DMT デシジョンツリーVer.1.3
使用マニュアル

2017 年 3 月 1 日

データマインテック株式会社

注：本マニュアル記載内容は予告なく変更される場合があります。

目次

1. 概要

1.1 DMT デシジョンツリーの概要

DMT デシジョンツリーは、予測モデル自動作成手法の 1 つである「デシジョンツリー」（または「決定木」、「判別ツリー」などと呼ばれる）を SAS または WPS 上で実行するアプリケーションプログラムです。取り扱えるモデルの予測対象は、カテゴリカル変数の特定カテゴリ（クラス）の出現率、連続変数の平均値、さらに、施策実施が有効／無効な集団を特定するための実施群と非実施群間の応答差（「アップリフト」と呼ばれる）です。

DMT デシジョンツリーは、予測モデルの自動作成の他に、説明変数の事前絞り込み機能、新しいデータに予測値を付与する機能、モデルの性能を精度や収益の観点から評価するさまざまな図表の作成機能、新しく出現したデータに基づくモデルの修正機能、スコアコード出力機能などを備えています。

1.2 デシジョンツリーの概要

デシジョンツリーが自動的に作成される仕組みは単純です。まず、全体を 1 つのノード（ルートノード）とみなして、どの説明変数のどの値を用いてこのノードに含まれるオブザベーションを 2 つのノードに分割すれば、"分割後の 2 つのノード間の目的変数の分布の違いが最大となる"かを、すべての説明変数について計算を行い、最も効果のある説明変数の値を分割条件に使って実際の分割を行います。分割後の各ノードについても、同様の処理を繰り返します。そして、各ノードは、"もはやこれ以上分割できない"と判断されると終端ノードとなります。

全ノードが終端ノードになったとき分割処理は終了し、決定木が完成します。1 回の分割が行われるたびに全ノード数が 1 個から 3 個、3 個から 5 個へ...といったように 2 つずつ増加し、最終的には階層的に分割された 1+2*（分割回数）個のノード数を持つ決定木が生成されます。そのうち終端ノードの数は（分割回数+1）個、中間ノード（ルートノードでも終端ノードでも無いノードのこと）の数は（分割回数-1）個となります。

なお、関心のある目的変数がカテゴリカル変数の場合は「分類木」と呼ばれ、ターゲット変数の値そのものを予測する場合は「回帰木」と呼ばれます。ターゲットが実施群（処理群）と非実施群（対照群）間の応答差（アップリフト）の場合は、本アプリケーションでは便宜的に「分類木アップリフト」、「回帰木アップリフト」と呼ぶことにします。

1.3 応用分野

業種や業務分野に関わらず、予測モデルの自動構築と、構築したモデルを用いて予測値の大きい対象を選別する（または除外する）という意思決定に広く利用することができます。また、実施施策の効果分析に用いることができます。

例えば、金融業においては顧客（企業・個人）に対する与信判断（新規および途上与信）や優良顧客の選別と離反防止、特定の金融商品推薦などに用いることができます。また、製造業においては生産工程上の歩留まり原因分析、建設業においては危険予知（ヒヤリハット）分析、流通・販売業においては商品購買分析や顧客維持分析、DM 送付先の適正化など、業務上のデータ分析の課題に幅広く適用できます。

1.4 動作環境

DMT デシジョンツリーVer.1.3 は、32 ビットまたは 64 ビット Windows 版 SAS[1] バージョン 9.2 以降の Base SAS, SAS/GRAPH および SAS/STAT プロダクト、または、WPS [2] バージョン 3.1 以降の WPS Core、WPS Graphing および WPS Statistics プロダクトが稼働している計算機システム上で動作します。

1.5 実行モード

本バージョンでは Windows デスクトップから独自の GUI 画面を起動し、画面から入力データやパラメータを選択・指定しながら分析を実行するモード（GUI 実行モード）と、SAS または WPS を対話モード（SAS ディスプレイマネージャ、または WPS ワークベンチ）で起動し、プログラムエディタ画面に本アプリケーションのマクロ呼び出しコマンドを入力し、実行するモード（コマンド実行モード）をサポートしています。
ただし、コマンド実行モードは SAS の SAS/Enterprise Guide 上では動作しない点に注意。SAS Foundations （SAS Display Manager）モードで起動できる環境が必要です。

1.6 構成要素

後述の導入方法に記した方法で本アプリケーション（DMT デシジョンツリーV1.3.exe）を導入すると、以下の SAS マクロで書かれた分析モジュール（マクロエントリ）を含むコンパイル済み SAS マクロカタログ（sasmacr.sas7bcat、または、SASMACR.wpccat）が指定したフォルダー内に生成されます。

（SAS マクロカタログに含まれる分析モジュール）
主要な分析モジュールは以下の 15 個です。
① DMT_CROSS ……説明変数とターゲット変数との関連性を AIC 統計量により評価します.

[1] SAS は米国 SAS Institute Inc. の登録商標です。
[2] WPS は英国 World Programming Ltd. の登録商標です。

② DMT_CROSSPLOT…… DMT_CROSS 実行結果をグラフ表示します
③ DMT_CROSSTAB …… DMT_CROSS 実行結果を表形式で表示します.
④ DMT_DATASAMP…… データセットのオブザベーションを分析用サンプルデータセットと検証用テストデータセットにランダムに振り分けます.
⑤ DMT_TREE …… ツリーモデルを作成します.
⑥ DMT_CVTREE…… ツリーモデルと交差検証法による検証結果を表すモデル形式データセットを作成します.
⑦ DMT_TREETAB …… ツリーモデルを分岐の仕方が分かる表形式で表示します.
⑧ DMT_NODETAB …… モデルの終端ノードをターゲット出現率（または平均値）の大きさの順にならべた上で、各ノードの属性定義とノード別統計量およびノード累積統計量を表示します.
⑨ DMT_TREECUT …… モデルの中間ノードの下のノードを削除（枝刈り）します.
⑩ DMT_TREEADD …… モデルの終端ノードに他のツリーモデルを接ぎ木（枝接ぎ）します.
⑪ DMT_TREESCORE…… データにモデルを適用し、個々のオブザベーションに予測値を付与します（スコアリング）。また、モデルの各ノードにおける該当件数とターゲット出現率を入力データに即して再計算したモデル形式データセットを作成します（検証モデルの作成、またはモデルの予測値の更新）。
⑫ DMT_GAINCHART …… モデルのゲインチャート(CAP 図)、ROC チャート、収益チャートを作成します.
⑬ DMT_COMPAREPLOT モデル予測値と実績値の比較プロットを作成します.
⑭ DMT_UPLIFTCHART… アップリフトモデル作成結果を図示します。
⑮ DMT_CORRECTTAB… 指定の予測確率をしきい値としたターゲット／非ターゲットの予測件数と実績件数のクロス集計表を作成し正答率を表示します.

なお、以上の他にいくつかのサブルーティンマクロが含まれます。

（GUI 実行機能）
DMT デシジョンツリーV1.3.exe には、SAS マクロカタログの生成を行うと共に、メニュー画面からマクロのパラメータを指定して実行する機能が含まれています。
この中には、マクロカタログに含まれるほとんどの分析機能に加えて、統計モデル作成機能や、データ入力・加工・ラベル定義やデータや結果ファイルの管理等の付加機能を備えています。
以下のような**「メニュー」**画面や各**「分析指定」**画面上でのマウス操作によるデータ分析が可能です。

「メニュー」画面

「デシジョンツリーモデル作成画面」

（サンプル実行プログラム）
DMT_TREE_VER1.3_SAMPLERUN.sas
DMT デシジョンツリーV1.3 をコマンド実行モードで実行する場合のサンプルプログラムです。

1.7 バージョン 1.3 の新機能および追加変更点

［新マクロモジュール］
DMT_CVTREE
……　全分析データを用いたモデル作成を行った後、交差検証法(Cross Validation)によるモデルの交差検証結果を表すモデル形式データセットを作成します。モデル検証用データを別途確保しておく必要がないため、特に、分析データ件数が少ないときに有用です。
DMT_UPLIFTCHART
……　施策実施効果（施策実施によるターゲット出現率またはターゲット平均値の実施しなかった場合に対する増加量）の分析結果を図示します。施策実施群（処理群）、施策非実施群（対照群）それぞれについて、施策実施効果

の大きい順（対照群においては逆順）にノードを並べたときの累積増加応答（累積アップリフト）のプロット図を作成します。

［機能追加変更マクロモジュール］

DMT_CROSS

…… (1) CONTROL= パラメータを追加
DATA= パラメータと CONTROL= パラメータを両方指定することにより、実施群(DATA= 入力データセット)と対照群(CONTROL= 入力データセット)のターゲット出現率（またはターゲット平均値）の差と各説明変数との関連の大きさを分析します。
(2) ORDER=パラメータを追加
分析結果表における説明変数カテゴリの表示順を制御します。
(3) PCTF=, MEANF= および AICF=パラメータを追加
統計量の表示フォーマットを指定します。
(4) &_XSEL、&_XDEL マクロ変数を出力
目的変数と関連があると判定された説明変数項目をグローバルマクロ変数 &_XSEL、関連が無いと判定された説明変数項目を &_XDEL にそれぞれ出力します。これらは同じ SAS セッションまたは WPS セッション内で、続いてツリーモデルを作成するとき説明変数指定を容易にするために用いることができます。

DMT_TREE

…… (1) CONTROL= パラメータを追加
DATA= パラメータと CONTROL= パラメータを両方指定することにより、実施群(DATA= 入力データセット)と対照群(CONTROL= 入力データセット)のターゲット出現率（またはターゲット平均値）の差のばらつきを説明変数ごとに AIC 値で評価した値を分割基準としたアップリフトモデルを作成します。

DMT_CROSSPLOT

…… (1) ORDER=パラメータを追加
分析結果図における説明変数カテゴリの表示順を制御します。
(2) NOLABEL= パラメータを追加
変数ラベルと文字変数値に定義されているフォーマットの使用を中止し、変数名と文字変数値をそのまま表示します。
※ 現行の WPS ではグラフィック上に日本語表示を行うことができないため、日本語ラベルや日本語フォーマットを定義している場合は NOLABEL=Y を指定します。
(3) PCTF=, MEANF= および AICF=パラメータを追加
統計量の表示フォーマットを指定します。
(4) GRAPH_LANGUAGE= パラメータを追加
グラフ出力言語を制御します。 ※ SAS では GRAPH_LANGUAGE=JAPANESE に設定可能です。WPS ではグラフ上に日本語表示ができないため、デフォルトの GRAPH_LANGUAGE=ENGLISH のままにしてください。

DMT_CROSSTAB

…… (1) ORDER=パラメータを追加
分析結果表における説明変数カテゴリの表示順を制御します。
(2) NOLABEL= パラメータを追加
変数ラベルと文字変数値に定義されているフォーマットの使用を中止し、変数名と文字変数値をそのまま表示します。
(3) PCTF=, MEANF= および AICF=パラメータを追加
統計量の表示フォーマットを指定します。

DMT_DATASAMP

…… (1) TESTRATE= パラメータを追加
テスト用データの抽出率の方を指定できるようにしました。

DMT_TREECUT

…… (1) TEST= パラメータを追加
TEST= には、モデルをテストデータに当てはめたときのモデル形式データセットを指定します。MODEL と TEST の中間ノードを比較して、その子ノード間のターゲット値の大きさの順が逆転している中間ノードを自動枝刈りしたモデルデータセットを作成します。

DMT_TREESCORE

…… (1) CONTROL= パラメータを追加
DATA= パラメータと CONTROL= パラメータを両方指定することにより、アップリフトモデルを新たな実施データ群(DATA=入力データセット)と対照データ群(CONTROL=入力データセット)に同時に当てはめた場合のモデル形式データセットを作成します。

DMT_TREETAB, DMT_NODETAB

…… (1) DETAIL= パラメータを追加
ノード統計量の表示項目数を制御します。DETAIL=Y と指定すると表示項目が増えます。
(2) PCTF=, MEANF= パラメータ
統計量の表示フォーマットを指定します。

DMT_GAINCHART , DMT_COMPAREPLOT

…… (1) GROUPNUM= パラメータを追加
モデル予測値の順に指定数のグループにデータを等件数分割した上で、各グループのモデル予測値と実際値に基づくプロット図を作成します。
(2) GROUPVAR= パラメータを追加
指定のグループ変数のカテゴリをグループ単位としたモデル予測値と実際値に基づくプロット図を作成します。

※ DMT デシジョンツリーVer1.2 の GROUPNODE=Y パラメータは廃止しました。代わりに、GROUPVAR=_NODE を使用します。

(3) AR_ROCF=, AMOUNTF=, R2F=, RMSEF= パラメータ

統計量の表示フォーマットを指定します。

(4) GRAPH_LANGUAGE= パラメータを追加

グラフ出力言語を制御します。

［入力データセットのデータセットオプション］

DATA=パラメータ、CONTROL=パラメータに指定する入力データセット名の後に、任意のデータセットオプションを指定可能にしました。

以下のマクロで利用可能です。

DMT_CROSS , DMT_DATASAMP , DMT_TREE, DMT_CVTREE, DMT_TREESCORE, DMT_GAINCHART, DMT_COMPAREPLOT, DMT_CORRECTTAB, DMT_UPLIFTCHART

1.8 無償提供評価版の制限

お客様が機能を評価することを目的として、無償提供する評価版の DMT デシジョンツリーVer.1.3 は、マクロカタログの一部のマクロ（DMT_TREE、DMT_CVTREE）に対して、入力データセット(data=データセットと control=データセット)のオブザベーション数に最大 **2,000 件まで**の制限を与えています。(※ この制限は where 条件指定とは無関係です。データセットに含まれる全件数で判断しています。)

他のマクロ（DMT_CROSS, DMT_CROSSTAB, DMT_CROSSPLOT, DMT_DATASAMP, DMT_TREETAB , DMT_NODETAB, DMT_TREECUT, DMT_TREEADD , DMT_TREESCORE , DMT_GAINCHART , DMT_COMPAREPLOT, DMT_CORRECTTAB, DMT_UPLIFTCHART）には無償提供評価版でもこの制限はありません。

1.9 処理するデータ件数を無制限とするアップグレードについて（有償）

有償で提供するライセンスコードを入力すると、マクロカタログの中にあるすべてのマクロを入力データセットの件数に制限おが無いものにアップグレードできます。なお、ライセンスコードは使用中の SAS または WPS のサイト番号でのみ有効です。料金、手続き等についてはお問い合わせください。

2. 導入方法

動作環境を確認の上、.以下のステップに従って導入してくださ い。

2.1 GUI 実行モードのセットアップ方法

2.1.1 ファイルのコピー

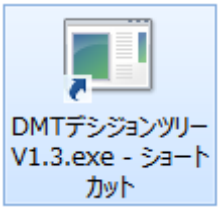
まず、弊社ウェブサイト（http://www.dataminetech.co.jp）からダウンロードしたプロダクトファイル（DMT_TREEV1.3_buildyyyymmdd.zip）（yyyymmdd には年月日が入ります）を任意の読み書き可能なユーザディレクトリにコピーし、そのディレクトリで解凍します。**ただし、ディレクトリパス名はすべて半角英数字のみで指定可能でなければならない点に注意してください。**

例えば、ユーザディレクトリ "C:¥Users¥ユーザプロファイル名"（ユーザプロファイル名の箇所は Windows ログインユーザ名）の中に DMT_TREEV1.3_buildyyyymmdd.zip ファイルをコピーしてその場所に解凍します。"DMT_TREEV1.3_buildyyyymmdd" という名前のディレクトリが生成され、その中に DMT デシジョンツリーV1.3.exe という名前のファイルが入っていることを確認します。

※ SAS ショートカット追記用_INITSTMT.txt, WPS ワークベンチ起動設定用_INITSTMT.txt, DMT_TREEV1.3_SAMPLERUN.sas の 3 ファイルはコマンド実行モード設定用のファイルです。

2.1.2 ショートカットの作成

"DMT デシジョンツリーV1.3.exe" のショートカットをデスクトップに作成します。



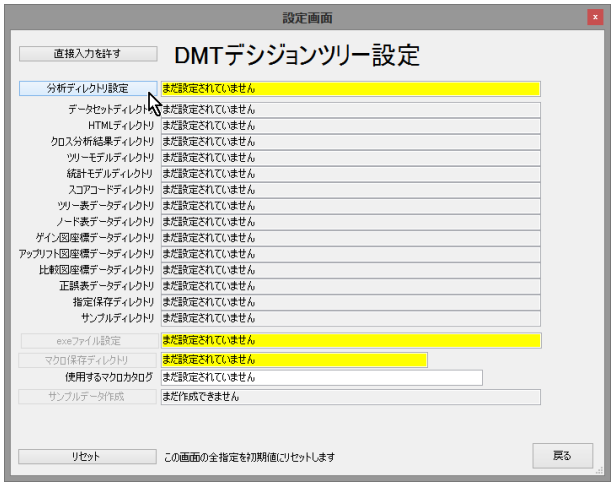
ショートカットをダブルクリックして、以下の**「メニュー」**画面が表示されることを確認します。

上記**「メニュー」**画面が表示されず、エラーが表示される場合は、GUI 画面の起動に必要なバージョンの.NET Framework 4.x が Windows マシンにインストールされていないことが原因の場合があります。マイクロソフト社のサイトから最新の .NET Framework 4.x を取得し、コンピュータにインストールした後、再度**「メニュー」**画面を起動してください。

2.1.3 初期設定

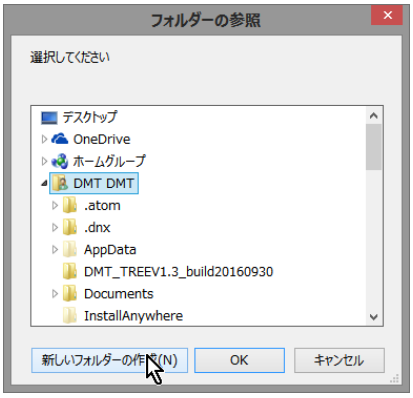
はじめて**「メニュー」**画面を起動した場合は、初期設定が必要 ボタンをクリックして**「DMT デシジョンツリー設定」**画面を開きます。

(1) 分析ディレクトリの設定

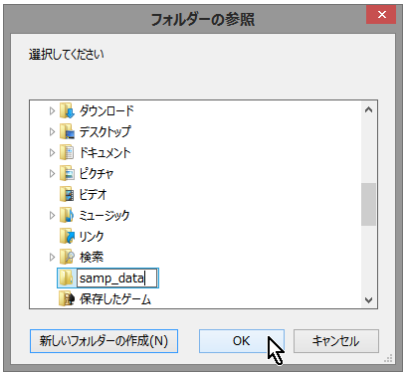


分析ディレクトリ設定 を押して、アプリケーションで使用するデータ、モデルデータ、HTML 出力、パラメータなどを保存するファイルの分析ルートディレクトリを指定します[3]。デフォルトではデスクトップを初期ディレクトリとして **フォルダーの参照** 画面が開きます。

ユーザディレクトリの下のデスクトップ（c:¥users¥ユーザプロファイル名¥desktop）やドキュメント（c:¥users¥ユーザプロファイル名¥documents）に分析ディレクトリを作成してもかまいませんが、ここでは、ユーザディレクトリを選択しておいて、新しいフォルダの作成(M) を押し、ユーザディレクトリの下に新しいフォルダを作成します。

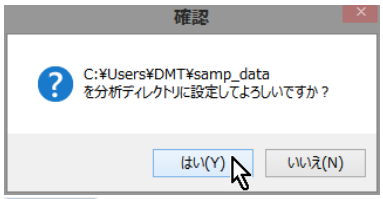


新しいフォルダに半角文字で名前をつけます。
（**重要な注意**）フォルダのパス名もすべて半角文字でなければなりません。全角文字が含まれる場合、エラーメッセージが出現しますので、指定し直してください）

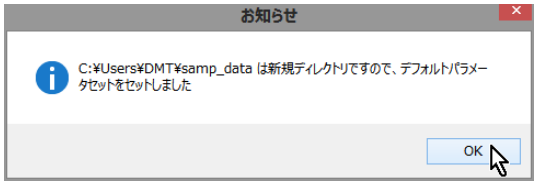


samp_data と名前をつけた後、OK を押します。

[3] メニュー操作によりディレクトリ設定ができない場合は、代替手段として、直接入力を許す を押し、ディレクトリパスをキーボード入力してから、設定 を押します。

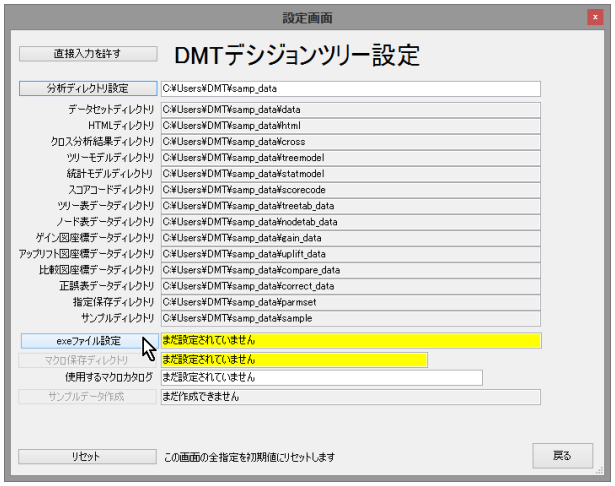


はい(Y) を押します。



OK を押して**「DMT デシジョンツリー設定」**画面に戻ります。

(2) exe ファイル設定

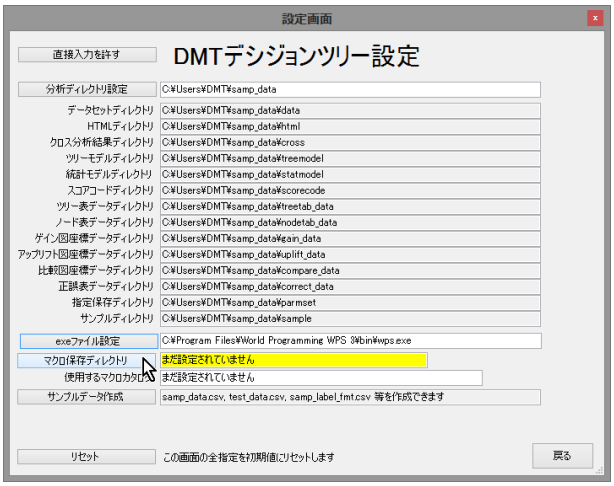


次に、exeファイル設定 ボタンを押して、導入されている SAS または WPS の実行ファイル（sas. exe または wps. exe）のパスを指定します。ファイル選択画面が C:¥Program Files ディレクトリを初期ディレクトリとして開きます。

通常、sas.exe ファイルは、C:¥ Program Files¥SASHome¥SASFoundation¥9.x¥sas.exe（ここで、9.x は SAS バージョンを表します）にあり、wps.exe は、C:¥Program Files¥World Programming WPS 3¥bin¥wps.exe にあります。ファイル選択画面のディレクトリパスを辿って指定します。ただし、インストール時の設定によって、実際の exe ファイルのパスは異なる場合もあります。

指定後、実行ファイルのバージョンチェックが行われ、完了すると exeファイル設定 ボタンの表示が exeファイル変更 に変化し、マクロ保存ディレクトリ ボタンが有効になります。

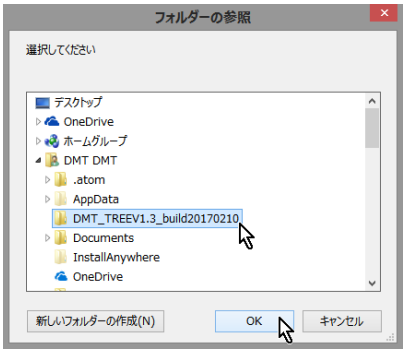
(3) マクロカタログファイル保存ディレクトリ設定



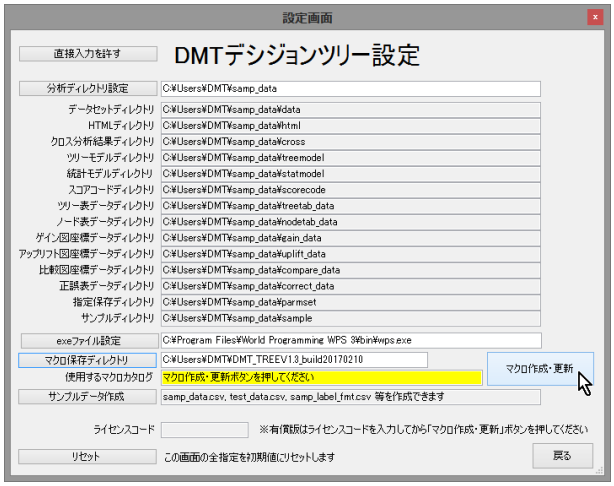
マクロ保存ディレクトリ ボタンを押して、**DMT デシジョンツリーマクロカタログ** を保存するディレクトリを、すべて半角英数字のみのパス名で指定します。

ディレクトリ選択画面が、**C:¥users¥ユーザプロファイル名** を初期ディレクトリとして開きます。

ここでは、**C:¥users¥ユーザプロファイル名¥DMT_TREEV1.3_buildyyyymmdd** ディレクトリを保存先ディレクトリに設定します。

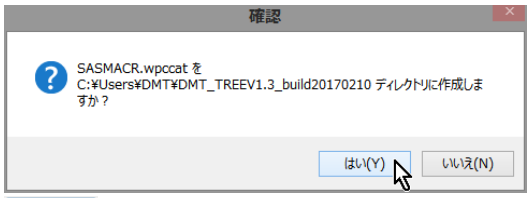


(4) マクロカタログの作成



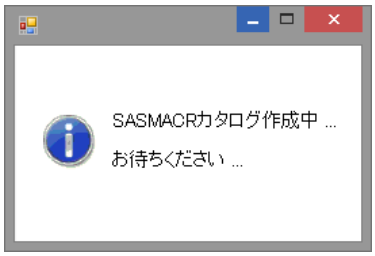
次に　マクロ作成・更新　ボタンを押します。

マクロカタログを作成するかどうかの確認画面が表示されます。



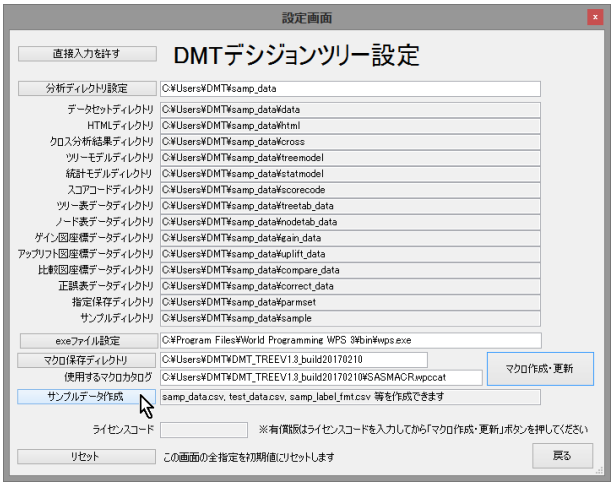
はい(Y)　を押します。

マクロカタログ作成中... のメッセージが出現します。



しばらくすると、終了のメッセージが表示されます。(※ マクロカタログの生成にはコンピュータ環境によっては、時間がかかる場合があります。)

OK　を押します。

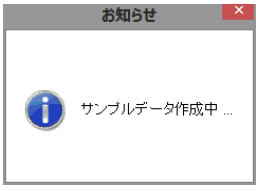


以上でマクロカタログの作成は終了です。

(5) サンプルデータの作成

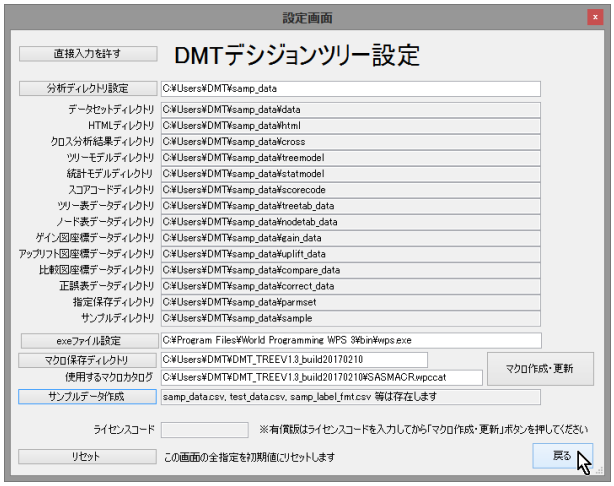
最後に、　サンプルデータの作成　を押し、**サンプルデータ**（CSV 形式、および、WPS データセット形式または SAS データセット形式）、**サンプルラベルフォーマット定義**（CSV 形式と SAS コード形式）を SAMPLE ディレクトリに作成しておきます。

はい(Y)　を押します。



サンプルデータ作成中のメッセージ画面が出現し、作成が終了すると画面は自動的に閉じます。

サンプルデータなどが生成されたというメッセージが出現します。　OK　を押して元の画面に戻ります。



以上で、GUI 実行モードの初期設定は終了です。

戻る ボタンを押して、**「メニュー」**画面に戻ります。

※ 途中で問題が起きた場合は、リセット を押して設定を最初からやり直してください。

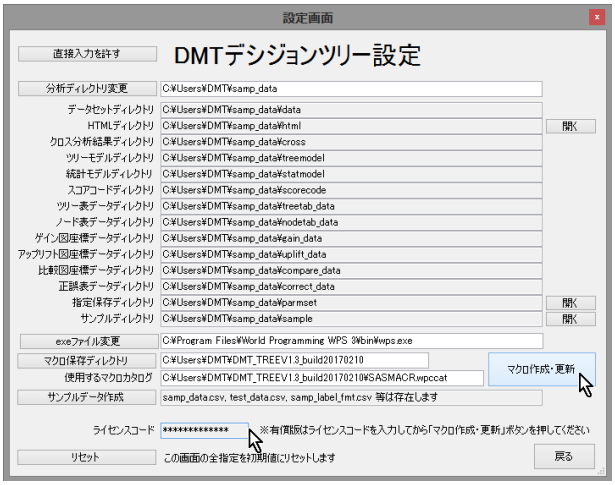
「メニュー」画面が各分析画面を呼び出す画面項目が選択できる状態になっていることを確認すると、GUI 実行モードの初期設定は完了です。

2.1.4 マクロカタログの更新方法

マクロカタログまたは GUI 実行アプリケーションの修正版がリリースされた場合は、"DMT デシジョンツリーV1.3.exe" の最新ビルドを含む DMT デシジョンツリーV1.3.zip が、弊社インターネットウェブサイト（http://www.dataminetech.co.jp）からダウンロード可能になります。そのときは、最新版をダウンロード、解凍し、設定画面から、既存の分析ディレクトリ設定、exe ファイル設定、マクロ保存ディレクトリ設定を行った上で、マクロ作成・更新 ボタンを押して、マクロカタログを最新版に更新してください。

2.1.5 有償版へのアップグレード方法

処理するオブザベーション件数に制限のない DMT デシジョンツリーV1.3 マクロカタログを生成または更新するには、設定画面の右下にある **ライセンスコード** 欄にライセンスコード（有償）を入力してから マクロ作成・更新 ボタンを押してください。



マクロ作成・更新完了メッセージが以下のように表示されることを確認してください。

ここから分析に進むこともできますが、 ボタンを押して一旦**「メニュー」**画面を終了し、コマンド実行モードのセットアップを行っておきましょう。

なお、メニュー画面を閉じて DMT デシジョンツリーを終了する際には、現在の設定を自動保存するかどうかがシステムから質問されます。

「はい」と答えると、現在の設定情報や分析画面のパラメータ指定全体が _LASTSAVE_ という名前で分析ディレクトリの parmset サブディレクトリ内に保存されます。（既存のものが存在する場合は上書き）
「いいえ」と答えると、分析ディレクトリに保存されている _LASTSAVE_ が維持されます。

ただし、「はい」と答える場合でも「いいえ」と答える場合でも、次回メニュー画面を起動したときは、前回の終了時点のパラメータ指定全体が残っています。_LASTSAVE_ の状態に戻すには、保存指定のロード を押して _LASTSAVE_ を選択します。
（※ パラメータのロード・保存 の項を参照）

2.2 SAS のコマンド実行モードのセットアップ方法

SAS ディスプレーマネージャのプログラムエディタに、マクロコマンドを入力しサブミットする方式で利用するモード(コマンド実行モード）の設定方法は以下のとおりです。

2.2.1 ファイルのコピー

設定に使用するモジュールは、GUI 実行モードの設定画面で作成した sasmacr.sas7bcat、SAS ショートカット追記用_INITSTMT.txt、そして DMT_TREE_VER1.3_SAMPLERUN.sas の 3 つです。

sasmacr.sas7bcat は GUI 実行モード用と同じものをコマンド実行用に SASUSER ディレクトリ（一般的には、c:¥users¥ユーザプロファイル名¥documents¥My SAS File¥9.x）に複製して用います。
ただし、SASUSER ライブラリの中に既存の sasmacr.sas7bcat ファイルが存在していた場合、置き換えるか、追加するかの選択があり、SASUSER ディレクトリの内容をまず確認します。

Windows エクスプローラで確認を行った結果、sasmacr.sas7bcat が存在しない、または DMT デシジョンツリーの旧バージョンが同じ名前で存在する場合は、新しい sasmacr.sas7bcat をコピーし既存のものがあれば上書きします。

＜参考＞
SASUSER ディレクトリ内に既に同じ名前の SAS マクロカタログが存在し、内容は残して、追加したい場合は、ファイル全体を置き換えてしまわないように、SAS を立ち上げてから、プログラム編集画面に以下のコマンドを入力しサブミットしてください。（このやり方により、新しい名前のマクロカタログエンテティは追加され、既存と同じ名前のマクロは更新されます。）

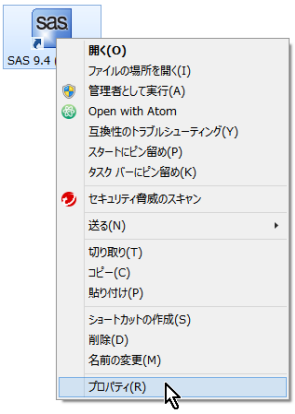
[C:¥users¥ユーザプロファイル名¥DMT_TREEV1.3_buildyyyymmdd ディレクトリに保存してある sasmacr.sas7bcat を SASUSER ディレクトリにコピー(新規は追加、既存は置換)する場合のコマンド]

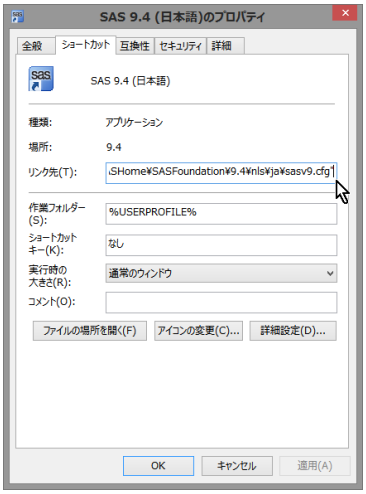
```
libname in " C:¥Users¥ユーザプロファイル名
¥DMT_TREEV1.3_buildyyyymmdd ¥DMT_TREEV1.3";
proc catalog cat=in.sasmacr;
   copy out=sasuser.sasmacr;
run;quit;
```

ここでユーザプロファイル名の個所には Windows のログインユーザ名としてください。

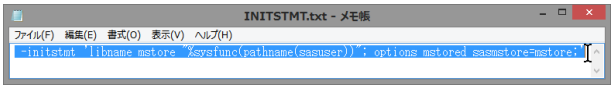
2.2.2 初期設定

sasmacr.sas7bcat の SASUSER ディレクトリへのコピーに続いて、コンパイル済みマクロカタログを SAS で利用可能にするため、SAS 起動ショートカットにオプションを設定します。
SAS 起動ショートカットを右クリックしてプロパティを開きます。

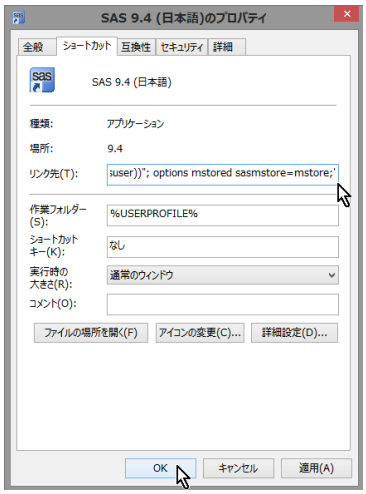


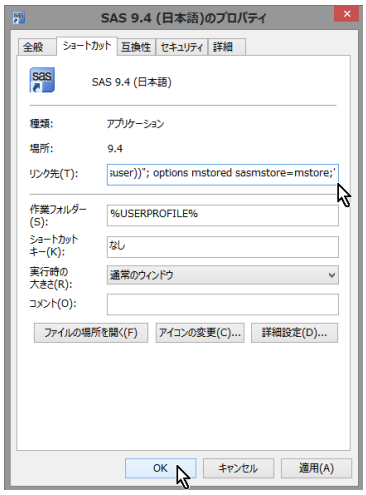


プロパティ画面のリンク先のテキストの末尾にカーソルを移動しておき、SAS ショートカット追記用_INITSTMT.txt ファイルを開きます。表示されるテキストをコピーし、リンク先のテキストの末尾に貼り付けます。



(注意) 先頭のスペース 1 個も忘れずに複写してください。





貼り付けた後、OK を押してショートカットのプロパティを閉じて、一旦 SAS を終了します。

次に、SAS 起動ショートカットをダブルクリックして SAS ディスプレイマネージャを起動します。
起動時のログ画面に、以下のようなメッセージが表示されていることを確認します。

```
NOTE: ライブラリ参照名 MSTORE は SASUSER と同じ物理ライブラリを参照しています。
NOTE: ライブラリ参照名 MSTORE を次のように割り当てました。
      エンジン: V9
      物理名: G:¥Users¥DMT¥Documents¥My SAS Files¥9.4
```

