第 262 講 目的に合う統計グラフを選ぶ

学習目標

- 比較、時間変化、構成、分布、二つの量の関係という目的に応じて、適切な統計グラフを選べる。
- 軸、単位、階級、割合の合計などを確認し、誤解を招かないグラフを作れる。
- 欠測値と 0 を区別し、グラフから言えることの範囲を説明できる。

1. 最初に「何を読み取りたいか」を決める

同じデータでも、目的によって表し方は変わる

統計グラフは、数値を飾るためのものではない。読み手に何を比べてほしいかを明確にする表現である。まず目的を一文にし、その後でグラフを選ぶのである。

読み取りたいこと	第一候補	選ぶ理由
項目どうしの大きさ	棒グラフ	共通の基準線から長さを比べられる
時間に伴う変化	折れ線グラフ	時間順の増減や転換を追える
全体に占める割合	帯グラフ・円グラフ	全体を 100% として構成を示せる
一つの量の分布	ヒストグラム	階級ごとの度数から形を見られる
二つの量の関係	散布図	各個体の値の組を点で示せる

一つのグラフですべてを示そうとすると、目的がぼやける。例えば、4 か月の各月における 3 項目の割合を比べるなら、時間変化と構成のどちらを重視するかを先に決める必要がある。

選択の手順

1. 観測の単位を確認する。人、日、地域など、1 行が何を表すかを明らかにする。
2. 変数が項目、時間、一つの数量、二つの数量のどれかを確認する。
3. 読み取りたいことを、比較、変化、構成、分布、関係から選ぶ。
4. 選んだグラフに必要な軸、単位、階級、凡例を付ける。

2. 項目の比較と時間変化

ある学習グループが、架空のアンケートで「復習に使いやすい教材」を 1 人一つ選んだとする。回答者は 40 人で、冊子 12 人、動画 18 人、音声 6 人、カード 4 人であった。

©ミライコーチング

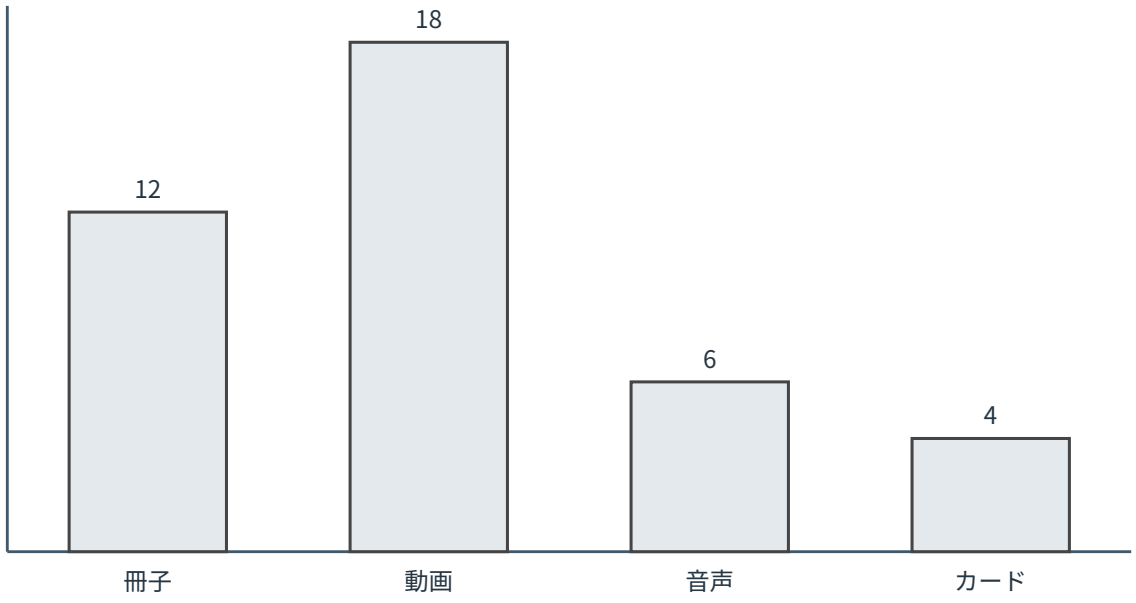

図 1　縦軸を 0 人から始めた、架空の教材別回答人数の棒グラフ

この目的は四つの項目の人数比較なので、棒グラフが適している。棒の幅をそろえ、縦軸を人数とし、基準線を 0 にする。値 12、18、6、4 の合計は 40 であり、回答者数と一致する。
一方、架空の学習記録として、ある 1 人の月ごとの読書時間が 4 月 120 分、5 月 150 分、6 月 135 分、7 月 180 分であったとする。

