中学レベルからはじめる

統計検定®3級対策 データ分析の基本

解答例

目次

第 1 章　データの種類 ･････ 4
1.1 質的変数と量的変数 ･････ 4
1.2 母集団と標本 ･････ 8
第 2 章　データの集め方とデータの中心 ･････ 11
2.1 実験研究と観察研究 ･････ 11
2.2 代表値の求め方 ･････ 13
第 3 章　データのばらつき ･････ 17
3.1 分散 ･････ 17
3.2 標準偏差 ･････ 20
第 4 章　データの変換とデータの標準化 ･････ 24
4.1 データの変換 ･････ 24
4.2 データの標準化 ･････ 27
第 5 章　相関 ･････ 34
5.1 共分散 ･････ 34
5.2 相関係数 ･････ 37
第 6 章　回帰分析 ･････ 41
6.1 回帰分析の基本 ･････ 41
6.2 回帰直線の推定 ･････ 42
第 7 章　質的データの表現 ･････ 48
7.1 質的変数を表すグラフ ･････ 48
7.2 質的変数の関係を表す表とグラフ ･････ 49
第 8 章　データの分布の形 ･････ 54
8.1 度数分布表とヒストグラム ･････ 54
8.2 四分位数と箱ひげ図 ･････ 59
第 9 章　確率の基礎 ･････ 65
9.1 確率の意味 ･････ 65
9.2 条件付き確率 ･････ 72
第 10 章　くり返し操作と確率 ･････ 82
10.1 くり返し操作における確率 ･････ 82
10.2 成功回数についての確率 ･････ 91
第 11 章　確率変数と確率分布 ･････ 101
11.1 確率変数の考え方 ･････ 101
11.2 確率変数の性質 ･････ 106
第 12 章　二項分布から正規分布へ ･････ 111
12.1 二項分布の確率計算 ･････ 111
12.2 正規分布の性質 ･････ 115
第 13 章　正規分布 ･････ 121
13.1 正規分布の確率計算 ･････ 121

13.2 標本平均・標本比率の性質 ・・ 132
第 14 章 推定と検定 ・・ 138
14.1 区間推定 ・・・ 138
14.2 仮説検定 ・・・ 142

第 1 章　データの種類

1.1　質的変数と量的変数

問題 1.1

(1) **身長： 量的変数**
　数値で表され，その数値のちがいが身長のちがいという量的な意味をもちます.

(2) **出身地： 質的変数**
　どこの地域かという種類のちがいだけを表し，もし数値化したとしても，その差は量として意味をもちません.

(3) **気温（℃）： 量的変数**
　数値のちがいが気温のちがいという量的な意味をもちます.

(4) **満足度（高い・ふつう・低い）： 質的変数**
　順序はありますが，もし数値化したとしても，その差は量として意味をもたないため質的変数に含めます.

(5) **曜日： 質的変数**
　月曜・火曜などの種類の区別を表すだけであり，もし数値化したとしても，その差は量として意味をもちません.

(6) **車の速度： 量的変数**
　数値で表され，その数値のちがいが速度のちがいという量的な意味をもちます.

(7) **職業： 質的変数**
　種類のちがいだけを表し，もし数値化したとしても，その差は量として意味をもちません.

(8) **兄弟姉妹の人数： 量的変数**
　人数を数で表し，その数値のちがいが人数のちがいという量的な意味をもちます.

(9) **車のボディタイプ（セダン・SUV・軽など）： 質的変数**
　車の種類のちがいを表すだけであり，もし数値化したとしても，その差は量として意味をもちません.

(10) **歩数： 量的変数**
　歩数を数で表し，その数値のちがいが歩いた量のちがいを表します.

(11) **成績評価（A・B・C・D）： 質的変数**
　順序はありますが，もし数値化したとしても，その差は量として意味をもたないため質的変数に分類します.

(12) **国籍： 質的変数**
　どの国籍かという種類のちがいだけを表し，もし数値化したとしても，その差は量として意味をもちません.

(13) **テストの得点： 量的変数**
　点数のちがいが得点のちがいという量的な意味をもちます.

(14) **天気（晴れ・くもり・雨）： 質的変数**

天気の種類のちがいを表すだけであり，もし数値化したとしても，その差は量として意味をもちません.

問題 1.2

(1) **身長： 比例尺度**

0 が絶対的なゼロを表し，差にも比にも量としての意味があります.

(2) **出身地： 名義尺度**

種類のちがいだけを表す変数であり，順序はなく，もし数値化したとしても，その差や大小には量としての意味がありません.

(3) **気温（℃）： 間隔尺度**

数値の差には量としての意味がありますが，0 ℃は絶対的なゼロではないため，もし比をとったとしても，その比には量としての意味がありません.

(4) **満足度（高い・ふつう・低い）： 順序尺度**

「高い」「ふつう」「低い」という評価の順序を表しますが，もし数値化したとしても，その差や比には量としての意味がありません.

(5) **曜日： 名義尺度**

種類の区別だけを表す変数であり，順序はなく，もし数値化したとしても，その差や大小には量としての意味がありません.

(6) **車の速度： 比例尺度**

0 が絶対的なゼロを表し，「2 倍の速さ」など比の比較が量として意味をもちます.

(7) **職業： 名義尺度**

種類のちがいだけを表す変数であり，順序はなく，もし数値化したとしても，その差や大小には量としての意味がありません.

(8) **兄弟姉妹の人数： 比例尺度**

0 が絶対的なゼロを表し，差にも比にも量としての意味があります.

(9) **車のボディタイプ（セダン・SUV・軽など）： 名義尺度**

種類のちがいを表すだけであり，順序はなく，もし数値化したとしても，その差や大小には量としての意味がありません.

(10) **歩数： 比例尺度**

0 歩は「まったく歩いていない」ことを表す絶対的なゼロであり，差にも比にも量としての意味があります.

(11) **成績評価（A・B・C・D）： 順序尺度**

「A」「B」「C」「D」という評価の順序を表しますが，もし数値化したとしても，その差や比には量としての意味がありません.

(12) **国籍： 名義尺度**

種類のちがいだけを表す変数であり，順序はなく，もし数値化したとしても，その差や大小には量としての意味がありません.

(13) **テストの得点： 間隔尺度**

点数の差には量としての意味がありますが，0 点は絶対的なゼロではないため，もし比をとったとしても，その比には量としての意味がありません.

(14) **天気（晴れ・くもり・雨）： 名義尺度**

種類のちがいを表すだけであり，順序はなく，もし数値化したとしても，その差や大小には量として

の意味がありません.

問題 1.3

(1) **身長： 連続変数**
測定によって得られる量であり，連続的な値として扱います.

(2) **出身地： 質的変数**
種類のちがいを表すだけであり，数量として扱いません.

(3) **気温（℃）： 連続変数**
測定値であり，連続的な値として扱います.

(4) **満足度（高い・ふつう・低い）： 質的変数**
順序はありますが，数量として扱いません.

(5) **曜日： 質的変数**
種類の区別だけを表し，数量として扱いません.

(6) **車の速度： 連続変数**
測定によって得られる値であり，連続的な値として扱います.

(7) **職業： 質的変数**
種類のちがいだけを表し，数量として扱いません.

(8) **兄弟姉妹の人数： 離散変数**
0 人，1 人，2 人のように，数えられる値です.

(9) **車のボディタイプ（セダン・SUV・軽など）： 質的変数**
種類のちがいを表すだけで，数量として扱いません.

(10) **歩数： 離散変数**
1 歩，2 歩，3 歩のように，数えられる量です.

(11) **成績評価（A・B・C・D）： 質的変数**
順序はありますが，数量として扱いません.

(12) **国籍： 質的変数**
種類のちがいだけを表し，数量として扱いません.

(13) **テストの得点： 離散変数**
0 点，1 点，2 点のように，通常は数えられる値として扱います.

(14) **天気（晴れ・くもり・雨）： 質的変数**
種類のちがいを表すだけで，数量として扱いません.

1.1 の確認テスト

問 1

各変数が質的変数か量的変数かを判断するときは，「値が種類を表すか」「数値のちがいが量のちがいを表すか」などを手がかりに考えます．とくに，数値で表されていても量として扱えない場合は質的変数になります.

(1) **店舗の所在地： 質的変数**
地名という種類のちがいを表す情報であり，もし数値化したとしても，その差は量として意味をもちません.

(2) **商品の重さ： 量的変数**

重さを測定した数値であり，数値のちがいが重量のちがいという量的な意味をもつため，量的変数に分類されます.

(3) **従業員の役職名： 質的変数**

役職という種類のちがいを表す情報であり，もし数値化したとしても，その差は量として意味をもちません.

(4) **在庫数： 量的変数**

個数として数えられる量であり，数値のちがいが在庫量のちがいという量的な意味をもつため，量的変数に分類されます.

問 2

各変数がどの尺度に属するかを判断するときは,「順序があるか」「差が意味をもつか」「比が意味をもつか」などを手がかりに考えます.

(1) **血液型（A・B・O・AB）： 名義尺度**

種類のちがいを表す情報であり，順序はなく，もし数値化したとしても，その差や大小には量としての意味がありません.

(2) **入学年度： 間隔尺度**

年度どうしの差（1 年の差など）には量としての意味がありますが，0 年が絶対的なゼロを表すわけではないため，もし比をとったとしても，その比には量としての意味がありません.

(3) **商品の価格： 比例尺度**

0 円は支払う金額がないことを表す絶対的なゼロであり，差にも比にも量としての意味があります.

(4) **配達希望時間帯： 順序尺度**

「8–12 時」「14–16 時」などのように順序を表しますが，その差がどれだけ時間の差を意味するかといった量的な解釈はできないため，順序はあるものの，量として扱うことはできません.

問 3

各変数がどの分類に属するかを判断するときは,「種類の区別か」「数えられる量か」「連続的な値として扱われる量か」を手がかりに考えます.

(1) **滞在日数： 離散変数**

日数として数えられる量です.

比例尺度

0 日は「滞在がない」ことを示す絶対的なゼロであり，差や比に意味があります.

(2) **利用端末（スマートフォン・タブレット・PC）： 質的変数**

種類のちがいを表します.

名義尺度

種類の区別だけを表します.

(3) **移動距離（km）： 連続変数**

測定値であり，連続的な値として扱います.

比例尺度

0 km は「移動していない」ことを示す絶対的なゼロであり，差や比に意味があります.

(4) **会員ランク（プラチナ・ゴールド・シルバー）： 質的変数**

ランクの種類のちがいを表し，数値としての扱いはしません.

順序尺度

上下の順序はありますが，その差や比を量として扱うことはできません．

(5) **体温（℃）： 連続変数**

測定値であり，連続的な値として扱います．

間隔尺度

数値の差には意味がありますが，0 ℃は絶対的なゼロではなく，比として扱うことはできません．

1.2 母集団と標本

問題 1.4

(1) **母集団： 市内の中学生全体， 標本： アンケートを行った 300 人**

市内の中学生全体が母集団であり，その一部である 300 人を標本としています．

(2) **母集団： この企業の従業員全体， 標本： 調査した 200 人**

従業員全体が母集団であり，そのなかから 200 人が標本として選ばれています．

(3) **母集団： 地域の農家全体， 標本： 訪問して調査した 100 戸**

地域の農家全体が母集団であり，その一部である 100 戸を標本としています．

問題 1.5

(1) **全数調査**

市が管理する公園すべてで利用者数を記録しており，対象全体を調べています．

(2) **標本調査**

投票を終えた有権者のなかから一部の人だけに協力してもらっており，対象の一部を調べています．

(3) **標本調査**

棚に並んでいる牛乳のなかから 40 本だけを選んで検査しており，対象の一部を調べています．

(4) **全数調査**

全社員にアンケートを実施しており，調査対象をもれなく調べています．

(5) **標本調査**

設置された 300 台のなかから 20 台だけを選んで点検しており，対象の一部を調べています．

問題 1.6

(1) **無作為抽出ではない**

午前中に市役所を訪れた人だけを対象にしており，時間帯による偏りが生じる可能性があるためです．

(2) **無作為抽出**

調査対象者の一覧に番号を付け，乱数を使って選んでおり，人の判断による偏りが入りにくい方法であるためです．

(3) **無作為抽出ではない**

公園で運動していた人だけを対象にしており，特定の行動をしている人に偏る可能性があるためです．

(4) **無作為抽出ではない**

協力すると申し出た人だけを対象にしており，自発的に参加する人に偏る可能性があるためです．

(5) **無作為抽出ではない**

立候補した人のなかから選んでおり，立候補するかどうかの段階で偏りが生じる可能性があるためです．

1.2 の確認テスト

問 1

母集団は「知りたい対象全体」，標本は「実際に調べた一部」です．

(1) **母集団： 市内の高校生全体，　　標本： 選ばれた 500 人**

全体の傾向を知るために，一部を取り出して調査しています．

(2) **母集団： 店舗の利用者全体，　　標本： 最近来店した 150 人**

「最近来店した人」は利用者全体の一部に当たります．

(3) **母集団： 工場で生産された製品全体，　　標本： 抜き取った 60 個**

品質確認のために，全体から一部を取り出して検査しています．

(4) **母集団： 全国の会員全体，　　標本： 地域ごとに比例して選ばれた 1200 人**

地域ごとに人数を調整して選んでいますが，全体の一部である点は変わりません．

問 2

全数調査は「対象全体を調べる」，標本調査は「一部を調べる」調査です．

(1) **全数調査**

参加者全員のデータを収集しています．

(2) **標本調査**

店舗ごとに 3 人だけを選んでおり，全体ではありません．

(3) **標本調査**

月ごとに一部（50 個）を抜き取って検査しています．

(4) **標本調査**

「過去 3 か月以内に 1 回以上購入した会員」は会員全体の一部に当たります．

（補足）購入経験のある会員だけを対象としているため，会員全体の購買傾向を推測する標本としては偏りがある可能性があります．

問 3

無作為抽出とは，乱数やくじ引きなどの偶然性を利用して，人の判断による偏りが入りにくいように標本を選ぶ方法です．

(1) **無作為抽出である**

番号札をよく混ぜて引いており，偶然性を利用して選んでいるため，人の判断による偏りが入りにくい方法です．

(2) **無作為抽出ではない**

午後に通りかかった人だけを対象にしており，時間帯による偏りが生じる可能性があるためです．

(3) **無作為抽出ではない**

早朝に来店していた人だけを対象にしており，時間帯による偏りが生じる可能性があるためです．

(4) **無作為抽出である**

会員番号を書いた番号札をよく混ぜて引いており，偶然性を利用して選んでいるため，人の判断による偏りが入りにくい方法です．

問 4

無作為抽出が重要なのは，標本の偏りを小さくし，母集団をより正しく推測するためです．

正解：(2)

(1) 作業のしやすさは利点の一つですが，重要とされる理由ではありません．

(2) **標本の偏りを小さくし，母集団の特徴をより正しく推測できるようにするため**

無作為抽出が重要とされる理由として最も適切です．

(3) 説明しやすさは利点の一つですが，本質的な理由ではありません．

(4) 母集団の大きさとは関係ありません．

(5) 無作為抽出でも結果が「必ず」正確になるわけではありません．

(6) 機械化しやすいかどうかは，本質的な理由ではありません．

第 2 章　データの集め方とデータの中心

2.1　実験研究と観察研究

問題 2.1

(1) **実験研究**

生徒をランダムに 2 つのグループに分け，学習アプリを使うかどうかという条件を調査者が操作して比較しています.

(2) **観察研究**

朝食の有無と集中度をアンケートでたずねており，条件を操作せず，もともとのちがいを観察しています.

(3) **実験研究**

教室ごとに空調設定をランダムに変えており，空調という条件を調査者が操作して比較しています.

(4) **観察研究**

降水量と図書館の利用者数について，過去の記録を用いて関係を調べており，条件を操作せずに観察しています.

問題 2.2

(1) **不適切**

朝型・夜型という別の要因が異なるため，睡眠改善方法のちがいによる影響と朝型・夜型のちがいによる影響を区別しにくくなり，適切な比較になっていません.

(2) **不適切**

水の温度という別の要因が異なるため，洗剤のちがいによる影響と水の温度のちがいによる影響を区別しにくくなり，適切な比較になっていません.

(3) **適切**

家庭をランダムに 2 つのグループに割り当てることで，条件以外の要因がグループ間で偏りにくくなるため，空気清浄機の有無による影響を比較しやすくなり，適切な比較になっています.

(4) **適切**

どちらのグループに入るかをくじで決めることで，条件以外の要因がグループ間で偏りにくくなり，成分を含まないクリームを用いることで，クリームを塗るという行為などの影響を考慮できるため，日焼け止めの効果を比較しやすくなり，適切な比較になっています.

問題 2.3

(1) **適切**

事前テストの点数で層に分け，各層内で無作為に割り当てているため，学習プログラムの影響を調べやすくなるからです.

(2) **適切**

年代で分けたうえで，各年代のなかで無作為に割り当てているため，学習プログラムの影響を調べやすくなるからです．

(3) **不適切**

英語を日常的に使うかどうかで割り当てが決まっており，「学習プログラムのちがい」の影響と「英語使用頻度のちがい」の影響が区別できなくなるからです．

2.1 の確認テスト

問 1

実験研究は「調査者が条件を操作する研究」，観察研究は「自然のままの状況を調べる研究」です．

(1) **実験研究**

児童を無作為に 2 つのグループに分け，新旧メニューを割り当てており，調査者が条件（メニューの種類）を操作しています．

(2) **実験研究**

患者を無作為に 2 群に分け，新旧の治療薬を割り当てており，調査者が条件（治療法）を操作しています．

(3) **観察研究**

条件を操作せず，自然に存在する通学時間と成績の関係を調べています．

(4) **観察研究**

条件を操作せず，自然に存在する歩数と血圧の関係を，健康アプリの記録を用いて調べています．

問 2

比較の仕方が適切かどうかは，「比較したい条件以外をできるだけ同じにできているか」で判断します．

(1) **不適切**

犬の体格（大きい／小さい）という別の要因が異なるため，えさ A とえさ B のちがいによる影響と体格のちがいによる影響を区別しにくくなるからです．

(2) **不適切**

国立大学と私立大学では学力や学習環境が異なる可能性があるため，教科書 A と教科書 B のちがいによる影響と，学力や学習環境のちがいによる影響を区別しにくくなるからです．

(3) **不適切**

日本国籍の学生と留学生では言語能力や学習背景が異なるため，授業形式（対面／録画）のちがいによる影響と，言語能力や学習背景のちがいによる影響を区別しにくくなるからです．

(4) **適切**

参加者にカーナビ A と B をランダムに割り当て，同じルートで比較しているため，条件以外の要因がグループ間で偏りにくくなり，カーナビのちがいによる影響を比較しやすくなるからです．

(5) **不適切**

記録時刻（朝 9 時／夕方 18 時）という別の要因が異なるため，地点のちがいによる影響と記録時刻のちがいによる影響を区別しにくくなるからです．

(6) **適切**

苗の高さで層に分け，各層内で無作為に割り当てているため，苗の高さのちがいがグループ間で偏りにくくなり，肥料の有無による影響を比較しやすくなるからです．

2.2　代表値の求め方

問題 2.4

(1) 5 日間の合計時間は,

$$72 \times 5 = 360$$

より 360 分となります.

(2) 1〜4 日目の合計時間は,

$$(75 \times 2) + (70 \times 2) = 150 + 140 = 290$$

より 290 分となります.

(3) 5 日目の勉強時間は,

$$360 - 290 = 70$$

より 70 分となります.

問題 2.5

平均値を求めると,

$$\frac{18+1+14+3+14+1+19+2+18+1+19}{11} = \frac{110}{11} = 10$$

より 10 分となります.

データを小さい順に並べると,

$$1,\ 1,\ 1,\ 2,\ 3,\ 14,\ 14,\ 18,\ 18,\ 19,\ 19$$

となります.

データは 11 個あるので，中央にくるのは左から数えて 6 番目の値です．6 番目の値は 14 なので，中央値は 14 分となります.

問題 2.6

(1) 小さい順に並べると,

$$7,\ 7,\ 7,\ 7,\ 7,\ 12,\ 13,\ 14,\ 14,\ 80$$

最頻値は最も多く出てくる値なので，最頻値は 7 時間です.

データは 10 個あるので，中央値は左から 5 番目と 6 番目の値の平均になります．それぞれ 7 と 12 なので,

$$\frac{7+12}{2} = 9.5$$

より，中央値は 9.5 時間です.

平均値は,

$$\frac{14+7+12+7+80+7+14+7+7+13}{10} = \frac{168}{10} = 16.8$$

より，16.8 時間となります．

(2) (a) ○：最頻値は「最も多い値」で決まり，外れ値 80 の影響を受けにくいです．

(b) ○：中央値は「位置」で決まるため，外れ値 80 があってもほとんど変化しません．

(c) ○：平均値は外れ値 80 によって大きく押し上げられており，多くの人のゲーム時間をよく表しているとはいえません．

(d) ×：この 10 人の典型的なゲーム時間を示す指標としては，外れ値の影響を受けにくい最頻値や中央値のほうが適しており，平均値が適しているとはいえません．

問題 2.7

(1) **平均値**

合計を人数で割って「1 人あたり」を求める場面なので，平均値が適切です．

(2) **中央値**

通勤時間が非常に長い人が一部いるため，外れ値の影響を受けにくい中央値が適します．

(3) **最頻値**

最も多く売れたパンの種類を知りたい状況なので，最頻値を用います．

2.2 の確認テスト

問 1

平均値は

$$\frac{3+0+1+3+0+2+0+3+1+0+5+3+2+0+1+3+0+3+0+0}{20} = \frac{30}{20} = 1.5$$

となります．

データを小さい順に並べると，

$$0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 1,\ 1,\ 2,\ 2,\ 3,\ 3,\ 3,\ 3,\ 3,\ 5$$

となります．10 番目と 11 番目の平均値は，

$$\frac{1+1}{2} = 1$$

となり，これが中央値です．

最も多く出ている値は 0 となり，これが最頻値です．

問 2

(1) 9 日分の合計は，

$$18 \times 9 = 162$$

となります．与えられた 8 日分の合計は，

$$19 + 15 + 22 + 17 + 20 + 16 + 18 + 17 = 144$$

となるので，残り 1 日は，

$$162 - 144 = 18$$

より 18℃ となります.

(2) 小さい順に並べると,

$$15,\ 16,\ 17,\ 17,\ 18,\ 18,\ 19,\ 20,\ 22$$

となり，5 番目の 18℃ が中央値です.

問 3

(1) 3 日が 2 回，5 日が 3 回なので，最頻値を 3 日と 5 日の 2 つにするには,

$$x = 3$$

とする必要があります.

(2) $x = 3$ を含めて小さい順に並べると,

$$1,\ 2,\ 3,\ 3,\ 3,\ 4,\ 4,\ 5,\ 5,\ 5,\ 6,\ 7,\ 8,\ 10$$

となります．7 番目と 8 番目の平均値は,

$$\frac{4+5}{2} = 4.5$$

となり，これが中央値です.

問 4

データの合計は 150，平均値は 10 なので，データの個数は

$$150 \div 10 = 15$$

となります.

したがって，中央値に位置するデータは小さい順に並べたときの 8 番目の値です.

問 5

(1) **平均値，中央値**

平均値は，データの合計を個数で割って求める値なので，計算結果がもとのデータのなかにない値になることがあります．たとえば，データが $1, 2, 3, 10$ のとき，平均値は

$$\frac{1+2+3+10}{4} = 4$$

となり，この値はデータのなかにありません.

また，このデータの中央値は，真ん中の 2 つの平均をとるので,

$$\frac{2+3}{2} = 2.5$$

となり，この値もデータのなかにありません.

一方，最頻値は「最も多く現れる値」なので，必ずデータのなかにある値になります.

(2) **平均値，中央値，最頻値**

平均値は，データの合計を個数で割って求める値です．すべてのデータを 2 倍すると合計も 2 倍になるため，平均値も 2 倍になります.

中央値は，データを小さい順に並べたときの真ん中の値です．すべてのデータを 2 倍しても並び順は変わらず，真ん中に来る値がそのまま 2 倍されるだけなので，中央値も 2 倍になります．

最頻値は，最も多く現れる値です．すべてのデータを 2 倍すると，どの値が何回出るかは変わらず，最も多い値が 2 倍されるだけなので，最頻値も 2 倍になります．

(3) **最頻値**

平均値は，データの合計を個数で割って求める値なので，同じデータであればいつも一つの値に決まります．

中央値も，データを小さい順に並べたときの真ん中の値として一つに決まります（個数が偶数のときは，真ん中の 2 つの平均をとると決めれば，やはり一つに定まります）．

一方，最頻値は「最も多く現れる値」です．たとえば，データが $0, 0, 1, 1$ のように，最も多い回数で現れる値が複数ある場合，最頻値は一つに定まりません．このように，最頻値は必ずしも一つに定まるとは限りません．

問 6

(1) ×：平均値は外れ値の影響を受けやすく，中央値は外れ値の影響を受けにくい代表値です．

(2) ×：最大値より大きい値を 1 つ追加しても，中央値が必ず大きくなるとは限りません．

(3) ×：新しい値が 1 回しか出てこなければ，最頻値は変化しないことがあります．

(4) ×：平均値が最適とは限らず，状況によっては中央値や最頻値のほうが適切なことがあります．

(5) ×：最小値より小さい値を 1 つ追加しても，中央値が必ず小さくなるとは限りません．

(6) ×：小さい順に並べたときの真ん中の値は中央値であり，平均値とは別の代表値です．

(7) ○：最頻値は「最も多く出てくる値」であり，数値でないデータ（例：好きな車種など）でも求められます．

(8) ×：「最も売れた価格帯」を知りたい場面では，最頻値を用いるのが適切です．

(9) ×：大きな遅延などの外れ値がある場合，典型的な値を表すには中央値が適切です．

(10) ×：「1 人あたり」の量を求めるときは，合計を人数で割る平均値を用いるのが適切です．

第 3 章　データのばらつき

3.1　分散

問題 3.1

(1) 売上個数の合計は

$$24 + 30 + 33 + 36 + 42 = 165$$

なので，平均値は

$$\frac{165}{5} = 33$$

です．

(2) 各値の偏差は，順に

$$24 - 33 = -9, \quad 30 - 33 = -3, \quad 33 - 33 = 0, \quad 36 - 33 = 3, \quad 42 - 33 = 9$$

となります．

(3) 偏差の合計は

$$-9 + (-3) + 0 + 3 + 9 = 0$$

なので，偏差の平均値は

$$\frac{0}{5} = 0$$

です．

問題 3.2

(1) 各値の偏差の 2 乗は，順に

$$(-9)^2 = 81, \quad (-3)^2 = 9, \quad 0^2 = 0, \quad 3^2 = 9, \quad 9^2 = 81$$

となります．

(2) 偏差の 2 乗の合計は

$$81 + 9 + 0 + 9 + 81 = 180$$

です．

(3) 分散は，偏差の 2 乗の平均値なので，

$$\frac{180}{5} = 36$$

です．

問題 3.3

(1) 各値を 2 乗します.

$$24^2 = 576,\ 30^2 = 900,\ 33^2 = 1089,\ 36^2 = 1296,\ 42^2 = 1764$$

(2)「値の 2 乗」の合計は

$$576 + 900 + 1089 + 1296 + 1764 = 5625$$

となります. したがって,「値の 2 乗」の平均値は

$$\frac{5625}{5} = 1125$$

となります.

(3) 平均値は問題 3.1 より

$$\frac{24 + 30 + 33 + 36 + 42}{5} = 33$$

なので, 平均値の 2 乗は

$$33^2 = 1089$$

です.

(4) 分散の公式

$$\text{分散} = \text{「値の 2 乗」の平均値} - (\text{平均値})^2$$

を使って分散を求めると,

$$1125 - 1089 = 36$$

となります. 問題 3.2 で求めた分散（36）と一致します.

3.1 の確認テスト

問 1

まず, 平均値を求めます.

$$\frac{9 + 4 + 12 + 6 + 3 + 10 + 7 + 6 + 8 + 5}{10} = 7$$

次に, 各値から平均値を引いて偏差を求めます.

$$2,\ -3,\ 5,\ -1,\ -4,\ 3,\ 0,\ -1,\ 1,\ -2$$

「偏差の 2 乗」は

$$4,\ 9,\ 25,\ 1,\ 16,\ 9,\ 0,\ 1,\ 1,\ 4$$

となります.

「偏差の 2 乗」の平均値を求めると,

$$\frac{4 + 9 + 25 + 1 + 16 + 9 + 0 + 1 + 1 + 4}{10} = \frac{70}{10} = 7$$

となります．したがって，このデータの分散は 7 です．

問 2

まず，平均値を求めます．データの合計は 3000，個数は 100 なので，

$$\frac{3000}{100} = 30$$

となります．

次に，「値の 2 乗」の平均値を求めます．「値の 2 乗」の合計は 95000 なので，

$$\frac{95000}{100} = 950$$

となります．

分散は

$$\text{「値の 2 乗」の平均値} - (\text{平均値})^2$$

で求められるので，

$$950 - 30^2 = 950 - 900 = 50$$

となります．したがって，このデータの分散は 50 です．

問 3

分散の公式より得られる

$$\text{「値の 2 乗」の平均値} = \text{分散} + (\text{平均値})^2$$

を用います．

A グループの「値の 2 乗」の平均値は

$$1 + 5^2 = 1 + 25 = 26$$

なので，「値の 2 乗」の合計は

$$8 \times 26 = 208$$

となります．

B グループの「値の 2 乗」の平均値は

$$1 + 10^2 = 1 + 100 = 101$$

なので，値の 2 乗の合計は

$$12 \times 101 = 1212$$

となります．

したがって，全体 20 個の「値の 2 乗」の平均値は

$$\frac{208 + 1212}{20} = \frac{1420}{20} = 71$$

となります．

次に，全体の平均値を求めます．A グループの合計は $8 \times 5 = 40$，B グループの合計は $12 \times 10 = 120$ なので，全体の平均値は

$$\frac{40 + 120}{20} = \frac{160}{20} = 8$$

となります．

したがって，全体の分散は

$$71 - 8^2 = 71 - 64 = 7$$

となります．

3.2 標準偏差

問題 3.4

問題 3.3 で求めた分散は 36 でした．標準偏差は，この分散のルートをとって求めます．

$$\sqrt{36} = 6$$

したがって，標準偏差は 6 です．

問題 3.5

問題 3.4 で求めた標準偏差は 6，問題 3.1 で求めた平均値は 33 でした．変動係数は，標準偏差を平均値で割って求めます．

$$\frac{6}{33} \approx 0.18$$

したがって，変動係数はおよそ 0.18 です．

3.2 の確認テスト

問 1

(1) 平均値が -1 であり，偏差が

$$2,\ -1,\ 3,\ 0,\ -3,\ 1,\ -2$$

なので，もとのデータは

$$-1+2=1,\quad -1-1=-2,\quad -1+3=2,\quad -1+0=-1,\quad -1-3=-4,\quad -1+1=0,\quad -1-2=-3$$

より，

$$1,\ -2,\ 2,\ -1,\ -4,\ 0,\ -3$$

です．

(2)「偏差の 2 乗」は

$$4,\ 1,\ 9,\ 0,\ 9,\ 1,\ 4$$

であり，その合計は

$$4+1+9+0+9+1+4=28$$

となります．したがって分散は

$$\frac{28}{7} = 4$$

となり，標準偏差は

$$\sqrt{4} = 2$$

となります．

問 2

(1) A の合計は

$$6 + 27 + 30 + 21 + 18 + 24 + 27 + 15 + 3 + 9 = 180$$

なので，A の平均値は

$$\frac{180}{10} = 18$$

です．

B の合計は

$$56 + 48 + 52 + 28 + 44 + 32 + 60 + 36 + 64 + 60 = 480$$

なので，B の平均値は

$$\frac{480}{10} = 48$$

です．

(2) A の各データの偏差とその 2 乗は次のようになります．

データ	6	27	30	21	18	24	27	15	3	9
偏差	-12	9	12	3	0	6	9	-3	-15	-9
偏差の 2 乗	144	81	144	9	0	36	81	9	225	81

「偏差の 2 乗」の合計は

$$144 + 81 + 144 + 9 + 0 + 36 + 81 + 9 + 225 + 81 = 810$$

なので，A の分散は

$$\frac{810}{10} = 81$$

標準偏差は

$$\sqrt{81} = 9$$

です．

B の各データの偏差とその 2 乗は次のようになります．

データ	56	48	52	28	44	32	60	36	64	60
偏差	8	0	4	-20	-4	-16	12	-12	16	12
偏差の 2 乗	64	0	16	400	16	256	144	144	256	144

「偏差の 2 乗」の合計は

$$64+0+16+400+16+256+144+144+256+144=1440$$

なので，B の分散は

$$\frac{1440}{10}=144$$

標準偏差は

$$\sqrt{144}=12$$

です．

(3) 変動係数は，標準偏差を平均値で割って求めます．

A の変動係数は

$$\frac{9}{18}=0.5$$

B の変動係数は

$$\frac{12}{48}=0.25$$

です．

よって，変動係数が大きい A のほうが，相対的にばらつきが大きいといえます．

問 3

(1) もとの 6 つのデータの平均値は 50 で，新しく追加された値も 50 なので，7 つのデータの合計は

$$6\times 50+50=350$$

となります．したがって，新しい 7 つのデータの平均値は

$$\frac{350}{7}=50$$

です．

(2) もとの 6 つのデータについて，分散は

6 個の「偏差の 2 乗」を全部足して，6 で割ったもの

です．つまり，

$$\text{分散}=\frac{\text{「偏差の 2 乗」の合計}}{6}$$

という関係になっています．これをいいかえると，

$$\text{「偏差の 2 乗」の合計}=\text{分散}\times 6$$

となります．

ここで，もとの 6 つのデータの標準偏差は $\sqrt{\frac{21}{2}}$ なので，分散は

$$\frac{21}{2}$$

です．したがって，「偏差の 2 乗」の合計は

$$\frac{21}{2} \times 6 = 63$$

です．

そして，新しく追加した 50 の偏差は

$$50 - 50 = 0$$

なので，新しい 7 つのデータの「偏差の 2 乗」の合計は変わらず 63 のままです．したがって，新しい 7 つのデータの分散は

$$\frac{63}{7} = 9$$

となります．よって，標準偏差は

$$\sqrt{9} = 3$$

となります．

(3) 変動係数は，標準偏差を平均値で割ることで求められます．新しい 7 つのデータの平均値は 50，標準偏差は 3 なので，変動係数は，

$$\frac{3}{50} = 0.06$$

となります．

問 4

(1) ○：偏差は「データ − 平均値」であり，すべての偏差を足すと必ず 0 になります．

(2) ×：0 が含まれていても，平均値と異なる値があればその偏差は 0 にならず，標準偏差も 0 にはなりません．

(3) ○：すべて同じ値であれば平均値もその値になり，偏差はすべて 0 となるため，分散も標準偏差も 0 になります．

(4) ○：平均値と同じ値を 1 つ追加する場合は，平均値は変わらず，その値の偏差は 0 になるため，偏差の 2 乗の合計は変わりません．

(5) ○：標準偏差は分散の（負でない）平方根であり，標準偏差を 2 乗すると分散になります．

(6) ×：平均値が大きくても，すべての値が平均値と等しければ標準偏差は 0 になるため，「平均値が大きいデータほど標準偏差も大きくなる」とは限りません．

第 4 章　データの変換とデータの標準化

4.1　データの変換

問題 4.1

問題 3.1 で求めた平均値は 33 でした.

