

第 43 講 相関と因果を区別する

学習目標

- ・ 相関と因果関係の違いを説明できる。
- ・ 相関が見られても、それだけで因果関係があるとは判断できないことを理解できる。
- ・ 2つの数量の関係について、共通の要因・逆向きの因果・偶然の可能性を考えられる。
- ・ データから示された事実と、そこから推測した因果関係を区別して主張を評価できる。

一緒に変化しても「原因」とは限らない

相関から分かるのは一緒に変化する傾向

2つの数量について、散布図が右上がりになったり、相関係数が正の大きな値になったりすれば、正の相関があると考えられます。

しかし、相関があることから直接、「一方が増えたから、もう一方が増えた」

と結論づけることはできません。

相関から分かるのは、2つの数量に一緒に変化する傾向が見られるということです。

その変化がどのような仕組みで起こったのかは、相関だけでは決まりません。

相関と因果関係

相関とは、2つの数量に一緒に変化する傾向が見られることです。

因果関係とは、一方の変化がもう一方の変化を引き起こす関係です。

相関が見られても、それだけで因果関係があるとは限りません。

相関と因果を分けて考える

データから、

「 x が大きいほど y も大きい傾向がある」

ことが分かっても、

「 x を大きくすれば、そのために y が大きくなる」

とまでは言えません。

一緒に変化していることと、原因になっていることは別の内容です。

共通の要因が隠れていることがある

第 3 の数量が両方に影響する

ある地域で、日ごとの「冷たい飲み物の販売数」と「冷房の使用量」を調べたところ、正の相関が見られたとします。

この結果だけを見て、

「冷たい飲み物を多く販売すると、冷房の使用量が増える」

と考えるのは適切ではありません。

たとえば、気温という別の要因を考えることができます。

気温が高い日は、冷たい飲み物が多く売れやすく、同時に冷房の使用量も増えやすくなります。

この場合、2つの数量の相関は、共通して影響を与える別の要因によって生じている可能性があります。

共通の要因

2つの数量の両方に影響を与え、見かけ上の相関を生じさせる別の要因が考えられることがあります。

相関を見つけたときは、「両方に影響する別の要因はないか」と考えることが大切です。

共通の要因を探す

2つの数量 x, y に相関が見られたら、
 x と y の両方に影響している別の数量はないか
を確認します。
見えている2つの数量だけで原因を決めないことが重要です。

原因と結果の向きが逆かもしれない

どちらが先に影響したのかを考える

ある道路について、「交通量」と「交通整理を行う人の人数」に正の相関が見られたとします。
ここから、
「交通整理を行う人が増えたから、交通量が増えた」
とは限りません。
実際には、交通量が多い場所や時間帯だから、交通整理を行う人を増やしている可能性があります。
つまり、
「交通整理を行う人の人数 → 交通量」
ではなく、
「交通量 → 交通整理を行う人の人数」
という逆向きの関係が考えられます。

因果の向きを確認する

相関がある2つの数量について、
「 x が y の原因」
と考える前に、
「反対に y が x に影響している可能性はないか」
を考えます。
相関係数の符号だけでは、原因と結果の向きは決まりません。

横軸と縦軸を入れ替えても原因は決まらない

散布図で x を横軸、 y を縦軸にしたからといって、 x が原因で y が結果になるわけではありません。
軸の置き方と因果関係は別の問題です。

偶然に相関らしく見えることもある

少ないデータでは特に注意する

たまたま集めた少数のデータで、2つの数量が右上がりに並ぶことがあります。

しかし、別の日や別の集団で調べると、その傾向が見られないこともあります。
このような場合、見えていた相関は偶然による可能性があります。
特にデータが少ないときは、一部の値の並び方によって、実際以上にはっきりした関係があるように見えることがあります。

偶然の可能性も残す

相関が見つかったときは、

- データの数は十分か
- 別のデータでも同じ傾向が見られるか

という点も考えます。
1 回の観察だけで因果関係を断定しないことが大切です。

主張を慎重に評価する

データが示す事実と解釈を分ける

たとえば、ある期間について、
「気温が高い日ほど、冷たい飲み物の販売数が多い傾向があった」
というデータが得られたとします。
このとき、
「気温と販売数に正の相関が見られた」
は、データから読み取った内容です。
一方、
「気温が上がることで販売数を増やした」
は原因についての解釈です。
因果関係を考えるには、相関以外の情報も必要になります。
データから直接分かることと、そこから推測したことを分けて表現します。

因果関係を考える 3 つの確認

相関があるという結果から因果関係を考えるときは、少なくとも次を確認します。

1. 2 つの数量の両方に影響する **共通の要因**はないか。
2. 原因と結果の **向きが逆**ではないか。
3. 見えている相関が **偶然**である可能性はないか。

これらを検討せずに「相関があるから原因である」と決めつけないようにします。

相関が強くても同じ

相関係数の絶対値が 1 に近く、強い直線的な相関が見られたとしても、それだけで因果関係が証明されたことにはなりません。
相関の強さと、因果関係があるかどうかは分けて考えます。

