練習問題の解答例

このファイルは，近代科学社より刊行されている分寺杏介「芯まで身につく はじめての統計学」の各章末にある練習問題の解答例です。

責任者：分寺杏介（bunji@bear.kobe-u.ac.jp）

※誤字脱字，内容の誤り，計算ミス等を見つけられた方は，何らかの方法で著者までお知らせいただけると幸いです。

【変更履歴】

2025/09/05	初版公開

第0章

(1) 日常生活やニュース……

特に「差がある」「効果があった」といった結論は，「仮説検定」という手法を用いて導かれていることが多いです。これは，本章で紹介したように（そして第 10, 11 章で詳しく説明する）統計的仮説検定が，差や効果の「有無」を検証する方法として用いられているためです。

【解答例】

- 朝食を食べる習慣がある人のほうがテストの成績が良い
- 女性役員・管理職の比率が高い企業ほど，収益性が高い
- 血液型と性格の関係（A 型の人は真面目？ O 型の人は……AB 型の人は……）

（他にも，いくらでも探せば見つかると思います。）

(2) 日常生活や社会の……

これも，考えれば考えるだけ無限に出てきますが，ここでは以下の疑問に答える場合について具体化した解答例を示してみます。

「アルバイトをしている大学生の方が，していない大学生よりも，就職活動で成功しやすいのではないか？」

(a)

0.2 節にあるように，「データを用いて検証可能な問い」にすることが重要です。「就職活動で成功しやすい」という表現は曖昧なので，具体的な指標に落とし込みます。すると，例えば以下のようないくつかの形で仮説を立てることができそうです。

【解答例】

1. 大学 4 年生において，在学中にアルバイト経験がある学生のほうが，ない学生に比べて，卒業までに内定を保有している割合は高いか？
2. 大学 4 年生において，在学中のアルバイトの総労働時間が長い学生ほど，最初に内定を得るまでの期間は短い傾向にあるか？
3. アルバイトで「リーダー」などの役職を経験した学生は，そうでない学生に比べて，希望する業界への就職率が高いか？

(b)

統計的な手法を用いて「問い」に答えるためには，ある程度多くのデータを集める必要があります（どれくらい必要になるかは場合によります：考え方の一例は 9.4 節や 11.6 節にて説明）。当然ながら，自分の近くにいる数人に聞いた結果だけでは統計的に何かを論じることはできません。

そして，今回の問いの場合には，基本的にアンケートによってデータを集めることになるでしょう。問いを 1. の形式に設定した場合には，大学 4 年生を対象に「アルバイトの経験があるか」「すでに内定を持っているか」などを尋ねるアンケートを用意したら良いと考えられます。もちろん問いの形式によって質問の内容は変わります。2. の場合には「アルバイトの総労働時間」および「最初に内定を獲得した時期」などになるはずです。

【解答例】
（1. の問いを検証する場合）大学 4 年生 200 名程度に，以下の項目を尋ねるアンケートを配布し，回答を収集すると良いと考えられる。

- **アルバイト経験の有無**：「大学 1 年〜4 年の間に，継続的なアルバイト経験はありますか？」（はい / いいえ）
- **内定の保有状況**：「卒業する 3 月末の時点で，就職先の内定を 1 社以上保有していますか？」（はい / いいえ）

そして，「アルバイト経験の有無」と「内定の保有状況」だけでは，本当にアルバイトが就活に影響しているのか，正確に評価することは難しいかもしれません。なぜなら，他の要因が両方に影響を与えている可能性があるためです（これを**交絡**と言います：詳しくは 3.7 節にて）。そこで，例えば以下のようなデータも同時に集めておくと，より信頼性の高い分析ができるようになる可能性があります。ただし，ここでどのようなデータを同時に集めておくべきかはやはりケースバイケースであり，以下はほんの一例です。

- **学業成績**：GPA（Grade Point Average）など。もともと真面目な学生が，アルバイトも学業も両立し，結果として就職もうまくいっているだけかもしれません。
- **所属学部**：文系か理系か，また学部によって就職のしやすさが異なる可能性があります。
- **サークル・部活動への参加状況**：アルバイト以外の課外活動が，コミュニケーション能力などを通じて就職活動に影響しているかもしれません。
- **インターンシップへの参加経験**：インターンシップは就職活動に直接的に影響を与える可能性が高いです。
- **性別，年齢**

また問いの形式によっては，アンケート以外の方法によってデータを収集することも可能です。例えば，(1) の解答例にあげた「女性役員・管理職の比率が高い企業ほど，収益性が高い」という問いを検討するために必要な「女性役員・管理職比率」や「収益性」の指標（営業利益や ROE，PER など）は，企業にアンケートを配布しなくても，データベース等にまとめられてい

る可能性が高いでしょう。

第1章

(1) 以下の各変数について……

(a)

「従業員数」は数値でその規模を表すので**量的変数**です。従業員数が 100 人の会社と 200 人の会社では，その差（100 人）に意味があります。また，「0 人」は「従業員が一人もいない」という絶対的な意味を持つため，掛け算や割り算にも意味があります（例：「従業員数が 2 倍」）。したがって，**比率尺度**に分類されます。

(b)

「学籍番号」は，学生一人ひとりを区別・識別するための番号（ラベル）であり，数値の大小に意味はありません。例えば，学籍番号 16 番の学生が 20 番の学生より偉い，といったことはありません。したがって，単なる分類のための**質的変数**であり，**名義尺度**に分類されます。

(c)

「AAA が最も信用度が高い」というように，格付けの間には順序や大小関係が存在します。しかし，「AAA と AA の信用の差」と「AA と A の信用の差」が等しいとは限りません。つまり，間隔は等しくありません。したがって，順序にのみ意味がある**質的変数**であり，**順序尺度**に分類されます。

(d)

「IQ」は知能の高さを数値で表すので**量的変数**です。IQ 100 と 110 の差 (10) と，IQ 110 と 120 の差 (10) は，等しい差であると考えられています。しかし，IQ が 0 は「知能が全くない」状態を意味するわけではなく，絶対的な原点ではありません。そのため，「IQ 120 の人は IQ 60 の人より 2 倍賢い」というような，掛け算・割り算を用いた解釈はできません。したがって，**間隔尺度**に分類されます。

【解答】

(a) 量的変数・比率尺度
(b) 質的変数・名義尺度
(c) 量的変数・順序尺度
(d) 量的変数・間隔尺度

(2) 問 1 で分類した各変数について……

(a)

「企業コード」は，学籍番号と同様に各企業を識別するための番号であり，**名義尺度**の変数です。名義尺度の数値には大小関係や優劣の意味は全くありません。したがって，コードの数字が小さいからといって，その会社が優れていると判断することはできません。

(b)

「信用格付け」は**順序尺度**の変数です。順序尺度では，「AA の方が BBB より信用度が高い」という大小関係は分かりますが，その「差」がどれくらいなのか，またそれが「何倍」なのかといった比率を議論することはできません。「2 番目」と「4 番目」という順序情報から，「2 倍のリスク」という比率の情報を導き出すことは，順序尺度の性質を超えた誤った解釈です。

(c)

「IQ」は**間隔尺度**の変数です。間隔尺度では，数値の差には意味があります（この場合，B さんは A さんより IQ が 20 高い）。しかし，絶対的な原点（0 が「無」を意味する点）が存在しないため，比率（掛け算・割り算）を計算することに意味はありません。「1.2 倍頭が良い」という解釈は，IQ が比率尺度であるかのような誤った使い方です。

【解答】

(a)　不適切である。 企業コードは名義尺度であり，数値の大小に意味はないため。
(b)　不適切である。 信用格付けは順序尺度であり，間隔や比率に意味はないため。
(c)　不適切である。 IQ は間隔尺度であり，比率で解釈することはできないため。

(3) 1.1.2 項で例として挙げた「学力……

大半のテストでは，「50 点から 60 点に上げる努力」と「90 点から 100 点に上げる努力」は質的に異なり，同じ 10 点でもその間の「学力の差」が等しいとは言えない，と考えられます。この観点に立てば，「学力テストの得点」は厳密には**順序尺度**と捉えるべきかもしれません。

順序尺度として解釈した場合に生じる問題点

一方で，もし「学力テストの得点」を厳密に順序尺度として解釈すると，私たちが普段当たり前のように使っている「平均点」を計算することすら統計学的に不適切となり，計算できる統計量は中央値などに限られます。そのため，一般的な分析手法の多くが使えなくなり（あるいは適用が難しくなり），データから引き出せる情報が少なくなってしまう可能性があります。

比率尺度として解釈した場合に生じる問題点

一方，「学力テストの得点」を比率尺度として解釈することには明確な問題があります。比率尺度は，0 が「全くない」という絶対的な原点を意味する必要がありますが，テストで 0 点を取ることが，その人に能力が「全くない（ゼロである）」ことを意味するようなテストはほとんど存在しません。そのため，「100 点を取った学生は，50 点を取った学生の 2 倍の学力がある」というような比を用いた解釈はできないと言えます。

まとめると，テストの得点が一般的に**間隔尺度**として扱われる主な理由は，**実用上の利便性が非常に高いため**です。もし「得点」を間隔尺度とみなすことが許されるならば，平均値や標準偏差を計算できるだけでなく，（11 章で紹介する t 検定など）より高度で強力な統計分析手法を用いることが可能になります。このように，間隔尺度と「みなす」ことで，データ分析の幅が格段に広がり，教育評価や指導改善に役立つ多くの情報を得ることができます。厳密性を少し犠牲にしてでも，得られる実用的なメリットが大きいため，慣習的に間隔尺度として扱われています。

【解答例】

- **間隔尺度として扱われる理由**：平均値や標準偏差が計算可能になり，t 検定などの強力な分析手法が使えるため，実用上のメリットが非常に大きいから。
- **順序尺度とした場合の問題点**：平均値が計算できず，分析手法が限定されるため，データから得られる情報が少なくなる。
- **比率尺度とした場合の問題点**：「0 点」が「学力が全くない」という絶対的な原点を意味しないため，「点数が 2 倍だから学力も 2 倍」のような比率による解釈ができない。

第2章

(1) 図 2.6 のグラフでは……

表現上の工夫の指摘

このグラフは，「体育が圧倒的に人気である」という印象を過度に強調するため，以下のような視覚的な工夫（トリック）を用いています。

3D（立体）表現と遠近法　グラフを立体的にし，手前にある「体育」の棒が，奥にある他の教科の棒よりも大きく見えるように描いています。これにより，実際の数値の差以上に，「体育」が大きく見え，圧倒的な印象を与えます。

視点（アングル）　グラフを斜め下から見上げるような視点で描くことにより，手前にある「体育」の高さがより強調されています。

グラフの始点　始点を 0 ではなく 10 にすることで，棒の長さどうしの倍率が実際よりも大きくなるようにしています。具体的に，元のデータでは体育は国語の 3 倍の人数（42 人 vs 14 人）なのですが，始点を 10 にすることで，棒の長さは 8 倍（32:4）となります。

棒の色　一般的に，白などの明るい色は膨張色とされます。したがって，体育だけが白い棒で表されることで，実際よりも大きく見えている可能性があります。ただこれは，体育の人気を「過度に」強調していると言えるほどかは分かりません。伝えたいメッセージ次第では妥当な装飾の範疇とも考えられます。

これらの表現は，見る人に正しいデータの比率を伝えることを妨げ，特定のメッセージを意図的に植え付けようとするものであり，誠実なデータ表現とは言えない可能性が高いでしょう。

客観的なグラフへの修正案

恣意的なメッセージ性を排し，データを客観的に伝えるためには，以下のような修正を行うべきです。

【解答例】

2D（平面）の棒グラフを用いる　3D 表現は視覚的な歪みを生むため，棒の高さ（長さ）が正確に量に対応する，シンプルな 2D の棒グラフにします。

縦軸の目盛りは 0 から始める　棒グラフやヒストグラムの縦軸（度数を表す軸）は，必ず 0 から始めます。これにより，量の比率が正しく表現されます。

色を調整する　ある科目に重点を置いた主張をすることが目的ではないならば（単に科目間のフェアな比較をしたいだけであれば），すべての棒の色は同じにしておくと良いでしょう。

(2) あるクラスの 8 人の……

(a)

平均値 (Mean)　全てのデータを合計し，データの個数で割ります。

$$\text{平均値} = \frac{85+72+68+92+68+83+75+68}{8} = \frac{611}{8} = 76.375$$

中央値 (Median)　まず，見やすいようにデータを小さい順に並べ替えます。

$$68, 68, 68, 72, 75, 83, 85, 92$$

今回はデータ数が 8 個（偶数）なので，中央に位置する 2 つの値（4 番目の 72 と 5 番目の 75）の平均値が中央値となります。

$$\text{中央値} = \frac{72+75}{2} = 73.5$$

最頻値 (Mode)　最も頻繁に出現する値です。「68」が 3 回出現しているため，最頻値は 68 です。

(b)

このデータセットに対する代表値としては，平均値または中央値が最も適切であると考えられます。

平均値が適切といえる理由　平均値はデータ内のすべてのケースの値を用いているため，多くの場合で代表値としてふさわしい可能性が高いでしょう。弱点は「外れ値の影響を受けやすい」ことですが，今回の 8 人には明らかな外れ値として注意すべき程のものはないと言えそうです（ケースバイケースですが）。

中央値が適切といえる理由　各個人の点数は比較的安定している一方で，一人だけ 92 点という高得点を獲得しています。このため，平均値はやや高めになっているという見方ができます。そこで，外れ値に頑健な中央値を採用することにも正当性があると言えるのです。

一方で，最頻値が適切であると考える根拠は薄いと考えられます。というのも，68 点の人が 3 人いたことは「たまたま」である側面が強く，ぴったり 68 であることに強い意味がない（67 でも 69 でもだいたい同じ）ためです。そのような状況で，たまたま 8 人の中の最低点が最も多かったことをもって「68」を代表値とすることにはあまり意味がないでしょう。

また，平均値と中央値ではどちらが良いか，という点はケースバイケースです。記述統計量を算出する目的や，外れ値に対する考え方等でも変わってきますが，今回の場合には平均値でも中央値でも基本的にはどちらでも問題はないと言えるでしょう。重要なのは，**なぜその代表値を使うことにしたのかを，きちんと説明できる**ことにあります。

(c)

分散は「各データの平均値からの差（偏差）の 2 乗」の平均値です。そこで，平均値（76.375 点）からの偏差およびその 2 乗をそれぞれ求めると以下のようになります（ただし偏差の 2 乗は細かすぎるので小数点第 3 位で四捨五入した値です）。

	1 番目	2 番目	3 番目	4 番目	5 番目	6 番目	7 番目	8 番目
もとの値 (x_i)	85	72	68	92	68	83	75	68
偏差 $(x_i - \bar{x})$	8.625	-4.375	-8.375	15.625	-8.375	6.625	-1.375	-8.375
偏差の 2 乗 $(x_i - \bar{x})^2$	74.39	19.14	70.14	244.14	70.14	43.89	1.89	70.14

分散　偏差平方和（偏差の 2 乗の合計）をデータの個数 (8) で割ります。

$$\sigma^2 = \frac{74.39 + 10.14 + 70.14 + 244.14 + 70.14 + 43.89 + 1.89 + 70.14}{8} = \frac{593.87}{8} \approx 74.23$$

標準偏差：分散の正の平方根をとります。

標準偏差 $(\sigma) = \sqrt{74.23} \approx 8.62$

【解答例】

(a)　平均値: 76.375 点，　中央値: 73.5 点，　最頻値: 68 点

(b)　**平均値**または**中央値**が最も適切と考えられる。

（平均値が良いと考えられる理由） データ全体の情報をきちんと用いて計算された代表値であり，データ内に明らかな外れ値も見られないため。

（中央値が良いと考えられる理由） (92 点を) 外れ値（とするならば，そ）の影響を受けにくく，データ全体の中心を安定して示せるため。

(c)　分散: 約 74.23，　標準偏差: 約 8.62 点

(3) 以下の問いに答えてください。……

(a)

データの個数が $n = 6$ で平均値が 10 であるという事実からは，**6 個の合計が** $6 \cdot 10 = 60$ **であ**るということが分かります。そして，今わかっている 5 個の値の合計が $3+7+9+12+16 = 47$ であることから，残りの 1 個は $60 - 47 = 13$ となります。

(b)

先ほどと同様に平均値の定義から，5 個の値の合計が $8 \cdot 5 = 40$ であると分かります。また，分散が 6 であるということからは，5 個の値について**平均値からの偏差の 2 乗の合計が**

$6 \cdot 5 = 30$ であると分かります。

ここで，残りの 2 個のデータの値を a, b として考えてみます。まず，平均値の情報から式を立てると，今わかっている 3 個の値の合計が $4 + 8 + 10 = 22$ であることから，$a + b = 18$（式 1）であると分かります。

次に，分散の情報から式を立てます。まず，偏差の 2 乗の合計を求める過程をまとめると，以下のようになります。

	1 番目	2 番目	3 番目	4 番目	5 番目
もとの値 (x_i)	4	8	10	a	b
偏差 $(x_i - 8)$	-4	0	2	$a - 8$	$b - 8$
偏差の 2 乗 $(x_i - \bar{x})^2$	16	0	4	$(a-8)^2$	$(b-8)^2$

ここから，偏差の 2 乗和の合計が 30 であるため，

$$16 + 0 + 4 + (a-8)^2 + (b-8)^2 = 30$$
$$\longrightarrow (a-8)^2 + (b-8)^2 = 10 \quad \text{（式 2）}$$

となります。

あとは，今立てた（式 1）と（式 2）を連立方程式として解くと，$(a, b) = (11, 7)$ であると分かります（順不同）。

【解答】

(a) 13
(b) 7 と 11

(4) ある変数の全体的な特徴を……

代表値と散布度だけでは不十分な理由は，**それらの要約統計量が同じであっても，元となるデータの分布形状が全く異なっている可能性があるから**です。わかりやすい例として，2.4 節の冒頭には，「平均値と中央値が同じ」いくつかの分布が載っています。これらは，「代表値だけで分布の特徴を把握するのは難しい」ケースの例です。同様に，図 2.21 や図 2.22 は，（適切なスケールのもとでは）「平均値と分散が同じ」分布の例となります。もちろん，同じように考えると「平均値と分散と歪度と尖度が同じ」異なる分布を考えることも理論上は可能です（図 1）。

（先取り）第 3 章では 2 変数の関係について考えますが，ここでも記述統計量だけでは変数間の関係性の全体的な傾向を掴むことはできません。その有名な例として，各変数の平均値・標準偏差および 2 変数間の相関係数がすべて同じになる様々な散布図の例を示した「アンスコムのカルテット」や「データサウルス」などがあります。興味があれば見てみてください。

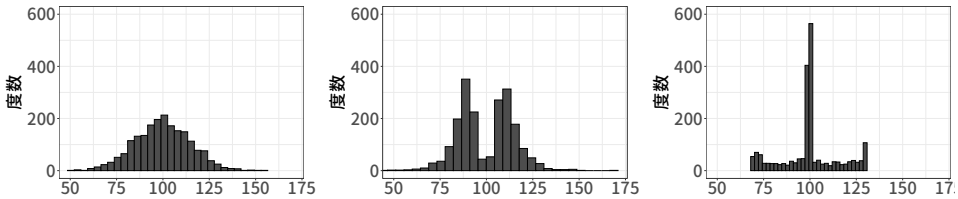

図 1　平均値・分散・歪度・尖度がいずれも同じヒストグラム

【解答例】平均値や分散などの記述統計量が全く同じでも，データの分布形状（外れ値の有無，関係性の形など）が大きく異なる場合があるため。データを可視化することで，要約統計量だけでは見えないデータの全体像を把握できるから。

