

# オンライン授業によるデータサイエンス教育

山守一徳<sup>†</sup>

## Data Science Education through Online Classes

Kazunori YAMAMORI<sup>†</sup>

<sup>†</sup>三重大大学教育学部

### 概要

データサイエンス教育は、多くの大学で教えようとし始めており、教育のできる専門家が少ないことから、教育内容を洗練できずに四苦八苦している状況にある。三重大では2020年から情報科学基礎の授業の中にデータサイエンス教育を埋め込み、1年生相手に授業をし始めたが、オンライン授業体制の中での授業であったため、充分教育できたか不安に感じる。本文では、2020年に行ったデータサイエンスの授業内容の紹介と共に、2021年以降に何を教えるべきかについて述べる。リテラシー教育にプラスアルファした程度のデータサイエンス教育でなく、確率統計学を学んでいない1年生前期の段階で、全学生に教えるべき内容について述べる。また、教育学部のオンライン授業で用いたテストの方法も紹介し、再びオンライン体制になった時に、試験をしっかりと行って教育の質を落とさないようにするための参考にしていただけたらと思う。

### キーワード

統計解析、Python、R、オンラインテスト

#### 1. はじめに

データサイエンスは、多くの大学で全学生に教えようとし始めている状況にある。ビッグデータが入手可能になり、データ解析をすることで、新しい知見を得たいとするニーズに応えるように、PythonやRのデータ解析ツールが出回っている。2016年12月に滋賀大学、大阪大学、北海道大学、東京大学、京都大学、九州大学を数理、データサイエンス教育強化の拠点校として文科省が認定し、他大学へ波及することを望んでいる。高校の「情報」の教科内容も、データサイエンス教

育の内容を盛り込み、日本全体の情報処理能力をレベルアップしようとしている。三重大でも、情報科学基礎の授業の内容をデータサイエンス教育に切り替えようとする取り組みを始めたところである。

2020年の情報科学基礎では、実データを利用して、EXCELで平均と分散、可能であれば相関などを教えると共に、グラフを描かせるという課題と統計ソフトRを体験させるという課題を盛り込み、15回の授業の中の5回はデータサイエンス教育の内容にするという変更が行われた。Rを教える回では、担当の先生が代わって授業分担することにな

った。その中でコロナウィルスの影響でオンライン授業になり 15 回授業が 12 回授業になったため、5 回予定のデータサイエンス教育が行われたのかどうか疑わしい。そして、2021 年の情報科学基礎に向けて、再び変更が検討されている状況にあるが、データサイエンスの内容を教えらるる教員が不足していることから、十分な内容に仕上がるのか不安な状況にある。本来なら教えた方が良い内容であっても、教えることができないという理由から、相変わらずのリテラシー教育が盛り込まれていると思われる。

1 年生前期の段階では、確率統計学も履修前であり、本来ならば、確率統計学を履修した後、統計解析の内容を履修し、一方で Python や R などのプログラミングも履修した後に、Python や R を使った統計解析の演習を行うのが理想である。全学生にそれだけの履修をさせるには、他の授業単位を削るだけでなく、教える教員を配備しないとけない。しかし、カリキュラム改革最小に、教員配備も最小にという条件の下で、どのように教えたら良いのかを述べたいと思う。

## 2. 高校の「情報Ⅱ」の内容

高校の普通教科「情報」の授業は、「社会と情報」と「情報の科学」の 2 種類存在し、高校ではどちらかを教えれば良いため、「社会と情報」が 8 割、「情報の科学」が 2 割の割合で高校が採択しているのが現状である。その内容が、2022 年から更新され、情報Ⅰ（仮称）と情報Ⅱ（仮称）の二つになると新しい学習指導要領に記されている。「情報Ⅰ」は全員必修で、おそらく 1 年生に担当され、情報Ⅱは」選択科目で、おそらく 2 年生に担当される。

