

39 データ分析の流れ

40 目的に合わせてデータの利用

この時間の目標

- データ分析の流れ/目的に合わせたデータの利用
- S データの分析や利用についてよく理解でき、配慮して活用しようと思った
- A データの分析や利用についてよく理解できた
- B データの分析や利用について理解できた
- C データの分析や利用について理解できなかった

39 データ分析の流れ

1 問題の明確化と計画

- 問題の明確化(problem)
 - 目的や解決すべきことを具体的に
 - 正確に把握
- 計画(plan)
 - 収集分析するデータを絞り込む
 - 新たなデータを収集/既存のデータを利用

2データの収集

- データの収集(data)
 - 必要なデータを収集する
 - 表などに整理する

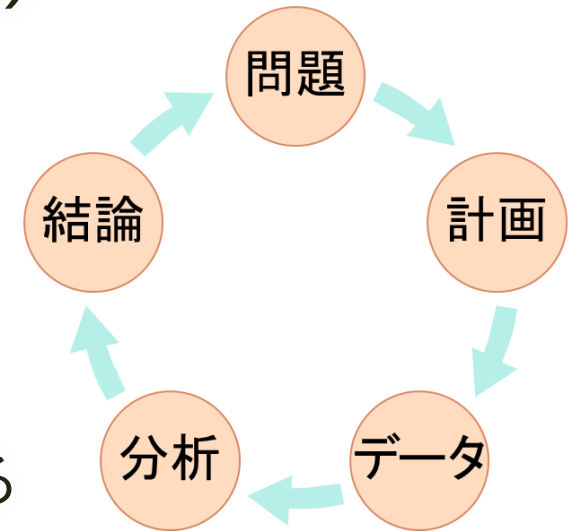
3データの分析

- 分析 (analysis)
 - グラフで可視化
 - 代表値 (平均値・中央値・最頻値)
 - クロス集計
 - 仮説検定・相関係数・標準偏差
- 結論 (conclusion)
 - 目的とした問題の解決
 - 新たな問題の発見

• 統計的探究プロセス (数学 I 201ページ)

• PPDACサイクル

- P(problem、問題)
 - 解決すべき事柄を把握し、統計で扱える問題を設定する
- P(plan、計画)
 - 設定した問題に対して、集めるべきデータとその集め方を考える
- D(data、データ収集)
 - 計画に従ってデータを集め、表などに整理する
- A(analysis、分析)
 - 目的やデータの種類に応じてグラフにまとめたり、データに関する数値を求めたりして、特徴や傾向を把握する
- C(conclusion、結論)
 - 見出した特徴や傾向から結論をまとめて表現したり、さらなる課題や改善点を見出したりする



40 目的に合わせた データの利用

1 欠損値と外れ値

- 欠損値: 数値が欠けている
- 外れ値: 他の値から大きく外れている

- 取り除くべきか・近似値などで埋めるか検討が必要

2分析の目的とデータの関係

- 目的に合わないデータ
 - 間違った結果を導き出す
- 分析したい内容と関係が強いデータを集める

3データの解釈

- 因果関係
 - かき氷の売り上げが増えると熱中症が増える
- 擬似相関
 - 因果関係がないのに因果関係があるように見える

分析に必要なデータ

バーコードデータ

- バーコードデータ
 - データクレンジングしました
- データクレンジングの方法
 - 桁数の違うバーコードデータを削除(8桁・13桁のみ)
 - 2で始まるバーコードデータを削除(ストアコード)
 - 本のバーコードデータを削除
 - チェックデジット異常のデータを削除
 - 価格異常のデータを削除
 - 品名は長い文字数を採用/価格は平均値

バーコードデータ

- バーコードv2
 - テーブルを分割
 - 購入データ: 購入者・バーコード番号
 - 商品データ: バーコード番号・商品名・価格

バーコードデータの分析

- 分析のために必要なデータを考える (PP)
 - ○○のデータがあれば□□について分析できる
 - ペアで相談(5分)・クラスで集約(10分)
 - 誰か仕切って
- クラスで手分けしてデータを追加? (D)
 - 2項目まで追加可能
- 分析までたどり着けるか? (AC)