

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

2022.02.06 SUN 1PM



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Rejou

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
HPI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。

セミナープログラムについて

統計の基礎

60分



データ演習

75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

- 統計を学ぼう！
- 演習①「e-Stat」を使ってみよう
- 身近なデータについて学ぼう！
- グラフについて
- 演習②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Zoomをオープンにして
質問やお話出来る時間を作りたいと思っています。

先生について



菅 由紀子 (かん ゆきこ) 先生

株式会社Rejou (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



データサイエンティスト協会



WOMEN IN DATA SCIENCE

2016年9月にRejou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。

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に話すやりとりがあります。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な。





グループに分かれて自己紹介をしよう

- ・お名前（ニックネームでOK）
- ・住んでいるところ
- ・どんな気持ちで参加していますか？
- ・どんなことを学ぼうと思っていますか？

皆さんのことを
教えてください！



6

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



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

【スライドの簡単な解説】

統計とは、たくさんの情報を集めて特徴を捉え、未来を予測する学問のことです。

【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、予測した結果の要因も分かりやすいものになります。
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】

統：たくさんの〇〇を集める（この場合の〇〇は情報）

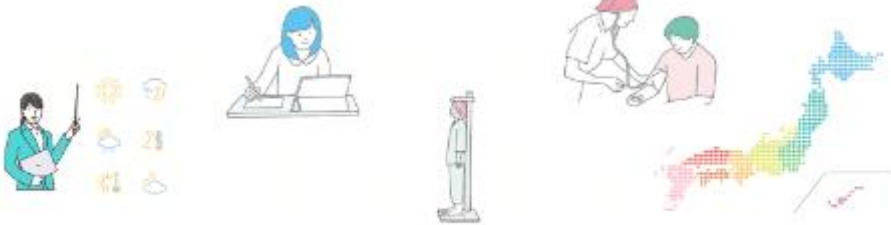
計：計算する・特徴をとらえる・未来を予測する

身近な統計について

【テキスト：2ページ】

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



【スライドの簡単な解説】

みなさんの身近なところに統計は存在し、生活に役立てられています。

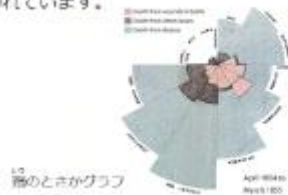
【このページのポイント】

たくさんのものから集められて計算されたデータは、私たちの生活を便利にするための大事な情報です。

ナイチンゲールは統計家だった！？

【テキスト：3ページ】

- ・ ナイチンゲールは1853年のクリミア戦争でイギリス軍の看護婦として従軍していましたが、実は統計を使って人命を助けた功績もあります。
- ・ 戦場でのけがよりも、伝染病や感染症で死亡する人の割合が高いこと、病院を清潔に保てば死亡者をおさえることができることをグラフを用いて統計的に証明したと言われています。



出典：看護省統計局HPより

12

【スライドの簡単な解説】

クリミア戦争の看護婦として有名なナイチンゲールは、統計を用いて人命を助けた一面も持っています。

【このページのポイント】

ナイチンゲールは、戦場において、戦いによるけがよりも感染症などの病気で死亡する人の割合が高いことをデータで証明しました。

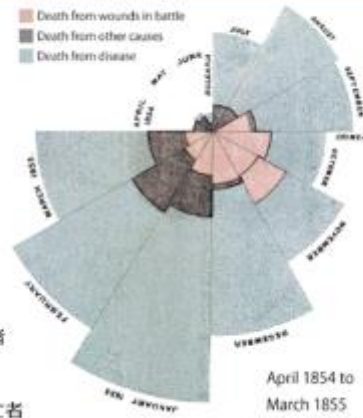
データを使って統計的に証明し、病院を清潔に保つよう働きかけたことで、当初は40%以上の人がなくなっていた病院の死亡率を、さまざまな対策によって2%まで下げることに成功しました。

鶏のとさかグラフ

【テキスト：3 ページ】

鶏のとさかグラフは円グラフの一種です。鶏のとさかグラフは、円グラフの角度は項目ごとに一定で、鼓の大小は半径で表される物を指します。

ナイチンゲールは鶏のとさかグラフを使い、クリミア戦争での負傷兵たちの死亡原因を、予防可能な疾病、負傷、その他に分けて視覚化しました。このグラフで、国の権力を持つ貴族（国会議員）に対して、感染症を防ぐための対策が必要であることを説明し、統計を用いて感染者を減少させることに貢献しました。



【スライドの簡単な解説】

「鶏のとさかグラフ」の解説。

青色は感染症での死亡者、赤色は戦闘による死亡者、（黒はその他）を表しています。グラフを見ると、青色（感染症による死亡者）が多いことが分かります。

【このページのポイント】

ナイチンゲールは、このグラフをもとに国の権力を持つ貴族（国会議員）に対して、感染症を防ぐための対策が必要であることを説明し、統計を用いて感染者を減少させることに貢献しました。

身近な統計データ

【テキスト：3ページ】



気象データの中に雨雲レーダーがあります。これは、にわか雨やゲリラ豪雨が発生しそうな時間や場所の予測に使われています。



テレビ番組の調査に使われる「視聴率」は、どのテレビ番組がどのくらい見られているかを調査した数値で、ここにも統計が使われています。視聴率の調査は、あらかじめ依頼を受けた家庭で、家族のだれが、いつ、どの番組を視聴したのかを記録することで行われています。



