なつやすみ
親子でデータサイエンス 2021

ワークガイド

はじめに

まずは統計を知ろう！

みなさんは**「統計」**をしってますか？
“計”と付いているくらいなので、何かを”計る“ことなのは間違いありません。
では、なにを“計る”のでしょう？
「統計」は、ある**“まとまったもの”**が、どんな**“特ちょう”**を持っているのかを、量や単位で**“計る”**ことです。といっても、少しわかりずらいですね。

たとえば、**自分のクラスには男子が多い**とすると、
自分のクラス＝まとまったもの
男子が多い＝特ちょう
となります。
そして、その特ちょうを量や単位で“計る”とすると
男子：２５人
女子：１８人
など、具体的な人数で比べてみることになります。

クラスの男子、女子の人数であれば、ひとりづつ数えていけば簡単にわかります。
では、**クラスで遅刻が多いのは男子、女子どちらでしょう？**と聞かれたら、
パッとすぐにわかるでしょうか？
すぐにわからないものは**“調べてみる”**ことが必要ですね。
一人ひとりに遅刻した回数を聞いて、それをメモして、
男子と女子で分けてまとめて、
遅刻した回数を比べてみて、はじめてどちらが遅刻が多いのかがわかります。

このように、知りたいことを一つひとつ集めて、それをまとめて、計算して、数字で計ることで、知りたかったことがわかる。これが「統計」なのです。

はじめに

「統計」はいろいろなものを調べることができます。

- 自分の家の前を通る車の数
- なつやすみの間に雨が降った日数
- 植物が成長するスピード
- 国の人口や、その増減について

他にもいっぱいありますが、この「統計」を使いながら、人が知りたいことを発見したり、人の役にたつことを考えたり実行できるようにするのが**「データサイエンス」**なのです。

次は、「統計」と「データサイエンス」はなにが違うのか？をお話ししますが、その前に、いくつかの言葉を説明しておきましょう。ここから先は、その言葉を使ってガイドを進めていくので、言葉の意味をしっかり覚えてください。

- 上で説明した**“まとまったもの”**のことを**集団や母集団**といいます。
- 集団の**“特ちょう”**は、ほかに**属性や傾向**という言葉でも表したりします。
- 上の説明で、クラスという集団の中の生徒**一人ひとり**が**要素**と呼ばれます。
- その一人ひとりの要素が男子か女子か、何回遅刻したか、といった量や単位を**データ**と言います。

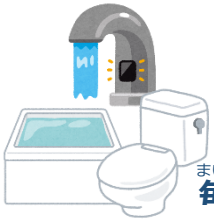
改めて、これらの言葉を使って「統計」を説明すると、

「統計」は、ある集団が、どんな属性・傾向を持っているのかを要素が持つデータを集めて調べること

となります。

ある道の交通量

毎日の降水量



毎月使う水の量

植物の成長記録

はじめに

データサイエンスってなんだろう？

データサイエンスは、統計を使って**“ひとの役に立つことを見つけたり、役に立つ使い方を考えたりする”**ことです。
役に立つことを見つけたり、役に立つ使い方を考えるときに重要なことは、統計を使って調べる対象や事象について**“調べる前に答えを予想してみて”**、それから統計を使って予想が結果と**“あっていたのか、ちがっていたのかを確認する”**ことです。
この“結果を予想すること“を**「仮説」**といいます。また、”あっていたのか、ちがっていたのかを確認する”ことを**「検証」**といいます。
「仮説」をたてて「検証」することで、調べたことが役に立つのか、どのように使えるのかがわかったり、こんなことが起こるのではないか、という予想ができるようになります。
世の中や会社が新しいことをするとき、または今までやってきたことの問題を解決しようとするとき、このわかったり、予想できたりすることが、とても役に立ちます。
データサイエンスは、世の中や会社の役にたつような「信じられる」結果を見つけることができる技術なのです。
ですから、データサイエンスを使えたり、データサイエンスのような考え方ができる人が求められているのです。

