

Rで統計解析入門

(番外篇) R Commander を用いた解析



本日のメニュー

1. R のセットアップ (Ver. 2.12.2) のメモ
2. R Commander の基礎
 - ▶ イントロ
 - ▶ データの読み込みと変数に対する処理
 - ▶ データのグラフ化
 - ▶ データ解析
3. パッケージによる機能追加



R のセットアップ

- ▶ CRAN（筑波大学）からダウンロード

<http://cran.md.tsukuba.ac.jp/bin/windows/base/old/2.12.2/R-2.12.2-win.exe>

R-2.12.2 for Windows (32/64 bit)

[Download R 2.12.2 for Windows](#) (37 megabytes, 32/64 bit) ここ

[Installation and other instructions](#)

New features in this version: [Windows specific](#), [all platforms](#).

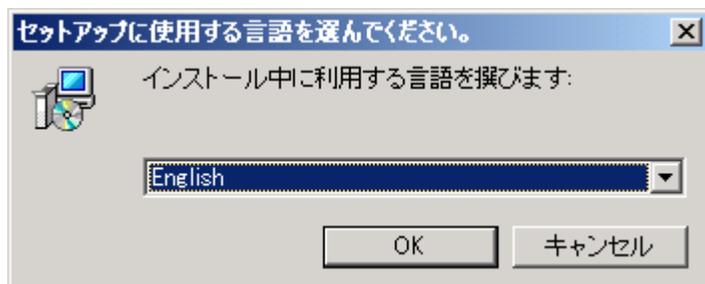
If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the [md5sum](#) of the .exe to the [true fingerprint](#). You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

- ▶ ダウンロードしたファイル R-2.12.2-win.exe をダブルクリック
（Vista / 7 の方は「右クリック 管理者権限として実行」

R-2.13.0 以降では日本語環境での不具合報告があるため、本資料では R-2.12.2 を使用する

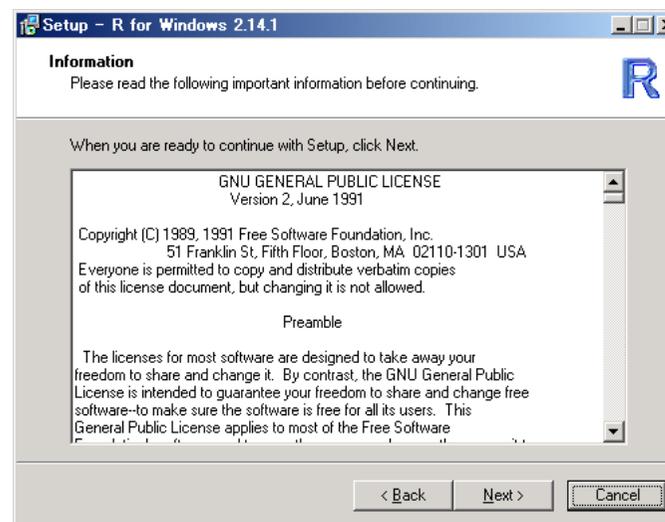
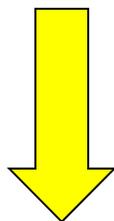


R のセットアップ



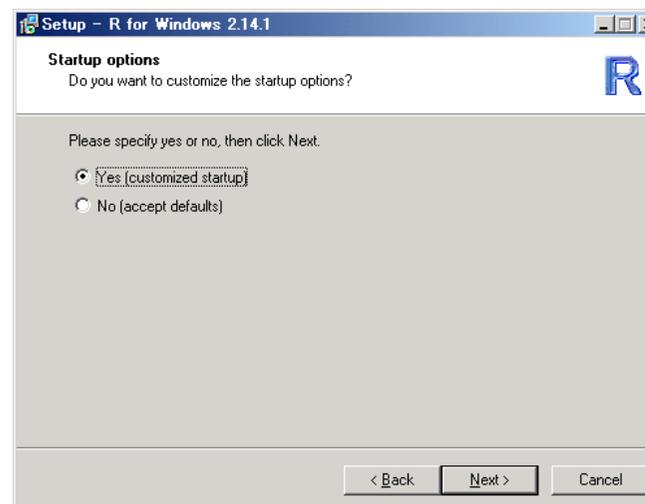
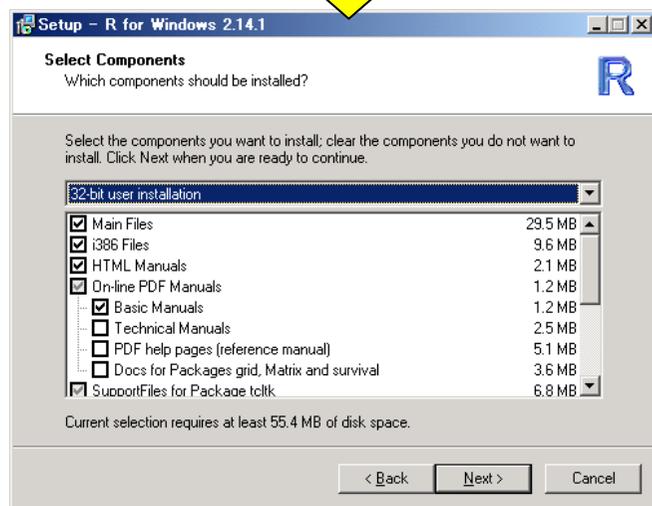
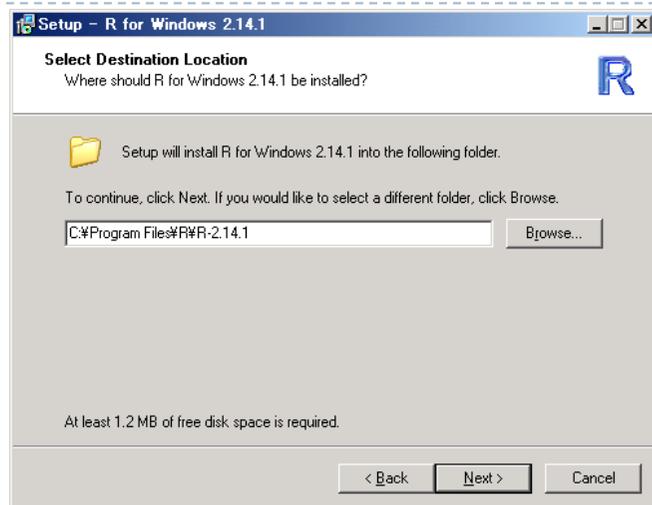
インストール中に使用する言語を
「English」に変更！！
(何もしないと文字化けします)

以降は「Next >」をクリックし続ける





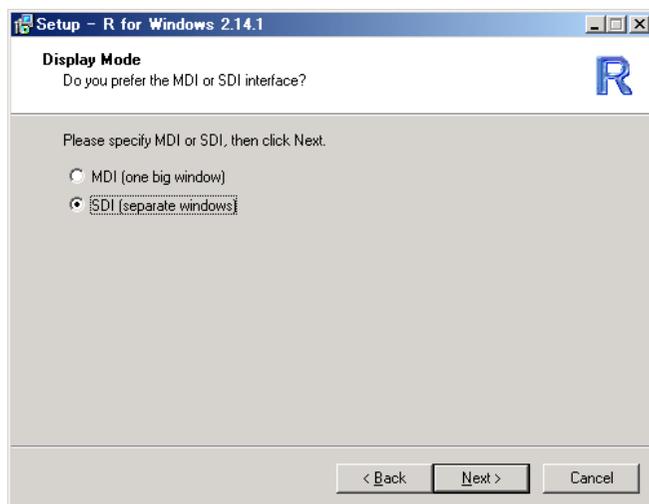
R のセットアップ



「Yes」を選択（カスタマイズする）



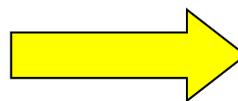
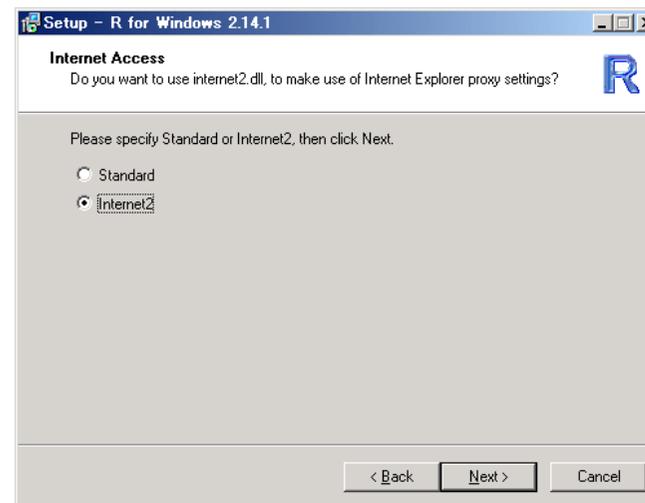
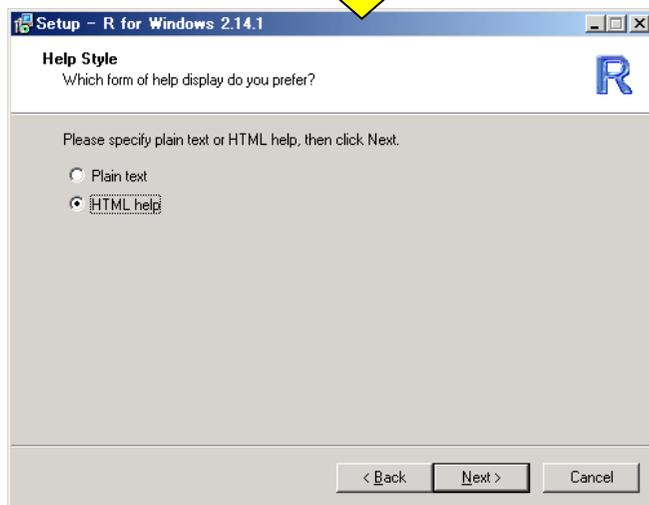
R のセットアップ



「SDI（各ウィンドウを分離）」

「ヘルプは HTML 形式」

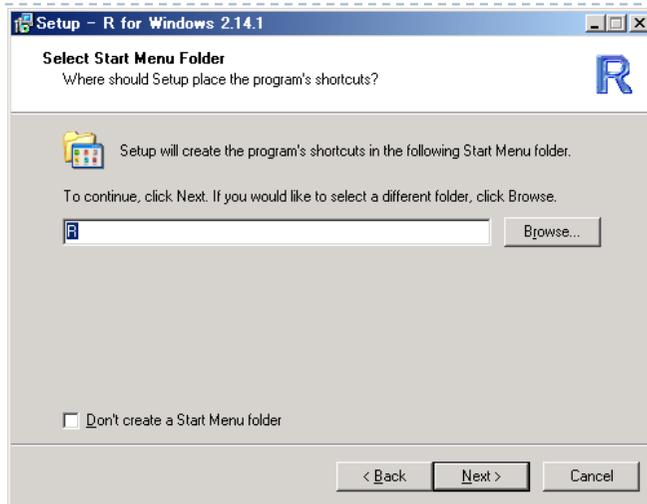
がお勧め



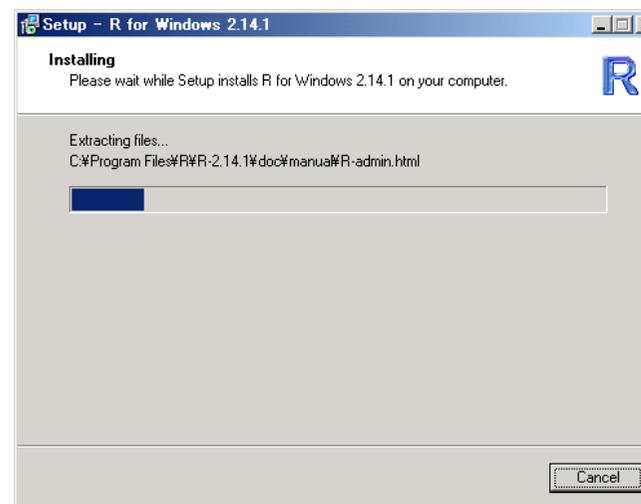
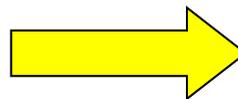
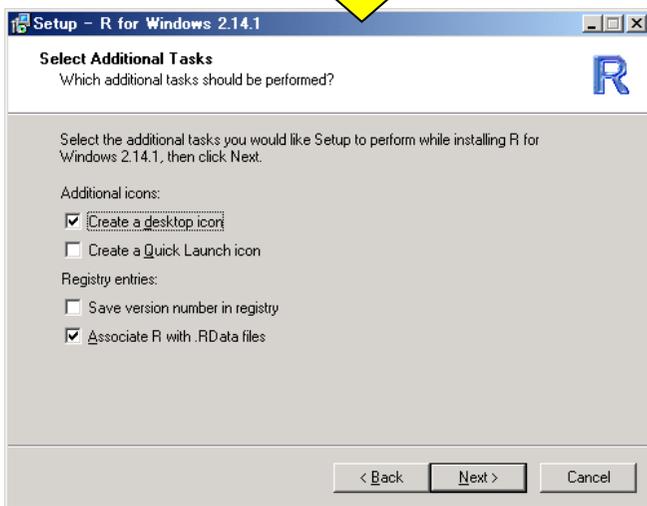
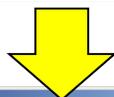
会社で R を使っている場合は
「internet2」を選択しておく



R のセットアップ



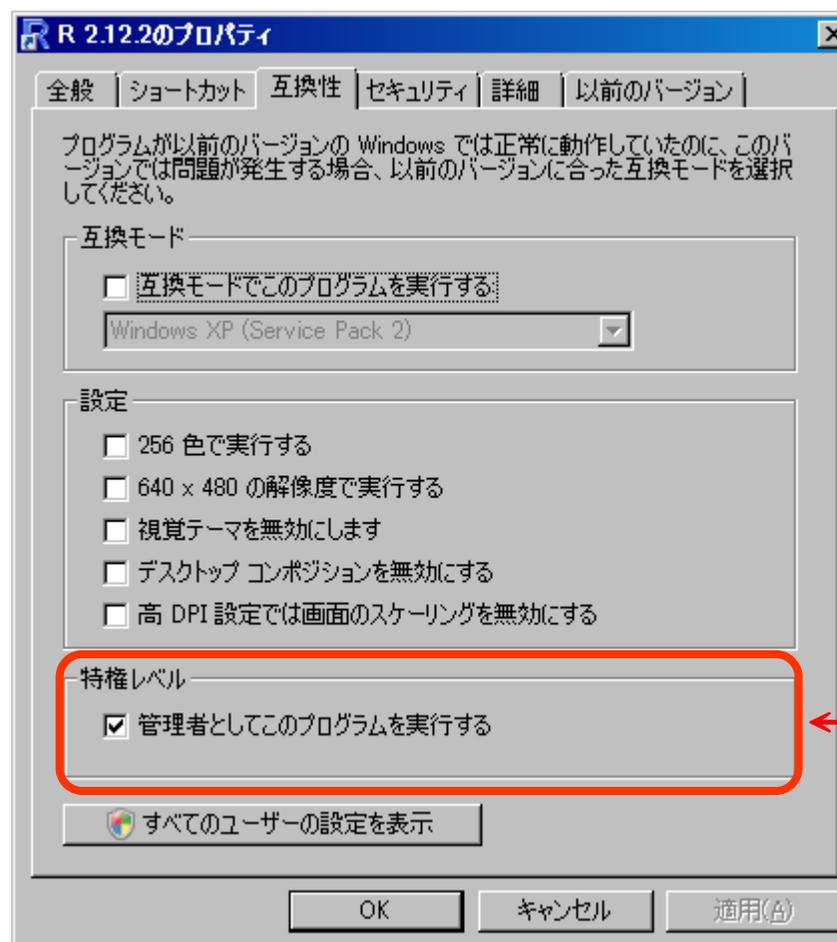
後は「Next >」をクリックし続けると
インストールが完了する





R のセットアップ

- ▶ R のショートカットのここをチェックすると起動が便利

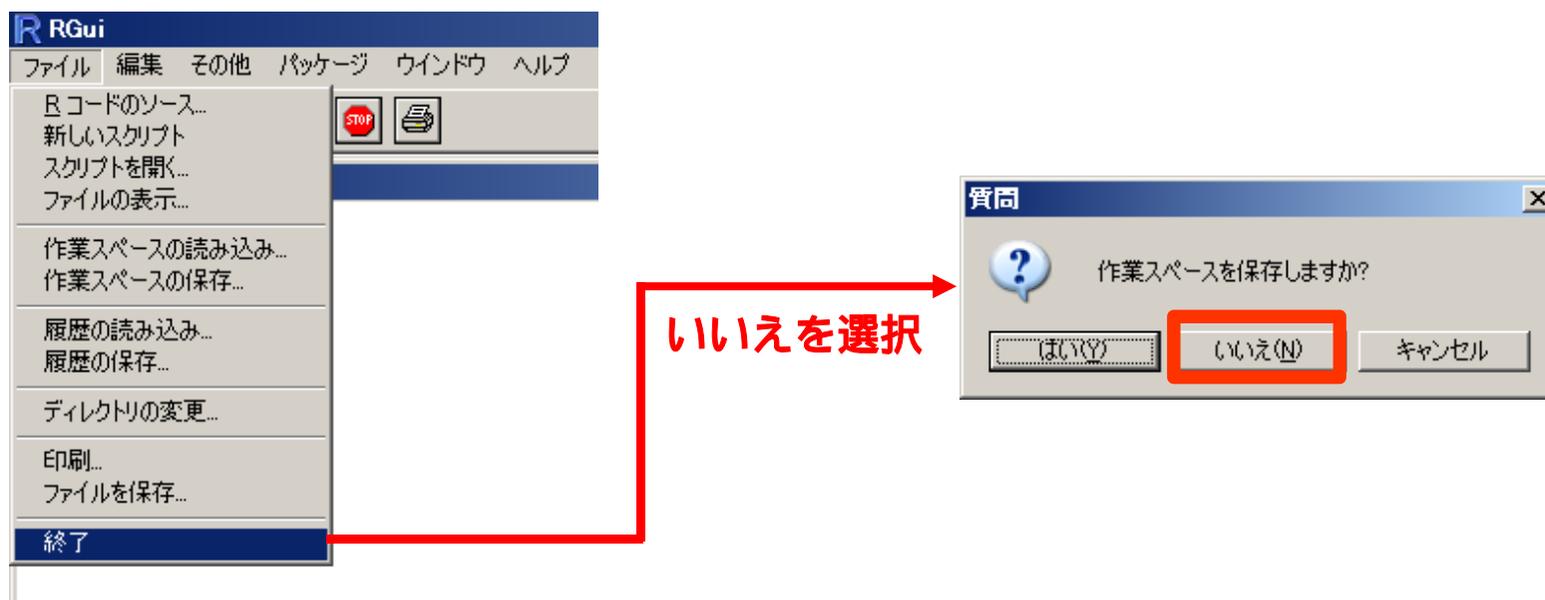




R Commander のセットアップ

- ▶ R を起動して以下を実行し、動作が落ち着いたら R を終了

```
> options(repos="http://cran.md.tsukuba.ac.jp")  
> install.packages("Rcmdr", dep=T)
```



▶ 9

R 起動時は「管理者として実行」する（前頁の作業が済んでいる場合は気にしなくて良い）
何故か既に R Commander が入っている場合は `remove.packages("Rcmdr", .libPaths())` で削除



本日のメニュー

1. R のセットアップ (Ver. 2.12.2) のメモ
2. R Commander の基礎
 - ▶ イントロ
 - ▶ データの読み込みと変数に対する処理
 - ▶ データのグラフ化
 - ▶ データ解析
3. パッケージによる機能追加



Rとは？

- ▶ オープンソース&フリーの統計解析用ソフト

【長所】

- ▶ 関数電卓, 数値計算, プログラミング, 統計解析, グラフィックスの機能があり, どの機能も充実している
- ▶ 機能拡張が容易に行える
- ▶ 使用人口が多いので, バグが少なく情報も豊富

【短所】

- ▶ EXCEL などの表計算ソフトに比べて GUI (マウス操作) の機能が劣っている (R の命令をひとつひとつ覚えなければいけない...)
- ▶ 大規模なデータを扱う場合は多少骨が折れる

GUI 版 R である R Commander の登場!



