1- 1 データ形式とは …… 1p
1- 2 データの貼り付け法 …… 3p
1- 3 データシートの基本操作 …… 5p
1- 4 群分け・統合 …… 7p
1- 5 ツールバーの機能 …… 9p
2- 1 システム設定 …… 11p
2- 2 変数情報設定 …… 13p
2- 3 群別描画条件の設定 …… 15p
2- 4 グラフ設定（座標・打点）…… 17p
2- 5 グラフ設定（座標・打点）…… 18p
2- 6 グラフ設定（一変量図）…… 19p
2- 7 グラフ設定（二変量図）…… 20p
2- 8 グラフ設定（統計量表示設定）…… 21p
2- 9 グラフ設定（全般・その他）…… 22p
3- 1 クロス集計ツール …… 23p
3- 2 ビジュアルクロス集計 …… 25p
3- 3 二変量統計 …… 27p
3- 4 分布型解析 …… 28p
3- 5 判断分析 …… 29p
3- 6 ダミー変数の作成 …… 31p
3- 7 重回帰分析（1）…… 33p
3- 8 重回帰分析（2）…… 35p
3- 9 多重ロジスティック分析（1）…… 37p
3-10 多重ロジスティック分析（2）…… 39p
3-11 傾向スコア分析…… 41p
3-12 Cox 比例ハッザード回帰分析（1）…… 43p
3-13 Cox 比例ハッザード回帰分析（2）…… 45p
3-14 生存曲線分析…… 46p
3-15 因子分析…… 47p
4- 1 グラフを編集ウインドウへ送る …… 49p
4- 2 グラフ編集ウインドウ …… 51p
4- 3 他ソフトウェアへのグラフの貼り付け …… 53p
5- 1 統計情報ウインドウ …… 54p



企画・開発

株式会社 医学統計研究所
〒755-0001 山口県宇部市沖宇部272-5
TEL.0836-35-5213　FAX.0836-35-5213

総販売元

株式会社 アーテック
〒532-0004 大阪市淀川区西宮原１丁目8-14-814
TEL.06-6915-5080（代）　FAX.06-6915-5081
E-mail:statflex@statflex.net


豊富な新機能を搭載して、「ビジュアル統計ソフト」がさらに進化

# StatFlex Ver.7

# スタットフレックス

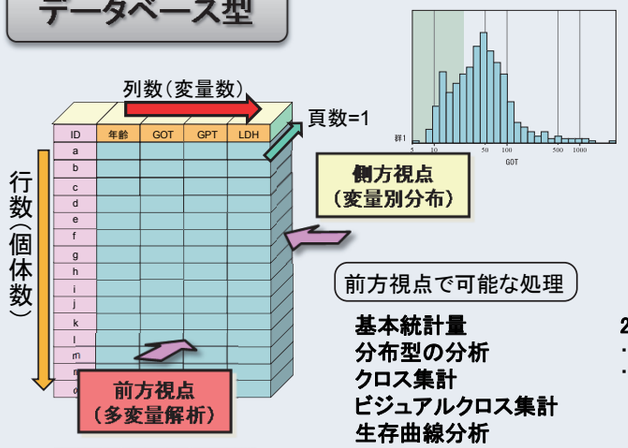
## 図説要点ガイド

http://www.statflex.net

# データ形式とは



## データベース型




側方視点で可能な処理

**基本統計量**

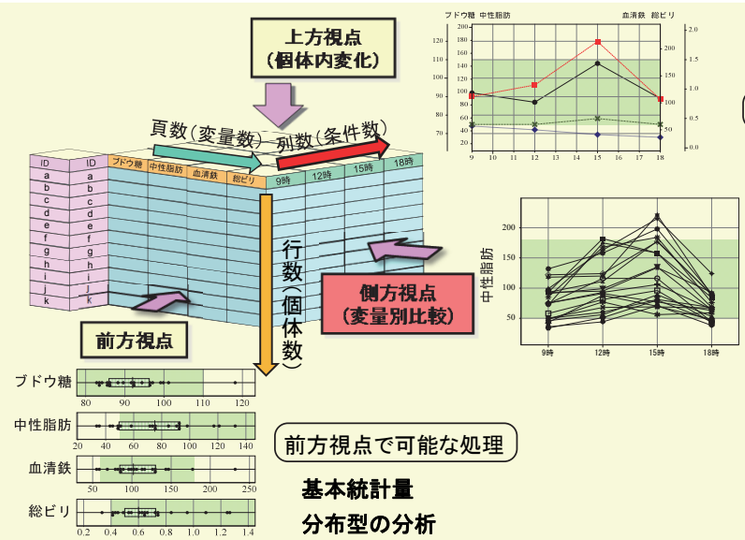
前方視点で可能な処理

**基本統計量**
**分布型の分析**
**クロス集計**
**ビジュアルクロス集計**
**生存曲線分析**

**多変量解析**
- 相関行列計算
- 重回帰分析
- 多重ロジスティック回帰分析
- Cox比例ハザード回帰分析
- 主成分分析
- 因子分析
- 1～3段階分散分析
- 階層クラスター分析

**2変量統計**
- 直線回帰
- 線形関係式
  - Deming回帰
  - 標準主軸回帰
  - Passing-Bablok回帰
- Altmanの偏差図
- Bootstrap法による回帰係数の標準誤差計算

## 関連多群型




側方視点で可能な処理

**基本統計量**

**関連2群・多群間の差の検定**
- 1標本t検定
- 符号検定
- 1標本wilcoxon検定
- 二元配置分散分析
- Friedman検定
- Bartlett検定
- 繰り返しのある二元配置分散分析
- 三元配置分散分析

**多重比較法**
- Tukey検定
- Scheffe検定
- Dunnett検定
- Dunn検定（2種）
- Newman-Keuls検定

前方視点で可能な処理

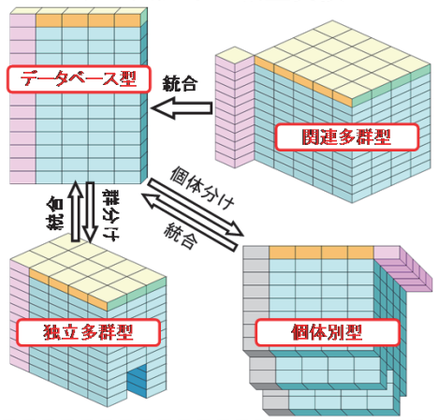
**基本統計量**
**分布型の分析**
**ビジュアルクロス集計**

**多変量解析**
**データベース型前方視点で可能な処理、および**
- Hotelling $T^2$検定
- 直線回帰係数の差の検定
- 2要因分散分析混合デザイン

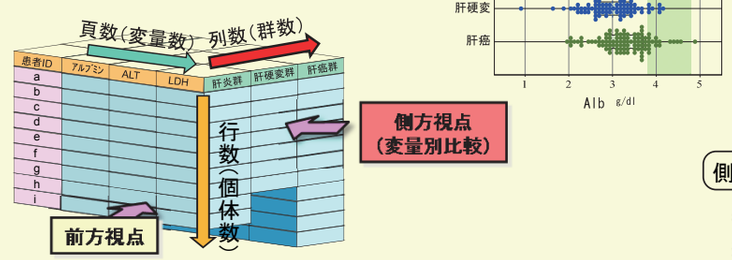
上方視点で可能な処理

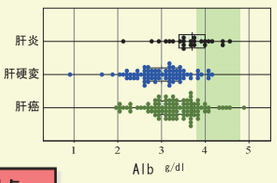
**基本統計量**
**分布型の分析**
**多変量解析**
**データベース型前方視点で可能な処理**

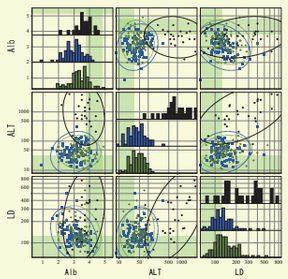
## データ形式 の相互変換




## 独立多群型




前方視点で可能な処理

**基本統計量**
**分布型の分析**
**クロス集計**
**ビジュアルクロス集計**
**生存曲線分析**
**多変量解析**
**データベース型前方視点で可能な処理、および**
- Hotelling $T^2$検定
- 回帰係数の差の検定
- 2要因分散分析混合デザイン

側方視点で可能な処理

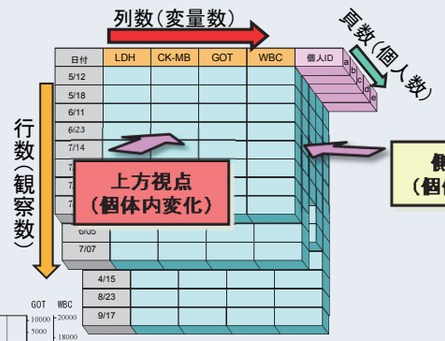
**基本統計量**

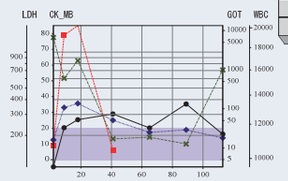
**独立2群・多群間の差の検定**
- 2標本t検定
- Mann-Whitney U検定
- F検定
- 一元配置分散分析
- Kruskal-Wallis検定
- Bartlett検定

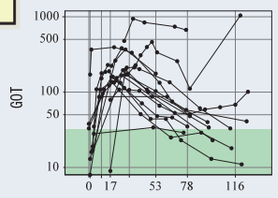
**多重比較法**
- Tukey検定
- Scheffe検定
- Dunnett検定
- Dunn検定（2種）
- Newman-Keuls検定

**ROC分析**

## 個体別型




上方視点で可能な処理

**基本統計量**
**分布型の分析**
**多変量解析**
**データベース型前方視点で可能な処理**

側方視点で可能な処理

**基本統計量**

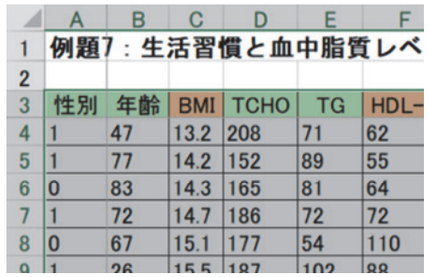
**独立2群・多群間の差の検定**
- 2標本t検定
- Mann-Whitney U検定
- F検定
- 一元配置分散分析
- Kruskal-Wallis検定
- Bartlett検定

**多重比較法**
- Tukey検定
- Scheffe検定
- Dunnett検定
- Dunn検定（2種）
- Newman-Keuls検定

※独立2群・多群間の差の検定と多重比較法は、個体を群と見なして処理します。

# 1-2 データの貼り付け法

## 手順1 Excelのデータをコピー



| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | 例題7：生活習慣と血中脂質レベ | | | | | |
| 2 | | | | | | |
| 3 | 性別 | 年齢 | BMI | TCHO | TG | HDL- |
| 4 | 1 | 47 | 13.2 | 208 | 71 | 62 |
| 5 | 1 | 77 | 14.2 | 152 | 89 | 55 |
| 6 | 0 | 83 | 14.3 | 165 | 81 | 64 |
| 7 | 1 | 72 | 14.7 | 186 | 72 | 72 |
| 8 | 0 | 67 | 15.1 | 177 | 54 | 110 |
| 9 | 1 | 26 | 15.5 | 187 | 102 | 88 |

貼り付けたい領域を選択し、コピー

マウスで操作：
領域をドラッグで選択→右クリック→「コピー」

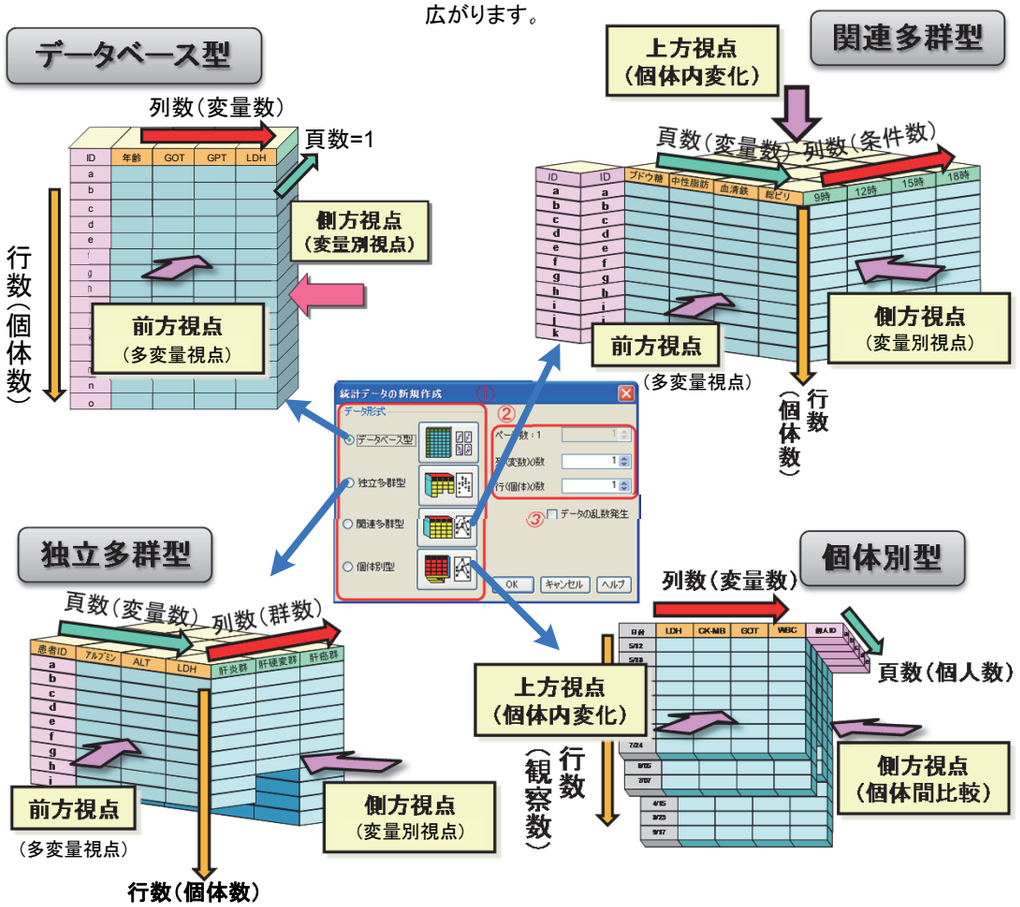
キーボードで操作：
表内の任意のセルに移動→ Ctrl + A → Ctrl + C
表を全選択　選択範囲をコピー

## 手順2 データシートを新規作成

①データ形式を選択
データ形式については下図を参照。

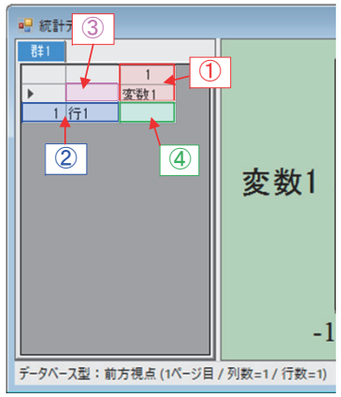
②頁、列、行数を指定
全て1のままで、シートは自動的に広がります。

③のチェックは外す

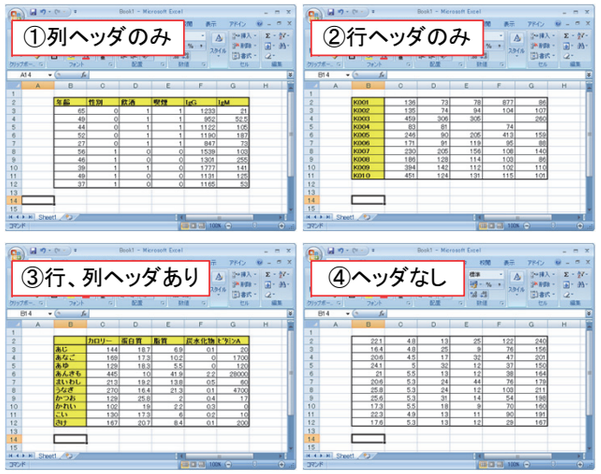



## 手順3 データシートへの貼り付け

データシートへの貼り付け位置

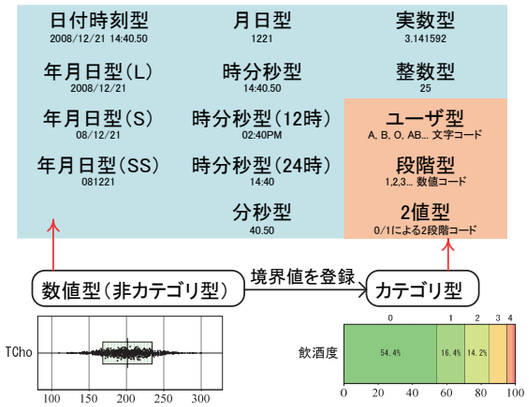



コピー元のデータによって異なります




## 手順4 変数型の確認

StatFlexは貼り付けられたデータの型を自動判定します。
もし、意に反する判定の場合は、変数設定1（下図）で変数型を修正します。




※文字型はグラフを描画できません

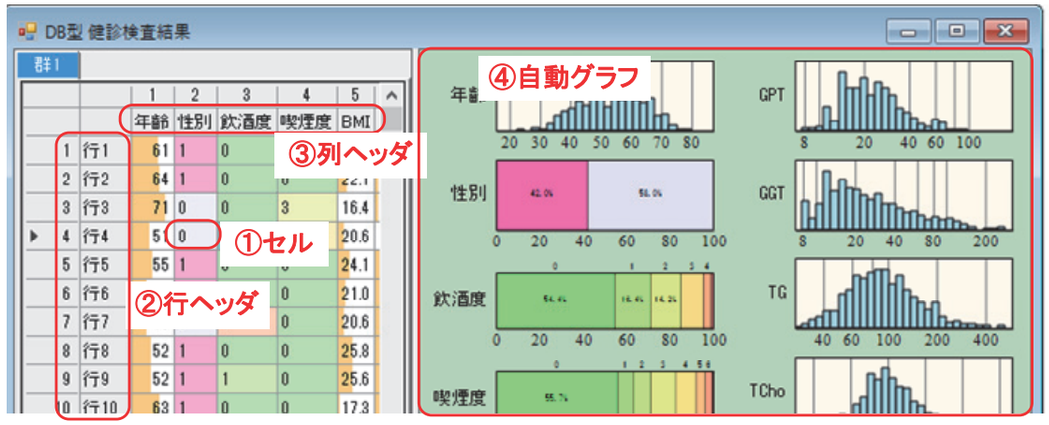
### 貼り付けエラーが発生した時は…

数値型と思われる変数(列)に、数値に変換できない文字型のデータが、10%未満と低い頻度で存在する場合に、エラーと判定します。

①エラー一覧が出ます　②正しい値に編集

④変数型を自動修正
修正しなければ、文字型として登録しますが、後で、Var1 で訂正可能です。

# 1-3 データシートの基本操作



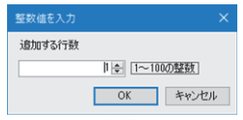

データシートでは、任意の位置を右クリックすることで、主な操作を実行できます。このページでは、データシート上の各位置での右クリックメニューを解説します。

## ① セル の右クリックメニュー

| | |
|---|---|
| コピー Ctrl + C | …クリップボードへコピー |
| 切り取り Ctrl + X | …コピーした後にクリア |
| 貼り付け Ctrl + V | …クリップボードから貼り付け |
| クリア | …選択したセルを欠測値にする |

## ② 行ヘッダ の右クリックメニュー

| | |
|---|---|
| コピー Ctrl + C | |
| 切り取り Ctrl + X | |
| 貼り付け Ctrl + V | |
| クリア | |
| 追加 | …終端に指定行数分を追加 |
| 挿入 | …選択位置に行を挿入 |
| 削除 | …選択行の削除 |




行数指定パネル

## ③ 列ヘッダ の右クリックメニュー
（前方・上方視点）

データ視点によって、列が変数である場合（前方・上方視点）と、列が群である場合（側方視点）があり、側方視点では「集計」「並べ替え」のみ利用できます。

**カテゴリ型の場合**

| | |
|---|---|
| コピー | Ctrl + C |
| 切り取り | Ctrl + X |
| 貼り付け | Ctrl + V |
| クリア | |
| 追加 | |
| 挿入 | |
| 削除 | |
| 表示しない | |
| 全て表示する | |
| 集計 | → A |
| 並べ替え | → B |
| フィルタ | → C |
| 変数設定 | → D |
| 変数変換 | → E |
| この変数で群分け | → G |
| カテゴリ情報設定 | → H |
| ダミー変数 | → I |
| 行データを縦に表示 | |

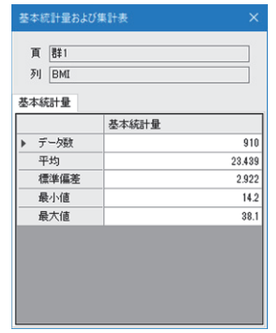
**数値型**

| | |
|---|---|
| コピー | Ctrl + C |
| 切り取り | Ctrl + X |
| 貼り付け | Ctrl + V |
| クリア | |
| 追加 | |
| 挿入 | |
| 削除 | |
| 表示しない | |
| 全て表示する | |
| 集計 | → A |
| 並べ替え | → B |
| フィルタ | → C |
| 変数設定 | → D |
| 変数変換 | → E |
| 境界値設定 | → F |
| この変数で群分け | → G |
| ダミー変数 | → I |
| 行データを縦に表示 | |

赤枠内は、境界値を登録済みの場合

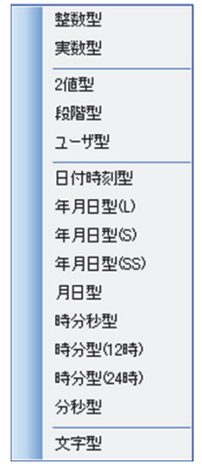
### B 並べ替え

### A 列集計パネル

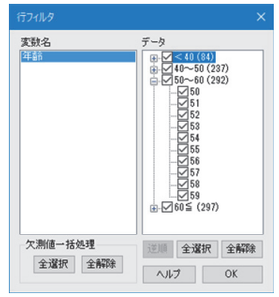



カテゴリ型では、カテゴリ集計にも対応

### E 変数変換

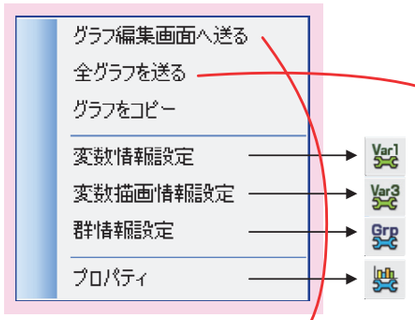



### C 行フィルタパネル




特定の値を持つ行を除外します

## ④ 自動グラフ の右クリックメニュー

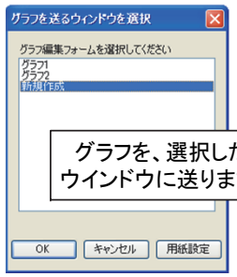



グラフに関する設定と、グラフ編集に送るための機能を提供します。

### 全グラフを送る

複数のページにあるグラフを、自在にレイアウトして、まとめてグラフ編集画面に送ります。

### グラフ編集画面へ送る

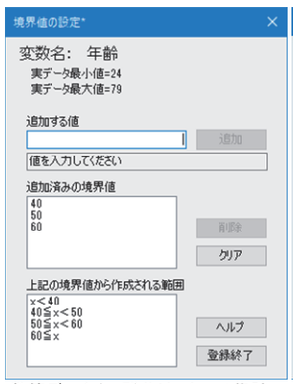



グラフを、選択したグラフ編集ウインドウに送ります。

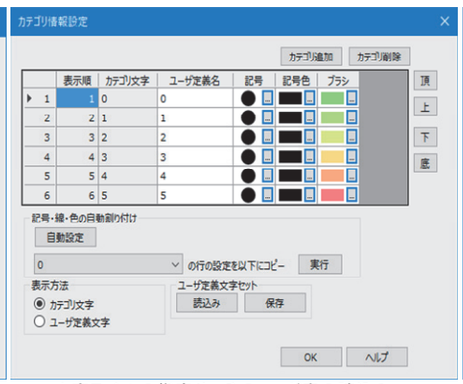
## D 変数基本情報設定パネル

数値データを、群分け、クロス集計、ダミー変数の作成に利用できるようにします

## F 境界値設定パネル

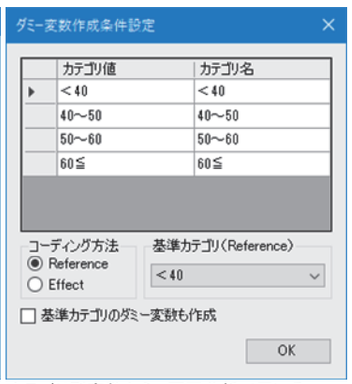



## H カテゴリ情報設定パネル




文字ラベルの指定や、そのコード表の読込み、帯グラフの並びや色を設定します

## I ダミー変数作成パネル




カテゴリ型変数から、要因分析で用いるダミー変数を自動作成します

# 1-4 群分け・統合

## 手順1 群分け可能なデータを用意する

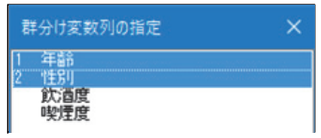
群分けを行うには、次の2つの条件が必要です。

**① データベース型である**
**② 群分け可能な列がある**

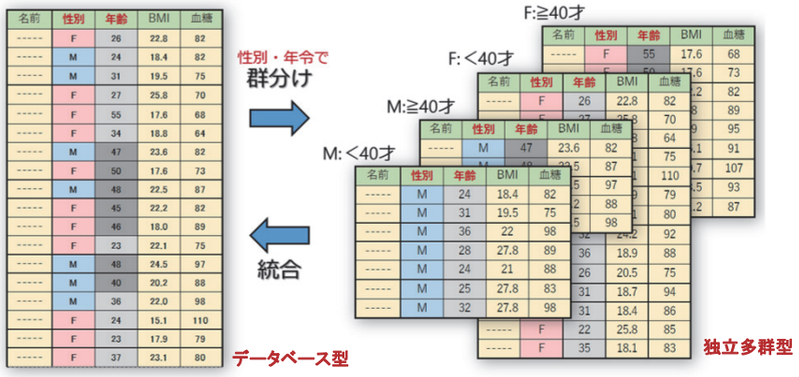
群分けに利用できる変数は、
・「ユーザ型」「段階型」「2値型」など、**カテゴリ情報**（コード化されたデータ）で構成されている変数です。
・「数値型」「時間型」の場合は、境界値を設定してカテゴリ情報を持たせると、群分けに利用できます。

