

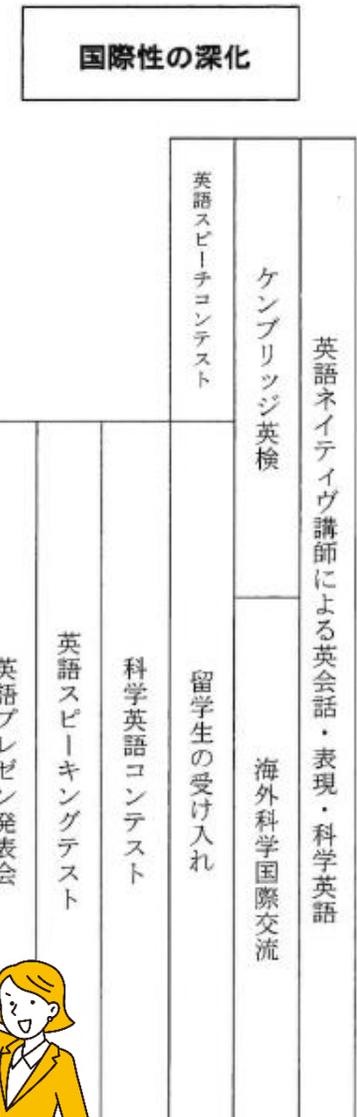
職員研修(情報科)

小中高一貫IKEDAデータサイエンスについて

担当：黒木

IKEDA データサイエンス

小学校を「探究力をつかむ」段階、中学校を「探究力を深める」段階、高等学校を「探究力を広げる」段階と位置付ける。各段階におけるSSH学校設定科目と研究活動(全員参加の自由研究および課題研究)を通して、児童生徒の探究力を育成する。**小中高一貫でデータサイエンスをカリキュラムとして位置付け**、あわせて国際化教育を展開することにより、高校でのSSH活動を開花させ、学園全体として国際性のある科学技術人材を育成する。
(年次報告書p.51より)



Data Science について、いろいろな意見があります

Data Scienceは、統計のことだよね。

プログラミングのことじゃないの？

理系のための情報処理技術だと思ってた。

株価を予測する情報処理技術ではないの？

専門家でないと使えない最新技術でしょ。

すぐに簡単に何でもできる魔法の技術だよね。



- Data Science はデータに関する総合的な情報処理の技術です。
- それは自然科学だけでなく、人文科学・社会科学でも活用されています。
- スマートフォンやネットショッピングも利用される身近な存在です。
- 「made in JAPAN」は Data Science がもたらした世界的な信用です。
- 実は、みなさんも小学生の頃から Data Science を学んでいます。



Data Science を目的別に分類

目的は5つ！

Point データを〇〇するために、どの方法を使う？

1. 一次データ収集 (Primary Data Collection)

アンケート調査、
インタビュー調査(個別面接、グループインタビュー)
観察、科学的な実験、センシング

2. 二次データ収集 (Secondary Data Collection)

