

StaatApp 操作マニュアル

1 はじめにお読みください

StaatApp 操作マニュアルを読む際に、はじめに確認する事項について説明します。

1-1 動作環境

StaatApp の動作環境は、以下の表のとおりです。

なお、本マニュアルでは Windows 10 での操作例を記載します。

OS	備考
Windows 10	64bit
Windows 11	64bit
macOS	-

1-2 対象読者

はじめて StaatApp を使う方

1-3 使用している画像

本マニュアルで使用している画像と、実際の画面が異なる場合があります。なお、機能に差異はありません。

1-4 手順で使用するサンプルデータ

マニュアルの手順では、統計解析用のサンプルデータを使用しています。サンプルデータは以下からダウンロードできます。

<https://1drv.ms/f/s!AoRSXWdvhdLQgZhYUHJkv-mui38RGA?e=mQPixZ>

1-5 マニュアル内の表記

マニュアル内で使用する表記の種類は、以下のとおりです。

- ■：重要事項です。操作上の注意事項や制限事項を説明します。必ず確認してください。
- ■：手順の参考となる情報です。
- **参考：(リンク)**：本マニュアル内または Web ページを参照します。
- 「」：画面名を示します。
- []：ボタン名/タブ名/キーボードキー名を示します。

目次

1	はじめにお読みください	1
1-1	動作環境	1
1-2	対象読者	1
1-3	使用している画像	1
1-4	手順で使用するサンプルデータ	1
1-5	マニュアル内の表記	1
2	基本操作	4
2-1	StaatApp を起動する (Windows)	4
2-2	StaatApp を起動する (macOS)	5
2-3	StaatApp のメニュー・画面について	7
2-4	データを読み込む (CSV ファイル)	8
2-5	データを読み込む (Excel ファイル)	9
2-6	CSV ファイルの出力フォルダを選択する	10
2-7	CSV ファイルを出力する	11
2-8	統計解析を実行する	12
2-9	変数を選択する/選択した変数を未選択に戻す	13
2-10	グラフの表示/保存/編集	14
3	データ操作	15
3-1	複数列を対象にしてソートする	16
3-2	置換する (表記ゆれの処理をする)	17
3-3	特定の行を削除する (不要データの削除・外れ値の処理)	19
3-4	欠測値の処理をする	20
3-5	ダミー変数へ変換する	22
3-6	異なるデータ型に変換する	23
4	解析手法別操作	25
4-1	基本統計量	27
4-2	相関係数	28
4-3	パラメトリック検定	29
4-3-1	対応のない t 検定	29
4-3-2	多元配置分散分析	30
4-4	マンホイットニーの U 検定	31
4-5	ウィルコクソンの符号順位和検定	32
4-6	クラスカル・ウォリス検定	33
4-7	フリードマン検定	34
4-8	チューキー・クレーマー法	35
4-9	スティール・ドゥワス検定	36
4-10	ダネット検定	37
4-11	正規性の確認	38
4-12	等分散性の検定	39
4-13	重回帰分析	40
4-14	数量化 II 類	41
4-15	ロジスティック回帰分析	43
4-15-1	ロジスティック回帰分析の実行	43
4-15-2	ロジスティック回帰分析の解析結果の見方	44
4-16	多項ロジスティック回帰分析	45
4-17	主成分分析	46
4-18	因子分析	49
4-19	k-means 法	51
4-20	階層クラスター分析	52
4-21	ヒストグラム	54
4-22	分布図	56
4-22-1	分類図を作成する	56
4-22-2	箱ひげ図の設定項目を確認する	58
4-22-3	バイオリンプロットの設定項目を確認する	58

4-2 2-4 ストリッププロット・スウォームプロットの設定項目を確認する	59
4-2 3 散布図行列・ヒートマップ	60
4-2 3-1 散布図行列を作成する	60
4-2 3-2 散布図行列の詳細設定とグラフ例	61
4-2 3-3 ヒートマップを作成する.....	62
4-2 3-4 ヒートマップの詳細設定とグラフ例.....	62
4-2 4 クロス集計表.....	64
4-2 5 生存曲線.....	66
4-2 6 Cox比例ハザード回帰.....	68
5 画面の表示設定.....	71
5-1 小数点の表示桁数を変更する	71
5-2 カラーモードを変更する.....	72
6 付録.....	73
6-1 起動時に発生するエラー.....	73
6-2 よくあるお問い合わせ.....	73

2 基本操作

StaadApp の基本的な操作方法について説明します。

2-1 StaatApp を起動する (Windows)

Windows で StaatApp を起動する手順を説明します。

指定の URL から実行ファイルをダウンロードし、ダウンロードした実行ファイルをダブルクリックすると StaatApp が起動します。

1 アプリをダウンロードする

1. [StaaApp](#) の製品ページにアクセスしてファイルをダウンロードし、任意の場所に保存します。

2 StaatApp を起動する

1. ダウンロードした StaatApp.exe をダブルクリックします。

StaatApp が起動します。

StaatApp の起動にかかる時間

StaatApp が起動するのに約 10 秒程度かかる場合があります。

2-2 StaatApp を起動する (macOS)

macOS で StaatApp を起動する手順を説明します。

macOS 版の StaatApp は、CPU が Apple 社製の M1/M2 チップの場合と、Intel 社製の x86_64 の場合で使用するアプリが異なります。使用している CPU に合ったアプリをご利用ください。

起動するアプリの種類

CPU が M1/M2 チップの場合、M1/M2 チップ用と x86_64用の両方のアプリを使用できます。ただし、x86_64用のアプリを使用すると StaatApp が起動するまでの時間が長くなる場合があるため、M1/M2 チップ用のアプリの使用をお勧めします。

CPU が x86_64 場合は、必ず x86_64用のアプリを実行してください。M1/M2 チップ用のアプリを使用しても、StaatApp を起動できません。

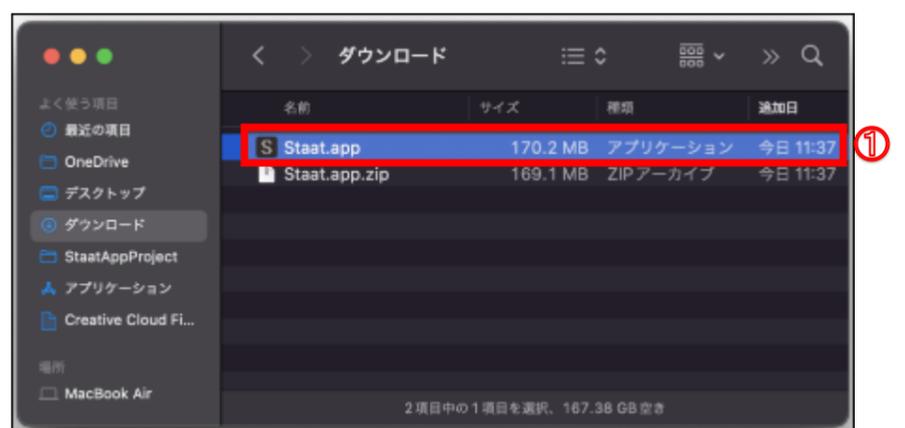
1 アプリをダウンロードする

1. [StaatApp](#) にアクセスしてアプリをダウンロードし、任意の場所に保存します。

2 StaatApp を起動する

1. ダウンロードした StaatApp.app を選択し、
[Control] キーを押しながらクリックします。

コンテキストメニューが表示されます。



2. [開く] をクリックします。

警告メッセージが表示されます。



警告メッセージ

macOS の仕様上、AppStore 以外から入手したアプリを実行しようとするすると警告メッセージが表示されますが、問題ありません。

3. [開く] ボタンをクリックします。
StaatApp が起動します。



StaatApp の起動にかかる時間

StaatApp が起動するのに約 3 分程度かかる場合があります。

2-3 StaatApp のメニュー・画面について

StaatApp のメニューと画面について説明します。



番号	名称	説明
①	メニューバー	各種統計解析, グラフの作成, StaatApp の環境設定などを行います。
②	ツールバー	データファイルの読み込み, CSV ファイルの出力, データクレンジングなどを行います。
③	データ操作画面	読み込んだデータファイルのデータが表示されます。

2-4 データを読み込む (CSV ファイル)

StaatApp で統計解析をする場合、統計解析をしたいデータを CSV ファイルまたは Excel ファイルにして StaatApp に読み込みます。

なお、CSV ファイルの作成方法は以下を参照してください。

参考：[CSV ファイルの作り方 | Staat](#)

ここでは、CSV ファイルを読み込む手順を説明します。

1 CSV ファイルを読み込む

1. ツールバーから [CSV ファイルの読込] ボタンをクリックします。

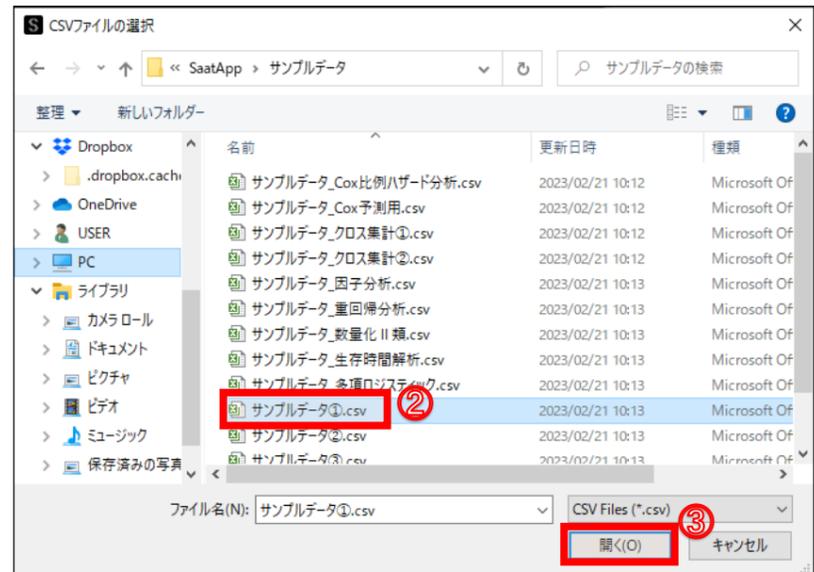
「CSV ファイルの選択」が表示されます。



2. CSV ファイルを選択します。

3. [開く] ボタンをクリックします。

データが表示されます。



1 行目のデータがヘッダーとして読み込まれ、列番号は左側に自動で表示されます。

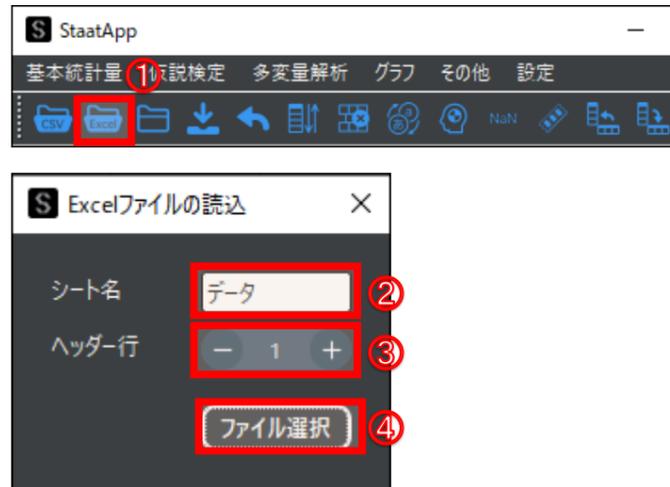
No.	継続期間	イベント発生	収入	性別	年齢
1	1	31	1	580 男性	
2	2	97	1	430 女性	
3	3	126	0	800 男性	
4	4	56	0	780 女性	
5	5	172	1	690 女性	
6	6	156	0	510 男性	
7	7	133	1	740 女性	
8	8	82	1	350 女性	
9	9	173	1	620 男性	
10	10	110	0	500 女性	
11	11	33	1	430 男性	
12	12	156	1	590 女性	
13	13	82	0	1200 男性	
14	14	112	1	810 男性	
15	15	63	1	620 女性	

2-5 データを読み込む (Excel ファイル)

Excel ファイルを読み込む手順を説明します。

1 Excel ファイルを読み込む

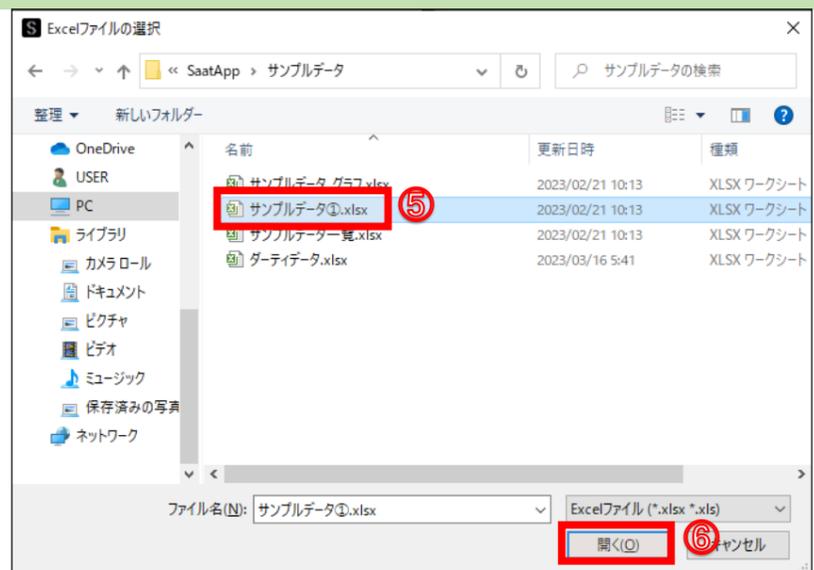
1. ツールバーから [Excel ファイルの読込] ボタンをクリックします。
「Excel ファイルの読込」が表示されます。
2. [シート名] 欄に、読み込む Excel ファイルのシート名を入力します。
3. [ヘッダー行] に、読み込むシートの列名 (変数名) がある行番号を入力します。
[-] または [+] ボタンをクリックして入力することもできます。
4. [ファイル選択] ボタンをクリックします。
「Excel ファイルの選択」が表示されます。



読み込み可能な Excel ファイルのシート数

1つの Excel ファイルの1つのシートのみ読み込めます。

5. Excel ファイルを選択します。
6. [開く] ボタンをクリックします。
データが表示されます。



ヘッダー行は最上部に、列番号は左側に表示されます。

	副乗有無	収入	性別	年齢	身長	体重
1	あり	580	男性	32	170	
2	なし	430	女性	28	156	
3	あり	800	男性	45	163	
4	あり	780	女性	36	161	
5	あり	690	女性	42	158	
6	なし	510	男性	30	172	
7	なし	740	女性	49	165	
8	なし	350	女性	23	161	
9	あり	620	男性	29	180	
10	なし	500	女性	25	150	
11	なし	430	男性	33	168	
12	あり	590	女性	36	159	
13	あり	1200	男性	51	169	
14	なし	810	男性	53	174	
15	なし	620	女性	46	160	

Excel ファイルが読み込めない

Excel ファイルが読み込めない場合は、StaatApp で使用できない文字 (スペースや記号など) が含まれている可能性があります。使用できない文字を削除/置換してください。データの整理方法については、以下を参照してください。

参考: [データの入力・加工方法 | Staat](#)

どうしても Excel ファイルが読み込めない場合は、Excel ファイルを CSV ファイルに変換して読み込みこませてください。

2-6 CSV ファイルの出力フォルダを選択する

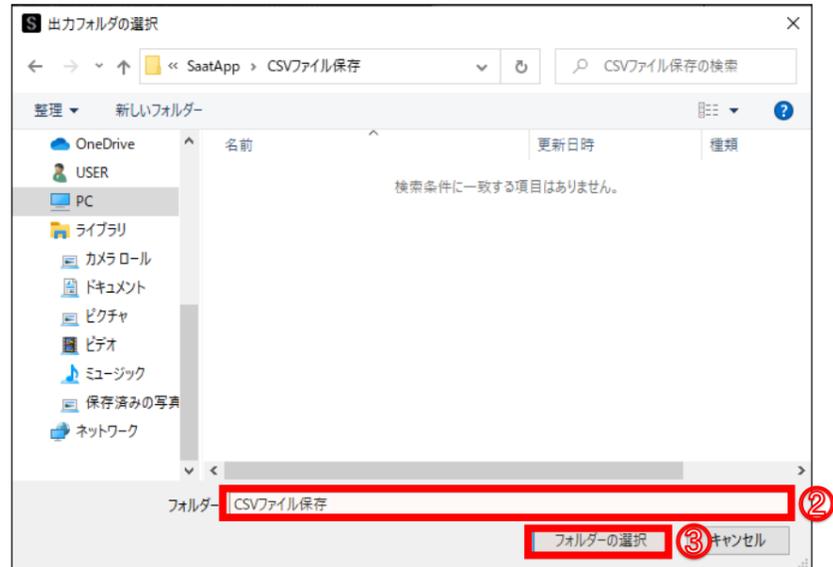
解析結果や加工したデータを CSV ファイルに出力できます。CSV ファイルを保存するには、あらかじめ保存先のフォルダを指定する必要があります。ここでは、CSV ファイルを保存するフォルダを指定する手順を説明します。

1 CSV ファイルの出力先フォルダを指定する

1. ツールバーから [出力フォルダ選択] ボタンをクリックします。
「出力フォルダの選択」が表示されます。



2. CSV ファイルを保存する任意のフォルダを選択します。
3. [フォルダーの選択] ボタンをクリックします。



2-7 CSV ファイルを出力する

解析結果加工したデータを CSV ファイルに出力できます。なお、CSV ファイルを保存するには、事前に CSV ファイルの保存先フォルダを指定する必要があります。以下の手順を実施してください。

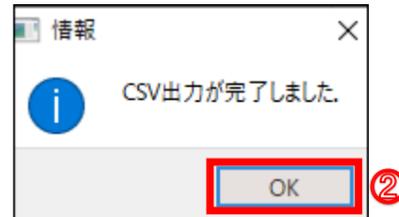
参考：2-6 CSV ファイルの出力フォルダを選択する (P.10)

1 CSV ファイルを出力する

1. ツールバーから [CSV 出力] ボタンをクリックします。
CSV ファイル出力完了のメッセージが表示されます。



2. [OK] ボタンをクリックします。
CSV ファイルの保存先フォルダに、CSV ファイルが保存されます。
データ操作画面で [CSV 出力] ボタンをクリックした場合は、“データ.csv”のファイル名で出力されます。



データ.csv ファイル保存の注意点

データ操作画面で [CSV 出力] ボタンをクリックした場合、出力されるファイル名は“データ.csv”固定です。

複数回 [CSV 出力] ボタンをクリックすると、最新のデータでデータ.csv ファイルが上書き保存されます。複数の異なる CSV ファイルを保存したい場合は、“データ.csv”ファイルからファイル名を変更するか、別のフォルダに CSV ファイルを保存してください。

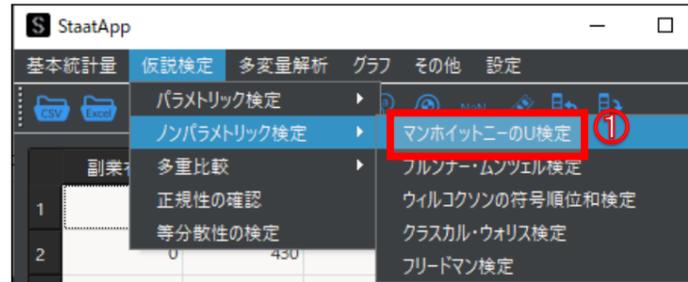
2-8 統計解析を実行する

統計解析の基本的な実行手順を説明します。ここでは例として、ノンパラメトリック検定の1つであるマンホイットニーのU検定の実行手順を説明します。手順では、サンプルデータ①.xlsxを使用し、性別列に入力された男性・女性の2群に対して収入を比較します。なお、事前に解析をするデータを読み込む必要があります。データを読み込む方法は、以下を参照してください。
参考：2-4 データを読み込む（CSV ファイル）（P.8）

1 マンホイットニーのU検定を実行する

1. メニューバーから、[仮説検定] → [ノンパラメトリック検定] → [マンホイットニーのU検定]を選択します。

「マンホイットニーのU検定」が表示されます。



2. [変数選択] の [群名列] プルダウンメニューから、比較する群名を選択します。

ここでは「性別」を選択します。

3. [群 A] プルダウンメニューから、比較の対象とする群名を選択します。

ここでは「男性」を選択します。

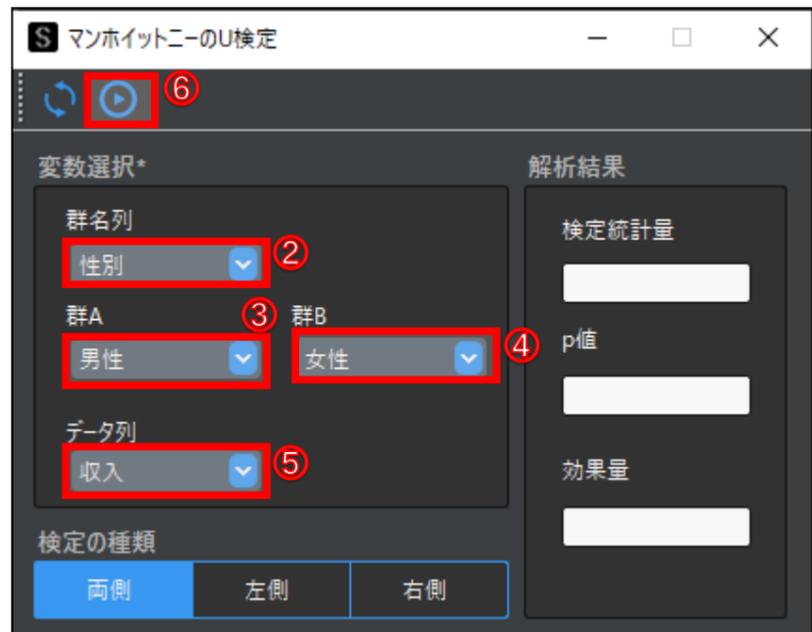
4. [群 B] プルダウンメニューから、比較の対象とする群名を選択します。

ここでは「女性」を選択します。

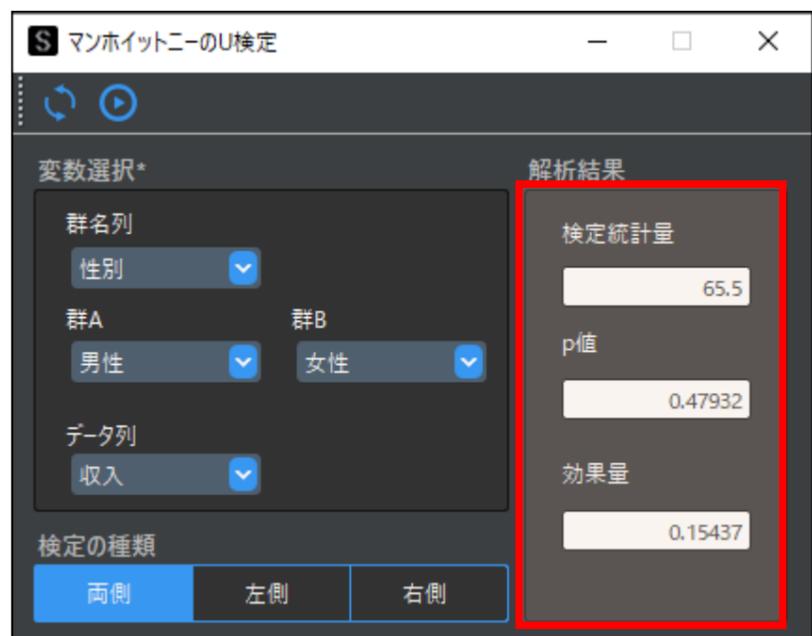
5. [データ列] プルダウンメニューから、比較する値が入力されているデータ（列名）を選択します。

