MAT220 第1講：統計学の概念

目次

これはMAT220 の第1講なので、私たちの目標は統計学の基本的な言語を確立することです。まず、統計学とは何か、そしてなぜ統計学が単なる公式の集まりではないのかを考えます。次に、データの種類について議論し、名義尺度、順序尺度、間隔尺度、比例尺度という四つの測定尺度を扱います。その後、標本抽出を導入し、いくつかの抽出方法を比較し、なぜ異なる方法が異なる結論につながるのかを説明します。最後に、データが私たちを誤解させる主な仕組み、つまり標本抽出の問題、非標本誤差、推論上の誤りについて議論します。

1　統計学とは何か？

統計学とは、データを数学的に研究する学問です。データとは、現実世界の何かから集められた情報だと考えることができます。統計学の目的は、この情報を収集し、整理し、分析し、解釈し、提示することで、よりよい意思決定ができるようにすることです。

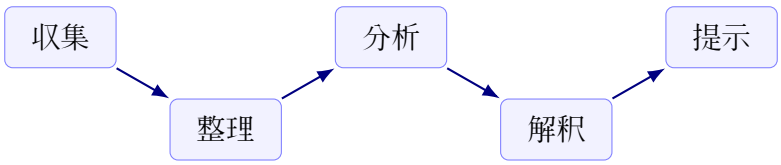
統計学は、科学、ビジネス、政府、医療、スポーツ、教育、技術など、あらゆる分野で使われています。誰かがデータを使って複雑な現実世界のシステムを理解しようとするとき、その人は統計的思考を使っています。

定義

統計学とは、データの研究であり、特にデータの収集、整理、分析、解釈、提示を扱う学問です。

1.1　統計研究の流れ

よい統計研究は、現実世界の文脈と数学的技法を組み合わせることで、あるシステムについて新しい情報を発見します。統計研究の基本的な流れの一つは次の通りです。



ここで：

- **収集**とは、現実世界のあるシステムを測定し、そこからデータを記録することです。
- **整理**とは、データを有用な形式に並べ替えることです。
- **分析**とは、適切な数学を用いてデータ集合の有用な特徴を記述することです。
- **解釈**とは、その結果が現実世界の文脈において何を意味するのかを説明することです。
- **提示**とは、結果を明確に、簡潔に、そして誠実に伝えることです。

上の流れは単なる数学的道具の集まりではないことに注意してください。むしろ、統計研究とは、数学的技法と現実世界の実践的な考慮を十分に理解したうえで組み合わせるものです。

1.2 なぜ統計学は退屈に感じられることがあるのか

統計学は公式に基づいているため、退屈で味気なく感じられることがあります。実際、統計学で使われる多くの公式の背後には深い数学理論がありますが、実用的な統計学ではその理論を直接使うことはあまりありません。数学的な洗練度という観点から見ると、統計学は非常に単純に感じられることがあります。というのも、よく使われる公式の多くは基本的な算術しか使わないからです。

実は、統計学はまさに同じ理由でとても面白いものでもあります。最も基本的な数学的操作から出発して、統計学はデータの全体的な構造を記述するための巧妙な公式を数多く作り出します。統計学の面白さは、これほど素朴な出発点から生まれたにもかかわらず、非常に強力で有用な道具であるという点にあります。統計学の実用的な重要性はいくら強調してもしすぎることはありません。世界は情報でできており、あなたや社会に関する多くの性質はデータに変換され、分析されうるのです。

1.3 記述統計と推測統計

統計学には二つの主要な柱があります。記述統計と推測統計です。

- **記述統計**とは、すでに持っているデータを記述する技術です。これには表、グラフ、平均、割合、ばらつき、その他の要約が含まれます。
- **推測統計**とは、少量のデータに基づいて、より大きな集団について結論を出す科学です。これには信頼区間、仮説検定、統計モデリングが含まれます。

例。 50 人の学生の通学時間を測定したとします。

- その50 人の学生の平均通学時間を計算するなら、私たちは**記述統計**を行っています。
- その50 人の学生を使って、大学全体の学生の平均通学時間を推定するなら、私たちは**推測統計**を行っています。

