

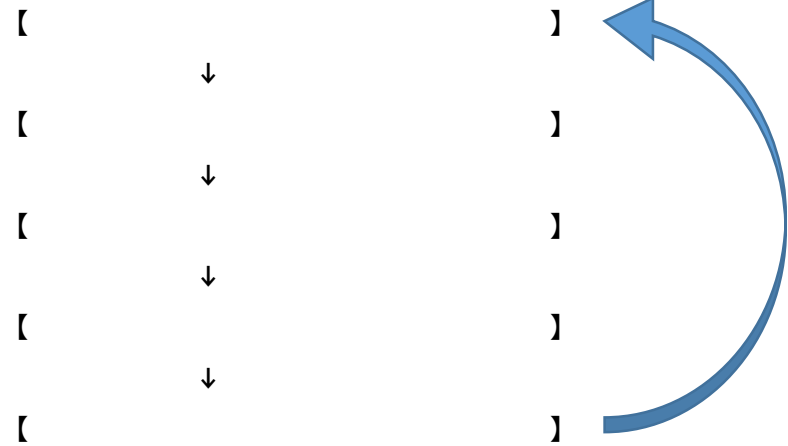
統計とは？

（１）統計とは

たくさんのデータ傾向について客観的に判断するための科学的な道具であり、データを集めて全体を把握・推定すること。統計は天気やお店の売り上げ、商圈分析、疫学、生態調査など様々な場面で使われます。

①PPDAC サイクルと課題研究

PPDAC サイクルとは？



あくまでもデータ分析は【 】ではなく、問題解決までの【 】です。実際に問題を発見し、解決までの計画があってこそそのデータ収集・分析です。この手順で必要になってくるのが統計学です。

②「エビデンス」って何？

「エビデンス」とは「根拠」のことで主観や直感とは違って、客観的な事実に対して使われる用語です。課題研究においても、もちろん大切です。データや客観的事実で示す根拠をもとに論文作成ができるように、正しいデータの集め方・分析の仕方・表し方を心得ておきましょう。

（２）全体を把握する

①記述統計

まず、統計の役割は主に2つあります。**1つ目が全体を把握する「記述統計」**です。つまり、データを集めて、表やグラフでデータの性質をまとめることです。つまり、調べた数値から、集団の性質や特徴を明らかにしていく統計です。

数値の大きい順や、年齢順、意図を持って並べたり、グラフを使ってデータの特徴がひと目でわかるようにしたりすることが大切です！

②平均とばらつき

「見える化」されたデータからさまざまな性質や特徴を読み取って分析していくことが記述統計の基本です。ここで知っておきたいのが「平均」と「ばらつき」です。

「平均」…集団を代表する中心的な値を【 】と言います。ですが、平均値＝代表値になるわけではありません。集めたデータの中で1つだけでも大きく外れる【 】があると平均値はこれに強く影響を受けてしまいます。また、集団の分布のちょうど中央に来る【 】、最も多く出現している値【 】についても頭のすみに置いておいてください。中央値はデータを並べて二等分した境界値になります。

