	40		16	25
9	50		13	16
10	60		19	17
11	70		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**

ピボットテーブルのフィールド

レポートに追加するフィールドを選択してください:

検索

☒ id
☒ 性別
☐ 年代
☐ 嗜好1
☒ 嗜好2

その他のテーブル...

①

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2

≡ 行

性別

Σ 値

個数 / id

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

≡ 行

性別 ▼

≡ 列

嗜好1 ▼

嗜好2 ▼

Σ 値

個数 / id ▼



	A	B	C	D	E	F	G	H
1								
2								
3	個数 / id	列ラベル ▼						
4		犬		犬 集計	猫		猫 集計	総計
5	行ラベル ▼	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

≡ 行

性別 ▼

年代 ▼

≡ 列

嗜好1 ▼

Σ 値

個数 / id ▼

「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

Excelでクロス集計

- 演習) 集計範囲を選択①し [挿入②]-[ピボットテーブル③], [OK④]

The screenshot shows the Excel interface with the 'Insert' tab selected. The 'PivotTable' button in the 'Tables' group is circled in red. A red circle highlights the data range B2:H202 in the worksheet. The 'PivotTable of the PivotTable' dialog box is open, showing the data range and the 'OK' button circled in red.

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(I): データ2!\$B\$2:\$H\$202

☐ 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

Excelでクロス集計

- 演習) 自由に目的をつくり, クロス集計しよう

ピボットテーブルのフィールド

レポートに追加するフィールドを選択してください:

検索

- ☐ id
- ☒ 商品
- ☐ 単価
- ☐ 数量
- ☒ 金額
- ☒ 仕入れ先
- ☐ 販売先

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

列ラベル

行ラベル

値

商品

仕入れ先

合計 / 金額

削除・修正したい場合は戻す

レイアウトの更新を保留する

更新

[ピボットテーブルのフィールド]

- 上半分に項目(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つを指定する
- 「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Excelシート内に完成
- 修正・編集も同様(ドラッグ&ドロップ)

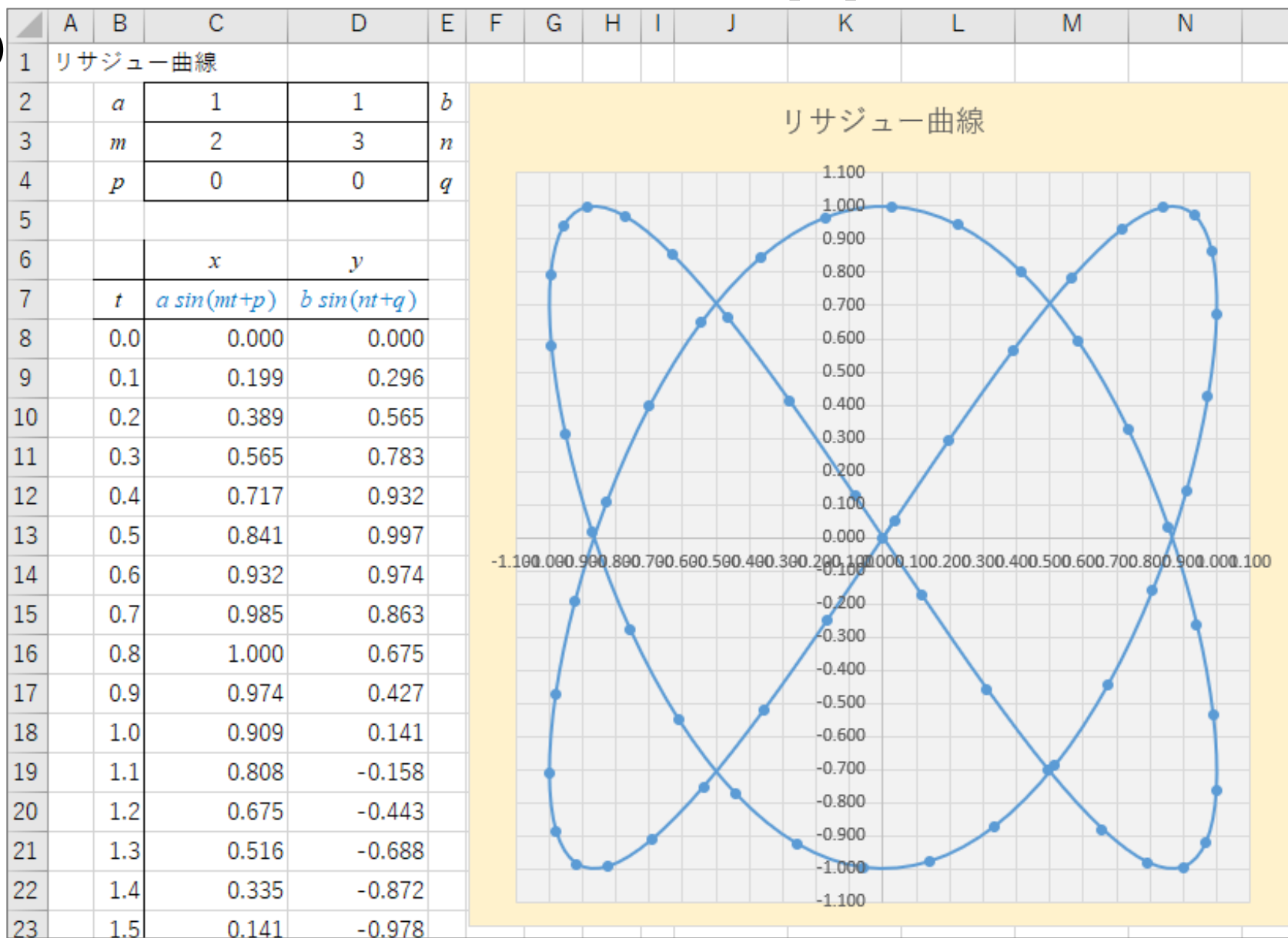
例) 左の設定でできたクロス集計表
「行」=「商品」, 「列」=「仕入れ先」, 「値」=「金額」合計