ここでは「収入」を選択します。

6. 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] に解析結果が表示されます。

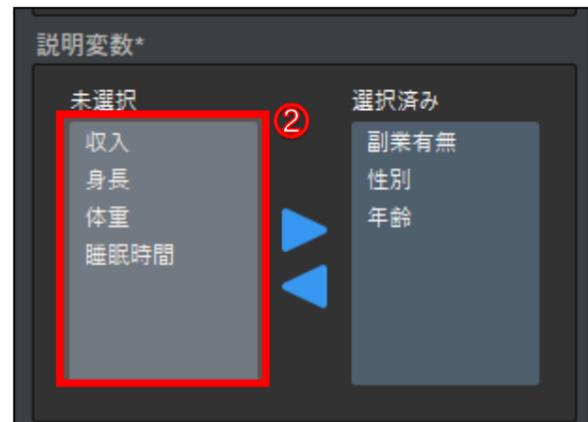


2-9 変数を選択する/選択した変数を未選択に戻す

検定の対象とする変数を選択する、または選択した変数を未選択に戻す手順を説明します。変数を1つずつ選択する方法と、すべての変数を一括で選択する方法があります。

1 変数を1つずつ選択する

1. 各種検定の実行画面を開きます。
2. [未選択] から、検定の対象とする変数を選択します。
選択した変数が [選択済み] に移動します。



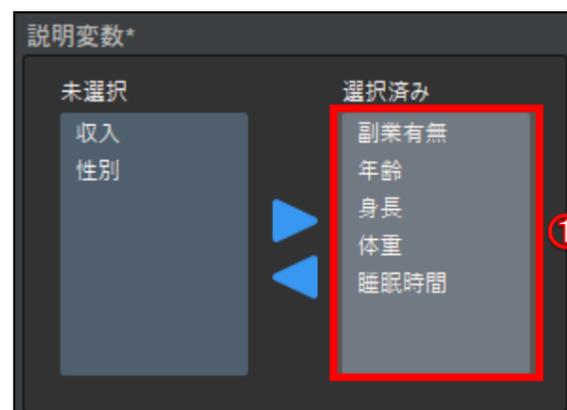
2 すべての変数を一括で選択する

1. 右矢印アイコンをクリックします。
[未選択] にあるすべての変数が [選択済み] に移動します。



3 選択した変数を1つずつ未選択に戻す

1. [選択済み] から、検定の対象外とする変数を選択します。
選択した変数が [未選択] に移動します。



4 すべての変数を一括で未選択に戻す

1. 左矢印アイコンをクリックします。
[選択済み] にあるすべての変数が [未選択] に移動します。



2-10 グラフの表示/保存/編集

各種統計解析結果のグラフや、ヒストグラム、分布図などのグラフを作成する場合、同じグラフを同時に複数表示できません。設定条件やグラフのレイアウトなどを変更したい場合は、現在表示しているグラフを閉じてからグラフを作成してください。なお、種類が異なるグラフは複数同時に表示できます。

また、作成したグラフはツールバーから画像として保存できます。[Savethefigure] ボタンをクリックすると、任意のフォルダに画像を保存できます。



3 データ操作

StaatApp では、置換機能や欠測値の処理機能を使用することで、データクレンジングやデータ操作ができます。データクレンジングの詳細については、以下を参照してください。

参考：[データクレンジング解説 | Staat](#)

以降で、データクレンジングとデータ操作に用いる以下の機能について説明します。

- ソート
- 置換
- 行削除
- 欠測値の処理
- ダミー変数作成
- データ型の変換

手順ではサンプルデータとして、以下のようにデータが入力されているダミーデータ.csv を使用します。

No.	副業有無	収入	性別	年齢	身長	体重	睡眠時間
1	有	580	男性	32	170	60	420
2	無	430	女性	28	156	43	480
3	有	800	男性	45	163	57	350
4	有	780	女	36	161	48	330
5	有	690	女性	42	158	49	350
6							
7	無	12000	女性	32	165	54	350
8	無	350	女性	23	161	45	440
9	有	620	男性	29	180	78	450
10	無	500	女性		150	42	380
11	無	430	男	33	168	69	430
12	有	590	女性	36	159	47	370
13	有	1200	男性	51	169	65	330
14	無	810	男性	53	174	70	340
15	無	620	女性				380
16	無	430	女性	31	156	45	400
17	有	570	女性	39	164	52	350
18	無	460	男	28	170	59	430
19	有	620	男性	30	178	71	340
20	無	320	男性	25	167	62	410
21	無	430	女性	41	166	58	400
22	有	570	女性	39	164	52	350

3-1 複数列を対象にしてソートする

ソート機能とは、数値データや文字列を、昇順/降順やアルファベット順などに並び替えることです。ソート機能を使ってデータを並び替えることで、データの表記ゆれを発見しやすくなるといったメリットがあります。StaatAppでは、単一列または複数列を基準にしてソートできます。

以降で、データをソートする手順を説明します。手順では、例として年齢>性別の順に並び替えします。

1 データをソートする（並び替える）

1. ツールバーから、[ソート] ボタンをクリックします。
「選択した変数でソート」が表示されます。



2. [変数一覧] から、基準とする変数（列名）を選択します。
選択した変数（列名）が[選択済み]に移動します。
[選択済み]の上位に表示されている変数（列名）ほど優先度が高くなります。
ここでは“年齢”と“性別”を選択し、年齢>性別の順にします。
3. [ソート順序] で [昇順] または [降順] にチェックを入れ、データの並び順を指定します。
4. [実行] ボタンをクリックします。



設定した内容でデータがソートされます。

No.	副業有無	収入	性別	年齢	身長
1	4 有	780.0	女	36.0	
2	8 無	350.0	女性	23.0	
3	2 無	430.0	女性	28.0	
4	16 無	430.0	女性	31.0	
5	7 無	12000.0	女性	32.0	
6	12 有	590.0	女性	36.0	
7	17 有	570.0	女性	39.0	
8	22 有	570.0	女性	39.0	
9	21 無	430.0	女性	41.0	
10	5 有	690.0	女性	42.0	
11	10 無	500.0	女性	nan	
12	15 無	620.0	女性	nan	
13	18 無	460.0	男	28.0	
14	11 無	430.0	男	33.0	
15	20 無	320.0	男性	25.0	

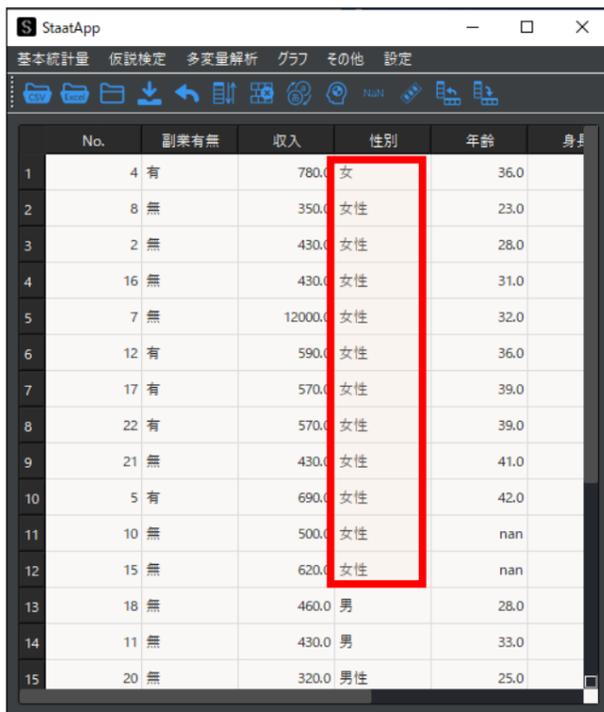
単一列でのソート

単一列（1つの列）でデータをソートする場合は、データ操作画面の列名部分を直接クリックすることも可能です。

3-2 置換する（表記ゆれの処理をする）

データに表記ゆれがある場合、数値や文字列を適切な値に置換して表記を統一できます。

手順では、例として“女”→“女性”に置換します。



No.	副業有無	収入	性別	年齢	身長
1	4 有	780.0	女	36.0	
2	8 無	350.0	女性	23.0	
3	2 無	430.0	女性	28.0	
4	16 無	430.0	女性	31.0	
5	7 無	12000.0	女性	32.0	
6	12 有	590.0	女性	36.0	
7	17 有	570.0	女性	39.0	
8	22 有	570.0	女性	39.0	
9	21 無	430.0	女性	41.0	
10	5 有	690.0	女性	42.0	
11	10 無	500.0	女性	nan	
12	15 無	620.0	女性	nan	
13	18 無	460.0	男	28.0	
14	11 無	430.0	男	33.0	
15	20 無	320.0	男性	25.0	

1 文字列を置換する

1. ツールバーから、[置換] ボタンをクリックします。

「置換」が表示されます。

2. [置換前の文字列] 欄に、置換元の文字列を入力します。

ここでは“女”と入力します。

3. [置換後の文字列] 欄に、置換後の文字列を入力します。

ここでは“女性”と入力します。

4. [対象列] から、置換の対象とする列を選択します。

ここでは“性別”を選択します。

5. 必要に応じて、[オプション] の [完全に一致] と [半角英数字に変換] をオンにします。

[完全に一致] をオンにすると、セル内の文字列が [置換前の文字列] に入力した文字列と一致する場合のみ置換します。

[半角英数字に変換] をオンにすると、全角英数字を半角英数字に変換します。

ここでは [完全に一致] をオンにします。

6. [実行] ボタンをクリックします。

“女”→“女性”に置換されます。



置換

置換前の文字列: 女 (2)

置換後の文字列: 女性 (3)

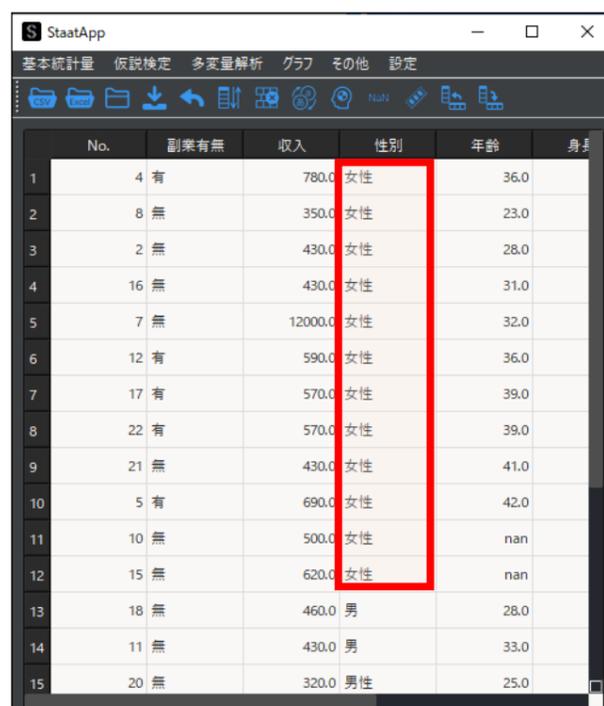
対象列: 性別 (4)

オプション

完全に一致 (5)

半角英数字に変換

実行 (6)



No.	副業有無	収入	性別	年齢	身長
1	4 有	780.0	女性	36.0	
2	8 無	350.0	女性	23.0	
3	2 無	430.0	女性	28.0	
4	16 無	430.0	女性	31.0	
5	7 無	12000.0	女性	32.0	
6	12 有	590.0	女性	36.0	
7	17 有	570.0	女性	39.0	
8	22 有	570.0	女性	39.0	
9	21 無	430.0	女性	41.0	
10	5 有	690.0	女性	42.0	
11	10 無	500.0	女性	nan	
12	15 無	620.0	女性	nan	
13	18 無	460.0	男	28.0	
14	11 無	430.0	男	33.0	
15	20 無	320.0	男性	25.0	

データを直接編集する

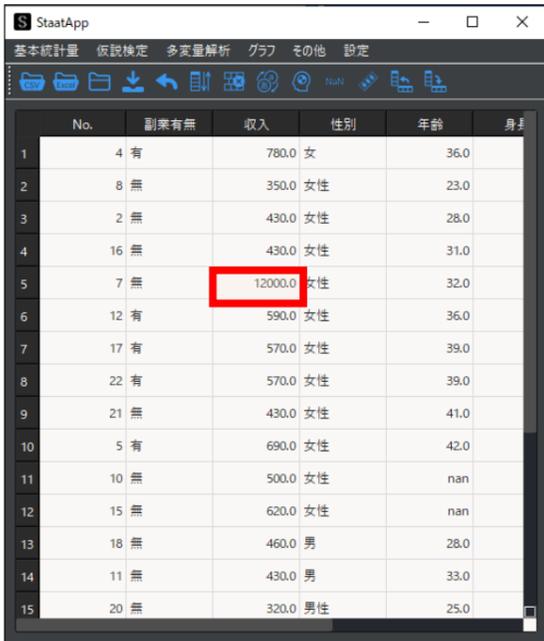
編集したいセルをダブルクリックすると、データを直接編集できます。

データ型が異なるデータの置換

データ型が異なるデータは置換できません。例えば、“収入”の“780”（数値型データ）を“なし”（文字列型データ）には置換できません。データ型が異なるデータを置換したい場合は、データ型を同一にしてから置換してください。

3-3 特定の行を削除する（不要データの削除・外れ値の処理）

不要なデータや外れ値（他のデータより大幅に数値が異なる値）がある場合、特定の行を削除できます。単一の行または連続した複数の行を削除できます。手順では、例として収入が他と比べて大きい（外れ値）5行目のデータを削除します。



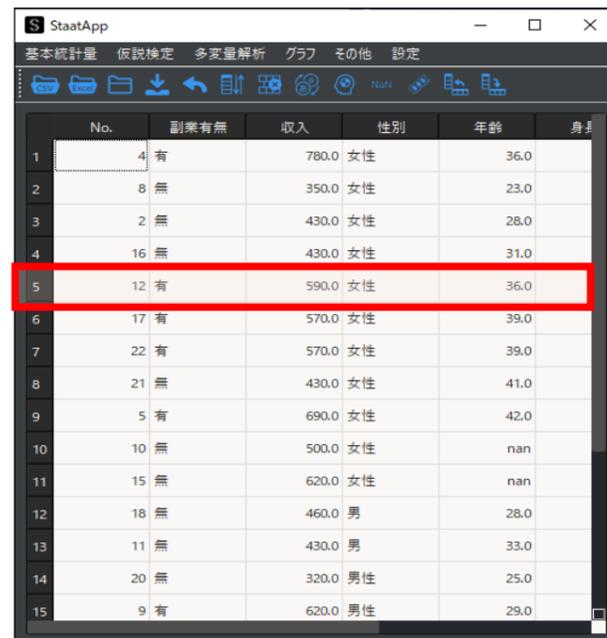
No.	副業有無	収入	性別	年齢	身長
1	4 有	780.0	女性	36.0	
2	8 無	350.0	女性	23.0	
3	2 無	430.0	女性	28.0	
4	16 無	430.0	女性	31.0	
5	7 無	12000.0	女性	32.0	
6	12 有	590.0	女性	36.0	
7	17 有	570.0	女性	39.0	
8	22 有	570.0	女性	39.0	
9	21 無	430.0	女性	41.0	
10	5 有	690.0	女性	42.0	
11	10 無	500.0	女性	nan	
12	15 無	620.0	女性	nan	
13	18 無	460.0	男性	28.0	
14	11 無	430.0	男性	33.0	
15	20 無	320.0	男性	25.0	

1 行を削除する

1. ツールバーから、[行の削除] ボタンをクリックします。
「行の削除」が表示されます。
2. [行番号（上限）] に、削除したい行番号を入力します。
入力した数値が自動で [行番号（下限）] に入力されます。
3. 連続した複数の行を削除したい場合は、[行番号（下限）] に削除したい末尾の行番号を入力します。
4. [実行] ボタンをクリックします。



5行目のデータが削除されます。



No.	副業有無	収入	性別	年齢	身長
1	4 有	780.0	女性	36.0	
2	8 無	350.0	女性	23.0	
3	2 無	430.0	女性	28.0	
4	16 無	430.0	女性	31.0	
5	12 有	590.0	女性	36.0	
6	17 有	570.0	女性	39.0	
7	22 有	570.0	女性	39.0	
8	21 無	430.0	女性	41.0	
9	5 有	690.0	女性	42.0	
10	10 無	500.0	女性	nan	
11	15 無	620.0	女性	nan	
12	18 無	460.0	男性	28.0	
13	11 無	430.0	男性	33.0	
14	20 無	320.0	男性	25.0	
15	9 有	620.0	男性	29.0	

3-4 欠測値の処理をする

StaatApp では、空白データなどの欠測値を読み込んだ場合“nan”と表示されます。“nan”データを一括処理する、欠測値の処理について説明します。

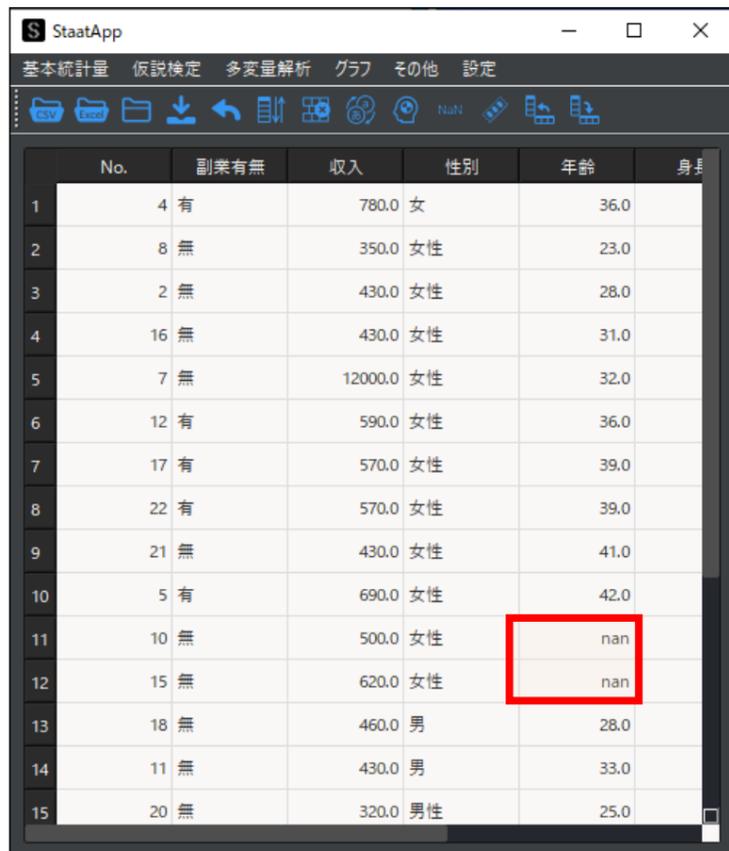
StaatApp での欠測値の処理方法には、以下の3つがあります。

- リストワイズ除去：欠測値がある行を削除する
- ペアワイズ除去：すべての変数が欠測値である行のみ削除する
- 補完：データ全体の平均値に変換する

参考：3 データ操作 (P.15)

手順では、例として年齢列の欠測値を平均値に補完します。

なお、文字列の平均値を求めることができないため、文字列データは補完できません。文字列データを修正したい場合は、直接データを編集/削除するか、リストワイズ除去で欠測値を削除してください。



No.	副業有無	収入	性別	年齢	身長
1	4 有	780.0	女	36.0	
2	8 無	350.0	女性	23.0	
3	2 無	430.0	女性	28.0	
4	16 無	430.0	女性	31.0	
5	7 無	12000.0	女性	32.0	
6	12 有	590.0	女性	36.0	
7	17 有	570.0	女性	39.0	
8	22 有	570.0	女性	39.0	
9	21 無	430.0	女性	41.0	
10	5 有	690.0	女性	42.0	
11	10 無	500.0	女性	nan	
12	15 無	620.0	女性	nan	
13	18 無	460.0	男	28.0	
14	11 無	430.0	男	33.0	
15	20 無	320.0	男性	25.0	

1 欠測値の処理をする

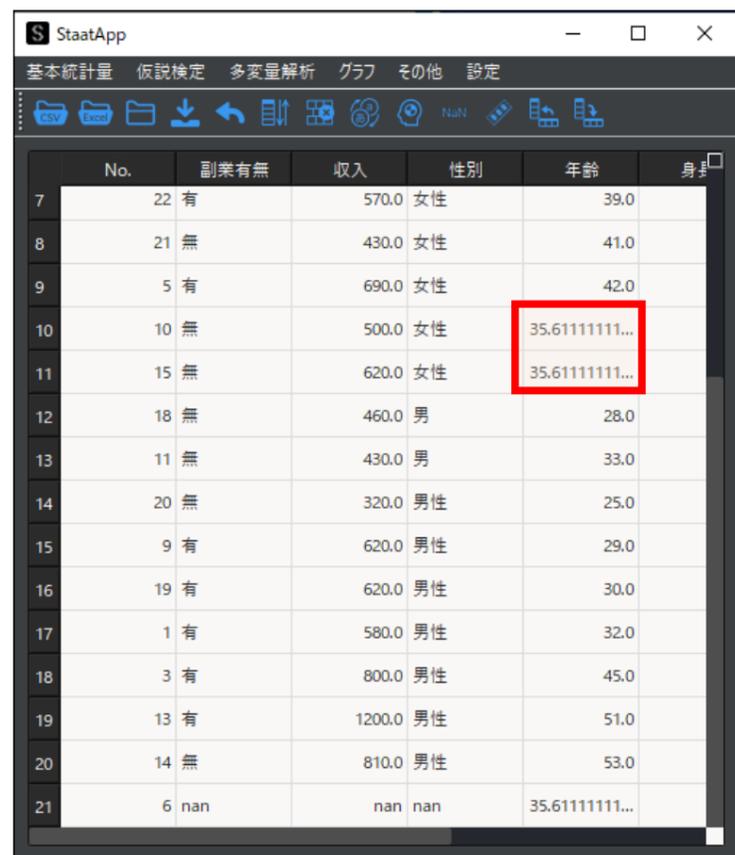
1. ツールバーから、[欠測値の処理] ボタンをクリックします。
「欠測値の処理」が表示されます。
2. [欠測値の検出] で欠測値の数を確認します。
3. [処理対象列] プルダウンメニューから、処理の対象とする列名を選択します。
ここでは“年齢”を選択します。
4. [処理方法] プルダウンメニューから、処理方法を選択します。
ここでは[補完]を選択します。
5. [補完する値] から、補完方法にチェックを入れます。
ここでは[平均値]にチェックを入れます。
6. [実行] ボタンをクリックします。



[補完する値] の各種補完方法

- 平均値：データ全体の平均値に変換します。
- 中央値：データを値が小さいものから順に並べたときに、真ん中にある値に変換します。
- 固定値：[固定値] 欄に入力した値に変換します。

年齢の欠測値が平均値に変換されます。



No.	副業有無	収入	性別	年齢	身長
7	22 有	570.0	女性	39.0	
8	21 無	430.0	女性	41.0	
9	5 有	690.0	女性	42.0	
10	10 無	500.0	女性	35.61111111111111	
11	15 無	620.0	女性	35.61111111111111	
12	18 無	460.0	男	28.0	
13	11 無	430.0	男	33.0	
14	20 無	320.0	男性	25.0	
15	9 有	620.0	男性	29.0	
16	19 有	620.0	男性	30.0	
17	1 有	580.0	男性	32.0	
18	3 有	800.0	男性	45.0	
19	13 有	1200.0	男性	51.0	
20	14 無	810.0	男性	53.0	
21	6 nan	nan	nan	35.61111111111111	

3-5 ダミー変数へ変換する

ダミー変数とは、文字列データ（数値以外のデータ）を0または1に変換した変数のことです。

例えば、“性別”列の“女性”というデータをダミー変数に変換した場合、“女性”列が追加され、0と1のデータが表示されます。この場合、1が女性、0が男性であることを表します。

以降で、文字列データをダミー変数に変換する手順を説明します。手順では例として、“性別”列の“男性”と“女性”をダミー変数に変換します。

1 ダミー変数に変換する

1. ツールバーから、[ダミー変数に変換] ボタンをクリックします。

「ダミー変数に変換」が表示されます。



2. [未選択] から、ダミー変数に変換する変数を選択します。

クリックした変数が [選択済み] に移動します。

ここでは“性別”を選択します。

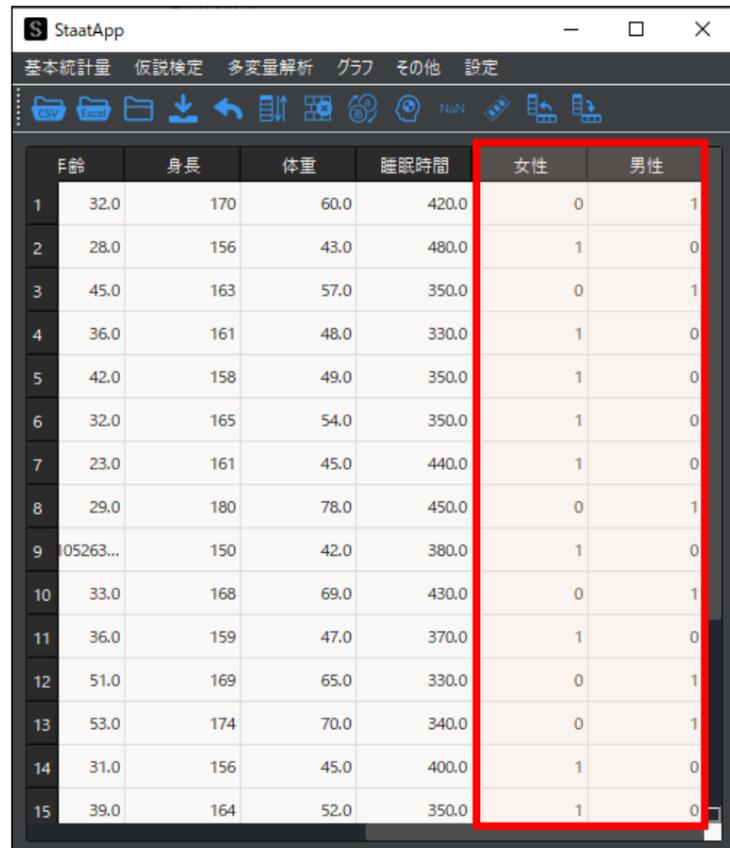
3. [実行] ボタンをクリックします。

[選択済み] に設定した列にある変数がダミー変数となり、データの最終列に追加されます。



“女性”列に0と表示されているのが“男性”，1と表示されているのが“女性”です。

“男性”列に0と表示されているのが“女性”，1と表示されているのが“男性”です。



	年齢	身長	体重	睡眠時間	女性	男性
1	32.0	170	60.0	420.0	0	1
2	28.0	156	43.0	480.0	1	0
3	45.0	163	57.0	350.0	0	1
4	36.0	161	48.0	330.0	1	0
5	42.0	158	49.0	350.0	1	0
6	32.0	165	54.0	350.0	1	0
7	23.0	161	45.0	440.0	1	0
8	29.0	180	78.0	450.0	0	1
9	05263...	150	42.0	380.0	1	0
10	33.0	168	69.0	430.0	0	1
11	36.0	159	47.0	370.0	1	0
12	51.0	169	65.0	330.0	0	1
13	53.0	174	70.0	340.0	0	1
14	31.0	156	45.0	400.0	1	0
15	39.0	164	52.0	350.0	1	0

