

# 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活用センター・和歌山県データ活用推進センター 運営(委託) | 株式会社Rejoui

## セミナープログラムについて

統計の基礎  
70分



データ演習  
70分

このセミナーでは、  
講師が説明する  
**「統計の基礎」**  
の時間と

グループに分かれて  
意見を発表する  
**「演習」**  
の時間があります

### 今日の授業内容

- 【統計の基礎】統計ってなんだろう  
(身近にある統計、データ)
- 【統計の基礎】データの可視化(グラフについて)
- 【演習】統計ダッシュボードで調べてみよう!
- 【演習】なりきり! データサイエンティスト

講義が終わった後しばらくはみなさまの質問にお答えしたり  
お話ししたりする時間を作りたいと思っています。

2

### 【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。

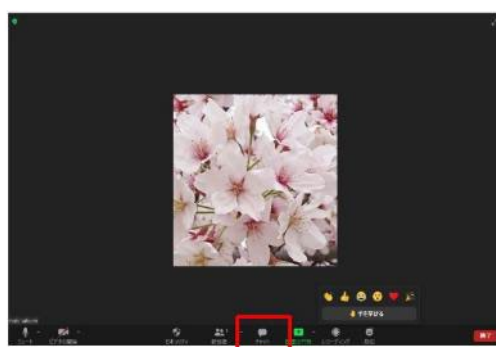
オンラインでご参加の皆さまへ

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！



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！

チャットを  
使って  
聞いてみよう





## 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

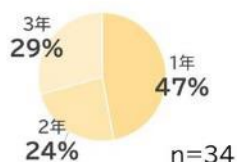


2016年9月に株式会社Rejouiという会社を作り、データサイエンティストの育成や学習カリキュラム・教材等を開発するお仕事をしています。

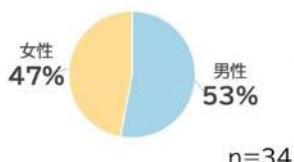
これまでデータサイエンスに関する学習講座やイベントをたくさん開催しており、小学生へ向けた「データサイエンスforキッズ」(自社開催)や「わくわく！統計アカデミー」(主催：総務省統計局)のほか、「機械学習・データサイエンス入門」(主催：総務省統計局統計研究研修所)では、高校の先生にデータサイエンスを教える先生として講師を務めました。子どもから大人までたくさんの人たちにデータサイエンスの楽しさを伝えています。

## 本日の参加者データ

### 学年



### 男女比



### 参加エリア



### 血液型は？ (わかる人のみ)



### 生まれの季節は？



### ペットを飼っている？



### 【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%が多い。
- 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したりと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。

一緒にたくさん話し合おう！

会場



Zoom



メンターが合図します！

皆さんのいろいろなアイデアを聞かせてください。

# 第一章

## 統計ってなんだろう？





統

統べて(たくさん)集める  
集めるもの=情報(データ)



計

計算する  
特徴や性質を推測  
未来を予測

## 【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

## 身近な統計について

【テキスト:3ページ】

身近なところにある統計データとしては、総務省が行なっている国勢調査や、身体測データ、天気データ、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。



11

### 【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

国勢調査は、日本に住んでいるすべての人を対象に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。データは行政の運営や街づくりに役立てられています。

人口動態調査は、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。



### 【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

## 人口統計について

【テキスト:3ページ】

人口統計は、生まれた子どもは何年後に何歳になるか、正確に予測できるように将来を予測しやすいという特性があります。人口動態は、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。

減っていく人口に対して統計データを用いて将来を予測し、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いるので、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生かされています。



13

### 【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】

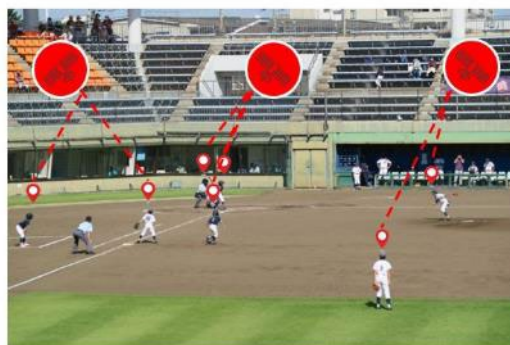
人口動態調査、人口統計、人口動態は経済の基礎、地方自治体の施策

## スポーツ統計について

【テキスト:4ページ】

身近なところにある統計データとして、スポーツ統計があります。

近年では、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。スポーツ統計では、科学的な見地からの戦術解析や選手起用・評価が行われています。戦術はもちろんのこと、対戦相手の評価、選手起用の判断が行われています。統計を用いることにより、感覚ではわからない価値を発見することができます。



14

### 【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

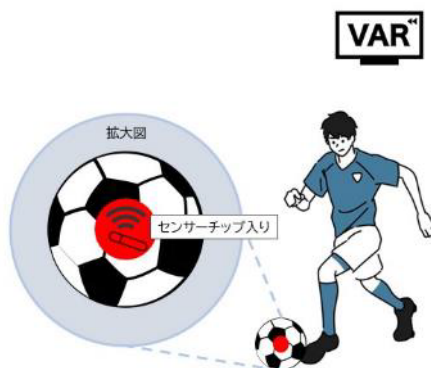
## サッカーにおける統計

【テキスト:4ページ】

サッカーにおけるデータ分析は、1950年ごろに取り入れられたと言われていました。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、高度な分析も実現可能です。ボールに触れていない選手の動きも重要とされており、戦術を考える上でも重要なデータとなっています。

<主な活用例>

- 選手ごとのシュート本数と枠内に飛んだ確率などを測定し、次にパスがくる方向を予測する。
- 選手ごとの走行距離、スプリント回数などのデータを用いて「ゴール期待値」を算出する。



15

【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した。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



## ラグビーにおける統計

【テキスト:4ページ】

2019年のワールドカップで活躍し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分析やトレーニングがありました。

ラグビーでは、動き出しの速さ＝「加速度」が勝負を分ける場面がいくつかあります。例えば、ボールをもらう瞬間や、相手との間合いを詰める速さ、ラック（※タックルされた選手が地面に置いたボールを、両チームのプレイヤーが組み合って奪い合うこと）の瞬間です。この加速度を高めていくために統計データを用いたトレーニングが有効です。



16

### 【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】

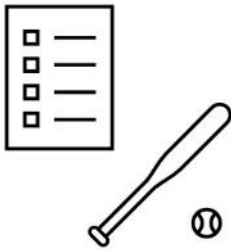
ラグビーにおける統計データと分析

## 野球における統計「セイバーメトリクス」

【テキスト:5ページ】

野球は性質上、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。勝ち負けだけでなく、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。



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率いるアスレックス(MLB)が、新たな統計手法「セイバーメトリクス」を用いて強いチームを作り上げたという実話があります。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。

17

### 【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

## フライボール革命『バレルゾーン』

【テキスト:5ページ】

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2015年から導入されました。打球のデータを分析したところ、打球速度が98マイル(約158km/h)以上、打球角度が26~30度のゾーンはバレルと名付けられ、このゾーンを意識して速い球を打つことでホームランの確率が上がることがわかりました。メジャーリーガーの間では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。



18

### 【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わかりました。

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わかりました。

### 【キーワード】

バレルゾーン、ホームラン、野球のデータ活用

# データとは

【テキスト:6ページ】

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、大きく「構造化データ」と「非構造化データ」の2種類に分けられます。

ある日時に体重がどれくらいの値であったかなど、「日時」と「体重」で表形式に変換できるデータを「構造化データ」。対して、映像や画像、音声など、そのままでは表形式に変換できないデータを「非構造化データ」といいます。



19

## 【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

### **国勢調査とは**

日本に居住しているすべての人を対象に、5年に1度実施する国の最も重要な統計調査で、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。これらは選挙を公平に行ったり、将来の人口の見通し、少子高齢化の予測など行政の運営に使われています。

また、企業活動や大学の学術研究などにも役立てられています。



#### 【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

## データとは(わたしたちの身近にあるデータ) 【テキスト:7ページ】

他にはこんなデータもあります



電車



カメラ  
サーモグラフィー



YouTube  
視聴履歴

データは生活のあらゆるところで収集されて  
安全を守ったり、生活を便利にするために役立てられています

21

### 【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

## データとは(わたしたちの身近にあるデータ) 【テキスト:8ページ】

### 経済センサスとは

日本の全産業分野の経済活動や産業構造を把握するため、日本国内にある全ての事業所・企業を対象に行う経済調査で、「経済の国勢調査」とも言われる重要な調査です。この経済センサスは「経済センサス-基礎調査」と「経済センサス-活動調査」の2つの調査で構成されています。「経済センサス」の調査結果は、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。また、調査結果により、日本全体の経済活動の変動や動向を把握しやすくなります。



22

#### 【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

## データとは(社会における統計の活用)

【テキスト:9ページ】

わたしたちの身近にあるデータは、社会の中のさまざまな場面で役立っています。



【交通量のデータを使って渋滞を和らげる】

過去のデータから渋滞しそうな場所と時間帯を予測して交通渋滞が起こらないように役立てられています。



【雨雲レーダーを使って台風やゲリラ豪雨を予測】

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被害が大きくなるように役立てられています。

23

### 【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、被害が大きくなる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



## データとは(社会における統計の活用)

【テキスト:9ページ】

わたしたちの身近にあるデータは、社会の中のさまざまな場面で役立っています。



### 【画像診断などにより病巣を診断】

医療データは、診察時に過去の診断結果から病状を予測して、誤診を防ぐだけでなく病気の早期発見や予防につながるように役立てられています。



総務省統計局は、国や企業が皆さんの暮らしを良くするための活動のために国勢調査等の統計調査を実施し、データを提供しています。具体的には、医療費や年金の給付額を決める・災害の対策に役立てる・企業が出店する場所を決めるときなどに役立てられています。