2 データの種類

2.1 データの取得

現実には、完全なデータ集合にアクセスできるとは限りません。むしろ通常は、自分たちが集めることのできるデータにしかアクセスできません。推測統計は、ある大きな集団についてすべてを完全に知らなくても、その集団について考えるための道具を作ります。これらは非常に強力で洞察に富んだ方法です。しかし、私たちのデータの質は、それを取得する方法の質に左右されます。そもそも悪いデータを集めてしまったなら、どれほど高度な数学を使っても分析には役立ちません。望ましくない影響が統計分析に偶然組み込まれてしまうことがあり、その結果、誤った結論が導かれることがあります。私たちの仕事は、こうした潜在的な影響を正しく特定し、できる限り小さくすることです。

重要な点

悪いデータは、通常、よい数学によって救うことはできません。データが混乱した方法、偏った方法、または不注意な方法で収集されたなら、数学的には正しくても、最終的な統計的結論は信頼できないかもしれません。

2.2 量的データとカテゴリーデータ

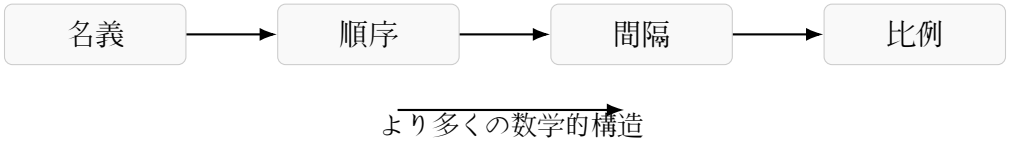
データは大まかに二つの広い種類に分けることができます。

- **量的データ**とは、数値データです。
- **カテゴリーデータ**とは、数値ではないデータです。

量的データには数値的構造があるため、多くの場合、算術演算を用いて分析できます。カテゴリーデータはラベル、名称、グループの一覧に近く、したがって扱える数学的構造ははるかに少なくなります。

2.3 測定尺度

現実世界で収集できるデータには多くの種類があります。したがって、どのような種類のデータにアクセスできるかによって、そのデータに対して特定の数学的操作を意味のある形で行える場合もあれば、行えない場合もあります。この潜在的に複雑な問題を整理する一つの方法が、Stevens が1946 年に導入した測定尺度です。測定尺度には四つの基本的な水準があります。



大まかに言えば、数学的構造が多いほど、データに対して実行できる操作も多くなります。ここから、それぞれの測定尺度を一つずつ確認します。

何が重要なのか？

測定尺度は、データについて直接かつ明確に説明する助けになります。自分たちが持っているデータの種類を分類することで、どのような数学的操作が意味を持ち、どのような主張が不適切なのかを聴衆に伝えることができます。

2.3.1 名義尺度

名義データは、内在的な順序を持たないラベルやカテゴリーから成ります。これはカテゴリーデータです。この測定尺度には数値的構造がないため、私たちにできるのは、二つの観測が同じか異なるかを判定し、その出現頻度を数えることだけです。しかし、使える所定の順序や算術はありません。

名義尺度

名義データは、自然な順序を持たない質的カテゴリーから成ります。例としては次のものがあります。

- 好きな色,
- 血液型,
- 国籍,
- 映画のジャンル,
- 通学に使う鉄道路線。

名義データについては、等しいかどうかを調べることと、頻度を数えることに意味があります。

2.3.2 順序尺度

“ordinal” という語は“order”、つまり順序という語に近い意味を持ちます。したがって、順序データとは、意味のある順序や順位を持つデータです。ただし、順位の間の差が等しい、または何らかの意味で測定可能であるとは仮定しません。

順序尺度

順序データは、意味のある順序を持つカテゴリーや値から成りますが、隣り合う値の間の距離は明確ではありません。
例：

- 競技者の順位,
- 痛みの尺度,
- 「非常に不満」「不満」「どちらでもない」「満足」「非常に満足」のような満足度の回答,
- 一年生、二年生、三年生、四年生のような学年。

順序データでは、中央値や百分位数はしばしば意味を持ちます。平均値も実務上使われることがありますが、隣り合うカテゴリーの間の距離が等しいとは限らないため、注意が必要です。

2.3.3 間隔尺度

間隔データには、意味のある順序と、値の間の等しい差があります。しかし、ゼロはその量が絶対的に存在しないことを表してはいません。そのため、「二倍多い」のような比率に関する主張は、一般には意味を持ちません。