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。

14

【スライドの簡単な解説】

統計データは、身近なところにもたくさんあり、生活に役立てられています。例として、気象データ・テレビの視聴率・国勢調査などがあります。

【このページのポイント】

気象データは、過去のデータをもとに未来の天気を予測することに役立てられています。

テレビの視聴率は、どのくらいの人番組を観ているのかを把握して、テレビ局の人のお仕事に役立てられています。（日本の全部の家を調べるわけにはいけないため、一部の世帯を対象に調査が行われています。）

国勢調査は、日本に住んでいる全ての人を対象に5年に1度調査が行われており、直近では2020年に実施されています。データは行政の運営や街づくりに役立てられています。従来の調査票（紙）による回答に加えて、現在はインターネットでの回答が可能になりました。

【キーワード】

国勢調査：総務省が日本に住む全ての人を対象に、5年に1度行っている調査です。世帯ごとの家族の人数や年齢・性別・就学・就業状況などの調査を行います。集められたデータは、将来の人口予測、行政サービス、学術の基礎資料として役立てられています。

身近な統計（ケトラー指数）

【テキスト：3ページ】

BMI計算式

$$\text{BMI} = \frac{\text{体重kg}}{\text{身長m} \times \text{身長m}}$$

- ・ 1835年に開発
- ・ 1972年体脂肪との相関が高いことが明らかに
- ・ 1985年肥満度の代替指数としての有効性検証

体重	BMI
低体重（やせ）	18.5未満
通常体重	18.5以上 25未満
肥満	25以上 30未満



近代統計学の父と言われる
アドルフ・ケトラー

ナイチンゲールもケトラーの著書である「社会物理学」を読み、統計の基礎を学習したと言われている。

【スライドの簡単な説明】

ケトラー指数とは、身長と体重の値を計算式にすることで、ヒトの肥満度合いを判定することができるものです。

【このページのポイント】

ケトラー指数はアドルフ・ケトラーという人が開発しました。

ケトラー指数はBMIとも呼ばれていて、肥満かどうかを判定する目安になっています。



演習その1：e-Stat で調べてみよう

【テキスト：4ページ】

政府統計の総合窓口「e-Stat」とは

日本の行政機関が公表するデータを一つにまとめたサイトです。統計データを検索したり、表などで見やすくなることができます。

e-Stat上にあるデータは、誰でもダウンロードしたり使ったりすることができます。



政府統計の総合窓口「e-Stat」

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

17

【スライドの簡単な解説】

政府統計の総合窓口「e-Stat（イースタット）」を使って自分の知りたい統計データを調べる方法を勉強します。

【このページのポイント】

「e-Stat」は日本の行政機関が公表するデータを1つにまとめたサイトです。

各府省等が独自に運用するWebサイトに散在していた統計関係情報を集約し、統計結果を誰でも利用しやすいかたちで提供することを目的に作られました。

統計データを探す方法としては、「すべてから探す」、「分野や組織から探す」、「キーワードで探す」などがあります。

統計データを視覚的にわかりやすい形にしたグラフや時系列表なども多数掲載されています。

【キーワード】

「e-Stat」…日本の行政機関が公表する統計データを検索したり表やグラフにしたりすることができるサイト。

演習その1：e-Stat で調べてみよう

【テキスト：5ページ】

実際に確かめてみよう！

まず初めに、「**落花生の生産量が最も多い都道府県**」を調べてみましょう。

A screenshot of the e-Stat website. The header includes the e-Stat logo and the text '統計で見る日本'. Below the header, there are navigation links. The main content area has two sections: '統計データを探す' (Search for statistical data) and '統計データを活用する' (Use statistical data). In the '統計データを探す' section, there are buttons for 'すべて' (All), '分野' (Field), and '地域' (Region). Below these, there is a search bar with the text 'キーワード検索' and a '検索' (Search) button. The '統計データを活用する' section has buttons for 'グラフ' (Graph), '時系列表' (Time series table), '地図' (Map), and '地域' (Region).A close-up of the search bar and search button from the e-Stat website. The search bar contains the text '落花生' (Peanut). The search button is labeled '検索'. A hand icon is pointing at the search button. There are two callout boxes: one pointing to the search bar with the text '「落花生」と入力' (Enter 'Peanut'), and another pointing to the search button with the text '「検索」を押す' (Press 'Search').

演習その1：e-Stat で調べてみよう

【テキスト：5 ページ】

実際に確かめてみよう！

検索結果の中から、「豆類の作付け面積および生産量/2/落花生/都道府県別生産状況」を選択しよう。

農水省 農産物 統計	農水省 農産物 統計	2020年1月	2021-11-17	EXCEL
貿易統計 (輸入) / 2 / 落花生 (調製したもの、炒ったもの)				
農水省 農産物 統計	農水省 農産物 統計	2021年9月	2021-11-11	EXCEL
貿易統計 (輸入) / 2 / 落花生 (調製したもの、炒ったもの)				
農水省 農産物 統計	農水省 農産物 統計	2005年	2011-07-11	EXCEL
豆類の作付け面積及び生産量 / 2 / 落花生 / 都道府県別生産状況				
農水省 農産物 統計	農水省 農産物 統計	2005年	2011-07-11	EXCEL
豆類の作付け面積及び生産量 / 2 / 落花生 / 品種別生産状況				

「豆類の作付け面積および生産量/2/落花生/都道府県別生産状況」のテキストを押す

【テキスト：6ページ】

EXCELファイルをダウンロードして、最も収穫量の多い都道府県を探してみよう。

最も収穫量が多い
都道府県はどこか
探してみよう！

演習その1：e-Stat で調べてみよう

【テキスト：7ページ】

実際に確かめてみよう！

次は、「**みかんの収穫量が最も多い都道府県**」を調べてみよう！

「分野」から、都道府県別のみかんの収穫量を把握する集計を行ってみよう。



The image shows a sequence of steps on the e-Stat website. On the left, the '分野' (Field) button is highlighted with a red box and a hand cursor, with a callout saying '「分野」を押す' (Press 'Field'). An arrow points to the right, showing a dropdown menu for '農林水産業' (Agriculture, Forestry, and Fisheries). In this menu, the '作物統計調査' (Crop Statistical Survey) option is highlighted with a red box and a callout saying '「作物統計調査」を押す' (Press 'Crop Statistical Survey').