プログラム編集画面に、以下のように入力してサブミットしてください。

```
%dmt_tree(help)
```

ログに DMT_TREE マクロの指定方法が表示されることを確認してください。

2.2.1 サンプルプログラムの実行

次にサンプルプログラム DMT_TREE_VER1.3_SAMPLERUN.sas をプログラム編集画面にコピーして、サブミットしてください。エラーなく実行できることをログと HTML 出力で確認してください。
実行内容については、次の「実行例」でほぼ同じ内容を説明しています。

2.3 WPS のコマンド実行モードセットアップ方法

WPS ワークベンチのプログラムエディタビューに、マクロコマンドを入力しサブミットする方式で利用するモード（コマンド実行モード）の設定方法を説明します。

2.3.1 ファイルのコピー

設定に使用するファイルは GUI 実行モードの設定画面で作成した SASMACR.wpccat、WPS ワークベンチ起動設定用 _INITSTMT.txt、そして DMT_TREE_VER1.3_SAMPLERUN.sas の 3 つです。

SASMACR.wpccat は GUI 実行モード用と同じものをコマンド実行用に SASUSER ディレクトリ（一般的には、c:¥users¥ユーザプロファイル名¥documents¥My WPS File）に複製して用います。
ただし、SASUSER ライブラリの中に既存の SASMACR.wpccat ファイルが存在していた場合、置き換えるか、追加するかの選択がありますので、SASUSER ディレクトリの内容をまず、確認します。

Windows エクスプローラで確認を行った結果、SASMACR.wpccat が存在しない、または DMT デシジョンツリーの旧バージョンが同じ名前で存在する場合は、新しい SASMACR.wpccat をコピーし既存のものがあれば上書きします。

＜参考＞
SASUSER ディレクトリ内に既に同じ名前の SAS マクロカタログが存在し、内容は残して、追加したい場合は、ファイル全体を置き換えてしまわないように、WPS を立ち上げてから、エディタビューに以下のコマンドを入力しサブミットしてください。（このやり方により、新しい名前のマクロカタログエンテティは追加され、既存と同じ名前のマクロは更新されます。）

[C:¥users¥ユーザプロファイル名¥DMT_TREEV1.3_buildyyyymmdd ディレクトリに保存してある SASMACR.wpccat を SASUSER ディレクトリにコピー(新規は追

加、既存は置換)する場合のコマンド]

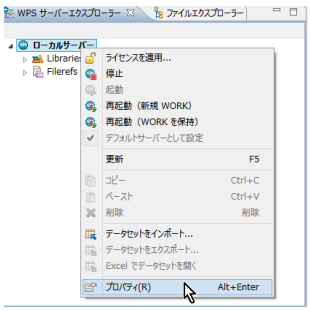
```
libname in " C:¥Users¥ユーザプロファイル名
¥DMT_TREEV1.3_buildyyyymmdd ¥DMT_TREEV1.3";
proc catalog cat=in.sasmacr;
   copy out=sasuser.sasmacr;
run;quit;
```

ここでユーザプロファイル名の個所にはWindows のログインユーザ名としてください。

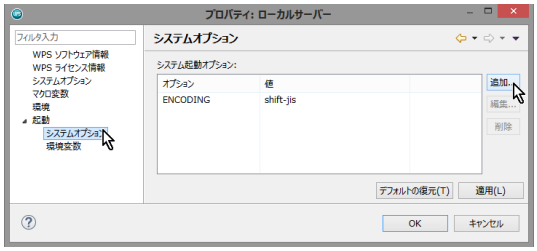
2.3.2 初期設定

SASMACR.wpccat の SASUSER ディレクトリへのコピーに続いて、マクロモジュールを WPS ワークベンチで利用可能にするための設定を行います。

WPS ワークベンチの WPS サーバーエクスプローラービューのローカルサーバー を右クリックし、プロパティを選択します。



プロパティ画面の左領域の 起動 を展開し、システムオプションを選択します。

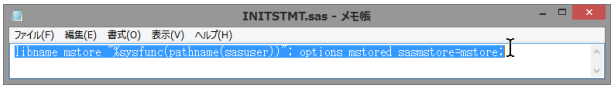


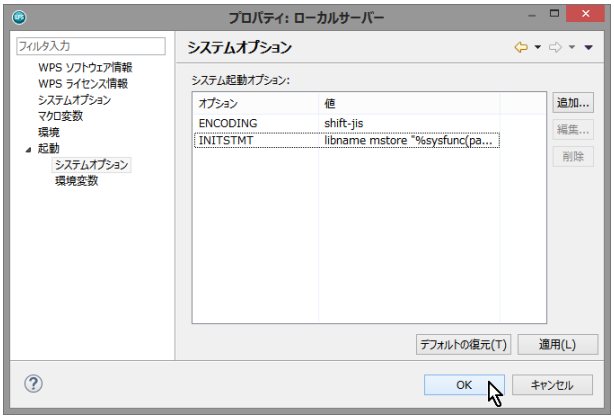
追加... を押します。

起動オプション画面が表示されます。
名前： には INITSTMT を選択し、値： には WPS ワークベンチ起動設定用_INITSTMT.txt ファイルの内容をコピーペーストして

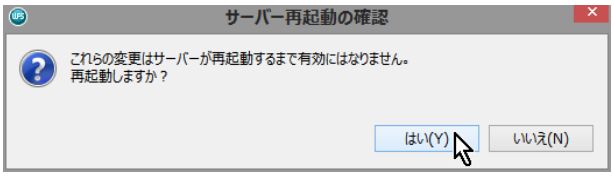
OK を押します。

(WPS ワークベンチ起動設定用_INITSTMT.txt の内容をコピーして値に貼り付ける)





システム起動オプションに INITSTMT が追加されたことを確認してから OK を押します。



オプションを有効にするため、再起動します。
その後、一旦 WPS ワークベンチをを終了します。

WPS ワークベンチを起動します。
起動時のログ画面に、以下のようなメッセージが表示されていることを確認します。

```
NOTE: Library mstore assigned as follows:
      Engine:        WPD
      Physical Name: G:¥Users¥DMT¥Documents¥My WPS Files
```

プログラム編集画面に、以下のように入力してサブミットしてください。

```
%dmt_tree(help)
```

ログに DMT_TREE マクロの指定方法が表示されることを確認します。

2.3.3 サンプルプログラムの実行

次にサンプルプログラム DMT_TREE_VER1.3_SAMPLERUN.sas をプログラム編集画面にコピーして、サブミットしてください。エラーなく実行できることをログと HTML 出力で確認してください。
実行内容については、次の「実行例」でほぼ同じ内容を説明しています。

3. 実行例

DMT_TREEV1.3_SAMPLERUN.sas プログラムは、それぞれ 2,000 件のオブザベーション数、12 項目の変数を持つサンプルデータ(samp_data)とテストデータ(test_data)を作成し、これらのデータを用いたDMTデシジョンツリーアプリケーションの使い方を例示します。

samp_data, test_data 項目

#	項目	ラベル	値	値のラベル
1	sei	性別	1	男性
			2	女性
2	nenrei	年齢	(数値)	
3	jukyo	住居区分	1	持家(自己所有)
			2	持家(家族所有)
			3	賃貸マンション
			4	借家
			5	アパート
			6	寮
			7	社宅
			欠損	不明
4	kazoku_kosei	家族構成	1	独身同居家族あり
			2	独身単身
			3	既婚子供あり
			4	既婚子供なし
			5	独身子供あり
			欠損	不明
5	gakureki	学歴	1	中学
			2	高校
			3	専門学校
			4	大学
			5	大学院
			欠損	不明
6	kinmusaki	勤務先	A	企業
			B	自営(法人)
			C	自営(個人)
			D	官公庁
			欠損	不明
7	gyoshu	業種	A	農林水産
			B	鉱業
			C	建設・土木業
			D	製造
			E	電気・ガス・水道
			F	運輸・通信
			G	卸売・小売
			H	金融・保険
			I	不動産
			J	ホテル・飲食
			K	医療・福祉
			L	その他サービス
			M	公務
			欠損	不明
8	shokushu	職種	1	営業
			2	販売
			3	経営・管理
			4	作業・清掃
			5	オペレータ・運転手
			6	事務
			7	技術・サポート
			欠損	不明
9	nenshu	年収	(数値)	
10	DM	DM送付有無フラグ	0	なし
			1	あり
11	flg	購入有無フラグ	0	なし
			1	あり
12	kingaku	購入金額	(数値)	

これらは、ある物品販売会社の 4000 件の会員データを表すものとします。（ここでは、便宜的に、あらかじめ 2000 件のオブザベーションを持つ samp_data と test_data にランダムに分けておき、samp_data を用いてモデル作成を行うようにしています。）

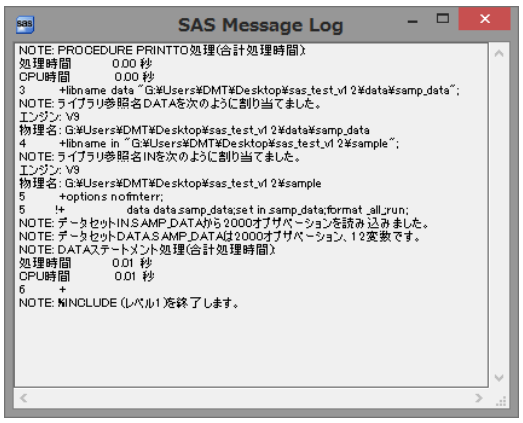
12 個の項目の中の最初の 9 項目は会員の属性項目（登録情報）であり、10 項目目の DM は会社の行動（直前のプロモーション）を表す変数です。11～12 項目目の flg と kingaku はそれぞれプロモーション実施後の一定期間内の会員の応答（購入有無と購入金額）を表しているものとします。

分析の目的は、購入率の分布を説明する顧客属性の組合せや、プロモーションの効果を評価することとします。以下、DMT デシジョンツリーV1.3 の利用方法を簡単に説明する目的で、以下の２種類の分析を行う手順と実行結果の一部を表示します。

（実行例１）優良顧客の判別
（実行例２）施策実施効果の分析

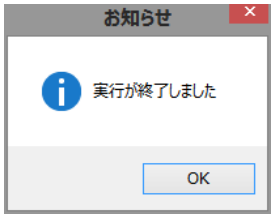
WPS での分析例を表示していますが、SAS では以下の注意と、グラフィック表示が日本語となっている点以外同じです。

(SAS での注意) SAS では、 実行 を押した後、「SAS Message Log」画面が出現し、実行ログが表示されます。



「NOTE: %INCLUDE(レベル 1)を終了します。」というメッセージが実行ログの最後に出現すれば実行は終了です。

を押して「SAS Message Log」画面を閉じると、以下の「お知らせ」画面が表示されます。



OK を押すと、分析画面で次の操作が可能になります。

3.1 （例1）優良顧客の判別

顧客の属性組合せによって、購入確率が高い優良顧客と購入確率が低い顧客を区別するための属性判別ルールを作成します。目的変数はクラス変数 flg、購入確率を求めたいクラスは flg=1（購入あり）です。

モデルは DM 送付有無フラグ別に作成することも考えられますが、ここでは DM 送付有無フラグは説明変数の 1 つとして用いることとし、変数 KINGAKU は説明変数から削除します。

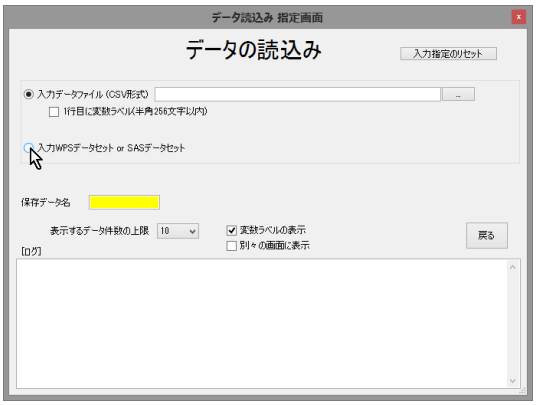
以下の分析手順を実行します。

3.1.1 データ読込

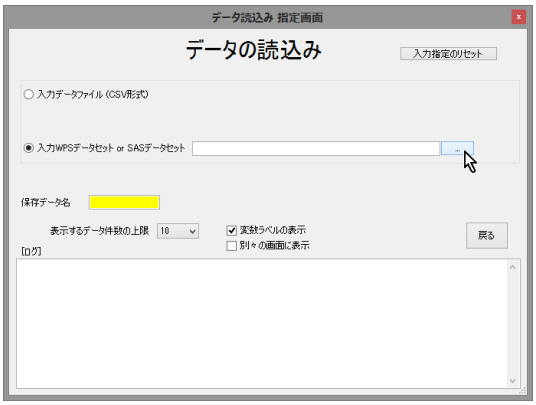
まず、本システムで分析を行うため、分析データ（samp_data と test_data）を data ディレクトリに読込みます。、

データ読込 をクリックします。

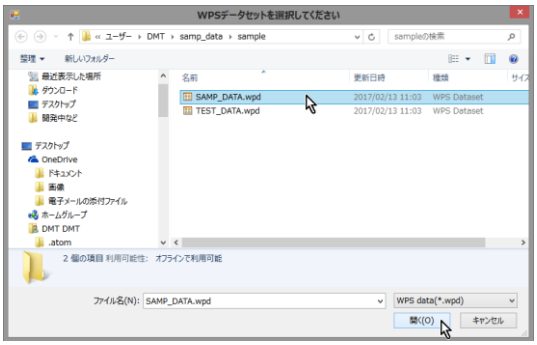
「データの読込み」 画面に切り替わります。



「入力WPS データセット or SAS データセット」ラジオボタンをクリックします。



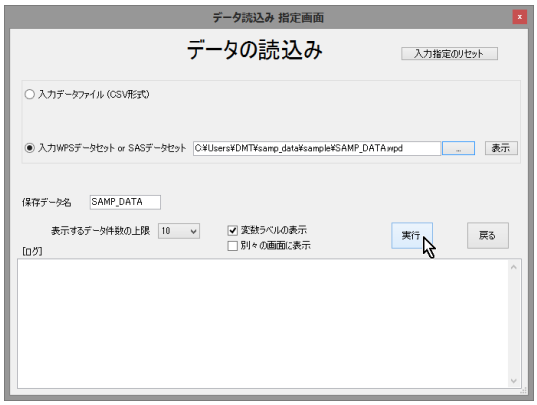
... ボタンを押し、入力 WPS データセットを選択します。



サンプル WPS データセットの選択画面が表示されます。

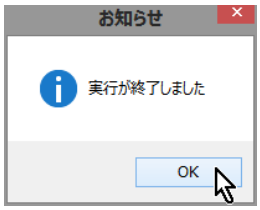
SAMP_DATA.wpd を選択し、 開く(O) を押します。

（Windows の設定により、.wpd のファイル拡張子部分は表示されない場合があります）



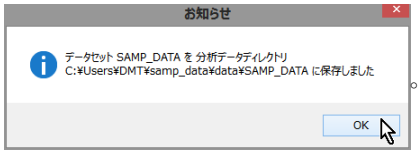
テキストボックスに読み取る WPS データセットファイル名がフルパスで表示されます。また、**保存データ名** に同じ名前が自動入力されます。

実行 を押します。

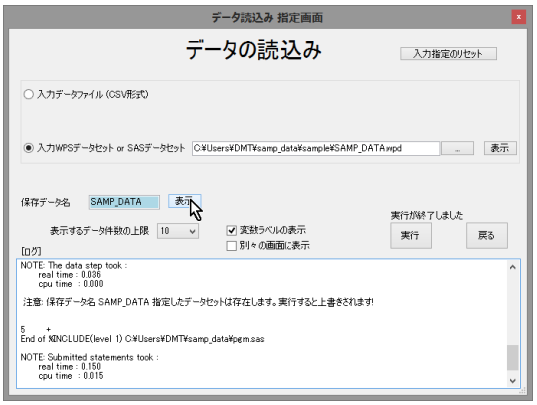


[**ログ**]に WPS 実行ログが表示され、**「実行終了」**のメッセージ画面が表示されます。

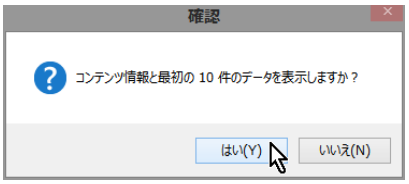
OK を押し、**「実行終了」**メッセージ画面を閉じます



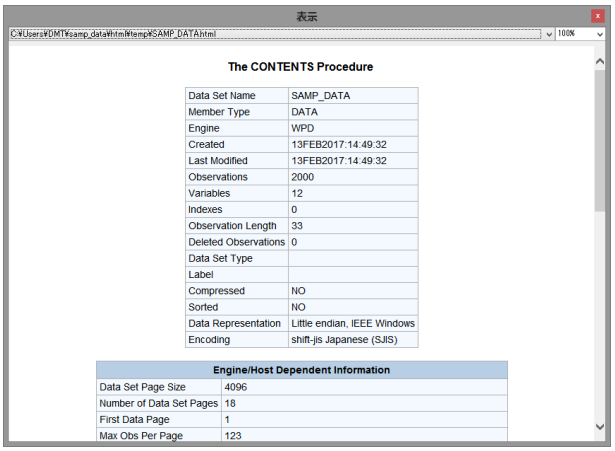
WPS データセット SAMP_DATA を保存したというメッセージが表示されます。 OK を押します。



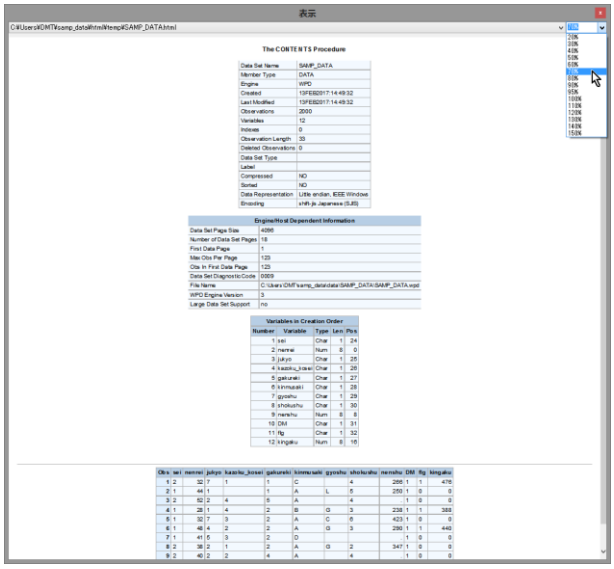
表示 を押して SAMP_DATA の内容を確認します。



はい(Y) を押します。

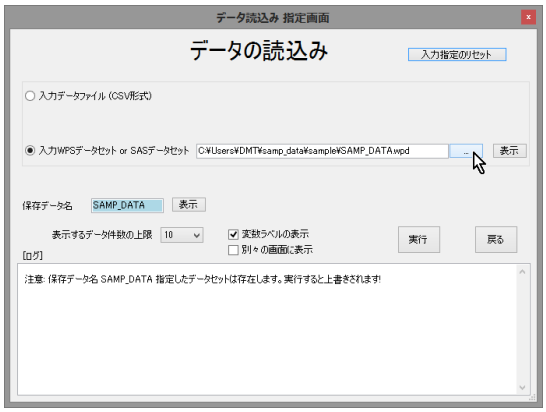


オブザベーション数が 2000 であることを確認します。「表示」バーをダブルクリックし、コンボボックスの「100%」を「60%」に変更して全体を表示してみます。

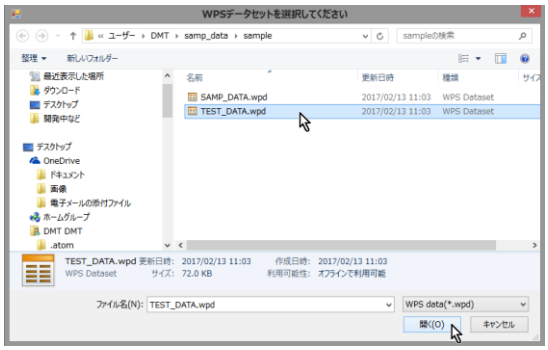


samp_data データセットのコンテンツ情報と**「DMT デシジョンツリー設定」**画面の「表示するデータ件数の上限」で設定してあるオブザベーション数のデータ値が表示されます。

ボタンを押してデータ表示を終了し、**「データの読込み」**画面に戻ります。



同様に、TEST_DATA を読み込みます。



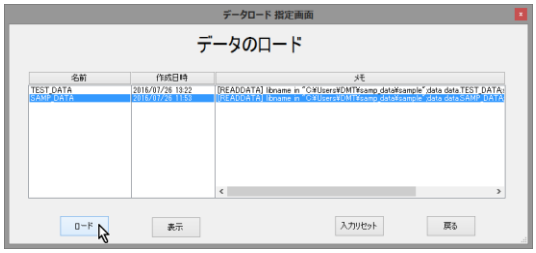
戻る　ボタンを押し **「データの読込み」**画面を終了し、**「メニュー」**画面　に戻ります。

3.1.2 ラベル付与

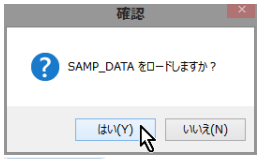
結果を見やすくするために変数と文字変数値にラベルを付けます。

ラベル付与 を押すと、**「分析データに変数ラベルと値ラベルを定義」** 画面に切り替わります。

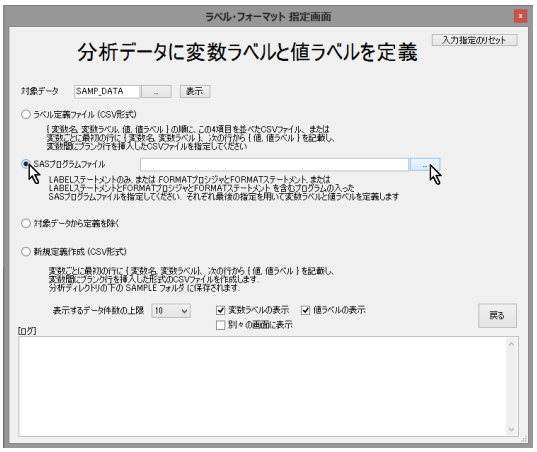
対象データ ... を選択します。



SAMP_DATA を選択し、ロード を押します。

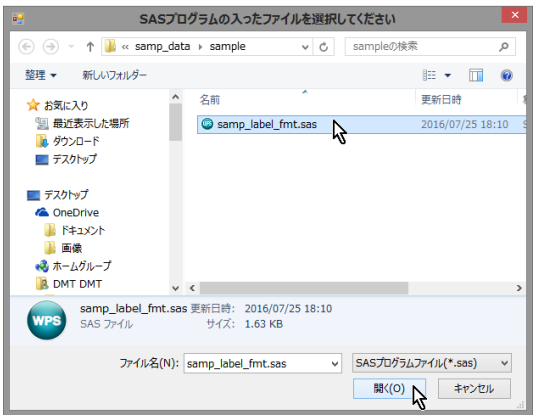


はい(Y) を押します。

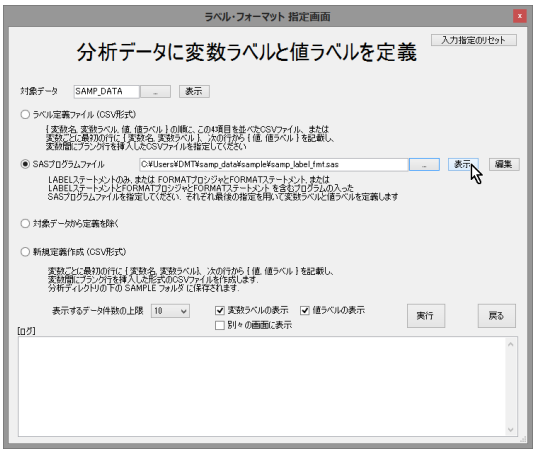


SAS プログラムファイル に切り替えます。

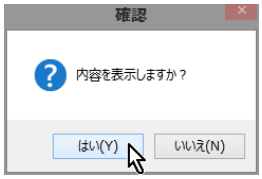
... を押します。



samp_label_fmt.sas を選択し、開く(O) を押します。

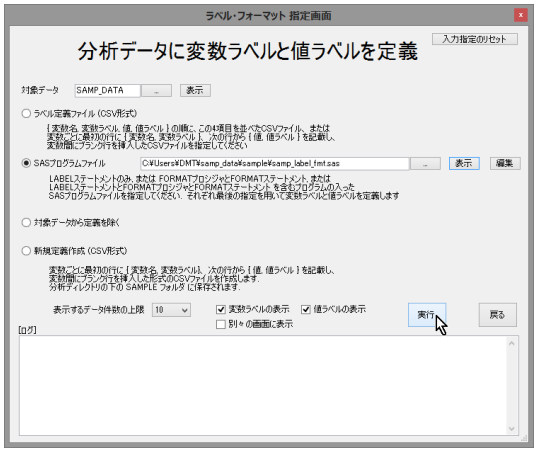


テキストボックスにsamp_label_fmt.sas ファイルのフルパスが表示されます。表示 を押して samp_label_fmt.sas の内容を確認します。



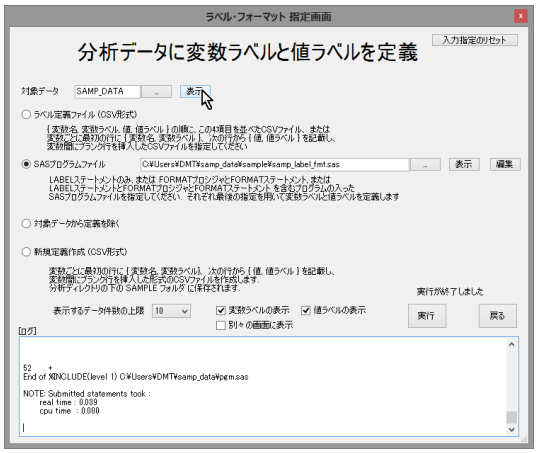
↓

SAS 言語の LABEL 文、PROC FORMAT 文、FORMAT ステートメントにより変数名、文字変数値にフォーマットを定義しているコードが表示されます。 ボタンを押してコード表示を終了します。

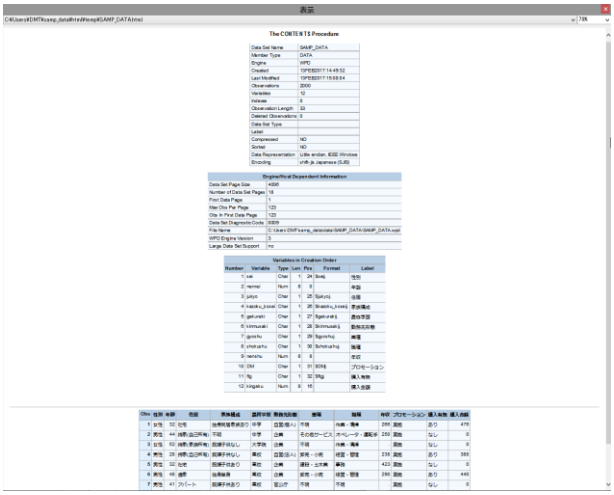


実行 を押します。

※ これ以降は、煩雑さを避けるため、実行後に出現する「ログ」画面、「実行完了確認画面」などの表示は基本的に省略します。



表示 を押して SAMP_DATA の内容を確認します。



変数にラベルが定義され、文字変数値にフォーマットが適用された表示になっていることを確認します。

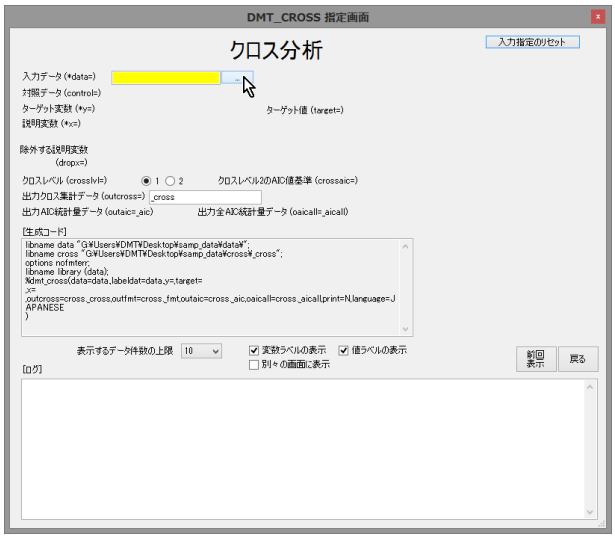
※ ここでは、TEST_DATA には変数ラベルと値ラベルの定義は故意に行わないことにします。

表示画面（ブラウザ）を閉じ、**「分析データに変数ラベルと値ラベルを定義 」**画面 を閉じて、**「メニュー」**画面 に戻ります。

3.1.3 項目分析

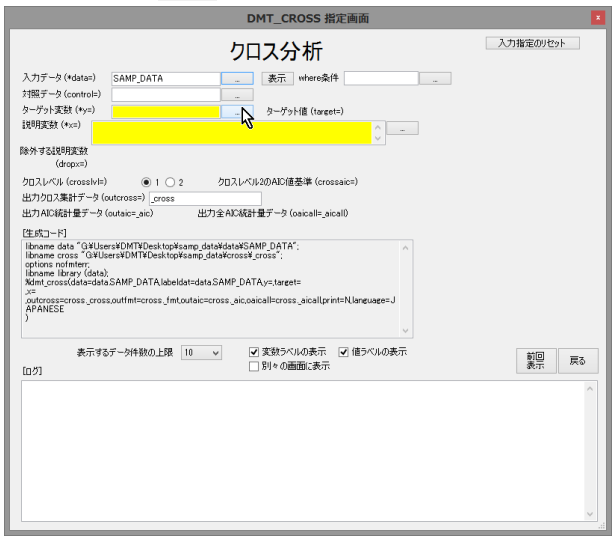
デシジョンツリーモデル作成前の事前分析として、説明変数とターゲット変数との関連性や説明変数分布の把握を行います。

クロス分析 を押すと、ターゲット変数と各説明変数間の関連分析を行う**「クロス分析」**画面が開きます。

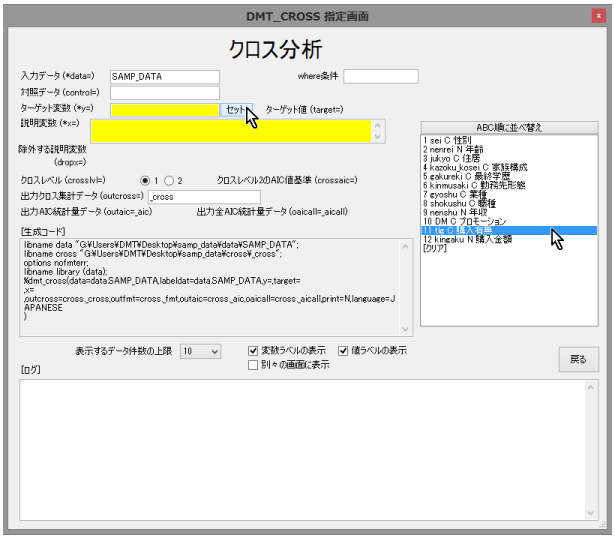


入力データ [...] を押し、**入力データ** として SAMP_DATA を選択し、ロードします。

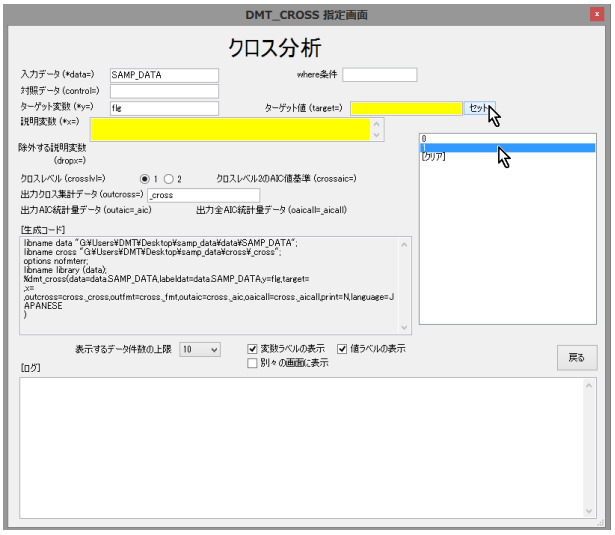
ターゲット変数 [...] を押します。



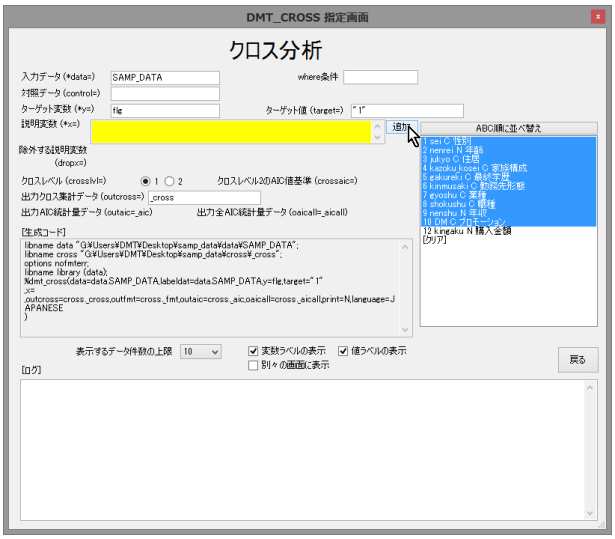
リストから **flg** を選択して [セット] を押します。



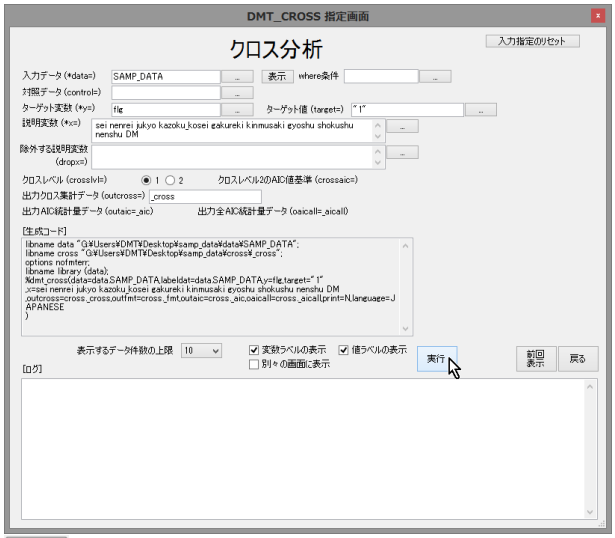
同様に、**ターゲット値** は **"1"** を選択します。



説明変数 は sei から DM までの 10 個の変数を選択し、[追加] を押します。

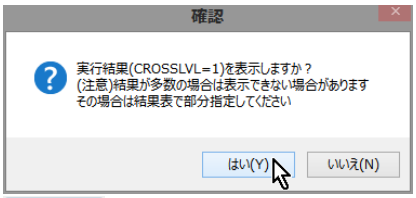


必須指定が完了すると 実行 ボタンが出現します。

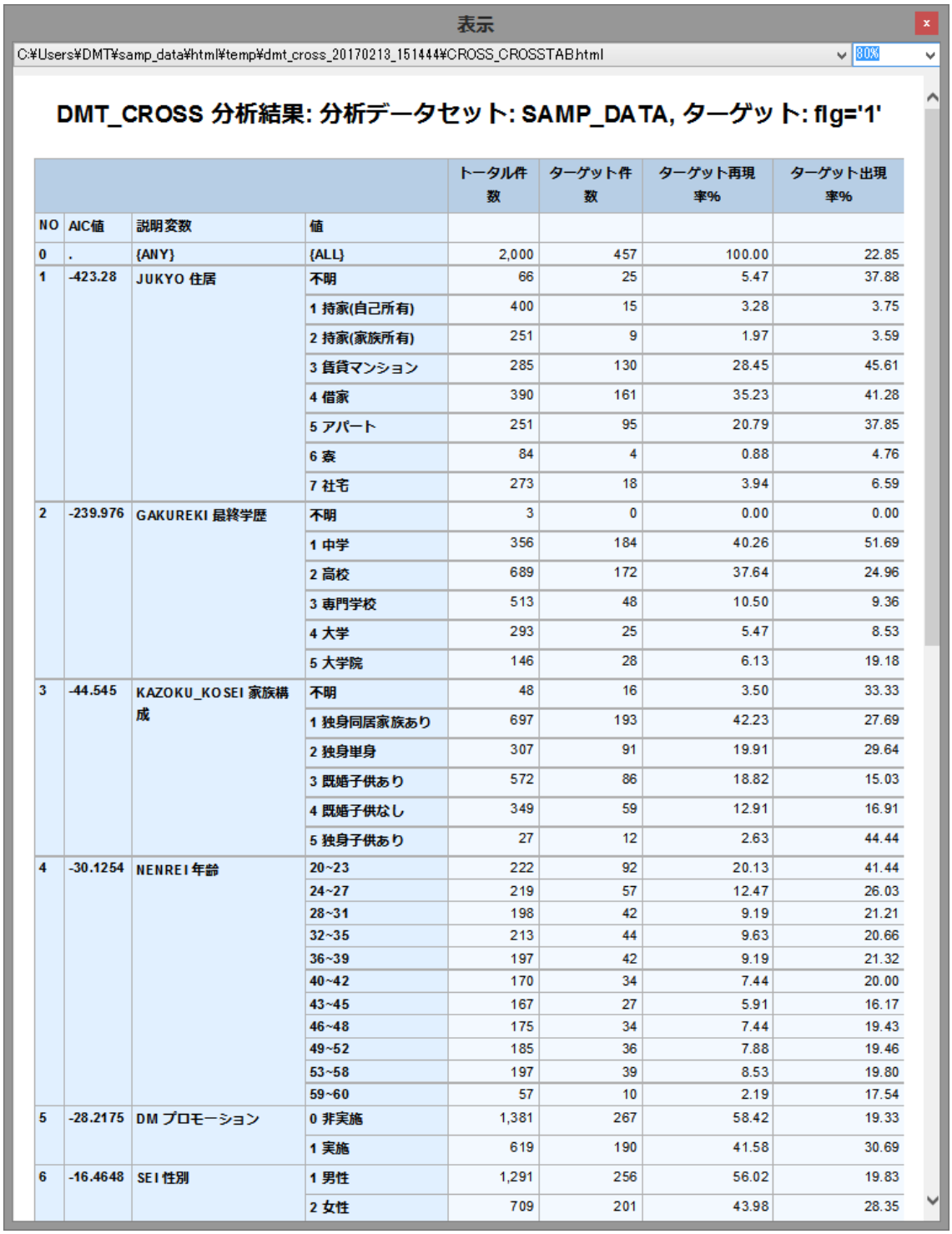


実行 を押します。

実行終了後、分析結果がデータ出力されたとのメッセージの後、
以下の出力表示確認画面が現れます。

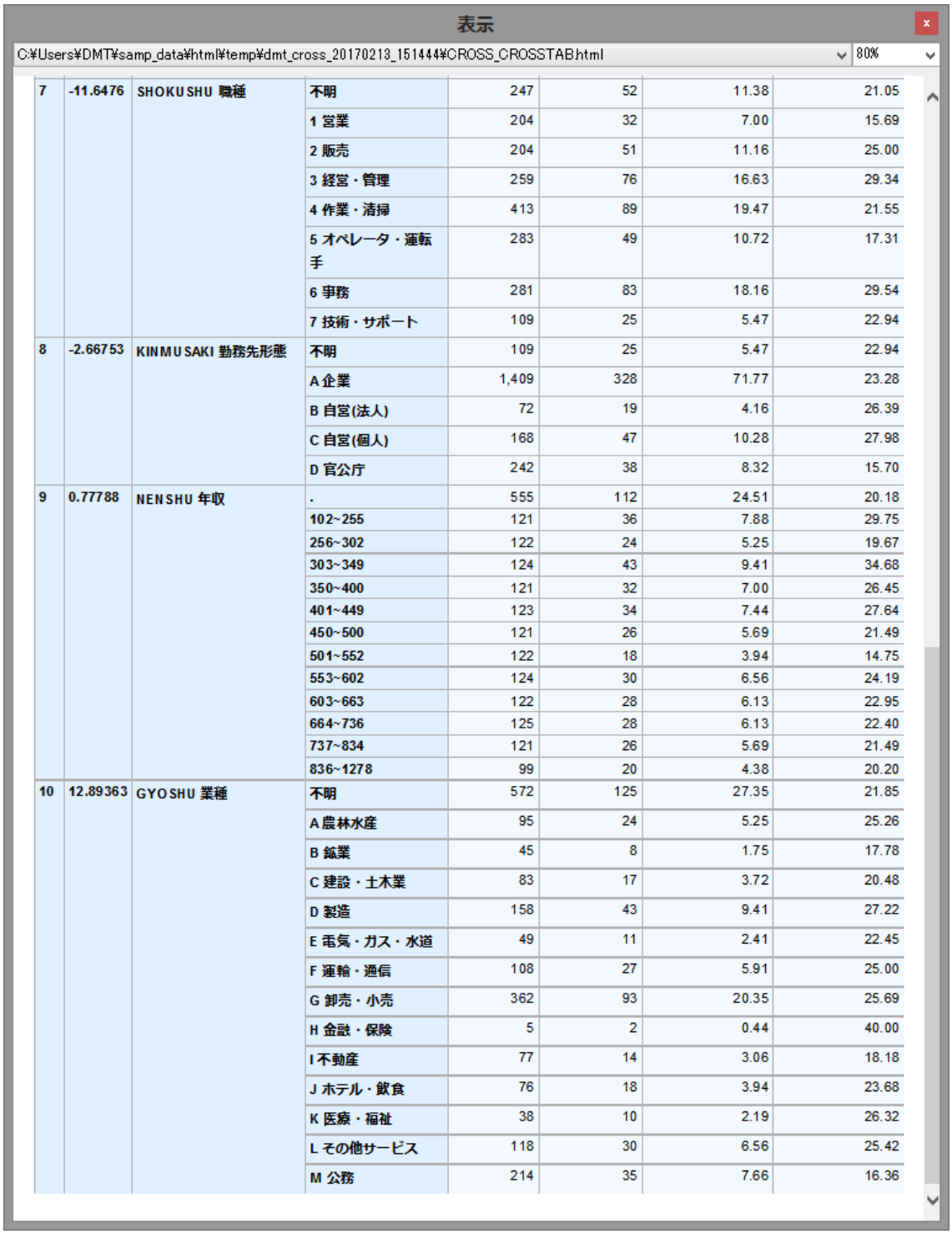


はい(Y) を押すと、分析結果表示画面が出現します。



DMT_CROSS 分析結果: 分析データセット: SAMP_DATA, ターゲット: flg='1'