24

### 【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

## データとは(データの種類)

【テキスト:10ページ】

数字で表されているデータは、以下のように分けることができます。

### ■名義尺度

出席番号・映画館や劇場での座席番号など、名前の代わりに他と区別するために存在する数のこと。



### ■順序尺度

順位(1位・2位・3位)や学年のように、数の大小に意味があるが、等間隔ではなく差や比率には意味がない数のこと。



25

【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】

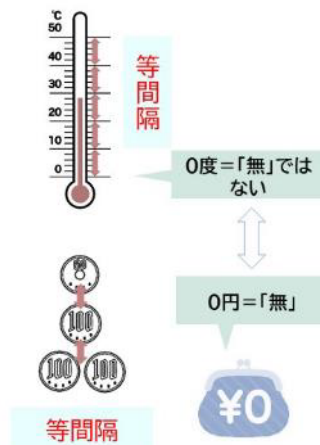
名義尺度、順序尺度

## ■間隔尺度

気温や時刻のように、目盛りが等間隔で差に意味がある数のこと。ゼロにも意味がある。

## ■比例尺度

身長や金額のように、目盛りが等間隔で差や比率に意味がある数のこと。ゼロは「無い」ことを意味する。



### 【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°Cと2°C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°C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

## データとは(データの種類)

【テキスト:11ページ】

クイズ！次のデータに用いられる尺度を当ててみよう！！

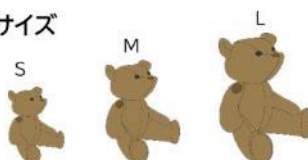
①アイスクリームの種類



②来場者の数



③ぬいぐるみのサイズ



ゼロに意味があるのか、ないのかにも注目して考えてみよう



27

### 【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。

## データとは(データの種類)

【テキスト:11ページ】

クイズ！次のデータに用いられる尺度を当ててみよう！！

①アイスクリームの種類



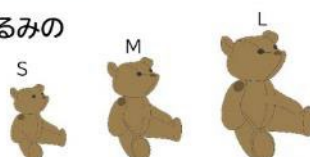
答え:名義尺度

②来場者の数



答え:比例尺度

③ぬいぐるみのサイズ



答え:順序尺度

28

### 【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、もし来場者数が0人だった場合。来場者は「いない」ということとなります。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。

# 基本統計量とは

【テキスト:12ページ】

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。

例えば、スポーツテストで50m走の結果についてのデータが得られたときに、同学年の男女の平均タイムを出すということはよく行われています。

また、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は、皆さんもよく聞く偏差値という概念があります。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。

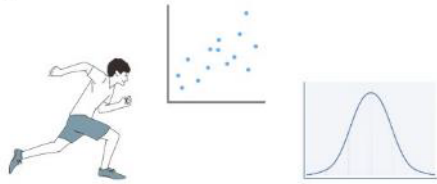
このように、データの基本的な特性は、平均のように分布全体を一つの数で表す「代表値」と、偏差値のようにデータのばらつきの大きさを表す「散布度」に大きく分けられます。

男子50m走(秒)

佐藤	6.6
鈴木	7.1
高橋	9.7
田中	8.4
渡辺	7.9
平均	7.94

女子50m走(秒)

伊藤	10.1
山本	7.9
中村	8.6
小林	9.3
加藤	8.9
平均	8.96



29

## 【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



## データの特徴を見よう

【テキスト:12ページ】

データの特徴をとらえやすくするためには、いくつかのポイントがあります。

ポイントが分かれば、たくさんのデータがあってもデータの特徴をとらえることができます。

表をぱっと見て、データの特徴は分かりますか？

図書館で借りた本の冊数							
A組	5	3	8	2	15	4	9
B組	6	2	4	5	3	8	5



A組のほうが多そう  
だけど..

31

### 【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



## データの特徴を見てみよう

【テキスト:12ページ】

では、これではどうでしょうか？

図書館で借りた本の冊数							
A組	2	3	1	2	15	4	9
B組	6	2	4	5	3	8	5
C組	4	7	15	2	1	3	16
D組	6	9	10	7	8	6	5
E組	11	2	3	20	4	1	5
F組	10	8	9	3	4	3	12
G組	8	15	4	7	2	6	3
H組	6	1	15	3	1	14	17
I組	16	12	10	7	17	2	1

見ただけでは  
分からないよ..!!



ビッグデータだったら？

たくさんのデータがあっても、データの特徴をとらえて分かりやすくするためのポイントをいくつか紹介します。

32

### 【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	$(140 + 135 + 145 + 155 + 160 + 150 + 165) \div 7 = 150(\text{cm})$
----	---

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

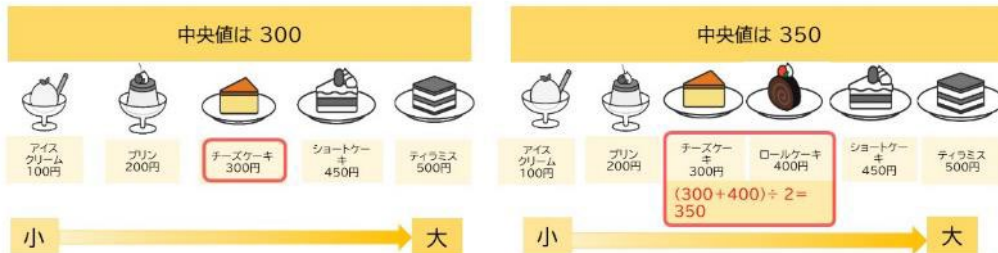
## データの特徴を見てみよう(中央値)

【テキスト:14ページ】

中央値は、データを小さい順に並べたときに真ん中に位置する数字のことを言います。

データが5個ある場合は、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。



34

### 【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。

$(300 + 400) \div 2 = 350$ 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

## データの特徴を見てみよう(最頻値)

【テキスト:15ページ】

最頻値は、データの中でもっとも頻繁に表れる値のことを言います。

以下の例の場合、数値(値段)ごとの登場回数を求めて、もっとも登場回数が多い値が最頻値です。

もっとも登場回数が多い値が2個以上ある場合は、それら全てが最頻値です。

(例1)

メニュー	値段 (円)
かつば巻き	100
納豆巻き	100
たまご	100
いなり寿司	100
いか	150
サーモン	150
たこ	150
えんがわ	200
えび	200
しめさば	300
あなご	400
たい	400

値段 (円)	回数
100	4
150	3
200	2
300	1
400	2

最頻値は 100

(例2)

メニュー	値段 (円)
かつば巻き	100
納豆巻き	100
たまご	100
いなり寿司	100
いか	150
サーモン	150
たこ	150
赤貝	150
えんがわ	200
えび	200
しめさば	300
あなご	400
たい	400

値段 (円)	回数
100	4
150	4
200	2
300	1
400	2

最頻値は  
100と150



35

### 【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

### ■代表値

ここまでで説明した「平均」「中央値」「最頻値」などの値を「代表値」と呼びます。  
代表値には以下のような種類があります。

平均

中央値

最頻値

最大値

最小値

#### 【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

## 第二章 データの可視化



## 可視化について

【テキスト:17ページ】

可視化とは、データを「目で見てわかりやすい形」に変えることです。

データ自体は文字や数字・記号などがただ並んだものであり、それだけ見ても、何が多いのか、どのような傾向があるのか、などをあまり読み取ることができず、せっかくデータがあっても役に立ちません。

そこで、図やグラフなどを使って、「見てもよくわからない形」であるデータを、「見てわかりやすい形」に変えることが可視化です。



38

### 【スライドの解説】

#### ■可視化とは

「見てもよくわからない」データを、図やグラフなどを使って、「目で見てわかりやすい形」に変えることです。

### 【このページのポイント】

図やグラフを用いると、データの全体像を分かりやすく伝えることができます。

### 【キーワード】

可視化

# 可視化することのメリット

【テキスト:18ページ】

可視化することのメリット1つ目は、「自分で見てわかりやすくなること」です。グラフで可視化することで、ある物事の変化を見たり、大きさや割合を比較したり、関係性を見つけたりしやすくなります。

2つ目は、「他の人に伝えやすくなること」です。データをそのまま見せるよりも、可視化した(わかりやすくまとめた)状態の方が、より素早く、正確に伝えることができます。

和歌山県 月別降水量の比較

	2020年の降水量合計(mm)	1990~2020年の降水量合計の平均(mm)
1月	78.5	36.2
2月	62.0	45.4
3月	89.0	82.5
4月	99.5	90.0
5月	113.5	112.6
6月	211.5	169.3
7月	438.5	177.4
8月	13.0	97.2
9月	153.5	142.2
10月	309.0	95.4
11月	62.5	53.3
12月	27.0	41.5

グラフで表すと..



一目で特徴が分かる!



