

バージョン 3.1.0

2023/10

作成：StaatApp 運営チーム

StaatApp

操作マニュアル

目次

用語説明	4
基本操作	5
1. データを読み込む	5
2. データを複製する.....	8
3. データを保存する.....	8
4. 解析用ウィンドウを表示する.....	9
5. 解析を実行する	9
6. 解析結果をポップアウトする	10
7. 解析結果をコピーする	10
8. 解析結果を保存する.....	11
9. グラフを画像形式で保存する	12
10. 表示桁数を変更する.....	13
データ操作	14
1. 複数列を対象にソート（並び替え）を行う	14
2. 不要な行または列を削除する	15
3. 任意の文字列・数値を置換する	16
4. 欠測値を削除・補完する	17
5. データ型を変換する.....	19
6. 2つのデータを結合する	20
7. 特定の条件でデータを抽出する	21
8. メニューバーのデータ操作機能の選択.....	22
9. 列名（変数名）の変更.....	22
10. 差分・対数変換	23
11. テストデータの作成.....	24
解析の実行・結果の表示.....	25
1. 相関係数	25
2. 対応のない場合の仮説検定	28
3. 対応のある場合の仮説検定.....	30
4. ロジスティック回帰分析（無料）	32
5. 数量化Ⅱ類	36
6. 主成分分析	39
7. 因子分析	44
8. 構造方程式モデリング.....	49
9. クラスター分析.....	55
10. 決定木（無料）	61
11. ARIMA（無料）	68
12. GARCH モデル	73
13. VAR モデル	78

14. Prophet	84
15. クロス集計表	88
16. 生存曲線	93
17. Cox 比例ハザード回帰	97
18. アソシエーション分析	101
グラフ作成	106
1. グラフ作成機能の共通操作（個別系列設定）	106
2. 折れ線グラフ（無料）	108
3. 散布図（無料）	110
4. 散布図行列（無料）	113
5. ヒストグラム	115
6. 分布図	117
7. エラーバー（無料）	122
自動考察	124
FAQ	128
1. Windows 版起動時にエラーが発生する	128
2. macOS 版の起動方法	131

用語説明

ホーム画面（データ操作画面）は、以下のような構成となります。

画面最上部のメニューバーを選択することで、ファイルの入出力やデータ操作、様々な解析用ウィンドウを表示することができます。

メニューバーの下の領域をツールバー言い、ボタンをクリックすることでデータ操作のダイアログを表示することができます。

データ表示域には読み込ませたデータが表示されます。データは同時に3つまで読み込むことができ、表示されているデータがアクティブデータとなります。データ操作や解析実行時はデフォルトでアクティブデータが対象になります。

The screenshot shows the StaatApp interface. At the top is a menu bar (メニューバー) with options: ファイル, データ操作, 基本統計量, 仮説検定, 多変量解析, 時系列分析, グラフ, その他, 設定. Below the menu bar is a toolbar (ツールバー) with icons for: 戻る, ソート, 削除, 置換, 欠測値, データ型, 結合, and フィルター. The main area is a data table (データ表示域) with columns: データ1 (副業有無), データ2 (収入), データ3 (性別), 年齢, 身長, and 体重. The table contains 14 rows of data. Annotations point to: 行番号 (row number) on the left, 列名 (変数名) (column name) on the right, and レコード (record) pointing to a row.

行番号	データ1 (副業有無)	データ2 (収入)	データ3 (性別)	年齢	身長	体重
1	あり	580	男性	32	170	60
2	なし	430	女性	28	156	43
3	あり	800	男性	45	163	57
4	あり	780	女性	36	161	48
5	あり	690	女性	42	158	49
6	なし	510	男性	30	172	61
7	なし	740	女性	49	165	54
8	なし	350	女性	23	161	45
9	あり	620	男性	29	180	78
10	なし	500	女性	25	150	42
11	なし	430	男性	33	168	69
12	あり	590	女性	36	159	47
13	あり	1200	男性	51	169	65
14	なし	810	男性	53	174	70

データ表示域の最上段には列名（変数名）が表示され、左側には行番号（インデックス）が表示されます。1行分のデータをレコードと言い、サンプルサイズ≒レコード数=行数となります。

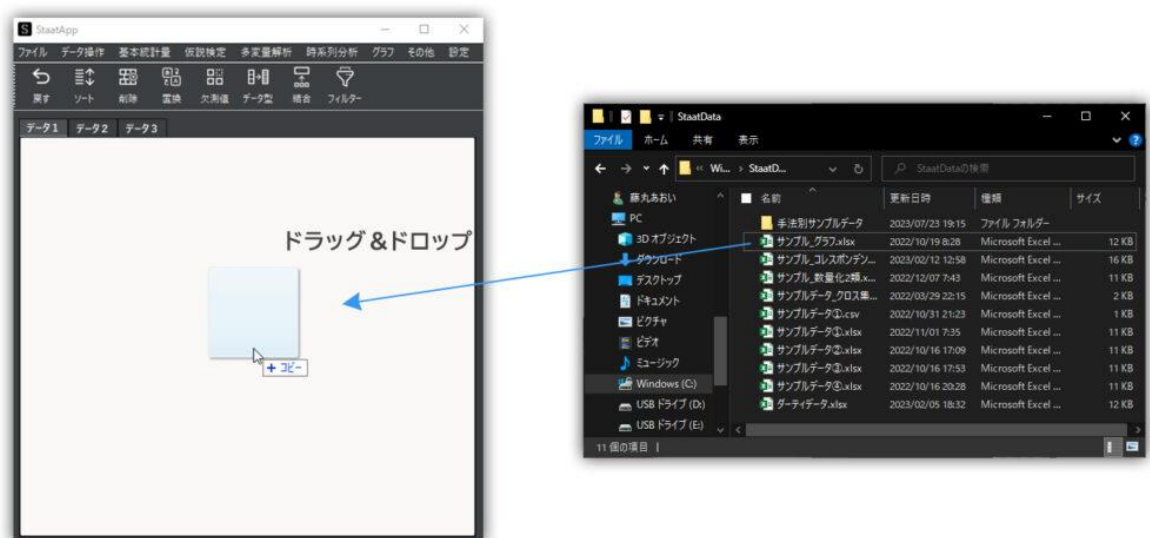
基本操作

1. データを読み込む

StaatApp は3つの方法でデータを読み込ませることができます。

① ファイルをドラッグ&ドロップする (CSV・Excelファイル限定)

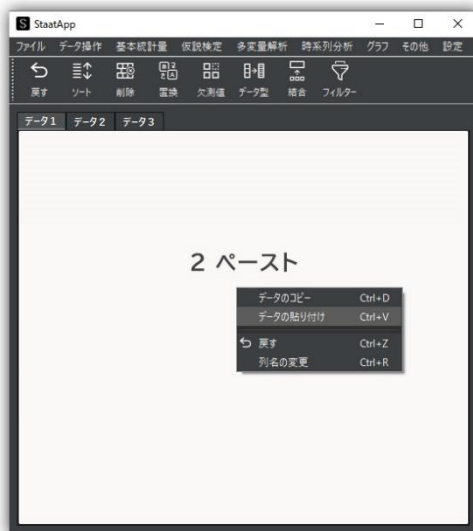
お使いのPCにあるファイルを StaatApp 上にドラッグ&ドロップすることで、データが StaatApp に取り込まれます。



② コピー&ペーストする

特定範囲のデータを読み込みたい場合は、表計算ソフト上で範囲選択後コピーを行い、StaatApp 上に貼り付けることで取り込むことができます。

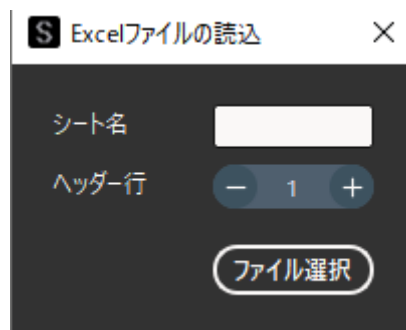
StaatApp では右クリックメニューもしくは、キーボードのコントロールキーと"V"を同時に押すことでクリップボードのデータを貼り付けることができます。



③ ファイルを選択して読み込む

ファイルに選択して読み込む場合は、メニューバーから「ファイル」→「CSVの読み込み」もしくは「Excelの読み込み」を選択します。

「Excelの読み込み」を選択した場合は、以下のダイアログが表示されるのでExcelファイルのシート名とヘッダー行を指定して、ファイルを選択します。



ヘッダー行に指定した行が列名に設定されます。特定のシート、特定の行から読み込むことができますが、基本的には②の方法で読み込む方が簡単です。

StaatApp で扱うデータ

StaatApp で扱うデータのデータ構造は、整然データとする必要があります。整然データについては以下のページで紹介しています。

[▷ 整然データとは](#)

StaatApp では列名（変数名）を用いて、分析対象のデータを選択するため扱うデータの1列目には列名を入力する必要があります。また、1つのデータに同じ列名がある場合、解析実行時に不具合が発生する可能性があるため、列名は1つのデータで一意となるようにしてください。

2. データを複製する

SaatApp内でデータを複製する場合は、右クリックメニューもしくはショートカットキーを用いて行います。アクティブデータのコピーは、キーボードのコントロールキーと"D"を同時にクリックすることでも行うことができます。



3. データを保存する

アクティブデータはメニューバーの「ファイル」→「CSV形式で保存」を選択することで任意のファイル名でCSVファイルとして保存することができます。



4. 解析用ウィンドウを表示する

読み込ませたデータに対して統計解析を行うためには、解析用ウィンドウを表示します。解析用ウィンドウはメニューバーから選択して表示することができます。

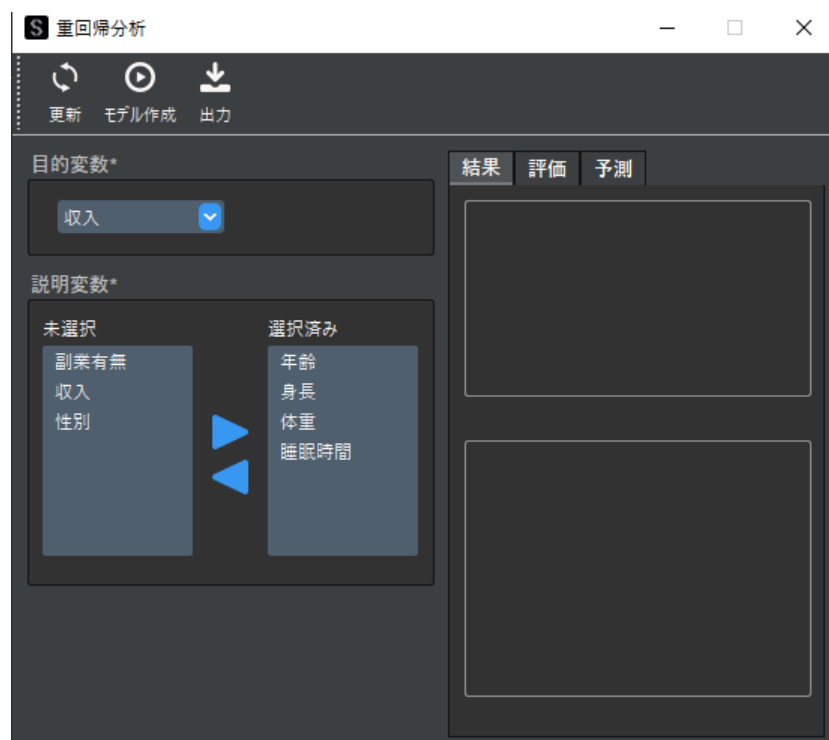
※ 有料版機能を表示するためには、ライセンス購入後に SaatApp から認証を行う必要があります。



5. 解析を実行する

解析を実行するためには、解析用ウィンドウの設定項目に変数名などを設定する必要があります。設定項目は解析手法によって異なりますが、全ての解析手法共通で分析対象の変数（列）を選択する必要があります。

以下は重回帰分析用ウィンドウの設定例です。



設定が完了したら再生ボタン（重回帰分析の場合は「モデル作成」ボタン）をクリックすると、解析が実行されます。基本的には1秒以内に解析結果が表示されますが、データ量や解析手法、使用するPCによっては数秒以上計算に要する場合があります。

6. 解析結果をポップアウトする

解析結果を別ウィンドウで表示する場合は、解析結果表示域の下にある「ポップアウト」ボタンをクリックします。



7. 解析結果をコピーする

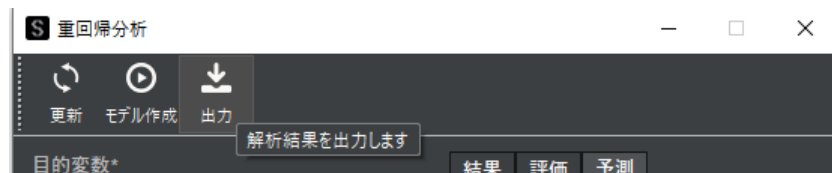
解析結果表示域の下にある「クリップボードにコピー」ボタンをクリックすると、対象の解析結果がクリップボードにコピーされます。

コピーした解析結果は、StaatAppのデータ表示域に貼り付けたり、Excelなどの表計算ソフトに貼り付けることができます。

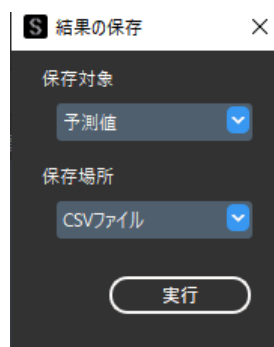
8. 解析結果を保存する

一部の解析結果は CSV ファイル形式やデータ表示域に保存することができます。

解析実行後に「出力」ボタンをクリックすると、出力用ダイアログが表示されます。



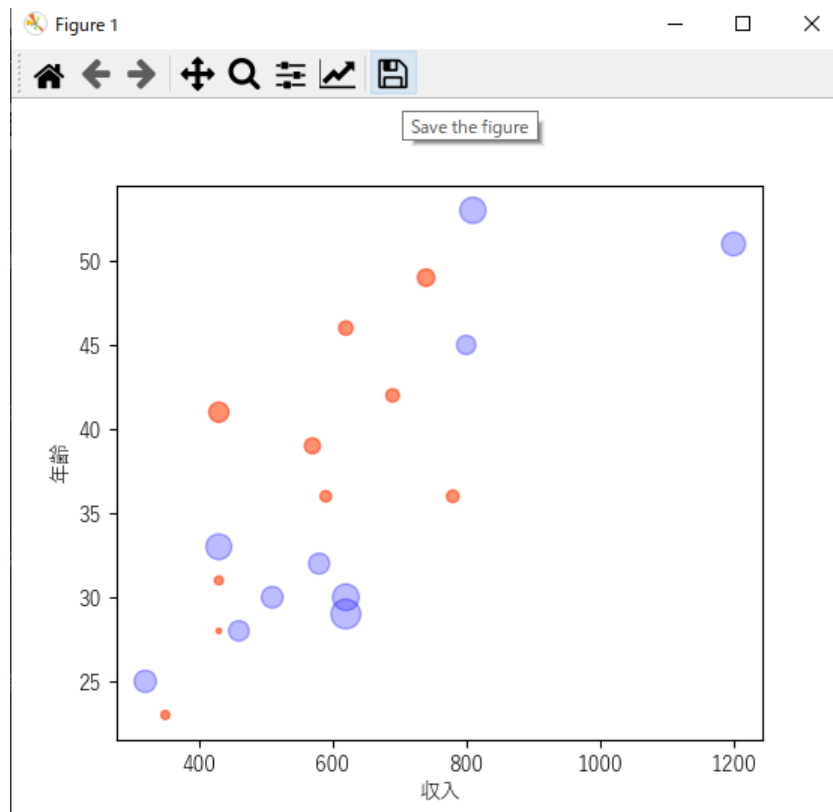
保存対象の解析結果と、保存場所を選択して「実行」ボタンをクリックすると、解析結果が保存されます。



9. グラフを画像形式で保存する

グラフ作成機能や一部の解析手法では、グラフ作成時・解析実行時に以下のようなグラフ表示ウィンドウが表示されます。

表示されたグラフは「画像保存」ボタンをクリックすることで、任意のファイル名で png 形式として保存することができます。



10. 表示桁数を変更する

デフォルトでは小数点以下の表示桁数は5桁に設定されているため、小数点第6位の値を四捨五入して表示されています。

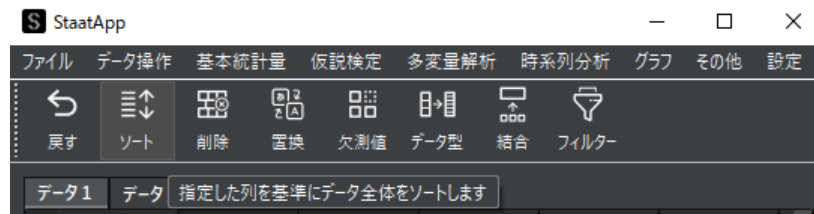
メニューバーの「設定」→「桁数の設定」で表示する小数点以下の桁数を変更することができます。



データ操作

1. 複数列を対象にソート（並び替え）を行う

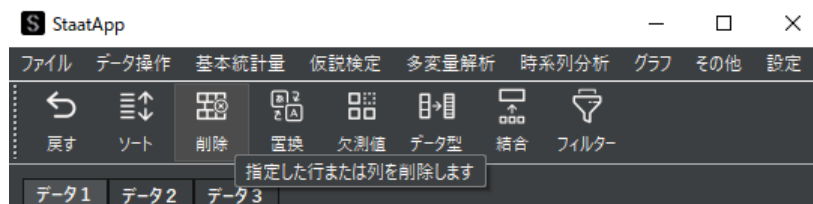
データ表示領域の列名部分をクリックすることでソートすることも可能ですが、数値データや複数列を基準として並び替えたい場合はツールバーの「ソート」機能を用います。



以下のようにデータを並び替えると、"性別"の列を見ると表記ゆれがあることがわかります。ソート機能では表記ゆれを簡単に発見できることも大きなメリットです。

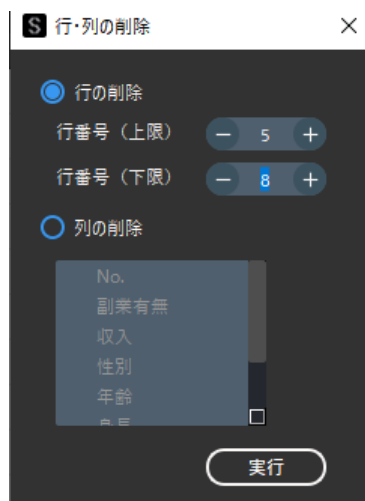
2. 不要な行または列を削除する

特定の行または列を削除する場合は「削除」機能を使います。

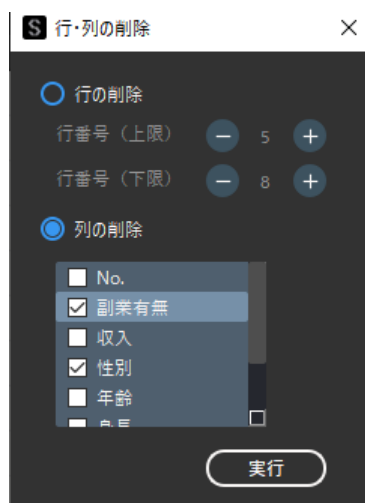


特定の行を削除したい場合は削除したい行番号を「行番号（上限）」に入力します。複数の行を一括で削除したい場合は、「行番号（上限）」に上側の行番号を、「行番号（下限）」に下側の行番号を入力します。

以下の設定では5行目から8行目が一括で削除されます。

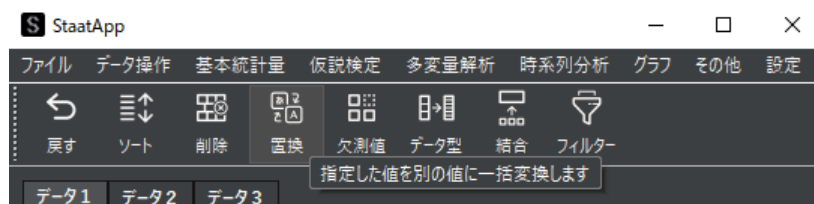


不要な列を削除した場合は、「列の削除」を選択して、削除したい列名にチェックを入れます。



3. 任意の文字列・数値を置換する

特定の文字列から特定の文字列に値を変換したい場合は、「置換」機能を用います。



置換用ダイアログから置換前後の文字列を入力します。オプションの「完全に一致」ではセル内の文字列が完全に一致した場合のみ置換します。「半角英数字に変換」では全角英数字を自動で半角英数字に変換します。

例では”性別”列の表記ゆれを修正するために、”女”→”女性”に置換します。



ダミー変数の作成

文字データからダミー変数を作成する場合は、「ワンホットエンコーディング」を選択して、変換対象の列を選択します。変換対象の列のカテゴリ数だけ新しい列が追加されます。

4. 欠測値を削除・補完する

StaatApp では空白データなどの欠測値を読み込ませた場合は，“nan”と表示されます。一括して“nan”の処理を行う場合は「欠測値」機能を用います。



表示されたダイアログでは、画面左側で変数ごとの欠測値の数を確認することができます。



”年齢”列では欠測値が3つあることがわかります。欠測値の処理方法として以下の3つの方法が可能です。

- リストワイズ除去
- ペアワイズ除去
- 補完

以下の設定では各変数の平均値で補完を行います。



欠測値が平均値で補完されています。数値データでない”副業有無”や”性別”列は平均値を求めることができないため、”nan”のままとなります。

カテゴリデータの欠測値に対しては、直接入力して値を変更もしくは除去することで対応します。

5. データ型を変換する

StaatApp では「数値型」「文字列型」の2つのデータ型が存在します。

数値型：半角数字で入力された値。演算など順序尺度以上の統計解析が可能。

文字列型：半角英字・全角文字で入力された値。クロス集計表などのカテゴリ変数に対する統計解析が可能。

▷ 統計学におけるデータの種類

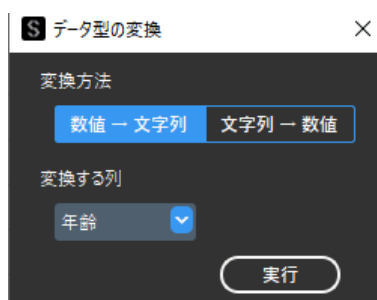
データ表示部分で値が左寄せされている列のデータ型は「文字列型」になります。逆に値が右寄せされている列は「数値型」になります。StaatApp ではデータが読み込まれた時点で自動で判定されます。

「数値型」から「文字列型」もしくは、半角数字だけが入力されている列を「文字列型」から「数値型」に変換することは可能です。

ツールバーの「データ型」機能を用います。



以下は”年齢”の列を「数値型」→「文字列型」に変換する例です。



※ 解析を実行した際に、データ型に関するエラーが発生する場合はこの機能を用いて変換してください。

6. 2つのデータを結合する

2つのデータを結合したい場合は、「結合」機能を使います。



データ1とデータ2を結合したい場合は、以下のように設定します。



デフォルトではデータは縦方向に結合されます。同じ列名を持つデータを結合したい場合、サンプルサイズを増やしたい場合に用います。

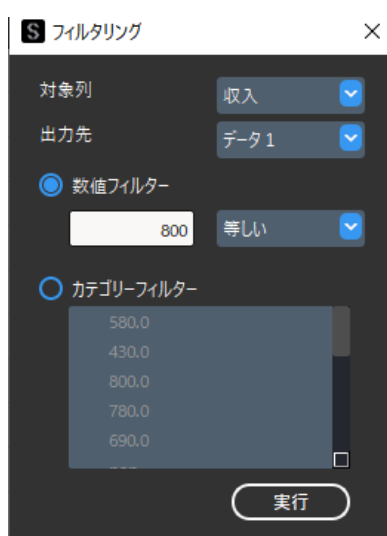
オプションの「横方向に結合」を選択すると、同じ行同士でデータが結合されます。変数を加える場合や、解析後に出力した予測結果を元のデータに結合したい場合に用います。

7. 特定の条件でデータを抽出する

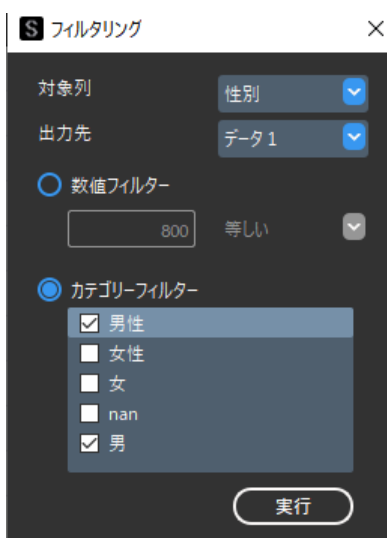
特定の条件に一致するデータのみを抽出したい場合は、「フィルター」機能を使います。



以下の例では"収入"列で、800 より小さいデータのみを抽出します。



特定の 카테고리（文字列）に一致するデータを抽出したい場合は、以下のように「カテゴリフィルター」を選択して、抽出したいカテゴリにチェックを入れます。



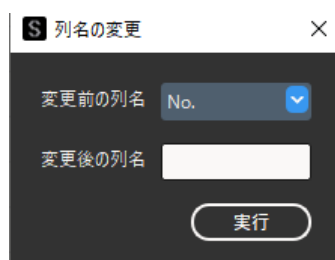
8. メニューバーのデータ操作機能の選択

メニューバーに設定されているデータ操作機能を用いる場合は、以下のように「データ操作」を選択して任意の機能を選択します。



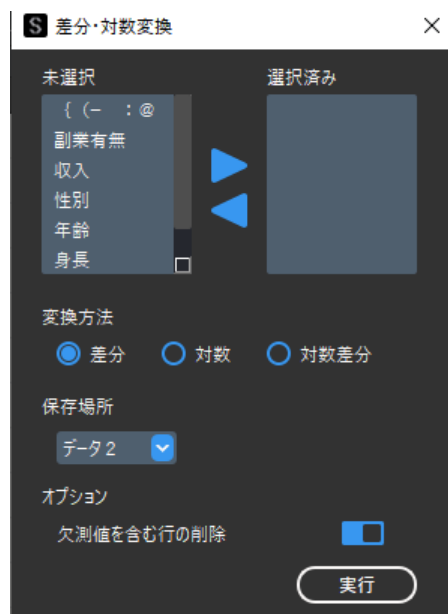
9. 列名（変数名）の変更

列名を変更する場合は、「列名の変更」機能を用います。「列名の変更」は「Ctrl+R」のショートカットキーでも実行可能です。



10. 差分・対数変換

特定の変数に対して、差分変換もしくは対数変換を行いたい場合は、「差分・対数変換」機能を用います。



この機能は時系列分析を行う場合や、回帰分析を行う場合に有効です。

11. テストデータの作成

既存のデータからテストデータを作成したい場合は、「テストデータの作成」機能を用います。



データの末尾から指定したサイズ分だけデータが分割されます。

オプションの「シャッフルして分割」を選択すると、データの末尾でなくランダムでレコードを選択して分割を行います。

解析の実行・結果の表示

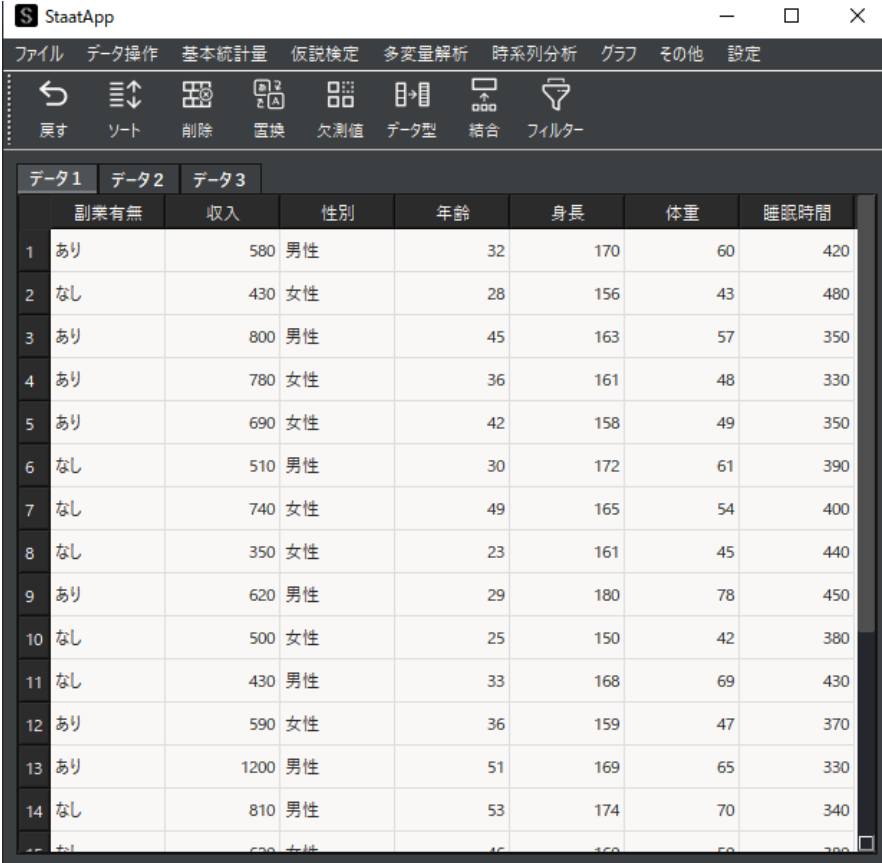
各解析手法の実行方法・結果の表示について、説明します。

本書に記載されていない解析手法の実行方法については、以下の解析手法一覧ページからご覧ください。

[▷ 解析手法一覧](#)

1. 相関係数

以下のサンプルデータを用いて解説します。

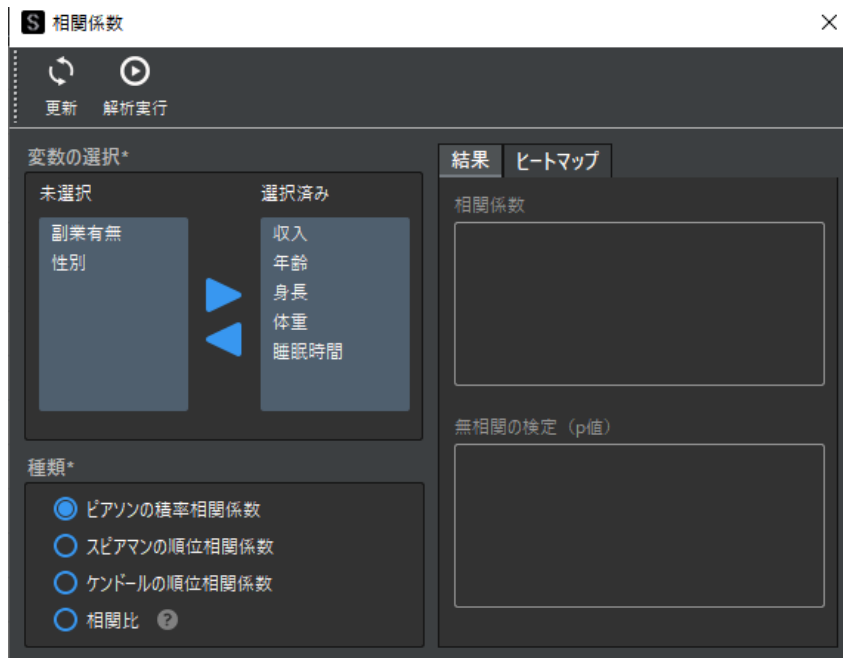


	データ1	データ2	データ3				
	副業有無	収入	性別	年齢	身長	体重	睡眠時間
1	あり	580	男性	32	170	60	420
2	なし	430	女性	28	156	43	480
3	あり	800	男性	45	163	57	350
4	あり	780	女性	36	161	48	330
5	あり	690	女性	42	158	49	350
6	なし	510	男性	30	172	61	390
7	なし	740	女性	49	165	54	400
8	なし	350	女性	23	161	45	440
9	あり	620	男性	29	180	78	450
10	なし	500	女性	25	150	42	380
11	なし	430	男性	33	168	69	430
12	あり	590	女性	36	159	47	370
13	あり	1200	男性	51	169	65	330
14	なし	810	男性	53	174	70	340

① 相関係数と無相関の検定結果の算出

メニューバーから「基本統計量」 → 「相関係数」を選択して、相関係数用ウィンドウを表示します。

「変数の選択」で相関係数の算出対象とする変数を全て選択します。



ピアソンの積率相関係数以外を算出したい場合は、「種類」から算出したい相関係数もしくは相関比を選択します。

※ 相関比を算出する場合は、1つ目の変数に文字型データ、2つ目の変数に数値データを選択してください。

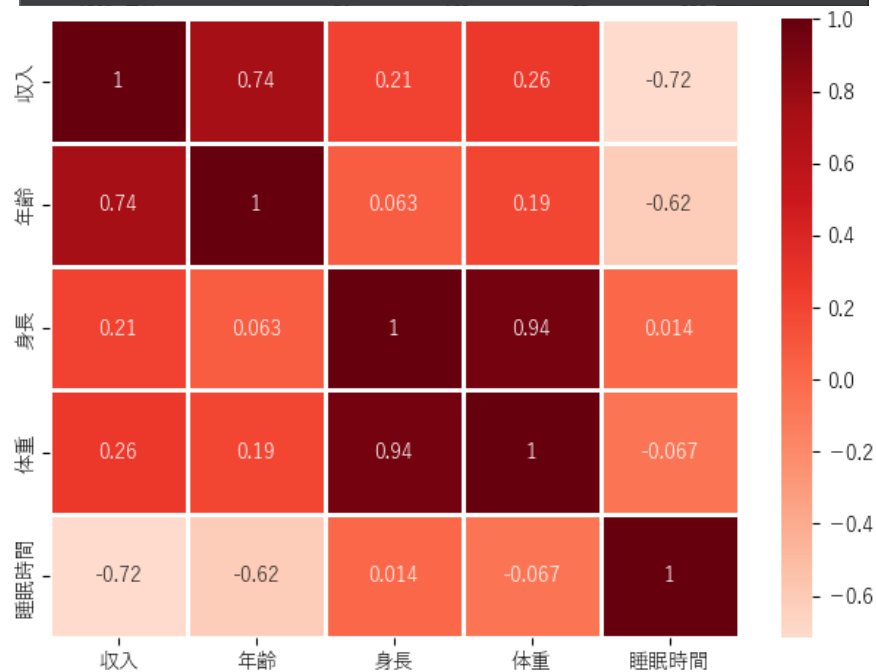
「解析実行」ボタンをクリックすると以下のように、相関行列と無相関の検定の検定結果行列が表示されます。

	収入	年齢	身長	体重	睡眠時間
収入	1.0	nan	nan	nan	nan
年齢	0.74233	1.0	nan	nan	nan
身長	0.21352	0.06346	1.0	nan	nan
体重	0.2633	0.18992	0.93823	1.0	nan
睡眠時間	-0.71611	-0.61619	0.01436	-0.06651	1.0

	収入	年齢	身長	体重	睡眠時間
収入	0.0	nan	nan	nan	nan
年齢	0.00012	0.0	nan	nan	nan
身長	0.35272	0.78465	0.0	nan	nan
体重	0.24883	0.40962	0.0	0.0	nan
睡眠時間	0.00026	0.00294	0.95075	0.77452	0.0

② ヒートマップの作成

相関係数を算出後にヒートマップタブを選択し、「描画」ボタンをクリックすることで以下のよう
な相関の強さを示すヒートマップを作成することができます。



2. 対応のない場合の仮説検定

対応のない場合の検定方法について解説します。対応のない場合とは、比較する標本（サンプル）を別の個体から得た場合のデータになります。

分析例として、社会人属性や収入などを示すデータから副業有無によって収入に差があるかを、マンホイットニーの U 検定を用いて判定します。

帰無仮説は「副業有無によって収入に差がない」と設定して、対立仮説は「副業有無によって収入に差がある」となります。

分析で使用するサンプルデータは以下のようになります。

データ1	データ2	データ3				
副業有無	収入	性別	年齢	身長	体重	
あり	580	男性	32	170	60	
なし	430	女性	28	156	43	1行 = 1個体
あり	800	男性	45	183	77	
あり	780	女性	36	161	48	
あり	690	女性	42	158	49	
なし	510	男性	30	172	61	
なし	740	女性	37	165	54	比較する群を示す属性とデータを縦方向に並べる
なし	350	男性	35	161	45	
あり	620	男性	29	180	78	
なし	500	女性	25	150	42	
なし	430	男性	33	168	69	
あり	590	女性	36	159	47	
あり	1200	男性	51	169	65	
なし	810	男性	53	174	70	

