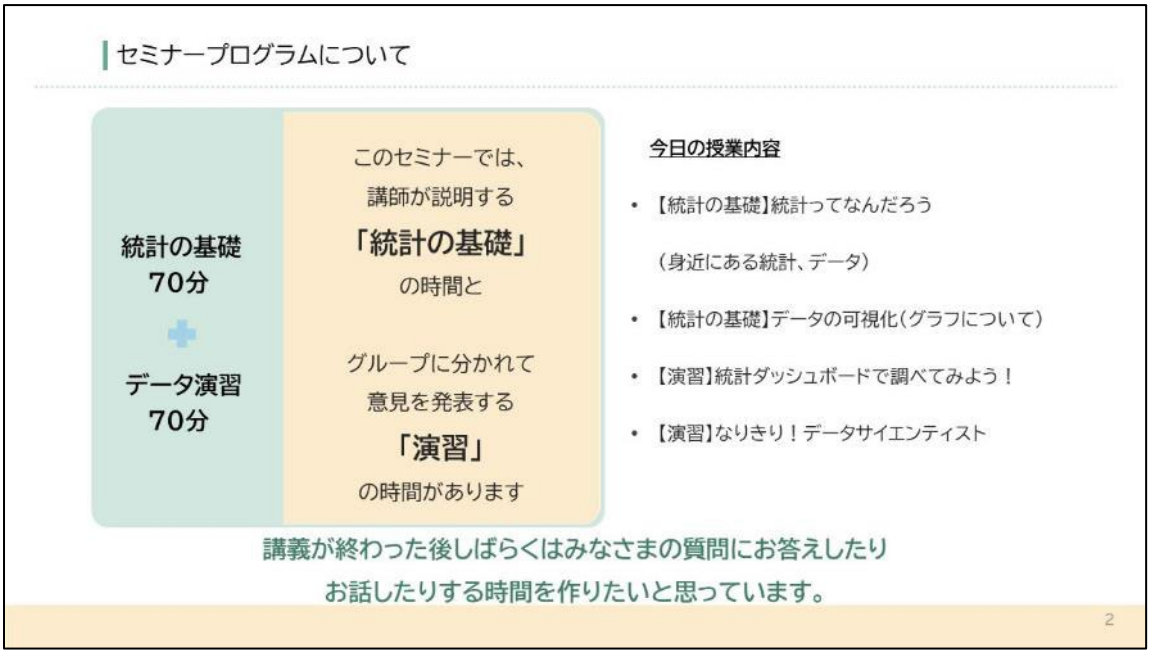
中学生限定！統計チャレンジセミナー
解説書

主催

総務省統計局・和歌山県


中学生限定
統計チャレンジセミナー
2023
主　催 ｜ 総務省統計局統計データ利活用センター・和歌山県データ利活用推進センター　運営 (委託) ｜ 株式会社Rejoui



【スライドの解説】
このセミナーは、「統計の基礎」を学び、学んだ内容を実際に試してみる「演習」を行います。「統計の基礎」では、統計とデータとその特徴を表す方法である統計について学びます。「演習」では、総務省統計局が提供している「統計ダッシュボード」で、さまざまなデータをグラフとして表します。また、データサイエンス体験として課題に挑戦します。

【このページのポイント】
このセミナーは、「統計の基礎」を学び、学んだ内容を実際に試してみる「演習」を行います。

【キーワード】
統計、データ、データの可視化

セミナーの撮影・動画配信のお知らせ

この講座の様子は録画および動画配信を行います

Zoomの表示名は、ニックネーム＋（お住まいの都道府県）にて
ご登録ください　【例】タロウ（和歌山県）





総務省統計局

総務省統計局の
HPにて後日動画配信

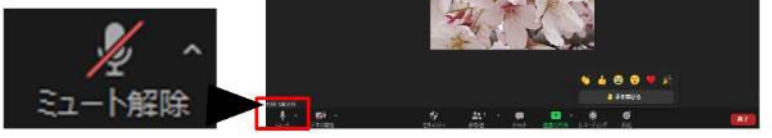
・この講座は、当日の様子を録画し、後日、総務省統計局のサイトで動画配信を行う予定です。あらかじめご了承ください。
・Zoom接続時の表示名は、本名フルネームを避けてニックネームにてご登録ください。

3

オンラインでご参加の皆さまへ

先生が話しているときは、ミュート(消音)の設定をお願いします。

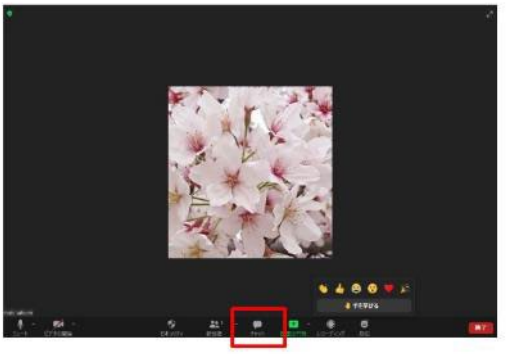
画面左下にある
マイクのボタンを押すと
ミュートにできるよ！



4

オンラインでご参加の皆さまへ

Zoomのチャット機能を使って、たくさん話しかけてください。



Zoom画面のチャットを押すと
チャット欄が表示されます。
是非、お気軽にお声がけください！



5

先生について

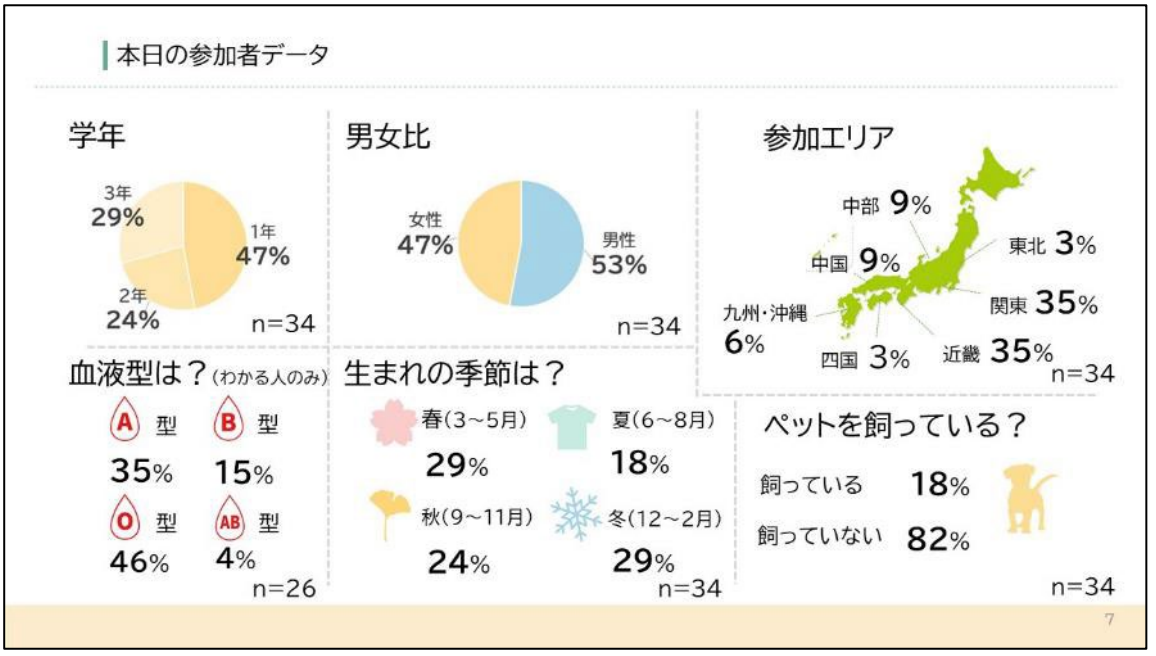


菅 由紀子(かん ゆきこ)先生



- 株式会社Rejoui(リジョウイ)代表取締役
- 広島大学 客員教授
- データサイエンティスト協会 スキル定義委員
- WiDS Ambassador

2016年9月に㈱Rejouiという会社を作り、データサイエンティストの育成や学習カリキュラム・教材等を開発するお仕事をしています。
これまでデータサイエンスに関する学習講座やイベントをたくさん開催しており、小学生へ向けた「データサイエンスforキッズ」(自社開催)や「わくわく！統計アカデミー」(主催:総務省統計局)のほか、「機械学習・データサイエンス入門」(主催:総務省統計局統計研究研修所)では、高校の先生にデータサイエンスを教える先生として講師を務めました。子どもから大人までたくさんの人たちにデータサイエンスの楽しさを伝えています。



【スライドの解説】
■回答者数：34名。約半分が、中学1年生、残りが中学2年生と3年生の割合。
■男女比：ほぼ同じ割合。
■参加エリア：関東と近畿が35%で多い。
■血液型：O型が46%、A型が35%、合わせて約80%。
■生まれの季節：ほぼ同じ割合で春夏秋冬に分かれた。
■ペットを飼っているかどうか：18%と少ない割合。

これらの結果は、今回セミナーに参加される皆さんに、事前にアンケートに回答いただき、回答データを統計で処理しグラフにしたものです。

【このページのポイント】
参加している生徒さんのイメージをつかむため、グラフにして共有する

【キーワード】
参加している生徒、学年、男女比、参加エリア

グループワーク(演習)について

セミナーの途中で、グループに分かれて自己紹介をしたり、自分が考えたことをなかまに話す
やりとりがあります。会場の皆さんもZoomで参加されている皆さんも、
先生やメンターの合図にしたがって、取り組んでくださいね。