SASの「なつやすみ親子でデータサイエンス」は、身近なものや興味のあるものを使って親子でデータサイエンスを体験してもらうためのイベントです。
ご家庭で取り組んでいただくために、どのようにデータサイエンスを進めていけばよいかをご紹介します。本ガイドを参考に、ぜひ楽しみながらデータサイエンスにチャレンジしてみてください！

データサイエンスのすすめかた

データサイエンスは、はじめに**知りたいことや困ったことを見つける**ことが大切です。
そして、もう一つ大切なことは、“きっとこうだろう”“たぶんこうじゃないかな”など**結果を予想してみる**ことです。
この予想した結果が、あっているのか、ちがっているのか、を確かめるためにデータを集めて、結果を探していきます。今回、みなさんには、このデータサイエンスの流れにそってワークをおこない、最後はポスターにまとめて発表してもらいます。

ワークの手順

① **テーマを決める**
知りたいこと、困っていることを見つけてみよう！

② **結果を予想してみる**
自分で結果を予想してみよう！

③ **調べかたを決める**
どうすれば予想した結果があっているのか、ちがっているのか、を確かめられるのかを考えてみよう！

④ **データを集める**
決めた調べかたで、欲しいデータを聞いたり、記録したり、探したりして集めよう！

⑤ **データを分析する**
集めたデータをグラフにして知りたいことを見つけよう！

⑥ **結果を確かめる**
自分が予想した結果はあっていた？ちがっていた？確かめてみよう！

⑦ **ワークをまとめる**
テーマを決める～結果を確かめるまでの流れをまとめよう！

⑧ **ポスターを考える**
まとめたもので、どんなポスターをつくるか考えてみよう！

Point ワークシートをつくりながら進めると良いよ！

ワークシートは、
作業を整理してまとめるのに役立ちます。
ノートや白紙、何でも良いので、作業をひとつづつ
まとめながら進めていきましょう！

ワーク❶　テーマを決める

知りたいこと、困っていることを見つけてみよう！

まずは、テーマを決めることからはじめましょう。
テーマは、みなさんの身近にあるもの、興味のあること、好きなもの、などから
“知りたいこと”“困っていること”を探してみてください。

テーマ例）

●きっかけとなる疑問	●テーマ
・自分の学校の通学路は安全かな？	・**通学路の安全**
・自分のおこづかいは多い？少ない？	・**正しいおこづかい**
・大人は子どもの時なりたかった仕事をしているの？	・**こどもの夢のかなう確率**
・仮面ライダーで一番強いのは何ライダー？	・**仮面ライダーランキング**
・天気予報って、どのくらい当たるの？	・**天気予報の確率**

Point　**いくつか案を出してから選んでみよう！**

データサイエンスの基本となる統計は、使えればとても便利なものですが、
向き不向きもあります。まずは、いくつかテーマになりそうなものを並べてみて、
統計で答えが出せそうなものを選びましょう。

例）

テーマ案

親が喜ぶ
プレゼントを見つけたい
→ 親や家族に聞けばわかる
自分の経験からわかる　✕

世の中の親って子供から
何をもらうと嬉しいの？
→ 友達の父兄や
子供のいる大人から
たくさんアンケートを取る　○

統計はデータをいっぱい集めないとわからないことを調べるのが得意！

ワークシートのまとめかた

テーマ案
・あああああああああああ
・あああああああああああ
・あああああああああああ
・あああああああああああ
・あああああああああああ

なぜそのテーマにしたの？

選んだテーマ

ワーク② 結果を予想してみる

自分で結果を予想してみよう！

決めたテーマに対して、自分で“こうじゃないかな？”“きっとこうだろう！”という結果を予想してみましょう。そのときに“なぜそう思うのだろう？”という理由もいっしょに考えることがたいせつです。

