

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

第1章 データの種類

1.1 質的変数と量的変数

例題 1.1 質的変数と量的変数のちがいを確認する

下の表は、中古車 6 台についてまとめた情報です．この表に示されている次の変数を，質的変数と量的変数に分類しましょう．

- | | |
|---------------|----------------------|
| (1) 車名 | (2) 年式 |
| (3) 排気量 | (4) 環境性能評価 (★ 1～★ 5) |
| (5) 気筒数 | (6) 走行距離 |
| (7) 修復歴 | (8) 駆動方式 |
| (9) トランスミッション | (10) 車両状態 (優・良・可) |

例題 1.1 質的変数と量的変数のちがいを確認する

車名	年式	排気量	環境性能評価	気筒数	走行距離
アイシス	2018	1500	★ 2	4	35000
ヴェゼル	2016	1500	★ 3	4	82000
デイズ	2020	660	★ 4	3	12000
アテンザ	2015	2200	★ 3	4	120000
スイフト	2019	1000	★ 4	3	58000
WRX STI	2017	2457	★ 1	4	48000

表 1.1 中古車データ（前半）

例題 1.1 質的変数と量的変数のちがいを確認する

車名	修復歴	駆動方式	トランスミッション	車両状態
アイシス	なし	FF	AT	優
ヴェゼル	あり	FF	CVT	良
デイズ	なし	FF	AT	優
アテンザ	あり	FR	MT	可
スイフト	なし	FF	CVT	良
WRX STI	なし	AWD	MT	優

表 1.1 中古車データ（後半）

例題 1.1 質的変数と量的変数のちがいを確認する

変数

変数とは、「どんな情報が入っているか」を表すものです。

たとえば「修復歴」という変数には、「あり」または「なし」という値が入ります。

例題 1.1 質的変数と量的変数のちがいを確認する

質的変数と量的変数

質的変数は、**種類を表す値**をとる変数です。

量的変数は、**数量として扱われる値**をとる変数です。

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**」

を手がかりに考えます。

例題 1.1 質的変数と量的変数のちがいを確認する

データの種類

変数について調べたときに得られる**値の集まり**をデータといいます。

質的変数に対するデータを**質的データ**といいます。

量的変数に対するデータを**量的データ**といいます。

例題 1.1 質的変数と量的変数のちがいを確認する

数値で表されていても質的変数の場合がある

たとえば，

車のナンバー，電話番号

は，数値で表されています．

しかし，これらの数値は種類を区別するためのものであり，

数値の差や比には意味がありません．

したがって，質的変数として扱います．

例題 1.1 質的変数と量的変数のちがいを確認する 順序があっても質的変数の場合がある

たとえば，

ホテルの星評価（★ 1～★ 5），検定の級（1 級・2 級・3 級）

には順序があります．

しかし，

その順序だけに意味があり， 数値として量的には扱いません．

したがって，これらも質的変数です．

例題 1.1 質的変数と量的変数のちがいを確認する

量的変数

たとえば，

面積，収入

は，対象の大きさや量を数値で表しています．

この場合は，

数値の大小や差に意味があります．

したがって，量的変数として扱います．

例題 1.1 質的変数と量的変数のちがいを確認する

(1) 車名

車名は、車を名前によって区別するための情報です。

「アイシス」「ヴェゼル」「デイズ」などは、車の種類のちがいを表しています。したがって、

質的変数です。

例題 1.1 質的変数と量的変数のちがいを確認する

(2) 年式

年式は年を数値で表すことができ、その数値のちがいは

年の経過という量的な意味

をもちます。したがって、

量的変数です。

例題 1.1 質的変数と量的変数のちがいを確認する

(3) 排気量

排気量は、エンジンの大きさを数値で表したものです。

数値のちがいが排気量のちがいという量的な意味をもちます。
したがって、

量的変数です。

例題 1.1 質的変数と量的変数のちがいを確認する

(4) 環境性能評価

環境性能評価の

★ 1, ★ 2, ★ 3, ★ 4, ★ 5

には順序があります。

しかし、星の数の差が環境性能の量のちがいを表しているわけではありません。したがって、

質的変数です。

例題 1.1 質的変数と量的変数のちがいを確認する

(5) 気筒数

気筒数は、気筒（シリンダー）の本数を数で表したものです。

数値のちがいが気筒数のちがいという量的な意味をもちます。
したがって、

量的変数です。

例題 1.1 質的変数と量的変数のちがいを確認する

(6) 走行距離

走行距離は、どれだけ走ったかを数値で表したものです。

数値のちがいが走行量のちがいを表します。したがって、

量的変数です。

例題 1.1 質的変数と量的変数のちがいを確認する

(7) 修復歴

修復歴は、

「あり」「なし」

という区別を表す情報です。

これは量のちがいではなく、種類のちがいです。したがって、

質的変数です。

例題 1.1 質的変数と量的変数のちがいを確認する

(8) 駆動方式

駆動方式は、

FF, FR, AWD

などの方式のちがいを表します。

これは量のちがいではなく、種類のちがいです。したがって、

質的変数です。

例題 1.1 質的変数と量的変数のちがいを確認する

(9) トランスミッション

トランスミッションは、

AT, MT, CVT

などの方式のちがいを表します。

これは量のちがいではなく、種類のちがいです。したがって、

質的変数です。

例題 1.1 質的変数と量的変数のちがいを確認する

(10) 車両状態

車両状態の

「優」「良」「可」

には順序があります。

しかし、数値化したとしても、その差が量として意味をもつわけではありません。したがって、

質的変数です。

例題 1.1 質的変数と量的変数のちがいを確認する

質的変数と量的変数に分類する

変数の種類	変数
質的変数	車名，環境性能評価，修復歴，駆動方式， トランスミッション，車両状態
量的変数	年式，排気量，気筒数，走行距離

このように，種類や区分を表しているか，数値そのものが大きさや量を表しているかに注目すると，質的変数と量的変数を区別できます。

例題 1.2 変数の尺度のちがいを確認する

例題 1.1 の「表 1.1 中古車データ」に含まれる次の変数について、それぞれがどの尺度に分類されるかを答えましょう。

- | | |
|---------------|----------------------|
| (1) 車名 | (2) 年式 |
| (3) 排気量 | (4) 環境性能評価 (★ 1～★ 5) |
| (5) 気筒数 | (6) 走行距離 |
| (7) 修復歴 | (8) 駆動方式 |
| (9) トランスミッション | (10) 車両状態 (優・良・可) |

次の 4 つの尺度から分類します。

名義尺度, 順序尺度, 間隔尺度, 比例尺度

例題 1.2 変数の尺度のちがいを確認する

4 つの尺度

質的変数は、名義尺度と順序尺度に、量的変数は、間隔尺度と比例尺度に分かれます。

このように、変数は値の性質に応じて次の 4 つの尺度に分類されます。

例題 1.2 変数の尺度のちがいを確認する

4 つの尺度

名義尺度は、種類のちがいだけを表します。

順序尺度は、種類のちがいに加えて、**順序**に意味があります。

間隔尺度は、順序に加えて、**差**に意味があります。

比例尺度は、順序・差に加えて、**比**にも意味があります。

例題 1.2 変数の尺度のちがいを確認する

尺度のちがいを整理すると、

尺度	順序	差	比	0 の意味
名義尺度	×	×	×	なし（分類のラベル）
順序尺度	○	×	×	なし（順序のみ）
間隔尺度	○	○	×	相対的な 0（任意の基準）
比例尺度	○	○	○	絶対的な 0（量の不存在）

例題 1.2 変数の尺度のちがいを確認する

判断するときは、

「順序」「差」「比」に意味があるか

を順番に考えます。

例題 1.2 変数の尺度のちがいを確認する

名義尺度

たとえば，

支払い方法

や

国道番号（1 号・23 号・41 号）

は，種類のちがいを表しています．

数値で表されている場合でも，

数値の大小や差に量としての意味はありません．

例題 1.2 変数の尺度のちがいを確認する

順序尺度

たとえば,

模試の順位 (1 位・2 位・3 位)

や

服のサイズ (S・M・L)

には順序があります.

しかし,

その差や比には意味がありません.

例題 1.2 変数の尺度のちがいを確認する

間隔尺度

たとえば,

西暦の年

では,

$$2020 - 2010 = 10$$

のように, 差には意味があります.

例題 1.2 変数の尺度のちがいを確認する

しかし，西暦の年には，「年の量がまったくない」ことを表す絶対的な 0 がないので，

2000 年を 1000 年の「2 倍」

と考えることに意味はありません．

つまり，

比には意味がありません．

例題 1.2 変数の尺度のちがいを確認する

比例尺度

たとえば，

人口，降水量

などでは，0 が「量がまったくない」ことを表します．そのため，

2 倍，半分

といった比較もできます．

つまり，

差にも比にも意味があります．

例題 1.2 変数の尺度のちがいを確認する

(1) 車名

車の種類のちがいを表し，順序はないので，
名義尺度です。

例題 1.2 変数の尺度のちがいを確認する

(2) 年式

年の差には意味がありますが、0 年は絶対的な 0 ではないので、
間隔尺度です。

例題 1.2 変数の尺度のちがいを確認する

(3) 排気量

差にも比にも意味があるので、

比例尺度です。

例題 1.2 変数の尺度のちがいを確認する

(4) 環境性能評価

★ 1～★ 5 には順序がありますが，星の数の差に量的な意味はないので，

順序尺度です．

例題 1.2 変数の尺度のちがいを確認する

(5) 気筒数

0 は「気筒がない」ことを表し、差にも比にも意味があるので、
比例尺度です。

例題 1.2 変数の尺度のちがいを確認する

(6) 走行距離

0 は「まったく走っていない」ことを表し、差にも比にも意味があるので、

比例尺度です。

例題 1.2 変数の尺度のちがいを確認する

(7) 修復歴

「あり」「なし」という種類のちがいだけを表すので、

名義尺度です。

例題 1.2 変数の尺度のちがいを確認する

(8) 駆動方式

FF, FR, AWD などの種類のちがいを表すので,

名義尺度です.

例題 1.2 変数の尺度のちがいを確認する

(9) トランスミッション

AT, MT, CVT などの種類のちがいを表すので,

名義尺度です.

例題 1.2 変数の尺度のちがいを確認する

(10) 車両状態

車両状態の

「優」「良」「可」

には順序があります。しかし、

「優」と「良」の差 と 「良」と「可」の差

を量として比較することはできません。したがって、

順序尺度です。

例題 1.2 変数の尺度のちがいを確認する

4 つの尺度に分類する

尺度	変数
名義尺度	車名，修復歴，駆動方式，トランスミッション
順序尺度	環境性能評価，車両状態
間隔尺度	年式
比例尺度	排気量，気筒数，走行距離

このように，順序があるか，差に意味があるか，比に意味があるかを順に確認することで，4 つの尺度を区別できます．

例題 1.3 離散変数と連続変数のちがいを確認する

例題 1.1 の「表 1.1 中古車データ」に含まれる以下の変数について、

質的変数, 離散変数, 連続変数

のどれに分類されるかを答えましょう。

- | | |
|---------------|----------------------|
| (1) 車名 | (2) 年式 |
| (3) 排気量 | (4) 環境性能評価 (★ 1~★ 5) |
| (5) 気筒数 | (6) 走行距離 |
| (7) 修復歴 | (8) 駆動方式 |
| (9) トランスミッション | (10) 車両状態 (優・良・可) |

例題 1.3 離散変数と連続変数のちがいを確認する

質的変数・離散変数・連続変数

質的変数は、種類や性質のちがいを表す変数です。

離散変数は、とびとびの値をとる量的変数です。

連続変数は、連続的な値をとる量的変数です。

例題 1.3 離散変数と連続変数のちがいを確認する

各変数を分類するときは、

種類の区別を表しているか

それとも、

数えられる量か

あるいは、

連続的な値として扱われる量か

を手がかりに考えます。

例題 1.3 離散変数と連続変数のちがいを確認する

離散変数

たとえば，

個数，回数

のように，

$0, 1, 2, 3, \dots$

と数える量は，**離散変数**です．

例題 1.3 離散変数と連続変数のちがいを確認する

連続変数

一方，

体積，面積，時間

のように，測定によって連続的な値をとる量は，

連続変数です．

例題 1.3 離散変数と連続変数のちがいを確認する

記録値と本来の変数

表に整数で記録されているからといって、必ずしも離散変数とは限りません．たとえば，走行距離が

35000, 82000, 12000

のように整数で記録されていても，走行距離そのものは測定される連続的な量です．したがって，

連続変数

として扱います．

例題 1.3 離散変数と連続変数のちがいを確認する

(1) 車名

名称としての区別を表すので、

質的変数です。

例題 1.3 離散変数と連続変数のちがいを確認する

(2) 年式

年単位のとびとびの値なので、

離散変数です。

例題 1.3 離散変数と連続変数のちがいを確認する

(3) 排気量

表では整数で記録されていますが、本来は測定値であり、連続的な値として扱うので、

連続変数です。

例題 1.3 離散変数と連続変数のちがいを確認する

(4) 環境性能評価

★の数には順序がありますが、量として扱う値ではないので、
質的変数です。

例題 1.3 離散変数と連続変数のちがいを確認する

(5) 気筒数

3 気筒・4 気筒など，本数として数えられる量なので，

離散変数です．

例題 1.3 離散変数と連続変数のちがいを確認する

(6) 走行距離

本来は測定される連続的な量なので、

連続変数です。

例題 1.3 離散変数と連続変数のちがいを確認する

(7) 修復歴

「あり」「なし」という区別を表すので、

質的変数です。

例題 1.3 離散変数と連続変数のちがいを確認する

(8) 駆動方式

FF, FR, AWD などの種類のちがいを表すので,

質的変数です.

例題 1.3 離散変数と連続変数のちがいを確認する

(9) トランスミッション

AT, MT, CVT などの種類のちがいを表すので,

質的変数です.

例題 1.3 離散変数と連続変数のちがいを確認する

(10) 車両状態

車両状態の

「優」「良」「可」

には順序があります。

しかし、これは量の大小ではなく、状態の種類の違いを表しています。したがって、

質的変数です。

例題 1.3 離散変数と連続変数のちがいを確認する

変数の種類	変数
質的変数	車名，環境性能評価，修復歴，駆動方式， トランスミッション，車両状態
離散変数	年式，気筒数
連続変数	排気量，走行距離

このように，まず質的変数か量的変数かを確認し，量的変数では値を1つずつ数えられるか，連続的に変化するかに注目すると，離散変数と連続変数を区別できます。

第1章 データの種類

1.2 母集団と標本

例題 1.4 母集団と標本のちがいを理解する

ある調査を行うとき，調べたい対象全体を母集団，その一部を取り出したものを標本といいます．

次の場面では，どれが母集団で，どれが標本に当たるかを考えてみましょう．

例題 1.4 母集団と標本のちがいを理解する

- (1) ある大学の学生の満足度を知るため、学生のなかから 500 人にアンケートを依頼した。
- (2) 市内の飲食店の売上傾向を調べるため、市内の飲食店のなかから 120 店を選んで調査した。
- (3) 日本に住む高校生の平均通学時間を調べるため、全国の高校生から 2000 人を選んで調査した。

例題 1.4 母集団と標本のちがいを理解する

母集団と標本

母集団とは、調査したい対象全体のことです。

標本とは、母集団から調査のために取り出した一部のことです。

なお、対象から得られるデータの集まりを母集団や標本とよぶこともあります。

母集団 $\supset$ 標本

例題 1.4 母集団と標本のちがいを理解する

標本の大きさ

標本に含まれる対象の個数を**標本の大きさ**といいます。

たとえば、100 人を標本として取り出した場合、

大きさ 100 の標本

とよびます。

標本の大きさは**標本サイズ**ともいいます。

例題 1.4 母集団と標本のちがいを理解する

(1) 大学の学生の満足度調査

母集団：この大学の学生全体

標本：アンケートに協力した 500 人

学生全体

から

500 人を取り出して調査

例題 1.4 母集団と標本のちがいを理解する

(2) 市内の飲食店の売上傾向調査

母集団：市内の飲食店全体

標本：調査した 120 店

飲食店全体

から

120 店を選んで調査

例題 1.4 母集団と標本のちがいを理解する

(3) 高校生の平均通学時間調査

母集団：日本に住む高校生全体

標本：調査した 2000 人

高校生全体

から

2000 人を選んで調査

例題 1.5 全数調査と標本調査のちがいを理解する

調査には，対象全体を調べる全数調査と，対象の一部を取り出して調べる標本調査があります．

次の場面では，どちらの調査方法が用いられているかを考えてみましょう．

例題 1.5 全数調査と標本調査のちがいを理解する

- (1) 空港で、持ち込み禁止物がないかを確認するため、全員の手荷物を検査した。
- (2) 工場で製品の品質を確認するため、生産された部品のなかから 120 個を選んで検査した。
- (3) 国勢調査を実施した。
- (4) 学校基本調査を実施した。
- (5) 家計調査を実施した。

例題 1.5 全数調査と標本調査のちがいを理解する

全数調査と標本調査

全数調査とは、調べたい対象全体（母集団）をもれなく調査する方法です。

標本調査とは、対象の一部（標本）を取り出して調査し、母集団の特徴を推測する方法です。

全体を調べる

全数調査

一部を調べる

標本調査

例題 1.5 全数調査と標本調査のちがいを理解する

全数調査は、対象全体を調べるため、

- 標本の選び方による偏りは生じません。
- 標本を一部だけ取り出すことによる誤差も生じません。

一方で、

時間や費用がかかる場合があります。

例題 1.5 全数調査と標本調査のちがいを理解する

標本調査は、対象の一部を調べるため、

効率的に調査できます。

ただし、標本によって結果が変わることがあるため、

標本の選び方に注意が必要です。

例題 1.5 全数調査と標本調査のちがいを理解する

(1) 空港で全員の手荷物を検査

空港で全員の手荷物を検査しているため、対象全体を調べています。
したがって、

全数調査です。

例題 1.5 全数調査と標本調査のちがいを理解する

(2) 部品 120 個を選んで検査

生産された部品のなかから 120 個を選んで検査しているため、対象の一部を調べています。

したがって、

標本調査です。

例題 1.5 全数調査と標本調査のちがいを理解する

(3) 国勢調査

国勢調査は、西暦の末尾が 0 または 5 の年に実施される調査です。

日本に住むすべての人と世帯を対象とし、年齢、性別、出生年月、就業状態、世帯人員などを調べます。

対象全体を調べています。したがって、

全数調査です。

例題 1.5 全数調査と標本調査のちがいを理解する

国勢調査について

- 日本に住むすべての人と世帯が対象です。
- 回答する義務があります。
- 調査結果はインターネット等で公開されます。

例題 1.5 全数調査と標本調査のちがいを理解する

(4) 学校基本調査

学校基本調査は，文部科学省が毎年実施する調査です。

日本の学校教育に関わるすべての学校を対象とし，
在籍する児童生徒・学生数，教職員数，進学状況，卒業後の進路などを調べます。

対象全体を調べています。したがって，

全数調査です。

例題 1.5 全数調査と標本調査のちがいを理解する

学校基本調査について

対象となる学校には、

幼稚園，小学校，中学校，高等学校，大学など

が含まれます。

調査結果は教育施策の基礎資料として用いられ，インターネット等で公開されます。

例題 1.5 全数調査と標本調査のちがいを理解する

(5) 家計調査

家計調査は，総務省統計局が毎月実施する調査です．

全国の一般世帯の一部を抽出して，
家計の収入や支出，貯蓄，消費行動などを調べます．

対象の一部を調べています．したがって，

標本調査です．

例題 1.5 全数調査と標本調査のちがいを理解する

家計調査について

家計調査は、家計の実態を把握するために行われます。
調査結果は、

物価指数や景気動向の分析など
に用いられ、インターネット等で公開されます。

例題 1.6 無作為抽出の考え方を理解する

ここでは，母集団を「学校全体の生徒」として考えます。
次の標本の選び方が，無作為抽出かどうかを考えてみましょう。

- (1) 学校の名簿に番号を付け，乱数を使って 10 人を選んだ。
- (2) 放課後に図書室にいた生徒だけを選んだ。
- (3) 担任が推薦した生徒だけを選んだ。
- (4) くじ引きで当たった生徒だけを選んだ。

例題 1.6 無作為抽出の考え方を理解する

標本調査では、母集団の特徴をできるだけ正しく推測するために、

特定の属性や状況に偏らないように

標本を選ぶことが大切です。

たとえば、学校全体の生徒について調べたいのに、図書室にいる生徒だけを選ぶと、

標本に偏りが生じる可能性があります。

例題 1.6 無作為抽出の考え方を理解する

無作為抽出

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。

無作為抽出は、

ランダムサンプリング

ともよばれます。

例題 1.6 無作為抽出の考え方を理解する

なぜ無作為に標本を選ぶのでしょうか。

無作為に選ぶことで、

標本に偏りが生じる可能性を小さくする

ことができます。その結果、

母集団の特徴をより正しく推測できる

ようになります。

例題 1.6 無作為抽出の考え方を理解する

単純無作為抽出法

単純無作為抽出法とは，母集団のすべての対象に番号を付けるなどして一覧にし，そのなかから乱数やくじ引きなどを使って標本を選ぶ方法です．

すべての対象を一覧にする

そのなかから

乱数やくじ引きで選ぶ

例題 1.6 無作為抽出の考え方を理解する

(1) 名簿に番号を付け，乱数で 10 人を選ぶ

無作為抽出です．

名簿に番号を付け，乱数で番号を選んでいるためです．

人の判断によって，選ぶ生徒を決めているわけではありません．

例題 1.6 無作為抽出の考え方を理解する

(2) 放課後に図書室にいた生徒だけを選ぶ

無作為抽出ではありません．

図書室にいる生徒に特定の傾向がある可能性があるためです．
学校全体の生徒を代表する標本としては，偏りが生じる可能性があります．

例題 1.6 無作為抽出の考え方を理解する

(3) 担任が推薦した生徒だけを選ぶ

無作為抽出ではありません。

生徒を選ぶときに、

選ぶ人の主観

が入るためです。

例題 1.6 無作為抽出の考え方を理解する

(4) くじ引きで当たった生徒だけを選ぶ

無作為抽出です。

くじ引きという偶然性を利用して生徒を選んでいるためです。
人の判断によって、選ぶ生徒を決めているわけではありません。

例題 1.6 無作為抽出の考え方を理解する

標本の選び方	無作為抽出か
名簿＋乱数	○
図書室にいた生徒だけ	×
担任が推薦した生徒だけ	×
くじ引き	○

判断のポイントは、
偶然性を利用して選んでいるか

第2章 データの集め方とデータを中心

2.1 実験研究と観察研究

例題 2.1 実験研究と観察研究のちがいを理解する

データの集め方には、

調査者が条件を操作する「実験研究」

と

自然に起こっている状況をそのまま観察する「観察研究」

があります。

次の研究がどちらに当たるのかを考えましょう。

例題 2.1 実験研究と観察研究のちがいを理解する

- (1) 学習方法 A と学習方法 B の効果を比べるために、生徒に 2 つの学習方法をランダムに割り当て、テストの得点のちがいを調べた。
- (2) スマートフォン利用時間と睡眠時間の関係を知るために、生徒にアンケートを行い、実際の生活状況を調べた。
- (3) 肥料のちがいが収穫量に与える影響を調べるために、肥料 A・肥料 B・肥料 C をランダムに畑の区画に割り当て、作物の収穫量を比較した。
- (4) 交通量と事故件数の関係を調べるために、地域ごとの記録を用いて、過去の状況を分析した。

例題 2.1 実験研究と観察研究のちがいを理解する

実験研究

実験研究とは，調査者が条件を意図的に設定したり割り当てたりして，その条件のちがいによって結果にどのような変化が生じるかを調べる研究方法です．

条件を調査者が変えることで，

条件の変化が結果にどのような影響を与えるか

を調べやすい特徴があります．

例題 2.1 実験研究と観察研究のちがいを理解する

観察研究

観察研究とは，調査者が条件を操作せず，実際に起こっている状況を観察してデータを集める研究方法です．

自然な状況でのデータを集めることができます。
ただし，

原因と結果の関係を断定することは
むずかしい場合があります．

例題 2.1 実験研究と観察研究のちがいを理解する

	実験研究	観察研究
条件の操作	する	しない
データの集め方	条件を設定して調べる	実際の状況を観察する

判断のポイント

調査者が条件を操作しているかどうか

例題 2.1 実験研究と観察研究のちがいを理解する

(1) 学習方法 A・B をランダムに割り当てる

実験研究です。

学習方法を調査者が割り当てて比較しているためです。
つまり，調査者が条件を操作しています。

例題 2.1 実験研究と観察研究のちがいを理解する

(2) スマートフォン利用時間と睡眠時間を調べる

観察研究です。

条件を操作せず，実際の生活状況をそのまま調べているためです。
もともとのちがいを観察しています。

例題 2.1 実験研究と観察研究のちがいを理解する

(3) 肥料 A・B・C をランダムに割り当てる

実験研究です.

肥料の種類を調査者が割り当てて比較しているためです。
つまり，調査者が条件を操作しています。

例題 2.1 実験研究と観察研究のちがいを理解する

(4) 地域ごとの過去の記録を分析する

観察研究です。

過去の記録を用いて、自然に起こった状況を調べているためです。
調査者は条件を操作していません。

例題 2.2 実験の比較の仕方が適切かどうかを考える

次の実験が，比較の仕方として適切かどうかを考えましょう．

- (1) 肥料 A と肥料 B の効果を比べるために，日当たりのよい区画に肥料 A を，日当たりの悪い区画に肥料 B をまいた．
- (2) 新しい学習方法の効果を調べるために，生徒をランダムに 2 つのクラスに分け，一方のクラスでは新しい学習方法を，もう一方のクラスでは従来の方法を用いた．

例題 2.2 実験の比較の仕方が適切かどうかを考える

(3) カフェイン飲料の効果を調べるために、男子をカフェイングループに、女子をノンカフェイングループに分けて比較した。

(4) 新しい薬の効果を調べるために、参加者をランダムに 2 つのグループに分け、一方に薬を、もう一方に薬の成分を含まない偽薬を与えた。

例題 2.2 実験の比較の仕方が適切かどうかを考える

実験の比較の仕方

実験では、比較したい条件以外の要因が結果に与える影響を、できるだけ小さくすることが大切です。

条件以外の要因まで異なっていると、

どの要因が結果に影響したのか

わからなくなることがあります。

例題 2.2 実験の比較の仕方が適切かどうかを考える

そのため、

比較したい条件以外の要因を
できるだけそろえる

ことが大切です。

または、ランダムな割り当てによって、

条件以外の要因が偏りにくくなる

ようにします。

例題 2.2 実験の比較の仕方が適切かどうかを考える

(1) 肥料 A と肥料 B を，日当たりの異なる区画で比較

不適切です．

肥料だけでなく，日当たりという別の要因も異なっています．

そのため，肥料のちがいによる影響と日当たりのちがいによる影響を区別しにくくなります．

例題 2.2 実験の比較の仕方が適切かどうかを考える

(2) 生徒をランダムに 2 つのクラスに分けて比較

適切です.

生徒をランダムに割り当てているため、条件以外の要因がクラス間で偏りにくくなります.

そのため、学習方法の影響を比較しやすくなります.

例題 2.2 実験の比較の仕方が適切かどうかを考える

(3) 男子と女子でカフェインの有無を分けて比較

不適切です.

カフェインの有無だけでなく、性別という別の要因も異なっています。そのため、カフェインの影響と性別による影響を区別しにくくなります。

例題 2.2 実験の比較の仕方が適切かどうかを考える

(4) ランダムに分け，一方に薬，他方に偽薬を与える

適切です．

参加者をランダムに割り当てることで，条件以外の要因が偏りにくくなります．

また，偽薬を用いることで，薬を飲むという行為などの影響も考慮できます．

例題 2.2 実験の比較の仕方が適切かどうかを考える

適切な比較のために

適切に比較するためには、

比較したい条件以外の要因をできるだけそろえる

または

ランダムな割り当てによって偏りにくくする

こうすることで、

調べたい条件の影響を比較しやすくなります。

例題 2.3 実験研究のグループ分けが適切かどうかを考える

学習法 A と学習法 B の効果を比べるため、グループ A の生徒には学習法 A、グループ B の生徒には学習法 B を用いることになりました。

グループ A、グループ B への次の分け方が適切かどうかを考えましょう。

例題 2.3 実験研究のグループ分けが適切かどうかを考える

- (1) 事前に行ったテストの点数で生徒をいくつかの層（似た特徴をもつ集団）に分け，各層のなかで無作為にグループ A とグループ B に分けた．
- (2) 成績上位の生徒をグループ A，成績下位の生徒をグループ B に分けた．
- (3) 学年と性別で分け，各組み合わせのなかで無作為にグループ A とグループ B に分けた．

例題 2.3 実験研究のグループ分けが適切かどうかを考える

実験研究のグループ分け

実験研究では，比較するグループができるだけ同じ条件になるように分け方を工夫することが重要です．

たとえば，

性別や学力など

結果に影響しそうな要因を考えます．

例題 2.3 実験研究のグループ分けが適切かどうかを考える

結果に影響しそうな要因がある場合には、

その要因で層に分ける

そのうえで、

各層のなかで無作為に割り当てる

という方法がよく用いられます。

例題 2.3 実験研究のグループ分けが適切かどうかを考える

(1) 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る
適切です．

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
です．

例題 2.3 実験研究のグループ分けが適切かどうかを考える

この分け方では、

成績のちがいがグループ間で偏るのを抑える

ことができます。

そのため、学習方法の影響を比較しやすくなります。

例題 2.3 実験研究のグループ分けが適切かどうかを考える

(2) 成績上位を A, 成績下位を B に分ける

不適切です.

この分け方では,

成績がグループ間で偏ってしまいます.

例題 2.3 実験研究のグループ分けが適切かどうかを考える

結果に差が生じても，

「学習方法のちがい」の影響

と

「成績のちがい」の影響

を区別できなくなります．

例題 2.3 実験研究のグループ分けが適切かどうかを考える

(3) 学年と性別で分け、各組み合わせ内で無作為に割り当てる

適切です。

学年と性別で分け、それぞれの組み合わせのなかで無作為に割り当てているためです。

例題 2.3 実験研究のグループ分けが適切かどうかを考える

この分け方では、

学年や性別のちがいによる影響を抑える

ことができます。

そのため、学習方法の影響を比較しやすくなります。

第2章 データの集め方とデータの中心

2.2 代表値の求め方

例題 2.4 平均値を求める

次の 5 人のテストの点数について考えます。

生徒	A	B	C	D	E
点数	88	74	30	90	76

- (1) 5 人の点数の平均値を求めましょう。
- (2) 他の人と比べて点数が大きく離れている生徒 C を除いた 4 人の平均値を求めましょう。
- (3) 生徒 F が新しく加わり、A～E と F の 6 人の平均値が 75 点になったとします。生徒 F の点数を求めましょう。

例題 2.4 平均値を求める

平均値は、

データの合計をそのままにして、
すべての値を同じ大きさにそろえたときの 1 つ分の値

と考えることができます。

たとえば、集めたものを全員に均等に分けたときの

「1 人分の量」

を考えるイメージです。

例題 2.4 平均値を求める

代表値

データの中心的な傾向を 1 つの値で表したものを代表値といいます。

代表値には、よく使われるものとして、

平均値

中央値

最頻値

などがあります。

例題 2.4 平均値を求める

平均値，中央値，最頻値は，いずれもデータの中心的な傾向を表す代表値です．

ただし，

どの代表値を用いるかによって，
データのどの側面に注目するかが変わります．

中央値と最頻値については，このあとの例題で確認します．

例題 2.4 平均値を求める

平均値

平均値とは、データの値をすべて合計し、その個数で割った値です。

$$\text{平均値} = \frac{\text{値の合計}}{\text{個数}}$$

例題 2.4 平均値を求める

(1) 5 人の点数の平均値

まず、5 人の点数を合計すると、

$$88 + 74 + 30 + 90 + 76 = 358$$

となります。合計は 358 点、データの個数は 5 なので、

$$\frac{358}{5} = 71.6$$

より、

平均値は 71.6 点です。

例題 2.4 平均値を求める

(2) 生徒 C を除いた 4 人の平均値

5 人の合計は 358 点でした．生徒 C の 30 点を除くと，

$$358 - 30 = 328$$

となります．合計は 328 点，データの個数は 4 なので，

$$\frac{328}{4} = 82$$

より，

平均値は 82 点です．

例題 2.4 平均値を求める

平均値と合計の関係

平均値と合計には，次の関係があります．

$$\text{平均値} = \frac{\text{合計}}{\text{個数}}$$

より

$$\text{合計} = \text{平均値} \times \text{個数}$$

例題 2.4 平均値を求める

(3) 生徒 F の点数

A～F の 6 人の平均値は 75 点です．したがって，6 人の合計は，

$$75 \times 6 = 450$$

となります．A～E の 5 人の合計は 358 点なので，

$$450 - 358 = 92$$

より，

生徒 F の点数は 92 点です．

例題 2.5 中央値を求める

ある日の 8 人の歩数は次のとおりです。

人	A	B	C	D	E	F	G	H
歩数	8200	7600	9000	8500	7900	8800	8100	12000

- (1) 8 人の歩数の中央値を求めましょう。
- (2) H を除いた 7 人の中央値を求めましょう。
- (3) 8 人の平均値を求め、中央値と比べてどちらが大きいかを確認しましょう。

例題 2.5 中央値を求める

中央値

中央値とは、データを小さい順に並べたときの中央に位置する値のことです。

代表値の 1 つで、データの中心をとらえる値です。

中央値は、たとえば、

「真ん中の支出額」 や 「真ん中のテストの点数」

を知りたいときに役立ちます。

例題 2.5 中央値を求める

データの個数が奇数の場合

小さい順に並べたときの

中央の 1 つの値

が中央値になります。

データの個数を n とすると、

$\frac{n+1}{2}$ 番目の値

です。

例題 2.5 中央値を求める

データの個数が偶数の場合

小さい順に並べたときの中央の 2 つの値の平均値が中央値になります。

データの個数を n とすると、

$$\frac{n}{2} \text{ 番目の値} \quad \text{と} \quad \left(\frac{n}{2} + 1\right) \text{ 番目の値}$$

の平均を求めます。

例題 2.5 中央値を求める

(1) 8 人の歩数の中央値

まず，小さい順に並べます．

7600, 7900, 8100, 8200,
8500, 8800, 9000, 12000

データは 8 個なので，中央にくるのは 4 番目と 5 番目の値です．

例題 2.5 中央値を求める

4 番目の値は 8200, 5 番目の値は 8500 です.

したがって,

$$\frac{8200 + 8500}{2} = 8350$$

より,

中央値は 8350 歩です.

例題 2.5 中央値を求める

(2) H を除いた 7 人の中央値

H の 12000 歩を除くと,

7600, 7900, 8100, 8200,
8500, 8800, 9000

データは 7 個になります.

例題 2.5 中央値を求める

データが 7 個なので，中央は左から 4 番目の値です．

4 番目の値は 8200

より，

中央値は 8200 歩です．

例題 2.5 中央値を求める

(3) 平均値を求め、中央値と比べる

8 人の歩数の合計は、

$$7600 + 7900 + 8100 + 8200 + 8500 + 8800 + 9000 + 12000 = 70100$$

となります。合計は 70100 歩、データの個数は 8 なので、

$$\frac{70100}{8} = 8762.5$$

より、

平均値は 8762.5 歩です。

例題 2.5 中央値を求める

このデータでは,

平均値 = 8762.5 歩

中央値 = 8350 歩

なので,

平均値のほうが大きくなっています.

例題 2.5 中央値を求める

平均値が中央値より大きくなったのは、

12000

のように、他の値より大きな値が含まれているためです。
この大きな値によってデータの合計が大きくなり、その結果、

平均値が押し上げられています。

例題 2.5 中央値を求める

平均値と中央値

平均値は，とても大きな値や小さな値があると，その影響を受けやすい代表値です．

一方，中央値は，並べたときの中央の位置で決まります．

例題 2.6 最頻値を求める

ある 11 人のアルバイトの時給（円）は次のとおりです。

人	A	B	C	D	E	F	G	H	I	J	K
時給	1400	1500	1200	9000	1500	1300	1500	1200	1500	1500	1500

- (1) 平均値を求めましょう。
- (2) 中央値を求めましょう。
- (3) 最頻値を求め、平均値、中央値、最頻値の 3 つの大きさを比べましょう。

例題 2.6 最頻値を求める

最頻値

最頻値とは，データのなかで最も多く現れる値のことです。
代表値の 1 つで，そのデータで「よく見られる値」を表します。

最頻値は，たとえば，

「よく選ばれる価格帯」 や 「アンケートで最も多い回答」
などを知りたいときに役立ちます。

例題 2.6 最頻値を求める

外れ値

外れ値とは，他の値と比べて極端に大きい，または小さい値のことです．

今回のデータでは，

9000 円

が，他の時給と比べて極端に大きな値になっています．

例題 2.6 最頻値を求める

外れ値があると、平均値は、データの合計が大きくなるため、

大きな影響を受けやすい

という特徴があります。

一方、

**中央値や最頻値は、
平均値に比べて外れ値の影響を受けにくい**

傾向があります。

例題 2.6 最頻値を求める

(1) 平均値

11 人の時給の合計は、

$$\begin{aligned} &1400 + 1500 + 1200 + 9000 + 1500 + 1300 \\ &\quad + 1500 + 1200 + 1500 + 1500 + 1500 = 23100 \end{aligned}$$

例題 2.6 最頻値を求める

時給の合計は 23100 円，データの個数は 11 なので，

$$\frac{23100}{11} = 2100$$

より，

平均値は 2100 円です．

例題 2.6 最頻値を求める

(2) 中央値

小さい順に並べると、

1200, 1200, 1300, 1400, 1500, 1500,
1500, 1500, 1500, 1500, 9000

データは 11 個あります。

例題 2.6 最頻値を求める

データが 11 個なので，中央にくるのは

6 番目の値です．

6 番目は 1500 なので，

中央値は 1500 円です．

例題 2.6 最頻値を求める

(3) 最頻値を求め、3 つの代表値を比べる

並べたデータを見ると、

1500 円が 6 回

現れています。最も多く現れる値なので、

最頻値は 1500 円です。

例題 2.6 最頻値を求める

この例では,

平均値 = 2100 円

中央値 = 1500 円

最頻値 = 1500 円

となり,

平均値だけが大きくなっています.

例題 2.6 最頻値を求める

平均値が大きくなっているのは、

9000 円

のように、他の値より大きな値が含まれているためです。
この大きな値によってデータの合計が大きくなり、その結果、

平均値が押し上げられています。

例題 2.6 最頻値を求める

このデータでは，ほとんどの人の時給は

1200～1500 円

の範囲にあります．そのなかに，9000 円という極端に大きな外れ値が 1 つ含まれています．このような場合には，

代表値の解釈に注意が必要です．

例題 2.6 最頻値を求める

平均値：2100 円

外れ値の影響を強く受けており，多くの人の時給よりかなり大きな値になっています．そのため，

「典型的な時給」を表す値として使うには，
慎重な判断が必要です．

例題 2.6 最頻値を求める

中央値：1500 円

外れ値の影響を受けにくく、

データを小さい順に並べたときの
中心の位置を示しています。

このグループの時給の中心を考える際の手がかりとして適しています。

例題 2.6 最頻値を求める

最頻値：1500 円

最も多くの人が得ている時給なので、

「よく見られる時給」

を表す値として参考になります。

ただし、最頻値だけでは、データのばらつきや分布の広がりまではわかりません。

例題 2.6 最頻値を求める

最頻値は，データによっては

- 最も多く現れる値が複数になる
- たまたま少し多く現れただけの値が最頻値になる

ことがあります．そのような場合には，

最頻値がデータの特徴を
うまく表さないこともあります．

例題 2.7 代表値の特徴を整理する

次の場面では、平均値、中央値、最頻値のうち、どの代表値を使うのが適切かを考えましょう。

- (1) アンケートで「好きな果物」をたずねたとき、いちばん人気のある果物を知りたい。
- (2) あるクラスのテストで、1 人だけ極端に高い点数の生徒がいるとき、クラスの「だいたいの真ん中の実力」を知りたい。
- (3) ある町で 1 か月に出たごみの総量を住民の人数で割って、「1 人あたりのごみ量」を知りたい。

例題 2.7 代表値の特徴を整理する

平均値

平均値は、すべての値を合計して、データの個数で割った値です。

たとえば、

「1 人あたり」 や 「1 日あたり」

などを考えるときに役立ちます。

例題 2.7 代表値の特徴を整理する

中央値

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。

データの中心的な位置

をとらえるのに適しています。

また、平均値に比べて外れ値の影響を受けにくいという特徴があります。

例題 2.7 代表値の特徴を整理する

最頻値

最頻値は、データのなかでいちばん多く現れる値です。

「人気のもの」 や 「よく見られる値」

を表すのに向いています。

例題 2.7 代表値の特徴を整理する

最も多く現れる値が複数ある場合には、

最頻値も複数あります。

また、最頻値は、

果物の種類、血液型など

数値でないデータ（名義尺度）にも用いることができます。

例題 2.7 代表値の特徴を整理する

(1) いちばん人気のある果物を知りたい

最頻値が適しています。

知りたいのは、

いちばん多く選ばれた果物だからです。

例題 2.7 代表値の特徴を整理する

(2) クラスの「だいたいの真ん中の実力」を知りたい

中央値が適しています。

1 人だけ極端に高い点数の生徒がいるため、

外れ値の影響を受けにくい

という中央値の特徴が重要になります。

例題 2.7 代表値の特徴を整理する

(3) 「1 人あたりのごみ量」を知りたい

平均値が適しています。

ごみの総量を住民の人数で割るという、

「1 人あたり」

を考える場面だからです。

例題 2.7 代表値の特徴を整理する

代表値の使い分けのまとめ

平均値： 「1 人あたり」「1 日あたり」 など

中央値： データの中心の位置を知りたいとき

最頻値： 最も多い値や人気のものを知りたいとき

第3章 データのばらつき

3.1 分散

例題 3.1 偏差を計算する

あるクラスで行ったテストの点数は次のとおりです。

学生	テストの点数	偏差
A	55	
B	35	
C	65	
D	45	
E	50	
合計 平均値		

例題 3.1 偏差を計算する

- (1) 5 人の平均値を求めましょう。
- (2) 各値の偏差を求めましょう。
- (3) 偏差の合計と偏差の平均値を求め、表を完成させましょう。

例題 3.1 偏差を計算する

平均値が同じでも，データのばらつき方が同じとは限りません．
たとえば，平均値が同じ 50 点でも，

全員が 50 点

と

0 点と 100 点が半分ずつ

では，データの広がり方は大きく異なります．

例題 3.1 偏差を計算する

そこで、データのばらつきを考えるために、

それぞれの値が平均値から
どのくらいずれているかに注目します。

この「平均値からのズレ」を表すのが偏差です。

例題 3.1 偏差を計算する

偏差

偏差とは、それぞれの値が平均値からどの方向にどれだけずれているかを表すものです。

$$\text{偏差} = \text{値} - \text{平均値}$$

例題 3.1 偏差を計算する

値が平均値より大きい



偏差は正の値となります。

一方，

値が平均値より小さい



偏差は負の値となります。

例題 3.1 偏差を計算する

値が平均値と同じときは、

$$\text{値} - \text{平均値} = 0$$

なので、

偏差は 0 です。

つまり、平均値からまったくずれていないことを表します。

例題 3.1 偏差を計算する

(1) 5 人の平均値

まず、5 人の点数をすべて足すと、

$$55 + 35 + 65 + 45 + 50 = 250$$

となります。合計は 250 点、データの個数は 5 なので、

$$\frac{250}{5} = 50$$

より、

平均値は 50 点です。

例題 3.1 偏差を計算する

(2) 各値の偏差

平均値は 50 点なので、各値から 50 を引くと、

$$\text{A さんの偏差} \quad 55 - 50 = 5$$

$$\text{B さんの偏差} \quad 35 - 50 = -15$$

$$\text{C さんの偏差} \quad 65 - 50 = 15$$

$$\text{D さんの偏差} \quad 45 - 50 = -5$$

$$\text{E さんの偏差} \quad 50 - 50 = 0$$

となります。

例題 3.1 偏差を計算する

偏差を見ると，平均値との位置関係がわかります．

偏差が正 $\implies$ 平均値より大きい

偏差が 0 $\implies$ 平均値と同じ

偏差が負 $\implies$ 平均値より小さい

また，偏差の絶対値が大きいほど，

平均値から大きく離れていることを表します．

例題 3.1 偏差を計算する

(3) 偏差の合計と平均値

求めた偏差をすべて足します.

$$5 + (-15) + 15 + (-5) + 0 = 0$$

より,

偏差の合計は 0 です.

