バージョン 4.2.0
2026/7
作成： StaatApp 運営チーム

StaatApp 操作マニュアル

目次

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用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．

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．

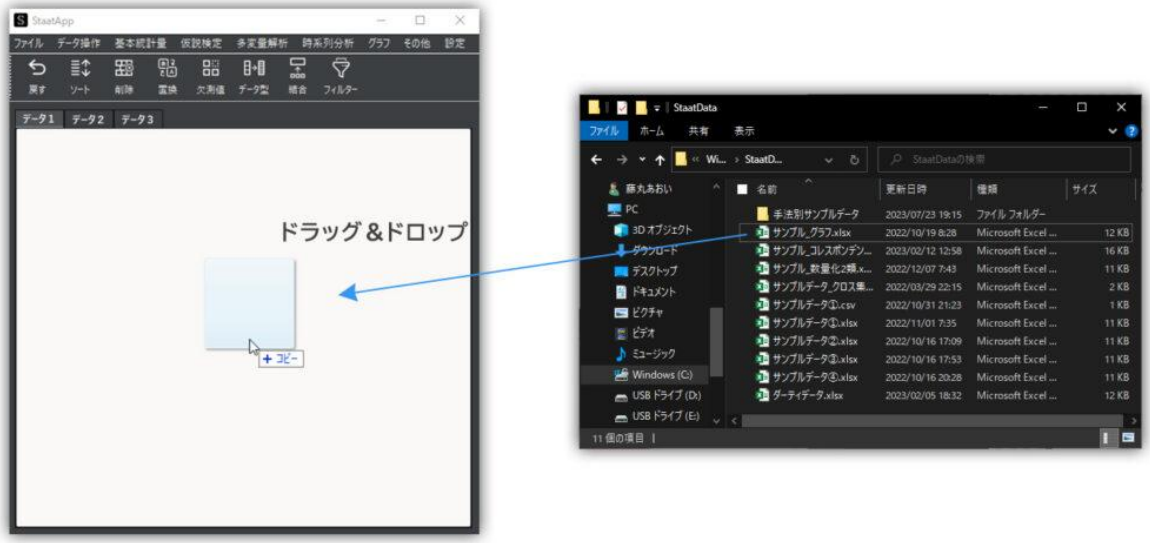
基本操作

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）

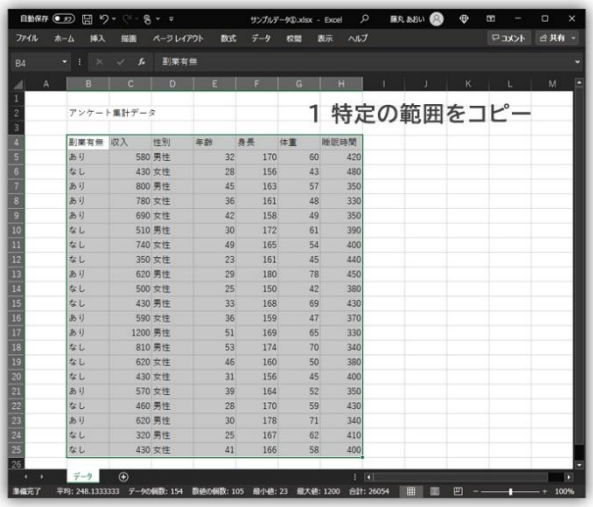
お使いの PC にあるファイルを StaatApp 上にドラッグ＆ドロップすることで，データが StaatApp に取り込まれます．



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．

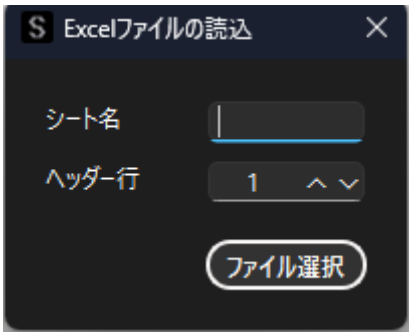
StaatApp では右クリックメニューもしくは，キーボードのコントロールキーと"V"を同時に押すことでクリップボードのデータを貼り付けることができます．



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

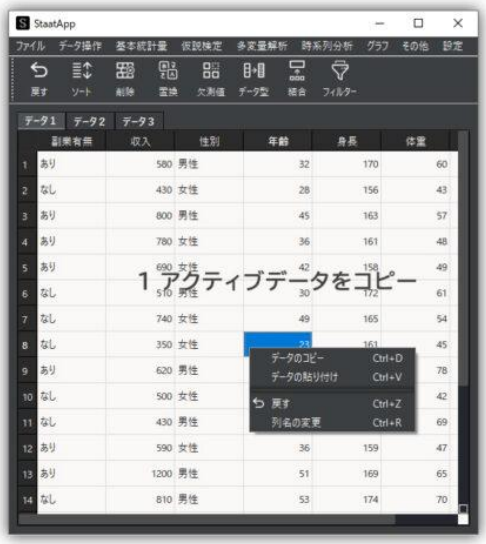
StaatApp で扱うデータのデータ構造は，整然データとする必要があります．整然データについては以下のページで紹介しています．

▷ 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．

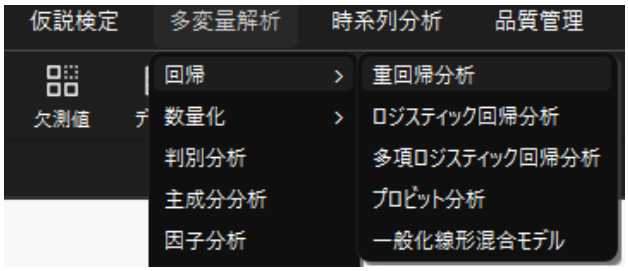


3．データを保存する

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

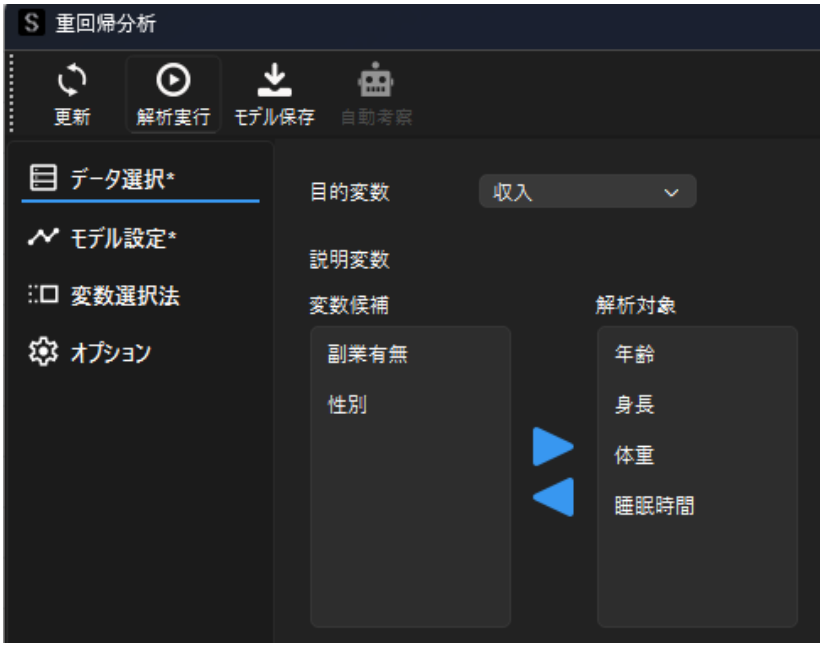
読み込ませたデータに対して統計解析を行うためには，解析用ウィンドウを表示します．解析用ウィンドウはメニューバーから選択して表示することができます．



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▶ボタン（重回帰分析の場合は「解析実行」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

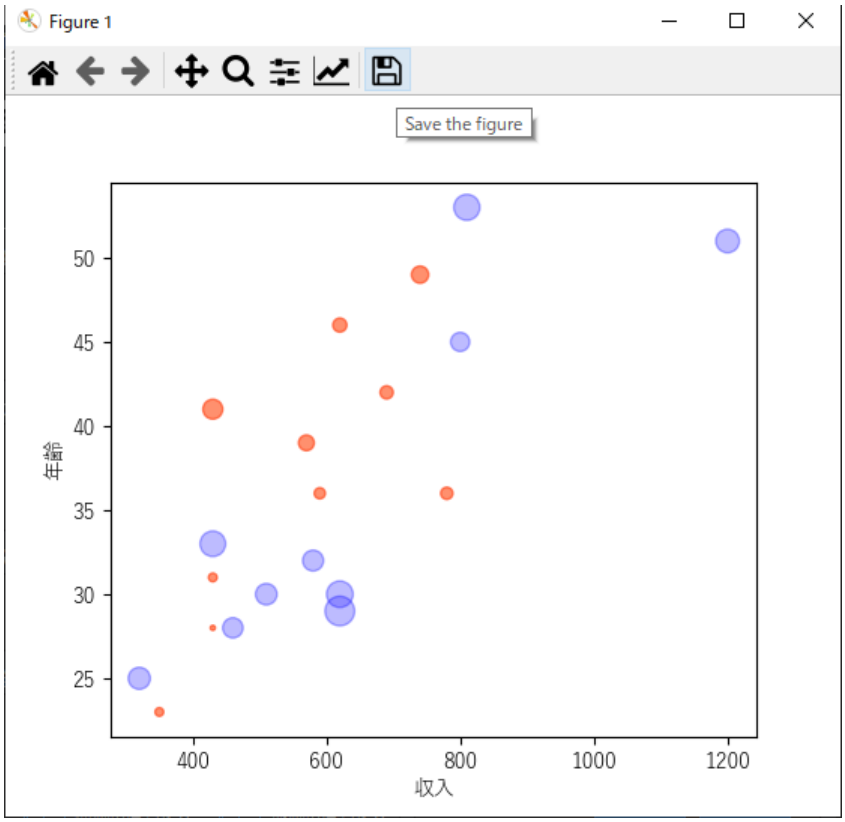
解析結果表示域の下にある「クリップボードにコピー」ボタンをクリックすると，対象の解析結果がクリップボードにコピーされます．

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. 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．

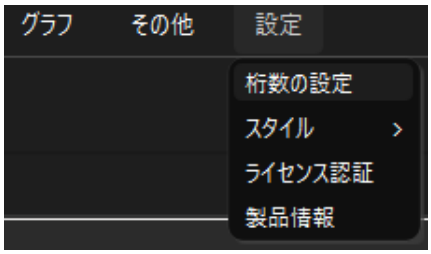
表示されたグラフは「画像保存」ボタンをクリックすることで，任意のファイル名で png 形式として保存することができます．



9. 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．

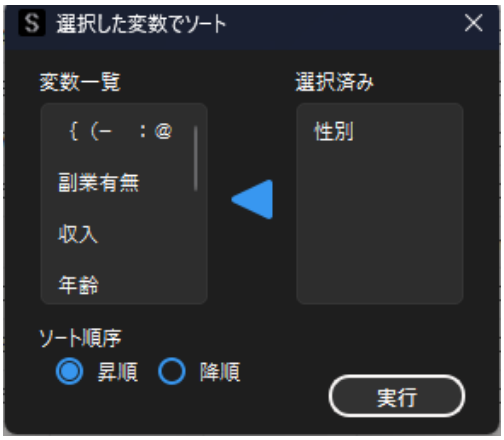


データ操作

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」機能を用います．

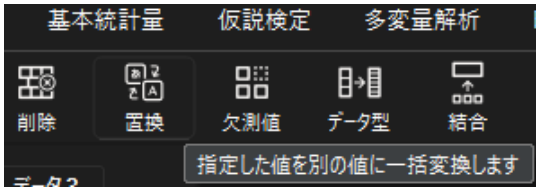
特定の行を削除したい場合は削除したい行番号を「行番号（上限）」に入力します．複数の行を一括で削除したい場合は，「行番号（上限）」に上側の行番号を，「行番号（下限）」に下側の行番号を入力します．

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「セル内容が完全一致したものを検索」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．

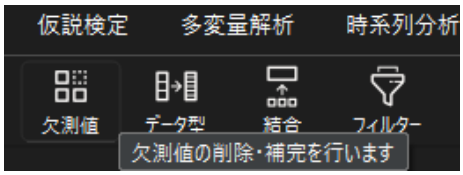
例では”性別”列の表記ゆれを修正するために，”女”→”女性”に置換します．

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，「One-Hot 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．変換対処の列のカテゴリー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．

カテゴリー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ー変数に対する統計解析が可能．

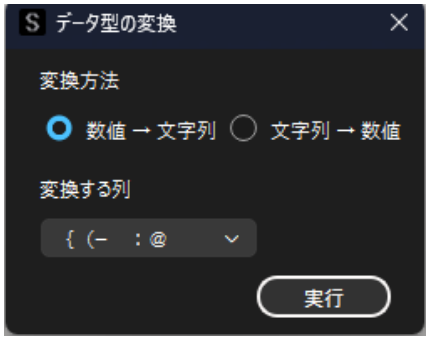
▷統計学におけるデータの種類

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．

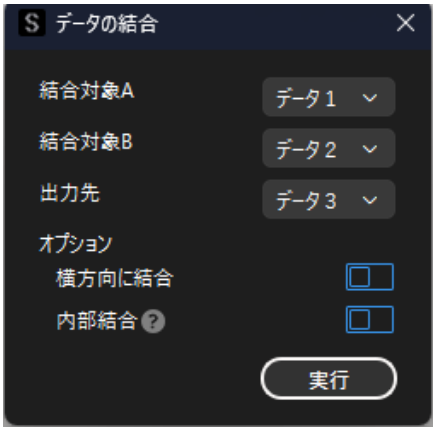


※解析を実行した際に，データ型に関するエラーが発生する場合はこの機能を用いて変換してください．

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」機能を用います．

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．

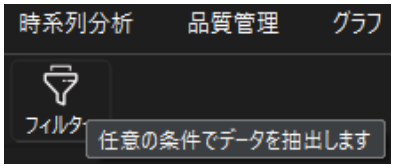


デフォルトではデータは縦方向に結合されます．同じ列名を持つデータを結合したい場合，サンプルサイズを増やしたい場合に用います．

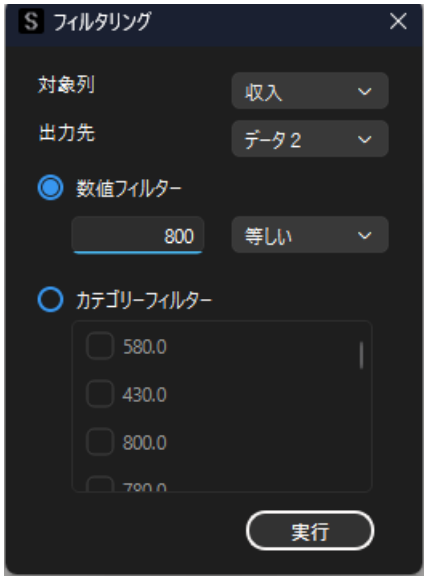
オプションの「横方向に結合」を選択すると，同じ行同士でデータが結合されます．変数を加える場合や，解析後に出力した予測結果を元のデータに結合したい場合に用います．

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」機能を用います．



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．



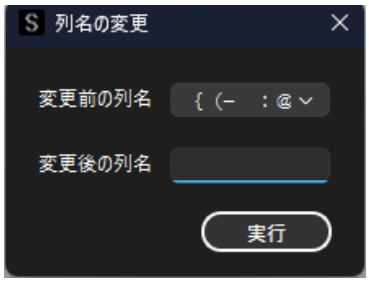
特定のカテゴリー（文字列）に一致するデータを抽出したい場合は，以下のように「カテゴリーフィルター」を選択して，抽出したいカテゴリーにチェックを入れます．

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．

	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．

「データの選択」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．

相関行列

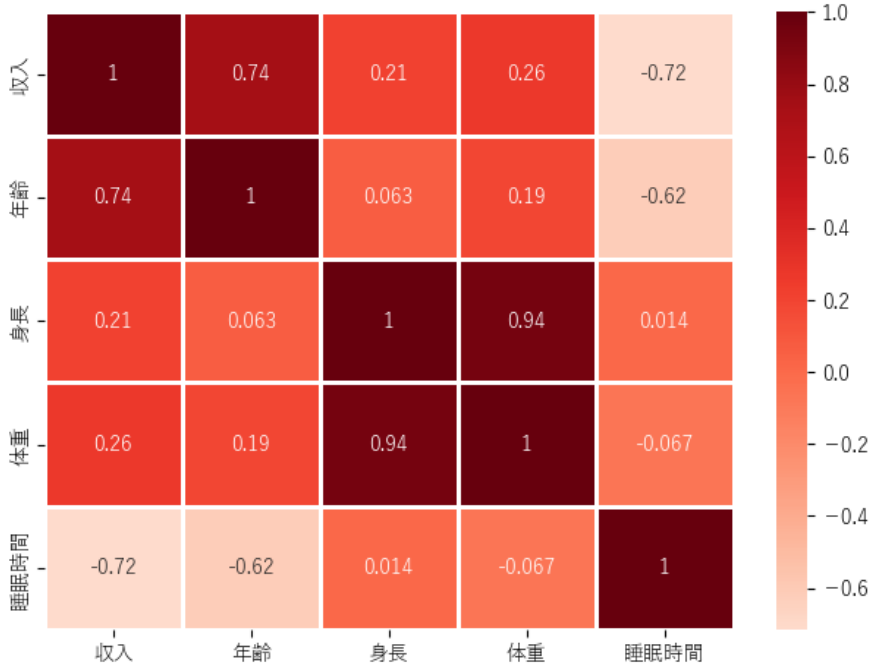
	収入	年齢	身長	体重	睡眠時間
収入	1.0	nan	nan	nan	nan
年齢	0.761	1.0	nan	nan	nan
身長	0.1985	0.0902	1.0	nan	nan
体重	0.2469	0.1565	0.9357	1.0	nan
睡眠時間	-0.6722	-0.5938	-0.009	-0.0401	1.0

無相関の検定（p値）

	収入	年齢	身長	体重	睡眠時間
収入	0.0	nan	nan	nan	nan
年齢	0.0001	0.0	nan	nan	nan
身長	0.3883	0.6973	0.0	nan	nan
体重	0.2805	0.498	0.0	0.0	nan
睡眠時間	0.0008	0.0045	0.9691	0.863	0.0

② ヒートマップの作成

ヒートマップ設定画面で「ヒートマップ」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．級内相関係数

級内相関係数は，ICC(1,1)，ICC(1,k)，ICC(2,1)，ICC(2,k)，ICC(3,1)，ICC(3,k) とそれぞれの信頼区間を算出することができます．

① ICC(1,1)の算出

以下のサンプルデータを用いて解説します．検者内信頼性を示す ICC(1,1)を算出する場合は，同じ評価者が複数回同じ被験者を観測したデータを読み込みます．

	被験者番号	1回目	2回目	3回目
1	1	54	56	60
2	2	64	63	65
3	3	70	65	54
4	4	71	70	70
5	5	84	82	90
6	6	68	70	56
7	7	77	76	69
8	8	92	90	79
9	9	53	55	61
10	10	82	80	90

級内相関係数用のウィンドウで，対象のデータを選択して「種別」にはデフォルトの「ICC(1,1)」を選択します．

「解析実行」ボタンをクリックすると，画面右側に級内相関係数と 95%信頼区間の値が表示されます．

② ICC(2,1)の算出

検者間信頼性を示す ICC(2,1)を算出する場合は，複数人の評価者が同じ被験者を観測したデータを読み込みます．

	被験者番号	評価者A	評価者B	評価者C
1	1	54	46	58
2	2	64	50	61
3	3	70	68	78
4	4	71	60	74
5	5	84	72	87
6	6	68	64	72
7	7	77	75	84
8	8	92	84	88
9	9	53	52	58
10	10	82	75	75

級内相関係数用のウィンドウで，対象のデータを選択して「種別」にはデフォルトの「ICC(2,1)」を選択します．

「解析実行」ボタンをクリックすると，画面右側に級内相関係数と 95%信頼区間の値が表示されます．

③ その他の級内相関係数の算出方法

ICC(1,k)は ICC(1,1)と同じデータの入力方法で，ICC(2,k)と ICC(3,1)，ICC(3,k)は ICC(2,1)と同じデータの入力方法で算出することができます．

3．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．

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「データ選択」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 α＝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．

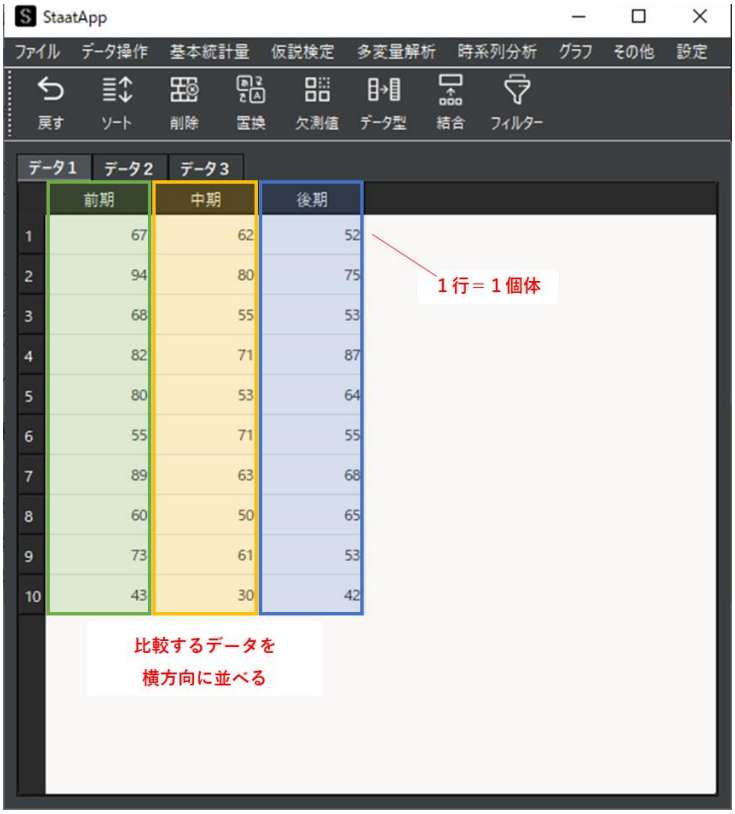
4．対応のある場合の仮説検定

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．



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，「データ選択」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 α＝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．

5. ロジスティック回帰分析

※ 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.