このような形式のデータをロングデータ（縦持ちデータ）と言います。ロングデータでは1行＝1個体のデータとして、所属する群（副業有無）を示す列と、比較対象のデータ（収入）が入力された列が必要です。

マンホイットニーの U 検定を行うために、マンホイットニーの U 検定用ウィンドウを表示します。ウィンドウが表示されたら「変数選択」の「群名列」に「副業有無」を、「群 A」に「あり」、「群 B」に「なし」、「データ列」に「収入」を選択します。



「解析実行」ボタンをクリックすると解析結果が表示されます。



検定結果は p 値 < 0.05 となるので、有意水準 $\alpha = 0.05$ において「副業有無によって収入に差がある」ということがわかります。

ここまでが対応のない検定の1つであるマンホイットニーのU検定の実行方法になります。

検定の種類

片側検定を行う場合は、以下のように設定してください。

「群A」が「群B」より小さいことを調べたい場合 → 右側

「群B」が「群A」より小さいことを調べたい場合 → 左側

StaatApp では対応のない場合の検定である一元配置分散分析や二元配置分散分析も、ロングデータ形式に対して実行することができます。

3. 対応のある場合の仮説検定

対応のある場合の検定方法について解説します。対応のある場合とは、比較する標本（サンプル）を同一の個体から得た場合のデータになります。

分析例として、学生のテストの得点を示すデータから受験時期によって得点に差があるかを、フリードマン検定を用いて判定します。

帰無仮説は「受験時期によって得点に差がない」と設定して、対立仮説は「受験時期によって得点に差がある」となります。

分析で使用するサンプルデータは以下のようになります。

	データ1	データ2	データ3	後期
1	67	62	52	
2	94	80	75	
3	68	55	53	
4	82	71	87	
5	80	53	64	
6	55	71	55	
7	89	63	68	
8	60	50	65	
9	73	61	53	
10	43	30	42	

このようなデータ形式をワイドデータ（横持ちデータ）と言います。ワイドデータでは1行=1個体のデータとして、比較対象のデータごとに入力された列が必要です。

フリードマン検定を行うために、フリードマン検定用ウィンドウを表示します。フリードマン検定用ウィンドウが表示されたら、「変数列選択」で検定対象のデータが入力された列を選択します。



「解析実行」ボタンをクリックすると解析結果が表示されます。



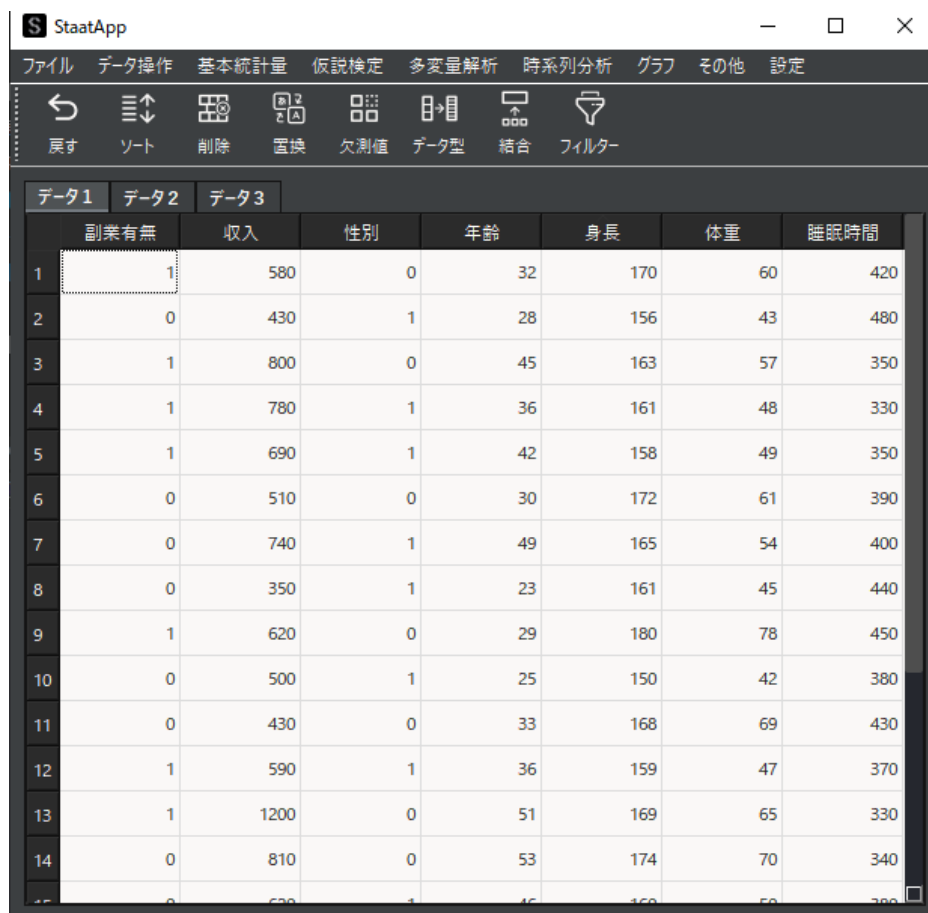
検定結果は p 値 < 0.05 となるので、有意水準 $\alpha = 0.05$ において「受験時期によってテストの得点に差がある」ということがわかります。

ここまでが対応のある検定の1つであるフリードマン検定の実行方法になります。

4. ロジスティック回帰分析（無料）

※重回帰分析・プロビット分析・多項ロジスティック回帰・判別分析についても同様の操作方法で実行可能です。

以下のサンプルデータを用いて解説します。

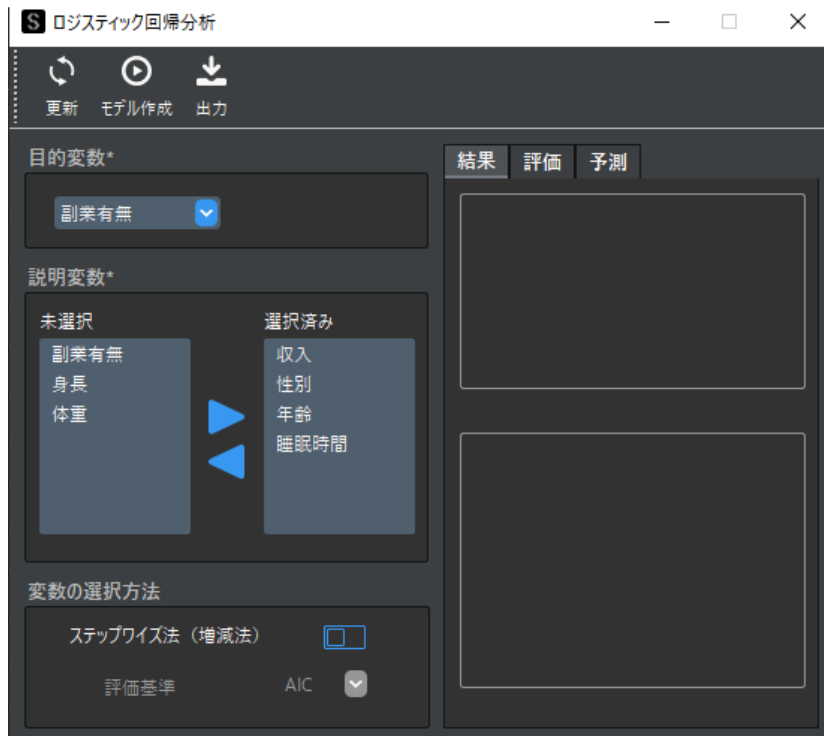


	データ1	データ2	データ3				
	副業有無	収入	性別	年齢	身長	体重	睡眠時間
1	1	580	0	32	170	60	420
2	0	430	1	28	156	43	480
3	1	800	0	45	163	57	350
4	1	780	1	36	161	48	330
5	1	690	1	42	158	49	350
6	0	510	0	30	172	61	390
7	0	740	1	49	165	54	400
8	0	350	1	23	161	45	440
9	1	620	0	29	180	78	450
10	0	500	1	25	150	42	380
11	0	430	0	33	168	69	430
12	1	590	1	36	159	47	370
13	1	1200	0	51	169	65	330
14	0	810	0	53	174	70	340

① モデルの作成

メニューバーから「多変量解析」→「回帰」→「ロジスティック回帰分析」を選択してロジスティック回帰分析用ウィンドウを表示します。

ロジスティック回帰分析用ウィンドウが表示されたら、目的変数と説明変数の設定を行います。ダミー変数を説明変数に選択する際は、多重共線性の問題を回避するために”1列分”を除いて選択します。



設定項目に入力が完了したら、ツールバーの「モデル作成」ボタンをクリックします。画面右側の「解析結果」にロジスティック回帰分析の結果が出力されます。

※ 説明変数に対してサンプルサイズが小さすぎると、モデルが作成できず結果は出力されません。



ステップワイズ法

「変数の選択方法」でステップワイズ法（増減法）を選択してモデルを作成すると、選択済み説明変数から増減法を用いて、最適な説明変数を自動で選択することが可能です。

モデルの評価基準は AIC（赤池情報量基準）もしくは BIC（ベイズ情報量基準）から選択可能で、増減法によって評価基準が最小となる説明変数が選択されます。

② 予測精度の評価（応用）

作成したモデルに対して、予測精度の評価を行う場合は「評価」タブで行います。事前にテストデータをモデル作成とは別のデータに入力しておき、「テストデータの選択」で対象のデータを選択します。

「算出」ボタンをクリックすると以下のように各評価指標と混合行列が表示されます。

※ テストデータには説明変数で選択した変数名（列名）を含む必要があります。

ロジスティック回帰分析

更新 モデル作成 出力

目的変数*
副業有無

説明変数*

未選択
副業有無
身長
体重

選択済み
収入
性別
年齢
睡眠時間

結果 評価 予測

テストデータの選択 データ2
算出

指標値

統計量	
正解率	0.85714
適合率	0.875
再現率	0.77778
F値	0.82353

混合行列

		予測結果	
		1	0
実測値	1	11	1
	0	2	7

③ 予測値の算出（応用）

作成したモデルを用いて予測値の算出を行います。「予測」タブから「予測用データの選択」を行い「算出」ボタンをクリックします。

※ 予測用データには説明変数で選択した変数名（列名）を含む必要があります。



	判定結果_副業有無	予測確率_副業有無
1	0	0.43261
2	0	0.03166
3	1	0.92853
4	1	0.99878
5	1	0.79926
6	0	0.27613
7	0	0.30617
8	0	0.04476
9	1	0.81118

予測用データに対して、判定結果と予測確率が算出されます。

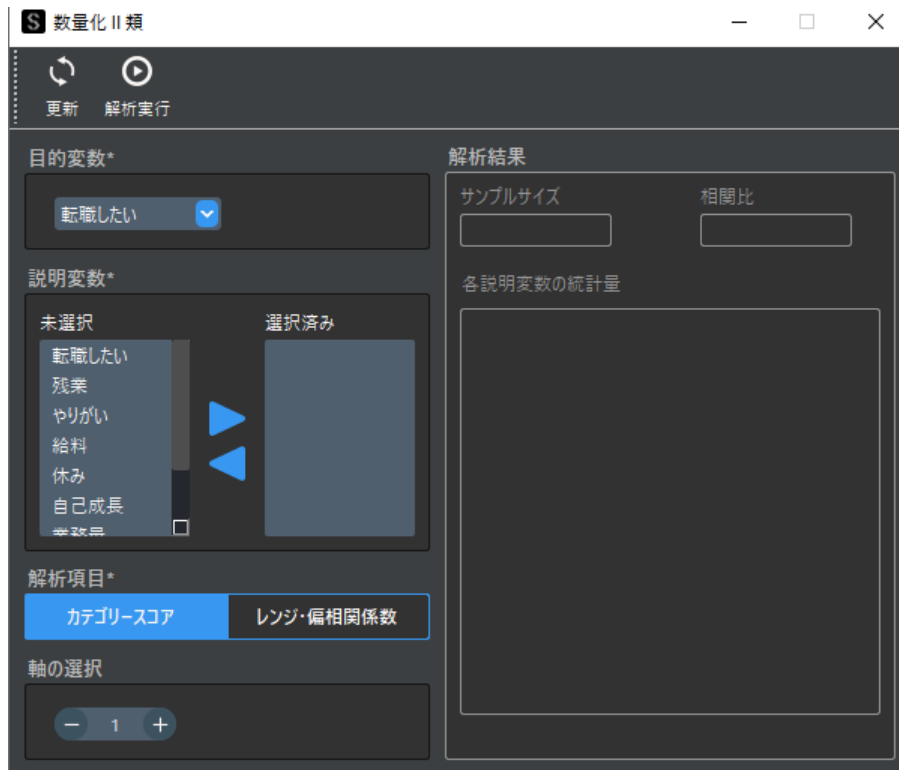
5. 数量化Ⅱ類

以下のサンプルデータを用いて解説します。サンプルデータは数値データですが、数量化Ⅱ類は文字列データに対しても実行可能です。



	データ1	データ2	データ3							
	転職したい	残業	やりがい	給料	休み	自己成長	業務量	仕事内容	人間関係	
1	4	2	4	4	1	3	2	2	1	
2	3	5	2	1	3	1	4	2	1	
3	5	2	3	4	3	4	2	4	5	
4	2	2	2	2	2	3	2	2	2	
5	5	5	4	3	4	4	5	4	3	
6	4	1	4	4	2	4	3	2	3	
7	1	3	1	1	1	1	2	1	1	
8	4	5	5	3	5	4	5	3	5	
9	1	1	3	5	1	3	1	3	1	
10	5	4	4	3	1	4	4	2	1	
11	2	3	3	3	2	4	4	4	2	
12	2	1	2	1	3	1	1	1	1	
13	3	5	2	3	5	2	5	1	3	
14	1	4	1	1	3	1	2	2	1	

メニューバーから「多変量解析」→「数量化Ⅱ類」を選択します。数量化Ⅱ類用ウィンドウが表示されたら、目的変数を選択します。例では"転職希望度"に影響を与える仕事の悩みを調べるために、目的変数は"転職したい"を選択します。



説明変数は、「未選択」から数量化II類の対象とする変数名をクリックして選択します。

数量化II類では固有値の数（目的変数のカテゴリー数 - 1）だけ解を求めることができます。一般的には1つ目の軸で結果を判断するため、特に意識しない方は「軸の選択」はデフォルト設定もままで問題ありません。

設定が完了したら、ツールバーの「解析実行」ボタンをクリックします。解析項目で「カテゴリースコア」を選択した場合は、説明変数に設定した変数のカテゴリーごとのカテゴリースコアが算出されます。

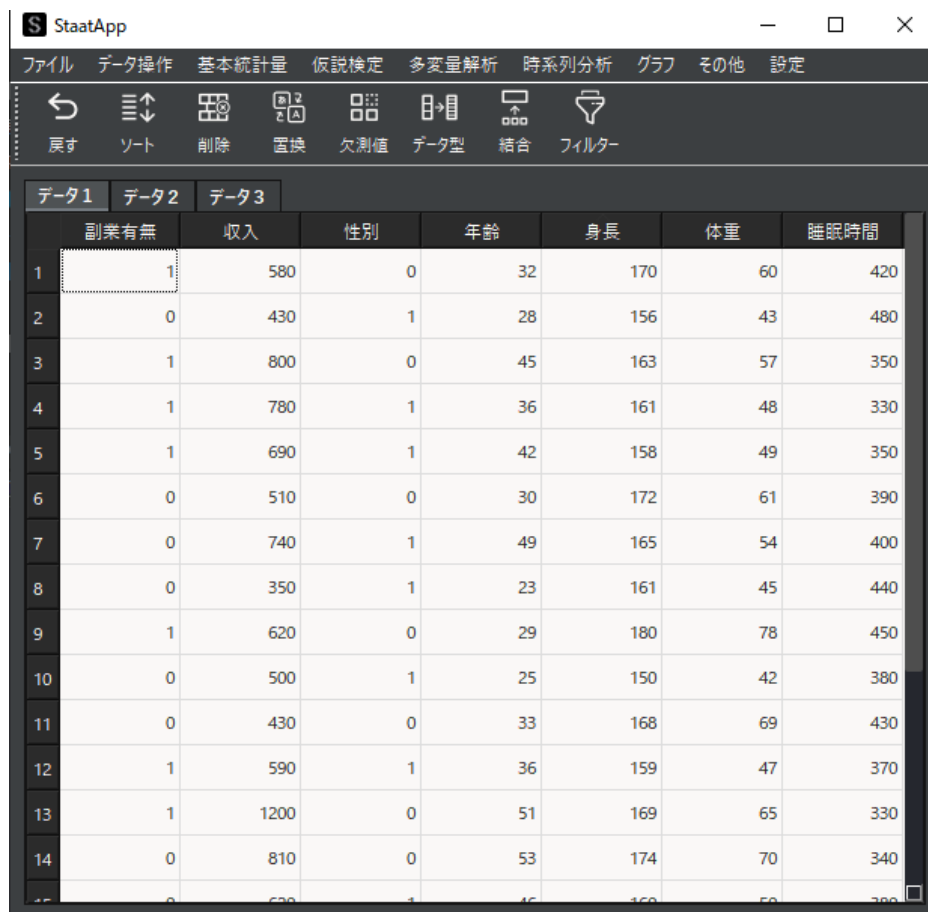


解析項目で「レンジ・偏相関係数」を選択した場合は、説明変数に設定した変数ごとのレンジと目的変数に対する偏相関係数が算出されます。



6. 主成分分析

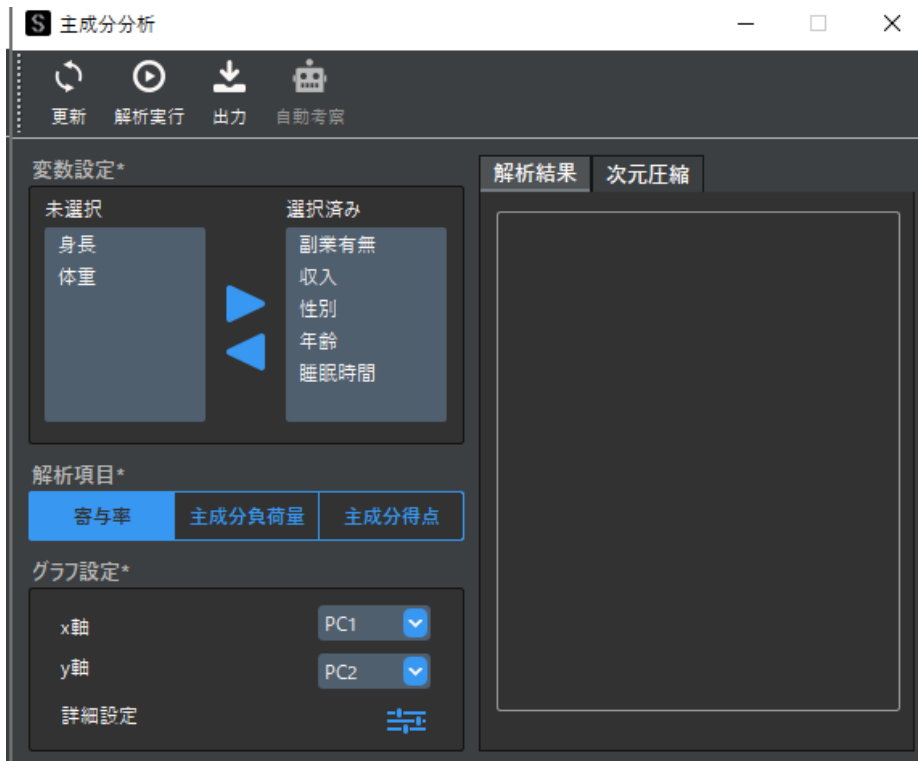
以下のサンプルデータを用いて解説します。



	データ1	データ2	データ3					
		副業有無	収入	性別	年齢	身長	体重	睡眠時間
1		1	580	0	32	170	60	420
2		0	430	1	28	156	43	480
3		1	800	0	45	163	57	350
4		1	780	1	36	161	48	330
5		1	690	1	42	158	49	350
6		0	510	0	30	172	61	390
7		0	740	1	49	165	54	400
8		0	350	1	23	161	45	440
9		1	620	0	29	180	78	450
10		0	500	1	25	150	42	380
11		0	430	0	33	168	69	430
12		1	590	1	36	159	47	370
13		1	1200	0	51	169	65	330
14		0	810	0	53	174	70	340
15		0	620	1	46	160	50	380

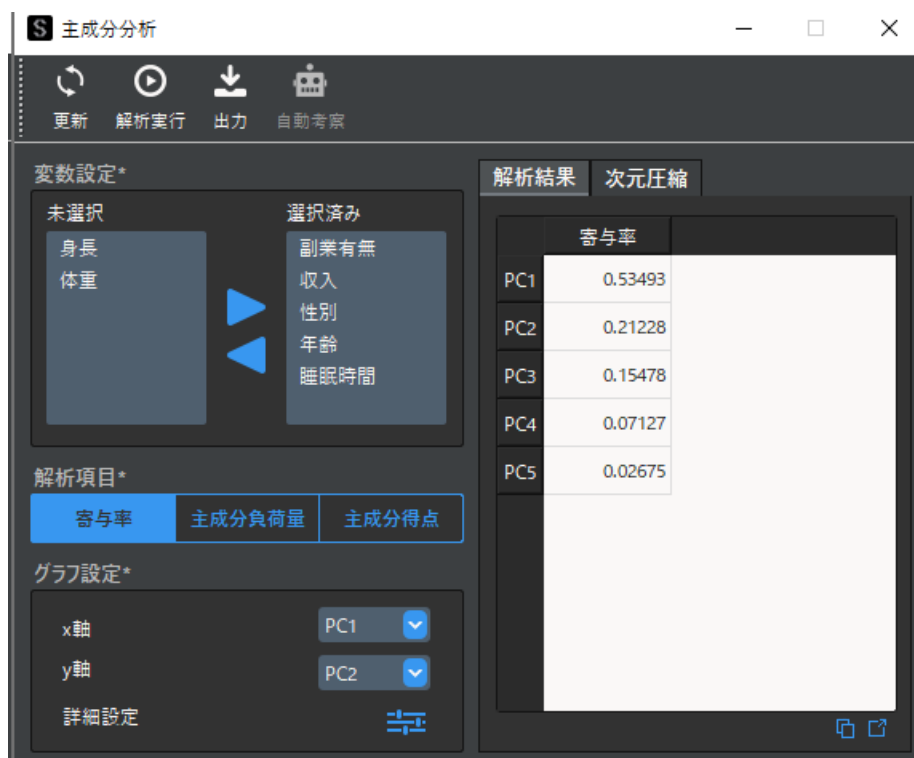
メニューバーから「多変量解析」→「主成分分析」を選択します。主成分分析用ウィンドウが表示されたら、未選択欄から主成分分析の対象とする変数名を選択します。

ダミー変数を選択する際は、多重共線性の問題を回避するために1列分を除いて選択します。サンプルデータは既に1列分除いたダミー変数になっています。



① 寄与率・累積寄与率の算出・グラフ作成

解析項目で「寄与率」を選択した状態で「解析実行」ボタンをクリックすると、主成分ごとの寄与率と累積寄与率のグラフが出力されます。

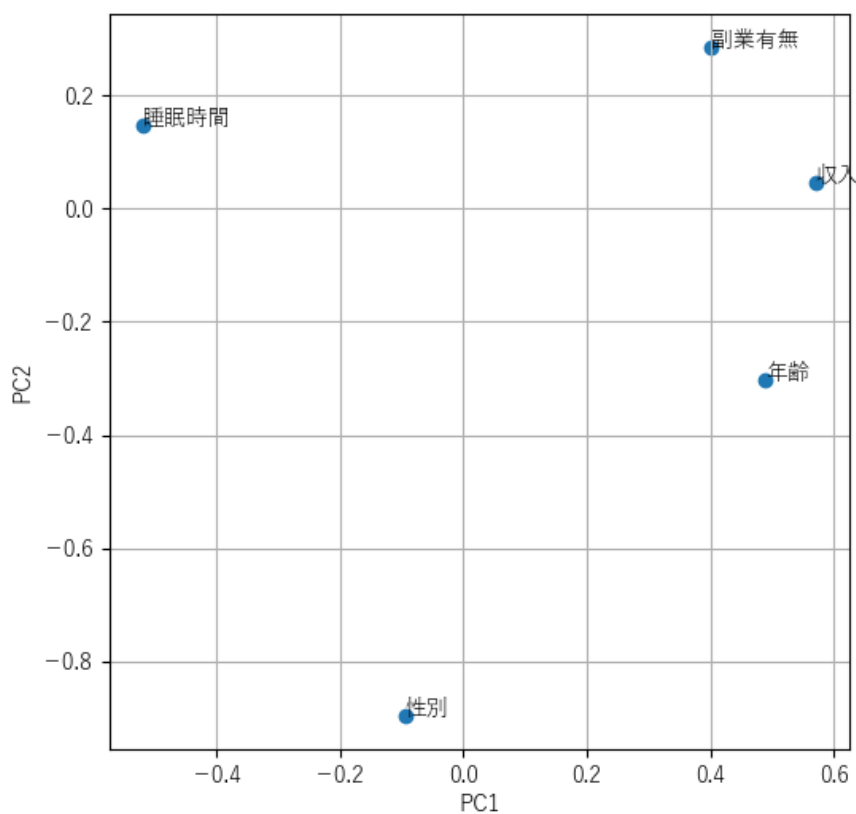


② 主成分負荷量の算出・グラフ作成

グラフ軸設定から「x 軸」と「y 軸」に設定する主成分を選択します。以下の図は「x 軸」は第一主成分 (=PC1) ，「y 軸」は第二主成分 (=PC2) を設定した例です。

解析項目で「主成分負荷量」を選択した状態で「解析実行」をクリックすると、各変数の主成分負荷量とプロットが出力されます。

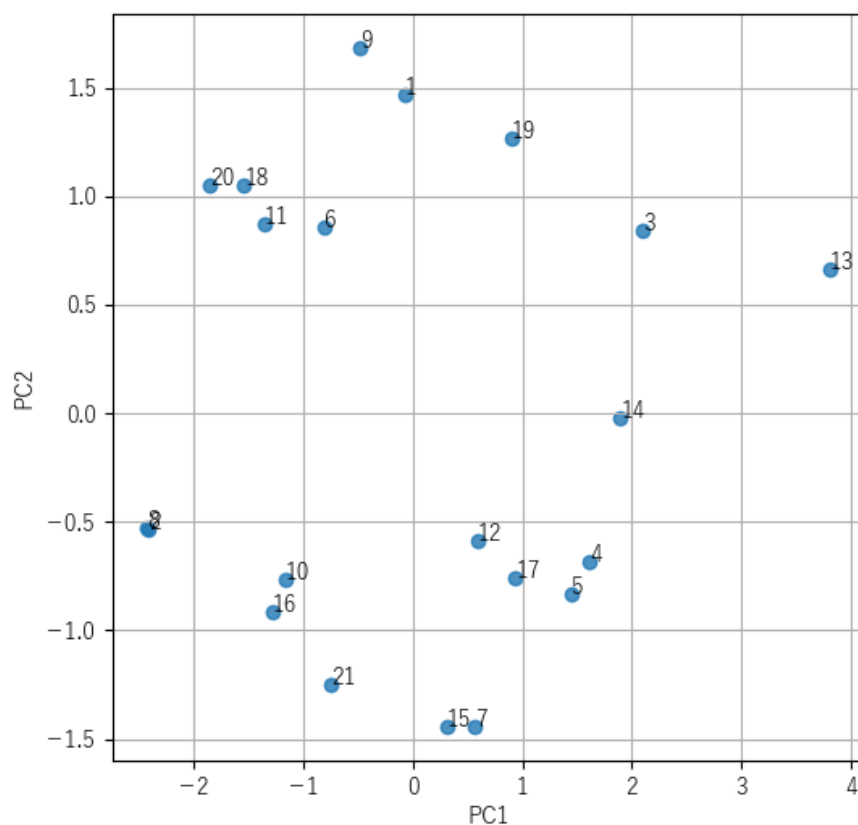
	副業有無	収入	性別	年齢	睡眠時間
PC1	0.39956	0.57029	-0.09431	0.48758	-0.51816
PC2	0.28506	0.04575	-0.89608	-0.30278	0.14836
PC3	-0.75514	0.14556	-0.38202	0.49896	0.11695
PC4	0.30113	0.33098	0.15531	0.2826	0.83414
PC5	-0.31333	0.73617	0.13446	-0.58461	-0.00597



③ 主成分得点のプロット作成

グラフ軸設定から「x 軸」と「y 軸」に設定する主成分を選択します。以下の図は「x 軸」は第一主成分 (=PC1)、「y 軸」は第二主成分 (=PC2) を設定した例です。

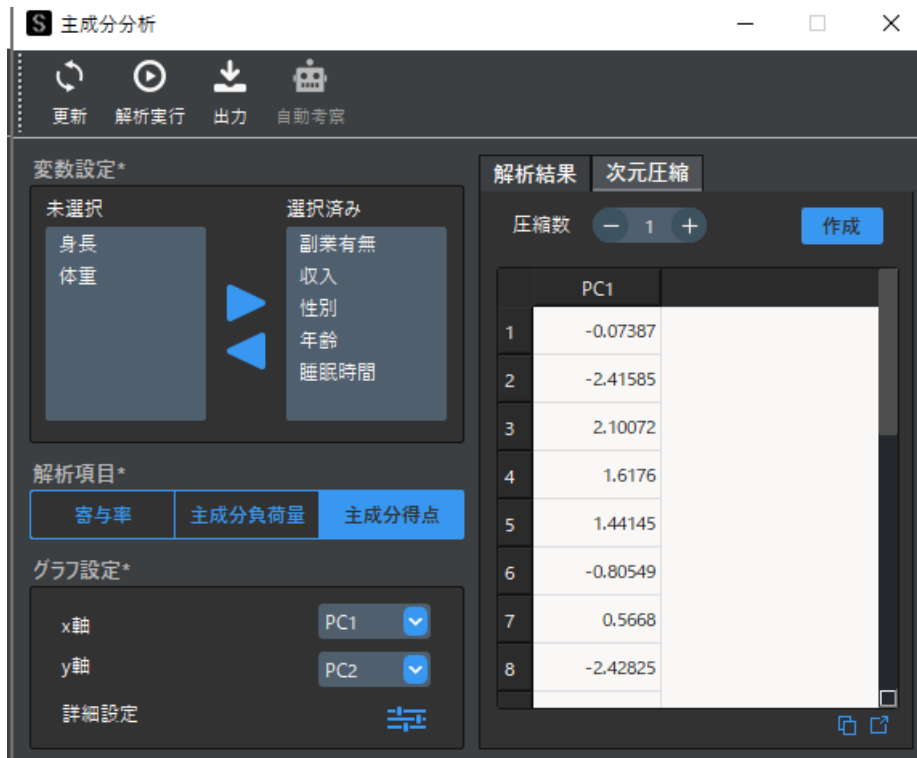
解析項目で「主成分得点」を選択した状態で「解析実行」をクリックすると、設定した主成分に対するサンプルごとのプロットが表示されます。



④ 次元圧縮（応用）

主成分分析は複数の変数を統合して、総合的な意味を持つ新しい変数を作成することができます。モデルに対する変数が減るので次元圧縮とも言い、StaatAppでも次元圧縮を行うことが可能です。

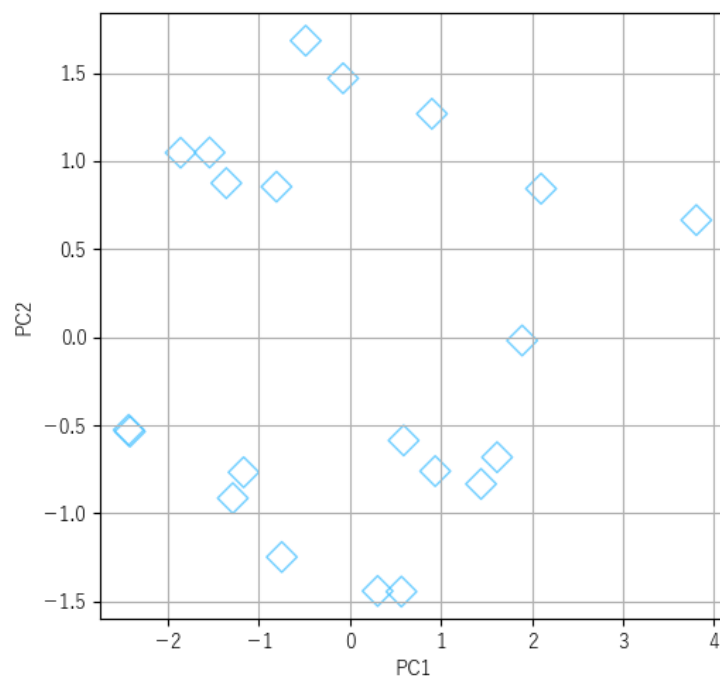
解析実行後に「次元圧縮」タブを表示して、圧縮数 (=作成する変数の数) を設定します。「作成」ボタンをクリックすると以下のように、設定した変数分だけ新しい変数が出力されます。



⑤ グラフ設定

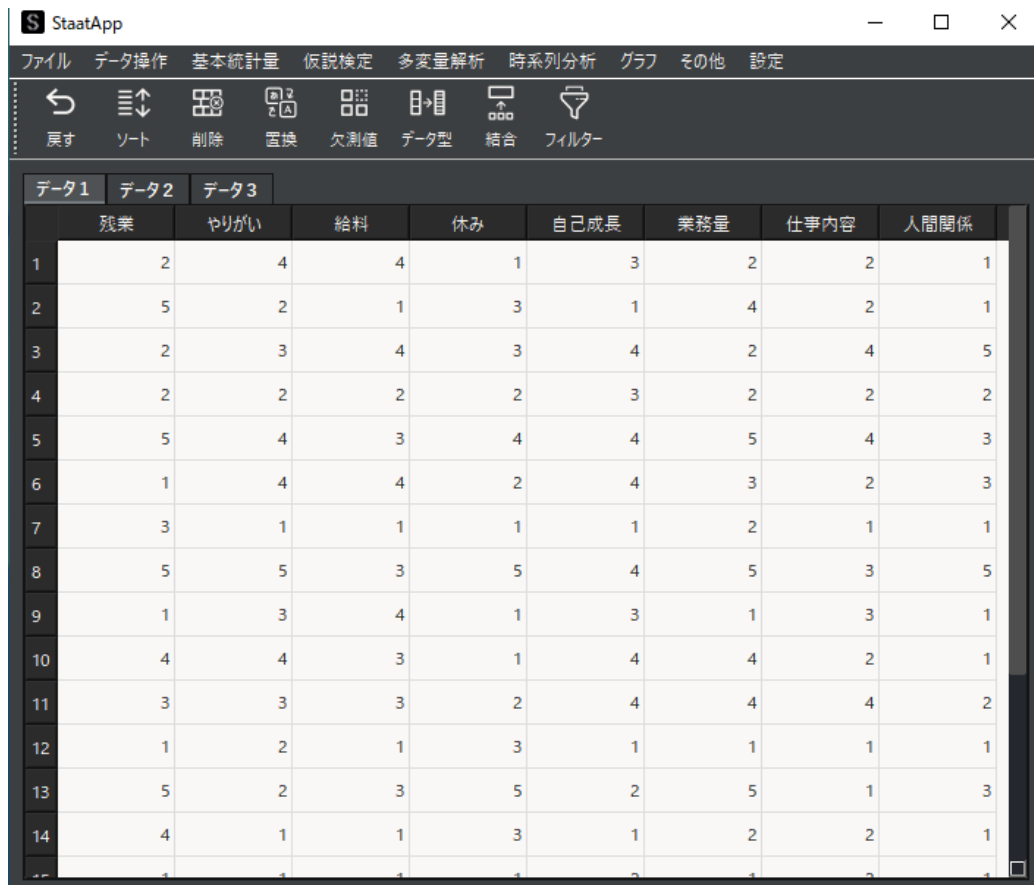
グラフ設定では「詳細設定」ボタンをクリックすると、以下のダイアログが表示されます。プロットされるマーカーのデザインや、ラベル（マーカー付近に表示される文字）を自由に変更することができます。