高等学校情報科「情報Ⅱ」教員研修用教材（本編）：文部科学省によると、3 章において、「情報とデータサイエンス」の章が設けられ、各項目からキーワードを抽出すると、以下の用語が記載されている。

### 第 3 章「情報とデータサイエンス」

#### 学習 11 データと関係データベース

データの信憑性と信頼性、選択バイアス、生存バイアス、ビッグデータ、JSON 形式、Python の pandas パッケージ

#### 学習 12 大量のデータの収集と整理・整形

Web スクレイピング、データクリーニング、ワイドフォーマット（横持ち形式）とロングフォーマット（縦持ち形式）、欠損値、異常値

#### 学習 13 重回帰分析とモデルの決定

重相関係数 R、寄与率（決定係数） $R^2$ 、自由度、統計ソフト R、増減法（ステップワイズ法）

#### 学習 14 主成分分析による次元削減

主成分得点、分散共分散行列、相関行列、固有値分解

#### 学習 15 分類による予測

教師あり学習、決定木、k-近傍法

#### 学習 16 クラスタリングによる分類

階層的クラスタリング、k-means 法、アソシエーション分析

#### 学習 17 ニューラルネットワークとその仕組み

勾配降下法、バックプロパゲーション（誤差逆伝播法）

#### 学習 18 テキストマイニングと画像認識

物体検出: Tiny YOLO

これらの用語を見ただけでも、教える内容がかなりレベル高いということがわかる。高校で、教えることのできる教員が見当たらないであろうと思われる点と、「情報Ⅱ」は、必修履修科目でも受験科目でもない点から、多くの高校で開設されない可能性が高い。しかし、文科省が求めているレベルがわかり、大学において、このレベルが教えられないとなると高校以下に下がってしまうことが起きる。

## 3. 「情報科学基礎」で教えてみた内容

2020年度の授業は、コロナウィルス対策のために、すべてオンライン授業となり、授業回数も12回しかなかった。その中で、データサイエンスの内容を盛り込む設計がなされ、従来のリテラシー教育も残しつつ、PythonとRを使った演習を行った。受講対象が、数学教育と情報教育コースの1年生相手であるため、TeXの演習も行った。

リテラシー教育では、WORDの演習1回、EXCELの演習1回、CSSを含むHTML作成の演習1回、情報倫理教育を1回行った。その中のEXCELの演習の中にデータサイエンス教育に関係付けさせるために、グラフ作成を行わせた。使うデータを総務省統計局のサイトからxls形式のファイルをダウンロードさせ、和暦を西暦のデータに直して、マーカー付き折れ線グラフを描かせた。このEXCELを使ってグラフを描かせる演習は、受講生の約半数が高校の情報の授業で演習したことがあると答えていた。

データサイエンス教育としては、Pythonを使って、箱ひげ図と線形単回帰分析の回帰直線を描かせる演習を1回と、k平均法を使うクラスタリングと主成分分析を行う演習を1回行った。また、Rを使って、決定係数と有意確率を求める演習を1回行った。

#### (1) Pythonを使った演習

Pythonのプログラミングは教えておらず、演習に使うPythonプログラムを渡し、その中の記述を変更させることで、課題を与えた。Pythonを稼働させるために、Anaconda3のインストール手順の動画を用意し、授業前までに自分のパソコン上へのインストール作業を各自で行わせた。Pythonのプログラムファイルも事前にダウンロードさせて、Jupyter Notebookを起動してPythonプログラムの動かし方も授業前までに確認させた。

Anaconda3をインストールせずに、Google Colaboratory <https://colab.research.google.com> に接続して、学籍番号@m.mie-u.ac.jpの

アカウントでログインして、現れたウィンドウにPythonプログラムを打ち込むと実行できるが、オーソドックスな方法を教えた。

Pythonは、空白文字の有無でエラーになることがあり、多くのプログラミング言語と比べると異質な言語である。そのため、最初に教えた方がよい。

箱ひげ図を描かせる演習では、成績のデータを使って、四方位範囲を教えた。第1四方位と第3四方位は高校では習っていないため、説明する必要があった。外れ値を持つデータを使うことで、箱ひげ図のひげの部分から外れた位置に○で表示されることも示した。また、標準偏差を平均値で割った変動係数も教えた。