39

## 【スライドの解説】

### ■可視化（目で見える表現に変換）することのメリット

メリット①：ひと目見ただけではっきりとわかるようになることです。グラフで可視化することで、ある物事の変化を発見したり、大きさや割合を比較したり、関係性に気がついたりすることが容易になります。

メリット②：他の人に伝わりやすくなることです。データをそのまま見せるよりも、ひと目見ただけではっきりとわかるような表現の方が、正確に伝わります。

## 【このページのポイント】

データを可視化することのメリットは、

- ・「自分で見てわかりやすくなること」
- ・「他の人に伝えやすくなること」

## 【キーワード】

可視化、グラフ



## グラフについて

先ほど見た表です。  
どちらの組がたくさん借りているかわかりますか？

図書館で借りた本の冊数							
A組	2	3	1	2	15	4	9
B組	6	2	4	5	3	8	5
C組	4	7	15	2	1	3	16
D組	6	9	10	7	8	6	5
E組	11	2	3	20	4	1	5
F組	10	8	9	3	4	3	12
G組	8	15	4	7	2	6	3
H組	6	1	15	3	1	14	17
I組	16	12	10	7	17	2	1

やっぱり  
見ただけでは  
分からないよ..!!



たくさんのデータ  
(ビッグデータ)だったら..?

40

### 【スライドの解説】

#### ■グラフについて

前にもお見せしたデータです。

A組からI組までの生徒たちが各月に図書館から借りた本の冊数を表にまとめた表です。

各クラスの借り出しにどんな特徴があるのかを考えてみましょう。

表ですと、特徴の前に、どのクラスが最も借りた冊数が多いのかが分かりません。

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

### 【このページのポイント】

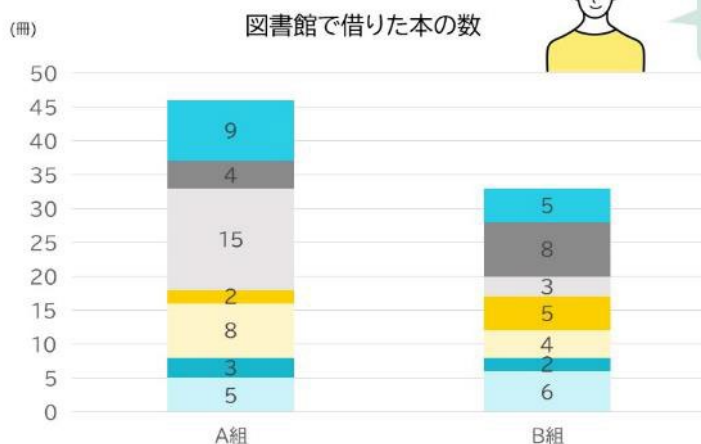
データの量が多いと、表の数値だけを見ただけでは特徴が把握し難いです。

### 【キーワード】

可視化、グラフ

## グラフについて

表をグラフにしてみると……



このように1本の棒に、複数のデータを積み上げて表示したグラフのことを積み上げ棒グラフといいます

41

【スライドの解説】

### ■データの可視化についての具体例①

・積み上げ棒グラフ：A組とB組の生徒たちが各月に図書館から借りた本の冊数をまとめた表

各月ごとの借りた冊数を色で分けて積み上げたグラフです。A組の借りた冊数は45冊と多いことが一目で分かります。B組の借りた冊数は33冊です。

【このページのポイント】

適切なグラフで可視化すると、得られたデータの特徴が目で把握できます。

【キーワード】

可視化、グラフ、積み上げ棒グラフ

## グラフについて

表をグラフにしてみると……

図書館で借りた本の数



グラフにすると、どの組が1番多く本を借りているかが分かりやすくなったね！



42

【スライドの解説】

### ■データの可視化についての具体例②

・積み上げ棒グラフ：A組からI組の生徒たちが各月に図書館から借りた本の冊数をまとめた表

各月ごとの借りた冊数を色で分けて積み上げたグラフです。I組の借りた冊数は65冊と多いことが一目で分かります。一番少ないのがB組33冊です。

【このページのポイント】

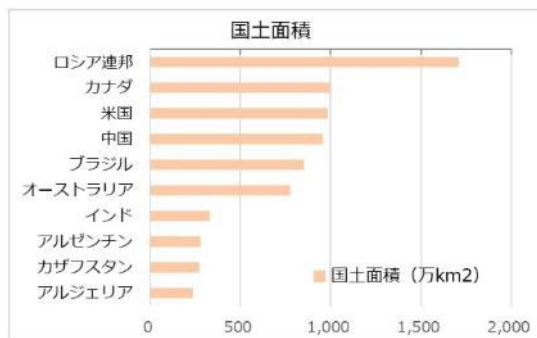
適切なグラフで可視化すると、得られたデータの特徴が目で把握できます。

【キーワード】

可視化、グラフ、積み上げ棒グラフ

## ■棒グラフ

「面積」「点数」「身長」など、同じものについて比較を行う場合、棒グラフが適しています。



## 【スライドの解説】

グラフにはいくつかの種類があります。何を表したいかによって適しているグラフは異なります。ここでは代表的なグラフをいくつか紹介します。

## ■棒グラフ

棒グラフとは数量を棒の長さで示したグラフのことです。棒グラフは、棒と棒の間をあけま

- ・例：国土面積を国別で表しています。国名は、名義尺度です。

## 【このページのポイント】

「面積」「点数」「身長」など一つの統計データの値を比較する場合は棒グラフが適していま

## 【キーワード】

棒グラフ

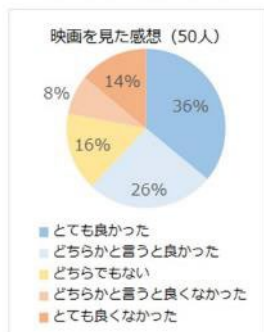
## グラフについて(さまざまなグラフ)(1)

【テキスト:19ページ】

### ■円グラフ(帯グラフ)

同じものの割合を見たいときは、円グラフを使います。

帯グラフも同じです。



割合をグラフに表すときは、データの数を必ず含めるようにしましょう。



44

### 【スライドの解説】

#### ■円グラフ

「映画を見た感想」など1つのことについての回答の割合を比較したいときは、円グラフや帯グラフを使います。

「映画を見た感想」の回答は、「とてもよかった」、「どちらかというよかった」などであり、大小関係にのみ意味がありますので順序尺度です。

### 【このページのポイント】

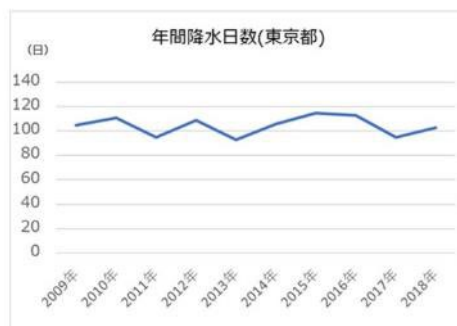
円グラフ、帯グラフ

### ■折れ線グラフ

数字を順番に、線で結んで表したものが折れ線グラフです。

「毎年の気温の変化」「毎日の体重の変化」など、主に関係のある2つの数について、連続して変化するデータの増減を見たいときに使います。

「毎年」「毎日」など、時間の経過とともに変化する数の様子を見ることに適しています。



### 【スライドの解説】

#### ■折れ線グラフ

気温や降水量などデータを時間の経過順に集めることがあります。時間の経過順に並んだデータを「時系列データ」と言います。時系列のデータのように時間と気温や降水量などの変化を見たい場合やデータ数が多いときには、数字を線で結んで表わす折れ線グラフで表現します。気温や降水量は、比例尺度です。

### 【このページのポイント】

時系列データのように、時間の経過とともに変化する「もの」の「変化」を見たいときに折れ線グラフで表わします。

時系列データを折れ線グラフで表す場合には、横軸に時間を、縦軸に時間の経過とともに変化する「もの」の値をとり、折れ線で結びます。

### 【キーワード】

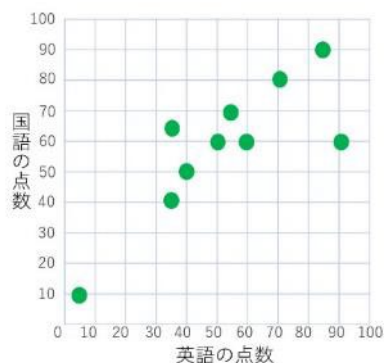
時系列データ、折れ線グラフ

### ■ 散布図

縦軸と横軸の2つの軸にそれぞれの要素(下の図では国語と英語の点数)をとり、データが当てはまるところに点を打って示すのが散布図です。

国語と英語の点数、身長と体重などの2つの量に関係があるかどうかをみるのに非常に便利なグラフです。また、2つの量には一方が増えともう一方はどのように変化するか、一方が減るともう一方はどのように変化するか、散布図から読みとることができます。

10人の国語と英語の成績の散布図



### 【スライドの解説】

#### ■ 散布図

国語と英語の点数、身長と体重などの2つの量にどのような関係があるかを把握するのに非常に便利なグラフです。散布図で表すと、2つの量の一方が増えともう一方はどのように変化するか、一方が減るともう一方はどのように変化するかを読みとることができます。

英語と国語という2つの得点データの散布図を描くには、横軸に英語の点数をとり、縦軸に国語の点数をとります。そして、生徒それぞれの国語と英語の点数が交わったところに点を打ちます。