8

第一章

統計ってなんだろう？





【スライドの解説】

■データについて

私たちは、興味のあるものを観察したり測定したりすることで、たくさんの情報を集めることができます。例えば、「毎日の体温を測定する」、「学校給食についてのアンケート調査を行う」といったことでたくさんの情報を得ることができます。これら情報を「データ」と言います。

■統計とは

「統計」とは、得たデータの特徴をわかりやすく数値で表す方法です。

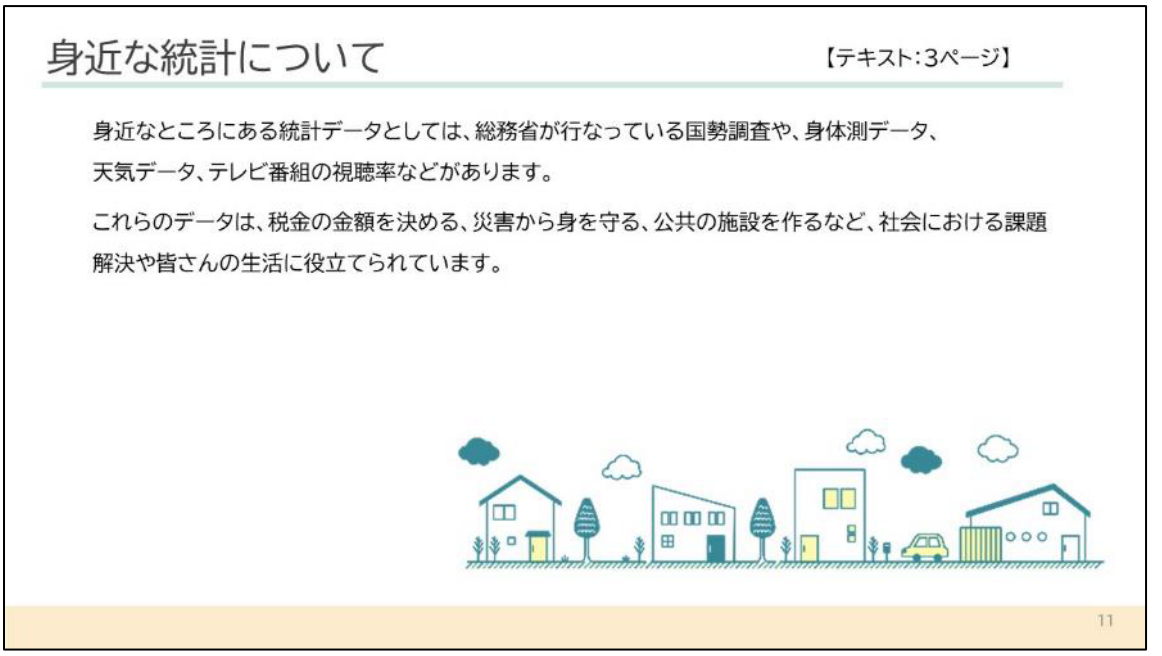
得られたデータは、さまざまな統計の方法で全体の特徴を推測したり、比較したり、未来を予測したりすることもできます。

【このページのポイント】

統計とは、AIを形作る技術の1つであり、未来に役立つ技術の基本となる学問です。
データを用いることで、勘や経験だけで決めるよりも確実な予測を立てられるようになります。
長い期間にわたってデータを蓄積することは、未来の予測の精度を上げるためにも非常に重要です。

【キーワード】

データ、統計、AI



【スライドの解説】

■身近なところにある統計データ

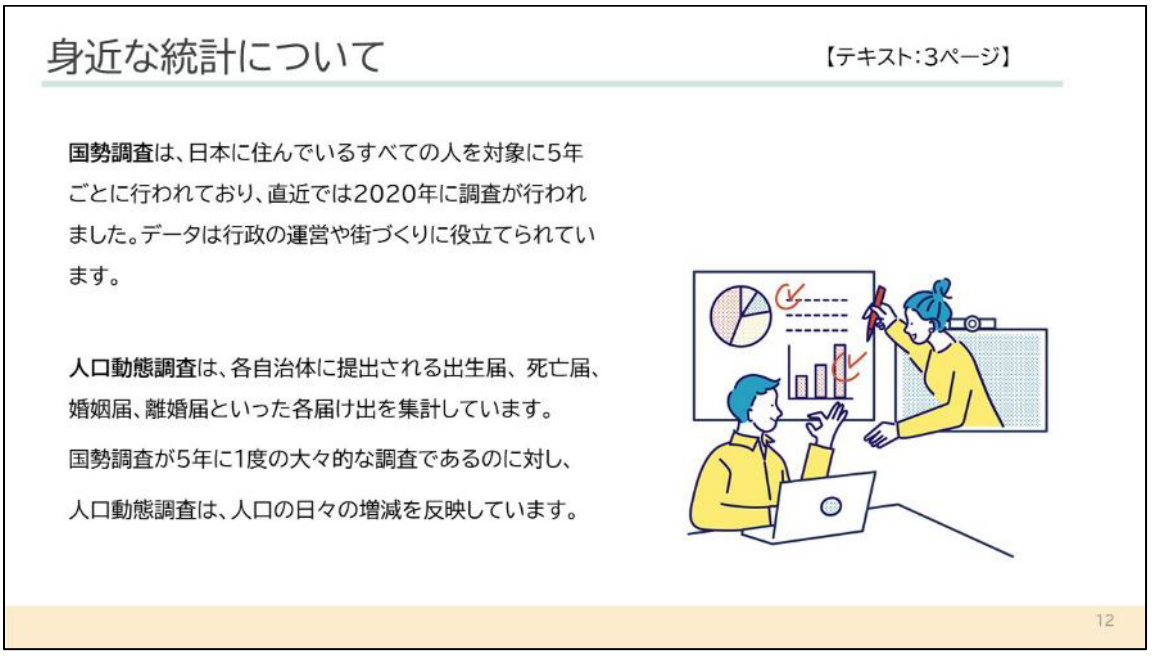
総務省が行なっている国勢調査や、厚生労働省が行っている人口動態調査などの人口統計、身体測定データ、気象データ、テレビ番組の視聴率などがあります。これらのデータは、税金の金額を決める、災害から身を守る、公共の施設を作るなど、社会における課題解決や皆さんの生活に役立てられています。

【このページのポイント】

総務省や厚生労働省などが調査し集められたデータは統計を用いて計算し、社会における課題解決や皆さんの生活に役立てられています。

【キーワード】

国勢調査、人口統計、身体測定データ、気象データ、テレビ番組の視聴率



【スライドの解説】

■国勢調査

5年ごとに行われており、直近では2020年に調査が行われました。日本に住んでいるすべての人を対象に、性別・年齢・家族の人数・住居の種類・通っている学校や仕事のことなどを調査しています。得られたデータは行政の運営や街づくりに役立てられています。

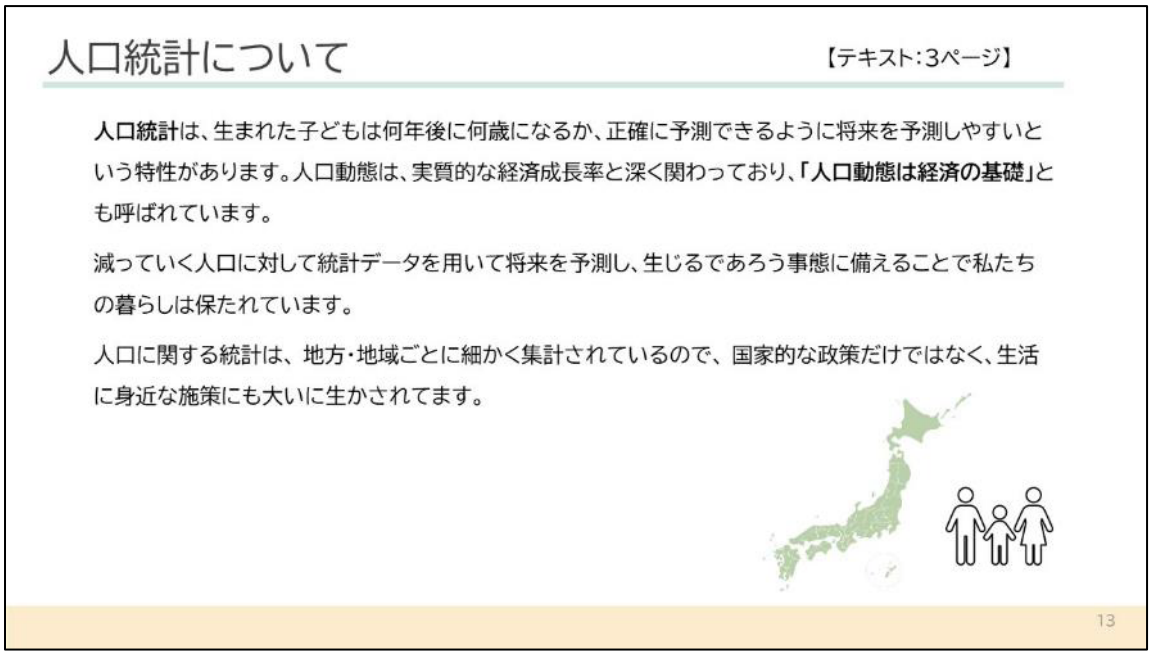
■人口動態調査

各自治体に提出される出生届、 死亡届、婚姻届、離婚届といった各届け出を集計し、毎月公表しています。

国勢調査が5年に1度の大々的な調査であるのに対し、人口動態調査は人口の日々の増減を反映しています。

【このページのポイント】
国勢調査は5年に1度の大々的な調査です。**人口動態調査**は、毎月公表されています。

【キーワード】
国勢調査、人口動態調査



【スライドの解説】

■人口動態調査による人口統計

・特徴

① 得られた人口データから将来の人口を予測することができます。
2020年に生まれた子どもは何年後に何歳になるかを正確に把握できます。
そのため、数年後の10歳の子どもの人口が予測できます。

