

0001

1-1 データと分布 > データソース・公的統計 | 難易度：基礎

都道府県別の人口を比較するレポートを作る。出典の信頼性と再現性を最優先するとき、最初に確認すべき資料はどれか。

- ア．政府統計の総合窓口等で、作成機関・調査名・時点・定義が確認できる公的統計表。
- イ．出典不明のSNS投稿を集計したランキング。
- ウ．検索結果の見出しだけを転記した一覧。
- エ．個人ブログの数値を、出典をたどらずそのまま採用した表。

答え・解説

正答：ア

- ア：公的統計では調査主体や定義などのメタデータを確認でき、比較条件を検証しやすい。
 - イ：作成方法や母集団が確認できず、一次情報源として不適切である。
 - ウ：元資料の定義や時点を確認できない。
 - エ：二次的な説明だけでは元統計の条件を検証できない。
- 総合解説：統計を利用するときは数値だけでなく、作成機関、調査名、対象、時点、単位、定義などのメタデータを確認する。

対象：2026年度 / 基準日 2026-09-08

0002

1-1 データと分布 > データソース・公的統計 | 難易度：基礎

同じ「人口」という名称の2表を比較したところ、一方は10月1日現在の人、他方は年度平均の千人単位だった。比較前の対応として最も適切なのはどれか。

- ア．表題が同じなので、そのまま差を計算する。
- イ．大きい方の値だけを採用する。
- ウ．基準時点と単位をそろえ、同じ概念を測る値か確認してから比較する。
- エ．単位だけを消して無次元量として扱う。

答え・解説

正答：ウ

- ア：表題が同じでも基準時点や単位が異なれば直接比較できない。
 - イ：大きさでデータの妥当性は決まらない。
 - ウ：時点・単位・定義が異なる値をそのまま比較すると誤解を招くため、条件の統一が必要である。
 - エ：単位を消しても測定条件の違いは解消しない。
- 総合解説：公的統計の比較では、対象期間・基準時点・単位・定義をそろえることが基本である。

対象：2026年度 / 基準日 2026-09-08

0003

1-1 データと分布 > データソース・公的統計 | 難易度：標準

月次統計について、速報公表後に確報で値が改訂された。過去の分析結果を更新する際の最も適切な対応はどれか。

- ア．速報値は一度公表されたので、確報を無視する。
- イ．速報と確報を無条件に平均する。
- ウ．分析目的に応じて最新の確報値を用い、使用した版や公表時点を記録する。
- エ．改訂があった月のデータをすべて削除する。

答え・解説

正答：ウ

ア：確報で改訂された事実を無視すると最新値と不整合になる。
イ：両者は同じ量の異なる版であり、平均する合理的根拠はない。
ウ：改訂のある統計では版の違いが再現性に影響するため、どの値を用いたか記録するのが適切である。
エ：改訂は通常、削除理由ではなく更新理由である。
総合解説：時点更新される統計では、データの版・取得日・確報／速報の区別を記録すると再現性が高まる。

対象：2026年度 / 基準日 2026-09-08

0004

1-1 データと分布 > データソース・公的統計 | 難易度：標準

消費者物価指数で基準時を100とし、ある時点の総合指数が106だった。この値の解釈として適切なものはどれか。

- ア．すべての商品価格が一律に6円上がったことを表す。
- イ．対象世帯の所得が6%増えたことを表す。
- ウ．物価上昇率が毎月必ず6%であることを表す。
- エ．基準時に対して、対象となる財・サービスの価格水準が指数上およそ6%高いことを表す。

答え・解説

正答：エ

ア：指数は円額の一律変化を意味しない。
イ：CPIは所得ではなく消費者物価の変動を示す。
ウ：基準時との水準比較であり、各月の上昇率を直接示す値ではない。
エ：指数106は基準100に対する相対水準を示す。個別価格の単純合計ではない。
総合解説：CPIは基準時を100とした価格水準の指数であり、指数値と各期間の変化率を区別する。

対象：2026年度 / 基準日 2026-09-08

0005

1-1 データと分布 > データソース・公的統計 | 難易度：標準

e-Statで「失業率」の長期推移を探したい。検索結果が多いとき、効率的な確認手順として最も適切なのはどれか。

- ア．検索結果の先頭だけを無条件で採用する。
- イ．統計名・作成機関・調査年月・地域などの条件を確認し、目的に合う統計表を選ぶ。
- ウ．数値が最も大きい表を採用する。
- エ．調査年月を確認せず、異なる年の表を結合する。

答え・解説

正答：イ

- ア：順位だけでは目的に合う系列か判断できない。
 - イ：検索語だけでなく統計名や時間・地域などのメタデータで絞ると目的の系列を特定しやすい。
 - ウ：値の大きさは検索目的との適合性を示さない。
 - エ：時点を確認しない結合は時系列比較を壊す。
- 総合解説：統計データ検索では、統計名、作成機関、対象地域、時点、表章項目等を確認する。

対象：2026年度 / 基準日 2026-09-08

0006

1-1 データと分布 > データソース・公的統計 | 難易度：標準

研究者が自ら質問票を設計して回答を収集したデータと、政府が既に公表している統計表を利用したデータの組合せとして適切なのはどれか。

- ア．前者も後者も必ず実験データである。
- イ．前者は研究者が直接収集した一次データ、後者は既存資料を利用する二次データとして扱える。
- ウ．前者は二次データ、後者は一次データである。
- エ．公的統計はデータではなく、分析に利用できない。

答え・解説

正答：イ

- ア：質問票調査や公表統計は実験とは限らない。
 - イ：自ら調査して収集したデータと、既存の統計資料を再利用するデータでは取得経路が異なる。
 - ウ：取得経路の説明が逆である。
 - エ：公的統計は代表的な二次データ源になり得る。
- 総合解説：データソースを明示すると、収集過程や利用上の制約を把握しやすい。

対象：2026年度 / 基準日 2026-09-08

0007

1-1 データと分布 > データソース・公的統計 | 難易度：応用

10年間の公的統計を折れ線で示す際、6年目から調査票の定義が変更されたことが注記にあった。最も適切な分析上の対応はどれか。

ア．変更内容を確認し、系列の比較可能性に影響するなら図や分析で区切り・注記を示す。
イ．注記は数値ではないので無視する。
ウ．変更前の値をすべて0に置き換える。
エ．6年目以降の値を自動的に2倍する。

答え・解説

正答：ア

ア：定義変更は見かけの増減を生むことがあるため、系列の接続可能性を検討する必要がある。
イ：定義変更は数値の意味に影響するため無視できない。
ウ：定義変更への適切な補正ではない。
エ：補正係数は変更内容の根拠なしに決められない。
総合解説：長期時系列では調査方法・定義・基準改定の有無を確認し、構造的な非連続をデータ変化と混同しない。

対象：2026年度 / 基準日 2026-09-08

0008

1-1 データと分布 > データソース・公的統計 | 難易度：応用

複数の公的統計を用いた分析を第三者が再現できるようにする。最低限残す情報として最も有用なのはどれか。

ア．最終グラフの画像だけを保存する。
イ．統計名、表番号や取得先、対象期間、取得日、使用変数、単位などを記録する。
ウ．分析者の感想だけを記録する。
エ．URLを省略し、検索語だけを記録する。

答え・解説

正答：イ

ア：元データや取得条件が分からず再現できない。
イ：利用したデータを一意に追跡できる情報を残すことが再現性につながる。
ウ：データ識別情報がないため再現性に寄与しない。
エ：検索結果は変化し得るため、具体的な出典情報が望ましい。
総合解説：再現可能な統計分析には、データの出典と取得条件を追跡できる記録が必要である。

対象：2026年度 / 基準日 2026-09-08

0009

1-1 データと分布 > 質的・量的変数・尺度水準 | 難易度：基礎

アンケートの「居住地域＝東・西・南・北」はどの尺度として扱うのが最も適切か。

- ア．順序尺度。北が東より常に大きい。
- イ．名義尺度。カテゴリ間に本質的な大小順序を置かない。
- ウ．間隔尺度。東と西の差が一定である。
- エ．比率尺度。北は南の2倍である。

答え・解説

正答：イ

ア：方角に統計的な大小順序はない。
イ：方角カテゴリは識別ラベルであり、通常は大小関係を前提としない。
ウ：カテゴリ差に数量的な等間隔は定義されない。
エ：意味のあるゼロや比は存在しない。
総合解説：尺度水準は数値コードの有無ではなく、値の意味と許される比較で判断する。

対象：2026年度 / 基準日 2026-09-08

0010

1-1 データと分布 > 質的・量的変数・尺度水準 | 難易度：基礎

顧客満足度を「不満・やや不満・普通・やや満足・満足」の5段階で記録した。最も適切な説明はどれか。

- ア．順序も意味を持たない名義尺度である。
- イ．必ず比率尺度であり、満足は普通の2倍である。
- ウ．順序は意味を持つが、隣り合うカテゴリ間の差が等しいとは限らない。
- エ．カテゴリ数が5なので連続変数である。

答え・解説

正答：ウ

ア：満足度には自然な順序がある。
イ：カテゴリ間の比に意味はない。
ウ：順序尺度では大小の順序は扱えるが、間隔の等しさは保証されない。
エ：カテゴリ数だけで連続性は決まらない。
総合解説：順序尺度では順位情報を利用できる一方、差や比の数量的解釈には注意する。

対象：2026年度 / 基準日 2026-09-08

0011 1-1 データと分布 > 質的・量的変数・尺度水準 | 難易度：基礎

摂氏温度20 と10 について「20 は10 の2倍の温かさ」と言えない主な理由はどれか。

ア．摂氏0 は量の完全な不在を表す絶対的ゼロではないから。

イ．摂氏温度には順序がないから。

ウ．摂氏温度は質的変数だから。

エ．20と10の差が計算できないから。

答え・解説 正答：ア

ア：比の解釈には意味のある絶対的ゼロが必要で、摂氏温度は間隔尺度として扱われる。

イ：温度の高低という順序はある。

ウ：数値として差を扱える量的変数である。

エ：差10 は意味を持つ。

総合解説：間隔尺度では差が意味を持つが、比率尺度のような倍率解釈は一般にできない。

対象：2026年度 / 基準日 2026-09-08

0012 1-1 データと分布 > 質的・量的変数・尺度水準 | 難易度：標準

次のうち、比率尺度として扱うのが最も自然な変数はどれか。

ア．荷物の質量（kg）。0kgが量の不在を表し、比も意味を持つ。

イ．西暦年。

ウ．学年（1年、2年、3年）。

エ．血液型コード（A=1,B=2等）。

答え・解説 正答：ア

ア：質量には絶対的ゼロがあり、2kgは1kgの2倍という比の解釈ができる。

イ：年の差は扱えるが、0年を絶対的ゼロとして倍率解釈する尺度ではない。

ウ：順序カテゴリであり、3年が1年の3倍という意味ではない。

エ：数値コードでも名義尺度の性質は変わらない。

総合解説：量的変数でも尺度水準は異なる。意味のあるゼロと比の解釈が比率尺度の特徴である。

対象：2026年度 / 基準日 2026-09-08

0013

1-1 データと分布 > 質的・量的変数・尺度水準 | 難易度：標準

- 「1日に受けた問い合わせ件数」と「1件あたりの対応時間」の型の組合せとして適切なのはどれか。
- ア．両方とも名義尺度である。
 - イ．件数は連続型、時間は離散型である。
 - ウ．両方とも必ず順序尺度である。
 - エ．件数は離散型、対応時間は測定精度の範囲で連続型として扱う。

答え・解説

正答：エ

ア：どちらも量的な測定値である。
イ：一般的な分類が逆である。
ウ：数量として差や比を扱える。
エ：件数は整数のカウント、時間は連続量としてモデル化するのが一般的である。
総合解説：離散型は可算的な値、連続型は区間内の値を連続的に取り得る量として考える。

対象：2026年度 / 基準日 2026-09-08

0014

1-1 データと分布 > 質的・量的変数・尺度水準 | 難易度：標準

- 部署を「営業=1、開発=2、総務=3」とコード化した。コード2の平均を計算することについて最も適切な説明はどれか。
- ア．1,2,3なので自動的に間隔尺度になる。
 - イ．3は1の3倍の部署規模を意味する。
 - ウ．名義尺度では度数も数えられない。
 - エ．コードは名義的なラベルなので、数値の平均に部署間の数量的意味は通常ない。

答え・解説

正答：エ

ア：コード値の並びだけでは尺度水準は変わらない。
イ：コードは規模を表していない。
ウ：カテゴリごとの度数は集計できる。
エ：数値で保存していても、部署コードに間隔や倍率の意味はない。
総合解説：分析方法は保存形式ではなく、変数が表す意味に基づいて選ぶ。

対象：2026年度 / 基準日 2026-09-08

0015

1-1 データと分布 > 質的・量的変数・尺度水準 | 難易度：標準

身長をcmからmに変換した。この操作について最も適切なのはどれか。

ア．同じ比率尺度の量を単位換算しただけで、個体の順序や比率関係は変わらない。

イ．量的変数が名義尺度へ変わる。

ウ．順位が必ず逆転する。

エ．標本数が100分の1になる。

答え・解説

正答：ア

ア：正の定数倍による単位換算では順序と比率関係が保たれる。

イ：単位換算で尺度の性質は変わらない。

ウ：正の定数倍なので順序は保たれる。

エ：値の単位を変えても観測数は変わらない。

総合解説：単位変換と尺度水準の変更を混同しない。

対象：2026年度 / 基準日 2026-09-08

0016

1-1 データと分布 > 質的・量的変数・尺度水準 | 難易度：標準

疾病の有無を0/1で記録した。平均値0.27が得られたときの解釈として適切なのはどれか。

ア．1を「疾病あり」と定義しているなら、平均0.27は標本中の疾病ありの割合0.27に等しい。

イ．疾病の重症度の平均が0.27である。

ウ．変数が比率尺度になったので「1は0の無限倍」と解釈する。

エ．0/1変数では平均を計算できない。

答え・解説

正答：ア

ア：0/1変数では標本平均は1の出現割合になる。

イ：0/1コードは有無を表し、重症度を測っていない。

ウ：カテゴリコードに倍率の意味はない。

エ：指示変数として平均を割合として利用できる。

総合解説：名義的な二値変数も0/1指示変数として数値計算に利用できるが、コードの意味を明示する。

対象：2026年度 / 基準日 2026-09-08

0017

1-1 データと分布 > 質的・量的変数・尺度水準 | 難易度：応用

「競技順位（1位、2位、...）」を要約する場合、尺度の性質に最も沿う考え方はどれか。

ア．順位は比率尺度なので、2位は1位の2倍悪いと解釈する。

イ．順位や中央値など順序情報を利用し、順位差を等間隔の数量差と当然視しない。

ウ．順位には順序がないので最頻値しか使えない。

エ．順位を西暦年と同じ間隔尺度として扱う必要がある。

答え・解説

正答：イ

ア：順位の比にそのような意味はない。

イ：順位は順序尺度なので、順序を使う要約が自然である。

ウ：順位そのものが順序情報を持つ。

エ：順位間の実力差が等しいとは限らない。

総合解説：順序尺度では順番は比較できるが、差や比の解釈には追加仮定が必要である。

対象：2026年度 / 基準日 2026-09-08

0018

1-1 データと分布 > 質的・量的変数・尺度水準 | 難易度：応用

年齢を実数値のままではなく「20代・30代・40代...」にまとめた。統計的な影響として最も適切なのはどれか。

ア．カテゴリ化すると必ず情報量が増える。

イ．カテゴリ化後も25歳と29歳の差4年がそのまま保持される。

ウ．元の細かな量的情報の一部が失われ、同じカテゴリ内の年齢差は表現されなくなる。

エ．カテゴリ化すると標本数が0になる。

答え・解説

正答：ウ

ア：細分値をまとめるため情報は通常減る。

イ：同じ20代カテゴリでは内部差が表現されない。

ウ：カテゴリ化は解釈を簡単にする一方、元の値が持つ細かな差を捨てる。

エ：観測数自体は変わらない。

総合解説：量的変数のカテゴリ化には情報損失があるため、目的に応じて使い分ける。

対象：2026年度 / 基準日 2026-09-08

0019 1-1 データと分布 > 度数分布・グラフ・分布形状 | 難易度：基礎

得点データ 41, 48, 52, 57, 63, 66, 74, 78 を「40以上50未満、50以上60未満、60以上70未満、70以上80未満」に区切る。度数の並びはどれか。

