

データサイエンス入門B

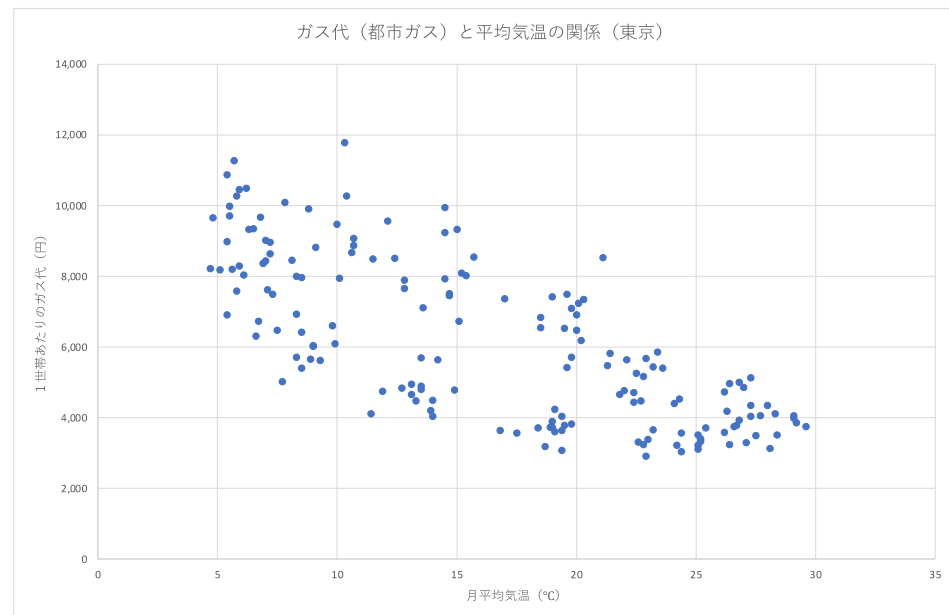
第10回 散布図と相関係数

高田 美樹

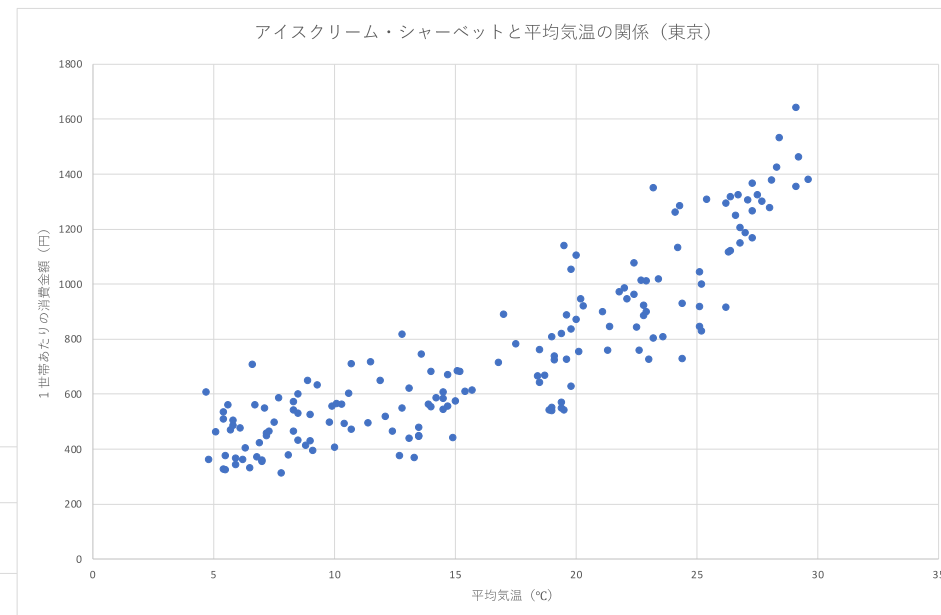
目次

1. 2つの変数の関係を可視化する散布図
2. 2つの変数の直線的な関係の強さを表す共分散と相関係数
3. 相関関係と因果関係
4. 擬似相関と潜在変数
5. 実験研究と観察研究
6. M字カーブの深さの要因

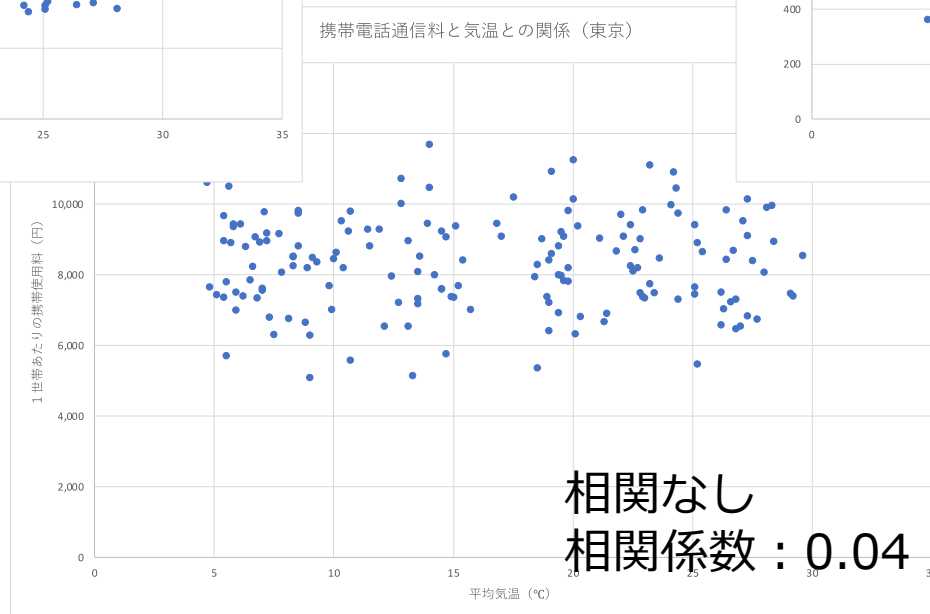
2つの変数の関係



負の相関
相関係数 : -0.74



正の相関
相関係数 : 0.88



相関なし
相関係数 : 0.04

家計データ

企業・家計・経済



主な調査

- ＞ 国民経済計算
- ＞ 個人企業経済調査
- ＞ 経済センサス-基礎調査
- ＞ 経済センサス-活動調査
- ＞ 経済構造実態調査
- ＞ **家計調査**
- ＞ 全国家計構造調査（旧全国消費実態調査）
- ＞ 小売物価統計調査
- ＞ 全国物価統計調査
- ＞ 産業連関表
- ＞ 法人企業統計調査
- ＞ 経済産業省企業活動基本調査

＞ すべて見る (87 調査)

家計調査	データベース	件数 更新日	ファイル	件数 更新日	概要
家計調査		112件 2023-09-05		23,545件 2023-09-05	

家計調査	公開（更新）日
------	---------

■家計収支編 [85件]

— 二人以上の世帯 [55件]

— 月次 [16件]

2023-09-05

— 四半期 [9件]

2023-08-08

— 年次 [30件]

2023-02-07

品目分類					
010	品目分類（2020年改定）（総数：金額）	-	2023-09-05		
011	品目分類（2020年改定）（総数：数量）	-	2023-09-05		

家計データ

項番	事項名	説明	選択/全項目 ?	
1/5	表章項目		1/1	項目を選択
2/5	品目分類（2020年改定）	ガス代・携帯電話料金・ アイスクリーム	3/690	項目を選択
3/5	世帯区分	2人以上世帯	1/4	項目を選択
4/5	地域区分	東京都区部	1/53	項目を選択
5/5	時間軸（月次）	全部	463/463	項目を選択

初期状態に戻す（全項目表示）

キャンセル

確定

家計データ

11									/品目分類 (2	10800150	30200000	70300030	
12	表章項目 コー	表章項目	世帯区分 コー	世帯区分	地域区分 コー	地域区分	時間軸 (月次	時間軸 (月次	/品目分類 (2	356 アイス	3.2 ガス代【円	763 携帯電話通信料【円】	
13	1	金額	3	二人以上の世	13003	13100 東京都	2007000909	2007年9月		831	3,460	5,474	
14	1	金額	3	二人以上の世	13003	13100 東京都	2007001010	2007年10月		553	3,994	7,227	
15	1	金額	3	二人以上の世	13003	13100 東京都	2007001111	2007年11月		370	4,695	5,139	
16	1	金額	3	二人以上の世	13003	13100 東京都	2007001212	2007年12月		525	6,490	5,098	
17	1	金額	3	二人以上の世	13003	13100 東京都	2008000101	2008年1月		368	8,592	7,005	
18	1	金額	3	二人以上の世	13003	13100 東京都	2008000202	2008年2月		325	10,066	5,704	
19	1	金額	3	二人以上の世	13003	13100 東京都	2008000303	2008年3月		472	9,496	5,584	
20	1	金額	3	二人以上の世	13003	13100 東京都	2008000404	2008年4月		556	7,760	5,768	
21	1	金額	3	二人以上の世	13003	13100 東京都	2008000505	2008年5月		643	7,223	5,370	
22	1	金額	3	二人以上の世	13003	13100 東京都	2008000606	2008年6月		759	5,825	6,677	
23	1	金額	3	二人以上の世	13003	13100 東京都	2008000707	2008年7月		1,188	5,152	6,550	
24	1	金額	3	二人以上の世	13003	13100 東京都	2008000808	2008年8月		1,206	4,119	6,479	
25	1	金額	3	二人以上の世	13003	13100 東京都	2008000909	2008年9月		730	3,779	7,300	
26	1	金額	3	二人以上の世	13003	13100 東京都	2008001010	2008年10月		549	4,357	6,930	
27	1	金額	3	二人以上の世	13003	13100 東京都	2008001111	2008年11月		440	5,287	6,542	
28	1	金額	3	二人以上の世	13003	13100 東京都	2008001212	2008年12月		497	6,836	7,693	
29	1	金額	3	二人以上の世	13003	13100 東京都	2008000101	2008年1月		370	4,695	5,139	

気温

国土交通省

 気象庁

Japan Meteorological Agency

ENGLISH

Other Languages

文字サイズ変更

標準

大

気象庁防災情報

Twitter

気象庁

Twitter

気象庁

短波・解説

気象庁

YouTube

Google 提供

検索

ホーム

防災情報

各種データ・資料

地域の気象

解説

各種申請・ご案内

気象

気象観測データ

最新の気象データ

過去の気象データ検索

過去の地点気象データ・ダウンロード

過去の地域平均気象データ検索



東京都全地点

埼玉県

茨城県

山梨県

神奈川県

千葉県

東京

東京

データの種類

☐ 時別値

☐ 日別値

☐ 2日別値

☐ 半旬別値

☐ 旬別値

☒ 月別値

☐ 3か月別値※

項目

気温

降水

日

☒ 月平均気温

☐ 日最高気温の月平均

☐ 日最低気温の月平均

☐ 月最高気温

☐ 月最低気温

☐ 日最高気温の月最低※

期間

☒ 連続した期間で表示する

最近1年

2007年9月から

2023年7月までの月別値を表示

☐ 特定の期間を複数年分、表示する

8月から7月の値を

2007年から2023年まで表示

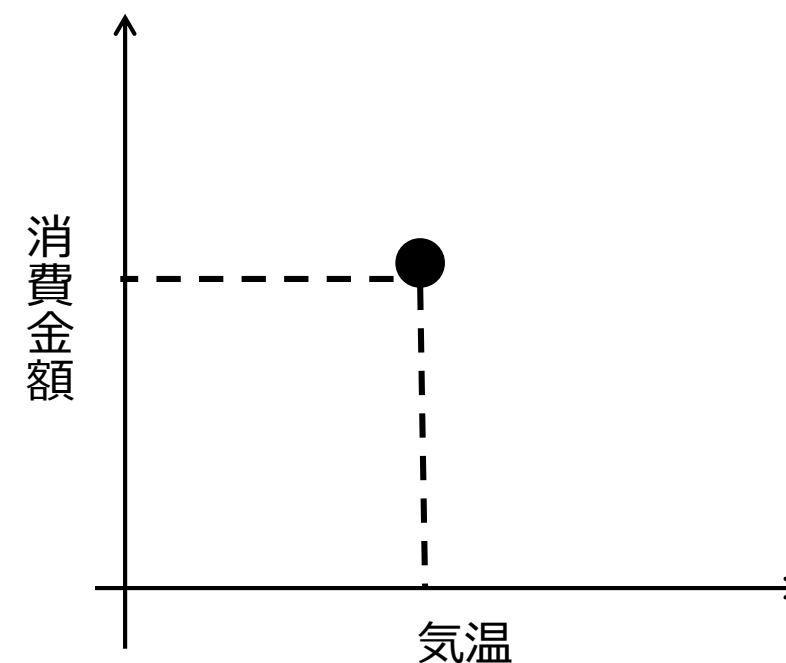
2007年9月 ~ 2023年7月

月別値一月平均気温

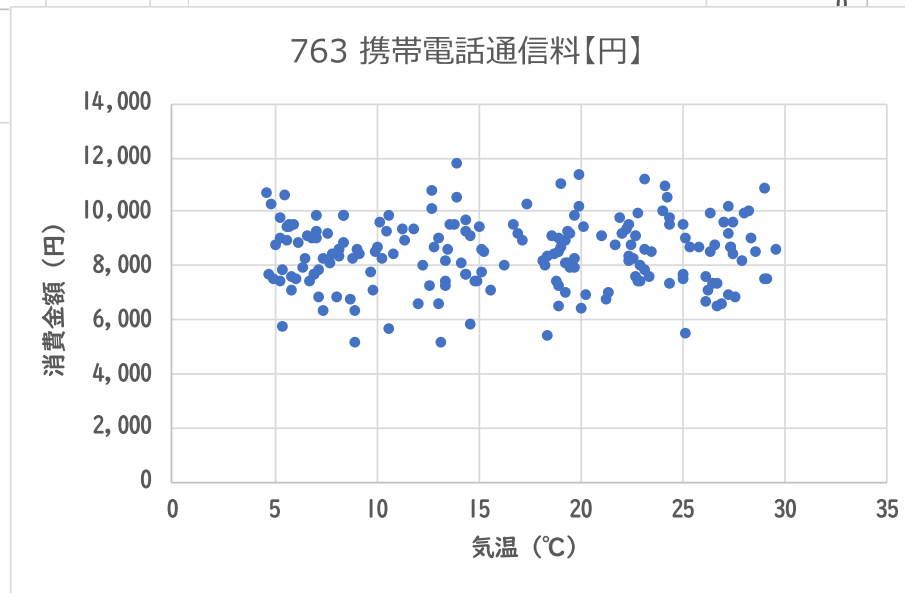
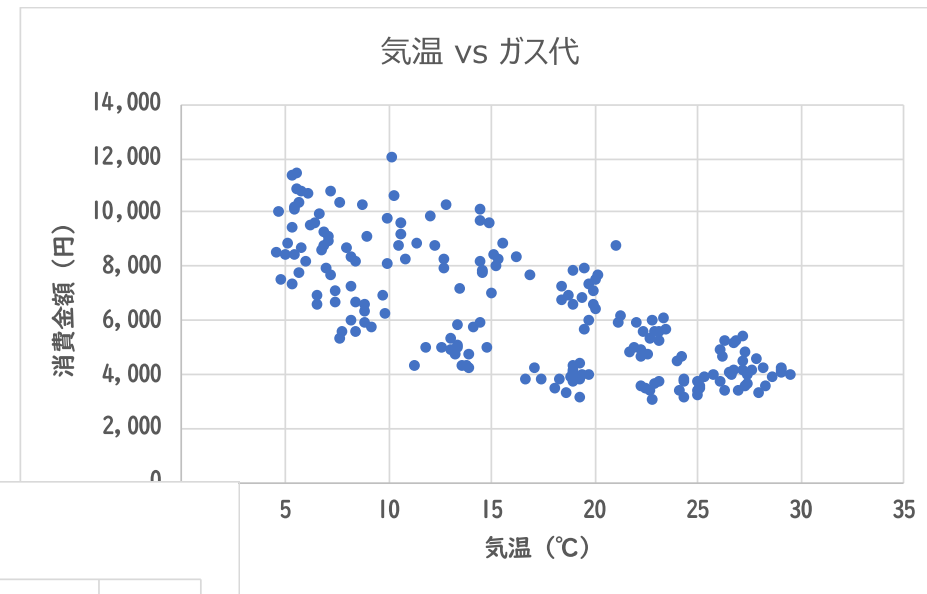
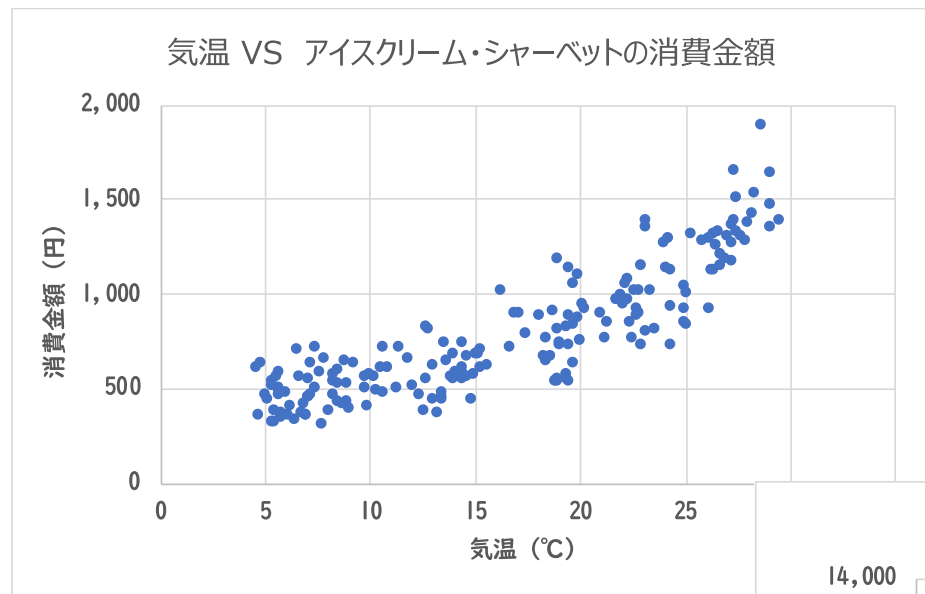
気温と消費支出