予想してみた結果例）

●テーマ	●予想してみた結果
・通学路の安全	車が多いし、安全じゃないと思う
・正しいおこづかい	小学生は４年生から多くなる
・こどもの夢のかなう確率	なんとなく50%くらい
・仮面ライダーランキング	大好きな〇〇ライダーが一番！
・天気予報の確率	晴れ、くもり、雨で、だいたい30%くらい

Point 結果の予想は、調べかた（次のワーク）を考えるためにも必要！

結果を予想することは、“何を見つければよいのか”をハッキリさせてくれます。データサイエンスでは、とても大切なものです。“なんとなくそう思う”“前に聞いた気がする”といったフワッとした理由でもかまいません。
まずは予想した結果を置いてみましょう！

ワークシートのまとめかた

テーマ

予想してみた結果

なぜそう思うのか？

ワーク③　調べかたを決める

どうすれば予想した結果があってるか、ちがっているか、を確かめられるのかを考えてみよう！

どんな調べかたをするかは、データサイエンスにおいてとても大事なことです。
この調べかたによっては、よいデータを集められなかったり、最後に見つかる結果が欲しいものではなかった、などワーク全体の結果を大きく左右します。

Point　**できるだけ多く、できるだけ広く調べよう！**

データサイエンスの基本となる統計では、集めるデータの数が多ければ多いほど、また調べる範囲が広ければ広いほど信じられる結果を見つけられるようになります。

例）

テーマ：コインを投げたときの表と裏の出やすさ
予想してみた結果：表が出やすいと思う。

調べかた：コインを投げて表が出たか裏が出たかを記録する。
10回投げるより100回投げるほうが結果を信じられる。

テーマ：こどもの夢のかなう確率
予想してみた結果：なんとなく50%くらい

調べかた：働いている大人の人にアンケートを取る。
親や先生だけでなくいろんな仕事の人に聞くほうが結果を信じられる

計る単位があいまいな時は基準を決めてから調べよう

たとえば「自分の学校の通学路は車の通る量が多い」と予想をたてたとき、１時間に５台の車が通ったら、それは多いのか、少ないのか？
よく考えるとあいまいなものは、「10台通れば多い」など基準を作ってから調べかたを決めていきましょう。

ワークシートのまとめかた

テーマ　　**予想してみた結果**

調べかた

（基準）あいまいな時は決める

ワーク④　データを集める

決めた調べかたで欲しいデータを聞いたり、記録したり、探したりして集めよう！

データとは「人に聞いたり記録をつけたり本やネットで探したりして集めたもの」のことです。自分が決めた調べかたが“○○な人に聞く”であれば、聞いた人一人ひとりの回答がデータになります。自分が欲しいデータを、必要なぶんだけ集めましょう！

Point　**データを整理する「表」も合わせて作りましょう！**

データがたくさん集まると、それを整理するものが必要になります。それが表です。データを集める前に表を作っておくと、自分がどんなデータを集めるつもりなのか、がよくわかるので、集めるデータが「ちょっとずれてるかも」「これでは欲しい結果が出せないかも」といったことにも気づいて直すこともできます。

例）

「コインを100回投げた結果」をまとめる表

投げた回数	結果
1回目	表
2回目	裏
3回目	表
4回目	表
⋮	⋮
100回目	裏

「大人は子どもの時なりたかった仕事をしているのか」を聞くアンケートをまとめる表

いまのお仕事	子どもの時なりたかったお仕事	いつなりたいと思ったのか	かなったかなわない	いまの仕事がすき？
先生	先生	中学生	かなった	すき
会社員	パイロット	小学生	かなわない	きらい
社長	野球選手	小学生	かなわない	すき
消防士	消防士	高校生	かなった	すき
公務員	公務員	高校生	かなった	すき
⋮	⋮	⋮	⋮	⋮

