

# IBM SPSS Modelerハンズオン 保守予測バージョン Ver2

2015年12月11日  
CAMSSセンター



## はじめに

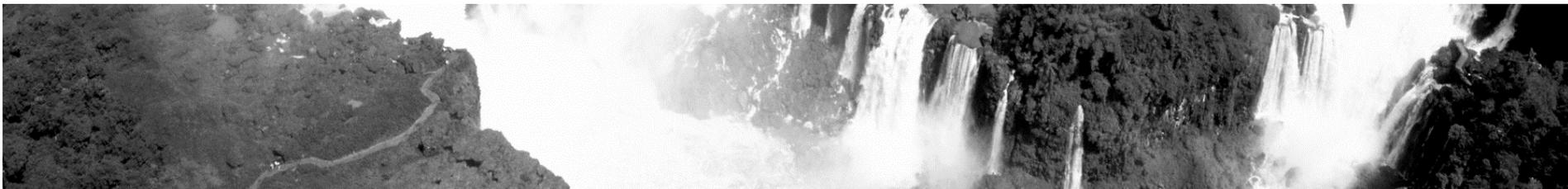
---

このハンズオン資料はIBM SPSS Modelerの操作を行いながら製品をご理解いただき、簡単なデモが実施できることを目的としています。

基本操作全般を習得するものや、資格取得を目的とするものではないことをご理解くださいますようお願いいたします。

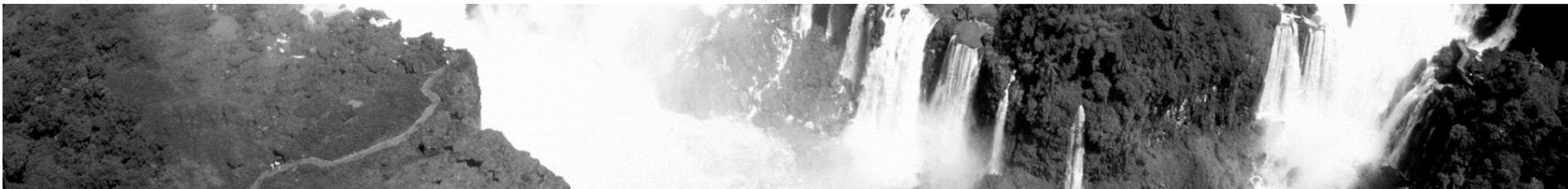


- ▶ データマイニングとIBM SPSS Modeler概要
- ▶ IBM SPSS Modeler ハンズオン 保守予測バージョン





- ▶ データマイニングとIBM SPSS Modeler概要
- ▶ IBM SPSS Modeler ハンズオン 保守予測バージョン



# データマイニングでできること

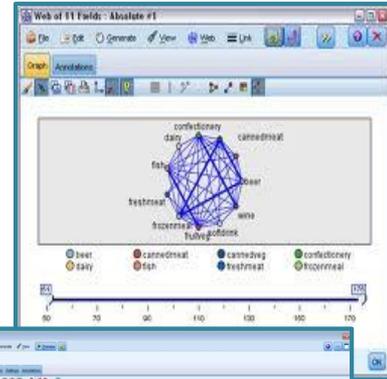
主にBIなどで分析できること以上の分析を行いたい場合に使います。

例えば、

## 1. パターン発見（アソシエーション）

POSの売上データや生産装置、カード履歴から特定のパターンを発見し、未知の法則を導き出す

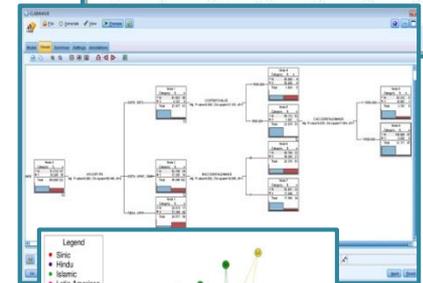
（例：Aという商品が何と一緒によく買われているか）



## 2. 予測（クラシフィケーション）

様々な過去の数値などを使って、未来の予測を行う

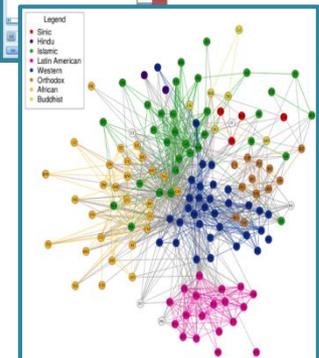
（例：平日来店し、Bという商品を一定数買う人はCを買う確率が80%）



## 3. 分類（セグメンテーション）

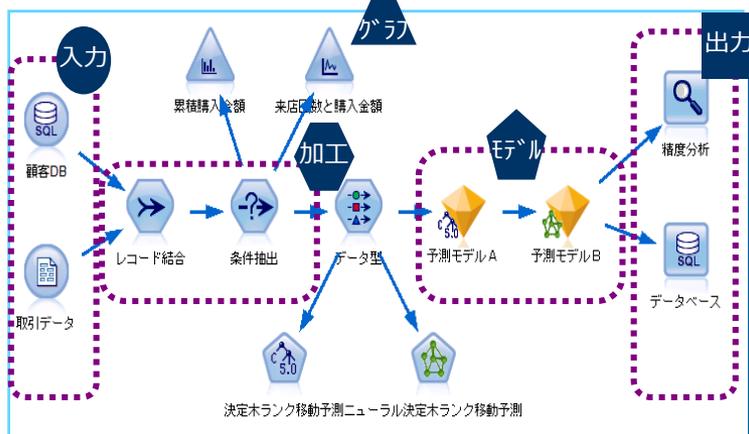
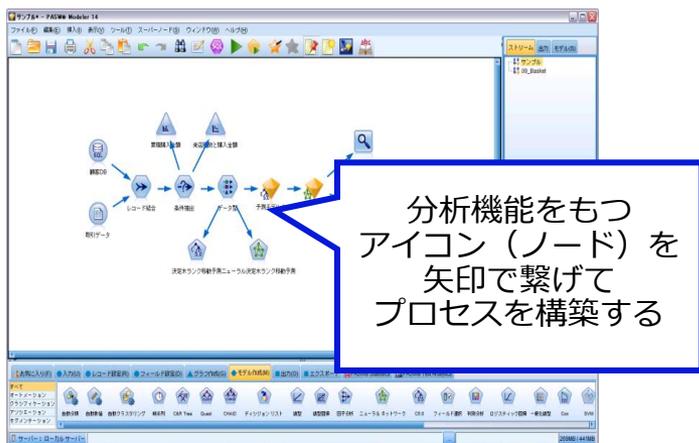
既知の情報（年齢、職業など）以外の情報を基に、新しいグループへの分類を行う

