

1 学習内容の説明 ⇒ 2 問題演習 ⇒ 3 振り返り（確認テスト・相互採点・リフレクションの記入）

【内容目標】 散らばりの度合いを表す値を求められるようになる

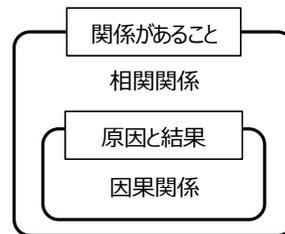
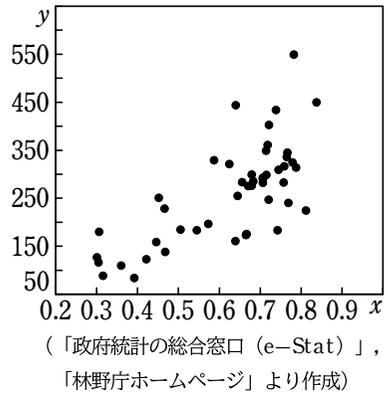
□ 相関関係と因果関係

47 都道府県において、2017 年度の森林面積の総面積に対する割合 x と、人口 100 万人あたりの郵便局の数 y のデータを調べ、散布図にしたところ、右の図のようになった。相関係数は 0.72 である。

x と y の間には正の相関関係が認められる。しかし、森林面積の割合が多いことが原因で郵便局が増えるとか、逆に、郵便局が多いことが原因で森林面積の割合が大きくなるということは断定できないであろう。つまり、一方が原因で他方が起こる

因果関係 があるとは断定できない。

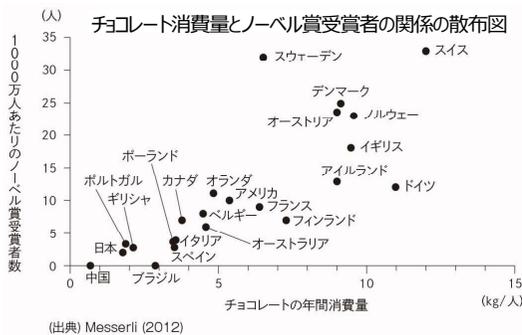
一般に、2 つの変量の間に関係があるからといって、必ずしも因果関係があるとはいえない。



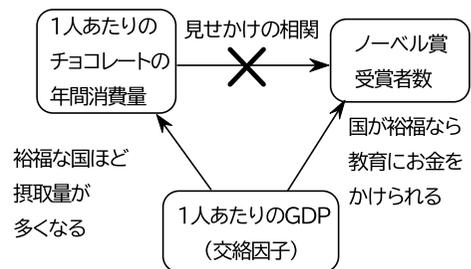
疑似相関とは？

因果関係がないのに、見えない要因によってあたかも因果関係があるように見える現象のこと。統計学でよく使われる言葉で、「見せかけの相関」「見かけ上の相関」とも言う。

一般的に有名な疑似相関の例としては、「チョコレートの消費量とノーベル賞の受賞者数」、「アイスクリームの売上と水難事故の数」、「年賀状を出す枚数と収入の高さ」等がある。



チョコレート消費量を増やしてもノーベル賞受賞者は増えない



チョコレートは生きていくのになくてもよいいわゆる贅沢品であるので、裕福な国ほど摂取量が多くなるのは当然である。また国が裕福になれば、教育にもお金をかけられるようになるので、ノーベル賞受賞者を輩出できる可能性は上がると考えられる。

このような「第3の変数」のことを「**交絡因子 (こらうくいんし)**」と呼ぶ。この交絡因子があると、相関関係にすぎないものがまるで因果関係のように見えてしまう。

□質的データをとる2つの変量の間の関係

これまで、身長や体重などのように数値として得られるデータを主に扱ってきた。このような数値として得られるデータを **量的データ** という。

これに対し、「A組」、「B組」などの所属クラスや、「神奈川県」や「京都府」などの都道府県のデータのように、数値ではないものとして得られるデータを **質的データ** という。

ここでは、質的データをとる2つの変量の間の関係を調べることについて考えよう。

例12)

