

統計学の講義資料を HTML で記述することの有効性

— R Markdown との連携 —

林 篤裕・名古屋工業大学
名古屋市昭和区御器所町
Tel/Fax: 052-735-5119
E-mail: hayashi.atsuhiko@nitech.ac.jp

1. はじめに

大人数を対象とした教室で、講義内容を学生に伝える方策として、かつては黒板への板書や OHP(Over Head Projector)といったアナログ的な手法が使われていたが、現在では電子資料を提供することが一般的になった。特に、2020 年初から感染が拡大した COVID-19 の脅威は教育の諸活動にも多大な影響を与え、教室に集っての開講を避ける必要からオンライン講義と呼ばれる運営形態が否応なく導入され、益々アナログ的な手法は採れなくなった。

その際に用いられたのが Moodle 等の学習管理システム(LMS, Learning Management System)で、これに電子資料や説明動画を掲載し、個々の学生にはインターネット環境下から参照させて学修を進める形態が普及した。元々、LSM は e ラーニングを実現するために開発されてきたプラットフォームであり、教員側からの情報提示だけでなく、学生からの質問・レポート等の提出や、個々の学生のアクセス記録に基づいて学修行動を把握する機能も備えており、オンライン講義との相性も良かったことから、これを機に導入した大学も多かったと思われる。

本発表では、統計学(やこれに類する科目)の講義における電子資料の作成に、HTML(Hyper Text Markup Language)が有効であることと、これを生成する機能を有している R Markdown を有機的に活用することを提案するものである。加えて、R Markdown が生成した HTML の一部を切り出して利用することができることも紹介する。

2. 統計学の講義資料に求められる構成要素

統計学の講義を念頭に考えた場合、講義内容の説明に必要な構成要素とも言うべき「部品」としては、文字(文章)、数式、画像、グラフ、表、他のサイトへのリンクぐらいが挙げられるであろう。特に文字については、文字サイズやフォント、色、下線等の修飾ができると効果的である。加えて、統計プログラムを学習内容に含んでいる講義の場合であれば、プログラムの掲載も必要になる。講義の目的にも依存するであろうが、プログラムを簡単にコピー&ペーストで取り込めるようになってくると講義時間中にプログラムの入力のための演習時間を

短くすることが出来るため、より高効率に講義を進行することが可能になる。

よって、講義資料を提示するツールにはこれらが容易に掲載・実現できることが求められる。表 1 には、講義資料を作成・提示するためのツール・環境に関する構成要素の掲載可能性と、幾つかの特徴をまとめた。

表 1. 作成ツールの比較表

| ツール・環境 | HTML / R Markdown | Power Point / MS-Word | LaTeX | テキストファイル |
|-------------|-------------------|-----------------------|--------|-------------|
| 文字 | ○(修飾可) | ○(修飾可) | ○(修飾可) | ○(修飾なし) |
| 数式 | ○ | ○ | ○ | △ |
| 画像 | ○ | ○ | ○ | × |
| グラフ | ○(Rで) | ○ | ○ | × |
| 表 | ○ | ○ | ○ | △ |
| プログラム | ○ | △ | △ | ○ |
| リンク | ○ | ○ | ○ | △ |
| 編集 | エディタ | 専用ソフト | エディタ | エディタ |
| 提示 | Webブラウザ PDF | 専用ソフト PDF | PDF | エディタ PDF |
| ページの概念 | ない | ある | ある | ない |
| Moodle への掲載 | HTML/PDF | PDF | PDF | テキスト/PDF |
| 費用 | 無料 | 有料 | 無料 | 無料/有料 |

3. 講義資料の作成ツールとしての HTML

世間には多くの LMS が発表・運用されており、中でも多くの大学で採用されている Moodle は、オープンソースであり、動画・音声等のマルチメディアや、PDF ファイル、Power Point、MS-Word 等だけでなく、HTML で記述したの電子ファイルも掲載可能である。