## 手順2 群分けに用いる変数を選択する

群分けボタン を押すと、**群分けパネル（右図）**に群分けに利用可能な変数が表示されます。




下の例では、ユーザ型である性別と、40歳を境界値とした年齢を群分け基準列としています。群分け基準として、性別が2通り、年齢が2通りになりますので、組み合わせで4通りの群に分けられます。

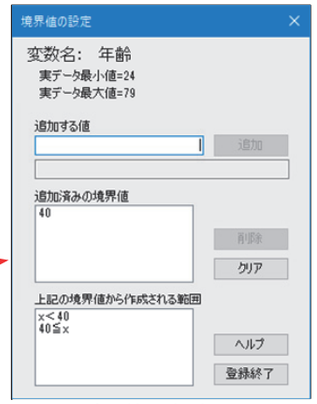



### 境界値設定について

「整数型」、「実数型」などの**数値型**、「日付時刻型」、「月日型」などの**時間型**など、連続値のデータは、カテゴリ情報を持ちません。ただし、変数設定1パネル内の「境界値」を設定することで、カテゴリ情報を作成することができ、群分け機能、ダミー変数の作成機能、データ列の生成機能などを使用できます。

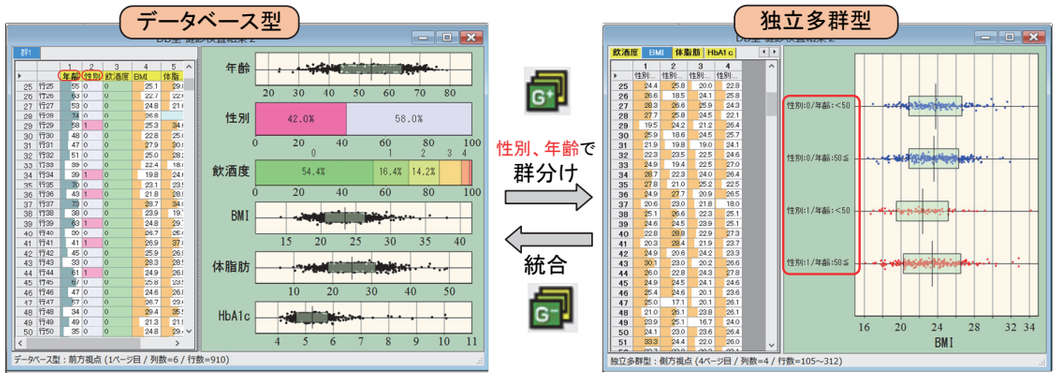
変数基本設定パネル

**境界値で区切られる各群は、下限を含み、上限を含みません。**右の例では、40を指定すると、40未満と40以上の群に分かれます。



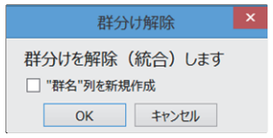

境界値設定パネル

## 実行例



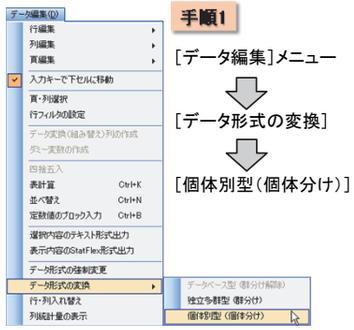

前方視点（多変量視点）　　側方視点（単変量視点）

統合ボタンを押すと、右のような確認が出ます。
最初から独立多群型でデータを新規作成した場合、**[群名列を新規作成]**をチェックします。これは、群分けの基準変数が存在せず、群情報が、失われるためです。**データベース型から群分けした場合、チェックは不要です。**



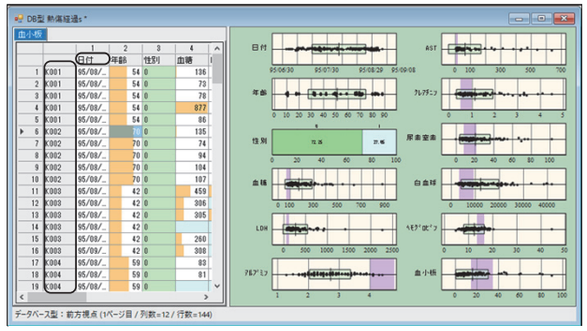

### 個体分け

データベース型のデータは、個体を表す変数と時系列を表す変数を指定すると、個体別型に変換できます。
個体分けを行うと、個体毎に頁分割され、個体別型上方視点になります。



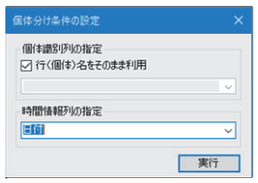

**手順1**
[データ編集]メニュー
↓
[データ形式の変換]
↓
[個体別型（個体分け）]

図は、ヘッダ列にIDが有り、日付があるデータです。

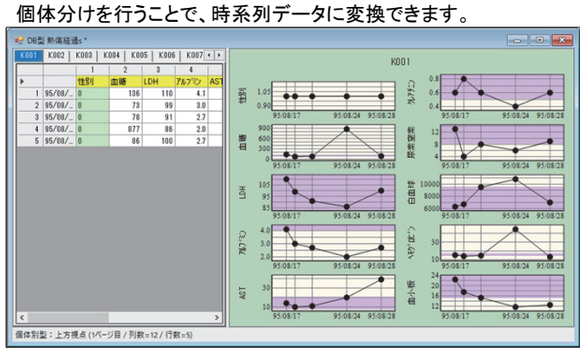


**手順2** 変数を指定

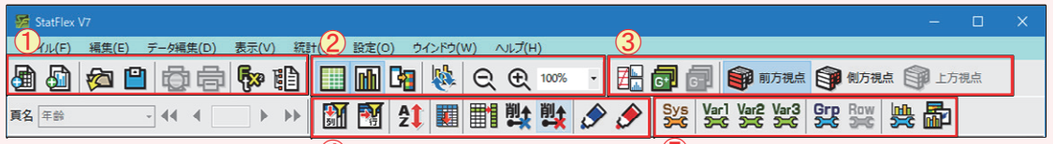
下図のような設定パネルで、個体を表す列と、時系列を表す列を指定します。




個体分けを行うことで、時系列データに変換できます。



## ■ メインウィンドウ




**① 基本操作ツールバー**
3種のウインドウ(データシート、グラフ編集、統計情報)の新規作成、ファイルの入出力、印刷など、基本操作用ボタン群です。

**② データ表示ツールバー**
データの表示切り換え用で、データシートと自動グラフの表示切り換え、データシートの文字サイズの変更用のボタンが配置されています。

**③ データ分析ツールバー**
ビジュアルクロス集計や群分け機能でデータを組み替えたり、3次元データシートの視点変更により、データを分析するためのボタン群です。視点変更ボタンの動作は、データ形式で異なります。

**④ データ操作ツールバー**
頁・列の絞り込み、行の絞り込み(オートフィルター)、並べ換えの操作、および特定のデータをマークしたり、一時削除するためのボタン群です。

**⑤ 設定ツールバー**
システム設定、変数設定、群設定、グラフ設定など、設定用ボタン群が配置されています。
スパナのマークが目印です。

- Sys 実行環境の設定: 統計情報出力法、データシート表示法、ファイルの作成法
- Var1 変数設定1: 変数・単位名、変数型、仕切り値、べき乗変換・小数位の登録
- Var2 変数設定2: グラフ表示域、度数分布図分割数、3つの参照域の登録
- Var3 変数設定3: 変数別に描画記号・線種の設定
- Grp 群情報設定: 群名の登録, 群別の描画記号、線や色の設定
- Row 行情報の設定: 行(個体)名の登録、行(個体)別の描画記号、線や色の設定
- グラフ設定: グラフの描画オプションで、計6 頁の設定パネルから構成
- グラフの種類設定: グラフの種類指定で、統計量表示や縦書き・横書きの切換

## ■ グラフ編集ウィンドウ 自動グラフや図形(直線、長方形、楕円、ラベル)を総称してオブジェクトと呼びます

- オブジェクトを選択するモードに切替
- 直線描画モードに切替
- 長方形描画モードに切替
- 楕円描画モードに切替
- ラベル描画モードに切替
- 図形(直線/長方形/楕円/ラベル)の設定
- 凡例の表示/非表示を切替
- 選択しているオブジェクトのグラフ書式をコピー
  (カーソルが変化した状態で、別のオブジェクトを選択すると、書式が貼り付けられます)
- グラフ書式のうち、貼り付け対象とする項目を限定
- 選択したオブジェクトの色を反転
- 選択したオブジェクトの自動配置
  (複数のオブジェクトを、均等配置する場合に使用)
- 直前操作の取消し(Undo)
- Undo の取消し(Redo)
- 複数のオブジェクトをグループ化
- グループ化されたオブジェクトのグループ化を解除
- 選択したオブジェクトを最前面に移動
- 選択したオブジェクトを一つ前面に移動
- 選択したオブジェクトを一つ後面に移動
- 選択したオブジェクトを最後面に移動
- 全てのオブジェクトを再描画
  (描画の乱れを修正)
- グリッドの表示/非表示を切替
- グラフ編集画面の設定
- 印刷用紙の設定

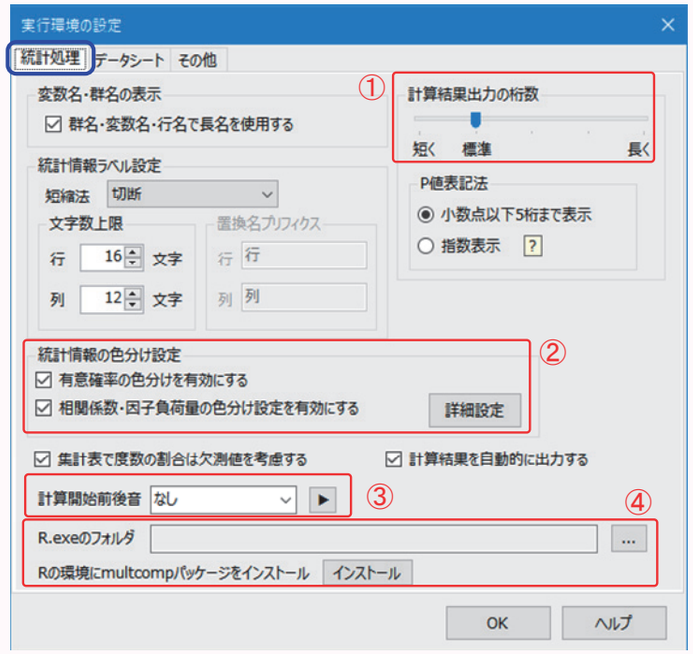
## ■ 統計情報ウインドウ

- 全ての統計情報を削除
- 現在開いている統計情報を閉じ、ファイルから統計情報を読込む
- 現在開いている統計情報に、ファイルから統計情報を追加読み込み
- チェックしている出力ブロックのみ保存
- ツリー上の出力ブロック全てを上書き保存
- 直前操作の取消し(Undo)
- Undoの取消し(Redo)
- 上下ブロックへ移動
- 統計情報ウインドウに関する設定

## StatFlexの主なショートカットキー

| | | | |
|---|---|---|---|
| Ctrl + C コピー | Ctrl + Z 直前操作の取り消し(Undo) | Ctrl + F 検索 | Ctrl + K 表計算 |
| Ctrl + V 貼り付け | Ctrl + Y Undoの取り消し | Ctrl + H 置換 | Ctrl + N 並べ替え |
| Ctrl + X 切り取り | Ctrl + A 全て選択 | Ctrl + B ブロック入力 (選択範囲に一定値を代入) | |

## (1)統計処理結果の出力に関する設定




### ■計算結果出力の桁数 ①

▶ 標準

基本統計量

パラメトリック法
［第1頁：群1］
N：行数
n：有効データ数

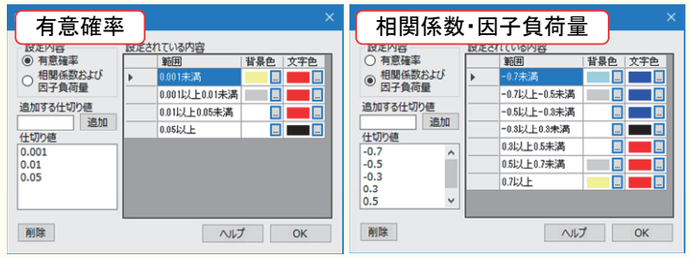
| 変数名 | n | Mean | SD | CV | SE | 最小値 | 最大値 | べき乗 | 変換原点 |
|---|---|---|---|---|---|---|---|---|---|
| 年齢 | 910 | 54.0 | 10.3 | 19.0 | 0.34 | 24.0 | 79.0 | 1.00 | 0.0 |
| 性別 | 910 | 0.4 | 0.5 | 117.6 | 0.02 | 0.0 | 1.0 | 1.00 | 0.0 |
| 飲酒度 | 910 | 1.0 | 1.3 | 132.4 | 0.04 | 0.0 | 5.0 | 1.00 | 0.0 |
| 喫煙度 | 910 | 1.4 | 1.9 | 133.2 | 0.06 | 0.0 | 7.0 | 1.00 | 0.0 |
| BMI | 910 | 23.44 | 2.92 | 12.5 | 0.097 | 14.20 | 38.10 | 1.00 | 0.00 |
| HbA1c | 910 | 5.24 | 0.57 | 10.9 | 0.019 | 4.10 | 10.10 | 1.00 | 0.00 |
| GPT | 910 | 25.8 | 14.4 | 56.0 | 0.48 | 7.0 | 172.0 | 0.00 | 0.0 |
| GGT | 910 | 35.1 | 35.6 | 101.3 | 1.18 | 7.0 | 282.0 | 0.00 | 0.0 |

▶ 標準+1桁

| 変数名 | n | Mean | SD | CV | SE | 最小値 | 最大値 | べき乗 | 変換原点 |
|---|---|---|---|---|---|---|---|---|---|
| 年齢 | 910 | 53.96 | 10.27 | 19.04 | 0.341 | 24.00 | 79.00 | 1.000 | 0.00 |
| 性別 | 910 | 0.42 | 0.49 | 117.63 | 0.016 | 0.00 | 1.00 | 1.000 | 0.00 |
| 飲酒度 | 910 | 0.96 | 1.27 | 132.44 | 0.042 | 0.00 | 5.00 | 1.000 | 0.00 |
| 喫煙度 | 910 | 1.41 | 1.88 | 133.22 | 0.062 | 0.00 | 7.00 | 1.000 | 0.00 |
| BMI | 910 | 23.439 | 2.922 | 12.47 | 0.0969 | 14.200 | 38.100 | 1.000 | 0.000 |
| HbA1c | 910 | 5.243 | 0.571 | 10.89 | 0.0189 | 4.100 | 10.100 | 1.000 | 0.000 |
| GPT | 910 | 25.77 | 14.43 | 56.00 | 0.478 | 7.00 | 172.00 | 0.000 | 0.00 |
| GGT | 910 | 35.12 | 35.57 | 101.27 | 1.179 | 7.00 | 282.00 | 0.000 | 0.00 |

### ■統計量の色分け表示設定 ②

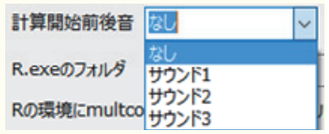
詳細設定 で下図のパネルが表示されます




上図のパネルで範囲とその色を設定し、システム設定パネルでチェックを付けて下さい

### ■計算開始・終了音の設定 ③

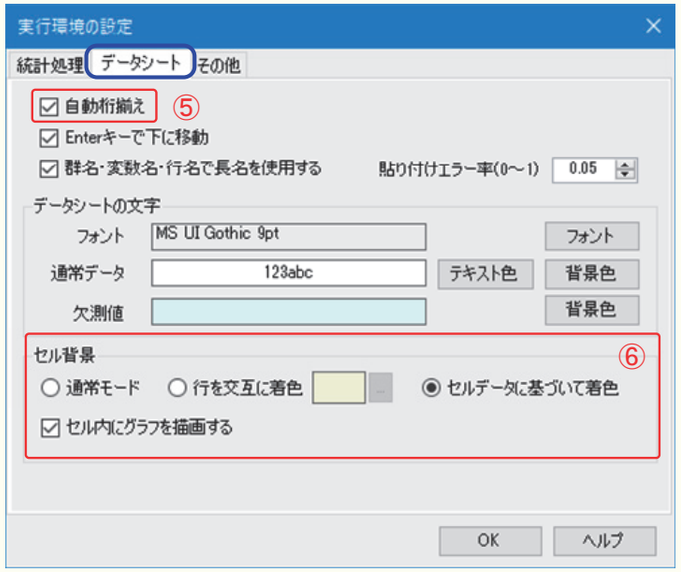
▶ 3種類のサウンドから選択できます




### ■Rの利用環境 ④

▶ 因子分析、一部検定でのP値の正確確率の計算で、Rを利用するため、Rのプログラムがインストールされている場所を表示します。多重比較に対しては、別途専用のパッケージ（multcomp）のインストールが必要となる場合があります。

## (2)データシートに関する設定




### ■自動桁揃え ⑤

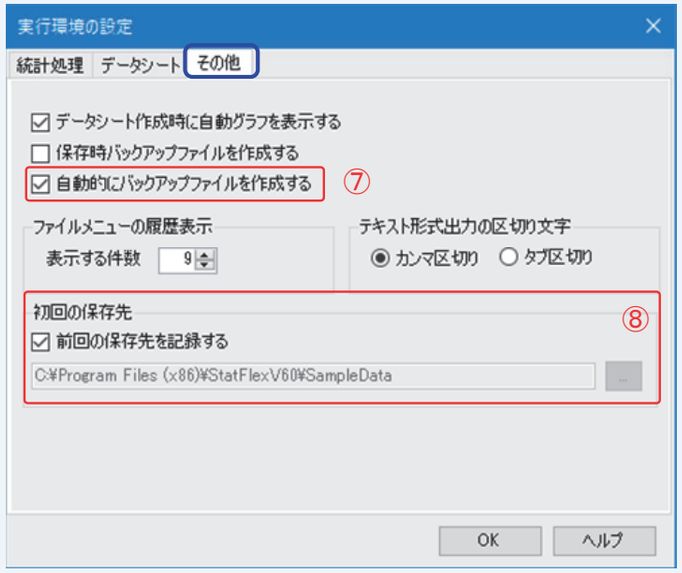
チェックを付けておくと、StatFlexは入力された数値の小数位を自動判定し、すそれに基づいて、データシート上で桁揃え表示しています。

自動桁揃えの小数位は、Var1 で確認できます。自動桁揃えがOFFの場合 Var1 で指定した小数位が、データシートの桁揃えに利用されます。

### ■セル背景 ⑥

セル内にグラフを描画

## (3)その他の設定（入出力ファイル関連）




### ■自動的にバックアップファイルを作成する ⑦

チェックしておくと、異常終了した場合でも、StatFlexの次回起動時にデータが自動復元されます

### ■初回の保存先 ⑧

［前回の保存先を記録する］にチェックしておくと、最後に使ったフォルダをに記録しておき、次回起動後、初回の保存や読み込み時に参照されます。チェックがない場合は、指定されたフォルダが参照されます。