(1) すべての値に 8 を加えると，平均値も 8 増えるので,

$$33 + 8 = 41$$

となります.

(2) すべての値を 3 倍すると，平均値も 3 倍になるので,

$$33 \times 3 = 99$$

となります.

問題 4.2

問題 3.3 で求めた分散は 36 でした.

(1) すべての値に 8 を加えても，分散は変わらないので，このときの分散は 36 となります.

(2) すべての値を 3 倍すると，分散は 3^2 倍，つまり，9 倍になるので，このときの分散は

$$36 \times 9 = 324$$

となります.

問題 4.3

問題 3.4 で求めた標準偏差は 6 でした.

(1) すべての値に 8 を加えても，標準偏差は変わらないので，このときの標準偏差は 6 となります.

(2) すべての値を 3 倍すると，標準偏差は 3 倍になるので，このときの標準偏差は

$$6 \times 3 = 18$$

となります.

4.1 の確認テスト

問 1

(1) このデータの平均値は

$$\frac{9+6+12+10+7+14+8+11+10+13}{10} = \frac{100}{10} = 10$$

です.

偏差は

$$-1,\ -4,\ 2,\ 0,\ -3,\ 4,\ -2,\ 1,\ 0,\ 3$$

となり,「偏差の 2 乗」の合計は

$$1+16+4+0+9+16+4+1+0+9=60$$

です．したがって，分散は

$$\frac{60}{10}=6$$

となります.

(2) すべての値に 11 を加えると，平均値も 11 増えるので，このときの平均値は

$$10+11=21$$

となります.

すべての値に 11 を加えても，分散は変わらないので，このときの分散は 6 のままです.

(3) すべての値を 7 倍すると，平均値も 7 倍になるので，このときの平均値は

$$10\times 7=70$$

となります.

すべての値を 7 倍すると，分散は 7^2 倍になるので，このときの分散は

$$6\times 7^2=6\times 49=294$$

となります.

(4) すべての値について「11 を加えたあとに 7 倍する」と，平均値は

$$(10+11)\times 7=147$$

となります.

すべての値について 11 を加えても分散は変わらず，そのあと 7 倍すると分散は 7^2 倍になります.
よって，このときの分散は

$$6\times 7^2=6\times 49=294$$

となります.

(5) すべての値について「7 倍したあとに 11 を加える」と，平均値は

$$(10\times 7)+11=81$$

となります.

すべての値を 7 倍すると分散は 7^2 倍になり，そのあと 11 を加えても分散は変わりません．よって，
このときの分散は

$$6\times 7^2=6\times 49=294$$

となります.

問 2

(1) このデータの平均値は

$$\frac{11+17+9+13+10}{5}=\frac{60}{5}=12$$

です．偏差は

$$-1,\ 5,\ -3,\ 1,\ -2$$

となり，「偏差の 2 乗」の合計は

$$1+25+9+1+4=40$$

です．したがって，分散は

$$\frac{40}{5}=8$$

となり，標準偏差は

$$\sqrt{8}=\sqrt{4\times 2}=\sqrt{4}\sqrt{2}=2\sqrt{2}=2\times 1.41=2.82$$

となります．

(2) すべての値に 20 を加えても，標準偏差は変わらないので，このときの標準偏差は 2.82 となります．

(3) すべての値を 100 倍すると，標準偏差も 100 倍されるので，このときの標準偏差は

$$2.82\times 100=282$$

となります．

問 3

「50 を引く」ということは「(-50) を加える」ことと同じです．また，「半分にする」ということは「1/2 倍する」ことと同じです．

(1) すべての値について (-50) を加えると，平均値は

$$500+(-50)=450$$

となります．

すべての値に (-50) を加えても標準偏差は変わらないので，このときの標準偏差は 10 となります．

(2) すべての値について 1/2 倍すると，平均値は

$$500\times\frac{1}{2}=250$$

となります．

すべての値を 1/2 倍すると標準偏差も 1/2 倍されるので，このときの標準偏差は

$$10\times\frac{1}{2}=5$$

となります．

(3) すべての値について「(-50) を加えたあとに 1/2 倍する」と，平均値は

$$(500+(-50))\times\frac{1}{2}=225$$

となります.

すべての値に (-50) を加えても標準偏差は変わらず，そのあと $1/2$ 倍すると標準偏差も $1/2$ 倍になるので，このときの標準偏差は

$$10 \times \frac{1}{2} = 5$$

となります.

分散は標準偏差の 2 乗なので,

$$5^2 = 25$$

となります.

4.2 データの標準化

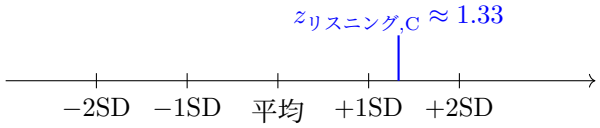
問題 4.4

(1) C さんのリスニング，リーディング，文法テストの点数を標準化して，それぞれの z 値を求めます．z 値は，平均との差を標準偏差で割った値で，平均値からどの方向に標準偏差の何個分ずれているかを表します.

まずリスニングについて，平均 62 点に対して 78 点は 16 点の差があります．この差を標準偏差 12 点で割って標準化すると,

$$z_{\text{リスニング,C}} = \frac{78 - 62}{12} \approx 1.33$$

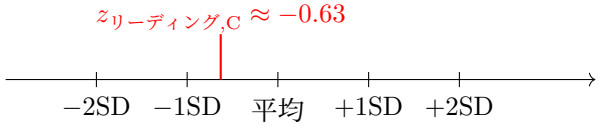
となり，平均より標準偏差（SD）の 1.33 個分だけ上に位置します.



次にリーディングについて，平均 70 点に対して 65 点は -5 点の差があります．この差を標準偏差 8 点で割って標準化すると,

$$z_{\text{リーディング,C}} = \frac{65 - 70}{8} \approx -0.63$$

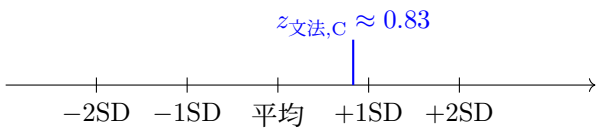
となり，平均より標準偏差（SD）の 0.63 個分だけ下に位置します.



最後に文法について，平均 55 点に対して 60 点は 5 点の差があります．この差を標準偏差 6 点で割って標準化すると,

$$z_{\text{文法,C}} = \frac{60 - 55}{6} \approx 0.83$$

となり，平均より標準偏差（SD）の 0.83 個分だけ上に位置します．



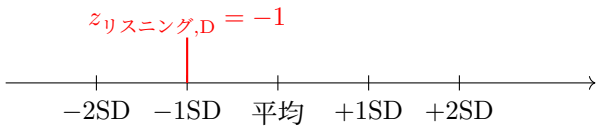
これら 3 つの z 値を比べると，C さんはリスニングの z 値が最も大きいので，この学年のなかで**相対的な位置が最も高い**といえるのは**リスニング**です．また，リーディングの z 値が最も小さいので，**相対的な位置が最も低い**といえるのは**リーディング**です．

(2) D さんのリスニング，リーディング，文法テストの点数についても，同様に標準化して z 値を求めます．

まずリスニングについて，平均 62 点に対して 50 点は −12 点の差があります．この差を標準偏差 12 点で割って標準化すると，

$$z_{\text{リスニング,D}} = \frac{50 - 62}{12} = -1$$

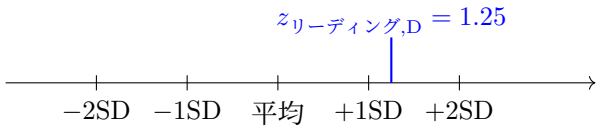
となり，平均より標準偏差（SD）の 1 個分だけ下に位置します．



次にリーディングについて，平均 70 点に対して 80 点は 10 点の差があります．この差を標準偏差 8 点で割って標準化すると，

$$z_{\text{リーディング,D}} = \frac{80 - 70}{8} = 1.25$$

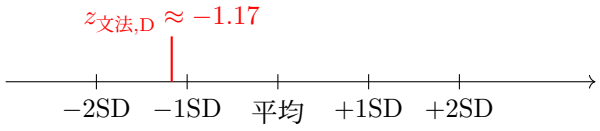
となり，平均より標準偏差（SD）の 1.25 個分だけ上に位置します．



最後に文法について，平均 55 点に対して 48 点は −7 点の差があります．この差を標準偏差 6 点で割って標準化すると，

$$z_{\text{文法,D}} = \frac{48 - 55}{6} \approx -1.17$$

となり，平均より標準偏差（SD）の 1.17 個分だけ下に位置します．



これら 3 つの z 値を比較すると，D さんはリーディングの z 値が最も大きいので，この学年のなかで**相対的な位置が最も高い**といえるのは**リーディング**です．また，文法の z 値が最も小さいので，**相対的な位置が最も低い**といえるのは**文法**です．

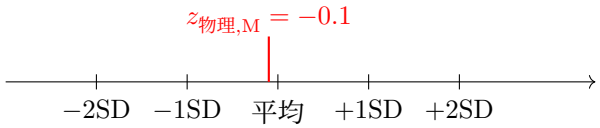
問題 4.5

(1) M さんの物理と化学の点数を標準化して，それぞれの z 値を求めます．z 値は，平均との差を標準偏差で割った値で，平均値からどの方向に標準偏差の何個分ずれているかを表します．

まず物理について，平均 60 点に対して 58 点は -2 点の差があります．この差を標準偏差 20 点で割って標準化すると，

$$z_{\text{物理,M}} = \frac{58 - 60}{20} = -0.1$$

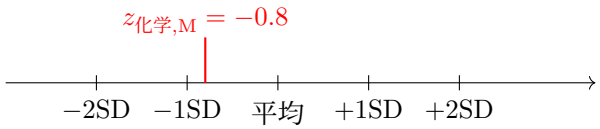
となり，平均より標準偏差（SD）の 0.1 個分だけ下に位置します．



次に化学について，平均 78 点に対して 70 点は -8 点の差があります．この差を標準偏差 10 点で割って標準化すると，

$$z_{\text{化学,M}} = \frac{70 - 78}{10} = -0.8$$

となり，平均より標準偏差（SD）の 0.8 個分だけ下に位置します．



(2) (1) で求めた z 値を用いて，M さんの物理と化学の偏差値を求めます．偏差値は $50 + 10z$ で求められます．

物理の偏差値は，

$$50 + 10 \times (-0.1) = 49$$

化学の偏差値は，

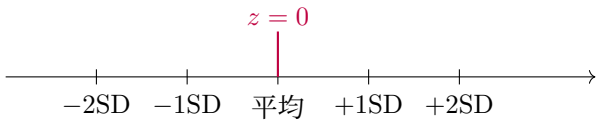
$$50 + 10 \times (-0.8) = 42$$

となります．

これらを比較すると，偏差値がより高い科目は「物理」であることがわかります．点数は物理のほうが低いにもかかわらず，平均値やばらつきのちがいにより，偏差値では物理のほうが高くなっています．

問題 4.6

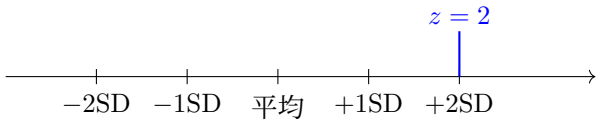
(1) $z = 0$ は「平均」と同じ位置です．よって，値は 110 となります．



(2) $z = 2$ は「平均より標準偏差 2 個分上」です．よって，値は

$$110 + 2 \times 30 = 170$$

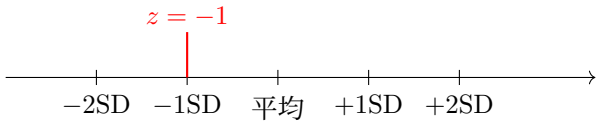
となります．



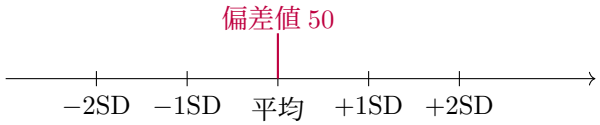
(3) $z = -1$ は「平均より標準偏差 1 個分下」です．よって，値は

$$110 - 1 \times 30 = 80$$

となります．



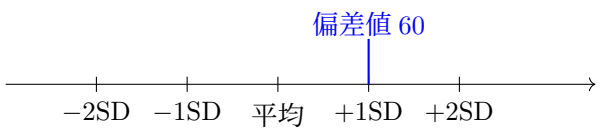
(4) 偏差値 50 は $z = 0$ に対応します．よって，値は 110 となります．



(5) 偏差値 60 は $z = 1$ に対応します．よって，値は

$$110 + 1 \times 30 = 140$$

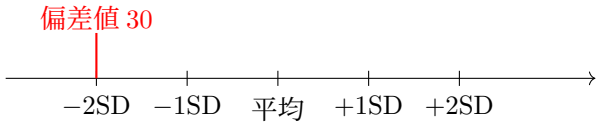
となります．



(6) 偏差値 30 は $z = -2$ に対応します．よって，値は

$$110 - 2 \times 30 = 50$$

となります.



4.2 の確認テスト

問 1

z 値は，平均との差を標準偏差で割った値で，平均値からどの方向に標準偏差の何個分ずれているかを表します.

(1) 身長 182 cm の z 値を求めます.

平均 169 cm に対して 182 cm は 13 cm の差があります．この差を標準偏差 8 cm で割って標準化すると，

$$z = \frac{182 - 169}{8} \approx 1.63$$

となります.

(2) 売上 33 個の日の z 値を求めます.

平均 40 個に対して 33 個は -7 個の差があります．この差を標準偏差 12 個で割って標準化すると，

$$z = \frac{33 - 40}{12} \approx -0.58$$

となります.

(3) 電気使用量 24.8 kWh の日の z 値を求めます.

平均 28 kWh に対して 24.8 kWh は -3.2 kWh の差があります．この差を標準偏差 3.5 kWh で割って標準化すると，

$$z_{=}\frac{24.8 - 28}{3.5} \approx -0.91$$

となります.

問 2

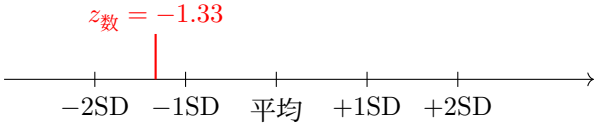
数学の z 値は

$$z_{数} = \frac{42 - 50}{6} \approx -1.33$$

なので，偏差値は

$$50 + 10(-1.33) = 36.7$$

となります.



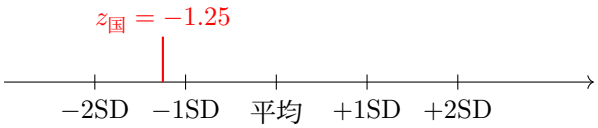
また，国語の z 値は

$$z_{国} = \frac{55 - 65}{8} = -1.25$$

なので，偏差値は

$$50 + 10(-1.25) = 37.5$$

となります．



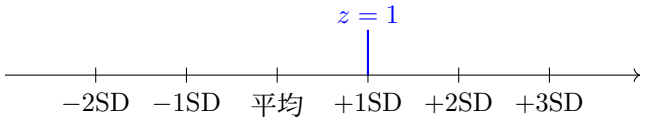
以上より，数学の偏差値は 36.7，国語の偏差値は 37.5 なので，**偏差値がより高いのは国語** です．

問 3

(1) $z = 1$ は「平均より標準偏差 1 個分上」です．よって，値は

$$70 + 1 \times 10 = 80$$

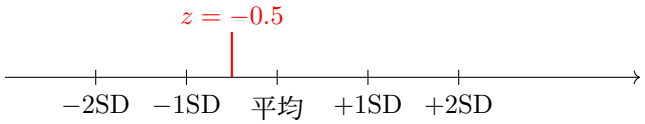
となります．



(2) $z = -0.5$ は「平均より標準偏差 0.5 個分下」です．よって，値は

$$70 - 0.5 \times 10 = 65$$

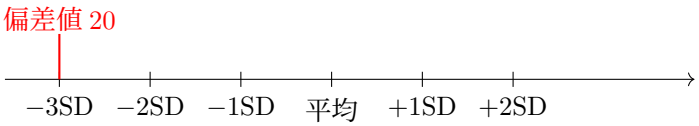
となります．



(3) 偏差値 20 は $z = -3$ に対応します．よって，値は

$$70 - 3 \times 10 = 40$$

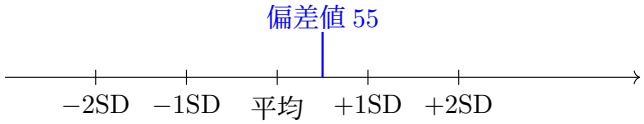
となります．



(4) 偏差値 55 は $z = 0.5$ に対応します．よって，値は

$$70 + 0.5 \times 10 = 75$$

となります．



問 4

(1) ×：偏差値は，得点を平均値と標準偏差で標準化した z 値から求めるため，得点だけで決まるわけではありません．

(2) ○：平均より標準偏差の 5 倍高い得点は $z = 5$ に相当し，偏差値は $50 + 10 \times 5 = 100$ になります．

(3) ○：偏差値 40 は $40 = 50 + 10z$ より $z = -1$ に相当し，平均より標準偏差 1 個分だけ低い位置にあります．

(4) ○：偏差値 70 は $70 = 50 + 10z$ より $z = 2$ に相当し，平均より標準偏差 2 個分だけ高い位置にあります．

(5) ×：偏差値は $50 + 10z$ で求められ，z が大きくなれば 100 を超え，小さくなれば 0 を下回るため，0〜100 に必ず収まるわけではありません．

第 5 章　相関

5.1　共分散

問題 5.1

(1) 点の分布は，左上から右下へ向かって全体として下がっているように並んでいます．このような並び方は，気温が高くなるほどホットコーヒーの売上個数が **減る傾向** を示しています．
(2) 点が右下がりに並んでいるため，気温と売上個数の間には **負の相関** があると判断できます．
(3) 気温 8℃で売上個数が 300 杯の点は，他の点の傾向から大きく外れており，**外れ値** と考えられます．

問題 5.2

身長の合計，体重の合計はそれぞれ

$$160+165+166+170+172+175=1008,\quad 49+55+54+55+53+58=324$$

となります．よって，身長の平均値，体重の平均値はそれぞれ

$$\frac{1008}{6}=168,\quad \frac{324}{6}=54$$

となります．

身長の偏差は，A から F の順にそれぞれ次のようになります．

$$160-168=-8,\ 165-168=-3,\ 166-168=-2,\ 170-168=2,\ 172-168=4,\ 175-168=7$$

体重の偏差は，A から F の順にそれぞれ次のようになります．

$$49-54=-5,\ 55-54=1,\ 54-54=0,\ 55-54=1,\ 53-54=-1,\ 58-54=4$$

よって，偏差どうしの積は，A から F の順にそれぞれ次のようになります．

$$(-8)\times(-5)=40,\ (-3)\times 1=-3,\ (-2)\times 0=0,\ 2\times 1=2,\ 4\times(-1)=-4,\ 7\times 4=28$$

偏差どうしの積の合計は

$$40-3+0+2-4+28=63$$

となります．

共分散は偏差どうしの積の平均なので

$$\frac{63}{6}=10.5$$

より，共分散は 10.5 です．

問題 5.3

問題 5.2 では，身長（cm）と体重（kg）の共分散は 10.5 でした．体重の単位を kg から g に直すと，各値は 1000 倍になります（1 kg = 1000 g）．

値が 1000 倍になると，平均値との差（偏差）もすべて 1000 倍になります．共分散は「偏差どうしの積」の平均なので，体重の偏差が 1000 倍になると，偏差どうしの積も 1000 倍になり，共分散も 1000 倍になります．

したがって，身長（cm）と体重（g）の共分散は

$$10.5 \times 1000 = 10500$$

となります．

5.1 の確認テスト

問 1

(1) 点の分布は，左下から右上へ向かって全体として上がっているように並んでいます．このような並び方は，気温が高くなるほどアイスクリームの売上が **増える傾向** を示しています．

(2) 点が右上がりに並んでいるため，気温と売上個数の間には**正の相関** があると判断できます．

(3)「気温 30.6℃，売上 43 個」を表す点は，他の点の傾向からやや外れています．この点（外れ値）を除くと，他の点はより帯状にまとまって並ぶため，**相関はより強くなる** と判断できます．

問 2

(1) 偏差どうしの積は，A から E の順にそれぞれ次のようになります．

$$(-0.5) \times 0.0 = 0.00, \quad 0.5 \times (-0.5) = -0.25, \quad (-1.5) \times 1.0 = -1.50,$$

$$1.5 \times (-1.0) = -1.50, \quad 0.0 \times 0.5 = 0.00$$

偏差どうしの積の合計は

$$0.00 - 0.25 - 1.50 - 1.50 + 0.00 = -3.25$$

となります．共分散は「偏差どうしの積」の平均なので，

$$\frac{-3.25}{5} = -0.65$$

となり，共分散は -0.65 です．

生徒	勉強時間	睡眠時間	勉強時間の偏差	睡眠時間の偏差	偏差どうしの積
A	2.0	7.0	-0.5	0.0	0.00
B	3.0	6.5	0.5	−0.5	−0.25
C	1.0	8.0	−1.5	1.0	−1.50
D	4.0	6.0	1.5	−1.0	−1.50
E	2.5	7.5	0.0	0.5	0.00
合計	12.5	35.0	0.0	0.0	−3.25
平均値	2.5	7.0	0.0	0.0	−0.65（共分散）

(2) 共分散が負の値であるため，勉強時間と睡眠時間の関係は **負の相関** であると考えられます．

問 3

(1) 身長の合計，座高の合計はそれぞれ

$$158+159+160+161+162+160=960,\quad 85+85+90+95+95+90=540$$

となります．よって，身長の平均値，座高の平均値はそれぞれ

$$\frac{960}{6}=160,\quad \frac{540}{6}=90$$

となります．

身長の偏差は，A から F の順にそれぞれ次のようになります．

$$158-160=-2,\ 159-160=-1,\ 160-160=0,\ 161-160=1,\ 162-160=2,\ 160-160=0$$

座高の偏差は，A から F の順にそれぞれ次のようになります．

$$85-90=-5,\ 85-90=-5,\ 90-90=0,\ 95-90=5,\ 95-90=5,\ 90-90=0$$

よって，偏差どうしの積は，A から F の順にそれぞれ次のようになります．

$$(-2)\times(-5)=10,\ (-1)\times(-5)=5,\ 0\times0=0,\ 1\times5=5,\ 2\times5=10,\ 0\times0=0$$

偏差どうしの積の合計は

$$10+5+0+5+10+0=30$$

となります．

共分散は偏差どうしの積の平均なので

$$\frac{30}{6}=5$$

より，身長（cm）と座高（cm）の共分散は 5 です．

(2) 座高の単位を cm から mm に変更すると，各値は 10 倍になります（1 cm = 10 mm）．値が 10 倍になると，平均値との差（偏差）もすべて 10 倍になり，偏差どうしの積も 10 倍になります．したがって，共分散も 10 倍になります．

よって，身長（cm）と座高（mm）の共分散は

$$5\times10=50$$

となります．

(3) (2) では座高だけを mm に直しましたが，ここでは身長も cm から mm に直すと，身長の値も 10 倍になります（1 cm = 10 mm）．値が 10 倍になると，平均値との差（偏差）もすべて 10 倍になります．したがって，共分散は (2) の値のさらに 10 倍になります．

よって，身長（mm）と座高（mm）の共分散は

$$50\times10=500$$

となります．

5.2　相関係数

問題 5.4

(1) 標準偏差は，偏差の 2 乗の平均値（分散）のルートをとった値です．表の「平均値の $\sqrt{}$」欄より，身長（cm）の標準偏差は 4.93，体重（kg）の標準偏差は 2.71 です．

(2) 問題 5.2 より，共分散は 10.5 です．よって，相関係数は

$$\frac{\text{共分散}}{\text{1 つ目のデータの標準偏差} \times \text{2 つ目のデータの標準偏差}}$$
$$= \frac{10.5}{4.93 \times 2.71} = \frac{10.5}{13.3603} = 0.785\ldots \approx 0.79$$

となります．

(3) 相関係数が 0.79 と 1 に近い正の値であることから，「身長が高い人ほど体重も大きい傾向があり，強い正の相関がみられる」と読み取れます．

問題 5.5

(1) 標準偏差は，偏差の 2 乗の平均値（分散）のルートをとった値です．表の「平均値の $\sqrt{}$」欄より，身長（cm）の標準偏差は 4.93，体重（g）の標準偏差は 2708 です．

(2) 問題 5.3 より，共分散は 10500 です（体重を kg から g にすると，共分散も 1000 倍になります）．したがって，相関係数は

$$\frac{\text{共分散}}{\text{1 つ目のデータの標準偏差} \times \text{2 つ目のデータの標準偏差}}$$
$$= \frac{10500}{4.93 \times 2708} = \frac{10500}{13350.44} = 0.786\ldots \approx 0.79$$

となります．

(3) 体重の単位を kg から g にしたため，共分散は 10.5 から 10500 と 1000 倍になりましたが，標準偏差も同じ割合で変化するため，**相関係数は問題 5.4 で求めた相関係数（体重の単位が kg）のときと同じく 0.79 のまま変わりません** でした．

5.2 の確認テスト

問 1

(1) 分散は，「偏差の 2 乗の平均値」で定義されます．X の偏差の 2 乗の合計は 250，データ数は 10 個なので，

$$X \text{ の分散} = \frac{250}{10} = 25$$

となります．標準偏差は分散のルートをとった値なので，

$$X \text{ の標準偏差} = \sqrt{25} = 5$$

です．

(2) 同様に，Y の偏差の 2 乗の合計は 360，データ数は 10 個なので，

$$Y \text{ の分散} = \frac{360}{10} = 36$$

となります．したがって，標準偏差は

$$Y \text{ の標準偏差} = \sqrt{36} = 6$$

です．

(3) 共分散は，「偏差どうしの積の平均値」で定義されます．X と Y の偏差どうしの積の合計は -225 なので，

$$X \text{ と } Y \text{ の共分散} = \frac{-225}{10} = -22.5$$

となります．

(4) 相関係数は，共分散をそれぞれの標準偏差の積で割ったものです．

$$\text{相関係数} = \frac{\text{共分散}}{X \text{ の標準偏差} \times Y \text{ の標準偏差}} = \frac{-22.5}{5 \times 6} = \frac{-22.5}{30} = -0.75$$

となります．

(5) 相関係数が -0.75 と -1 に近い負の値であることから，X（値段）が大きくなるほど Y（売上個数）が小さくなるという **強い負の相関** があると読み取れます．

問 2

(1) **正解：a**

散布図 A では，右上がりの形になっているので，勉強時間が長いほど点数が高くなる傾向があり，相関係数は正の値になります．

(2) 各文について，○ × と理由は次のとおりです．

(a) ×：相関係数の符号は，共分散の符号で決まります．変数を入れ替えても偏差どうしの積は同じなので，相関係数の符号は変わりません．

(b) ○：相関係数は，2 つの変数を入れ替えても値が変わりません．分母の標準偏差の積も，分子の共分散も同じ値になるからです．

(c) ×：変数を入れ替えても，相関係数の値は変わらないため，1 に近づくことはありません．

(d) ×：変数を入れ替えても，相関係数の値は変わらないため，0 に近づくこともありません．

(3) 各文について，○ × と理由は次のとおりです．

(a) ○：点数に一定の値（5 点）を加えると，全体が上に平行移動するだけで，点の並び方や形は変わりません．

(b) ×：平行移動では，各点の間の距離やばらつきは変わらないので，点のばらつきは大きくなりません．

(c) ×：点数に一定の値を加えても，X と Y の増減の向きは変わらないので，相関係数の符号は変わりません．

(d) ○：平行移動では，点の並び方はそのままなので，相関係数の大きさ（絶対値）は変わりません．

(e) ○：点数に一定の値を加えると平均値は変わりますが，偏差は変わらないため，分散，標準偏差，共分散も変化しません．したがって，相関係数も変わりません．

(4) 各文について，○ × と理由は次のとおりです．

(a) ○：相関係数は，どんな 2 つの変数についても -1 から 1 の間の値をとります．

(b) ×：相関係数の符号は，共分散の符号と常に一致します．共分散が正なら相関係数も正，負なら相関係数も負になります．

(c) ○：2 つの変数の関係が曲線状の形をしている場合では，直線的な関係が弱くなるため，相関係数の絶対値が小さくなることがあります．

(d) ○：相関係数の大きさ（絶対値）は，2 つの変数の直線的な関係の強さを表します．

(e) ○：相関係数は，外れ値があるとその値に強く引きずられやすく，外れ値の影響を受けやすい指標です．

(f) ○：変数を入れ替えても，相関係数は変わりません．また，平行移動では偏差は変わらず，正の定数倍では偏差や標準偏差，共分散が同じ割合で変化するため，相関係数は変わりません．

(g) ×：片方の変数を負の定数倍すると，その変数の偏差の符号がすべて反転し，偏差どうしの積も反転します．そのため，相関係数の符号も反転します．

(h) ×：相関係数が大きくても，それだけで因果関係があるとはいえません．相関と因果関係は別の概念であり，混同しないよう注意が必要です．

問 3

(1) 点 P は，他の点の右上がりの相関から外れた位置にある外れ値で，全体の正の相関を弱めています．点 P を除くと，残りの点は右上がりの傾向がそろい，直線的なまとまりが強くなるため，**相関係数は大きくなります**．

(2) 集団 Q は，右上がりに直線状に並んだ，強い正の相関をつくる点の集まりです．この集団を除くと，残りの点も右上がりではありますが，雲状に広がっており，直線的なまとまりが弱くなります．そのため，**相関係数は小さくなります**．

問 4

(1) **正解：a**

縦軸の値をすべて 2 倍にすると，偏差も 2 倍になり，偏差どうしの積も 2 倍になります．そのため，共分散は 2 倍になります．

(2) **正解：b**

相関係数は，共分散を標準偏差どうしの積で割ったものです．縦軸の値を 2 倍にすると，共分散もその変数の標準偏差も同じ割合（2 倍）で変化します．分子と分母が同じ割合で変化するため，相関係数の値は変わりません．

(3) **正解：d**

縦軸の値をすべて −2 倍にすると，偏差も −2 倍になり，偏差どうしの積も −2 倍になります．そのため，共分散は −2 倍になります．

(4) **正解：c**

相関係数は，共分散を標準偏差どうしの積で割ったものです．縦軸の値をすべて −2 倍にすると，共分散は −2 倍になります（つまり，符号が反転して 2 倍の大きさになります）．一方，標準偏差は分散の（負でない）平方根として定義されるため，2 倍になるだけで常に 0 以上の値をとります．その結果，相関係数は分子だけ符号が反転し，大きさ（絶対値）は変わりません．

(5) **正解：a**

縦軸の単位を cm から m に変えると，値は 100 分の 1 になります．偏差も 100 分の 1 になり，偏差どうしの積も 100 分の 1 になります．そのため，共分散は 100 分の 1 になります．

(6) **正解：b**

相関係数は，共分散を標準偏差どうしの積で割ったものです．縦軸の単位を cm から m に変えると，共分散もその変数の標準偏差も同じ割合（100 分の 1）で変化します．分子と分母が同じ割合で変化するため，相関係数の値は変わりません．

第 6 章　回帰分析

6.1　回帰分析の基本

問題 6.1

(1) 最寄り駅からの距離をもとに家賃を予測するため，**最寄り駅からの距離を説明変数，家賃を目的変数**とみなします.
(2) 気温をもとに暖房の電気使用量を予測するため，**気温を説明変数，暖房の電気使用量を目的変数**とみなします.
(3) 読書時間をもとに読んだページ数を予測するため，**読書時間を説明変数，読んだページ数を目的変数**とみなします.

問題 6.2

外気温（℃）を x，売上個数を y とすると，

$$y = 120 + 8x$$

と表され，切片は $a = 120$，傾きは $b = 8$ です.

(1) 傾きが 8 であるため，外気温が 1 ℃ 上がると，売上個数は **8 個増える** と予測されます.
(2) 切片が 120 であるため，外気温が 0 ℃ のときの売上個数は **120 個** と予測されます.
(3) 外気温が 10 ℃ のときは $x = 10$ なので，これを $y = 120 + 8x$ に代入すると，

$$y = 120 + 8 \times 10 = 200$$

となります．よって，このときの売上個数は **200 個** と予測されます.
(4) 外気温が 30 ℃ のときは $x = 30$ なので，これを $y = 120 + 8x$ に代入すると，

$$y = 120 + 8 \times 30 = 360$$

となります．よって，このときの売上個数は **360 個** と予測されます.

6.1 の確認テスト

問 1

(1) 降水量をもとに日照時間を予測するため，**降水量を説明変数，日照時間を目的変数** とみなします.
(2) 歩く速さをもとに 1 日の歩数を予測するため，**歩く速さを説明変数，1 日の歩数を目的変数** とみなします.
(3) 部屋の温度をもとに飲み物を飲む量を予測するため，**部屋の温度を説明変数，飲み物を飲む量を目的変数** とみなします.

