R Maniax Advance

フリーの統計ソフト「R」をさらに使いこなす本

著:後藤和智（後藤和智事務所OffLine）／ イラスト:くっく（エラー404）

後藤和智事務所OffLine

R Maniax Advance

フリーの統計ソフト「R」を
さらに使いこなす本

著：後藤和智（後藤和智事務所 OffLine）
表紙イラスト：くっく（エラー 404）

まえがき

43 冊目の同人誌となります、後藤和智です。本書は、2013 年の「コミックマーケット 84」（2013 年夏コミ）にて出した同人誌『R Maniax――フリーの統計ソフト「R」を使いこなす本』（後藤和智事務所 OffLine、2013 年（コミックマーケット 84））の続編にあたる同人誌で、サブタイトルにもあるとおりフリーソフト「R」の解説書です。本書では基礎的な事項が中心であった前著よりもさらに高度なデータ分析に踏み込み、より複雑な分析を行いたい方の要求に応えるものとなっています（たぶん）。R はオープンソースで開発されている統計解析ソフトであり、現在もアップデートが重ねられています。特に 2013 年 4 月に 8 年ぶりに行われた、R2.15.3 から R3.0.0 へのメジャーアップデートでは大きなベクトルデータにも対応することが可能になりました（参考：http://sourceforge.jp/magazine/13/04/08/153000）。

私も前著を出して以降、2014 年 1 月の「艦隊これくしょん～艦これ～」（以下、「艦これ」）オンリーイベント「海ゆかば 2」で出した『提督のための統計学――艦隊決戦統計解析論序説』（後藤和智事務所 OffLine、2014 年（海ゆかば 2）。蛇足ですけど同書は 2011 年の東方ジャンル進出から 3 年ぶりとなる新ジャンル進出ということになります）や、同年 4 月の「仙台コミケ 216」で出した『「ヤンキー」論の奇妙な位相――平成日本若者論史 9』（後藤和智事務所 OffLine、2014 年（仙台コミケ 216））などにおいて、新しい手法をいろいろと試しています。具体的には前者で実験計画法、後者でテキストマイニング及びフリーのテキストマイニングソフト「KH Coder」の導入です。また 5 月の「杜の奇跡 22」で出した『青少年政策の計量分析 2――平成日本若者論史 10』（後藤和智事務所 OffLine、2014 年（杜の奇跡 22））では、従来の sem パッケージよりもより簡単に共分散構造分析（パス解析）を行うことができるパッケージ「lavaan」を使ったパス解析を導入しております。そのほか『改訂増補版 紅魔館の統計学なティータイム――市民のための統計学 Special2』（後藤和智事務所 OffLine、2013 年（コミックマーケット 85））では R に連動するソフトである「RStudio」の解説も行っております。これらの成果についても、本書に「逆輸入」させております。

R の解説書は現在も多く刊行されており、コミケで出される統計本においても R を使ったものが多く見られるようになりました。R は無料の統計ソフトとしてはかなり多くの分析が行えると共に、拡張性も高く、なぜか商用版まで出てしまっている（「Revolution R Enterprise」http://www.revolutionanalytics.com/revolution-r-enterprise。日本語サイトもあり。http://www.r-analytics.jp/）という状況です。「ビッグデータ」への注目などで統計学が採り上げられている状況の中で、やはり自分で分析することにより統計学の価値（と限界）を学ぶことができる「R」の重要性は以前よりもさらに増している、というのは本シリーズ及び前著の元となった『三訂版・市民のための統計解析』（後藤和智事務所 OffLine、2011 年（コミックマーケット 81））/『統合版・市民のための統計学』として『市民のための〈基礎から学ぶ〉統計学』（後藤和智事務所 OffLine、2010 年（サンシャインクリエイション 49）ブクログのパブーにて配信中）から変わっ

4

ておりません。

ついでですが、本書ではRなどの既存の統計ソフトに取って代わるとまで言われている、プログラミング言語「Python」（パイソン）での統計処理についても簡単に触れます。Pythonは、人によってはただの統計ソフトともあるいはプログラミング言語とも解釈が分かれるRとは違って、れっきとしたプログラミング言語であり、また科学計算の分野やプログラマーなどから近年絶大な支持を得ています。もちろん現時点でRでパッケージなどを使ってできる豊富な統計処理でも、Pythonではできないものも多いようですが、この2つを学べば、将来のデータサイエンスや統計同人という点でかなり強力なツールを得ることができると思います。