他の教員がどのようなツールを用いて電子的な講義資料を作成しているかを網羅的に把握しているわけではないが、COVID-19 対応としてリモート講義を強いられた際、本学では Moodle 上に講義資料を掲載することが推奨され、Power Point で作成した資料をアップロードすることが例示されていた。その理由は Power Point にはスライド毎に音声を録音する機能があるため、画面に説明資料を提示し、それに呼応したナレーションを流すことによって講義の進行を模擬的に実現できたからであろう。事実、この方法でリモート講義を運営していた同僚も少なからずいた。

著者も当初この推奨された方法を試してはみたものの、

Power Point で提示情報(文字や図表)を作成後、スライド 1 枚ごとに対応する説明を音声で録音していく作業は非常に煩雑で、かつページが変わる度に説明を中断する必要もあったので、個人的には非常にストレスフルであった。加えて、Power Point のスライドにはページ概念があるので、講義の進度や理解度に合わせて、より詳しい説明を追記しようとするとう当該のページから溢れてしまい、文字を小さくしてページ内に収めたり、次ページとの調整が必要であったりと講義内容自身ではなく、体裁の整形に費やす時間が想像以上に多く必要になった。Power Point が多機能であることは事実であるが、比較的動作の重いソフトウェアであることもネックとなった。

他にも LaTeX で講義資料を作成されている方を存じ上げているし、MS-Word で作成することも考えなくは無かったが、どうも納得が行くものではなかった。

著者はこれまでも講義資料を HTML で記述して Web サーバーに掲載してきた。学生にはこれを印刷したものを配布すると共に、サーバーの URL も共有していたので、欠席した場合等には各自で印刷することも可能であった。また、統計分析用言語である SAS 言語や R 言語のソースコードも掲載していたので、学生はいちいち印刷資料の文字(コマンド等)を手入力する必要はなく、Web サーバー上の資料からコピー&ペーストで統計分析用言語のソースコードを入手することができるというメリットにも重宝していた。

HTML は World Wide Web(いわゆるインターネット)の普及と呼応するように整備・拡張されてきた計算機言語で、その名称からも推察できる通り、構造化された文書表現するためのものであり、Web ページの記述言語として広く利用されている。Google Chrome 等の Web ブラウザは、HTML で記述された電子情報を表示するプログラムとして重宝されているが、視点を変えると、サーバーから HTML で記述された電子情報をダウンロードしてきて、それを規定されたルールに則って、画面に表示するツールとも言える。HTML にはページ概念がなく、希望する場所に新たな記述を追記すれば、押し出された情報は下側に移動するだけである。

また、HTML は数式を表示する機能も有しており、しかもその表記法は LaTeX のそれに準じるものなので、LaTeX ユーザーであれば容易に記述することができると思われる。

計算機言語と言っても、Web ブラウザ側の解釈に冗長なところがあるので、文法上の厳密なルールに規定された記述がなされていなくても目的の表示ができるのも、

表 2. 代表的な HTML のタグ

| |
|--|
| ul, ol, li, h, b, p, br, font, pre, code, hr, href, img, a |
|--|

個人的には好都合であった。HTML には多くのタグが規定されているが、その中でも表 2 に示したタグを駆使すれば構造化された文章の多くを記述することが可能であると考えている。加えて、表 1 に示したように、HTML は統計学の講義資料に求められる構成要素を記述することに何の障害もない。

副次的なメリットとしては、上述のように HTML で記述された講義資料は、Web ブラウザに転送されて表示されるので、興味を持った表示の実現方法を知りたいければ、そのソースコードを見ることによって把握することができる。いわばオープンソースであり、同様の表示を必要とした際の参考になる。

当初取っ付き難いと感じた Moodle であったが、使い込むに従って、強力な支援ツールであることも判った。

4. HTML も生成可能な R Markdown

今更言うまでもないであろうが、統計解析用プログラミング言語である R 言語には、多くの支援ツールが開発・提供されており、その中の一つに RStudio という統合環境がある。名称からは想像し難いが、RStudio は R 言語に特化されたものではなく、多くのプログラミング言語用の開発環境であるが、統計学関係ではもっぱら R 言語の統合環境として利用されていると思われる(ちなみに手元の RStudio では 52 の言語をサポートしている)。R 言語を利用して統計分析を進める環境として優れている RStudio だが、それ以外に R 言語の計算結果を取り込んで報告書を作成するレポート機能も有しており、この機能を利用するには R Markdown と呼ばれるマークアップ言語で記述する必要がある。RStudio の左上ペインに、データを読み込んだり分析したりするコマンド群を R Markdown の文法に則ったソースコード(.rmd ファイル)として記述し、コンパイル(knit と呼ばれる)することによって、計算結果を MS-Word や PDF(LaTeX 経由で)、そして HTML のフォーマットで出力することができる。