問 2

(1) ○ ：説明変数と目的変数の選び方は，分析の目的によって変わることがあります．
(2) ○ ：説明変数と目的変数の関係があっても，因果関係があるとは限りません．
(3) × ：相関が強くても，因果関係があるとは限りません．
(4) ○ ：説明変数と目的変数は，表や散布図でどちらを横軸・縦軸に置いたかによって決まるものではありません．
(5) × ：説明変数と目的変数の選び方は，相関だけで自然に決まるわけではありません．

問 3

(1) 所得をもとに消費支出を予測するため，**所得を説明変数，消費支出を目的変数** とみなします．
(2) 傾きが 0.68 であるため，所得が 1 万円増えると，消費支出は **0.68 万円（6800 円）増える** と予測されます．
(3) 切片が 5.2 であるため，所得が 0 万円のときの消費支出は **5.2 万円** と予測されます．
(4) 所得が 40 万円のときは

$$5.2 + 0.68 \times 40 = 32.4$$

となるため，消費支出は **32.4 万円** と予測されます．

6.2　回帰直線の推定

問題 6.3

(1) 共分散が -0.15，x の標準偏差が 0.5 なので，傾き b は

$$b = \frac{-0.15}{0.5^2} = \frac{-0.15}{0.25} = -0.6$$

となります．
(2) x の平均値が 2.0，y の平均値が 1.5 なので，切片 a は

$$a = 1.5 - (-0.6 \times 2.0) = 1.5 + 1.2 = 2.7$$

となります．
したがって，回帰直線は

$$y = 2.7 - 0.6x$$

となります．
(3) K くんが塾で 3 時間勉強した日は $x = 3$ なので，これを $y = 2.7 - 0.6x$ に代入すると，

$$y = 2.7 - 0.6 \times 3 = 0.9$$

となります．よって，このときの自宅学習時間は **およそ 0.9 時間** と予測されます．
(4) K くんが塾で勉強しなかった日は $x = 0$ なので，これを $y = 2.7 - 0.6x$ に代入すると，

$$y = 2.7 - 0.6 \times 0 = 2.7$$

となります．よって，このときの自宅学習時間は **およそ 2.7 時間** と予測されます．

問題 6.4

(1) 相関係数が -0.40，x の標準偏差が 0.4，y の標準偏差が 0.5 なので，傾き b は

$$b = -0.40 \times \frac{0.5}{0.4} = -0.40 \times 1.25 = -0.50$$

となります．したがって，傾きは -0.50 です．

(2) x の平均値が 1.5，y の平均値が 1.2 なので，切片 a は

$$a = 1.2 - (-0.50) \times 1.5 = 1.95$$

となります．

したがって，回帰直線は

$$y = 1.95 - 0.50x$$

となります．

(3) T くんが塾で 3 時間勉強した日は $x = 3$ なので，これを $y = 1.95 - 0.50x$ に代入すると，

$$y = 1.95 - 0.50 \times 3 = 0.45$$

となります．よって，このときの自宅学習時間は **およそ 0.45 時間** と予測されます．

(4) T くんが塾で勉強しなかった日は $x = 0$ なので，これを $y = 1.95 - 0.50x$ に代入すると，

$$y = 1.95 - 0.50 \times 0 = 1.95$$

となります．よって，このときの自宅学習時間は **およそ 1.95 時間** と予測されます．

問題 6.5

もとの回帰直線は

$$y = -20 + 0.45x$$

です．

(1) 身長を m で表すために

$$x' = \frac{x}{100}$$

とおくと，$x = 100x'$ となります．これを回帰直線の式に代入すると，

$$y = -20 + 0.45(100x')$$

つまり，

$$y = -20 + 45x'$$

となります．これが，身長（m） x' から体重（kg） y を予測する回帰直線の式ということになります．

(2) 体重を g で表すために

$$y' = 1000y$$

とおきます．もとの回帰直線

$$y = -20 + 0.45x$$

の両辺に 1000 をかけると，

$$y' = 1000y = 1000(-20 + 0.45x)$$

つまり，

$$y' = -20000 + 450x$$

となります．これが，身長（cm） x から体重（g） y' を予測する回帰直線の式ということになります．

(3) 身長を m，体重を g で表すために

$$x' = \frac{x}{100}, \qquad y' = 1000y$$

とおきます．(1) の結果より，

$$y = -20 + 45x'$$

であるので，この両辺に 1000 をかけると，

$$y' = 1000y = 1000(-20 + 45x')$$

つまり，

$$y' = -20000 + 45000x'$$

となります．これが，身長（m） x' から体重（g） y' を予測する回帰直線の式ということになります．

6.2 の確認テスト

問 1

回帰直線を $y = a + bx$ とおきます．

相関係数が 0.60，x の標準偏差が 10，y の標準偏差が 8 なので，傾き b は次のように求められます．

$$b = \text{相関係数} \times \frac{y \text{ の標準偏差}}{x \text{ の標準偏差}} = 0.60 \times \frac{8}{10} = 0.48$$

また，x の平均値 $\bar{x}$ が 50，y の平均値 $\bar{y}$ が 70 であることから，切片 a は

$$a = \bar{y} - b\bar{x} = 70 - 0.48 \times 50 = 46$$

となります．したがって，回帰直線は

$$y = 46 + 0.48x$$

となります．

問 2

(1) 標準偏差は，偏差の 2 乗の平均値（分散）のルートをとった値です．表の「平均値の $\sqrt{}$」欄より，読書時間（分）の標準偏差は 11.90，国語テストの得点（点）の標準偏差は 7.30 です．

(2) 共分散は，「偏差どうしの積の平均値」で定義されます．偏差どうしの積の合計は，表の偏差を用いて次のように求められます．

$$5 \times 2 + (-20) \times 0 + 15 \times 14 + (-10) \times (-10) + 10 \times (-2) + 0 \times (-4) = 300$$

したがって，共分散は

$$\frac{300}{6} = 50$$

となります．

(3) 相関係数は，共分散をそれぞれの標準偏差の積で割ったものです．

$$\text{相関係数} = \frac{\text{共分散}}{x \text{ の標準偏差} \times y \text{ の標準偏差}} = \frac{50}{11.90 \times 7.30} = \frac{50}{86.87} \approx 0.58$$

となります．

(4) 回帰直線 $y = a + bx$ の傾き b は，x の分散が $850/6$ であることから，

$$b = \frac{\text{共分散}}{x \text{ の分散}} = \frac{50}{850/6} = \frac{300}{850} = \frac{6}{17} \approx 0.35$$

となります．

また，読書時間の平均値 $\bar{x}$ が 45，国語テストの平均値 $\bar{y}$ が 70 なので，切片 a は，丸める前の傾き $b = 6/17$ を用いて

$$a = \bar{y} - b\bar{x} = 70 - \frac{6}{17} \times 45 = \frac{920}{17} \approx 54.1$$

となります．

したがって，回帰直線は

$$y = 54.1 + 0.35x$$

となります．

(5) 1 週間の平均読書時間が 50 分のときは $x = 50$ です．予測値の計算では，丸める前の傾きと切片を用いると，

$$y = \frac{920}{17} + \frac{6}{17} \times 50 = \frac{1220}{17} \approx 71.8$$

となります．

よって，このときの国語テストの得点は **71.8 点**と予測されます．

問 3

(1) ○：傾きは，説明変数が 1 増えたときに目的変数がどれだけ変化するとみなせる量です．

(2) ×：切片は，説明変数が 0 のときの目的変数の予測値を表します．

(3) ×：相関係数の大きさと傾きの大きさは別の概念です．

(4) ○：観測範囲の外側で予測すると，直線関係が続くとは限りません．

(5) ○：グループ差や外れ値があると，回帰直線の形が大きく影響を受けることがあります．
(6) ○：関係が直線で近似できない場合，曲線のほうが当てはまりがよいことがあります．
(7) ×：説明変数の単位を変えると，1 単位あたりの変化量が変わるため，傾きも変わります．
(8) ×：目的変数の単位を変えると，傾きと切片は同じ倍率で変わります．

問 4

(1) 気温を華氏で表すために

$$x' = 1.8x + 32$$

とおきます．この式を x について解くと，

$$x = \frac{x' - 32}{1.8}$$

となります．これをもとの回帰直線

$$y = -30 + 5x$$

に代入すると，

$$y = -30 + 5\left(\frac{x' - 32}{1.8}\right)$$

つまり，

$$y = -\frac{1070}{9} + \frac{25}{9}x'$$

となります．これが，華氏で表した気温 x' から売上個数 y を予測する回帰直線の式ということになります．

(2) 売上個数を箱で表すために

$$y' = \frac{y}{6}$$

とおきます（1 箱 6 個入り）．もとの回帰直線

$$y = -30 + 5x$$

の両辺を 6 で割ると，

$$y' = \frac{y}{6} = \frac{-30 + 5x}{6}$$

つまり，

$$y' = -5 + \frac{5}{6}x$$

となります．これが，気温（℃） x から箱で表した売上 y' を予測する回帰直線の式ということになります．

(3) 気温を華氏，売上を箱で表すために

$$x' = 1.8x + 32, \qquad y' = \frac{y}{6}$$

とおきます．(1) の結果より，

$$y = -\frac{1070}{9} + \frac{25}{9}x'$$

であるので，この両辺を 6 で割ると，

$$y' = \frac{y}{6} = \frac{-\frac{1070}{9} + \frac{25}{9}x'}{6}$$

つまり，

$$y' = -\frac{1070}{54} + \frac{25}{54}x'$$

となります．これが，華氏で表した気温 x' から箱で表した売上 y' を予測する回帰直線の式ということになります．

第 7 章　質的データの表現

7.1　質的変数を表すグラフ

問題 7.1

(1) **レーダーチャート**
　放射状の軸上に値をとり，家電製品の複数の特徴を同時に比較できます．多面的な比較に適したグラフです．

(2) **棒グラフ**
　棒の高さを使って，5 種類の飲料の販売本数を直接比較できます．

(3) **円グラフ（または，帯グラフ）**
　円を分割して，家庭の年間支出が全体のなかでどれだけの割合を占めるかを表します．割合の構成を表すには帯グラフを用いることもでき，複数の家庭を比較したい場合には帯グラフがとくに有効です．

(4) **帯グラフ（または，積み上げ棒グラフ）**
　4 つの地域それぞれについて，交通手段の利用割合のちがいを，全体の長さをそろえた帯で比較できます．一方，地域ごとの人数の大小も表したい場合には，全体の長さが異なる積み上げ棒グラフを用いることができます．

(5) **積み上げ棒グラフ（または，棒グラフ）**
　5 つの企業について，職種別従業員数（営業・事務・技術など）を比較できます．合計人数と内訳を同時に表したい場合には積み上げ棒グラフが適しており，職種ごとの人数を直接比べたい場合には棒グラフを用いることができます．

（ここに示した解答は一例であり，これに限られるものではありません）

問題 7.2

(1) ○：棒グラフより，情報学部 240 人は法学部 210 人より多く，経済学部 280 人より少ないことがわかります．

(2) ×： 円グラフより，A 店 45%と C 店 20%の合計は 65%であり，全体の 7 割（70%）は超えていません．

(3) ×：帯グラフより，20 代の割合は地域 X が 30%，地域 Y が 25%であり，地域 X のほうが低いという読み取りは誤りです．

(4) ○： レーダーチャートより，価格 3 より評価が高い項目は，走行距離 4，内装 5，装備 4 の 3 つあります．

(5) ○：積み上げ棒グラフより，経営学部では女子学生 120 人／ 320 人＝ 37.5%であり，女子学生の割合は 4 割未満になっています．

7.1 の確認テスト

問 1

(1) **積み上げ棒グラフ（または，棒グラフ）**

部署ごとの従業員数を，正社員・契約社員・パートなどの内訳とともに比較できます．合計人数と内訳を同時に表したい場合には積み上げ棒グラフが適しています．

(2) **棒グラフ**

棒の高さを使って，7 種類の野菜の販売量を直接比較できます．数量の大小を表すのに適したグラフです．

(3) **円グラフ（または，帯グラフ）**

円を分割して，年間支出の内訳が全体のなかでどれだけの割合を占めるかを表します．割合の構成を表すには帯グラフを用いることもでき，複数の企業を比較したい場合には帯グラフがとくに有効です．

(4) **帯グラフ（または，積み上げ棒グラフ）**

4 つの学年それぞれについて，クラブ活動の参加状況の割合のちがいを，全体の長さをそろえた帯で比較できます．一方，各学年の人数の大小も表したい場合には，全体の長さが異なる積み上げ棒グラフを用いることができます．

(5) **レーダーチャート**

放射状の軸上に値をとり，製品の複数の特徴を同時に比較できます．多面的な評価を視覚的に示すのに適したグラフです．

問 2

(1) ○： 太陽光は 2026 年 20%から 2030 年 33%へ増加しています．

(2) ×： 風力は 2027 年 17%，2028 年 20%であり，2027 年のほうが大きいわけではありません．

(3) ×： 水力は 5 年間を通して常に最も大きいわけではなく，2030 年には太陽光 33%が水力 30%より大きくなっています．

(4) ×： その他は 20%，18%，16%，14%，12%と年々減少しています．

(5) ○： 太陽光と風力の合計は 2026 年 35%から 2030 年 58%へ増加しています．

7.2　質的変数の関係を表す表とグラフ

問題 7.3

(1) 全体 200 人のうち，週 3 回以上運動している人は 40 人なので，全体のうち週 3 回以上運動している人の割合は

$$\frac{40}{200} = \frac{1}{5}$$

(2) 40 代 80 人のうち，運動をほとんどしない人は 50 人なので，40 代の人のうち運動をほとんどしない人の割合は

$$\frac{50}{80} = \frac{5}{8}$$

(3) 運動をほとんどしない 90 人のうち，20 代の人は 20 人なので，運動をほとんどしない人のうち 20 代の

人の割合は

$$\frac{20}{90} = \frac{2}{9}$$

(4) 30 代 60 人のうち，週 1〜2 回運動している人は 28 人なので，30 代の人のうち週 1〜2 回運動している人の割合は

$$\frac{28}{60} = \frac{7}{15}$$

問題 7.4

(1) ライブ配信中心 80 人のうち，「とても満足」は 25 人，「不満」は 10 人です．したがって，人数の比は

$$25 : 10 = 5 : 2$$

となります．

(2) 録画視聴中心 60 人のうち，「あまり満足していない」15 人と「不満」7 人を合わせて

$$15 + 7 = 22$$

です．よって割合は

$$\frac{22}{60} = \frac{11}{30}$$

となります．

(3)「とても満足」60 人と「まあ満足」60 人を合わせて

$$60 + 60 = 120$$

です．そのうち録画視聴中心の学生は

$$20 + 18 = 38$$

となります．したがって割合は

$$\frac{38}{120} = \frac{19}{60}$$

となります．

(4)「不満」と答えた学生は 21 人であり，全体 180 人のうちの割合は

$$\frac{21}{180} = \frac{7}{60}$$

となります．

(5) (a) ○ :「とても満足」の割合は，ライブ配信中心 25 ／ 80 ＝ 31.25%，録画視聴中心 20 ／ 60 ＝ 33.3%，対面併用 15 ／ 40 ＝ 37.5%となり，ライブ配信中心が最も低くなっています．

(b) × :録画視聴中心 60 人のうち，「とても満足」20 人，「まあ満足」18 人，「あまり満足していない」15 人，「不満」7 人となっています．割合はそれぞれ 60 人で割って求めますが，分母が同じなので，人数が最も多い「とても満足」の割合が最も高く，「まあ満足」ではありません．

(c) ○ :ライブ配信中心では，「とても満足」25 人と「まあ満足」30 人の合計は 55 人，「あまり満足していない」15 人と「不満」10 人の合計は 25 人です．

これらの割合はどちらも 80 人で割って求めるため，分母が同じときは人数の比と割合の比が一致します．そこで，

$$\frac{55}{25} = 2.2$$

となるので，割合も 2 倍以上になっていることがわかります．

問題 7.5

(1) ○：小学生では，水・お茶・ジュースはいずれも高さが 1/3 です．同じ小学生のなかでの割合が等しいため，それぞれを選んだ人数も同じです．

(2) ○：中学生の高さは，水が 1/6，お茶が 1/3，ジュースが 1/2 です．よって，水を選んだ人数が最も少ないです．

(3) ×：高校生の高さは，水が 1/6，お茶が 1/3 です．よって，お茶より水を選んだ人数のほうが少ないです．

(4) ×：小学生のジュースの人数に対応する面積は

$$\frac{1}{2} \times \frac{1}{3} = \frac{1}{6},$$

中学生のジュースの面積は

$$\frac{1}{3} \times \frac{1}{2} = \frac{1}{6}.$$

よって両者は等しく，小学生のほうが多いとはいえません．

(5) ○：中学生のお茶の面積は

$$\frac{1}{3} \times \frac{1}{3} = \frac{1}{9},$$

高校生のお茶の面積は

$$\frac{1}{6} \times \frac{1}{3} = \frac{1}{18}.$$

よって，中学生のお茶を選んだ人数のほうが多いです．

(6) ×：全体のお茶の面積は

$$\frac{1}{2} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{3} + \frac{1}{6} \times \frac{1}{3} = \frac{1}{3},$$

全体のジュースの面積は

$$\frac{1}{2} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{2} + \frac{1}{6} \times \frac{1}{2} = \frac{5}{12}.$$

よって，全体ではお茶を選んだ人数のほうが少ないです．

7.2 の確認テスト

問 1

(1) 改善した患者は 80 人なので，全体 180 人のうち改善した患者の割合は

$$\frac{80}{180} = \frac{4}{9}$$

(2) 治療法 B を用いた患者 60 人のうち，改善した患者は 25 人，改善しなかった患者も 25 人です．したがって，人数の比は

$$25 : 25 = 1 : 1$$

となります．

(3) 治療法 A を用いた患者 70 人のうち，改善した患者 40 人とどちらともいえない患者 10 人の合計は 50 人なので，その割合は

$$\frac{50}{70} = \frac{5}{7}$$

となります．

(4) 治療法 B または治療法 C を用いた患者の人数は

$$60 + 50 = 110$$

そのうち改善しなかった患者の人数は

$$25 + 15 = 40$$

したがって割合は

$$\frac{40}{110} = \frac{4}{11}$$

となります．

(5) 改善した患者は 80 人であり，そのうち副作用が出た患者の人数は

$$15 + 10 + 5 = 30$$

したがって割合は

$$\frac{30}{80} = \frac{3}{8}$$

となります．

問 2

(1) モザイク図の横幅は，各学年が全体に占める人数の割合を表しています．1 年生の幅は 3/8 で，2 年生と 3 年生はともに 2/8，4 年生は 1/8 です．したがって，最も幅が大きい 1 年生の学生が最も多いといえます．

(2) 1 年生のなかで，平日にアルバイトをしない学生の割合は 2/8 です．一方，平日にアルバイトをする学生の割合は

$$\frac{2}{8} + \frac{2}{8} + \frac{2}{8} = \frac{6}{8}$$

となります．同じ 1 年生のなかなので，人数はこれらの割合に比例します．したがって，人数の比は

$$\frac{2}{8} : \frac{6}{8} = 1 : 3$$

となります．

(3) 3 年生の学生のうち，「1〜3 時間」と答えた学生の割合はモザイク図の高さから

$$\frac{3}{8}$$

となります．

(4) モザイク図では，長方形の面積がその区分に属する人数に比例します．このことから，「5 時間以上」と答えた学生についても，対応する長方形の面積を比べれば人数の比較ができます．1 年生では

$$\frac{3}{8} \times \frac{2}{8} = \frac{6}{64},$$

4 年生では

$$\frac{1}{8} \times \frac{2}{8} = \frac{2}{64}$$

となります．よって，1 年生のほうが人数が多いといえます．

(5) 1 年生と 4 年生では，全体に占める人数の割合（幅）が 3/8 と 1/8 で異なります．したがって，「5 時間以上」と答えた学生の人数そのものを比べるだけでは，学年による傾向のちがいを公平に比べているとはいえません．よって，この比較は適切とはいえません．

第 8 章　データの分布の形

8.1　度数分布表とヒストグラム

問題 8.1

度数を数える前に，データを小さい順に並べ替えておくと，どの階級にいくつ入るかを確認しやすくなります．今回のデータを並べ替えると，次のようになります．

$$\{2.8,\ 2.9,\ 3.2,\ 3.3,\ 3.5,\ 3.7,\ 3.9,\ 4.0,\ 4.1,\ 4.3,\ 4.4,\ 4.8,\ 5.0,\ 5.1,\ 5.2\}$$

階級を決めるときには，データの最小値と最大値を見て，どの範囲を区切ればよいかを考えます．このデータでは，最小値は 2.8，最大値は 5.2 です．

階級幅を 0.5 とすると，階級はたとえば次のように設定できます．

2.5以上3.0未満, 3.0以上3.5未満, 3.5以上4.0未満, 4.0以上4.5未満, 4.5以上5.0未満, 5.0以上5.5未満

次に，各階級に入るデータの個数（度数）を数えます．たとえば「2.5 以上 3.0 未満」には

$$2.8,\ 2.9$$

が入り，度数は 2 になります．すべての階級について度数を数えると，合計が 15 個になることを確認できます．

以上より，次のような度数分布表になります．

階級	度数
2.5 以上 3.0 未満	2
3.0 以上 3.5 未満	2
3.5 以上 4.0 未満	3
4.0 以上 4.5 未満	4
4.5 以上 5.0 未満	1
5.0 以上 5.5 未満	3
合計	15

問題 8.2

(1) 度数が最も大きい階級は「80 以上 100 未満」（度数 10）です．

(2)「80 以上 100 未満」に 10，「100 以上 120 未満」に 8 が入るので，

$$10 + 8 = 18$$

より，80 以上 120 未満に入る度数は 18 です．

(3) 全体のデータ数が 30 なので，相対度数は「度数 $\div$ 30」で求めます．累積度数は，小さい階級から順に度数を足し合わせたもの，累積相対度数は，相対度数を小さい階級から順に足し合わせたものです．

まとめると次のようになります.

階級	度数	相対度数	累積度数	累積相対度数
40 以上 60 未満	3	3/30	3	3/30
60 以上 80 未満	7	7/30	10	10/30
80 以上 100 未満	10	10/30	20	20/30
100 以上 120 未満	8	8/30	28	28/30
120 以上 140 未満	2	2/30	30	30/30

(4) 階級値は，階級の下端と上端の平均をとって求めます．この度数分布表では，小さい階級から順に

$$\frac{40+60}{2}=50,\ \frac{60+80}{2}=70,\ \frac{80+100}{2}=90,\ \frac{100+120}{2}=110,\ \frac{120+140}{2}=130$$

となります.

平均値は,「全体の合計 ÷ データ数」で求めます．度数分布表では，もとのデータの合計がわからないので，各階級のデータを階級値で代表させて合計の近似値を計算します.

まず,「階級値 × 度数」をすべて足し合わせて，全体の合計のおおよその値を求めます.

$$50\times 3+70\times 7+90\times 10+110\times 8+130\times 2=2680$$

平均値の近似値は，この合計をデータ数 30 で割って求めるので,

$$\frac{2680}{30}\approx 89.3$$

より，89.3 となります.

(5) 中央値は，データを小さい順に並べたときの真ん中の値です．全体が 30 個なので，中央値は 15 番目と 16 番目の位置にあたります.

累積度数を小さい階級から順に計算すると,

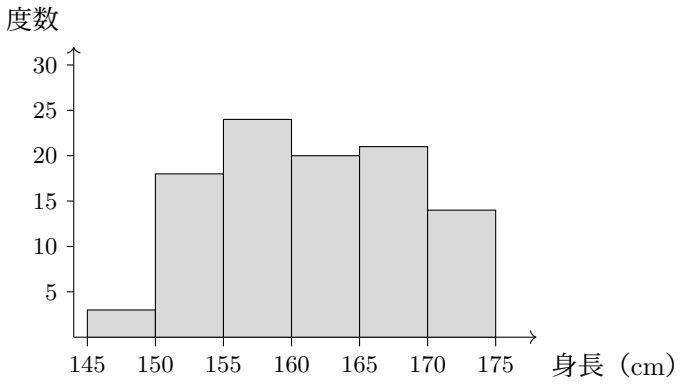
$$\begin{aligned}
&40\text{以上}60\text{未満}\to 3\\
&60\text{以上}80\text{未満}\to 3+7=10\\
&80\text{以上}100\text{未満}\to 10+10=20\\
&100\text{以上}120\text{未満}\to 20+8=28\\
&120\text{以上}140\text{未満}\to 28+2=30
\end{aligned}$$

となります.

15 番目と 16 番目は，累積度数が 10 を超えて 20 になる「80 以上 100 未満」の階級に入ります．この階級の階級値は 90 なので，中央値のおおよその値は 90 です.

問題 8.3

100 人の身長の度数分布表にもとづくヒストグラムは，次のようになります.



8.1 の確認テスト

問 1

データを小さい順に並べ替えると，次のようになります．

$$\left\{\begin{array}{l} 3,\ 6,\ 11,\ 13,\ 14,\ 20,\ 21,\ 23,\ 24,\ 26, \\ 27,\ 31,\ 33,\ 35,\ 35,\ 35,\ 38,\ 40,\ 41,\ 43, \\ 44,\ 46,\ 50,\ 50,\ 50,\ 51,\ 51,\ 52,\ 53,\ 54 \end{array}\right\}$$

(1) 階級を決めるときには，データの最小値と最大値を見て，どの範囲を区切ればよいかを考えます．このデータでは，最小値は 3，最大値は 54 です．

階級幅を 10 とすると，階級はたとえば次のように設定できます．

0以上10未満, 10以上20未満, 20以上30未満,

30以上40未満, 40以上50未満, 50以上60未満

次に，各階級に入るデータの個数（度数）を数えます．たとえば「0 以上 10 未満」には

3, 6

が入り，度数は 2 になります．

すべての階級について度数を数えると，合計が 30 個になることを確認できます．

以上より，次のような度数分布表になります．

階級	度数
0 以上 10 未満	2
10 以上 20 未満	3
20 以上 30 未満	6
30 以上 40 未満	6
40 以上 50 未満	5
50 以上 60 未満	8
合計	30

(2) 度数分布表を見ると，最も度数が大きい階級は「50 以上 60 未満」（度数 8） です．

(3)「30 以上 40 未満」に 6 個，「40 以上 50 未満」に 5 個，「50 以上 60 未満」に 8 個あるので，

$$6+5+8=19$$

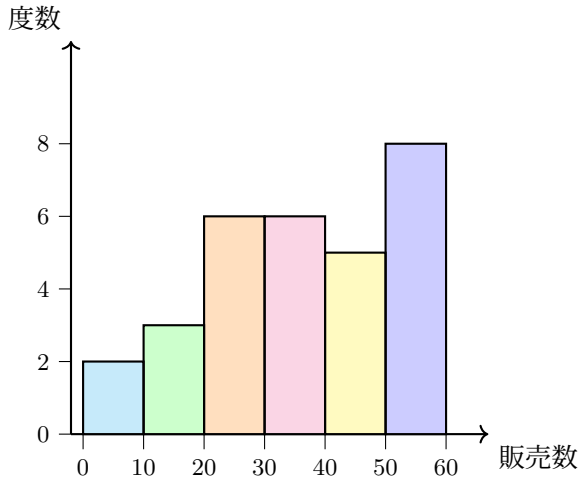
より，販売数が 30 個以上の月は 19 個あることがわかります．

(4) 相対度数は「度数 ÷ 30」で求めます．累積度数は小さい階級から順に足し合わせ，累積相対度数は相対度数を順に足し合わせて求めます．

まとめると次のようになります．

階級	度数	相対度数	累積度数	累積相対度数
0 以上 10 未満	2	2/30	2	2/30
10 以上 20 未満	3	3/30	5	5/30
20 以上 30 未満	6	6/30	11	11/30
30 以上 40 未満	6	6/30	17	17/30
40 以上 50 未満	5	5/30	22	22/30
50 以上 60 未満	8	8/30	30	30/30

(5) 度数分布表にもとづくヒストグラムは，次のようになります．



問 2

(1) まず，全体のデータ数を求めます．度数をすべて足し合わせると，

$$8+10+13+10+8=49$$

なので，データ数は 49 です．

階級値は，階級の下端と上端の平均をとって求めます．この度数分布表では，小さい階級から順に

$$\frac{0+100}{2}=50,\ \frac{100+200}{2}=150,\ \frac{200+300}{2}=250,\ \frac{300+400}{2}=350,\ \frac{400+500}{2}=450$$

となります．

平均値は，「全体の合計 ÷ データ数」で求めます．度数分布表では，もとのデータの合計がわからないので，各階級のデータを階級値で代表させて合計の近似値を計算します．

まず，「階級値 × 度数」をすべて足し合わせて，全体の合計のおおよその値を求めます．

$$50\times 8+150\times 10+250\times 13+350\times 10+450\times 8=12250$$

平均値の近似値は，この合計をデータ数 49 で割って求めるので，

$$\frac{12250}{49} = 250$$

より，250 となります．

(2) 中央値は，データを小さい順に並べたときの真ん中の値です．全体が 49 個なので，中央値は 25 番目の位置にあたります．

累積度数を小さい階級から順に計算すると，

$$\begin{aligned}
&0\text{以上}100\text{未満} \ : \ 8,\\
&100\text{以上}200\text{未満} \ : \ 8+10=18,\\
&200\text{以上}300\text{未満} \ : \ 18+13=31,\\
&300\text{以上}400\text{未満} \ : \ 31+10=41,\\
&400\text{以上}500\text{未満} \ : \ 41+8=49
\end{aligned}$$

となります．

25 番目は，累積度数が 18 を超えて 31 になる「200 以上 300 未満」の階級に入ります．この階級の階級値は 250 なので，中央値のおおよその値は 250 です．

(3) 度数が最も大きい階級は，「200 以上 300 未満」（度数 13）です．

問 3

(1) 棒グラフの高さを比べると，「150 以上 200 未満」の区間の棒が最も高くなっています．したがって，最も人数が多い水分摂取量の区間は「150 以上 200 未満」です．

(2) 1 日の水分摂取量が 200 mL 未満の人は，「0 以上 50 未満」「50 以上 100 未満」「100 以上 150 未満」「150 以上 200 未満」の 4 つの区間に入っている人の合計です．

棒の高さから人数を読むと，

$$\begin{aligned}
&0\text{以上}50\text{未満} \ : \ 4\text{人},\\
&50\text{以上}100\text{未満} \ : \ 8\text{人},\\
&100\text{以上}150\text{未満} \ : \ 14\text{人},\\
&150\text{以上}200\text{未満} \ : \ 18\text{人}
\end{aligned}$$

なので，合計は

$$4+8+14+18=44$$

となります．よって，1 日の水分摂取量が 200 mL 未満の人は 44 人です．

(3) 中央値は，データを小さい順に並べたときの真ん中の値です．全体が 80 人なので，中央値は 40 番目と 41 番目の位置にあたります．

各区間までの累積人数を，小さい区間から順に求めると，

$$\begin{aligned}
0\text{以上}50\text{未満} &: 4,\\
50\text{以上}100\text{未満} &: 4+8=12,\\
100\text{以上}150\text{未満} &: 12+14=26,\\
150\text{以上}200\text{未満} &: 26+18=44,\\
200\text{以上}250\text{未満} &: 44+16=60,\\
250\text{以上}300\text{未満} &: 60+12=72,\\
300\text{以上}350\text{未満} &: 72+6=78,\\
350\text{以上}400\text{未満} &: 78+2=80
\end{aligned}$$

となります．

40 番目と 41 番目は，累積人数が 26 を超えて 44 になる「150 以上 200 未満」の区間に入ります．この区間の階級値は

$$\frac{150+200}{2}=175$$

なので，中央値のおおよその値は 175 mL です．

8.2　四分位数と箱ひげ図

問題 8.4

まず第 2 四分位数，つまり，中央値を求めます．

データ全体の中央値は，20 個の真ん中（10 番目と 11 番目の平均）で求めます．10 番目と 11 番目の値はそれぞれ 57，58 なので，

$$\frac{57+58}{2}=57.5$$

より，第 2 四分位数は 57.5 となります．

次に，この中央値の「位置」（10 番目と 11 番目の間）で，20 個のデータを下半分と上半分に分けます．

- **下半分の 10 個のデータ** は下記です．

  48, 49, 50, 51, 52, 53, 54, 55, 56, 57

  この 10 個のデータの中央値（5 番目と 6 番目の平均）が第 1 四分位数です．5 番目と 6 番目の値は 52，53 なので，

  $$\text{第 1 四分位数}=\frac{52+53}{2}=52.5$$

  となります．
- **上半分の 10 個のデータ** は下記です．

  58, 59, 60, 61, 62, 63, 64, 65, 66, 68

  この 10 個のデータの中央値（5 番目と 6 番目の平均）が第 3 四分位数です．5 番目と 6 番目の値は 62，63 なので，

  $$\text{第 3 四分位数}=\frac{62+63}{2}=62.5$$

となります.

よって,

四分位範囲 = 第3四分位数 − 第1四分位数 = 62.5 − 52.5 = 10

となります.

問題 8.5

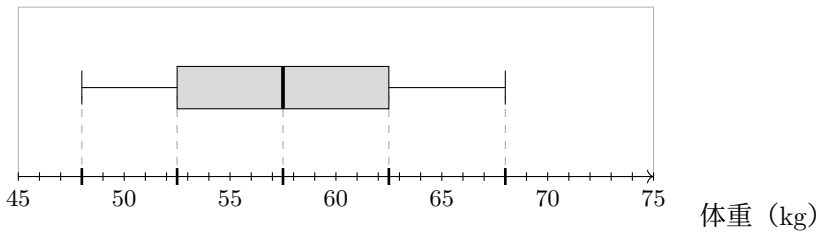
問題 8.4 で求めた値を整理すると,

最小値 = 48, 第 1 四分位数 = 52.5, 中央値 = 57.5, 第 3 四分位数 = 62.5, 最大値 = 68

となります. 箱ひげ図は, これらの 5 つの値を横軸上にとって作成します.

- 横軸上に 48, 52.5, 57.5, 62.5, 68 の位置をとる.
- 第 1 四分位数 52.5 から第 3 四分位数 62.5 までを箱で表す.
- 箱の中の, 第 2 四分位数（中央値）57.5 の位置に縦の線を引く.
- 箱の左端（52.5）から最小値 48 まで, 右端（62.5）から最大値 68 までひげ（線）を伸ばす.

このようにして, 20 人の体重データの箱ひげ図を作成すると, 下記のようになります.



問題 8.6

このデータは 20 点台から 80 点台まであるので, 幹は

2, 3, 4, 5, 6, 7, 8

となります. 各点数の 1 の位を対応する幹の行に葉として並べ, 葉を小さい順に整えます. すべてのデータを並べると, 次のような幹葉表示になります.

幹	葉
2	2 4 8
3	0 2 4 8
4	0 2 4 6 8 8
5	0 2 4 6 6 6 8 8
6	0 0 2 4 4 6 8 8
7	0 2 2 4 6 8
8	0 2 4 6 8

次に, この幹葉表示を使って四分位数を求めます. データは 40 個なので, 小さい順に並べたときの位置は次のようになります.

- **第 2 四分位数（中央値）**
  全体の真ん中にある **20 番目と 21 番目の値の平均**です.
- **第 1 四分位数**
  下半分（1〜20 番目）の真ん中にあたる値です．下半分は 20 個なので，真ん中は **10 番目と 11 番目の値の平均**になります.
- **第 3 四分位数**
  上半分（21〜40 番目）の真ん中にあたる値です．上半分も 20 個なので，真ん中は **30 番目と 31 番目の値の平均**になります.

そこで，幹葉表示の各行に含まれる葉の個数と累計を表に追加すると，次のようになります.