（例：A～Gまでの商品カテゴリの購入金額を基に、グループを作成する）



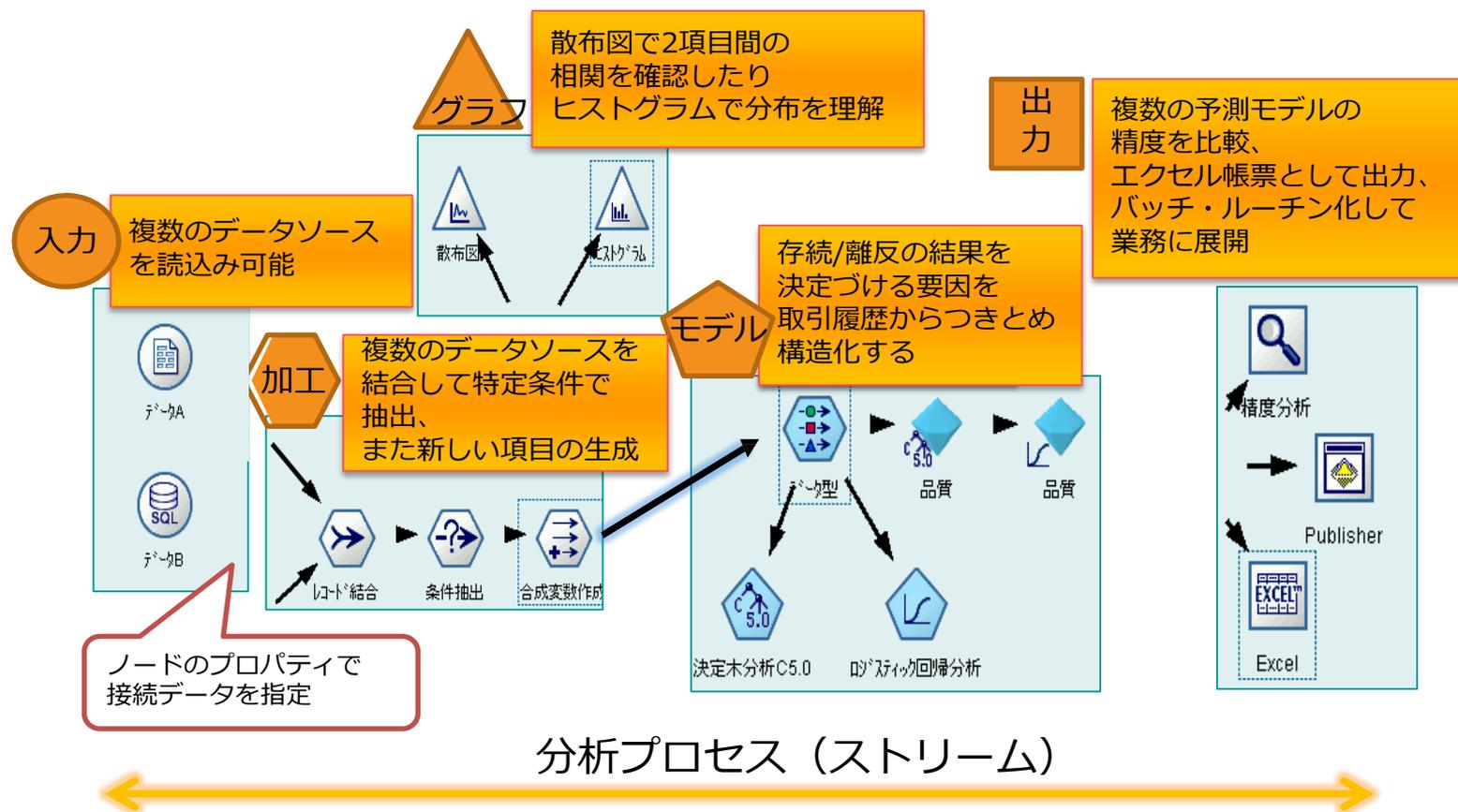
# SPSS Modelerとは

- 予測分析を行うためのツール、ワークベンチ（作業台）です
- データの入力から加工・視覚化・モデリング・出力までをサポートします



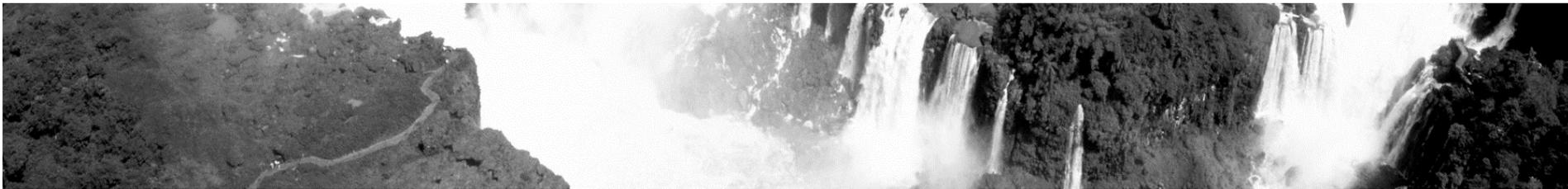
# 分析プロセス（ストリーム）

- アイコン（機能）と矢印をマウスで繋げて分析プロセスを作成します
- 試行錯誤をサポートし、分析再現性を確保します

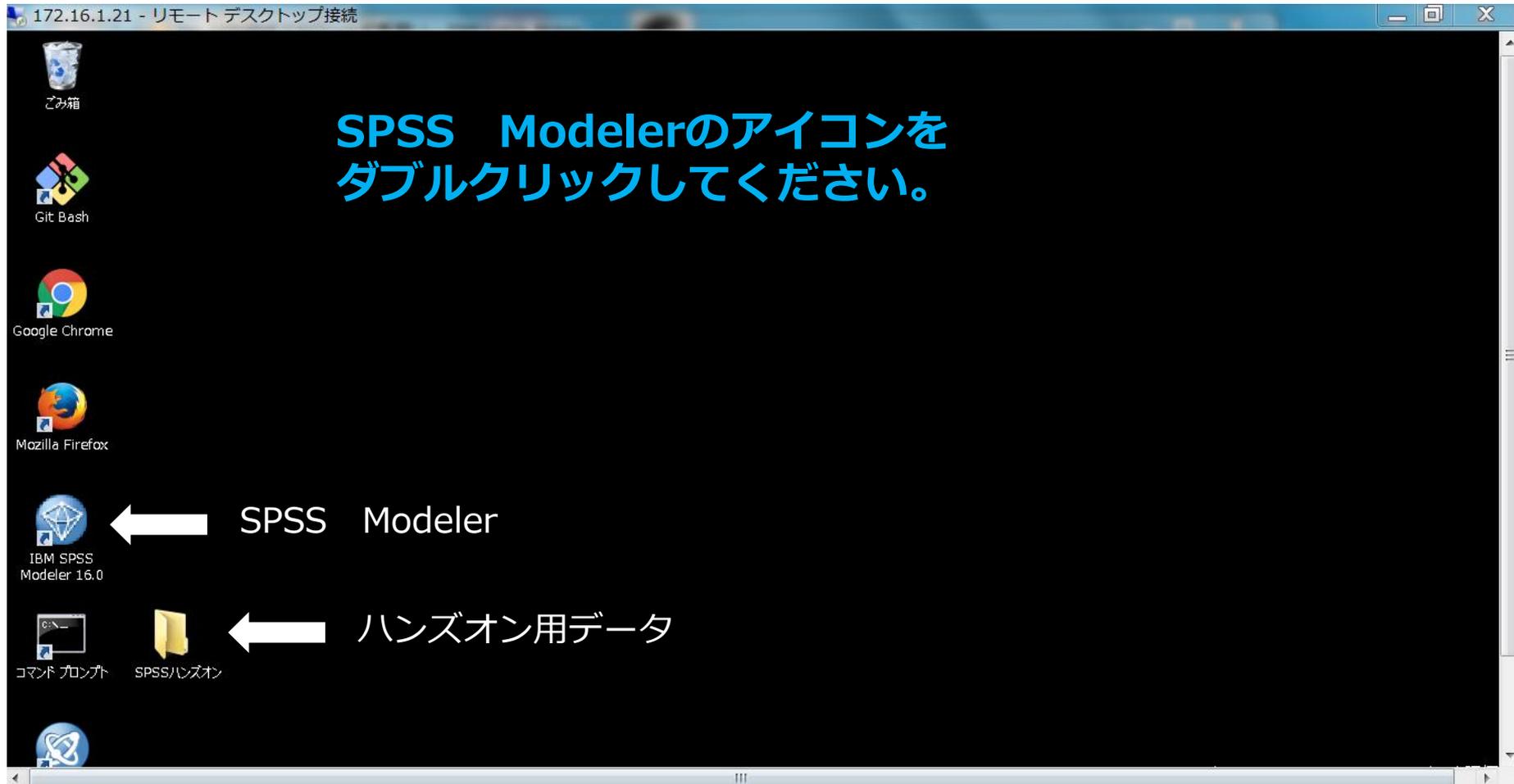




- ▶ データマイニングとIBM SPSS Modeler概要
- ▶ IBM SPSS Modeler ハンズオン 保守予測バージョン



# リモートデスクトップ接続後の画面



# 画面構成と用語

The screenshot shows the IBM SPSS Modeler interface. The main workspace is labeled "ストリーム領域" (Stream Area). The top menu bar includes "ファイル(F)", "編集(E)", "挿入(I)", "表示(O)", "ツール(T)", "スーパーノード(S)", "ウィンドウ(W)", and "ヘルプ(H)". The toolbar contains various icons for file operations, editing, and modeling. The right-hand side features a "マネージャ" (Manager) pane with tabs for "ストリーム", "出力", and "モデル(S)". The "ストリーム" tab is active, showing a project tree with "SPSS\_HandsOn\_demo" and "ストリーム1". Below this is a "CRISP-DM クラス" (CRISP-DM Class) pane showing a project structure: "(未保存のプロジェクト)" (Unsaved Project) containing "ビジネスの理解", "データの理解", "データの準備", "モデリング", "評価", and "展開". The bottom toolbar includes buttons for "お気に入り", "入力", "レコード設定", "フィールド設定", "グラフ作成", "モデル作成", "出力", and "エクスポート". The "モデル作成" button is highlighted with a callout box labeled "ノード". Below the toolbar is a palette labeled "パレット" (Palette) containing various modeling nodes such as "オートメーション", "クラシフィケーション", "アソシエーション", "セグメンテーション", "自動分類", "自動数値", "自動クラスタリング", "時系列", "C&R Tree", "QUEST", "CHAID", "ディビジョンリスト", "線型", "C5.0", "線型回帰", "因子分析", "ニューラルネット", "フィールド選択", "判別分析", "ロジスティック回帰", "一般化線型", "GLMM", "Cox", "SVM", "ベイズ", and "SLRM". The status bar at the bottom indicates "サーバー: ローカル サーバー" and "148MB / 271MB".

ストリーム1 - IBM® SPSS® Modeler

ファイル(F) 編集(E) 挿入(I) 表示(O) ツール(T) スーパーノード(S) ウィンドウ(W) ヘルプ(H)

ストリーム 出力 モデル(S)

SPSS\_HandsOn\_demo  
ストリーム1

マネージャ

ストリーム領域

ノード

パレット

お気に入り 入力 レコード設定 フィールド設定 グラフ作成 モデル作成 出力 エクスポート IBM® SPSS® Statistics

すべて  
オートメーション  
クラシフィケーション  
アソシエーション  
セグメンテーション  
自動分類 自動数値 自動クラスタリング 時系列 C&R Tree QUEST CHAID ディビジョンリスト 線型 C5.0 線型回帰 因子分析 ニューラルネット フィールド選択 判別分析 ロジスティック回帰 一般化線型 GLMM Cox SVM ベイズ SLRM

