

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)

(hayashi.atsumi@nitech.ac.jp (mailto:hayashi.atsumi@nitech.ac.jp))

【お知らせ】本発表の資料は以下URLで公開しています。後ほど再掲します。
<http://stat.web.nitech.ac.jp/haifu/#StatEdu2403>

(<http://stat.web.nitech.ac.jp/haifu/#StatEdu2403>)



1. はじめに

- 講義での情報伝達:
 - 時代とともに: 板書、OHP(Over Head Projector) ==> 電子資料の配布
 - オンライン講義
 - COVID-19の影響も: 集うことを避ける必要
- 学習管理システム(LMS, Learning Management System)への掲載
 - eラーニングのプラットフォーム
 - Moodle, Blackboard, Google Classroom等
 - 教員と学生間の情報やりとり: 講義資料、レポート提出、質問等
 - アクセスログの解析: 統計学的にも面白そうであるが
- 統計学の講義における電子資料の作成

【疑問】 効率的な作成ツールはないものか？

【提案】 HTMLが有効ではないか: Hyper Text Markup
Language

- R Markdownと連携しながら

- [私的気付き] グラフを抜き出して利用することも可能
- 周知の情報ばかりで既に多くの実践者がおられるのかもしれない。

【参考】 この資料もR Markdownで作成してみました。


2. 統計学の講義資料に求められる構成要素

- 統計学の説明に用いられる「部品」
 - 文字(文章)、数式、画像、グラフ、表、他のサイトへのリンク
- 修飾: **文字サイズ** やフォント、色、上線 等
- 統計解析用プログラム: 演習として計算機言語を取り扱う場合
 - コピー&ペーストができれば入力作業を短縮できる
(講義の開講目的にも依る)
- 「横書き」であること
- 統計学の電子資料:
 - これら「部品」が表現できるかが作成ツール選択の条件
- 【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
提示内容	文字	○(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	専用ソフト		エディタ
	PDF	PDF	PDF	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

- Power Point
 - 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
 - 洗練された発表、アニメーション、多くのユーザー、...
 - 音声も保存可。オンライン講義で重宝。 <=== 本学のガイドブック
 - 【私見】使いづらさ: 講義資料の作成・提示ツールとして
 - 高機能ゆえ? 専用ソフト: 重い
 - 「ページ」の概念: 整形作業に時間を割かれる
<===> 内容の充実
 - 体裁調整: ページから溢れたら。次ページとの関係。...
- LaTeXやMS-Wordも考えたが...: どうもシックリ来ない
- Moodle: 本学も採用。COVID-19対応として利用を求められた。

- オープンソースのLMS
- 掲載可能情報: マルチメディア(動画・音声等)、PDFファイル、
Power Point、MS-Word等、HTML
- 「手書き資料をスキャンしてPDFで掲載」という同僚も
- 『30年近いHTMLの作成経験と蓄積が活かそうだ!!』 
 - これまでHTMLで講義資料を作成してきた。
 - 講義資料を印刷して配布するとともに、
Webサーバーに公開して学生にURLを通知。<=== 各自で印刷も可
- Webブラウザ: Google Chrome, FireFox, Edge, Safari,...
 - Webページの「閲覧ソフト」 <=== 多くの方の認識
 - 別の見方: 「HTMLの表示器」
 - サーバーからダウンロードしてきてHTMLを解釈・整形表示
 - ダウンロードされたHTMLは閲覧可能 <=== 参考になる
- スクロール方向の転換(意識改革): 統計学＝横書き
 - 板書やPower Point: 横長の『シングルページ』
 - Webページ: 縦長の『巻物』、長さはほぼ無限
 - 「ページ」の概念がない。希望する場所に加筆・修正できる。
- 【表2】 代表的なHTMLのタグ:
ほとんどの構造化された説明資料の作成が可能。

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

4. HTMLも生成可能なR Markdown

- R言語の支援ツール: 多数発表されている
 - R Commander, R AnalyticFlow, 各種パッケージ, ...
 - RStudio、およびその中の R Markdown に注目
- RStudio:
 - R言語の統合開発環境。奥が深そう。

- R Markdown:
 - R言語を用いた 分析から計算結果のレポート生成まで。
 - R言語のプログラムのみならず、統計量(数値)やグラフまで
 - 分析の再現性記録にも
 - [蛇足] R言語以外にも延べ52言語に対応。名称から想像できないが。
 - R Markdownも(HTMLも)テキストファイル: エディタで編集=軽い
 - 出力: MS-Word、PDF(LaTeX経由)、そして HTML



引用: 「R Markdown入門 (https://kazutan.github.io/kazutanR/Rmd_intro.html)」

- R Markdownの記法:
 - Markdown由来のものもあるが、HTMLのタグもほぼ利用できる
 - Markdown: 文書を記述するための軽量マークアップ言語
 - 「R言語用に拡張されたHTML」とも言えるのではない【私見】
 - R Markdownで解釈できないタグはそのまま出力されている?
 - HTMLのタグを用いてドンドン書き込んで試してみる
 - たまにうまく表現されないこともある
[know-how] p タグを入れると解決することも
 - 数式はLaTeX記法で記述する (HTMLの機能かも)
 - 生成されたHTMLに手を加えることも可能【奥の手】
- chunk(チャンク): R言語への指定 <=== 新たに覚える必要
 - プログラム自身の評価を行うか
 - 計算結果を表示させるか等
 - 難しいものではない。指定の On/Off を指示。

◎ 統計学の学修: 身近な生きた最新データで例示するのが有効
【私見】

- 毎年更新されるデータ: ファイル名を変更するだけで最新のデータに対する講義資料を作成できる <=== 作業の省力化

5. グラフの有効活用

- R言語の実行結果: 数値情報 と 図形情報
- どのように HTML で実現されているか? 生成されたHTMLを覗いてみる。
- 図形情報
 - 画像フォーマット: png(Portable Network Graphics) (変更も可)
 - img タグで貼り付けられている
 - 個々に取り出しが可能 ==> 他のHTMLに移植
- R言語とR Markdownを「作図ツール」と捉えることもできる

The image displays a web browser window on the left showing 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 plotted through the data points. Below the plot, there are two bullet points:

- 「誤差が一番小さい」直線が良さそうだ。
- では、その「誤差」って何? どう定義する?
- また、「一番小さい」って何? どう定義する?
- ユニーク(唯一、一意)に決めるには何らかの定義が要るよね?

On the right, there are two overlapping windows showing R Markdown source code and its rendered HTML output. The top window shows R code for fitting a linear model and plotting it:

```

114 <LI>回帰分析とは?↓
115 <p>↓
116 <LI>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?↓
117 <LI>それはどうやって決めたのか?↓
118 </UL>↓
119 ↓↓
120 [ ,echo=FALSE]↓
121 RresultTmp.<-lm(Weight~Height,.data=StudTmp)↓
122 ##.RresultTmp↓
123 ##.summary(RresultTmp)↓
124 ↓
125 plot(Weight~Height,.data=StudTmp,↓
126       xlab="Height",ylab="Weight",↓
127       main="Scatter_Plot_of_Shintyou_and_Taijyuu")↓
128 abline(RresultTmp)↓
129 abline(h=seq(40,100,10),lty=3)↓
130 abline(v=seq(140,190,5),lty=3)↓
131