幹	葉	個数	累計
2	2 4 8	3	3
3	0 2 4 8	4	7
4	0 2 4 6 8 8	6	13
5	0 2 4 6 6 6 8 8	8	21
6	0 0 2 4 4 6 8 8	8	29
7	0 2 2 4 6 8	6	35
8	0 2 4 6 8	5	40

四分位数は次のように読み取ります.

- **第 1 四分位数（10・11 番目）**
  10 番目と 11 番目の値は 4 の行に含まれます．3 の行までで 7 個あるので，4 の行では

  $$10-7=3, \qquad 11-7=4$$

  より，左から 3 番目と 4 番目の葉が 10 番目と 11 番目に対応します．値は 44 と 46 なので,

  $$\frac{44+46}{2}=45$$

  となります．これが第 1 四分位数です.
- **第 2 四分位数（20・21 番目）**
  20 番目と 21 番目の値は 5 の行に含まれます．4 の行までで 13 個あるので，5 の行では

  $$20-13=7, \qquad 21-13=8$$

  より，左から 7 番目と 8 番目の葉が 20 番目と 21 番目に対応します．値はいずれも 58 なので,

  $$\frac{58+58}{2}=58$$

  となります．これが第 2 四分位数です.
- **第 3 四分位数（30・31 番目）**
  30 番目と 31 番目の値は 7 の行に含まれます．6 の行までで 29 個あるので，7 の行では

  $$30-29=1, \qquad 31-29=2$$

  より，左から 1 番目と 2 番目の葉が 30 番目と 31 番目に対応します．値は 70 と 72 なので,

  $$\frac{70+72}{2}=71$$

となります．これが第 3 四分位数です．

8.2 の確認テスト

問 1

(1) まず第 2 四分位数，つまり，中央値を求めます．

データは 27 個なので，中央値は 14 番目の値です．14 番目の値は 142 なので，第 2 四分位数は 142 となります．

次に，この中央値の「位置」（14 番目）で，27 個のデータを下半分と上半分に分けます（中央値の値そのものは上下どちらにも含めません）．

- **下半分の 13 個のデータ** は下記です．

$$102,\ 110,\ 115,\ 118,\ 120,\ 122,\ 125,\ 128,\ 130,\ 133,\ 135,\ 138,\ 140$$

この 13 個のデータの中央値（7 番目の値）が第 1 四分位数です．7 番目の値は 125 なので，第 1 四分位数は 125 となります．

- **上半分の 13 個のデータ** は下記です．

$$145,\ 147,\ 150,\ 152,\ 155,\ 158,\ 160,\ 162,\ 165,\ 168,\ 170,\ 175,\ 180$$

この 13 個のデータの中央値（7 番目の値）が第 3 四分位数です．7 番目の値は 160 なので，第 3 四分位数は 160 となります．

(2) 四分位範囲は，第 3 四分位数から第 1 四分位数を引いた値なので，

$$\text{四分位範囲} = \text{第 3 四分位数} - \text{第 1 四分位数} = 160 - 125 = 35$$

となります．

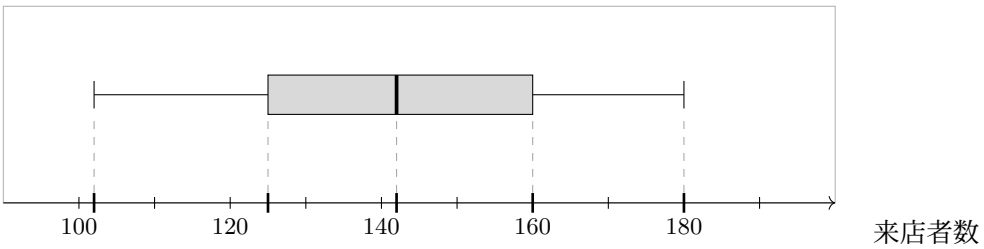
(3)「真ん中の 50%」に当たる部分は，第 1 四分位数から第 3 四分位数までの区間です．第 1 四分位数が 125，第 3 四分位数が 160 なので，真ん中の 50%に入るおおよその範囲は 125 以上 160 以下となります．

(4) このデータでは，

$$\textbf{最小値} = 102,\quad \textbf{第 1 四分位数} = 125,\quad \textbf{中央値} = 142,\quad \textbf{第 3 四分位数} = 160,\quad \textbf{最大値} = 180$$

となります．箱ひげ図は，これらの 5 つの値を横軸上にとって作成します．

- 横軸上に 102，125，142，160，180 の位置をとる．
- 第 1 四分位数 125 から第 3 四分位数 160 までを箱で表す．
- 箱の中の，中央値 142 の位置に縦の線を引く．
- 箱の左端（125）から最小値 102 まで，右端（160）から最大値 180 までひげ（線）を伸ばす．



問 2

(1) 箱ひげ図では，それぞれ次のような意味があります．

- 左のひげの先：最小値
- 箱の左端：第 1 四分位数
- 箱の中の太い線：中央値
- 箱の右端：第 3 四分位数
- 右のひげの先：最大値

図をよく見ると，左のひげの先は 3，箱の左端は 5，箱の中の太い線は 6，箱の右端は 8，右のひげの先は 12 の位置にあります．

したがって，

$$\textbf{最小値} = 3,\quad \textbf{第 1 四分位数} = 5,\quad \textbf{中央値} = 6,\quad \textbf{第 3 四分位数} = 8,\quad \textbf{最大値} = 12$$

となります．

(2)「真ん中の 50%」に当たる部分は，第 1 四分位数から第 3 四分位数までの区間です．この箱ひげ図では，第 1 四分位数が 5，第 3 四分位数が 8 なので，真ん中の 50%に入るおおよその範囲は 5 以上 8 以下となります．

(3) 上位 25%に当たる部分は，第 3 四分位数以上の値です．この箱ひげ図では，第 3 四分位数が 8 なので，上位 25%に入るのは，およそ 8 時間以上の人です．

問 3

(1) **正解：C**

問 2 の箱ひげ図に最も対応しているヒストグラムは，C です．

(2) **正解：a**

箱ひげ図では，四分位数によってデータがおよそ 25%ずつの 4 つの部分に分けられています．この図では，4 つの部分の区間の幅は次のようになっています．

- 下位 25%：3〜5（幅 2）
- 次の 25%：5〜6（幅 1）
- 次の 25%：6〜8（幅 2）
- 上位 25%：8〜12（幅 4）

同じおよそ 25%のデータでも，区間の幅がせまければ，その範囲にデータが比較的集中していることを表します．一方，区間の幅が広ければ，その範囲にデータが広く散らばっていることを表します．

この箱ひげ図では，5〜8 の範囲に真ん中のおよそ 50%のデータが入っている一方で，8〜12 の範囲には上位のおよそ 25%のデータが広く散らばっています．

したがって，5〜8 付近にデータが多く集まり，その右側では 12 まで広く散らばっている分布と考えられます．この特徴に最もよく合うのがヒストグラム C です．

問 4

データは 30 個なので，小さい順に並べたときの 15 番目と 16 番目の値の平均が中央値になります（全体が 30 個なので，真ん中にある 2 つの値の平均が中央値になります）．

そこで，幹葉表示の各行に含まれる葉の個数と累計を表にすると，次のようになります．

幹	葉	個数	累計
2	0 2 5	3	3
3	0 1 3	3	6
4	0 0 2 2 4 5	6	12
5	0 0 2 3 5 8 8	7	19
6	0 0 2 3 3	5	24
7	8 8	2	26
8	0 0 2 4	4	30

15 番目と 16 番目の値は 5 の行に含まれます．4 の行までで 12 個あるので，5 の行では

$$15-12=3, \qquad 16-12=4$$

より，左から 3 番目と 4 番目の葉が 15 番目と 16 番目の値に対応します．値は 52 と 53 なので，中央値は

$$\frac{52+53}{2}=52.5$$

となります．

第 9 章　確率の基礎

9.1　確率の意味

問題 9.1

ここでの試行は「サイコロを 1 回振る」ということです．全事象は $\{1, 2, 3, 4, 5, 6\}$ で，6 通りあります．

(1)「2 の目が出る」という事象は $\{2\}$ で，1 通りです．したがって，その確率は

$$P(\text{「2 の目が出る」}) = \frac{\text{事象の要素数}}{\text{全事象の要素数}} = \frac{1}{6}$$

となります．

(2)「奇数の目が出る」という事象は $\{1,\ 3,\ 5\}$ で，3 通りです．したがって，その確率は

$$P(\text{「奇数の目が出る」}) = \frac{\text{事象の要素数}}{\text{全事象の要素数}} = \frac{3}{6} = \frac{1}{2}$$

となります．

問題 9.2

ここでの試行は「大小 2 つのサイコロを同時に 1 回振る」ということです．

全事象は，下記のように，大きいサイコロの目と小さいサイコロの目の組をすべて並べたものとして表します．

$$\left\{\begin{array}{l}(1,1),\ (1,2),\ (1,3),\ (1,4),\ (1,5),\ (1,6),\\ (2,1),\ (2,2),\ (2,3),\ (2,4),\ (2,5),\ (2,6),\\ (3,1),\ (3,2),\ (3,3),\ (3,4),\ (3,5),\ (3,6),\\ (4,1),\ (4,2),\ (4,3),\ (4,4),\ (4,5),\ (4,6),\\ (5,1),\ (5,2),\ (5,3),\ (5,4),\ (5,5),\ (5,6),\\ (6,1),\ (6,2),\ (6,3),\ (6,4),\ (6,5),\ (6,6)\end{array}\right\}$$

大きいサイコロの目が 1 のときには，小さいサイコロの目は 1〜6 の 6 個あります．同じように，大きいサイコロの目が 2 のときにも 6 個，3 のときにも 6 個，... ... ，6 のときにも 6 個あります．このようにして，6 個が 6 回あるので，

$$6 \times 6 = 36$$

より，36 が全事象の要素数です．

ここで，事象の要素数を数えて確率を求めます．

(1)「2 つのサイコロの目の和が 2 になる」事象は

$$\{(1,1)\}$$

の 1 個です．したがって，その確率は

$$P(\text{「2」}) = \frac{1}{36}$$

となります．

(2)「2 つのサイコロの目の和が 4 になる」事象は

$$\{(1,3),(2,2),(3,1)\}$$

の 3 個です．したがって，その確率は

$$P(\text{「4」}) = \frac{3}{36} = \frac{1}{12}$$

となります．

(3)「2 つのサイコロの目の和が 6 になる」事象は

$$\{(1,5),(2,4),(3,3),(4,2),(5,1)\}$$

の 5 個です．したがって，その確率は

$$P(\text{「6」}) = \frac{5}{36}$$

となります．

(4)「2 つのサイコロの目の和が偶数になる」事象は，和が 2，4，6，8，10，12 になる場合です．

和が 2，4，6 になる場合の要素数は，(1) 〜 (3) で求めたように

$$1,\ 3,\ 5$$

です．

和が 8 になる組は

$$\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$$

の 5 個です．

和が 10 になる組は

$$\{(4,6),(5,5),(6,4)\}$$

の 3 個です．

和が 12 になる組は

$$\{(6,6)\}$$

の 1 個です．

これら 6 つの要素数を合わせると，

$$1+3+5+5+3+1=18$$

となり，求める確率は

$$P(\text{「偶数」}) = \frac{18}{36} = \frac{1}{2}$$

となります.

(5)「2 つのサイコロの目の和が奇数になる」事象は,「偶数になる」事象の余事象です. したがって, その確率は

$$P(\text{「奇数」}) = 1 - P(\text{「偶数」}) = 1 - \frac{1}{2} = \frac{1}{2}$$

となります.

(6)「2 つのサイコロの目の和が 5 以下になる」事象は, 和が 2, 3, 4, 5 になる場合です.

和が 2 と 4 になる場合の要素数は, (1) と (2) で求めたように

$$1,\ 3$$

です.

和が 3 になる組は

$$\{(1,2),(2,1)\}$$

の 2 個です.

和が 5 になる組は

$$\{(1,4),(2,3),(3,2),(4,1)\}$$

の 4 個です.

これら 4 つの要素数を合わせると,

$$1+2+3+4=10$$

となり, 求める確率は

$$P(\text{「5 以下」}) = \frac{10}{36} = \frac{5}{18}$$

となります.

(7)「2 つのサイコロの目の和が 6 以下になる」という事象は,「和が 5 以下になる」事象と「和が 6 になる」事象のどちらかが起こる場合です. これら 2 つの事象は同時には起こらない排反な事象なので, 確率の加法定理より, 求める確率は

$$\begin{aligned} P(\text{「6 以下」}) &= P(\text{「5 以下」 または 「6」}) \\ &= P(\text{「5 以下」}) + P(\text{「6」}) \\ &= \frac{10}{36} + \frac{5}{36} = \frac{5}{12} \end{aligned}$$

となります ($P(\text{「5 以下」})$ と $P(\text{「6」})$ は (6), (3) で求めています).

(8)「2 つのサイコロの目の和が 7 以上になる」事象は,「6 以下になる」事象の余事象です. したがって, その確率は

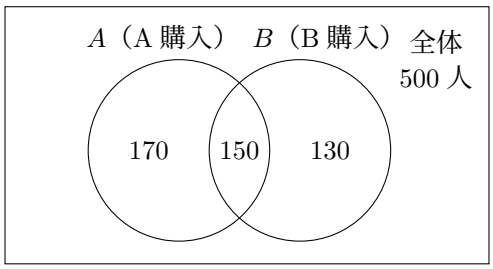
$$\begin{aligned} P(\text{「7 以上」}) &= 1 - P(\text{「6 以下」}) \\ &= 1 - \frac{15}{36} = \frac{7}{12} \end{aligned}$$

となります.

問題 9.3

集合 A を「商品 A を購入した利用者の集合」，集合 B を「商品 B を購入した利用者の集合」とします．

まず，商品 A と商品 B の両方を購入した利用者が 150 人いるので，重なりの部分に 150 を入れます．次に，商品 A を購入した 320 人からこの 150 人を引いて，商品 A だけを購入した利用者は 170 人，商品 B を購入した 280 人から同じ 150 人を引いて，商品 B だけを購入した利用者は 130 人になります．これらをベン図に配置すると，下のようになります．



(1) 商品 A だけを購入した利用者の人数は，商品 A を購入した 320 人から両方購入した 150 人を除いた

$$|A \setminus B| = 320 - 150 = 170$$

です．

(2) 商品 B だけを購入した利用者の人数は，商品 B を購入した 280 人から両方購入した 150 人を除いた

$$|B \setminus A| = 280 - 150 = 130$$

です．

(3)「どちらか一方だけを購入した利用者」は，「A だけ」か「B だけ」のどちらかに当てはまる利用者です．したがって，その人数は

$$170 + 130 = 300$$

です．

(4)「商品 A または商品 B を購入した利用者」は，ベン図でいうと，「A だけ」「両方購入」「B だけ」の 3 つの部分すべてを合わせたものです．したがって，その人数は

$$|A \cup B| = 170 + 150 + 130 = 450$$

です．

(5) 商品 A も商品 B も購入していない利用者の人数は，全体 500 人から「A または B を購入した利用者」450 人を除いた

$$500 - 450 = 50$$

です．

(6) 利用者全体から 1 人を選んだとき，その人が商品 A を購入している確率は，商品 A を購入した 320 人を全体 500 人で割ればよいので，

$$\frac{|A|}{500} = \frac{320}{500} = \frac{16}{25}$$

となります．

(7) 利用者全体から 1 人を選んだとき，その人が商品 A または商品 B を購入している確率は，「A または B を購入した利用者」450 人を全体 500 人で割ればよいので，

$$\frac{|A \cup B|}{500} = \frac{450}{500} = \frac{9}{10}$$

となります．

(8) 利用者全体から 1 人を選んだとき，その人が商品 A と商品 B の両方を購入している確率は，両方購入した 150 人を全体 500 人で割ればよいので，

$$\frac{|A \cap B|}{500} = \frac{150}{500} = \frac{3}{10}$$

となります．

9.1 の確認テスト

問 1

ここでの試行は「1〜12 のカードから 1 枚を無作為に選ぶ」ということです．全事象は

$$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

の 12 通りです．

(1)「3 の倍数のカードを引く」という事象は

$$\{3, 6, 9, 12\}$$

の 4 通りです．したがって，その確率は

$$P(\text{「3 の倍数」}) = \frac{\text{事象の要素数}}{\text{全事象の要素数}} = \frac{4}{12} = \frac{1}{3}$$

となります．

(2)「偶数のカードを引く」という事象は

$$\{2, 4, 6, 8, 10, 12\}$$

の 6 通りです．したがって，その確率は

$$P(\text{「偶数」}) = \frac{6}{12} = \frac{1}{2}$$

となります．

(3)「3 の倍数である偶数のカードを引く」という事象は

$$\{6, 12\}$$

の 2 通りです．したがって，その確率は

$$P(\text{「3 の倍数かつ偶数」}) = \frac{2}{12} = \frac{1}{6}$$

となります．

問 2

ここでの試行は「6 個のボールのなかから同時に 2 個を取り出す」ということです．ボールは，赤が 1 個，青が 2 個，黄が 3 個あります．

青いボールは 2 個あるので，区別のために青 1，青 2 とします．黄色いボールも同様に，黄 1，黄 2，黄 3 とします．

全事象は，2 個のボールの組をすべて挙げたもので，次の 15 通りです．

$$\left\{\begin{array}{l}(\text{赤}, \text{青 } 1),\ (\text{赤}, \text{青 } 2),\ (\text{赤}, \text{黄 } 1),\ (\text{赤}, \text{黄 } 2),\ (\text{赤}, \text{黄 } 3),\\ (\text{青 } 1, \text{青 } 2),\ (\text{青 } 1, \text{黄 } 1),\ (\text{青 } 1, \text{黄 } 2),\ (\text{青 } 1, \text{黄 } 3),\\ (\text{青 } 2, \text{黄 } 1),\ (\text{青 } 2, \text{黄 } 2),\ (\text{青 } 2, \text{黄 } 3),\\ (\text{黄 } 1, \text{黄 } 2),\ (\text{黄 } 1, \text{黄 } 3),\\ (\text{黄 } 2, \text{黄 } 3)\end{array}\right\}$$

したがって，全事象の要素数は 15 通りです．

(1)「取り出したボールがどちらも青である」という事象は

$$\{(\text{青 } 1, \text{青 } 2)\}$$

の 1 通りです．したがって，その確率は

$$P(\text{「どちらも青である」}) = \frac{\text{事象の要素数}}{\text{全事象の要素数}} = \frac{1}{15}$$

となります．

(2)「取り出したボールのうちどちらかは青でない」という事象は，「取り出したボールがどちらも青である」事象の余事象です．したがって，その確率は

$$P(\text{「どちらかは青でない」}) = 1 - P(\text{「どちらも青である」}) = 1 - \frac{1}{15} = \frac{14}{15}$$

となります．

(3)「取り出したボールがどちらも青でない」という事象は

$$\{(\text{赤}, \text{黄 } 1),\ (\text{赤}, \text{黄 } 2),\ (\text{赤}, \text{黄 } 3),\ (\text{黄 } 1, \text{黄 } 2),\ (\text{黄 } 1, \text{黄 } 3),\ (\text{黄 } 2, \text{黄 } 3)\}$$

の 6 通りです．したがって，その確率は

$$P(\text{「どちらも青でない」}) = \frac{\text{事象の要素数}}{\text{全事象の要素数}} = \frac{6}{15} = \frac{2}{5}$$

となります．

(4)「取り出したボールのうちどちらかは青である」という事象は，「取り出したボールがどちらも青でない」事象の余事象です．したがって，その確率は

$$P(\text{「どちらかは青である」}) = 1 - P(\text{「どちらも青でない」}) = 1 - \frac{2}{5} = \frac{3}{5}$$

となります．

問 3

集合 A を「犬を飼っている生徒の集合」，集合 B を「猫を飼っている生徒の集合」とします．

まず，犬を飼っている生徒が 70 人，猫を飼っている生徒が 55 人います．この 2 つの人数をそのまま足すと

$$70 + 55 = 125$$

となりますが，この中には「犬と猫の両方を飼っている生徒」が左の円と右の円の両方で 2 回数えられています．

一方，「犬または猫を飼っている生徒」は 100 人であることがわかっています．したがって，125 人のうち，2 回数えられている人数は

$$125 - 100 = 25$$

となり，犬と猫の両方を飼っている生徒は 25 人です．

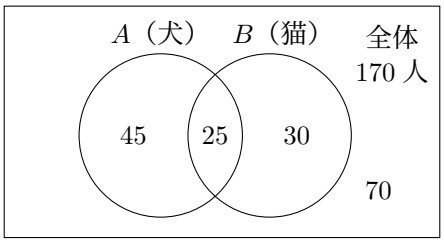
よって，犬だけを飼っている生徒は $70 - 25 = 45$ 人，猫だけを飼っている生徒は $55 - 25 = 30$ 人になります．

また，犬も猫も飼っていない生徒は 70 人います．犬または猫を飼っている生徒は 100 人なので，この学年の生徒は全部で

$$100 + 70 = 170$$

となります．

これらをベン図に配置すると，下のようになります．



(1) 犬と猫の両方を飼っている生徒の人数は

$$|A \cap B| = 25$$

です．

(2) 犬だけを飼っている生徒の人数は，犬を飼っている 70 人から両方を飼っている 25 人を除いた

$$|A \setminus B| = 70 - 25 = 45$$

です．

(3) 猫だけを飼っている生徒の人数は，猫を飼っている 55 人から両方を飼っている 25 人を除いた

$$|B \setminus A| = 55 - 25 = 30$$

です．

(4) この学年の生徒の人数は，犬または猫を飼っている 100 人と，犬も猫も飼っていない 70 人を合わせた

$$100 + 70 = 170$$

です．

（補足）上のように，ベン図で考えると，犬を飼っている人数 $|A|$ と猫を飼っている人数 $|B|$ をそのまま足すと，両方を飼っている部分 $|A \cap B|$ が 2 回数えられてしまいます．

そこで，2 回数えた分を 1 回分引くと，「犬または猫を飼っている生徒」の人数 $|A \cup B|$ が求められます．

一般に，2 つの集合 A, B について次の関係が成り立ちます．

$$|A \cup B| = |A| + |B| - |A \cap B|$$

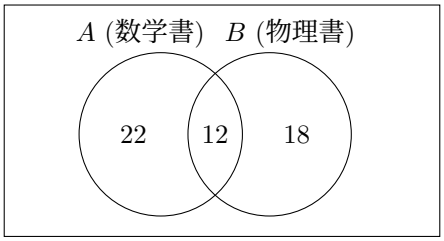
これはベン図での数え方を整理した結果として自然に出てくる関係です．

9.2　条件付き確率

問題 9.4

数学書を借りたことを事象 A，物理書を借りたことを事象 B とします．

まず，状況をベン図で整理します．



(1) 1 件選んだときに，その記録が数学書を借りていることがわかったので，**全体は数学書を借りた 34 件**になります．そのうち物理書も借りている記録は 12 件なので，求める確率は

$$P(B \mid A) = \frac{\text{事象 } A \cap B \text{ の要素数}}{\text{事象 } A \text{ の要素数}} = \frac{12}{34} = \frac{6}{17}$$

となります．

(2) 1 件選んだときに，その記録が物理書を借りていることがわかったので，**全体は物理書を借りた 30 件**になります．そのうち数学書も借りている記録は 12 件なので，求める確率は

$$P(A \mid B) = \frac{\text{事象 } A \cap B \text{ の要素数}}{\text{事象 } B \text{ の要素数}} = \frac{12}{30} = \frac{2}{5}$$

となります．

問題 9.5

スーツケースに fragile のステッカーが貼られていることを事象 A，priority のステッカーが貼られていることを事象 B とします．

ここでは

$$P(A),\ P(B),\ P(A \cap B)$$

といった確率そのものが与えられています．したがって，条件付き確率は次の公式を用いて求めます．

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}, \qquad P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

与えられている確率は

$$P(A) = \frac{1}{5}, \qquad P(B) = \frac{1}{10}, \qquad P(A \cap B) = \frac{1}{25}$$

です．これをもとに計算していきます．

(1) 選んだスーツケースに fragile のステッカーが貼られていること（事象 A）がわかったので，求める確率は

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{25}}{\frac{1}{5}} = \frac{1}{5}$$

となります．

(2) 選んだスーツケースに priority のステッカーが貼られていること（事象 B）がわかったので，求める確率は

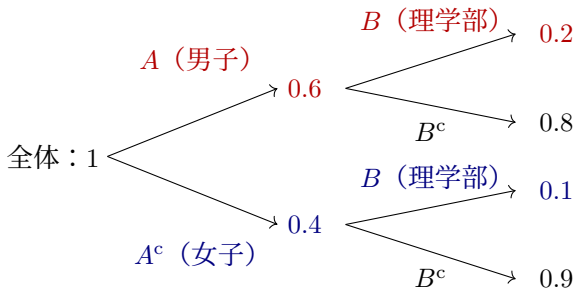
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{25}}{\frac{1}{10}} = \frac{2}{5}$$

となります．

問題 9.6

男子学生であることを事象 A，理学部に所属していることを事象 B とします．

次に，割合を樹形図で整理します．



(1) 理学部に所属している学生は，男子で理学部に所属している学生と，女子で理学部に所属している学生の 2 つに分けられます．

- 男子学生は全体の 0.6 です．そのうち理学部に所属しているのは 0.2 なので，

$$P(A \cap B) = P(A) \times P(B \mid A) = 0.6 \times 0.2 = 0.12$$

- 女子学生は全体の 0.4 です．そのうち理学部に所属しているのは 0.1 なので，

$$P(A^{\mathrm{c}} \cap B) = P(A^{\mathrm{c}}) \times P(B \mid A^{\mathrm{c}}) = 0.4 \times 0.1 = 0.04$$

ここで，「理学部に所属している」という事象は，「男子で理学部」事象と「女子で理学部」事象のどちらかが起こる場合であり，これら 2 つの事象は排反です．

よって，「この大学の 1 年生から 1 人選んだとき，その人が理学部に所属している確率」は，加法定理より

$$P(B) = P(A \cap B) + P(A^{\mathrm{c}} \cap B) = 0.12 + 0.04 = 0.16$$

となります．

(2) (1) より，

$$P(A \cap B) = 0.12, \qquad P(B) = 0.16$$

がわかっています．

したがって，「理学部に所属している学生のなかから 1 人選んだとき，その人が男子学生である」確率は

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.12}{0.16} = \frac{3}{4}$$

となります．

問題 9.7

大小 2 つのサイコロの出る目の組は全部で 36 通りあります（例題 9.2 参照）．以下では，各事象が起こる結果の数を数えて確率を求め，$P(A \cap B) = P(A)P(B)$ などが成り立つかどうかを調べます．

(1) $P(A \cap B)$ と $P(A)P(B)$ を比較します．

- **$P(A)$ を求めます．**

  大きいサイコロの目が 4 以下になるのは

  $$\{1, 2, 3, 4\}$$

  の 4 通りで，それぞれに対して小さいサイコロの目は 1〜6 の 6 通りあります．したがって，

  $$P(A) = \frac{4 \times 6}{36} = \frac{24}{36} = \frac{2}{3}$$

- **$P(B)$ を求めます．**

  小さいサイコロの目が奇数になるのは

  $$\{1, 3, 5\}$$

  の 3 通りで，それぞれに対して大きいサイコロの目は 1〜6 の 6 通りあります．したがって，

  $$P(B) = \frac{3 \times 6}{36} = \frac{18}{36} = \frac{1}{2}$$

- **$P(A \cap B)$ を求めます．**

  「大きいサイコロの目が 4 以下」かつ「小さいサイコロの目が奇数」になる組は

  $$\{(1,1), (1,3), (1,5), (2,1), (2,3), (2,5), (3,1), (3,3), (3,5), (4,1), (4,3), (4,5)\}$$

  の 12 個です．したがって，

  $$P(A \cap B) = \frac{12}{36} = \frac{1}{3}$$

- **$P(A)P(B)$ を求めます．**

  $$P(A)P(B) = \frac{2}{3} \times \frac{1}{2} = \frac{1}{3}$$

以上より，

$$P(A \cap B) = P(A)P(B)$$

となるので，**A と B は独立です**.

(2) $P(C \cap D)$ と $P(C)P(D)$ を比較します.

- **$P(C)$ を求めます.**

  大きいサイコロの目が 5 または 6 になるのは

  $$\{5, 6\}$$

  の 2 通りで，それぞれに対して小さいサイコロの目は 1〜6 の 6 通りあります．したがって，

  $$P(C) = \frac{2 \times 6}{36} = \frac{12}{36} = \frac{1}{3}$$

- **$P(D)$ を求めます.**

  和が 9 以上になる組は

  $$\{(3, 6), (4, 5), (4, 6), (5, 4), (5, 5), (5, 6), (6, 3), (6, 4), (6, 5), (6, 6)\}$$

  の 10 個です．したがって，

  $$P(D) = \frac{10}{36} = \frac{5}{18}$$

- **$P(C \cap D)$ を求めます.**

  「大きいサイコロが 5 または 6」かつ「和が 9 以上」になる組は

  $$\{(5, 4), (5, 5), (5, 6), (6, 3), (6, 4), (6, 5), (6, 6)\}$$

  の 7 個です．したがって，

  $$P(C \cap D) = \frac{7}{36}$$

- **$P(C)P(D)$ を求めます.**

  $$P(C)P(D) = \frac{1}{3} \times \frac{5}{18} = \frac{5}{54}$$

以上より，

$$P(C \cap D) \neq P(C)P(D)$$

となるので，**C と D は独立ではありません**.

(3) $P(E \cap F)$ と $P(E)P(F)$ を比較します.

- **$P(E)$ を求めます.**

  大きいサイコロの目が 1,3,5 のいずれかであるのは

  $$\{1, 3, 5\}$$

  の 3 通りで，それぞれに対して小さいサイコロの目は 1〜6 の 6 通りあります．したがって，

  $$P(E) = \frac{3 \times 6}{36} = \frac{1}{2}$$

- **$P(F)$ を求めます.**

  積が 12 の倍数になる組は

  $$\{(2, 6), (3, 4), (4, 3), (4, 6), (6, 2), (6, 4), (6, 6)\}$$

の 7 個です．したがって，

$$P(F) = \frac{7}{36}$$

- **$P(E \cap F)$ を求めます．**

「大きいサイコロの目が 1,3,5」かつ「積が 12 の倍数」になる組は

$$\{(3,4)\}$$

の 1 個です．したがって，

$$P(E \cap F) = \frac{1}{36}$$

- **$P(E)P(F)$ を求めます．**

$$P(E)P(F) = \frac{1}{2} \times \frac{7}{36} = \frac{7}{72}$$

以上より，

$$P(E \cap F) \neq P(E)P(F)$$

となるので，**E と F は独立ではありません．**

9.2 の確認テスト

問 1

統計学を履修していることを事象 A，プログラミングを履修していることを事象 B とします．

まず，統計学を履修している学生が 70 人，プログラミングを履修している学生が 65 人います．この 2 つの人数をそのまま足すと

$$70 + 65 = 135$$

となりますが，この中には「統計学とプログラミングの両方を履修している学生」が 2 回数えられています．

一方，「統計学またはプログラミングを履修している学生」は 95 人であることがわかっています．したがって，135 人のうち，2 回数えられている人数は

$$135 - 95 = 40$$

となり，統計学とプログラミングの両方を履修している学生は 40 人であることがわかります．

(1) 統計学とプログラミングの両方を履修している学生の人数は

$$|A \cap B| = 40$$

です．

(2) 統計学を履修している学生のなかから 1 人選ぶので，**全体は統計学を履修している 70 人** になります．

そのうちプログラミングも履修している学生は 40 人なので，求める確率は

$$P(B \mid A) = \frac{|A \cap B|}{|A|} = \frac{40}{70} = \frac{4}{7}$$

となります.

(3) プログラミングを履修している学生のなかから 1 人選ぶので，**全体はプログラミングを履修している 65 人** になります.

そのうち統計学も履修している学生は 40 人なので，求める確率は

$$P(A \mid B) = \frac{|A \cap B|}{|B|} = \frac{40}{65} = \frac{8}{13}$$

となります.

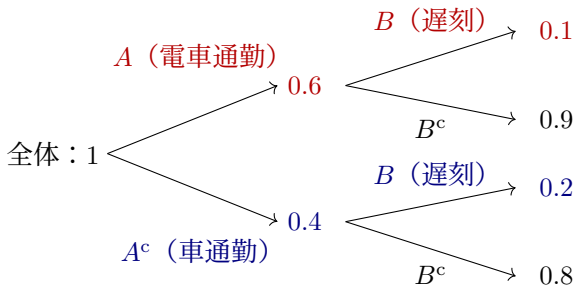
問 2

電車通勤であることを事象 A，過去 1 年で遅刻したことを事象 B とします.

電車通勤と車通勤の比は「6：4」であるので，電車通勤の社員は全体の 0.6，車通勤の社員は全体の 0.4 です.

また，電車通勤の社員が過去 1 年で遅刻した割合は 0.1，車通勤の社員が過去 1 年で遅刻した割合は 0.2 です.

ここで，割合を樹形図で整理します.



(1) 遅刻した社員は，電車通勤で遅刻した社員と，車通勤で遅刻した社員の 2 つに分けられます.

- 電車通勤の社員は全体の 0.6 です．そのうち遅刻するのは 0.1 なので，

$$P(A \cap B) = 0.6 \times 0.1 = 0.06$$

- 車通勤の社員は全体の 0.4 です．そのうち遅刻するのは 0.2 なので，

$$P(A^{\mathrm{c}} \cap B) = 0.4 \times 0.2 = 0.08$$

ここで，「遅刻した」という事象 B は，「電車通勤で遅刻した」事象と「車通勤で遅刻した」事象のどちらかが起こる場合であり，これら 2 つの事象は排反です.

よって，「この会社から 1 人選んだとき，その人が過去 1 年で遅刻した」確率は，加法定理より

$$P(B) = P(A \cap B) + P(A^{\mathrm{c}} \cap B) = 0.06 + 0.08 = 0.14$$

となります.

(2) (1) より，

$$P(A^{\mathrm{c}} \cap B) = 0.08, \qquad P(B) = 0.14$$

がわかっています.

