

統計基礎 第11回 (2017.12.18)

テーマ：相関関係と因果関係

1 はじめに

- ・ 冬休みの宿題は、冬休み明けの最初の授業（1/15）に提出して下さい。
台風で休講になった回の授業の代わりです。
- ・ 第15回目（1/29）にテスト（総合演習）を行います。PCを使った問題を含んでいます。持ち込み可です。冒頭に、座席を指定しますのでご協力下さい。

演習1 100円あたりの期待値は、おおよそどの位だと考えますか？

（用語）期待値：その試行を行ったとき、その結果として得られる値の平均値のこと。ギャンブルの場合は、掛け金に対して戻ってくる見込みの金額のこと。

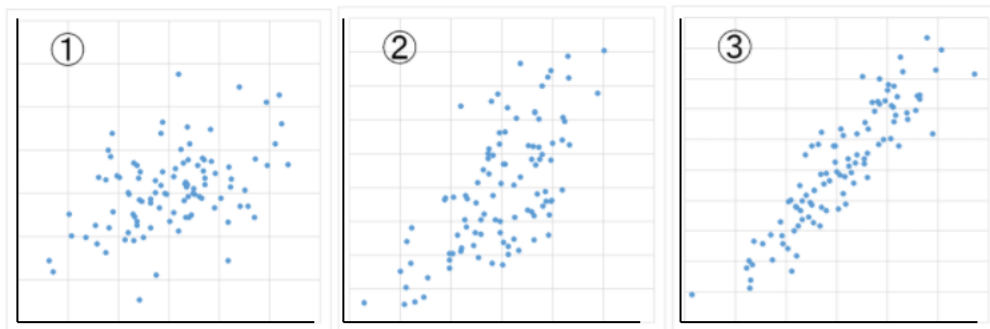
(1) 宝くじ

ア 100円、 イ 75円、 ウ 50円

(2) 競馬、競輪

ア 100円、 イ 75円、 ウ 50円

演習2 次の3つの散布図のうち、相関係数の値が一番大きいものはどれですか？



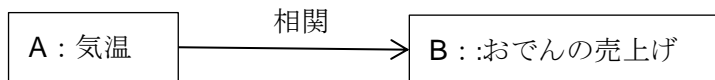
2 相関関係と因果関係

- ・ ポイント：相関関係があっても因果関係があるとは限らない

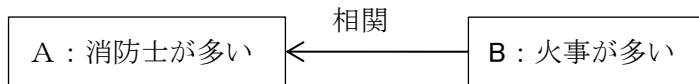
相関関係があるということは因果関係があるということですか？

- ・ 2つの変数（A, B）間に相関が見られる場合、偶然によるものを除けば、以下の3つの可能性があります。

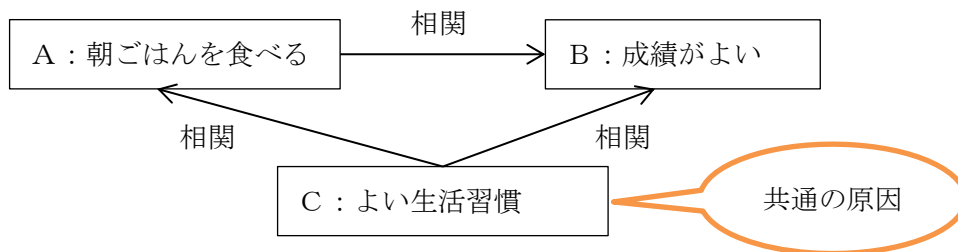
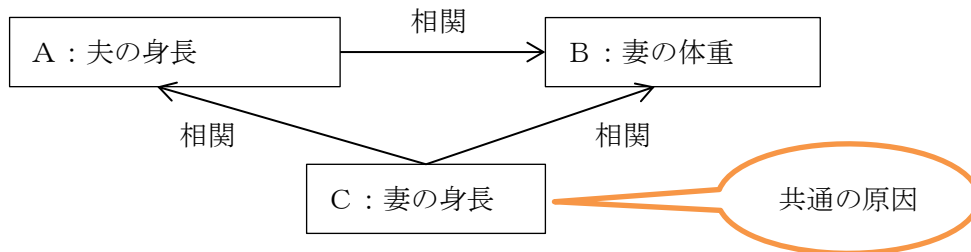
① Aが原因