```

The bottom window shows the rendered HTML output, including the same text as the source code and a base64-encoded image of the scatter plot:

```

8599 <LI>↓
8600 回帰分析とは?↓
8601 <p>↓
8602 <LI>↓
8603 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?↓
8604 <LI>↓
8605 それはどうやって決めたのか?↓
8606 </UL>↓
8607 <p><img src="data:image/png;base64,iVBORwOKGgoAAAANSUHEUgAABUAAAAAPACAM#
8608 AAA21BMVEUJAAAAADoAAGYA0joA0mYADpAAZrY6AA60gA60jo60mY6ZmY6ZpA6ZrY6kJA6
8609 ABm0gBm0jpmZjpmZmZmZpBmkJBmkLZmkNtmt rZmtttmtv+Q0gCQZgCQZjgQZmaGkGaGkJC
8610 raQt tu0tv+0Z9u0Z/+ZgCZjg2kDg2kGa2kJC2kLa2tpC2tra2ttu2Z9u2Z/+2///bkDrt
8611 pDbt rbb25Db27bb29vb2//b/7bb/9vb///tmb/25D/27b/29v//7b//9v///8B3sFYAAA#
8612 B2HAAAdhwGP5fFIAAgAEIEQVR4n029C5vcNoZgySzL Tsm12pbtymr1Wh65t8fatVzacVou
8613 cn/4s2GBEKL4ALEADgZ5zz2QoGA7gX4OME+AhmYMAQBbY0g0AANgCBQAIBMECgCQCQIF
8614 IJAAQyQaAAAJkgUACATBAoAEAmCBQAIBMECgCQCQIFAMgEgQIAZIJAQAyQaAAAJkgUAC#
8615 BQAIBMECgCQCQIFAMgEgQIAZIJAQAyQaAAAJkgUACATBAoAEAmCBQAIBMECgCQCQIFAMgE
8616 QAYQaAAAJkgUACATBAoAEAmCBQAIBMECgCQCQIFAMgEgQIAZIJAQAyQaAAAJkgUACATBAc
8617 BMECgCQCQIFAMgEgQIAZIJAQAyQaAAAJkgUACATBAoAEAmCBQAIBMECgCQCQIFAMgEgQIA#

```

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 three overlapping windows illustrating the workflow from R Markdown to a web browser. The top-left window shows the R Markdown source code for a section on regression, including text, mathematical formulas for the regression equation and error terms, and the sum of squared residuals. The top-right window shows the corresponding HTML code generated from the R Markdown, using LaTeX-style syntax for mathematical expressions and HTML tags for structure. The bottom-left window shows the final rendered HTML in a web browser, displaying the text and mathematical formulas as they appear to a user.

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)

6. まとめに代えて

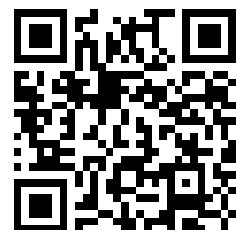
- 統計学の講義における電子資料の効率的な作成
 - HTML も選択肢になり得る
 - R Markdownとの連携 <=== 強力な支援ツール
- 特徴
 - 統計学の説明に使う構成要素は表現可能
 - 縦スクロールの巻物 (ページの概念なし)
 - ハンドリング: テキストファイルの軽快さ
 - Webブラウザで表示・閲覧可
 - 無料
 - (HTMLが登録できるLMSであれば活用可能)
- HTMLの可能性(再認識): リモート会議での提示にも有効?
 - 【例】DNC CBT報告書解説
(http://stat.web.nitech.ac.jp/haifu/DNC_CBT2208/index.html)

- 【掲載記事】「大学入試におけるC B Tが秘める可能性と今後の課題 (https://view-next.benesse.jp/view_section/bkn-hs/article12780/)」, 『VIEW next』 高校版 2022年度 10月号 (https://view-next.benesse.jp/view/bkn-hs/article12546/), PP24-26。
- 「必要は発明の母」というか、「火事場の馬鹿力」というか
 - 個人的な動機: COVID-19が猛威を振るっていた2020年
 - [序] リモート講義 & Moodleの使用を強いられた
 - [破] 講義資料の作成だけで疲弊 (講義運営が困難に陥りかけた)
 - [急] R MarkdownとそのHTML生成機能を知り、大いに助かった
 - 窮地を救ってくれたのは RStudio であった
 - 語弊を覚悟して言えば「COVID-19 には感謝」かも
- 『HTMLのすゝめ』、『R Markdownのすゝめ』
 - 統計学の講義資料の作成ツールとして
 - 今回紹介した技(?)に言及した情報に出逢えなかったので発表してみました。皆さんには既知の事項ばかりであればお恥ずかしい次第です。
 - 私が気付いていない大きな落とし穴があるのかも。
 - より良い方策をご存知であれば、ご教授いただければ幸いです。
もし、参考になる部分があれば、ご自由にお持ちください。

◎ 【お知らせ】 本発表の資料は以下URLで公開しています。

<http://stat.web.nitech.ac.jp/haifu/#StatEdu2403>

(<http://stat.web.nitech.ac.jp/haifu/#StatEdu2403>)



99. 参考文献

- 24 R Markdown [基礎], <https://www.jaysong.net/RBook/rmarkdown.html> (<https://www.jaysong.net/RBook/rmarkdown.html>).
- MathJax, <https://www.mathjax.org/> (<https://www.mathjax.org/>).
- R Markdown, <https://rmarkdown.rstudio.com/> (<https://rmarkdown.rstudio.com/>).
- R Markdown入門, https://kazutan.github.io/kazutanR/Rmd_intro.html (https://kazutan.github.io/kazutanR/Rmd_intro.html).
- Yihui Xie, J. J. Allaire, Garrett Grolemond, R Markdown: The Definitive Guide, <https://bookdown.org/yihui/rmarkdown/> (<https://bookdown.org/yihui/rmarkdown/>).
- Yihui Xie, Christophe Dervieux, Emily Riederer, R Markdown Cookbook, <https://bookdown.org/yihui/rmarkdown-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/> (<https://gedevan-aleksizde.github.io/rmarkdown-cookbook/>).

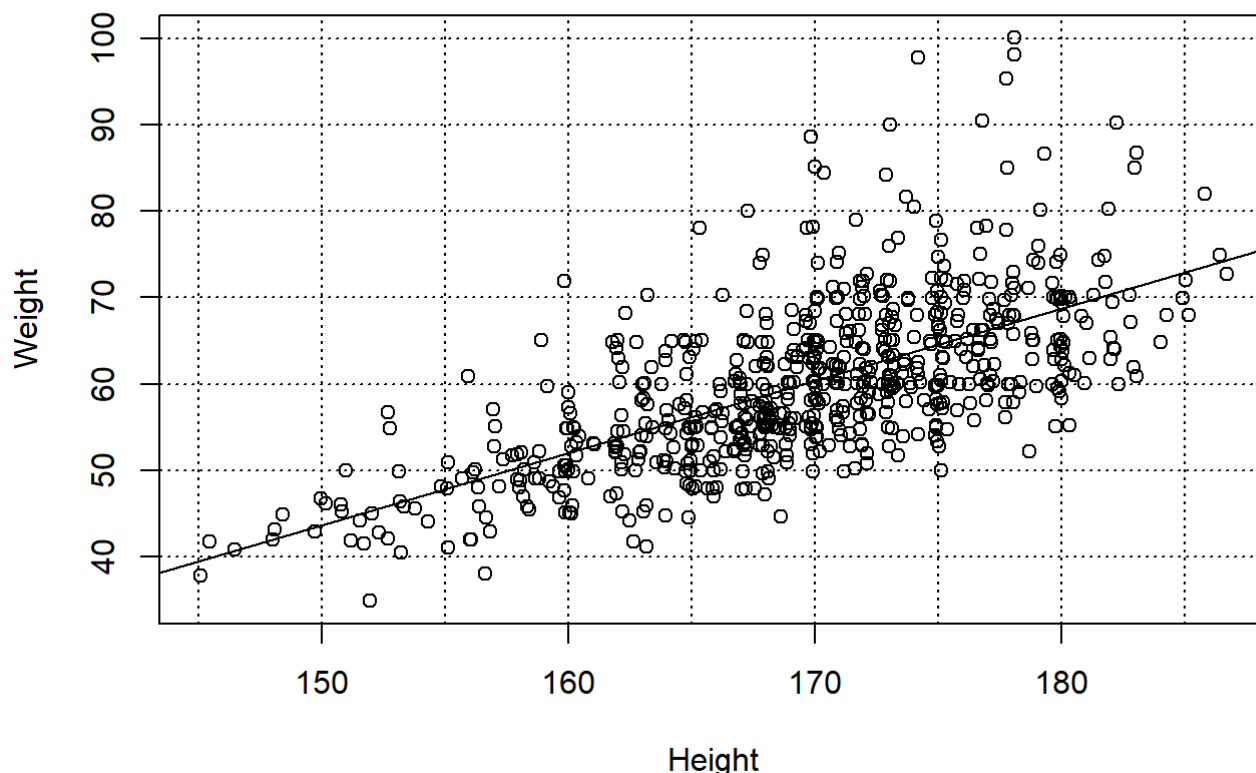
(URLは何れも2024年2月28日現在)