線形単回帰分析の演習では、共分散行列、相関係数を教え、寄与率とも呼ばれ、予測式が実測値をどれほどうまく予測しているのかを示す決定係数 $R^2$ 、偶然に実際と反した数値が統計量として計算されてしまう確率を示すP値も教えた。これらの数値は、Pythonのプログラミングを使うと容易に求めることができる。

k平均法を使うクラスタリングでは、k平均法のアルゴリズムを教え、クラスタリングで色分けされた散布図を表示させた。kの数値によって、きれいに分類されなくなることを示した。

主成分分析の演習では、30次元の乳がんデータを使って、第1主成分と第2主成分の2次元散布図を表示させ、分散が最大となる方向を理解させた。

#### (2) Rを使った演習

Rのプログラミングを教えておらず、演習に使うRプログラムを渡し、その中の記述を変更させることで、課題を与えた。

Rを稼働させるために、RStudioのインストール手順の動画を用意し、授業前までに自分のパソコン上へのインストール作業を各自で行わせた。Rのプログラムファイルも事前にダウンロードさせて、Rの動かし方の

PowerPoint も事前に与え、授業前までに R の動作の確認をさせた。

決定係数と有意確率を求める演習では、東京都の 2003 年から 2012 年の天候データとアイスクリームの売り上げのデータを用意し、自由度調整済み決定係数を教えた。モデル全体の有意確率である P 値が、0.05 を下回ると「この結果は偶然とは言えない」を意味する「有意である」という用語も教えた。パラメータの推定値を標準偏差で割った t 値や、各パラメータの有意確率である  $\Pr(>|t|)$  の値についても教え、 $\Pr(>|t|)$  の値は 0.05 を下回ると「有意である」と言えることを教えた。

## 4. 教えるべき内容

### (1) 統計的仮説検定

帰無仮説と対立仮説を立てて、いずれが正しいかの仮説検定をする方式は、様々な分野で利用されるため、理系学生も文系学生も全員が学習すべき内容であると思われる。否定したい相手の主張が帰無仮説、通したい自分の主張が対立仮説である。母集団に関する仮説を標本から得た情報に基づき検証する。同じ帰無仮説でも、対立仮説の取り方で両側検定になったり片側検定になったりして逆の結果が得られることがある。

よく出される問題は、母平均、母分散の分かっている母集団から取ってきた標本の標本平均が母平均と差があった時、この差に意味があるかどうか、有意差であると言えるかどうかという問題が出される。

この問題を解かせるために、正規分布から標準正規分布に置き直すために、確率変数の  $X$  を  $\frac{x-\mu}{\sigma}$  に置き換えるというテクニックを教える必要がある。  $X$  が離散型変数でも連続型変数でも、 $Y=aX+b$  の線形変換を行うと、期待値  $E[aX+b]=aE[X]+b=a\mu+b$ 、分散  $V[aX+b]=a^2V[X]=a^2\sigma^2$  となることは式の

変形から確認ができる。そこで、 $a=\frac{1}{\sigma}$ 、 $b=\frac{-\mu}{\sigma}$

とすると、期待値 0、分散 1 の標準正規分布になる訳である。ここで、注意しないといけないのが、標本数  $n$  の場合、 $X$  を  $\frac{x-\mu}{\sqrt{\sigma^2/n}}$  に置き換えることである。

### (2) 不偏分散、標本分散

母集団から標本  $n$  個取り出した時、不偏分散は  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 、標本分散は  $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  であるが、高校では分散の式を後者で教えているため、前者を使うことを教える必要がある。不偏分散の平均は  $E(s^2) = \sigma^2$  となり、母数と等しくなるため、不偏性を持つと言われる。