例題 3.1 偏差を計算する

偏差の合計は 0，データの個数は 5 なので，

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

より，

偏差の平均値も 0 です．

例題 3.1 偏差を計算する

学生	テストの点数	偏差
A	55	5
B	35	-15
C	65	15
D	45	-5
E	50	0
合計	250	0
平均値	50	0

例題 3.1 偏差を計算する

偏差の合計

今回の例では，偏差の合計は

$$5 + (-15) + 15 + (-5) + 0 = 0$$

となりました．正の偏差と負の偏差が打ち消し合っています．

一般に，偏差の合計は 0 になります．

例題 3.2 分散を計算する

例題 3.1 のデータについて，次の手順で分散を求めましょう．

学生	テストの点数	偏差	偏差の 2 乗
A	55	5	
B	35	-15	
C	65	15	
D	45	-5	
E	50	0	
合計	250	0	
平均値	50	0	

例題 3.2 分散を計算する

- (1) 各値の「偏差の 2 乗」を求めましょう。
- (2) 「偏差の 2 乗」の合計を求めましょう。
- (3) 分散（「偏差の 2 乗」の平均値）を求め、表を完成させましょう。

例題 3.2 分散を計算する

データが、

平均値からどれくらい散らばっているか

を表すことを考えます。

このばらつきの大きさを表すために、

分散

という指標を使います。

例題 3.2 分散を計算する

偏差をそのまま足すと，正の偏差と負の偏差が打ち消し合うため，

ばらつきの大きさを表すことができません．

そこで，偏差が打ち消し合わないように，

偏差を 2 乗して，すべて 0 以上の値にします．

例題 3.2 分散を計算する

分散

分散とは、偏差の 2 乗を平均した値のことです。

$$\text{分散} = \frac{\text{偏差の 2 乗の合計}}{\text{個数}}$$

分散が大きいほどばらつきが大きく、分散が小さいほどばらつきが小さいと考えます。

例題 3.2 分散を計算する

(1) 偏差の 2 乗

それぞれの偏差を 2 乗すると,

$$\text{A さん} \quad 5^2 = 25$$

$$\text{B さん} \quad (-15)^2 = 225$$

$$\text{C さん} \quad 15^2 = 225$$

$$\text{D さん} \quad (-5)^2 = 25$$

$$\text{E さん} \quad 0^2 = 0$$

となります。

例題 3.2 分散を計算する

(2) 偏差の 2 乗の合計

偏差の 2 乗をすべて足すと,

$$25 + 225 + 225 + 25 + 0 = 500$$

したがって,

偏差の 2 乗の合計は 500 です.

例題 3.2 分散を計算する

(3) 分散

偏差の 2 乗の合計は 500，データの個数は 5 なので，

$$\frac{500}{5} = 100$$

より，

分散は 100 です．

例題 3.2 分散を計算する

学生	テストの点数	偏差	偏差の 2 乗
A	55	5	25
B	35	-15	225
C	65	15	225
D	45	-5	25
E	50	0	0
合計	250	0	500
平均値	50	0	100 (分散)

例題 3.2 分散を計算する

分散の意味

分散は、偏差を 2 乗してから平均することで、正負の偏差が打ち消し合うことなく、平均値からのばらつきの大きさを表すことができます。

例題 3.2 分散を計算する

分散の単位

分散は偏差を 2 乗して平均をとって求めるため、単位も、もとの単位の 2 乗になります。

そのため、分散はばらつきを表せますが、もとのデータと単位が異なり、直感的に解釈しにくい場合があります。

例題 3.2 分散を計算する

そこで，ばらつきをもとのデータと同じ単位で表すために，

分散の平方根

を考えます．

これが，このあと扱う標準偏差です．

例題 3.3 分散の公式を使ってみる

例題 3.1, 例題 3.2 で扱ったテストの点数について考えます.

学生	テストの点数 点数の 2 乗
A	55
B	35
C	65
D	45
E	50
合計	250
平均値	50

例題 3.3 分散の公式を使ってみる

- (1) 各値を 2 乗しましょう。
- (2) 「値の 2 乗」の合計と「値の 2 乗」の平均値を求め、表を完成させましょう。
- (3) 平均値の 2 乗を求めましょう。
- (4) 次の公式を使って、分散を求めましょう。

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

例題 3.3 分散の公式を試してみる

(1) 各値を 2 乗する

学生	テストの点数	点数の 2 乗
A	55	$55^2 = 3025$
B	35	$35^2 = 1225$
C	65	$65^2 = 4225$
D	45	$45^2 = 2025$
E	50	$50^2 = 2500$

例題 3.3 分散の公式を試してみる

(2) 「値の 2 乗」の合計と平均値

まず、点数の 2 乗をすべて足します。

$$3025 + 1225 + 4225 + 2025 + 2500 = 13000$$

したがって、点数の 2 乗の合計は 13000 です。

例題 3.3 分散の公式を試してみる

点数の 2 乗の合計は 13000，データの個数は 5 なので，

$$\frac{13000}{5} = 2600$$

より，

「値の 2 乗」の平均値は 2600 です．

例題 3.3 分散の公式を使ってみる

学生	テストの点数	点数の 2 乗
A	55	3025
B	35	1225
C	65	4225
D	45	2025
E	50	2500
合計	250	13000
平均値	50	2600

例題 3.3 分散の公式を使ってみる

(3) 平均値の 2 乗

例題 3.1 より，平均値は 50 です．

したがって，

$$50^2 = 2500$$

より，

平均値の 2 乗は 2500 です．

例題 3.3 分散の公式を使ってみる

(4) 公式を使って分散を求める

公式

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

に代入すると、

$$2600 - 2500 = 100$$

となるので、

分散は 100 です。

これは、例題 3.2 で偏差の 2 乗から求めた分散と同じ値です。

例題 3.3 分散の公式を使ってみる

分散の 2 つの求め方

分散は、次の 2 つの方法で求められます。

方法 1 偏差の 2 乗を使う

「偏差の 2 乗」の平均値

方法 2 値の 2 乗を使う

「値の 2 乗」の平均値 $-$ (平均値)²

第3章 データのばらつき

3.2 標準偏差

例題 3.4 標準偏差を計算する

例題 3.1～例題 3.3 で扱ったテストの点数について、標準偏差を求めましょう。

学生	テストの点数	偏差	偏差の 2 乗
A	55	5	25
B	35	-15	225
C	65	15	225
D	45	-5	25
E	50	0	0
合計	250	0	500
平均値	50	0	100 (分散)

例題 3.4 標準偏差を計算する

標準偏差

標準偏差とは，分散の（負でない）平方根のことです．

$$\text{標準偏差} = \sqrt{\text{分散}}$$

例題 3.4 標準偏差を計算する

分散は、偏差を 2 乗してから平均をとるため、

単位も、もとの単位の 2 乗になります。

そのため、分散の値だけでは、

平均値からどれくらい離れているのか

を直接イメージしにくい面があります。

例題 3.4 標準偏差を計算する

標準偏差は、分散の平方根をとって求めます．分散では、単位がもとの単位の 2 乗になっていましたが、平方根をとることで、

もとのデータと同じ単位

でばらつきの大きさを表すことができます．おおまかには、

**平均値のまわりにどの程度散らばっているかを表す、
ズレの大きさの目安と考えることができます．**

例題 3.4 標準偏差を計算する

標準偏差が大きいほど、

値は平均値から大きく離れる傾向

があります。

標準偏差が小さいほど、

値は平均値の近くに集まる傾向

があります。

例題 3.4 標準偏差を計算する

例題 3.2, 例題 3.3 で求めた分散は,

$$100$$

でした. 100 の (負でない) 平方根をとると,

$$\sqrt{100} = 10$$

より,

標準偏差は 10 です.

例題 3.4 標準偏差を計算する

このデータでは、

平均値は **50 点**

標準偏差は **10 点**

です。標準偏差 10 は、

平均点 **50 点**を中心とした

ばらつきの大きさがおよそ 10 点という目安になります。

例題 3.5 変動係数を計算して比較する

工場で製造された製品 A，製品 B の長さ（mm）は次のとおりです。

製品 A	110	70	130	100	90
製品 B	1010	970	1030	990	1000

- (1) 製品 A，製品 B の長さの平均値をそれぞれ求めましょう。
- (2) 製品 A，製品 B の長さの標準偏差をそれぞれ求めましょう。
- (3) 製品 A，製品 B の長さの変動係数を求め，平均値に対してどちらが相対的にばらつきが大きいかを考えましょう。

例題 3.5 変動係数を計算して比較する

変動係数

変動係数とは、標準偏差を平均値で割ったもので、平均値に対する相対的なばらつきを表す指標です。

$$\text{変動係数} = \frac{\text{標準偏差}}{\text{平均値}}$$

例題 3.5 変動係数を計算して比較する

変動係数は、

0 が意味のある比例尺度のデータなどについて、
相対的なばらつきを比較するときに用います。

例題 3.5 変動係数を計算して比較する

変動係数には注意点があります。

平均値が 0 の場合には、

変動係数を定義できません。

また、平均値が 0 に近い場合には、

適切な指標とならないことがあります。

例題 3.5 変動係数を計算して比較する

(1) 製品 A の平均値

$$\frac{110 + 70 + 130 + 100 + 90}{5} = \frac{500}{5} = 100$$

より,

製品 A の平均値は 100 mm です.

例題 3.5 変動係数を計算して比較する

(1) 製品 B の平均値

$$\frac{1010 + 970 + 1030 + 990 + 1000}{5} = \frac{5000}{5} = 1000$$

より,

製品 B の平均値は 1000 mm です.

例題 3.5 変動係数を計算して比較する

(2) 製品 A の偏差

製品 A の平均値は 100 mm なので、各値から 100 を引いて偏差を求めると、

$$110 - 100 = 10$$

$$70 - 100 = -30$$

$$130 - 100 = 30$$

$$100 - 100 = 0$$

$$90 - 100 = -10$$

となります。

例題 3.5 変動係数を計算して比較する

(2) 製品 A の分散

偏差の 2 乗を平均すると,

$$\frac{10^2 + (-30)^2 + 30^2 + 0^2 + (-10)^2}{5} = \frac{2000}{5} = 400$$

となります。

例題 3.5 変動係数を計算して比較する

(2) 製品 A の標準偏差

分散は 400 なので、

$$\sqrt{400} = 20$$

より、

製品 A の標準偏差は 20 mm です。

例題 3.5 変動係数を計算して比較する

(2) 製品 B の偏差

製品 B の平均値は 1000 mm なので、各値から 1000 を引いて偏差を求めると、

$$1010 - 1000 = 10$$

$$970 - 1000 = -30$$

$$1030 - 1000 = 30$$

$$990 - 1000 = -10$$

$$1000 - 1000 = 0$$

となります。

例題 3.5 変動係数を計算して比較する

(2) 製品 B の分散

偏差の 2 乗を平均すると、

$$\frac{10^2 + (-30)^2 + 30^2 + (-10)^2 + 0^2}{5} = \frac{2000}{5} = 400$$

となります。

例題 3.5 変動係数を計算して比較する

(2) 製品 B の標準偏差

分散は 400 なので,

$$\sqrt{400} = 20$$

より,

製品 B の標準偏差は 20 mm です.

例題 3.5 変動係数を計算して比較する

製品 A と製品 B は、

どちらも標準偏差は 20 mm です.

しかし、平均値は、

製品 A : 100 mm

製品 B : 1000 mm

と大きく異なります.

例題 3.5 変動係数を計算して比較する

標準偏差だけを見ると，

「ばらつきの大きさは同じ」

と判断してしまいそうです．

そこで，

平均値の大きさを考慮してばらつきを比較するために，
変動係数を用います．

例題 3.5 変動係数を計算して比較する

(3) 製品 A の変動係数

標準偏差は 20 mm，平均値は 100 mm なので，

$$\frac{20}{100} = 0.2$$

より，

製品 A の変動係数は 0.2 です。

例題 3.5 変動係数を計算して比較する

(3) 製品 B の変動係数

標準偏差は 20 mm，平均値は 1000 mm なので，

$$\frac{20}{1000} = 0.02$$

より，

製品 B の変動係数は 0.02 です。

例題 3.5 変動係数を計算して比較する

	製品 A	製品 B
平均値	100 mm	1000 mm
標準偏差	20 mm	20 mm
変動係数	0.2	0.02

この例では、標準偏差は同じですが、

平均値を基準にしたばらつきは
製品 A のほうが大きい

ことがわかります。

例題 3.5 変動係数を計算して比較する

変動係数の特徴

変動係数は、

$$\frac{\text{標準偏差}}{\text{平均値}}$$

という比なので、単位をもちません．そのため、変動係数を用いることが適切なデータでは、

- 平均値の大きさが異なるデータ
- 単位が異なるデータ

についても、相対的なばらつきを比較できます．

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

例題 4.1 値を変換したときの平均値の変化を調べる

あるクラスの 5 人の通学時間（分）は次のとおりです．

学生	通学時間（分）
A	65
B	5
C	35
D	50
E	95

例題 4.1 値を変換したときの平均値の変化を調べる

- (1) 5 人の平均通学時間を求めましょう。
- (2) すべての通学時間に 10 分を加えたとき、平均値がどのように変化するかを調べましょう。
- (3) すべての通学時間を 2 倍したとき、平均値がどのように変化するかを調べましょう。

例題 4.1 値を変換したときの平均値の変化を調べる

(1) 5 人の平均通学時間

まず、5 人の通学時間をすべて足します。

$$65 + 5 + 35 + 50 + 95 = 250$$

したがって、通学時間の合計は 250 分です。

例題 4.1 値を変換したときの平均値の変化を調べる

通学時間の合計は 250 分，データの個数は 5 なので，

$$\frac{250}{5} = 50$$

より，

平均通学時間は 50 分です。

例題 4.1 値を変換したときの平均値の変化を調べる

(2) 10 分を加えたときの平均値

通学時間は、

65, 5, 35, 50, 95

から

75, 15, 45, 60, 105

に変わります。

例題 4.1 値を変換したときの平均値の変化を調べる

もとの 5 人の合計は 250 分です．5 人それぞれに 10 分を加えるので，新しい合計は

$$250 + 10 \times 5 = 300$$

となります．したがって，合計は

$$10 \times 5 = 50 \text{ 分増えます.}$$

例題 4.1 値を変換したときの平均値の変化を調べる

変換後の合計は 300 分，データの個数は 5 なので，

$$\frac{300}{5} = 60$$

より，

平均値は 60 分です．

例題 4.1 値を変換したときの平均値の変化を調べる

もとの平均値は 50 分，変換後の平均値は 60 分です．

したがって，

$$50 \longrightarrow 60$$

平均値も 10 分増えました．

例題 4.1 値を変換したときの平均値の変化を調べる

(3) 2 倍したときの平均値

通学時間は、

65, 5, 35, 50, 95

から

130, 10, 70, 100, 190

に変わります。

例題 4.1 値を変換したときの平均値の変化を調べる

もとの 5 人の合計は 250 分です．すべての値を 2 倍するので，新しい合計は

$$250 \times 2 = 500$$

となります．したがって，合計は

2 倍になります．

例題 4.1 値を変換したときの平均値の変化を調べる

変換後の合計は 500 分，データの個数は 5 なので，

$$\frac{500}{5} = 100$$

より，

平均値は 100 分です．

例題 4.1 値を変換したときの平均値の変化を調べる

もとの平均値は 50 分，変換後の平均値は 100 分です．

したがって，

$$50 \longrightarrow 100$$

平均値も 2 倍になりました．

例題 4.1 値を変換したときの平均値の変化を調べる

値を加えたとき

すべての値に一定の数 a を加えると,

平均値にも a が加わります.

今回の例では,

$$50 + 10 = 60$$

となりました.

例題 4.1 値を変換したときの平均値の変化を調べる

一定の数を使ったとき

すべての値を一定の数 b 倍すると、

平均値も b 倍になります。

今回の例では、

$$50 \times 2 = 100$$

となりました。

例題 4.2 値を変換したときの分散の変化を調べる

例題 4.1 と同じ通学時間データについて考えます.

学生	通学時間 (分)
A	65
B	5
C	35
D	50
E	95
平均値	50

例題 4.2 値を変換したときの分散の変化を調べる

- (1) もとのデータの分散を求めましょう。
- (2) すべての通学時間に 10 分を加えたとき、分散がどのように変化するかを調べましょう。
- (3) すべての通学時間を 2 倍したとき、分散がどのように変化するかを調べましょう。

例題 4.2 値を変換したときの分散の変化を調べる

(1) もとのデータの偏差

平均値は 50 分なので、各値から 50 を引いて偏差を求めると、

$$65 - 50 = 15$$

$$5 - 50 = -45$$

$$35 - 50 = -15$$

$$50 - 50 = 0$$

$$95 - 50 = 45$$

となります。

例題 4.2 値を変換したときの分散の変化を調べる

(1) もとのデータの分散

偏差の 2 乗を平均すると、

$$\frac{15^2 + (-45)^2 + (-15)^2 + 0^2 + 45^2}{5} = \frac{4500}{5} = 900$$

となります。したがって、

分散は 900 です。

例題 4.2 値を変換したときの分散の変化を調べる

(2) 10 分を加えたときの分散

すべての値に 10 分を加えると、平均値にも 10 分が加わります。
もとの平均値は 50 分なので、

$$50 + 10 = 60$$

より、変換後の平均値は 60 分です。

例題 4.2 値を変換したときの分散の変化を調べる

すべての値と平均値に同じ 10 分が加わるので、

$$\text{変換後の偏差} = (\text{値} + 10) - (\text{平均値} + 10) = \text{値} - \text{平均値}$$

となります。つまり、

偏差は変わりません。

例題 4.2 値を変換したときの分散の変化を調べる

実際、変換前の偏差は、

$$15, \quad -45, \quad -15, \quad 0, \quad 45$$

でしたが、10 分を加えたあとも、

$$15, \quad -45, \quad -15, \quad 0, \quad 45$$

のままです。

例題 4.2 値を変換したときの分散の変化を調べる

偏差が変わらないので、偏差の 2 乗も変わりません。したがって、

変換後の分散も 900 のまま

です。つまり、

すべての値に 10 を加えても、
分散は変化しません。

例題 4.2 値を変換したときの分散の変化を調べる

(3) 2 倍したときの分散

すべての値を 2 倍すると、平均値も 2 倍になります.

もとの平均値は 50 分なので,

$$50 \times 2 = 100$$

より、変換後の平均値は 100 分です.

例題 4.2 値を変換したときの分散の変化を調べる

すべての値と平均値が 2 倍になるので、

$$\text{変換後の偏差} = 2 \times \text{値} - 2 \times \text{平均値} = 2(\text{値} - \text{平均値})$$

となります。したがって、

偏差も 2 倍になります。

例題 4.2 値を変換したときの分散の変化を調べる

実際、もとの偏差は、

$$15, \quad -45, \quad -15, \quad 0, \quad 45$$

でしたが、2 倍すると、

$$30, \quad -90, \quad -30, \quad 0, \quad 90$$

となります。

例題 4.2 値を変換したときの分散の変化を調べる

変換後の偏差は，変換前の偏差の 2 倍なので，

$$(\text{変換後の偏差})^2 = (2 \times \text{変換前の偏差})^2 = 4 \times (\text{変換前の偏差})^2$$

となります．したがって，

偏差の 2 乗は 4 倍になります．

例題 4.2 値を変換したときの分散の変化を調べる

分散は「偏差の 2 乗」の平均値なので、分散も 4 倍になります。

$$900 \times 4 = 3600$$

より、

変換後の分散は 3600

になります。

例題 4.2 値を変換したときの分散の変化を調べる

一定の数を加えたとき

すべての値に一定の数 a を加えても、

分散は変わりません。

値と平均値が同じだけ変化するため、偏差が変わらないからです。

例題 4.2 値を変換したときの分散の変化を調べる

一定の数を使ったとき

すべての値を一定の数 b 倍すると、

分散は b^2 倍になります。

偏差が b 倍になり、偏差の 2 乗が b^2 倍になるためです。

例題 4.3 値を変換したときの標準偏差の変化を調べる

例題 4.1, 例題 4.2 と同じ通学時間データについて考えます。

学生	通学時間 (分)
A	65
B	5
C	35
D	50
E	95
平均値	50
分散	900

例題 4.3 値を変換したときの標準偏差の変化を調べる

- (1) もとのデータの標準偏差を求めましょう。
- (2) すべての通学時間に 10 分を加えたとき、標準偏差がどのように変化するかを調べましょう。
- (3) すべての通学時間を 2 倍したとき、標準偏差がどのように変化するかを調べましょう。

例題 4.3 値を変換したときの標準偏差の変化を調べる

(1) もとの標準偏差

分散は 900 なので,

$$\sqrt{900} = 30$$

より,

標準偏差は 30 分です.

例題 4.3 値を変換したときの標準偏差の変化を調べる

(2) 10 分を加えたときの標準偏差

例題 4.2 で確認したように、すべての値に一定の数を加えても、

分散は変わりません。

したがって、分散は 900 のままです。

例題 4.3 値を変換したときの標準偏差の変化を調べる

変換後の分散も 900 なので，変換後の標準偏差は

$$\sqrt{900} = 30$$

となります．つまり，

変換後の標準偏差も 30 分のままです．

例題 4.3 値を変換したときの標準偏差の変化を調べる

このように，すべての値に 10 分を加えても，

標準偏差は変化しません．

これは，すべての値と平均値が同じだけ移動し，平均値からのばらつきの大きさは変わらないためです．

例題 4.3 値を変換したときの標準偏差の変化を調べる

(3) 2 倍したときの標準偏差

例題 4.2 より，すべての値を 2 倍すると，分散は $2^2 = 4$ 倍になります．

もとの分散は 900 なので，変換後の分散は

$$900 \times 4 = 3600$$

となります．

例題 4.3 値を変換したときの標準偏差の変化を調べる

変換後の分散は 3600 なので,

$$\sqrt{3600} = 60$$

より,

変換後の標準偏差は 60 分です.

例題 4.3 値を変換したときの標準偏差の変化を調べる

(3) なぜ 2 倍になるのか

分散が 4 倍になると，標準偏差の倍率は 4 の負でない平方根なので，

$$\sqrt{4} = 2$$

より，

標準偏差も 2 倍になります。

今回の例では，

30 分 から 60 分

となりました。

例題 4.3 値を変換したときの標準偏差の変化を調べる

一定の数を加えたとき

すべての値に一定の数 a を加えても、

標準偏差は変わりません。

データ全体の位置は移動しますが、ばらつきの大きさは変わらないためです。

例題 4.3 値を変換したときの標準偏差の変化を調べる

一定の数のかけたとき

すべての値を一定の数 b 倍すると、

標準偏差は $|b|$ 倍になります。

分散が b^2 倍になり、標準偏差は $\sqrt{\text{分散}}$ なので、

$$\sqrt{b^2} = |b|$$

より、標準偏差は $|b|$ 倍になります。

第4章 データの変換とデータの標準化

4.2 データの標準化

例題 4.4 z 値を用いて比較する

ある高校で、1 年生の身長と体重について調べたところ、次のような平均値と標準偏差が得られました。

	平均値	標準偏差
身長 (cm)	165	4
体重 (kg)	55	10

A さんは、身長 167 cm、体重 72 kg で、B さんは、身長 158 cm、体重 45 kg でした。

例題 4.4 z 値を用いて比較する

- (1) A さんの身長 167 cm の z 値と，体重 72 kg の z 値を求めましょう。
- (2) (1) で求めた z 値を比べて，A さんについて，この学年のなかで相対的な位置がより高いといえるのは，身長と体重のどちらか答えましょう。
- (3) B さんの身長 158 cm の z 値と，体重 45 kg の z 値を求めましょう。
- (4) (3) で求めた z 値を比べて，B さんについて，この学年のなかで相対的な位置がより低いといえるのは，身長と体重のどちらか答えましょう。

例題 4.4 z 値を用いて比較する

標準化

標準化とは、標準偏差が 0 でないデータを、

平均値 0，標準偏差 1 となるように変換すること

です．

例題 4.4 z 値を用いて比較する

標準化を行うのは、

平均やばらつきが異なるデータどうしても、
値の位置を同じ基準で比べられるようにするためです。

例題 4.4 z 値を用いて比較する

z 値

標準化によって得られる値を z 値といいます。

z 値は、

$$z = \frac{\text{値} - \text{平均値}}{\text{標準偏差}}$$

で求めます。

例題 4.4 z 値を用いて比較する

z 値は,

その値が平均値よりどれだけ上か下かを,
標準偏差を単位として表したものです.

例題 4.4 z 値を用いて比較する

z 値の符号を見ると，平均値との位置関係がわかります．

$$z > 0 \implies \text{平均値より上}$$

$$z = 0 \implies \text{平均値と同じ}$$

$$z < 0 \implies \text{平均値より下}$$

例題 4.4 z 値を用いて比較する

また， z 値の絶対値は，

平均値から標準偏差何個分だけ
離れているか

を表します．たとえば，

$$z = 1.5$$

なら，平均値より標準偏差 1.5 個分だけ上に位置します．

例題 4.4 z 値を用いて比較する

(1) A さんの身長 z 値

身長の平均値は 165 cm，標準偏差は 4 cm です．

A さんの身長は 167 cm なので，

$$z_{\text{身長},A} = \frac{167 - 165}{4} = 0.5$$

となります．

例題 4.4 z 値を用いて比較する

A さんの身長 z 値は,

$$z_{\text{身長},A} = 0.5$$

なので,

**A さんの身長は、平均より
標準偏差 0.5 個分だけ上に位置します。**

例題 4.4 z 値を用いて比較する

(1) A さんの体重の z 値

体重の平均値は 55 kg，標準偏差は 10 kg です．

A さんの体重は 72 kg なので，

$$z_{\text{体重},A} = \frac{72 - 55}{10} = 1.7$$

となります．

例題 4.4 z 値を用いて比較する

A さんの体重の z 値は,

$$z_{\text{体重},A} = 1.7$$

なので,

**A さんの体重は, 平均より
標準偏差 1.7 個分だけ上に位置します.**

例題 4.4 z 値を用いて比較する

(2) A さんの身長と体重を比較する

A さんの z 値は,

$$z_{\text{身長},A} = 0.5, \quad z_{\text{体重},A} = 1.7$$

です. $1.7 > 0.5$ なので,

A さんでは, 相対的な位置がより高いのは体重

です.

例題 4.4 z 値を用いて比較する

(3) B さんの身長 z 値

B さんの身長は 158 cm なので、

$$z_{\text{身長},B} = \frac{158 - 165}{4} = -1.75$$

となります。

B さんの身長は、平均より標準偏差 1.75 個分だけ下に位置します。

例題 4.4 z 値を用いて比較する

(3) B さんの体重の z 値

B さんの体重は 45 kg なので、

$$z_{\text{体重},B} = \frac{45 - 55}{10} = -1$$

となります。

B さんの体重は、平均より標準偏差 1 個分だけ下に位置します。

例題 4.4 z 値を用いて比較する

(4) B さんの身長と体重を比較する

B さんの z 値は,

$$z_{\text{身長},B} = -1.75, \quad z_{\text{体重},B} = -1$$

です. $-1.75 < -1$ なので,

B さんでは, 相対的な位置がより低いのは身長です.

例題 4.4 z 値を用いて比較する

z 値を使った比較

z 値を使うと、身長と体重のように、単位やばらつきが異なるデータであっても、

平均値からの相対的な位置を
同じ基準で比較できます。

例題 4.5 偏差値を用いて比較する

ある高校で、2 年生の数学と地理のテストについて、次のような平均値と標準偏差が得られました。

	平均値	標準偏差
数学（200 点満点）	132	28
地理（100 点満点）	68	12

K くんは、数学で 160 点、地理で 75 点を取りました。

例題 4.5 偏差値を用いて比較する

- (1) K さんの数学と地理の z 値をそれぞれ求めましょう。なお、 z 値は、それまでに割り切れなかった場合は小数第 1 位まで求めます。
- (2) (1) で求めた z 値を用いて、数学と地理の偏差値を求め、偏差値がより高い科目を答えましょう。

例題 4.5 偏差値を用いて比較する

偏差値

データを,

平均値 50, 標準偏差 10

となるように変換した値を偏差値といいます.

例題 4.5 偏差値を用いて比較する

偏差値は、

$$\text{偏差値} = 50 + 10z$$

で求めます。つまり、

z 値を 10 倍して 50 を加える

という変換です。

例題 4.5 偏差値を用いて比較する

z 値と偏差値の対応を確認すると、

z 値	偏差値	意味
0	50	平均と同じ
1	60	平均より 1 SD 上
-1	40	平均より 1 SD 下

なお、SD は標準偏差を表します。

例題 4.5 偏差値を用いて比較する

(1) 数学の z 値

数学の平均値は 132 点，標準偏差は 28 点です．

K くんは 160 点なので，

$$z_{\text{数学},K} = \frac{160 - 132}{28} = 1$$

となります．

例題 4.5 偏差値を用いて比較する

数学の z 値は,

$$z_{\text{数学},K} = 1$$

なので,

**K さんの数学の点数は、平均より
標準偏差 1 個分だけ上に位置します。**

例題 4.5 偏差値を用いて比較する

(1) 地理の z 値

地理の平均値は 68 点，標準偏差は 12 点です．

K くんは 75 点なので，

$$z_{\text{地理},K} = \frac{75 - 68}{12} = 0.58 \dots \approx 0.6$$

となります．

例題 4.5 偏差値を用いて比較する

地理の z 値は,

$$z_{\text{地理},K} \approx 0.6$$

なので,

**K さんの地理の点数は、平均より
標準偏差 0.6 個分だけ上に位置します。**

例題 4.5 偏差値を用いて比較する

(2) 数学の偏差値

数学の z 値は 1 なので,

$$50 + 10 \times 1 = 60$$

より,

数学の偏差値は 60 です.

例題 4.5 偏差値を用いて比較する

(2) 地理の偏差値

地理の z 値は約 0.6 なので、

$$50 + 10 \times 0.6 = 56$$

より、

地理の偏差値は 56 です。

例題 4.5 偏差値を用いて比較する

求めた偏差値は、

数学：60

地理：56

です。したがって、

偏差値がより高いのは数学です。

例題 4.5 偏差値を用いて比較する

z 値と偏差値

z 値は、平均値からどの方向にどれだけずれているかを、標準偏差を単位として表したものです。

偏差値は、その z 値を平均値 50、標準偏差 10 となるように表し直したものです。

例題 4.6 z 値や偏差値からもとの値を読み取る

あるデータの平均値は 60，標準偏差は 20 です．次の z 値や偏差値に対応するもとの値を求めましょう．

(1) z 値が 0 のときの値

(2) z 値が 1 のときの値

(3) z 値が -2 のときの値

(4) 偏差値が 50 のときの値

(5) 偏差値が 70 のときの値

(6) 偏差値が 40 のときの値

例題 4.6 z 値や偏差値からもとの値を読み取る

標準化された値の位置づけ

z 値と偏差値は，どちらもデータの値が平均値に対してどの位置にあるかを表す値です．

z 値は，平均値が 0，標準偏差が 1 となるように変換した値で，偏差値は，平均値が 50，標準偏差が 10 となるように変換した値です．

例題 4.6 z 値や偏差値からもとの値を読み取る

z 値から，平均値との位置関係を読み取れます．

z 値	位置
0	平均と同じ
1	平均より 1 SD 上
-1	平均より 1 SD 下
2	平均より 2 SD 上
-2	平均より 2 SD 下

例題 4.6 z 値や偏差値からもとの値を読み取る

偏差値との対応は、次のようになります。

z 値	-2	-1	0	1	2
偏差値	30	40	50	60	70

つまり、

偏差値 50 が平均に対応

します。

例題 4.6 z 値や偏差値からもとの値を読み取る

z 値からもとの値を求める

標準化の式

$$z = \frac{\text{値} - \text{平均値}}{\text{標準偏差}}$$

を逆向きに使うと、

$$\text{値} = \text{平均値} + z \times \text{標準偏差}$$

となります。

例題 4.6 z 値や偏差値からもとの値を読み取る

偏差値から z 値を求める

偏差値の式

$$\text{偏差値} = 50 + 10z$$

を逆向きに使うと、

$$z = \frac{\text{偏差値} - 50}{10}$$

となります。

例題 4.6 z 値や偏差値からもとの値を読み取る

今回のデータでは、平均値が 60，標準偏差が 20 なので，

$$\text{もとの値} = 60 + z \times 20$$

と考えることができます．

つまり， z 値 1 につき，もとの値では 20 ずつ移動します．

例題 4.6 z 値や偏差値からもとの値を読み取る

(1) $z = 0$ のとき

$z = 0$ は,

平均と同じ位置です.

平均値は 60 なので,

値は 60 です.

例題 4.6 z 値や偏差値からもとの値を読み取る

(2) $z = 1$ のとき

$z = 1$ は、平均より標準偏差 1 個分上です。したがって、もとの値は

$$60 + 1 \times 20 = 80$$

と求められます。つまり、

値は 80 です。

例題 4.6 z 値や偏差値からもとの値を読み取る

(3) $z = -2$ のとき

$z = -2$ は，平均より標準偏差 2 個分下です．したがって，もとの値は

$$60 - 2 \times 20 = 20$$

と求められます．つまり，

値は 20 です．

例題 4.6 z 値や偏差値からもとの値を読み取る

(4) 偏差値 50 のとき

偏差値 50 は,

$$z = 0$$

に対応します. つまり平均と同じ位置なので,

値は 60 です.

例題 4.6 z 値や偏差値からもとの値を読み取る

(5) 偏差値 70 のとき

偏差値 70 は,

$$z = 2$$

に対応します. したがって, もとの値は

$$60 + 2 \times 20 = 100$$

と求められます. つまり,

値は 100 です.

例題 4.6 z 値や偏差値からもとの値を読み取る

(6) 偏差値 40 のとき

偏差値 40 は、

$$z = -1$$

に対応します。したがって、もとの値は

$$60 - 1 \times 20 = 40$$

と求められます。つまり、

値は 40 です。

例題 4.6 z 値や偏差値からもとの値を読み取る

今回の結果を並べると、

z 値	-2	-1	0	1	2
偏差値	30	40	50	60	70
もとの値	20	40	60	80	100

第5章 相関

5.1 共分散

例題 5.1 散布図から相関を読み取る

同じクラスの生徒 11 人について、1 日の勉強時間とテストの点数を調べたところ、次のようになりました。

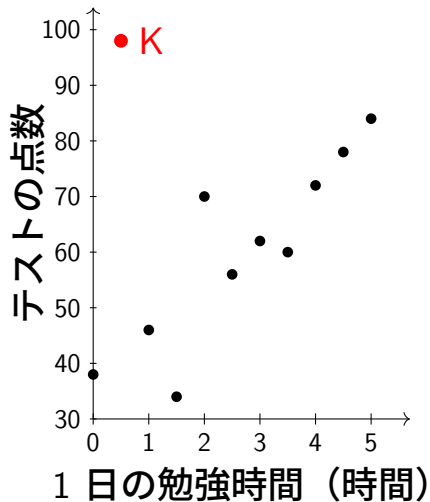
例題 5.1 散布図から相関を読み取る

生徒	勉強時間（時間）	テストの点数
A	2.5	56
B	0.0	38
C	4.0	72
D	1.0	46
E	5.0	84
F	3.5	60
G	1.5	34
H	2.0	70
I	4.5	78
J	3.0	62
K	0.5	98

例題 5.1 散布図から相関を読み取る

この 2 つの変数の関係を，散布図を使って見てみましょう．

例題 5.1 散布図から相関を読み取る



例題 5.1 散布図から相関を読み取る

(1) この散布図は，正の相関と負の相関のどちらを表しているように見えるでしょうか．

(2) 生徒 K のデータ

(0.5 時間, 98 点)

を除くと，相関は強くなるか弱くなるか判断しましょう．

例題 5.1 散布図から相関を読み取る

散布図

散布図は、2 つの変数についてのデータの組（ペア）を，平面上に点として表すグラフです．

点の位置や並び方を見ることで，

2 つの変数の関係の傾向

を視覚的にとらえることができます．

例題 5.1 散布図から相関を読み取る

散布図では、

- 2 つの変数がどのように関係しているか
- 点がどの程度まとまって並んでいるか
- 他の点から大きく外れた点がないか

などを確認できます。

例題 5.1 散布図から相関を読み取る

相関

相関とは、

2 つの変数の間にある直線的な関係を表すものです。

点が一直線に近く並ぶほど相関は強く，点の並びがばらつくほど相関は弱くなります。

例題 5.1 散布図から相関を読み取る

正の相関

横軸の値が大きくなるにつれて、縦軸の値も大きくなる傾向があるとき、**正の相関**があるといえます。

右上がり

に点が並ぶイメージです。

例題 5.1 散布図から相関を読み取る

負の相関

横軸の値が大きくなるにつれて、縦軸の値が小さくなる傾向があるとき、**負の相関**があるといえます。

右下がり

に点が並ぶイメージです。

例題 5.1 散布図から相関を読み取る

(1) 相関の向きを判断する

全体として、勉強時間が長くなるほどテストの点数も高くなる傾向があります。

点はおおむね右上がりに並んでいるので、

正の相関があると判断できます。

例題 5.1 散布図から相関を読み取る

(2) 相関の強さの変化を判断する

生徒 K のデータは、

$(0.5, 98)$

です。この点は、

他の点の傾向から大きく外れています。

散布図では、このような外れた点も見つけやすくなります。

例題 5.1 散布図から相関を読み取る

生徒 K の点を除くと，残りの点は，より一直線に近く並びます．

したがって，

相関はより強くなる

と判断できます．

例題 5.1 散布図から相関を読み取る

相関を見るときに注意

2 つの変数の間に相関があっても、

一方が他方の原因であるとは限りません。

相関と因果関係は区別して考える必要があります。

第三の要因が影響している場合もあります。

例題 5.2 共分散を求める

5 人の生徒について、1 日の勉強時間とテストの点数を調べました。
テストは 1 問 2 点で採点しています。

生徒	勉強時間	点数 (1 問 2 点)
A	8	72
B	10	80
C	12	88
D	6	64
E	14	96

例題 5.2 共分散を求める

例題 5.1 では、散布図を見て 2 つの変数の関係を調べました。
今度は、

2 つの変数の関係を数値で表す

ことを考えます。

そのために用いるのが共分散です。

例題 5.2 共分散を求める

- (1) 勉強時間と点数の平均値を求めましょう。
- (2) 各生徒について、勉強時間の偏差と点数の偏差を求めましょう。
- (3) 各生徒について、2 つの偏差の積を求めましょう。
- (4) 偏差どうしの積をすべて足し、人数で割って共分散を求めましょう。
- (5) 求めた共分散の符号から、勉強時間と点数の関係は**正の相関**と**負の相関**のどちらと考えられるか判断しましょう。

例題 5.2 共分散を求める

共分散

2 つの変数について、それぞれの偏差どうしを掛け合わせ、その平均値を求めたものを共分散といいます。

$$\text{共分散} = \frac{\text{偏差どうしの積の合計}}{\text{データの個数}}$$

例題 5.2 共分散を求める

偏差どうしの積の符号に注目します。

偏差がどちらも正（平均値より大きい），または，どちらも負（平均値より小さい）なら，

偏差どうしの積は正

になります。一方の偏差が正でもう一方が負なら，

偏差どうしの積は負

になります。

例題 5.2 共分散を求める

偏差の符号がそろっているときは，2 つの値が

同じ動き

をし，符号がそろわないときは，

逆の動き

をします．そのため，偏差どうしの積の平均である共分散の符号から，

相関の向き

を読み取ることができます．

例題 5.2 共分散を求める

(1) 勉強時間の平均値

5 人の勉強時間の合計を 5 で割ると、

$$\frac{8 + 10 + 12 + 6 + 14}{5} = 10$$

より、

平均値は 10 時間です。

例題 5.2 共分散を求める

(1) 点数の平均値

5 人の点数の合計を 5 で割ると、

$$\frac{72 + 80 + 88 + 64 + 96}{5} = 80$$

より、

平均値は 80 点です。

例題 5.2 共分散を求める

(2) それぞれの偏差

それぞれの値から平均値を引いて、偏差を求めると、次のようになります。

生徒	勉強時間の偏差	点数の偏差
A	-2	-8
B	0	0
C	2	8
D	-4	-16
E	4	16

例題 5.2 共分散を求める

たとえば A さんについて、勉強時間の偏差は、

$$8 - 10 = -2$$

点数の偏差は、

$$72 - 80 = -8$$

です。2つの偏差は同じ符号なので、

勉強時間と点数は同じ動きをしています。

例題 5.2 共分散を求める

(3) 偏差どうしの積

$$\text{A さん} \quad (-2) \times (-8) = 16$$

$$\text{B さん} \quad 0 \times 0 = 0$$

$$\text{C さん} \quad 2 \times 8 = 16$$

$$\text{D さん} \quad (-4) \times (-16) = 64$$

$$\text{E さん} \quad 4 \times 16 = 64$$

例題 5.2 共分散を求める

(4) 共分散を求める

偏差どうしの積を足すと，

$$16 + 0 + 16 + 64 + 64 = 160$$

となります．共分散は，偏差どうしの積の平均なので，

$$\frac{160}{5} = 32$$

より，

共分散は 32 です．

例題 5.2 共分散を求める

生徒	勉強時間	点数	勉強時間の偏差	点数の偏差	偏差どうしの積
A	8	72	-2	-8	16
B	10	80	0	0	0
C	12	88	2	8	16
D	6	64	-4	-16	64
E	14	96	4	16	64
合計	50	400	0	0	160
平均値	10	80	0	0	32

例題 5.2 共分散を求める

(5) 共分散の符号から判断する

今回の共分散は 32 で、正の値です。

したがって、

正の相関があると考えられます。

例題 5.2 共分散を求める

共分散の符号と相関

共分散が正 $\implies$ 正の相関

共分散が 0 $\implies$ 直線的な関係がない（無相関）

共分散が負 $\implies$ 負の相関

共分散が 0 のときは，正または負の直線的な関係がとらえられないと考えます．

例題 5.3 単位を変えると共分散はどう変わるかを調べる

例題 5.2 と同じ 5 人について，テストを今度は 1 問 1 点で採点します．つまり，例題 5.2 の点数をすべて半分にしたデータです．

生徒	勉強時間	点数 (1 問 1 点)
A	8	36
B	10	40
C	12	44
D	6	32
E	14	48

例題 5.3 単位を変えると共分散はどう変わるかを調べる

例題 5.2 では、

1 問 2 点

でしたが、今回は、

1 問 1 点

です。したがって、点数はすべて

$\frac{1}{2}$ 倍

になっています。

例題 5.3 単位を変えると共分散はどう変わるかを調べる

- (1) 点数の平均値を求めましょう。
- (2) 各生徒について、勉強時間の偏差と点数の偏差を求めましょう。
- (3) 各生徒について、2 つの偏差の積を求めましょう。
- (4) 偏差どうしの積をすべて足し、人数で割って共分散を求めましょう。
- (5) 例題 5.2 の共分散と比べて、どのように変化したかを考えましょう。

例題 5.3 単位を変えると共分散はどう変わるかを調べる

(1) 点数の平均値

5 人の点数の合計を 5 で割ると、

$$\frac{36 + 40 + 44 + 32 + 48}{5} = 40$$

より、

平均値は 40 点です。

例題 5.3 単位を変えると共分散はどう変わるかを調べる

例題 5.2 では，点数の平均値は 80 点でした．

今回は 40 点なので，

$$80 \longrightarrow 40$$

となっています．点数を半分にすると，

平均値も半分になります．

例題 5.3 単位を変えると共分散はどう変わるかを調べる

(2) それぞれの偏差

それぞれの値から平均値を引いて、偏差を求めると、次のようになります。

生徒	勉強時間の偏差	点数の偏差
A	-2	-4
B	0	0
C	2	4
D	-4	-8
E	4	8

例題 5.3 単位を変えると共分散はどう変わるかを調べる

点数の偏差を比べてみましょう。

例題 5.2 では、

$$-8, \quad 0, \quad 8, \quad -16, \quad 16$$

今回は、

$$-4, \quad 0, \quad 4, \quad -8, \quad 8$$

です。

点数の偏差もすべて半分になっています。

例題 5.3 単位を変えると共分散はどう変わるかを調べる

(3) 偏差どうしの積

$$\text{A さん} \quad (-2) \times (-4) = 8$$

$$\text{B さん} \quad 0 \times 0 = 0$$

$$\text{C さん} \quad 2 \times 4 = 8$$

$$\text{D さん} \quad (-4) \times (-8) = 32$$

$$\text{E さん} \quad 4 \times 8 = 32$$

例題 5.3 単位を変えると共分散はどう変わるかを調べる

偏差どうしの積を比べてみましょう。

例題 5.2 では、

16, 0, 16, 64, 64

今回は、

8, 0, 8, 32, 32

です。

偏差どうしの積もすべて半分になっています。

例題 5.3 単位を変えると共分散はどう変わるかを調べる

(4) 共分散を求める

偏差どうしの積を足すと，

$$8 + 0 + 8 + 32 + 32 = 80$$

となります．共分散は，偏差どうしの積の平均なので，

$$\frac{80}{5} = 16$$

より，

共分散は 16 です．

例題 5.3 単位を変えると共分散はどう変わるかを調べる

生徒	勉強時間	点数	勉強時間の偏差	点数の偏差	偏差どうしの積
A	8	36	-2	-4	8
B	10	40	0	0	0
C	12	44	2	4	8
D	6	32	-4	-8	32
E	14	48	4	8	32
合計	50	200	0	0	80
平均値	10	40	0	0	16

例題 5.3 単位を変えると共分散はどう変わるかを調べる

(5) 共分散を比較する

例題 5.2 では,

$$\text{共分散} = 32$$

今回は,

$$\text{共分散} = 16$$

です。したがって,

共分散も半分になりました。

例題 5.3 単位を変えると共分散はどう変わるかを調べる

点数を半分にすると，点数の偏差も半分になります．

そのため，偏差どうしの積も半分になり，共分散も

半分になります．

例題 5.3 単位を変えると共分散はどう変わるかを調べる

共分散は、

$$32 \longrightarrow 16$$

と変化しましたが，どちらも正の値です．

したがって，関係の向きはどちらも，

正の相関のままです．

例題 5.3 単位を変えると共分散はどう変わるかを調べる

共分散と単位

片方の変数の値を k 倍すると、その変数の偏差も k 倍になり、

共分散も k 倍になります。

例題 5.3 単位を変えると共分散はどう変わるかを調べる

共分散の注意点

共分散の値の大きさは、

データの単位に左右されます。

そのため、共分散の大きさだけで相関の強さを比較することはできません。

例題 5.3 単位を変えると共分散はどう変わるかを調べる

では,

単位の影響を受けずに,
相関の強さを表すにはどうすればよいでしょうか.