聞くことを増やせばわかることも増えるよ！

ワークシートのまとめかた

調べかた

データ集めでやったこと

データを集めた期間

データを集めて感じたこと、苦労したこと

世の中にもみんなが使えるオープンデータがある

国や企業がいろんなテーマで調べたデータを誰もが見て使えるように公開しているオープンデータというものがあります。自分が使いたいものがあれば利用してみましょう。

ワーク⑤　データを分析する

集めたデータをグラフにして知りたいことを見つけよう！

ここで分析という言葉が出てきましたが、みなさんは**集めたデータから知りたいことを見つける作業のこと**と覚えておいてください。
前のワークで集めたデータを表にまとめたら、グラフを作りましょう。グラフにすることで、集めたデータがあらわすものを目で見て確認することができます。
また、そこから**「平均値（へいきんち）」「中央値（ちゅうおうち）」「最頻値（さいひんち）」**など、集めたデータの特ちょうを見つけ出すことができます。

Point　**グラフの種類と使い方を覚えよう！**

グラフは、集めたデータがあらわすものをわかりやすくしてくれます。
大きさや多さを比べたり増えたり減ったりといった変化を見るのにとても便利です。
グラフには、種類と使いかたがあるので、自分が集めたデータや、調べたいことに合ったグラフを選んで作りましょう。