	A	B	C	D	E	F	G
1							
2							
3	合計 / 金額	列ラベル					
4	行ラベル	農協JB	農協JC	農場α	農場β	農場γ	総計
5	じゃがいも	¥227,990	¥209,870	¥416,070	¥181,680	¥301,090	¥1,336,700
6	たまねぎ	¥632,040	¥209,560	¥223,420	¥208,200	¥316,740	¥1,589,960
7	にんじん	¥455,810	¥291,930	¥492,780	¥443,520	¥208,200	¥1,892,240
8	はくさい	¥359,360	¥435,900	¥61,860	¥398,340	¥172,360	¥1,427,820
9	れんこん	¥425,520	¥88,400	¥285,020	¥437,740	¥617,690	¥1,854,370
10	総計	¥2,100,720	¥1,235,660	¥1,479,150	¥1,669,480	¥1,616,080	¥8,101,090
11							

• 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(媒介変数 t , パラメータ a, b, m, n, p, q)

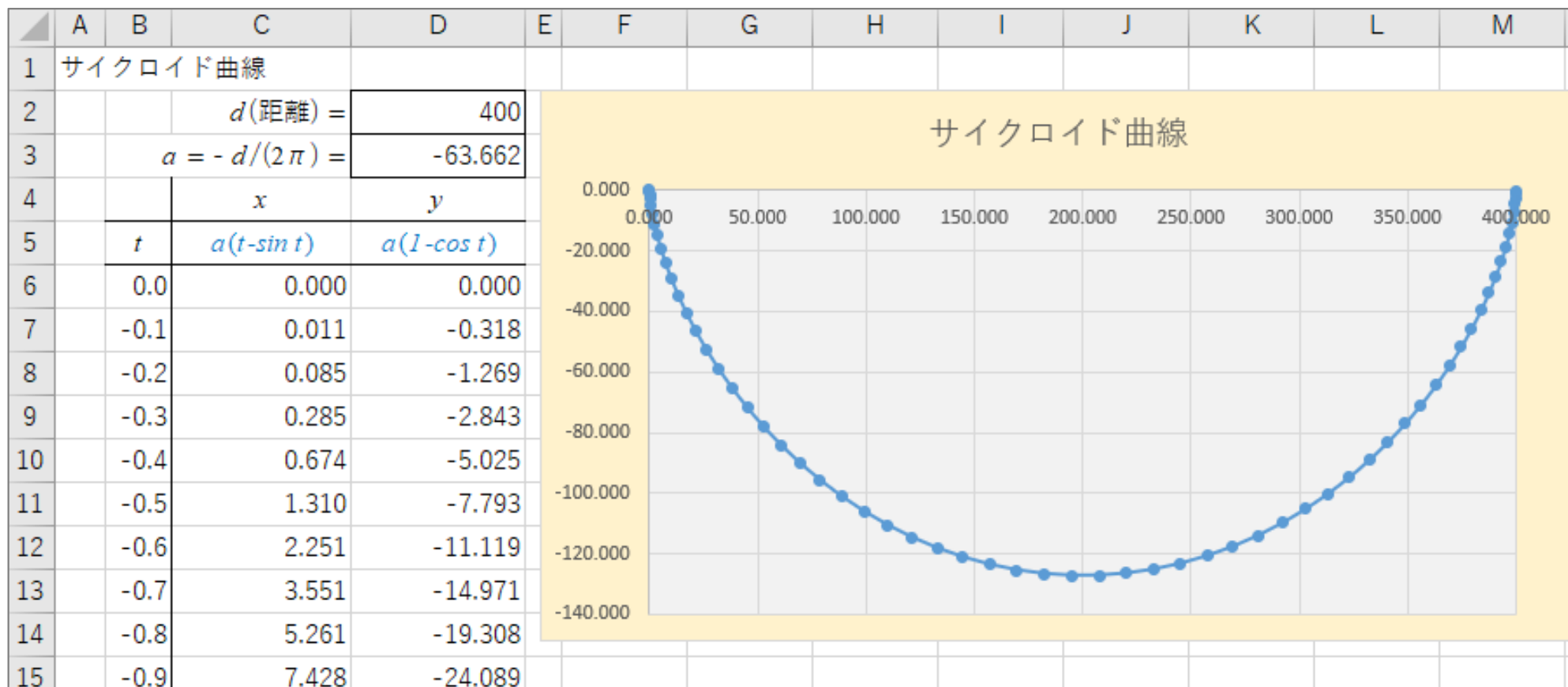
$$\begin{cases} x = a \sin(mt + p) \\ y = b \sin(nt + q) \end{cases}$$



• 媒介変数表記とExcel散布図(平滑線)による関数の描画

– サイクロイド曲線(媒介変数 t , パラメータ a)

$$\begin{cases} x = a(t - \sin t) \\ y = a(1 - \cos t) \end{cases}$$



• 媒介変数表記とExcel散布図(平滑線)による関数の描画

– 2次曲線(双曲線・楕円・円・放物線)(媒介変数 t , パラメータ a, b, p, q)