(5) ある企業の 2 つの部署で…

単純な平均値の比較だけで「部署 B の方が充実度が高い」と結論付けられない理由は，主に以下の 2 点が考えられます。

1. **データのばらつき（標準偏差）が大きく異なる点**：部署 A は標準偏差が 0.5 点と小さく，従業員の回答が平均値 3.5 点の周りに密集している（多くの従業員が似たような満足度）ことを示唆しています。一方，部署 B は標準偏差が 1.0 点と大きく，回答が広く散らばっていることを意味します。これは満足している人としていない人がどちらもいることを意味しています。平均的に満足度が高いとしても，部署内に深刻な不満を持つ従業員が多数いるとしたら，一概に「良い結果」とは言えません。
2. **平均値の差が統計的に意味のある差とは限らない点**：観測された平均値の差 (0.5 点) は，調査対象者がこの 30 人だったから偶然生じた「誤差」の範囲内かもしれません。この差が単なる偶然なのか，それとも部署間に本当に意味のある（**統計的に有意な**）差があるのかを判断するためには，統計的仮説検定（第 11 章で紹介する t 検定など）を行う必要があります。観測された結果（0.5 点の差）だけをもって，そこに本質的な意味があると断定することは危険です（仮説検定したら万事解決というわけでもありませんが）。

【解答例】

1. **ばらつきの違い**：部署 B は標準偏差が大きく，満足している人と不満な人が混在している両極端な状況である可能性があり，平均値だけでは実態を判断できない。
2. **統計的有意性**：観測された平均値の差が，偶然による誤差なのか，本当に意味のある差なのかは統計的検定などを行わないと判断できない。

(6) 2 つの全国模試……

【解説】

(a)

- **X 模試**：平均点が高く標準偏差が小さい。これは，問題が比較的易しく，多くの受験者が高得点帯に集中したため，**得点のばらつきが小さく，差がつきにくい模試**であったことを示唆します。
- **Y 模試**：平均点が低く標準偏差が大きい。これは，問題の難易度が高く，学力に応じて点差が大きく開いたため，**得点のばらつきが大きく，差がつきやすい模試**であったことを示唆します。

(b)

標準化得点 $z = \dfrac{\text{得点} - \text{平均点}}{\text{標準偏差}}$，偏差値 $= 10 \cdot z + 50$ を用いて計算します。

- **X 模試**：

$$z = \frac{75 - 70}{8} = 0.625$$

$$\text{偏差値} = 10 \cdot 0.625 + 50 = 56.25$$

- **Y 模試**：

$$z = \frac{75 - 60}{15} = 1.0$$

$$\text{偏差値} = 10 \cdot 1.0 + 50 = 60.0$$

(c)

同じ 75 点でも，偏差値は Y 模試の方が高く出ています。これは，A さんの学力が，それぞれの模試の受験者集団の中で占める**相対的な位置**が異なることを意味します。大学の合否は受験生の中での相対的な順位で決まるため，合格可能性を判断する際には以下の点に注意すべきです。

異なる模試の比較には偏差値を用いる　平均点やばらつきが異なるため，素点での比較は無意味です。同じ母体が実施する模試であれば，基本的に各回の難易度は概ね同じになることを目指して調整されているかもしれませんが，実際の難易度（あるいは受験者のレベル）が同じであるかはまた別の話なのです。そのため，偏差値という共通の物差しで評価する必要があります。

偏差値の変化と学力の変化は別の話　偏差値はあくまでも相対的なものなので，偏差値の上がり幅によって「学力がどの程度上昇したか」を評価するのは難しいでしょう。

模試の受験者層を考慮する　自分の志望校のレベルと，その模試の受験者層がどれだけ一致しているかが重要です。難関大学を目指すなら，ハイレベルな受験者が多い模試（差がつきやすい Y 模試のような）での偏差値の方が，より信頼できる指標となります。

【解答例】

(a) X 模試は比較的易しくばらつきが小さい「差がつきにくい」模試。Y 模試は平均点が低くばらつきが大きい「差がつきやすい」模試。
(b) X 模試: 標準化得点 0.625, 偏差値 56.25。 Y 模試: 標準化得点 1.0, 偏差値 60.0。
(c) 異なるテストの成績を比較する際は，素点ではなく必ず偏差値を用いるべき。また，その偏差値はあくまでも相対的な指標であることに注意し，どのような受験者集団の中での相対的な位置を示しているのか（模試のレベル）を考慮することが重要。

(7) 以下の 3 つのヒストグラム……

ある変数のヒストグラムの形状を想像することは，特に統計的推測の準備段階で非常に重要となります。もちろん最終的には，実際に観測されたデータをもとにヒストグラムを作成してみることが必須なのですが，様々な変数について「**データを集めてみたらこんな形になりそうな気がする**」という直感力を養うと，データ分析においてはかなり役に立つでしょう。以下は，筆者が考えたロジックに過ぎません。もしかしたら，別の考え方もあるかもしれないので，「こういう考え方はどうですか？」など思いついたら，何らかの方法で筆者にお知らせいただけると幸いです。

ヒストグラム A　(2) ある大学の入学試験の点数
理由：入学試験の得点は，学力を反映したものです。一般的に学力をはじめとした「人間の特性」のようなものは，A のように**左右対称の山の形をしている**ことが多いとされています。すなわち，平均値くらいの人が多く，平均値から離れるほどその割合は減っていくでしょう。

ヒストグラム B　(3) 成人男性の年間所得
理由：所得の分布は，多くの人が比較的低い所得層に集中している一方で，ごく一部の人は非常に高い所得を得ているでしょう。したがって，**左側に山があり，右側に長く裾を引く「右に歪んだ」分布**になる B の分布のような形になると考えられます。

ヒストグラム C　(1) 駅前のコンビニの 1 時間ごとの来店客数
理由：駅前のコンビニの来店客数は，（駅によりますが）通勤ラッシュ・帰宅ラッシュ時には多くの人が利用すると考えられる一方で，昼過ぎや終電後の利用者は相対的にかなり少ないと考えられます。したがって，全体としては C のように**山が 2 つある「二峰性」の分布**になるかもしれません。

【解答例】

ヒストグラム A　→ (2) ある大学の入学試験の点数：（理由）学力などの人間の特性の多く

は，左右対称の山のような分布になると考えられるため。

ヒストグラム B → (3) 成人男性の年間所得：（理由）所得は多くの人が一定の範囲内にある一方で，ごく少数の人達が圧倒的に高く，分布としては左側にピークがあり右に長く裾を引くような形になると考えられるため。

ヒストグラム C → (1) 駅前のコンビニの 1 時間ごとの来店客数：（理由）ピーク時と深夜など，時間帯によって二極化すると考えられるため。

第 3 章

(1) 以下の 2 変数について……

学習時間を変数 x, テスト得点を変数 y とします。データ数は $n = 9$ です。まず，各変数の平均値をそれぞれ計算すると，

$$\bar{x} = \frac{4+1+5+2+3+5+1+2+4}{9} = \frac{27}{9} = 3 \tag{1}$$

$$\bar{y} = \frac{66+53+97+32+64+80+42+82+78}{9} = \frac{594}{9} = 66 \tag{2}$$

となります。

これを用いて，各変数の偏差および偏差の積をそれぞれ計算した結果が以下の表です。

学習時間（x）	4	1	5	2	3	5	1	2	4
$x - \bar{x}$	1	−2	2	−1	0	2	−2	−1	1
$(x - \bar{x})^2$	1	4	4	1	0	4	4	1	1
テスト得点（y）	66	53	97	32	64	80	42	82	78
$y - \bar{y}$	0	−13	31	−34	−2	14	−24	16	12
$(y - \bar{y})^2$	0	169	961	1156	4	196	576	256	144
$(x - \bar{x})(y - \bar{y})$	0	26	62	34	0	28	48	−16	12

あとは，この結果から相関係数を計算するだけです。そのために，まずは各変数の標準偏差と共分散をそれぞれ計算してみます。

$$\begin{aligned}
&(x \text{ の標準偏差}) \quad s_x = \sqrt{\frac{1+4+4+1+0+4+4+1+1}{9}} = \sqrt{\frac{20}{9}} \approx 1.49 \\
&(y \text{ の標準偏差}) \quad s_y = \sqrt{\frac{0+169+961+1156+4+196+576+256+144}{9}} \\
&\qquad\qquad = \sqrt{\frac{3462}{9}} \approx 19.61 \\
&(\text{共分散}) \quad s_{xy} = \frac{0+26+62+34+0+28+48-16+12}{9} = \frac{194}{9} \approx 21.56
\end{aligned} \tag{3}$$

以上より，相関係数は，

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{21.56}{1.49 \cdot 19.61} \approx 0.738 \tag{4}$$

となります。

【解答】相関係数 $r \approx 0.738$

(2) ある企業の 3 つの部署……

相関比 η^2 は，

$$\eta^2 = \frac{\mathrm{SS}^2_{\text{群間}}}{\mathrm{SS}^2_x} = \frac{\mathrm{SS}^2_{\text{群間}}}{\mathrm{SS}^2_{\text{群間}} + \mathrm{SS}^2_{\text{群内}}} \tag{3.6}$$

と計算します。そして，上の式中に登場する SS は偏差平方和[1]，すなわち**分散の計算時に，サンプルサイズで割る前の値**を表していました。

以上を踏まえて，必要な要素をそれぞれ順に計算していきましょう。

$\mathrm{SS}^2_{\text{群間}}$ の計算　これは，「全体平均」と「各群の平均値」の差の二乗和でした。各部署の平均値は表に記載されているため，あとは「全体平均」があれば計算可能です。

全体平均 $\bar{x}$ の計算　3 つの部署のサンプルサイズ（回答数）の合計は $n = 30+40+35 = 105$ 人です。そして全体平均は，各部署の平均値をサンプルサイズで重み付けした

$$\bar{x} = \frac{(30 \cdot 65) + (40 \cdot 75) + (35 \cdot 85)}{30 + 40 + 35} = \frac{7925}{105} \approx 75.48 \tag{5}$$

と求められます。

$\mathrm{SS}^2_{\text{群間}}$ の計算　分散の計算時と同様に，「全体平均」からの偏差の二乗和を求めたら良いので，

$$\begin{aligned} \mathrm{SS}^2_{\text{群間}} &= \underbrace{30 \cdot (65 - 75.48)^2}_{\text{A 部署のぶん}} + \underbrace{40 \cdot (75 - 75.48)^2}_{\text{B 部署のぶん}} + \underbrace{35 \cdot (85 - 75.48)^2}_{\text{C 部署のぶん}} \\ &\approx 3294.9 + 9.2 + 3172.1 \\ &= 6476.2 \end{aligned} \tag{6}$$

となります。

$\mathrm{SS}^2_{\text{群内}}$ の計算　これは，単純に各群の分散の「サンプルサイズで割る前の値」を足したら良いの

[1] 本文では SS^2 と表していますが，SS（Sum of Square）自体に「二乗」という意味が含まれているため，少し変な表現になってしまっていることに気づきました。いつか直せたら直します。

で，

$$\mathrm{SS}^2_{\text{群内}} = \underbrace{30 \cdot 100}_{\text{A 部署のぶん}} + \underbrace{40 \cdot 64}_{\text{B 部署のぶん}} + \underbrace{35 \cdot 81}_{\text{C 部署のぶん}} = 8395 \tag{7}$$

となります。

相関比の計算　あとはそれぞれの値を式に入れると，$\eta^2 = \dfrac{6476.2}{6476.2 + 8395} \approx 0.436$ と求めることができます。

【解答】相関比 $\eta^2 \approx 0.436$

(3) 以下は，ある高校で……

クラメールの連関係数 V の計算の手順は以下のとおりです。

1. **連関が無いとした場合の期待クロス表を作る**　今回は，「運動部所属」と「運動部非所属」が 1:1 になっているため，各列を半分に分割したら良いと分かります。

学業成績	1	2	3	4	5	合計
運動部所属	7.5	17.5	37.5	25	12.5	100
運動部非所属	7.5	17.5	37.5	25	12.5	100
合計	15	35	75	50	25	200

2. **期待クロス表との偏差を取る**　期待クロス表との偏差が大きいほど，連関が大きいと見なすことができます。まずは単純に偏差を計算しましょう。

学業成績	1	2	3	4	5
運動部所属	−2.5	−2.5	2.5	0	2.5
運動部非所属	2.5	2.5	−2.5	0	−2.5

3. **偏差を調整する**　各セルについて，$\dfrac{(\text{観測度数} - \text{期待度数})^2}{\text{期待度数}}$ をそれぞれ計算していきます。

例えば左上の「運動部 · 成績 1」のセルの場合には，$\dfrac{(-2.5)^2}{7.5} \approx 0.833$ となります。同様にして，各セルで値を計算すると，

学業成績	1	2	3	4	5
運動部所属	0.833	0.357	0.167	0	0.5
運動部非所属	0.833	0.357	0.167	0	0.5

となり，これをすべて足すと，$\chi^2 \approx 3.714$ となります。

4. χ^2 値を標準化する　V は，

$$V = \sqrt{\frac{\chi^2}{n \cdot (\text{行と列のカテゴリ数の少ない方} - 1)}} \tag{3.7}$$

と定義される値です。今回の場合，カテゴリ数はそれぞれ 2 と 5 なので，

$$V = \sqrt{\frac{3.714}{200 \cdot (2-1)}} = \sqrt{\frac{3.714}{200}} \approx \sqrt{0.01857} \approx 0.136$$

と求められます。

【解答】クラメールの連関係数 $V \approx 0.136$

(4) 以下に示したケースが発生しうる……

(a)

因果関係が存在しないにも関わらず，相関係数の値が大きくなる代表的な状況は，**疑似相関**がある場合や，**ただの偶然**があります。

疑似相関の具体例　「年間のモバイルバッテリーの売上」と「年間の SNS の投稿数」。

説明　単位を「年」として考えると，これらの背後にはどちらも「スマートフォンの普及」という第 3 の変数を考えることができそうです。SNS の利用頻度が増えたことが原因で（充電が不足するようになり）モバイルバッテリーを購入した，と考えることもできるかもしれませんが，直接的な因果関係というよりは一つの要因に過ぎないでしょう。仮に SNS の投稿をやめ（させ）たとしても，いずれにせよきっと動画を見たりなどの理由でモバイルバッテリーは必要になるはずです。

ただの偶然の具体例　これも色々とあるのですが，https://www.tylervigen.com/spurious-correlations などに色々と掲載されています。本当に偶然に過ぎないのか，あるいはまだ見ぬ第 3 の変数が隠れているのかは分かりません……

(b)

因果関係が想定されるにも関わらず，相関係数がゼロに近くなる代表的な状況は，**非線形の相関関係**がある場合が考えられます。相関係数は，あくまでも直線的な関係を測る指標であるために，増加部分と減少部分が打ち消し合うと，結果として相関係数はほぼゼロになることがあります。

U 字の相関の具体例　「ゲームの難易度」と「プレイヤーの離脱率」。

説明　ゲームは，簡単すぎる場合にはすぐ飽きてしまって離脱する可能性が高い一方で，難しすぎる場合にも諦めてしまうと考えられます。ゲームに限った話ではないかもしれませんが，タスクの難易度はその人にとってちょうど良いほどやる気を引き出させるでしょう。

逆 U 字の相関の具体例　「肥料の量」と「作物の収穫量」

説明　肥料を増やすと収穫量は増えますが，ある点を超えて与えすぎると逆に収穫量は減ってしまいます。ヒトに投与する薬などとおなじように，「過ぎたるは及ばざるが如し」なものについて相関係数を計算すると，相関関係を正しく反映しないものとなってしまうかもしれません。

(c)

相関関係も連関関係も，本質的には同じ**2 変数の共変関係**を表しているに過ぎません。そのため，相関係数のときと同じように**疑似（無）相関**とも呼べるような状況を引き起こす第 3 の変数が存在している可能性が考えられるでしょう。

具体例　「新しい AI ツールの導入有無」と「納期の遵守」。

説明　一般的には，新しい AI ツールを導入することで業務は効率化され，結果的に納期を遵守できる割合は改善すると考えられます。しかし，もしもある部署では AI あるいは IT に不慣れな場合には，むしろ新しいツールに慣れないために納期の遵守率が低下してしまうかもしれません。このように，第 3 の変数（この例では「IT リテラシー」）で層ごとに分けたときに，正の効果がみられる層と負の効果が見られる層がある場合には，これらを合わせて連関係数を計算すると，効果が打ち消し合ってゼロに近い値になってしまう可能性があります（シンプソンのパラドックスも参照）。

【解答例】

(a)**疑似相関**のケース。例：「年間のモバイルバッテリーの売上」と「年間の SNS の投稿数」。第 3 の変数「スマートフォンの普及率」が両方に影響している。
(b)**非線形の関係**のケース。例：「肥料の量」と「作物の収穫量」。（逆）U 字型の関係では，直線的な相関は 0 に近くなる。
(c)**第 3 の変数**によって効果が変化するケース。例：「新しい AI ツールの導入有無」と「納期の遵守」。層ごとに分析した結果と，全体を統合した結果は大きく異なることがある。

(5) あるデータ分析で，「アプリの……

結論付けてしまうことの問題点

「相関関係は因果関係を意味しない」ため，この結論は早計です。観測された負の相関は，少なくとも以下の 2 つの別の可能性を排除できていません。

因果関係が逆 「睡眠時間が短い」ことが原因で，「アプリの使用時間」が結果となっている可能性（例：寝付けないから暇つぶしにアプリを見てしまう）。

第三の変数（交絡因子）の存在 「強いストレス」のような第三の変数が原因で，「気分転換のためのアプリ使用時間が増え」，同時に「ストレスで睡眠時間が短くなる」という両方を引き起こしている可能性（疑似相関）。

必要な追加の調査や分析の提案

因果関係を示すための J.S. ミルの 3 つの条件を考えると，より信頼性の高い因果関係の結論を導くためには，例えば以下のようなアプローチが考えられます。

1. **介入研究（ランダム化比較試験，RCT）** 第 3 の変数の影響を取り除くための方法としてゴールドスタンダードとされているのが，対象者をランダムに群に分け，異なる介入を行う，というものです。今回の例では，対象者をランダムに 2 群に分け，一方には「アプリ使用時間を減らす」ような介入を行う[2]一方で，もう一方には何もしない，というデザインが考えられます。参加者を十分に多く集めたうえで，完全ランダムに群分けができれば，理論上は「アプリ利用時間」以外の「第 3 の変数になりうるすべての変数」については均等に割り振られていることになるため，単純に両群の睡眠時間を比較することで因果関係により近づけると考えられます。

縦断的調査（パネルデータ分析） 個人のスマートフォンに対して，アプリの利用時間を制限するような介入を行うことはあまり現実的ではないかもしれません。このような場合には，観察されたデータをもとに，因果効果を推定する必要が出てきます（統計的因果推論）。因果関係を示すためには，「原因変数が結果変数に先行している」ようなデータを集める必要があるため，例えば同じ対象者を長期間追跡調査し，個人内でのアプリ使用時間の変化と，**その後の**睡眠時間の変化の関係を見ると良いかもしれません。ただしこの場合には，以下に説明するように，第 3 の変数となりうる交絡要因をきちんと統制することが重要となります。

交絡因子の統制 縦断的なデータの取得が困難（1 時点でしか取れない）場合には，因果関係の検証は更に難しくなります。まずは，ストレスレベル，年齢，職業など，睡眠時間に影響を与えそうな他の変数のデータも収集し，重回帰分析などの手法でそれらの影響を取り除いた上で，アプリ使用時間と睡眠時間の純粋な関係性（相関関係）を分析するのが良いかもしれません。そもそもこの 2 つの変数が無関係であるならば，因果関係も無いと言えそうです。