間隔尺度

間隔データは、意味のある順序と意味のある差を持ちますが、本当のゼロを持ちません。
例：

- 摂氏で測った温度,
- 生年。

例えば、20°C は10°C の「二倍暑い」と物理的に意味のある形で言うことはできません。なぜなら、0°C は温度が存在しないことを意味しないからです。しかし、20°C と30°C の差は、30°C と40°C の差と同じ大きさです。

2.3.4 比例尺度

比例データには、順序、等しい間隔、そしてその量の不在を表す本当のゼロがあります。そのため、差も比も意味を持ちます。

比例尺度

比例データは、意味のある順序、意味のある差、そして本当のゼロを持ちます。
例：

- 体重,
- 身長,
- 長さ,
- 時間の長さ,
- 教室にいる学生の人数,
- 給与。

例えば、60 分の通学時間は、30 分の通学時間の本当に二倍の長さです。なぜなら、基準として使える絶対的なゼロ、つまり0 分が存在するからです。この比較が意味を持つのは、0 分が時間がまったく経過していないことを意味するからです。

2.3.5 測定尺度：まとめ

尺度	主な構造	可能な比較	例
名義	カテゴリーのみ	同じか異なるか	血液型
順序	順序を持つカテゴリー	大きいか小さいか	満足度評価
間隔	等しい差、本当のゼロなし	差	摂氏温度
比例	等しい差と本当のゼロ	差と比	身長

練習

それぞれの変数を、名義、順序、間隔、比例のいずれかに分類しなさい。

(a) 学生が最も好きなコンビニエンスストア。
(b) 1 から5 までの満足度評価。
(c) 摂氏で測った温度。
(d) キャンパスまで通学するのにかかる時間。

解答

(a) 名義。回答はカテゴリーのラベルだからです。
(b) 順序。回答には順序がありますが、評価の間隔が等しいとは限らないからです。
(c) 間隔。差には意味がありますが、摂氏ゼロ度は温度が存在しないことを本当に意味しないからです。
(d) 比例。時間の長さには本当のゼロがあり、比にも意味があるからです。

3 概念としての標本抽出

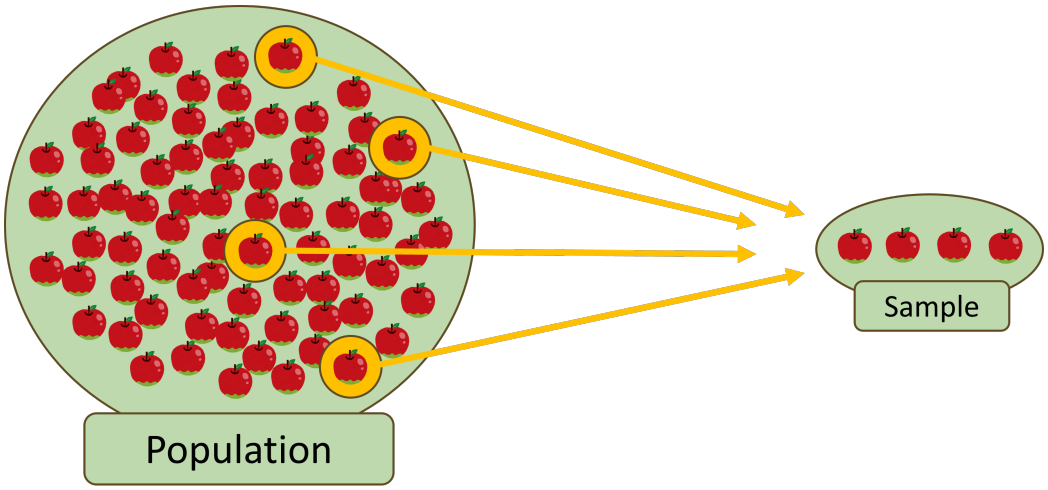
3.1 基本的な定義

すでに述べたように、現実世界を研究するとき、研究したい対象の集まりが大きすぎて、そのまま扱えないことが非常によくあります。そのため、実際にはその一部だけを取り出して、それを研究することが多くなります。この現実的で実践的な事情が、次の用語を動機づけます。