				トータル件数	ターゲット件数	ターゲット再現率%	ターゲット出現率%
NO	AIC値	説明変数	値				
0	.	{ANY}	{ALL}	2,000	457	100.00	22.85
1	-423.28	JUKYO 住居	不明	66	25	5.47	37.88
			1 持家(自己所有)	400	15	3.28	3.75
			2 持家(家族所有)	251	9	1.97	3.59
			3 賃貸マンション	285	130	28.45	45.61
			4 借家	390	161	35.23	41.28
			5 アパート	251	95	20.79	37.85
			6 寮	84	4	0.88	4.76
			7 社宅	273	18	3.94	6.59
2	-239.976	GAKUREKI 最終学歴	不明	3	0	0.00	0.00
			1 中学	356	184	40.26	51.69
			2 高校	689	172	37.64	24.96
			3 専門学校	513	48	10.50	9.36
			4 大学	293	25	5.47	8.53
			5 大学院	146	28	6.13	19.18
3	-44.545	KAZOKU_KOSEI 家族構成	不明	48	16	3.50	33.33
			1 独身同居家族あり	697	193	42.23	27.69
			2 独身単身	307	91	19.91	29.64
			3 既婚子供あり	572	86	18.82	15.03
			4 既婚子供なし	349	59	12.91	16.91
			5 独身子供あり	27	12	2.63	44.44
4	-30.1254	NENREI 年齢	20~23	222	92	20.13	41.44
			24~27	219	57	12.47	26.03
			28~31	198	42	9.19	21.21
			32~35	213	44	9.63	20.66
			36~39	197	42	9.19	21.32
			40~42	170	34	7.44	20.00
			43~45	167	27	5.91	16.17
			46~48	175	34	7.44	19.43
			49~52	185	36	7.88	19.46
			53~58	197	39	8.53	19.80
			59~60	57	10	2.19	17.54
5	-28.2175	DM プロモーション	0 非実施	1,381	267	58.42	19.33
			1 実施	619	190	41.58	30.69
6	-16.4648	SEI 性別	1 男性	1,291	256	56.02	19.83
			2 女性	709	201	43.98	28.35



7	-11.6476	SHOKUSHU 職種	不明	247	52	11.38	21.05
			1 営業	204	32	7.00	15.69
			2 販売	204	51	11.16	25.00
			3 経営・管理	259	76	16.63	29.34
			4 作業・清掃	413	89	19.47	21.55
			5 オペレータ・運転手	283	49	10.72	17.31
			6 事務	281	83	18.16	29.54
			7 技術・サポート	109	25	5.47	22.94
8	-2.66753	KINMUSAKI 勤務先形態	不明	109	25	5.47	22.94
			A 企業	1,409	328	71.77	23.28
			B 自営(法人)	72	19	4.16	26.39
			C 自営(個人)	168	47	10.28	27.98
			D 官公庁	242	38	8.32	15.70
9	0.77788	NENSHU 年収	.	555	112	24.51	20.18
			102~255	121	36	7.88	29.75
			256~302	122	24	5.25	19.67
			303~349	124	43	9.41	34.68
			350~400	121	32	7.00	26.45
			401~449	123	34	7.44	27.64
			450~500	121	26	5.69	21.49
			501~552	122	18	3.94	14.75
			553~602	124	30	6.56	24.19
			603~663	122	28	6.13	22.95
			664~736	125	28	6.13	22.40
			737~834	121	26	5.69	21.49
			836~1278	99	20	4.38	20.20
10	12.89363	GYOSHU 業種	不明	572	125	27.35	21.85
			A 農林水産	95	24	5.25	25.26
			B 鉱業	45	8	1.75	17.78
			C 建設・土木業	83	17	3.72	20.48
			D 製造	158	43	9.41	27.22
			E 電気・ガス・水道	49	11	2.41	22.45
			F 運輸・通信	108	27	5.91	25.00
			G 卸売・小売	362	93	20.35	25.69
			H 金融・保険	5	2	0.44	40.00
			I 不動産	77	14	3.06	18.18
			J ホテル・飲食	76	18	3.94	23.68
			K 医療・福祉	38	10	2.19	26.32
			L その他サービス	118	30	6.56	25.42
			M 公務	214	35	7.66	16.36

クロス分析 結果表は、10 個の説明変数を、ターゲット変数 flg と関連が強い順（AIC 値の小さい順）に表示します。結果から、jukyo, gakureki, kazoku_kosei, nenrei, DM, sei ,shokushu, kinmusaki の 8 個の変数は、AIC 値が負の値となっており、flg と関連があることを示しています。 一方、表の末尾の NENSHU と GYOSHU については AIC 値がプラスとなっており、flg との関連性が認められないことを表しています。

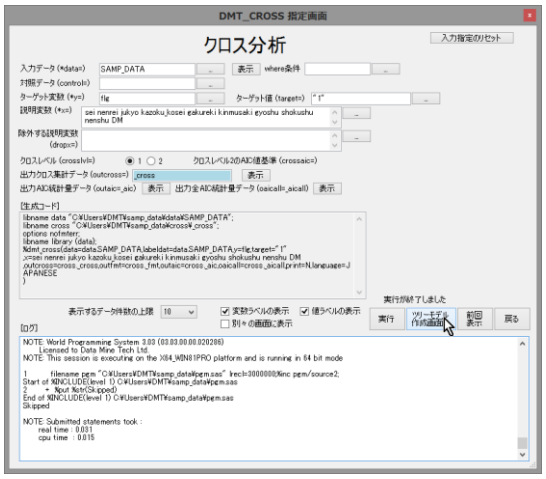
また、各変数カテゴリ別の該当度数、ターゲット件数、ターゲット再現率（=ターゲット件数／総ターゲット件数*100）と出現率（=ターゲット件数／該当件数*100）が表示されます。 文字タイプ説明変数のカテゴリ値とその該当件数、数値タイプ説明変数の存在範囲、外れ値

や欠損値の存在割合などが把握できます。

ボタンを押して **クロス分析結果表示** を終了し、**「クロス分析」**画面に戻ります。

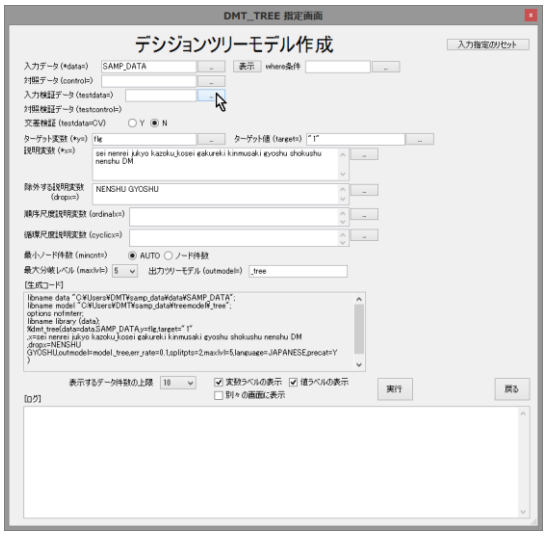
3.1.4 ツリーモデルの作成

「クロス分析」画面で ツリーモデル作成画面へ を押します。

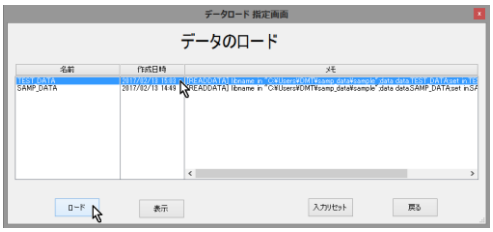


クロス分析画面で指定した入力データ、目的変数、そして分析結果に基づき、目的変数との関連性が見られた変数のみを説明変数に指定した**「デシジョンツリーモデル作成」**画面に切り替わります。(※ **除外する説明変数** に関連が無いとみなされた 2 つの変数 NENSHU, GYOSHU が自動指定されます)

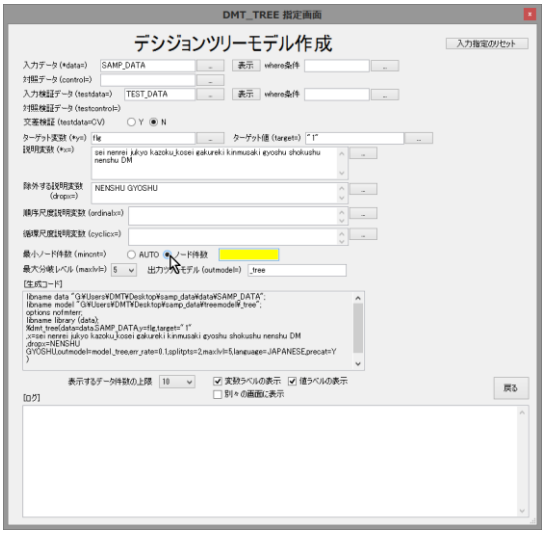
入力検証データ の ... を押します。



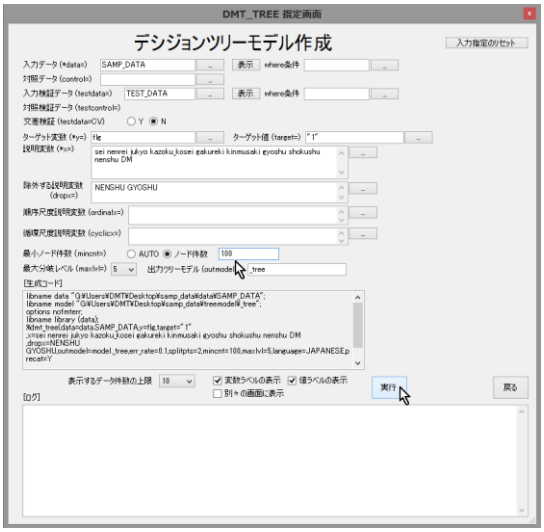
TEST_DATA を選択し、ロードします。



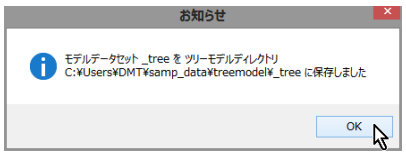
最小ノード件数 の指定を **自動** から **ノード件数** に切り替えます。



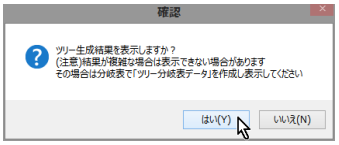
ノード件数の値 に 100 と入力してから 実行 を押します。



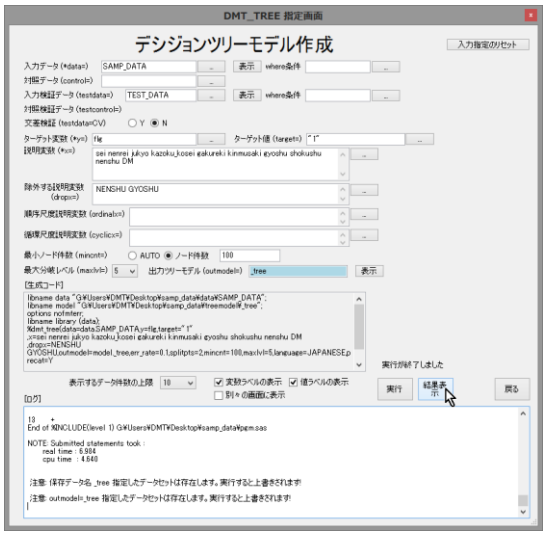
分析が実行され、しばらくすると終了します。
作成されたモデルが既定の _tree という名前で システムに保存されます。



分類木モデルの場合、ツリー分岐表、ゲインチャート、比較プロットが表示できます。

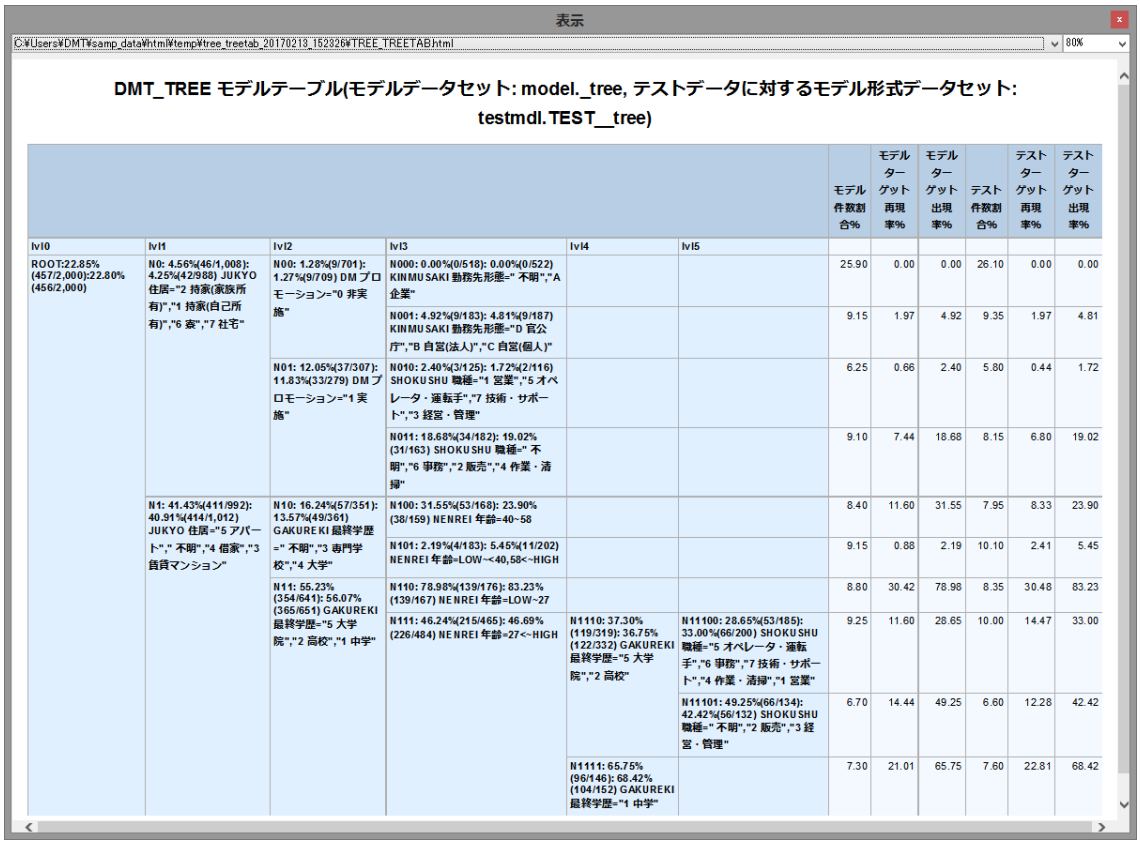


3.1.5 ツリーモデルの表示(ツリー分岐表)



ツリー分岐表 の表示

結果表示 を押します。



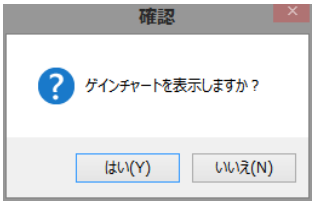
						モデル件数割合%	モデルターゲット再現率%	モデルターゲット出現率%	テスト件数割合%	テストターゲット再現率%	テストターゲット出現率%
lvl0	lvl1	lvl2	lvl3	lvl4	lvl5						
ROOT:22.85% (457/2,000):22.80% (456/2,000)	N0: 4.56%(46/1,008): 4.25%(42/988) JUKYO 住居="2 持家(家族所有)","1 持家(自己所有)","6 寮","7 社宅"	N00: 1.28%(9/701): 1.27%(9/709) DM プロモーション="0 非実施"	N000: 0.00%(0/518): 0.00%(0/522) KINMUSAKI 勤務先形態=" 不明","A 企業"			25.90	0.00	0.00	26.10	0.00	0.00
			N001: 4.92%(9/183): 4.81%(9/187) KINMUSAKI 勤務先形態="D 官公庁","B 自営(法人)","C 自営(個人)"			9.15	1.97	4.92	9.35	1.97	4.81
		N01: 12.05%(37/307): 11.83%(33/279) DM プロモーション="1 実施"	N010: 2.40%(3/125): 1.72%(2/116) SHOKUSHU 職種="1 営業","5 オペレータ・運転手","7 技術・サポート","3 経営・管理"			6.25	0.66	2.40	5.80	0.44	1.72
			N011: 18.68%(34/182): 19.02% (31/163) SHOKUSHU 職種=" 不明","6 事務","2 販売","4 作業・清掃"			9.10	7.44	18.68	8.15	6.80	19.02
	N1: 41.43%(411/992): 40.91%(414/1,012) JUKYO 住居="5 アパート","" 不明","4 借家","3 賃貸マンション"	N10: 16.24%(57/351): 13.57%(49/361) GAKUREKI 最終学歴=" 不明","3 専門学校","4 大学"	N100: 31.55%(53/168): 23.90% (38/159) NENREI 年齢=40~58			8.40	11.60	31.55	7.95	8.33	23.90
			N101: 2.19%(4/183): 5.45%(11/202) NENREI 年齢=LOW~<40,58<~HIGH			9.15	0.88	2.19	10.10	2.41	5.45
		N11: 55.23% (354/641): 56.07% (365/651) GAKUREKI 最終学歴="5 大学院","2 高校","1 中学"	N110: 78.98%(139/176): 83.23% (139/167) NENREI 年齢=LOW~27			8.80	30.42	78.98	8.35	30.48	83.23
			N111: 46.24%(215/465): 46.69% (226/484) NENREI 年齢=27<~HIGH	N1110: 37.30% (119/319): 36.75% (122/332) GAKUREKI 最終学歴="5 大学院","2 高校"	N11100: 28.65%(53/185): 33.00%(66/200) SHOKUSHU 職種="5 オペレータ・運転手","6 事務","7 技術・サポート","4 作業・清掃","1 営業"	9.25	11.60	28.65	10.00	14.47	33.00
					N11101: 49.25%(66/134): 42.42%(56/132) SHOKUSHU 職種=" 不明","2 販売","3 経営・管理"	6.70	14.44	49.25	6.60	12.28	42.42
				N1111: 65.75% (96/146): 68.42% (104/152) GAKUREKI 最終学歴="1 中学"		7.30	21.01	65.75	7.60	22.81	68.42

ツリー分岐表 には、ノード分岐に採用された説明変数値とターゲット出現率（ターゲット件数／ノード件数）が分岐ノードごとに表示されます。また、終端ノードについては、**「件数割合%」**、**「ターゲット再現率%」**、**「ターゲット出現率%」** が右側に表示されます。

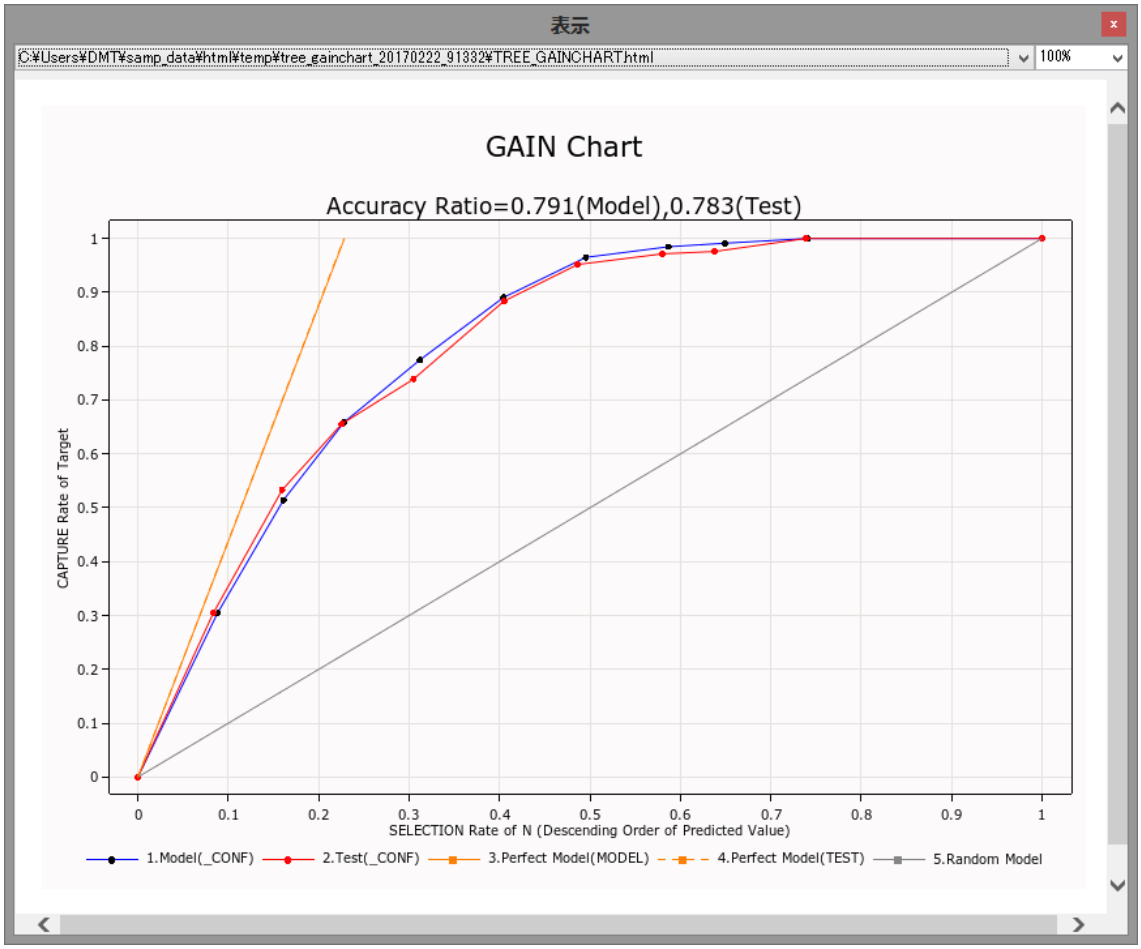
ツリー分岐表 は、本アプリケーションのツリー生成アルゴリズムに従って、自動的に出現率（購入率）の高低の差ができるだけ顕著となるように、分析対象データを逐次的に分けていく過程が表示されています。なお、ここでは、ツリー生成条件として、最小ノード件数=100、最大分岐レベル=5（既定値）をセットしています。

出現率の分布は、まず **住居区分** の違いによって最も大きくなっており、持家系のグループ（1,008 件）では 4.56% の出現率（平均の 22.85% の約 1/5）、賃貸系のグループ（992 件）では 41.43%（平均の約 2 倍）の出現率を示しています。さらに、持家系のグループは DM **プロモーション有無** によって分かれ、**プロモーション** 実施 グループは 12.05%、**プロモーション非**実施 グループは 1.28%の出現率となっています。その他のグループも、出現率の高低が最も際立つように自動的に選ばれた項目値によって分かれていきます。最終的に 10 個のグループ（終端ノード）が生成されており、各ノードの出現率は 0%～78.98% の範囲に分布しています。

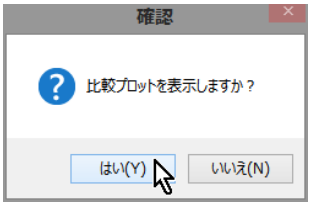
3.1.6 ツリーモデルの評価（ゲインチャート）



ゲインチャートの表示

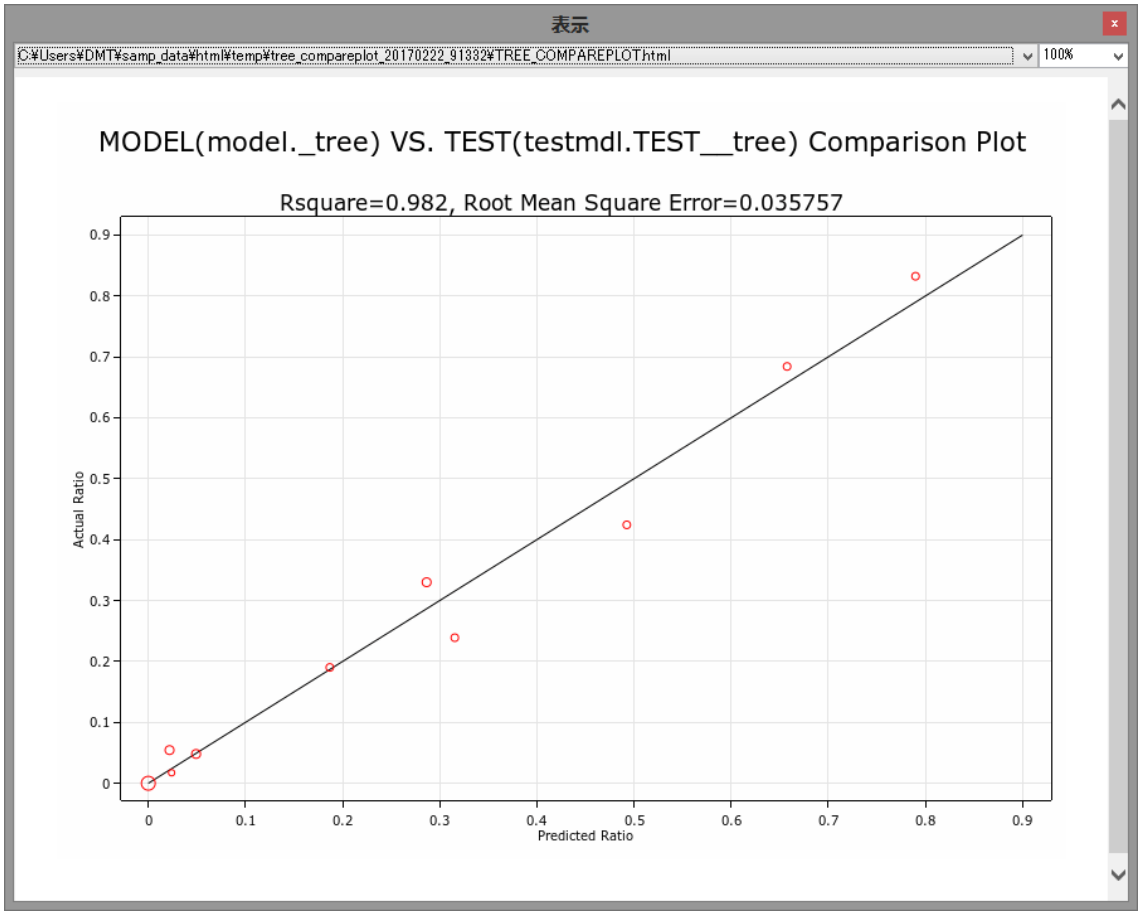


ゲインチャート はモデルの予測出現率の順位と実績出現率との関連を評価するモデルの精度指標の 1 つです。左上に膨らんだ曲線になっているほど、モデルの精度（ここでは予測確率の大きさと実際のイベント出現率との関連性を意味します）が高いことを表し、テストデータにモデルを当てはめたときの曲線との差が小さいほどモデル精度の安定性（汎化性能）が高いことを表します。この結果例では、まずまずの精度と安定性を示しています。



比較プロット（予測値と実際値の散布図）の表示

3.1.7 ツリーモデルの評価(比較プロット)



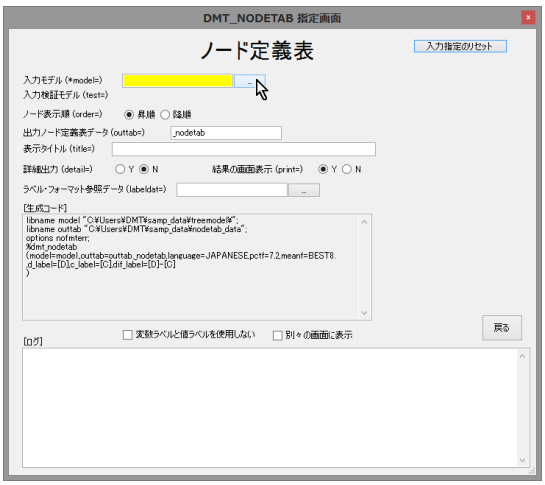
比較プロット はモデルの予測値と実績値の差（誤差）の大きさを評価します。TEST_ DATA にモデルを当てはめた場合の、10 個の終端ノードの予測出現率と実績出現率の散布図が表示されます。終端ノードを表す赤い円が 0 から 0.65 の範囲に広がり、いずれも対角線上の近くにプロットされていますので、検証データにおけるツリーモデルの予測値は実績値に近かったことがわかります。

「デシジョンツリーモデル作成」 画面を終了し、**「メニュー」**画面 に戻ります。

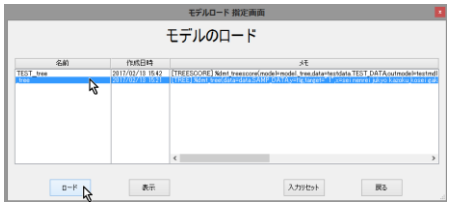
3.1.8 ツリーノードの表示（ノード定義表）

ノード表 を押すと、**「ノード定義表」**画面に切り替わります。

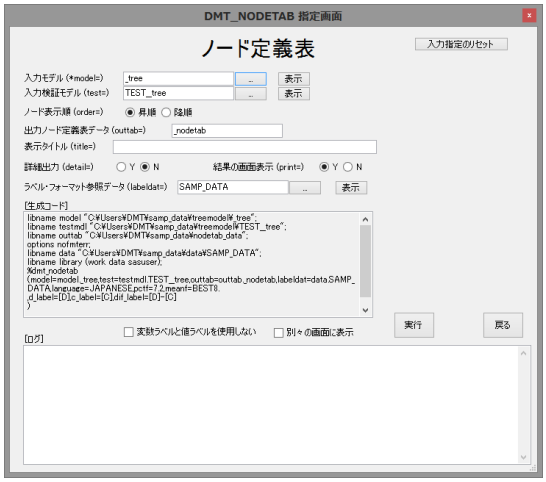
既存のツリーモデルに対し、各終端ノードの説明変数組合せ定義が分かる形式でモデルの内容を表示します。



入力モデル を選択します。

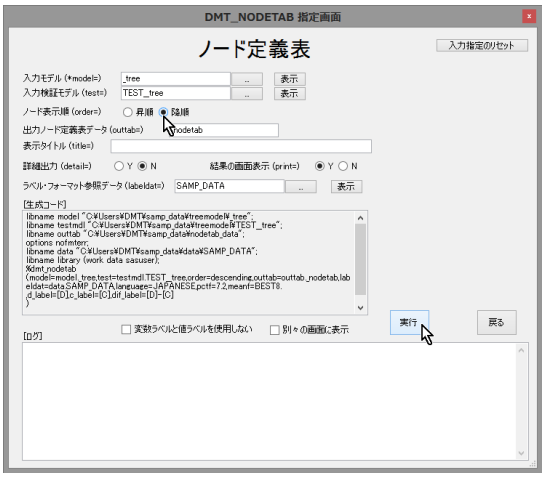


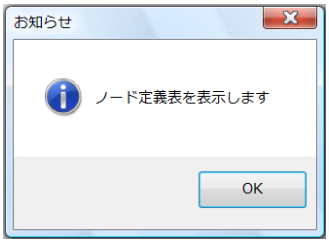
_tree を選択し、ロードします。



入力検証モデル には**「デシジョンツリーモデル作成」**において、入力検証データに指定された TEST_DATA にモデル _tree を適用した モデル形式データセット TEST__tree が自動入力されます。
また、出力ノード定義表データ、ラベル・フォーマット参照データの項目にもそれぞれ、 _nodetab, SAMP_DATA が自動入力されます。