例題

例題 1

ある地域で、気温が高い日ほど冷たい飲み物の販売数が多く、同時にプールの利用者数も多いというデータが得られました。

この結果から、

「冷たい飲み物の販売数が増えると、プールの利用者数が増える」

と結論づけてよいでしょうか。理由を説明しなさい。

例題 1 の解答・解説

その結論を相関だけから断定することはできません。

気温が高いことによって、冷たい飲み物の販売数とプールの利用者数の両方が増えている可能性があります。

この場合、気温という共通の要因が考えられます。

したがって、2つの数量に相関が見られても、一方が他方の原因とは限りません。

例題 2

ある場所で、交通量が多い時間帯ほど交通整理を行う人の人数も多いという正の相関が見られました。

「交通整理を行う人を増やすと交通量が増える」と考えることの問題点を説明しなさい。

例題 2 の解答・解説

原因と結果の向きが逆である可能性があります。

交通整理を行う人が増えたため交通量が増えたのではなく、交通量が多いため交通整理を行う人を増やしていることが考えられます。

したがって、正の相関だけから因果の向きを決めることはできません。

例題 3

5 日間だけ調べたところ、ある店では店内の時計が示す時刻の数字が大きいほど、その時間帯の販売個数も多い傾向が見られました。

この 5 日間の結果だけから「時刻の数字が大きくなることが販売個数を増やす原因である」と結論づけることが適切でない理由を 1 つ答えなさい。

例題 3 の解答・解説

少ないデータでたまたま相関のような並びが生じた可能性があります。

また、時間帯による人の流れなど、別の要因が販売個数に影響している可能性もあります。

したがって、このデータだけから時刻の数字そのものが原因だとは判断できません。

演習問題

練習問題 1

次の主張が適切かどうか答え、理由を説明しなさい。

「2つの数量の相関係数が 0.92 だった。強い正の相関があるので、一方がもう一方の原因である。」

練習問題 2

ある地域では、気温が高い日ほどアイスクリームの販売数と電力使用量の両方が多い傾向がありました。

この 2つの数量の相関を考えると、共通の要因として考えられるものを答えなさい。

練習問題 3

ある店舗では、来客数が多い日ほど配置される店員数も多いという相関が見られました。

次の 2つの説明のうち、データだけではどちらも可能性があると考えべきものを選びなさい。

- ・ (ア) 店員数を増やしたことが来客数の増加につながった。
- ・ (イ) 多くの来客が予想される日に店員数を増やした。

練習問題 4

ある小さなデータでは、数量 x と数量 y に強い正の相関が見られました。

同じ種類のデータをさらに多く集めることには、どのような意味がありますか。「偶然」という言葉を使って説明しなさい。

練習問題 5

次のうち、相関が見られたときに確認するとよいものをすべて選びなさい。

- ・ (ア) 両方の数量に影響する別の要因がないか。
- ・ (イ) 原因と結果の向きが逆ではないか。
- ・ (ウ) 偶然に相関が見えている可能性はないか。
- ・ (エ) 相関係数が正なら必ず因果関係があると考えてよいか。

練習問題 6

ある調査で、屋外施設の利用者数と日焼け止めの販売数に正の相関が見られました。

「屋外施設の利用者数を増やせば、必ず日焼け止めの販売数が増える」と断定できない理由を、共通の要因に注目して説明しなさい。

練習問題 7

次の 2 つの文を、「データから直接読み取った内容」と「因果についての解釈」に分けなさい。

- ・ (ア) 運動施設の利用時間が長い日ほど、飲料の販売数が多い傾向が見られた。
- ・ (イ) 運動施設の利用時間が長くなったことが、飲料の販売数を増やした。

練習問題 8

ある生徒は、
「相関係数が -0.95 なら強い負の相関なので、一方が増えることがもう一方を減らす原因だ」と考えました。
この考えのどこが不十分か説明しなさい。

演習問題の解答・解説

練習問題 1 の解答・解説

適切ではありません。

$r = 0.92$ から強い正の直線的な相関があることは読み取れますが、それだけでは因果関係は分かりません。

共通の要因、逆向きの因果、偶然などの可能性を検討する必要があります。

練習問題 2 の解答・解説

共通の要因として気温が考えられます。

気温が高くなることでアイスクリームの販売数が増え、同時に冷房などによる電力使用量も増える可能性があります。

練習問題 3 の解答・解説

データだけでは (ア) と (イ) のどちらの可能性も考える必要があります。

相関があることだけでは、店員数と来客数のどちらが原因側なのかを決めることはできません。

練習問題 4 の解答・解説

データが少ないと、偶然によって相関が強く見える可能性があります。

より多くの同じ種類のデータを調べることで、見えている傾向が少数のデータだけで偶然に生じたものではないかを考えやすくなります。

練習問題 5 の解答・解説

(ア)、(イ)、(ウ) は確認するとよい内容です。

(エ) は適切ではありません。相関係数が正であることは正の相関を表しますが、因果関係があることを保証しません。

したがって、選ぶのは (ア)、(イ)、(ウ) です。

練習問題 6 の解答・解説

たとえば、天候や気温などの共通の要因が考えられます。

晴れて暑い日には屋外施設の利用者数が増えやすく、同時に日焼け止めの販売数も増えやすい可能性があります。

そのため、屋外施設の利用者数そのものが販売数を増やしたとは、相関だけから断定できません。

練習問題 7 の解答・解説

(ア) は、2 つの数量にどのような傾向が見られたかを述べているので、データから直接読み取った内容です。

(イ) は、一方の変化がもう一方を引き起こしたと述べているので、因果についての解釈です。

相関の観察と因果の主張を区別して読むことが大切です。

練習問題 8 の解答・解説

$r = -0.95$ から、強い負の直線的な相関があることは分かります。

しかし、相関係数の値だけでは因果関係があるかどうかは判断できません。

共通の要因がある可能性や、原因と結果の向きが逆である可能性、偶然の可能性などを考える必要があります。

したがって、「強い負の相関がある」ことと「一方が他方を減らす原因である」ことを同じものとして扱うのが不十分です。