



R Commander を用いたデータ解析

長島 健悟
城北大学 薬学部

Kengo NAGASHIMA
Laboratory of Biostatistics,
Department of Pharmaceutical Technochemistry,
Joshi University

2010年1月5日

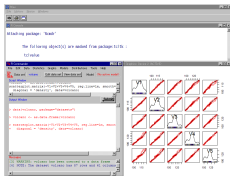
- オープンソースの統計解析システムで、統計計算とグラフィックスのための言語・環境
- "The Comprehensive R Archive Network (CRAN)" というユーザ団体により管理・メンテナンスされている、信頼性も高い
- 様々な分野・領域で用いられている (統計的手法の研究・データ解析)
- 追加パッケージを利用した拡張が可能、パッケージは世界中の R ユーザが開発して登録している (2009/12/24 日現在, 2115 個)
- R 言語 (プログラミング言語) によって処理を記述する。プログラミングが行えなくても、R Commander という GUI フロントエンド (マウス操作で処理できる) が開発されている
- フリーウェアであり、Web 上の解説ページ・書籍も多い、インターネット接続環境があれば、いつでも利用できる

1 / 82

2 / 82

R Commander とは

- Professor, John Fox (McMaster University) が R 用に開発した追加パッケージ
- マウス操作で R を用いたデータ解析や、グラフの作成ができる
- もちろんフリーウェア



R や R Commander のメリットとデメリット

メリット

- フリーウェアであり、おそらくどんな環境でも、パソコンとインターネット接続環境があれば利用できる

デメリット

- ユーザが有志で開発するため、市販品より不親切な部分もある
- ヘルプファイルは英語のものが多く、非常にユーザフレンドリーな設計とも言えない

3 / 82

4 / 82

Rのインストール (1-1)

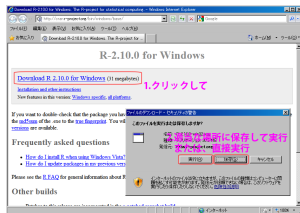
Windows 版 R のインストール手順です

- ローカルファイルからインストールする場合
 - セットアップファイル"R-2.10.0-win32.exe"をダブルクリックして実行し、スライド スライド 7 に進んでください
- インターネットからインストールする場合
 - ダウンロードの手順に進みます、スライド 6 に進んでください

5 / 82

Rのインストール (1-2)

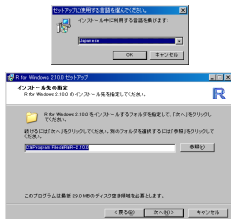
- CRAN (<http://cran.r-project.org/bin/windows/base/>) から、Windows 版 R をダウンロードして実行します



6 / 82

Rのインストール (2)

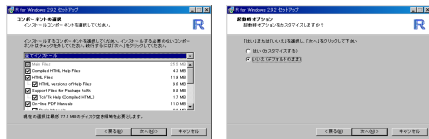
- セットアップ言語を選択し、次へをクリックしていきます
- インストール先指定の画面が出ますので、必要ならば変更してください



7 / 82

Rのインストール (3)

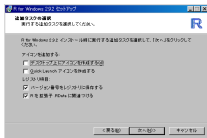
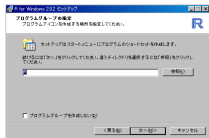
- インストールするコンポーネントを指定します、特に理由がなければ、"利用者向け"か"全て"のどちらかが良いでしょう
- 起動オプションはこのままで、次へをクリックします



8 / 82

Rのインストール (4)

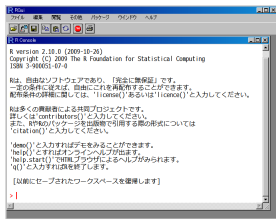
- プログラムグループも特に理由がなければこのままで、次へをクリックします
- 追加タスクの設定では、必要であればデスクトップ/Quick Launch アイコンの作成にチェックし、下の2つにチェックが入っている事を確認し、次へをクリックします
- クリックするとインストールが開始されます、しばらく待っているとインストールが終了します