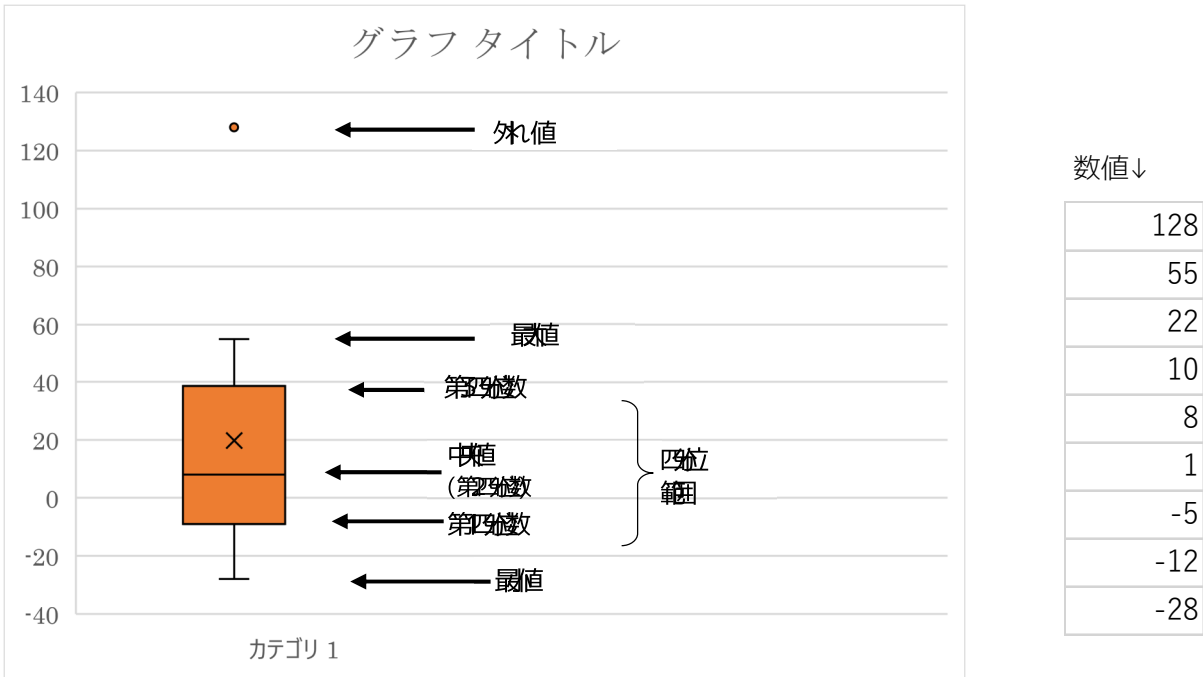
「ばらつき」…集めたデータが平均に集中しているのか、広く散らばっているのかは平均だけ見てもわかりません。データがどれくらい散らばっているかを調べる指標が【 】で、それを数値化したものが【 】です。

③データを整理する

すでにグラフについては学習しましたが、どんなグラフがあったか覚えていますか？

今回はすでに数学のデータの分析でも学習した**箱ひげ図 Boxplot**について用語をおさらいしておきます。

箱ひげ図とはデータ値のばらつきを見るためのグラフです。データの分布を「箱」と「ひげ」で表したグラフで、データがどのあたりの値に集中しているかをひと目でとらえることができます。