定義

- **母集団：** 知りたいと思っている人や物の完全な集団。
- **標本：** 母集団から選ばれ、実際にデータを集める小さな集団。
- **母集団パラメータ：** 母集団全体を記述する真の数値。正確には知らない場合もあります。
- **標本統計量：** 標本から計算される数値。多くの場合、母集団パラメータを推定するために使われます。

言い換えると、母集団とは、私たちが実際に研究したい集まりです。発見したい事実は母集団パラメータと呼ばれますが、母集団が大きすぎるため、これらを直接発見できないことがよくあります。そこで代わりに、全体の母集団から選ばれた小さな集団である部分集合を取ります。この小さな集団に対して、より簡単な計算を行うことができ、その結果として標本統計量と呼ばれる部分的な情報が得られます。標本の視覚的な例を示します。



3.2　例題：東京の平均給与

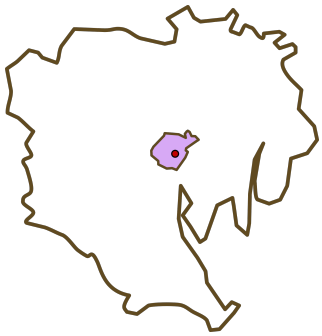
母集団から標本を得る過程を、母集団からの標本抽出と呼びます。標本抽出は統計学において非常に重要な部分であり、この授業の多くは標本と母集団全体の関係に充てられます。しかし、標本抽出を行うときには十分に注意しなければなりません。なぜなら、真の母集団を代表しない影響を偶然導入してしまう可能性があるからです。ここでは、この考えを例で説明します。

Jenny が東京都心の千代田区にある金融会社のインターンだとします。彼女は上司から、東京都民の平均給与を調べるように指示されました。そのため、Jenny はこれを調べるために統計研究を行う必要があります。Jenny は自治体で働いているわけではないので、東京都の母集団に関する完全な給与データにはアクセスできません。そこで彼女は、東京都民30 人の標本を取り、その人たちの平均給与を計算することにしました。

統計研究を行う際、Jenny は結果が誰に尋ねるかに大きく依存することに気づきます。例えば、標本は次のような要因によって異なります。

- 標本を取る時間帯,
- 調査を行う場所,
- 関係する人々の年齢,

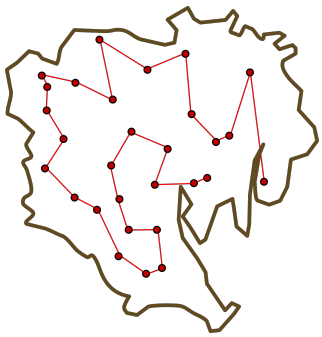
などです。以下に、Jenny が30 人の標本を取る可能性のある四つの方法と、それぞれに内在する問題を示します。



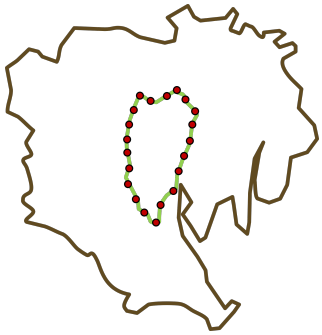
方法1： 千代田区にある自分のオフィスから30 人の参加者を選ぶ。

多様性の欠如： 自分の近所の人々から標本を取るのは確かに便利です。しかし、その数値は東京全体を代表しないかもしれません。地域によって、豊かさや貧しさ、年齢層、仕事の種類が異なるからです。したがって、たとえ30 人に尋ねたとしても、東京全体ではなく、その地域の特徴を偶然測定してしまう可能性があります。千代田区の場合、この標本は平均給与を過大評価しやすいでしょう。

方法2： 東京のどこからでも30 人の参加者をランダムに選ぶ。



実行上の困難： 都内全体から人々をランダムに抽出することは、偏りがないように感じられるかもしれません。しかし、実際に行うには費用も時間もかかります。ランダムに選ばれたさまざまな地域へ移動することが難しいだけでなく、誰を代表者として尋ねるかによって、他の潜在的な偏りも生じます。それでも、紙の上では偏りがないように見える計画でも、資源の制約によって実際には失敗することがあります。

