

R2-6

表計算・プログラミング両面による統計分析指導実践省察

私立佼成学園中学校・高等学校非常勤講師 布村 寛

要旨 「情報の科学」では、問題解決のデータサイエンスに係る指導を、表計算ソフトによる実験を軸に実施してきた。一方でプログラミング指導に関しては、近年は Web 理解を優先して JavaScript を採用してきたが、機械学習への進展を視野に見直した。「情報 1」移行期である昨年度、同一集団に対する Excel, Python 両面の統計実験指導を実施した。数学 I との連携に鑑みながら、その比較省察を行う。

1. 旧課程における指導計画と表計算統計指導

旧課程における年間計画概略は「情報とコンピュータ(約 20 時間), 「ネットワーク～DB～統計～シミュレーション(約 20 時間)」 「プログラミング(Pov-Ray, VBA, JavaScript 等約 10 時間)」の流れで実験中心の授業展開で実施してきた。

表題の指導は、問題解決の項目として DB 実験からデータを出力し、表計算ソフトの数式・関数・マクロ・分析ツールを活用し実施してきた。数学 I での統計指導が 1 月以降なので、情報が先行していた。対象学年：1 学年、教科書：実教出版

昨年の全高情研実践報告に対する、プログラミングを用いたデータサイエンスに言及するご助言があり、後半の指導計画を修正して Python を活用したデータサイエンスまで範囲を拡大した。

1.1 表計算によるガウス分布

乱数と VBA を組み合わせた確率分布実験によるガウス曲線・パレート図理解から始め、数学 I の統計範囲へ展開し、新体力テスト分析へ繋いできた。応用では回帰・移動平均などへ展開した。

コイン 10,000 個を投げて表なら 1、裏なら 0 として確率 1/2 の乱数を B 列に発生させる。これを 10,000 回試行した結果を F 列に記録し、ヒストグラムを作成する。右下のスイッチを押して若干時間経過後に処理が終了したら F2～F10001 の範囲でヒストグラムを作成する。

coin1	=RANDBETWEEN(0,1)	trial1	4968
coin2	=RANDBETWEEN(0,1)	trial2	4985
coin3	=RANDBETWEEN(0,1)	trial3	4976
coin4	=RANDBETWEEN(0,1)	trial4	5071
coin5	=RANDBETWEEN(0,1)	trial5	5012
coin6	=RANDBETWEEN(0,1)	trial6	4996
coin7	=RANDBETWEEN(0,1)	trial7	4916
coin8	=RANDBETWEEN(0,1)	trial8	5025

**10,000回
試行スイッチ**

```
Dim a As Integer
For a = 2 To 10001
Cells(a, 6) = Cells(10002, 2)
Next a
```

coin100	=RANDBETWEEN(0,1)	trial100	5000
sum	=SUM(B2:B10001)	max	=MAX(F2:F10001)
		min	=MIN(F2:F10001)
		med	=MEDIAN(F2:F10001)
		ave	=AVERAGE(F2:F10001)
		stdev	=STDEV.P(F2:F10001)

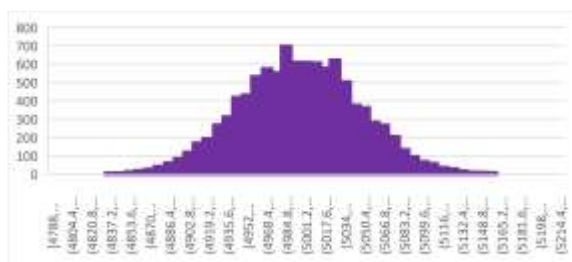


図 1 (表計算ソフトを用いたガウス分布実験)

1.2 基礎統計

実習を伴う形で諸統計値、標準偏差、ヒストグラム・パレート、相関、回帰・チャート・移動平均分析、各種モデル化・シミュレーションに進む。

1.3 モデル化・シミュレーション

モンテカルロ法における π 算出などの試行実験では、セル中の乱数処理、数式及び反復マクロの説明で実習進行の数学的な意味を理解させた。

1.4 移動平均・回帰・チャート

データをセルにインポートした後、初回は手入力操作で数式処理を実施させ、次に分析ツールを用いた可視化で処理の意味を吟味させた。

2. Python による問題解決

クラウド活用やコンパイルまでの一括処理の利点から、開発環境を Google Colaboratory に設定。以下 2-3 以降は表計算実施済データを使用した。

2.1 言語作法目録

(1)数値型、演算子 (2)テキスト型 (3)ブール型 (4)型変換 (5)変数 (6)input 関数 (7)条件分岐 (8)反復 (9)乱数 (10)描画ツール (11)関数モジュール (12)配列 ①基本②集合③リスト型④辞書型 (13)統計用 statistics モジュール (14)文字数 (16)ヒストグラム ax.hist (16-1)正規分布ヒストグラム (16-2)乱数による正規分布作成 (16-3)正規分布からヒストグラム (17)散布図 ax.scatter (18)箱ひげ図 ax.boxplot (19)並べ替え・ソート sort (20)数学用 math モジュール (21)チャート分析・ローソク足 (22)移動平均 (23)～現在教材作成中 (1)～(12)は全体共通で同時授業展開を行い、(13)以降は各生徒の自由学習とした。

2.2 思考コーディング課題例

- ・対話的プロンプト(入力ダイアログ)活用
- ・年齢を回答させて、還暦までの年数計算を返す。
- ・身長、体重を回答させて、BMI を算出し以下の判定を返す (bmi25 以上 over weight, 18.5 未満 under weight, 18.5 以上 25 未満 healthy weight)
- ・入れ子課題として x 年 y 組 z 番の全表示。
- ・n 角形描画、一辺の長さを m とする次の図形。
- ・円：中心座標(200,200)、半径 r=100 の円