==== 【以下は講義資料中での使用例】

====

1.1. アイディア 【グラフ】

Scatter Plot of Shintyou and Taijyuu



7.1. 定式化: 式の展開、解法 【数式】

回帰平面の方程式:

$$y = a_0 + \sum_{j=1}^p a_j x_j^t$$

測定値と予測値のズレ(残差):

$$\varepsilon_i = y_i - (a_0 + \sum_{j=1}^p a_j x_{ij}) \quad (i = 1, 2, \dots, n)$$

残差の2乗和を最小に:

$$S = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \{y_i - (a_0 + \sum_{j=1}^p a_j x_{ij})\}^2 \rightarrow \min$$

行列表記:

$$\mathbf{Y} = \mathbf{XA}$$

ここで、

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ 1 & x_{21} & x_{23} & \cdots & x_{2j} & \cdots & x_{2p} \\ 1 & x_{31} & x_{32} & \cdots & x_{3j} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 1 & x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & \vdots & \vdots & & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_j \\ \vdots \\ a_p \end{pmatrix}.$$

\mathbf{A} を求めるには、左側から \mathbf{X}^t をかけて、

$$\mathbf{X}^t \mathbf{Y} = \mathbf{X}^t \mathbf{X} \mathbf{A}$$

その後、今度は左側から $(\mathbf{X}^t \mathbf{X})^{-1}$ をかけて、

$$\mathbf{A} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}$$

よって、 \mathbf{A} を求めるには、 $(\mathbf{X}^t \mathbf{X})$ が逆行列を持つことが条件となる。

==== 【参考: 本発表の予稿原稿】 =====

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 (mailto: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

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の概念があるので、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、構造化された文書を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にはページの概念がなく、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どところがあるので、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、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。

表2. 代表的な HTML のタグ

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、動作が軽いのも魅力である。

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るとは思えないかと、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によって、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。

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)

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)

参考文献

- 24 R Markdown [基礎], <https://www.jaysong.net/RBook/rmarkdown.html> (<https://www.jaysong.net/RBook/rmarkdown.html>).
- MathJax, <https://www.mathjax.org/> (<https://www.mathjax.org/>).
- R Markdown, <https://rmarkdown.rstudio.com/> (<https://rmarkdown.rstudio.com/>).

4. R Markdown入門, https://kazutan.github.io/kazutanR/Rmd_intro.html (https://kazutan.github.io/kazutanR/Rmd_intro.html).
5. Yihui Xie, J. J. Allaire, Garrett Golemud, R Markdown: The Definitive Guide, <https://bookdown.org/yihui/rmarkdown/> (<https://bookdown.org/yihui/rmarkdown/>).
6. Yihui Xie, Christophe Dervieux, Emily Riederer, R Markdown Cookbook, <https://bookdown.org/yihui/rmarkdown-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/> (<https://gedevan-aleksizde.github.io/rmarkdown-cookbook/>).

(URLは何れも2024年2月28日現在)

==== 【参考: 実際に講義で使用した資料の一部】 ====

1. 回帰分析

今回は、多変量解析の代表的な手法である回帰分析について解説する。

工学系や農学系の実験等を行う領域では頻繁に使用される手法であるが、日常的な話題の中でも概念は広く利用されているので、取っ付き易い手法ではないだろうか。近年は機械学習等でも取り扱われていると思う。過去のデータからその構造を把握し、新規に測定されたデータに対する予測を行ないたいと言うときに、回帰分析は有用である。構造のシンプルな単回帰分析でこの手法の原理を理解し、複数の説明変数を用いた重回帰分析に拡張する。残差の取り方や、その二乗和を最少にするという考えは同じである。

1.1. アイディア

散布図にもっともらしい「直線」を当てはめたい。

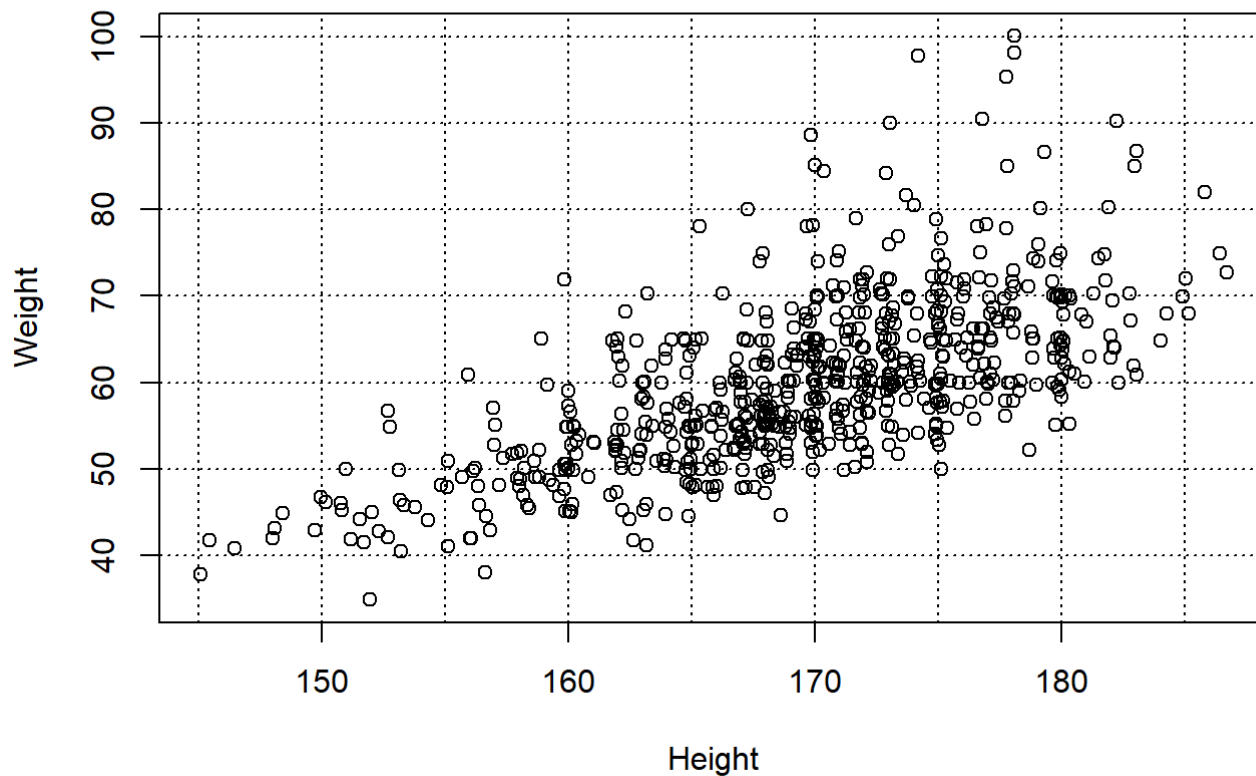
- 何に使う?
- 何に使える?
- 散布図: 縦軸=体重、横軸=身長

皆さんから収集した体格データを用いて、散布図を描いてみよう。どういう直線が「もっともらしい」と考えるか?