9 / 82

Rの起動

- スタートメニュー等から"R 2.10.0" (Rバージョン番号) をクリックして起動します



10 / 82

R Commanderのインストール

- Rを起動できたら、次はR Commanderをインストールしていきます
- ローカルファイルからインストールする場合
 - ローカルファイルからインストールする手順に進みます、スライド 12に進んでください
- インターネットからインストールする場合
 - ダウンロードの手順に進みます、スライド 15に進んでください

R Commanderのインストール (1-1)

- パッケージ→ローカルにある... をクリックします



- ファイル (Rcmdr_1.5-3.zip) を選択し、開くをクリックします
- 同様に、car_1.2-16.zip, RcmdrPlugin.EstOR_1.0-1.zip, もインストールします



11 / 82

12 / 82

R Commander のインストール (1-2)

- 次に, "Rcmdr"パッケージをロードします. R Console に

```
library(Rcmdr)
```

と入力するか, パッケージ → パッケージの読み込み → Rcmdr → OK とクリックし



パッケージをロードします

- 追加でパッケージをインストールする必要があるため, OK をクリックしインストールを開始します



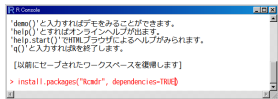
13 / 82

R Commander のインストール (2-1)

- ネットワーク経由でインストールする場合についてです
- R Console に

```
install.packages("Rcmdr", dependencies = TRUE)
```

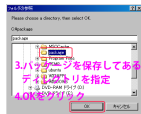
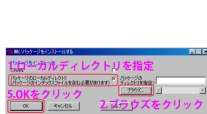
と入力し, Enter キーを押して実行してください
- これは, "Rcmdr"という名前のパッケージをインストールしてください, という意味の命令になっています



15 / 82

R Commander のインストール (1-3)

- まず, ローカルディレクトリを指定します
- ブラウズをクリックし, パッケージが保存してあるディレクトリを指定します
- OK をクリックすると, インストールが開始されます

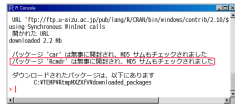


- しばらく待っているとインストールが終了します
- スライド 18 に進んでください

14 / 82

R Commander のインストール (2-2)

- 先ほどの命令を実行すると, ミラーサイトの選択画面が表示されます
- 負荷分散のため, 日本からアクセスする場合は Japan のサイトを選択します
- ミラーサイトを選択すると, 自動的にダウンロードが始まり, パッケージのインストールが行われます



- 正常に終了すると, 「パッケージ 'Rcmdr' は無事に開封され, MD5 サムもチェックされました」と表示されます

16 / 82

R Commander のインストール (2-3)

- 最後に, "Rcmdr"パッケージをロードします. R Console に

```
library(Rcmdr)
```

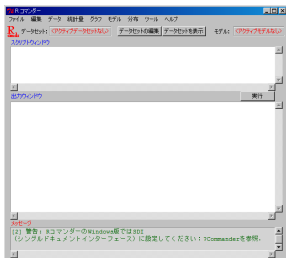
と入力するか, パッケージ → パッケージの読み込み → Rcmdr → OK とクリックし



パッケージをロードします

17 / 82

R Commander の起動画面



18 / 82

R Commander の終了

- ウィンドウの × ボタンで終了できます
- 未保存の解析結果やスクリプトがある場合, 保存するかどうか聞いてくれます
- 終了後におかしなメッセージが出てしまう場合や, もう一度起動しなくなってしまう場合, R Console に

```
detach("package:Rcmdr")
```