例えば、主成分得点については以下のようなグラフを作成することも可能です。



7. 因子分析

以下のサンプルデータを用いて解説します。



The screenshot shows the StaatApp software interface. The menu bar includes: ファイル, データ操作, 基本統計量, 仮説検定, 多変量解析, 時系列分析, グラフ, その他, 設定. The toolbar contains icons for: 戻す, ソート, 削除, 置換, 欠測値, データ型, 結合, and フィルター. Below the toolbar, there are tabs for 'データ1', 'データ2', and 'データ3'. The main data table has the following structure:

	データ1	データ2	データ3	データ4	データ5	データ6	データ7	データ8	データ9
	残業	やりがい	給料	休み	自己成長	業務量	仕事内容	人間関係	
1	2	4	4	1	3	2	2	1	
2	5	2	1	3	1	4	2	1	
3	2	3	4	3	4	2	4	5	
4	2	2	2	2	3	2	2	2	
5	5	4	3	4	4	5	4	3	
6	1	4	4	2	4	3	2	3	
7	3	1	1	1	1	2	1	1	
8	5	5	3	5	4	5	3	5	
9	1	3	4	1	3	1	3	1	
10	4	4	3	1	4	4	2	1	
11	3	3	3	2	4	4	4	2	
12	1	2	1	3	1	1	1	1	
13	5	2	3	5	2	5	1	3	
14	4	1	1	3	1	2	2	1	

メニューバーから「多変量解析」→「因子分析」を選択します。因子分析用ウィンドウが表示されたら、未選択一覧から因子分析の対象とする変数名を選択します。

変数を選択したら、選択した変数分の固有値が算出されます。固有値は因子数を決定する際に用いることがあります。



因子数・軸の回転方法・因子抽出方法の選択を行います。デフォルトでは以下の値が設定されています。

- 因子数：2
- 軸の回転方法：varimax（バリマックス回転）

① 因子負荷量・共通性の算出

解析項目で「因子負荷量・共通性」を選択した状態で「解析実行」をクリックすると、変数ごとの因子負荷量と共通性が算出されます。

「ML1」とは第一因子を意味します。

S 因子分析

更新 解析実行 自動考察

変数設定*

未選択

選択済み

残業
やりがい
給料
休み
自己成長
業務量

解析項目*

因子負荷量・共通性 寄与率・累積寄与率

固有値

	1	2	3	4
固有値	3.61634	2.11925	0.87876	0.48565

オプション

因子数 - 2 +

軸の回転方法 varimax

解析結果

	ML1	ML2	共通
残業	-0.11119	0.88498	0.7
やりがい	0.84449	0.16724	0.7
給料	0.76706	-0.08845	0.5
休み	-0.01935	0.50512	0.2
自己成長	0.94465	0.06509	0.8
業務量	0.27652	0.86941	0.8
仕事内容	0.6884	0.01902	0.4

パイプロット

x軸 ML1

y軸 ML2

グラフ設定

描画

② 寄与率・累積寄与率の算出

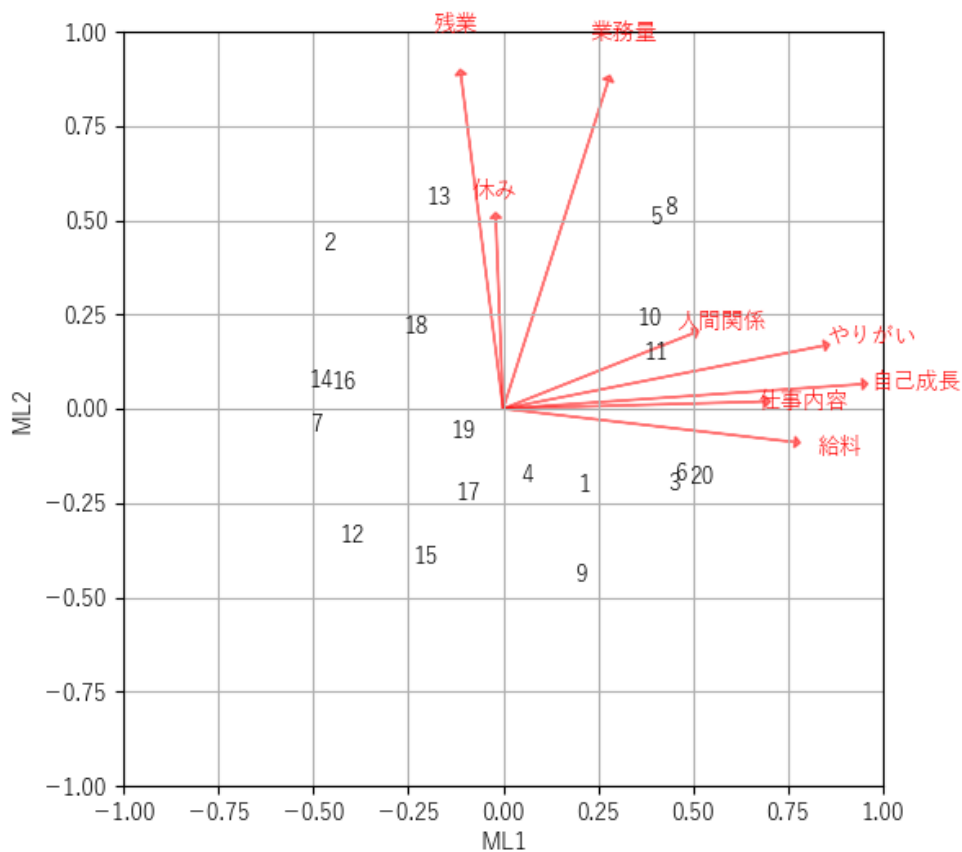
解析項目で「寄与率・累積寄与率」を選択した状態で「解析実行」をクリックすると、因子ごとの因子負荷量の二乗和・寄与率・累積寄与率が算出されます。



③ バイプロットの作成

バイプロットのドロップダウンメニューから「x 軸」と「y 軸」に設定する因子を選択します。以下の図は「x 軸」は第一因子 (=ML1) ，「y 軸」は第二因子 (=ML2) を設定した例です。

「グラフ作成」を押すと以下のような散布図が表示されます。



④ バイプロットの設定

バイプロットの表示設定は、「グラフ設定」をクリックすることで以下のダイアログが表示され、プロットや矢印のデザインを変更することができます。



8. 構造方程式モデリング

以下のサンプルデータを用いて解説します。



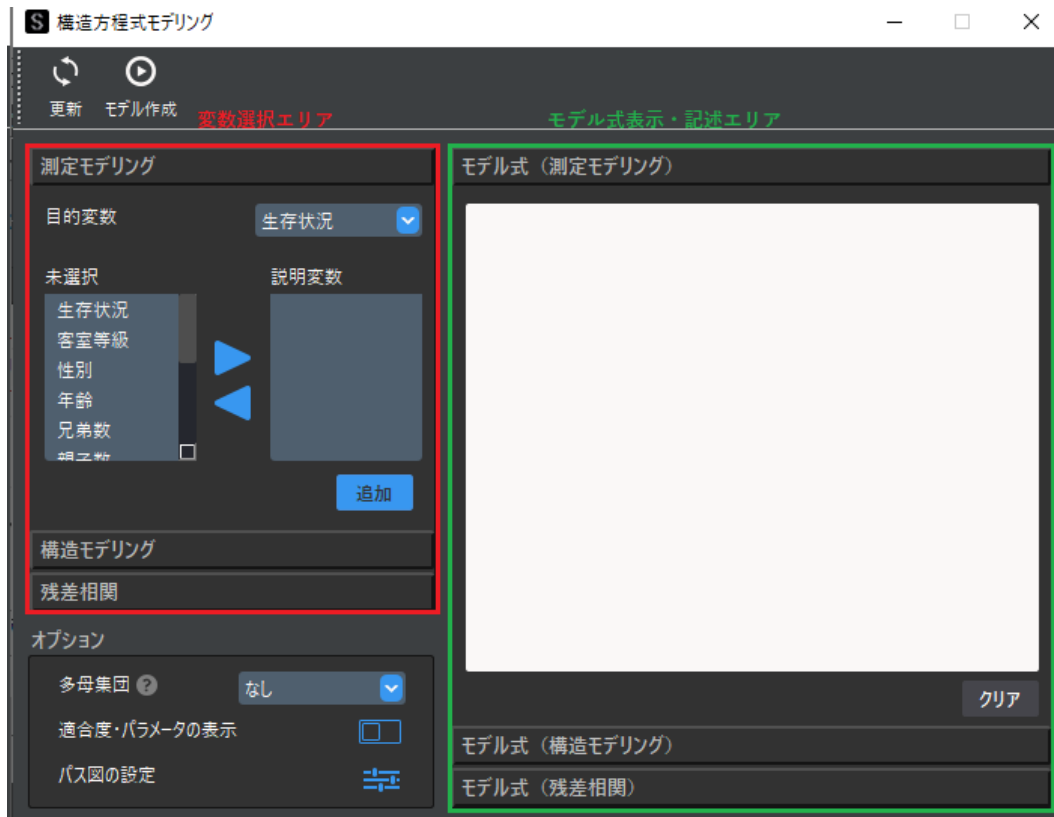
	データ1	データ2	データ3				
	生存状況	客室等級	性別	年齢	兄弟数	親子数	
1	0	3	male	22.0	1	0	
2	1	1	female	38.0	1	0	
3	1	3	female	26.0	0	0	
4	1	1	female	35.0	1	0	
5	0	3	male	35.0	0	0	
6	0	1	male	54.0	0	0	
7	0	3	male	2.0	3	1	
8	1	3	female	27.0	0	2	
9	1	2	female	14.0	1	0	
10	1	3	female	4.0	1	1	
11	1	1	female	58.0	0	0	
12	0	3	male	20.0	0	0	
13	0	3	male	39.0	1	5	
14	0	3	female	14.0	0	0	
15	1	2	female	55.0	0	0	
16	0	3	male	2.0	4	1	

① 構造方程式モデリング用ウィンドウ

メニューバーから「多変量解析」→「構造方程式モデリング」を選択して構造方程式モデリング用ウィンドウを表示します。

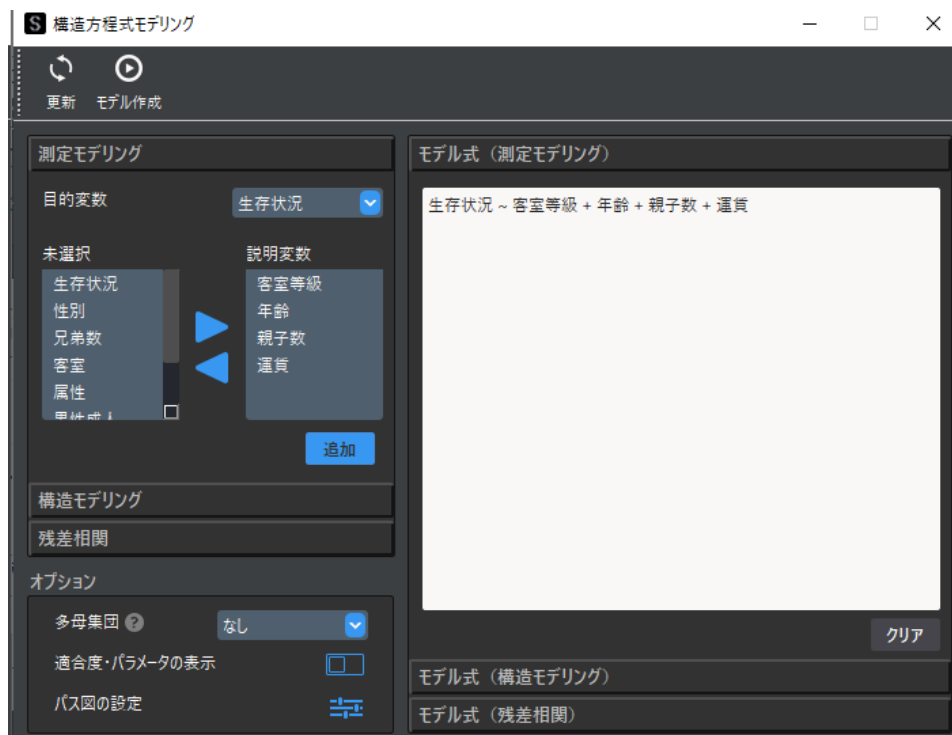
構造方程式モデリング用ウィンドウは、以下のような構成となっています。変数選択エリアで、変数名を選択してモデル式を追加します。

モデル式表示・記述エリアでは追加したモデル式が表示されます。モデル式表示・記述エリアでは自由に入力ができるので、テキストを修正したりコピー＆ペーストすることが可能です。



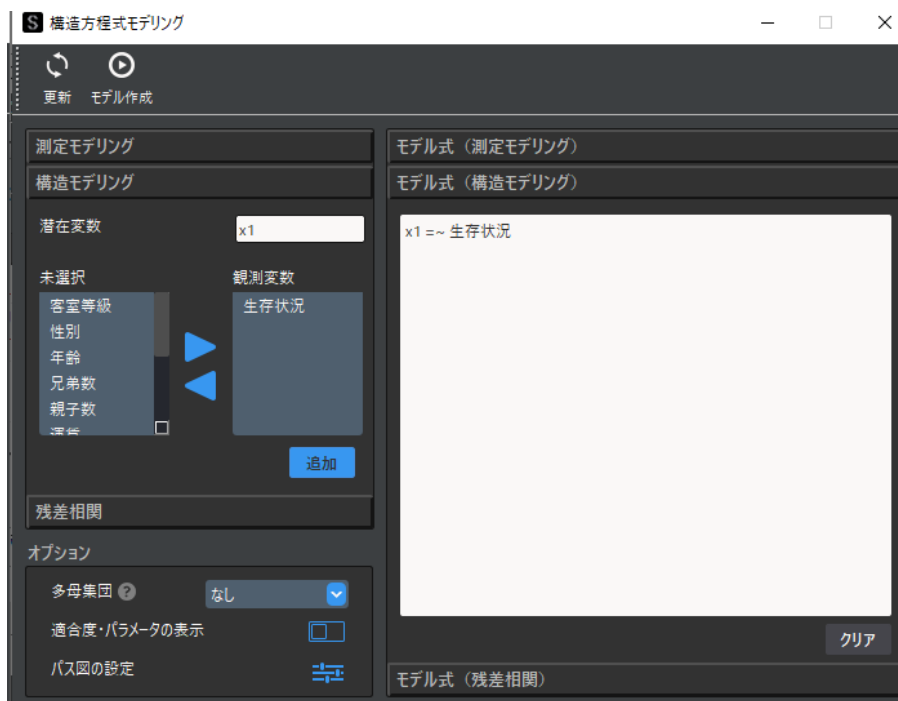
② 測定モデリングの定義

測定モデリング（回帰関係）のモデル式を追加する場合は、「目的変数」と「説明変数」を選択して、「追加」ボタンをクリックすると以下の用にモデル式に追加されます。



③ 構造モデリングの定義

構造モデリング（潜在変数）を追加する場合は、潜在変数名を入力して、「観測変数」を選択してから「追加」ボタンをクリックします。



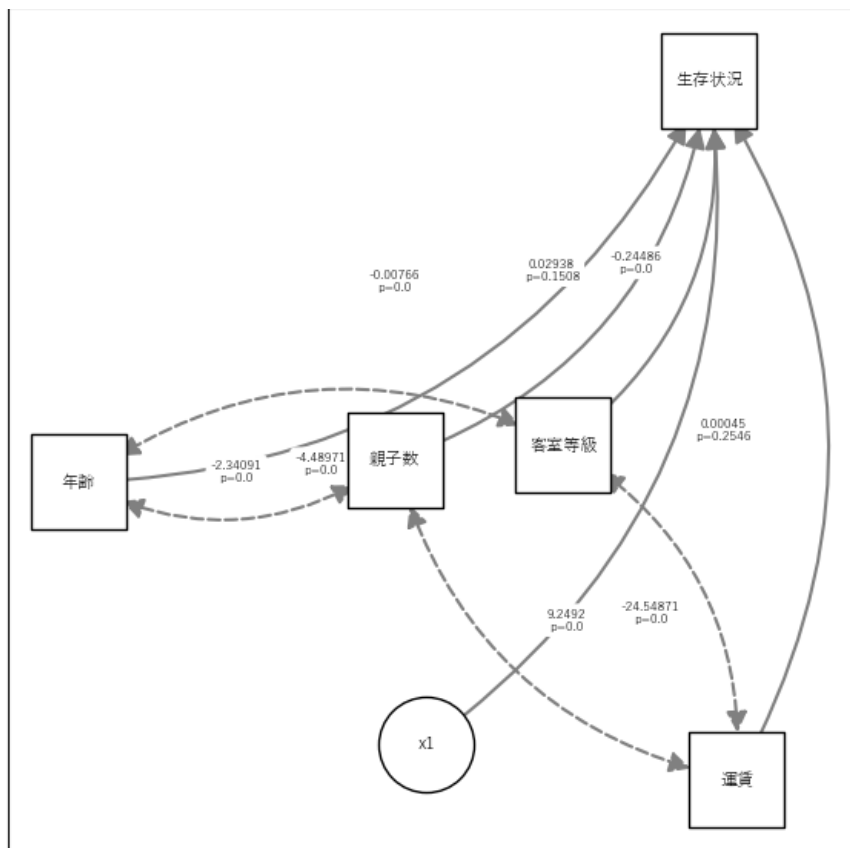
④ 残差相関の定義

残差相関（共分散関係）を追加する場合は、変数を選択して「追加」ボタンをクリックすると、選択した変数の全ての組み合わせについてのモデル式が追加されます。



⑤ パス図の表示（モデル作成）

モデル式の追加・記述が完了したら、ツールバーの「モデル作成」ボタンをクリックすると以下のようなパス図が表示されます。



パス図において四角形のノードが観測変数を示し、円のノードが潜在変数を示します。また、点線の両矢印は共分散関係を示します。

※ R の Lavaan パッケージと同等の表示内容です。

⑥ モデル式の記述方法

モデル式の記述方法は、構造方程式モデリングで最も使用されている R の Lavaan パッケージと同じです。

記述方法は以下のページで、丁寧に紹介されているのでお読みください。

[Lavaan チュートリアル \(関西大学\)](#)

StaatApp のモデル式表示・記述エリアは、測定モデリングと構造モデリング、残差相関に分かれています。どれかの表示・記述エリアに正しくテキストを入力すれば解析されます。テキスト

トエディターなどでモデル式を作成して、コピペする場合は適当なエリアに貼り付けてください。

⑦ 多母集団同時解析（応用）

多母集団同時解析を行なう場合は、「オプション」の「多母集団」から多母集団を示す属性が入力された列を選択してください。

多母集団同時解析を行なうと、選択した属性の種類数だけモデリングされパス図が表示されます。

⑧ 適合度・パラメータの表示

モデルの適合度やパラメータを表示する場合は、「オプション」の「適合度・パラメータの表示」にチェックを入れてください。

チェックを入れた状態で、モデル作成を行なうと以下のような結果用ウィンドウが表示されます。

適合度の指標		
X ² 統計量	0.00031	
RMSEA	0.0	
GFI	1.0	
AGFI	1.0	
CFI	1.00855	
AIC	20.0	

パラメータ			推定値	標準誤差	z値	p値
生存状況	~	客室等級	-0.24486	0.02586	-9.46774	0.0
生存状況	~	年齢	-0.00766	0.00126	-6.08746	0.0
生存状況	~	親子数	0.02938	0.02045	1.43673	0.1508
生存状況	~	運賃	0.00045	0.00039	1.13925	0.2546
生存状況	~	x1	1.0	nan	nan	nan
客室等級	~~	年齢	-4.48971	0.30976	-14.49411	0.0
年齢	~~	親子数	-2.34091	0.39649	-5.90408	0.0
親子数	~~	運賃	9.2492	1.28468	7.19964	0.0
運賃	~~	客室等級	-24.54871	0.9369	-26.20206	0.0
x1	~~	x1	0.06432	0.0052	12.37102	0.0
生存状況	~~	生存状況	0.13215	0.0052	25.41787	0.0

※ パス図のエッジ付近に表示される値は、パラメータの推定値と p 値になります。

⑨ パス図の設定（応用）

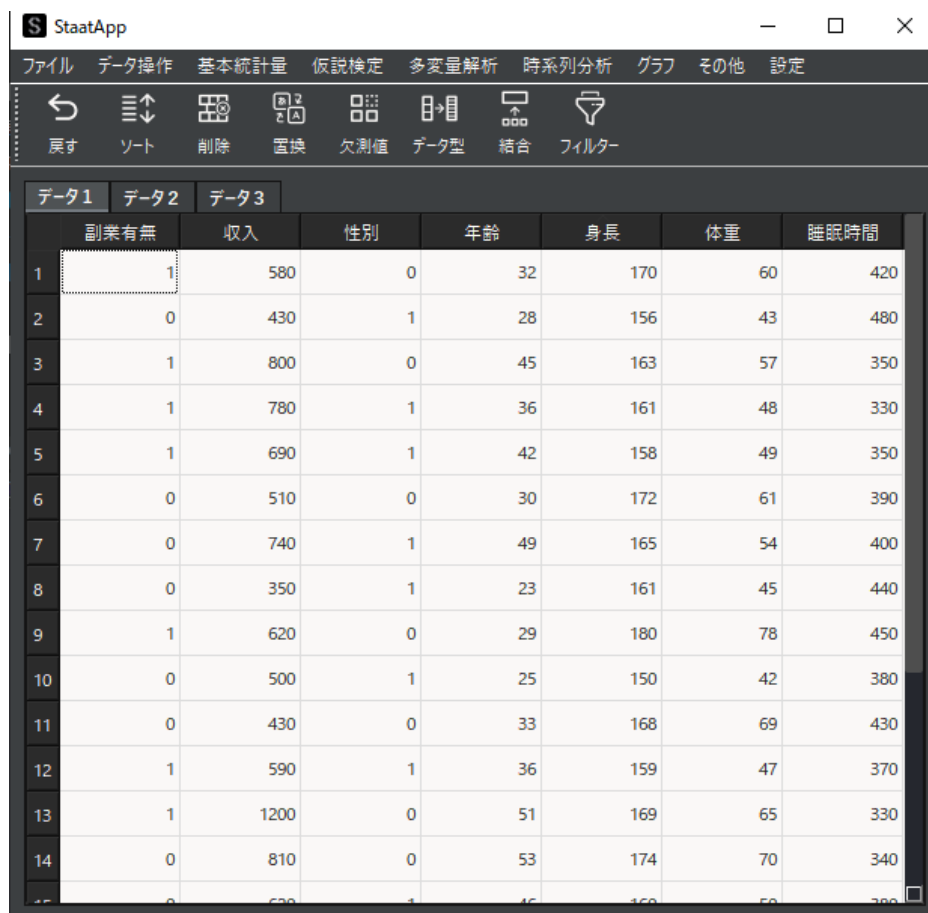
パス図の表示設定を変更する場合は、設定ボタンをクリックして、以下のようなグラフ設定用ウィンドウで変更します。



パス図のレイアウト（ノードの配置）を変更する場合は、「全体」の「レイアウト」で変更します。パス図はレイアウトの設定に加えて、モデル作成時にある程度ランダムな配置で作成されるので、好みのレイアウトにならない場合は、モデル作成を繰り返し行ってください。

9. クラスター分析

以下のサンプルデータを用いて解説します。



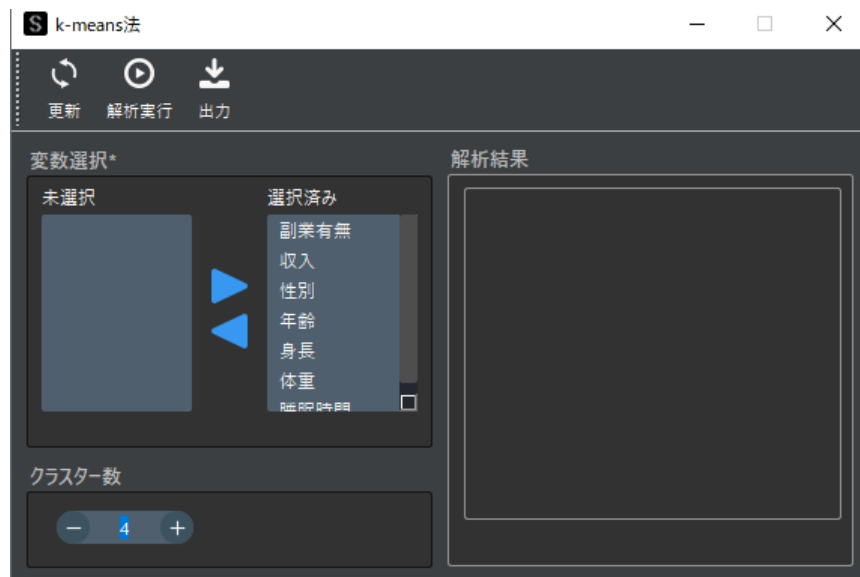
	データ1	データ2	データ3					
		副業有無	収入	性別	年齢	身長	体重	睡眠時間
1		1	580	0	32	170	60	420
2		0	430	1	28	156	43	480
3		1	800	0	45	163	57	350
4		1	780	1	36	161	48	330
5		1	690	1	42	158	49	350
6		0	510	0	30	172	61	390
7		0	740	1	49	165	54	400
8		0	350	1	23	161	45	440
9		1	620	0	29	180	78	450
10		0	500	1	25	150	42	380
11		0	430	0	33	168	69	430
12		1	590	1	36	159	47	370
13		1	1200	0	51	169	65	330
14		0	810	0	53	174	70	340
15		0	620	1	46	160	50	380

① k-means 法

メニューバーから「多変量解析」→「クラスタリング」→「k-means 法」を選択します。k-means 法用のウィンドウが表示されたら、クラスタリング対象を選択します。

ダミー変数を選択する際は、多重共線性の問題を回避するために1列分を除いて選択します。サンプルデータは既に1列分除いたダミー変数になっています。

クラスタリングを行う集団の数（クラスター数）を設定します。以下の画像ではクラスター数を4と設定しました。



設定が完了したらツールバーの「解析実行」ボタンをクリックして k-means 法を実行します。実行結果は以下のように表示されます。

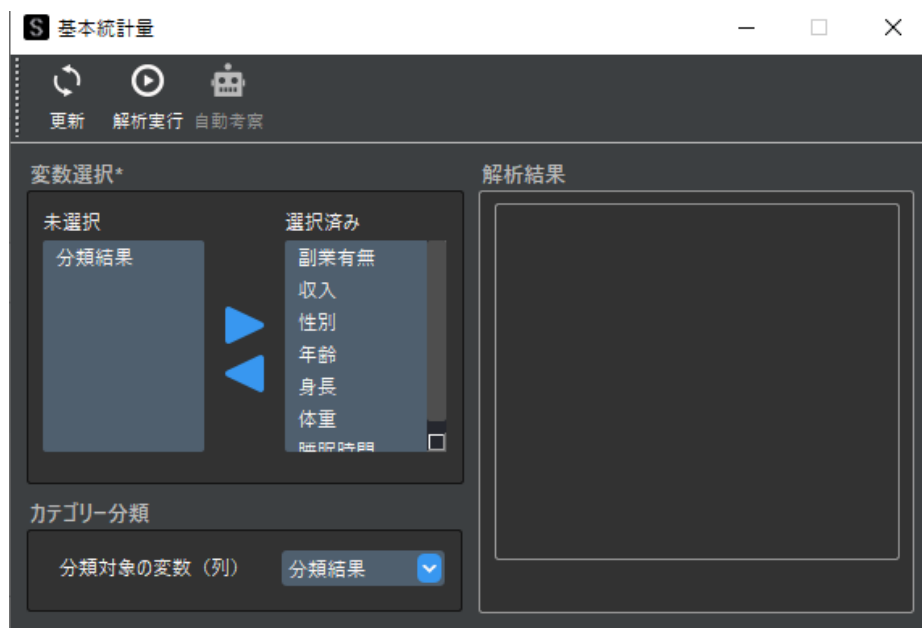
S クラスタリング結果							
	分類結果	副業有無	収入	性別	年齢	身長	体重
1	クラスター-2	1	580	0	32	170	60
2	クラスター-1	0	430	1	28	156	43
3	クラスター-4	1	800	0	45	163	57
4	クラスター-4	1	780	1	36	161	48
5	クラスター-4	1	690	1	42	158	49
6	クラスター-2	0	510	0	30	172	61
7	クラスター-4	0	740	1	49	165	54
8	クラスター-1	0	350	1	23	161	45
9	クラスター-2	1	620	0	29	180	78
10	クラスター-2	0	500	1	25	150	42

② クラスタリング結果の分析（応用）

クラスタリング結果は、「分類結果」列のクラスター番号を用いて様々な分析が可能です。追加で分析を行なう場合は出力機能を用いて、データ操作画面に保存します。

基本統計量（平均値や中央値）を、各クラスターの特徴を把握したい場合は、基本統計量機能（無料）を用います。

基本統計量機能で以下のように、「カテゴリー分類」にクラスター番号が入った列を指定します。



「解析実行」をクリックすると、変数・クラスターごとの統計量が表示されます。

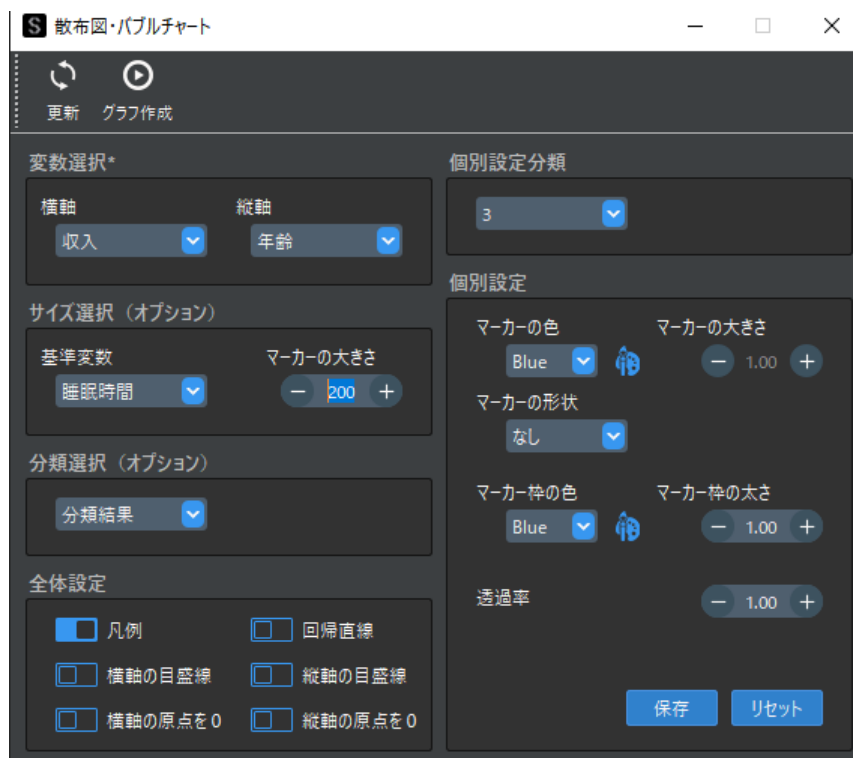


	副業有無	収入	性別	年齢	身長
クラスター1_サンプルサイズ	7.0	7.0	7.0	7.0	7.0
クラスター1_平均	0.0	407.14286	0.57143	29.85714	163.42857
クラスター1_標準偏差	0.0	51.22313	0.53452	5.95619	5.76938
クラスター1_最小値	0.0	320.0	0.0	23.0	156.0
クラスター1_第一四分位数	0.0	390.0	0.0	26.5	158.5
クラスター1_中央値	0.0	430.0	1.0	28.0	166.0
クラスター1_第三四分位数	0.0	430.0	1.0	32.0	167.5
クラスター1_最大値	0.0	460.0	1.0	41.0	170.0
クラスター2_サンプルサイズ	8.0	8.0	8.0	8.0	8.0
クラスター2_平均	0.625	576.25	0.5	33.375	166.625

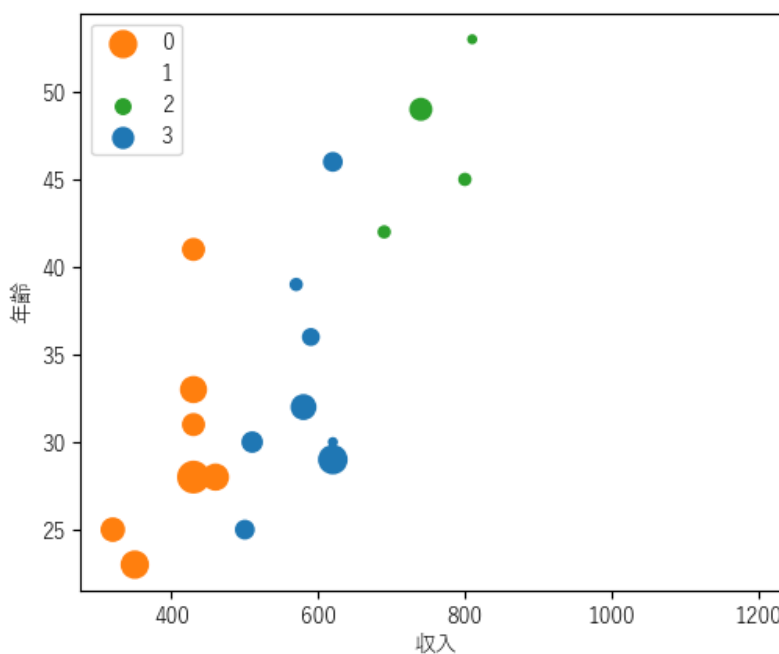
自動考察機能を用いると、クラスターごとのデータの特徴についての考察を得ることができます。

③ クラスタリング結果の可視化（応用）

クラスタリング結果は散布図やバブルチャートなどで可視化することができます。保存したデータを対象に散布図・バブルチャート機能（無料）で以下のように設定をします。



以下のようにクラスタリング結果ごとに色付けしたバブルチャートを作成することができます。



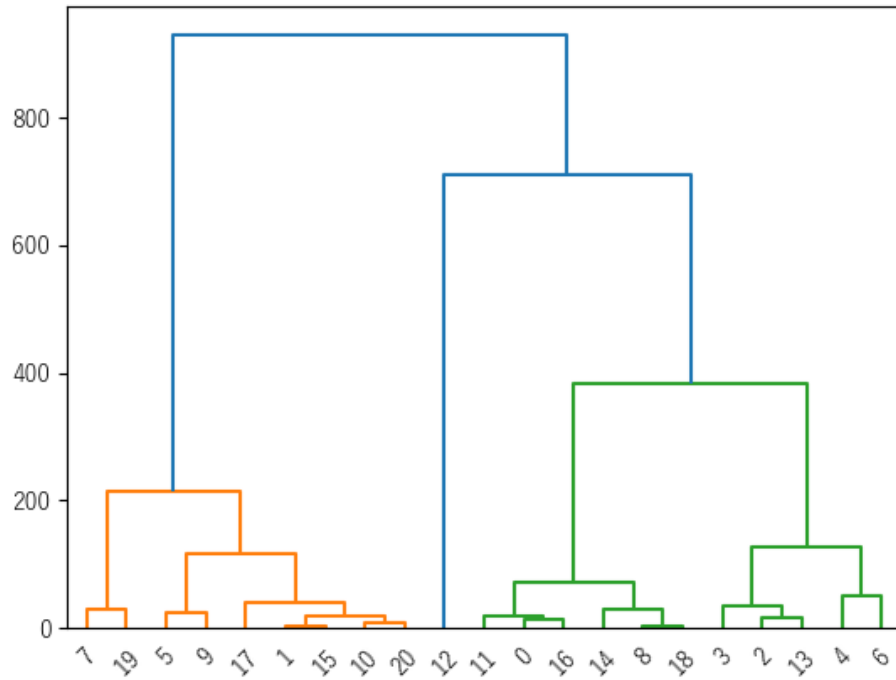
④ 階層クラスター分析

メニューバーから「多変量解析」→「クラスタリング」→「階層クラスター分析」を選択します。階層クラスター分析用のウィンドウが表示されたら、クラスタリング対象とする変数名を選択します。