② Bが原因



③ Cが原因



(重要) 相関関係があっても、因果関係があるとは限らない

AとB：相関関係あり
AとB：因果関係なし } ⇒ AとB：疑似相関である
(見かけの相関)

演習2 カゼ薬への支出額とセーターの支出額に、下図のとおりです。以下の問に答えて下さい。

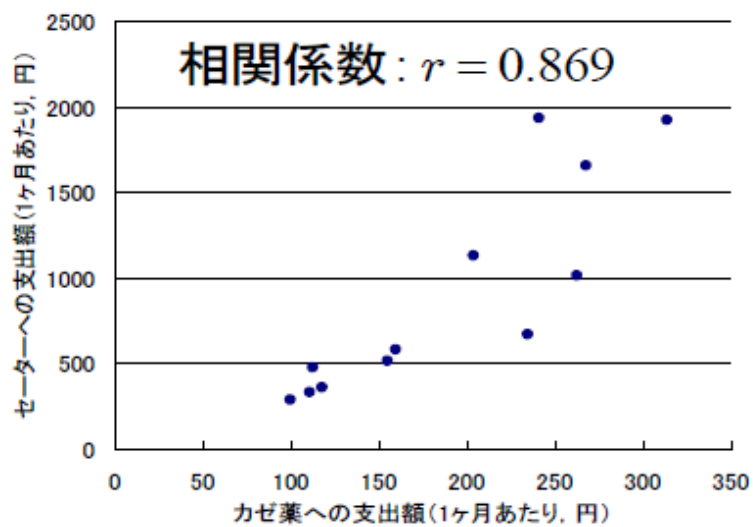
(1) 相関関係

ア あり イ なし

(2) 因果関係

ア あり イ なし

(3) (2)でイと答えた人へ。因果関係がないとすると、共通の原因は何と考えますか？



演習3 次の2つの量には、

- (1) 相関関係があると思いますか？
- (2) 因果関係があると思いますか？
- (3) (2) でなしの場合、共通の原因は何であると考えますか？

	事例	相関関係 ○：あり ×：なし	因果関係 ○：あり ×：なし	(因果関係がない場合) 共通の原因は？
(1)	アイスクリームの売上げと 熱中症の患者数			
(2)	小学生の足の大きさと 小学生の知っている漢字の数			
(3)	体力と年収			

因果関係の有無を調べる方法

- ・ **体力と年収**に相関関係があることがわかっています。
因果関係の有無を調べるにはどのような方法ありますか？

(方法1) 因果関係の理屈を考える

- ◇ **体力**がある → 仕事を頑張れる → **年収**が大きい ?
- ◇ **年収**が大きい人 → 運動習慣がある → **体力**がある ?

(方法2) 他の項目の値を一定にして相関関係を調べる

⇒ 同じ**年齢**の人について、両者の関係を調べる

(方法3) 複数の相関関係を調べる

⇒ **体力と年収** の相関関係
年齢と体力 の相関関係、
年齢と年収 の相関関係を調べる

演習 4

- (1) 以下の 30 人のデータから、**体力**と**年収**について、散布図を描いて下さい。
- (2) ①**体力**と**年収**、②**年齢**と**体力**、③**年齢**と**年収**、の相関係数をそれぞれ求めてください。
- (3) **体力**と**年収**の因果関係は大きいと考えますか、小さいと考えますか？
そのように考えた理由を書いて下さい。

体力と年収の関係			
サンプル 番号	体力(点)	年収 (万円)	年齢(歳)
1	81	391	18
2	81	385	24
3	89	396	17
4	87	437	19
5	72	443	20
6	75	386	27
7	85	449	30
8	81	480	32
9	72	482	27
10	73	455	28
11	73	412	34
12	76	523	24
13	71	516	43
14	71	499	40
15	70	477	40
16	62	497	39
17	70	435	34
18	68	552	39
19	70	627	55
20	58	659	48
21	60	626	50
22	62	599	47
23	56	625	46
24	56	602	44
25	61	574	65
26	66	579	49
27	48	610	66
28	57	589	50
29	60	618	55
30	56	567	63