### 【このページのポイント】

散布図は、2つの量に関係があるかどうかをみるのに非常に便利なグラフです。

### 【キーワード】

散布図、散布図の書き方

### 散布図の書き方

散布図を作成するには2ステップがあります。

- 1、関係を調べたいデータを縦軸と横軸にとる
- 2、プロット(打点)する

右に10人の国語と英語のテストの成績があります。この10人のデータを使った散布図を見てみましょう。そして国語と英語の点数にどのような関係があるのかを考えてみましょう。

10人の国語と英語の成績

	国語	英語
後藤	50点	40点
佐藤	80点	70点
山崎	10点	5点
鈴木	60点	50点
伊藤	60点	60点
清水	60点	90点
三上	90点	85点
遠藤	65点	35点
小川	70点	55点
戸田	40点	35点

### 【スライドの解説】

#### ■散布図の書き方 (2ステップ)

- step1. 関係を調べたい、英語と国語の点数といったようなそれぞれの2つのデータを用意し、横軸に英語の点数をとり、縦軸に国語の点数をとります。
- step2. 生徒それぞれの国語と英語の点数が交わったところに点を打ちます。

### 【このページのポイント】

散布図を書くには2つのステップがあります。

### 【キーワード】

散布図の書き方



## グラフについて(さまざまなグラフ)(3)

【テキスト:22ページ】

### 散布図の書き方

#### 1、関係を調べたいデータを用意し、縦軸と横軸のメモリを決める

散布図の縦軸を国語の点数とし、横軸を英語の点数とします。縦軸、横軸に0点から100点のメモリを10点刻みで書き込みます。

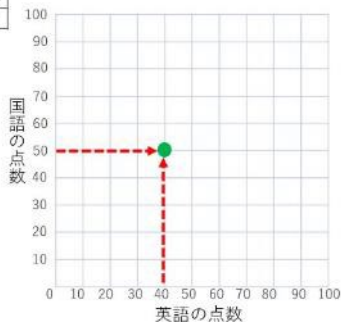
#### 2、プロット(打点)する

ごとうさんの成績を打点してみましょう。ごとうさんは国語50点、英語40点です。その2つが交わる場所に点を打ちます。

\*点を打つことをプロットといいます。

10人の国語と英語の成績

	国語	英語
後藤	50点	40点
佐藤	80点	70点
山崎	10点	5点
鈴木	60点	50点
伊藤	60点	60点



48

【スライドの解説】

### ■散布図の書き方(詳細)

散布図を書くには、関係を調べたい、英語と国語の点数といった生徒それぞれの2つのデータが必要です。横軸に英語の点数の目盛りとして0点から100点を10点刻みにとり、縦軸に国語の点数の目盛りとして0点から100点を10点刻みにとります。そして、生徒それぞれの国語と英語の点数が交わったところに点を打ちます。この例では、後藤さんの国語と英語の点数を打っています。点を打つことを「プロット」と言います。

【このページのポイント】

点を打つことを「プロット」と言います。

【キーワード】

プロット

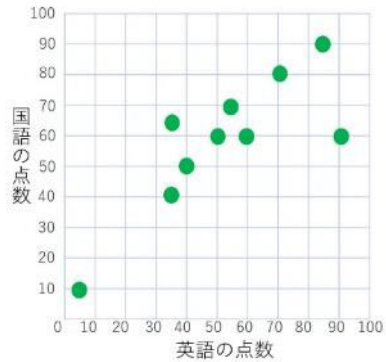
## グラフについて(さまざまなグラフ)(3)

【テキスト:22ページ】

### 散布図の書き方

プロットを繰り返し、10人の国語の点数と英語の点数を打点していくと、右のような散布図が完成します。

できた散布図を見ると、打点が右上がりになっており、英語の点数が良いほど、国語の点数も良くなる関係が見られます。



49

### 【スライドの解説】

#### ■散布図の書き方（詳細・続き）

10人の生徒それぞれの国語の点数と英語の点数をプロットしていくと、上のような散布図が完成します。

英語の点数が高いほど、国語の点数も高くなる関係が見られます。

### 【このページのポイント】

散布図は2つの量にどのような関係があるかをみるのに非常に便利なグラフです。

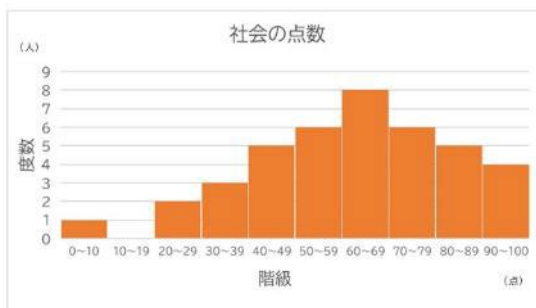
### 【キーワード】

散布図から読み取れること

### ■ヒストグラム

データをいくつかの区間に分け、横軸にデータの区間を、縦軸にその区間に含まれるデータの数をとって示すのがヒストグラムです。

棒グラフは質的データに用いますが、ヒストグラムは量的データに対して用います。データを分ける区間を適切に設定すれば、データのおおまかな分布を読みとることができます。



### 【スライドの解説】

#### ■ヒストグラム

ヒストグラムは、社会の点数のように、連続する数値のデータを表すグラフです。棒グラフと異なり、棒と棒の間が空いていません。

ヒストグラムは、データをいくつかの区間に分け、各区間に含まれるデータの数を表します。

### 【このページのポイント】

ヒストグラムは、連続する数値のデータ(量的データ)を表すグラフです。

棒グラフは国名、スイーツの種類など数値ではないデータ(質的データ)に用います。

### 【キーワード】

ヒストグラム、棒グラフ、量的データ、質的データ

## グラフについて(さまざまなグラフ)(3)

【テキスト:23ページ】

### ヒストグラムの書き方

ヒストグラムを作成するには2ステップがあります。

- 1、度数分布表を作成する
- 2、度数分布表に基づきヒストグラムを作成する

右に40人の社会のテストの成績があります。

このデータを使ったヒストグラムを見てみましょう。

40人の社会の成績

名前	社会の点数 (点)
青木	78
伊藤	83
大橋	64
神田	47
⋮	⋮
望月	68
山崎	92

51

### 【スライドの解説】

#### ■ヒストグラムの書き方(2ステップ)

- step1. 度数分布表を作成する。
- step2. 度数分布表に基づきヒストグラムを作成する。

### 【このページのポイント】

ヒストグラムを作成するには2ステップがあります。

### 【キーワード】

ヒストグラムの作成手順

## ヒストグラムの書き方

### 1、度数分布表を作成する

下の図のように点数をある区間ごとに区切って整理したものを度数分布表といいます。

階級とは、度数分布表のひとつの区間のことをいいます。

また、各階級のデータの個数を度数といいます。

度数分布表

階級	度数
90点~100点	4
80~89	5
70~79	6
60~69	8
50~59	6
40~49	5
30~39	3
20~29	2
10~19	0
0~10	1
合計	40

### 【スライドの解説】

#### ■ヒストグラムの書き方（詳細）

step1. 度数分布表を作成する

上の図のように社会科の点数0点から100点までを10等分します。10等分した各区間、0点以上から10点未満、10点以上から20点未満などを**階級**と言います。各階級に含まれる生徒の数(データ数)が**度数**です。**度数分布表**は階級と度数からなります。

### 【このページのポイント】

階級と度数からなる表が**度数分布表**です。

### 【キーワード】

度数分布表、階級、度数

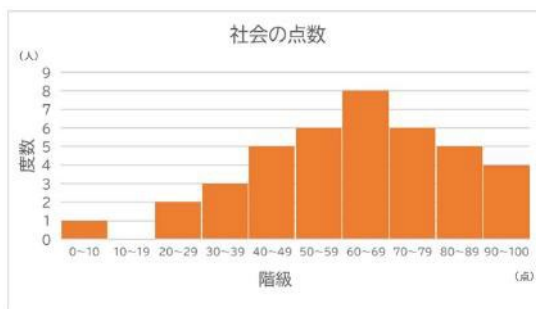
### ヒストグラムの書き方

#### 2、度数分布表に基づきヒストグラムを作成する

度数分布表に従ってヒストグラムを作成します。

ヒストグラムは一見棒グラフとよく似ていますが、棒グラフが高さで度数を表すのに対し、ヒストグラムは面積が度数を表しています。

そのため階級の幅が一定でない場合には高さに注意する必要があります。



完成したヒストグラムをみると、60~70点の人数が最も多く、おおよそ左右対称であることがわかります。

### 【スライドの解説】

#### ■ヒストグラムの書き方（詳細・続き）

step2. 度数分布表に基づきヒストグラムを作成する

度数分布表に従ってヒストグラムを作成します。ヒストグラムは一見棒グラフとよく似ていますが、棒グラフが高さで度数を表すのに対し、ヒストグラムは面積が度数を表しています。そのため階級の幅が一定でない場合には高さに注意する必要があります。

### 【このページのポイント】

ヒストグラムは一見棒グラフとよく似ていますが、棒グラフが高さで度数を表すのに対し、ヒストグラムは面積が度数を表しています。

### 【キーワード】

度数分布表、階級、度数、ヒストグラム

## データとは(データの種類)

【テキスト:25ページ】

クイズ!どのグラフだと分かりやすいか考えてみよう。

第1問:2020年時点において、日本にいる外国人の国籍の割合が知りたい場合、どの種類のグラフだと分かりやすいでしょうか?