- ア . 1, 3, 2, 2。52を40台に含める。
- イ . 2, 2, 2, 2。各階級に2個ずつ入る。
- ウ . 2, 1, 3, 2。57を60台に含める。
- エ . 2, 2, 1, 3。66を70台に含める。

答え・解説 正答：イ

ア：52は50以上60未満の階級に入る。
イ：境界を左閉右開として数えると各階級2個である。
ウ：57は50以上60未満に入る。
エ：66は60以上70未満に入る。
総合解説：階級境界の含み方を明確にして度数を数える。

対象：2026年度 / 基準日 2026-09-08

0020 1-1 データと分布 > 度数分布・グラフ・分布形状 | 難易度：基礎

50人中、ある階級の度数が15人だった。相対度数として適切なのはどれか。

- ア . 0.15。度数15を100で割る。
- イ . 3.33。50 ÷ 15で求める。
- ウ . 0.30 (30%)。15 ÷ 50で求める。
- エ . 35%。50 - 15を百分率にする。

答え・解説 正答：ウ

ア：総度数50で割る必要がある。
イ：比の向きが逆である。
ウ：相対度数は階級度数を総度数で割る。
エ：これは他の階級の合計割合に対応する。
総合解説：相対度数の合計は原則1 (100%) となる。

対象：2026年度 / 基準日 2026-09-08

0021

1-1 データと分布 > 度数分布・グラフ・分布形状 | 難易度：基礎

「60点未満」の累積相対度数が0.72だった。最も適切な解釈はどれか。

- ア．60点の観測だけが72％である。
- イ．観測値の28％が必ず60点ちょうどである。
- ウ．平均点が72点である。
- エ．観測値の72％が60点未満に含まれる。

答え・解説

正答：エ

ア：累積値は1点だけの比率ではない。
イ：補数は60点以上の割合に対応し、60点ちょうどとは限らない。
ウ：累積相対度数と平均は別の指標である。
エ：累積相対度数は指定した上限までの観測割合を表す。
総合解説：累積相対度数は分布の下側にどれだけ観測が含まれるかを示す。

対象：2026年度 / 基準日 2026-09-08

0022

1-1 データと分布 > 度数分布・グラフ・分布形状 | 難易度：基礎

連続量である製品重量の分布形状を確認したい。最も適切な図はどれか。

- ア．製品番号を横軸にした円グラフ。
- イ．重量をカテゴリ名として並べたレーダーチャート。
- ウ．重量を適切な階級に分けたヒストグラム。
- エ．中央値だけを1本の棒で示す棒グラフ。

答え・解説

正答：ウ

ア：分布形状を把握しにくい。
イ：連続量の度数分布を見る目的には不適切である。
ウ：連続量の分布形状を見るには階級ごとの度数を連続した長方形で表すヒストグラムが適する。
エ：1つの代表値だけでは分布形状が分からない。
総合解説：図は変数型と分析目的に合わせて選ぶ。

対象：2026年度 / 基準日 2026-09-08

0023

1-1 データと分布 > 度数分布・グラフ・分布形状 | 難易度：標準

同じデータから階級幅2と階級幅20の2種類のヒストグラムを作った。最も適切な説明はどれか。

- ア．階級幅を変えても棒の形は必ず完全に同じになる。
- イ．階級幅は必ず1にしなければならない。
- ウ．階級幅が広すぎると細かな構造が隠れ、狭すぎると偶然変動が強調されることがある。
- エ．階級幅を広げるほど標本数が増える。

答え・解説

正答：ウ

ア：集約方法が変わるため形は変わり得る。
イ：唯一の固定幅が常に最適とは限らない。
ウ：ヒストグラムの見え方はビン幅に依存するため、目的に応じた幅の検討が必要である。
エ：観測数自体は変わらない。
総合解説：ヒストグラムは階級設定に敏感であり、複数の幅を確認することも有用である。

対象：2026年度 / 基準日 2026-09-08

0024

1-1 データと分布 > 度数分布・グラフ・分布形状 | 難易度：標準

階級Aは幅5・度数20、階級Bは幅10・度数30である。面積を度数に比例させるヒストグラムの高さ（度数密度）はどれか。

- ア．Aは4、Bは3。度数を階級幅で割る。
- イ．Aは20、Bは30。度数をそのまま高さにする。
- ウ．Aは0.25、Bは0.33。幅を度数で割る。
- エ．Aは100、Bは300。度数と幅を掛ける。

答え・解説

正答：ア

ア： $20/5=4$ 、 $30/10=3$ であり、面積はそれぞれ20と30になる。
イ：幅が異なる場合、面積が度数に比例しない。
ウ：度数密度は度数÷幅である。
エ：高さの定義が逆である。
総合解説：不等階級幅では高さではなく長方形の面積が度数を表すよう度数密度を使う。

対象：2026年度 / 基準日 2026-09-08

0025

1-1 データと分布 > 度数分布・グラフ・分布形状 | 難易度：標準

多くの値が小さい側に集まり、少数の非常に大きい値が右側へ長く伸びているヒストグラムである。形状の表現として適切なのはどれか。

- ア．左右完全対称の分布。
- イ．左裾が長い分布。
- ウ．必ず二峰性の分布。
- エ．右裾が長い（正の歪みをもつ）分布。

答え・解説

正答：エ

ア：右側の尾が長いので対称ではない。
イ：尾の方向が逆である。
ウ：裾の長さだけから山が2つとは言えない。
エ：大きい側に長い尾がある分布は右に歪んだ分布と表現する。
総合解説：分布形状では中心だけでなく、対称性、裾、山の数などを見る。

対象：2026年度 / 基準日 2026-09-08

0026

1-1 データと分布 > 度数分布・グラフ・分布形状 | 難易度：標準

ヒストグラムに明瞭な山が2つあり、その中間の度数が少ない。まず考えるべきこととして適切なのはどれか。

- ア．平均値が必ず0であると判断する。
- イ．異なる集団が混在している可能性を含め、層別要因を確認する。
- ウ．標本数が必ず2であると判断する。
- エ．全データが同じ値であると判断する。

答え・解説

正答：イ

ア：山の数から平均0とは言えない。
イ：二峰性は複数のサブグループの混合などで生じることがある。
ウ：2つの山と標本数は無関係である。
エ：同値なら山は通常1つである。
総合解説：分布の複数ピークは層別や生成過程の違いを調べる手掛かりになる。

対象：2026年度 / 基準日 2026-09-08

0027

1-1 データと分布 > 度数分布・グラフ・分布形状 | 難易度：標準

ヒストグラムでは主なデータが0～20に集中し、100付近に1件だけ離れた棒がある。次の対応として最も適切なのはどれか。

- ア．図から離れているので自動的に削除する。
- イ．100を20に置き換えて分布を滑らかにする。
- ウ．入力ミスか実在する極端値かを原データで確認し、理由なく削除しない。
- エ．外れ値が1件あると平均は計算不能になる。

答え・解説

正答：ウ

ア：外れ値候補であることだけでは削除根拠にならない。
イ：恣意的な値変更になる。
ウ：離れた値はエラーの場合も実値の場合もあるため、根拠を確認する。
エ：平均は計算できるが影響を受ける可能性がある。
総合解説：外れ値候補はデータ品質確認と分析上の感度検討の対象である。

対象：2026年度 / 基準日 2026-09-08

0028

1-1 データと分布 > 度数分布・グラフ・分布形状 | 難易度：標準

標本数100の学校Aと標本数500の学校Bで得点分布を比較したい。階級別の構成比を比べる目的で最も適切なのはどれか。

- ア．両校の度数をそのまま比較し、標本数の差は無視する。
- イ．各学校で相対度数を求め、同じ階級境界で比較する。
- ウ．Aの全度数を5倍すれば必ず同じ分布になる。
- エ．階級境界を学校ごとに無関係に変える。

答え・解説

正答：イ

ア：度数は標本数に強く依存する。
イ：標本数が異なると生の度数は規模の影響を受けるため、割合で比較すると構成を見やすい。
ウ：倍率補正だけで分布形状が同じとは限らない。
エ：同じ区間でないと階級別比較が難しい。
総合解説：標本規模の異なる分布比較では同一階級と相対度数が有用である。

対象：2026年度 / 基準日 2026-09-08

0029

1-1 データと分布 > 度数分布・グラフ・分布形状 | 難易度：応用

2群の得点について「70点以下の割合」を直接比較したい。度数分布表から最も直接的に使える量はどれか。

- ア．70点を含む階級の度数だけ。
- イ．最大値だけ。
- ウ．70点以下までの累積相対度数。
- エ．階級幅の平均だけ。

答え・解説

正答：ウ

- ア：それより低い階級を含まない。
 - イ：閾値以下の割合は分らない。
 - ウ：閾値以下の割合は累積相対度数で直接表せる。
 - エ：観測割合とは無関係である。
- 総合解説：累積相対度数は特定の値以下（または未満）の割合を読むのに適する。

対象：2026年度 / 基準日 2026-09-08

0030

1-1 データと分布 > 度数分布・グラフ・分布形状 | 難易度：応用

年度別割合が48%, 50%, 52%と変化した棒グラフで、縦軸を47%から始めると変化が非常に大きく見えた。最も適切な評価はどれか。

- ア．縦軸を47%から始めると元データも自動的に47%減る。
- イ．割合のグラフでは軸目盛を表示してはいけない。
- ウ．軸範囲は図の印象に一切影響しない。
- エ．切断軸は差を視覚的に強調するため、軸範囲を明示し、目的に応じて誤解のない表現を選ぶ必要がある。

答え・解説

正答：エ

- ア：軸表示と元データは別である。
 - イ：目盛は解釈に重要である。
 - ウ：視覚的な差の大きさに影響する。
 - エ：軸の切断は数値自体を変えないが、見かけの差を拡大できる。
- 総合解説：可視化では尺度・軸・基準線などの選択が解釈に与える影響を意識する。

対象：2026年度 / 基準日 2026-09-08

0031

1-2 代表値・散布度 > 平均・中央値・最頻値 | 難易度：基礎

午前クラス20人の平均が70点、午後クラス30人の平均が80点だった。50人全体の平均として適切なものはどれか。

- ア . 75点。2クラスの平均を単純平均する。
- イ . 76点。人数で重み付けして $(20 \times 70 + 30 \times 80) / 50$ とする。
- ウ . 80点。人数が多いクラスの平均だけを採る。
- エ . 150点。2クラスの平均を足す。

答え・解説

正答：イ

ア：人数が異なるため全体平均にはならない。
イ：クラス人数が異なるため単純に75点とせず人数加重する。
ウ：午前クラスの得点も含める必要がある。
エ：平均の和は全体平均ではない。
総合解説：群平均から全体平均を求めるときは各群の標本数で加重する。

対象：2026年度 / 基準日 2026-09-08

0032

1-2 代表値・散布度 > 平均・中央値・最頻値 | 難易度：基礎

6個の測定値を小さい順に並べると、3番目が8、4番目が12だった。この標本の中央値はどれか。

- ア . 8。小さい方の中央値だけを採る。
- イ . 12。大きい方の中央値だけを採る。
- ウ . 11。最小値3と最大値21の平均を使う。
- エ . 10。中央2値8と12の平均である。

答え・解説

正答：エ

ア：偶数個では2値の平均を使う。
イ：同様に2値の平均が必要である。
ウ：これはミッドレンジであり中央値ではない。
エ：偶数個では並べた中央2値の平均を中央値とする。
総合解説：中央値は順序統計量で、外れ値に平均ほど強く影響されない。

対象：2026年度 / 基準日 2026-09-08

0033

1-2 代表値・散布度 > 平均・中央値・最頻値 | 難易度：基礎

データ 1, 1, 2, 3, 3, 4 について最頻値の説明として適切なのはどれか。

- ア．2だけが最頻値である。
- イ．平均値と同じ値だけを最頻値と呼ぶ。
- ウ．1と3の2つが最頻値である。
- エ．最頻値は必ず1つに決まる。

答え・解説

正答：ウ

ア：2は1回しか出現していない。
イ：最頻値は出現頻度で定義される。
ウ：1と3はいずれも2回出現し、最大頻度が同じである。
エ：複数の値が同じ最大頻度を持つ場合がある。
総合解説：最頻値は最も頻繁に現れる値であり、複数存在することもある。

対象：2026年度 / 基準日 2026-09-08

0034

1-2 代表値・散布度 > 平均・中央値・最頻値 | 難易度：基礎

5世帯の月収が 25, 28, 30, 32, 300（万円）だった。典型的な世帯の中心を表す際、平均より中央値を併記する利点はどれか。

- ア．300万円の極端値の影響を受けにくく、中央的な位置を示せる。
- イ．中央値は必ず平均より大きくなる。
- ウ．中央値はすべての値を足して求めるため外れ値に敏感である。
- エ．中央値を使うと標本数が減る。

答え・解説

正答：ア

ア：中央値は値の順位に基づくため極端な大値の影響が比較的小さい。
イ：大小関係は分布により異なる。
ウ：中央値は合計を用いない。
エ：標本数は変わらない。
総合解説：歪みや外れ値がある場合、平均と中央値の違いが分布の特徴を示す。

対象：2026年度 / 基準日 2026-09-08

0035

1-2 代表値・散布度 > 平均・中央値・最頻値 | 難易度：標準

全観測値を「新しい値=2×元の値+5」で変換する。元の平均が12なら、新しい平均はどれか。

- ア . 24。定数5の加算を無視する。
- イ . 17。倍率2を無視する。
- ウ . 34。5も2倍して加える。
- エ . 29。平均にも同じ線形変換を適用できる。

答え・解説

正答：エ

ア：加算も平均に反映される。
イ：2倍の効果を含める必要がある。
ウ：変換式は $2x+5$ であり $2(x+5)$ ではない。
エ：平均は線形性を持つので $2 \times 12+5=29$ である。
総合解説：全観測に $ax+b$ を施すと平均も $a \times \text{平均}+b$ に変換される。

対象：2026年度 / 基準日 2026-09-08

0036

1-2 代表値・散布度 > 平均・中央値・最頻値 | 難易度：標準

5人の測定値が 10, 12, 欠測, 14, 16 だった。欠測を0とみなす根拠がないとき、観測済み4値の平均はどれか。

- ア . 10.4。欠測を0として5で割る。
- イ . 13。観測済みの合計52を4で割る。
- ウ . 52。合計をそのまま平均とする。
- エ . 14。最大値と最小値の平均を使う。

答え・解説

正答：イ

ア：欠測を0とする根拠がない。
イ：欠測とゼロは意味が異なるため、根拠なく0代入しない。
ウ：平均では観測数で割る。
エ：それは観測済み値の平均ではない。
総合解説：欠測処理は分析上の仮定を伴う。欠測を観測値0と混同しない。