	東京	10800150	30200000	70300030
時間軸（月次）	平均気温(°C)	356 アイスクリーム代【円】	3.2 ガス代【円】	763 携帯電話通信料【円】
2007年9月	25.2	831	3,460	5,474
2007年10月	19.4	553	3,994	7,227
2007年11月	13.3	370	4,695	5,139
2007年12月	9.8	525	6,490	5,098
2008年1月	5.9	368	8,592	7,005
2008年2月	5.5	325	10,066	5,704
2008年3月	10.7	472	9,496	5,584
2008年4月	14.7	556	7,760	5,768
2008年5月	18.5	643	7,223	5,370
2008年6月	21.3	759	5,825	6,677
2008年7月	27.0	1,188	5,152	6,550
2008年8月	26.8	1,206	4,119	6,479
2008年9月	24.4	730	3,779	7,300
2008年10月	19.4	549	4,357	6,930
2008年11月	13.1	440	5,287	6,542
2008年12月	9.8	497	6,836	7,693

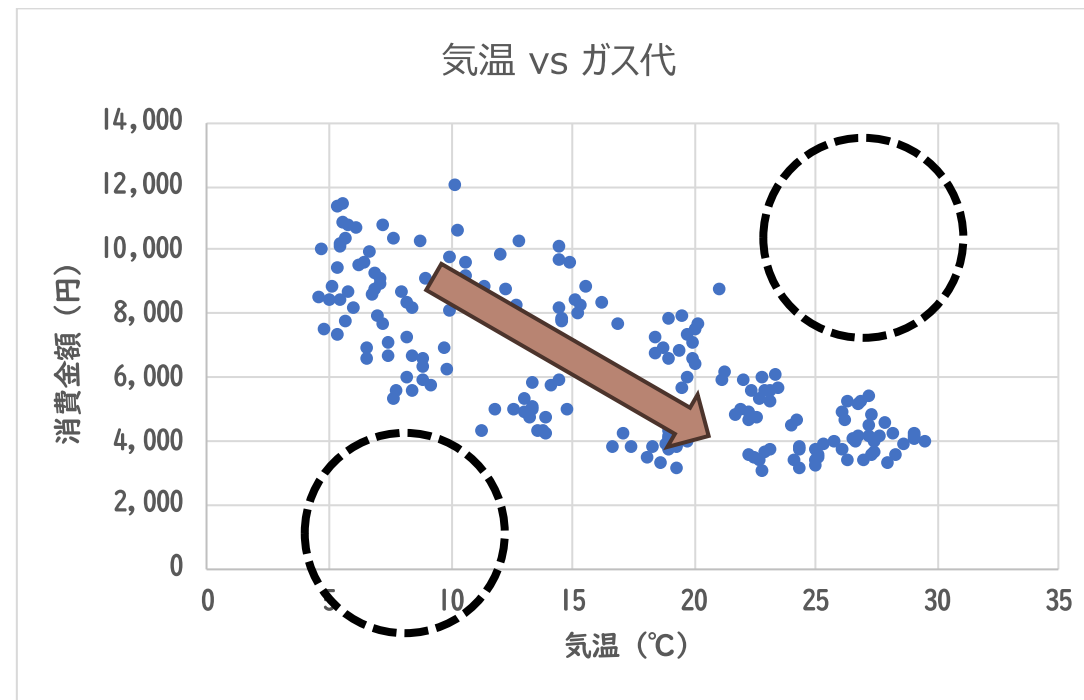
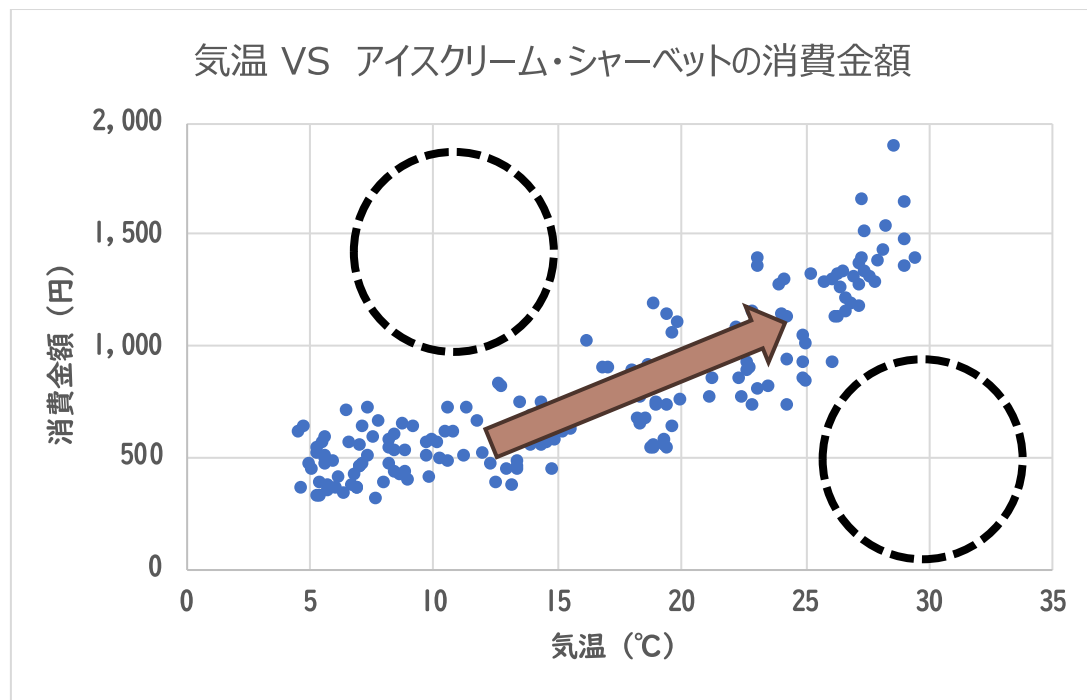
散布図



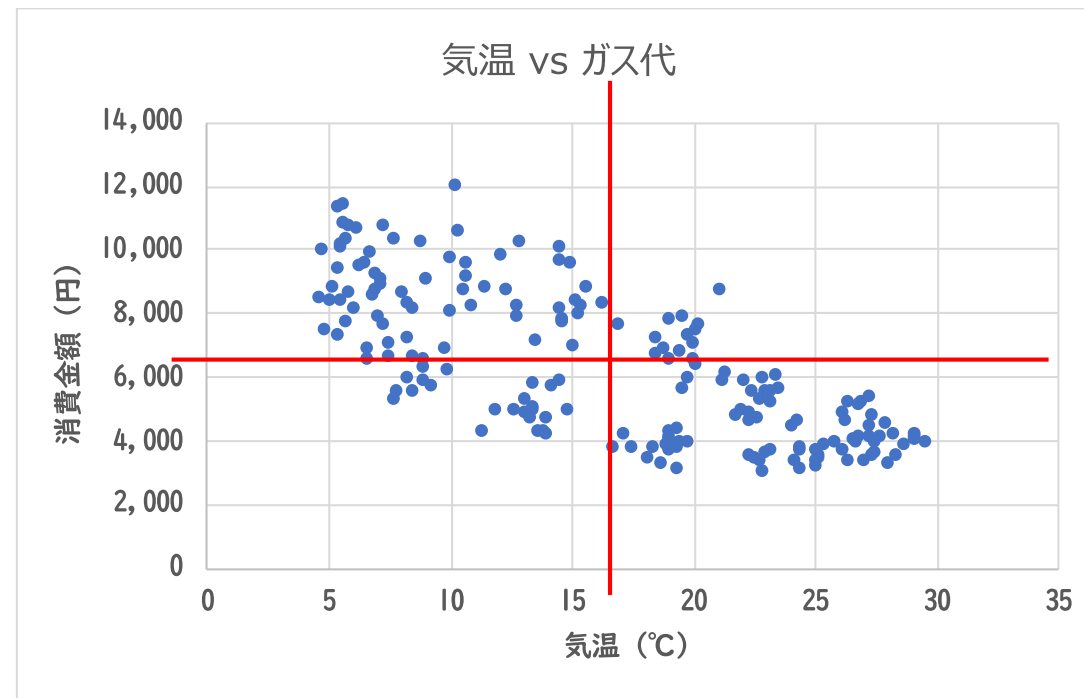
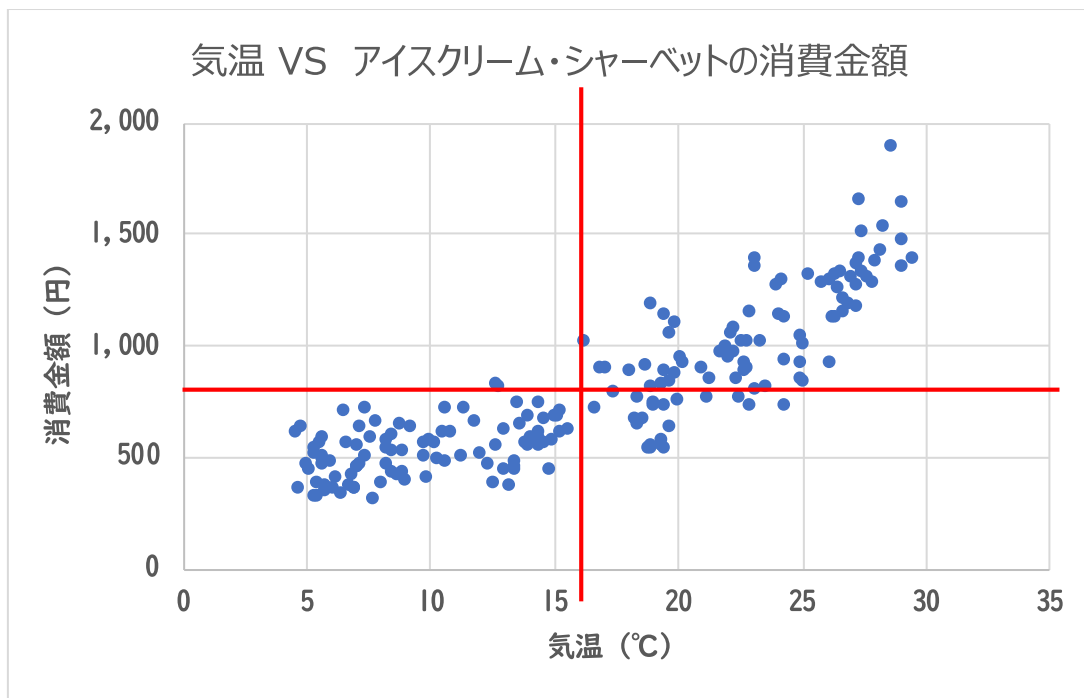
気温vs消費金額



右上がり・右下がり

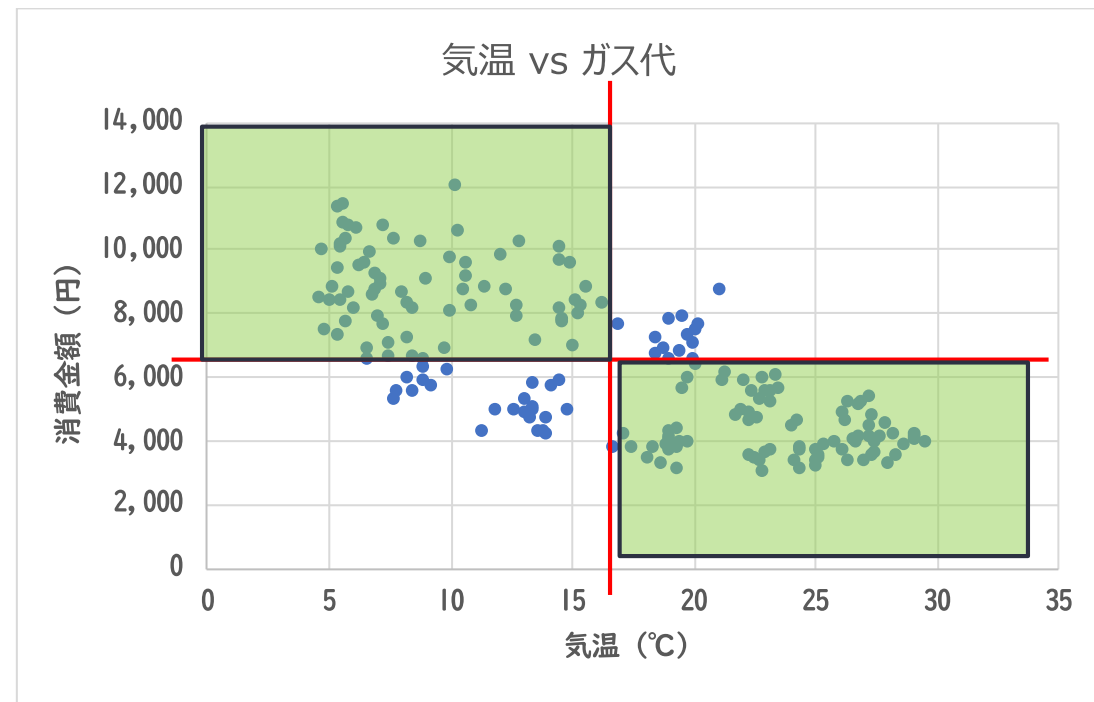
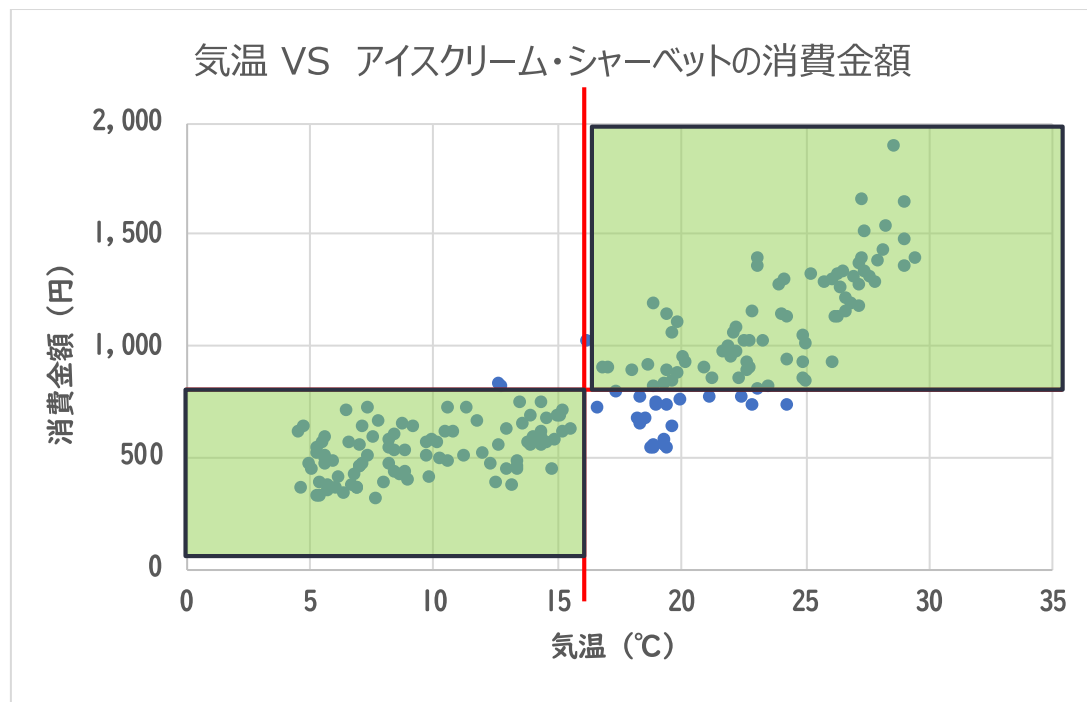