[注意1] 散布図を丁寧に観るとデータの稠密な部分と疎な部分があることを理解せよ。

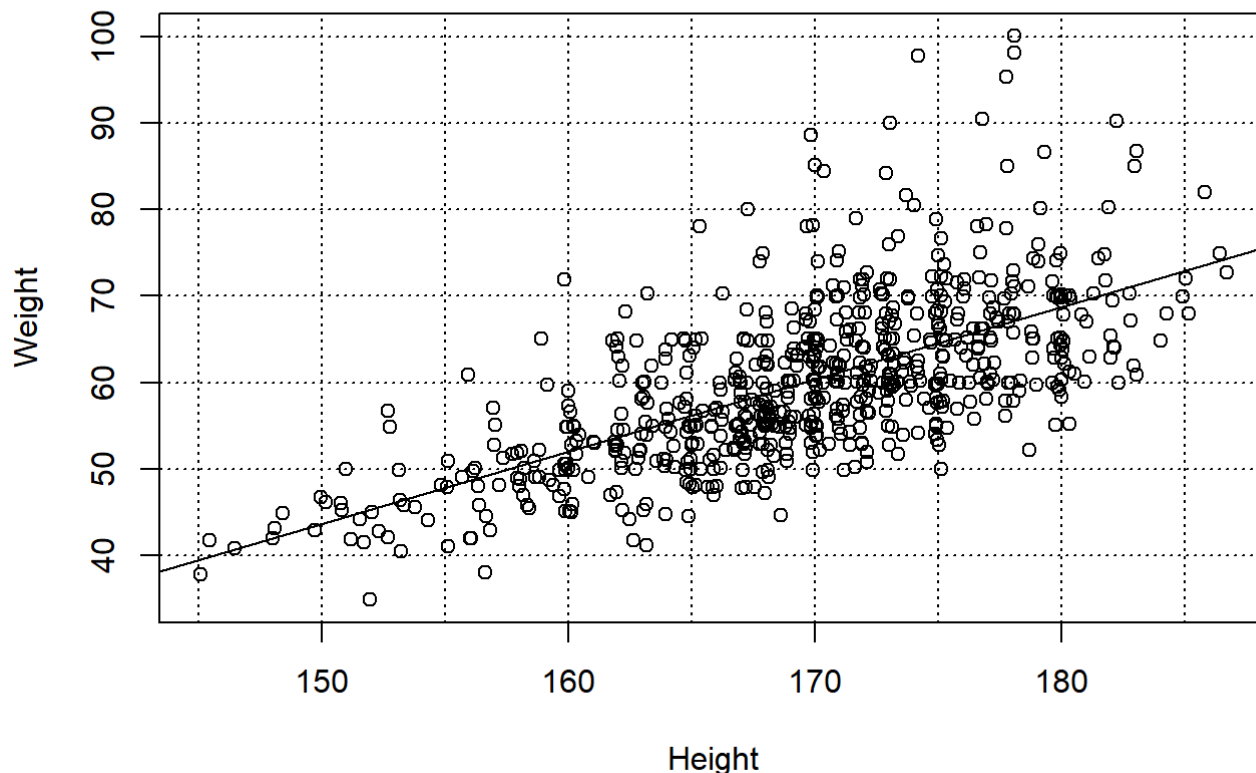
[注意2] ここに掲載している散布図は、視覚的に密度を理解し易くするために、第3回の第3節で紹介した、乱数を用いて描画点をブレさせて表示してある。

Scatter Plot of Shintyuu and Taijyuu



- どの様に直線の位置決めをすれば良い?
- 一意に決めるにはどのようなアイデアがある?
- そもそも、どういう考えから導き出された?
- 回帰分析とは?
- 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?
- それはどうやって決めたのか?

Scatter Plot of Shintyou and Taijyuu



- 「誤差が一番小さい」直線が良さそうだ。
- では、その「誤差」って何? どう定義する?
- また、「一番小さい」って何? どう定義する?
- ユニーク(唯一、一意)に決めるには何らかの定義が要るよね?

1.2. 前準備: 完全データの抽出

データに欠損値が含まれていると、後々の分析(残差プロット等)時に、毎回欠損値を除去する手順を組み込む必要が生じて面倒なので、分析対象の変数だけを先に切り出し、その上で、欠損値を除外(完全データ)しておく。

```
# データの切り出し
colnames(Student23)
```

```
## [1] "No"      "Sex"      "Height"    "Weight"    "Chest"
## [6] "Residence" "Remittance" "Carrier"    "Fee"
```

```
StudTmp1<-Student23[,3:4] # 身長(3番目)と体重(4番目)だけを切り出す
StudTmp1[1:15,]
```

```
##   Height Weight
## 1  145.0   38
## 2  145.5   42
## 3  146.7   41
## 4  148.0   42
## 5  148.0   43
## 6  148.9  NA
## 7  149.0   45
## 8  150.0   43
## 9  150.0   46
##10  150.0   47
##11  151.0   42
##12  151.0   45
##13  151.0   46
##14  151.0   50
##15  151.0  NA
```

```
dim(StudTmp1)
```

```
## [1] 772  2
```

元々の身長と体重のデータは 772例だったが、欠損値を含むデータが96例あり、残ったのは676例である。

```
StudHW<-na.omit(StudTmp1) # 欠損値を含むデータを削除
dim(StudHW)
```

```
## [1] 676  2
```

```
colnames(StudHW)
```

```
## [1] "Height" "Weight"
```

```
StudHW[1:15,]
```

```
##   Height Weight
## 1  145.0  38.0
## 2  145.5  42.0
## 3  146.7  41.0
## 4  148.0  42.0
## 5  148.0  43.0
## 7  149.0  45.0
## 8  150.0  43.0
## 9  150.0  46.0
##10  150.0  47.0
##11  151.0  42.0
##12  151.0  45.0
##13  151.0  46.0
##14  151.0  50.0
##16  151.7  41.5
##17  152.0  35.0
```

以後は、Student23に代えて、欠損値を除去して完全データにしたStudHWを使って分析を行う。

【蛇足: パズルの要領で】上記では、欠損値を含むデータの例数を、96例=772-676として求めたが、欠損値を含むデータの例数を直接的に数える式を作ってみると、例えば以下のようなになる(他にも実現方法はあると思う)。sum()関数は、本来は引数(カッコ内の式)の合計値を計算する関数だが、論理値 TRUE/FALSEの場合はTRUEを1、FALSEを0と捉えてその総数(=個数)を算出してくれる。2通りの方法で実現してみたが、それぞれがどのように機能しているか理解できるであろうか? なお、コメントアウト(行頭に#を付けてある)行は確認のために変数の内容を表示させる行である。表示させなければ、行頭の#を削除すれば良い。

[実現方法1]

```
# 欠損値を含むデータの例数カウント(方法1)
StudTmp2<-!is.na(StudTmp1)
# StudTmp2[1:15,]
StudTmp3<-StudTmp2[,1] & StudTmp2[,2]
# StudTmp3[1:15]
# (StudTmp3==F)[1:15]
sum(StudTmp3==F)
```

```
## [1] 96
```

[実現方法2]

```
# 欠損値を含むデータの例数カウント(方法2)
StudTmp2<-is.na(StudTmp1)
# StudTmp2[1:15,]
StudTmp3<-StudTmp2[,1] | StudTmp2[,2]
# StudTmp3[1:15]
sum(StudTmp3) ## これと同じ意味 ==> sum(StudTmp3==T)
```

```
## [1] 96
```

2. 単回帰分析：予測等に使う、連続変数の関係

- 体重を身長で説明(回帰)したい：[体重]=a+b[身長]：回帰係数
- 関係性は直線＝線形を仮定する
- 説明される変数：目的変数、従属変数、dependent variable
- 説明する変数：説明変数、独立変数、independent variable
- 誤差の取り方：指定された独立変数における測定誤差。だから誤差は垂直方向に分布する。
- 回帰直線からのズレ(誤差)のことを回帰分析では「残差」と呼ぶ。
- 残差の分布は正規分布を仮定する
- 測定場所によらず、残差は同一の分布であることを仮定する
- どうやって直線を決める?：予測誤差の2乗和を最小にする

2.1. 定式化

回帰平面の方程式:

$$y = a + bx$$

測定値と予測値のズレ:

$$\varepsilon_i = y_i - (a + bx_i) \quad (i = 1, 2, \dots, n)$$

ズレの2乗の和を最小に:

$$S = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \{y_i - (a + bx_i)\}^2 \rightarrow \min$$

- [余談] 回帰分析では「2乗和」を最小にすることを考えるが、「絶対値和」とか「符号付き和(1乗和)」を最小にする方法もアイデアとしてはあり得る。2乗和だと式の展開が楽になる。
- 説明する変数の個数
 - 一つ: 単回帰分析: 身長で説明したい: [体重]=a+b[身長]
 - 複数: 重回帰分析: 身長と胸囲で説明したい: [体重]=a+b[身長]+c[胸囲]

```
# 回帰分析(指定方法1)
ResultHW <- lm(Weight ~ Height,
               data=StudHW) # 線形モデル, Linear Model
ResultHW # 簡略化した計算結果表示
```

```
##
## Call:
## lm(formula = Weight ~ Height, data = StudHW)
##
## Coefficients:
## (Intercept)      Height
##    -81.5878      0.8352
```

