

Python を用いた重回帰分析をテーマとしたレベル判別テストの試作

上野 春毅*¹ 長谷川 理*² 小松川 浩*¹ 山川 広人*¹

*¹ 公立千歳科学技術大学 理工学部

*² 武蔵野大学 データサイエンス学部

h-ueno@photon.chitose.ac.jp

本研究は、各学校段階のプログラミング教育の接続を意識した教材開発の一環として、Python を用いた重回帰分析をテーマとしたレベル判別テストを試作し、これを報告するものである。テストは CBT システムを使用し、学習者の知識習得度を 7 段階の難度レベルで判定した。本稿は、その実践の結果について検討を図る。

1. はじめに

各学校段階でのプログラミング教育の全面実施が進む中で、高校情報 I・II ではプログラミング言語を用いてデータの分析・活用の基礎を学ぶ内容が取り上げられる。さらに大学では初年次教育とも連係した数理データサイエンス・AI 教育が重要視されている。こうした中で、回帰分析を用いたデータ分析手法とプログラミング言語を用いた実践は、高校・大学の両段階で接続的に学ぶ内容として捉えることができる。本稿は、各学校段階のプログラミング教育の接続を意識した教材開発の一環として、Python を用いた重回帰分析をテーマとしたレベル判別テストを試作し、これを報告するものである。

2. ベースシステムと試作した教材

本稿で報告するテストは、筆者らが開発した CBT システム⁽¹⁾をベースとして、学習者の Python による重回帰分析を対象とした理解度をレベル別で判定できるよう狙ったものである。ベースシステムでは、教員が事前に 7 段階の難度にレベル分けされたテスト問題を一定数用意することで CBT を実施できる。受験者が CBT で出題された問題に 1 問ずつ答案を送信するたび、システムはその正否情報から受験者が正答できる確率の高い難度レベルを推定し、難度レベルに応じた問題を新たに出題する。これを繰り返すことで、CBT の終了時に受験者が安定し正答できる難度レベルを判定結果(知識習得度)として示す。

テスト問題は、7 段階の難度(レベル)に沿って整備した。問題の内容は、回帰分析のプログラミングに焦点をあてて、用語やライブラリなどの知識を修得して様々な事柄に応用できることを目標とする。表 1 に既に開発した CBT 教材⁽²⁾と本稿の段階で開発した重回帰分析の単元の教材一覧を示す。各単元を共通して、最も低いレベル 1 は用語の定義や理解を問う問題とした。具体的には、

表 1 整備したテストの単元と問題数

単元名	総問題数(問)
1. Python 入門	54
2. 条件分岐	71
3. 反復処理	57
4. リスト・タプル	98
5. 辞書・集合	68
6. 関数	60
7. クラス	61
8. ライブラリの利用	81
9. 重回帰分析 ※本稿段階で追加	72

回帰分析において、ある変数に対して影響を及ぼし、要因となると考えられる変数を【1】という。



レベル1

scikit-learnを用い、説明変数x、目的変数yとして重回帰モデルmodelを学習させる。以下のコードを完成させなさい。

```
from sklearn.linear_model import LinearRegression
```

```
model = LinearRegression()
```

```
model.【1】(x, y)
```

【1】

レベル4

説明変数をx、目的変数をyとして説明変数を標準化し、重回帰分析を実行する。また、重回帰モデルを用いてテストデータの予測を行う。以下のコードを完成させなさい。

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.linear_model import LinearRegression
```

```
from sklearn.【1】 import StandardScaler
```

レベル7

図 1 上から、重回帰分析レベル 1, 4, 7 の設問例

説明変数や目的変数といった用語の定義の理解を問う問題とした。レベル 2 では、具体例をもとに適する用語を選択させる問題とした。レベル 3 で

は、分析に用いる Python のライブラリの名称や関数名を答えさせる問題とした。レベル 4 ではレベル 3 のライブラリの名称などをプログラムに当てはめる問題とした。具体的な例として、import 文の穴埋め問題があげられる。レベル 5 では、データの読み込みから回帰のモデルを生成するプログラムの穴埋め問題とした。レベル 6 では、訓練データ・テストデータといった前処理を加えた穴埋め問題とした。レベル 7 では、データ読み込み～モデル生成～モデルを用いた予測までの一通りのプログラムの穴埋め問題とした。

3. CBT の検証結果

2章で整備した教材を、A大学の授業に適用して検証する。A大学は理系の単科大学であり、授業は2年次のデータサイエンス系の必修科目である。学力の背景として対象者は1年次に情報系の科目でC言語の変数・繰り返し文・条件文を学んでいる。よって、Pythonは初めて用いるものの、他のプログラミング言語の既習経験が一定程度ある前提で、データサイエンスについてプログラミング実習を通じて学ぶ授業である。この中で、授業進度にあわせたCBTの教材による予習と、授業直前の知識習得度のチェックを課した。2022年度および2023年度のCBTの結果を表2・3に示す。ここでの「Python基本文法」とは表1の「1. Python入門」～「5. 辞書・集合」をひとまとめに出題したCBTに対応する。「Pythonライブラリの利用」はデータを扱うライブラリであるNumpyやPandasに関する教材である。

表3の結果から、CBTの高いレベルを取得できていることがわかった。昨年度の表2と比較しても同程度の高い結果となっていることがわかり、重回帰分析部分の利用の前提となる「Pythonライ

ブラリの利用」の単元までに対し、両年度の学生の知識習得度は概ね同様の結果となることがわかった。

なお本稿では重回帰分析部分の単元を追加したが、この単元の授業での試用は執筆時点で開始段階にある。発表時に、重回帰分析部分の試用結果までを踏まえた検証結果を報告する予定である。

4. まとめ

本稿は、各学校段階のプログラミング教育の接続を意識した教材開発の一環として、Pythonを用いた重回帰分析をテーマとしたレベル判別テストを試作し、試用を通じた検証を開始した。高校情報I・IIや、初年次教育と連係した数理・データサイエンス・AI教育の実践を考えると、検証環境を理系だけでなく文系などの幅広い利用環境での実践結果と改善が求められる。

謝辞

本研究はJSPS科研費JP20K03234の助成を受けたものである。

参考文献

- (1) 上野春毅ら：段階的な学習目標を持つ反転学習モデルのための適応型学習システムの開発, 教育システム情報学会紙, Vol.37, No.3, pp. 212-217 (2020)
- (2) 山川広人ら：Pythonプログラミングの知識習得度テストの試作と実践結果の検討, 日本情報科教育学会第15回全国大会公演論文集, pp. 2-3 (2022)

表 2 2022 年度 A 大学データサイエンス系実習科目の CBT の結果

	レベル 1	レベル 2	レベル 3	レベル 4	レベル 5	レベル 6	レベル 7
基本文法	0%	1%	1%	0%	4%	17%	77%
関数	1%	1%	2%	1%	5%	12%	78%
ライブラリ の利用	0%	0%	2%	2%	9%	11%	76%

表 3 2023 年度 A 大学データサイエンス系実習科目の CBT の結果

	レベル 1	レベル 2	レベル 3	レベル 4	レベル 5	レベル 6	レベル 7
基本文法	0%	0%	2%	2%	5%	9%	82%
関数	1%	2%	1%	2%	4%	7%	84%
ライブラリ の利用	0%	0%	1%	2%	6%	4%	86%