R Markdown の文法は HTML のそれと非常に類似しており、(これは既に広く知られていることなのかもしれないが)、R Markdown で定められていない(と思われる)タグもそのまま HTML に吐き出してくれるので、個人的には「R 言語用に拡張された HTML」といった捉え方をしている。希望に副わない表示をしてしまう HTML が生成されてしまった場合にも、直接 HTML に手を加えることで修正することが可能である。

少し追加で習得する必要があるのは、chunk(チャンク)と呼ばれる R 言語への指定であり、プログラム自身の評価や、その計算結果を表示させるか等を指定する命令群である。しかし、難しいものではないので、一度理解してしまえば以後は使い回すことができる。

加えて、R Markdown のソースコードも HTML のコードも何れもテキストファイルなので、編集には手慣れたテキストエディタを利用すれば良く、動作が軽い

も魅力である。

統計学の講義では、身近な生きたデータを用いるのが学生の理解を促進すると考えているが、毎年更新されるデータに基づいた講義を行っているような場合、新しい計算結果をこれまで書き溜めた講義資料に取り込もうとすると、R Markdown で記述しておけばデータファイルの名前を変更するだけで最新のデータに基づいた講義資料を生成することができるというメリットもある。

5. グラフの有効活用

前節で述べたように R Markdown を用いれば、R 言語で記述した分析命令とその計算結果をまとめた Web ページを HTML で生成することができる。R Markdown が内部でどのような処理を行っているかを正確には把握出来ていないのだが、R 言語の実行結果である統計量等の数値情報だけでなくヒストグラムや箱ひげ図、散布図等の図形情報も含まれる。これらがどのように HTML で実現されているかに興味を持ったので、生成された HTML を覗いてみたところ、図形情報については png という画像フォーマット (Portable Network Graphics) で作成されており、それが img という HTML のタグに引用される形の記述となっていることが判った。つまり、生成された Web ページに含まれるグラフ群の中から、必要と考えるグラフだけを容易に抜き出すことができることを意味しており、事実、当該のグラフ部分だけを他の HTML に挿入してみると期待通りの表示となった。このテクニックはグラフ以外にも有効なので、R 言語の出力を独立して利用したいような場合、重宝すると思われる。

なおデフォルトの画像フォーマットは png であったが、陽に指定すればそれ以外のフォーマットでも作成することができるとなっている。

参考までに、著者が回帰分析の講義で用いた散布図部分(図 1)と数式部分(図 2)の、R Markdown のソースコード(右上)と生成された HTML(右下)、および Web ブラウザの表示(左)を以下に示しておく。

6. まとめに代えて

Power Point は、プレゼンテーションツールとしてデファクトスタンダードの地位を確立したと言っても過言ではないであろう。多機能であり洗練された提示が行えることから学会発表等で多用されていることも納得できる。しかし、それを講義資料の作成・提示ツールとして用いるには、やや重厚過ぎるように感じると共に、講義内容の洗練に注力すべき労力が奪われるように危惧している。

講義資料をどのような方法で準備するかは、それぞれの教員が判断することであることは改めて言うまでもないことではあるが、こと「統計学の講義資料」という限定的な目的であれば、HTML もその選択肢に加えていただけないのではないかと、今回紹介することにした次第