てか、PythonでRも動かせるようですし…。（参考：「RjpWiki」より「PythonでR」http://www.okada.jp.org/RWiki/?Python%20%A4%C7%A1%A1R）

本書の表紙イラストは、主に「東方Project」の古明地こいし（「東方地霊殿」EXステージボスなど）などのデフォルメイラストや4コマ漫画を描かれている、サークル「エラー404」のくっく氏に担当していただきました。氏の漫画ではかわいらしいデフォルメの絵柄とは対照的に（？）、意外と理不尽な展開も多くて、振り回される役である姉・古明地さとり（「東方地霊殿」4面ボスなど）が不憫に思えてきます…（苦笑）。また氏は2012年冬コミよりイラスト集も刊行されているので興味を持った方は是非とも手に取られてみてはいかがでしょうか。

統計学を使った同人誌も少しずつではありますが増えてきており、それらの傍らに本書や前著を置いていただけると幸いです。また本書が刊行される「コミックマーケット86」では、サークル「でいひま」が主宰される艦これ統計合同誌『**統計これくしょん〜統これ〜』（でいひま、2014年（コミックマーケット86））**にも寄稿しております。同書は、同サークル主宰の牧田翠氏をはじめ、「Paradoxical Library」のありらいおん氏、「久幸繙文」の久樹輝幸氏などといった豪華なメンバーが揃っているのでこちらもお見逃しなく。

注意

※1 本書はR3.1.0に対応しております。

※2 本書は、既刊同人誌『提督のための統計学——艦隊決戦統計解析論序説』（後藤和智事務所OffLine、2014年（海ゆかば2））のデータを一部流用している箇所があります。同書のメイキングを少し含みますので、あらかじめご了承ください。

目次

6

第 1 章 R の導入・セットアップのおさらい

1.1 はじめに

本書の解説に入る前に、「R」のセットアップについてもう一度おさらいしておきます。まえがきでも書いたとおり、「R」はオープンソースで開発されているフリーソフトであり、そのためインターネットに繋がっていさえいれば誰でも無料で導入することができます（かつては R の解説書に R をインストールするための CD-ROM も附属していることが多かったですが、最近はほとんどありませんね）。

ただ R は極めて便利ですが、その便利さを支えているのが様々なユーザーによって開発されたパッケージであり、R で満足のいく統計解析を行うためにはパッケージは必須となります。一部の支援ソフトも、パッケージによって提供されているものがあります。また R で統計解析を行う際には、「RjpWiki」をはじめとするオンラインの意見交換の場もありますが、できれば自分に合った解説書を購入し、オフラインでも参照できるようにしましょう。本書でおすすめする解説書は次の通りです。

舟尾暢男『The R Tips 第 2 版――データ解析環境 R の基本技・グラフィックス活用集』オーム社、2009 年

前著や『市民のための統計解析』シリーズでも散々採り上げている、もうおなじみの R ユーザー必携の解説書ですね。統計解析はもとより、グラフィックやプログラミング、シミュレーションまで、基本的な事項のほとんどを説明しております。

石田基広『改訂 2 版 R 言語逆引きハンドブック』C&R 研究所、2014 年

目的に応じて引く R の新しい解説書ですが、このたびバージョン 3 以降に対応するために改訂版が刊行されました。こちらは目的別に記されているので、『The R Tips』よりも実用性が高くなっております。少々分厚いですが『The R Tips』とは違い A5 版なので、持ち運びもしやすくなっております。

山田剛史ほか『R によるやさしい統計学』オーム社、2008 年

こちらは検定や種々の多変量解析などに対応した、分析に使う人向けの解説書です。Kindle でも配信が始まっております。

Winston Chang：著、石井弓美子ほか：訳『R グラフィックスクックブック――ggplot2 によるグラフ作成のレシピ集』オライリージャパン、2013 年

8

本書でも触れることになる、Rの高度なグラフィックパッケージggplot2の解説書です。ggplot2では従来の（『The R Tips』などで紹介されている）plotコマンドではできない高度なグラフィックもいろいろできるので、ggplot2でのグラフ作成も試してみてはいかがでしょうか。