ほとんどの場合において，因果関係を検証する最良の方法は，RCT の実施だと考えられます。

[2] 強制的にアプリ使用時間を減らさせる介入としては，例えば「1 日 2 時間以上アプリを開いたら自動的に電源が落ちるように仕掛ける」などがあるかもしれません。倫理的に許されるかは別の話ですが……

ただし，現実的にはそのような介入は不可能であることも多い（国や自治体レベルの分析など）ため，「どのようなデータが取れるならば，どのような手法によって因果関係を評価できるか」を知っておくことで，最適な分析デザインが提案できるようになるでしょう。

【解答】

- **問題点**：相関関係は因果関係を意味しない。「逆の因果関係」や「第三の変数（交絡因子）」の可能性を排除できないため。
- **追加提案**：因果関係を検証するために，(1) 介入研究 (RCT) の実施，(2) 縦断的調査，(3) 交絡因子を統制した多変量解析などが必要となる。

第 4 章

(1) ある会社の月間売上高……

(a) 回帰式の解釈

切片の意味　切片の「1000」は，広告費と従業員数が両方とも 0 のときの売上高の予測値です。つまり，広告を全く出さず，従業員も 0 人だったとしても，月間 1000 万円の売上が予測されることを意味します。ただし，これはモデル上の計算値であり，現実的な状況から外れた予測である点には注意が必要です。今回のケースでは，「回帰直線の当てはまりを良くするための調整役」以上の意味は無いと考えても差し支えないでしょう[3]。

広告費の効果　広告費の係数「2.5」は，**従業員数が一定であれば**，広告費を 1 万円増やすと，売上高が平均して 2.5 万円増加することを示します。一般的には，広告費が大きい企業ほど従業員数も多い傾向にあるため，2 つの説明変数間の正の相関関係を統制した効果の大きさが「2.5」と表れています。ただしこの係数に関しては，売上が際限なく大きくなっていくとは考えられません。あくまでも**回帰係数の算出に用いたデータにある範囲においては**広告費 1 万円あたり 2.5 万円の効果がある，という点には注意が必要です。

従業員数の効果　従業員数の係数「30」は，広告費が一定であれば，従業員を 1 人増やすと，売上高が平均して 30 万円増加することを示します。ただしこれも，むやみに従業員数を増やしまくって良い，という話ではありません。

モデルの全体的な予測精度　決定係数 $R^2 = 0.75$ は，このモデルが売上高のばらつきの 75% を，広告費と従業員数で説明できていることを意味します。1 に近いほど予測精度が高いとされ，0.75 は比較的高く，このモデルがある程度の予測精度を持つことを示唆します。

[3] ただし，説明変数がいずれも中心化されている場合には，切片は「平均的な広告費で，平均的な従業員数のときの売上高の予測値」となるため，これ自体が意味を持つようになります（4.2 節の**【コラム】 切片に意味を持たせるためには**を参照）

(b) 予測売上高の計算

回帰式に，広告費 $= 200$，従業員数 $= 25$ を代入して計算するだけです。

$$\begin{aligned}\widehat{\text{売上高}} &= 1000 + 2.5 \cdot 200 + 30 \cdot 25 \\ &= 1000 + 500 + 750 \\ &= 2250\end{aligned}$$

【解答】

(a)**切片**: 広告費・従業員数ともに 0 の場合，1000 万円の売上が予測される。
広告費の効果: 1 万円増やすと売上が 2.5 万円増加。
従業員数の効果: 1 人増やすと売上が 30 万円増加。
予測精度: 売上高のばらつきの 75% を説明できている。
(b)2250 万円

(2) 以下のデータに対して……

(a)

回帰直線を $\hat{y} = \beta_0 + \beta_1 x$ とします。最小二乗法では，実測値 y と予測値 $\hat{y}$ の差の二乗和が最小になる β_0, β_1 の値を求めるため，順に計算していきましょう。

データ番号	x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
1	1	2	$\beta_0 + \beta_1$	$2 - \beta_0 - \beta_1$	$\beta_0^2 + \beta_1^2 + 2\beta_0\beta_1 - 4\beta_0 - 4\beta_1 + 4$
2	2	3	$\beta_0 + 2\beta_1$	$3 - \beta_0 - 2\beta_1$	$\beta_0^2 + 4\beta_1^2 + 4\beta_0\beta_1 - 6\beta_0 - 12\beta_1 + 9$
3	3	5	$\beta_0 + 3\beta_1$	$5 - \beta_0 - 3\beta_1$	$\beta_0^2 + 9\beta_1^2 + 6\beta_0\beta_1 - 10\beta_0 - 30\beta_1 + 25$
4	4	4	$\beta_0 + 4\beta_1$	$4 - \beta_0 - 4\beta_1$	$\beta_0^2 + 16\beta_1^2 + 8\beta_0\beta_1 - 8\beta_0 - 32\beta_1 + 16$

以上について，$(y - \hat{y})^2$ をすべて足すと，$4\beta_0^2 + 30\beta_1^2 + 20\beta_0\beta_1 - 28\beta_0 - 78\beta_1 + 54$ となります。

(b)

偏微分では，一方の変数を定数とみなした微分をそれぞれ行っていきます。まず，β_1 について偏微分を行う場合，β_0 を定数とみなして，β_1 の次数の順に式を整理すると，

$$\underbrace{30\beta_1^2}_{\text{2 次の項}} + \underbrace{(20\beta_0 - 78)\beta_1}_{\text{1 次の項}} + \underbrace{(4\beta_0^2 - 28\beta_0 + 54)}_{\text{定数項}}$$

となります。この式を β_1 について微分すると，

$$60\beta_1 + (20\beta_0 - 78)$$

という式が得られます。

同様に，β_0 について偏微分を行う場合には，β_1 を定数とみなせばよいので，

$$\underbrace{4\beta_0^2}_{\text{2 次の項}} + \underbrace{(20\beta_1 - 28)\beta_1}_{\text{1 次の項}} + \underbrace{(30\beta_1^2 - 78\beta_0 + 54)}_{\text{定数項}}$$

を β_0 について微分することで，

$$8\beta_0 + (20\beta_1 - 28)$$

という式が得られます。

いま，それぞれの偏微分について得られた式がどちらも 0 になる β_0, β_1 が最小二乗法の解となるため，

$$\begin{cases} 60\beta_1 + 20\beta_0 - 78 = 0 \\ 8\beta_0 + 20\beta_1 - 28 = 0 \end{cases}$$

を解くと，$(\beta_0, \beta_1) = (1.5, 0.8)$ となります。

(c)

一般的に，単回帰分析について最小二乗法で求めた回帰係数は，各変数の平均値 $\bar{x}, \bar{y}$ および x の分散 s_x^2，共分散 s_{xy} を用いると，

$$\beta_0 = \bar{y} - \beta_1 \bar{x}, \qquad \beta_1 = \frac{s_{xy}}{s_x^2} \tag{4.8}$$

と表すことができます。ここに各値を当てはめて，回帰係数を求めてみましょう。まずは計算に必要な平均値などを求めます。

$$\bar{x} = \frac{1+2+3+4}{4} = 2.5, \qquad \bar{y} = \frac{2+3+5+4}{4} = 3.5$$

これを利用すると，s_x^2 および s_{xy} はそれぞれ

$$\begin{aligned} s_x^2 &= \frac{1}{4}\left\{(1-2.5)^2 + (2-2.5)^2 + (3-2.5)^2 + (4-2.5)^2\right\} \\ &= \frac{1}{4}(2.25 + 0.25 + 0.25 + 2.25) \\ &= 1.25 \end{aligned}$$

$$\begin{aligned} s_{xy} &= \frac{1}{4}\{(1-2.5)(2-3.5) + (2-2.5)(3-3.5) + (3-2.5)(5-3.5) + (4-2.5)(4-3.5)\} \\ &= \frac{1}{4}(2.25 + 0.25 + 0.75 + 0.75) \\ &= 1 \end{aligned}$$

となります。以上より，

$$\beta_1 = \frac{1}{1.25} = 0.8, \qquad \beta_0 = 3.5 - 0.8 \cdot 2.5 = 1.5$$

となり，もちろん **(b)** で得られた結果と一致します。

【解答】

(a)$4\beta_0^2 + 30\beta_1^2 + 20\beta_0\beta_1 - 28\beta_0 - 78\beta_1 + 54$
(b)(c)$\hat{y} = 1.5 + 0.8x$

(3) ある小学校の 1 年生から……

(a)

B さんの提言は，**相関関係と因果関係を混同**しており，統計データの解釈を誤っています。このデータには，「**年齢（学年）**」という第 3 の変数が隠れています。高学年になるほど身長は高くなり，学習内容も高度になるためテストの点数も高くなる傾向があります。このように，1 年生から 6 年生までのデータを一緒に分析したことで，「年齢」という交絡因子が原因となり，見かけ上の相関（**疑似相関**）が生じていると考えられます。したがって，「牛乳を飲んで身長を伸ばせば成績が上がる」という結論には飛躍があります。

(b)

因果関係があるかを考える際には，まず「どのようなロジックが考えられるか」を想像します。実際の研究やビジネス場面では，因果関係は過去の研究や，専門領域に蓄積された知見をもとに構成していくことになります。そして，想像された因果関係を，実際にデータによって検証する（RCT や統計的因果推論などを利用する）ことで，因果関係を明らかにしていくのです。

今回のケースでは，無理矢理感がてみると，以下のようなロジックが考えられます。もちろん，他の変数が影響している可能性もいくらでも考えられます。

- **身長→学力**：身体的な成長が早い児童は，脳の発達も早く認知能力が高いため，学力も高くなる。
- **学力→身長**：学力が高い児童は，自己管理能力が高く，食事や睡眠にも気を配るため，身体的な発育も良くなる。

このように，多くの場合には，想像のレベルならばどちらの方向の因果関係も考えることができるでしょう。したがって，厳密に因果関係を議論する場合には，できるだけ**これらの変数の時間的関係も考慮に入れた**データが取れるとベターです。

(c) 因果関係を検証するためのデータ

因果関係を特定するためには，まず疑似相関の原因となっている「年齢」の影響を取り除くことが重要です。このためには，**同じ学年の児童に限定**してデータを収集・分析すると良いでしょう。これにより「年齢」という要因は統制されます。ただ，これだけでは因果関係までは分かりません。

より詳細な分析のためには，

- 他に第 3 の変数となりうる（できるだけ多くの）変数：性別，家庭環境，生活習慣といっ

た，身長と学力の両方に関係しそうな変数
- 因果関係のロジックを示すために有用な変数：**(b)** で挙げた例で言えば，「認知能力」や「自己管理能力」「睡眠時間」など

もあわせて収集すると良いでしょう。欲を言えば，因果関係を示すための変数については，時間的な前後関係があるとなお良いです。例えば，「4 月時点での身長・認知能力」と「7 月時点での学力」の間に，（他の変数の影響を統制したうえで）正の相関関係があれば，これは「4 月時点で身長が高かった人ほど認知能力も高く，その結果 7 月時点では（4 月時点で身長が低かった人と比べて）学力に差が出ている」というわけなので，単純な相関よりは因果関係に近いものが示せている，と言えるかもしれません。

【解答例】

(a)B さんの提言は誤りである可能性が非常に高い。相関と因果を混同しており，この相関は「年齢（学年）」などの交絡因子による疑似相関である可能性が高い。
(b)（例）身長→学力：身体的成長と脳の発達の連動。 学力→身長：自己管理能力の高さが生活習慣の良さに繋がり発育を促進。
(c)「年齢」の影響を統制するため，同じ学年のみを対象に分析する。また，性別や生活習慣など交絡要因となりうる変数や，認知能力など因果関係を示すのに有用と思われる変数を追加収集すると良いと考えられる。

(4) 「家の広さ (X)」から……

言うまでもないことですが，家の価格は「家の広さ」以外にも様々な要因によって決まります。そして，回帰分析では，**分析者がモデルに入れた説明変数**によって被説明変数の値の変動を説明します。このとき，残差はただの誤差というだけでなく「モデルに含まれている説明変数（この場合は『家の広さ』）だけでは説明しきれなかった部分」という意味を持ちます。

残差分析を行うことで，ある回帰分析のモデルを当てはめたときに，具体的にどのケースにおいて大きな残差が見られるかをチェックすることができます。これは，現在の回帰モデルが見落としている，家の価格に影響を与える重要な要因を発見できる可能性を持っています。今回の例の場合，具体的には，以下のような新しい知見や仮説が考えられます。

- **立地の重要性**：A さんの家は，最寄り駅に近い，人気の学区内にある，といった「立地条件」が非常に良いのではないか。
- **築年数や建物の質**：A さんの家は，築年数が非常に浅い，最近リフォームされた，耐震性能が高い，といった「建物の質」に関する要因があるのではないか。
- **付加価値の存在**：日当たりが良い，眺望が素晴らしい，広い庭がある，といった他の家にはない「特別な付加価値」があるのではないか。

このように，残差分析は，モデルを改善するための次のステップ，すなわち「**次に追加すべき**

重要な説明変数は何か」という仮説を与えてくれる，非常に有用な分析手法なのです。

【解答例】

- **ズレの意味**：A さんの家には，「家の広さ」という説明変数だけでは説明できない，価格を押し上げる特別なプラスの価値が存在することを示唆している。
- **得られる知見**：モデルに新たに追加すべき重要な説明変数の仮説が得られる。例えば，「立地条件（駅からの距離など）」「建物の質（築年数など）」「付加価値（日当たり，眺望など）」といった変数が，家の価格に影響を与えているのではないかという仮説を立てることができる。

第 5 章

(1) 次に挙げる調査ケースのうち……

調査には，対象となる集団の全員を調べる「**全数調査**」と，集団の一部を抜き出して調べる「**標本調査**」がありました。どちらが適切かは，調査の目的や対象集団の特性によって決まります。ただし，基本的には**全数調査が容易にできるならば，全数調査をしたほうが良い**と考えられます。今回の調査ケースをそれぞれ見てみると，

小学校のクラスの身長測定　調査対象の集団（母集団）が 40 人と非常に小さいため，全員を調査するのにかかる**時間，費用，手間がごくわずか**です。このような場合には，全数調査を問題なく実施できるでしょう。

新製品の破壊検査　この調査は，製品を壊して強度などを調べるものです。もし全数調査を行ってしまうと，**販売する製品が一つもなくなってしまいます**。このような調査では，標本調査が不可欠です。

全国の高校生の学習時間調査　調査対象の母集団が約 300 万人と非常に大規模です。もし全数調査を行おうとすると，**膨大な費用と時間，労力がかかり，現実的ではありません**。このような場合は標本調査が適しています。

【解答】

- **最も適切なケース**：小学校のクラス（40 人）の身長測定
- **理由**：調査対象となる母集団が 40 人と非常に小さく，全員を調査するためのコスト（時間，費用，手間）が非常に低いため。逆に，破壊検査は全数調査が不可能であり，全国調査はコスト面から全数調査が非現実的である。

(2) ある大学（学生数 4000 人）の……

標本調査で重要なのは，母集団の縮図となるような，**偏りのない標本**を抽出することです。究極的な理想は，母集団全体の中から，くじ引きに類する方法で，ランダムに対象者を選び出すことです。これが難しい場合には，自分が採用したサンプリングの方法がどのような標本の偏りを生む可能性があるかを慎重に考えましょう。もしも，その偏りが分析結果にクリティカルに影響する場合には，別の方法を検討すべきかもしれません。

最寄り駅の改札で学生に聞く　この場合，対象者は**電車通学の生徒に偏ります**。徒歩，自転車，バスなどで通学している生徒が調査対象から完全に抜け落ちてしまいます。そのため，大学全体の平均値と比べると，通学時間が長めに出てしまうかもしれません。

文学部の学生から無作為に選んだ人たちに聞く　**標本での学部が偏ってしまいます**。一般的に，多くの大学では学部によってライフスタイルが異なる傾向にあり，例えば実験等で大学に行く必要日数が多い理系学部と比べると，相対的に大学より遠くに住んでいるなどの可能性も考えられます。

ある週の月曜から金曜 1 限の講義の出席者に聞く　**1 限に出席できるような真面目な学生に偏ります**[4]。また，通学時間が非常に長い生徒は最初から 1 限の履修を避けていたり，調査の日に交通トラブルがあれば，回答できない人がいるかも知れません。その結果，通学時間が長い生徒が調査から抜け落ち，平均通学時間が実際よりも短く算出される可能性があります。

【解答例】

- **駅の改札**：電車通学の生徒に偏り，他の通学手段の生徒が含まれない結果，平均値が母数よりも長くなってしまう可能性がある。
- **文学部**：特定の学部の学生に偏るため，大学全体の平均値とは離れる可能性がある。
- **朝の授業**：遅刻や欠席をしない生徒に偏り，通学時間が長い生徒などが除外される結果，平均値が母数よりも短くなってしまう可能性がある。

(3) 高校生の身長に関する……

サンプルサイズを増やすメリット

サンプルサイズを大きくする最大のメリットは，**母集団の真の値をより正確に推定できるようになる**ことです。サンプルが大きいほど偶然による偏りが減り，標本から計算した統計量（平均値など）が母集団の値に近づきます（詳しくは第 8 章で）。統計的には，標本平均と母平均の差は「標準誤差」(9 章）と呼ばれ，多くの場合には，これがサンプルサイズに応じて小さくなると

[4] もちろん，1 限を履修しているならば出席するのが普通です。が，大学生の現実として，なかなか 1 限には出席できない人がいるのもまた事実です。

いう形で表されます。

サンプルサイズを決める際の現実的な制約

ということで，理論的にはサンプルサイズは大きいほど良いのですが，現実には様々な制約があります。

- **費用**：調査票の印刷代，郵送費，謝礼など，サンプルサイズが大きくなるほど費用は増大します。
- **時間**：調査対象者全員から回答を回収し，集計・分析するには時間がかかります。
- **労力**：調査を実施・管理・分析するための人手にも限界があります。

どの程度の精度で推定する必要があるか，によって必要なサンプルサイズを事前に決めることも可能（詳しくは第 9 章）なので，そういった方法を駆使して，必要十分なサンプルサイズを用意することが重要となります。

多段抽出法の具体的な手順

「**多段抽出法**」は，母集団全体（今回のケースでは全国の高校生：約 300 万人）からランダムサンプリングする代わりに，複数の層におけるランダムサンプリングを繰り返す方法です。具体的には，例えば以下のように行うことができます。

1. **第 1 段階（地域の抽出）**：全国の 47 都道府県から，ランダムに **10 の都道府県**を抽出します。
2. **第 2 段階（学校の抽出）**：選ばれた 10 都道府県内の全ての高校のリストから，各都道府県あたり **5 つの高校**をランダムに抽出します（合計 50 校）。
3. **第 3 段階（生徒の抽出）**：選ばれた 50 校のそれぞれについて，全校生徒の名簿から **20 人の生徒**をランダムに抽出します。

この手順により，最終的に $50 \text{ 校} \cdot 20 \text{ 人} = 1000 \text{ 人}$ の標本が得られます。全国の高校生から完全ランダムに 1000 人を選び出す場合には，確実に数百の高校に依頼する必要があり大きな手間がかかりますが，多段抽出法を利用することで，このコストを 50 校に減らすことができます。