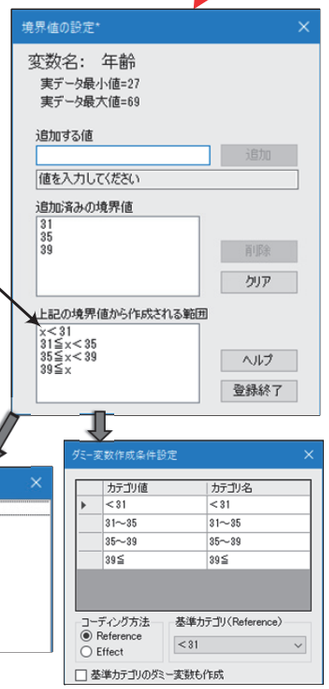
# 変数情報設定

## 変数基本情報設定（Var1）

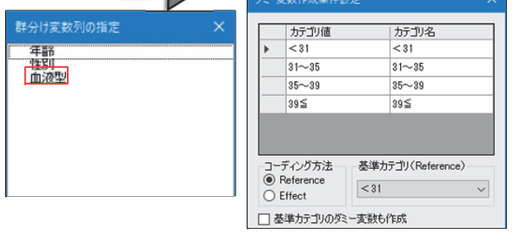
### 境界値の設定

数値型をカテゴリ型として扱えます。

境界値で区切られる各群は、**下限値を含み、上限値を含みません。**

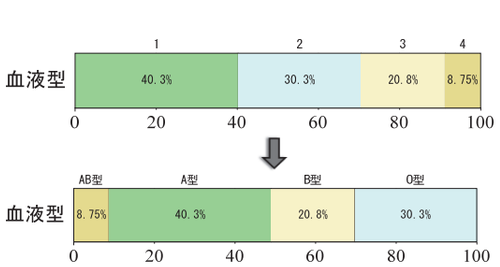



境界値の設定で、数値型でも群分けやダミー変数の作成が可能になります。




### カテゴリ情報編集

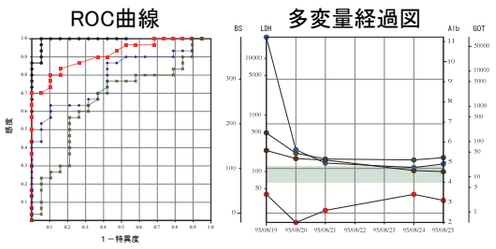
カテゴリ表示名や帯グラフの色分けを変更



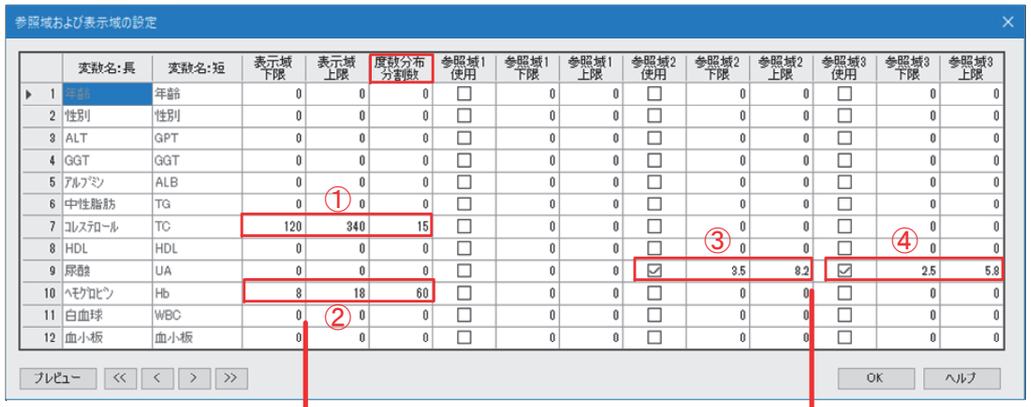

段階型の場合、数値ではないユーザ定義名を設定すると、ユーザ型になります。

## 変数別描画条件の設定（Var3）

変数毎に描画条件を指定します。自動設定で自動割り付けできます。
この指定はROC曲線と多変量経過図で有効です。

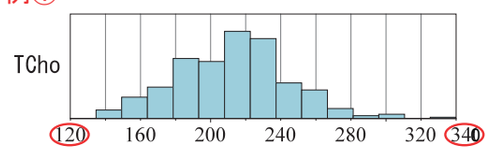



## 参照域および表示域の設定（Var2）

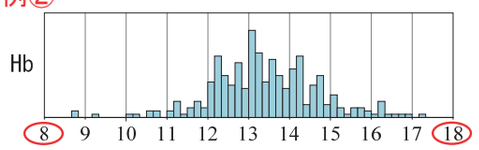



表示域の下限・上限と分割数を設定すれば、度数分布図の形状を変数毎に個別に設定できます。

例①

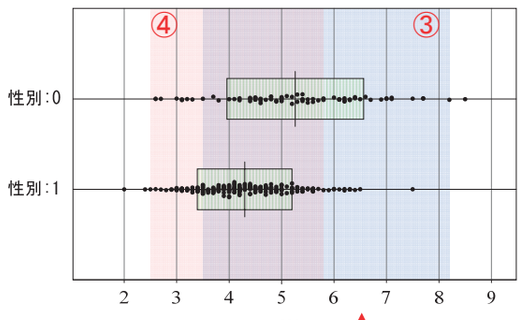



例②



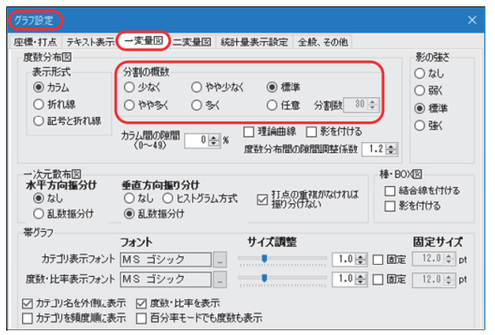

参照域は変数毎に3セット登録できます。下図は③と④で各々男性と女性の基準範囲を設定した例です。

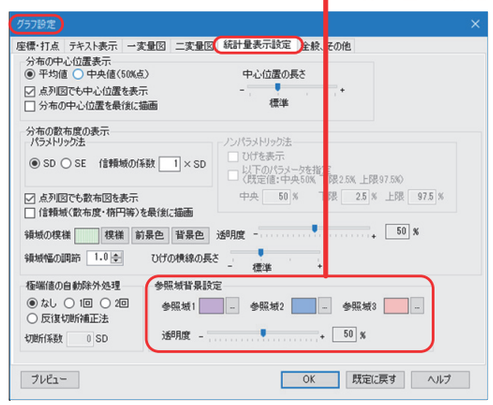
色と透明度の指定は の統計量の表示設定ページ（下図）で行います。



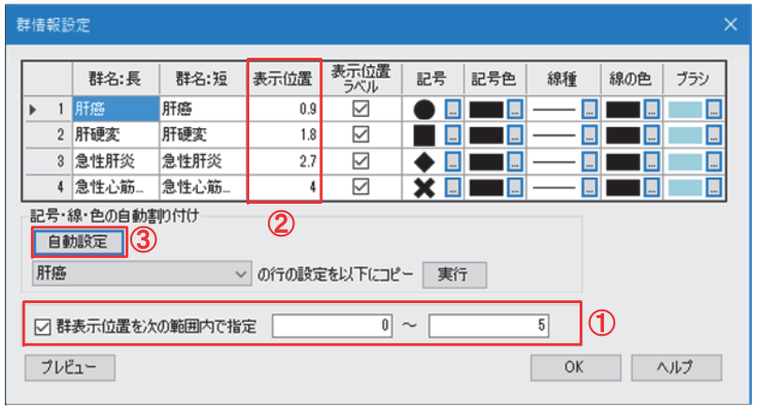

領域色の指定

⑤度数分布図分割数の指定がない場合、[設定]－[グラフ設定] の[一変量図]の度数分布図設定（下図）の指定が適用されます。



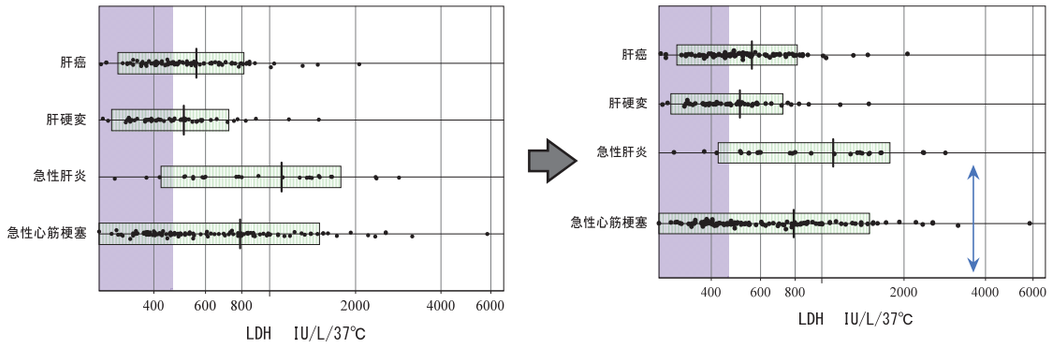



# 群別描画条件の設定 Grp



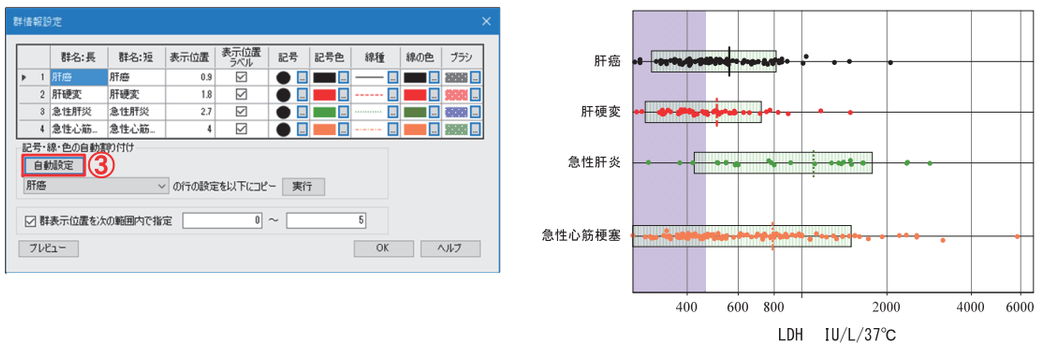

## ■群別に表示位置を調整するには

①で全体としての表示位置の範囲を指定
②で各群の表示位置を指定



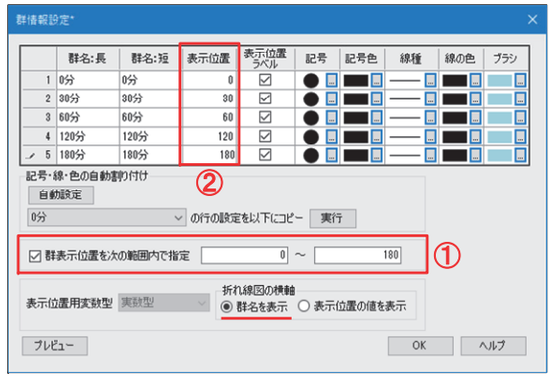

## ■群別に記号や色を変えるには

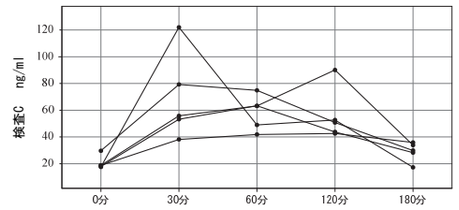
③ 自動設定 ボタンが便利です。任意の色に設定することもできます。




## ■折れ線グラフで条件(時間)軸の表示位置を調整するには

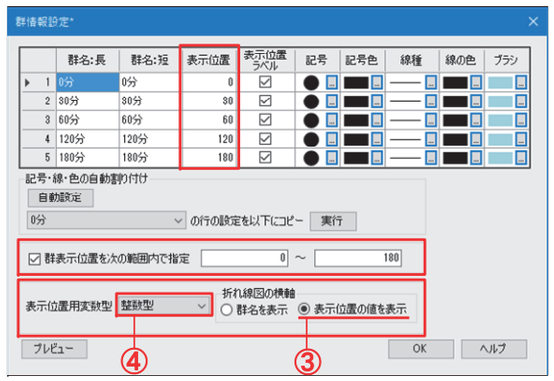
### ▶条件軸に群(列)名をそのまま表示する場合



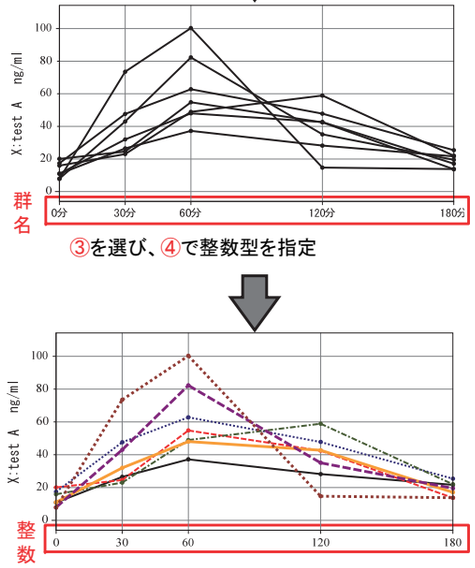




①で全体としての表示位置の範囲を指定
②で各群の表示位置を指定

### ▶条件軸に時間などの数値を表示する場合




時間をそのまま整数として表示した例です。



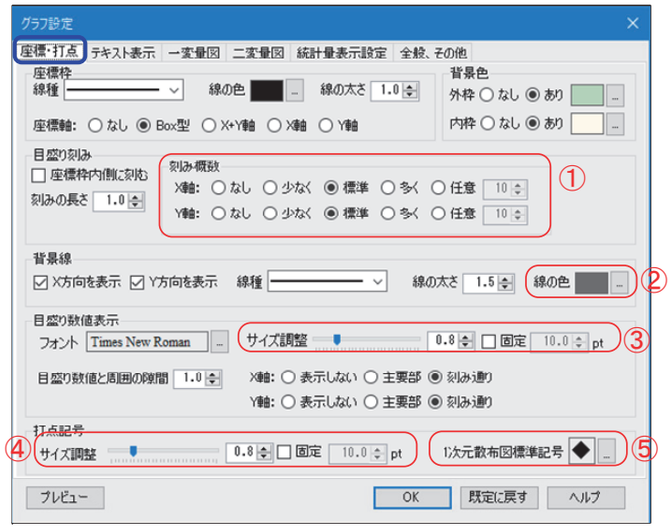

③を選び、④で整数型を指定




折れ線のデータ(個体)別の記号や線は Row で変更します

描画条件を設定するには、自動設定 ボタンが便利です

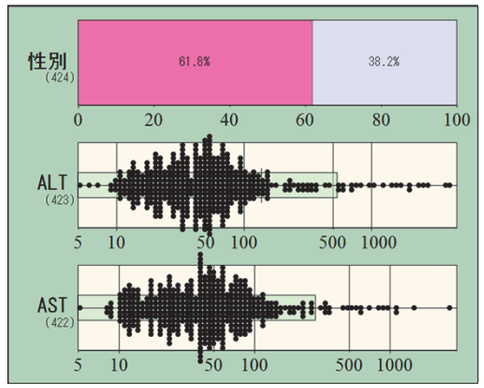
# グラフ設定（座標・打点）



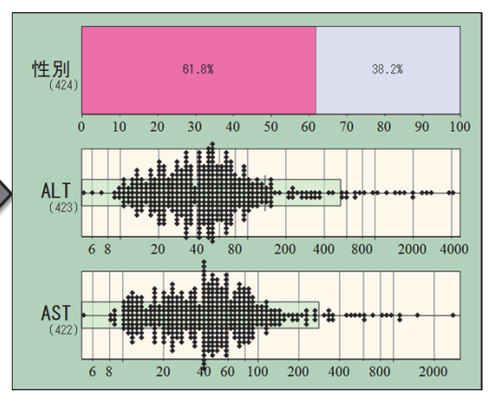

**設定例**　①～⑤を変更する前後のグラフを下に例示します。

| 変更した箇所 | 変更前 | 変更後 |
|---|---|---|
| ①目盛りの刻み概数 | 少なく | 標準 |
| ②背景線の色 | 黒 | 灰色 |
| ③目盛り数値のフォントサイズ | 1.0 | 0.8 |
| ④打点記号のサイズ | 1.0 | 0.8 |
| ⑤打点記号 | ● | ◆ |

変更前



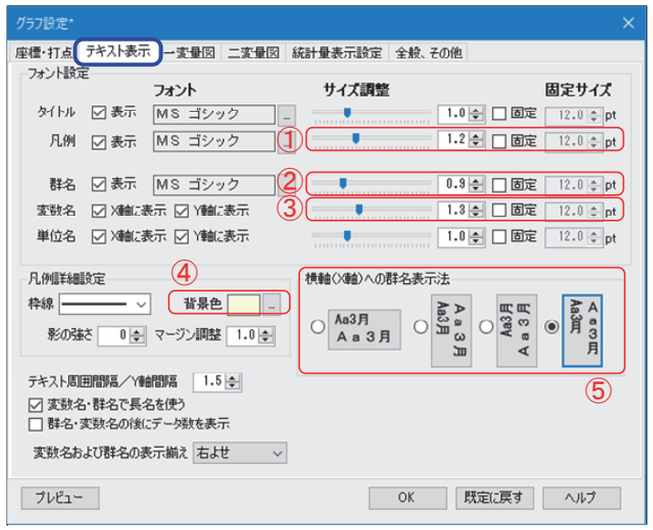

変更後







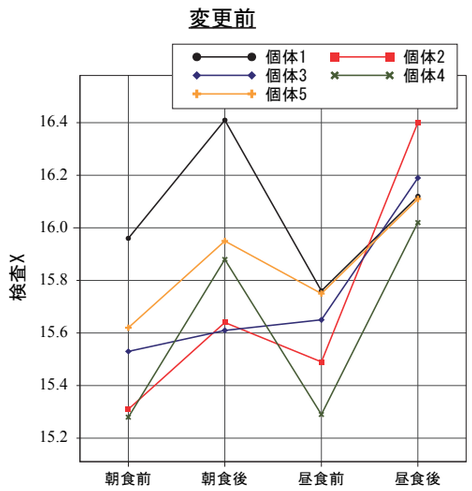

# グラフ設定（テキスト表示）



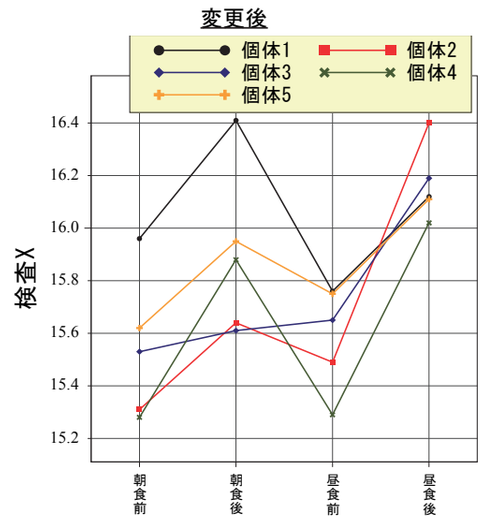

**設定例**　上図①～⑤を変更する前後のグラフを下に例示します。

| 変更した箇所 | 変更前 | 変更後 |
|---|---|---|
| ①凡例のフォントサイズ | 1.0 | 1.2 |
| ②群名のフォントサイズ | 1.0 | 0.9 |
| ③変数名のフォントサイズ | 1.0 | 1.3 |
| ④凡例の背景色 | 白 | 薄い黄色 |
| ⑤横軸への群名の表示法 | Aa3月 A a 3 月 | A a 3 月 Aa3月 |

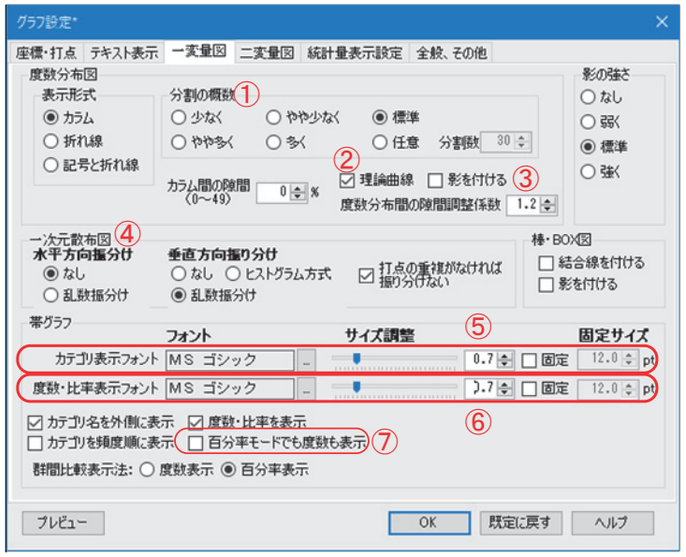
変更前




変更後



# グラフ設定（一変量図）

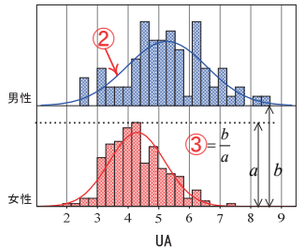



### ②理論曲線

チェックすると平均値と標準偏差に基づいて正規分布曲線を表示します

### ③度数分布間の隙間調整

複数の度数分布図を上下に配置する場合、その間隔を最大度数の何倍とするかを指定します（側方視点の場合に有効）




## ■ 度数分布図

### ①分割の概数

度数分布図の分割概数を指定します。

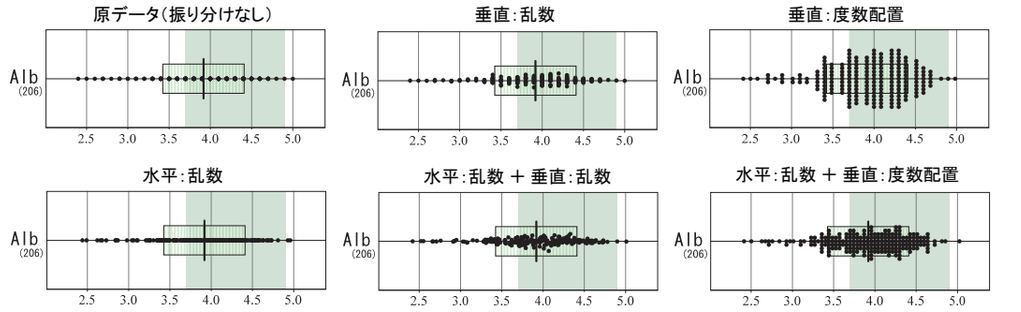
Var2で変数別指定があれば、それを優先

一般に、このVar2で変数毎に、表示域と分割数の指定をお勧めします。

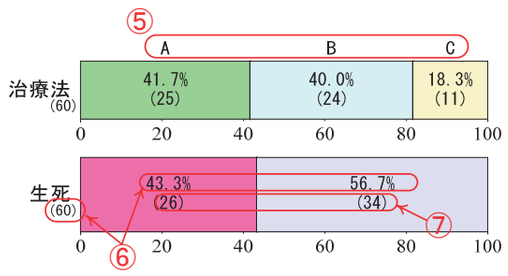
## ■ 一次元散布図

### ④打点の乱数振り分け機能

水平乱数では、小数位の1つ下の桁に乱数を加えてグラフ表示します。例えば、小数位が1で、0.1刻みで記録された数値のとき、−0.05～0.04の範囲の一様乱数を付加して、グラフ表示し、点の重なりを防ぎます。




## ■ 帯グラフ




### ⑤カテゴリ名フォント

カテゴリ名表示用のフォントとそのサイズを指定します

### ⑥度数・比率表示フォント

カテゴリの度数・比率、群名・変数名の後のデータ数表示用のフォントとそのサイズを指定します。
群名・変数名の後にデータ数を表示するには、「テキスト表示」の「群名・変数名の後にデータ数を表示」をチェックして下さい。

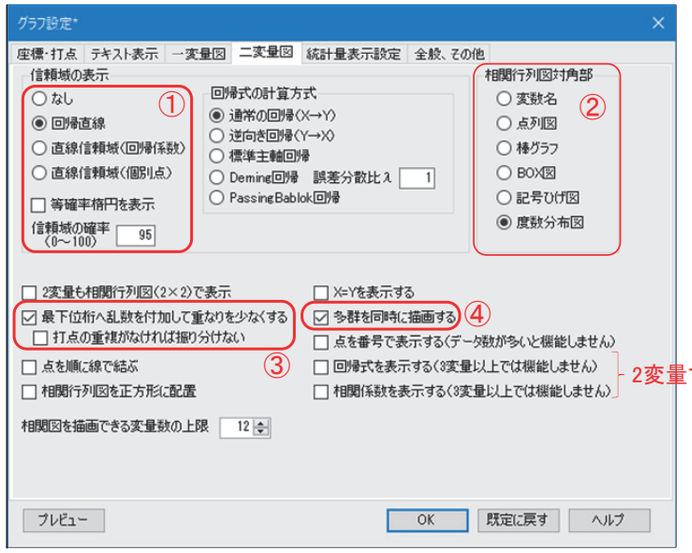
### ⑦百分率モードでも度数を表示

百分率と度数を併記します





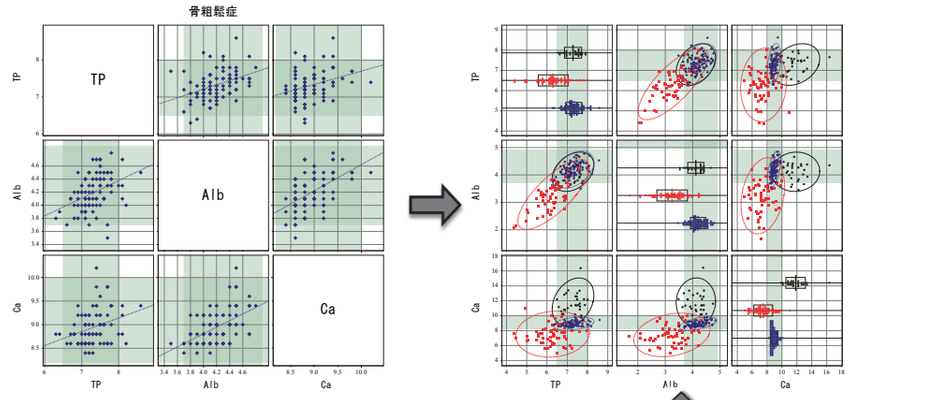
# グラフ設定（二変量図）



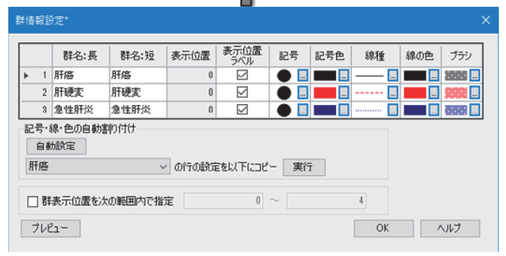

## 設定例

①～④を変更する前後のグラフを例示します

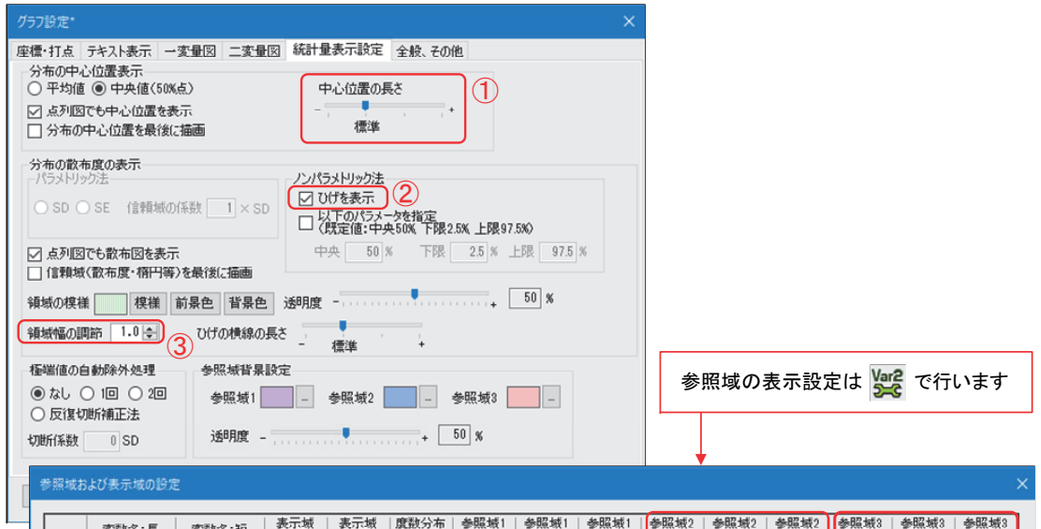
| 変更した箇所 | 変更前 | 変更後 |
|---|---|---|
| ①信頼域の表示方法 | 回帰直線 | 等確率楕円 |
| ②相関行列図対角部 | 変数名 | 点列図 |
| ③最下位桁へ乱数を付加して… | チェックなし | チェックあり |
| ④多群を同時に描画する | チェックなし | チェックあり |




群別の記号や色はGrpで設定します



# グラフ設定（統計量表示設定）

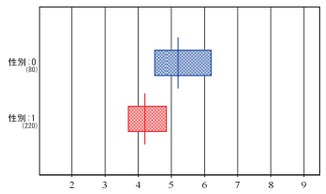



参照域の表示設定は Var2 で行います

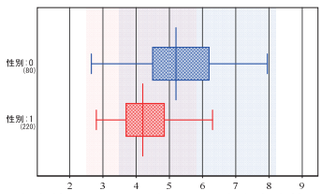
**設定例**　①～④を変更する前後のグラフを例示します

| 変更した箇所 | 変更前 | 変更後 |
|---|---|---|
| ①中心位置の長さ | 標準 | 標準＋1 |
| ②ひげを表示（※） | チェックなし | チェックあり |
| ③領域幅の調節 | 1.0 | 1.3 |
| ④参照域2使用、参照域3使用 | チェックなし | チェックあり |

