

# IoT、ビッグデータにも活用 すぐに使える統計解析

統計を何にどう使うか

1

## 目次

- 本研修について.....1
- 三日間の流れ.....2
- プロローグ.....5
- 第一章.....8
- 第二章.....15
- 第三章.....18
- 第四章.....22
- 第五章.....24
- 第六章.....28
- 第七章.....31
- プロローグ.....33

2

## 1. 本研修について

- (1)教科書についての解説
- (2)本研修の趣旨

## 2. 三日間の流れ(1)

- 一日目
  - この本って何？
  - プロローグ
  - 第一章:データ分析を設計する
  - 第二章:データを事前にチェックする
  - 第三章:分析方法を選ぶ

### 3. 三日間の流れ(2)

- 二日目

- 第四章: ケース実習「新商品の配置問題」
- 第五章: 標準偏差を使おう
- 第六章: グループ間の差の確からしさを検証する

### 4. 三日間の流れ(3)

- 三日目

- 第六章続き
- 第七章: 分析結果の受け止め方と伝え方
- エピローグ

## 5.プロローグ：データ分析の心構え(1)

- 1. よく起こる問題
  - 共有すべきは「データ分析の一連のストーリー」である
    - (1)何のために、何を知ろうとしたか
    - (2)そのためにどんな仮定を置き、どの範囲を考えに入れたか
    - (3)どんなデータを使って、どんな意味合いの数字を出したか
  - 問題が起こらないようにするためには
    - 組織のメンバー一人一人が「データ分析ストーリー」の重要性を認識すること
    - いくつかの基本的な心構えを持つこと
- 2. データ分析って何？
  - データを政治利用しない
  - データ分析とは「発見があること」「知識が得られること」

## 6.プロローグ：データ分析の心構え(2)

- 3. データから考えない
  - データアプローチではなく、課題アプローチをすべきである
  - データアプローチとは
    - (1)こんなデータがある
    - (2)何に使えるか？どう分析すれば使えるか？
  - 課題アプローチとは
    - (1)「課題」を決める
    - (2)課題を解決するための「問い」を決める
    - (3)問に答えるために「データと分析方法」を決める

## 7. プロローグ：データ分析の心構え(3)

- 4. プロセスが全て
  - 計算する前の3つのステップ
    - 第一ステップ：データ分析のストーリーを設計する
      - (1)何のために、何を知ろうとするのか
      - (2)そのためにどんな仮定を置き、どの範囲を考えに入れるのか
      - (3)どんなデータを使って、どんな意味合いの数字を出すのか
    - 第二ステップ：データの信頼性と傾向をチェックする
    - 第三ステップ：「知りたいこと」と「データの傾向」を照らし合わせて、適切な分析方法を選ぶ
  - 計算後の2つのステップ
    - 第一ステップ：分析の結果を正しく評価し、解釈する
    - 第二ステップ：分析の結果を正しく使ってもらえるように表現する

## 8. 第1章：データ分析を設計する(1)

- 1. 分析ストーリーの見える化
  - データ分析ストーリー(再掲)
    - (1)何のために、何を知ろうとするのか
    - (2)そのためにどんな仮定を置き、どの範囲を考えに入れるのか
    - (3)どんなデータを使って、どんな意味合いの数字を出すのか
  - 分析の概念図を描く
    - 問題領域、評価軸、要因はそれぞれ何か？
    - 分析の概念図を描くことによって、「要因の見落とし」や「評価軸の見落とし」を防ぐことができる。



## 9.第1章:データ分析を設計する(2)

- 2. なぜ「分析の概念図」を描くのか？
  - － 分析の概念図は必ず紙に書く。
  - － 紙に書くことで、データ分析の一連のストーリーを見える化し、共有化できる
    - (1)データ分析で重大な見落としを防ぐ
    - (2)無駄なデータ収集を防ぐ
    - (3)作業の途中で迷子になるのを防ぐ
    - (4)結果を相手に納得してもらい、使ってもらえるようになる
    - (5)そもそも「データアプローチ」になるのを防ぐ
- 3. 分析の概念図はこう描く
  - － 分析の概念図を描く6つのステップ
    - (1)問題領域の決定
    - (2)評価軸の決定
    - (3)問題の具体的記述
    - (4)要因の列挙
    - (5)要因の選択
    - (6)部品の配置・連結