右上がり・右下がり



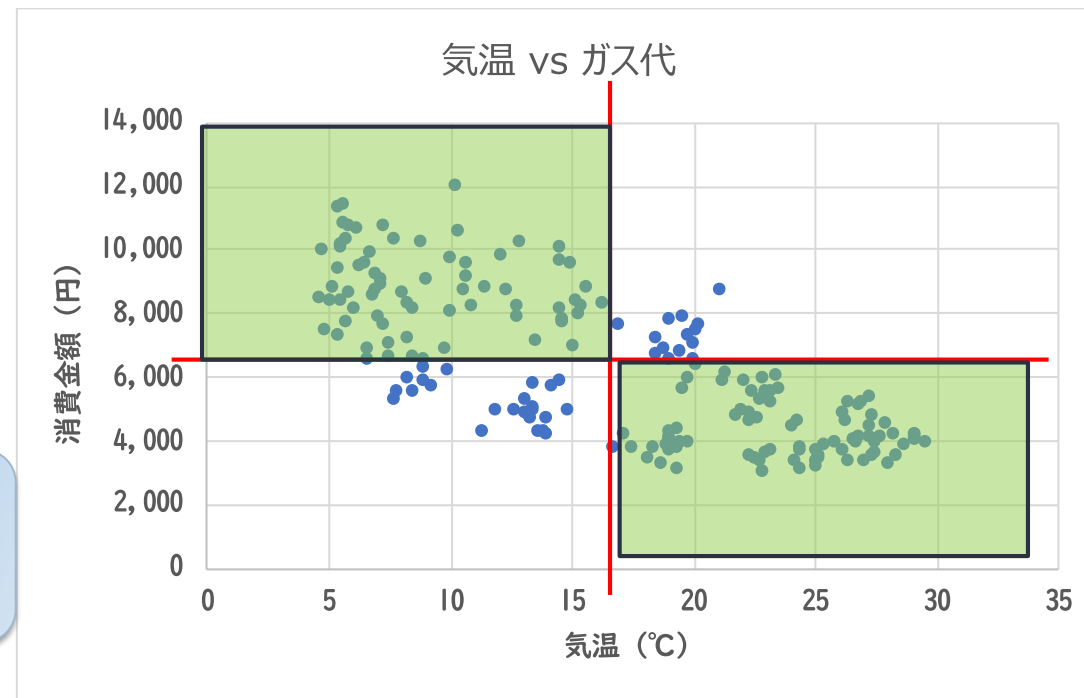
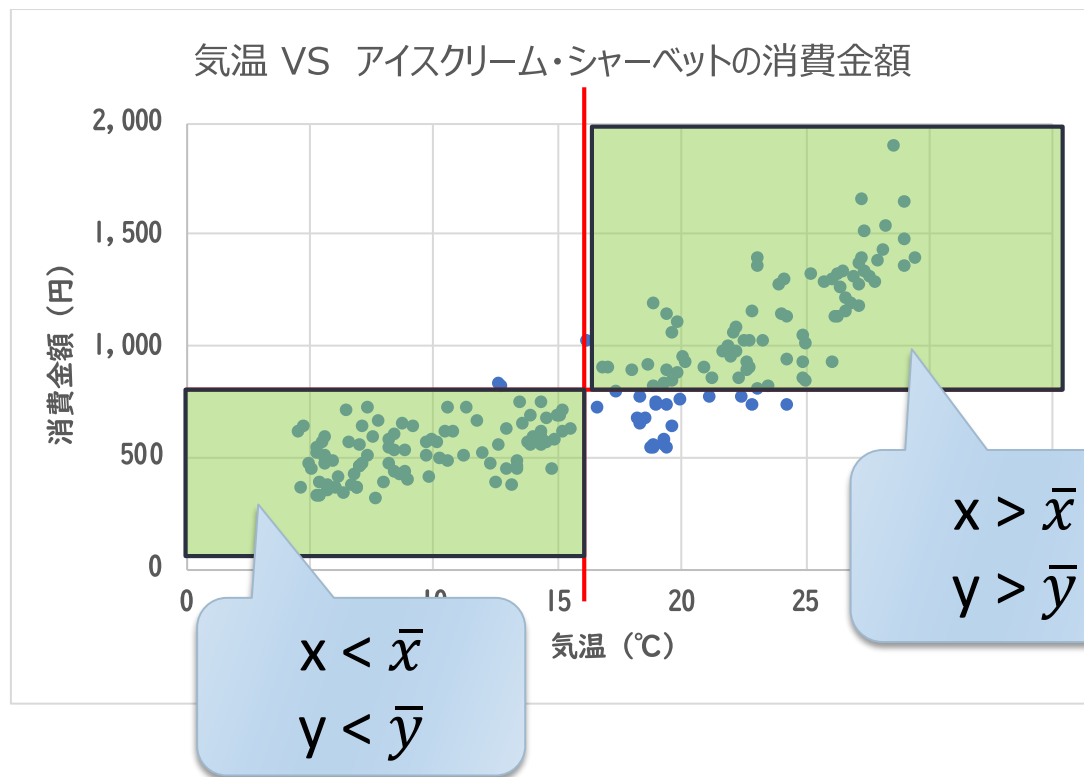
	平均気温(°C)	アイスクリーム	ガス代	携帯電話通信料
平均値	16.51	781.19	6285.93	8378.07

右上がり・右下がり



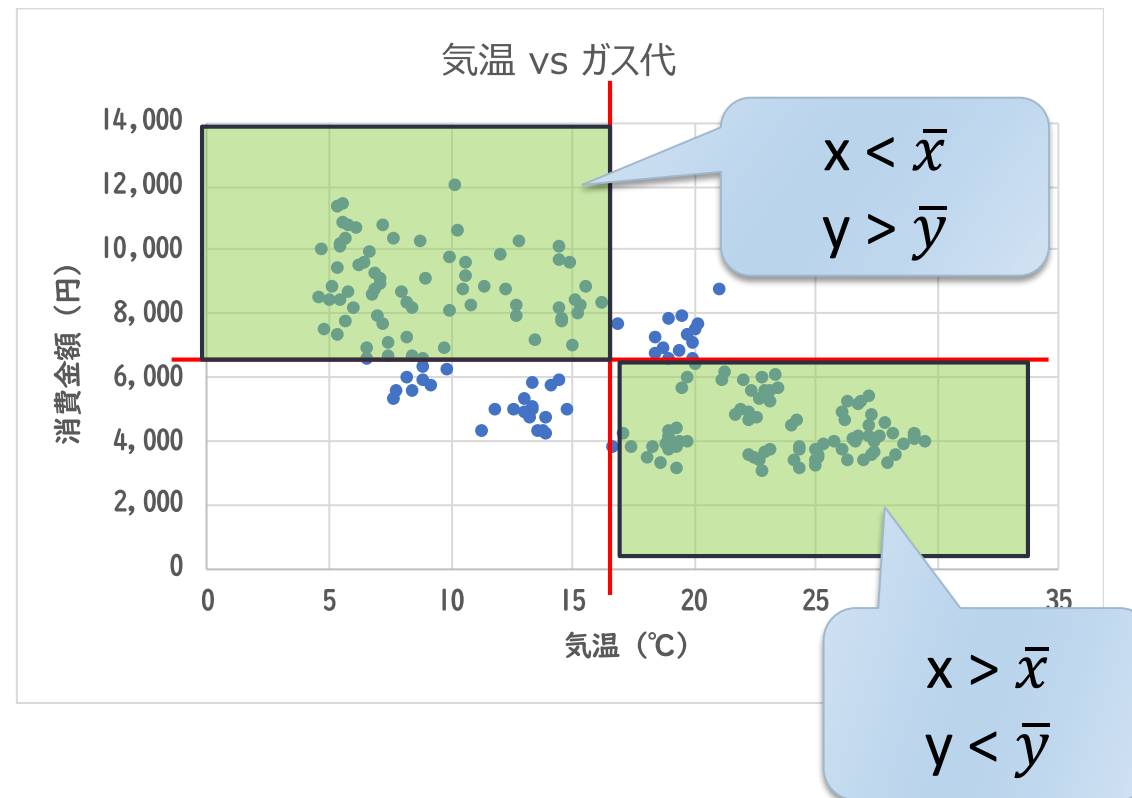
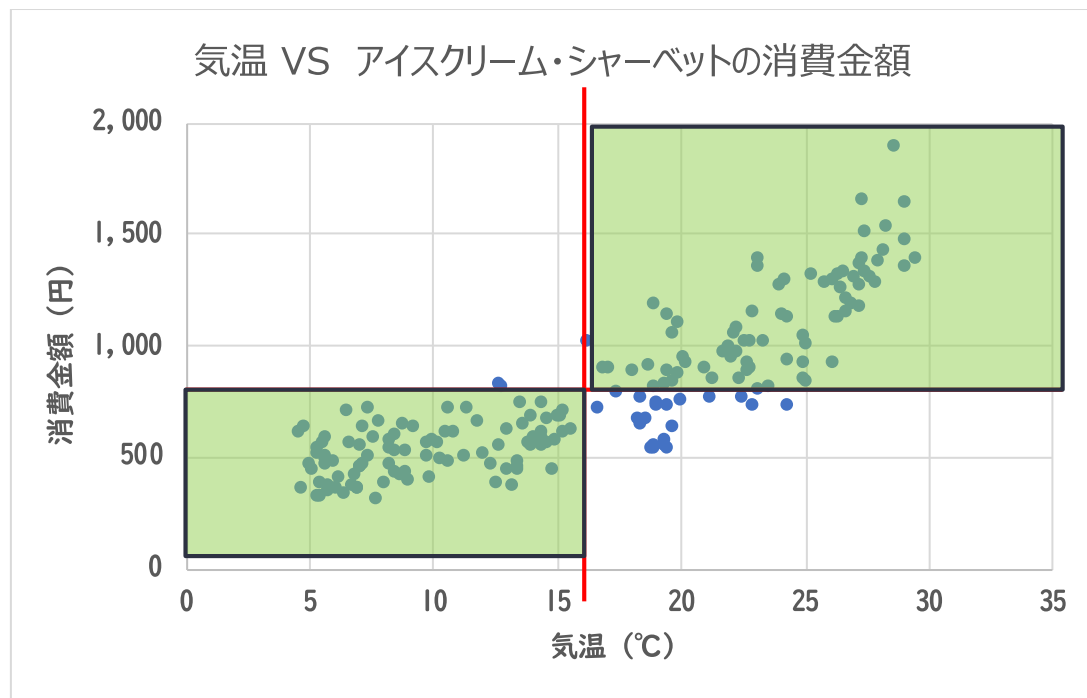
	平均気温(°C)	アイスクリーム	ガス代	携帯電話通信料
平均値	16.51	781.19	6285.93	8378.07

右上がり・右下がり



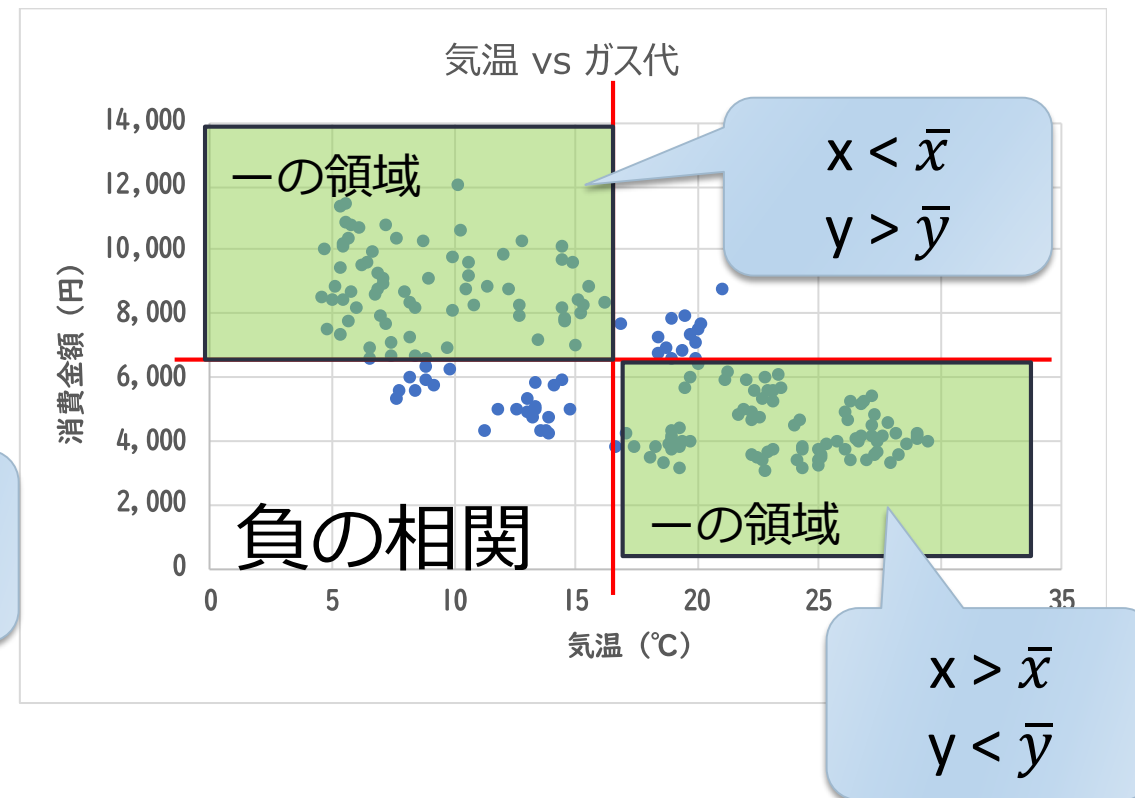
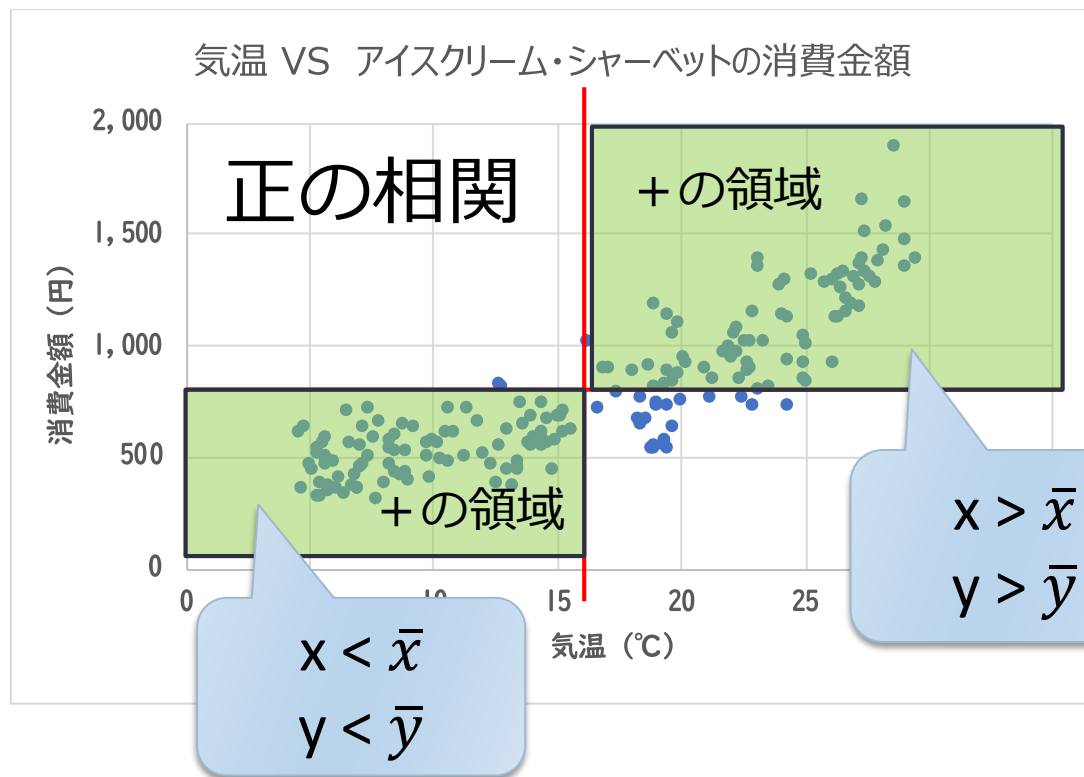
	平均気温(°C)	アイスクリーム	ガス代	携帯電話通信料
平均値	16.51	781.19	6285.93	8378.07

右上がり・右下がり



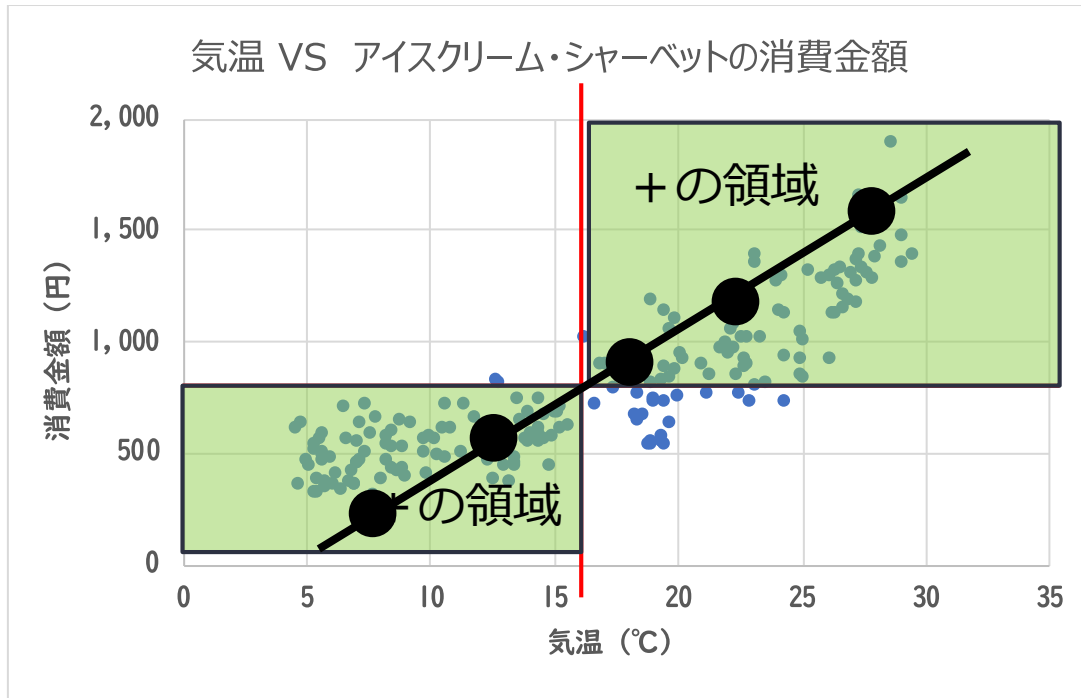
	平均気温(°C)	アイスクリーム	ガス代	携帯電話通信料
平均値	16.51	781.19	6285.93	8378.07