	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．

解析結果

モデル全体の指標

n	対数尤度	AIC	BIC
21	-6.7762	23.5524	28.775

係数

説明変数	偏回帰係数	Wald統計量	p値	調整オッズ比	95%CI下限	95%CI上限
切片	3.2668	0.2653	0.7908	26.2265	0.0	79222354...
収入	0.0293	1.9991	0.0456	1.0297	1.0006	1.0596
性別	0.887	0.5623	0.5739	2.4278	0.1103	53.4437
年齢	-0.3828	-1.9619	0.0498	0.6819	0.4652	0.9996
睡眠時間	-0.0197	-0.8139	0.4157	0.9805	0.9352	1.0281

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 モデルの評価・予測（応用）

作成したモデルに対して，予測精度の評価や発生確率の予測を行う場合は「モデル保存」ボタンをクリックしてモデルの保存を行います．

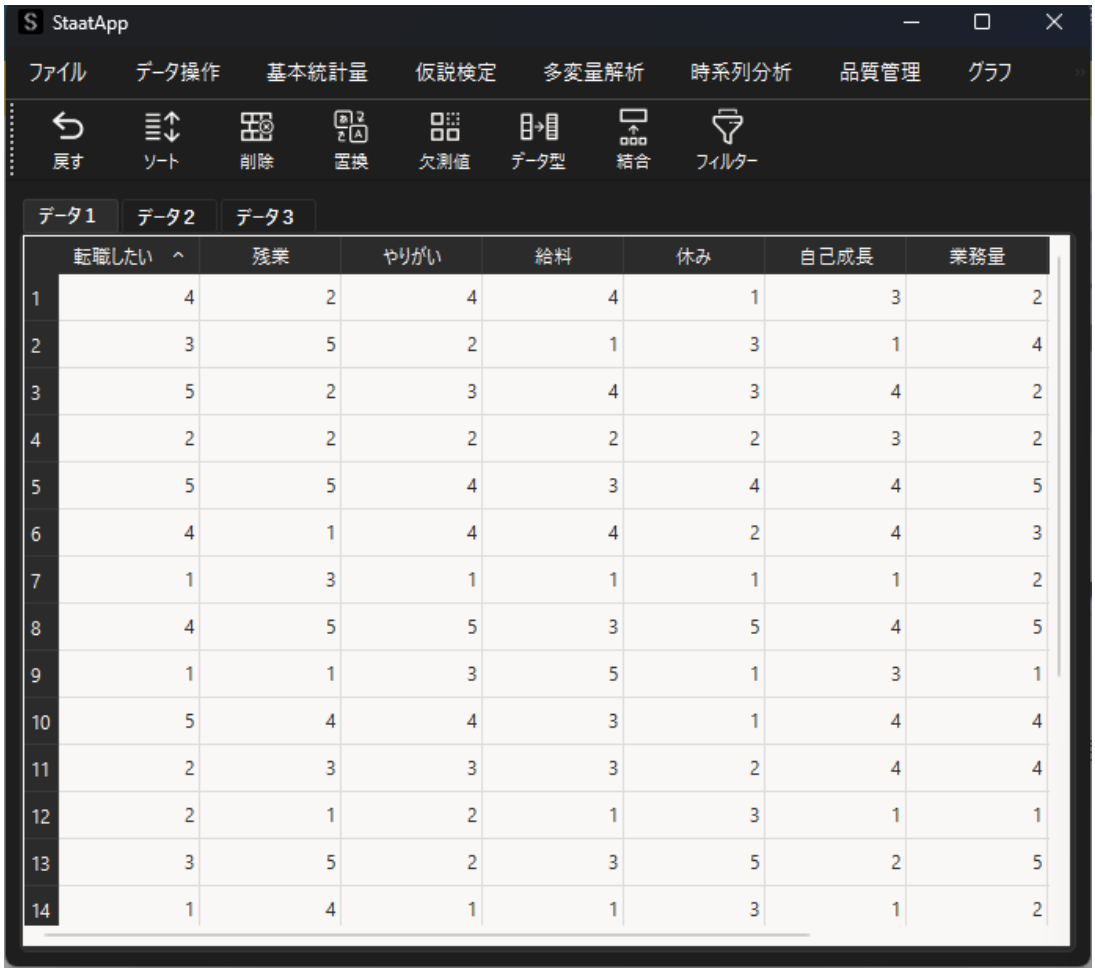
モデルの評価・予測方法の詳細は以下の Web サイトをご覧ください．

▷モデルの評価・予測

6．数量化 II 類

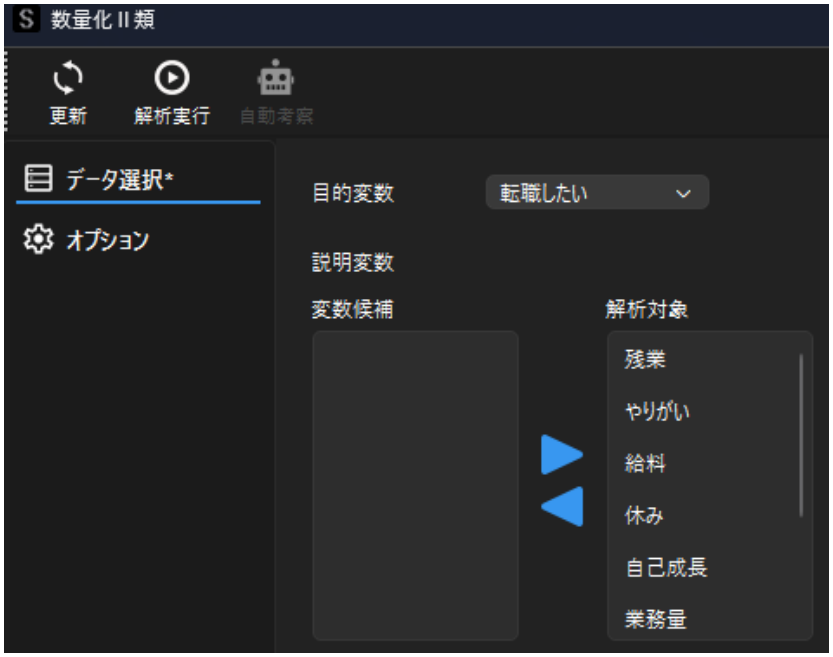
以下のサンプルデータを用いて解説します．サンプルデータは数値データですが，数量化 II 類は文字列データに対しても実行可能です．

※ 数量化 II 類では数値データ（例えば，1,2,3）といった値も，全て文字列データ（例えば，A,B,C）として扱い計算を行います．



	転職したい	残業	やりがい	給料	休み	自己成長	業務量
1	4	2	4	4	1	3	2
2	3	5	2	1	3	1	4
3	5	2	3	4	3	4	2
4	2	2	2	2	2	3	2
5	5	5	4	3	4	4	5
6	4	1	4	4	2	4	3
7	1	3	1	1	1	1	2
8	4	5	5	3	5	4	5
9	1	1	3	5	1	3	1
10	5	4	4	3	1	4	4
11	2	3	3	3	2	4	4
12	2	1	2	1	3	1	1
13	3	5	2	3	5	2	5
14	1	4	1	1	3	1	2

メニューバーから「多変量解析」→「数量化」→「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，「変数候補」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．

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．解析結果にはモデル全体の指標と軸ごとの結果が表示されます．軸ごとの結果はタブ内に表示されます．

解析結果

モデル全体の指標

サンプル数	群数	相関比
20	5	0.8

各軸の情報　説明変数の重要度　カテゴリースコア

軸番号	固有値	寄与率	累積寄与率
1	1.0	0.25	0.25
2	1.0	0.25	0.5
3	1.0	0.25	0.75
4	1.0	0.25	1.0
5	0.0	0.0	1.0
6	0.0	0.0	1.0
7	0.0	0.0	1.0
8	0.0	0.0	1.0
9	0.0	0.0	1.0
10	0.0	0.0	1.0
11	0.0	0.0	1.0

解析結果タブで「カテゴリースコア」を選択した場合は，カテゴリーごとのカテゴリースコアとサンプルサイズが算出されます．

解析結果

モデル全体の指標

サンプル数	群数	相関比
20	5	0.8

各軸の情報　説明変数の重要度　カテゴリースコア

変数名	カテゴリ名	カテゴリスコア	n	%
残業	1	-0.1501	6	30.0
残業	2	-0.2849	3	15.0
残業	3	-0.3144	5	25.0
残業	4	0.0873	2	10.0
残業	5	0.7882	4	20.0
やりがい	1	0.2432	5	25.0
やりがい	2	-0.9754	6	30.0
やりがい	3	-0.2274	3	15.0
やりがい	4	0.0579	5	25.0
やりがい	5	5.0291	1	5.0
給料	1	-0.1558	6	30.0

7. 数量化III類

以下のサンプルデータを用いて解説します．

	性別	副業有無	業界	結婚有無	居住地
1	男性	なし	製造	既婚	地方
2	男性	あり	流通	未婚	地方
3	男性	なし	製造	既婚	地方
4	女性	あり	金融	既婚	都市
5	男性	なし	公務員	既婚	地方
6	男性	あり	IT	既婚	都市
7	男性	なし	公務員	既婚	都市
8	女性	なし	流通	既婚	都市
9	女性	なし	金融	既婚	地方
10	男性	あり	製造	既婚	都市
11	男性	なし	製造	既婚	地方
12	女性	あり	流通	未婚	地方
13	男性	あり	IT	既婚	都市
14	女性	なし	公務員	未婚	地方

メニューバーから「多変量解析」→「数量化」→「数量化III類」を選択します．数量化III類用ウィンドウが表示されたら，変数を選択します．例ではサンプルデータの全ての変数（列）を選択します．

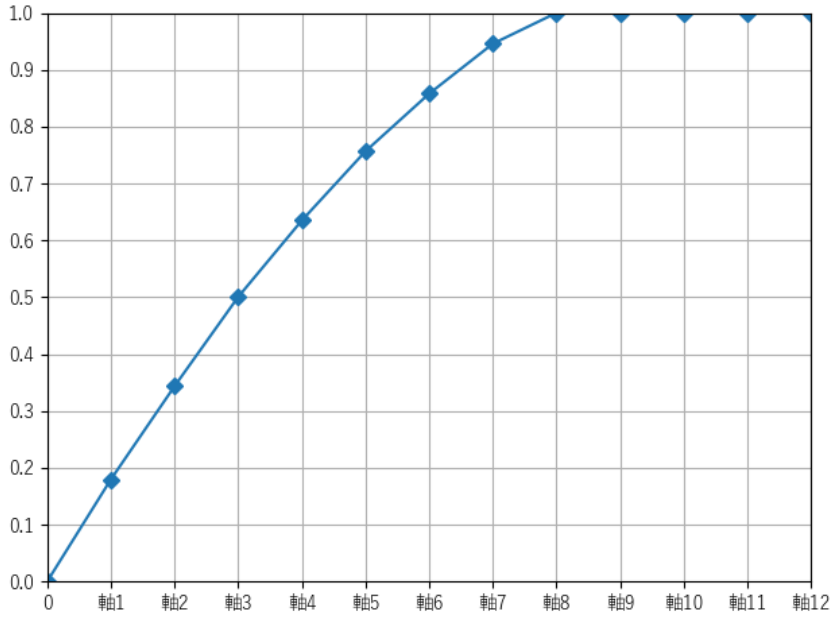
グラフ設定画面の詳細設定では，散布図のマーカーの色や種類を変更することができます．

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ます．解析結果は「各軸の情報」「カテゴリースコア」「サンプルスコア」のタブごと表示されます．

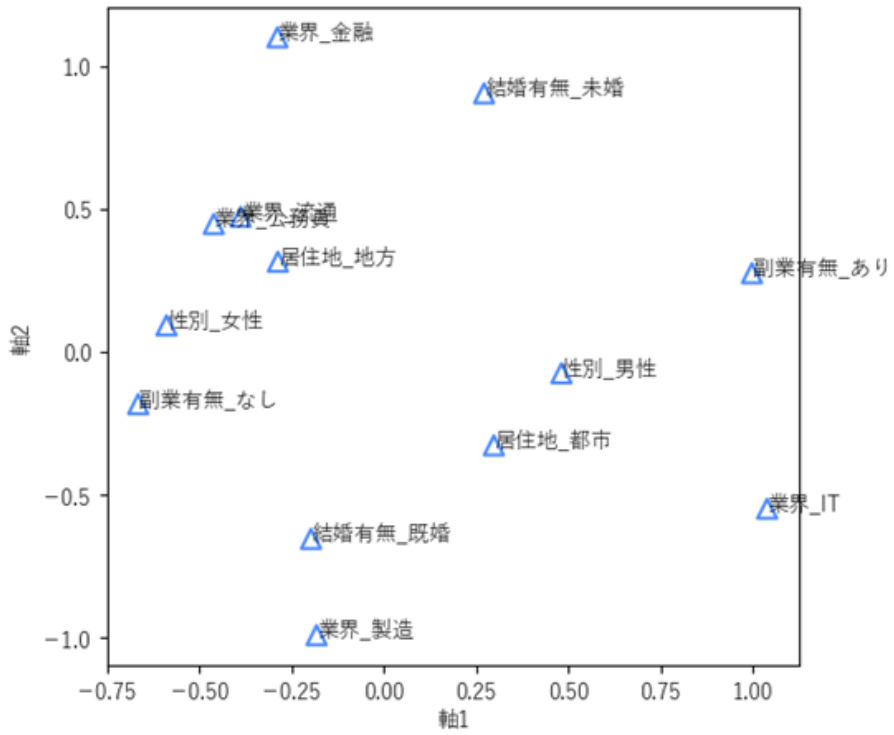
	固有値	寄与率	累積寄与率
軸1	0.287	0.1794	0.1794
軸2	0.262	0.1639	0.3433
軸3	0.251	0.157	0.5004
軸4	0.215	0.1346	0.6349
軸5	0.195	0.1216	0.7566
軸6	0.162	0.1013	0.8579
軸7	0.141	0.0878	0.9457
軸8	0.087	0.0543	1.0
軸9	0.0	0.0	1.0
軸10	0.0	0.0	1.0
軸11	0.0	0.0	1.0

グラフ設定画面で，「寄与率」「カテゴリースコア」「サンプルスコア」のグラフ出力の有無を選択可能です．それぞれのグラフは以下のように出力されます．
※ グラフの表示設定はデフォルトから変更しています．

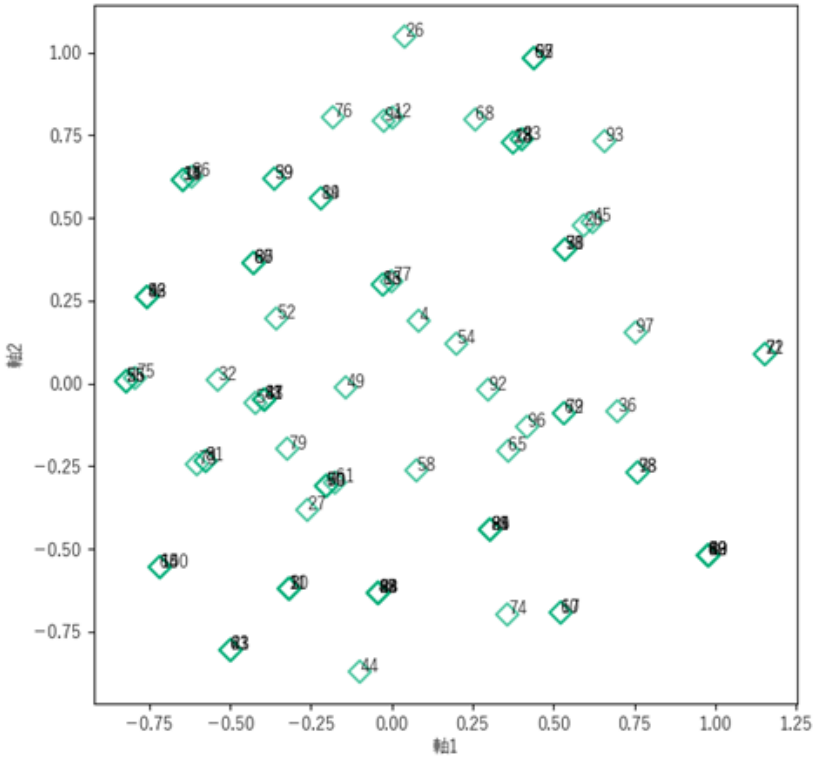
寄与率のプロット例



カテゴリースコアのプロット例



サンプルスコアのプロット例



解析結果はそれぞれクリップボードにコピーすることが可能なので，解析結果は Excel などを用いてグラフ化することも可能です．

8. 主成分分析

以下のサンプルデータを用いて解説します．

	副業有無	収入	性別	年齢	結婚
1	有	580	男性	32	未婚
2	無	430	女性	28	未婚
3	有	800	男性	45	結婚
4	有	780	女性	36	未婚
5	有	690	女性	42	結婚
6	無	510	男性	30	結婚
7	無	740	女性	49	結婚
8	無	350	女性	23	未婚
9	有	620	男性	29	結婚
10	無	500	女性	25	未婚
11	無	430	男性	33	未婚
12	有	590	女性	36	結婚
13	有	1200	男性	51	結婚
14	無	810	男性	53	未婚

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，主成分分析の対象とする変数名を選択します．

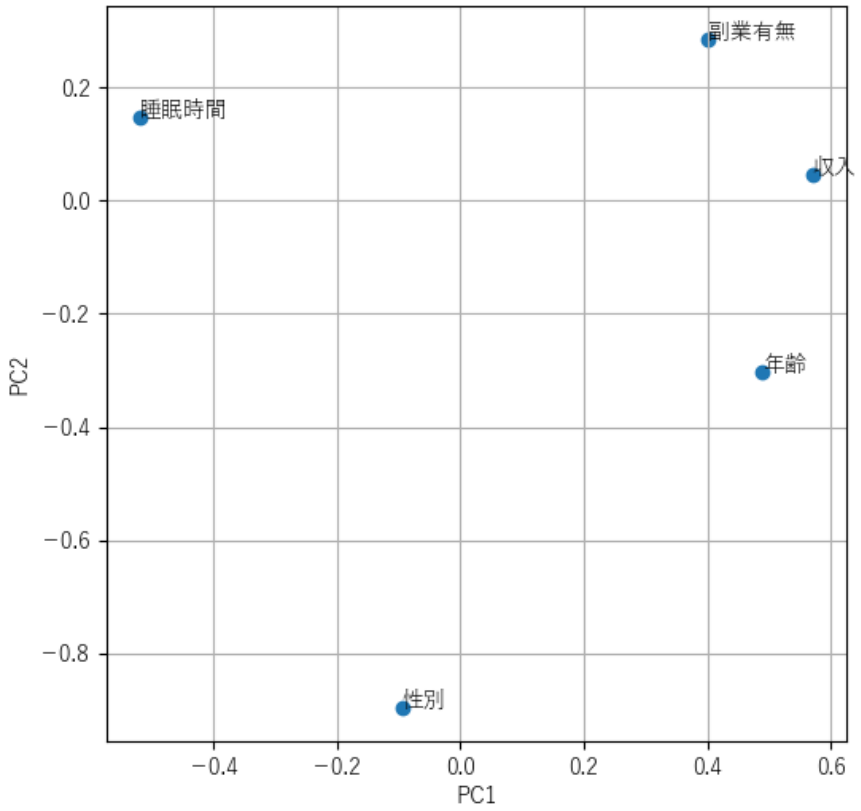
ダミー変数を選択する際は，多重共線性の問題を回避するために１列分を除いて選択します．サンプルデータは既に１列分除いたダミー変数になっています．

「解析実行」ボタンをクリックすると，各タブに解析結果の表示とグラフ設定画面で選択したグラフが表示されます．
デフォルトでは主成分負荷量のプロットと主成分得点のプロットが表示されます．

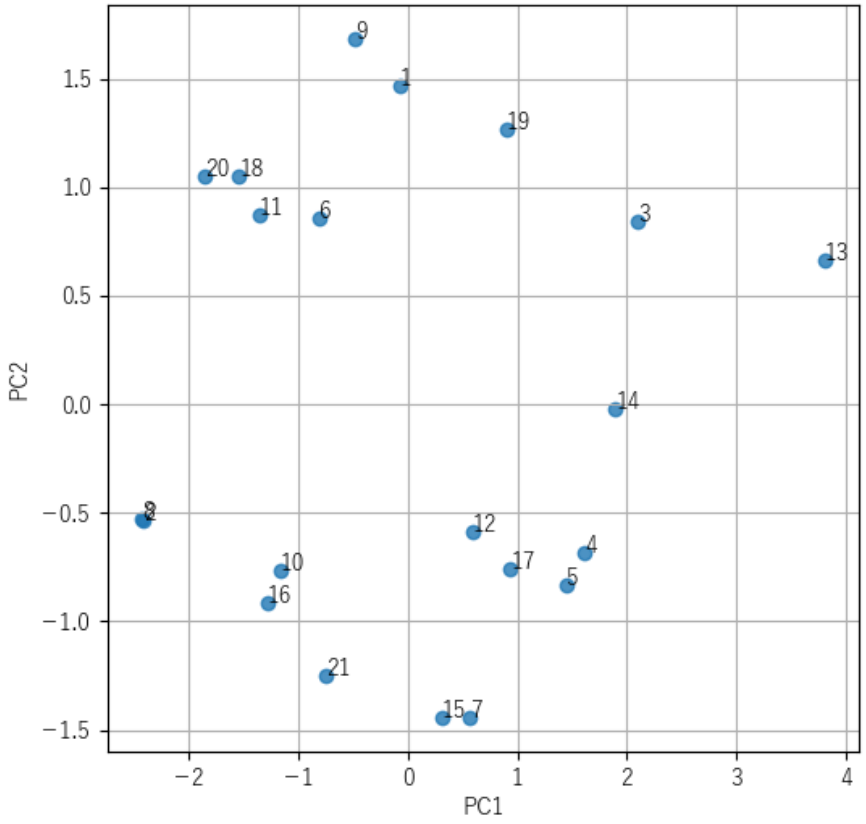
主成分負荷量の解析結果は以下のようになります．

	副業有無	収入	性別	年齢	睡眠時間
PC1	0.3996	0.5703	-0.0943	0.4876	-0.5182
PC2	-0.2851	-0.0458	0.8961	0.3028	-0.1484
PC3	0.7551	-0.1456	0.382	-0.499	-0.1169
PC4	0.3011	0.331	0.1553	0.2826	0.8341
PC5	-0.3133	0.7362	0.1345	-0.5846	-0.006

主成分負荷量のプロット例です．描画対象の軸（主成分）は，グラフ設定画面の軸設定で変更します．例はデフォルト設定で，第一主成分（PC1）と第二主成分（PC2）を選択した場合になります．



主成分得点のプロット例



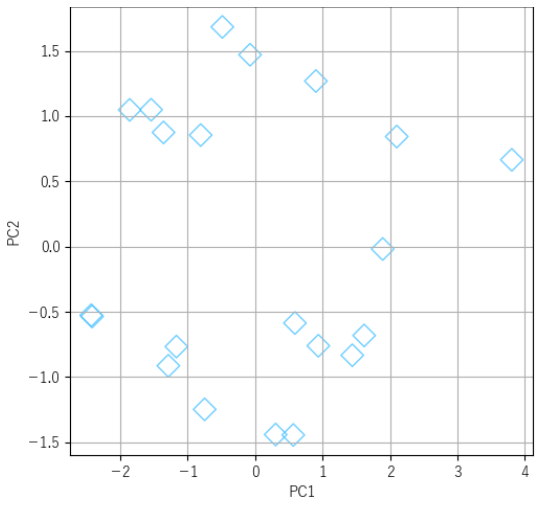
次元圧縮（応用）

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．

オプション画面で「次元圧縮」を ON にして，圧縮数（＝作成する変数の数）を設定します．「解析実行」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．

③ 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．



9. 因子分析

以下のサンプルデータを用いて解説します．

	残業	やりがい	給料	休み	自己成長	業務量
1	2	4	4	1	3	2
2	5	2	1	3	1	4
3	2	3	4	3	4	2
4	2	2	2	2	3	2
5	5	4	3	4	4	5
6	1	4	4	2	4	3
7	3	1	1	1	1	2
8	5	5	3	5	4	5
9	1	3	4	1	3	1
10	4	4	3	1	4	4
11	3	3	3	2	4	4
12	1	2	1	3	1	1
13	5	2	3	5	2	5
14	4	1	1	3	1	2

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，「モデル設定」画面に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．