【解答例】

- **サンプルサイズを増やすメリット**：標本から得られる推定値の信頼性が高まり，推定の誤差が小さくなる。
- **現実的な制約**：費用，時間，労力（人的資源）といったリソースの限界。
- **多段中手法の手順例**：(1) 全国の都道府県から一部を無作為抽出 $\rightarrow$ (2) 選ばれた都道府県内の高校から一部を無作為抽出 $\rightarrow$ (3) 選ばれた高校の生徒から一部を無作為抽出。

第6章

(1) サイコロを2回投げたときの……

サイコロを2回投げるとき，起こりうるすべての目の出方は $6 \cdot 6 = 36$ 通りです。あとは，各設問の状況に合致するパターンの数を数えるだけです。

(a)

2回とも同じ目が出る事象は，(1,1), (2,2), (3,3), (4,4), (5,5), (6,6) の **6通り**です。よって，確率は $\dfrac{6}{36} = \dfrac{1}{6}$ です。

(b)

ここでは，余事象（「和が9以上になる」確率）を考えます[5]。和が9以上になるのは，(3,6), (4,5), (4,6), (5,4), (5,5), (5,6), (6,3), (6,4), (6,5), (6,6) の **10通り**です。よって，「和が9以上になる」確率は $\dfrac{10}{36}$ です。これを1から引くと，求める確率は $1 - \dfrac{10}{36} = \dfrac{26}{36} = \dfrac{13}{18}$ です。

(c)

余事象（「1回も6が出ない」確率）を考えます。1回目に6以外が出る確率は $\dfrac{5}{6}$，2回目も6以外が出る確率は $\dfrac{5}{6}$ です。よって，「2回とも6が出ない」確率は $\dfrac{5}{6} \cdot \dfrac{5}{6} = \dfrac{25}{36}$ です。最終的に求める確率は $1 - \dfrac{25}{36} = \dfrac{11}{36}$ です。

【解答】

(a) $\dfrac{1}{6}$

(b) $\dfrac{13}{18}$

(c) $\dfrac{11}{36}$

(2) 大学生100人に1週間の……

[5] 2つのサイコロの出目の和は，2から12になりますが，ちょうど中央値になるのは7です。そのため，「2から8」の全パターンを数えるよりも，「9から12」の全パターンを数えるほうが早いのです。

(a)

平均値の計算では，すべての値を足す必要がありますが，同じ値が複数回登場する場合には，これを掛け算で表したほうが楽になります。そこで，x の値が同じ人をまとめて表すと，平均値は

$$\bar{x} = \frac{(0 \cdot 20) + (1 \cdot 30) + (2 \cdot 25) + (3 \cdot 15) + (4 \cdot 7) + (5 \cdot 3)}{100} = \frac{168}{100} = 1.68$$

となります。

(b)

分散の計算も，平均値のときと同じように，x の値が同じ人はまとめて表すと楽になります。

$$\begin{aligned} s_x^2 &= \frac{20 \cdot (0 - 1.68)^2 + 30 \cdot (1 - 1.68)^2 + 25 \cdot (2 - 1.68)^2 + 15 \cdot (3 - 1.68)^2 + 7 \cdot (4 - 1.68)^2 + 3 \cdot (5 - 1.68)^2}{100} \\ &\approx \frac{56.45 + 13.87 + 2.56 + 26.136 + 37.68 + 33.07}{100} \\ &\approx 1.698 \end{aligned}$$

【別の解法】変数 x の分散は，実は $s_x^2 = \overline{x^2} - (\bar{x})^2$，すなわち「2 乗の平均」マイナス「平均の 2 乗」によっても求められます。

$$\begin{aligned} \overline{x^2} &= \frac{(0^2 \cdot 20) + (1^2 \cdot 30) + (2^2 \cdot 25) + (3^2 \cdot 15) + (4^2 \cdot 7) + (5^2 \cdot 3)}{100} \\ &= \frac{0 + 30 + 100 + 135 + 112 + 75}{100} = \frac{452}{100} = 4.52 \\ \text{分散} &= 4.52 - (\bar{x})^2 = 4.52 - 2.8224 = 1.6976 \end{aligned}$$

(c)

週 2 回以上運動する学生の割合は，単純に標本中の該当人数を求めるだけです。$25 + 15 + 7 + 3 = 50$ 人より，割合は $\frac{50}{100} = 0.5$ です。もちろんこれは，この標本における割合なので，母集団での割合はまた別の値となるでしょう。ただし，標本のサイズが大きくなるほど，標本でこのように計算した割合と，母集団での実際の割合のズレは小さくなっていきます。

(d)

この標本からランダムに 1 人選んだときに，その人が週 3 回以上運動している確率は，（無作為抽出が完璧にできるならば）もちろん「標本の中で週 3 回以上運動している人の割合」に一致します。この該当人数は $15 + 7 + 3 = 25$ 人であることから，求めたい確率は $\frac{25}{100} = 0.25$ です。

【解答】

(a)1.68 回
(b)1.6976
(c)0.5 (50%)

(d)0.25 ($\frac{1}{4}$)

(3) 次の表は，あるコンビニエンスストアの……

期待値や分散の計算の仕方は，(2) の度数分布表での計算と基本的には同じです。「度数」が「確率」に変わっていますが，言ってしまえば「総度数 n が 1 になるように調整されている」だけなので，むしろ「n で割る」という作業がなくなってラク，とも言えるかもしれません。

(a)

$$\begin{aligned} E[X] &= \sum x \cdot P(X = x) \\ &= (0 \cdot 0.3) + (1 \cdot 0.3) + (2 \cdot 0.2) + (3 \cdot 0.1) + (4 \cdot 0.1) \\ &= 0 + 0.3 + 0.4 + 0.3 + 0.4 \\ &= 1.4 \end{aligned}$$

(b)

$$\begin{aligned} V[X] &= \sum (x - E[X])^2 \cdot P(X = x) \\ &= \{(0 - 1.4)^2 \cdot 0.3\} + \{(1 - 1.4)^2 \cdot 0.3\} + \{(2 - 1.4)^2 \cdot 0.2\} + \{(3 - 1.4)^2 \cdot 0.1\} + \{(3 - 1.4)^2 \cdot 0.1\} \\ &= 0.588 + 0.048 + 0.072 + 0.256 + 0.676 \\ &= 1.64 \end{aligned}$$

【別の解法】この場合でも，「2 乗の平均」マイナス「平均の 2 乗」の公式を利用して解くことが可能です。すなわち，$V[X] = E[X^2] - (E[X])^2$ です。

$$\begin{aligned} E[X^2] &= (0^2 \cdot 0.3) + (1^2 \cdot 0.3) + (2^2 \cdot 0.2) + (3^2 \cdot 0.1) + (4^2 \cdot 0.1) \\ &= 0 + 0.3 + 0.8 + 0.9 + 1.6 = 3.6 \\ V[X] &= E[X^2] - (E[X])^2 = 3.6 - (1.4)^2 = 3.6 - 1.96 = 1.64 \end{aligned}$$

(c)

これは単純に $P(X \geq 2)$ を求めるだけです。離散確率変数ならば，該当する確率をすべて足し合わせたら良いので，$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4) = 0.2 + 0.1 + 0.1 = 0.4$ となります。

(d)

これも該当する確率をすべて足し合わせるだけです。$P(X \leq 1) = P(X = 0) + P(X = 1) =$

$0.3 + 0.3 = 0.6$

【解答】

(a)1.4 個
(b)1.64
(c)0.4
(d)0.6

(4) 以下の図は，ある連続型確率変数……

連続型確率分布の性質は以下の 2 つです。

- 特定の一点の値をとる確率（例：$P(X = 0)$）は，常に **0** になります。
- ある範囲の確率（例：$P(0 \leq X \leq 1)$）は，確率密度関数のグラフと X 軸で囲まれた部分の**面積**に等しくなります。

この性質に基づいて，それぞれの確率を評価すると，

- $P(X = 0)$ と $P(X = 3)$：特定の値の確率なので，両方とも **0** です。
- $P(-2 \leq X \leq -1)$：グラフの-2 から-1 までの区間の面積です。この部分は山から離れており，面積は**小さい正の値**です。
- $P(0 \leq X \leq 1)$：グラフの 0 から 1 までの区間の面積です。この区間は $P(-2 \leq X \leq -1)$ と横幅は同じですが，それよりも分布の山の中心に近く，関数値が最も高い部分を含んでいるため，面積は相対的に**より大きな正の値**になります。

【解答】 $P(X = 0) = P(X = 3) < P(-2 \leq X \leq -1) < P(0 \leq X \leq 1)$

(5) 以下の確率変数について，その確率分布が……

【解説】

(a)

一般的な（よく考えて作られた）テストでは，多くの学生が平均点周辺に集まり，極端に高い・低い点数になる生徒は少なくなるため，**左右対称の山のような形**になると予想されます。取りうる値は 0 から 100 の整数なので，離散確率分布となります（が，実際の分析では，これだけカテゴリ数が多い場合には連続確率分布を置くことも可能です）。

ただし，もちろんテストの点数の分布は，テストの性質にもよります。例えばテストの難易度

が非常に高く，ごく一部の生徒のみが対応できるようなテストであれば，左右対称よりも左に裾が長いような形になるかもしれません。また，問題数が極端に少ない（例：4 問で構成，1 問 25 点）場合には，左右対称とは言いづらい確率分布になる可能性もあるでしょう。

(b)

街中で偶然見かける友人の数は，0 人や 1〜2 人という日は多いものの，多数になる日は滅多にないでしょう。そのため，小さい値の頻度が高く，値が大きくなるにつれて急激に頻度が下がる，**右に長く裾を引く分布（右に歪んだ分布）**になると考えられます。取りうる値はもちろん 0 以上の整数のみですが，最も起こりやすいのは 0（か場合によっては 1-2 くらい）かもしれません。

(c)

「今」のバッテリーの残量自体がランダムだとすると，「次に充電が必要になるまでの時間」は比較的なだらかに広く分布すると考えられます。ただし，パソコンのバッテリーの残量が少ない状態で「今から使い始める」というケースはあまりないかもしれません。比較的残量が多い状態にあることが多い（右のほうにピークがあり，左に裾が長い）と考えると，そして一部のタスクでは急激にバッテリーを消費すると考えると，稀に極端に時間が短い場合が考えられるため，結果的に「次に充電が必要になるまでの時間」の分布は**右のほうにピークがあり，左に裾が長い分布**になると考えられます[6]。

(d)

大きな臨時収入や支出がなければ，多くの月は平均的な増減額の周辺に収まるでしょう。中心は個人によって異なるはずです。毎月いくら程度貯蓄しようと決めているかによって平均値は決まりますが，いずれにせよ，**その中心の周りである程度左右対称な分布**になる可能性が高いです。そして，場合によっては 0 以下（負）の値を取る可能性ももちろんあります。

【解答例】

(a)左右対称に近い山のような形
(b)右に長く裾を引く分布（右に歪んだ分布）
(c)右のほうにピークがあり，左に裾が長い分布
(d)中心（平均増減額）の周りである程度左右対称な分布

(6) 6.3 節の最後【コラム】では，……

[6] ただしこれは，前提の置き方や考え方等によっていろいろな答えが出てくると思います。統計学的な思考のトレーニングなのです。

メリット

連続変数を離散的に扱うメリットは，統計的な性質というよりも，調査時や結果の解釈のわかりやすさにあります。

1. **回答の得やすさ**：アンケート調査などで正確な金額を答えることは，源泉徴収票が手元になかったりしないと難しく，また回答にも抵抗があることが多いでしょう。この場合も，カテゴリであれば回答しやすい傾向があります。
2. **分かりやすさと解釈の容易さ**：例えば「年収の平均が 352 万円」という連続的な示し方よりも，特に数字が苦手な人にとっては，「301-400 万円の層が最も多い（最頻値）」といった表し方をした方が，直感的に理解しやすくなるかもしれません。
3. **外れ値に対する頑健性**：統計的には，端のカテゴリを「1000 万円以上」とすることで，場合によっては極端に高い年収の値が全体の平均値を大きく引き上げるなどの影響を緩和できます。ただしこれが分析の目的に沿って問題ない変更であるかは要注意です。

デメリット

1. **情報の損失**：離散化する場合の最大のデメリットです。例えば「301-400 万円」というカテゴリでは，301 万円と 399 万円の差が失われ，同じものとして扱われてしまいます。一方で，本来は「399 万円と 401 万円」はほとんど同じであるにもかかわらず，これらは明確に別のカテゴリとして（つまり 301 万円と 399 万円の差よりも大きな差として）扱われてしまいます。
2. **区切り方の恣意性**：カテゴリの境界をどこに設定するかは分析者が決める必要があり，この区切り方次第で分析結果が変わってしまう可能性があります。ヒストグラムも，区切り幅次第である程度形を変えられることがあり，場合によっては分析者にとって都合の良い解釈ができるように区切られてしまうかもしれません。
3. **統計分析手法の制限**：統計的には，相関係数や回帰分析といった強力な手法が使えなってしまいます。できないことはないのですが，カテゴリの分け方による影響で，結果が正しくなくなってしまう可能性があります。

【解答例】

- **メリット**：(1) アンケートなどで回答を得やすくなる，(2) 解釈が直感的で分かりやすくなる，(3) 外れ値の影響を緩和できる。
- **デメリット**：(1) 元データが持つ詳細な情報が失われる，(2) カテゴリの区切り方に恣意性が入り込む，(3) 使える統計分析手法が制限される。

第7章

(1) 以下の事象について，……

【解説】

(a)

ある学生が毎日同じルートで通学したときの所要時間は，信号待ち，交通量，天候など，多くの独立した小さな要因の積み重ねで決まると考えられます。中心極限定理により，多くの独立な確率変数の和は正規分布に近似するため，通学時間も平均値を中心とした左右対称の釣鐘型の分布（正規分布）になると考えられます。この際に必要な仮定としては，「信号に引っかかるかどうかはランダムに決まる」「交通量はランダムに決まる」など，様々な要因が無作為に決定しているということです。もしも「火曜日だけは必ず信号に引っかかる」などがあると，「火曜日の所要時間」と「それ以外の曜日の所要時間」の分布が異なるため，例えば 2 つの正規分布を混ぜたような形になるかもしれません。

(b)

「二度寝する」か「しない」かの 2 択の結果が得られる試行を，1 ヶ月（約 30 日）という決まった回数だけ繰り返すと考えられます。各日の二度寝の確率が（平日も休日も，前の日に寝た時間などにも関係なく）一定で，毎日の結果が互いに独立であると仮定すれば，二項分布が当てはまります。

(c)

このように，「ある事象が初めて起こるまでの時間」のモデル化には，指数分布がよく使われます。具体的には，探し物が「見つかる」という事象が，いつ起きるか分からないが，どの瞬間にも一定の確率で起こりうると仮定した場合（すなわち，発見率が一定の場合）には，指数分布を当てはめることが適切となります。

(d)

ポアソン分布は，「一定の期間や範囲内で，ある事象が平均して何回起こるか」を表す分布です。バグが互いに独立に，かつ一定の発生率で起こるという仮定のもとでは，「1 日」という一定期間に発生するバグの数をモデル化するために，ポアソン分布が適切と考えられます。

(e)

体内時計による時間測定の誤差は，多くの生理的・心理的な要因が偶然に影響し合った結果と考えられます。そのため，真の 1 分（あるいは個人が平均的に感じる 1 分）を中心として，測定結果が左右対称にばらつく正規分布に従うと考えるのが自然です。

【解答例】

(a)**正規分布**。（仮定）所要時間が多くの独立かつランダムな要因の和で決まる。
(b)**二項分布**。（仮定）毎日の二度寝の有無が独立で，その確率は一定。
(c)**指数分布**。（仮定）探し物が見つかる確率（発見率）が時間によらず一定。
(d)**ポアソン分布**。（仮定）バグが互いに独立に，かつ一定の発生率で起こる。
(e)**正規分布**。（仮定）測定誤差が多くの偶然的な要因の和で決まる。

(2) ある大学のテニスサークルでは……

これは「成功確率 $p = 0.3$ の試行を n 回行ったときに k 回成功する」という**二項分布**の問題です。

(a)

$n = 20, p = 0.3$ で，$k = 5$ となる確率を求めます。二項分布の確率質量関数は $P(X = k) = {}_n\mathrm{C}_k p^k (1-p)^{n-k}$ です。

$$\begin{aligned} P(X = 5) &= {}_{20}\mathrm{C}_5 \cdot (0.3)^5 \cdot (1 - 0.3)^{20-5} \\ &= 15504 \cdot (0.3)^5 \cdot (0.7)^{15} \\ &\approx 15504 \cdot 0.00243 \cdot 0.00475 \approx 0.1789 \end{aligned}$$

(b)

$n = 50, p = 0.3$ です。このように n が大きい場合には，「0 人が入部する」から「19 人が入部する」をすべて計算して足し合わせるのは（手計算では）現実的ではないため，正規分布で近似して求めましょう。正規近似では，サンプルサイズが十分に大きい場合，もとの二項分布と同じ期待値・分散をもつ正規分布に近似できます。そこで，まずはもとの二項分布 $B(50, 0.3)$ の期待値と分散を求めると，

- 期待値 $\mu = np = 50 \cdot 0.3 = 15$
- 分散 $\sigma^2 = np(1-p) = 50 \cdot 0.3 \cdot 0.7 = 10.5$

となります。あとは，「20 人以上が入部する」確率 $P(X \geq 20)$ を，式変形して標準正規分布表に当てはめていくだけです。

$$\begin{aligned} P(X \geq 20) &= P\left(\frac{X - 15}{\sqrt{10.5}} \geq \frac{20 - 15}{\sqrt{10.5}} \right) \\ &\approx P\left(z \geq 1.54 \right) \end{aligned}$$

あとは，標準正規分布表から $P(Z \geq 1.54) \approx 0.0618$ となります。

【補足】二項分布に基づいてそのまま確率を計算すると，およそ 0.0848 となります。この値に少しでも近づけるために**連続性の補正**を行う場合には，0.5 を引いた $P(X \geq 19.5)$ として計算

します。

$$P(X \geq 19.5) = P\left(\frac{X-15}{\sqrt{10.5}} \geq \frac{19.5-15}{\sqrt{10.5}}\right)$$
$$\approx P\left(z \geq 1.39\right)$$

この場合，標準正規分布表から $P(Z \geq 1.39) \approx 0.0823$ となります。

【解答】

(a)約 17.89%
(b)約 6.18%（連続性の補正をした場合には約 8.23%）

(3) 学生食堂では，昼休みの……

【解説】これは「単位時間あたりに平均 λ 回起こる事象」を扱う**ポアソン分布**の問題です。1 分あたりの平均来店数は $\frac{40}{60} = \frac{2}{3}$ 人です。

(a)

15 分間の平均来店数は，$\frac{2}{3} \cdot 15 = 10$ 人です。したがって，「15 分間に来店する学生数」の確率分布は，$Po(10)$ となります。そして，ポアソン分布では，期待値と分散はともにパラメータ λ と等しくなるため，期待値・分散ともに $\lambda = 10$ となります。

(b)

$\lambda = 10$ のポアソン分布で，$x = 0$ となる確率を求めます。ポアソン分布の確率質量関数は $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ です。ここに $x = 0, \lambda = 10$ を代入すると，

$$P(X = 0) = \frac{e^{-10} \cdot 10^0}{0!} = e^{-10} \approx 0.0000454$$

となります。

(c)

1 時間（60 分）の平均来店数 $\frac{2}{3} \cdot 60 = 40$ です。このように λ が大きい場合には，正規分布で近似ができるようになります。ポアソン分布 $Po(40)$ と同じ期待値・分散を持つため，

- 期待値 $\mu = \lambda = 40$
- 標準偏差 $\sigma = \sqrt{\lambda} = \sqrt{40} \approx 6.32$