豊田秀樹ほか『データマイニング入門』東京図書（Rで学ぶ最新データ解析）、2008年

東京図書から出ている「Rで学ぶ最新データ解析」シリーズはどれも解説書としてよくできていますが、とりわけおすすめしたいのが同書です。同書は種々の多変量解析やデータマイニング、そして一部グラフィックまで含む複雑な統計処理を紹介しており、多変量解析を使いたいユーザーは必携でしょう。

西内啓（ひろむ）『統計学が最強の学問である』ダイヤモンド社、2013年

今の統計学ブームを巻き起こした一冊。Rも少し出てきますが、モチベーション向上用に。Kindle版あり。

森田果（はつる）『実証分析入門――データから「因果関係」を読み解く作法』日本評論社、2014年

Rの解説書ではありませんが、統計学の解説書として。章のサブタイトルに詰め込まれたネタ（「高校時代に逢った、ような…」など）で話題をさらいましたが、社会科学における統計学的な考え方の解説書として極めていい仕上がりになっております。

豊田秀樹『購買心理を読み解く統計学――実例で見る心理・調査データ解析28』東京図書、2006年

こちらもRの解説書ではなく、統計学の解説書ですが、マーケティングや心理学・社会学に使う多変量解析・データマイニングの手法が網羅的に紹介されています。Rでできるものもいくつかありますが、それらは自分で調べるのがいいでしょう。

後藤和智『統計同人誌をつくろう！――調べて、分析して、書きたい人のために』後藤和智事務所OffLine、2013年（コミックマーケット85）

統計学に基づいた同人誌の製作という方向性を重視した、統計学の解説書です。基本的な手法の紹介の他、RやLaTeXなどのソフト使い方を解説しております。『エロマンガ統計』シリーズの著者、牧田翠氏との対談も収録しております。

1.2 Rをインストールする

リアル及びパソコン内の環境、そしてモチベーションを整えたらRをインストールしましょう。RはCRANというRのサーバー（ないしそのミラーサイト）から入手することができますが、日本でRを使いたいならそのオンライン上の最大のサイトである**「RjpWiki」**（http://www.okada.jp.org/RWiki/index.php?RjpWiki）をブックマークしておきましょう。Rの解説も兼ねております

ので、インターネットに接続できる環境があれば適宜参照できるようにしておくのがいいでしょう（「書籍」ページではなぜか弊サークルの同人誌も複数採り上げられています…。恐縮恐縮アンド恐縮です…）。

Windows の場合は、RjpWiki の「R のインストール」のページにある「最新版はこちらからダウンロードしてください」の「こちら」をクリックすると、筑波大学（2014 年 7 月現在）にある CRAN ミラーサイトの R の配布ページに飛ぶことができるのでそこから最新版を選択します。R 本体は「base」のページにあります（CRAN で配布されていない一部パッケージには R の最新版に対応しきれていないというものもありますが、それ以外の理由で敢えて旧版を入手するメリットは皆無に等しいです。最新版を入手しましょう。どうしても旧版が欲しいなら「Previous releases」のページに旧版があります）。最新版のインストーラは本体入手ページの一番上にあるのでわかりやすいと思います。

R をインストールしたら、とりあえず開いてみましょう。最初に開かれるのはコンソールと言い、この部分にコマンドを入力します。また操作に応じてグラフィックなどのウインドウが開かれます。

1.3 パッケージをインストールする

R をインストールしたら、次に R の豊富な機能を使うことができるようにパッケージをインストールしましょう。パッケージはインターネットを通じて CRAN から入手します。大量のデータのやりとりを行いますので、それなりに長い時間と速い回線が必要になります。

その前に、パッケージをインストールする際には、R を「管理者として」起動しましょう（R が Program Files フォルダ内にある場合は特に）。そのまま起動しているとインストールできないパッケージもいくつかあり、そこでインストールが止まってしまうのですが、管理者としてなら全てのパッケージのインストールを行えると思います。

管理者として R を起動する場合は、スタートメニューで R のアイコンを右クリックします。そうするとメニューが表示されるので、その中から「管理者として起動」を選択します。そこから R を管理者として起動できます。