距離測定法とクラスタリング手法の設定が可能です。デフォルトでは階層クラスター分析で最も一般的な手法が設定されているため、特にこだわりがない場合は設定する必要はありません。

設定が完了したらツールバーの「解析実行」ボタンをクリックして階層クラスター分析を実行します。

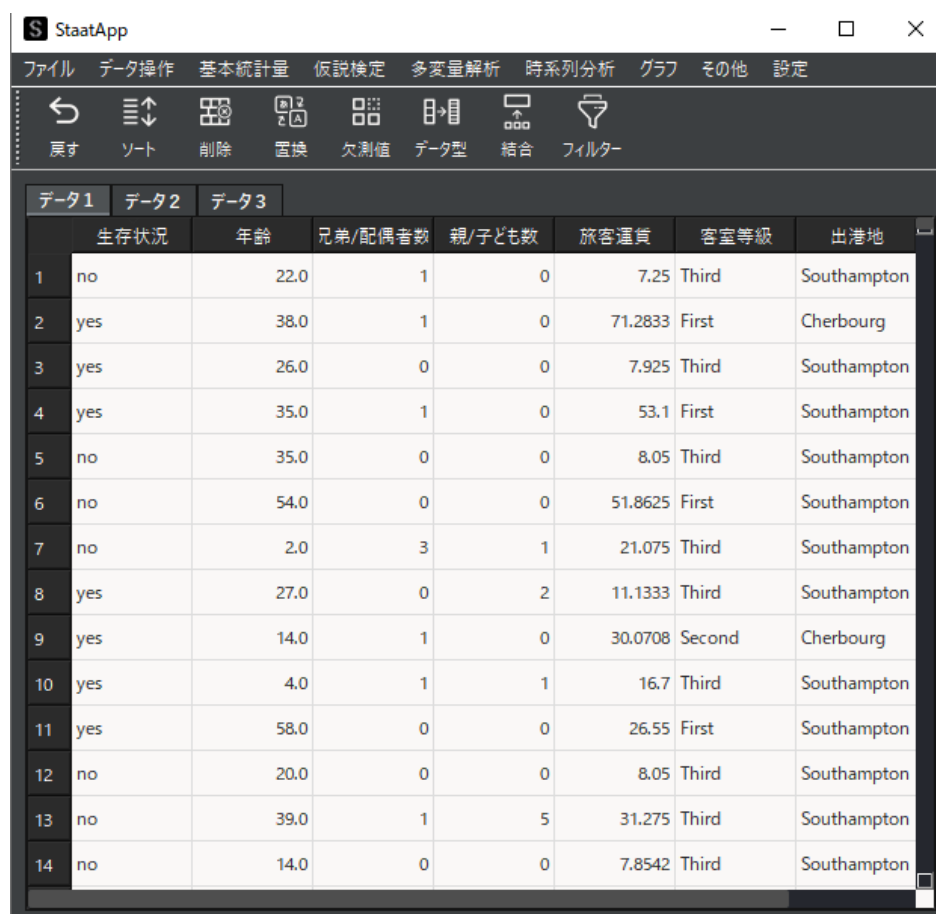


実行すると画像のようなデンドログラム（樹形図）が表示されます。樹形図の分岐に近いサンプルほど類似したクラスターに分類されていることがわかります。

10. 決定木（無料）

※ ランダムフォレスト・勾配ブースティング決定木についても同様の操作方法で実行可能です。

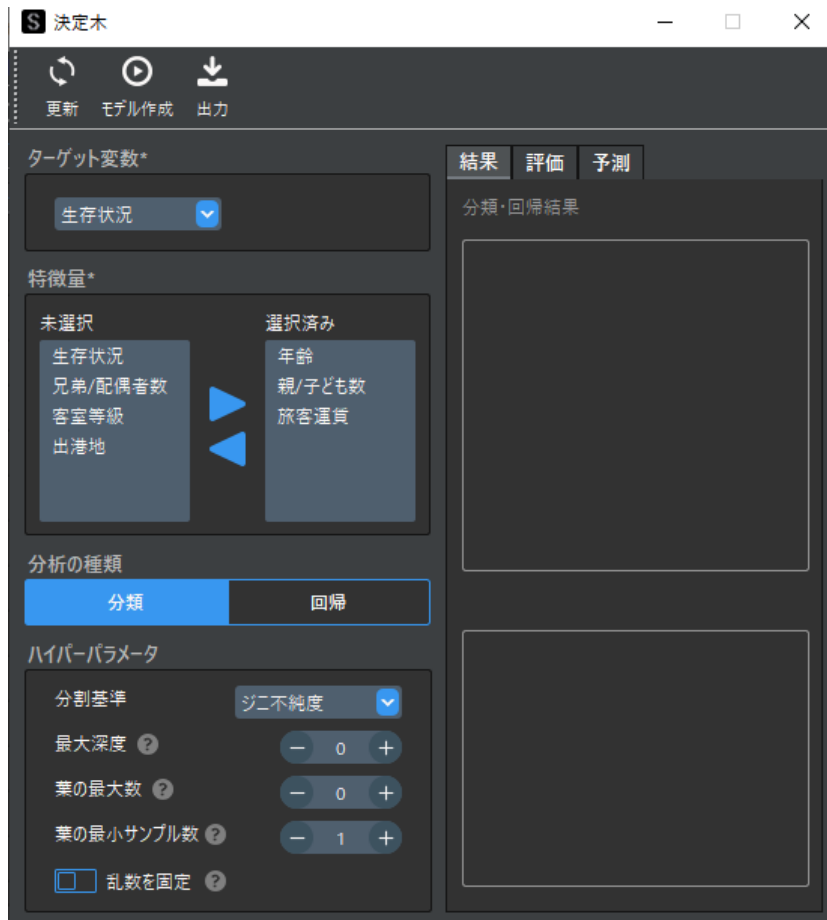
以下のサンプルデータを用いて解説します。



	データ1	データ2	データ3					
	生存状況	年齢	兄弟/配偶者数	親/子ども数	旅客運賃	客室等級	出港地	
1	no	22.0	1	0	7.25	Third	Southampton	
2	yes	38.0	1	0	71.2833	First	Cherbourg	
3	yes	26.0	0	0	7.925	Third	Southampton	
4	yes	35.0	1	0	53.1	First	Southampton	
5	no	35.0	0	0	8.05	Third	Southampton	
6	no	54.0	0	0	51.8625	First	Southampton	
7	no	2.0	3	1	21.075	Third	Southampton	
8	yes	27.0	0	2	11.1333	Third	Southampton	
9	yes	14.0	1	0	30.0708	Second	Cherbourg	
10	yes	4.0	1	1	16.7	Third	Southampton	
11	yes	58.0	0	0	26.55	First	Southampton	
12	no	20.0	0	0	8.05	Third	Southampton	
13	no	39.0	1	5	31.275	Third	Southampton	
14	no	14.0	0	0	7.8542	Third	Southampton	

メニューバーから「多変量解析」→「機械学習」→「決定木」を選択します。

決定木用ウィンドウを表示したらターゲット変数（目的変数）には分類を行いたい変数を，特徴量（説明変数）には分類するための基準とする変数を選択します。



ハイパーパラメータを調整することで過学習や未学習を避け、モデルの予測精度を向上させることができます。例ではデフォルト設定に加えて最大深度を”3”に設定します。

① モデルの作成

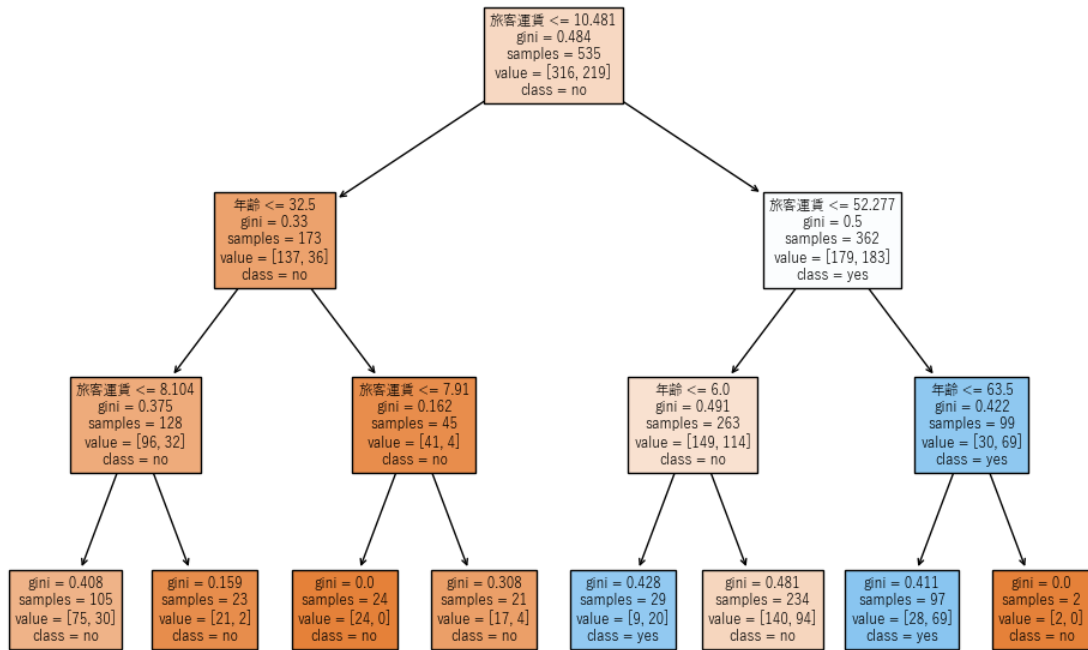
ツールバーの「モデル作成」ボタンでモデルの作成を実行します。分類問題では以下のように、木構造のグラフが作成されます。

木構造の見方としては一番上の枠（ノード）の1行目に書かれている特徴量が、最も分類結果に影響を与えている特徴量になります。例では生存有無に最も影響を与えているのは旅客運賃となります。

3行目の「samples」このノードでのサンプル数を示し、4行目の「value」は現段階での構成比（生存有無の no=316, yes=219）を示します。

1つ目の条件が”True”であった場合に左のノードに進み、”False”であった場合に右のノードに進みます。例では旅客運賃が10.484ポンド以下であった場合に、左のノードに進みます。

ノードの色は構成比の偏りを示し、生存有無で no が多いほどオレンジ色，yes が多いほど青色になります。



画面右側の解析結果にはサンプルごとの分類結果と特徴量ごとの重要度（影響度）が表示されます。

決定木

更新 モデル作成 出力

ターゲット変数*
生存状況

特徴量*
未選択: 生存状況, 兄弟/配偶者数, 客室等級, 出港地
選択済み: 年齢, 親/子ども数, 旅客運賃

分析の種類
分類 回帰

ハイパーパラメータ
分割基準: ジニ不純度
最大深度: 3
葉の最大数: 0
葉の最小サンプル数: 1
 乱数を固定

結果 評価 予測

分類・回帰結果

	予測結果	生存状況	年齢
1	no	no	
2	yes	yes	
3	no	yes	
4	yes	yes	
5	no	no	
6	no	no	
7	yes	no	

重要度

特徴量	重要度
年齢	0.19492
親/子ども数	0.0
旅客運賃	0.80508

ハイパーパラメータ

- **分割基準**：分類問題では”ジニ係数”・”エントロピー”・”シャノンゲイン情報”から選択します。ジニ係数の方が連続データに強く、カテゴリーデータに対してはエントロピーが強いとされています。回帰問題では”二乗誤差”・”平均二乗誤差”・”絶対誤差”・”ポアソン逸脱度”から選択可能です。
- **最大深度**：ツリーの最大深度は値が大きいほど、深く分割を行います。デフォルト設定では全ての葉が1になるまで分割されます。過学習を防ぐためには最大値を設定します。また、過学習の防止に最も効果がある場合があります。
- **葉の最大数**：最大の葉の数が多いほど分割されます。デフォルト設定では葉の数は無制限です。過学習を防ぐためには値を小さくして、分割回数を制限します。
- **葉の最小サンプル数**：葉を構成するのに必要な最小限のサンプルの数になります。値が小さいと過学習になる可能性があります。
- **乱数の固定**：学習ごとに用いられる乱数を固定します。乱数を固定した場合、同じデータで学習した場合の演算結果が等しくなります。

② 予測精度の評価

作成したモデルに対して、テストデータを用いてモデルの精度評価を行います。

※ テストデータは事前に用意もしくは、テストデータ作成機能で作成する必要があります。

評価タブを表示してテストデータに「データ2」を選択します。「算出」ボタンをクリックすると以下のように4つの評価指標と混合行列が表示されます。

評価指標はそれぞれ値が大きいほどモデルの予測精度が高いことを意味します。混合行列はテストデータ（実測値）と予測値を比較した行列で、生存有無が”yes”の場合はほとんど予測が的中していることがわかります。



③ 予測値の算出

作成したモデルに対して予測用データを読み込ませて、ターゲット変数の予測を行います。

予測タブを表示し予測用データの選択、「算出」ボタンをクリックすると予測結果が表示されます。

予測値

	予測値_生存状況
1	no
2	no
3	yes
4	yes
5	no
6	no
7	no
8	no
9	no

④ 回帰問題の実行

回帰問題を行う場合はターゲット変数に数値データを選択して、分析の種類に回帰を選択します。回帰問題では予測値が連続データとして算出されます。

The screenshot shows the '決定木' (Decision Tree) interface in StaatApp. The 'ターゲット変数*' (Target Variable) is set to '年齢' (Age). Under '特微量*' (Features), '生存状況', '兄弟/配偶者数', '客室等級', '出港地', and '年齢' are listed as '未選択' (Not Selected), while '親/子ども数' and '旅客運賃' are listed as '選択済み' (Selected). The '分析の種類' (Analysis Type) is set to '回帰' (Regression). Under 'ハイパーパラメータ' (Hyperparameters), '分割基準' (Split Criterion) is '二乗誤差' (Squared Error), '最大深度' (Maximum Depth) is 3, '葉の最大数' (Maximum Number of Leaves) is 0, and '葉の最小サンプル数' (Minimum Number of Samples per Leaf) is 1. The '乱数を固定' (Fix Randomness) checkbox is unchecked. The '結果' (Results) tab is active, showing a table of predicted values, actual age, and the number of family members. Below the table is a '重要度' (Importance) table.

	予測結果	年齢	親/子ども数
1	28.4472	22.0	
2	34.67742	38.0	
3	28.4472	26.0	
4	34.67742	35.0	
5	28.4472	35.0	
6	43.23529	54.0	
7	17.81515	2.0	

	重要度
親/子ども数	0.50824
旅客運賃	0.49176

モデルの評価指標は分類問題とは異なり RMSE（平均平方二乗誤差）や MAE（平均絶対誤差）で評価します。これらの値は予測結果と実測値との誤差を示し、値が小さいほどモデルの予測精度が高いと言えます。

R^2 （決定係数）は 0~1 の間で算出され、1 に近いほど当てはまりがよいとされます。

決定木

更新 モデル作成 出力

ターゲット変数*
年齢

特徴量*
未選択: 生存状況, 兄弟/配偶者数, 客室等級, 出港地, 年齢
選択済み: 親/子ども数, 旅客運賃

分析の種類
分類 回帰

ハイパーパラメータ
分割基準: 二乗誤差
最大深度: 3
葉の最大数: 0
葉の最小サンプル数: 1
 乱数を固定

結果 評価 予測

テストデータの選択: データ2
算出

指標値

	指標値
RMSE	12.55264
MAE	9.827
決定係数	0.21774

混合行列

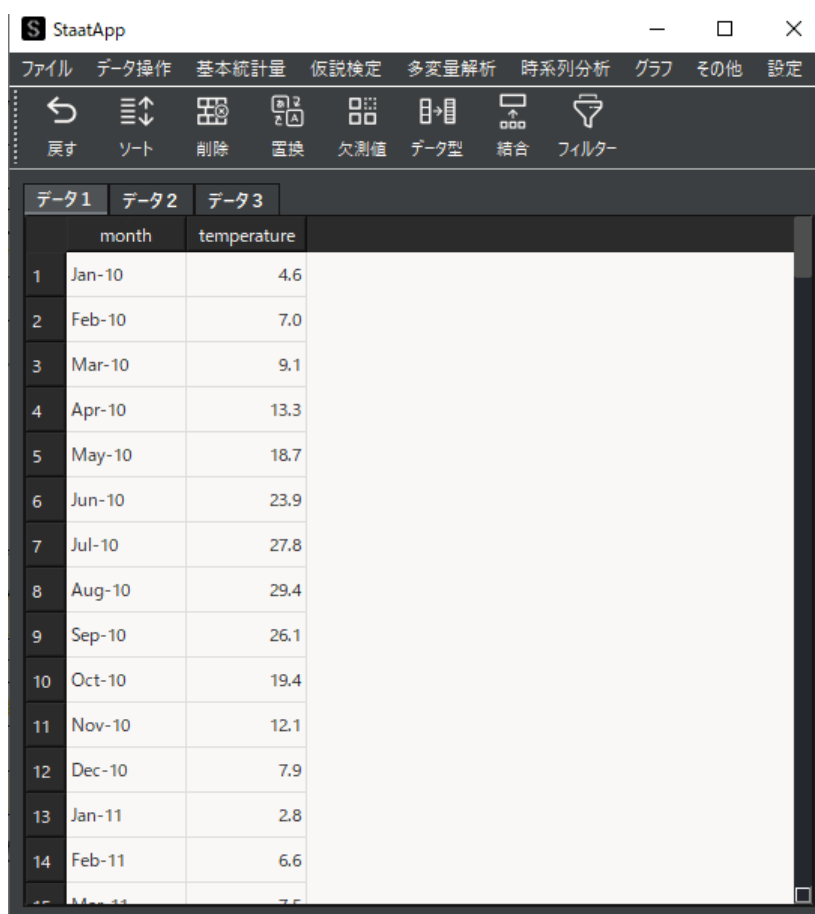
予測結果

	yes	no
yes	101	7
no	35	36

実測値

11. ARIMA（無料）

以下のサンプルデータ（名古屋市の月別平均気温）を用いて解説します。



The screenshot shows the StaatApp interface with a data table. The table has two columns: 'month' and 'temperature'. The data is as follows:

	month	temperature
1	Jan-10	4.6
2	Feb-10	7.0
3	Mar-10	9.1
4	Apr-10	13.3
5	May-10	18.7
6	Jun-10	23.9
7	Jul-10	27.8
8	Aug-10	29.4
9	Sep-10	26.1
10	Oct-10	19.4
11	Nov-10	12.1
12	Dec-10	7.9
13	Jan-11	2.8
14	Feb-11	6.6
15	Mar-11	7.5

※ StaatApp では、折れ線グラフ作成機能や自己相関係数機能で、ARIMA モデル作成前に時系列データの特徴を把握することができます。

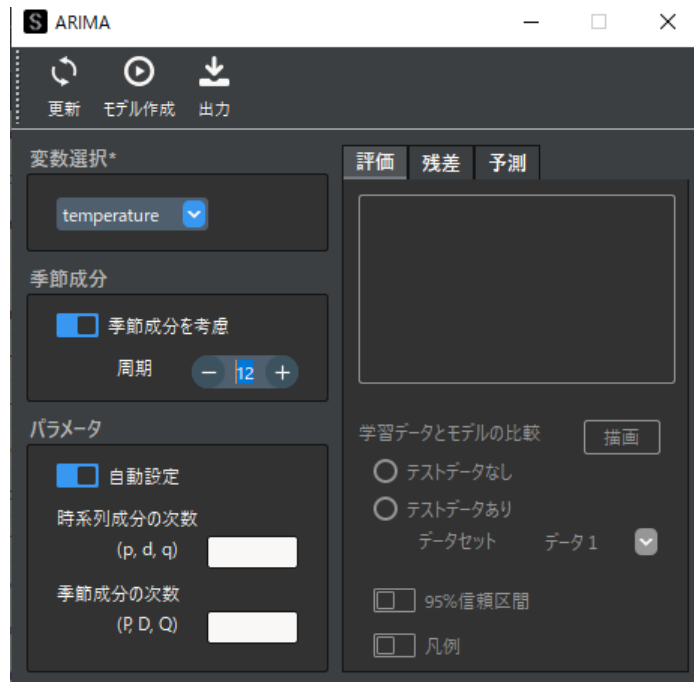
▷ StaatApp で行う時系列分析（ボックス・ジェンキンス法）

① モデルの作成

メニューバーから「時系列分析」→「モデル」→「ARIMA」を選択します。

変数選択で対象のデータ列名を選択します。次に季節成分では「季節成分を考慮」を選択して、周期を 12 に設定します。ARIMA 機能では「季節成分を考慮」を選択すると SARIMA モデルが選択されます。

パラメータは自動設定（デフォルト）のままとします。



「モデル作成」ボタンをクリックするとモデルが作成され評価指標が表示されます。

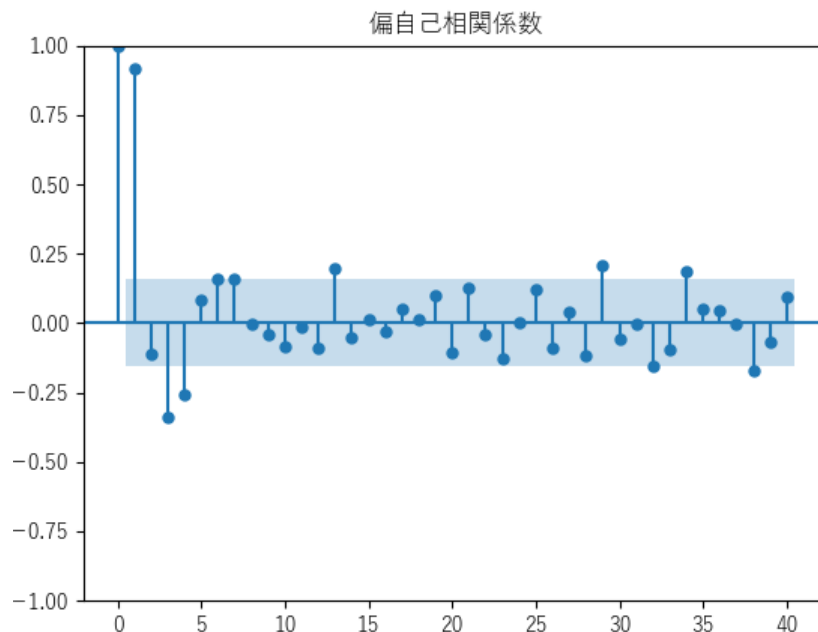
② モデルの評価

評価指標を用いてモデルの評価を行います。ARIMA 機能で算出される評価指標は全て絶対的な基準はありません。各指標とも値が小さいほど、良いモデルであると考えることができます。

StaatApp ではモデル作成時のパラメータ自動計算において、AIC が最小になるような値を求めて設定しています。



残差分析は残差タブの「コレログラム」選択して残差の偏自己相関係数のコレログラムを作成します。



青帯の範囲の外にある値は、有意水準 $\alpha=0.05$ で無相関の検定における有意を示します。ラグ5以下の値を除いて、周期ごとに偏自己相関係数が有意となっていますがそれほど大きな値ではないため比較的あてはまりのよいモデルと評価することができます。

③ 将来予測

作成したモデルから将来予測を行います。予測タブから任意の時点数の選択して、将来予測を行います。

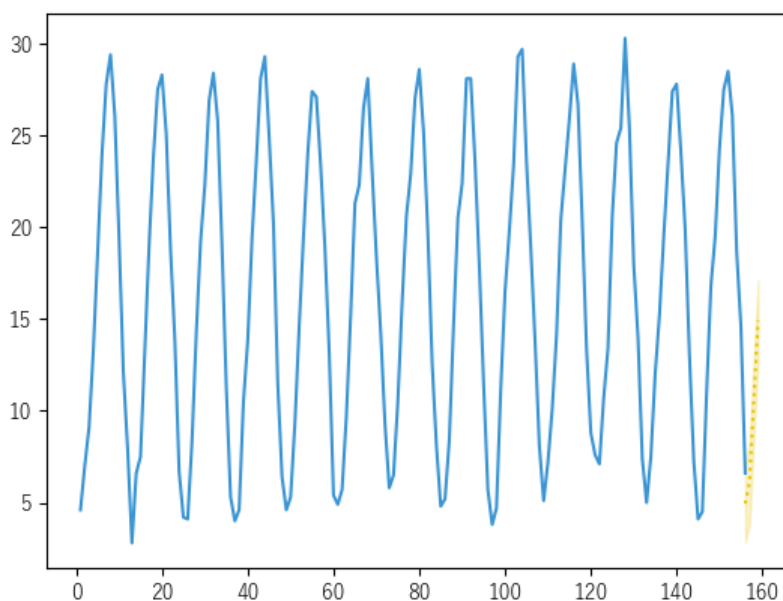


The screenshot shows the ARIMA software interface with the following settings:

- 変数選択*: temperature
- 季節成分: 季節成分を考慮, 周期: 12
- パラメータ: 自動設定, 時系列成分の次数 (p, d, q): (0, 0, 1), 季節成分の次数 (P, D, Q): (2, 1, 1)
- 予測タブ: 予測時点数: 4, 予測値: 算出
- 予測値表:

予測時点	値
1	4.95045
2	6.08915
3	10.58905
4	14.91255

Additional options: 予測グラフ: 95%信頼区間, 描画

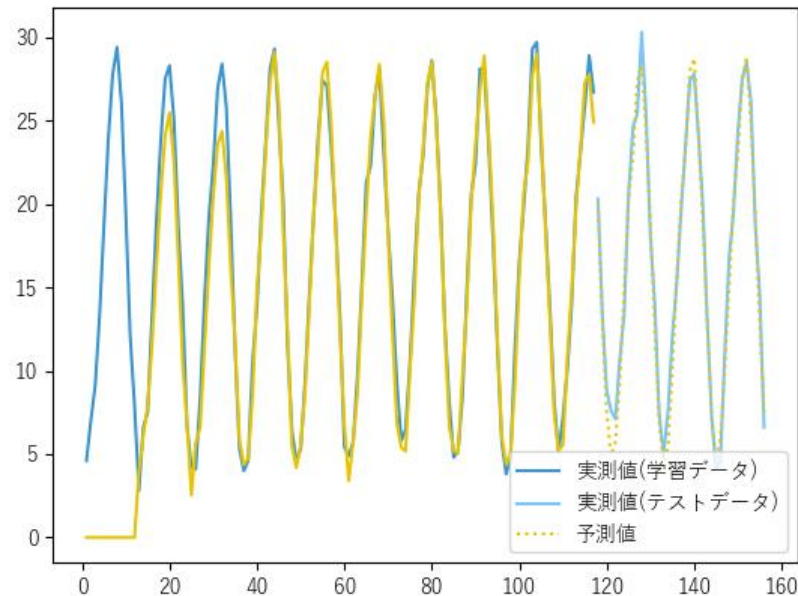


④ ホールドアウト検証

StaatApp ではモデル評価の方法として、事前に学習データとテストデータを分割して、学習データから作成したモデルの予測値とテストデータを比較するホールドアウト検証が可能です。

ARIMA 機能を用いる前に、読み込ませたデータを「テストデータの作成」機能で分割を行います。

分割を行ったら学習データでモデルを作成して、モデルの評価タブにある「学習データとモデルの比較」で「テストデータあり」を選択すると以下のようなグラフを作成することができます。



⑤ AR・MA モデル

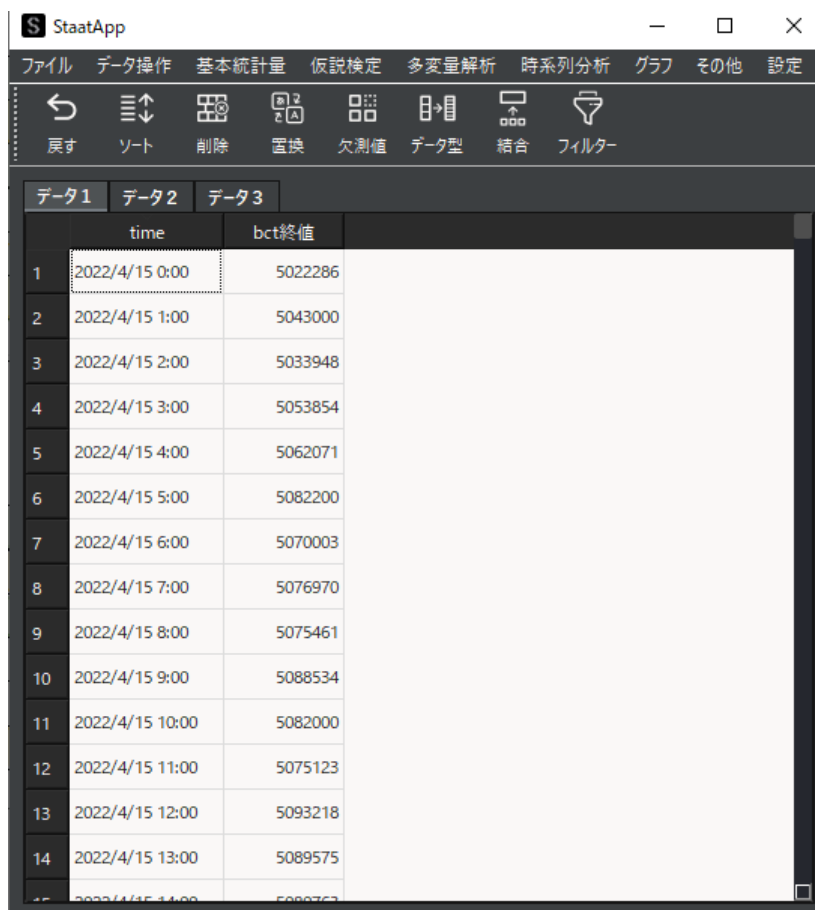
ARIMA 機能ではデフォルトでパラメータが自動設定されますが、自動設定を行わず任意のパラメータを設定することで AR モデルや MA モデルを用いることが可能です。

例えば、AR(1)モデルを用いる場合はパラメータ設定の「時系列成分の次数」に(1, 0, 0)と入力してください。

MA(1)モデルを用いる場合は、(0, 0, 1)と入力します。

12. GARCH モデル

以下のサンプルデータ（2022 年第 1 四半期の 1 時間ごとのビットコインの価格データ）を用いて解説します。



The screenshot shows the StaatApp interface with a data table. The table has two columns: 'time' and 'bct終値'. The data rows are numbered 1 through 14, representing hourly data from 2022/4/15 0:00 to 2022/4/15 13:00. The 'bct終値' column contains numerical values representing Bitcoin prices.

	time	bct終値
1	2022/4/15 0:00	5022286
2	2022/4/15 1:00	5043000
3	2022/4/15 2:00	5033948
4	2022/4/15 3:00	5053854
5	2022/4/15 4:00	5062071
6	2022/4/15 5:00	5082200
7	2022/4/15 6:00	5070003
8	2022/4/15 7:00	5076970
9	2022/4/15 8:00	5075461
10	2022/4/15 9:00	5088534
11	2022/4/15 10:00	5082000
12	2022/4/15 11:00	5075123
13	2022/4/15 12:00	5093218
14	2022/4/15 13:00	5089575

① 定常性の確認

ADF 検定を行い定常過程かどうか判定します。



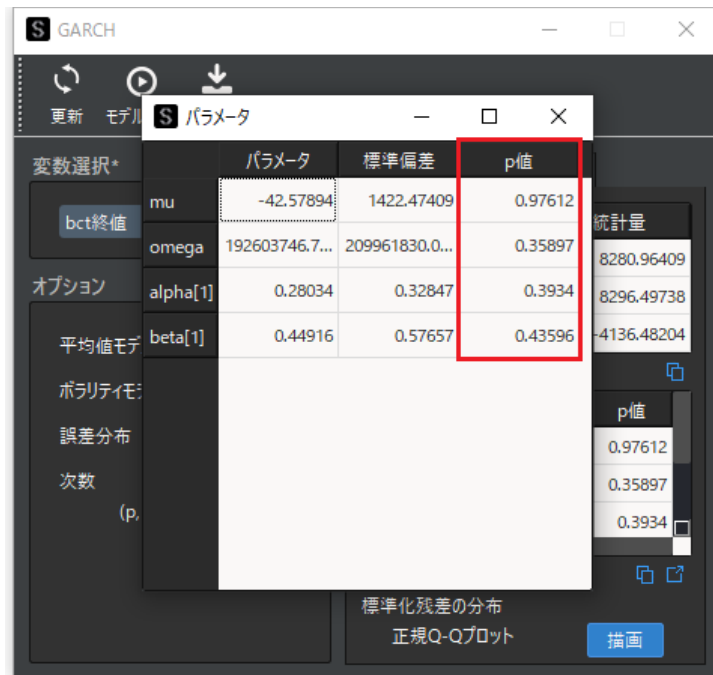
ADF 検定の値が全て有意水準=0.05 より大きいため、BTC 価格は単位根過程（≒非定常過程）であると言えます。GARCH モデルは定常過程の時系列データに対して有効なため、差分変換を行います。



② モデルの作成・評価

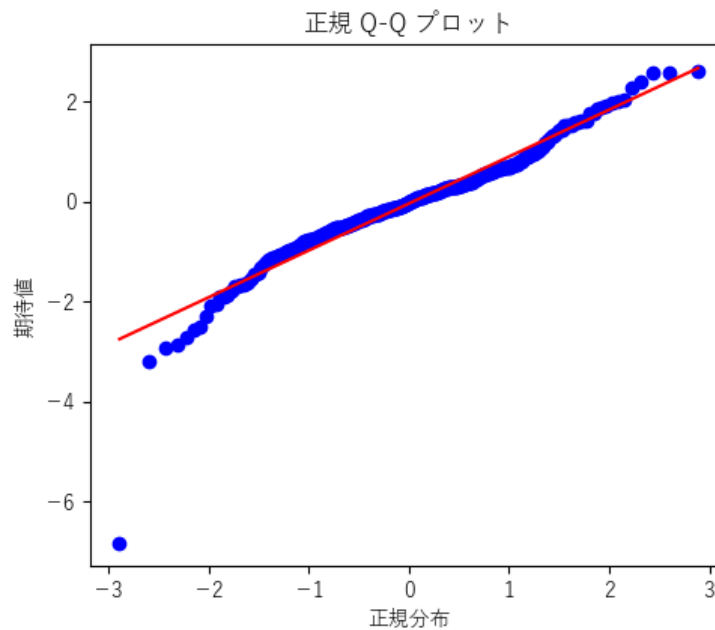
GARCH 機能でモデル作成を行います。分析対象のデータを選択して、「モデル作成」ボタンをクリックするとモデルが作成されます。