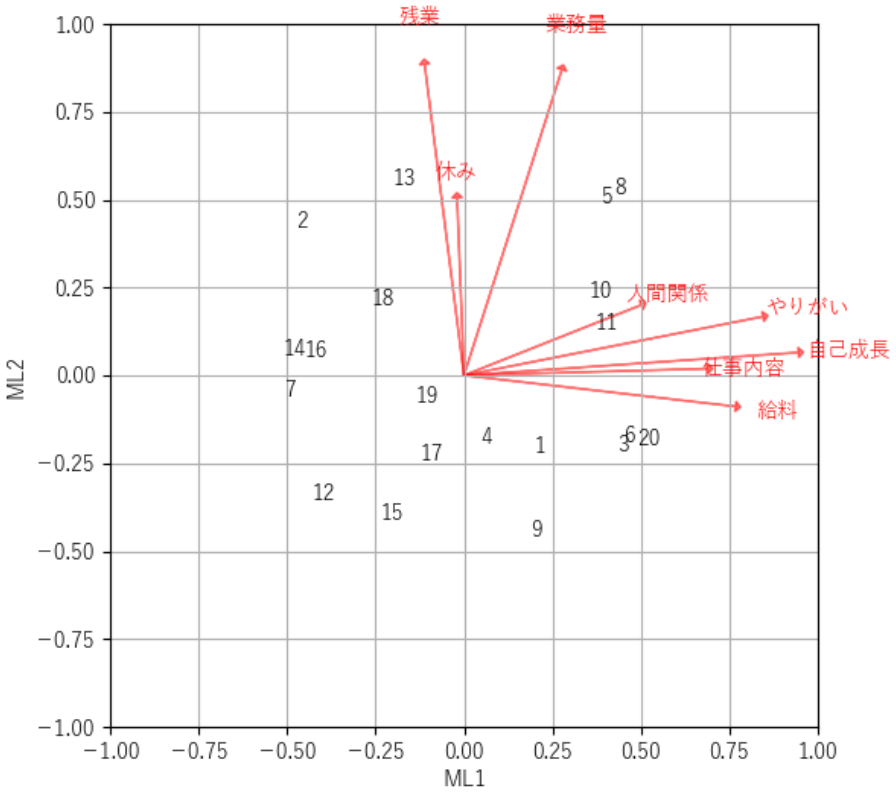
「グラフ設定」画面では因子数・軸の回転方法・因子抽出方法の選択を行います．デフォルトでは以下の値が設定されています．

・因子数：2
・軸の回転方法：varimax（バリマックス回転）

「解析実行」をクリックすると，各タブに解析結果が表示されます．

グラフ設定画面では解析実行時に表示するグラフを設定できます．「寄与率」「因子負荷量」「因子得点」のプロットを表示することができます．

因子得点（バイプロット）の例です．描画対象の軸（主成分）は，グラフ設定画面の軸設定で変更します．例はデフォルト設定で，第一因子（ML1）と第二因子（ML2）を選択した場合になります．



10. 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．

	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

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，変数選択エリアの「測定モデリング」「構造モデリング」「残差相関」画面で変数名を選択してモデル式を追加します．

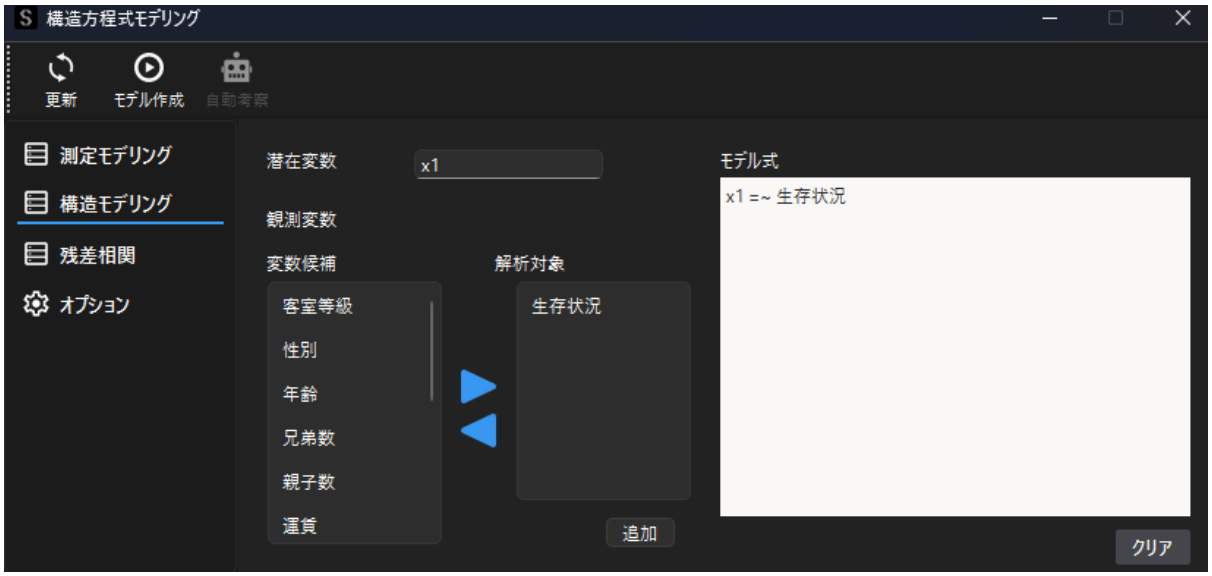
モデル式表示・記述エリアでは追加したモデル式が表示されます．モデル式表示・記述エリアでは自由に入力ができるので，テキストの修正やコピー＆ペーストすることが可能です．

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．

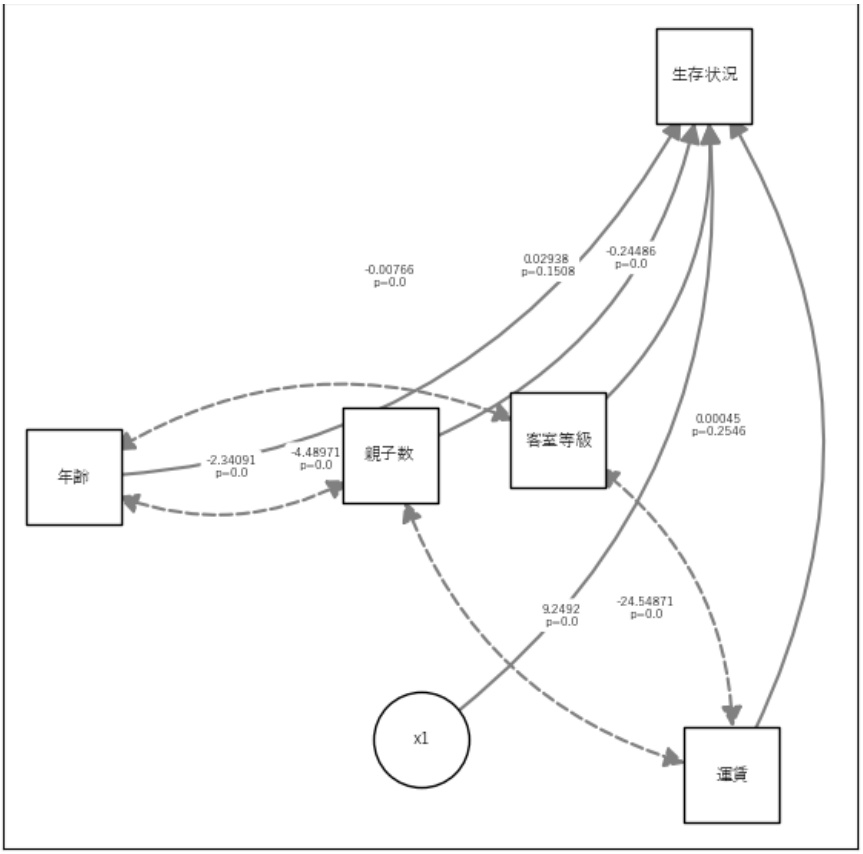


④ 残差相関の定義

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が，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

モデル作成を行なうと以下のような解析結果画面にモデルの適合度やパラメータが表示されます．

※ 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．

11. 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．

	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 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．

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解析結果をコピーして，ホーム画面にペーストします．

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．

基本統計量

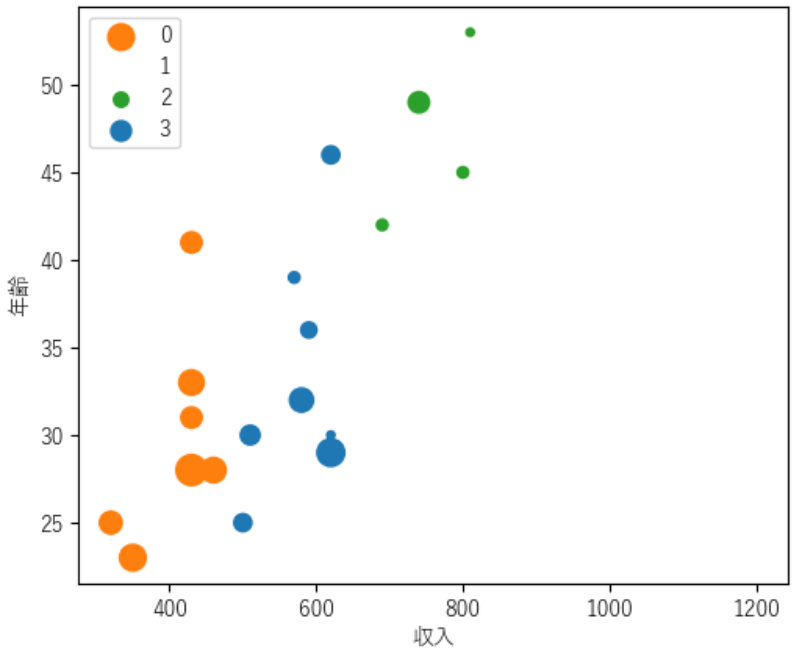
	副業有無	収入	性別	年齢	身長
クラスター1_サンプルサイズ	11.0	11.0	11.0	11.0	11.0
クラスター1_平均	0.3636	557.2727	1.0	36.0	159.6364
クラスター1_標準偏差	0.5045	141.6397	0.0	8.4735	4.6319
クラスター1_最小値	0.0	350.0	1.0	23.0	150.0
クラスター1_第一四分位数	0.0	430.0	1.0	29.5	157.0
クラスター1_中央値	0.0	570.0	1.0	36.0	160.0
クラスター1_第三四分位数	1.0	655.0	1.0	41.5	162.5
クラスター1_最大値	1.0	780.0	1.0	49.0	166.0
クラスター2_サンプルサイズ	10.0	10.0	10.0	10.0	10.0
クラスター2_平均	0.5	635.0	0.0	35.6	171.1
クラスター2_標準偏差	0.527	251.2303	0.0	10.1346	5.1088
クラスター2_最小値	0.0	320.0	0.0	25.0	163.0
クラスター2_第一四分位数	0.0	472.5	0.0	29.25	168.25
クラスター2_中央値	0.5	600.0	0.0	31.0	170.0
クラスター2_第三四分位数	1.0	755.0	0.0	42.0	173.5

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．

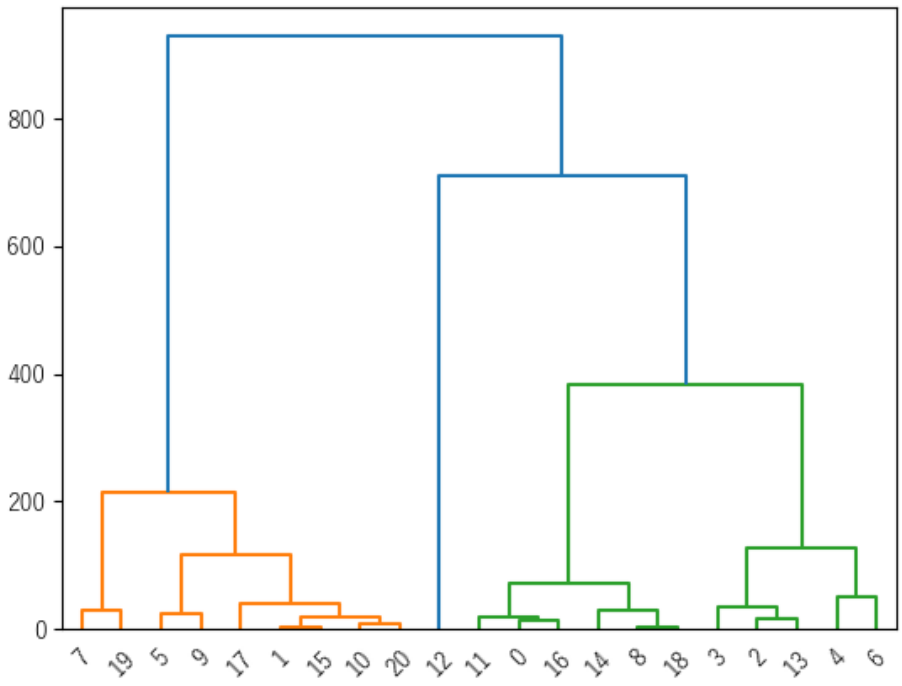


④ 階層クラスター分析

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．

「モデル設定」画面では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．樹形図の分岐が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．

12. 決定木

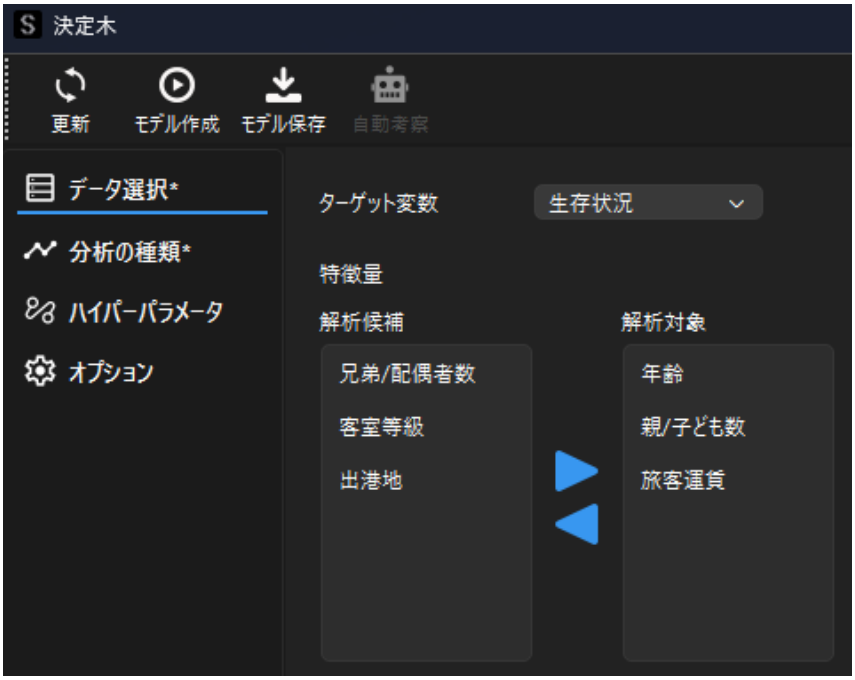
※ ランダムフォレスト・勾配ブースティング決定木についても同様の操作方法で実行可能です．

以下のサンプルデータを用いて解説します．

	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」画面では，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

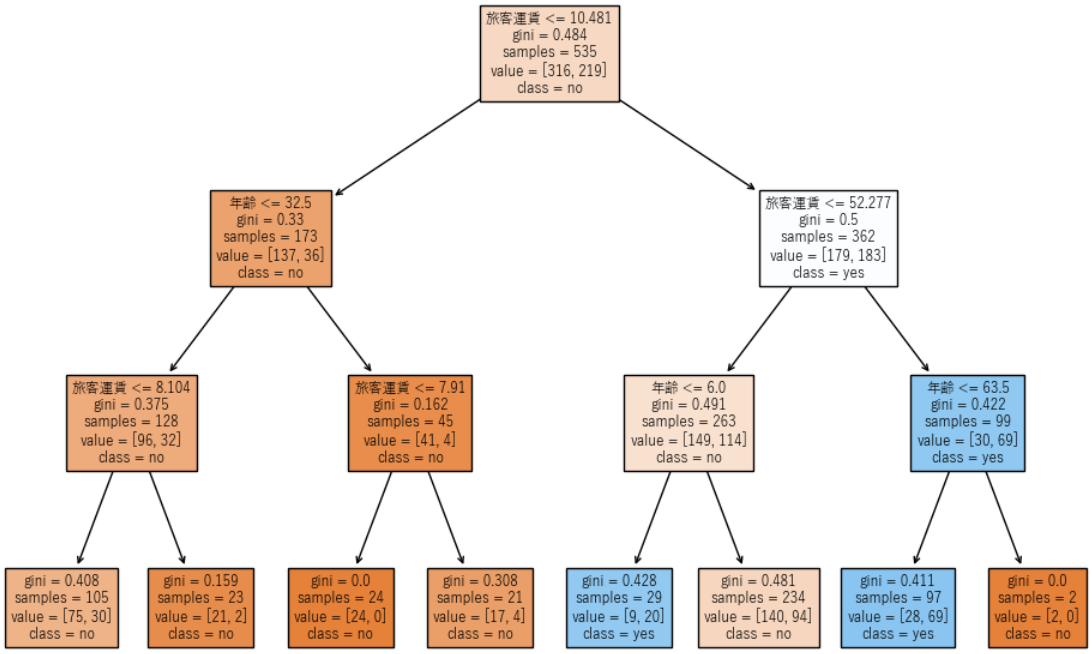
ツールバーの「モデル作成」ボタンでモデルの作成を実行します．分類問題では以下のように，木構造のグラフが作成されます．

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．

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多いです．

- **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・予測値の算出

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う場合は，「モデルの保存」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．

モデルの評価・予測方法の詳細は以下の Web サイトをご覧ください．

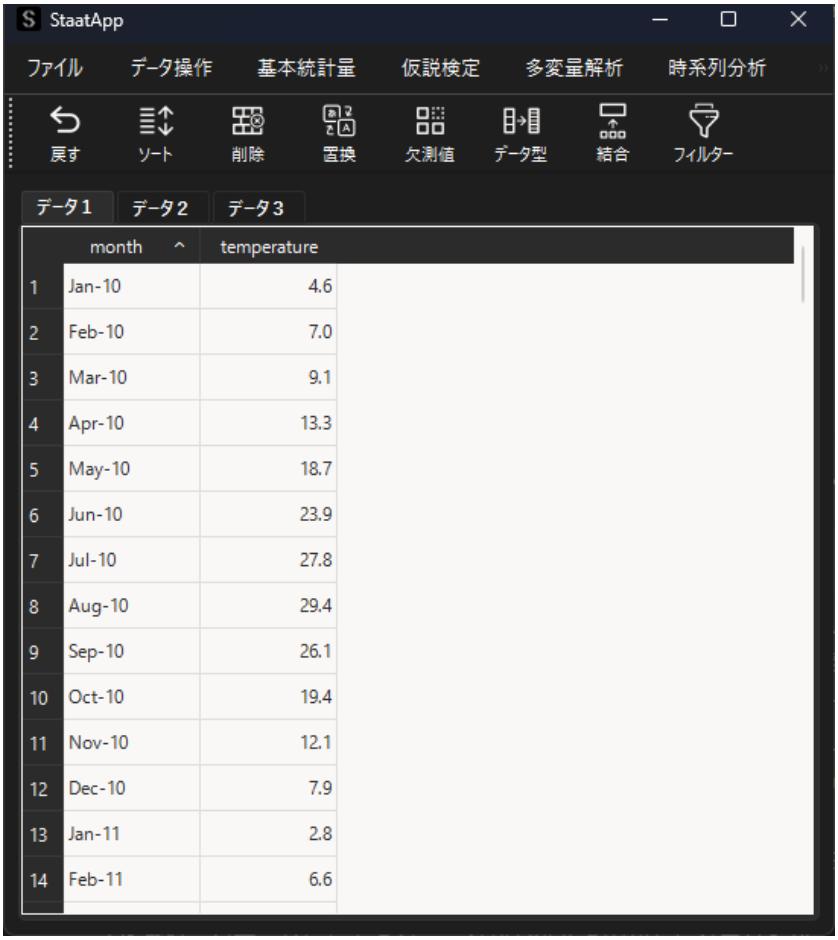
▷モデルの評価・予測

③ 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．回帰問題でも分類問題と同様に重要度が算出されます．

13. ARIMA

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．



	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

※ 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）

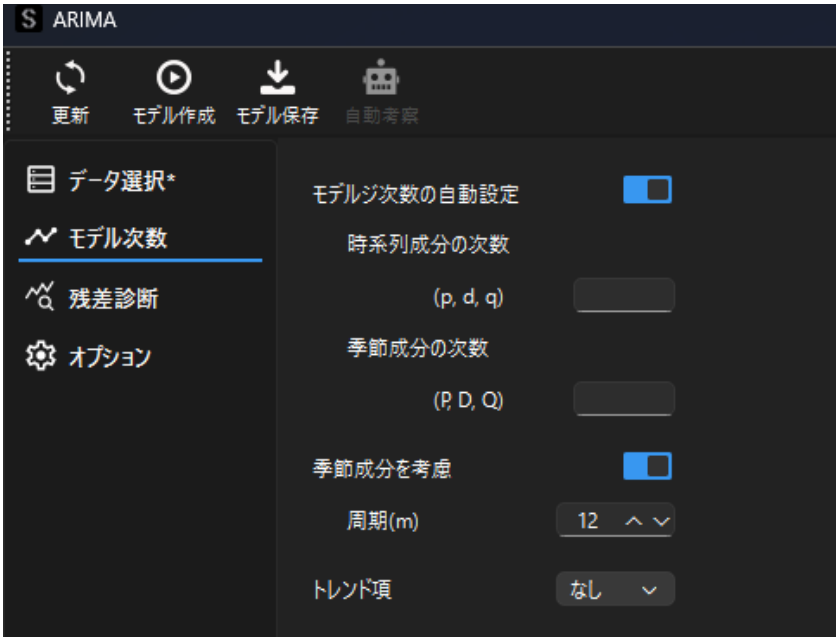
① モデルの作成

メニューバーから「時系列分析」→「モデル」→「ARIMA」を選択します．

「データ選択」画面で目的変数を選択します．外生変数（ARIMAX）を設定する場合は，変数候補から選択を行います．今回は外生変数を設定せずに，気温を目的変数として解析を行います．

次に「モデル次数」画面で，モデルの次数を設定します．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．
※自動設定では，AIC が最小になるような次数が設定されます．



「モデル作成」ボタンをクリックするとモデルが作成され解析結果が表示されます．

解析結果

モデルの仕様

(p,d,q)	(P,D,Q,m)	トレンド	外生変数
(0, 0, 1)	(2, 1, 1, 12)	なし	なし

モデルの評価指標

n	logLik	AIC	BIC	HQIC	MAE	MSE
156	-181.6838	373.3675	387.305	379.0276	2.2649	28.3335