9

## 10.第1章:データ分析を設計する(3)

- 4. 問題領域を決める
  - － 問題領域を挙げる
    - ワーク1-1
  - 初めに思いついた問題領域だけで決めない
    - 1)もっと広い可能性と、もっと狭い可能性を「書き出す」  
→見える化
  - － 問題領域を選ぶ
    - 2)どの問題領域が今の課題に最適か、比べて選ぶ
    - 関係者と議論して合意しておく

10

## 11.第1章:データ分析を設計する(4)

- 5. 評価軸を決める
    - 評価軸を挙げる
      - ワーク1-2
      - (1)ボトムアップ式
      - (2)いもづる式
    - 評価軸を選ぶ
      - 評価軸が多すぎると、分析が複雑になり意思決定も難しくなる。
      - 関係者と議論して合意しておく。
  - 6. 問題を文で表す
    - ワーク1-3
- 主語、問題領域、評価軸をつなげて、1つの文で表す
- 言語化することで、やるべきことが明確化、ブレない
  - データを集めた後では、後戻りの手間は大きい

## 12.第1章:データ分析を設計する(5)

- 7. 要因を挙げる
  - ワーク1-4
  - 質より量
  - 漏れなく列挙する
- 8. 要因を選ぶ
  - 重要度で仕分けする(ワーク1-5)
  - 入手しやすさで仕分けする(ワーク1-6)
  - 分析する要因を決める

## 13.第1章:データ分析を設計する(6)

### • 9. 部品をつなげる

- (1)問題領域と評価軸を書く
- (2)評価軸に要因をつなげる
- (3)要因をグルーピングする
- (4)要因同士の関係を推測してつなげる
- (5)分析の流れを説明できるか確認する

(ワーク1-7)

## 14.第1章の理解度チェック

- ①分析の概念図を描くべき5つの理由は何ですか？
- ②分析の概念図を描くための大きな6つのステップは何ですか？
- ③問題領域の決定における注意点は何ですか？
- ④評価軸の決定における注意点は何ですか？
- ⑤要因の列挙における注意点は何ですか？
- ⑥要因の選択を行うための3つの手順は何ですか？
- ⑦要員の配置・連結を行うための5つの手順は何ですか？
- ⑧要員の配置・連結を行うための5つの手順は何ですか？
- ⑨分析の概念図作成における最も重要な注意点は何ですか？



## 15.第2章:データを事前にチェックする(1)

- 1. なぜ事前にチェックするのか？
- 2. データの出所をチェックする
  - (1)データの5W1H
  - (2)一次情報かどうか
- 3. データの全体概要をチェックする(ワーク2-1)  
データ全体を、なじみ深く感じるまで見る
  - (1)データのサイズを確認する
  - (2)データがどのように並んでいるかを見る
  - (3)それぞれのデータの意味合いを理解する
  - (4)値の代替の規模(桁)と、単位を確認する
  - (5)欠損値、外れ値の「有無・多少」を見る
  - (6)その他、目立つ特徴を頭に叩き込む

## 16.第2章:データを事前にチェックする(2)

- 4. 個別の値をチェックする
  - (1)欠損値をチェックする(ワーク2-2,2-3)
  - (2)外れ値をチェックする(ワーク2-4)
  - (3)データの方角をチェックする
  - (4)データをクレンジングする(ワーク2-5)
    - ①欠損値の扱いの判断
    - ②外れ値の扱いの判断
    - ③データの方角の逆転
- 5. データの傾向をチェックする(ワーク2-6)
  - (1)データを再び可視化する
  - (2)値がどの辺に分布しているかチェックする
  - (3)どんな形に分布しているかチェックする