R を管理者として起動したら、パッケージを持ってくるサイトを選択します。メニューバーの「パッケージ」から「ダウンロードサイトの選択」を選びます。Shift キーや Ctrl キーで複数選択できますが、おすすめは「Bioc Software」系のもの以外をすべて選択することです。そしてダウンロードサイトを選んだら、パッケージをインストールしましょう。メニューバーの「パッケージ」から「パッケージのインストール」を選びます。ここで CRAN ミラーサイトの選択が求められることがありますので、適当なものを選択しておきましょう。CRAN ミラーサイトを選択したらパッケージの一覧（ダウンロードサイトの選択によって変わってきます）が現れますので、R をインストールして最初のパッケージのインストールならば、Shift キーなどで全てのパッケージを選択しましょう。「OK」ボタンをクリックしたら、インストールが始まります。全てのパッケージをインストールする場合、相当時間がかかるのでその間は別の作業をするなりしましょう（目を離す場合は他の人にコンピュータを触らせないようにしてください。管理者として R を起動している場合

10

は特に注意してください)。

個別にパッケージをインストールする場合は、インストールしたいパッケージを選択して（これも Shift や Ctrl などで複数選択可）インストールします。もし目的のパッケージが中ったら、ダウンロードサイトの設定を変えて試してみましょう。それでも駄目な場合はそのパッケージは CRAN では提供されていないので配布者のサイトから直接取りに行く必要があります。それを R で使えるようにするにはメニューバーの**「パッケージ」の「ローカルにある zip ファイルからのパッケージのインストール」**でダウンロードしたパッケージを選択することによって初めて使えるようになります。

1.4 関連ソフト

本節では、R と一緒に使うと便利な関連ソフトの紹介と、その入手方法について解説していきます。R は本章冒頭などでも述べたとおり（Python などと比べて）統計ソフト・プログラミング言語（？）としての独自性が強く、またインターフェイス自体も決して親切とは言えないため、その習得コストは比較的大きなものとなっております（まあ一度習得してしまえばあとはその有り余る能力を使うことができるので得られるものも大きいですが)。そのような R の問題点を解決するため、R では様々な支援ソフトが開発されています。

また R とは関係ないソフトの中にも、R と連携することができるソフトもあります（本書の後の章で解説する MeCab や Python もその中に含まれます)。

・R コマンダー

R Commander（R コマンダー）は、2004 年から開発されている R の入力支援ソフトの一つです。CRAN でパッケージで提供されております。R コマンダーは、これさえ起動していれば、データの読み込みから分析、グラフィックなどが簡単な操作でできてしまうという優れものです。

データはテキストファイルの他、Excel や SPSS などのファイル、また Excel のクリップボードからも作成することができます。分析は回帰分析やクラスター分析などの各種多変量解析に対応、またグラフも様々なものを作成することができます（第 2 章で一部を紹介します)。いくつかの分析については、前著『R Maniax』にて解説しておりますのでそちらをご参照ください。

・Rz

Rz は大阪大学（2014 年 7 月現在）の**林真広**氏によって開発されている支援ソフトで、こちらも CRAN にてパッケージで提供されております。Rz はクロス集計や 2 変量の高度なプロットなどといった集計分野で高い性能を発揮します。こちらも『R Maniax』で少し採り上げておりますのでご参照ください。

Rz 解説ページ（日本語）：http://m884.jp/Rz_Ja.html

・RStudio

RStudio はオープンソースで開発されている R の入力支援ソフトです。デスクトップ用とサーバー用の 2 種類があり、それぞれに無料版と商用版の 2 つが存在します。

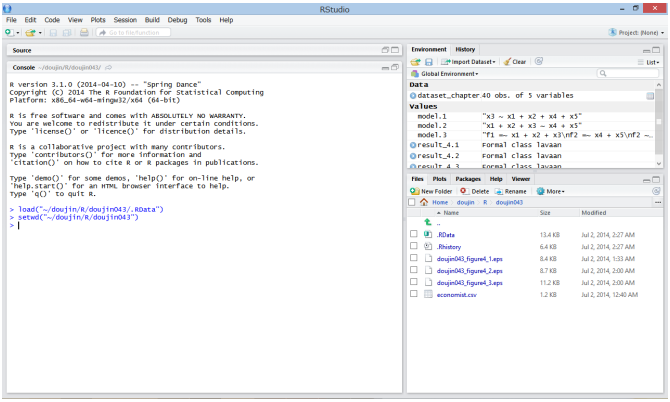

図 1-1 RStudio

RStudio はコンソール、データの一覧表示、ファイル・フォルダの管理などを一つの画面の中で簡単に行うことができるツールであり、R の操作はもちろん、ファイルの閲覧、パッケージの管理なども行うことができます。R コマンダーのような分析支援の機能はありませんが、パッケージで提供されているものは RStudio 上でも動かすことができます（ただし R コマンダーの場合、R コマンダー上で計算などが行われるのではなく、直接 RStudio のコンソールにコマンドが渡されることになります）。