R Commander とは？

- ▶ John Fox 教授（カナダ McMaster 大学）が開発した GUI 版 R のこと
- ▶ マウス操作で R を使うことが出来る
（それほど R の命令を覚えなくても R の出力が得られる）
- ▶ 2005 年頃より，関西大学の荒木孝治先生が主体となって R Commander のメッセージ翻訳がなされ...
- ▶ R Commander はバージョン 1.1-1 より本格的に日本語化された！
 - ▶ 最新版はバージョン 1.8-3 だが，日本語環境でのバグが多いかも
 - ▶ 「日本語環境でのバグが少ない」ものでの最新のバージョンは 1.6-3
- ▶ EZR という亜種も登場！
（自治医科大学附属さいたま医療センター・神田先生作）



R Commander の起動

- ▶ R のアイコン  をクリック or スタートメニューから起動
- ▶ Vista / 7 をお使いの方は、アイコンを右クリックして「管理者権限として実行」を選択して起動してください

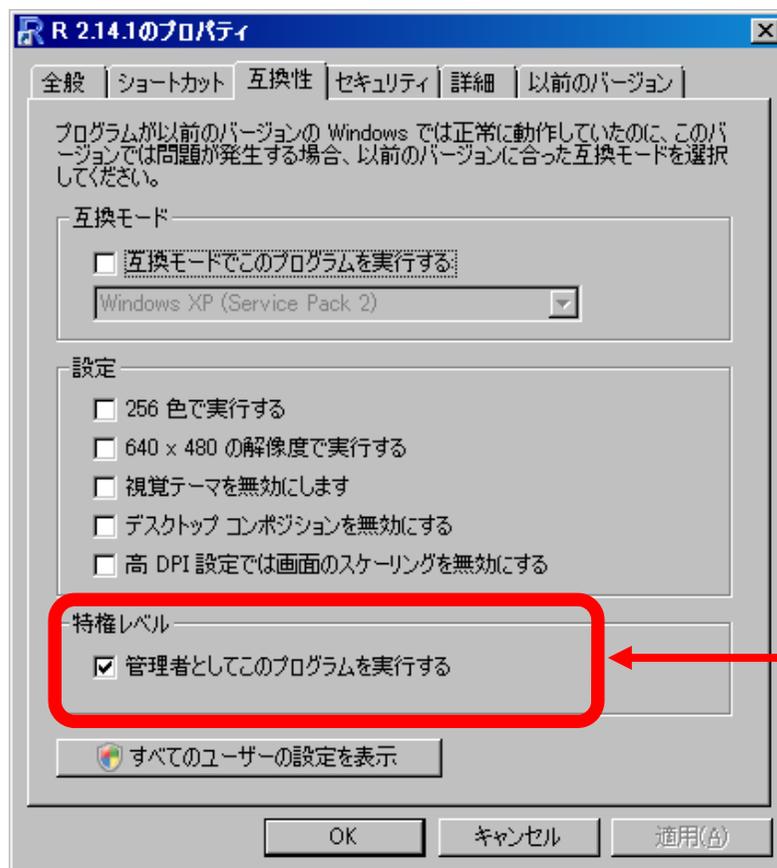


クリック！



R Commander の起動

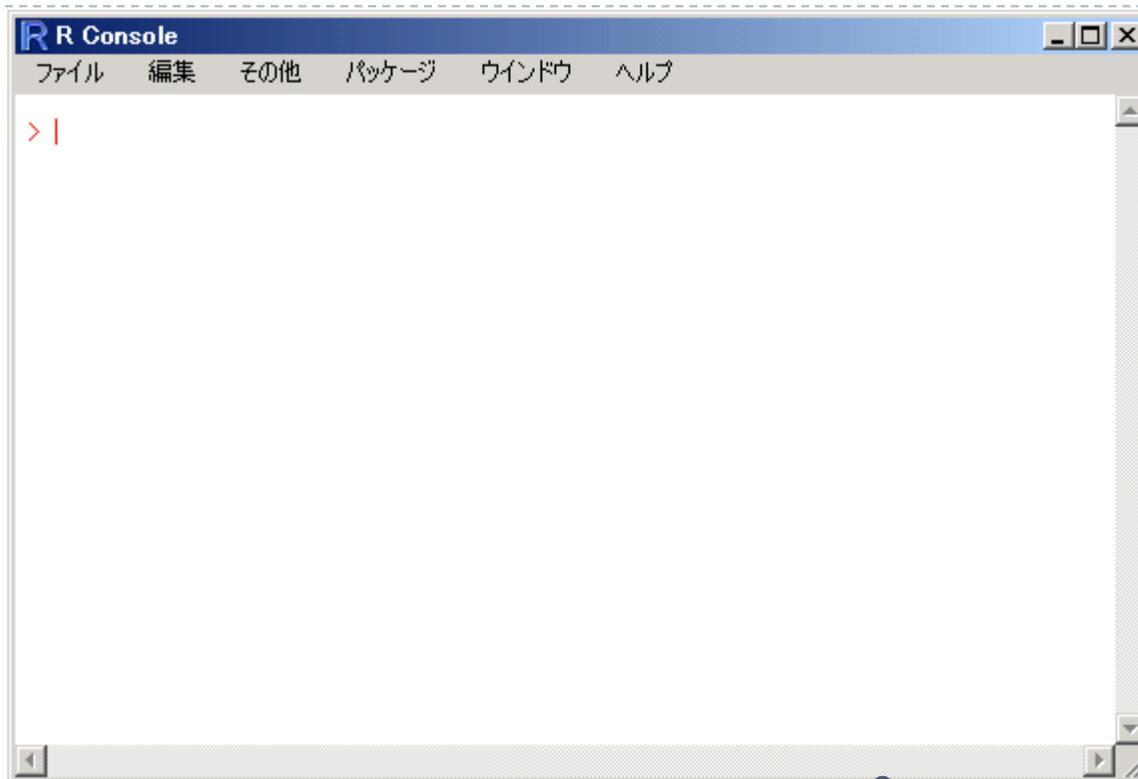
- ▶ Vista / 7 をお使いの方は、アイコン  を右クリック プロパティから「互換性」 「管理者として...」をチェックしておく则毎回の起動が楽



チェック！



R Console から R Commander を起動する場合



初回 : `library(Rcmdr)`
2 回目以降 : `Commander()`



R Commander の起動

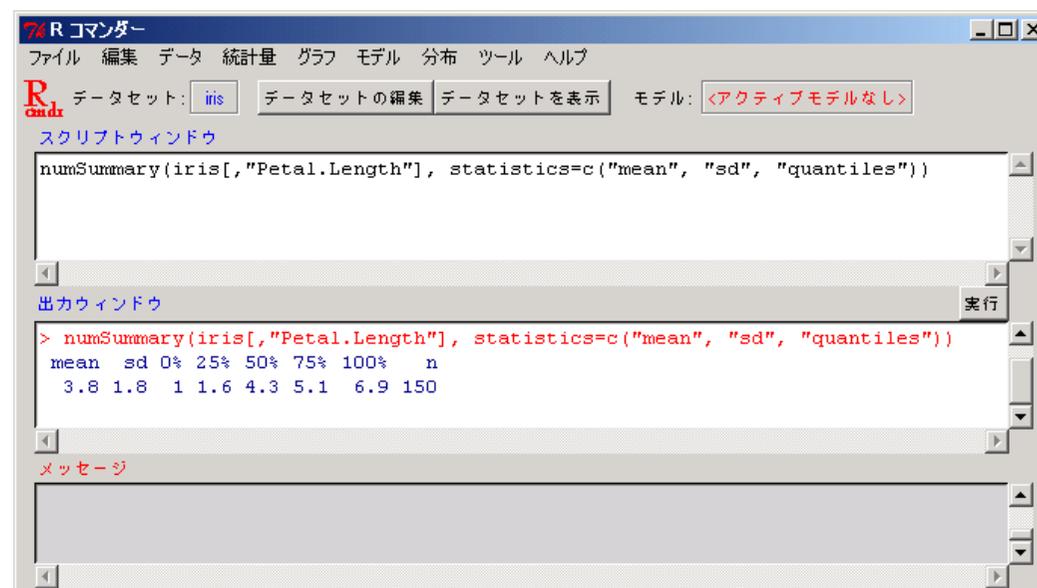
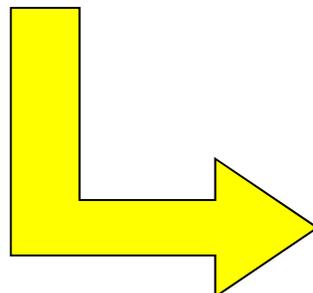
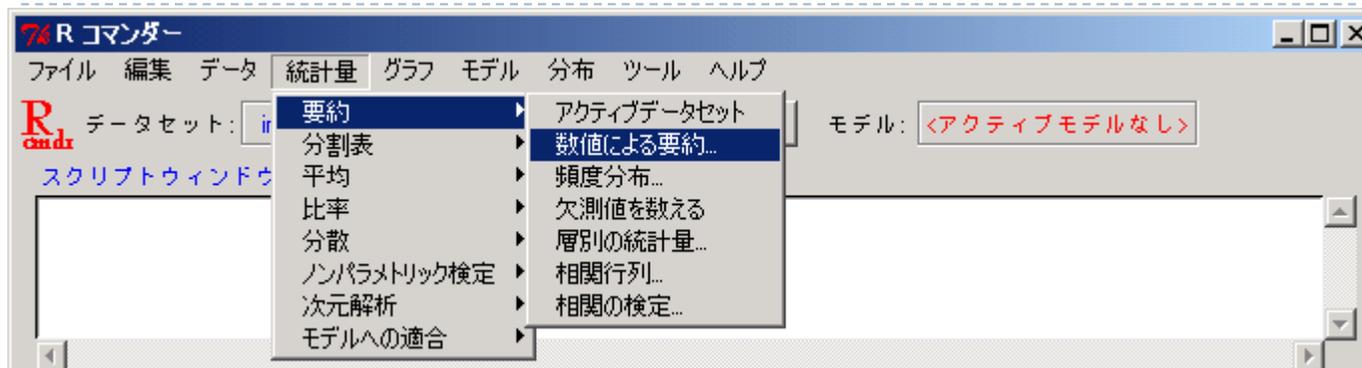
The image shows two windows from the R Commander application. The top window is titled "R Console" and contains a simple prompt "> |". The bottom window is titled "R コマンダー" and is the main interface. It features a menu bar with options like "ファイル", "編集", "データ", "統計量", "グラフ", "モデル", "分布", "ツール", and "ヘルプ". Below the menu bar, there are fields for "データセット:" (currently showing "<アクティブデータセットなし>") and "モデル:" (currently showing "<アクティブモデルなし>"). There are also buttons for "データセットの編集" and "データセットを表示". The main area is divided into three sections: "スクリプトウィンドウ" (empty), "出力ウィンドウ" (empty), and "メッセージ" (containing the message "[1] メモ: Rコマンドーのバージョン 1.6-3: Sun Apr 01 10:34:34 2012").

← 最小化する

メインはこれ →



R Commander でデータ解析を行う流れ



- ① メニューから機能を選択
- ② スクリプトウィンドウには実行した R のコマンドが出力される
- ③ 出力ウィンドウには、実行結果が出力される
- ④ メッセージにはエラーや警告が出力される



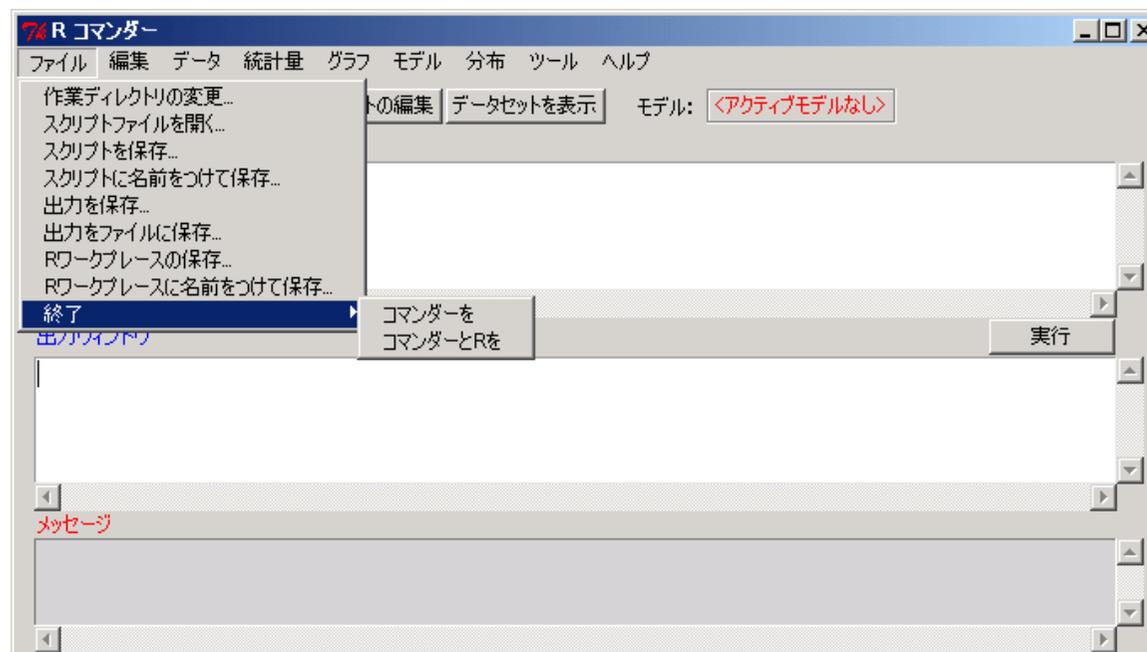
R Commander の機能一覧

- ▶ ファイル操作
- ▶ データの入出力, データの編集
- ▶ 統計量の算出, 検定の実行
- ▶ 様々なグラフ描画
- ▶ モデル解析
- ▶ 分位点, 確率点の算出, 分布のプロット, 乱数生成
- ▶ 実行環境の設定機能, ヘルプ機能
- ▶ R Commander への機能追加



R Commander の終了

- ▶ R または R Commander の右上の [×] 印をクリックする
- ▶ メニューから選択する
- ▶ 関数 `q()` または `quit()` を実行する





本日のメニュー

1. R のセットアップ (Ver. 2.12.2) のメモ

2. R Commander の基礎

- ▶ イントロ

- ▶ データの読み込みと変数に対する処理

- ▶ データのグラフ化

- ▶ データ解析

3. パッケージによる機能追加



架空のデータ「DEP」

- ▶ データ「DEP」は以下からダウンロード可能（要解凍）
<http://www.cwk.zaq.ne.jp/fkhud708/files/R-cmdr-intro.zip>
- ▶ うつ病を患っている患者さんに薬剤治療を行った後，QOLの点数を測定
- ▶ QOL（Quality of Life；生活の質）の点数：以下の架空のアンケート票を使って患者さんに回答してもらい，各質問項目で回答した番号を合計したものを当該患者さんの点数とする

No	質問	当てはまらない (1点)	あまり当てはまらない (2点)	やや当てはまる (3点)	当てはまる (4点)
1	起床時に気分が良い				
2	朝食は美味しい				
3	学校/会社に行きたい				
:	:	:	:	:	:



架空のデータ「DEP」の変数

- ▶ **GROUP** : 薬剤の種類 (A, B, C)
- ▶ **QOL** : QOL の点数 (数値) 点数が大きい方が良い
- ▶ **EVENT** : 改善の有無 (1 : 改善あり, 2 : 改善なし)
 QOLの点数が5点以上である場合を「改善あり」とする
- ▶ **DAY** : 観察期間 (数値, 単位は日)
- ▶ **PREDRUG** : 前治療薬の有無 (YES : 他の治療薬を投与したことあり,
 NO : 投与したことなし)
- ▶ **DURATION** : 罹病期間 (数値, 単位は年)



架空のデータ「DEP」

GROUP	QOL	EVENT	DAY	PREDRUG	DURATION
A	15	1	50	NO	1
A	13	1	200	NO	3
A	11	1	250	NO	2
A	11	1	300	NO	4
A	10	1	350	NO	2
A	9	1	400	NO	2
A	8	1	450	NO	4
A	8	1	550	NO	2
A	6	1	600	NO	5
A	6	1	100	NO	7
A	4	2	250	NO	4
A	3	2	500	NO	6
A	3	2	750	NO	3
A	3	2	650	NO	7
A	1	2	1000	NO	8
A	6	1	150	YES	6
A	5	1	700	YES	5
A	4	2	800	YES	7
A	2	2	900	YES	12
A	2	2	950	YES	10
B	13	1	380	NO	9
B	12	1	880	NO	5
B	11	1	940	NO	2
B	4	2	20	NO	7
B	4	2	560	NO	2
B	5	1	320	YES	11
B	5	1	940	YES	3
B	4	2	80	YES	6
B	3	2	140	YES	7
B	3	2	160	YES	13

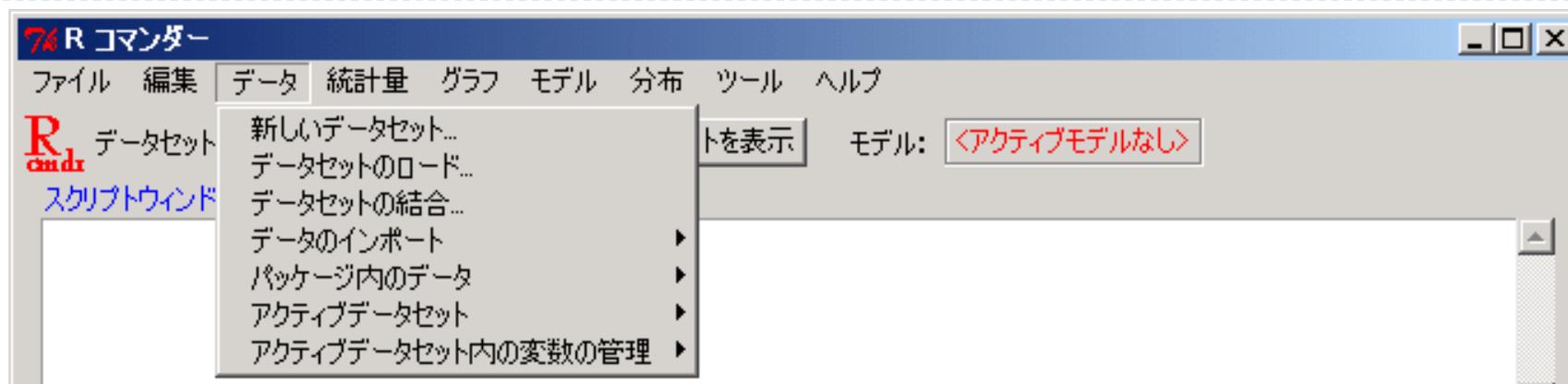


架空のデータ「DEP」

GROUP	QOL	EVENT	DAY	PREDRUG	DURATION
B	3	2	240	YES	15
B	2	2	280	YES	9
B	2	2	440	YES	8
B	2	2	520	YES	7
B	2	2	620	YES	9
B	2	2	740	YES	8
B	2	2	860	YES	2
B	1	2	880	YES	10
B	0	2	920	YES	8
B	0	2	960	YES	4
C	9	1	170	NO	1
C	7	1	290	NO	4
C	5	1	430	NO	2
C	3	2	610	NO	4
C	2	2	110	NO	5
C	2	2	410	NO	2
C	1	2	530	NO	7
C	1	2	580	NO	2
C	0	2	810	NO	3
C	0	2	990	NO	10
C	6	1	30	YES	1
C	5	1	830	YES	6
C	3	2	70	YES	16
C	2	2	310	YES	9
C	2	2	370	YES	18
C	1	2	490	YES	7
C	1	2	690	YES	10
C	0	2	730	YES	3
C	0	2	770	YES	12
C	0	2	910	YES	8



メニュー〔データ〕

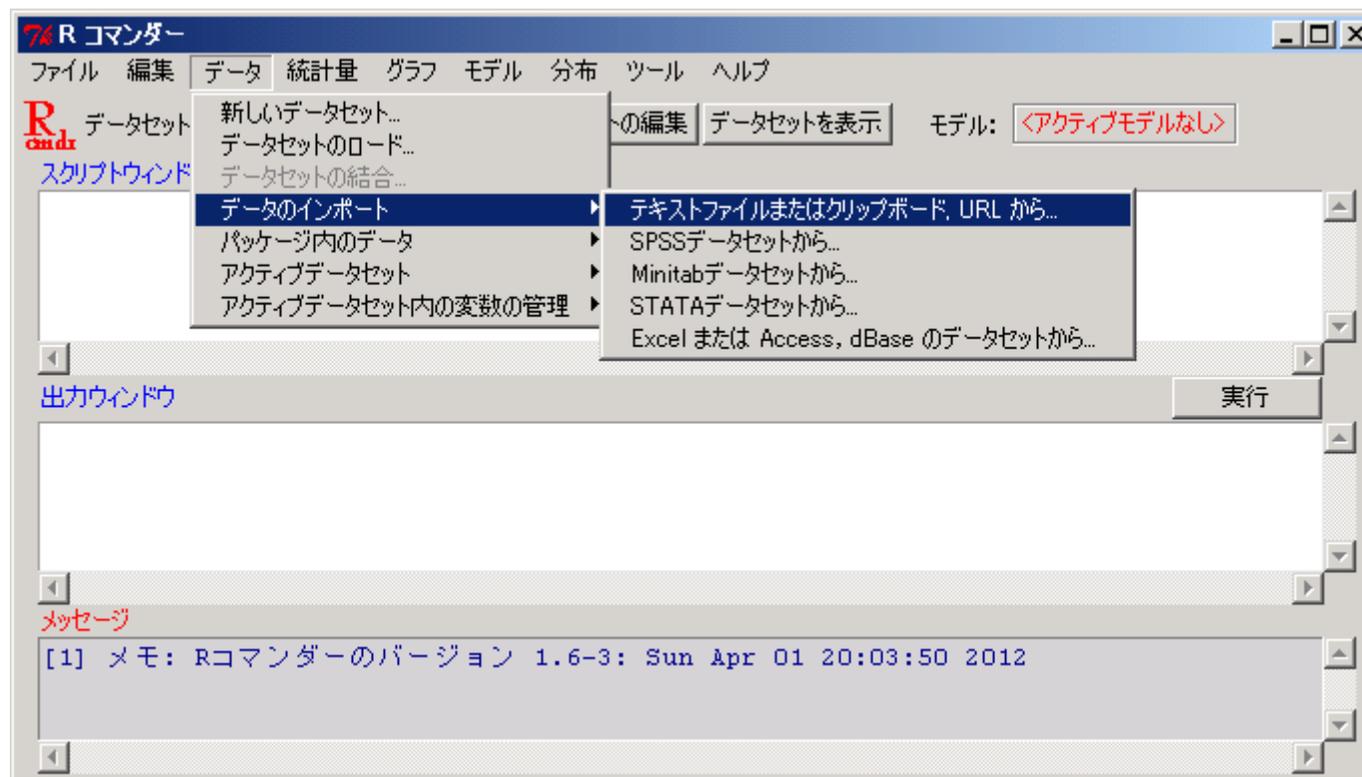


- ▶ データの読み込みを行う
 - ▶ 解析に使うデータセットを選択する
 - ▶ 読み込んだデータセットを表示する
 - ▶ データの一部（部分集合）を切り取る
 - ▶ データの変数に対して処理を行う
 - ▶ データの要約を行う
- etc...



データ「DEP」の読み込み①（CSV形式）

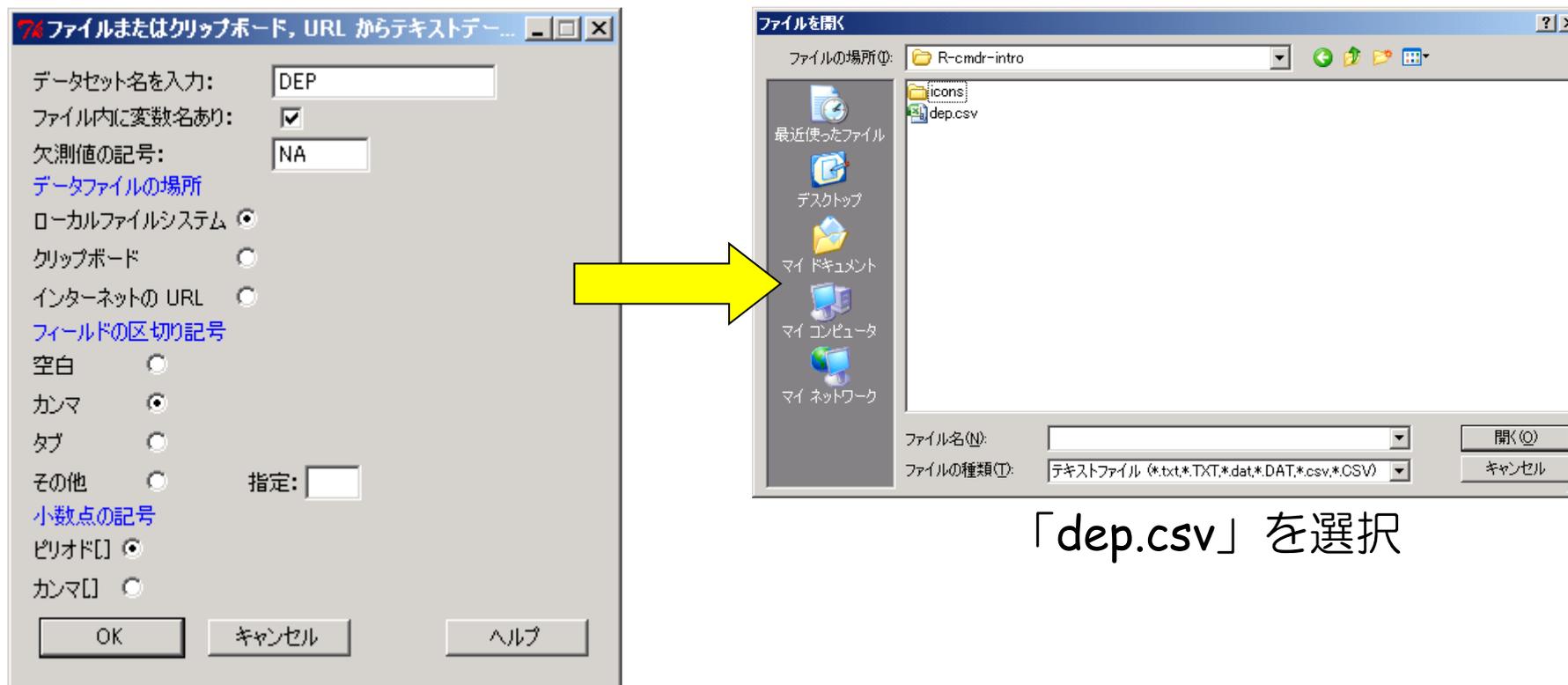
- ▶ CSV形式のファイル「dep.csv」を読み込む
- ▶ [データ] [データのインポート] [テキストファイルまたは...]