② 人口動態は、景気動向や経済成長の目安となる指標(実質的な経済成長率)と
深く関わっており、「人口動態は経済の基礎」とも呼ばれています。

日本は少子化の影響で労働人口は年々減少しており、少ない労働人口で経済力を維持するために1人当たりの生産性を高める必要があります。

また、社会の無駄をできるだけ少なくして、豊かさが広く人々に行きわたるようにする努力も求められます。減少する人口に対してデータと統計を用いて将来を予測し、今後生じるであろう事態に備えることで私たちの暮らしは保たれています。

人口に関する統計は、 地方・地域ごとに細かく集計されており、 国家的な政策だけではなく、生活に身近な施策にも大いにいかされています。

【このページのポイント】
人口動態調査による人口統計がほかの統計と異なり、将来の人口を予測することができます。

【キーワード】
人口動態調査、人口統計、人口動態は経済の基礎、地方自治体の施策



【スライドの解説】

■スポーツ統計

例：野球、サッカー、バスケットボール、アイスホッケーなど

活用方法：戦術、対戦相手の評価、選手起用の判断に役立てられています。

軍事用のトラッキング（追跡システム）を活用し、様々なデータを収集し高度な分析がなされています。

データを収集し統計を用いて分析することで、感覚ではわからない価値を発見したり、人の感覚を裏付けたりすることができます。

【このページのポイント】

身近なところにある統計データの一つに、スポーツ統計があります。
数多くのスポーツにおいて軍事用のトラッキング（追跡システム）を活用し様々なデータを収集し高度な分析がなされています。

【キーワード】
スポーツ統計



【スライドの解説】

■サッカーにおける統計

1950年ごろに取り入れられたと言われています。現代では、カメラやセンサーを用いたデータ取得が行われており、高度な分析も行われています。

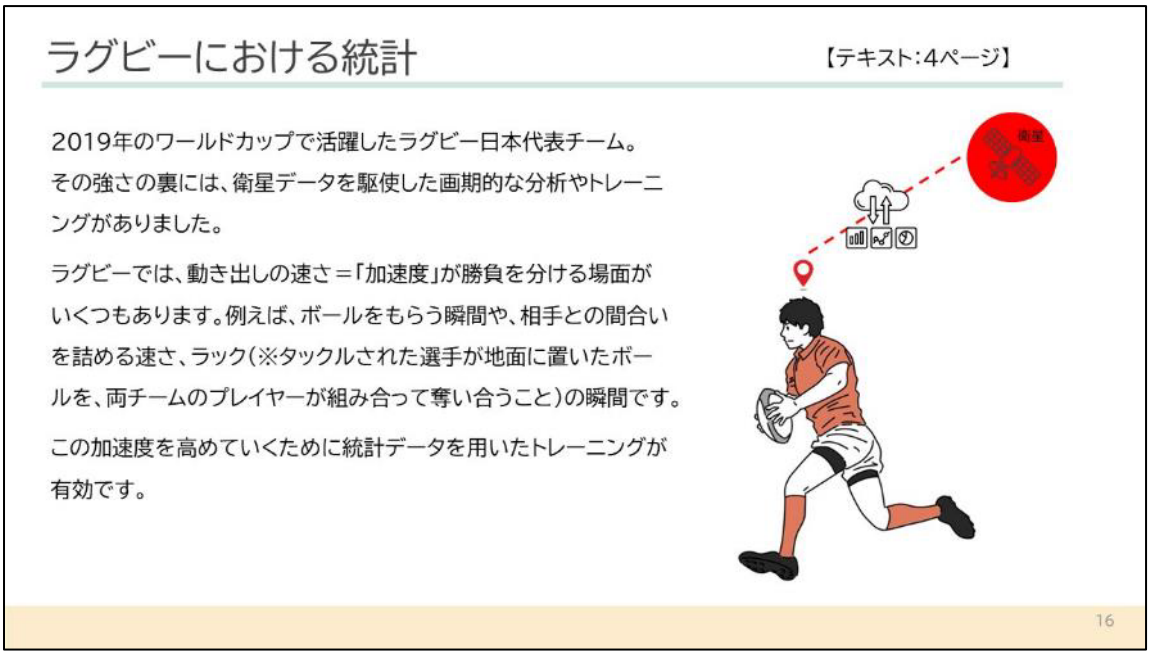
また、戦術を考える上でボールに触れていない選手の動きも重要とされ、すべての選手のデータが収集され重要なデータとなっています。

【このページのポイント】

サッカーにおけるデータ分析は、スポーツにおける分析の先駆けと言われています。

【キーワード】

サッカーにおける統計データと分析



【スライドの解説】

■ラグビーにおける統計

2019年のワールドカップで世界の強豪国と熱戦を繰り広げたラグビー日本代表チーム。

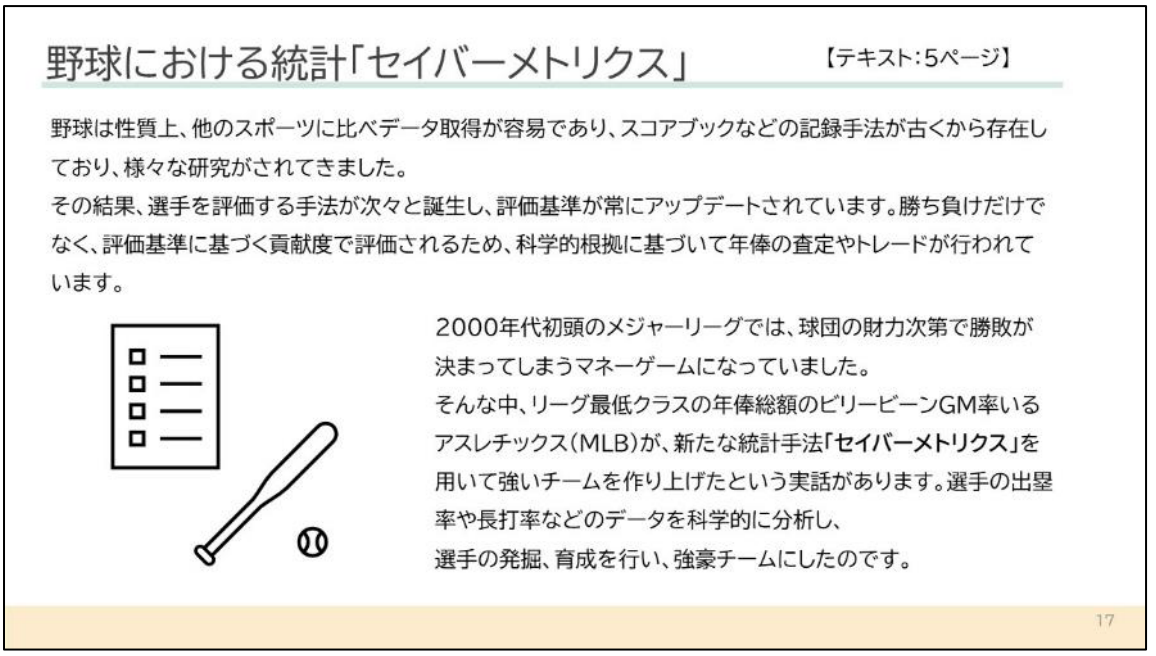
その強さの裏には、衛星データを駆使した画期的なデータ収集やデータ分析、および科学的なトレーニングがありました。GPSを用いて選手の総移動距離、加速回数、スプリント回数などの身体の動きのデータを取得し分析したり、トレーニングの負荷を分析して怪我を防止したりすることで、練習の質が格段に上がりました。

【このページのポイント】

2019年のワールドカップで世界の強豪国と熱戦を繰り広げたラグビー日本代表チームの強さの裏にはスポーツ統計がありました。

【キーワード】

ラグビーにおける統計データと分析



【スライドの解説】

■野球における統計「セイバーメトリクス」

性質上、他のスポーツに比べデータ取得が容易であり、スコアブックなどの記録手法が古くから存在しており、様々な研究がされてきました。その結果、選手を評価する手法が次々と誕生し、評価基準が常にアップデートされています。勝ち負けにこだわり過ぎず、評価基準に基づく貢献度で評価されるため、科学的根拠に基づいて年俸の査定やトレードが行われています。

■参考：映画「マネーボール」（2011年公開）

2000年代初頭のメジャーリーグでは、球団の財力次第で勝敗が決まってしまうマネーゲームになっていました。そんな中、リーグ最低クラスの年俸総額のビリービーンGM率いるアスレチックス（MLB）が、新たな統計手法「セイバーメトリクス」を用い、選手の出塁率や長打率などのデータを科学的に分析し、選手の発掘、育成を行い、強豪チームにしたのです。この実話は、映画「マネーボール」（2011年公開）になっています。