そのために用いるのが,

相関係数です.

第5章 相関

5.2 相関係数

例題 5.4 相関係数を求める

例題 5.2 と同じデータについて，以下の問いに答え，勉強時間とテストの点数（1 問 2 点）の相関係数を求めてみましょう．

生徒	勉強時間	点数	勉強時間の偏差	点数の偏差
A	8	72	-2	-8
B	10	80	0	0
C	12	88	2	8
D	6	64	-4	-16
E	14	96	4	16

例題 5.4 相関係数を求める

共分散は 32 であることがわかっています。

- (1) 勉強時間の標準偏差と、点数（1 問 2 点）の標準偏差を求めましょう。
- (2) 勉強時間と点数（1 問 2 点）の相関係数を求めましょう。
- (3) 求めた相関係数から、勉強時間と点数（1 問 2 点）の関係について読み取れることを述べましょう。

例題 5.4 相関係数を求める

まず、相関係数がどのような値かを確認します。

例題 5.3 では、点数を半分になると、

$$32 \longrightarrow 16$$

と、共分散も半分になりました。

つまり、共分散の大きさは、

データの単位の影響を受けます。

例題 5.4 相関係数を求める

そこで、

単位の影響を受けずに、
相関の強さを表す

ために用いるのが、

相関係数です。

例題 5.4 相関係数を求める

相関係数

2 つの変数の標準偏差がともに 0 でないとき，相関係数 r は，共分散をそれぞれの標準偏差の積で割ったものです。

$$r = \frac{\text{共分散}}{1 \text{ 目データの標準偏差} \times 2 \text{ 目データの標準偏差}}$$

例題 5.4 相関係数を求める

相関係数は、

2 つの変数の相関の強さを表す指標です。

また、相関係数は、

単位の影響を受けません。

例題 5.4 相関係数を求める

相関係数 r の値は、つねに

$$-1 \leq r \leq 1$$

の範囲に収まります。

$$\textcolor{red}{-1} \quad \longleftarrow \quad 0 \quad \longrightarrow \quad \textcolor{blue}{1}$$

例題 5.4 相関係数を求める

相関係数の符号を見ると，相関の向きがわかります．

$$r > 0 \implies \text{正の相関}$$

$$r < 0 \implies \text{負の相関}$$

例題 5.4 相関係数を求める

相関係数が 1, 0, -1 のどの値に近いかを見ると, 相関の強さがわかります.

$r = 1$ に近い $\implies$ 強い正の相関

$r = 0$ に近い $\implies$ 相関が弱い

$r = -1$ に近い $\implies$ 強い負の相関

例題 5.4 相関係数を求める

相関係数を見るとき注意

相関係数は、直線的な関係の強さを表す指標です。

そのため、曲線的な関係があっても、相関係数が 0 に近い値になることがあります。散布図もあわせて確認することが重要です。

また、相関係数は、**外れ値の影響を受けやすい**という特徴があります。

例題 5.4 相関係数を求める

(1) 標準偏差を求める

まず、それぞれの偏差の 2 乗を求め、その合計を計算します。

生徒	勉強時間の偏差の 2 乗	点数の偏差の 2 乗
A	4	64
B	0	0
C	4	64
D	16	256
E	16	256
合計	40	640

例題 5.4 相関係数を求める

偏差の 2 乗の平均が分散なので、勉強時間の分散は、

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

点数の分散は、

$$\frac{640}{5} = 128$$

となります。

例題 5.4 相関係数を求める

標準偏差は $\sqrt{\text{分散}}$ なので、

勉強時間の標準偏差は $\sqrt{8}$ です.

点数（1 問 2 点）の標準偏差は $\sqrt{128}$ です.

例題 5.4 相関係数を求める

(2) 相関係数を求める

勉強時間の標準偏差は $\sqrt{8}$ ，点数（1 問 2 点）の標準偏差は $\sqrt{128}$ ，また，これらの共分散は 32 なので，相関係数は

$$\begin{aligned} r &= \frac{32}{\sqrt{8}\sqrt{128}} = \frac{32}{(2\sqrt{2})(8\sqrt{2})} = \frac{32}{2 \times 8 \times (\sqrt{2}\sqrt{2})} \\ &= \frac{32}{16 \times 2} = 1 \end{aligned}$$

となります。

例題 5.4 相関係数を求める

(3) 相関係数から読み取る

相関係数は 1 です.

これは,

最も強い正の相関を表します.

例題 5.4 相関係数を求める

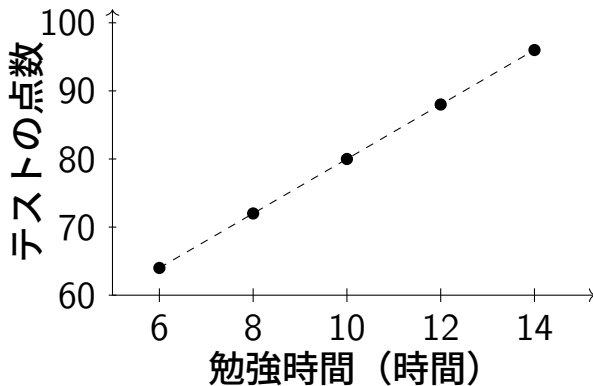
今回の 5 つのデータでは，勉強時間が長くなるほど，点数も高くなっています．

しかも，すべての点が

1 本の直線上に並んでいます．

そのため，相関係数は最大の 1 になります．

例題 5.4 相関係数を求める



すべての点が一直線上に並んでいます。

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

例題 5.3 では、点数を 1 問 1 点に換算しました。

このとき、共分散は、

$$32 \longrightarrow 16$$

と半分になりました。

では、相関係数はどのように変化するでしょうか。

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

例題 5.3 で求めた値を確認します。

生徒	勉強時間の偏差	点数の偏差
A	-2	-4
B	0	0
C	2	4
D	-4	-8
E	4	8

共分散は 16 です。

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

- (1) 点数（1 問 1 点）の標準偏差を求めましょう。
- (2) 勉強時間と点数（1 問 1 点）の相関係数を求めましょう。
- (3) 例題 5.4 で求めた 1 問 2 点の場合の相関係数と比べ、どのように変化したかを考えましょう。

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

(1) 点数の標準偏差を求める

点数の偏差は、

$$-4, \quad 0, \quad 4, \quad -8, \quad 8$$

なので、偏差の 2 乗は、

$$16, \quad 0, \quad 16, \quad 64, \quad 64$$

となります。

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

偏差の 2 乗の合計は、

$$16 + 0 + 16 + 64 + 64 = 160$$

です。したがって、点数の分散は

$$\frac{160}{5} = 32$$

となります。

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

標準偏差は $\sqrt{\text{分散}}$ なので、

点数（1 問 1 点）の標準偏差は $\sqrt{32}$ となります．

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

例題 5.4 では、点数（1 問 2 点）の標準偏差は、 $\sqrt{128}$ でした。
今回は、 $\sqrt{32}$ です。

ここで、

$$\sqrt{32} = \sqrt{\frac{128}{4}} = \frac{1}{2}\sqrt{128}$$

なので、

標準偏差も半分になっています。

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

(2) 相関係数を求める

勉強時間の標準偏差は $\sqrt{8}$ ，点数（1 問 1 点）の標準偏差は $\sqrt{32}$ ，
また，これらの共分散は 16 なので，相関係数は

$$\begin{aligned} r &= \frac{16}{\sqrt{8}\sqrt{32}} = \frac{16}{(2\sqrt{2})(4\sqrt{2})} = \frac{16}{2 \times 4 \times (\sqrt{2}\sqrt{2})} \\ &= \frac{16}{8 \times 2} = \frac{16}{16} = 1 \end{aligned}$$

となります。

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

(3) 相関係数を比較する

つまり，1 問 2 点の場合も，1 問 1 点の場合も， $r = 1$ であり，

相関係数は変わりません．

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

点数を半分になると,

共分散も半分

$$32 \longrightarrow 16$$

標準偏差も半分

$$\sqrt{128} \longrightarrow \sqrt{32} \left(= \frac{1}{2} \sqrt{128} \right)$$

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

相関係数は、

$$r = \frac{\text{共分散}}{\text{標準偏差の積}}$$

です。

共分散と、変換した点数の標準偏差がともに同じ割合で変化するため、その影響が分子と分母で打ち消されます。

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

相関係数と単位

片方の変数を正の数倍しても、

相関係数は変わりません。

そのため、相関係数は、

単位の影響を受けない指標です。

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

縦軸と横軸を入れ替える

2 つの変数の役割を入れ替えると、それぞれの標準偏差は入れ替わりますが、共分散は変わりません．したがって、

相関係数は変わりません．

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

同じ値を足す・引く

データ全体に同じ値を足したり引いたりすると，平均値も同じだけ移動するため，平均値からのズレである偏差は変わりません．

したがって，

相関係数も変わりません．

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

片方のデータを正の数倍する

片方のデータを正の数 k 倍すると、その変数の標準偏差も k 倍になり、共分散も k 倍になります。

相関係数

$$r = \frac{\text{共分散}}{\text{標準偏差の積}}$$

では、分子と分母が同じ割合で変化するので、

相関係数は変わりません。

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

ただし，片方のデータを負の数倍した場合は異なります．

平均値からのズレの向きがすべて反対になるため，偏差どうしの積の符号も反対になります．したがって，

相関係数の符号が反転します．

例題 5.5 単位を変えると相関係数はどう変わるかを調べる

変換	相関係数
縦軸と横軸を入れ替える	変わらない
同じ値を足す・引く	変わらない
片方を正の数倍する	変わらない
片方を負の数倍する	符号が反転

第6章 回帰分析

6.1 回帰分析の基本

例題 6.1 説明変数か目的変数かを判断する

次の 3 つの文は、2 つの変数の関係について述べたものです。ここで示されている関係は、必ずしも因果関係を意味するわけではありません。

文の内容から、どちらを説明変数、どちらを目的変数とみなすかを判断しましょう。

- (1) 「勉強時間が長いほど、テストの点数が高くなる傾向がある」
- (2) 「年齢が高いほど、身長が高くなる」
- (3) 「商品の価格が高いほど、売上個数は少なくなる」

例題 6.1 説明変数か目的変数かを判断する

説明変数と目的変数

ある変数の値をもとに，別の変数の値を予測するとき，
予測に用いる側の変数を，説明変数 x とよびます．
予測の対象となる変数を，目的変数 y とよびます．

説明変数 x をもとに，目的変数 y を予測します．

例題 6.1 説明変数か目的変数かを判断する

説明変数と目的変数の区別は，

分析の目的によって決まります．

したがって，どちらを説明変数とするかがいつでも一意に決まるとは限りません．

例題 6.1 説明変数か目的変数かを判断する

また，説明変数と目的変数を区別したからといって，

**それだけで因果関係があると
判断することはできません．**

あくまで，予測の際の変数の役割を区別しています．

例題 6.1 説明変数か目的変数かを判断する

(1) 勉強時間とテストの点数

勉強時間をもとに，テストの点数を予測すると考えます．

説明変数：勉強時間

目的変数：テストの点数

例題 6.1 説明変数か目的変数かを判断する

(2) 年齢と身長

年齢をもとに，身長を予測すると考えます．

説明変数：年齢

目的変数：身長

例題 6.1 説明変数か目的変数かを判断する

(3) 商品の価格と売上個数

商品の価格をもとに，売上個数を予測すると考えます．

説明変数：商品の価格

目的変数：売上個数

例題 6.2 回帰直線の式の意味を読み取る

あるクラスの生徒について，1 日の勉強時間とテストの点数の関係を調べたところ，次の回帰直線が得られました．

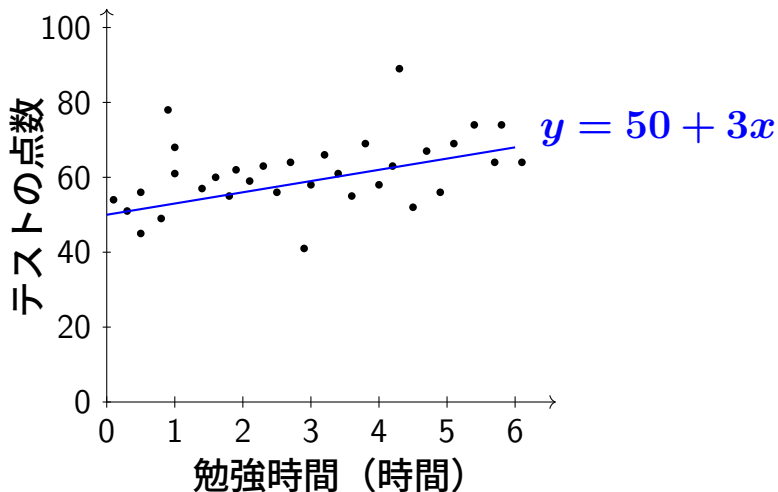
$$y = 50 + 3x$$

ここで， x は説明変数で，1 日の勉強時間（時間）， y は目的変数で，テストの点数です．

例題 6.2 回帰直線の式の意味を読み取る

- (1) 勉強時間が 1 時間増えると、テストの点数はどのように変化すると予測されるでしょうか。
- (2) 勉強時間が 0 時間のとき、テストの点数はどれくらいと予測されるでしょうか。
- (3) 勉強時間が 5 時間のとき、テストの点数はどれくらいと予測されるでしょうか。

例題 6.2 回帰直線の式の意味を読み取る



例題 6.2 回帰直線の式の意味を読み取る

回帰分析

回帰分析では、2 つの変数の関係を直線で近似し、その直線を使って目的変数の値を予測します。

例題 6.2 回帰直線の式の意味を読み取る

回帰直線

2 つの変数の関係を近似した直線を，回帰直線といいます．

回帰直線は，一般に，

$$y = a + bx$$

と表します．

例題 6.2 回帰直線の式の意味を読み取る

回帰直線

$$y = a + bx$$

では、 x は説明変数、 y は目的変数です。

つまり、

x の値をもとに y の値を予測します。

例題 6.2 回帰直線の式の意味を読み取る

傾き

回帰直線

$$y = a + bx$$

の b を傾きといいます。

傾き b は、

x が 1 増えると、
 y がどれだけ変化するとみなせるかを表します。

例題 6.2 回帰直線の式の意味を読み取る

切片

回帰直線

$$y = a + bx$$

の a を切片といいます。

切片 a は、

$x = 0$ のときの y の予測値を表します。

例題 6.2 回帰直線の式の意味を読み取る

今回の回帰直線は、

$$y = 50 + 3x$$

です。この式から、

切片は 50，傾きは 3

と読み取れます。

例題 6.2 回帰直線の式の意味を読み取る

(1) 点数の変化を予測する

回帰直線は、

$$y = 50 + 3x$$

なので、傾きは

3

です。したがって、

勉強時間が1時間増えると、
テストの点数は 3 点高くなると予測されます。

例題 6.2 回帰直線の式の意味を読み取る

(2) 勉強時間が 0 時間の点数を予測する

$y = 50 + 3x$ に $x = 0$ を代入すると,

$$y = 50 + 3 \times 0$$

つまり,

$$y = 50$$

となります。したがって、勉強時間が 0 時間のときのテストの点数は,

50 点と予測されます。

例題 6.2 回帰直線の式の意味を読み取る

(3) 勉強時間が 5 時間の点数を予測する

$y = 50 + 3x$ に $x = 5$ を代入すると、

$$y = 50 + 3 \times 5$$

つまり、

$$y = 65$$

となります。したがって、勉強時間が 5 時間のときのテストの点数は、

65 点と予測されます。

例題 6.2 回帰直線の式の意味を読み取る

回帰直線の読み方

回帰直線

$$y = a + bx$$

傾き b : x が 1 増えたときの y の予測値の変化量

切片 a : $x = 0$ のときの y の予測値

例題 6.2 回帰直線の式の意味を読み取る

回帰直線を利用するときの注意

回帰直線は，直線的な関係をもとに予測します．

- 直線よりも曲線のほうが当てはまりがよい場合があるため，散布図の形を確認する
- グループ差や外れ値の影響を受けることがあるため，必要に応じてグループごとの分析や外れ値の扱いを検討する

例題 6.2 回帰直線の式の意味を読み取る

回帰直線を利用するときの注意

- 観測された範囲を大きく外れた予測（外挿）には注意する
- 相関係数が 0 に近い場合は，直線的な関係が弱いため，回帰直線による予測には注意する

第6章 回帰分析

6.2 回帰直線の推定

例題 6.3 共分散から回帰直線を求める

あるクラスの生徒について，1 日の運動時間（分）を x ，体力テストの得点を y とすると，次の値が得られました．

$$x \text{ の平均値 } \bar{x} = 30, \quad y \text{ の平均値 } \bar{y} = 55$$

$$x \text{ の標準偏差} = 10, \quad \text{共分散} = 60$$

このとき， y を x の値から予測するための回帰直線を求めましょう．

例題 6.3 共分散から回帰直線を求める

回帰直線の傾き

回帰直線

$$y = a + bx$$

の傾き b は,

$$b = \frac{\text{共分散}}{(x \text{ の標準偏差})^2}$$

で求めます。

例題 6.3 共分散から回帰直線を求める

ここで,

$$(x \text{ の標準偏差})^2 = x \text{ の分散}$$

なので,

$$b = \frac{\text{共分散}}{x \text{ の分散}}$$

と表すこともできます.

例題 6.3 共分散から回帰直線を求める

回帰直線の切片

傾き b が求まったら、切片 a は、

$$a = \bar{y} - b\bar{x}$$

で求めます。

ここで、 $\bar{x}$ は x の平均値、 $\bar{y}$ は y の平均値を表します。

例題 6.3 共分散から回帰直線を求める

切片の式

$$a = \bar{y} - b\bar{x}$$

を変形すると,

$$\bar{y} = a + b\bar{x}$$

となります.

つまり, 回帰直線 $y = a + bx$ は,

$(\bar{x}, \bar{y})$ を通ります.

例題 6.3 共分散から回帰直線を求める

傾き b を求める

x の標準偏差は 10，共分散は 60 なので，

$$b = \frac{60}{10^2} = \frac{60}{100} = 0.6$$

より，

傾きは $b = 0.6$ です．

例題 6.3 共分散から回帰直線を求める

切片 a を求める

$$a = \bar{y} - b\bar{x}$$

に,

$$\bar{x} = 30, \quad \bar{y} = 55, \quad b = 0.6$$

を代入すると,

$$a = 55 - 0.6 \times 30 = 55 - 18 = 37$$

より,

切片は $a = 37$ です.

例題 6.3 共分散から回帰直線を求める

$y = a + bx$ に $a = 37$, $b = 0.6$ を代入すると,

$$y = 37 + 0.6x$$

となります．これが，1 日の運動時間 x から体力テストの得点 y を予測する回帰直線です．

例題 6.3 共分散から回帰直線を求める

回帰直線の求め方のまとめ

まず傾き b を,

$$b = \frac{\text{共分散}}{(x \text{ の標準偏差})^2} = \frac{\text{共分散}}{x \text{ の分散}}$$

で求め、次に切片 a を, $a = \bar{y} - b\bar{x}$ で求めます.

そして、求めた a と b を $y = a + bx$ に代入して、回帰直線を求めます.

例題 6.4 相関係数から回帰直線を求める

あるクラスの生徒について，1 日の運動時間（分）を x ，体力テストの得点を y とすると，次の値が得られました．

$$x \text{ の平均値 } \bar{x} = 30, \quad y \text{ の平均値 } \bar{y} = 55$$

$$x \text{ の標準偏差} = 10, \quad y \text{ の標準偏差} = 12$$

$$\text{相関係数} = 0.50$$

このとき， y を x の値から予測するための回帰直線を求めましょう．

例題 6.4 相関係数から回帰直線を求める

相関係数から傾きを求める

回帰直線

$$y = a + bx$$

の傾き b は,

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

で求めることもできます。

例題 6.4 相関係数から回帰直線を求める

傾き b を求める

相関係数は 0.50, x の標準偏差は 10, y の標準偏差は 12 なので,

$$b = 0.50 \times \frac{12}{10} = 0.50 \times 1.2 = 0.6$$

より,

傾きは $b = 0.6$ です.

例題 6.4 相関係数から回帰直線を求める

切片は,

$$a = \bar{y} - b\bar{x}$$

で求めます.

$$\bar{x} = 30, \quad \bar{y} = 55, \quad b = 0.6$$

を代入すると,

$$a = 55 - 0.6 \times 30 = 55 - 18 = 37$$

より,

切片は $a = 37$ です.

例題 6.4 相関係数から回帰直線を求める

$y = a + bx$ に $a = 37$, $b = 0.6$ を代入すると,

$$y = 37 + 0.6x$$

となります．これが，1 日の運動時間 x から体力テストの得点 y を予測する回帰直線です．

例題 6.4 相関係数から回帰直線を求める

回帰直線の傾きの 2 つの求め方のまとめ

共分散から求める

$$b = \frac{\text{共分散}}{(x \text{ の標準偏差})^2}$$

相関係数から求める

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

例題 6.4 相関係数から回帰直線を求める

相関係数と回帰直線の傾き

相関係数は、直線的な関係の強さを表しますが、回帰直線の傾きは、説明変数が 1 増えたときの目的変数の変化量を表します。

したがって、

相関係数の大きさと傾きの大きさは、
別の概念です。

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

あるクラスの生徒について，1 日の運動時間（分）を x ，体力テストの得点を y としたところ，次の回帰直線が得られました．

$$y = 37 + 0.6x$$

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

(1) 運動時間を「分」ではなく「時間」で表すことにします.

そこで,

$$x' = \frac{x}{60}$$

とおきます.

運動時間 (時間) x' から得点 y を予測する回帰直線を求めましょう.

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

(2) 体力テストの得点を 10 分の 1 にした値

$$y' = \frac{y}{10}$$

を考えます。

運動時間 (分) x から得点 y' を予測する回帰直線を求めましょう。

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

(1) 運動時間を「時間」にした回帰直線を求める

$$x' = \frac{x}{60} \quad \text{なので} \quad x = 60x'$$

となります。これを，もとの回帰直線

$$y = 37 + 0.6x$$

に代入します。

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

$x = 60x'$ を代入すると,

$$y = 37 + 0.6(60x')$$

となります. 整理すると,

$$y = 37 + 36x'$$

となります. これが, 運動時間 (時間) x' から体力テストの得点 y を予測する回帰直線です.

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

もとの回帰直線は、

$$y = 37 + 0.6x$$

変換後は、

$$y = 37 + 36x'$$

です。したがって、傾きは、

$$0.6 \longrightarrow 36$$

と変化します。

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

「分」では 1 単位は 1 分ですが、「時間」では 1 単位は 1 時間 = 60 分です。

つまり、説明変数が 1 単位増えることは、もとの単位では 60 増えることに対応します。

そのため、1 単位あたりの変化量である傾きも、

60 倍になります。

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

一方，切片は，

$$37 \longrightarrow 37$$

と変わりません．これは，

$x = 0$ と $x' = 0$ が同じ状態だからです．

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

説明変数の単位変換

説明変数 x の単位を比例的に変えると、

傾きだけが変わります。

切片は変わりません。

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

(2) 得点を 10 分の 1 にした回帰直線を求める

$$y' = \frac{y}{10}$$

とします.

もとの回帰直線は,

$$y = 37 + 0.6x$$

です.

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

もとの式

$$y = 37 + 0.6x$$

の両辺を 10 で割ると,

$$\frac{y}{10} = \frac{37}{10} + \frac{0.6}{10}x$$

となります.

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

ここで,

$$y' = \frac{y}{10}$$

なので,

$$y' = 3.7 + 0.06x$$

となります。これが、運動時間（分） x から得点 y' を予測する回帰直線です。

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

もとの回帰直線は、

$$y = 37 + 0.6x$$

変換後は、

$$y' = 3.7 + 0.06x$$

です。したがって、傾きと切片は、それぞれ

$$0.6 \longrightarrow 0.06 \qquad 37 \longrightarrow 3.7$$

と変化します。

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

目的変数の単位変換

目的変数 y の単位を比例的に変えると,

傾きと切片の両方が同じ倍率で変わります.

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

	傾き	切片
もとの式	0.6	37
x : 分 $\rightarrow$ 時間	36	37
y : 10 分の 1	0.06	3.7

例題 6.5 単位を変えると回帰直線はどう変わるかを確認する

単位変換が回帰直線に与える影響のまとめ

説明変数 x の単位を比例的に変えると,

傾きだけが変わる

目的変数 y の単位を比例的に変えると,

傾きと切片の両方が同じ倍率で変わる

第7章 質的データの表現

7.1 質的変数を表すグラフ

例題 7.1 質的変数を表すグラフの種類と用途を確認する

次の目的に適したグラフの種類は何かを考えましょう。なお、複数あってもかまいません。

- (1) 5 つの学部 of 学生数を比較したい。
- (2) ある商品の売上構成（A 店・B 店・C 店の割合）を示したい。
- (3) 2 つの地域の年代別人口構成を比較したい。
- (4) 中古車の特徴（走行距離・外装・内装・装備・価格）をそれぞれ 5 段階で評価して、多面的に比較したい。
- (5) 複数の学部について、男女別の学生数を比較したい。

例題 7.1 質的変数を表すグラフの種類と用途を確認する

データの特徴を表すグラフには、いくつかの種類があります。
大切なのは、

何を表したいのかに応じて、
適したグラフを選ぶことです。

例題 7.1 質的変数を表すグラフの種類と用途を確認する

(1) 5 つの学部在学生数を比較する

5 つの学部在学生数を，棒の高さを使って比較できるため，

棒グラフが適しています．

例題 7.1 質的変数を表すグラフの種類と用途を確認する

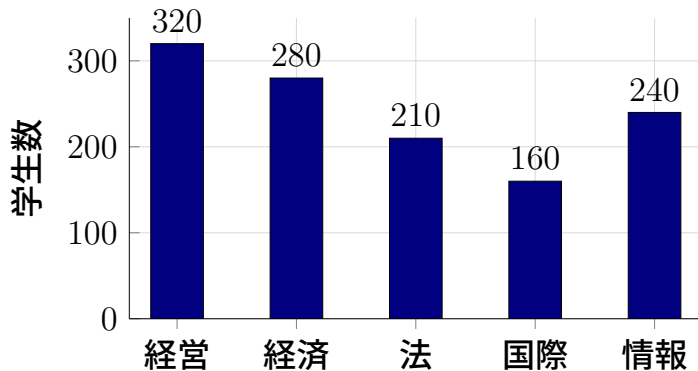


図 7.1 棒グラフ：学部ごとの学生数

例題 7.1 質的変数を表すグラフの種類と用途を確認する

棒グラフ

棒グラフは、棒の高さを使って、

項目ごとの数量を比較するためのグラフです。

例題 7.1 質的変数を表すグラフの種類と用途を確認する

(2) 売上構成を示す

A 店・B 店・C 店の各売上が、全体のなかでどれだけの割合を占めるかを表したいので、

円グラフが適しています。

帯グラフを使うこともできます。

例題 7.1 質的変数を表すグラフの種類と用途を確認する

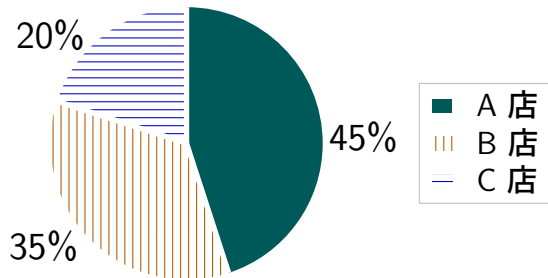


図 7.2 円グラフ：売上構成

例題 7.1 質的変数を表すグラフの種類と用途を確認する

円グラフ

円グラフは、1 つの集団について、

全体に対する割合の構成を表すグラフです。

例題 7.1 質的変数を表すグラフの種類と用途を確認する

(3) 2つの地域の年代別人口構成を比較する

2つの地域について、年代別人口の割合を表し、その構成のちがいを比較したいので、

帯グラフが適しています。

例題 7.1 質的変数を表すグラフの種類と用途を確認する

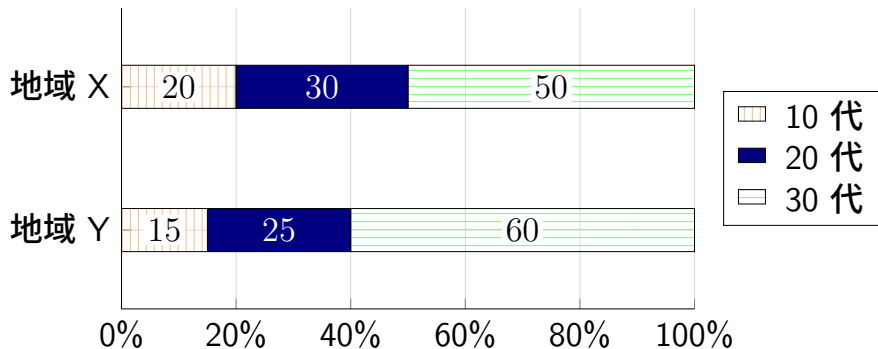


図 7.3 帯グラフ：地域別年代構成

例題 7.1 質的変数を表すグラフの種類と用途を確認する

帯グラフ

帯グラフは、帯の長さの比で割合の構成を表します。複数の集団について、

構成割合を比較する

こともできます。

例題 7.1 質的変数を表すグラフの種類と用途を確認する

帯グラフでは、各集団の全体を同じ長さにそろえます。そのため、
割合の構成のちがいを比較しやすくなります。

例題 7.1 質的変数を表すグラフの種類と用途を確認する

一方、年代ごとの人数の大小も表したい場合には、項目ごとの数量を積み上げて表す

積み上げ棒グラフを使うこともできます。

例題 7.1 質的変数を表すグラフの種類と用途を確認する

(4) 中古車の特徴を多面的に比較する

中古車の 5 つの特徴を 5 段階で評価し，放射状に配置すると，項目ごとの強みや弱みを多面的に比較できます．

レーダーチャートが適しています．

例題 7.1 質的変数を表すグラフの種類と用途を確認する

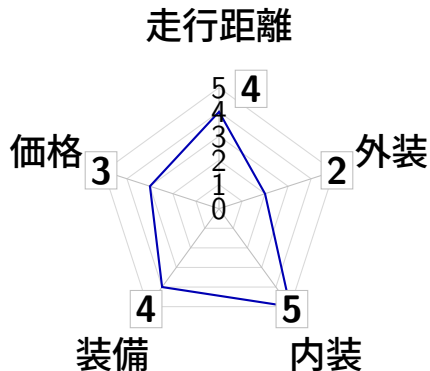


図 7.4 レーダーチャート：中古車の特徴

例題 7.1 質的変数を表すグラフの種類と用途を確認する

レーダーチャート

レーダーチャートは，放射状の軸上に値をとり，

複数の項目を同時に比較するためのグラフです．

例題 7.1 質的変数を表すグラフの種類と用途を確認する

(5) 学部ごとの男女別学生数を比較する

各学部の学生数を男女別に表すと，学部全体の学生数と，その内訳を同時に比較できます．

積み上げ棒グラフが適しています．

例題 7.1 質的変数を表すグラフの種類と用途を確認する

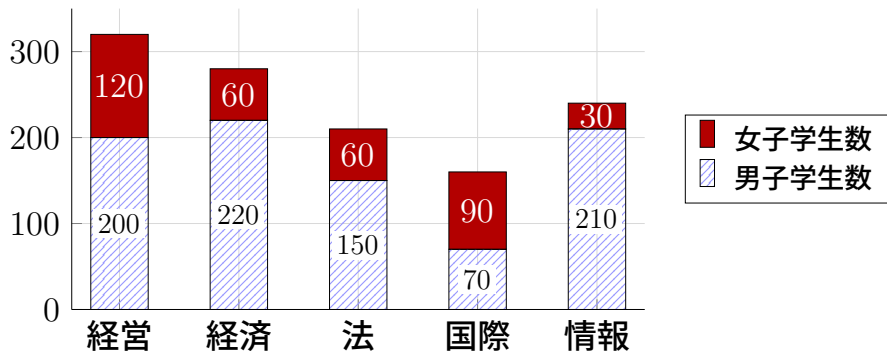


図 7.5 積み上げ棒グラフ：学部ごとの男女別学生数

例題 7.1 質的変数を表すグラフの種類と用途を確認する

積み上げ棒グラフ

積み上げ棒グラフは、棒を項目ごとに積み上げて、

全体の数量とその内訳を表すグラフです。

例題 7.1 質的変数を表すグラフの種類と用途を確認する

積み上げ棒グラフでは、

棒全体の高さ

が学部全体の学生数を表し、それぞれの部分が、

男女別の内訳

を表しています。

例題 7.1 質的変数を表すグラフの種類と用途を確認する

男女の人数をそれぞれ別の棒で表して、直接比較したい場合には、
棒グラフを使うこともできます。

例題 7.2 質的変数のグラフを読み取る

次の記述が，例題 7.1 の図 7.1～図 7.5 のグラフから読み取れる内容として適切かどうかを判断しましょう．適切である場合は○，適切でない場合は×とします．

- (1) 棒グラフ（図 7.1）より，法学部の学生数は，国際学部より多い．
- (2) 棒グラフ（図 7.1）より，経営学部と情報学部の学生数を合わせると，全体の半分以上になる．
- (3) 円グラフ（図 7.2）より，A 店の売上割合は，B 店と C 店の合計より小さい．

例題 7.2 質的変数のグラフを読み取る

- (4) 円グラフ（図 7.2）より，B 店と C 店の売上割合の合計は，全体の 6 割以上である．
- (5) 帯グラフ（図 7.3）より，地域 X では 30 代の割合は，10 代と 20 代の合計と同じである．
- (6) 帯グラフ（図 7.3）より，30 代の割合は地域 X のほうが地域 Y より高い．

例題 7.2 質的変数のグラフを読み取る

- (7) レーダーチャート（図 7.4）より，最も評価が高い項目は走行距離である．
- (8) レーダーチャート（図 7.4）より，外装より評価が低い項目は存在しない．
- (9) 積み上げ棒グラフ（図 7.5）より，情報学部では男子学生が 8 割以上を占める．
- (10) 積み上げ棒グラフ（図 7.5）より，女子学生数が最も多い学部は経営学部である．

例題 7.2 質的変数のグラフを読み取る

(1) 棒グラフを読み取る

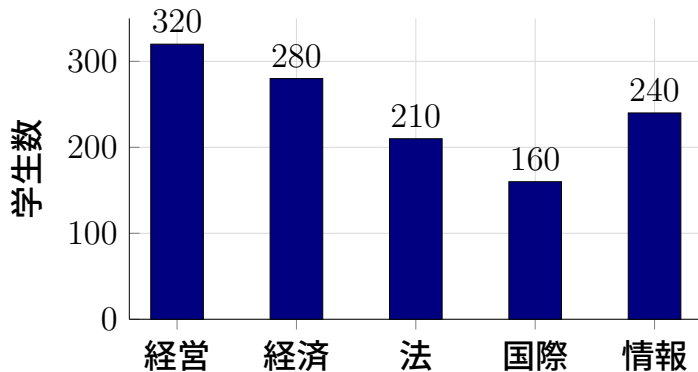


図 7.1 棒グラフ：学部ごとの学生数

例題 7.2 質的変数のグラフを読み取る

(1) 法学部と国際学部を比較する

棒グラフより、

法学部は 210 人、 国際学部は 160 人

です.

$$210 > 160$$

なので、



例題 7.2 質的変数のグラフを読み取る

(2) 経営学部と情報学部を合わせる

2 学部の学生数は、

$$320 + 240 = 560$$

です。5 学部全体では、

$$320 + 280 + 210 + 160 + 240 = 1210$$

なので、その半分は、

$$\frac{1210}{2} = 605$$

です。

例題 7.2 質的変数のグラフを読み取る

経営学部と情報学部の合計は 560 人，全体の半分は 605 人です．

$$560 < 605$$

なので，



例題 7.2 質的変数のグラフを読み取る

(3) 円グラフを読み取る

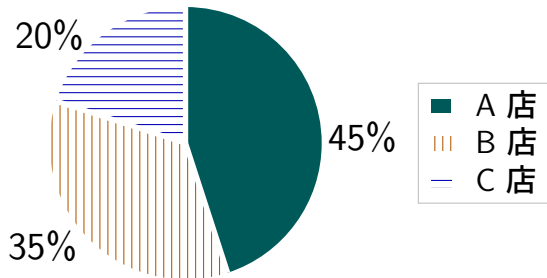


図 7.2 円グラフ：売上構成

例題 7.2 質的変数のグラフを読み取る

(3) A 店と B 店・C 店を比較する

A 店は、45% です。一方、B 店と C 店の合計は、

$$35\% + 20\% = 55\%$$

です。

$$45\% < 55\%$$

なので、



例題 7.2 質的変数のグラフを読み取る

(4) B 店と C 店の合計

B 店と C 店の売上割合の合計は、

$$35\% + 20\% = 55\%$$

です。6 割は 60% であり、

$$55\% < 60\%$$

なので、



例題 7.2 質的変数のグラフを読み取る

(5) 帯グラフを読み取る

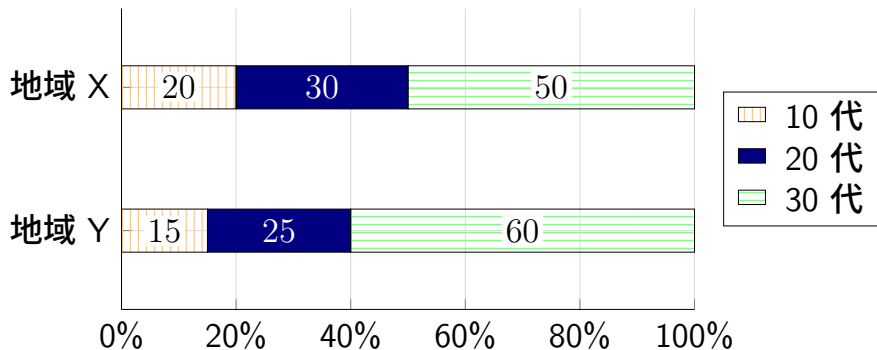


図 7.3 帯グラフ：地域別年代構成

例題 7.2 質的変数のグラフを読み取る

(5) 地域 X の年代構成

地域 X では、

10代は 20%, 20代は 30%

であり、

$$20\% + 30\% = 50\%$$

です。30代も 50% なので、



例題 7.2 質的変数のグラフを読み取る

(6) 30 代の割合を比較する

30 代の割合は,

地域 X は 50%, 地域 Y は 60%

です.

$$50\% < 60\%$$

なので,



例題 7.2 質的変数のグラフを読み取る

(7) レーダーチャートを読み取る

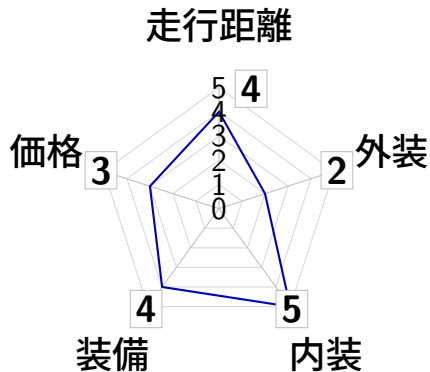


図 7.4 レーダーチャート：中古車の特徴

例題 7.2 質的変数のグラフを読み取る

(7) 最も評価が高い項目

走行距離の評価は、

4

です。一方、内装の評価は、

5

です。最も評価が高いのは内装なので、



例題 7.2 質的変数のグラフを読み取る

(8) 外装より低い評価があるか

5 つの評価は、

4, 2, 5, 4, 3

です．外装の評価は 2 であり，これより低い評価はありません．
したがって，



例題 7.2 質的変数のグラフを読み取る

(9) 積み上げ棒グラフを読み取る

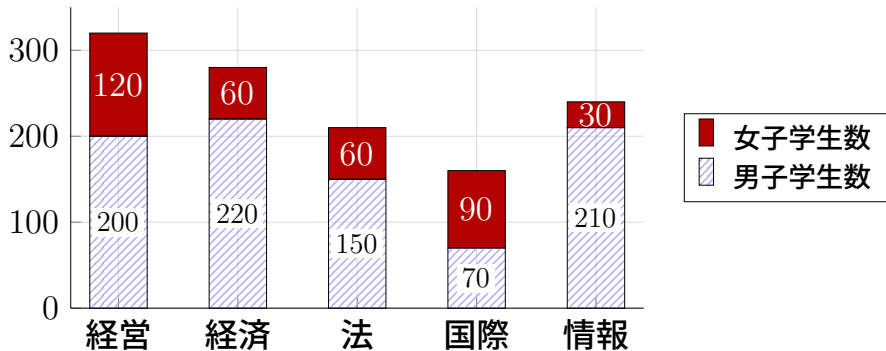


図 7.5 積み上げ棒グラフ：学部ごとの男女別学生数

例題 7.2 質的変数のグラフを読み取る

(9) 情報学部男子学生の割合

情報学部では、

男子は 210 人、女子は 30 人

なので、合計は、

$$210 + 30 = 240$$

人です。

例題 7.2 質的変数のグラフを読み取る

男子学生の割合は、

$$\frac{210}{240} = 0.875 = 87.5\%$$

です。87.5% は 8 割以上なので、



例題 7.2 質的変数のグラフを読み取る

(10) 女子学生数を比較する

女子学生数は、

経営	経済	法	国際	情報
120	60	60	90	30

です。最も多いのは経営学部の 120 人なので、



第7章 質的データの表現

7.2 質的変数の関係を表す表とグラフ

例題 7.3 クロス集計表から割合を計算する

次の表は、勉強法 A・勉強法 B を用いた人について、学習の効果ごとの人数をまとめたものです。

	効果あり	効果なし	合計
勉強法 A	30	70	100
勉強法 B	20	40	60
合計	50	110	160

例題 7.3 クロス集計表から割合を計算する

このクロス集計表を用いて，次の割合を求めましょう．

- (1) 勉強法 A を用いた人のうち，効果があった人の割合
- (2) 勉強法 B を用いた人のうち，効果があった人の割合
- (3) 効果があった人のうち，勉強法 A を用いた人の割合
- (4) 全体のうち，効果があった人の割合