平均値は母平均も標本平均も  $\mu$  となり、違和感がないが、母分散  $\sigma^2$  の時、標本平均の分散は、 $\frac{\sigma^2}{n}$  となることから、混乱しやすい箇所である。母集団の分布が正規分布であれば、 $\bar{X}$  の分布も正規分布になるが、標本数を大きくすれば、母集団分布が正規分布でない場合でも、 $\bar{X}$  の分布は正規分布で近似できる。この中心極限定理も教えておいた方がよい。

### (3) 区間推定

標本から母集団の平均や分散を予測する区間推定は、統計解析の基礎であろう。母分散が既知の場合は、正規分布に従い、母平均の 95% 信頼区間は、 $[\bar{X} - \frac{\sigma}{\sqrt{n}} \times 1.96, \bar{X} + \frac{\sigma}{\sqrt{n}} \times 1.96]$  となる。

母分散が未知の場合は、不偏分散  $s^2$  を使って、自由度が  $n-1$  の t 分布に従う。母平均の 95% 信頼区間は、 $[\bar{X} - \frac{s}{\sqrt{n}} \times 2.262, \bar{X} + \frac{s}{\sqrt{n}} \times 2.262]$  となる。

母集団が正規分布で、母平均、母分散不明の時、母分散を推定する場合、不偏分散  $s^2$  を

使って、 $\frac{(n-1)s^2}{\sigma^2}$ が自由度(n-1)の $\chi^2$ 分布に従うことを利用し、例えば95%信頼区間は、n=10の場合  $2.700 \leq \frac{9s^2}{\sigma^2} \leq 19.02$  から、 $s^2$ を計算した後 $\sigma^2$ の信頼区間を求める。 $\chi^2$ 分布は自由度によって形状が大きく変わり左右非対称である。分布表の上側%点と下側%点の読み取り方も教える必要がある。標準正規分布に従う独立なn個の変数の2乗和が自由度nの $\chi^2$ 分布に従い、不偏分散は平均との差を使うので、自由度が1つ減るのである。

(4) 有意水準

5%で両側検定せよと言った問題は非常に多く見受けられるため、棄却するとか受容するとか採択するという用語と共に、有意水準を教える必要がある。有意水準は相手の主張が真実の場合に誤って出してしまう確率であり、危険率とも言う。正規分布の確率密度関数  $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$  の形は、高校で習っているため、棄却域の面積はすぐに理解できると思われる。

帰無仮説が正しいのに棄却してしまう第1種の誤り(危険率と同じ)、対立仮説が正しいのに帰無仮説を採択してしまう第2種の誤りを教える時に、Pythonプログラムで教えると、出力の行列の並びが、過誤を教える時の表の並びと違うので、誤解しないようにしなければならない。具体的には、過誤を教える時の表は、表1の形を使うが、Pythonプログラムで `confusion_matrix` 関数の出力で

表1 混同行列の教える時の形

	正例 (予測)	負例 (予測)
正例 (実績)	True Positive	False Negative
負例 (実績)	False Positive	True Negative

表2 プログラムの出力される形

	負例 (予測)	正例 (予測)
負例 (実績)	48	5
正例 (実績)	8	82

は、  
`m <- confusion_matix(y_test, y_pred)`  
`print(format(m))`

のプログラムによって  
 [[48 5]

[8 82]]

と出力されたとすると、表2の順番になるため、適合率  $\frac{TP}{TP+FP}$  や再現率(真陽性率)  $\frac{TP}{TP+FN}$ 、偽陽性率  $\frac{FP}{FP+TN}$  などの計算を間違えることが起きる。

医療現場ではROC曲線がよく出てくるので、真陽性率と偽陽性率を教えた時にはROC曲線も教えて、検出漏れは防ぎたいという目指す形状のROC曲線を教えておきたいところである。