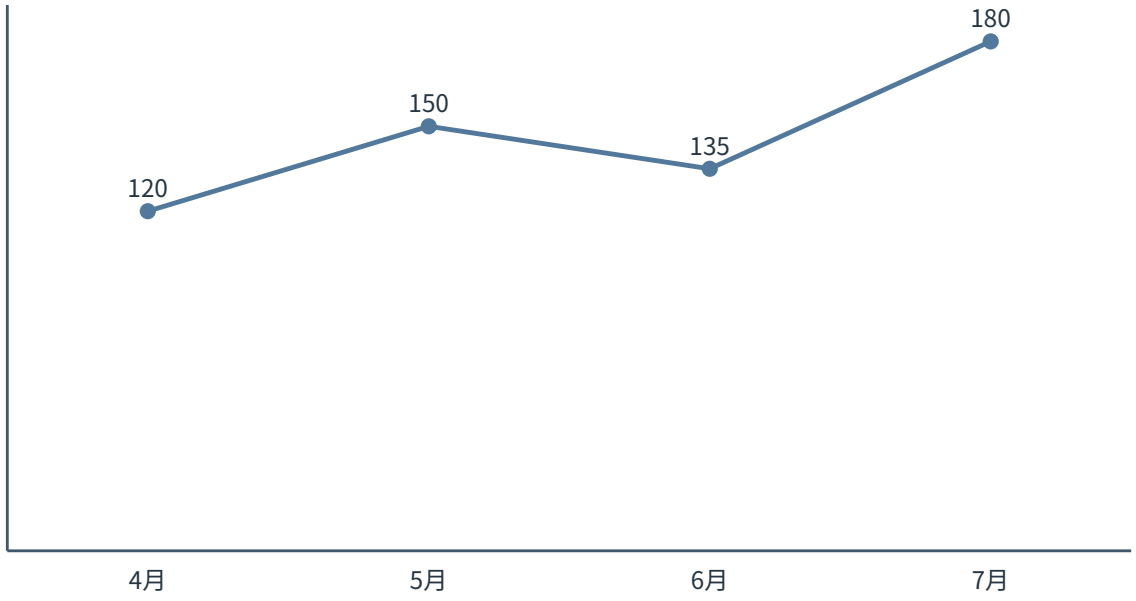

図 2　縦軸を読書時間（分）とした、ある 1 人の月別記録の折れ線グラフ

この目的は時間順の変化なので、折れ線グラフが適している。4 月から 5 月は 30 分増え、5 月から 6 月は 15 分減り、6 月から 7 月は 45 分増えている。ただし、点と点の間を結んだ線が、月の途中の値を実測したことを意味するわけではない。

©ミライコーチング

例題 1：比較か変化か

次の二つの架空データについて、棒グラフと折れ線グラフのどちらを第一候補とすべきか。理由も述べよ。
(1) 同じ日に測った A 地点、B 地点、C 地点の気温が、それぞれ 18℃、21℃、16℃である。
(2) A 地点で正午に測った気温が、月曜日 18℃、火曜日 20℃、水曜日 17℃、木曜日 21℃である。

例題 1 の解答・解説

(1) 棒グラフである。同じ日における地点という項目どうしの大きさを比べる目的だからである。横軸を地点、縦軸を気温（℃）とする。
(2) 折れ線グラフである。曜日という時間順に沿った増減を見る目的だからである。横軸を曜日、縦軸を気温（℃）とする。ただし、各値は正午の測定値であり、一日中の気温を表すものではない。

軸を切り詰めると差が大きく見える

棒グラフは棒の長さで量を比べるため、数量の軸は原則として 0 から始める。例えば 98 と 100 だけを示すために縦軸を 97 から始めると、2 の差が非常に大きく見える。やむを得ず軸を省略するなら、省略を明示し、実際の差や割合も併記する必要がある。

3. 全体の構成を割合で表す

構成を表すときは、各項目が同じ全体の一部であることを確認する。先の 40 人の回答を割合に直すと、冊子 30%、動画 45%、音声 15%、カード 10% である。