したがって，「遅刻した社員のなかから 1 人選んだとき，その人が車通勤である」確率は

$$P(A^{\mathrm{c}} \mid B) = \frac{P(A^{\mathrm{c}} \cap B)}{P(B)} = \frac{0.08}{0.14} = \frac{4}{7}$$

となります．

(3) 遅刻した社員のなかでの電車通勤の社員の割合は「遅刻した社員のなかから 1 人選んだとき，その人が電車通勤である」確率

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.06}{0.14} = \frac{3}{7}$$

となります．

同様に，遅刻した社員のなかでの車通勤の社員の割合は「遅刻した社員のなかから 1 人選んだとき，その人が車通勤である」確率です．これは (2) より 4/7 であることがわかっています．

ここで，

$$\frac{3}{7} < \frac{4}{7}$$

なので，**遅刻した社員のなかでは車通勤の社員のほうが多い** といえるということになります．

問 3

2 枚のコイン（コイン 1 とコイン 2）を同時に 1 回投げるときの出る面の組は

(表, 表), (表, 裏), (裏, 表), (裏, 裏)

の 4 通りあります．以下では，各事象が起こる結果の数を数えて確率を求め，$P(A \cap B) = P(A)P(B)$ などが成り立つかどうかを調べます．

(1) $P(A \cap B)$ と $P(A)P(B)$ を比較します．

- **$P(A)$ を求めます．**

  事象 A（コイン 1 が表である）が起こるのは

  $\{(表, 表), (表, 裏)\}$

  の 2 通りです．したがって，

  $$P(A) = \frac{2}{4} = \frac{1}{2}$$

- **$P(B)$ を求めます．**

  事象 B（2 枚のコインが同じ面である）が起こるのは

  $\{(表, 表), (裏, 裏)\}$

  の 2 通りです．したがって，

  $$P(B) = \frac{2}{4} = \frac{1}{2}$$

- **$P(A \cap B)$ を求めます．**

  「コイン 1 が表である」かつ「同じ面である」となるのは

  $\{(表, 表)\}$

の 1 通りです．したがって，

$$P(A \cap B) = \frac{1}{4}$$

- **$P(A)P(B)$ を求めます．**

$$P(A)P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

以上より，

$$P(A \cap B) = P(A)P(B)$$

となるので，**A と B は独立です**．

(2) $P(A \cap C)$ と $P(A)P(C)$ を比較します．

- **$P(C)$ を求めます．**

事象 C（表が 1 枚だけである）が起こるのは

$$\{(\text{表}, \text{裏}), (\text{裏}, \text{表})\}$$

の 2 通りです．したがって，

$$P(C) = \frac{2}{4} = \frac{1}{2}$$

- **$P(A \cap C)$ を求めます．**

「コイン 1 が表である」かつ「表が 1 枚だけである」となるのは

$$\{(\text{表}, \text{裏})\}$$

の 1 通りです．したがって，

$$P(A \cap C) = \frac{1}{4}$$

- **$P(A)P(C)$ を求めます．**

$$P(A)P(C) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

以上より，

$$P(A \cap C) = P(A)P(C)$$

となるので，**A と C は独立です**．

(3) $P(B \cap C)$ と $P(B)P(C)$ を比較します．

- **$P(B \cap C)$ を求めます．**

「同じ面である」かつ「表が 1 枚だけである」となるのは 0 通りです（同じ面なら (表, 表) または (裏, 裏) であり，表は 0 枚か 2 枚になるため）．したがって，

$$P(B \cap C) = \frac{0}{4} = 0$$

- **$P(B)P(C)$ を求めます．**

$$P(B)P(C) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

以上より，

$$P(B \cap C) \neq P(B)P(C)$$

となるので，**B と C は独立ではありません．**

(4) $P(A \cap D)$ と $P(A)P(D)$ を比較します．

- **$P(D)$ を求めます．**

  事象 D（表が 2 枚である）が起こるのは

  $$\{(\text{表}, \text{表})\}$$

  の 1 通りです．したがって，

  $$P(D) = \frac{1}{4}$$

- **$P(A \cap D)$ を求めます．**

  「コイン 1 が表である」かつ「表が 2 枚である」となるのは

  $$\{(\text{表}, \text{表})\}$$

  の 1 通りです．したがって，

  $$P(A \cap D) = \frac{1}{4}$$

- **$P(A)P(D)$ を求めます．**

  $$P(A)P(D) = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$$

以上より，

$$P(A \cap D) \neq P(A)P(D)$$

となるので，**A と D は独立ではありません．**

問 4

箱の中には，当たりくじ 3 本，はずれくじ 17 本が入っており，全部で 20 本です．

(1) 引いたくじを元に戻すという試行を考えます．1 回目に当たりくじを引く事象を A_1，2 回目に当たりくじを引く事象を B_1 とします．

1 回目に当たりくじを引く確率は

$$P(A_1) = \frac{3}{20}$$

となります．

引いたくじを元に戻すので，2 回目の条件は 1 回目の結果によって変わりません．したがって，

$$P(B_1 \mid A_1) = P(B_1)$$

が成り立ちます．

2 回目も 20 本のなかから引くので，

$$P(B_1) = \frac{3}{20}$$

となります.

よって,「2 回とも当たりくじである」確率は

$$P(A_1 \cap B_1) = P(A_1) \times P(B_1) = \frac{3}{20} \times \frac{3}{20} = \frac{9}{400}$$

となります.

(2) 引いたくじを元に戻さないという試行を考えます. 1 回目に当たりくじを引く事象を A_2, 2 回目に当たりくじを引く事象を B_2 とします.

1 回目に当たりくじを引く確率は

$$P(A_2) = \frac{3}{20}$$

となります.

引いたくじを元に戻さないので, 2 回目の条件は 1 回目の結果によって変わります.

1 回目に当たりくじを引いたとき, 残りは当たりくじ 2 本, はずれくじ 17 本の計 19 本になるので,

$$P(B_2 \mid A_2) = \frac{2}{19}$$

となります.

したがって,「2 回とも当たりくじである」確率は

$$P(A_2 \cap B_2) = P(A_2) \times P(B_2 \mid A_2) = \frac{3}{20} \times \frac{2}{19} = \frac{3}{190}$$

となります.

(3) 引いたくじを元に戻さないという試行を考えます. 1 回目に当たりくじを引く事象を A_2, 2 回目に当たりくじを引く事象を B_2 とします.

1 回目にはずれくじを引く確率は

$$P(A_2^{\mathrm{c}}) = \frac{17}{20}$$

となります.

1 回目にはずれくじを引いたとき, 残りは当たりくじ 3 本, はずれくじ 16 本の計 19 本なので,

$$P(B_2 \mid A_2^{\mathrm{c}}) = \frac{3}{19}$$

となります.

よって,「1 回目にはずれくじを引き, 2 回目に当たりくじを引く」確率は

$$P(A_2^{\mathrm{c}} \cap B_2) = P(A_2^{\mathrm{c}}) \times P(B_2 \mid A_2^{\mathrm{c}}) = \frac{17}{20} \times \frac{3}{19} = \frac{51}{380}$$

となります.

ここで,「2 回目に当たりくじを引く」という事象 B_2 は,「1 回目が当たりで 2 回目も当たり」である事象 $A_2 \cap B_2$ と,「1 回目がはずれで 2 回目が当たり」である事象 $A_2^{\mathrm{c}} \cap B_2$ のどちらかが起こる場合です. 1 回目は「当たり」か「はずれ」のどちらか一方であり, 両方同時には起こらないので, これら 2 つの事象は排反です.

したがって,「2 回目に当たりくじを引く」確率は, 加法定理より

$$P(B_2) = P(A_2 \cap B_2) + P(A_2^{\mathrm{c}} \cap B_2) = \frac{3}{190} + \frac{51}{380} = \frac{57}{380} = \frac{3}{20}$$

となります.

第 10 章　くり返し操作と確率

10.1　くり返し操作における確率

問題 10.1

「1 個目が不良品である」という事象を A_1,「2 個目が不良品である」という事象を A_2,「3 個目が不良品である」という事象を A_3 とします.

製品 1 個が不良品である確率は一定で 1/10 なので,

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{10}$$

です.

その余事象 $A_1^{\mathrm{c}}, A_2^{\mathrm{c}}, A_3^{\mathrm{c}}$ は，それぞれ「1 個目が不良品でない」という事象,「2 個目が不良品でない」という事象,「3 個目が不良品でない」という事象であり,

$$P(A_1^{\mathrm{c}}) = P(A_2^{\mathrm{c}}) = P(A_3^{\mathrm{c}}) = \frac{9}{10}$$

です.

(1)「3 個とも不良品である」という事象は，3 つの事象 A_1, A_2, A_3 がすべて起こることです．各製品が不良品であるかどうかは互いに独立なので，これら 3 つの事象は独立です.

したがって,「3 個とも不良品である」確率は,

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) = \frac{1}{10} \times \frac{1}{10} \times \frac{1}{10} = \frac{1}{1000}$$

となります.

(2)「少なくとも 1 個は不良品が出る」という事象の余事象は,「1 個も不良品が出ない」という事象です.

「1 個も不良品が出ない」という事象は，3 つの事象 $A_1^{\mathrm{c}}, A_2^{\mathrm{c}}, A_3^{\mathrm{c}}$ がすべて起こることです．各製品が不良品であるかどうかは互いに独立なので，これら 3 つの事象は独立です.

したがって,「1 個も不良品が出ない」確率は,

$$P(A_1^{\mathrm{c}} \cap A_2^{\mathrm{c}} \cap A_3^{\mathrm{c}}) = P(A_1^{\mathrm{c}})P(A_2^{\mathrm{c}})P(A_3^{\mathrm{c}}) = \left(\frac{9}{10}\right)^3 = \frac{729}{1000}$$

となります．したがって,

$$P(\text{少なくとも 1 個は不良品が出る}) = 1 - P(\text{1 個も不良品が出ない}) = 1 - \frac{729}{1000} = \frac{271}{1000}$$

となります.

問題 10.2

「1 個目が不良品である」という事象を A_1,「2 個目が不良品である」という事象を A_2,「3 個目が不良品である」という事象を A_3 とします.

(1)「不良品が 1 個だけ出る」という事象は，次の 3 つの排反な場合に分けられます．

$$A_1 \cap A_2^c \cap A_3^c, \qquad A_1^c \cap A_2 \cap A_3^c, \qquad A_1^c \cap A_2^c \cap A_3$$

各製品が不良品であるかどうかは互いに独立なので，A_1, A_2^c, A_3^c は互いに独立です．したがって，

$$P(A_1 \cap A_2^c \cap A_3^c) = P(A_1)P(A_2^c)P(A_3^c) = \frac{1}{10} \times \frac{9}{10} \times \frac{9}{10} = \frac{81}{1000}$$

となります．また同様に，

$$P(A_1^c \cap A_2 \cap A_3^c) = P(A_1^c)P(A_2)P(A_3^c) = \frac{9}{10} \times \frac{1}{10} \times \frac{9}{10} = \frac{81}{1000},$$
$$P(A_1^c \cap A_2^c \cap A_3) = P(A_1^c)P(A_2^c)P(A_3) = \frac{9}{10} \times \frac{9}{10} \times \frac{1}{10} = \frac{81}{1000}$$

となります．

これら 3 つの事象は互いに排反であり，いずれか 1 つが起これば「不良品が 1 個だけ出る」ことになります．したがって，加法定理より

$$\begin{aligned} P(\text{不良品が 1 個だけ出る}) &= P(A_1 \cap A_2^c \cap A_3^c) + P(A_1^c \cap A_2 \cap A_3^c) + P(A_1^c \cap A_2^c \cap A_3) \\ &= \frac{81}{1000} + \frac{81}{1000} + \frac{81}{1000} \\ &= \frac{243}{1000} \end{aligned}$$

となります．

(2)「不良品が 2 個だけ出る」という事象は，次の 3 つの排反な場合に分けられます．

$$A_1 \cap A_2 \cap A_3^c, \qquad A_1 \cap A_2^c \cap A_3, \qquad A_1^c \cap A_2 \cap A_3$$

各製品が不良品であるかどうかは互いに独立なので，A_1, A_2, A_3^c は互いに独立です．したがって，

$$P(A_1 \cap A_2 \cap A_3^c) = P(A_1)P(A_2)P(A_3^c) = \frac{1}{10} \times \frac{1}{10} \times \frac{9}{10} = \frac{9}{1000}$$

となります．

同様に，

$$P(A_1 \cap A_2^c \cap A_3) = P(A_1)P(A_2^c)P(A_3) = \frac{1}{10} \times \frac{9}{10} \times \frac{1}{10} = \frac{9}{1000},$$
$$P(A_1^c \cap A_2 \cap A_3) = P(A_1^c)P(A_2)P(A_3) = \frac{9}{10} \times \frac{1}{10} \times \frac{1}{10} = \frac{9}{1000}$$

となります．

これら 3 つの事象は互いに排反であり，いずれか 1 つが起これば「不良品が 2 個だけ出る」ことになります．したがって，加法定理より

$$\begin{aligned} P(\text{不良品が 2 個だけ出る}) &= P(A_1 \cap A_2 \cap A_3^c) + P(A_1 \cap A_2^c \cap A_3) + P(A_1^c \cap A_2 \cap A_3) \\ &= \frac{9}{1000} + \frac{9}{1000} + \frac{9}{1000} \\ &= \frac{27}{1000} \end{aligned}$$

となります．

(3)「不良品が 2 個以上出る」という事象は，「不良品が 2 個だけ出る」という事象と「不良品が 3 個出る」という事象のいずれかが起こることです．これら 2 つの事象は互いに排反です．

ここで,「不良品が 2 個だけ出る」確率は (2) より 27/1000,「不良品が 3 個出る」確率は問題 10.1(1) より 1/1000 です.

したがって，加法定理より

$$P(\text{不良品が 2 個以上出る}) = P(\text{不良品が 2 個だけ出る}) + P(\text{不良品が 3 個出る}) = \frac{27}{1000} + \frac{1}{1000} = \frac{28}{1000}$$

となります.

問題 10.3

1 回目に当たりを引くという事象を A_1，2 回目に当たりを引くという事象を A_2，3 回目に当たりを引くという事象を A_3 とします.

このとき，1 回目にはずれを引く事象は A_1^{c}，2 回目にはずれを引く事象は A_2^{c}，3 回目にはずれを引く事象は A_3^{c} となります.

(1) 1 回目に当たりを引いたとき，残りは当たりが 2 本，はずれが 7 本となるので，このとき 2 回目にも当たりを引く確率は

$$P(A_2 \mid A_1) = \frac{2}{9}$$

となります.

(2) 1 回目に当たりを引く確率は

$$P(A_1) = \frac{3}{10}$$

です.

1 回目に当たりを引いたとき，2 回目にも当たりを引く確率は（(1) より）

$$P(A_2 \mid A_1) = \frac{2}{9}$$

となります.

1 回目と 2 回目に当たりを引いたとき，残りは当たりが 1 本，はずれが 7 本となるので，このとき 3 回目に当たりを引く確率は

$$P(A_3 \mid A_1 \cap A_2) = \frac{1}{8}$$

となります.

したがって，3 回とも当たりを引く確率は

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)\,P(A_2 \mid A_1)\,P(A_3 \mid A_1 \cap A_2) = \frac{3}{10} \times \frac{2}{9} \times \frac{1}{8} = \frac{1}{120}$$

となります.

(3)「少なくとも 1 本は当たりである」事象の余事象は，「3 本ともはずれである」事象 $A_1^{\mathrm{c}} \cap A_2^{\mathrm{c}} \cap A_3^{\mathrm{c}}$ です.

まず，この確率を求めます.

1 回目にはずれを引く確率は

$$P(A_1^{\mathrm{c}}) = \frac{7}{10}$$

です.

1 回目にはずれを引いたとき，残りははずれ 6 本，当たり 3 本なので，このとき 2 回目にはずれを引く確率は

$$P(A_2^{\mathrm{c}} \mid A_1^{\mathrm{c}}) = \frac{6}{9} = \frac{2}{3}$$

となります.

1 回目と 2 回目にはずれを引いたとき，残りははずれ 5 本，当たり 3 本なので，このとき 3 回目にはずれを引く確率は

$$P(A_3^{\mathrm{c}} \mid A_1^{\mathrm{c}} \cap A_2^{\mathrm{c}}) = \frac{5}{8}$$

となります.

したがって，3 回ともはずれを引く確率は

$$P(A_1^{\mathrm{c}} \cap A_2^{\mathrm{c}} \cap A_3^{\mathrm{c}}) = P(A_1^{\mathrm{c}})\, P(A_2^{\mathrm{c}} \mid A_1^{\mathrm{c}})\, P(A_3^{\mathrm{c}} \mid A_1^{\mathrm{c}} \cap A_2^{\mathrm{c}}) = \frac{7}{10} \times \frac{2}{3} \times \frac{5}{8} = \frac{7}{24}$$

となります.

よって，引いた 3 本のうち，少なくとも 1 本は当たりである確率は

$$P(\text{少なくとも 1 本当たり}) = 1 - P(\text{3 回ともはずれ}) = 1 - P(A_1^{\mathrm{c}} \cap A_2^{\mathrm{c}} \cap A_3^{\mathrm{c}}) = 1 - \frac{7}{24} = \frac{17}{24}$$

となります.

10.1 の確認テスト

問 1

「1 回目に成功する」という事象を A_1，「2 回目に成功する」という事象を A_2，「3 回目に成功する」という事象を A_3 とします.

1 回の挑戦で成功する確率は一定で 1/5 なので，

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{5}$$

です.

その余事象 $A_1^{\mathrm{c}}, A_2^{\mathrm{c}}, A_3^{\mathrm{c}}$ は，それぞれ「その回に失敗する」という事象であり，

$$P(A_1^{\mathrm{c}}) = P(A_2^{\mathrm{c}}) = P(A_3^{\mathrm{c}}) = \frac{4}{5}$$

です.

挑戦を行うたびの結果は互いに影響しないので，これらの事象はすべて独立です.

(1)「3 回とも成功する」という事象は A_1, A_2, A_3 がすべて起こることです.

したがって，その確率は

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) = \left(\frac{1}{5}\right)^3 = \frac{1}{125}$$

となります.

(2) 成功が 2 回以上続けて起こるのは，次の 3 つの排反な場合です.

$$A_1 \cap A_2 \cap A_3^{\mathrm{c}}, \qquad A_1^{\mathrm{c}} \cap A_2 \cap A_3, \qquad A_1 \cap A_2 \cap A_3$$

まず,

$$P(A_1 \cap A_2 \cap A_3^{\mathrm{c}}) = P(A_1)P(A_2)P(A_3^{\mathrm{c}}) = \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right) = \frac{4}{125}$$

となります．同様に,

$$P(A_1^{\mathrm{c}} \cap A_2 \cap A_3) = P(A_1^{\mathrm{c}})P(A_2)P(A_3) = \left(\frac{4}{5}\right) \left(\frac{1}{5}\right)^2 = \frac{4}{125}$$

となります．また,

$$P(A_1 \cap A_2 \cap A_3) = \left(\frac{1}{5}\right)^3 = \frac{1}{125}$$

となります．これらは排反なので，加法定理より

$$P(2 \text{ 回以上連続成功}) = \frac{4}{125} + \frac{4}{125} + \frac{1}{125} = \frac{9}{125}$$

となります.

(3)「最初に成功するのが 2 回目」というのは,

$$A_1^{\mathrm{c}} \cap A_2$$

が起こることです.

したがって，その確率は

$$P(A_1^{\mathrm{c}} \cap A_2) = P(A_1^{\mathrm{c}})P(A_2) = \frac{4}{5} \times \frac{1}{5} = \frac{4}{25}$$

となります.

問 2

1 回目に赤い玉が出るという事象を A_1，2 回目に赤い玉が出るという事象を A_2，3 回目に赤い玉が出るという事象を A_3，4 回目に赤い玉が出るという事象を A_4 とします.

(1) 1 回目に赤い玉が出る確率は

$$P(A_1) = \frac{6}{10}$$

です.

1 回目に赤い玉が出たとき，袋の中身は赤い玉 5 個，白い玉 4 個になるので，このとき 2 回目に赤い玉が出る確率は

$$P(A_2 \mid A_1) = \frac{5}{9}$$

となります.

したがって，1 回目と 2 回目に赤い玉が出る確率は

$$P(A_1 \cap A_2) = P(A_1)\, P(A_2 \mid A_1) = \frac{6}{10} \times \frac{5}{9} = \frac{1}{3}$$

となります.

(2)「少なくとも 1 回は白い玉が出る」事象の余事象は,「4 回とも赤い玉が出る」事象です.

1 回目に赤い玉が出る確率は

$$P(A_1) = \frac{6}{10}$$

です.

1 回目に赤い玉が出たとき，2 回目にも赤い玉が出る確率は（(1) より）

$$P(A_2 \mid A_1) = \frac{5}{9}$$

です.

1 回目と 2 回目に赤い玉が出たとき，残りは赤い玉が 4 個，白い玉が 4 個となるので，このとき 3 回目に赤い玉が出る確率は

$$P(A_3 \mid A_1 \cap A_2) = \frac{4}{8}$$

となります.

1 回目から 3 回目まで赤い玉が出たとき，残りは赤い玉が 3 個，白い玉が 4 個となるので，このとき 4 回目に赤い玉が出る確率は

$$P(A_4 \mid A_1 \cap A_2 \cap A_3) = \frac{3}{7}$$

となります.

したがって，4 回とも赤い玉が出る確率は

$$P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1)\,P(A_2 \mid A_1)\,P(A_3 \mid A_1 \cap A_2)\,P(A_4 \mid A_1 \cap A_2 \cap A_3) = \frac{6}{10} \times \frac{5}{9} \times \frac{4}{8} \times \frac{3}{7} = \frac{1}{14}$$

となります.

したがって，少なくとも 1 回は白い玉が出る確率は

$$\begin{aligned} P(\text{少なくとも 1 回白}) &= 1 - P(\text{4 回とも赤}) \\ &= 1 - P(A_1 \cap A_2 \cap A_3 \cap A_4) \\ &= 1 - \frac{1}{14} \\ &= \frac{13}{14} \end{aligned}$$

となります.

(3)「赤い玉がちょうど 1 回だけ出る」事象は，次の 4 つの排反な場合に分けられます.

$$A_1 \cap A_2^{\mathrm{c}} \cap A_3^{\mathrm{c}} \cap A_4^{\mathrm{c}}, \quad A_1^{\mathrm{c}} \cap A_2 \cap A_3^{\mathrm{c}} \cap A_4^{\mathrm{c}}, \quad A_1^{\mathrm{c}} \cap A_2^{\mathrm{c}} \cap A_3 \cap A_4^{\mathrm{c}}, \quad A_1^{\mathrm{c}} \cap A_2^{\mathrm{c}} \cap A_3^{\mathrm{c}} \cap A_4$$

まず,「1 回目だけ赤」の場合を考えます.

1 回目に赤い玉が出る確率は

$$P(A_1) = \frac{6}{10}$$

です.

1 回目に赤い玉が出たとき，残りは赤 5 個，白 4 個となるので，このとき 2 回目に白い玉が出る確率は

$$P(A_2^{\mathrm{c}} \mid A_1) = \frac{4}{9}$$

となります.

1 回目に赤，2 回目に白が出たとき，残りは赤 5 個，白 3 個となるので，このとき 3 回目に白い玉が出る確率は

$$P(A_3^{\mathrm{c}} \mid A_1 \cap A_2^{\mathrm{c}}) = \frac{3}{8}$$

となります.

1 回目に赤，2 回目と 3 回目に白が出たとき，残りは赤 5 個，白 2 個となるので，このとき 4 回目に白い玉が出る確率は

$$P(A_4^{\mathrm{c}} \mid A_1 \cap A_2^{\mathrm{c}} \cap A_3^{\mathrm{c}}) = \frac{2}{7}$$

となります.

したがって，「1 回目だけ赤」の確率は

$$P(A_1 \cap A_2^{\mathrm{c}} \cap A_3^{\mathrm{c}} \cap A_4^{\mathrm{c}}) = P(A_1)\, P(A_2^{\mathrm{c}} \mid A_1)\, P(A_3^{\mathrm{c}} \mid A_1 \cap A_2^{\mathrm{c}})\, P(A_4^{\mathrm{c}} \mid A_1 \cap A_2^{\mathrm{c}} \cap A_3^{\mathrm{c}}) = \frac{6}{10} \times \frac{4}{9} \times \frac{3}{8} \times \frac{2}{7} = \frac{1}{35}$$

となります.

同様に，「2 回目だけ赤」「3 回目だけ赤」「4 回目だけ赤」の場合も，その確率はいずれも 1/35 となります.

これら 4 つの事象は互いに排反であり，いずれか 1 つが起これば「赤い玉がちょうど 1 回だけ出る」ことになります.

したがって，求める確率は，加法定理より

$$\begin{aligned} P(\text{赤がちょうど 1 回}) &= P\big(A_1 \cap A_2^{\mathrm{c}} \cap A_3^{\mathrm{c}} \cap A_4^{\mathrm{c}}\big) + P\big(A_1^{\mathrm{c}} \cap A_2 \cap A_3^{\mathrm{c}} \cap A_4^{\mathrm{c}}\big) \\ &\quad + P\big(A_1^{\mathrm{c}} \cap A_2^{\mathrm{c}} \cap A_3 \cap A_4^{\mathrm{c}}\big) + P\big(A_1^{\mathrm{c}} \cap A_2^{\mathrm{c}} \cap A_3^{\mathrm{c}} \cap A_4\big) \\ &= \frac{1}{35} + \frac{1}{35} + \frac{1}{35} + \frac{1}{35} \\ &= \frac{4}{35} \end{aligned}$$

となります.

問 3

1 回目に運動部の生徒が選ばれる事象を U_1，1 回目に文化部の生徒が選ばれる事象を C_1，1 回目にどちらの部活動にも入っていない生徒が選ばれる事象を N_1 とします.

同様に，2 回目，3 回目についても U_2, C_2, N_2，U_3, C_3, N_3 を定めます.

クラスの内訳は次のとおりです.

- 運動部の生徒：5 人
- 文化部の生徒：10 人
- どちらの部活動にも入っていない生徒：15 人

合計 30 人のなかから，重複なく 3 人を選びます.

(1)「3 人ともどちらの部活動にも入っていない」という事象は

$$N_1 \cap N_2 \cap N_3$$

です.

1 回目にどちらの部活動にも入っていない生徒が選ばれる確率は

$$P(N_1) = \frac{15}{30}$$

です.

1 回目にどちらの部活動にも入っていない生徒が選ばれたとき，残りはどちらの部活動にも入っていない生徒 14 人，その他 15 人となるので，このとき 2 回目にもどちらの部活動にも入っていない生徒が選ばれる確率は

$$P(N_2 \mid N_1) = \frac{14}{29}$$

となります.

さらに，1 回目と 2 回目にどちらの部活動にも入っていない生徒が選ばれたとき，残りはどちらの部活動にも入っていない生徒 13 人，その他 15 人となるので，このとき 3 回目にもどちらの部活動にも入っていない生徒が選ばれる確率は

$$P(N_3 \mid N_1 \cap N_2) = \frac{13}{28}$$

となります.

したがって，3 人ともどちらの部活動にも入っていない確率は

$$\begin{aligned} P(N_1 \cap N_2 \cap N_3) &= P(N_1)\, P(N_2 \mid N_1)\, P(N_3 \mid N_1 \cap N_2) \\ &= \frac{15}{30} \times \frac{14}{29} \times \frac{13}{28} \\ &= \frac{13}{116} \end{aligned}$$

となります.

(2)「少なくとも 1 人は運動部か文化部に入っている」という事象の余事象は，「3 人ともどちらの部活動にも入っていない」という事象 $N_1 \cap N_2 \cap N_3$ です．その確率は (1) より 13/116 です.

したがって,

$$\begin{aligned} P(\text{少なくとも 1 人は運動部か文化部}) &= 1 - P(\text{3 人ともどちらの部活動にも入っていない}) \\ &= 1 - P(N_1 \cap N_2 \cap N_3) \\ &= 1 - \frac{13}{116} \\ &= \frac{103}{116} \end{aligned}$$

となります.

(3)「3 人とも同じ種類の生徒」である場合は，次の 3 つの排反な場合に分けられます.

$$U_1 \cap U_2 \cap U_3, \qquad C_1 \cap C_2 \cap C_3, \qquad N_1 \cap N_2 \cap N_3$$

まず，3 人とも運動部の生徒である場合を考えます．1 回目に運動部の生徒が選ばれる確率は

$$P(U_1) = \frac{5}{30}$$

です.

1 回目に運動部の生徒が選ばれたとき，残りの運動部の生徒は 4 人，その他は 25 人となるので，このとき 2 回目にも運動部の生徒が選ばれる確率は

$$P(U_2 \mid U_1) = \frac{4}{29}$$

となります.

1 回目と 2 回目に運動部の生徒が選ばれたとき，残りの運動部の生徒は 3 人，その他は 25 人となるので，このとき 3 回目にも運動部の生徒が選ばれる確率は

$$P(U_3 \mid U_1 \cap U_2) = \frac{3}{28}$$

となります.

したがって，3 人とも運動部の生徒である確率は

$$\begin{aligned} P(U_1 \cap U_2 \cap U_3) &= P(U_1)\, P(U_2 \mid U_1)\, P(U_3 \mid U_1 \cap U_2) \\ &= \frac{5}{30} \times \frac{4}{29} \times \frac{3}{28} \\ &= \frac{1}{406} \end{aligned}$$

となります.

次に，3 人とも文化部の生徒である場合を考えます．1 回目に文化部の生徒が選ばれる確率は

$$P(C_1) = \frac{10}{30}$$

です.

1 回目に文化部の生徒が選ばれたとき，残りの文化部の生徒は 9 人，その他は 20 人となるので，このとき 2 回目にも文化部の生徒が選ばれる確率は

$$P(C_2 \mid C_1) = \frac{9}{29}$$

となります.

1 回目と 2 回目に文化部の生徒が選ばれたとき，残りの文化部の生徒は 8 人，その他は 20 人となるので，このとき 3 回目にも文化部の生徒が選ばれる確率は

$$P(C_3 \mid C_1 \cap C_2) = \frac{8}{28}$$

となります.

したがって，3 人とも文化部の生徒である確率は

$$\begin{aligned} P(C_1 \cap C_2 \cap C_3) &= P(C_1)\, P(C_2 \mid C_1)\, P(C_3 \mid C_1 \cap C_2) \\ &= \frac{10}{30} \times \frac{9}{29} \times \frac{8}{28} \\ &= \frac{6}{203} \end{aligned}$$

となります.

3 人ともどちらの部活動にも入っていない確率は，(1) より

$$P(N_1 \cap N_2 \cap N_3) = \frac{13}{116}$$

です.

これら 3 つの事象は互いに排反であり，いずれか 1 つが起これば「3 人とも同じ種類の生徒が選ばれる」ことになります.

したがって，加法定理より，3 人とも同じ種類の生徒である確率は

$$\begin{aligned} P(\text{3 人とも同じ種類}) &= P(U_1 \cap U_2 \cap U_3) + P(C_1 \cap C_2 \cap C_3) + P(N_1 \cap N_2 \cap N_3) \\ &= \frac{1}{406} + \frac{6}{203} + \frac{13}{116} \\ &= \frac{117}{812} \end{aligned}$$

となります.

10.2　成功回数についての確率

問題 10.4

(1) **順番をつけて選ぶ場合**

最初に選ぶ記号ごとに整理して書き出します.

- A を最初に選ぶ場合：(A,B)，(A,C)，(A,D)，(A,E)
- B を最初に選ぶ場合：(B,A)，(B,C)，(B,D)，(B,E)
- C を最初に選ぶ場合：(C,A)，(C,B)，(C,D)，(C,E)
- D を最初に選ぶ場合：(D,A)，(D,B)，(D,C)，(D,E)
- E を最初に選ぶ場合：(E,A)，(E,B)，(E,C)，(E,D)

したがって，順番をつけて 2 個を選ぶときの選び方は

$$5 \times 4 = 20 \text{ 通り}$$

あります.

(2) **順番をつけずに選ぶ場合**

(1) で求めたように，A と B の 2 個を選んだときには，(A,B) と (B,A) のように，同じ 2 個の組が 2×1 通りずつ現れます．ここで，

$$2 \times 1$$

は，2 個の記号から 2 個を順番をつけて選ぶときの選び方の総数です．1 つ目の位置には 2 通り，2 つ目の位置には残り 1 通りしかないためです.

このような重複を取り除くために，(1) で順番をつけて数えた 20 通りを 2×1 で割ります.

したがって，順番を区別しない選び方の総数は

$$\frac{20}{2 \times 1} = 10$$

となります.

問題 10.5

(1) 7 人のなかから 3 人を選ぶときの選び方の総数は

$$\binom{7}{3} = \frac{7 \times 6 \times 5}{3 \times 2 \times 1} = 35$$

です（7 人のなかから 3 人を選ぶとき，1 人目の選び方は 7 通り，2 人目は 6 通り，3 人目は 5 通りなので，分子は 7×6×5 の 3 数の積になります）.

(2) 8 個の色玉（赤・青・黄・緑・白・黒・紫・橙）のなかから 4 個を選ぶときの選び方の総数は

$$\binom{8}{4} = \frac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2 \times 1} = 70$$

です（8 個の色玉のなかから 4 個を選ぶとき，1 個目の選び方は 8 通り，2 個目は 7 通り，3 個目は 6 通り，4 個目は 5 通りなので，分子は 8×7×6×5 の 4 数の積になります）.

(3) 12 種類の果物のなかから 4 種類を選ぶときの選び方の総数は

$$\binom{12}{4} = \frac{12 \times 11 \times 10 \times 9}{4 \times 3 \times 2 \times 1} = 495$$

です（12 種類の果物のなかから 4 種類を選ぶとき，1 種類目の選び方は 12 通り，2 種類目は 11 通り，3 種類目は 10 通り，4 種類目は 9 通りなので，分子は 12×11×10×9 の 4 数の積になります）.

(4) 6 種類のカードのなかから 1 種類を選ぶときの選び方の総数は

$$\binom{6}{1} = 6$$

です（6 種類のカードのなかから 1 種類を選ぶとき，選び方は 6 通りです）.

また，6 種類のカードのなかから 5 種類を選ぶときの選び方の総数は

$$\binom{6}{5} = \frac{6 \times 5 \times 4 \times 3 \times 2}{5 \times 4 \times 3 \times 2 \times 1} = 6$$

です（6 種類のカードのなかから 5 種類を選ぶとき，1 種類目の選び方は 6 通り，2 種類目は 5 通り，3 種類目は 4 通り，4 種類目は 3 通り，5 種類目は 2 通りなので，分子は 6×5×4×3×2 の 5 数の積になります）.

以上より，

$$\binom{6}{1} = \binom{6}{5}$$

が成り立ちます．これは，「1 種類を選ぶこと」は「選ばない 5 種類を決めること」と同じであると考えれば明らかです.

問題 10.6

(1) 裏が出る確率は 1/2 なので，8 回とも裏になる確率は

$$\left(\frac{1}{2}\right)^8 = \frac{1}{256}$$

です.

(2) 表がちょうど 4 回になるとき，表が出る回の選び方は

$$\binom{8}{4} = \frac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2 \times 1} = 70$$

通りあります.

たとえば「1 回目から 4 回目だけ表」の確率は

$$\left(\frac{1}{2}\right)^4 \times \left(\frac{1}{2}\right)^4 = \frac{1}{256}$$

です.