（＊上の図は用語解説のためのものでなんの数値でもありません）

×は平均値を表します。箱の部分は四分位数（データを4等分したときの区切りの値）、ひげの両端は最大値と最小値を表します。

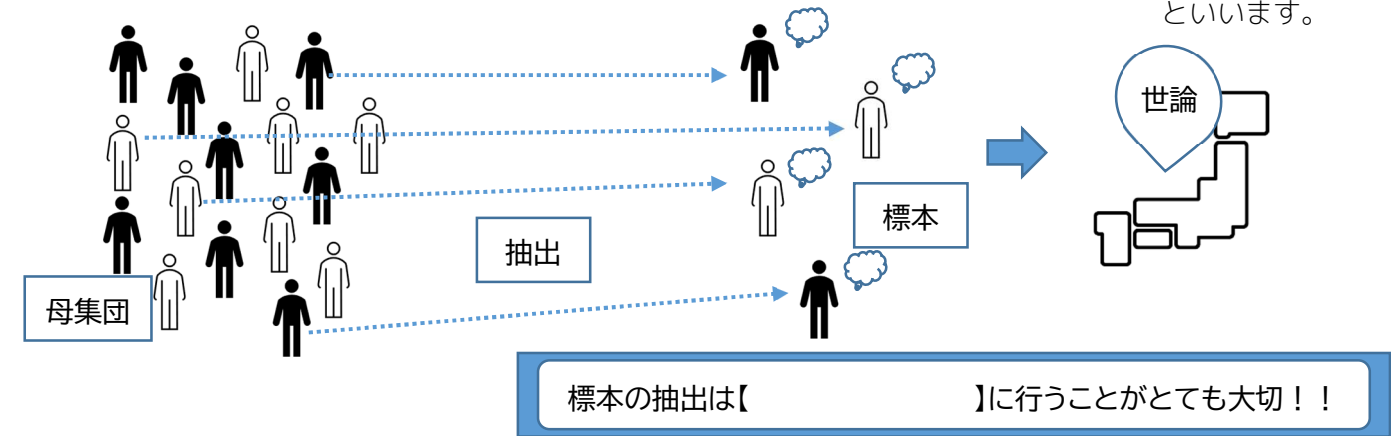
(3) 全体を推定する

①推測統計

統計の役割の2つ目が「推測統計」です。ある集団の一部のデータを集めて、集団全体を推定する統計のことです。

例えば…世論調査をするときに、日本に住むすべての人に意見を訪ねて回することは不可能ではないかもしれませんが、現実的ではありません。こんなときに、一部の人【 】の意見を聞いて集計し、それを日本人全体【 】の意見として推定し、世論として発表しています。これを【 】

といいます。



対象となる集団をすべて調べる【 】に比べて時間と費用は減らせるが、どうしても【 】は出てしまいます。その誤差についてはまた後程…

②標本の落とし穴

先ほどの例の続きで…世論調査で「内閣の支持率」を調査するとき。どのような抽出方法が良いと思いますか？例えば、駅前で街頭アンケートを行った。これはどうでしょうか？考えてみてください。

標本（サンプル）の抽出で大切なのは、標本が偏らないことです。世論調査で言えば、それが日本国民という母集団の縮図になっているかということ。街頭アンケートとは【 】のひとつです。つまり、その街やその駅を利用する人々の意見や動向を知るための調査です。

考えてみよう！

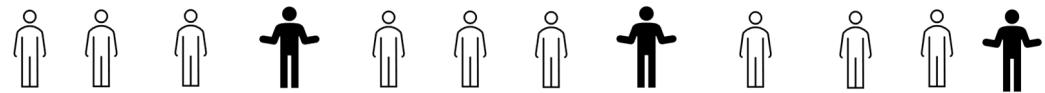
【問い】次の調査には、無作為抽出という観点から見て、どんな問題があるでしょうか？

- ア．自分の開発したソフトウェアの使い勝手のよさを調べるために、同じ研究室のメンバーに頼んでソフトを実際に使ってもらいアンケートをとった。
- イ．血液型性格診断のためのデータ（血液型と性格特性）を、血液型性格診断のベストセラー書籍の巻末にあるアンケート1万人以上から集めた。
- ウ．スマホの使用頻度についてのアンケートをたくさんの人からとりたいので、SNS上でアンケートを拡散した。
- エ．高校生の環境に対する意識と行動について調べるために、刈谷高校から無作為に3クラス選んでアンケートをとった。

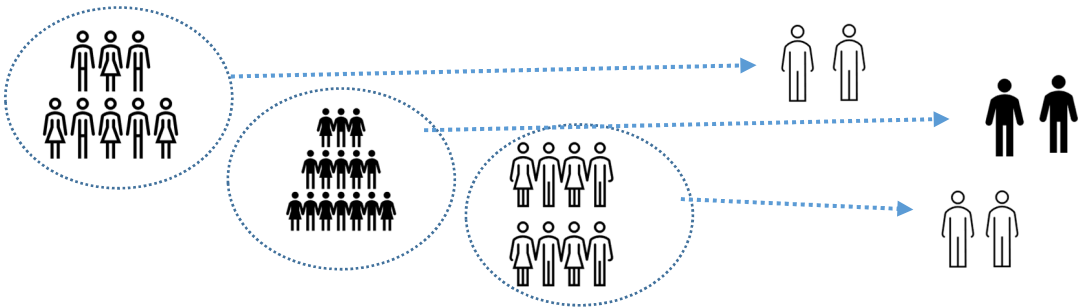
③標本の選び方

標本には偏りがないように。ランダムでないと間違った憶測につながってしまいます。

【 】完全にランダムに、乱数や〇〇人目ごとになどで抽出する方法



【 】母集団を地域や属性などの層で分け、層ごとに標本を抽出する方法



母集団によっては数こそ小さいものの全体への影響が無視できない重要な階層が含まれていることもあります。単純な無作為だとその小さな階層から入らない可能性があり、データが大きく左右されてしまいます。そんなときに仕えるのがこの方法です。

！統計データにだまされない・だましてしまわないように！

参考 佐々木弾『イラスト＆図解 知識ゼロでも楽しく読める！統計学のしくみ』（西東社）