教材	冊子	動画	音声	カード	合計
人数（人）	12	18	6	4	40
割合（%）	30	45	15	10	100

百分率は「項目の人数 ÷ 全体の人数 × 100」で求める。例えば冊子は $12 \div 40 \times 100 = 30\%$ である。円グラフの中心角には百分率を 100 で割った割合を用いるので、動画は 45% を 0.45 と直し、$0.45 \times 360^\circ = 162^\circ$ となる。複数の集団で構成を横に並べて比べるなら、同じ長さを 100% とした帯グラフが適している。

例題 2：構成比を確かめる

架空の 30 冊の本を分類すると、物語 12 冊、科学 9 冊、歴史 6 冊、芸術 3 冊であった。
(1) 各分類の割合を求めよ。
(2) 1 冊を一つの分類だけに数えたとき、割合の合計を確かめよ。
(3) 項目間の冊数の差を正確に比べる目的と、全体に占める構成を見る目的に、それぞれ適するグラフを答えよ。

©ミライコーチング

例題 2 の解答・解説

(1) 物語は $\frac{12}{30} \times 100 = 40\%$、科学は 30%、歴史は 20%、芸術は 10% である。
(2) $40 + 30 + 20 + 10 = 100\%$ であり、全体と一致する。
(3) 冊数の差を比べるなら棒グラフ、全体に占める構成を見るなら円グラフまたは帯グラフが適する。目的が違うので、同じデータでも第一候補が変わるのである。

割合を足せる条件

各個体が一つの項目だけに属し、項目が全体を漏れなく覆うとき、割合の合計は 100% になる。複数回答なら、1 人が複数項目に数えられるため、合計が 100% を超えても誤りとは限らない。丸めた割合では合計が 99% や 101% になることもあるので、丸め方を記す。

4. 分布を見るヒストグラム

次の 20 個の値は、架空の小テスト得点である。

$$23, 27, 31, 34, 36, 38, 41, 42, 44, 46, 47, 49, 52, 54, 56, 58, 61, 63, 67, 72$$

幅 10 点の階級を $20 \le x < 30, 30 \le x < 40, \ldots, 70 \le x < 80$ とすると、度数は順に 2、4、6、4、3、1 であり、合計は 20 である。