RStudio はパッケージで提供されていません。そのため公式サイト（http://www.rstudio.com/）から入手する必要があります（「RStudio 統計」などで検索すればすぐに見つかります。なお 2014 年 7 月現在英語版しかないもよう）。公式サイトからインストーラを入手し、インストールしたらあとは終了です。追加の設定は不要です。

RStudio をそのまま起動すると、最初は何もない真っ白な画面が現れますが、コンピュータにインストールされている最新の R を自動的に起動し、RStudio が本格的に使えるようになれます（もしその直後にファイルの保存を求められた場合でも、キャンセルしましょう）。左側はコンソール、右上はグローバル環境（ワークスペースに保存されているデータ）、右下はファイル一覧やパッケージ一覧、グラフィックなどです。特にパッケージ一覧では、パッケージの名前のチェックボックスをオンにするだけでパッケージを呼び出すことができるので、大文字と小文字の打ち間違いでパッケージが呼び出せない、という事態を回避できます。

RStudio はグラフィックの出力などでも大きさを簡単に指定することができるので（もちろん通常の R のコンソールによる操作でもできますが、少し面倒です）、特に EPS で図を出力したい場合には打ってつけでしょう（Adobe Bridge などの EPS ファイルが簡単に見られるソフトと併用するといいでしょう）。

・MeCab、RMeCab

MeCab は Google 日本語入力の開発者が作成したオープンソースの日本語形態素解析エンジンであり、RMeCab はそれを R 上で動かすためのパッケージです。R でテキストマイニングを行いたいという場合は、MeCab をインストールし、さらに RMeCab もインストールする必要があります。

12

MeCab はフリーソフトなので開発者のサイト（http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html）から直接入手することができます。また **RMeCab** はパッケージ開発者の**石田基広**氏のサイト（http://rmecab.jp/wiki/index.php?RMeCab）から入手します。RMeCab を使う場合には、石田氏の**『R によるテキストマイニング入門』（森北出版、2008 年）**または**『R で学ぶ日本語テキストマイニング』（ひつじ書房、2013 年）**を買っておくのがいいでしょう。特に後者は、前者に比べて分析事例が豊富なので、テキストマイニングによる解析のレシピ集として打ってつけです。

なお LaTeX や MeCab などといったように、設定ファイルをテキストエディタなどで変える必要があるものは、Program Files のフォルダにインストールしない方が便利です。Program Files のフォルダは、その多くが管理者権限を持っていないと変更することができませんし、管理者権限で変更しても設定が反映されていないということもありますので注意が必要です。

・Ruby、Python

R は確かに便利かつ高性能なソフトなのですが、先に述べたとおりプログラミング言語としての独自性が高いため、その教育コストが高かったり（まあその分リターンも大きいのですが）するため、特に既存の他のプログラミング言語に慣れ親しんだ人に取ってはいろいろと苦労するところもあると思います。そのため R を他のプログラミング言語と並行させて使うというものも提唱されております。

例えば **Ruby** を併用するパターンです。**Sau Sheong Chang『R と Ruby によるデータ解析入門』（瀬戸山雅人、河内崇、高野雅典：訳、オライリージャパン、2013 年）**では、プログラミング言語である Ruby を併用したデータ解析が紹介されています。また本書では **Python** にも少しだけ触れますが、Python と R を併用するという考え方もあります。

何回か述べている通り、R はプログラミング言語としての独自性が強いため習得コストが高く、特に FORTRAN などで科学データ・経済データを扱ってきた方にとってはそれが R を忌避する原因になっているという側面もあるようです（なお筆者は FORTRAN は学生・院生のときに少しだけ使ったが、結局 R ばかり使っていたもよう）。ただ R のデータ解析能力は間違いなく魅力的なものです。経済やマーケティングなどで「ビッグデータ」を扱う人などは、これらのプログラミング言語と R の併用、というものも検討してみるといいかもしれません。