合否が判定されるある試験において、受験者100人全員を対象に、教材Aを使用して学習したか調べたところ、その人数は表1のようになった。例12の表1のような表を **分割表** という。クロス集計表ともいう。

表1 **分割表**

	合	否	計
Aの使用：有	9	5	14
Aの使用：無	42	44	86
計	51	49	100

$9 \div 14 = 0.6428\dots$ $5 \div 14 = 0.3571\dots$
 $42 \div 86 = 0.4883\dots$ $44 \div 86 = 0.5116\dots$

表1において、教材Aを使用した者、使用していない者のそれぞれにおいて、合格者、不合格者が占める割合を計算すると、表2のようになる。

表2

	合	否
A：有	64%	36%
A：無	49%	51%

表2だけを見ると、教材Aの使用が合否に影響を及ぼしているように見える。ここでさらに、教材Bを使用して学習したかも調べてみたところ、その人数は表3のようになった。

表3

		合	否
A：有	B：有	6	1
A：有	B：無	3	4
A：無	B：有	31	9
A：無	B：無	11	35

練習15) 表3をもとに、次の問いに答えよ。
 (1) 表4の空らん適切な数を入れよ。

表4

	合	否	計
B：有	37	10	47
B：無	14	39	53
計	51	49	100

$37 \div 47 = 0.7872\dots$ $10 \div 47 = 0.2127\dots$
 $14 \div 53 = 0.2641\dots$ $39 \div 53 = 0.7358\dots$

(2) 教材Bを使用した者、使用していない者のそれぞれにおいて、合格者、不合格者の占める割合を計算して表2のようにまとめよ。

	合	否
B：有	79%	21%
B：無	26%	74%

深める 上で得られたデータから、教材A、Bのどちらの方が、この試験の合否により影響を及ぼしていると予想できるだろうか。

	合	否		合	否		合	否
A：有	64%	36%	B：有	79%	21%	A：有 B：有	6	1
A：無	49%	51%	B：無	26%	74%	A：有 B：無	3	4
						A：無 B：有	31	9
						A：無 B：無	11	35

1 学習内容の説明 ⇒ 2 問題演習 ⇒ 3 振り返り（確認テスト・相互採点・リフレクションの記入）

【内容目標】 散らばりの度合いを表す値を求められるようになる

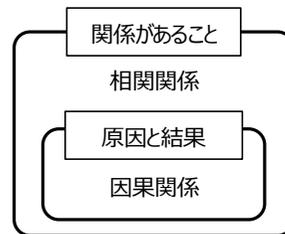
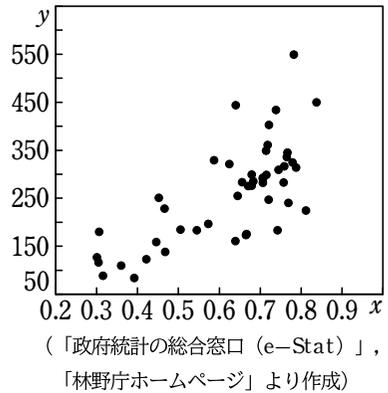
□ 相関関係と因果関係

47 都道府県において、2017 年度の森林面積の総面積に対する割合 x と、人口 100 万人あたりの郵便局の数 y のデータを調べ、散布図にしたところ、右の図のようになった。相関係数は 0.72 である。

x と y の間には正の相関関係が認められる。しかし、森林面積の割合が多いことが原因で郵便局が増えるとか、逆に、郵便局が多いことが原因で森林面積の割合が大きくなるということは断定できないであろう。つまり、一方が原因で他方が起こる

因果関係 があるとは断定できない。

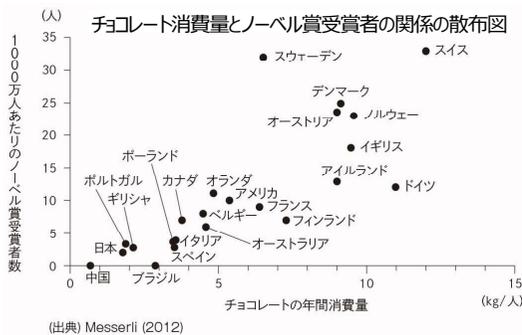
一般に、2つの変量の間に関連関係があるからといって、必ずしも因果関係があるとはいえない。



