

令和 6 年 6 月 25 日現在

機関番号：53302

研究種目：基盤研究(C) (一般)

研究期間：2021～2023

課題番号：21K02894

研究課題名(和文) 気象と飲食を主題としたデータ分析実践学習プログラムの構築と電子教材の開発

研究課題名(英文) Developing learning programs and e-learning materials for data analysis practice on the theme of weather and food and drink

研究代表者

今澤 明男 (Imazawa, Akio)

国際高等専門学校・国際理工学科・客員教授

研究者番号：20148141

交付決定額(研究期間全体)：(直接経費) 1,600,000円

研究成果の概要(和文)：本研究の目的は、現実のデータを用いた演習を中心とするデータ分析「実践」学習環境を構築し、社会人学習者が、各自の目的や持ち時間に合わせてデータ分析を効率的に学べるようにすることである。これを実現するため、1台のパーソナル・コンピュータ上で、実際の気象データを多数格納し簡単な操作で自由に取り出せるデータベース・システムと、それらのデータを対象とした分析を体験するための課題およびそのための分析用プログラムを掲載した例題集と、実際の分析のためのデータ解析環境、の三つを学習者が同時に操作しつつ「実践」の多様な体験が効率的にできる環境を開発した。なお、飲食に関するデータも分析対象に加えた。

研究成果の学術的意義や社会的意義

データ分析の基礎的な理論や解析言語の基礎的な使い方を学んだ社会人が、時間的制約のある中でデータ分析の実践について実際のデータを用いて学べる環境を提示できた。またその中で、日本の実際の気象データや飲食データがデータ分析の実践学習に有効に利用できることが示せた。さらに、教材のデジタル化によって学習者の目的や持ち時間に合わせた多様な学習方法が可能となることも示せた。

研究成果の概要(英文)：The purpose of this study is to create a "practical" learning environment for data analysis centered on exercises using real data so that working adults can learn data analysis efficiently according to their own goals and time. To achieve this, we created an environment on a single computer that allows learners to simultaneously operate three things: 1) a database system that stores large amounts of actual meteorological data and allows it to be freely retrieved with simple operations, 2) a sample problem collection that includes tasks to experience analyzing that data and analysis programs for that purpose, and 3) a data analysis environment for carrying out actual analysis, allowing for efficient experience of various "practical" activities. Food and beverage data was also included in the analysis subjects.

研究分野：管理工学

キーワード：データ分析学習 気象データ 飲食データ

様式 C-19、F-19-1 (共通)

1. 研究開始当初の背景

1.1 社会人のためのデータ分析学習環境充実の必要性

近年、世界の多くの企業においてデータ分析技術を新製品開発やマーケティング等に活用して競争力を高める動きが顕著になってきている。行政の分野でも、データ分析技術を積極的に活用する動きがある。一方、これに呼応して教育機関、特に大学等において「データ分析」についての学習の場が、理・工学領域に留まらず社会科学分野や人文科学分野を含む幅広い領域の学生に向けて設けられつつある。しかし、世界の急激な動きに伍していくためには、企業の現場で現在働いている社会人のためのデータ分析学習環境の充実も必要である。

1.2 データ分析「実践」学習の必要性

「データ分析」についての学習には、①「理論」、②「操作」、③「実践」の少なくとも3種類があると考えられる。このうち「理論」の学習とは、データ分析の理論的側面(主として数理的側面)について学ぶことである。「操作」の学習とは、統計解析用コンピュータ言語や統計解析アプリ等を用いて実際にデータ分析をする際のコンピュータ操作を身につけることである。

一方、「実践」についての学習とは、理論や操作をある程度学んだ学習者が、データ分析技術を効果的に用いて対象の本質に洞察力をもって迫り、目的の情報を導出する過程や、その際の問題点と解決方法を学ぶことである。この学習ためには少なくとも、性質の異なる各種の実際のデータを対象とした分析を多数回繰り返して失敗や成功を体験できる学習教材・環境が必要である。また、忙しい社会人学習者の要求や持ち時間に合わせて学習内容や学習時間を調整でき、かつ関心と意欲を持って学習できる教材・環境が望まれる。

2. 研究の目的

本研究の目的は、現実のデータを用いた演習を中心とするデータ分析「実践」学習環境を構築し、時間のない社会人が、各自の目的や持ち時間に合わせて効率的に学べるようにすることである。具体的には、1台のパーソナル・コンピュータ上で、①実際のデータを多数格納し自由に取り出せるデータベースと、②それらのデータを対象とした分析の「実践」を体験するための例題集と、③実際の分析のためのデータ解析環境、の三つを学習者が同時に操作して「実践」の体験学習を効率的にできる環境の開発ある。

3. 研究の方法

3.1 使用データの選定

(1) 気象データの利用

本研究では、分析対象データとして、まず日本国内の気象データを用いる。気象に関しては、日本では過去1世紀以上の大量のデータが気象庁によって収集・整理され、同庁のホームページなどを通じて広く一般に公開されている。一方、例えば気温は、地点・季節・時刻・天候により異なるとともに、緯度や標高など基本的決定要因との間に相関を持ち、1日や1年といったサイクル(循環)や、温暖化・寒冷化などのトレンド(傾向)も有している。さらに降水量などは、いわゆるロングテールの分布に従っていて、単なる平均値での分析の妥当性が低い。このように、気象データは分析対象として豊富な特性を持っている。さらに、たとえば気温の場合、これを温度、すなわち数値で表すことに抵抗を感じる人はいない。また、測定値は「暑い」「寒い」などと直感的に理解でき、緯度や標高、季節や時刻などの基本的決定要因についても広く知られている。すなわち、広く一般の人々にとっても現象の数値化に抵抗がなく、測定値の意味を容易に理解でき、「洞察力」を持って分析・解釈ができる対象である。