第2問:2022年1月において、和歌山県の気温の変化を日毎に表す場合、どの種類のグラフだと分かりやすいでしょうか?

第3問:2021年における都道府県別のみかんの生産量を表す場合、どの種類のグラフだと分かりやすいでしょうか?

54

### 【スライドの解説】

#### ■データの種類とグラフについての問題

適したグラフを選んでみましょう。

第1問 割合を示すにはどのようなグラフが適しているでしょうか。

第2問 日付や時間とともに変化することを示すにはどのようなグラフが適しているでしょうか。

第3問 都道府県別のみかんの生産量を表すにはどのようなグラフが適しているでしょうか。

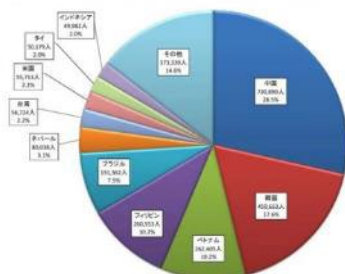
### 【このページのポイント】

示したことに応じた適切なグラフがあります。

## クイズ!どのグラフだと分かりやすいか当ててみよう!

第1問:2020年時点において、日本にいる外国人の国籍の割合が知りたい場合、どの種類のグラフだと分かりやすいでしょうか?

在留外国人の構成比(令和2年6月末)



出典:出入国在留管理庁

円グラフ  
(帯グラフ)

### 【スライドの解説】

#### ■答え

外国人の国籍という1つのもの「割合」を表すには、円グラフや帯グラフです。国籍は名義尺度です。

### 【このページのポイント】

1つのもの「割合」を表すには、円グラフや帯グラフが適しています。

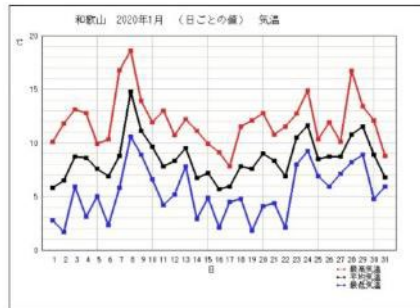


## データとは(データの種類)答え

【テキスト:25ページ】

クイズ!どのグラフだと分かりやすいか当ててみよう!

第2問:2022年1月において、和歌山県の気温の変化を毎日に表す場合、どの種類のグラフだと分かりやすいでしょうか?



折れ線グラフ

【スライドの解説】

### ■答え

和歌山県の気温の変化のデータは、時系列データです。

時系列データには折れ線グラフが適しています。

折れ線グラフとは、数字を順番に、線で結んで表したグラフです。

【このページのポイント】

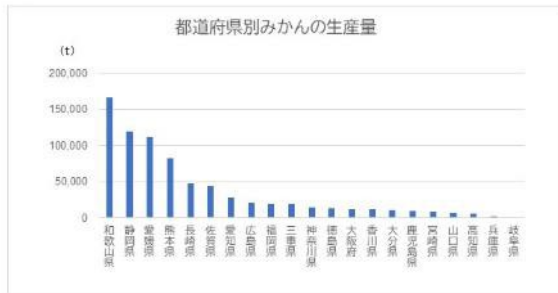
「毎年」「毎日」など、時間の経過とともに変化する数の様子を表すには折れ線グラフが便利です。

## データとは(データの種類)答え

【テキスト:25ページ】

クイズ!どのグラフだと分かりやすいか当ててみよう!

第3問:2021年における都道府県別のみかんの生産量を表す場合、どの種類のグラフだと分かりやすいでしょうか?



出典:農林水産省「作物統計」

棒グラフ

57

【スライドの解説】

### ■答え

都道府県別のミカンの生産量を表すグラフは、棒グラフです。

横軸には都道府県、縦軸にはミカンの生産量をとりま。

都道府県は名義尺度で質的データです。量的データではありませんので棒グラフとなります。

【このページのポイント】

横軸の都道府県は順序を入れ替えても問題ありません。上の棒グラフは、ミカンの生産量が多い都道府県で並んでいます。

## その他のデータ表現

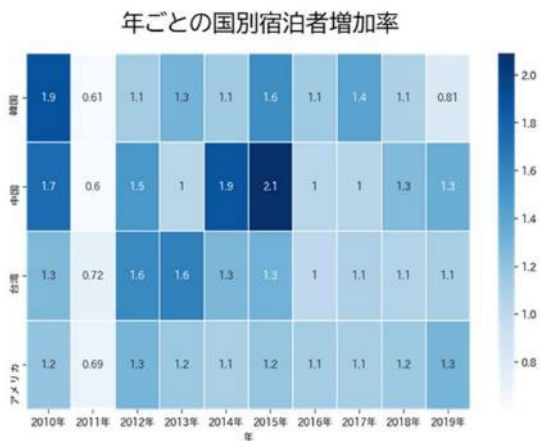
【テキスト:26ページ】

今回勉強したグラフ以外にも、データを表現する方法はたくさんあります。

たとえば、データの軸をオセロのマスのように変化させ、値の大きさを色で表現する「**ヒートマップ**」という方法があります。

軸を年月日などの時間にする事で、色の違いから値の変化を見ることができます。

また、ヒートマップは地図と相性が良く、人口や人の動きを表すのにも使われています。



58

### 【スライドの解説】

これまで学んだグラフ以外にも、データを表現する方法はたくさんあります。

### ■ヒートマップ

たとえば、データの値の大小を色の濃さで表現する「**ヒートマップ**」があります。

### 【このページのポイント】

「**ヒートマップ**」は値の大きさを色で表現します。

ヒートマップは地図と相性が良く、人口や人の動きを表すのにも使われています。

### 【キーワード】

ヒートマップ

## 統計ダッシュボードとは

【テキスト:27ページ】

### 「統計ダッシュボード」とは、

国や民間企業等が提供しているデータをグラフに加工してまとめたサイトです。

データの経年変化(時系列)や地域による差の比較等に役立ちます。統計ダッシュボード上にあるデータは、誰でも簡単に利用できます。

<https://dashboard.e-stat.go.jp/>

#### こんな時に便利！

- ・統計データを簡単にわかりやすいグラフにしたい
- ・世界(または地域別)の人口の推移をグラフで見たい
- ・Excelを使わずにブラウザ上でグラフ化したい
- ・大量のデータを定期的に利用したい



59

### 【スライドの解説】

#### ■統計ダッシュボード

「統計ダッシュボード」とは、国や民間企業等が提供しているデータをグラフに加工してまとめたサイトです。

データの経年変化(時系列)や地域による差の比較等に役立ちます。

### 【このページのポイント】

統計ダッシュボード上にあるデータは、誰でも簡単に利用できます。

### 【このページのポイント】

統計ダッシュボード：<https://dashboard.e-stat.go.jp/>

## 演習その1(統計ダッシュボードで調べてみよう) 【テキスト:28ページ】

コロナ禍の入帰国者数のデータを調べてみよう！

### 【お題1】

2020年4月～2022年9月までの日本における入(帰)国者数の推移を見てみよう！

1. 「グラフ検索」の入力欄に「入国者数」と入力して検索を押す。
2. 出入国者と表記のあるデータを開く。

The image shows a two-step process on a dashboard. On the left, the search interface has a search bar with '入国者数' (Immigration) entered and a '検索' (Search) button. A callout bubble points to the search bar with the text '①「入国者数」と入力' (Enter 'Immigration'). Another callout bubble points to the search button with the text '②「検索」を押す' (Click 'Search'). An arrow points to the right, showing the search results page. The results page has a table with '出入国者' (Immigration) selected. A callout bubble points to this selection with the text '「出入国者」を押してみよう。' (Click 'Immigration').

60

【スライドの解説】

### ■演習①

コロナ禍の入帰国者数を調べてみます。

【お題1】 2020年4月～2022年9月までの日本への入(帰)国者数の推移を見てみましょう。

1. 「グラフ検索」の入力欄に「入国者数」と入力して検索をクリックします。
2. 「出入国者」と表記のあるデータをクリックします。2021年4月からの棒グラフが現れます。