$$\begin{cases} x = a(\cos t)^p \\ y = \frac{b \sin t}{(\cos t)^q} \end{cases}$$

楕円の例

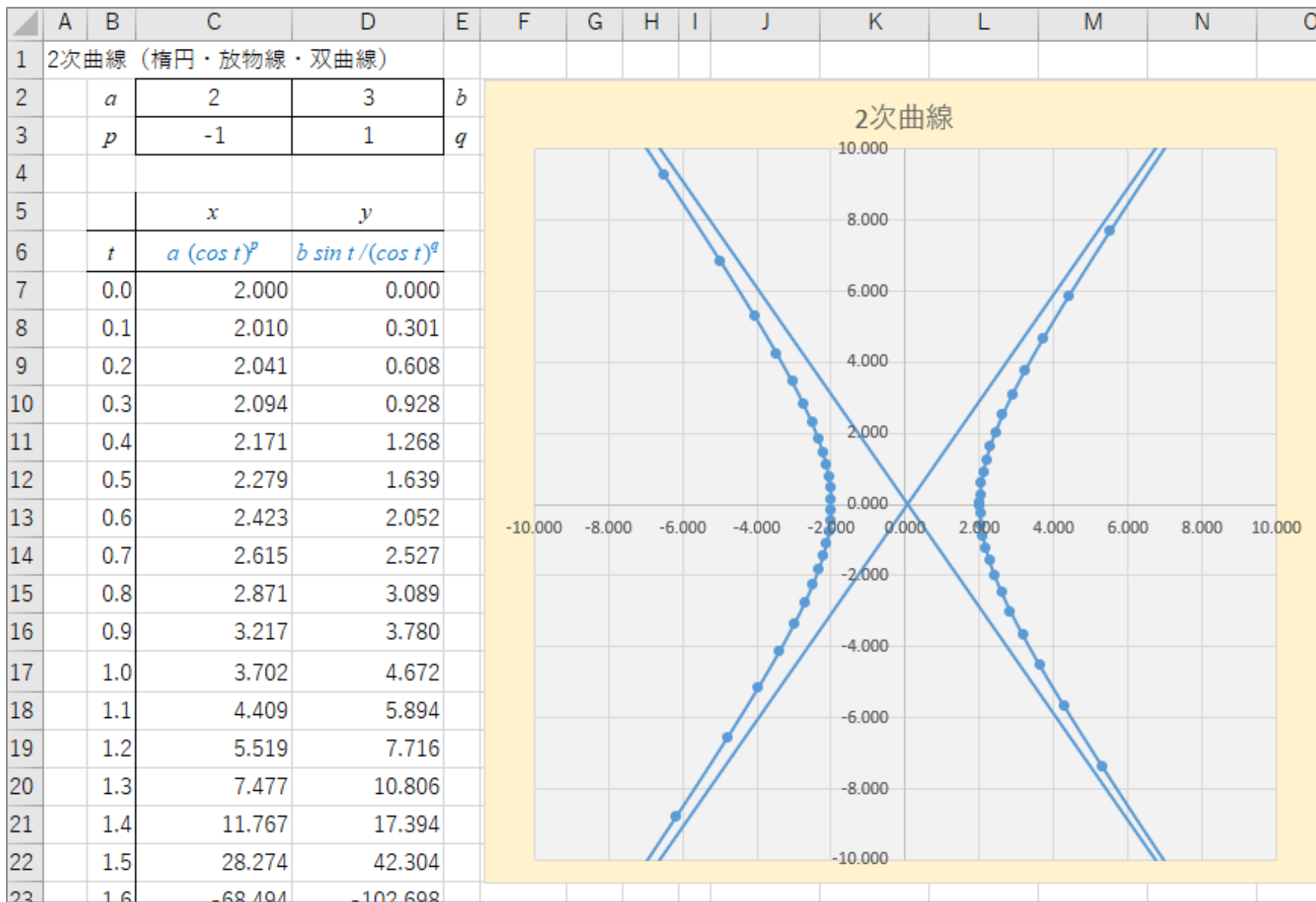
$$\begin{matrix} a & 7 & 3 & b \\ p & 1 & 0 & q \end{matrix}$$

円の例

$$\begin{matrix} a & 7 & 7 & b \\ p & 1 & 0 & q \end{matrix}$$

放物線の例

$$\begin{matrix} a & 7 & 4 & b \\ p & 2 & 0 & q \end{matrix}$$



Coffee Break!

• 媒介変数表記とExcel散布図(平滑線)による関数

- ハイポサイクロイド[デルトイド($a=3, b=1$)・アステロイド($a=4, b=1$)]

$$\begin{cases} x = (a - b) \cos t + b \cos\left(\frac{a-b}{b}t\right) \\ y = (a - b) \sin t + b \sin\left(\frac{a-b}{b}t\right) \end{cases}$$

- エピサイクロイド[カージオイド($a=1, b=1$)/ネフロイド($a=2, b=1$)]

$$\begin{cases} x = (a + b) \cos t - b \cos\left(\frac{a+b}{b}t\right) \\ y = (a + b) \sin t - b \sin\left(\frac{a+b}{b}t\right) \end{cases}$$

- レムニスケート($a=3$)