データ「DEP」の読み込み①（CSV形式）

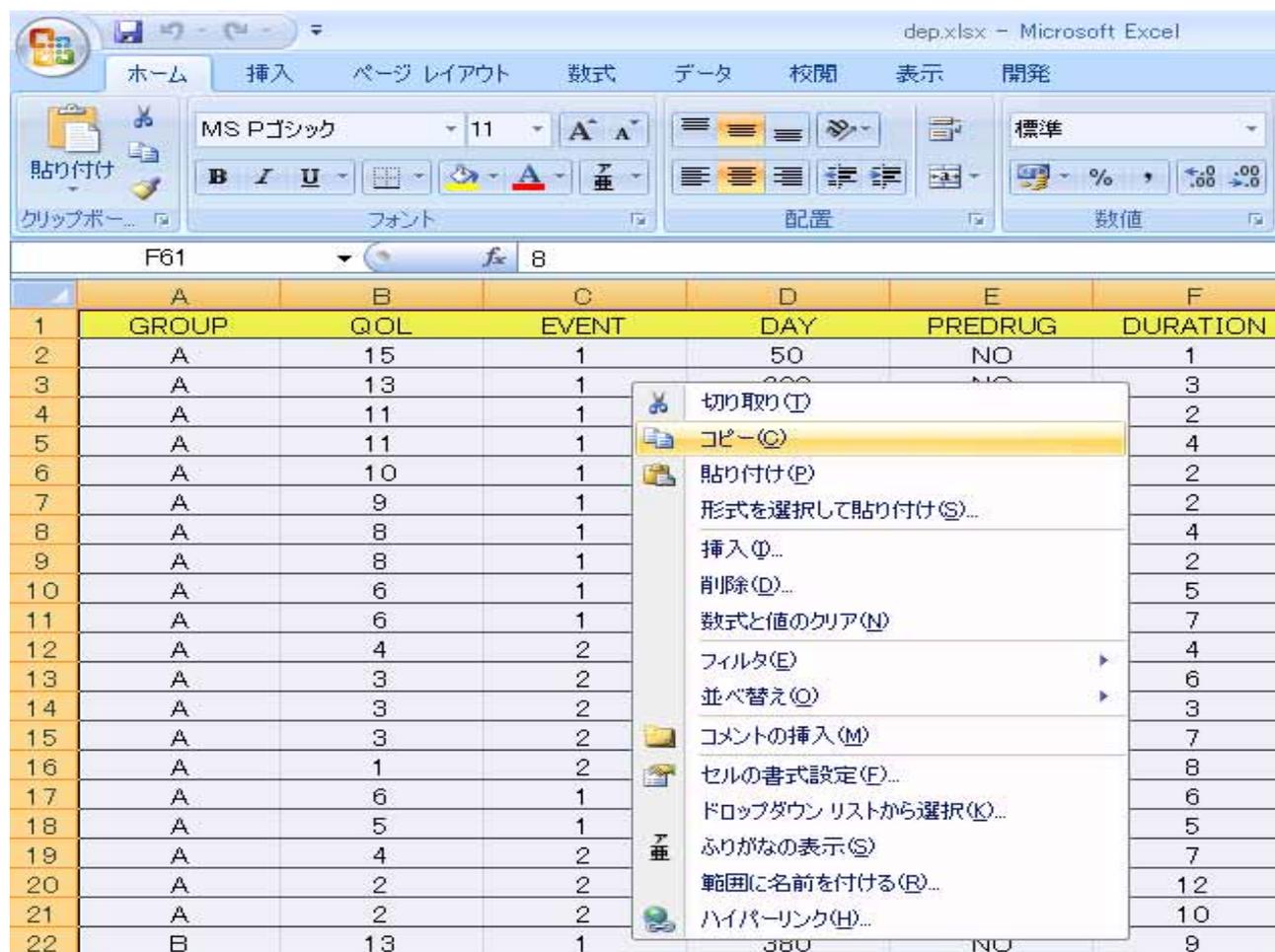
- ▶ データセット名：[DEP](#)
- ▶ データファイルの場所：[ローカルファイルシステム](#)
- ▶ フィールドの区切り文字：[カンマ](#)





データ「DEP」の読み込み②（クリップボード）

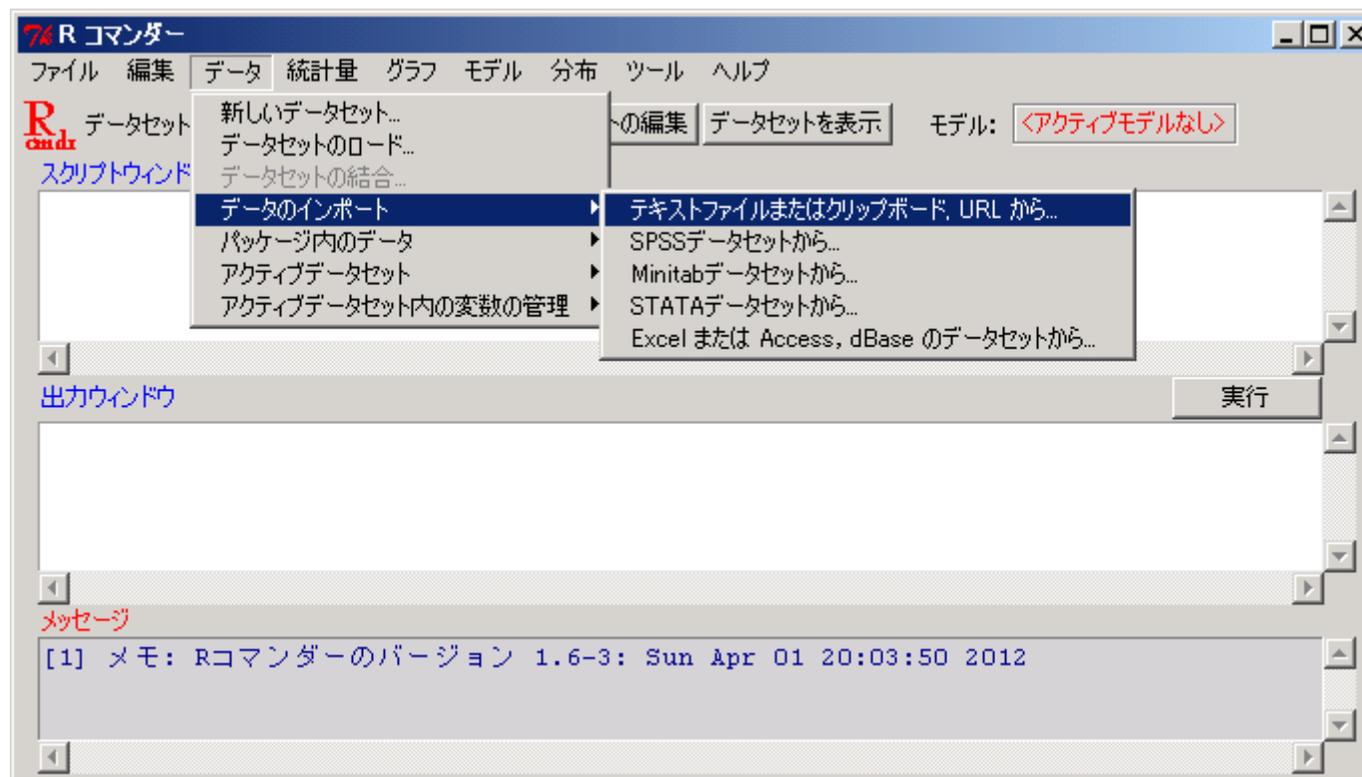
- ▶ EXCEL ファイル「dep.xlsx」を開いてデータがある範囲をコピーする





データ「DEP」の読み込み②（クリップボード）

- ▶ コピーした（クリップボードから）データを読み込む
- ▶ [データ] [データのインポート] [テキストファイルまたは...]





データ「DEP」の読み込み②（クリップボード）

- ▶ データセット名：DEP
- ▶ データファイルの場所：クリップボード

74 ファイルまたはクリップボード, URL からテキストデー...

データセット名を入力:

ファイル内に変数名あり:

欠測値の記号:

データファイルの場所

ローカルファイルシステム

クリップボード

インターネットの URL

フィールドの区切り記号

空白

カンマ

タブ

その他 指定:

小数点の記号

ピリオド[]

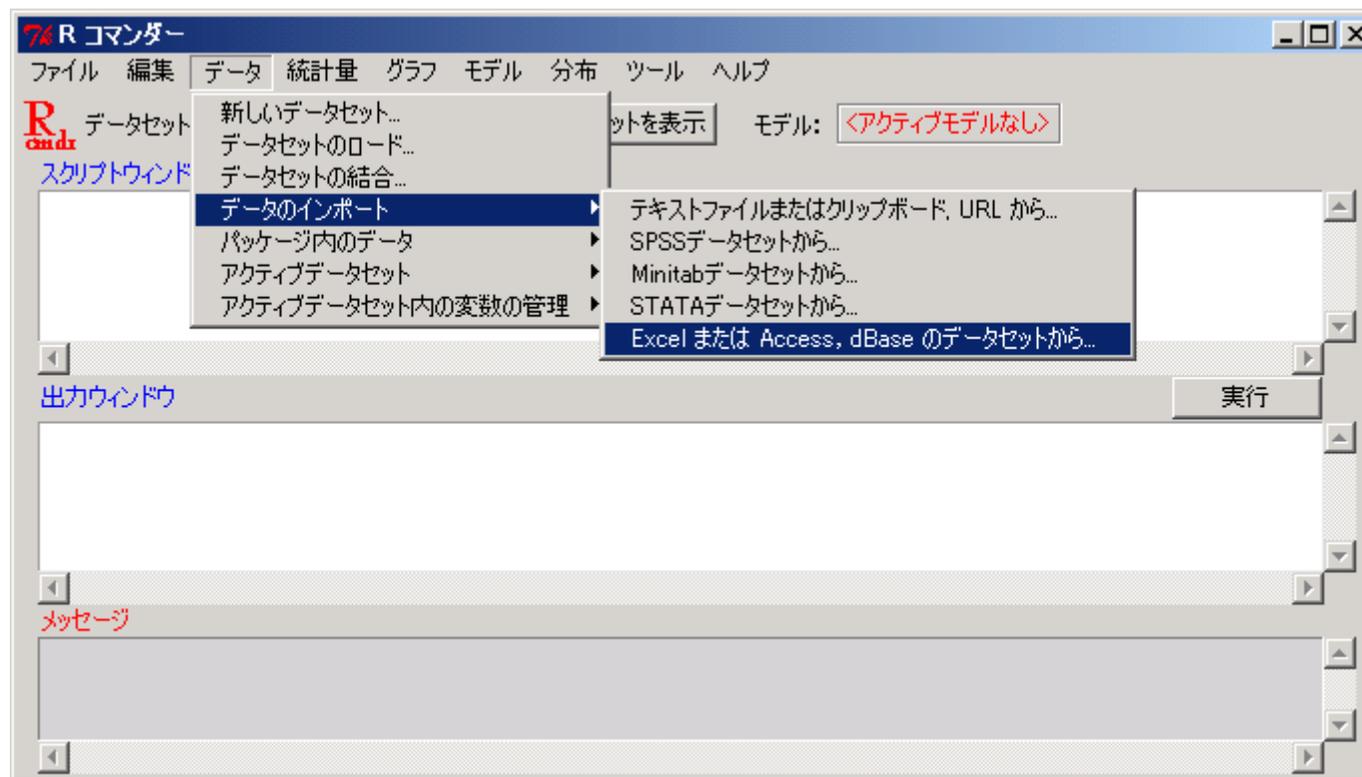
カンマ[]

OK キャンセル ヘルプ



データ「DEP」の読み込み③ (EXCEL)

- ▶ EXCEL ファイル「dep.xlsx」を読み込む
- ▶ [データ] [データのインポート] [Excel または...]





データ「DEP」の読み込み③ (EXCEL)

- ▶ データセット名：DEP
- ▶ ダイアログから「dep.xlsx」を選択

注：ファイルのパスに日本語が入っていると上手くいきません...



「dep.xlsx」を選択

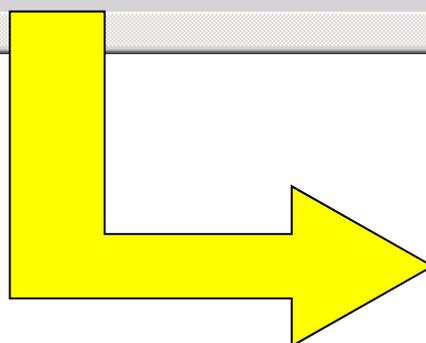


データ「DEP」の表示

① データを選択

② ここをクリック

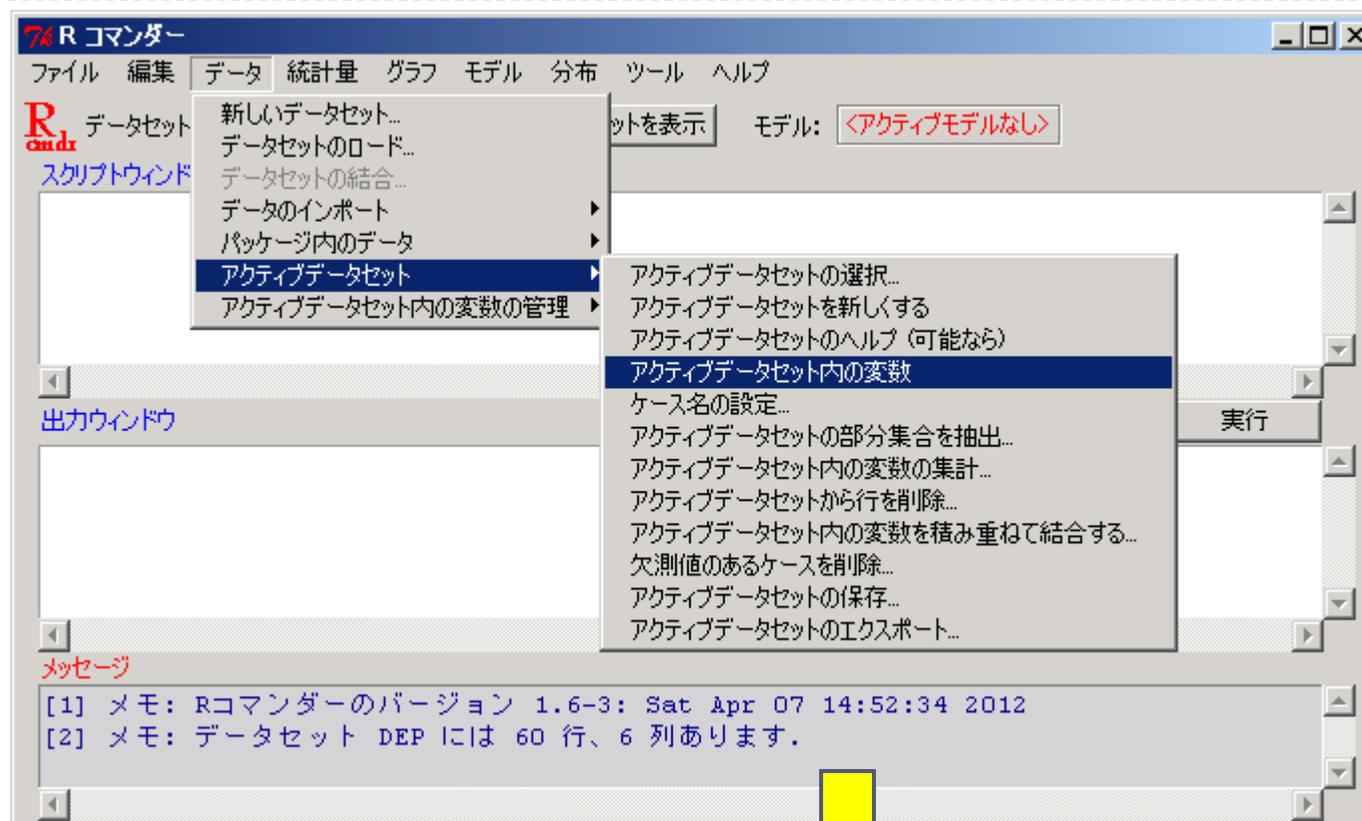
選択されたデータを
「アクティブデータ」
と呼ぶ



	GROUP	QOL	EVENT	DAY	PREDRUG	DURATION
1	A	15	1	50	NO	1
2	A	13	1	200	NO	3
3	A	11	1	250	NO	2
4	A	11	1	300	NO	4
5	A	10	1	350	NO	2
6	A	9	1	400	NO	2
7	A	8	1	450	NO	4
8	A	8	1	550	NO	2
9	A	6	1	600	NO	5
10	A	6	1	100	NO	7
11	A	4	2	250	NO	4
12	A	3	2	500	NO	6
13	A	3	2	750	NO	3
14	A	3	2	650	NO	7
15	A	1	2	1000	NO	8



アクティブデータの変数を表示



```
> names(DEP)
[1] "GROUP"      "QOL"        "EVENT"      "DAY"        "PREDRUG"   "DURATION"
```



ある条件に合致したデータのみ抽出

The screenshot shows the R Commander interface. The 'データ' (Data) menu is open, and 'アクティブデータセットの部分集合を抽出...' (Partial extraction from active data set...) is selected. This opens a dialog box titled 'データセットから部分的に抽出' (Partial extraction from data set). The dialog box has the following settings:

- すべての変数を含む (Include all variables)
- または (or)
- 変数(1つ以上選択) (Select 1 or more variables): DAY, DURATION, EVENT, GROUP
- 部分集合の表現 (Expression for partial set): GROUP != "C"
- 新しいデータセットの名前 (New data set name): AB

Buttons: OK, キャンセル (Cancel), ヘルプ (Help)

A large yellow arrow points from the dialog box back to the main window, indicating the result of the operation.

データセット DEP から薬剤 C
を除いたデータを「AB」という
名前で保存する場合は
「GROUP != "C"」と表現する



【参考】データを抽出する時の演算子

比較演算子

記号	==	!=	>=	>	<=	<
意味	等しい	≠		>		<

GROUP == "A" # GROUP が A であるレコードを抽出
GROUP != "C" # GROUP が C でないレコードを抽出
QOL >= 5 # QOL が 5 以上であるレコードを抽出

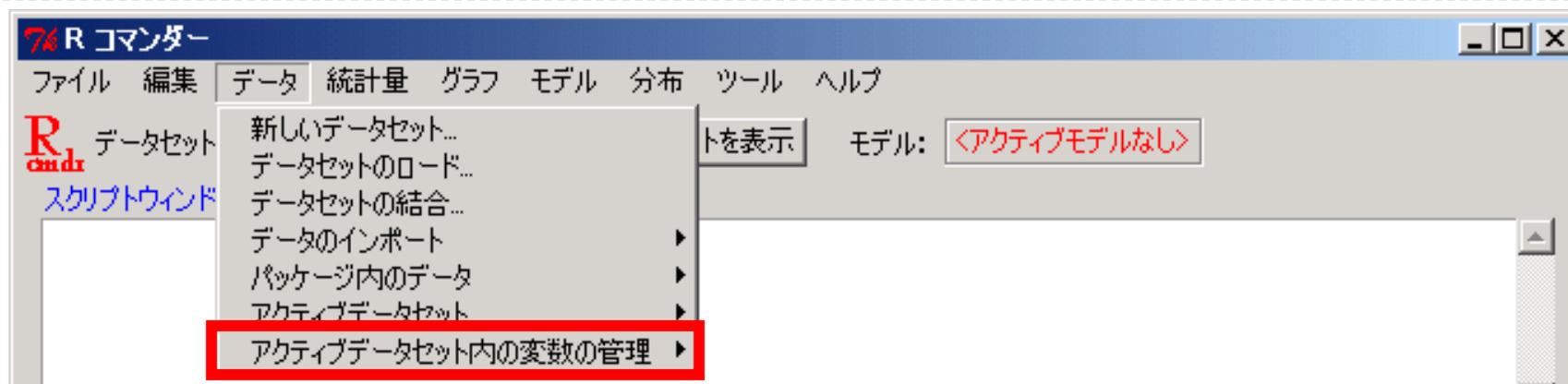
複数の条件を重ねるときの演算子

記号	!	&	
意味	≠	かつ	又は

GROUP == "A" & QOL >= 5 # 「GROUP が A」かつ「QOL が 5 以上」
GROUP == "A" | QOL >= 5 # 「GROUP が A」又は「QOL が 5 以上」



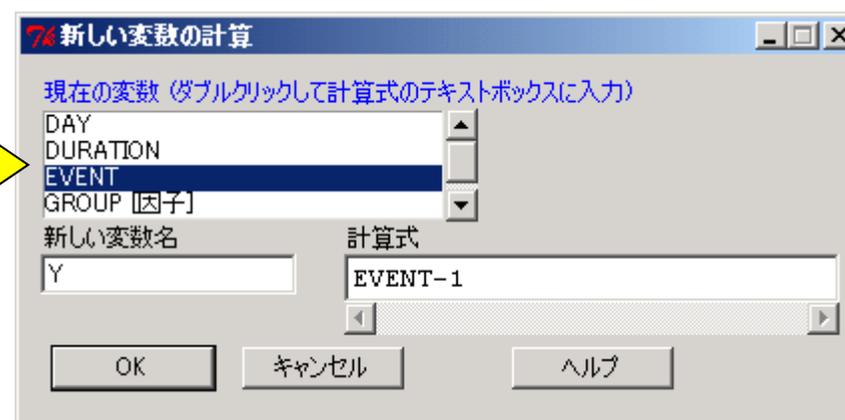
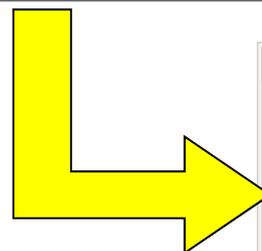
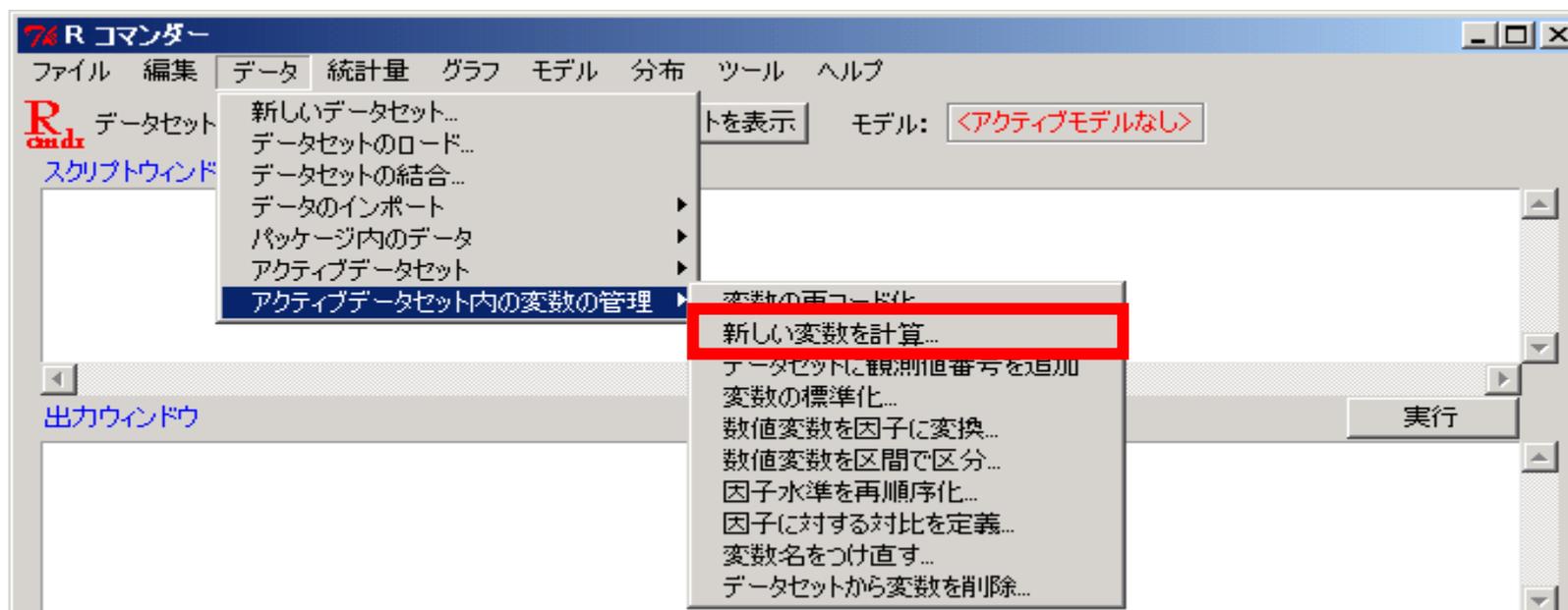
アクティブデータ内の変数の管理



- ▶ 新しい変数を追加する
 - ▶ 数値変数をカテゴリ変数（因子）に変換する
 - ▶ 変数名を変更する
 - ▶ 変数を削除する
- etc...