【テキスト：8ページ】

「データベース」の下を選択し、「作況調査（果樹）」にある「長期累年」の下に表示されている「年次」を押します。

演習その1：e-Stat で調べてみよう

【テキスト：8ページ】

実際に確かめてみよう！

「みかん」の「⇒DB」（データベース形式）を押して、データを表示させる。

データセット一覧

一覧をばで表示

品名統計名	作物統計調査	年度
統計統計名	作物統計調査	
統計小種別	作物調査（国産）	
統計小種別	果樹調査	
統計小種別	果樹	

品名	統計名	年度	日付（更新）	DB形式・ダウンロード
1	みかん	2018年	2019-05-20	⇒DB 一覧
2	なつあかり	2018年	2019-05-20	⇒DB 一覧
3	はっさく	2018年	2019-05-20	⇒DB 一覧
4	いよかん	2018年	2019-05-20	⇒DB 一覧

「みかん」の
「⇒DB」を押す

演習その1：e-Stat で調べてみよう

【テキスト：9ページ】

実際に確かめてみよう！

アクセスすると、表頭にみかんの結果樹面積や収穫量など、表側には年次が表示されている集計表が出てきます。このような集計表のことをクロス集計表といいます。現在のレイアウトでは都道府県ごとの収穫量を把握することができないため、レイアウトを変更する必要があります。

「レイアウト設定」
を押す

都道府県	結果樹面積	収穫量	年次
北海道	11,400	854,100	777,800
青森県	43,200	277,800	884,800
岩手県	42,800	274,700	780,200
宮城県	44,700	489,800	854,400
秋田県	44,800	414,400	267,200
山形県	45,300	319,400	819,800
福島県	46,100	284,800	258,100
茨城県	47,400	1,019,800	919,400
栃木県	48,200	204,100	887,800
群馬県	49,300	1,089,800	958,500
埼玉県	55,200	881,800	214,200
千葉県	57,500	1,132,800	1,059,800
東京都	54,200	1,089,800	858,800

演習その1：e-Stat で調べてみよう

【テキスト：9ページ】

実際に確かめてみよう！

最初のレイアウトで【ページ上部】に表示されている「全国、都府県、全国農業地域」と【行】に表示されている「時間軸（年次）」を入れ替えると、クロス集計を行うことができます。

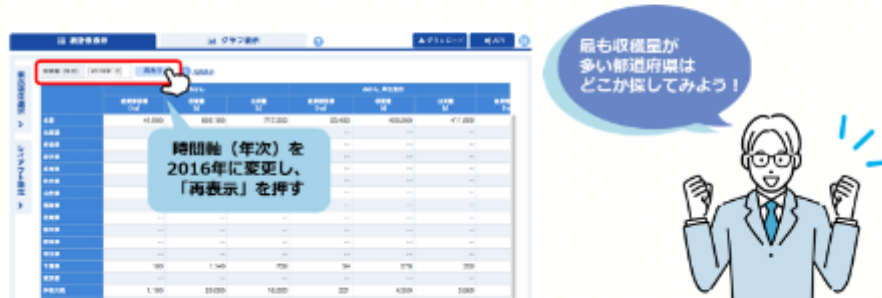
The image contains two screenshots of the e-Stat web application interface. The left screenshot shows a list of categories on the left side of the page. A red double-headed arrow indicates a swap operation between the top category '全国、都府県、全国農業地域' and the bottom category '時間軸（年次）'. A callout bubble points to the top category with the text: 「全国、都府県、全国農業地域」が行に、「時間軸（年次）」がページ上部になるようにクリックしたまま移動させる. The right screenshot shows the same interface after the swap. A callout bubble points to the '設定して表示を更新' (Set and update display) button with the text: 入れ替えたら「設定して表示を更新」を押す. The page number '25' is visible in the bottom right corner of the right screenshot.

演習その1：e-Stat で調べてみよう

【テキスト：10ページ】

実際に確かめてみよう！

これで都道府県ごとの収穫量を見ることができるようになりました。



【スライドの簡単な解説】

「e-Stat」にアクセスし、みかんの収穫量が最も多い都道府県を調べます。

【このページのポイント】

「e-Stat」にアクセスし、みかんの収穫量が最も多い都道府県を調べます。

その後、グループワークでグループのみんなでみかんの収穫量の調べ方や結果について共有します。



第二章

データについて



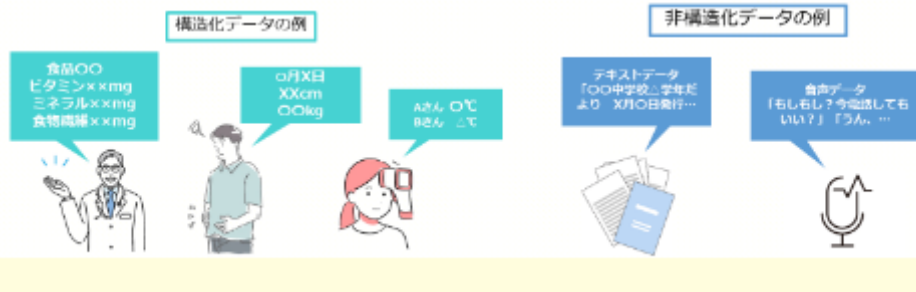
28

データとは

【テキスト：12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。



【スライドの簡単な解説】

さまざまな形式で記録された情報を、総合して「データ」といいます。

成績表や店の売上金額などのように数字で表せられるものを「構造化データ」といい、動画配信サイトで視聴ができるような「音声」や「映像」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。