1. おまけ 統計学を楽しむためのおすすめ本 | サイト

統計学の有効性に疑問を持ったり、あるいは同じデータばかり眺めて飽きた場合は、いろいろなデータに触れてみるのもいいでしょう。世の中には統計学やコンピュータを用いたいろいろな研究もあり、それらを眺めることによって楽しむのもまたモチベーションを保つことに繋がります。たぶん。

吉川（きっかわ）徹『現代日本の「社会の心」――計量社会意識論』有斐閣、2014 年

長期的な意識調査の分析から現代の日本を読み解くという計量社会学の模範的な研究書です。専門書に比べて安価で手に入るので、社会学における統計学の使い方を考えるにはちょうどいい教材となっています。もう少し教育・青少年関係が欲しい、という方は、**渡辺秀樹ほか『勉強と居場所――学校と家族の日韓比較』（勁草書房、2013 年）**あたりをおすすめしておきます。ちょっと値は張りますが…。

谷本奈穂「ポピュラー音楽の歌詞における携帯電話の意味」（中村隆志：編著『恋愛ドラマとケータイ』（青弓社、2014 年）所収）

AKB48 などの歌詞をテキストマイニングによって分析し、ポピュラー音楽における内容を統計学的に分析しています。それ以外にも本書は統計学を用いた分析も多く、作品の浅薄な読解から現代社会の「リアル」なるものにこじつけるような「批評」とは一線を画しています。

サークル「でいひま」の同人誌

統計を扱った同人誌は多数ありますが、その中でも特におすすめなのがサークル「でいひま」の『エロマンガ統計』シリーズです。成人向け漫画から始まり、最近ではアニメやライトノベル、アダルトゲームなどにも手を伸ばし、解析手法も最初は単純集計と対応分析くらいだったものが、ベイズ統計、テキストマイニングなども導入するようになり、積極的に新領域・新手法に手を出していく様もまた面白いです。書籍は即売会や COMIC ZIN で入手できるほか、電子版もあるようなので、サークルのサイト（http://ch.nicovideo.jp/dayhima）を是非チェックしてみてください。

忍殺語形態素解析辞書「チャドー」アカウント @njdict_Chado

アイエエエ！　ニンジャ！？　ニンジャナンデ！？　ツイッター上（@NJSLYR）で連載されている、〈間違った日本観〉をベースにし、そのストーリーと独特な翻訳で話題をさらっている人気小説**『ニンジャスレイヤー』（書籍版は KADOKAWA（エンターブレイン）より刊行中）**の分析に特化した形態素解析辞書を作成しておられる方がいます（もちろん原作者や版元とは関係ない二次創作ですが）。こちらは忍殺（「ニンジャスレイヤー」の略称）の言葉を MeCab や KH Coder で分析しているようで、統計学から見た忍殺という楽しみ方を配信しておられます。忍殺と統計学、テキストマイニングに興味があったらフォロー重点な。

14

第 2 章 グラフィック

2.1 はじめに

本章では、データの可視化であるグラフィックについて説明します。データの分析した結果について、表などの形式で示すのもいいですが、やはり一番読者に対して訴求力があるのはグラフでしょう。もちろんグラフを使うことで簡単に読者などを「騙す」ことも可能なのですが、逆に言えばグラフによる「騙し方」を知ることにより、グラフを使った「騙し」に対する抵抗力をつける、という考え方も可能です（蛇足ですけど、ツイッターなどを見る限り、円グラフは最近の統計ユーザーの間では結構評判が悪いようです）。

前著『R Maniax』においてはグラフィックに触れたのは R コマンダーの部分くらいでほとんど触れなかったので、本章ではグラフィックについていくつか解説しようと思います。R によるグラフィックについては、基本的なグラフ描画コマンドである plot や、種々の分析に対応した独自の描画コマンド、そして ggplot2 という高度なグラフィックパッケージもあり、データの可視化という点でも R は優れた機能を持っています。