なお、以下のように記述しても同じ結果が得られる。こちらの方が表記としては理解し易いかもしれないが、長くなるのが難点か。

```
# 回帰分析(指定方法2)
ResultTmp1 <- lm(StudHW$Weight ~ StudHW$Height,
                 data=StudHW) # 線形モデル, Linear Model
ResultTmp1 # 簡略化した計算結果表示
```

- 回帰直線の係数はそれぞれ -81.6と0.835と求められた。
- これを使えば、教室の入り口にメジャーを貼り付けておけば体重を言い当てられるはず。

[演習2.1-1] 試しに自分の体重がうまく予測できているか計算してみよ。

- $-81.6 + 0.835 \times 178 = 67.0$ Kg。
- 残差は $87.6 - 67.0 = 20.6$ Kg。
- うーん、メタボ健診に引っかかるわけだ。
- 部分的に係数や残差を取り出して表示することもできる。

```
coef <- ResultHW$coefficient # 計算結果の取り出し(回帰係数)
coef # 回帰係数
```

```
## (Intercept)      Height
## -81.587782      0.835239
```

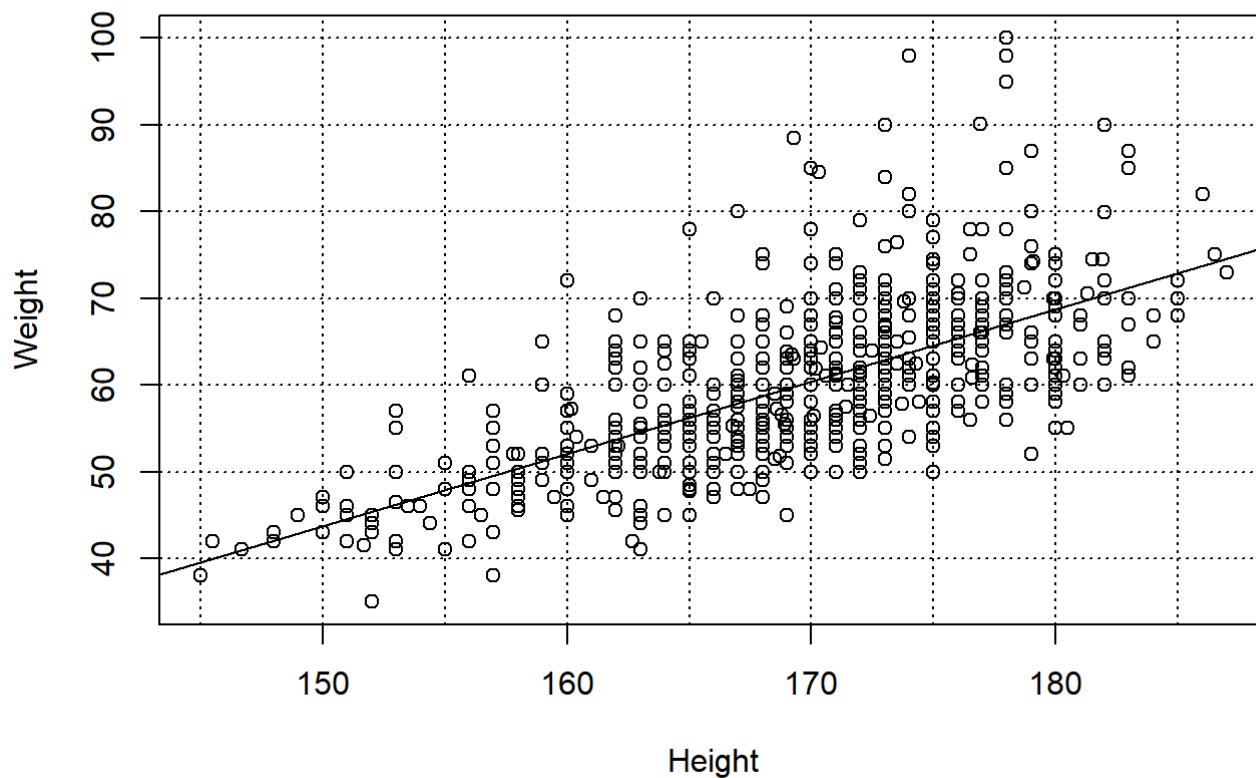
```
resid <-RresultHW$residuals # 計算結果の取り出し(残差)
resid[1:20]                 # 残差(先頭の20例)
```

```
##          1          2          3          4          5          7
## -1.52187498  2.06050551  0.05821870 -0.02759202  0.97240798  2.13716897
##          8          9          10         11         12         13
## -0.69807004  2.30192996  3.30192996 -2.53330906  0.46669094  1.46669094
##          14         16         17         18         19         20
##  5.46669094 -3.61797637 -10.36854807 -2.36854807 -1.36854807 -0.36854807
##          21         22
## -5.20378708 -4.20378708
```

- 散布図に回帰直線を付加する方法。

```
plot(Weight ~ Height, data=StudHW,
     xlab = "Height", ylab = "Weight",
     main="Scatter Plot of Shintyou and Taijyuu") # 散布図の描画(これまでと別の指定方法)
abline(RresultHW)                               # 回帰直線を付加
abline(h=seq(40, 100, 10), lty=3)
abline(v=seq(140, 190, 5), lty=3)
```

Scatter Plot of Shintyou and Taijyuu



3. 回帰係数以外の重要な情報

回帰分析について、回帰係数以外にも知っておくべき情報がある。それらは「summary(RresultHW)」の出力中に表示されている。

```
summary(RresultHW) # より詳細な計算結果表示
```

```
##
## Call:
## lm(formula = Weight ~ Height, data = StudHW)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.920  -4.732  -0.917   3.435  34.256
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -81.58778     6.25615  -13.04  <2e-16 ***
## Height       0.83524     0.03688   22.65  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.166 on 674 degrees of freedom
## Multiple R-squared:  0.4322, Adjusted R-squared:  0.4313
## F-statistic: 513 on 1 and 674 DF,  p-value: < 2.2e-16
```

- 決定係数 : Adjusted R-squared
 - 目的変量が説明変量でどの程度説明しているかの割合。
 - 1に近いほど当てはまりが良いと言える。

```
Adjusted R-squared: 0.4306
```

- 説明変量が予測に役立っているかの検定
 - 回帰に役立っているか : p-value : 小さいと有意(役立っている)

```
F-statistic: 462.2 on 1 and 609 DF,  p-value: < 2.2e-16
```

- 回帰係数がゼロではないかの検定
 - 回帰係数の検定(係数=0か?): Pr(>|t|) : 小さいと有意(係数=0ではないと言える)

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -80.70745     6.54697  -12.33  <2e-16 ***
Height       0.82931     0.03857   21.50  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