例題 7.3 クロス集計表から割合を計算する

クロス集計表

クロス集計表とは、2 つの質的変数を組み合わせて、

各組み合わせに属する度数をまとめた表です。

例題 7.3 クロス集計表から割合を計算する

今回のクロス集計表では、1 つ目の質的変数は、

勉強法

で、「A」「B」に分類されています。

2 つ目の質的変数は、

学習の効果

で、「効果あり」「効果なし」に分類されています。

例題 7.3 クロス集計表から割合を計算する

度数

度数とは、ある分類や項目に当てはまる

データの個数

のことです。

たとえば、「勉強法 A を用いた人数」や「勉強法 A で効果があった人数」などが度数です。

例題 7.3 クロス集計表から割合を計算する

割合は、

$$\text{割合} = \frac{\text{該当するデータの個数（度数）}}{\text{基準となるデータの個数（度数）}}$$

で求めます。

大切なのは、

何を基準にするのかを確認することです。

例題 7.3 クロス集計表から割合を計算する

(1) 勉強法 A を用いた人のうち、効果があった人の割合

「勉強法 A を用いた人のうち」とあるので、基準となるのは、

勉強法 A を用いた 100 人です。

そのうち、効果があった人は 30 人です。

例題 7.3 クロス集計表から割合を計算する

したがって、求める割合は、

$$\frac{30}{100} = \frac{3}{10}$$

となります。

例題 7.3 クロス集計表から割合を計算する

(2) 勉強法 B を用いた人のうち、効果があった人の割合

「勉強法 B を用いた人のうち」とあるので、基準となるのは、

勉強法 B を用いた 60 人です。

そのうち、効果があった人は 20 人です。

例題 7.3 クロス集計表から割合を計算する

したがって、求める割合は、

$$\frac{20}{60} = \frac{1}{3}$$

となります。

例題 7.3 クロス集計表から割合を計算する

勉強法 A と勉強法 B について、「効果あり」の割合を比べると、

$$\text{勉強法 A : } \frac{30}{100} = \frac{3}{10}$$

$$\text{勉強法 B : } \frac{20}{60} = \frac{1}{3}$$

となります。

例題 7.3 クロス集計表から割合を計算する

効果があった人数だけを見ると，

$$30 > 20$$

なので，勉強法 A のほうが多く見えます．

しかし，勉強法 A と B では，

もとの人数が異なります．

そのため，効果を比較するときには，それぞれの勉強法を用いた人のなかでの割合を見ることが大切です．

例題 7.3 クロス集計表から割合を計算する

(3) 効果があった人のうち、勉強法 A を用いた人の割合

「効果があった人のうち」とあるので、基準となるのは、

効果があった 50 人です。

そのうち、勉強法 A を用いた人は 30 人です。

例題 7.3 クロス集計表から割合を計算する

したがって、求める割合は、

$$\frac{30}{50} = \frac{3}{5}$$

となります。

例題 7.3 クロス集計表から割合を計算する

(4) 全体のうち，効果があった人の割合

「全体のうち」とあるので，基準となるのは，

全体の 160 人です．

そのうち，効果があった人は 50 人です．

例題 7.3 クロス集計表から割合を計算する

したがって、求める割合は、

$$\frac{50}{160} = \frac{5}{16}$$

となります。

例題 7.3 クロス集計表から割合を計算する

同じクロス集計表でも，問いによって基準となる人数が変わります．

問い	基準となる人数
A を用いた人のうち	100
B を用いた人のうち	60
効果があった人のうち	50
全体のうち	160

「～のうち」に注目すると，分母を判断しやすくなります．

例題 7.4 クロス集計表を読み取る

ある大学の 1 年生 180 人について、「所属学部」と「昼食のとり方」を調べました。

	弁当	購買	学食	外食	合計
経営学部	20	25	20	15	80
法学部	18	20	12	10	60
情報学部	15	15	7	3	40
合計	53	60	39	28	180

例題 7.4 クロス集計表を読み取る

- (1) 昼食を弁当にしている学生の割合を求めましょう。
- (2) 法学部の学生について、弁当にしている学生の人数と学食にしている学生の人数の比を求めましょう。
- (3) 経営学部の学生のうち、昼食を弁当または学食にしている学生の割合を求めましょう。
- (4) 法学部または情報学部に所属している学生のうち、昼食を購入にしている学生の割合を求めましょう。

例題 7.4 クロス集計表を読み取る

- (5) 昼食を弁当または購買にしている学生の割合を求めましょう。
- (6) 昼食を外食にしている学生のうち、法学部または情報学部に所属している学生の割合を求めましょう。
- (7) 弁当または学食にしている学生について、情報学部 to 所属している学生の人数と法学部に所属している学生の人数の比を求めましょう。

例題 7.4 クロス集計表を読み取る

クロス集計表を読み取るときは、

どの値が「該当する度数」か

と、

どの値が「基準となる度数」か

を正確に確認することが大切です。

例題 7.4 クロス集計表を読み取る

(1) 弁当にしている学生の割合

弁当にしている学生は，

53 人

です．全体は，

180 人

なので，求める割合は，

$$\frac{53}{180}$$

となります．

例題 7.4 クロス集計表を読み取る

(2) 法学部の弁当と学食の人数の比

法学部では、

弁当は 18 人, 学食は 12 人

です。したがって、求める比は、

$$18 : 12 = 3 : 2$$

です。

例題 7.4 クロス集計表を読み取る

(3) 経営学部で弁当または学食の割合を求める

経営学部の学生は 80 人です。そのうち、

弁当は 20 人, 学食は 20 人

であり、これらの合計は、

$$20 + 20 = 40$$

人です。

例題 7.4 クロス集計表を読み取る

したがって、求める割合は、

$$\frac{40}{80} = \frac{1}{2}$$

です。

例題 7.4 クロス集計表を読み取る

(4) 法学部または情報学部で購買の割合を求める

法学部と情報学部の学生数は、

$$60 + 40 = 100$$

人です。そのうち購買にしている学生は、

$$20 + 15 = 35$$

人です。

例題 7.4 クロス集計表を読み取る

したがって、求める割合は、

$$\frac{35}{100} = \frac{7}{20}$$

です。

例題 7.4 クロス集計表を読み取る

(5) 弁当または購買の割合を求める

弁当にしている学生は 53 人，購買にしている学生は 60 人です．

したがって，弁当または購買にしている学生は

$$53 + 60 = 113$$

人です．

例題 7.4 クロス集計表を読み取る

全体は 180 人なので，求める割合は，

$$\frac{113}{180}$$

となります．

例題 7.4 クロス集計表を読み取る

(6) 外食にしている学生のうち、法学部または情報学部割合を求める

外食にしている学生は、

28 人

です。そのうち、法学部または情報学部の学生は、

$$10 + 3 = 13$$

人です。

例題 7.4 クロス集計表を読み取る

したがって、求める割合は、

$$\frac{13}{28}$$

です。

例題 7.4 クロス集計表を読み取る

(7) 弁当または学食の学生のうち、情報学部と法学部の人数の比を求める

弁当または学食にしている学生のうち、情報学部に所属している学生は、

$$15 + 7 = 22$$

人です。法学部に所属している学生は、

$$18 + 12 = 30$$

人です。

例題 7.4 クロス集計表を読み取る

したがって、これらの人数の比は、

$$22 : 30$$

です．約分すると、

$$11 : 15$$

となります．

例題 7.5 モザイク図を読み取る

ある大学の学科 A・学科 B の学生について、ふだん利用している通学手段を調べました。

通学手段は、

電車 ・ 自動車 ・ その他

の 3 種類です。

学科 A の学生は全体の 40%，学科 B の学生は全体の 60% です。

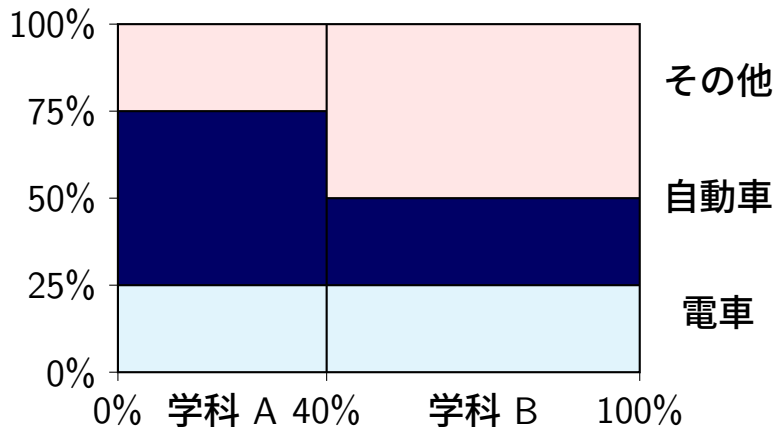
例題 7.5 モザイク図を読み取る

調査結果は、次のとおりです。

	電車	自動車	その他	合計
学科 A	20	40	20	80
学科 B	30	30	60	120
合計	50	70	80	200

この情報をモザイク図で表します。

例題 7.5 モザイク図を読み取る



例題 7.5 モザイク図を読み取る

次の記述が，モザイク図から読み取れる内容として適切かどうかを判断しましょう．

- (1) 学科 A では，自動車の割合が最も大きい．
- (2) 学科 B では，その他の割合が最も大きい．
- (3) 学科 A と学科 B では，最も割合が大きい通学手段が同じである．

例題 7.5 モザイク図を読み取る

- (4) 電車を利用している人数は，学科 A より学科 B のほうが多い．
- (5) その他の人数は，学科 A と学科 B で同じである．
- (6) 全体として，自動車を利用している人数は，電車を利用している人数より多い．

例題 7.5 モザイク図を読み取る

モザイク図

モザイク図は，2 つの質的変数の各組み合わせに属するデータの割合を，

長方形の面積で表すグラフです．

例題 7.5 モザイク図を読み取る

モザイク図では、

幅： 一方の変数の各分類が全体に占める割合

高さ： その分類のなかで、他方の変数の各分類が占める割合

であり、長方形の面積は、

$$\text{幅} \times \text{高さ}$$

です。

例題 7.5 モザイク図を読み取る

したがって、

面積

は、

その組み合わせが全体に占める割合

を表します。

例題 7.5 モザイク図を読み取る

この例題では、

クロス集計表の数値を使わず

モザイク図の幅・高さ・面積だけから判断します。

例題 7.5 モザイク図を読み取る

モザイク図の読み取り

モザイク図では、幅・高さ・面積に注目して読み取ります。

幅の比較： 学科ごとの学生数の割合

高さの比較： 各学科のなかでの通学手段の割合

面積の比較： 各組み合わせに属する学生数の比較

例題 7.5 モザイク図を読み取る

(1) 学科 A で最も割合が大きい通学手段

学科 A では、

その他は 25%

自動車は 50%

電車は 25%

です。

例題 7.5 モザイク図を読み取る

ここで、

$$50\% > 25\%$$

なので、学科 A では、自動車の割合が最も大きくなっています。

したがって、



例題 7.5 モザイク図を読み取る

(2) 学科 B で最も割合が大きい通学手段

学科 B では、

その他は 50%

自動車は 25%

電車は 25%

です。

例題 7.5 モザイク図を読み取る

学科 B では、その他の割合が 50% で最も大きくなっています。
したがって、



例題 7.5 モザイク図を読み取る

(3) 最も割合が大きい通学手段を比較する

学科 A では、

自動車が最も大きく、

学科 B では、

その他が最も大きくなっています。

例題 7.5 モザイク図を読み取る

学科 A と学科 B では，最も割合が大きい通学手段が異なります。
したがって，



例題 7.5 モザイク図を読み取る

(4) 電車を利用している人数を比較する

電車の高さは，学科 A も学科 B も

25%

と同じです．しかし，学科の幅は，

学科 A は 40%， 学科 B は 60%

です．

例題 7.5 モザイク図を読み取る

電車の部分の面積は，

$$\text{学科 A : } 0.40 \times 0.25 = 0.10$$

$$\text{学科 B : } 0.60 \times 0.25 = 0.15$$

です．学科 B のほうが面積が大きいので，電車を利用している人数も学科 B のほうが多いことがわかります．したがって，



人数を比べるときは面積に注目します．

例題 7.5 モザイク図を読み取る

(5) その他の人数を比較する

「その他」の高さは,

学科 A は 25%, 学科 B は 50%

です. さらに, 幅は,

学科 A は 40%, 学科 B は 60%

です.

例題 7.5 モザイク図を読み取る

「その他」の面積は，

$$\text{学科 A : } 0.40 \times 0.25 = 0.10$$

$$\text{学科 B : } 0.60 \times 0.50 = 0.30$$

です．学科 B のほうが面積が大きいので，「その他」の人数は，学科 B のほうが多いことがわかります．したがって，



例題 7.5 モザイク図を読み取る

(6) 自動車と電車を全体で比較する

自動車の面積は，

$$\text{学科 A : } 0.40 \times 0.50 = 0.20$$

$$\text{学科 B : } 0.60 \times 0.25 = 0.15$$

です．したがって，自動車全体の割合は，

$$0.20 + 0.15 = 0.35$$

です．

例題 7.5 モザイク図を読み取る

一方、電車の面積は、

$$\text{学科 A : } 0.40 \times 0.25 = 0.10$$

$$\text{学科 B : } 0.60 \times 0.25 = 0.15$$

です。したがって、電車全体の割合は、

$$0.10 + 0.15 = 0.25$$

です。

例題 7.5 モザイク図を読み取る

ここで、

$$0.35 > 0.25$$

なので、自動車を利用している人数のほうが多いことがわかります。

したがって、



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

例題 8.1 度数分布表を作る

あるクラスの生徒 40 人について、1 週間の平均読書時間（分）を調べました．

12, 25, 30, 18, 42, 55, 60, 33, 28, 20,
15, 38, 45, 50, 22, 35, 40, 48, 52, 27,
10, 18, 24, 32, 36, 41, 47, 53, 58, 29,
14, 19, 26, 31, 37, 43, 49, 56, 62, 34

例題 8.1 度数分布表を作る

このデータについて、次のように階級幅 10 の度数分布表を作りましょう。

- (1) 「10 以上 20 未満」「20 以上 30 未満」などのように、左端を「以上」、右端を「未満」として階級を設定しましょう。
- (2) 各階級に入るデータの個数を数え、度数分布表にまとめましょう。

例題 8.1 度数分布表を作る

度数分布表

度数分布表とは，データをいくつかの区間に分け，

それぞれの区間に含まれる
データの個数をまとめた表です．

例題 8.1 度数分布表を作る

階級と度数

階級とは、データを区間ごとに区切ったときのそれぞれの区間のことです。

度数とは、それぞれの階級に入るデータの個数のことです。

例題 8.1 度数分布表を作る

度数を数える前に、データを小さい順に並べ替えると、どの階級に入るかを確認しやすくなります。

10, 12, 14, 15, 18, 18, 19, 20, 22, 24,
25, 26, 27, 28, 29, 30, 31, 32, 33, 34,
35, 36, 37, 38, 40, 41, 42, 43, 45, 47,
48, 49, 50, 52, 53, 55, 56, 58, 60, 62

例題 8.1 度数分布表を作る

並べ替えたデータから，

$$\text{最小値} = 10$$

$$\text{最大値} = 62$$

であることがわかります．

今回は階級幅が 10 なので，10 から 70 までを 10 ずつ区切って考えます．

例題 8.1 度数分布表を作る

(1) 階級を設定する

たとえば，

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

60 以上 70 未満

と区切ることができます．

例題 8.1 度数分布表を作る

これで、最小値 10 から最大値 62 までをすべて含めることができます．

例題 8.1 度数分布表を作る

たとえば，

10 以上 20 未満

という階級には，

10 は入りますが，20 は入りません．

20 は，次の

20 以上 30 未満

の階級に入ります．

例題 8.1 度数分布表を作る

(2) 各階級の度数を数える

「10 以上 20 未満」に入るデータは、

10, 12, 14, 15, 18, 18, 19

です。したがって、

度数は 7 です。

例題 8.1 度数分布表を作る

同じように数えると，すべての階級の度数を求めることができます．
度数分布表は，次になります．

例題 8.1 度数分布表を作る

階級	度数
10 以上 20 未満	7
20 以上 30 未満	8
30 以上 40 未満	9
40 以上 50 未満	8
50 以上 60 未満	6
60 以上 70 未満	2
合計	40

例題 8.1 度数分布表を作る

度数の合計を確認

度数の合計は，

$$7 + 8 + 9 + 8 + 6 + 2 = 40$$

となります．

もとのデータ数も 40 なので，

度数の合計が，もとのデータ数と
一致していることを確認できました．

例題 8.2 度数分布表から平均値と中央値の近似値を求める

例題 8.1 で作成した度数分布表を使います。

階級	度数
10 以上 20 未満	7
20 以上 30 未満	8
30 以上 40 未満	9
40 以上 50 未満	8
50 以上 60 未満	6
60 以上 70 未満	2
合計	40

例題 8.2 度数分布表から平均値と中央値の近似値を求める

- (1) 度数が最も大きい階級（最頻階級）はどこか答えましょう。
- (2) 平均読書時間が 30 分以上 50 分未満の生徒は何人いるでしょうか。
- (3) 各階級の相対度数，累積度数，累積相対度数を求め，表に追記しましょう。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

- (4) 階級値を使って、平均値の近似値を求めましょう.
- (5) 度数分布表から、中央値のおおよその値を求めましょう.
- (6) (4), (5) の近似値を, もとのデータから求めた

$$\text{平均値} = 35.35, \quad \text{中央値} = 34.5$$

と比べて, 差の大きさを確かめましょう.

例題 8.2 度数分布表から平均値と中央値の近似値を求める

(1) 最頻階級を求める

階級	度数
10 以上 20 未満	7
20 以上 30 未満	8
30 以上 40 未満	9
40 以上 50 未満	8
50 以上 60 未満	6
60 以上 70 未満	2

例題 8.2 度数分布表から平均値と中央値の近似値を求める

最頻階級

度数が最も大きい階級を、**最頻階級**といいます。

今回、度数の最大値は 9 なので、

30 以上 40 未満が最頻階級です。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

(2) 30 分以上 50 分未満の人数を求める

「30 以上 40 未満」の度数は 9, 「40 以上 50 未満」の度数は 8 です。
したがって, 30 分以上 50 分未満の生徒は

$$9 + 8 = 17$$

人です.

例題 8.2 度数分布表から平均値と中央値の近似値を求める

相対度数

相対度数は、各階級の度数を全体のデータ数で割った値です。

相対度数は、

$$\text{相対度数} = \frac{\text{各階級の度数}}{\text{全体のデータ数}}$$

で求めます。

今回はデータが 40 個なので、分母は 40 です。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

累積度数

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です。

たとえば、

$$7$$

$$7 + 8 = 15$$

$$7 + 8 + 9 = 24$$

のように求めます。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

累積相対度数

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。

最後の階級では、

1

になります。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

(3) 相対度数・累積度数・累積相対度数を求める

各階級の相対度数は、

$$\frac{7}{40}, \quad \frac{8}{40}, \quad \frac{9}{40}, \quad \frac{8}{40}, \quad \frac{6}{40}, \quad \frac{2}{40}$$

です。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

累積度数は

$$7$$

$$7 + 8 = 15$$

$$15 + 9 = 24$$

$$24 + 8 = 32$$

$$32 + 6 = 38$$

$$38 + 2 = 40$$

となります。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

累積相対度数は、

$$\frac{7}{40}, \quad \frac{7}{40} + \frac{8}{40} = \frac{15}{40}, \quad \frac{15}{40} + \frac{9}{40} = \frac{24}{40},$$

$$\frac{24}{40} + \frac{8}{40} = \frac{32}{40}, \quad \frac{32}{40} + \frac{6}{40} = \frac{38}{40}, \quad \frac{38}{40} + \frac{2}{40} = \frac{40}{40}$$

となります。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

階級	度数	相対度数	累積度数	累積相対度数
10 以上 20 未満	7	$7/40$	7	$7/40$
20 以上 30 未満	8	$8/40$	15	$15/40$
30 以上 40 未満	9	$9/40$	24	$24/40$
40 以上 50 未満	8	$8/40$	32	$32/40$
50 以上 60 未満	6	$6/40$	38	$38/40$
60 以上 70 未満	2	$2/40$	40	$40/40$

例題 8.2 度数分布表から平均値と中央値の近似値を求める

階級値

階級値とは，その階級を代表する値です．

階級の下端と上端の平均をとって，

$$\text{階級値} = \frac{\text{下端} + \text{上端}}{2}$$

で求めます．

例題 8.2 度数分布表から平均値と中央値の近似値を求める

たとえば、「10 以上 20 未満」の階級値は、

$$\frac{10 + 20}{2} = 15$$

「20 以上 30 未満」の階級値は、

$$\frac{20 + 30}{2} = 25$$

です。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

すべての階級の階級値は，次になります．

階級	階級値
10 以上 20 未満	15
20 以上 30 未満	25
30 以上 40 未満	35
40 以上 50 未満	45
50 以上 60 未満	55
60 以上 70 未満	65

例題 8.2 度数分布表から平均値と中央値の近似値を求める

(4) 平均値の近似値を求める

度数分布表では、各階級に入っているデータをすべて階級値で代表させます。

そこで、

$$\text{階級値} \times \text{度数}$$

を足し合わせることで、全体の合計を近似します。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

各階級の「階級値 × 度数」を足し合わせると、

$$15 \times 7 + 25 \times 8 + 35 \times 9 + 45 \times 8 + 55 \times 6 + 65 \times 2 = 1440$$

となります。これは、全体の合計の近似値です。

データの個数は 40 なので、平均値の近似値は、

$$\frac{1440}{40} = 36$$

分です。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

(5) 中央値の近似値を求める

データは全部で 40 個なので、中央値は、

20 番目と 21 番目の位置にあたります。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

累積度数を見ると，

階級	累積度数
10 以上 20 未満	7
20 以上 30 未満	15
30 以上 40 未満	24

です．20 番目と 21 番目は，

30 以上 40 未満の階級に入ります．

例題 8.2 度数分布表から平均値と中央値の近似値を求める

「30 以上 40 未満」の階級値は、

$$\frac{30 + 40}{2} = 35$$

です。

この階級に入るデータを 35 分で代表させるので、中央値の近似値は 35 分です。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

(6) もとのデータから求めた値と比較する

もとのデータでは、

$$\text{平均値} = 35.35, \quad \text{中央値} = 34.5$$

でした。度数分布表から求めた近似値は、

$$\text{平均値} = 36, \quad \text{中央値} = 35$$

です。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

平均値の差の大きさは、

$$|35.35 - 36| = 0.65$$

なので、平均値の差は 0.65 分です。

中央値の差の大きさは、

$$|34.5 - 35| = 0.5$$

なので、中央値の差は 0.5 分です。

例題 8.2 度数分布表から平均値と中央値の近似値を求める

度数分布表では、各階級のデータをすべて

1 つの階級値で代表させています。

そのため、

もとのデータから求めた値と
少しずれることがあります。

例題 8.3 ヒストグラムを作る

あるクラスの生徒 50 人について、1 日の睡眠時間（時間）を調べました．この度数分布表にもとづいて、ヒストグラムを作成しましょう．

階級	度数
4 以上 5 未満	4
5 以上 6 未満	10
6 以上 7 未満	16
7 以上 8 未満	12
8 以上 9 未満	8
合計	50

例題 8.3 ヒストグラムを作る

ヒストグラム

ヒストグラムは、量的データをいくつかの区間（階級）に分け、

各階級の度数を長方形の面積で表したグラフです。

長方形の幅は階級の幅に対応し、長方形の面積がその階級の度数に比例するように表します。

例題 8.3 ヒストグラムを作る

今回は，すべての階級の幅が1時間 で同じです．そのため，

長方形の高さで度数を表せます．

横軸には階級，縦軸には度数をとり，長方形の高さを

4, 10, 16, 12, 8

とします．

例題 8.3 ヒストグラムを作る

ヒストグラムでは，階級を

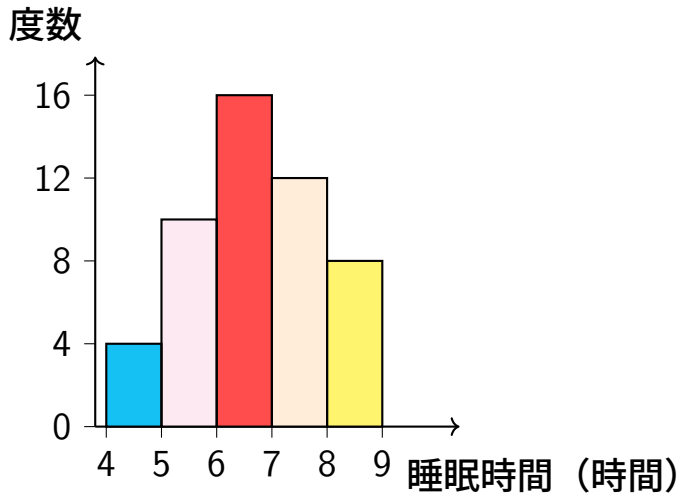
4 以上 5 未満 $\rightarrow$ 5 以上 6 未満 $\rightarrow \dots$

のように，連続して設定します．

そのため，

長方形どうしを隙間なく並べます．

例題 8.3 ヒストグラムを作る



例題 8.3 ヒストグラムを作る

ヒストグラムを見るときは、

- 分布の形や偏り
- どの階級が最も多いか
- 山の位置やデータの広がり

に注目します。

例題 8.3 ヒストグラムを作る

今回のヒストグラムは、

全体として 1 つの山をもつ形

で、大きく偏った形ではありません。

また、最も度数が大きいのは 6 以上 7 未満 の階級なので、

6 以上 7 未満が最頻階級です。

第8章 データの分布の形

8.2 四分位数と箱ひげ図

例題 8.4 四分位数と四分位範囲を求める

次の 30 か月分の販売数データを使います．

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

データはすでに小さい順に並べてあります．

例題 8.4 四分位数と四分位範囲を求める

このデータについて、

第 1 四分位数， 第 2 四分位数（中央値）， 第 3 四分位数
を求めましょう。さらに、

四分位範囲

を求めましょう。

例題 8.4 四分位数と四分位範囲を求める

四分位数

四分位数とは，データを小さい順に並べたとき，

データの位置をおよそ 4 等分する
区切りとなる値です．

例題 8.4 四分位数と四分位範囲を求める

四分位数には、

第 1 四分位数： 下からおよそ 25%の位置

第 2 四分位数： 下から 50%の位置 (中央値)

第 3 四分位数： 下からおよそ 75%の位置

があります。

例題 8.4 四分位数と四分位範囲を求める

四分位範囲

第 3 四分位数から第 1 四分位数を引いた値を，四分位範囲といいます．

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

四分位範囲は，

真ん中のおよそ 50% のデータの広がりを表します．

例題 8.4 四分位数と四分位範囲を求める

本書では、まず、

全体の中央値

を求め、これを第 2 四分位数とします。

そのあと、中央値の位置で、

下半分 と 上半分

に分けます。

例題 8.4 四分位数と四分位範囲を求める

そして、下半分の中央値を

第 1 四分位数

とし、上半分の中央値を

第 3 四分位数

とします。

例題 8.4 四分位数と四分位範囲を求める

データを上下に分ける方法

- データ数が偶数の場合
中央の 2 つの値の位置で、下半分と上半分に分けます。
- データ数が奇数の場合
中央値の位置を境界として分け、**中央値そのものは上下どちらにも含めません。**

例題 8.4 四分位数と四分位範囲を求める

第 2 四分位数を求める

データは 30 個なので、中央にあるのは、

15 番目と 16 番目の 2 つの値です。

例題 8.4 四分位数と四分位範囲を求める

15 番目の値は，

35

16 番目の値も，

35

です．したがって，第 2 四分位数は，

$$\frac{35 + 35}{2} = 35$$

です．

例題 8.4 四分位数と四分位範囲を求める

第 1 四分位数を求める

中央値の位置でデータを分けます．

下半分の 15 個は，

3, 6, 11, 13, 14, 20, 21, 23,
24, 26, 27, 31, 33, 35, 35

です．

この 15 個の中央値を求めます．

例題 8.4 四分位数と四分位範囲を求める

15 個のデータの中央は、

8 番目です。

8 番目の値は、

23

なので、

第 1 四分位数は 23 です。

例題 8.4 四分位数と四分位範囲を求める

第 3 四分位数を求める

上半分の 15 個は,

35, 38, 40, 41, 43, 44, 46, 50,
50, 50, 51, 51, 52, 53, 54

です.

この 15 個の中央値を求めます.

例題 8.4 四分位数と四分位範囲を求める

15 個のデータの中央は、

8 番目です。

8 番目の値は、

50

なので、

第 3 四分位数は 50 です。

例題 8.4 四分位数と四分位範囲を求める

求めた四分位数は、

第 1 四分位数： 23

第 2 四分位数： 35

第 3 四分位数： 50

です。

例題 8.4 四分位数と四分位範囲を求める

四分位範囲は、

第 3 四分位数 − 第 1 四分位数

なので、

$$50 - 23 = 27$$

です。

したがって、四分位範囲は 27 で、真ん中のおよそ 50% のデータの広がりを表しています。

例題 8.5 箱ひげ図を作る

例題 8.4 と同じ 30 か月分の販売数データを使います。

例題 8.4 で求めた値は、

$$\text{最小値} = 3$$

$$\text{第 1 四分位数} = 23, \quad \text{中央値} = 35$$

$$\text{第 3 四分位数} = 50, \quad \text{最大値} = 54$$

です。

これらの値をもとに、箱ひげ図を作成しましょう。

例題 8.5 箱ひげ図を作る

箱ひげ図

箱ひげ図は，データのばらつきや中心の位置をひと目で確認できるように表した図です．

次の 5 つの値を使って描きます．

最小値

第 2 四分位数（中央値）

最大値

第 1 四分位数

第 3 四分位数

例題 8.5 箱ひげ図を作る

箱ひげ図では、5 つの値を次の位置に表します。



例題 8.5 箱ひげ図を作る

箱は、

第 1 四分位数

から

第 3 四分位数

までの範囲を表します。つまり、

真ん中のおよそ 50% のデータが
含まれる範囲です。

例題 8.5 箱ひげ図を作る

箱の左端は第 1 四分位数，右端は第 3 四分位数なので，箱の長さは，

$$\text{第 3 四分位数} - \text{第 1 四分位数}$$

となります．したがって，箱の長さは，

四分位範囲を表しています．

今回は，

$$50 - 23 = 27$$

です．

例題 8.5 箱ひげ図を作る

今回使う 5 つの値は、

最小値	第 1 四分位数	中央値	第 3 四分位数	最大値
3	23	35	50	54

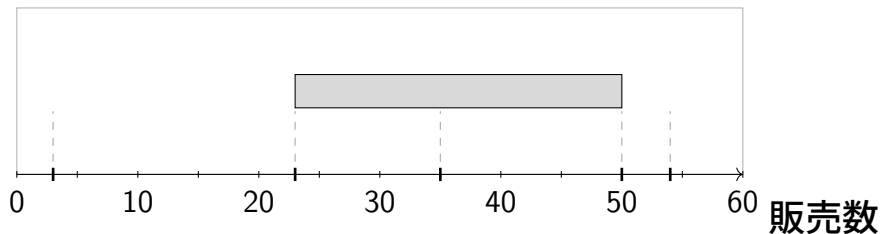
です。

例題 8.5 箱ひげ図を作る

まず、横軸上に 3, 23, 35, 50, 54 の位置をとります。

第 1 四分位数 23 と第 3 四分位数 50 を両端として、

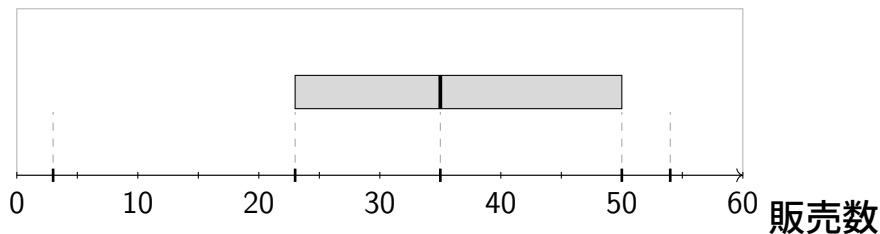
箱を描きます。



例題 8.5 箱ひげ図を作る

次に、箱の中の中央値 35 の位置に、

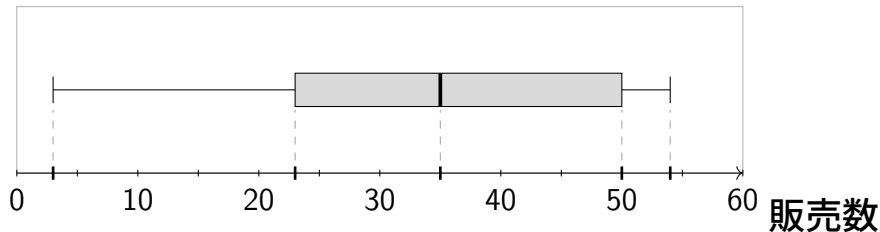
縦の線を引きます。



例題 8.5 箱ひげ図を作る

最後に，箱の左端 23 から最小値 3 まで，箱の右端 50 から最大値 54 まで，

ひげを伸ばします．



例題 8.5 箱ひげ図を作る

箱ひげ図では、箱の長さを見ることで、

真ん中のおよそ 50% のデータの広がりを
確認できます。

また、最小値から最大値までを見ることで、

データ全体の広がりも確認できます。

例題 8.6 幹葉表示を作る

あるクラス 50 人の数学の点数データがあります。

72, 45, 63, 34, 61, 75, 47, 58, 32, 84,
63, 55, 36, 73, 52, 47, 75, 66, 34, 88,
63, 55, 36, 42, 70, 79, 52, 67, 75, 61,
45, 58, 36, 73, 50, 64, 81, 57, 70, 84,
35, 45, 53, 60, 55, 68, 61, 78, 38, 90

例題 8.6 幹葉表示を作る

このデータについて、次の問いに答えましょう。

(1) 幹を「10 の位」、葉を「1 の位」として、幹葉表示を作りましょう。

(2) 作成した幹葉表示をもとに、

第 1 四分位数，第 2 四分位数（中央値），第 3 四分位数
を求めましょう。

例題 8.6 幹葉表示を作る

幹葉表示

幹葉表示（幹葉図）は，データの値そのものを使って分布の形を表す図です．

各データを，

幹と葉に分けて並べます．

例題 8.6 幹葉表示を作る

今回の点数では,

幹は, 10 の位を表します.

葉は, 1 の位を表します.

たとえば,

47

なら,

幹 = 4, 葉 = 7

です.

例題 8.6 幹葉表示を作る

幹葉表示では，幹は，

上から下へ小さい順

葉は，

左から右へ小さい順

に並べます。

例題 8.6 幹葉表示を作る

まず，データを小さい順に並べます．

32, 34, 34, 35, 36, 36, 36, 38, 42, 45,
45, 45, 47, 47, 50, 52, 52, 53, 55, 55,
55, 57, 58, 58, 60, 61, 61, 61, 63, 63,
63, 64, 66, 67, 68, 70, 70, 72, 73, 73,
75, 75, 75, 78, 79, 81, 84, 84, 88, 90

例題 8.6 幹葉表示を作る

データは、30 点台から 90 点台まであります。

したがって、幹は、

3, 4, 5, 6, 7, 8, 9

となります。

例題 8.6 幹葉表示を作る

(1) 幹葉表示を作る

まず、30 点台のデータは、

32, 34, 34, 35, 36, 36, 36, 38

です。したがって、

3 | 2 4 4 5 6 6 6 8

となります。

例題 8.6 幹葉表示を作る

同じように，40 点台から 90 点台までを整理すると，

4		2 5 5 5 7 7
5		0 2 2 3 5 5 5 7 8 8
6		0 1 1 1 3 3 3 4 6 7 8
7		0 0 2 3 3 5 5 5 8 9
8		1 4 4 8
9		0

となります．

例題 8.6 幹葉表示を作る

以上をまとめると，

幹	葉
3	2 4 4 5 6 6 6 8
4	2 5 5 5 7 7
5	0 2 2 3 5 5 5 7 8 8
6	0 1 1 1 3 3 3 4 6 7 8
7	0 0 2 3 3 5 5 5 8 9
8	1 4 4 8
9	0

例題 8.6 幹葉表示を作る

ここで、たとえば、

$$4 \mid 2 \ 5 \ 5 \ 5 \ 7 \ 7$$

は、

$$42, 45, 45, 45, 47, 47$$

を表しています。

幹葉表示では、

元のデータの値を残したまま
分布を見ることができます。

例題 8.6 幹葉表示を作る

(2) 四分位数を求める

データは全部で 50 個です．第 2 四分位数（中央値）は，

25 番目と 26 番目の平均です．

例題 8.6 幹葉表示を作る

下半分は 25 個なので，その中央は 13 番目です。
したがって，第 1 四分位数は，

全体の 13 番目です。

同じように，第 3 四分位数は，

全体の 38 番目です。

例題 8.6 幹葉表示を作る

幹ごとの個数と累計を確認します.

幹	個数	累計
3	8	8
4	6	14
5	10	24
6	11	35
7	10	45
8	4	49
9	1	50

例題 8.6 幹葉表示を作る

第 1 四分位数

第 1 四分位数は、全体の 13 番目の値です。

30 点台までに 8 個あるので、全体の 13 番目は、40 点台の

$$13 - 8 = 5 \text{ 番目}$$

の値です。

例題 8.6 幹葉表示を作る

40 点台の行は,

4 | 2 5 5 5 7 7

です.

左から 5 番目の葉は 7 なので, 全体の 13 番目の値は

47

です. つまり,

第 1 四分位数は 47

例題 8.6 幹葉表示を作る

第 2 四分位数 (中央値)

第 2 四分位数は、全体の 25 番目と 26 番目の平均です.

50 点台までに 24 個あるので、全体の 25 番目と 26 番目は、60 点台の

$$25 - 24 = 1 \text{ 番目}, \quad 26 - 24 = 2 \text{ 番目}$$

の値です.

例題 8.6 幹葉表示を作る

60 点台の行は、

6 | 0 1 1 1 3 3 3 4 6 7 8

です。

1 番目の値は 60, 2 番目の値は 61 なので、

$$\frac{60 + 61}{2} = 60.5$$

より、

第 2 四分位数は 60.5

例題 8.6 幹葉表示を作る

第 3 四分位数

第 3 四分位数は、全体の 38 番目の値です。

60 点台までに 35 個あるので、全体の 38 番目は、70 点台の

$$38 - 35 = 3 \text{ 番目}$$

の値です。

例題 8.6 幹葉表示を作る

70 点台の行は、

7 | 0 0 **2** 3 3 5 5 5 8 9

です。

左から 3 番目の葉は 2 なので、全体の 38 番目の値は

72

です。つまり、

第 3 四分位数は 72

例題 8.6 幹葉表示を作る

求めた四分位数は、

第 1 四分位数： 47

第 2 四分位数： 60.5

第 3 四分位数： 72

です.

第9章 確率の基礎

9.1 確率の意味

例題 9.1 確率の基本を理解する

サイコロを振ると，出る目は 1～6 のいずれかです．以下では，どの目も同じように出やすいとします．

次の問いに答え，確率をどのように求めるかを整理しましょう．

- (1) サイコロを 1 回振ったとき，「1 の目が出る」確率を求めましょう．
- (2) サイコロを 1 回振ったとき，「3 以下の目が出る」確率を求めましょう．

例題 9.1 確率の基本を理解する

試行

試行とは、

結果が偶然に決まるような行い、または観測

のことです。

今回の試行は、

「サイコロを 1 回振る」ということです。

例題 9.1 確率の基本を理解する

全事象

全事象とは，試行で起こりうる結果全体を集めた集合のことです．

サイコロを 1 回振る場合，起こりうる結果は 1～6 なので，

$$U = \{1, 2, 3, 4, 5, 6\}$$

です．

例題 9.1 確率の基本を理解する

事象

事象とは，全事象の部分集合として表される結果の集まりのことです．

たとえば，

$$A = \{1\}$$

は，「1 の目が出る」という事象に対応します．

例題 9.1 確率の基本を理解する

根元事象

根元事象とは、1 つの結果だけからなる事象のことです。

たとえば、

$\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5\}$, $\{6\}$

は、それぞれ根元事象です。

例題 9.1 確率の基本を理解する

サイコロの各目のように，全事象に含まれる各根元事象が同じ程度に起こりやすいと考えられるとき，

同様に確からしいといいます．

例題 9.1 確率の基本を理解する

確率

確率とは、事象の起こりやすさを数で表したものです。

事象 A が起こる確率を、

$$P(A)$$

と表します。

例題 9.1 確率の基本を理解する

全事象 U は，試行を行えば必ず起こります．したがって，

$$P(U) = 1$$

です．つまり，

全事象が起こる確率は 1 です．

例題 9.1 確率の基本を理解する

根元事象が同様に確からしいとき，事象 A が起こる確率は，

$$P(A) = \frac{\text{事象 } A \text{ の要素数}}{\text{全事象 } U \text{ の要素数}}$$

で求めることができます．

例題 9.1 確率の基本を理解する

今回の試行は、

「サイコロを 1 回振る」

ということです。全事象は、

$$U = \{1, 2, 3, 4, 5, 6\}$$

なので、

全事象の要素数は 6 です。

例題 9.1 確率の基本を理解する

(1) 「1 の目が出る」確率

「1 の目が出る」という事象は，

$$\{1\}$$

です．したがって，この事象の要素数は，

$$1$$

です．

例題 9.1 確率の基本を理解する

よって,

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

となります.

例題 9.1 確率の基本を理解する

(2) 「3 以下の目が出る」確率

3 以下の目は、

$$1, \quad 2, \quad 3$$

です。したがって、「3 以下の目が出る」という事象は、

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

で、要素数は 3 です。

例題 9.1 確率の基本を理解する

よって,

$$\begin{aligned} P(\text{「3 以下の目が出る」}) &= \frac{\text{事象の要素数}}{\text{全事象の要素数}} \\ &= \frac{3}{6} \\ &= \frac{1}{2} \end{aligned}$$

となります.

例題 9.2 場合の数にもとづいて確率を計算する

大小 2 つのサイコロを同時に 1 回振ったとき，次の確率を求めましょう．

- (1) 2 つのサイコロの目が同じになる確率
- (2) 2 つのサイコロの目が異なる確率
- (3) 2 つのサイコロの目の和が 11 になる確率
- (4) 2 つのサイコロの目の和が 12 になる確率
- (5) 2 つのサイコロの目の和が 11 以上になる確率

例題 9.2 場合の数にもとづいて確率を計算する

今回の試行は、大小 2 つのサイコロを同時に 1 回振るということです。

結果は、

(大きいサイコロの目, 小さいサイコロの目)

という組で表します。

例題 9.2 場合の数にもとづいて確率を計算する

全事象は,

$$\begin{aligned} &\{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), \\ &\quad (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), \\ &\quad (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), \\ &\quad (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), \\ &\quad (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), \\ &\quad (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\} \end{aligned}$$

です.

例題 9.2 場合の数にもとづいて確率を計算する

大きいサイコロの目は 6 通り，小さいサイコロの目も 6 通りです。
したがって，全事象の要素数は

$$6 \times 6 = 36$$

です。

例題 9.2 場合の数にもとづいて確率を計算する

(1) 2 つのサイコロの目が同じになる確率を求める

該当する結果は、

$$\{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$$

です。全部で 6 通りあります。

したがって、

$$P(\text{「同じになる」}) = \frac{6}{36} = \frac{1}{6}$$

となります。

例題 9.2 場合の数にもとづいて確率を計算する

(2) 2 つのサイコロの目が異なる確率を求める

全体は 36 通りで、目が同じになる結果は 6 通りでした.

$$36 - 6 = 30$$

より、目が異なる結果は 30 通りです.

したがって、

$$P(\text{「異なる」}) = \frac{30}{36} = \frac{5}{6}$$

となります.

例題 9.2 場合の数にもとづいて確率を計算する

余事象

余事象とは，ある事象が起こらない場合を表す事象のことです。
事象 A の余事象を A^c と表すと，

$$P(A^c) = 1 - P(A)$$

となります。

例題 9.2 場合の数にもとづいて確率を計算する

「目が異なる」は、「目が同じになる」の余事象なので、

$$\begin{aligned} P(\text{「異なる」}) &= 1 - P(\text{「同じになる」}) \\ &= 1 - \frac{1}{6} \\ &= \frac{5}{6} \end{aligned}$$

と求めることもできます。

例題 9.2 場合の数にもとづいて確率を計算する

(3) 目の和が 11 になる確率を求める

和が 11 になる結果は、

$$(5, 6), \quad (6, 5)$$

の 2 通りです.

したがって、

$$P(\text{「和が 11」}) = \frac{2}{36} = \frac{1}{18}$$

となります.

例題 9.2 場合の数にもとづいて確率を計算する

(4) 目の和が 12 になる確率を求める

和が 12 になる結果は、

$$(6, 6)$$

の 1 通りだけです。

したがって、

$$P(\text{「和が 12」}) = \frac{1}{36}$$

となります。

例題 9.2 場合の数にもとづいて確率を計算する

(5) 目の和が 11 以上になる確率を求める

サイコロ 2 つの目の和は、最大でも 12 です。

したがって、「和が 11 以上」とは、

「和が 11」または「和が 12」ということです。

例題 9.2 場合の数にもとづいて確率を計算する

「和が 11 になる」と「和が 12 になる」は、

同時には起こりません。

このような 2 つの事象を、

排反な事象といいます。

例題 9.2 場合の数にもとづいて確率を計算する

排反な事象と加法定理

排反な事象 A, B について、 A または B が起こる確率は、

$$P(A \text{ または } B) = P(A) + P(B)$$

となります。

例題 9.2 場合の数にもとづいて確率を計算する

したがって、

$$\begin{aligned}P(\text{「和が 11 以上」}) &= P(\text{「和が 11」}) + P(\text{「和が 12」}) \\&= \frac{2}{36} + \frac{1}{36} \\&= \frac{3}{36} \\&= \frac{1}{12}\end{aligned}$$

となります。

例題 9.3 ベン図を使って確率を求める

あるクラスの生徒 40 人について、次のことがわかっています。

- 猫が好きな生徒は 22 人
- 犬が好きな生徒は 18 人
- 猫と犬の両方が好きな生徒は 10 人

この状況をベン図で整理して、問いに答えましょう。

例題 9.3 ベン図を使って確率を求める

- (1) 猫だけが好きな生徒は何人でしょうか。
- (2) 犬だけが好きな生徒は何人でしょうか。
- (3) 猫か犬のどちらか一方だけが好きな生徒は何人でしょうか。
- (4) 猫または犬が好きな生徒は何人でしょうか。
- (5) 猫も犬も好きではない生徒は何人でしょうか。
- (6) このクラスから 1 人を選ぶとき、猫または犬が好きである確率を求めましょう。

例題 9.3 ベン図を使って確率を求める

2 つの事象 A, B について,

$$A \cup B$$

は,

少なくとも一方が起こる事象です.