対象：2026年度 / 基準日 2026-09-08

0037

1-2 代表値・散布度 > 平均・中央値・最頻値 | 難易度：標準

8個の観測値の平均が15である。観測値の合計として正しいのはどれか。

- ア . 120。平均×個数=15×8である。
- イ . 23。平均と個数を足す。
- ウ . 1.875。平均を個数で割る。
- エ . 105。平均から個数を引く。

答え・解説

正答：ア

ア：算術平均=合計/個数なので合計は平均×個数で求める。
イ：平均の定義に合わない。
ウ：合計の逆算ではない。
エ：平均の定義と無関係である。
総合解説：平均の定義から合計や欠けた値を逆算できる。

対象：2026年度 / 基準日 2026-09-08

0038

1-2 代表値・散布度 > 平均・中央値・最頻値 | 難易度：標準

5個の値の平均が20で、4個が 12, 18, 22, 24 と分かっている。残り1個はどれか。

- ア . 20。平均値をそのまま欠測値とする。
- イ . 4。100を既知4値で割る。
- ウ . 24。全体合計100から既知の合計76を引く。
- エ . 76。既知値の合計を答える。

答え・解説

正答：ウ

ア：既知値の合計を考慮する必要がある。
イ：求めるのは残差である。
ウ： $5 \times 20 = 100$ 、 $12 + 18 + 22 + 24 = 76$ なので残りは24である。
エ：残り1値ではない。
総合解説：全体平均と標本数から全体合計を求め、既知合計との差を使う。

対象：2026年度 / 基準日 2026-09-08

0039

1-2 代表値・散布度 > 平均・中央値・最頻値 | 難易度：標準

ある所得データで平均が620万円、中央値が430万円だった。これだけから断定はできないが、示唆として最も自然なのはどれか。

- ア．一部の大きな所得が平均を押し上げる右裾の長い分布の可能性がある。
- イ．必ず左右完全対称である。
- ウ．全員の所得が620万円である。
- エ．平均が中央値より大きいと標本数は必ず偶数である。

答え・解説

正答：ア

ア：右に歪んだ分布では大きな値が平均を上方へ引きやすい。
イ：平均と中央値の差はむしろ非対称性の手掛かりになる。
ウ：中央値430万円と両立しない。
エ：標本数の偶奇とは直接関係しない。
総合解説：代表値の位置関係は分布形状を考える補助情報になるが、それだけで形状を確定はできない。

対象：2026年度 / 基準日 2026-09-08

0040

1-2 代表値・散布度 > 平均・中央値・最頻値 | 難易度：標準

極端な測定誤差の影響を抑えつつ平均的水準を見たいので、両端5%ずつを除いて平均する方法を検討した。この要約の考え方として適切なものはどれか。

- ア．トリム平均は極端値の影響を弱めるための代表値の一つである。
- イ．トリム平均は常に中央値と完全一致する。
- ウ．トリム平均はデータを大きい順に並べるだけで値を計算しない。
- エ．トリム平均を使うと分散が必ず0になる。

答え・解説

正答：ア

ア：端の観測を一定割合除外して平均することで外れ値への感度を下げる。
イ：一般には一致しない。
ウ：除外後に平均を計算する。
エ：散らばりが消えるわけではない。
総合解説：中心の指標には平均・中央値・最頻値などがあり、目的や分布形状に応じて選ぶ。

対象：2026年度 / 基準日 2026-09-08

0041

1-2 代表値・散布度 > 平均・中央値・最頻値 | 難易度：応用

データの算術平均を $\bar{x}$ とする。各観測値から平均を引いた偏差の総和 $\sum (x_i - \bar{x})$ について正しいのはどれか。

- ア．0になる。
- イ．必ず標本数nになる。
- ウ．必ず標準偏差に等しい。
- エ．常に正になる。

答え・解説

正答：ア

ア：算術平均の定義から、偏差の正負が総和で相殺される。
イ：偏差和はnではない。
ウ：標準偏差は二乗偏差を基にする。
エ：正負の偏差があり総和は0となる。
総合解説：偏差和が0になることは平均の基本的性質である。

対象：2026年度 / 基準日 2026-09-08

0042

1-2 代表値・散布度 > 平均・中央値・最頻値 | 難易度：応用

靴のサイズ別販売数から「最も多く売れたサイズ」を発注の参考にしたい。最も直接的な代表値はどれか。

- ア．最頻値。販売数が最大のサイズを示す。
- イ．算術平均だけ。
- ウ．範囲。
- エ．標準偏差。

答え・解説

正答：ア

ア：カテゴリや離散値で最も頻度の高い値を知りたい目的には最頻値が直接対応する。
イ：平均サイズは実際に存在しない値になる場合があり「最多」とは別概念である。
ウ：範囲は散らばりを示す。
エ：標準偏差は散布度であり最多カテゴリを示さない。
総合解説：代表値は分析目的に応じて使い分ける。

対象：2026年度 / 基準日 2026-09-08

0043

1-2 代表値・散布度 > 分散・標準偏差・範囲・四分位範囲 | 難易度：基礎

ある日の気温観測で最低7、最高22 だった。この日の気温の範囲（最大値 - 最小値）はどれか。

- ア . 29。最大値と最小値を足す。
- イ . 7。最小値だけを使う。
- ウ . 22。最大値だけを使う。
- エ . 15。最大22から最小7を引く。

答え・解説

正答：エ

ア：範囲は差である。
イ：範囲では最大値も必要である。
ウ：範囲は最大値そのものではない。
エ：範囲=max - minである。
総合解説：範囲は計算が簡単だが、最小値・最大値だけに依存し外れ値の影響を受けやすい。

対象：2026年度 / 基準日 2026-09-08

0044

1-2 代表値・散布度 > 分散・標準偏差・範囲・四分位範囲 | 難易度：基礎

ある箱ひげ図で箱の下端が24、上端が38だった。箱の長さに対応する四分位範囲IQRはいくつか。

- ア . 62。Q1とQ3を足す。
- イ . 7。差をさらに2で割る。
- ウ . 14。38 - 24である。
- エ . 24。Q1そのもの。

答え・解説

正答：ウ

ア：IQRは差である。
イ：IQRの定義では2で割らない。
ウ：IQR=Q3 - Q1で、中央50%の広がりを示す。
エ：Q1は位置指標でIQRではない。
総合解説：IQRは中央半分の散らばりを表し、極端値に比較的頑健である。

対象：2026年度 / 基準日 2026-09-08

0045

1-2 代表値・散布度 > 分散・標準偏差・範囲・四分位範囲 | 難易度：基礎

母集団データ 2, 4, 6 の母平均は4である。母分散（分母3）として正しいのはどれか。

- ア . 8/3。二乗偏差4,0,4の合計8を3で割る。
- イ . 4。偏差の絶対値平均を用いる。
- ウ . 8。二乗偏差和を割らない。
- エ . 0。偏差の和が0だから分散も0とする。

答え・解説

正答：ア

ア：母分散は母集団全体の二乗偏差平均である。
イ：分散は二乗偏差を使う。
ウ：母分散は観測数で割る。
エ：偏差を二乗するため0にはならない。
総合解説：分散では正負の偏差を相殺させず二乗して平均する。

対象：2026年度 / 基準日 2026-09-08

0046

1-2 代表値・散布度 > 分散・標準偏差・範囲・四分位範囲 | 難易度：基礎

分散が49のデータの標準偏差として正しいのはどれか。

- ア . 49。分散をそのまま使う。
- イ . 2401。分散を二乗する。
- ウ . 24.5。分散を2で割る。
- エ . 7。分散の正の平方根である。

答え・解説

正答：エ

ア：標準偏差は平方根を取る。
イ：逆方向の操作である。
ウ：標準偏差の定義ではない。
エ：標準偏差= $\sqrt{\text{分散}}$ である。
総合解説：標準偏差は元データと同じ単位を持つため散らばりを直感的に解釈しやすい。

対象：2026年度 / 基準日 2026-09-08

0047

1-2 代表値・散布度 > 分散・標準偏差・範囲・四分位範囲 | 難易度：標準

身長をcmで測定したとき、分散と標準偏差の単位の説明として適切なのはどれか。

- ア．分散も標準偏差も必ず無単位である。
- イ．分散はcm、標準偏差はcm²である。
- ウ．分散はcm²、標準偏差はcmである。
- エ．どちらもcm³である。

答え・解説

正答：ウ

ア：元データの単位に依存する。
イ：逆である。
ウ：分散は偏差を二乗するため単位も二乗され、平方根を取った標準偏差は元単位に戻る。
エ：三乗は用いない。
総合解説：分散と標準偏差では単位が異なる点に注意する。

対象：2026年度 / 基準日 2026-09-08

0048

1-2 代表値・散布度 > 分散・標準偏差・範囲・四分位範囲 | 難易度：標準

すべての観測値に100を加えた。分散と標準偏差はどうなるか。

- ア．分散は100増え、標準偏差も100増える。
- イ．分散だけ0になる。
- ウ．どちらも変わらない。
- エ．標準偏差だけ100倍になる。

答え・解説

正答：ウ

ア：平行移動は散らばりを変えない。
イ：偏差は保持される。
ウ：全値への同一定数加算は偏差を変えないため散布度は不変である。
エ：定数加算は倍率変換ではない。
総合解説：位置の移動と散らばりの変化を区別する。

対象：2026年度 / 基準日 2026-09-08

0049 1-2 代表値・散布度 > 分散・標準偏差・範囲・四分位範囲 | 難易度：標準

全観測値を3倍した。元の標準偏差が4なら、新しい標準偏差と分散の倍率はどれか。

ア．標準偏差は7になり、分散は3倍になる。

イ．標準偏差は12になり、分散は元の9倍になる。

ウ．標準偏差も分散も3倍になる。

エ．標準偏差は4のままで、分散だけ9倍になる。

答え・解説 正答：イ

ア：加算のように扱っている。

イ：a倍すると標準偏差は|a|倍、分散は a^2 倍になる。

ウ：分散は倍率の二乗で変化する。

エ：倍率変換では標準偏差も変わる。

総合解説：線形変換 $ax+b$ ではbは散布度に影響せず、aが標準偏差・分散を拡大縮小する。

対象：2026年度 / 基準日 2026-09-08

0050 1-2 代表値・散布度 > 分散・標準偏差・範囲・四分位範囲 | 難易度：標準

標本から母分散を推定する際の不偏分散で、偏差平方和を通常 $n - 1$ で割る主な理由として適切なのはどれか。

ア． $n - 1$ で割ると必ず分散が0になるから。

イ．標本平均を推定に使ったことで自由度を1失い、母分散に対する不偏性を得るためである。

ウ．標本数を1件削除したから。

エ．分散の単位を元の単位に戻すため。

答え・解説 正答：イ

ア：むしろnより小さい分母なので正の二乗偏差和なら大きくなる。

イ：標本平均からの偏差には1本の制約があり、 $n - 1$ で割る補正が用いられる。

ウ：観測自体を削除するわけではない。

エ：単位は分母補正では戻らない。

総合解説：母分散と標本分散の定義・目的を区別する。

対象：2026年度 / 基準日 2026-09-08

0051 1-2 代表値・散布度 > 分散・標準偏差・範囲・四分位範囲 | 難易度：標準

データに非常に大きな外れ値が1つ追加された。範囲・標準偏差・IQRについて一般的に最も適切なのはどれか。

ア．IQRだけが必ず無限大になる。
イ．3指標とも外れ値の影響を全く受けない。
ウ．範囲と標準偏差は大きく影響を受け得るが、IQRは比較的影響を受けにくい。
エ．範囲だけが必ず小さくなる。

答え・解説 正答：ウ

ア：IQRは極端値に比較的頑健である。
イ：範囲や標準偏差は影響を受ける。
ウ：IQRは中央50%に基づくため極端値への感度が低い。
エ：大きな外れ値なら範囲は通常広がる。
総合解説：散布度の指標ごとに外れ値への感度が異なる。

対象：2026年度 / 基準日 2026-09-08

0052 1-2 代表値・散布度 > 分散・標準偏差・範囲・四分位範囲 | 難易度：標準

同じ標本数の2群A,Bで、各群内標準偏差は同じだが群平均が大きく離れている。2群をまとめた全体の分散について最も適切なのはどれか。

ア．群内標準偏差が同じなら全体分散は必ず同じ値になる。
イ．群内のばらつきに加えて群平均間の違いも全体のばらつきへ寄与する。
ウ．群平均の差は分散に一切影響しない。
エ．2群をまとめると分散は必ず0になる。

答え・解説 正答：イ

ア：群平均間の差も影響する。
イ：全体の二乗偏差には群内変動と群間の位置差の両方が含まれる。
ウ：全体平均からの偏差に反映される。
エ：異なる値があれば0にはならない。
総合解説：データを統合した分散には群内散布と群間差の両方が影響する。

対象：2026年度 / 基準日 2026-09-08

0053

1-2 代表値・散布度 > 分散・標準偏差・範囲・四分位範囲 | 難易度：応用

右裾が長く外れ値もある診療費データの「典型的な散らばり」を要約したい。最も妥当な組合せはどれか。

- ア．最大値と範囲だけを示せば中心も十分分かる。
- イ．標準偏差は外れ値に影響されないので、それだけで十分である。
- ウ．散らばりを示す指標は不要で、平均だけでよい。
- エ．中央値とIQRを中心に示し、必要に応じ平均・標準偏差も補助的に示す。

答え・解説

正答：エ

ア：中心情報が不足する。
イ：標準偏差は外れ値の影響を受ける。
ウ：同じ平均でも分布の広がりは異なり得る。
エ：歪みと外れ値がある場合、中央値・IQRは頑健な要約となる。
総合解説：中心と散らばりは分布形状に応じて組み合わせて報告する。

対象：2026年度 / 基準日 2026-09-08

0054

1-2 代表値・散布度 > 分散・標準偏差・範囲・四分位範囲 | 難易度：応用

あるデータ集合の分散が0だった。このとき必ず成り立つことはどれか。

- ア．すべての観測値が同じ値である。
- イ．平均が必ず0である。
- ウ．標本数が1に限られる。
- エ．最大値と最小値が異なる。

答え・解説

正答：ア

ア：二乗偏差の平均が0になるには各偏差が0でなければならない。
イ：共通値が0である必要はない。
ウ：複数観測が同値でも分散0になる。
エ：分散0なら最大値=最小値である。
総合解説：分散0は観測値にばらつきがないことを意味する。

対象：2026年度 / 基準日 2026-09-08

0055

1-2 代表値・散布度 > 箱ひげ図・視覚的比較 | 難易度：基礎

箱ひげ図を構成する五数要約のうち、箱の内部の線として通常描かれる中心位置はどれか。

- ア．中央値（第2四分位数）。
- イ．必ず平均値。
- ウ．標準偏差。
- エ．標本数。

答え・解説

正答：ア

ア：標準的な箱ひげ図では箱内の線が中央値を示す。
イ：平均を追加表示する図もあるが標準要素ではない。
ウ：標準偏差を線で示すものではない。
エ：標本数は位置線では表さない。
総合解説：箱ひげ図は四分位数を中心に分布の位置と広がりを要約する。

対象：2026年度 / 基準日 2026-09-08

0056

1-2 代表値・散布度 > 箱ひげ図・視覚的比較 | 難易度：基礎

Q1=20、Q3=32の箱ひげ図で1.5IQR則を用いる。上側フェンスはいくつか。

- ア．44。Q3+IQRとする。
- イ．38。Q3+0.5IQRとする。
- ウ．18。Q3 - 1.5IQRとする。
- エ．50。IQR=12なので32+1.5 × 12=50。

答え・解説

正答：エ

ア：1.5倍が必要である。
イ：係数が違う。
ウ：これは上側ではなく別方向の計算である。
エ：上側フェンス=Q3+1.5IQRである。
総合解説：箱ひげ図の外れ値ルールはデータエラーを断定する規則ではなく、外れ値候補を識別する目安である。