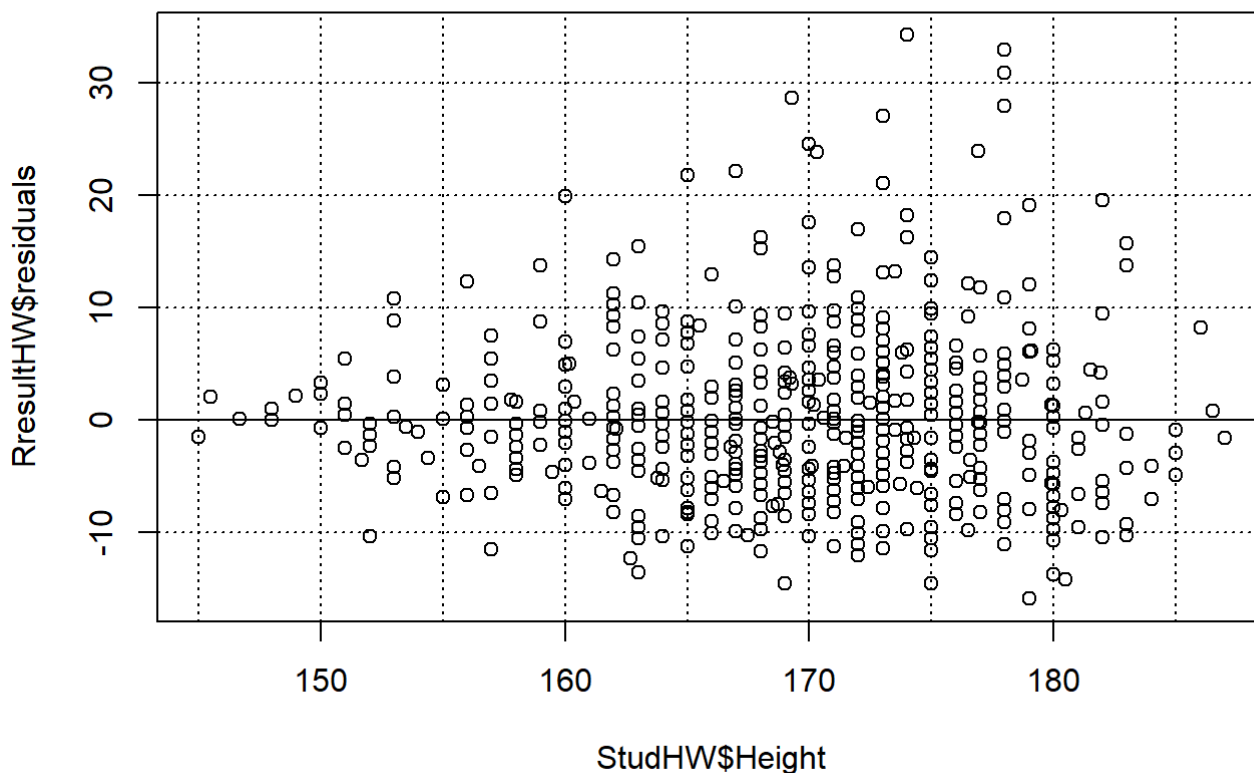
4. 残差分析もお忘れなく

- 回帰直線の係数が求まったらそれでおしまい?
- 仮定が3つあったはず
 - 関係性は直線=線形を仮定する
 - 誤差の分布は正規分布を仮定する

- 測定場所によらず、残差は同一の分布(正規分布)であることを仮定する
- 仮定が成り立っているかは確認せねば。どうやって？
 - もし、残差が何らかの傾向を示しているなら、直線＝線形関係ではないのかもしれないと疑う必要がある。==> バナナカーブ
 - 測定する場所で残差の傾向が異なるなら、残差の分布が同一とは言えないかもしれない。==> 順に広がっているラッパカーブ
 - 残差のヒストグラム、箱ひげ図で偏りが判定できる。
 - Q-Q プロットの斜め直線に乗っていれば正規分布と言える。ズレが大きいようだとの検討の余地がある。
 - 程度問題ではあるが、少しぐらいの逸脱は許容範囲と言える。
 - 残差が正規分布をするということを意識してもらえば良い。

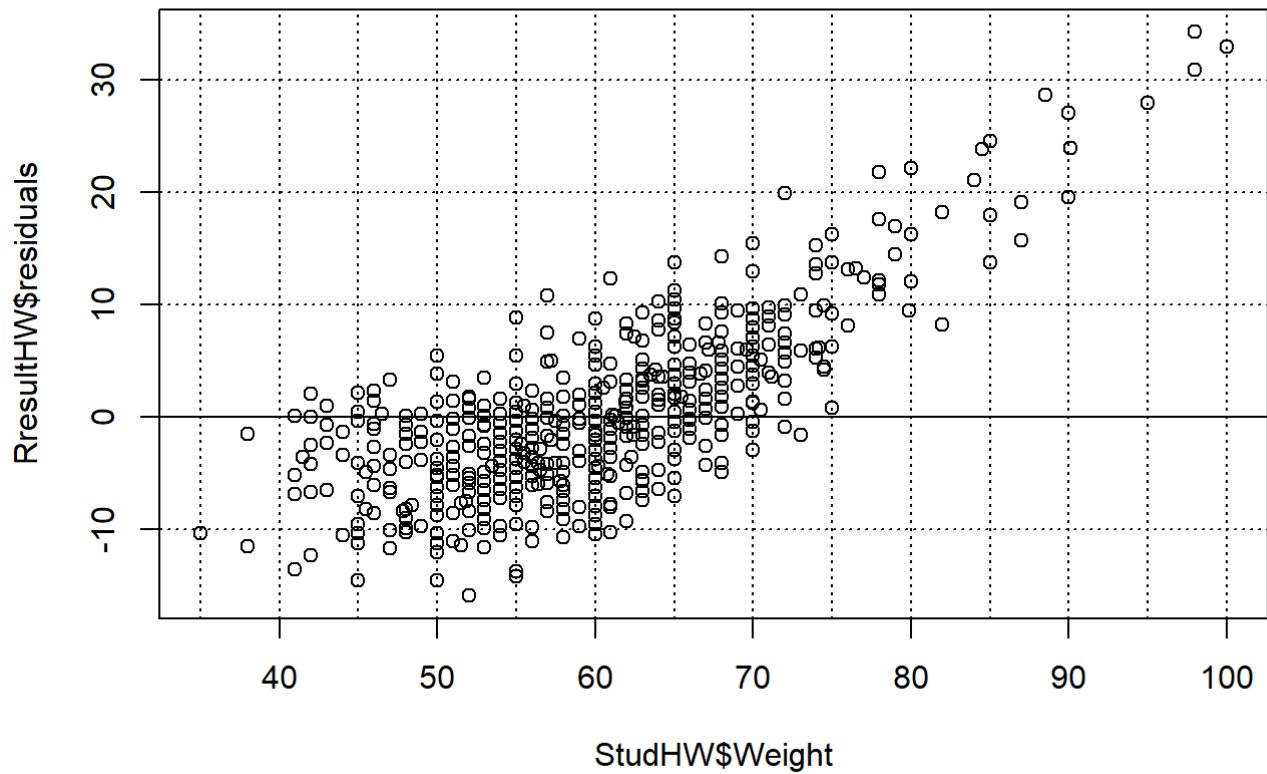
残差プロット: 対身長に対して

```
plot(StudHW$Height, ResultHW$residuals) # 残差プロット(対身長)
abline(v=seq(140, 190, 5), lty=3)      # 点線を追記
abline(h=seq(-10, 30, 10), lty=3)     # 点線を追記
abline(h=0, lty=1)                    # 原点を追記
```