この事象を和事象といいます.

例題 9.3 ベン図を使って確率を求める

$$A \cap B$$

は、

両方が起こる事象です。

この事象を積事象といいます。

例題 9.3 ベン図を使って確率を求める

また,

$$A \setminus B$$

は,

A が起こり, B が起こらない事象 (A と B の差) です.

さらに,

$$A^c$$

は,

A が起こらない事象 (A の余事象) を表します.

例題 9.3 ベン図を使って確率を求める

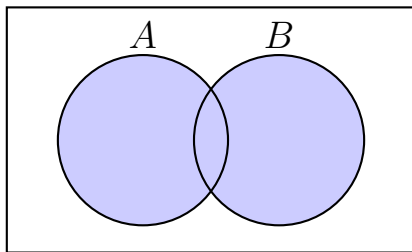
ベン図

ベン図は、複数の集合の関係を図で表す方法です。
長方形は全事象、円はそれぞれの事象を表します。

例題 9.3 ベン図を使って確率を求める

事象のベン図による表し方

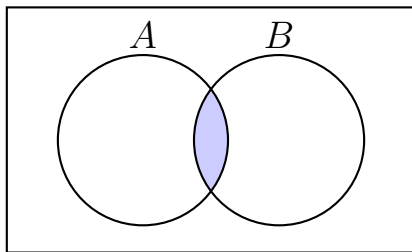
和事象 $A \cup B$ は、少なくとも一方が起こるという事象です。



例題 9.3 ベン図を使って確率を求める

事象のベン図による表し方

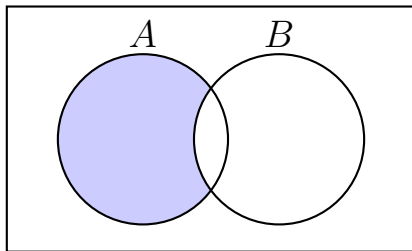
積事象 $A \cap B$ は、両方が起こるという事象です。



例題 9.3 ベン図を使って確率を求める

事象のベン図による表し方

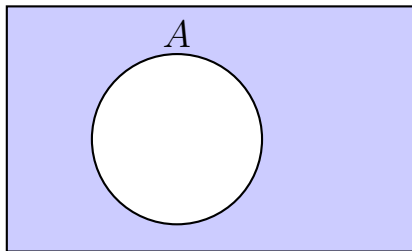
事象の差 $A \setminus B$ は、 A が起こり、かつ B が起こらないという事象です。



例題 9.3 ベン図を使って確率を求める

事象のベン図による表し方

余事象 A^c は、 A が起こらないという事象です。

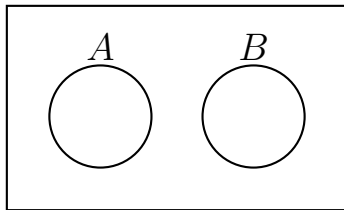


例題 9.3 ベン図を使って確率を求める

事象のベン図による表し方

排反な事象は，事象 A と B が同時には起こらない，つまり重ならないことを表します．

$$A \cap B = \emptyset$$



例題 9.3 ベン図を使って確率を求める

集合の要素数

集合 A に含まれる要素の個数を,

$$|A|$$

と表します. たとえば, $A = \{1, 2, 3\}$ なら,

$$|A| = 3$$

です.

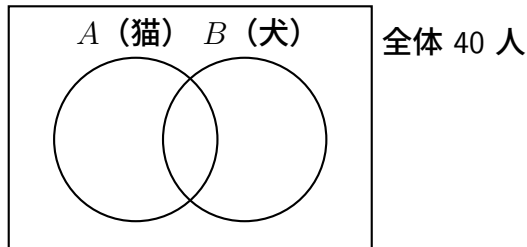
例題 9.3 ベン図を使って確率を求める

今回は,

$$A = \{ \text{猫が好きな生徒} \}$$

$$B = \{ \text{犬が好きな生徒} \}$$

とします. まず, 空のベン図を用意します.

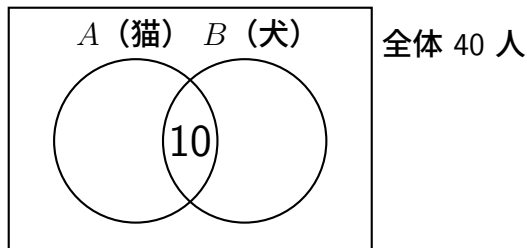


例題 9.3 ベン図を使って確率を求める

猫と犬の両方が好きな生徒は 10 人なので、

$$|A \cap B| = 10$$

です。重なるの部分に 10 を書き込みます。



例題 9.3 ベン図を使って確率を求める

(1) 猫だけが好きな生徒の人数を求める

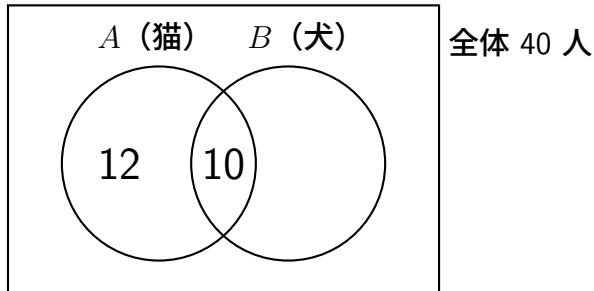
猫が好きな 22 人のうち、10 人は両方好きな生徒です。
したがって、猫だけが好きな生徒の人数は、

$$22 - 10 = 12$$

となります。

例題 9.3 ベン図を使って確率を求める

猫だけの部分に 12 を書き込みます。



例題 9.3 ベン図を使って確率を求める

(2) 犬だけが好きな生徒の人数を求める

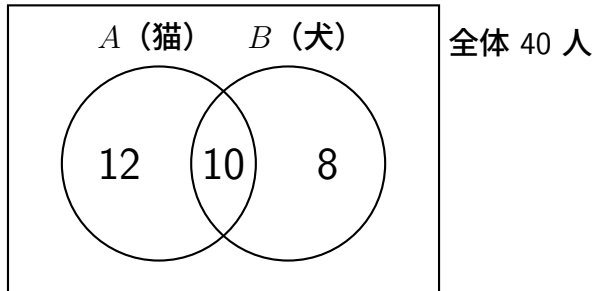
犬が好きな 18 人のうち、10 人は両方好きな生徒です。
したがって、犬だけが好きな生徒の人数は、

$$18 - 10 = 8$$

となります。

例題 9.3 ベン図を使って確率を求める

犬だけの部分に 8 を書き込みます。



例題 9.3 ベン図を使って確率を求める

(3) どちらか一方だけが好きな生徒の人数を求める

どちらか一方だけが好きな生徒は、「猫だけ」と「犬だけ」を合わせた生徒です。

したがって、その人数は、

$$12 + 8 = 20$$

となります。

どちらか一方だけが好きな生徒は 20 人です。

例題 9.3 ベン図を使って確率を求める

(4) 猫または犬が好きな生徒の人数を求める

「猫または犬が好き」は、ベン図の円の中全体です。
したがって、猫または犬が好きな生徒の人数は、

$$12 + 10 + 8 = 30$$

となります。

例題 9.3 ベン図を使って確率を求める

集合の要素数では、

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

という関係があります。

今回なら、

$$|A \cup B| = 22 + 18 - 10 = 30$$

となります。

例題 9.3 ベン図を使って確率を求める

(5) 猫も犬も好きではない生徒の人数を求める

全体は 40 人で，猫または犬が好きな生徒は 30 人です。
したがって，猫も犬も好きではない生徒の人数は，

$$40 - 30 = 10$$

となります。

例題 9.3 ベン図を使って確率を求める

(6) 猫または犬が好きである確率

猫または犬が好きな生徒は 30 人，全体は 40 人なので，

$$\frac{30}{40} = \frac{3}{4}$$

より，

求める確率は $\frac{3}{4}$ です．

例題 9.3 ベン図を使って確率を求める

確率の加法定理

2つの事象 A, B について、少なくとも一方が起こる確率は、

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

と表されます。

例題 9.3 ベン図を使って確率を求める

今回の場合，

$$P(A) = \frac{22}{40}, \quad P(B) = \frac{18}{40}, \quad P(A \cap B) = \frac{10}{40}$$

です．

例題 9.3 ベン図を使って確率を求める

したがって、

$$\begin{aligned}P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\&= \frac{22}{40} + \frac{18}{40} - \frac{10}{40} \\&= \frac{3}{4}\end{aligned}$$

となります。

第9章 確率の基礎

9.2 条件付き確率

例題 9.4 条件付き確率を理解する

あるクラスで、コーヒーが好きな人と紅茶が好きな人について調べたところ、次のことがわかりました。

- コーヒーが好きな人は 24 人
- 紅茶が好きな人は 20 人
- コーヒーと紅茶の両方が好きな人は 8 人

例題 9.4 条件付き確率を理解する

- (1) 1 人選んだときに、その人がコーヒーが好きであることがわかったとします。このとき、その人が紅茶も好きである確率を求めましょう。
- (2) 1 人選んだときに、その人が紅茶が好きであることがわかったとします。このとき、その人がコーヒーも好きである確率を求めましょう。

例題 9.4 条件付き確率を理解する

条件付き確率では，ある条件がわかったとき，

考える全体をその条件に限定します．

つまり，もとの全体すべてを基準にするではありません．

例題 9.4 条件付き確率を理解する

条件付き確率

事象 A が起こったという条件のもとで、事象 B が起こる確率を、

$$P(B \mid A)$$

と表し、これを条件付き確率といいます。

例題 9.4 条件付き確率を理解する

$$P(B \mid A)$$

は、

「 A が起こったという条件のもとで、 B が起こる確率」
を表します。

記号 $\mid$ は、

「～という条件のもとで」という意味です。

例題 9.4 条件付き確率を理解する

一般に,

$$P(A) > 0$$

のとき,

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

となります.

例題 9.4 条件付き確率を理解する

起こりうる根元事象が同様に確からしいとみなせるときは、

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

と考えることもできます。つまり、

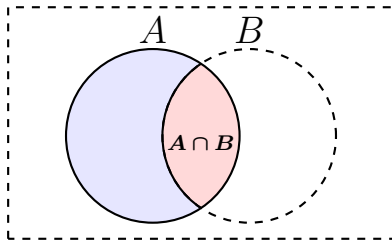
分母は条件となる事象 A です。

例題 9.4 条件付き確率を理解する

条件付き確率をベン図で考える

A が起こったという条件がつくと、

考える全体は A の内部に限定されます。



例題 9.4 条件付き確率を理解する

今回は,

$$A = \{ \text{コーヒーが好き} \}$$

$$B = \{ \text{紅茶が好き} \}$$

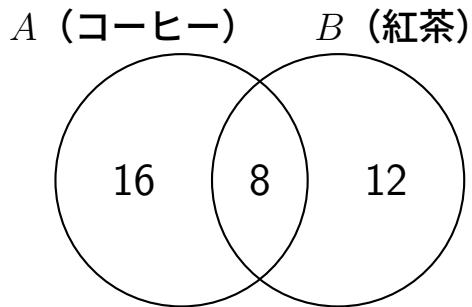
とします.

与えられている人数は,

$$|A| = 24, \quad |B| = 20, \quad |A \cap B| = 8$$

です.

例題 9.4 条件付き確率を理解する



$$24 - 8 = 16, \quad 20 - 8 = 12$$

例題 9.4 条件付き確率を理解する

(1) コーヒーが好きな人が紅茶も好きである確率

「コーヒーが好きであることがわかった」という条件があるので、考える全体は、

コーヒーが好きな 24 人になります。

例題 9.4 条件付き確率を理解する

コーヒーが好きな 24 人のうち、紅茶も好きな人は、

8 人

です。したがって、

$$P(B \mid A) = \frac{8}{24} = \frac{1}{3}$$

となります。

例題 9.4 条件付き確率を理解する

(2) 紅茶が好きな人がコーヒーも好きである確率

今度は、「紅茶が好きであることがわかった」という条件なので、考える全体は、

紅茶が好きな 20 人になります。

例題 9.4 条件付き確率を理解する

紅茶が好きな 20 人のうち、コーヒーも好きな人は、

8 人

です。したがって、

$$P(A \mid B) = \frac{8}{20} = \frac{2}{5}$$

となります。

例題 9.4 条件付き確率を理解する

今回，

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

ですが，

$$P(A \mid B) = \frac{2}{5}$$

です．

$P(B \mid A)$ と $P(A \mid B)$ は同じ確率ではありません．

例題 9.4 条件付き確率を理解する

ちがいは、

どの集団を全体として考えるかにあります。

$P(B \mid A)$ では A を全体とし、 $P(A \mid B)$ では B を全体とします。

例題 9.5 確率から条件付き確率を求める

あるクラスからランダムに 1 人を選ぶとき、次のことがわかっています。

- コーヒーが好きである確率は 0.6
- 紅茶が好きである確率は 0.5
- コーヒーと紅茶の両方が好きである確率は 0.2

例題 9.5 確率から条件付き確率を求める

- (1) 1 人選んだときに，その人がコーヒーが好きであることがわかったとします．このとき，その人が紅茶も好きである確率を求めましょう．
- (2) 1 人選んだときに，その人が紅茶が好きであることがわかったとします．このとき，その人がコーヒーも好きである確率を求めましょう．

例題 9.5 確率から条件付き確率を求める

例題 9.4 では,

人数 (要素数)

を使って条件付き確率を求めました.

今回は, 人数の代わりに,

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

という確率そのものが与えられています.

例題 9.5 確率から条件付き確率を求める

コーヒーが好きであることを事象 A ，紅茶が好きであることを事象 B とします．

与えられている確率は，

$$P(A) = 0.6, \quad P(B) = 0.5, \quad P(A \cap B) = 0.2$$

です．

例題 9.5 確率から条件付き確率を求める

条件付き確率の公式は，

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \quad (P(A) > 0)$$

です． A と B を入れ替えると，

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad (P(B) > 0)$$

となります．

例題 9.5 確率から条件付き確率を求める

(1) コーヒーが好きな人が紅茶も好きである確率

コーヒーが好きであること、つまり事象 A が起こったことがわかっています。

求めるのは、

$$P(B \mid A)$$

です。

例題 9.5 確率から条件付き確率を求める

公式に代入すると,

$$\begin{aligned} P(B \mid A) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{0.2}{0.6} \\ &= \frac{1}{3} \end{aligned}$$

となります.

例題 9.5 確率から条件付き確率を求める

(2) 紅茶が好きな人がコーヒーも好きである確率

紅茶が好きであること，つまり事象 B が起こったことがわかっています．

求めるのは，

$$P(A \mid B)$$

です．

例題 9.5 確率から条件付き確率を求める

公式に代入すると,

$$\begin{aligned} P(A \mid B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{0.2}{0.5} \\ &= \frac{2}{5} \end{aligned}$$

となります.

例題 9.5 確率から条件付き確率を求める

人数で考える場合も，確率で考える場合も，

条件付き確率の考え方は同じです．

人数では，

$$\frac{|A \cap B|}{|A|}$$

確率では，

$$\frac{P(A \cap B)}{P(A)}$$

となります．これは，人数を全体の人数で割ると，それぞれの確率になるためです．

例題 9.6 逆向きの条件付き確率を求める

ある集団について、喫煙とある病気の関係性を調べたところ、次のことがわかりました。

- 喫煙者数と非喫煙者数の比は 3 : 7
- 喫煙者がこの病気にかかる確率は 0.05
- 非喫煙者がこの病気にかかる確率は 0.01

例題 9.6 逆向きの条件付き確率を求める

- (1) この集団から 1 人選んだとき，その人がこの病気にかかっている確率を求めましょう。
- (2) この病気にかかっている人のなかから 1 人選んだとき，その人が喫煙者である確率を求めましょう。

例題 9.6 逆向きの条件付き確率を求める

喫煙者であることを事象 A ，この病気にかかっていることを事象 B とします．

喫煙者と非喫煙者の比は $3 : 7$ なので，

$$P(A) = 0.3, \quad P(A^c) = 0.7$$

です．

例題 9.6 逆向きの条件付き確率を求める

与えられている発病確率は、

$$P(B \mid A) = 0.05$$

$$P(B \mid A^c) = 0.01$$

です。これらはそれぞれ、

喫煙者のなかで病気にかかる割合

非喫煙者のなかで病気にかかる割合

を表しています。

例題 9.6 逆向きの条件付き確率を求める

条件付き確率についての関係式

条件付き確率は、

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \quad (P(A) > 0)$$

でした．

例題 9.6 逆向きの条件付き確率を求める

この式を $P(A \cap B)$ について解くと,

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

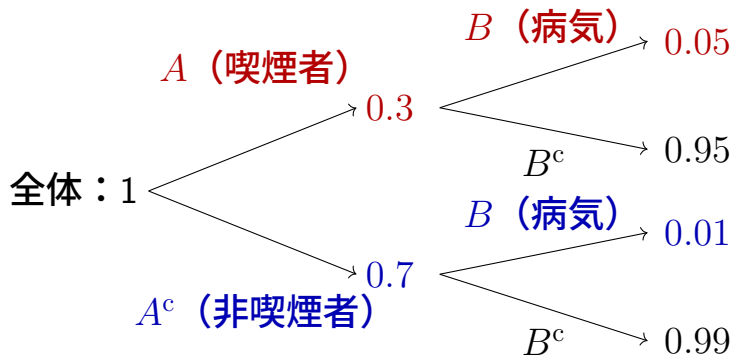
となります。つまり,

「ある集団に属する確率」 × 「その集団のなかでの確率」

で、枝の先まで進んだ確率を求められます。

例題 9.6 逆向きの条件付き確率を求める

割合を樹形図で整理します.



例題 9.6 逆向きの条件付き確率を求める

樹形図では，ある枝をたどった先の確率は，

枝に書かれた確率をかける

ことで求めます．たとえば，事象 A の枝を通り，続いて事象 B の枝を通る場合は，

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

です．

例題 9.6 逆向きの条件付き確率を求める

(1) 病気にかかっている確率を求める

まず,

喫煙者で、病気にかかっている

確率を求めます.

$$\begin{aligned}P(A \cap B) &= P(A)P(B \mid A) \\&= 0.3 \times 0.05 \\&= 0.015\end{aligned}$$

例題 9.6 逆向きの条件付き確率を求める

次に，

非喫煙者で，病気にかかっている

確率を求めます．

$$\begin{aligned}P(A^c \cap B) &= P(A^c)P(B \mid A^c) \\&= 0.7 \times 0.01 \\&= 0.007\end{aligned}$$

例題 9.6 逆向きの条件付き確率を求める

「病気にかかっている」人は、

喫煙者で病気 または 非喫煙者で病気

のどちらかです。

この 2 つは同時には起こらないので、排反です。

例題 9.6 逆向きの条件付き確率を求める

したがって，加法定理より，

$$\begin{aligned} P(B) &= P(A \cap B) + P(A^c \cap B) \\ &= 0.015 + 0.007 \\ &= 0.022 \end{aligned}$$

となります．

例題 9.6 逆向きの条件付き確率を求める

(2) 病気の人の中で喫煙者である確率

今度は、

病気にかかっていることが条件です。

したがって、求めるのは、

$$P(A \mid B)$$

です。

例題 9.6 逆向きの条件付き確率を求める

(1) より,

$$P(A \cap B) = 0.015, \quad P(B) = 0.022$$

なので, 条件付き確率の公式より,

$$\begin{aligned} P(A \mid B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{0.015}{0.022} = \frac{15}{22} \end{aligned}$$

となります.

例題 9.6 逆向きの条件付き確率を求める

この例では、 $P(B \mid A) = 0.05$ は、

喫煙者のなかで病気にかかる確率

でした．一方、今回求めた $P(A \mid B) = \frac{15}{22}$ は、

病気にかかっている人のなかで喫煙者である確率

です．

このように、条件の向きを逆にすると、表す確率も変わります．

例題 9.7 事象の独立性を理解する

大小 2 つのサイコロを同時に 1 回振ります．

それぞれのサイコロについて，1～6 のどの目も同じように出やすいとします．

次のそれぞれの小問で与えられた 2 つの事象が，独立であるかどうかを調べましょう．

例題 9.7 事象の独立性を理解する

- (1) 事象 A : 大きいサイコロの目が偶数である
事象 B : 小さいサイコロの目が 3 の倍数である
- (2) 事象 C : 大きいサイコロの目が 1~3 のいずれかである
事象 D : 2 つのサイコロの目の和が 7 である
- (3) 事象 E : 大きいサイコロの目が偶数である
事象 F : 2 つのサイコロの目の積が偶数である

例題 9.7 事象の独立性を理解する

事象の独立

2 つの事象 A , B について、 $P(A) > 0$ のとき、 A が起こったという条件をつけても、 B の確率が変わらないとき、 A と B は独立であるといいます。

例題 9.7 事象の独立性を理解する

この関係は、

$$P(B \mid A) = P(B) \quad (P(A) > 0)$$

と表すことができます。つまり、

**A が起こったという条件をつけても、
 B の確率は変わらない**

ということです。

例題 9.7 事象の独立性を理解する

条件付き確率の公式

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

を使うと， A と B が独立であることは，

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

と同値です．

例題 9.7 事象の独立性を理解する

今回は各小問で、

$$P(A \cap B) \quad \text{と} \quad P(A)P(B)$$

を求めて比較します。

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

なら独立、

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

なら独立ではありません。

例題 9.7 事象の独立性を理解する

大小 2 つのサイコロの目の組は、

$$6 \times 6 = 36$$

通りあります。

したがって、以下では、

全事象は 36 通りとして確率を求めます。

例題 9.7 事象の独立性を理解する

(1) 事象 A と B が独立か調べる

まず、 $P(A)$ を求めます．大きいサイコロの目が偶数になるのは、

$$\{2, 4, 6\}$$

の 3 通りです．また、小さいサイコロの目は 6 通りあるので、

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

です．

例題 9.7 事象の独立性を理解する

次に、 $P(B)$ を求めます．

小さいサイコロの目が 3 の倍数になるのは，

$$\{3, 6\}$$

の 2 通りです．また，大きいサイコロの目は 6 通りあるので，

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

です．

例題 9.7 事象の独立性を理解する

$A \cap B$ は、大きいサイコロが偶数かつ小さいサイコロが 3 の倍数となる事象です。

該当する組は、

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

の 6 通りです。

例題 9.7 事象の独立性を理解する

したがって、

$$P(A \cap B) = \frac{6}{36} = \frac{1}{6}$$

です。一方、

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

です。よって、

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

となります。

事象 A と B は独立です。

例題 9.7 事象の独立性を理解する

(2) 事象 C と D が独立か調べる

C は、大きいサイコロの目が

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

のいずれかになる事象です。したがって、

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

です。

例題 9.7 事象の独立性を理解する

D は、2 つの目の和が 7 になる事象です。

該当する組は、

$$\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$$

の 6 通りです。したがって、

$$P(D) = \frac{6}{36} = \frac{1}{6}$$

です。

例題 9.7 事象の独立性を理解する

$C \cap D$ は、大きいサイコロが 1~3 かつ 2 つの目の和が 7 となる事象です.

該当する組は,

$$\{(1, 6), (2, 5), (3, 4)\}$$

の 3 通りなので,

$$P(C \cap D) = \frac{3}{36} = \frac{1}{12}$$

です.

例題 9.7 事象の独立性を理解する

一方,

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

です。したがって,

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

となります。

事象 C と D は独立です。

例題 9.7 事象の独立性を理解する

(3) 事象 E と F が独立か調べる

E は、大きいサイコロの目が偶数になる事象です。

偶数は、

$$\{2, 4, 6\}$$

の 3 通りなので、

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

です。

例題 9.7 事象の独立性を理解する

F は、2 つの目の積が偶数になる事象です。

積が偶数にならないのは、両方の目が奇数のときだけです。それぞれのサイコロで、奇数は

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

の 3 通りなので、積が奇数になる組は、

$$3 \times 3 = 9$$

通りです。

例題 9.7 事象の独立性を理解する

全体は 36 通りなので、積が偶数になる組は、

$$36 - 9 = 27$$

通りになります。

したがって、

$$P(F) = \frac{27}{36} = \frac{3}{4}$$

です。

例題 9.7 事象の独立性を理解する

大きいサイコロの目が偶数なら，小さいサイコロの目が何であっても，2 つの目の積は必ず偶数になります．

したがって，

$$P(E \cap F) = P(E) = \frac{1}{2}$$

となります．

一方，

$$P(E)P(F) = \frac{1}{2} \times \frac{3}{4} = \frac{3}{8}$$

です．

例題 9.7 事象の独立性を理解する

$$\frac{1}{2} \neq \frac{3}{8}$$

より,

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

となります.

事象 E と F は独立ではありません.

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

例題 10.1 事象の独立性を用いて確率を求める

サイコロを 3 回続けて振ったとき、次の確率を求めましょう。

- (1) 3 回とも 1 の目が出る確率
- (2) 1 回も 1 の目が出ない確率
- (3) 少なくとも 1 回は 1 の目が出る確率

例題 10.1 事象の独立性を用いて確率を求める

「1 回目に 1 が出る」という事象を A_1

「2 回目に 1 が出る」という事象を A_2

「3 回目に 1 が出る」という事象を A_3

とします．サイコロを 1 回振ったとき，1 が出る確率は， $\frac{1}{6}$ なので，

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{6}$$

です．

例題 10.1 事象の独立性を用いて確率を求める

また，余事象については，

$$P(A_1^c) = P(A_2^c) = P(A_3^c) = \frac{5}{6}$$

となります．

例題 10.1 事象の独立性を用いて確率を求める

事象の独立性と積事象の確率

2 つの事象 A, B が独立であるとき,

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

が成り立ちます.

例題 10.1 事象の独立性を用いて確率を求める

サイコロをくり返し振る場合のように，各回の結果が互いに影響しないとき，各回の結果に対応する事象は互いに独立です．

このとき，

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n)$$

となります．

例題 10.1 事象の独立性を用いて確率を求める

(1) 3 回とも 1 の目が出る確率を求める

「3 回とも 1 が出る」という事象は、

$$A_1 \cap A_2 \cap A_3$$

です。

サイコロを振るたびの結果は互いに影響しないので、 A_1 , A_2 , A_3 は互いに独立です。

例題 10.1 事象の独立性を用いて確率を求める

したがって、

$$\begin{aligned}P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2)P(A_3) \\&= \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \\&= \frac{1}{216}\end{aligned}$$

となります。

例題 10.1 事象の独立性を用いて確率を求める

(2) 1 回も 1 の目が出ない確率を求める

「1 回も 1 が出ない」という事象は、

$$A_1^c \cap A_2^c \cap A_3^c$$

です。

サイコロを振るたびの結果は互いに影響しないので、 A_1^c , A_2^c , A_3^c は互いに独立です。

例題 10.1 事象の独立性を用いて確率を求める

したがって、

$$\begin{aligned}P(A_1^c \cap A_2^c \cap A_3^c) &= P(A_1^c)P(A_2^c)P(A_3^c) \\&= \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \\&= \frac{125}{216}\end{aligned}$$

となります。

例題 10.1 事象の独立性を用いて確率を求める

(3) 少なくとも 1 回は 1 の目が出る確率を求める

「少なくとも 1 回は 1 が出る」の余事象は、「1 回も 1 が出ない」です。

その確率は、(2) より、

$$\frac{125}{216}$$

です。

例題 10.1 事象の独立性を用いて確率を求める

したがって、

$$\begin{aligned} P(\text{少なくとも 1 回は 1 が出る}) &= 1 - P(1 \text{ 回も 1 が出ない}) \\ &= 1 - \frac{125}{216} \\ &= \frac{91}{216} \end{aligned}$$

となります。

例題 10.2 成功回数についての確率を求める

サイコロを 3 回続けて振ったとき，次の確率を求めましょう．

- (1) 1 の目がちょうど 1 回だけ出る確率
- (2) 1 の目がちょうど 2 回だけ出る確率

例題 10.2 成功回数についての確率を求める

「1 回目に 1 が出る」という事象を A_1

「2 回目に 1 が出る」という事象を A_2

「3 回目に 1 が出る」という事象を A_3

とします。それぞれ、

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{6}$$

であり、また、

$$P(A_1^c) = P(A_2^c) = P(A_3^c) = \frac{5}{6}$$

です。

例題 10.2 成功回数についての確率を求める

(1) 1 の目がちょうど 1 回だけ出る確率を求める

「ちょうど 1 回だけ」ということは、1 が出る位置によって、

$$A_1 \cap A_2^c \cap A_3^c$$

$$A_1^c \cap A_2 \cap A_3^c$$

$$A_1^c \cap A_2^c \cap A_3$$

の 3 通りに分けられます。

例題 10.2 成功回数についての確率を求める

この 3 つの場合は同時には起こりません。したがって、3 つの事象は互いに排反です。

それぞれの確率を求めて、最後に足し合わせます。

例題 10.2 成功回数についての確率を求める

まず,

$$A_1 \cap A_2^c \cap A_3^c$$

を考えます.

これは, 1 回目だけ 1 が出て, 2 回目と 3 回目は 1 が出ないという
場合です.

例題 10.2 成功回数についての確率を求める

サイコロを振るたびの結果は互いに影響しないので,

$$A_1, \quad A_2^c, \quad A_3^c$$

は互いに独立です.

例題 10.2 成功回数についての確率を求める

したがって、

$$\begin{aligned}P(A_1 \cap A_2^c \cap A_3^c) &= P(A_1)P(A_2^c)P(A_3^c) \\&= \frac{1}{6} \times \frac{5}{6} \times \frac{5}{6} \\&= \frac{25}{216}\end{aligned}$$

です。

例題 10.2 成功回数についての確率を求める

2 回目だけ 1 が出る場合は,

$$P(A_1^c \cap A_2 \cap A_3^c) = \frac{5}{6} \times \frac{1}{6} \times \frac{5}{6} = \frac{25}{216}$$

3 回目だけ 1 が出る場合は,

$$P(A_1^c \cap A_2^c \cap A_3) = \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} = \frac{25}{216}$$

です.

例題 10.2 成功回数についての確率を求める

3 つの場合は互いに排反なので，加法定理より，

$$\begin{aligned} P(1 \text{ がちょうど } 1 \text{ 回}) &= \frac{25}{216} + \frac{25}{216} + \frac{25}{216} \\ &= \frac{75}{216} \end{aligned}$$

となります．

例題 10.2 成功回数についての確率を求める

(2) 1 の目がちょうど 2 回だけ出る確率を求める

今度は、1 が出ない位置によって、

$$A_1 \cap A_2 \cap A_3^c$$

$$A_1 \cap A_2^c \cap A_3$$

$$A_1^c \cap A_2 \cap A_3$$

の 3 通りに分けられます。

例題 10.2 成功回数についての確率を求める

まず,

$$A_1 \cap A_2 \cap A_3^c$$

を考えます.

これは, 1 回目と 2 回目に 1 が出て, 3 回目だけ 1 が出ないという
場合です.

例題 10.2 成功回数についての確率を求める

サイコロを振るたびの結果は互いに影響しないので,

$$A_1, \quad A_2, \quad A_3^c$$

は互いに独立です.

例題 10.2 成功回数についての確率を求める

したがって、

$$\begin{aligned}P(A_1 \cap A_2 \cap A_3^c) &= P(A_1)P(A_2)P(A_3^c) \\&= \frac{1}{6} \times \frac{1}{6} \times \frac{5}{6} \\&= \frac{5}{216}\end{aligned}$$

です。

例題 10.2 成功回数についての確率を求める

2 回目だけ 1 が出ない場合は,

$$P(A_1 \cap A_2^c \cap A_3) = \frac{1}{6} \times \frac{5}{6} \times \frac{1}{6} = \frac{5}{216}$$

1 回目だけ 1 が出ない場合は,

$$P(A_1^c \cap A_2 \cap A_3) = \frac{5}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{5}{216}$$

です.

例題 10.2 成功回数についての確率を求める

3 つの場合は互いに排反なので，加法定理より，

$$\begin{aligned} P(1 \text{ がちょうど } 2 \text{ 回}) &= \frac{5}{216} + \frac{5}{216} + \frac{5}{216} \\ &= \frac{15}{216} \end{aligned}$$

となります．

例題 10.2 成功回数についての確率を求める

今回、「1 がちょうど 1 回」でも「1 がちょうど 2 回」でも、場合の数は 3 通りでした。

たとえば「1 がちょうど 1 回」なら、3 回のうち、どの 1 回で 1 が出るかを選ぶことになります。

この考え方は、このあとの組合せにつながります。

例題 10.3 独立でない場合の積事象の確率を求める

赤い玉が 3 個，白い玉が 2 個入った袋があります．

この袋から 1 個ずつ，続けて 3 回，玉を取り出します．

ただし，

取り出した玉は袋に戻しません．

例題 10.3 独立でない場合の積事象の確率を求める

次の確率を求めましょう。

- (1) 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
- (2) 1 回目と 2 回目に赤い玉が出る確率
- (3) 3 回とも赤い玉が出る確率
- (4) 少なくとも 1 回は白い玉が出る確率

例題 10.3 独立でない場合の積事象の確率を求める

「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

とします.

最初, 袋の中には,

赤 3 個, 白 2 個

の合計 5 個があります.

例題 10.3 独立でない場合の積事象の確率を求める

(1) 1 回目赤のとき、2 回目も赤である確率を求める

1 回目に赤い玉が出ると、袋の中は、

赤 2 個，白 2 個

になります。

つまり、残っている玉は 4 個です。

例題 10.3 独立でない場合の積事象の確率を求める

したがって、1 回目に赤が出たという条件のもとで、2 回目にも赤が出る確率は、

$$P(A_2 \mid A_1) = \frac{2}{4} = \frac{1}{2}$$

となります。

例題 10.3 独立でない場合の積事象の確率を求める

玉を袋に戻さないので、1 回目に何色の玉が出たかによって、

2 回目に赤が出る確率が変わります。

したがって、各回の事象は独立ではありません。

例題 10.3 独立でない場合の積事象の確率を求める

独立でないときの積事象の確率

事象 A が起こったかどうかによって事象 B の確率が変わるとき、 A と B は独立ではありません。

このとき、

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

と求めます。

例題 10.3 独立でない場合の積事象の確率を求める

3 つ以上の事象についても、順に条件付き確率を使います。

たとえば、

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2 \mid A_1) \times P(A_3 \mid A_1 \cap A_2)$$

となります。

例題 10.3 独立でない場合の積事象の確率を求める

(2) 1 回目と 2 回目に赤い玉が出る確率を求める

まず、1 回目に赤い玉が出る確率は、

$$P(A_1) = \frac{3}{5}$$

です。また、(1) より、

$$P(A_2 \mid A_1) = \frac{1}{2}$$

です。

例題 10.3 独立でない場合の積事象の確率を求める

したがって、

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1)P(A_2 \mid A_1) \\ &= \frac{3}{5} \times \frac{1}{2} \\ &= \frac{3}{10} \end{aligned}$$

となります。

例題 10.3 独立でない場合の積事象の確率を求める

(3) 3 回とも赤い玉が出る確率を求める

1 回目と 2 回目に赤い玉が出ると、袋の中は、

赤 1 個，白 2 個

になるので，この条件のもとで 3 回目に赤が出る確率は，

$$P(A_3 \mid A_1 \cap A_2) = \frac{1}{3}$$

です．

例題 10.3 独立でない場合の積事象の確率を求める

したがって、

$$\begin{aligned}P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \\&= \frac{3}{5} \times \frac{1}{2} \times \frac{1}{3} \\&= \frac{1}{10}\end{aligned}$$

となります。

例題 10.3 独立でない場合の積事象の確率を求める

(4) 少なくとも 1 回は白い玉が出る確率を求める

「少なくとも 1 回は白い玉が出る」という事象の余事象は、

「3 回とも赤い玉が出る」です。

その確率は、(3) より、

$$\frac{1}{10}$$

です。

例題 10.3 独立でない場合の積事象の確率を求める

したがって、

$$\begin{aligned} P(\text{少なくとも 1 回白}) &= 1 - P(3 \text{ 回とも赤}) \\ &= 1 - \frac{1}{10} \\ &= \frac{9}{10} \end{aligned}$$

となります。

例題 10.3 独立でない場合の積事象の確率を求める

例題 10.1 では、各回の結果が互いに影響しなかったので、

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

と計算しました。

今回のように、前の回の結果によって次の回の確率が変わる場合は、

その時点までの結果を条件にした
条件付き確率を使います。

第10章 くり返し操作と確率

10.2 成功回数についての確率

例題 10.4 順列と組合せの総数を書き出して確かめる

6 個の記号

A, B, C, D, E, F

について、次の問いに答えましょう。

- (1) このなかから 2 個を順番をつけて選ぶときの選び方の総数を、書き出して求めましょう。
- (2) このなかから 2 個を順番をつけずに選ぶときの選び方の総数を求めましょう。

例題 10.4 順列と組合せの総数を書き出して確かめる

(1) 順番をつけた選び方の総数を求める

選び方を書き出すと、

A を最初に選ぶ場合： (A, B) , (A, C) , (A, D) , (A, E) , (A, F)

B を最初に選ぶ場合： (B, A) , (B, C) , (B, D) , (B, E) , (B, F)

C を最初に選ぶ場合： (C, A) , (C, B) , (C, D) , (C, E) , (C, F)

D を最初に選ぶ場合： (D, A) , (D, B) , (D, C) , (D, E) , (D, F)

E を最初に選ぶ場合： (E, A) , (E, B) , (E, C) , (E, D) , (E, F)

F を最初に選ぶ場合： (F, A) , (F, B) , (F, C) , (F, D) , (F, E)

例題 10.4 順列と組合せの総数を書き出して確かめる

最初に選ぶ記号は 6 通りあり、そのそれぞれについて、2 個目の選び方は 5 通りあります。

したがって、選び方の総数は、

$$6 \times 5 = 30$$

となります。

例題 10.4 順列と組合せの総数を書き出して確かめる

(2) 順番をつけない選び方の総数を求める

たとえば,

$$(A, B) \quad \text{と} \quad (B, A)$$

は, (1) では別の選び方として数えています.

しかし, 順番をつけずに選ぶ場合は, どちらも A と B を選んだという同じ組です.

例題 10.4 順列と組合せの総数を書き出して確かめる

2 個の記号を並べる順番は、

$$2 \times 1 = 2$$

通りあります。つまり、(1) の 30 通りでは、同じ組を 2 通りずつ数えています。

したがって、順番をつけずに選ぶときの選び方の総数は、

$$\frac{6 \times 5}{2 \times 1} = 15$$

となります。

例題 10.5 組合せの総数を求める

次の問いに答えましょう。

ここでは，順番を区別せずに「どれを選ぶか」だけを考えます。

- (1) 4 人のなかから 2 人を選ぶときの選び方の総数を求めましょう。
- (2) 平面上に異なる 6 個の点 A, B, C, D, E, F があり，どの 3 点も一直線上にないとします．3 点を選んで三角形を作るときの選び方の総数を求めましょう。

例題 10.5 組合せの総数を求める

(3) 20 冊の本の中から 3 冊を選ぶときの選び方の総数を求めましょう.

(4) 5 科目の中から 2 科目を選ぶ場合と, 3 科目を選ぶ場合で, 選び方の総数が同じになることを確認しましょう.

例題 10.5 組合せの総数を求める

n 個のなかから k 個を選ぶ組合せの総数

を,

$$\binom{n}{k}$$

と表します.

例題 10.5 組合せの総数を求める

n 個から k 個を順番をつけて選ぶと,

$$n \times (n - 1) \times \cdots \times (n - k + 1)$$

通りあります.

一方, 選んだ k 個の並べ方は,

$$k \times (k - 1) \times \cdots \times 2 \times 1$$

通りあります.

例題 10.5 組合せの総数を求める

したがって、

$$\binom{n}{k} = \frac{n \times (n-1) \times \cdots \times (n-k+1)}{k \times (k-1) \times \cdots \times 2 \times 1}$$

となります。

たとえば、例題 10.4 では、

$$\binom{6}{2} = \frac{6 \times 5}{2 \times 1} = 15$$

です。

例題 10.5 組合せの総数を求める

組合せの対称性

選ぶ k 個を決めることは、選ばない $(n - k)$ 個を決めることと同じです.

したがって,

$$\binom{n}{k} = \binom{n}{n - k}$$

が成り立ちます.

例題 10.5 組合せの総数を求める

(1) 4 人から 2 人を選ぶ選び方の総数を求める

$$\binom{4}{2} = \frac{4 \times 3}{2 \times 1} = 6$$

です.

例題 10.5 組合せの総数を求める

(2) 6 点から三角形の作り方の総数を求める

$$\binom{6}{3} = \frac{6 \times 5 \times 4}{3 \times 2 \times 1} = 20$$

です。どの 3 点も一直線上にないので、3 点を選ぶたびに 1 つの三角形ができます。

例題 10.5 組合せの総数を求める

(3) 20 冊から 3 冊を選ぶ選び方の総数を求める

$$\binom{20}{3} = \frac{20 \times 19 \times 18}{3 \times 2 \times 1} = 1140$$

です.

例題 10.5 組合せの総数を求める

(4) 選び方の総数が同じになることを確認する

$$\binom{5}{2} = \frac{5 \times 4}{2 \times 1} = 10$$

$$\binom{5}{3} = \frac{5 \times 4 \times 3}{3 \times 2 \times 1} = 10$$

となります。したがって、次が成り立ちます。

$$\binom{5}{2} = \binom{5}{3}$$

例題 10.6 成功回数についての確率を求める

ある作業を 1 回行ったとき，成功する確率を

$$\frac{1}{3}$$

とします．この作業を互いに独立に 5 回続けて行うとき，次の確率を求めましょう．

- (1) 5 回とも成功する確率
- (2) 少なくとも 1 回は成功する確率
- (3) 成功がちょうど 2 回だけ起こる確率

例題 10.6 成功回数についての確率を求める

成功回数についての確率

同じ操作を n 回行い、各回で「成功」か「失敗」かが決まる場合を考えます。

各回の結果が他の回の結果によって影響を受けないとき、各回の結果に対応する事象は互いに独立です。

したがって、1 つの並びが起こる確率は、各回の確率を掛け合わせて求めます。

例題 10.6 成功回数についての確率を求める

n 回のうち、成功がちょうど k 回となる場合を考えます。

そのうち、1つの並びを考えると、

成功，失敗，成功，失敗，失敗，成功，...

のようなものがあります。

例題 10.6 成功回数についての確率を求める

成功する確率を p とすると，失敗する確率は $(1 - p)$ なので，並び

成功，失敗，成功，失敗，失敗，成功，...

が起こる確率は，

$$p \times (1 - p) \times p \times (1 - p) \times (1 - p) \times p \times \dots$$

となります．成功が k 回，失敗が $(n - k)$ 回 なので，これは，

$$p^k (1 - p)^{n-k}$$

となります．

例題 10.6 成功回数についての確率を求める

成功がちょうど k 回となる並びは「 n 回のうち成功する k 回を選ぶ」と考えるので、

$$\binom{n}{k}$$

通りあります．

どの並びも、成功が k 回、失敗が $(n - k)$ 回なので、それが起こる確率は

$$p^k(1 - p)^{n-k}$$

です．

例題 10.6 成功回数についての確率を求める

成功がちょうど k 回起こる確率

各回の成功・失敗に対応する事象が互いに独立で、各回の成功確率が p で一定のとき、 n 回のうち成功がちょうど k 回起こる確率は、

$$\binom{n}{k} p^k (1 - p)^{n-k}$$

です。

例題 10.6 成功回数についての確率を求める

(1) 5 回とも成功する確率を求める

成功確率は $1/3$ なので,

$$\left(\frac{1}{3}\right)^5 = \frac{1}{243}$$

です.