3-6 異なるデータ型に変換する

StaatApp では、数値型と文字列型の 2 つのデータ型を定義します。

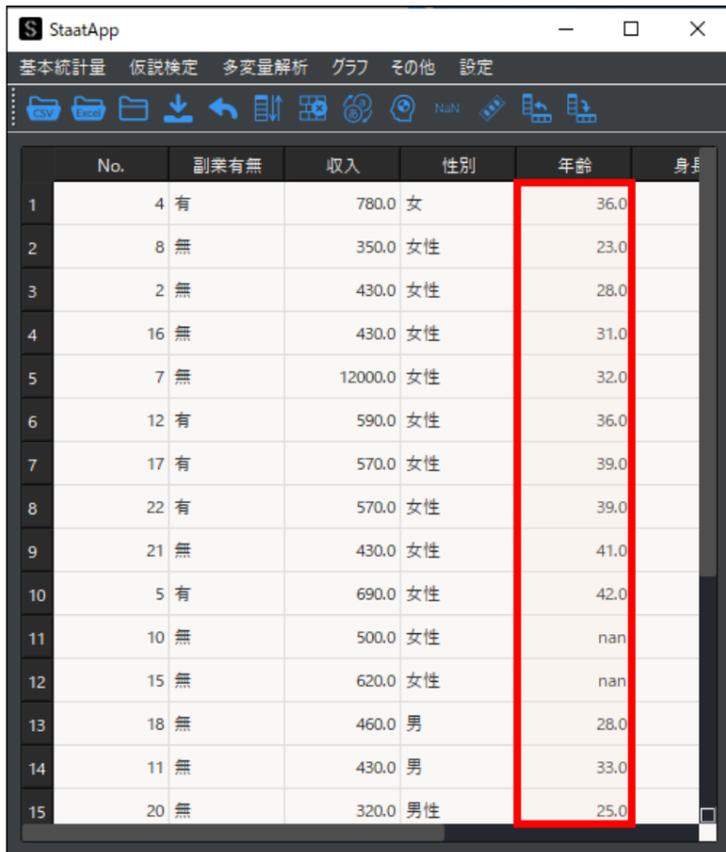
- 数値型：半角数字で入力された値のことで、演算など順序尺度以上の統計解析ができます。
- 文字列型：半角英字または全角文字で入力された値のことで、クロス集計表などのカテゴリ変数に対する統計解析ができます。

StaatApp ではデータを読み込んだ際に自動でデータ型を判別し、文字列型は値が左寄せ、数値型は値が右寄せで表示されます。

手順では、例として“年齢”の数値型データを文字列型に変換します。

データ型の種類の詳細については、以下を参照してください。

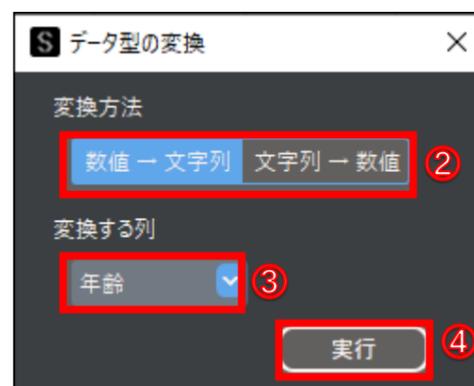
参考：[統計学での尺度とは | Staat](#)



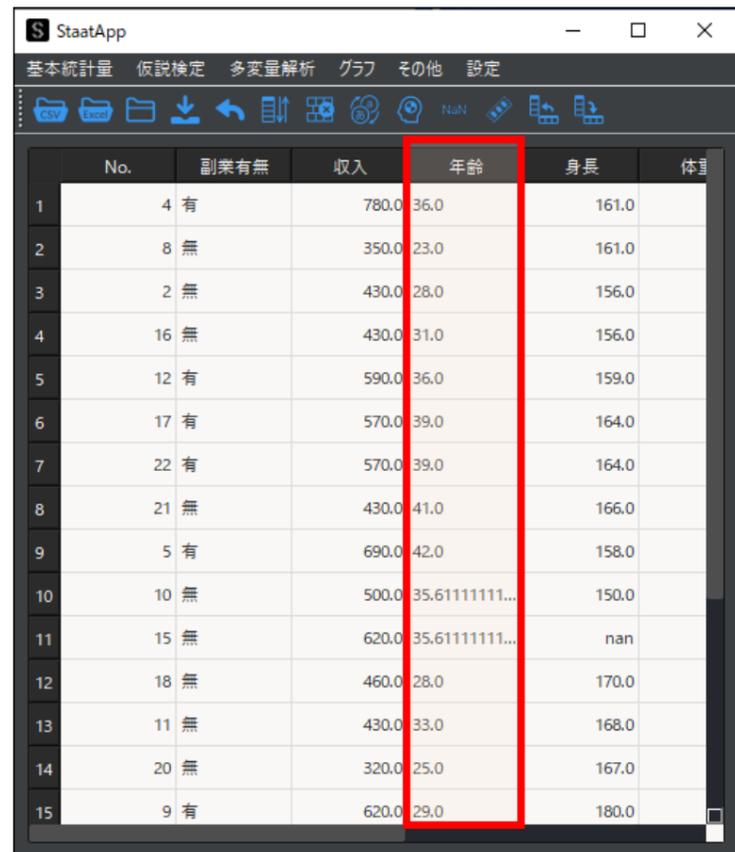
No.	副業有無	収入	性別	年齢	身長
1	4 有	780.0	女	36.0	
2	8 無	350.0	女性	23.0	
3	2 無	430.0	女性	28.0	
4	16 無	430.0	女性	31.0	
5	7 無	12000.0	女性	32.0	
6	12 有	590.0	女性	36.0	
7	17 有	570.0	女性	39.0	
8	22 有	570.0	女性	39.0	
9	21 無	430.0	女性	41.0	
10	5 有	690.0	女性	42.0	
11	10 無	500.0	女性	nan	
12	15 無	620.0	女性	nan	
13	18 無	460.0	男	28.0	
14	11 無	430.0	男	33.0	
15	20 無	320.0	男性	25.0	

1 データ型を変換する

1. ツールバーから、[データ型の変換] ボタンをクリックします。
「データ型の変換」が表示されます。
2. [変換方法] で変換方法を選択します。
ここでは [数値→文字列] を選択します。
3. [変換する列] プルダウンメニューから、データ型を変換する列を選択します。
ここでは“年齢”を選択します。
4. [実行] ボタンをクリックします。



“年齢”列のデータ型が文字列型に変更され、値が左寄せに表示されます。



The screenshot shows the StaatApp interface with a data table. The '年齢' column is highlighted with a red box, showing values that are left-aligned, indicating they are now strings. The table has columns for No., 副業有無, 収入, 年齢, 身長, and 体重.

No.	副業有無	収入	年齢	身長	体重
1	4 有	780.0	36.0	161.0	
2	8 無	350.0	23.0	161.0	
3	2 無	430.0	28.0	156.0	
4	16 無	430.0	31.0	156.0	
5	12 有	590.0	36.0	159.0	
6	17 有	570.0	39.0	164.0	
7	22 有	570.0	39.0	164.0	
8	21 無	430.0	41.0	166.0	
9	5 有	690.0	42.0	158.0	
10	10 無	500.0	35.611111111...	150.0	
11	15 無	620.0	35.611111111...	nan	
12	18 無	460.0	28.0	170.0	
13	11 無	430.0	33.0	168.0	
14	20 無	320.0	25.0	167.0	
15	9 有	620.0	29.0	180.0	

4 解析手法別操作

各種解析手法の手順を説明します。StaatApp で扱う解析手法は、以下の表のとおりです。

大分類	中分類	解析手法	
基本統計量	基本統計量	平均値・標準偏差・四分位数など	
	相関係数	ピアソンの積率相関係数	
		スピアマンの順位相関係数	
		ケンドールの順位相関係数	
		相関比	
仮説検定	パラメトリック検定	対応のない t 検定	
		ウェルチの t 検定	
		対応のある t 検定	
		一元配置分散分析	
		反復測定分散分析	
		多元配置分散分析	
	ノンパラメトリック検定	マンホイットニーの U 検定	
		ブルンナー・ムンチェル検定	
		ウィルコクソン符号順位和検定	
		クラスカル・ウォリス検定	
		フリードマン検定	
	多重比較	テューキー・クレーマー法	
		スティーラー・ドゥワース検定	
		ダネット検定	
	正規性の確認	正規 Q-Q プロット	
		シャピロ・ウィルク検定	
		コルモゴロフ・スミルノフ検定	
	等分散性の検定	F 検定	
		バートレット検定	
		ルビーン検定	
	多変量解析	-	重回帰分析
			数量化 II 類
			ロジスティック回帰分析
			多項ロジスティック回帰分析

多変量解析	-	主成分分析
		因子分析
		K-means 法
		階層クラスター分析
グラフ	-	ヒストグラム
		分布図
		散布図行列・ヒートマップ
その他	クロス集計表	クロス集計表
		カイ二乗検定
		フィッシャーの正確確率検定
		クラメールの連関係数
		オッズ比とリスク比
		コレスポンデンス分析
	生存時間解析	生存曲線
		Cox比例ハザード回帰

なお、解析を実行する前に、解析するデータのデータクレンジング及びデータの読み込みが必要です。

4-1 基本統計量

基本統計量を求める手順を説明します。基本統計量では、平均値や標準偏差を求められます。手順では、例としてサンプルデータ①.csvを使用します。

1 基本統計量を算出する

1. メニューバーから、[基本統計量] → [基本統計量] を選択します。
「基本統計量」が表示されます。



2. [変数選択] の [未選択] から、基本統計量を算出したい変数を選択します。
選択した変数が [選択済み] に移動します。
ここではすべての変数を選択します。

3. ツールバーから、[解析実行] ボタンをクリックします。

変数ごとの基本統計量が算出され、[解析結果] に平均値や標準偏差が表示されます。



	No.	収入	年齢
サンプルサイズ	22.0	21.0	19
平均	11.5	1133.33333	35.421
標準偏差	6.49359	2497.60752	8.288
最小値	1.0	320.0	23
第一四分位数	6.25	430.0	25
中央値	11.5	580.0	33
第三四分位数	16.75	690.0	40
最大値	22.0	12000.0	53

4-2 相関係数

相関係数とは、2つの変数の関係性を調べる統計量です。2つの変数のデータの種類によって、相関係数を使い分ける必要があります。

相関係数の使い分けの詳細については、以下を参照してください。

参考：[相関係数 | Staat](#)

以降で、相関係数を求める手順を説明します。手順では、例としてサンプルデータ①.csvを使用します。

1 相関係数を算出する

1. メニューバーから、[基本統計量] → [相関係数] を選択します。

「相関係数」が表示されます。



2. [変数選択] の [変数 A] プルダウンメニューから、相関係数を計算する変数 (列) を選択します。

ここでは「収入」を選択します。

3. [変数選択] の [変数 B] プルダウンメニューから、相関係数を計算する変数 (列) を選択します。

ここでは「年齢」を選択します。

4. [種類] から、算出したい相関係数の種類を選択します。

ここでは [スピアマン] を選択します。

5. [解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。



相関比を計算する

相関比を計算したい場合は、[変数 A] でカテゴリ変数を選択してください。

相関係数の種類

各種相関係数の詳細については、以下を参照してください。

参考：[スピアマンの順位相関係数 | Staat](#)

参考：[ケンドールの順位相関係数 | Staat](#)

参考：[相関比 | Staat](#)

4-3 パラメトリック検定

パラメトリック検定とは、データの母集団が正規分布に従う場合に行う仮説検定です。データの母集団が正規分布に従うと仮定できない場合は、ノンパラメトリック検定を実行します。

StaatApp で実行できるパラメトリック検定は、以下のとおりです。

- 対応のない t 検定
- ウェルチの t 検定
- 対応のある t 検定
- 一元配置分散分析
- 反復測定分散分析
- 多元配置分散分析

以降で、対応のない t 検定と多元配置分散分析の解析手順を分けて説明します。手順では、例としてダーティデータ.csv を使用します。

パラメトリック検定

パラメトリック検定の各手法の詳細については、以下を参照してください。

参考：[対応のない t 検定 | Staat](#), 4-3-1 対応のない t 検定 (P.29)

参考：[ウェルチの t 検定 | Staat](#)

参考：[対応のある t 検定 | Staat](#)

参考：[分散分析 | Staat](#)

参考：[テューキー法とテューキー・クレーマー法 | Staat](#)

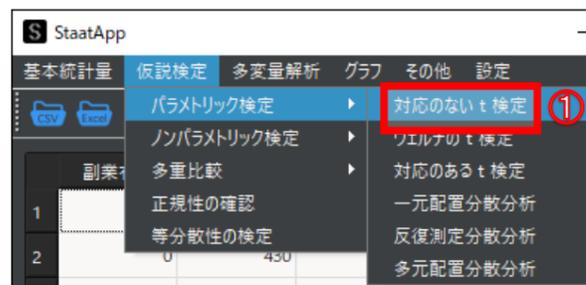
4-3-1 対応のない t 検定

対応のない t 検定の実行手順を説明します。ここでは例として、副業の有無によって収入に違いがあるかどうかを解析します。

1 対応のない t 検定を実行する

1. メニューバーから、[仮説検定] → [パラメトリック検定] → [対応のない t 検定] を選択します。

「対応のない t 検定」が表示されます。



2. [変数選択] の [群名列] プルダウンメニューから、検定の対象とする変数 (列) を選択します。

ここでは“副業有無”を選択します。

3. [群 A] プルダウンメニューから、比較の対象とする変数 (列) を選択します。

ここでは“有”を選択します。

4. [群 B] プルダウンメニューから、比較の対象とする変数 (列) を選択します。

ここでは“無”を選択します。

5. [データ列] プルダウンメニューから、比較したい変数 (列) を選択します。

ここでは“収入”を選択します。



6. [検定の種類] から検定方法を選択します。

ここでは“群 A が群 B より大きい”ことを比較したいため、[右側] を選択します。

7. ツールバーから、[解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。

解析結果から、“副業をしている人としていない人で年収に差がある”ことが分かります。

検定の種類

どのように比較したいかによって、検定の種類が異なります。

- 両側：[群 A] と [群 B] の違いを比較する
- 左側：[群 A] が [群 B] より小さいことを調べる
- 右側：[群 B] が [群 A] より大きい/小さいことを調べる

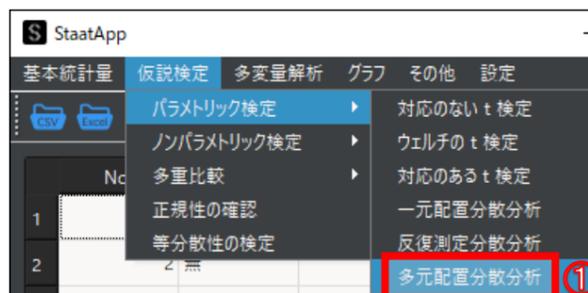
4-3-2 多元配置分散分析

多元配置分散分析の実行手順を説明します。以降で例として、“副業有無”と“性別”の2つの要因による収入の影響・交互作用を調べます。

1 多元配置分散分析を実行する

1. メニューバーから、[仮説検定] → [パラメトリック検定] → [多元配置分散分析] を選択します。

「多元配置分散分析」が表示されます。



2. [データ列選択] プルダウンメニューから、検定の対象とする変数（列）を選択します。

ここでは“収入”を選択します。

3. [変数選択] の [一覧] から、要因とする変数（列）を選択します。

クリックした変数（列）が [選択済み] に移動します。

ここでは“副業有無”と“性別”を選択します。

2つ選択した場合は2元配置分散分析が、3つ選択した場合は3元配置分散分析が実行されます。

4. ツールバーから、[解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。

解析結果の [副業有無:性別] は交互作用を示します。

解析結果から、“副業有無と性別に交互作用がない”ことが分かります。



その他の解析結果の見方や二元配置分散分析の考え方は、以下を参照してください。

[参考：二元配置分散分析 | Staat](#)

4-4 マンホイットニーの U 検定

マンホイットニーの U 検定とは、対応のない 2 群の差の検定（対応のない t 検定）のノンパラメトリック版です。順序尺度以上のデータに対して用いることができます。ウィルコクソンの順位和検定と呼ばれることもありますが、手法自体は同じです。

マンホイットニーの U 検定は、サンプル数によって p 値の計算方法が異なります。StaatApp では、サンプル数に合った計算方法を自動で選択して計算できます。

マンホイットニーの U 検定の考え方や用いるデータの前提条件の詳細については、以下を参照してください。

参考：[マンホイットニーの U 検定（小標本） | Staat](#)

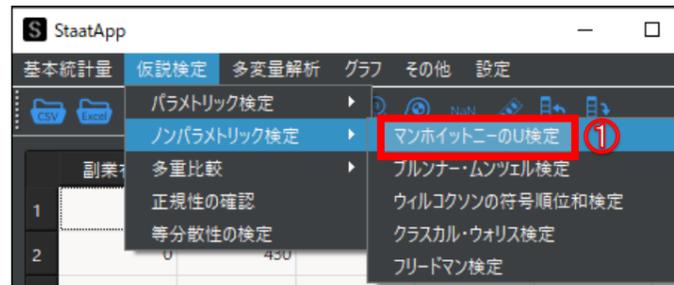
参考：[マンホイットニーの U 検定（大標本） | Staat](#)

以降で、マンホイットニーの U 検定の実行手順を説明します。手順では、例としてサンプルデータ①.xlsx を使用し、性別によって男女の年齢に違いがあるかどうかを解析します。

1 マンホイットニーの U 検定を実行する

1. メニューバーから、[仮説検定] → [ノンパラメトリック検定] → [マンホイットニーの U 検定] を選択します。

「マンホイットニーの U 検定」が表示されます。



2. [変数選択] の [群列名] プルダウンメニューから、検定の対象とする変数（列）を選択します。

ここでは「性別」を選択します。

3. [群 A] プルダウンメニューから、比較の対象とする変数（列）を選択します。

ここでは「女性」を選択します。

4. [群 B] プルダウンメニューから、比較の対象とする変数（列）を選択します。

ここでは「男性」を選択します。

5. [データ列] プルダウンメニューから、比較したい変数（列）を選択します。

ここでは「年齢」を選択します。



6. [検定の種類] から検定方法を選択します。

ここでは [両側] を選択します。

片側検定を行いたい場合は、[左側] または [右側] を選択します。

7. ツールバーから、[解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。

解析結果の [副業有無:性別] は交互作用を示します。

解析結果から、「年齢に差があるとは言えない」ことがわかります。

検定の種類

どのように比較したいかによって、検定の種類が異なります。

- 両側： [群 A] と [群 B] の違いを比較する
- 左側： [群 A] が [群 B] より小さいことを調べる
- 右側： [群 B] が [群 A] より大きいことを調べる

4-5 ウィルコクソンの符号順位和検定

ウィルコクソンの符号順位和検定とは、対応のある2群の差の検定（対応のあるt検定）のノンパラメトリック版です。順序尺度以上のデータに対して用いることができます。

ウィルコクソンの符号順位和検定は、サンプル数によってp値の計算方法が異なります。StaatAppでは、サンプル数に合った計算方法を自動で選択して計算できます。

検定の考え方や用いるデータの前提条件の詳細については、以下を参照してください。

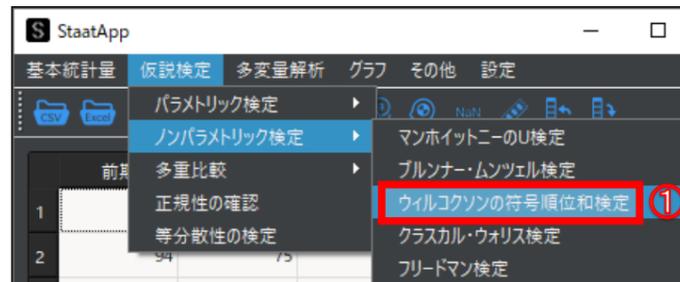
参考：[ウィルコクソンの符号順位和検定 | Staat](#)

以降で、ウィルコクソンの符号順位和検定の実行手順を説明します。手順では、例としてサンプルデータ②.csvを使用し、前期と後期でテスト結果に違いがあるかどうかを解析します。

1 ウィルコクソンの順位符号和検定を実行する

1. メニューバーから、[仮説検定] → [ノンパラメトリック検定] → [ウィルコクソンの順位符号和検定] を選択します。

「ウィルコクソンの順位符号和検定」が表示されます。



2. [変数列選択] の [群 A] プルダウンメニューから、比較の対象とする変数（列）を選択します。

ここでは「前期」を選択します。

3. [変数列選択] の [群 B] プルダウンメニューから、比較の対象とする変数（列）を選択します。

ここでは「後期」を選択します。

4. [検定の種類] から検定方法を選択します。

ここでは [両側] を選択します。

片側検定を行いたい場合は、[左側] または [右側] をクリックします。

5. ツールバーから、[解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。

解析結果から、「前期と後期の点数に差がある」ことがわかります。



検定の種類

どのように比較したいかによって、検定の種類が異なります。

- 両側： [群 A] と [群 B] の違いを比較する
- 左側： [群 A] が [群 B] より小さいことを調べる
- 右側： [群 B] が [群 A] より大きいことを調べる

4-6 クラスカル・ウォリス検定

クラスカル・ウォリス検定とは、対応のない3群以上の差の検定（一次元配置分散分析）のノンパラメトリック版です。順序尺度以上のデータに対して用いることができます。

クラスカル・ウォリス検定は、サンプル数によって p 値の計算方法が異なります。StaatApp では、サンプル数に合った計算方法を自動で選択して計算できます。

検定の考え方や用いるデータの前提条件の詳細については、以下を参照してください。

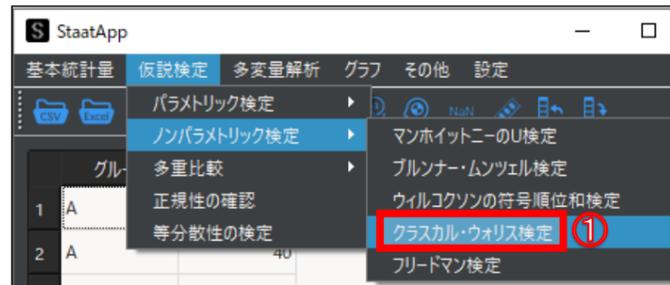
参考：[クラスカル・ウォリス検定 | Staat](#)

以降で、クラスカル・ウォリス検定の実行手順を説明します。手順では、例としてサンプルデータ③.csv を使用し、クラスによって点数に違いがあるかどうかを解析します。