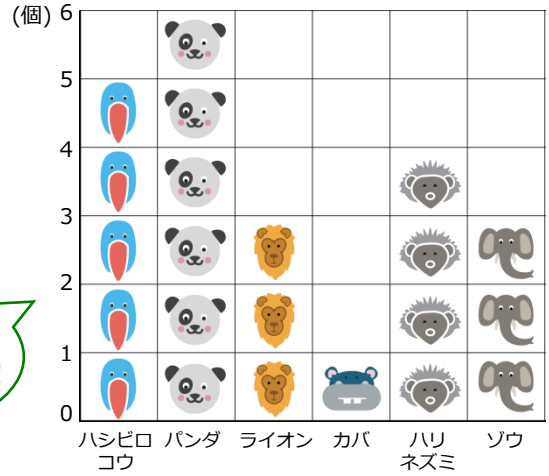
2つ以上の要素の大小を比べるグラフ

●絵グラフ

つかいかた

集めたデータを絵や記号で表します。
一目でイメージが伝わるグラフです。
個数で数えられる数量を表すのに向いています。

例）動物園の好きな動物

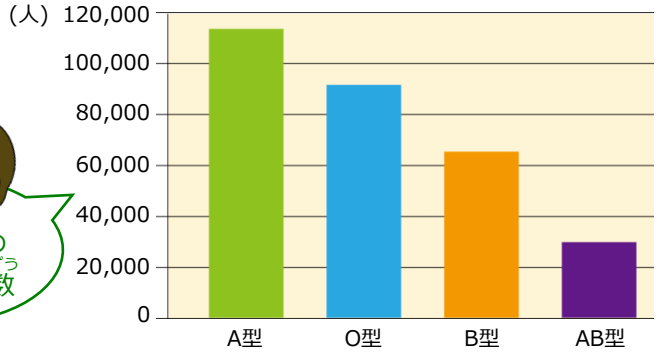


●棒グラフ

つかいかた

2つ以上の要素の数量をならべて大きさを比べるときに使います。
大きさ順にならべれば順位も表せます。

例）日本人（30万1959人）の血液型別の人数



ワーク⑤　データを分析する

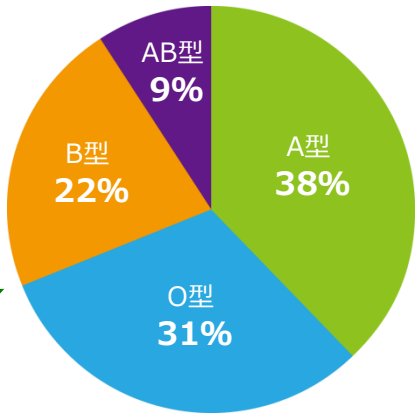
わりあいを見るグラフ

●円グラフ

つかいかた

集団の中で、どの要素がどのくらい多いのか、少ないのかを表すときに使います。
※統計では、これを割合（わりあい）といいます。

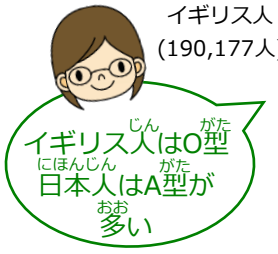
例）日本人（30万1959人）の血液型の割合



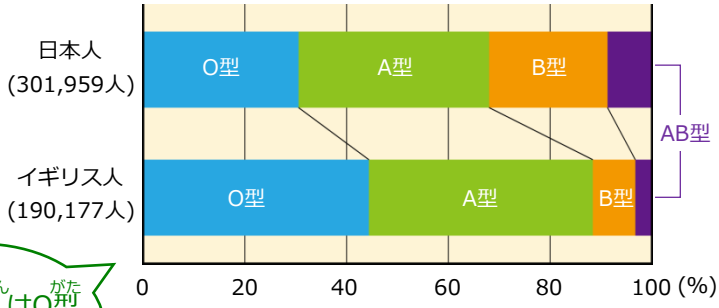
●帯グラフ

つかいかた

円グラフとおなじように集団全体を100%としてわりあいを表します。
複数の要素を比べるときは帯グラフのほうが向いています。



例）日本人とイギリス人の血液型の割合



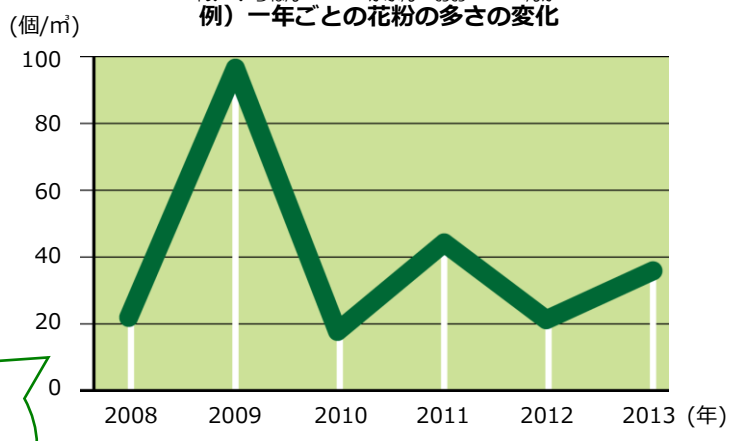
へんかを見るグラフ

●折れ線グラフ

つかいかた

時間が進むあいだに起こる変化を表すときに使います。
過去の記録を表すだけでなく将来どうなるかを予測したいときにも使われます。

例）一年ごとの花粉の多さの変化



ワーク⑤　データを分析する

Point　集団の特ちょうの見つけかたを覚えよう！

集計したり、グラフにまとめたりすると、その集団の特ちょうを表す数値を見つけることができます。統計では「平均値（へいきんち）」「中央値（ちゅうおうち）」「最頻値（さいひんち）」の３つが、よく使われます。

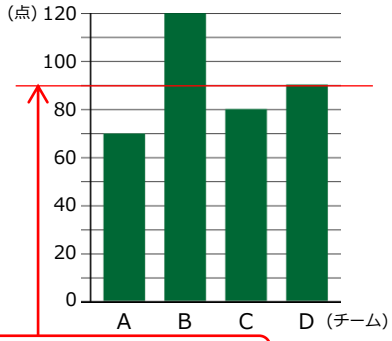
集団の特ちょうを表す数値

●平均値（へいきんち）

集めたデータ全体の数値を平らにならした値が平均値です。
平均値は、集計したデータ全体の基準として使えます。

平均値の計算方法
データの値の合計 ÷ データの個数

チーム	得点
A	70
B	120
C	80
D	90
合計	360



360÷4＝90　平均得点は90点

●中央値（ちゅうおうち）

集めたデータの数値を大きさ順にならべたとき、ちょうど真ん中にくる値を中央値といいます。
もしデータの数が偶数のときは真ん中の２つの値を足して２で割った数値が中央値になります。
中央値は、平均値と同じく集めたデータ全体の基準として使えます。平均値との違いは、異常な値の影響を受けにくい、ということです。