残差診断

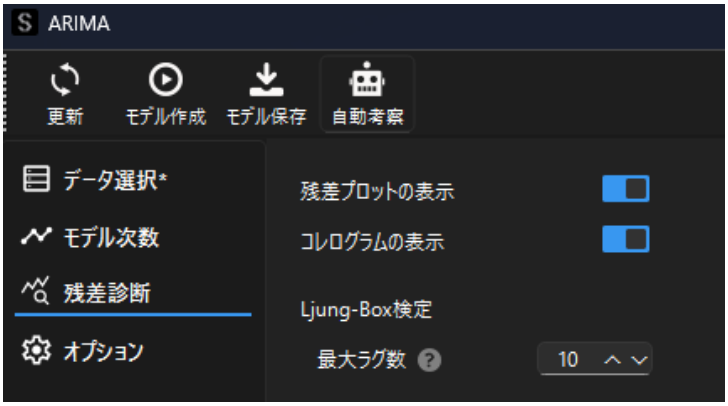
検定	統計量	p値
Ljung-Box	396.0373	0.0

係数

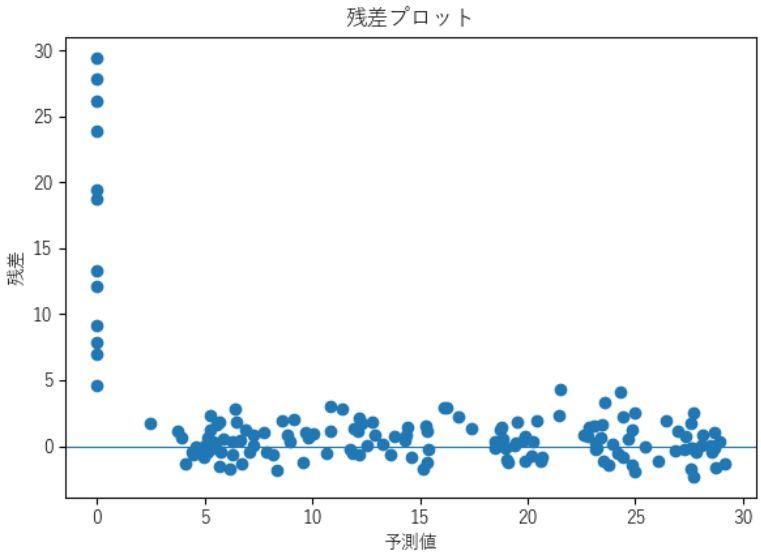
種類	パラメータ	推定値	標準誤差	z	p値	CI下限	CI上限
MA	ma.L1	0.3127	0.0869	3.5993	0.0003	0.1424	0.48
季節AR	ar.S.L12	-0.2043	0.1629	-1.2541	0.2098	-0.5235	0.11
季節AR	ar.S.L24	-0.2922	0.1376	-2.1235	0.0337	-0.5618	-0.022
季節MA	ma.S.L12	-0.5603	0.1657	-3.3823	0.0007	-0.885	-0.235

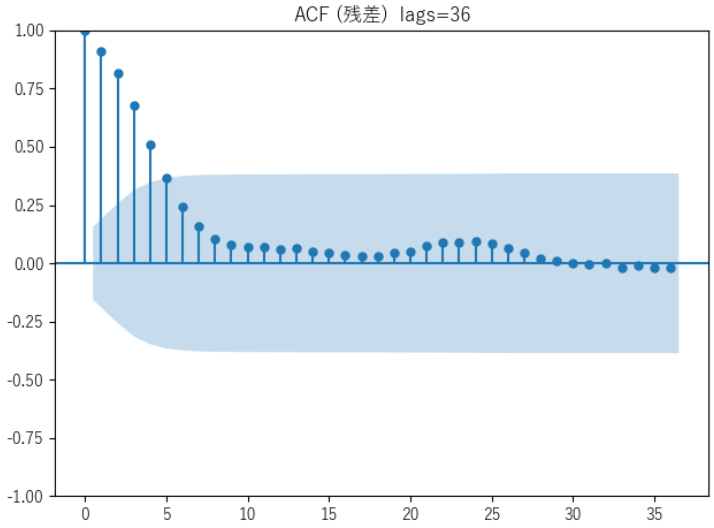
② 残差診断

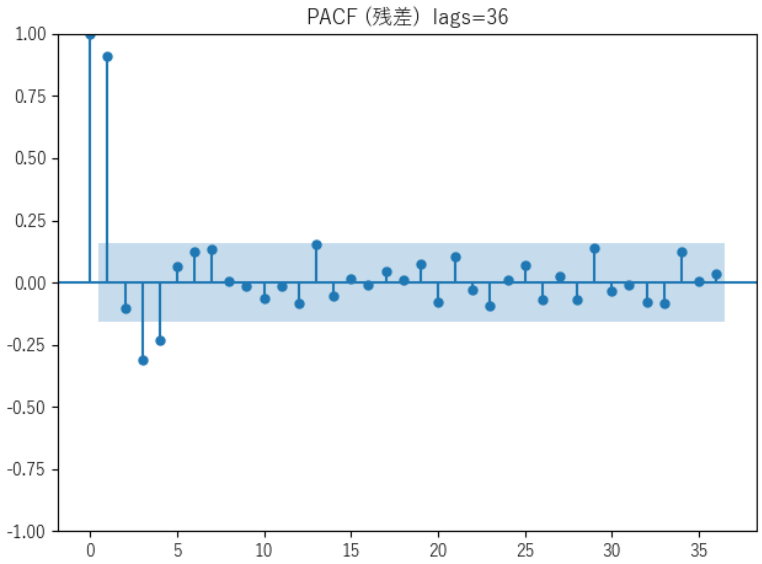
残差プロットやコレログラムを作成する場合は，「残差診断」画面でそれぞれ ON にします．



モデル作成時に以下のように残差プロットと ACF・PACF のコレログラムが表示されます．



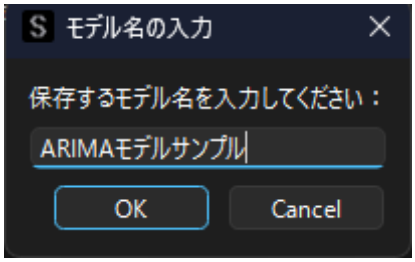




青帯の範囲の外にある値は，有意水準α=0.05 で無相関の検定における有意を示します．ラグ 5 以下の値を除いて，周期ごとに偏自己相関係数が有意となっていますがそれほど大きな値ではないため比較的あてはまりのよいモデルと評価することができます．

③ 将来予測

作成したモデルから将来予測を行います．予測機能を用いるためには作成したモデルを保存します．「モデル保存 」ボタンを押して，モデル名を入力します．



「OK」ボタンを押して，保存先のフォルダを選択します．他の端末などに保存したモデルを共有したい場合は，「Yes」選択して保存するフォルダを選択してください．
指定しない場合は，StaatApp のデフォルトフォルダに保存されます．

保存が完了したらホーム画面から「時系列分析」→「時系列モデルを用いた予測」を選択して，時系列モデルの予測用画面を表示します．

一覧から先程保存したモデルを選択します．過去に保存したモデルも選択可能です．

将来予測を行う場合，「予測設定」から予測ステップ数を選択します．予測開始時点は「最新時点」を選択します．

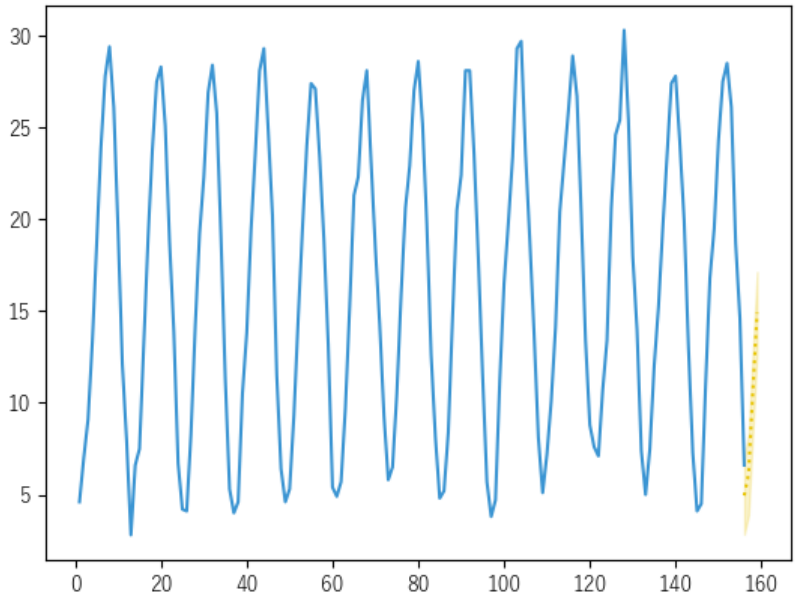
予測グラフに学習データを含めたい場合は，アクティブデータに学習時のデータをセットして「比較データ」で比較対象に選択します．
予測値のみ表示する場合は，データ及び選択は不要です．

外生変数（ARIMAX）を含めてモデルを作成した場合，作成時と同じ列名 を含む予測ステップ数（行数）の外生変数をセットして，データセットとして選択します．もしくは，予測用の外生変数がない場合は，「最終値を保持」を選択します．

「実行」ボタンをクリックすると，予測値とグラフが表示されます．

予測結果

ステップ	予測値	95%CI 下限	95%CI 上限
157	4.9505	2.8361	7.0648
158	6.0891	3.8738	8.3045
159	10.589	8.3737	12.8044
160	14.9125	12.6972	17.1279
161	20.1572	17.9419	22.3725
162	23.8517	21.6364	26.067
163	26.7258	24.5105	28.9411
164	29.1942	26.9789	31.4095
165	25.3905	23.1752	27.6058
166	18.8292	16.6138	21.0445



④ ホールドアウト検証

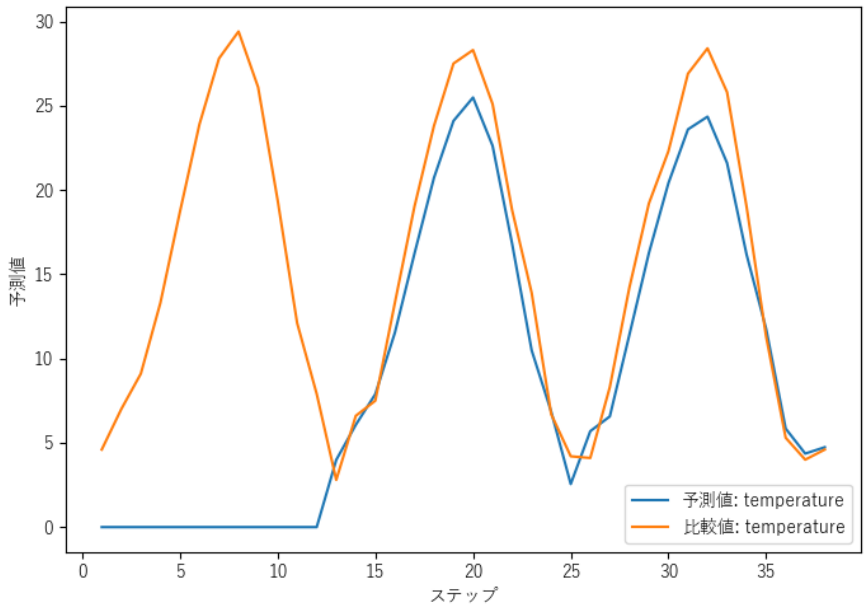
StaatApp ではモデル評価の方法として，事前に学習データとテストデータを分割して，学習データから作成したモデルの予測値とテストデータを比較するホールドアウト検証が可能です．

ARIMA 機能を用いる前に，読み込ませたデータを「データ操作」→「テストデータの作成」機能で分割を行います．学習時は分割したデータでモデルを作成し，保存します．

アクティブデータに比較用のテストデータを設定して，「予測設定」で比較ステップ数（テストデータのステップ数）だけ「予測ステップ数」に設定します．「予測開始時点」は「任意時点」として，1 を入力して「比較データ」にテストデータの比較する列名を選択します．

今回であれば，「テストデータの作成」で 25%を分割してテストデータにしたので，テストデータは 38 ステップあるため，以下のように「予測設定」で設定します．

「比較データ」に比較対象を選択して実行すると，以下のようにテストデータと予測データ描画したグラフが表示されます．



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．

14. 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．

	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

「時系列分析」→「ADF 検定」を選択して，ADF 検定を行い定常過程かどうか判定します．

	c	ct	ctt
bct終値	0.1732	0.342	0.5518

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，「データ操作」→「差分対数変換」を選択して差分変換を行います．

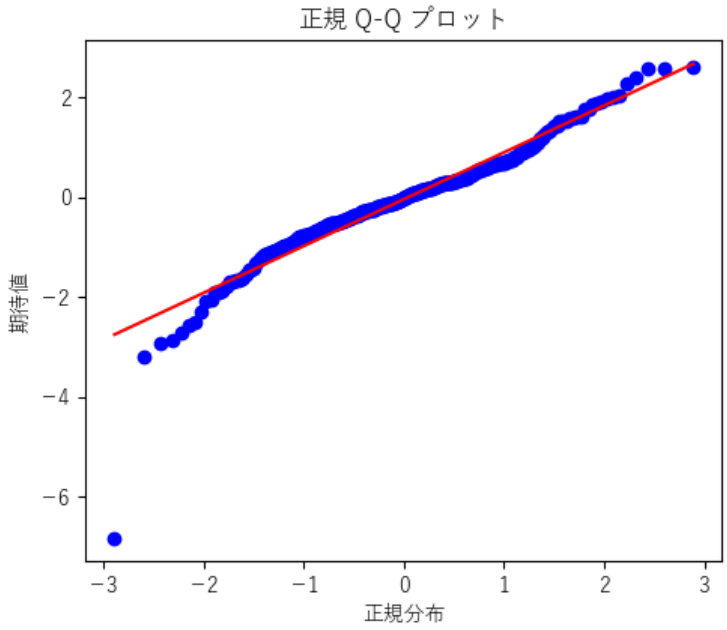
② モデルの作成・評価

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．

「モデル設定」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．

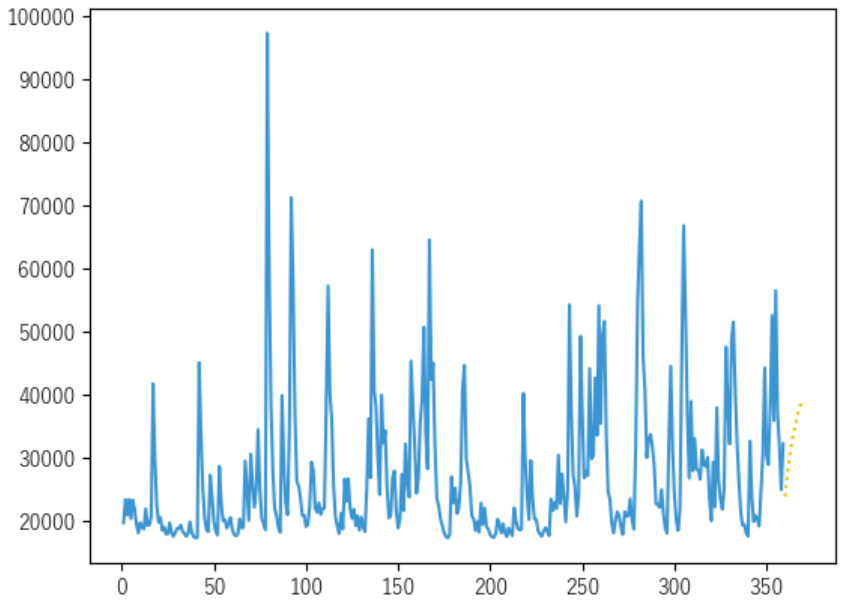
モデルが作成されたら評価指標を見ます．特に GARCH モデルではモデル式の係数の p 値に注目します．この値が小さいほどモデル式の各項は統計学的に意味がある値であることを示し，モデルの当てはまりが良い言えます．

モデル作成時に「オプション」画面で「標準化残差の分布」を ON にすると，正規 Q-Q プロットで正規分布と比較した標準化残差の偏りを確認できます．



③ ボラティリティの描画

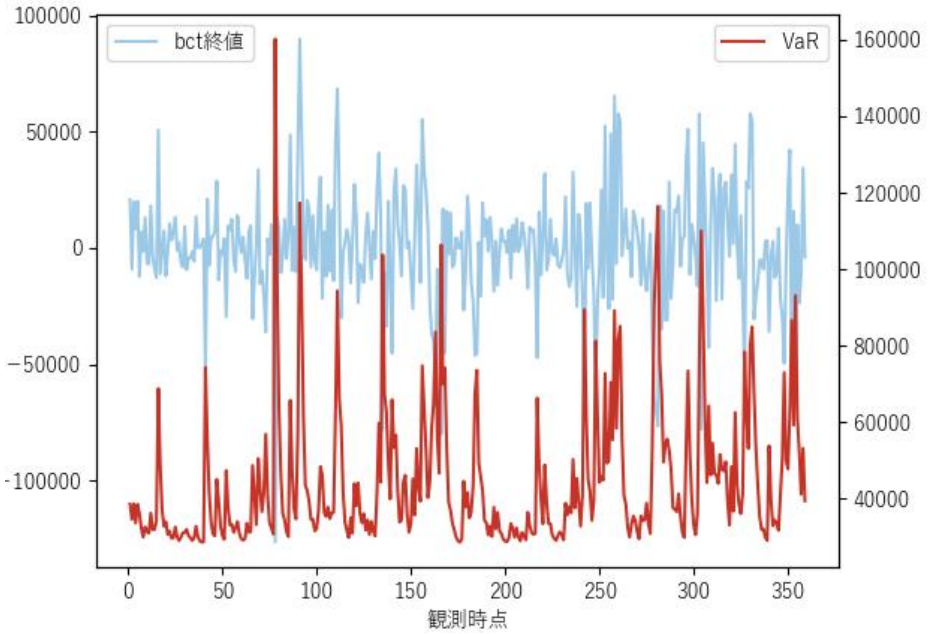
ボラティリティはモデル作成時に自動で算出 されます．グラフ作成したい場合は，「ボラティリティ」画面で，「ボラティリティの描画」を ON にします．



④ VaR の算出

VaR の将来予測（観測時点の次の時点の値）は解析結果に表示されます．
※ StaatApp では正規分布を用いて VaR を算出しています．

グラフ作成する場合は，「VaR」画面で「VaR の描画」を ON にしてください．



有意水準α=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．

15. 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．

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

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．

「データ選択」で 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ることができます．


解析結果

モデルの仕様

n	ラグ選択基準	採用ラグ	トレンド	変数
199	aic	3	c	realgdp, realcons, realinv

モデルの評価指標

AIC	BIC	FPE	HQIC	LogLik	det(Sigma_u)
20.1135	20.6099	543598711....	20.3144	-2818.3963	469244066....

残差診断

系列	ラグ	LB統計量	p値
realgdp	10	5.9431	0.82
realcons	10	3.8013	0.9559
realinv	10	8.8019	0.551

グレンジャー因果性検定

原因	検定統計量	p値	自由度1
realcons	21.5822	0.0	3
realinv	2.3187	0.0745	3

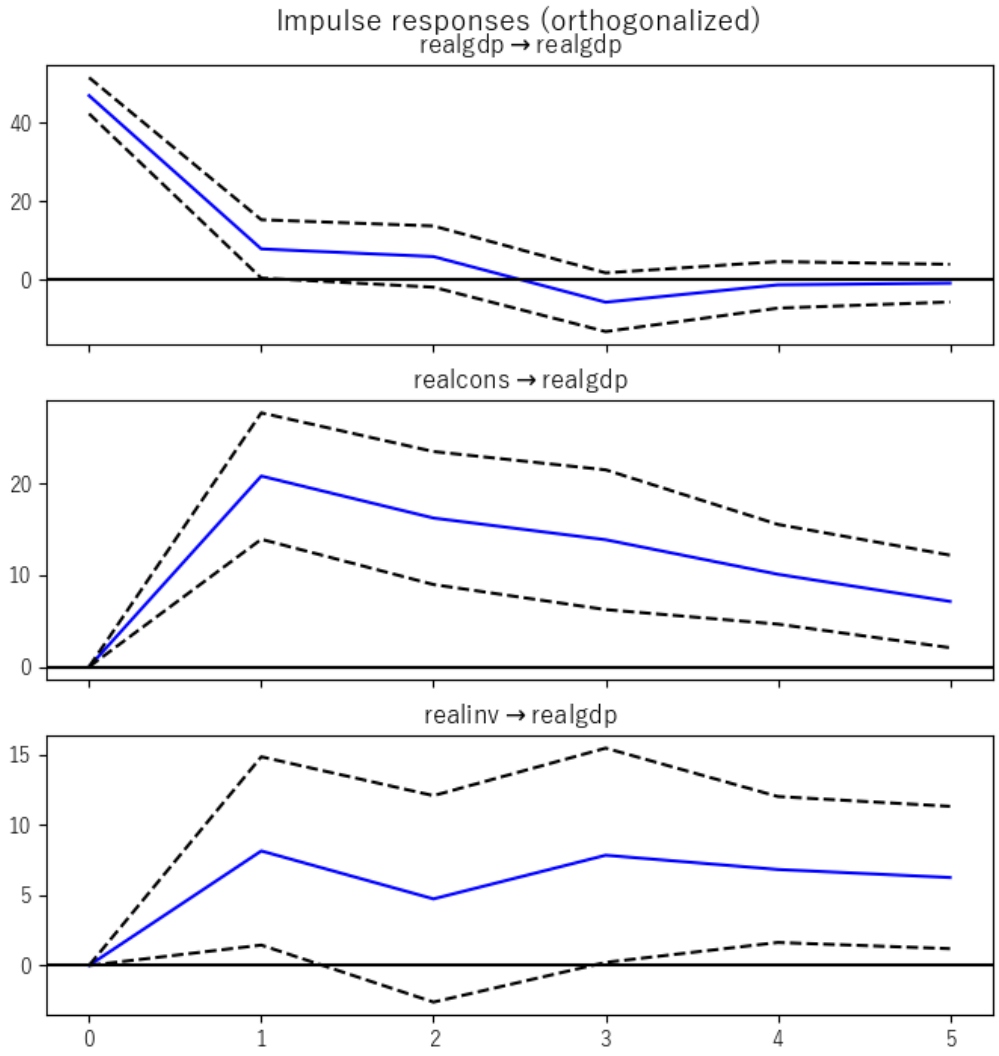
VAR モデル機能では ARIMA 機能と同様に将来予測やホールドアウト検証や残差分析も行うことが可能です．