右上がり・右下がり

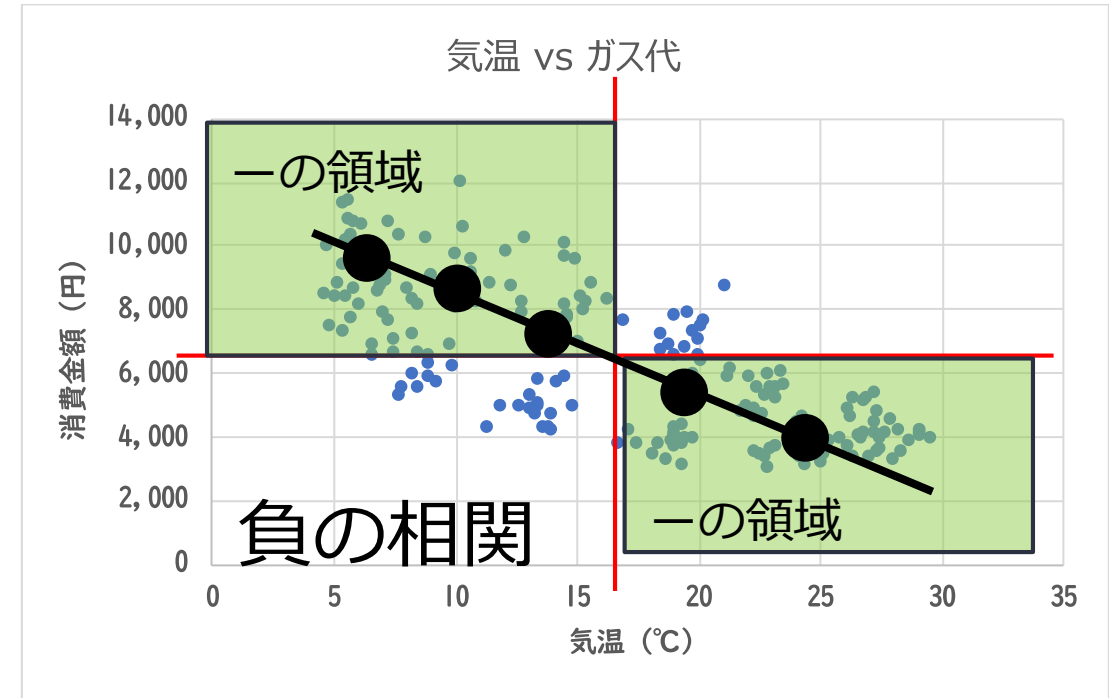


	平均気温(°C)	アイスクリーム	ガス代	携帯電話通信料
平均値	16.51	781.19	6285.93	8378.07

直線のあてはまり



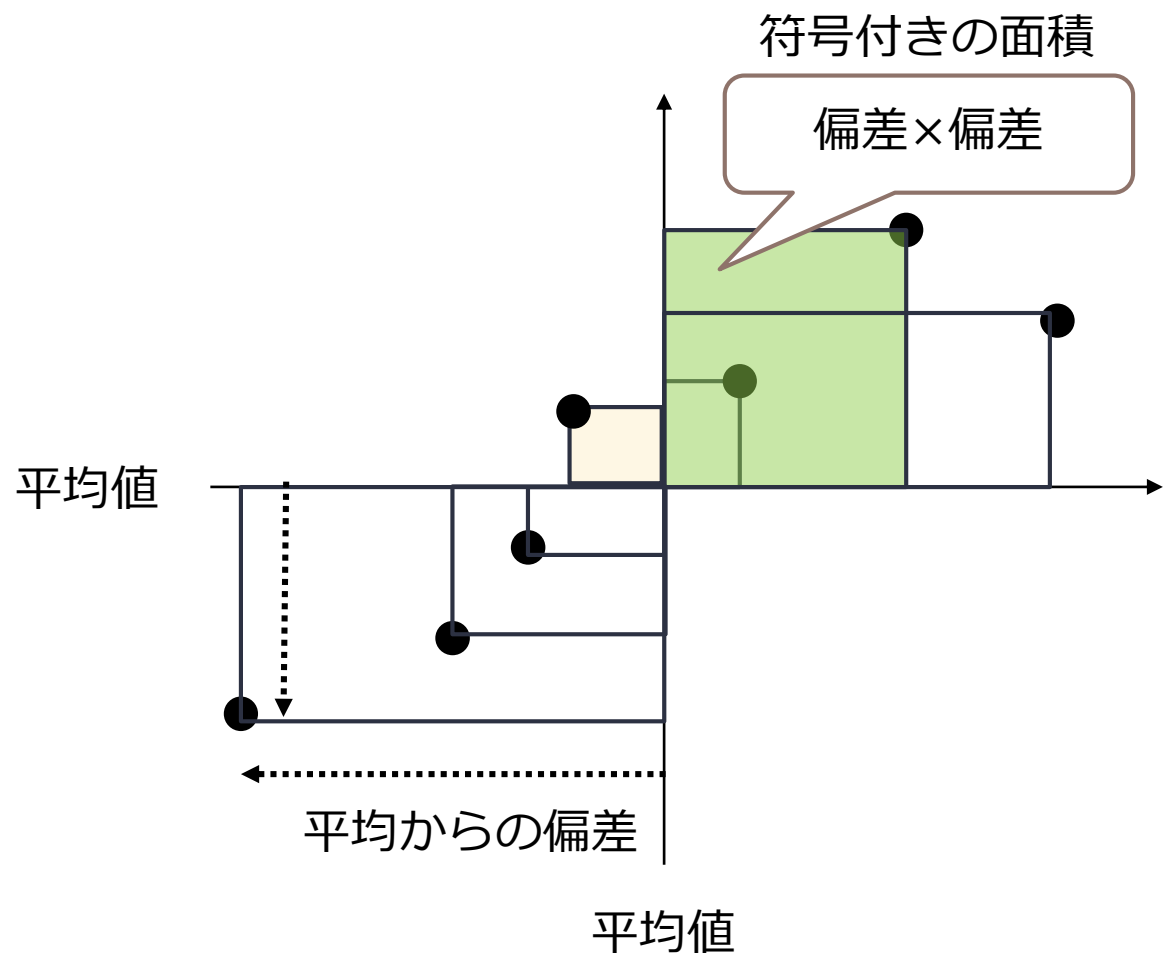
完全な正の相関



完全な負の相関

	平均気温(°C)	アイスクリーム	ガス代	携帯電話通信料
平均値	16.51	781.19	6285.93	8378.07

各点の面積

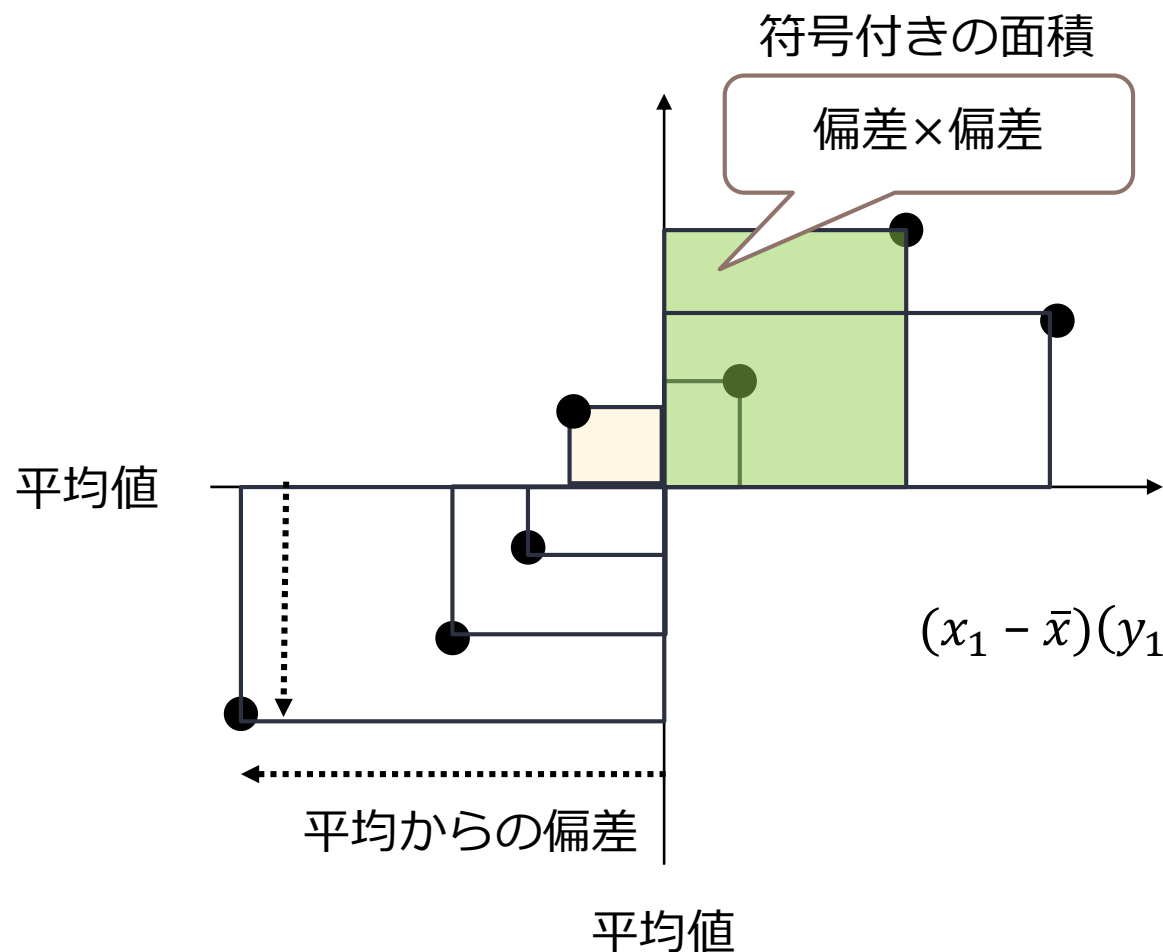


$$\tilde{x}_i = x_i - \bar{x}$$

$$\tilde{y}_i = y_i - \bar{y}$$

$$\tilde{x}_i \cdot \tilde{y}_i = (x_i - \bar{x})(y_i - \bar{y})$$

各点の面積



$$\tilde{x}_i = x_i - \bar{x}$$

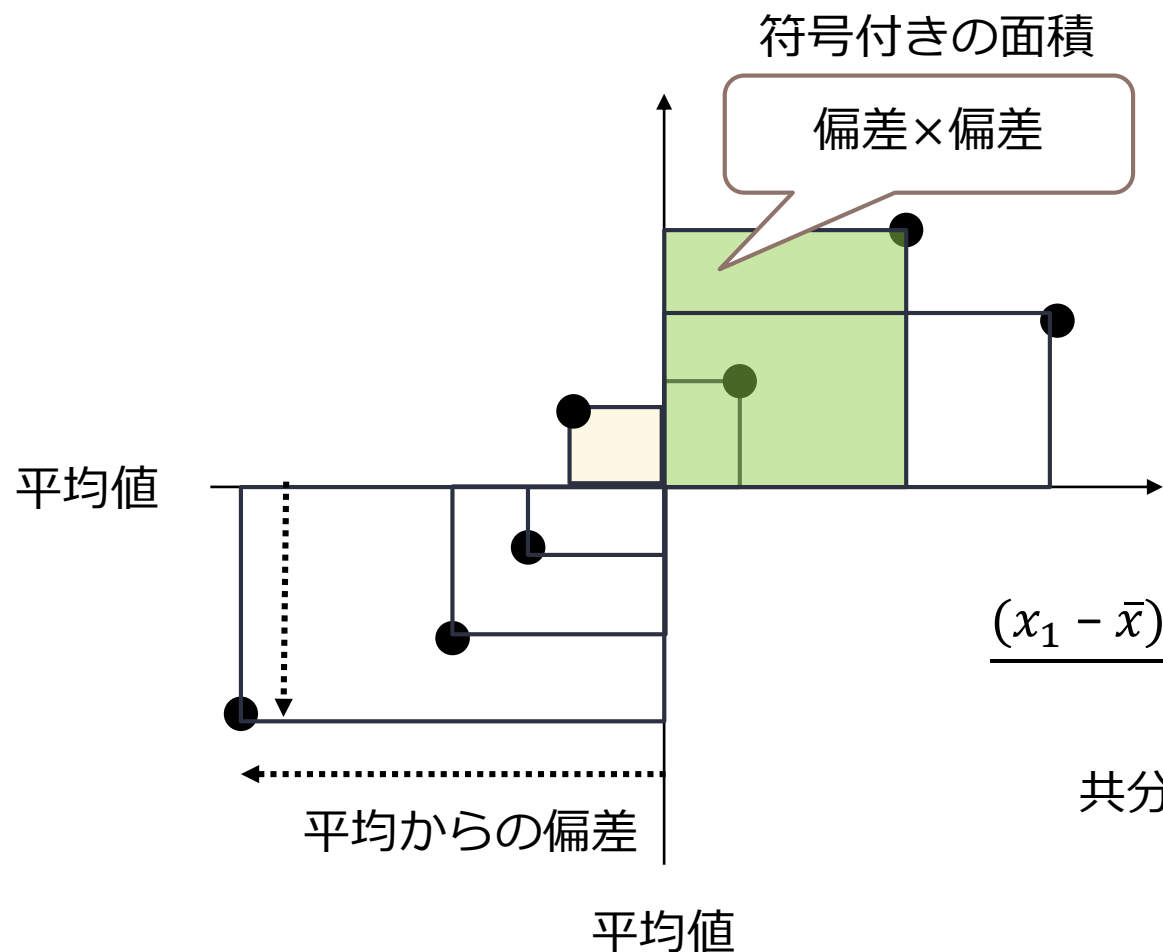
$$\tilde{y}_i = y_i - \bar{y}$$

$$\tilde{x}_i \cdot \tilde{y}_i = (x_i - \bar{x})(y_i - \bar{y})$$

$$(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})$$

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

各点の面積



$$\tilde{x}_i = x_i - \bar{x}$$

$$\tilde{y}_i = y_i - \bar{y}$$

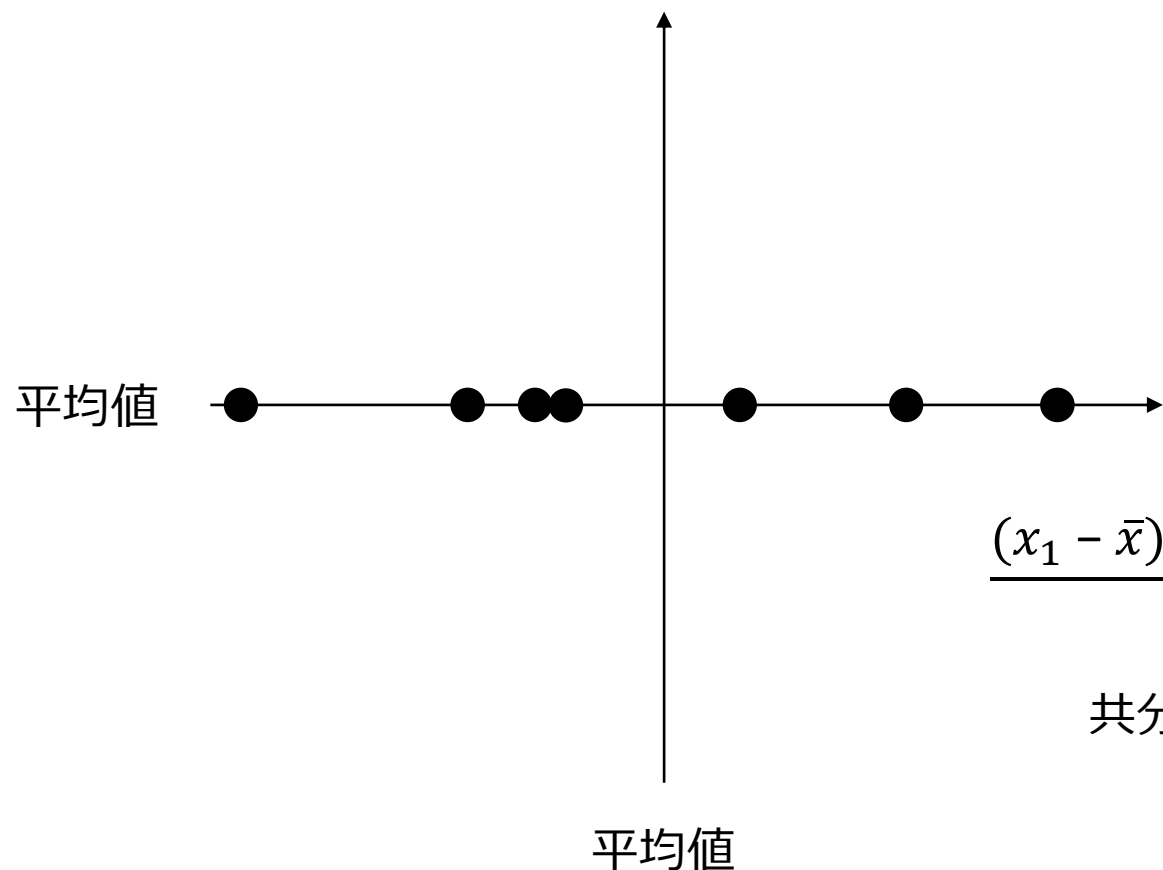
$$\tilde{x}_i \cdot \tilde{y}_i = (x_i - \bar{x})(y_i - \bar{y})$$