新しい変数を計算



アクティブデータに新たに変数を追加
(既存の変数から計算してもよいし、
既存の変数を使わずに計算してもよい)



【参考】 R の計算で使用出来る演算子と数学関数

```
> 1+sqrt(2) ⏎
```

```
[1] 2.414214
```

演算子

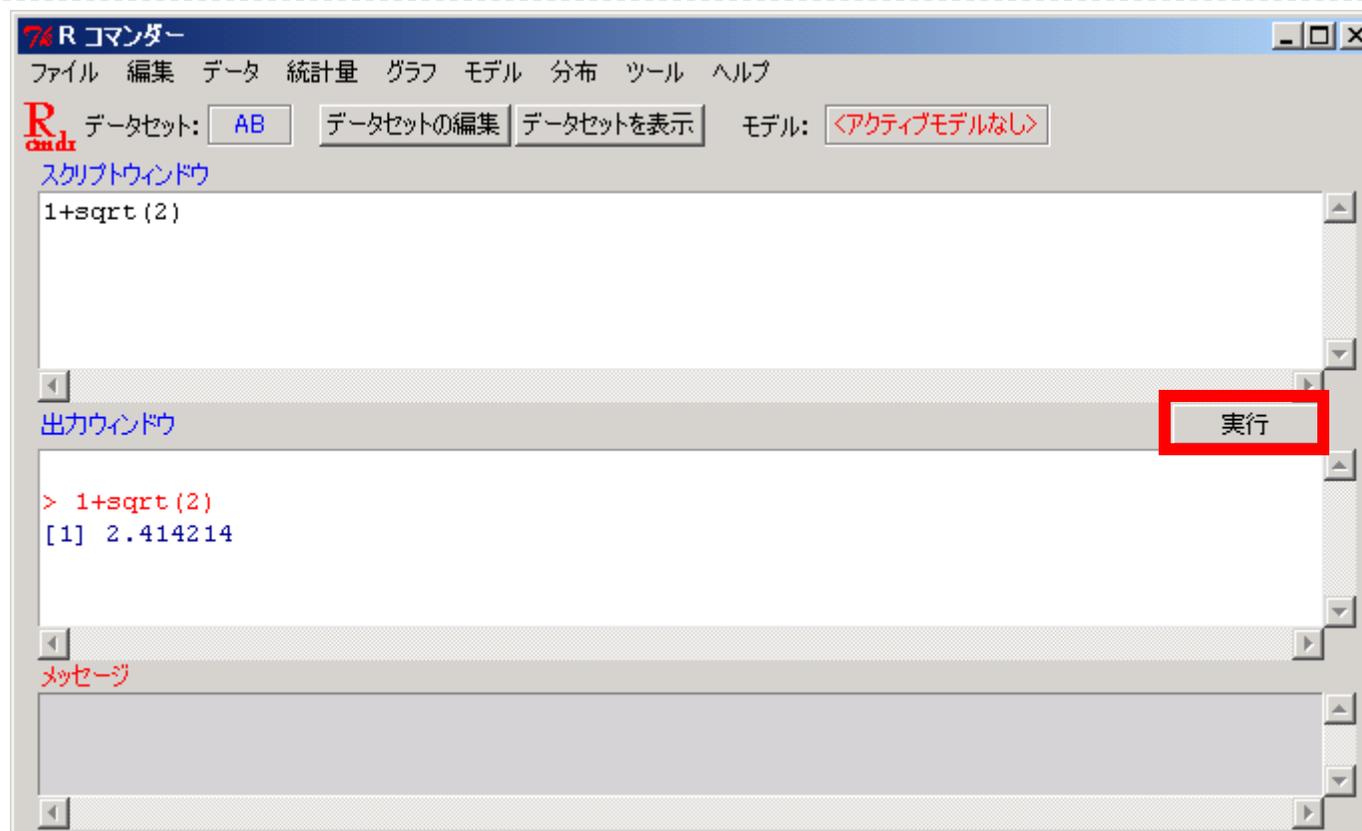
演算子	+	-	*	/	^
意味	加算	減算	乗算	除算	累乗

数学関数

関数	$\sin(x)$	$\cos(x)$	$\tan(x)$	$\log(x)$	$\log_{10}(x)$
意味	$\sin x$	$\cos x$	$\tan x$	$\log_e x$	$\log_{10} x$
関数	$\sinh(x)$	$\cosh(x)$	$\tanh(x)$	$\exp(x)$	$\text{sqrt}(x)$
意味	$\sinh x$	$\cosh x$	$\tanh x$	e^x	ルート x
関数	$\text{abs}(x)$	$\text{trunc}(x)$	$\text{round}(x)$	$\text{floor}(x)$	$\text{ceiling}(x)$
意味	絶対値	整数部分	丸め	切り下げ	切り上げ



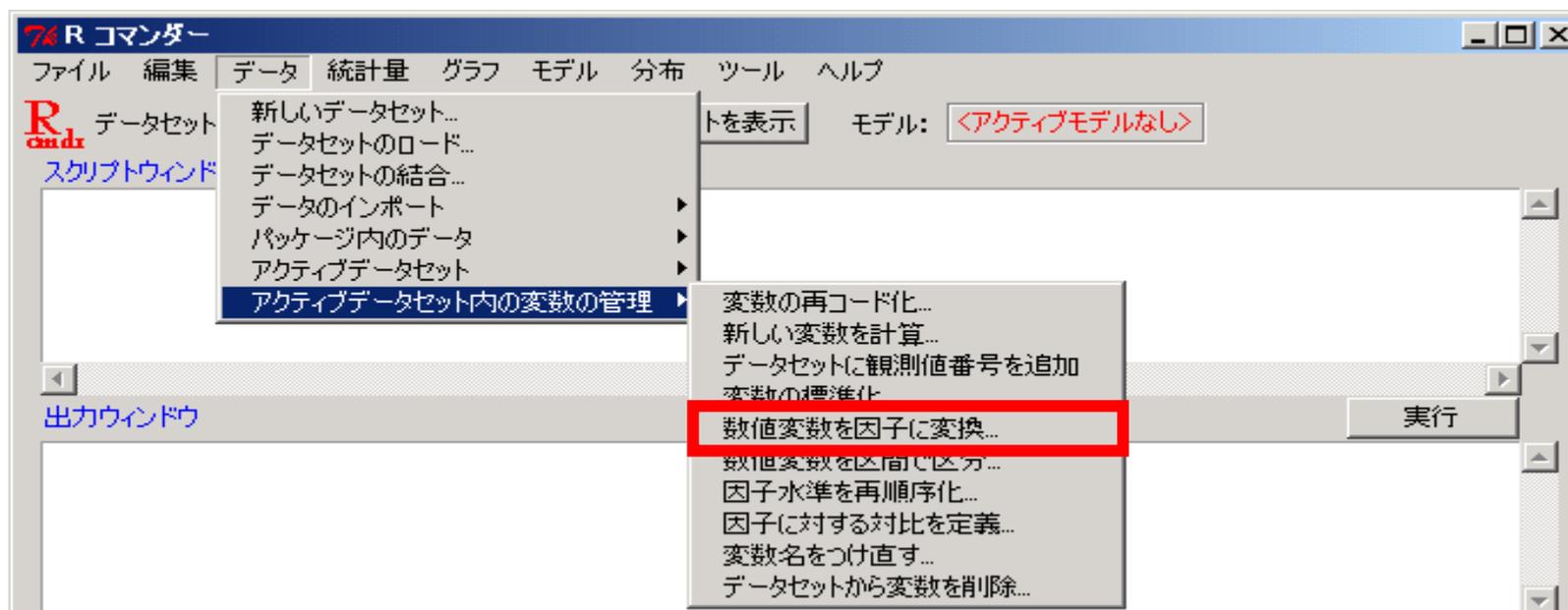
【参考】 R Commander で関数電卓



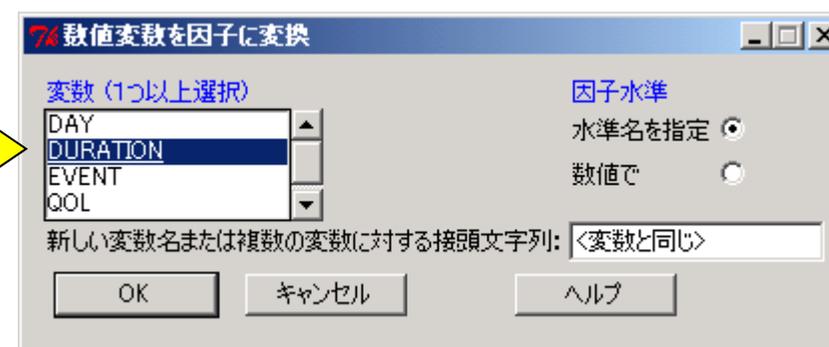
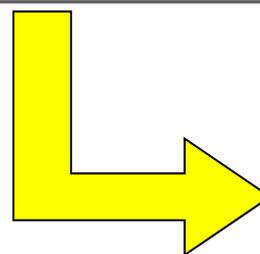
- ▶ スクリプトウィンドウに数式を入れた後（必要に応じて，計算したい部分のみマウスで選択してから）【実行】をクリックする



数値変数をカテゴリ変数（因子）に変換

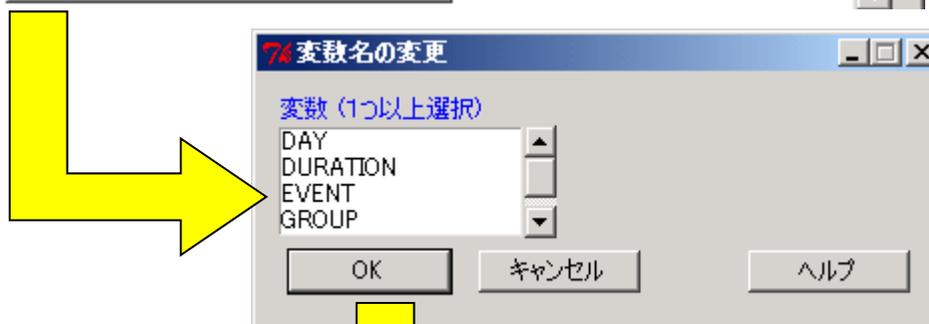
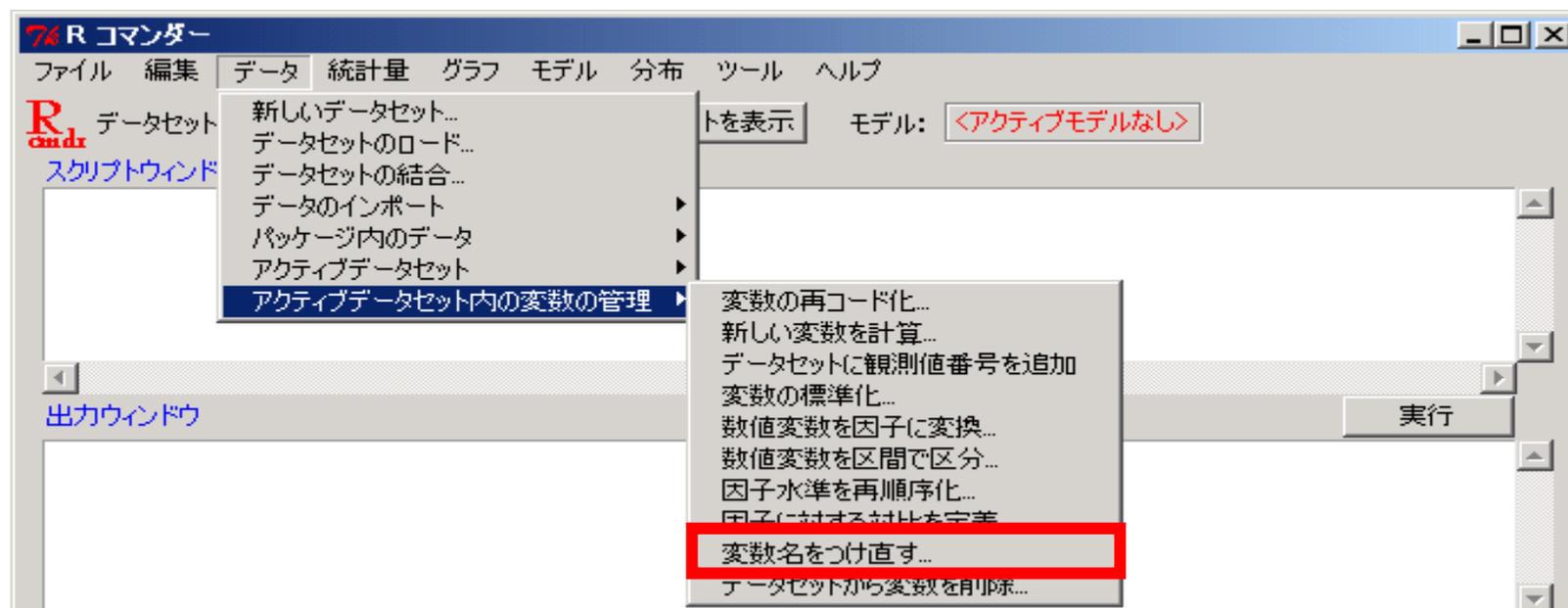


回帰分析等ではカテゴリ変数は必ず因子に変換しなければいけないのでカテゴリ変数として扱いたい変数はあらかじめカテゴリ変数（因子）に変換



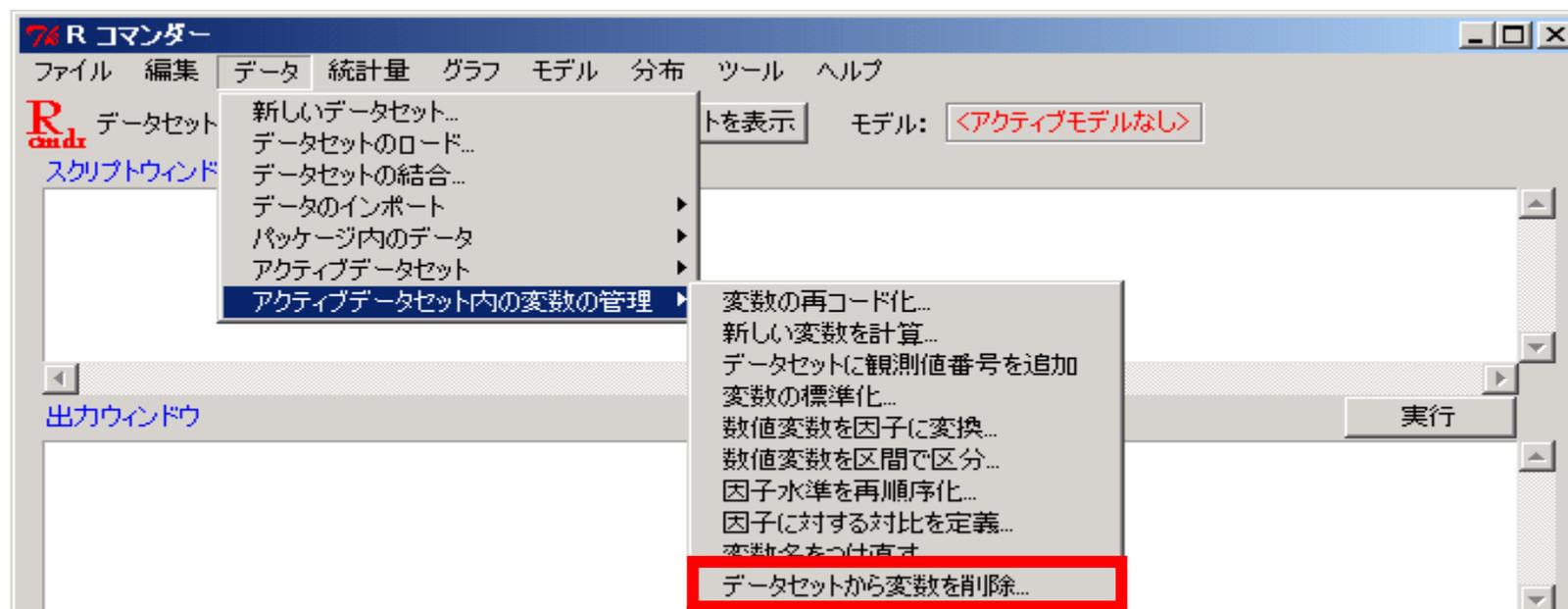


変数名を変更する





変数を削除する





本日のメニュー

1. R のセットアップ (Ver. 2.12.2) のメモ

2. R Commander の基礎

- ▶ イントロ

- ▶ データの読み込みと変数に対する処理

- ▶ データのグラフ化

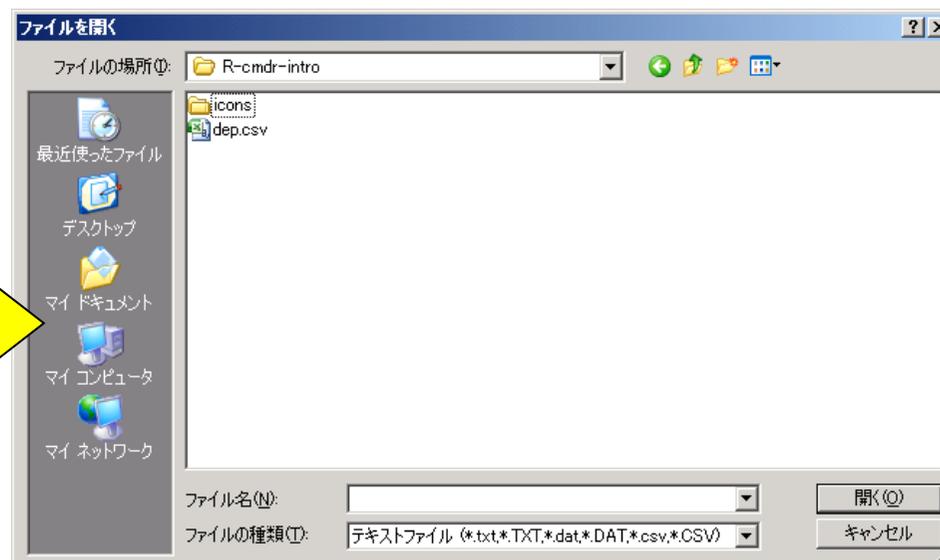
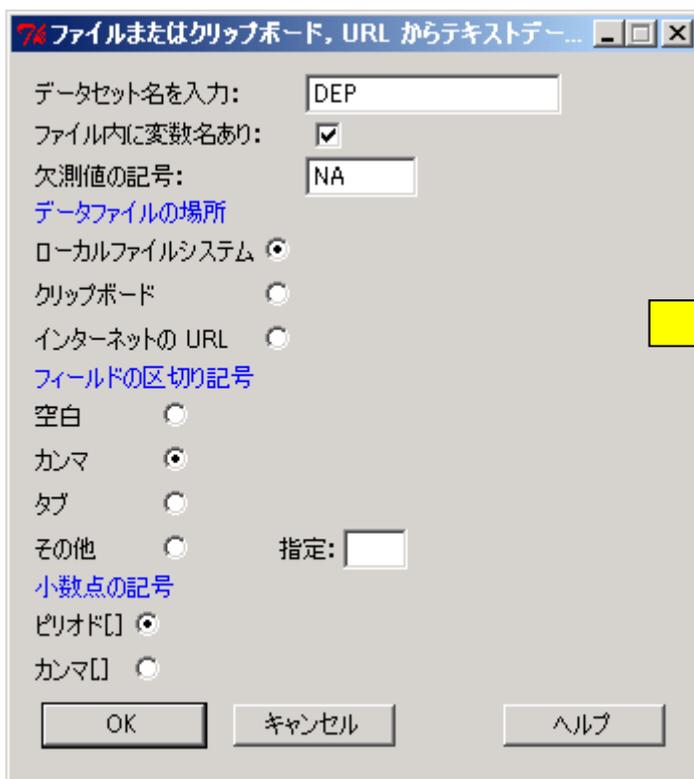
- ▶ データ解析

3. パッケージによる機能追加



【準備】 データ「DEP」の読み込み（CSV形式）

- ▶ データセット名：[DEP](#)
- ▶ データファイルの場所：[ローカルファイルシステム](#)
- ▶ フィールドの区切り文字：[カンマ](#)



「dep.csv」を選択

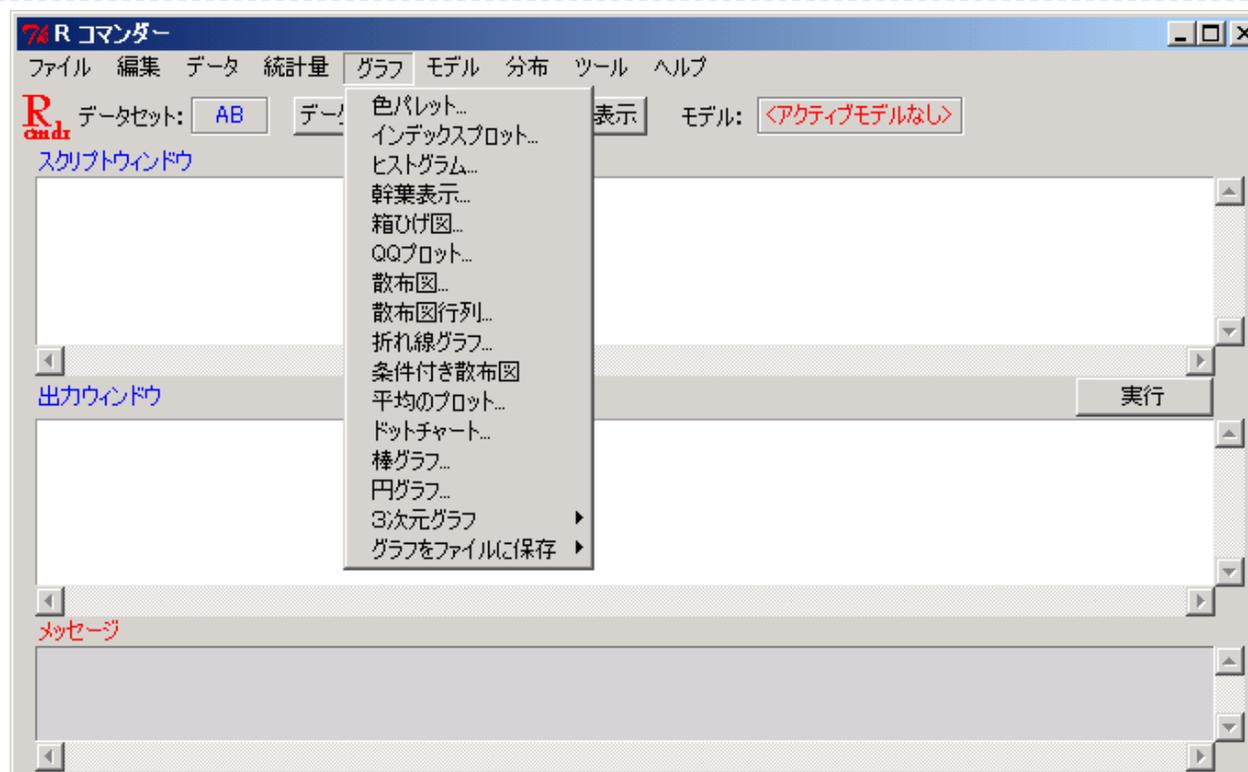


データ「DEP」の変数

- ▶ **GROUP** : 薬剤の種類 (**A**, **B**, **C**)
- ▶ **QOL** : QOL の点数 (数値) 点数が大きい方が良い
- ▶ **EVENT** : 改善の有無 (1 : 改善あり, 2 : 改善なし)
 QOLの点数が 5 点以上である場合を「改善あり」とする
- ▶ **DAY** : 観察期間 (数値, 単位は日)
- ▶ **PREDRUG** : 前治療薬の有無 (**YES** : 他の治療薬を投与したことあり,
 NO : 投与したことなし)
- ▶ **DURATION** : 罹病期間 (数値, 単位は年)