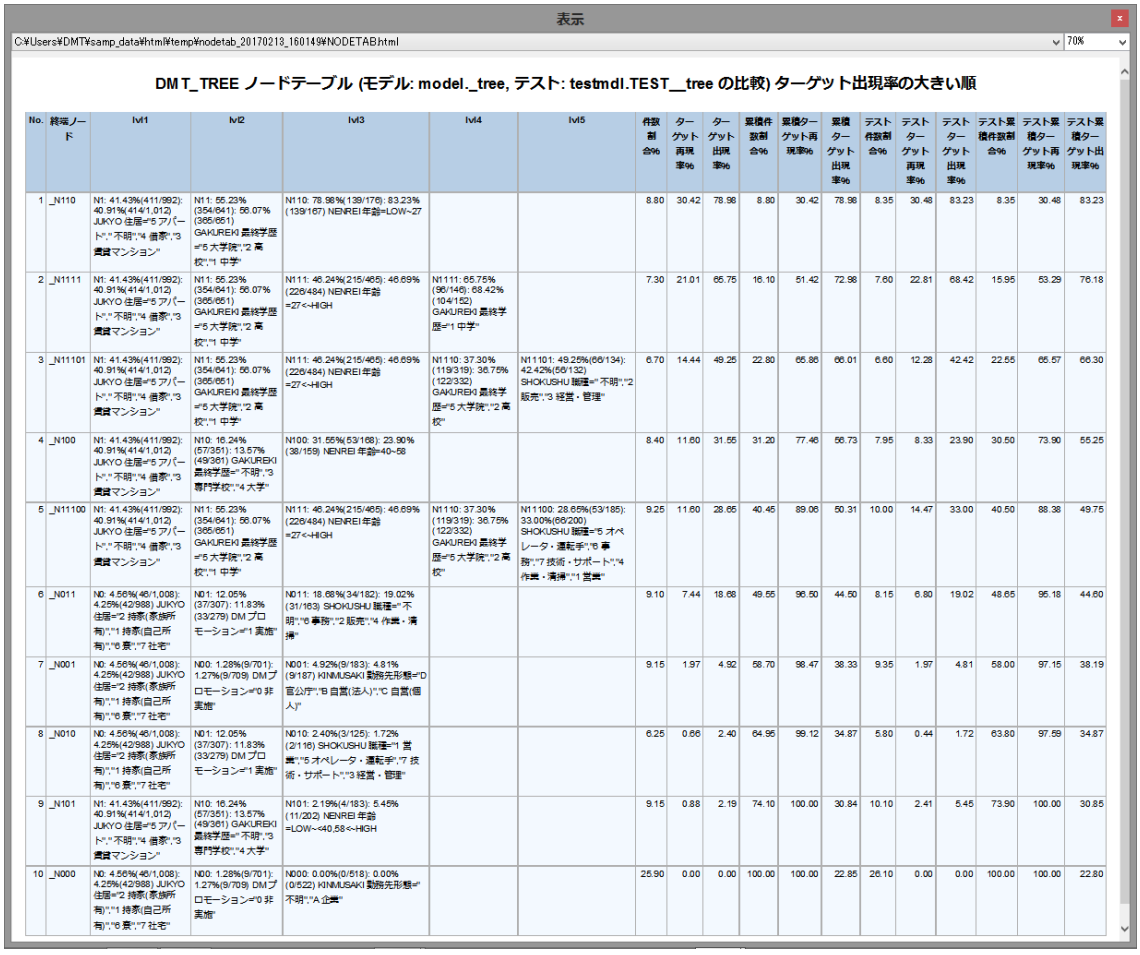
ノード表示順 を 降順 にセットして、ノードの並びを出現率の小さい順（デフォルトの 昇順） から 大きい順（降順） に変更し、 実行 を押します。





実行終了後、上記画面が表示されますので、 OK を押します。

ノード定義表 の表示



No.	終端ノード	lvl1	lvl2	lvl3	lvl4	lvl5	件数割合%	ターゲット再現率%	ターゲット出現率%	累積件数割合%	累積ターゲット再現率%	累積ターゲット出現率%	テスト件数割合%	テストターゲット再現率%	テストターゲット出現率%	テスト累積件数割合%	テスト累積ターゲット再現率%	テスト累積ターゲット出現率%
1	_N110	N1: 41.43%(411/992): 40.91%(414/1,012) JUKYO 住居="5 アパート","" 不明","4 借家","3 賃貸マンション"	N11: 55.23% (354/641): 56.07% (365/651) GAKUREKI 最終学歴 ="5 大学院","2 高校","1 中学"	N110: 78.98%(139/176): 83.23% (139/167) NENREI 年齢=LOW~27			8.80	30.42	78.98	8.80	30.42	78.98	8.35	30.48	83.23	8.35	30.48	83.23
2	_N1111	N1: 41.43%(411/992): 40.91%(414/1,012) JUKYO 住居="5 アパート","" 不明","4 借家","3 賃貸マンション"	N11: 55.23% (354/641): 56.07% (365/651) GAKUREKI 最終学歴 ="5 大学院","2 高校","1 中学"	N111: 46.24%(215/465): 46.69% (226/484) NENREI 年齢 =27<-HIGH	N1111: 65.75% (96/146): 68.42% (104/152) GAKUREKI 最終学歴="1 中学"		7.30	21.01	65.75	16.10	51.42	72.98	7.60	22.81	68.42	15.95	53.29	76.18
3	_N11101	N1: 41.43%(411/992): 40.91%(414/1,012) JUKYO 住居="5 アパート","" 不明","4 借家","3 賃貸マンション"	N11: 55.23% (354/641): 56.07% (365/651) GAKUREKI 最終学歴 ="5 大学院","2 高校","1 中学"	N111: 46.24%(215/465): 46.69% (226/484) NENREI 年齢 =27<-HIGH	N1110: 37.30% (119/319): 36.75% (122/332) GAKUREKI 最終学歴="5 大学院","2 高校"	N11101: 49.25%(66/134): 42.42%(56/132) SHOKUSHU 職種="" 不明","2 販売","3 経営・管理"	6.70	14.44	49.25	22.80	65.86	66.01	6.60	12.28	42.42	22.55	65.57	66.30
4	_N100	N1: 41.43%(411/992): 40.91%(414/1,012) JUKYO 住居="5 アパート","" 不明","4 借家","3 賃貸マンション"	N10: 16.24% (57/351): 13.57% (49/361) GAKUREKI 最終学歴="" 不明","3 専門学校","4 大学"	N100: 31.55%(53/168): 23.90% (38/159) NENREI 年齢=40~58			8.40	11.60	31.55	31.20	77.46	56.73	7.95	8.33	23.90	30.50	73.90	55.25
5	_N11100	N1: 41.43%(411/992): 40.91%(414/1,012) JUKYO 住居="5 アパート","" 不明","4 借家","3 賃貸マンション"	N11: 55.23% (354/641): 56.07% (365/651) GAKUREKI 最終学歴 ="5 大学院","2 高校","1 中学"	N111: 46.24%(215/465): 46.69% (226/484) NENREI 年齢 =27<-HIGH	N1110: 37.30% (119/319): 36.75% (122/332) GAKUREKI 最終学歴="5 大学院","2 高校"	N11100: 28.65%(53/185): 33.00%(66/200) SHOKUSHU 職種="5 オペレータ・運転手","6 事務","7 技術・サポート","4 作業・清掃","1 営業"	9.25	11.60	28.65	40.45	89.06	50.31	10.00	14.47	33.00	40.50	88.38	49.75
6	_N011	N0: 4.56%(46/1,008): 4.25%(42/988) JUKYO 住居="2 持家(家族所有)","1 持家(自己所有)","6 寮","7 社宅"	N01: 12.05% (37/307): 11.83% (33/279) DM プロモーション="1 実施"	N011: 18.68%(34/182): 19.02% (31/163) SHOKUSHU 職種="不明","6 事務","2 販売","4 作業・清掃"			9.10	7.44	18.68	49.55	96.50	44.50	8.15	6.80	19.02	48.65	95.18	44.60
7	_N001	N0: 4.56%(46/1,008): 4.25%(42/988) JUKYO 住居="2 持家(家族所有)","1 持家(自己所有)","6 寮","7 社宅"	N00: 1.28%(9/701): 1.27%(9/709) DM プロモーション="0 非実施"	N001: 4.92%(9/183): 4.81% (9/187) KINMUSAKI 勤務先形態="D 官公庁","B 自営(法人)","C 自営(個人)"			9.15	1.97	4.92	58.70	98.47	38.33	9.35	1.97	4.81	58.00	97.15	38.19
8	_N010	N0: 4.56%(46/1,008): 4.25%(42/988) JUKYO 住居="2 持家(家族所有)","1 持家(自己所有)","6 寮","7 社宅"	N01: 12.05% (37/307): 11.83% (33/279) DM プロモーション="1 実施"	N010: 2.40%(3/125): 1.72% (2/116) SHOKUSHU 職種="1 営業","5 オペレータ・運転手","7 技術・サポート","3 経営・管理"			6.25	0.66	2.40	64.95	99.12	34.87	5.80	0.44	1.72	63.80	97.59	34.87
9	_N101	N1: 41.43%(411/992): 40.91%(414/1,012) JUKYO 住居="5 アパート","" 不明","4 借家","3 賃貸マンション"	N10: 16.24% (57/351): 13.57% (49/361) GAKUREKI 最終学歴="" 不明","3 専門学校","4 大学"	N101: 2.19%(4/183): 5.45% (11/202) NENREI 年齢 =LOW~<40.58<-HIGH			9.15	0.88	2.19	74.10	100.00	30.84	10.10	2.41	5.45	73.90	100.00	30.85
10	_N000	N0: 4.56%(46/1,008): 4.25%(42/988) JUKYO 住居="2 持家(家族所有)","1 持家(自己所有)","6 寮","7 社宅"	N00: 1.28%(9/701): 1.27%(9/709) DM プロモーション="0 非実施"	N000: 0.00%(0/518): 0.00% (0/522) KINMUSAKI 勤務先形態="" 不明","A 企業"			25.90	0.00	0.00	100.00	100.00	22.85	26.10	0.00	0.00	100.00	100.00	22.80

ノード定義表 には、**終端ノード** 別の生成規則（説明変数値の組合せ方）を表す **ノードの定義**（この例では「M1」~「M5」の最大 5 つの変数値の組合せ）、と各ノードのターゲット値に関する統計量が表示されます。統計量としては、ノードごとの**「件数割合%」**、**「ターゲット再現率%」**、**「ターゲット出現率%」** がノード分岐表の場合と同じく表示され、さらに、その右側に、No1 からそのノードの No までの累積値も表示されます。また、今回のように検証データ（TEST=パラメータ）を指定した場合は、モデルを検証データに適用した場合の統計量も表示されます。

ノード定義表を見ると、優良顧客（または不良（不芳）顧客）のイメージをノードの説明変数値の組合せによって把握することができます。
また、優良顧客や休眠顧客を対象として、さまざまな施策（営業促進施策や与信施策など）を実施する場合、ノード定義表で集計表示された各種統計量は、**施策実施範囲（累積件数割合）**や**施策実施効果（累積ターゲット再現率と累積ターゲット出現率）**を検討するために用いることもできます。

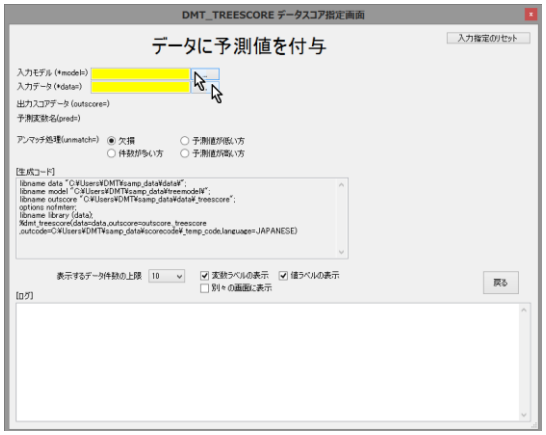
例えば、この結果から、上位 3 個の終端ノードに該当する顧客のみを対象として、新たな施策を実施する場合、施策実施対象者の分析母集団全体に対する割合（**「累積件数割合%」**）は 22.8%、施策実施により応答するであろう顧客の分析母集団全体に対する捕捉割合（**「累積ターゲット再現率%」**）は 66.86%、期待出現率（**「累積ターゲット出現率%」**）は 66.01%と見積もることができます。つまり、全体の売上件数の 6 割を稼ぐ 2 割の優良顧客を特定することが出来たということを示しています。

3.1.9 モデル予測値の付与（スコアリング）

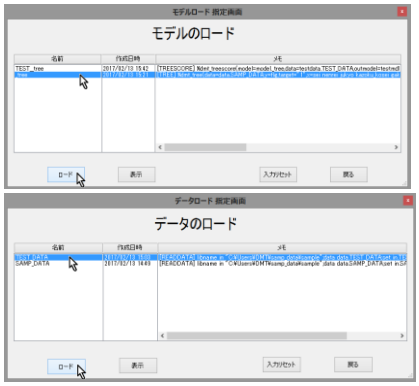
分析結果画面、デシジョンツリーモデル作成画面を閉じて、**「メニュー」** 画面 に戻ります。
次の分析のために、検証用データ（TEST_DATA）にモデル予測値を付与します。

予測付与 を押します。

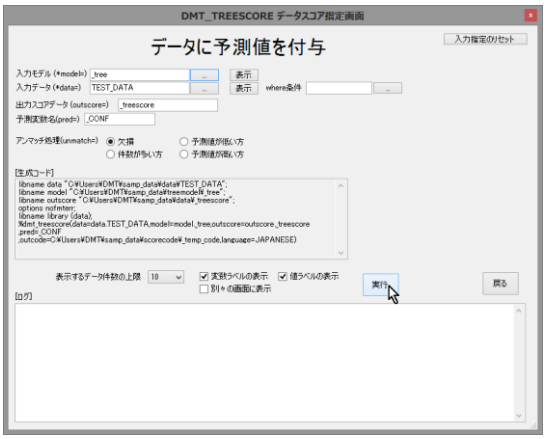
「データに予測値を付与」画面に切り替わります。



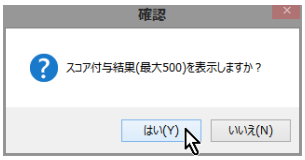
入力モデル に _tree をロードし、**入力データ** に TEST_DATA をロードします。



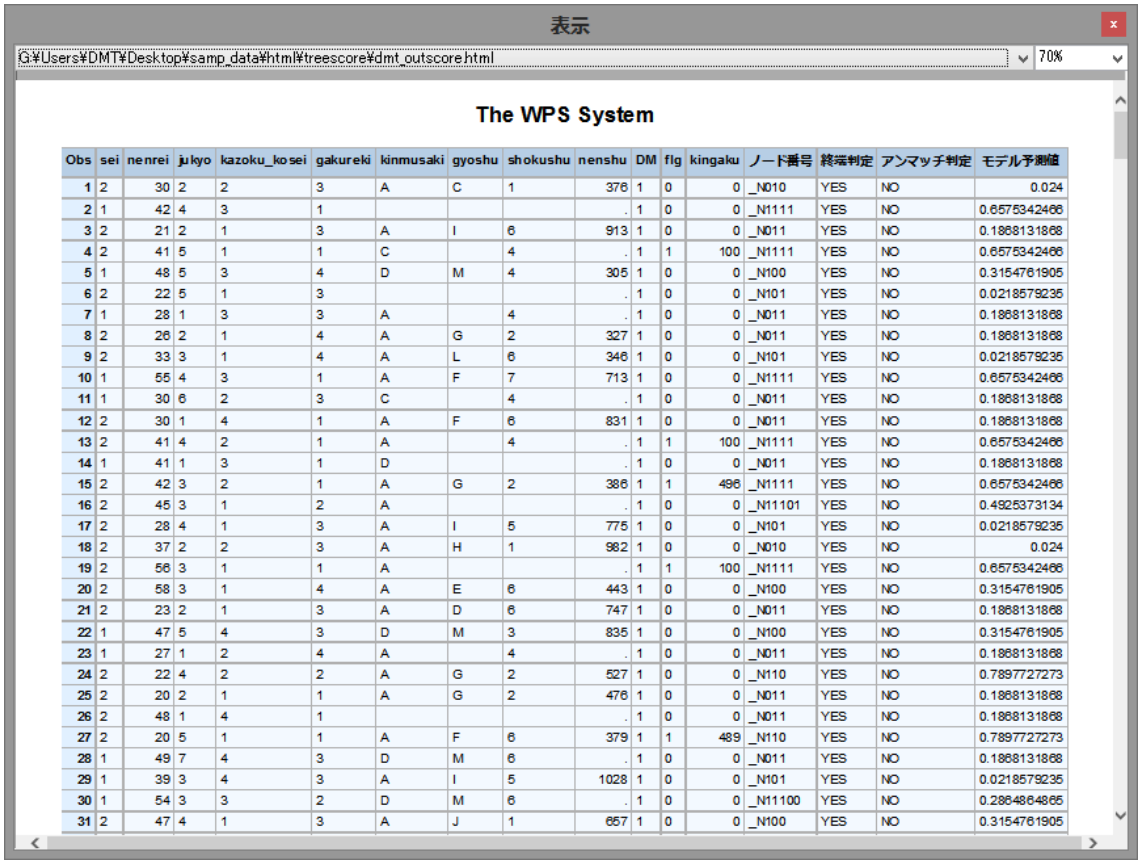
実行 を押します。



予測値表示 を押します。



予測値付与結果 の表示



Obs	sei	nenrei	jukyo	kazoku_kosei	gakureki	kinmusaki	gyoshu	shokushu	nenshu	DM	flg	kingaku	ノード番号	終端判定	アンマッチ判定	モデル予測値
1	2	30	2	2	3	A	C	1	376	1	0	0	_N010	YES	NO	0.024
2	1	42	4	3	1				.	1	0	0	_N1111	YES	NO	0.6575342466
3	2	21	2	1	3	A	I	6	913	1	0	0	_N011	YES	NO	0.1868131868
4	2	41	5	1	1	C		4	.	1	1	100	_N1111	YES	NO	0.6575342466
5	1	48	5	3	4	D	M	4	305	1	0	0	_N100	YES	NO	0.3154761905
6	2	22	5	1	3				.	1	0	0	_N101	YES	NO	0.0218579235
7	1	28	1	3	3	A		4	.	1	0	0	_N011	YES	NO	0.1868131868
8	2	26	2	1	4	A	G	2	327	1	0	0	_N011	YES	NO	0.1868131868
9	2	33	3	1	4	A	L	6	346	1	0	0	_N101	YES	NO	0.0218579235
10	1	55	4	3	1	A	F	7	713	1	0	0	_N1111	YES	NO	0.6575342466
11	1	30	6	2	3	C		4	.	1	0	0	_N011	YES	NO	0.1868131868
12	2	30	1	4	1	A	F	6	831	1	0	0	_N011	YES	NO	0.1868131868
13	2	41	4	2	1	A		4	.	1	1	100	_N1111	YES	NO	0.6575342466
14	1	41	1	3	1	D			.	1	0	0	_N011	YES	NO	0.1868131868
15	2	42	3	2	1	A	G	2	386	1	1	496	_N1111	YES	NO	0.6575342466
16	2	45	3	1	2	A			.	1	0	0	_N11101	YES	NO	0.4925373134
17	2	28	4	1	3	A	I	5	775	1	0	0	_N101	YES	NO	0.0218579235
18	2	37	2	2	3	A	H	1	982	1	0	0	_N010	YES	NO	0.024
19	2	56	3	1	1	A			.	1	1	100	_N1111	YES	NO	0.6575342466
20	2	58	3	1	4	A	E	6	443	1	0	0	_N100	YES	NO	0.3154761905
21	2	23	2	1	3	A	D	6	747	1	0	0	_N011	YES	NO	0.1868131868
22	1	47	5	4	3	D	M	3	835	1	0	0	_N100	YES	NO	0.3154761905
23	1	27	1	2	4	A		4	.	1	0	0	_N011	YES	NO	0.1868131868
24	2	22	4	2	2	A	G	2	527	1	0	0	_N110	YES	NO	0.7897727273
25	2	20	2	1	1	A	G	2	476	1	0	0	_N011	YES	NO	0.1868131868
26	2	48	1	4	1				.	1	0	0	_N011	YES	NO	0.1868131868
27	2	20	5	1	1	A	F	6	379	1	1	489	_N110	YES	NO	0.7897727273
28	1	49	7	4	3	D	M	6	.	1	0	0	_N011	YES	NO	0.1868131868
29	1	39	3	4	3	A	I	5	1028	1	0	0	_N101	YES	NO	0.0218579235
30	1	54	3	3	2	D	M	6	.	1	0	0	_N11100	YES	NO	0.2864864865
31	2	47	4	1	3	A	J	1	657	1	0	0	_N100	YES	NO	0.3154761905

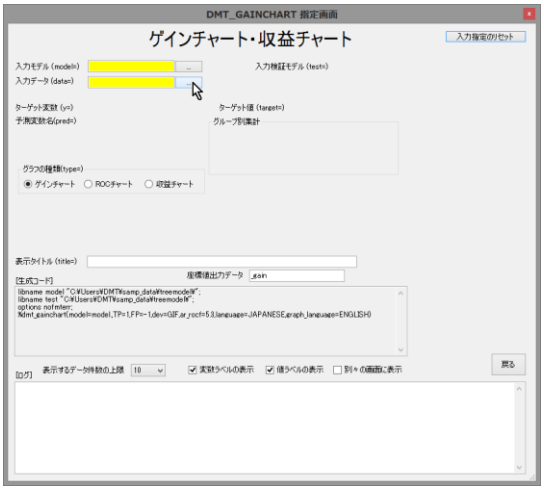
3.1.10 収益チャート

検証データに付与したモデル予測出現率を使って、出現率が高い方からどの出現率までの終端ノードに対して営業施策を実施すると最大収益が得られるかを計算します。ただし、この営業施策の１件当たりのコストは50、購入発生の場合の収益は検証データの実績購入金額（変数 kingaku の値）とみなします。

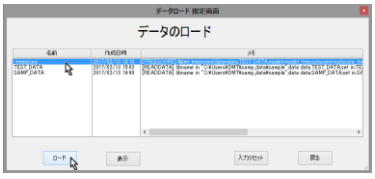
ゲイン・収益 を押します。

「ゲインチャート・収益チャート」画面に切り替わります。

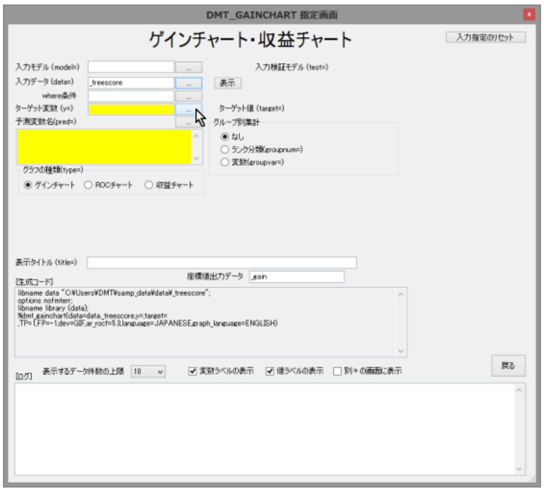
入力データ の ... を押します。



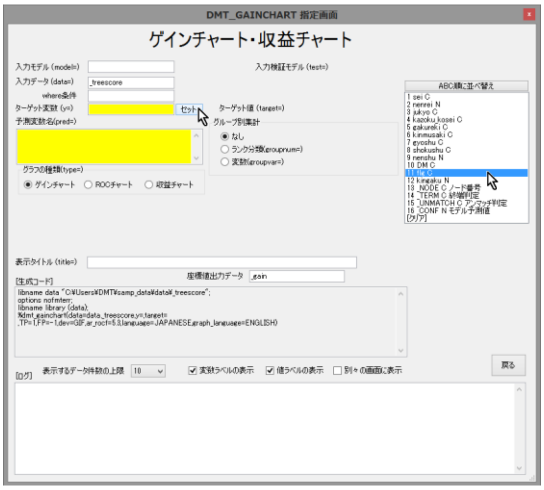
_treescore をロードします。



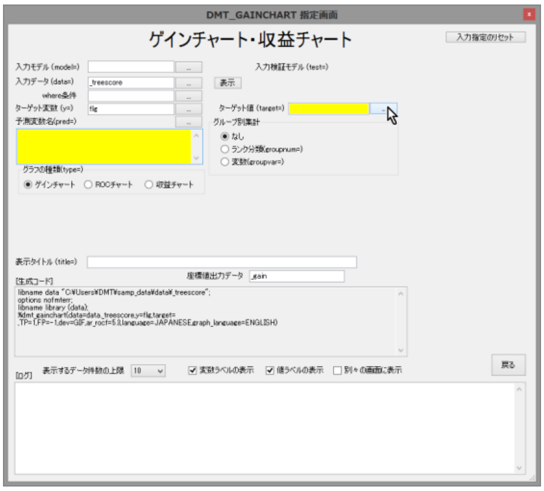
ターゲット変数 に flg 、**ターゲット値**に 1 、**予測変数名** に _CONF をセットします。



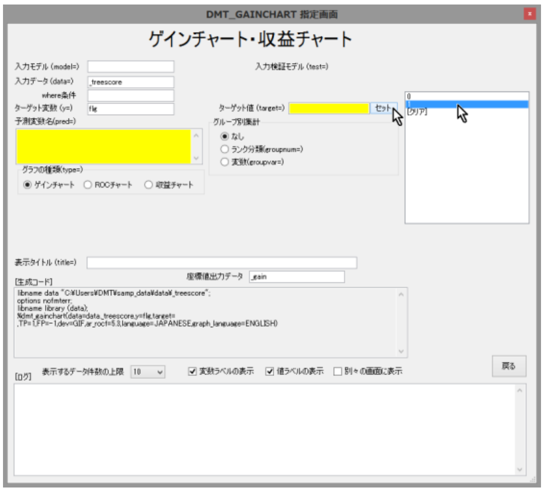
↓



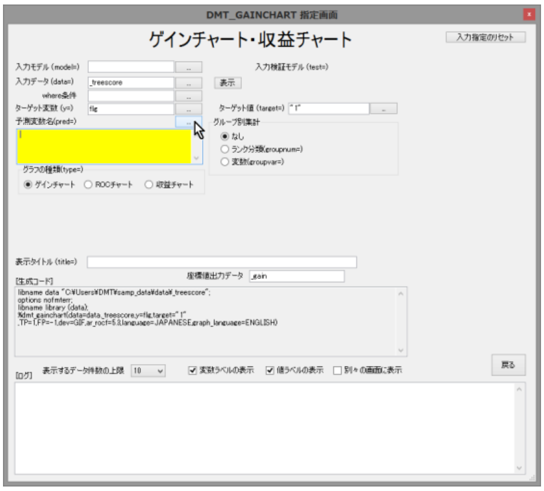
↓



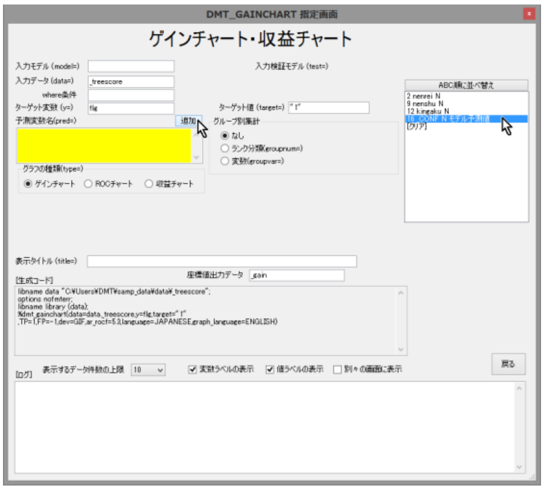
↓



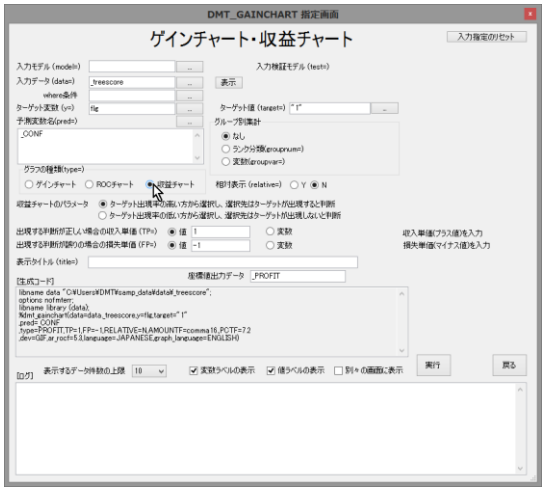
↓



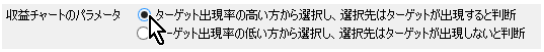
↓



グラフの種類 を **収益チャート** に変更します。



今回は、購入することを期待して出現率（購入率）が高い方から施策実施対象を選択するので、**ターゲット出現率の高い方から選択** の設定のままにしておきます。

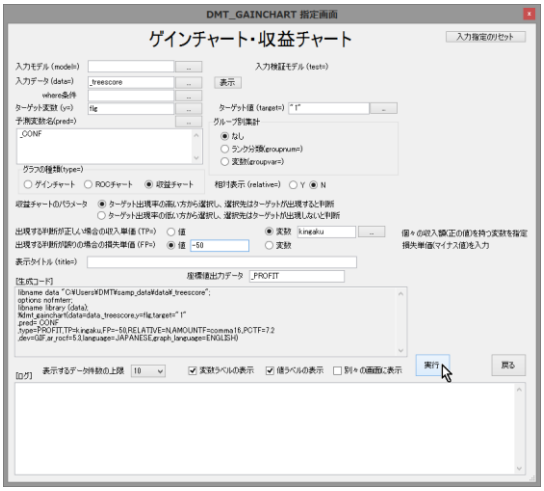


購入するだろうという**判断が正しかった場合**の施策実施顧客からの収益は、購入実績金額（変数 kingaku）を選択します。

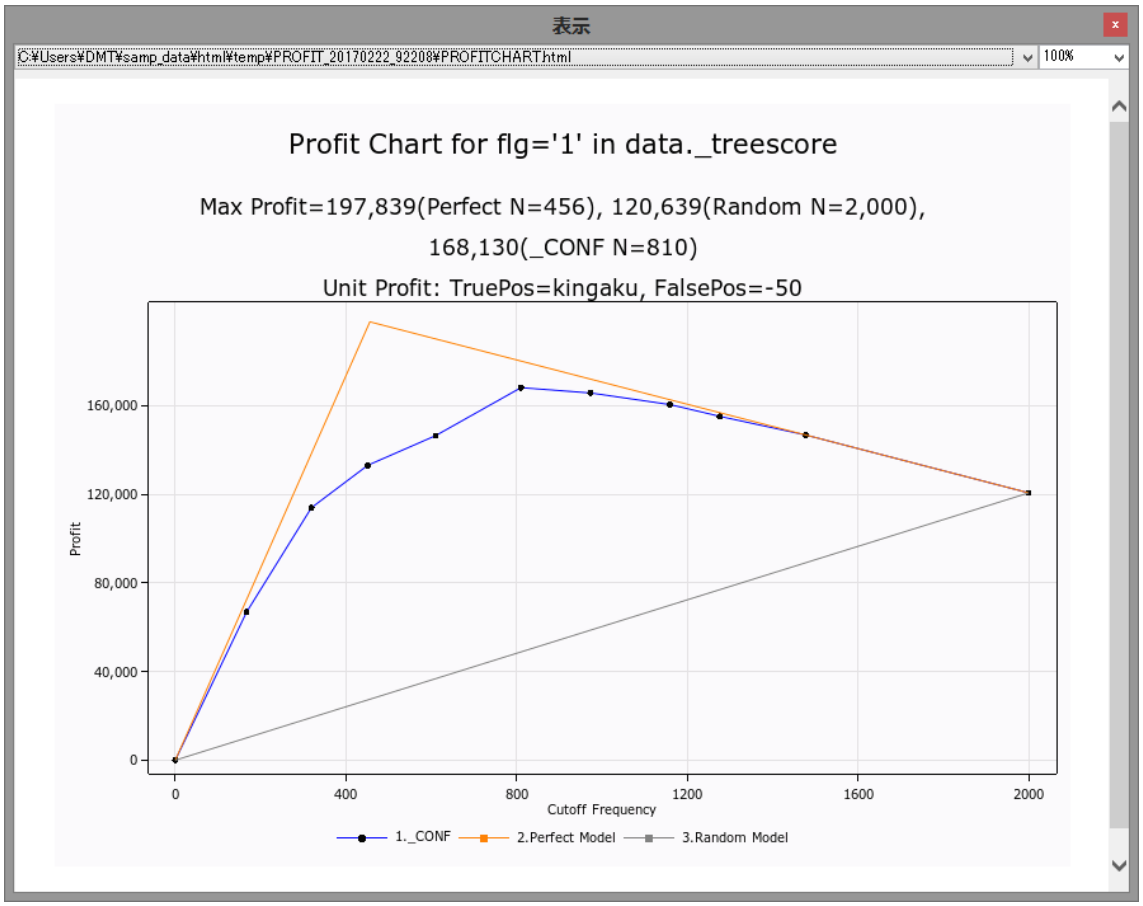
↓

一方、購入するだろうという**判断が誤っていた場合**の施策実施顧客にかかるコストは、今回は一律 -50 とします

以上のパラメータ設定後、 実行 を押します。



収益チャートの表示



収益チャート の**横軸は予測出現率の大きい順に終端ノードを並べたときの累積件数（施策選択対象件数）**を表し、**縦軸はその累積件数から得られる合計収益額** を表します。

図の左端の点は施策実行対象を全く選択しなかった場合を表し、常に収益=0 となります。一方、図の右端の点は全部のノード（全員）を選択した場合を表し、どのモデルを用いても同じ値になります。（値は収益とコストの関係で決まります。負の値になる場合もあります。）図から、出現率の大きい方から 5 個の終端ノードまでを施策実施対象として選択した場合に **最大収益** が得られることがわかります。（実施対象件数 810 件、期待収益額 168,130）
このように、実務的な収益の観点から最適な施策実施対象を定義することが可能です。

なお、456 件の購入あり顧客のみを施策実施対象として選択する **完全モデル** の収益額は 197,839 です。一方、ランダムモデル（あてずっぽうモデル） を使う場合は、全員を施策実施対象とする場合が最大収益が得られ、収益額は 120,639 となります。

3.2 （例 2）施策実施効果の分析

施策効果が大きい／小さい顧客の判別ルールを作成します。目的変数はクラス変数 flg、購入確率を求めたいクラスは flg=1（購入あり）で、施策実施 / 非実施のデータ区分は、変数 DM の値（実施：DM="1", 非実施：DM="0"）で識別されています。

以下の分析手順を実行します。

3.2.1 データ読込

分析に用いるデータ（SAMP_DATA）とモデル検証に用いるデータ（TEST_DATA）は 3.1.1 で既に読み込まれています。

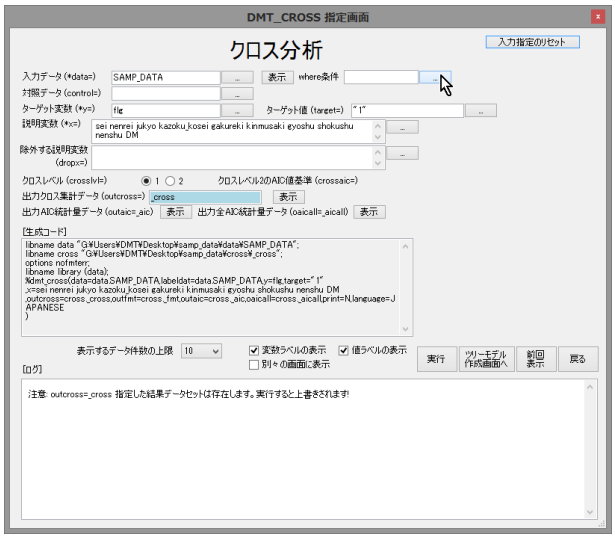
3.2.2 ラベル付与

SAMP_DATA には 3.1.2 で既に変数と文字変数値にラベルが付けられています。

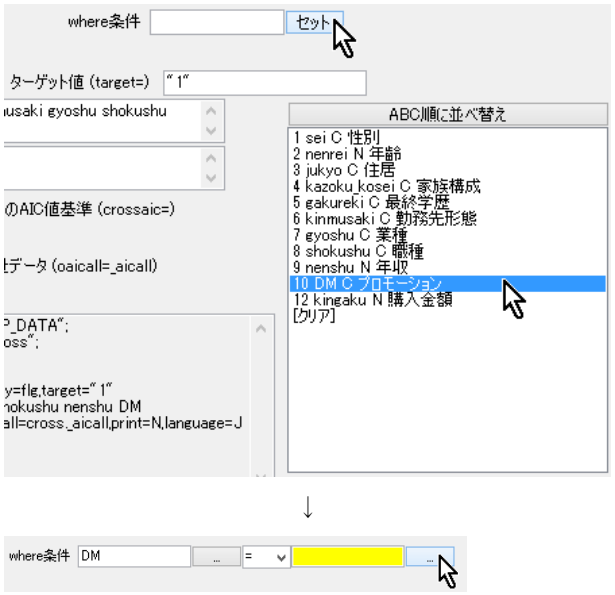
3.2.3 項目分析

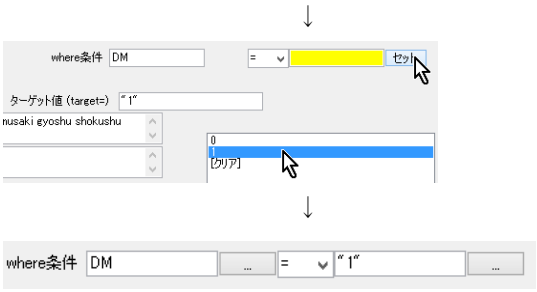
デシジョンツリーモデル作成前の事前分析として、説明変数とターゲット変数との関連性や説明変数分布の把握を行います。

クロス分析 を押します。**「クロス分析」**画面が前回指定したパラメータが指定された状態で開きます。

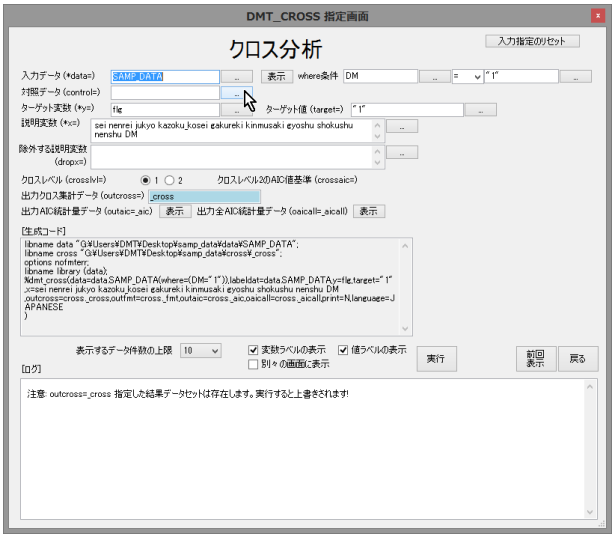


where **条件** ... を押し、SAMP_DATA の中で、**変数 DM の値が "1" の条件を満たすオブザベーションを施策実施データとして入力するよう指定します。**