$$\frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})}{n}$$

共分散

$$v_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

共分散の性質



$$\tilde{x}_i = x_i - \bar{x}$$

$$\tilde{y}_i = y_i - \bar{y} = 0$$

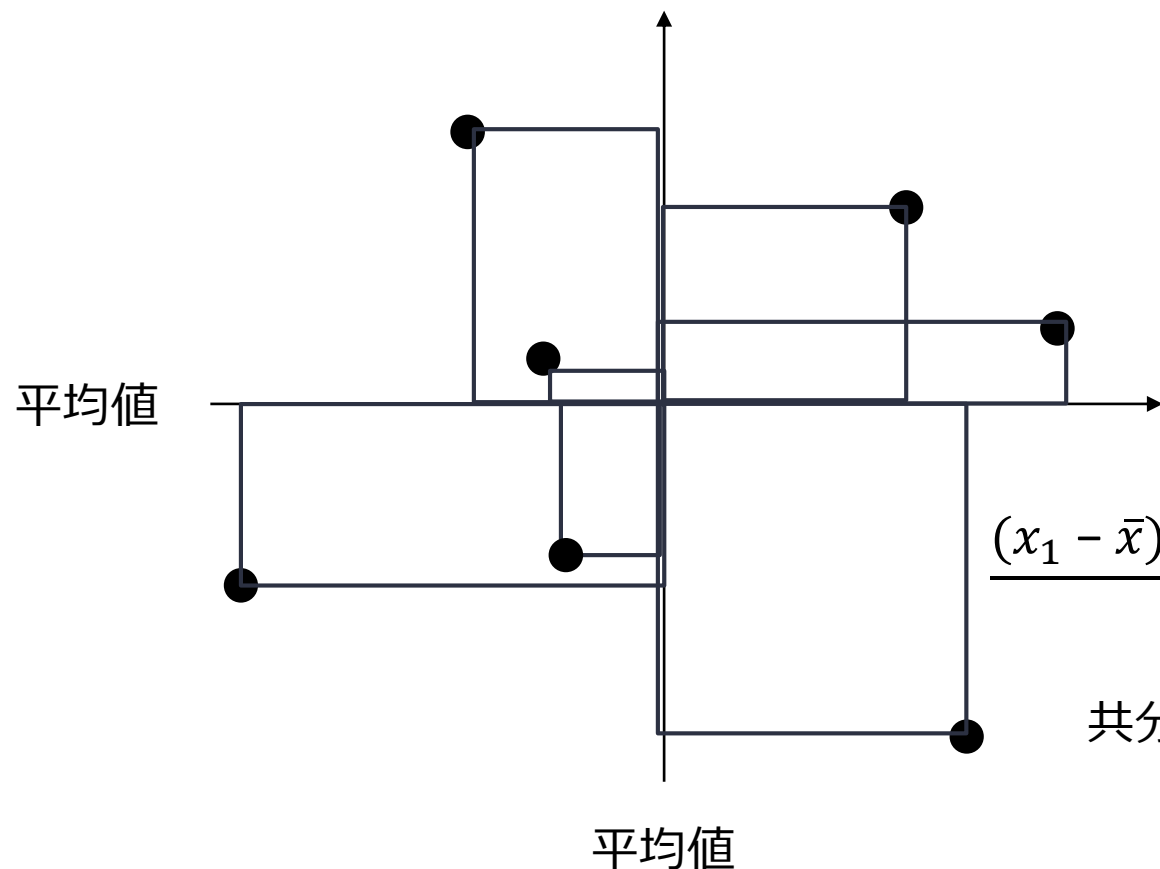
$$\tilde{x}_i \cdot \tilde{y}_i = (x_i - \bar{x})(y_i - \bar{y})$$

$$\frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})}{n}$$

共分散

$$v_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 0$$

共分散の性質



$$\tilde{x}_i = x_i - \bar{x}$$

$$\tilde{y}_i = y_i - \bar{y} = 0$$

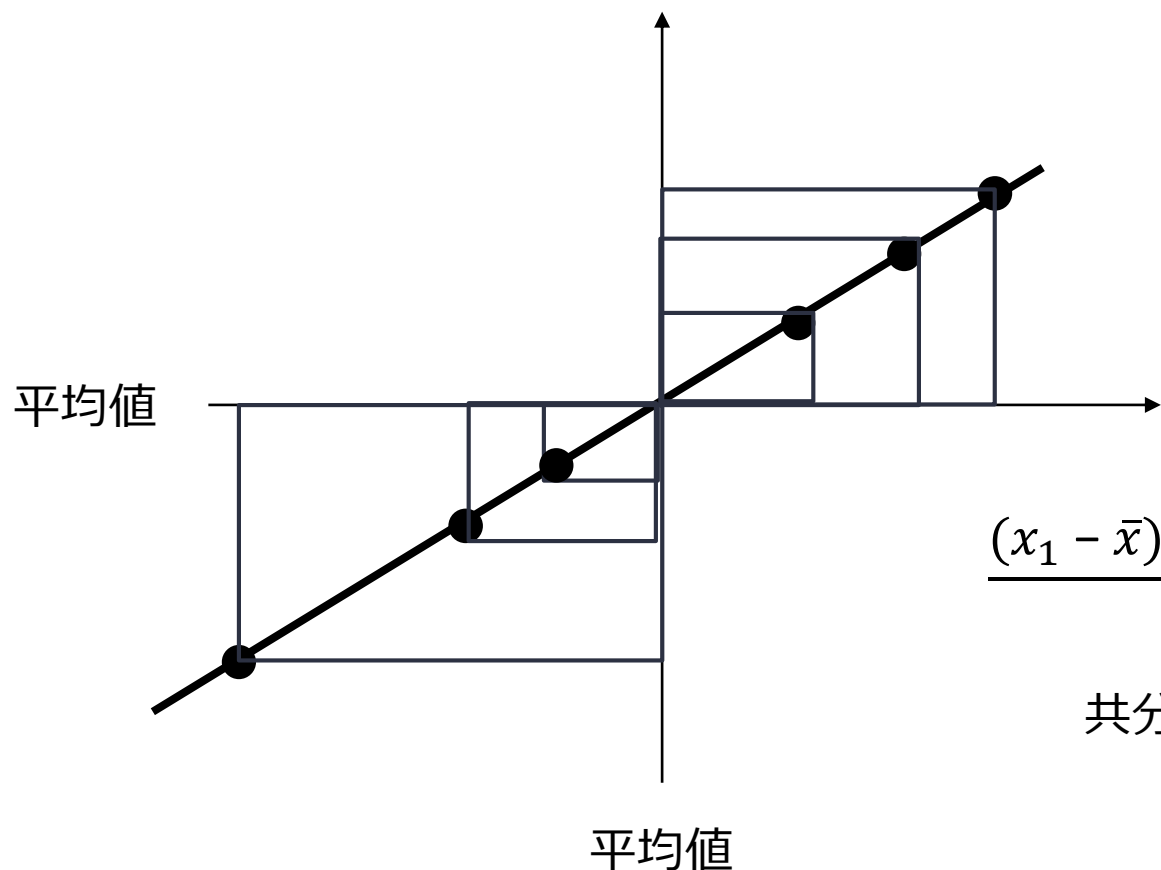
$$\tilde{x}_i \cdot \tilde{y}_i = (x_i - \bar{x})(y_i - \bar{y})$$

$$\frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})}{n}$$

共分散

$$v_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

共分散の性質



$$\tilde{x}_i = x_i - \bar{x}$$

$$\tilde{y}_i = y_i - \bar{y} = 0$$

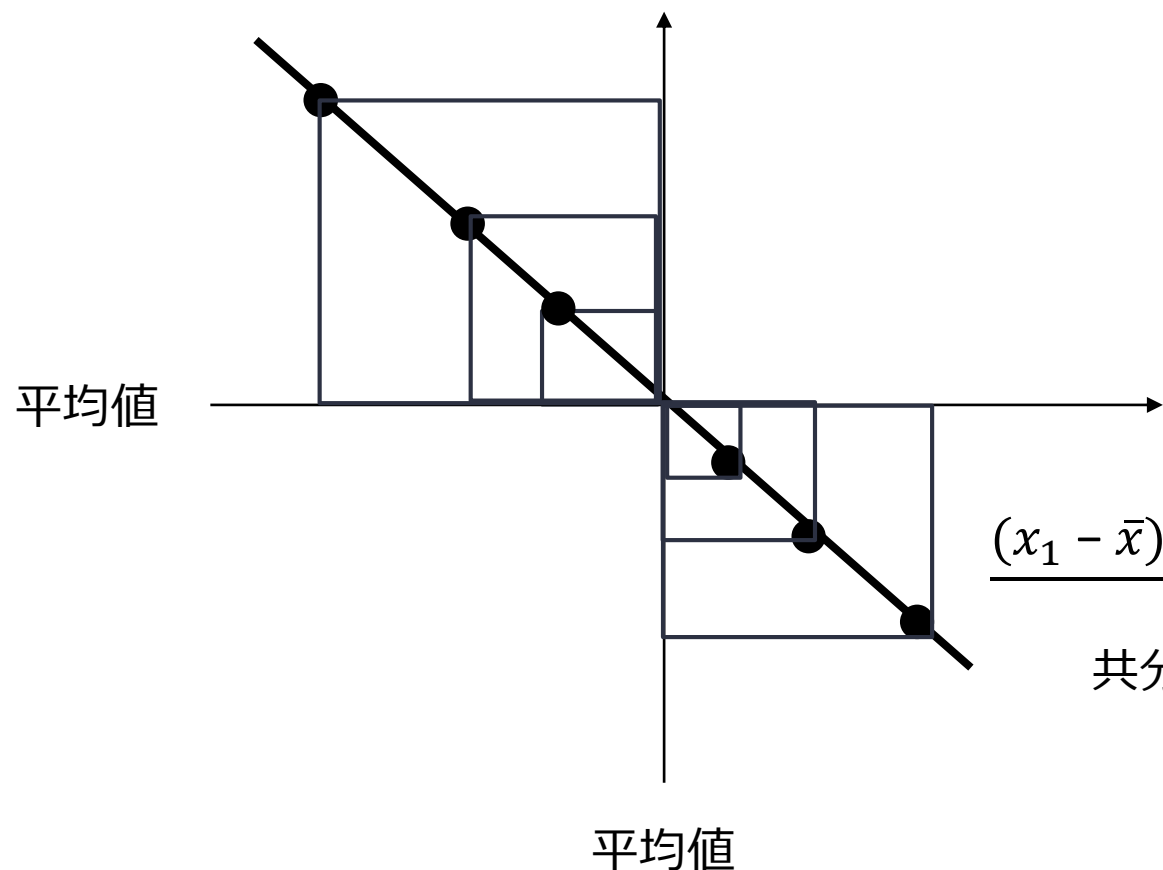
$$\tilde{x}_i \cdot \tilde{y}_i = (x_i - \bar{x})(y_i - \bar{y})$$

$$\frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})}{n}$$

共分散

$$v_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

共分散の性質



$$\tilde{x}_i = x_i - \bar{x}$$

$$\tilde{y}_i = y_i - \bar{y} = 0$$

$$\tilde{x}_i \cdot \tilde{y}_i = (x_i - \bar{x})(y_i - \bar{y})$$

$$\frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})}{n}$$

共分散

$$v_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

zスコアと偏差値（第9回スライド）

$$z_i = \frac{x_i - \mu}{\sigma}$$

μ : 平均
 σ : 標準偏差

$$\text{偏差値} = z_i \times 10 + 50$$

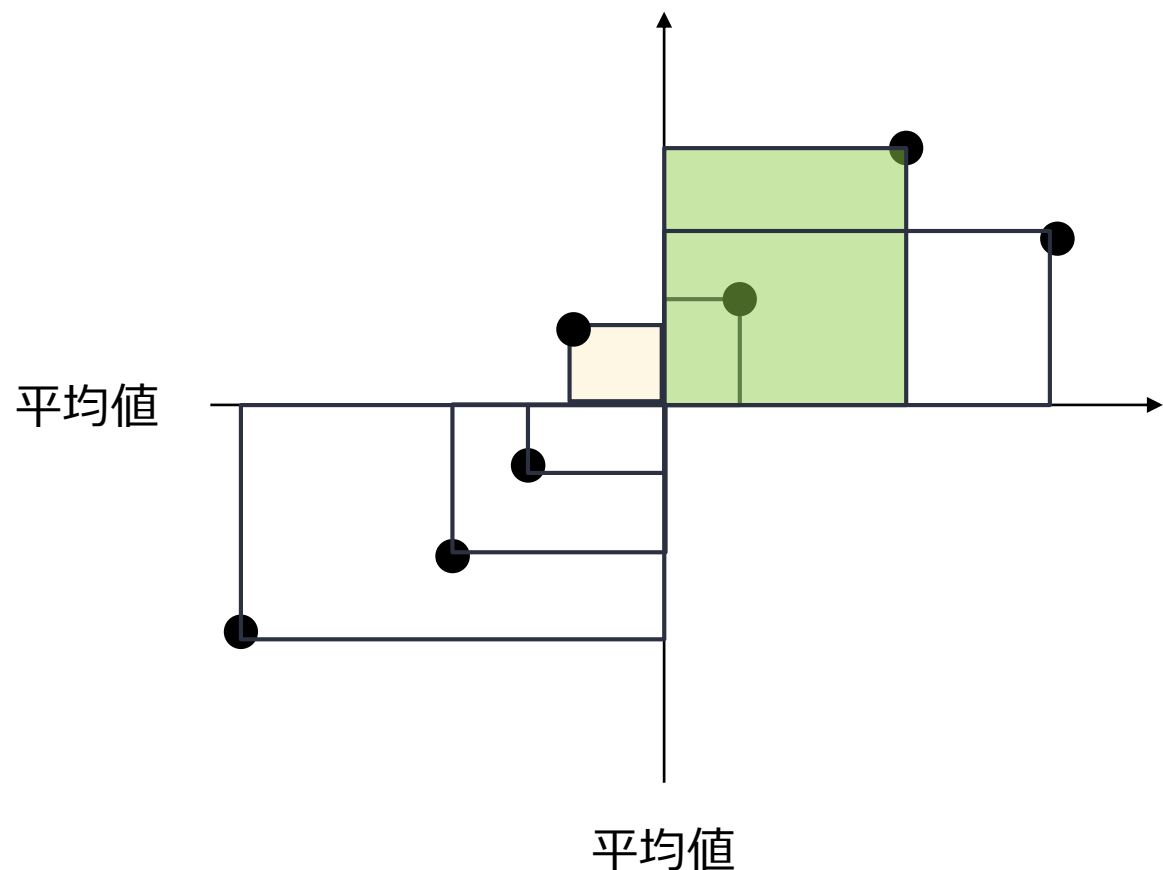
男子の握力25kg

握力	zスコア	偏差値
25	-0.54	44.55

女子の握力25kg

握力	zスコア	偏差値
25	0.39	53.90

相関係数



$$\frac{x_i - \bar{x}}{\sigma_x}$$

$$\frac{y_i - \bar{y}}{\sigma_y}$$

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$



標準化

$$v_{xy} = \frac{1}{n} \sum_{i=1}^n \frac{x_i - \bar{x}}{\sigma_x} \cdot \frac{y_i - \bar{y}}{\sigma_y}$$

相関係数

$$r = \frac{v_{xy}}{\sigma_x \sigma_y}$$

相関係数

標準偏差	7.51	330.74		相関係数	0.86257465	0.86257465
平均	16.51	781.19		共分散	2142.85094	
	東京	10800150				
時間軸（月次	平均気温(°C)	356 アイスク	xの偏差	yの偏差	積	
2007年9月	25.2	831	8.69	49.81	432.708226	
2007年10月	19	553	2.49	-228.19	-567.48444	
2007年11月	13.3	370	-3.21	-411.19	1321.18519	
2007年12月	9	525	-7.51	-256.19	1924.76686	
2008年1月	5.9	368	-10.61	-413.19	4385.20613	
2008年2月	5.5	325	-11.01	-456.19	5024.04435	
2008年3月	10.7	472	-5.81	-309.19	1797.34016	
2008年4月	14.7	556	-1.81	-225.19	408.28676	
2008年5月	18.5	643	1.99	-138.19	-274.56821	
2008年6月	21.3	759	4.79	-22.19	-106.21429	

=CORREL(x , y)

相関係数のだいたいの目安

相関係数	だいたいの目安
0.7 ~ 1.0	かなり強い正の相関がある
0.4 ~ 0.7	正の相関がある
0.2 ~ 0.4	弱い正の相関がある
-0.2 ~ 0.2	ほとんど相関がない
-0.4 ~ -0.2	弱い負の相関がある
-0.7 ~ -0.4	負の相関がある
-1.0 ~ -0.7	かなり強い負の相関がある