対象：2026年度 / 基準日 2026-09-08

0057

1-2 代表値・散布度 > 箱ひげ図・視覚的比較 | 難易度：標準

同じ尺度で描いた2群の箱ひげ図で、群Bの中央値線が群Aより明らかに高い位置にある。直接言えることはどれか。

- ア．群Bの平均が必ず群Aより大きい。
- イ．群Bの標本数が必ず多い。
- ウ．群Bの中央値が群Aより大きい。
- エ．群Bの分散が必ず0である。

答え・解説

正答：ウ

ア：箱ひげ図の中央値だけから平均の大小は必ずしも分らない。
イ：標準箱ひげ図の位置から標本数は決まらない。
ウ：共通軸上の中央値線の位置から中央値の大小を比較できる。
エ：中央値の位置と分散0は無関係である。
総合解説：箱ひげ図で何を直接読み取れるかと、推測に過ぎないことを区別する。

対象：2026年度 / 基準日 2026-09-08

0058

1-2 代表値・散布度 > 箱ひげ図・視覚的比較 | 難易度：標準

共通の縦軸で描いた縦型箱ひげ図で、群Aの箱の高さが群Bより大きい。最も直接的な解釈はどれか。

- ア．群Aの平均が群Bより大きい。
- イ．群Aの範囲が必ず群Bより大きい。
- ウ．群AのIQRが群Bより大きい。
- エ．群Aの標本数が必ず多い。

答え・解説

正答：ウ

ア：箱の高さは平均ではない。
イ：ひげや外れ値を含む全範囲とは別である。
ウ：箱の長さはQ3 - Q1、すなわちIQRに対応する。
エ：通常の箱の長さは標本数を表さない。
総合解説：箱の長さは中央50%の散らばりの比較に使える。

対象：2026年度 / 基準日 2026-09-08

0059

1-2 代表値・散布度 > 箱ひげ図・視覚的比較 | 難易度：標準

ある箱ひげ図で上側のひげが下側よりかなり長く、中央値も箱の下側に寄っている。示唆として自然なのはどれか。

- ア．必ず完全対称な正規分布である。
- イ．大きい値側に長い尾をもつ右に歪んだ分布の可能性がある。
- ウ．左側にのみ外れ値が存在することが確定する。
- エ．標本数が2である。

答え・解説

正答：イ

ア：描写は非対称である。
イ：箱やひげの非対称性は歪みの手掛かりになる。
ウ：ひげ形状だけで外れ値位置を確定できない。
エ：箱ひげ図の形から標本数は決まらない。
総合解説：箱ひげ図は分布の非対称性を概観できるが、詳細形状はヒストグラム等も併用するとよい。

対象：2026年度 / 基準日 2026-09-08

0060

1-2 代表値・散布度 > 箱ひげ図・視覚的比較 | 難易度：標準

1.5IQR型箱ひげ図で、ひげの外側に点が描かれている。この点について適切な説明はどれか。

- ア．統計的ルールで外れ値候補とされた観測であり、誤データと断定はできない。
- イ．必ず入力ミスなので削除する。
- ウ．必ず平均値を表す。
- エ．ひげの外側の点は観測値ではない。

答え・解説

正答：ア

ア：外れ値フラグは要確認を意味し、実在する極端値である可能性もある。
イ：検証なしの削除は不適切である。
ウ：外側の点は平均の標準表示ではない。
エ：通常は実際の観測値を示す。
総合解説：外れ値候補は原データや測定過程を確認して扱う。

対象：2026年度 / 基準日 2026-09-08

0061

1-2 代表値・散布度 > 箱ひげ図・視覚的比較 | 難易度：応用

2つのデータ集合が全く同じ五数要約を持つとする。箱ひげ図について最も適切なのはどれか。

- ア．箱ひげ図は同じ形になり得るが、ヒストグラムなどで見る詳細な分布形状は異なる場合がある。
- イ．五数要約が同じなら元データも必ず完全一致する。
- ウ．箱ひげ図だけで各観測値をすべて復元できる。
- エ．五数要約が同じなら標本数も必ず同じである。

答え・解説

正答：ア

ア：箱ひげ図は情報を五数要約へ圧縮するため、山の数などは失われる。
イ：異なるデータが同じ五数要約を持ち得る。
ウ：個々の値は保持されない。
エ：標本数が異なる場合もあり得る。
総合解説：箱ひげ図は比較に有用だが、分布の詳細をすべて表現するものではない。

対象：2026年度 / 基準日 2026-09-08

0062

1-2 代表値・散布度 > 箱ひげ図・視覚的比較 | 難易度：応用

店舗AとBの待ち時間の箱ひげ図を比較する。Aは0～30分、Bは0～120分の別軸で描かれている。最も適切な改善はどれか。

- ア．各軸の目盛を消して箱の長さだけを見る。
- イ．待ち時間を店舗名の名義尺度に変換する。
- ウ．中央値線だけを削除する。
- エ．同じ時間尺度・共通軸にそろえて並べ、位置とIQRを比較する。

答え・解説

正答：エ

ア：さらに誤解しやすくなる。
イ：量的情報が失われる。
ウ：比較に有用な情報を失う。
エ：軸範囲が違っていると見かけの長さが比較できないため共通尺度が望ましい。
総合解説：複数群の視覚比較では同じ単位・尺度を使う。

対象：2026年度 / 基準日 2026-09-08

0063

1-3 標準化・分布特性 > 偏差・標準化（z得点） | 難易度：基礎

平均72、標準偏差8の試験で84点を取った。z得点として正しいのはどれか。

- ア . 10.5。84 ÷ 8とする。
- イ . 12。84 - 72だけを使う。
- ウ . 1.5。(84 - 72)/8=1.5である。
- エ . 0.67。8 ÷ 12とする。

答え・解説

正答：ウ

ア：平均を引く必要がある。
イ：標準偏差で割る必要がある。
ウ：z得点は観測値が平均から標準偏差何個分離れているかを表す。
エ：比の向きが逆である。
総合解説： $z = (x - \text{平均}) / \text{標準偏差}$ で標準化する。

対象：2026年度 / 基準日 2026-09-08

0064

1-3 標準化・分布特性 > 偏差・標準化（z得点） | 難易度：基礎

ある観測のz得点が - 1.8だった。最も適切な解釈はどれか。

- ア . 平均より1.8標準偏差だけ高い。
- イ . 観測値そのものが - 1.8である。
- ウ . 平均より1.8標準偏差だけ低い位置にある。
- エ . 標準偏差が - 1.8である。

答え・解説

正答：ウ

ア：符号の解釈が逆である。
イ：zは無次元の標準化値である。
ウ：zの符号は平均の上下、絶対値は標準偏差単位の距離を示す。
エ：標準偏差は非負である。
総合解説：標準化により異なる尺度の位置を標準偏差単位で比較できる。

対象：2026年度 / 基準日 2026-09-08

0065

1-3 標準化・分布特性 > 偏差・標準化（z得点） | 難易度：基礎

A試験は平均50・標準偏差10で65点、B試験は平均80・標準偏差4で86点だった。相対的位置の比較として適切なのはどれか。

- ア．Aの方が65<86なので相対的位置も必ず低い。
- イ．Bだけ $z=6$ である。
- ウ．両者とも $z=1.5$ で、平均からの相対的位置は同じである。
- エ．Aだけ $z=15$ である。

答え・解説

正答：ウ

ア：尺度が異なる得点の生値比較は適切でない。
イ：偏差6を標準偏差4で割ると1.5である。
ウ：Aは $(65 - 50) / 10 = 1.5$ 、Bは $(86 - 80) / 4 = 1.5$ である。
エ：偏差を標準偏差で割る必要がある。
総合解説：z得点は平均と標準偏差の異なる尺度間で相対的位置を比較する手段となる。

対象：2026年度 / 基準日 2026-09-08

0066

1-3 標準化・分布特性 > 偏差・標準化（z得点） | 難易度：標準

各観測値を同じデータ集合の平均と標準偏差でz標準化した。標準化後の中心と尺度について正しいのはどれか。

- ア．平均も標準偏差も元の値のままである。
- イ．平均は0になり、用いた標準偏差の定義に対応して標準偏差は1になる。
- ウ．平均は1、標準偏差は0になる。
- エ．すべてのz得点が0になる。

答え・解説

正答：イ

ア：標準化で位置と尺度が変わる。
イ：各偏差を標準偏差で割るため中心0、尺度1へ変換される。
ウ：逆である。
エ：平均と等しい観測だけが0になる。
総合解説：z標準化は中心と尺度をそろえるが分布形状の情報は残る。

対象：2026年度 / 基準日 2026-09-08

0067

1-3 標準化・分布特性 > 偏差・標準化（z得点） | 難易度：標準

標準偏差が正のデータをz標準化した。元データの大小順位について正しいのはどれか。

- ア．順位は必ず逆転する。
- イ．順位は保たれる。
- ウ．全観測が同順位になる。
- エ．順位は標本数が偶数のときだけ保たれる。

答え・解説

正答：イ

ア：正の係数の変換なので逆転しない。
イ： $z=(x - \text{平均})/\text{正の標準偏差}$ は単調増加の線形変換である。
ウ：値の差は標準偏差単位で残る。
エ：標本数の偶奇は無関係である。
総合解説：正の線形変換は順序を保持する。

対象：2026年度 / 基準日 2026-09-08

0068

1-3 標準化・分布特性 > 偏差・標準化（z得点） | 難易度：標準

平均100、標準偏差15の変数で $z = -0.8$ の観測値 x はいくつか。

- ア．112。符号を逆にして加える。
- イ．88。 $x=100+(-0.8)\times 15=88$ である。
- ウ．92。標準偏差に0.8を足し引きする。
- エ．-12。 $z\times$ 標準偏差だけを答える。

答え・解説

正答：イ

ア： z は負なので平均より下である。
イ： $z=(x - \mu)/\sigma$ を $x=\mu+z\sigma$ に変形する。
ウ： z は標準偏差の倍数で使う。
エ：これは平均からの偏差であり観測値ではない。
総合解説：標準化と逆標準化を相互に扱えるようにする。

対象：2026年度 / 基準日 2026-09-08

0069

1-3 標準化・分布特性 > 偏差・標準化（z得点） | 難易度：標準

入力データの1件だけz得点が9.4で、他は - 2 ~ 2程度だった。最初の対応として適切なのはどれか。

- ア．zが3を超えたら無条件に削除する。
- イ．9.4を自動的に2へ置換する。
- ウ．z得点は外れ値確認には一切使えない。
- エ．入力・単位・測定条件を確認し、実値ならそのまま分析上の影響も検討する。

答え・解説

正答：エ

- ア：機械的削除は不適切である。
 - イ：恣意的な改変になる。
 - ウ：標準偏差単位で極端さを見る手掛かりになる。
 - エ：非常に大きなzは異常候補を示すが、誤りとは限らない。
- 総合解説：標準化値は異常候補の探索に利用できるが、処理にはデータ生成過程の確認が必要である。

対象：2026年度 / 基準日 2026-09-08

0070

1-3 標準化・分布特性 > 偏差・標準化（z得点） | 難易度：標準

全観測値が同じで標準偏差が0だった。このデータを通常の式でz標準化しようとするとうなるか。

- ア．すべて自動的にz=1になる。
- イ．すべてz=0と定義されるのが数式上必然である。
- ウ．0で割ることになるため、通常のz得点は定義できない。
- エ．標準偏差を - 1と置けばよい。

答え・解説

正答：ウ

- ア：0除算を回避できない。
 - イ：式は0/0となり通常は未定義である。
 - ウ：z得点の分母には正の標準偏差が必要である。
 - エ：標準偏差を恣意的に変更できない。
- 総合解説：標準化の適用条件も確認する。

対象：2026年度 / 基準日 2026-09-08

0071

1-3 標準化・分布特性 > 偏差・標準化（z得点） | 難易度：応用

身長をcmからmへ変換してからz得点を計算した。cmで計算したz得点と比べてどうなるか。

- ア．同じになる。平均と標準偏差も同じ正の倍率で変わるため比が相殺される。
- イ．100倍になる。
- ウ．1/100になる。
- エ．符号が必ず逆になる。

答え・解説

正答：ア

ア：正の線形尺度変換ではz得点是不変である。
イ：分子・分母の双方が1/100になるため相殺される。
ウ：同様に相殺される。
エ：正の倍率変換なので符号は変わらない。
総合解説：z得点は単位に依存しない無次元量である。

対象：2026年度 / 基準日 2026-09-08

0072

1-3 標準化・分布特性 > 偏差・標準化（z得点） | 難易度：応用

2つのデータ集合をそれぞれz標準化したところ、平均0・標準偏差1になった。ここから確実に言えないことはどれか。

- ア．両方の標準化後平均が0であること。
- イ．2つの分布形状が完全に同一であること。
- ウ．両方の標準化後尺度が1であること。
- エ．z得点が無次元であること。

答え・解説

正答：イ

ア：標準化の基本性質である。
イ：標準化は位置と尺度をそろえるが、歪みや多峰性などの形状は同一にしない。
ウ：同じ定義の標準偏差を使えば基本性質である。
エ：単位は分子・分母で相殺される。
総合解説：標準化と正規化・正規分布化を混同しない。

対象：2026年度 / 基準日 2026-09-08

0073

1-3 標準化・分布特性 > 変動係数・指数化 | 難易度：基礎

平均40、標準偏差6の正の比率尺度データについて、変動係数CVを百分率で表すとどれか。

- ア . 6.7%。40/6を百分率とする。
- イ . 34%。40 - 6を百分率とする。
- ウ . 15%。6/40 × 100である。
- エ . 240%。40 × 6を百分率とする。

答え・解説

正答：ウ

ア：比が逆である。
イ：差ではない。
ウ：CVは標準偏差を平均で割った相対的散布度である。
エ：積ではない。
総合解説：CVは平均水準に対する標準偏差の割合を示す。

対象：2026年度 / 基準日 2026-09-08

0074

1-3 標準化・分布特性 > 変動係数・指数化 | 難易度：基礎

機械Aの平均寿命100時間・標準偏差5時間、機械Bの平均寿命20日・標準偏差2日である。相対的なばらつきをCVで比較するとどちらが大きいか。

- ア . A。平均100が20より大きいから。
- イ . 両者同じ。標準偏差がどちらも一桁だから。
- ウ . 比較不能。CVは同一単位のデータにしか使えない。
- エ . B。Aは5%、Bは10%である。

答え・解説

正答：エ

ア：単位が異なる平均の生値は比較できず、CVではA=5%である。
イ：A=5%、B=10%で異なる。
ウ：正の比率尺度で意味がある場合、CVは単位に依存しない相対比較に使える。
エ：同じ各データ内の単位で標準偏差/平均を取ればCVは無次元となり、Bの相対散布が大きい。
総合解説：CVは平均に対する散らばりを比較する指標で、絶対的な標準偏差だけでは比較しにくい尺度間で有用である。

対象：2026年度 / 基準日 2026-09-08

0075

1-3 標準化・分布特性 > 変動係数・指数化 | 難易度：基礎

平均が0.02、標準偏差が1.5の測定値でCVを主要指標にしようとしている。注意点として最も適切なのはどれか。

- ア．平均が小さいほどCVは必ず0に近づく。
- イ．CVは標準偏差を使わないので問題ない。
- ウ．平均が0付近なら分散も必ず0である。
- エ．平均が0に近いとCVが非常に大きく不安定になり、相対散布の解釈が適切でないことがある。