② 因果分析

解析結果ではグレンジャー因果性検定の結果が表示されます．

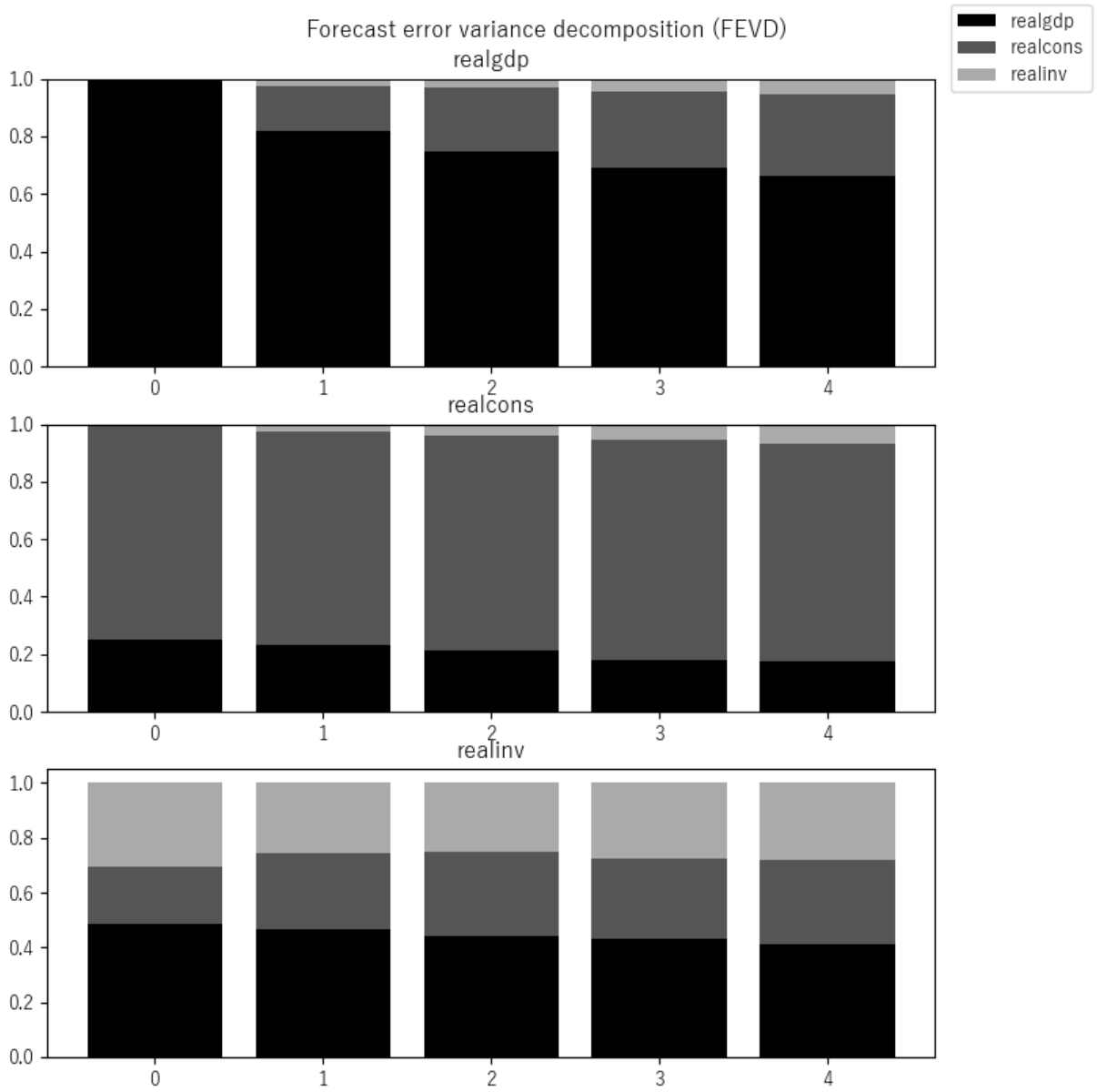
インパルス応答分析を行いたい場合は，「因果分析」画面で「インパルス応答分析」を ON にします．インパルス応答関数は，一つの時系列変数に対する"衝撃"が他の時系列変数に与える効果を時間経過に従って追跡するための分析手法です．

VAR 機能の設定で「評価・予測対象」に"realgdp"を設定しているので，"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）を行う場合は，「予測誤差分散分解（FEVD）」を ON にします． 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．

16. Prophet

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．

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

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．

「データ操作」→「日時データの作成」で以下のダイアログを表示します．

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．

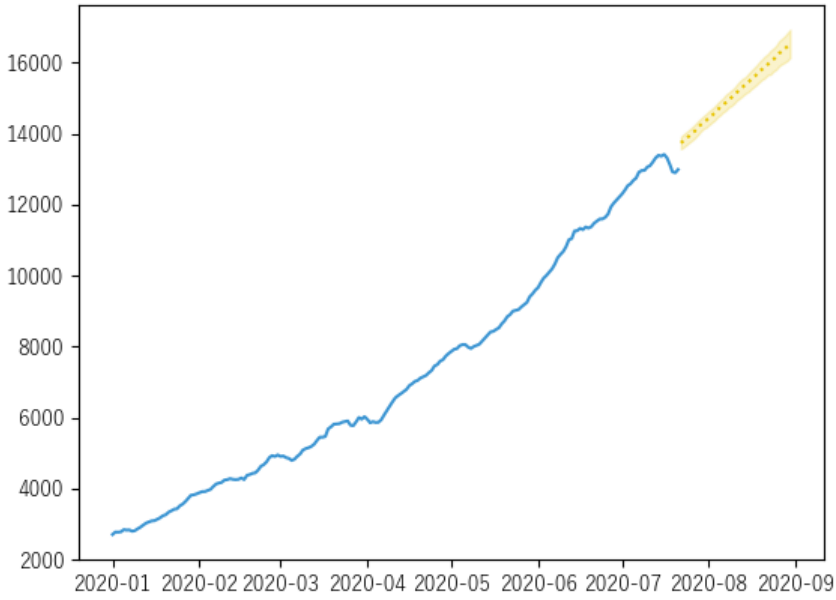
因みに Prophet で推奨されている日時データの形式は，"YYYY-MM-DD"もしくは"YYYY-MM-DD hh:mm:ss"になります．

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）や予測値が表示されます．

「予測設定」で「予測グラフの作成」を ON すると，以下のような学習データ後の予測データを折れ線グラフで描画することができます．



17. 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．

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

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．

プレビュー画面に，クロス集計表が表示されます．

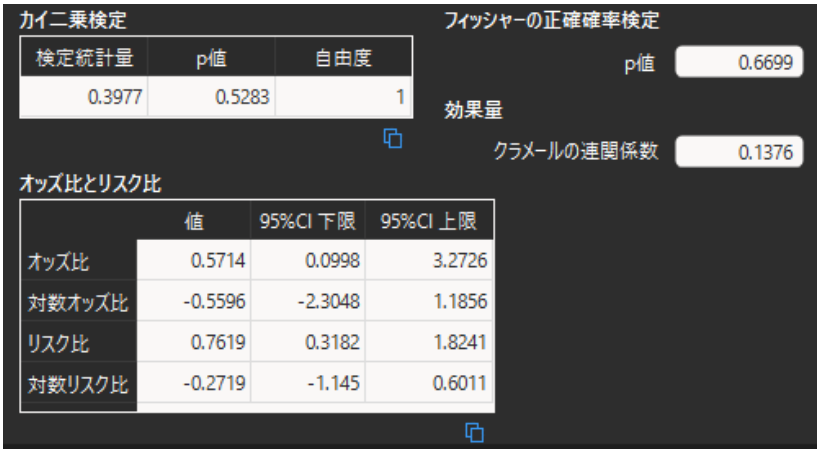
「解析実行」ボタンをクリックすると，解析結果画面に表示されます．

② カイ二乗検定

カイ二乗検定はデフォルトでクロス集計表の作成時に実行されます．効果量としてクラメールの連関係数も算出されます．

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 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．2×2 クロス集計表を作成した場合は，解析実行時にフィッシャー正確確率検定の p 値が算出されます．



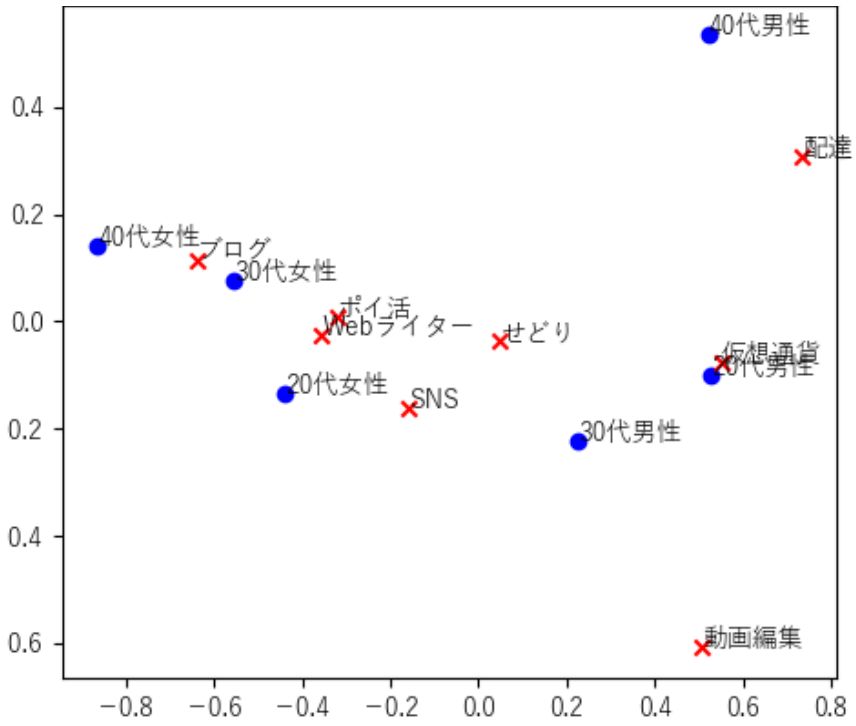
④ オッズ比とリスク比

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．

⑤ コレスポンデンス分析

コレスポンデンス分析を実行する場合は，クロス集計表の「オプション」画面から「コレスポンデンス分析の実行」を ON にします．

ON にした状態で，「解析実行」を行うと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，クロス集計表のプレビューに貼り付けます．

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		
		Yes	No	合計
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．

18. 生存曲線

以下のサンプルデータを用いて解説します．

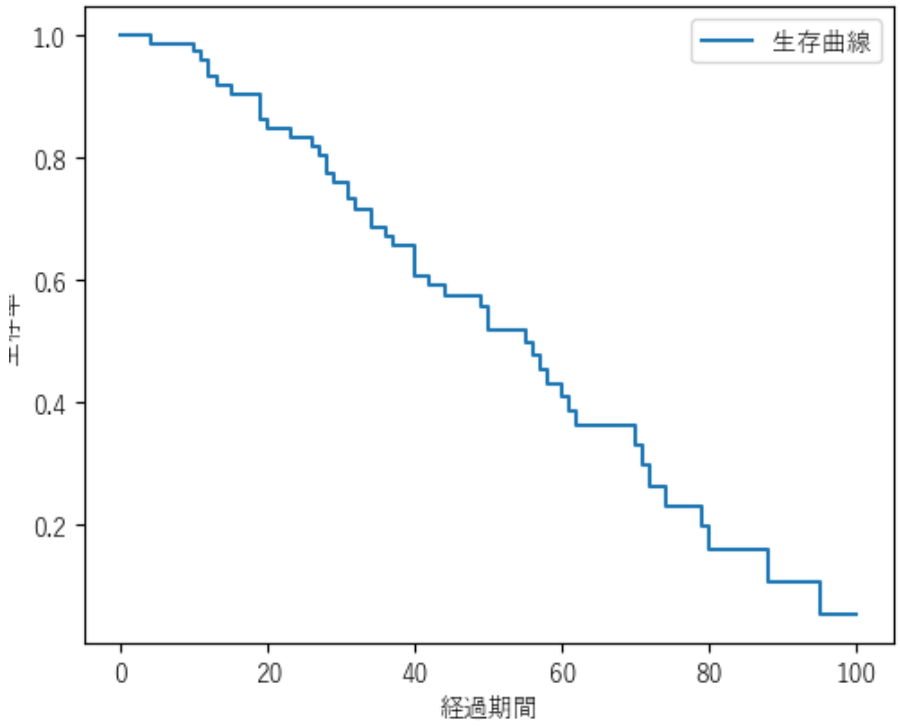
生存曲線を作成するためには，イベント発生までの経過時間を表すデータ（経過日数）と，イベント発生の有無を示すデータ（死亡状況）が必要です．イベント発生の有無を示すデータは，イベント発生した場合→"1"，打ち切りの場合"0"とします．

	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

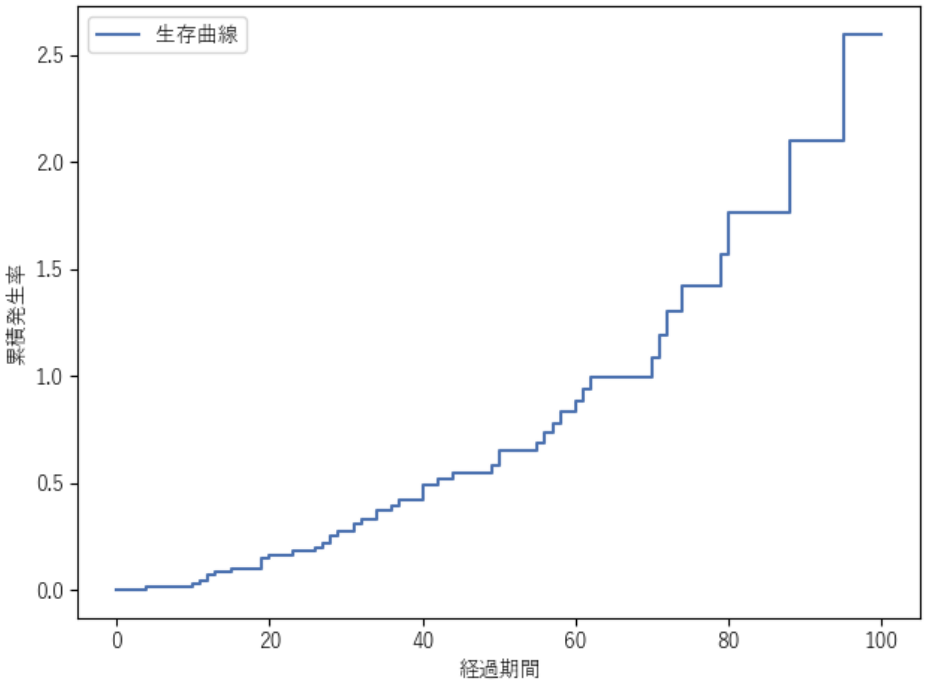
① 生存曲線の作成

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」と「イベント発生」として設定する変数をコンボボックスから選択します．

グラフ設定画面で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．

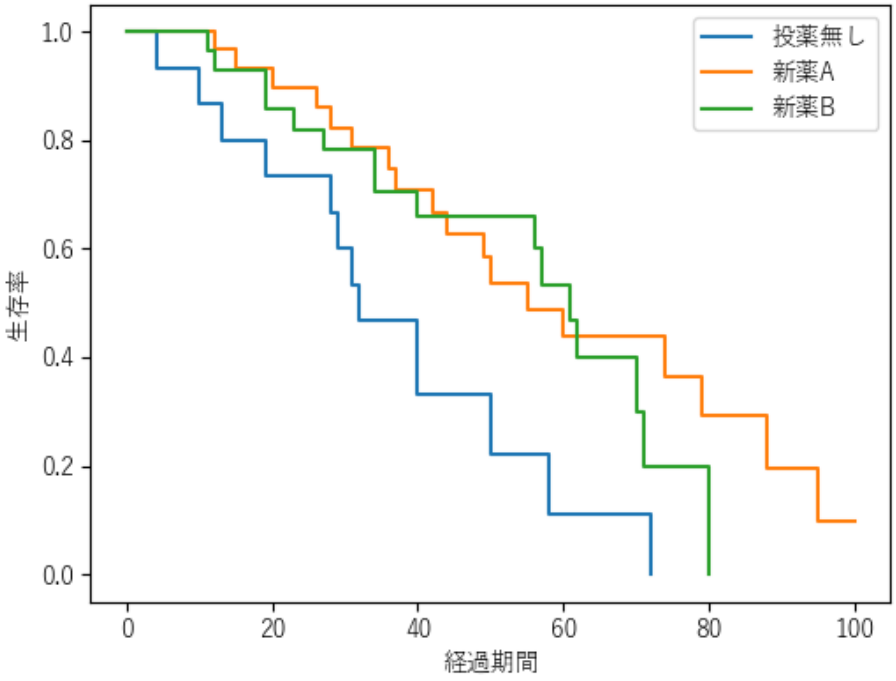


グラフ設定画面で累積発生率を選択した場合は，以下のようなグラフが作成されます．

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「群列名」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．



解析結果

追跡期間の要約

群	n	イベント数	打ち切り数	中央値	最小	最大
投薬無し	15	13	2	32.0	4.0	72.0
新薬A	30	18	12	46.5	12.0	100.0
新薬B	30	16	14	43.0	6.0	80.0

中央値生存期間

群	中央値生存期間	95%CI下限	95%CI上限
投薬無し	32.0	13.0	50.0
新薬A	55.0	37.0	88.0
新薬B	61.0	34.0	71.0

ログランク検定

検定統計量	9.1894
p値	0.0101

一般化ウィルコクソン検定

検定統計量	7.0923
p値	0.0288

19. 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．

	No.	継続期間	イベント発生	収入	性別	年齢	結婚
1	1	31	1	580	男性	32	未婚
2	2	97	1	430	女性	28	未婚
3	3	126	0	800	男性	45	既婚
4	4	56	0	780	女性	36	未婚
5	5	172	1	690	女性	42	既婚
6	6	156	0	510	男性	30	既婚
7	7	133	1	740	女性	49	既婚
8	8	82	1	350	女性	23	未婚
9	9	173	1	620	男性	29	既婚
10	10	110	0	500	女性	25	未婚
11	11	33	1	430	男性	33	未婚
12	12	156	1	590	女性	36	既婚
13	13	82	0	1200	男性	51	既婚
14	14	112	1	810	男性	53	未婚

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．

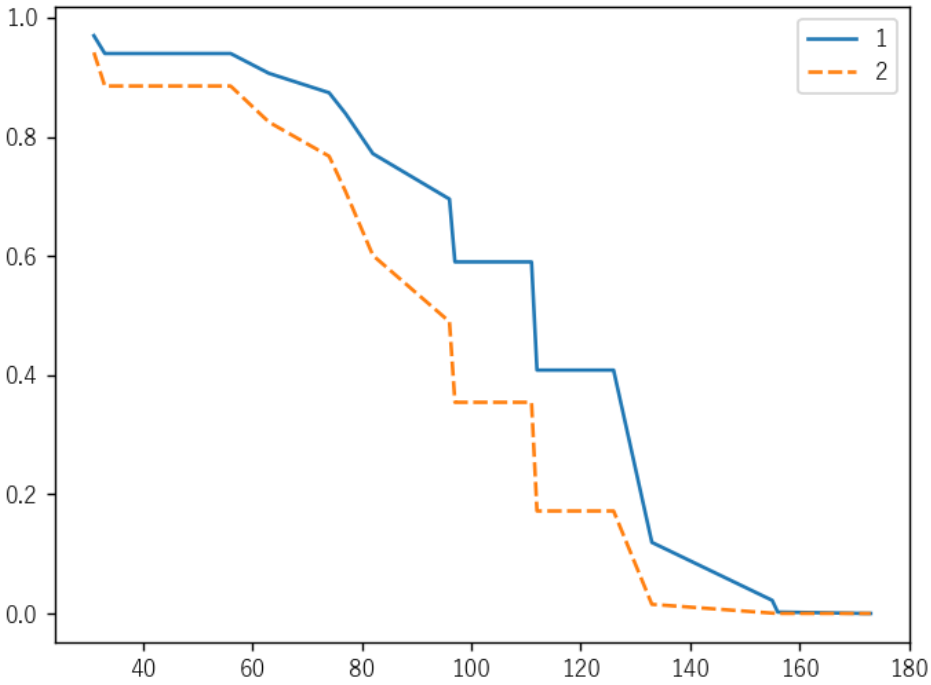
② 予測値の算出（応用）

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．


1.0
0.8
0.6
0.4
0.2
0.0
40
60
80
100
120
140
160
180
1
2

20. 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．

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

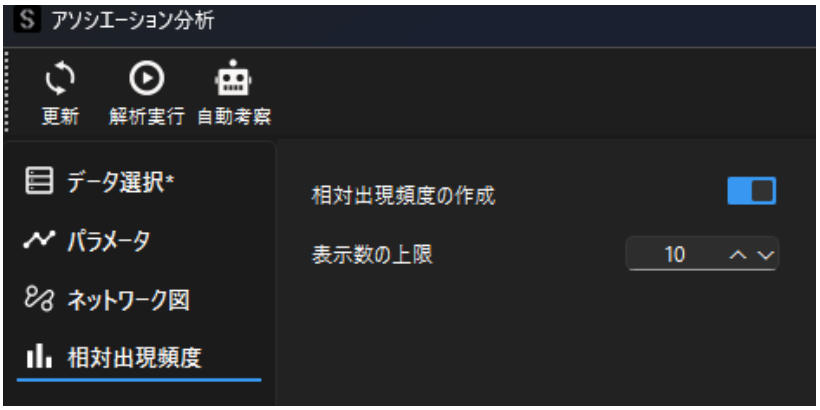
「解析実行」ボタンをクリックすると，アソシエーション分析の解析結果が出力されます．

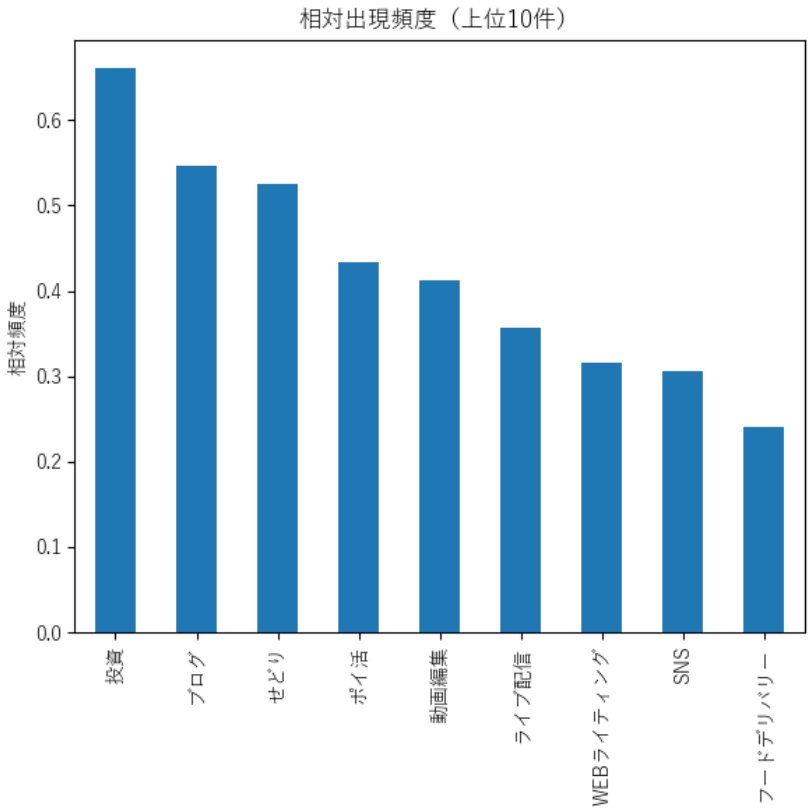
解析結果はアイテムセットごとに，代表的なアソシエーションルールの算出結果が表示されます．

	始点	終点	始点の支持度	終点の支持度	支持度	信頼度	Lift	epresentativity	Leverage	Conviction
1	ボイ活	せどり	0.433	0.525	0.275	0.6351	1.2097	1.0	0.0477	1.3017
2	せどり	ボイ活	0.525	0.433	0.275	0.5238	1.2097	1.0	0.0477	1.1907
3	ライブ配信	ブログ	0.357	0.547	0.249	0.6975	1.2751	1.0	0.0537	1.4974
4	ブログ	ライブ配信	0.547	0.357	0.249	0.4552	1.2751	1.0	0.0537	1.1803
5	WEBライティ...	ブログ	0.316	0.547	0.229	0.7247	1.3248	1.0	0.0561	1.6454
6	ブログ	WEBライティ...	0.547	0.316	0.229	0.4186	1.3248	1.0	0.0561	1.1766
7	動画編集	SNS	0.412	0.306	0.189	0.4587	1.4991	1.0	0.0629	1.2822
8	SNS	動画編集	0.306	0.412	0.189	0.6176	1.4991	1.0	0.0629	1.5378
9	フードデリバリー	ボイ活	0.241	0.433	0.156	0.6473	1.4949	1.0	0.0516	1.6076
10	ボイ活	フードデリバリー	0.433	0.241	0.156	0.3603	1.4949	1.0	0.0516	1.1865
11	ライブ配信	SNS	0.357	0.306	0.15	0.4202	1.3731	1.0	0.0408	1.1969
12	SNS	ライブ配信	0.306	0.357	0.15	0.4902	1.3731	1.0	0.0408	1.2613