※点列図の場合にも散布度を箱で表示できますが、ひげは表示しません



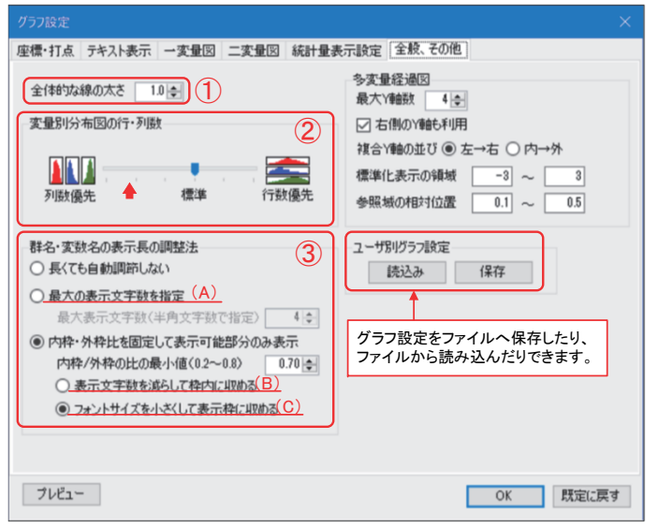

中央値、IQRを表す箱図




中央値、IQR、2.5%点、97.5%点を表す箱ひげ図

この[統計量表示設定]の一部は、
グラフ形式設定パネルからも行えます

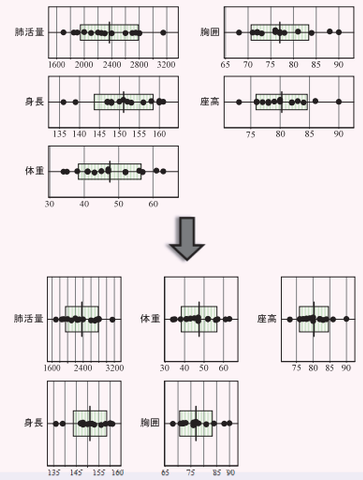
# グラフ設定（全般・その他）



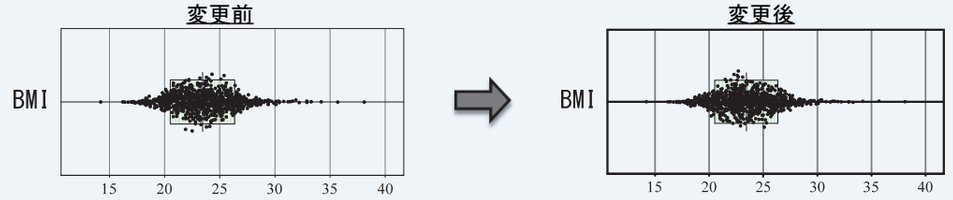

グラフ設定をファイルへ保存したり、
ファイルから読み込んだりできます。

**設定例2**

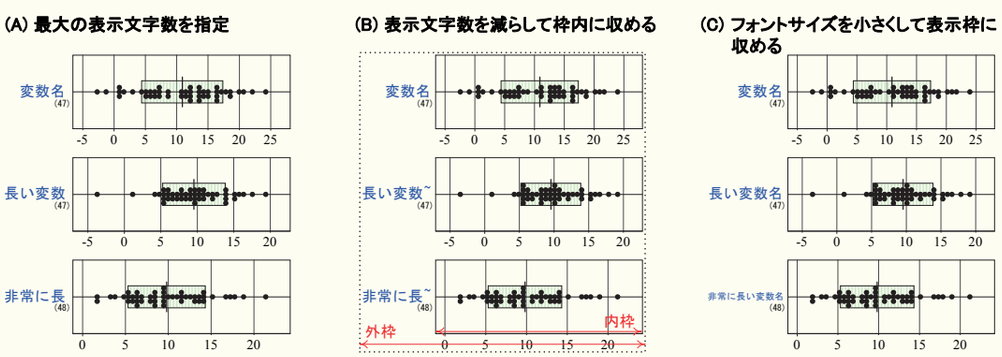
②で多変量図の配置を[標準]から
列数優先へ変更




**設定例1**　①で全体的な線の太さを1.0⇒2.0に変更

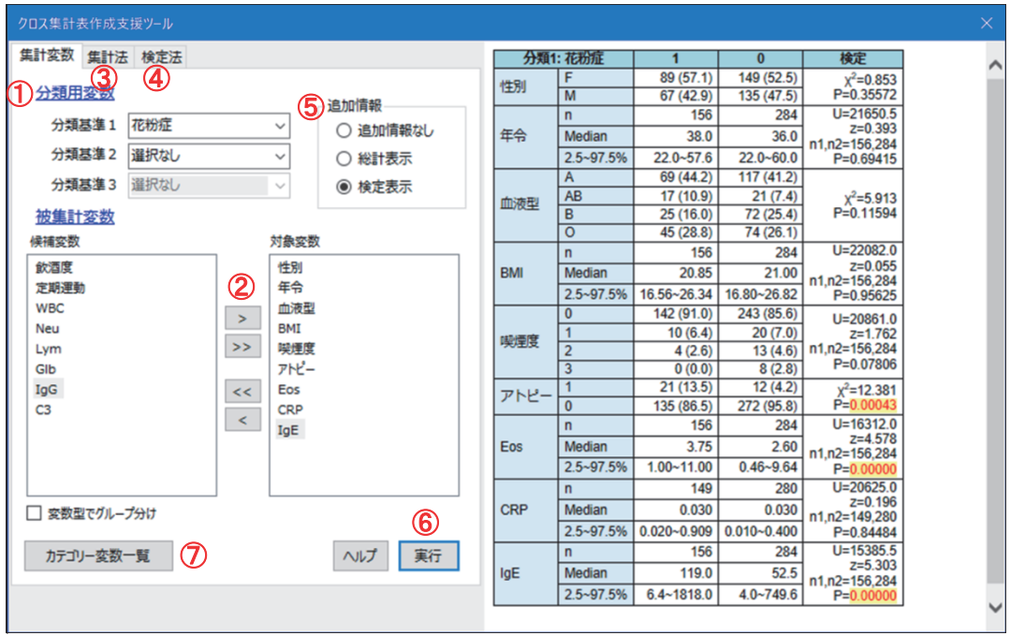



**設定例3**　③で変数名の表示長を(A)、(B)、(C)の3通りに設定



# 3-1 クロス集計ツール

分類用変数でデータを群分けし、他の変数の値を群毎に集計する機能をクロス集計と呼びます。被集計変数の型が異なっていても、自動判定により一気に集計でき、群間差の検定も同時に実行できます。




| 分類1: 花粉症 | | 1 | 0 | 検定 |
|---|---|---|---|---|
| 性別 | F | 89 (57.1) | 149 (52.5) | $\chi^2$=0.853 |
| | M | 67 (42.9) | 135 (47.5) | P=0.35572 |
| 年令 | n | 156 | 284 | U=21650.5 |
| | Median | 38.0 | 36.0 | z=0.393 |
| | 2.5~97.5% | 22.0~57.6 | 22.0~60.0 | n1,n2=156,284 P=0.69415 |
| 血液型 | A | 69 (44.2) | 117 (41.2) | |
| | AB | 17 (10.9) | 21 (7.4) | $\chi^2$=5.913 |
| | B | 25 (16.0) | 72 (25.4) | P=0.11594 |
| | O | 45 (28.8) | 74 (26.1) | |
| BMI | n | 156 | 284 | U=22082.0 |
| | Median | 20.85 | 21.00 | z=0.055 |
| | 2.5~97.5% | 16.56~26.34 | 16.80~26.82 | n1,n2=156,284 P=0.95625 |
| 喫煙度 | 0 | 142 (91.0) | 243 (85.6) | U=20861.0 |
| | 1 | 10 (6.4) | 20 (7.0) | z=1.762 |
| | 2 | 4 (2.6) | 13 (4.6) | n1,n2=156,284 |
| | 3 | 0 (0.0) | 8 (2.8) | P=0.07806 |
| アトピー | 1 | 21 (13.5) | 12 (4.2) | $\chi^2$=12.381 |
| | 0 | 135 (86.5) | 272 (95.8) | P=0.00043 |
| Eos | n | 156 | 284 | U=16312.0 |
| | Median | 3.75 | 2.60 | z=4.578 |
| | 2.5~97.5% | 1.00~11.00 | 0.46~9.64 | n1,n2=156,284 P=0.00000 |
| CRP | n | 149 | 280 | U=20625.0 |
| | Median | 0.030 | 0.030 | z=0.196 |
| | 2.5~97.5% | 0.020~0.909 | 0.010~0.400 | n1,n2=149,280 P=0.84484 |
| IgE | n | 156 | 284 | U=15385.5 |
| | Median | 119.0 | 52.5 | z=5.303 |
| | 2.5~97.5% | 6.4~1818.0 | 4.0~749.6 | n1,n2=156,284 P=0.00000 |

## 手順1 分類変数を指定①

分類基準を設定します。3つまで指定できます。各基準のカテゴリー数を$k_1$, $k_2$, $k_3$ とすると、出力表の列数は、$k_1 \times k_2 \times k_3$となり、総計や検定結果欄を含めると$(k_1+1)\times(k_2+1)\times(k_3+1)$なります。

## 手順2 被集計変数を指定②

クロス集計に組み込む対象変数を、候補から選択します。

● 変数型でグループ分け： 数値型やカテゴリー型順に並べ替えられて表示されます。

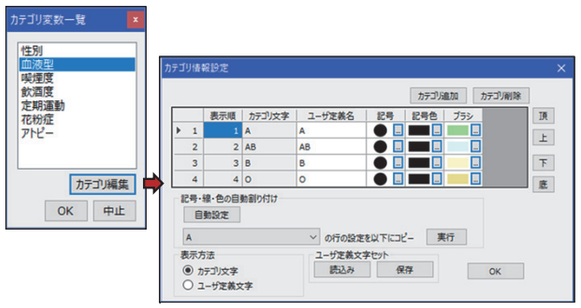
## 手順3 計算・出力方法を指定

集計用統計量を指定③ ：数値型とカテゴリー型に分けて指定（右頁参照）
群間差検定の方式を指定④ ：変数型に応じた検定法を指定（右頁参照）
追加情報（出力形式）を指定⑤ ：総計や群間差検定結果の結果を出力するか指定

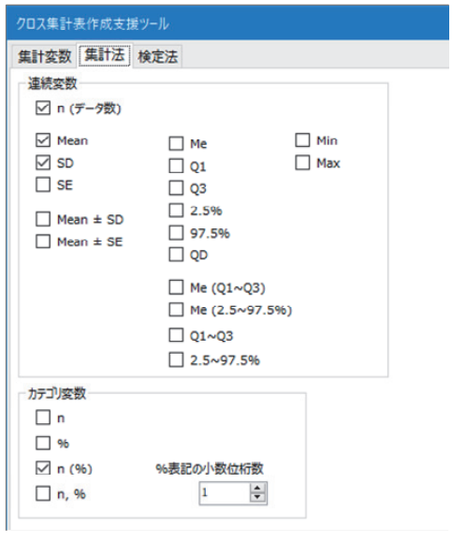
## 手順4 計算を実行する⑥

## 手順5 カテゴリー名や並びを編集⑦

クロス集計表に表示する、カテゴリー名やその並びを指定できます




● 集計法③




連続変数では、データ数や平均、SD以外に、中央値や四分位範囲、パーセンタイル値など任意の数の統計量を指定できます。有効桁数は、システム設定で調整します。
カテゴリ型では、度数(n)とその比率(%)の組合せを指定します。%表示に対してはその表示桁数を設定できます。

● 検定法④

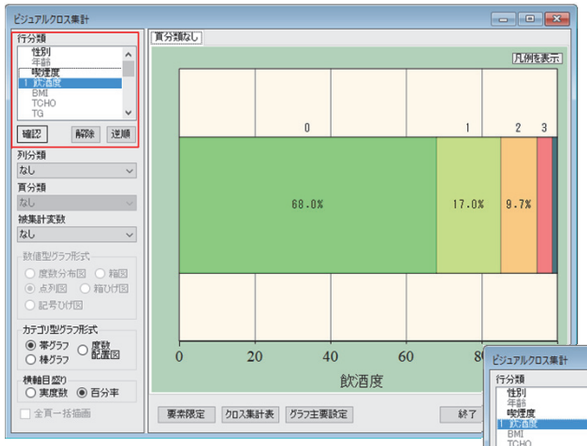
2値型やユーザー型では検定方法は固定です。
段階型と整数型・実数型では、パラメトリック検定法とノンパラメトリック検定法を選択できます。

● 2つの基準変数でクロス集計した例

年代（2層）、花粉症の有無で4通りに集計し、検定結果も同時出力、数値型はノンパラメトリック法で検定

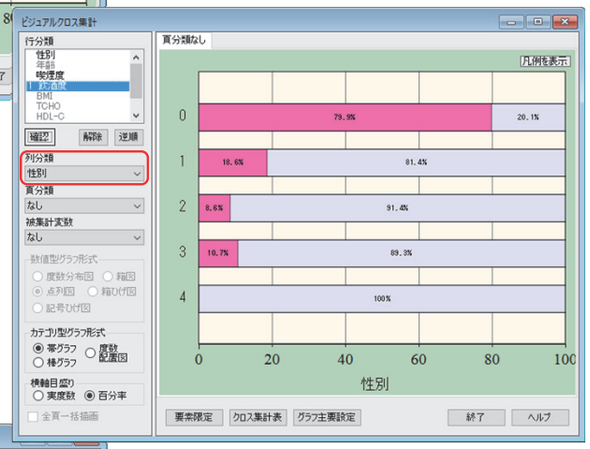
| 分類1: 年令2 | | <40 | | | 40≦ | | | 総計 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 分類2: 花粉症 | | 1 | 0 | 検定 | 1 | 0 | 検定 | 1 | 0 | 検定 |
| 性別 | F | 48 (57.1) | 87 (54.0) | $\chi^2$=0.215 | 41 (56.9) | 62 (50.4) | $\chi^2$=0.779 | 89 (57.1) | 149 (52.5) | $\chi^2$=0.853 |
| | M | 36 (42.9) | 74 (46.0) | P=0.64273 | 31 (43.1) | 61 (49.6) | P=0.37745 | 67 (42.9) | 135 (47.5) | P=0.35572 |
| 血液型 | A | 34 (40.5) | 74 (46.0) | | 35 (48.6) | 43 (35.0) | | 69 (44.2) | 117 (41.2) | |
| | AB | 8 (9.5) | 8 (5.0) | $\chi^2$=5.957 | 9 (12.5) | 13 (10.6) | $\chi^2$=4.819 | 17 (10.9) | 21 (7.4) | $\chi^2$=5.913 |
| | B | 13 (15.5) | 39 (24.2) | P=0.11373 | 12 (16.7) | 33 (26.8) | P=0.18554 | 25 (16.0) | 72 (25.4) | P=0.11594 |
| | O | 29 (34.5) | 40 (24.8) | | 16 (22.2) | 34 (27.6) | | 45 (28.8) | 74 (26.1) | |
| BMI | n | 84 | 161 | U=6462.0 | 72 | 123 | U=4195.0 | 156 | 284 | U=22082.0 |
| | Median | 20.25 | 20.00 | z=0.570 | 21.60 | 21.90 | z=0.613 | 20.85 | 21.00 | z=0.055 |
| | 2.5~97.5% | 16.64~25.80 | 16.51~26.77 | n1,n2=84,161 P=0.56877 | 16.52~26.92 | 17.60~26.83 | n1,n2=72,123 P=0.54005 | 16.56~26.34 | 16.80~26.82 | n1,n2=156,284 P=0.95625 |
| 喫煙度 | 0 | 72 (85.7) | 134 (83.2) | U=6588.0 | 70 (97.2) | 109 (88.6) | U=4036.0 | 142 (91.0) | 243 (85.6) | U=20861.0 |
| | 1 | 8 (9.5) | 17 (10.6) | z=0.520 | 2 (2.8) | 3 (2.4) | z=2.166 | 10 (6.4) | 20 (7.0) | z=1.762 |
| | 2 | 4 (4.8) | 10 (6.2) | n1,n2=84,161 | 0 (0.0) | 3 (2.4) | n1,n2=72,123 | 4 (2.6) | 13 (4.6) | n1,n2=156,284 |
| | 3 | 0 (0.0) | 0 (0.0) | P=0.60327 | 0 (0.0) | 8 (6.5) | P=0.03031 | 0 (0.0) | 8 (2.8) | P=0.07806 |
| アトピー | 1 | 12 (14.3) | 10 (6.2) | $\chi^2$=4.403 | 9 (12.5) | 2 (1.6) | Fisher両側 | 21 (13.5) | 12 (4.2) | $\chi^2$=12.381 |
| | 0 | 72 (85.7) | 151 (93.8) | P=0.03587 | 63 (87.5) | 121 (98.4) | P=0.00243 | 135 (86.5) | 272 (95.8) | P=0.00043 |
| Eos | n | 84 | 161 | U=4884.0 | 72 | 123 | U=3331.0 | 156 | 284 | U=16312.0 |
| | Median | 3.65 | 2.60 | z=3.567 | 3.85 | 2.70 | z=2.885 | 3.75 | 2.60 | z=4.578 |
| | 2.5~97.5% | 1.00~10.52 | 0.45~9.65 | n1,n2=84,161 P=0.00036 | 1.03~11.48 | 0.46~9.04 | n1,n2=72,123 P=0.00391 | 1.00~11.00 | 0.46~9.64 | n1,n2=156,284 P=0.00000 |

# 3-2 ビジュアルクロス集計



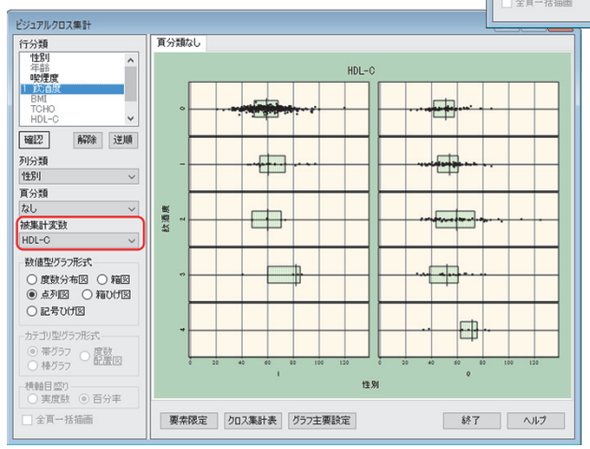

## 手順1 行分類変数を選択

行分類変数のカテゴリの構成を帯グラフ、棒グラフ、度数表示形式図のいずれかで図示します。




## 手順2 列分類変数を選択（省略可）

行分類変数のカテゴリを縦軸に配置し、列分類変数のカテゴリの構成を帯グラフ、棒グラフ、度数表示形式のいずれかで図示します。



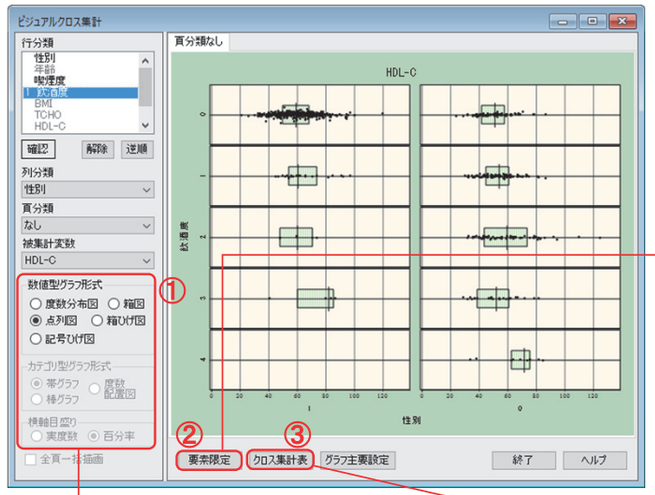

## 手順3 被集計変数を選択（省略可）

分類変数で作成される区画（左の例では5行×2列の10区画）ごとに、被集計変数に関するグラフを作成します。
被集計変数が数値型の時は、点列図、度数分布図、記号ひげ図、箱図、箱ひげ図のいずれかで分布を図示します。

被集計変数がカテゴリ型の場合は、被集計変数のカテゴリの構成を帯グラフ、棒グラフ、度数表示形式ずのいずれかで図示します。

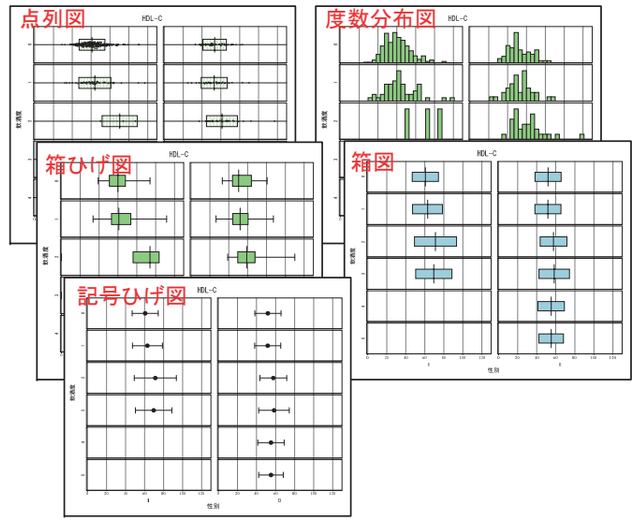
## 手順4 頁分類変数を選択（省略可）

さらに、行や列と同じ感覚でページ方向に分割できます。

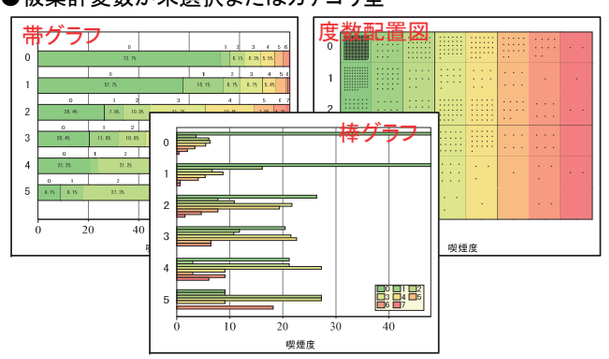



## ①グラフ形式を変更

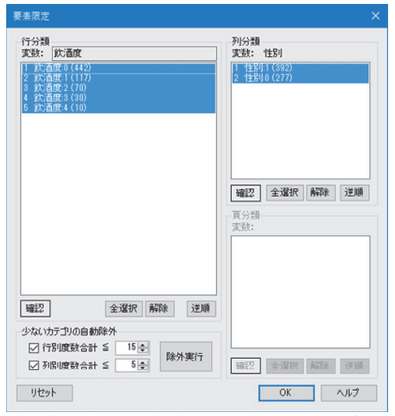
●被集計変数が数値型




●被集計変数が未選択またはカテゴリ型



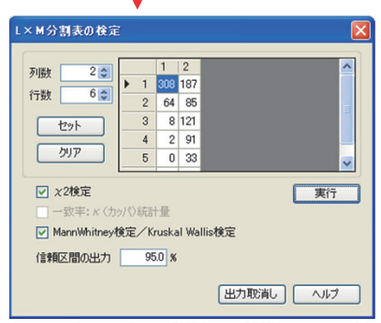

## ②要素限定



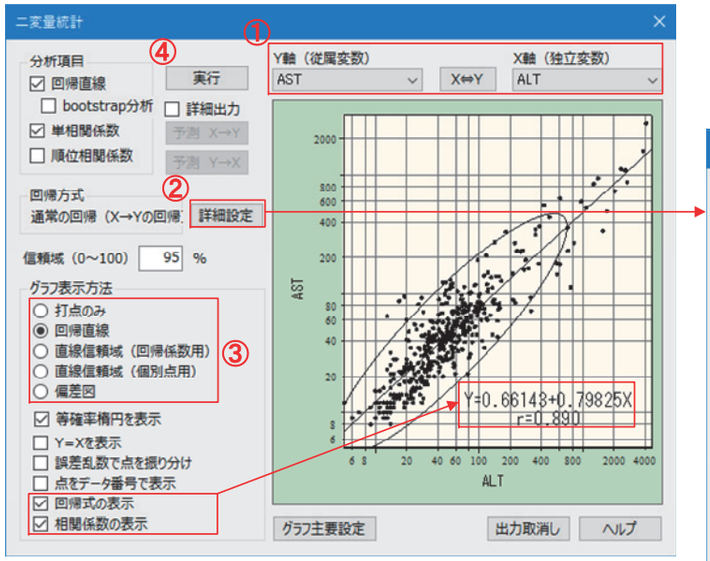

行、列、頁分類変数の表示するカテゴリとその順序を指定できます。

## ③クロス集計表を作成

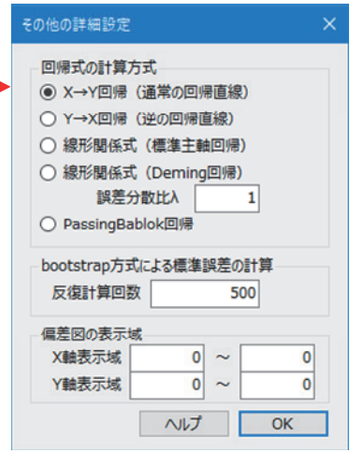
指定された分類変数、被集計変数に従い、クロス集計表を作成できます。




クロス集計表の[検定]ボタンで計数値の検定ができます。（度数の場合のみ）




回帰方式の指定

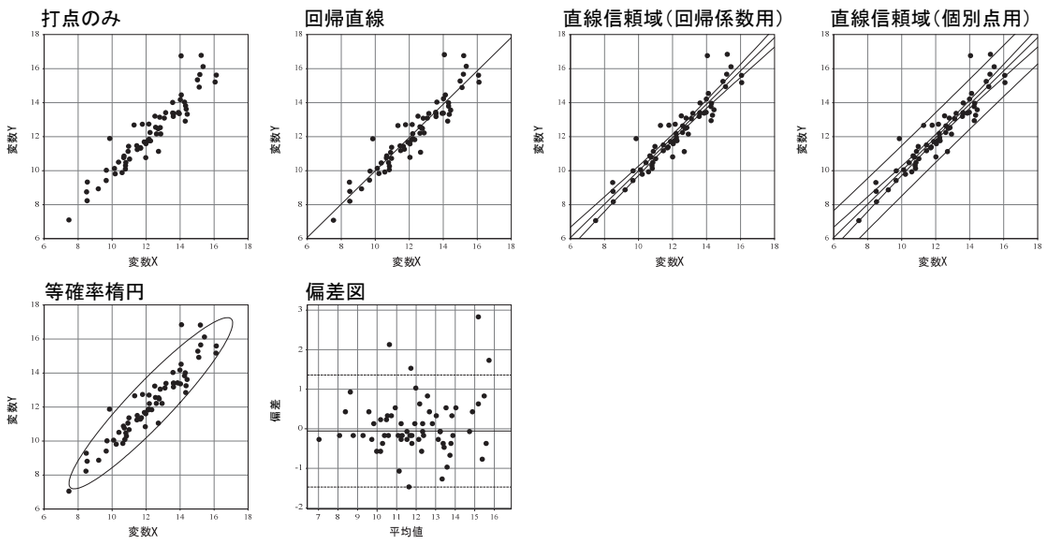



**手順1** ①変数を決める

x軸とy軸の変数を決めます。中央の X⇔Y で入れ換えできます。

**手順2** ②で回帰方式を指定

**手順3** ③グラフの表示方法を指定（線形関係式では直線信頼域は表示されません）




**手順4** ④で分析出力

チェックした項目について分析結果を出力します。[詳細出力]をチェックすると回帰係数の信頼区間を出力します（線形関係式の場合は、bootstrapもチェック）。







変量別（群別）に、データの分布型（正規性）の評価に利用します。
正規性の検定とべき乗変換による分布の正規化に対応しています。

実在データの最小値
（べき乗変換の変換原点は、最小値より小さく指定）

▶ 極端値マーキング

極端値検出法(3法)を選んで、[極端値マーク]ボタンを押すと、べき乗変換指定に基づき極端値を検出し、ヒストグラム（散布図）やデータ表にマークします。
[マーク値を除外]ボタンでそのデータを削除できます。