【このページのポイント】

実際に例題を使用しながら作成していきます。

【このページのポイント】

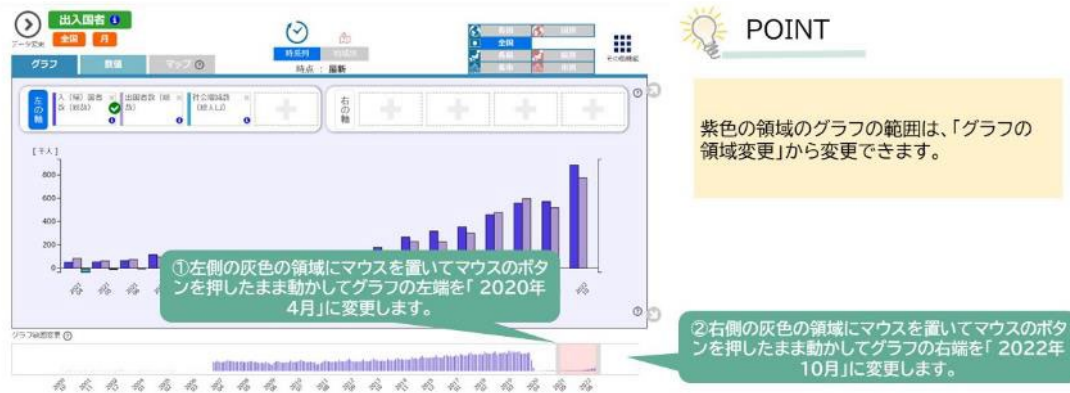
統計ダッシュボード、グラフ検索



## 演習その1(統計ダッシュボードで調べてみよう) 【テキスト:29ページ】

コロナ禍の入帰国者数のデータを調べてみよう！

3. グラフ範囲が「2020年4月～2022年9月」になるように調整する。



61

【スライドの解説】

### ■演習①(続き)

3. グラフ範囲が「2020年4月～2022年9月」になるように調整します。

【このページのポイント】

グラフの範囲の変更は、「**グラフ範囲変更**」からマウスで操作できます。

【このページのポイント】

統計ダッシュボード、グラフの範囲の変更

## 演習その1(統計ダッシュボードで調べてみよう) 【テキスト:30ページ】

コロナ禍の入帰国者数のデータを調べてみよう！

【お題2】

2019年10月～2022年10月までの日本における入(帰)国者数の推移を見てみよう！

1. 自己紹介
2. グラフを見て気づいたことを発表しましょう



62

【スライドの解説】

### ■演習①

【お題2】

2019年10月～2022年10月までの日本における入(帰)国者数の推移を見て、グループ内で気がついたことを話し合ってみましょう。

## グループに分かれて自己紹介をしよう

---

こんなことを話してみましよう！

- ・お名前(ニックネームでOK)
- ・住んでいるところ
- ・ここまでのセミナーの感想など

皆さんのことを  
教えてください！





## ブレイクアウトルームへの入り方

---



参加を押してね

## ブレイクアウトルームへの入り方



## グループワーク、お疲れさまでした！

どんなことを話せましたか？

グループで出た意見や感想を是非教えてください。

オンラインでご参加の皆さんは、

是非チャットで送ってくださいね！



66

### 【スライドの解説】

2019年10月～2022年10月までの日本における入（帰）国者数の推移を見て、グループ内で話し合われた内容の一部を参考に載せておきますので、みなさんもぜひ考えてみてください。

### ■受講者からのご意見

- ・2019年10月～2022年10月までの日本における入（帰）国者数の推移グラフから、日本への入（帰）国者数がコロナの影響を受けていることに気がつきました。

### 【このページのポイント】

日本における入（帰）国者数の推移のようなグラフのデータは、**時系列データ**と呼ばれます。時系列データを見たときは、コロナウイルス感染を防止するための水際対策(感染レベルが高い国からの入国を認めないなど)の政策や社会の動きを時系列にそってグラフに書き込んでみましょう。そこから気がつくことがあります。

### 【キーワード】

時系列データのグラフの見方

## ダッシュボードの使い方(例)

【テキスト:31ページ】



「データ変更」からデータ周期や階級を変えられます



「数値」から選択している期間の数値データが見られます



67

### 【スライドの解説】

#### ■ダッシュボードの使い方

- ・「データ変更」からデータの周期や階級を変えることができます。

データ周期を「年」にしてみましょう。2019年まで入出国数が増えてきていましたが、2020年に大きく減少したことが見えます。2021年にはさらに減少しています。

- ・「数値」から選択している期間の数値データが見られます。

2019年から2021年度までの数値です。2021年の減少が大きいことが分かります。

時点	入(帰)国者数(総数)【人】	出国者数(総数)【人】	社会増減数【人】
2019	51,409,309	51,229,818	208,783
2020	8,045,450	7,910,567	41,907
2021	904,325	1,084,938	-35,188

### 【このページのポイント】

「データ変更」からデータの周期や階級を変えることができます。

「数値」から選択している期間の数値データが見られます。

### 【キーワード】

ダッシュボードの使い方。

## ダッシュボードの使い方(例)

【テキスト:32ページ】



各国や都道府県、市区町村の比較が簡単にできます



都道府県別のデータがある場合は、マップも参照できます



68

### 【スライドの解説】

#### ■ダッシュボードの使い方の具体例

- ・各国や都道府県、市区町村の比較が簡単にできます
- ・都道府県別のデータがある場合は、マップも参照できます

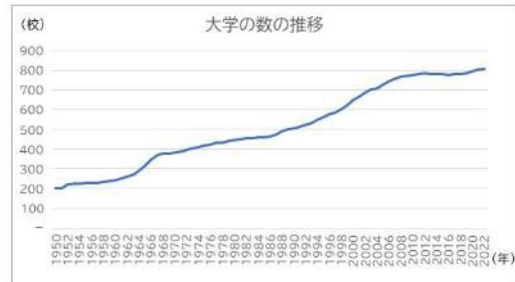
### 【このページのポイント】

ダッシュボードは、各国や都道府県、市区町村の比較が簡単にできます。  
年度を変えることで時系列の変化が見られます。

## データの見方

【テキスト:33ページ】

二つのグラフを見て、気づいた点を挙げてみよう。



気づいたことをメモしてみよう

69

### 【スライドの解説】

#### ■データの見方

データを見るときには、コツがあります。

大学の在籍者数を表した棒グラフと大学進学率を表した赤い色の折れ線グラフを同時に表わしたグラフと、大学の数の推移を表した折れ線グラフがあります。どちらも時系列データのグラフです。

大学の在籍者数の棒グラフは左側の縦軸が目盛(人)です。大学進学率の折れ線グラフは右側の縦軸が目盛(%)です。2021年の大学進学率は60%近いです。

この2つのグラフを見て、気づいたことを挙げましょう。

#### ■受講者からのご意見

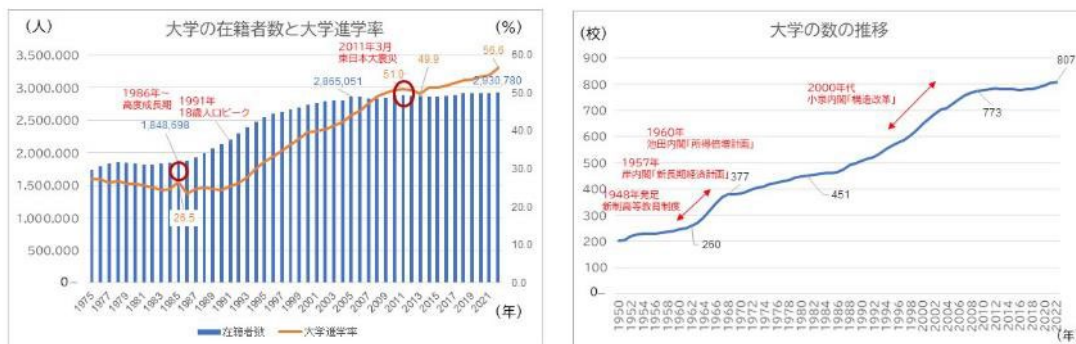
- ・ 大学進学率は上がっているのに、大学の在学者はそこまで増えていない。
- ・ 大学進学率と大学の数の増え方が似ている。

### 【このページのポイント】

二つのグラフを関連づけて、気がついたことを挙げてみましょう。

## データの見方

グラフに書き込みをしたり、その時にあった社会の動きなどを調べてみると、特徴や周期が見えてくることもあります。次の章では、グラフを読み解くときに重要なポイントなどを学びながら、実際のデータサイエンティストが行うデータの読み解きを一緒に体験していきましょう。



70

### 【スライドの解説】

#### ■データの見方

社会の動きや震災などをグラフに書き加えていくと、データの特徴や周期が見えてくることもあります。

### 【このページのポイント】

#### ■体験・データサイエンティストが時系列データをどのように読み解くか

- ・大学の在籍者数と大学進学率のグラフ（赤い字→その時の社会の動き）

- ①高度成長期の始まり(1986年)から大学の在籍者数も大学進学率も増加しています。大学の数のグラフを見ると、1962年から68年にかけて、大学の数の折れ線グラフの線の傾きが急になっています。
- ②第二次世界大戦が1945年9月に終結し、1947年から1949年に出生数が大きく増えました(第1次ベビーブーム)。その子たちが大学進学するのが1965年から1968年とすると、第1次ベビーブーム世代の進学先として、この時期に大学の数を100校以上増やしたと考えられます。
- ③1991年に日本の18歳人口が最大になっていますが、それ以降も大学の在籍者数は増え続け、2005年ごろから大きな変化がありません。しかし、大学進学率は上昇しています。少子化の影響という受講者からの意見がありました。その通りです。18歳人口が減少しているにもかかわらず大学在籍者にはあまり変化がないということは、18歳人口の大学進学率が上がっていることとなります。



次の章では、グラフを読み解くときに重要なポイントなどを学びながら、実際のデータサイエンティストが行うデータの読み解きを一緒に体験していきましょう。



## 第三章

なりきり！

データサイエンティスト



突然ですが・・・画像の食べ物、何と呼びますか？



73

【スライドの解説】

### ■データサイエンス体験

画像の食べ物、何と呼びますか？「ホットケーキ」、「パンケーキ」と呼ぶ人がいます。地域によっても異なるかもしれません。考えてみましょう。

## データサイエンス体験 (Googleトレンドとは) 【テキスト:36ページ】

### Googleトレンドとは

ある単語がGoogleでどれだけ検索されているかというトレンドをグラフで見ることができるツールです。アカウント登録不要かつ無料で利用でき、参照データはリアルタイムであるため、最新の検索動向が分かることが特徴です。過去1時間を始め、最大過去5年間、または2004年から現在までといった、幅広い期間におけるキーワードの検索推移を閲覧することが可能です。