の正規分布によって近似が可能です。あとは，「60 人以上が来店する」確率 $P(X \geq 60)$ を，式変形して標準正規分布表に当てはめていくだけです。

$$P(X \geq 60) = P\left(\frac{X-40}{6.32} \geq \frac{60-40}{6.32}\right)$$
$$\approx P\left(z \geq 3.17\right)$$

あとは，標準正規分布表から $P(Z \geq 3.17) = 0.0008$ となります。

【補足】ポアソン分布に基づいてそのまま確率を計算すると，およそ 0.0017 となります。この値に少しでも近づけるために**連続性の補正**を行う場合には，0.5 を引いた $P(X \geq 59.5)$ として計算します。

$$P(X \geq 59.5) = P\left(\frac{X-40}{6.32} \geq \frac{59.5-40}{6.32}\right)$$
$$\approx P\left(z \geq 3.09\right)$$

この場合，標準正規分布表から $P(Z \geq 3.09) = 0.0010$ となります。

【解答】

(a)**期待値**: 10, **分散**: 10
(b)**約** 0.0045%
(c)**約** 0.08% （**連続性の補正をした場合には約** 0.1%）

(4) 潜在的な顧客にダイレクトメール……

これは「成功確率 $p = 0.05$ の試行で，初めて成功するまでに何回かかるか」という**幾何分布**の問題です。

(a)

幾何分布の期待値は $E[X] = \dfrac{1}{p}$ で計算できます。

$$E[X] = \frac{1}{0.05} = 20$$

(b)

これは「1 通目でクリックする確率」＋「2 通目でクリックする確率」＋「3 通目でクリックする確率」の和です。

- $P(X = 1) = p = 0.05$
- $P(X = 2) = (1-p)p = 0.95 \cdot 0.05 = 0.0475$
- $P(X = 3) = (1-p)^2 p = (0.95)^2 \cdot 0.05 \approx 0.0451$

これらの和は $0.05 + 0.0475 + 0.0451 = 0.1426$ です。

【解答】

(a)20 通
(b)約 14.3%

(5) ある果物の重さ (g) が正規分布……

【解説】平均 $\mu = 60$, 標準偏差 $\sigma = 8$ の正規分布に基づいて，指示された値や確率を求める問題です。

(a)

求めたい確率は，$P(X \geq 70)$ です。標準正規分布表に当てはめられるように標準化していきます。

$$\begin{aligned} P(X \geq 70) &= P\left(\frac{X-60}{8} \geq \frac{70-60}{8}\right) \\ &\approx P\left(z \geq 1.25\right) \end{aligned}$$

あとは，標準正規分布表から $P(Z \geq 1.25) \approx 0.1056$ となります。

(b)

この問題は，それぞれ $P(X > x) = 0.20$, $P(X < x) = 0.20$ となる x を求める問題に相当します。まず，$P(X > x) = 0.20$ となる x を求めるために，標準正規分布表で $P(Z > z) = 0.20$ となる z の値を探すと，およそ 0.84 あたりになることがわかります。これは，いかなる正規分布 $N(\mu, \sigma^2)$ についても，$P(X > \mu + 0.84\sigma) = 0.2$ になる，ということを意味しています。したがって，標準化を戻すように操作すると，

$$\begin{aligned} P(Z > z) &= P\left(\mu + Z\sigma > \mu + z\sigma\right) \\ &= P\left(X > 60 + 0.84 \cdot 8\right) \\ &= P(X > 66.72) = 0.2 \end{aligned}$$

となることがわかります。同様に，$P(X < x) = 0.20$ となる x については，符号が入れ替わるだけなので，$x = 60 - 0.84 \cdot 8 = 53.28$ となります。

(c)

平均値 60 を中心とした左右対称な区間で，確率が 0.05 となる範囲を求めます。これは，正規分布の中心で分割すると，まず上側については $P(\mu \leq X \leq x) = 0.025$ となるような x の値を求めることに相当します。正規分布の対称性を踏まえると，

$$
\begin{aligned}
0.5 - P(\mu \leq X \leq x) &= P(X > x) \\
&= P\left(\frac{X-60}{8} > \frac{x-60}{8}\right) \\
&= P\left(Z > \frac{x-60}{8}\right) = 0.475
\end{aligned}
$$

となるような x を求める問題に変換できます。標準正規分布表から，$P(Z > z) = 0.475$ に近い値を探すと，$z \approx 0.06$ であることから，

$$
\frac{x-60}{8} = 0.06
$$

となり，これを変形させることで $x = 60.48$ とわかります。同様に，5% 区間の下側は，途中の式変形の符号が変わるだけなので，$x = 60 - 0.48 = 59.52$ となります。

【解答】

(a)約 10.56%
(b)（上位 20%）約 66.72 g，（下位 20%）約 53.28 g
(c)59.52 g から 60.48 g

(6) ある選挙で，……

(a)

$n = 10, p = 0.5$ の二項分布 $B(10, 0.5)$ で，$P(X \geq 6)$ を求めます。頑張って一つ一つの確率を計算してみると，

$$
\begin{aligned}
P(X \geq 6) &= P(X=6) + P(X=7) + P(X=8) + P(X=9) + P(X=10) \\
&= \sum_{k=6}^{10} {}_{10}\mathrm{C}_k (0.5)^k (0.5)^{10-k} = (0.5)^{10} \sum_{k=6}^{10} {}_{10}\mathrm{C}_k \\
&= \frac{1}{1024}(210 + 120 + 45 + 10 + 1) = \frac{386}{1024} \approx 0.377
\end{aligned}
$$

となります。今回の計算では，$P(X = x)$ の計算が，x の値によらず 0.5^{10} を含んでいることから，少しだけ楽に計算が可能です。

(b)

二項分布 $B(10, 0.5)$ の期待値は $\mu = np = 5$, また標準偏差は $\sigma = \sqrt{np(1-p)} = \sqrt{2.5} \approx 1.581$ です。ここから，$P(X \geq 6)$ を標準正規分布表に当てはめて計算すると，

$$
\begin{aligned}
P(X \geq 6) &= P\left(\frac{X-5}{1.581} \geq \frac{6-5}{1.581}\right) \\
&\approx P\left(Z \geq 0.63\right)
\end{aligned}
$$

あとは，標準正規分布表から $P(Z \geq 0.63) \approx 0.2643$ となります。

(c)

採用すべき結果：(a) の**二項分布を用いて計算した正確な値（0.377）**を採用すべきです。**理由と注意点**：正規分布による近似は，あくまでサンプルサイズ n が大きい（かつ成功確率 p が極端な値ではない）場合に有効な「近似計算」です。今回のケースでは $n = 10$ と比較的小さいために，近似計算の誤差が大きくなってしまいました。このように，近似は計算を簡略化する強力なツールですが，可能であれば（特に n が小さい場合），元の離散分布を使って正確に計算する方が常に望ましいです。近似を用いる際には，その適用条件（n が十分に大きいかなど）を満たしているかを確認し，結果が「おおよその値」であることを認識しておく必要があります。ちなみに今回の場合，連続性の補正を用いて，$P(X \geq 5.5)$ を計算すると，およそ 0.3745 となり，(a) の値にかなり近くなります。

【解答】

(a)約 37.7%
(b)約 26.4%
(c)(a) の正確な値を採用すべき。正規近似はサンプルサイズが小さいと誤差が大きくなるため。近似を用いる際は，その適用条件を満たしているかを確認し，結果が近似値であることを認識する必要がある。

(7) あるコールセンターへの……

コールセンターの着信数が，どの 1 時間をとっても同程度の期待値になることは，現実的にはあまり考えられません。むしろ，1 日の中でも時間帯によって大きく変動するのが普通です。例えば，多くの企業のコールセンターでは，昼休み（12 時〜13 時）や業務終了間際（16 時〜17 時）に着信が集中し，早朝や深夜はほとんど着信がない，といったパターンが見られます。この場合，「発生率が一定」というポアソン分布の仮定は満たされません。

そのような状況であるにもかかわらず，もし，平均着信率が 1 日を通して変わらないとみなして，単一のパラメータ λ を用いてポアソン分布モデルを当てはめてしまうと，モデルは時間帯による繁閑の差を全く捉えることができません。その結果，

- **ピーク時の呼量を過小評価**：モデルは着信が集中する時間帯の「本当の忙しさ」を予測できず，必要なオペレーターの人数を少なく見積もってしまう可能性があります。
- **閑散時の呼量を過大評価**：逆に，着信が少ない時間帯の呼量を多めに見積もってしまい，不必要な数のオペレーターを配置してしまう可能性があります。

というように，実際の着信数をあまり反映できていない見積もりになってしまう可能性があります。

統計モデルを現実に適用する際は，そのモデルが持つ**仮定が，対象とする現象の実態と合っているか**を常に批判的に検討する必要があります。そして今回の例のように仮定が満たされていない場合は，分析の目的にもよりますが，例えば「何人のオペレーターを用意したらよいか」を検討するような場合には，時間帯を区切って分析する（「12-13 時」と「19 時」を別々に分析する）と良いかもしれません。

【解答】

- **満たされない状況**：昼休みや業務終了間際など，時間帯によって電話の着信率が大きく変動する状況では，ポアソン分布の仮定が満たされているとは言えない。
- **影響と注意点**：単一の平均値でモデル化すると，ピーク時の呼量を過小評価し，閑散時を過大評価するなど，実態から乖離した予測をしてしまう。モデルの仮定が現実と合っているかを常に確認し，合わない場合は時間帯を区切るなど分析方法を工夫する必要がある。

(8) 大学生の「1 ヶ月の書籍購入費」……

正規分布に従うと仮定できる確率変数は，その性質として「平均値周辺に多く分布し，左右に均等に広がっている」必要があります。仮に，大学生の「1 ヶ月の書籍購入費」が正規分布に従うとしたら，例えば「1 ヶ月の平均値は 3000 円で，1000 円の人と 5000 円の人は同程度いる」といったデータが観測されると考えられます。しかし実際には，書籍購入費はこのような分布にはなっていないと考えられます。具体的には，

1. 多くの学生は毎月本を買うわけではないので，購入費が **0 円の確率が最も高く，大きな山が 0 の位置にできる**と考えられます。
2. 少額の購入者はいるものの，金額が上がるにつれて人数は急激に減少します。
3. 一方で，学期のはじめには専門の教科書をまとめて購入するため，数万円といった**非常に高額な支出をする学生が少数ながら存在**します。

これらの特徴を総合すると，分布の形状は，0 円に最も高いピークを持ち，そこから急激に減少した後，**右側に非常に長く裾を引く（右に歪んだ）分布**になると予想されます。

確率分布の（想像される）特徴の正規分布との違いを並べると，

1. **対称性**：正規分布は左右対称ですが，この分布は著しく**右に歪んで**います。
2. **値の範囲**：正規分布は理論上マイナスの値も取りえますが，購入費は必ず **0 円以上**です。
3. **中心の位置**：正規分布は平均値・中央値・最頻値が一致しますが，この分布では**最頻値は 0 円**，平均値は一部の高額購入者に引っ張られて中央値よりもさらに高い値になるはずです。

といった感じになるかもしれません。

【解答例】正規分布には従わない。具体的には，(1) 分布が左右対称ではなく，0 円を最頻値として**右に著しく歪んだ**形になると考えられる。また，(2) 値が必ず 0 以上であり，(3) 最頻値 (0 円)，中央値，平均値が一致しないと考えられる。

第 8 章

(1) 統計的推測を行う上では，……

統計的推測の目的は，手元にある 1 つの「標本」から，その背景にある「母集団」全体の性質を推測することです。ここで登場する**「標本の分布」**は，私たちが**実際に観測した 1 つの標本データ**の分布です。これは，たまたま身長が高い人が多く集まるなど，**偶然による偏りを含んでいる可能性**があります。一方**「標本分布」**は，もし母集団から**標本を何度も何度も抽出し続けたとしたら，その都度計算される統計量（例えば標本平均）がどのような分布を描くか**という，**理論上の確率分布**です。

以上を踏まえて**なぜ「標本分布」が重要なのか**を考えてみます。私たちが持っているのはたった 1 つの標本平均ですが，それが真の母平均とどれくらい近いのか，あるいは離れているのかを知りたいわけです。「標本分布」は，**「標本統計量が，偶然によってどの程度ばらつくのか」**という程度を示してくれます。特に**中心極限定理**により，「標本平均の標本分布は正規分布に近似する」ことが分かっています。この理論的な「標本分布」の性質が分かっているおかげで，私たちはたった 1 つの標本から，「真の母平均は 95% の確率でこの範囲にあるだろう」といった区間推定や仮説検定を行うことができるのです。つまり，「標本分布」は，**偶然の産物である標本と，知りたい母集団との間の関係性を，確率を通して考える**ために不可欠なものである，ということができます。これは，単一の，実際に観測された「標本の分布」だけでは考えようのない話なのです。

【解答例】手元の「標本の分布」は偶然による偏りを含み，それだけで統計的推測を行うことはできないが，「標本分布」は標本統計量（例：標本平均）が偶然によってどの程度ばらつくかを示す理論上の確率分布だから。この理論的な分布の性質（特に中心極限定理）が分かっているおかげで，私たちはたった 1 つの標本からでも，母集団の性質を確率的に推測（区間推定や仮説検定）することができる。

(2) ある果物の重さ（g）が……

標本平均 $\bar{X}$ の標本分布が正規分布 $N\left(\mu, \dfrac{\sigma^2}{n}\right)$ に従うことを利用して，それぞれの値を求め

ていきます。

標本平均の期待値

標本平均の期待値 $E[\bar{X}]$ は，母平均 μ と等しくなります。

$$E[\bar{X}] = \mu = 60$$

標本平均の標準偏差

標本平均の標準偏差は，正規母集団の場合は $\dfrac{\sigma}{\sqrt{n}}$ で計算されます。

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{10}} \approx 2.53$$

標本平均が 55g 以上 65g 以下となる確率

求めたい確率は，$P(55 \leq \bar{X} \leq 65)$ です。ここで，標本平均 $\bar{X}$ の標本分布が，先ほど求めたように正規分布 $N\left(60, 2.53^2\right)$ となることを踏まえて，値を標準化していきます。

$$\begin{aligned} P(55 \leq \bar{X} \leq 65) &= P\left(\frac{55-60}{2.53} \leq \frac{X-60}{2.53} \leq \frac{65-60}{2.53}\right) \\ &\approx P\left(-1.98 \leq Z \leq 1.98\right) \\ &= 2 \cdot P(0 \leq Z \leq 1.98) \\ &= 2\left\{0.5 - P(Z > 1.98)\right\} \end{aligned}$$

あとは，標準正規分布表から $P(Z \geq 1.98) \approx 0.0239$ となるため，求めたい確率は 0.9522 となります。

【解答】

- 期待値: 60 g
- 標準偏差: 約 2.53 g
- 確率: 約 95.22%

(3) あるバスケットボールチームの……

これは標本比率の標本分布の問題です。1 人あたりの成功回数が二項分布 $B(10, 0.3)$ に従い，サンプルサイズが $n = 5$ であると考えます。ここで，中心極限定理より，5 人の成功回数の平均値は，

$$\bar{X} \sim N\left(3, \frac{2.1}{5}\right)$$

によって近似することが可能です。

このとき，求めたい確率「10 回ずつ打ったときの平均成功率が 40% 以上となるおおよその確

率」は，$P(\bar{X} \geq 4)$ と表すことができるため，これを標準正規分布表によって求めます。

$$\begin{aligned} P(X \geq 4) &= P\left(\frac{\bar{X}-3}{0.648} \geq \frac{4-3}{0.648}\right) \\ &\approx P\left(Z \geq 1.54\right) \end{aligned}$$

あとは，標準正規分布表から $P(Z \geq 1.54) \approx 0.0618$ となります。

【別の解法】すべての試行（3 ポイントシュート）がそれぞれ同じ成功確率のベルヌーイ試行なのであれば，「5 人が 10 回ずつ」を「1 人が 50 回」とみなしてしまっても問題はないはずです。ということで，5 人分の計 50 回のシュートにおける成功回数を X とすると，これは二項分布 $B(50, 0.3)$ と表されます。この二項分布は，n が十分に大きいため，正規近似によって，同じ期待値と分散を持つ正規分布 $N(15,)$ に近似されます。ここで求めたい確率は，成功確率 40% 以上，すなわち「50 回中 20 回以上成功」と表せるため，$P(X \geq 20)$ となります。あとは，標準正規分布表を用いて確率を求めるだけです。

$$\begin{aligned} P(X \geq 20) &= P\left(\frac{\bar{X}-15}{3.24} \geq \frac{20-15}{3.24}\right) \\ &\approx P\left(Z \geq 1.54\right) \end{aligned}$$

あとは，標準正規分布表から $P(Z \geq 1.54) \approx 0.0618$ となります。

【解答】約 6.18%

(4) ある工場では，……

規格外品が生み出される確率が常に一定である（ポアソン過程に従う）と仮定するならば，1 日の規格外品の個数は，平均 $\lambda = 30$ のポアソン分布に従うと考えられます。このとき，1 日あたりの平均個数 $\bar{X}$ の標本分布は，サンプルサイズ $n = 7$ であることから，中心極限定理により，

- 標本平均の期待値：$E[\bar{X}] = \mu = \lambda = 30$
- 標本平均の分散：$\dfrac{\sigma^2}{n} = \dfrac{\lambda}{7} \approx 4.286$

の正規分布 $N(30, 4.286)$ で近似されます。そして，求めたい確率は，$P(\bar{X} \geq 35)$ です。あとは，標準正規分布表を用いて確率を求めるだけです。

$$\begin{aligned} P(X \geq 35) &= P\left(\frac{\bar{X}-30}{2.07} \geq \frac{35-30}{2.07}\right) \\ &\approx P\left(Z \geq 2.42\right) \end{aligned}$$

あとは，標準正規分布表から $P(Z \geq 2.42) \approx 0.0078$ となります。

【解答】約 0.78%

(5) 体育の授業で測定した……

母集団が正規分布に従う場合，その確率変数を**標本平均で**標準化した値の二乗和 $\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2$ の標本分布が，自由度 $n-1$ の χ^2 分布に従うことが知られていました。ここで，自由度 $n-1$ の χ^2 分布の期待値および分散がそれぞれ $n-1, 2(n-1)$ で表されることを踏まえると，標本分散 s_x^2 の期待値および分散はそれぞれ

期待値

$$E\left[\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2\right](=n-1) = E\left[\frac{n}{\sigma^2}\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]$$
$$= \frac{n}{\sigma^2}E\left[s_x^2\right]$$

と変換できることから，$E\left[s_x^2\right] = \frac{n-1}{n}\sigma^2 = \frac{8}{9}\cdot 6^2 = 32$ となります。

分散　同じように式変形させると，

$$V\left[\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2\right](=2n-2) = V\left[\frac{n}{\sigma^2}\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]$$
$$= \frac{n^2}{\sigma^4}V\left[s_x^2\right]$$

と変換できることから，$V\left[s_x^2\right] = \frac{2(n-1)}{n^2}\sigma^4 = \frac{16}{81}\cdot 6^4 = 256$ と求められます。

【解答】

- 標本分散の期待値: 32
- 標本分散の分散: 256

(6) ある保険会社では，……

今回のケースでは，具体的な確率分布の情報はわかりませんが，母平均と母標準偏差のみは分かっている状況です。大数の法則および中心極限定理は，母平均と母標準偏差が定義されていれば適用可能な定理なので，これだけの情報でも（理論上は）標本平均の標本分布を求めることが可能です。

(a)