【このページのポイント】

紙に書かれた情報は、コンピュータが読み取れる状態に電子化することで「データ」と呼ばれるようになります。

数字、記号のほか文字や画像、映像なども「データ」の1つであり、日常生活のいたるところで、データは取得されています。（駅の改札、病院での検温など）

【キーワード】

構造化データ…成績表や店の売上金額などのように数字で表せられるデータ

非構造化データ…「音声」や「映像」など、そのままでは表形式に変換できないデータ

データとは（わたしたちの身近にあるデータ）【テキスト：13ページ】

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年に一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あらゆることを調査しています。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、企業活動や大学の調査などにも役立てられています。



31

【スライドの簡単な解説】

国勢調査は、5年に1回行われる国の最も重要な統計調査であり、日本に住んでいる人全員を対象にした大規模な調査です。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。

【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、住みよいまちづくり、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）

データとは（わたしたちの身近にあるデータ）【テキスト：13ページ】



【スライドの簡単な解説】

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。交通機関や街中の映像、動画配信サービスなどもその1つです。

動画サービスなどでは、いつ誰が何を見たかのデータを「ログ」と呼び、それらを計算し、個人の好きなものを予想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（社会における統計の活用）

【テキスト：15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。

33

【スライドの簡単な解説】

生活の中で使われるデータで、特にみなさんに身近なもの1つに「交通」「天気」に関するデータがあります。

渋滞を避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。

データとは（社会における統計の活用）

【テキスト：15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。



【医療データを使って疾病を診断】

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。

34

【スライドの簡単な解説】

病院などにおいても、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。

例えば、過去にガンになった人のデータは、これから病気を防いだり治療をしたりするためにとても重要な役割をもっています。

【このページのポイント】

データを集めることで、未来の人々の健康や安全を守るために活用することができます。

データとは（データの種類）

【テキスト：16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。



【スライドの簡単な解説】

データは、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：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：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注意が必要です。

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

データとは（データの種類）

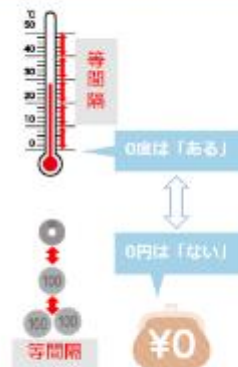
【テキスト：16ページ】

■ 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。



【スライドの簡単な解説】

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、差の大きさに意味を持つ。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：17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ピザのサイズ



②来場者の数



③時刻



【スライドの簡単な解説】

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】

- ①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。
- ②・③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。

データとは（データの種類）

【テキスト：17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ピザのサイズ

 S
 M
 L

順序尺度

②来場者の数



比例尺度

③時刻






間隔尺度

【スライドの簡単な解説】

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のピザのサイズは、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。

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」ということになります。よって正解は比例尺度です。

③の時刻は、例えば0時の場合。これは時間が存在しないという意味ではありません。よって間隔尺度であることが分かります。

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。



39

【スライドの簡単な解説】

データの基本的な特性を表す基本統計量の解説。

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、データの特性は「代表値」と「散布度」の二つに分けられます。

【このページのポイント】

データの基本的な特性は、平均のようにデータの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」の二つに分けられます。

【キーワード】

「代表値」…データの分布全体を一つの数で表す。

「散布度」…データのばらつきの大きさを表す。

データの特徴を見てみよう

【テキスト：19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・・・

40

【スライドの簡単な解説】

データの取り扱い方についての解説。

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。

データの特徴を見てみよう

【テキスト：19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

AN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見やすくするためのポイントをいくつか紹介します。

【スライドの簡単な解説】

データの取り扱い方についての解説。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、医療の画像からガン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

データの特徴を見てみよう（平均値）

【テキスト：19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	$(140 + 135 + 145 + 155 + 160 + 150 + 165) \div 7 = 150(\text{cm})$
----	---

42

【スライドの簡単な解説】

たくさんのデータから特徴をとらえる方法の1つである「平均」についての解説。

【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）

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。



43

【スライドの簡単な解説】

たくさんのデータから特徴をとらえる方法の1つである「中央値」についての解説。

【このページのポイント】

左の図のように5種類のスイーツが値段の安いものから高いものの順に並んでいます。この場合は真ん中にくるチーズケーキの200円が中央値です。

また、データが偶数個の場合は真ん中同士を足して2で割ると中央値が求められます。

右の図のようにスイーツが6種類並んでいる場合は、真ん中である3番目(チーズケーキ200円)と4番目(ロールケーキ400円)を足した数字を2で割った数値が中央値です。

つまり $(200 + 400) \div 2 = 300$ となるので中央値は300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をいう。

ワーク：（中央値を出してみよう！）

【テキスト：21ページ】

次の表はとある中学校の2年生でハンドボール投げをした結果です。

中央値を求めてみましょう。

記録（m）	
男子	20m, 25m, 18m, 21m, 26m, 22m, 17m, 15m, 30m
女子	14m, 15m, 12m, 10m, 17m, 20m, 8m, 13m, 12m, 17m