●データ数が奇数の場合

1	2	2	5	6	7	7

中央値は5

●データ数が偶数の場合

1	2	2	5	6	7	7	8

中央値は5.5

●異常値があった場合の平均値との違い

1	2	2	5	6	7	50

中央値は5　**異常値**
平均値は約30

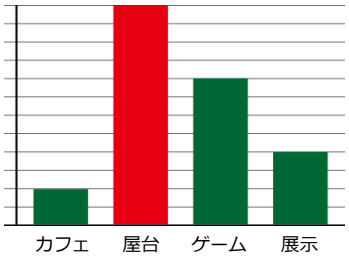
中央値は異常値の影響を受けにくい

●最頻値（さいひんち）

集めたデータの数値の中で、いちばん多く表れた数値を最頻値といいます。
最頻値はデータ全体の“多数派”を表します。
絵グラフや棒グラフにすると、一番高い山になるところが最頻値になります。

【文化祭の出し物希望】

出し物	票数
カフェ	4
屋台	24
ゲーム	18
展示	8



最頻値は屋台

ワーク⑤　データを分析する

集団の特ちょうを表す数値の違いとそれぞれの意味

例）あるクラスのテストの結果

順位	点数
1	100
2	95
3	95
4	92
5	90
6	85
7	83
8	80
9	77
10	75
11	73
12	67
13	62
14	60
15	58
16	58
17	58
18	58
19	57
20	52
21	50
22	50
23	45
24	45
25	43
26	42
27	40