(5) 母平均の差の検定

2つのグループに差があるかどうか、2つの標本から検定することもよく行われる。2つの標本のデータが、対のデータになっている場合とそうでない場合に分かれ、母分散が既知の場合と未知の場合に分かれ、未知でも母分散の値が等しいという場合と等しいか不明という場合に分かれる。

母分散が既知の場合は、 $T = \frac{\bar{X}_A - \bar{X}_B - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}}$  の

式を用い、標準正規分布を適用するが、母分散が未知で等分散の場合、不偏分散  $U_A^2$ 、 $U_B^2$  を

使って、 $T = \frac{\bar{X}_A - \bar{X}_B - (\mu_A - \mu_B)}{\sqrt{(\frac{1}{n_A} + \frac{1}{n_B}) \frac{(n_A-1)U_A^2 + (n_B-1)U_B^2}{n_A + n_B - 2}}}$  の式が、自

由度  $n_A + n_B - 2$  のt分布に従うことを適用する。母分散が未知で等分散でない場合、

$T = \frac{\bar{X}_A - \bar{X}_B - (\mu_A - \mu_B)}{\sqrt{\frac{U_A^2}{n_A} + \frac{U_B^2}{n_B}}}$  の式が、自由度 f の t 分布に

従うことを適用する。この時 f は、

$\frac{(\frac{U_A^2}{n_A} + \frac{U_B^2}{n_B})^2}{\frac{1}{n_A-1}(\frac{U_A^2}{n_A})^2 + \frac{1}{n_B-1}(\frac{U_B^2}{n_B})^2}$  に一番近い整数である。

#### (6) 正規分布以外の分布

正規分布は高校で学習済みである。母集団の分散が分かっていない時に、母平均を推定する時に使う  $t$  分布は教える必要がある。また、不偏分散の分布は、 $\chi^2$  分布になり、分散を検定する時に用いられる。2つの分散の比は  $F$  分布になり、等分散検定をする時に用いられる。

一方、待ち行列問題など社会現象を取り上げると使われるのが、ポアソン分布である。ポアソン分布は、交通事故の発生件数や機械の故障回数など、稀な現象が一定の割合で、単位時間にランダムに発生する時の回数を示す分布である。発生する時間間隔は、指数分布になる。何に用いられるのかの用途を示しながら教えるのが重要である。

#### (7) Python プログラミング

**Shift** キーを押しながら **Enter** キーを押すとコードが実行されるという操作に慣れる必要がある。

また、**if** 文を書く時に、字下げが無くなった行から、**if** 文とは別の処理と見做されるため、字下げが文法的に意味ある言語になっている。**for** 文の繰り返しの終わりも字下げの終わりによって示され、字下げは重要な記述法である。

珍しい関数としては、**zip** 関数があり、異なるリストから同時に要素を取り出していきたい時に **for** 文に関連して使われる。例えば、`for one,two in zip([1,2,3], [11,12,13]):`

```
print(one, 'と', two)
```

また、無名関数という関数に名前を付けなくても定義できる **lambda** 記法がある。関数を引数や戻り値として使う関数である **map** 関数もあり、例えば、

```
list(map(lambda x:x*2, [1, 2, 3, 4]))
```

によって `[2,4,6,8]` のリストを作り出すことができる。

**filter** 関数も珍しい関数である。例えば、`list(filter(lambda x:x>=60,[29,71,57,92]))` によって、`[71,92]` のリストを作り出すことができる。

**Python** は、統計解析用のライブラリが多数用意されているため、統計解析の演習をさせるのに便利である。ただし、**Python** のバージョンを揃えないと、**seed** 関数の引数を同じにしても乱数発生の結果が異なって、演習結果に違いが出ることもある。