1 クラスカル・ウォリス検定を実行する

1. メニューバーから、[仮説検定] → [ノンパラメトリック検定] → [クラスカル・ウォリス検定] を選択します。

「クラスカル・ウォリス検定」が表示されます。



2. [変数選択] の [群名列] プルダウンメニューから、比較する変数（列）を選択します。

[群選択] の [未選択] に、選択した変数（列）にある値が表示されます。

ここでは“グループ”を選択します。

3. [変数選択] の [データ列] プルダウンメニューから、比較の対象とする変数（列）を選択します。

ここでは“データ”を選択します。

4. [群選択] の [未選択] から、比較の対象とする値を選択します。

選択した値が [選択済み] に移動します。

ここではすべての値を選択します。

5. ツールバーから、[解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。

解析結果から、“クラス間に点数の差がある”ことが分かります。



4-7 フリードマン検定

フリードマン検定とは、対応のある3群以上の差の検定（対応のある一次元配置分散分析）のノンパラメトリック版です。順序尺度以上のデータに対して用いることができます。

フリードマン検定は、サンプル数によってp値の計算方法が異なります。StaatAppでは、サンプル数に合った計算方法を自動で選択して計算できます。

検定の考え方や用いるデータの前提条件の詳細については、以下を参照してください。

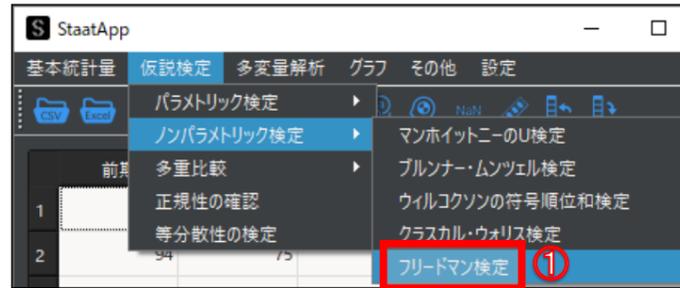
参考：[フリードマン検定 | Staat](#)

以降で、フリードマン検定の実行手順を説明します。手順では、例としてサンプルデータ②.csvを使用し、前期・中期・後期のテストの点数に違いがあるかどうかを解析します。

1 フリードマン検定を実行する

1. メニューバーから、[仮説検定] → [ノンパラメトリック検定] → [フリードマン検定] を選択します。

「フリードマン検定」が表示されます。



2. [変数列選択] の [未選択] から、比較の対象とする変数（列）をクリックします。

クリックした変数（列）が [選択済み] に移動します。

ここではすべての変数（列）を選択します。

3. ツールバーから、[解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。

解析結果から、“前期・中期・後期のテストの点数に差がある”ことがわかります。



4-8 テューキー・クレーマー法

テューキー法とは、対応のない t 検定を 3 群以上の標本に対して使えるようにした検定方法です。検出力が高いため、多重比較の中で最も一般的に使用されます。比較方法では、すべての 2 群同士を平均の差が大きいかどうかで判断します。

テューキー・クレーマー法は Excel などを用いて計算する場合、検定統計量と専用の分布表を用いる必要があります。StaatApp では、サンプル数に合った計算方法を自動で選択して計算できます。

検定の考え方や用いるデータの前提条件の詳細については、以下を参照してください。

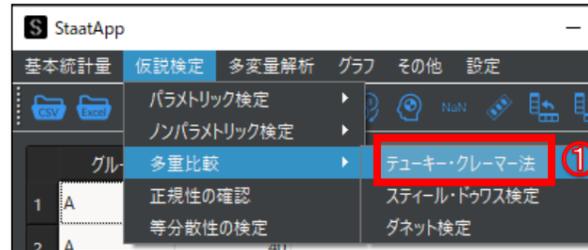
参考：[テューキー・クレーマー法 | Staat](#)

以降で、テューキー・クレーマー法の実行手順を説明します。手順では、例としてサンプルデータ③.csv を使用し、クラスによって点数に違いがあるかどうかを解析します。

1 テューキー・クレーマー法を実行する

1. メニューバーから、[仮説検定] → [多重比較] → [テューキー・クレーマー法] を選択します。

[テューキー・クレーマー法] が表示されます。



2. [変数選択] の [群名列] プルダウンメニューから、比較する変数 (列) を選択します。

[群選択] の [未選択] に、選択した変数 (列) がある値が表示されます。

ここでは“グループ”を選択します。

3. [変数選択] の [データ列] プルダウンメニューから、比較の対象とする変数 (列) を選択します。

ここでは“データ”を選択します。

4. [群選択] の [未選択] から、比較の対象とする値を選択します。

選択した値が [選択済み] に移動します。

ここではすべての値を選択します。

5. ツールバーから、[解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。

解析結果から、“A と B、B と C に点数の差がある”ことが分かります。



4-9 スティール・ドゥワス検定

スティール・ドゥワス検定とは、マンホイットニーの U 検定を 3 群以上の標本に対して使えるようにした方法です。多重比較の中で最も一般的な手法である、チューキー法のノンパラメトリック版でもあります。比較方法では、すべての 2 群同士を代表値の差が大きいかどうかで判断します。

検定の考え方や用いるデータの前提条件の詳細については、以下を参照してください。

参考：[スティール・ドゥワス検定 | Staat](#)

以降で、スティール・ドゥワス検定の実行手順を説明します。手順では、例としてサンプルデータ③.csv を使用し、クラスによって点数に違いがあるかどうかを解析します。

1 スティール・ドゥワス検定を実行する

1. メニューバーから、[仮説検定] → [多重比較] → [スティール・ドゥワス検定] を選択します。

「スティール・ドゥワス検定」が表示されます。



2. [変数選択] の [群名列] プルダウンメニューから、比較する変数 (列) を選択します。

[群選択] の [未選択] に、選択した変数 (列) がある値が表示されます。

ここでは「グループ」を選択します。

3. [変数選択] の [データ列] プルダウンメニューから、比較の対象とする変数 (列) を選択します。

ここでは「データ」を選択します。

4. [群選択] の [未選択] から、比較の対象とする値を選択します。

選択した値が [選択済み] に移動します。

ここではすべての値を選択します。

5. ツールバーから、[解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。

解析結果から、「A と B, B と C に点数の差がある」ことが分かります。



4-10 ダネット検定

ダネット検定は、特定のグループと他のグループを比較したい場合に用いる多重比較です。

検定の考え方や用いるデータの前提条件の詳細については、以下を参照してください。

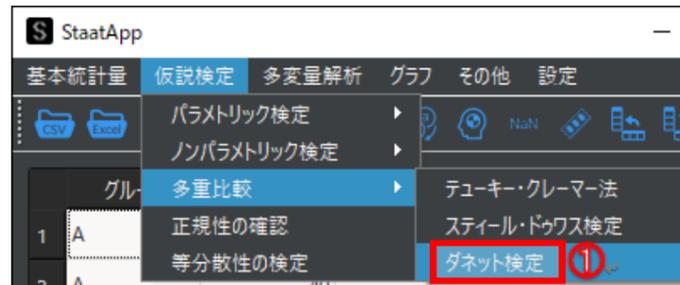
参考：[ダネット検定 | Staat](#)

以降で、ダネット検定の実行手順を説明します。手順では、例としてサンプルデータ③.csv を使用し、クラス A のテストの点数を基準として、クラス B とクラス C のテストの点数に違いがあるかどうかを解析します。

1 ダネット検定を実行する

1. メニューバーから、[仮説検定] → [多重比較] → [ダネット検定] を選択します。

「ダネット検定」が表示されます。



2. [変数選択] の [群名列] プルダウンメニューから、比較する変数 (列) を選択します。

[処理群選択] の [未選択] に、選択した変数 (列) がある値が表示されます。

ここでは“グループ”を選択します。

3. [変数選択] の [データ列] プルダウンメニューから、比較の対象とする変数 (列) を選択します。

ここでは“データ”をクリックします。

4. [対象群選択] プルダウンメニューから、比較の基準とする変数 (列) を選択します。

ここでは“A”を選択します。

5. [処理群選択] の [未選択] から、比較の対象とする値を選択します。

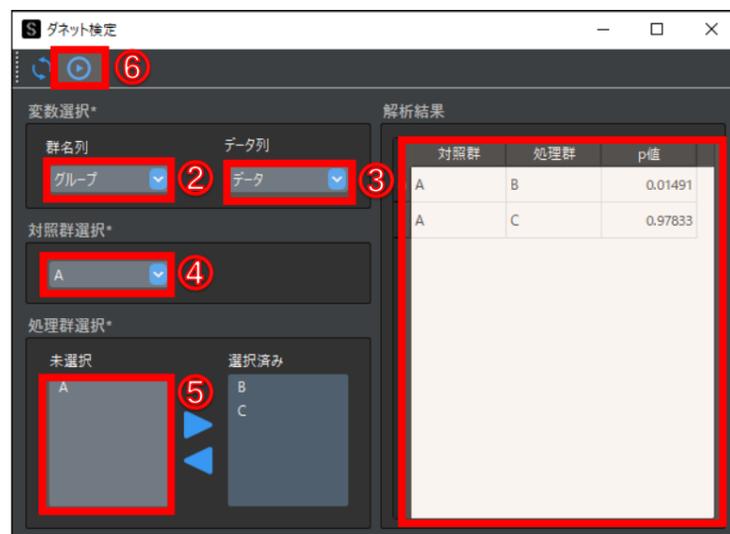
選択した値が [選択済み] に移動します。

ここでは“B”と“C”を選択します。

6. ツールバーから、[解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。

解析結果から、“A と B に点数の差がある”ことが分かります。



4-1-1 正規性の確認

正規性の確認とは、t検定や分散分析などのパラメトリック検定を行うための前提条件である、母集団が正規分布に従う（正規性がある）かどうかを調べることです。

正規性の確認の考え方や用いるデータの前提条件の詳細については、以下を参照してください。

参考：[正規性の確認 | Staat](#)

以降で、正規性の確認の実行手順を説明します。手順では、例としてサンプルデータ①.csvを使用し、収入の母集団が正規分布に従うかどうかを調べます。

1 正規性の確認を実行する

1. メニューバーから、[仮説検定] → [正規性の確認] を選択します。
「正規性の確認」が表示されます。



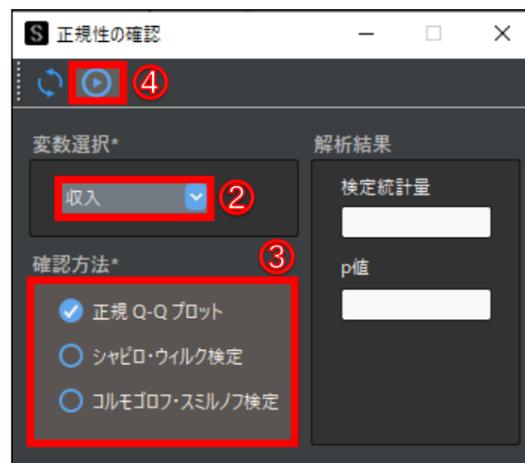
2. [変数選択] プルダウンメニューから、比較する変数（列）を選択します。

ここでは“収入”を選択します。

3. [確認方法] から確認方法を選択します。

ここでは [正規 Q-Q プロット] にチェックを入れます。

4. [解析実行] ボタンをクリックします。

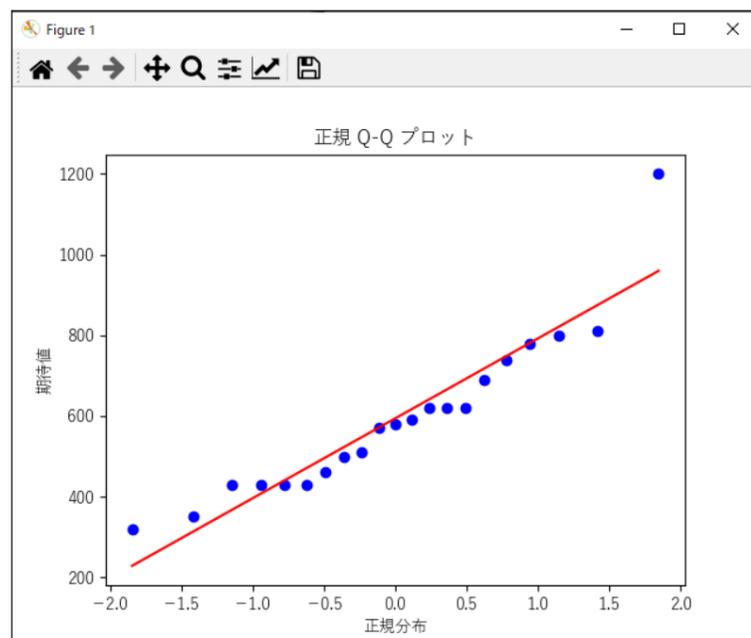


[確認方法]

[シャピロ・ウィルク検定]、[コルモゴロフ・スミルノフ検定] にチェックを入れて [解析実行] ボタンをクリックすると、[解析結果] に以下の解析結果が表示されます。

- シャピロ・ウィルク検定：検定統計量と p 値を求めます。
- コルモゴロフ・スミルノフ検定：正規分布と比較する 1 標本です。

正規 Q-Q プロットが表示されます。



4-12 等分散性の検定

等分散性の検定とは2項目（2群以上）のデータに対して、帰無仮説を各群の母集団の分散が等しいとして行う仮説検定です。主にt検定などの前提条件を確認するために行います。

以降で、等分散性の検定の実行手順を説明します。手順では、例としてダティデータ.csvを使用し、副業有無の違いによって収入の分散に差があるかどうかを解析します。

1 等分散性の検定を実行する

1. メニューバーから、[仮説検定] → [等分散性の検定] を選択します。
「等分散性の検定」が表示されます。



2. [変数選択] の [群名列] プルダウンメニューから、比較する変数（列）を選択します。

[群選択] の [未選択] に、選択した変数（列）にある値が表示されます。

ここでは“副業有無”を選択します。

3. [変数選択] の [データ列] プルダウンメニューから、比較の対象とする変数（列）を選択します。

ここでは“収入”を選択します。

4. [群選択] の [未選択] から、比較の対象とする値を選択します。

選択した値が [選択済み] に移動します。

ここではすべての値を選択します。

5. [検定方法] から [F検定] を選択します。

6. [解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。



解析結果から、“副業有無によって収入の分散に差があるとは言えない”ということが分かります。

[検定方法]

比較したい対象の数によって、検定方法が異なります。

- F検定：2群を比較します。
- バートレット検定：3群以上を比較します。
- ルビーン検定：3群以上かつ正規性の仮定が不要な比較をします。

4-13 重回帰分析

重回帰分析とは、説明変数が複数ある場合に行う回帰分析です。目的変数に対する説明変数の影響度を調べる、または得られた回帰式を用いて目的変数の予測をするために行います。

検定の考え方や用いるデータの前提条件の詳細については、以下を参照してください。

参考：[重回帰分析 | Staat](#)

以降で、重回帰分析の実行手順を説明します。手順では、例としてサンプルデータ_重回帰分析.csv を使用します。

1 重回帰分析を実行する

1. メニューバーから、[多変量解析] → [重回帰分析] を選択します。
「重回帰分析」が表示されます。



2. [目的変数] プルダウンメニューから、比較する変数（列）を選択します。
ここでは“年収”を選択します。

参考：[説明変数と目的変数 | Staat](#)

3. [説明変数] の [未選択] から、比較の対象とする変数（列）を選択します。

選択した値が [選択済み] に移動します

ここでは“年齢”、“通勤時間”、“扶養人数”、“大学”を選択します。

ダミー変数を選択する場合は、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。例では、ダミー変数はありません。

参考：[ダミー変数 | Staat](#)

参考：[多重共線性 | Staat](#)

4. [解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。

解析結果の見方は、以下の表のとおりです。



出力項目	概要	解釈例
サンプルサイズ	各変数の観測数	例では 10人のデータを用いたため、n=10
自由度調整済み決定係数	回帰式の当てはまり度	0.9365 という値は当てはまりがよい
偏回帰係数	該当する変数の影響度	年齢の偏回帰係数は 11.53298 であり、正の影響を与える
t (検定統計量)	偏回帰係数の検定統計量	大学(偏差値)は 2 以上であり、統計学的に有意な偏回帰係数である
p 値	偏回帰係数の検定の p 値	通勤時間(分)は $p < 0.05$ であるため、統計学的に有意ではない
標準偏回帰係数	変数間で影響度を比較する指標	扶養人数に対して、年齢の値が大きいため影響度が大きい
VIF	多重共線性の指標	すべての変数が 10 未満のため、問題はない

4-14 数量化Ⅱ類

数量化Ⅱ類は、カテゴリーデータ（質的変数）に対して行う多変量解析です。重回帰分析と同様に目的変数に対する説明変数の各カテゴリーの影響度や、作成した予測式に予測用データを代入することで結果を予測します。

参考：[統計学での尺度とは | Staat](#)

参考：[説明変数と目的変数 | Staat](#)

数量化Ⅱ類の計算式は、StaatApp 独自に作成しています。計算結果の妥当性の検証やアルゴリズムの参考にした文献は、以下を参照してください。

参考：[質的データの判別分析数量化Ⅱ類 | Staat](#)

検定の考え方や用いるデータの前提条件の詳細については、以下を参照してください。

参考：[数量化Ⅱ類 | Staat](#)

以降で、数量化Ⅱ類の実行手順を説明します。手順では、例としてサンプルデータ_数量化Ⅱ類.csv を使用し、転職希望度に影響を与える仕事の悩みを調べます。

サンプルデータ_数量化Ⅱ類.csv

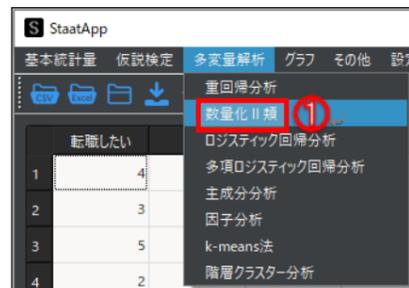
社会人20人に対してアンケートを取り、転職の希望度と仕事の悩みについて当てはまる～当てはまらないを、1～5段階で評価しています。アンケート項目（変数）は、以下のとおりです。

- y：転職したい
- x1：残業が多い
- x2：仕事にやりがいを感じない
- x3：給料が少ない
- x4：休みが少ない、または取りにくい
- x5：単純作業が多くスキルアップする見込みがない
- x6：業務量が多い
- x7：やりたい仕事をさせてもらえない
- x8：人間関係がうまくいかない

1 数量化Ⅱ類を実行する

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」が表示されます。



2. [目的変数] プルダウンメニューから、比較する変数（列）を選択します。

ここでは“転職したい”を選択します。

3. [説明変数] の [未選択] から、比較の対象とする変数（列）を選択します。

ここでは、“残業”、“やりがい”、“給料”、“人間関係”を選択します。

4. [解析項目] から、[カテゴリースコア] または [レンジ・偏相関係数] を選択します。

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を算出する場合は、[カテゴリースコア] を選択します。

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を算出する場合は、[レンジ・偏相関係数] を選択します。

ここでは [カテゴリースコア] を選択します。

5. [軸の選択] で、固有値の数（目的変数のカテゴリー数）を入力します。

[-] または [+] ボタンをクリックして入力することもできます。

一般的には1つ目の軸で結果を判断するため、“1”（デフォルト設定）のままにします。

ここでは“1”（デフォルト設定）を入力します。

6. [解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。

Quantitative II 類

目的変数*
転職したい

説明変数*
未選択: 転職したい, 休み, 自己成長, 業務量, 仕事内容
選択済み: 残業, やりがい, 給料, 人間関係

解析項目*
カテゴリースコア

軸の選択
- 1 +

解析結果
サンプルサイズ: 20
相関比: 1.0
各説明変数の統計量

	カテゴリースコア
残業_1	-0.0
残業_2	-0.0
残業_3	0.0
残業_4	0.0
残業_5	0.0
やりがい_1	-5.60112
やりがい_2	-5.60112
やりがい_3	-5.60112
やりがい_4	-5.60112

4-15 ロジスティック回帰分析

ロジスティック回帰分析は、複数の要因（説明変数）から結果（目的変数）が起きる確率を説明・予測する統計的手法です。ロジスティック回帰分析と似た手法として、重回帰分析があります。ロジスティック回帰分析と重回帰分析との違いは、目的変数です。ロジスティック回帰分析は、目的変数が“Yes/No”のような2つの値のみを取得する場合に用います。目的変数が量的変数である場合は、重回帰分析が適しています。

以降で、ロジスティック回帰分析の実行手順と、解析結果の見方を説明します。

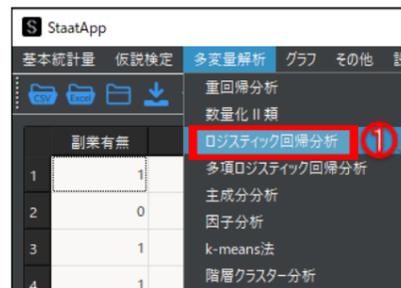
4-15-1 ロジスティック回帰分析の実行

ロジスティック回帰分析の実行手順を説明します。手順では、例としてサンプルデータ①.csvを使用します。

1 ロジスティック回帰分析を実行する

1. メニューバーから、[多変量解析] → [ロジスティック回帰分析] を選択します。

「ロジスティック回帰分析」が表示されます。



2. [目的変数] プルダウンメニューから、比較する変数（列）を選択します。

ここでは“副業有無”を選択します。

参考：[説明変数と目的変数 | Staat](#)

3. [説明変数] の [未選択] から、比較の対象とする変数（列）を選択します。

ここでは、“収入”、“性別”、“睡眠時間”を選択します。

ダミー変数を選択する場合は、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。例では、あらかじめ1列分を除いたダミー変数になっています。

参考：[ダミー変数 | Staat](#)

参考：[多重共線性 | Staat](#)

4. [解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。



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が表示されません。

4-15-2 ロジスティック回帰分析の解析結果の見方

ロジスティック回帰分析の解析結果では、以下の5つの指標を用います。

1.対数尤度

尤度（ゆうど）とは尤（もっと）もらしさを示す指標です。対数尤度は、値が大きいほど推定した回帰モデルの当てはまりが良いことを表します。

説明変数の候補となるデータが複数あり、複数のパターンで回帰モデルを推定する場合は、対数尤度が最も大きい組み合わせが最も良い回帰モデルであると言えます。手順例のように説明変数の組み合わせが1つの場合は、対数尤度の確認は不要です。

2.偏回帰係数

偏回帰係数が0の場合、対象の説明変数は推定した回帰モデルに影響を与えないことを意味します。値が正の場合は確率を上げる方向に、負の場合は確率を下げる方向に影響を与えます。

手順例の“収入”の説明変数は正の値であるため、収入が多いほど副業をしている確率が上がることを意味します。“年齢”は負の値であるため、年齢が上がるほど副業をしている確率が下がることを意味します（切片は、回帰モデルの定数項です）。

3.Wald 統計量

Wald 統計量の絶対値が大きいほど、目的変数に影響を与えていることを意味します。手順例で Wald 統計量の絶対値が最も大きいのは“収入”であるため、年収の多さが副業の有無に最も影響を与えていると言えます。

4.p 値

p 値とは、偏回帰係数の Wald 統計量についての p 値です。p 値が有意水準を下回れば、対象の説明変数は有意な偏回帰係数であると言えます。

統計学では、一般的に有意水準 $\alpha = 0.05$ とされます。有意水準 $\alpha = 0.05$ とは、実際に関連性がない場合でも関連性が存在すると結論付けるリスクが5%である、という意味です。有意水準 $\alpha = 0.05$ とすると、“収入”と“年齢”のみが統計的に有意であり、目的変数に関連性があると言えます。

5.調整オッズ比と 95%信頼区間 (95%CI)

調整オッズ比は、偏回帰係数の自然対数として求められます。調整オッズ比は、値が1未満の場合は目的変数に対して負の方向に、1以上の場合は正の方向に影響することを意味します。オッズ比の95%信頼区間は、下限値と上限値が“1”を跨がない場合、対象の説明変数は統計的に有意であると言えます。

参考：[オッズ比とは | Staat](#)

手順例の解析結果では、p 値と同じ説明変数が有意になっていることが分かります。

4-16 多項ロジスティック回帰分析

多項ロジスティック回帰分析は、複数の要因（説明変数）から結果（目的変数）が起きる確率を説明・予測する統計的手法です。多項ロジスティック回帰分析と似た手法として、重回帰分析があります。多項ロジスティック回帰分析と重回帰分析との違いは、目的変数です。多項ロジスティック回帰分析は、目的変数がカテゴリーデータの場合に用います。目的変数がカテゴリーデータであるかつ2値（0 or 1）である場合は、ロジスティック回帰分析が適しています。

実行手順と解析結果の見方は、ほぼロジスティック回帰分析と同様です。詳細手順はロジスティック回帰分析を参照してください。

説明変数	偏回帰係数	Wald統計量	p値	調整オッズ比
1 切片	-6.44929	-1.63294	0.10248	0.0015
1 収入	0.03163	2.20265	0.02762	1.0321
1 性別	1.10742	0.73869	0.4601	3.0265
1 年齢	-0.36767	-1.96003	0.04999	0.6923

ロジスティック回帰分析と多項ロジスティック回帰分析の違い

多項ロジスティック回帰分析では、目的変数の“カテゴリー数-1”のみ回帰式が作成されます。手順例では性別の“無回答”と“男性”カテゴリーに対する回帰式が作成されます。