(2) 飲食データの利用

気象や気候と食との関係がしばしば指摘されている。例えば気温と冷たい飲食物の売上との関係がそうである。しかも、各食品の購入額などのデータは、気象データと同様に比較的入手しやすく、また学習者の親近感や関心も高い。そこで、飲食のデータも教材に取り込むことにした。

3.2 解析環境

パーソナル・コンピュータ上での解析環境としては、統計解析向けのプログラミング言語及びその開発実行環境である「R言語(以下、R)」を用いる。Rはオープンソース・フリーソフトウェアであり、インストール時の設定も比較的簡単である。また、もともと統計解析・データ解析のために開発されたものであり、解析のための豊富なライブラリーが用意され、煩雑な解析作業が簡単に実行可能である。信頼性も高い。

3.3 演習課題（例題）集作成方針

ただ単に与えられたデータに一つの解析手法を一回適用して終わりとなるような課題ではなく、複数回の異なった分析の積み重ねを通してデータが持つ本質に迫る課題を目指す。さらに、Rのプログラムも掲載し、電子化された課題集からコピーしてRのコンソールに貼り付けて実行するだけでも分析ができるようにする。また、各分析結果の概要となるグラフなども掲載し、プログラムの実行をせずとも通読するだけでも学習できるようにする。

3.4 データベース・システムの作成方針

気象データベースからは、簡単な操作で目的とするデータが取り出せるようにし、学習の効率化を図る。さらに、取り出したデータは csv 形式で保存し、Rのみならず Excel にも簡単に取り込めるようにして作業の簡便化とシームレス化を図る。

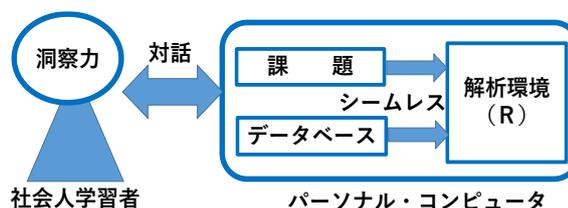


図1 学習環境のイメージ

4. 研究成果

4.1 データベース・システムの開発

(1) データ

データベースには以下の気象データを格納した。

- ア) 観測地点：①日本の平均気温算出に使われる 15 地点と、②各都道府県を代表する 47 地点の合計 59 地点（4 地点が重複）。
 - イ) 気温に関するデータ：毎時の気温、毎日の平均気温、毎月の平均気温、毎年の平均気温、
 - ウ) 降水量に関するデータ：毎時の降水量、毎日の降水量、毎月の降水量、毎年の降水量
 - エ) データの対象期間：毎時（2023 年 1 月 1 日 1 時～12 月 31 日 24 時）、毎日（2022 年 1 月 1 日～2023 年 12 月 31 日）、毎月（2019 年 1 月～2023 年 12 月）、毎年（1876 年～2023 年）。
- なお飲食データは、このデータベースではなく、利用する各課題の専用フォルダに置いた。

(2) 操作

操作画面で、取り出したいデータの、①保存先フォルダ、②データの種類（1 時間ごと、日ごと、月ごと、年ごとの別）、③観測地点（複数選択可）、④データ項目（気温または降水量、複数選択可）、⑤期間または時点、の各項目についてクリックにより選択すると、指定保存先に csv 形式のデータファイルとして保存される。ファイル名は自動で付され、データの種類と項目および始点と終点で決定される。ただし、ファイル名を手動で変更することは可能である。

4.2 演習課題（例題）集の作成

(1) 課題内容

課題は以下の 4 種に大別できる。括弧内は主たる学習内容である。

- ア) 気象データの時系列分析（トレンドの抽出、周期の推定、成分分解、数理モデルの当てはめと予測など）
- イ) 食品購入データと気象データの相関の把握（相関係数、回帰分析など）
- ウ) 食品購入データに基づく主要都市の分類と地域性の把握（距離算出、クラスター分析など）
- エ) 気象の時系列データ間の変動パターンの変動に基づく観測地点の分類（標準化、距離算出、クラスター分析など）

なお、以上の課題実践においては、分布形の確認、データの変換、階差の利用、定常性の確認などが適宜含まれている。

(2) 課題の構成要素と操作

学習者は、①課題の内容、②使用データ、③使用する主たる分析方法、から学習したい課題を選択できるようになっている。また、各課題は最終目標達成に向け、複数のステップごとの小課題で構成される。学習者は各小課題を実践しながら、対象データに対する理解を深めつつ最終目標に到達できるようにしてある。一方、小課題ごとの R でのプログラムならびに実行結果の概要も掲載してある。プログラムはコメントを十分に付し、内容がすぐ理解でき、学習者による自分に合わせた改変も可能なようにしてある。

4.3 学習方法と演習の拡張性

学習者に時間がない場合は、関心のある課題に目を通す。小課題・プログラム・実行結果が掲載されているので、これだけでも学習できる。やや時間に余裕がある場合は、プログラムをコピーして実行すると、内容を具体的に確認できる。余裕があれば、課題で指定されたデータとは、観測地点・時点や期間などが異なるデータを自らの関心に合わせてデータベース等から取得して分析を実行するとよい。学習者にとっての新しい発見が期待できる。習熟してきたら、プログラムを自らに合わせた改変して分析を実行すると実践力がより高まると期待される。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 Akio Imazawa and Nobuyuki Naoe
2. 発表標題 Air Temperature Data as a Learning Material for Time-series Analysis
3. 学会等名 IEEE 12th International Conference on Engineering Education (ICEED 2023) (国際学会)
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------