サンプルサイズ $n=100$ は十分に大きいため，母集団が正規分布でなくても，**中心極限定理**により標本平均 $\bar{X}$ の標本分布は正規分布で近似できます。

- 標本平均の期待値：$E[\bar{X}] = \mu = 50$

- 標本平均の標準偏差：$\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{30}{\sqrt{100}} = 3$

求めたい確率は $P(45.5 \leq \bar{X} \leq 54.5)$ です。標準正規分布表に当てはめるために，式の変形をしていきましょう。

$$\begin{aligned} P(45.5 \leq \bar{X} \leq 55.5) &= P\left(\frac{45.5-50}{3} \leq \frac{X-50}{3} \leq \frac{55.5-50}{3}\right) \\ &\approx P\left(-1.5 \leq Z \leq 1.5\right) \\ &= 2 \cdot P(0 \leq Z \leq 1.5) \\ &= 2\left\{0.5 - P(Z > 1.5)\right\} \end{aligned}$$

あとは，標準正規分布表から $P(Z \geq 1.5) \approx 0.0668$ となるため，求めたい確率は 0.8664 となります。

(b)

もしサンプルサイズが 25 件だった場合，(a) で求めた確率の**信頼性は著しく低下します**。その理由は，今回の計算が**中心極限定理**に依存しているからです。中心極限定理は，サンプルサイズ n が**十分に大きければ**成立します。具体的にどの程度のサンプルサイズが必要であるかは場合によります。もとの確率分布の形が正規分布から離れているほど多くのサンプルサイズが必要となるのですが，一般的には $n \geq 30$ が目安とされているようです。したがって，今回の $n = 25$ はこの目安を下回っており，特に元の母集団が正規分布ではないため，標本平均の分布が正規分布に近似できるという保証がありません。その結果，正規分布を前提として計算した確率は，実際の確率とは大きく異なっている可能性があります。

また，中心極限定理に従うと，サンプルサイズが小さくなることで標本分布の分散が大きくなってしまいます。これにより，標本平均がある区間に収まる確率も小さくなります。ただ，これは「標本平均の推定値としての信頼性の低下」ではある一方で，「求めた確率そのものの信頼性」とは別の話とも言えます。

【解答例】

(a)約 86.64%
(b)信頼性は著しく低下する。サンプルサイズが 25 では中心極限定理が十分に機能する保証がなく，特に母集団が正規分布でないため，標本平均が正規分布に従うという仮定が成り立たない可能性が高いから。

(7) 母集団分布が正規分布に……

元のサンプルサイズを n とします。サンプルサイズを 2 倍，つまり $2n$ にしたときに標本分布の分散がどのようになるかを計算してみましょう。

標本平均の標本分布の分散

標本平均 $\bar{X}$ の標本分布の分散は $V[\bar{X}] = \dfrac{\sigma^2}{n}$ です。サンプルサイズを $2n$ にすると，新しい分散は

$$V'[\bar{X}] = \frac{\sigma^2}{2n} = \frac{1}{2}\frac{\sigma^2}{n} = \frac{1}{2}V[\bar{X}] \tag{8}$$

となります。

不偏標本分散の標本分布の分散

標本分散 s^2 の標本分布の分散は $V\left[s_x^2\right] = \dfrac{2(n-1)}{n^2}\sigma^4$ です。サンプルサイズを $2n$ にすると，新しい分散は

$$V'[s_x^2] = \frac{2(2n-1)}{(2n)^2}\sigma^4 = \frac{2n-1}{4(n-1)}\frac{2(n-1)}{n^2}\sigma^4 = \frac{2n-1}{4(n-1)}V[s_x^2] \tag{9}$$

となります。

【解答】

- 標本平均の標本分布の分散： $\dfrac{1}{2}$ 倍
- 不偏標本分散の標本分布の分散： $\dfrac{2n-1}{4(n-1)}$ 倍

第 9 章

(1) 9.1.1 項のはじめに例示した……

不偏性とは「推定量の期待値が真の値と等しくなる」性質です。そして，ある確率変数の期待値は，「母集団からランダムに 1 つ選ぶことを繰り返したときの平均値」と言えます。したがってこの推定量の期待値 $E[x_i]$ は，母集団分布にかかわらず母平均 μ に等しいので，**不偏性は持っています**。

有効性とは「他の不偏推定量と比べて分散が小さい」性質です。この推定量の分散 $V[x_i]$ は，母集団分布の分散 σ^2 です。これよりも分散が小さくなる標本統計量があるならば有効性がないことが示されますが，ここで標本平均 $\bar{x}$ の分散は，中心局限定理が適用できるならば $\dfrac{\sigma^2}{n}$ となります。したがって，$n \geq 2$ ならば必ず $\sigma^2 > \dfrac{\sigma^2}{n}$ となります。よって，標本平均よりは分散が大きくなるため，**有効性は持っていません**。

一致性は「サンプルサイズを大きくすると真の値に近づく」性質です。この推定量は，サンプルサイズ n を大きくしても，その標本の中からランダムに 1 つを選んで推定量としている以上，推

定量を決定しているサンプルサイズは常に 1 のままです。そして，上で見たように分散は σ^2 のままで減少しないため，真の値に近づく保証はありません。よって，**一致性は持っていません**。

【解答】

- **不偏性**：持っている。
- **有効性**：持っていない。
- **一致性**：持っていない。

(2) 日本人大学生の 1 日の……

母平均の不偏推定量

母平均の不偏推定量としては，**標本平均** $\bar{x}$ を使用するのがセオリーです。

$$\bar{x} = \frac{2+3+3+4+4+5+5+6+8+10}{10} = \frac{50}{10} = 5$$

母分散の不偏推定量

母分散の不偏推定量は，**不偏分散** $\dfrac{n}{n-1}s_x^2$ です。今回は，標本分散を求めるところから始める必要があるので，最後に偏差平方和を n ではなく $n-1$ で割って求めましょう。以上より，

もとの値 (x_i)	2	3	3	4	4	5	5	6	8	10
偏差 $(x_i - \bar{x})$	-3	-2	-2	-1	-1	0	0	1	3	5
偏差の 2 乗 $(x_i - \bar{x})^2$	9	4	4	1	1	0	0	1	9	25

不偏分散は

$$\text{不偏分散} = \frac{9+4+4+1+1+0+0+1+9+25}{9} = \frac{54}{9} = 6 \tag{10}$$

と求められます。

【解答】

- 母平均の不偏推定値: 5 時間
- 母分散の不偏推定値: 6

(3) サンプルサイズが十分に大きい……

不偏分散ではなく標本分散を用いても大きな問題がないと考えられる大きな理由は，標本分散が母分散の**最尤推定量**であり，最尤推定量はサンプルサイズが大きい場合に**一致性**や**漸近正規**

性という優れた性質を持つからです。

一致性　最尤推定量は一般に「一致性」を持っています。一致性とは，「サンプルサイズ n を無限に大きくしていくと，推定量は真の値（この場合は母分散 σ^2）に収束していく」という性質です。

標本分散と不偏分散の関係　不偏分散 (s^2) と標本分散 (S^2) の間には，$s^2 = \dfrac{n}{n-1}S^2$ という関係があります。そして，サンプルサイズ n が十分に大きいとき，係数 $\dfrac{n}{n-1}$ は限りなく 1 に近づきます（例 $n = 1000$ なら $\dfrac{1000}{999} \approx 1.001$）。これは，サンプルサイズが大きくなると，標本分散と不偏分散の値の差が実用上無視できるほど小さくなることを意味します。このために，標本分散と不偏分散のどちらを利用しても**大きな問題はない**，と言えるのです。

漸近正規性　さらに，左右推定量には，サンプルサイズが大きくなると，その標本分布が正規分布によって近似できる，という性質があります（この性質自体は正規母集団における不偏分散も持っています）。これにより，標準正規分布表を用いた簡単な計算や，「95% 信頼区間がおよそ ± 1.96 標準偏差で表せる」といったことが言えるようになり，実用上いろいろと楽なのです。

結論として，サンプルサイズが大きい場合，（最尤推定量である）標本分散も（不偏推定量である）不偏分散も，どちらも一致性によって真の母分散に近づいていき，両者の値も非常に近くなるため，どちらを用いても大きな問題はないと言えます。

【解答例】標本分散は母分散の最尤推定量であり，最尤推定量はサンプルサイズが大きいときに真の値に収束する「一致性」を持つから。また，サンプルサイズが大きいと，標本分散と不偏分散の値の差（$\dfrac{n}{n-1}$ 倍）が実用上無視できるほど小さくなるため。

(4) ある農場で，収穫した……
※訂正：(b) に書いてある「なお，糖度のばらつき（母集団標準偏差）は 0.8 度であるとわかっているものとします。」は，この問題全体に共通する前提としないと解けませんでした。いつか修正できたら修正します。

(a)

母平均を μ，母分散を 0.8^2 と置いたとき，求めたいのは $P(\mu - 0.1 \leq \bar{X} \leq \mu + 0.1) > 0.70$ となる最小の n です。まず，標準正規分布表から $P(-z \leq Z \leq z) = 0.70$ となる z を求めます。

$$P(-z \leq Z \leq z) = 2 \cdot P(0 \leq Z \leq z)$$
$$= 2 \cdot (0.5 - P(Z \geq z))$$

となることから，$P(Z \geq z) = 0.15$ となる z を探すと，最も近いのは $z \approx 1.04$ となります。したがって，任意の正規分布 $N(\mu, \sigma^2)$ に対して，$P(\mu - 1.04\sigma \leq X \leq \mu + 1.04\sigma) \approx 0.7$ となる，ということです。

そして，標本平均の標本分布は，正規分布 $N\left(\mu, \frac{0.8^2}{n}\right)$ と表すことができます。以上より，条件を満たすサンプルサイズは，

$$1.04\frac{0.8}{\sqrt{n}} < 0.1$$

という不等式で表されます。あとは，この不等式を解くと，$69.22 < n$ となります。ただし，サンプルサイズはかならず整数なので，答えとしては「$69.22 < n$ を満たす最小の整数」すなわち 70 個となります。

(b)

$n = 70, \bar{x} = 12, \sigma = 0.8$ を用いて信頼区間を計算していきます。今回は母分散既知の状況なので，標準正規分布を用います。この場合標本平均の標本分布は，正規分布 $N\left(12, \frac{0.8^2}{70}\right)$ と表せます。

求めたい確率は $P(\mu_L \leq \bar{X} \leq \mu_U) = 0.95$ となるような μ_L, μ_U の値です。この式を，先ほど求めた標本平均の標本分布にしたがって標準化させていくと，

$$P(\mu_L \leq \bar{X} \leq \mu_U) = P\left(\frac{\mu_L - 12}{0.8/\sqrt{70}} \leq Z \leq \frac{\mu_U - 12}{0.8/\sqrt{70}}\right) = 0.95$$

となります。また，標準正規分布表から，標準正規分布において

$$P(-1.96 \leq Z \leq 1.96) \approx 0.95$$

となります。以上の 2 つの式の対応する項をつなげると，

$$\begin{cases} \frac{\mu_L - 12}{0.8/\sqrt{70}} = -1.96 \\ \frac{\mu_U - 12}{0.8/\sqrt{70}} = 1.96 \end{cases}$$

となります。最終的に，95% 信頼区間は

$$[\mu_L, \mu_U] = \left[12 - 1.96\frac{0.8}{\sqrt{70}}, 12 + 1.96\frac{0.8}{\sqrt{70}}\right]$$
$$= [12 - 0.187, 12 + 0.187]$$
$$= [11.813, 12.187]$$

と求めることができます。

(c)

$n = 5, \bar{x} = 12, s_x = 0.8$ です。今回は，先ほどと異なり母分散未知，かつサンプルサイズが小

さいので t 分布を用いて信頼区間を求めます。母分散が既知の場合，母分散を用いて標準化した $\dfrac{\bar{X}-\mu}{\sigma/n}$ は標準正規分布に従う一方で，代わりに不偏分散（$\hat{s}_x^2$ と表しておきます）を用いて標準化した $\dfrac{\bar{X}-\mu}{\hat{s}_x/n}$ は，自由度 $n-1$ の t 分布に従いました。

以上を用いて信頼区間を求めていきましょう。まず，不偏分散は $\hat{s}_x^2 = \frac{5}{4}\cdot 0.8 = 1$ となります。これを用いて求めたい確率を標準化すると，

$$P(\mu_L \leq \bar{X} \leq \mu_U) = P\left(\frac{\mu_L - 12}{1/\sqrt{5}} \leq t \leq \frac{\mu_U - 12}{1/\sqrt{5}}\right) = 0.95$$

となります。また t 分布表からは，自由度 $df = n-1 = 4$ の t 分布において

$$P(-2.776 \leq Z \leq 2.776) \approx 0.95$$

と分かります。以上の 2 つの式の対応する項をつなげると，

$$\begin{cases} \frac{\mu_L - 12}{1/\sqrt{5}} = -2.776 \\ \frac{\mu_U - 12}{1/\sqrt{5}} = 2.776 \end{cases}$$

となります。最終的に，95% 信頼区間は

$$\begin{aligned}[\mu_L, \mu_U] &= \left[12 - 2.776\frac{1}{\sqrt{5}}, 12 + 2.776\frac{1}{\sqrt{5}}\right] \\ &= [12 - 1.241, 12 + 1.241] \\ &= [10.759, 13.241]\end{aligned}$$

と求めることができます。

【解答】

(a)70 個
(b)**点推定値**: 12 度，　95% **信頼区間**: [11.813, 12.187]
(c)**点推定値**: 12 度，　95% **信頼区間**: [10.759, 13.241]

(5) ある製品の不良率を……

標本比率の信頼区間を求める問題です。与えられた値は $n = 100, x = 5$ です。まず**点推定値**は，標本不良率がそのまま不偏推定量として利用可能です。 したがって，$\hat{p} = \dfrac{x}{n} = \dfrac{5}{100} = 0.05$ です。

また今回はサンプルサイズが $n = 100$ と十分に大きいため，**95% 信頼区間**は，正規近似を用いて計算が可能です。一つ一つの製品の出来上がりについて，母集団分布をベルヌーイ分布 $B(1, p)$ とおくと，中心極限定理より，標本平均（標本比率）の標本分布は

$$\bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

と表すことができます。ここで，(4)-(b) で見たように，標準正規分布に基づいて計算される 95% 信頼区間は，**（標本平均）±1.96（標準誤差）**で表されることから。

$$\begin{aligned}
\text{信頼区間} &= \hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\
&= 0.05 \pm 1.96\sqrt{\frac{0.05(0.95)}{100}} \\
&\approx 0.05 \pm 1.96 \cdot 0.0218 \\
&\approx 0.05 \pm 0.0427
\end{aligned}$$

となります。

【解答】

- **点推定値**: 0.05 (5%)
- 95% **信頼区間**: [0.0073, 0.0927] （**約** 0.7% **から** 9.3%）

(6) 母集団分布が正規分布に従うとき……

95% 信頼区間の幅は，標本平均の標準偏差の $(2 \cdot 1.96 =)3.92$ 倍です。そして，標本平均の標準偏差は $\frac{\sigma}{\sqrt{n}}$ と表されました。したがって，95% 信頼区間の幅は，サンプルサイズ n に対して $3.92\frac{\sigma}{\sqrt{n}}$ と表されます。

ここでサンプルサイズを n から $2n$ へと 2 倍にすると，幅は

$$3.92\frac{\sigma}{\sqrt{2n}} = \frac{1}{\sqrt{2}} \cdot 3.92\frac{\sigma}{\sqrt{n}}$$

と，もとの信頼区間の幅の $\frac{1}{\sqrt{2}}$ 倍になります。

【解答】 $\boldsymbol{\frac{1}{\sqrt{2}}}$ **倍**（約 0.707 倍）になる。

(7) ある世論調査では，……

(5) で見たように，標本比率の信頼区間は，標準正規分布を用いて算出されます。そして，そ

の区間は（**標本平均**）$\pm$**1.96**（**標準誤差**）で表されます。したがって，95% 信頼区間の幅が 6 ということから，1.96（標準誤差）が 0.3 であることがわかります。標本比率の標準誤差が $\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ であることから，

$$\begin{aligned} 0.03 &= 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ &= 1.96\sqrt{\frac{0.42(1-0.42)}{n}} \end{aligned}$$

という等式が成り立ちます。あとは，これを n について解けば

$$\begin{aligned} \sqrt{n} &= \frac{1.96}{0.03}\sqrt{0.42 \cdot 0.58} \approx 65.33 \cdot 0.4936 \approx 32.24 \\ n &\approx (32.24)^2 \approx 1039.4 \end{aligned}$$

と求めることができます。

このように，信頼区間の中には「どの程度のサンプルサイズで推定した結果であるか」という情報も含まれていることから，推定の精度を示す指標の一つとして報告することが推奨されているのです。

【解答】およそ **1039 - 1040 人** を対象に行われたと推定できる。

(8) 区間推定において，……

結論としては，**99% 信頼区間**のほうが，95% 信頼区間よりも幅が広くなります。というのも「95% 信頼区間」とは，「この区間が本当に知りたい母数を捉えている」という主張に対する確信度（そのような標本が得られる割合）が 95% であることを意味します。同様に「99% 信頼区間」は確信度が 99% です。より高い確信度（99%）で「真の母数を捉えている」あるいは「より多くの標本できちんと母数を捉える区間が生成できる」と主張するためには，**予測を外すリスクをより小さくする**必要があります。そのためには，より「安全」な予測，すなわち**より広い範囲をカバーして，真の母数を「逃さない」ようにする**必要があります。魚を捕る網に例えるなら，「99% の確率で魚を捕まえる網」は「95% の確率で捕まえる網」よりも，より大きな網でなければならないのと同じと言えるでしょう。

【解答例】

- **幅が広いのは**：99% 信頼区間
- **理由**：「99%」というより高い確信度で母数を含む区間を作れるようにするためには，真の値を外しにくくする（逃さないようにする）必要があり，そのためにはより広い範囲をカバーする必要があるから。

第10章

(1) あるスマートフォンアプリの……

(a)

仮説検定では，多くの場合まず「基準となる値」を帰無仮説として立てます。この場合，開発者が検証したいのは「目標である 30 分に到達しているか？」ということです。したがって，以下のようにそれぞれ仮説を設定することができます。

- **帰無仮説 (H_0)**：アプリの平均利用時間は 30 分である ($\mu = 30$)。
- **対立仮説 (H_1)**：アプリの平均利用時間は 30 分より短い ($\mu < 30$)。

(b)

片側検定

- **メリット**：関心のある方向（今回は「30 分より短い」）の変化を検出する力（**検出力**）が，両側検定よりも高くなります。
- **デメリット**：設定した方向と逆の変化（平均利用時間が想定外に 30 分を大幅に上回っているなど）を検出することができません。それは良いとしても，両側検定と比べると，さほど強くない変化であっても有意になってしまう可能性が高まることが，場合によってはデメリットになり得ます。

両側検定 (対立仮説を $H_1 : \mu \neq 30$ とする)

- **メリット**：平均が 30 分からズレてさえいれば，それが上であっても下であっても検出でき，また棄却域も上下で分割されるため，より保守的で客観的な検定と言えます。
- **デメリット**：「30 分より短い」という特定の方向の変化に対する検出力は，片側検定に比べて低くなります。

今回の開発者の目的は「目標である 30 分を**下回っていないか**」を検証することであり，関心の方向が明確です。したがって，その方向の変化をより高い感度で捉えられる**片側検定**を採用することもあり得るでしょう。一方で，保守的な態度を示す（より顕著な差が出なければ有意と認めない）ならば，両側検定を使用しても全く問題はありません。