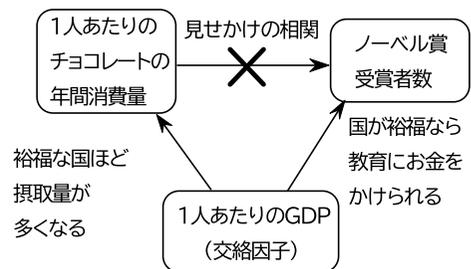
疑似相関とは？

因果関係がないのに、見えない要因によってあたかも因果関係があるように見える現象のこと。統計学でよく使われる言葉で、「見せかけの相関」「見かけ上の相関」とも言う。

一般的に有名な疑似相関の例としては、「チョコレートの消費量とノーベル賞の受賞者数」、「アイスクリームの売上と水難事故の数」、「年賀状を出す枚数と収入の高さ」等がある。



チョコレート消費量を増やしてもノーベル賞受賞者は増えない



チョコレートは生きていくのになくてもよいいわゆる贅沢品であるので、裕福な国ほど摂取量が多くなるのは当然である。また国が裕福になれば、教育にもお金をかけられるようになるので、ノーベル賞受賞者を輩出できる可能性は上がると考えられる。

このような「第3の変数」のことを「**交絡因子 (こらうくいんし)**」と呼ぶ。この交絡因子があると、相関関係にすぎないものがまるで因果関係のように見えてしまう。

□質的データをとる2つの変量の間の関係

これまで、身長や体重などのように数値として得られるデータを主に扱ってきた。このような数値として得られるデータを **量的データ** という。

これに対し、「A組」、「B組」などの所属クラスや、「神奈川県」や「京都府」などの都道府県のデータのように、数値ではないものとして得られるデータを **質的データ** という。

ここでは、質的データをとる2つの変量の間の関係を調べることについて考えよう。

例12)

合否が判定されるある試験において、受験者100人全員を対象に、教材Aを使用して学習したか調べたところ、その人数は表1のようになった。例12の表1のような表を **分割表** という。クロス集計表ともいう。

表1 **分割表**

	合	否	計
Aの使用：有	9	5	14
Aの使用：無	42	44	86
計	51	49	100

$$9 \div 14 = 0.6428\dots \quad 5 \div 14 = 0.3571\dots$$

$$42 \div 86 = 0.4883\dots \quad 44 \div 86 = 0.5116\dots$$

表1において、教材Aを使用した者、使用していない者のそれぞれにおいて、合格者、不合格者が占める割合を計算すると、表2のようになる。

表2

	合	否
A：有	64%	36%
A：無	49%	51%

表2だけを見ると、教材Aの使用が合否に影響を及ぼしているように見える。ここでさらに、教材Bを使用して学習したかも調べてみたところ、その人数は表3のようになった。

表3

		合	否
A：有	B：有	6	1
A：有	B：無	3	4
A：無	B：有	31	9
A：無	B：無	11	35

練習15) 表3をもとに、次の問いに答えよ。
(1) 表4の空らん適切な数を入れよ。

表4

	合	否	計
B：有	37	10	47
B：無	14	39	53
計	51	49	100

$$37 \div 47 = 0.7872\dots \quad 10 \div 47 = 0.2127\dots$$

$$14 \div 53 = 0.2641\dots \quad 39 \div 53 = 0.7358\dots$$

(2) 教材Bを使用した者、使用していない者のそれぞれにおいて、合格者、不合格者の占める割合を計算して表2のようにまとめよ。

	合	否
B：有	79%	21%
B：無	26%	74%

深める 上で得られたデータから、教材A、Bのどちらの方が、この試験の合否により影響を及ぼしていると

解答

予想できるだろうか。

	合	否		合	否		合	否	
A：有	64%	36%	B：有	79%	21%	A：有	B：有	6	1
A：無	49%	51%	B：無	26%	74%	A：有	B：無	3	4
						A：無	B：有	31	9
						A：無	B：無	11	35

使用している受験者の合格する割合が多いことから、Bの方がより影響を及ぼしていると予想できる。(Aは無くても受かっている受験者も多く、そもそも使用している人数が少なく判断しづらい)