男子の中央値： _____ m

女子の中央値： _____ m

44

【スライドの簡単な解説】

（ワーク）「中央値」を出してみよう。

【このページのポイント】

男子と女子で計測した人数が違うことにも注目してみましょう。

ワーク：（中央値を出してみよう！） 答え

【テキスト：21ページ】

①小さい順に並べる

記録 (m)										
男子	15m	17m	18m	20m	21m	22m	25m	26m	30m	
女子	8m	10m	12m	12m	13m	14m	15m	17m	17m	20m

小 大

②真ん中（3個目と4個目）の平均を求める

記録 (m)										
男子	15m	17m	18m	20m	21m	22m	25m	26m	30m	
女子	8m	10m	12m	12m	13m	14m	15m	17m	17m	20m

(13m+14m)÷2=13.5m

【スライドの簡単な解説】

「中央値」の問題の答えを解説します。

【このページのポイント】

まず、小さいものから大きいものの順に並べ直します。

男子：データの個数は奇数(9個)であり、記録を小さい順に並べると、

15m, 17m, 18m, 20m, 21m, 22m, 25m, 26m, 30m

→真ん中に21mの記録がくるため、中央値は21mです。

女子：データの個数は偶数(10個)であり、記録を小さい順に並べると、

8m, 10m, 12m, 12m, 13m, 14m, 15m, 17m, 17m, 20m

→真ん中にくる13mと14mの記録を足して2で割ると中央値がでます。

つまり $(13+14) \div 2 = 13.5$ となるので、中央値は13.5mです。

データの特徴を見てみよう（最頻値）

【テキスト：22ページ】

最頻値は、データの中でもっとも登場回数が多い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
鯖巻	100
たまご	100
いなり寿司	100
いわし	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
鯖巻	100
たまご	100
いなり寿司	100
いわし	150
サーモン	150
たこ	150
えんがわ	150
えび	200
しめさば	200
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

【スライドの簡単な解説】

たくさんのデータから特徴をとらえる方法の1つである「最頻値」についての解説。

【このページのポイント】

値段がいくらのお寿司が一番多いのかを調べます。

(例1)では、100円のお寿司が一番多いことがわかります。

つまりここでの最頻値は100円です。

(例2)では、100円と150円のお寿司が1番多いことがわかります。

1番が2つの場合でも、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



47

【スライドの簡単な解説】

代表値についての解説。

【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。
もっと統計学やビッグデータについてたくさん勉強するようになると、ここに書いてある「最小値」や「最大値」もチェックします。
これらを合わせて、そのデータを代表する値として「代表値」といいます。

【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…データを説明する「平均」「中央値」「最頻値」などの総称。



第三章

データの可視化



可視化について

【テキスト：24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。



50

【スライドの簡単な解説】

- ・データの可視化についての解説。

【このページのポイント】

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」に変えることを可視化といいます。

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。

この章では、グラフの種類とその特性、書くためのコツなどを勉強していきます。

【キーワード】

可視化…データを「目で見てわかりやすい形」に変えること。

グラフについて

先ほど見た表です。
表をばっと見て、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	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・・・

51

【スライドの簡単な解説】

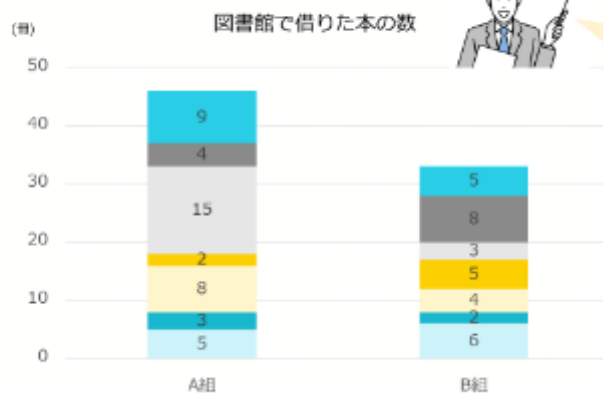
図書館で借りた本の冊数(2クラス分)を比較してみて、どちらのクラスが多く本を借りているかを考えます。

【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、数字からなんとなくA組の借りた本の数が多いことがわかります。

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

52

【スライドの簡単な解説】

前のページの表を積み上げ棒グラフにして見てみます。

【このページのポイント】

グラフで表すことによって、表で見るよりもさらにはっきりとどちらのクラスが多く本を借りているかが分かりやすくなります。

【キーワード】

積み上げ棒グラフ…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。

グラフについて

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



見ただけではよく
分からない・・・！

たくさんのデータ
(ビッグデータ)
だったら？

53

【スライドの簡単な解説】

図書館で借りた本の冊数(9クラス分)を比較してみて、どのクラスが多く本を借りているかを考えます。

【このページのポイント】

データがたくさんあると、ぱっと見ただけでは比較が難しいことがわかります。

グラフについて（さまざまなグラフ）（1） 【テキスト：25ページ】

■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。



55

【スライドの簡単な解説】

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。

【このページのポイント】

「面積」「点数」「身長」など、同じものについて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す。

グラフについて（さまざまなグラフ）（1） 【テキスト：25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。



56

【スライドの簡単な解説】

円グラフや帯グラフは、割合の大きさをグラフに表したものです。

【このスライドのポイント】

同じものについての割合を見たいときは円グラフ・帯グラフを使います。

グラフについて（さまざまなグラフ）（2） 【テキスト：26ページ】

■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、変化の様子を見るときに使います。

「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。



57

【スライドの簡単な解説】

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。

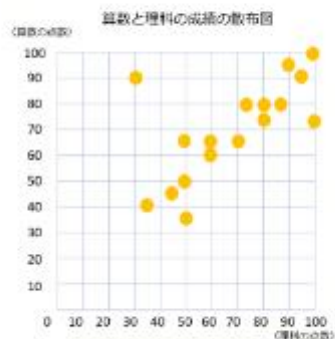
グラフについて（さまざまなグラフ）（3）

【テキスト：27ページ】

■ 散布図

縦軸と横軸の2つの軸にそれぞれの要素（下の図では算数と理科の点数）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。