サーバー: ローカル サーバー 148MB / 271MB

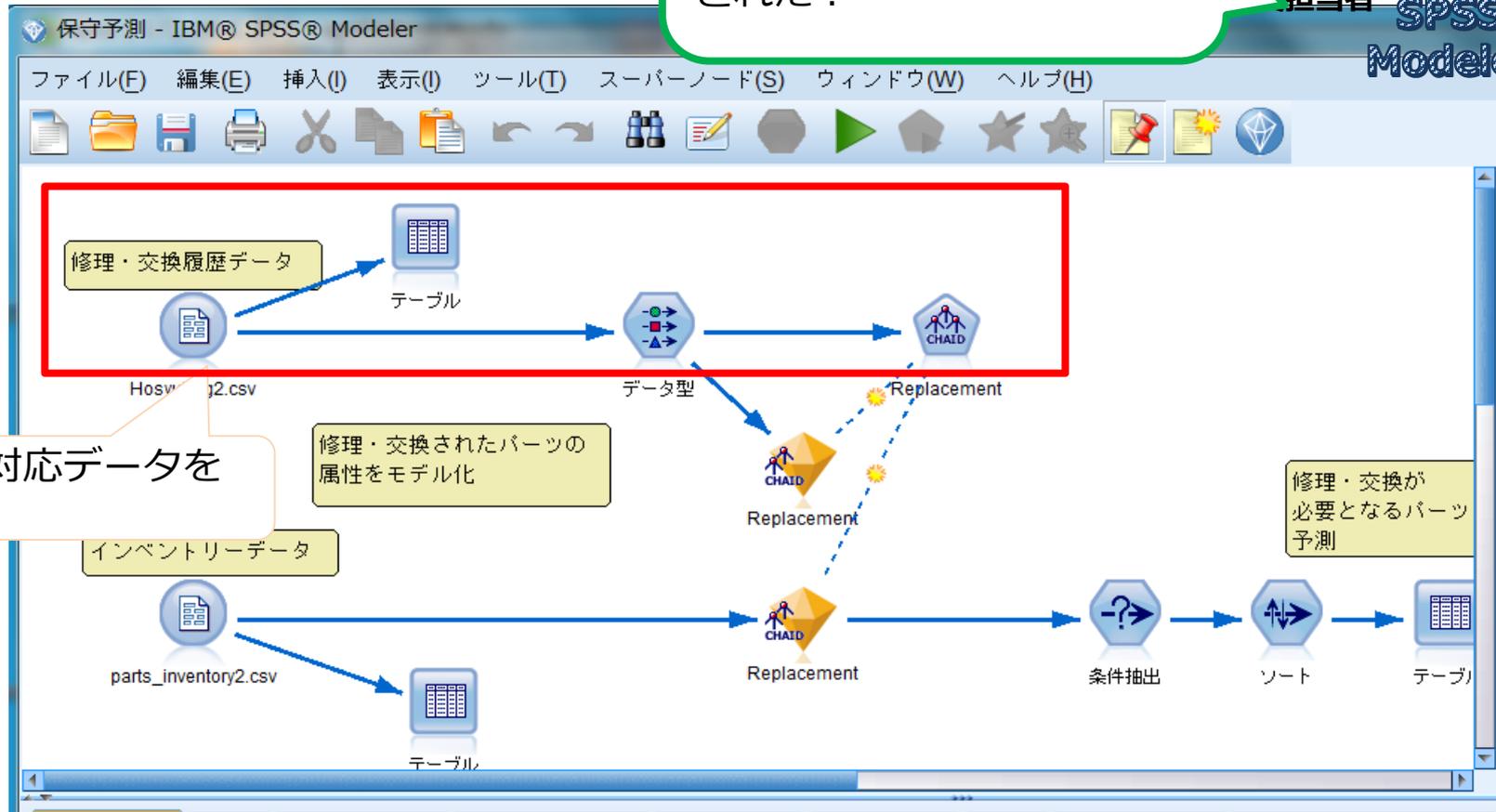
# 修理対応データを元に保守予測

修理対応データから何か分からないかな・・・。  
 現在販売した車のクラッチで修理交換が必要なものはどれだ？

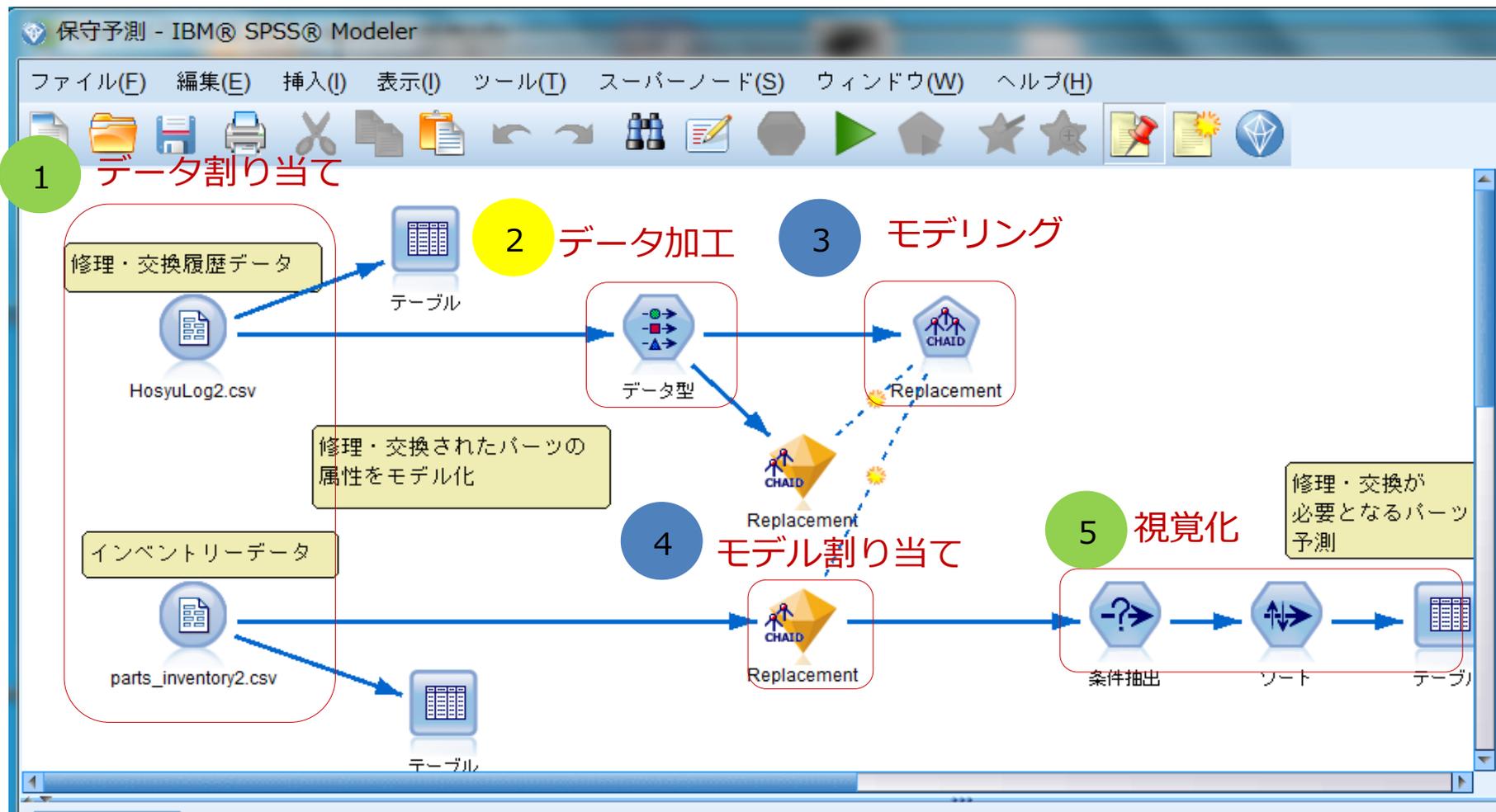


担当者

SPSS  
Modeler



# 今回作成するストリーム



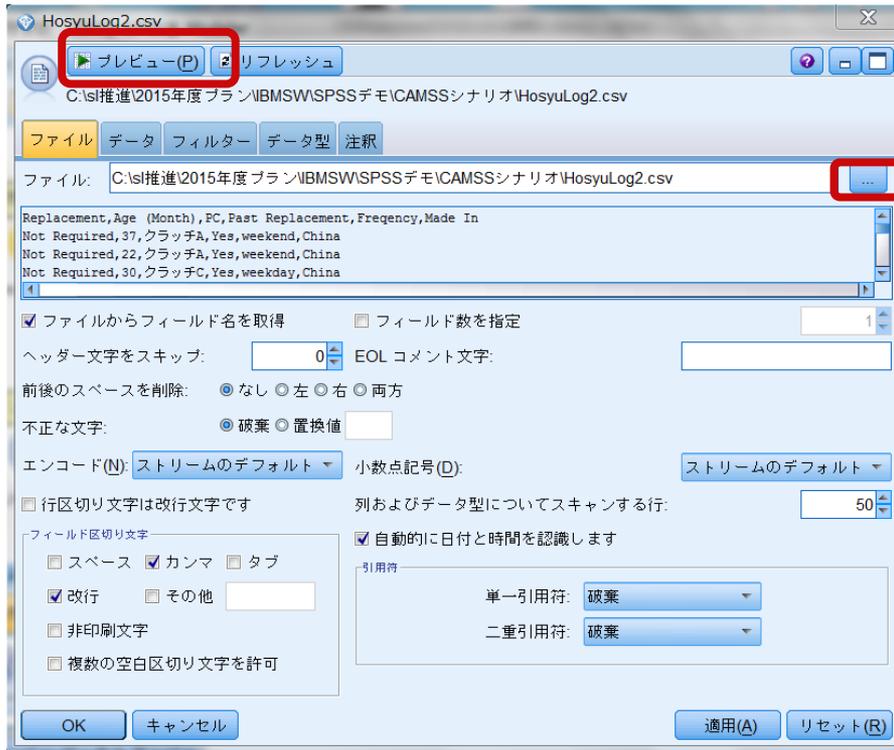
# ①データ割り当て 1 (CSVファイル1つ)