Excel の基本操作

(1) 離れた2つのセル範囲を選択する方法（復習）

- ① 1つ目のセル範囲を選択
- ② Ctrl キーを押したまま2つ目のセル範囲を選択

(2) 散布図のX軸（横軸）とY軸（縦軸）を交換する方法

（注）散布図では、原因となるものをX軸とし、結果となるものをY軸とします。

- ・ グラフツール→「デザイン」タブ→「データの選択」
- ・ 「データソースの選択」のダイアログボックスで「編集」ボタンをクリック
 - 系列名： 系列Yの項目名のセルを選択
 - 系列Xの値の欄： セルの範囲を選択（項目名は含めない）
 - 系列Yの値の欄： セルの範囲を選択（項目名は含めない）

(3) X軸ラベル、Y軸ラベルに素早く入力する方法

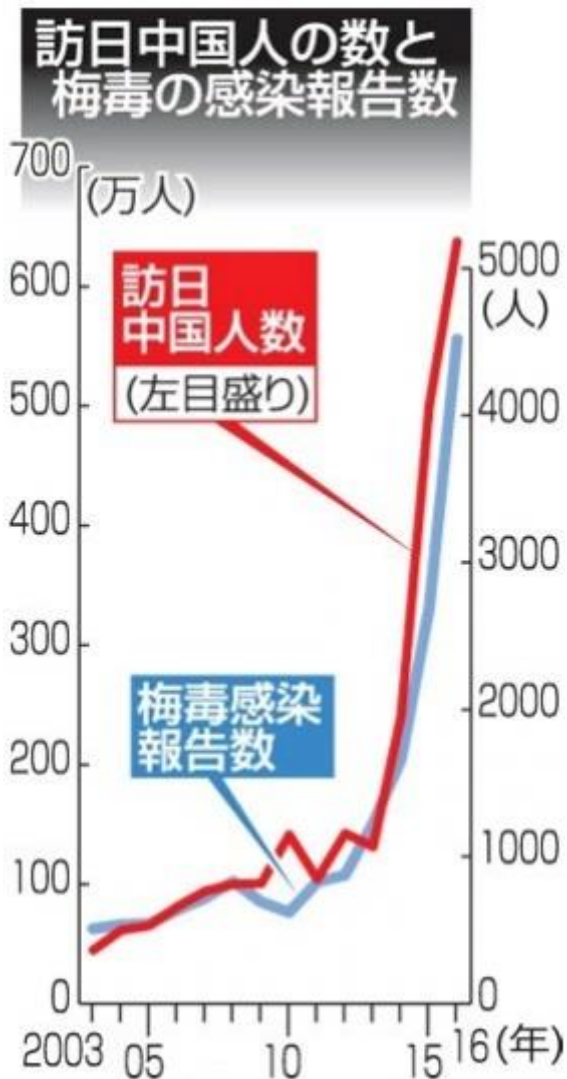
- ・ （グラフの余白をクリック）→グラフツール→デザイン→グラフのレイアウト
→（グラフのX軸ラベルに「軸ラベル」、グラフのY軸ラベルに「軸ラベル」という文字が入力されているレイアウトを選択）

(4) セル内で途中改行する方法

- ・ Alt キーを押したまま Enter キーを押す

報道 梅毒急増で根拠薄い投稿拡散 葛飾区議がツイート 「差別あおる」指摘も

(2017年12月08日、西日本新聞)