オプションでは平均値モデルやボラティリティモデル、誤差分布を選ぶことができますが基本的にはデフォルト設定の GARCH(1,1)で大丈夫です。



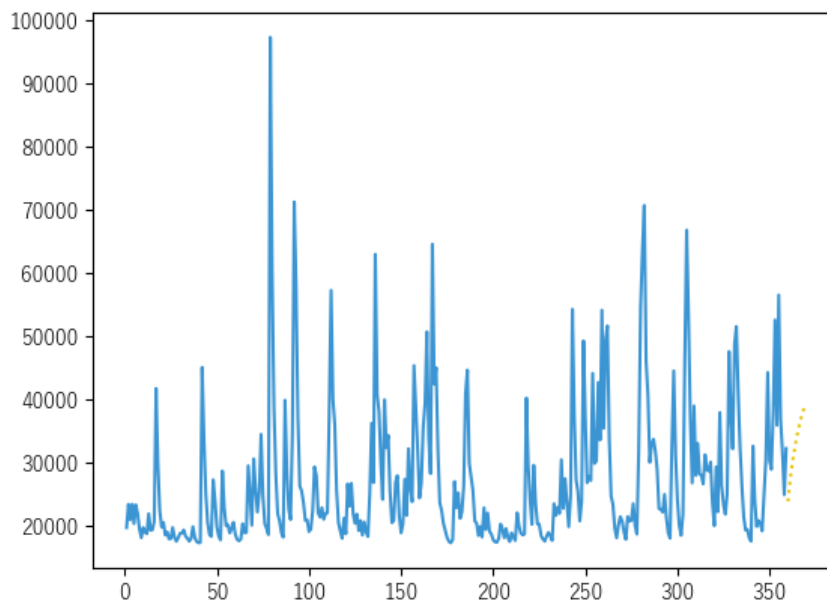
モデルが作成されたら評価指標を見ます。特に GARCH モデルではモデル式の変数の p 値に注目します。この値が小さいほどモデル式の各項は統計学的に意味がある値であることを示し、モデルの当てはまりが良いと言えます。

標準化残差の分布では残差はほぼ正規分布に従い偏りが無いということがわかります。



③ ボラティリティの算出

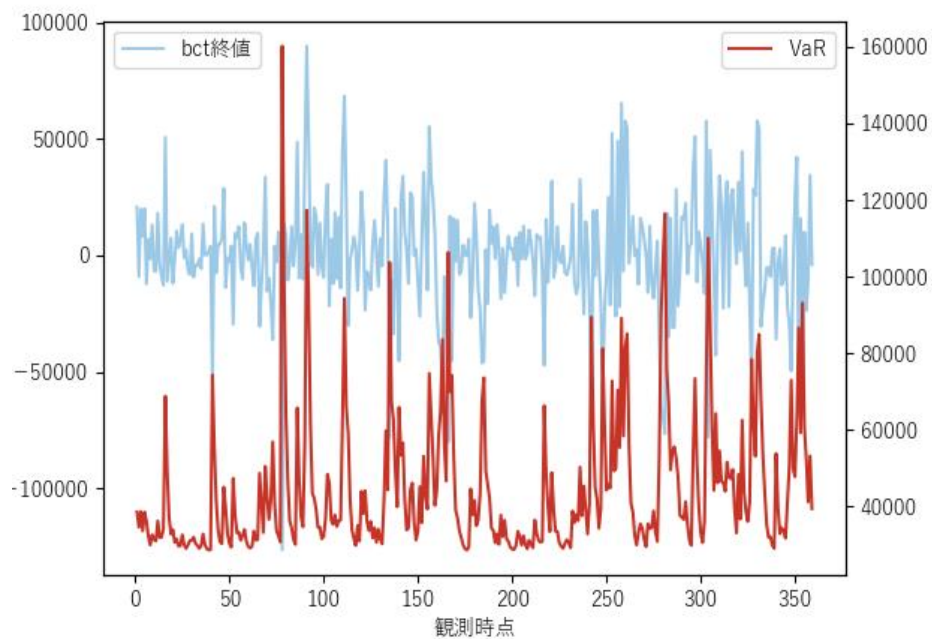
ボラティリティの算出と将来予測を行います。「ボラティリティ」タブの「予測」をオンにして予測時点数を”10”に設定します。「描画」ボタンをクリックするとボラティリティが算出され、将来予測値を含むグラフが表示されます。



④ VaR の算出

VaR の将来予測（観測時点の次の時点の値）を算出します。

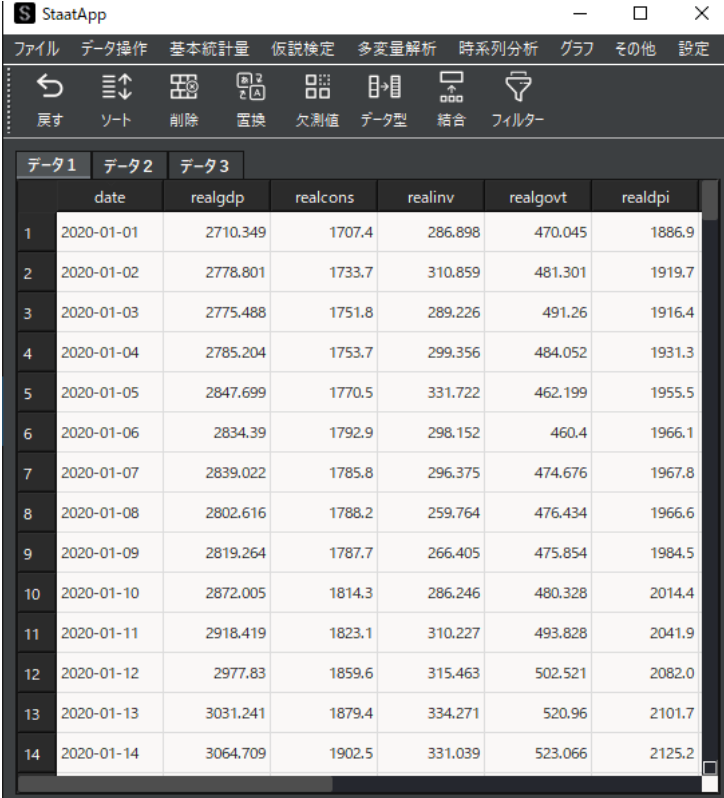
※ StaatApp では正規分布を用いて VaR を算出しています。



有意水準 $\alpha=0.05$ で 39413.43...という値になります．今回の分析例では BCT の価格変動データの差分系列に対しての VaR になるので，1 時間後には 5% の確率で 1 単位あたり 39413.43 円以上の損失が発生することを意味します．

13. VAR モデル

以下のサンプルデータ（マクロ経済学データセット macrodata）を用いて解説します。



データ1	データ2	データ3				
	date	realgdp	realcons	realinv	realgovt	realdpi
1	2020-01-01	2710.349	1707.4	286.898	470.045	1886.9
2	2020-01-02	2778.801	1733.7	310.859	481.301	1919.7
3	2020-01-03	2775.488	1751.8	289.226	491.26	1916.4
4	2020-01-04	2785.204	1753.7	299.356	484.052	1931.3
5	2020-01-05	2847.699	1770.5	331.722	462.199	1955.5
6	2020-01-06	2834.39	1792.9	298.152	460.4	1966.1
7	2020-01-07	2839.022	1785.8	296.375	474.676	1967.8
8	2020-01-08	2802.616	1788.2	259.764	476.434	1966.6
9	2020-01-09	2819.264	1787.7	266.405	475.854	1984.5
10	2020-01-10	2872.005	1814.3	286.246	480.328	2014.4
11	2020-01-11	2918.419	1823.1	310.227	493.828	2041.9
12	2020-01-12	2977.83	1859.6	315.463	502.521	2082.0
13	2020-01-13	3031.241	1879.4	334.271	520.96	2101.7
14	2020-01-14	3064.709	1902.5	331.039	523.066	2125.2

サンプルデータには長期変動があるため、「差分・対数変換」機能で差分変換を行ってから VAR モデルの作成を行います。

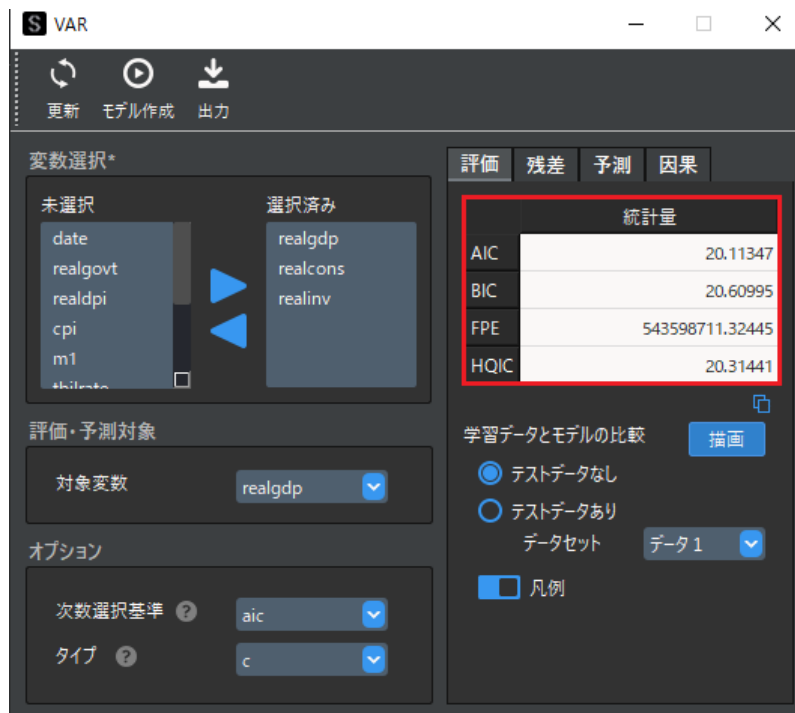
定常過程の確認は、「折れ線グラフ」機能や「ADF 検定」機能でも確認することができます。

① モデルの作成・評価

VAR モデル機能を用いてパラメータを自動推定して、モデルを作成します。

「変数選択」3つの変数を選択して、「モデル作成」ボタンをクリックするとモデルが作成されます。VAR モデル機能ではパラメータは次数選択基準に従って最適な値が設定されます。（デフォルトの基準は AIC：赤池情報量基準になります）

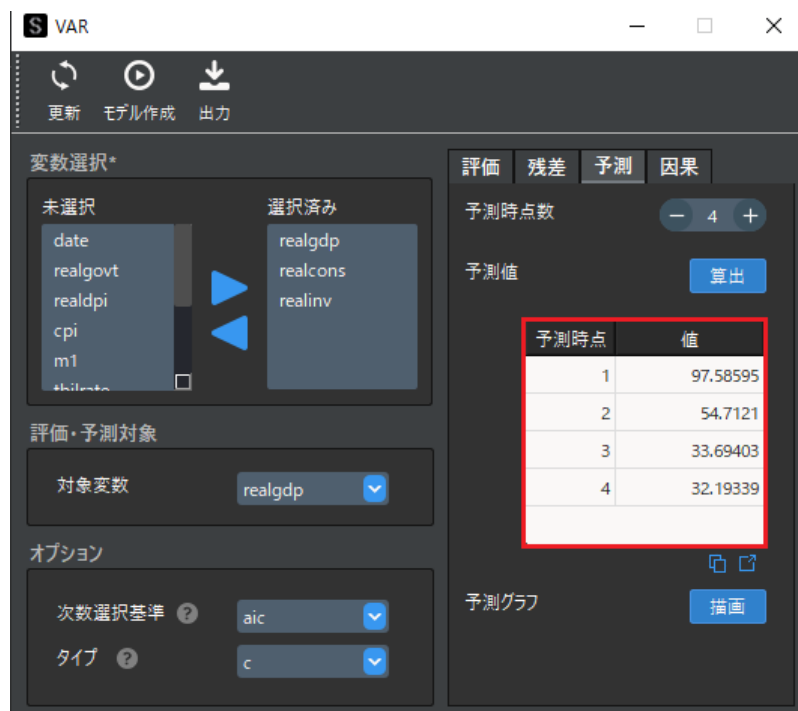
算出された評価指標を用いてモデルの評価を行います。VAR 機能で算出される評価指標は全て絶対的な基準はありません。各指標とも値が小さいほど、良いモデルであると考えられます。

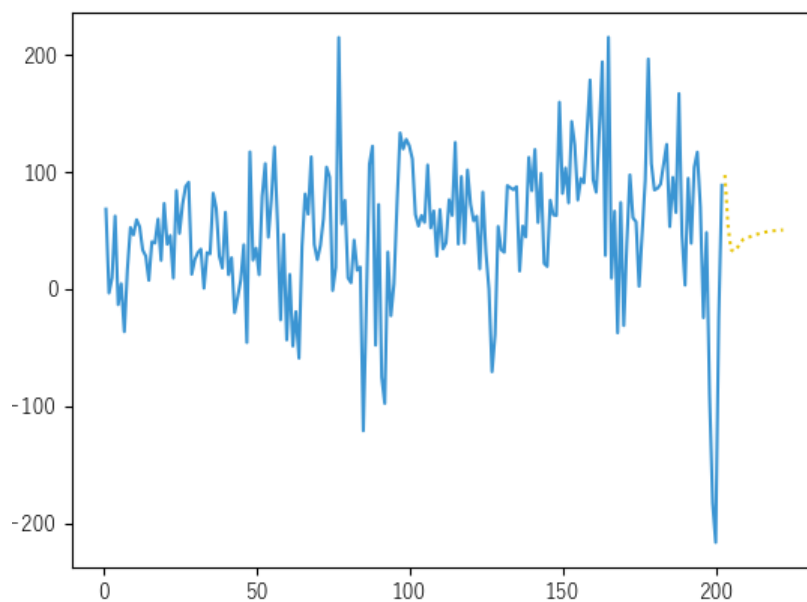


VAR モデル機能では ARIMA 機能と同様にホールドアウト検証や残差分析も行うことが可能です。

② 将来予測

作成したモデルから将来予測を行います。予測タブから任意の時点数の選択して、将来予測を行います。



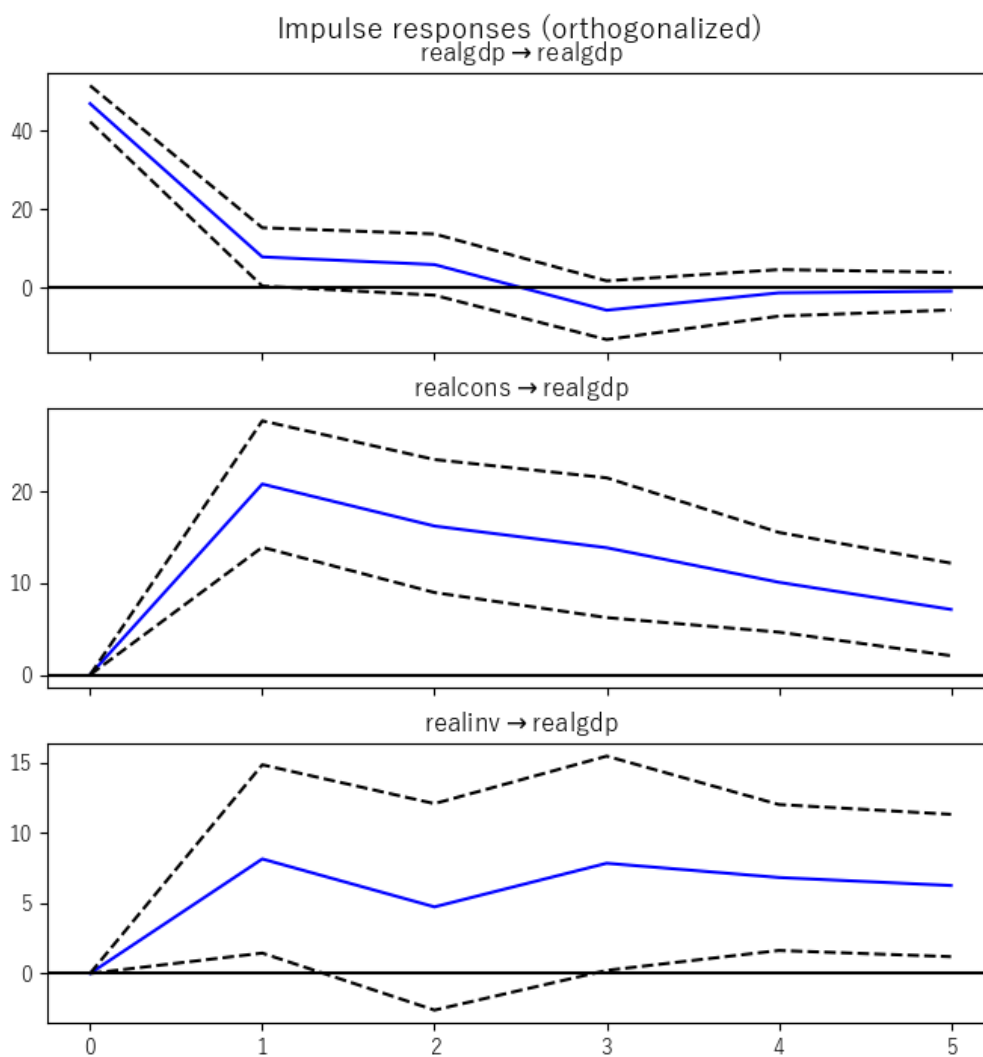


③ 因果分析

「因果」タブでは VAR モデルを用いて、3つの手法で因果関係の分析が可能です。

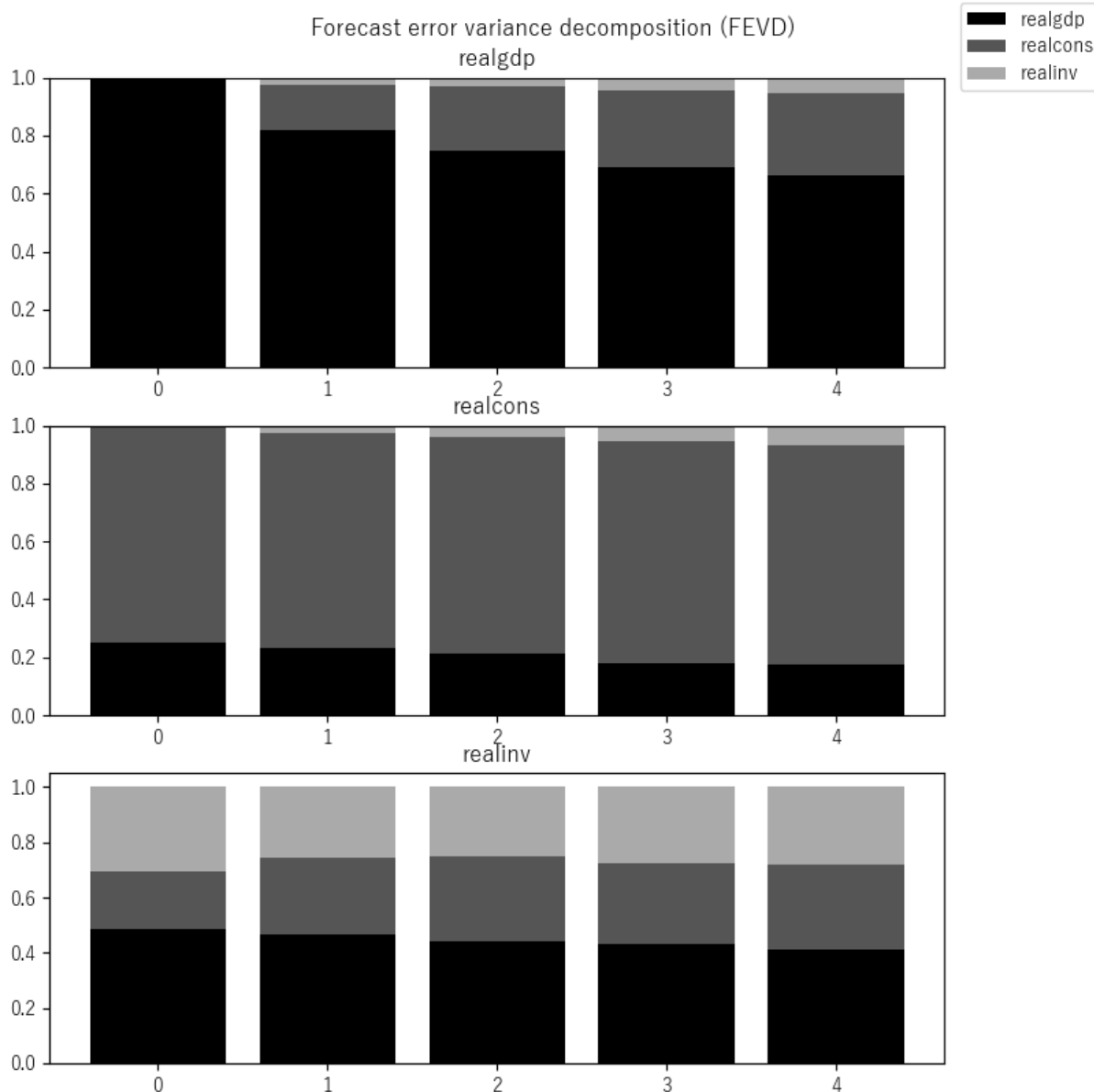
インパルス応答関数のグラフを表示します。インパルス応答関数は、一つの時系列変数に対する"衝撃"が他の時系列変数に与える効果を時間経過に従って追跡するための分析手法です。

VAR 機能の設定で「評価・予測対象」に"realgdp"を設定しているため、"realgdp"に対して与える影響がグラフで表示されます。



インパルス応答分析の結果から，“realcons”は“realinv”と比較して“realgdp”を増加させる影響が大きいということがわかります。

続いてVEFD（予測誤差分散分解）を行い，グラフを表示します．VEFDは各変数が各変数に対する相対的な影響力を示します．



“realgdp”と“realcons”は自身への影響が最も大きいことがわかります。“realinv”は“realgdp”への影響が最も大きいことがわかります。

最後にグレンジャー因果性検定を行い“realgdp”に対する“realcons”の“realinv”の影響を調べます。

“realcons”の p 値が 0.0 で有意水準 $\alpha=0.05$ より小さいことから「realcons は realgdp に対してグレンジャー因果性がある」といえることが言えます。

※ 表示桁数設定で、小数点以下の表示桁数を増やすことも可能です。

VAR

更新 モデル作成 出力

変数選択*

未選択
date
realgovt
realdpi
cpi
m1
skilrate

選択済み
realgdp
realcons
realinv

評価・予測対象
対象変数 realgdp

オプション
次数選択基準 aic
タイプ c

評価 残差 予測 因果

分析手法 解析

インパルス応答分析
 直交化

予測誤差分散分解(FEVD)

グレンジャー-因果性検定

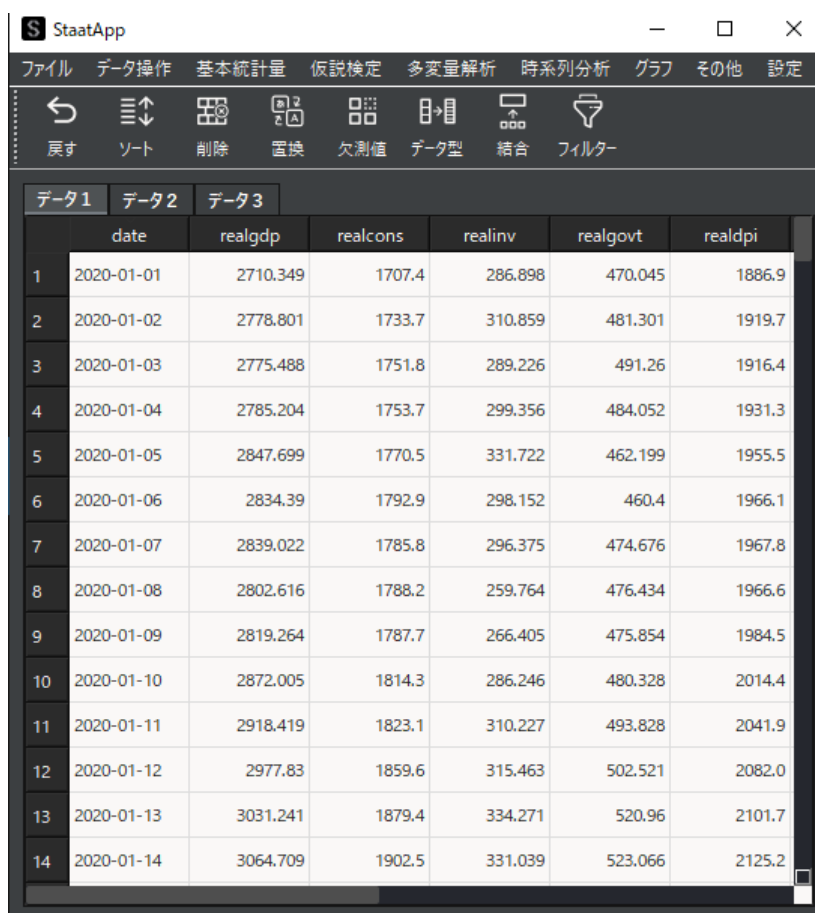
	p値
realcons	0.0
realinv	0.0745

オプション
グラフサイズ (横, 縦) (6, 6)
予測時点数 5

14. Prophet

※ Prophet はサブスクリプション版限定の機能です。

以下のサンプルデータ（マクロ経済学データセット macrodata）を用いて解説します。



The screenshot shows the StaatApp interface with a menu bar and a toolbar. The main area displays a table with the following data:

	データ1	データ2	データ3			
	date	realgdp	realcons	realinv	realgovt	realdpi
1	2020-01-01	2710.349	1707.4	286.898	470.045	1886.9
2	2020-01-02	2778.801	1733.7	310.859	481.301	1919.7
3	2020-01-03	2775.488	1751.8	289.226	491.26	1916.4
4	2020-01-04	2785.204	1753.7	299.356	484.052	1931.3
5	2020-01-05	2847.699	1770.5	331.722	462.199	1955.5
6	2020-01-06	2834.39	1792.9	298.152	460.4	1966.1
7	2020-01-07	2839.022	1785.8	296.375	474.676	1967.8
8	2020-01-08	2802.616	1788.2	259.764	476.434	1966.6
9	2020-01-09	2819.264	1787.7	266.405	475.854	1984.5
10	2020-01-10	2872.005	1814.3	286.246	480.328	2014.4
11	2020-01-11	2918.419	1823.1	310.227	493.828	2041.9
12	2020-01-12	2977.83	1859.6	315.463	502.521	2082.0
13	2020-01-13	3031.241	1879.4	334.271	520.96	2101.7
14	2020-01-14	3064.709	1902.5	331.039	523.066	2125.2

① 前処理（日時データの作成）

サンプルデータには Prophet で使える形式の日時データが含まれていないため、日時データの作成を行います。また、Prophet では将来予測を行う場合の単位が、1日もしくは1時間になります。サンプルデータは1959年の第一四半期から四半期ごとのデータですが、将来予測を適切に行うために意図的に以下のように設定を行い、日時データを作成します。



日時データ作成機能では、基準日時から任意の間隔の日時データを対象のアクティブデータ行数分作成して、アクティブデータセットの最初の列に追加します。

因みに Prophet で推奨されている日時データの形式は、"YYYY-MM-DD"もしくは"YYYY-MM-DD hh:mm:ss"になります。

② モデルの作成・評価

Prophet 機能でモデル作成を行います。「予測対象データ」に予測を行いたい変数を、「時間データ」に日時データを選択します。日時データは①で作成したデータを選択しています。

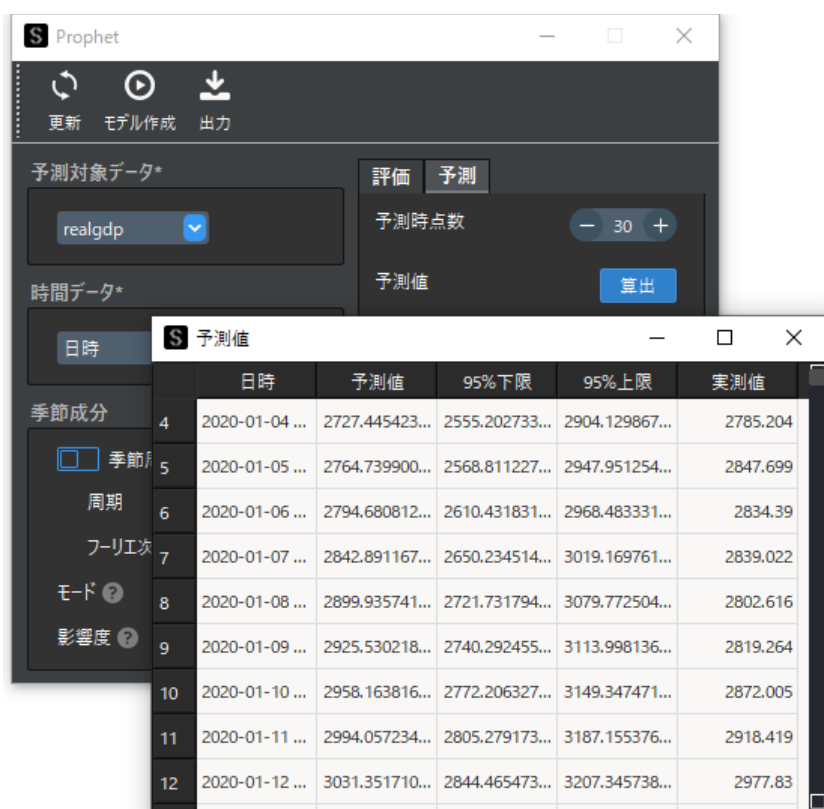
「モデル作成」ボタンをクリックするとモデルが作成されて、モデルの評価指標 (MAE・MSE) が表示されます。



Prophet 機能では ARIMA 機能と同様にホールドアウト検証や残差分析も行うことが可能です。

③ 将来予測

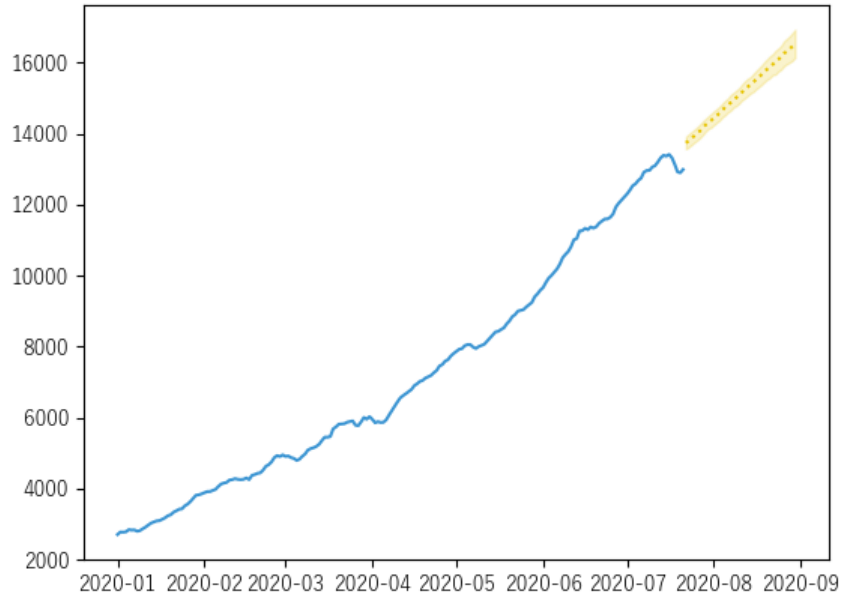
「予測」タブで将来予測を行います。任意の予測時点数を選択して、「算出」をクリックすると予測値が表示されます。



The screenshot shows the Prophet application interface. The main window has a top bar with '更新' (Refresh), 'モデル作成' (Model Creation), and '出力' (Output) buttons. Below this, there are controls for '予測対象データ*' (Prediction Target Data) set to 'realgdp', '時間データ*' (Time Data), and '予測時点数' (Prediction Time Points) set to 30. A '算出' (Calculate) button is visible. A secondary window titled 'S 予測値' (S Prediction Values) is overlaid, displaying a table of prediction results.