例題 10.6 成功回数についての確率を求める

(2) 少なくとも 1 回は成功する確率を求める

「少なくとも 1 回成功する」の余事象は、

5 回とも失敗することです。

失敗する確率は $2/3$ なので、5 回とも失敗する確率は、

$$\left(\frac{2}{3}\right)^5 = \frac{32}{243}$$

です。

例題 10.6 成功回数についての確率を求める

したがって、少なくとも 1 回成功する確率は、

$$1 - \frac{32}{243} = \frac{211}{243}$$

となります。

例題 10.6 成功回数についての確率を求める

(3) 成功がちょうど 2 回だけ起こる確率を求める

5 回のうち成功する 2 回を選ぶので、成功と失敗の並びは、

$$\binom{5}{2} = \frac{5 \times 4}{2 \times 1} = 10$$

通りあります。

たとえば、「1 回目と 2 回目だけ成功」という並びを考えます。

例題 10.6 成功回数についての確率を求める

「1 回目と 2 回目だけ成功」の確率は、

$$\left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^3 = \frac{8}{243}$$

です。

同じ確率となる並びが 10 通りあるので、求める確率は、

$$10 \times \frac{8}{243} = \frac{80}{243}$$

となります。

例題 10.6 成功回数についての確率を求める

公式を使えば,

$$\binom{5}{2} \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^3 = \frac{80}{243}$$

とまとめて求めることもできます.

例題 10.7 復元しないときの成功回数の確率を求める

10 本のくじがあります。そのうち、

当たり 3 本、はずれ 7 本

です。このなかから、くじを戻さずに 6 回続けて引きます。
次の確率を求めましょう。

- (1) 当たりがちょうど 1 回だけ出る確率
- (2) 当たりがちょうど 2 回だけ出る確率
- (3) 当たりがちょうど 3 回だけ出る確率

例題 10.7 復元しないときの成功回数の確率を求める

くじを引いたあとに元に戻さない場合を考えます。

この場合，当たりを引く確率は，

$$\frac{\text{残っている当たりの本数}}{\text{残っているくじの本数}}$$

となります。はずれを引く確率は，

$$\frac{\text{残っているはずれの本数}}{\text{残っているくじの本数}}$$

となります。

例題 10.7 復元しないときの成功回数の確率を求める

復元しない場合は、くじを引くたびに残っている本数が変わります。

したがって、各回の

当たりを引く確率

や

はずれを引く確率

は、その時点で残っている本数によって変わります。

例題 10.7 復元しないときの成功回数の確率を求める

たとえば，

10 本のくじ （当たり 3 本・はずれ 7 本）

から，復元しないで 6 回引く場合を考えます．

当たりがちょうど 2 回出る並びとして，

当・当・は・は・は・は

を考えてみましょう．

例題 10.7 復元しないときの成功回数の確率を求める

並び

当・当・は・は・は・は

が起こる確率は、

$$\frac{3}{10} \times \frac{2}{9} \times \frac{7}{8} \times \frac{6}{7} \times \frac{5}{6} \times \frac{4}{5}$$

です。

くじを戻さないなので、分母も分子も途中で変化しています。

例題 10.7 復元しないときの成功回数の確率を求める

一方,

は・は・は・は・当・当

という並びが起こる確率は,

$$\frac{7}{10} \times \frac{6}{9} \times \frac{5}{8} \times \frac{4}{7} \times \frac{3}{6} \times \frac{2}{5}$$

となります.

先ほどとは, 掛ける確率の並び方が異なります.

例題 10.7 復元しないときの成功回数の確率を求める

2 つの式を比べると、**当たり 2 回** について分子に現れる値は、

3, 2

はずれ 4 回 について分子に現れる値は、

7, 6, 5, 4

です。また、分母に現れる値は、どちらも、

10, 9, 8, 7, 6, 5

です。

例題 10.7 復元しないときの成功回数の確率を求める

つまり，並びによって，掛ける確率の順序は異なりますが，分子と分母に現れる値は同じなので，

掛け合わせた結果は同じになります．

したがって，当たりが同じ回数だけ出る並びは，どの並びも同じ確率で起こります．

例題 10.7 復元しないときの成功回数の確率を求める

復元しないときの成功回数の確率

当たりがちょうど k 回となる並びは、

$$\binom{n}{k}$$

通りあります。したがって、成功回数の確率は次で求めます。

$$\binom{n}{k} \times (1 \text{ つの並びの確率})$$

例題 10.7 復元しないときの成功回数の確率を求める

(1) 当たりがちょうど 1 回だけ出る確率を求める

6 回のうち，当たりが出る 1 回を選ぶので，当たりがちょうど 1 回だけ出る並びは，

$$\binom{6}{1} = 6$$

通りあります．たとえば，

当・は・は・は・は・は

という並びを考えます．

例題 10.7 復元しないときの成功回数の確率を求める

「1 回目だけ当たり」という並びが起こる確率は、

$$\frac{3}{10} \times \frac{7}{9} \times \frac{6}{8} \times \frac{5}{7} \times \frac{4}{6} \times \frac{3}{5} = \frac{1}{20}$$

です。

同じ確率となる並びが 6 通りあるので、求める確率は、

$$6 \times \frac{1}{20} = \frac{3}{10}$$

となります。

例題 10.7 復元しないときの成功回数の確率を求める

(2) 当たりがちょうど 2 回だけ出る確率を求める

6 回のうち、当たりが出る 2 回を選ぶので、当たりがちょうど 2 回だけ出る並びは、

$$\binom{6}{2} = \frac{6 \times 5}{2 \times 1} = 15$$

通りあります。たとえば、

当・当・は・は・は・は

という並びを考えます。

例題 10.7 復元しないときの成功回数の確率を求める

「1 回目と 2 回目だけ当たり」という並びが起こる確率は、

$$\frac{3}{10} \times \frac{2}{9} \times \frac{7}{8} \times \frac{6}{7} \times \frac{5}{6} \times \frac{4}{5} = \frac{1}{30}$$

です。

同じ確率となる並びが 15 通りあるので、求める確率は、

$$15 \times \frac{1}{30} = \frac{1}{2}$$

となります。

例題 10.7 復元しないときの成功回数の確率を求める

(3) 当たりがちょうど 3 回だけ出る確率を求める

6 回のうち、当たりが出る 3 回を選ぶので、当たりがちょうど 3 回だけ出る並びは、

$$\binom{6}{3} = \frac{6 \times 5 \times 4}{3 \times 2 \times 1} = 20$$

通りあります。たとえば、

当・当・当・は・は・は

という並びを考えます。

例題 10.7 復元しないときの成功回数の確率を求める

「1 回目から 3 回目までだけ当たり」という並びが起こる確率は、

$$\frac{3}{10} \times \frac{2}{9} \times \frac{1}{8} \times \frac{7}{7} \times \frac{6}{6} \times \frac{5}{5} = \frac{1}{120}$$

です。

同じ確率となる並びが 20 通りあるので、求める確率は、

$$20 \times \frac{1}{120} = \frac{1}{6}$$

となります。

例題 10.8 復元しないときの選び方と確率を求める

ある会社では、社員向けのアンケートを作るために、13 個の質問候補の中から 4 個を無作為に選ぶことにしました。

この 13 個のうち、

業務に関する質問： 6 個

職場環境に関する質問： 7 個

です。

例題 10.8 復元しないときの選び方と確率を求める

次の問いに答えましょう。

- (1) 4 個の選び方の総数を求めましょう。
- (2) 選んだ 4 個のうち、業務に関する質問がちょうど 2 個になる選び方は何通りあるでしょうか。
- (3) 4 個を無作為に選ぶとき、業務に関する質問がちょうど 2 個になる確率を求めましょう。

例題 10.8 復元しないときの選び方と確率を求める

ここでは、復元しないで選ぶときの

順番を区別しない選び方を考えます。

順番を区別しないので、組合せを用いて数えます。

例題 10.8 復元しないときの選び方と確率を求める

全体から n 個を選び，特定の種類をちょうど k 個選ぶとします．

特定の種類のものを k 個選ぶ方法は，

$$\binom{\text{特定の種類の個数}}{k}$$

通りあります．そして，残りの $n - k$ 個をそれ以外の種類のものから選ぶ方法は，

$$\binom{\text{それ以外の種類の個数}}{n - k}$$

通りあります．

例題 10.8 復元しないときの選び方と確率を求める

したがって、条件を満たす選び方は、

$$\binom{\text{特定の種類の個数}}{k} \times \binom{\text{それ以外の種類の個数}}{n - k}$$

通りあります。

一方、全体から n 個を選ぶ方法は、

$$\binom{\text{全体の個数}}{n}$$

通りあります。

例題 10.8 復元しないときの選び方と確率を求める

復元しないときの選び方と確率

したがって、全体から n 個を選び、特定の種類のものをちょうど k 個選ぶ確率は、

$$\frac{\binom{\text{特定の種類の個数}}{k} \times \binom{\text{それ以外の種類の個数}}{n-k}}{\binom{\text{全体の個数}}{n}}$$

で求められます。

例題 10.8 復元しないときの選び方と確率を求める

(1) 4 個の選び方の総数を求める

13 個の質問候補の中から 4 個を選ぶので、選び方の総数は、

$$\binom{13}{4} = \frac{13 \times 12 \times 11 \times 10}{4 \times 3 \times 2 \times 1} = 715$$

より、715 通りです。

例題 10.8 復元しないときの選び方と確率を求める

(2) 業務に関する質問がちょうど 2 個になる選び方を求める

まず、業務に関する質問 6 個のなかから 2 個を選びます。

$$\binom{6}{2} = \frac{6 \times 5}{2 \times 1} = 15$$

より、業務に関する質問を 2 個選ぶ方法は、15 通りです。

例題 10.8 復元しないときの選び方と確率を求める

残りの 2 個は，職場環境に関する質問から選びます．

職場環境に関する質問 7 個のなかから 2 個を選ぶので，その選び方は，

$$\binom{7}{2} = \frac{7 \times 6}{2 \times 1} = 21$$

より，21 通りです．

例題 10.8 復元しないときの選び方と確率を求める

「業務 2 個」の選び方は 15 通り,「職場環境 2 個」の選び方は 21 通りです.

この 2 つの選び方を組み合わせるので,条件を満たす選び方の総数は,

$$15 \times 21 = 315$$

より, 315 通りです.

例題 10.8 復元しないときの選び方と確率を求める

(3) 業務に関する質問がちょうど 2 個になる確率を求める

4 個の選び方は全部で 715 通りあり，そのうち，業務に関する質問がちょうど 2 個になる選び方は 315 通りです．

したがって，

$$P(\text{業務がちょうど 2 個}) = \frac{315}{715} = \frac{63}{143}$$

です．

第11章 確率変数と確率分布

11.1 確率変数の考え方

例題 11.1 確率分布を求める

コインを 2 回投げたときに表が出た回数を表す確率変数を X とします.

コインの表と裏は，どちらも同じように出やすいとします.

- (1) 確率変数 X のとりうる値をすべて書き出しましょう.
- (2) それぞれの値をとる確率を求め，確率分布表を作りましょう.

例題 11.1 確率分布を求める

まず，この例題で扱う状況を整理しておきましょう．

コインを 2 回投げて，出た結果から「表が出た回数」を数えます．

この「結果に数に対応させる決め方」に X という名前をつけています．

例題 11.1 確率分布を求める

確率変数

確率変数とは，起こりうるそれぞれの結果に数に対応させる「対応の決め方」のことです．

その対応させた数には確率が決まっています．

例題 11.1 確率分布を求める

今回の X は、「表が出た回数」を表すので、

$$\text{表} \cdot \text{表} \quad \Longrightarrow \quad X = 2$$

$$\text{表} \cdot \text{裏} \quad \Longrightarrow \quad X = 1$$

$$\text{裏} \cdot \text{表} \quad \Longrightarrow \quad X = 1$$

$$\text{裏} \cdot \text{裏} \quad \Longrightarrow \quad X = 0$$

となります。

このように、それぞれの結果に数を対応させます。

例題 11.1 確率分布を求める

今回の X は、とりうる値を 0, 1, 2 のように数え上げることができます。

離散型確率変数

とりうる値を数え上げることができる確率変数を、

離散型確率変数

といいます。

例題 11.1 確率分布を求める

離散型確率変数 X が値 k をとる確率を,

$$P(X = k)$$

と書きます。たとえば,

$$P(X = 2)$$

は、今回の場合、「表・表」が出る確率を表しています。

例題 11.1 確率分布を求める

離散型確率分布

離散型確率変数 X のとりうる値には、それぞれ、その値をとる確率が決まっています。

この「とりうる値とそれぞれの値をとる確率との対応」を、

離散型確率分布

といいます。

例題 11.1 確率分布を求める

確率分布表

離散型確率分布の対応関係を表の形に整理したものを、

確率分布表

といいます。

例題 11.1 確率分布を求める

(1) X のとりうる値を書き出す

コインを 2 回投げるとき、表が出る回数は、

$$0, 1, 2$$

の 3 通りです。

したがって、 X のとりうる値は、0, 1, 2 です。

例題 11.1 確率分布を求める

(2) 確率を求め，確率分布表を作る

2 回の結果の並び方は，

表・表， 表・裏， 裏・表， 裏・裏

の 4 通りです．どの並びも同じ確率

$$\frac{1}{4}$$

で起こります．

例題 11.1 確率分布を求める

それぞれの並びで表が出る回数を数えると、

並び	X
表・表	2
表・裏	1
裏・表	1
裏・裏	0

となります。

例題 11.1 確率分布を求める

したがって、

$$P(X = 0) = \frac{1}{4}$$

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

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

となります。

例題 11.1 確率分布を求める

これらを確認率分布表に整理すると、

x	0	1	2
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

となります。

例題 11.1 確率分布を求める

離散型確率分布では，とりうるすべての値の確率を足すと 1 になります．

今回も，

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

となっています．

これは確率の基本的な性質です．

例題 11.2 期待値を求める

例題 11.1 で扱った，コインを 2 回投げたときに表が出た回数を表す確率変数 X の確率分布は次のとおりです．

x	0	1	2
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

このとき，確率変数 X の期待値を求めましょう．

例題 11.2 期待値を求める

期待値は、確率で重みづけした平均値です。

これは、同じ試行を何度もくり返したときの平均的な値を表します。

例題 11.2 期待値を求める

期待値

離散型の確率変数 X の期待値とは、

「値」 × 「その値が出る確率」

をそれぞれ計算し、すべて足し合わせたものです。

確率変数 X の期待値を、

$$E(X)$$

と表します。

例題 11.2 期待値を求める

期待値は平均的な値を表すため、

実際にとりうる値になるとは限りません。

例題 11.2 期待値を求める

今回、確率変数 X は、

$$0, \quad 1, \quad 2$$

の 3 つの値をとります。

それぞれの確率は、

$$P(X = 0) = \frac{1}{4}, \quad P(X = 1) = \frac{1}{2}, \quad P(X = 2) = \frac{1}{4}$$

です。

例題 11.2 期待値を求める

したがって、

$$\begin{aligned} E(X) &= 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} \\ &= 0 + \frac{1}{2} + \frac{2}{4} \\ &= 1 \end{aligned}$$

となります。つまり、

期待値は 1 です。

例題 11.2 期待値を求める

$E(X) = 1$ ということは、コインを 2 回投げるという試行を何度もくり返したとき、

長い目で見れば、平均して
1 回くらい表が出ることが期待される

ということです。

例題 11.3 分散と標準偏差を求める

例題 11.1 で扱った，コインを 2 回投げたときに表が出た回数を表す確率変数 X の確率分布は次のとおりです。

x	0	1	2
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

このとき，確率変数 X の分散と標準偏差を求めましょう。
ただし， $\sqrt{2} = 1.41$ とし，標準偏差は小数第 2 位まで求めましょう。

例題 11.3 分散と標準偏差を求める

分散と標準偏差は、確率変数が期待値のまわりにどれくらい散らばっているかを表す指標です。

第 3 章で扱ったデータの分散と標準偏差と同じ考え方で定義されます。

例題 11.3 分散と標準偏差を求める

分散

離散型の確率変数 X の分散とは、

「値と期待値の差の 2 乗」 × 「その値が出る確率」

をそれぞれ計算し、すべて足し合わせたものです。

確率変数 X の分散を、

$$V(X)$$

と表します。

例題 11.3 分散と標準偏差を求める

標準偏差

分散の負でない平方根を，標準偏差といいます．

標準偏差は，散らばりの大きさを確率変数 X と同じ単位で表します．

例題 11.3 分散と標準偏差を求める

例題 11.2 より， X の期待値は

$$E(X) = 1$$

です。したがって，

$$0, \quad 1, \quad 2$$

のそれぞれについて，

$$(\text{値} - \text{期待値})^2 \times \text{確率}$$

を計算し，すべて足し合わせて X の分散を求めます。

例題 11.3 分散と標準偏差を求める

$$\begin{aligned} V(X) &= (0 - 1)^2 \times \frac{1}{4} + (1 - 1)^2 \times \frac{1}{2} + (2 - 1)^2 \times \frac{1}{4} \\ &= 1 \times \frac{1}{4} + 0 \times \frac{1}{2} + 1 \times \frac{1}{4} \\ &= \frac{1}{2} \end{aligned}$$

例題 11.3 分散と標準偏差を求める

標準偏差は、分散の負でない平方根なので、

$$\sqrt{V(X)} = \sqrt{\frac{1}{2}} = \frac{1}{\sqrt{2}} = \frac{1}{1.41} = 0.709 \dots \approx 0.71$$

となります。

例題 11.3 分散と標準偏差を求める

以上より,

分散は $\frac{1}{2}$

標準偏差は 0.71

となります.

例題 11.3 分散と標準偏差を求める

この標準偏差 0.71 という値は、コインを 2 回投げるという試行をくり返したとき、

表の回数が期待値 1 のまわりに
どの程度散らばるかを表す目安です。

第11章 確率変数と確率分布

11.2 確率変数の性質

例題 11.4 期待値の性質を確かめる

サイコロを 1 回投げ，出た目を表す確率変数を X とします．

ここで，新しい確率変数 Y を，

$$Y = 2X + 1$$

と定めます．

このとき，次の問いに答えましょう．

例題 11.4 期待値の性質を確かめる

- (1) 確率変数 Y の確率分布表を作しましょう.
- (2) 期待値の定義に従って, Y の期待値 $E(Y)$ を求めましょう.
- (3) 期待値の性質

$$E(2X + 1) = 2E(X) + 1$$

の左辺と右辺が一致することを, 実際に計算して確かめましょう.

例題 11.4 期待値の性質を確かめる

まず、確率変数 X の確率分布を確認します。

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

これより、 X の期待値は、

$$E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{7}{2}$$

です。

例題 11.4 期待値の性質を確かめる

(1) Y の確率分布表を作る

X は、サイコロの出た目をそのまま数値として対応させた確率変数です.

この X の値を 2 倍して 1 を足した値を対応させることで、

$$Y = 2X + 1$$

という新しい確率変数をつくることができます.

例題 11.4 期待値の性質を確かめる

$Y = 2X + 1$ なので、 X の値に応じて Y の値は、

X	1	2	3	4	5	6
Y	3	5	7	9	11	13

となります。たとえば、

$$X = 1 \implies Y = 2 \times 1 + 1 = 3$$

です。

例題 11.4 期待値の性質を確かめる

どの値も確率 $\frac{1}{6}$ で起こるので、 Y の確率分布表は、

y	3	5	7	9	11	13
$P(Y = y)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

となります。

例題 11.4 期待値の性質を確かめる

(2) Y の期待値を求める

期待値の定義に従って計算すると、 Y の期待値は、

$$E(Y) = 3 \times \frac{1}{6} + 5 \times \frac{1}{6} + 7 \times \frac{1}{6} + 9 \times \frac{1}{6} + 11 \times \frac{1}{6} + 13 \times \frac{1}{6} = 8$$

となります。

例題 11.4 期待値の性質を確かめる

期待値の線形性

期待値には、次の線形性があります。

確率変数 X から、新しい確率変数

$$Y = aX + b$$

をつくったとき、次が成り立ちます。

$$E(aX + b) = aE(X) + b$$

例題 11.4 期待値の性質を確かめる

これは、確率変数を

一定倍する

または

一定の数を足す

といった変換をしたとき、**その変換がそのまま期待値にも反映される**という性質です。

例題 11.4 期待値の性質を確かめる

(3) 期待値の性質を確かめる

期待値の性質 $E(2X + 1) = 2E(X) + 1$ について、まず左辺を確認します。

$Y = 2X + 1$ なので、(2) より、

$$E(2X + 1) = E(Y) = 8$$

です。

例題 11.4 期待値の性質を確かめる

次に、右辺を計算します．

$$E(X) = \frac{7}{2}$$

なので，

$$2E(X) + 1 = 2 \times \frac{7}{2} + 1 = 8$$

となります．

例題 11.4 期待値の性質を確かめる

左辺は,

$$E(2X + 1) = 8$$

右辺は,

$$2E(X) + 1 = 8$$

となりました．したがって,

$$E(2X + 1) = 2E(X) + 1$$

が確かに成り立っています．

例題 11.5 公式を使って分散を求める

例題 11.3 では，コインを 2 回投げたときに表が出た回数を表す確率変数 X の分散を，分散の定義にしたがって計算しました．

ここでは，同じ確率変数 X について，分散の公式を使って分散を求めましょう．

例題 11.5 公式を使って分散を求める

この方法では、

$$E(X^2)$$

を使います。

$E(X^2)$ を求めるために、まず、 X^2 の確率分布を考えます。

X^2 は、確率変数 X の値を 2 乗して得られる

新しい確率変数

です。

例題 11.5 公式を使って分散を求める

分散の公式

離散型の確率変数 X の分散は、

$$V(X) = E(X^2) - \{E(X)\}^2$$

と求めることもできます。

この形を使うと、分散を手早く計算できることがあります。

例題 11.5 公式を使って分散を求める

例題 11.1 より，確率変数 X の確率分布は，

x	0	1	2
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

です．また，例題 11.2 より， X の期待値は，

$$E(X) = 1$$

です．

例題 11.5 公式を使って分散を求める

X の値を 2 乗すると,

X	0	1	2
X^2	0	1	4
確率	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

となります.

確率は, もとの X の各値に対応する確率と同じです.

例題 11.5 公式を使って分散を求める

X^2 の期待値は,

$$E(X^2) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 4 \times \frac{1}{4} = \frac{3}{2}$$

となります.

例題 11.5 公式を使って分散を求める

分散の公式

$$V(X) = E(X^2) - \{E(X)\}^2$$

に代入すると,

$$V(X) = \frac{3}{2} - 1^2 = \frac{1}{2}$$

となります.

これは, 例題 11.3 で定義から求めた X の分散と一致しています.

例題 11.6 分散の性質を確かめる

例題 11.4 と同じく，サイコロを 1 回投げたときに出た目を表す確率変数 X を用いて，

$$Y = 2X + 1$$

とおきます．

このとき，次の問いに答えましょう．

例題 11.6 分散の性質を確かめる

(1) 分散の公式

$$V(Y) = E(Y^2) - \{E(Y)\}^2$$

を使って， Y の分散 $V(Y)$ を求めましょう．

(2) 分散の性質

$$V(2X + 1) = 2^2 V(X)$$

の左辺と右辺が一致することを，実際に計算して確かめましょう．

例題 11.6 分散の性質を確かめる

まず、確率変数 X の確率分布を確認します。

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

また、 X の期待値は、

$$E(X) = \frac{7}{2}$$

です（例題 11.4 参照）。

例題 11.6 分散の性質を確かめる

(1) Y の分散を求める

例題 11.4 より,

$$Y = 2X + 1$$

の確率分布は,

y	3	5	7	9	11	13
$P(Y = y)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

です.

例題 11.6 分散の性質を確かめる

Y の値を 2 乗すると,

Y	3	5	7	9	11	13
Y^2	9	25	49	81	121	169
確率	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

となります。

例題 11.6 分散の性質を確かめる

したがって、

$$\begin{aligned} E(Y^2) &= 9 \times \frac{1}{6} + 25 \times \frac{1}{6} + 49 \times \frac{1}{6} + 81 \times \frac{1}{6} + 121 \times \frac{1}{6} + 169 \times \frac{1}{6} \\ &= \frac{227}{3} \end{aligned}$$

となります。

例題 11.6 分散の性質を確かめる

また、例題 11.4 より、 Y の期待値は

$$E(Y) = 8$$

です。分散の公式

$$V(Y) = E(Y^2) - \{E(Y)\}^2$$

に代入すると、

$$V(Y) = \frac{227}{3} - 8^2 = \frac{35}{3}$$

となります。

例題 11.6 分散の性質を確かめる

分散の性質

分散には、次の性質があります。

確率変数 X を用いて、

$$Y = aX + b$$

としたとき、次が成り立ちます。

$$V(aX + b) = a^2 V(X)$$

例題 11.6 分散の性質を確かめる

この性質は、確率変数を a 倍すると、

分散は a^2 倍

になることを表します。

一方、一定の数 b を足しても、分散は変わりません。

例題 11.6 分散の性質を確かめる

(2) 分散の性質を確かめる

分散の性質 $V(2X + 1) = 2^2 V(X)$ について、まず左辺を確認します。

$Y = 2X + 1$ なので、(1) より、

$$V(2X + 1) = V(Y) = \frac{35}{3}$$

です。

例題 11.6 分散の性質を確かめる

次に、 X の期待値は $E(X) = \frac{7}{2}$ なので、 X の分散は、

$$\begin{aligned} V(X) &= \left(1 - \frac{7}{2}\right)^2 \times \frac{1}{6} + \left(2 - \frac{7}{2}\right)^2 \times \frac{1}{6} + \left(3 - \frac{7}{2}\right)^2 \times \frac{1}{6} \\ &\quad + \left(4 - \frac{7}{2}\right)^2 \times \frac{1}{6} + \left(5 - \frac{7}{2}\right)^2 \times \frac{1}{6} + \left(6 - \frac{7}{2}\right)^2 \times \frac{1}{6} \\ &= \frac{35}{12} \end{aligned}$$

となります。

例題 11.6 分散の性質を確かめる

したがって、右辺は、

$$\begin{aligned} 2^2 V(X) &= 4 \times \frac{35}{12} \\ &= \frac{35}{3} \end{aligned}$$

となります。

例題 11.6 分散の性質を確かめる

左辺は,

$$V(2X + 1) = \frac{35}{3}$$

右辺も,

$$2^2 V(X) = \frac{35}{3}$$

となりました. したがって,

$$V(2X + 1) = 2^2 V(X)$$

が確かに成り立っています.

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

例題 12.1 二項分布の確率分布表を求める

ある作業を 1 回行ったとき，成功する確率を

$$\frac{1}{3}$$

とします．

この作業を 5 回くり返し，各回の成功・失敗に関する事象は互いに独立であるとします．

成功した回数を表す確率変数を X とするとき， X の確率分布表を作りましょう．

例題 12.1 二項分布の確率分布表を求める

1 回の作業で成功する確率は,

$$\frac{1}{3}$$

なので, 失敗する確率は,

$$1 - \frac{1}{3} = \frac{2}{3}$$

です.

5 回の結果について, 成功した回数ごとに確率を調べます.

例題 12.1 二項分布の確率分布表を求める

ここで「並び」とは、5 回の結果を、

1 回目, 2 回目, 3 回目, 4 回目, 5 回目

の順に並べたものです.

たとえば,

成功・失敗・失敗・成功・失敗

のような並びです.

例題 12.1 二項分布の確率分布表を求める

二項分布

同じ操作を n 回行い、各回で「成功」か「失敗」かが決まるとします。

各回の成功確率が一定で p であり、各回の成功・失敗に関する事象が互いに独立であるとき、

成功した回数を表す確率変数 X の確率分布を二項分布といいます。

例題 12.1 二項分布の確率分布表を求める

二項分布は、

$$B(n, p)$$

と表します。ここで、

$$n = \text{試行回数}, \quad p = \text{成功確率}$$

です。

確率変数 X の確率分布が $B(n, p)$ であるとき、

X は二項分布 $B(n, p)$ に従う

といいます。

例題 12.1 二項分布の確率分布表を求める

今回の作業では、

$$n = 5, \quad p = \frac{1}{3}$$

です。したがって、成功した回数を表す確率変数 X は、

$$B \left(5, \frac{1}{3} \right) \text{ に従います。}$$

例題 12.1 二項分布の確率分布表を求める

X が二項分布 $B(n, p)$ に従うとき，成功回数がちょうど k 回となる確率は，

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

です（例題 10.6 参照）．

例題 12.1 二項分布の確率分布表を求める

$X = 0$ の場合

5 回すべて失敗する 1 つの並びの確率は、

$$\left(\frac{2}{3}\right)^5 = \frac{32}{243}$$

です。このような並びは、

$$\binom{5}{0} = 1$$

通りあります。

例題 12.1 二項分布の確率分布表を求める

したがって、 $X = 0$ となる確率は、

$$\begin{aligned} P(X = 0) &= \binom{5}{0} \times \left(\frac{1}{3}\right)^0 \times \left(\frac{2}{3}\right)^5 \\ &= 1 \times \frac{32}{243} \\ &= \frac{32}{243} \end{aligned}$$

となります。

例題 12.1 二項分布の確率分布表を求める

$X = 1$ の場合

たとえば，1 回目だけ成功する 1 つの並びの確率は，

$$\frac{1}{3} \times \left(\frac{2}{3}\right)^4 = \frac{16}{243}$$

です．成功する 1 回の選び方は，

$$\binom{5}{1} = 5$$

通りあります．

例題 12.1 二項分布の確率分布表を求める

したがって、 $X = 1$ となる確率は、

$$\begin{aligned} P(X = 1) &= \binom{5}{1} \times \frac{1}{3} \times \left(\frac{2}{3}\right)^4 \\ &= 5 \times \frac{16}{243} \\ &= \frac{80}{243} \end{aligned}$$

となります。

例題 12.1 二項分布の確率分布表を求める

$X = 2$ の場合

たとえば，1 回目と 2 回目だけ成功する 1 つの並びの確率は，

$$\left(\frac{1}{3}\right)^2 \times \left(\frac{2}{3}\right)^3 = \frac{8}{243}$$

です．成功する 2 回の選び方は，

$$\binom{5}{2} = \frac{5 \times 4}{2 \times 1} = 10$$

通りあります．

例題 12.1 二項分布の確率分布表を求める

したがって、 $X = 2$ となる確率は、

$$\begin{aligned} P(X = 2) &= \binom{5}{2} \times \left(\frac{1}{3}\right)^2 \times \left(\frac{2}{3}\right)^3 \\ &= 10 \times \frac{8}{243} \\ &= \frac{80}{243} \end{aligned}$$

となります。

例題 12.1 二項分布の確率分布表を求める

$X = 3$ の場合

たとえば，1～3 回目が成功する 1 つの並びの確率は，

$$\left(\frac{1}{3}\right)^3 \times \left(\frac{2}{3}\right)^2 = \frac{4}{243}$$

です．成功する 3 回の選び方は，

$$\binom{5}{3} = \frac{5 \times 4 \times 3}{3 \times 2 \times 1} = 10$$

通りあります．

例題 12.1 二項分布の確率分布表を求める

したがって、 $X = 3$ となる確率は、

$$\begin{aligned} P(X = 3) &= \binom{5}{3} \times \left(\frac{1}{3}\right)^3 \times \left(\frac{2}{3}\right)^2 \\ &= 10 \times \frac{4}{243} \\ &= \frac{40}{243} \end{aligned}$$

となります。

例題 12.1 二項分布の確率分布表を求める

$X = 4$ の場合

たとえば，1～4 回目が成功する 1 つの並びの確率は，

$$\left(\frac{1}{3}\right)^4 \times \frac{2}{3} = \frac{2}{243}$$

です．成功する 4 回の選び方は，

$$\binom{5}{4} = \frac{5 \times 4 \times 3 \times 2}{4 \times 3 \times 2 \times 1} = 5$$

通りあります．

例題 12.1 二項分布の確率分布表を求める

したがって、 $X = 4$ となる確率は、

$$\begin{aligned} P(X = 4) &= \binom{5}{4} \times \left(\frac{1}{3}\right)^4 \times \frac{2}{3} \\ &= 5 \times \frac{2}{243} \\ &= \frac{10}{243} \end{aligned}$$

となります。

例題 12.1 二項分布の確率分布表を求める

$X = 5$ の場合

5 回すべて成功する 1 つの並びの確率は、

$$\left(\frac{1}{3}\right)^5 = \frac{1}{243}$$

です。このような並びは、

$$\binom{5}{5} = 1$$

通りあります。

例題 12.1 二項分布の確率分布表を求める

したがって、 $X = 5$ となる確率は、

$$\begin{aligned} P(X = 5) &= \binom{5}{5} \times \left(\frac{1}{3}\right)^5 \times \left(\frac{2}{3}\right)^0 \\ &= 1 \times \frac{1}{243} \\ &= \frac{1}{243} \end{aligned}$$

となります。

例題 12.1 二項分布の確率分布表を求める

したがって、確率分布表は、

x	0	1	2	3	4	5
$P(X = x)$	$\frac{32}{243}$	$\frac{80}{243}$	$\frac{80}{243}$	$\frac{40}{243}$	$\frac{10}{243}$	$\frac{1}{243}$

となります。

例題 12.1 二項分布の確率分布表を求める

ここまで、 $X = 0, 1, \dots, 5$ の場合について、それぞれ確率を求めてきました。

ここで、これまでの計算をまとめて考えます。

k を $0, 1, 2, 3, 4, 5$ のいずれかとするとき、成功する回数が k 回である並びは、

$$\binom{5}{k}$$

通りあります。

例題 12.1 二項分布の確率分布表を求める

1つの並びが起こる確率は,

$$\left(\frac{1}{3}\right)^k \times \left(\frac{2}{3}\right)^{5-k}$$

の形になります.

したがって, $X = k$ である確率は,

$$P(X = k) = \binom{5}{k} \times \left(\frac{1}{3}\right)^k \times \left(\frac{2}{3}\right)^{5-k}$$

と表されます.

例題 12.2 二項分布の期待値を求める

ある作業を 1 回行ったとき，成功する確率を

$$\frac{1}{3}$$

とします．

この作業を 5 回くり返し，各回の成功・失敗に関する事象は互いに独立であるとします．

成功した回数を表す確率変数を X とします．

例題 12.2 二項分布の期待値を求める

例題 12.1 では、 X の確率分布表をつくりました。

ここでは、 X の期待値を求めます。

(1) 確率分布表を用いて、 X の期待値 $E(X)$ を定義に従って求めましょう。

(2) 二項分布の期待値を求める公式

$$E(X) = np$$

を用いて、 X の期待値をもう一度求めましょう。

例題 12.2 二項分布の期待値を求める

(1) 確率分布表から期待値を求める

例題 12.1 で求めた確率分布表は、

x	0	1	2	3	4	5
$P(X = x)$	$\frac{32}{243}$	$\frac{80}{243}$	$\frac{80}{243}$	$\frac{40}{243}$	$\frac{10}{243}$	$\frac{1}{243}$

です。

例題 12.2 二項分布の期待値を求める

期待値は、

「値」×「その値が出る確率」をそれぞれ計算し、
すべて足し合わせて求めます。

したがって、 X の期待値は、

$$\begin{aligned} E(X) &= 0 \times \frac{32}{243} + 1 \times \frac{80}{243} + 2 \times \frac{80}{243} + 3 \times \frac{40}{243} + 4 \times \frac{10}{243} + 5 \times \frac{1}{243} \\ &= \frac{5}{3} \end{aligned}$$

となります。

例題 12.2 二項分布の期待値を求める

二項分布の期待値

確率変数 X が二項分布

$$B(n, p)$$

に従うとき、 X の期待値は

$$E(X) = np$$

と求めることができます。

例題 12.2 二項分布の期待値を求める

(2) 公式を使って期待値を求める

今回の X は,

$$B\left(5, \frac{1}{3}\right)$$

に従います。したがって,

$$n = 5, \quad p = \frac{1}{3}$$

です。

例題 12.2 二項分布の期待値を求める

公式

$$E(X) = np$$

に代入すると,

$$E(X) = 5 \times \frac{1}{3} = \frac{5}{3}$$

となります.

例題 12.2 二項分布の期待値を求める

(1) では，期待値の定義から，

$$E(X) = \frac{5}{3}$$

となりました．(2) でも，

$$E(X) = np = \frac{5}{3}$$

となりました．

2 つの方法で求めた結果は一致しています．

例題 12.3 二項分布の分散を求める

ある作業を 1 回行ったとき，成功する確率を

$$\frac{1}{3}$$

とします．

この作業を 5 回くり返し，各回の成功・失敗に関する事象は互いに独立であるとします．

成功した回数を表す確率変数を X とします．

例題 12.3 二項分布の分散を求める

例題 12.1 では、 X の確率分布表をつくりました。

ここでは、 X の分散を求めます。

(1) 確率分布表を用いて、 X の分散 $V(X)$ を定義に従って求めましょう。

(2) 二項分布の分散を求める公式

$$V(X) = np(1 - p)$$

を用いて、 X の分散をもう一度求めましょう。

例題 12.3 二項分布の分散を求める

(1) 分散の定義から分散を求める

例題 12.1 で求めた確率分布表は、

x	0	1	2	3	4	5
$P(X = x)$	$\frac{32}{243}$	$\frac{80}{243}$	$\frac{80}{243}$	$\frac{40}{243}$	$\frac{10}{243}$	$\frac{1}{243}$

です．また，例題 12.2 より， X の期待値は次であることがわかっています．

$$E(X) = \frac{5}{3}$$

例題 12.3 二項分布の分散を求める

分散は、「値と期待値の差の 2 乗」×「その値が出る確率」をそれぞれ計算し、すべて足し合わせて求めます。

したがって、 X の分散は次になります。

$$\begin{aligned} V(X) &= \left(0 - \frac{5}{3}\right)^2 \times \frac{32}{243} + \left(1 - \frac{5}{3}\right)^2 \times \frac{80}{243} + \left(2 - \frac{5}{3}\right)^2 \times \frac{80}{243} \\ &\quad + \left(3 - \frac{5}{3}\right)^2 \times \frac{40}{243} + \left(4 - \frac{5}{3}\right)^2 \times \frac{10}{243} + \left(5 - \frac{5}{3}\right)^2 \times \frac{1}{243} \\ &= \frac{10}{9} \end{aligned}$$

例題 12.3 二項分布の分散を求める

二項分布と分散

確率変数 X が二項分布

$$B(n, p)$$

に従うとき、 X の分散は

$$V(X) = np(1 - p)$$

と求めることができます。

例題 12.3 二項分布の分散を求める

(2) 公式を使って分散を求める

今回の X は,

$$B\left(5, \frac{1}{3}\right)$$

に従います。したがって,

$$n = 5, \quad p = \frac{1}{3}$$

です.

例題 12.3 二項分布の分散を求める

公式

$$V(X) = np(1 - p)$$

に代入すると,

$$V(X) = 5 \times \frac{1}{3} \times \left(1 - \frac{1}{3}\right) = \frac{10}{9}$$

となります.

例題 12.3 二項分布の分散を求める

(1) では、分散の定義から、

$$V(X) = \frac{10}{9}$$

となりました。(2) でも、

$$V(X) = np(1 - p) = \frac{10}{9}$$

となりました。

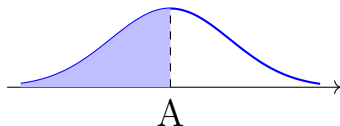
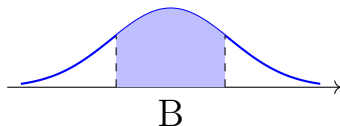
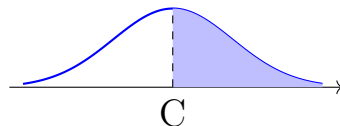
2 つの方法で求めた結果は一致しています。

第12章 二項分布から正規分布へ

12.2 正規分布の性質

例題 12.4 連続型確率分布の確率の表し方を理解する

(1) 連続型確率変数 X の確率分布を表す 3 つの図について、塗られた部分に対応する確率を表す式を正しく割り当てたものを選びましょう。



(a) $A : P(X < b)$, $B : P(a < X < b)$, $C : P(X > a)$

(b) $A : P(X > a)$, $B : P(a < X < b)$, $C : P(X < b)$

(c) $A : P(a < X < b)$, $B : P(X < b)$, $C : P(X > a)$

例題 12.4 連続型確率分布の確率の表し方を理解する

(2) 連続型確率変数 X について,

$$P(X > a)$$

は、次のどれと等しくなるでしょうか．

(a) $1 - P(a < X < b)$

(b) $P(X < a)$

(c) $1 - P(X \leq a)$

例題 12.4 連続型確率分布の確率の表し方を理解する

(3) ある測定の誤差を表す確率変数 X (cm) は,

−2 cm から 2 cm

のあいだで生じるものとします．この範囲では，区間の幅が同じなら，その区間に入る確率も同じとします．

次の確率を求めましょう．

- (a) $P(X > 1)$
- (b) $P(-1 < X < 0.5)$
- (c) $P(X \leq -1.5)$

例題 12.4 連続型確率分布の確率の表し方を理解する

連続型確率変数と確率の表し方

確率変数のとりうる値が連続的に変化し、数え上げることができない場合、このような確率変数を**連続型確率変数**といいます。

連続型確率変数では、ある範囲に入る確率をグラフの下の面積として表します。

例題 12.4 連続型確率分布の確率の表し方を理解する

離散型では、

$$P(X = k)$$

のように、それぞれの値に確率を割り当てます。

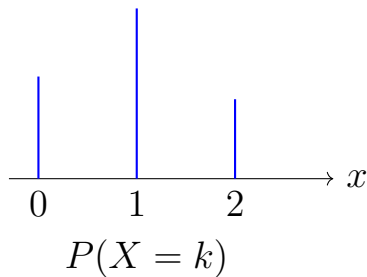
一方、連続型では、

$$P(X < a), \quad P(X > a), \quad P(a < X < b)$$

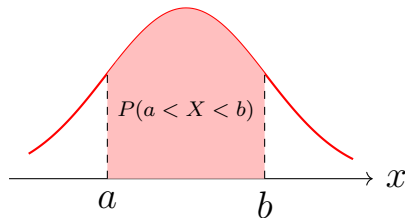
のように、ある範囲に入る確率を考えます。

例題 12.4 連続型確率分布の確率の表し方を理解する

離散型



連続型



連続型では、**曲線下の面積**がその区間に入る確率を表します。

例題 12.4 連続型確率分布の確率の表し方を理解する

連続型確率分布とは、連続型確率変数について、各区間に入る確率がどのように定まるかを表したものです。

したがって、図の塗られた部分を見れば、

左側, 2 つの値のあいだ, 右側

のどの範囲の確率を表しているかを読み取ることができます。

例題 12.4 連続型確率分布の確率の表し方を理解する

(1) 図に対応する確率の式を選ぶ

図 A は、ある値より左側が塗られているので、 $P(X < b)$

図 B は、2 つの値のあいだが塗られているので、 $P(a < X < b)$

図 C は、ある値より右側が塗られているので、 $P(X > a)$

と表すことができます。したがって、

正解は (a) です。

例題 12.4 連続型確率分布の確率の表し方を理解する

連続型では、確率はグラフの下での面積で表されます。

全事象が起こる確率は 1 なので、

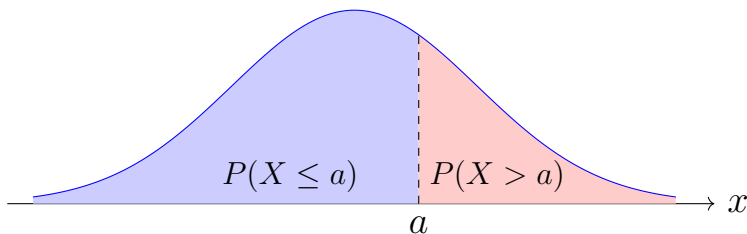
$$\text{グラフ全体の下での面積} = 1$$

となります。また、1 点だけでは面積が 0 なので、

$$P(X = a) = 0$$

です。

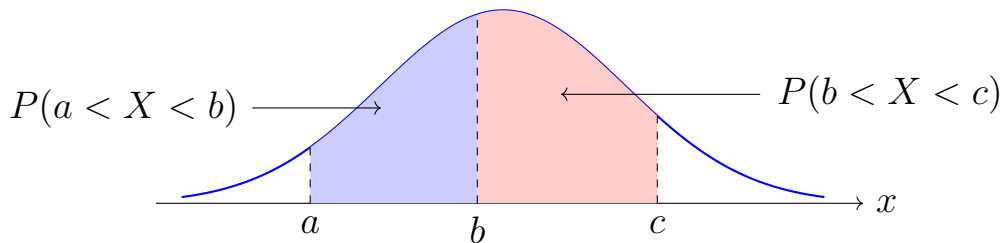
例題 12.4 連続型確率分布の確率の表し方を理解する



2 つの部分を含わせると，グラフ全体の面積になります．

例題 12.4 連続型確率分布の確率の表し方を理解する

また，連続型では，区間を分けると，確率も面積の和として考えることができます．



例題 12.4 連続型確率分布の確率の表し方を理解する

$a < b < c$ のとき,

$$P(a < X < c) = P(a < X < b) + P(b < X < c)$$

となります。また,

$$P(a < X < b)$$

は, $P(X < b)$ から $P(X < a)$ を取り除いた面積なので,

$$P(a < X < b) = P(X < b) - P(X < a)$$

となります。

例題 12.4 連続型確率分布の確率の表し方を理解する

(2) $P(X > a)$ と等しい確率を考える

$X \leq a$ の場合と $X > a$ の場合を合わせると、全体になります。

したがって、

$$P(X \leq a) + P(X > a) = 1$$

なので、

$$P(X > a) = 1 - P(X \leq a)$$

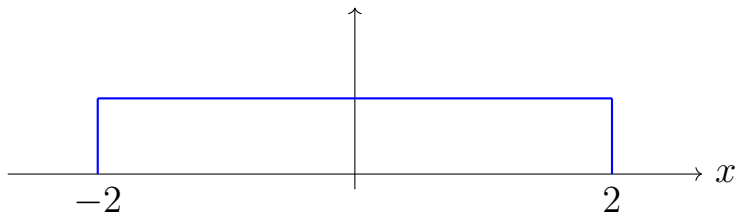
となります。

正解：(c) $1 - P(X \leq a)$

例題 12.4 連続型確率分布の確率の表し方を理解する

(3) 分布の形を考える

X は -2 から 2 までの値をとり、区間の幅が同じなら確率も同じなので、分布の高さは一定になります。



全体の面積が 1 であることを使います。

例題 12.4 連続型確率分布の確率の表し方を理解する

−2 から 2 までの区間の長さは,

$$2 - (-2) = 4$$

です. 全体の面積は 1 なので,

$$4 \times \text{高さ} = 1$$

となります. したがって,

$$\text{高さ} = \frac{1}{4}$$

です.