② 相対出現頻度

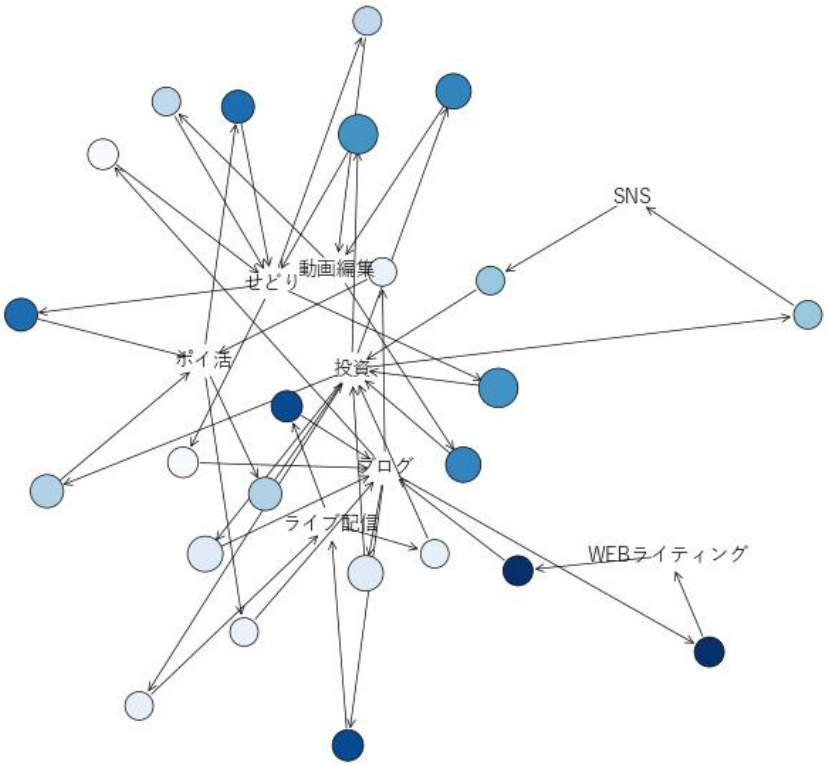
「相対出現頻度」画面を表示して，「相対出現頻度の作成」を ON にして，「解析実行」することで以下のようにアイテムの相対出現頻度を示した，棒グラフが表示されます．





③ ネットワーク図

ネットワーク図はデフォルト設定で解析実行時に作成されます．「ネットワーク図」画面で中間ノードの濃淡にする値やネットワークの形状を設定できます．



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．

21. メタアナリシス

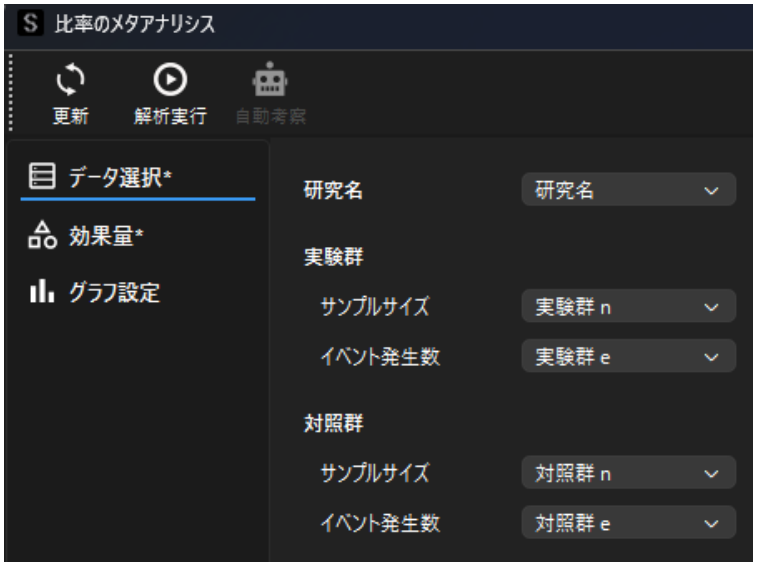
① カテゴリーデータに対するメタアナリシス

カテゴリーデータつまり，オッズ比やリスク比を用いたメタアナリシスの実行方法を説明します．

分析用のデータとして，以下のように研究名と比較する 2 群（実験群・対照群）ごとのサンプルサイズ，イベント発生数を用意します．

	研究名	実験群 n	実験群 e	対照群 n	対照群 e
1	研究A	50	30	50	25
2	研究B	125	55	125	45
3	研究C	35	7	35	10
4	研究D	70	43	70	38
5	研究E	20	5	20	7
6	研究F	100	74	100	63

メニューバーから「その他」→「メタアナリシス」→「比率のメタアナリシス」を選択します．
比率のメタアナリシス用ウィンドウが表示されたら，各項目に分析対象の変数を選択します．



「解析実行」ボタンをクリックすると，メタアナリシスの結果が画面右側に表示されます．選択した効果量と，固定効果モデル・変量効果モデルによる統合効果，重みなどが表示されます．

解析結果

	n	効果量	95% 下限	95% 上限	重み(固定) %	重み(変量) %
研究A	100	1.5	0.6793	3.3121	12.5807	13.5377
研究B	250	1.3968	0.8404	2.3216	31.7032	32.909
研究C	70	0.625	0.2067	1.8894	10.0645	6.9408
研究D	140	1.3411	0.6843	2.6284	18.4396	18.761
研究E	40	0.619	0.1578	2.4289	6.6048	4.5455
研究F	200	1.6716	0.914	3.0571	20.6071	23.3059
固定効果モデル	800	1.3271	0.9933	1.7731	100.0	nan
変量効果モデル	800	0.2853	-0.0062	0.5767	nan	100.0

解析結果画面全体では，モデルの全体効果やコクランの Q 検定の結果や異質性尺度 I^2 の結果が表示されます．

解析結果

効果量

	n	効果量	95% 下限	95% 上限	重み(固定) %	重み(変量) %
研究A	100	1.5	0.6793	3.3121	12.5807	13.5377
研究B	250	1.3968	0.8404	2.3216	31.7032	32.909
研究C	70	0.625	0.2067	1.8894	10.0645	6.9408
研究D	140	1.3411	0.6843	2.6284	18.4396	18.761
研究E	40	0.619	0.1578	2.4289	6.6048	4.5455

全体効果

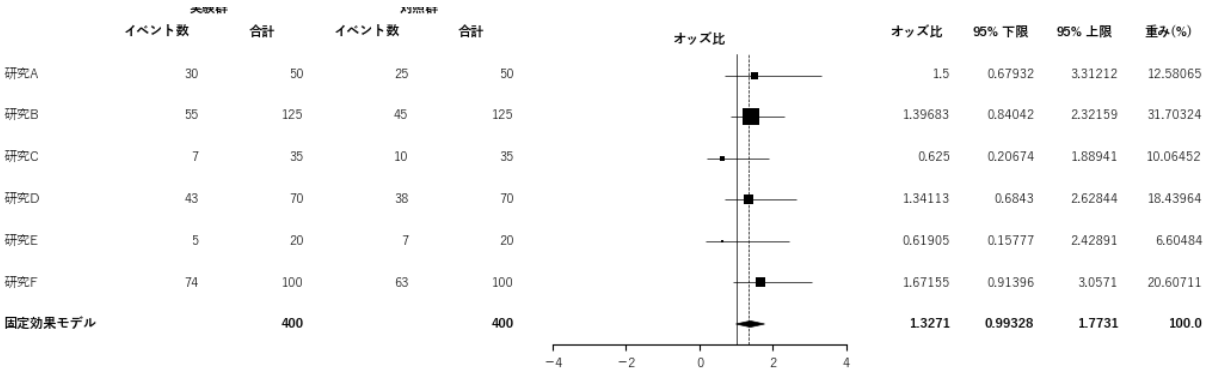
	z	p値
固定効果モデル	1.9144	0.0558
変量効果モデル	1.9183	0.0553

異質性指標

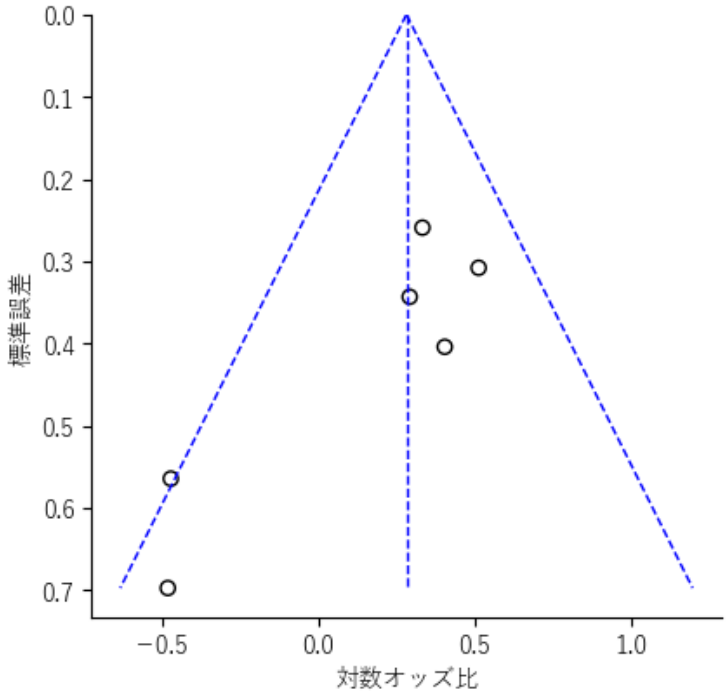
コクランのQ検定 統計量 X^2	3.6683
コクランのQ検定 p値	0.5982
異質性尺度 I^2	0.0
研究間分散 τ^2	0.0

「グラフ設定」画面では設定を変更することで，フォレストプロットとファンネルプロットを作成することができます．

フォレストプロットではモデルを選択して，「作成」ボタンをクリックすると以下のようなフォレストプロットが表示されます．



ファンネルプロットの「作成」ボタンをクリックすると，以下のようなファンネルプロットが表示されます．



リスク比・リスク差についても効果量の設定を変更することで，同様の解析を行うことが可能です．

② 連続データに対するメタアナリシス

連続データつまり，平均差や標準化平均差を用いたメタアナリシスの実行方法を説明します．

分析用のデータとして，以下のように研究名と比較する 2 群ごとのサンプルサイズ，平均値，標準誤差を用意します．

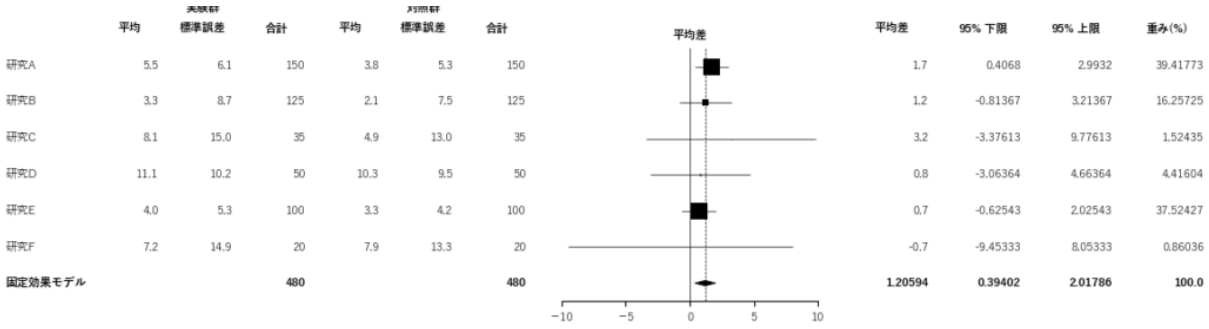
※ 標準誤差は文献に記載されている，信頼区間から算出することができます．

	研究名	実験群 平均	実験群 se	実験群 n	対照群 平均	対照群 se	対照群 n
1	研究A	5.5	6.1	150	3.8	5.3	150
2	研究B	3.3	8.7	125	2.1	7.5	125
3	研究C	8.1	15.0	35	4.9	13.0	35
4	研究D	11.1	10.2	50	10.3	9.5	50
5	研究E	4.0	5.3	100	3.3	4.2	100
6	研究F	7.2	14.9	20	7.9	13.3	20

メニューバーから「その他」→「メタアナリシス」→「平均値のメタアナリシス」を選択して，分析用ウィンドウを表示します．

分析対象の変数を選択して，「解析実行」ボタンをクリックすると計算結果が表示されます．

計算結果の表示内容と，グラフの作成方法は比率のメタアナリシスと同様になります．フォレストプロットについては，表示項目が比率のメタアナリシスとは異なり，以下のようになります．



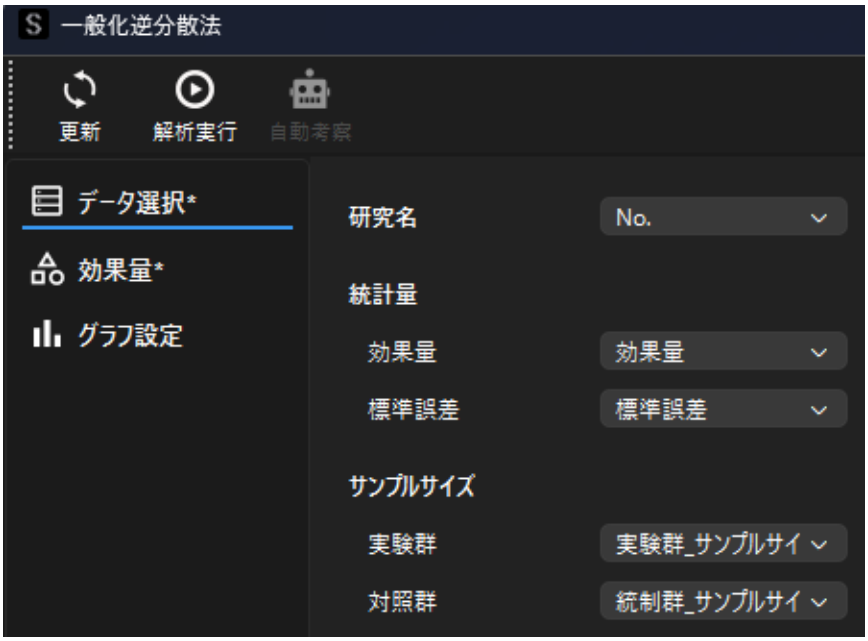
標準化平均差についても効果量の設定を変更するだけで，同様の解析を行うことが可能です．

③ 一般化逆分散法によるメタアナリシス

一般化逆分散法を用いたメタアナリシスでは，ここまで紹介したカテゴリーデータと連続データの両方のデータを扱うことができます．

分析画面は以下のようになっており，研究ごとに効果量，標準誤差，2 群ごとのサンプルサイズが必要になります．

例としてリスク比を用いたメタアナリシスを行いたい場合は，効果量として各研究のリスク比をデータとして用意して，以下のように変数と効果量を選択します．



一般化逆分散法では，群ごとのイベント発生数や平均値，標準誤差が必要ないので，文献から得られるデータに制限や不整合がある場合に，役立つことが多いです．

一般化逆分散法では，オッズ比，リスク比，リスク差，平均差，標準化平均差を効果量として扱うことができます．

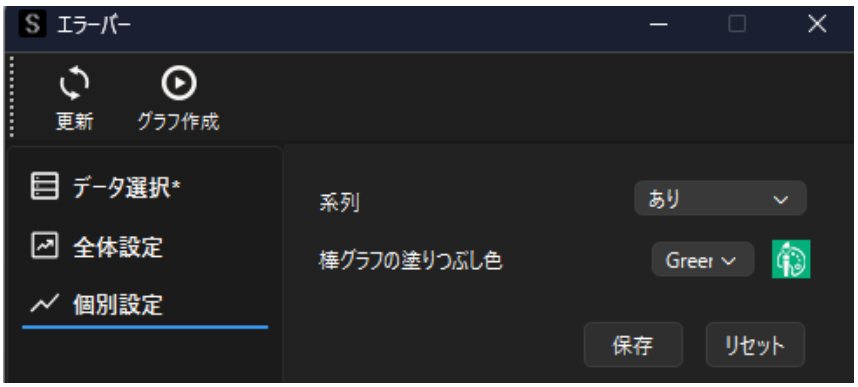
グラフ作成

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．

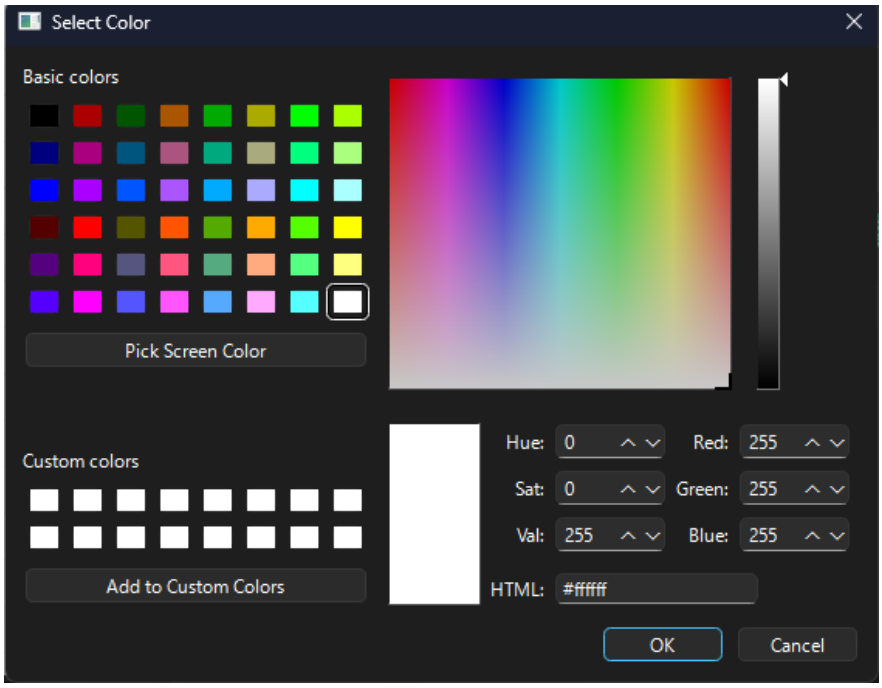
1．グラフ作成機能の共通操作（個別系列設定）

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，「個別設定」画面で対象の描画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，設定したい変数を選択した状態で，棒グラフの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．

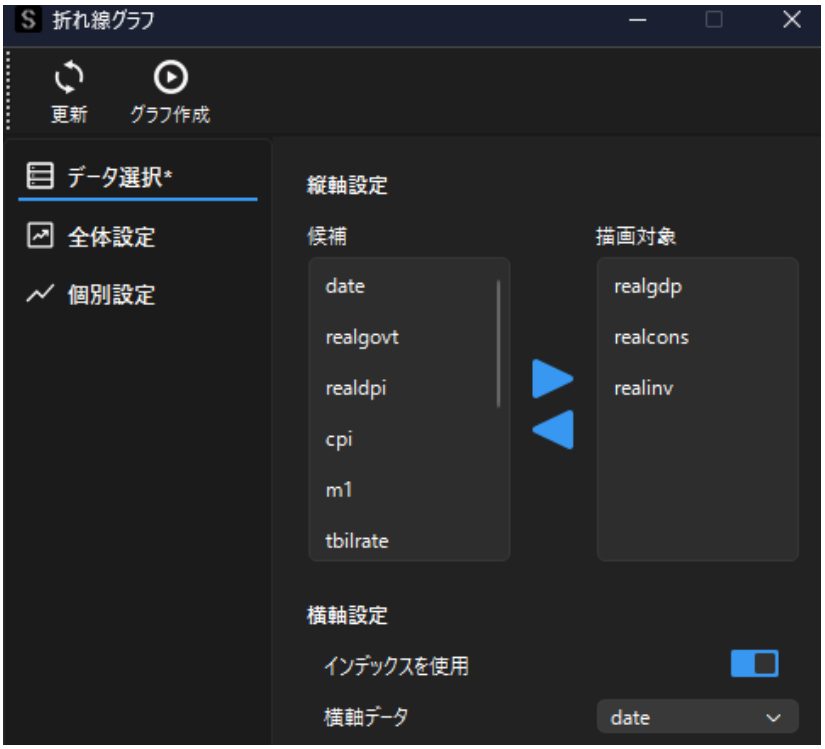
個別設定は「保存」ボタンをクリックすることで，設定値が保存されてグラフ作成時に反映されます．
全体設定などは「保存」を行なう必要はありません．

2．折れ線グラフ

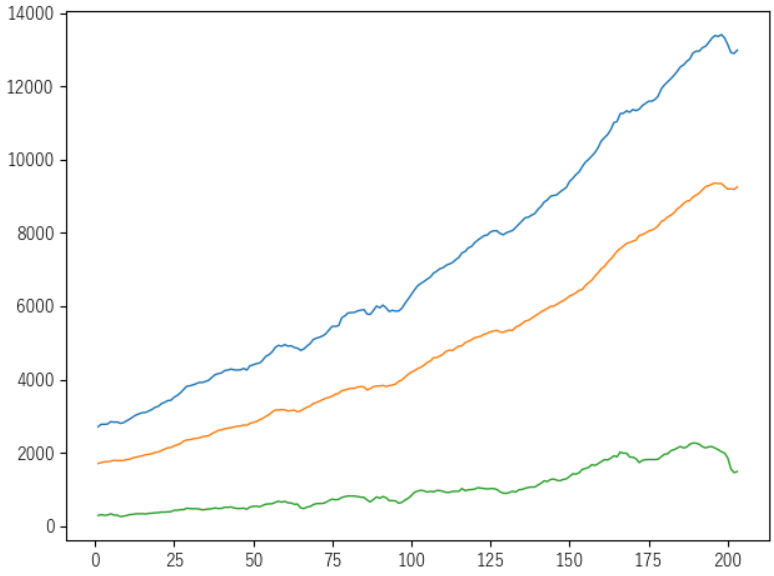
折れ線グラフは時系列データなど，変化を表すデータに有効なグラフです．

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，「縦軸設定」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．