#### (8) R プログラミング

**RStudio** の使い方を教える必要がある。**Console** と書かれた部分に式を打ち込んで、**Enter** キーを押すと実行される。または、プログラムを書いた拡張子 **.R** のファイルを読み込んで、左上の **Source** ペインと呼ばれる区画の中のカーソル位置を先頭行に移動してから **Control** キーを押しながら **Enter** キーを押してカーソル行の位置のコードを実行させる。カーソルの位置を下へ移動させながら **Control** キー+**Enter** キーを押すことを繰り返していくのが面倒な場合は、**Control** キー+**Alt** キー+**R** キーを押すと、先頭行から最後の行まで実行してくれる。

## 5. オンライン授業におけるテスト実施方法

オンライン授業では、テストの実施方法に苦勞をする。東京大学では、学部配属に影響するため、成績付けが重要であり、テストも監視カメラのように **Web** カメラかスマフォカメラを設置し、解答作成中の状態を撮影しながらテストを受けるという体制が取られた授業もある。三重大教育学部ではレポート提出に切り替えた授業が多く、テストを実施した授業は少なかったようである。オンラインでテストを行う時に学生間でのカンニングが行われないかが問題であり、学生ごとに試験問題を変える手法を用意にできないかと考えた。

(1)学籍番号を使って出題を変え、手書きさせる方法

1つの授業（具体的には、ネットワーク基礎）では、学籍番号を問題の中に埋め込み、解答が同一にならないように出題した。10問中最後の2問は学生間で同じ解答になったが、8問は正解が異なり、試験時間60分の中で解答に掛かる時間が切迫しているため、学生間のカンニングは起きていないと思われる。解答は手書きで行い、試験終了後に解答用紙をフマフオのカメラで撮影しメールで送付する手法を採用した。オンラインのカメラは、試験終了後の不審な動きをしないかを見るために使用した。出題は、MoodleへWordファイルをアップする方法で行い、アップと同時に、Zoomの共有画面で共有提示した。

(2)学籍番号末尾を使って出題を少し変え、パソコンで打たせる方法

プログラミングの授業のテストでは、2問出題し、クラス名や変数名に学籍番号を付けさせ、学籍番号の末尾1桁によって、学籍番号を付けさせるものを変えて、正解が同じにならないようにした。解答はパソコンに打ち込み、解答ファイルをMoodle内の提出場所へ提出させる方法を行った。提出の締切時間を設定することで、試験終了後の不正に備えた。オンラインのカメラは、試験中の不審な動きをしないかを見るために使用した。プログラミングの問題は、正解に辿り付ける学生が少なく、学生間のカンニングは起きていないと思われる。

(3)出題順番を一人一人変えて、手書きさせる方法

1つの授業（具体的には、マルチメディア概論）では、量多く記述させる問題を7問出題し、その問題の順番を学生一人一人異なるように出題した。7つの出題文をメール本文に作成し、学生にメールで送る方式を採用した。用いたプログラムを図1に示す。

```
Public Sub SendGmail()
    Dim oApp As Object
    Worksheets("Sheet1").Select
    x = 1
    n = Sheets("Sheet1").Cells
(Rows.Count, 1).End(xlUp).Row
    MsgBox n
    Do While x <= n
        list = Array(1, 2, 3, 4, 5, 6, 7)
        list = Shuffle(list)
        i = 0
        body = ""
        Worksheets("Sheet2").Select
        Do While i <= 6
            body = body & "問" & (i + 1)
& ":" & Sheets("Sheet2").Cells(list(i), 1).
Value & vbCrLf
            i = i + 1
        Loop
        Set mITEM=App.CreateItem(0)
        With mITEM
            .To = Sheets("Sheet1").
Cells(x, 1).Value
            .cc = "yamamori@edu.mie-
u.ac.jp"
            .Subject = "試験問題"
            .body = body
            '.Attachments.Add "E:¥添付.docx"
            .Send
        End With
        x = x + 1
    Loop
    If MsgBox("OutLook を終了します
か?", 4) = 6 Then ' vbYes=6[はい]ならば
        oApp.Quit 'アウトルックを終了
    End If
End Sub

Private Function Shuffle(list As Variant)
As Variant
```

```

Dim i
For i = 0 To UBound(list)
  Randomize
  rn = Int(UBound(list) * Rnd)
  tmp = list(i)
  list(i) = list(rn)
  list(rn) = tmp
Next
Shuffle = list
End Function