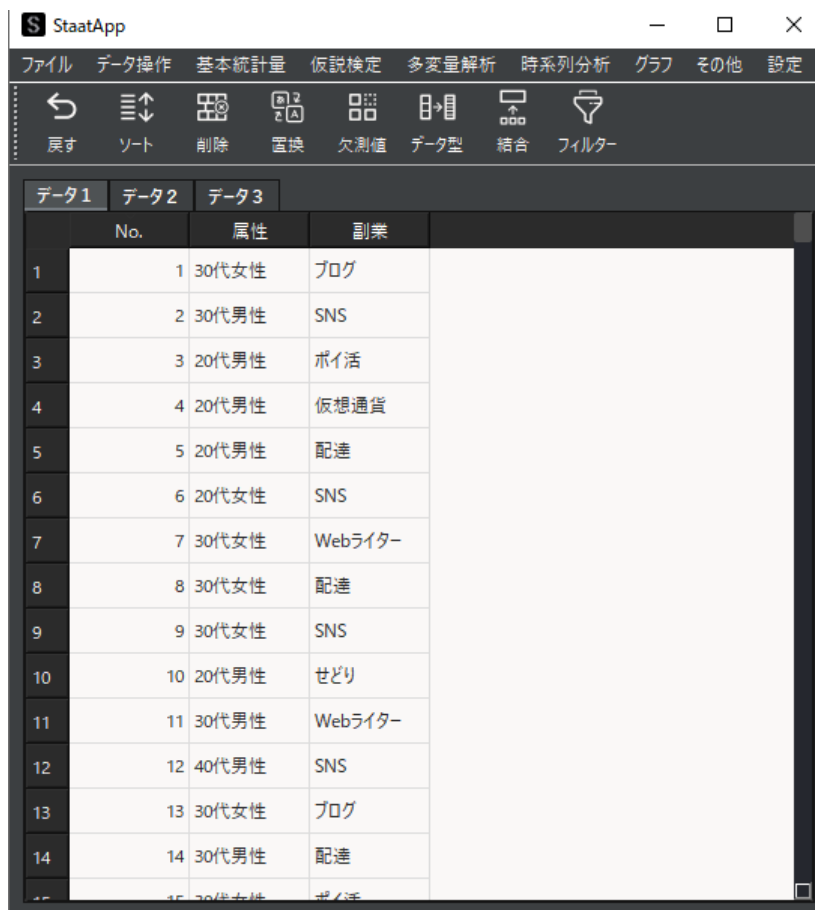
	日時	予測値	95%下限	95%上限	実測値
4	2020-01-04 ...	2727.445423...	2555.202733...	2904.129867...	2785.204
5	2020-01-05 ...	2764.739900...	2568.811227...	2947.951254...	2847.699
6	2020-01-06 ...	2794.680812...	2610.431831...	2968.483331...	2834.39
7	2020-01-07 ...	2842.891167...	2650.234514...	3019.169761...	2839.022
8	2020-01-08 ...	2899.935741...	2721.731794...	3079.772504...	2802.616
9	2020-01-09 ...	2925.530218...	2740.292455...	3113.998136...	2819.264
10	2020-01-10 ...	2958.163816...	2772.206327...	3149.347471...	2872.005
11	2020-01-11 ...	2994.057234...	2805.279173...	3187.155376...	2918.419
12	2020-01-12 ...	3031.351710...	2844.465473...	3207.345738...	2977.83

「描画」ボタンをクリックすると将来予測値を含むグラフが表示されます。



15. クロス集計表

以下のサンプルデータを用いて解説します。



The screenshot shows the StaatApp interface with a data table. The table has three columns: 'No.', '属性', and '副業'. The data rows are as follows:

データ1	データ2	データ3
No.	属性	副業
1	30代女性	ブログ
2	30代男性	SNS
3	20代男性	ポイ活
4	20代男性	仮想通貨
5	20代男性	配達
6	20代女性	SNS
7	30代女性	Webライター
8	30代女性	配達
9	30代女性	SNS
10	20代男性	せどり
11	30代男性	Webライター
12	40代男性	SNS
13	30代女性	ブログ
14	30代男性	配達
15	30代女性	ポイ活

① クロス集計

メニューバーから「その他」→「クロス集計表」を選択します。

クロス集計表用のウィンドウを表示したら、表側と表頭の項目の設定を行います。例では表側に"属性"の列を、表頭に"副業"の列を設定しています。

設定が完了したら、ツールバーの「集計実行」ボタンをクリックすることで以下のようにクロス集計表が作成されます。

クロス集計表

ファイル 分析

更新 集計実行 出力 自動考察

クロス集計設定

表側 (行項目) 属性
表頭 (列項目) 副業

クロス集計表

	SNS	Webライター	せどり	ブログ	ポイ活	仮想通貨	動画編集	配達	合計
20代女性	8	8	3	13	7	2	3	2	46
20代男性	9	5	6	7	5	13	12	24	81
30代女性	7	10	4	21	12	3	1	3	61
30代男性	6	7	3	6	10	11	8	10	61
40代女性	3	4	1	12	5	0	0	0	25
40代男性	2	3	2	4	4	6	0	16	37
合計	35	37	19	63	43	35	24	55	311

② カイ二乗検定

クロス集計表ウィンドウのメニューバー「分析」→「カイ二乗検定」を選択します。

カイ二乗検定用のウィンドウが表示されたらツールバーの「解析実行」ボタンをクリックします。以下のように検定統計量と p 値、クラメールの連関係数（効果量）が表示されます。

カイ二乗検定

解析実行

解析結果

検定統計量	0.39773	p値	0.52827
クラメールの連関係数	0.13762		

※ 画像はサンプルデータとは関係がない、算出例です。

③ フィッシャーの正確確率検定

フィッシャーの正確確率検定はカイ二乗検定と同じ手順で実行可能です。フィッシャーの正確確率検定は2×2クロス集計表に対してのみ実行できるので注意してください。今回紹介したようなクロス集計表に対しては実行できません。



※ 画像はサンプルデータとは関係がない、算出例です。

④ オッズ比とリスク比

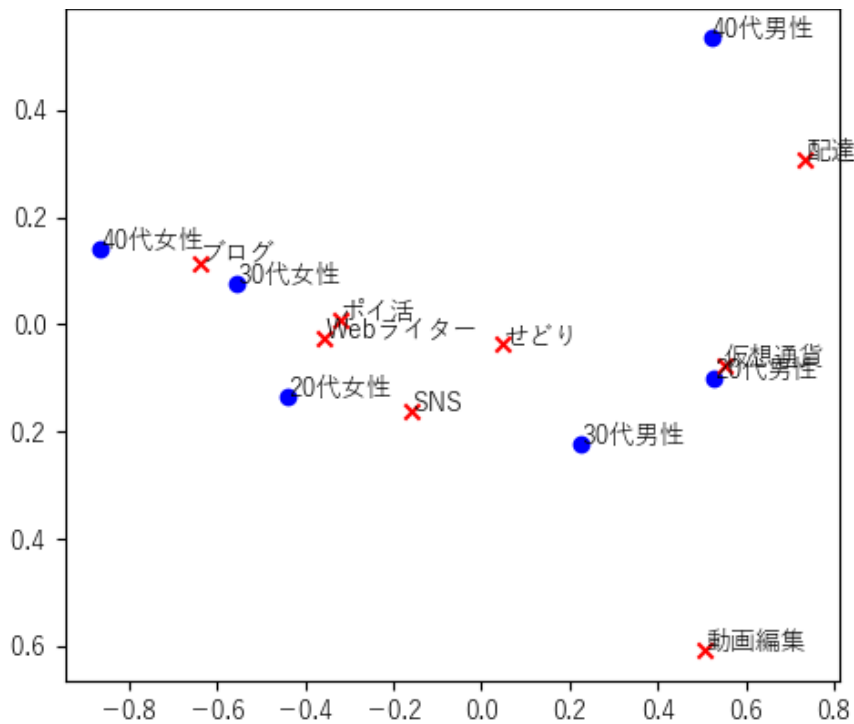
オッズ比とリスク比も2×2クロス集計表に対してのみ実行可能です。解析結果には、95%信頼区間（95%CI 下限，95%CI 上限）も算出されます。

	値	95%CI 下限	95%CI 上限
オッズ比	0.57143	0.09978	3.2726
対数オッズ比	-0.55962	-2.30482	1.18559
リスク比	0.7619	0.31823	1.82413
対数リスク比	-0.27193	-1.14497	0.6011

⑤ コレスポネンス分析

コレスポネンス分析を実行する場合は、クロス集計表ウィンドウのメニューバーから「分析」→「コレスポネンス分析」を選択します。

以下のように項目間の類似度を示した散布図が表示されます。



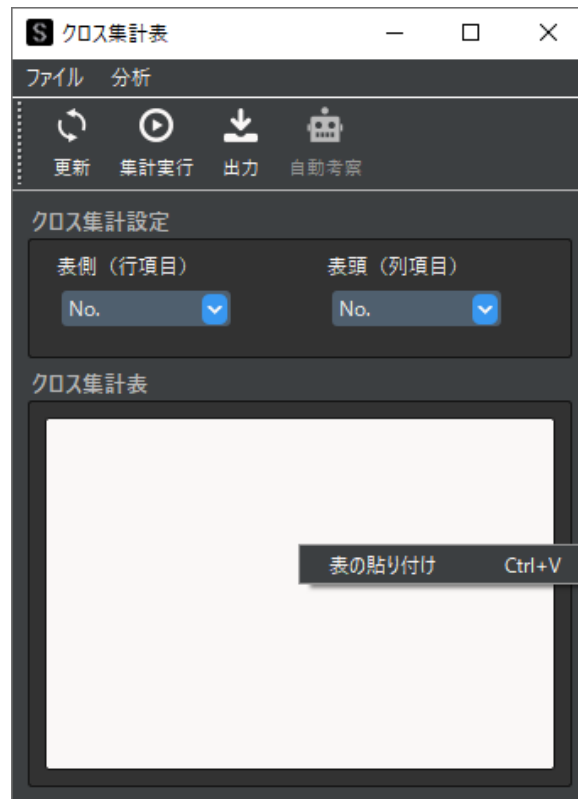
⑥ クロス集計表の貼り付け（応用）

クロス集計を行わずに、既に集計されたクロス集計表について分析を行いたい場合は、表計算ソフトの対象のセルをコピーして、クロス集計表ウィンドウに貼り付けます。

Excel のようは表計算ソフトの対象セルを選択してコピーします。

		副業		合計
		Yes	No	
性別	男性	18	15	33
	女性	8	24	32
合計		26	39	65

貼り付けは右クリックメニューもしくは、「Ctrl+V」で行なうことができます。



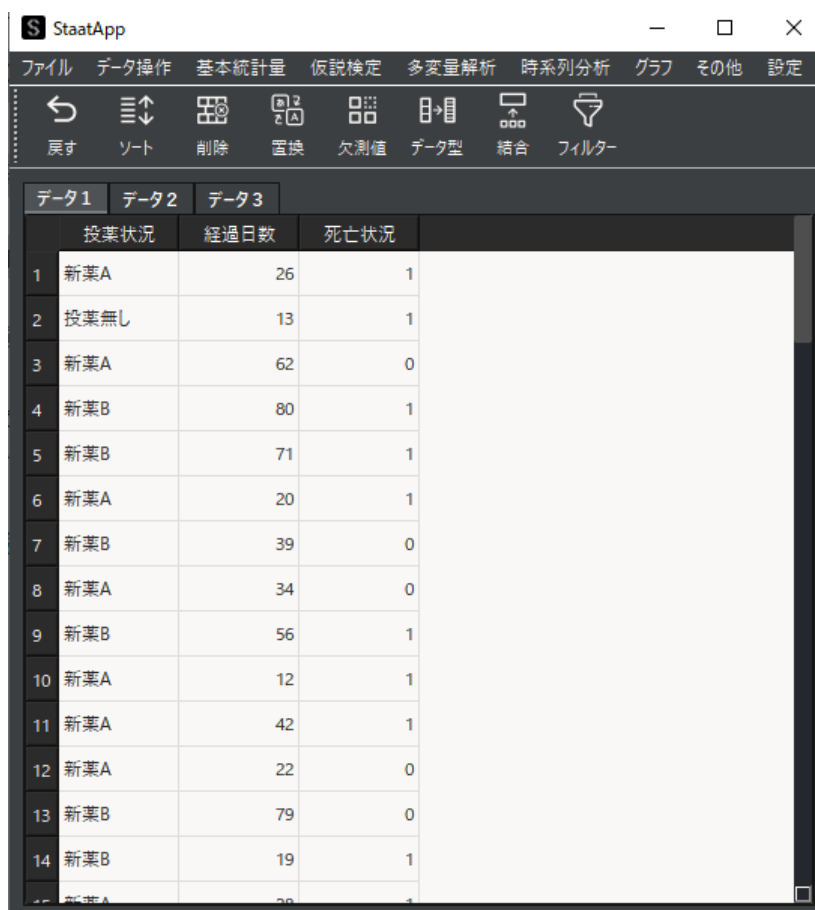
クロス集計表を貼り付ける際の注意点として、必ず“合計”を含むクロス集計表をコピー&ペーストしてください。

“合計”を含まないクロス集計表の場合、クロス集計表の分析で形式エラーが発生します。

16. 生存曲線

以下のサンプルデータを用いて解説します。

生存曲線を作成するためには、イベント発生までの経過時間を表すデータ（経過日数）と、イベント発生の有無を示すデータ（死亡状況）が必要です。イベント発生の有無を示すデータは、イベント発生した場合→"1"，打ち切りの場合"0"とします。



データ1	データ2	データ3
投薬状況	経過日数	死亡状況
1 新薬A	26	1
2 投薬無し	13	1
3 新薬A	62	0
4 新薬B	80	1
5 新薬B	71	1
6 新薬A	20	1
7 新薬B	39	0
8 新薬A	34	0
9 新薬B	56	1
10 新薬A	12	1
11 新薬A	42	1
12 新薬A	22	0
13 新薬B	79	0
14 新薬B	19	1
15 新薬A	28	1

① 生存曲線の作成

メニューバーから「その他」→「生存時間解析」→「生存曲線」を選択します。生存曲線用のウィンドウが表示されたら、「経過時間データ」と「イベント発生データ」として設定する変数をコンボボックスから選択します。

S 生存曲線
×

↻ 更新
▶ 解析実行

変数列選択*

経過時間
イベント発生

経過日数 ▼
死亡状況 ▼

群名列

1群 ▼

ログランク検定の結果

検定統計量

p値

一般化ウィルコクソン検定の結果

検定統計量

p値

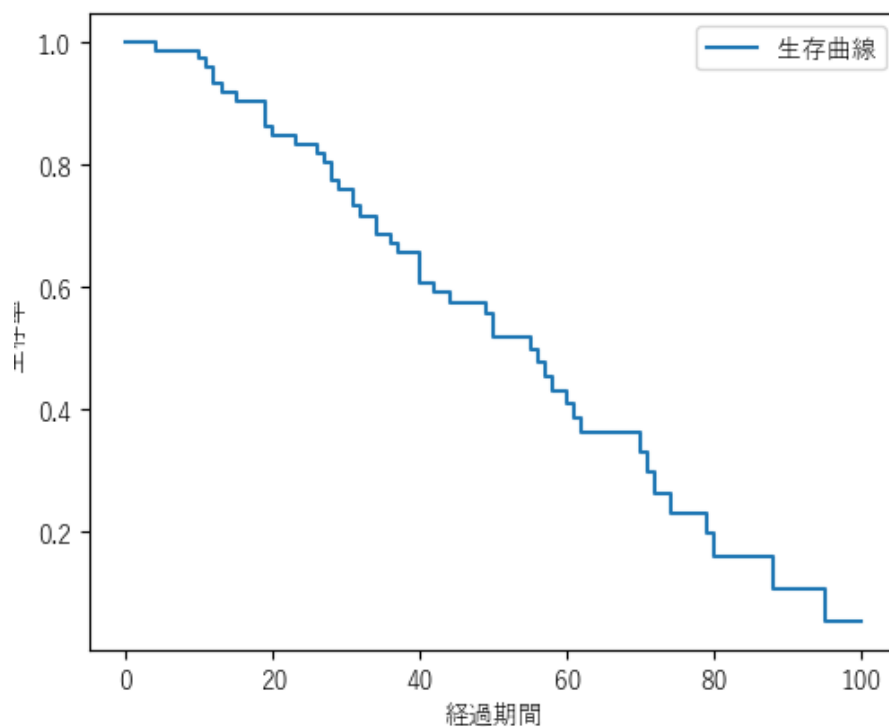
グラフオプション

打ち切りデータ

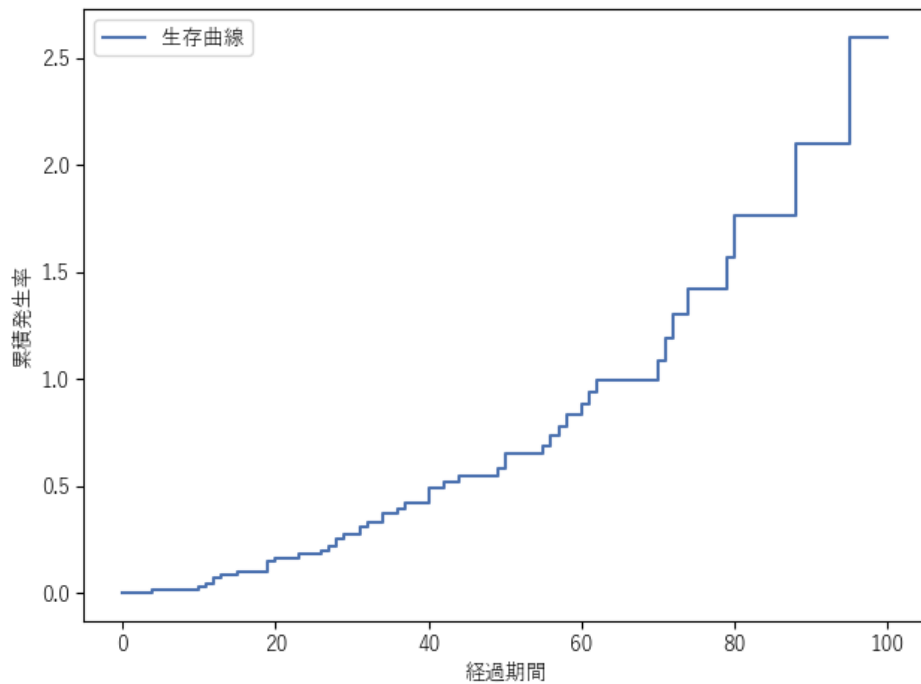
95%信頼区間

累積発生率

生存曲線に打ち切りデータ，95%信頼区間を表示する場合はチェックボックスをクリックします．設定が完了したらツールバーの「解析実行」ボタンをクリックして生存曲線を作成します．以下のようにグラフが表示されます．



打ち切りデータと 95%信頼区間を表示する場合は以下のようなグラフになります．



累積発生率を選択した場合は、以下のようなグラフが作成されます。

② 生存曲線の検定

2つ以上の生存曲線を比較する方法として、ログランク検定と一般化ウィルコクソン検定が可能です。2群の比較を行う場合は「群名列」のコンボボックスで、属する群を示すカテゴリーデータが入力された列を選択します。

「解析実行」ボタンをクリックすると、画面右側に検定結果と生存曲線が作成されます。

🔍 生存曲線
✕

🔄 更新
▶ 解析実行

変数列選択*

経過時間: 経過日数 ▼

イベント発生: 死亡状況 ▼

群名列: 投薬状況 ▼

ログランク検定の結果

検定統計量: 9.18945

p値: 0.01011

グラフオプション

打ち切りデータ:

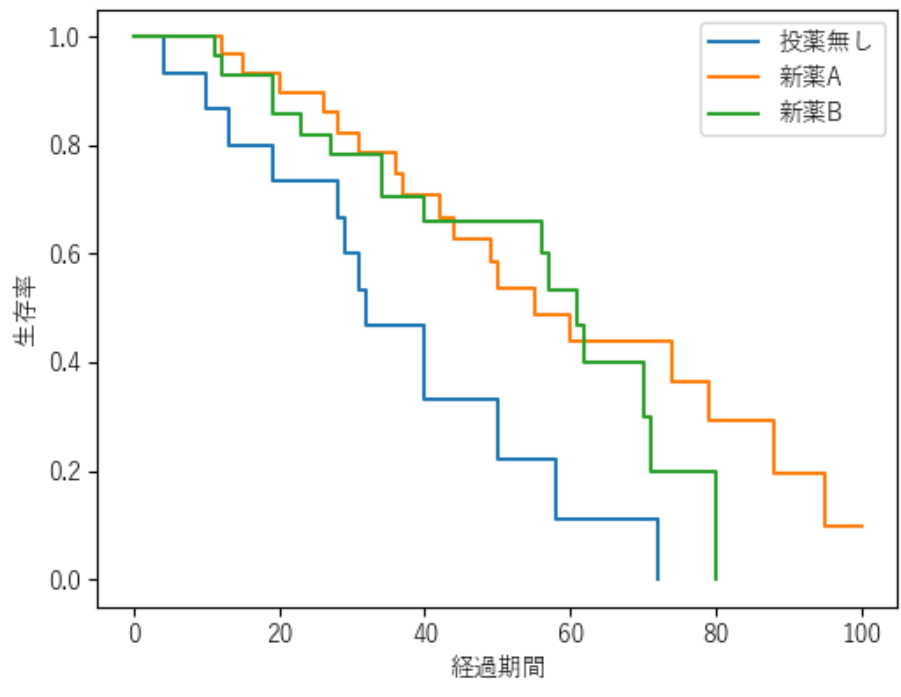
95%信頼区間:

累積発生率:

一般化ウィルコクソン検定の結果

検定統計量: 7.09234

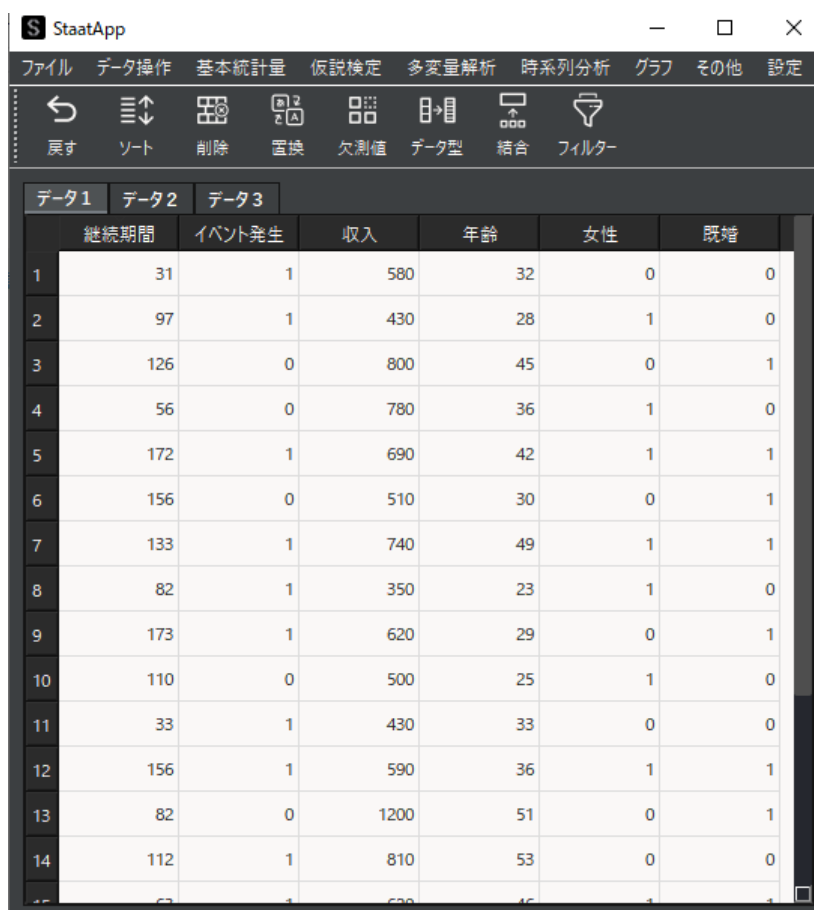
p値: 0.02883



17. Cox 比例ハザード回帰

以下のサンプルデータを用いて解説します。

イベント発生の有無を示すデータは、イベント発生した場合→"1"、発生していない場合"0"とします。



The screenshot shows the StaatApp software interface. The main window displays a data table with the following columns: 継続期間 (Continuation Period), イベント発生 (Event Occurrence), 収入 (Income), 年齢 (Age), 女性 (Female), and 既婚 (Married). The data rows are numbered 1 through 14.

データ1	データ2	データ3				
	継続期間	イベント発生	収入	年齢	女性	既婚
1	31	1	580	32	0	0
2	97	1	430	28	1	0
3	126	0	800	45	0	1
4	56	0	780	36	1	0
5	172	1	690	42	1	1
6	156	0	510	30	0	1
7	133	1	740	49	1	1
8	82	1	350	23	1	0
9	173	1	620	29	0	1
10	110	0	500	25	1	0
11	33	1	430	33	0	0
12	156	1	590	36	1	1
13	82	0	1200	51	0	1
14	112	1	810	53	0	0

① Cox 比例ハザード回帰

メニューバーから「その他」→「生存時間解析」→「Cox 比例ハザード回帰」を選択します。

Cox 比例ハザード回帰用のウィンドウが表示されたら、目的変数のコンボボックスからの経過時間データとイベント発生データとする変数名（列名）を選択します。

説明変数は未選択欄に表示されている変数名をクリックして選択します。ダミー変数を選択する際は、多重共線性の問題を回避するために 1 列分を除いて選択します。

設定が完了したらツールバーの「解析実行」ボタンをクリックして解析を実行します。実行結果は以下のように画面右側に表示されます。

S Cox比例ハザード回帰

更新 解析実行 出力 自動考察

目的変数*

経過時間データ 継続期間

イベント発生データ イベント発生

説明変数 (共変量) *

未選択 No.

選択済み 収入 年齢 女性 既婚

結果 ハザード比 予測

統計量

C Index	0.86029
対数尤度	-20.93476
Partial AIC	49.86952

	偏回帰係数	ハザード比	標準誤差
収入	-0.00892	0.99112	0.00439
年齢	0.13284	1.14207	0.06459
女性	0.12983	1.13863	0.63077
既婚	-2.34108	0.09622	1.13923

② ハザード比の検証

ハザード比タブを選択すると、変数ごとのハザード比の検定統計量と p 値を確認することが可能です

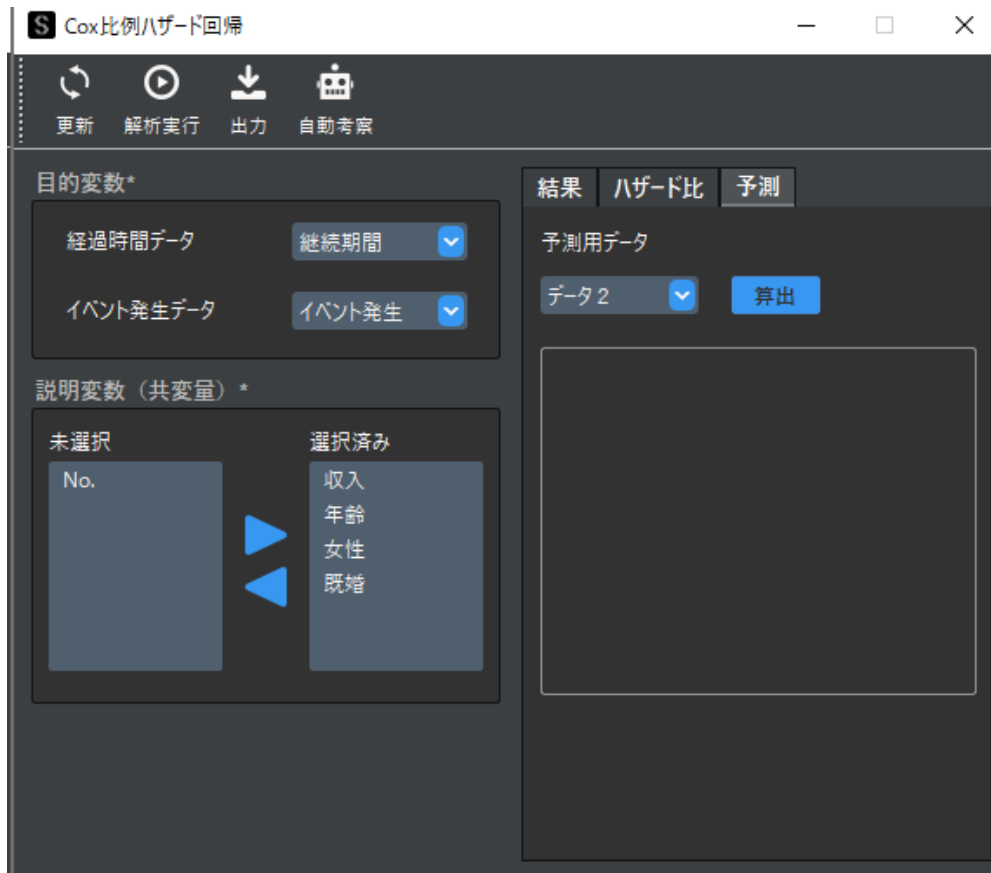
	検定統計量	p値
収入	1.95836	0.16169
女性	0.88945	0.34563
年齢	1.435	0.23095
既婚	0.77166	0.3797

③ 予測値の算出（応用）

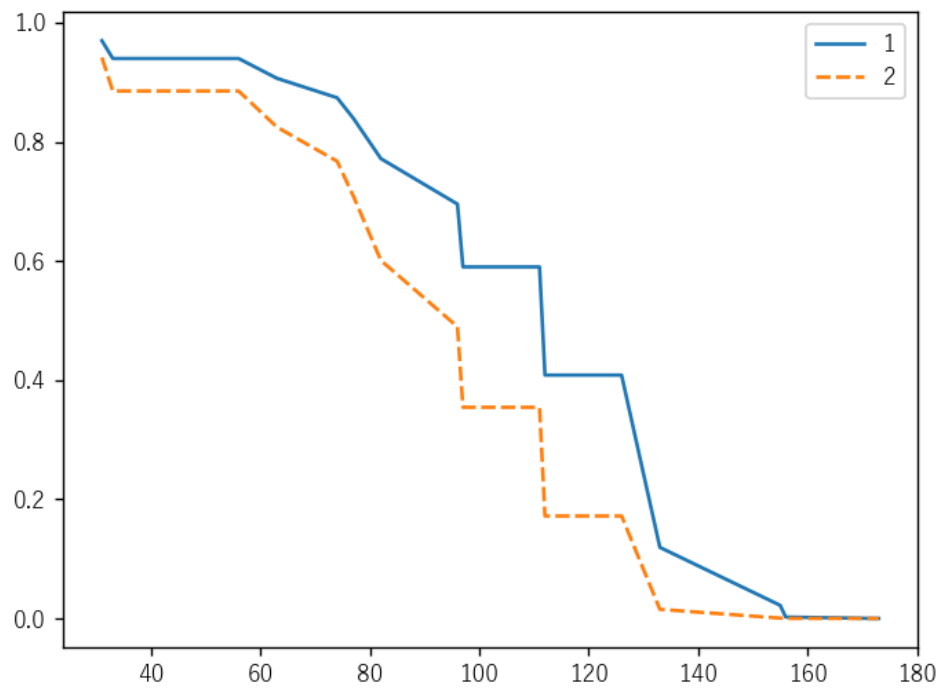
Cox 比例ハザード回帰で作成したモデルと予測用データを用いて、予測値の算出・生存曲線の予測を行います。

予測用データは入力データと同様に Excel などを用いて以下のように作成します。列名は解析実行時に選択した列名（変数名）と一致させる必要があります（順不同）。

予測用データ StaatApp に取り込み、予測用データを保存したデータを選択します。以下の例では、「データ 2」に予測用データを取り込んでいます。




「算出」ボタンをクリックすると、以下のように生存曲線が作成・表示されます。



18. アソシエーション分析

以下のサンプルデータ (n = 1000) を用いて解説します。



The screenshot shows the StaatApp interface with a data table. The table has columns labeled 'データ1' through 'データ3' and '0' through '7'. The data rows contain various items and their relationships, with many cells containing 'nan'.

	データ1	データ2	データ3	0	1	2	3	4	5	6	7
1	投資	動画編集	SNS	せどり	nan	nan	nan	nan	nan	nan	nan
2	ライブ配信	ブログ	nan	nan	nan	nan	nan	nan	nan	nan	nan
3	投資	ポイ活	動画編集	せどり	ブログ	nan	nan	nan	nan	nan	nan
4	ブログ	WEBライティ...	動画編集	nan	nan	nan	nan	nan	nan	nan	nan
5	フードデリバリー	投資	せどり	nan	nan	nan	nan	nan	nan	nan	nan
6	ポイ活	フードデリバリー	ブログ	nan	nan	nan	nan	nan	nan	nan	nan
7	せどり	投資	nan	nan	nan	nan	nan	nan	nan	nan	nan
8	WEBライティ...	ブログ	フードデリバリー	nan	nan	nan	nan	nan	nan	nan	nan
9	ポイ活	ブログ	SNS	nan	nan	nan	nan	nan	nan	nan	nan
10	ブログ	ライブ配信	投資	ポイ活	SNS	nan	nan	nan	nan	nan	nan
11	投資	せどり	nan	nan	nan	nan	nan	nan	nan	nan	nan
12	フードデリバリー	せどり	投資	WEBライティ...	動画編集	nan	nan	nan	nan	nan	nan
13	ポイ活	動画編集	投資	せどり	nan	nan	nan	nan	nan	nan	nan
14	投資	動画編集	せどり	ポイ活	フードデリバリー	nan	nan	nan	nan	nan	nan
15	投資	ポイ活	nan	nan	nan	nan	nan	nan	nan	nan	nan

アソシエーション分析を行う場合は1つのサンプル・トランザクションを1レコードとして、アイテム・商品は列ごとに入力してください。

① アソシエーションルールの算出

メニューバーから「その他」→「アソシエーション分析」を選択してアソシエーション分析用ウィンドウを表示します。

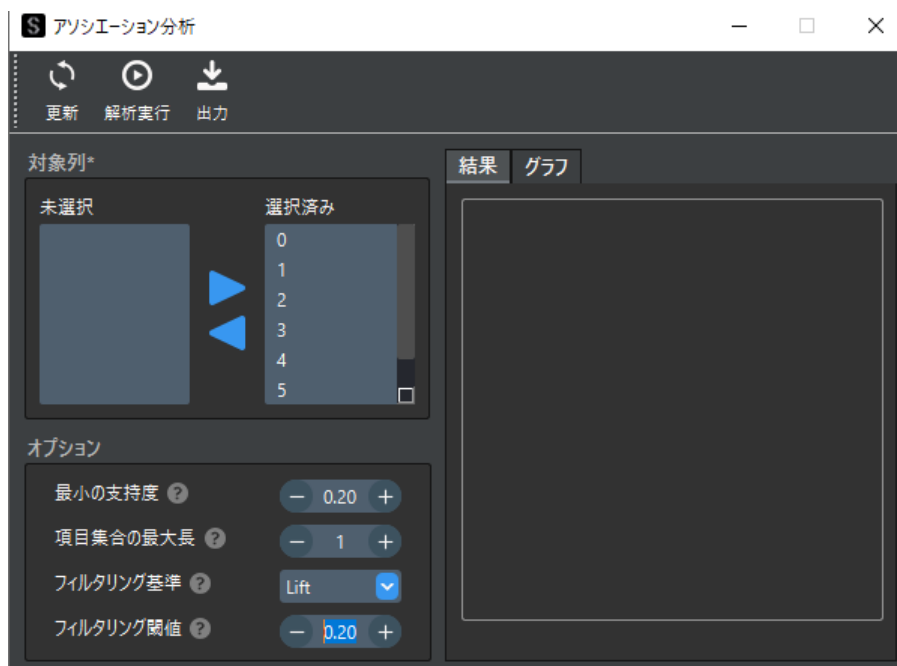
アソシエーション分析用ウィンドウが表示されたら、アソシエーション分析の対象とする列を選択します。

オプションではアソシエーション分析で抽出するアイテムセットに対して、しきい値やアイテムセットに含めるアイテム数を設定できます。

オプション

- **最小の支持度**：設定した値の支持度以上のアイテムセットが抽出されます。値が小さいほど許容する支持度が小さくなるので、より多くのアイテムセットが抽出されます。
- **項目集合の最大長**：アイテムセットに含めるアイテム数に最大値になります。1の場合は1対1のアイテムセットが抽出されます。3と設定した場合、例えば"SNS"と"WEB ライティング・せどり"といったアイテムセットも抽出されるようになります。
- **フィルタリング基準・フィルタリング閾値**：最小の支持度とは別に、アイテムセットを抽出する条件を設定します。フィルタリング閾値に設定したアソシエーションルールとフィルタリング閾値を基準に、アイテムセットを抽出します。

分析例では、以下のように設定します。



「解析実行」ボタンをクリックすると、アソシエーション分析の解析結果が出力されます。

解析結果はアイテムセットごとに、代表的なアソシエーションルールの算出結果が表示されます。

	始点	終点	始点の支持度	終点の支持度	支持度	信頼度	Lift	Leverage	Conviction
1	SNS	投資	0.306	0.661	0.207	0.67647	1.0234	0.00473	1.04782
2	投資	SNS	0.661	0.306	0.207	0.31316	1.0234	0.00473	1.01043
3	ブログ	WEBライティ...	0.547	0.316	0.229	0.41865	1.32483	0.05615	1.17657
4	WEBライティ...	ブログ	0.316	0.547	0.229	0.72468	1.32483	0.05615	1.64538
5	ブログ	せどり	0.547	0.525	0.24	0.43876	0.83573	-0.04718	0.84634
6	せどり	ブログ	0.525	0.547	0.24	0.45714	0.83573	-0.04718	0.83447
7	ポイ活	せどり	0.433	0.525	0.275	0.6351	1.20972	0.04768	1.30174
8	せどり	ポイ活	0.525	0.433	0.275	0.52381	1.20972	0.04768	1.1907
9	動画編集	せどり	0.412	0.525	0.209	0.50728	0.96625	-0.0073	0.96404
10	せどり	動画編集	0.525	0.412	0.209	0.3981	0.96625	-0.0073	0.9769
11	投資	せどり	0.661	0.525	0.395	0.59758	1.13825	0.04797	1.18036
12	せどり	投資	0.525	0.661	0.395	0.75238	1.13825	0.04797	1.36904
13	ブログ	ポイ活	0.547	0.433	0.206	0.3766	0.86975	-0.03085	0.90953

② 相対出現頻度（応用）

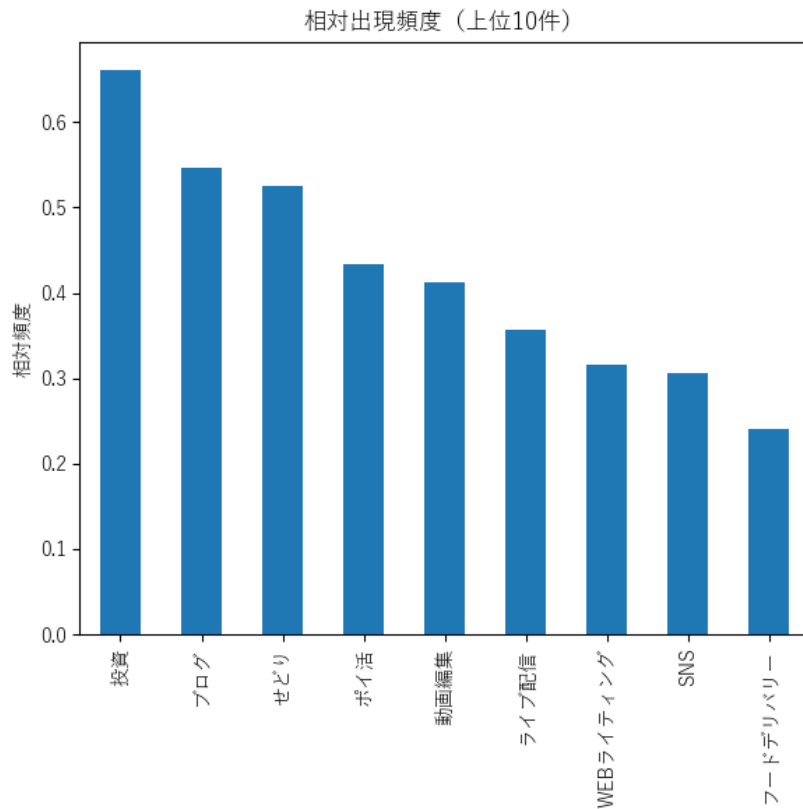
「グラフ」タブを表示して、「グラフの種類」を相対出現頻度に設定し、「作成」ボタンをクリックすることで以下のようにアイテムの相対出現頻度を示した、棒グラフが表示されます。

The screenshot shows the 'アソシエーション分析' (Association Analysis) window. At the top, there are icons for '更新' (Refresh), '解析実行' (Execute Analysis), and '出力' (Output). Below this, the '対象列*' (Target Columns) section has '未選択' (Not Selected) and '選択済み' (Selected) areas. The 'オプション' (Options) section includes:

- 最小の支持度 (Minimum Support): 0.20
- 項目集合の最大長 (Maximum Length of Item Sets): 1
- フィルタリング基準 (Filtering Criterion): Lift
- フィルタリング閾値 (Filtering Threshold): 0.20

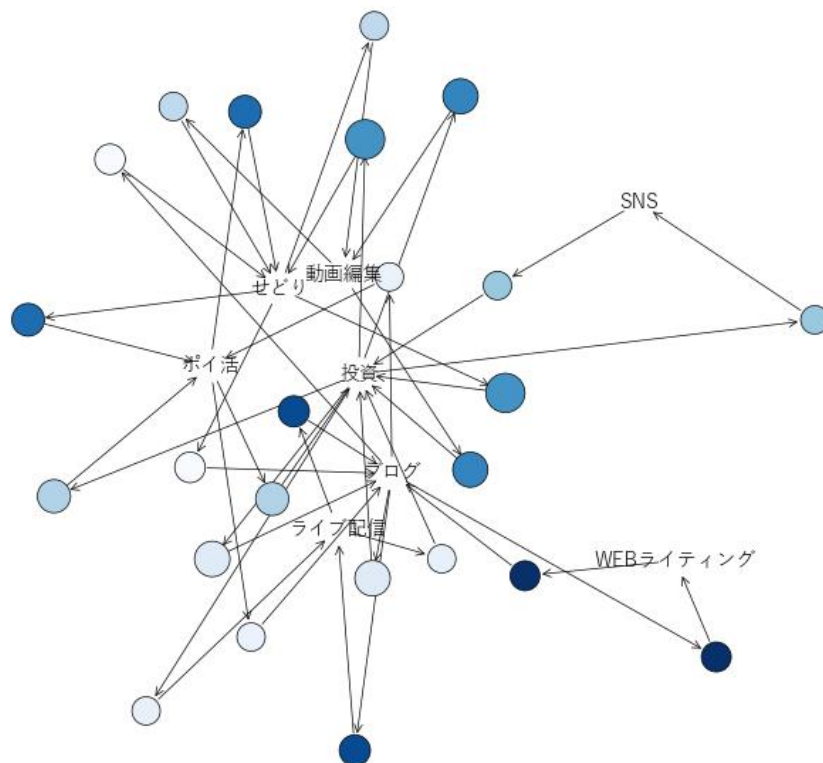
 The '結果' (Result) tab is selected, showing 'グラフの種類' (Graph Type) set to '相対出現頻度' (Relative Frequency) and a '作成' (Create) button. Below this, the 'オプション' (Options) for the graph are shown:

- 相対出現頻度 (Relative Frequency): 表示数の上限 (Upper Limit of Number of Displays) set to 10.
- ネットワーク図 (Network Graph): 中間ノードの濃淡 (Intermediate Node Density) set to '信頼度' (Trust), and ラベルサイズ (Label Size) set to 10.
- 共通 (Common): グラフサイズ (横, 縦) (Graph Size (Width, Height)) set to (6, 6).