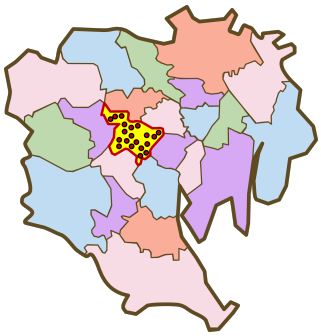


方法3： 山手線に乗り、30 駅それぞれで降りて、各駅で一人に尋ねる。

アクセス可能性の欠如： 「一駅につき一人」を抽出すると、東京に住んでいない観光客や外部の人が含まれる可能性があります。また、観光地の駅、通勤の拠点、ビジネス街に偏り、特定の種類の労働者に標本が偏る可能性もあります。別の言い方をすれば、この標本は特定の種類の東京在住者、つまり山手線沿線で働いたり移動したりする人々に偏る可能性があります。

方法4： 東京23 区のうち一つをランダムに選び、その区内から30 人の参加者をランダムに選ぶ。

多様性の欠如： 東京全体を移動するよりは便利であり、Jenny の地元である千代田区だけで標本を取るよりは偏りが少ないとはいえ、この種の標本にはそれでも多様性の欠如があります。ランダム性は、小さな標本が母集団の多様性を完全に代表することを保証しません。むしろ、それは選択の第一段階をランダムにするだけです。



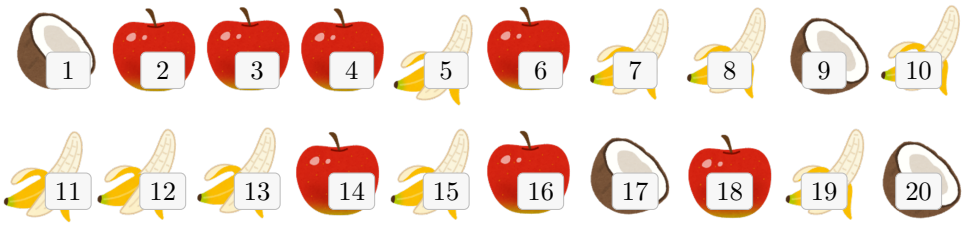
では、この場合の最良の標本抽出方法は何でしょうか。実際には、それはJenny が何をしようとしているかによります。実務上、唯一の「最良の方法」が存在することはあまりありません。むしろ、標本抽出はしばしば資源と誤差の間のバランスを取る作業であり、ときには妥協しなければなりません。理想的には、これらの偏りは研究を行う研究者によって理解されているべきであり、結果を伝える際にはできるだけ明確に示されるべきです。

実践における標本抽出

標本抽出は、しばしば資源と誤差の間でバランスを取る作業です。最も正確なデータ収集方法は、通常、より高価で、より難しく、より時間がかかります。ときには妥協が必要ですが、結果を提示するときにはその妥協を明確にすべきです。

4　標本抽出の方法

前の例のJenny に動機づけられて、ここからは一般的な標本抽出方法をいくつか確認します。この節では、母集団が20 個の果物から成ると仮定します。内訳は、7 個のリンゴ、9 本のバナナ、4 個のココナッツです。



私たちの目標は、異なる標本抽出方法を使って、この母集団から大きさ5 の標本を選ぶことです。

4.1　単純無作為標本

定義：単純無作為標本

単純無作為標本とは、与えられた大きさの可能な標本すべてが等しい確率で選ばれるようにして得られる標本です。

例えば、私たちの果物の母集団から、大きさ5 の単純無作為標本の一つは次のようになります。

標本 5 9 10 12 18

別の大きさ5 の単純無作為標本は次のようになります。

標本 2 4 8 14 17

単純無作為標本では、標本にリンゴ、バナナ、ココナッツがいくつ入るかを意図的に制御しません。選択規則としては公平ですが、偶然によって標本が不均衡になることはあります。

4.2　系統標本

定義：系統標本

系統標本とは、ランダムな開始点を選び、その後、母集団のk 番目ごとのメンバーを選ぶことで得られる標本です。

例えば、1 から始めて4 番目ごとに選ぶと、次の標本が得られます。

標本 1 5 9 13 17

3 から始めて4 番目ごとに選ぶと、次の標本が得られます。

標本 3 7 11 15 19

系統抽出は、特に母集団がすでにリストとして並んでいる場合に便利です。しかし、そのリスト自体に繰り返しのパターンがある場合には危険です。例えば、工場のリストで4 番目ごとの品物がすべて同じ機械から来ているなら、4 番目ごとに選ぶことで、偶然にも一つの機械だけを標本にしてしまう可能性があります。