2.2 plot コマンドなどによる基礎的なグラフィック

2.2.1 散布図に回帰直線を追加する

ここでは、回帰分析において、散布図に回帰直線と補助線を追加してみるという操作をやってみます。まず、相関関係を持ちそうな適当なデータを作成するプログラムを下記のように作ります。

```
chapter2_make <- function() {
    N <- matrix(0,ncol=2,nrow=100)
    for (i in 1 : 100) {
      N[i,1] <- rnorm(1,mean=3,sd=3)
      N[i,2] <- rnorm(1,mean=N[i,1],sd=2)
      }
    colnames(N) <- c("x","y")
    N <- data.frame(N)
    return(N)
}
```

このプログラムでは、x を平均 3・標準偏差 3 の正規分布に従う乱数 100 個とし、y は平均を対

16

応する x、標準偏差 2 の正規分布に従う乱数としています（rnorm は正規分布に従う乱数を生成するプログラムです。rnorm(乱数の数 ,mean= 平均 ,sd= 標準偏差) と入力し、平均を省略した場合は 0、標準偏差を省略した場合は 1 になります）。このプログラムを使って、適当な x と y の組を作成します。

```
> dataset_c2 <- chapter2_make()
> plot(dataset_c2)
```

ここで生成した図を plot コマンドを使って描画すると、図 2-1 のようなものになります（もちろん乱数に基づいているので図は毎回変わりますが、基本的に形状は似たようなものになると思います）。しかしこれでは、両軸の目盛りがあってもわかりづらいです。

ここでは、x 軸と y 軸を描画してみます。x 軸は y=0 という式で示される直線であり、y 軸は x=0 の直線ですが、ここでは直線を引くコマンド abline を使って引いてみましょう（パッケージ不要）。abline コマンドの使い方は次の通りです。

```
abline(a= 傾き ,b=y 切片 ,lty= 線の種類 ,lwd= 線の太さ ) # 一般的な直線を引く場合
abline(h=y 位置 ,lty= 線の種類 ,lwd= 線の太さ ) # 水平な直線 y=h を引く場合
abline(v=x 位置 ,lty= 線の種類 ,lwd= 線の太さ ) # 垂直な直線 x=v を引く場合
```

x 軸と y 軸を追加するには、次のように入力します。

```
abline(h=0)
abline(v=0)
```

これで、x 軸と y 軸にあたる箇所に直線が引かれました（図 2-2）。次に回帰直線を引きます。y を x で回帰する単回帰モデルを、次のように作ります。

```
> lm_c2 <- lm(y~x, data=dataset_c2)
> lm_c2 # 確認

Call:
lm(formula = y ~ x, data = dataset_c2)

Coefficients:
(Intercept)            x
    -0.3820       0.9785

> cor(dataset_c2) # 念のため相関係数
```

```
          x         y
x 1.0000000 0.7885067
y 0.7885067 1.0000000
```

abline コマンドは回帰直線を引く場合にも使えます。次のように入力します。

```
abline( 回帰分析の結果 ,lty= 線の種類 ,lwd= 線の太さ ) #
    回帰分析に基づく直線を引く場合
```

ここでは次のように入力して、回帰直線を点線で引いてみます。点線は lty のところを 2 に指定すれば引くことができます。

```
abline(lm_c2, lty=2)
```

このようにしてできた図が図 2-3 となります。簡単な作図ですが、軸と回帰直線を引くと、データの性質がよりわかりやすくなると思います。

また、座標内に線分を引くには、lines コマンド（パッケージ不要）を使います。

```
lines(c(x 始点 ,x 終点 ),c(y 始点 ,y 終点 ),lty= 種類 ,lwd=
    太さ )
```

なお、参考までに、lty と lwd について、1 ～ 6 の値を入れたときの表示を図 2-4 に示してみました。線を引くときはこれをご参考にしていただけると幸いです。こちらのプロットもプログラムを使っています。