解析結果は、対数尤度と回帰式ごとの各説明変数の統計量が出力されます。

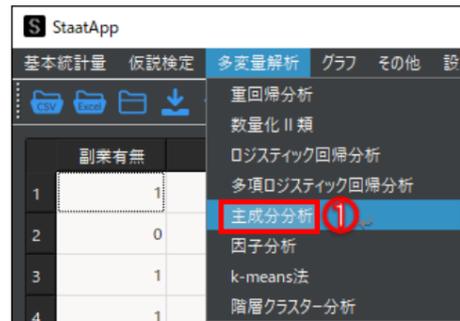
4-17 主成分分析

主成分分析とは、多数の変数に含まれている情報を要約して、少数の変数で表す手法のことです。変数が多くデータの特徴が分かりにくい場合に主成分分析を行うことで、変数（次元）を少なくしてデータの特徴を明確にできます。2つの主成分に要約した場合は、平面の散布図としてデータを確認することもできます。

以降で、主成分分析の実行手順を説明します。手順では、例としてサンプルデータ①.csvを使用します。

1 主成分分析を実行する

1. 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」が表示されます。



2. [変数設定] の [未選択] から、比較の対象とする変数（列）を選択します。

ここでは、“収入”、“年齢”、“副業有無”、“性別”、“睡眠時間”を選択します。

ダミー変数を選択する場合は、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。例では、あらかじめ1列分を除いたダミー変数になっています。

参考：[ダミー変数 | Staat](#)

参考：[多重共線性 | Staat](#)

3. [解析項目] から、[寄与率] / [主成分負荷量] / [主成分得点] を選択します。

ここでは [寄与率] を選択します。

4. [グラフ軸設定] の [x 軸] プルダウンメニューから、第一主成分を選択します。

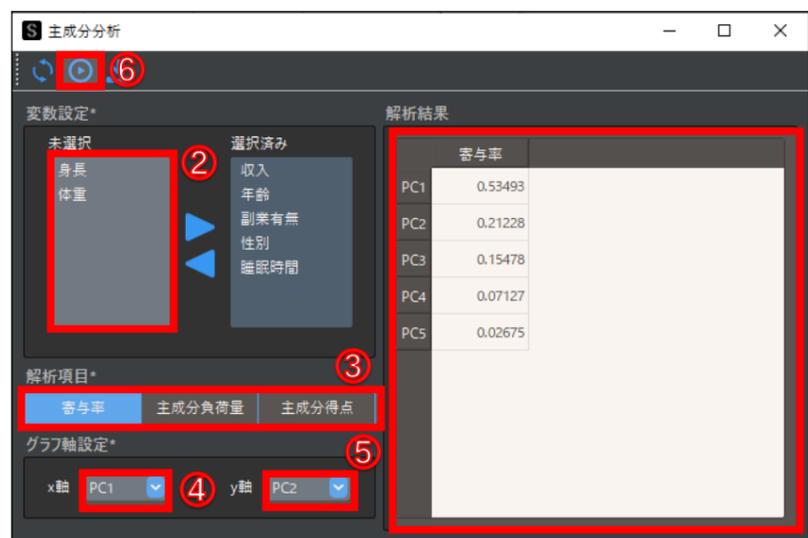
ここでは“PC1”を選択します。

5. [グラフ軸設定] の [y 軸] プルダウンメニューから、第二主成分を選択します。

ここでは“PC2”を選択します。

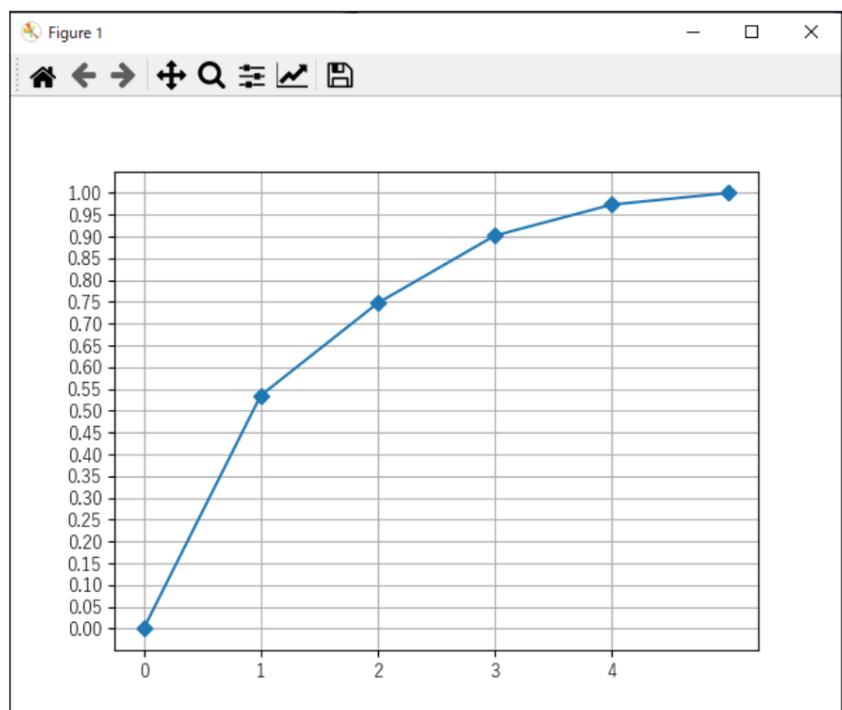
6. [解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。



主成分ごとの寄与率と累積寄与率のグラフが表示されます。

解析結果の概要については、以下の表を参考にしてください。



出力項目	説明	解釈例
主成分得点	各主成分を軸とした場合の各サンプルの点数です。	年収が高い人ほど第一主成分の正の方向にプロットされている →第一主成分は収入の多さを示す指標と解釈できる
寄与率	各主成分の説明力です。	第一主成分の寄与率 = 0.534 → 約 53% の情報を説明できている

累積寄与率	寄与率の合算値です.	第三主成分までで約 90%の情報を説明できている
主成分負荷量	各主成分に対する各変数の影響度です.	第一主成分に対して主成分負荷量が高い項目は、年収と年齢 →第一主成分は社会的地位を表す指標と解釈できる

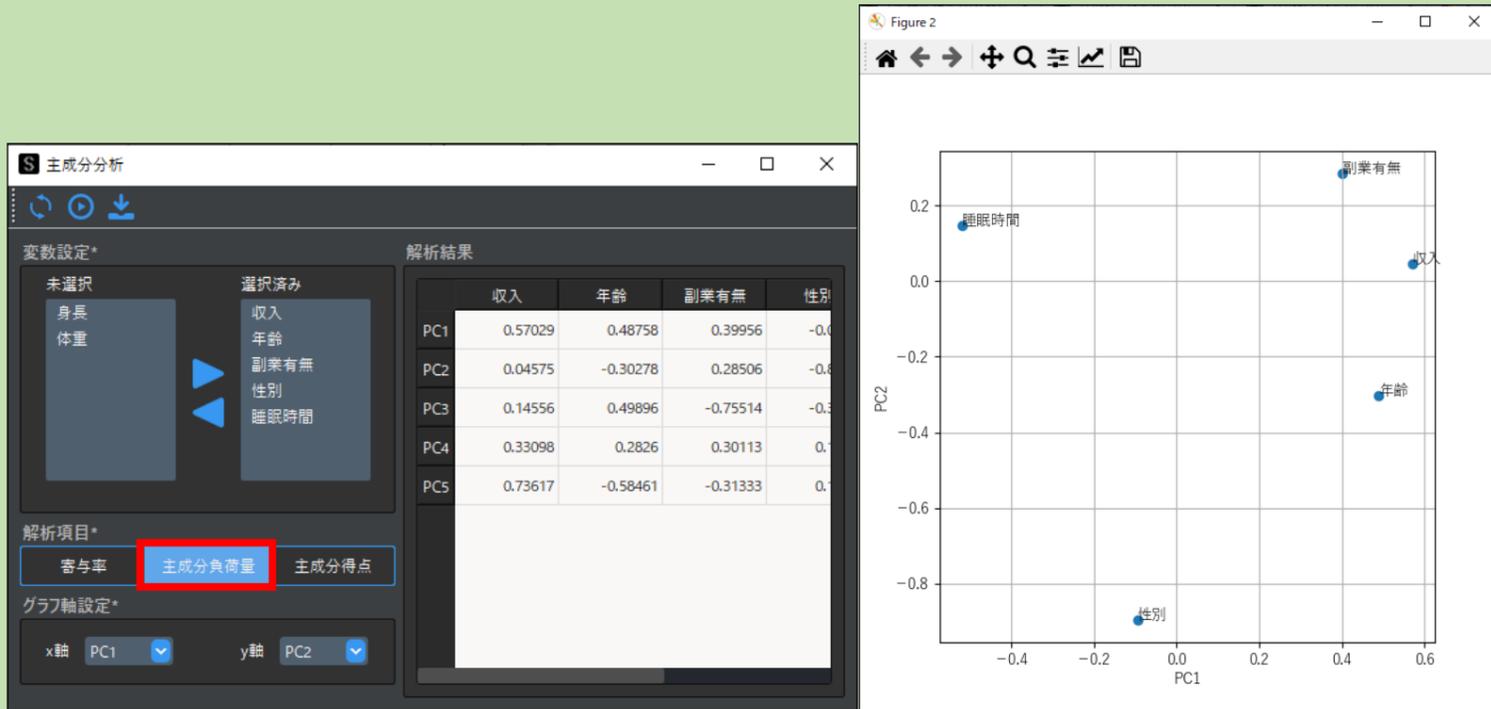
解析結果の解釈の詳細については、以下を参照してください。

参考：[主成分分析 | Staat](#)

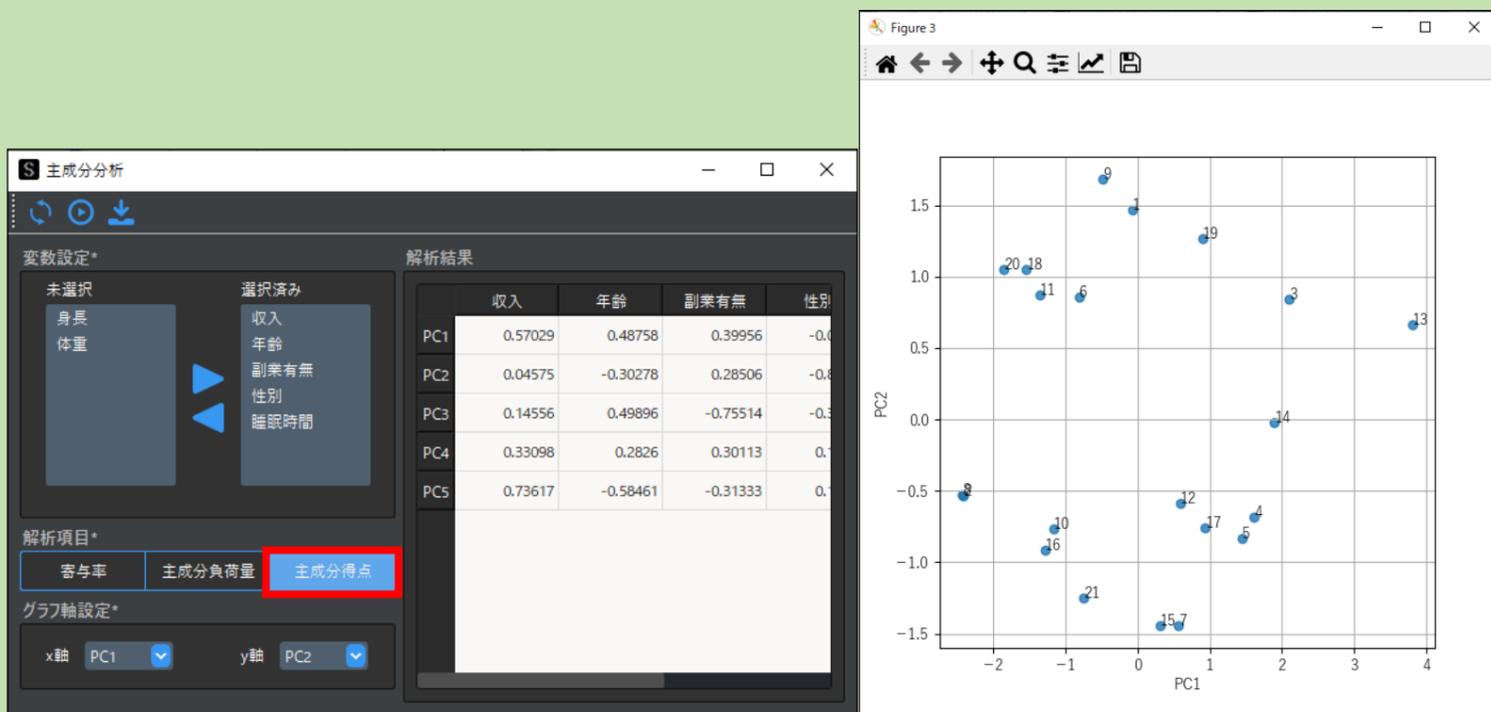
主成分負荷量/主成分得点の解析結果

[解析項目] で [主成分負荷量] または [主成分得点] を選択した場合の解析結果について説明します。

[主成分負荷量] を選択した場合は、各変数の主成分負荷量とプロットが表示されます。



[主成分得点] を選択した場合は、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。

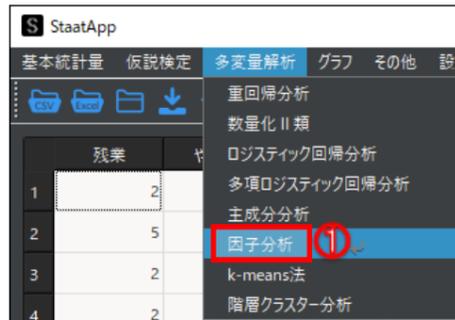


4-18 因子分析

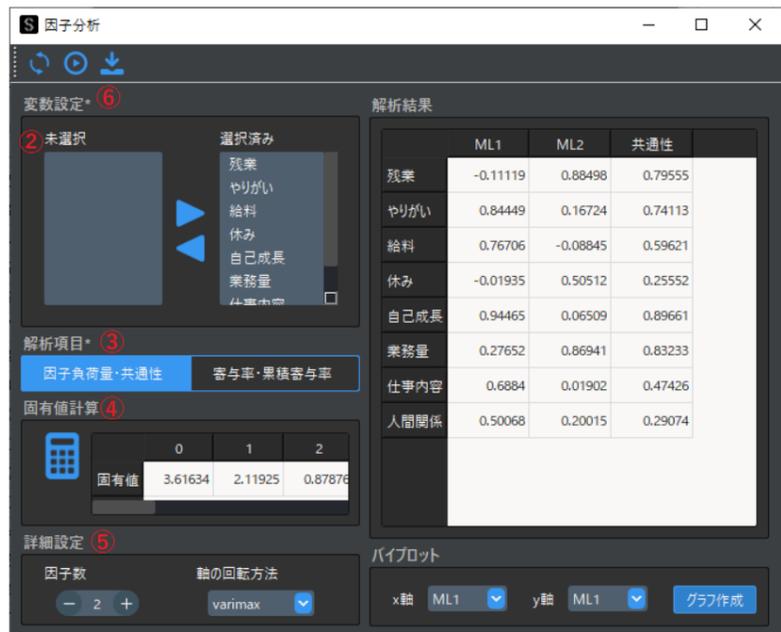
因子分析とは、変数に共通して影響を与えている概念（共通因子）を抽出し、変数間の関係性を調べる手法です。以降で、因子分析の実行手順を説明します。手順では、例としてサンプルデータ_因子分析.csvを使用します。

1 因子分析を実行する

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」が表示されます。



2. [変数設定] の [未選択] から、分析とする変数（列）を選択します。
ここではすべての値を選択します。
3. [解析項目] から、[因子負荷量・共通性] または [寄与率・累積寄与率] を選択します。
ここでは [因子負荷量・共通性] を選択します。
4. 選択した変数分の固有値を求める場合は、[固有値計算] の電卓アイコンをクリックします（固有値は因子数を決定する際に用います）。
5. [詳細設定] で [因子数]、[軸の回転方法]、[因子抽出方法] をそれぞれ指定します。



デフォルトでは以下に設定されています（一般的な値を設定しています）。

- 因子数：2
 - 軸の回転方法：varimax（バリマックス回転）
 - 因子抽出方法：最尤法
6. [解析実行] ボタンをクリックします。
[解析結果] に解析結果が表示されます。

[詳細設定] の各種項目

[詳細設定] の各種項目の詳細については、以下のとおりです。

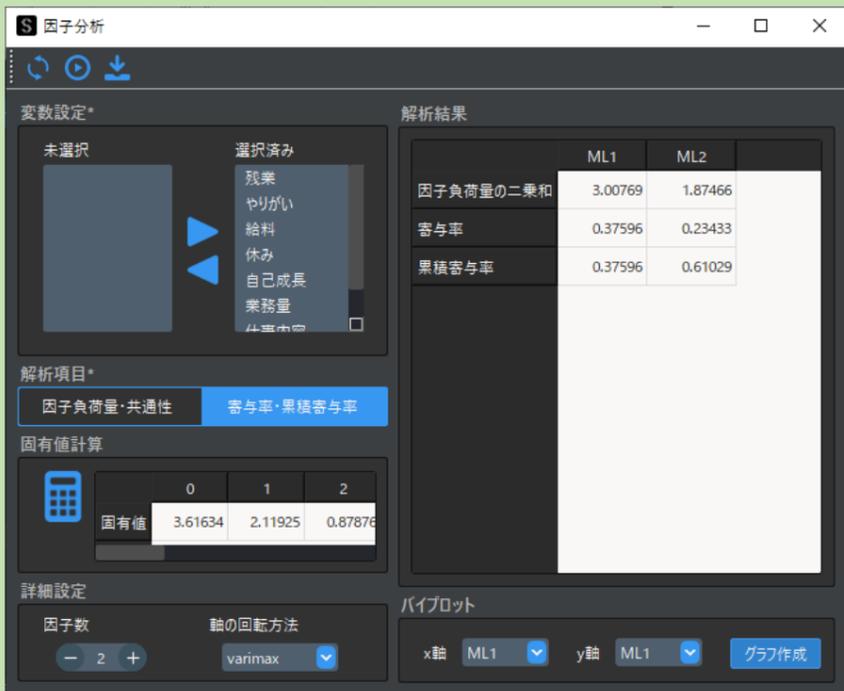
- 因子数：因子数は、固有値を計算した上で決定します。一般的には、固有値が2以上の変数の数を因子数とします。手順例では、固有値が2以上の変数が2つのため、因子数を2とします
- 軸の回転方法：軸の回転は、変数に対して因子の当てはまりが良くなるように原点を中心に座標軸を回転させることです。回転方法には主に直交回転と斜交回転があります。軸の回転方法やその他の回転方法についての詳細は、以下を参照してください。
参考：[因子分析とは | Staat](#)
- 因子抽出方法：因子抽出方法は因子モデルの計算方法です。代表的な方法は、以下の2つです。
 - ・最尤法：多変量正規分布を用いた最尤推定を行います。StaatAppでは最尤法で因子分析を行います。
 - ・最小残差法：残差を最小にするように計算します。

解析結果の概要については、以下の表を参考にしてください。

出力項目	概要	解釈例
因子負荷量	各因子と各変数の関連性の大きさ	第一因子では自己成長の因子負荷量が大きい →第一因子は自己成長と関連性が深い因子
共通性	各変数の因子モデルへの影響度	休みの共通性が小さい →因子分析の対象から外しても問題はなさそうである
寄与率	各因子の説明力	第一因子の寄与率=0.402 →約40%の情報を説明できている
累積寄与率	寄与率の合算値	第二因子までで約65%の情報を説明できている →60%を超えているため良い因子モデルであると考えられる

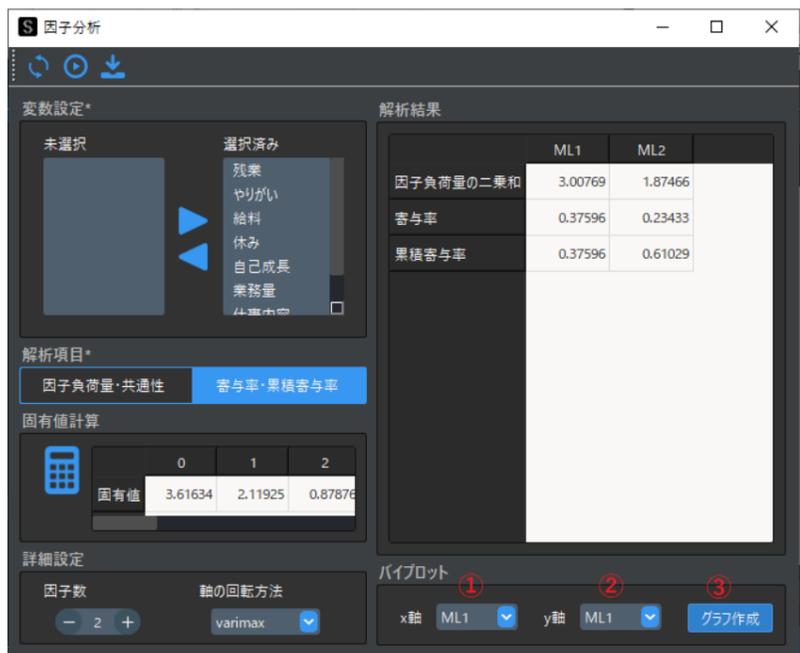
寄与率・累積寄与率の解析結果

[解析項目] で [寄与率・累積寄与率] を選択した場合の解析結果は、以下のように表示されます。



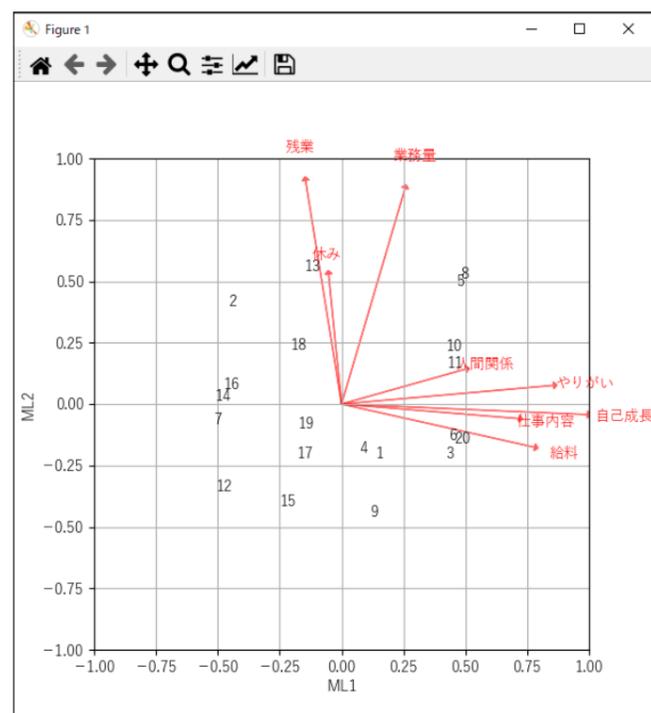
2 グラフを作成する

1. [パイプロット] の [x 軸] プルダウンメニューから、第一因子を選択します。
ここで“ML1”を選択します。
2. [パイプロット] の [y 軸] プルダウンメニューから、第二因子を選択します。
ここでは“ML2”を選択します。
3. [グラフ作成] ボタンをクリックします。



散布図が表示されます。

[変数設定] や [詳細設定] を変更して異なるパイプロットを作成したい場合は、各設定を変更後、[解析実行] ボタンをクリックしてから [グラフ作成] ボタンをクリックします。



パイプロット

パイプロットは、各因子の因子負荷量のベクトル表記と、因子得点をプロットした図です。パイプロットの見方やパイプロットを用いた軸のネーミングの詳細については、以下を参照してください。

参考：[因子分析とは | Staat](#)

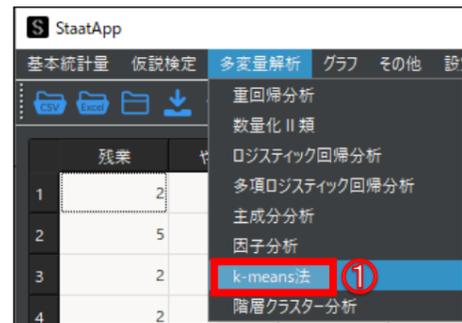
4-19 k-means 法

k-means 法とは、データを設定したクラスターごとに大まかに分類し、各データとクラスターの重心の距離が別のクラスターの重心より小さくなるようにデータを分類する方法です。

以降で、k-means 法の実行手順を説明します。手順では、例としてサンプルデータ①.csv を使用します。

1 K-means 法を実行する

1. メニューバーから、[多変量解析] → [k-means 法] を選択します。
「k-means 法」が表示されます。



2. [変数設定] の [未選択] から、比較の対象とする変数 (列) を選択します。
選択した変数が [選択済み] に移動します。

ここでは、「収入」、「年齢」、「副業有無」、「性別」、「身長」を選択します。

ダミー変数を選択する場合は、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。

例では、あらかじめ 1 列分を除いたダミー変数になっています。

参考：[ダミー変数 | Staat](#)