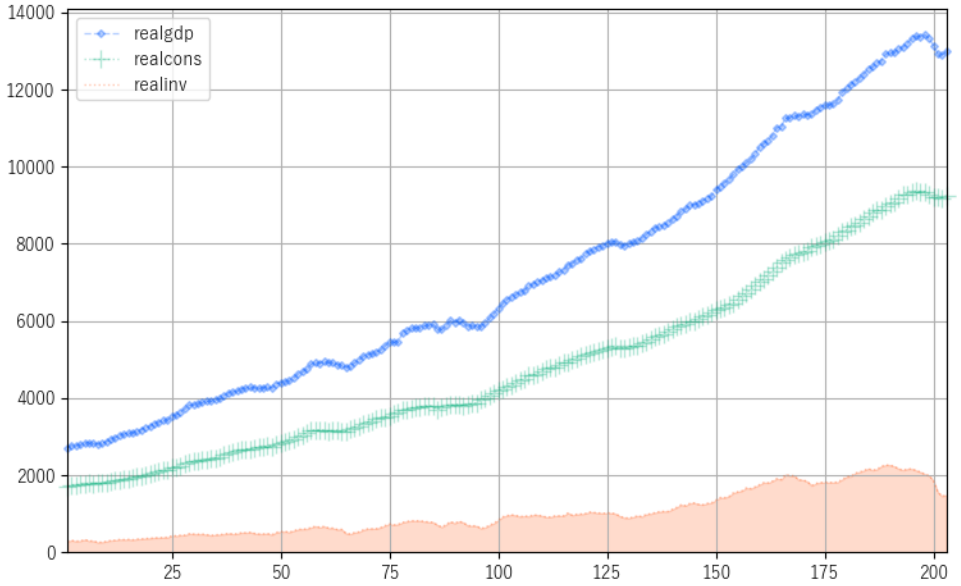


② 横軸の設定

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．



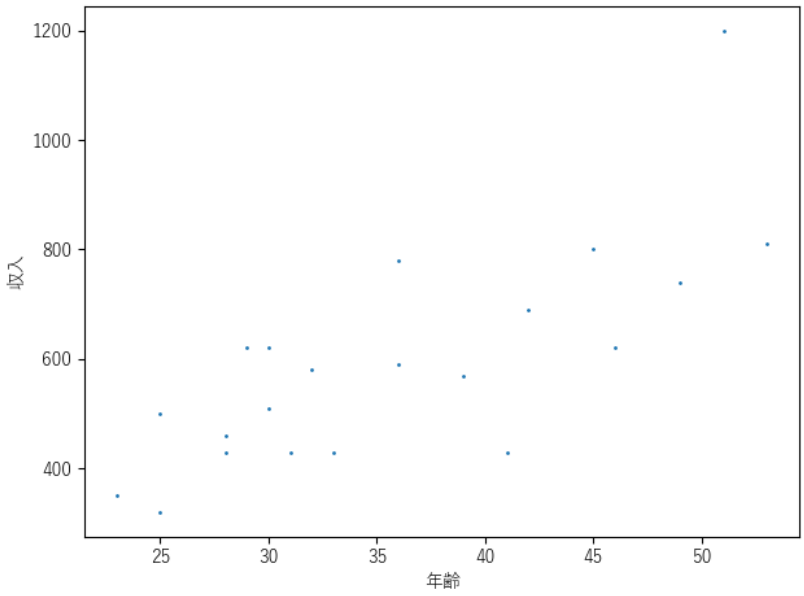
3．散布図

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

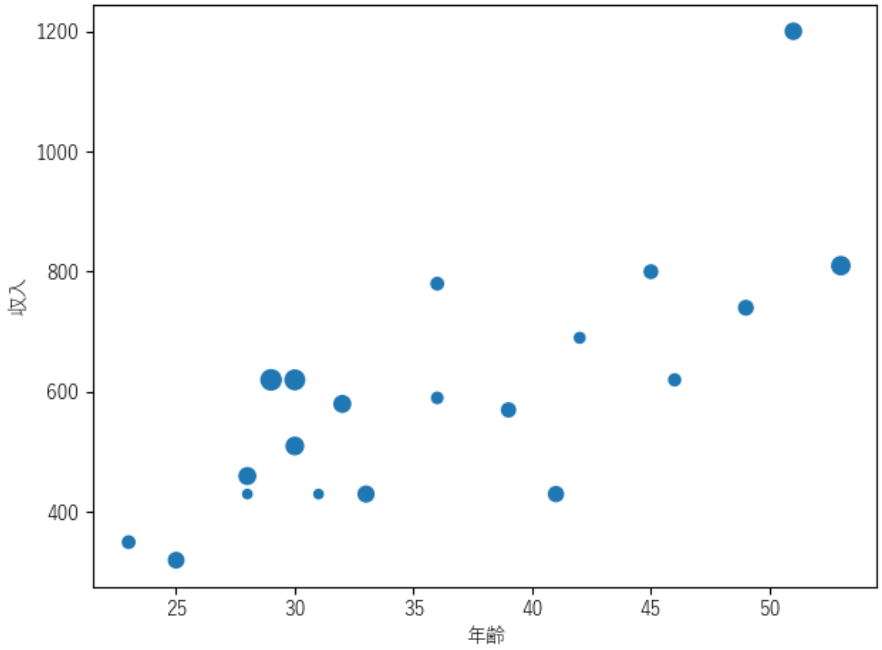
メニューバーから「グラフ」→「散布図」を選択します．以下のような散布図ウィンドウが表示されたら，横軸と縦軸に描画対象としたい変数（列名）を選択します．

「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．



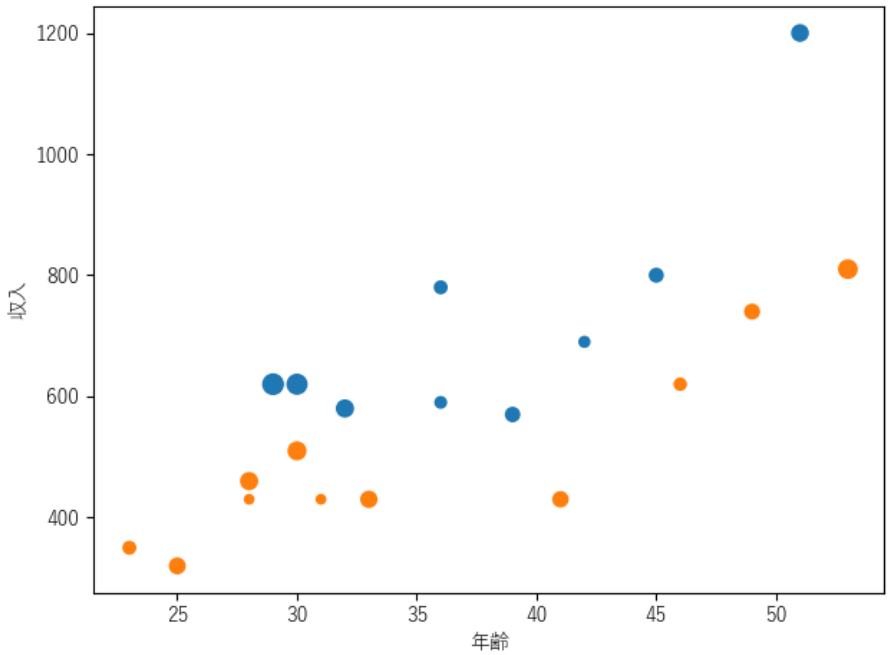
② サイズ選択（バブルチャート）

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．



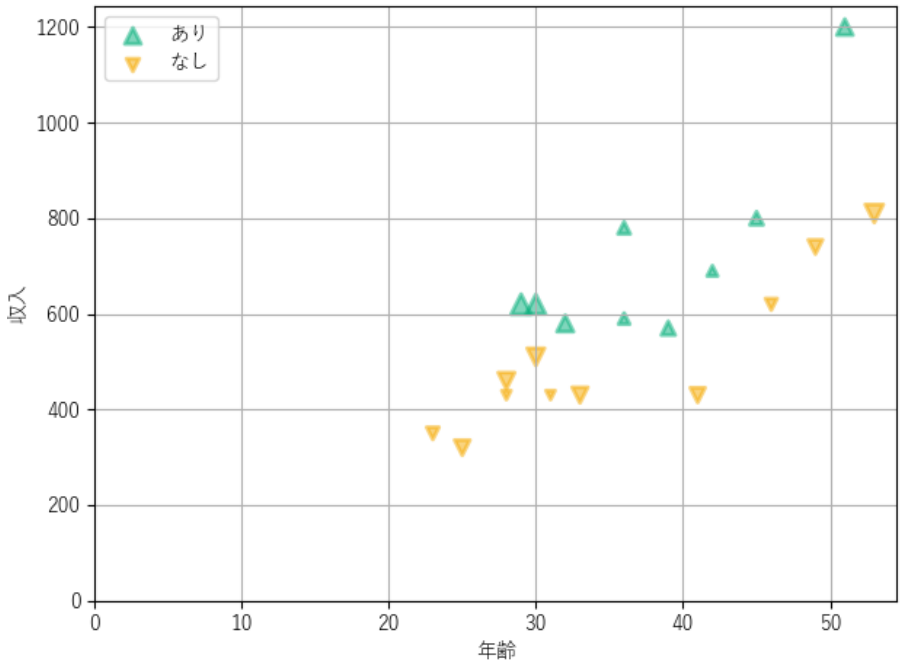
③ カテゴリーによる色分け

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．

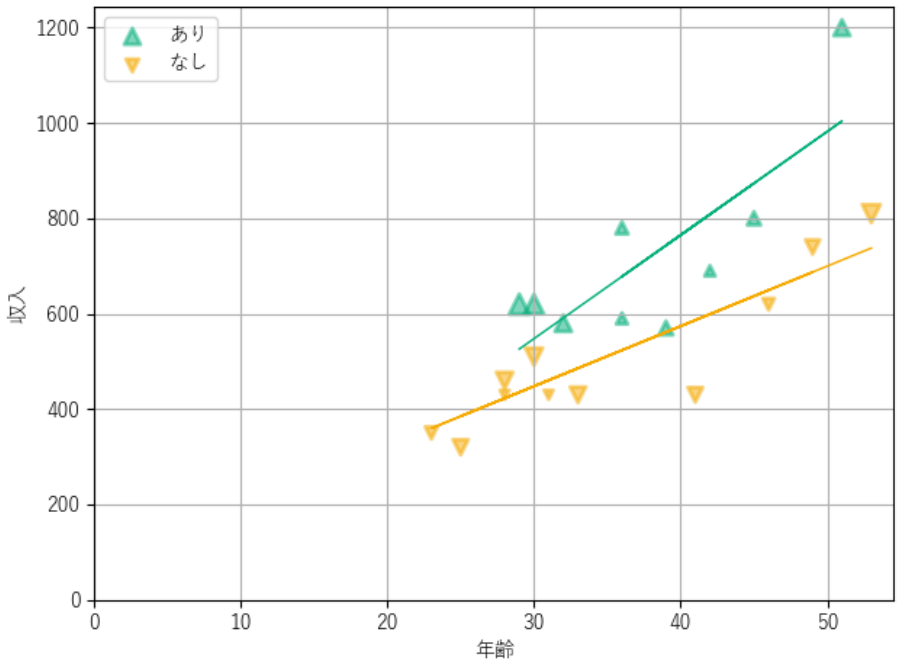


④ その他の設定

「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.



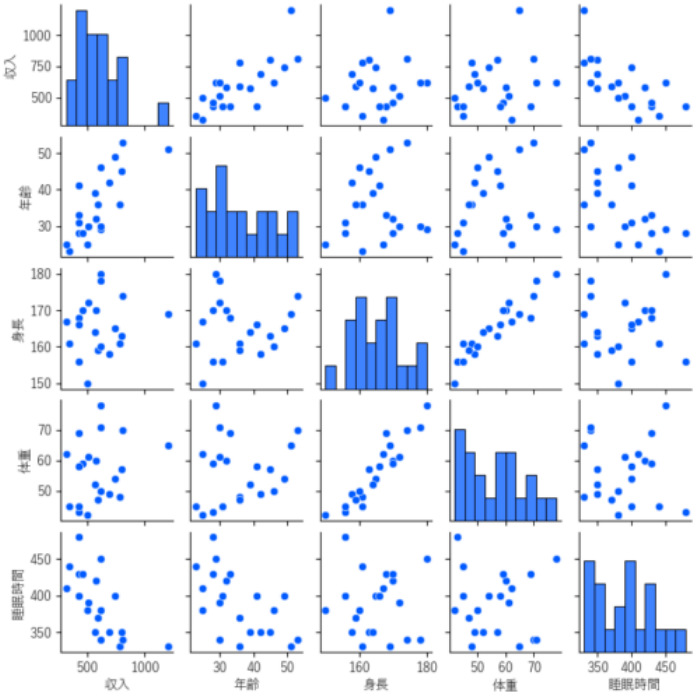
4．散布図行列

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．

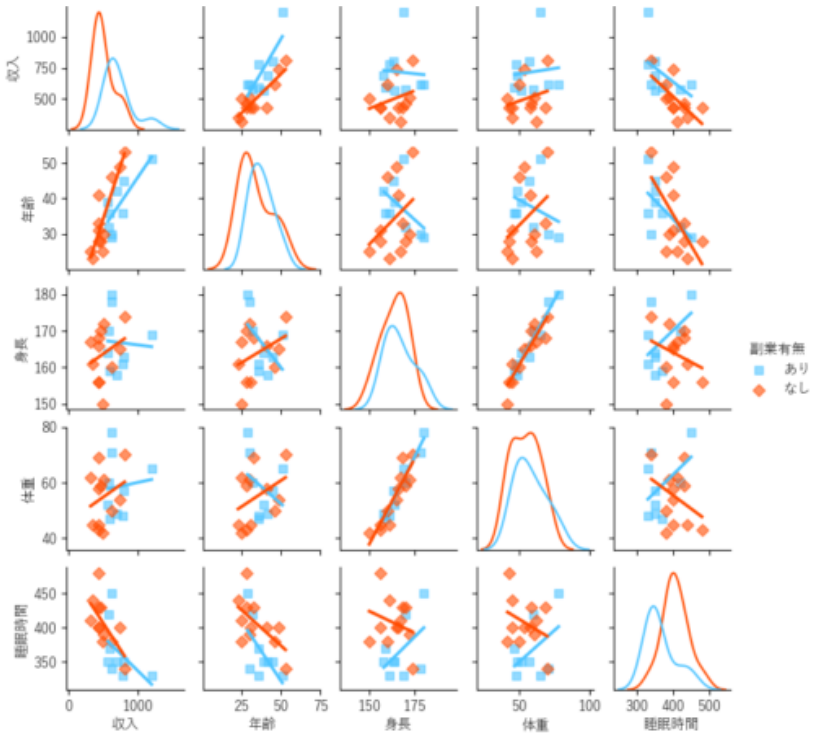


② カテゴリーによる色分け

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．

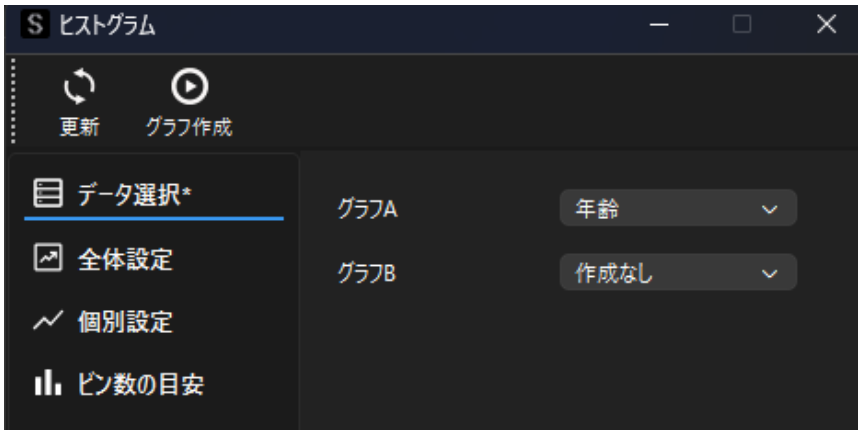


5. ヒストグラム

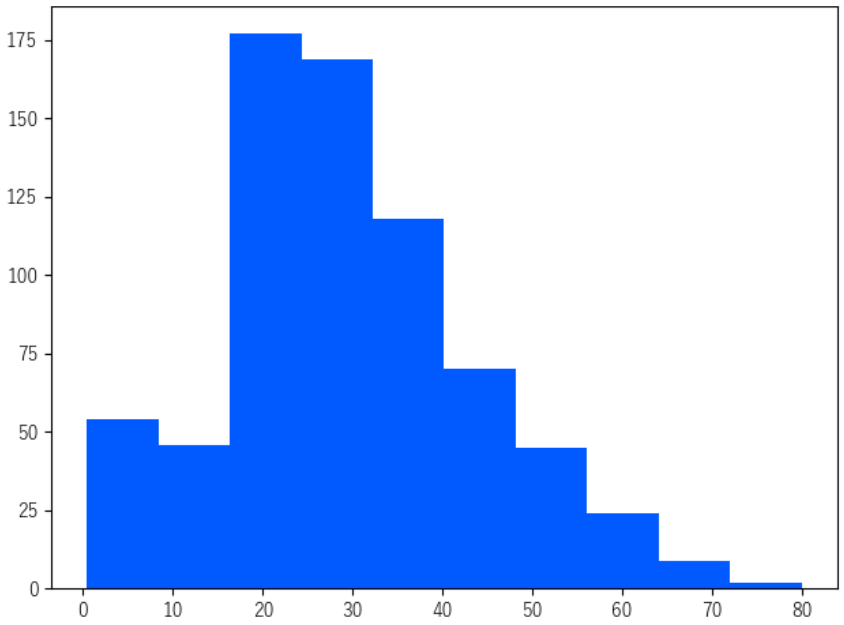
ヒストグラムは１つの変数の分布を示す最も基本的なグラフです．

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．

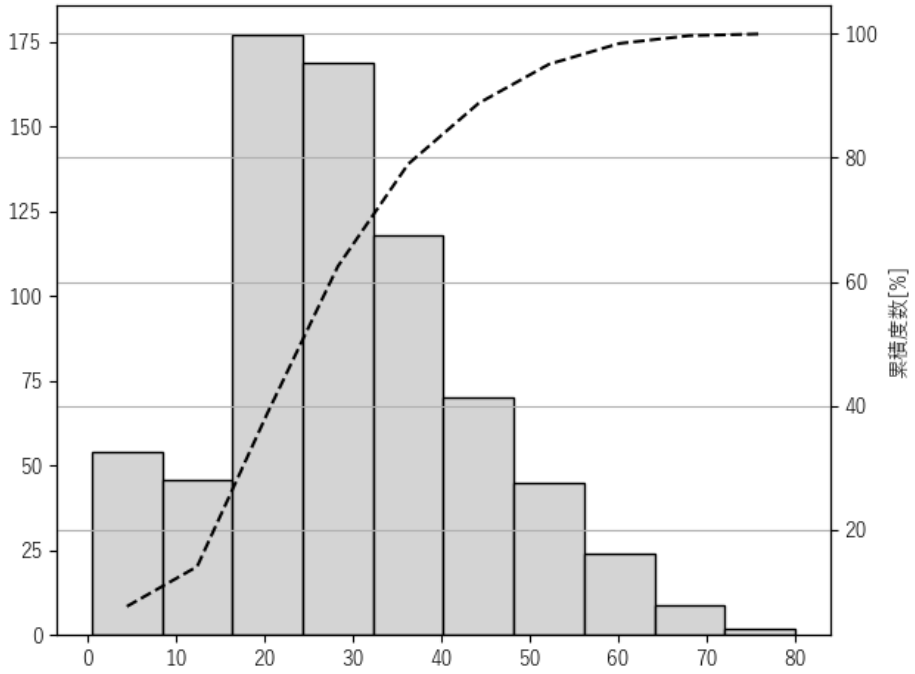


② ２つのヒストグラムを表示する

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.