と入力し, R Commander を完全に終了させて下さい
終了後にまた起動する場合も

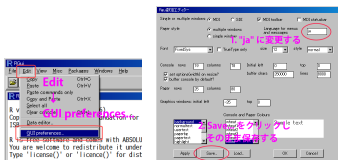
```
library(Rcmdr)
```

を入力します

19 / 82

R の表示を日本語化する

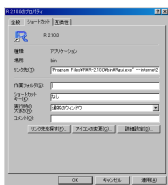
- "Edit" → "GUI preferences..." と選択
- "Language for menu and message" を "ja" に変更
- "Save" をクリックすると "Rconsole" というファイル (R の設定ファイル) を保存する画面が開くので, マイドキュメントに保存する (通常はそのまま保存するとマイドキュメントに保存される)
- R を再起動すると, メニュー画面とメッセージがほぼ日本語化される



20 / 82

パッケージのダウンロードに失敗する場合

- "cran.r-project.org" をポート 80 でコネクトできません"と表示され、インストールに失敗する
- 大学内などから使用し、プロキシの設定が必要な場合
- 起動用ショートカットを右クリックしてプロパティを開き、リンク先の



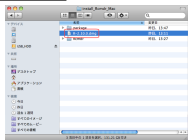
|--internet2

を追加して保存する

Rのインストール (1-1)

Macintosh 版 R (Mac OS X 10.4.4 以上用) のインストール手順です

- ローカルファイルからインストールする場合
 - セットアップファイル "R-2.10.0.dmg" をダブルクリックして実行し、スライド スライド 24 に進んでください



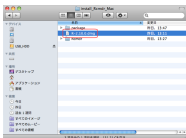
- インターネットからインストールする場合
 - ダウンロードの手順に進みます、スライド 23 に進んでください

21 / 82

22 / 82

Rのインストール (1-2)

- CRAN (<http://cran.r-project.org/bin/macosx/>) から、Macintosh 版 R をダウンロードして実行します



Rのインストール (2)

- Rのインストーラが起動します、ダブルクリックして進みます
- システムの必要条件等が表示されます、"続ける"をクリックします

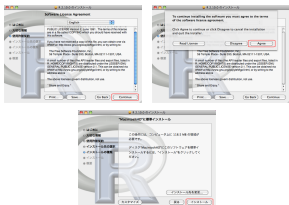


23 / 82

24 / 82

Rのインストール (3)

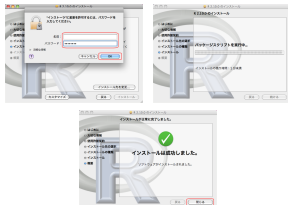
- "使用許諾契約"が表示されます, "Continue", "Agree"をクリックします
- 必要ならインストール先を変更し, "インストール"をクリックします



25 / 82

Rのインストール (4)

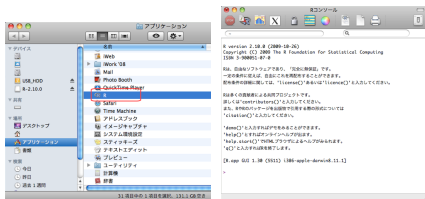
- パスワードの入力を求められるので, 入力してOKをクリックします
- インストール作業を行い, 正常に終了すると「インストールは成功しました。」と表示されます



26 / 82

Rの起動

- Finder 等から "R" をダブルクリックして起動します



27 / 82

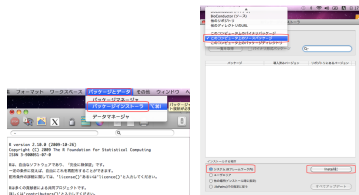
R Commanderのインストール

- R を起動できたら, 次は R Commander をインストールしていきます
- ローカルファイルからインストールする場合
 - ローカルファイルからインストールする手順に進みます, スライド 29に進んでください
- インターネットからインストールする場合
 - ダウンロードの手順に進みます, スライド 34に進んでください

28 / 82

R Commander のインストール (1-1)