1. 入力ノードパレットから可変長ファイルノードをストリーム領域にドラッグしダブルクリック
2.  ボタンを押しHosyuLog2.csvを割り当てる
3. プレビューを実行⇒データの内容を確認



可変長ファイル

## 2.CSVファイルの割り当て



## プレビュー実行結果

	Replacement	Age (Month)	PC	Past Replacement	Freqency	Made In
1	Not Required	37	クラッチA	Yes	weekend	China
2	Not Required	22	クラッチA	Yes	weekend	China
3	Not Required	30	クラッチC	Yes	weekday	China
4	Not Required	33	クラッチC	Yes	weekend	Taiwan
5	Not Required	37	クラッチA	Yes	weekend	China
6	Not Required	40	クラッチA	Yes	weekend	China
7	Not Required	32	クラッチA	Yes	weekend	China
8	Not Required	35	クラッチC	Yes	weekday	China
9	Not Required	32	クラッチC	Yes	weekday	China
10	Not Required	25	クラッチA	Yes	weekend	China

## <今回使用するCSVファイル>

- ① HosyuLog2.csv
- ② parts\_inventory2.csv

DesktopのSPSSハンズオンフォルダーに格納しています。

## ②データ加工

1. フィールド設定パレットのデータ型ノードをストリーム領域にドラッグ
2. HosyuLog2.csvのノードから接続（以下のいずれかの操作方法）

操作方法 1：マウスの中央ボタンを押しながらドラッグ

操作方法 2：.csvのノードを右クリック⇒F2キー押す⇒レコード結合ノードを右クリック

3. データ型ノードをダブルクリック
4. 「値の読み込み」をクリック
5. ロールを以下の図に従い変更（ロールの各フィールドをクリックし選択変更）

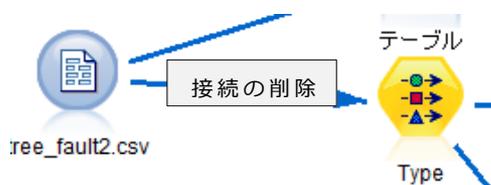
**Point** パーツを交換する（目的変数）要因（説明変数）を求めるため、パーツ交換の有無が分析のキーとなる

5. OKボタンをクリック

2. 操作方法 2 F2を押したときの画面



- 2.接続の解除（矢印上を右クリック）

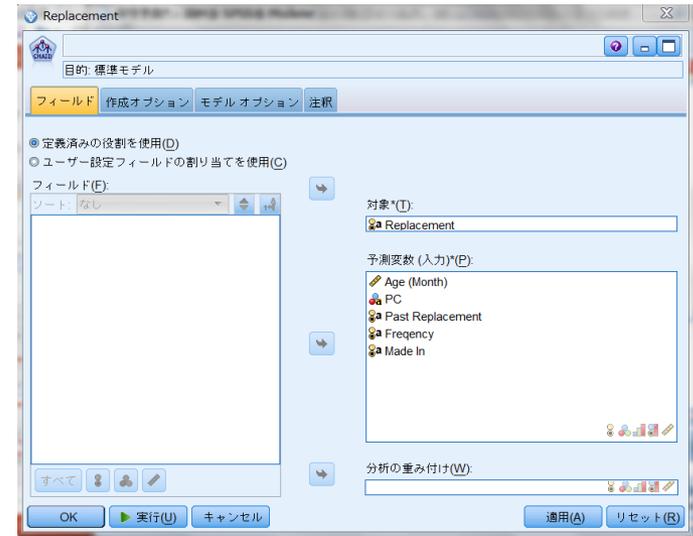
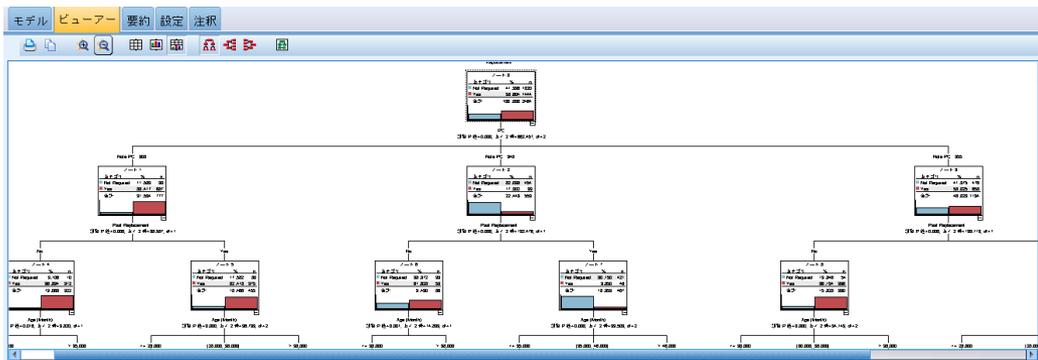


### ③モデリング パターン発見 (CHAID 決定木分析)

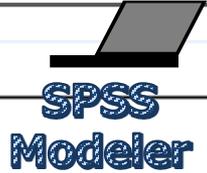
1. モデル作成パレットのCHAIDノードをストリームに配置し作成したデータ型ノードとリンク
2. CHAIDノードをダブルクリック
3. 右下の図のように設定 (PartsReplacementフラグをキーとし、それに結びつくパターンを発見する)
4. 実行ボタンをクリック (モデルナゲットが作成される)
5. PartsReplacementモデルナゲットをダブルクリック
6. ビューワータブをクリック
7. ビューワータブの表示  
ディビジョンツリーにより、交換・修理を行ったパーツとその要因をマイニング  
使用期間 (月) が要因のほとんどだがマザーボードは生産国にも要因がある。



↑モデルナゲット



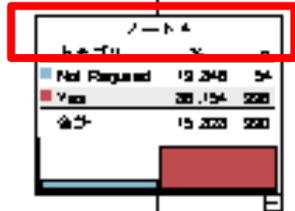
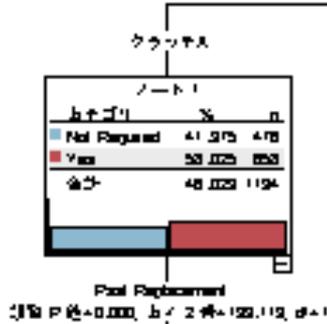
# 修理対応データを元に傾向分析