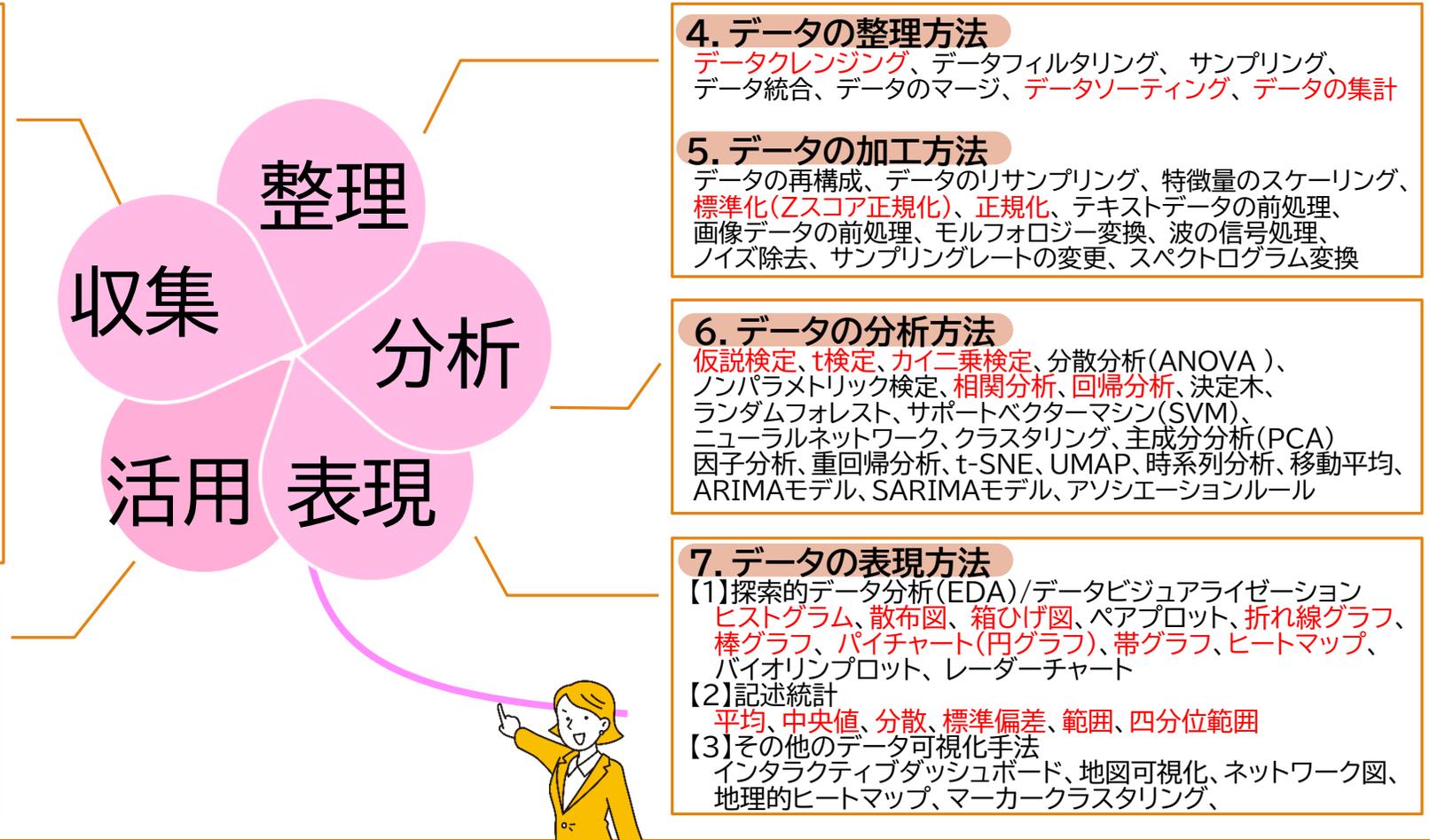
公的データベース、
商業データベース、
学術論文

3. その他

ウェブスクレイピング(ウェブサイトからデータを自動収集)、
API(Twitter API, Google Maps API)、
IoTデバイス(センサーやスマートデバイス)、
ウェブサーバーやアプリケーションサーバーログファイル、
ユーザーアクティビティログ、リアルタイムデータストリーム、
クラウドソーシングプラットフォーム、
ソーシャルメディアデータ、オープンデータポータル、
ECサイトやマーケットプレイスからの取引データ、
CRMシステム、ERPシステム

8. データの活用方法

予測、データマイニング、シーケンスマイニング、
自然言語処理、テキスト分類、感情分析、
トピックモデル、異常検知、最適化、
ヒューリスティック(遺伝的アルゴリズムなど)、レポート生成、
ビジネスインテリジェンス(BI)、KPIモニタリング、
データシミュレーション、エージェントベースシミュレーション





Point

データを〇〇するために、どの方法を使う？



7. データの表現方法

- 【1】探索的データ分析(EDA)/データビジュアライゼーション
ヒストグラム、散布図、箱ひげ図、ペアプロット、折れ線グラフ、棒グラフ、パイチャート(円グラフ)、帯グラフ、ヒートマップ、バイオリンプロット、レーダーチャート
- 【2】記述統計
平均、中央値、分散、標準偏差、範囲と四分位範囲
- 【3】その他のデータ可視化手法
インタラクティブダッシュボード、地図可視化、ネットワーク図、地理的ヒートマップ、マーカークラスタリング、

箱ひげ図を使って、データの分布を表現する



Point

データを〇〇するために、どの方法を使う？

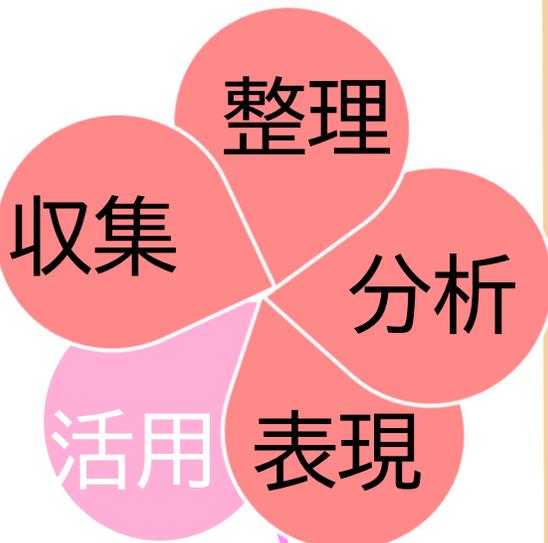


6. データの分析方法

仮説検定、t検定、カイ二乗検定、分散分析(ANOVA)、ノンパラメトリック検定、相関分析、回帰分析、決定木、ランダムフォレスト、サポートベクターマシン(SVM)、ニューラルネットワーク、クラスタリング、主成分分析(PCA) 因子分析、重回帰分析、t-SNE、UMAP、時系列分析、移動平均、ARIMAモデル、SARIMAモデル、アソシエーションルール

t検定を使って、データの平均の差があるかを分析する

ハクセンシオマネキの 雌にもてる雄は4.2秒間のダンスを休まずに続ける



- ・モルフォロジー変換で、不要なエリアの削除して台形補正をした
- ・PCにカニを認識させ、追跡と重心検出、行動判断を自動でさせた
- ・関心円の増減を画像解析して、求愛ダンスの個体差を確認した
- ・誤差グラフから、求愛では雌と一定の距離を保つことを確認した
- ・バイオリンプロットから、小さなハサミ側のメスに求愛中と確認した
- ・時系列で関心円の面積を移動平均して、求愛ダンスを波に変換した
- ・短時間フーリエ変換による信号処理を使い、波の周波数を検出した
- ・周波数とそのパワーを使って、スペクトログラムを作成した
- ・モテるオス達のスペクトログラムから、4.2秒間を確認した