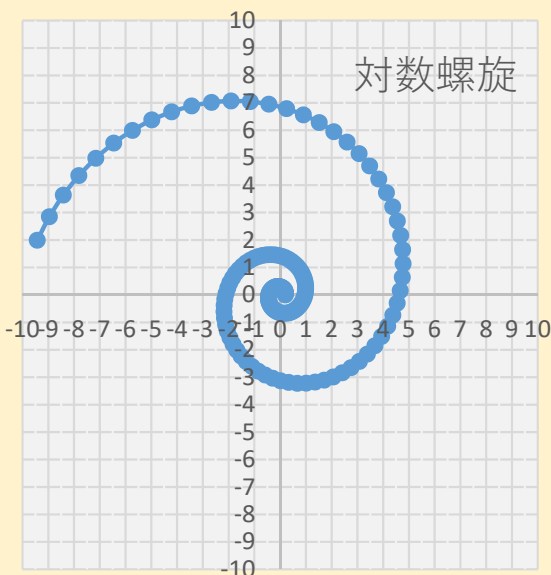
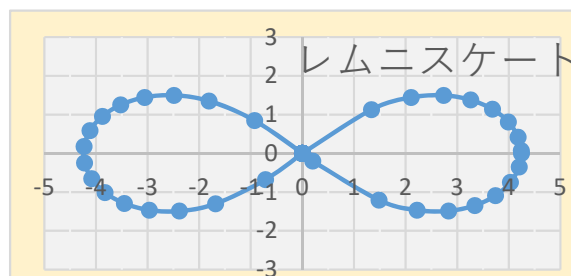
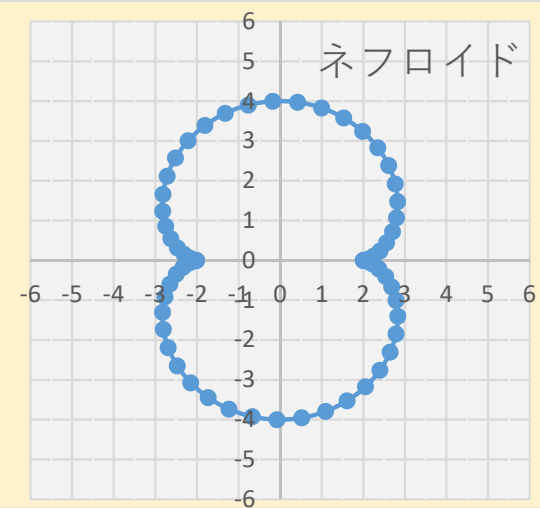
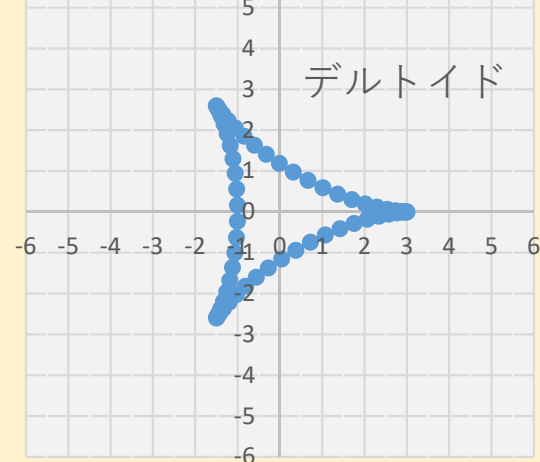
$$\begin{cases} x = \sqrt{2}a \cos t \sqrt{\cos 2t} \\ y = \sqrt{2}a \sin t \sqrt{\cos 2t} \end{cases}$$

- リマソン($a=1, b=2$)

$$\begin{cases} x = (a + b \cos t) \cos t \\ y = (a + b \cos t) \sin t \end{cases}$$

- 対数螺旋($a=0.2, b=0.25$)

$$\begin{cases} x = ae^{bt} \cos t \\ y = ae^{bt} \sin t \end{cases}$$



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)×y(質的)式

	紅茶	緑茶	珈琲	計
男	1	2	3	6
女	2	0	2	4
計	3	2	5	10

連関係数

クロス集計
から
理論度数
を
求める

	紅茶	緑茶	珈琲	計
男	1.8	1.2	3.0	6
女	1.2	0.8	2.0	4
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

$$\left\{ \begin{aligned} \chi^2 &= \frac{(0-4)^2}{4} + \frac{(3-2)^2}{2} + \frac{(9-6)^2}{6} \\ &\quad + \frac{(6-2)^2}{2} + \frac{(0-1)^2}{1} + \frac{(0-3)^2}{3} \\ &= 18 \\ n &= 18 \\ m &= \min\{2-1, 3-1\} = 1 \end{aligned} \right.$$