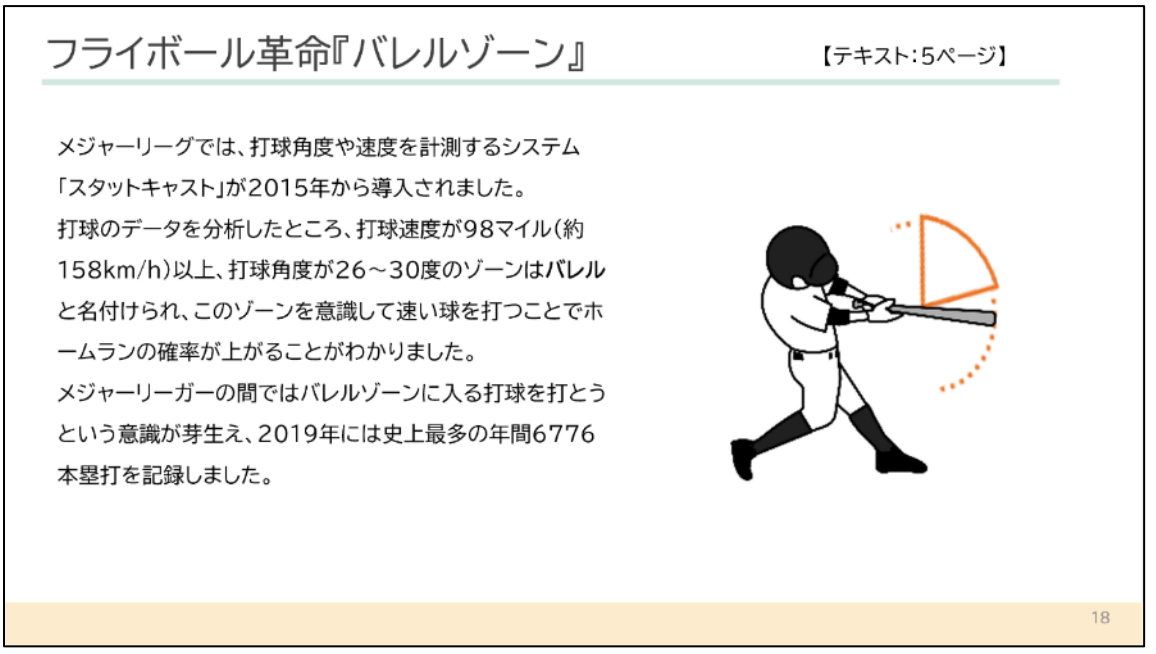
【このページのポイント】

野球では、選手を科学的な根拠に基づいて評価しています。

2000年代初頭に、メジャーリーグの最低年俸総額のアスレチックス（MLB）は、統計データを用いて選手を育成し強豪チームとなりました。

【キーワード】

野球における統計データと分析活用



【スライドの解説】

■野球のデータ分析活用：フライボール革命「バレルゾーン」

2015年からメジャーリーグでは、打球角度や速度を計測するシステム「スタットキャスト」が導入されました。打球のデータを分析した結果、ゴロよりもフライの方が打率と長打率が高く、約153km/h（打球速度）を超えるフライはホームランになる確率が非常に高くなることも判明しました。

この結果から、**バレルゾーン（打球速度が98マイル（約158km/h）以上、打球角度が26～30度）**と名付け、この打球角度と打球速度を意識して打つことでホームランの確率が上がることが分かりました。

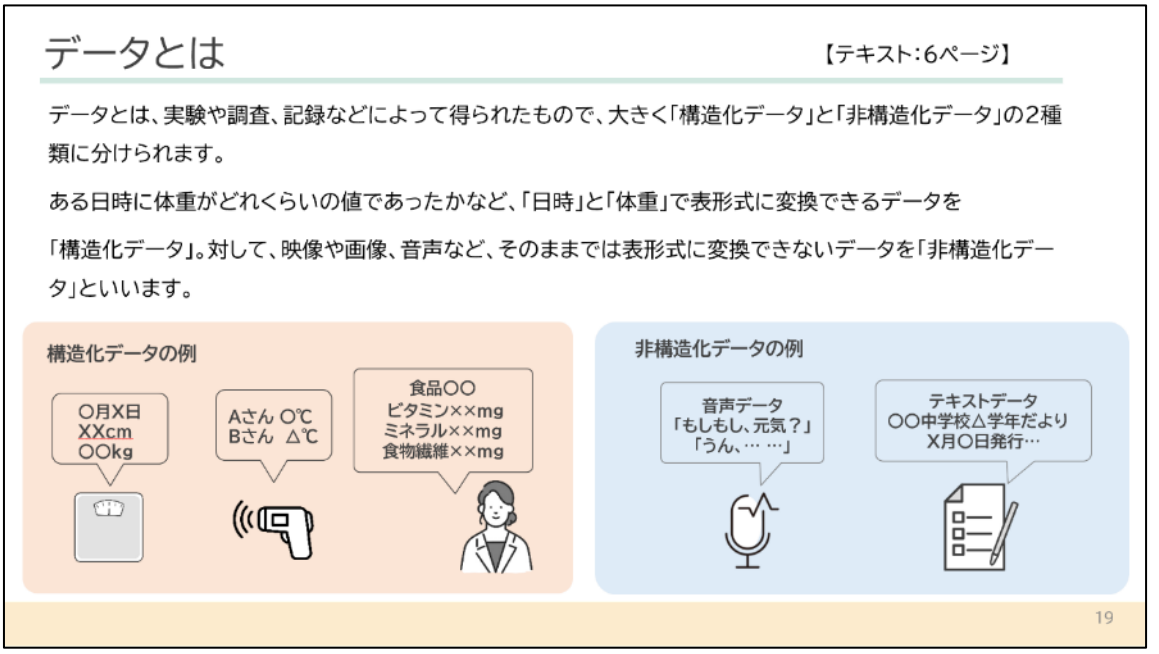
メジャーリーガーの間でバレルゾーンに入る打球を打とうという意識が芽生え、2019年には史上最多の年間6776本塁打を記録しました。

【このページのポイント】

メジャーリーグでは、打球角度や速度を計測するシステム「スタットキャスト」が導入されました。このデータを分析し、バレルゾーン(打球速度が98マイル（約158km/h）以上、打球角度が26～30度)を意識して打つことでホームランの確率が上がることが分かりました。