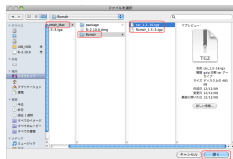
- パッケージとデータ → パッケージインストーラをクリックします
- このコンピュータのソースパッケージ、システムにチェックし、インストールをクリックします



29 / 82

R Commander のインストール (1-2)

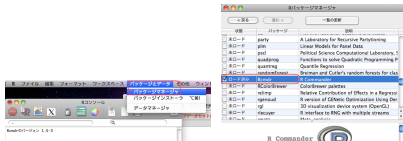
- ファイル選択画面で、"Rcmdr_1.5-3.tgz"を選択し、"開く"をクリックします
- 同様の操作を繰り返し、"car_1.2-16.tgz"を選択し、"開く"をクリックします
- 同様の操作を繰り返し、"RcmdrPlugin.EstOR_1.0-1.tar.gz"を選択し、"開く"をクリックします



30 / 82

R Commander のインストール (1-3)

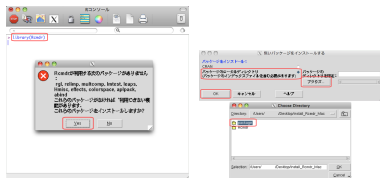
- 次に、"Rcmdr"パッケージをロードします。R Console に
| library(Rcmdr)
と入力するか、パッケージマネージャから Rcmdr をクリックし



パッケージをロードします

R Commander のインストール (1-4)

- 追加でパッケージをインストールする必要があるため、OK をクリックしインストールを開始します
- ローカルディレクトリにチェックし、ブラウズをクリックし、パッケージが保存してあるディレクトリを指定します
- OK をクリックすると、インストールが開始されます

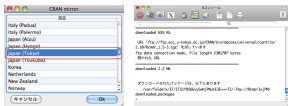


31 / 82

32 / 82

R Commander のインストール (1-5)

- しばらく待っているとインストールが終了します
- スライド 37に進んでください



- しばらくすると、右上のような画面が表示されます

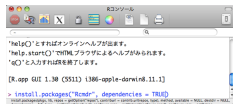
33 / 82

R Commander のインストール (2-2)

- 先ほどの命令を実行すると、ミラーサイトの選択画面が表示されます
- 負荷分散のため、日本からアクセスする場合は Japan のサイトを選択します
- ミラーサイトを選択すると、自動的にダウンロードが始まり、パッケージのインストールが行われます

R Commander のインストール (2-1)

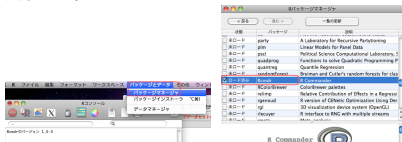
- ネットワーク経由でインストールする場合についてです
- R Console に
`install.packages("Rcmdr", dependencies = TRUE)`
と入力し、Enter キーを押して実行してください
- これは、"Rcmdr"という名前のパッケージをインストールしてください、という意味の命令になっています



34 / 82

R Commander のインストール (2-3)

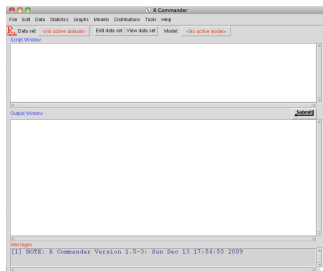
- 最後に、"Rcmdr"パッケージをロードします。R Console に
`library(Rcmdr)`
と入力するか、パッケージマネージャから Rcmdr をクリックし



- パッケージをロードします

35 / 82

36 / 82



41 / 82

使用するデータについて

表 2. リンパ節転移で層別したタモキシフェン使用と乳がんの再発

タモキシフェン	リンパ節転移あり			リンパ節転移なし		
	再発あり	再発なし	対象者数	再発あり	再発なし	対象者数
使用	368	847	1215	96	1238	1334
非使用	253	507	760	171	1421	1592
合計	621	1354	1975	267	2659	2926
	再発割合の差 = -3%			再発割合の差 = -3.5%		

- 再発割合 (リンパ節転移あり)
タモキシフェン使用 368/1215=30.3%, 非使用 253/760=33.3%
- 再発割合 (リンパ節転移なし)
タモキシフェン使用 464/2549=7.2%, 非使用 424/2352=10.7%

使用するデータについて

- Matsuyama Y, et al. Second cancers after adjuvant tamoxifen therapy for breast cancer in Japan. Ann Oncol. 2000; 11: 1537-1543.