③ ネットワーク図（応用）

ネットワーク図を作成する場合は、「グラフの種類」をネットワーク図に設定し、「作成」ボタンをクリックします。



矢印はアソシエーションルールの方向性を示し、アイテムセット間の円の大きさはそのアイテムセットの支持度の大きさ、円の濃淡はグラフ作成時に設定したアイテムルールの大きさを示します。

円の濃淡が濃いほど、設定したアイテムルールの値が大きいことを意味します。アイテムルールは信頼度・リフト値・Conviction・Leverage から選択することが可能です。

上記の例ではアイテムルールをリフト値に設定しているので、"WEB ライティング"は"ブログ"を選択した場合に選択される可能性が高いということが読み取れます。

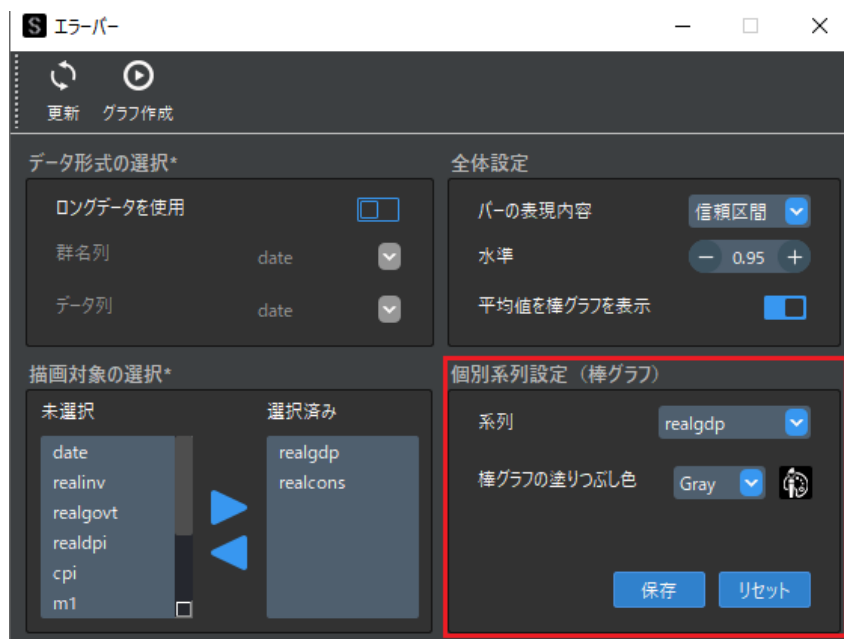
グラフ作成

グラフ作成機能を用いて、グラフを作成する方法を説明します。

1. グラフ作成機能の共通操作（個別系列設定）

グラフ作成機能において、デフォルト設定からグラフの色などを変更したい場合は、「個別系列設定」で対象の変数ごとに設定を行います。

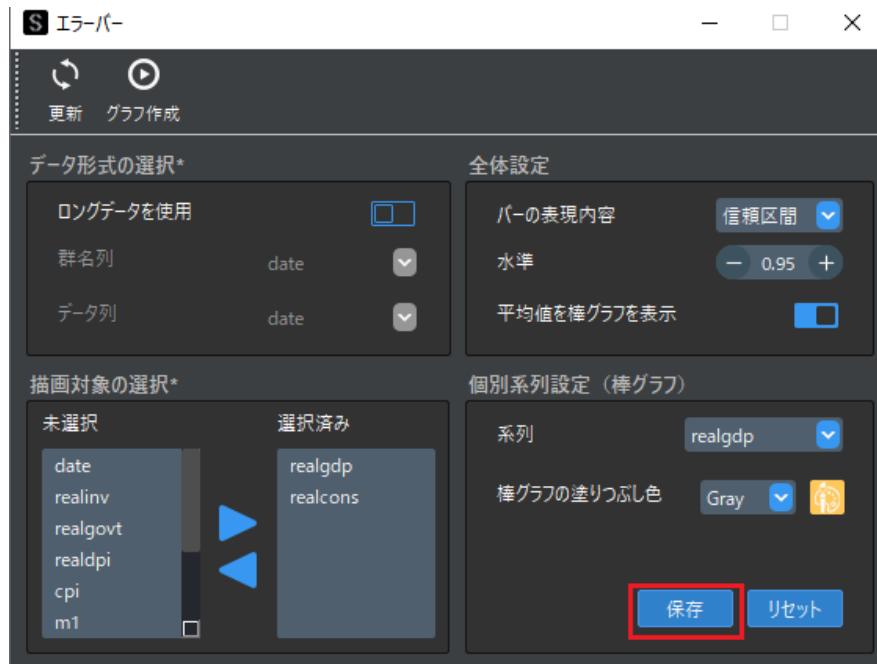
エラーバー作成機能の例では、赤枠が個別系列設定を行なう画面になります。設定したい変数を選択した状態で、色を自由に設定することができます。設定中の色は「パレット」ボタンに反映されます。



色の設定は基本色以外からも自由に設定可能で、「パレット」ボタンをクリックすると以下のような色選択ウィンドウが表示されます。

カラーパレットからの選択や、RGB 値から設定することが可能です。

※ 「Add to Custom Colors」ボタンは使用できません。



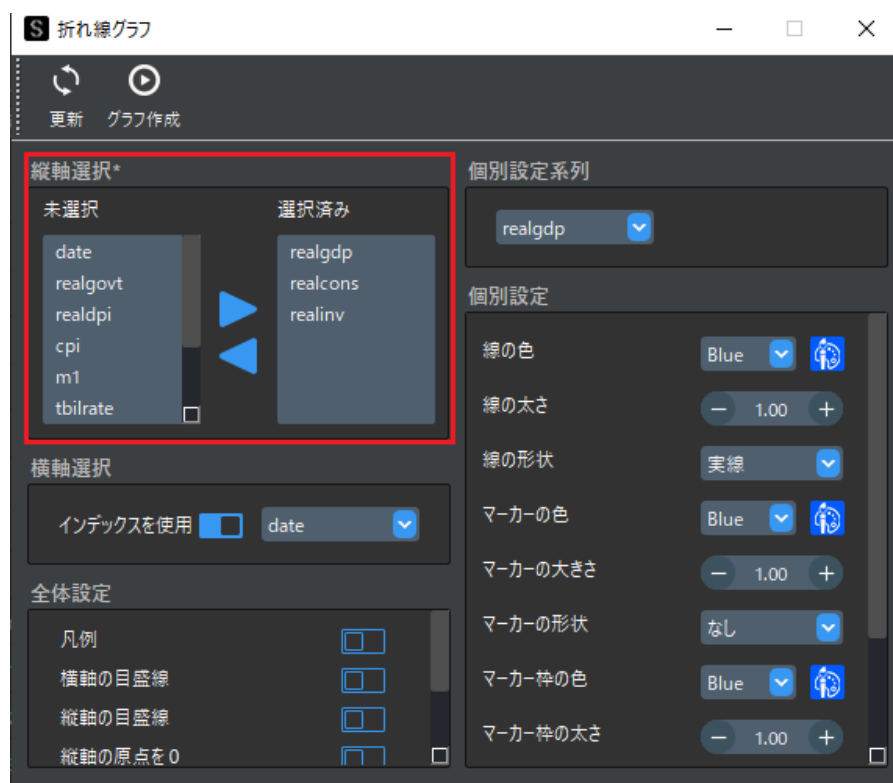
全体設定などは「保存」を行なう必要はありません。

2. 折れ線グラフ（無料）

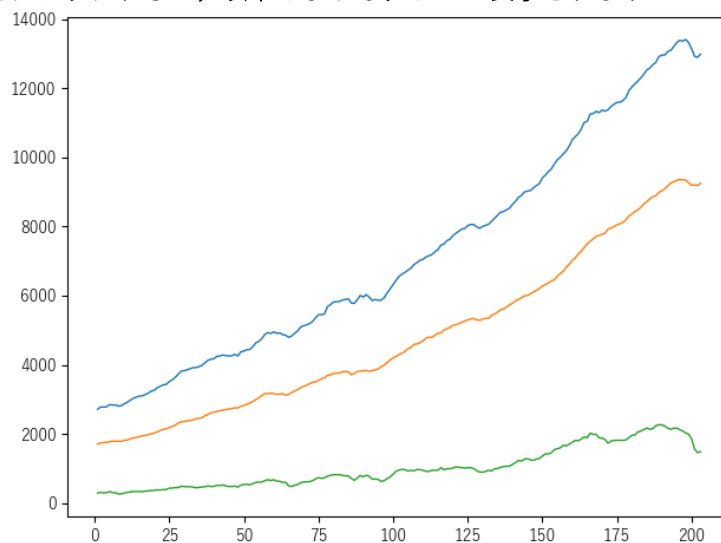
折れ線グラフは時系列データなど、変化を表すデータに有効なグラフです。

① 基本操作

メニューバーから「グラフ」→「折れ線グラフ」を選択します。以下のような折れ線グラフウィンドウが表示されたら、「縦軸対象」で折れ線グラフの描画対象としたい変数（列名）を選択します。



「グラフ作成」をクリックすると、以下のようなグラフが表示されます。

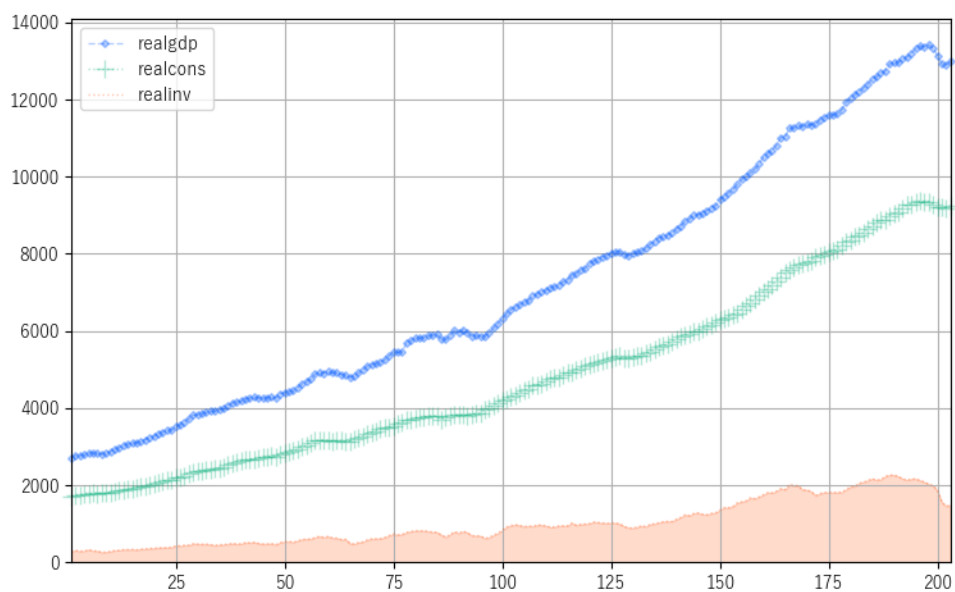


② 横軸の設定

横軸の値を設定する場合は、「横軸選択」で「インデックスを使用」を OFF にして横軸に設定する変数を選択します。

③ その他の設定

「全体設定」や個別系列ごとに設定を変更することで、以下のようなグラフも作成可能です。

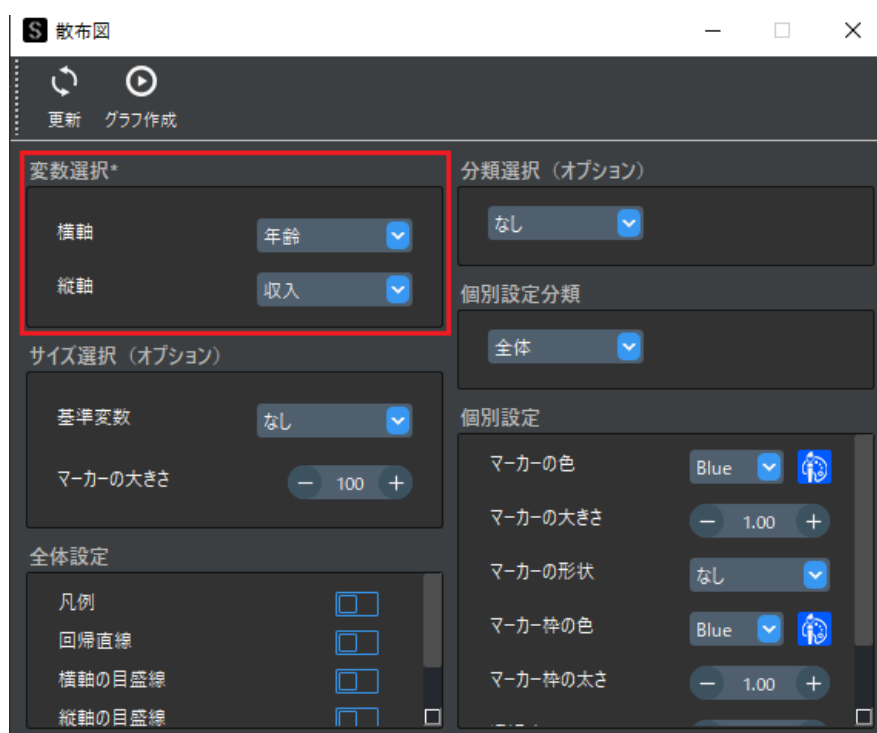


3. 散布図（無料）

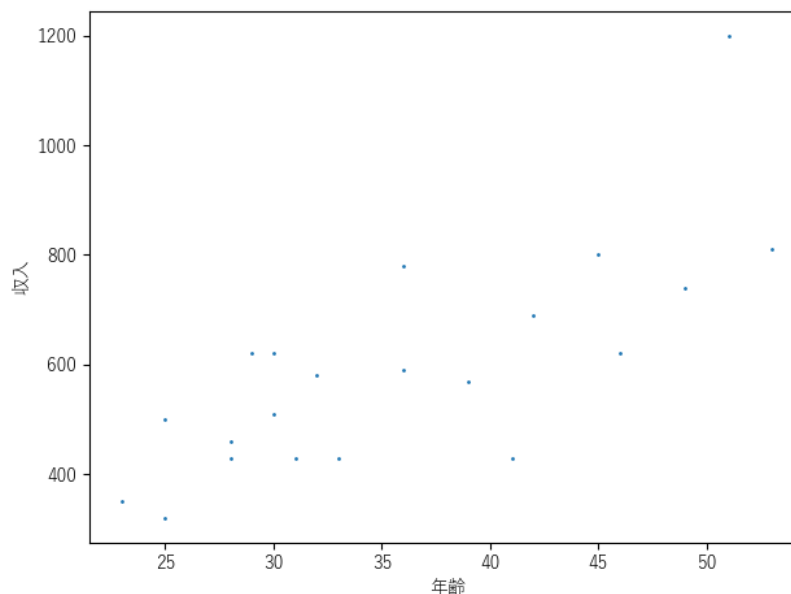
散布図は2つの量的データの関係性を示すのに有効なグラフです。3つの変数の関係性を示す場合は、バブルチャートが有効です。

① 基本操作

メニューバーから「グラフ」→「散布図」を選択します。以下のような散布図ウィンドウが表示されたら、横軸と縦軸に描画対象としたい変数（列名）を選択します。

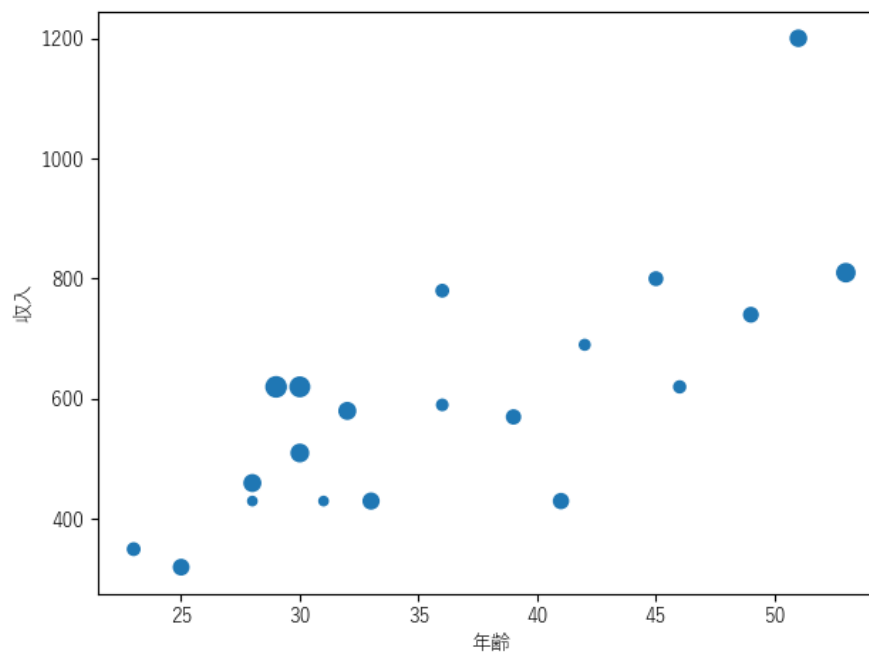


「グラフ作成」をクリックすると、以下のようなグラフが表示されます。



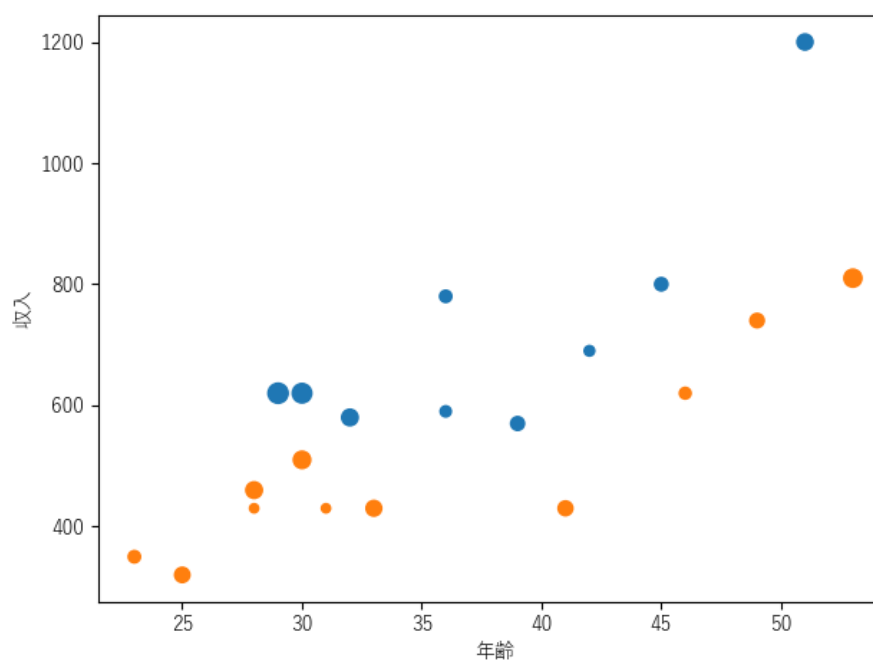
② サイズ選択（バブルチャート）

プロットのサイズで量的データを示したい場合は、「サイズ選択（オプション）」で、「基準変数」とする変数（列名）を選択して、グラフを作成します。



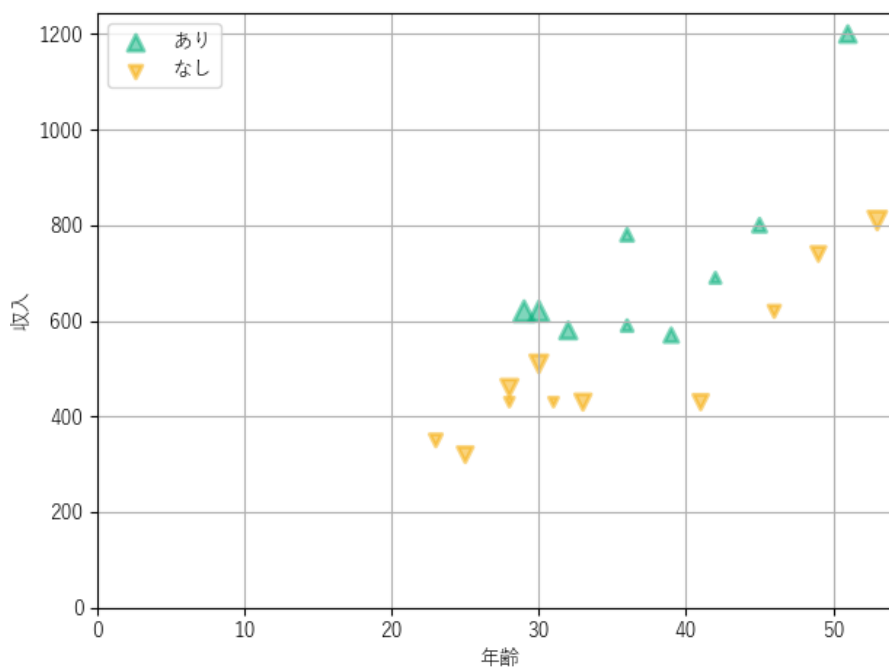
③ カテゴリーによる色分け

プロットの色をカテゴリーによって色分けしたい場合は、「分類選択（オプション）」でカテゴリーを示す変数（列名）を選択します。

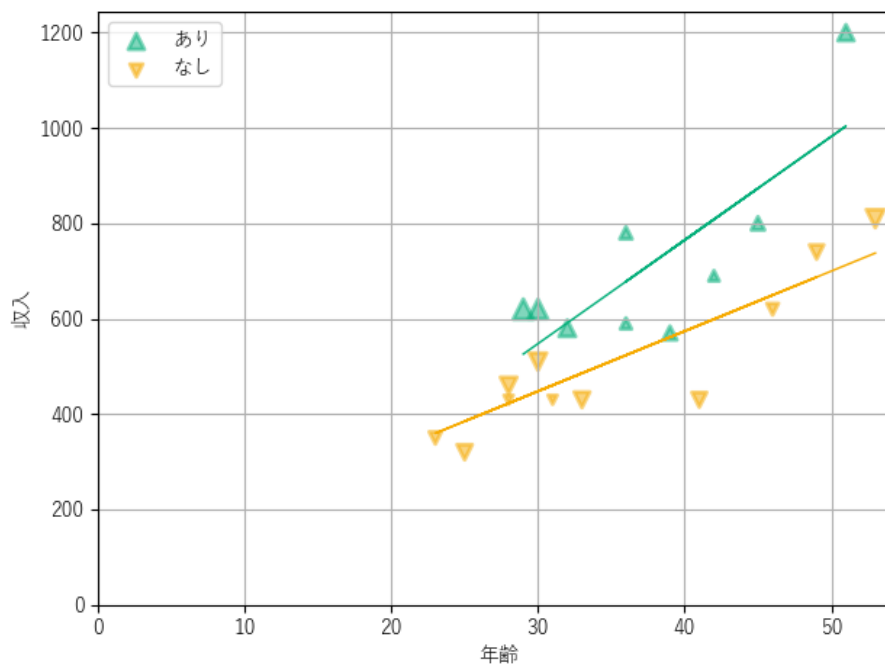


④ その他の設定

「全体設定」や個別系列ごとに設定を変更することで以下のようなグラフも作成可能です。



「全体設定」の「回帰直線」を ON にすると分類ごとに回帰直線が表示されます。



4. 散布図行列（無料）

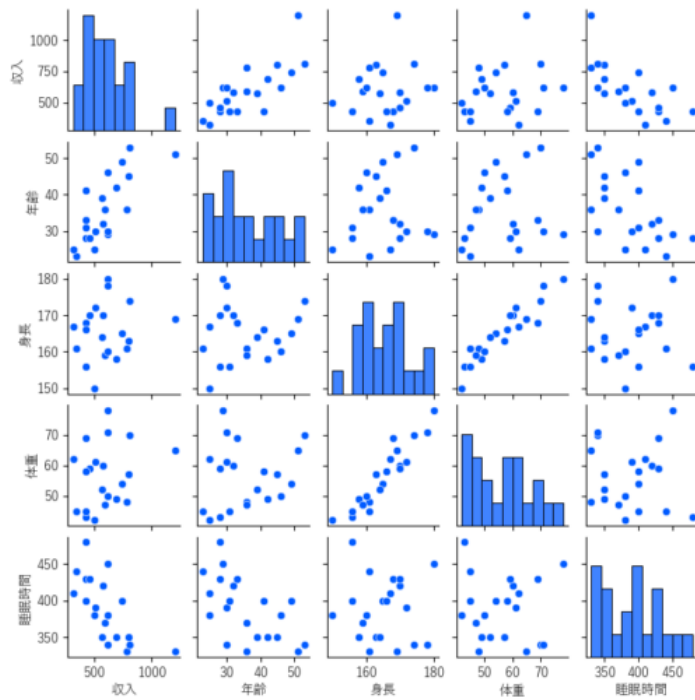
散布図行列は複数の変数に対して、各変数ペア間の相関関係を調べるのに有効なグラフです。重回帰分析などの多変量解析を実行する前に必ず行いたい手続きになります。

① 基本操作

メニューバーから「グラフ」→「散布図行列」を選択します。以下のような散布図行列ウィンドウが表示されたら、描画対象としたい変数（列名）を選択します。



「グラフ作成」をクリックすると、以下のようなグラフが表示されます。

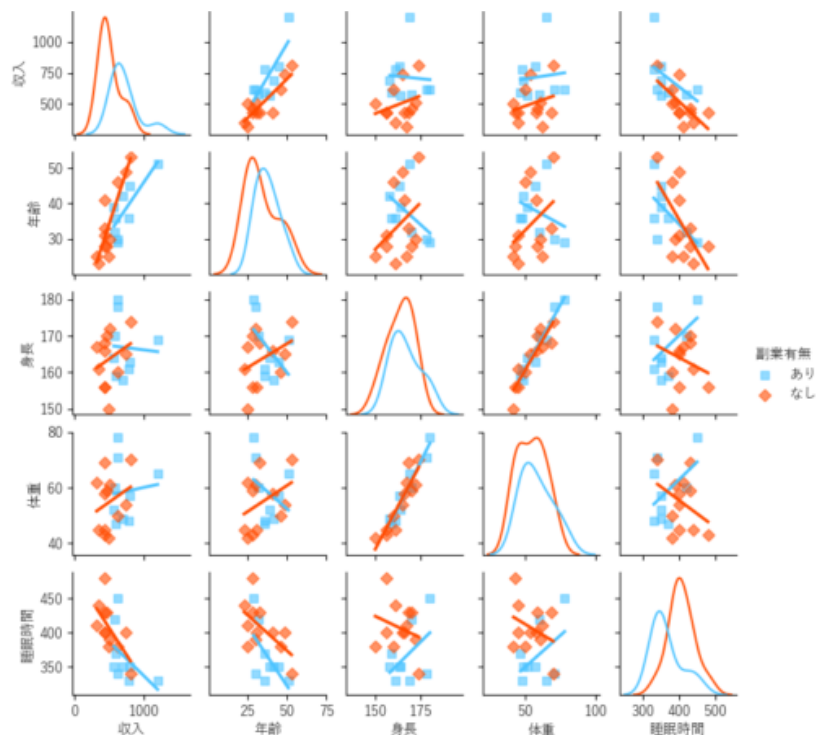


② カテゴリーによる色分け

プロットの色をカテゴリーによって色分けしたい場合は、「分類選択」でカテゴリーを示す変数（列名）を選択します。

③ その他の設定

「全体設定」や個別系列ごとに設定を変更することで以下のようなグラフも作成可能です。

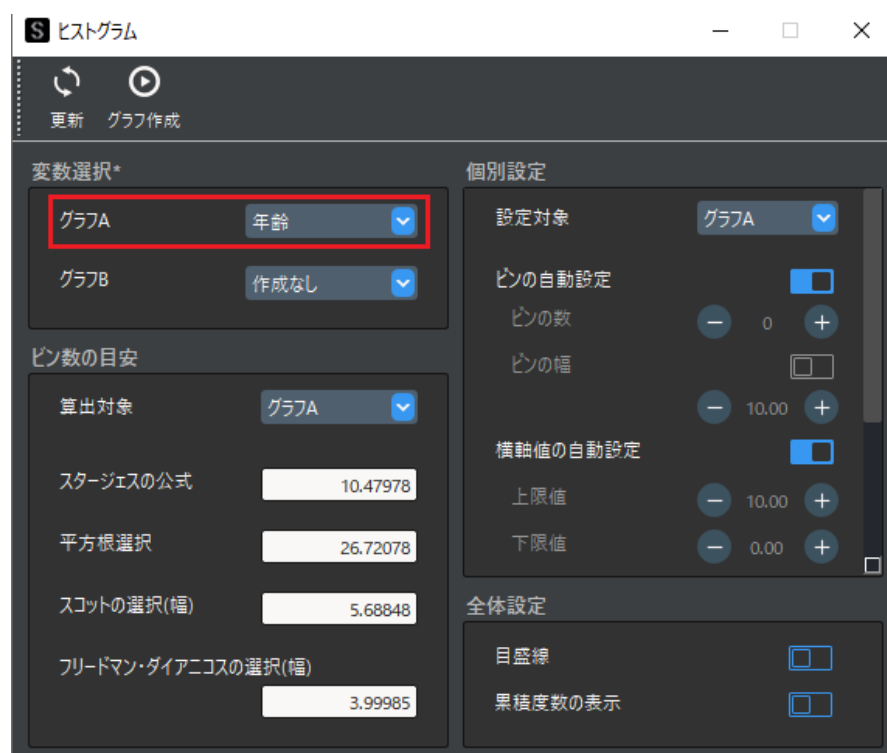


5. ヒストグラム

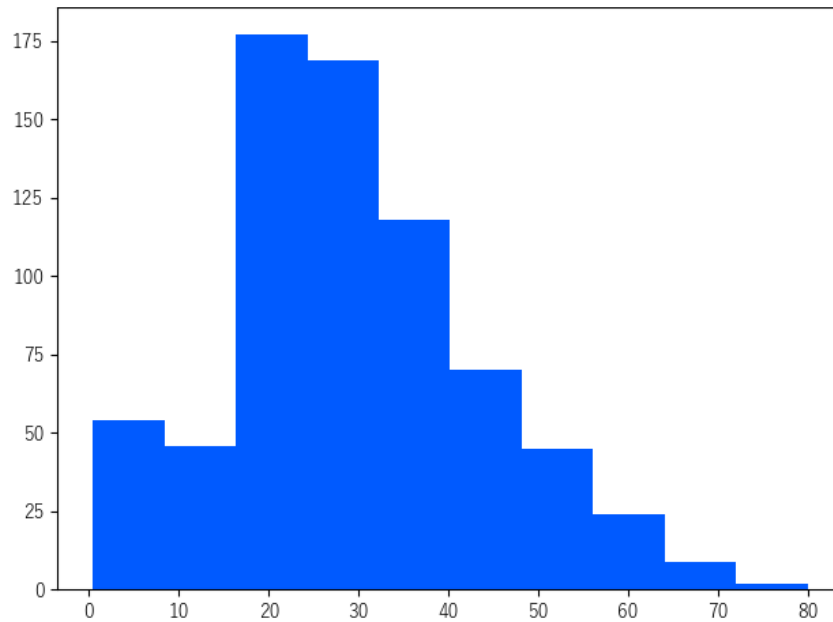
ヒストグラムは1つの変数の分布を示す最も基本的なグラフです。

① 基本操作

メニューバーから「グラフ」→「ヒストグラム」を選択します。以下のようなヒストグラムウィンドウが表示されたら、グラフ A に描画対象としたい変数（列名）を選択します。



変数を選択すると「ビン数の目安」に4つの基準・公式に沿ったビン数の目安が表示されます。基本的にはビン数は自動調整されるので、このまま「グラフ作成」ボタンをクリックします。