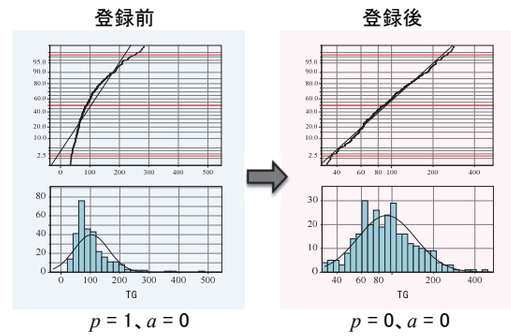
【注】本機能は、分布の正規性を前提としています。必ず分布の中央部が正規分布と見なせることを確認して利用して下さい。

**手順1** ①で対象変数を選択 →
正規分布
**手順2** 分布の正規性の判定
正規分布でない
**手順3** べき乗値⑥、変換原点⑦を指定（下記枠内参照）
**手順4** ⑧で変換を登録
**手順5** ⑨を押すと、べき乗変換の登録がデータシートに反映されます。
終了

変換前②、変換後③の分布図、正規確率紙④、正規検定⑤を利用します。

▶ 分布の正規性の簡易判定法

1）|歪度|<0.5 かつ |尖度|<0.5
2）正規確率紙の縦軸(累積度数率)の10～90の範囲に直線性を認める

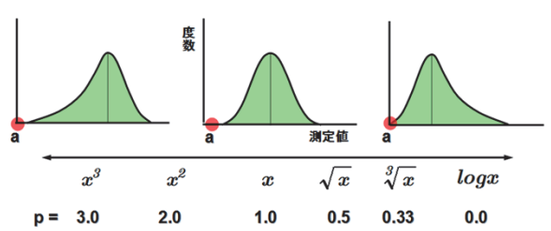



▶ Box-Cox方式によるべき乗変換式

$$X = \begin{cases} \dfrac{(x-a)^p - 1}{p} & (p \neq 0 \text{ の場合}) \\ \log(x-a) & (p = 0 \text{ の場合}) \end{cases}$$

xは変換前の値、pはべき乗値、aは変換原点、Xは変換後の値です。

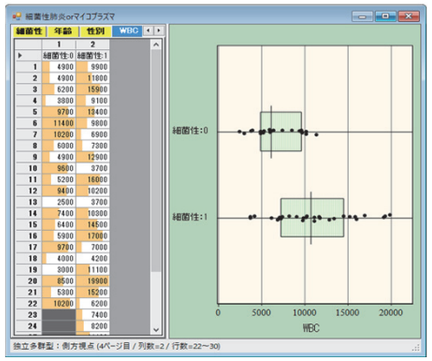
べき乗値pには、変換がない場合1.0、分布が高値側へ広がる場合は1.0よりも小さな値を、逆に低値側に広がる場合は1.0よりも大きな値を指定します。変換原点aは、最小値よりも小さな値を設定します。



# 3-5 判断分析

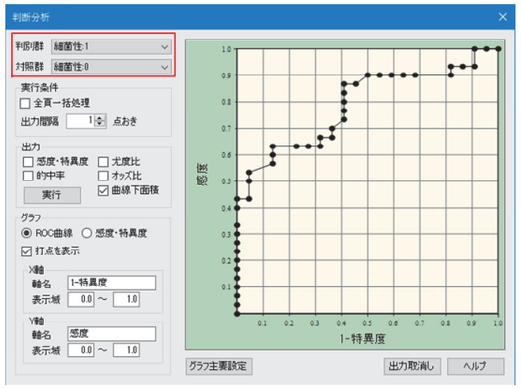
当該変数の値で、任意の2群をどの程度の精度で分別できるかを分析する機能を提供します。
カットオフ値を段階的に変化させた場合に、2群の判別度（感度、特異度、的中率、オッズ比、尤度比）の変化の一覧およびROC曲線を表示します。

## 手順1 データ形式の確認



独立多群型の側方視点で実行できます。
データベース型から、群分けした場合は、側方視点から、そのまま実行できます。

## 手順2 判別群・対照群の指定



対象となる2群を指定します。

## 手順3 実行条件の指定

### 全頁一括処理

グラフ表示・出力において、全変数を対象にするかを設定します。チェックすると、ROC曲線を描ける変数が複数ある場合に、それらを同時描画します。

各変数の打点記号や線の色や種類を変更するには、グラフ上で右クリックし、「**変数別描画条件の設定**」を実行してください。

### 出力間隔

カットオフ値を移動する間隔を指定します。
1の場合、重複のないカットオフ値を全て利用します。2以上にすると、その割合でデータを間引きます。例えば3とした場合、カットオフ値を3つおきに利用します。但し、出力表は小さくなりますが、曲線下面積の計算精度が低下します。

## 手順4 出力

出力項目にチェックを入れ、[**計算出力]ボタン**を押すと、右頁の「結果出力」が統計情報に記録されます。

### 分析結果

各カットオフ値に対し、以下の指標を出力します。用語説明は、判別群=疾患群、対照群=非疾患群としています。

**感度・特異度**：感度は疾患群の陽性率、特異度は非疾患群の陰性率
**+的中率**　：陽性者の中の疾患を有する人の割合
**−的中率**　：陽性者の中の疾患を有する人の割合
**尤度比**　：疾患群の陽性率/非疾患群の偽陽性率
**オッズ比**　：疾患群の陽性オッズ/非疾患群が偽陰性オッズ
**a ～ d**　：**右の**2×2分割表のa～dに相当

$$感度=\frac{a}{a+c} \quad 特異度=\frac{d}{b+d} \quad 陽性適中率=\frac{a}{a+b}$$

$$オッズ比=\frac{a/c}{b/d} \quad 尤度比=\frac{a/(a+c)}{b/(b+d)} \quad 陰性適中率=\frac{d}{c+d}$$

| | 疾患群 | 非疾患群 |
|---|---|---|
| 陽性 | a | b |
| 陰性 | c | d |

### 曲線化面積（AUC）

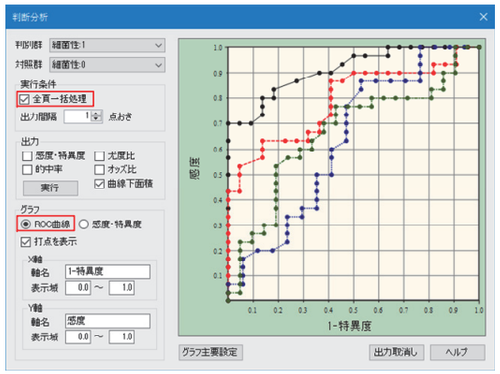
ROC曲線の右下側の面積の値で、当該変数により、2群の判別がどの程度うまく行えるかを示します。0.5～1.0の値をとり、1.0に近いほど判別度が高いことになります。
②の実行条件にて全頁一括処理を選択し、かつ変数が複数ある場合に、全ての変数に対してAUCを計算し、さらに全変数間でAUCの差の検定を行います。

## ROC曲線 / 感度・特異度曲線の切り替え

### ROC曲線

縦軸を感度、横軸を1−特異度とするROC曲線を描画します。複数の変数を同時に描画できます。

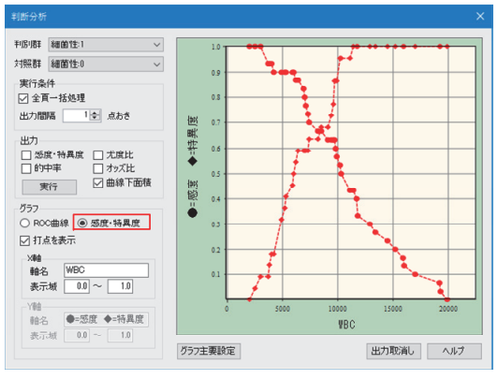
**打点を表示：**
カットオフ値を打点で表すかどうかを指定します。



ROC曲線（多変量一括処理）

### 感度・特異度

縦軸を比率、横軸をカットオフ値とする、感度と特異度のグラフを表示します。複数の変数に対する重ね合わせ表示はできません。



感度・特異度曲線

## 結果出力

手順4でチェックした項目に従い、以下のように出力されます。
（[全頁一括処理]をチェックした場合の出力例の一部を、例示します）

| No. | cutoff | 感度 | 1-特異度 | +的中率 | -的中率 | 尤度比 | オッズ比 | a | b | c | d |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 71.0 | 0.06667 | 0.04762 | 0.66667 | 0.41667 | 1.40000 | 1.42857 | 2 | 1 | 28 | 20 |
| 2 | 80.0 | 0.23333 | 0.04762 | 0.87500 | 0.46512 | 4.90000 | 6.08696 | 7 | 1 | 23 | 20 |
| 3 | 90.0 | 0.30000 | 0.14286 | 0.75000 | 0.46154 | 2.10000 | 2.57143 | 9 | 3 | 21 | 18 |
| 4 | 95.0 | 0.43333 | 0.19048 | 0.76471 | 0.50000 | 2.27500 | 3.25000 | 13 | 4 | 17 | 17 |
| 5 | 99.0 | 0.56667 | 0.23810 | 0.77273 | 0.55172 | 2.38000 | 4.18462 | 17 | 5 | 13 | 16 |
| 6 | 107.0 | 0.63333 | 0.33333 | 0.73077 | 0.56000 | 1.90000 | 3.45455 | 19 | 7 | 11 | 14 |
| 7 | 118.0 | 0.73333 | 0.42857 | 0.70968 | 0.60000 | 1.71111 | 3.66667 | 22 | 9 | 8 | 12 |
| 8 | 133.0 | 0.76667 | 0.57143 | 0.65714 | 0.56250 | 1.34167 | 2.46429 | 23 | 12 | 7 | 9 |
| 9 | 141.0 | 0.80000 | 0.71429 | 0.61538 | 0.50000 | 1.12000 | 1.60000 | 24 | 15 | 6 | 6 |
| 10 | 153.0 | 0.86667 | 0.85714 | 0.59091 | 0.42857 | 1.01111 | 1.08333 | 26 | 18 | 4 | 3 |
| 11 | 205.0 | 0.96667 | 0.90476 | 0.60417 | 0.66667 | 1.06842 | 3.05263 | 29 | 19 | 1 | 2 |
| 12 | 254.0 | 1.00000 | 0.95238 | 0.60000 | 1.00000 | 1.05000 | | 30 | 20 | 0 | 1 |

**< 感度=特異度となるカットオフ値 >**
カットオフ値 = 108.9
感度(=特異度)= 0.65041

[全頁一括処理]をチェックした場合、曲線下面積および差の検定結果を表示します。

| 頁 | 変数名 | 被判別群n | 対照群n | 曲線下面積 | 標準誤差 |
|---|---|---|---|---|---|
| 1 | 年齢 | 30 | 21 | 0.91270 | 0.03948 |
| 2 | WBC | 30 | 21 | 0.78492 | 0.06347 |
| 3 | LDH | 30 | 21 | 0.66508 | 0.07827 |

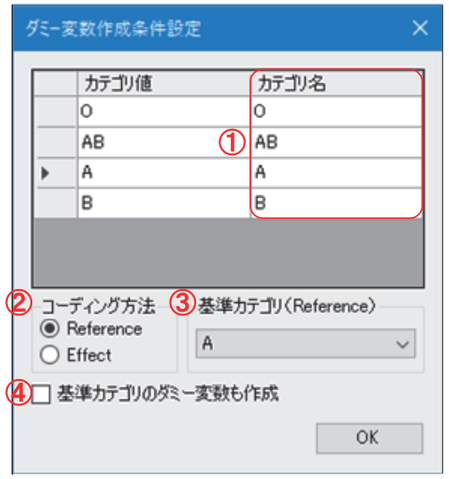
| 変数名 | 変数名 | 平均相関 | 平均面積 | z | 両側確率 |
|---|---|---|---|---|---|
| 年齢 | WBC | 0.04179 | 0.84881 | 1.57210 | 0.11593 |
| 年齢 | LDH | 0.04564 | 0.79048 | 1.92751 | 0.05392 |
| WBC | LDH | 0.03466 | 0.72659 | 0.81837 | 0.41315 |

# 3-6 ダミー変数の作成

## 手順1 ダミー変数作成パネル表示

- カテゴリ型（2値型、段階型、ユーザ型）、境界値を登録した数値型の列ヘッダを右クリック
- 表示されたメニューから「ダミー変数」を選択

## 手順2 ダミー変数の設定




**① カテゴリ名（ダミー変数名）**

作成するダミー変数の名前を設定します。
標準ではカテゴリ名がそのまま使用されます。

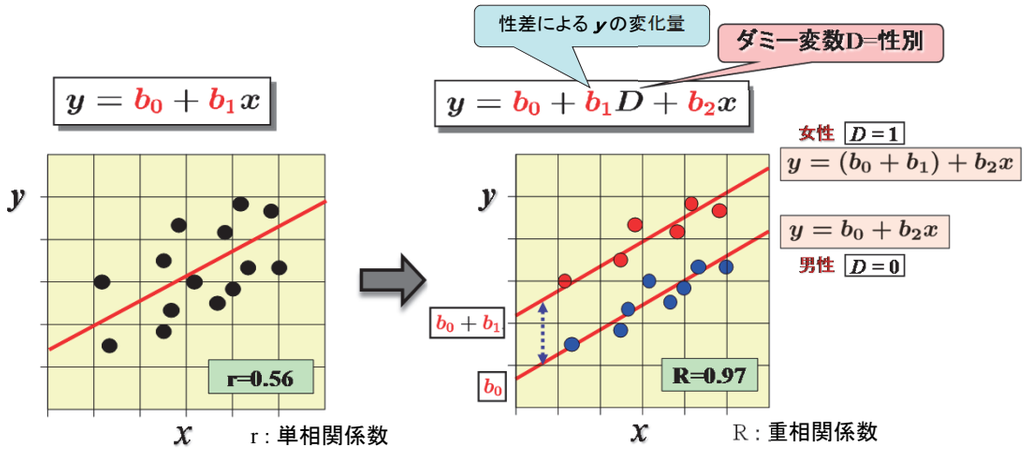
**② コーディング方法の指定**

コーディング方法（別枠）を指定します

**③基準カテゴリの指定**

基準カテゴリを選択します。
基準カテゴリに対してもダミー変数を作成する場合には④をチェック。
（ただし、reference coding場合のみ有効）

## ダミー変数（$D$）のグラフ上の意味



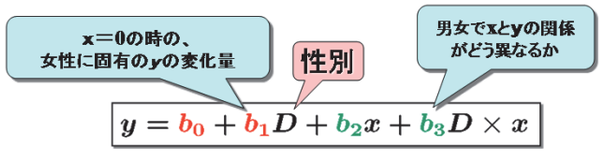

r：単相関係数

xだけではyを十分説明できていません

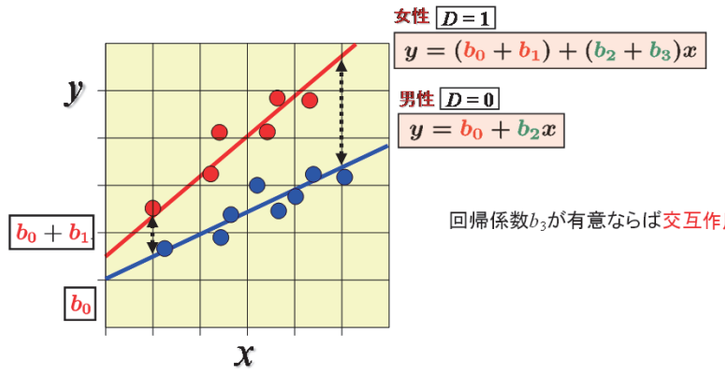
R：重相関係数

Dを加えることで、yをより的確に説明できるようになっています

## ダミー変数（$D$）のグラフ上の意味：交互作用有りの場合




交互作用とは、この例ではyとxの関係（傾き）が、変数Dの値によって変化することを指します




回帰係数$b_3$が有意ならば交互作用があることになります

## 2種類のコーディング方法

基準カテゴリ

### Reference codingの場合

▶2カテゴリの場合（1変数作成）

| 元の変数 性別 | ダミー変数 女性 |
|---|---|
| 男性 | 0 |
| 女性 | 1 |

▶kカテゴリの場合（$k$=4の場合 → 3変数作成）

| 元の変数 血液型 | 作成する変数 B型 | O型 | AB型 |
|---|---|---|---|
| A型 | 0 | 0 | 0 |
| B型 | 1 | 0 | 0 |
| O型 | 0 | 1 | 0 |
| AB型 | 0 | 0 | 1 |

### Effect codingの場合

▶2カテゴリの場合

| 元の変数 性別 | ダミー変数 女性 |
|---|---|
| 男性 | −1 |
| 女性 | 1 |

▶kカテゴリの場合

| 元の変数 血液型 | 作成する変数 B型 | O型 | AB型 |
|---|---|---|---|
| A型 | −1 | −1 | −1 |
| B型 | 1 | 0 | 0 |
| O型 | 0 | 1 | 0 |
| AB型 | 0 | 0 | 1 |

カテゴリがk種類ある場合は、k-1個のダミー変数が作成されます。
（だたし、Reference codingを選択し④にチェックした場合はk個作成されます）

# 3-7 重回帰分析（1）

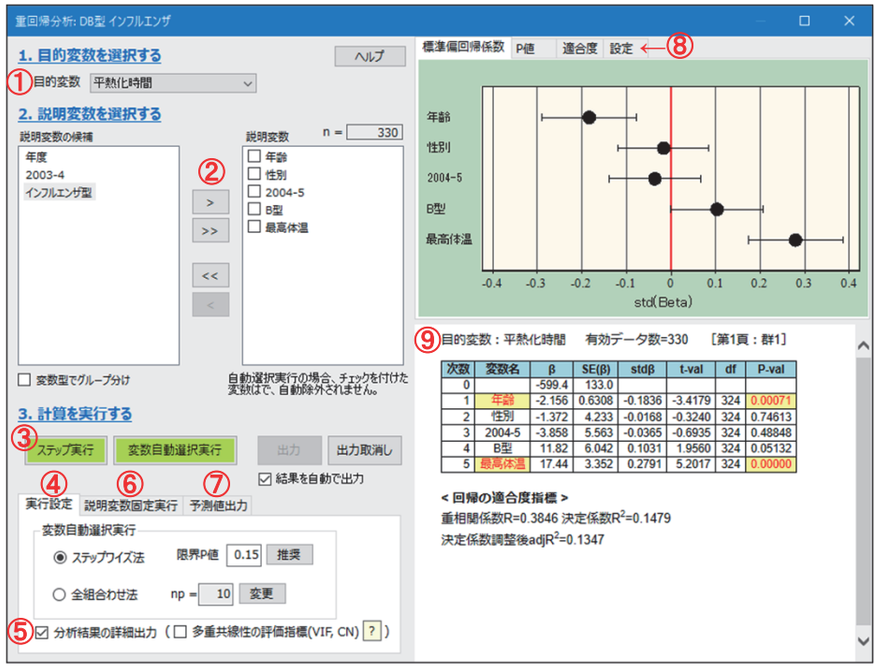
重回帰分析は、一群の説明変数($x_1$ ～$x_k$)と目的変数($y$) の関係を分析する場合に利用します。

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

例えば、HDL-C(善玉コレステロール) を目的変数とし、説明変数として設定した年齢、性別、喫煙度、飲酒度、肥満度(BMI) のどれが直接的に関係するかを分析する場合に利用します。このとき、HDL-C と各説明変数の関係は次式で表されます。

$$\text{HDL-C} = \beta_0 + \beta_1(\text{年齢}) + \beta_2(\text{性別}) + \beta_3(\text{飲酒度}) + \beta_4(\text{喫煙度}) + \beta_5(\text{BMI})$$

ここで、係数$\beta_1$ ～ $\beta_5$ は偏回帰係数と呼ばれ、個々の説明変数（年齢～BMI）と目的変数(HDL-C) の関係の強さを、他の説明変数の影響がない形で表しています。例えば、$\beta_1$ は年齢とHDL-C の関係を、性別・喫煙度・飲酒度・BMI の差異に影響されない形で表しています。一方、$\beta_0$ は定数項で、説明変数で説明できない部分の大きさを表します。

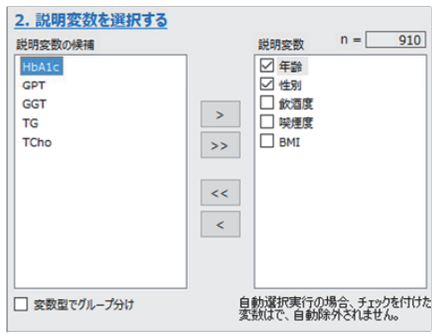



**手順1** 目的変数を選択する ①　　回帰モデルの目的変数を選択します。

**手順2** 説明変数を選択する ②

回帰モデルに組み込む説明変数を、候補から選択します。ステップ実行時の有効データ数は「n = □」に表示されます。

［＞］：候補中の選択項目を追加
［＞＞］：全ての候補を追加
［＜＜］：全ての説明変数を候補に戻す
［＜］：選択した説明変数を候補に戻す




変数名の前のチェックを入れると、「**変数自動選択実行**」時に除外されなくなります。年齢、性別など、他の説明変数との関連が想定される変数は、制御変数としてチェックを入れます。

## ③ 計算を実行する

**手順3** a. ステップ実行

選択された説明変数からなる重回帰モデルを作成し、分析結果（重回帰分析表）を出力します。出力される情報と、その解釈は「結果出力」の通りです。

**手順3'** b. 変数自動選択実行

選択された説明変数の中から、最適なモデルを自動的に選びます。実行法などの設定は④ [変数自動選択実行]タブで設定します。

## オプション設定

### ④ 実行設定

● ステップワイズ法
選択されている説明変数を、選択順にモデルに組み込み、P値が⑥の限界値よりも小さい変数をモデルに組み込みます。この過程を、3回繰り返すことで、最適なモデルを構築する方法です。

● 全組み合わせ法
全説明変数の組合せで回帰モデルを作り、最も適合度の良いものを自動選択します。計算所用時間を考慮し、説明変数の最大数は、初期値が10です。

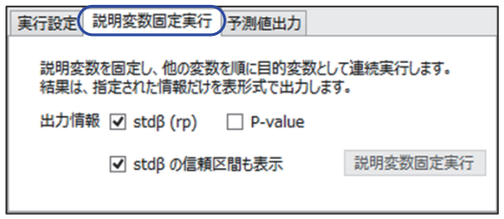
● 限界P値
ステップワイズ法で、説明変数をモデルに残す限界とする有意確率を指定します。