答え・解説

正答：エ

ア：逆に大きくなりやすい。
イ：標準偏差/平均で定義される。
ウ：中心位置と散らばりは別である。
エ：CVは平均を分母にするため、平均0付近では解釈が不安定になる。
総合解説：CVは正の比率尺度で平均が0から十分離れている場合に使いやすい。

対象：2026年度 / 基準日 2026-09-08

0076

1-3 標準化・分布特性 > 変動係数・指数化 | 難易度：標準

売上を基準年=100として指数化する。基準年売上500、今年売上650なら指数はいくつか。

- ア．150。650 - 500だけを使う。
- イ．130。650/500×100である。
- ウ．76.9。500/650×100とする。
- エ．115。500と650の平均を100で割る。

答え・解説

正答：イ

ア：差ではなく比で指数化する。
イ：基準比を100倍して指数化する。
ウ：比の向きが逆である。
エ：指数の定義ではない。
総合解説：指数化により異なる時点の相対水準を基準値100で表せる。

対象：2026年度 / 基準日 2026-09-08

0077

1-3 標準化・分布特性 > 変動係数・指数化 | 難易度：標準

基準年=100の生産指数が92になった。基準年との比較として適切なのはどれか。

- ア．基準年より92%低い。
- イ．基準年より8倍高い。
- ウ．基準年より指数上8%低い水準である。
- エ．8単位の絶対量が減ったと必ず言える。

答え・解説

正答：ウ

ア：指数92は残存水準92%を表す。
イ：倍率解釈が誤っている。
ウ： $92/100=0.92$ なので基準比で8%低い。
エ：指数は相対水準で、元の単位量の差とは限らない。
総合解説：指数値と変化率を区別する。

対象：2026年度 / 基準日 2026-09-08

0078

1-3 標準化・分布特性 > 変動係数・指数化 | 難易度：標準

同じ時系列指数の基準年を2015年=100から2020年=100へ変更した。最も適切な説明はどれか。

- ア．基準年変更で元の実データそのものが変わる。
- イ．新基準では過去の順位が必ず逆転する。
- ウ．各指数値は変わるが、同じ2時点間の比率関係はリベースによって本質的には保たれる。
- エ．基準変更すると指数はすべて100になる。

答え・解説

正答：ウ

ア：元の観測量は変わらない。
イ：正の共通倍率なので時点の相対順位は保たれる。
ウ：基準変更は系列全体を共通倍率で再尺度化する操作である。
エ：100になるのは新しい基準時点である。
総合解説：指数の基準変更は表示尺度の変更であり、元データの関係を変えるものではない。

対象：2026年度 / 基準日 2026-09-08

0079

1-3 標準化・分布特性 > 変動係数・指数化 | 難易度：標準

商品Aは2010年=100、商品Bは2020年=100の指数で現在値がそれぞれ120と110だった。現在の
上昇率を直接比較する前に必要なことはどれか。

- ア．120>110なのでAの現在価格が必ず高い。
- イ．2つの指数を足して230とする。
- ウ．指数なら基準年は無関係である。
- エ．基準時点と指数の定義をそろえるか、各商品の対応する基準からの変化率として解釈する。

答え・解説

正答：エ

- ア：指数は価格水準の絶対額ではない。
 - イ：基準の異なる指数の和に意味はない。
 - ウ：指数の解釈には基準時点が重要である。
 - エ：基準が違う指数の生値120と110だけでは同じ期間の変化を比較できない。
- 総合解説：指数比較では基準、対象範囲、ウェイト等の定義を確認する。

対象：2026年度 / 基準日 2026-09-08

0080

1-3 標準化・分布特性 > 変動係数・指数化 | 難易度：標準

ある量が1年目に10%増え、2年目に10%減った。初期値100を指数化して追うと2年後はいくつか。

- ア．100。+10%と-10%を単純に相殺する。
- イ．99。100×1.10×0.90=99である。
- ウ．90。2年目の10%減だけを初期値に適用する。
- エ．110。1年目の増加だけを残す。

答え・解説

正答：イ

- ア：割合は異なる基準に掛かる。
 - イ：同率の増減でも基準となる水準が変わるため元に戻らない。
 - ウ：1年目の増加を反映していない。
 - エ：2年目の減少も反映する必要がある。
- 総合解説：連続する率変化は加算ではなく倍率の積で扱う。

対象：2026年度 / 基準日 2026-09-08

0081

1-3 標準化・分布特性 > 変動係数・指数化 | 難易度：応用

「店舗ごとの売上の相対的なばらつき」と「2019年を100とした各年売上水準」をそれぞれ示したい。最も適切な組合せはどれか。

- ア．前者はCV、後者は指数化を用いる。
- イ．前者は指数化、後者はCVだけを用いる。
- ウ．両方とも最頻値だけで表す。
- エ．両方とも範囲だけで表す。

答え・解説

正答：ア

ア：CVは相対散布、指数化は基準に対する相対水準を表す。
イ：目的が逆である。
ウ：目的に対応しない。
エ：時系列の基準比は範囲では表せない。
総合解説：似た「相対」という言葉でも、散布と水準では指標の目的が異なる。

対象：2026年度 / 基準日 2026-09-08

0082

1-3 標準化・分布特性 > 変動係数・指数化 | 難易度：応用

摂氏温度のCVを都市間比較に用いることが不適切になり得る理由として最も適切なのはどれか。

- ア．摂氏温度は絶対的ゼロを持つ比率尺度ではなく、原点を変えると平均に対する比が変わるから。
- イ．摂氏温度には平均が計算できないから。
- ウ．摂氏温度の標準偏差は必ず0だから。
- エ．CVはカテゴリデータ専用だから。

答え・解説

正答：ア

ア：CVは意味のあるゼロを持つ比率尺度で解釈しやすい。
イ：平均は計算できる。
ウ：温度が変動すれば標準偏差は正になり得る。
エ：CVは量的データの相対散布指標である。
総合解説：CVの意味は尺度水準に依存する。原点任意の間隔尺度では倍率解釈に注意する。

対象：2026年度 / 基準日 2026-09-08

0083

1-3 標準化・分布特性 > ローレンツ曲線・ジニ係数・歪度・尖度 | 難易度：基礎

所得が全世帯で完全に等しいとき、ローレンツ曲線はどこに位置するか。

- ア．横軸上に一致する。
- イ．原点と(1,1)を結ぶ45度の完全平等線上に一致する。
- ウ．縦軸上に一致する。
- エ．完全平等線より常に上側へ出る。

答え・解説

正答：イ

ア：累積所得比率が0のままではない。
イ：各累積世帯比率と累積所得比率が等しくなるためである。
ウ：累積世帯比率の増加を無視している。
エ：標準的な並べ方では完全平等が基準線となる。
総合解説：ローレンツ曲線は分配の集中度を累積比率で表す。

対象：2026年度 / 基準日 2026-09-08

0084

1-3 標準化・分布特性 > ローレンツ曲線・ジニ係数・歪度・尖度 | 難易度：基礎

ジニ係数が0に近い状態の解釈として適切なのはどれか。

- ア．ローレンツ曲線が完全平等線に近く、分配が比較的均等である。
- イ．1人にすべてが集中している状態に近い。
- ウ．平均が0であることを意味する。
- エ．標本数が0であることを意味する。

答え・解説

正答：ア

ア：ジニ係数0は完全平等に対応する。
イ：それはジニ係数が大きい側の極端例である。
ウ：ジニ係数は平均値そのものではない。
エ：標本数を表す指標ではない。
総合解説：ジニ係数は分配の不均等度を0に近いほど平等、1に近いほど集中という方向で表す。

対象：2026年度 / 基準日 2026-09-08

0085

1-3 標準化・分布特性 > ローレンツ曲線・ジニ係数・歪度・尖度 | 難易度：標準

同じ完全平等線に対し、曲線Aが曲線Bより全体として大きく下方へ膨らんでいる。一般的な解釈として適切なのはどれか。

- ア．Aの方が分配の不平等が大きいと考えられる。
- イ．Aの方が必ず平均所得が高い。
- ウ．Bの方が必ず標本数が多い。
- エ．両者のジニ係数は必ず0である。

答え・解説

正答：ア

ア：完全平等線からの乖離が大きいほど不平等度が高い方向を示す。
イ：ローレンツ曲線は平均所得水準を直接比較しない。
ウ：曲線形状から標本数は決まらない。
エ：完全平等線と一致していない。
総合解説：ローレンツ曲線は水準ではなく分配構造を比較する。

対象：2026年度 / 基準日 2026-09-08

0086

1-3 標準化・分布特性 > ローレンツ曲線・ジニ係数・歪度・尖度 | 難易度：標準

歪度が大きな正の値を示すデータの形状として典型的なのはどれか。

- ア．右側に長い裾をもつ分布。
- イ．左側に長い裾をもつ分布。
- ウ．必ず完全対称である。
- エ．すべての観測値が等しい。

答え・解説

正答：ア

ア：正の歪度は大きい値側の長い尾に対応する。
イ：負の歪度の方向である。
ウ：歪度が大きな正なら非対称性がある。
エ：その場合歪度の通常の解釈自体ができない。
総合解説：歪度は分布の非対称性の方向と程度を表す。

対象：2026年度 / 基準日 2026-09-08

0087

1-3 標準化・分布特性 > ローレンツ曲線・ジニ係数・歪度・尖度 | 難易度：標準

試験が易しすぎて多くの受験者が高得点に集中し、少数の低得点者が左へ尾を作った。歪度の符号として自然なのはどれか。

- ア．正。高得点が多いから必ず正である。
- イ．負。小さい値側に長い尾がある。
- ウ．必ず0。得点データだから。
- エ．歪度はカテゴリ数を表すので符号を持たない。

答え・解説

正答：イ

ア：符号は山の位置より尾の方向で考える。
イ：左裾が長い分布は負の歪みと呼ばれる。
ウ：非対称なら0とは限らない。
エ：歪度は数値的な形状指標である。
総合解説：歪度の符号は長い尾の方向で判断する。

対象：2026年度 / 基準日 2026-09-08

0088

1-3 標準化・分布特性 > ローレンツ曲線・ジニ係数・歪度・尖度 | 難易度：標準

あるソフトは正規分布の尖度を3、別のソフトは超過尖度を0とする。両者の関係として適切なのはどれか。

- ア．超過尖度は通常の尖度から3を引いた定義である。
- イ．超過尖度は通常の尖度を3倍した値である。
- ウ．尖度と歪度は同じ量である。
- エ．正規分布の尖度はどの定義でも必ず1である。

答え・解説

正答：ア

ア：正規分布を基準0にするため、 $kurtosis - 3$ をexcess kurtosisとする定義が使われる。
イ：定義が違う。
ウ：対称性と尾・尖りの性質で異なる指標である。
エ：代表的定義では3または超過尖度0である。
総合解説：ソフトウェア出力では尖度の定義が通常尖度か超過尖度か確認する。

対象：2026年度 / 基準日 2026-09-08

0089

1-3 標準化・分布特性 > ローレンツ曲線・ジニ係数・歪度・尖度 | 難易度：応用

地域AとBの平均所得は同じだが、Aのジニ係数がBより大きい。最も適切な解釈はどれか。

- ア．Aの平均所得が必ずBより高い。
- イ．ジニ係数が大きいほど標本数が多い。
- ウ．平均が同じならジニ係数も必ず同じ。
- エ．平均水準が同じでも、Aの所得分配はBより不均等である可能性が高い。

答え・解説

正答：エ

- ア：平均は同じと与えられている。
 - イ：標本数の指標ではない。
 - ウ：異なる分配でも同じ平均を持ち得る。
 - エ：平均は中心水準、ジニ係数は分配の不均等度を表すため別の情報である。
- 総合解説：中心水準と格差・散布の指標は別々に見る必要がある。

対象：2026年度 / 基準日 2026-09-08

0090

1-3 標準化・分布特性 > ローレンツ曲線・ジニ係数・歪度・尖度 | 難易度：応用

2地域のローレンツ曲線が途中で交差している。曲線だけから「一方がすべての累積比率でより平等」と順位付けできるか。

- ア．必ず交点の多い地域が平等である。
- イ．交差したら両地域の分配は完全に同じである。
- ウ．交差したら両者の平均が0になる。
- エ．できない。交差しているため一方が全域で完全平等線に近いとは言えず、要約指標等も併せて検討する。

答え・解説

正答：エ

- ア：交点数は平等度の定義ではない。
 - イ：一部の累積比率で一致だけで全体同一とは限らない。
 - ウ：平均水準とは無関係である。
 - エ：ローレンツ優越による一意比較には曲線が交差しないことが重要である。
- 総合解説：ローレンツ曲線が交差する場合、単純な全域優越比較はできない。

対象：2026年度 / 基準日 2026-09-08

0091

1-4 2変数データ > 散布図・共分散・相関係数 | 難易度：基礎

Xが小さいとYが低く、中程度のXでYが高く、さらにXが大きくなるとYが再び低くなる弓形の散布図が得られた。最も適切な判断はどれか。

ア．弓形なら必ず相関係数は+1になる。
イ．明確な非線形関係があり得るため、相関係数だけで関係の強さを要約すると見落としが生じる。
ウ．相関係数が0に近ければXとYは必ず無関係である。
エ．散布図はカテゴリデータにしか使えない。

答え・解説

正答：イ

ア：完全な直線関係でない限り+1とはならない。
イ：Pearson相関は線形関係を要約するので、曲線関係では小さくても関連が存在し得る。
ウ：非線形関係が残る可能性がある。
エ：2つの量的変数の関係を見る代表的な図である。
総合解説：散布図で形状を確認してから相関係数を解釈する。

対象：2026年度 / 基準日 2026-09-08

0092

1-4 2変数データ > 散布図・共分散・相関係数 | 難易度：基礎

Xが平均より大きい観測ではYも平均より大きいことが多く、Xが平均より小さい観測ではYも小さいことが多い。このとき共分散の符号として自然なのはどれか。

ア．負になりやすい。同じ向きに動くほど負になる。
イ．必ず0。平均との差は相殺されるから。
ウ．正になりやすい。偏差の積が正になる観測が多いから。
エ．必ず1。共分散は相関係数と同じ範囲だから。

答え・解説

正答：ウ

ア：同方向なら共分散は正の方向である。
イ：偏差そのものではなく偏差の積を平均する。
ウ：同符号の偏差の積は正であり、その平均的な大きさが共分散に反映される。
エ：共分散は単位に依存し±1に限定されない。
総合解説：共分散は2変数が同じ向き・逆向きに動く傾向を偏差積で表す。

対象：2026年度 / 基準日 2026-09-08

0093

1-4 2変数データ > 散布図・共分散・相関係数 | 難易度：基礎

母集団として3組 $(X,Y)=(1,2),(2,4),(3,6)$ を考える。 X 平均2、 Y 平均4のとき母共分散はどれか。

- ア . 4/3。偏差積は2,0,2で合計4、3で割る。
- イ . 0。 X と Y の各偏差和が0だから。
- ウ . 4。偏差積和をそのまま使う。
- エ . 2。 X と Y の平均の差を使う。

答え・解説

正答：ア

ア：母共分散は偏差積の平均である。
イ：偏差積の和は0ではない。
ウ：母共分散では観測数で割る。
エ：共分散の定義ではない。
総合解説：共分散計算では各組の偏差の積を平均する。

対象：2026年度 / 基準日 2026-09-08

0094

1-4 2変数データ > 散布図・共分散・相関係数 | 難易度：基礎

Pearsonの相関係数 r が取り得る範囲として正しいのはどれか。

- ア . 0以上1以下だけ。
- イ . $-\infty$ から ∞ まで。
- ウ . -1 以上1以下。
- エ . 標本数と同じ範囲。

答え・解説

正答：ウ

ア：負の線形関係では負になる。
イ：共分散とは異なり標準化される。
ウ：相関係数は共分散を両変数の標準偏差で標準化した無次元量である。
エ：標本数とは無関係である。
総合解説： $r=+1$ は完全な正の直線関係、 $r=-1$ は完全な負の直線関係を表す。