算数と理科の点数、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。



【スライドの簡単な解説】

散布図とは、縦軸と横軸にそれぞれの要素をとり、データが当てはまるところに点を打って示したものです。

【このページのポイント】

2つの要素にどのような関係があるかを読み取るのに便利なグラフです。

グラフについて（さまざまなグラフ）（3）

【テキスト：27ページ】

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人のデータを使って散布図を作ってみましょう。
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
ごとう	50点	40点
さとう	80点	70点
やまざき	10点	5点
すずき	60点	50点
いとう	60点	60点
しみず	60点	90点
みかみ	90点	85点
えんどう	65点	35点
おがわ	70点	55点
とだ	40点	35点

59

【スライドの簡単な解説】

散布図の書き方説明です。

【このページのポイント】

散布図を作成する際は、まず関係を調べたい要素を縦軸と横軸にとります。次にデータが該当する場所にプロット（打点）します。

グラフについて（さまざまなグラフ）（3）

【テキスト：28ページ】

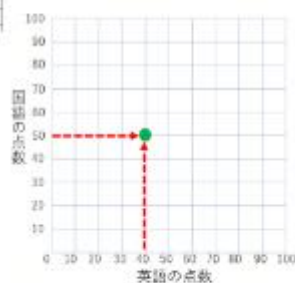
散布図の書き方

1. 関係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。

20人の国語と英語の成績

	国語	英語
ごとう	50点	40点
さとう	80点	70点
やまざき	70点	50点
すずき	60点	90点
いとう	80点	80点
1人ずつ	40点	60点



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。

【スライドの簡単な解説】

前のページの説明に習い、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】

ここでは、ごとうさんの点数をプロットしてみます。

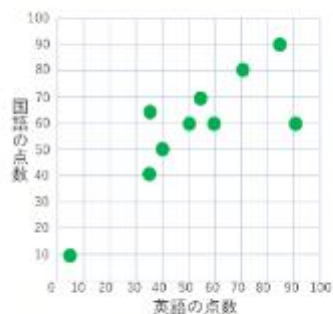
グラフについて（さまざまなグラフ）（3）

【テキスト：28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、打点が右肩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。



61

【スライドの簡単な解説】

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】

全員の点数をプロットするとこのような散布図が完成します。

散布図を見ると右肩上がりにプロットされており、英語の点数が良いほど、国語の点数が良くなるという傾向がみられることが分かります。

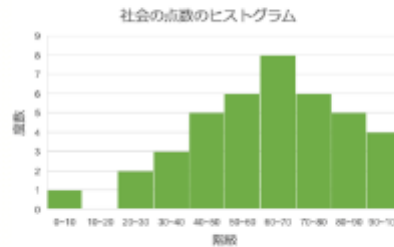
グラフについて（さまざまなグラフ）（4）

【テキスト：29ページ】

■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。



62

【スライドの簡単な解説】

ヒストグラムを作成する際は、データをいくつかの区間に分けていきます。
横軸にデータの区間、縦軸にその区間に含まれるデータの数をとって示します。

【このページのポイント】

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。

【キーワード】

質的データ…好きなスポーツ、血液型、自動車のナンバーなど、単に分類や種類を区別するためのデータや、順位、学年など順序に意味があるデータ。

量的データ…枚数、身長、金額など、数値で推し測ることができ、数字の大小に意味をもつデータ。

グラフについて（さまざまなグラフ）（4）

【テキスト：27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。

このデータを使ってヒストグラムを作ってみましょう。

40人の社会の成績

社会の点数（点）	人数（人）
100	2
95	1
90	1
⋮	⋮
5	1
0	0
合計	40

63

【スライドの簡単な解説】

ヒストグラムの書き方説明です。

【このページのポイント】

ヒストグラムを作成する際は、まず度数分布表を作成し、定めたデータ区分に何名が割り振られるのかを明らかにします。

その後、度数分布表に基づいてグラフを作成していきます。

【キーワード】

度数分布表…データがどのように散らばっているかを示す表のこと。値自身や値の階級に対して、その範囲にいくつデータがあるかの頻度（度数）を表す。

グラフについて（さまざまなグラフ）（4）

【テキスト：30ページ】

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
点以上 点未満 90~100	4(100点を含む)
80~90	5
70~80	6
60~70	8
50~60	6
40~50	5
30~40	3
20~30	2
10~20	0
0~10	1
合計	40

64

【スライドの簡単な解説】

前のページの説明に習い、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】

度数分布表の作成にあたり、今回は点数の階級を10点ずつで区切ります。

度数分布表の1つの区間のことを階級といい、各階級のデータの個数を度数といいます。

グラフについて（さまざまなグラフ）（4）

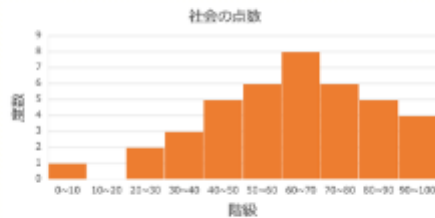
【テキスト：30ページ】

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。



65

【スライドの簡単な解説】

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】

作成した度数分布表に従い、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で度数を表します。

そのため、ヒストグラムの場合は階級が一定であるかも大切なポイントであり、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。