参考：[多重共線性 | Staat](#)

3. [クラスター数] で、クラスタリングを行う集団の数 (クラスター数) を入力します。
[-] または [+] ボタンをクリックして入力することもできます。
ここでは「4」を入力します。

4. [解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。

選択した変数データの先頭列に、「class」という行が追加されて出力されます。

「class」列にある値が、クラスタリングした結果の属する集団を示す番号です。同じ「1」を持つサンプル (行) は、同じクラスターであることが分かります。

解析結果の詳細については、以下を参照してください。

参考：[K-means 法 \(非階層クラスター分析\) | Staat](#)



4-20 階層クラスター分析

階層クラスター分析とは、データ群の中で最も近いデータでグループ分けすることを繰り返し、クラスターの数を減らしていく手法です。以降で、階層クラスター分析の実行手順を説明します。手順では、例としてサンプルデータ①.csvを使用します。

階層クラスター分析と非階層クラスター分析の違い

階層クラスター分析は計算量が膨大になるため、ビッグデータの分析には向いていません。階層クラスター分析を行うサンプル数の目安は、100以下です。

また、非階層クラスター分析では、樹形図を作成しません。クラスター分析を行う前に、グループ数を決定する必要があります。

1 階層クラスター分析を実行する

1. メニューバーから、[多変量解析] → [階層クラスター分析] を選択します。

「階層クラスター分析」が表示されます。



2. [変数選択] の [未選択] から、比較の対象とする変数 (列) を選択します。

選択した変数が [選択済み] に移動します。

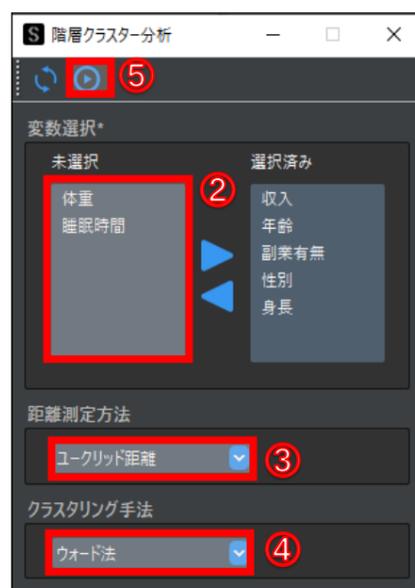
ここでは、“収入”、“年齢”、“副業有無”、“性別”、“身長”を選択します。

3. [距離測定方法] プルダウンメニューから、距離測定方法を選択します。ここではデフォルトのまま [ユークリッド距離] を選択します。

4. [クラスタリング手法] プルダウンメニューから、クラスタリング手法を選択します。

ここではデフォルトのまま [ワード法] を選択します。

5. [解析実行] ボタンをクリックします。

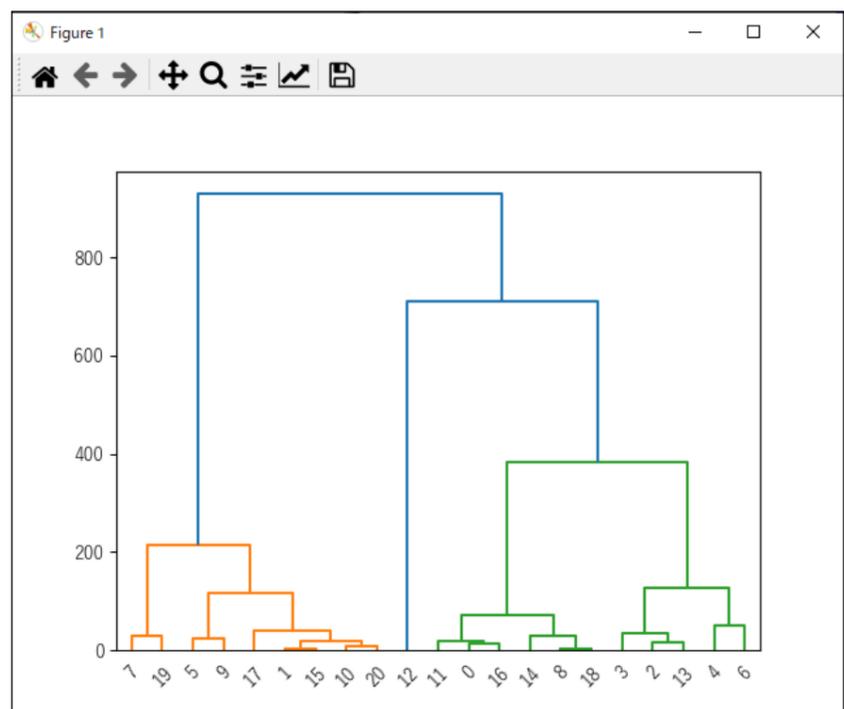


デンドログラム (樹形図) が表示されます。

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ることが分かります。

解析結果の詳細については、以下を参照してください。

参考：[階層クラスター分析 | Staat](#)



[距離測定方法] と [クラスタリング手法]

StaatApp で設定できる [距離測定方法] と [クラスタリング手法] には、以下の方法があります。

距離測定法	クラスタリング手法
ユークリッド距離 (デフォルト)	[ワード法] / [最短距離法] / [最長距離法] / [重心法] / [群平均法] / [メディアン法]
標準化ユークリッド距離	[最短距離法] / [最長距離法] / [群平均法]
マハラノビスの距離	
市街地距離	
最大座標差	
ミンコフスキー距離	

4-21 ヒストグラム

ヒストグラムとは、量的データの分布を表すグラフです。データは棒（ビン）で表され、棒の面積は階級ごとの度数を示します。棒グラフと異なる点は、ヒストグラムは棒の横幅にも意味があることです。棒グラフは棒の高さのみでデータを表しますが、ヒストグラムは棒の面積でデータの階級や度数を表します。以降で、ヒストグラムの作成手順を説明します。手順では、例としてサンプルデータ_グラフ.xlsx を使用し、30歳の給料と40歳の給料を比較します。

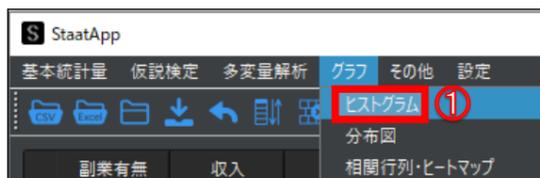
各種グラフの特徴や選び方

StaatApp で作成できるグラフの特徴や選び方の詳細については、以下を参照してください。

参考：[グラフによる分布の可視化方法 | Staat](#)

1 ヒストグラムを作成する

1. メニューバーから、[グラフ] → [ヒストグラム] を選択します。
「ヒストグラムの作成」が表示されます。



2. [変数選択] の [グラフ A] プルダウンメニューから、比較の対象とする変数（列）を選択します。

ここでは“30歳_給料”を選択します。

3. [変数選択] の [グラフ B] プルダウンメニューから、比較の対象とする変数（列）を選択します。

ここでは“40歳_給料年齢”を選択します。

4. ビン数の目安を求める場合は、[ビン数の目安] で電卓アイコンをクリックします。★画像：差し替え

[スタージェスの公式]、[平方根選択]、[スコットの選択(幅)]、[フリードマン・ダイアコニスの選択(幅)] にそれぞれ値が入力されます。

各公式の詳細については、以下を参照してください。

参考：[ヒストグラムの見方・かき方 | Staat](#)

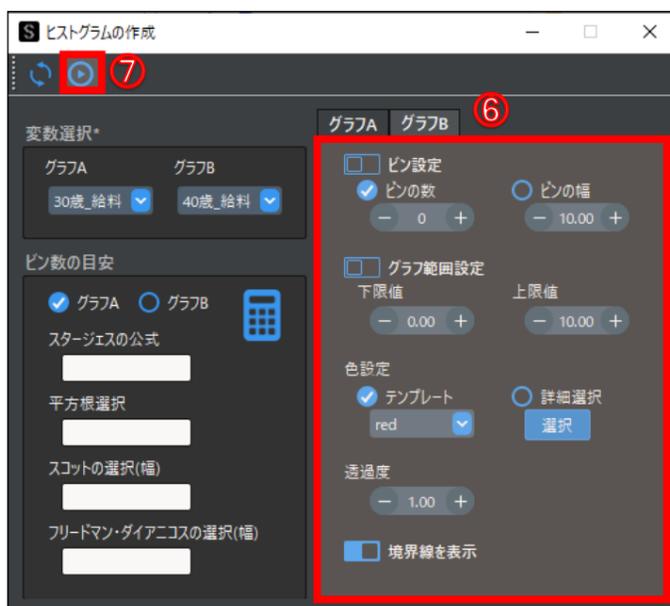
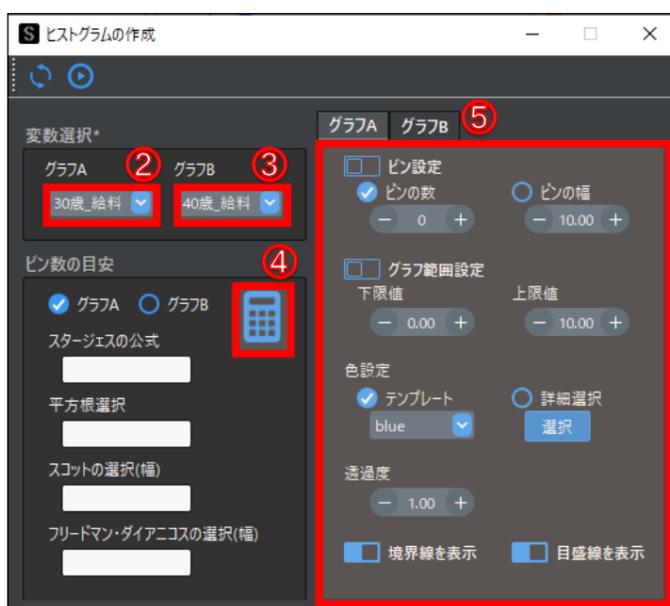
5. [グラフ A] タブで、以下のように設定します。

- [色設定] : blue
- [境界線を表示] : オンにする
- [目盛線を表示] : オンにする

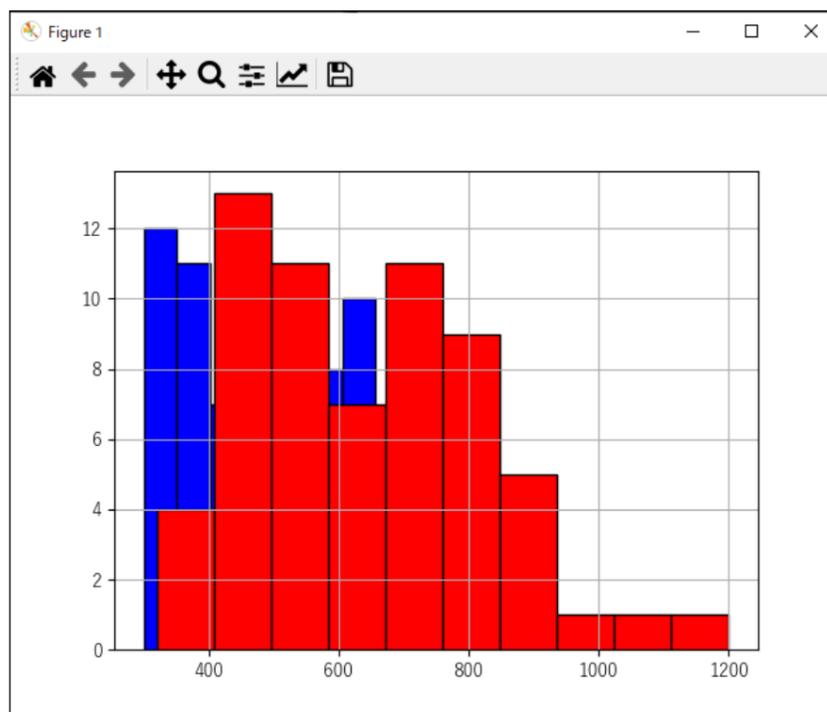
6. [グラフ B] タブで、以下のように設定します。

- [色設定] : red
- [境界線を表示] : オンにする

7. [解析実行] ボタンをクリックします。



2つのヒストグラムが重ねて表示されます.



4-22 分布図

分布図とは、データの分布を調べる/表すための図のことで、データの傾向を視覚的に捉えることができます。

以降で、分布図の作成手順と、各種グラフの詳細設定について説明します。

StaatApp で作成できるグラフの特徴や選び方の詳細については、以下を参照してください。

参考：[グラフによる分布の可視化方法 | Staat](#)

4-22-1 分布図を作成する

分布図の作成手順を説明します。手順では、例としてサンプルデータ_グラフ.xlsx を使用し、企業 A、企業 B、企業 C に勤める 30 歳の給料を比較します。

1 分布図を作成する

1. メニューバーから、[グラフ] → [分布図] を選択します。

「分布図の作成」が表示されます。



2. [グラフ選択] から、作成するグラフの種類を選択します。

ここでは [箱ひげ] のままにします。

複数のグラフを選択すると、グラフを重ねて作成できます。

3. [軸選択] の [横軸(項目)] プルダウンメニューから、横軸に表示する項目を選択します。

[横軸項目選択] の [未選択] に、選択した項目の値が表示されます。

ここでは「企業名」を選択します。

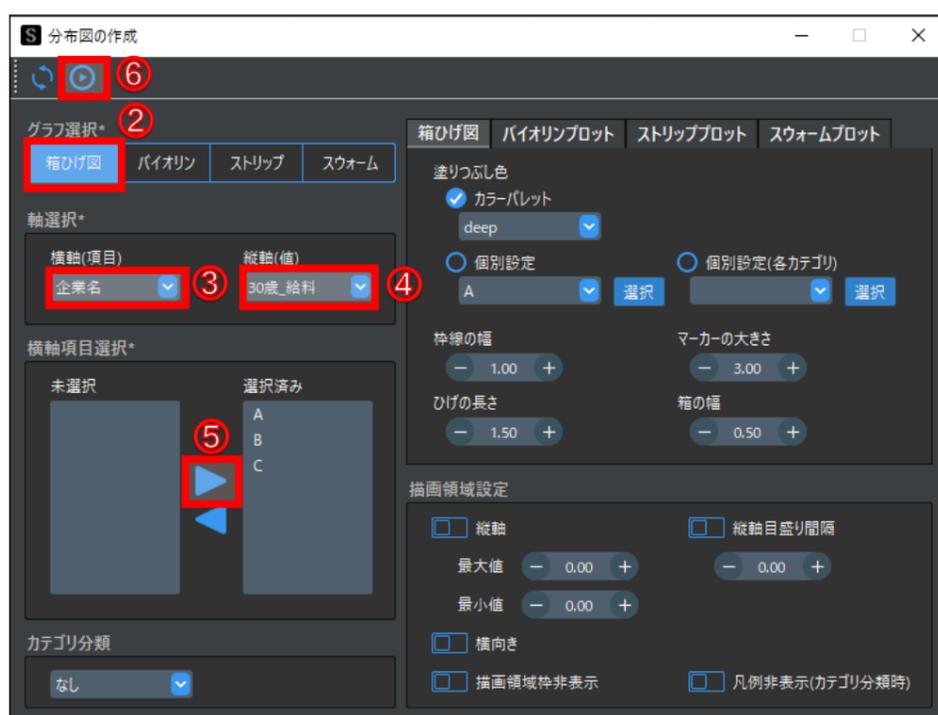
4. [軸選択] の [縦軸(値)] プルダウンメニューから、縦軸に表示する項目を選択します。

ここでは「30 歳_給料」を選択します。

5. [横軸項目選択] の [未選択] から、横軸に表示する値を選択します。

ここではすべての値を選択します。

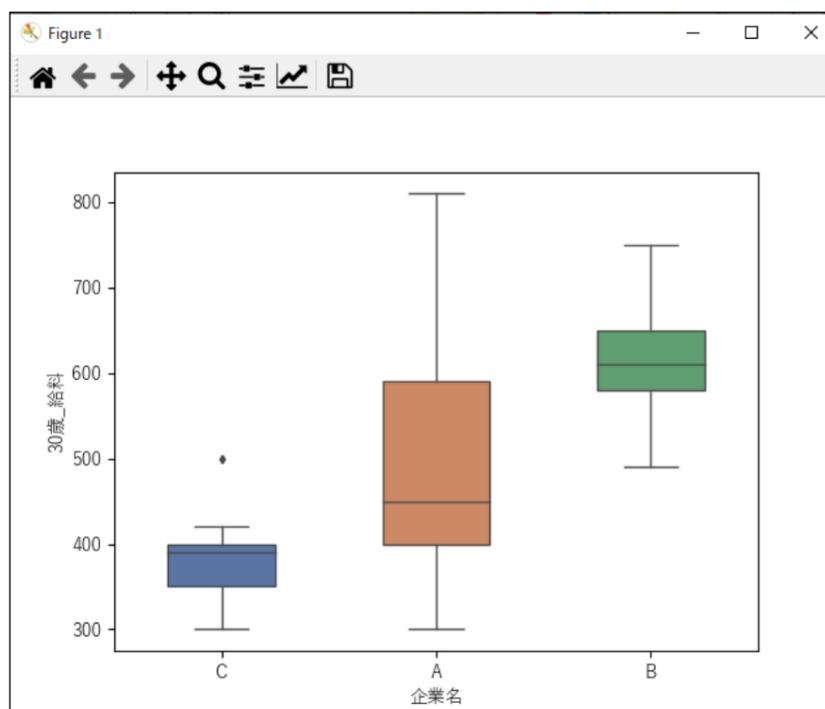
6. [解析実行] ボタンをクリックします。



箱ひげ図が表示されます。

横軸項目の並び順は、選択した値の出現順です。並び順を変えたい場合は、データ操作機能で並び替えてください。

参考：[3 データ操作 \(P.15\)](#)



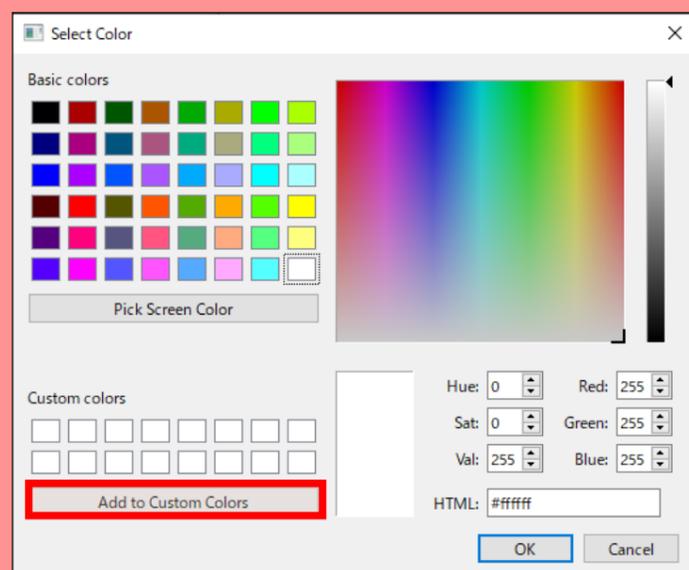
描画領域設定

[描画領域設定] では、数値軸（縦軸）の設定やグラフの向きの設定ができます。

項目名	説明
縦軸	縦軸の最大値と最低値を指定します。
縦軸目盛り間隔	縦軸の最大値と最低値を指定した場合、目盛り間隔を指定します。
横向き	グラフの向きを右に 90° 回転します。横軸と縦軸が自動で入れ替わります。
描画領域枠非表示	描画領域の上側と右側の枠線を非表示にします。
凡例非表示(カテゴリ分類時)	カテゴリ分類時に自動作成される凡例を削除します。

カラー設定の注意点

各種グラフのタブで [選択] ボタンをクリックすると「Select Color」が表示され、グラフのカラー設定ができます。ただし、[Add to Custom Colors] をクリックするとグラフが作成されなくなるため、クリックしないでください。



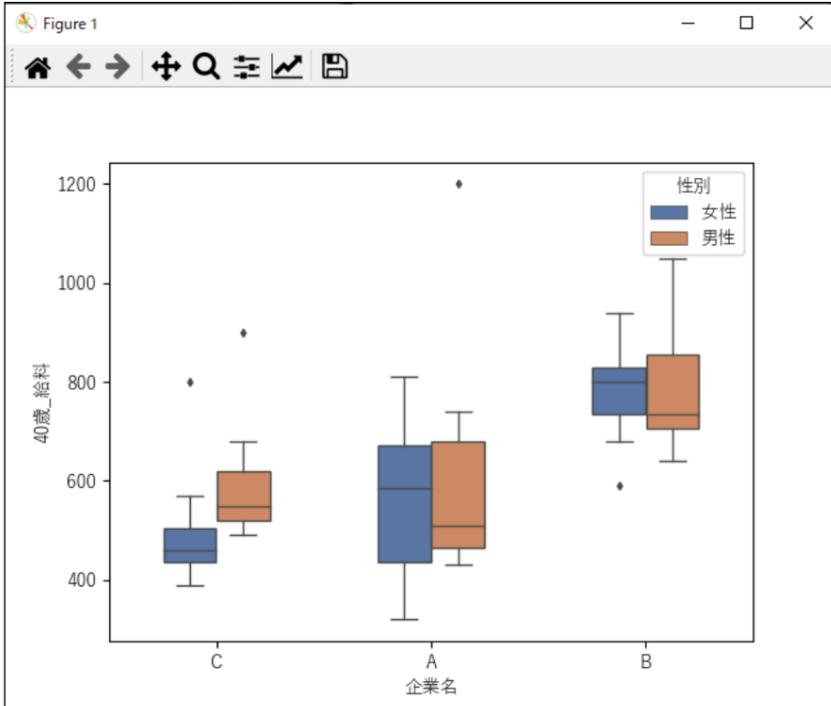
4-2 2-2 箱ひげ図の設定項目を確認する

箱ひげ図の各設定値は、以下のとおりです。

- 枠線の幅：枠線・ひげの線の太さ
- マーカーの大きさ：外れ値を示すマーカーの大きさ
- ひげの長さ：ひげの長さの上限値（外れ値としない範囲）。箱の長さ（IQR）×設定値＝ひげの長さ箱の幅：箱部分の幅

箱ひげ図の詳細については、以下を参照してください。

参考：[箱ひげ図 | Staat](#)