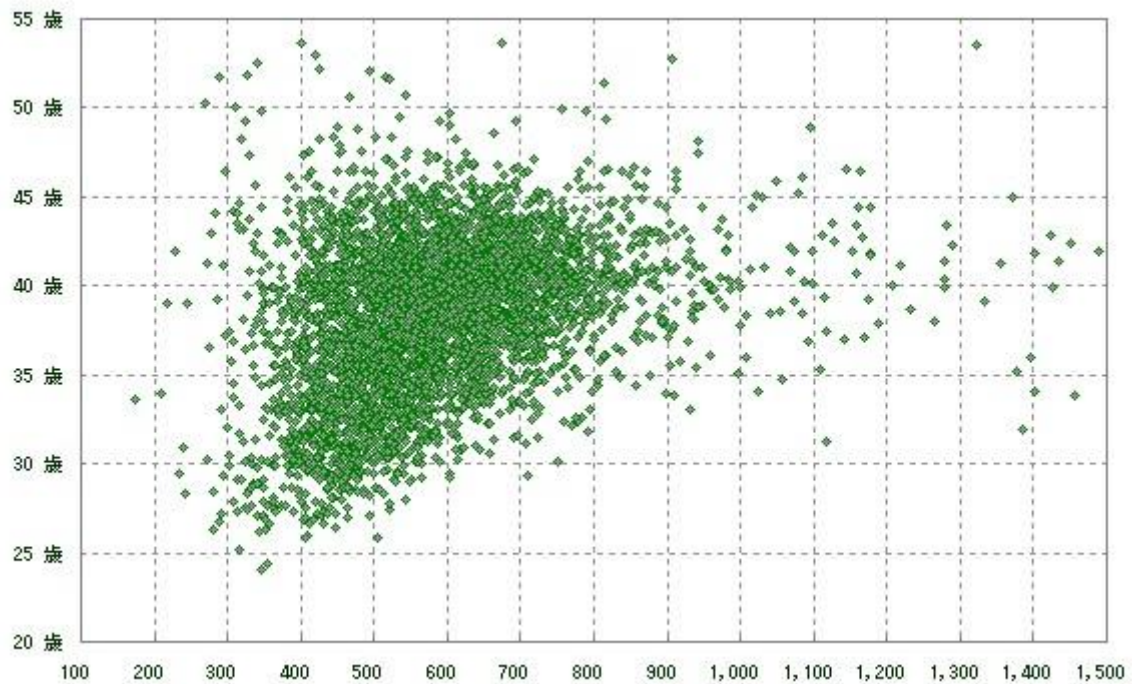


- ・ 国内で急増する梅毒感染について、訪日中国人観光客の増加が原因だと地方議員がツイッターで指摘し、インターネット上で拡散している。議員はネット情報などを根拠に挙げ、賛同するブログもあるが、専門機関によれば、因果関係を裏付けるデータはない。
- ・ 国立感染症研究所が、今年の梅毒感染者が5053人に達したと発表した11月28日。東京都葛飾区議の鈴木信行氏はツイッターで「誰が日本に持ち込んだか分かるじゃん。一番日本に来ている外国人の支那人だよ。」と発信した。感染者が5千人を超えるのは1973年以来、44年ぶりだった。

- ・ 鈴木氏はブログでも、訪日中国人が10年間で急増したことを示す統計を紹介し、感染報告数の伸びと一致すると指摘。訪日外国人旅行者の増加が梅毒急増の「外的要因」と類推する医師のブログの一部を引用するなどして持論を展開した。
- ・ 外国人差別につながりかねない内容だが、そもそも真偽はどうか。
- ・ 厚生労働省によると、医師が梅毒患者を診察した場合、感染経路や感染地域を届け出る必要がある。ところが、国籍に関する情報は届け出項目になく、担当者は「中国人観光客との因果関係は調査しないと分からない」と話す。
- ・ 国立感染症研究所も「国外から梅毒の細菌が流入するリスクはあるが、中国からは断じられない」と否定的。鈴木氏が引用したブログに関し、執筆者の医師が勤務する病院のスタッフは「可能性のある要因の一つとして挙げただけ」と困惑する。医師が診察で梅毒感染を見逃さなくなったり、検査を受ける人が増えたりしたとの見方もあり、感染急増の原因はよく分かっていないのが実情だ。
- ・ 訪日中国人数と梅毒感染数の推移を示すデータを比べても、疑問が残る。確かに近年、増加傾向の一致が目立つが、中国人数が増えた2008～10年は、梅毒報告数は逆に減っている。

参考 上場企業社員に関する年齢と年収の散布図

(野村総合研究所「生活者1万人アンケート」)



解答例

【演習2】

- (1) ア、 (2) イ、 (3) その月の気温

「カゼ薬への支出額」と「セーターの支出額」の散布図では、点が右上がりとなっており、両者には正の相関関係がある。しかしながら、一般的に考えて、両者に直接の因果関係があるとは考えにくく、擬似相関であると考えられる。①「その月の気温」→「カゼ薬への支出額」、②「その月の気温」→「セーターの購入額」という因果関係なっているものと推測される。

【演習3】

- (1) ○、×、気温
(2) ○、×、年齢
(3) ○、×、年齢