1982年から1990年にかけて日本の9施設で手術を受けた乳がん患者4901名について、手術後にタモキシフェンという女性ホルモンによく似た乳がんの治療薬を使用したかどうかと、その後の乳がんの再発との関係を1996年まで追跡して調べた結果の一部である (佐藤俊雄, 交絡事実と反事実の比較, 岩波科学 2008年4月号, より引用)

表 1. タモキシフェン使用と乳がんの再発

タモキシフェン	再発あり	再発なし	対象者数
使用	464	2085	2549
非使用	424	1928	2352
合計	888	4013	4901
	再発割合の差 = 0.2%		

- 再発割合
タモキシフェン使用 464/2549=18.2%, 非使用 424/2352=18.0%

42 / 82

プラグインのロード

- オッズ比の区間推定のためのプラグイン"RcmdrPlugin.EstOR"をロードしておきます
- R Commanderの画面から、ツール → Rcmdr プラグインのロードを選択します
- プラグイン"RcmdrPlugin.EstOR"を選択し、OKをクリックします
- 再起動の指示が出ますので、OKをクリックします



43 / 82

44 / 82

データの読み込み 1 (Windows の場合)

- "tamoxifen.xls" というファイルを使います
- データ → データのインポート → Excel または... をクリックします



45 / 82

データの読み込み 3 (Windows の場合)

- 表 (シート) の選択画面が出ますので, "data" を選択し, OK をクリックします
- 以上で, データの読み込みが完了します



- Excel ファイル以外にも, R のデータファイル, Access や他の統計解析パッケージのデータファイルも読み込むことができます

47 / 82

データの読み込み 2 (Windows の場合)

- データセット名を指定できます, 今回は "Dataset" のまま OK をクリックします
- ファイルの選択画面が出ますので, "tamoxifen.xls" を選択し, 開くをクリックします



46 / 82

データの読み込み 1 (Macintosh の場合)

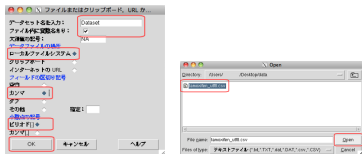
- "tamoxifen_utf8.csv" というファイルを使います
- データ → データのインポート → テキストファイルまたは... をクリックします



48 / 82

データの読み込み 2 (Macintosh の場合)

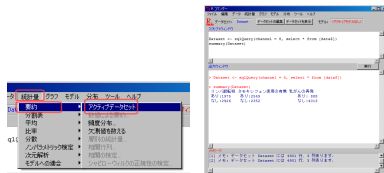
- データ名は"Dataset", データファイルの場所は"ローカルファイルシステム", フィールドの区切り記号は"カンマ", 小数点の記号は"ピリオド"にチェックが入っている事を確認します
- OK をクリックすると、ファイルの選択画面が開くので、"tamoxifen_utf8.csv" を指定します
- 以上で、データの読み込みが完了します



49 / 82

要約統計量の計算

- 各変数の要約統計量を計算します
- 統計量 → 要約 → アクティブデータセットを選択します
- 連続変数の場合の要約統計量は、最小値, 25%点, 中央値, 平均値, 75%点, 最大値を出力します。カテゴリ変数の場合の要約統計量は、各水準の度数を出力します