【キーワード】

バレルゾーン、ホームラン、野球のデータ活用



【スライドの解説】

■データとは

実験や調査、記録などによって得られたもので、2種類に分けられます。

・構造化データ

表形式に変換できるデータです。例えばある日時に体重がどれくらいの値であったかなどは「日時」と「体重」で表として表すことができます。

・非構造化データ

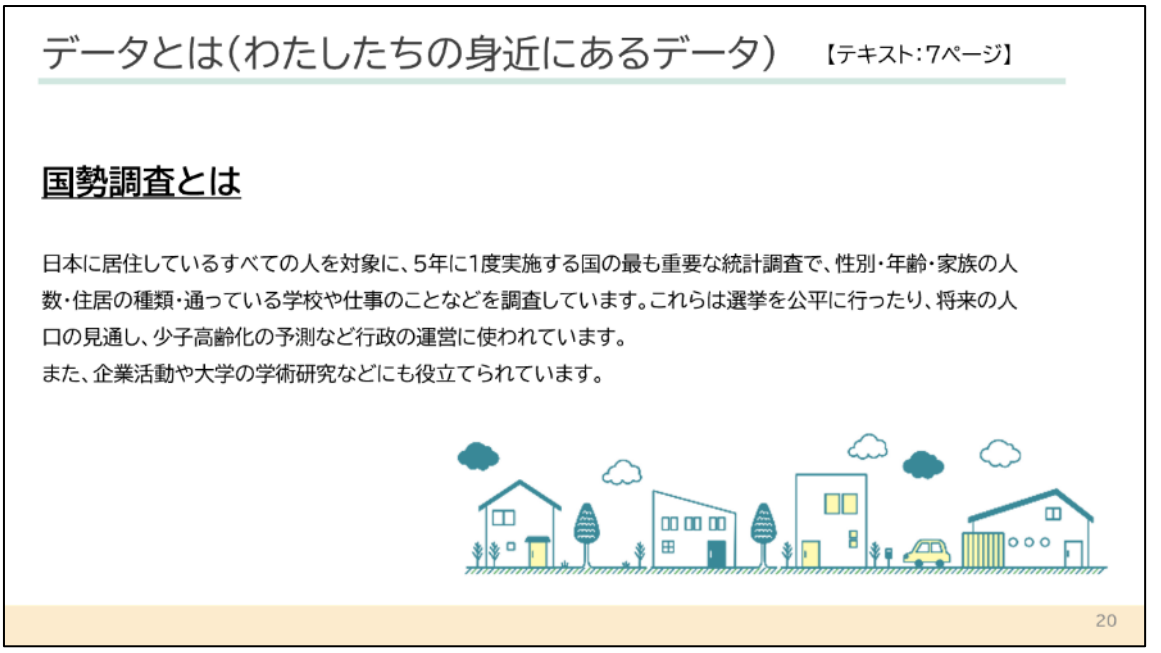
映像や画像、音声など、そのままでは表として表すことができないデータです。

【このページのポイント】

データとは、実験や調査、記録などによって得られたもので、構造化データ、非構造化データに分けられます。

【キーワード】

構造化データ、非構造化データ



【スライドの解説】

■国勢調査

5年に1回行われる日本に住んでいる人全員を対象にした調査です。みなさんの生活について調査を行うことで、より住みやすい国づくりに役立てられます。

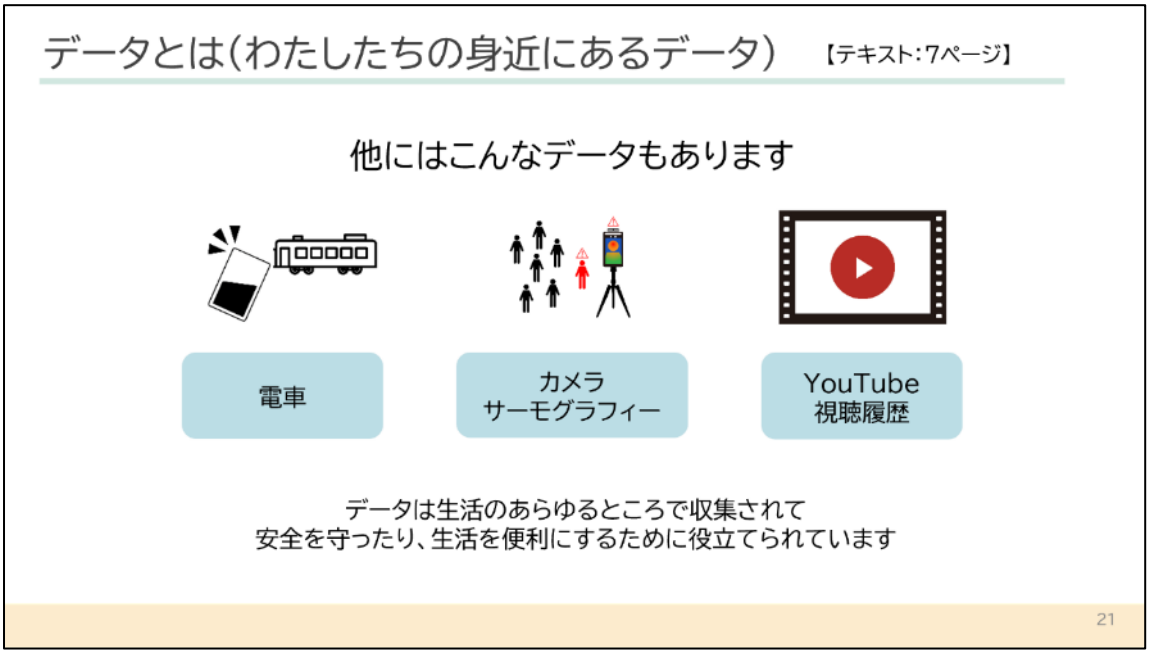
国勢調査の結果は、構造化データであり、表形式で提供されています。

【このページのポイント】

私たちが日々生活している中には、国勢調査で集められたデータを活用して、住みよい街づくり、防災対策などに使われています。

【キーワード】

国勢調査（こくせいちょうさ）、構造化データ



【スライドの解説】

■わたしたちの身近にあるデータ

現代は、ありとあらゆるものが「データ」として取得されています。

例：交通機関や街中の映像（非構造化データ）、動画配信サービスの履歴（構造化データ）など。

動画配信サービスなどでは、「ログ」と呼ばれる、いつ誰が何を見たかのデータを蓄積しています。「ログ」は構造化データです。動画配信サービスでは、「ログ」を分析し、個人の好きなものを推測し、推薦するシステムなどもあります。

【このページのポイント】

私たちの身近には、人のあらゆる行動の情報をデータとして取得する仕組みがたくさんあります。

【キーワード】

行動データ＝ログデータ



【スライドの解説】

■「経済センサス」とは

「経済の国勢調査」とも言われる重要な調査です。調査結果は、国及び地方公共団体における行政施策の立案や、民間企業における経営計画の策定など、社会経済の発展を支える基礎資料として広く活用されています。

■経済センサスの構成

・「経済センサス‐基礎調査」

　活動状況等の基本的構造を明らかにする基礎調査です。

・「経済センサス‐活動調査」

　事業内容、売上・費用、設備投資などの経済活動の状況を明らかにする活動調査です。

　経済センサスの調査結果は、構造化データです。

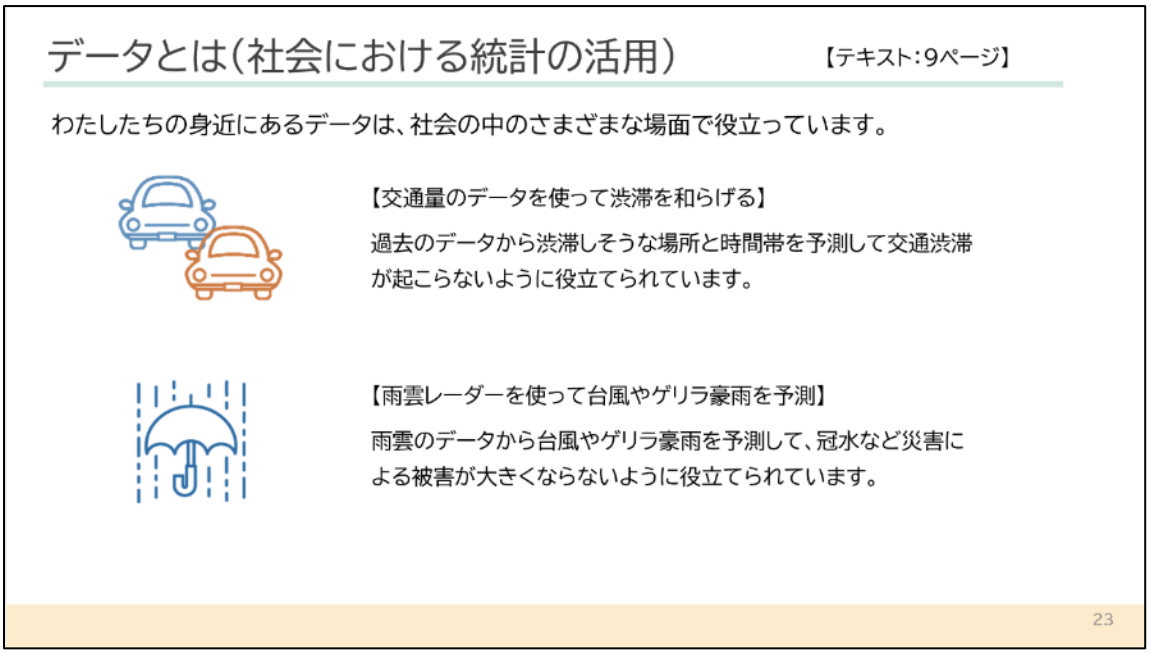
【このページのポイント】

「経済センサス」とは、「経済の国勢調査」とも言われる重要な調査です。

調査結果は、国及び地方公共団体における行政施策の立案や、民間企業における経営計画の策定など、社会経済の発展を支える基礎資料として広く活用されています。

【キーワード】

経済センサス、経済の国勢調査



【スライドの解説】

■生活の中で使われるデータ

みなさんに身近なもののひとつに「交通」「天気」に関するデータがあります。

交通渋滞をさけたり解消したりするために、過去に取得したデータが活用されています。

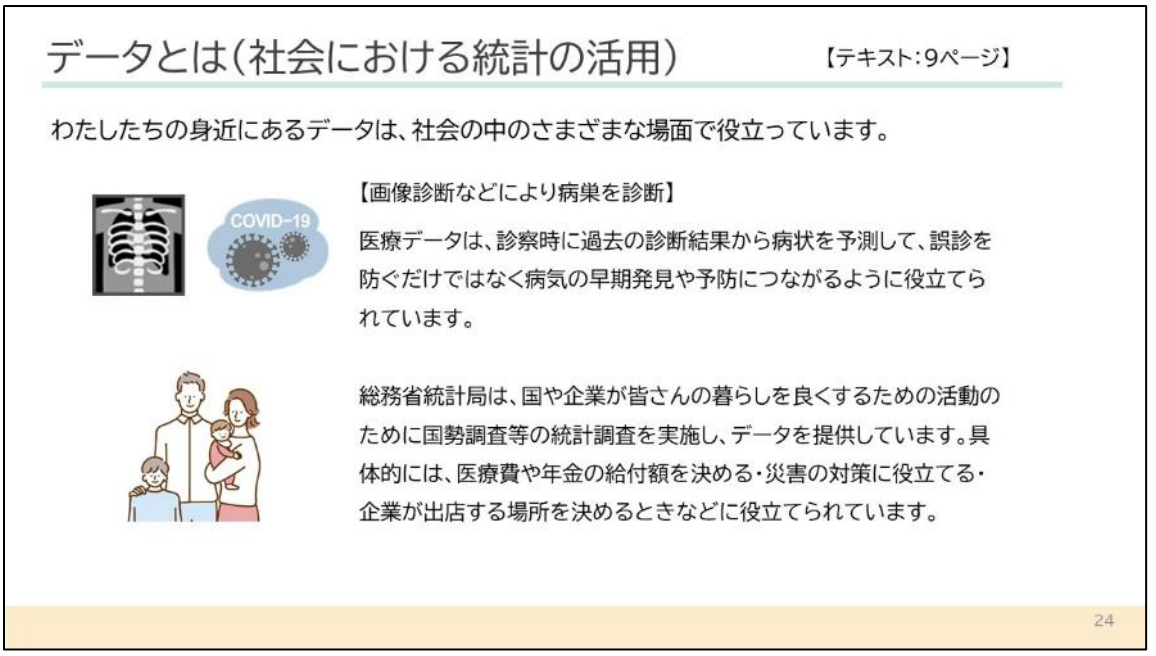
雨雲のデータから台風やゲリラ豪雨を予測して、冠水など災害による警報が出され、被害が大きくならないように役立てられています。