4.3　層化標本

定義：層化標本

層化標本とは、まず関連する特徴に基づいて母集団を**層**と呼ばれるグループに分け、その後、それぞれの層から無作為に抽出することで得られる標本です。

私たちの果物の例では、自然な層は3 種類の果物です。

$$\text{リンゴ}, \qquad \text{バナナ}, \qquad \text{ココナッツ}.$$

層化標本では、いくつかのリンゴ、いくつかのバナナ、いくつかのココナッツを意図的に選び、標本が三つすべての果物の種類を代表するようにすることがあります。

大きさ5 の層化標本の例：

標本 2 6 9 14 19

ここでは、それぞれの果物の種類から標本を取っています。目標は単なるランダム性ではなく、重要なグループ全体にわたる代表性です。

層化抽出は、母集団に重要な部分集団があり、それぞれの部分集団が標本に含まれることを確実にしたい場合に有用です。

4.4　クラスター標本

定義：クラスター標本

クラスター標本とは、母集団を通常は自然な、または便利なまとまりに基づくクラスターに分け、その後、一つ以上のクラスター全体を無作為に選ぶことで得られる標本です。

例えば、果物のリストが大きさ5 のクラスターに分けられているとします。

$$\{1,2,3,4,5\}, \quad \{6,7,8,9,10\}, \quad \{11,12,13,14,15\}, \quad \{16,17,18,19,20\}.$$

このとき第二クラスターを無作為に選ぶと、標本は次のようになります。

標本 6 7 8 9 10

第三クラスターを無作為に選ぶと、標本は次のようになります。

標本 11 12 13 14 15

クラスター抽出は、母集団全体に散らばった個人を標本にすることが高価または困難な場合に有用です。例えば、研究者はいくつかの教室を無作為に選び、その教室のすべての学生に調査を行うかもしれません。

4.5　よくある混同：層化抽出とクラスター抽出

層化抽出とクラスター抽出は、どちらも母集団をグループに分けるため、似ているように感じられるかもしれません。違いは、そのグループが果たす役割にあります。

方法	グループの作り方	標本の選び方
層化抽出	果物の種類、性別、学年、区のような重要な特徴を用いてグループを作る。	すべての層から無作為に抽出する。
クラスター抽出	教室、地域、リストの区画のような、自然または便利なまとまりであることが多い。	クラスター全体を無作為に選ぶ。

要するに、**層化抽出**ではすべてのグループから標本を取りますが、**クラスター抽出**ではグループ全体そのものを標本として取ります。

4.6　その他の標本抽出方法

標本抽出方法は他にもたくさんあります。実務上役に立つものもありますが、多くの場合、重大な限界も伴います。

方法	基本的な考え方
便宜抽出	最も接触しやすい人を標本にする。
自発的回答標本	人々が参加するかどうかを自分で選ぶ。
パネル抽出	同じ集団を時間を追って追跡する。
場当たり的抽出	明確な無作為規則なしに、非形式的または不注意な方法で標本を取る。
スノーボール抽出	既存の参加者が将来の参加者の募集を助ける。
極端事例抽出	最小値や最大値のような極端な事例を意図的に標本にする。

これらの方法は自動的に間違っているわけではありませんが、慎重に解釈しなければなりません。例えば、便宜標本は安く早く得られるかもしれませんが、それが教えてくれるのは、母集団全体についてというよりも、誰が接触しやすかったかについてかもしれません。

4.7　Jenny の東京研究に戻る

ここで、Jenny の方法を標本抽出の用語で説明できます。

- 方法1：**千代田区のオフィス近くで30 人に尋ねる。** これは便宜標本に最も近いです。主な懸念は、千代田区が東京を代表していないかもしれないことです。

- 方法2：**東京から30 人を無作為に標本抽出する。** これは単純無作為標本に最も近いです。主な懸念は、原理的には正確ですが、実践上は難しいことです。
- 方法3：**山手線の各駅で一人に尋ねる。** これは経路に基づく便宜標本に最も近いです。主な懸念は、住民ではなく通勤者や観光客を標本にしてしまう可能性があることです。
- 方法4：**一つの区を無作為に選び、その中で標本を取る。** これはクラスター型の標本に最も近いです。主な懸念は、一つの区がすべての区を代表しないかもしれないことです。