### ⑤ 分析結果の詳細出力

標準偏回帰係数の信頼区間や多重共線性指標が必要な場合にチェックします。説明変数間の相関が強いと多重共線性により異常解となります。その状況は、**多重共線性評価指標(VIF, CN)から推定できます。**分散拡張係数VIF>10や条件番号CN>30の場合、多重共線性が生じている可能性が高くなります。

### ⑥ 説明変数固定実行

説明変数の候補として上がっている全て項目について、選択された説明変数を固定して自動的に重回帰分析を一括して行う機能です。出力の最後に、std$\beta$ やP値の一覧表を出力します。



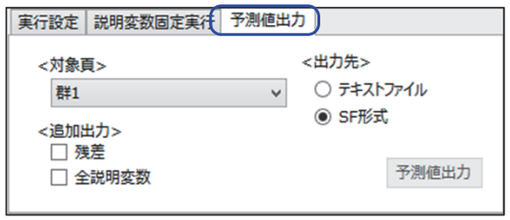

### ⑦ 予測値出力

全説明変数を指定すると、回帰に用いた説明変数の値を同時に出力します。

● **対象頁**　の指定で、得られた回帰式を、異なる頁のデータに適用して予測値を算出できます。

● **出力先**
テキストファイル：ファイル名の指定が必要です。
SF形式：StatFlexのデータシートが作成されます。

● **追加出力**
残差（予測値ー実測値）や説明変数の値も同時出力するかを指定します。

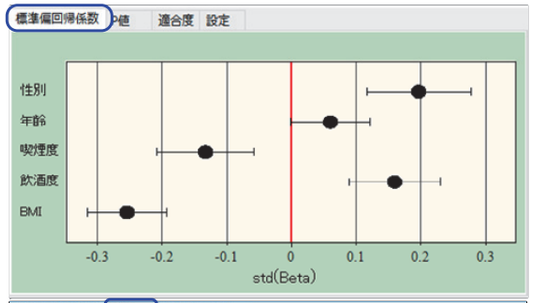


# 重回帰分析（2）

## ⑧ グラフ出力

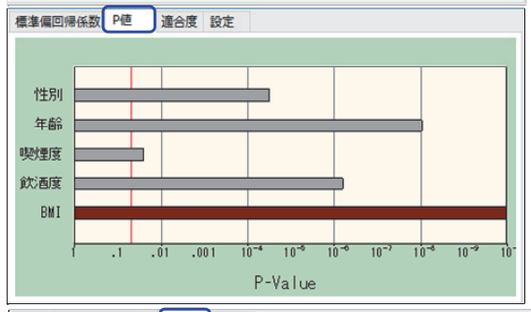
### a. 標準偏回帰係数比較図

選択された**説明変数**に対して、**std(β)値**を出力した記号髭を作成します。髭の幅は信頼区間で、設定値に従って表示されます。



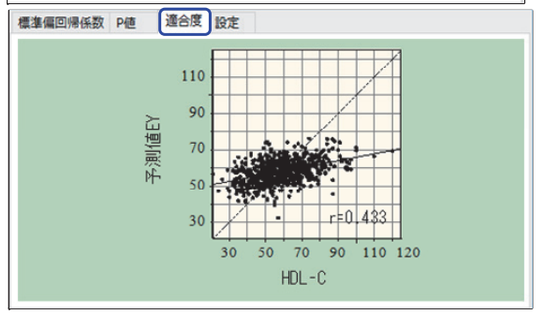

### b. P値比較図

選択された**説明変数**に対して、各P**値**を棒グラフで表示します。横軸は対数目盛でP値が0.001未満は指数形式表示します。$10^{-10}$未満の値は赤色の帯を表示します。




### c. 予測値適合度

**目的変数の観測値Y**と、回帰モデルから**予測値されたYの値(EY)**の関係を2次元散布図で表示します。相関係数は、重相関係数に相当し、回帰モデルの適合度を表します。




### d. 分析結果図表示設定

● std$\beta$ 表示域自動設定

チェックを外すと任意の範囲でstd$\beta$を表示することができます。

● P値表示域自動設定

チェックを外すと任意の範囲でP値を表示することができます。

## ⑨ 結果出力

目的変数：HDL-C　有効データ数=910　［第1頁：群1］

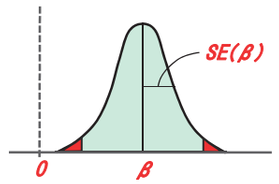
| 次数 | 変数名 | β | SE(β) | stdβ | t-val | df | P-val |
|---|---|---|---|---|---|---|---|
| 0 | | 80.18 | 4.574 | | | | |
| 1 | 性別 | 5.878 | 1.228 | 0.1963 | 4.7859 | 904 | 0.00000 |
| 2 | 年齢 | 0.08624 | 0.04508 | 0.0599 | 1.9131 | 904 | 0.05605 |
| 3 | 喫煙度 | -1.050 | 0.3034 | -0.1332 | -3.4596 | 904 | 0.00057 |
| 4 | 飲酒度 | 1.852 | 0.4195 | 0.1592 | 4.4161 | 904 | 0.00001 |
| 5 | BMI | -1.289 | 0.1573 | -0.2547 | -8.1933 | 904 | 0.00000 |

回帰の適合度指標
重相関係数R=0.3754 決定係数$R^2$=0.1409
決定係数調整後adj$R^2$=0.1362

**βとt値の関係式**

$$t = \frac{\beta}{SE(\beta)}$$

**βと有意性の関係図**




**変数名**　：説明変数の名前
**β**　：偏回帰係数を表し、個々の説明変数と目的変数との関連の強さを示す
**SE(β)**　：βの標準誤差
**stdβ**　：標準偏回帰係数を表し、単位に依存しない。多重共線性、交互作用なければ −1≦stdβ≦1
**t-val**　：目的変数との関連の有意性を表す値で、関連性が高いほど高値
df　：t値の自由度
**P-val**　：t値を有意確率に置き換えた値で、関連が強いほど低値

**重相関係数（R）**：目的変数と、回帰モデルによる予測値との相関係数　（0≦R≦1）

**決定係数調整後**（$R^2$adj）：決定係数(=重相関係数$^2$) を、説明変数の数を考慮して調整した値、モデルの適合度が高いほど高値となります。説明変数の選択では一般に、この値が増えれば残し、低下すれば除外します。

## ⑤ 多重共線性評価指標

分散拡大係数VIF：説明変数間の相関性の強さを表す指標でVIFはvariance inflation factorの略。p個の説明変数のうち1つ($x_i$)[i=1,2,.p]を目的変数に他のp-1個を説明変数として重回帰分析を行い、重相関係数(Ri)を求めれば、

$$\mathrm{VIF} = \frac{1}{1 - {R_i}^2}$$

として計算できます。VIFが10以上の場合、当該説明変数が関係した多重共線性により異常解が生じている可能性が高いと判断されます。またVIFの逆数は、トレランス(tolerance;Tol)と呼ばれ、小さいほど異常と判断します。

条件番号CN：説明変数間の相関性の強さを表す指標でCNはcondition numberの略。主成分分析を利用した、変数間の相互関連性（共通性）の強さを判定する方法です。p個の説明変数を使って主成分分析を行えば、元の変数との総合的な関連の強さを表すp個の固有値が得られます。p個の説明変数間の相関が強いほど、最大の固有値$\lambda_{max}$、と最小の固有値$\lambda_{min}$のギャップが大きくなり、次式のCNをその指標とします。

$$CN = \sqrt{\lambda_{max}/\lambda_{min}}$$

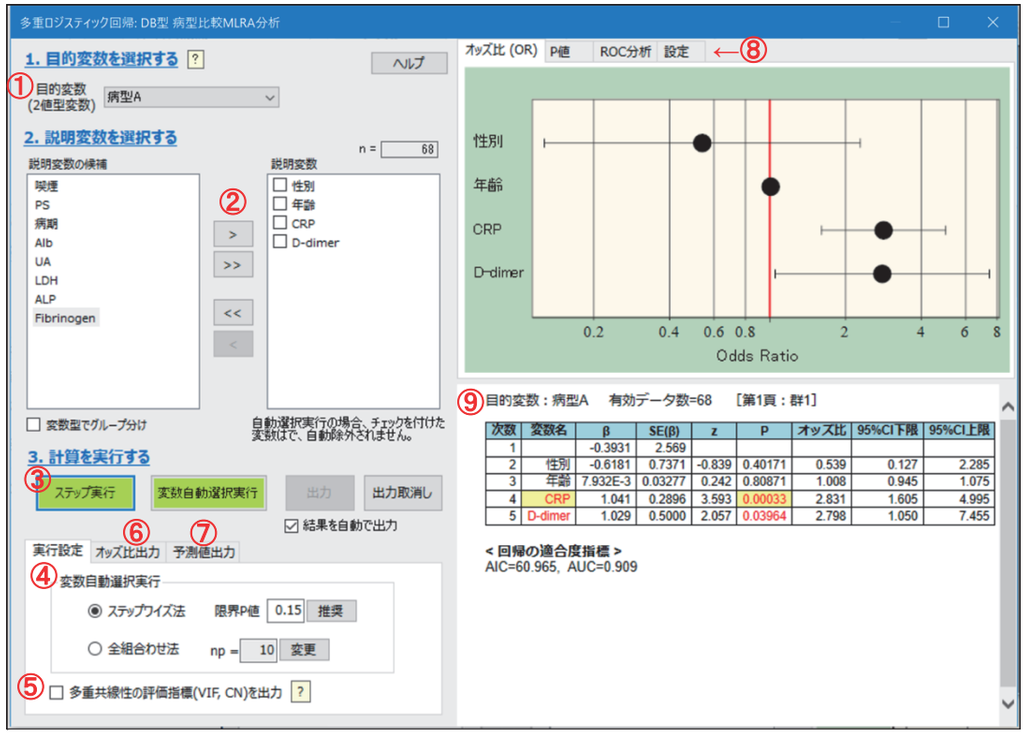
CNが30以上のとき、総合的に多重共線性が生じている可能性が高くなります。


統計

# 3-9 多重ロジスティック分析（1）

疾患群に属する確率 $p$ を目的変数とし、それに関連する要因（説明変数）を次式で組み合わせたのが多重ロジスティック回帰モデルです。各説明変数 $x_i$ の寄与度は回帰係数 $\beta_i$ に示されます。その有意性は $\beta_i$ をその標準誤差 $SE(\beta)$ で割った値 $z$ が、正規分布することを利用して判定します。

$$p = \frac{1}{1+e^{-X}} \qquad X = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$



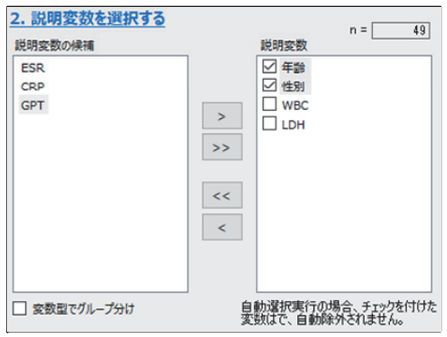

**手順1** 目的変数を選択する ①　　回帰モデルの目的変数を選択します。

**手順2** 説明変数を選択する ②

回帰モデルに組み込む説明変数を、候補から選択します。
ステップ実行時の有効データ数は「n = □」に表示されます。

［＞］：候補中の選択項目を追加
［＞＞］：全ての候補を追加
［＜＜］：全ての説明変数を候補に戻す
［＜］：選択した説明変数を候補に戻す

変数名の前のチェックを入れると、「**変数自動選択実行**」時に除外されなくなります。年齢、性別など、他の説明変数との関連が想定される変数は、制御変数としてチェックを入れます。




## ③ 計算を実行する

**手順3** a. ステップ実行

選択された説明変数からなる回帰モデルを作成し、分析結果（回帰分析表）を出力します。出力される情報と、その解釈は「結果出力」の通りです。

**手順3'** b. 変数自動選択実行

選択された説明変数の中から、最適なモデルを自動的に選びます。実行法などの設定は④［変数自動選択実行］タブで設定します。

## オプション設定

### ④ 実行設定

● ステップワイズ法

選択されている説明変数を、選択順にモデルに組み込み、P値が⑥の限界値よりも小さい変数をモデルに組み込みます。この過程を、3回繰り返すことで、最適なモデルを構築する方法です。

● 限界P値

ステップワイズ法で、説明変数をモデルに残す限界とする有意確率を指定します。

● 全組み合わせ法

全説明変数の組合せで回帰モデルを作り、最も適合度の良いものを自動選択します。計算所用時間を考慮し、説明変数の最大数は、初期値が10です。

### ⑤ 多重共線性評価指標(VIF, CN)

説明変数間の相関が強いと多重共線性により異常解となります。分散拡張係数VIF>10や条件番号CN>30の場合、多重共線性の可能性が高くなります。

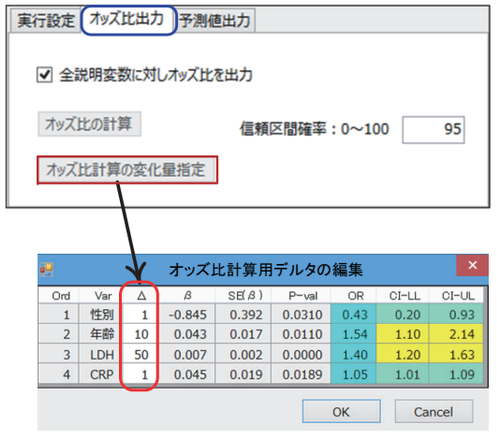
### ⑥ オッズ比出力

● 全説明変数に対しオッズ比(OR)を出力

チェックをはずと、2値型、段階型の説明変数に限定して、変化量1に対するORを計算します。

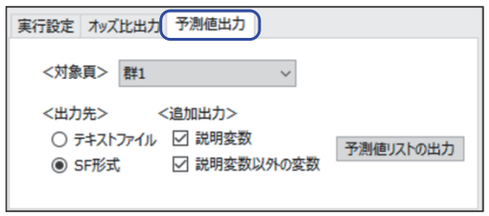
● オッズ比変化量Δの指定

ORの計算は、初期設定では、各説明変数が1変化した場合（Δ=1）について計算されます。Δの欄に実用上意味のある変化量を指定すれば、以降その値がORの計算で利用されます。




### ⑦予測値出力

回帰モデルにより計算される**予測値EP**（被検群に属する確率）を出力します。モデルを作成した群とは、別の群に対して予測値を出力することも可能です。




● 対象頁

予測値リストを作成する対象となる群を指定します。

● 出力先

テキストファイル：テキストファイルに出力
SF形式：StatFlexのデータシート形式で出力

● 追加出力

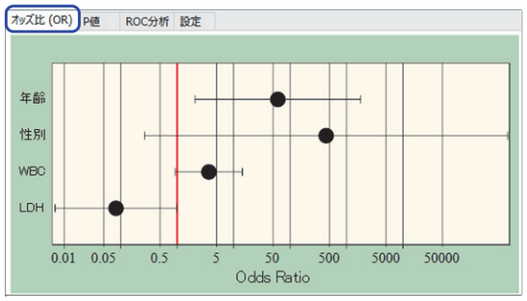
説明変数：回帰モデルに含まれる全説明変必要で数を追加出力
説明変数以外の変数：回帰モデルに含まない説明変数も追加出力、傾向スコア分析で必要になります

# 3-10 多重ロジスティック分析（2）

## ⑧ グラフ出力

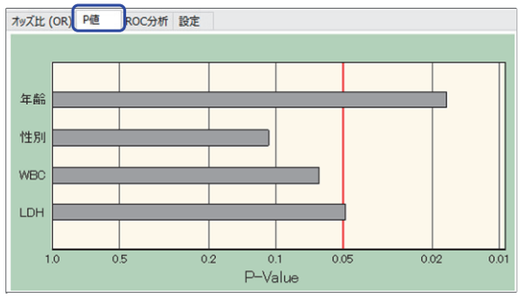
### a. オッズ比（OR)比較図

選択された**説明変数**に対して、オッズ比の信頼区間を作成します。任意のデルタ値に対応します。



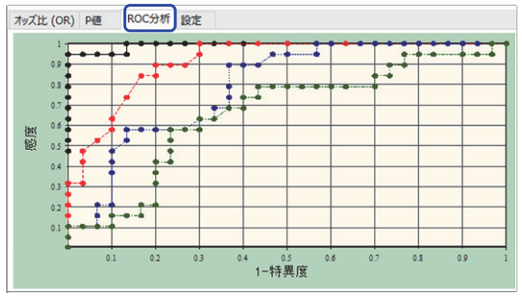

### b. P値比較図

選択された**説明変数**に対して、各P**値**を棒グラフで表示します。横軸は対数目盛でP値が0.001未満は指数形式表示します。$10^{-12}$未満の値は赤色の帯を表示します。




### c. ROC分析

回帰モデルから予測されるEP（被検群に属する確率）からROC曲線を描いて表示します。またメニュー設定により、モデル中の全説明変数について個別のROC曲線を重ねて描画できます。




### d. 分析結果図表示設定

#### ● OR表示域自動設定

チェックを外すと任意の範囲でOR表示域の指定できます。

#### ● P値表示域自動設定

チェックを外すと任意の範囲でP値を表示できます。

#### ● 説明変数に対してもROC曲線を表示

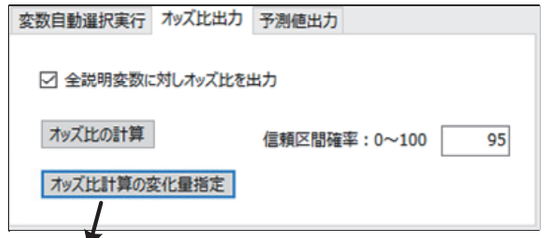
回帰予測値EPに対するROC曲線のみに限定する場合にはチェックを外します

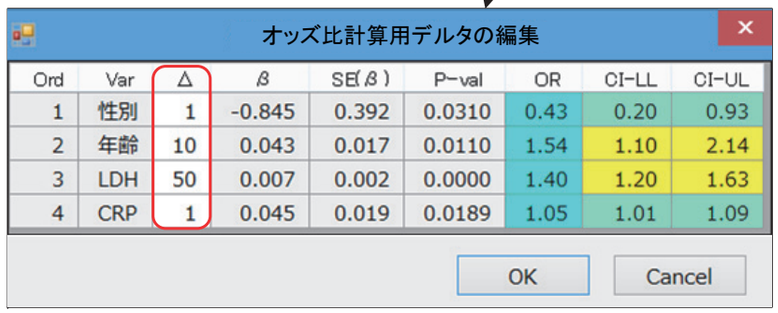
#### ● ROC分析モード

ROC曲線を描くモードと感度・特異度曲線を描くモードを切り替えます。後者はEPに対してのみ有効です。

## オッズ比の比較図作成上の注意点

オッズ比(OIR)は、自動では説明変数の変化量（Δ）が1の場合について計算されます。ダミー変数の場合はOKですが、一般には計測単位によってORの意味が異なるため、説明変数毎に意味のあるΔ値を指定すべきです。これには、[オッズ比計算の変化量指定]ボタンを押し、Δ値を変更します。




## ⑨ 結果出力

目的変数：細菌性　有効データ数=49　[第1頁：群1]

| 次数 | 変数名 | β | SE(β) | z | P | オッズ比 | Δβ | 95%CI下限 | 95%CI上限 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | -20.08 | 10.81 | | | | | | |
| 2 | 年齢 | 0.4104 | 0.1723 | 2.382 | 0.01721 | 60.603 | 10 | 2.070 | 1774.179 |
| 3 | 性別 | 6.076 | 3.777 | 1.609 | 0.10771 | 435.293 | 1 | 0.265 | 714455.044 |
| 4 | WBC | 1.292E-3 | 6.981E-4 | 1.851 | 0.06412 | 3.642 | 1000 | 0.927 | 14.307 |
| 5 | LDH | -0.1248 | 0.06333 | -1.971 | 0.04874 | 0.082 | 20 | 0.007 | 0.986 |

回帰の適合度指標
AIC=18.791, AUC=0.993

変数名　：説明変数の名前
$\beta$　：目的変数との関連の強さを表す数値で、モデル中の回帰係数に相当
SE($\beta$)　：$\beta$の標準誤差
z値　：$z = \beta / SE(\beta)$。説明変数の有意性を表し、有意性が高いほど高値
P　：z値を有意確率に置き換えた値で、関連が強いほど、低値
オッズ比　：$e^{\beta}$により計算される値で、当該説明変数の診断（判別）能を表す
95%信頼区間　：オッズ比の信頼区間

AIC　：モデルの適合度を表す指標です。良いモデルほど低値となります。
標準化された指標ではないため、異なるデータセット間で、AICの比較は無効です。

AUC　：作成したモデルによる予測値EPを使えば、2群をどの程度適確に判別できるかを表した値です。
0.5から1.0の範囲をとり、**1.0に近いほど**回帰モデルの診断的有用性が高いことになります。

# 3-11 傾向スコア分析

臨床研究において、後ろ向き調査で、2つの治療法の有効性の比較分析を行う場合、2治療法を適用条件が異なっておれば、平等な比較になりません。そこで、治療法の適用に影響する要因(背景因子)を多重ロジスティック回帰分析で調べ、一方の治療法を適用する確率を傾向スコアとして求めます。そして、2つの治療群から、傾向スコアが同程度である症例を抽出し、それに限定して2治療効果の比較を行います。 傾向スコア分析の応用範囲は広く、異なる治療法のデータ統合以外にも、異なる施設の臨床データの統合にも利用できるます。

手順1:クロス集計表で、2つの治療法(A,B)群の背景因子の差違を単変量解析

手順2:多重ロジスティック分析で2治療群の背景因子を分析

手順3:治療Aが使われる確率:傾向スコア(EP*)を、全説明変数と共に出力

手順4:傾向スコアを目安に、マッチング機能で2治療群からマッチしたデータを抽出

手順5:クロス集計表で、マッチング後の2治療群の背景因子および予後を単変量解析

手順6:必要に応じて、多変量解析で、2治療法における予後の差違を分析

* **EP**: estimated probability to be assigned to therapy A

**例題**:新薬の有無と、生命予後の関係を、後ろ向き調査で調べたデータ。新薬の転帰(死亡)との関連を、治療の背景因子(5変数)を揃えて比較したい。Kereiakes DJ. Am Heart J 2000;140:603-10

| 死亡 | 新薬 | 性別 | ステント | 糖尿病 | 心筋梗塞 | 駆出率 |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 | 0 | 56 |
| 1 | 0 | 1 | 0 | 1 | 0 | 50 |
| 1 | 0 | 1 | 0 | 1 | 0 | 30 |
| 1 | 0 | 1 | 0 | 0 | 0 | 51 |
| 1 | 1 | 1 | 0 | 0 | 0 | 50 |
| 0 | 1 | 0 | 1 | 1 | 1 | 49 |
| 0 | 1 | 0 | 1 | 0 | 1 | 45 |
| 0 | 1 | 0 | 1 | 0 | 1 | 15 |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 |

①

## ① クロス集計で2治療の背景因子を分析

| 新薬 | | 1 | 0 | 検定 |
|---|---|---|---|---|
| 性別 | 1 | 231 (33.1) | 115 (38.6) | $\chi^2$=2.783 |
| | 0 | 467 (66.9) | 183 (61.4) | P=0.09530 |
| ステント | 1 | 492 (70.5) | 174 (58.4) | $\chi^2$=13.796 |
| | 0 | 206 (29.5) | 124 (41.6) | P=0.00020 |
| 糖尿病 | 1 | 143 (20.5) | 80 (26.8) | $\chi^2$=4.859 |
| | 0 | 555 (79.5) | 218 (73.2) | P=0.02750 |
| 心筋梗塞 | 1 | 125 (17.9) | 18 (6.0) | $\chi^2$=23.922 |
| | 0 | 573 (82.1) | 280 (94.0) | P=0.00000 |
| 駆出率 | n | 698 | 298 | t=-2.625 |
| | Mean | 50.4 | 52.3 | df=994 |
| | SD | 10.4 | 10.3 | P=0.00880 |
| 死亡 | 1 | 11 (1.6) | 15 (5.0) | $\chi^2$=9.821 |
| | 0 | 687 (98.4) | 283 (95.0) | P=0.00173 |

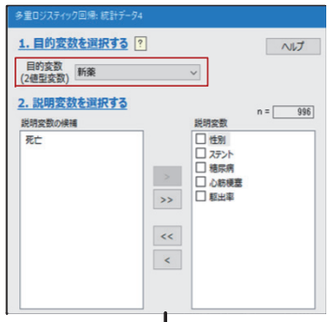
治療法の選択に関わる可能性のある変数に偏りがないかチェックします。

転帰(死亡)との関係はマッチング後に判定

②

## ② 多重ロジスティック回帰で新薬利用の背景因子を分析

目的変数に治療法(新薬)を選択。
説明変数に、転帰(死亡)以外を選択

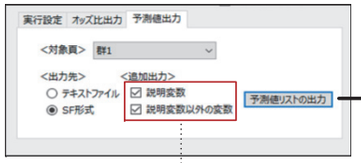



目的変数:新薬　有効データ数=996　[第1頁 : 群1]

| 次数 | 変数名 | β | SE(β) | z | P | OR | OR下限 | OR上限 |
|---|---|---|---|---|---|---|---|---|
| 0 | | 1.289 | 0.3991 | | | | | |
| 1 | 性別 | -0.2035 | 0.1482 | -1.373 | 0.16975 | 0.816 | 0.610 | 1.091 |
| 2 | ステント | 0.5363 | 0.1469 | 3.652 | 0.00026 | 1.710 | 1.282 | 2.280 |
| 3 | 糖尿病 | -0.3532 | 0.1660 | -2.127 | 0.03341 | 0.702 | 0.507 | 0.973 |
| 4 | 心筋梗塞 | 1.178 | 0.2665 | 4.421 | 0.00001 | 3.247 | 1.926 | 5.475 |
| 5 | 駆出率 | -0.01475 | 7.26E-03 | -2.033 | 0.04210 | 0.985 | 0.971 | 0.999 |