なぜ可視化が必要なのか

アンスコムの四重奏

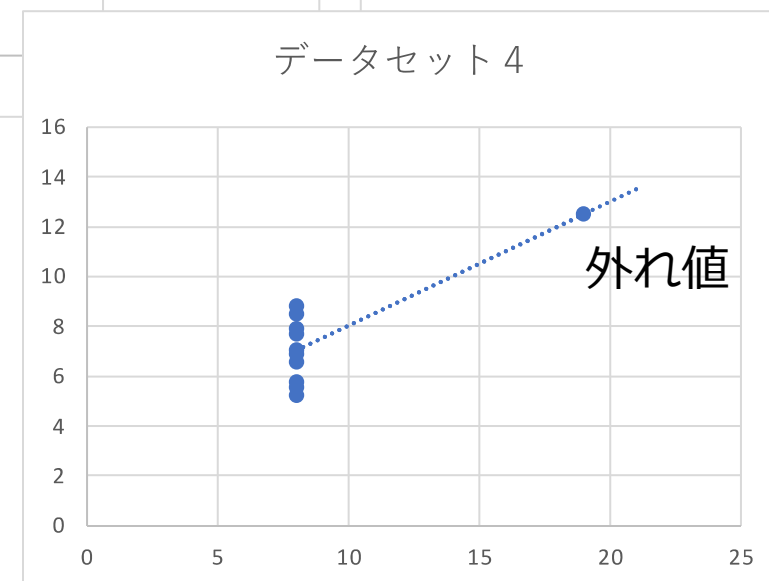
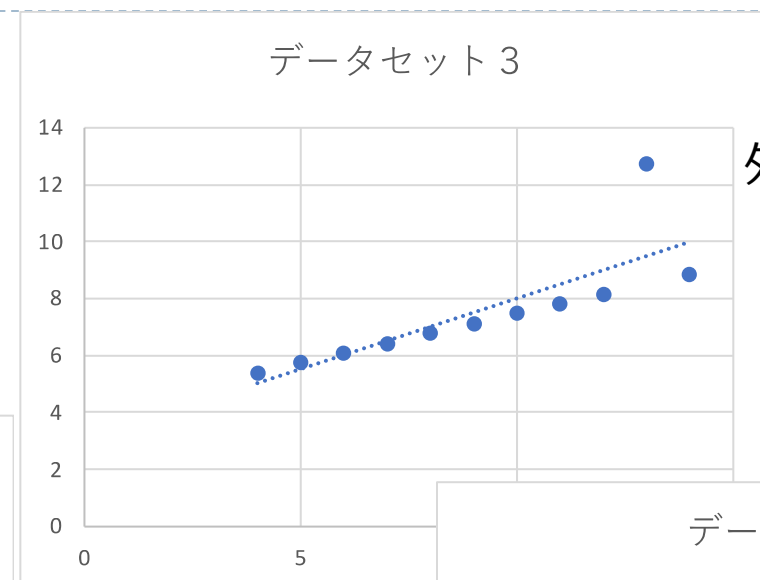
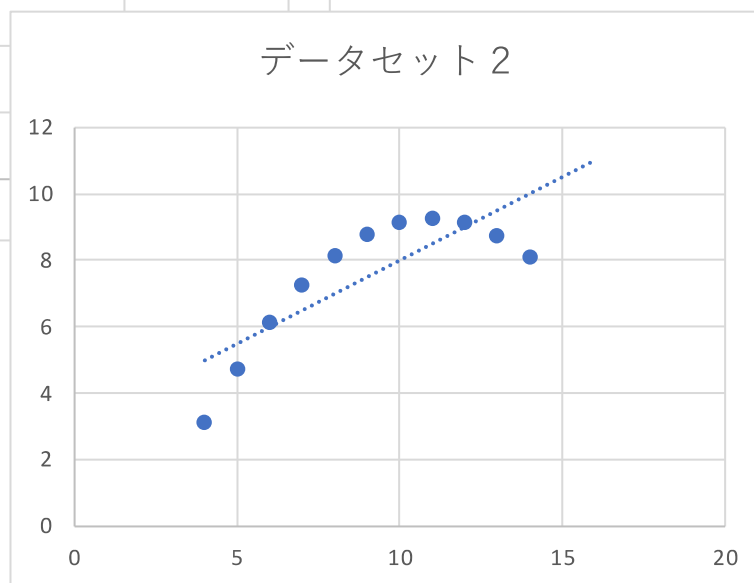
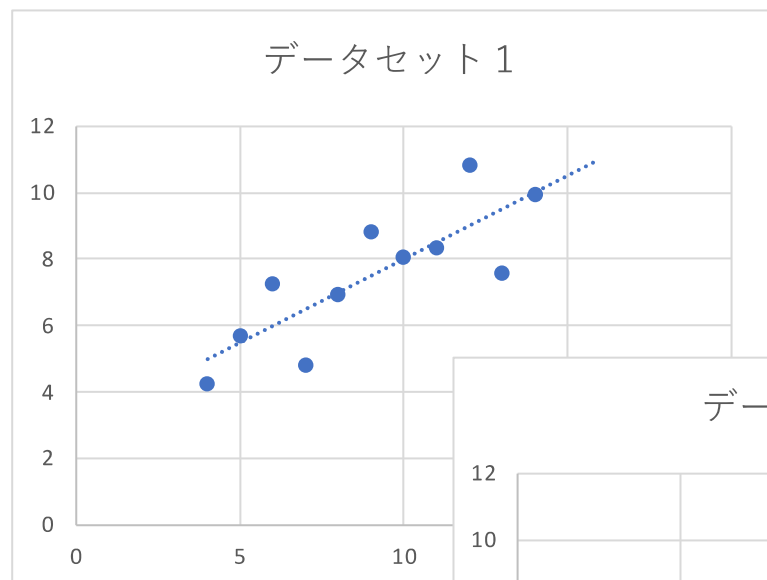
dataset	x	y
I	10	8.04
I	8	6.95
I	13	7.58
I	9	8.81
I	11	8.33
I	14	9.96
I	6	7.24
I	4	4.26
I	12	10.84
I	7	4.82
I	5	5.68
相関係数		

dataset	x	y
II	10	9.14
II	8	8.14
II	13	8.74
II	9	8.77
II	11	9.26
II	14	8.1
II	6	6.13
II	4	3.1
II	12	9.13
II	7	7.26
II	5	4.74
相関係数		

dataset	x	y
III	10	7.46
III	8	6.77
III	13	12.74
III	9	7.11
III	11	7.81
III	14	8.84
III	6	6.08
III	4	5.39
III	12	8.15
III	7	6.42
III	5	5.73
相関係数		

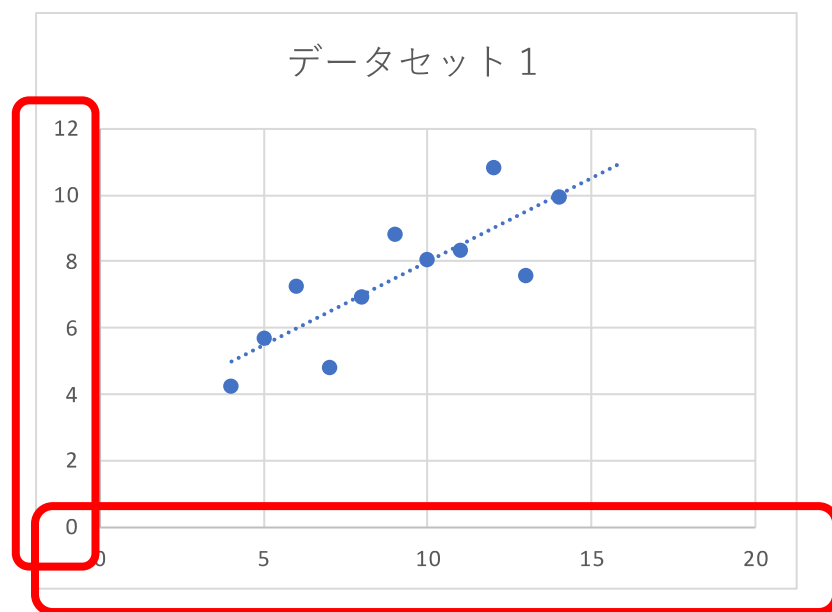
dataset	x	y
IV	8	6.58
IV	8	5.76
IV	8	7.71
IV	8	8.84
IV	8	8.47
IV	8	7.04
IV	8	5.25
IV	19	12.5
IV	8	5.56
IV	8	7.91
IV	8	6.89
相関係数		

なぜ可視化が必要なのか



相関関係と因果関係

▶ 相関関係



直線的な関係

▶ 因果関係

原因



結果

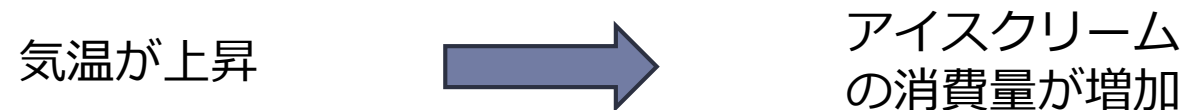
気温が上昇



アイスクリームの消費量が増加

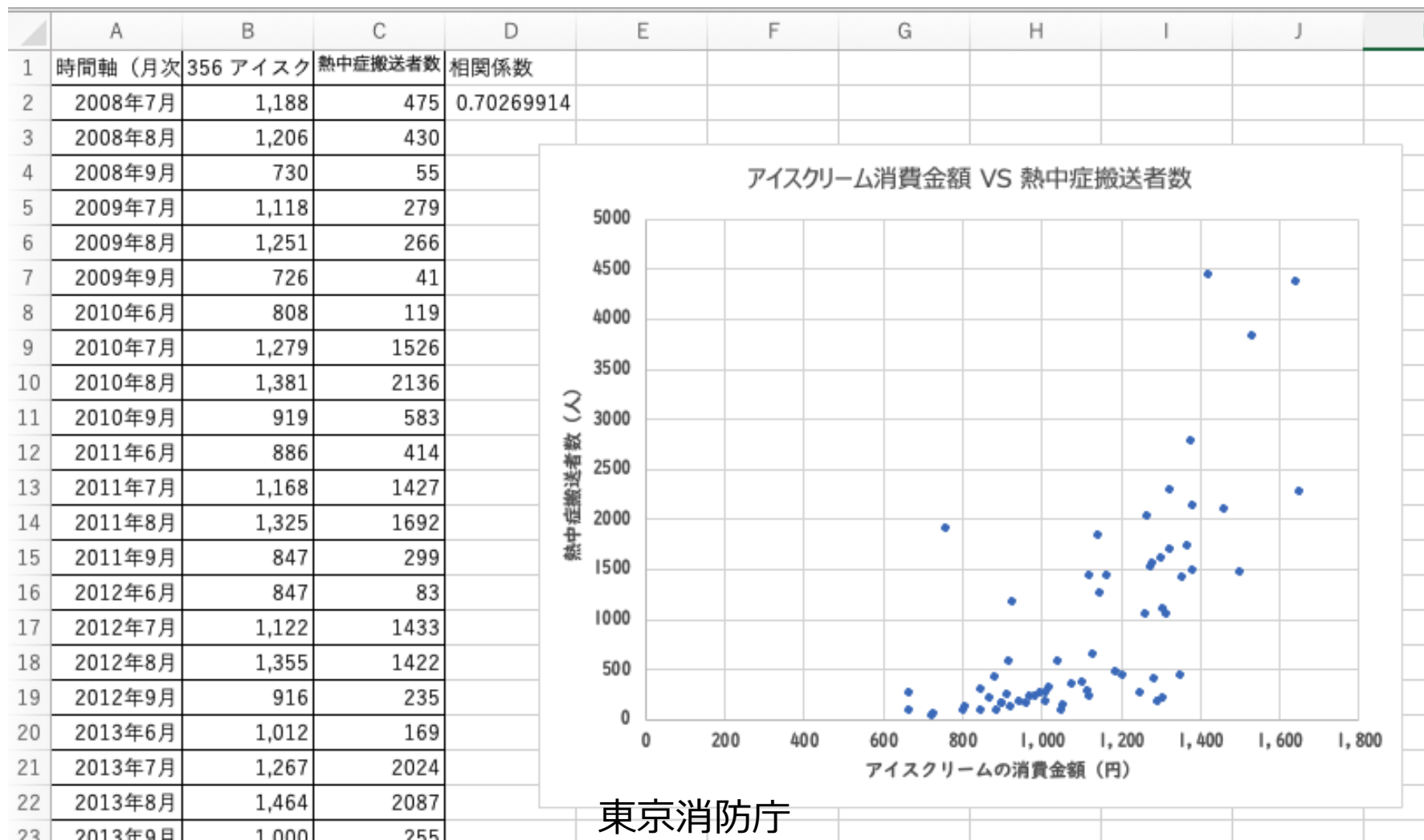
相関関係と因果関係

- ▶ 因果関係があれば相関関係がある



- ▶ 相関関係があっても因果関係があるとは限らない

相関関係と因果関係

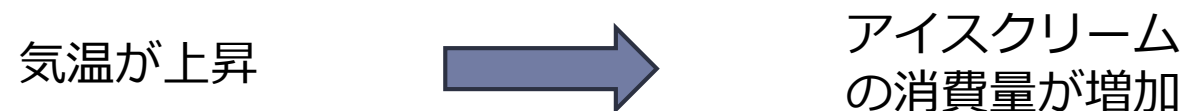


東京消防庁

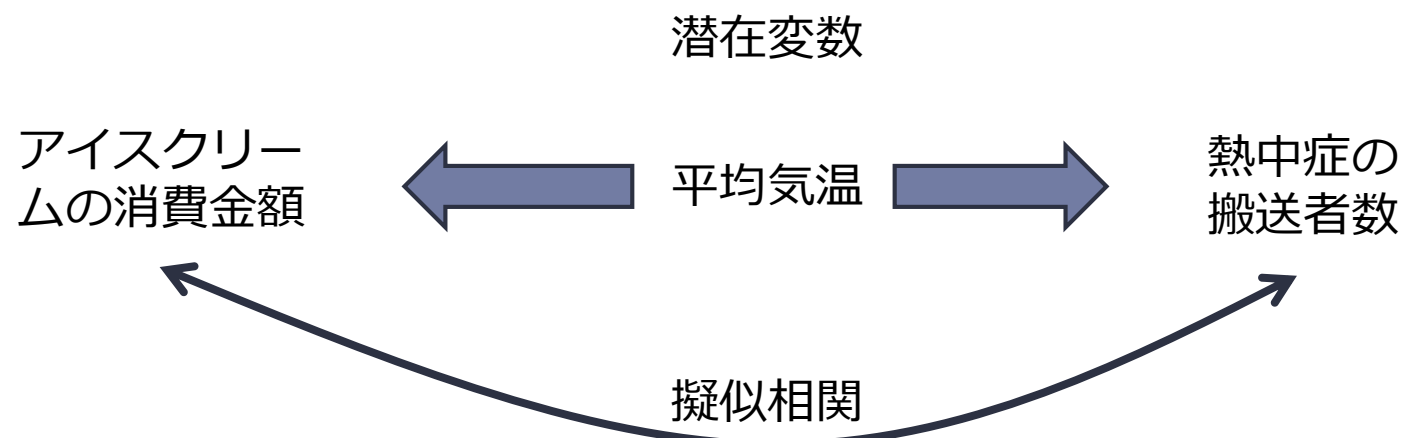
<https://www.fdma.go.jp/disaster/heatstroke/post4.html>