<https://trends.google.co.jp/trends/?geo=JP>

こんな時に便利！

- ・話題のキーワードをいち早く調査したい
- ・複数のキーワードの推移を比較したい
- ・ある国における急上昇ワードを知りたい
- ・年別の検索キーワードのランキングを知りたい



74

### 【スライドの解説】

#### ■Googleトレンドとは

みなさんは、何かを調べようとするときに、Googleで検索すると思います。

ある単語がGoogleでどれだけ検索されているかを調べるサイトがあります。そのサイトをGoogleトレンドと言います。

### 【このページのポイント】

Googleトレンドとは、ある単語がGoogleでどれだけ検索されているかというトレンド(傾向)をグラフで見ることができるツールです。

### 【キーワード】

Googleトレンド : <https://trends.google.co.jp/trends/?geo=JP>

# データサイエンス体験 (Googleトレンドとは) 【テキスト:36ページ】

Googleトレンド: <https://trends.google.co.jp/trends/?geo=JP>



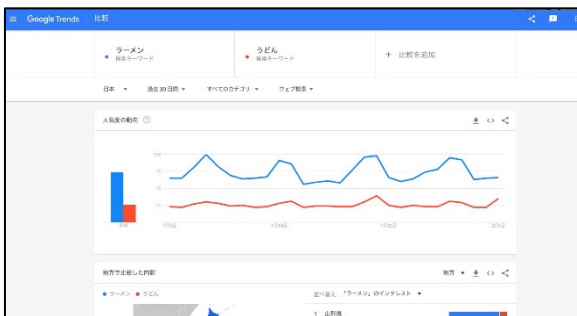
75

## 【スライドの解説】

### ■Googleトレンドの使い方

Googleトレンドにアクセスしてみましょう。

調べたい言葉を入力すると、その言葉がどれくらい検索されているかが1年間分表示されます。



「ラーメン」と「うどん」の検索数を比較してみましょう。全体的にラーメンの検索数が多いですが、「ラーメン」と「うどん」の折れ線グラフの波形は似ています。

## 【このページのポイント】

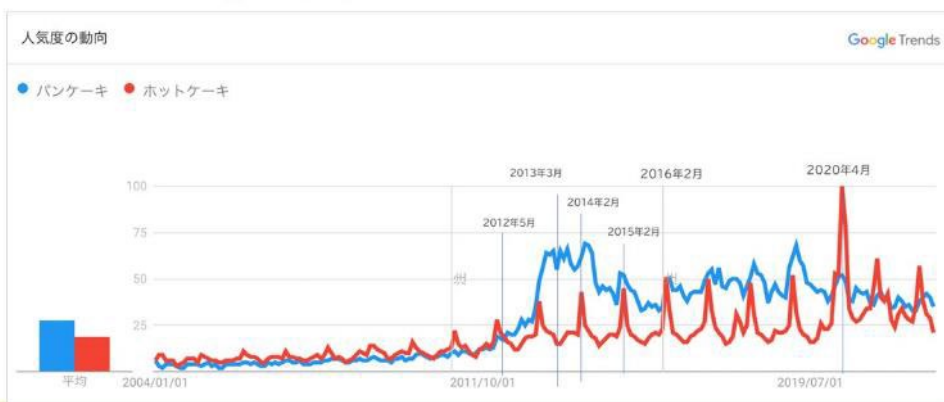
自分の調べたい、気になる様々な言葉をGoogleトレンドで入力してみましょう。データを比較することで、新しい発見があるかもしれません。

## 【キーワード】

Googleトレンド

Googleトレンドで2004年から現在までの「パンケーキ」と「ホットケーキ」を比較したデータ(検索総数の相対値)があります。

このデータからわかることは何でしょうか。



76

## 【スライドの解説】

### ■パンケーキとホットケーキのデータ比較

Googleトレンドを使って調べた、2004年から現在までの「パンケーキ」と「ホットケーキ」の検索数を比較したデータ（検索総数の相対値）があります。このデータから気がついたことを挙げましょう。

## 【このページのポイント】

Googleトレンドの「比較」をクリックして言葉を追加することで複数のもののトレンドの比較ができます。

## 【キーワード】

Googleトレンド、パンケーキ、ホットケーキ

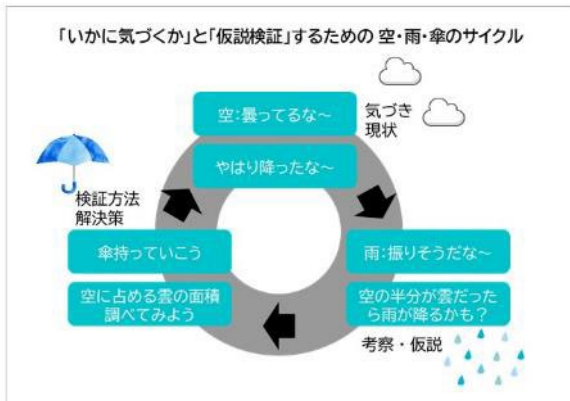
### ■受講者からのご意見

- ・パンケーキの検索数は、2015年から少なくなっている
- ・ほぼ毎年2月頃にホットケーキの検索数が増えている。
- ・2020年の4月にホットケーキの検索数が突出して増えている。
- ・2013年にパンケーキが多く調べられている。

## データサイエンス体験（空雨傘のサイクル）

【テキスト：38ページ】

与えられたデータやグラフを読み解く際には、与えられたデータやグラフからわかる「事実」と、その事実に対する「解釈」、そこから考えることのできる「解決策」を明確に区別する必要があります。



空

いまどんな状況か？事実を確認する

雨

事実をどう解釈するかを考える

傘

問題解決策、次の行動を提示する

77

### 【スライドの解説】

#### ■空雨傘サイクル

みなさんは、「空」を見て曇っていたならば、「雨」が降りそうだと推測し、「傘」を持っていこうと考えます。

データサイエンティストも、また同じことを行います。

与えられたデータやグラフから得ることができる「**事実**」と、その事実から考えられる「**解釈(仮説)**」、そして、そこから考えることのできる「**解決策(検証方法)**」を考えます。

### 【このページのポイント】

空雨傘のサイクル

### 【キーワード】

空、雨、傘

## データサイエンス体験 (空雨傘のサイクル)

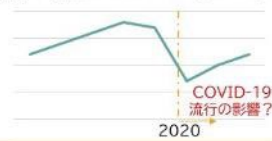
### ■ 特徴的な箇所を見つける

- ・グラフの突出したところ
- ・データがないところ
- ・値が大きいところ・小さいところ



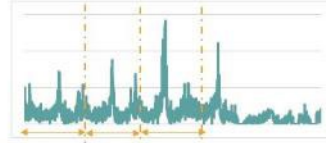
### ■ 線を引いて分類する

- ・時間や値の大きさを分割
- ・仮説や検証アイデアを考えやすい



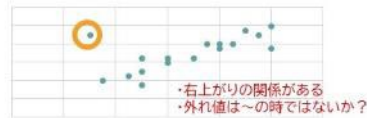
### ■ パターンや周期性を見つける

- ・曜日の周期
- ・季節の周期
- ・何かしらのイベント



### ■ 想定や仮説を書き出す

- ・こうなったはずでは？
- ・解決策にもつながりやすい



78

### 【スライドの解説】

#### ■空雨傘サイクル・具体例

事実である「空」から、「雨」が降りそうかどうかを推測するために注目すべき点を4つ学びます。

- ・ 特徴的な箇所を見つけること
- ・ パターンや周期性を見つけること
- Googleトレンドのラーメンとうどんの周期は似ていましたね。
- ・ 線を引いて分類するなどの事実の整理
- ・ 考察や仮説の書き出し

### 【このページのポイント】

データやグラフから読み取れることはもちろん、社会情勢や時事問題との関連性を見極めるところや逆説で考えてみることも非常に重要です。

### 【キーワード】

特徴的な箇所を見つけること、パターンや周期性を見つけること、線を引いて分類するなどの事実の整理、考察や仮説の書き出し

## データサイエンス体験（空雨傘シートの作成） 【テキスト:40ページ】

Googleトレンドで2004年から現在までの「パンケーキ」と「ホットケーキ」を比較したデータ(検索総数の相対値)をもとに、空雨傘シートを埋めてみましょう。

空(気づき)	雨(考察・仮説)	傘(打ち手)
<ul style="list-style-type: none"><li>● 現在日本ではパンケーキで検索する人が多い。</li></ul>	<ul style="list-style-type: none"><li>● 2012年5月前後に何らかのトレンドがあつて「パンケーキ」が定着したのではないか？</li></ul>	<ul style="list-style-type: none"><li>● 2012年前半に日本で話題になった「パンケーキ」に関連する情報を把握する</li></ul>