50 / 82

要約統計量の計算結果

- このような結果が出力されます

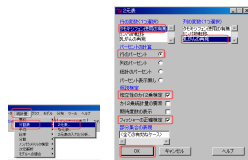
```
> summary(Dataset)
リンパ節転移 タモキシフェン使用の有無 乳がんの再発
あり:1975   あり:2349   あり: 888
なし:2926   なし:2352   なし:4013
```

- 連続変数の場合

```
> summary(BOD)
      Time      demand
Min.   11.000   Min.    8.30
1st Qu. 12.250   1st Qu. 11.42
Median 13.500   Median 13.00
Mean   13.667   Mean   14.03
3rd Qu. 14.750   3rd Qu. 15.25
Max.   17.000   Max.   19.00
```

二元分割表の作成

- タモキシフェン使用の有無と、乳がんの再発をクロス集計した、二元分割表を作成します
- オプションで仮説検定を指定できます
- 統計量 → 分割表 → 2元表を選択します
- 行変数: タモキシフェン使用の有無, 列変数: 乳がんの再発, 行のパーセント, 検定のオプションを選択し, OK をクリックします



51 / 82

52 / 82

二元分割表の計算結果

- 表 1 と同じ結果が得られます

```
> .Table <- xtabs(~タモキシフェン使用の有無+乳がんの再発, data=dataset)
> .Table
      乳がんの再発
タモキシフェン使用の有無 あり なし
あり      464 2085
なし      424 1928

> rowPercentages(.Table) # Row Percentages
      乳がんの再発
タモキシフェン使用の有無 あり なし Total Count
あり 18.2 81.8 100 2549
なし 18.0 82.0 100 2352
```

表 1. タモキシフェン使用と乳がんの再発

タモキシフェン	再発あり	再発なし	対象者数
使用	464	2085	2549
非使用	424	1928	2352
合計	888	4013	4901
	再発割合の差 = 0.2%		

53 / 82

二元分割表の計算結果

- 再発割合が異なるかどうかを、ピアソンのカイニ乗検定を用いて検定します ($\alpha = 0.05$ とする)

$$\begin{cases} H_0: & \text{タモキシフェン使用の有無で再発割合は変わらない} \\ H_1: & \text{タモキシフェン使用の有無で再発割合が異なる} \end{cases}$$

```
> .Test <- chisq.test(.Table, correct=FALSE)
> .Test
      Pearson's Chi-squared test

data: .Table
X-squared = 0.0255, df = 1, p-value = 0.873
```

- P -value=0.873 であり、 $\alpha = 0.05$ のもとで帰無仮説は棄却できませんでした

54 / 82

二元分割表の計算結果

- 再発オッズ比は 1 であるかどうかを、Fisher の正確検定を用いて検定します ($\alpha = 0.05$ とする)

$$\begin{cases} H_0: & \text{タモキシフェン使用の有無に対する再発オッズ比は 1 である} \\ H_1: & \text{タモキシフェン使用の有無に対する再発オッズ比は 1 でない} \end{cases}$$

```
> fisher.test(.Table)
      Fisher's Exact Test for Count Data

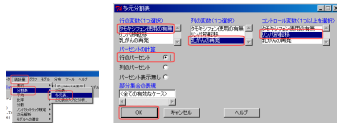
data: .Table
p-value = 0.882
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.072582 1.173706
sample estimates:
 odds ratio |
 1.011935
```

- P -value=0.882 であり、 $\alpha = 0.05$ のもとで帰無仮説は棄却できませんでした
- オッズ比の推定値は 1.01 で、95%信頼区間は [0.87, 1.17] でした

55 / 82

多元分割表の作成

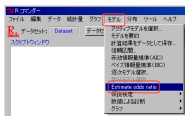
- タモキシフェン使用の有無と、乳がんの再発の分割表を、リンパ節転移の有無で層別した、多元分割表を作成します
- 統計量 → 分割表 → 多元表を選択します
- 行変数: タモキシフェン使用の有無, 列変数: 乳がんの再発, コントロール変数: リンパ節転移, 行のパーセントを選択し、OK をクリックします