例題 12.4 連続型確率分布の確率の表し方を理解する

(3)(a) $P(X > 1)$

1 から 2 までの区間の長さは,

$$2 - 1 = 1$$

です。したがって,

$$P(X > 1) = \frac{1}{4} \times 1 = \frac{1}{4}$$

となります。

例題 12.4 連続型確率分布の確率の表し方を理解する

(3)(b) $P(-1 < X < 0.5)$

区間の長さは,

$$0.5 - (-1) = 1.5$$

です. したがって,

$$P(-1 < X < 0.5) = \frac{1}{4} \times 1.5 = \frac{3}{8}$$

となります.

例題 12.4 連続型確率分布の確率の表し方を理解する

(3)(c) $P(X \leq -1.5)$

−2 から −1.5 までの区間の長さは、

$$-1.5 - (-2) = 0.5$$

です。したがって、

$$P(X \leq -1.5) = \frac{1}{4} \times 0.5 = \frac{1}{8}$$

となります。

例題 12.5 正規分布の形と標準化を理解する

ある学校の生徒の身長を表す確率変数 X は、期待値 165 cm、標準偏差 5 cm の正規分布に従うことがわかりました。

- (1) 正規分布の形として正しいものを選びましょう。
 - (a) 左にだけ大きく伸びた片側だけの形である。
 - (b) 右上がりの直線の形である。
 - (c) 左右対称で、中央がいちばん高い山の形である。

例題 12.5 正規分布の形と標準化を理解する

(2) 期待値 165 cm は、分布のどの部分を表しているでしょうか。

- (a) 山のいちばん高い位置
- (b) 分布の右端の位置
- (c) 分布の広がり（ばらつき）の大きさ

(3) 標準偏差 5 cm は、何を表しているでしょうか。

- (a) 分布の最大値
- (b) 分布の中心の位置
- (c) 分布の広がり（ばらつき）の大きさ

例題 12.5 正規分布の形と標準化を理解する

(4) 次の身長について、それぞれ z 値を求めましょう.

(a) 165 cm

(b) 170 cm

(c) 160 cm

(d) 175 cm

(e) 155 cm

例題 12.5 正規分布の形と標準化を理解する

正規分布

正規分布は，左右対称で中央がいちばん高い山の形をした連続型確率分布です．

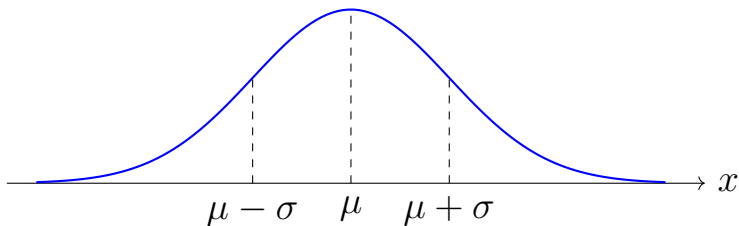
期待値 μ は山の中心の位置を表し，標準偏差 σ は，期待値 μ のまわりにどの程度ばらついているかを表します．標準偏差が大きいほど，山は横に広がります．

例題 12.5 正規分布の形と標準化を理解する

正規分布は、期待値 μ ，分散 σ^2 を使って、

$$N(\mu, \sigma^2)$$

と表します。



期待値 μ が山の中心となり，その左右に同じ形で広がっています。

例題 12.5 正規分布の形と標準化を理解する

(1) 正規分布の形

正規分布は、左右対称で、中央がいちばん高い山の形をしています。
したがって、正解は、

(c) 左右対称で、中央がいちばん高い山の形である。

例題 12.5 正規分布の形と標準化を理解する

(2) 期待値が表すもの

正規分布では、期待値 μ が分布の中心にあり、その位置で曲線が最も高くなります。

したがって、正解は、

(a) 山のいちばん高い位置です。

例題 12.5 正規分布の形と標準化を理解する

(3) 標準偏差が表すもの

標準偏差は、期待値のまわりにどの程度ばらついているかを表します。

したがって、正解は、

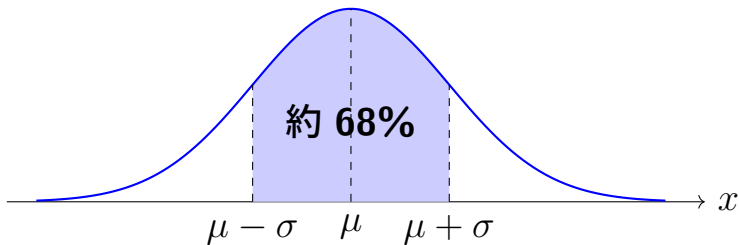
(c) 分布の広がり（ばらつき）の大きさです。

例題 12.5 正規分布の形と標準化を理解する

正規分布では，期待値 μ を中心として，

$$\mu - \sigma \quad \text{から} \quad \mu + \sigma$$

までの範囲に，全体のおよそ 68% が含まれます．



例題 12.5 正規分布の形と標準化を理解する

正規分布は、期待値や標準偏差が変わると、山の位置や広がりも変わります。

そこで、異なる正規分布を共通の尺度で扱えるように、確率変数を変換して、

期待値が 0，標準偏差が 1

になるように尺度をそろえることを考えます。

例題 12.5 正規分布の形と標準化を理解する

標準正規分布

どの正規分布でも共通の尺度で扱えるように，確率変数を変換して，期待値が 0 ，標準偏差が 1 になるように尺度をそろえる操作を標準化といいます．

標準化によって得られる期待値 0 ，標準偏差 1 の正規分布を標準正規分布といいます．

例題 12.5 正規分布の形と標準化を理解する

確率変数 X が正規分布

$$N(\mu, \sigma^2)$$

に従うとき、 X のとる値 x を

$$z = \frac{x - \mu}{\sigma}$$

と変換します。

このようにして得られる値を z 値といいます。

例題 12.5 正規分布の形と標準化を理解する

z 値は,

その値が期待値よりどれだけ上か下かを,
標準偏差を単位として表したもの

です. z 値の符号を見ると,

$z > 0 \Rightarrow$ 期待値より上

$z = 0 \Rightarrow$ 期待値と同じ

$z < 0 \Rightarrow$ 期待値より下

となります.

例題 12.5 正規分布の形と標準化を理解する

また、 z 値の絶対値は、

期待値から標準偏差何個分だけ離れているか

を表します。たとえば、

$$z = 1$$

なら、期待値より標準偏差 1 個分上、

$$z = -2$$

なら、期待値より標準偏差 2 個分下です。

例題 12.5 正規分布の形と標準化を理解する

確率変数 X そのものを標準化すると,

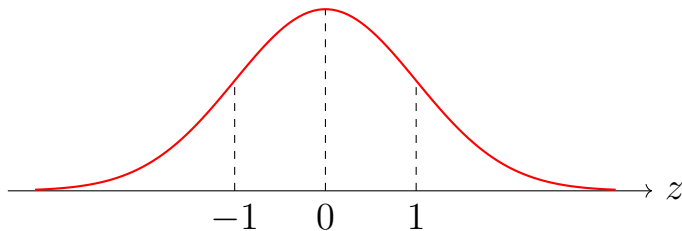
$$Z = \frac{X - \mu}{\sigma}$$

となります. この Z は,

期待値 0, 標準偏差 1

の正規分布, つまり, 標準正規分布 $N(0, 1)$ に従います.

例題 12.5 正規分布の形と標準化を理解する



標準正規分布 $N(0, 1)$

例題 12.5 正規分布の形と標準化を理解する

(4)(a) 身長 165 cm の z 値を求める

$$z = \frac{x - \mu}{\sigma}$$

において，この例では，

$$\mu = 165, \quad \sigma = 5$$

となります．よって，身長 165 cm の z 値は，

$$z = \frac{165 - 165}{5} = 0$$

です．

例題 12.5 正規分布の形と標準化を理解する

(4)(b) 身長 170 cm の z 値を求める

身長 170 cm の z 値は,

$$z = \frac{170 - 165}{5} = 1$$

です.

例題 12.5 正規分布の形と標準化を理解する

(4)(c) 身長 160 cm の z 値を求める

身長 160 cm の z 値は,

$$z = \frac{160 - 165}{5} = -1$$

です.

例題 12.5 正規分布の形と標準化を理解する

(4)(d) 身長 175 cm の z 値を求める

身長 175 cm の z 値は,

$$z = \frac{175 - 165}{5} = 2$$

です.

例題 12.5 正規分布の形と標準化を理解する

(4)(e) 身長 155 cm の z 値を求める

身長 155 cm の z 値は,

$$z = \frac{155 - 165}{5} = -2$$

です.

例題 12.5 正規分布の形と標準化を理解する

以上をまとめると、

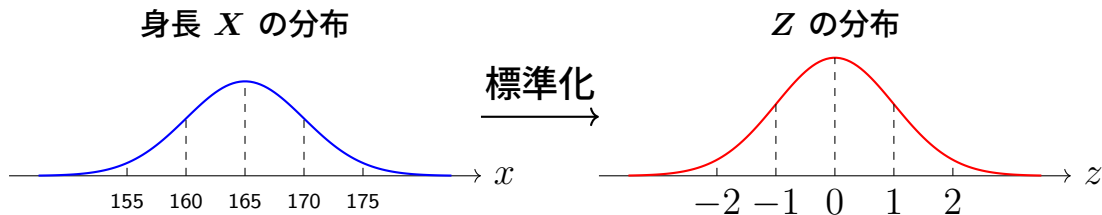
x	155	160	165	170	175
z	-2	-1	0	1	2

となります。たとえば、

$$170 \text{ cm} \longleftrightarrow z = 1$$

と対応します。

例題 12.5 正規分布の形と標準化を理解する



横軸を「期待値から標準偏差何個分か」という尺度に変換

第13章 正規分布

13.1 正規分布の確率計算

例題 13.1 標準正規分布の確率を求める

確率変数 Z は標準正規分布に従うとします．標準正規分布表（巻末）を使って，次の確率を求めましょう．

(1) $P(Z > 1.00)$

(2) $P(Z \leq 1.57)$

(3) $P(Z < -1.00)$

(4) $P(Z > -0.33)$

(5) $P(-0.50 \leq Z \leq 1.70)$

例題 13.1 標準正規分布の確率を求める

標準正規分布表

標準正規分布表（上側確率）には，標準正規分布における確率変数 Z の値がある値 u より大きくなる確率，

$$\text{上側確率 } P(Z > u)$$

が載っています．

行と列で u の値を指定すると，対応する上側確率 $P(Z > u)$ を読み取ることができます．

例題 13.1 標準正規分布の確率を求める

連続型確率分布の不等号

連続型確率分布では、ある 1 点をとる確率は 0 です。

$$P(Z = u) = 0$$

したがって、

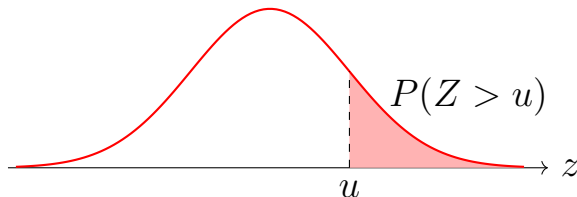
$$P(Z < u) = P(Z \leq u), \quad P(Z > u) = P(Z \geq u)$$

となり、不等号に「=」が付いていても付いていなくても確率は同じです。

例題 13.1 標準正規分布の確率を求める

正の上側確率 $P(Z > u)$

まず、 $u > 0$ のときの $P(Z > u)$ を図で確認します。



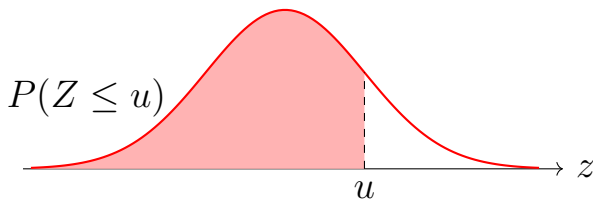
標準正規分布表から、この右側の面積を直接読み取ります。

例題 13.1 標準正規分布の確率を求める

正の下側確率 $P(Z \leq u)$

$u > 0$ のとき、 $P(Z \leq u)$ は $P(Z > u)$ の補集合なので、

$$P(Z \leq u) = 1 - P(Z > u)$$

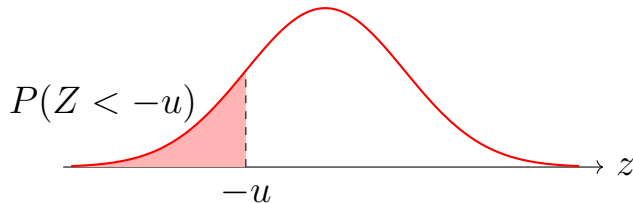


例題 13.1 標準正規分布の確率を求める

負の下側確率 $P(Z < -u)$

標準正規分布は原点を中心に左右対称なので、

$$P(Z < -u) = P(Z > u) \quad (u > 0)$$

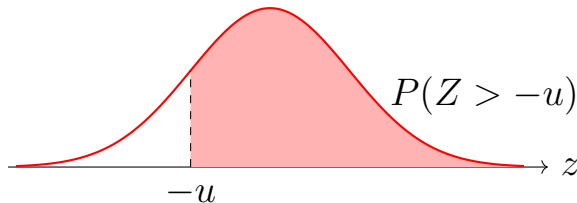


例題 13.1 標準正規分布の確率を求める

負の上側確率 $P(Z > -u)$

左右対称性と補集合の関係より、

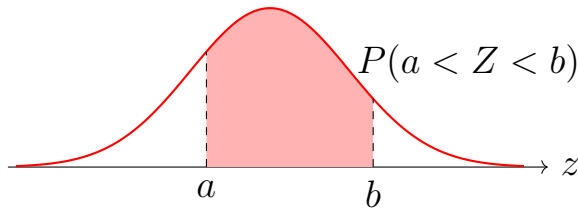
$$P(Z > -u) = P(Z < u) = 1 - P(Z > u) \quad (u > 0)$$



例題 13.1 標準正規分布の確率を求める

区間確率 $P(a < Z < b)$

区間確率 $P(a < Z < b)$ は、面積の差を使って求めます。



$$P(a < Z < b) = P(Z > a) - P(Z > b)$$

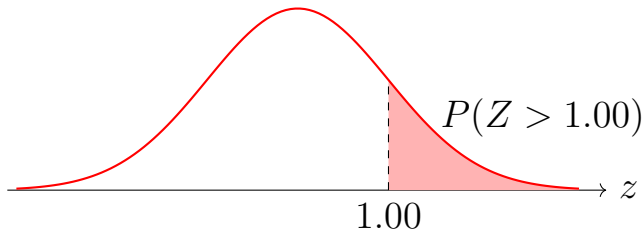
例題 13.1 標準正規分布の確率を求める

(1) $P(Z > 1.00)$ を求める

標準正規分布表の行「1.0」，列「.00」から，

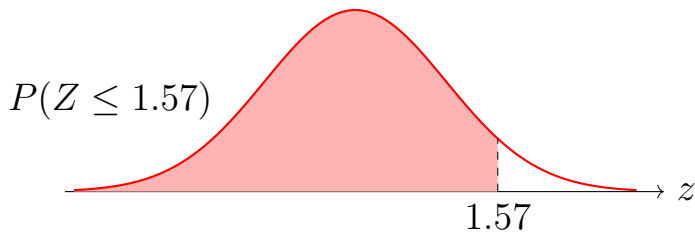
$$P(Z > 1.00) = 0.1587$$

と読み取れます。



例題 13.1 標準正規分布の確率を求める

(2) $P(Z \leq 1.57)$ を求める



例題 13.1 標準正規分布の確率を求める

標準正規分布表から，

$$P(Z > 1.57) = 0.0582$$

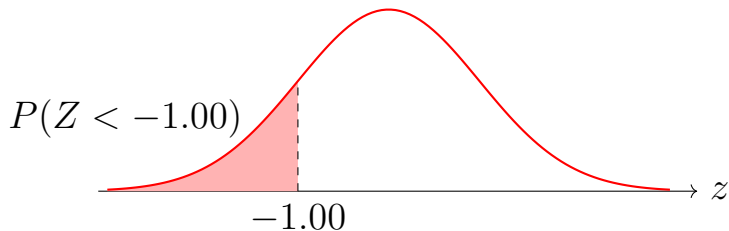
と読み取れます．したがって，

$$\begin{aligned} P(Z \leq 1.57) &= 1 - P(Z > 1.57) \\ &= 1 - 0.0582 \\ &= 0.9418 \end{aligned}$$

となります．

例題 13.1 標準正規分布の確率を求める

(3) $P(Z < -1.00)$ を求める



例題 13.1 標準正規分布の確率を求める

標準正規分布は左右対称なので，

$$P(Z < -1.00) = P(Z > 1.00)$$

となります．標準正規分布表より，

$$P(Z > 1.00) = 0.1587$$

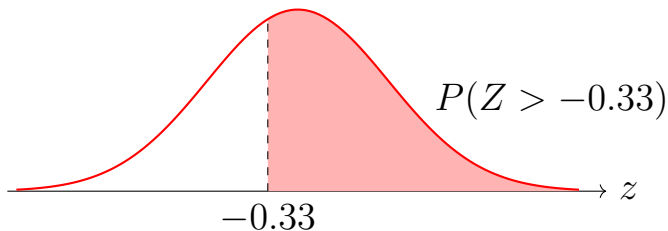
なので，

$$P(Z < -1.00) = 0.1587$$

となります．

例題 13.1 標準正規分布の確率を求める

(4) $P(Z > -0.33)$ を求める



例題 13.1 標準正規分布の確率を求める

標準正規分布は左右対称なので、

$$P(Z > -0.33) = P(Z < 0.33)$$

となります。標準正規分布表より、

$$P(Z > 0.33) = 0.3707$$

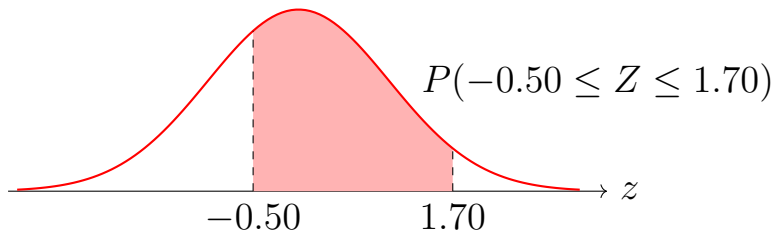
なので、

$$\begin{aligned} P(Z > -0.33) &= P(Z < 0.33) \\ &= 1 - P(Z > 0.33) \\ &= 1 - 0.3707 = 0.6293 \end{aligned}$$

となります。

例題 13.1 標準正規分布の確率を求める

(5) $P(-0.50 \leq Z \leq 1.70)$ を求める



例題 13.1 標準正規分布の確率を求める

まず,

$$P(Z > -0.50) = 1 - P(Z < -0.50)$$

となります。左右対称性より,

$$P(Z < -0.50) = P(Z > 0.50)$$

であり, 標準正規分布表から次がわかります。

$$P(Z > 0.50) = 0.3085$$

例題 13.1 標準正規分布の確率を求める

したがって、

$$\begin{aligned} P(Z > -0.50) &= 1 - P(Z < -0.50) = 1 - P(Z > 0.50) \\ &= 1 - 0.3085 = 0.6915 \end{aligned}$$

となります。また、標準正規分布表より、

$$P(Z > 1.70) = 0.0446$$

です。

例題 13.1 標準正規分布の確率を求める

したがって、

$$\begin{aligned} P(-0.50 \leq Z \leq 1.70) &= P(Z > -0.50) - P(Z > 1.70) \\ &= 0.6915 - 0.0446 \\ &= 0.6469 \end{aligned}$$

となります。

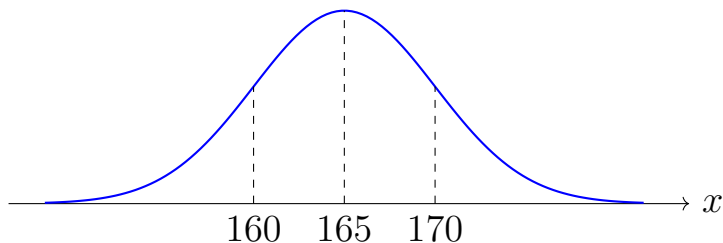
例題 13.2 正規分布の確率を求める

ある学校の生徒の身長を表す確率変数 X は、期待値 165 cm、標準偏差 5 cm の正規分布に従うことがわかりました。

無作為に 1 人選んだこの学校の生徒について、標準正規分布表（巻末）を使って次を求めましょう。

- (1) 身長が 168 cm 以上である確率
- (2) 身長が 157 cm 以下である確率
- (3) 身長が 162 cm 以上である確率
- (4) 身長が 160 cm 以上 170 cm 以下である確率

例題 13.2 正規分布の確率を求める



この学校の身長分布（イメージ）

例題 13.2 正規分布の確率を求める

標準化して確率を求める

確率変数 X が正規分布

$$N(\mu, \sigma^2)$$

に従うとき、

$$Z = \frac{X - \mu}{\sigma}$$

と変換して標準化します。

例題 13.2 正規分布の確率を求める

標準化すると、横軸が

期待値 0，標準偏差 1

の共通の尺度にそろえられます。

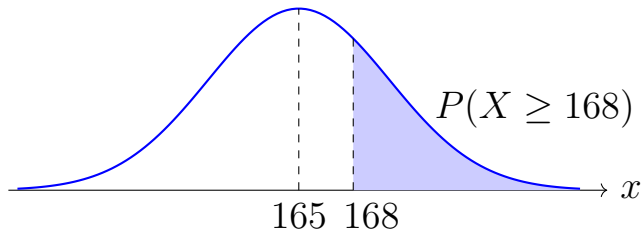
横軸の尺度は変わりますが、

対応する範囲の曲線下の面積（＝確率）は変わりません。

したがって、標準化したあとは標準正規分布表を使って確率を求めることができます。

例題 13.2 正規分布の確率を求める

(1) 身長が 168 cm 以上である確率を求める



例題 13.2 正規分布の確率を求める

168 cm を標準化すると,

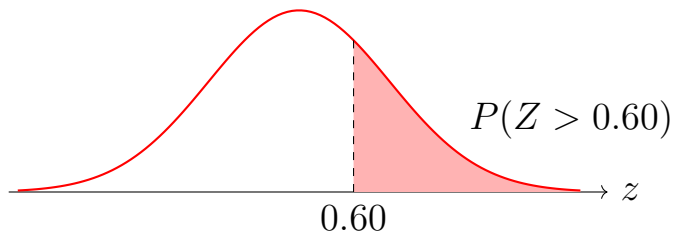
$$\begin{aligned} P(X \geq 168) &= P\left(Z \geq \frac{168 - 165}{5}\right) \\ &= P(Z \geq 0.60) \end{aligned}$$

となります. つまり, 身長 168 cm は, 標準化すると

$$168 \text{ cm} \quad \longleftrightarrow \quad z = 0.60$$

に対応します.

例題 13.2 正規分布の確率を求める



標準正規分布表より,

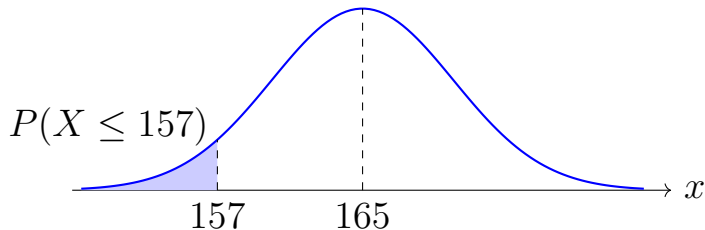
$$P(Z > 0.60) = 0.2743$$

です。したがって,

$$P(X \geq 168) = 0.2743$$

例題 13.2 正規分布の確率を求める

(2) 身長が 157 cm 以下である確率を求める



例題 13.2 正規分布の確率を求める

157 cm を標準化すると,

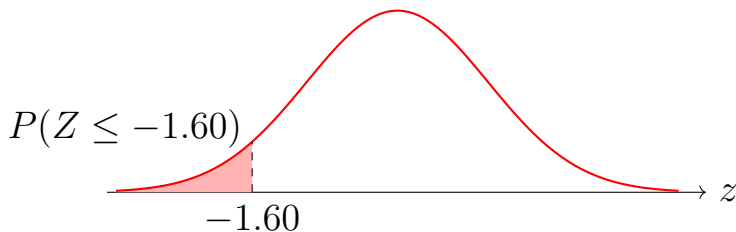
$$\begin{aligned} P(X \leq 157) &= P\left(Z \leq \frac{157 - 165}{5}\right) \\ &= P(Z \leq -1.60) \end{aligned}$$

となります。左右対称性より,

$$P(Z \leq -1.60) = P(Z \geq 1.60)$$

です。

例題 13.2 正規分布の確率を求める



標準正規分布表より、

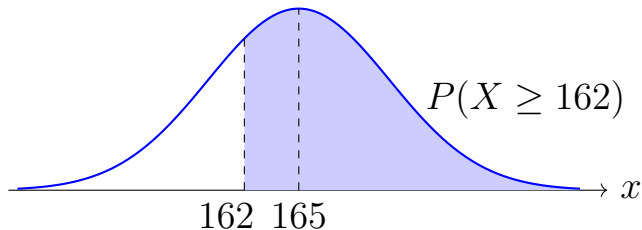
$$P(Z > 1.60) = 0.0548$$

です。したがって、

$$P(X \leq 157) = 0.0548$$

例題 13.2 正規分布の確率を求める

(3) 身長が 162 cm 以上である確率を求める



例題 13.2 正規分布の確率を求める

162 cm を標準化すると,

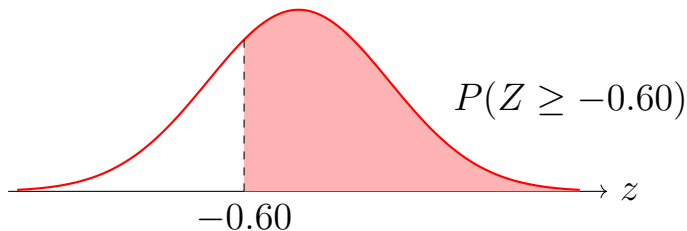
$$\begin{aligned} P(X \geq 162) &= P\left(Z \geq \frac{162 - 165}{5}\right) \\ &= P(Z \geq -0.60) \end{aligned}$$

となります。左右対称性より,

$$P(Z \geq -0.60) = P(Z \leq 0.60)$$

です。

例題 13.2 正規分布の確率を求める



標準正規分布表より、

$$P(Z > 0.60) = 0.2743$$

です。

例題 13.2 正規分布の確率を求める

したがって,

$$\begin{aligned}P(Z \geq -0.60) &= P(Z \leq 0.60) \\&= 1 - P(Z > 0.60) \\&= 1 - 0.2743 \\&= 0.7257\end{aligned}$$

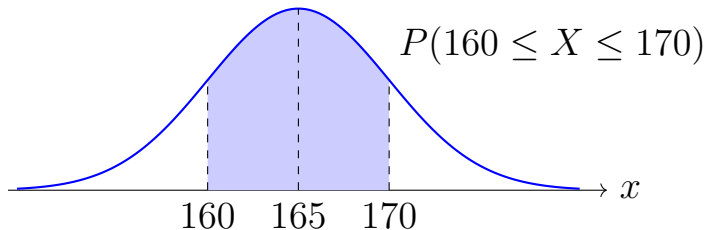
となります。よって、求める確率は、

$$P(X \geq 162) = 0.7257$$

です。

例題 13.2 正規分布の確率を求める

(4) 身長が 160 cm 以上 170 cm 以下である確率を求める



例題 13.2 正規分布の確率を求める

区間の両端を標準化すると,

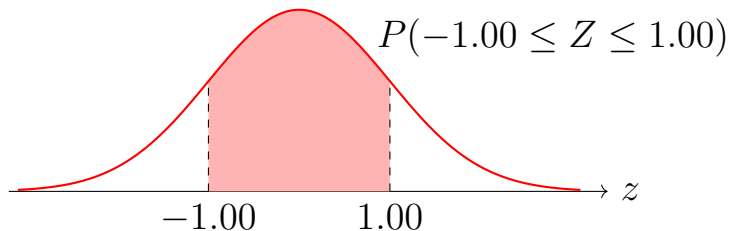
$$\begin{aligned}P(160 \leq X \leq 170) &= P\left(\frac{160 - 165}{5} \leq Z \leq \frac{170 - 165}{5}\right) \\&= P(-1.00 \leq Z \leq 1.00)\end{aligned}$$

となります。つまり,

$$160 \text{ cm} \longleftrightarrow -1.00, \quad 170 \text{ cm} \longleftrightarrow 1.00$$

です。

例題 13.2 正規分布の確率を求める



求める区間の確率は，上側確率の差を使って，

$$P(-1.00 \leq Z \leq 1.00) = P(Z > -1.00) - P(Z > 1.00)$$

と求めます。

例題 13.2 正規分布の確率を求める

標準正規分布表より，

$$P(Z > 1.00) = 0.1587$$

です．また，左右対称性と補集合の関係より，

$$\begin{aligned} P(Z > -1.00) &= 1 - P(Z < -1.00) \\ &= 1 - P(Z > 1.00) \\ &= 1 - 0.1587 \\ &= 0.8413 \end{aligned}$$

となります．

例題 13.2 正規分布の確率を求める

したがって、

$$\begin{aligned}P(-1.00 \leq Z \leq 1.00) &= P(Z > -1.00) - P(Z > 1.00) \\&= 0.8413 - 0.1587 \\&= 0.6826\end{aligned}$$

です。よって、求める確率は、

$$P(160 \leq X \leq 170) = 0.6826$$

です。

例題 13.3 二項分布を正規分布で近似して確率を求める

144 個の製品を検査したときの合格品の個数を表す確率変数を X とします。ただし、各製品が合格する確率は $1/2$ であり、各製品の合否は互いに独立であるとしします。

(1) 確率変数 X の期待値，分散，標準偏差を求めましょう。

(2) 正規分布による近似を用いて，合格品が 75 個以上となる確率 $P(X \geq 75)$ を求めましょう。

ただし，確率は小数第 2 位まで求めましょう。

例題 13.3 二項分布を正規分布で近似して確率を求める

1 個の製品について、

$$\text{合格する確率} = \frac{1}{2}$$

であり、各製品の合否は互いに独立です。

したがって、144 個の製品のうち合格した個数を表す X は、

二項分布 $B\left(144, \frac{1}{2}\right)$ に従います。

例題 13.3 二項分布を正規分布で近似して確率を求める

(1) 期待値・分散・標準偏差を求める

二項分布 $B(n, p)$ に従う確率変数 X の期待値と分散は,

$$E(X) = np, \quad V(X) = np(1 - p)$$

で求められます.

ここでは,

$$n = 144, \quad p = \frac{1}{2}$$

です.

例題 13.3 二項分布を正規分布で近似して確率を求める

したがって、期待値は

$$E(X) = 144 \times \frac{1}{2} = 72$$

となります。また、分散は、

$$V(X) = 144 \times \frac{1}{2} \times \left(1 - \frac{1}{2}\right) = 36$$

であり、標準偏差は、

$$\sqrt{36} = 6$$

となります。

例題 13.3 二項分布を正規分布で近似して確率を求める

二項分布の正規近似

正規近似とは，ある確率分布を正規分布で近似して扱う方法です．二項分布 $B(n, p)$ は，試行回数 n が大きくなるにつれて，正規分布に近づくことが知られています．

例題 13.3 二項分布を正規分布で近似して確率を求める

二項分布の正規近似

二項分布 $B(n, p)$ を近似する正規分布の期待値と分散は、

$$\text{期待値 } np, \quad \text{分散 } np(1 - p)$$

です。そのため、

$B(n, p)$ を $N(np, np(1 - p))$ で近似して扱う

ことがあります。

例題 13.3 二項分布を正規分布で近似して確率を求める

(2) 75 個以上となる確率を求める

二項分布を正規分布で近似するときは、

np と $n(1 - p)$ がともに十分大きい

ことが目安になります。

今回の場合は、

$$np = 144 \times \frac{1}{2} = 72, \quad n(1 - p) = 144 \times \frac{1}{2} = 72$$

であり、ここでは正規近似ができるものとして考えます。

例題 13.3 二項分布を正規分布で近似して確率を求める

(1) より,

$$E(X) = 72, \quad V(X) = 36, \quad \sigma = 6$$

でした.

そこで, 二項分布

$$B\left(144, \frac{1}{2}\right)$$

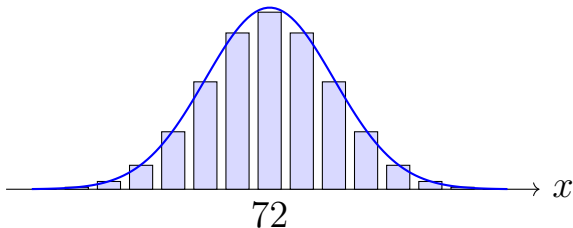
を,

$$N(72, 36)$$

で近似して扱います.

例題 13.3 二項分布を正規分布で近似して確率を求める

二項分布は離散型ですが，試行回数が大きくなると，正規分布に近い形になります．



二項分布を正規分布で近似

例題 13.3 二項分布を正規分布で近似して確率を求める

近似した正規分布の期待値は 72，標準偏差は 6 なので，

$$Z = \frac{X - 72}{6}$$

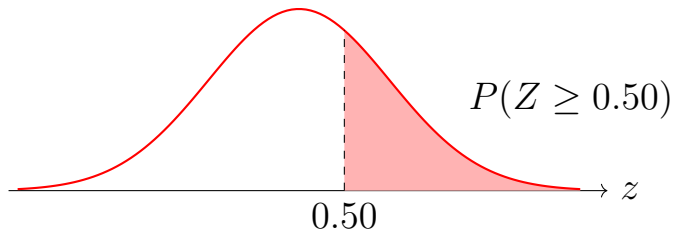
と標準化します．

求めたい確率は $P(X \geq 75)$ です．したがって，

$$P(X \geq 75) = P\left(Z \geq \frac{75 - 72}{6}\right) = P(Z \geq 0.50)$$

となります．

例題 13.3 二項分布を正規分布で近似して確率を求める



標準正規分布表より、

$$P(Z > 0.50) = 0.3085$$

です。

例題 13.3 二項分布を正規分布で近似して確率を求める

したがって、

$$P(X \geq 75) = 0.3085 \approx 0.31$$

となります。

よって、正規分布による近似では、合格品が 75 個以上となる確率は
約 0.31 です。

第13章 正規分布

13.2 標本平均・標本比率の性質

例題 13.4 標本平均の分布と確率を求める

ある科目の試験について考えます．ここでは，16 人の点数の平均がどのような値になりやすいかを調べることにします．

試験を行う前には，各生徒が何点を取るかわからないため，それぞれの生徒の点数を表す確率変数を

$$X_1, X_2, \dots, X_{16}$$

とします．

例題 13.4 標本平均の分布と確率を求める

16 人の点数は母集団から無作為に選ばれたものと考えます。
どの生徒の点数も、

期待値 50, 標準偏差 12

の正規分布に従うとします。

16 人の平均点を表す確率変数を

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{16}}{16}$$

とします。

例題 13.4 標本平均の分布と確率を求める

このとき，次の問いに答えましょう．

- (1) ある生徒の点数 X_1 が 65 点以上である確率を求めましょう．
- (2) 16 人の平均点 $\bar{X}$ が 54 点以上である確率を求めましょう．

例題 13.4 標本平均の分布と確率を求める

母平均と標本平均

母集団の平均値を母平均といい、通常 μ と書きます。
また、母集団の分散と標準偏差をそれぞれ

母分散, 母標準偏差

といい、通常

σ^2 , σ

と書きます。

例題 13.4 標本平均の分布と確率を求める

母平均や母分散など、

母集団の確率分布を特徴づける量を「母数」といいます。

今回の母集団では、

$$\mu = 50, \quad \sigma = 12$$

です。

例題 13.4 標本平均の分布と確率を求める

一般に、母集団から無作為に n 個のデータを取り出し、それぞれを

$$X_1, X_2, \dots, X_n$$

とすると、その平均を表す確率変数

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

を標本平均といいます。

例題 13.4 標本平均の分布と確率を求める

標本平均 $\bar{X}$ は,

母平均 μ を推定するための量

です. ここでは,

$$X_1, X_2, \dots, X_n$$

は互いに独立であるものとします.

各 X_i がいずれも期待値 μ , 分散 σ^2 の正規分布に従うとき, 標本平均 $\bar{X}$ も正規分布に従います.

例題 13.4 標本平均の分布と確率を求める

標本平均の期待値・分散・標準偏差

母平均を μ ，母標準偏差を σ ，標本サイズを n とすると，標本平均 $\bar{X}$ の期待値と分散は，

$$E(\bar{X}) = \mu, \quad V(\bar{X}) = \frac{\sigma^2}{n}$$

となります．したがって，標本平均の標準偏差は，次になります．

$$\frac{\sigma}{\sqrt{n}}$$

例題 13.4 標本平均の分布と確率を求める

(1) ある生徒が 65 点以上となる確率を求める

ある生徒の点数 X_1 は,

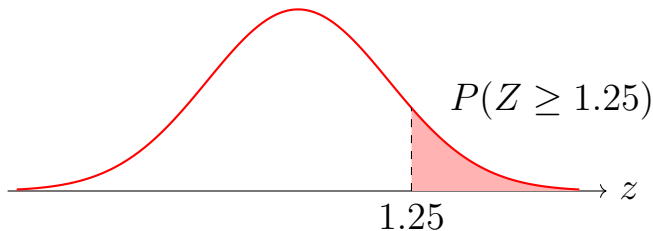
期待値 50, 標準偏差 12

の正規分布に従います. 求める確率 $P(X_1 \geq 65)$ を標準化すると,

$$\begin{aligned} P(X_1 \geq 65) &= P\left(Z \geq \frac{65 - 50}{12}\right) \\ &= P(Z \geq 1.25) \end{aligned}$$

となります.

例題 13.4 標本平均の分布と確率を求める



標準正規分布表より,

$$P(Z > 1.25) = 0.1056$$

です。したがって,

$$P(X_1 \geq 65) = 0.1056$$

例題 13.4 標本平均の分布と確率を求める

(2) 標本平均が 54 点以上となる確率を求める

母平均は $\mu = 50$ なので,

$$E(\bar{X}) = 50$$

です. また, 標本サイズは $n = 16$ なので,

$$V(\bar{X}) = \frac{12^2}{16} = \frac{144}{16} = 9$$

となります.

例題 13.4 標本平均の分布と確率を求める

よって、標本平均 $\bar{X}$ の標準偏差は、

$$\sqrt{9} = 3$$

です。

したがって、標本平均 $\bar{X}$ は、

期待値 50，標準偏差 3 の正規分布に従います。

例題 13.4 標本平均の分布と確率を求める

1 人の点数 X_1 は,

期待値 50, 標準偏差 12

でした. 一方, 16 人の標本平均 $\bar{X}$ は,

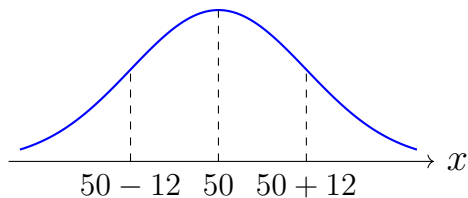
期待値 50, 標準偏差 3

です.

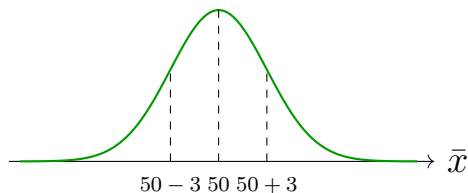
中心は同じですが,
標本平均のほうがばらつきが小さくなります.

例題 13.4 標本平均の分布と確率を求める

母集団の点数の分布



標本平均の分布



標本平均の分布は、1 人分の点数の分布よりも**中心付近に集中**しています。

例題 13.4 標本平均の分布と確率を求める

求める確率は、

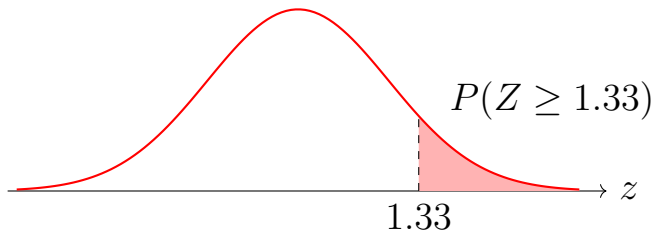
$$P(\bar{X} \geq 54)$$

です。標本平均の期待値は 50，標準偏差は 3 なので、

$$\begin{aligned} P(\bar{X} \geq 54) &= P\left(Z \geq \frac{54 - 50}{3}\right) \\ &= P(Z \geq 1.333\dots) \\ &\approx P(Z \geq 1.33) \end{aligned}$$

となります。

例題 13.4 標本平均の分布と確率を求める



標準正規分布表より,

$$P(Z > 1.33) = 0.0918$$

です。したがって,

$$P(\bar{X} \geq 54) = 0.0918$$

例題 13.5 標本比率の分布と確率を求める

ある公園で来園者を調べたところ、「犬を連れている人」の割合は母集団では $\frac{1}{2}$ であることがわかっています。

この公園に来た人のなかから 100 人を無作為に選び、そのうち何人が犬を連れているかを調べます。その人数を表す確率変数を X とします。

100 人のうち犬を連れていた人の割合を表す確率変数を、

$$\hat{p} = \frac{X}{100}$$

とします。

例題 13.5 標本比率の分布と確率を求める

このとき，次の問いに答えましょう．

- (1) 標本比率 $\hat{p}$ の期待値と標準偏差を求めましょう．
- (2) 標本比率 $\hat{p}$ が 0.55 以上となる確率を，正規分布による近似を用いて，小数第 2 位まで求めましょう．

例題 13.5 標本比率の分布と確率を求める

母比率

母集団において，ある事象が起こる割合を

母比率といい，通常 p と書きます．

例題 13.5 標本比率の分布と確率を求める

標本比率

母集団から無作為に n 個を取り出し，そのうちある事象が起こった個数を確率変数 X とすると，その割合を表す確率変数を，

$$\hat{p} = \frac{X}{n}$$

と書きます．この $\hat{p}$ を標本比率といいます．

例題 13.5 標本比率の分布と確率を求める

今回，

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

は母集団全体における「犬を連れている人」の割合です。
一方，

$$\hat{p} = \frac{X}{100}$$

は，無作為に選んだ 100 人のなかでの「犬を連れている人」の割合です．

例題 13.5 標本比率の分布と確率を求める

標本比率の期待値・分散・標準偏差

母比率を p ，標本サイズを n とすると，標本比率 $\hat{p}$ の期待値と分散は，

$$E(\hat{p}) = p, \quad V(\hat{p}) = \frac{p(1-p)}{n}$$

となります．したがって，標本比率 $\hat{p}$ の標準偏差は，次になります

$$\sqrt{\frac{p(1-p)}{n}}$$

例題 13.5 標本比率の分布と確率を求める

標本サイズ n が十分大きいときには、標本比率 $\hat{p}$ の分布を、

$$\text{期待値 } p, \quad \text{分散 } \frac{p(1-p)}{n}$$

の正規分布で近似して扱うことができます。

例題 13.5 標本比率の分布と確率を求める

正規分布で近似するときは、

$$np, \quad n(1 - p)$$

がともに十分大きいことが目安です。

今回の場合は、

$$np = 100 \times \frac{1}{2} = 50, \quad n(1 - p) = 100 \times \frac{1}{2} = 50$$

であり、ここでは正規近似ができるものとして考えます。

例題 13.5 標本比率の分布と確率を求める

(1) 標本比率の期待値と標準偏差を求める

母比率は $p = \frac{1}{2}$ なので,

$$E(\hat{p}) = p$$

より,

$$E(\hat{p}) = \frac{1}{2}$$

となります。

例題 13.5 標本比率の分布と確率を求める

標本比率 $\hat{p}$ の標準偏差は,

$$\sqrt{\frac{p(1-p)}{n}}$$

より,

$$\sqrt{\frac{\frac{1}{2} \left(1 - \frac{1}{2}\right)}{100}} = \sqrt{\frac{1}{400}} = \frac{1}{20}$$

となります。

例題 13.5 標本比率の分布と確率を求める

したがって、標本比率 $\hat{p}$ について、

$$\text{期待値} = \frac{1}{2}, \quad \text{標準偏差} = \frac{1}{20}$$

となります。

例題 13.5 標本比率の分布と確率を求める

(2) 標本比率が 0.55 以上となる確率を求める

標本比率 $\hat{p}$ の分布を,

$$\text{期待値 } \frac{1}{2}, \quad \text{標準偏差 } \frac{1}{20}$$

の正規分布で近似します.