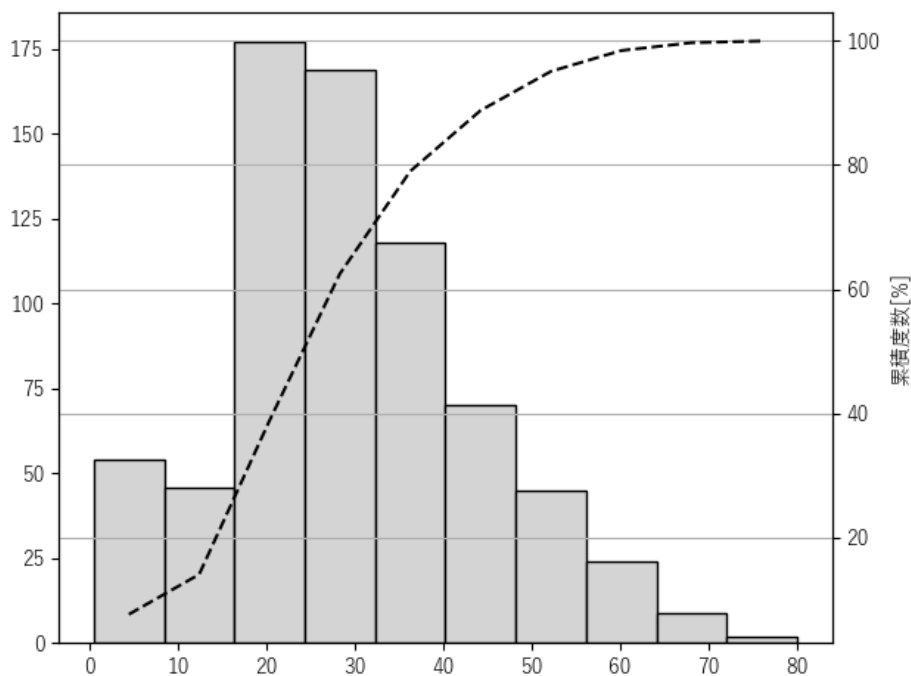


② 2つのヒストグラムを表示する

2つのヒストグラムを表示する場合は、グラフ B に対象の変数を選択してください。

③ その他の設定

「全体設定」やヒストグラムごとに設定を変更することで以下のようなグラフも作成可能です。

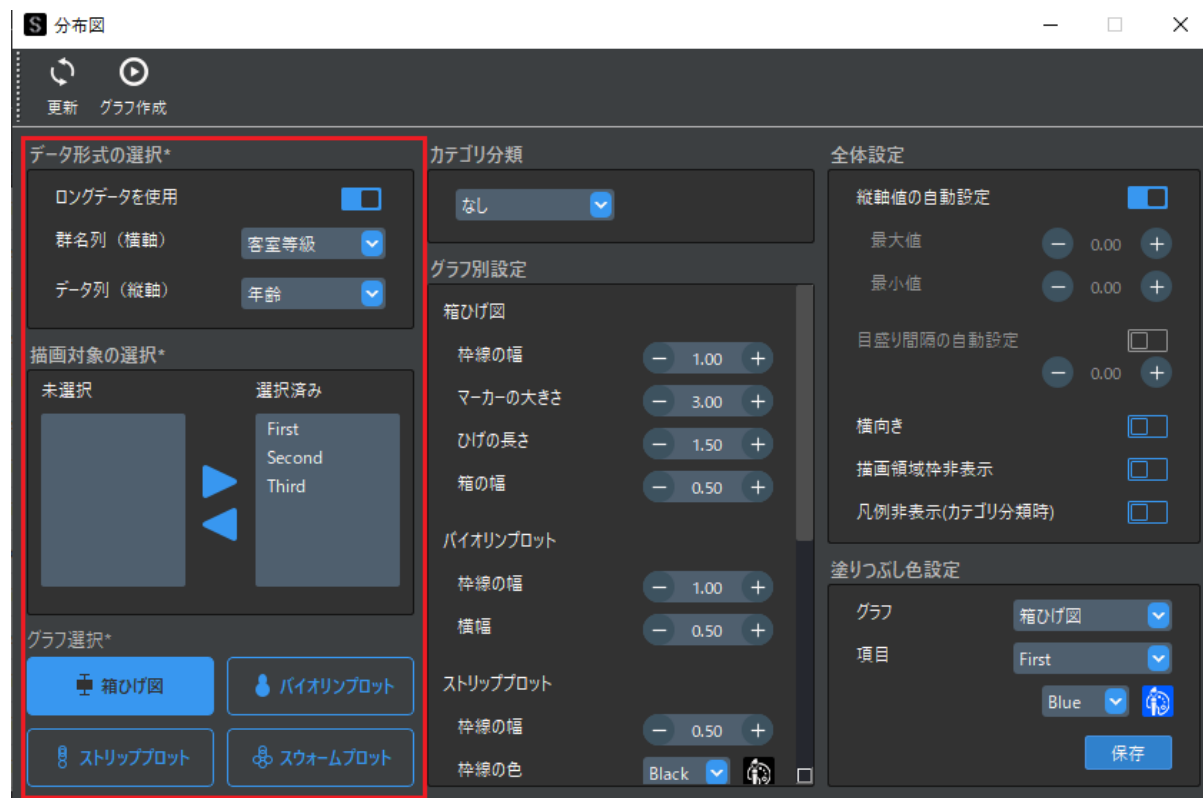


6. 分布図

分布図は複数のデータ分布を比較したい場合に有効です。

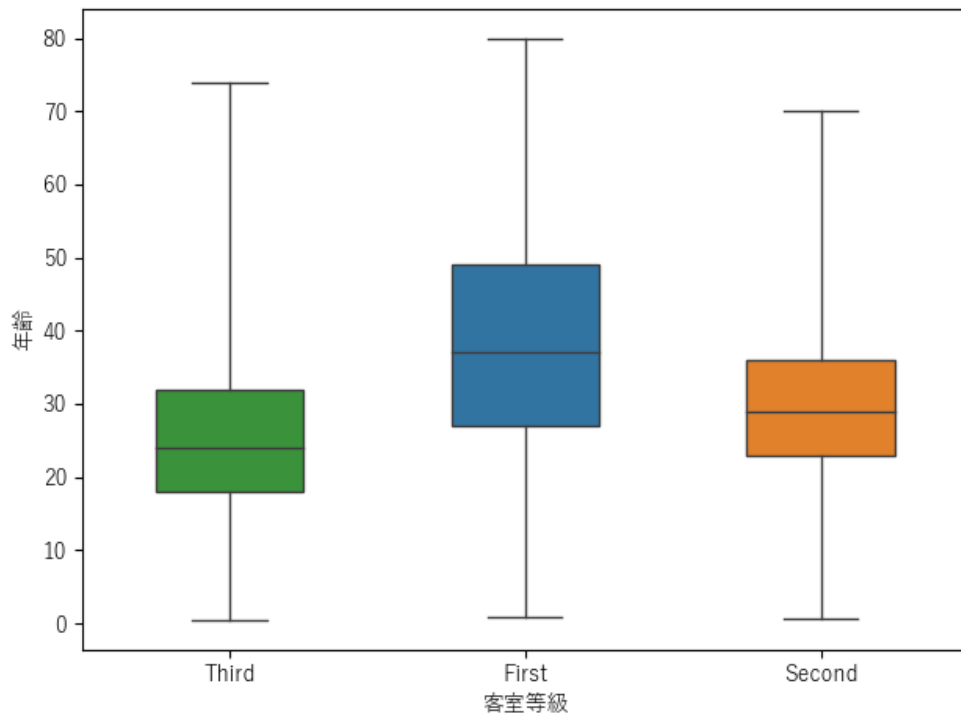
① 基本操作

メニューバーから「グラフ」→「分布図」を選択します。以下のような分布図ウィンドウが表示されたら、「データ形式」「描画対象」「グラフ」を選択します。

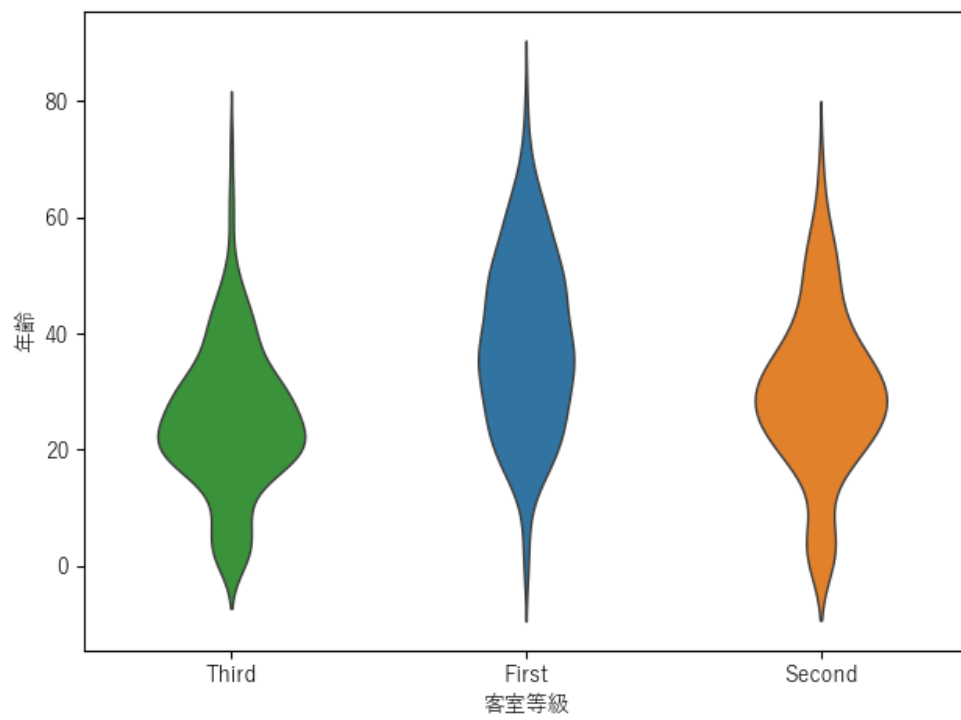


各グラフを選択して「グラフ作成」をクリックした場合、以下のようにグラフが表示されます。

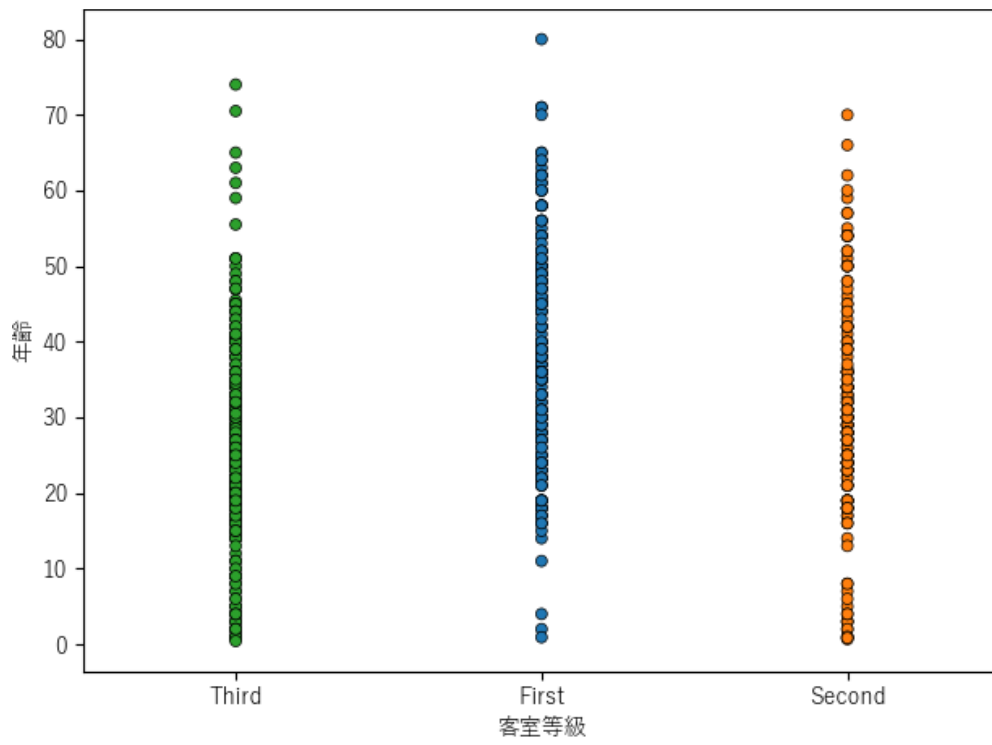
箱ひげ図



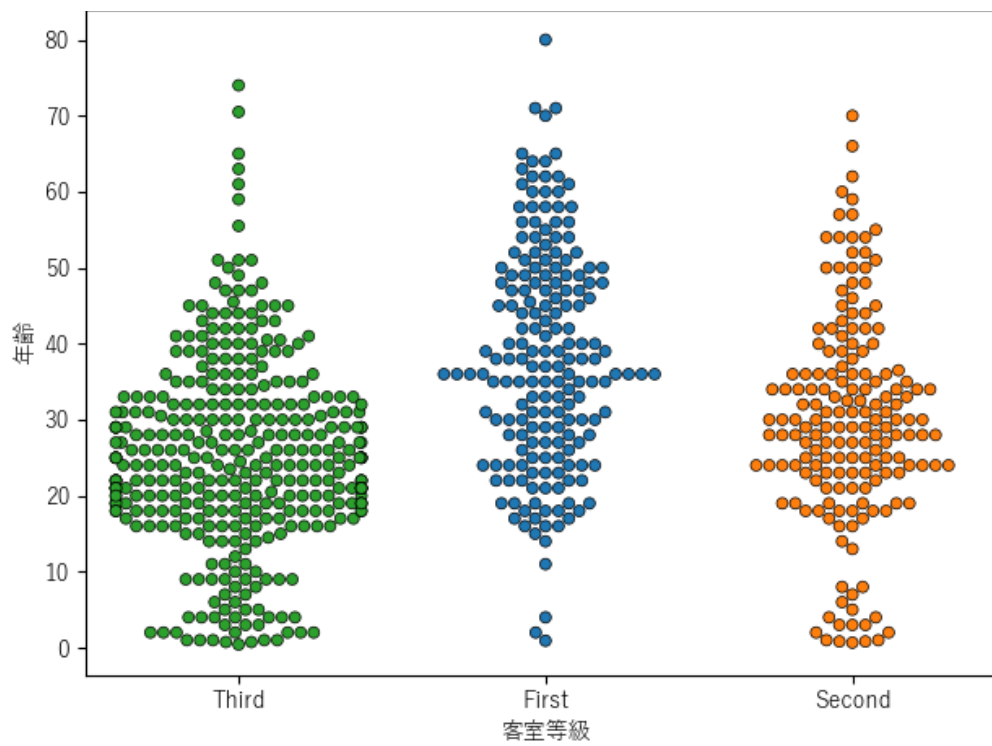
バイオリンプロット



ストリッププロット

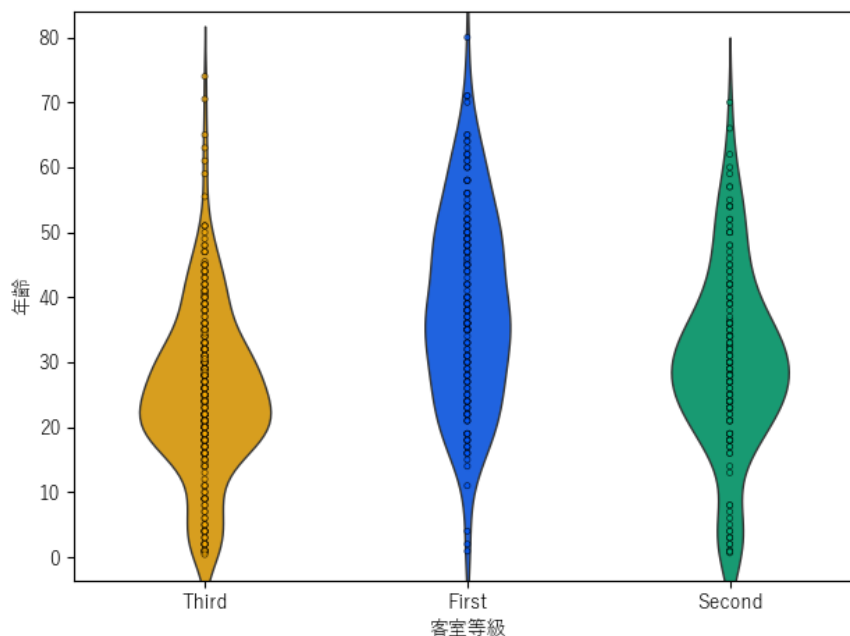


スウォームプロット



② 複数のグラフを組み合わせる

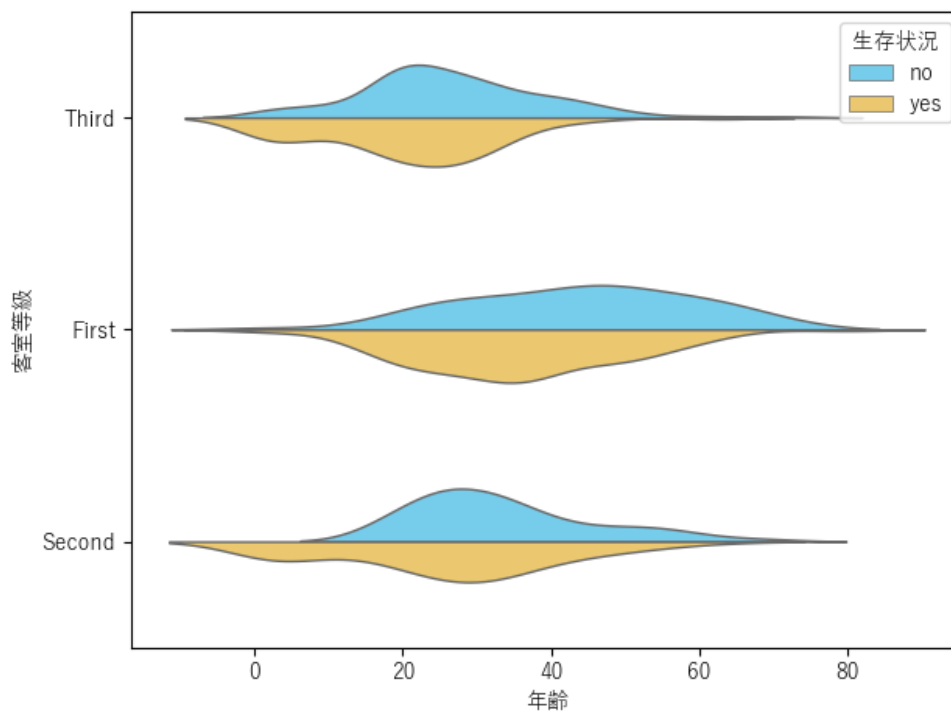
グラフ選択で複数のグラフを選択した場合、グラフが重なって表示されます。バイオリンプロットとストリッププロットを組み合わせたグラフの作成例は以下になります。



③ カテゴリーによる色分け

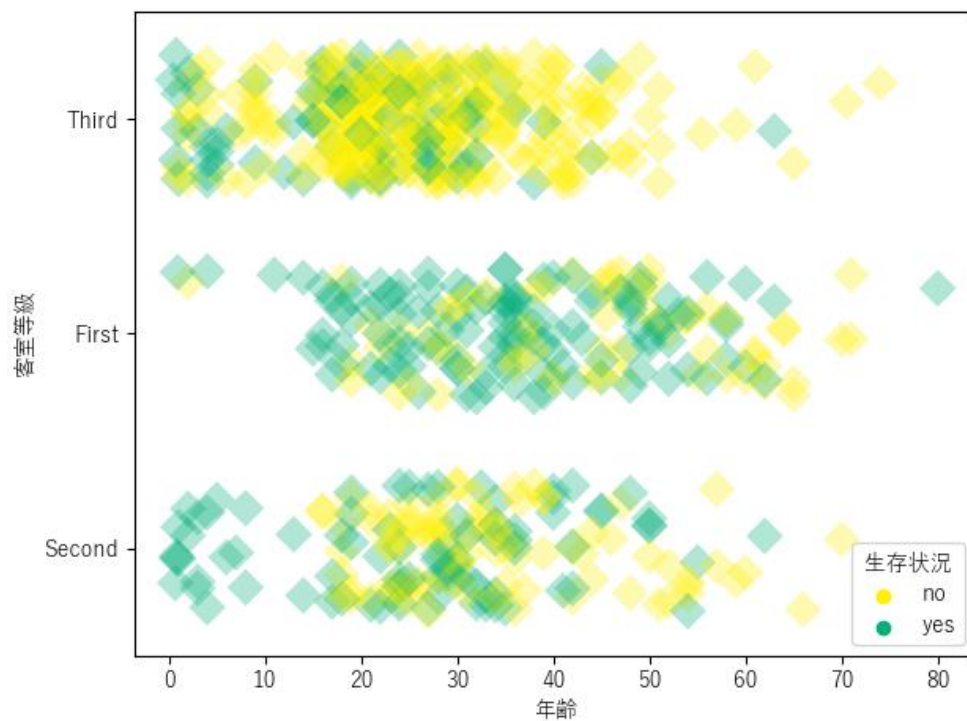
グラフの色をカテゴリーによって色分けしたい場合は、「カテゴリー分類」でカテゴリーを示す変数（文字型データ）を選択します。

※バイオリンプロットはカテゴリー数が2つの場合のみ分類可能です。



④ グラフ別の設定

「グラフ別設定」でグラフごとに様々な設定が変更可能です。特にストリッププロットの設定を変更することで、以下のような非常に自由度が高い様々なグラフを作成することが可能です。



7. エラーバー（無料）

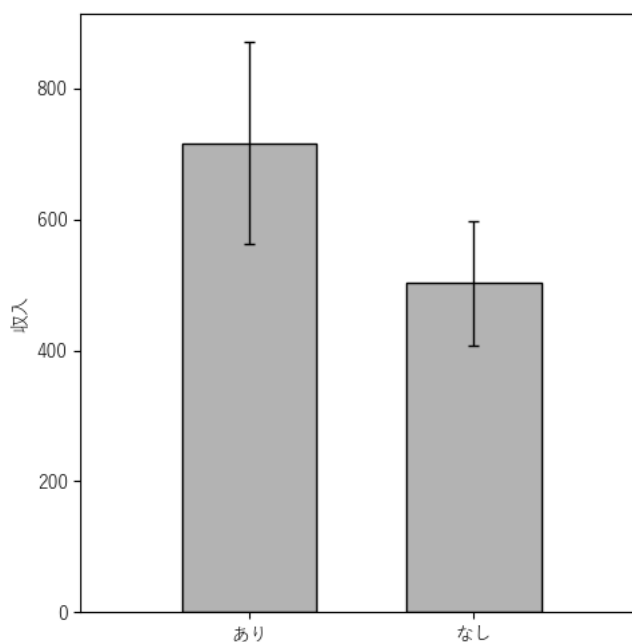
エラーバーはt検定の結果を示す際によく用いられます。

① 基本操作

メニューバーから「グラフ」→「エラーバー」を選択してエラーバー用ウィンドウを表示します。エラーバー用ウィンドウが表示されたら、描画対象のデータ設定を行います。以下の設定では"副業有無"の"あり""なし"ごとに,"収入"の値をエラーバーで表現します。



「グラフ作成」をクリックすると以下のようなグラフが作成されます。



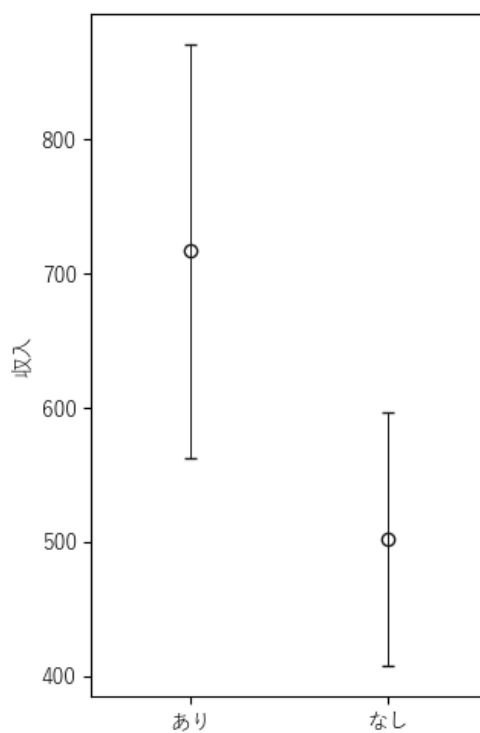
② エラーバーの値

全体設定では、エラーバーに使用する値を設定します。エラーバーの使用する値は以下の3つから選択可能です。

- 信頼区間
- 標準偏差
- 標準誤差

③ エラーバーだけ表示する

全体設定の「平均値を棒グラフに表示」にチェックを入れることで、以下のように棒グラフを除いたグラフが作成されます。エラーバーの中心円は平均値になります。



自動考察

StaatApp では ChatGPT API を用いた自動考察が可能です。

自動考察機能の使い方と注意点について説明します。

① 解析の実行（ロジスティック回帰分析）

ロジスティック回帰分析の結果から自動考察を行います。

メニューバーから「多変量解析」→「回帰」→「ロジスティック回帰分析」を選択してロジスティック回帰分析用ウィンドウを表示します。

ロジスティック回帰分析用ウィンドウが表示されたら、目的変数と説明変数の設定を行います。「モデル作成」をクリックすると以下のように、結果が表示されます。

統計量	
サンプルサイズ	21.0
対数尤度	-6.77619
AIC(赤池情報量基準)	23.55238
BIC(バイズ情報量基準)	-35.15998

	偏回帰係数	Wald統計量
切片	3.26677	0.26533
収入	0.02925	1.99906
性別	0.88699	0.56232
年齢	-0.38281	-1.96193
睡眠時間	-0.01966	-0.81392

② 自動考察の実行

解析を実行したら、ツールバーの「自動考察」ボタンをクリックします。自動考察中は以下のよう
にボタンが黄色くなります。

※ 自動考察は約 1 分要します。自動考察中はアプリの操作はしないようにしてください。



自動考察が完了すると考察結果が記載されたウィンドウが表示されます。考察内容の"データフレーム"は自動考察を行なうためのインプット形式なので、読み飛ばしてください。



考察結果の注意点

自動考察機能による考察結果は統計学的・数学的解釈でしかないので記載内容が定性的に正しいのか、相関関係・因果関係が現実的に正しいのかは十分に吟味してください。

精度については考察対象のデータ数（表のセル数）が多いと下がる傾向にあります。特に基本統計量やクロス集計表で、大量の変数を用いた場合は回答内容が正しいか十分に確認するようにしてください。その他の解析結果に対する考察の精度は非常に高くなっています。

③ 追加解説（応用）

考察結果が記載されたウィンドウでは、表示された考察内容についてより詳細な解説を得ることができます。

詳しく知りたい"変数名"や"統計用語"について、追加解説用の枠に入力して「質問」ボタンをクリックすることで、以下のようにさらに詳しい解説が表示されます。

追加解説

収入 について詳しく [質問](#)

収入の偏回帰係数は0.02925であり、p値が0.04560となっています。また、調整オッズ比は1.02969であり、95%CIの下限と上限は1.00057と1.059650e+00です。

偏回帰係数は、説明変数（この場合は収入）が1単位変化すると、目的変数（副業有無）がどれだけ変化するかを示しています。この場合、収入が1単位増加すると、副業有無の対数オッズが0.02925増加する傾向があります。

p値は、仮説検定の結果、説明変数が目的変数に統計的に有意な影響を与えるかどうかを示しています。この場合、p値が0.04560となっており、有意水準（一般的には0.05）以下であるため、収入は副業有無に統計的に有意な影響を与えと言えます。

調整オッズ比は、ある説明変数のオッズ比が1単位変化した場合の目的変数のオッズ比の変化を示しています。この場合、収入の調整オッズ比が1.02969となっており、収入が増えると副業の有無のオッズが1.02969倍増加する傾向があるとと言えます。

95%CI（95%信頼区間）は、説明変数の効果の真の値が含まれる範囲を示しています。下限と上限の値を見ると、収入が増えると副業の有無のオッズが1倍以上1.059650倍以下になることが95%の信頼度で言えるということです。

以上の結果から、収入が副業有無に統計的に有意な影響を与える予測変数であり、収入が増えると副業の有無が増える傾向があると考えられます。ただし、効果の大きさは0.02925と比較的小さいため、他の要因との相互作用や他の変数の影響と併せて考える必要があります。

※ 追加解説の結果については再度、解説を求めることはできません。

④ 自動考察機能を使用できる解析手法一覧

自動考察を行える解析手法一覧は以下になります。

解析手法	考察対象
基本統計量	各変数の統計量
相関係数	相関関係・無相関の検定結果
分析分析	分析分析表
重回帰分析	モデルの適合度・各説明変数の統計量
ロジスティック回帰分析	モデルの適合度・各説明変数の統計量
プロビット分析	モデルの適合度・各説明変数の統計量
一般化線形混合モデル	モデルの適合度・ランダム効果・固定効果
判別分析	モデルの適合度・各説明変数の統計量
主成分分析	主成分負荷量
因子分析	因子負荷量
クロス集計表	クロス集計結果
Cox 比例ハザード回帰	各説明変数の統計量
アソシエーション分析	アソシエーションルール

⑤ 自動考察機能の仕様

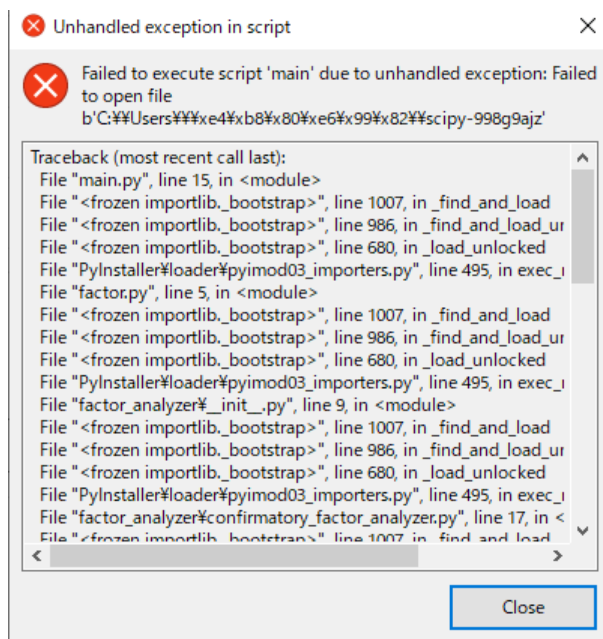
自動考察機能は、OpenAI 社が開発した ChatGPT API (AI) を利用しています。

解析結果を API に入力することで、その回答を得ています。StaatApp で解析を行ったデータ自体は ChatGPT に入力していないので、データ流出はしません。

FAQ

1. Windows 版起動時にエラーが発生する

StaatApp を起動した際に、以下の画面が表示されて起動しない場合があります。



原因は起動しているユーザ名に日本語が含まれているためです。上記のエラーが発生した場合、紹介する2つの対応方法のうちどちらかを行うことで起動することができます。

① ユーザ名が英数字のユーザで実行する

Windows のユーザで英数字のみで表記されたユーザがある場合、そのユーザでログインして StaatApp を実行します。

英数字のみで表記されたユーザが存在しない場合は、新しくユーザを作成するか対応方法②を実施してください。

② 環境変数を変更する

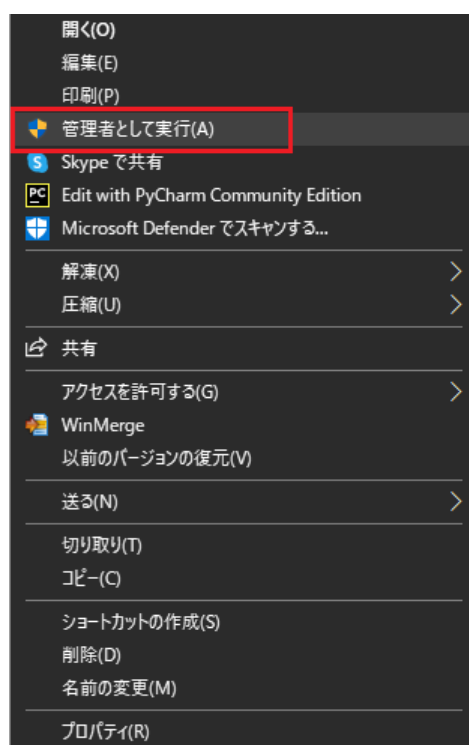
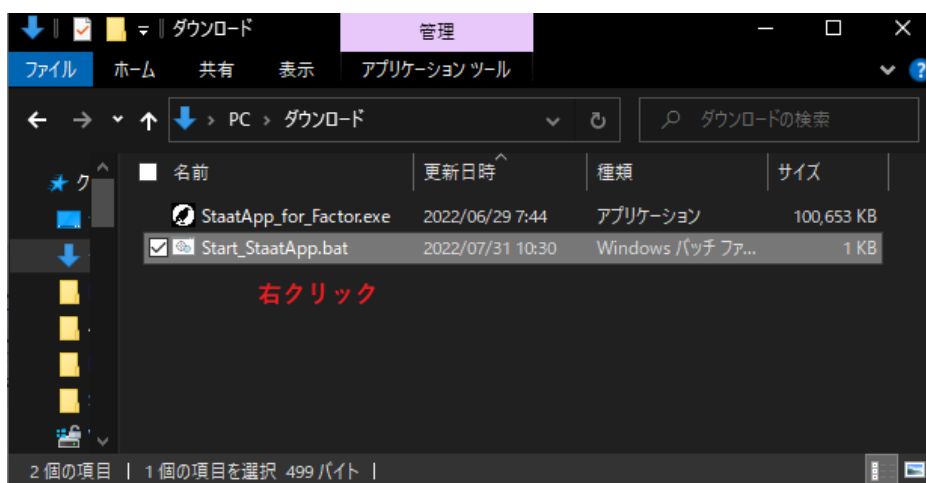
エラーの原因は StaatApp 起動時に使用する temp フォルダのに日本語が含まれることです。以下のページを参考に環境変数 TEMP と TMP を、日本語が含まれないパスに変更することで対応可能です。

[▶ 環境変数の変更方法](#)

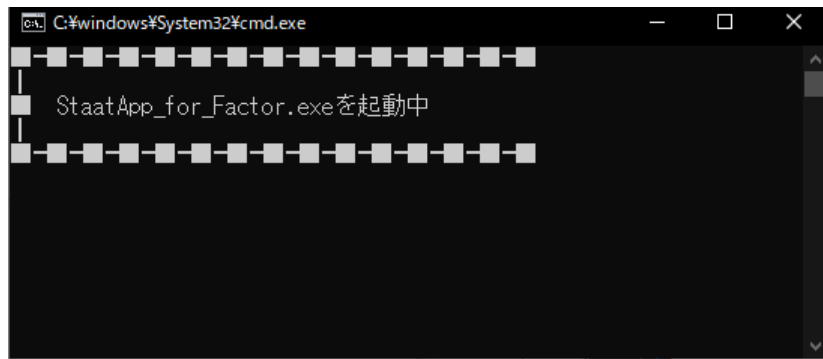
Windows にあまり詳しくない方は、StaatApp 起動用の bat ファイルを [Github](#) からダウンロードしてお使いください。一番手軽かつ PC の設定に影響を与えないためおすすめの方法です。

Github から”Start_StaatApp.zip”をダウンロードしたら、zip ファイルを解凍します。解凍後、同じフォルダに使用する StaatApp を配置してください。

同じフォルダに StaatApp と Start_StaatApp.bat を配置したら、Start_StaatApp.bat を右クリックして「管理者として実行」を選択します。



「このアプリがデバイスに変更を加えることを許可しますか。」といった画面が表示されたら「はい」を選択します。以下の画面が表示後、数秒後に StaatApp が起動して画面が表示されます。



▷ 起動用バッチファイルのダウンロード

2. macOS 版の起動方法

ダウンロードした app ファイルを「Control + 右クリック（ダブルタップ）」します。



表示されたメニューウィンドウから「開く」を選択します。



以下の画像のような警告が表示されるので「開く」を選択します。

※ macOS 仕様上、AppStore 以外から入手したアプリは警告が表示されますが正常な動作のため問題ありません。



起動まで最大4分程度要するので、気長にお待ちください。以下の画面が表示されたら起動完了です。

macOS 版の UI（ユーザインターフェース）は操作マニュアルとは異なりますが、使用方法や動作は Windows 版と同等です。