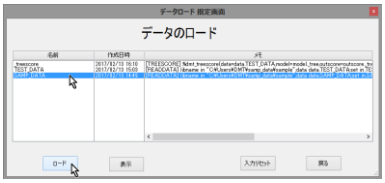


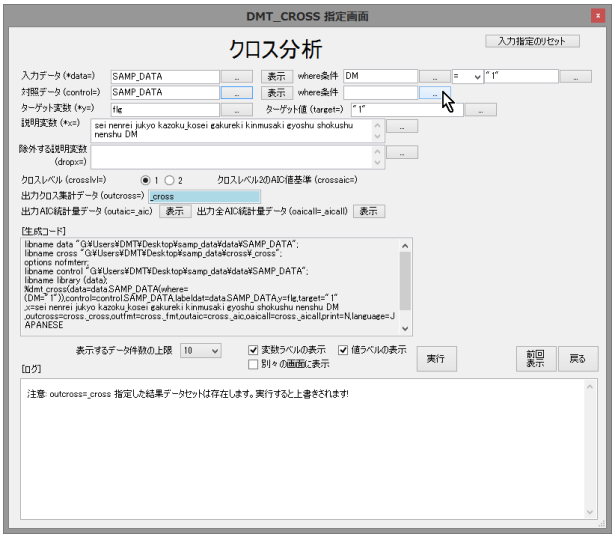


対照データ ... を押します。

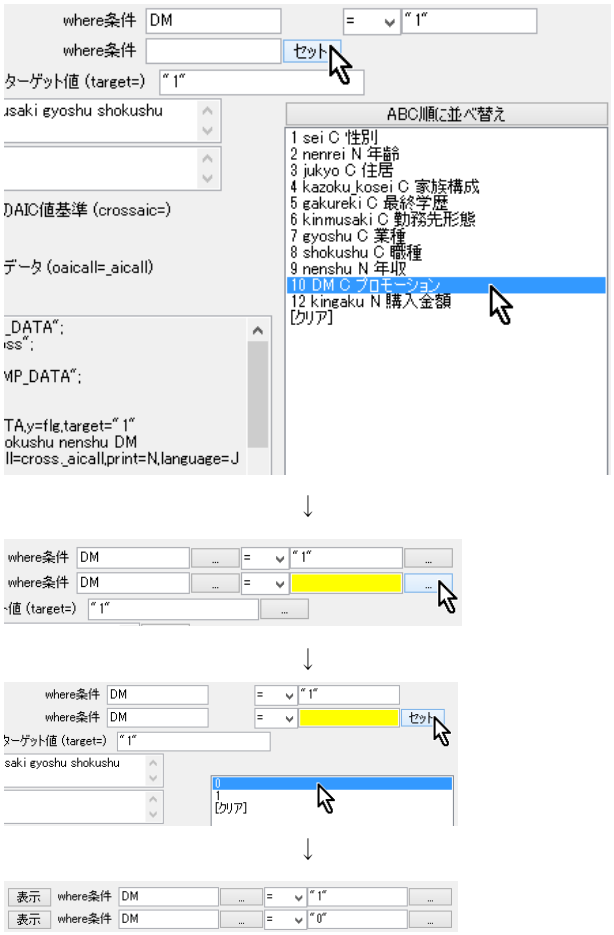


SAMP_DATA をロードします。

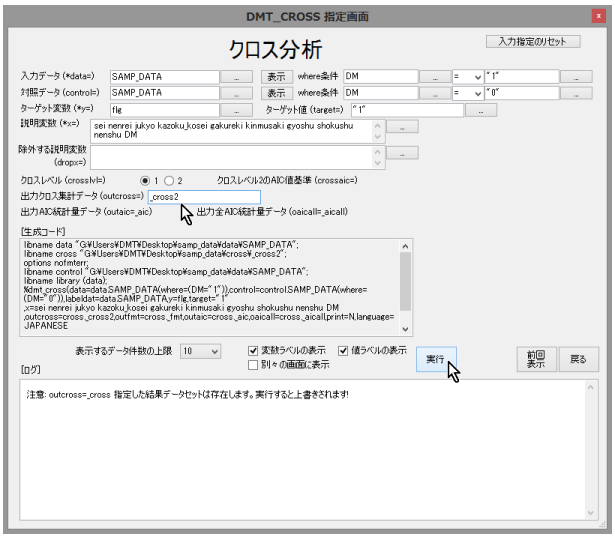




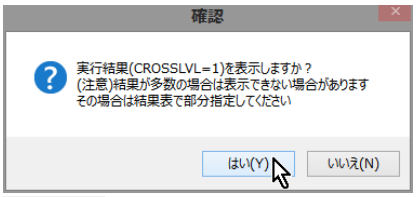
where 条件 [...] を押し、SAMP_DATA の中で、**変数 DM の値が "0" の条件を満たすオブザベーションを対照（施策非実施）データとして入力するよう指定します。**



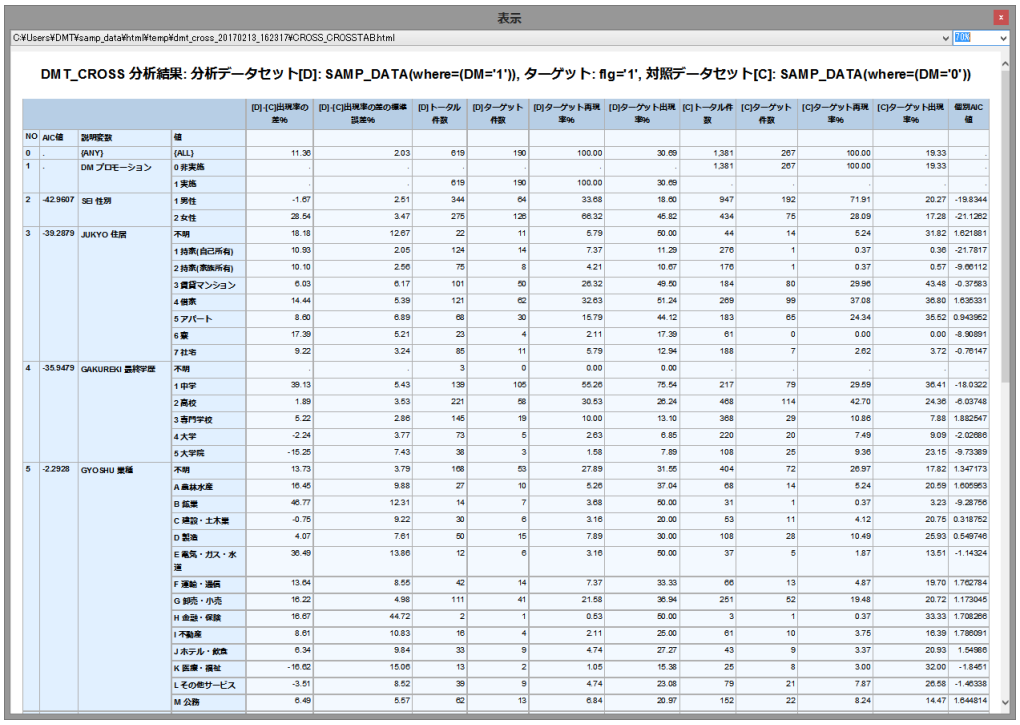
出力クロス集計データ を _cross2 に変更します。



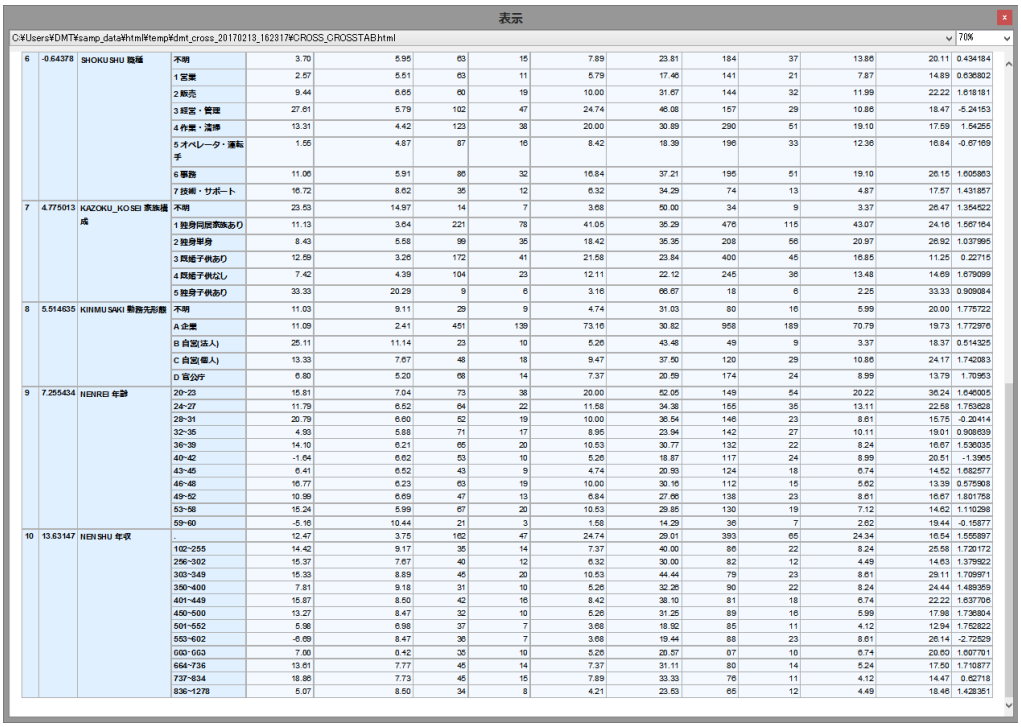
実行 を押します。



はい(Y) を押すと、分析結果表示画面が出現します。



NO	AIC値	説明変数	値	[D]-[C]出現率の差%	[D]-[C]出現率の差の標準誤差%	[D]トータル件数	[D]ターゲット件数	[D]ターゲット再現率%	[D]ターゲット出現率%	[C]トータル件数	[C]ターゲット件数	[C]ターゲット再現率%	[C]ターゲット出現率%	個別AIC値
0	.	(ANY)	(ALL)	11.36	2.03	619	190	100.00	30.69	1,381	267	100.00	19.33	.
1	.	DM プロモーション	0 非実施	.	.	.	.	.	.	1,381	267	100.00	19.33	.
			1 実施	.	.	619	190	100.00	30.69	.	.	.	.	.
2	-42.9607	SEI 性別	1 男性	-1.67	2.51	344	64	33.68	18.60	947	192	71.91	20.27	-19.8344
			2 女性	28.54	3.47	275	126	66.32	45.82	434	75	28.09	17.28	-21.1262
3	-39.2879	JUKYO 住居	不明	18.18	12.67	22	11	5.79	50.00	44	14	5.24	31.82	1.621881
			1 持家(自己所有)	10.93	2.05	124	14	7.37	11.29	276	1	0.37	0.36	-21.7817
			2 持家(家族所有)	10.10	2.56	75	8	4.21	10.67	176	1	0.37	0.57	-9.66112
			3 賃貸マンション	6.03	6.17	101	50	26.32	49.50	184	80	29.96	43.48	-0.37583
			4 借家	14.44	5.39	121	62	32.63	51.24	269	99	37.08	36.80	1.636331
			5 アパート	8.60	6.89	68	30	15.79	44.12	183	65	24.34	35.52	0.943952
			6 寮	17.39	5.21	23	4	2.11	17.39	61	0	0.00	0.00	-8.90891
			7 社宅	9.22	3.24	85	11	5.79	12.94	188	7	2.62	3.72	-0.76147
4	-35.9479	GAKUREKI 最終学歴	不明	.	.	3	0	0.00	0.00	.	.	.	.	.
			1 中学	39.13	5.43	139	105	55.26	75.54	217	79	29.59	36.41	-18.0322
			2 高校	1.89	3.53	221	58	30.53	26.24	468	114	42.70	24.36	-6.03748
			3 専門学校	5.22	2.86	145	19	10.00	13.10	368	29	10.86	7.88	1.882547
			4 大学	-2.24	3.77	73	5	2.63	6.85	220	20	7.49	9.09	-2.02686
			5 大学院	-15.25	7.43	38	3	1.58	7.89	108	25	9.36	23.15	-9.73389
5	-2.2928	GYOSHU 業種	不明	13.73	3.79	168	53	27.89	31.55	404	72	26.97	17.82	1.347173
			A 農林水産	16.45	9.88	27	10	5.26	37.04	68	14	5.24	20.59	1.605963
			B 鉱業	46.77	12.31	14	7	3.68	50.00	31	1	0.37	3.23	-9.28756
			C 建設・土木業	-0.75	9.22	30	6	3.16	20.00	53	11	4.12	20.75	0.318752
			D 製造	4.07	7.61	50	15	7.89	30.00	108	28	10.49	25.93	0.549746
			E 電気・ガス・水道	36.49	13.86	12	6	3.16	50.00	37	5	1.87	13.51	-1.14324
			F 運輸・通信	13.64	8.55	42	14	7.37	33.33	66	13	4.87	19.70	1.762784
			G 卸売・小売	16.22	4.98	111	41	21.58	36.94	251	52	19.48	20.72	1.173045
			H 金融・保険	16.67	44.72	2	1	0.53	50.00	3	1	0.37	33.33	1.708266
			I 不動産	8.61	10.83	16	4	2.11	25.00	61	10	3.75	16.39	1.786091
			J ホテル・飲食	6.34	9.84	33	9	4.74	27.27	43	9	3.37	20.93	1.54966
			K 医療・福祉	-16.62	15.06	13	2	1.05	15.38	25	8	3.00	32.00	-1.8451
			L その他サービス	-3.51	8.52	39	9	4.74	23.08	79	21	7.87	26.58	-1.46338
			M 公務	6.49	5.57	62	13	6.84	20.97	152	22	8.24	14.47	1.644814



6	-0.64378	SHOKUSHU 職種	不明	3.70	5.95	63	15	7.89	23.81	184	37	13.86	20.11	0.434184
			1 営業	2.57	5.51	63	11	5.79	17.46	141	21	7.87	14.89	0.636802
			2 販売	9.44	6.65	60	19	10.00	31.67	144	32	11.99	22.22	1.618181
			3 経営・管理	27.61	5.79	102	47	24.74	46.08	157	29	10.86	18.47	-5.24153
			4 作業・清掃	13.31	4.42	123	38	20.00	30.89	290	51	19.10	17.59	1.54255
			5 オペレータ・運転手	1.55	4.87	87	16	8.42	18.39	196	33	12.36	16.84	-0.67169
			6 事務	11.06	5.91	86	32	16.84	37.21	195	51	19.10	26.15	1.605863
			7 技術・サポート	16.72	8.62	35	12	6.32	34.29	74	13	4.87	17.57	1.431857
7	4.775013	KAZOKU_KOSEI 家族構成	不明	23.53	14.97	14	7	3.68	50.00	34	9	3.37	26.47	1.354522
			1 独身同居家族あり	11.13	3.64	221	78	41.05	35.29	476	115	43.07	24.16	1.567164
			2 独身単身	8.43	5.58	99	35	18.42	35.35	208	56	20.97	26.92	1.037995
			3 既婚子供あり	12.59	3.26	172	41	21.58	23.84	400	45	16.85	11.25	0.22715
			4 既婚子供なし	7.42	4.39	104	23	12.11	22.12	245	36	13.48	14.69	1.679099
			5 独身子供あり	33.33	20.29	9	6	3.16	66.67	18	6	2.25	33.33	0.909084
8	5.514635	KINMUSAKI 勤務先形態	不明	11.03	9.11	29	9	4.74	31.03	80	16	5.99	20.00	1.775722
			A 企業	11.09	2.41	451	139	73.16	30.82	958	189	70.79	19.73	1.772976
			B 自営(法人)	25.11	11.14	23	10	5.26	43.48	49	9	3.37	18.37	0.514325
			C 自営(個人)	13.33	7.67	48	18	9.47	37.50	120	29	10.86	24.17	1.742083
			D 官公庁	6.80	5.20	68	14	7.37	20.59	174	24	8.99	13.79	1.70963
9	7.255434	NENREI 年齢	20~23	15.81	7.04	73	38	20.00	52.05	149	54	20.22	36.24	1.646005
			24~27	11.79	6.52	64	22	11.58	34.38	155	35	13.11	22.58	1.753628
			28~31	20.79	6.60	52	19	10.00	36.54	146	23	8.61	15.75	-0.20414
			32~35	4.93	5.88	71	17	8.95	23.94	142	27	10.11	19.01	0.908639
			36~39	14.10	6.21	65	20	10.53	30.77	132	22	8.24	16.67	1.536035
			40~42	-1.64	6.62	53	10	5.26	18.87	117	24	8.99	20.51	-1.3965
			43~45	6.41	6.52	43	9	4.74	20.93	124	18	6.74	14.52	1.682577
			46~48	16.77	6.23	63	19	10.00	30.16	112	15	5.62	13.39	0.575908
			49~52	10.99	6.69	47	13	6.84	27.66	138	23	8.61	16.67	1.801758
			53~58	15.24	5.99	67	20	10.53	29.85	130	19	7.12	14.62	1.110298
			59~60	-5.16	10.44	21	3	1.58	14.29	36	7	2.62	19.44	-0.15877
10	13.63147	NENSHU 年収	.	12.47	3.75	162	47	24.74	29.01	393	65	24.34	16.54	1.555897
			102~255	14.42	9.17	35	14	7.37	40.00	86	22	8.24	25.58	1.720172
			256~302	15.37	7.67	40	12	6.32	30.00	82	12	4.49	14.63	1.379922
			303~349	15.33	8.89	45	20	10.53	44.44	79	23	8.61	29.11	1.709971
			350~400	7.81	9.18	31	10	5.26	32.26	90	22	8.24	24.44	1.489359
			401~449	15.87	8.50	42	16	8.42	38.10	81	18	6.74	22.22	1.637706
			450~500	13.27	8.47	32	10	5.26	31.25	89	16	5.99	17.98	1.736804
			501~552	5.98	6.98	37	7	3.68	18.92	85	11	4.12	12.94	1.752822
			553~602	-6.69	8.47	36	7	3.68	19.44	88	23	8.61	26.14	-2.72529
			603~663	7.00	0.42	35	10	5.26	28.57	67	10	6.74	20.60	1.607701
			664~736	13.61	7.77	45	14	7.37	31.11	80	14	5.24	17.50	1.710877
			737~834	18.86	7.73	45	15	7.89	33.33	76	11	4.12	14.47	0.62718
			836~1278	5.07	8.50	34	8	4.21	23.53	65	12	4.49	18.46	1.428351

クロス分析 結果表は、10 個の説明変数を、実施群と対照群間のカテゴリ別のターゲット変数 flg の分布の差が総合的に大きい順(AIC 値の小さい順)に表示します。ただし、リストの最初にある変数 DM は、施策実施データと対照データを分けるための識別変数ですので、AIC 値は欠損です。

結果から、sei ,jukyo, gakureki, gyoshu, shokushu の 5 個の変数は、AIC 値が負の値となっており、flg の差が施策実施有無と関連があることを示しています。（ただし、shokushu の AIC 値は 0 に近い負の値ですので、強い関連ではありません）No.0 の全体を見ると、実施群と対照群間の出現率の差は 11.36%　（標準誤差 2.14%）ですので、施策効果はあったと認められます。しかし、いくつかのカテゴリ（例えば、sei の男性や gakureki の大学院生では、実施群と対照群間の出現率の差が負の値とな

っており、この施策は男性や大学院卒に対しては無効（むしろ逆効果）であったことを意味しています。一方、残りの変数については AIC 値がプラスとなっており、flg の差と施策実施有無との関連性は認められないことを表しています。

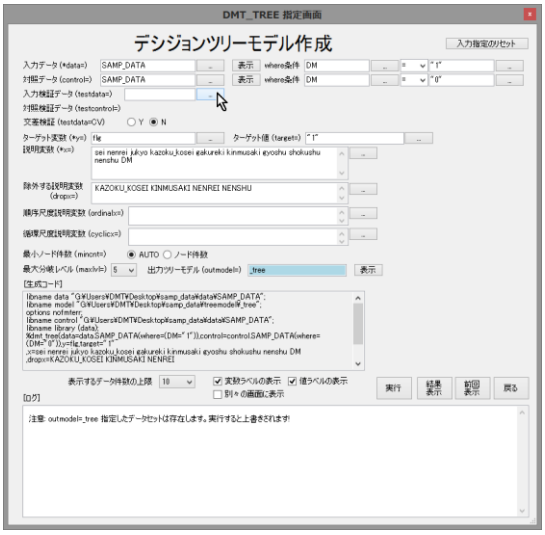
表には、各変数カテゴリ別の出現率の差、出現率の差の標準誤差、実施群と対照群それぞれにおける、該当度数、ターゲット件数、ターゲット再現率（=ターゲット件数／総ターゲット件数*100）と出現率（=ターゲット件数／該当件数*100）が表示されます。 そして、表の一番右には、カテゴリ単位で評価した flg の差と施策実施有無との関連性を表す個別 AIC 値が表示されます。

ボタンを押して **クロス分析結果表示** を終了し、**「クロス分析」**画面に戻ります。

3.2.4 ツリーモデルの作成

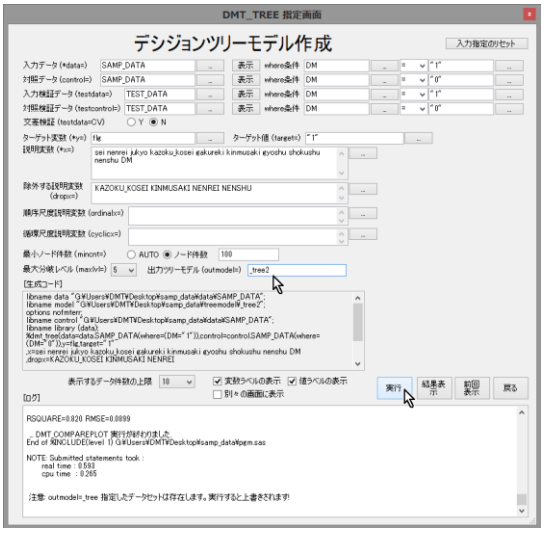
「クロス分析」画面で ツリーモデル作成画面へ を押します。

クロス分析画面で指定した入力データ、目的変数、そして分析結果に基づき、目的変数との関連性が見られた変数のみを説明変数に指定した**「デシジョンツリーモデル作成」**画面に切り替わります。(※ **除外する説明変数** に関連が無いとみなされた変数が自動指定されます)



入力検証データ に TEST_DATA を指定し、where 条件（DM = "1"）を追加指定します。
同様に、**対照検証データ** にも TEST_DATA をロードし、where 条件（DM = "0"）を追加指定します。

そして、
最小ノード件数 を AUTO から 100 件に切り替え、
出力ツリーモデル を _tree2 に変更した後、
実行 を押します。



(※ 最小ノード件数を AUTO （既定）に設定すると、ツリー分岐生成条件である分岐後の各ノードに含まれるデータの必要最小件数が、一定件数ではなく、分岐後の各ノードのターゲット出現率の標準誤差の指定の誤差率内に収まるように設定されます。 誤差率が小さいほど分岐が起こりにくくなりますが、この例では、サンプル数が少ないので、既定値（0.1）のままではツリーが生育しにくいため一定のデータ件数を最小ノード件数として指定しています。 なお、 AUTO 指定のときの誤

差率はオプション画面で設定値を変更できます。）

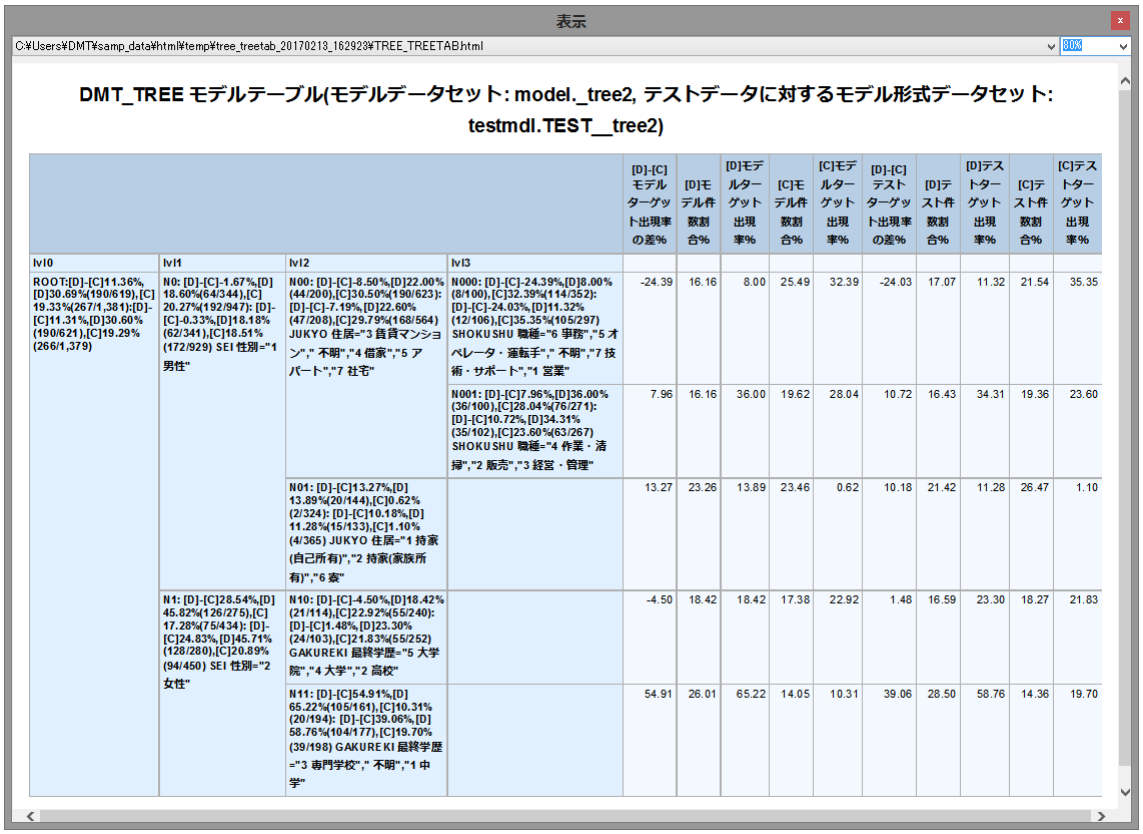
分析が実行され、しばらくすると終了します。
作成されたモデルが _tree2 に保存されます。

結果表示 を押します。

分類木アップリフトモデルの場合、ツリー分岐表、アップリフトチャート、比較プロットが表示可能です。

ツリー分岐表 の表示

3.2.5 アップリフトツリーモデルの表示(ツリー分岐表)



DMT_TREE モデルテーブル(モデルデータセット: model._tree2, テストデータに対するモデル形式データセット: testmdl.TEST__tree2)

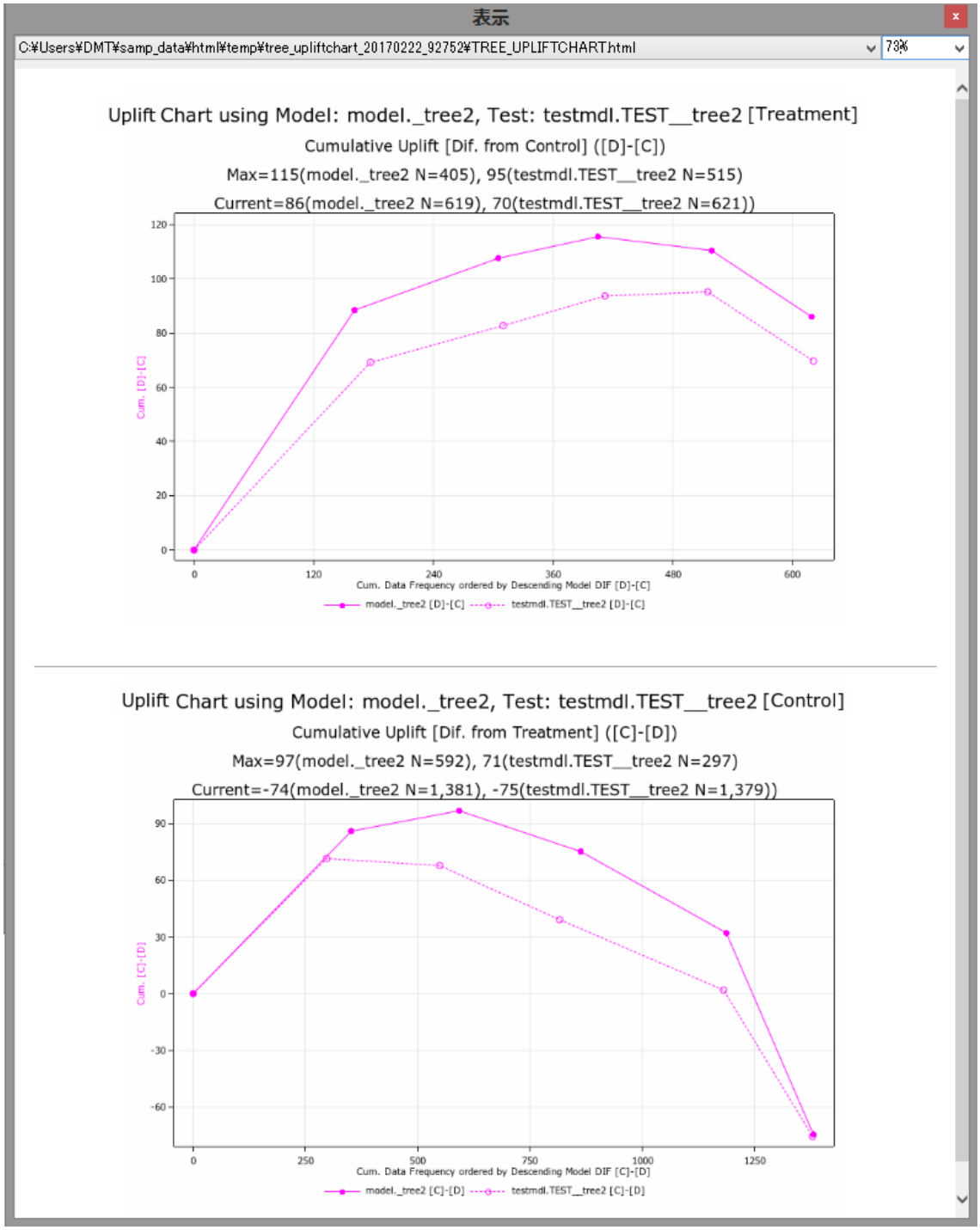
				[D]-[C]モデルターゲット出現率の差%	[D]モデル件数割合%	[D]モデルターゲット出現率%	[C]モデル件数割合%	[C]モデルターゲット出現率%	[D]-[C]テストターゲット出現率の差%	[D]テスト件数割合%	[D]テストターゲット出現率%	[C]テスト件数割合%	[C]テストターゲット出現率%
lvl0	lvl1	lvl2	lvl3										
ROOT:[D]-[C]11.36%,[D]30.69%(190/619),[C]19.33%(267/1,381):[D]-[C]11.31%,[D]30.60%(190/621),[C]19.29%(266/1,379)	N0: [D]-[C]-1.67%,[D]18.60%(64/344),[C]20.27%(192/947): [D]-[C]-0.33%,[D]18.18%(62/341),[C]18.51%(172/929) SEI 性別="1 男性"	N00: [D]-[C]-8.50%,[D]22.00%(44/200),[C]30.50%(190/623): [D]-[C]-7.19%,[D]22.60%(47/208),[C]29.79%(168/564) JUKYO 住居="3 賃貸マンション"," 不明","4 借家","5 アパート","7 社宅"	N000: [D]-[C]-24.39%,[D]8.00%(8/100),[C]32.39%(114/352): [D]-[C]-24.03%,[D]11.32%(12/106),[C]35.35%(105/297) SHOKUSHU 職種="6 事務","5 オペレータ・運転手"," 不明","7 技術・サポート","1 営業"	-24.39	16.16	8.00	25.49	32.39	-24.03	17.07	11.32	21.54	35.35
			N001: [D]-[C]7.96%,[D]36.00%(36/100),[C]28.04%(76/271): [D]-[C]10.72%,[D]34.31%(35/102),[C]23.60%(63/267) SHOKUSHU 職種="4 作業・清掃","2 販売","3 経営・管理"	7.96	16.16	36.00	19.62	28.04	10.72	16.43	34.31	19.36	23.60
		N01: [D]-[C]13.27%,[D]13.89%(20/144),[C]0.62%(2/324): [D]-[C]10.18%,[D]11.28%(15/133),[C]1.10%(4/365) JUKYO 住居="1 持家(自己所有)","2 持家(家族所有)","6 寮"		13.27	23.26	13.89	23.46	0.62	10.18	21.42	11.28	26.47	1.10
	N1: [D]-[C]28.54%,[D]45.82%(126/275),[C]17.28%(75/434): [D]-[C]24.83%,[D]45.71%(128/280),[C]20.89%(94/450) SEI 性別="2 女性"	N10: [D]-[C]-4.50%,[D]18.42%(21/114),[C]22.92%(55/240): [D]-[C]1.48%,[D]23.30%(24/103),[C]21.83%(55/252) GAKUREKI 最終学歴="5 大学院","4 大学","2 高校"		-4.50	18.42	18.42	17.38	22.92	1.48	16.59	23.30	18.27	21.83
		N11: [D]-[C]54.91%,[D]65.22%(105/161),[C]10.31%(20/194): [D]-[C]39.06%,[D]58.76%(104/177),[C]19.70%(39/198) GAKUREKI 最終学歴="3 専門学校"," 不明","1 中学"		54.91	26.01	65.22	14.05	10.31	39.06	28.50	58.76	14.36	19.70

ツリー分岐表 には、ノード分岐に採用された説明変数値と実施群（[D]）、対照群（[C]）間のターゲット出現率の差（[D]-[C]）、そして群別のターゲット出現率、件数割合、ターゲット再現率、ターゲット出現率が分岐ノードごとに表示されます。モデル検証用テストデータを分析画面で指定した場合は、:（コロン）の後に、検証データにおける各統計量も表示されます。また、終端ノードについては、**「ターゲット出現率の差%」**と実施群、対照群別の**「件数割合%」**と**「ターゲット出現率%」**の値がモデル作成用データおよびテストデータ別に右側に表示されます。

実施群と対照群間の出現率の差は、クロス分析で見たように最も関連性が高い **性別** の違いによって、最初にもたらされ、その値によって 2 つのノードに分岐しています。そして、男性は住居区分と職種、女性は学歴の違いによってそれぞれさらに分岐し、最終的に 5 個のグループ（終端ノード）が生成されています。終端ノードの実施群と対照群間の出現率の差（アップリフト）は -24.39%～54.91% の範囲に分布しています。

3.2.6 ツリーモデルの評価(アップリフトチャート)

アップリフトチャートの表示



アップリフトチャート は横軸にモデルの予測出現率の差が大きい順に実施データ、対照データをそれぞれ並べて、対応するアップリフト（予測出現率の差の累積値＝予測追加出現数）を縦軸にプロットした図です。実施データ（DATA=入力データ）では、予測出現率の差を **施策を実施した場合の予測出現率－施策を実施しなかった場合の予測出現率**（既定では [D]-[C] と表示）と定義し、対照データ（CONTROL=入力データ）では、逆に、施策を実施しなかった場合の予測出現率－施策を実施した場合の予測出現率の差（既定では [C]-[D] と表示）と定義しています。

アップリフトチャートから、以下のことがわかります。

［実施データについて］
・実施データを、すべて実施しなかったとした場合と比

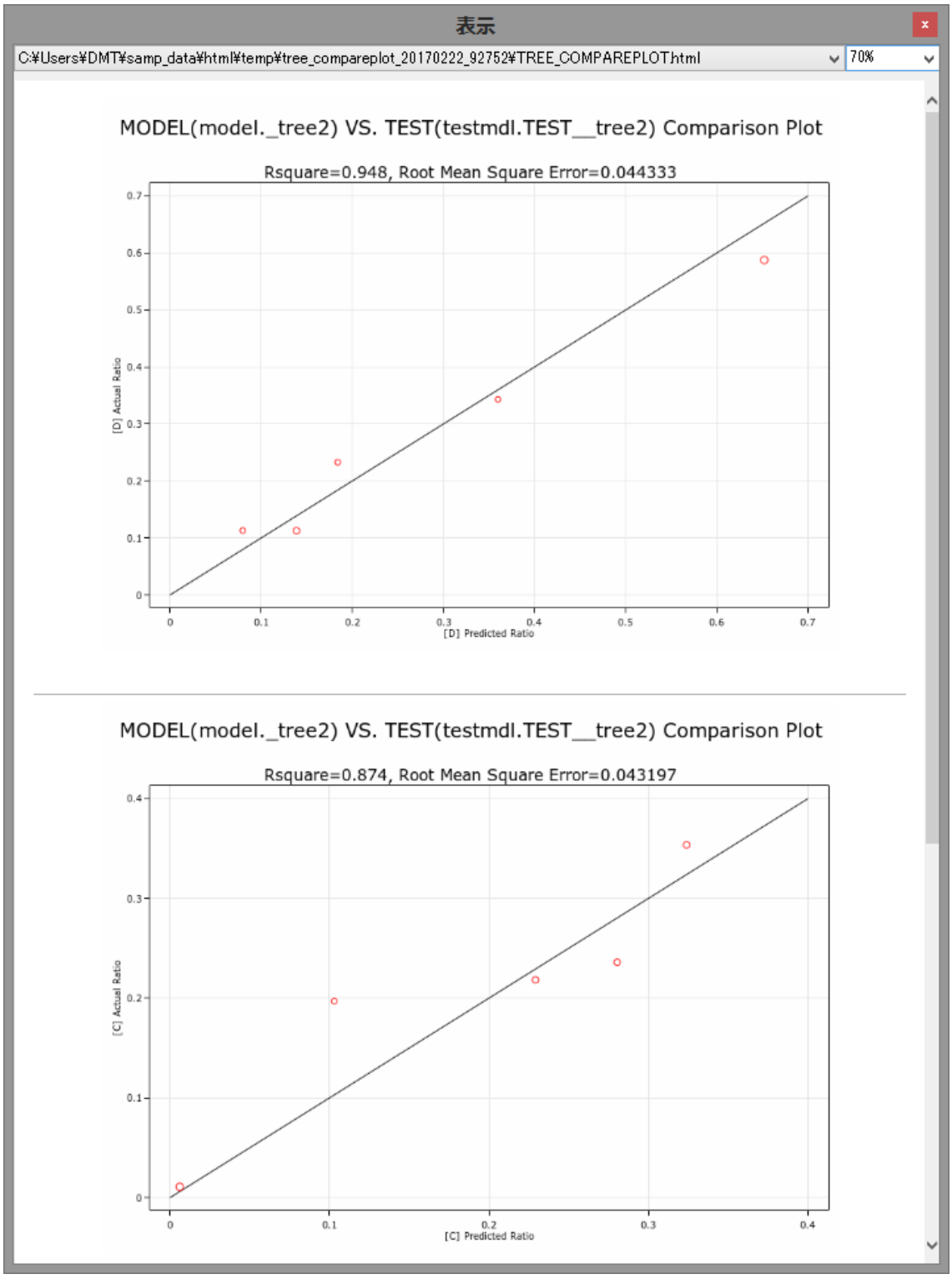
較した、全体の実施効果は、モデル作成データでは +86（619 件）、テストデータでは +70（621 件）と見積もられる。（Current の累積 Uplift 値）
・実施データでは、[D]-[C]の予測値が正の値であったノードのみを実施したとすれば、計算上の最大の実施効果（モデル作成データでは +115（405 件）、テストデータでは +95（515 件））が得られる。（Max の累積 Uplift 値）
・したがって、[D]-[C]の予測値が正の値であったノードのみを実施すれば、モデル作成データでは 115-86=+29、テストデータでは 95-70=+25 だけ現状の全部実施の場合より出現数が増えることが期待されます。