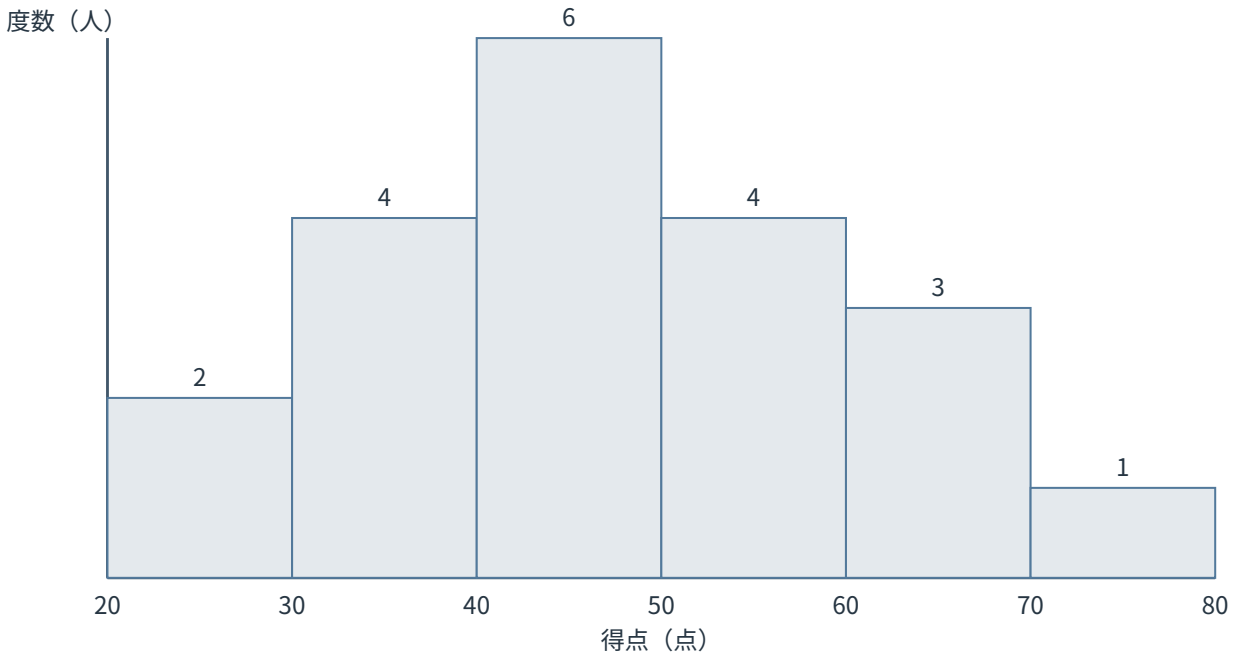

図 3　架空の小テスト得点 20 人分の分布

ヒストグラムでは横軸が連続した数量であり、隣り合う階級の柱を接して描く。例えば 40 点は $40 \le x < 50$ に入れ、$30 \le x < 40$ には入れない。境界値を二重に数えないため、階級の端を含むかどうかを決めておく。

例題 3：階級と度数

架空の待ち時間（分）12 個を

$$2, 4, 5, 7, 8, 9, 10, 12, 14, 15, 17, 19$$

とする。x を待ち時間（分）とし、階級を $0 \le x < 5, 5 \le x < 10, 10 \le x < 15, 15 \le x < 20$ として答えよ。
(1) 各階級の度数を求めよ。
(2) 目的が待ち時間の分布の形を見ることであるとき、適するグラフを答えよ。
(3) 値 10 と 15 が入る階級を答えよ。

例題 3 の解答・解説

(1) 度数は順に 2、4、3、3 である。合計は $2+4+3+3=12$ となり、データ数と一致する。
(2) ヒストグラムが適する。一つの数量を階級に分け、度数の分布を見る目的だからである。
(3) 10 は $10 \le x < 15$、15 は $15 \le x < 20$ に入る。各階級を「以上、未満」で統一したためである。

階級幅が異なる場合

この講では階級幅をそろえたヒストグラムを扱う。階級幅が異なるとき、柱の高さを度数のままにすると面積が度数に対応しない。度数を階級幅で割った度数密度を高さに用いる必要がある。

複数の集団について、中央値や散らばりを簡潔に比べる目的には箱ひげ図も使える。次は、二つの架空の組について、最小値、第 1 四分位数、中央値、第 3 四分位数、最大値を並べたものである。

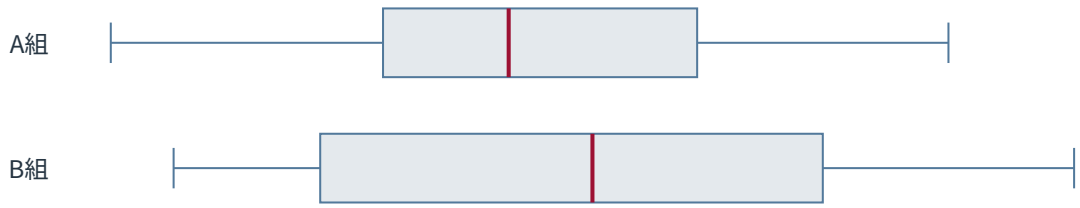

図 4　架空の 2 組の五数要約を表す箱ひげ図

次の二本の数直線は、五数要約の値を確かめるための別図であり、それぞれ 30 点から 80 点までを数値付きで示す。箱ひげ図とは表示幅が異なるため、紙面上の横位置を直接対応させない。A 組の五数要約は 32、45、51、60、72 点、B 組は 35、42、55、66、78 点であり、箱ひげ図の読み取りにはこれらの数値を用いる。

©ミライコーチング



図 5　A 組の五数要約（点）



図 6　B 組の五数要約（点）

A 組の中央値は 51 点、B 組の中央値は 55 点である。また、四分位範囲は A 組が $60-45=15$ 点、B 組が $66-42=24$ 点であり、中央 50% の散らばりは B 組の方が大きい。ただし、箱ひげ図だけでは山の数や個々の値までは分からない。分布の詳しい形を見る目的なら、ヒストグラムなども併用するのである。