## 17.第2章の理解度チェック

- ①データの事前チェックが必要な2つの理由は何ですか？
- ②データをチェックする4つの大きな視点とは何ですか？
- ③個別の値をチェック・クレンジングする3つのポイントは何ですか？
- ④エクセルで欠損値を特定する汎用的な2つの方法は何ですか？
- ⑤外れ値を特定するための3つのステップは何ですか？
- ⑥欠損値の扱いを判断する3つの手順は何ですか？
- ⑦外れ値の扱いを判断する3つの手順は何ですか？
- ⑧データの方向を機械的に逆転する方法は何ですか？
- ⑨データの傾向をチェックする3つのステップは何ですか？

## 18.第3章：分析方法を選ぶ

- 1. 代表値を使い分ける
  - (1)3つの代表値を知る
    - たくさんのデータを、あえて1つの数字で代表させたもの
      - 平均値、中央値、最頻値のこと(ワーク3-1)
  - (2)代表値の得手・不得手
    - 1)平均値が有効な4条件
      - ①(見かけでなく)本質的に数量の情報
      - ②分布の山が1つ
      - ③極端に非対称な分布ではない
      - ④極端な外れ値がない、または少ない

## 19.第3章:分析方法を選ぶ

- 2)中央値が有効な条件
  - ①ある値が、真ん中より上か下かを知りたい
  - ②平均値より中央値の方が、分布の山に近い
- 3)最頻値が有効な条件
  - ①本質的に名前や言葉を表す分類データ
  - ②数値データを等しい間隔に区切って作った分類データ
- 万能な代表値はない
- どんなに適切な代表値を選んでも、生データに比べれば多くの情報が失われる

## 20.第3章:分析方法を選ぶ

- クロス集計する
  - クロス集計とは?(ワーク3-2,3-3)
    - ①データが要因にどう影響されているか、という構造がわかる。
    - ②潜在的な要因の発見に役立つ
  - クロス集計の注意点(ワーク3-4)
    - 1)先入観を捨てる
    - 2)密接に関連する項目は全て出す
    - 3)データ数や構成比も出す

## 21. 第3章の理解度チェック

- ①代表値を3つ挙げて、意味を説明してください。
- ②平均値が有効な4条件は何ですか？
- ③平均値が有効でないのは、どのような場合ですか？
- ④中央値が有効なのは、どのような場合ですか？
- ⑤最頻値が有効なのは、どのような場合ですか？
- ⑥クロス集計を行う効果は何ですか？
- ⑦クロス集計を行う際の3つの注意点は何ですか？

## 22. 第4章: ケース実習

### 「新商品の配置問題」(1)

- 1. データ分析を設計する
  - (1)問題領域を決める(ワーク4-1)
  - (2)評価軸を決める(ワーク4-2)
  - (3)問題を文で表す(ワーク4-3)
  - (4)原因を挙げる(ワーク4-4)
  - (5)要因を選ぶ(ワーク4-5)
    - 重要度と入手しやすさを同時に考えない
    - 深く悩みすぎない
  - (6)部品をつなげる(ワーク4-6)



## 23.第4章:ケース実習

### 「新商品の配置問題」(2)

- 2. データを事前にチェックする(ワーク4-7)
- 3. 分析方法を選ぶ(ワーク4-8, 4-9)
- 4. 分析を実行する(ワーク4-10)

## 24.第5章:標準偏差を使おう(1)

- 1. 標準偏差って何？
  - データのバラつき(=分布の幅広さ)を、数字で表したもの
  - 「各データと平均値の乖離」から、分布の幅広さを考える
- 2. 標準偏差はこう使う
  - (1)多様性や格差を定量化する、比較する
  - (2)不確実性を定量化する、比較する
  - (3)リスクを定量化する、比較する
  - (4)平均値の信頼性を判断する、比較する
  - (5)品質を管理する