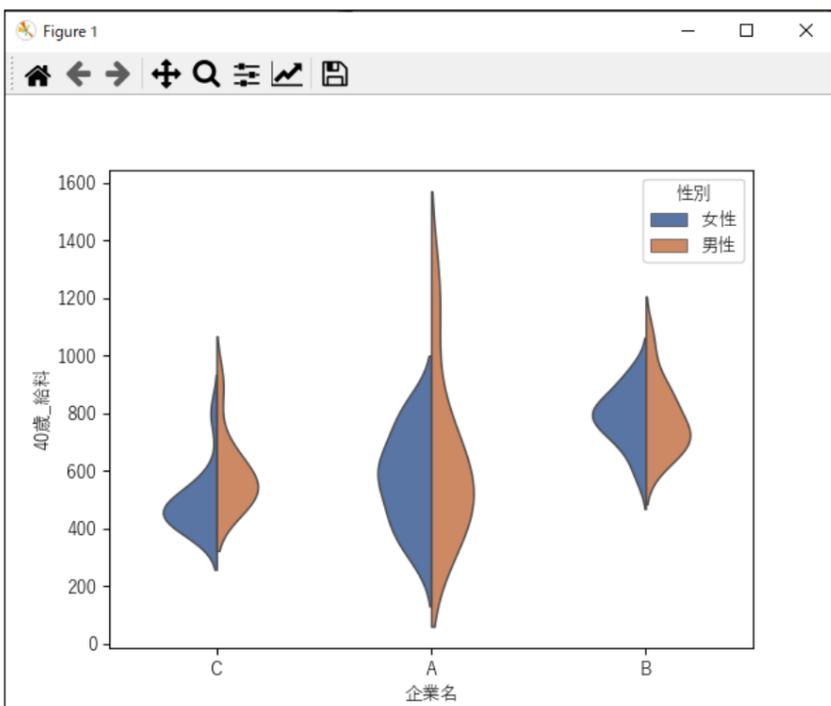
4-2 2-3 バイオリンプロットの設定項目を確認する

バイオリンプロットの各設定値は、以下のとおりです。

- 枠線の幅：枠線の太さ
- 横幅：膨らみ部分の幅

バイオリンプロットの詳細については、以下を参照してください。

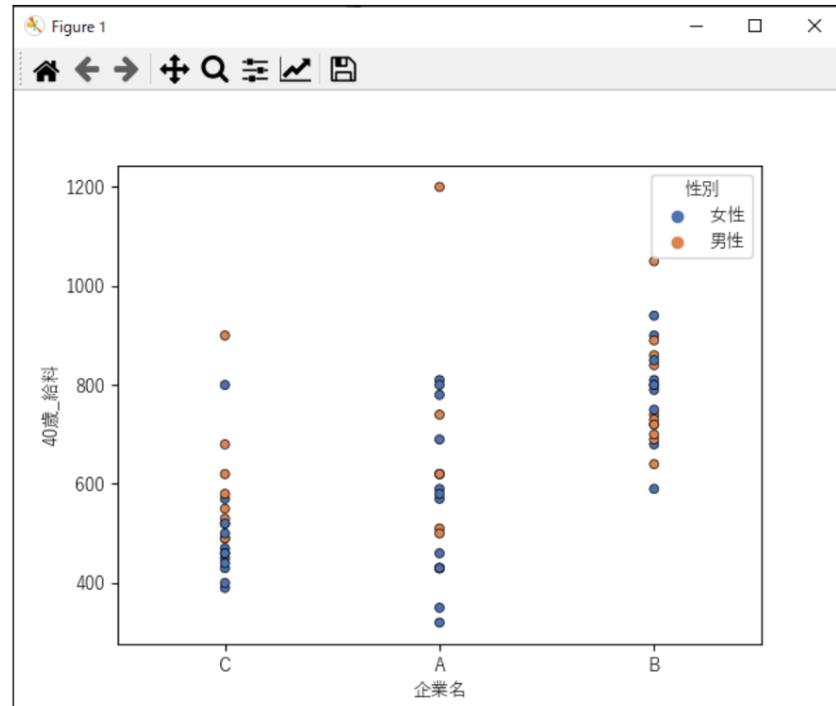
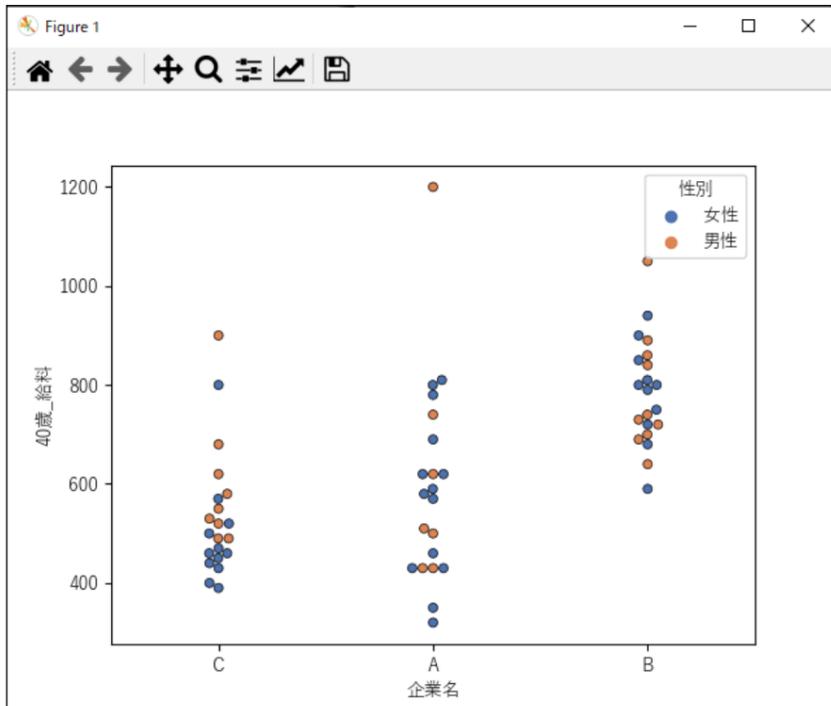
参考：[バイオリンプロット | Staat](#)



4-2 2-4 ストリッププロット・スウォームプロットの設定項目を確認する

ストリッププロット及びスウォームプロットバイオリンプロットの各設定値は、以下のとおりです。

- 枠線の幅：マーカーの枠線の太さ。枠線を消す場合は“0”を指定する
- 枠線の色：マーカーの枠線の色
- マーカーの大きさ：マーカーの大きさ
- 横幅（ストリッププロットのみ）：マーカーが重なる場合の横の広がり具合。直線状に分布させる場合は“0”を指定する



4-23 散布図行列・ヒートマップ

散布図とは、2つの量的変数の分布を表すグラフです。横軸と縦軸にそれぞれ変数を設定して作成します。散布図では、外れ値の検出や2つの変数の相関関係を調べられます。

ヒートマップとは、数値の強弱で色分けをし、視覚的にデータを確認できるグラフです。

以降で、散布図行列とヒートマップの作成手順を説明します。手順では、例としてサンプルデータ①.csvを使用します。

なお、StaatAppで作成できるグラフの特徴や選び方の詳細については、以下を参照してください。

参考：[グラフによる分布の可視化方法 | Staat](#)

4-23-1 散布図行列を作成する

散布図行列は多変量分析の基本的な手法で、扱うデータごとの相関関係を調べられます。各変数の相関関係は、重回帰分析等の多変量解析を行うための重要な情報となります。

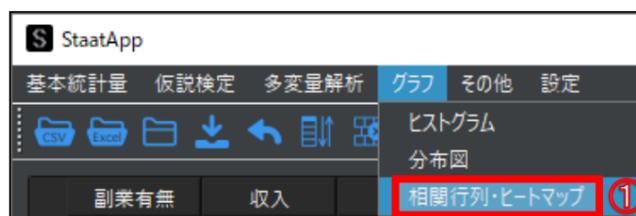
論文等でも、多変量解析に用いた値の特徴を示すために用いられることがあります。

以降で、散布図行列の作成手順を説明します。

1 散布図行列を作成する

1. メニューバーから、[グラフ] → [散布図行列・ヒートマップ] を選択します。

「散布図行列・ヒートマップの作成」が表示されます。



2. [変数選択] の [未選択] から、散布図行列に用いる項目（変数）を選択します。

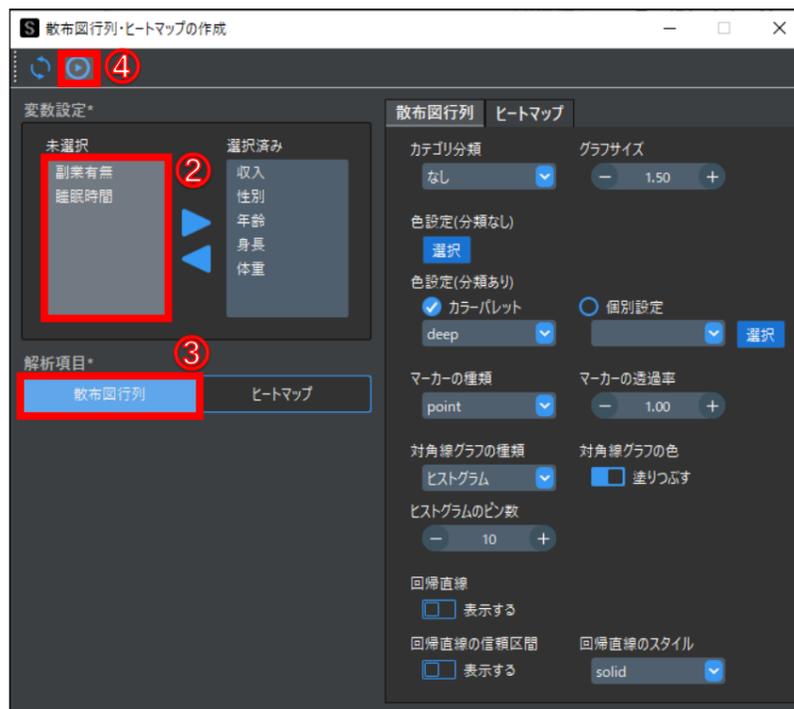
選択した変数が [選択済み] に移動します。

ここでは、“収入”、“性別”、“年齢”、“身長”、“体重”を選択します。

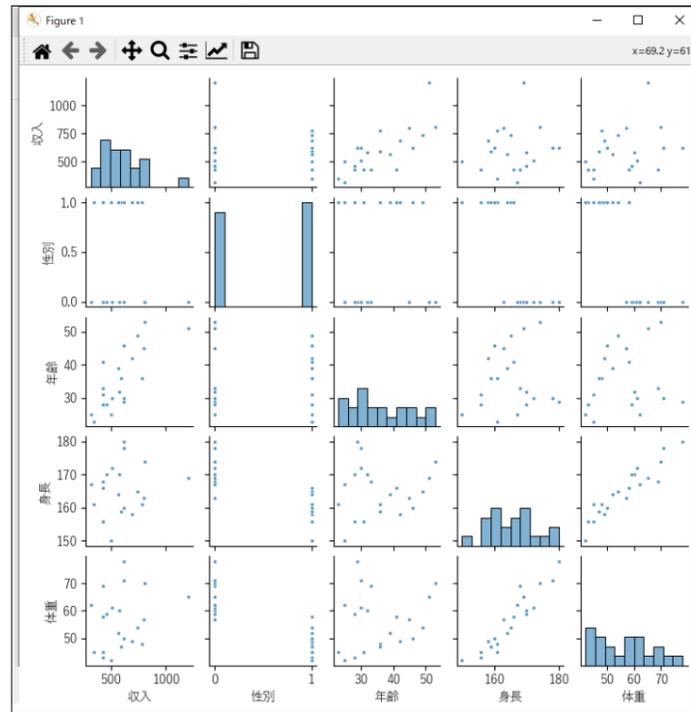
3. [解析項目] で解析項目を選択します。

ここではデフォルトのまま [散布図行列] を選択します。

4. [解析実行] ボタンをクリックします。

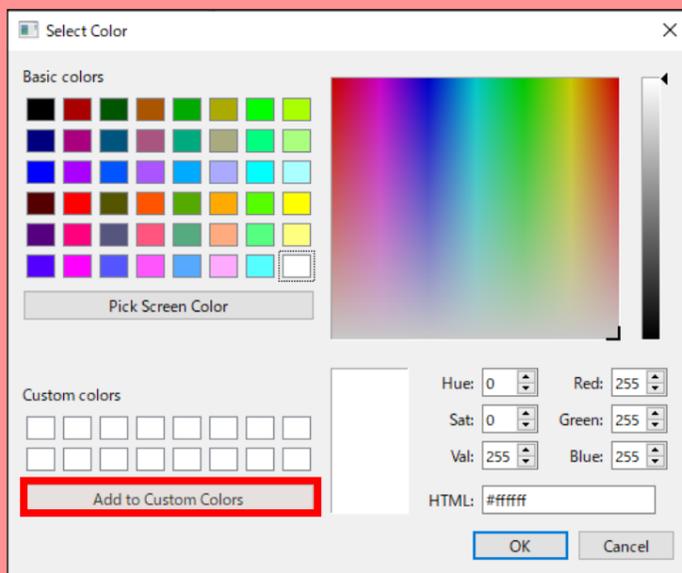


散布図行列が表示されます。



カラー設定の注意点

各種グラフのタブで [選択] ボタンをクリックすると「Select Color」が表示され、グラフのカラー設定ができます。ただし、[Add to Custom Colors] をクリックするとグラフが作成されなくなるため、クリックしないでください。



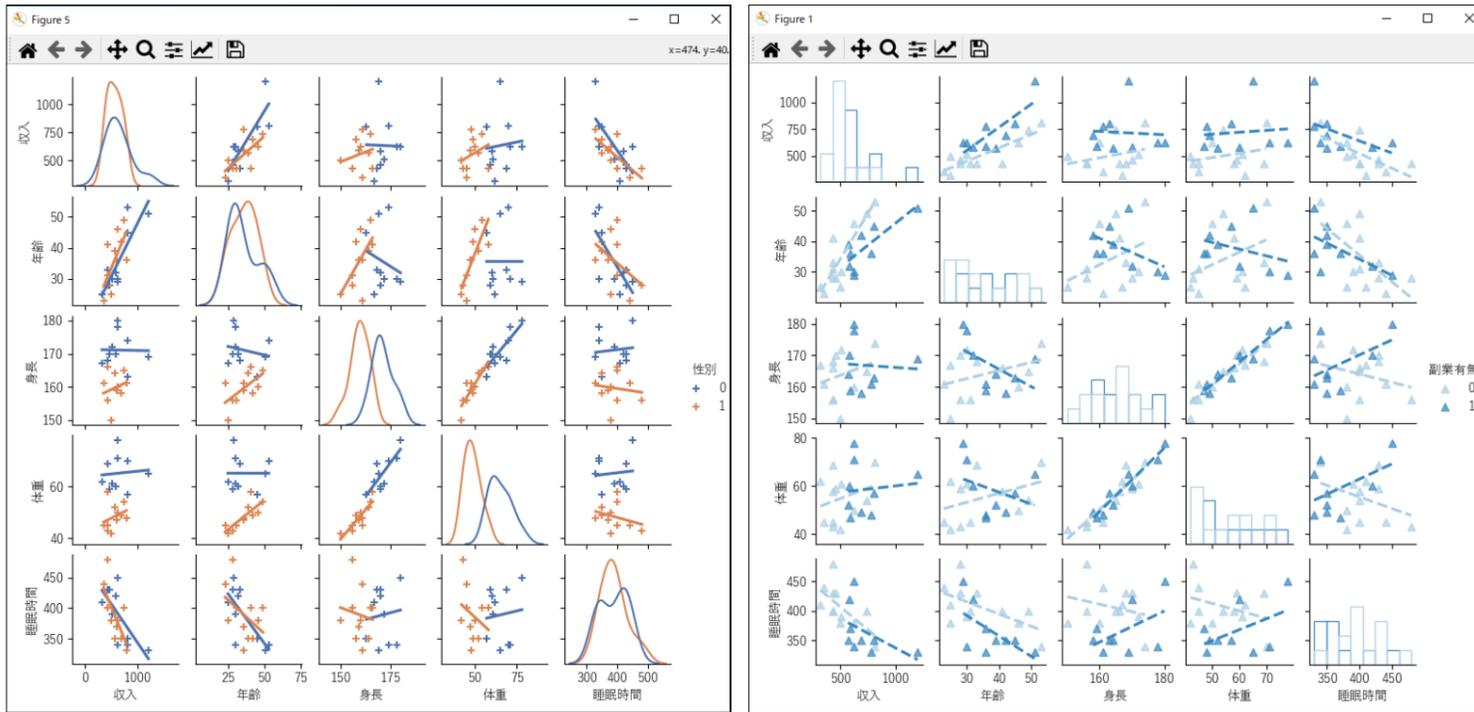
4-23-2 散布図行列の詳細設定とグラフ例

[散布図行列] タブで設定できる詳細設定の項目については、以下の表を参照してください。

設定項目	説明
カテゴリ分類	選択したカテゴリ列の値で、プロットを色分けします。
グラフサイズ	出力するグラフの大きさを指定します。
色設定(分類なし)	カテゴリ分類を行わない場合、プロットの色を指定します。
色設定(分類あり)	カテゴリ分類を行う場合、色を指定します。カラーパレットを用いる場合は [カラーパレット] にチェックを入れて色を指定します。各カテゴリに任意の色を選択する場合は、 [個別設定] にチェックを入れて色を指定します。
マーカーの種類	マーカーの種類を指定します。
マーカーの透過率	マーカーの透過率を指定します。
対角線グラフの種類	対角線上に配置するグラフの種類 (ヒストグラム/カーネル密度) を指定します。
対角線グラフの色	対角線上に配置するグラフを塗りつぶす場合、オンにします。
ヒストグラムのビン数	対角線上にヒストグラムを表示する場合、ビン数を指定します。

回帰直線	散布図上に回帰直線を表示する場合、オンにします。
回帰直線の信頼区間	回帰直線の95%信頼区間を表示する場合、オンにします。
回帰直線のスタイル	回帰直線のスタイルを指定します。

散布図行列の作成例は、以下のようになります。



4-23-3 ヒートマップを作成する

ヒートマップの作成手順は、詳細設定以外は散布図行列と同様です。項目の選択とグラフの選択方法については、以下を参照してください。

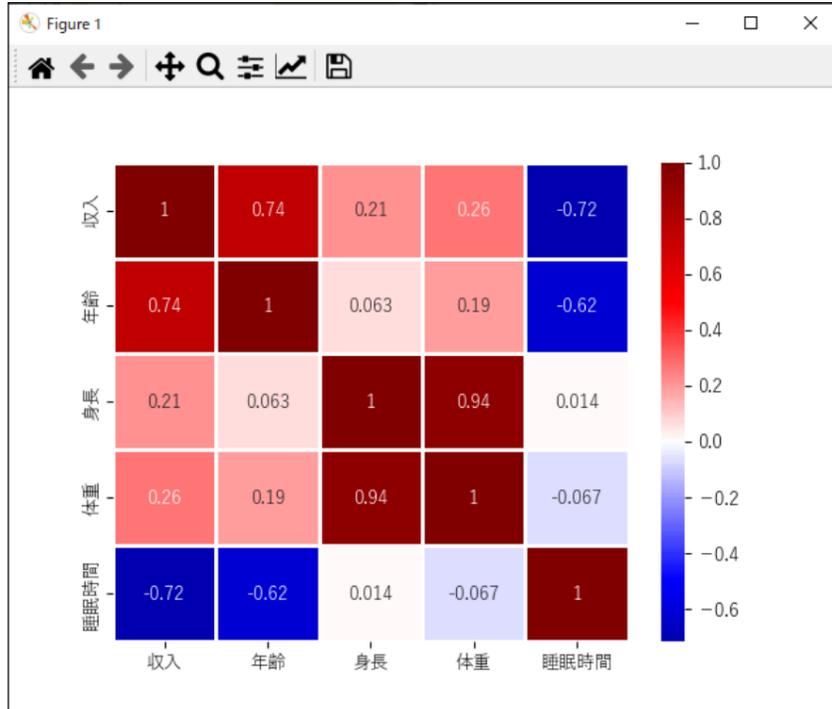
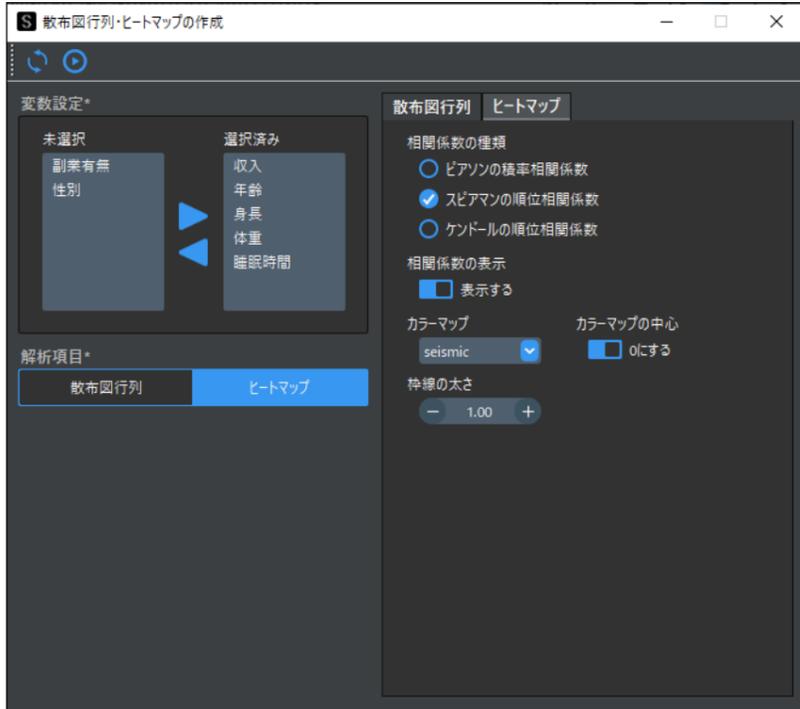
参考：4-23-1 散布図行列を作成する (P.60)

4-23-4 ヒートマップの詳細設定とグラフ例

[ヒートマップ] タブで設定できる詳細設定の項目については、以下の表を参照してください。

設定項目	説明
相関係数の種類	算出する相関係数の種類を選択します。相関係数の種類の詳細については、以下を参照してください。 参考： 相関係数 Staat
相関係数の表示	ヒートマップ上に相関係数を表示する場合、オンにします。
カラーマップ	使用するカラーマップを指定します。
カラーマップの中心	カラーマップの中心を0にする場合、オンにします。
枠線の太さ	枠線の太さを指定します。デフォルトでは“0”が指定されています。“0”の場合、枠線は表示されません。

相関係数の絶対値が大きいほど、色が濃くなるヒートマップを作成する場合の設定例とヒートマップは、以下のようになります。



4-24 クロス集計表

クロス集計表では、以下の4つの分析が可能です。

- カイ二乗検定
- フィッシャーの正確確率検定
- オッズ比とリスク比
- コレスポネンス分析

クロス集計表の詳細については、以下を参照してください。

参考：[クロス集計表の分析方法 | Staat](#)

手順では、例としてサンプルデータ_クロス集計①.csv を使用し、属性と副業の関連性を分析します。

1 クロス集計表を作成する

1. 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」が表示されます。



2. [クロス集計設定] の [表側 (行項目)] プルダウンメニューから、行に表示する項目 (変数) を選択します。
ここでは「属性」を選択します。
3. [クロス集計設定] の [表頭 (列項目)] プルダウンメニューから、列に表示する項目 (変数) を選択します。
ここでは「副業」を選択します。
4. [解析実行] ボタンをクリックします。
クロス集計表が表示されます。
次に、各種分析を順に実行します。



2 カイ二乗検定を実行する

1. メニューバーから、[分析] → [カイ二乗検定] を選択します。
「カイ二乗検定」が表示されます。



2. [解析実行] ボタンをクリックします。
[検定統計量]、[p 値]、[クラメールの連関係数] にそれぞれ値が表示されます。



3 フィッシャーの正確確率検定を実行する

1. メニューバーから、[分析] → [フィッシャーの正確確率検定] を選択します。
「フィッシャーの正確確率検定」が表示されます。

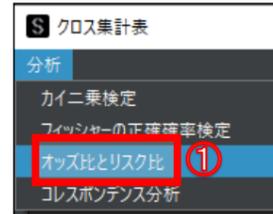


2. [解析実行] ボタンをクリックします。
[p 値] , [クラメールの連関係数] にそれぞれ値が表示されます。



4 オッズ比とリスク比を実行する

1. メニューバーから, [分析] → [オッズ比とリスク比] を選択します。
「オッズ比とリスク比」が表示されます。



2. [解析実行] ボタンをクリックします。
95%信頼区間 (95%CI 下限・95%CI 上限) が算出されます。
参考：[オッズ比とは | Staat](#)

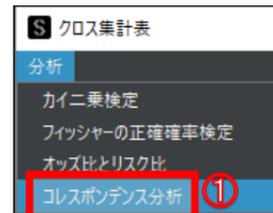
オッズ比とリスク比

解析結果

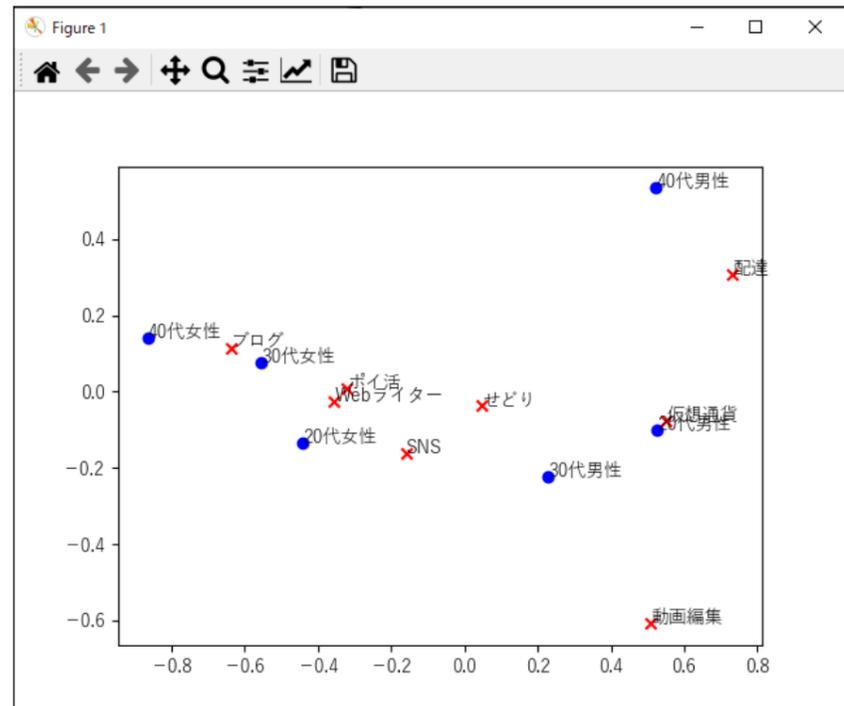
	値	95%CI 下限	95%CI 上限
オッズ比	3.6	1.25544	10.32309
対数オッズ比	1.28093	0.22748	2.33438
リスク比	1.65	1.07989	2.52109
対数リスク比	0.50078	0.07686	0.92469