【このページのポイント】

わたしたちの身近なデータは、私たちの暮らしに役に立っています。

【キーワード】

交通量データ、雨雲データ



【スライドの解説】

■病院などにおけるデータ活用

病気の診断をするために過去のデータが活用されています。

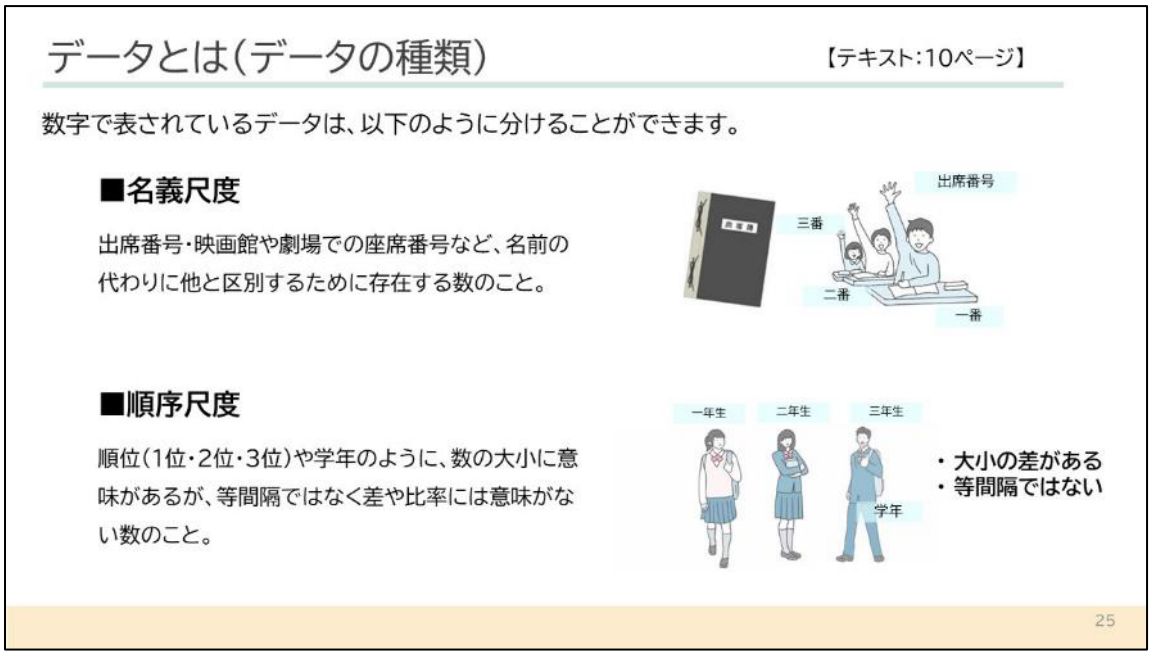
過去にガンになった人の情報は、ガンの早期発見や治療に重要な役割をもっています。

【このページのポイント】

データを集めて分析することで、未来の人々の健康や安全を守るために役立てられています。

【キーワード】

医療データ、早期発見、治療



【スライドの解説】

■データの種類

その役割によって4つに分類されます。このページでは、そのうち2つを説明します。

①**名義尺度（めいぎしゃくど）**：分類するためだけにつけられたデータ（出席番号、座席番号など）

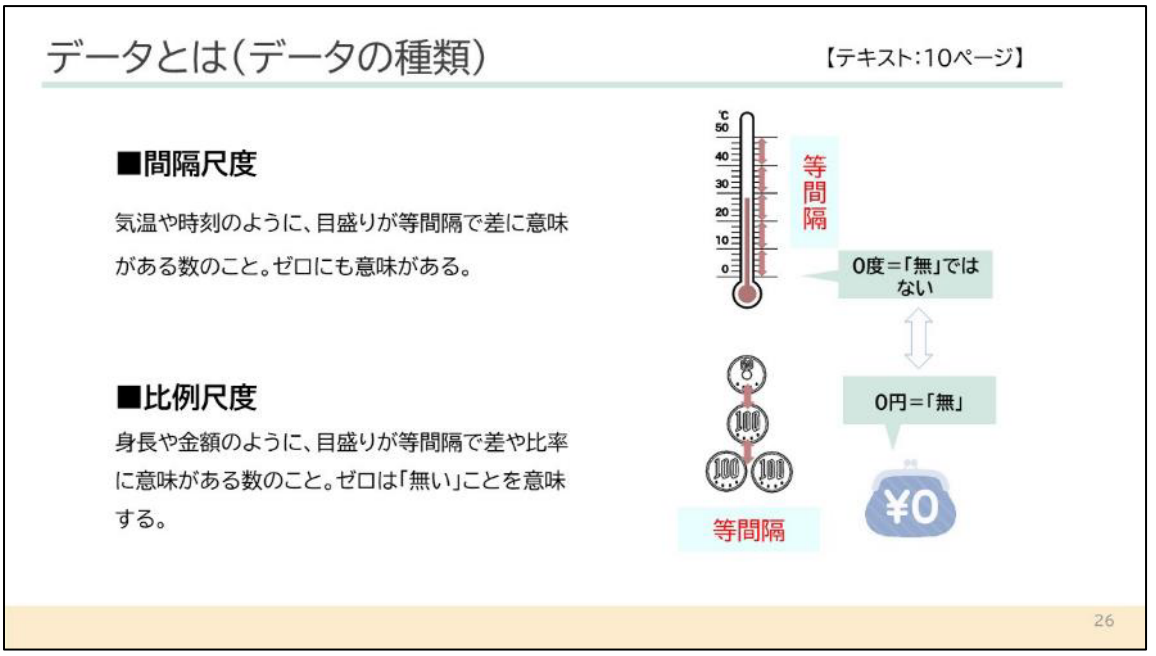
②**順序尺度（じゅんじょしゃくど）**：順番に意味をもつが、データ同士の距離・間隔には意味をもたないデータ（順位・学年など）

【このページのポイント】

順序尺度は、次のページに出てくる間隔尺度と間違えやすいので要注意です！
数字の順番に意味があり、数字同士の差に意味がないものを「順序尺度」、数字の順番にも差にも意味があるものを「間隔尺度」と覚えましょう。

【キーワード】

名義尺度、順序尺度



【スライドの解説】

1つ前のページで紹介した4つの分類のうち、残りの2つを説明します。

③**間隔尺度（かんかくしゃくど）**：気温の1℃と2℃のように、決まった間隔で並び、差の大きさに意味があります。温度にはマイナスがあるので、0℃にも数字として意味があります。

④**比例尺度（ひれいしゃくど）**：身長、年齢のように値の大小関係、差の大きさに意味があり、さらに0（ゼロ）が「ない」を意味するもの。（身長0cm、金額0円=ない）