$$\rightarrow V = \sqrt{\frac{18}{18 \cdot 1}} = 1 \quad \text{嗜好と性別は} \quad \text{完全相関}$$

$$\left\{ \begin{aligned} \chi^2 &= \frac{(3-4)^2}{4} + \frac{(1-2)^2}{2} + \frac{(8-6)^2}{6} \\ &\quad + \frac{(3-2)^2}{2} + \frac{(2-1)^2}{1} + \frac{(1-3)^2}{3} \\ &= 17/4 \\ n &= 18 \\ m &= \min\{2-1, 3-1\} = 1 \end{aligned} \right.$$

$$\rightarrow V = \sqrt{\frac{17/4}{18 \cdot 1}} \approx 0.49 \quad \text{嗜好と性別} \quad \text{は多少相関}$$

$$\left\{ \begin{aligned} \chi^2 &= \frac{(4-4)^2}{4} + \frac{(2-2)^2}{2} + \frac{(6-6)^2}{6} \\ &\quad + \frac{(2-2)^2}{2} + \frac{(1-1)^2}{1} + \frac{(3-3)^2}{3} \\ &= 0 \\ n &= 18 \\ m &= \min\{2-1, 3-1\} = 1 \end{aligned} \right.$$

$$\rightarrow V = \sqrt{\frac{0}{18 \cdot 1}} = 0 \quad \text{嗜好と性別} \quad \text{は無相関}$$

2変数の関係

□ 2変数の関係1：x(質的)×y(質的)式

□ クラメルの変関係数 *Cramer's coefficient of association*

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1																		
2		観察度数					理論度数					(観察度数-理論度数) ² /理論度数						[H4] = H\$6 * \$J4 / \$J\$6
3			犬	猫	総計			犬	猫	総計			犬	猫	総計			→[H4:I5]へコピー
4		女性	46	48	94		女性	45.12	48.88	94		女性	0.017	0.016	94			[M4] = (C4-H4)^2/H4
5		男性	50	56	106		男性	50.88	55.12	106		男性	0.015	0.014	106			→[H4:I5]へコピー
6		総計	96	104	200		総計	96	104	200		総計	96	104	200			
7																		[M8] = SUM(M4:N5)
8								ピアソンのχ ² 統計量				χ ² =	0.062					[M9] = E6
9												n=	200	m=	1			[N9] = MIN(2-1, 2-1)
10								クラメルの連関係数				V=	0.018					[M10] = SQRT(M8 / (M9 * O9))
11																		
12								理論度数					(観察度数-理論度数) ² /理論度数					
13			犬	猫	総計			犬	猫	総計			犬	猫	総計			
14		紅茶	58	52	110		紅茶	52.80	57.20	110		紅茶	0.512	0.473	110			
15		珈琲	38	52	90		珈琲	43.20	46.80	90		珈琲	0.626	0.578	90			
16		総計	96	104	200		総計	96	104	200		総計	96	104	200			
17																		
18								ピアソンのχ ² 統計量				χ ² =	2.189					
19												n=	200	m=	1			
20								クラメルの連関係数				V=	0.105					

参考文献

(推測)統計学について学びたい場合は

- ◆ 村上 雅人『なるほど統計学』海鳴社(2002)
- ◆ 石井 俊全『1冊でマスター 大学の統計学』技術評論社(2018)

もっと知りたい人へ

- 関連する経営学科の授業
 - 「基礎統計」(1/2セメ)
 - 「基礎統計演習」(3/4セメ)
 - 「データ処理応用」(2/3セメ)
 - 「統計モデル分析」(5セメ)
 - 「ビッグデータ・AI演習」(6セメ)

2変数の相関(参考)

尺度によって
分析法が変わ
ることに注意

- 2変数 x, y 間の相関を調べる方法(図表と式)