である。しかも、R Markdown がその作成を強力にバックアップしてくれるのが嬉しい点である。

利用するファイルはテキストファイルであるので、軽快に編集ができ、しかも何れのツールも無料で利用できることもメリットではないだろうか。

ここに至るまでの過程では、講義資料の作成に HTML を用いている人がおられるのではないかと検索してみたが、的確なものはヒットしなかった。その意味では著者の気付いていない大きな欠点があるのかもしれない。一方で、著者の調査不足から、もう既に広く利用されていたり、より高度な実装方法が実現されていたりするのかもしれない。そのような際には是非ご教授いただければ幸いである。

個人的な動機としては、リモート講義を強いられたことにより講義の運営が難しくなりかけたところに、R Markdown とその HTML 生成機能を知ること、大いに救われた。「必要は発明の母」の諺ではないが、窮地を救ってくれたのは RStudio であったことから、語弊を覚悟して言えば COVID-19 には感謝しないといけないのかもしれない(著者も 1 回罹患したが)。

本稿が統計学やデータサイエンスを題材とした講義の運営の参考になれば幸いである。

参考文献

- 1) 24 R Markdown [基礎], <https://www.jaysong.net/RBook/rmarkdown.html>.
- 2) MathJax, <https://www.mathjax.org/>.
- 3) R Markdown, <https://rmarkdown.rstudio.com/>.
- 4) R Markdown 入門, https://kazutan.github.io/kazutanR/Rmd_intro.html.
- 5) Yihui Xie, J. J. Allaire, Garrett Golemund, R Markdown: The Definitive Guide, <https://bookdown.org/yihui/rmarkdown/>.
- 6) Yihui Xie, Christophe Dervieux, Emily Riederer, R Markdown Cookbook, <https://bookdown.org/yihui/rmarkdown-cookbook/>, [和訳版] R Markdown クックブック, <https://gedevan-aleksizde.github.io/rmarkdown-cookbook/>.

(URL は何れも 2024 年 2 月 28 日現在)

The figure shows a web browser on the left displaying a scatter plot titled "Scatter Plot of Shintyou and Taijyuu". The x-axis is labeled "Height" and ranges from 150 to 180. The y-axis is labeled "Weight" and ranges from 40 to 100. A regression line is drawn through the data points. Below the plot, there are two bullet points in Japanese: "「誤差が一番小さい」直線が良さそうだ。" and "では、その「誤差」って何? どう定義する?".

On the right, there are two windows showing R Markdown source code. The top window shows the code for creating the scatter plot with a regression line using R's `plot()` and `abline()` functions. The bottom window shows the text content of the R Markdown document, including the question "なぜ僕らは分布の中央付近を通過する直線が相応しいと感じるのか?" and the answer "それはどうやって決めたのか?".

図 1. 回帰直線を説明する部分の R Markdown ソース(右上)、生成された HTML(右下)と Web ブラウザの表示(左)

The figure shows a web browser on the left displaying a page titled "7.1. 定式化: 式の展開、解法". The page contains text in Japanese: "測定の場所によらず、残差は同一の分布であることを仮定する" and "残差の二乗和を最小にする(最小二乗法)". Below this, it says "回帰平面の方程式:" followed by the equation
$$y = a_0 + \sum_{j=1}^p a_j x_j^i$$
. Then it says "測定値と予測値のスレ(残差):" followed by the equation
$$\epsilon_i = y_i - (a_0 + \sum_{j=1}^p a_j x_{ij}) \quad (i = 1, 2, \dots, n)$$
. Next, it says "残差の2乗和を最小に:" followed by the equation
$$S = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n \{y_i - (a_0 + \sum_{j=1}^p a_j x_{ij})\}^2 \rightarrow \min$$
. Finally, it says "行列表記:" followed by the equation
$$Y = XA$$
. Below this, it says "ここで," followed by the matrix equation
$$Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2p} \\ 1 & x_{31} & x_{32} & \cdots & x_{3j} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{pmatrix}$$
.

On the right, there are two windows showing R Markdown source code. The top window shows the code for the text and equations, using `math display="..."` for the formulas. The bottom window shows the HTML output of the R Markdown document, including the section title "7.1. 定式化: 式の展開、解法" and the rendered equations.

図 2. 回帰式を説明する部分の R Markdown ソース(右上)、生成された HTML(右下)と Web ブラウザの表示(左)