同じ計算で確率が求められる並びが 70 通りあるので，表がちょうど 4 回になる確率は

$$70 \times \frac{1}{256} = \frac{70}{256} = \frac{35}{128}$$

となります.

(3) 表が 6 回以上になる確率は，「表が 6 回」「表が 7 回」「表が 8 回」のそれぞれの確率をすべて足し合わせて求められます.

まず，表がちょうど 6 回になる場合を考えます．表が出る回の選び方は

$$\binom{8}{6} = \frac{8\times7\times6\times5\times4\times3}{6\times5\times4\times3\times2\times1} = 28$$

通りあります．

たとえば「1 回目から 6 回目が表，7 回目と 8 回目が裏」の確率は

$$\left(\frac{1}{2}\right)^6 \times \left(\frac{1}{2}\right)^2 = \left(\frac{1}{2}\right)^8 = \frac{1}{256}$$

です．

同じ計算で確率が求められる並びが 28 通りあるので，表がちょうど 6 回になる確率は

$$28 \times \frac{1}{256} = \frac{28}{256}$$

となります．

次に，表がちょうど 7 回になる確率は

$$\binom{8}{7} \times \frac{1}{256} = 8 \times \frac{1}{256} = \frac{8}{256}$$

です．

また，表がちょうど 8 回になる確率は

$$\binom{8}{8} \times \frac{1}{256} = 1 \times \frac{1}{256} = \frac{1}{256}$$

です．

したがって，表が 6 回以上になる確率は

$$\frac{28}{256} + \frac{8}{256} + \frac{1}{256} = \frac{37}{256}$$

となります．

問題 10.7

(1) 白い花が 1 本も観察されない場合を考えます．このときは，1 本目から 5 本目まで全部赤い花が観察されるという並びしかありません．

この並びが起こる確率は

$$\frac{8}{12} \times \frac{7}{11} \times \frac{6}{10} \times \frac{5}{9} \times \frac{4}{8} = \frac{7}{99}$$

です．

したがって，白い花が 1 本も観察されない確率は

$$\frac{7}{99}$$

となります．

(2) 白い花がちょうど 2 本だけ観察されるとき，白い花が観察される本の選び方は

$$\binom{5}{2} = \frac{5\times4}{2\times1} = 10$$

通りあります．

たとえば「1 本目と 3 本目だけ白い花が観察される」場合を考えます．この並びが起こる確率は

$$\frac{4}{12}\times\frac{8}{11}\times\frac{3}{10}\times\frac{7}{9}\times\frac{6}{8}=\frac{7}{165}$$

です．

同じ形の計算で確率が求められる並びが 10 通りあるので，白い花がちょうど 2 本だけ観察される確率は

$$10\times\frac{7}{165}=\frac{14}{33}$$

となります．

(3) 白い花が 3 本以上観察される確率は，「白い花が 3 本」「白い花が 4 本」のそれぞれの確率を足し合わせて求められます．

まず，白い花がちょうど 3 本だけ観察される場合を考えます．白い花が観察される本の選び方は

$$\binom{5}{3}=\frac{5\times4\times3}{3\times2\times1}=10$$

通りあります．

たとえば「1 本目から 3 本目までだけ白い花が観察される」場合を考えます．この並びが起こる確率は

$$\frac{4}{12}\times\frac{3}{11}\times\frac{2}{10}\times\frac{8}{9}\times\frac{7}{8}=\frac{7}{495}$$

です．

同じ計算で確率が求められる並びが 10 通りあるので，白い花がちょうど 3 本だけ観察される確率は

$$10\times\frac{7}{495}=\frac{14}{99}$$

となります．

次に，白い花がちょうど 4 本だけ観察される場合を考えます．白い花が観察される本の選び方は

$$\binom{5}{4}=\frac{5\times4\times3\times2}{4\times3\times2\times1}=5$$

通りあります．

たとえば「1 本目から 4 本目までだけ白い花が観察される」場合を考えます．この並びが起こる確率は

$$\frac{4}{12}\times\frac{3}{11}\times\frac{2}{10}\times\frac{1}{9}\times\frac{8}{8}=\frac{1}{495}$$

です．

同じ計算で確率が求められる並びが 5 通りあるので，白い花がちょうど 4 本だけ観察される確率は

$$5\times\frac{1}{495}=\frac{1}{99}$$

となります．

したがって，白い花が 3 本以上観察される確率は

$$\frac{14}{99}+\frac{1}{99}=\frac{5}{33}$$

となります．

問題 10.8

この問題では，全体の個数が 16 であり，そのうち 4 本を選ぶ状況を考えます．ここで，赤ワインは 10 本，白ワインは 6 本あります．

(1) 16 本のなかから 4 本を選ぶ選び方の総数は

$$\binom{16}{4} = \frac{16 \times 15 \times 14 \times 13}{4 \times 3 \times 2 \times 1} = 1820$$

です．

(2) 赤ワインがちょうど 2 本になる場合を考えます．

まず，赤ワイン 10 本のなかから 2 本を選ぶときの選び方の総数は

$$\binom{10}{2} = \frac{10 \times 9}{2 \times 1} = 45$$

です．

次に，残りの 2 本は白ワイン 6 本のなかから 2 本を選びます．白ワイン 6 本のなかから 2 本を選ぶときの選び方の総数は

$$\binom{6}{2} = \frac{6 \times 5}{2 \times 1} = 15$$

です．

ここで，「赤ワインを 2 本選ぶ 45 通り」と「白ワインを 2 本選ぶ 15 通り」が組み合わさって，「赤ワインがちょうど 2 本になるような 4 本の選び方」になります．つまり，その選び方の総数は

$$(\text{赤 2 本の選び方の総数}) \times (\text{白 2 本の選び方の総数}) = 45 \times 15 = 675$$

と計算できるので，675 通りあることがわかります．

したがって，4 本を無作為に選ぶとき，赤ワインがちょうど 2 本になる確率は

$$\frac{675}{1820} = \frac{135}{364}$$

となります．

(3) 4 本とも赤ワインになる場合を考えます．

赤ワイン 10 本のなかから 4 本を選ぶときの選び方の総数は

$$\binom{10}{4} = \frac{10 \times 9 \times 8 \times 7}{4 \times 3 \times 2 \times 1} = 210$$

です．白ワインからは選ばないので，「4 本とも赤ワインになるような 4 本の選び方の総数」は 210 通りです．

したがって，4 本を無作為に選ぶとき，4 本とも赤ワインになる確率は

$$\frac{210}{1820} = \frac{3}{26}$$

となります．

10.2 の確認テスト

問 1

(1) 青で通過できる確率は 2/3 なので，これを「成功」とみなすと，成功確率は 2/3 です．

青で通過できる信号がちょうど 3 個であるとき，青になる信号の選び方は

$$\binom{7}{3} = \frac{7 \times 6 \times 5}{3 \times 2 \times 1} = 35$$

通りあります．

たとえば「1 つ目から 3 つ目だけ青」の確率は

$$\left(\frac{2}{3}\right)^3 \times \left(\frac{1}{3}\right)^4 = \frac{8}{2187}$$

です．

同じ計算で確率が求められる並びが 35 通りあるので，青で通過できる信号がちょうど 3 個である確率は

$$35 \times \frac{8}{2187} = \frac{280}{2187}$$

となります．

(2) 青で通過できる信号が 1 個もないということは，7 個すべての信号が赤であるということです．

赤になる確率は 1/3 なので，7 個すべて赤である確率は

$$\left(\frac{1}{3}\right)^7 = \frac{1}{2187}$$

となります．

(3) 青で通過できる信号の個数が 2 個以下になる確率は，「青が 0 個」「青が 1 個」「青が 2 個」のそれぞれの確率をすべて足し合わせて求められます．

まず，青が 0 個である確率は（(2) より）

$$\left(\frac{1}{3}\right)^7 = \frac{1}{2187}$$

です．

次に，青がちょうど 1 個である場合を考えます．青になる信号の選び方は

$$\binom{7}{1} = 7$$

通りあります．

たとえば「1 つ目だけ青」の確率は

$$\left(\frac{2}{3}\right)^1 \times \left(\frac{1}{3}\right)^6 = \frac{2}{2187}$$

です．

同じ計算が 7 通りあるので，

$$7 \times \frac{2}{2187} = \frac{14}{2187}$$

となります．

次に，青がちょうど 2 個である場合を考えます．青になる信号の選び方は

$$\binom{7}{2} = \frac{7 \times 6}{2 \times 1} = 21$$

通りあります．

たとえば「1 つ目と 2 つ目だけ青」の確率は

$$\left(\frac{2}{3}\right)^2 \times \left(\frac{1}{3}\right)^5 = \frac{4}{2187}$$

です．

同じ計算が 21 通りあるので，

$$21 \times \frac{4}{2187} = \frac{84}{2187}$$

となります．

したがって，青で通過できる信号の個数が 2 個以下になる確率は

$$\frac{1}{2187} + \frac{14}{2187} + \frac{84}{2187} = \frac{99}{2187} = \frac{11}{243}$$

となります．

問 2

(1) 黒インクのペンがちょうど 3 本になる場合を考えます．

まず，黒インク 4 本のなかから 3 本を選ぶときの選び方の総数は

$$\binom{4}{3} = \frac{4 \times 3 \times 2}{3 \times 2 \times 1} = 4$$

です．

次に，残りの 2 本は青インク 6 本のなかから選びます．青インク 6 本のなかから 2 本を選ぶときの選び方の総数は

$$\binom{6}{2} = \frac{6 \times 5}{2 \times 1} = 15$$

です．

ここで，「黒インク 3 本を選ぶ 4 通り」と「青インク 2 本を選ぶ 15 通り」が組み合わさって，「黒インクがちょうど 3 本になるような 5 本の選び方」になります．つまり，その選び方の総数は

$$\text{（黒 3 本の選び方の総数）} \times \text{（青 2 本の選び方の総数）} = 4 \times 15 = 60$$

と計算できるので，60 通りあることがわかります．

また，10 本のなかから 5 本を選ぶ選び方の総数は

$$\binom{10}{5} = \frac{10 \times 9 \times 8 \times 7 \times 6}{5 \times 4 \times 3 \times 2 \times 1} = 252$$

です．

よって，黒インクのペンがちょうど 3 本である確率は

$$\frac{60}{252} = \frac{5}{21}$$

となります．

(2) 黒インクの本数が青インクより多くなるのは，「黒が 3 本，青が 2 本」または「黒が 4 本，青が 1 本」の場合です．

まず，10 本のなかから 5 本を選ぶ選び方の総数は，(1) で求めたように 252 です．

黒が 3 本，青が 2 本になる場合の選び方は，(1) より

$$60 \text{ 通り}$$

あります．

次に，黒が 4 本，青が 1 本になる場合を考えます．

黒インク 4 本のなかから 4 本を選ぶ選び方は

$$\binom{4}{4} = 1$$

通りあり，青インク 6 本のなかから 1 本を選ぶ選び方は

$$\binom{6}{1} = 6$$

通りあります．

したがって，黒が 4 本，青が 1 本になる選び方の総数は

$$1 \times 6 = 6$$

通りです．

よって，黒インクの本数が青インクより多くなる選び方の総数は

$$60 + 6 = 66$$

通りです．

したがって，求める確率は

$$\frac{66}{252} = \frac{11}{42}$$

となります．

(3) 黒インクと青インクがどちらも少なくとも 1 本ずつ含まれる場合を考えます．

黒インクは全部で 4 本しかないため，5 本すべてが黒インクになることはありません．したがって，5 本すべてが青インクになる場合だけを除けばよいことになります．

まず，10 本のなかから 5 本を選ぶ選び方の総数は，(1) で求めたように 252 です．

青インクだけ 5 本になる場合の選び方は

$$\binom{6}{5} = \frac{6 \times 5 \times 4 \times 3 \times 2}{5 \times 4 \times 3 \times 2 \times 1} = 6$$

通りあります．

したがって，「黒インクと青インクがどちらも少なくとも 1 本ずつ含まれる」選び方の総数は

$$252 - 6 = 246$$

通りです．

よって，求める確率は

$$\frac{246}{252} = \frac{41}{42}$$

となります.

問 3

(1) 最初の 2 回の取り出しで得られた 2 枚のカードの「穴の数の合計」が 1 になる場合を考えます.

まず，穴が 1 つあるカードは 4 枚あるので，そのなかから 1 枚を選ぶときの選び方の総数は 4 です.

次に，穴が 0 のカードは 5 枚あるので，そのなかから 1 枚を選ぶときの選び方の総数は 5 です.

ここで,「穴が 1 つあるカードを 1 枚選ぶ 4 通り」と「穴が 0 のカードを 1 枚選ぶ 5 通り」が組み合わさって,「穴の数の合計が 1 になる 2 枚の選び方」になります．つまり，その選び方の総数は

$$(\text{穴 } 1 \text{ の選び方の総数}) \times (\text{穴 } 0 \text{ の選び方の総数}) = 4 \times 5 = 20$$

と計算できるので，20 通りあることがわかります.

10 枚のカードのなかから 2 枚を選ぶときの選び方の総数は

$$\binom{10}{2} = \frac{10 \times 9}{2 \times 1} = 45$$

なので，求める確率は

$$\frac{20}{45} = \frac{4}{9}$$

となります.

(2) 最初の 2 回の取り出しで得られた 2 枚のカードの「穴の数がちがう」場合を考えます.

このような穴の数の組は

$$\{2,1\}, \quad \{2,0\}, \quad \{1,0\}$$

の 3 通りです.

まず，穴が 2 つあるカードは 1 枚，穴が 1 つあるカードは 4 枚あるので，{2,1} の場合の選び方は

$$1 \times 4 = 4$$

通りあります.

次に，穴が 2 つあるカードは 1 枚，穴が 0 のカードは 5 枚あるので，{2,0} の場合の選び方は

$$1 \times 5 = 5$$

通りあります.

また，穴が 1 つあるカードは 4 枚，穴が 0 のカードは 5 枚あるので，{1,0} の場合の選び方は

$$4 \times 5 = 20$$

通りあります.

したがって，穴の数がちがう 2 枚のカードの選び方の総数は

$$4 + 5 + 20 = 29$$

です.

10 枚のカードのなかから 2 枚を選ぶときの選び方の総数は，(1) で求めたように 45 なので，求める確率は

$$\frac{29}{45}$$

です.

(3) 最初の 2 回の取り出しで得られた 2 枚のカードの「穴の数の差の絶対値」が 1 になる場合を考えます.

このような穴の数の組は

$$\{2,1\},\quad \{1,0\}$$

の 2 通りです.

まず，穴が 2 つあるカードは 1 枚，穴が 1 つあるカードは 4 枚あるので，{2,1} の場合の選び方は

$$1\times 4=4$$

通りあります.

また，穴が 1 つあるカードは 4 枚，穴が 0 のカードは 5 枚あるので，{1,0} の場合の選び方は

$$4\times 5=20$$

通りあります.

したがって，穴の数の差の絶対値が 1 になる 2 枚のカードの選び方の総数は

$$4+20=24$$

です.

10 枚のカードのなかから 2 枚を選ぶときの選び方の総数は，(1) で求めたように 45 なので，求める確率は

$$\frac{24}{45}=\frac{8}{15}$$

となります.

(4) 最初の 3 回の取り出しで得られた 3 枚のカードの「穴が 1 つあるカードの枚数」がちょうど 2 枚になる場合を考えます.

まず，穴が 1 つあるカードは 4 枚あります．また，それ以外のカード（穴が 2 つまたは 0 のカード）は

$$1+5=6$$

枚あります.

穴が 1 つあるカードを 2 枚，それ以外のカードを 1 枚選ぶときの選び方の総数は

$$\binom{4}{2}\times\binom{6}{1}=\frac{4\times 3}{2\times 1}\times 6=6\times 6=36$$

です.

10 枚のカードのなかから 3 枚を選ぶときの選び方の総数は

$$\binom{10}{3}=\frac{10\times 9\times 8}{3\times 2\times 1}=120$$

なので，求める確率は

$$\frac{36}{120}=\frac{3}{10}$$

となります.

第 11 章　確率変数と確率分布

11.1　確率変数の考え方

問題 11.1

(1) サイコロを 1 回投げるとき，出る目は

$$1,\ 2,\ 3,\ 4,\ 5,\ 6$$

の 6 通りです．どの目も同じ確率で出るので，それぞれの確率は

$$\frac{1}{6}$$

です．

したがって，確率変数 X のとりうる値は

$$1,\ 2,\ 3,\ 4,\ 5,\ 6$$

であり，この 6 通りはすべて同じ確率 1/6 で起こるので，

- 目が 1 になる事象は 1 通りあり，その確率は 1/6
- 目が 2 になる事象は 1 通りあり，その確率は 1/6
- 目が 3 になる事象は 1 通りあり，その確率は 1/6
- 目が 4 になる事象は 1 通りあり，その確率は 1/6
- 目が 5 になる事象は 1 通りあり，その確率は 1/6
- 目が 6 になる事象は 1 通りあり，その確率は 1/6

となります．したがって，

$$P(X=1)=\frac{1}{6},\ P(X=2)=\frac{1}{6},\ P(X=3)=\frac{1}{6},\ P(X=4)=\frac{1}{6},\ P(X=5)=\frac{1}{6},\ P(X=6)=\frac{1}{6}$$

となります．これらを整理すると，次のような確率分布表になります．

x	1	2	3	4	5	6
$P(X=x)$	1/6	1/6	1/6	1/6	1/6	1/6

(2) サイコロを 2 回投げたときに出た目の組と，その合計を表にすると次のようになります．

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

この表を見ると，合計が 2 のマスは 1 個，合計が 3 のマスは 2 個，合計が 4 のマスは 3 個，…… というように数えられます.

この 36 個のマスはすべて同じ確率 1/36 で起こるので,

- 合計が 2 になる事象は 1 通りあり，その確率は 1/36
- 合計が 3 になる事象は 2 通りあり，その確率は 2/36
- 合計が 4 になる事象は 3 通りあり，その確率は 3/36
- 合計が 5 になる事象は 4 通りあり，その確率は 4/36
- 合計が 6 になる事象は 5 通りあり，その確率は 5/36
- 合計が 7 になる事象は 6 通りあり，その確率は 6/36
- 合計が 8 になる事象は 5 通りあり，その確率は 5/36
- 合計が 9 になる事象は 4 通りあり，その確率は 4/36
- 合計が 10 になる事象は 3 通りあり，その確率は 3/36
- 合計が 11 になる事象は 2 通りあり，その確率は 2/36
- 合計が 12 になる事象は 1 通りあり，その確率は 1/36

となります．したがって，確率変数 Y のとりうる値は

$$2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ 11,\ 12$$

であり，次のような確率分布表になります.

y	2	3	4	5	6	7	8	9	10	11	12
$P(Y=y)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

問題 11.2

(1) 問題 11.1 で求めたように，確率変数 X のとりうる値は 1，2，3，4，5，6 の 6 通りであり，どの目も同じ確率で出るので，確率分布は次のようになります.

x	1	2	3	4	5	6
$P(X=x)$	1/6	1/6	1/6	1/6	1/6	1/6

確率変数の期待値は,「値」と「その値が出る確率」を掛け合わせ，それらをすべて足し合わせて求めるので,

$$E(X) = 1\times\frac{1}{6} + 2\times\frac{1}{6} + 3\times\frac{1}{6} + 4\times\frac{1}{6} + 5\times\frac{1}{6} + 6\times\frac{1}{6} = \frac{21}{6} = \frac{7}{2}$$

となります.

(2) 問題 11.1 で，出た目の組とその合計を調べることで，確率変数 Y の確率分布は次のようになることがわかりました.

y	2	3	4	5	6	7	8	9	10	11	12
$P(Y=y)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

この確率分布を使って，期待値を求めます．期待値は，それぞれの合計の値について「値 $\times$ 確率」を計算し，それらをすべて足し合わせて求めるので,

$$\begin{aligned} E(Y) &= 2\times\frac{1}{36} + 3\times\frac{2}{36} + 4\times\frac{3}{36} + 5\times\frac{4}{36} + 6\times\frac{5}{36} + 7\times\frac{6}{36} \\ &\quad + 8\times\frac{5}{36} + 9\times\frac{4}{36} + 10\times\frac{3}{36} + 11\times\frac{2}{36} + 12\times\frac{1}{36} \\ &= 7 \end{aligned}$$

となります.

問題 11.3

(1) 問題 11.1 より，確率変数 X の確率分布は次のとおりです.

x	1	2	3	4	5	6
$P(X=x)$	1/6	1/6	1/6	1/6	1/6	1/6

期待値は，問題 11.2 より $E(X)=7/2$ です.

したがって，分散は，1，2，3，4，5，6 のそれぞれについて「(値−期待値)^2×確率」を計算し，それらを足し合わせて求めるので,

$$\begin{aligned} V(X) &= \left(1-\frac{7}{2}\right)^2\times\frac{1}{6}+\left(2-\frac{7}{2}\right)^2\times\frac{1}{6}+\left(3-\frac{7}{2}\right)^2\times\frac{1}{6} \\ &\quad+\left(4-\frac{7}{2}\right)^2\times\frac{1}{6}+\left(5-\frac{7}{2}\right)^2\times\frac{1}{6}+\left(6-\frac{7}{2}\right)^2\times\frac{1}{6} \\ &= \frac{35}{12} \end{aligned}$$

となります.

標準偏差は，分散の（負でない）平方根なので,

$$\sqrt{V(X)}=\sqrt{\frac{35}{12}}=\frac{\sqrt{35}}{\sqrt{12}}=\frac{\sqrt{35}}{2\sqrt{3}}=\frac{\sqrt{35\times3}}{2\times3}=\frac{\sqrt{105}}{6}=\frac{10.25}{6}=1.708\ldots\approx1.71$$

となります.

(2) 問題 11.1 より，確率変数 Y の確率分布は次のとおりです.

y	2	3	4	5	6	7	8	9	10	11	12
$P(Y=y)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

期待値は，問題 11.2 より $E(Y)=7$ です.

したがって，分散は，2，3，4，...，12 のそれぞれについて「(値 − 期待値)^2 × 確率」を計算し，それらを足し合わせて求めるので,

$$\begin{aligned} V(Y) &= (2-7)^2\times\frac{1}{36}+(3-7)^2\times\frac{2}{36}+(4-7)^2\times\frac{3}{36}+(5-7)^2\times\frac{4}{36}+(6-7)^2\times\frac{5}{36} \\ &\quad+(7-7)^2\times\frac{6}{36}+(8-7)^2\times\frac{5}{36}+(9-7)^2\times\frac{4}{36}+(10-7)^2\times\frac{3}{36}+(11-7)^2\times\frac{2}{36}+(12-7)^2\times\frac{1}{36} \\ &= \frac{35}{6} \end{aligned}$$

となります.

標準偏差は，分散の（負でない）平方根なので,

$$\sqrt{V(Y)}=\sqrt{\frac{35}{6}}=\frac{\sqrt{35}}{\sqrt{6}}=\frac{\sqrt{35\times6}}{6}=\frac{\sqrt{210}}{6}=\frac{14.49}{6}=2.415\ldots\approx2.42$$

となります.

11.1 の確認テスト

問 1

(1) 1 回目に赤が出る確率は

$$\frac{1}{3}$$

青が出る確率は

$$\frac{2}{3}$$

です．毎回玉を戻すので，2 回目も同じ確率になります．

2 回の操作で出る色の組み合わせは

(赤，赤), (赤，青), (青，赤), (青，青)

の 4 通りです．それぞれの確率は

$$P(\text{赤，赤}) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9},$$
$$P(\text{赤，青}) = \frac{1}{3} \times \frac{2}{3} = \frac{2}{9},$$
$$P(\text{青，赤}) = \frac{2}{3} \times \frac{1}{3} = \frac{2}{9},$$
$$P(\text{青，青}) = \frac{2}{3} \times \frac{2}{3} = \frac{4}{9}$$

となります．

赤が出た回数を表す確率変数 X のとりうる値は

0, 1, 2

の 3 通りです．それぞれについて，起こりうる結果とその確率を整理すると，

- 赤が 0 回出る事象は (青，青) の 1 通りであり，その確率は 4/9
- 赤が 1 回出る事象は (赤，青), (青，赤) の 2 通りであり，それぞれの確率は 2/9 なので，足し合わせると 4/9
- 赤が 2 回出る事象は (赤，赤) の 1 通りであり，その確率は 1/9

となります．したがって，

$$P(X=0) = \frac{4}{9}, \quad P(X=1) = \frac{4}{9}, \quad P(X=2) = \frac{1}{9}$$

となります．これらを整理すると，次のような確率分布表になります．

x	0	1	2
$P(X=x)$	4/9	4/9	1/9

(2) 期待値は，「値」と「その値が出る確率」を掛け合わせ，それらをすべて足し合わせて求めるので，

$$\begin{aligned} E(X) &= 0 \times \frac{4}{9} + 1 \times \frac{4}{9} + 2 \times \frac{1}{9} \\ &= 0 + \frac{4}{9} + \frac{2}{9} \\ &= \frac{6}{9} = \frac{2}{3} \end{aligned}$$

となります.

分散は，0，1，2 のそれぞれについて「(値 − 期待値)^2 × 確率」を計算し，それらを足し合わせて求めるので,

$$\begin{aligned} V(X) &= \left(0-\frac{2}{3}\right)^2 \times \frac{4}{9} + \left(1-\frac{2}{3}\right)^2 \times \frac{4}{9} + \left(2-\frac{2}{3}\right)^2 \times \frac{1}{9} \\ &= \frac{4}{9} \times \frac{4}{9} + \frac{1}{9} \times \frac{4}{9} + \frac{16}{9} \times \frac{1}{9} \\ &= \frac{16}{81} + \frac{4}{81} + \frac{16}{81} \\ &= \frac{36}{81} = \frac{4}{9} \end{aligned}$$

となります.

標準偏差は，分散の（負でない）平方根なので,

$$\sqrt{V(X)} = \sqrt{\frac{4}{9}} = \frac{2}{3}$$

となります.

問 2

サイコロを 1 回投げたときに，出た目が偶数になるのは，2，4，6 が出たときであり，奇数になるのは，1，3，5 が出たときです．どの目も同じ確率で出るので,

$$P(X=2) = \frac{3}{6} = \frac{1}{2}, \qquad P(X=0) = \frac{3}{6} = \frac{1}{2}$$

となります．したがって，確率分布は次のようになります.

x	0	2
$P(X=x)$	1/2	1/2

期待値は,「値」と「その値が出る確率」を掛け合わせ，それらを足し合わせて求めるので,

$$E(X) = 0 \times \frac{1}{2} + 2 \times \frac{1}{2} = 1$$

となります.

分散は,「(値 − 期待値)^2 × 確率」をそれぞれについて計算し，足し合わせて求めるので,

$$\begin{aligned} V(X) &= (0-1)^2 \times \frac{1}{2} + (2-1)^2 \times \frac{1}{2} \\ &= 1 \times \frac{1}{2} + 1 \times \frac{1}{2} \\ &= 1 \end{aligned}$$

となります.

標準偏差は，分散の（負でない）平方根なので,

$$\sqrt{V(X)} = \sqrt{1} = 1$$

となります.

問 3

(1) 100 円のくじは全部で 10 本あります．くじ全体は 20 本なので，

$$P(X=100)=\frac{10}{20}=\frac{1}{2}$$

となります．

(2) くじの本数は全部で 20 本であり，それぞれの賞金額とその確率は次のようになります．

x	0	100	200	1000
$P(X=x)$	3/20	10/20	5/20	2/20

期待値は，「値」と「その値が出る確率」を掛け合わせ，それらをすべて足し合わせて求めるので，

$$\begin{aligned}E(X)&=0\times\frac{3}{20}+100\times\frac{10}{20}+200\times\frac{5}{20}+1000\times\frac{2}{20}\\&=0+50+50+100\\&=200\end{aligned}$$

となります．

(3) 分散は，0，100，200，1000 のそれぞれについて「(値 − 期待値)^2 × 確率」を計算し，それらを足し合わせて求めるので，

$$\begin{aligned}V(X)&=(0-200)^2\times\frac{3}{20}+(100-200)^2\times\frac{10}{20}+(200-200)^2\times\frac{5}{20}+(1000-200)^2\times\frac{2}{20}\\&=40000\times\frac{3}{20}+10000\times\frac{10}{20}+0\times\frac{5}{20}+640000\times\frac{2}{20}\\&=6000+5000+0+64000\\&=75000\end{aligned}$$

となります．

11.2 確率変数の性質

問題 11.4

商品の支払額（円）を表す確率変数 Z は

$$Z=1000-80X$$

と定められています．これは，「出た目を 80 円ずつ割引し，1000 円から引いた金額」を表しています．

ここでは，期待値の線形性

$$E(aX+b)=aE(X)+b$$

を使って，$E(Z)$ を求めます．

まず，

$$E(Z)=E(1000-80X)$$

となります．期待値の線形性をそのまま用いると，

$$E(1000-80X)=E(-80X+1000)=-80E(X)+1000$$

となります．例題 11.2 より $E(X) = 7/2$ なので，

$$-80E(X) + 1000 = -80 \times \frac{7}{2} + 1000 = -280 + 1000 = 720$$

となります．

したがって，Z の期待値は

$$E(Z) = 720$$

となります．

問題 11.5

(1) 問題 11.1 より，確率変数 X の確率分布は次のとおりです．

x	1	2	3	4	5	6
$P(X = x)$	1/6	1/6	1/6	1/6	1/6	1/6

期待値は，問題 11.2 より $E(X) = 7/2$ です．

まず，X の値に応じて X^2 の値がどのようになるかを確認します．また，それぞれの値をとる確率もあわせてまとめると，次のようになります．

X	1	2	3	4	5	6
X^2	1	4	9	16	25	36
確率	1/6	1/6	1/6	1/6	1/6	1/6

これを使って，X^2 の期待値 $E(X^2)$ を求めます．期待値は，「値 × 確率」をそれぞれ計算し，それらを足し合わせて求めるので，

$$E(X^2) = 1 \times \frac{1}{6} + 4 \times \frac{1}{6} + 9 \times \frac{1}{6} + 16 \times \frac{1}{6} + 25 \times \frac{1}{6} + 36 \times \frac{1}{6} = \frac{91}{6}$$

となります．

したがって，

$$V(X) = E(X^2) - \{E(X)\}^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

となります．

(2) 問題 11.1 より，確率変数 Y の確率分布は次のとおりです．

y	2	3	4	5	6	7	8	9	10	11	12
$P(Y = y)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

期待値は，問題 11.2 より $E(Y) = 7$ です．

まず，Y の値に応じて Y^2 の値がどのようになるかを確認します．また，それぞれの値をとる確率もあわせてまとめると，次のようになります．

Y	2	3	4	5	6	7	8	9	10	11	12
Y^2	4	9	16	25	36	49	64	81	100	121	144
確率	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

これを使って，Y^2 の期待値 $E(Y^2)$ を求めます．期待値は，「値 $\times$ 確率」をそれぞれ計算し，それらを足し合わせて求めるので，

$$\begin{aligned} E(Y^2) &= 4 \times \frac{1}{36} + 9 \times \frac{2}{36} + 16 \times \frac{3}{36} + 25 \times \frac{4}{36} + 36 \times \frac{5}{36} + 49 \times \frac{6}{36} \\ &\quad + 64 \times \frac{5}{36} + 81 \times \frac{4}{36} + 100 \times \frac{3}{36} + 121 \times \frac{2}{36} + 144 \times \frac{1}{36} \\ &= \frac{329}{6} \end{aligned}$$

となります．

したがって，

$$V(Y) = E(Y^2) - \{E(Y)\}^2 = \frac{329}{6} - 7^2 = \frac{35}{6}$$

となります．

問題 11.6

確率変数 X の分散は，問題 11.3 より $V(X) = 35/12$ です．

確率変数 Z は

$$Z = 1000 - 80X$$

と表されるので，分散の性質 $V(aX+b) = a^2\,V(X)$ を用いると

$$V(Z) = V(1000 - 80X) = V(-80X + 1000) = (-80)^2 V(X) = 6400 \times \frac{35}{12} = \frac{56000}{3}$$

となります．

11.2 の確認テスト

問 1

(1) 確率変数 X は，0〜9 の数字カードを無作為に 1 枚引いたときに出た数字を表しています．どのカードも同じ確率で選ばれるので，X のとりうる値とその確率は次のようになります．

x	0	1	2	3	4	5	6	7	8	9
$P(X=x)$	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10

期待値は，「値」と「その値が出る確率」を掛け合わせ，それらをすべて足し合わせて求めるので，

$$E(X) = 0 \times \frac{1}{10} + 1 \times \frac{1}{10} + 2 \times \frac{1}{10} + \cdots + 9 \times \frac{1}{10} = \frac{45}{10} = \frac{9}{2}$$

となります．

(2) 期待値の線形性

$$E(aX+b) = aE(X) + b$$

を用いると，

$$E(Y) = E(1000X) = 1000E(X) = 1000 \times \frac{9}{2} = 4500$$

となります．

問 2

(1) 確率変数 X は，成功したときに 5 点，失敗したときに 1 点をとるので，とりうる値とその確率は次のようになります．

x	1	5
$P(X=x)$	3/10	7/10

期待値は，「値 $\times$ 確率」をそれぞれ計算し，それらを足し合わせて求めるので，

$$E(X) = 1 \times \frac{3}{10} + 5 \times \frac{7}{10} = \frac{3}{10} + \frac{35}{10} = \frac{38}{10} = \frac{19}{5}$$

となります．

(2) まず，X の値に応じて X^2 の値がどのようになるかを確認します．また，それぞれの値をとる確率もあわせてまとめると，次のようになります．

X	1	5
X^2	1	25
確率	3/10	7/10

これを使って，X^2 の期待値 $E(X^2)$ を求めます．期待値は，「値 $\times$ 確率」をそれぞれ計算し，それらを足し合わせて求めるので，

$$E(X^2) = 1 \times \frac{3}{10} + 25 \times \frac{7}{10} = \frac{3}{10} + \frac{175}{10} = \frac{178}{10} = \frac{89}{5}$$

となります．

(3) 分散は，次の公式

$$V(X) = E(X^2) - \{E(X)\}^2$$

を用いて求めます．これまでに求めた値を代入すると，

$$V(X) = \frac{89}{5} - \left(\frac{19}{5}\right)^2 = \frac{89}{5} - \frac{361}{25} = \frac{445}{25} - \frac{361}{25} = \frac{84}{25}$$

となります．

問 3

(1) 確率変数 X は，くじに書かれている

$$0,\ 200,\ 300,\ 400,\ 600$$

のいずれかの値をとり，どの値も同じ確率 1/5 で選ばれます．よって，X の確率分布は次のとおりです．

x	0	200	300	400	600
$P(X=x)$	1/5	1/5	1/5	1/5	1/5

まず，X の値に応じて X^2 の値がどのようになるかを確認します．また，それぞれの値をとる確率もあわせてまとめると，次のようになります．

X	0	200	300	400	600
X^2	0	40000	90000	160000	360000
確率	1/5	1/5	1/5	1/5	1/5

期待値は，「値 × 確率」をそれぞれ計算し，それらを足し合わせて求めるので，X の期待値は

$$E(X) = 0 \times \frac{1}{5} + 200 \times \frac{1}{5} + 300 \times \frac{1}{5} + 400 \times \frac{1}{5} + 600 \times \frac{1}{5} = 300$$

となり，X^2 の期待値は

$$E(X^2) = 0 \times \frac{1}{5} + 40000 \times \frac{1}{5} + 90000 \times \frac{1}{5} + 160000 \times \frac{1}{5} + 360000 \times \frac{1}{5} = 130000$$

となります.