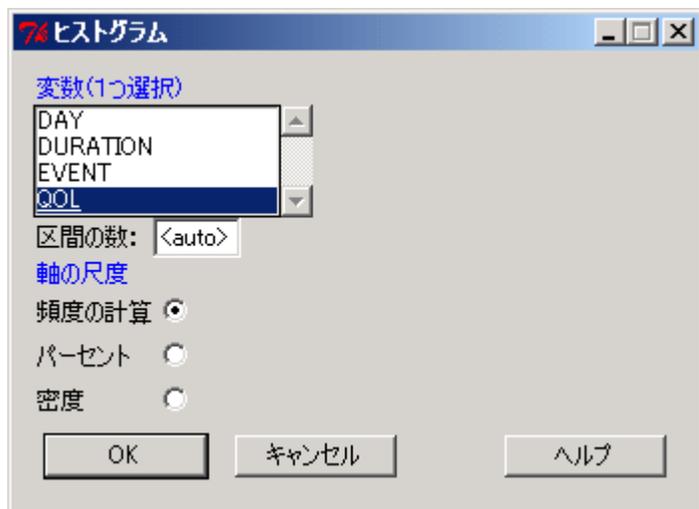
メニュー〔グラフ〕



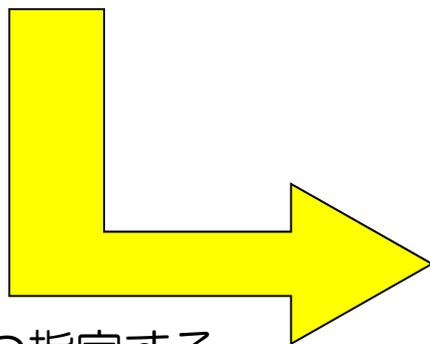
- ▶ 様々な種類のグラフが描け、グラフをファイルに保存することも可能
- ▶ 1変数のグラフ化に関する内容は「1つのデータの要約」を参照のこと：
http://www.occn.zaq.ne.jp/cuhxr802/R-stat-intro_03.pdf



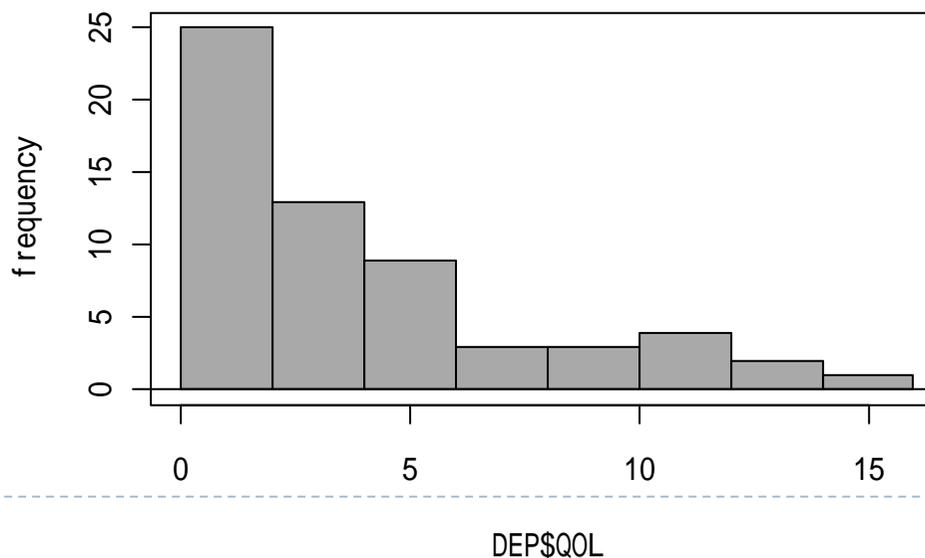
QOL のヒストグラム



- ▶ ヒストグラムはデータ全体に対するものなので、この場合は薬剤 A も薬剤 B も薬剤 C も混ぜたヒストグラムとなっている
- ▶ 薬剤別にグラフを描く場合は、「薬剤 A のみのデータ」「薬剤 B のみのデータ」「薬剤 C のみのデータ」を作成してグラフを描けばよい

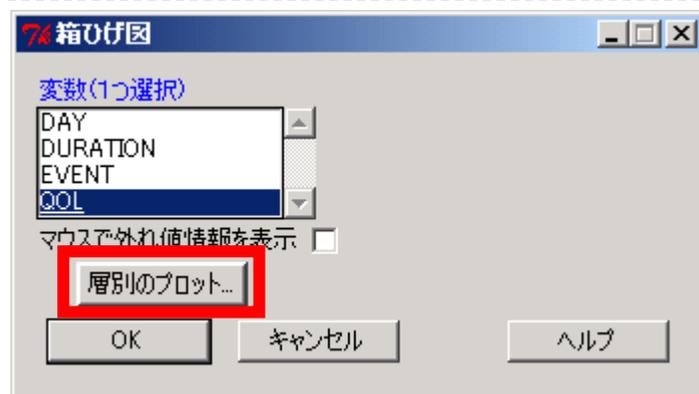


変数を 1 つ指定する

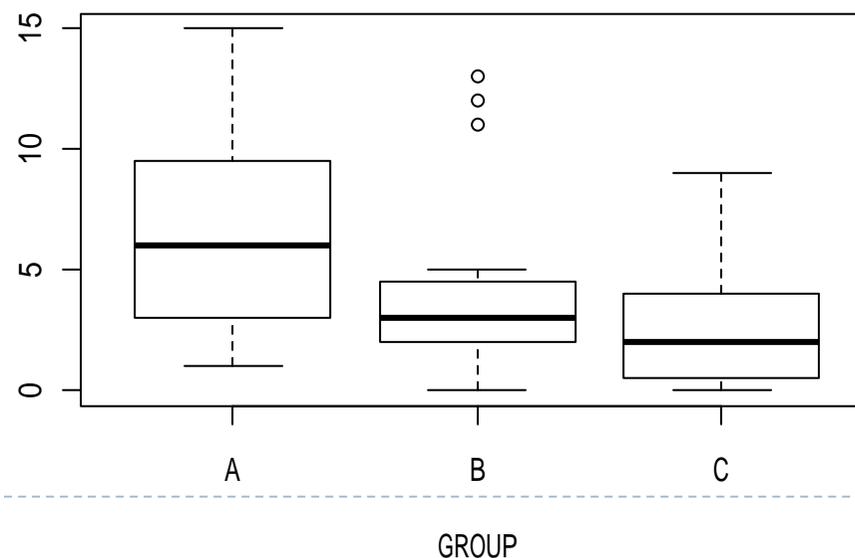
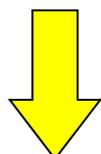




QOL の箱ひげ図

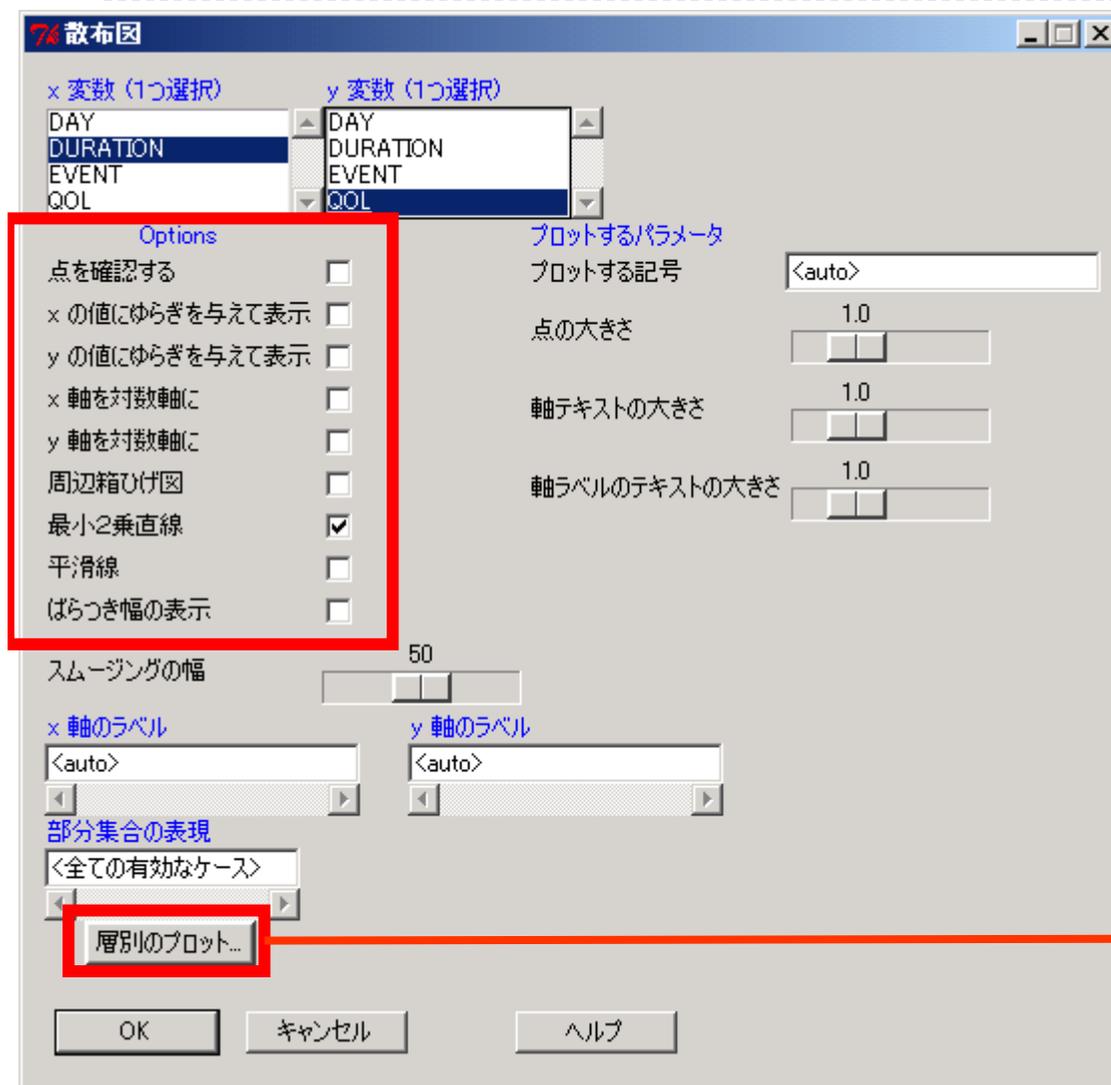


- ▶ 「層別のプロット」をクリック
- ▶ 層別変数として **GROUP** を指定することで薬剤別の QOL の箱ひげ図が作成される

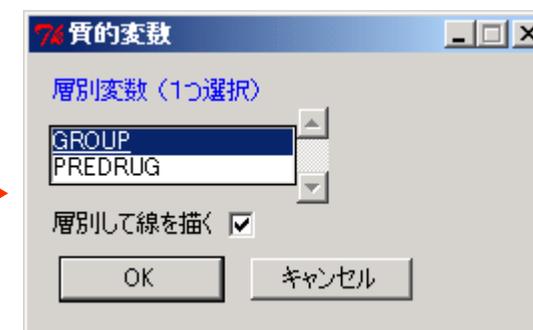




DURATION（罹病期間）と QOL の散布図



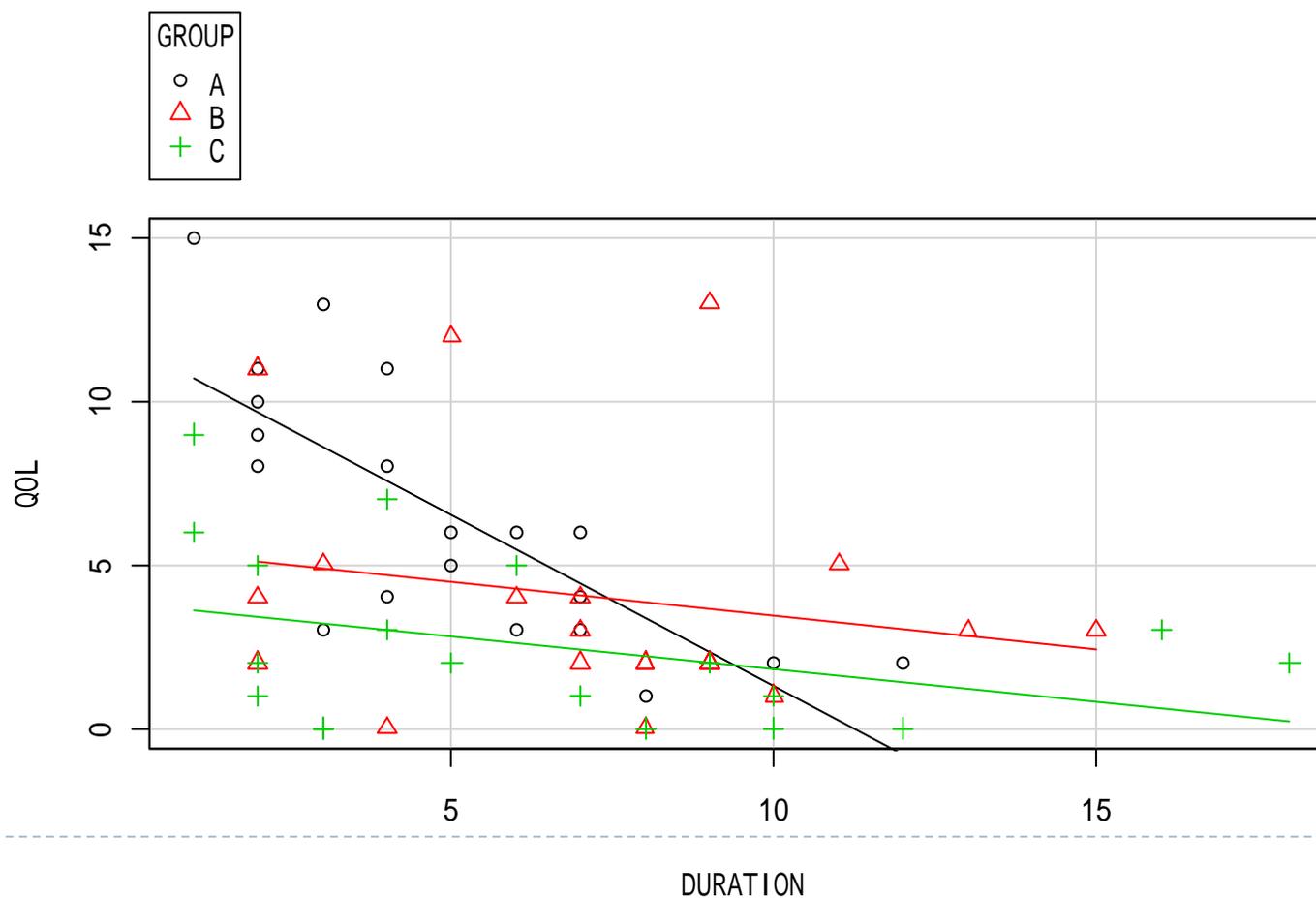
- ▶ 「層別のプロット」をクリック
- ▶ 層別変数として **GROUP** を指定することで、薬剤別の QOL の箱ひげ図が作成される
- ▶ オプションを多くつけない方がよい（最小 2 乗直線のみ）





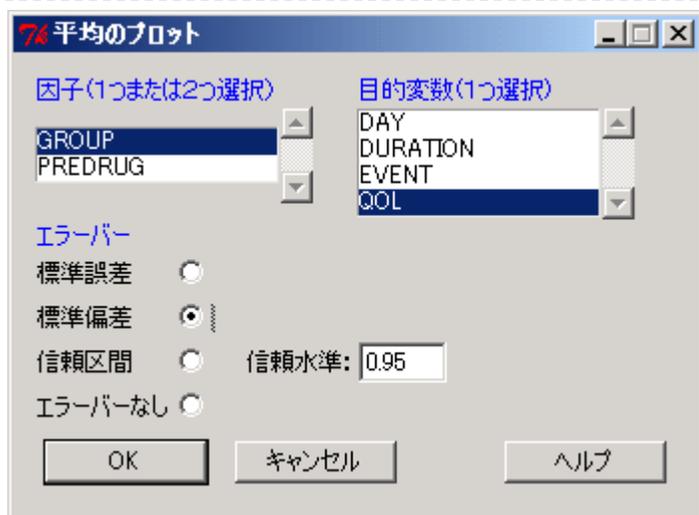
DURATION（罹病期間）と QOL の散布図

- ▶ 解釈については「散布図と回帰直線と相関係数」を参照のこと：
http://www.occn.zaq.ne.jp/cuhxr802/R-stat-intro_04.pdf