<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
		クロス集計					連関係数				

<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
		点グラフ					相関比				

<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
		散布図					相関係数				

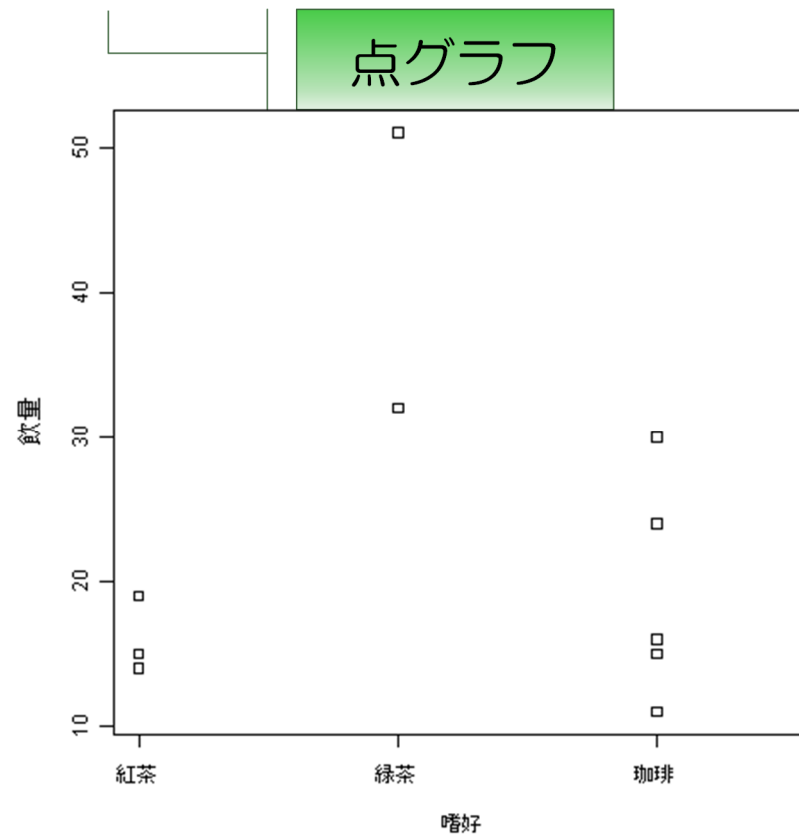
2変数の関係

□ 2変数の関係2 : x (量的) $\times y$ (質的) 図

	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

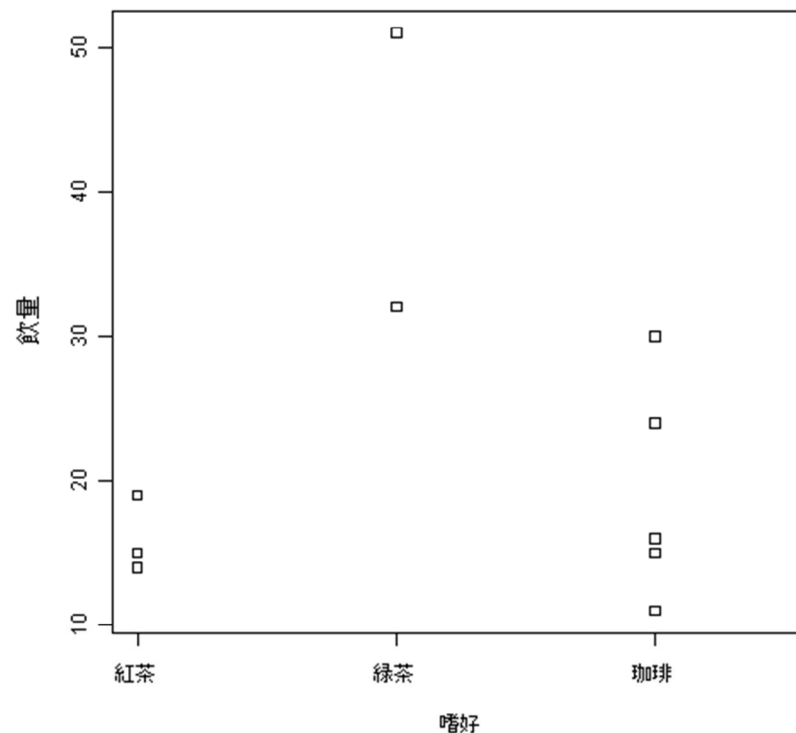
	A	B	C	D	E	F	G	H	I	J	
飲量 x	15	32	16	30	50	12	14	24	18	19	量的
嗜好 y	紅茶	緑茶	珈琲	珈琲	緑茶	珈琲	紅茶	珈琲	珈琲	紅茶	質的

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
平均	16	41	20
偏差平方	49	324	9

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

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

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

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

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

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	
			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