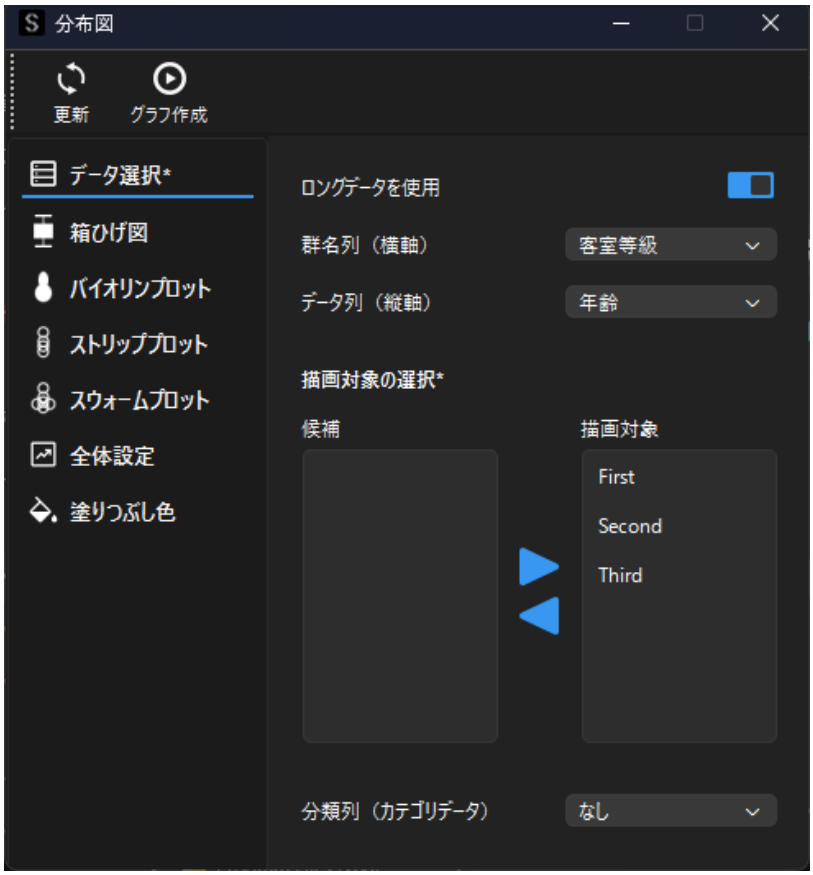


6．分布図

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．

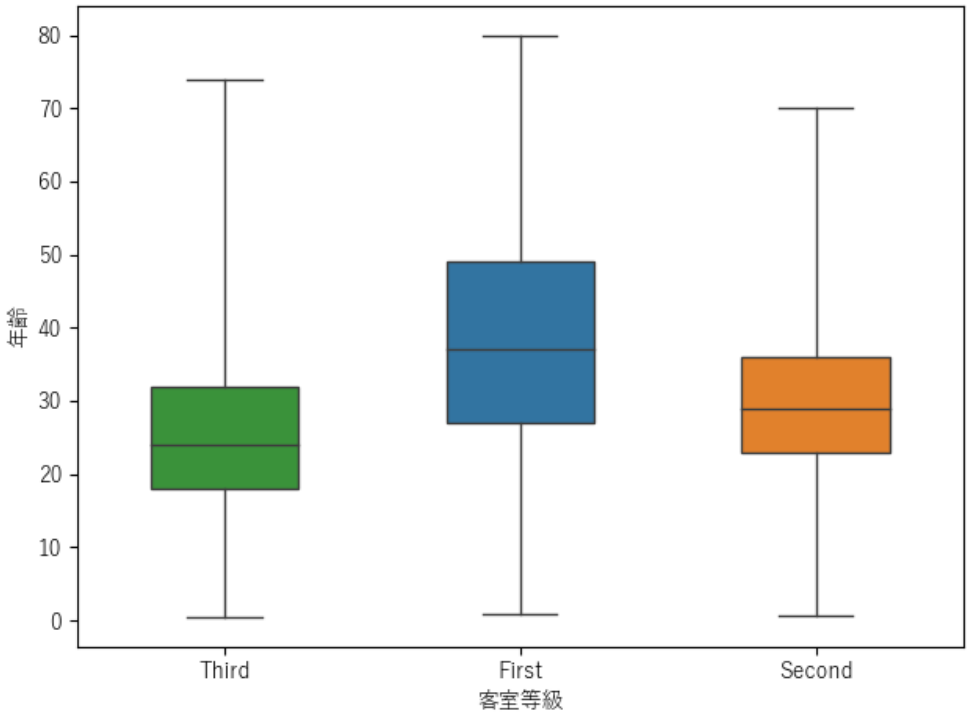
① 基本操作

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．

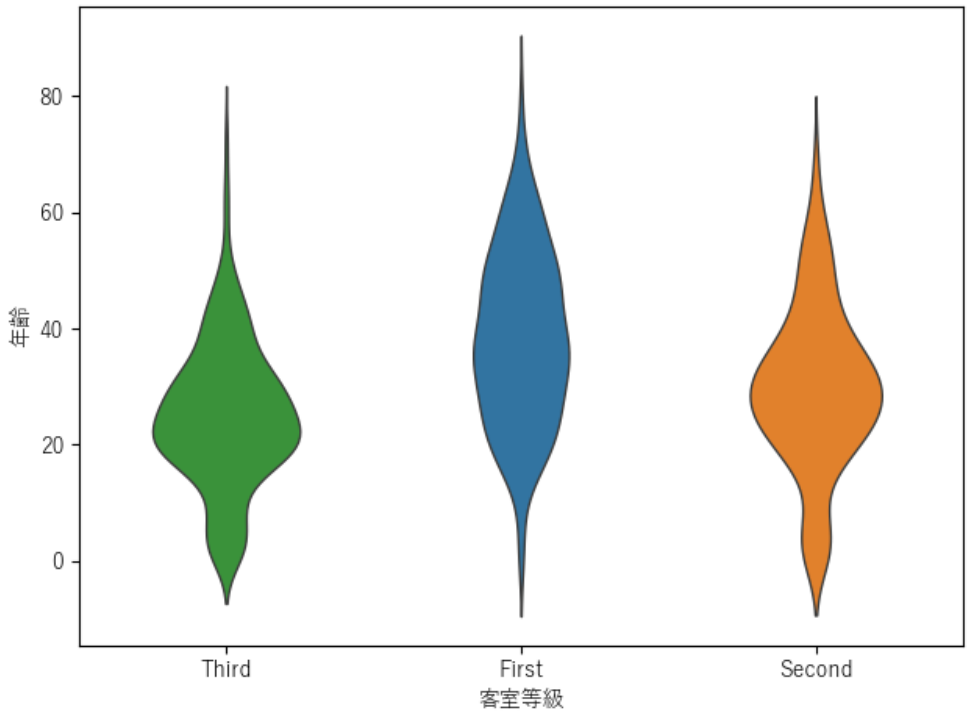


各グラフ画面で描画するグラフを ON にして「グラフ作成」をクリックした場合，以下のようにグラフが表示されます．

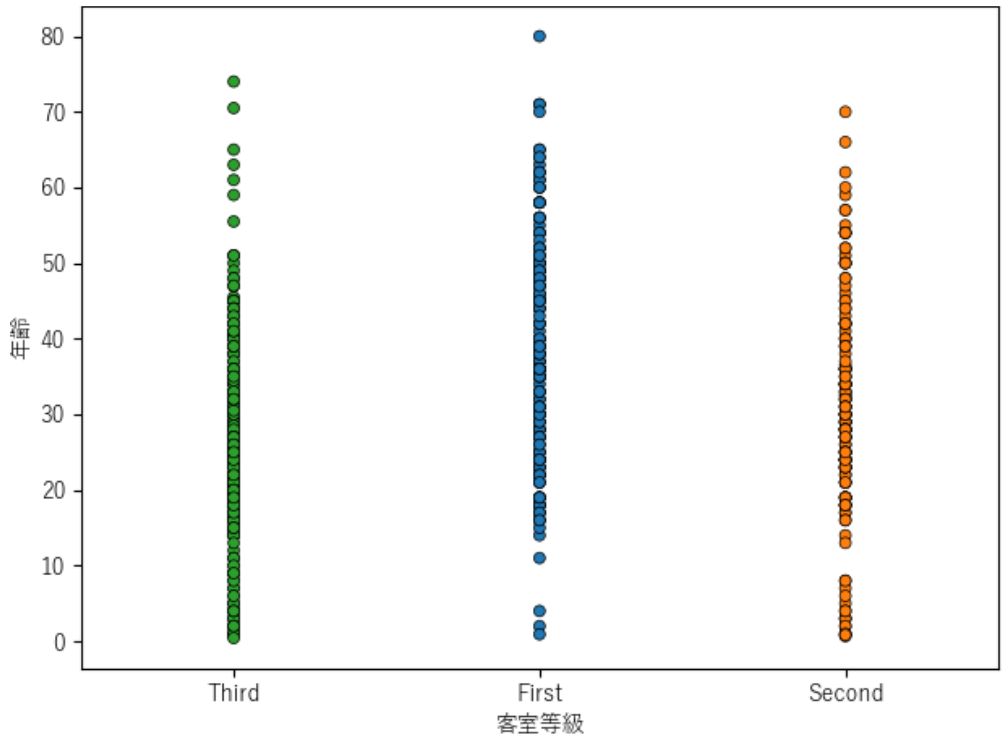
箱ひげ図



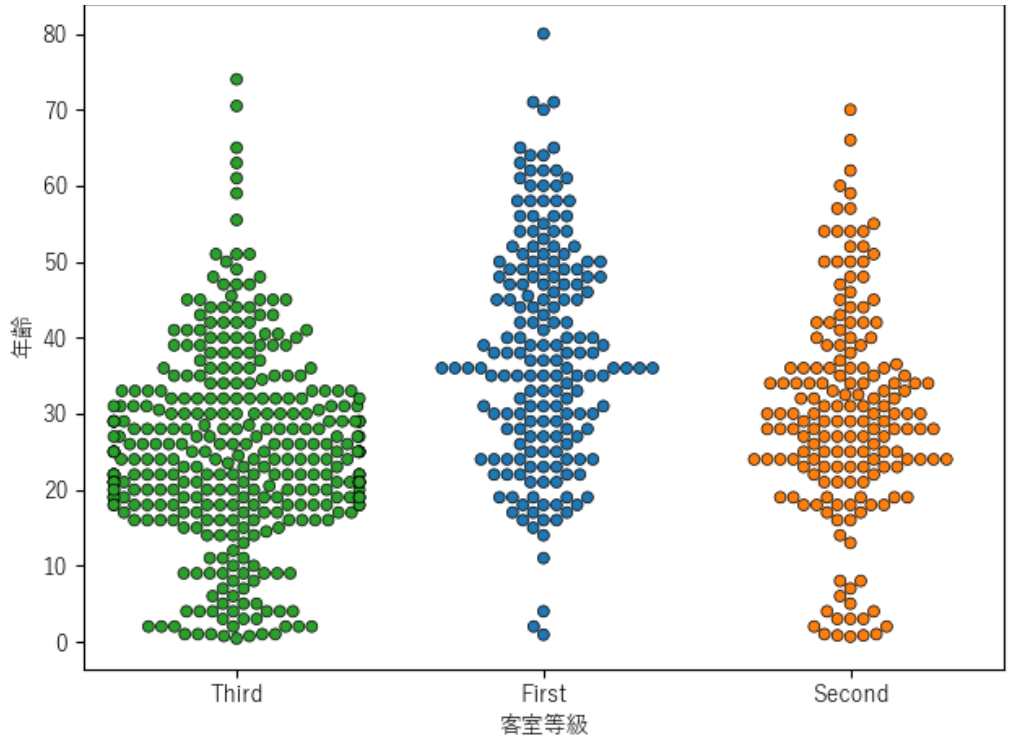
バイオリンプロット



ストリッププロット

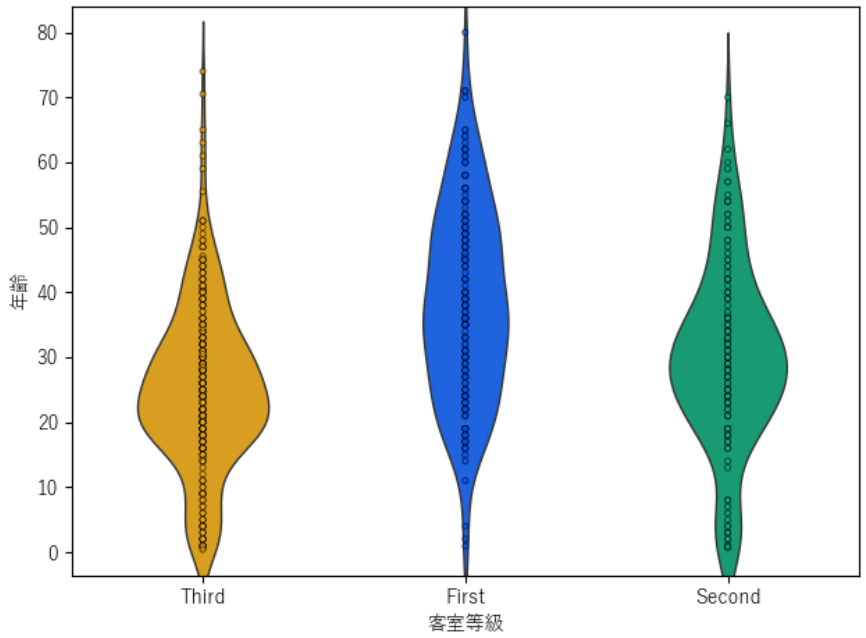


スウォームプロット



② 複数のグラフを組み合わせる

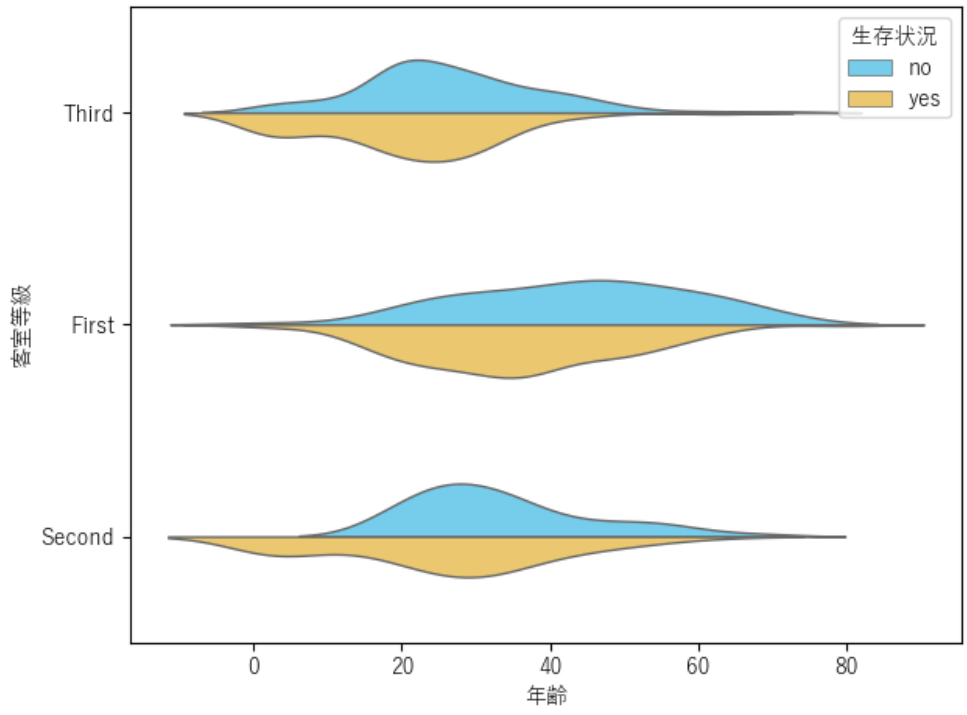
グラフ選択で複数のグラフを選択した場合，グラフが重なって表示されます．バイオリンプロットとストリッププロットを組み合わせたグラフの作成例は以下になります．



③ カテゴリーによる色分け

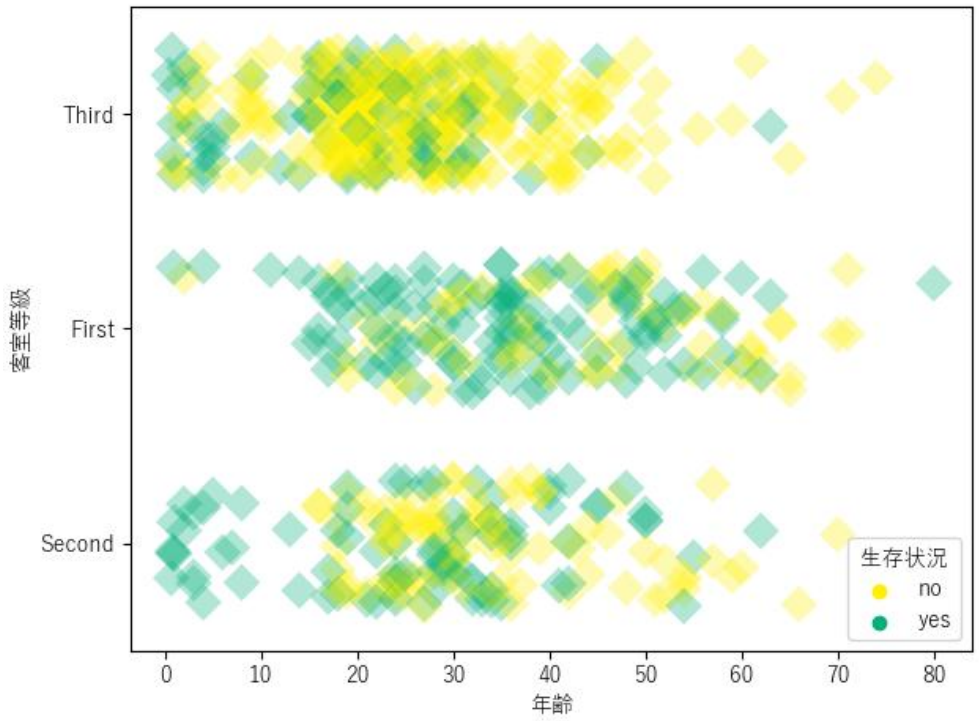
グラフの色をカテゴリーによって色分けしたい場合は，「カテゴリ分類」でカテゴリーを示す変数（文字型データ）を選択します．

※ 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．



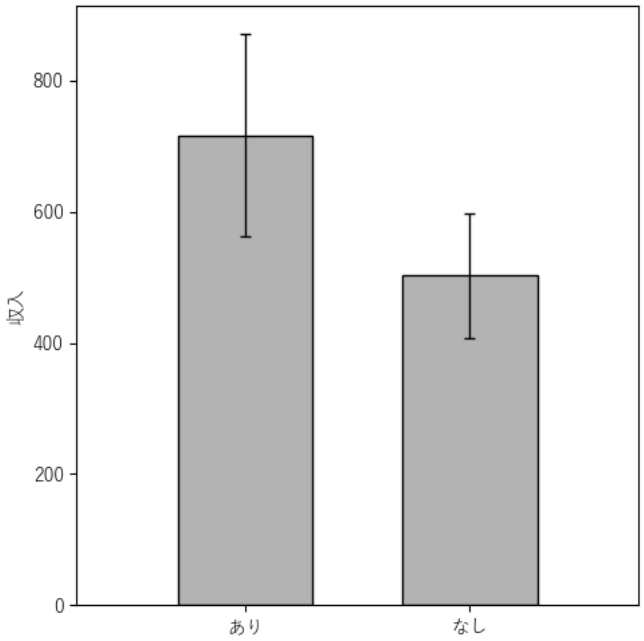
7. エラーバー

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．



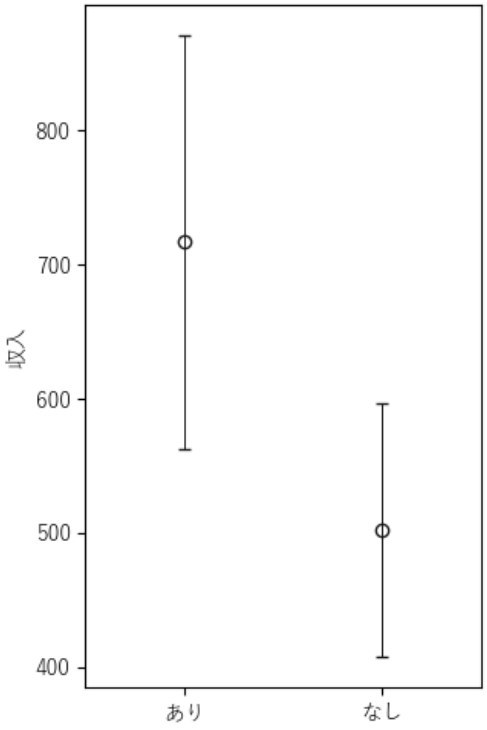
② エラーバーの値

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．

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．

※ 自動考察は数十秒要します．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．

自動考察結果は以下のような新しいウィンドウに表示されます．ロジスティック回帰分析の場合は，総評，各結果の解釈，モデルに注意点・改善点で構成されています．

考察結果の注意点

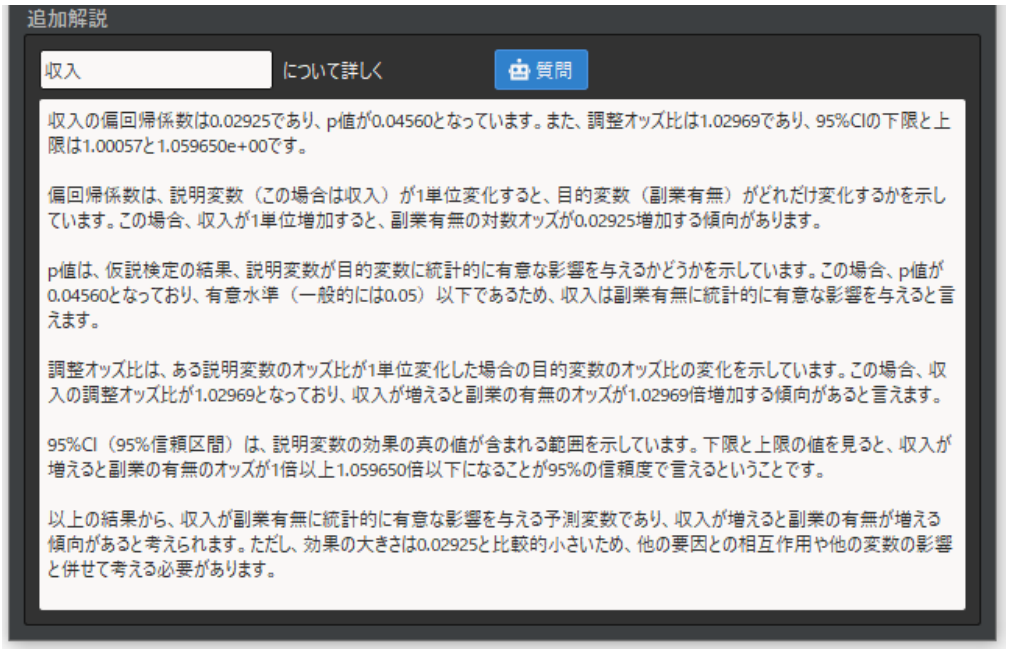
自動考察機能による考察結果は統計学的・数学的解釈でしかないので記載内容が定性的に正しいのか，相関関係・因果関係が現実的に正しいのかは十分に吟味してください．

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

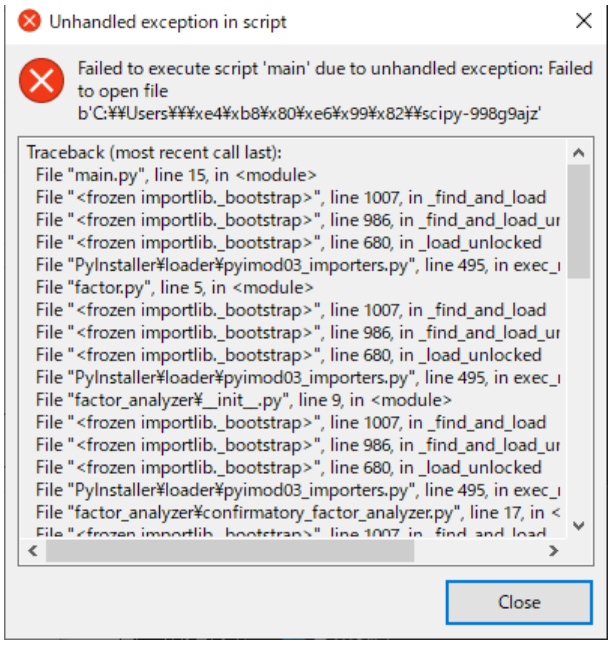
自動考察機能は，OpenAI 社が開発した ChatGPT API （AI）を利用しています．

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

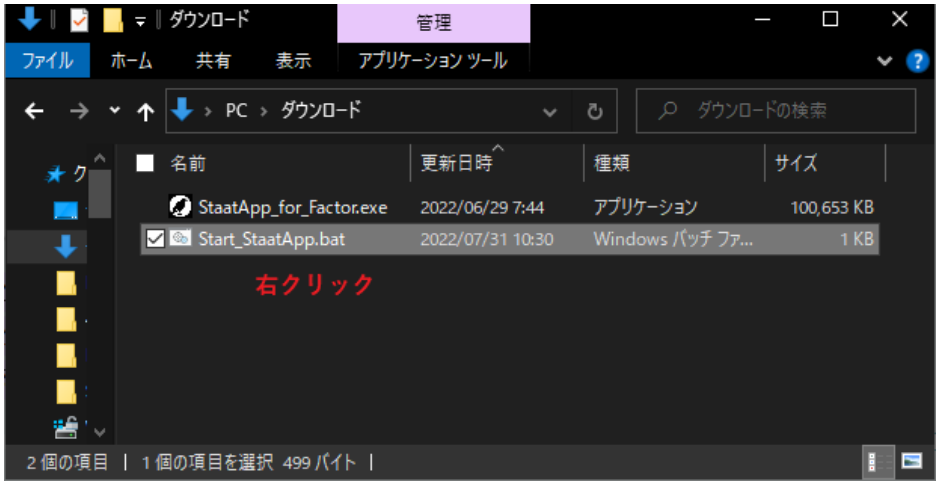
エラーの原因は StaatApp 起動時に使用する temp フォルダのに日本語が含まれることです．以下のページを参考に環境変数 TEMP と TMP を，日本語が含まれないパスに変更することで対応可能です．

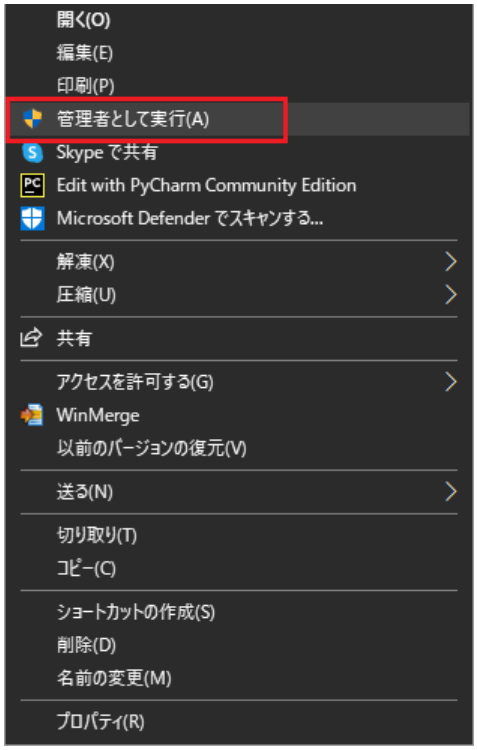
▷環境変数の変更方法

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．

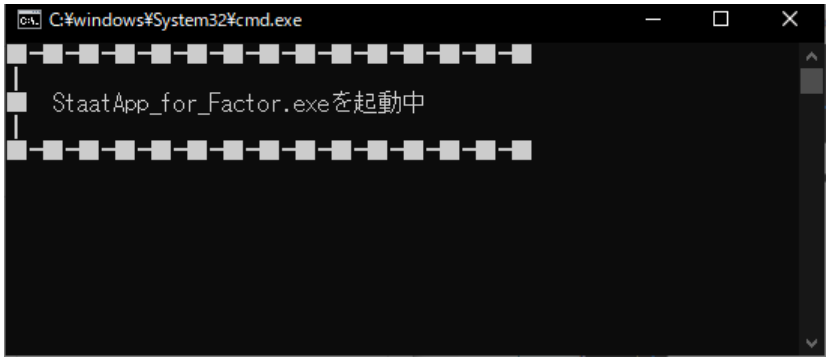
Github から”Start_StaatApp.zip”をダウンロードしたら，zip ファイルを解凍します．解凍後，同じフォルダに使用する StaatApp を配置してください．

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な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，使用方法や実行結果は Windows 版と同等です．