この語彙だけで実践上の問題が解決するわけではありませんが、研究の強みと弱みを明確に説明する助けになります。

5　データが私たちを誤解させるとき

統計的結論が失敗する主な理由は三つあります。

統計的結論が失敗する三つの方法

1. **標本抽出の問題：** 標本が母集団を正直に代表していない。
2. **非標本誤差：** 標本がうまく選ばれていても、データそのものが歪んでいる。
3. **推論上の誤り：** 結論が、データが実際に支持する範囲を超えている。

よい統計研究は、この三つすべてに対処しなければなりません。標本が大きいだけでは十分ではなく、正しい公式を使うだけでも十分ではありません。推論の連鎖全体が信頼できるものでなければなりません。

5.1　標本誤差と標本抽出バイアス

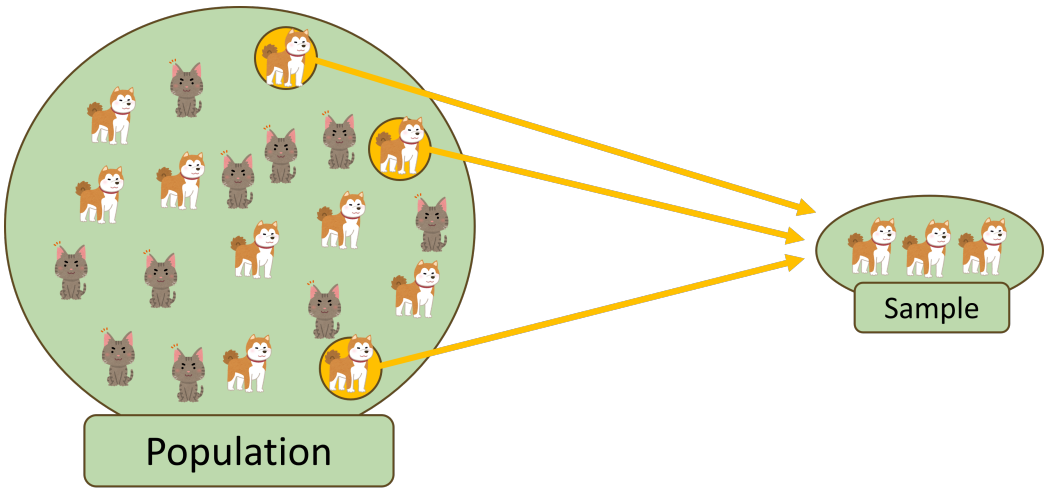
標本抽出の問題には、重要な二つの種類があります。

標本誤差と標本抽出バイアス

- **標本誤差**とは、標本と母集団の間に生じるランダムな不一致です。抽出方法が公平であっても起こることがあります。
- **標本抽出バイアス**とは、標本抽出方法における体系的な不公平です。方法そのものが、母集団の一部を含まれやすくしたり、含まれにくくしたりします。

5.1.1　例：標本誤差

母集団には猫と犬が同じ数だけ含まれているとします。大きさ3 の単純無作為標本を取り、偶然、選ばれた3 匹がすべて犬だったとします。



これは標本誤差です。方法は公平でしたが、その特定の標本が運悪く偏っていたからです。

5.1.2 例：標本抽出バイアス

Jenny が千代田区のような裕福な地域の人々にだけ給与について尋ねるとします。これは単なる不運ではありません。抽出方法そのものが体系的に不公平です。なぜなら、裕福な人々を過剰に代表しているからです。

これは標本抽出バイアスです。Jenny が同じ方法を何度も繰り返すと、東京の平均給与を過大評価する傾向のある標本を取り続けることになります。

5.2 非標本誤差

標本がうまく選ばれていても、データそのものが歪むことがあります。このような歪みを**非標本誤差**と呼びます。

定義：非標本誤差

非標本誤差とは、標本をランダムに選ぶ行為によって生じたものではない、統計研究における誤差です。

よくある非標本誤差には次のものがあります。

- 不明確または誘導的な質問，
- 不正確な測定，
- 不誠実な回答，
- 回答の欠落，
- 記録の誤り，
- データ入力の誤り，
- 故障した機器，

などです。重要なのは、標本抽出方法自体が妥当であっても、これらの誤差が起こりうるという点です。

5.3　例：偏った調査質問

東京の通勤者に調査を行い、次のように尋ねたとします。

「今朝の電車はどれほどひどくて最悪でしたか？」

これは悪い質問です。なぜなら、その電車移動がひどかった、または最悪だったということを前提にしており、回答者を否定的な回答へ誘導しているからです。よりよい言い方は次のようになります。