相関関係と因果関係

- ▶ 因果関係があれば相関関係がある

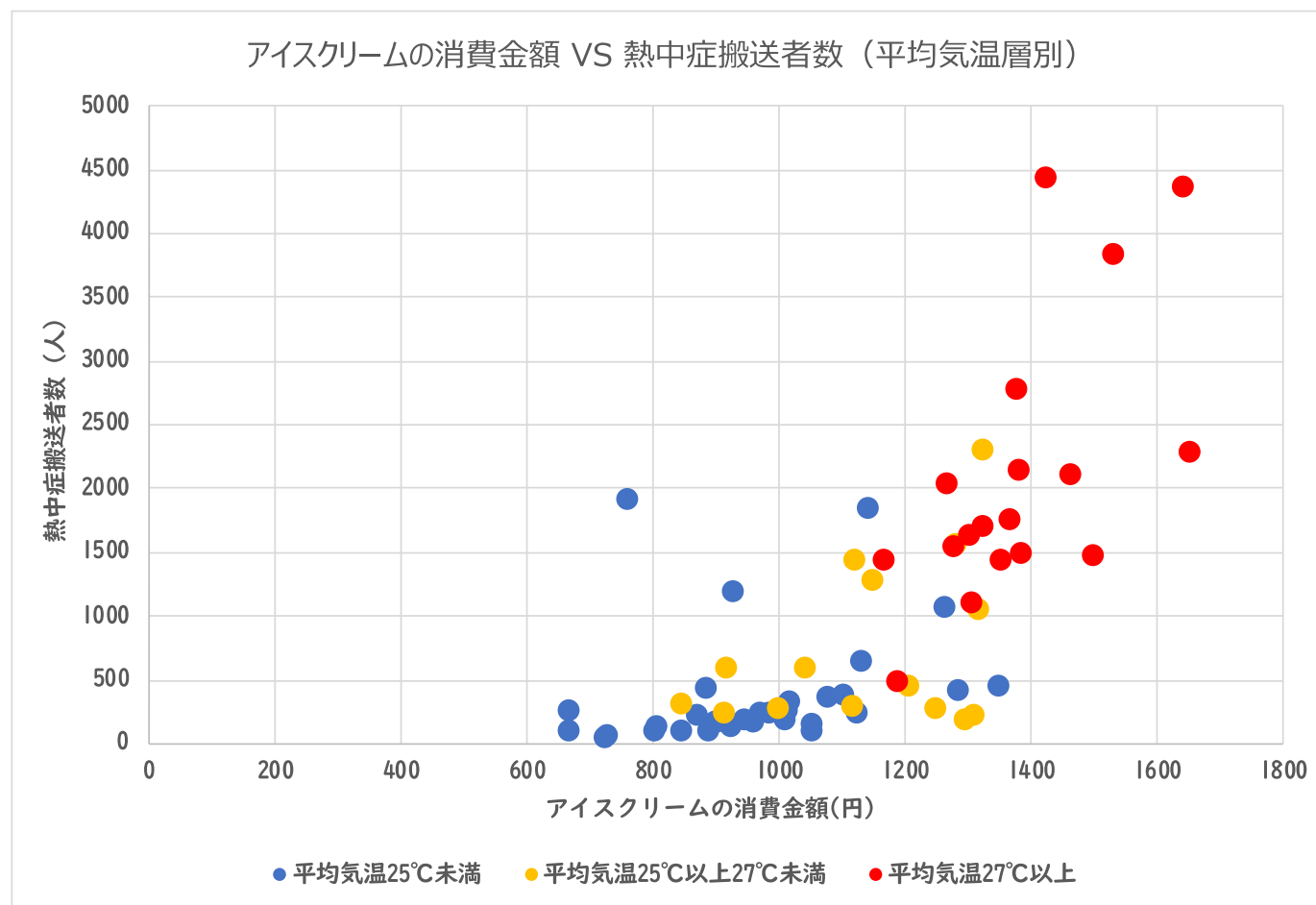


- ▶ 相関関係があっても因果関係があるとは限らない



相関関係と因果関係

▶ 平均気温で層化

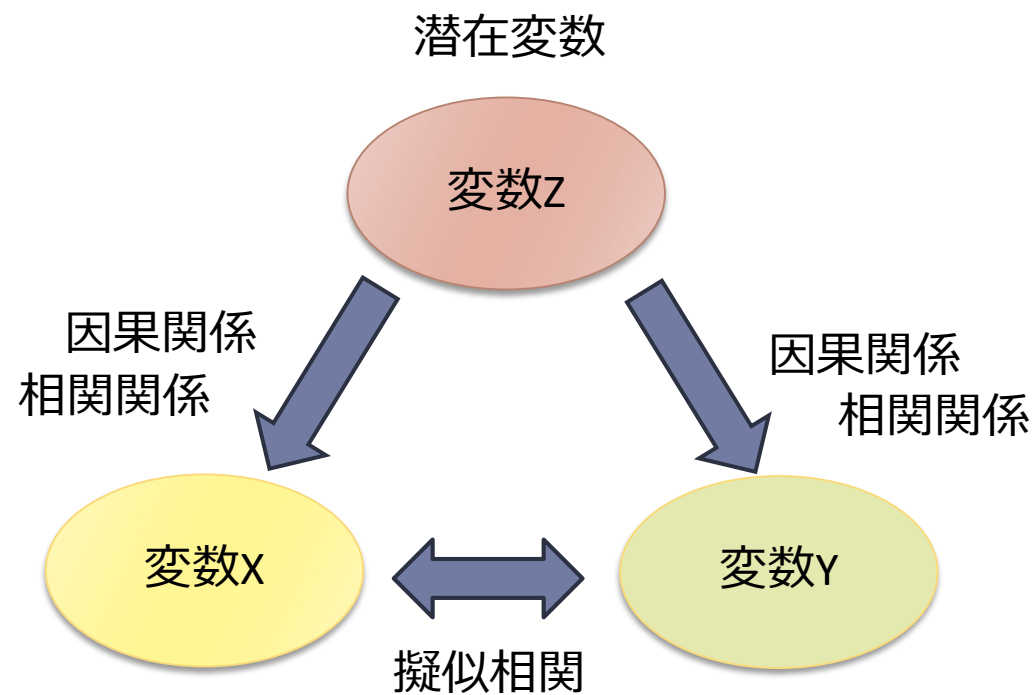
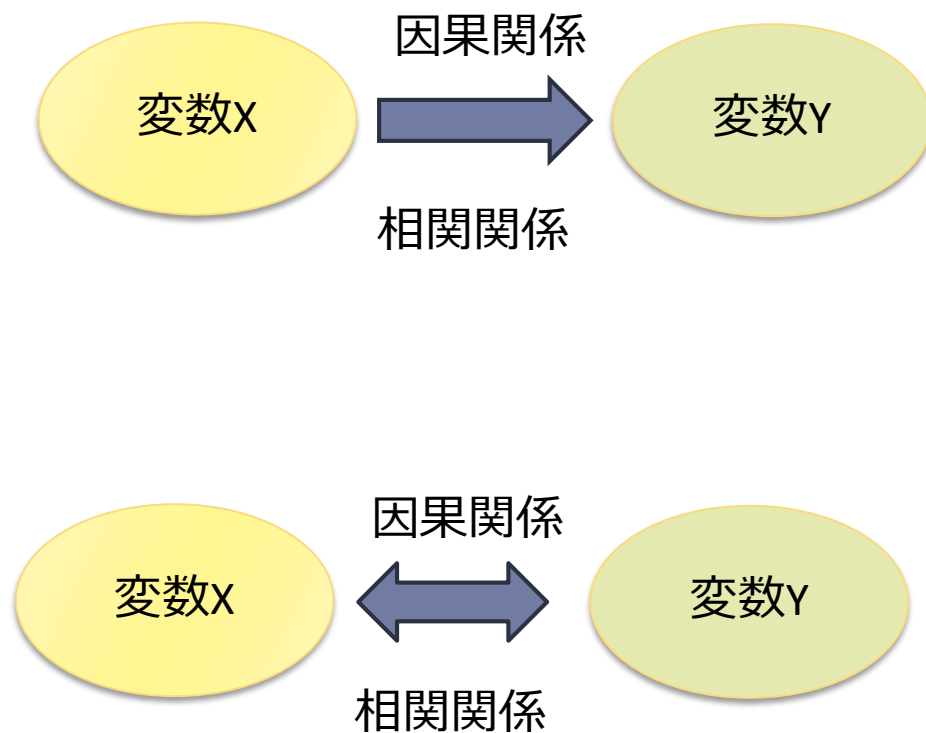


層	相関係数
25℃未満	0.23
25℃以上27℃未満	0.39
27℃以上	0.64
全体	0.70

【因果関係の特徴】

- ・ 時間順序
- ・ 直接制
- ・ 相関関係あり

因果関係



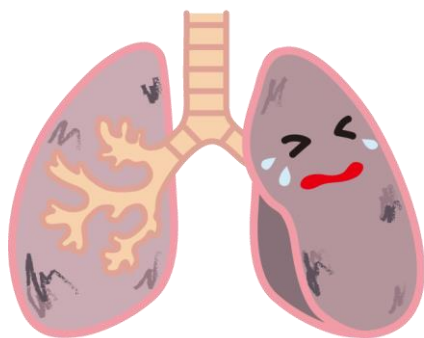
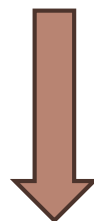
観察研究と実験研究

▶ 観察研究

タバコを吸うグループ

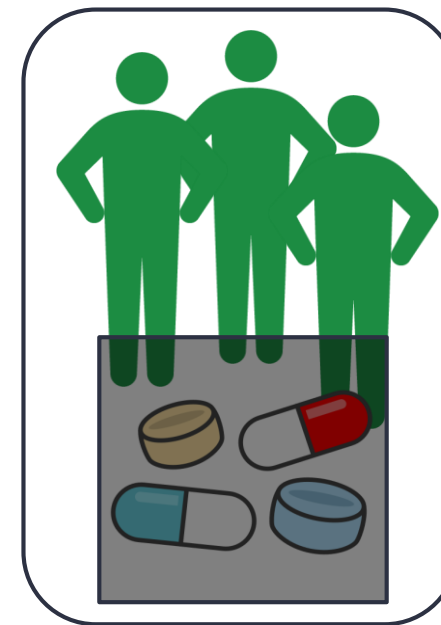


タバコを吸わないグループ



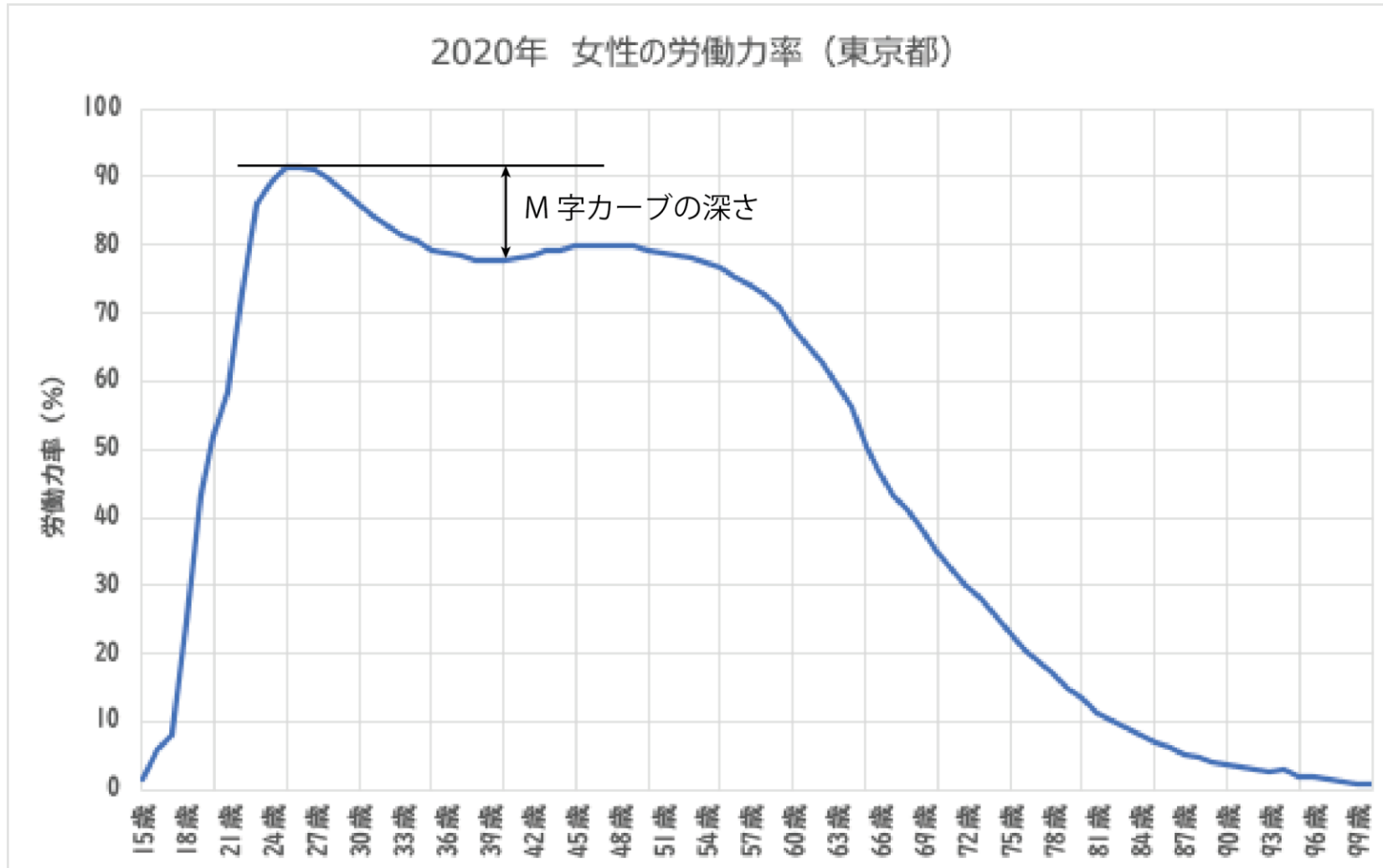
肺がん発生率

▶ 実験研究

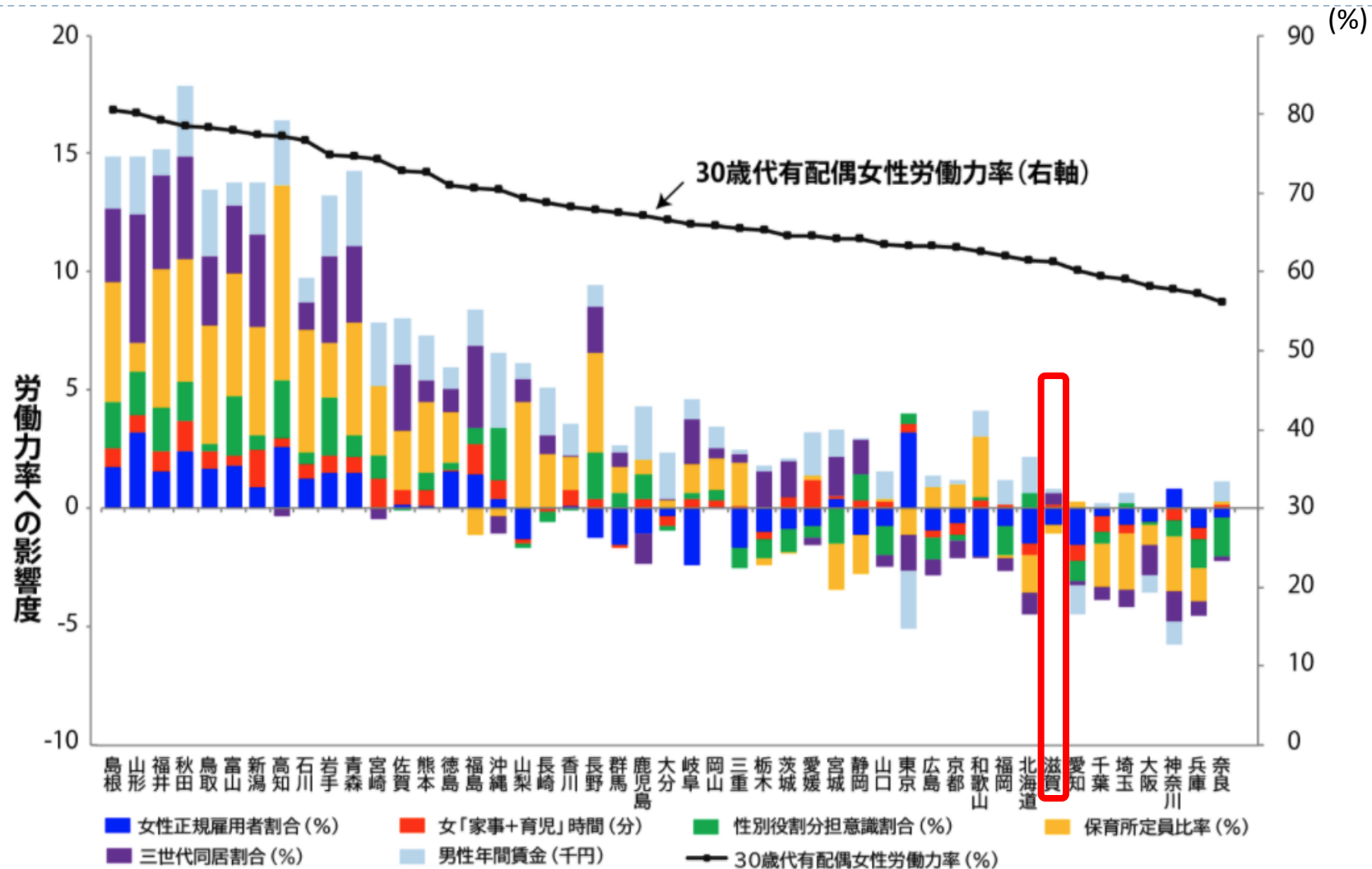


プラセボ
(偽薬)

M字カーブ



事例 Analysis (データサイエンス入門Aのスライド)



都道府県の各変数の影響度(全国平均を0としている)

<https://www.stat.go.jp/dstart/point/seminar2/c.html>

M字カーブ

人口・世帯



主な調査

- 国勢調査
- 人口推計
- 人口動態調査
- 生命表
- 国民生活基礎調査

➤ すべて見る (21 調査)

国勢調査	データベース	件数 更新日	ファイル	件数 更新日	概要
時系列データ		62件 2022-03-31		226件 2022-03-31	
都道府県・市区町村別の主な結果				5件 2022-07-22	
令和2年国勢調査		544件 2022-12-27		8,474件 2023-03-10	

令和2年国勢調査

公開（更新）日

■人口等基本集計（主な内容：男女・年齢・配偶関係、世帯の構成、住居の状態、母子・父子世帯、国籍など）[225件]

2022-12-27

■就業状態等基本集計（主な内容：労働力状態、就業者の産業・職業、教育など）[182件]

2022-12-27

■抽出詳細集計（主な内容：就業者の産業・職業（小・中分類）など）[47件]

2022-12-27

1-1-2	男女、年齢（各歳）、国籍総数か日本人別労働力率（15歳以上）－全国、都道府県、21大都市、特別区、人口50万以上の市	2020年10月	2022-05-27	➔ DB	🔗 API
-------	--	----------	------------	----------------------	-----------------------

M字カーブ

項番	事項名	説明	選択/全項目 ?	
1/6	表章事項		1/1	項目を選択
2/6	国籍総数か日本人	国籍総数	1/2	項目を選択
3/6	男女	女性	1/3	項目を選択
4/6	年齢	15-100歳	87/92	項目を選択
5/6	全国～人口50万以上の市	都道府県	47/99	項目を選択
6/6	時間軸（年次）		1/1	項目を選択

M字カーブ

ページ上部 (欄外)

▼

時間軸 (年次)

^

▼

▼

国籍総数が日本人

^

▼

▼

男女

^

▼

列

▼

表章事項

^

▼

▼

全国～人口50万以上の市

^

▼

都道府県

行

▼

年齢

^

▼

年齢

ページ上部

列

行

キャンセル

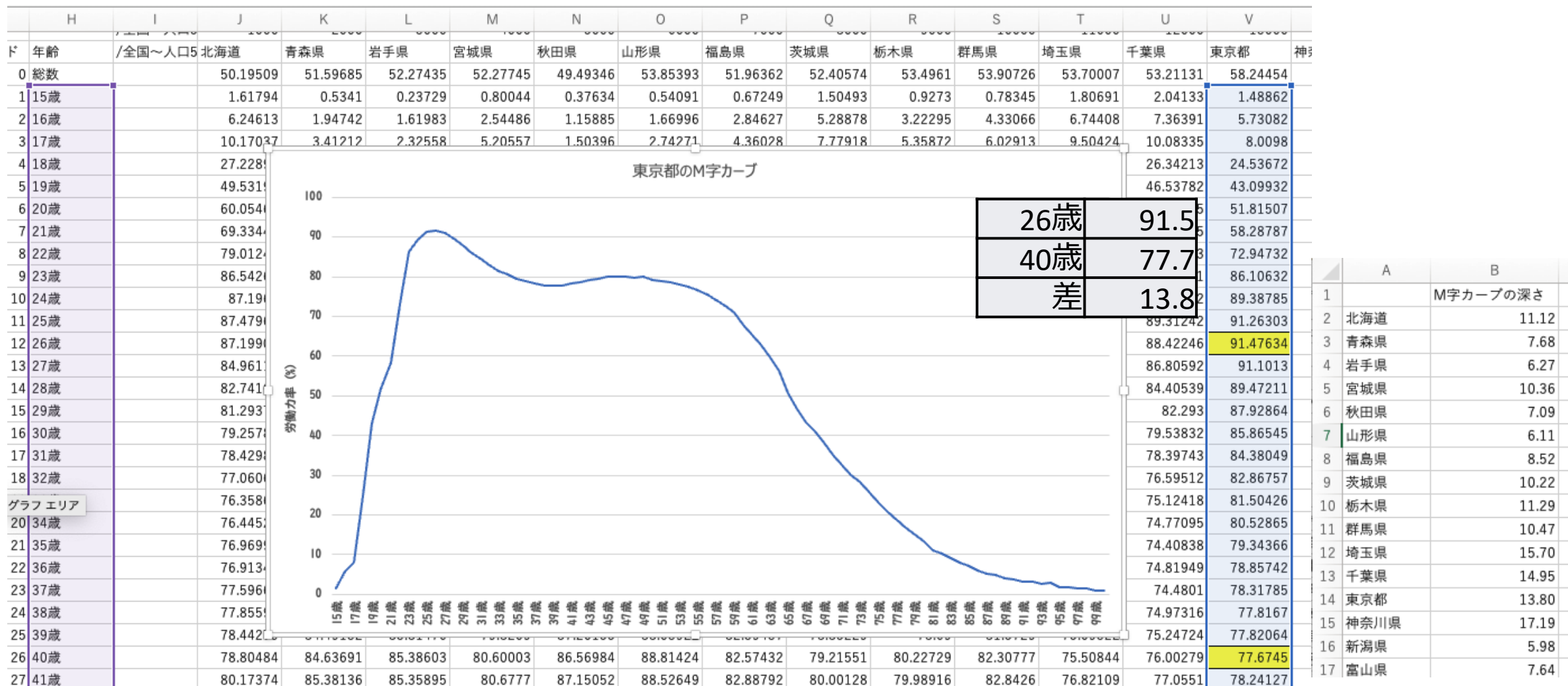
設定して表示を更新

画面表示セル数：4,089

画面総セル数：4,089

(統計表は最大100,000セルまで表示可能、
グラフは最大15,000セルまで表示可能)