対象：2026年度 / 基準日 2026-09-08

0095

1-4 2変数データ > 散布図・共分散・相関係数 | 難易度：標準

各観測で $Y=20-3X$ が厳密に成り立ち、 X にばらつきがある。Pearson相関係数はどうなるか。

- ア . +1。直線関係なら符号に関係なく+1。
- イ . -1。負の傾きをもつ完全な直線関係である。
- ウ . 0。切片20があるから。
- エ . -3。傾きがそのまま相関係数になる。

答え・解説

正答：イ

ア：傾きが負なので符号は負である。

イ：定数項は相関に影響せず、負の比例係数をもつ完全線形関係なので $r=-1$ である。

ウ：切片は相関に影響しない。

エ：相関係数は傾きとは異なり $-1 \sim 1$ に標準化される。

総合解説：完全線形関係でも傾きの符号により相関係数の符号が決まる。

対象：2026年度 / 基準日 2026-09-08

0096

1-4 2変数データ > 散布図・共分散・相関係数 | 難易度：標準

X をcmからmへ、 Y を円から千円へそれぞれ正の倍率で単位変換した。Pearson相関係数はどうなるか。

- ア . 変わらない。正の線形尺度変換では相関係数は不変である。
- イ . 1000倍になる。
- ウ . $1/100$ になる。
- エ . 必ず符号が逆転する。

答え・解説

正答：ア

ア：共分散と標準偏差が同じ倍率で変化し標準化で相殺される。

イ：相関係数は無次元である。

ウ：単位倍率は相殺される。

エ：正の倍率では符号は変わらない。

総合解説：相関係数は測定単位に依存しないが、負の倍率変換では符号が変わる。

対象：2026年度 / 基準日 2026-09-08

0097

1-4 2変数データ > 散布図・共分散・相関係数 | 難易度：標準

新変数 $Z = -2Y$ を作った。 X と Y の相関が0.6なら、 X と Z の相関はどれか。

- ア . - 0.6。 Y に負の倍率を掛けると相関の符号が反転する。
- イ . 1.2。 0.6を2倍する。
- ウ . 0.3。 0.6を2で割る。
- エ . 0.6。 負号も相殺される。

答え・解説

正答：ア

ア：正の倍率では不変、負の倍率では方向だけ反転する。
イ：相関係数は倍率の大きさでは変わらない。
ウ：倍率の大きさは標準化で相殺される。
エ：負号は関係方向を逆転させるため符号が変わる。
総合解説：相関係数の大きさは単位倍率に不変だが、負の尺度反転で符号が反転する。

対象：2026年度 / 基準日 2026-09-08

0098

1-4 2変数データ > 散布図・共分散・相関係数 | 難易度：標準

大部分の点は弱い関係しかないが、右上に1点だけ極端な観測があり $r=0.85$ となった。最も適切な対応はどれか。

- ア . 散布図と外れ値の妥当性を確認し、その点を含む / 除く場合の相関の感度も検討する。
- イ . $r=0.85$ なら散布図を見る必要はない。
- ウ . 極端な点は必ず削除する。
- エ . 相関係数は外れ値の影響を全く受けない。

答え・解説

正答：ア

ア：Pearson相関は極端な点の影響を受けることがある。
イ：外れ値や非線形性を確認するため散布図が重要である。
ウ：実在する観測なら削除根拠が必要である。
エ：大きく変化する場合がある。
総合解説：数値要約だけでなく散布図を併用する。

対象：2026年度 / 基準日 2026-09-08

0099

1-4 2変数データ > 散布図・共分散・相関係数 | 難易度：標準

学習時間と試験得点に $r=0.70$ の正相関があった。これだけから結論できることはどれか。

ア．学習時間を1時間増やせば必ず得点が0.70点上がる。

イ． r が0.70なら第三の変数は存在しない。

ウ．正相関なら2変数の平均は等しい。

エ．両者に正の線形関連があることは示唆されるが、学習時間が得点上昇を引き起こしたと断定はできない。

答え・解説

正答：エ

ア： r は回帰係数ではなく因果効果でもない。

イ：相関だけでは交絡を排除できない。

ウ：平均の大小とは無関係である。

エ：相関は関連を示すが、交絡や逆因果などを排除しない。

総合解説：相関係数を因果効果や回帰傾きと混同しない。

対象：2026年度 / 基準日 2026-09-08

0100

1-4 2変数データ > 散布図・共分散・相関係数 | 難易度：標準

X が $-2, -1, 0, 1, 2$ を等確率に取り、 $Y=X^2$ とする。 X と Y の線形相関が0になることがあるが、最も適切な説明はどれか。

ア．相関0なら Y は X と必ず独立である。

イ． Y は X から完全に決まる非線形関係をもつため、相関0でも独立とは限らない。

ウ． $Y=X^2$ なら必ず相関+1である。

エ．相関係数は非線形関係を必ず1で表す。

答え・解説

正答：イ

ア：相関0は一般に独立を意味しない。

イ：対称性により線形共分散が相殺されても二次関係は残る。

ウ：+1は正の完全直線関係の場合である。

エ：Pearson相関は線形関係の尺度である。

総合解説：相関係数は線形関連を要約するため、散布図で非線形性も確認する。

対象：2026年度 / 基準日 2026-09-08

0101

1-4 2変数データ > 散布図・共分散・相関係数 | 難易度：応用

共分散と相関係数の違いとして最も適切なのはどれか。

- ア．共分散も相関も必ず - 1 ~ 1 の範囲である。
- イ．相関係数だけが単位を持つ。
- ウ．共分散は符号を持たない。
- エ．共分散は測定単位に依存するが、相関係数は標準化され単位に依存しない。

答え・解説

正答：エ

ア：共分散にはその制約がない。
イ：相関係数は無次元である。
ウ：逆方向に動けば負になり得る。
エ：相関係数は共分散を標準偏差の積で割って標準化する。
総合解説：比較可能性の点で相関は標準化された共分散と捉えられる。

対象：2026年度 / 基準日 2026-09-08

0102

1-4 2変数データ > 散布図・共分散・相関係数 | 難易度：応用

多数の観測が同じ整数座標に重なり、散布図では点が少なく見えた。関係を見やすくする工夫として適切なのはどれか。

- ア．重なった観測を1件だけ残して他を削除する。
- イ．点の透明度やジッター、度数を表す工夫を用いて重なりを可視化する。
- ウ．XとYをすべて同じ値に置換する。
- エ．散布図の軸を消せば重なりが解決する。

答え・解説

正答：イ

ア：データの度数情報を失う。
イ：重複点が隠れるoverplottingでは表示方法を工夫する。
ウ：関係情報を破壊する。
エ：重複点そのものは解消しない。
総合解説：散布図は点の密度や重なりにも注意して解釈する。

対象：2026年度 / 基準日 2026-09-08

0103

1-4 2変数データ > 層別・相関行列・擬相関・偏相関 | 難易度：基礎

全体ではXとYにほぼ相関がないが、年齢層別に見ると各層で強い正相関があった。最も適切な対応はどれか。

ア．全体集計だけで結論せず、年齢層による構成差を考慮して層別結果を検討する。
イ．全体相関が0なら層別結果は無効である。
ウ．層別すると必ず因果関係が証明される。
エ．年齢層を無視して観測を半分削除する。

答え・解説

正答：ア

ア：集約により層内の関係が弱まったり逆転したりすることがある。
イ：集約と層内で異なる関係は起こり得る。
ウ：層別は交絡検討に役立つが因果証明そのものではない。
エ：構成要因を理解する方法ではない。
総合解説：全体と層別の両方を比較して交絡や集約の影響を確認する。

対象：2026年度 / 基準日 2026-09-08

0104

1-4 2変数データ > 層別・相関行列・擬相関・偏相関 | 難易度：基礎

2治療法の成功率を病院別に見るとAがBより高いのに、全病院を合計するとAがBより低くなった。最も考えるべきことはどれか。

ア．成功率は割合なので集約しても逆転は数学的に不可能である。
イ．病院ごとの症例構成や割付割合が異なり、集約で関係が逆転している可能性。
ウ．病院別の結果をすべて誤差として捨てる。
エ．AとBの平均年齢が同じなら必ず逆転する。

答え・解説

正答：イ

ア：シンプソンのパラドックスとして起こり得る。
イ：層ごとの重みが異なると集約率の大小が層内比較と逆転することがある。
ウ：層別情報は重要である。
エ：逆転条件は層別構成に依存し、平均年齢だけでは決まらない。
総合解説：層別変数の分布が群間で異なる場合、集約結果の解釈に注意する。

対象：2026年度 / 基準日 2026-09-08

0105

1-4 2変数データ > 層別・相関行列・擬相関・偏相関 | 難易度：基礎

4変数の相関行列について必ず成り立つ性質はどれか。

- ア．すべての非対角要素も1である。
- イ．対角要素は必ず0である。
- ウ．対角要素は1で、行列は対称である。
- エ．上三角と下三角は符号が逆になる。

答え・解説

正答：ウ

ア：異なる変数間の相関は様々な値を取る。
イ：自己相関は1である。
ウ：各変数自身との相関は1、 $r(X,Y)=r(Y,X)$ である。
エ：相関行列は対称で同じ値である。
総合解説：相関行列は多数の変数ペアの線形関連を一覧化する。

対象：2026年度 / 基準日 2026-09-08

0106

1-4 2変数データ > 層別・相関行列・擬相関・偏相関 | 難易度：標準

相関行列で $r(A,B)=0.82$, $r(A,C)=-0.15$, $r(B,C)=0.05$ だった。線形関係の絶対的な強さが最も大きいペアはどれか。

- ア．AとC。負の相関は必ず正より強い。
- イ．BとC。0.05が0に近いから強い。
- ウ．AとB。|0.82|が最も大きい。
- エ．3ペアは同じ。符号が異なるだけ。

答え・解説

正答：ウ

ア：強さは絶対値で比較し、0.15は小さい。
イ：0に近いほど線形関連は弱い。
ウ：相関の強さは符号ではなく絶対値も見て比較する。
エ：絶対値が異なる。
総合解説：相関係数の符号は方向、絶対値は線形関連の強さを示す。

対象：2026年度 / 基準日 2026-09-08

0107

1-4 2変数データ > 層別・相関行列・擬相関・偏相関 | 難易度：標準

地域ごとの冷房使用量と水難事故件数に正相関があった。最も自然な第三変数候補はどれか。

- ア．相関係数そのもの。
- イ．標本平均。
- ウ．気温。暑い時期には冷房使用も水辺活動も増える可能性がある。
- エ．単位名。

答え・解説

正答：ウ

ア：相関係数は第三の原因変数ではない。
イ：平均は共通原因の候補ではない。
ウ：共通原因が2変数を同時に変化させると擬相関が生じ得る。
エ：単位表記が両現象を生じさせるわけではない。
総合解説：観察データの相関では第三変数・共通トレンドなどを検討する。

対象：2026年度 / 基準日 2026-09-08

0108

1-4 2変数データ > 層別・相関行列・擬相関・偏相関 | 難易度：標準

XとYのZを制御した偏相関の説明として適切なのはどれか。

- ア．Zの値をすべて0に書き換えて計算する。
- イ．XとYからZと線形に関連する部分を調整した後に残る線形関連を要約する。
- ウ．XとYの因果効果を必ず表す。
- エ．単純相関を必ず2倍した値である。

答え・解説

正答：イ

ア：データ改変ではない。
イ：偏相関は共通する第三変数の線形影響を取り除いた残差同士の相関として理解できる。
ウ：交絡を調整しても因果性が自動的に証明されるわけではない。
エ：一般的な関係ではない。
総合解説：偏相関は特定の変数を線形に調整した関連を見る指標である。

対象：2026年度 / 基準日 2026-09-08

0109

1-4 2変数データ > 層別・相関行列・擬相関・偏相関 | 難易度：標準

XとYの単純相関0.74に対し、年齢Zを制御した偏相関は0.18だった。最も適切な解釈はどれか。

- ア．年齢がXとYの因果関係を完全に証明した。
- イ．単純相関0.74は計算ミスである。
- ウ．偏相関0.18ならXとYは必ず独立である。
- エ．単純な関連のかなりの部分が年齢との線形な関係で説明される可能性がある。

答え・解説

正答：エ

- ア：偏相関だけでは因果証明できない。
 - イ：調整前後で値が異なること自体は不自然でない。
 - ウ：小さな線形偏相関は独立を保証しない。
 - エ：Z調整後に相関が大きく低下したことは、Zが両変数と関連していた可能性を示す。
- 総合解説：調整前後の相関差は第三変数の影響を考える手掛かりになる。

対象：2026年度 / 基準日 2026-09-08

0110

1-4 2変数データ > 層別・相関行列・擬相関・偏相関 | 難易度：標準

100変数の相関行列から絶対値が最大の1組だけを選び「真の強い関係」と報告した。注意点として適切なのはどれか。

- ア．多数の組合せを探索すると偶然大きな相関が選ばれやすく、再現性や検証用データを考慮すべきである。
- イ．相関行列では偶然変動が存在しない。
- ウ．変数が100個なら全相関は必ず0である。
- エ．最大相関は必ず因果関係を意味する。

答え・解説

正答：ア

- ア：大量のペアから最大値を選ぶと選択バイアスが生じる。
 - イ：標本相関には標本変動がある。
 - ウ：変数数から相関値は決まらない。
 - エ：相関と因果は別である。
- 総合解説：探索的な相関発見は独立データや理論的根拠で確認することが望ましい。

対象：2026年度 / 基準日 2026-09-08

0111

1-4 2変数データ > 層別・相関行列・擬相関・偏相関 | 難易度：応用

細かく20層に分けた結果、各層の観測が2～3件しかなく相関が非常に不安定になった。最も適切な評価はどれか。

ア．層を細かくするほど相関推定は必ず正確になる。
イ．各層2件なら相関は必ず0になる。
ウ．層別は交絡確認に有用だが、過度に細分すると各層の標本数が減り推定が不安定になる。
エ．層別した時点で標本誤差は消える。

答え・解説

正答：ウ

ア：小標本では変動が大きくなる。
イ：2点では相関が極端になりやすく、安定しない。
ウ：層別には情報の均質化と標本数確保のトレードオフがある。
エ：標本変動は残る。
総合解説：層別分析では各層の標本サイズと推定精度も確認する。

対象：2026年度 / 基準日 2026-09-08

0112

1-4 2変数データ > 層別・相関行列・擬相関・偏相関 | 難易度：応用

相関行列で $r(P,Q) = -0.91$ だった。この値の意味として適切なのはどれか。

ア．PとQに強い正の関係がある。
イ．PとQの平均値の差が-0.91である。
ウ．Pが大きいほどQが小さい強い負の線形関係が示唆される。
エ．QがPの91%を必ず引き起こす。

答え・解説

正答：ウ

ア：符号の方向が逆である。
イ：相関は平均差ではない。
ウ：-1に近い相関は強い負の直線的関連を表す。
エ：因果割合を意味しない。
総合解説：相関の符号と絶対値を分けて解釈する。

対象：2026年度 / 基準日 2026-09-08

0113

1-4 2変数データ > カテゴリカルデータ・度数表・2元クロス表 | 難易度：基礎

回答「賛成・反対・未定」を集計する基本表として最も適切なのはどれか。

- ア．各カテゴリの度数と相対度数を並べる一元度数表。
- イ．回答を1,2,3として平均だけを示す。
- ウ．各回答の標準偏差だけを示す。
- エ．散布図だけを作る。