「今朝の電車での移動はどうでしたか？」

非常に悪い	悪い	普通	良い	非常に良い

この表現はずっと中立的であり、回答者に特定の価値判断を促していません。

5.4　結論が偏るよくある方法

一部のバイアスは単なる技術的誤りではありません。むしろ、証拠を選び、記憶し、解釈する方法における失敗である場合があります。この現象のよくある例を三つ挙げます。

生存者バイアス

生存者バイアスは、ある選別過程を生き残った例だけを見るときに起こります。
例：「昔の建築はとても美しかったのに、現代の建物はとても醜い。」
これは誤解を招く可能性があります。なぜなら、多くの醜い古い建物は取り壊され、美しい古い建物は保存されたからです。現存する建物は、過去の建物のランダムな標本ではありません。

選択バイアス

選択バイアスは、データが特定の結論に有利になるような形で選ばれるときに起こります。
例：「歌舞伎町の人々に、東京は騒がしく混沌としているか尋ねた。彼らはそうだと言った。したがって、東京は騒がしく混沌としている。」
歌舞伎町は東京を代表する標本ではありません。この結論は、東京全体についてというよりも、歌舞伎町について多くを語っているのかもしれません。

資金提供バイアスまたは利益相反

資金提供バイアスは、ある特定の結論から利益を得る団体によって研究が資金提供される場合に起こることがあります。
例えば、ある業界が自分たちの製品が有害かどうかに関する研究に資金を出している場合、研究設計、データ、分析、解釈に注意深く目を向ける必要があります。

5.5　推論上の誤り

標本が妥当で、データが正確に記録されていても、最終的な結論が間違っていることがあります。これは、推論が証拠によって支持される範囲を超えてしまうときに起こります。

よくある推論上の誤りには次のものがあります。

- 小さな標本から巨大な母集団へ一般化すること，
- 相関を因果関係として扱うこと，
- 潜伏変数を無視すること，
- 交絡変数を無視すること，
- 統計量を正しく使っているが、それを誤って解釈すること，
- 証拠が正当化するよりも強い結論を出すこと。

5.6　相関と因果関係

二つの変数が関係していても、一方が他方の原因であるとは限りません。二つのことが一緒に起こるからといって、自動的に一方が他方の原因であるとは言えません。

例。 東京の非常に暑い日に、コンビニエンスストアでより多くのペットボトルの水が売れ、同時に熱中症を報告する人も増えるとします。
ペットボトルの水を買うことが熱中症の原因だと結論するのは愚かです。より合理的な説明は、暑い天気が両方の変数に影響しているというものです。

この例では、ペットボトルの水の売上と熱中症は相関していますが、因果的な説明には別の変数、つまり気温が関わっています。

5.7　統計的主張を信頼するためのチェックリスト

統計的結論を信頼する前に、次の質問をしましょう。

1. **母集団：** その主張は本当はどの母集団について述べているのか？
2. **標本：** 実際に標本にされたのは誰、または何か？
3. **標本抽出方法：** 標本のメンバーはどのように選ばれたのか？
4. **データの質：** 測定、質問、記録は信頼できるものだったか？
5. **欠測データ：** 誰が回答しなかったのか。そして、それは重要かもしれないか？
6. **変数：** 潜伏変数や交絡変数は存在するか？
7. **結論：** 結論はデータから従っているのか、それとも言い過ぎているのか？

まとめ

統計的主張の強さは、その背後にある全体の連鎖の強さに依存します。

$$\text{収集} \longrightarrow \text{整理} \longrightarrow \text{分析} \longrightarrow \text{解釈} \longrightarrow \text{提示}.$$

連鎖の一部が弱ければ、最終的な結論も弱くなる可能性があります。