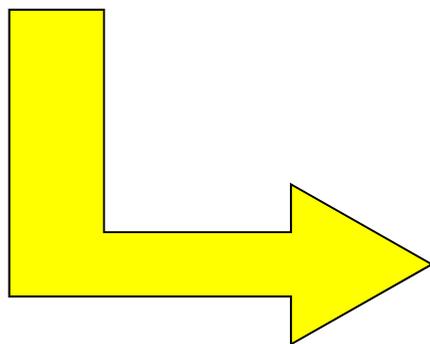




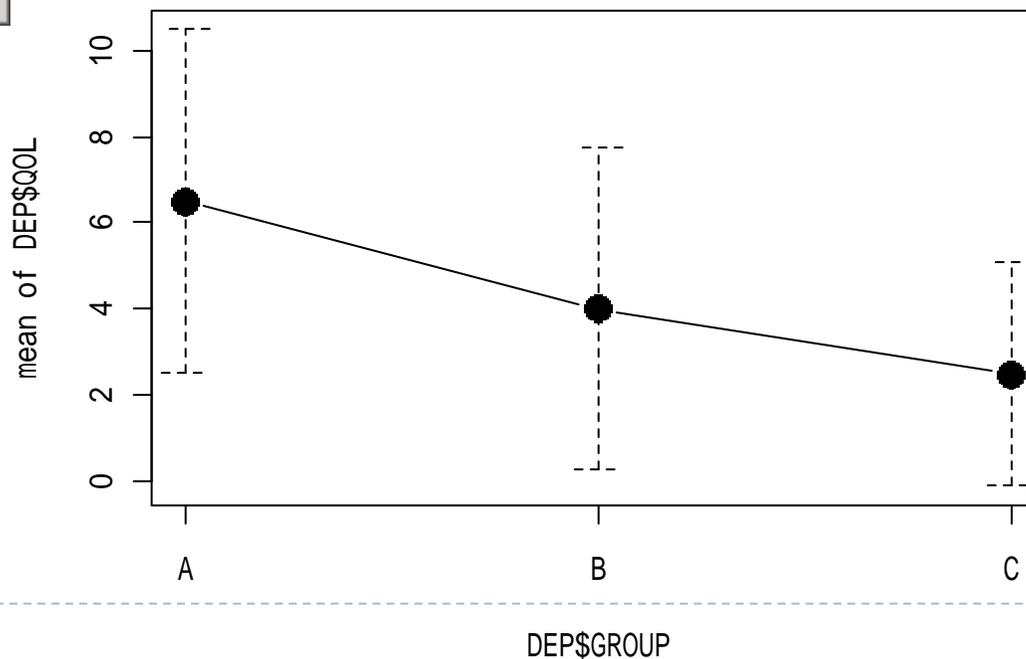
QOL に関する平均・標準偏差の図



- ▶ 薬剤別に QOL の平均値の図が作成される
- ▶ エラーバーは「標準偏差」の代わりに「標準誤差」や「●%信頼区間」でも可

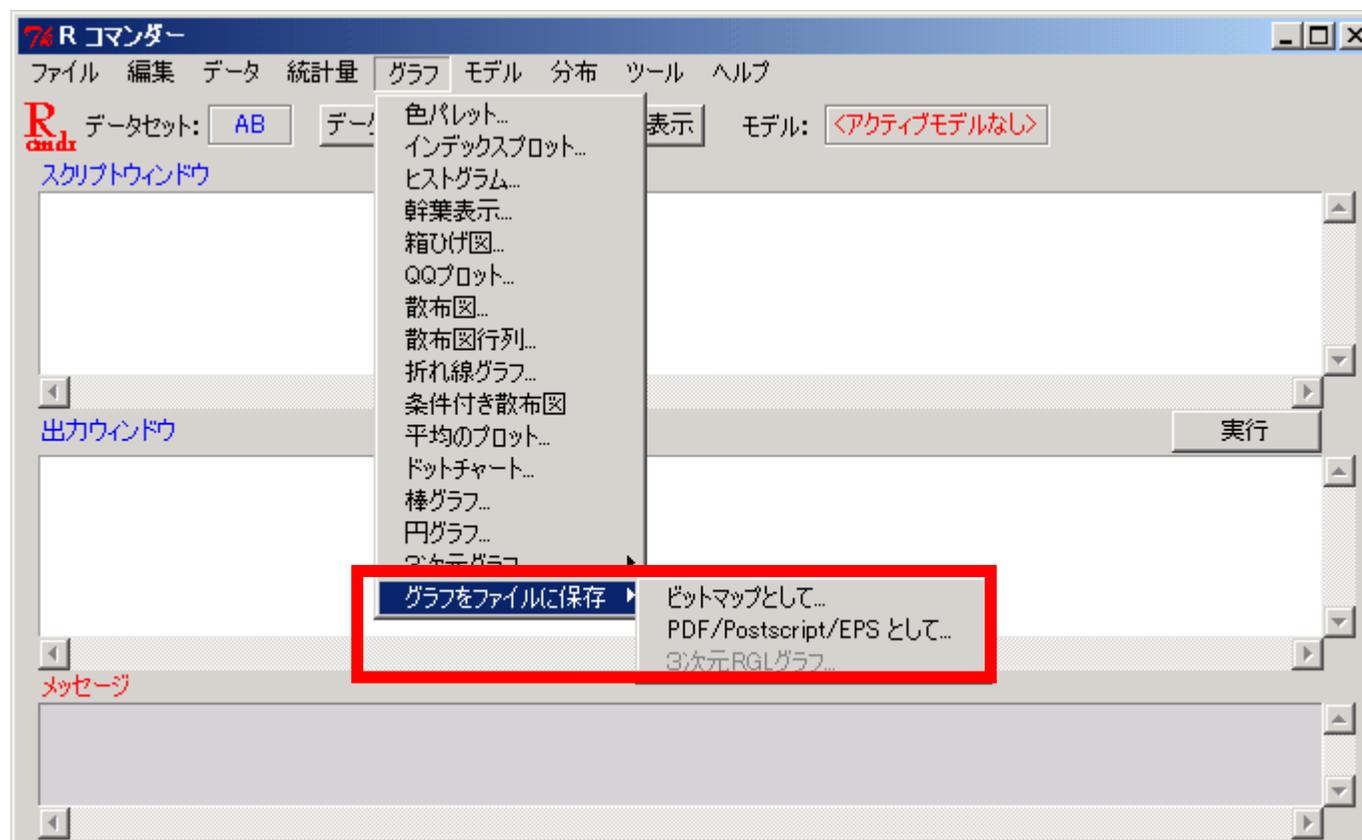


Plot of Means





グラフの保存



- ▶ ビットマップ（実際は png と jpeg ）， pdf ，ポストスクリプト ， eps 形式でグラフを保存することが出来る



本日のメニュー

1. R のセットアップ (Ver. 2.12.2) のメモ

2. R Commander の基礎

- ▶ イントロ
- ▶ データの読み込みと変数に対する処理
- ▶ データのグラフ化

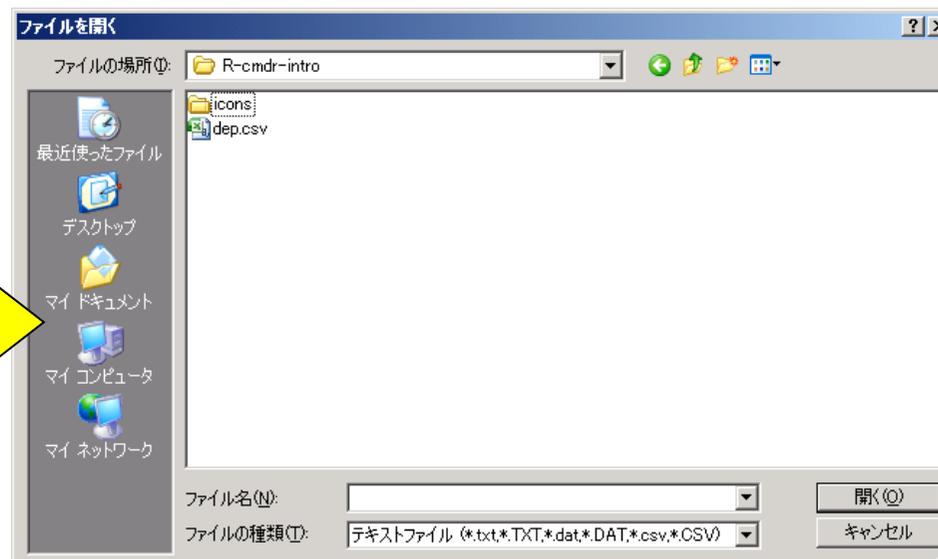
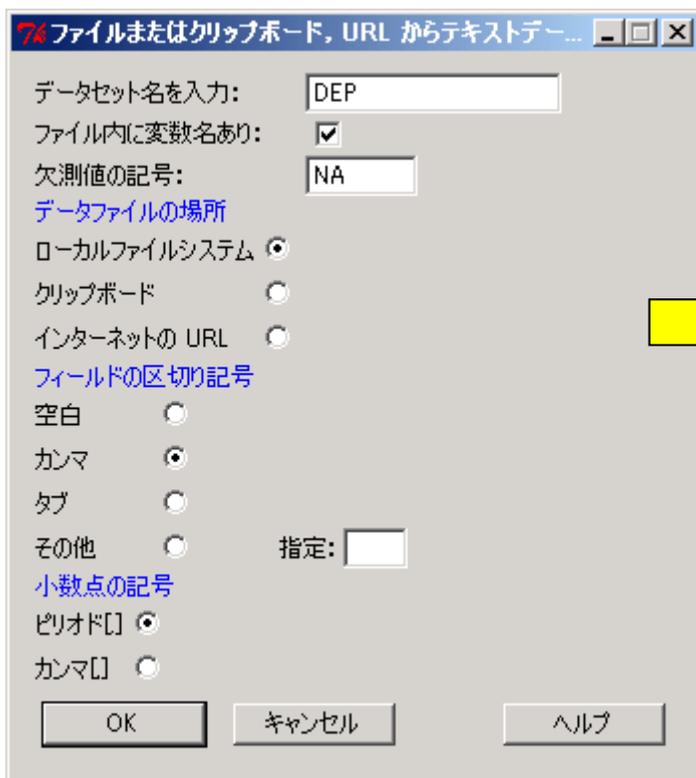
▶ データ解析

3. パッケージによる機能追加



【準備】データ「DEP」の読み込み（CSV形式）

- ▶ データセット名：[DEP](#)
- ▶ データファイルの場所：[ローカルファイルシステム](#)
- ▶ フィールドの区切り文字：[カンマ](#)



「dep.csv」を選択



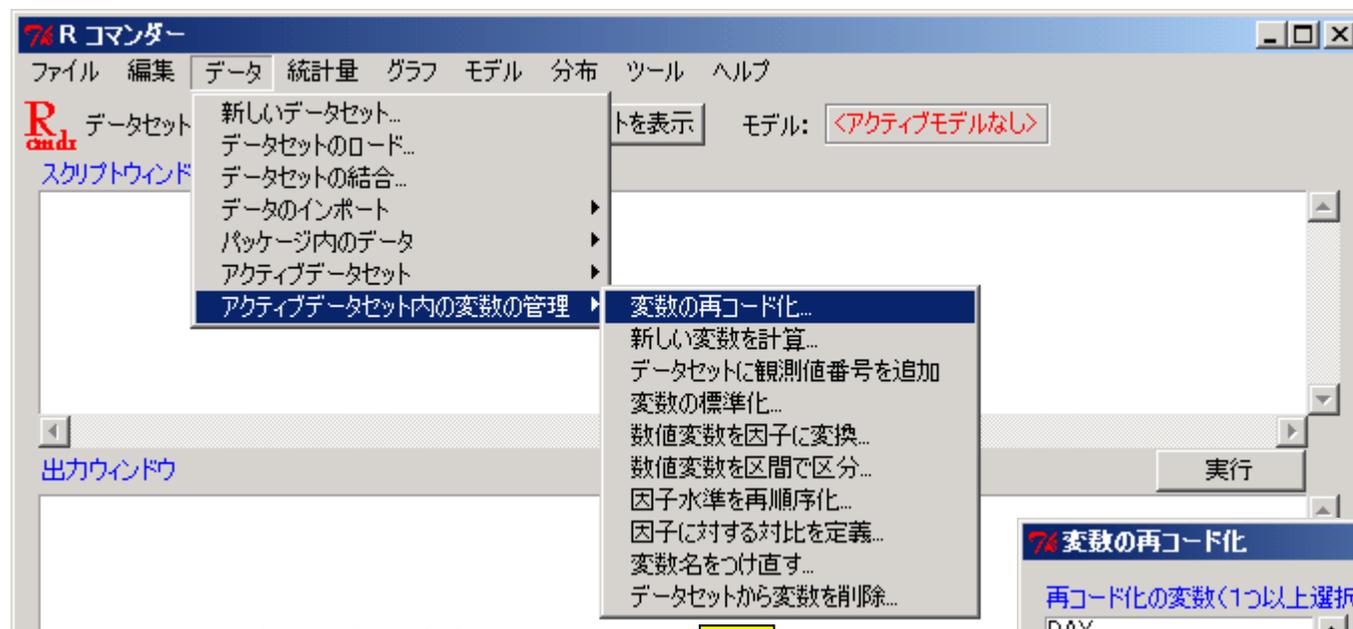
【準備】薬剤 A と薬剤 B のデータを抽出 「AB」

The screenshot shows the R Commander interface. The 'Active Data Set' menu is open, and the option 'Active Data Set Partial Extraction...' is selected. A yellow arrow points from this menu item to the 'Partial Extraction from Data Set' dialog box. In the dialog box, the 'Include all variables' checkbox is checked. Under 'or select variables (select 1 or more)', the variables DAY, DURATION, EVENT, and GROUP are listed. The 'Partial set expression' field contains 'GROUP != "C"'. The 'New data set name' field contains 'AB'. The 'OK' button is highlighted.

データセット DEP から薬剤 C を除いたデータを「AB」という名前で保存する場合は「GROUP != "C"」と表現する



【準備】薬剤 A と薬剤 B のデータを抽出 「AB」

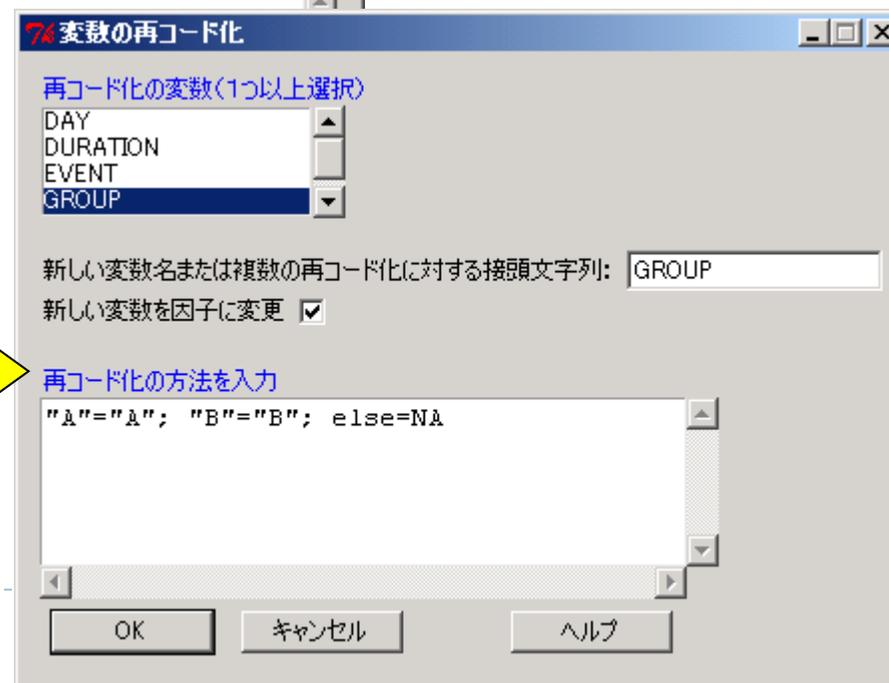
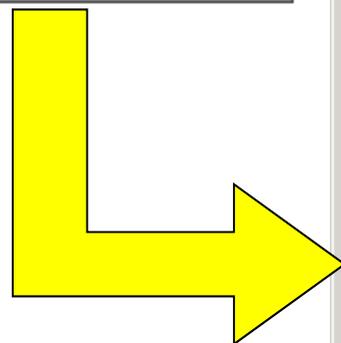


【バグ】

変数 GROUP には "C" というデータは無いはずだが、元の「DEP」に "C" があることが原因で "C" というデータが存在することになっている...

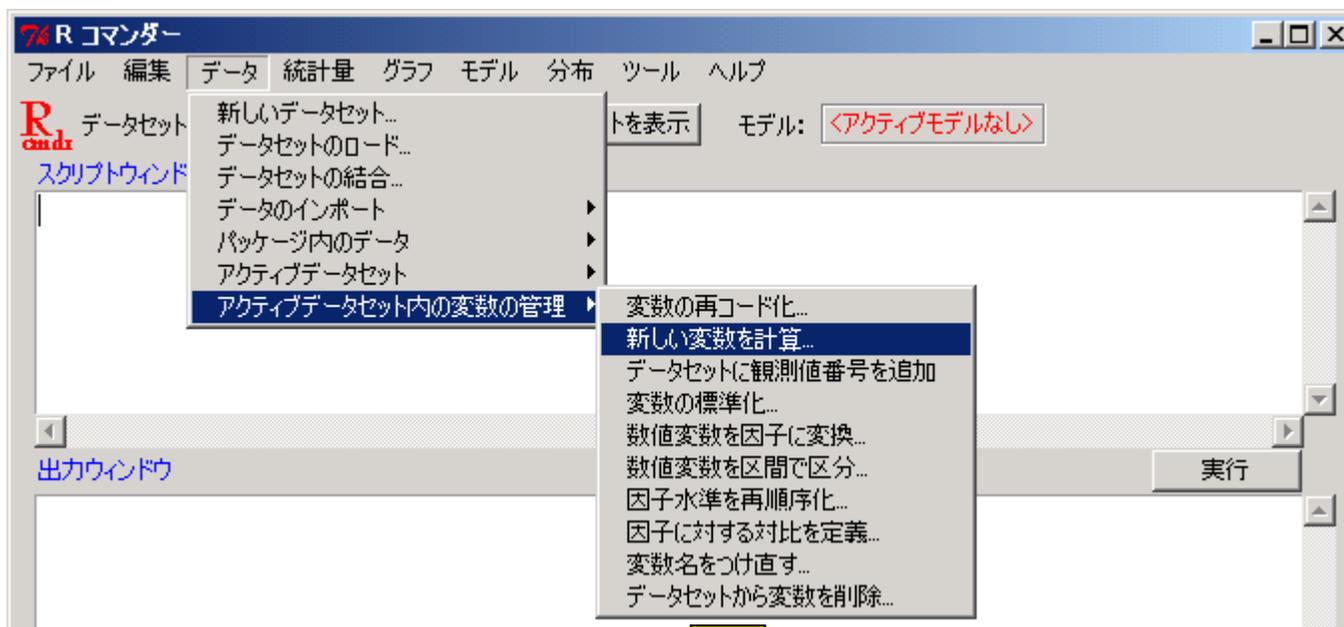
【バグ対応】

- ① 「変数の再コード化」を選択
- ② 変数：GROUP
新しい変数名...：GROUP
再コード化の方法...
"A"="A"; "B"="B"; else=NA
と入力
- ③ [OK] をクリック





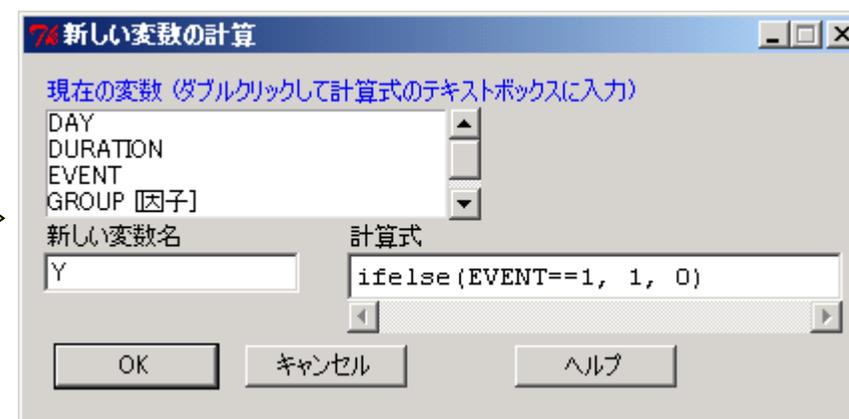
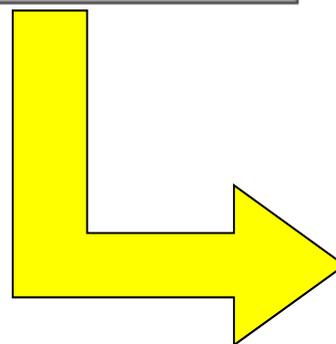
【準備】変数 Y (0 or 1 の変数) の作成



変数 EVENT が 1 ならば 1,
変数 EVENT が 2 ならば 0
という変数 Y を作成するために

`ifelse(EVENT==1, 1, 0)`

なる計算式を入力する



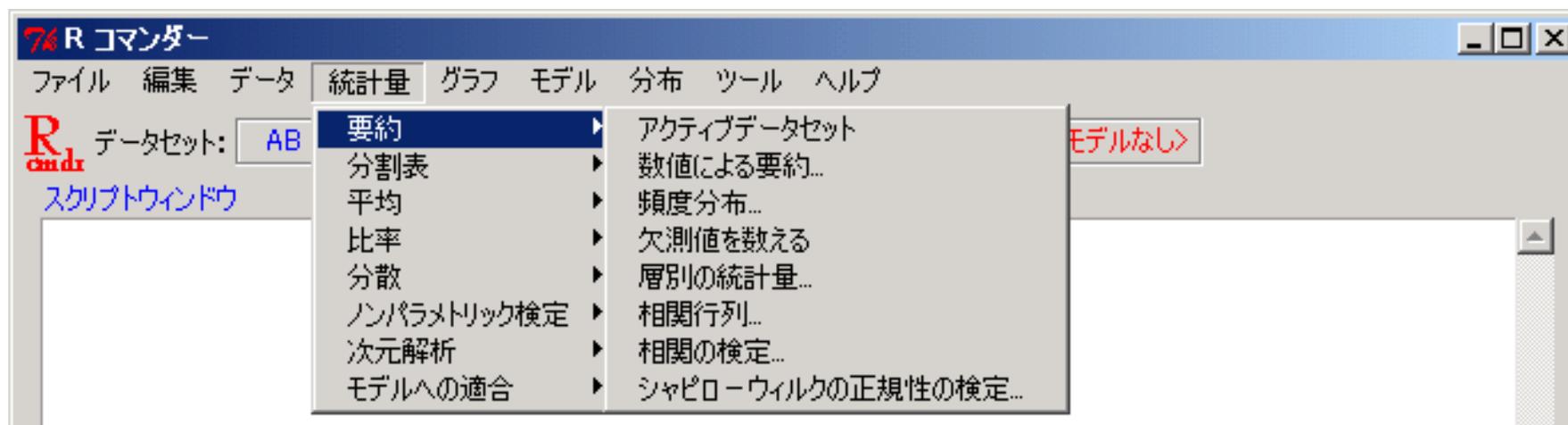


データ「AB」の変数

- ▶ **GROUP** : 薬剤の種類 (**A**, **B**)
- ▶ **QOL** : QOL の点数 (数値) 点数が大きい方が良い
- ▶ **EVENT** : 改善の有無 (**1** : 改善あり, **2** : 改善なし)
 QOLの点数が **5** 点以上である場合を「改善あり」とする
- ▶ **DAY** : 観察期間 (数値, 単位は日)
- ▶ **PREDRUG** : 前治療薬の有無 (**YES** : 他の治療薬を投与したことあり,
 NO : 投与したことなし)
- ▶ **DURATION** : 罹病期間 (数値, 単位は年)
- ▶ **Y** : 改善の有無 (**1** : 改善あり, **0** : 改善なし)
 QOLの点数が **5** 点以上である場合を「改善あり」とする



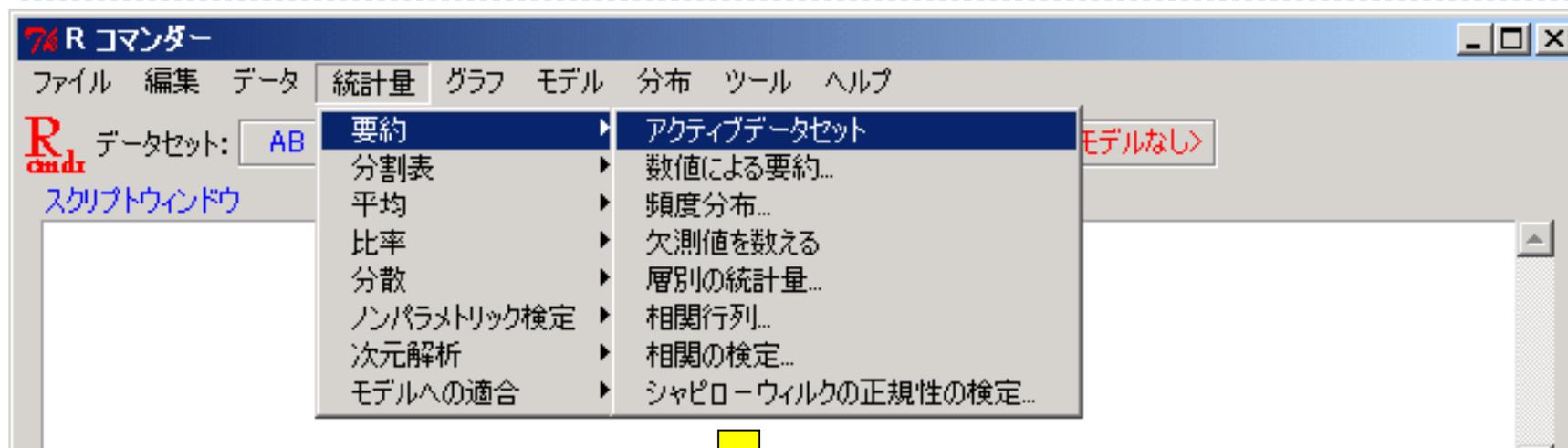
メニュー〔統計量〕 〔要約〕



- ▶ 要約：要約統計量や相関係数の算出，頻度集計を行う
- ▶ 連続変数の要約に関する内容は「1つのデータの要約」を参照のこと：
http://www.occn.zaq.ne.jp/cuhxr802/R-stat-intro_03.pdf
- ▶ 2標本 t 検定に関する内容は「2標本 t 検定と回帰分析」を参照のこと：
http://www.occn.zaq.ne.jp/cuhxr802/R-stat-intro_05.pdf



メニュー〔統計量〕 〔要約〕

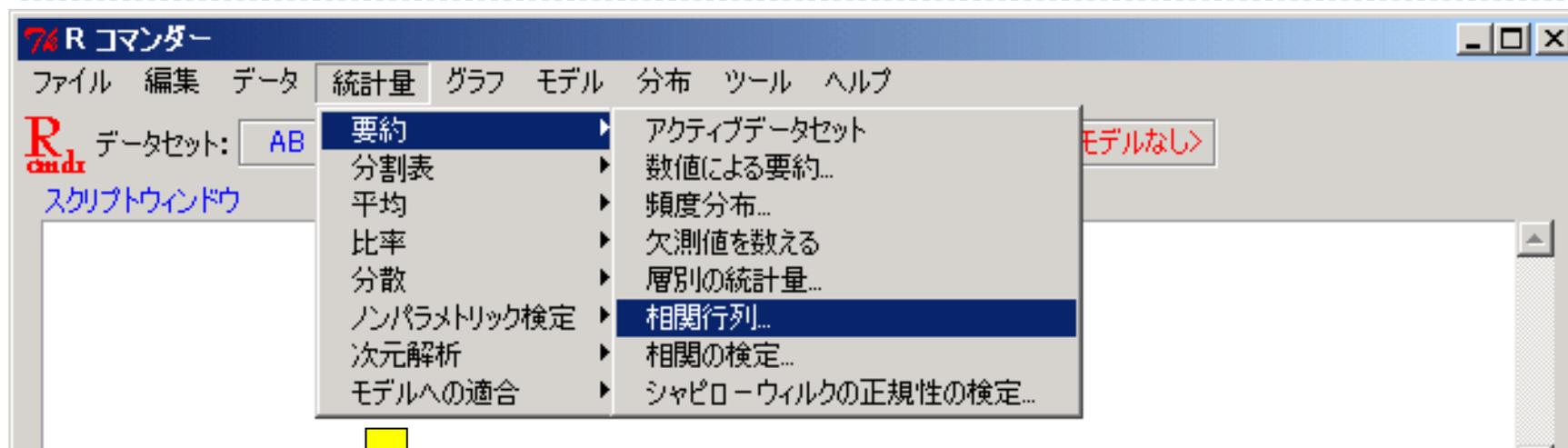


とりあえずデータセットの要約を行う

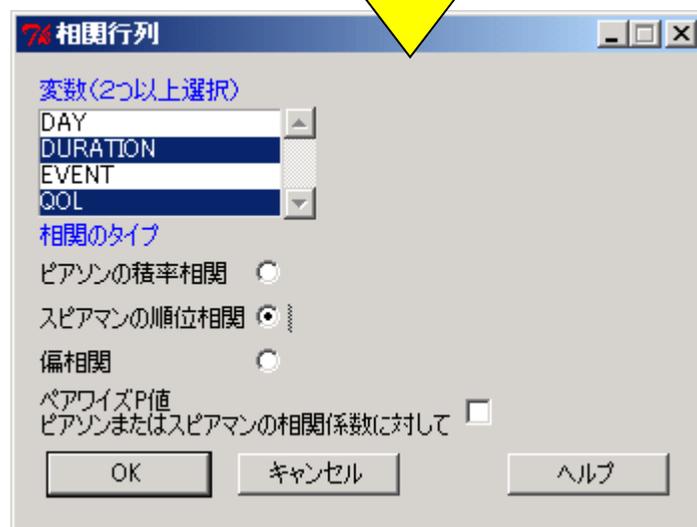
GROUP	QOL	EVENT	DAY	PREDRUG	DURATION
A:20	Min. : 0.00	Min. :1.000	Min. : 20.0	NO :20	Min. : 1.000
B:20	1st Qu.: 2.00	1st Qu.:1.000	1st Qu.: 250.0	YES:20	1st Qu.: 3.000
C: 0	Median : 4.00	Median :2.000	Median : 510.0		Median : 6.000
	Mean : 5.25	Mean :1.575	Mean : 519.5		Mean : 6.125
	3rd Qu.: 8.00	3rd Qu.:2.000	3rd Qu.: 815.0		3rd Qu.: 8.000
	Max. :15.00	Max. :2.000	Max. :1000.0		Max. :15.000