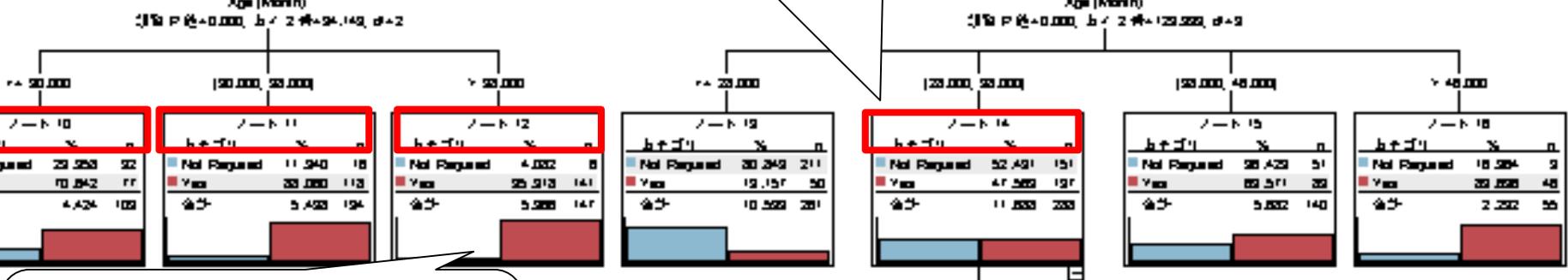
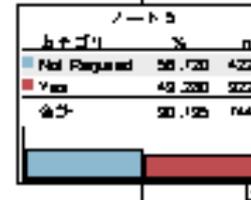
品質担当者

過去交換したことがないものは故障しやすい

ほほう、こういう傾向があるのか・・・。



今まで交換したもののでも28ヶ月以上使用すると故障率が上がる



クラッチAは38か月以上使用するとほとんど交換が必要

# ①データ割り当て 2 現在稼働しているパーツデータを追加

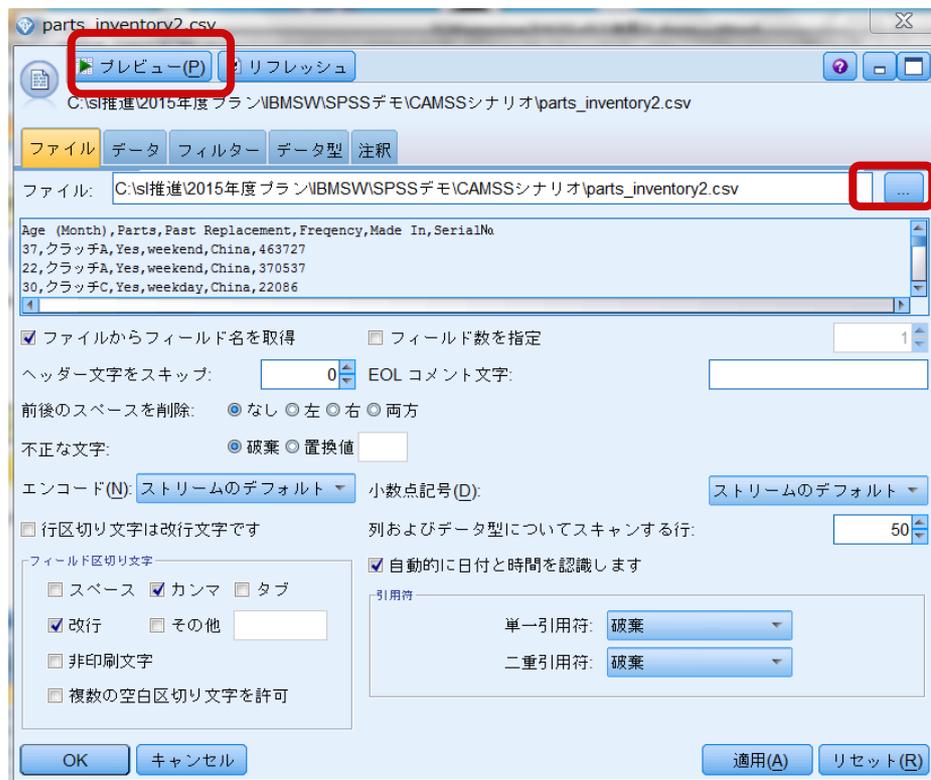
1. 入力ノードパレットから可変長ファイルノードをストリーム領域にドラッグしダブルクリック
2.  ボタンを押しparts\_inventory2.csvを割り当てる
3. プレビューを実行⇒データの内容を確認