56 / 82

ロジスティック回帰モデルによる単変量解析と区間推定

- モデル → Estimate odds ratio とクリックします



- 推定するためには、プラグイン"RcmdrPlugin.EstOR"をロードしておく必要があります

ロジスティック回帰モデルによる単変量解析と区間推定 (解析結果)

- 先ほどの結果からオッズ比の点推定と区間推定を行った結果が得られます

```
> EstOR.estimate(GLM.1)
95% Wald confidence interval
(Intercept)
OR 2.5% 97.5%
4.494 4.063 4.969

95% Wald confidence interval
タモキシフェン使用の有無 [T.なし]
OR 2.5% 97.5%
1.012 0.875 1.170
```

- タモキシフェン使用の再発オッズ比の推定値は 1.01 で、95%信頼区間は [0.87, 1.17] でした

61 / 82

62 / 82

単変量解析の結果から

表 1. タモキシフェン使用と乳がんの再発

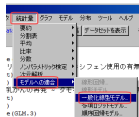
タモキシフェン	再発あり	再発なし	対象者数
使用	464	2085	2549
非使用	424	1928	2352
合計	888	4013	4901
	再発割合の差 = 0.2%, 再発オッズ比 1.01		

- 再発割合の差は非常に小さく、再発オッズ比も 1 に近かった
- 仮説検定の結果や、区間推定の結果から、統計的に意味がある差も見られなかった

タモキシフェンの乳がん再発予防効果はない？

ロジスティック回帰モデルによる多変量解析と区間推定 1

- タモキシフェン使用の有無と、乳がんの再発の関係を、ロジスティック回帰モデルにより解析します
- 統計量 → モデルへの適合 → 一般化線型モデルを選択します



63 / 82

64 / 82

どちらが正しいのか

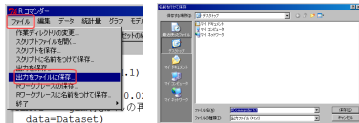
- 表 2 をみると、リンパ節転移がある場合、タモキシフェンの使用が多く、リンパ節転移がない場合は逆に少なくなっている
- リンパ節転移が再発を引き起こす原因になっているとしたら、タモキシフェンを使用した集団に、再発しやすい人が沢山含まれていることになる (実際、リンパ節転移は再発のリスク因子)
- こういう場合に単変量解析を行ってしまうと、再発割合の差はおかしな事になる (今回は同じぐらいになってしまった)
- リンパ節転移の様な因子は交絡因子であり、交絡因子を調整しない推定結果は正しくない
- 多変量解析を用いて、交絡因子の状態を仮想的に同じとした場合の、タモキシフェン使用の再発予防効果をみるべき

タモキシフェンの乳がん再発予防効果はある

69 / 82

解析結果の保存

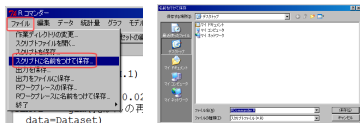
- ファイル→出力をファイルに保存とクリックします
- ファイル名を指定して、保存をクリックします



70 / 82

スクリプトの保存

- ファイル→スクリプトに名前を付けて保存とクリックします
- ファイル名を指定して、保存をクリックします

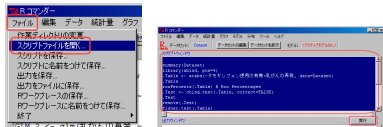


- スクリプトファイルは今まで解析した内容を記録したファイルです
- データ (名前は同じにする) とプラグインの読み込みをしておけば、スクリプトウィンドウの内容を Ctrl+A 等で選択し、実行ボタンを押すことでマウス操作を省略して解析結果が得られます