平均値
66.3点

中央値
60点

最頻値
58点

平均値

このクラスのテストの平均点は66.3点です。

取った点数で上位、下位を分けられる。

中央値

このクラスのテストをランキングすると、ちょうど真ん中になるのは60点です。

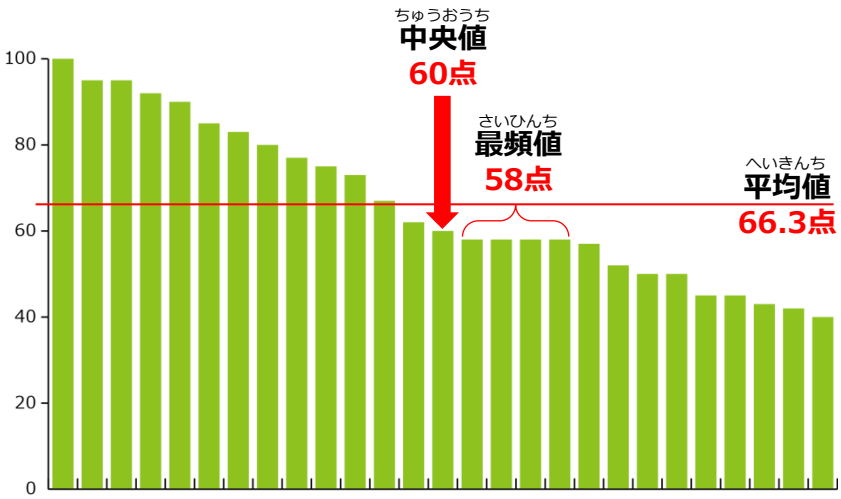
人の順位で上位、下位を分けられる。

最頻値

このクラスのテストでいちばん多くの人が取った点数は58点です。

このテストで取りやすい点数（むずかしい、やさしい）を判断できる。

グラフにする



グラフや特ちょうを表す数値あれこれ

ここで紹介したもの以外にもグラフの種類や特ちょうを表す数値の種類はいろいろあります。
興味があれば、書籍やネットなどで調べて、使ってみてください。

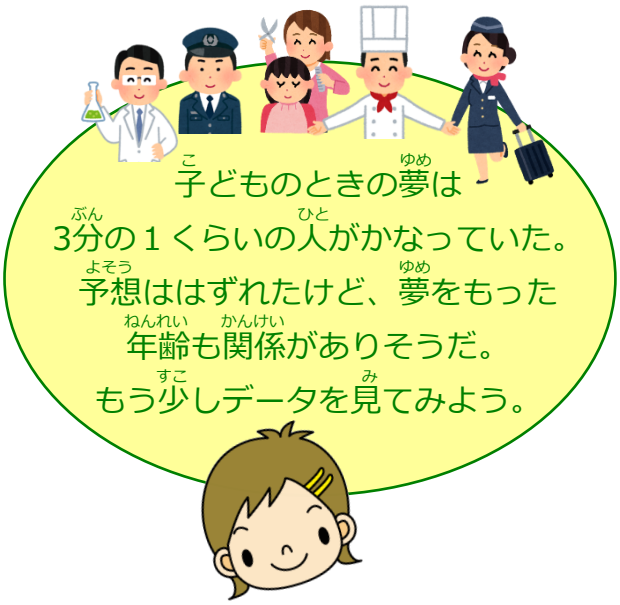
ワーク⑥　結果を確かめる

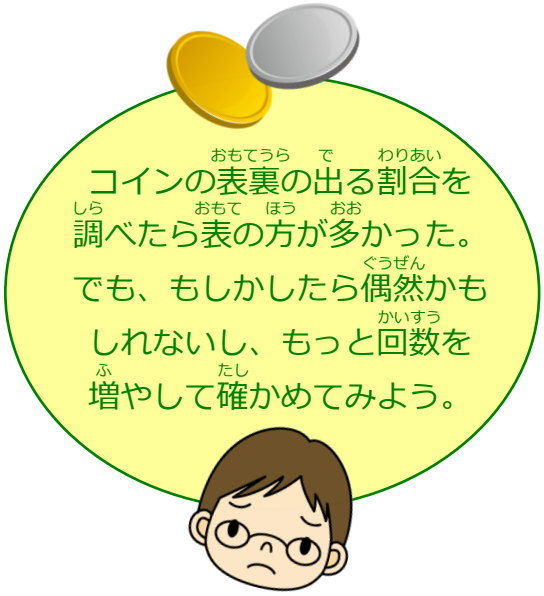
予想した結果はあっていた？ちがっていた？確かめて結果を出そう！

予想してみた結果と、分析して分かったことを比べながら、自分が考えたとおりだったのか、ちがっていたのか、を確認しましょう。

Point　**あっていても、ちがっていても、調べてわかることが大切！**

データサイエンスは、数多くのデータから、普段はわからないことがわかるところに意味があります。予想した結果があっていたか、ちがっていたかを確認したら、そこで終わらずに、ワークを通して新しくわかったこと、新しく発見できたことにも注目してみましょう！
また、一度の結果に満足せず、時間のゆるすかぎり、さらに結果を正確にするために調べなおしてみたり、別の調べかたをためしてみることもおススメします。