2.2.2 主成分分析・対応分析のプロット（バイプロット）

バイプロットとは、主成分分析や対応分析において、行（個票）の分析結果と列（パラメータ）の分析結果を同じ平面上に配置したものを指します。

主成分分析を例にとって説明すると、prcomp コマンドないし princomp コマンドによって作成された主成分

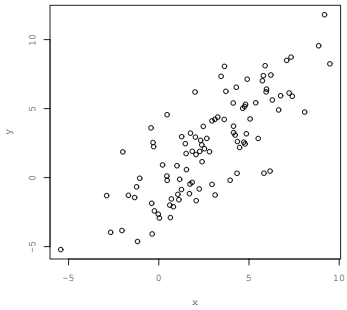

図 2-1 散布図プロット

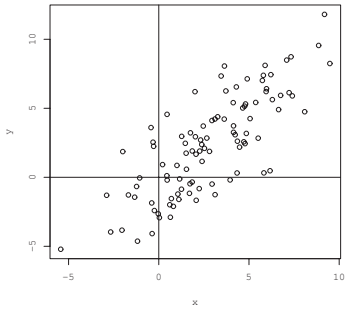

図 2-2 散布図に x,y 軸を追加

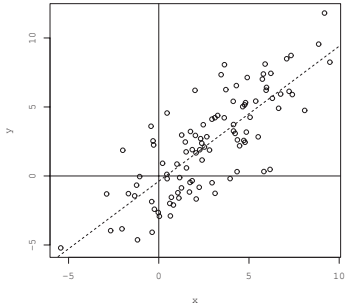

図 2-3 散布図に回帰直線を追加

18

分析の結果を用いて、次のように入力します。

biplot(結果 , choice= プロットしたい主成分)

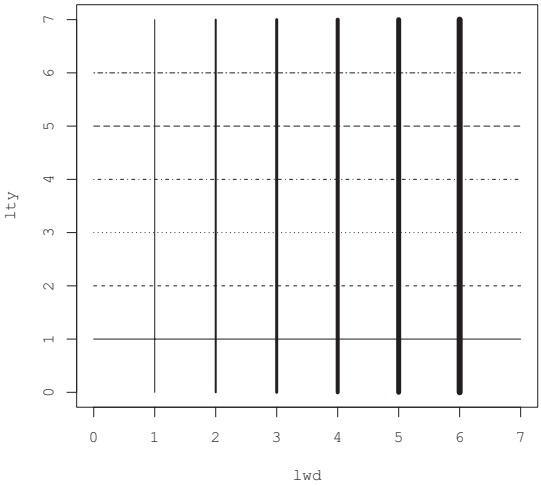

図 2-4 線の太さ（lwd）と種類（lty）

ここでは、2013 年プロ野球セ・リーグにおける規定投球回に到達した投手 17 人の防御率・WHIP（1 イニングあたりにどれだけ走者を出すか）・QS 率（全先発登板の内、6 回までに 3 自責点までで抑えた割合）・被打率・被 BABIP（BABIP とは、本塁打でない打数がどれだけフェアになるかの指標）・被出塁率・被長打率・K/BB（1 四死球あたりにどれだけ三振を奪うか）・奪三振率（K/9）・与四球率（BB/9）・被本塁打率（HR/9）の 11 指標について主成分分析をしようと思います（データは「プロ野球ヌルデータ置き場 2013 年版」http://lcom.sakura.ne.jp/NulData/2013/index.html より）。なお長打率と BABIP の計算式は次のようになります。

長打率＝塁打数 / 打席数 = (単打 + 二塁打 × 2 + 三塁打 × 3 + 本塁打 × 4) / 打席数
BABIP = (安打 - 本塁打) / (打数 - 三振 - 本塁打 + 犠飛)
参考：「Baseball Concrete」＞「用語解説」http://baseballconcrete.web.fc2.com/glossary.html

このデータを「baseball2013.csv」という CSV ファイルに入力し、次のように計算を行います。

```
> dataset_c2_bb <- data.frame(read.csv("baseball2013.csv", header=T, row.name=1))
> dataset_c2_bb
           防御率 WHIP  QS 率 被打率 被 BABIP 被出塁率 被長打率 K.BB  K.9 BB.9 HR.9
前田健太     2.10 0.96 0.769  0.203   0.243    0.251    0.290 3.95 8.09 2.05 0.67
能見篤史     2.69 1.08 0.800  0.232   0.255    0.275    0.349 3.10 6.33 2.04 0.90
スタンリッジ 2.74 1.29 0.731  0.266   0.304    0.320    0.367 2.64 6.95 2.63 0.67
メッセンジャー 2.89 1.17 0.690  0.241   0.290    0.297    0.341 3.27 8.39 2.57 0.60
小川泰弘     2.93 1.12 0.731  0.235   0.275    0.284    0.329 3.00 6.83 2.28 0.46
菅野智之     3.12 1.15 0.692  0.249   0.299    0.291    0.332 4.19 7.93 1.89 0.51
```