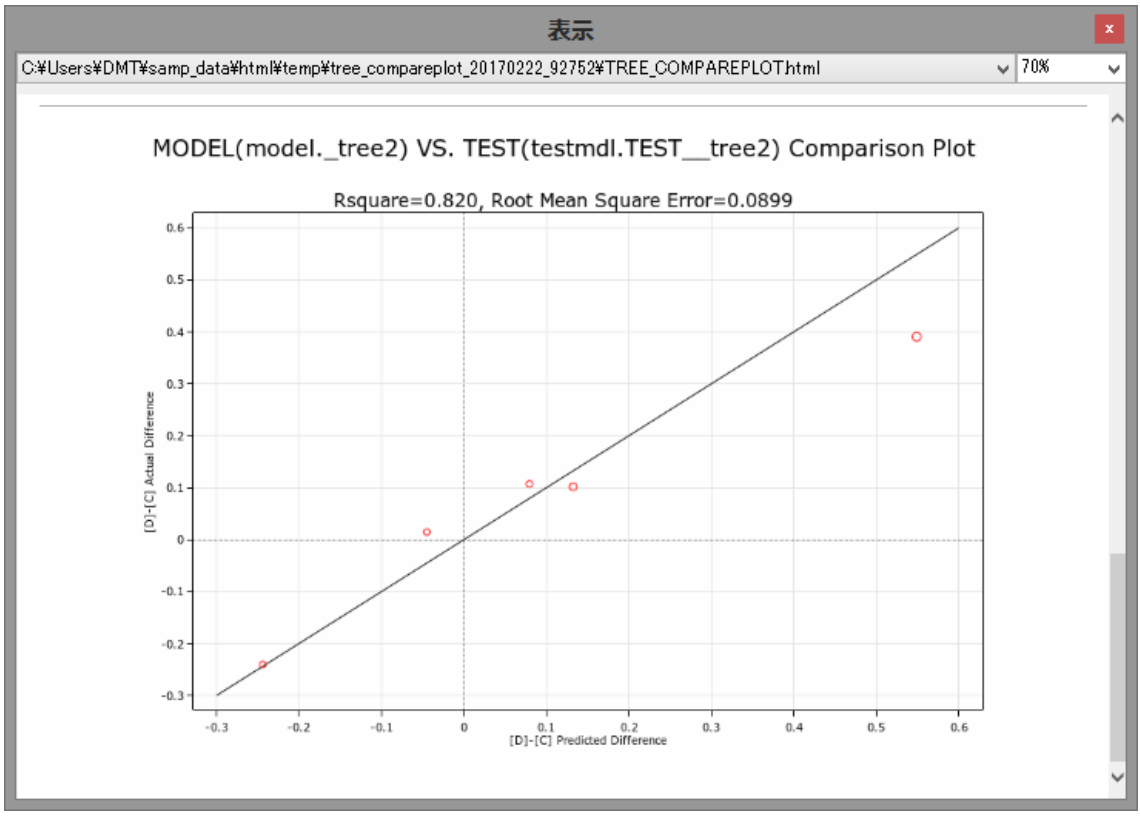
［対照（非実施）データについて］
・対照データを、すべて実施した場合と比較した場合の全体の非実施効果は、モデル作成データでは -74、テストデータでは -75 と見積もられる。（Current の累積 Uplift 値）すべて実施したとすれば、符号を変えた数だけ出現数が増える計算になる。
・対照データでは[C]-[D]の予測値が正の値であったノードのみを非実施とし、残りをすべて実施したとすれば、計算上の最大の非実施効果（モデル作成データでは +97（592 件）、テストデータでは +71（297 件））が得られる。（Max の累積 Uplift 値）
・したがって、[C]-[D]の予測値が正の値であったノードのみを非実施とし、残りをすべて実施したとすれば、モデル作成データでは 90+74=+164、テストデータでは 71+75=+146 だけ現状の全部非実施の場合より出現数が増えることが期待されます。

このように、アップリフトチャートによって、施策実施先を最適化すると、どれだけ出現数が増えるかを見積もることができます。

3.2.7 ツリーモデルの評価（比較プロット）

比較プロット（予測値と実際値の散布図）の表示


表示
C:¥Users¥DMT¥samp_data¥html¥temp¥tree_compareplot_20170222_92752¥TREE_COMPAREPLOThtml
70%
MODEL(model._tree2) VS. TEST(testmdl.TEST__tree2) Comparison Plot
Rsquare=0.948, Root Mean Square Error=0.044333
[D] Actual Ratio
[D] Predicted Ratio
MODEL(model._tree2) VS. TEST(testmdl.TEST__tree2) Comparison Plot
Rsquare=0.874, Root Mean Square Error=0.043197
[C] Actual Ratio
[C] Predicted Ratio



比較プロット はモデルの予測値と実績値の差（誤差）の大きさを評価します。実施データにおける出現率（[D]）、対照データにおける出現率（[C]）、各ノードにおける 2 つの出現率の差（[D]-[C]）を TEST_DATA にモデルを当てはめた場合の値と比較した散布図がそれぞれ表示されます。[C]の散布図において 1 個のノードが対角線より離れていますが、残りは対角線上の近くにプロットされていますので、検証データにおけるツリーモデルの予測値は実績値に近かったことがわかります。

「デシジョンツリーモデル作成」 画面を終了し、**「メニュー」**画面 に戻ります。

4. アルゴリズム

4.1 ノード分割アルゴリズム

DMTデシジョンツリーは、親ノード集団を2つの子ノード集団に分割する処理を繰り返し行います。
PRECAT=Y 指定（デフォルト）の場合は、以下の（１）を最初に一度だけ実行し（２）～（４）の処理を親ノード集団ごとに行います。
PRECAT=N 指定の場合は、以下の（１）～（４）を親ノード集団ごとに行います。

（１）　数値説明変数のカテゴライズ
（２）　AIC基準による候補分岐変数の決定
（３）　候補分岐変数のカテゴリ分類パターンの計算
（４）　最小ノード件数を満たす候補分岐変数とカテゴリ分類パターンの選択

4.1.1 数値説明変数のカテゴライズ

説明変数ごとに、スタージェスの公式を用いて、欠損で無いオブザベーション件数Nに対する階級数を決定します。この階級数以下の種類数の値しか持たない数値変数は、個々の存在する値そのものが個々のカテゴリとして定義されます。なお、nomergen=パラメータに任意の数値を指定することにより、個々の値を個々のカテゴリとする階級数の上限を明示的に与えることも可能です。スタージェスの公式、もしくは nomergen=パラメータにより明示的に与えられた階級数を超える種類数の値を持つ数値変数は、件数Nを階級数で除した1階級平均件数を必要件数とし、カテゴライズのしきい値を、最小値の方から上記1階級必要件数に達するまでを1つのカテゴリとして逐次決定していきます。

Nを欠損を除くオブザベーション件数、log2()を2を底とする対数関数、ceil()を整数値への切り上げ関数とすると、以下の計算式により1階級必要件数を決定しています。

階級数=ceil(1+log2(N))
1階級必要件数=ceil(N/階級数)

なお、最後のカテゴリが1階級必要件数に達していない場合、1つ前のカテゴリに併合するかどうかを選択できます（lastcatm=Y/Nオプション）。デフォルトは併合しない（lastcatm=N）設定です。したがって、当該数値説明変数にタイが全く存在しない場合は、1番目から最後から1つ前のカテゴリの該当件数はすべて等しくなり、最後のカテゴリのみそれ以下の該当件数を持つことになります。（タイが存在する場合は各カテゴリの該当件数は等しくなりません。）
この数値説明変数のカテゴライズ処理はターゲット変数とは無関係に行われます。

4.1.2 欠損が多い説明変数のカテゴライズについて

lastcatm=N（デフォルト）の場合は、欠損でないオブザベーション件数が1階級必要件数に満たない変数でも欠損と1つの有効な値の範囲を示すカテゴリの2つが生成されます。しかし、DMTデシジョンツリーのノード分割アルゴリズムでは、数値タイプ説明変数については、有効な値で作成されるカテゴリ数が2個以上存在しないと、その説明変数は分析から除外するように制御しています。（数値説明変数の場合は「欠損」と「それ以外」というツリー分岐が発生しないようにするため。文字タイプ変数の場合は、常に欠損は有効なカテゴリとして取り扱うため、「欠損」と「それ以外」というツリー分岐が発生する可能性があります。）なお、lastcatm=Yとすると、欠損でないオブザベーション件数が2階級必要件数に満たない変数は分析対象から除外されます。もしも欠損が多い数値タイプ説明変数の欠損と欠損以外の違いに意味があると考える場合は、欠損とそれ以外という2つの値を持つ文字タイプ説明変数を作成して、その変数を分析に用いるようにしてください。

4.1.3 AIC 基準による候補分岐採用説明変数の決定

ターゲット値の出現率を予測する分類木の場合は、ターゲット変数（y=パラメータ）を、ターゲット値（target=パラメータを満たす値）と非ターゲット値の2値変数とみなして、これと個々の説明変数（x=パラメータ）との間の分割表モデルにおける統計的関連性をAIC値により評価します。

ターゲット変数の値を予測する回帰木の場合は、ターゲット変数（y=パラメータ）を目的変数、個々の説明変数（x=パラメータ）を処理変数とみなした一元配置分散分析モデルにおけるAIC値により評価し

ます。

アップリフトモデルの場合は、まず、各変数の実施群と対照群の各カテゴリの出現率または平均値を、親ノードにおいては差が無くなるように調整します。その上でカテゴリ単位の実施群と対照群間の出現率の差、または平均値の差の有意性に関するAIC値（個別AIC値）を算出します。説明変数ごとのAIC値は、個別AIC値を変数単位に合算した値-2 と定義し評価に用いています。

AIC値最小、すなはち、最もターゲット変数の分布と関連が高いとみなされた説明変数を親ノード集団を2つの子ノードに分岐させる第一候補説明変数、2番目に関連が高いとみなされた説明変数を第2候補説明変数、 ... k番目に関連が高いとみなされた説明変数を第k候補説明変数というように決定します。ただし、いずれの候補もAIC値が負であることを条件とします。
なお、nomergen=パラメータを用いて数値説明変数のカテゴライズ方法をカテゴリ数が多くなるように指定すると、AIC値が上昇するため、分岐説明変数に採用されにくくなります。

4.1.4 2 分岐属性値範囲の決定

4.1.3で選択された候補説明変数が**文字タイプ変数**の場合は、標準では**「名義尺度」**（個々の値の並び方に制約がまったく無い尺度）とみなして、ターゲット比率の大きさ、もしくはターゲット平均値の順にすべてのカテゴリを並べておいた上で、2つに分ける最適位置をエントロピー最小基準、偏差平方和最小基準、またはAIC値基準により探索します。k個のカテゴリが存在する場合、k-1通りの計算結果を比較することになります。
しかし、ordinalx=パラメータに指定した文字タイプ変数については**「順序尺度」**（ソート順に値が並ぶという隣接制約がある尺度）、cyclicx=パラメータに指定した文字タイプ変数については**「循環尺度」**（個々の値にはソート順の隣接制約があるが、両端の値の間にも隣接関係があるとする尺度）とみなして以下のように処理しています。「順序尺度」の場合は、値のソート順にカテゴリを並べておいた上で、2つに分ける最適位置を探索します。k個のカテゴリが存在する場合、名義尺度の場合と同じく、k-1通りの計算結果を比較することになります。「循環尺度」の場合は、値のソート順にカテゴリを並べておいた上で、（あ）2つに分ける場合、（い）3つに分けた上で1番目と3番目の併合カテゴリと2番目のカテゴリに2分する場合の可能な全パターンを計算した上で、最適な分割方法を探索します。k個のカテゴリが存在する場合、(k-1)+{1+2+...+(k-2)}通りの計算結果を比較することになります。

ただし、文字タイプ説明変数の場合の欠損カテゴリは単なる1個のカテゴリとして、有効なカテゴリと同列に取り扱います。（これは順序尺度、または循環尺度の指定の場合でも同様です。）なおカテゴリ数が一定の値（デフォルトはmaxcatn=1000）を超える異なる値を持つ文字タイプ説明変数は分析対象から除外されます。

候補説明変数が**数値タイプ変数**の場合は、標準（splitpts=2）では**「循環尺度」**とみなし、（１）の方法でカテゴライズされた値のリストを（あ）2つに分けた場合、（い）3つに分けた上で1番目と3番目の併合カテゴリと2番目のカテゴリに2分し、さらに欠損値が存在する場合は分岐後のどちらかのノードに含むことを考慮した上で、可能な全パターンを計算し、分岐後のターゲット比率に関してエントロピー最小となるパターンを探索しています。（１）の方法によりk個のカテゴリを持つようにカテゴライズされた数値変数の場合、(k-1)+{1+2+...+(k-2)}通りの計算結果を比較することになります。しかし、ordinalx=パラメータに指定した数値タイプ変数については**「順序尺度」**とみなして（あ）の方法による併合パターンのみを探索するよう指定することも可能です。ordinalx=パラメータに指定した数値タイプ変数は(k-1)通りの計算結果のみを比較することになります。

ここで、splitpts=1 と指定すると、全数値タイプ説明変数を標準では**「順序尺度」**とみなして（あ）の方法による併合パターンのみを探索するよう切り替わります。そして、cyclicx=パラメータに指定した数値タイプ変数については**「循環尺度」**とみなし（あ）と（い）の両パターンの計算結果を比較して併合パターンを決定します。

4.1.5 最小ノード件数を満たす分岐説明変数の選択

第一候補説明変数の各カテゴリを2つのノードに振

り分けたとき、2つのノードが共に最小件数基準（mincnt=パラメータ）を満たす[4]場合は、その分け方をノード分岐方法として採用します。もしも、その分け方がノード最小件数基準を満たさない場合は、ノード最小件数を満たす分け方が存在すれば、その中で最適な分け方を保存しておきます。

次に、第二候補説明変数について同様の計算を行います。第二候補説明変数の分け方が最小件数基準を満たす場合、第一候補説明変数で保存しておいた分け方が存在すればそれと比較を行い、良い方の分け方をノード分岐方法として採用します。第一候補説明変数で保存しておいた分け方が存在しない場合は、第二候補説明変数の分け方を採用します。
第二候補説明変数の分け方が最小件数基準を満たさない場合は、第一候補説明変数で保存しておいた分け方と比較を行い、良い方を保存しておき、第三候補説明変数について同様の計算を行います。

AIC値<0の条件を満たす候補説明変数が尽きるまで以上の計算を行い、最後に保存されていた分け方が存在する場合、その分け方を分岐方法として採用します。存在しない場合は「分岐不能」として親ノードを終端ノードにします。

4.2 終端条件

2種類の終端条件が設定可能です。
（１）　ノード最小件数（mincnt=パラメータ）
（２）　分割の最大階層（maxlvl=パラメータ）

4.2.1 ノード最小件数（mincnt=パラメータ）

mincnt=パラメータの値は正の整数（1～n）、またはキーワードAUTO（デフォルト）です。
mincnt=AUTOとは、分類木モデルの場合、分岐先の2つの子ノードの該当件数をそれぞれN1,N2、ターゲット出現率をそれぞれp1,p2とすると、以下の条件を満たす分岐候補説明変数が全く存在しない場合に親ノードを終端ノードに設定します。

SQRT{p1*(1-p1)/N1}<=err_rate*p1　かつ
SQRT{p2*(1-p2)/N2}<=err_rate*p2

これらの式の左辺は、それぞれ、N1個、N2個の抽出データ上で観測されたターゲット出現率p1,p2を、それぞれのノード定義における母集団出現率の推計値とした場合の標準誤差を表しています。これらの標準誤差が右辺の観測比率pのerr_rate倍以内に収まるようなノード件数N1,N2になっているかどうかをチェックします。
上式をN1,N2についてそれぞれ解くと、

N1>=p1*(1-p1)/(err_rate*p1*err_rate*p1)
N2>=p2*(1-p2)/(err_rate*p2*err_rate*p2)

を同時に満たす2つの子ノードのみを生成します。
（出現率の差に関するアップリフトモデルの場合は実施群、対照群ともに上記条件を満たす必要があります）
(ただし p1>0.999のときp1=1,p1<0.001のときp1=0.001.　p2も同様)

なお、err_rateは 0<err_rate<1 の範囲で指定可能です。

回帰木モデルの場合は、分岐先の2つの子ノードの該当件数をN1,N2、ターゲット平均値をm1,m2、ターゲット標準偏差をs1,s2、許容誤差率をerr_rate（ERR_RATE=パラメータで指定します）とすると、以下の条件を満たすノードのみを生成します。

s1/SQRT(N1)<=err_rate*m1　かつ
s2/SQRT(N2)<=err_rate*m2

これらの式の左辺は、それぞれ、N1個、N2個の抽出データ上で観測されたターゲット平均値m1,m2の標準誤差を表しています。この標準誤差が右辺の観測平均値mのerr_rate倍以内に収まるようなノード件数N1,N2になっているかどうかをチェックします。
上式をN1,N2についてそれぞれ解くと、

N1>=s1*s1/(err_rate*err_rate*m1*m1)
N2>=s2*s2/(err_rate*err_rate*m2*m2)
(ただし |m1|<0.001のとき|m1|=0.001.　m2も同様)

しかしながら、上式では、s1=0,s2=0なら
N1>=0,N2>0となってしまうので、

M1>=max(N1,OYA_N/10,10)

[4] アップリフトモデルでは、実施群、対照群の両方で最小件数基準を満たす必要があります。

M2>=max(N2,OYA_N/10,10)

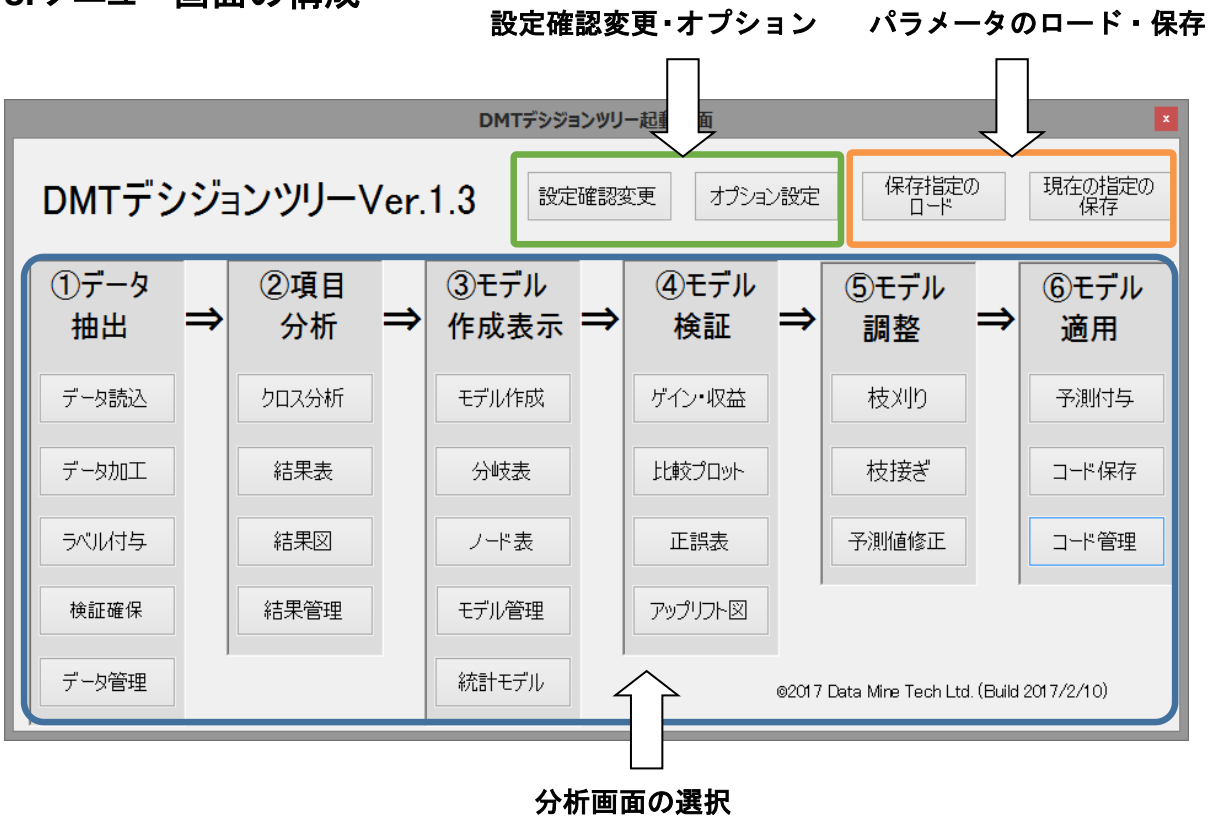
ただし、N1,N2は上記算式による、OYA_Nは親ノード件数です。このM1,M2を同時に満たす2つの子ノードのみを生成します。（平均値の差に関するアップリフトモデルの場合は実施群、対照群ともに上記条件を満たす必要があります）

デフォルトは、mincnt=AUTO,err_rate=0.1 に設定しています。

4.2.2 分割の最大階層（maxlvl=パラメータ）

分割の最大階層に達したノードは強制的に終端ノードになります。maxlvl=パラメータは1～20 の範囲の整数で指定できます。デフォルトは、maxlvl=5 です。

5. メニュー画面の構成



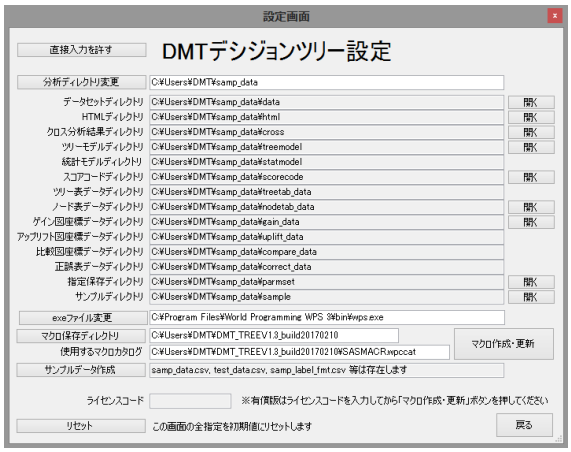
「メニュー」画面 には、**設定確認変更ボタン、オプション設定ボタン、パラメータのロードボタン・パラメータの保存ボタン、** そして、 **分析画面選択ボタン** が配置されています。

5.1 設定確認変更

設定確認変更 を押すと、GUI実行モードでDMTデシジョンツリーを実行するために必要な初期設定と設定変更を行う「設定画面」が開きます。
メニュー画面を最初に起動した場合は、以下のように 初期設定が必要 ボタンのみが表示されています。

初期設定方法については、導入方法をご参照ください。

GUI実行モードのセットアップ方法 に従って初期設定完了後に 実行例 を実行後に「設定画面」を開くと、以下のようになっています。



5.1.1 直接入力を許す

直接入力を許す ボタンを押すとボタン表示が 直接入力を禁止する に切り替わり、設定すべき 分析ディレクトリ設定 、 exeファイル設定 、 マクロ保存ディレクトリ の 3 箇所のディレクトリまたはファイルパス名がテキストボックスに直接入力可能になります。
（※ システムの制限等の理由により、各ボタンを押しても、ディレクトリやファイル選択ウィザードが開かない場合に有用です。）

テキストボックスにパス名を入力後、右に出現する 設定 ボタンを押すと入力が確定され、存在がチェックされます。
すべての入力設定完了後に、直接入力を禁止する を押すと、元の状態に戻ります。

5.1.2 分析ディレクトリの変更

分析ディレクトリ変更 を押すと、現在の分析ディレクトリを新しい分析ディレクトリ、または、既存の別の分析ディレクトリへ切り替えることができます。

新たなデータ分析を行う場合は、新しい分析ディレクトリを作成し、その中に分析結果を保存すると良いでしょう。

既存の分析ディレクトリに変更すると、その分析ディレクトリに保存されているすべての実行結果ディレクトリと _LASTSAVE_ パラメータが自動的にセットされ、その分析ディレクトリで実行した最後の状態から分析を継続することができます。

5.1.3 exe ファイルの変更

exeファイル変更 を押すと、導入されている SAS または WPS の実行ファイル（sas.exe または wps.exe）を変更できます。
ファイル選択画面は C:¥Program Files ディレクトリを初期ディレクトリとして開きます。

通常、sas.exe ファイルは、C:¥Program Files¥SASHome¥SASFoundation¥9.x¥sas.exe（ここで、9.x は SAS バージョンを表します）にあり、wps.exe は、C:¥Program Files¥World Programming WPS 3¥bin¥wps.exe にあります。ファイル選択画面のディレクトリパスを辿って指定します。ただし、インストール時の設定によって、実際の exe ファイルのパスは異なる場合もあります。

ただし、SAS から WPS へ、または WPS から SAS へ exc ファイルを変更する場合は リセット を行ってから行う必要があります。（※ 既存の分析フォルダーは SAS または WPS のいずれかを用いるかを決定済みであるため）

5.1.4 マクロ保存ディレクトリ

マクロ保存ディレクトリ ボタンを押すと、**DMT デシジョンツリーマクロカタログ** を保存するディレクトリを変更できます。
ディレクトリ選択画面は、**C:¥users¥ユーザプロファイル名** を初期ディレクトリとして開きます。

指定したディレクトリ内に、コンパイル済みマクロカタログ（sasmacr.sas7bcat または SASMACR.wpccat）が存在すれば、使用するマクロカタログ に自動的に指定されます。
（注意： もしもそのマクロカタログが本メニューの マクロ作成・更新 で作成されたもので無い場合は、別のディレクトリを指定し、新たにマクロカタログを作成し、それを用いてください。）

5.1.5 マクロ作成・更新

マクロ作成・更新 ボタンを押すと、本メニューに組み込

まれているソースプログラムからコンパイル済みマクロカタログが生成され、マクロ保存ディレクトリ内に保存されます。(既存のものは上書き)

注意： 本アプリケーションの最新のビルドを入手したときは、ファイルのコピー 記載の方法により、ファイルを適切な場所に解凍・保存した後、ショートカットのリンク先を最新版の "DMT デシジョンツリーV1.3.exe" に変更するなどの方法により、最新ビルドの DMT デシジョンツリーV1.3 を起動し、既存の分析ディレクトリ、exe ファイル、マクロ保存ディレクトリをそれぞれ指定した後、 マクロ作成・更新 を押して、マクロカタログも最新版に更新してください。(※ 動作を確認した後、古いビルドファイルは削除します。)

5.1.6 サブディレクトリを開く

分析ディレクトリの各サブディレクトリ内にディレクトリやファイルが存在する場合、ディレクトリパスが表示されたテキストボックスの右側に 開く ボタンが現れます。 開く を押すと、該当ディレクトリの中に保存されているディレクトリやファイルを確認できます。

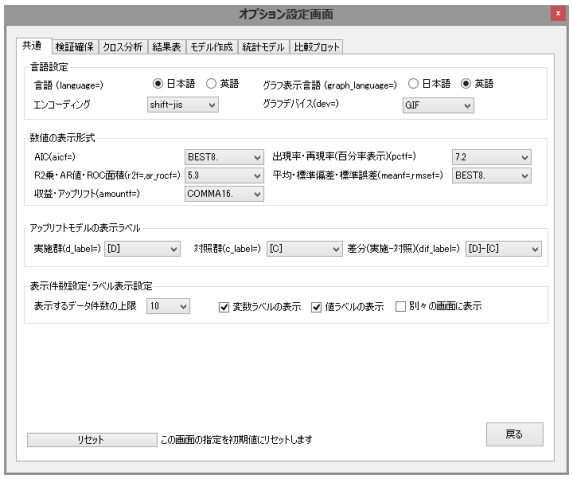
※ 起動画面の「データ管理」、「結果管理」、「モデル管理」、「コード管理」メニューでサポートしていない、複数アイテムをまとめて削除することが可能です。

注意：
①内容を削除する場合は、ディレクトリ単位で保存されているものはディレクトリ単位で削除してください。ディレクトリ内の一部のファイルのみ削除すると、動作しなくなります。
②ディレクトリ名やファイル名の変更は動作しなくなる原因になりますので、行わないでください。

5.2 オプション設定

オプション設定 を押すと、各分析画面で共通なオプションと各分析画面でのみ有効なオプションの設定値の確認と変更ができます。

5.2.1 共通オプション



5.2.1.1. 言語(language=)

すべてのマクロ分析モジュールの共通パラメータ language= の値を指定します。実行ログメッセージ、実行結果画面項目名に表示する言語（日本語か英語のいずれか）を選択します。language=JAPANESE が既定です。

5.2.1.2. グラフ表示言語(graph_language=)

グラフィック出力を行うマクロ分析モジュール（dmt_crossplot, dmt_gainchart, dmt_compareplot.dmt_upliftchart）の共通パラメータ graph_language= の値を指定します。グラフィック出力画面に表示する既定のタイトルや軸ラベル等に表示する言語として用いられます。graph_language=ENGLISH が既定です。※ 現行 WPS ではグラフ上に日本語が表示できませんので、graph_language=ENGLISH の設定を変更しないでください。
※分析データに日本語ラベルや文字変数ラベルが定義されている場合は、グラフテキストのカナ化けを避けるため、nolabel=Y オプションを指定します。

5.2.1.3. エンコーディング

SAS または WPS の実行コマンド wps.exe に付随する -encoding オプションを指定します。※ 現行では -encoding shift-jis 以外はサポートしていません。将来、utf-8 エンコードに対応する予定です。(時期未定)

5.2.1.4. グラフデバイス(dev=)

グラフィック出力を行うマクロ分析モジュールの共通パラメータ dev= の値を指定します。dev=GIF が既定です。それ以外には dev=JPEG を指定できます。

5.2.1.5. 数値の表示形式

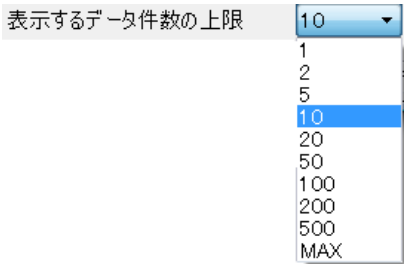
aic 値(aicf=)、出現率・再現率(pctf=)、R2 乗 (r2f=)、AR 値・ROC 面積(ar_rocf=)、平均・標準偏差・平均値の標準誤差(meanf=)、収益・アップリフト(amountf=)をそれぞれ指定します。

5.2.1.6. アップリフトモデルの表示ラベル

アップリフトモデルの結果表示に使用する施策実施群、対照群（施策非実施群)、およびその差を意味する表示ラベルを設定します。適用場面に応じてドロップダウンリストから選択します。これ以外の表示が必要であればコマンド実行方式でパラメータ指定してください。

5.2.1.7. 表示するデータ件数の上限

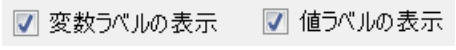
各分析画面で 表示 ボタンを押すと表示されるデータの最大表示件数を設定します。



既定は 10 です。データセットのコンテンツやデータ値は HTML 形式で表示されます。この設定値を大きくすると表示に時間がかかります。

5.2.1.8. 変数ラベルの表示、値ラベルの表示

データセットのコンテンツやデータ値の表示時に変数ラベル、値ラベル（本システムの **「ラベル付与画面」**画面で、分析データセットの文字変数値に１対１に対応させて定義した出力フォーマット）を使用するかどうかをそれぞれ設定します。

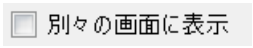


既定はいずれも使用するに設定しています。

※ この設定は、本システムの**「ラベル付与画面」**においてデータセットに定義された変数ラベル、値ラベルが存在するデータセット表示の場合にのみ適用されます。

5.2.1.9. 別々の画面に表示

各分析画面において、表示 ボタンや 前回表示 ボタンを押したときに出現する分析結果表示画面の操作モードを選択します。



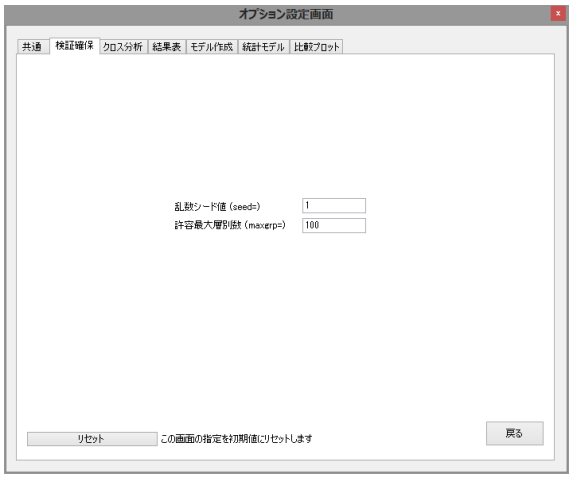
チェックが外れた状態で、表示 ボタンや 前回表示 ボタンを押すと、分析結果表示画面が出現しますが、その表示画面を閉じないと次の操作ができないモードです。
チェックが入った状態では、表示 ボタンや 前回表示 ボタンを押すごとに別々の分析結果表示画面が出現し、分析画面と表示画面いずれの画面も操作できるモードになります。ただし、分析画面を閉じると全ての表示画面は閉じられます。

なお、**表示するデータ件数の上限、 変数ラベルの表示、値ラベルの表示、別々の画面に表示** は各分析画面にも配置されており、どの分析画面で変更しても変更効果は残ります。

5.2.2 各分析画面で有効なオプション

以下のオプションは、煩雑さを避ける目的で、各分析画面においては指定できないオプションです。必要に応じて設定値を変更します。

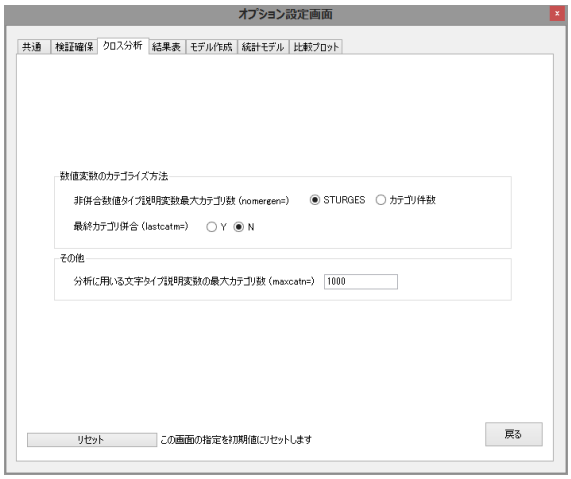
5.2.2.1. 検証確保画面



乱数シード値 (seed=1)
正の整数値を指定すると、同じシード値に対して常に同じコンピュータ乱数系列が生成されます。一方、値0を指定すると、生成されるコンピュータ乱数系列は実行するたびに異なるものとなります。分析結果の再現性を求める場合は、シード値は0以外に指定してください。

許容最大層別数 (maxgrp=100)
非常にたくさんのカテゴリを持つ層別変数を誤って指定した場合に実行を行わないようにするためのオプションです。指定の値を超える場合はエラーとして分析を中断します。問題がない場合は、値を大きくして再実行してください。

5.2.2.2. クロス分析



非併合数値タイプ説明変数最大カテゴリ数 (nomergen=STURGES)
個々の数値タイプ説明変数のカテゴライス方法に関して、欠損値を除いた値の種類数がこの値以下の場合、その数値説明変数は個々の値をカテゴリとみなすように指定します。デフォルトはスタージェスの公式により計算された値です。

CEIL(1+log2(N))