## 25. 第5章:標準偏差を使おう(2)

- 3.標準偏差を計算する
- 4.標準偏差で意思決定する
  - (1)標準偏差を使ってみる(ワーク5-1)
  - (2)外れ値を客観的な基準で特定する(ワーク5-2)
    - ①最大値、最小値、平均値( $m$ )、標準偏差( $\sigma$ )、下限値( $m-3\sigma$ )、上限値( $m+3\sigma$ )を計算(場合によっては $2\sigma$ 基準を決める)
    - ②平均値から $3\sigma$ (または $2\sigma$ )離れているデータを、外れ値と判断
      - ☆「条件付き書式」で、外れ値の所在を一目瞭然にする
    - ③元データ全体をコピー
    - ④コピー先のデータで、「除いても問題ない」と判断した外れ値を除く
    - ⑤外れ値を除くと、新たな外れ値が現れることもある
      - クレンジングを中止するか、外れ値がなくなるまで続けるか、二者択一で方針を決定する
  - (3)2種類の標準偏差

25

## 26. 第5章:標準偏差を使おう(3)

- (3)2種類の標準偏差
  - 「標本標準偏差」と「母集団の標準偏差」があるが、ネット・書籍ともに用語の混乱があるので(Wikipediaにそれなりに詳細な解説がある)、資料や参考書等を読むときは十分に注意されたい。

26

## 27. 第5章の理解度チェック

- ①標準偏差の「具体的な意味」を説明してください
- ②ビジネスにおける標準偏差の有用性を3つ以上挙げてください
- ③標準偏差は、どのような計算式で定義されますか？
- ④エクセルで標準偏差を求める2通りの方法は何ですか？

## 28. 第6章:グループ間の差の確からしさを検証する(1)

- 1.グループ間の大小関係は正しいとは限らない
- 2.大小関係の確からしさをどう考える？
  - (1)確率で考える
    - ①サンプルの平均値は信頼できるとは限らない
    - ②グループ間の大小関係が、全体でもサンプルでも同じであるとは限らない
    - ③確定的には言えないが、全体でも同じ大小関係である「確率」を考えることはできる
    - ④両グループのサンプル分布の重なり割合が小さいほど、全体でもサンプルと同じ大小関係である確率が高い
  - (2)大小関係が確からしい確率は何で決まる？

## 29.第6章:グループ間の差の確からしさを検証する(2)

### – (2)大小関係が確からしい確率は何で決まる？

- ⑤グループ間の平均値の差が大きいほど、また各グループ内のバラつきが小さいほど、両グループのサンプル分布の重なり具合は小さくなる(つまり、全体でもサンプルと同じ大小関係である可能性が高い)
- ⑥データ数が大きいほど、全体でもサンプルと同じ大小関係である確率が高い

### • 3.大小関係の確からしさを判断する

- (1)危険率を見る
- (2)何%以下なら確からしいか？(ワーク6-1)
- (3)「対応なし」と「対応あり」

## 30.第6章の理解度チェック

- ①サンプルデータの平均値に不確実性があるのはなぜですか？
- ②サンプルデータをグループごとに平均して得られた大小関係が、正しいとは限らないのはなぜですか？
- ③グループ間の大小関係の確からしさに影響する3つの要因は何ですか？
- ④グループ間の大小関係の確からしさを定量的に検証できますか？
- ⑤大小関係の確からしさの検証結果「p値」はどういう意味ですか？
- ⑥p値が何%以下なら「大小関係は確からしい」と判断できますか？
- ⑦「対応あり」と「対応なし」を使い分ける基準は何ですか？

### 31.第7章:分析結果の受け止め方と伝え方(1)

- 1. 結果の解釈はここに注意
  - (1)仮説確認バイアス(ワーク7-1)
  - (2)アンカリング(ワーク7-2)
  - (3)フレーミング(ワーク7-3)
  - (4)プライミング(ワーク7-4)
  - (5)疑似相関(ワーク7-5)
  - (6)まとめ

### 32.第7章:分析結果の受け止め方と伝え方(2)

- 2. 結果の表現はここに注意
  - (1)データの集め方(ワーク7-6,7-7)
  - (2)グラフの見せ方(ワーク7-8)
  - (3)言葉の表現の仕方(ワーク7-9)
    - ①統計的な数字も、相手に刺さる実質的な意味合いを伝えないと、説得力を持たない
    - ②比率や倍率のような抽象的な数字だけでは、問題の真の深刻さは伝わりにくい
  - (4)まとめ



### 33.エピローグ

- 全体を振り返って
- さらなる学習のために