

データの分析

第6回

23j1-306

この時間の目標

- データの分析
- S データ分析の手順や方法が活用でき、積極的に取り組んだ
- A データ分析の手順や方法が活用できた
- B データ分析の手順や方法が利用できた
- C データ分析の手順や方法が利用できなかった

テキストマイニング体験

情報 I 1・2学期の感想を分析

テキストマイニング

- 文章から傾向を読み取る
 - 日本語は難しい
 - まだ完全に解決していない
- テキストマイニングツール「ユーザーローカル」
 - テキストを貼り付けて解析
 - 情報 I の「質問・感想より」でおなじみ
- 文書を解析
 - 情報 I 1・2学期の感想

実習:テキストマイニング1

- 文書を用意する
 - 自分のドライブのファイルを開く
 - 自己評価の感想:年組席番感想.xlsx
 - ユーザーローカルを開く
- 文章を貼り付ける
- [ワードクラウドを作成]をクリック

実習: テキストマイニング2

- 中央下[画像ダウンロード]をクリック
- [PNG画像ダウンロード]をクリック
- 自分のドライブに保存
 - ファイル名は 23j1-306 とする



- 情報 I にどのように取り組んでいたか文章にまとめ授業の振り返り[設問11]に回答

データの分析1

全3回

授業計画

進め方はスライドを参考に
進行管理を確実に

- 条件
 - SSDSE(教育用標準データセット)を利用
 - PPDACの手法で
- 全3回
 - 時間もないのでスゴい発見はできない
 - Pまずデータの概要をつかむ/視点を決めて仮説を立てる
 - P必要なデータが何か明らかに 1回目
 - D必要なデータを集めまとめる
 - A分析(統計的手法を使う) 2回目
 - C結論(スライド作成) 3回目

分析するデータ

- SSDSE(教育用標準データセット)
 - データ分析のための汎用素材
 - 独立行政法人統計センターが作成・公開
 - 主要な公的統計を地域別に一覧
 - 表形式のデータセット
 - 解説書付き
- 関連する他の調査データの収集・利用も可

1. P(problem)問題

- まずデータの概要をつかむ/視点を決めて仮説を立てる
- SSDSEにどのようなデータがあるか
 - ○○と□□の関係を知りたい→データ分析の目的
- 仮説を立てる
 - ○○と□□には△△の関係があるのではないか
 - この仮説の真偽を分析して解き明かす
 - [明らかにできない]という結論も
 - [明らかにできないということ]を明らかにした

2. P(plan)計画

- 必要なデータが何か明らかに
- 利用するデータを具体的に決める
- どのような統計的手法を使うかを決める

3. D(data)データ収集

- 必要なデータを集めまとめる
- 不必要な列や行の削除
- 比較対象などのデータ入手
- 処理しやすい形の表にまとめる

- 関連する他の調査データの収集・利用も可

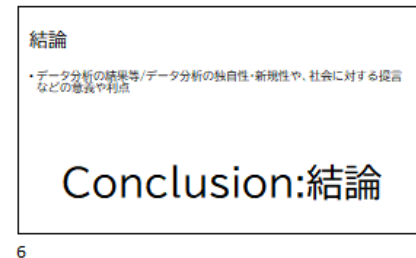
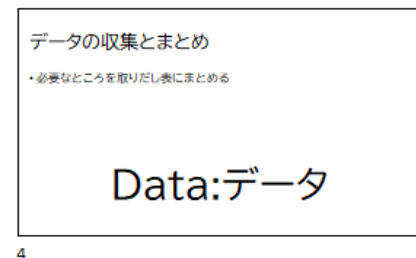
4. A(analysis)分析

- 分析(統計的手法を使う)
- グラフにして全体像をつかむ
- 代表値などをもとめ、データの傾向を見る
- 相関係数を求めて関連を見る
- 標準偏差を求めて散らばりを見る
- 仮説検定によってデータに語るせる

5. C(conclusion)結論

- 結論(スライド作成)
 - テンプレートを使用して作成

- 仮説は検証できたのか
- 発見はあったか
 - 新たな問題点
 - 新たな視点



データやファイルの共有

- 保存場所
 - Teamsのクラスのチーム
 - 情報 I のチャンネル → ファイル → データの分析 → グループ
- 保存するデータ
 - 分析するデータファイル
 - 加工してからアップロード
 - まとめのスライド
 - Teams上で手分けして入力

優れた分析ができれば

- 統計データ分析コンペティション
 - <https://www.nstac.go.jp/statcompe/>
 - 論文を書いて応募してみよう

授業計画 SSDSEでPPDAC

進め方はスライドを参考に
進行管理を確実に

- 条件

- SSDSE(教育用標準データセット)を利用
- PPDACの手法で

- 全3回

- 時間もないのでスゴい発見はできない
- Pまずデータの概要をつかむ/視点を決めて仮説を立てる
- P必要なデータが何か明らかに 1回目
- D必要なデータを集めまとめる
- A分析(統計的手法を使う) 2回目
- C結論(スライド作成) 3回目