県立宮崎北高校科学部がMATLABで
カニの求愛行動を解析！
(MATLAB EXPO 2020最優秀賞受賞)



無限に広がるMATLABの可能性を
目撃する、ライトニングトーク 第2弾
(MATLAB EXPO 2021最優秀賞受賞)





池田学園独自のData Scienceカリキュラムをつくる

- 小学校5年生から高校1年生までの生徒が情報の授業でData Scienceを学習する

※ 今年度は全学年が同じ内容
「データの表現方法」を学んでいる

年次進行で段階的に
つくっていく

各学年で使った
教材をストック

- 小中高の教科「情報」の担当者による定例会

※ カリキュラムや教材や
指導方法について情報共有

数学や理科の授業との
合科も視野に入れる

センシングや画像解析、
信号処理も検討する

情報リテラシー&IKEDA データサイエンス進捗計画

※各学年の枠内の数値は和暦を示す

	扱うデータ	理科の素材	数学の進度	情報	小5	小6	中1	中2	中3	高1
データを集める・データを表現する	A: 表計算ソフトを使えるようになるよう! PCの使い方(マウスなど) Excelの基本(選択・コピー&ペースト・参照)				6 7 8 ↑↑↑					
	B: グラフを作ろう! 4グラフ(棒・折れ線・円・帯)の特徴 最適なグラフを選んで、Excelでグラフ作成	← 実数値				6 7 8 ↑↑↑				
	C: 表計算ソフトで計算しよう! 四則演算 割合を求める	← 相対値					6 7 8 ↑↑↑	6 7 8 ↑↑↑	6 7 8 ↑↑↑	6 7 8 ↑↑↑
	D: レアデータから集計しよう! カウント count(), countif() 平均値 average() 丸目処理 round(), roundup(), rounddown()	平均値			四則演算子			6 7 8 ↑↑↑	6 7 8 ↑↑↑	6 7 8 ↑↑↑
データを分析する	E: 1変数のデータの分布を可視化しよう! 母集団と標本の違い 度数分布表とヒストグラムの作成 関数 median(), mode()	階級値 代表値						6 7 8 ↑↑↑	6 7 8 ↑↑↑	6 7 8 ↑↑↑
	F: 1変数のデータの分布を数字で表現しよう! 代表値とばらつき 分散・標本標準偏差・不偏標準偏差の計算	分散 標準偏差						6 7 8 ↑↑↑	6 7 8 ↑↑↑	6 7 8 ↑↑↑
	G: 外すべきデータを見つけよう! 箱ひげ図と外れ値 欠損値や異常値との違い	四分位数 外れ値 異常値			条件分岐			6 7 8 ↑↑↑	6 7 8 ↑↑↑	6 7 8 ↑↑↑
データを分析する・データを活用する	H: 2変数のデータの分布を表現しよう! ヒストグラムと散布図の関連性 散布図と回帰直線の作成 相関係数と重相関係数 相関関係と因果関係	相関係数 重相関係数						6 7 8 ↑↑↑	6 7 8 ↑↑↑	6 7 8 ↑↑↑
	I: 2つの平均値に差があるか確認しよう! 帰無仮説と対立仮説 平均の差の検定 一群のt検定・対応のないt検定・対応のあるt検定	t値 p値						6 7 8 ↑↑↑	6 7 8 ↑↑↑	6 7 8 ↑↑↑
	J: 2つの比率に差があるか確認しよう! 比率の差の検定 χ^2 検定・z検定	χ^2 値 z値 p値						6 7 8 ↑↑↑	6 7 8 ↑↑↑	6 7 8 ↑↑↑
	H: 確率分布から推定しよう! 標準正規分布からの推定 回帰式からの推定							6 7 8 ↑↑↑	6 7 8 ↑↑↑	6 7 8 ↑↑↑

小5~中2

中3・高1

☆目標☆
小6年生以降は県統計コンクールに出展(9月上旬)

特に初年度は、生徒の様子を観察しながら、次年度以降の授業進度の調整を図らなければなりません。小学校・中学校・高等学校の担当者で定期的に会議を開けたらと思います。
また、新たな取り組みに対する不安感を解消し、相互理解を深め、柔軟な発想で楽しい授業を構築していくために、生徒だけでなく私たちもノウハウの共有するOJTやアイデアを互いに出し合い、スキルアップしましょう。

※ 環境が整い、授業時間に余裕があれば、探求型の実習、マイコンを使ったセンシング、画像解析や信号処理を組み込む。