```

図1 Gmail 発信する EXCEL マクロ

Gmail のメールサーバを使う場合、ユーザ認証の方法が難しいため、OutLook のメールソフトを利用してメール発信させる。OutLook では、Gmail のメールを読み書きできるように設定しておく。Sheet1 には、届け先のメールアドレスを Cells(x, 1) から縦に並べておく。Sheet2 には、試験問題文を 7 問、Cells(list(i), 1) から縦に並べておく。list(i) は list = Array(1, 2, 3, 4, 5, 6, 7) と list = Shuffle(list) によって、1 から 7 の数字がシャッフルされ並び順が変わって代入されるため、body = body & "問" & (i + 1) & "：" & Sheets("Sheet2").Cells(list(i), 1).Value & vbCrLf によって、メール本文に問 1 から問 7 の試験問題文が並べられることになる。"OutLook を終了しますか?" と尋ねるのは動作確認のために入れている。

'Attachments.Add "E:¥添付.docx" の先頭のコメント化文字( ) を外すと、添付ファイルが付いてメール発信されるが、添付ファイルがあると、学生に届くメールに約 10 分の時間遅れが発生し、添付ファイルの利用はお勧めでない。

解答は手書きとし、試験終了後に解答用紙をフマフォのカメラで撮影しメールで送付する手法を採用した。試験終了後には、学生間で試験問題の順番が違うことに気付いたようであったが、試験中には手書きで

書く時間に追われて、学生間のカンニングは起きていないと思われる。この方式は試験問題の難しさに公平性を保つことができる。

## 6. まとめ

WORD、EXCEL、PowerPoint の使い方などのリテラシーを教えてきている教員から見ると、データサイエンス教育の内容は、統計解析の内容であるため、教えられないと言われるのも無理はない。データサイエンスの内容が教えられる教員がほとんど居ない中で、求められる内容を高度化するのは無理である。その結果が、データサイエンスと称する授業の内容が、EXCEL のグラフ作成程度で終わってしまうことに成り兼ねない。EXCEL のグラフ作成は、私の受講生の半数の学生が高校で習得済みであった。もう大学で教えるレベルではないと思われる。大学ではより高度な内容を教えなければならない。2020 年からは高校「情報」のレベルも上がるので、高校で教えてもらってくる学生の方が、データサイエンスのことをよく知っているということが起きる。大学でのデータサイエンスと称する授業が、レベルの低い恥ずかしい内容にならないことを切に願う。

## 参考文献

- [1] 塚本邦尊、山田典一、大澤文孝著：東京大学のデータサイエンス育成講座、マイナビ出版
- [2] 滋賀大学データサイエンス学部編：大学生のためのデータサイエンス (I) オフィシャルスタディノート、一般財団法人日本統計協会発行
- [3] 柴田望洋著：新明解 Python 入門、SB クリエイティブ
- [4] 石井俊全著：大学の統計学、技術評論社
- [5] 倉田博史、星野崇宏著：入門統計解析、新



世社

[6] 涌井良幸、涌井貞美著、統計学の図鑑、技術評論社

[7] 松村優哉、湯谷啓明、紀ノ定保礼、前田和寛著、R ユーザのための RStudio 実践入門、技術評論社

[8] 高等学校情報科「情報Ⅱ」教員研修用教材（本編）、[https://www.mext.go.jp/a\\_menu/shotou/zyouhou/detail/mext\\_00742.html](https://www.mext.go.jp/a_menu/shotou/zyouhou/detail/mext_00742.html)（2020年9月現在）

[9] 東京大学教養学部統計学教室編：統計学入門、東京大学出版会

[10] 安井翔太著、効果検証入門、技術評論社

[11] 涌井良幸、涌井貞美著、統計解析がわかる、技術評論社