可変長ファイル

CSVファイルの割り当て

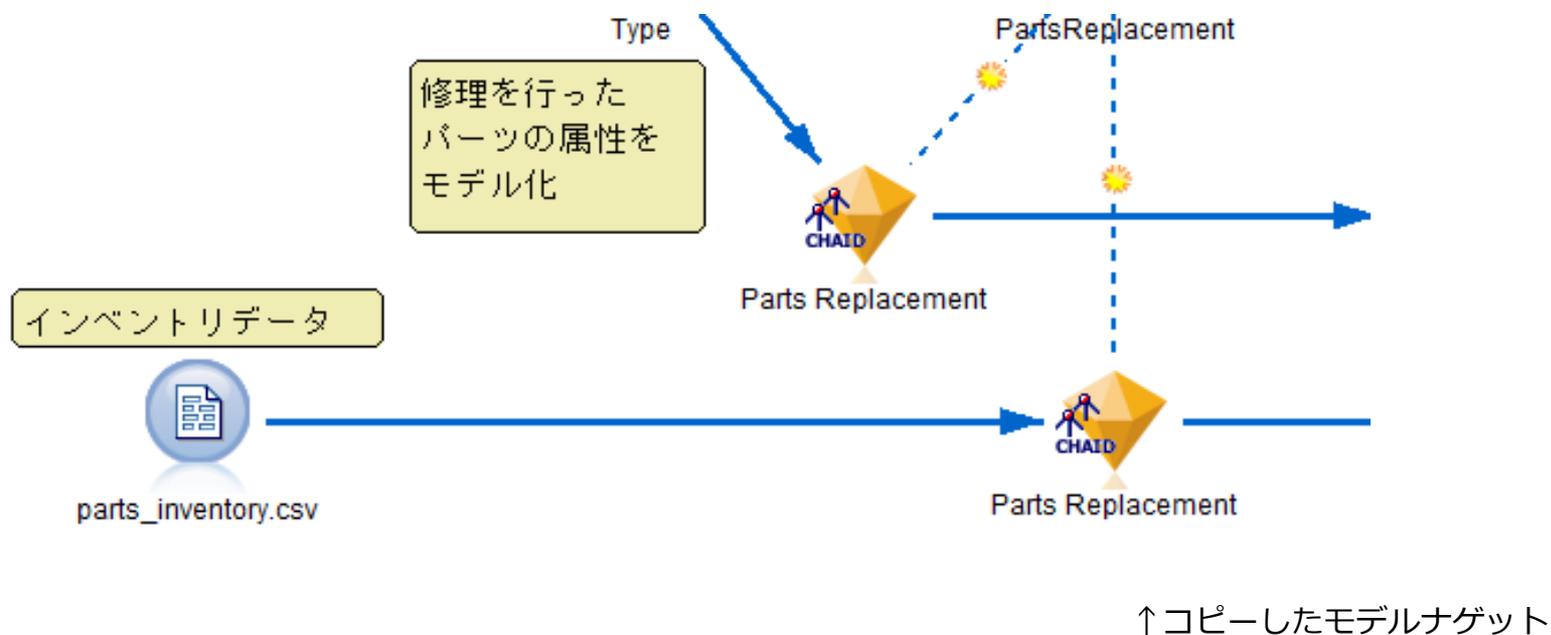
プレビュー実行結果



	Age (Month)	Parts	Past Replacement	Frequency	Made In	SerialNo.
1	37	クラッチA	Yes	weekend	China	463727
2	22	クラッチA	Yes	weekend	China	370537
3	30	クラッチC	Yes	weekday	China	22086
4	33	クラッチC	Yes	weekend	Taiwan	59607
5	37	クラッチA	Yes	weekend	China	452145
6	40	クラッチA	Yes	weekend	China	721373
7	32	クラッチA	Yes	weekend	China	656953
8	35	クラッチC	Yes	weekday	China	851087
9	32	クラッチC	Yes	weekday	China	622508
10	25	クラッチA	Yes	weekend	China	756610

## ④モデリング 1 作成したモデルを割り当て (流用)

1. 先に作成されたモデルナゲットをコピー & ペースト
2. parts\_inventory2.csvから接続



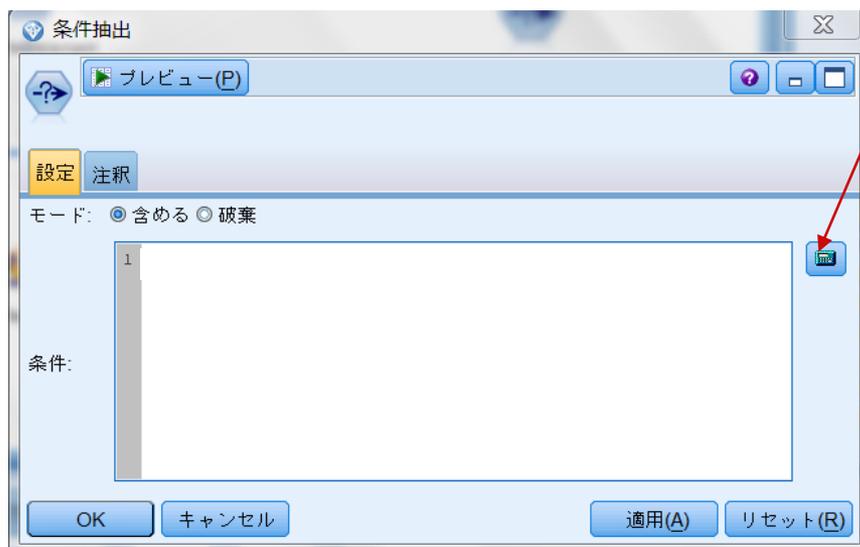
## ⑤視覚化 交換修理が必要と予測されたパーツのみを抽出

1. レコード設定パレットから条件抽出ノードをストリーム領域に配置しモデルナゲットとリンク
2. 条件抽出ノードをダブルクリック
3. 式入力ボタン  を押し、Clem式ビルダーのウィンドウが立ち上がる
4. 右下フィールドウインドウの\$R-Parts Replacement ダブルクリック
5. 上部の窓に' \$R-Parts Replacement 'と記述されるので、続いて='Yes'と入力（右下画面のとおり）
6. 検査ボタンを押す（念のため、構文に間違いがあると式が赤文字になります）
7. OKボタンを押す
8. 条件抽出ウインドウでOKボタンを押す