【このページのポイント】

「間隔尺度」「比例尺度」とも目盛りが等間隔という点は同じですが、0（ゼロ）に意味があるかどうかで区別することが出来ます。

【キーワード】

間隔尺度、比例尺度



【スライドの解説】

前のページで学習した尺度についてのクイズです。

【このページのポイント】

②はゼロに意味があるかないかに注目して考えてみましょう。

③はサイズの大小に意味はあるかないかに注目して考えてみましょう。



【スライドの解説】

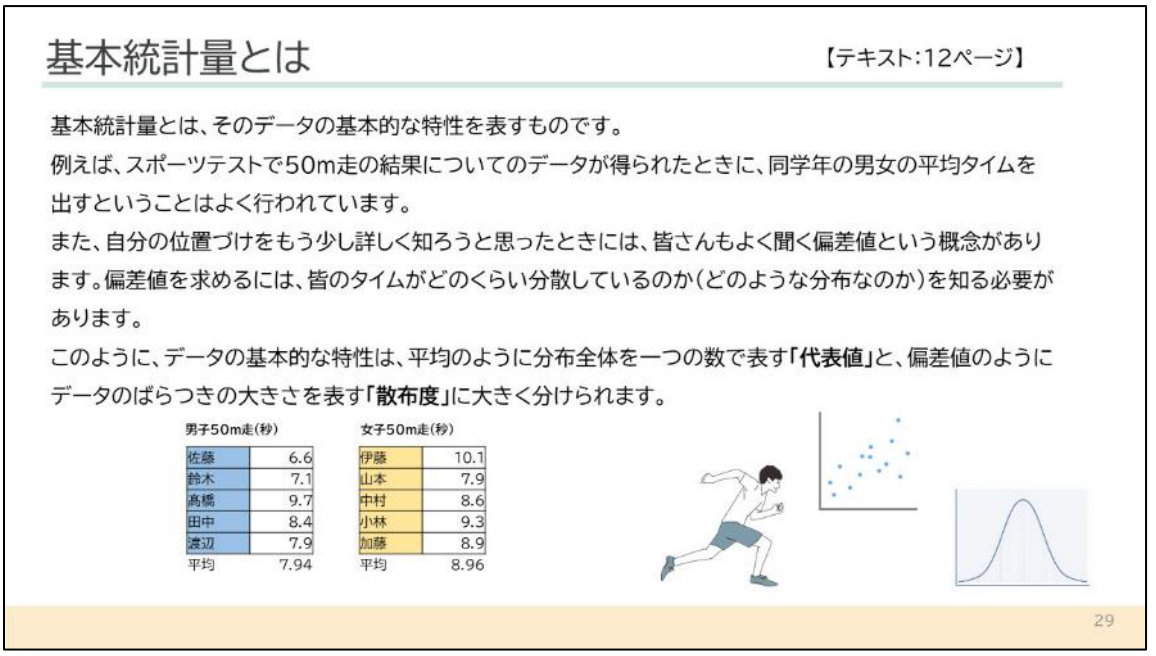
尺度当てクイズの答えを解説します。

【このページのポイント】

①のアイスクリームの種類は、単にアイス同士を分類する尺度なので、正解は、名義尺度です。

②の来場者の数は、もし来場者数か0人だった場合。来場者は「いない」ということになります。よって正解は比例尺度です。

③のぬいぐるみのサイズは、S・M・Lと大きさの大小に意味あることがわかります。よって正解は順序尺度です。



【スライドの解説】

■基本統計量とは

基本統計量とは、そのデータの基本的な特性を表すものです。

・例：

① スポーツテスト：50m走の結果についてのデータ→同学年の男女の平均タイムを出す。

② 偏差値：テスト結果などで、自分の位置づけをもう少し詳しく知ろうと思ったときに用いられる。
偏差値を求めるには、皆のタイムがどのくらい分散しているのか（どのような分布なのか）を知る必要があります。

■データの基本的な特性

・**「代表値」**平均のように分布全体を一つの数で表す。

・**「散布度」**偏差値のようにデータのばらつきの大きさを表す

【このページのポイント】

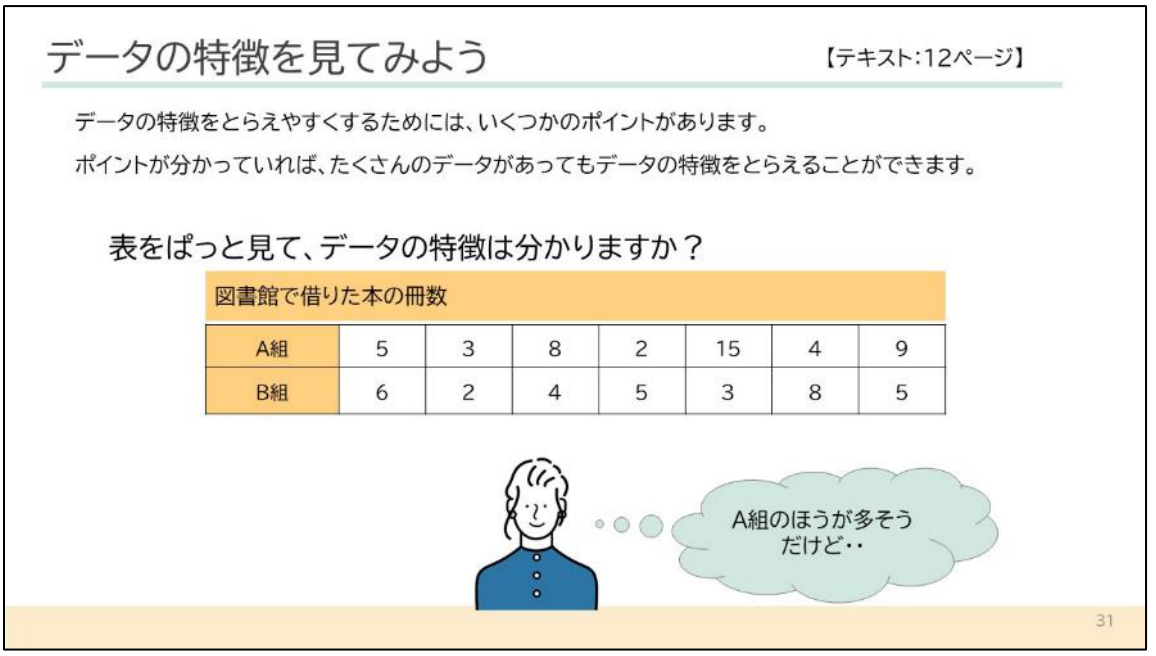
基本統計量とは、そのデータの基本的な特性を表すものです。

データの基本的な特性は、平均のように分布全体を一つの数で表す**「代表値」**と、偏差値のようにデータのばらつきの大きさを表す**「散布度」**に大きく分けられます。

【キーワード】

基本統計量、代表値、散布度


休憩中
休憩が終わったら、データについて
詳しくお話していきます♫
30



【スライドの解説】

■データの取り扱い方について（データ量が少ない場合）

図書館で借りた本の冊数(2クラス分)を比較してみて、どんな特徴があるのかを考えます。

【このページのポイント】

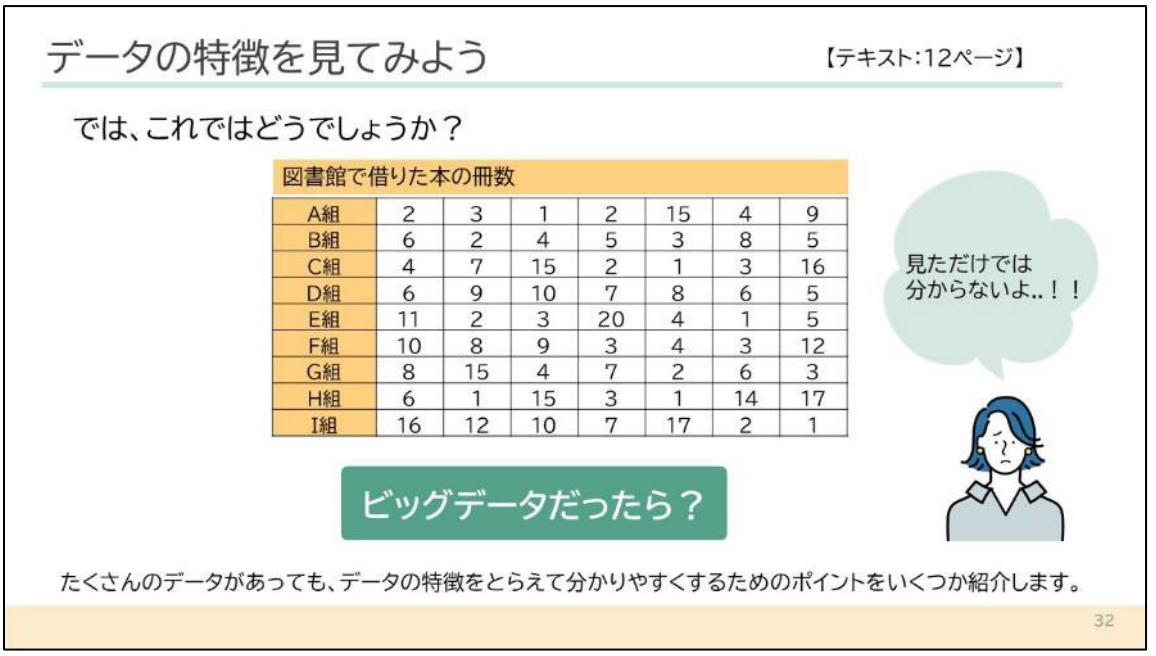
2クラス分のデータを比較したときに、ぱっと見てA組の借りた本の数が多いことがわかります。

データの量が少ないと、数字だけを見ただけでも特徴が把握しやすいです。

【合計冊数】

A組　46冊

B組　33冊



【スライドの解説】

■データの取り扱い方について（データ量が多い場合）

図書館で借りた本の冊数(9クラス分)を比較してみて、どんな特徴があるのかを考えます。

【このページのポイント】

9クラス分のデータを比較したときに、すぐに特徴をとらえることは難しいことが分かります。

交通渋滞の予測や、医療の画像から癌（がん）かどうかを判別するような時にはたくさんのデータが必要です。このようなビッグデータから特徴をすぐに把握するためのポイントがいくつかあります。

【キーワード】

ビッグデータ：世の中にあるたくさんのデータ

A組36冊　B組33冊　C組48冊　D組51冊　E組46冊
F組49冊　G組45冊　H組57冊　I組65冊

データの特徴を見てみよう(平均値)

【テキスト:13ページ】

平均とは、いくつかの数値に対して、それぞれをすべて足し合わせた後、数値の個数で割った値のことをいいます。

例えば7人の身長が以下のような場合、
7人の身長の合計値を人数で割った値が平均身長となります。

【例1】

名前	身長(cm)
Aさん	140
Bさん	135
Cさん	145
Dさん	155
Eさん	160
Fさん	150
Gさん	165



平均	(140+135+145+155+160+150+165)÷7=150(cm)