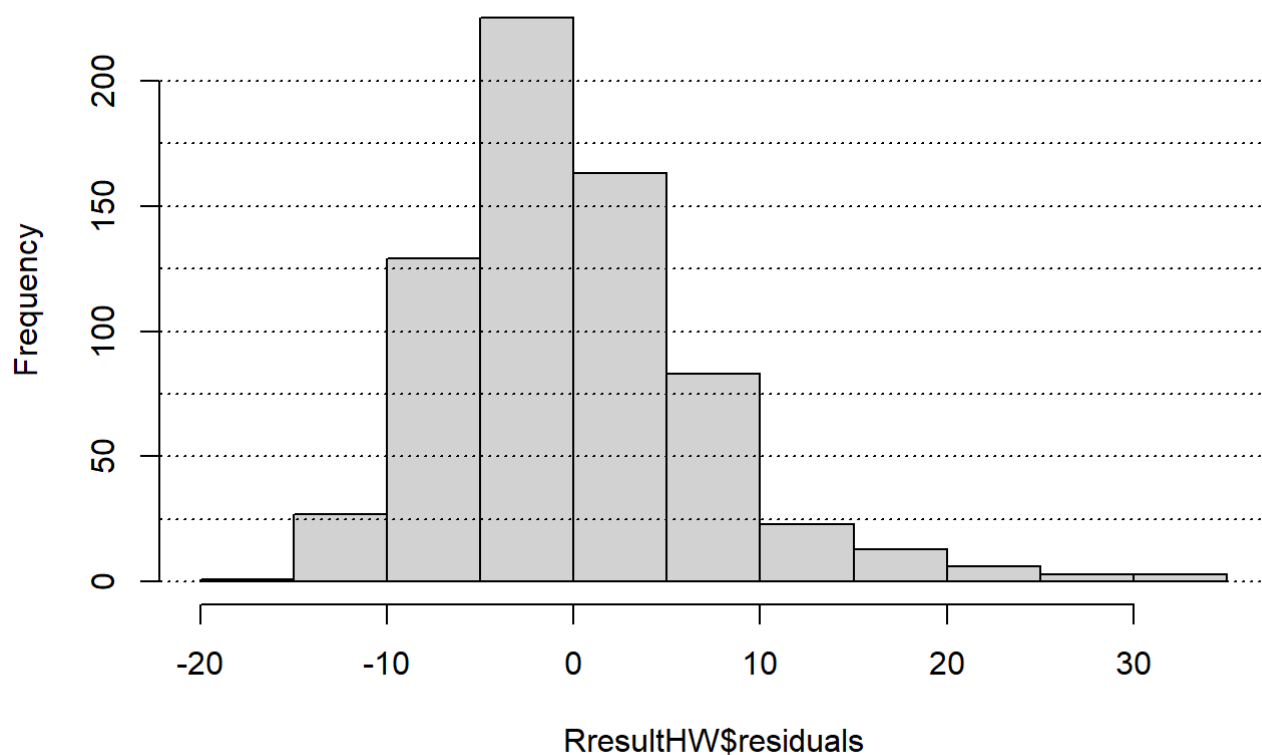
残差プロット: 対体重に対して

```
plot(StudHW$Weight, ResultHW$residuals) # 残差プロット(対体重)
abline(v=seq(30, 100, 5), lty=3)      # 点線を追記
abline(h=seq(-10, 30, 10), lty=3)     # 点線を追記
abline(h=0, lty=1)                    # 原点を追記
```

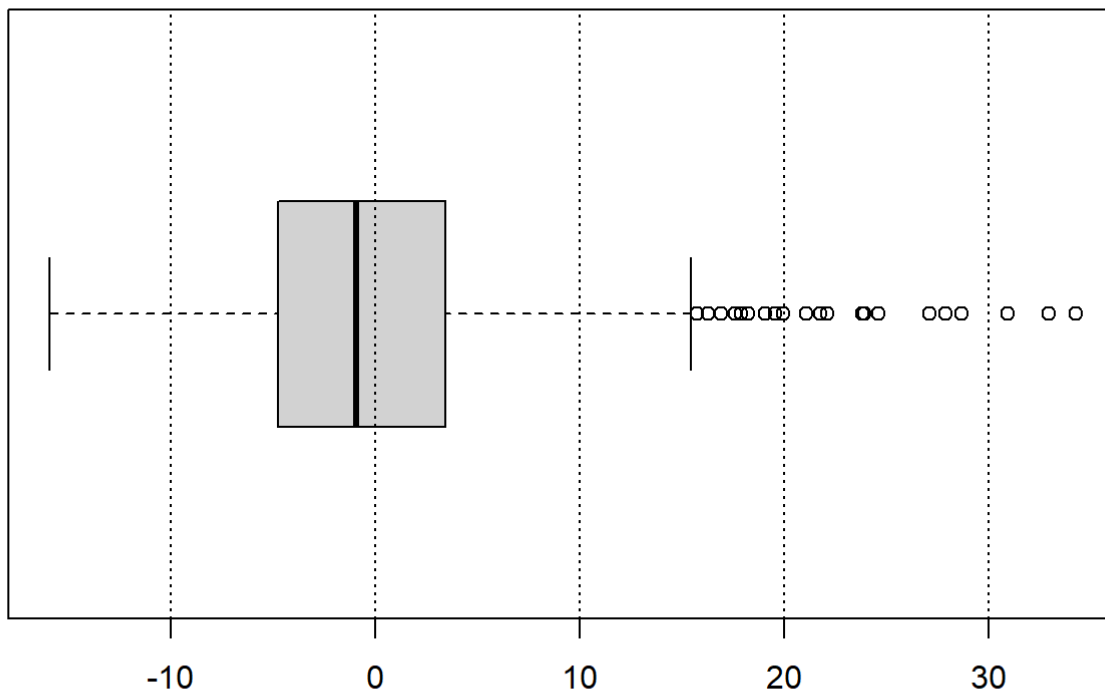


```
# 残差のヒストグラム  
hist(RresultHW$residuals, right=F)  
abline(h=seq(0, 200, 25), lty=3) # 点線を追記
```

Histogram of RresultHW\$residuals



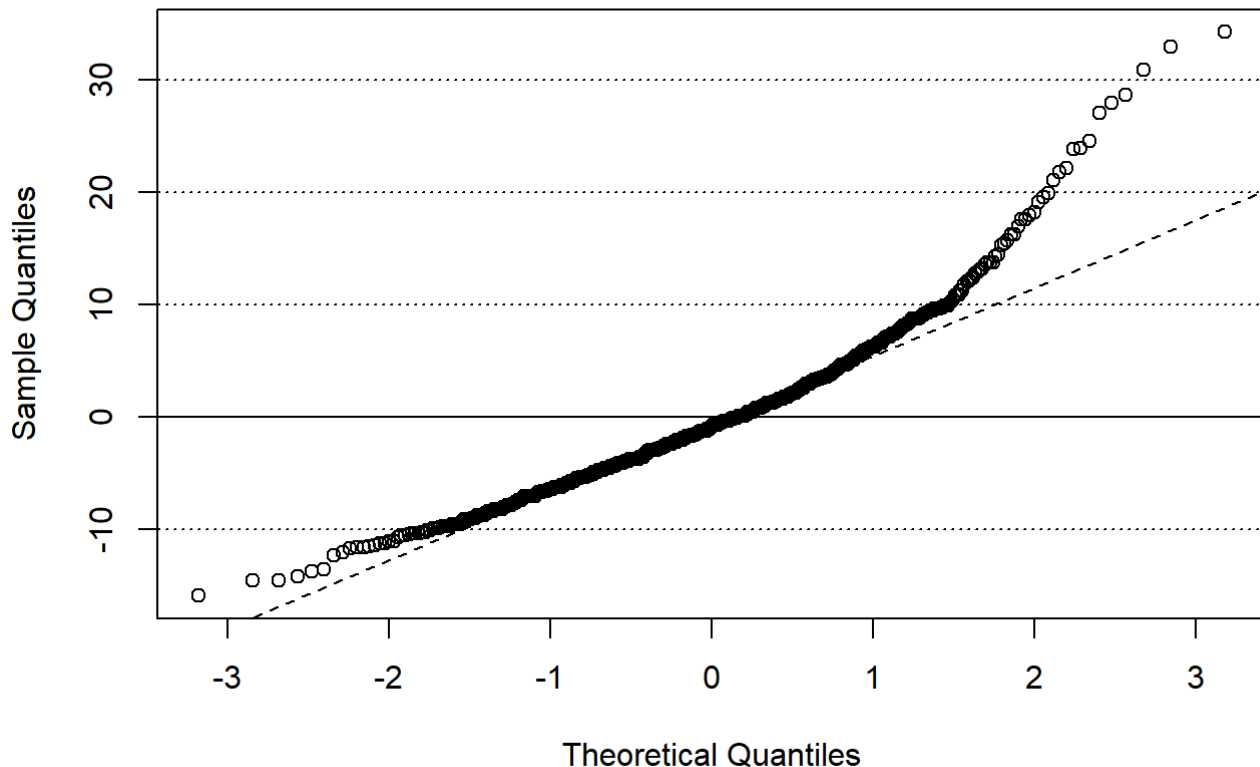
```
# 残差の箱ひげ図  
boxplot(RresultHW$residuals, horizontal=T)  
abline(v=seq(-10, 30, 10), lty=3) # 点線を追記
```