したがって，X の分散は

$$V(X) = E(X^2) - \{E(X)\}^2 = 130000 - 300^2 = 130000 - 90000 = 40000$$

となります.

標準偏差は，分散の（負でない）平方根なので，X の標準偏差は

$$\sqrt{V(X)} = \sqrt{40000} = 200$$

となります.

(2) 分散の性質

$$V(aX + b) = a^2\, V(X)$$

を用いると，Y の分散は

$$V(Y) = V\left(\frac{5}{4}X + 10\right) = \left(\frac{5}{4}\right)^2 V(X) = \frac{25}{16} \times 40000 = 62500$$

となります.

標準偏差は，分散の（負でない）平方根なので，Y の標準偏差は

$$\sqrt{V(Y)} = \sqrt{62500} = 250$$

となります.

問 4

(1) ○： 離散型確率分布では，とりうるすべての値の確率の和は 1 になるという性質があります.

(2) ×： 確率変数を a 倍して b を足したときの期待値は，「a の 2 乗倍」ではなく，a 倍されてさらに b が加わります.

(3) ×： 確率変数を a 倍すると，分散は a 倍ではなく a^2 倍になります.

(4) ○： 確率変数に一定の数を足すと，期待値もその分だけ増えます.

(5) ×： 確率変数に一定の数を足しても，分散は変わりません.

第 12 章　二項分布から正規分布へ

12.1　二項分布の確率計算

問題 12.1

成功する確率が 1/4 の作業を 4 回行ったときに成功した回数を表す確率変数 X の値ごとに，並びの数とその確率を調べます．

1 回の作業で成功する確率は 1/4 なので，失敗する確率は $(1-1/4=)\,3/4$ です．

ここで，「並び」とは，4 回の結果を「1 回目」「2 回目」...「4 回目」の順に並べたものを指します．たとえば「成功・失敗・失敗・成功」のように，4 つの結果を順番つきで書いたものです．

- **$X=0$ の場合**

  「4 回すべて失敗」の確率は

  $$\left(\frac{3}{4}\right)^4=\frac{81}{256}$$

  です．失敗が 4 回になる並びは

  $$\binom{4}{0}=1$$

  通りあり，同じ計算で求められる並びがそれだけあるということなので，$X=0$ となる確率は，次になります．

  $$\binom{4}{0}\times\left(\frac{3}{4}\right)^4=1\times\frac{81}{256}=\frac{81}{256}$$

- **$X=1$ の場合**

  たとえば「1 回目だけ成功」の確率は

  $$\frac{1}{4}\times\left(\frac{3}{4}\right)^3=\frac{27}{256}$$

  です．成功する回の選び方は

  $$\binom{4}{1}=4$$

  通りあり，同じ計算で求められる並びがそれだけあるということなので，$X=1$ となる確率は，次になります．

  $$\binom{4}{1}\times\frac{1}{4}\times\left(\frac{3}{4}\right)^3=4\times\frac{27}{256}=\frac{108}{256}$$

- **$X=2$ の場合**

  たとえば「1 回目と 2 回目だけ成功」の確率は

  $$\left(\frac{1}{4}\right)^2\times\left(\frac{3}{4}\right)^2$$

です．成功する回の選び方は

$$\binom{4}{2} = \frac{4 \times 3}{2 \times 1} = 6$$

通りあり，同じ計算で求められる並びがそれだけあるということなので，$X = 2$ となる確率は，次になります．

$$\binom{4}{2} \times \left(\frac{1}{4}\right)^2 \times \left(\frac{3}{4}\right)^2 = 6 \times \frac{9}{256} = \frac{54}{256}$$

- **$X = 3$ の場合**

  たとえば「1〜3 回目が成功」の確率は

$$\left(\frac{1}{4}\right)^3 \times \frac{3}{4}$$

です．

成功する回の選び方は

$$\binom{4}{3} = \frac{4 \times 3 \times 2}{3 \times 2 \times 1} = \frac{4}{1} = 4$$

通りあり，同じ計算で求められる並びがそれだけあるということなので，$X = 3$ となる確率は，次になります．

$$\binom{4}{3} \times \left(\frac{1}{4}\right)^3 \times \frac{3}{4} = 4 \times \frac{3}{256} = \frac{12}{256}$$

- **$X = 4$ の場合**

  「4 回すべて成功」の確率は

$$\left(\frac{1}{4}\right)^4 = \frac{1}{256}$$

です．成功する回の選び方は

$$\binom{4}{4} = 1$$

通りあり，同じ計算で求められる並びがそれだけあるということなので，$X = 4$ となる確率は，次になります．

$$\binom{4}{4} \times \left(\frac{1}{4}\right)^4 = 1 \times \frac{1}{256} = \frac{1}{256}$$

したがって，確率分布表は次のようになります．

x	0	1	2	3	4
$P(X = x)$	$\frac{81}{256}$	$\frac{108}{256}$	$\frac{54}{256}$	$\frac{12}{256}$	$\frac{1}{256}$

問題 12.2

X が二項分布 $B(n, p)$ に従う場合，X の期待値は

$$E(X) = np$$

という簡単な形で求めることができます.

(1) X は二項分布 $B(10,\,2/5)$ に従います．公式より，その期待値は

$$E(X) = 10 \times \frac{2}{5} = 4$$

となります.

(2) 当たりの確率は

$$p = \frac{3}{3+1} = \frac{3}{4}$$

なので，X は二項分布 $B(8,\,3/4)$ に従います．公式より，その期待値は

$$E(X) = 8 \times \frac{3}{4} = 6$$

となります.

(3) 赤玉が出る確率は

$$p = \frac{3}{3+6+11} = \frac{3}{20}$$

なので，X は二項分布 $B(15,\,3/20)$ に従います．公式より，その期待値は

$$E(X) = 15 \times \frac{3}{20} = \frac{45}{20} = \frac{9}{4}$$

となります.

問題 12.3

X が二項分布 $B(n,p)$ に従う場合，X の分散は

$$V(X) = np(1-p)$$

という簡単な形で求めることができます.

(1) X は二項分布 $B(50,\,1/5)$ に従います．公式より，その分散は

$$V(X) = 50 \times \frac{1}{5} \times \left(1 - \frac{1}{5}\right) = 50 \times \frac{1}{5} \times \frac{4}{5} = 8$$

となります.

(2) X は二項分布 $B(20,\,1/2)$ に従います．公式より，その分散は

$$V(X) = 20 \times \frac{1}{2} \times \left(1 - \frac{1}{2}\right) = 20 \times \frac{1}{2} \times \frac{1}{2} = 5$$

となります.

(3) 1〜200 の自然数が書かれたカードは全部で 200 枚あり，このうち 10 の倍数は 20 枚です．したがって，1 回取り出す操作で 10 の倍数が出る確率は

$$p = \frac{20}{200} = \frac{1}{10}$$

となります．よって，X は二項分布 $B(30,\,1/10)$ に従います．公式より，その分散は

$$V(X) = 30 \times \frac{1}{10} \times \left(1 - \frac{1}{10}\right) = 30 \times \frac{1}{10} \times \frac{9}{10} = \frac{27}{10}$$

となります.

12.1 の確認テスト

問 1

サイコロで 4 の目が出る確率は

$$p = \frac{1}{6}$$

なので，X は二項分布 $B(12, 1/6)$ に従います．公式より，その期待値は

$$E(X) = 12 \times \frac{1}{6} = 2$$

となります.

問 2

X は二項分布 $B(25, 4/5)$ に従います．公式より，その分散は

$$V(X) = 25 \times \frac{4}{5} \times \left(1 - \frac{4}{5}\right) = 25 \times \frac{4}{5} \times \frac{1}{5} = 4$$

となります.

標準偏差は，分散の（負でない）平方根なので，X の標準偏差は

$$\sqrt{V(X)} = \sqrt{4} = 2$$

となります.

問 3

- 赤玉が出る確率は

  $$p = \frac{3}{3+5+12} = \frac{3}{20}$$

  なので，X は二項分布 $B(40, 3/20)$ に従います．公式より，その期待値は

  $$E(X) = 40 \times \frac{3}{20} = \frac{120}{20} = 6$$

  その分散は

  $$V(X) = 40 \times \frac{3}{20} \times \left(1 - \frac{3}{20}\right) = 40 \times \frac{3}{20} \times \frac{17}{20} = \frac{51}{10}$$

  となります.
- 青玉が出る確率は

  $$p = \frac{5}{3+5+12} = \frac{5}{20} = \frac{1}{4}$$

  なので，Y は二項分布 $B(40, 1/4)$ に従います．公式より，その期待値は

  $$E(Y) = 40 \times \frac{1}{4} = 10$$

  その分散は

  $$V(Y) = 40 \times \frac{1}{4} \times \left(1 - \frac{1}{4}\right) = 40 \times \frac{1}{4} \times \frac{3}{4} = \frac{15}{2}$$

となります.

- 赤玉または青玉が出る確率は

$$p = \frac{3+5}{3+5+12} = \frac{8}{20} = \frac{2}{5}$$

なので，$X+Y$ は二項分布 $B(40,\, 2/5)$ に従います．公式より，その期待値は

$$E(X+Y) = 40 \times \frac{2}{5} = 16$$

その分散は

$$V(X+Y) = 40 \times \frac{2}{5} \times \left(1 - \frac{2}{5}\right) = 40 \times \frac{2}{5} \times \frac{3}{5} = \frac{48}{5}$$

となります.

問 4

最初の送信で成功する確率は $9/10$ であり，最初は失敗して（確率 $1/10$）再送で成功する確率は $8/10$ なので，1 通のメールが「成功」となる確率は

$$p = \frac{9}{10} + \left(1 - \frac{9}{10}\right) \times \frac{8}{10} = \frac{9}{10} + \frac{8}{100} = \frac{49}{50}$$

となります．したがって，X は二項分布 $B(50,\, 49/50)$ に従います.

公式より，その期待値は

$$E(X) = 50 \times \frac{49}{50} = 49$$

分散は

$$V(X) = 50 \times \frac{49}{50} \times \left(1 - \frac{49}{50}\right) = 50 \times \frac{49}{50} \times \frac{1}{50} = \frac{49}{50}$$

となります.

12.2　正規分布の性質

問題 12.4

(1) **正解：(a)，(b)，(c)，(d)**

連続型確率変数では，確率はグラフの下の「面積」として表されます．区間を分けると，その確率も面積の和や差として考えることができます.

連続型では 1 点だけに対応する面積は 0 であるため，区間の端点を含むかどうかは確率に影響しません．したがって，以下では端点のちがいを気にせず議論します.

(a) $a < b < c$ のとき，区間 (a, c) の部分は (a, b) と (b, c) に分けられます．したがって，

$$P(a < X < c) = P(a < X < b) + P(b < X < c)$$

が成り立ちます.

(b) $P(X < b)$ は，$(-\infty, a)$ と (a, b) に分けられます.

$$P(X < b) = P(X < a) + P(a < X < b)$$

が成り立ちます.

(c) (b) の関係式を変形すると,

$$P(a < X < b) = P(X < b) - P(X < a)$$

となります．右辺は,「左側の面積 $P(X < b)$ から，さらに左側の面積 $P(X < a)$ を引いたもの」として理解できます.

(d) 全事象が起こる確率は 1 であるため,

$$1 = P(X < a) + P(a < X < b) + P(X > b)$$

が成り立ちます．これを $P(a < X < b)$ について解くと,

$$P(a < X < b) = 1 - P(X < a) - P(X > b)$$

となり，この等式も正しいです.

(e) $P(X < a)$ と $P(X > b)$ は，それぞれ左側と右側の「外側」の部分を表します．これらを足すと，全体のうち真ん中の区間 (a, b) を除いた部分になるので,

$$P(a < X < b) = P(X < a) + P(X > b)$$

とはなりません．よってこれは誤りです.

(2) この分布では，100 から 160 までのあいだでグラフの高さが一定になっています．横の長さは

$$160 - 100 = 60$$

です．連続型では，グラフの下の面積が全体で 1 になるように表します．したがって，高さは

$$\frac{1}{60}$$

となります.

(a) $P(T > 150)$ は区間 $(150, 160)$ に対応する部分の面積です．この区間の長さは

$$160 - 150 = 10$$

なので,

$$P(T > 150) = \frac{1}{60} \times 10 = \frac{1}{6}$$

となります.

(b) $P(120 < T < 145)$ は区間 $(120, 145)$ に対応する部分の面積です．この区間の長さは

$$145 - 120 = 25$$

なので,

$$P(120 < T < 145) = \frac{1}{60} \times 25 = \frac{5}{12}$$

となります.

(c) $|T - 130| < 10$ は，まず

$$-10 < T - 130 < 10$$

と書けます．これを T について解くと，

$$120 < T < 140$$

となります．この区間の長さは

$$140 - 120 = 20$$

なので，

$$P(|T - 130| < 10) = \frac{1}{60} \times 20 = \frac{1}{3}$$

となります．

(d) $T \leq 110$ または $T \geq 155$ に対応する部分は，左端から 110 までの長さと，155 から右端までの長さを合わせたものになります．

左側の長さは

$$110 - 100 = 10$$

右側の長さは

$$160 - 155 = 5$$

です．2 つを合わせると

$$10 + 5 = 15$$

なので，

$$P(T \leq 110 \text{ または } T \geq 155) = \frac{1}{60} \times 15 = \frac{1}{4}$$

となります．

問題 12.5

(1) **正解：(a)，(c)**

- **(a) について**

  正規分布は左右対称の山形で，中心付近の値が最も多くなります．期待値 60 点が分布の中心なので，この説明は正しいです．

- **(b) について**

  「0 点から 100 点まで一様にばらついている」は一様分布の説明であり，正規分布の特徴とは一致しません．したがって誤りです．

- **(c) について**

  標準偏差が 10 点なので，期待値を中心とした「±1 標準偏差」の範囲は

  $$60 - 10 = 50, \qquad 60 + 10 = 70$$

  より，50 点から 70 点のあいだです．この範囲に全体のおよそ 68% が含まれます．よってこの説明は正しいです．

(2) 期待値 60 点，標準偏差 10 点なので，点数 x の z 値は

$$z = \frac{x - 60}{10}$$

で求められます.

(a) 62 点のとき

$$z = \frac{62 - 60}{10} = \frac{2}{10} = 0.2$$

(b) 58 点のとき

$$z = \frac{58 - 60}{10} = \frac{-2}{10} = -0.2$$

(c) 63 点のとき

$$z = \frac{63 - 60}{10} = \frac{3}{10} = 0.3$$

(d) 57 点のとき

$$z = \frac{57 - 60}{10} = \frac{-3}{10} = -0.3$$

12.2 の確認テスト

問 1

(1) **正解：(c)**

図 A は，左側がすべて塗られています．これは，「X がある境界値より小さい値をとる場合」を表しており，次のような形になります.

$$P(X < -2)$$

図 B は，中央の区間が塗られています．これは，「X が 2 つの境界値のあいだに入る場合」を表しており，次のような形になります.

$$P(-1 < X < 1)$$

図 C は，右側がすべて塗られています．これは，「X がある境界値より大きい値をとる場合」を表しており，次のような形になります.

$$P(X > 2)$$

以上より，適切な選択肢は (c) です.

(2) **正解：(a)，(b)，(d)，(e)，(g)**

連続型確率変数では，確率はグラフの下の「面積」として表されます．区間を分けると，その確率も面積の和や差として考えることができます.

連続型では 1 点だけに対応する面積は 0 であるため，区間の端点を含むかどうかは確率に影響しません．したがって，以下では端点のちがいを気にせず議論します.

(a) 右辺は，左側の面積 $P(X < b)$ から，さらに左側の面積 $P(X \leq a)$ を引いたものです．この差が区間 (a, b) に対応する面積となるため，正しいです.

(b) 連続型では 1 点だけをとる区間の長さは 0 であるため，どのような値 b に対しても

$$P(X = b) = 0$$

となります．

(c) 右辺は，全体から左側の面積 $P(X < b)$ を引いたもので，$X \geq b$ に対応します．左辺の $P(X > a)$ は $X > a$ に対応する面積であり，$X \geq b$ とは区間が一致しません．したがって誤りです．

(d) 全体の面積は 1 であり，それは左側の面積 $P(X \leq a)$，真ん中の区間 (a, c) の面積 $P(a < X < c)$，右側の面積 $P(X \geq c)$ に分かれます．したがって

$$1 = P(X \leq a) + P(a < X < c) + P(X \geq c)$$

が成り立ち，これを $P(a < X < c)$ について解くと与式になります．正しいです．

(e) 面積 $P(X > b)$ の中には，区間 (b, c) に対応する面積 $P(b < X < c)$ が含まれています．この部分を引くと，右側の区間 (c, ∞) に対応する面積 $P(X > c)$ が残るため，正しいです．

(f) 面積 $P(X < b)$ と $P(X > a)$ は，区間 (a, b) に対応する部分をどちらも含んでいます．そのため，これらを足すと区間 (a, b) の部分が二重に数えられ，$P(a < X < b)$ とは一致しません．したがって，この等式は誤りです．

(g) 全体の面積は 1 であり，それは左側の面積 $P(X < a)$ と右側の面積 $P(X \geq a)$ に分かれます．したがって

$$1 = P(X < a) + P(X \geq a)$$

が成り立ち，これを $P(X < a)$ について解くと与式になります．正しいです．

問 2

(1) この分布では，0.2 秒から 1.4 秒までのあいだでグラフの高さが一定になっています．横の長さは

$$1.4 - 0.2 = 1.2$$

です．連続型では，グラフの下の面積が全体で 1 になるように表します．したがって，高さは

$$\frac{1}{1.2}$$

となります．

(a) $P(Y > 1.0)$ は区間 $(1.0, 1.4)$ に対応する部分の面積です．この区間の長さは

$$1.4 - 1.0 = 0.4$$

なので，

$$P(Y > 1.0) = \frac{1}{1.2} \times 0.4 = \frac{1}{3}$$

となります．

(b) $P(0.5 < Y < 1.2)$ は区間 $(0.5, 1.2)$ に対応する部分の面積です．この区間の長さは

$$1.2 - 0.5 = 0.7$$

なので，

$$P(0.5 < Y < 1.2) = \frac{1}{1.2} \times 0.7 = \frac{7}{12}$$

となります.

(c) $P(Y \leq 0.3)$ は区間 $(0.2, 0.3)$ に対応する部分の面積です．この区間の長さは

$$0.3 - 0.2 = 0.1$$

なので,

$$P(Y \leq 0.3) = \frac{1}{1.2} \times 0.1 = \frac{1}{12}$$

となります.

(2) 期待値 12 mm，標準偏差 2 mm なので，値 x の z 値は

$$z = \frac{x - 12}{2}$$

で求められます.

(a) 10 mm の z 値は

$$z = \frac{10 - 12}{2} = \frac{-2}{2} = -1$$

となります.

(b) 12 mm の z 値は

$$z = \frac{12 - 12}{2} = \frac{0}{2} = 0$$

となります.

(c) 14 mm の z 値は

$$z = \frac{14 - 12}{2} = \frac{2}{2} = 1$$

となります.

(d) 15 mm の z 値は

$$z = \frac{15 - 12}{2} = \frac{3}{2} = 1.5$$

となります.

第 13 章　正規分布

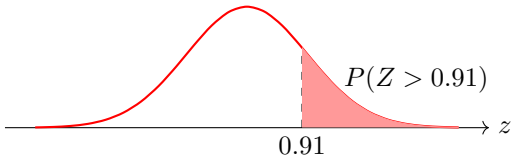
13.1　正規分布の確率計算

問題 13.1

(1) $\boldsymbol{P(Z > 0.91)}$

標準正規分布表（上側確率）には，Z の 0 以上の値 u に対する上側確率 $P(Z > u)$ が載っています．行「0.9」，列「.01」から

$$\boldsymbol{P(Z > 0.91) = 0.1814}$$



(2) $\boldsymbol{P(Z < -0.62)}$

標準正規分布は原点を中心に左右対称です．そのため，

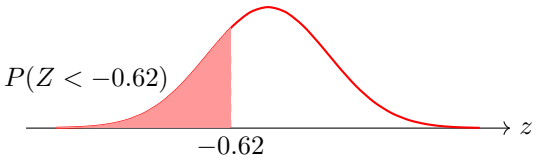
$$\boldsymbol{P(Z < -0.62) = P(Z > 0.62)}$$

となります．ここで，標準正規分布表から

$$P(Z > 0.62) = 0.2676$$

であることがわかるので

$$P(Z < -0.62) = 0.2676$$



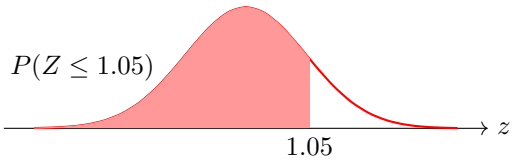
(3) $\boldsymbol{P(Z \leq 1.05)}$

標準正規分布表（上側確率）には，Z の 0 以上の値 u に対する上側確率 $P(Z > u)$ が載っています．そのため，まず

$$P(Z > 1.05) = 0.1469$$

を読み取ります．$P(Z \leq 1.05)$ は $P(Z > 1.05)$ の補集合なので，

$$\boldsymbol{P(Z \leq 1.05) = 1 - P(Z > 1.05)} = 1 - 0.1469 = 0.8531$$



(4) $\boldsymbol{P(Z > -0.48)}$

標準正規分布は左右対称なので，

$$\boldsymbol{P(Z > -0.48) = P(Z < 0.48)}$$

ここで，正規分布表から

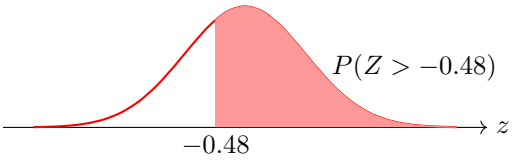
$$P(Z > 0.48) = 0.3156$$

であることがわかるので，

$$P(Z < 0.48) = 1 - P(Z > 0.48) = 1 - 0.3156 = 0.6844$$

したがって

$$P(Z > -0.48) = 0.6844$$



(5) $\boldsymbol{P(-1.10 \leq Z \leq 0.55)}$

まず，

$$\boldsymbol{P(Z > -1.10) = 1 - P(Z < -1.10)}$$

となります．ここで，標準正規分布は左右対称なので，

$$\boldsymbol{P(Z < -1.10) = P(Z > 1.10)}$$

となります．標準正規分布表から

$$P(Z > 1.10) = 0.1357$$

であることがわかるので，

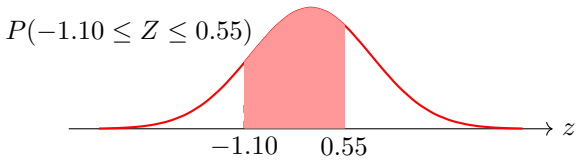
$$\begin{aligned}\boldsymbol{P(Z > -1.10)} &= \boldsymbol{1 - P(Z < -1.10)} \\ &= \boldsymbol{1 - P(Z > 1.10)} = 1 - 0.1357 = 0.8643\end{aligned}$$

また，標準正規分布表から

$$P(Z > 0.55) = 0.2912$$

であることがわかるので，

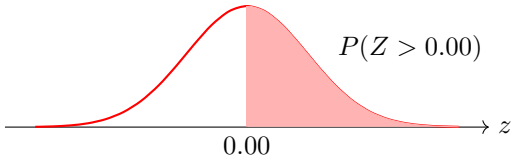
$$\begin{aligned} \boldsymbol{P(-1.10 \leq Z \leq 0.55) = P(Z > -1.10) - P(Z > 0.55)} \\ &= 0.8643 - 0.2912 \\ &= 0.5731 \end{aligned}$$



問題 13.2

(1) 点数が 70 点以上である確率 $P(X \geq 70)$ を標準化すると，

$$P(X \geq 70) = P\left(Z \geq \frac{70-70}{8}\right) = P(Z \geq 0.00)$$



標準正規分布表より

$$P(Z > 0.00) = 0.5000$$

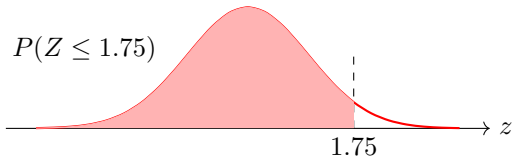
よって，求める確率は

$$P(X \geq 70) = 0.5000$$

となります．およそ 50% 程度の生徒が 70 点以上であることがわかります．

(2) 点数が 84 点以下である確率 $P(X \leq 84)$ を標準化すると，

$$P(X \leq 84) = P\left(Z \leq \frac{84-70}{8}\right) = P(Z \leq 1.75)$$



標準正規分布表より

$$P(Z > 1.75) = 0.0401$$

よって，

$$P(Z \leq 1.75) = 1 - P(Z > 1.75) = 1 - 0.0401 = 0.9599$$

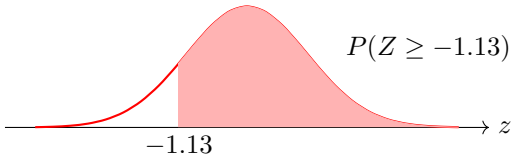
したがって，求める確率は

$$P(X \leq 84) = 0.9599$$

となります．およそ 96% 程度の生徒が 84 点以下であることがわかります．

(3) 点数が 61 点以上である確率 $P(X \geq 61)$ を標準化すると，

$$P(X \geq 61) = P\left(Z \geq \frac{61 - 70}{8}\right) = P(Z \geq -1.125) \approx P(Z \geq -1.13)$$



標準正規分布は左右対称なので，

$$P(Z \geq -1.13) = P(Z \leq 1.13) = 1 - P(Z > 1.13) = 1 - 0.1292 = 0.8708$$

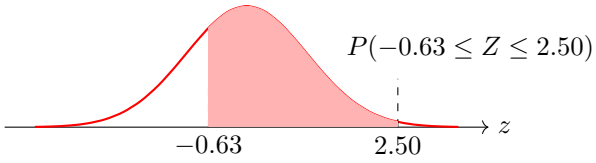
よって，求める確率は

$$P(X \geq 61) = 0.8708$$

となります．およそ 87% 程度の生徒が 61 点以上であることがわかります．

(4) 点数が 65 点以上 90 点以下である確率 $P(65 \leq X \leq 90)$ を標準化すると，

$$P(65 \leq X \leq 90) = P\left(\frac{65 - 70}{8} \leq Z \leq \frac{90 - 70}{8}\right) = P(-0.625 \leq Z \leq 2.50) \approx P(-0.63 \leq Z \leq 2.50)$$



標準正規分布は左右対称なので，

$$\begin{aligned} P(Z > -0.63) &= 1 - P(Z < -0.63) \\ &= 1 - P(Z > 0.63) = 1 - 0.2643 = 0.7357 \end{aligned}$$

また，

$$P(Z > 2.50) = 0.0062$$

なので，

$$\begin{aligned} P(-0.63 \leq Z \leq 2.50) &= P(Z > -0.63) - P(Z > 2.50) \\ &= 0.7357 - 0.0062 = 0.7295 \end{aligned}$$

よって，求める確率は

$$P(65 \leq X \leq 90) = 0.7295$$

となります．およそ 73% 程度の生徒がこの範囲に収まることがわかります．

問題 13.3

一般に，1 回の操作で成功する確率を p とします．この操作を n 回くり返し，各回の結果は互いに独立であるとします．成功した回数を確率変数 X とすると，X は二項分布 $B(n,p)$ に従います．

ここでは，1 枚のコインを 100 回投げたときの，表が出た回数を表す確率変数を X として考えます．コイン 1 回で表が出る確率は $1/2$ とみなせるので，X は二項分布 $B(100,1/2)$ に従います．

二項分布 $B(n,p)$ に従う確率変数 X の期待値 $E(X)$ と分散 $V(X)$ は

$$E(X) = np, \qquad V(X) = np(1-p)$$

で求められます（例題 12.2，例題 12.3 参照）．

この問題では，$n = 100$，$p = 1/2$ なので，

$$E(X) = 100 \times \frac{1}{2} = 50, \qquad V(X) = 100 \times \frac{1}{2} \times \frac{1}{2} = 25$$

となります．標準偏差は分散の（負でない）平方根なので，

$$\sqrt{25} = 5$$

です．

二項分布 $B(n,p)$ は，試行回数 n が大きくなるにつれて，期待値 np，分散 $np(1-p)$ の正規分布に近づくことが知られています．そのため，n が十分大きいときには，二項分布 $B(n,p)$ を正規分布 $N(np,\ np(1-p))$ で近似することがあります．

ここでは，試行回数 $n = 100$ が十分大きいとして，X を期待値 50，標準偏差 5 の正規分布で近似して扱うことにします．

そこで，

$$Z = \frac{X - 50}{5} \qquad \left(Z = \frac{X - \text{期待値}}{\text{標準偏差}} \right)$$

とおき，標準正規分布表を使って確率を求めます．

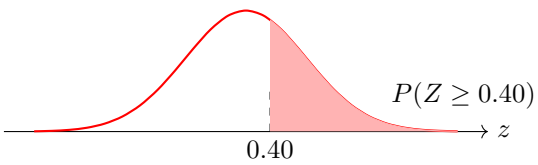
求めたい確率は「52 回以上」となる確率

$$P(X \geq 52)$$

です．標準化すると，

$$P(X \geq 52) = P\left(Z \geq \frac{52 - 50}{5} \right) = P(Z \geq 0.40)$$

となります．



ここで，標準正規分布表より

$$P(Z > 0.40) = 0.3446$$

なので，

$$P(X \geq 52) = 0.3446$$

となります．よって，正規分布による近似では，表が 52 回以上となる確率は，約 0.34 です．

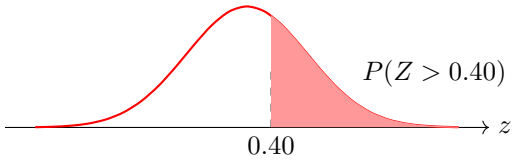
13.1 の確認テスト

問 1

(1) $\boldsymbol{P(Z > 0.40)}$

標準正規分布表（上側確率）には，Z の 0 以上の値 u に対する上側確率 $P(Z > u)$ が載っています．行「0.4」，列「.00」から

$$\boldsymbol{P(Z > 0.40) = 0.3446}$$



(2) $\boldsymbol{P(Z < -1.11)}$

標準正規分布は原点を中心に左右対称です．そのため，

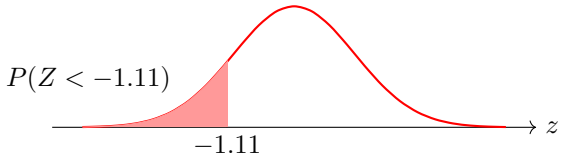
$$\boldsymbol{P(Z < -1.11) = P(Z > 1.11)}$$

となります．ここで，標準正規分布表から

$$P(Z > 1.11) = 0.1335$$

であることがわかるので，

$$P(Z < -1.11) = 0.1335$$



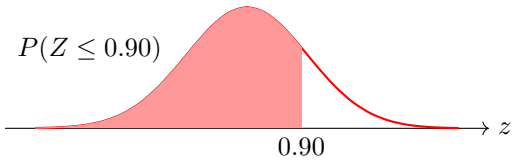
(3) $\boldsymbol{P(Z \leq 0.90)}$

標準正規分布表（上側確率）には，Z の 0 以上の値 u に対する上側確率 $P(Z > u)$ が載っています．行「0.9」，列「.00」から

$$P(Z > 0.90) = 0.1841$$

を読み取ります．$P(Z \leq 0.90)$ は $P(Z > 0.90)$ の補集合なので，

$$\boldsymbol{P(Z \leq 0.90) = 1 - P(Z > 0.90)} = 1 - 0.1841 = 0.8159$$



(4) $\boldsymbol{P(Z \geq -0.32)}$

標準正規分布は左右対称なので，

$$\boldsymbol{P(Z \geq -0.32) = P(Z \leq 0.32)}$$

標準正規分布表から

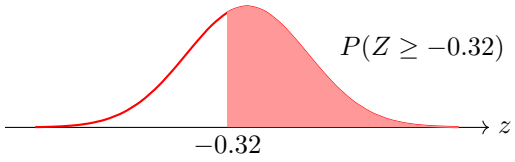
$$P(Z > 0.32) = 0.3745$$

であることがわかるので，

$$\boldsymbol{P(Z \leq 0.32) = 1 - P(Z > 0.32)} = 1 - 0.3745 = 0.6255$$

したがって，

$$P(Z \geq -0.32) = 0.6255$$



(5) $\boldsymbol{P(-0.22 < Z < 1.59)}$

まず，

$$\boldsymbol{P(Z > -0.22) = 1 - P(Z < -0.22)}$$

となります．標準正規分布は左右対称なので，

$$\boldsymbol{P(Z < -0.22) = P(Z > 0.22)}$$

となります．標準正規分布表から

$$P(Z > 0.22) = 0.4129$$

であることがわかるので，

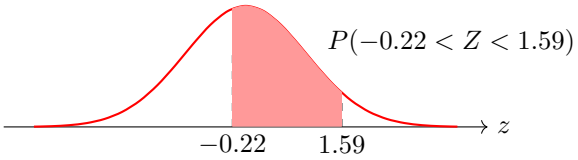
$$\begin{aligned}\boldsymbol{P(Z > -0.22)} &= \boldsymbol{1 - P(Z < -0.22)} \\ &= \boldsymbol{1 - P(Z > 0.22)} = 1 - 0.4129 = 0.5871\end{aligned}$$

また，標準正規分布表より

$$P(Z > 1.59) = 0.0559$$

であることがわかるので，

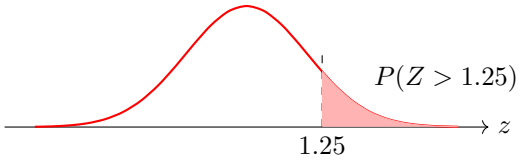
$$\begin{aligned} \boldsymbol{P(-0.22 < Z < 1.59) = P(Z > -0.22) - P(Z > 1.59)} \\ &= 0.5871 - 0.0559 \\ &= 0.5312 \end{aligned}$$