可視化することのメリット

【テキスト：31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(分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

	1951年の降水量の 平均 (mm)	1996～2020年の 降水量の平均 (mm)
1月	28.9	36.2
2月	82.8	49.4
3月	89.8	82.9
4月	99.9	90.9
5月	113.5	112.6
6月	211.5	166.2
7月	436.9	127.9
8月	25.9	97.2
9月	255.9	142.2
10月	309.0	95.4
11月	62.9	53.3
12月	27.8	41.3

グラフで表すと・・・

一目で特徴が分かる！



参照元 <https://www.jma.go.jp/jma/index.html>

(6)

【スライドの簡単な解説】

可視化することのメリットについての解説。

【このページのポイント】

可視化することのメリットとしては、

- ・「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」

の2点があります。

グラフ当てクイズ

【テキスト：32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？

【スライドの簡単な解説】

学習したグラフについてのクイズです。

【このページのポイント】

第1問は、「割合が知りたい」という点に注目して考えてみましょう。

第2問は、「変化を日ごとに表す」というのがポイントです。

第3問は、「都道府県ごとの生産量（数量）を比較できる」グラフを考えてみましょう。

グラフ当てクイズ 答え

【テキスト：32ページ】

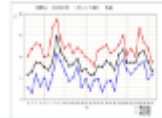
クイズ！どのグラフだと分かりやすいか考えてみよう！

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？



円グラフ
(帯グラフ)

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



折れ線グラフ

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？



棒グラフ

【スライドの簡単な解説】

グラフ当てクイズの答えを解説します。

【このページのポイント】

第1問：割合の大きさを表す円グラフもしくは帯グラフが正解です。

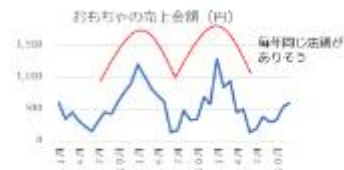
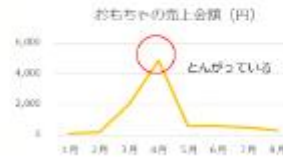
第2問：時間の経過とともに変化する数の様子を表す折れ線グラフが正解です。

第3問：同じものについての統計データの値を比較できる棒グラフが正解です。

グラフを見るときポイント

【テキスト：33ページ】

- 特徴的なところ（とんがっているところ、他と差が大きいところ）を探す（丸をつけよう）
- パターン（法則）があるかどうかを探してみよう



【スライドの簡単な解説】

とんがっている、他と差が大きい、パターンなど、グラフを見るときポイントについていくつか説明します。

【このページのポイント】

とんがっている、差が大きい、パターンがあるなどの特徴があるところには何かが起きていることが多いです。

特徴があるところに何が起こったのか考えてみると良い気づきを得ることができます。

グラフについて（グラフの読み方）

【テキスト：33ページ】



気づいたことを書いてみよう

どうしてそれが起きたのが考えてみよう

73

【スライドの簡単な解説】

アイスクリームの月別支出金額と、東京の最高気温の折れ線グラフを見て気づいたことと、どうしてそれが起きているのかを考えるワークシートです。

【このページのポイント】

グラフを見て気づいたことと、それがなぜ起きているのかをセットで考えると分かってくることがあって良いです。



第四章

なりきり！データサイエンティスト



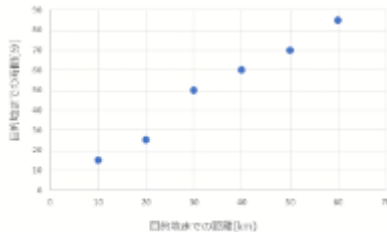
72

回帰分析とは

【テキスト：35ページ】

回帰分析とは、「数値データ」と「その結果となる数値」から、どのような関係があるかを明らかにする統計手法のことです。

例えば、皆さんが車でどこかに出かけるとします。その際「目的地までの距離(数値データ)」と「目的地までにかかる時間(結果となる数値)」を調べてみたところ、下のグラフのようになりました。



グラフから、「遠くに行こうとすればするほど時間がかかる」という関係性がわかります。

この関係性をもとにすると、目的地までの距離がわかれば、目的地までのおおよその時間が予測できます。

これも簡単な回帰分析といえます。

73

【スライドの簡単な解説】

「回帰分析」についての解説。

【このページのポイント】

回帰分析とは、「数値データ」と「その結果となる数値」から、どのような関係があるかを明らかにする統計手法のことです。

グラフから、「遠くに行こうとすればするほど時間がかかる」という関係性がわかり、目的地までの距離がわかれば、目的地までのおおよその時間が予測できることもわかります。