条件抽出

式入力ボタン



## ⑤視覚化 確率順にソート

レコード設定パレットからソートノードをストリーム領域に条件抽出ノードとリンク

ソートノードをダブルクリック

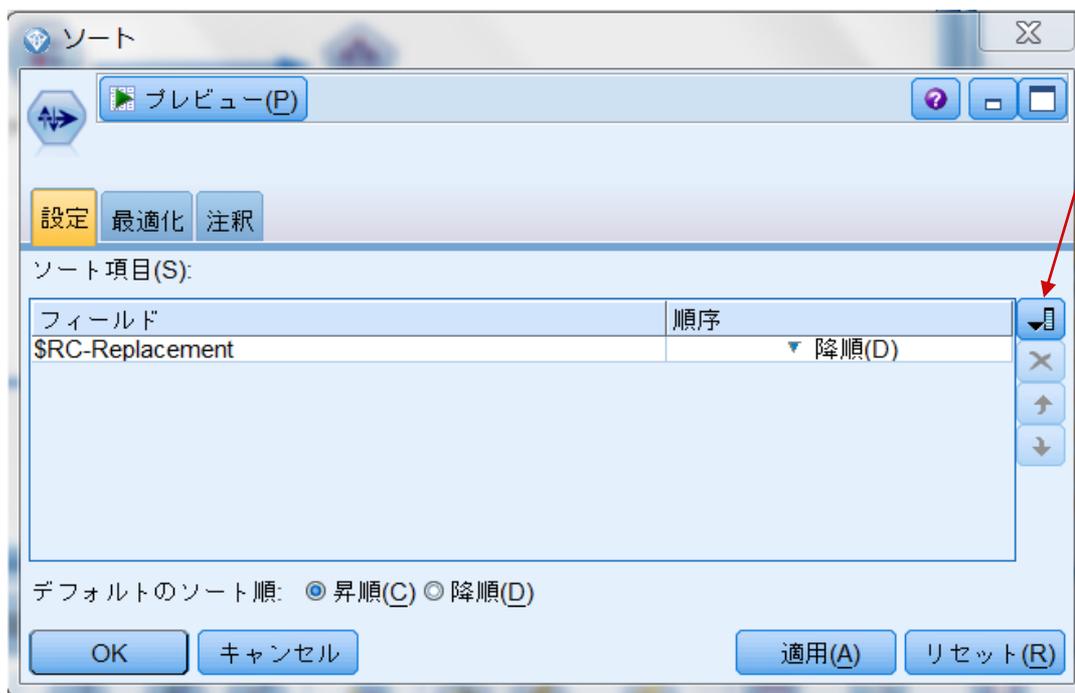
フィールド選択ボタン  を押し、フィールド選択のウィンドウが立ち上がる

ウィンドウの中の\$RC-Parts Replacement ダブルクリック

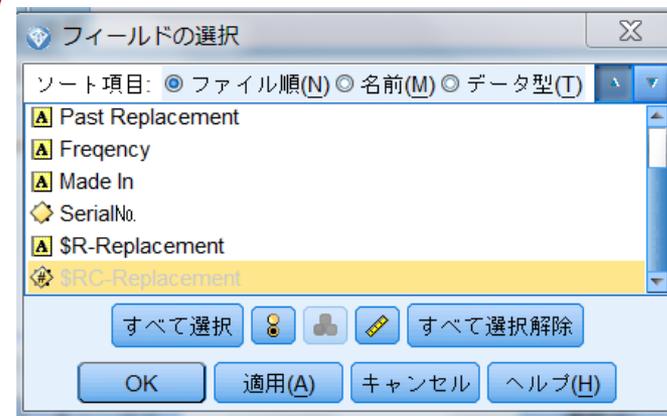
順序を降順に設定

OKボタンを押す

ソートウィンドウでOKボタンを押す



フィールド選択ボタン



## ⑤視覚化 (確率付き予測)

1. 出カパレットからテーブルノードをストリーム領域に配置しソートノードとリンク
2. テーブルノードをダブルクリック
4. 実行ボタンをクリック



テーブル

交換修理が必要となるパーツの確率付き一覧表をOUTPUT

シリアル番号をもとに先行保守を行うことで、保守パーツ在庫の削減、保守要員配置の効率化が可能となる。

テーブル (8 フィールド、1,425 レコード) #2

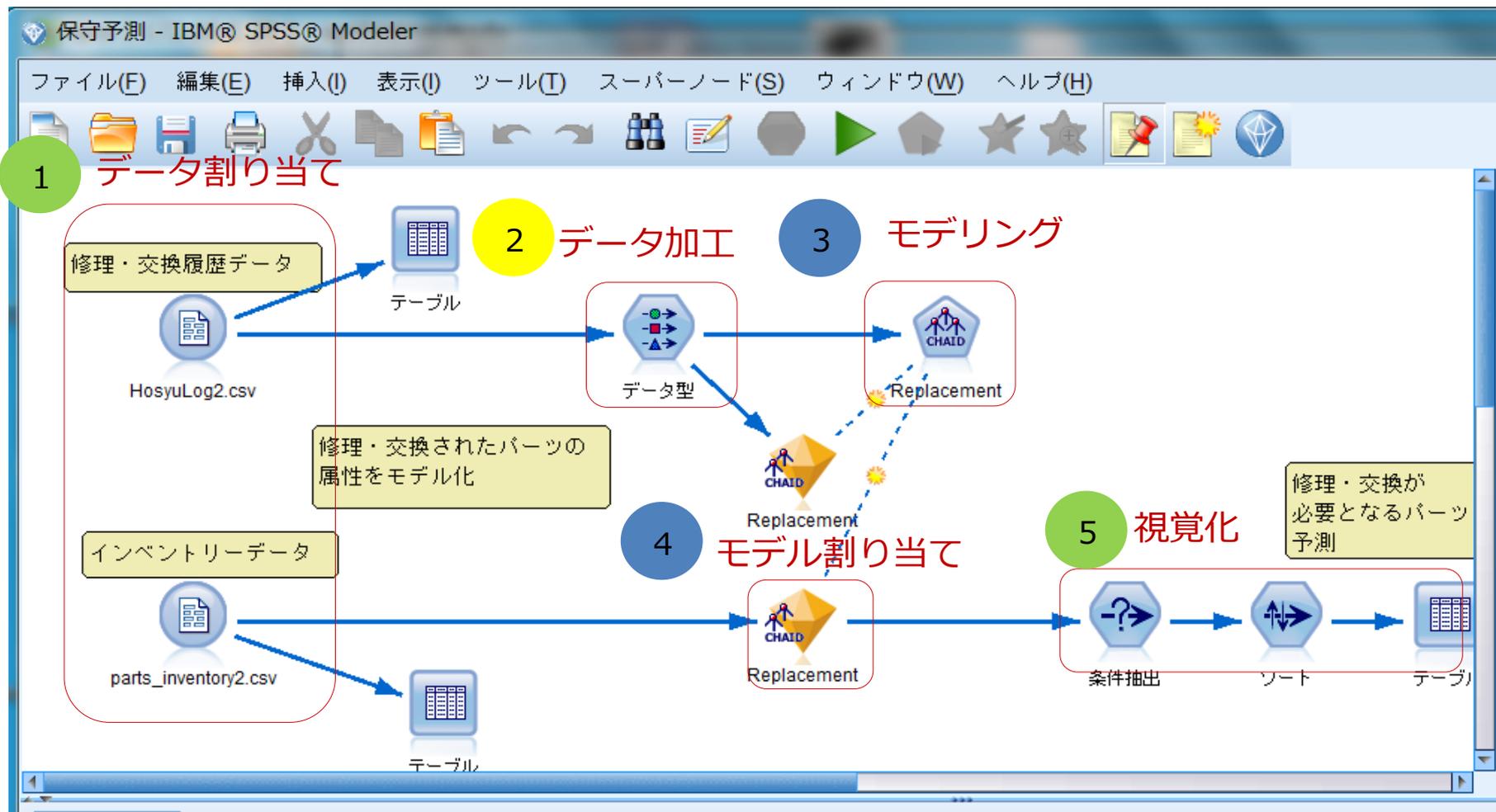
ファイル(F) 編集(E) 生成(G)

テーブル 注釈

	Age (Month)	Parts	Past Replacement	Freqency	Made In	SerialNo.	\$R-Replacement	\$SRC-Replacement
1	58	ク...	No	weekend	China	629821	Yes	0.994
2	55	ク...	No	weekday	Taiwan	156360	Yes	0.994
3	36	ク...	No	weekday	Taiwan	134261	Yes	0.994
4	43	ク...	No	weekday	Taiwan	983117	Yes	0.994
5	50	ク...	No	weekend	Taiwan	952348	Yes	0.994
6	39	ク...	No	weekend	Taiwan	762670	Yes	0.994
7	40	ク...	No	weekday	Taiwan	56814	Yes	0.994
8	43	ク...	No	weekday	Taiwan	274983	Yes	0.994
9	38	ク...	No	weekday	Taiwan	351894	Yes	0.994
10	39	ク...	No	weekend	Taiwan	564850	Yes	0.994
11	41	ク...	No	weekend	Taiwan	73984	Yes	0.994
12	61	ク...	No	weekday	Taiwan	181781	Yes	0.994
13	36	ク...	No	weekend	Taiwan	452180	Yes	0.994
14	40	ク...	No	weekday	China	889927	Yes	0.994
15	54	ク...	No	weekday	Taiwan	995959	Yes	0.994
16	49	ク...	No	weekday	Taiwan	965554	Yes	0.994
17	44	ク...	No	weekend	Taiwan	230210	Yes	0.994
18	42	ク...	No	weekday	Taiwan	995942	Yes	0.994
19	45	ク...	No	weekday	China	354527	Yes	0.994
20	40	ク...	No	weekday	Taiwan	156952	Yes	0.994

OK

# 今回作成したストリーム



ありがとうございました。

**i**GUAZU