79

### 【スライドの解説】

#### ■空雨傘シートの作成

事実である「空」から、「雨」が降りそうかどうかを推測するために4つの注目すべき点に注目して、空雨傘シートを作成してみましょう。

### 【このページのポイント】

空雨傘シート

### 【キーワード】

特徴的な個所を見つけること、パターンや周期性を見つけること、線を引いて分類するなどの事実の整理、考察や仮説の書き出し



## データサイエンス体験(空雨傘シートの作成)

【テキスト:40ページ】

Googleトレンドで2004年から現在までの「パンケーキ」と「ホットケーキ」を比較したデータ(検索総数の相対値)をもとに、グループのみんなと一緒に空雨傘シートを埋めてみましょう。



80

### 【スライドの解説】

#### ■空雨傘シートの作成

グループで空雨傘シートを作成してみましょう。

### 【このページのポイント】

空雨傘シート

### 【キーワード】

特徴的な個所を見つけること、パターンや周期性を見つけること、線を引いて分類するなどの事実の整理、考察や仮説の書き出し

## グループワーク、お疲れさまでした！

---

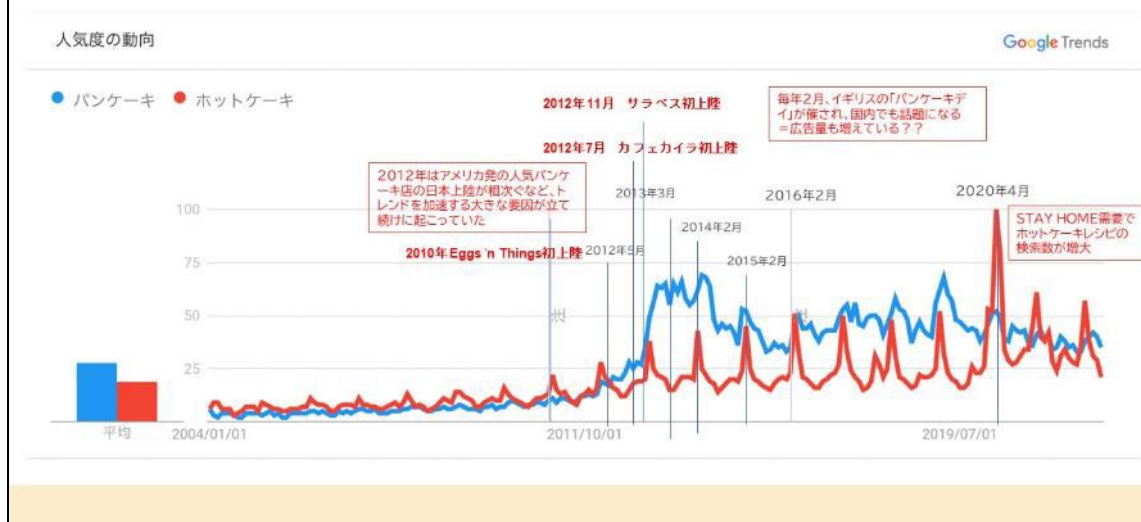
どんなことを話せましたか？

いくつかのグループに発表してもらいたと思います。

「発表したい！」という方は是非挙手をお願いします！



## データサイエンス体験 (データの見方 例)



### 【スライドの解説】

#### ■データの見方まとめ

##### 【例】

- ・トレンドはイベントに関係します。
- ・2012年にアメリカの人気パンケーキの店が日本に上陸しました。
- ・毎年2月に検索数が増えています。これは、毎年4月にイギリスで「パンケーキデイ」が開催され、日本でも話題になっていることが推測できます。
- ・2020年の4月のホットケーキの検索数がピークになっているのは、受講生のみなさんが気づいたように、コロナ禍により自宅でホットケーキを作ることになったからと推測できます。

データサイエンティストは、みなさんが今体験したことを行っています。

##### 【このページのポイント】

空雨傘シート

##### 【キーワード】

データサイエンティストの体験

## データサイエンス体験 (空雨傘シート回答例) 【テキスト:40ページ】

Googleトレンドで2004年から現在までの「パンケーキ」と「ホットケーキ」を比較したデータ(検索総数の相対値)をもとに、空雨傘シートを埋めてみましょう。

空(気づき)	雨(考察・仮説)	傘(打ち手)
<ul style="list-style-type: none"><li>● 現在日本ではパンケーキで検索する人の方が多。(若年層を中心に)</li><li>● 2012年5月からパンケーキの検索数がホットケーキを逆転した。</li><li>● 2014年3月を天井にパンケーキの検索数が減少してきている</li><li>● 緩やかながらホットケーキで検索するユーザーは増えてきている。</li><li>● 2020年4月、「ホットケーキ」検索数が増加している</li><li>● 毎年2月にホットケーキの検索件数が増えるトレンド傾向がある</li></ul>	<ul style="list-style-type: none"><li>● 2012年5月前後に何らかのトレンドがあって「パンケーキ」が定着したのではないか？</li><li>● パンケーキブームが、現在終わりつつある？ (代替となるトレンドが出てきているのでは？)</li><li>● 2月にホットケーキと関連のあるイベントがあるのではないか？(季節？)</li><li>● 2020年4月はこれまでのトレンドと異なるホットケーキに対する需要の増加があった？</li></ul>	<ul style="list-style-type: none"><li>● 2012年前半に日本で話題になった「パンケーキ」に関連する情報を把握する(メディアでの掲載、SNSによるトレンド、新たな飲食店の出店と人気状況)</li><li>● パンケーキと同等のファストフード等におけるトレンドの移り変わり(特に若年層)</li><li>● 国内・海外におけるホットケーキを食べる習慣があるイベントの調査(1月から3月の期間)</li><li>● 2020年4月のホットケーキミックスの売れ行き、トレンド状況を調べる</li></ul>

83

### 【スライドの解説】

みなさんの空雨傘シートをまとめました。

### 【このページのポイント】

空雨傘シート

### 【キーワード】

データサイエンティストの体験

## データサイエンティストとは

【テキスト:41ページ】

データサイエンティストとは、「データサイエンス力とデータエンジニアリング力をベースに、データから価値を創出し、ビジネス課題に答えを出すプロフェッショナル」のことを指します。(※ここでの「ビジネス」とは社会に役に立つ意味のある活動全般を指す)

### データサイエンティストの仕事:業界別

#### 例)医療現場



患者のカルテや画像診断のデータを使い、これからかかりそうな病気を予測したり、治療法の検討に役立つ分析を行っている

#### 例)食品製造工場



工場内の機械の状況をデータから把握し、故障やトラブルを事前に防いだり、不良品の検出をAIで行っている

#### 例)イベント運営



ライブやスポーツイベントなどの来場者データを活用し、人気のプログラムを把握したり、混雑を事前に予測するなど、より良いイベントを企画するための分析を行う

#### 例)スポーツチームの育成



試合中の選手の動きや試合展開のデータを分析し、勝率をあげるための戦略作りに役立てたり、選手の育成の手助けをする

84

### 【スライドの解説】

#### ■データサイエンティストの仕事・職業別

データサイエンティストはさまざまな職場で活躍しています。

たとえば、医療現場におけるデータサイエンティストの活躍は最初に説明しました。その他、工場では不良品を見つけ出すAIを開発しているデータサイエンティスト、ライブ会場等では照明を調整するデータサイエンティストが活躍しています。

また、冒頭でサッカーや野球のデータ活用についてお話をしましたが、スポーツの分野でもたくさんの方のデータサイエンティストが活躍しています。

### 【このページのポイント】

データサイエンティストと職場

### 【キーワード】

データサイエンティスト

# データサイエンティストとは

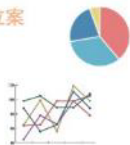
【テキスト:42ページ】

こんなお仕事もデータサイエンティストがしています！

## ■マーケティング担当

### データに基づくマーケティング施策の立案

- ・新商品の開発にアンケートデータを解析する
- ・過去の売り上げ実績から未来の売り上げを予測する



### 広告配信の計画

インターネット広告の配信実績を解析し、効果の高いバナー広告を選ぶ



## ■データ分析コンサルタント

### データ分析プロジェクトにおける管理監督

- ・どんなデータを使って解析すればいいかを監督する
- ・メンバーがスケジュール通り仕事ができているかを監督する



### AI導入コンサルティング(相談相手)

- ・どんなAI技術を使うと良いか選定し、依頼主に伝える
- ・AIに投入するデータを選ぶ



データサイエンティストの活躍の場は、年々広がっています！

85

## 【スライドの解説】

### ■データサイエンティストの仕事(続き)

アンケートを基にどのような新商品を開発すればよいかを考えるデータサイエンティストや効果の高い広告やその配信を考えるデータサイエンティストがいます。政府の施策立案をお手伝いしているデータサイエンティストもいます。

## 【このページのポイント】

データサイエンティストの活躍の場

## 【キーワード】

データサイエンティスト

## おわり

---

本日はありがとうございました！

この後しばらくはZoomをオープンにしているので、先生へ質問をしたり、みんなと会話したり自由にお使いください。会場の皆さんはこの後各端末でアンケートにお答えいただいた後、ご質問にお答えしたり交流したりする時間をご用意しています。

なお、このイベントは総務省統計局のサイトにて公開予定となっております！

オンラインでご参加の皆さんには、登録のメールアドレスにアンケートをお送りしています！



会場の皆さんは、PCのデスクトップにある【アンケート】をクリックして回答してね！