これも簡単な回帰分析といえます。

【キーワード】

回帰分析

回帰分析とは

【テキスト：36ページ】

元々の数値データのことを「説明変数」、その結果となる数値のことを「目的変数」と言います。

「目的変数」は、私たちが知りたい、つまり目的となる変数のことです。

「説明変数」は、どのような関係があるかを明らかにする、つまり関係性を説明するための変数のことを指します。

さきほどの例では、「目的地までの距離」が説明変数、「目的地までの時間」が目的変数となります。



74

【スライドの簡単な解説】

「説明変数」と「目的変数」についての解説。

【このページのポイント】

元々の数値データ（関係性を説明するための変数）のことを「説明変数」、結果となる数値（私たちが知りたい、目的となる変数）のことを「目的変数」と言います。

今回の例の場合、「目的地までの距離」が説明変数、「目的地までの時間」が目的変数となります。

【キーワード】

説明変数

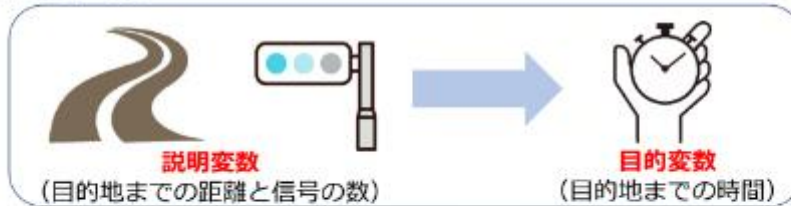
目的変数

回帰分析とは

【テキスト：36ページ】

ここで、2つ目の説明変数として「信号の数」を加えてみましょう。

信号の数が多いと、それだけ信号を待つ時間が増えるので、より時間がかかってしまいます。「目的地までの距離」と「信号の数」の2つを説明変数として使うことで、より正確に「目的地までの時間」を予測できそうです。



説明変数を1つ（「距離」だけ）で行う分析を「単回帰分析」

説明変数2つ以上（「距離」と「信号の数」）の複数で行う分析を「重回帰分析」といいます。

【スライドの簡単な解説】

単回帰分析と重回帰分析の解説。

【このページのポイント】

2つ目の説明変数として「信号の数」を加え、「目的地までの距離」と「信号の数」の2つを説明変数として使うことで、より正確に「目的地までの時間」を予測できる可能性が高くなります。

説明変数を1つ（「距離」だけ）で行う分析を「単回帰分析」、

説明変数2つ以上（「距離」・「信号の数」）で行う分析を「重回帰分析」といいます。

【キーワード】

単回帰分析

重回帰分析

演習その2：仮説立案をしてみよう！

【テキスト：37ページ】

仮説とは、【確からしい仮の結論】のことをいいます。確かかどうかを確認していくことを仮説検証といい、データを活用することは仮説検証の精度を高めてくれます。

データを活用した課題解決を目指すとき、意味のないデータを処理することにならないように「こうかな？」と思う「仮説」を組み立ててからその仮説を裏付けるデータを集めます。



76

【スライドの簡単な解説】

データを使った仮説検証についての解説。

仮説とは、「確からしい仮の結論」のことをいい、確かかどうかを確認していくことを仮説検証といいます。

【このページのポイント】

データを活用した課題解決を目指すときには、すぐにデータを集めたり分析したりするのではなく、必ず仮説を組み立ててから裏付けをしていきます。

【キーワード】

仮説検証

演習その2：仮説立案をしてみよう！

【テキスト：38ページ】

例題：総務省統計局のTwitterフォロワー数に影響を与えるのはどのような要素があると思いますか？仮説を立ててみましょう！



フォロワーが増える時はどのような時かな？
今日の統計チャレンジセミナーもフォロワーが増える要素になっているかも！



総務省統計局のTwitterアカウントがどのようなことを呟いていて、
どれくらい反応があるか調べれば仮説が立てやすそうだね！

77

【スライドの簡単な解説】

総務省統計局のTwitterフォロワー数に影響を与えるのはどのような要素があるか、仮説を立てます。

【このページのポイント】

総務省統計局のフォロワーが増えるのはどんな時か、反応の良い投稿とはどんなものか等を考えてみましょう。

演習その2：仮説立案をしてみよう！

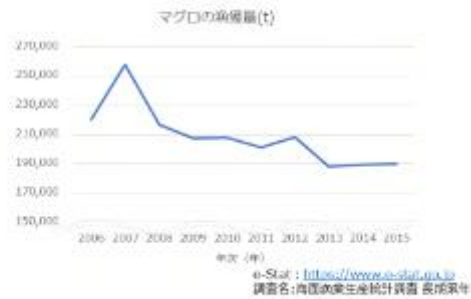
【テキスト：38、39ページ】

演習問題：マグロの漁獲量に影響を与えるのはどのような要素があるでしょうか？

仮説を立ててみましょう。

【参考】2006年から10年間の
マグロ漁獲量（全国）

年次（年）	マグロの漁獲量(t)
2006	220,331
2007	257,655
2008	216,885
2009	207,436
2010	206,059
2011	201,203
2012	206,172
2013	188,036
2014	189,705
2015	189,972



78

【スライドの簡単な解説】

マグロの漁獲量には何が影響しているか、仮説を立てます。

【このページのポイント】

2006年から10年間のマグロの漁獲量を表す表やグラフを参考にしながら、仮説を考えてみましょう。

演習その2：仮説立案をしてみよう！

【テキスト：40ページ】

仮説を見つけるヒント

■特徴的な箇所を見つける

グラフの突出しているところ/データが離れているところ/集計表の値が大きいところ・小さいところを見つける。

■社会情勢、時事問題などと照らし合わせてみる

災害、経済情勢などと関連性はあるか？

■線を引いて分類する

同じ傾向を示す箇所はないか？どのような特徴があるか？

■想定や仮説を書き足す

こうだったはずでは？本来あるべき状態とどのように異なるかを考え、整理する。

■逆説で考える

ここまで挙げたアイデアと反対の意見で考えてみる。



考えた仮説は
メモしておこう！

79

【スライドの簡単な解説】

仮説立案をするためのヒントについて解説。

【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ことや逆説で考えてみることも非常に重要です。

今日の感想

【テキスト：41ページ】

今日のセミナーの内容をグループでまとめてみよう！

- このセミナーのポイント
- これからの学習に活かそうなこと
- もっと学んでみたいこと



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。

このイベントは、後日総務上統計局のサイトにて公開予定です。