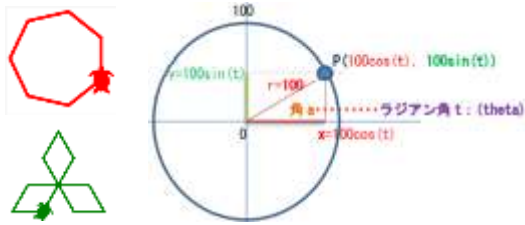


図 1 Python を用いた図形描画コーディング

2.3 Python によるガウス分布

・構文を転記させ、パラメータを可変させる。

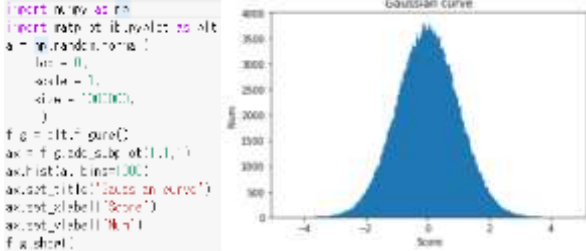


図 2 Python を用いたガウス分布実験

2.4 Python による基礎統計例

統計用モジュール導入 import statistics を用いたヒストグラム描画

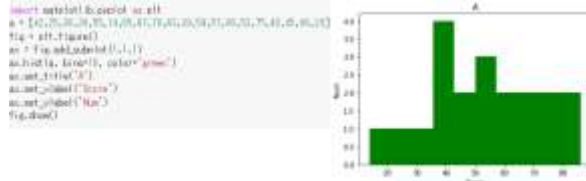


図 3 Python を用いたヒストグラム描画実験
統計用モジュール導入 matplotlib.pyplot を用



図 4 Python を用いた箱ひげ図描画実験

2.5 Python 応用例

表計算で指導した移動平均による平滑化とトレンド把握を Python にも拡張した。

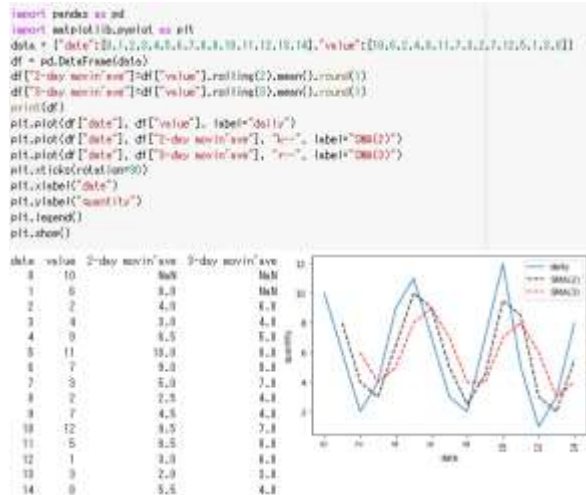


図 5 Python を用いた移動平均分析

3. 数学 I との教科横断を踏まえた省察

情報の科学における統計分析指導は、数学 I 第 5 章「分析」への導入という形で、その内容に関するほぼ全てにおける実習を、可視化を含めて担ってきた。その視点で考察すると、カリキュラム上の進度において数学で既習ならば、Python によるコーディング指導の方が、教科連携を含む様々な学習を効率的で発展性なものにするであろう。逆に情報が先なら、表計算によりセルに直接データや数式を入力できる点が、概念理解を支援するであろう。モジュールやライブラリ活用以前の順次・分岐・反復や変数・関数処理までのアルゴリズム実装であれば順不問と考える。

3.1 表計算によるデータサイエンス指導の省察

数学 I での未習段階で、確率分布(ガウス曲線)、分散・標準偏差とヒストグラム、偏差値、四分位数と箱ひげ図・相関係数・回帰などの概念に対して数式を用いた実証実験の展開が可能である。手計算に近く、結果として数学 I の学習への実戦的な導入となる。分析ツール・グラフ描画ツールに関しては、Python におけるモジュールやライブラリと同じくブラックボックスとして活用できる。

3.2 Python によるデータサイエンス指導の省察

モジュールやライブラリを用いない範囲の図形描画、関数処理は思考を伴う学習として思考的展開を全体で進められた。一方、それらを活用した統計・分析においては、数学や物理の公式を用いた演繹的な解法に近い印象で、帰納法的に原理を発見させる手順に至らなかった。集団全体でゼロベースからコーディング作法に取り組ませ、関数設定以降は、進捗を生徒個々の能力に委ね書写・データ入替で理解を促した。課題の最後まで辿りつき Python の利点をコメントできたのは約 270 名の内、興味関心の特に強い 2 名であった。ライブラリ活用でのエラー対応の指導では、小論文指導並みの困難があり、1 科目内での限界を感じた。

3.3 Python と学習環境

Python では Google 上に開発環境を設定可能であり、DX の方向性を鑑みると利がある。しかし数学 I の学習範囲の実証という点では、表計算における直接操作にスピード感がある。本件に対する二面的指導は年間授業時間の 3 割に達し、冗長であった。

参考文献

- (1) 片岡巖発行、データサイエンティスト養成講座
- (2) 高橋淳一 野村嗣 西村隆宏 水上ひろき他 著 データサイエンティスト養成講座登竜門編 (1)(2)共に 技術評論社