ただし、CEILは整数値への切り上げ関数、log2は2を底とする対数関数、Nは欠損値を除くデータ件数を表します。

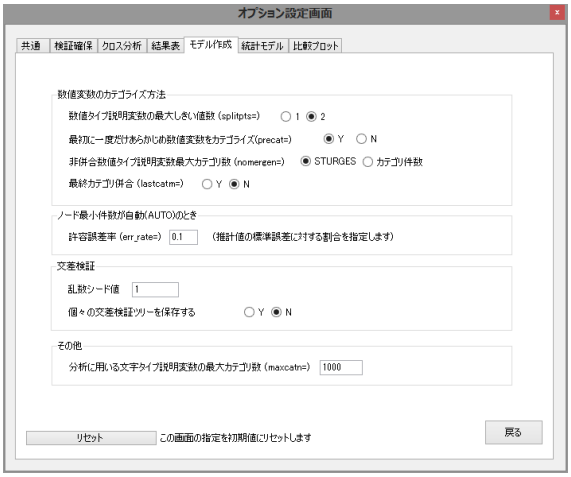
最終カテゴリ併合 (lastcatm=N)
数値タイプ説明変数のカテゴライズ方法に関して、最後のカテゴリを最後から2番目のカテゴリに併合するか否かを指定します。デフォルトはN（併合しない）です。
「ノード分割アルゴリズム」の「(１) 数値説明変数のカテゴライズ」に記載したように、一般にタイが存在する数値変数（たとえば年齢）の場合、カテゴライズ結果は最後にカテゴリのみ他のカテゴリより件数がかなり少なくなる可能性があります。そのため最後のカテゴリを1つ前のカテゴリと併合する方がモデルの安定性が高まる場合があります。

分析に用いる文字タイプ説明変数の最大カテゴリ数 (maxcatn=1000)
この指定は文字タイプ変数が単なるオブザベーション識別変数であって分析対象では無いとみなすためのパラメータです。デフォルトは1000。文字タイプ説明変数のカテゴリ数が指定の数を超える場合、その文字タイプ説明変数は分析対象から除外されます。

5.2.2.3. 結果表

全体平均値の表示 (no0=Y)
ターゲット値の全体出現率またはターゲット変数の全体平均値の集計結果を表す行を最初の行に表示するか否かを選択します。デフォルトは no0=Y （表示する）です。no0=YまたはNを指定します。

5.2.2.4. モデル作成



数値タイプ説明変数の最大しきい値数 (splitpts=2)
数値説明変数が分岐候補説明変数に選択された場合

のカテゴリ併合方法を指定します。1または2を指定できます。(2がデフォルト)。1を指定するとk個のカテゴリを2つに分ける(k-1)通りの併合パターンのみを計算し、採用された場合あるしきい値の前後に分かれることになります。(すべての数値説明変数がデフォルトで順序尺度とみなされます) 2(デフォルト)の場合は、2つに分けるパターンと3つに分けて最初と最後を一緒にするパターンの両方を計算し、最適な併合パターンを探索します.(すべての数値説明変数がデフォルトで循環尺度とみなされます)

最初に一度だけあらかじめ数値変数をカテゴライズ (precat=Y)
分析開始時にあらかじめ1度だけすべての数値タイプ説明変数をまとめてカテゴライズする(Y)か否(N)かを選択します。precat=Y がデフォルト。
precat=N を指定すると、ノード分割を行うたびに数値説明変数のカテゴライズが行われます。precat=N を指定するとモデルの精度が良くなる可能性がありますが、相対的に実行時間が増加します。

非併合数値タイプ説明変数最大カテゴリ数 (nomergen=STURGES)
個々の数値タイプ説明変数のカテゴライス方法に関して、欠損値を除いた値の種類数がこの値以下の場合、その数値説明変数は個々の値をカテゴリとみなすように指定します。デフォルトはスタージェスの公式で計算された値です。

CEIL(1+log2(N))

ただし、CEILは整数値への切り上げ関数、log2は2を底とする対数関数、Nは欠損値を除くデータ件数を表します。

最終カテゴリ併合 (lastcatm=N)
数値タイプ説明変数のカテゴライズ方法に関して、最後のカテゴリを最後から2番目のカテゴリに併合するか否かを指定します。デフォルトはN(併合しない)です。
一般にタイが存在する数値変数(たとえば年齢)の場合、カテゴライズ結果は最後のカテゴリのみ他のカテゴリより件数がかなり少なくなる可能性があります。そのため最後のカテゴリを1つ前のカテゴリと併合する方がモデルの安定性が高まる場合があります。

許容誤差 (err_rate=0.1)
err_rateは mincnt=AUTO 指定の場合に有効です。0<err_rate<1 の範囲で指定可能です。1に近い値を指定することは、分類木モデルでは許容する誤差範囲(標準誤差)を予測値(0から1の範囲であることに注意)と同じ程度に設定することを意味しますので、予測値のブレが非常に大きなモデルが出来てしまう危険性が高くなります。逆に0に近い値を指定することは、相対的に誤差が小さいノードを生成することにつながりますが、ターゲット出現率の値が0または1に近いノードは非常に多くのノード件数が必要となりますので、そのようなノードは生成されにくくなります。
回帰木モデルの場合も平均値の標準誤差が平均値の err_rate 倍に収まるために必要な件数を計算して mincntの値を動的に決定します。
入力データセットの件数があまり豊富で無い場合は、このパラメータ値を大きくするか、mincnt=指定に定数値を指定します。

乱数シード値 (seed=1)
交差検証実行時のデータ分割に用いる乱数シード値を指定します。正の整数値を指定すると、同じシード値に対して常に同じコンピュータ乱数系列が生成されます。一方、値0を指定すると、生成されるコンピュータ乱数系列は実行するたびに異なるものとなります。分析結果の再現性を求める場合は、シード値は0以外に指定してください。

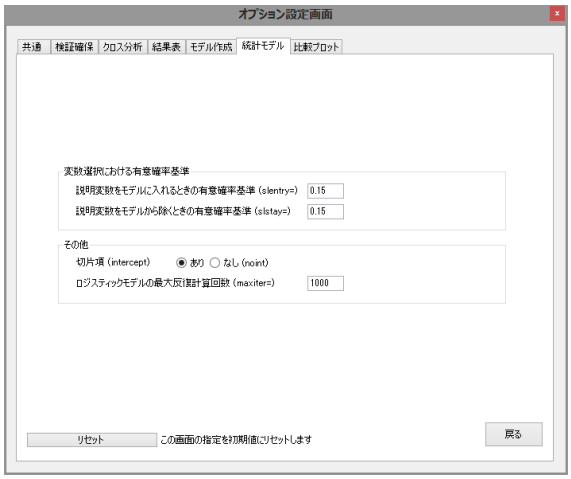
個々の交差検証ツリーを保存 (Y/N)
交差検証実行時に作成されるfold=パラメータ指定数個の個々の交差検証用ツリーモデルをモデル管理画面に登録して参照可能とするか否かを指定します。Nがデフォルトです。デフォルトではoutmodel=パラメータに指定した分析結果出力モデルと出力モデル名の後に _CV の接尾辞のついた検証用モデル形式データセットの2つのツリーモデルが出力されます。

Yを指定すると、上記2つのツリーモデルの他に、出力モデル名の後に _CV1, _CV2, ... , _CVfold (foldはfoldパラメータの値)の接尾辞が付いた個々の交差検証モデルも出力されます。これらの出力ツリーモデルは、モデル分岐表作成やゲインチャート作成など、他のモデルと同様の操作が可能です。

なお、**個々の交差検証ツリーを保存** (Y/N) の指定に関わらず、outmodel=パラメータに指定した分析結果出力モデルが入ったディレクトリ内に以下のデータセットが保存されます。(**「設定」画面**の**「ツリーモデルディレクトリ」**の**「表示」ボタン**から検索することができます。)

分析に用いる文字タイプ説明変数の最大カテゴリ数 (maxcatn=1000)
このパラメータは文字タイプ変数が単なるオブザベーション識別変数であって分析対象では無いとみなすためのパラメータです。デフォルトは1000です。文字タイプ説明変数のカテゴリ数が指定の数を超える場合、その文字タイプ説明変数は分析対象から除外されます。**2～5000の範囲で指定可能です。**

5.2.2.5. 統計モデル



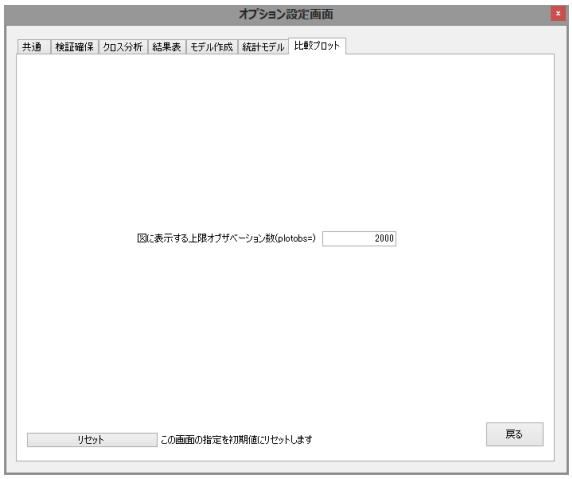
説明変数をモデルに入れるときの有意確率基準 (slentry=0.15)
モデルに含まれていない説明変数の中からモデルに追加するときの有意確率基準を指定します。

説明変数をモデルから除くときの有意確率基準 (slstay=0.15)
モデルに含まれている説明変数の中でモデルから除くための有意確率基準を指定します。なお、slstayはモデルに残るための基準という意味です。

切片項 (intercept)
モデルに切片項パラメータを含むか否かを指定します.（含む（「あり」）がデフォルト）

ロジスティックモデルの最大反復計算回数 (maxiter=100)
最尤法によるパラメータ推計時の最大反復計算回数を指定します。反復回数が十分で無い場合、最尤法によるパラメータ推計は収束に至らない場合があります。変数選択を指定した場合は、
各変数選択段階でのパラメータが収束しないまま、次の変数選択段階に進む場合があります。このような場合、このオプションの値を大きくするとパラメータ推計結果が収束する場合があります。

5.2.2.6. 比較プロット

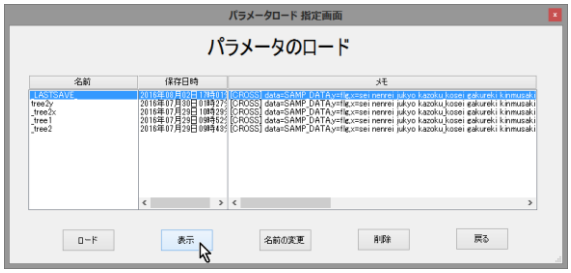


図に表示する上限オブザベーション数 (plotobs=2000)
data= 入力データセットに含まれるデータから図に表示する上限オブザベーション数を正の整数値で指定します。デフォルトは5000です。入力データセットのオブザベーション数がこの上限を超える場合はランダム抽出を行い上限数のデータのみプロットの対象にしています。なお、R2乗値の計算は全オブザベーションから計算しています。

5.3 パラメータのロード･保存

5.3.1 保存指定のロード

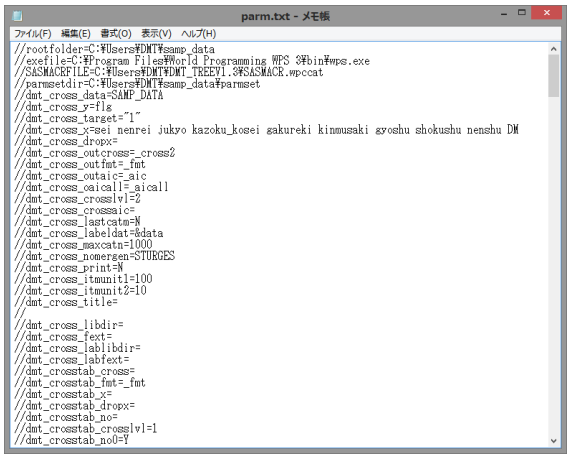
保存指定のロード を押すと、メニュー画面を終了する際に _LASTSAVE_ という名前で自動保存された、または次の 現在の指定の保存 を押して明示的に名前を付けて保存した各分析画面の全部の入力パラメータセットを、ロードまたは名前の変更または削除します。



操作したいパラメータ保存アイテム名をクリックすると、操作ボタンが表示されますので、表示・名前の変更・削除の操作を行います。

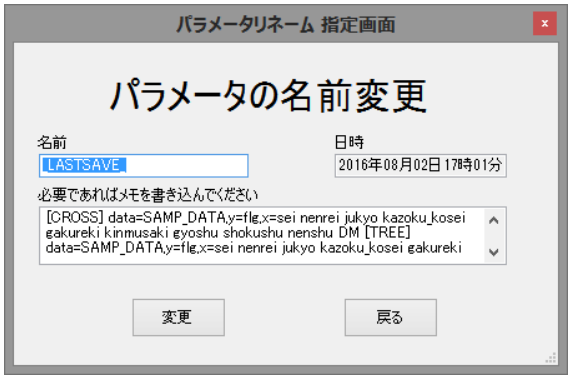
ロード 選択したパラメータセットを有効にします。

表示 選択したパラメータセットの内容をリスト表示します。



パラメータセットを表示した例
設定画面や各分析画面の入力項目ごとに値が保存されています。

名前の変更　選択したパラメータセットの名前を変更します。



パラメータセットの名前変更画面の例
名前とメモの項目を変更可能です。

削除　選択したパラメータセットを削除します。

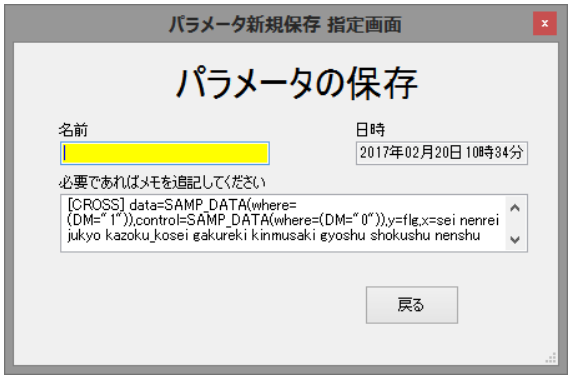
なお、リストボックスの上部にある、

名前　保存日時
メモ

ボタンを押すと、それぞれの項目の並び順にアイテムをソートして表示できます。(押すごとに昇順、降順が切り替わります)

5.3.2 現在の指定の保存

現在の指定の保存　を押すと、最後に実行した状態で残されている、全分析画面の全部の入力パラメータセットを、名前を付けて新規保存します。



※ 既存の名前は指定できません。先に削除してから指定してください。
※ メモ欄には既定でクロス分析[CROSS]で始まる）とツリー作成（[TREE]で始まる）で最後に実行したパラメータが表示され、これを見ると、どのデータでどの指定を行ったかがわかります。メモを入力する場合は、これらの情報は消さないで、末尾に追記することをおすすめします。

5.4 分析ディレクトリのファイル表示

本システムの**「設定」**画面の指定、および各**「分析」**画面の実行により生成されるディレクトリ・ファイル等は以下のとおりです。

「設定」画面において、分析ディレクトリの各ディレクトリ内にファイルやディレクトリが存在する場合は　開く　ボタンが表示されます。必要に応じて、保存されているディレクトリやファイルの確認が可能です。※ htmlディレクトリ内のサブディレクトリは名前の変更や削除は行わないでください。また、cross、data、parmset、scorecode、statmodel、treemodel の各ディレクトリについては､ディレクトリ内に保存された個々のデータやモデルをディレクトリ単位に削除することは問題ありませんが、ディレクトリ内の個々のファイルの名前の変更、削除、内容の編集等を行うとシステムが起動しなくなる恐れがあります。もしも、既存分析ディレクトリのファイル構造に問題が発生したときは、新しい分析ディレクトリを作成し、その中に使用したい既存のデータやモデルデータセットファイルをコピーして用いてください。

以下の図は、**設定画面**で作成した分析ルートディレクトリ（ここでは root と表示）の下に作成されるディレクトリと主要なファイルの一覧を示しています。

分析ディレクトリをルートとするDMTデシジョンツリーGUI実行モードディレクトリ一覧

5.5 各分析画面の処理の流れ

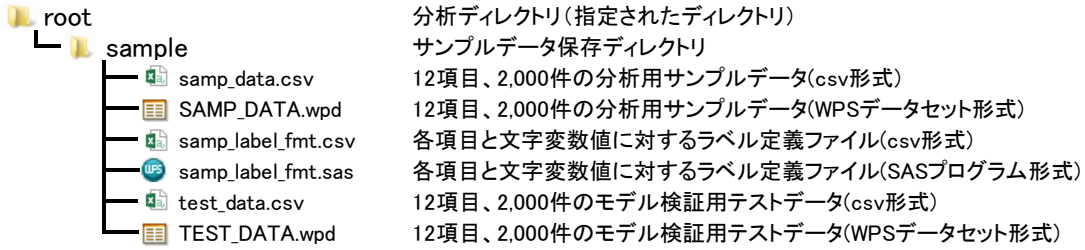
各分析画面にある 実行 ボタンを押すと、画面指定により生成されたSASコードがパラメータとしてシステムに保存され、分析ルートディレクトリの下にある pgm.sas ファイルにコピーされます。pgm.sas ファイルへのSASコードのコピーが終わると、ただちに submit_sas.bat（SAS上で動かす場合）、または、submit_wps.bat（WPSの場合） が起動し、SASサーバまたはWPSサーバがバッチ型動作モードで pgm.sas に書かれたSASコードを実行に移します。なお、incpgm.sas ファイルは pgm.sas ファイルのを %inc コマンドで読み取って実行するように指定した SASステートメント が記述されており、

submit_sas.bat（SASの場合）、または submit_wps.bat（WPSの場合）はいずれも incpgm.sas　をバッチ実行するように指定したバッチファイルです。

5.6 サンプルデータ

「設定」画面において、

サンプルデータの作成　ボタンを押すと、本マニュアルの実行例に示したサンプルデータを、分析ディレクトリの下の SAMPLE ディレクトリに作成します。



サンプルディレクトリのファイル一覧

5.7 分析画面

「起動」画面 から いずれかの**「分析」**画面選択ボタン を押すと、各分析画面に切り替わります。

5.7.1 ①データ抽出

データ読込　CSV形式、またはSASデータまたはWPSデータ形式の分析対象データを本システムで利用できるように、分析ディレクトリに読み込みます。SASデータまたはWPSデータを読み取った段階では、データに定義されている変数ラベルはコピーされますが、変数値に定義されたユーザ定義フォーマットはすべて削除されます。あらためて、ラベル付与　画面で文字変数値に対する1対1のフォーマット（これを値ラベルと呼んでいます。）を定義してください。

データ加工　入力データを加工（変数のタイプ変換、加工変数の作成、オブザベーションの条件抽出など）を行います。
特に、本システムでデータ分析を行う場合は、説明変数のタイプ（文字タイプか数値タイプか）は分析上、また値ラベル付与上重要です。値が少数の離散的な値しかとらないような数値変数があれば、この画面で文字タイプに変換しておくと良いでしょう。

ラベル付与　分析結果を分かりやすく表示するために、変数名と文字変数の個々の値に1対1で対応する説明ラベルを定義します。
特定の決まりで入力されたCSV形式のファイル、もしくは SAS言語のLABELステートメント、FORMATステートメント、FORMATプロシジャのコードを入力に用いることが可能です。

検証確保　分析対象として入力したデータをランダムに2分して、モデル作成用データとモデル検証用データを確保します。

データ管理　システムに保存したデータの名前・作成日時・メモ（作成方法など）を表示し、内容の確認・名前の変更・削除を行います。
なお、データセット名の変更や削除はそのデータセットを参照している他のプログラムやパラメータ値には波及されませんので、ご注意ください。

5.7.2 ②項目分析

クロス分析　ターゲット変数と説明変数間の関連性を分析し、関連の強い順（AIC）に説明変数のカテゴリ（数値変数は範囲）別のターゲット変数の分布（ターゲットがクラス変数の場合は出現率、連続変数の場合は平均・標準偏差）を集計します。ターゲット変数との関連性のみならず、デシジョンツリー

分析に用いる説明変数の状況（カテゴリの存在範囲や件数バランス、欠損値の割合など）を事前チェックすることができます。

結果表　クロス分析結果を表の形で表示します。

結果図　クロス分析結果を図示します。

結果管理　クロス分析結果データセットの名前・作成日時・メモ（作成方法など）を表示し、内容の確認・名前の変更・削除を行います。
なお、結果データセット名の変更や削除はその結果データセットを参照している他のプログラムやパラメータ値には波及されませんので、ご注意ください。

5.7.3 ③モデル作成表示

モデル作成　モデル作成用データを使ってデシジョンツリーモデルを作成します。

分岐表　作成したデシジョンツリーモデルを分岐の仕方がわかる形式で表示します。

ノード表　作成したデシジョンツリーモデルを終端ノードごとの説明変数の組合せ定義や件数比率・ターゲット件数比率（ターゲット再現率）・ターゲット出現率を出現率の大きさの順に並べて表示します。

モデル管理　作成したモデルの名前・作成日時・メモ（作成方法など）を表示し、内容の確認・名前の変更・削除を行います。
なお、モデルデータセット名の変更や削除はそのモデルデータセットを参照している他のプログラムやパラメータ値には波及されませんので、ご注意ください。

統計モデル　モデル作成用データを使って統計モデルを作成します。

5.7.4 ④モデル検証

ゲイン・収益　モデルの予測値の順位と実際のターゲット出現有無との関連の強さを表すゲインチャート（CAP曲線）やROC曲線を描きます。また、損益見込み額を計算する収益チャートを描きます。分類木モデル、または出現率の差を目的変数とする差分分類木（アップリフト分類木）の場合のみ作成可能です。

比較プロット　実際値とモデルの予測値との誤差が把握できる散布図を描きます。分類木、回帰木、アップリフトモデル、いずれの場合でも作成可能です。

正誤表　ターゲットが出現するかしないかの予測と実際の2*2のクロス度数集計表を作成し、正答率を表示します。分類木モデルの場合のみ作成可能です。

アップリフト図　実施データと対照データそれぞれについて、モデル予測値に基づくツリーノード別実施効果（累積アップリフト）を図示します。

5.7.5 ⑤モデル調整

枝刈り　モデルの当てはまりを改善する目的で、当てはまりの悪いツリーモデルの一部を削除し、モデルを簡素化します。

枝接ぎ　モデルの精度や納得性を高める目的で、指定の終端ノードに別のツリーモデルを接ぎ足して、モデルを複雑化します。

予測値修正　モデルの分岐の仕方・形状は変えずに、新たなデータにモデルを適用したときの、ノードごとのターゲット出現率またはターゲット平均値を、新たな予測値とするモデルを作成します。

5.7.6 ⑥モデル適用

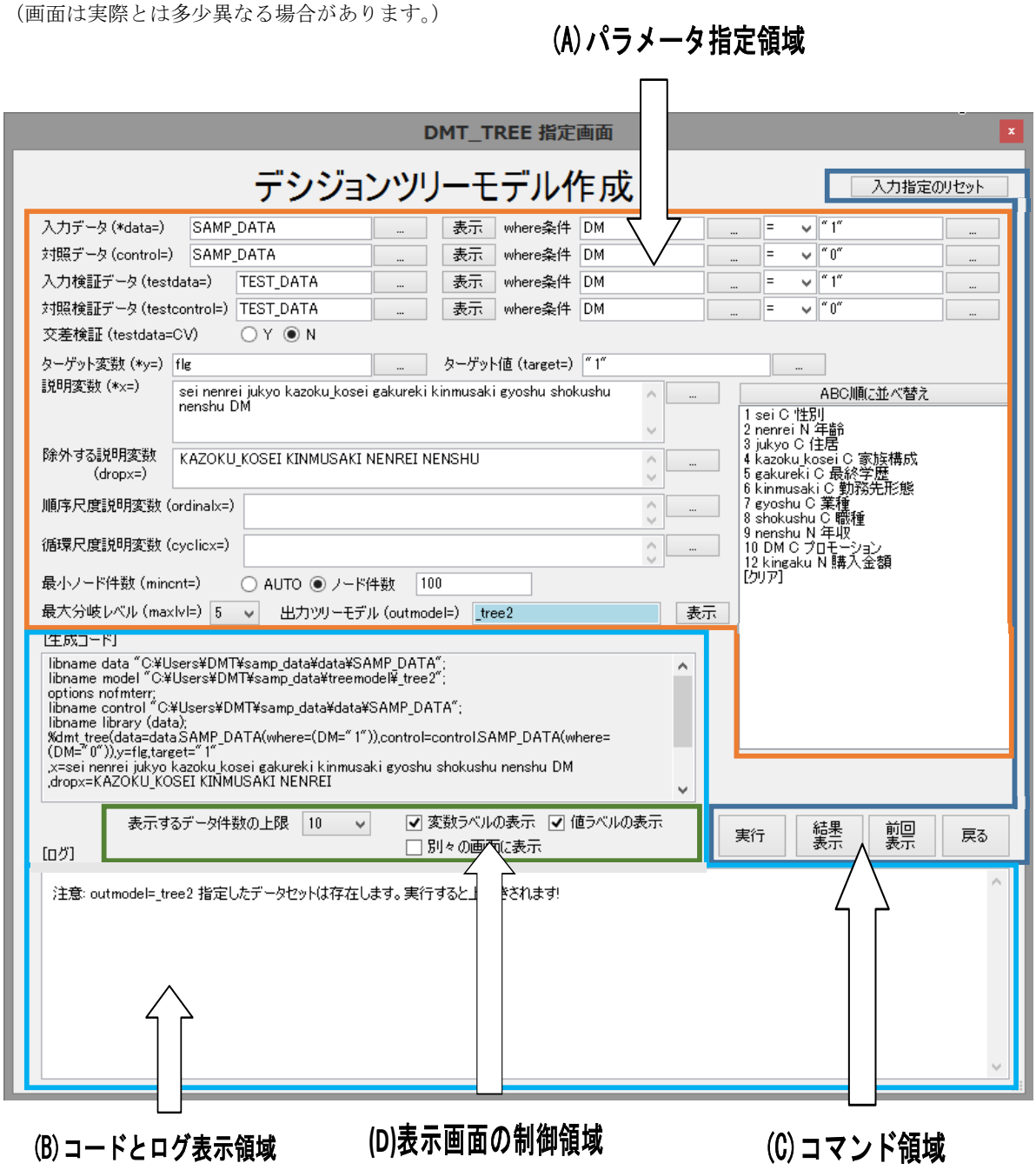
予測付与　モデルをデータに適用し、各オブザベーションに対して、所属ノード番号や予測値（分類木の場合はターゲット出現率、回帰木の場合はターゲット変数平均値）を付与します。

コード保存　モデルからモデル予測値を計算するSASプログラムコードをファイルに出力します。

コード管理　モデル予測値を計算するSASプログラムコードの名前・作成日時・メモ（作成方法など）を表示し、内容の確認・名前の変更・削除を行います。

6. 分析画面の構成

各分析画面は、基本的に、以下の図に示すように、(A) パラメータ指定領域 、(B) コードとログ表示領域 、(C) コマンド領域 、(D) 表示画面の制御領域の 4 つの領域で構成されています。
（画面は実際とは多少異なる場合があります。）



6.1（A）パラメータ指定領域

各分析画面に固有のパラメータを指定する領域です。以下の要素（オブジェクト）が配置されています。

6.1.1 パラメータ(パラメータ名=)

（入力データ(*data=) など） 入力するパラメータの日本語ラベル、および、カッコの中にDMTデシジョンツリープロダクトの該当するマクロ定義の中のパラメータ名を表示しています。カッコの中の*（アスタリスク）で始まるパラメータ名は、マクロ定義における必須入力パラメータを表します。

6.1.2 テキストボックス

（　） パラメータ入力値を表示します。 テキストボックスに直接入力可能な場合と、テキストボックスのすぐ右に配置されている 選択**ボタン**（...） を押して入力しなければならない場合があります。また、テキストボックスの背景色が黄色・赤色・水色の場合は、それぞれ、以下の意味を表します。

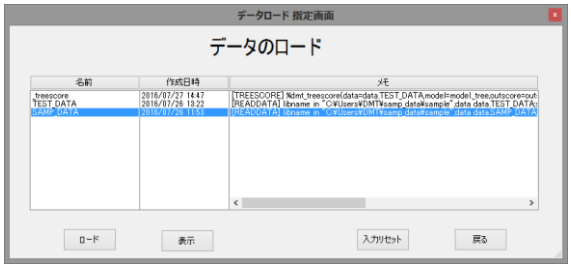
黄色背景（　） パラメータ入力が必須、かつ優先的に入力しなければならないことを表します。
赤色背景（samp_data など） パラメータ値が無効であり、値を変更しなければならないことを表します。
水色背景（_tree など） パラメータ値に指定されたデータが既に存在しており、そのまま実行するとデータの内容が上書きされることを警告しています。同時にデータ内容を表示できることも表しています。

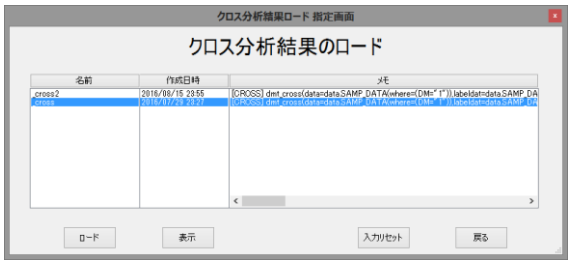
6.1.3 選択ボタン

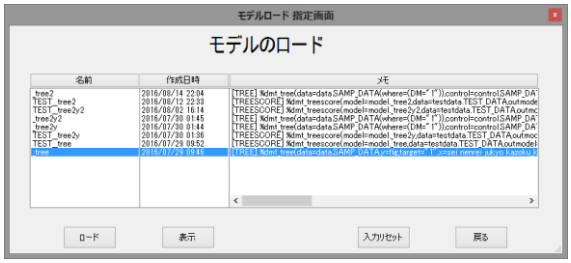
（...） パラメータ入力をボタンで行わなければならない場合、または行える場合に配置されています。押すとデータセットを選択するためのエクスプローラ画面、データやモデルなどをロードする画面、変数や値を選択するリストボックスなどが開きます。

6.1.4 既存のデータやモデルのロード画面

選択ボタン（...）を押すと、システムに保存されているデータ、クロス分析結果データ、ツリーモデルをロードする画面が出現する場合があります。



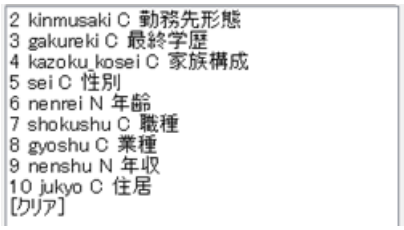




ロード ... 選択したアイテムをロードします。
表示 ... 選択したアイテムの内容を確認します。
入力リセット ... 指定済みのテキストボックスの内容をブランクにリセット（クリア）します。
戻る ... ロード画面を終了します。

6.1.5 リストボックス

（下図 など） 変数や変数値を指定する選択ボタン（...） を押すと出現するアイテム選択リストです。



選択できるアイテム数は1個のみの場合と複数個選択可能の場合があります。
複数個選択可能な場合は、**拡張選択**（Shiftキーや Cntl キーを押しながら複数アイテムを選択する操作）が可能です。また、リストの最後の[クリア]を選択してセット または 追加 を押すと、テキストボックスの

内容がクリアされます。

リストボックス でアイテム選択後、**選択ボタン**（ ... ） は **セットボタン**（ セット ） または **追加ボタン**（ 追加 ）のいずれかに変わります。

6.1.6 セットボタン

（ セット ） テキストボックスの値が 選択されたアイテムに置き換わります。この表示に変わるテキストボックスは、基本的に、**手入力不可** です。

6.1.7 追加ボタン

（ 追加 ） テキストボックスの値の末尾に選択されたアイテムが追加されます。(アイテム間の区切り文字としてブランクが入ります) この表示に変わるテキストボックスは **手入力可能** です。

6. 1. 8 リストボックスの上にソートボタン

（ ABC順に並べ替え ） が配置されている場合があります。ソートボタン を押すと、押すたびに、リストボックス のアイテムがアルファベット順、またはその逆順に、並べ替えられて表示されますので、選択したいアイテムを見つけやすくなります。

6.1.9 表示ボタン

（ 表示 ） 指定されたデータが存在する場合に出現し、押すとデータ内容が表示されます。表示画面は (D) 表示画面の制御領域のコマンドによって設定されます。

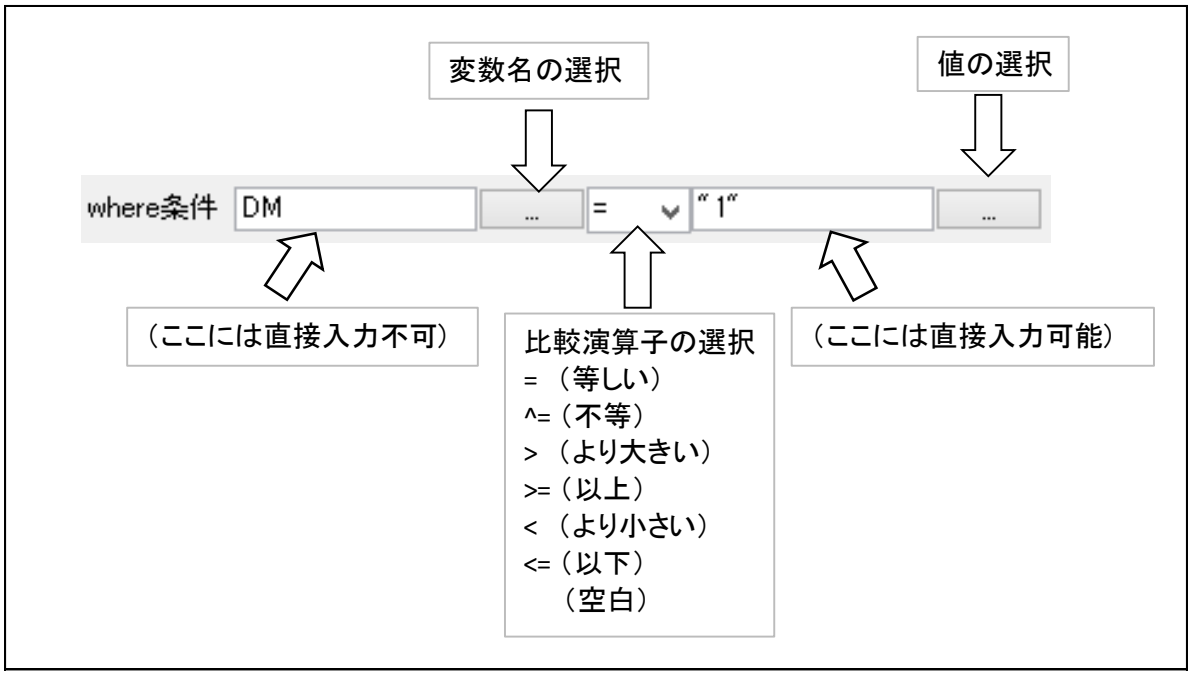
6.1.10 ラジオボタンとチェックボックス

Y または N のような排他的選択を行うパラメータの選択の場合に配置されています。排他的でない ON/OFF パラメータセットの場合は、**チェックボックス**（ ☑ ） が配置されている場合があります。

6.1.11 where 条件式の指定

いくつかの分析画面においては、入力データセットに対してwhere条件式によるオブザベーション抽出指定が可能です。

where条件には 変数名 と 演算子 と 値 の3つを指定します。



where条件の変数名部分の指定は、選択ボタンを押して出現する入力データセットに含まれる変数リストから 1 つの変数を選択します。選択された変数名がテキストボックスに表示されます。

演算子の部分は、コンボボックスから比較演算子を１つ選択します。ただし、最後の空白を選択すると、比較演算子を選択されません。このときは、右側の入力可能なテキストボックスに独自の条件式（例えば、 in 演算子や contains 演算子を使った抽出条件式）を入力指定できます。

値の部分は、選択ボタンを押して値を1つ選択することできますが、入力データセットのオブザベーションが多いと値の検索に時間がかかる場合があります。テキストボックスに直接値を入力することもできます。（直接入力する場合は、文字値の場合は値を引用符で囲んで指定します）

6.2（B）コードとログ表示領域

［**生成コード**］にはパラメータを指定していくに従って生成されるマクロ呼び出しコードが表示されます。入力パラメータが正しくコードに反映されているかどうか確認できます。また、SASディスプレイマネージャまたはWPSワークベンチで実行するために、生成されたコードをコピーしておくこともできます。

［**ログ**］には実行後のバッチジョブログやパラメータ入力エラーや警告その他のメッセージ、およびWPS実行中のログが表示されます。※ SAS実行ログは実行中に出現する 「SAS Message Log」画面に表示されます。

6.3（C）コマンド領域

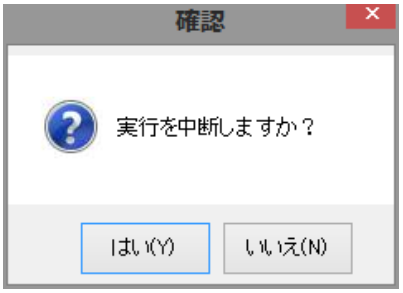
実行、戻る、前回表示、結果表示（モデル作成画面・予測値付与・コード保存のみ）、入力指定のリセット などのボタンが配置されています。

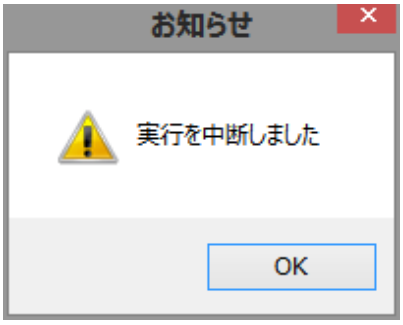
6.3.1 実行

. . . 指定のパラメータ条件で実行を開始します。実行すると、実行結果の成否（エラーの有無）や中断指定の有無にかかわらず、最新のパラメータセットの値は実行時の設定値に上書きされます。

6.3.2 実行の中断

実行ボタンを押してしばらくすると（1秒後に設定）、ボタンの表示が **実行** から **中断** に変化します。**中断** と表示されているときにボタンを押すと 中断するかどうかを尋ねる **ダイアログボックス** が出現します。中断したい場合は、 **はい**（Y）を押してください。