メニュー〔統計量〕 〔要約〕



データセット内の変数間の相関係数を算出する

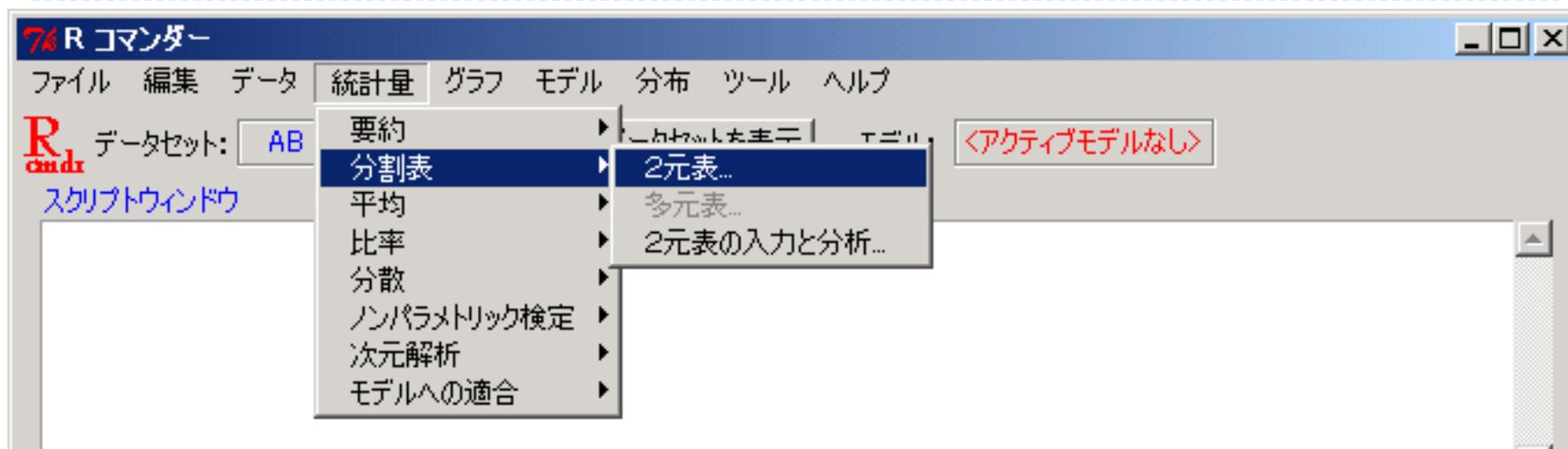


変数を 2 つ以上選択し「スピアマンの順位相関」を選択（変数を選択する際は [Ctrl] を押しながら選択）

	DURATION	QOL
DURATION	1.0000000	-0.5581124
QOL	-0.5581124	1.0000000



メニュー〔統計量〕 〔分割表〕

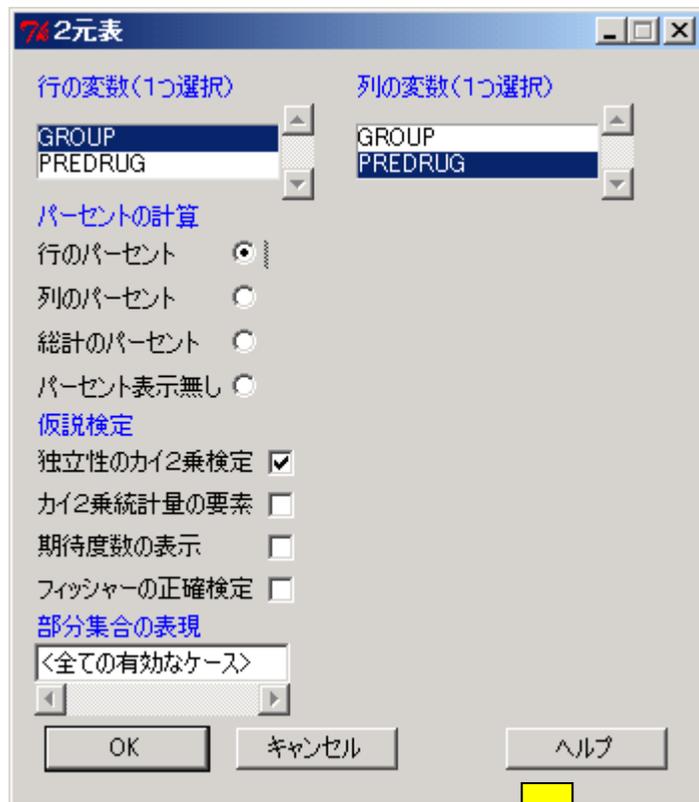


- ▶ 分割表：2 × 2 分割表（クロス表）を作成する
- ▶ 2 値データの要約や（独立性の）カイ 2 乗検定に関する内容は「2 値データの要約」を参照のこと：

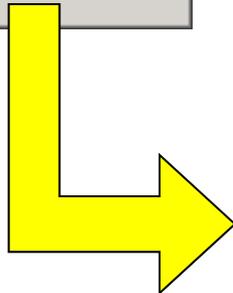
http://www.occn.zaq.ne.jp/cuhxr802/R-stat-intro_08.pdf



メニュー〔統計量〕〔分割表〕



行の変数：GROUP
列の変数：PREDRUG
パーセントの計算：行...
仮説検定：
独立性的カイ2乗検定



```
> .Table
      PREDRUG
GROUP NO YES
  A  15   5
  B   5  15

> rowPercents(.Table) # Row Percentages
      PREDRUG
GROUP NO YES Total Count
  A  75  25   100     20
  B  25  75   100     20

> .Test
      Pearson's Chi-squared test
data:  .Table
X-squared = 10, df = 1, p-value = 0.001565
```



メニュー〔統計量〕 平均, 比率, 分散

74 R コマンダー

ファイル 編集 データ 統計量 グラフ モデル 分布 ツール ヘルプ

データセット: データセットを表示

モデル: <アクティブモデルなし>

要約
分割表
平均
比率
分散
ノンパラメトリック検定
次元解析
モデルへの適合

1標本t検定...
独立サンプルt検定...
対応のあるt検定...
1元配置分散分析...
多元配置分散分析...

74 R コマンダー

ファイル 編集 データ 統計量 グラフ モデル 分布 ツール ヘルプ

データセット: データセットを表示

モデル: <アクティ>

要約
分割表
平均
比率
分散
ノンパラメトリック検定
次元解析
モデルへの適合

1標本比率の検定...
2標本比率の検定...

74 R コマンダー

ファイル 編集 データ 統計量 グラフ モデル 分布 ツール ヘルプ

データセット: データセットを表示

モデル: <アクティ>

要約
分割表
平均
比率
分散
ノンパラメトリック検定
次元解析
モデルへの適合

分散の比のF検定...
バートレットの検定...
ルビーンの検定...

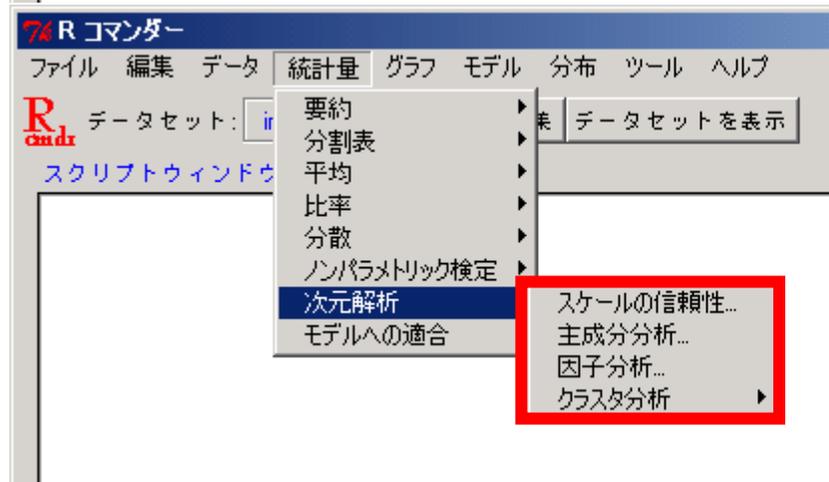
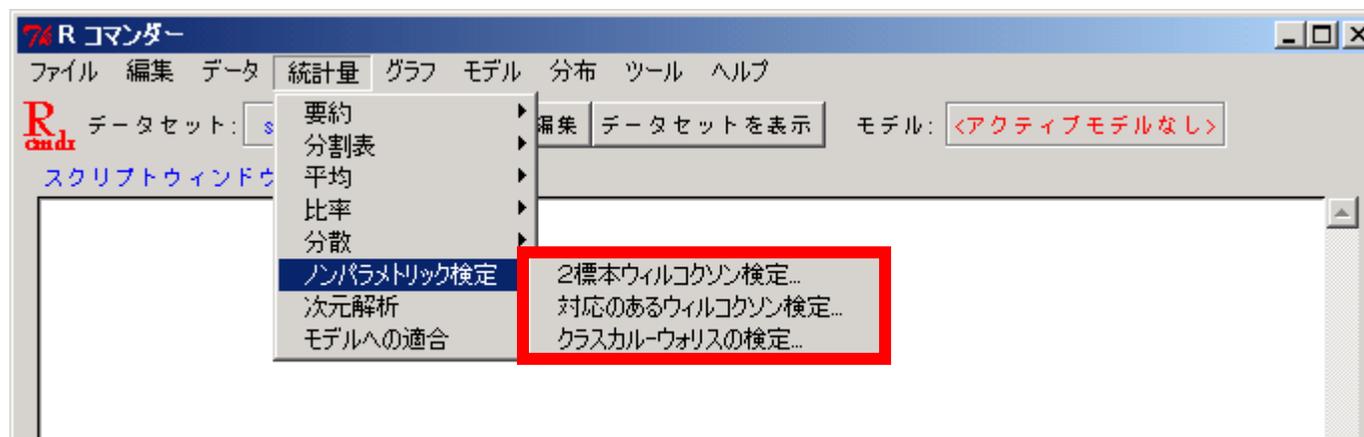
前ページと同様の操作をすることで

- 独立サンプル（2標本）t検定
- 対応のある（1標本）t検定
- 分散分析
- 比率に関する検定
- 分散の比の検定（F検定）
- バートレットの検定
- ルビーンの検定

を実行することが出来る



メニュー〔統計量〕 ノンパラ, その他の解析

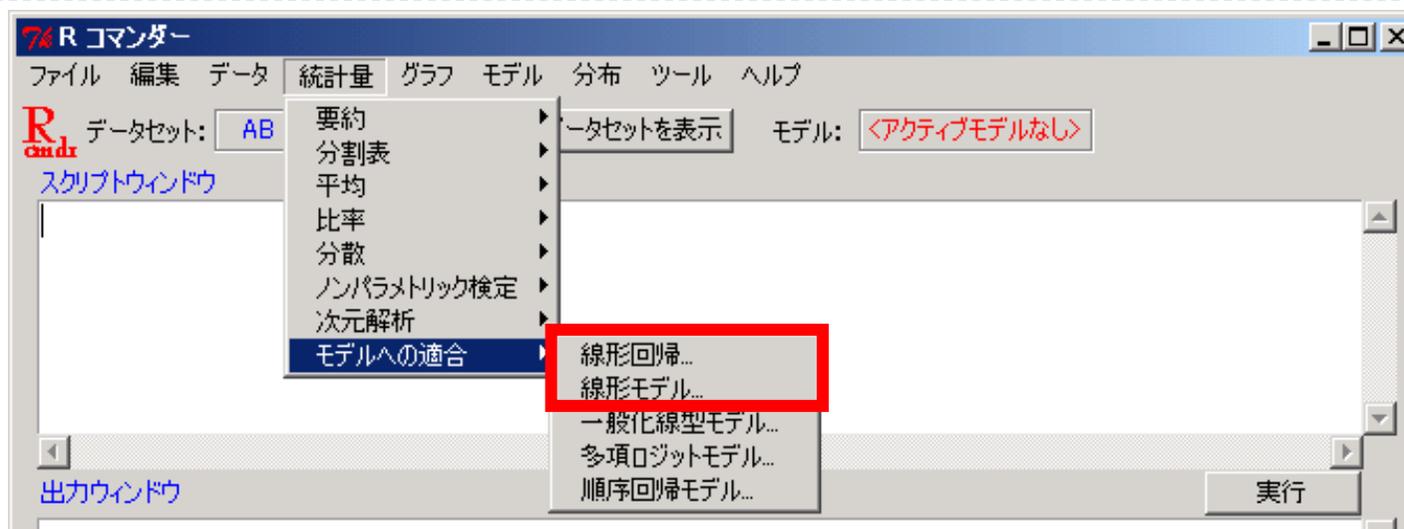


前々ページと同様の操作をすることで

- 2標本ウィルコクソン検定
 - 対応のある（1標本）ウィルコクソン検定
 - クラスカル・ウォリス検定
 - スケールの信頼性（クローンバックの α ）
 - 主成分分析
 - 因子分析
 - クラスタ分析
- を実行することが出来る



メニュー〔統計量〕 〔モデルへの適合〕



- ▶ 線形回帰：単回帰分析については「散布図と回帰直線と相関係数」を参照のこと：
http://www.occn.zaq.ne.jp/cuhxr802/R-stat-intro_04.pdf
- ▶ 線形モデル：回帰分析と分散分析については以下を参照のこと：
「2 標本 t 検定と回帰分析」
http://www.occn.zaq.ne.jp/cuhxr802/R-stat-intro_05.pdf
「分散分析と共分散分析」
http://www.occn.zaq.ne.jp/cuhxr802/R-stat-intro_06.pdf



メニュー〔統計量〕

〔モデルへの適合〕

〔線形回帰〕

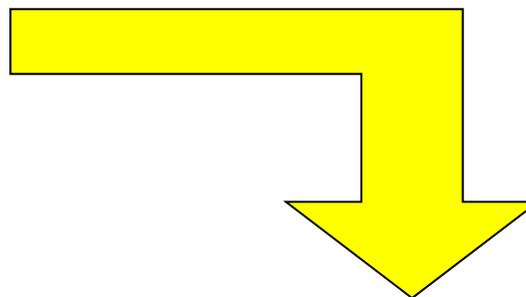


〔線形回帰〕を選択した後

目的変数：QOL

説明変数：DURATION（罹病期間）

を選択する



Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	8.9334	1.1450	7.802	2.08e-09	***
DURATION	-0.6014	0.1638	-3.672	0.000738	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.491 on 38 degrees of freedom

Multiple R-squared: 0.2619, Adjusted R-squared: 0.2424

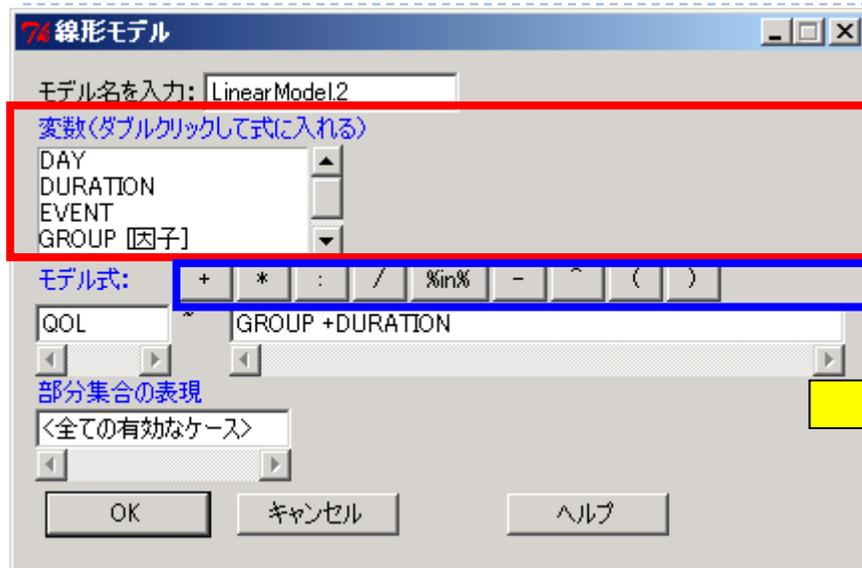
F-statistic: 13.48 on 1 and 38 DF, p-value: 0.0007379



メニュー〔統計量〕

〔モデルへの適合〕

〔線形モデル〕



変数をダブルクリックして式に追加する

演算子をクリックして式に追加する

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	9.1873	1.1646	7.889	1.92e-09	***
GROUP[T.B]	-1.2907	1.1678	-1.105	0.27619	
DURATION	-0.5375	0.1732	-3.102	0.00367	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.481 on 37 degrees of freedom

Multiple R-squared: 0.2855, Adjusted R-squared: 0.2468

F-statistic: 7.391 on 2 and 37 DF, p-value: 0.001993



【参考】モデル式を指定する書式

- ▶ モデル式の例とその意味（ ε は誤差項）

$$Y \sim X : Y = a + bX + \varepsilon$$

$$Y \sim X_1 + X_2 : Y = a + b_1X_1 + b_2X_2 + \varepsilon$$

$$Y \sim X_1 + X_2 - 1 : Y = b_1X_1 + b_2X_2 + \varepsilon \quad (\text{切片がないモデル})$$

$$Y \sim . : Y = (\text{Y 以外の変数を全て説明変数に指定}) + \varepsilon$$

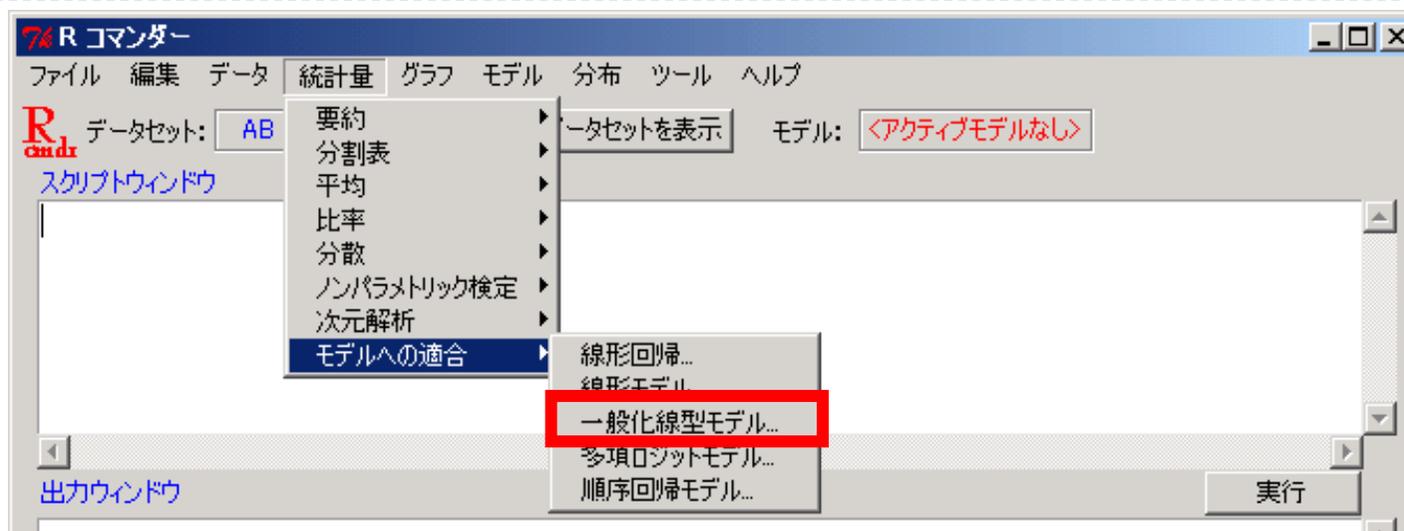
$$Y \sim X_1 * X_2 : Y = a + b_1X_1 + b_2X_2 + b_3X_1X_2 + \varepsilon \quad (\text{交互作用モデル})$$

$$Y \sim X_1 + X_2 + X_1 * X_2 : \text{上と同じ交互作用モデル}$$

$$Y \sim (X_1 + X_2)^2 : \text{上と同じ交互作用モデル}$$



メニュー〔統計量〕 〔モデルへの適合〕



- ▶ 一般化線形モデル：例えばロジスティック回帰分析を行う場合に選択する
詳しくは「ロジスティック回帰分析」を参照のこと：

http://www.occn.zaq.ne.jp/cuhxr802/R-stat-intro_09.pdf



メニュー〔統計量〕 〔モデルへの適合〕 〔一般化線形モデル〕



変数をダブルクリックして式に追加する

演算子をクリックして式に追加する

リンク関数族: **binomial**
リンク関数: **logit**
を選択することでロジスティック回帰となる

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.8367	0.5261	1.590	0.1118
GROUP[T.B]	-0.8099	0.7913	-1.024	0.3060
PREDRUG[T.YES]	-1.6523	0.7971	-2.073	0.0382 *