5. 二つの量の関係を見る散布図

次の表は、架空の 8 日について、学習時間と確認テスト得点を同じ日に記録したものである。

日	1	2	3	4	5	6	7	8
学習時間（分）	20	30	35	45	50	60	70	80
得点（点）	48	55	52	63	68	66	78	82

©ミライコーチング

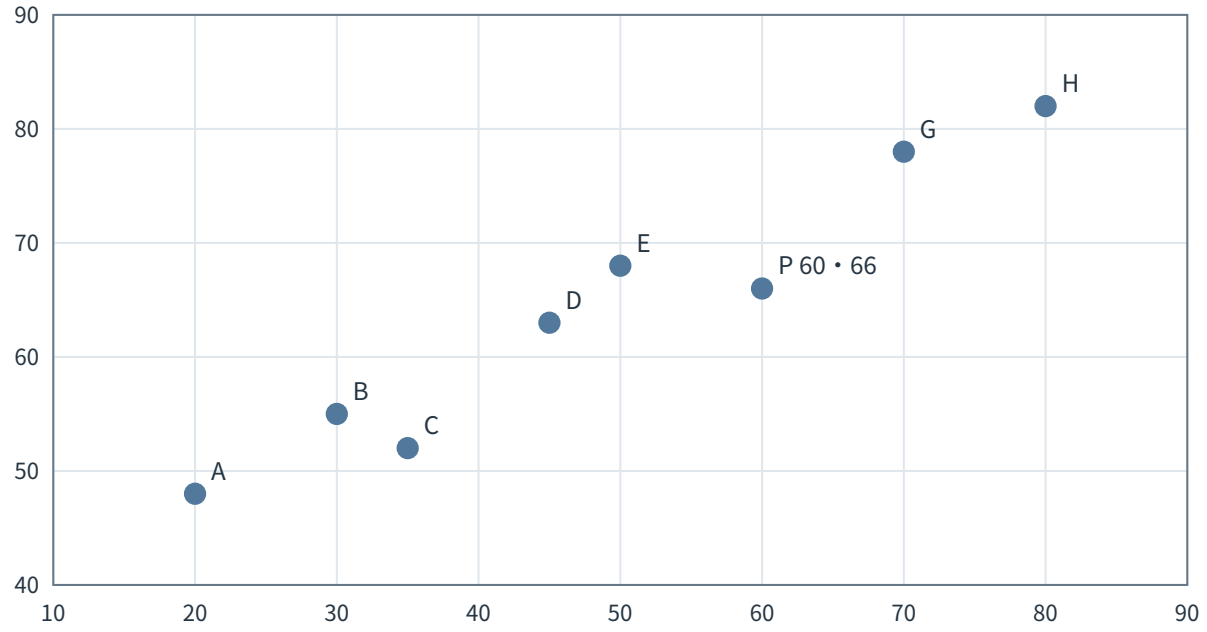

図 7　横軸を学習時間（分）、縦軸を得点（点）とした架空の 8 日分の散布図

右上がりの傾向があるので、この 8 日では学習時間が長い日ほど得点が高い傾向が見られる。しかし、散布図だけから「学習時間を増やしたことが得点上昇の原因である」とは断定できない。睡眠、問題の難しさ、復習内容など、別の要因が関係する可能性があるからである。

例題 4：散布図から言えること

上の散布図について答えよ。
(1) 横軸と縦軸が表す量と単位を答えよ。
(2) 点 $(60, 66)$ の意味を答えよ。
(3) この散布図だけから「60 分学習すれば必ず 66 点になる」と言えない理由を述べよ。

例題 4 の解答・解説

(1) 横軸は学習時間（分）、縦軸は得点（点）である。
(2) ある 1 日は、学習時間が 60 分で得点が 66 点であったことを表す。一つの点は同じ観測単位から得た値の組である。
(3) 点は過去の架空の 8 日分の観測であり、同じ学習時間でも得点が必ず一定になるとは示していない。また、散布図の傾向は因果関係を保証しないからである。

6. 欠測値、0、対象範囲を明記する

0 は「測定した結果が 0」であり、欠測は「値が得られていない」である。この二つを同じ 0 としてグラフに入れてはいけない。例えば月曜日から金曜日の歩数が 4200、5100、欠測、4800、0 であるとする。水曜日は記録なし、金曜日は測定できて実際に 0 歩であったという条件なら、意味が異なる。

折れ線グラフで欠測の前後を線で結ぶと、測っていない値が存在したように見える。欠測の位置で線を切り、注記する。さらに、標本が誰または何を含むか、期間、単位、複数回答の可否も題名や注記に示すのである。