- 残差のヒストグラム、箱ひげ図で偏りが判定できる。

```
# 残差のQQプロット: 正規確率プロット
qqnorm(RresultHW$residuals) # QQプロット(正規確率プロット)
qqline(RresultHW$residuals, lty=2) # 基準線の描画
abline(h=0, lty=1) # 点線を追記
abline(h=seq(-10, 30, 10), lty=3) # 点線を追記
```

Normal Q-Q Plot



- Q-Q プロットの斜め直線に乗っていれば正規分布と言える。ズレが大きいようだとの検討の余地がある。
- [検討事項] 残差が20を超える辺りの(少なくとも)8例ほどは分布を乱しているように見える。====> 外れ値か? 吟味が必要。次回に解説。

5. この分析のまとめ(結果の見方)

- 対象になったのは 676名。
- 説明変数が予測に役立っているか?
 - 回帰に役立っているか : p-value : 小さいと有意(役立っている)
 [この例] 1% 未満(0.01%) なので役に立っていると言える。
- 決定係数 : Adjusted R-squared: (相関係数 : R)
 - 目的変数が説明変数でどの程度説明しているかの割合。
 - 1に近いほど当てはまりが良いと言える。
 [この例] 43% 程(約半分)を説明できている。
- 回帰係数 : Estimate
 [この例] $a=-81.6$, $b=0.835$
- 説明変数が予測に役立っているか?
 回帰係数の検定(係数=0 か?) : $\Pr(>|t|)$: 小さいと有意(係数=0ではないと言える)
 [この例] 両者とも 1% 未満(0.01%) なので回帰係数はゼロではない(何らかの意味がある数字と言える)。
- 残差の性質 ====> 正規性 : 残差プロット、残差解析
 - 残差(予測誤差)は正規分布をしていると仮定してモデルが構築されている。
 - この仮定が覆ると、回帰分析として成立していないことになる。
 - 残差が正規分布をしているか確認する必要がある。

- 均等に散らばっているか?
- 傾向はないか?
- ...

[この例] 残差には概ね傾向は見られない。ただし体重の大きい5例程度は要確認。場合によっては外れ値として除外も。

[演習5-1] 上述では[体重]を[身長]で説明する回帰式の例で説明した。では、[体重]を[胸囲]で説明するとどうなるか? 各自でやってみよ。

[演習5-2] スマホ月額料金を仕送り額で説明する回帰分析等、興味のある組み合わせで実行してみよ。

6. 有効桁数に注意せよ：どこまでが「意味ある桁」か?

測定精度上回る計算結果は表示させることはできても、意味はない。プログラムの出力をそのまま書き写さないように。

[重要な注意] 統計ソフトは単なる道具。使いこなすのは各自。

[例1] 四捨五入の数値で考えてみれば：精度(正確さ)が異なることに注意

12.3 <==== 12.25~12.34
 12 <==== 11.5 ~12.4
 67.8 <==== 67.75~67.84
 68 <==== 67.5 ~68.4

[例2] 日本の観測史上の最高気温は、2018(平成30)年7月23日に熊谷市で観測された41.1度であり、最低気温は、1902(明治35)年1月25日に北海道旭川市の-41度であった。====> -41.0度

[例3] 2001年のイチロー選手の打率は3割5分であった。2006年は3割3分1厘であった。====> 3割5分0厘

【今週の思い】 回帰分析のアイデアを理解し、分析手順を習得せよ。

81. 参考

このページで取り扱ったプログラムだけを抜き出して以下に列挙しておく。

```
## 1.1. アイディア
tmpRand<-rnorm(2000)
tmpRand[1:40]

StudTmp<-Student23
StudTmp$Height <-Student23$Height+tmpRand[1:598]*0.2
StudTmp$Weight <-Student23$Weight+tmpRand[1001:1598]*0.2

# 乱数で位置をずらした散布図
plot(StudTmp$Height, StudTmp$Weight,
      xlab = "Height", ylab = "Weight",
      main="Scatter Plot of Shintyou and Taijyuu")
abline(h=seq(40,100,10), lty=3)
abline(v=seq(140,190,5), lty=3)

## # 元データそのものの散布図(点が重なって判読できない)
## plot(Student23$Height, Student23$Weight)
## abline(h=seq(40,100,10), lty=3)
## abline(v=seq(140,190,10), lty=3)

RresultTmp <- lm(Weight ~ Height, data=StudTmp)
## RresultTmp
## summary(RresultTmp)

plot(Weight ~ Height, data=StudTmp,
      xlab = "Height", ylab = "Weight",
      main="Scatter Plot of Shintyou and Taijyuu")
abline(RresultTmp)
abline(h=seq(40,100,10), lty=3)
abline(v=seq(140,190,5), lty=3)

## 1.2. 前準備(Rでの利用に際して)
# データの切り出し
colnames(Student23)
StudTmp1<-Student23[,2:3] # 身長(2番目)と体重(3番目)だけを切り出す
StudTmp1[1:15,]
dim(StudTmp1)

StudHW<-na.omit(StudTmp1) # 欠損値を含むデータを削除
dim(StudHW)
colnames(StudHW)
StudHW[1:15,]

# 欠損値を含むデータの例数カウント(方法1)
StudTmp2<-!is.na(StudTmp1)
# StudTmp2[1:15,]
StudTmp3<-StudTmp2[,1] & StudTmp2[,2]
# StudTmp3[1:15]
# (StudTmp3==F)[1:15]
sum(StudTmp3==F)

# 欠損値を含むデータの例数カウント(方法2)
StudTmp2<-is.na(StudTmp1)
# StudTmp2[1:15,]
```

```
StudTmp3<-StudTmp2[,1] | StudTmp2[,2]
# StudTmp3[1:15]
sum(StudTmp3) ## これと同じ意味 ==> sum(StudTmp3==T)

## 2.1. 定式化
# 回帰分析の表記方法(その1)
ResultHW <- lm(Weight ~ Height,
               data=StudHW) # 線形モデル, Linear Model
ResultHW # 簡略化した計算結果表示

# 回帰分析の表記方法(その2)
ResultTmp1 <- lm(StudHW$Weight ~ StudHW$Height,
                 data=StudHW) # 線形モデル, Linear Model
ResultTmp1 # 簡略化した計算結果表示

coef <-RresultHW$coefficient # 計算結果の取り出し(回帰係数)
coef # 回帰係数

resid <-RresultHW$residuals # 計算結果の取り出し(残差)
resid[1:20] # 残差(先頭の20例)

# 散布図に回帰直線を付加する方法
plot(Weight ~ Height, data=StudHW,
     xlab = "Height", ylab = "Weight",
     main="Scatter Plot of Shintyou and Taijyuu") # 散布図の描画(これまでと別の指定方法)
abline(RresultHW) # 回帰直線を付加
abline(h=seq(40, 100, 10), lty=3)
abline(v=seq(140, 190, 5), lty=3)

## 3. その他の重要な情報
summary(RresultHW) # より詳細な計算結果表示

## 4. 残差分析もお忘れなく
# 残差プロット: 対身長に対して
plot(StudHW$Height, RresultHW$residuals) # 残差プロット(対身長)
abline(v=seq(140, 190, 5), lty=3) # 点線を追記
abline(h=seq(-10, 30, 10), lty=3) # 点線を追記
abline(h=0, lty=1) # 原点を追記

# 残差プロット: 対体重に対して
plot(StudHW$Weight, RresultHW$residuals) # 残差プロット(対体重)
abline(v=seq(30, 100, 5), lty=3) # 点線を追記
abline(h=seq(-10, 30, 10), lty=3) # 点線を追記
abline(h=0, lty=1) # 原点を追記

# 残差のヒストグラム
hist(RresultHW$residuals, right=F)
abline(h=seq(0, 200, 25), lty=3) # 点線を追記

# 残差の箱ひげ図
boxplot(RresultHW$residuals, horizontal=T)
abline(v=seq(-10, 30, 10), lty=3) # 点線を追記

# 残差のQQプロット: 正規確率プロット
qqnorm(RresultHW$residuals) # QQプロット(正規確率プロット)
qqline(RresultHW$residuals, lty=2) # 基準線の描画
```

```
abline(h=0, lty=1)      # 点線を追記  
abline(h=seq(-10, 30, 10), lty=3)  # 点線を追記
```