33

【スライドの解説】

■平均

たくさんのデータから中心的な位置をとらえる方法の１つに「平均」があります。

平均は最も頻繁に使われ、みなさんも良く知っているもので、代表値の一つです。

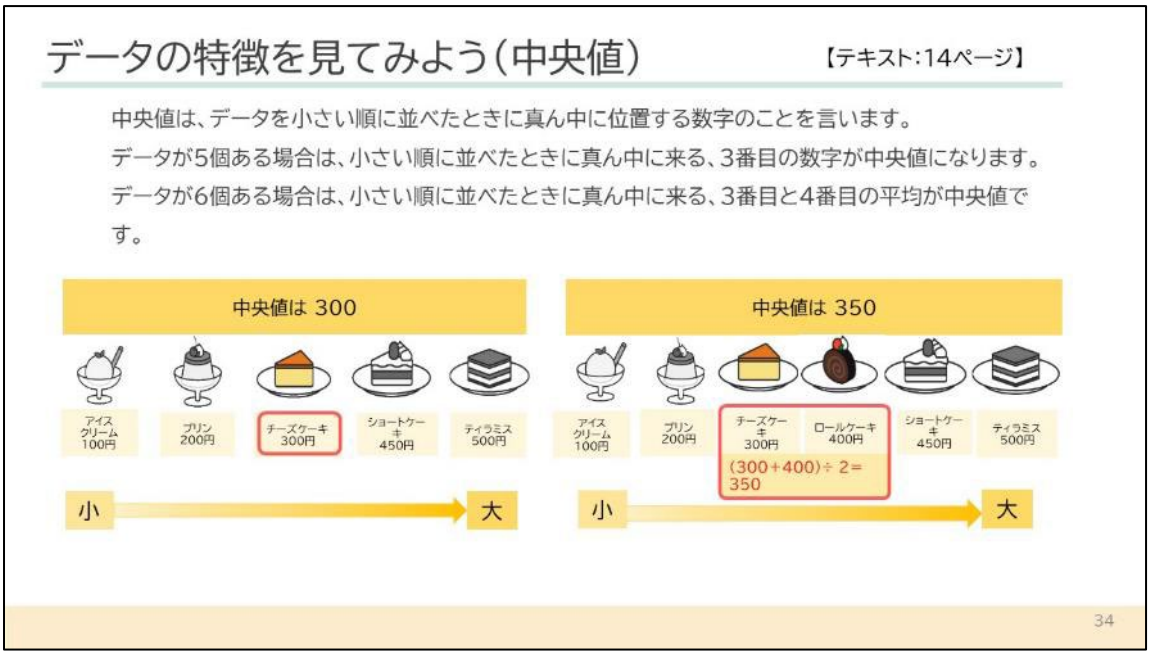
平均は、すべてのデータをすべて足し合わせて、それをデータの数で割った値です。

【このページのポイント】

このページでは身長の平均を求めます。表にのっているAさん～Gさんの身長を足し合わせ、それを人数(7人)で割った値が平均身長となります。

【キーワード】

平均：いくつかの数値を全部足して、その個数で割ったもの



【スライドの解説】

■中央値

平均以外の重要な「代表値」は「中央値」です。「中央値」とは、データを小さい順に並べたときに真ん中に位置する数値です。データが5個(奇数)ある場合は、小さい順に並べたときに真ん中に来る、3番目の数字が中央値になります。データが6個(偶数)ある場合は、小さい順に並べたときに真ん中に来る、3番目と4番目の平均が中央値です。

【このページのポイント】

■左の図（データ数が奇数の場合）

５種類のスイーツが左から右に、値段の安いものから高いものへと順番に並んでいます。この場合はデータ数が奇数なので、真ん中にくるチーズケーキの300円が中央値です。

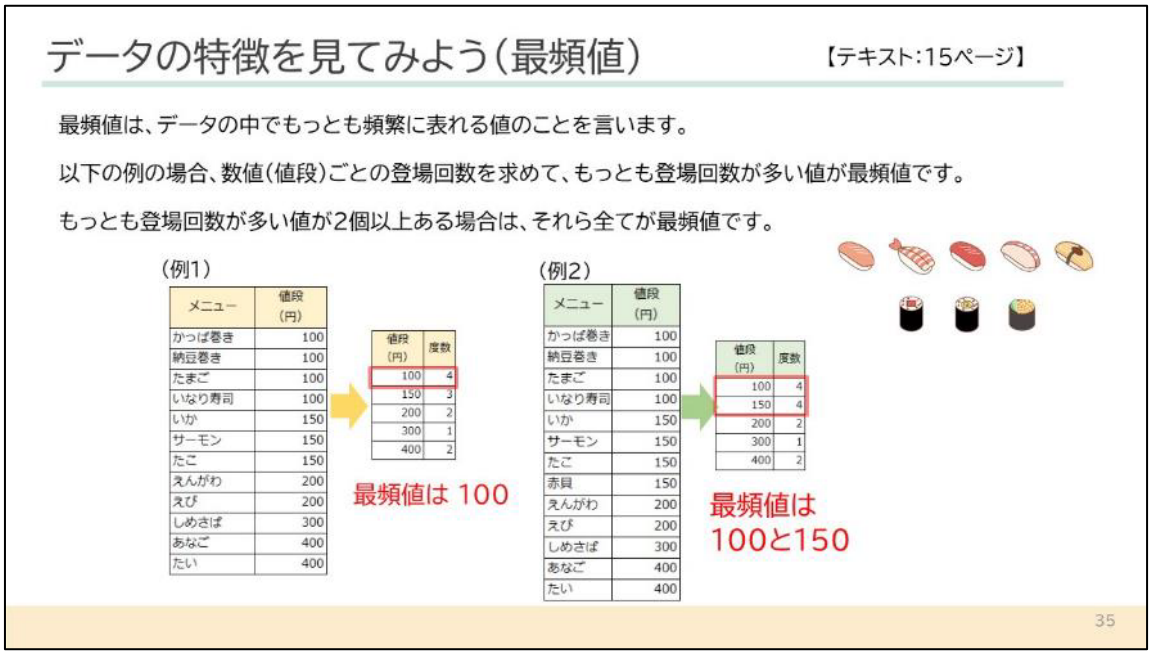
■右の図（データ数が偶数の場合）

スイーツが6種類だった場合、データ数が偶数なので、中央値は真ん中の２つの平均となります。つまり、チーズケーキの300円とロールケーキの400円の平均です。

(300＋400)÷2＝350となるので、中央値は350円です。

【キーワード】

中央値：データを小さい順から大きい順に、または大きい順から小さい順に並べたときの真ん中の値



【スライドの解説】

■最頻値

平均、中央値以外にもよく使われる「代表値」があります。「最頻値」です。

最頻値とは、データの中で最も頻繁に表われる値のことを言います。最も頻繁に表われる値が2つ以上あった場合には、それら全てが最頻値となります。

【このページのポイント】

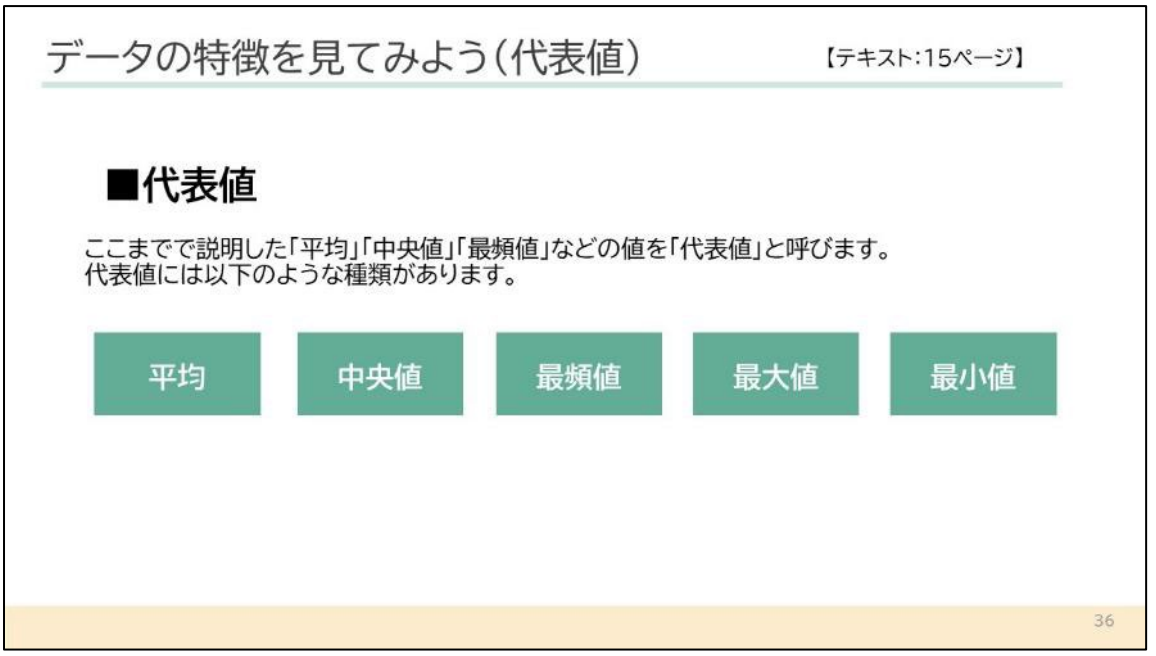
２つのお店で、いくらの値段のお寿司が一番多いのかを調べます。

(例1)では、100円のお寿司が最も多いことがわかります。つまり、最頻値は100円です。

(例2)では、100円と150円のお寿司が最も多いことがわかります。どちらも最頻値といえるので、最頻値は100円と150円となります。

【キーワード】

最頻値：データの中でもっとも登場回数が多い値



【スライドの解説】

■代表値

代表値には、「平均」、「中央値」、「最頻値」があります。

【このページのポイント】

「平均」、「中央値」、「最頻値」の3つはどんなデータを見るときも必ずチェックします。

データの特徴をさらに把握するためには、「最小値」や「最大値」もチェックします。

【キーワード】

最大値：データの中でもっとも大きい値

最小値：データの中でもっとも小さい値

代表値：データを代表する値である「平均」「中央値」「最頻値」の総称