M字カーブ



3 世帯の割合

人口・世帯



主な調査

- 国勢調査
- 人口推計
- 人口動態調査
- 生命表
- 国民生活基礎調査

➤ すべて見る (21 調査)

国勢調査	データベース	件数 更新日	ファイル	件数 更新日	概要
時系列データ		62件 2022-03-31		226件 2022-03-31	
都道府県・市区町村別の主な結果				5件 2022-07-22	
令和2年国勢調査		544件 2022-12-27		8,474件 2023-03-10	
平成27年国勢調査		381件 2020-05-22		11,414件 2021-11-30	
平成22年国勢調査		432件		12,615件	

都道府県・市区町村別の主な結果	公開（更新）日
■ 都道府県・市区町村別の主な結果 [5件]	2022-07-22

表番号	統計表	調査年月	公開（更新）日	
令和2年				
	都道府県・市区町村別の主な結果	2020年10月	2022-07-22	EXCEL

3世帯同居の割合

major_results_2020.xlsx

「a」: 全国、都道府県 「0」: 区 「1」: 政令市(特別区部を含む) 「2」: 市 「3」: 町、村	
都道府県・市区町村名 (英語)	市などの別 (地域識別 コード)
Tokyo-to	a
Kanagawa-ken	a
Niigata-ken	a
Toyama-ken	a
Ishikawa-ken	a

	A	B	C	D	E
1	都道府県名	都道府県・市 区町村名	一般世帯	(再掲) 3世代世帯	3世帯の割合
2	00_全国	00000_全国	55,704,949	2,030,030	4.20%
3	01_北海道	01000_北海道	2,469,063	59,601	2.41%
4	02_青森県	02000_青森県	509,649	42,895	8.42%
5	03_岩手県	03000_岩手県	490,828	46,934	9.56%
6	04_宮城県	04000_宮城県	980,549	67,505	6.88%
7	05_秋田県	05000_秋田県	383,531	39,453	10.29%
8	06_山形県	06000_山形県	396,792	55,238	13.92%
9	07_福島県	07000_福島県	740,089	69,042	9.33%
10	08_茨城県	08000_茨城県	1,181,598	80,078	6.78%
11	09_栃木県	09000_栃木県	795,449	56,778	7.14%
12	10_群馬県	10000_群馬県	803,215	43,722	5.44%
13	11_埼玉県	11000_埼玉県	3,157,627	105,717	3.35%
14	12_千葉県	12000_千葉県	2,767,661	95,699	3.46%
15	13_東京都	13000_東京都	7,216,650	97,110	1.35%
16	14_神奈川県	14000_神奈川県	4,210,122	93,407	2.22%
17	15_新潟県	15000_新潟県	862,796	90,547	10.49%

<https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200521&tstat=000001049104&cycle=0&tclass1=000001049105&tclass2val=0>

保育園

●統計データを探す (政府統計の調査結果を探します) その他の絞り込み

 **すべて**
政府統計一覧の中から探します

 **分野**
17の統計分野から探します

 **組織**
統計を作成した府省等から探します

キーワード検索: 検索

●統計データを活用する

 **グラフ**
主要指標をグラフで表示
(統計ダッシュボード)

 **時系列表**
主要指標を時系列表で表示
(統計ダッシュボード)

 **地図**
地図上に統計データを表示

 **地域**
都道府県、市区町村の
主要データを表示

☒ 都道府県データ

☐ 市区町村データ

 **データ表示**

 **地域ランキング**

1 絞り込み

地域区分 ☒ 都道府県 ☐ 地方 類型 III 地方の選び方

2 地域候補

地域候補 ☐ 地域

選択中地域 48地域 (残り 2452 地域選択可能) クリア

地域を選択 >

全て選択 >

< 地域を削除

< 全て削除

キャンセル 確定

00000_全国
01000_北海道
02000_青森県
03000_岩手県
04000_宮城県
05000_秋田県
06000_山形県
07000_福島県
08000_茨城県
09000_栃木県
10000_群馬県

1 絞り込み

データ ☒ 基礎データ ☐ 指標データ

分野

大分類

小分類

キーワード検索 実行

2 項目候補

項目候補 568項目

項目を選択 >

全て選択 >

< 項目を削除

< 全て削除

A1405	0~5歳人口 (人)
J2505	保育所等定員数 (詳細票) (人)
J250502	保育所等利用待機児童数 (人)
J2506	保育所等在所見数 (詳細票) (人)

保育園

調査年 2020年度 ▾ 並べ替え ☒ なし ☐ あり 降順 ▾ 項目 A7101

	A710101 一般世帯数 【世帯】	J250502 保育所等利用待機児童数 【人】	J2506 保育所等在所児数（詳細票） 【人】
00000_全国	55,704,949	12,439	2,624,335
01000_北海道	2,469,063	134	79,404
02000_青森県	509,649	0	30,723
03000_岩手県	490,828	58	28,895
04000_宮城県	980,549	340	39,497
05000_秋田県	383,531	22	22,621
06000_山形県	396,792	0	24,954
07000_福島県	740,089	141	31,614
08000_茨城県	1,181,598	193	57,353
09000_栃木県	795,449	34	40,394
10000_群馬県	803,215	14	46,241
11000_埼玉県	3,157,627	1,083	112,929
12000_千葉県	2,767,661	833	100,009

2 表示年度

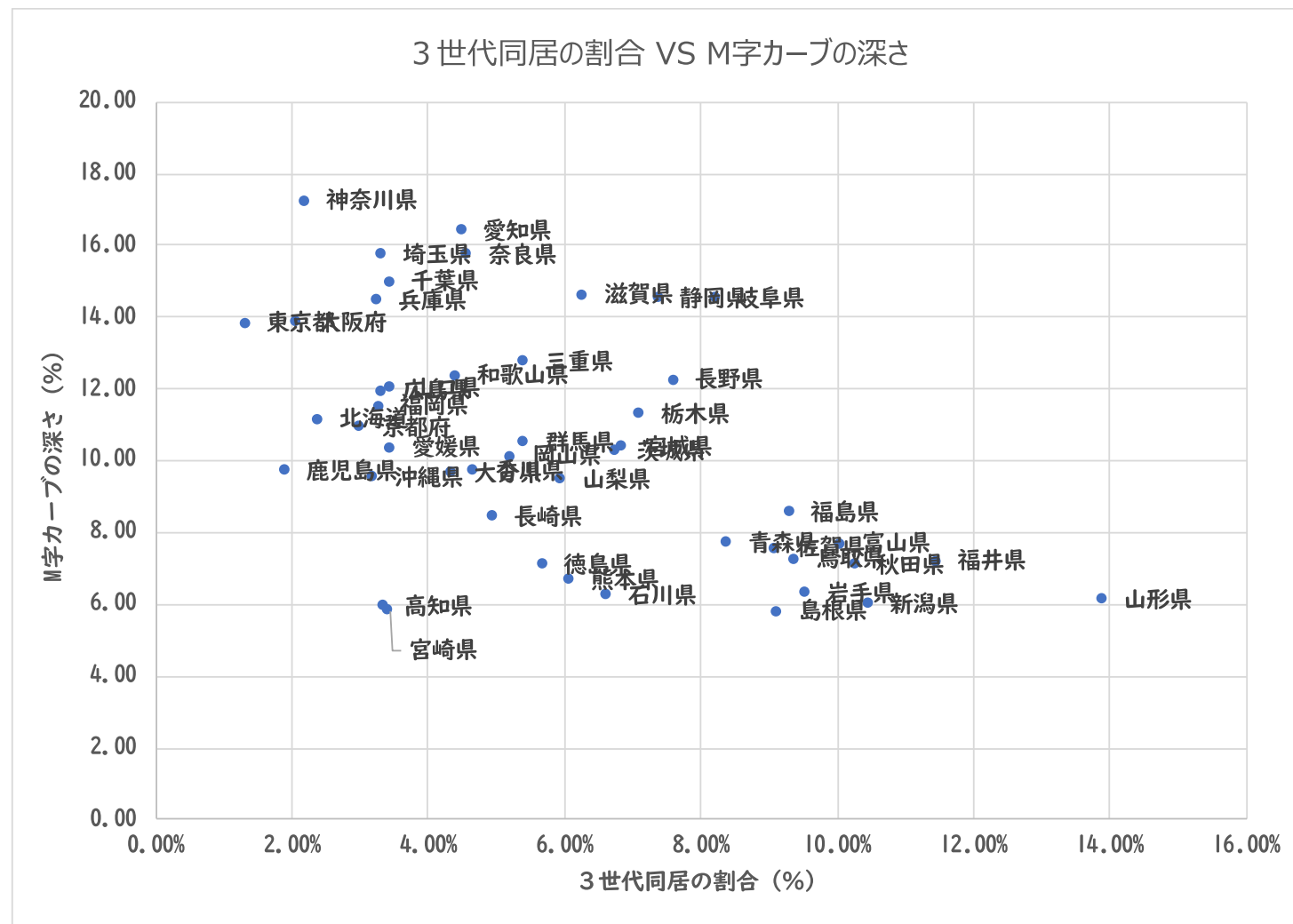
西暦年 2020 年から 2020 年まで

※都道府県データの提供年度は1975年以降となります。

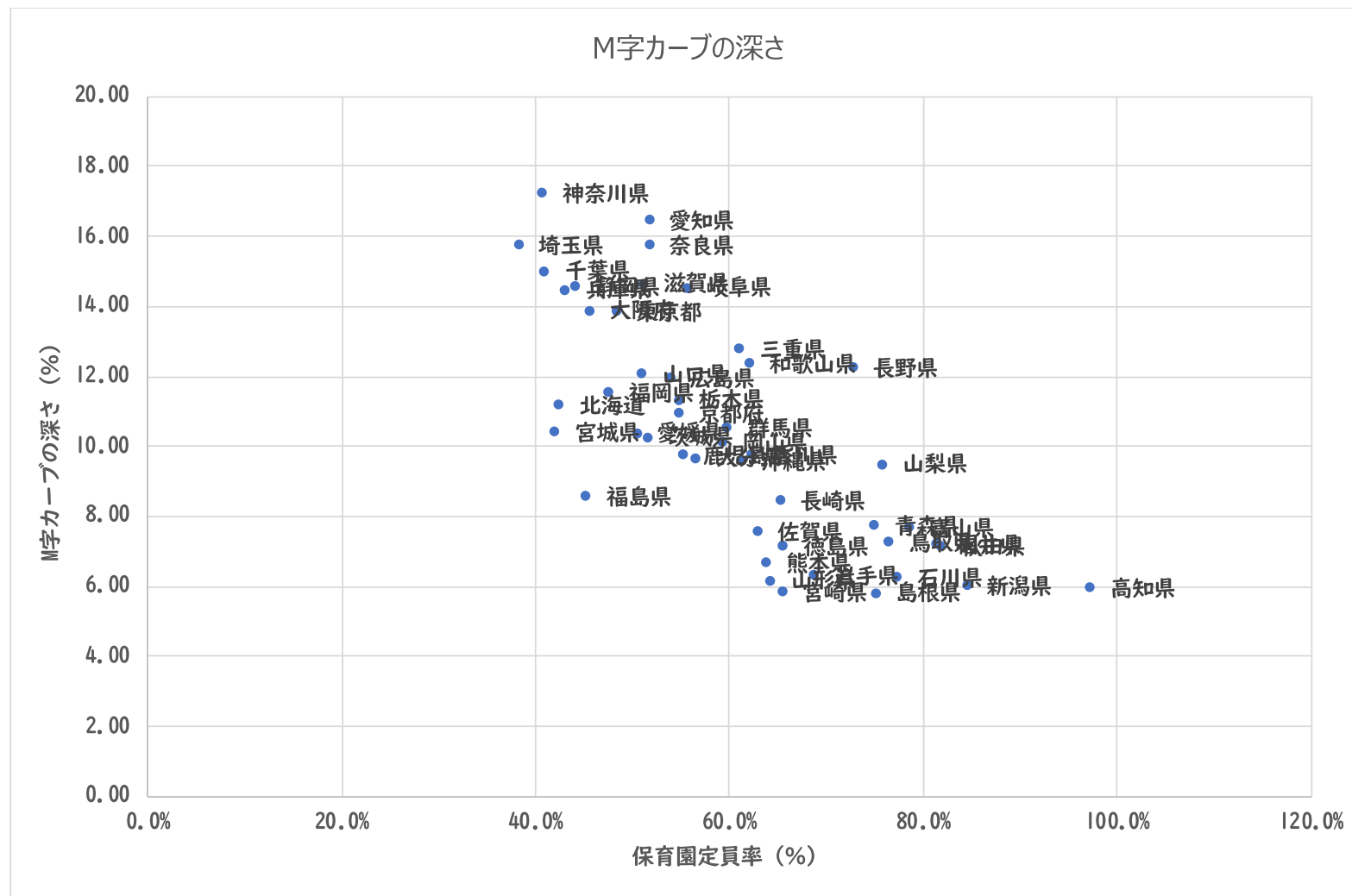
定員/幼児人口

	A1405	J2505	J250502	J2506	定員/幼児人口
地域	A1405_0~5歳	保育所等定員	保育所等利用	保育所等在所	
全国	5,514,746	2,858,117	12,439	2,624,335	
北海道	198,869	84,821	134	79,404	42.7%
青森県	45,552	34,236	0	30,723	75.2%
岩手県	46,031	31,709	58	28,895	68.9%
宮城県	97,187	40,995	340	39,497	42.2%
秋田県	31,138	25,515	22	22,621	81.9%
山形県	42,084	27,124	0	24,954	64.5%
福島県	74,375	33,797	141	31,614	45.4%
茨城県	118,241	61,249	193	57,353	51.8%
栃木県	81,362	44,731	34	40,394	55.0%
群馬県	79,015	47,408	14	46,241	60.0%
埼玉県	316,674	121,700	1,083	112,929	38.4%
千葉県	270,201	111,235	833	100,009	41.2%
東京都	630,803	306,055	2,343	287,079	48.5%
神奈川県	404,239	165,323	496	159,800	40.9%
新潟県	87,193	73,818	3	62,210	84.7%
富山県	41,303	32,511	0	29,101	78.7%
石川県	50,089	38,830	0	34,108	77.5%
福井県	34,316	27,969	0	25,632	81.5%
山梨県	32,961	25,039	0	19,107	76.0%
長野県	85,849	62,598	46	52,058	72.9%
岐阜県	84,109	46,892	0	37,467	55.8%
静岡県	155,452	68,800	122	62,833	44.3%
愛知県	364,976	189,374	155	157,939	51.9%
三重県	74,994	45,963	81	39,969	61.3%
滋賀県	69,597	35,753	495	34,237	51.4%
京都府	1,070,020	587,776	1,083	514,202	51.8%

M字カーブの深さ vs 3世帯同居の割合



M字カーブの深さ vs 保育園の定員割合



相関係数

3世代同居率 VS M字カーブの深さ	0.53	
保育園定員率 VS M字カーブの深さ	0.73	

まとめ

項目	内容
散布図	2変数の関係を可視化する
共分散	$v_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$
相関係数	$r = \frac{v_{xy}}{\sigma_x \sigma_y} \quad -1 \leq r \leq 1$
相関関係と因果関係	<div><p>The diagram illustrates the relationship between two variables, 変数X (Variable X) and 変数Y (Variable Y). It shows two types of relationships: a causal relationship (因果関係) indicated by a solid blue arrow pointing from X to Y, and a correlation relationship (相関関係) indicated by a dashed blue arrow pointing from X to Y.</p></div> <div><p>The diagram illustrates a latent variable (潜在変数) influencing two observed variables, 変数X (Variable X) and 変数Y (Variable Y). The latent variable is shown in a red oval at the top, with arrows pointing down to the observed variables. The arrows are labeled with '因果関係' (Causal relationship) and '相関関係' (Correlation relationship). The observed variables are shown in yellow and green ovals, respectively, with a double-headed arrow between them labeled '擬似相関' (Spurious correlation).</p></div>

問題

企業・家計・経済



主な調査

- 国民経済計算
- 個人企業経済調査
- 経済センサス-基礎調査
- 経済センサス-活動調査
- 経済構造実態調査
- **家計調査**
- 全国家計構造調査（旧全国消費実態調査）
- 小売物価統計調査
- 全国物価統計調査
- 産業連関表
- 法人企業統計調査
- 経済産業省企業活動基本調査

➤ **すべて見る (87 調査)**

家計調査	データベース	件数 更新日	ファイル	件数 更新日	概要
家計調査		112件 2023-09-05		23,545件 2023-09-05	

家計調査	公開（更新）日
------	---------

■家計収支編 [85件]

— 二人以上の世帯 [55件]

— 月次 [16件]

2023-09-05

— 四半期 [9件]

2023-08-08

— 年次 [30件]

2023-02-07

品目分類					
010	品目分類（2020年改定）（総数：金額）	-	2023-09-05		
011	品目分類（2020年改定）（総数：数量）	-	2023-09-05		

家計データ

項番	事項名	説明	選択/全項目 ?	
1/5	表章項目		1/1	項目を選択
2/5	品目分類（2020年改定）	問題文に指定された項目	3/690	項目を選択
3/5	世帯区分	2人以上世帯	1/4	項目を選択
4/5	地域区分	東京都区部	1/53	項目を選択
5/5	時間軸（月次）	全部	463/463	項目を選択

初期状態に戻す（全項目表示）

キャンセル

確定

参考文献

- ▶ 「身近な統計」 石崎克也・渡辺美智子 放送大学教育振興会
- ▶ 「大学生のためのデータサイエンス1」 オフィシャル スタディノート
ト 総務省統計局