【解答例】

(a)**帰無仮説 (H_0)**: $\mu = 30$ (平均利用時間は 30 分)， **対立仮説 (H_1)**: $\mu < 30$ (平均利用時間は 30 分未満)

(b)**片側検定**を用いる。「目標を下回っているか」という関心の方向が明確であり，その変化を検出する力が高いため。

【別案】：**両側検定**を用いる。そのほうがより保守的に結果を判断できると考えられるため。

(2) 以下の状況において，……

この検定の仮説は，帰無仮説（H_0）：新薬に効果はない， 対立仮説（H_1）：新薬に効果がある，となります。

(a)

「実際には効果がある」ということは，「帰無仮説 (H_0) が誤っている」状況です。その誤った帰無仮説を棄却せず，「効果なし」と結論付けてしまったため，これは「**誤った帰無仮説を採択してしまう誤り**」，すなわち**第 2 種の誤り**にあたります。

(b)

第 2 種の誤りの確率（β）を減らすことは，検定の**検定力**（$1-\beta$）を高めることと同じです。検定力が向上する条件を踏まえると，考えられる主な対策は以下の通りです。

- **サンプルサイズを大きくする**：調査対象者の数を増やすことが最も基本的な対策です。必要最低限の検出力を得るために必要なサンプルサイズを，データ収集前に決めておくのがおすすめです。
- **有意水準を大きく（緩く）する**：例えば，有意水準 α を 5% から 10% に変更します。これにより帰無仮説を棄却しやすくなります。ただ，これは反対に，実質的にあまり効果がない場合でも「効果あり」と判断してしまう（第 1 種の誤り）可能性を高めてしまい，また見方によっては，分析結果を恣意的に操作しているようにも感じられるかもしれないので，基本的には行うべきではないでしょう。
- **測定の精度を上げる**：薬効を何で測定するかにもよりますが，基本的には標準誤差が小さくなる，つまりデータのばらつきを小さくすることで，帰無分布の幅が小さくなるために，小さな効果でも検出しやすくなります。

(c)

(b) の対策には，それぞれ以下のような注意点（トレードオフ）があります。

- **サンプルサイズを大きくする場合**：調査にかかる**費用，時間，労力が大幅に増加**します。
- **有意水準を大きくする場合**：第 2 種の誤りを減らせる一方で，**第 1 種の誤り（実際には効果がないのに「効果あり」と結論付けてしまう誤り）の確率が増加**してしまいます。第 1 種と第 2 種の誤りはトレードオフの関係にあります。
- **測定の精度を上げる場合**：測定法の改善が必要になるため，かなりの労力を要するでしょう。そして，そもそも測定精度が上げられるのであれば，はじめからその方法を採用していた可能性が高いと考えられます。

以上をまとめると，検定力を高めるためには，やはり**サンプルサイズを大きくする**のが基本だということです。

【解答例】

(a)**第 2 種の誤り**
(b)サンプルサイズを大きくする（有意水準を大きくする，測定の精度を上げる）。
(c)サンプルサイズを増やすとコストが増大する。（有意水準を大きくすると，第 1 種の誤りの確率が増加するというトレードオフがある。）

(3) p 値に関する以下の解釈……

ここに示した 3 つの解釈は，よく見られる誤解の例です。以下で一つずつ説明していきましょう。

(a)

これは p 値の最も典型的な誤解です。p 値は「**もし帰無仮説が正しいとした場合に，観測されたデータか，それ以上に極端なデータが得られる確率**」を意味します。すなわち p 値は，帰無仮説が正しいという**仮定の上で**計算された確率であり，「帰無仮説が誤っている確率」そのものを教えてくれるものではありません。

(b)

p 値は，結果の「**統計的な有意性（珍しさ）**」の度合いを示すものであり，実際の「**効果の大きさ（effect size）**」を示すものではありません。強いて言うなら，「**観測された**効果の大きさ」が大きいほど p 値は小さくなりますが，それ以外にもサンプルサイズの影響も受けるため，やはり p 値のみによって効果の大きさを比べることはできません。

(c)

有意水準を 5% とした場合，p 値が 0.06 であれば「帰無仮説を棄却できない」という結論になります。しかし，これは「帰無仮説が正しいことを証明した」わけでは決してありません。単に「**帰無仮説が誤っていると結論付けるほどの十分な証拠が，今回のデータからは得られなかった**」ことを意味するだけです。そもそも背理法的に仮説検定を行う以上は，「帰無仮説が正しい」と積極的に主張することはできないのです。

【解答例】

(a)**正しくない**。p 値は「帰無仮説が正しいと仮定した上での，観測されたデータと同等以上に極端なデータが出現する確率」であり，「帰無仮説が誤っている確率」を表すものではない。
(b)**正しくない**。p 値は帰無仮説のもとで計算された「統計的な（手続き上の）有意性」を示すものであり，「効果の大きさ」を示すものではない。

(c)**正しくない**。「帰無仮説を棄却できない」ことは「帰無仮説が正しい」ことを意味しない。単に「帰無仮説を棄却するほどの十分な証拠が得られなかった」ことを意味するに過ぎない。

(4) あるウェブメディアが，……

これは**多重比較**に関するです。

(a)

- 1 回の検定で誤って「有意差あり」と判断する確率（第 1 種の誤り）は，有意水準 $\alpha = 0.05$ です。
- 1 回の検定で正しく「有意差なし」と判断する確率は $1 - 0.05 = 0.95$ です。
- もしもすべての検定が独立であるとするならば，20 回とも正しく「有意差なし」と判断する確率は $(0.95)^{20} \approx 0.358$ です。
- したがって，「少なくとも 1 回が有意差ありと判断してしまう」確率は，この余事象です。

$$1 - (0.95)^{20} \approx 1 - 0.358 = 0.642$$

(b)

この編集長の判断には，主に以下の 2 つの統計的な問題点があります。

1. **多重比較の問題を無視している**：(a) で計算したように，実際には効果がない施策を多数検定した場合，**高い確率で，偶然による「当たり」（偽の陽性結果）が出てしまいます**。編集長が見ている結果は，この偶然の産物である可能性が非常に高いでしょう。もしも同じ手続きでデータを再度集めた場合には，きっと別の種類において $p < 0.05$ となる可能性が高いと想像できます。
2. **再現性を確認していない**：ということで，たった 1 回の検定結果だけを根拠に全社的な方針を変更するのは早計です。意思決定の影響範囲が大きくなるほど，その採用には慎重になるべきです。したがって，その結果が他の記事や異なる状況でも再現できるかどうかを，**改めて別のデータで追試・検証**するべきです。そして，可能ならば「なぜ『？』で終わるタイトルはクリック率が高くなるのか」という因果関係のメカニズムについても検討すると，結果の説得力が高まるでしょう。

【解答例】

(a)約 **64.2%**
(b)**多重比較の問題**を無視している。多数の検定を行えば偶然有意な結果が出ることはあり，今回の結果がその偶然の産物である可能性が高い。また，1 回の結果だけで判断せず，再現性を確認するべき。

第11章

(1) 以下の 3 つの状況について，……

(a)

1 つの標本（あるクラス 25 人）の平均値を，比較対象となる母平均（全国平均 65 点）と比較しています。この例では母集団（全国の学生）の分散が不明であるため，標本から計算した不偏分散を用いて検定を行う **1 標本の t 検定**が適切です。

(b)

同じユーザーに対して「利用前」と「利用後」の 2 つの時点でデータを取得しており，2 つのデータ群に明確な対応関係があります。この場合，各ユーザーのスコアの変化量（差）を算出し，その変化量の平均が 0 と有意に異なるかどうかを検定する**対応のある t 検定**が適切です。

(c)

「文系学生」と「理系学生」という，互いに独立した 2 つのグループ（標本）の平均値を比較しています。2 つのグループの学生間には (b) のような対応関係はありません。そのため，**対応のない t 検定**が適切です。

【解答】

(a)**1 標本の t 検定**：1 つの標本の平均値を，基準となる母平均と比較するため。
(b)**対応のある t 検定**：同じ対象者に対する，施策前後の変化を比較するため。
(c)**対応のない t 検定**：互いに独立な 2 つのグループの平均値を比較するため。

(2) ある従業員の業務効率化……

これは，同じ対象者における前後比較なので，「対応のある t 検定」を行います。まず，帰無仮説と対立仮説はそれぞれ

【帰無仮説】 $\mu_d = 0$（プログラム前後の母平均は等しい）
【対立仮説】 $\mu_d > 0$（受講後の母平均のほうが大きい）

となります。

検定統計量 t を求めるために，まずは x_d の不偏分散およびその正の平方根 $\hat{\sigma}_d$ を計算します。今回は標本標準偏差が 20，サンプルサイズが $n = 16$ なので，

$$\hat{\sigma}_d = \sqrt{\hat{\sigma}_d^2} = \sqrt{\frac{16}{15} \cdot 20^2} \approx \sqrt{426.67} \approx 20.66$$

これを用いると，検定統計量 t は

$$t = \frac{\bar{X}_d - 0}{\frac{20.66}{\sqrt{16}}} = \frac{12}{5.165} \approx 2.323$$

となります。

続いて棄却域を求めます。有意水準 $\alpha = 0.05$ の両側検定であることから，t 分布表より，棄却域は，自由度 $df = n - 1 = 15$ のときの上側 2.5% 点である $t_{0.025}(15) = 2.131$ となります（$|t| \geq 2.131$）。したがって，先ほど計算した検定統計量 $t = 2.323$ は，2.131 より大きいので，帰無仮説は棄却されます。

【解答】検定統計量 $t = 2.323$ であり，自由度 15，有意水準 5% の両側検定における棄却域 ($|t| \geq 2.131$) に入るため，帰無仮説は棄却される。結論として，**このプログラムには効果があると言える**。

(3) 1 標本の母平均の検定について，……

その明確な理由は，t 検定よりも標準正規分布に基づく検定（z 検定）の方が，**統計的検出力が高い**からです。というのも t 検定は，母分散 σ^2 が未知であるという「不確かさ」を考慮して作られた検定であり，その分だけ正規分布よりも棄却域が狭く（棄却のハードルが少し高く）なっています。これは，t 分布表において自由度が小さいほど，示されている値が 1.96 よりも大きくなっていることに表れています。

一方母分散が既知であれば，この「不確かさ」は存在しません。したがって，母分散が既知であるにも関わらず t 検定を使うのは，必要以上に保守的な判断をしていることになります。z 検定の方が t 検定よりも棄却域がわずかに広いため，帰無仮説を棄却しやすくなり，検出力が高まります。母分散が既知という貴重な情報を持っているのに，それを使わずに t 検定を行うのは，検定のパワーを自ら落としてしまう行為なのです。実際には「母分散が既知」という状況があまり見られないのも事実ですが，必要に応じて使い分けられるようになりましょう。

【解答】標準正規分布に基づく検定（z 検定）の方が，t 検定よりも**統計的検出力が高い**から。母分散が既知であれば，その不確かさを考慮した t 検定を用いる必要はなく，より検出力の高い z 検定を用いるべきである。

(4) ある大学の講義で，……

これは，2 つの独立したグループの平均値を比較する「対応のない t 検定」の問題です。まず，帰無仮説と対立仮説はそれぞれ

【帰無仮説】 $\mu_A = \mu_B$（2 つのグループの母平均は等しい）

【対立仮説】 $\mu_A \neq \mu_B$（2 つのグループの母平均は異なる）

となります。

検定統計量 t を求めるために，まず，2 群の分散をプールした分散 $\hat{\sigma}^2$ を計算します。

$$\hat{\sigma}^2 = \frac{(20-1)\cdot 15^2 + (40-1)\cdot 12^2}{20+40-2} = \frac{4275+5616}{58} \approx 170.53$$

これを用いると，検定統計量 t は以下のように計算できます。

$$t = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}} = \frac{(72-78)-0}{\sqrt{170.53\left(\frac{1}{20}+\frac{1}{40}\right)}} = \frac{-6}{\sqrt{12.79}} \approx -1.678$$

続いて棄却域を求めます。有意水準 $\alpha = 0.05$ の両側検定であることから，棄却域は，自由度 $df = n_A + n_B - 2 = 58$ のときの上側 2.5% 点です。t 分布表には $df = 58$ のときの値はないため，最も近い $df = 60$ のときの値で代用します[7]は，$t_{0.025}(60) \approx 2.001$ となります（$|t| \geq 2.001$）。したがって，先ほど計算した検定統計量 $t \approx -1.678$ は，-2.001 より大きいので，帰無仮説は棄却されません。

【解答例】検定統計量 $t \approx -1.678$ であり，自由度 58，有意水準 5% の両側検定における棄却域（$|t| \geq 2.001$）に入らないため，帰無仮説は棄却されない。結論として，**新しい教授法に効果があるとは言えない**。

(5) 対応のない 2 標本の……

検定力は，検定統計量の値が大きくなるほど高くなります。対応のない t 検定の統計量は，分子が平均値の差，分母が標準誤差です。

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}}$$

観測される平均値の差 $\bar{X}_A - \bar{X}_B$ とプールされた分散 $\hat{\sigma}^2$ が一定だとすると，t の値を大きくするには，分母の標準誤差をできるだけ小さくする必要があります。これは，$\left(\frac{1}{n_A}+\frac{1}{n_B}\right)$ の部分を最小化することと同じです。

サンプルサイズの合計 $N = n_A + n_B$ が一定という制約のもとで $\left(\frac{1}{n_A}+\frac{1}{n_B}\right)$ を最小にするのは，$n_A = n_B$ のときです。例えば，2 群のサンプルサイズの合計が $N = 40$ の場合，

- **均等** $(n_A = 20, n_B = 20)$: $\frac{1}{20}+\frac{1}{20} = 0.1$

[7] 自由度を実際よりも少し大きく見積もる代用では，棄却域の幅が僅かに広くなります。そのため，棄却域の端が検定統計量 t に相当近い場合には，統計ソフトウェアなどを用いて厳密に $df = 58$ のときの棄却域を確認したほうが良いかもしれません。

- **不均等** $(n_A = 10, n_B = 30)$: $\dfrac{1}{10} + \dfrac{1}{30} \approx 0.133$

このように，サンプルサイズが不均等になるほど標準誤差は大きくなり，t 値は小さくなって検定力は低下します。したがって，同じ総サンプルサイズであれば，各グループの数を均等に近づけるほど，検定力は高くなります。

【解答例】検定統計量の分母である標準誤差が，$\left(\dfrac{1}{n_A} + \dfrac{1}{n_B}\right)$ という項に比例するため。合計サンプルサイズが一定のとき，この項はサンプルサイズが均等な $(n_A = n_B)$ 場合に最小となる。標準誤差が最小になることで t 値が最大化され，検定力が高くなるから。

(6) 基本的に全ての統計的仮説検定は，……

無作為抽出でない場合に生じる問題

無作為抽出ができていない場合，得られた標本が母集団を代表していない「**偏った標本（バイアスのある標本）**」である可能性が高くなります。そもそも仮説検定の枠組み（帰無仮説が正しいときの検定統計量の確率分布）は，母集団からの無作為抽出が行われた際に観測されうる検定統計量の確率分布です。そのため，偏った標本で仮説検定を行うと，そもそもの検定統計量の標本分布が理論的な分布からズレてしまい，設定した有意水準（例えば 5%）が全く意味をなさなくなります。したがって，第 1 種・第 2 種の誤りの確率が制御不能になってしまいます。

無作為抽出できていない標本分布の例

例えば，大学の満足度調査を図書館前で行った場合，大学の図書館を日頃から利用するような，満足度の高いと考えられる学生が集まりやすいと考えられます。その結果，標本平均の標本分布の中心は，真の母平均よりも高い値に**シフト**してしまいます。その結果，例えば「A 大学の満足度は例年と変わらない」といった帰無仮説を検証しようとした場合，実際には帰無仮説が正しいとしても，標本の偏りによって，帰無仮説を棄却できるほど「満足度が高い」という結果が出る（第 1 種の誤りを犯す）確率が非常に高くなってしまいます。

【解答例】標本に偏りが生じ，検定結果を母集団全体に一般化できなくなる。また，検定統計量が理論上の分布に従わなくなるため，第 1 種・第 2 種の誤りを犯す確率が制御できなくなる。

(7) 以下のような行為は ……

この行為が「p-hacking」として禁じられている理由は，**検定の前提を崩し，第 1 種の誤り（本**

当は相関がないのに，偶然の変動を捉えて「相関がある」と誤って結論づけてしまうこと）の確率を高めてしまうからです。

仮説検定における有意水準 5% という基準は，「**もし 1 回だけ検定を行うなら**，第 1 種の誤りを犯す確率は 5% に抑えられる」という約束事のもとで成り立っています。しかし，この研究者は，最初の検定の**結果を見てから**，「もう少しで有意になりそうだから」という理由でサンプルを追加するという意思決定をしています。このように，データや分析の途中結果を「つまみ食い」しながら，分析方針（この場合はサンプルサイズの決定）を後付けで変更すると，自分の望む結果が出るまで試行を繰り返すことができてしまいます。もしも帰無仮説が本当に正しいならば，データの偶然のばらつきによって，どこかのタイミングで p 値が 0.05 を下回る可能性は十分にあり，その「都合の良い偶然」だけを最終結果として報告することができてしまいます。明らかにこれは恣意的な結果の選択（チェリーピッキング）です。専門的な言い方をするならば，第 1 種の誤りの確率を % よりも高くすることで，帰無仮説を無理やり棄却しようとしている，というわけで，このような行為は禁止されているのです。

【解答】検定結果を途中でのぞき見て，その結果に応じてサンプルサイズを追加するなどの分析方針の変更を行うと，検定を複数回行っているのと実質的に同じになり，第 1 種の誤り（偶然有意な結果を得てしまうこと）の確率が，設定した有意水準（例：5%）よりも大幅に高くなってしまうから。

(8) ある研究で，「2 つの教授法の……

デザイン (1)：同じ学生に両方試す（対応のあるデザイン）

- **統計的メリット**：個人差が完全に統制されるため，対応のないデザインと比べて**検定力が非常に高くなります**。
- **実験デザイン上のデメリット**：2 つの教授法を順番に実施すると，**順序効果**（1 番目に受けた教授法の影響が 2 番目に残ってしまう，2 番目の教授法を受ける頃には疲れてしまって集中できていない，など）が生じる可能性があります。これを打ち消すには，例えば受ける教授法を「A → B の順」と「B → A の順」にするグループを同数用意する「カウンターバランス」が必要になり，デザインが複雑になります。

デザイン (2)：2 グループに分ける（対応のないデザイン）

- **統計的デメリット**：グループ間の個人差（例えば，たまたま教授法 A を受けたグループのほうが基礎学力が平均的に高かった，など）がノイズとなるため，デザイン (1) に比べて**検定力は低くなります**。
- **実験デザイン上のメリット**：順序効果を心配する必要がなく，デザインがシンプルです。
- **実験デザイン上のデメリット**：「統計的デメリット」にも繋がりますが，検定の妥当性が，グループのランダムな割り当てに大きく依存します。そのため，割当がランダムかつもともとの学力が均等でないと，グループ間の元々の差なのか教授法の効果なのか区別がつかなく

なります。

【解答】

- **デザイン (1)（対応あり）**：
  ・メリット：個人差が統制され，**検定力が高い**。少ないサンプルで済む。
  ・デメリット：**順序効果**（学習効果や疲労効果）が発生するリスクがある。
- **デザイン (2)（対応なし）**：
  ・メリット：順序効果がなく，デザインがシンプル。
  ・デメリット：個人差がノイズとなり**検定力が低い**。効果の妥当な評価には**無作為な割り当て**が不可欠。