作成後の点検

- 題名から対象、期間、目的が分かるか。
- 横軸と縦軸に量の名前と単位があるか。
- 棒の基準線、時間順、階級の端、値の組が正しいか。
- 度数の合計がデータ数と一致するか。構成比は条件に応じて 100% になるか。
- 0 と欠測を区別し、観測していない部分を補ったように見せていないか。
- 傾向を因果関係として言い換えていないか。

演習問題

問題 1：目的から選ぶ

次の目的に最も適するグラフを、棒グラフ、折れ線グラフ、帯グラフ、ヒストグラム、散布図から一つずつ選べ。
(1) 5 種類の文房具の販売個数を比べる。
(2) ある日の 1 時間ごとの気温の変化を見る。
(3) 二つのクラスで通学手段の割合を比べる。
(4) 50 人の身長の分布を見る。
(5) 50 人それぞれの身長と体重の関係を見る。

問題 2：割合と合計

架空の 50 人の希望を一人一つ集計すると、A が 20 人、B が 15 人、C が 10 人、D が 5 人であった。各項目の割合を求め、合計を確かめよ。また、構成を見るのに適するグラフを一つ答えよ。

問題 3：ヒストグラムの度数

架空のデータ

$$11, 14, 18, 20, 22, 25, 27, 31, 33, 39$$

を、$10 \le x < 20, 20 \le x < 30, 30 \le x < 40$ の三階級に分ける。各階級の度数を求め、合計を確かめよ。また、値 20 が入る階級を答えよ。

問題 4：欠測と 0

ある機器の 4 日間の出力回数が、1 日目 8 回、2 日目は記録なし、3 日目 0 回、4 日目 5 回であった。2 日目を 0 として扱ってはいけない理由と、折れ線グラフを作るときの表し方を述べよ。

問題 5：誤解を招く軸

商品 A の販売数は 98 個、商品 B は 100 個であった。縦軸を 97 個から始める棒グラフが差を実際以上に大きく見せやすい理由を述べ、適切な直し方を答えよ。

©ミライコーチング

問題 6：散布図の解釈

架空の散布図で、運動時間が長い人ほど測定値が高い右上がりの傾向が見られた。この事実だけから「運動時間を増やせば、すべての人の測定値が必ず高くなる」と結論してよいか。理由を述べよ。

解答・解説

問題 1 の解答・解説

(1) 棒グラフである。項目別の個数を比べるためである。
(2) 折れ線グラフである。時間順の増減を見るためである。
(3) 帯グラフである。各クラスを 100% として構成を横に比べるためである。
(4) ヒストグラムである。一つの連続的な数量を階級に分け、分布を見るためである。
(5) 散布図である。同じ 50 人から得た二つの量の値の組を点で表し、関係を見るためである。

問題 2 の解答・解説

A は $\frac{20}{50}\times 100=40\%$、B は 30%、C は 20%、D は 10% である。合計は $40+30+20+10=100\%$ となる。一人一つの回答で全員を分類している条件と整合する。構成を見るなら円グラフまたは帯グラフが適する。

問題 3 の解答・解説

$10\le x<20$ は 11、14、18 の 3 個、$20\le x<30$ は 20、22、25、27 の 4 個、$30\le x<40$ は 31、33、39 の 3 個である。度数は 3、4、3、合計は 10 でデータ数と一致する。20 は左端を含む $20\le x<30$ に入る。

問題 4 の解答・解説

2 日目は値が得られていない欠測であり、測定して 0 回だった 3 日目とは意味が異なる。2 日目を 0 にすると、実際に 0 回だったという誤った情報になる。折れ線グラフでは 2 日目を欠測と注記し、その位置で線を切る。1 日目と 3 日目を直接結ばないのである。

問題 5 の解答・解説

棒は基準線からの長さで量を表す。97 を基準にすると、見かけの長さが A は 1、B は 3 となり、実際の比 98 : 100 より差が大きく見える。縦軸を 0 から始め、98 個と 100 個という値も表示するのが適切である。

問題 6 の解答・解説

結論してはいけない。右上がりは観測された範囲で二つの量に関係がある傾向を示すが、因果関係や全員についての必然性は示さない。年齢、生活習慣など別の要因が両方に関係している可能性もあり、点の散らばりや対象範囲も確認する必要がある。