71 / 82

スクリプトの実行

- ファイル→スクリプトファイルを開くとクリックします
- ファイル名を指定して、開くをクリックします (画像省略)
- スクリプトウィンドウの内容を全て選択し、実行ボタンを押します



- スクリプトファイルに記録された処理を実行し、結果を出力してくれます

72 / 82

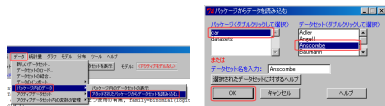
- R のセットアップ+R 入門
<http://cwoweb2.bai.ne.jp/~jgb11101/files/cart/cart.html>
【R 入門】第 6 回: R Commander 入門
<http://www.ocn.naq.ne.jp/cuhxr802/R-intro-06.pdf>
 統計解析フリーソフト R 入門 [GUI 版 R] R Commander の説明書
<http://cwoweb2.bai.ne.jp/~jgb11101/files/R-commander.pdf>
 舟尾暢男氏 (武田薬品工業 (株)) のホームページ
- The R Commander: A Basic-Statistics GUI for R
<http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/>
 R Commander の公式ページ, 英語だが有益な情報が多数公開されている
- R-Tips (中央農業研究センター 竹澤邦夫 先生)
<http://cse.naro.affrc.go.jp/takezawa/r-tips/r.html>
- R による統計処理 (群馬大学 社会情報学部 青木繁伸 先生)
<http://aoki2.si.gunma-u.ac.jp/R/>
- 統計処理ソフトウェア R についての Tips (群馬大学 大学院医学系研究科 中澤港 先生)
<http://phi.med.gunma-u.ac.jp/swtips/R.html>
- 生態学のデータ解析 (北海道大学 大学院環境科学院・地球環境科学研究所 久保拓弥 先生)
<http://hoshoees.hokudai.ac.jp/~kubo/ce/FrontPage.html>

73 / 82

データ"Anscombe"について

- 1970 年のアメリカの公立高校における支出額調査 (Anscombe FJ. *Computing in Statistical Science Through APL*. Springer-Verlag, 1981.)
- education
一人あたりの教育支出額 (\$)
- income
一人あたりの収入 (\$)
- young
18 歳未満の割合 (1000 人あたり)
- urban
都会の割合 (1000 あたり); 詳細は不明 (人口が多い都市の割合?)

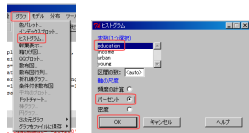
- データ → パッケージ内のデータ → アタッチされた... を開くとクリックします
- car をダブルクリック → Anscombe をダブルクリックし, OK をクリックします
- これでデータの読み込みが完了します



74 / 82

ヒストグラムの作成

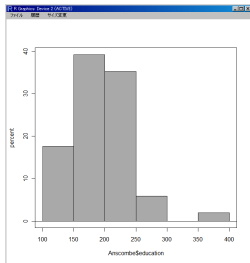
- グラフ → ヒストグラムとクリックします
- グラフを描く変数と, 軸の尺度を選び, OK をクリックします



75 / 82

76 / 82

ヒストグラム



箱ひげ図の作成

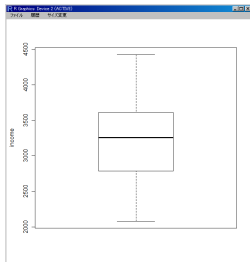
- グラフ → 箱ひげ図をクリックします
- グラフを描く変数を選び、OKをクリックします
- はずれ値を表示するオプションもあり、グループ変数で層別した箱ひげ図も描くことができます



77 / 82

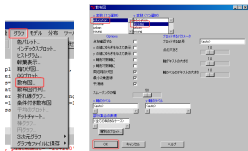
78 / 82

箱ひげ図



散布図の作成

- グラフ → 散布図をクリックします
- x 軸 (横軸) の変数, y 軸 (縦軸) の変数, いくつかのオプションを選び、OKをクリックします



79 / 82

80 / 82