5 コレスポンデンス分析を実行する

1. メニューバーから, [分析] → [コレスポンデンス分析] を選択します。



項目間の類似度を示した散布図が表示されます。



4-25 生存曲線

生存曲線とは、イベント発生までの時間と生存確率の関係を表した図です。イベント発生までの時間とは、以下のような状況を指します。

- 死亡するまでの時間
- 顧客が離反するまでの時間
- 機械が故障するまでの時間

イベントの発生と生存時間の関係を分析する統計手法が、生存時間解析です。生存時間解析の一般的な手法として、 Kaplan-Meier法があります。 Kaplan-Meier法では、観測期間内にイベントが発生しなかった場合等を打ち切りデータとして扱えます。 StaatApp では、 Kaplan-Meier法で解析します。 また、2群以上の生存曲線を比較するログランク検定や、一般化ウィルコクソン検定も可能です。

生存時間解析や生存曲線の見方の詳細については、以下を参照してください。

参考：[生存時間解析 | Staat](#)

以降で、生存曲線の作成手順を説明します。手順では、例としてサンプルデータ_生存時間解析.csv を使用し、死亡するまでの経過日数と、生存している、または死亡した状況の関係を解析します。

分析に必要なデータ

生存曲線を作成するには、イベント発生までの経過時間を表すデータと、イベント発生の有無を示すデータが必要です。以降の手順例では、イベント発生までの経過時間を表すデータが“経過日数”、イベント発生の有無を示すデータが“死亡状況”になります。

イベント発生の有無を示すデータは、イベントが発生した場合を“1”、打ち切りの場合を“0”とします。

1 生存曲線を表示する

1. 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」が表示されます。



2. [変数列選択] の [経過時間] プルダウンメニューから、グラフに表示する変数を選択します。

ここでは“経過日数”を選択します。

3. [変数列選択] の [イベント発生] プルダウンメニューから、グラフに表示する変数を選択します。

ここでは“死亡状況”を選択します。

4. 生存曲線に打ち切りデータを表示する場合は、 [グラフ表示項目] の [打ち切りデータ] をオンにします。

ここではデフォルト (オフ) 設定にします。

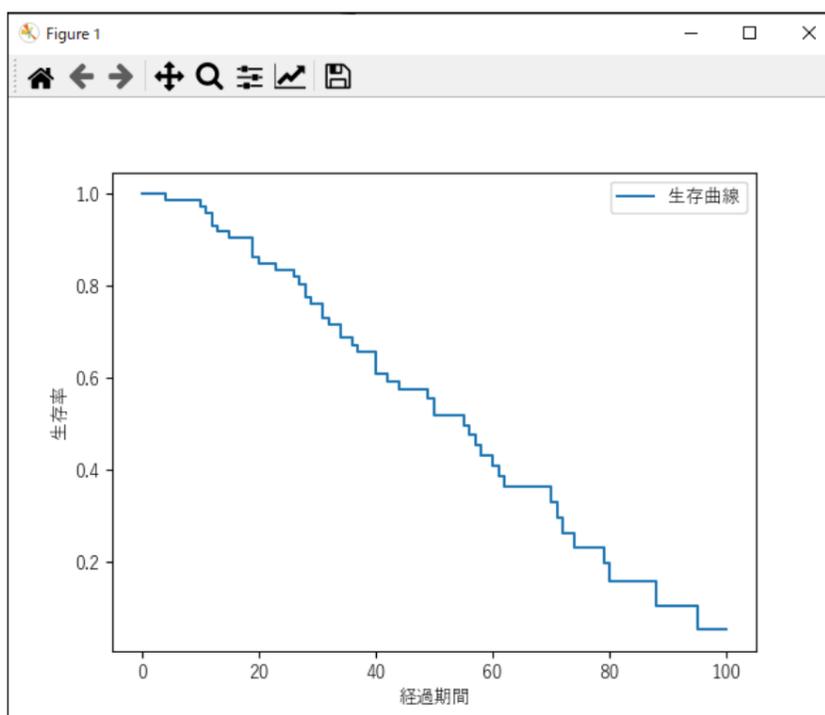
5. 生存曲線に 95%信頼区間を表示する場合は、 [グラフ表示項目] の [95%信頼区間] をオンにします。

ここではデフォルト (オフ) 設定にします。

6. [解析実行] ボタンをクリックします。

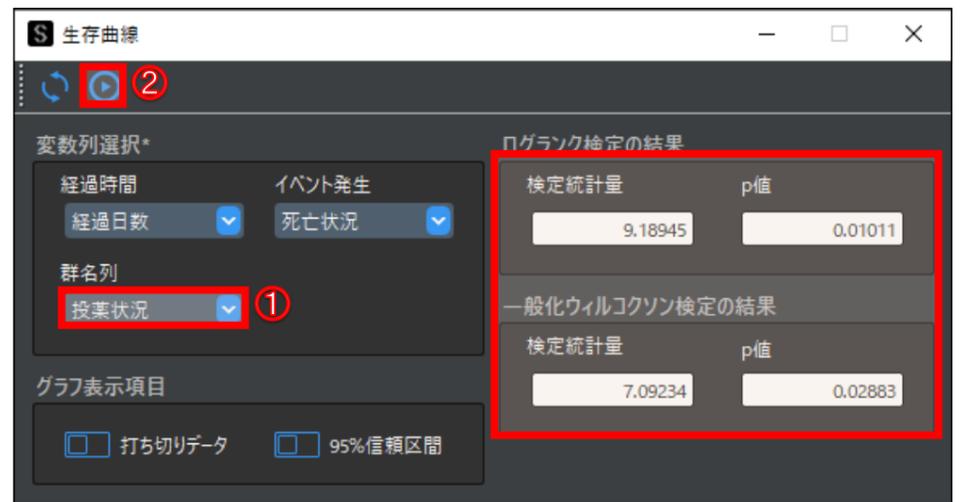
生存曲線のグラフが表示されます。

次に、ログランク検定と一般化ウィルコクソン検定を用いて2つ以上の生存曲線を比較します。

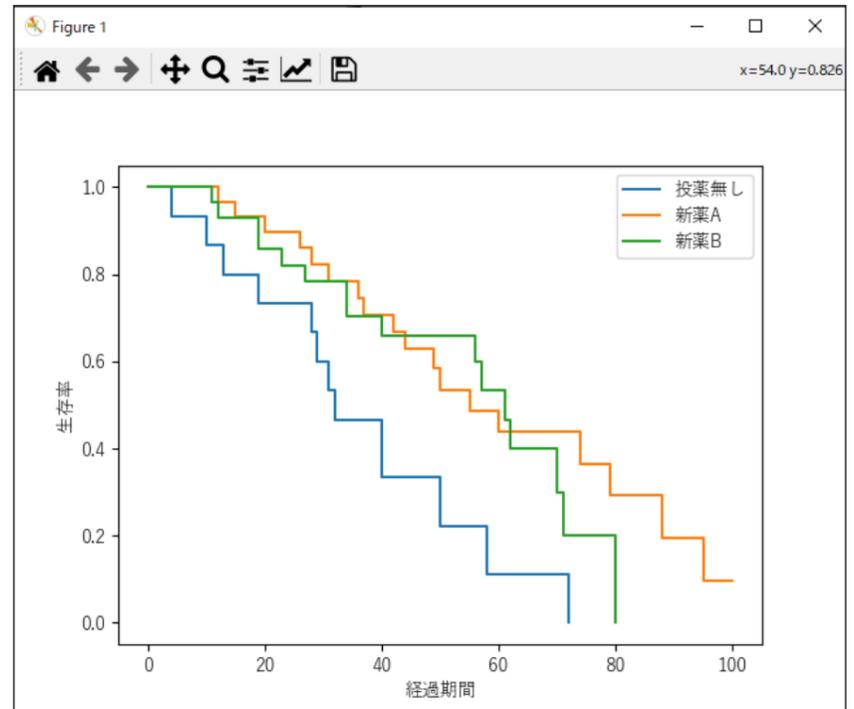


2 2つ以上の生存曲線を比較する

1. [変数選択] の [群名列] プルダウンメニューから、属する群を示すカテゴリデータが入力された変数 (列名) を選択します。
ここでは“投薬状況”を選択します。
2. [解析実行] ボタンをクリックします。
[ログランク検定の結果] 及び [一般化ウィルコクソン検定の結果] の [検定統計量] と [p 値] に、検定結果が表示されます。



同時に、生存曲線のグラフが表示されます。



4-26 Cox比例ハザード回帰

Cox 比例ハザード回帰とは、生存時間データのための重回帰分析です。

生存時間データは生存曲線で表され、イベント発生までの時間を示します。Cox 比例ハザード回帰を行うことで、生存時間データに影響を与えている要因（共変量）を調べたり、生存曲線を予測するモデルを作成したりできます。

Cox 比例ハザード回帰や解析結果の見方の詳細については、以下を参照してください。

参考：[Cox 比例ハザード回帰 | Staat](#)

以降で、Cox 比例ハザード回帰の実行手順と、予測モデルを使用した生存曲線の作成手順を説明します。

手順では、例としてサンプルデータ_Cox比例ハザード分析.csv を使用します。あらかじめ“性別”と“結婚”をダミー変数に変換しています（性別→女性/男性、結婚→既婚/未婚）。

参考：[エラー! 参照元が見つかりません。エラー! 参照元が見つかりません。\(P.エラー! ブックマークが定義されていません。\)](#)

1 Cox比例ハザード回帰を実行する

1. 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] が表示されます。



2. [目的変数] の [経過時間データ] プルダウンメニューから、経過時間データとする変数名（列名）を選択します。

ここでは“継続時間”を選択します。

3. [目的変数] の [イベント発生データ] プルダウンメニューから、発生データとする変数名（列名）を選択します。

ここでは“イベント発生”を選択します。

4. [説明変数] の [未選択] から、変数名を選択します。

選択した変数が [選択済み] に移動します。

ここでは、“収入”、“女性”、“既婚”を選択します。

ダミー変数を選択する場合は、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。

参考：[多重共線性 | Staat](#)

5. [解析項目] で解析方法を選択します。

ここでは [Cox比例ハザード回帰] を選択します。

6. [解析実行] ボタンをクリックします。

[解析結果] に解析結果が表示されます。



次に、Cox比例ハザード回帰で作成したモデルを用いて生存曲線の予測を行います。なお、生存曲線の予測を行うには、事前に解析が実行済みである必要があります。

[比例ハザード性の検証]

[比例ハザード性の検証] を選択した場合は、以下のような解析結果になります。



2 予測用データを作成する

1. StaatApp に読み込ませる予測用のデータを作成します。

Excel または csv 形式で作成します。

ここではサンプルデータ_Cox 予測用.csv を使用します。

例)

収入	年齢	性別_女性	結婚_既婚
580	32	1	0
430	28	0	0

予測用データの列名・カテゴリーデータ

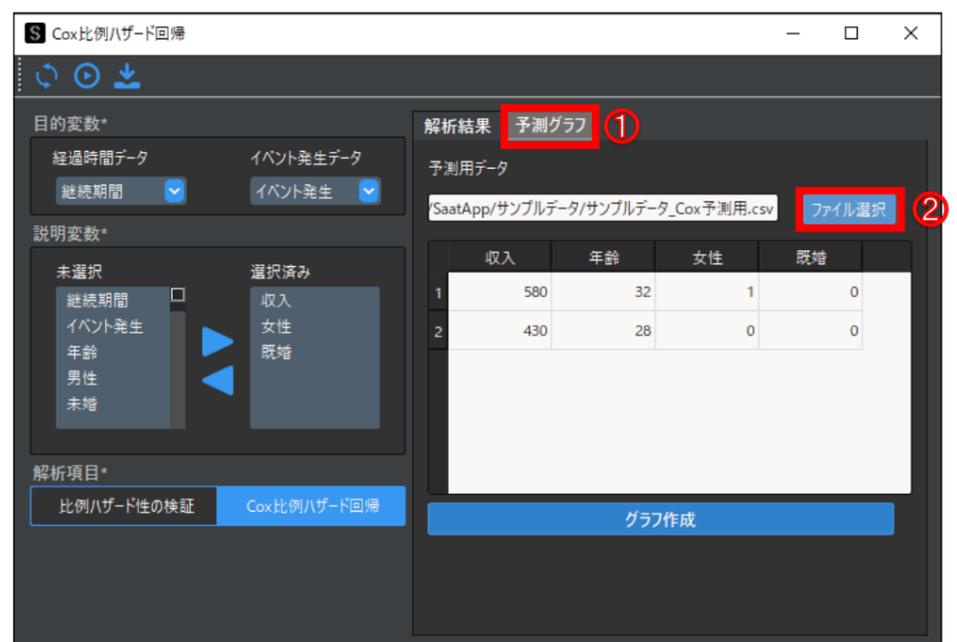
- 予測用データの列名は、1 Cox比例ハザード回帰を実行するの手順 4 で選択した変数名と同じにしてください。順不同です。
- カテゴリーデータは、事前にダミー変数に変換してください。

3 予測した生存曲線を作成する

1. [予測グラフ] タブをクリックします。

2. [ファイル選択] ボタンをクリックします。

「予測用データの選択」が表示されます。

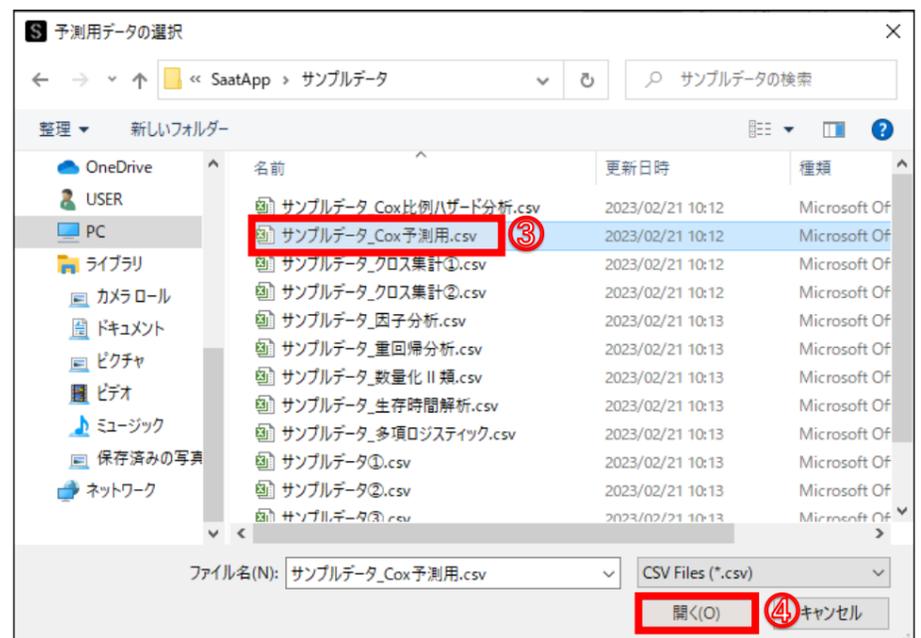


3. 予測用データのファイルを選択します。

ここでは“サンプルデータ_Cox 予測用.csv”を選択します。

4. [開く] ボタンをクリックします。

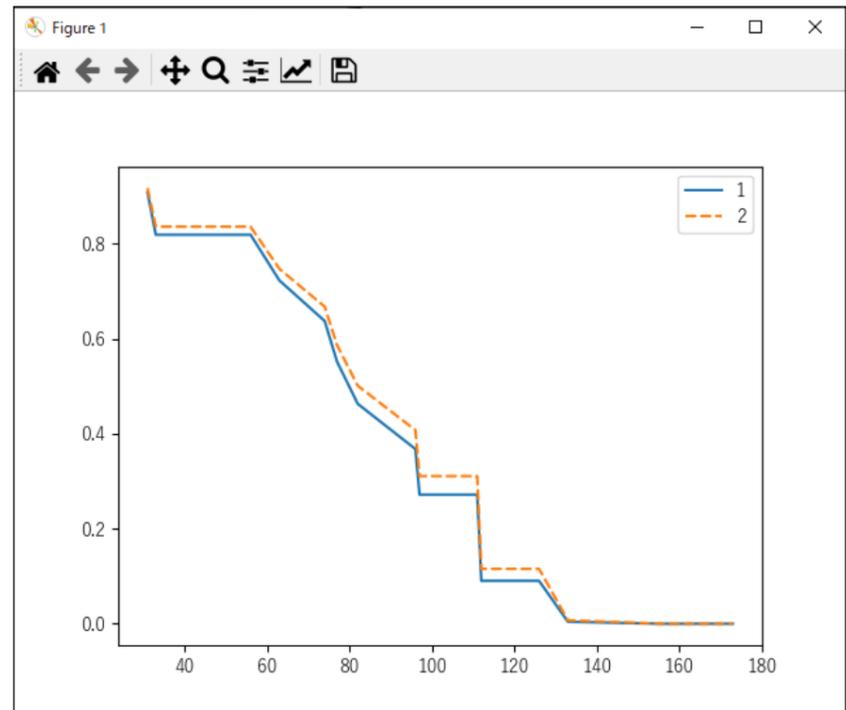
「Cox比例ハザード回帰」に戻り、読み込んだファイルのデータが表示されます。



5. [グラフ作成] ボタンをクリックします。

	収入	年齢	女性	既婚
1	580	32	1	0
2	430	28	0	0

生存曲線が表示されます。



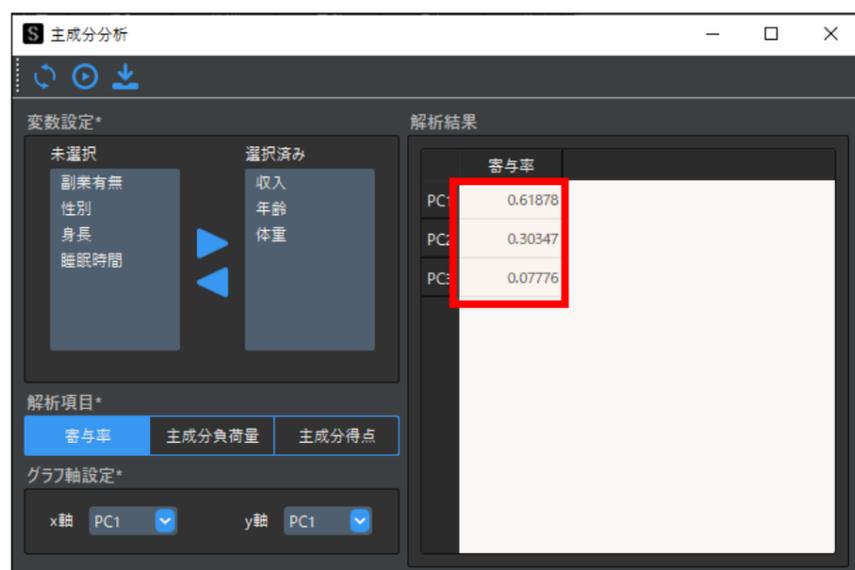
5 画面の表示設定

小数点の表示桁数と、StaatApp のカラーモードを変更できます。

5-1 小数点の表示桁数を変更する

各種検定の解析結果の数値は、デフォルトでは小数点以下 5 桁まで表示されます。小数点第 6 位の値は、四捨五入して表示されます。以降で、小数点以下の表示桁数を変更する手順を説明します。なお、データ操作画面の数値の表記に、本設定は反映されません。

手順では、例として小数点の表示桁数を 5 桁から 2 桁に変更します。

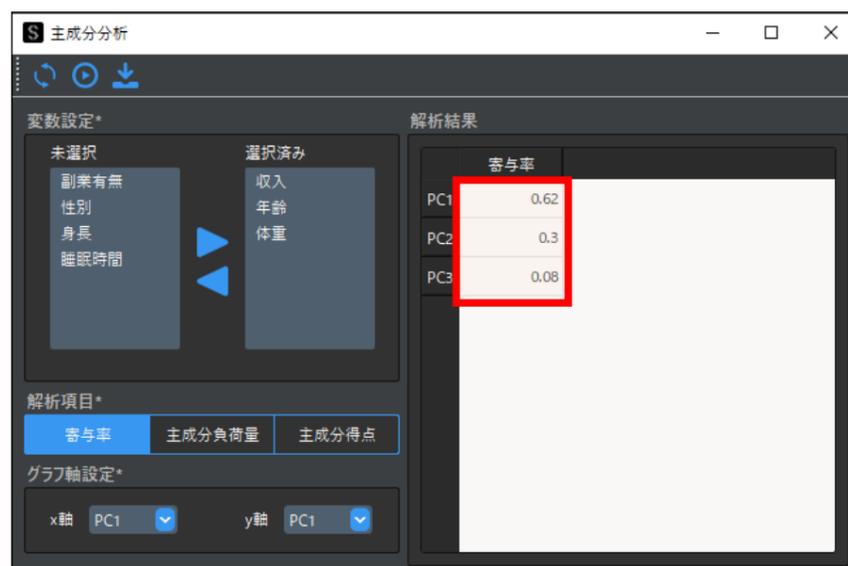


1 小数点の表示桁数を変更する

1. メニューバーから、[設定] → [桁数の設定] を選択します。
設定画面が表示されます。
2. [表示する小数の桁数] に、表示する小数点以下の桁数を入力します。
[-] または [+] ボタンをクリックして入力することもできます。
ここでは“2”と入力します。
3. [設定] ボタンをクリックします。



再度解析を実行すると、解析結果の小数点の桁数が 2 まで表示されます。



5-2 カラーモードを変更する

StaatApp のカラーモードには、ダークモードとノーマルモードがあります。デフォルトではダークモードに設定されています。以降で例として、カラーモードをノーマルモードに変更する手順を説明します。

1 カラーモードを変更する

1. メニューバーから [設定] → [スタイル] → [ノーマルモード] を選択します。メニューバーから、[設定] → [桁数の設定] を選択します。



カラーモードがノーマルモードに変更されます。

The screenshot shows the StaatApp application window displaying a data table. The table has columns for '副業有無' (Part-time job), '収入' (Income), '性別' (Gender), '年齢' (Age), '身長' (Height), and '体' (Body). The data is as follows:

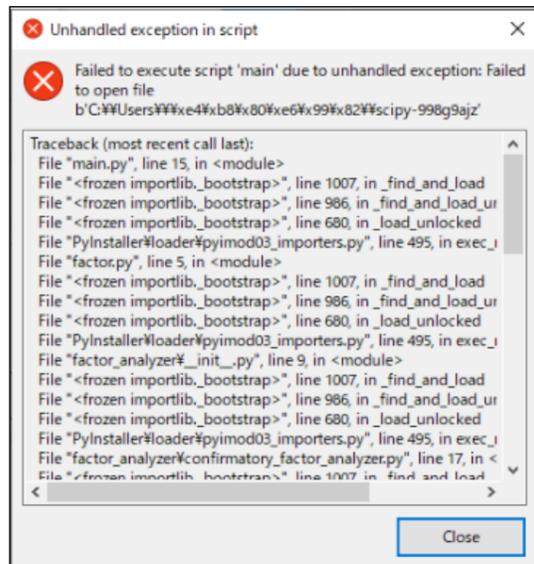
	副業有無	収入	性別	年齢	身長	体
1	1	580	0	32	170	
2	0	430	1	28	156	
3	1	800	0	45	163	
4	1	780	1	36	161	
5	1	690	1	42	158	
6	0	510	0	30	172	
7	0	740	1	49	165	
8	0	350	1	23	161	
9	1	620	0	29	180	
10	0	500	1	25	150	
11	0	430	0	33	168	
12	1	590	1	36	159	
13	1	1200	0	51	169	
14	0	810	0	53	174	
15	0	620	1	46	160	

6 付録

解析に使用するライブラリ、起動時のエラー、よくあるお問い合わせについて説明します。

6-1 起動時に発生するエラー

Windows のログオンユーザー名に日本語が含まれている場合、StaatApp を起動しようとすると、以下の画面が表示されて起動できません。



対処方法については、以下を参照してください。

参考：[StaatApp が起動しない場合の対応方法 | Staat](#)

なお、macOS では上記のエラーは発生しません。

6-2 よくあるお問い合わせ

StaatApp に関する質問と、質問に対する回答を記載します。

なお、その他の仕様や機能に関する質問は、Discord で受け付けています。以下を参照してください。

参考：[Discord で質問・相談 | Staat](#)

Q：データ数が多くても計算できますか？

A：サンプルサイズが 1 万程度のデータでも正常に計算できます。なお、計算速度は PC のスペックによって異なります。

Q：論文に結果を使用できますか？

A：Python の主要なライブラリを利用して計算しているため、論文に結果を利用できます。論文内で手法や参考文献を記載する際は、以下を参考に記載してください。

参考：[StaatApp の仕様詳細 | Staat](#)