問 2

(1) 重量が 105 g 以上である確率 $P(X \geq 105)$ を標準化すると，

$$P(X \geq 105) = P\left(Z \geq \frac{105 - 100}{4}\right) = P(Z \geq 1.25)$$



標準正規分布表より

$$P(Z > 1.25) = 0.1056$$

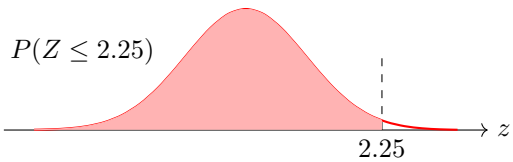
よって，求める確率は

$$P(X \geq 105) = 0.1056$$

となります．およそ 11% 程度の製品が 105 g 以上であることがわかります．

(2) 重量が 109 g 以下である確率 $P(X \leq 109)$ を標準化すると，

$$P(X \leq 109) = P\left(Z \leq \frac{109 - 100}{4}\right) = P(Z \leq 2.25)$$



標準正規分布表より

$$P(Z > 2.25) = 0.0122$$

なので，

$$P(Z \leq 2.25) = 1 - P(Z > 2.25) = 1 - 0.0122 = 0.9878$$

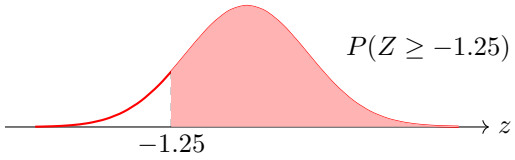
したがって，求める確率は

$$P(X \leq 109) = 0.9878$$

となります．およそ 99% 程度の製品が 109 g 以下であることがわかります．

(3) 重量が 95 g 以上である確率 $P(X \geq 95)$ を標準化すると，

$$P(X \geq 95) = P\left(Z \geq \frac{95-100}{4}\right) = P(Z \geq -1.25)$$



標準正規分布は左右対称なので，

$$P(Z \geq -1.25) = P(Z \leq 1.25) = 1 - P(Z > 1.25) = 1 - 0.1056 = 0.8944$$

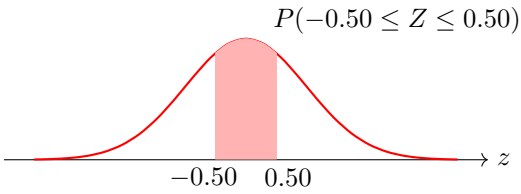
よって，求める確率は

$$P(X \geq 95) = 0.8944$$

となります．およそ 89% 程度の製品が 95 g 以上であることがわかります．

(4) 重量が 98 g 以上 102 g 以下である確率 $P(98 \leq X \leq 102)$ を標準化すると，

$$P(98 \leq X \leq 102) = P\left(\frac{98-100}{4} \leq Z \leq \frac{102-100}{4}\right) = P(-0.50 \leq Z \leq 0.50)$$



標準正規分布は左右対称なので，

$$\begin{aligned} P(Z > -0.50) &= 1 - P(Z < -0.50) \\ &= 1 - P(Z > 0.50) = 1 - 0.3085 = 0.6915 \end{aligned}$$

また，

$$P(Z > 0.50) = 0.3085$$

したがって，

$$\begin{aligned} P(-0.50 \leq Z \leq 0.50) &= P(Z > -0.50) - P(Z > 0.50) \\ &= 0.6915 - 0.3085 = 0.3830 \end{aligned}$$

よって，求める確率は

$$P(98 \leq X \leq 102) = 0.3830$$

となります．およそ 38% 程度の製品がこの範囲に収まることがわかります．

問 3

赤玉が 2 個，白玉が 3 個入っているので，1 回の操作で赤玉が出る確率は

$$p = \frac{2}{5}$$

となります．150 回の操作のうち，赤玉が出た回数を表す確率変数を X とすると，X は二項分布 $B(150, 2/5)$ に従います．

二項分布 $B(n, p)$ に従う確率変数 X の期待値 $E(X)$ と分散 $V(X)$ は

$$E(X) = np, \qquad V(X) = np(1-p)$$

で求められます．

この問題では，$n = 150$，$p = 2/5$ なので，

$$E(X) = 150 \times \frac{2}{5} = 60, \qquad V(X) = 150 \times \frac{2}{5} \times \frac{3}{5} = 36$$

となります．標準偏差は分散の（負でない）平方根なので，

$$\sqrt{36} = 6$$

です．

二項分布 $B(n, p)$ は，試行回数 n が大きくなるにつれて，期待値 np，分散 $np(1-p)$ の正規分布に近づくことが知られています．そのため，n が十分大きいときには，二項分布 $B(n, p)$ を正規分布 $N(np,\ np(1-p))$ で近似することがあります．

ここでは，試行回数 $n = 150$ が十分大きいとして，X を期待値 60，標準偏差 6 の正規分布で近似して扱うことにします．

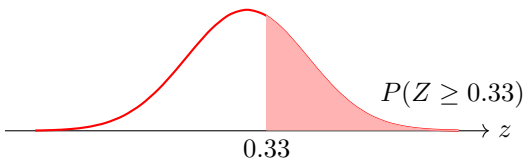
そこで，

$$Z = \frac{X - 60}{6} \qquad \left(Z = \frac{X - \text{期待値}}{\text{標準偏差}} \right)$$

とおき，標準正規分布表を使って確率を求めます．求めたい確率は「62 回以上」となる確率 $P(X \geq 62)$ です．標準化すると，

$$P(X \geq 62) = P\left(Z \geq \frac{62 - 60}{6} \right) = P(Z \geq 0.333\ldots) \approx P(Z \geq 0.33)$$

となります．



ここで，標準正規分布表より

$$P(Z > 0.33) = 0.3707$$

なので，

$$P(X \geq 62) = 0.3707$$

となります．つまり，赤玉が 62 回以上出る確率は，約 0.37 です．

問 4

1 回サイコロを振ったとき，「5 以上の目」が出る確率は

$$p = \frac{2}{6} = \frac{1}{3}$$

です．サイコロを 72 回振ったとき，「5 以上の目」が出た回数を表す確率変数を X とすると，X は二項分布 $B(72, 1/3)$ に従います．

二項分布 $B(n, p)$ に従う確率変数 X の期待値 $E(X)$ と分散 $V(X)$ は

$$E(X) = np, \qquad V(X) = np(1-p)$$

で求められます．

この問題では，$n = 72$，$p = 1/3$ なので，

$$E(X) = 72 \times \frac{1}{3} = 24, \qquad V(X) = 72 \times \frac{1}{3} \times \frac{2}{3} = 16$$

となり，標準偏差は

$$\sqrt{16} = 4$$

です．

ここでは，試行回数 $n = 72$ が十分大きいとして，X を期待値 24，標準偏差 4 の正規分布で近似して扱うことにします．

そこで，

$$Z = \frac{X - 24}{4} \qquad \left(Z = \frac{X - \text{期待値}}{\text{標準偏差}} \right)$$

とおき，標準正規分布表を使って確率を求めます．

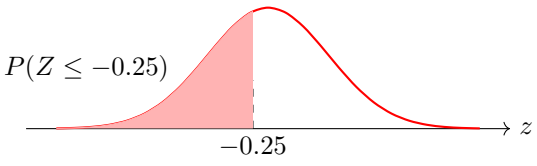
(1) 求めたい確率は

$$P(X \leq 23)$$

です．標準化すると，

$$P(X \leq 23) = P\left(Z \leq \frac{23 - 24}{4} \right) = P(Z \leq -0.25)$$

となります．



標準正規分布は左右対称なので，

$$P(Z \leq -0.25) = P(Z \geq 0.25) = P(Z > 0.25) = 0.4013$$

となります．よって，求める確率は

$$P(X \leq 23) = 0.4013$$

となります．つまり，成功する回数が 23 以下である確率は，約 0.40 です．

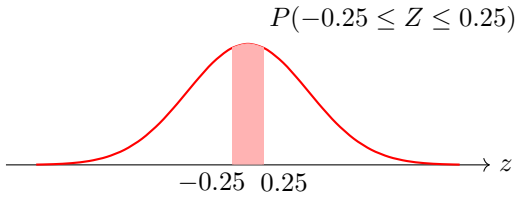
(2) 求めたい確率は

$$P(23 \leq X \leq 25)$$

です．標準化すると，

$$P(23 \leq X \leq 25) = P\left(\frac{23-24}{4} \leq Z \leq \frac{25-24}{4}\right) = P(-0.25 \leq Z \leq 0.25)$$

となります．



ここで，(1) より

$$P(Z \leq -0.25) = 0.4013$$

なので，

$$P(Z > -0.25) = 1 - P(Z \leq -0.25) = 1 - 0.4013 = 0.5987$$

となります．

また，

$$P(Z > 0.25) = 0.4013$$

なので，

$$P(-0.25 \leq Z \leq 0.25) = P(Z > -0.25) - P(Z > 0.25) = 0.5987 - 0.4013 = 0.1974$$

となります．

よって，求める確率は

$$P(23 \leq X \leq 25) = 0.1974$$

となります．つまり，成功する回数が 23 以上 25 以下である確率は，約 0.20 です．

13.2　標本平均・標本比率の性質

問題 13.4

(1) **1 日の来館者数 X_1 が 550 人以上である確率**

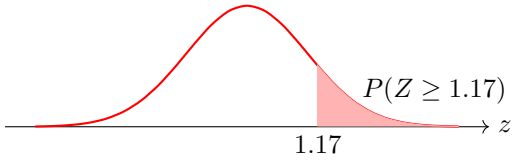
1 日の来館者数 X_1 は，期待値 480，標準偏差 60 の正規分布に従います．

正規分布の確率を求めるときには，標準化して標準正規分布の確率に読み替えます．

1 日の来館者数が 550 人以上になる確率 $P(X_1 \geq 550)$ を標準化すると，

$$P(X_1 \geq 550) = P\left(Z \geq \frac{550-480}{60}\right) = P(Z \geq 1.166\ldots) \approx P(Z \geq 1.17)$$

となります．



標準正規分布表より

$$P(Z > 1.17) = 0.1210$$

なので，

$$P(X_1 \geq 550) = 0.1210$$

となります．つまり，550 人以上となる確率は 0.1210 です．

(2) **標本平均 $\bar{X}$ が 510 人以上である確率**

各日の来館者数は期待値 480，標準偏差 60 の正規分布に従います．標本サイズは 25 なので，標本平均

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{25}}{25}$$

の期待値 $E(\bar{X})$ と分散 $V(\bar{X})$ は

$$E(\bar{X}) = 480, \qquad V(\bar{X}) = \frac{60^2}{25} = 144$$

となり，標本平均の標準偏差は

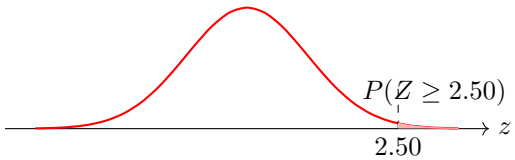
$$\sqrt{144} = 12$$

となります．つまり，標本平均 $\bar{X}$ は期待値 480，標準偏差 12 の正規分布に従います．

標本平均が 510 人以上になる確率 $P(\bar{X} \geq 510)$ を標準化すると，

$$P(\bar{X} \geq 510) = P\left(Z \geq \frac{510 - 480}{12}\right) = P(Z \geq 2.50)$$

となります．



標準正規分布表より

$$P(Z > 2.50) = 0.0062$$

なので，

$$P(\bar{X} \geq 510) = 0.0062$$

となります．つまり，標本平均が 510 人以上となる確率は 0.0062 です．

問題 13.5

(1) **標本比率 $\hat{p}$ の期待値と分散**

母比率は

$$p = 0.3$$

です.

よって，標本比率 $\hat{p} = X/84$ の期待値は,

$$E(\hat{p}) = 0.3$$

となります．つまり，標本比率の期待値は母比率（＝母集団における「メールを開封した人」の割合）と同じ 0.3 です.

また，標本比率 $\hat{p} = X/84$ の分散は,

$$V(\hat{p}) = \frac{p(1-p)}{84} = \frac{0.3 \times 0.7}{84} = 0.0025$$

となります．したがって，標本比率の標準偏差は

$$\sqrt{V(\hat{p})} = \sqrt{0.0025} = 0.05$$

となります.

(2) **標本比率 $\hat{p}$ が 0.24 以下である確率**

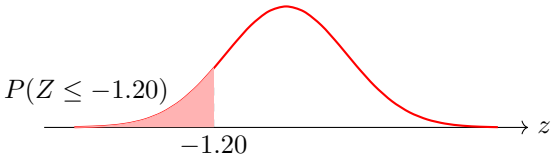
ここでは，標本サイズ $n = 84$ が十分大きいとして，標本比率 $\hat{p}$ を期待値 0.3，標準偏差 0.05 の正規分布で近似して扱うことにします.

正規分布の確率を求めるときには，標準化して標準正規分布の確率に読み替えます.

標本比率が 0.24 以下になる確率 $P(\hat{p} \leq 0.24)$ を標準化すると,

$$P(\hat{p} \leq 0.24) = P\left(Z \leq \frac{0.24 - 0.30}{0.05}\right) = P(Z \leq -1.20)$$

となります.



標準正規分布は左右対称なので,

$$P(Z \leq -1.20) = P(Z \geq 1.20) = 0.1151$$

となります．よって，求める確率は

$$P(\hat{p} \leq 0.24) = 0.1151$$

となります．つまり，標本比率が 0.24 以下となる確率は約 0.12 です.

13.2 の確認テスト

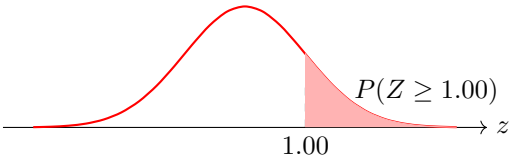
問 1

(1) **測定値 X_1 が 518 mm 以上である確率**

1 個の測定値 X_1 は，期待値 500，標準偏差 18 の正規分布に従います．
正規分布の確率を求めるときには，標準化して標準正規分布の確率に読み替えます．
測定値が 518 mm 以上になる確率 $P(X_1 \geq 518)$ を標準化すると，

$$P(X_1 \geq 518) = P\left(Z \geq \frac{518-500}{18}\right) = P(Z \geq 1.00)$$

となります．



標準正規分布表より

$$P(Z > 1.00) = 0.1587$$

なので，

$$P(X_1 \geq 518) = 0.1587$$

となります．つまり，518 mm 以上となる確率は 0.1587 です．

(2) **標本平均 $\bar{X}$ が規格を満たす確率**

各測定値は期待値 500，標準偏差 18 の正規分布に従います．測定した製品は 9 個なので，標本平均

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_9}{9}$$

の期待値 $E(\bar{X})$ と分散 $V(\bar{X})$ は

$$E(\bar{X}) = 500, \qquad V(\bar{X}) = \frac{18^2}{9} = 36$$

となり，標本平均の標準偏差は

$$\sqrt{36} = 6$$

です．したがって，標本平均 $\bar{X}$ は期待値 500，標準偏差 6 の正規分布に従います．
規格を満たすとは

$$488 \text{ mm} \leq \bar{X} \leq 512 \text{ mm}$$

であることなので，この確率を標準化すると，

$$P(488 \leq \bar{X} \leq 512) = P(-2.00 \leq Z \leq 2.00)$$

となります．

標準正規分布は左右対称なので,

$$P(Z > -2.00) = 1 - P(Z < -2.00) = 1 - P(Z > 2.00)$$

です．標準正規分布表より

$$P(Z > 2.00) = 0.0228$$

なので,

$$P(Z > -2.00) = 1 - 0.0228 = 0.9772$$

となります．また,

$$P(Z > 2.00) = 0.0228$$

なので,

$$P(-2.00 \leq Z \leq 2.00) = P(Z > -2.00) - P(Z > 2.00) = 0.9772 - 0.0228 = 0.9544$$

となります．
したがって,

$$P(488 \leq \bar{X} \leq 512) = 0.9544$$

となります．つまり，標本平均 $\bar{X}$ が規格を満たす確率は 0.9544 です．

問 2

(1) 標本比率 $\hat{p}$ の期待値と標準偏差

母比率は

$$p = 0.6$$

です．
よって，標本比率 $\hat{p} = X/150$ の期待値は,

$$E(\hat{p}) = 0.6$$

となります．つまり，標本比率の期待値は母比率と同じ 0.6 です．
また，標本比率 $\hat{p} = X/150$ の分散は,

$$V(\hat{p}) = \frac{0.6(1-0.6)}{150} = \frac{0.24}{150} = 0.0016$$

となります．したがって，標本比率の標準偏差は

$$\sqrt{V(\hat{p})} = \sqrt{0.0016} = 0.04$$

となります．

(2) **レジ袋が足りなくなる確率**

レジ袋を 99 枚用意しているので，レジ袋が足りなくなるのは

$$X \geq 100$$

となる場合です．つまり，

$$\frac{X}{150} \geq \frac{100}{150}$$

となり，標本比率は $\hat{p} = X/150$ なので，

$$\hat{p} \geq \frac{100}{150} = \frac{2}{3}$$

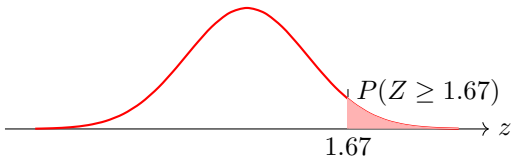
と書けます．

ここでは，標本サイズ $n = 150$ が十分大きいとして，標本比率 $\hat{p}$ を期待値 0.6，標準偏差 0.04 の正規分布で近似して扱うことにします．

標本比率が 2/3 以上になる確率 $P(\hat{p} \geq 2/3)$ を標準化すると，

$$P\left(\hat{p} \geq \frac{2}{3}\right) = P\left(Z \geq \frac{\frac{2}{3} - 0.60}{0.04}\right) = P\left(Z \geq \frac{\frac{2}{3} - \frac{3}{5}}{\frac{4}{100}}\right) = P\left(Z \geq \frac{\frac{1}{15}}{\frac{4}{100}}\right) = P\left(Z \geq \frac{1}{15} \cdot \frac{100}{4}\right)$$
$$= P\left(Z \geq \frac{5}{3}\right) = P(Z \geq 1.666\ldots) \approx P(Z \geq 1.67)$$

となります．



標準正規分布表より

$$P(Z > 1.67) = 0.0475$$

なので，

$$P\left(\hat{p} \geq \frac{2}{3}\right) = 0.0475$$

となります．つまり，150 人の来店に対してレジ袋が足りなくなる確率は約 0.05 です．

第 14 章　推定と検定

14.1　区間推定

問題 14.1

(1) 母平均 μ の 95% 信頼区間は

$$50 - 1.96 \times \frac{10}{\sqrt{25}} \ \leq \ \mu \ \leq \ 50 + 1.96 \times \frac{10}{\sqrt{25}}$$

つまり，

$$50 - 1.96 \times 2 \ \leq \ \mu \ \leq \ 50 + 1.96 \times 2$$

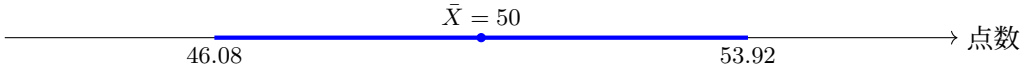
より，

$$46.08 \ \leq \ \mu \ \leq \ 53.92$$

となります．区間の中心は

$$\frac{46.08 + 53.92}{2} = 50$$

となり，標本平均 $\bar{X} = 50$ と一致していることがわかります．



95% 信頼区間の中心が標本平均になっているイメージ

また，区間の幅（右端 − 左端）は

$$53.92 - 46.08 = 7.84$$

となります．

(2) 母標準偏差を $\sigma = 20$ とすると，母平均 μ の 95% 信頼区間は

$$50 - 1.96 \times \frac{20}{\sqrt{25}} \ \leq \ \mu \ \leq \ 50 + 1.96 \times \frac{20}{\sqrt{25}}$$

つまり，

$$50 - 7.84 \ \leq \ \mu \ \leq \ 50 + 7.84$$

より，

$$42.16 \ \leq \ \mu \ \leq \ 57.84$$

となります．区間の幅（右端 − 左端）は

$$57.84 - 42.16 = 15.68$$

となり，(1) の幅 7.84 のちょうど 2 倍になっています．σ を 10 から 20 に 2 倍にすると，区間の幅も 2 倍に広がることがわかります．

(3) 標本サイズを $n = 100$ とすると，母平均 μ の 95% 信頼区間は

$$50 - 1.96 \times \frac{10}{\sqrt{100}} \ \leq \ \mu \ \leq \ 50 + 1.96 \times \frac{10}{\sqrt{100}}$$

つまり，

$$50 - 1.96 \ \leq \ \mu \ \leq \ 50 + 1.96$$

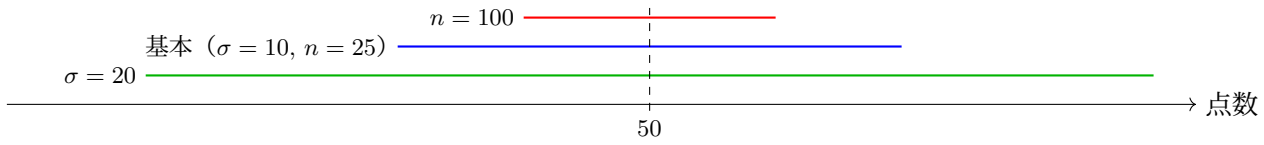
より，

$$48.04 \ \leq \ \mu \ \leq \ 51.96$$

となります．区間の幅（右端 − 左端）は

$$51.96 - 48.04 = 3.92$$

となり，(1) の幅 7.84 の半分になっています．n を 25 から 100 に 4 倍にすると，$\sqrt{n}$ は 5 から 10 に 2 倍になり，区間の幅は半分になることがわかります．



σ や n を変えたときの 95% 信頼区間の比較

問題 14.2

分散（標準偏差）が既知のとき，母平均 μ の 95% 信頼区間は次の式で求められます．

$$\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} \ \leq \ \mu \ \leq \ \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

ここで，$\bar{X}$ は標本平均，σ は母標準偏差，n は標本サイズです．

本問では $\bar{X} = 40,\ \sigma = 21,\ n = 49$ なので，母平均 μ の 95% 信頼区間は

$$40 - 1.96 \times \frac{21}{\sqrt{49}} \ \leq \ \mu \ \leq \ 40 + 1.96 \times \frac{21}{\sqrt{49}}$$

つまり，

$$40 - 1.96 \times 3 \ \leq \ \mu \ \leq \ 40 + 1.96 \times 3$$

より，

$$34.12 \ \leq \ \mu \ \leq \ 45.88$$

となります．

問題 14.3

不合格品の標本比率は

$$\hat{p} = \frac{450}{900} = \frac{1}{2}$$

です．また，標本サイズは $n = 900$ なので，母比率 p の 95% 信頼区間は

$$\frac{1}{2} - 1.96 \times \sqrt{\frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{900}} \ \leq \ p \ \leq \ \frac{1}{2} + 1.96 \times \sqrt{\frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{900}}$$

つまり，

$$\frac{1}{2} - 1.96 \times \sqrt{\frac{1}{3600}} \ \leq \ p \ \leq \ \frac{1}{2} + 1.96 \times \sqrt{\frac{1}{3600}}$$

より，

$$\frac{1}{2} - 1.96 \times \frac{1}{60} \ \leq \ p \ \leq \ \frac{1}{2} + 1.96 \times \frac{1}{60}$$

となります．分母を 6000 にそろえると，

$$\frac{3000}{6000} - \frac{196}{6000} \ \leq \ p \ \leq \ \frac{3000}{6000} + \frac{196}{6000}$$

つまり，

$$\frac{2804}{6000} \ \leq \ p \ \leq \ \frac{3196}{6000}$$

となり，

$$0.4673 \ \leq \ p \ \leq \ 0.5327$$

となります．これを小数第 3 位まで示すと，

$$0.467 \ \leq \ p \ \leq \ 0.533$$

となります．

14.1 の確認テスト

問 1

母分散（標準偏差）が既知のとき，母平均 μ の 95% 信頼区間は次の式で求められます．

$$\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} \ \leq \ \mu \ \leq \ \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

ここで，$\bar{X}$ は標本平均，σ は母標準偏差，n は標本サイズです．

本問では $\bar{X} = 210,\ \sigma = 20,\ n = 64$ なので，母平均 μ の 95% 信頼区間は

$$210 - 1.96 \times \frac{20}{\sqrt{64}} \ \leq \ \mu \ \leq \ 210 + 1.96 \times \frac{20}{\sqrt{64}}$$

つまり，

$$210 - 1.96 \times 2.5 \ \leq \ \mu \ \leq \ 210 + 1.96 \times 2.5$$

より，

$$205.10 \ \leq \ \mu \ \leq \ 214.90$$

となります.

問 2

右巻きの貝殻の標本比率は

$$\hat{p} = \frac{90}{150} = \frac{3}{5}$$

です．また，標本サイズは $n = 150$ なので，母比率 p の 95% 信頼区間は

$$\frac{3}{5} - 1.96 \times \sqrt{\frac{\frac{3}{5}\left(1-\frac{3}{5}\right)}{150}} \leq p \leq \frac{3}{5} + 1.96 \times \sqrt{\frac{\frac{3}{5}\left(1-\frac{3}{5}\right)}{150}}$$

つまり，

$$\frac{3}{5} - 1.96 \times \sqrt{\frac{1}{625}} \leq p \leq \frac{3}{5} + 1.96 \times \sqrt{\frac{1}{625}}$$

より，

$$\frac{3}{5} - 1.96 \times \frac{1}{25} \leq p \leq \frac{3}{5} + 1.96 \times \frac{1}{25}$$

となります．分母を 625 にそろえると，

$$\frac{375}{625} - \frac{49}{625} \leq p \leq \frac{375}{625} + \frac{49}{625}$$

つまり，

$$\frac{326}{625} \leq p \leq \frac{424}{625}$$

となり，

$$0.5216 \leq p \leq 0.6784$$

となります．これを小数第 3 位まで示すと，

$$0.522 \leq p \leq 0.678$$

となります.

問 3

例題 14.3 の設定（**標本比率 0.20，標本サイズ 400 による 95% 信頼区間 0.161〜0.239**）を例として挙げながら説明します.

(1) ×：95% 信頼区間は「母比率を含むと期待される区間」です．しかし，1 回の標本から作った区間が本当に母比率を含んでいるかどうかは断定できません.

（例）標本比率が $\hat{p} = 0.20$，標本サイズ $n = 400$ のとき，95% 信頼区間は 0.161〜0.239 と計算されますが，この区間が母比率を含んでいるかどうかはわかりません.

(2) ×：母比率 p は固定された 1 つの値です．したがって，「母比率が信頼区間のどこかにある確率が 0.95」という解釈は誤りです．確率が関係するのは標本比率のゆれのほうです.

（例）信頼区間が 0.161〜0.239 と計算されても，「母比率がこの区間にある確率が 0.95」という意味ではありません.

(3) ○：同じ条件で標本を多数取り直して，そのつど

$$\hat{p} - 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \ \leq \ p \ \leq \ \hat{p} + 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

の形で 95% 信頼区間を作る手続きには，多数の区間のうち約 95% が母比率 p を含むという理論的な性質があります．

（例）同じ条件で 400 人を取り直すと，そのたびに標本比率 $\hat{p}$ の値は母比率 p のまわりで少しずつ変わります．そのつど

$$\hat{p} - 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{400}} \ \leq \ p \ \leq \ \hat{p} + 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{400}}$$

の区間が得られます．この手続きには，多数の区間のうち約 95% が母比率を含むという理論的な性質があります．

(4) ○：標本サイズ n が大きいほど，$\sqrt{\hat{p}(1-\hat{p})/n}$ が小さくなるため，信頼区間は狭くなります．

（例）

- $n = 400$ → 片側の広がり $1.96 \times \sqrt{\hat{p}(1-\hat{p})/400}$
- $n = 800$ → 片側の広がり $1.96 \times \sqrt{\hat{p}(1-\hat{p})/800}$

同じ $\hat{p}$ なら，$n = 800$ のほうが区間は狭くなります．

(5) ○：母比率の信頼区間は標本比率 $\hat{p}$ を中心に左右に広げたものです．

（例）標本比率が 0.20 のとき，母比率 p の 95% 信頼区間は

$$0.20 - 1.96 \times \sqrt{\frac{0.20 \times 0.80}{400}} \ \leq \ p \ \leq \ 0.20 + 1.96 \times \sqrt{\frac{0.20 \times 0.80}{400}}$$

となり，この区間の中心は標本比率 $\hat{p} = 0.20$ です．

14.2　仮説検定

問題 14.4

(1) 検定統計量（標本平均の z 値）は，

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - 12.0}{2.0/5} = \frac{\bar{X} - 12.0}{0.4}$$

の式に代入して求めます．

地域 A（平均 12.7 kWh）では，

$$z_A = \frac{12.7 - 12.0}{0.4} = \frac{0.7}{0.4} = 1.75$$

地域 B（平均 12.9 kWh）では，

$$z_B = \frac{12.9 - 12.0}{0.4} = \frac{0.9}{0.4} = 2.25$$

地域 C（平均 11.4 kWh）では，

$$z_C = \frac{11.4 - 12.0}{0.4} = \frac{-0.6}{0.4} = -1.5$$

となります．

(2) 有意水準 5% の両側検定では，棄却域は

$$|z| \geq 1.96$$

です．この基準に従って判断します．

地域 A は，計算した z 値が 1.75 であり，1.96 より小さいため，帰無仮説を **棄却しません**．

地域 B は，計算した z 値が 2.25 であり，1.96 より大きいため，帰無仮説を **棄却します**．

地域 C は，計算した z 値が -1.5 であり，絶対値は 1.96 より小さいため，帰無仮説を **棄却しません**．

問題 14.5

まず，それぞれの検定における棄却域を確認します．

- 5% 両側検定：棄却域は $|Z| \geq 1.96$
- 1% 両側検定：棄却域は $|Z| \geq 2.58$
- 5% 片側検定（上側）：棄却域は $Z \geq 1.64$

標本 A，B，C の z 値は

$$Z_A = -2.10, \quad Z_B = 2.20, \quad Z_C = 2.70$$

です．

(1) 5% 両側検定では，$|Z| \geq 1.96$ が棄却域なので，

$$|Z_A| = 2.10 \geq 1.96, \quad |Z_B| = 2.20 \geq 1.96, \quad |Z_C| = 2.70 \geq 1.96$$

より，**H_0 を棄却するのは標本 A，標本 B，標本 C です．**

(2) 1% 両側検定では，$|Z| \geq 2.58$ が棄却域なので，

$$|Z_A| = 2.10 < 2.58, \quad |Z_B| = 2.20 < 2.58, \quad |Z_C| = 2.70 \geq 2.58$$

より，**H_0 を棄却するのは標本 C です．**

(3) 5% 片側検定（上側）では，$Z \geq 1.64$ が棄却域なので，

$$Z_A = -2.10 < 1.64, \quad Z_B = 2.20 \geq 1.64, \quad Z_C = 2.70 \geq 1.64$$

より，**H_0 を棄却するのは標本 B と標本 C です．**

以上の結果を表にまとめると，次のようになります（○ は「棄却する」，× は「棄却しない」を表します）．

標本	z 値	5% 両側	1% 両側	5% 片側（上側）
A	-2.10	○	×	×
B	2.20	○	×	○
C	2.70	○	○	○

(4) (a)○：第 1 種の過誤は「本当は帰無仮説が正しいのに棄却してしまう誤り」です．この調査では「本当は母平均が 100 のままで変わっていないのに，帰無仮説を棄却してしまう」状況を指しており，記述は定義と一致しています．

(b)○：第 2 種の過誤は「本当は帰無仮説が誤りなのに棄却しない誤り」です．この調査では「本当は母平均が 100 から変化しているのに，帰無仮説を棄却しない」状況を指しており，記述は定義と一致しています．

(c)○：有意水準を大きくすると棄却域は広くなり，帰無仮説を棄却しやすくなります．そのため，「本当は帰無仮説が正しいのに棄却してしまう」第 1 種の過誤は起こりやすくなります．

(d)×：有意水準を大きくすると棄却域は広くなり，帰無仮説を棄却しやすくなります．したがって，第 2 種の過誤（本当は帰無仮説が誤りなのに棄却しない誤り）はむしろ起こりにくくなります．「起こりやすくなる」という記述は誤りです．

問題 14.6

(1) ×：会場 A の z 値は 2.01 であり，5% 両側検定の棄却域（$z \leq -1.96$ または $z \geq 1.96$）に入ります．帰無仮説は棄却されるため，「棄却されず，例年の回答率と異なるとはいえない」という説明は不適切です．

(2) ×：会場 B の z 値は -1.61 であり，棄却域には入りません．帰無仮説は棄却されないため，「棄却され，例年の回答率と異なるといえる」という説明は不適切です．

(3) ×：会場 A の z 値は 2.01 で棄却域に入り帰無仮説を棄却しますが，会場 B の z 値は -1.61 で棄却域に入らず帰無仮説を棄却しません．よって「異なると判断しやすい」のは会場 A のほうであり，「会場 A よりも会場 B のほうが判断しやすい」という説明は不適切です．

14.2 の確認テスト

問 1

(1) 節水施策が効果をもつなら平均使用量は小さくなると考えているため，対立仮説は

$$H_1 : \ \mu < 480$$

となっています．したがって，**この検定は下側片側検定** です．

(2) **正解：(a)**

標準化した値は $z \approx -2.28$ です．

また，5% 下側片側検定の棄却域は $Z \leq -1.64$，1% 下側片側検定の棄却域は $Z \leq -2.33$ です．

$$-2.28 \leq -1.64 \quad \text{かつ} \quad -2.28 > -2.33$$

より，次がわかります．

- 5% 有意である（棄却域に入る）
- 1% 有意ではない（棄却域に入らない）

よって，適切なのは (a)（5% 有意であるが 1% 有意ではない）となります．

問 2

まず，標本比率は

$$\hat{p} = \frac{90}{300} = \frac{3}{10}$$

です．母比率は $p_0 = 1/4$，また，標本サイズは $n = 300$ です．

標本比率を標準化した値は

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

で与えられます.

これらを代入すると,

$$z = \frac{\frac{3}{10} - \frac{1}{4}}{\sqrt{\frac{1}{4} \cdot \frac{3}{4}/300}} = \frac{\frac{1}{20}}{\frac{1}{40}} = 2$$

となります.

ここで，有意水準 5% の両側検定では，棄却域は

$$z \leq -1.96 \quad \text{または} \quad z \geq 1.96$$

です.

今回の z 値は

$$z = 2 \geq 1.96$$

であり，棄却域に入ります.

したがって，帰無仮説

$$H_0 : \ p = 1/4$$

を棄却します．レビュー率は通常の $p = 1/4$ から変化したといえます.