求める確率は,

$$P(\hat{p} \geq 0.55)$$

です.

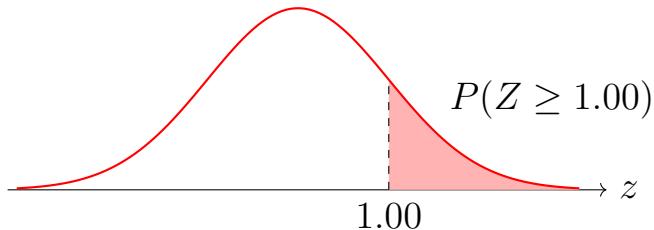
例題 13.5 標本比率の分布と確率を求める

標準化すると,

$$\begin{aligned} P(\hat{p} \geq 0.55) &= P\left(Z \geq \frac{0.55 - \frac{1}{2}}{\frac{1}{20}}\right) \\ &= P(Z \geq 1.00) \end{aligned}$$

となります.

例題 13.5 標本比率の分布と確率を求める



標準正規分布表より,

$$P(Z > 1.00) = 0.1587$$

です。したがって,

$$P(\hat{p} \geq 0.55) = 0.1587 \approx 0.16$$

第14章 推定と検定

14.1 区間推定

例題 14.1 信頼区間の考え方を理解する

ある高校で、全国模試（100 点満点）が実施されました．模試の採点結果は外部業者から順次返却されるため、この時点では 3 年生全員の点数はまだそろっていません．

そこで、先に返却された 1 クラス 36 人の点数を用いて、3 年生全体の平均点（母平均） μ を推定することにします．

3 年生のクラス分けは無作為に行われているものとします．

例題 14.1 信頼区間の考え方を理解する

このクラスの平均点は、

$$\bar{X} = 70$$

でした。また、全国の過去の受験データから、模試の点数の標準偏差は一定とみなせることがわかっており、

$$\sigma = 12$$

とされています。

ここでは、一つの値だけで推定するのではなく、「母平均を推定するための区間」を考えます。

例題 14.1 信頼区間の考え方を理解する

3 年生全体の平均点（母平均） μ の 95% 信頼区間は，

$$70 - 1.96 \times \frac{12}{\sqrt{36}} \leq \mu \leq 70 + 1.96 \times \frac{12}{\sqrt{36}}$$

です。したがって，

$$70 - 3.92 \leq \mu \leq 70 + 3.92$$

より，

$$66.08 \leq \mu \leq 73.92$$

となります。以下では，端点を四捨五入して**およそ 66～74** と表します。

例題 14.1 信頼区間の考え方を理解する

この具体例を参考にして，次の記述が母平均の 95% 信頼区間の一般的な性質として適切かどうかを判断しましょう．適切なら○，適切でなければ×をつけます．

- (1) 母平均は，計算された 95% 信頼区間に含まれている．
- (2) 母平均が，計算された 95% 信頼区間に入る確率は 0.95 である．
- (3) 同じ手順で多数の標本を取り直して 95% 信頼区間を作ると，その手続きは理論上，約 95% の割合で母平均を含むように定められている．

例題 14.1 信頼区間の考え方を理解する

- (4) 信頼区間の中心は、母平均と一致する。
- (5) 信頼区間の幅は、標本平均の値によって変化する。
- (6) 母標準偏差が大きいほど、信頼区間は広くなる。
- (7) 標本サイズを大きくすると、信頼区間は狭くなる。
- (8) 信頼区間の中心は、標本平均である。

例題 14.1 信頼区間の考え方を理解する

区間推定と母平均の 95% 信頼区間

区間推定とは、未知の母数の値を 1 点で推定するのではなく、

その母数が含まれると考えられる区間

を定めて推定する方法です。

母平均の 95% 信頼区間とは、同じ手順で多数の標本を取り直して区間を作ったとき、そのおよそ 95% が本当の母平均を含むように定められた区間のことです。

例題 14.1 信頼区間の考え方を理解する

第 13 章で学んだように、各 X_i が互いに独立で、いずれも期待値 μ 、標準偏差 σ の正規分布に従うとき、標本平均 $\bar{X}$ は、

$$\text{期待値 } \mu, \quad \text{標準偏差 } \frac{\sigma}{\sqrt{n}}$$

の正規分布に従います。

ここで、 μ は母平均、 σ は母標準偏差です。

例題 14.1 信頼区間の考え方を理解する

また，母集団が正規分布に従わない場合でも，標本サイズ n が十分大きいときには，近似的にこの正規分布に従います．

したがって，

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

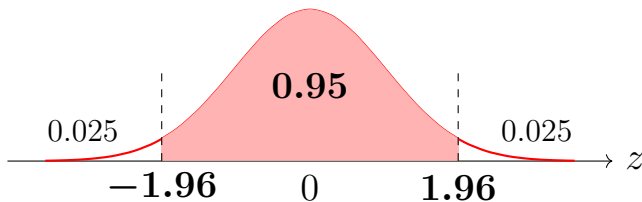
と標準化すると，前者では標準正規分布に従い，後者では近似的に標準正規分布に従います．

例題 14.1 信頼区間の考え方を理解する

標準正規分布では、

$$-1.96 \leq Z \leq 1.96$$

となる確率が 0.95 です。



例題 14.1 信頼区間の考え方を理解する

$$-1.96 \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq 1.96$$

を μ について解くと,

$$\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

となります.

これが、母標準偏差 σ が既知のときの母平均 μ の 95% 信頼区間です.

例題 14.1 信頼区間の考え方を理解する

(1) 1 回の信頼区間と母平均

95% 信頼区間は、母平均を区間で推定したものです。しかし、1 回の標本から作った区間が実際に母平均を含んでいるかどうかは断定できません。

今回、

66～74

という区間が得られましたが、この区間が実際に母平均を含んでいるかどうかはわかりません。したがって、

×

例題 14.1 信頼区間の考え方を理解する

(2) 母平均が区間に入る確率

母平均 μ は、固定された 1 つの値です。したがって、

「母平均が計算された区間に入る確率は 0.95」

という解釈は適切ではありません。

確率に関係するのは、標本を取り直したときに変化する標本平均や信頼区間のほうです。したがって、



例題 14.1 信頼区間の考え方を理解する

(3) 多数の信頼区間を作ったとき

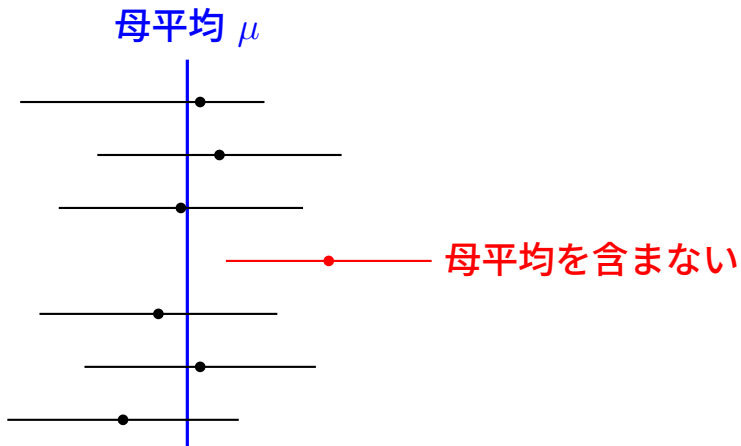
同じ条件で標本を多数取り直し，そのつど

$$\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

の形で 95% 信頼区間を作ると，多数の区間のうち約 95% が母平均 μ を含むという性質があります．したがって，



例題 14.1 信頼区間の考え方を理解する



例題 14.1 信頼区間の考え方を理解する

標本を取り直すたびに標本平均が変わるため、作られる信頼区間も変わります。

このため、多数の信頼区間を作ると、母平均 μ を含む区間もあれば、含まない区間もあります。

例題 14.1 信頼区間の考え方を理解する

(4) 信頼区間の中心は母平均か

母平均の 95% 信頼区間は,

$$\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

です。したがって、区間の中心は母平均 μ ではなく,

標本平均 $\bar{X}$

です。よって,



例題 14.1 信頼区間の考え方を理解する

(5) 信頼区間の幅

信頼区間の幅は、

$$2 \times 1.96 \times \frac{\sigma}{\sqrt{n}}$$

です。したがって、信頼区間の幅は母標準偏差 σ と標本サイズ n によって決まり、 $\bar{X}$ の値そのものには依存しません。
よって、

×

例題 14.1 信頼区間の考え方を理解する

今回，

$$1.96 \times \frac{12}{\sqrt{36}} = 1.96 \times \frac{12}{6} = 3.92$$

なので，信頼区間の幅は，

$$3.92 \times 2 = 7.84$$

です．したがって，標本平均が 70 でも 80 でも，

信頼区間の幅は 7.84 で変わりません．

例題 14.1 信頼区間の考え方を理解する

(6) 母標準偏差と信頼区間の幅

信頼区間の片側の広がりとは、

$$1.96 \times \frac{\sigma}{\sqrt{n}}$$

です。したがって、母標準偏差 σ が大きいほど、信頼区間は広くなります。したがって、



例題 14.1 信頼区間の考え方を理解する

標本サイズを $n = 36$ とすると,

$$\sigma = 12 \text{ のとき} \quad 1.96 \times \frac{12}{6} = 3.92,$$

$$\sigma = 24 \text{ のとき} \quad 1.96 \times \frac{24}{6} = 7.84$$

となります.

母標準偏差が 2 倍になると, 片側の広がりも 2 倍になります.

例題 14.1 信頼区間の考え方を理解する

(7) 標本サイズと信頼区間の幅

信頼区間の片側の広がりとは、

$$1.96 \times \frac{\sigma}{\sqrt{n}}$$

です。標本サイズ n が大きくなるほど $\sigma/\sqrt{n}$ は小さくなるため、信頼区間は狭くなります。したがって、



例題 14.1 信頼区間の考え方を理解する

母標準偏差を $\sigma = 12$ とすると,

$$n = 36 \text{ のとき} \quad 1.96 \times \frac{12}{6} = 3.92,$$

$$n = 100 \text{ のとき} \quad 1.96 \times \frac{12}{10} = 2.352 \approx 2.35$$

となります.

標本サイズが大きくなると, 左右の広がりは小さくなります.

例題 14.1 信頼区間の考え方を理解する

(8) 信頼区間の中心は標本平均か

母平均の 95% 信頼区間は,

$$\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

です. この式から, 信頼区間は標本平均 $\bar{X}$ を中心に左右に同じだけ広げた区間であることがわかります. したがって,



例題 14.1 信頼区間の考え方を理解する

今回の標本平均は,

$$\bar{X} = 70$$

です. 95% 信頼区間は,

$$70 - 3.92 \leq \mu \leq 70 + 3.92$$

なので,

$$[66.08, 73.92]$$

となります.

例題 14.1 信頼区間の考え方を理解する

区間の中心を確認すると、

$$\frac{66.08 + 73.92}{2} = 70 = \bar{X}$$

となります。

区間の中心は標本平均 $\bar{X}$ と一致しています。

例題 14.2 母平均の 95% 信頼区間を求める

ある大学で、オンライン講義の 1 回あたりの視聴時間（分）について、受講者全体の平均視聴時間（母平均） μ を推定したいと考えています。

過去のログデータから、視聴時間のばらつきは一定であり、母標準偏差は

$$\sigma = 20 \text{ 分}$$

であるとわかっています。

例題 14.2 母平均の 95% 信頼区間を求める

受講者のなかから 100 人を無作為に抽出したところ、平均視聴時間は

$$\bar{X} = 50 \text{ 分}$$

でした。

受講者全体の平均視聴時間（母平均） μ の

95% 信頼区間を求めましょう。

例題 14.2 母平均の 95% 信頼区間を求める

母標準偏差が既知のとき，母平均 μ の 95% 信頼区間は，

$$\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

で求められます．

本問では，

$$\bar{X} = 50, \quad \sigma = 20, \quad n = 100$$

です．

例題 14.2 母平均の 95% 信頼区間を求める

母平均の 95% 信頼区間

公式に代入すると,

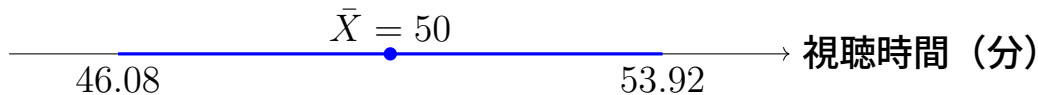
$$50 - 1.96 \times \frac{20}{\sqrt{100}} \leq \mu \leq 50 + 1.96 \times \frac{20}{\sqrt{100}}$$

となり, これを計算すると,

$$46.08 \leq \mu \leq 53.92$$

となります.

例題 14.2 母平均の 95% 信頼区間を求める



母平均の 95% 信頼区間は、
標本平均 $\bar{X} = 50$ を中心に
左右へ同じだけ広がっています。

例題 14.3 母比率の 95% 信頼区間を求める

ある大学で、図書館を「週 1 回以上利用している」学生の割合（母比率） p を推定したいと考えています。

無作為に抽出した 400 人のうち、週 1 回以上利用している学生は 80 人でした。

このとき、正規分布による近似を用いて、週 1 回以上利用している学生の割合（母比率） p の 95% 信頼区間を求めましょう。

計算結果は、小数第 3 位まで求めましょう。

例題 14.3 母比率の 95% 信頼区間を求める

母比率の 95% 信頼区間

母比率の 95% 信頼区間とは、同じ手順で多数の標本を取り直して区間を作ったとき、

そのおよそ 95% が本当の母比率を含むように
定められた区間

のことです。

例題 14.3 母比率の 95% 信頼区間を求める

標本サイズ n が十分大きいとき、標本比率 $\hat{p}$ の分布は、

$$\text{期待値 } p, \quad \text{標準偏差 } \sqrt{\frac{p(1-p)}{n}}$$

の正規分布で近似して扱うことができます。

例題 14.3 母比率の 95% 信頼区間を求める

標本比率 $\hat{p}$ を標準化すると,

$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

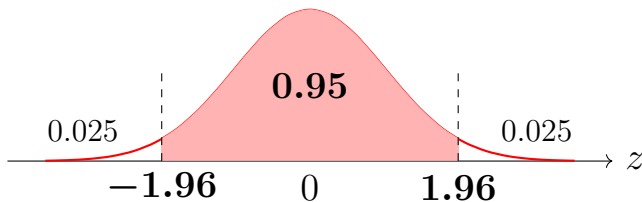
は近似的に標準正規分布に従います.

例題 14.3 母比率の 95% 信頼区間を求める

標準正規分布の中央 95% は、

$$-1.96 \leq Z \leq 1.96$$

です。



例題 14.3 母比率の 95% 信頼区間を求める

母比率 p は未知なので，分散に含まれる p を標本比率 $\hat{p}$ で近似して扱います．

その結果，母比率 p の 95% 信頼区間は，

$$\hat{p} - 1.96 \times \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \leq p \leq \hat{p} + 1.96 \times \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

となります．

例題 14.3 母比率の 95% 信頼区間を求める

母比率の 95% 信頼区間

400 人のうち、週 1 回以上利用している学生は 80 人なので、

$$\hat{p} = \frac{80}{400} = \frac{1}{5}$$

です。また、

$$n = 400$$

です。

例題 14.3 母比率の 95% 信頼区間を求める

公式に代入すると,

$$\frac{1}{5} - 1.96 \times \sqrt{\frac{\frac{1}{5} \left(1 - \frac{1}{5}\right)}{400}} \leq p \leq \frac{1}{5} + 1.96 \times \sqrt{\frac{\frac{1}{5} \left(1 - \frac{1}{5}\right)}{400}}$$

となります.

例題 14.3 母比率の 95% 信頼区間を求める

計算すると,

$$\frac{1}{5} - 1.96 \times \frac{1}{50} \leq p \leq \frac{1}{5} + 1.96 \times \frac{1}{50}$$

より,

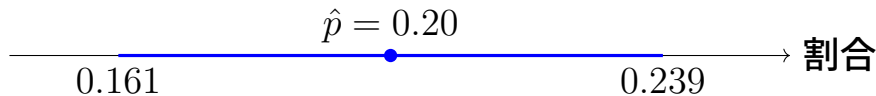
$$0.1608 \leq p \leq 0.2392$$

となります. 小数第 3 位まで示すと,

$$0.161 \leq p \leq 0.239$$

です.

例題 14.3 母比率の 95% 信頼区間を求める



母比率の 95% 信頼区間は、
標本比率 $\hat{p}$ を中心に
左右へ同じだけ広がっています。

第14章 推定と検定

14.2 仮説検定

例題 14.4 母平均の仮説検定を行う

ある高校で、数学のテスト（100 点満点）が実施されました。
例年の平均点（母平均）は

$$\mu = 50.0$$

で安定していると知られています。また、点数のばらつきは一定で、
母標準偏差は

$$\sigma = 12.0$$

であることがわかっています。

例題 14.4 母平均の仮説検定を行う

今年度、3 年生のクラスを無作為に 2 つ抽出して平均点を調べたところ、次のようになりました．

	人数	平均点
クラス A	36	53.6
クラス B	36	55.0

3 年生のクラス分けは無作為に行われているものとします．

例題 14.4 母平均の仮説検定を行う

ここでは、例年の平均点 50.0 点と比べて、**今年度の平均点が高いか低いかを含めて変わったといえるか**を調べます。

帰無仮説 H_0 と対立仮説 H_1 を、

$$H_0 : \mu = 50$$

$$H_1 : \mu \neq 50$$

とし、有意水準 5% の両側検定を行います。

例題 14.4 母平均の仮説検定を行う

次の記述が適切である場合は○，適切でない場合は×をつけましょう．

- (1) クラス A は，標本平均が 50 を上回っているため，帰無仮説を棄却すると判断できる．
- (2) クラス A の標本平均の z 値は， -1.96 から 1.96 の範囲にあるので，帰無仮説を棄却しない．
- (3) クラス B は，標本平均の z 値が棄却域に入るため，帰無仮説を棄却する．

例題 14.4 母平均の仮説検定を行う

- (4) 母平均の検定では，標本平均そのものではなく，標本平均を標準化した検定統計量（ z 値など）で判断する．
- (5) 有意水準 5% の両側検定では，標準正規分布を用いる場合の棄却域は $z \geq 1.96$ である．
- (6) 「帰無仮説を棄却しない」とは，帰無仮説が正しいと判断したという意味である．

例題 14.4 母平均の仮説検定を行う

仮説検定

仮説検定または検定とは、母数について立てた仮定（帰無仮説）と観測されたデータがどの程度整合しているかを数値的に判断する方法です。

帰無仮説とは、母数について立てる仮定のうち、検定の出発点として置く仮定です。

対立仮説とは、帰無仮説に対立する主張です。

例題 14.4 母平均の仮説検定を行う

今回の帰無仮説は，

$$H_0 : \mu = 50$$

であり，「母平均は 50.0 点である（例年どおりといえる）」という仮定です．これに対する対立仮説は，

$$H_1 : \mu \neq 50$$

であり，「母平均は 50.0 点ではない（例年どおりとはいえない）」という主張です．

例題 14.4 母平均の仮説検定を行う

有意水準とは、**どれくらい厳しく極端な値を「めったに起こらない」とみなすか**をあらかじめ決めるための基準です。

この極端な値の領域を棄却域といいます。

有意水準 5% とは、帰無仮説が正しいときにそれを誤って棄却する確率を 5% 以下に抑えるという意味です。

例題 14.4 母平均の仮説検定を行う

観測されたデータが，帰無仮説のもとでどれくらい極端なズレを示しているかを数値化した量を検定統計量といいます．

母分散既知の母平均の検定では，

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

という z 値を検定統計量として用います．ここで，各記号は次のものを表します．

$\bar{X}$: 標本平均

μ_0 : 帰無仮説で仮定した母平均

σ : 母標準偏差

n : 標本サイズ

例題 14.4 母平均の仮説検定を行う

検定統計量が棄却域に入れば、帰無仮説のもとではめったに起こらないほど極端と判断して、

帰無仮説を棄却

します。棄却域に入らなければ、帰無仮説を棄却するだけの根拠が得られなかったとして、

帰無仮説を棄却しない

と判断します。

例題 14.4 母平均の仮説検定を行う

今回は，帰無仮説で仮定した母平均，母標準偏差，標本サイズは，それぞれ

$$\mu_0 = 50.0, \quad \sigma = 12.0, \quad n = 36$$

です．これらの値と，各クラスの標本平均を，

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

に代入して，検定統計量を求めます．

例題 14.4 母平均の仮説検定を行う

クラス A の検定統計量

クラス A では,

$$\bar{X} = 53.6$$

なので,

$$z_A = \frac{53.6 - 50.0}{12.0/\sqrt{36}} = \frac{3.6}{2.0} = 1.8$$

となります。

例題 14.4 母平均の仮説検定を行う

クラス B の検定統計量

クラス B では,

$$\bar{X} = 55.0$$

なので,

$$z_B = \frac{55.0 - 50.0}{12.0/\sqrt{36}} = \frac{5.0}{2.0} = 2.5$$

となります。

例題 14.4 母平均の仮説検定を行う

有意水準 5% の両側検定では、標準正規分布の左右の 2.5% ずつが棄却域になります。

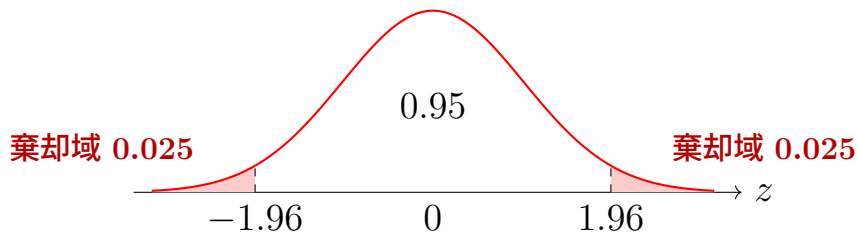
$$|Z| \geq 1.96$$

つまり、

$$z \leq -1.96 \quad \text{または} \quad z \geq 1.96$$

が棄却域です。

例題 14.4 母平均の仮説検定を行う



両側検定では、**正方向・負方向のどちらの大きなズレも**問題にします。

例題 14.4 母平均の仮説検定を行う

(1) 標本平均が 50 を上回れば棄却できるか

クラス A の検定統計量は、 $z_A = 1.8$ です．有意水準 5% の両側検定の棄却域は、

$$z \leq -1.96 \quad \text{または} \quad z \geq 1.96$$

なので、 z_A は棄却域に入りません．

標本平均が 50 を上回っているだけでは、帰無仮説を棄却することはできません．したがって、



例題 14.4 母平均の仮説検定を行う

(2) クラス A の z 値と棄却域

クラス A の検定統計量は、 $z_A = 1.8$ です.

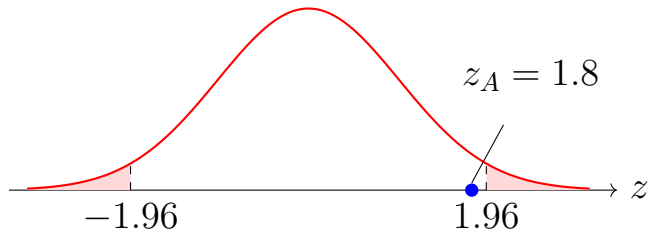
$$-1.96 < 1.8 < 1.96$$

なので、 z_A は棄却域には入りません.

したがって、帰無仮説は棄却されません.



例題 14.4 母平均の仮説検定を行う



$z_A = 1.8$ は棄却域の外にあります。

例題 14.4 母平均の仮説検定を行う

(3) クラス B の z 値と棄却域

クラス B の検定統計量は、 $z_B = 2.5$ です.

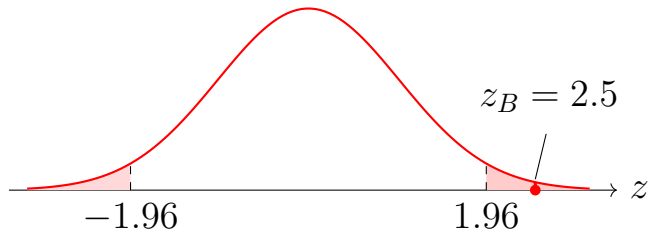
$$2.5 > 1.96$$

なので、 z_B は棄却域に入ります.

したがって、帰無仮説を棄却し、対立仮説を支持します.



例題 14.4 母平均の仮説検定を行う



$z_B = 2.5$ は**棄却域**に入っています。

例題 14.4 母平均の仮説検定を行う

(4) 検定で用いる値

母平均の検定では，標本平均そのものの大きさだけでは判断しません．

標本平均を，帰無仮説で仮定した母平均 μ_0 を用いて，

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

と標準化して得られる検定統計量を用いて判断します．したがって，



例題 14.4 母平均の仮説検定を行う

(5) 5% 両側検定の棄却域

5% 両側検定では、左右の 2.5% ずつが棄却域です。したがって、

$$z \leq -1.96 \quad \text{または} \quad z \geq 1.96$$

が棄却域です。

$z \geq 1.96$ だけでは、右側の棄却域しか表していません。したがって、



例題 14.4 母平均の仮説検定を行う

(6) 「帰無仮説を棄却しない」の意味

「帰無仮説を棄却しない」とは、**帰無仮説が正しいと判断したという意味ではありません**．今回のデータからは、帰無仮説を棄却するだけの根拠が得られなかったという意味です．したがって、



例題 14.4 母平均の仮説検定を行う

仮説検定の基本的な流れ

仮説検定は、次の 4 つの手順で進めます。

(1) 仮説を立てる

帰無仮説と対立仮説を設定します。

(2) 検定の設定を決める

両側検定か片側検定かを選び、有意水準を定めます。

例題 14.4 母平均の仮説検定を行う

仮説検定の基本的な流れ

(3) 検定統計量を計算する
観測されたデータから検定統計量を求めます。

(4) 結論を出す
検定統計量が棄却域に入るかどうかで判断します。

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

例題 14.4 で扱った高校の数学テストと同じデータを用います。

以下に，設定とデータを再掲します。

例年の平均点は

$$\mu = 50.0$$

母標準偏差は

$$\sigma = 12.0$$

です。

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

今年度の2つのクラスのデータは，次のとおりです．

	人数	平均点
クラス A	36	53.6
クラス B	36	55.0

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

今年度の平均点が、

「高いか低いかを含めて変わった」

といえるかを調べるときは、両側検定を行います。一方、

「高くなっている」

といえるかを調べるときは、片側検定（上側）を行います。

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

両側検定と片側検定の帰無仮説 H_0 と対立仮説 H_1 は、それぞれ次のようになります。

$$\text{両側検定：} \quad H_0 : \mu = 50.0, \quad H_1 : \mu \neq 50.0$$

$$\text{片側検定（上側）：} \quad H_0 : \mu = 50.0, \quad H_1 : \mu > 50.0$$

また、例題 14.4 で求めた検定統計量は、

$$z_A = 1.80, \quad z_B = 2.50$$

です。

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

(1) クラス A・クラス B の z 値について、次の 4 つの検定で帰無仮説 H_0 を棄却するかどうかを判断しましょう。

○は「棄却する」，×は「棄却しない」を表します。

クラス	z 値	5%両側	1%両側	5%片側（上側）	1%片側（上側）
A	1.80				
B	2.50				

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

(2) 次の記述が適切なら○，適切でなければ×をつけましょう。

- (a) 片側検定と両側検定のどちらを選ぶかは，有意になりやすいほうはどちらかという基準で決める．
- (b) 両側検定では，「高い場合」と「低い場合」の両方のズレを問題にする．
- (c) 有意水準を 5% から 1% にすると，棄却域は狭くなる．

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

- (d) 標本サイズを固定したまま有意水準を小さくすると、第 1 種の過誤は起こりやすくなる。
- (e) 標本サイズを固定したまま有意水準を小さくすると、第 2 種の過誤は起こりやすくなる。
- (f) 第 1 種の過誤とは、本当は帰無仮説が正しいのに棄却してしまう誤りである。

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

- (g) 第 2 種の過誤とは、本当は帰無仮説が誤りなのに棄却しない誤りである。
- (h) この例における第 1 種の過誤とは、本当は母平均は 50.0 から変わっているのに、「変わっていない」と判断する誤りである。
- (i) この例における第 2 種の過誤とは、本当は母平均は 50.0 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誤りである。

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

両側検定と片側検定

両側検定は，帰無仮説で仮定した値からのズレを，両方向について考える検定です．

片側検定は，帰無仮説で仮定した値からのズレを，一方の方向について考える検定です．

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

有意水準 5% の両側検定

左右に 2.5% ずつ棄却域をとるので、

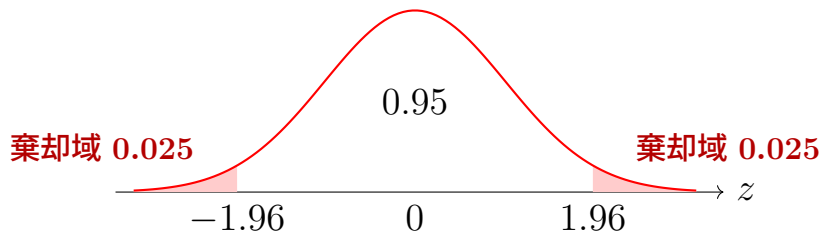
$$|z| \geq 1.96$$

つまり、

$$z \leq -1.96 \quad \text{または} \quad z \geq 1.96$$

が棄却域です。

例題 14.5 両側検定・片側検定と有意水準のちがいを考える



例題 14.5 両側検定・片側検定と有意水準のちがいを考える

有意水準 5% の片側検定

上側片側検定では，上側 5% を棄却域とするので，

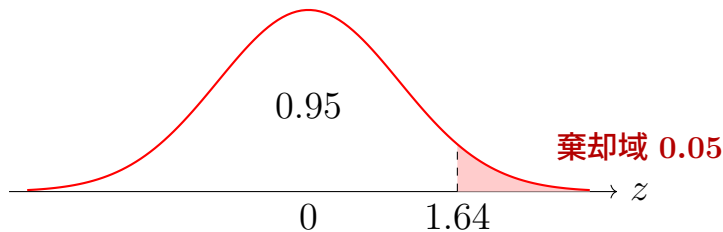
$$z \geq 1.64$$

です．下側片側検定では，下側 5% を棄却域とするので，

$$z \leq -1.64$$

です．

例題 14.5 両側検定・片側検定と有意水準のちがいを考える



高い方向だけを問題にするときには、**上側片側検定**を用います。

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

片側検定と両側検定の選択は、

対立仮説の内容
(どの方向のズレを問題にするか)

によって決まります。

なお、検定統計量とその検定で採用した有意水準の棄却域に入ったとき、**有意である**といいます。

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

有意水準と過誤の関係

仮説検定で判断を誤ることを過誤といいます。

第 1 種の過誤とは、本当は帰無仮説が正しいのに、帰無仮説を棄却してしまう誤りです。

第 2 種の過誤とは、本当は帰無仮説が誤りなのに、帰無仮説を棄却しない誤りです。

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

今回の帰無仮説は，

$$H_0 : \mu = 50$$

です．

したがって，この例における第 1 種の過誤は，

本当は母平均が 50 点のままなのに，
「変わった」と判断する誤り

です．

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

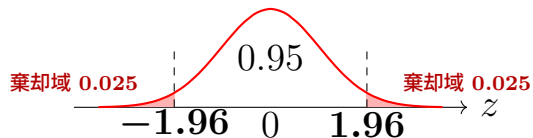
この例における第 2 種の過誤は、

**本当は母平均が 50 から変化しているのに、
帰無仮説を棄却しない誤りです。**

つまり、実際には変化しているのに、「変化したとはいえない」と判断する場合です。

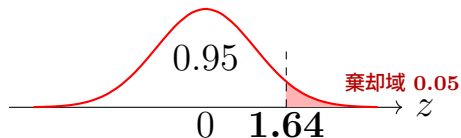
例題 14.5 両側検定・片側検定と有意水準のちがいを考える

両側検定・片側検定と有意水準の比較



有意水準 5% の両側検定

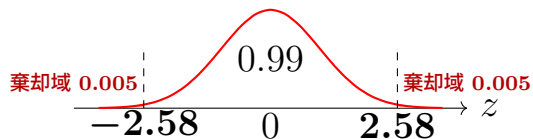
棄却域： $|z| \geq 1.96$



有意水準 5% の片側検定（上側）

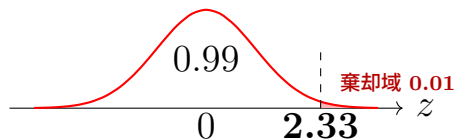
棄却域： $z \geq 1.64$

例題 14.5 両側検定・片側検定と有意水準のちがいを考える



有意水準 1% の両側検定

棄却域： $|z| \geq 2.58$



有意水準 1% の片側検定（上側）

棄却域： $z \geq 2.33$

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

今回使う棄却域を整理すると、

検定	棄却域
5% 両側	$ z \geq 1.96$
1% 両側	$ z \geq 2.58$
5% 片側（上側）	$z \geq 1.64$
1% 片側（上側）	$z \geq 2.33$

有意水準が小さいほど、棄却するためには、より極端な z 値が必要になります。

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

標本サイズを固定したまま有意水準を小さくすると、

より慎重に帰無仮説を棄却することになります。

そのため、

第 1 種の過誤は起こりにくくなり、
第 2 種の過誤は起こりやすくなります。

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

(1) クラス A の検定結果

クラス A では、 $z_A = 1.80$ です。したがって、

検定	比較	判定
5% 両側	$1.80 < 1.96$	×
1% 両側	$1.80 < 2.58$	×
5% 片側（上側）	$1.80 > 1.64$	○
1% 片側（上側）	$1.80 < 2.33$	×

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

(1) クラス B の検定結果

クラス B では、 $z_B = 2.50$ です。したがって、

検定	比較	判定
5% 両側	$2.50 > 1.96$	○
1% 両側	$2.50 < 2.58$	×
5% 片側（上側）	$2.50 > 1.64$	○
1% 片側（上側）	$2.50 > 2.33$	○

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

(1) 検定結果をまとめる

クラス	z 値	5%両側	1%両側	5%片側（上側）	1%片側（上側）
A	1.80	×	×	○	×
B	2.50	○	×	○	○

同じ z 値でも、両側・片側のちがいや有意水準によって判断が変わることがわかります。

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

(2)(a) 片側検定と両側検定の選択

片側検定と両側検定のどちらを選ぶかは、**対立仮説の内容**によって決まります．

「有意になりやすいほう」を後から選ぶことはできません．したがって、



例題 14.5 両側検定・片側検定と有意水準のちがいを考える

(2)(b) 両側検定で考えるズレ

両側検定では、 $\mu > 50$ となる高い方向だけでなく、 $\mu < 50$ となる低い方向のズレも問題にします。

つまり、**両方向のズレを調べます**。したがって、



例題 14.5 両側検定・片側検定と有意水準のちがいを考える

(2)(c) 有意水準と棄却域

有意水準 5% の両側検定では、 $|z| \geq 1.96$ が棄却域です。一方、有意水準 1% では、 $|z| \geq 2.58$ が棄却域です。

したがって、有意水準を小さくすると棄却域は狭くなります。



例題 14.5 両側検定・片側検定と有意水準のちがいを考える

(2)(d) 有意水準と第 1 種の過誤

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。

有意水準を小さくすると、帰無仮説をより慎重に棄却するため、第 1 種の過誤は起こりにくくなります。したがって、



例題 14.5 両側検定・片側検定と有意水準のちがいを考える

(2)(e) 有意水準と第 2 種の過誤

有意水準を小さくすると，帰無仮説を棄却しにくくなります．

そのため，本当は帰無仮説が誤っていても棄却できない場合が増え，第 2 種の過誤は起こりやすくなります．したがって，



例題 14.5 両側検定・片側検定と有意水準のちがいを考える

(2)(f) 第 1 種の過誤

第 1 種の過誤とは、本当は帰無仮説が正しいのに、帰無仮説を棄却してしまう誤りです。したがって、



例題 14.5 両側検定・片側検定と有意水準のちがいを考える

(2)(g) 第 2 種の過誤

第 2 種の過誤とは、本当は帰無仮説が誤りなのに、帰無仮説を棄却しない誤りです。したがって、



例題 14.5 両側検定・片側検定と有意水準のちがいを考える

(2)(h) 過誤の種類を判断する

「本当は母平均が 50.0 から変わっているのに、帰無仮説を棄却しない」というのは、**本当は帰無仮説が誤りなのに棄却しない**場合です。

これは第 2 種の過誤です。したがって、



例題 14.5 両側検定・片側検定と有意水準のちがいを考える

(2)(i) 過誤の種類を判断する

「本当は母平均が 50.0 のままなのに、帰無仮説を棄却してしまう」というのは、**本当は帰無仮説が正しいのに棄却する**場合です。

これは第 1 種の過誤です。したがって、



例題 14.5 両側検定・片側検定と有意水準のちがいを考える

両側検定・片側検定・有意水準の整理

何を決めるか	基準
両側検定か片側検定か	対立仮説の内容
有意水準	どれだけ慎重に棄却するか
棄却するかどうか	検定統計量が棄却域に入るか

例題 14.5 両側検定・片側検定と有意水準のちがいを考える

とくに、

片側検定か両側検定かは、
結果を見てから選ぶものではありません。

検定を行う前に、対立仮説の内容に応じて決めます。

例題 14.6 母比率の仮説検定を行う

ある国のパスポート所持率（母比率）は，例年 $p = 0.30$ といわれています．

今年，2 つの地域でそれぞれ 200 人を無作為に抽出して調査したところ，次の結果が得られました．

	所持者数	標本比率
地域 A	66	0.33
地域 B	74	0.37

例題 14.6 母比率の仮説検定を行う

例年の所持率 0.30 と比べて、今年所持率が**高いか低いかを含めて異なるといえるか**を調べます。

帰無仮説 H_0 と対立仮説 H_1 を、

$$H_0 : p = 0.30, \quad H_1 : p \neq 0.30$$

とし、正規分布による近似を用いて、有意水準 5% の両側検定を行います。

例題 14.6 母比率の仮説検定を行う

次の記述が適切である場合は○，適切でない場合は×をつけましょう．

- (1) 地域 A の標本比率の z 値は正なので，帰無仮説を棄却する．
- (2) 地域 A は，標本比率が 0.30 を上回っているため，帰無仮説を棄却する．
- (3) 地域 A の標本比率の z 値は， -1.96 から 1.96 の範囲にあるので，帰無仮説を棄却しない．

例題 14.6 母比率の仮説検定を行う

(4) 地域 B は、標本比率の z 値が棄却域に入るので、帰無仮説を棄却する。

(5) 「帰無仮説を棄却しない」とは、「母比率が 0.30 であると判断した」という意味である。

例題 14.6 母比率の仮説検定を行う

母比率の仮説検定

母比率の仮説検定では、帰無仮説で仮定した母比率を p_0 とすると、帰無仮説のもとで、標本比率 $\hat{p}$ の分布は、

$$\text{期待値 } p_0, \quad \text{標準偏差 } \sqrt{\frac{p_0(1 - p_0)}{n}}$$

となります。

例題 14.6 母比率の仮説検定を行う

母比率の仮説検定

したがって、母比率の仮説検定では、

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

を検定統計量として用います。

例題 14.6 母比率の仮説検定を行う

標本比率 $\hat{p}$ は、標本を取り直すたびにゆらぎます。

そのため、観測されたズレが、

「 $p = 0.30$ と仮定しても起こりやすいズレ」

なのか、

「 $p = 0.30$ では起こりにくいほど大きなズレ」

なのかを、 z 値で判断します。

例題 14.6 母比率の仮説検定を行う

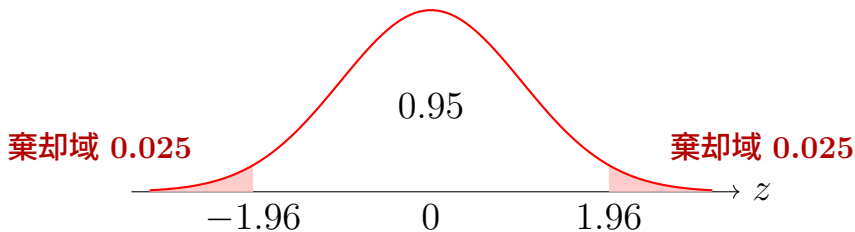
有意水準 5% の両側検定では、

$$|Z| \geq 1.96$$

つまり、

$$z \leq -1.96 \quad \text{または} \quad z \geq 1.96$$

が棄却域です。



例題 14.6 母比率の仮説検定を行う

地域 A の検定統計量

地域 A では,

$$\hat{p} = 0.33, \quad p_0 = 0.30, \quad n = 200$$

です. したがって,

$$z_A = \frac{0.33 - 0.30}{\sqrt{0.30 \times (1 - 0.30)/200}} = 0.926 \dots \approx 0.93$$

となります.

例題 14.6 母比率の仮説検定を行う

地域 B の検定統計量

地域 B では,

$$\hat{p} = 0.37, \quad p_0 = 0.30, \quad n = 200$$

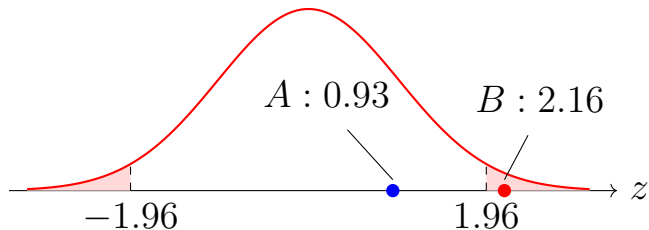
です。したがって,

$$z_B = \frac{0.37 - 0.30}{\sqrt{0.30 \times (1 - 0.30)/200}} = 2.160 \dots \approx 2.16$$

となります。

例題 14.6 母比率の仮説検定を行う

地域 A・地域 B の z 値の位置



地域 A は棄却域の外，地域 B は棄却域の中にあります。

例題 14.6 母比率の仮説検定を行う

(1) z 値が正なら棄却するか

地域 A の z 値は, $z_A = 0.93$ で, 確かに正です. しかし, 棄却するかどうかは z 値の符号ではなく,

棄却域に入っているか

で判断します.

0.93 は棄却域に入りません. したがって,



例題 14.6 母比率の仮説検定を行う

(2) 標本比率が 0.30 を上回れば棄却できるか

地域 A の標本比率は、 $\hat{p} = 0.33$ であり、0.30 を上回っています。

しかし、 $z_A = 0.93$ であり、棄却域には入りません。標本比率が 0.30 を上回っているだけでは、帰無仮説を棄却することはできません。したがって、



例題 14.6 母比率の仮説検定を行う

(3) 地域 A の z 値と棄却域

地域 A では、 $z_A = 0.93$ です.

$$-1.96 < 0.93 < 1.96$$

なので、 z_A は棄却域に入りません.

したがって、帰無仮説は棄却されません.



例題 14.6 母比率の仮説検定を行う

(4) 地域 B の z 値と棄却域

地域 B では、 $z_B = 2.16$ です.

$$2.16 > 1.96$$

なので、 z_B は棄却域に入ります.

したがって、帰無仮説を棄却し、対立仮説を支持します.



例題 14.6 母比率の仮説検定を行う

(5) 「帰無仮説を棄却しない」の意味

「帰無仮説を棄却しない」とは、

母比率が 0.30 であると判断したという意味ではありません。

今回のデータからは、帰無仮説を棄却するだけの十分な根拠が得られなかったという意味です。したがって、