答え・解説

正答：ア

ア：1つのカテゴリ変数はカテゴリ別件数と割合で要約できる。
イ：コードに数量的間隔の意味がない場合、平均は中心的要約にならない。
ウ：カテゴリ頻度を直接示さない。
エ：1変数カテゴリの度数把握には適さない。
総合解説：カテゴリカルデータは度数・割合が基本要約となる。

対象：2026年度 / 基準日 2026-09-08

0114

1-4 2変数データ > カテゴリカルデータ・度数表・2元クロス表 | 難易度：基礎

性別（男・女）×購入（有・無）の2元クロス表で「女・購入有」のセルが35である。この35は何を表すか。

- ア．女性かつ購入有の両条件を満たす観測数。
- イ．女性の総数。
- ウ．購入有の総数。
- エ．全標本数。

答え・解説

正答：ア

ア：クロス表の内部セルは2つのカテゴリ条件の同時度数を表す。
イ：女性行の周辺度数とは限らない。
ウ：購入有列の周辺度数とは限らない。
エ：表全体の合計とは異なる。
総合解説：内部セルと周辺度数を区別する。

対象：2026年度 / 基準日 2026-09-08

0115

1-4 2変数データ > カテゴリカルデータ・度数表・2元クロス表 | 難易度：基礎

あるクロス表でA群100人中、商品購入者が40人だった。A群内購入率はどれか。

- ア．60%。非購入者割合を答える。
- イ．40%。40/100である。
- ウ．140%。人数を足す。
- エ．2.5%。100/40を百分率化する。

答え・解説

正答：イ

ア：購入率ではない。
イ：行内割合はその行の合計を分母にする。
ウ：割合の計算ではない。
エ：比の向きが逆である。
総合解説：条件付き割合を比較するときにはどの周辺合計を分母にしているか明確にする。

対象：2026年度 / 基準日 2026-09-08

0116

1-4 2変数データ > カテゴリカルデータ・度数表・2元クロス表 | 難易度：標準

「喫煙者のうち疾病ありの割合」と「疾病ありのうち喫煙者の割合」は同じか。

- ア．必ず同じ。クロス表の同じセルを使うから。
- イ．どちらも全標本数を分母にする。
- ウ．一般には異なる。前者は喫煙者総数、後者は疾病あり総数を分母にする。
- エ．カテゴリデータでは割合を計算できない。

答え・解説

正答：ウ

ア：分子が同じでも分母が異なる。
イ：条件付き割合では周辺合計を分母にする。
ウ：条件の向きが違いため分母が異なる。
エ：割合は基本的な要約である。
総合解説：条件付き確率・割合では「何のうち」を明確にする。

対象：2026年度 / 基準日 2026-09-08

0117

1-4 2変数データ > カテゴリカルデータ・度数表・2元クロス表 | 難易度：標準

2×2表のセルが [[30,20],[10,40]] である。第1行の周辺度数と第2列の周辺度数の組合せはどれか。

- ア．第1行30、第2列40。
- イ．第1行60、第2列50。
- ウ．第1行100、第2列100。
- エ．第1行50、第2列60。

答え・解説

正答：エ

ア：内部セルだけを採っている。
イ：行・列の合計を取り違えている。
ウ：総数を双方に使っている。
エ：第1行30+20=50、第2列20+40=60である。
総合解説：周辺度数は行または列の内部セルを合計して求める。

対象：2026年度 / 基準日 2026-09-08

0118

1-4 2変数データ > カテゴリカルデータ・度数表・2元クロス表 | 難易度：標準

2×2表で第1行合計60、第1列合計50、総数100である。独立を仮定した第1行×第1列セルの期待度数はどれか。

- ア．30。60×50/100である。
- イ．110。行合計と列合計を足す。
- ウ．50。列合計だけを使う。
- エ．10。100 - 60 - 50とする。

答え・解説

正答：ア

ア：独立時の期待度数は行合計×列合計/総数で求める。
イ：期待度数の式ではない。
ウ：行割合を反映していない。
エ：セル期待値の算出方法ではない。
総合解説：クロス表の独立性の検討では周辺度数から期待度数を求める。

対象：2026年度 / 基準日 2026-09-08

0119

1-4 2変数データ > カテゴリカルデータ・度数表・2元クロス表 | 難易度：標準

2つの2 × 2表が行合計・列合計とも全く同じだった。最も適切な説明はどれか。

- ア．周辺合計が同じなら内部セルも必ず同じ。
- イ．周辺合計が同じなら2変数は必ず独立。
- ウ．周辺合計が同じなら総数は異なる。
- エ．内部セルの配分が異なれば、2変数の関連の強さは異なり得る。

答え・解説

正答：エ

ア：自由度があり異なる配分が可能である。
イ：独立性は内部セルと期待度数の一致で判断する。
ウ：両方の周辺合計が同じなら総数も同じである。
エ：周辺分布だけでは共同分布の内部構造は一意に決まらない。
総合解説：カテゴリ変数の関連を見るには内部セルの組合せを見る必要がある。

対象：2026年度 / 基準日 2026-09-08

0120

1-4 2変数データ > カテゴリカルデータ・度数表・2元クロス表 | 難易度：標準

5段階の回答を「肯定・中立・否定」の3カテゴリへまとめた。最も適切な説明はどれか。

- ア．カテゴリをまとめると必ず関連が強くなる。
- イ．元の5カテゴリを完全に復元できる。
- ウ．表は簡潔になるが、元の細かなカテゴリ差の情報は失われる。
- エ．標本数が3/5に減る。

答え・解説

正答：ウ

ア：強弱はデータ次第である。
イ：統合後の表だけから内部内訳は復元できない。
ウ：カテゴリ統合は情報を圧縮するため結果の見え方が変わることがある。
エ：観測数ではなくカテゴリ数が減る。
総合解説：再分類は分析目的に沿って根拠を明確にする。

対象：2026年度 / 基準日 2026-09-08

0121

1-4 2変数データ > カテゴリカルデータ・度数表・2元クロス表 | 難易度：応用

アンケートの無回答を除外して割合を計算した。報告上の注意として適切なのはどれか。

- ア．無回答は自動的に「反対」とみなす。
- イ．有効回答数を分母として明示し、無回答数・割合も必要に応じて示す。
- ウ．無回答があっても分母は必ず配布数と同じにする。
- エ．カテゴリデータでは欠測は発生しない。

答え・解説

正答：イ

ア：根拠のない代入である。
イ：欠測除外により分母が変わるため、集計対象を明示する必要がある。
ウ：指標の定義に応じ分母を明示すべきである。
エ：無回答など欠測は起こり得る。
総合解説：クロス集計の割合は分母の定義で値が変わるため明記する。

対象：2026年度 / 基準日 2026-09-08

0122

1-4 2変数データ > カテゴリカルデータ・度数表・2元クロス表 | 難易度：応用

職業カテゴリと疾病有無のクロス表で大きな割合差が見られた。これだけから言えることとして適切なのはどれか。

- ア．割合差があれば因果関係が自動的に証明される。
- イ．カテゴリ変数間では関連を評価できない。
- ウ．割合差があると2群の人数は必ず同じである。
- エ．両変数に関連がある可能性は示せるが、職業が疾病を引き起こしたと因果断定はできない。

答え・解説

正答：エ

ア：研究デザインと交絡の検討が必要である。
イ：クロス表やカイ二乗検定等で評価できる。
ウ：人数の等しさとは無関係である。
エ：クロス表は関連を記述するが、観察データでは交絡等を排除しない。
総合解説：カテゴリカルデータでも関連と因果を区別する。

対象：2026年度 / 基準日 2026-09-08

0123

1-5 データ収集・実験計画 > 観察研究・実験研究・調査設計 | 難易度：基礎

研究者が参加者の食事内容を変更せず、普段の食習慣と血圧を記録した。この研究デザインとして適切なのはどれか。

- ア．無作為化比較実験。
- イ．観察研究。研究者が曝露を割り付けていない。
- ウ．完全無作為化実験。
- エ．ブロック実験。

答え・解説

正答：イ

ア：食事条件を無作為に割り付けていない。
イ：自然に存在する食習慣を観察して関連を調べている。
ウ：介入の割付がない。
エ：処理条件をブロック内に割り付けた実験ではない。
総合解説：研究者が要因を操作・割付するかどうかで観察研究と実験研究を区別する。

対象：2026年度 / 基準日 2026-09-08

0124

1-5 データ収集・実験計画 > 観察研究・実験研究・調査設計 | 難易度：基礎

新しい学習法と従来法を比較するため、生徒を無作為に2群へ割り付けて8週間後の得点を比較した。最も適切な分類はどれか。

- ア．横断的な観察研究。
- イ．単純無作為抽出だけを行った標本調査。
- ウ．ケースシリーズ。
- エ．無作為化比較実験。

答え・解説

正答：エ

ア：研究者が学習法を割り付けている。
イ：対象抽出ではなく処理割付が中心である。
ウ：比較群と無作為割付がある。
エ：研究者が処理を割り付け、その割付に無作為化を用いている。
総合解説：無作為割付は処理群間の背景因子の偏りを偶然化し、因果比較を支える。

対象：2026年度 / 基準日 2026-09-08

0125 1-5 データ収集・実験計画 > 観察研究・実験研究・調査設計 | 難易度：基礎

ある1時点で1000人の運動習慣と肥満の有無を同時に調べた。デザインの特徴として適切なのはどれか。

- ア．必ず前向きコホート研究である。
- イ．肥満を研究者が割り付けた実験である。
- ウ．横断研究なら因果の向きが必ず分かる。
- エ．横断的観察研究で、関連は調べられるが時間的な前後関係の判断には限界がある。

答え・解説 正答：エ

- ア：追跡をしていない。
 - イ：疾病状態を割り付けていない。
 - ウ：同時測定では前後関係を確定しにくい。
 - エ：曝露と結果を同時点で測る横断研究では時間順序が不明瞭になりやすい。
- 総合解説：観察研究のデザインごとに得られる情報と限界が異なる。

対象：2026年度 / 基準日 2026-09-08

0126 1-5 データ収集・実験計画 > 観察研究・実験研究・調査設計 | 難易度：標準

喫煙状況を基準時に記録し、その後10年間追跡して疾病発生を比較した。最も適切な説明はどれか。

- ア．喫煙を無作為割付した実験である。
- イ．1時点だけの横断研究である。
- ウ．追跡研究では脱落は起こり得ない。
- エ．時間的な追跡を行う観察研究で、曝露後の疾病発生を比較できる。

答え・解説 正答：エ

- ア：曝露を研究者が操作していない。
 - イ：10年間の追跡がある。
 - ウ：長期追跡では脱落がバイアス要因になり得る。
 - エ：研究者が喫煙を割り付けず、基準時の曝露から結果を追跡している。
- 総合解説：縦断観察では時間順序を把握しやすい一方、交絡や脱落に注意する。

対象：2026年度 / 基準日 2026-09-08

0127

1-5 データ収集・実験計画 > 観察研究・実験研究・調査設計 | 難易度：標準

コーヒー摂取量と心疾患に関連が見られたが、コーヒー摂取量の多い群では喫煙率も高かった。喫煙について最も適切な説明はどれか。

- ア．交絡とは無関係で、必ず媒介変数である。
- イ．喫煙率が違っても観察研究には影響しない。
- ウ．交絡は実験研究でしか生じない。
- エ．コーヒー摂取と心疾患の両方に関連するなら交絡要因になり得る。

答え・解説

正答：エ

ア：因果構造を確認せず媒介と断定できない。
イ：群間の背景差は解釈に影響する。
ウ：観察研究で特に重要な問題である。
エ：第三変数が曝露と結果の双方に関連すると観察された関連を歪める可能性がある。
総合解説：観察研究の因果解釈では交絡候補を測定・調整する。

対象：2026年度 / 基準日 2026-09-08

0128

1-5 データ収集・実験計画 > 観察研究・実験研究・調査設計 | 難易度：標準

健康アプリ利用者だけにオンライン回答を募り「全国成人の平均運動時間」を推定した。主な懸念として適切なものはどれか。

- ア．標本数が多ければ選択バイアスは必ず消える。
- イ．オンライン調査なら無作為抽出が自動的に成立する。
- ウ．選択バイアスは測定単位だけの問題である。
- エ．健康意識の高い人が参加しやすく、対象母集団を代表しない選択バイアスが生じ得る。

答え・解説

正答：エ

ア：偏った抽出は大標本でも偏りを残す。
イ：募集方法次第である。
ウ：標本への入り方の偏りである。
エ：参加機会や回答意欲が運動習慣と関連すると標本が偏る。
総合解説：代表性は標本数だけでなく抽出・参加過程に依存する。

対象：2026年度 / 基準日 2026-09-08

0129

1-5 データ収集・実験計画 > 観察研究・実験研究・調査設計 | 難易度：標準

「健康のため当然毎日運動すべきだと思いますが、あなたも運動していますか」と質問した。問題点として最も適切なのはどれか。

- ア．質問が長いほど必ず無作為化される。
- イ．誘導的な質問で、回答を特定方向へ偏らせる可能性がある。
- ウ．誘導質問は標本誤差だけを小さくする。
- エ．質問票では文言は結果に影響しない。

答え・解説

正答：イ

ア：質問長と無作為化は別である。
イ：価値判断を含む文言は社会的望ましさや同意方向のバイアスを生み得る。
ウ：測定上の非標本誤差を増やし得る。
エ：質問の表現は回答に影響し得る。
総合解説：調査設計では質問文、順序、回答選択肢などによる測定誤差にも注意する。

対象：2026年度 / 基準日 2026-09-08

0130

1-5 データ収集・実験計画 > 観察研究・実験研究・調査設計 | 難易度：標準

新薬試験で対照群を設ける主な目的として適切なのはどれか。

- ア．対照群を設けると標本数が必ず1になる。
- イ．対照群にはデータを記録しない。
- ウ．対照群があれば無作為化は不要である。
- エ．処理を受けた群と比較する基準を作り、自然変動や時間経過などと処理効果を区別しやすくする。

答え・解説

正答：エ

ア：標本数とは無関係である。
イ：比較には同様に結果測定が必要である。
ウ：無作為化と対照は別の設計要素である。
エ：比較対象があることで観測変化の背景要因を評価しやすい。
総合解説：適切な対照条件は処理効果の比較基準となる。

対象：2026年度 / 基準日 2026-09-08

0131

1-5 データ収集・実験計画 > 観察研究・実験研究・調査設計 | 難易度：応用

治療効果の評価者が参加者の割付群を知ると主観評価が影響される恐れがある。評価者盲検化の目的として適切なものはどれか。

- ア．標本を母集団から無作為抽出するため。
- イ．処理群の人数を自動的に等しくするため。
- ウ．割付情報による評価の偏りを減らす。
- エ．欠測値を0に置換するため。

答え・解説

正答：ウ

ア：盲検化は抽出法ではない。
イ：群サイズ調整の方法ではない。
ウ：群情報を知らないことで期待や先入観が測定に入り込むのを抑える。
エ：欠測処理とは無関係である。
総合解説：実験の内部妥当性には割付だけでなく測定過程のバイアス低減も重要である。

対象：2026年度 / 基準日 2026-09-08

0132

1-5 データ収集・実験計画 > 観察研究・実験研究・調査設計 | 難易度：応用

観察研究で強い関連があり、無作為化実験でも同方向の差が再現された。因果解釈について最も適切なものはどれか。

- ア．相関が強ければ研究デザインは一切不要である。
- イ．無作為化実験は観察研究より交絡を抑えやすく、因果的解釈を強める証拠になるが、実施・測定上の限界も確認する。
- ウ．無作為化実験では測定誤差も脱落も絶対に起きない。
- エ．観察研究と実験研究は因果推論上まったく同じ情報量である。