【再掲】モデル式を指定する書式

- ▶ モデル式の例とその意味（ ε は誤差項）

$$Y \sim X : Y = a + bX + \varepsilon$$

$$Y \sim X_1 + X_2 : Y = a + b_1X_1 + b_2X_2 + \varepsilon$$

$$Y \sim X_1 + X_2 - 1 : Y = b_1X_1 + b_2X_2 + \varepsilon \quad (\text{切片がないモデル})$$

$$Y \sim . : Y = (\text{Y 以外の変数を全て説明変数に指定}) + \varepsilon$$

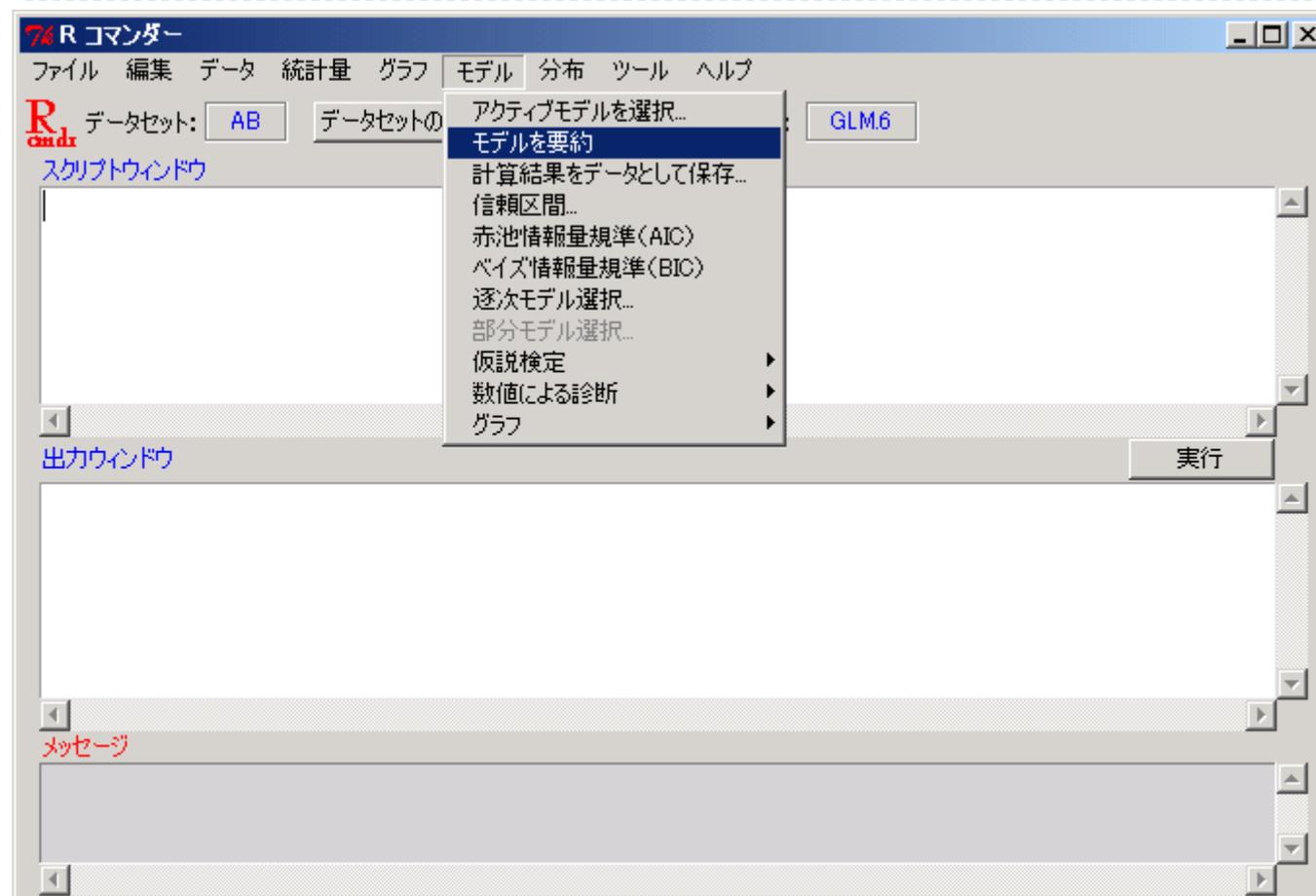
$$Y \sim X_1 * X_2 : Y = a + b_1X_1 + b_2X_2 + b_3X_1X_2 + \varepsilon \quad (\text{交互作用モデル})$$

$$Y \sim X_1 + X_2 + X_1 * X_2 : \text{上と同じ交互作用モデル}$$

$$Y \sim (X_1 + X_2)^2 : \text{上と同じ交互作用モデル}$$



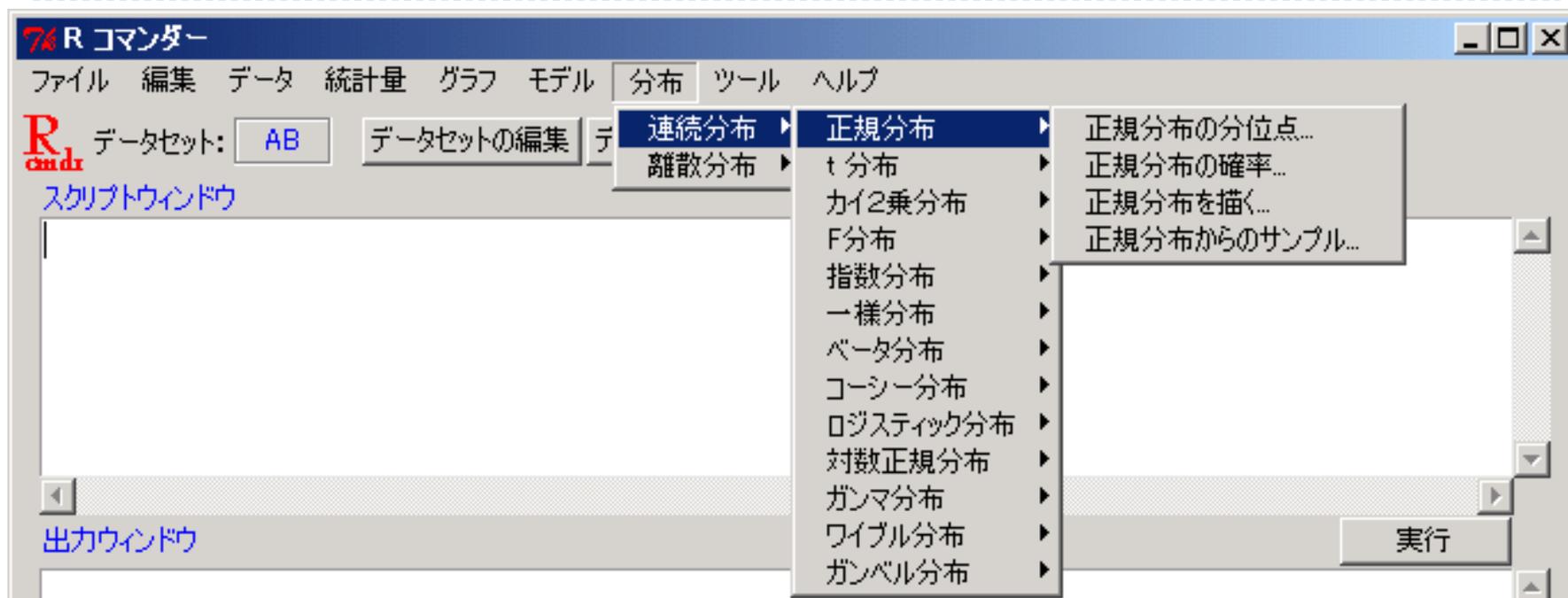
メニュー〔モデル〕



- ▶ モデルによる解析を行った後は、様々なモデルに関する情報を出力することが出来る（詳細は省略）



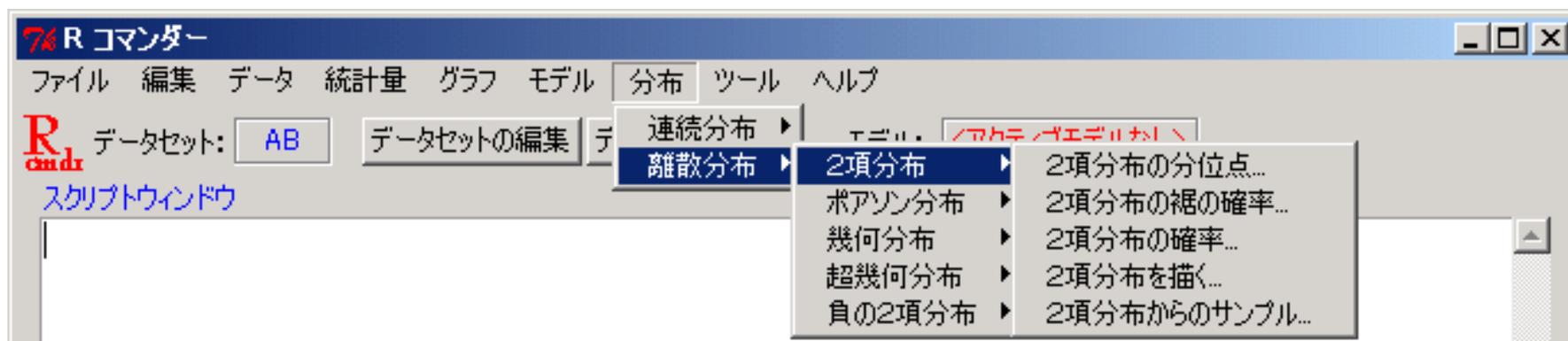
メニュー〔分布〕



- ▶ 連続分布：正規分布，t 分布， χ^2 分布，F 分布，指数分布，一様分布，ベータ分布，コーシー分布，ロジスティック分布，対数正規分布，ガンマ分布，ワイブル分布，ガンベル分布（二重指数分布）について...
 - 累積分布の算出，確率点の算出，乱数の算出，グラフの描画を行う



メニュー〔分布〕



- ▶ 離散分布：2 項分布，ポアソン分布，幾何分布，超幾何分布，負の 2 項分布について...

累積分布の算出，確率点の算出，確率，乱数の算出，グラフの描画を行う



〔分布〕 正規分布の（裾の）確率と分位点を計算

正規確率

変数の値 1.96

μ (平均) 0

σ (標準偏差) 1

下側確率

上側確率

OK キャンセル ヘルプ

正規分布の分位点

確率 0.025

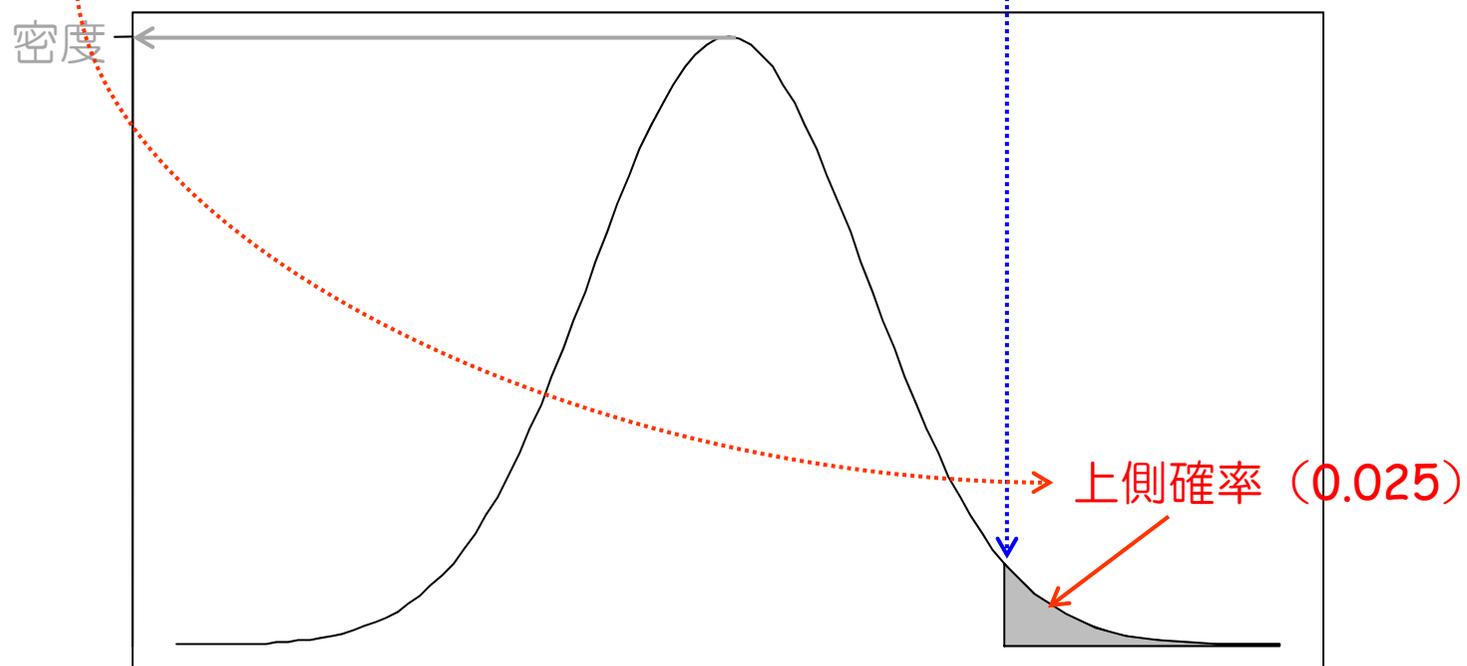
μ (平均) 0

σ (標準偏差) 1

下側確率

上側確率

OK キャンセル ヘルプ





〔分布〕 二項分布の確率と確率分布の図を生成

2項確率

試行回数 10
成功の確率 0.5

OK キャンセル ヘルプ

2項分布

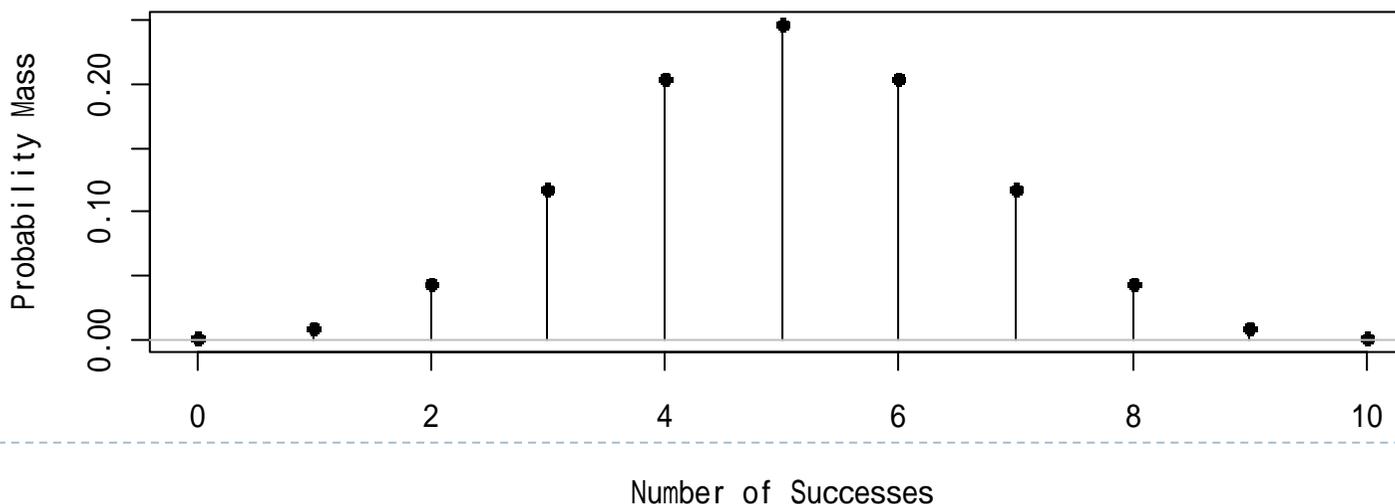
試行回数 10
成功の確率 0.5

確率関数のプロット
分布関数のプロット

OK キャンセル ヘルプ

Pr	
0	0.0009765625
1	0.0097656250
2	0.0439453125
3	0.1171875000
4	0.2050781250
5	0.2460937500
6	0.2050781250
7	0.1171875000
8	0.0439453125
9	0.0097656250
10	0.0009765625

Binomial Distribution: Trials = 10, Probability of success = 0.5





本日のメニュー

1. R のセットアップ (Ver. 2.12.2) のメモ

2. R Commander の基礎

- ▶ イントロ
- ▶ データの読み込みと変数に対する処理
- ▶ データのグラフ化
- ▶ データ解析

3. パッケージによる機能追加



パッケージとは

- ▶ R は関数とデータを機能別に分類して「パッケージ」という形で用意
- ▶ どのようなパッケージがあるのかは関数 `library()` を実行すると表示

パッケージ名	解説
boot	ブートストラップに関するパッケージ
foreign	R 以外のデータファイルを読み込むためのパッケージ
lattice	ラティス・グラフィックス関数パッケージ
nlme	線形&非線形混合効果モデル用のパッケージ
nnet	ニューラル・ネットワーク用のパッケージ
rpart	CART に関するパッケージ
splines	スプライン回帰用のパッケージ
survival	生存時間解析用のパッケージ

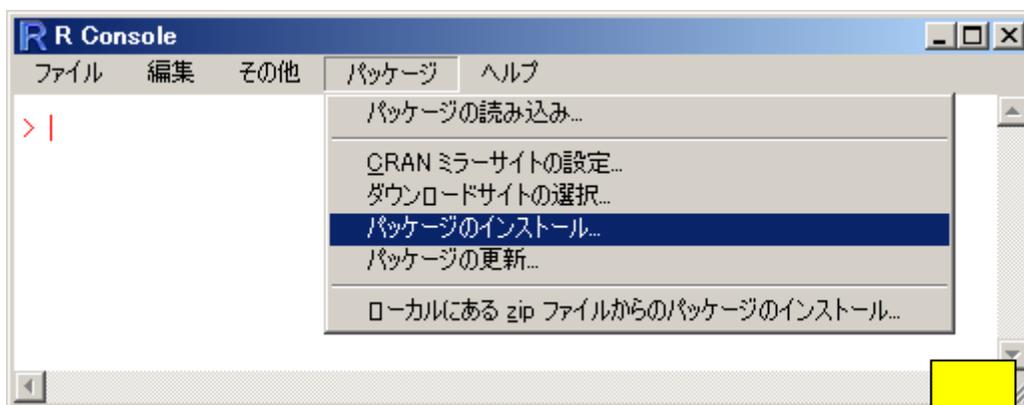


追加パッケージのインストール

- ▶ コマンドでパッケージ「 RcmdrPlugin.SurvivalT 」をインストール

```
> install.packages("RcmdrPlugin.SurvivalT", dep=T)
```

- ▶ R Console からパッケージ「 RcmdrPlugin.SurvivalT 」をインストール



- ① R Console のメニュー「パッケージ」から「パッケージのインストール」を選択
- ② 「Japan(Tsukuba)」 [OK] をクリック
- ③ インストールするパッケージ（ RcmdrPlugin.SurvivalT ）を選択して[OK] をクリック



パッケージの呼び出し

- ▶ コマンドでパッケージ「RcmdrPlugin.SurvivalT」を呼び出す場合：

```
> library(RcmdrPlugin.SurvivalT) # パッケージを呼び出す  
> library(help="RcmdrPlugin.SurvivalT") # パッケージのヘルプ
```

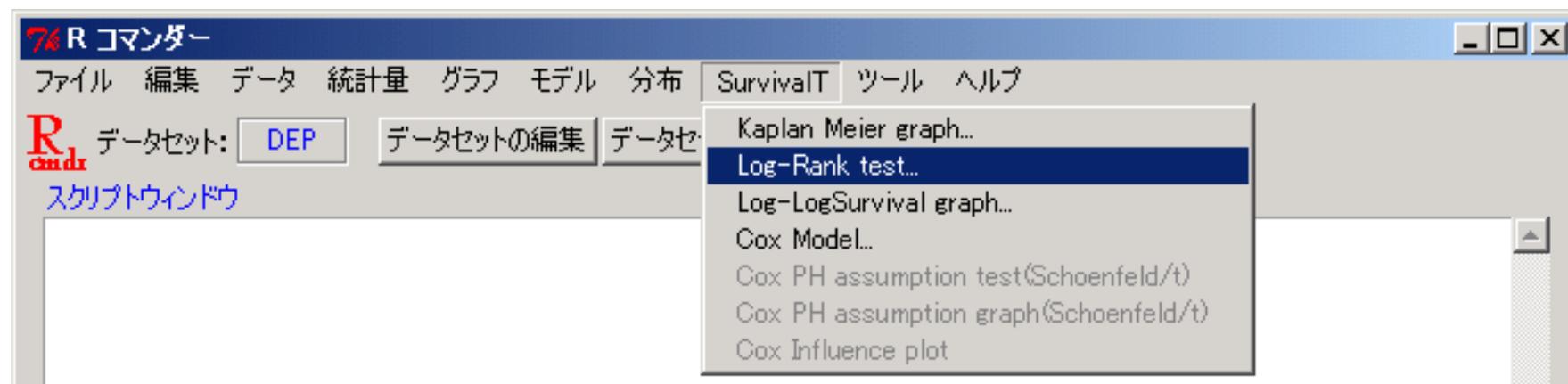
- ▶ メニューからパッケージ「RcmdrPlugin.SurvivalT」を呼び出す場合：

The screenshot shows the R Commander interface. The 'Tools' menu is open, and the 'Rcmdrプラグインのロード...' option is selected. A yellow arrow points from this option to the 'Load Plug-ins' dialog box. The dialog box shows a list of plugins, with 'RcmdrPlugin.SurvivalT' selected.

- ① メニュー〔ツール〕から「Rcmdr プラグインのロード」を選択
- ② 読み込むパッケージ名を選択して [OK] を選択
(今回は RcmdrPluginSurvivalT を選択)



パッケージ「RcmdrPlugin.SurvivalT」



- ▶ 生存時間解析に関するメニューが追加される
- ▶ カプラン・マイヤー法とログランク検定については「生存時間解析〔前篇〕」を参照のこと：

http://www.occn.zaq.ne.jp/cuhxr802/R-stat-intro_11.pdf

- ▶ Cox 回帰分析については「生存時間解析〔中篇〕」を参照のこと：

http://www.occn.zaq.ne.jp/cuhxr802/R-stat-intro_12.pdf



R Commander 用のパッケージの一部

パッケージ名	解説
RcmdrPlugin.coin	データ解析手法の追加（中央値の検定等）
RcmdrPlugin.epack	時系列解析用の手法
RcmdrPlugin.Export	解析結果を LaTeX や html のソースに変換
RcmdrPlugin.HH	データ解析手法やグラフの種類を追加
RcmdrPlugin.KMggplot2	綺麗なグラフ（ Kaplan-Meier プロット等）
RcmdrPlugin.mosaic	モザイクプロットや assoc plot を描く機能を追加
RcmdrPlugin.MAc RcmdrPlugin.MAd	メタアナリシス用のデータ解析手法
RcmdrPlugin.qcc	品質管理に関する解析手法
RcmdrPlugin.TeachingDemos	統計に関する楽しいデモンストレーション
RcmdrPlugin.TextMining	テキストマイニング用の手法



本日のメニュー

1. R のセットアップ (Ver. 2.12.2) のメモ
2. R Commander の基礎
 - ▶ イントロ
 - ▶ データの読み込みと要約
 - ▶ データのグラフ化
 - ▶ データ解析
3. パッケージによる機能追加



参考文献

- ▶ R コマンダー入門 (John Fox 著, 荒木先生訳)
<http://www.ec.kansai-u.ac.jp/user/arakit/documents/Getting-Started-with-the-Rcmdr-ja.pdf>
- ▶ R Commander によるデータ解析 (大森 崇 他著, 共立出版)
- ▶ R と R コマンダーではじめる多変量解析
(荒木 孝治 他著, 日科技連出版社)
- ▶ R と R コマンダーではじめる実験計画法
(荒木 孝治 他著, 日科技連出版社)
- ▶ EZR でやさしく学ぶ統計学 ~ EBM の実践から臨床研究まで ~
(神田 善伸 著, 中外医学社)
- ▶ R Commander の説明書 (ただし古めの情報)
<http://cwoweb2.bai.ne.jp/~jgb11101/files/R-commander.pdf>
- ▶ R で統計解析入門
<http://cwoweb2.bai.ne.jp/~jgb11101/files/R-stat-intro/index.html>

Rで統計解析入門

終