ワークシートのまとめかた

テーマ

予想してみた結果　　**統計でわかった結果**

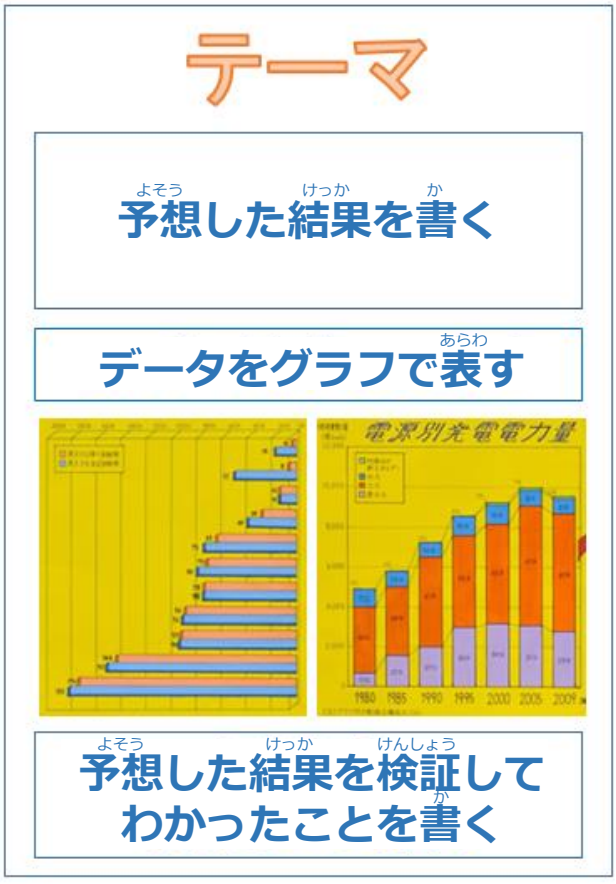
ワークを通して感じたこと、発見したこと

ワーク⑦　ポスターを考える

「まとめシート」をもとに
どんなポスターをつくるか考えてみよう！

みなさんの自宅でのワークは以上になりますが、もし時間があれば、イベント当日につくるポスターをどんなものにするか、イメージして考えてみてください。イベント当日は、ポスターを作る前にSASのサポートスタッフからアドバイスをもらいつつ、自宅でのワークの流れと分析方法を確認してからポスターを作っていただきます。

ポスターの例）



統計局のサイトに、全国のこどもたちが作ったポスターが紹介されていますので、いろいろな作品を参考にするのも良いでしょう。

「なるほど統計学園　統計グラフコンクール」

http://www.stat.go.jp/naruhodo/c1conc.htm

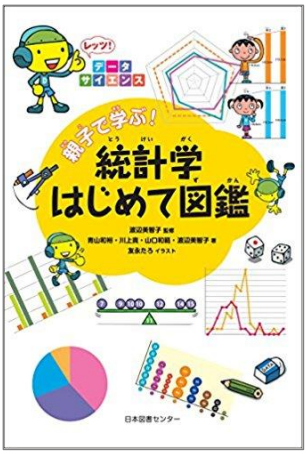
どんな形にレイアウトするかはアイデア次第！
自分だけのポスターをいろいろ考えてみてね！

ワークの参考になるもの

本やネット上の情報など、ワークの助けになりそうなものをご紹介します。

統計についての本

親子で学ぶ！統計学はじめて図鑑



（日本図書センター）

まなびのずかん　統計学の図鑑

（技術評論社）

統計についてのネット情報

- **小学生のための統計学習　まなぼう統計**
  http://www.toukei.metro.tokyo.jp/manabou/syou/ma1204s001.htm
- **なるほど統計学園**
  http://www.stat.go.jp/naruhodo/index.htm
- **統計学習の指導のために（先生向け）** ※補助教材やリンク集が参考になります。
  http://www.stat.go.jp/teacher/index.htm

オープンデータの一部

- **DATA GO JP** ※日本政府のオープンデータリンク集
  https://www.data.go.jp/
- **自治体オープンデータ** ※自治体の公共データカタログ
  https://www.open-governmentdata.org/

さいごに　ワークをしていて困ったら

「なつやすみ親子でデータサイエンス」運営事務局では、参加者の方からのワークに関するご質問、お問い合わせ窓口を開設しています。
イベントの目的として、できる限り親子で力を合わせてデータサイエンスにチャレンジしていただきたいと思いますが、どうしてもわからないことや、確認したいことがある場合は、窓口までお問い合わせください。

「なつやすみ親子でデータサイエンス」参加者用窓口

E-mail : SAS@eventdesk.jp

Let’s Do Data Science !