答え・解説

正答：イ

ア：関連の強さだけでは因果性を保証しない。
イ：研究デザインにより交絡やバイアスへの耐性が異なる。
ウ：実験にも実務上の限界はある。
エ：処理割付の制御が大きな違いである。
総合解説：結論の強さはデータだけでなく、どのようにデータが生成されたかに依存する。

対象：2026年度 / 基準日 2026-09-08

0133

1-5 データ収集・実験計画 > 無作為抽出・標本抽出法・標本誤差 | 難易度：基礎

母集団名簿5000人から乱数を用いて200人を選ぶ方法として最も適切な名称はどれか。

- ア．便宜抽出。
- イ．単純無作為抽出。
- ウ．割当法。
- エ．全数調査。

答え・解説

正答：イ

ア：利用しやすい人だけを選ぶ方法ではない。

イ：各個体を無作為な仕組みで標本へ選ぶ基本的な抽出法である。

ウ：カテゴリ別人数を任意に割り当てる方法ではない。

エ：5000人全員を調べていない。

総合解説：無作為抽出は標本誤差を確率的に評価する基盤となる。

対象：2026年度 / 基準日 2026-09-08

0134

1-5 データ収集・実験計画 > 無作為抽出・標本抽出法・標本誤差 | 難易度：基礎

名簿から無作為に開始位置を1つ選び、その後20人おきに抽出する方法はどれか。

- ア．層化抽出。
- イ．系統抽出。
- ウ．クラスター抽出。
- エ．有意抽出。

答え・解説

正答：イ

ア：母集団を層に分けて各層から抽出する方法ではない。

イ：ランダムスタート後に一定間隔で抽出する。

ウ：集団単位を抽出する方法ではない。

エ：一定間隔の確率的手続きとは異なる。

総合解説：系統抽出では名簿の周期性が抽出間隔と重なる場合に注意する。

対象：2026年度 / 基準日 2026-09-08

0135

1-5 データ収集・実験計画 > 無作為抽出・標本抽出法・標本誤差 | 難易度：基礎

全国調査で地域ごとの標本数を確保したい。地域を層に分け、各地域から無作為抽出する利点として適切なのはどれか。

- ア．層化すると無回答が必ず0になる。
- イ．各層を確実に標本へ含められ、層内が均質なら推定精度向上も期待できる。
- ウ．層化すると標本誤差が数学的に必ず0になる。
- エ．層化では各層から調査しない。

答え・解説

正答：イ

ア：回答率とは別問題である。
イ：層化は重要なサブグループの代表性確保と分散低減に利用される。
ウ：標本調査なので標本変動は残る。
エ：各層から抽出する設計である。
総合解説：層化抽出は母集団構造を利用して抽出を設計する。

対象：2026年度 / 基準日 2026-09-08

0136

1-5 データ収集・実験計画 > 無作為抽出・標本抽出法・標本誤差 | 難易度：標準

母集団2000人がA層1200人、B層800人からなる。標本200人を比例配分するとA層とB層は何人が。

- ア．A120人、B80人。母集団比60％:40％を標本にも適用する。
- イ．A100人、B100人。
- ウ．A80人、B120人。
- エ．A200人、B200人。

答え・解説

正答：ア

ア： $1200/2000=0.6$ 、 $800/2000=0.4$ なので200人を比例配分する。
イ：これは等配分であり比例配分ではない。
ウ：層比を逆にしている。
エ：標本総数400人になってしまう。
総合解説：比例配分では各層標本数を母集団層比に合わせる。

対象：2026年度 / 基準日 2026-09-08

0137

1-5 データ収集・実験計画 > 無作為抽出・標本抽出法・標本誤差 | 難易度：標準

全国の小学校から学校を無作為に選び、選ばれた学校の児童を調査する方法に近いのはどれか。

- ア．クラスター抽出。学校という集団単位を抽出する。
- イ．単純無作為抽出で全国児童名簿から個人を直接選ぶ方法。
- ウ．層化抽出で全学校から必ず一部児童を取る方法。
- エ．全数調査。

答え・解説

正答：ア

- ア：個人でなく自然な集団を一次抽出単位にする。
 - イ：記述では学校単位を選んでいる。
 - ウ：選ばれない学校がある点が異なる。
 - エ：全学校・全児童を調査していない。
- 総合解説：クラスター抽出は調査コストを下げやすい一方、同一クラスター内の類似性で精度が低下することがある。

対象：2026年度 / 基準日 2026-09-08

0138

1-5 データ収集・実験計画 > 無作為抽出・標本抽出法・標本誤差 | 難易度：標準

まず市区町村を抽出し、次に抽出市区町村内の世帯を抽出し、最後に世帯員を選ぶ方法はどれか。

- ア．一段の単純無作為抽出だけ。
- イ．全数調査。
- ウ．層化だけで、段階抽出ではない。
- エ．多段抽出。

答え・解説

正答：エ

- ア：複数段階の抽出単位がある。
 - イ：母集団全員ではない。
 - ウ：層化と多段は別概念で、記述は明確に多段である。
 - エ：抽出を複数の段階に分けて行っている。
- 総合解説：大規模調査ではコストと実施可能性のため多段抽出がよく用いられる。

対象：2026年度 / 基準日 2026-09-08

0139

1-5 データ収集・実験計画 > 無作為抽出・標本抽出法・標本誤差 | 難易度：標準

同じ母集団から同じ方法で無作為標本を繰り返し取ると推定値が少しずつ変わる。この変動は主に何と呼ばれるか。

- ア．必ず測定器の故障。
- イ．非回答バイアスだけ。
- ウ．データ改ざん。
- エ．標本誤差（標本変動）。

答え・解説

正答：エ

- ア：測定誤差とは別である。
 - イ：回答欠落による非標本誤差とは別である。
 - ウ：確率標本でも自然に生じる変動である。
 - エ：母集団の一部だけを観測することによる偶然変動である。
- 総合解説：標本誤差と非標本誤差を区別する。

対象：2026年度 / 基準日 2026-09-08

0140

1-5 データ収集・実験計画 > 無作為抽出・標本抽出法・標本誤差 | 難易度：標準

他の条件が同じで単純無作為抽出の標本数を100から400へ増やした。標準誤差の一般的な変化として最も自然なのはどれか。

- ア．4倍になる。
- イ．おおむね1/2になる方向。標準誤差は多くの場合 $1/\sqrt{n}$ に比例する。
- ウ．必ず0になる。
- エ．標本数と標準誤差は一切関係しない。

答え・解説

正答：イ

- ア：標本数増加で通常は精度が上がる。
 - イ：標本数4倍で平方根は2倍なので標準誤差は半分程度になる。
 - ウ：有限標本では一般に0にはならない。
 - エ：多くの推定量でn増加により標準誤差は小さくなる。
- 総合解説：標本数を増やす効果は平方根則で逡減することが多い。

対象：2026年度 / 基準日 2026-09-08

0141

1-5 データ収集・実験計画 > 無作為抽出・標本抽出法・標本誤差 | 難易度：応用

無作為抽出は適切だったが、質問文が曖昧で回答者が異なる意味に解釈した。これは主にどの種類の問題か。

- ア．標本誤差だけ。
- イ．有限母集団補正だけ。
- ウ．無作為化による処理効果。
- エ．測定に起因する非標本誤差。

答え・解説

正答：エ

ア：標本の偶然変動ではない。
イ：抽出率の補正とは無関係である。
ウ：実験の処理割付ではない。
エ：抽出以外の質問・測定過程で生じる誤差である。
総合解説：標本数を増やしても非標本誤差の偏りは自動的には解消しない。

対象：2026年度 / 基準日 2026-09-08

0142

1-5 データ収集・実験計画 > 無作為抽出・標本抽出法・標本誤差 | 難易度：応用

無作為に5000人へ調査票を送り、500人だけ回答した。回答者と非回答者で意見が異なる可能性が高い。最も重要な懸念はどれか。

- ア．500人もいれば標本誤差もバイアスも必ず0である。
- イ．無作為抽出なら回答率は結果に影響しない。
- ウ．非回答者は自動的に反対意見として計上する。
- エ．非回答バイアスにより、回答者の結果が母集団を代表しない可能性。

答え・解説

正答：エ

ア：標本数だけでは非回答バイアスを消せない。
イ：非回答が結果と関連すれば影響する。
ウ：根拠のない代入である。
エ：抽出時に無作為でも、回答段階で系統的な欠落が起きれば偏り得る。
総合解説：調査精度には抽出設計だけでなく回答過程も重要である。

対象：2026年度 / 基準日 2026-09-08

0143

1-5 データ収集・実験計画 > フィッシャーの3原則・処理群 / 対照群 | 難易度：基礎

フィッシャーの実験計画の3原則として適切な組合せはどれか。

- ア．無作為化・反復・局所管理（ブロック化）。
- イ．平均・分散・中央値。
- ウ．推定・検定・相関。
- エ．抽出・欠測・補完。

答え・解説

正答：ア

ア：処理割付の無作為化、誤差評価のための反復、既知の変動要因を抑える局所管理が基本である。
イ：記述統計量であり実験計画の3原則ではない。
ウ：統計分析手法の分類である。
エ：3原則ではない。
総合解説：実験計画ではデータ生成段階で比較の妥当性と精度を高める。

対象：2026年度 / 基準日 2026-09-08

0144

1-5 データ収集・実験計画 > フィッシャーの3原則・処理群 / 対照群 | 難易度：基礎

実験で処理A/Bを実験単位へ無作為に割り付ける主な理由として最も適切なのはどれか。

- ア．各群の結果を必ず同じにするため。
- イ．既知・未知の背景要因が特定処理へ系統的に偏るのを抑え、処理比較の妥当性を高める。
- ウ．標本数を必ず2にするため。
- エ．測定器の精度を自動的に上げるため。

答え・解説

正答：イ

ア：無作為化は結果の同一化ではない。
イ：無作為化は交絡を偶然化し、統計的推論の基盤を作る。
ウ：群数や標本数とは別である。
エ：測定精度を直接改善する操作ではない。
総合解説：無作為化は処理群間の系統的な割付偏りを減らす。

対象：2026年度 / 基準日 2026-09-08

0145

1-5 データ収集・実験計画 > フィッシャーの3原則・処理群 / 対照群 | 難易度：標準

同じ処理条件を複数の独立な実験単位で繰り返す反復の主な目的はどれか。

- ア．偶然変動を評価し、処理効果の推定精度を高める。
- イ．同じ測定値をコピーして標本数を増やす。
- ウ．処理の種類を1つに減らす。
- エ．無作為化を禁止する。

答え・解説

正答：ア

ア：複数の独立な実験単位があることで誤差分散を評価できる。
イ：コピーは独立な反復ではない。
ウ：比較処理数とは別である。
エ：反復と無作為化は併用できる。
総合解説：反復は擬似反復と区別し、独立な実験単位で行う。

対象：2026年度 / 基準日 2026-09-08

0146

1-5 データ収集・実験計画 > フィッシャーの3原則・処理群 / 対照群 | 難易度：標準

農地の東西で土壌条件が異なるため、似た場所をブロックにまとめ、各ブロック内で肥料A/Bを比較した。ブロック化の目的はどれか。

- ア．肥料Aだけを東側に割り付ける。
- イ．土壌差を無視して標本数を減らす。
- ウ．ブロック内では処理を無作為化してはいけない。
- エ．土壌差による変動をブロック要因として分離し、肥料比較の精度を高める。

答え・解説

正答：エ

ア：処理と位置が完全交絡する。
イ：変動要因を利用するのがブロック化である。
ウ：通常はブロック内で無作為化する。
エ：既知の大きな変動要因を局所的にそろえることで誤差を減らす。
総合解説：ブロック化と無作為化を組み合わせると既知の異質性を抑えつつ割付偏りを防げる。

対象：2026年度 / 基準日 2026-09-08

0147

1-5 データ収集・実験計画 > フィッシャーの3原則・処理群 / 対照群 | 難易度：標準

1枚の培養皿に処理Aを施し、同じ皿から10回測定して「Aの反復数10」とした。問題点として適切なのはどれか。

ア．10回測れば実験単位も自動的に10になる。
イ．同じ皿の測定は必ずすべて独立である。
ウ．独立な実験単位は1枚なので、10測定を独立な処理反復とみなすと擬似反復になる。
エ．反復測定はどんな目的でも禁止される。

答え・解説

正答：ウ

ア：実験単位は処理が独立に割り付けられる単位で決まる。
イ：共通環境による相関があり得る。
ウ：同一実験単位内の反復測定は処理割付の独立反復とは異なる。
エ：測定精度評価には有用だが独立反復と区別する。
総合解説：反復の単位を正しく定義し、観測数と独立実験単位数を混同しない。

対象：2026年度 / 基準日 2026-09-08

0148

1-5 データ収集・実験計画 > フィッシャーの3原則・処理群 / 対照群 | 難易度：標準

実験単位がほぼ均質で、既知の大きな位置差もない。3処理を比較する最も単純な設計として適切なのはどれか。

ア．必ず100個のブロックを作る。
イ．処理ごとに実施日を固定して無作為化しない。
ウ．処理を実験単位へ無作為に割り付ける完全無作為化設計。
エ．結果を見てから処理を割り付ける。

答え・解説

正答：ウ

ア：不要な細分化は設計を複雑にする。
イ：日差と処理が交絡する可能性がある。
ウ：均質な実験単位なら追加ブロックを設けず無作為割付する設計が自然である。
エ：事後割付は比較を歪める。
総合解説：設計は既知の変動構造に応じて選ぶ。

対象：2026年度 / 基準日 2026-09-08

0149

1-5 データ収集・実験計画 > フィッシャーの3原則・処理群 / 対照群 | 難易度：応用

処理Aをすべて午前、処理Bをすべて午後に実施した。室温が午後に高くなるときの主な問題はどれか。

ア．午前と午後を固定すると無作為化より必ず精度が高い。

イ．時間帯は処理と無関係なので問題にならない。

ウ．処理が2種類なら交絡は定義されない。

エ．処理と時間帯が交絡し、差が処理効果が室温差が区別しにくい。

答え・解説

正答：エ

ア：時間差が処理差に混入する。

イ：室温変化が結果に影響する可能性がある。

ウ：2処理でも交絡は起こる。

エ：処理割付と外部条件が完全に結び付くと因果解釈が困難になる。

総合解説：実施順の無作為化やブロック化で時間的変動との交絡を抑える。

対象：2026年度 / 基準日 2026-09-08

0150

1-5 データ収集・実験計画 > フィッシャーの3原則・処理群 / 対照群 | 難易度：応用

4種類の製造条件を3日間比較する。日をブロックとし、各日で4条件の順序を無作為化し、各条件を複数の独立試料で反復した。評価として最も適切なものはどれか。

ア．ブロック化したので無作為化と反復は不要である。

イ．反復があるので日差は自動的に完全消失する。

ウ．日差をブロック化し、順序の無作為化と独立反復も行うため、3原則を組み合わせた設計である。

エ．4条件あるため対照比較は統計的に不可能である。

答え・解説

正答：ウ

ア：3原則は相補的である。

イ：ブロック化で日差を明示的に扱う意義がある。

ウ：既知の変動要因を制御しつつ割付偏りを防ぎ、誤差評価のための反復を持つ。

エ：複数処理の比較は実験計画・分散分析で扱える。

総合解説：無作為化・反復・局所管理は異なる役割を持ち、適切に組み合わせる。

対象：2026年度 / 基準日 2026-09-08