(9) 10.2.2 節では，一般的に……

検定力が高ければ，同じ効果を検出するために必要なサンプルサイズは小さくて済みます。そして，他の条件が同じであれば，**片側検定は両側検定よりも検定力が高くなります**。

検定力とは，「対立仮説が真であるときに，正しく帰無仮説を棄却できる確率」です。

- **両側検定**：有意水準 α（例えば 5%）を両側の裾に分割するため（左右それぞれ 2.5%），棄却のハードルとなる臨界値はより極端な値になります。
- **片側検定**：有意水準 α を関心のある片側の裾に集中させるため（例えば右側 5%），臨界値は両側検定ほど極端ではない値になります。

ということで，実際に 11.6 節の数値例について，片側検定の場合に検定力 90% に到達するために必要なサンプルサイズを求めてみましょう。これは，(11.39) 式をもとに計算可能です。片側検定ということで，左側の棄却域（右辺の第 1 項）がなくなり，右辺の棄却域が「標準誤差の 1.96 倍」から「標準誤差の **1.64 倍**」に変わるため，

$$
\begin{aligned}
1-\beta &= P\left[\bar{X} \geq \left(500 + 1.64\frac{5}{\sqrt{n}}\right)\right] \\
&= P\left[\frac{\bar{X}-500.5}{5/\sqrt{n}} \geq \frac{\left(500 + 1.64\frac{5}{\sqrt{n}}\right) - 500.5}{5/\sqrt{n}}\right] \\
&= P\left[Z \geq 1.64 - 0.5\frac{\sqrt{n}}{5}\right] \geq 0.9
\end{aligned}
\tag{11}
$$

という不等式が成り立ちます。あとは，(11.41) 式に当てはめて

$$
1.64 - 0.5\frac{\sqrt{n}}{5} < -1.28
$$

という不等式を解くと，$n > 852.64$ となります。この値は，両側検定のときに必要な $n > 1049.76$ と比べて確かに小さくなっています。

【解答例】他の条件が同じ場合，片側検定は両側検定よりも**検定力が高くなる**。なぜなら，片側検定は有意水準を片側の裾に集中させるため，帰無仮説を棄却するためのハードルが両側検定よりも低くなるからである。11.6 節の数値例では，片側検定の場合に必要なサンプルサイズは $n > 852.64$ となり，両側検定のときの値よりも小さくなっていることがわかる。

(10) ある市内の小学 6 年生……

(a)

最も大きな問題点は，「**相関関係と因果関係を混同している**」ことです。第 3 章の練習問題でも確認しましたが，これは統計的仮説検定であっても変わりません。統計的に有意な相関は，「運動時間が長い生徒は BMI が低い傾向にある」という関連性（相関関係）の有意性を示しているだけであり，「運動時間を増やすこと（原因）が BMI を低下させる（結果）」という因果関係を証明するものではありません。具体的には，以下の可能性が見過ごされています。

- **逆の因果関係**：BMI が低い（痩せている）から，体を動かすのが苦手ではないために，結果として運動時間が長くなる。
- **第三の変数の影響（交絡）**：家庭の食生活や親の健康意識といった第三の変数が，「運動時間」と「BMI」の両方に影響を与えている（疑似相関）。

(b)

因果関係に迫るためには，以下のようなアプローチが考えられます。これらは，第 3 章で見たものと概ね同じです。仮説検定の結果を盲信しないように気をつけましょう。

1. **介入研究（ランダム化比較試験）の実施**：子供たちを無作為に 2 群に分け，一方に運動プログラムを課し，もう一方は何もしない。その後 BMI の変化を比較する。
2. **縦断的調査（パネルデータ分析）**：同じ子供たちを長期間追跡調査し，「運動時間の変化」がその後の「BMI の変化」に繋がっているかといった時間的な前後関係を分析する。
3. **交絡因子の測定と統計的統制**：食事内容，保護者の BMI など，考えられる交絡因子のデータを収集し，**重回帰分析**などの手法でそれらの影響を統計的に統制した上で，運動時間と BMI の純粋な関係性を評価する。

【解答】

(a)**相関関係と因果関係の混同**という問題点がある。統計的仮説検定の結果は，因果関係の存在を証明するものではない。具体的には，「逆の因果関係」や「第三の変数（交絡）」の可能性を考慮していない。

(b)(1) 介入研究（ランダム化比較試験），(2) 縦断的調査，(3) 交絡因子を統制した重回帰分析などのアプローチが考えられる。

(11) ある製薬会社が，……

(a)

今回の仮説検定で得られた p 値の意味を丁寧に記述すると，「**もし帰無仮説（新薬とプラセボの効果に差はない）が正しいとしたら，今回観測された 0.04 日という平均差か，それ以上に大きな差が，全くの偶然によって生じる確率は 2%（p=0.02）である**」となります。そしてこの確率が事前に定めた有意水準（通常 5%）より小さいため，「これは単なる偶然とは考えにくい」と判断し，帰無仮説を棄却して「新薬とプラセボの効果には（偶然とは言えない）差がある」と結論付けます。「統計的に有意」とは，**帰無仮説を棄却できるようなデータが得られた**という意味でしかありません。

(b)

この薬は，「**効果的な薬」とは評価されない可能性が高い**でしょう。この結果は，「**統計的有意性**」と「**実用的・臨床的重要性**」が全く別のものであるという典型的な例です。p 値が 0.02 であることから，この薬には「効果がゼロではない」ことが統計的には示唆されています。しかし，その効果の大きさ（**効果量**）は「平均 0.04 日」，つまり 1 時間にも満たない症状期間の短縮です。ほとんどの患者にとって，その差を体感することすらできず，実用的な意味はほぼないと言えるでしょう。この臨床試験はサンプルサイズが非常に大きかった（大規模臨床試験の）ため，このような実務的には無意味なほど小さな差でも，統計的には有意な結果として検出できてしまったのだと推測されます。p 値だけを見て判断を誤ってはならず，常に効果の大きさもセットで評価することが重要です。

【解答例】

(a)p 値は「もし新薬に効果がないとしたら，観測された 0.04 日以上の差が偶然生じる確率は 2% である」ことを意味する。「統計的に有意」とは，これが十分に小さい確率であるため，偶然ではないと判断し「差がある」と結論付けることを示す。

(b)「効果的な薬」とは評価しない。統計的には有意であっても，症状が 0.04 日（約 1 時間）短くなるという平均効果は**実用的にほとんど意味がない**ため。統計的有意性と実用的な重要性は別であり，効果量も合わせて評価する必要がある。

第 12 章

(1) 第 12 章の最初に挙げた……

ある事象（ここでは商品の提供）が発生するまでの時間のデータがある場合には，指数分布を当てはめられるケースが多く見られます。これ自体はよくあることなのですが，他の確率分布と

同様に，**その確率分布の背後にある仮定が満たされているか**をよく考えることが，統計モデリングにおける重要なポイントです。

指数分布を仮定した場合のデータ生成メカニズム

「提供にかかる時間」が指数分布に従うと仮定することは，「**ある時点まで商品が提供されていなくても，そこからさらに待つ時間は，それまでの待ち時間とは無関係に決まる**」という仮定[8]を置くことになります。言い換えると，「商品の提供」というイベントが，どの瞬間においても一定の確率で発生する（一瞬一瞬をベルヌーイ試行とみなしている）ことを意味します。

仮定が現実的に満たされているかについての考察

この「無記憶性」の仮定は，ファストフード店の実際のオペレーションを考えると，**現実的には満たされていない可能性が高い**と考えられます。

- **最低提供時間の存在**：注文を受けてから調理するには物理的に最低限の時間が必要です。しかし，指数分布はパラメータの値によらず 0 で確率密度が最大になります。つまり，0 に近い時間で発生する確率が最も高いため，この点で現実と乖離しています。
- **待ち時間と提供完了確率の関係**：現実には，注文から時間が経てば経つほど，商品は完成に近づいているはずです。つまり，「長く待てば待つほど，次の瞬間には提供される確率が高まる」と考えるのが自然であり，確率が一定であるとする指数分布の仮定と矛盾します。

したがって，「提供にかかる時間」のモデルとして指数分布をそのまま用いるのは，現実のプロセスを過度に単純化しすぎている可能性があります。

とはいえ，実際の分析では，他の変数を組み合わせたり（一般化線形モデルと呼ばれます），あるいは仮定を多少無視して分析を行うことも多々あります。例えば，上に挙げた「待ち時間と提供完了確率の関係」は，「ある病気にかかってから死亡するまで」などの時間のデータに対しても同じ矛盾が指摘できますが，このデータに対しては指数分布を当てはめる分析（生存時間分析）がメジャーです。このあたりは，各分野のセオリーなども影響するため，一概に「過程が満たされないから○○分布は使えない」というように決め打ちで考えるのも早計かもしれません。

【解答】

- **仮定**：「無記憶性」を仮定している。これは，ある時間待った後，さらに待つ時間はそれまでの待ち時間とは無関係である，つまり提供完了の発生率が常に一定であるという仮定。
- **考察**：**満たされていない可能性が高い**。現実には調理に必要な最低時間が存在し，また時間が経つほど提供される確率は高まっていくと考えられるため，無記憶性の仮定は現実と合わない。

[8] 指数分布（および幾何分布）がもつこの性質は「**無記憶性**」と呼ばれます。

(2) あるチョコレート菓子の企業で，……

まず，回帰係数の点推定値については，標本で計算された回帰係数が不偏推定量であるため，これをそのまま用いて $\hat{\beta}_0 = 0.8, \hat{\beta}_1 = 0.4$ とします。

(a) 回帰係数の 95% 信頼区間

1. β_1 **の信頼区間** $\hat{\beta}_1$ の標本分布は，

$$\hat{\beta}_1 \sim N(\left(\beta_1, \frac{\sigma_\varepsilon^2}{n s_x^2}\right)$$

となりました（12.10 式）。ただし，σ_ε^2 が未知であるため，不偏分散を計算します。これは，(12.14) 式より

$$\hat{\sigma}_\varepsilon^2 = \frac{n}{n-2} s_e^2 = \frac{10}{8} \cdot 1.2 = 1.5$$

と求められます。したがって，β_1 の標本分布の標準偏差は

$$\sqrt{\frac{\hat{\sigma}_\varepsilon^2}{n s_x^2}} = \sqrt{\frac{1.5}{10 \cdot 3.2}} \approx 0.217$$

となります。そして，信頼区間は自由度 $n-2=8$ の t 分布に基づいて計算されます。t 分布表から，自由度 8 の t 分布における上側確率が 2.5% となる値が $t = 2.306$ であるとわかるため，95% 信頼区間は

$$[0.4 - 2.306 \cdot 0.217, 0.4 + 2.306 \cdot 0.217] = [0.4 - 0.500, 0.4 + 0.500] = [-0.1, 0.9]$$

となります。

2. β_0 **の信頼区間** $\hat{\beta}_0$ の標本分布は，

$$\hat{\beta}_0 \sim N(\left(\beta_0, \sigma_\varepsilon^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{n s_x^2}\right]\right)$$

となりました（12.10 式）。σ_ε^2 は先ほど計算した（1.5）ため，これを用いて β_0 の標本分布の標準偏差を計算すると，

$$\sigma_\varepsilon^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{n s_x^2}\right] = 1.5 \left[\frac{1}{10} + \frac{4^2}{10 \cdot 3.2}\right] = 1.5 \cdot 0.6 = 0.9$$

となります。β_1 と同様に，信頼区間は自由度 $n-2=8$ の t 分布に基づいて計算されるため，95% 信頼区間は

$$[0.8 - 2.306 \cdot 0.9, 0.8 + 2.306 \cdot 0.9] \approx [0.8 - 2.075, 0.8 + 2.075] = [-2.075, 2.875]$$

となります。

(b) β_1 の仮説検定

(a) で求めた $\hat{\sigma}_\varepsilon^2$ を用いて，$\hat{\beta}_1$ を標準化した検定統計量 t は，

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}_\varepsilon^2}{n s_x^2}}} = \frac{0.4}{\sqrt{\frac{1.5}{10 \cdot 3.2}}} \approx \frac{0.4}{0.217} \approx 1.843$$

となります。そして，棄却域はやはり自由度 $n-2=8$ の t 分布に基づいて $|t| \geq 2.306$ となるため，今回の検定統計量 t は，棄却域には含まれません。この結果は，(a) で求めた β_1 の 95% 信頼区間が 0 を含んでいることからも明らかです。

【解答】

(a)β_0 の 95% 信頼区間: [-2.075, 2.875]， β_1 の 95% 信頼区間: [-0.1, 0.9]
(b)検定統計量 $t \approx 1.843$ であり，自由度 8，有意水準 5% の棄却限界値 2.306 を超えないため，帰無仮説は棄却されない。したがって，広報費が販売個数に影響を与えるとは言えない。

(3) ある研究者が，……

(a)

目的変数である生産量 Y を対数変換した理由としては，主に以下の 2 点が考えられます。

1. **分布の形状の補正**：生産量のような変数は「右に裾の長い分布」になりがちですが，対数変換を施すことで正規分布に近い左右対称の分布に近づけることができます。これは，回帰分析の誤差項が正規分布に従うという仮定を満たす上で役立ちます。12.2 節の「正規性の仮定」のところでも説明したように，回帰分析の統計的推測のためには，誤差項が正規分布に従っていれば問題はないのですが，これは目的変数の元々の分布が正規分布に近い場合には，より成立しやすくなると考えられます。
2. **モデルの解釈のしやすさ**：回帰分析において説明変数と目的変数がいずれも対数に変換されていると，その係数 β_1 は「**弾力性**」として解釈できます。これは，「ある変数の**変化率**が別の変数の**変化率**にどの程度影響するか」を示すものです。すなわち，「機械の使用時間 (X_1) が 1% 増加すると，生産量 (Y) が β_1% 変化する」という意味になり，ビジネス上有用な示唆を与えることが多くあります。

(b)

この分析に「作業者の年齢」を新たな説明変数として加えることには，**多重共線性（multicolinearity）**という統計的な問題を引き起こす可能性が高いです。モデルには既に「作業者の経験年数 (X_2)」が含まれており，一般的に「年齢」と「経験年数」の間には非常に強い正の相関関係が存在します。このように相関が極めて高い説明変数を同時にモデルに投入すると，モデルはどちらの変数が生産量に影響を与えているのかをうまく区別できなくなり，回帰係数の推定値が非常に不安定になります（図 4.13）。その結果，各変数が生産性に与える個々の影響を正しく評価することが困難になります。多くの場合，これは標準誤差が大きくなってしまう，という形で表れるため，信頼区間がめちゃめちゃに広くなったり，仮説検定の検定力が著しく低下する，などの問題を引き起こします。

【解答】

(a)(1) 右に歪んだ生産量の分布を，正規分布に近い形に補正することで，仮定を満たしやすくため。(2) 係数 β_1 を「X_1 が 1% 変化したときの Y の変化率 (%)」という弾力性として解釈しやすくするため。

(b)**多重共線性**の問題。既にある説明変数「経験年数」と新たに追加する「年齢」の間に非常に強い相関があるため，それぞれの変数が生産性に与える影響を正しく分離して推定することが困難になる。

(4) ある金融機関が，……

【解説】

(a)

モデル A（機械学習）

- **長所**：**極めて高い予測性能**を持つため，説明変数間の複雑な関係性を捉えて，かなりの高精度で貸し倒れリスクを判定可能です。
- **短所**：一番の問題は，**解釈可能性が低い（ブラックボックス性を持つ）**点です。例えばある顧客に対して「貸し倒れリスクが高い」と判定されたことを理由に融資を断る場合に，何を根拠にそのような予測が行われたかが説明できなければ，顧客は納得できないかもしれません。機械学習モデルは有用ですが，最終的な意思決定は人間の手に委ねられるべきなのです。

モデル B（重回帰）

- **長所**：回帰分析モデルは**高い解釈可能性**を持つことがポイントです。回帰係数は，非常に明快なやり方で，どの変数がどの程度影響しているかを教えてくれます。また，モデルがシンプルであるために，複雑な機械学習モデルと比べると，サンプルサイズが小さくても比較的安定した結果を得ることができるのも長所と言えます。
- **短所**：**予測性能が比較的低い**です。単純な線形関係しか仮定できないため，U 字の相関のようなものですら正確に反映することができません[9]。そして，機械学習モデルと比べると予測精度が低いため，貸し倒れリスクを誤って判断してしまうことは直接的なデメリットとなります。

(b)

どちらのモデルを使うべきかは，そのモデルを利用する「**目的**」によって決まります。

- **モデル A を使うべき場面**：目的が純粋な「**予測**」であり，その根拠の説明が不要な場合。例えば，大量の申請を自動的にスクリーニングするシステムなど。
- **モデル B を使うべき場面**：目的が「**解釈**」や「**説明**」，「**要因の理解**」である場合。例えば，顧客に融資を断る理由を説明する必要がある場合や，モデルの公平性・透明性が求められる

[9] あらかじめ二乗の項を入れておく，などである程度対策は可能です

場合。

金融機関の融資判断のように，結果の説明責任が伴う場面では，予測精度が多少劣っても解釈可能性の高いモデル B が好まれることが多いと考えられますが，ケースバイケースでしょう。

【解答】

(a)**モデル A** 長所: 高い予測性能，短所: 低い解釈性。 **モデル B** 長所: 高い解釈性，短所: 低い予測性能。

(b)**予測精度**が最優先で理由の説明が不要な場面（例：自動スクリーニング）ではモデル A を，**予測の根拠説明**や要因の理解，公平性の担保が重要な場面（例：顧客への説明）ではモデル B を使うべき。

(5)「統計モデルはすべて…

(a)

いかなる統計モデルも「間違っている」と言えるのは，**モデルが本質的に「現実の単純化」に過ぎないから**です。現実の世界は無数の要因が複雑に絡み合って生じていますが，統計モデルはその複雑な現実を，人間が理解できる範囲まで単純化・抽象化したものです。例えば，変数間の関係を直線で近似したり，正規分布という理想的な分布を仮定したりします。しかし，現実はモデルが仮定するような単純な数式や分布に完璧に従っているわけではありません。その意味で，いかなるモデルも現実そのものを**寸分違わず再現しているわけではない**ため，厳密に言えばすべて「間違っている」のです。

(b)

モデルが「間違っている（現実の完全なコピーではない）」としても，それが「**分析の目的に対して，現実の重要な側面を十分に捉え，有益な示唆を与えてくれる**」ならば，そのモデルは「有用である」と言えます。モデルの価値は，その絶対的な正しさにあるのではなく，**特定の目的を達成するための道具としての有効性**にあります[10]。統計モデルも，シンプルな構造（例えば説明変数が数個の重回帰分析）でそれなりの当てはまりや予測性能を叩き出すのであれば，「大体の予測をする」という目的に対しては十分な働きをしてくれるかもしれません。その意味で「有用」と言えるのです。ジョージ・ボックスのこの言葉は，「モデルの限界を認識しつつ，**自分の目的にとってそのモデルが十分に役立つかどうか**」という実用的な視点を持つことの重要性を教えてくれるものなのです。

[10] これは，例えばプラモデルなどでも同じことが言えます。プラモデルに対して「本物と同じように動かないからダメだ」などという人はいません。これは，プラモデルの目的が「動くこと」よりも「本物と同じ見た目であること」にあるため，と言えるでしょう。

【解答】

(a)**モデルは複雑な現実を単純化・抽象化したものであり，現実そのものを完璧に再現しているわけではないから。**

(b)**モデルが完璧でなくても，分析の目的（予測や要因理解など）を達成する上で，現実の重要な側面を十分に捉え，有益な示唆を与えてくれる道具としては十分に役立つという意味。**