AIC=1175.656, AUC=0.618

## ③ 傾向スコア(EP)を出力



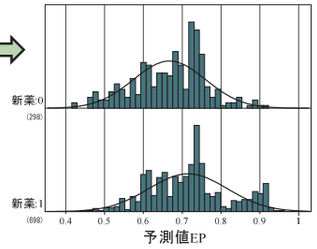

説明変数以外の変数も含めて予測値EPをStatFlex形式で出力

③

| 予測値EP | 新薬 | 性別 | ステント | 糖尿病 | 心筋梗塞 | 駆出率 | 死亡 |
|---|---|---|---|---|---|---|---|
| 0.477 | 1 | 1 | 0 | 1 | 0 | 56 | 1 |
| 0.499 | 0 | 1 | 0 | 1 | 0 | 50 | 1 |
| 0.572 | 0 | 1 | 0 | 1 | 0 | 30 | 1 |
| 0.583 | 0 | 1 | 0 | 0 | 0 | 51 | 1 |
| 0.586 | 1 | 1 | 0 | 0 | 0 | 50 | 1 |
| 0.642 | 0 | 1 | 0 | 1 | 0 | 10 | 1 |
| 0.660 | 1 | 0 | 1 | 1 | 0 | 55 | 1 |

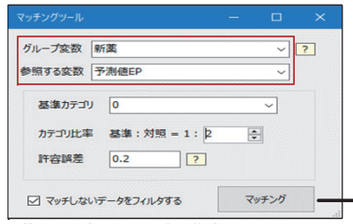
④

治療法(新薬)で群分けして予測値EPを比較




## ④ EPを目安に2治療群をマッチング

グループ変数に治療法(新薬)、参照変数に予測値EPを選択、
2治療法のデータ数比3:7に対し、1:2でのマッチング例です。




基準カテゴリでは、比率の指定で、0と1のどちらを基準にするかを決めます。

| ペア番号 | 予測値EP | 新薬 | 性別 | ステント | 糖尿病 | 心筋梗塞 | 駆出率 | 死亡 |
|---|---|---|---|---|---|---|---|---|
| 10 | 0.477 | 1 | 1 | 0 | 1 | 0 | 56 | 1 |
| 1 | 0.499 | 0 | 1 | 0 | 1 | 0 | 50 | 1 |
| 2 | 0.583 | 0 | 1 | 0 | 0 | 0 | 51 | 1 |
| 20 | 0.586 | 1 | 1 | 0 | 0 | 0 | 50 | 1 |
| 245 | 0.642 | 0 | 1 | 0 | 1 | 0 | 10 | 1 |
| 3 | 0.66 | 1 | 0 | 1 | 1 | 0 | 55 | 1 |
| 34 | 0.612 | 1 | 1 | 1 | 1 | 0 | 55 | 1 |

⑤

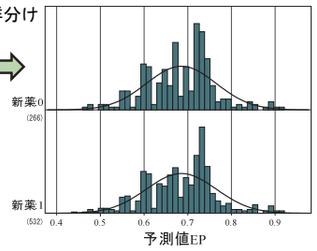
## ⑤ マッチング後のクロス集計表

| 新薬 | | 1 | 0 | 検定 |
|---|---|---|---|---|
| 予測値EP | n | 532 | 266 | t=-0.003 |
| | Mean | 1 | 1 | df=796 |
| | SD | 0 | 0 | P=0.99775 |
| 性別 | 1 | 173 (32.5) | 92 (34.6) | $\chi^2$=0.342 |
| | 0 | 359 (67.5) | 174 (65.4) | P=0.55879 |
| ステント | 1 | 356 (66.9) | 172 (64.7) | $\chi^2$=0.403 |
| | 0 | 176 (33.1) | 94 (35.3) | P=0.52553 |
| 糖尿病 | 1 | 111 (20.9) | 57 (21.4) | $\chi^2$=0.034 |
| | 0 | 421 (79.1) | 209 (78.6) | P=0.85386 |
| 心筋梗塞 | 1 | 31 (5.8) | 18 (6.8) | $\chi^2$=0.272 |
| | 0 | 501 (94.2) | 248 (93.2) | P=0.60213 |
| 駆出率 | n | 532 | 266 | t=0.732 |
| | Mean | 52.1 | 51.6 | df=796 |
| | SD | 8.6 | 10.1 | P=0.46410 |
| 死亡 | 1 | 8 (1.5) | 14 (5.3) | $\chi^2$=9.349 |
| | 0 | 524 (98.5) | 252 (94.7) | P=0.00223 |

⑥

マッチング後は治療法によって関連変数に差違がないか確認します。

治療法で群分け




## ⑥ マッチング後の予後分析

転帰(死亡)を目的変数に、多重ロジスティック分析を行い、治療法との関係判定します。

目的変数:死亡　有効データ数=798

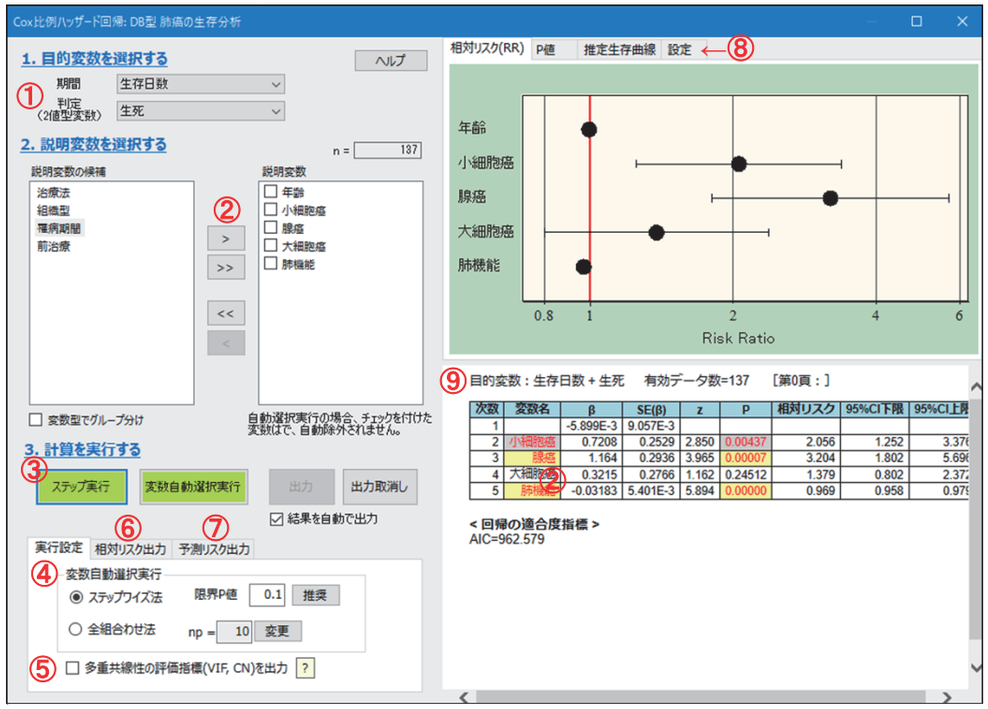
| 次数 | 変数名 | β | SE(β) | z | P | OR | Δ | OR下限 | OR上限 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | | -1.3550 | 0.9714 | | | | | | |
| 1 | 新薬 | -1.2010 | 0.4610 | -2.605 | 0.00918 | 0.301 | 1 | 0.122 | 0.743 |
| 2 | 性別 | 0.4907 | 0.4528 | 1.084 | 0.27843 | 1.634 | 1 | 0.673 | 3.968 |
| 3 | ステント | 0.4008 | 0.4922 | 0.814 | 0.41555 | 1.493 | 1 | 0.569 | 3.918 |
| 4 | 糖尿病 | 0.8079 | 0.4736 | 1.706 | 0.08802 | 2.243 | 1 | 0.887 | 5.675 |
| 5 | 心筋梗塞 | 1.3800 | 0.6164 | 2.238 | 0.02521 | 3.973 | 1 | 1.187 | 13.300 |
| 6 | 駆出率 | -0.04958 | 0.01805 | -2.748 | 0.00600 | 0.609 | 10 | 0.428 | 0.868 |

AIC=189.172, AUC=0.783

# 3-12 Cox 比例ハッザード回帰分析（1）

疫学研究の対象となる1群の集団（コホート）を経時的に観察し、ある特定の病態・状態が起こるまでの期間に関与する要因を分析するのに利用します。Cox比例ハッザード回帰により、特定の要因の有無で、時点tにおいて、ある病態（状態）の生じるリスク（ハッザード）*h(t)*がどの程度高いかを次式により定量的に評価できます。p個の要因（説明変数：$x_1, x_2, \ldots x_p$）の重要性は、相対リスク(RR)として、回帰係数βi から計算できます。

$$h(t) = h_0(t)e^{\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}$$




**手順1** 目的変数を選択①　観察期間と判定情報（2値型）の2つの指定が必要です。

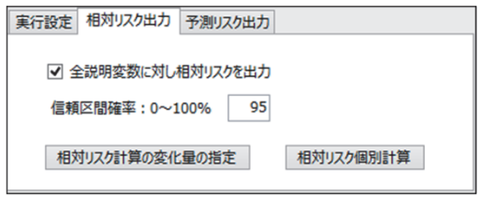
**手順2** 説明変数を選択②

**手順3** 計算を実行　a. ステップ実行③　選択後の説明変数から回帰モデルを作成

b. 変数自動選択実行④　選択後の説明変数から最適モデルを自動作成

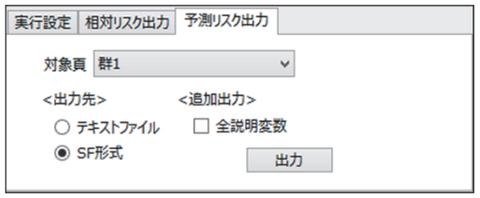
■ 多重共線性評価指標(VIF, CN)　⑤　説明変数間の相関が強いと多重共線性により異常解となります。分散拡張係数VIF>10や条件番号CN>30の場合、多重共線性の可能性が高くなります。

■ 相対リスク(RR)の計算と調整⑥




説明変数毎に、適切な変化量を指定してRRを計算

■予測リスクの出力⑦

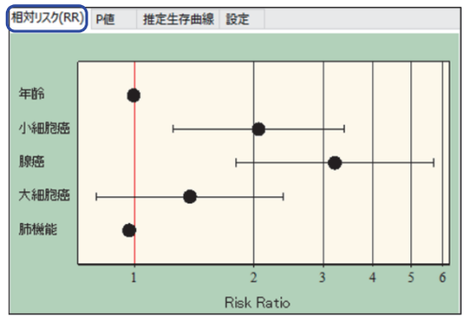



回帰モデルから、各個人の予測リスクを出力

■ グラフ出力⑧

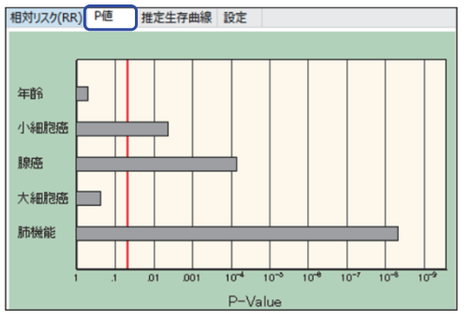
a. 相対リスク（RR）比較図

選択された**説明変数**に対して、相対リスクとその信頼区間を表示します。任意のデルタ値の変化量の指定(7に対応します。



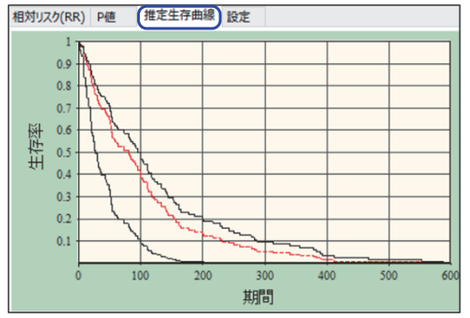

b. P値比較図

選択された**説明変数**に対して、各P**値**を棒グラフで表示します。横軸は対数目盛でP値が0.001未満は指数形式表示します。$10^{-12}$未満の値は赤色の帯を表示します。




c. 推定生存曲線

回帰モデルから予測されるEP（被検群に属する確率）からROC曲線を描いて表示します。またメニュー設定により、モデル中の全説明変数について個別のROC曲線を重ねて描画できます。




d. 分析結果図表示設定

● RR表示域自動設定
チェックを外すと任意の範囲でRR表示域を指定できます。

● P値表示域自動設定
チェックを外すと任意の範囲でP値を表示できます。ただし$10^{-10}$が下限です。

● 生存曲線表示域自動設定
チェックを外すと縦軸、横軸の範囲を任意に指定できます。

● 推定生存曲線の群分け
群分け用変数：指定された2値型変数に応じて、生存曲線を分割して表示
群分けしない生存曲線も表示：
群分け前の曲線を赤色の曲線で表示

■ 結果出力

**<< Cox比例ハザード回帰 >>**

目的変数：生存日数＋生死　有効データ数=137　［第0頁：］

| 次数 | 変数名 | β | SE(β) | z | P | 相対リスク | 95%CI下限 | 95%CI上限 |
|---|---|---|---|---|---|---|---|---|
| 1 | 年齢 | -5.899E-3 | 9.057E-3 | 0.651 | 0.51486 | 0.994 | 0.977 | 1.012 |
| 2 | 小細胞癌 | 0.7208 | 0.2529 | 2.850 | 0.00437 | 2.056 | 1.252 | 3.376 |
| 3 | 腺癌 | 1.164 | 0.2936 | 3.965 | 0.00007 | 3.204 | 1.802 | 5.696 |
| 4 | 大細胞癌 | 0.3215 | 0.2766 | 1.162 | 0.24512 | 1.379 | 0.802 | 2.372 |
| 5 | 肺機能 | -0.03183 | 5.401E-3 | 5.894 | 0.00000 | 0.969 | 0.958 | 0.979 |

**< 回帰の適合度指標 >**
AIC=962.579

**変数名**　：**説明変数の名前**
**β**　：**目的変数との関連の強さを表す数値で、モデル中の回帰係数に相当**
**SE(β)**　：**βの標準誤差**
**z値**　：**z = β/SE(β)。説明変数の有意性を表し、有意性が高いほど高値**
**P**　：**z値を有意確率に置き換えた値で、関連が強いほど、低値**
**相対リスク**　：**$e^{\beta}$により計算される値で、当該説明変数の診断（判別）能を表す**
**95%信頼区間**：**相対リスクの信頼区間**

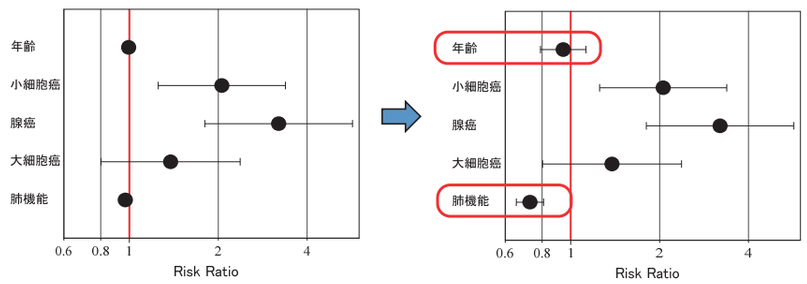
AIC　：モデルの適合度を表す指標です。良いモデルほど低値となります。
　　標準化された指標ではないため、異なるデータセット間で、AICの比較は無効です。

## 相対リスク(RR)を作図する際の注意点

相対リスクは、自動計算の場合、変化量デルタ（Δ）が1の場合の結果になります。ダミー変数ではOKですが、適切な計算には、説明変数毎に、臨床的に意味のある変化量を指定する必要があります。これには、⑥で相対リスク個別計算ボタンで、Δ値を変更します。下の例は、肺機能の値が0～100の値を取るため、Δを1から10の変化に、年齢はΔを1から10才の変化に指定した例です。

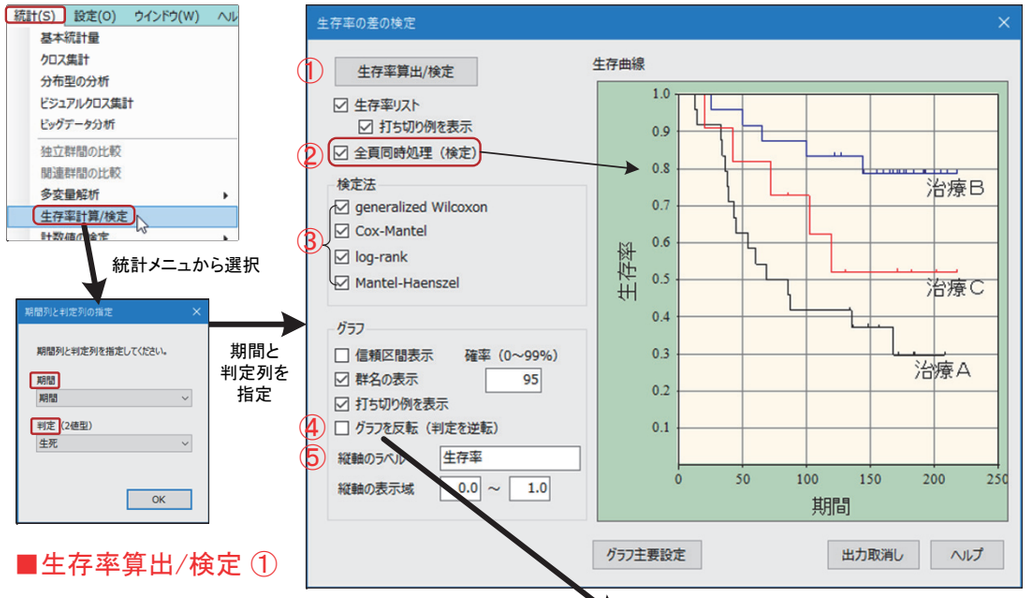
目的変数：生存日数＋生死　有効データ数=137　［第0頁：］

| 次数 | 変数名 | β | SE(β) | z | P | 相対リスク | Δβ | 95%CI下限 | 95%CI上限 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 年齢 | -5.899E-3 | 9.057E-3 | 0.651 | 0.51486 | 0.943 | 10 | 0.789 | 1.126 |
| 2 | 小細胞癌 | 0.7208 | 0.2529 | 2.850 | 0.00437 | 2.056 | 1 | 1.252 | 3.376 |
| 3 | 腺癌 | 1.164 | 0.2936 | 3.965 | 0.00007 | 3.204 | 1 | 1.802 | 5.696 |
| 4 | 大細胞癌 | 0.3215 | 0.2766 | 1.162 | 0.24512 | 1.379 | 1 | 0.802 | 2.372 |
| 5 | 肺機能 | -0.03183 | 5.401E-3 | 5.894 | 0.00000 | 0.727 | 10 | 0.654 | 0.809 |






カプランマイヤー(Kaplan-Meier)方式で生存率の経時変化を計算し、生存曲線を描画します。
計算に必要な情報は、生存期間を記録した変数列と、生死判定を記録した変数列（必ず2値型）です。群分け後のデータでは、全群について分析し、曲線の差を検定できます。コホート研究に広く適用でき、発症の有無、治療の成否などの判定情報と、判定までの期間の記録から分析できます。




### ■生存率算出/検定 ①

**< 生存率表：Survival Table >**

NA＝対象数(N at risk)　NE＝イベント数(N events)　NC＝打切り数

| SQN | 期間 | 生存率 | 標準誤差 | 95%信頼区間 | | NA | NE | NC |
|---|---|---|---|---|---|---|---|---|
| 1 | 12 | 0.9600 | 0.03919 | 0.88319 | 1.00000 | 25 | 1 | 1 |
| 2 | 14 | 0.9183 | 0.05542 | 0.80963 | 1.00000 | 23 | 1 | 0 |
| 3 | 33 | 0.8765 | 0.0668 | 0.7456 | 1.00000 | 22 | 1 | 0 |
| 4 | 34 | 0.8348 | 0.07554 | 0.68673 | 0.98284 | 21 | 1 | 0 |
| 5 | 36 | 0.7930 | 0.08249 | 0.63136 | 0.95472 | 20 | 1 | 0 |
| 14 | 87 | 0.4174 | 0.10068 | 0.22005 | 0.61473 | 11 | 1 | 0 |
| 15 | 134 | 0.4174 | | | | 10 | 0 | 1 |
| 16 | 135 | 0.371 | 0.09961 | 0.17579 | 0.56624 | 9 | 1 | 0 |
| 17 | 136 | 0.371 | | | | 8 | 0 | 1 |

中央生存期間 median survival time＝85.00 95%CI＝43.00～168.00

### ■生存曲線の差の検定 ③（②で全頁同時処理の場合に有効）

< 1:治療A > vs. <2:治療C>

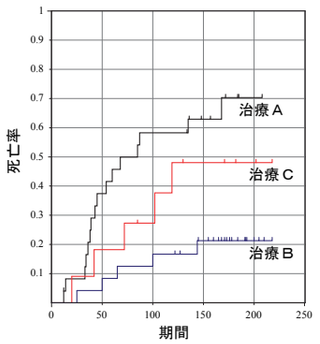
| | | | |
|---|---|---|---|
| GW test | z=1.290 | P=0.19718 | W=69, SD(W)=2862.619 |
| CM test | z=1.202 | P=0.22935 | U=-2.641, SD(U)=2.197 |
| LR test | $\chi^2$=1.435 | P=0.23095 | df=1 |
| MH test | $\chi^2$=0.950 | P=0.32983 | df=1 OR=1.857 |

< 1:治療A > vs. <3:治療B>

| | | | |
|---|---|---|---|
| GW test | z=3.224 | P=0.00126 | W=280, SD(W)=7541.327 |
| CM test | z=3.366 | P=0.00076 | U=-7.500, SD(U)=2.228 |
| LR test | $\chi^2$=11.119 | P=0.00085 | df=1 |
| MH test | $\chi^2$=9.870 | P=0.00168 | df=1 OR=5.061 |

GW : generalized Wilcoxon test　LR : log-rank test
CM : Cox-Mantel test　MH : Mantel-Haenszel test

### ■グラフの反転と表示域設定 ④

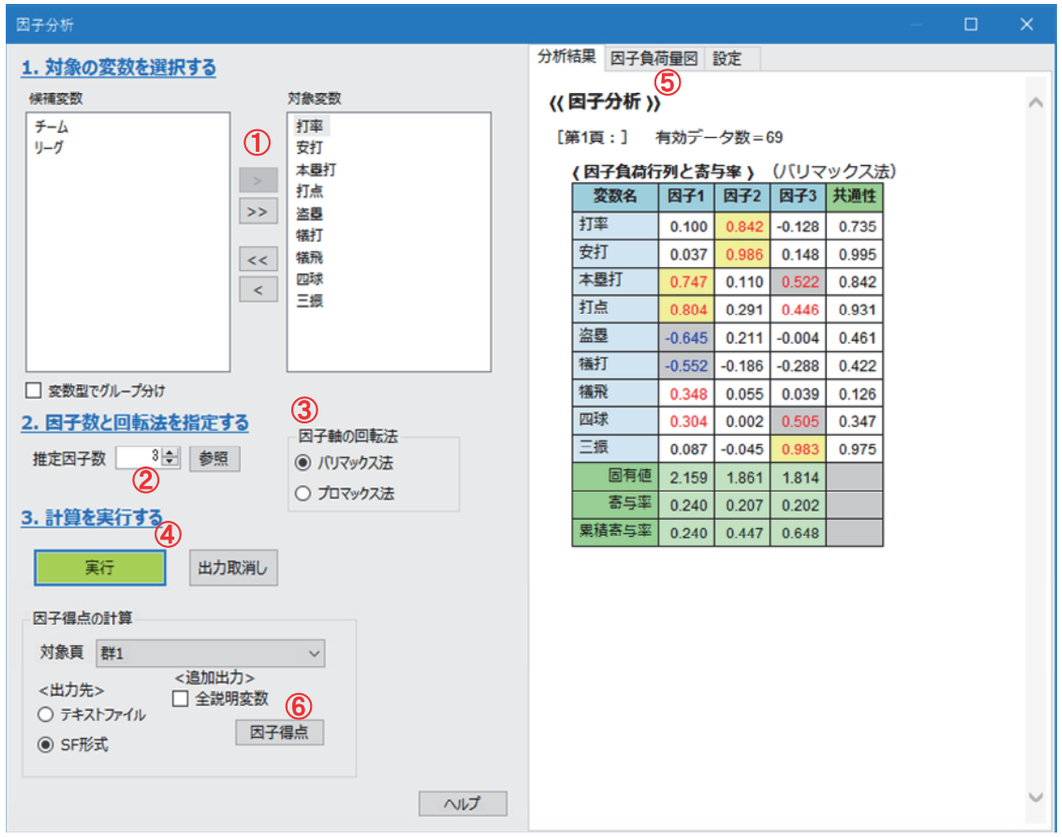



### ■縦軸ラベルの変更 ⑤

反転時に指定します。一般に、判定は生死に限らず、再発の有無や、成否などデータに応じて修正してください。

# 3-15 因子分析

複数の変数間の相互関係を分析し、変数間にどのような共通性のある要素（因子）が存在するかを調べる手段を提供します。例えば、野球選手の打撃成績変数（本塁打数、安打数、打点数、打率、犠飛数、盗塁数）の相関関係から、各変数値に共通性の高いものとして長打力や巧打力などの因子が括り出されます。そして、その2つの因子を元の変数がどの程度持ち合わせるかを調べることで、変数をグループ分けできるようになります。たとえば、本塁打数・打点数・犠飛数は長打力の因子と関連が強く、安打数・打率・盗塁数は好打力の因子が強いことになります（本節の例題を参照）。また、また各因子を個々の選手がどの程度持っているかを、因子得点の形で計算することが出来ます。




## ① 対象変数を選択する

分析に用いる変数を4つの矢印ボタンで選択します。最低2つ以上の変数を指定する必要があります。
選択順には意味はありません。

## ② 推定因子数を選択する

対象変数の数より小さ目の値を指定します。参照ボタンを押すと、次の3通りの方法で推定因子数を求めることができます。

1. 相関行列（R）から求めた固有値＞ 1.0 の数
2. R対角部を決定係数で置換し求めた固有値＞ 0 の数
3. 芝の経験則

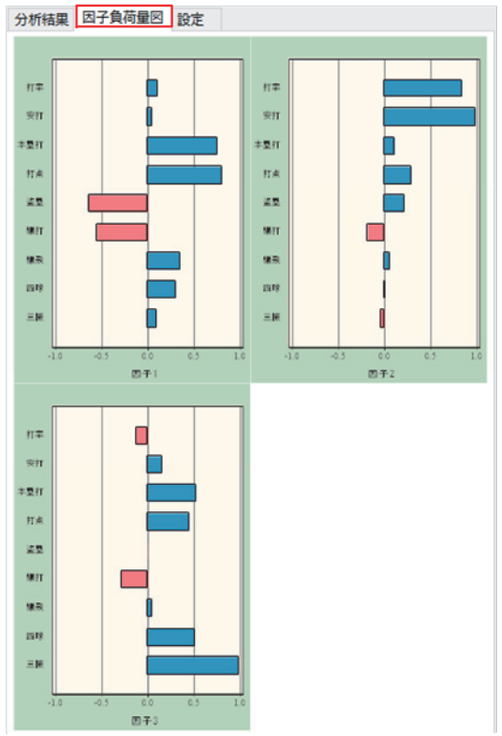
## ③ 因子軸の回転法を指定

バリマックス法またはプロマックス法を選択します。
因子負荷量からみて解釈しやすい因子の得られる方法を探索的に選ぶことになります。

## ④ 計算を実行

## ⑤ 表示オプションを設定

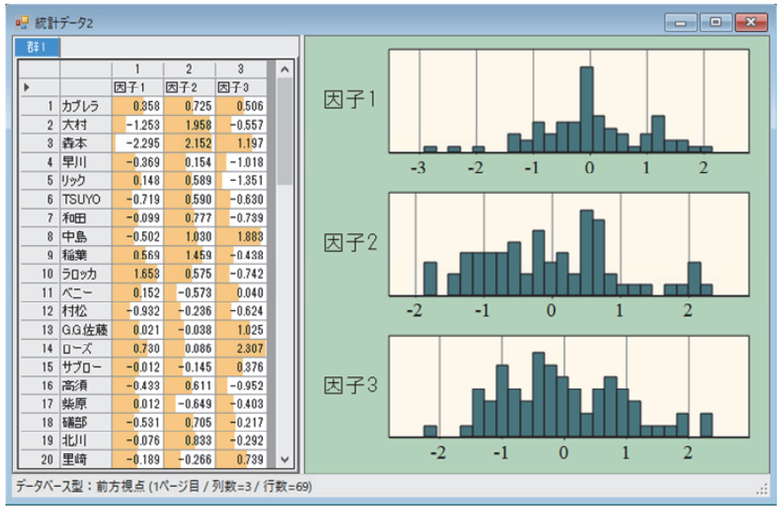
・因子分析表（左図）
・因子負荷量図（右図）



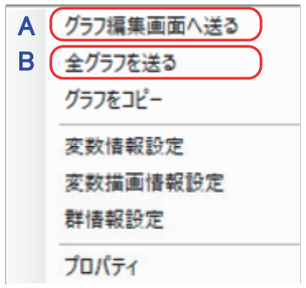

## ⑥ 因子得点を出力

○ テキストファイル　指定ファイル名でテキスト形式で出力します。
○ SF 形式ファイル　新しいデータシートに出力します。
□ 全説明変数にチェックを入れると、各説明変数の値が追加出力されます。

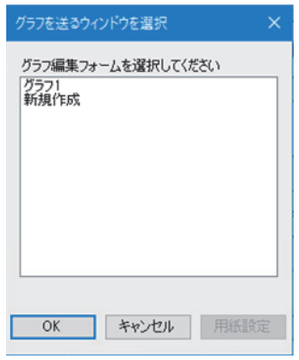
（下図は、チェックせずSF形式で出力した例）



# 4-1 グラフを編集ウインドウに送る



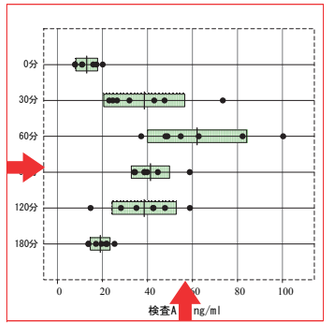

## A. グラフ編集画面へ送る



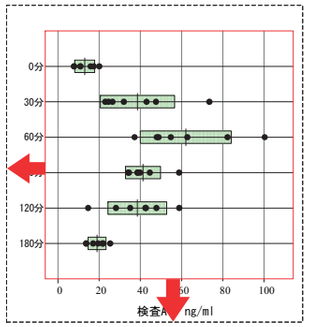

送り先のグラフ編集ウインドウを指定します。

既にグラフ編集ウインドウが存在する場合は、それらが全てリストに表示されます。新しいグラフ編集ウインドウに送る場合は、[新規作成]を選びます。

外枠を基準に描画




内枠を基準に描画




指定されたサイズで赤色の領域を確保した後、目盛り数値やラベルなどのグラフ部品を描画します。

**任意のサイズ**

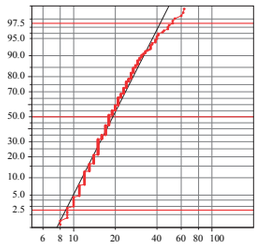
OKボタンを押したあと、グラフ編集ウインドウが表示されますので、そこでマウスでサイズを指定します。

**長さを指定**

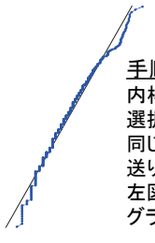
ここで指定したサイズで、グラフ編集ウインドウに送られます。

**グラフの重ね書き**

側方視点のグラフは、あらかじめ Var2 で表示域を揃えておくとグラフを重ね書きできます。

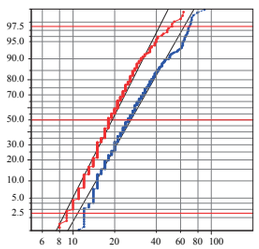



**手順i**
長さを指定してグラフ編集へ送ります。



**手順ii**
内枠基準で、[重ね書き]を選択し、手順iと同じサイズ、同じグラフ編集ウインドウへ送ります。
左図のような座標枠のないグラフが送られます。

**手順iii**
グラフ編集ウインドウ内で、手順iと手順iiで送られたグラフを移動させ、グラフを重ね合わせます。




## B. 全グラフを送る

### ① 対象グラフ
編集ウインドウへ送るグラフを選択します。番号順に配置されます。

### ② 行数・列数
グラフ編集ウィンドウに配置するグラフの行数と列数を指定します。
[行列数を指定]のチェックを外すと、行数および列数は自動的に決まります。

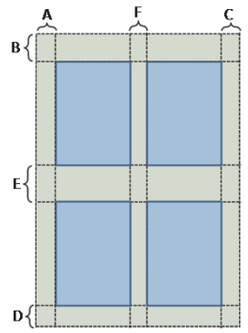
### ③ 配置順
①で指定したグラフをどう配置するか決めます。

### ④ 配置基準
各グラフを以下のように揃えます。
[左上]：左揃えかつ上揃えでグラフを配置します。
[中心]：水平方向と垂直方向ともにセンタリングしてグラフを配置します。

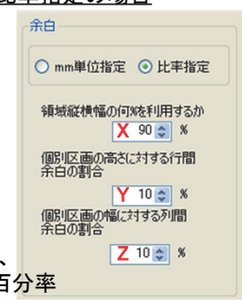
### ⑥ プレビュー
グラフがどのように編集インドウへ送られるかを確認できます。

### ⑤ 余白
用紙の4隅、行間、列間のマージンを指定します。Mm単位指定と比率指定の2通りの方法があります。
下図は、mm単位指定の場合の、それぞれの数値がどこの余白に対応するかを示します。

**mm単位指定の場合**

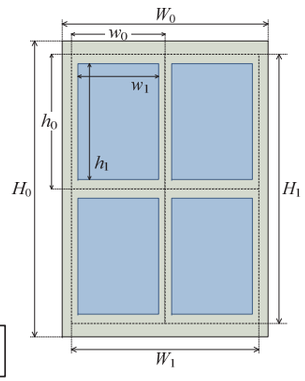



**比率指定の場合**



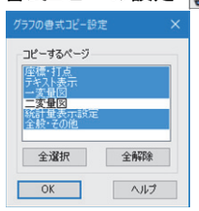

X, Y, Zは、
x, y, zの百分率

$$x = \frac{W_1}{W_0} = \frac{H_1}{H_0} \qquad y = \frac{h_0 - h_1}{h_0} \qquad z = \frac{w_0 - w_1}{w_0}$$



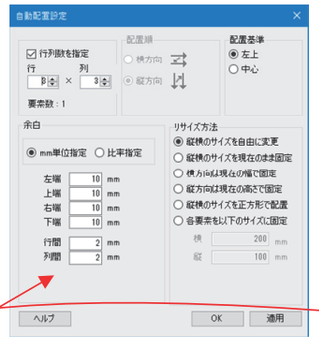
# 4-2 グラフ編集ウインドウ

## 書式コピーの設定



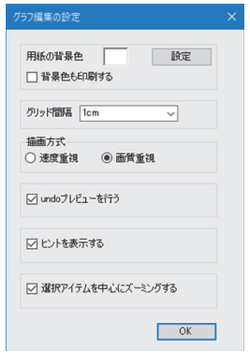

グラフ編集パネルのどのページの書式をコピーするか設定します。

## 自動配置



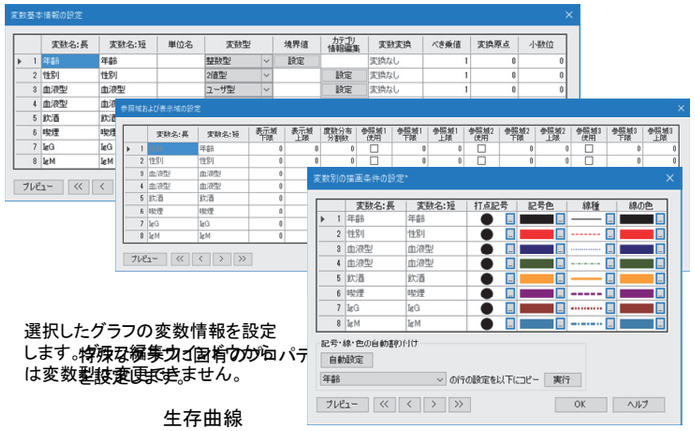

選択したグラフを一つ以上のグラフを格子状に整列するための設定です。

## グラフ編集ウィンドウの設定




背景色やグリッドの間隔など、グラフ編集ウインドウの実行環境設定です。

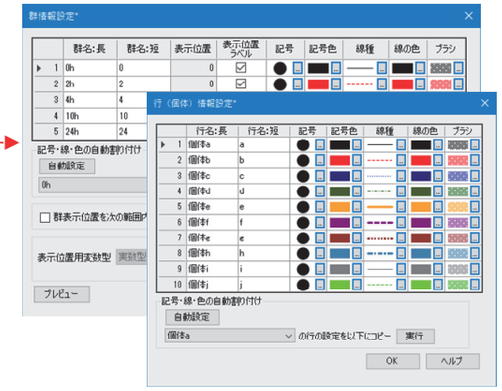
## 各種変数情報設定




選択したグラフの変数情報を設定します。特殊な編集以外はこのパネルは変数型は変更しません。

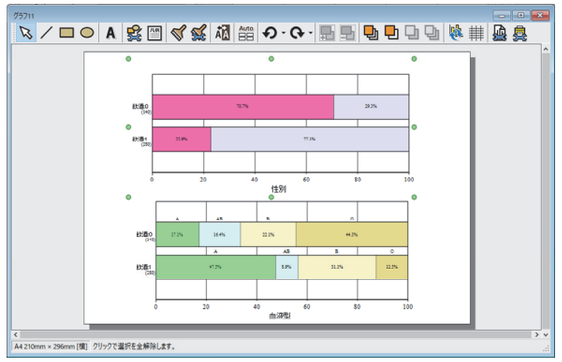
生存曲線

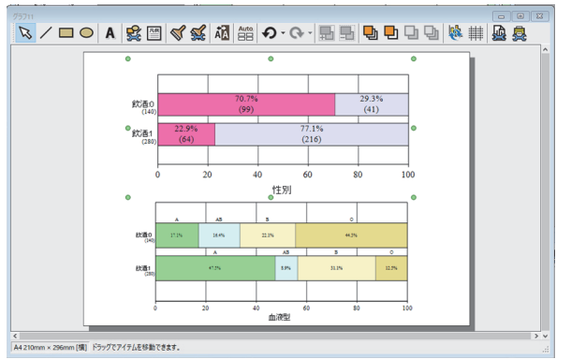
## 群情報および行情報設定



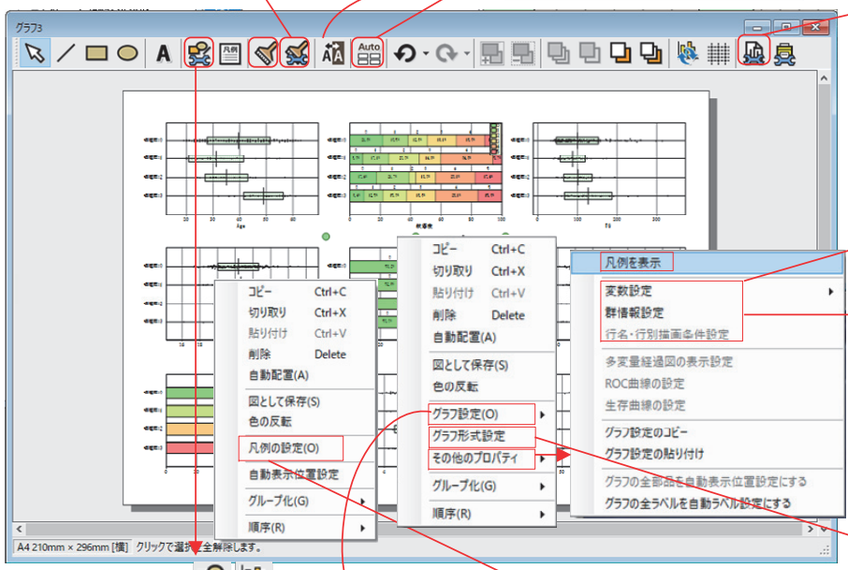

選択したグラフの群情報と行情報を設定します。

## 書式コピー

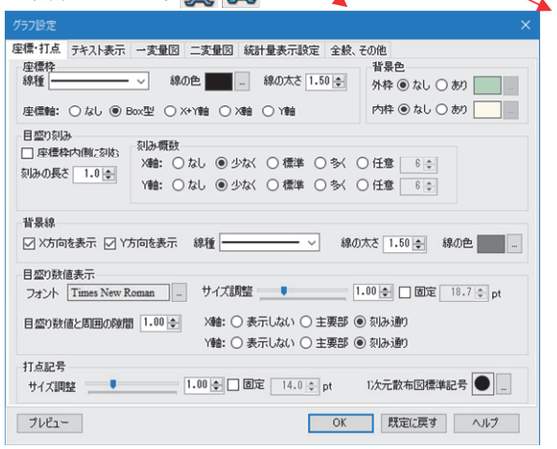



1. 書式をコピーしたいコピー元となるグラフを選び、 をクリック。
2. マウスのカーソルが か に変わったことを確認し、コピー先のグラフをクリックする。




## 図形プロパティ




選択しているものによって、表示されるパネルが異なります。
グラフの場合⇒グラフ設定パネル
図形（文字、線、楕円、矩形）の場合⇒図形の設定パネル
凡例の場合⇒凡例設定パネル（右）

## 凡例の設定

選択した凡例の書式を設定します。

凡例に対応したグラフ
・ROC曲線
・生存曲線
・相関行列図
（複数頁同時描画時）
・折れ線図

一部の情報はグラフ設定パネルの[テキスト表示]ページでも設定できます。

## グラフ形式設定

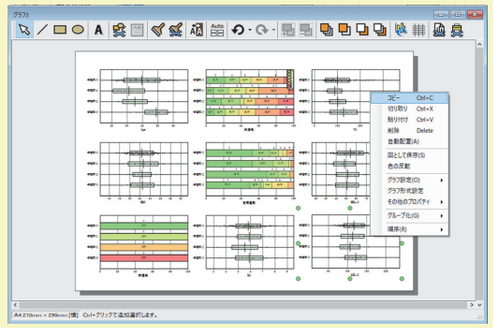
# 他のソフトウエアへのグラフ貼り付け



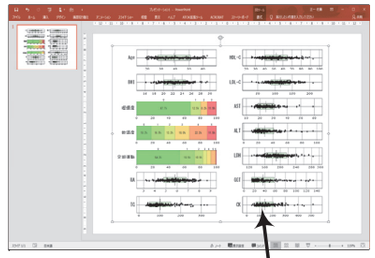
StatFlexでグラフをコピー

自動グラフから直接コピー

グラフ編集結果のコピー

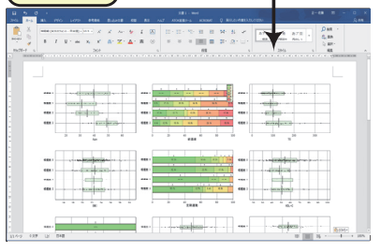


Powerpoint

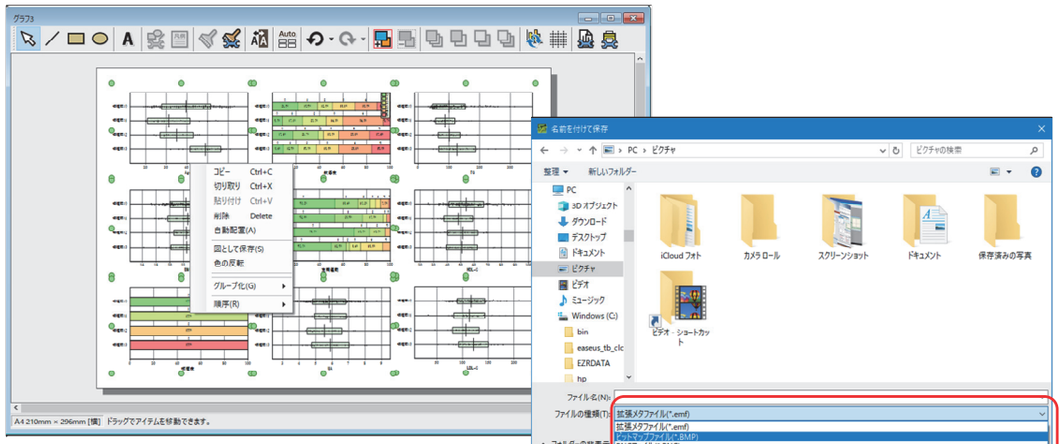


グループ化解除で部品化
すれば各要素を編集できます

Word

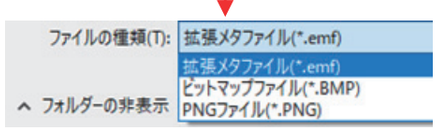


作成したグラフをStatFlex上でコピーした場合、そのデータは内部的には拡張メタファイル形式として保持されます。したがって、拡大縮小しても画像の劣化が起こりません。



作成したグラフはグラフ編集ウィンドウへ送ると、画像ファイルとして保存できます。
サポートしているファイル形式は以下の3種類です。

- 拡張メタファイル（拡張子はemf）
- ビットマップ（拡張子はbmp）
- Portable network graphics形式（拡張子はpng）

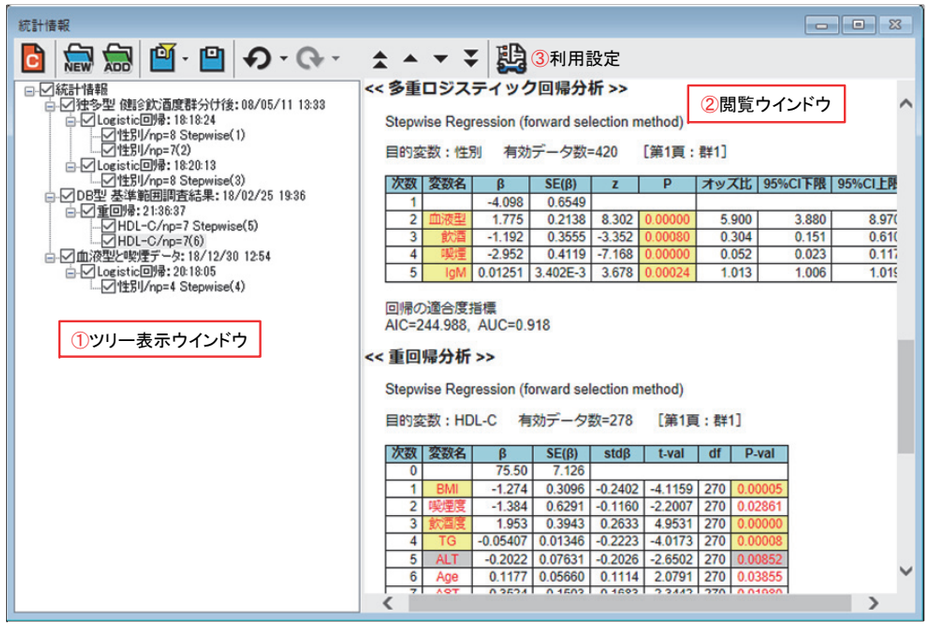






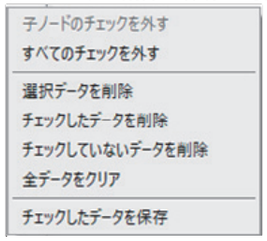





## ①ツリー表示ウインドウ

［統計情報］の1つ下の階層に、ファイル単位でまとめられます。次の右クリックメニューで選択したデータの削除や保存が可能です。




## ②閲覧ウインドウ

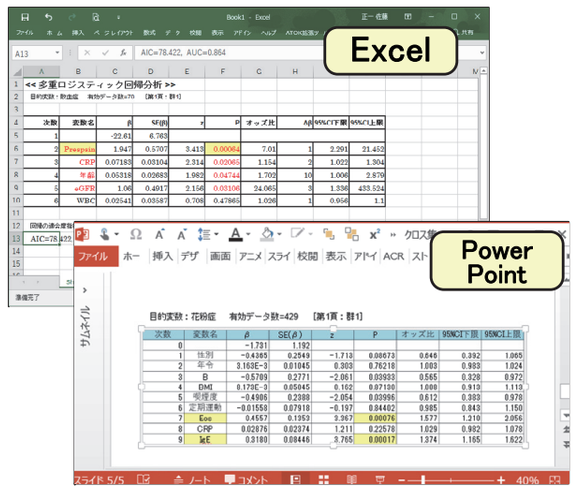
▲ ▼ 表示されている前/次の出力へ移動

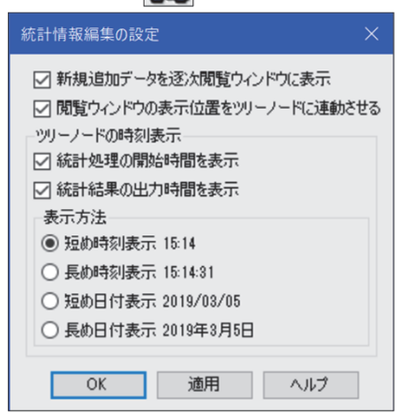
⏫ ⏬ 表示されている先頭/末尾の出力へ移動

Ctrl + F で検索機能の呼び出し

Ctrl + P で印刷機能を呼び出し

分析結果はほとんど表形式で出力されます。
必要な部分をコピーすればExcelやPowerPointに書式を保持したまま貼り付けできます。




## ③利用設定