ただし、**はい**（Y） を押してから中断処理を行っている間にSASまたはWPSの実行が完了した場合は、実行終了のメッセージが表示されます。

6.3.3 前回表示

. . . 図表を表示する画面の場合は、最後に表示したhtmlファイルは、分析フォルダの下のhtmlディレクトリに保存されており、再実行することなく **前回表示** ボタンを押すことで 再表示することができます。

6.3.4 戻る

. . . ターゲット変数名、ターゲット変数値、説明変数リストなどを選択するリストボックス表示中にこのボタンを押すと、リストボックスを閉じて選択中の状態を解除します。そうでない場合は、その**「分析」**画面を終了し**「メニュー」**画面に戻ります。なお、 ボタンは、常に、アクティブな画面を終了させます。

6.3.5 入力指定のリセット

. . . その分析画面の指定パラメータを一旦すべて初

期値にリセットします。ただし、リセットした段階ではまだパラメータセットは保存されていません。（実行ボタンを押さない限りパラメータセットは変更されません。）分析画面を終了して起動画面に戻ってから再度同じ分析画面に切り替えるとリセット前のパラメータが復元されます。
リセットボタンは実行やパラメータ入力に何か問題が起きた際に押して、パラメータを再入力してください。

6.4 （D） 表示画面（ブラウザ）の制御領域

（A）パラメータ指定領域に配置された**表示**ボタン、または（B）コマンド領域に配置された**前回表示**ボタンを押したときに表示するデータ件数やラベル表示の有無、そして出現する表示画面（ブラウザ）のモードを制御します。

入力データや出力データの表示オブザベーション件数は、下部にある、表示するデータ件数の上限で制御します。**コンボボックス** 下矢印のボタン（▾）を押すと選択候補アイテムが表示され、その中から表示するデータ件数値を選択します。

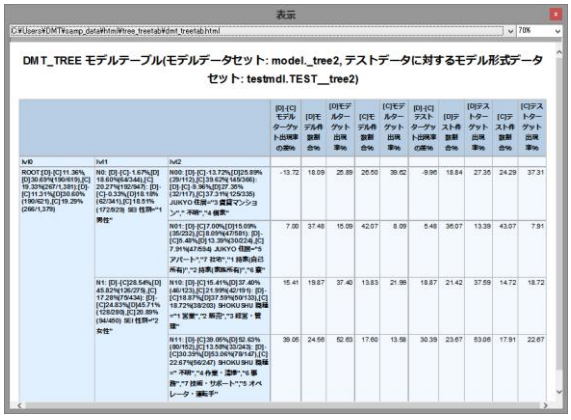
☑ 変数ラベルの表示　☑ 値ラベルの表示 のチェックボックスにチェックの有無により、変数ラベル、値ラベル（個々の文字変数値に定義されたフォーマットのこと）が定義されていた場合にそれらを用いるか否かを選択できます。

なお、表示するデータ件数の上限 指定と ☑ 変数ラベルの表示　☑ 値ラベルの表示 指定は、分析結果を表すクロス分析結果データ、モデル作成結果データ、ゲインチャートなどの座標値データおよび予測値付与スコアコードファイルの表示には適用されません。

☐ 別々の画面に表示 チェックボックスにチェックを入れると、その分析画面において、表示 や 前回表示 を押すたびに新しい画面がオープンし、複数の結果を同時表示できるようになります。ただし、その分析画面を閉じるとすべての表示画面は自動的にクローズされます。

7. 表示画面（ブラウザ）の操作

各分析画面の表示ボタンまたは前回表示ボタンを押すと、リクエストに応じて、入力データの内容、分析結果の図表、分析結果ファイルの内容などを表示する画面（ブラウザ）が出現します。



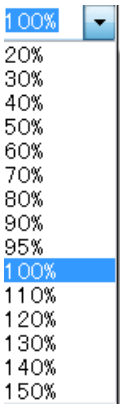
表示画面（ブラウザ）は、以下のように操作できます。

7.1 画面の拡大・縮小およびスクロール

画面右下角にマウスカーソルを置くと、カーソルの形状が ⤡ に変わります。このとき、マウスをドラグすることにより、表示画面の大きさをテンポラリに変更できます。画面最上部のウインドウタイトル（表示と書かれた部分）をダブルクリックすることにより、「全画面化」／「元の大きさに戻す」の切り替えができます。また、画面右側に配置されているスクロールバーを動かすことにより、表示をスクロールできます。

7.2 表示の拡大・縮小

画面右上に配置されているコンボボックスの ▾ を押して、表示の拡大率を変更できます。

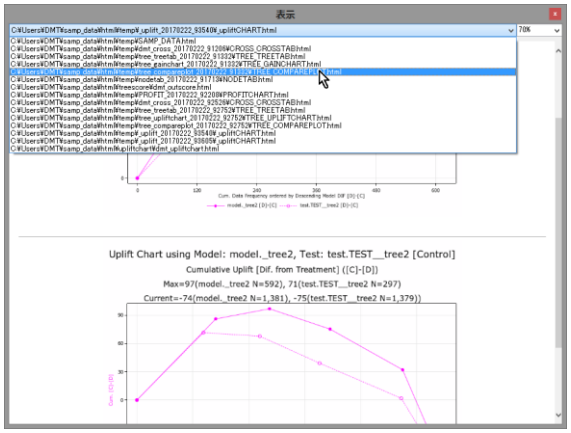


拡大率を小さくすると、画面に表示できる情報を増やすことができます。
なお、任意の拡大率を直接入力することもできます。（例: 121%と入力しエンターキーを押します）直前に変更された拡大率は保持されます。

7.3 過去の表示項目の再表示

タイトルバーの下に配置されているコンボボックス

は現在表示されているファイルのhtml出力ファイルのフルパスを表しています。このコンボボックスバーを押すと、本アプリケーションを起動してから表示したすべての表示履歴がリストされますので、任意の表示履歴を選択することにより再表示可能です。

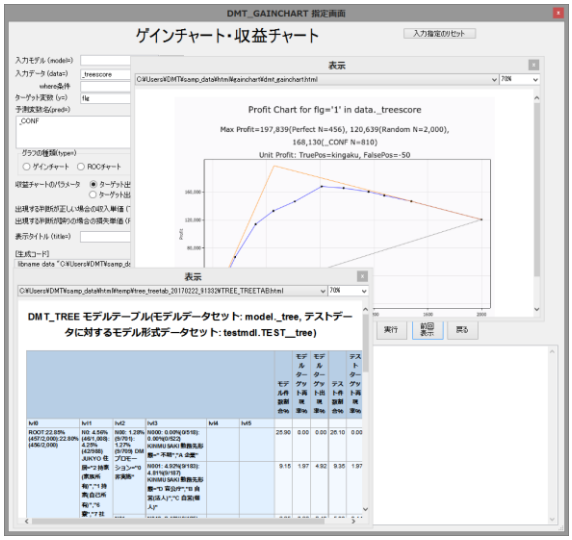


表示リクエストは、すべて分析ディレクトリ¥html¥tempフォルダの中に一時的にコピーされ、起動画面を閉じるまで保持され、 アプリケーションを終了すると、分析ディレクトリ¥html¥tempフォルダは初期化されます。

なお、 別々の画面に表示 にチェックを入れてから表示させた個々の画面については、表示時点以前の履歴のみを保持しています。

7.4 表示画面の複数表示

別々の画面に表示 にチェックを入れた状態で表示または再表示をリクエストすると、表示画面が出現された後、表示画面を閉じなくてもその分析画面の別の指定を行うことができるモードになります。
再び表示を行い、その画面に別の分析結果を履歴から選択表示させることにより、複数の分析結果を同時に閲覧できるので便利です。



各画面は個別に閉じることができますが、表示画面を呼び出した分析画面を閉じるとすべての表示画面も閉じます。

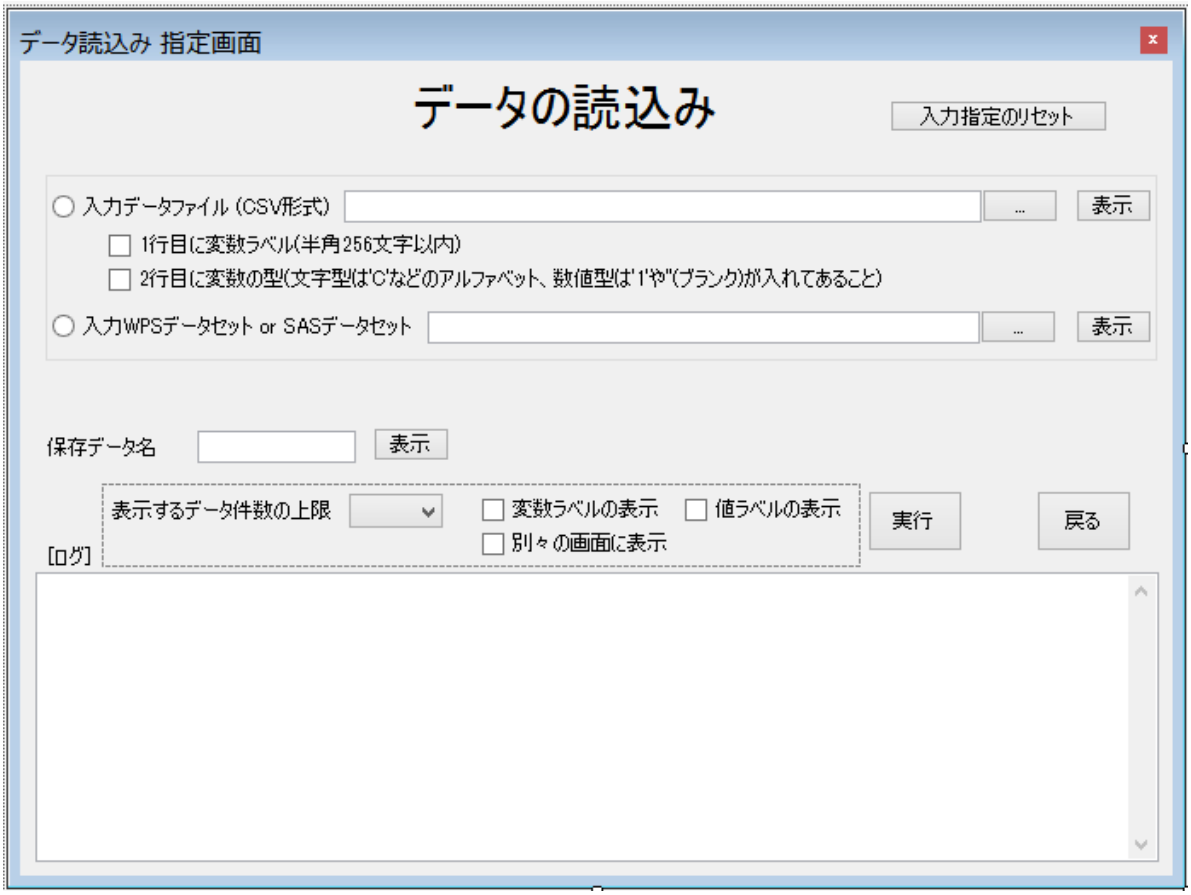
7.5 表示画面のクローズ

表示画面を閉じるには画面右上の ボタンを押します。 別々の画面に表示 のチェックが外れた状態でオープンした表示画面は閉じないと先に進めません。
別々の画面に表示 にチェックを入れた状態でオープンした表示画面は放置したままで分析画面の別の指示を行うことができますが、分析画面を閉じると自動的にすべてクローズされます。

8. 分析画面 ①データ抽出

データ入力・加工・変数ラベルや値ラベルの定義を行います。

8.1 データ読み込み



8.1.1 概要

本アプリケーションで分析を行うデータセットを入力指定します。CSV形式（カンマで区切られた可変長テキスト形式）ファイル、またはSASではV7以降のSASデータセット（sas7bdatファイル形式）、WPSではWPSデータセット（WPDファイル形式）を読み取ることができます。CSVファイル入力の場合は、1行目が項目名を表すかどうかを自動判定します。

8.1.2 指定方法

この機能はマクロモジュールには含まれていません。GUI実行モードでのみ指定可能です。

（必須指定）

以下の指定は必須です。ただし、(1),(2)はいずれか1つを選択します。

(1) 入力データファイル（CSV形式）
　　.... CSV形式ファイルを入力する場合に選択します。テキストボックスの右の … ボタンを押してファイルを選択します。テキストボックスにファイルのパス名が表示されると、表示 ボタンが出現し、表示ボタンを押すことにより、ファイルの中身を確認できます。

(2) 入力WPSデータセットorSASデータセット
　　… WPSデータセットまたはSASデータセット形式ファイルを入力する場合に選択します。テキストボックスの右の … ボタンを押してWPSデータセットまたはSASデータセットを選択します。テキストボックスにファイルのパス名が表示されると、表示 ボタンが出現し、表示ボタンを押すことにより、ファイルの中身を確認できます。

保存データ名
　　… 本システムに保存するデータセット名を指

定します。入力データファイル、または入力WPSデータセットまたはSASデータセットの拡張子を除いたファイル名が有効なSAS名であれば、ファイル選択時に自動入力されますが、任意の有効なSAS名に変更できます。

（任意指定）
以下の指定は任意ですが、(1) の入力に伴い、正しくデータ読み取りを行うためにチェックが必要な場合があります。

1行目に変数ラベル（半角256文字以内）
…入力CSVファイルの最初の行に変数名もしくは変数ラベルが入っているかどうかを指定します。ただし、入力CSVファイルを選択した後に、1行目と2～20行目までのデータ項目を比較して、1行目が項目名を表すかどうかをシステムが自動判定します。 項目名を表すと判断した場合、自動的にこのチェックボックスにチェックが入ります。
もしも自動判定が正しくない場合は手動で制御してください。

8.1.3 イニシャルディレクトリ

CSVファイル、またはWPSデータセットまたはSASデータセットの選択画面のイニシャルディレクトリはユーザプロファイル¥分析ルートディレクトリ¥sample に設定しています。他のディレクトリのファイルを入力したい場合は、ダイアログの左側のドック部分から目的ファイルのディレクトリを辿って選択してください

8.1.4 変数名、変数ラベル、フォーマットについて

CSVファイル読み取りの場合の保存データセットの変数名は、1行目が変数名で無い場合、または無効な変数名の場合、VARk （kは項目の定義順序を表す、k=1,2,…）という名前の変数名がつきます。無効な変数名（例えばabc*123）の場合は、”VARk” + “ “ + “abc*123” という変数ラベルが設定されます。

（例1）

```
ID,店舗,商品ITEM,amount
0001,東京,A4ノート,100
0002,大阪,万年筆,50
```

上記CSVファイルを読むと、変数名は ID, VAR2, VAR3, amount となり、変数 VAR2, VAR3 には それぞれ、”VAR2 店舗”, “var3 商品ITEM” という変数ラベルが付きます。

Variables in Creation Order

Number	Variable	Type	Len	Pos	Format	Informat	Label
1	ID	Char	5	0	$5.	$5.	
2	VAR2	Char	4	5	$4.	$4.	VAR2 店舗
3	VAR3	Char	8	9	$8.	$8.	VAR3 商品ITEM
4	amount	Char	6	17	$6.	$6.	

Obs	ID	VAR2 店舗	VAR3 商品ITEM	amount
1	0001	東京	A4ノート	100
2	0002	大阪	万年筆	50

1行目がデータの場合は、すべての変数名が VARk となり、変数ラベルはつきません。

（例2）

```
0001,東京,A4ノート,100
0002,大阪,万年筆,50
```

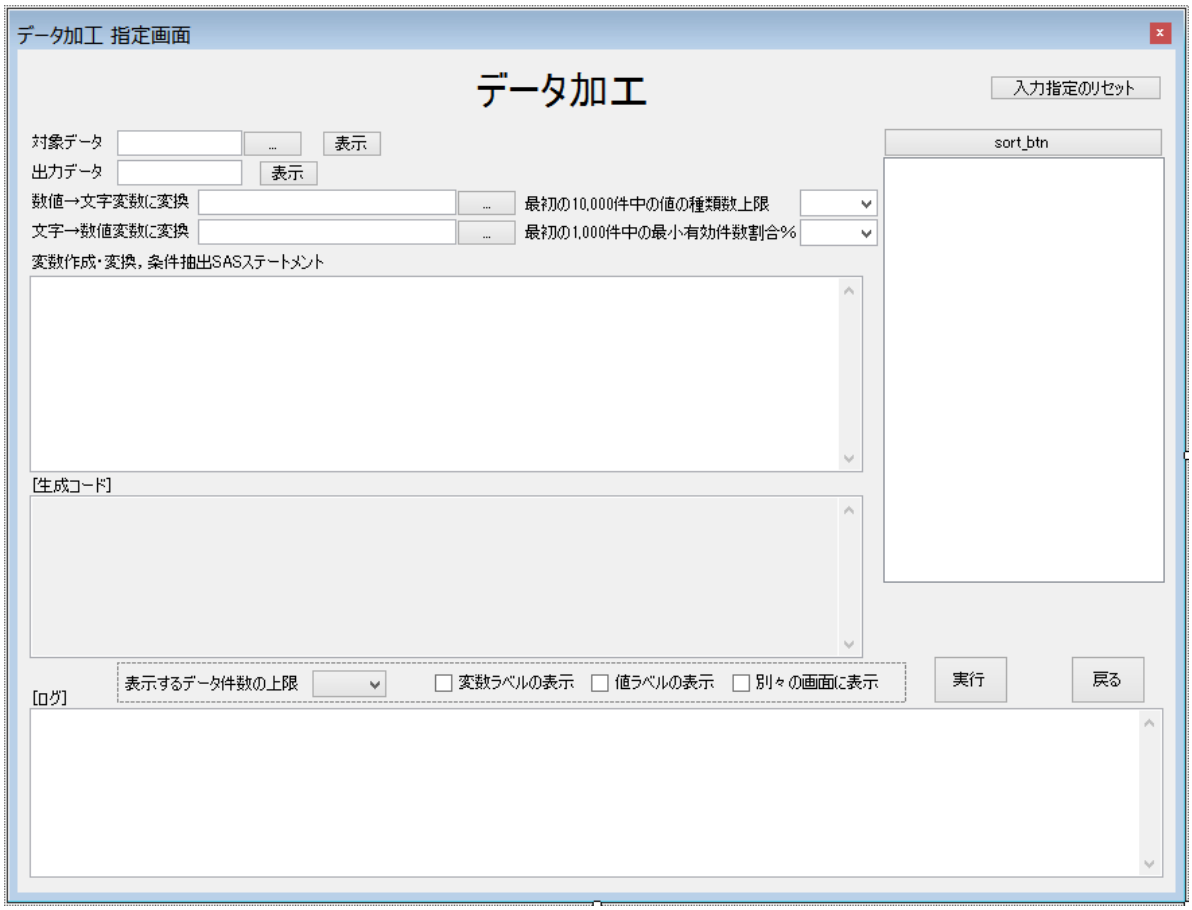
Variables in Creation Order

Number	Variable	Type	Len	Pos	Format	Informat
1	VAR1	Num	8	0	BEST12.	BEST32.
2	VAR2	Char	4	16	$4.	$4.
3	VAR3	Char	8	20	$8.	$8.
4	VAR4	Num	8	8	BEST12.	BEST32.

Obs	VAR1	VAR2	VAR3	VAR4
1	1	東京	A4ノート	100
2	2	大阪	万年筆	50

WPSデータセットまたはSASデータセット読み取りの場合は、読み取り後のデータセットの変数名、変数ラベルはそのまま入力データセットの変数名、変数ラベルがコピーされます。しかし、**変数フォーマットはすべて削除されます**ので、**ラベル付与** 画面で 改めて 値ラベル として定義してください。

8.2 データ加工



8.2.1 概要

入力したデータセットの変数タイプの変更（数値タイプから文字タイプへ、またその逆）、分析に用いる変数の選択、新変数の作成、条件抽出などのデータ加工が行えます。

8.2.2 指定方法

この機能はマクロモジュールには含まれていません。GUI実行モードでのみ指定可能です。

（必須指定）

対象データ
　.... 入力データセットを選択します。

出力データ
　… 出力するデータセット名を入力します。入力する対象データセット名と異なる名前を付ける必要があります。

（任意指定）

数値→文字変数に変換
　… 数値タイプから文字タイプに変換する変数名のリストを選択または指定します。

最初の10,000件中の値の種類数上限
　… 数値タイプから文字タイプに変換する変数の選択条件を設定します。指定の値種類数以下を持つ数値タイプ変数のみ選択されるように設定します。デフォルトは50。

文字→数値変数に変換
　… 文字タイプから数値タイプに変換する変数名のリストを選択または指定します。

最初の10,000件中の最小有効件数割合%
　… 文字タイプから数値タイプに変換する変数の選択条件を設定します。値を数値タイプに変換しても有効な値となる割合%が指定の割合%以上である文字変数のみリストボックスに表示されるように設定します。デフォルトは50%。

変数作成・変換，条件抽出SASステートメント
　… 変数作成・変換、条件抽出などの目的で、自由にSASコードを記述できます。

8.2.3 生成コードの構造

この画面指定で生成されるSASコードの構造は、以下のとおりです。

(1) ファイル割り当て、オプション設定などの定型前処理部分（libname,optionsステートメント）
(2) 数値→文字変換、文字→数値変換の指定が存在する場合は、変数ラベル保存処理部分（data _null_ で始まる DATAステップ）
(3) データ加工処理部分（data outdata.xxxxで始まるDATAステップ）

［生成コード］で生成されたSASコードを確認します。

(1) ファイル割り当て、オプション設定などの定型前処理部分

```
libname indata
"C:¥Users¥DMT¥samp_data¥data¥samp_data_csv";
libname outdata
"C:¥Users¥DMT¥samp_data¥data¥samp_data_csv2
";
options nofmterr;
libname library (outdata);
```

(2) 数値→文字変換、文字→数値変換指定がある場合の変数ラベル保存処理部分

```
data _null_;set indata.samp_data_csv(obs=1);
length _labelvar $256;
 if vname(sei) ne vlabel(sei) then
_labelvar=vlabel(sei);else _labelvar='';
 call symput("_label_0", _labelvar);
 if vname(shokushu) ne vlabel(shokushu) then
_labelvar=vlabel(shokushu);else _labelvar='';
 call symput("_label_1", _labelvar);
;run;
```

(3) データ加工処理部分

```
data
outdata.samp_data_csv2(rename=(_dummy_0=sei
_dummy_1=shokushu));
set indata.samp_data_csv;
 _dummy_0=left(put(sei,best12.));if _dummy_0='.'
then _dummy_0='';
 _dummy_1=left(put(shokushu,best12.));if
_dummy_1='.' then _dummy_1='';
 label _dummy_0="&_label_0";
 label _dummy_1="&_label_1";
drop sei shokushu;
if nenrei>=40 then nenrei_kbn="中高年";
else nenrei_kbn="若年";
run;
```

drop sei shokushu;とrun;の間に、**変数作成・変換, 条件抽出SASステートメント** に入力したテキストがそのまま挿入されます。

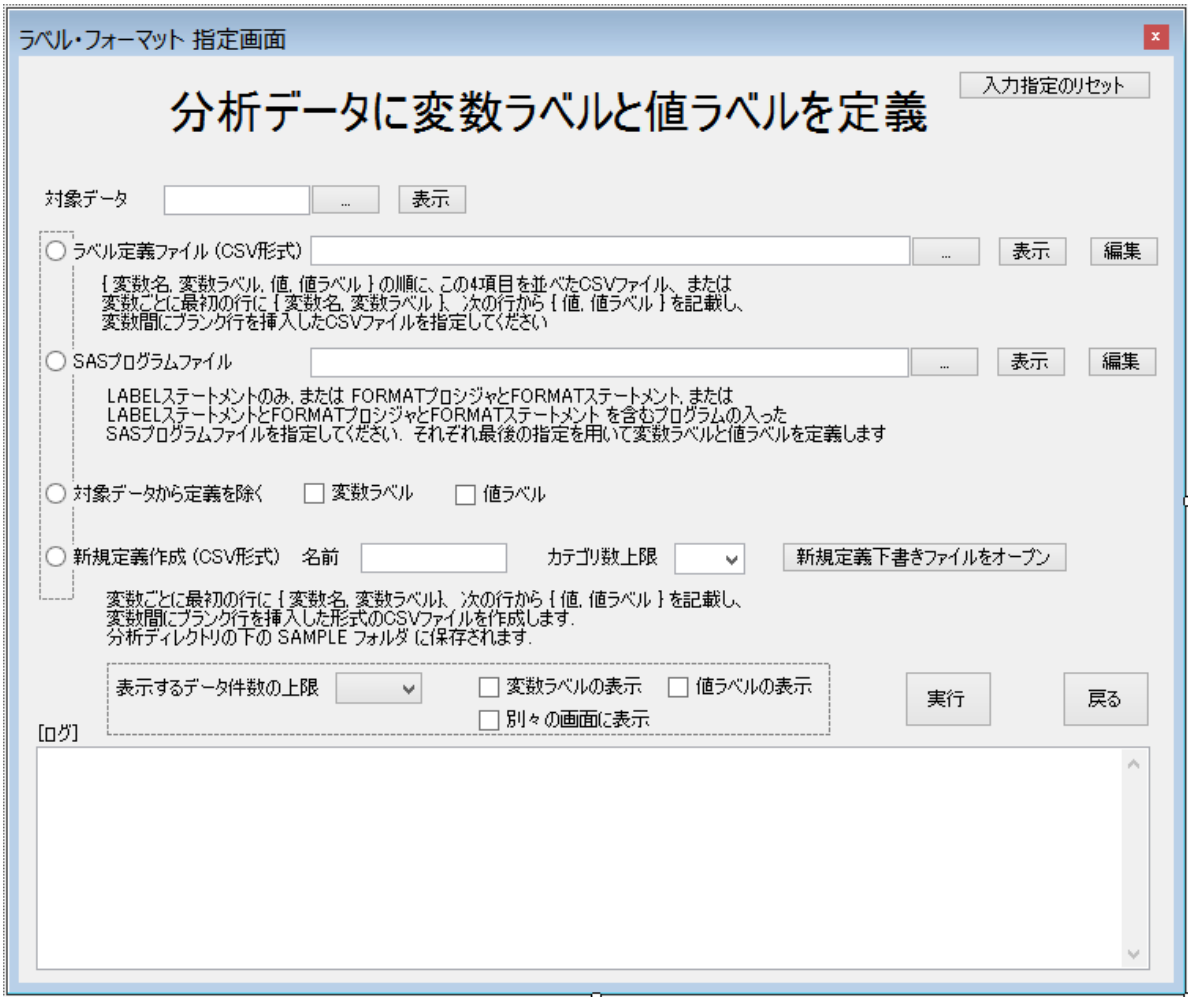
（TIPS）以下の例はCARDS文で入力したデータを読み取り、分析ルートディレクトリのDATAライブラリの中に保存します。

(1) 入力データには存在するデータを適当に指定する。ここでは samp_data
(2) **出力データ**に、**作成するデータセット名**を入れる。ここでは samp2
(2) **変数作成・変換・条件抽出SASステートメント テキストボックス**には以下のように入力する、
(2-1) 最初の行に、 stop;run; と入れる。
(2-2) 次に、data outdata.samp2; で始まるデータステップを入力する。このとき、ライブラリ名はoutdataで固定、**作成するデータセット名**は(1) の**出力データ テキストボックス**に入れた名前と同じにする。
ここでは、 data outdata.samp2; に続いて、cards文でデータを読み取るプログラムを書いています。

［生成コード］

```
libname indata
"C:¥Users¥DMT¥samp_data¥data¥samp_data ";
libname outdata
"C:¥Users¥DMT¥samp_data¥data¥samp2";
options nofmterr;
libname library (outdata);
data outdata.samp2;
set indata.samp_data;
stop;run;

data outdata.samp2;
   input id $ x1 x2;
cards;
01 1 2
02 2 4
03 3 10
;
run;
```

8.3 ラベル付与



8.3.1 概要

変数の意味や文字変数値の値の意味を分かりやすく表示するためのラベルを定義します。本システムにおいては、変数につけるラベルを **変数ラベル** 、文字変数の個々の値に1対1でつけるラベルを **値ラベル** と呼んでいます。値ラベルは、文字変数の個々の値ごとに定義されたフォーマット値のことです。

8.3.2 指定方法

この機能はマクロモジュールには含まれていません。GUI実行モードでのみ指定可能です。

（必須指定）
対象データ
　　.... **入力データセットを選択します。**

以下の(1),(2),(3),(4)のいずれかが必須指定です。

(1) **ラベル定義ファイル (CSV形式)**
　　.... **{ 変数名, 変数ラベル, 値, 値ラベル } の順に、この4項目を並べたCSVファイル、または変数ごとに最初の行に { 変数名, 変数ラベル }, 次の行から続けて { 値, 値ラベル } を記載し、変数間にブランク行を挿入したCSVファイルを入力する場合に選択します。**

以下の①、②のいずれかのパターンのCSVファイルをあらかじめ作成しておき、作成したCSVファイルをここで指定します。

① { 変数名, 変数ラベル, 値, 値ラベル } の順に、4項目を並べたCSVファイル

（例）

```
flg,応答,1, あり
flg,応答,0, なし
kinmusaki,勤務先形態,A,企業
kinmusaki,勤務先形態,B,自営(法人)
kinmusaki,勤務先形態,C,自営(個人)
kinmusaki,勤務先形態,D,官公庁
kinmusaki,勤務先形態,,不明
```

```
gakureki,最終学歴,1,中学
gakureki,最終学歴,2,高校
gakureki,最終学歴,3,専門学校
gakureki,最終学歴,4,大学
gakureki,最終学歴,5,大学院
gakureki,最終学歴,,不明
kazoku_kosei,家族構成,1,独身同居家族あり
kazoku_kosei,家族構成,2,独身単身
kazoku_kosei,家族構成,3,既婚子供あり
kazoku_kosei,家族構成,4,既婚子供なし
kazoku_kosei,家族構成,5,独身子供あり
kazoku_kosei,家族構成,,不明
sei,性別,1,男性
sei,性別,2,女性
nenrei,年齢,,
```

なお、同じ変数名、変数ラベルが続く行については、以下のように、重複する変数名、変数ラベルはヌル値で入力されていてもかまいません。

（重複した変数名、変数ラベルをヌル指定した例）

```
flg,応答,1,あり
,,0,なし
kinmusaki,勤務先形態,A,企業
,,B,自営(法人)
,,C,自営(個人)
,,D,官公庁
,,,不明
gakureki,最終学歴,1,中学
,,2,高校
,,3,専門学校
,,4,大学
,,5,大学院
,,,不明
kazoku_kosei,家族構成,1,独身同居家族あり
,,2,独身単身
,,3,既婚子供あり
,,4,既婚子供なし
,,5,独身子供あり
,,,不明
sei,性別,1,男性
,,2,女性
nenrei,年齢,,
```

注意：
・数値タイプ変数には変数ラベルのみ指定可能です。上記の変数 nenrei のように指定します。
・文字タイプ変数の欠損値に対する値ラベルは上記の「不明」という値ラベルを定義しているように指定します。
・文字タイプ変数には、変数ラベルのみ指定してもかまいませんし、一部の値に対して値ラベルを指定してもかまいません。
・厳密な指定の有効性チェックは行っておりません。対象データに存在しない変数が指定された場合は、以下のようになります。

変数ラベル … 存在する変数については有効、存在しない変数については無視されます。
値ラベル … 存在する変数については有効、存在しない変数については無視されます。

・**編集** ボタンを押すと、編集可能なメモ帳が開きます。一部変更・削除・追加を行う場合に便利です。
・結果がおかしい場合は、他の方法を試してください。

②変数ごとに最初の行に { 変数名, 変数ラベル }, 次の行から続けて { 値, 値ラベル } を記載し、変数間にブランク行を挿入したCSVファイル

（例）

```
flg, 応答
1, あり
0, なし

kinmusaki, 勤務先形態
A, 企業
B, 自営(法人)
C, 自営(個人)
D, 官公庁
, 不明

gakureki, 最終学歴
1, 中学
2, 高校
3, 専門学校
4, 大学
5, 大学院
,不明

kazoku_kosei, 家族構成
1, 独身同居家族あり
2, 独身単身
3, 既婚子供あり
4, 既婚子供なし
5, 独身子供あり
, 不明

sei, 性別
1, 男性
2, 女性

nenrei, 年齢
```

注意：
・数値タイプ変数には変数ラベルのみ指定可能です。上記の変数 nenrei のように指定します。
・文字タイプ変数の欠損値に対する値ラベルは上記の「不明」という値ラベルを定義しているように指定します。

・文字タイプ変数には、変数ラベルのみ指定してもかまいませんし、一部の値に対して値ラベルを指定してもかまいません。
・**空白行を1つの変数の指定の区切りとして認識しますので、必ず空白行を変数定義ごとに挿入してください。**
逆に、**1つの変数の指定の途中で空白行を挿入しない**でください。
・厳密な指定の有効性チェックは行っておりません。対象データに存在しない変数が指定された場合は、以下のようになります。

変数ラベル … 存在する変数については有効、存在しない変数については無視されます。
値ラベル … 存在する変数については有効、存在しない変数については無視されます。

・**編集** ボタンを押すと、編集可能なメモ帳が開きます。一部変更・削除・追加を行う場合に便利です。
・結果がおかしい場合は、他の方法を試してください。

(2) SASプログラムファイル
… LABELステートメントのみ,またはFORMATプロシジャとFORMATステートメント，または、LABELステートメントとFORMATプロシジャとFORMATステートメント を含むプログラムの入った　SASプログラムファイルを用いる場合に選択します。それぞれ最後の指定を用いて変数ラベルと値ラベルを定義します。

SASプログラムコードを使って変数ラベル、値ラベルを定義します。

変数ラベル … LABELステートメントから定義されます。複数の LABELステートメントが存在する場合は、最後の LABELステートメントのみ用いられます。LABELステートメントの中に、**対象データに存在しない変数の変数ラベル定義は指定しないでください**。(エラーメッセージがログに出現します)

値ラベル … PROC FORMATステートメントからRUNステートメントまでの範囲とFORMATステートメントから定義されます。文字タイプフォーマットのみ利用されます。複数存在する場合は、最後のPROC FORMATステートメントからRUNステートメントまでの範囲、最後のFORMATステートメントのみが、それぞれ用いられます。FORMATステートメントに定義された変数の中に、**対象データに存在しない変数のフォーマット定義は指定しないでください**。(エラーメッセージがログに出現します)

（例）

```
label flg='応答' kinmusaki='勤務先形態' gakureki='最終学歴'　sei='性別' nenreix='年齢';

proc format library=library;
value $flgj '1'='あり' '0'='なし';
value $kinmuj 'A'='企業' 'B'='自営(法人)' 'C'='自営(個人)' 'D'='官公庁' other='不明';
value $gakuj '1'='中学' '2'='高校' '3'='専門学校' '4'='大学' '5'='大学院' other='不明';
value $kazokuj '1'='独身同居家族あり' '2'='独身単身' '3'='既婚子供あり' '4'='既婚子供なし' '5'='独身子供あり' other='不明';
value $seij '1'='男性' '2'='女性' other='不明';
run;

format flg $flgj. kinmusaki $kinmuj. gakureki $gakuj. kazoku_kosei $kazokuj. sei $seij.;
```

注意：
・数値タイプ変数には変数ラベルのみ指定可能です。上記の変数 nenrei のように指定します。
・FORMATプロシジャは VALUEステートメントで元の変数値に対する1対1のフォーマット値のみを指定してください。(other=指定は使用しないでください)
・文字タイプ変数には、変数ラベルのみ指定してもかまいませんし、一部の値に対して値ラベルを指定してもかまいません。
・厳密な指定の有効性チェックは行っておりません。対象データに存在しない変数を指定するとエラーとなります。
・**編集** ボタンを押すと、編集可能なメモ帳が開きます。一部変更・削除・追加を行う場合に便利です。
・結果がおかしい場合は、他の方法を試してください。

(3) **対象データから定義を除く**
… 変数ラベル、値ラベルそれぞれの定義を削除したい場合に選択します。

これを選択し、実行すると、チェックボックスにチェックが入った変数ラベルと文字変数のフォーマット定義はデータセットから削除されます。

なお、データに変数ラベルや値ラベルが定義されたままでも、多くの場合で、表示する時点で、ラベル表示を行う／行わないを選択することができますので、一度データにラベル定義や値ラベルを定義したら、訂正を行いたい場合以外は削除する必要はあまりありません。

(4) **新規定義作成** (CSV形式)
… 新たに変数ラベル、値ラベルを定義したい場合に選択します。入力対象データに定義されている変数ラベル、値ラベルを初期値として、変数ラベル、値ラベルを編集する画面が開きます。

対象データの任意の項目に変数ラベル、値ラベルを定義したい場合に選択します。

名前 を入力し、**新規定義下書きファイルをオープン**を押すと、変数については、対象データの全変数の

{ **変数名**, **変数ラベル** } が表示されます。**カテゴリ上限** 以下の値の種類数を持つ文字タイプ変数については、続いて { **値**, **値ラベル** } が存在する値の数だけ表示され、変数の区切りとしてブランク行が１行追加されます。この内容を編集したCSVファイルは **名前** に保存されます。
こうして作成したCSVファイルは **(1) ラベル定義ファイル (CSV形式) ラジオボタン** の入力として用いることができ、編集した内容で対象データに変数ラベル、値ラベルを定義できます。

（オープンしたときの例）

```
flg, flg
0, 0
1, 1

kinmusaki, kinmusaki
A, A
B, B
C, C
D, D
 ,

gakureki, gakureki
1, 1
2, 2
3, 3
4, 4
5, 5
 ,

kazoku_kosei, kazoku_kosei
1, 1
2, 2
3, 3
4, 4
5, 5
 ,

sei, 性別
1, 男性
2, 女性

nenrei, 年齢
```

（編集後の例）

```
flg, 応答有無
0, なし
1, あり

kinmusaki, 勤務先
A, 企業
B, 自営(法人)
C, 自営(個人)
D, 官公庁
 , 不明

gakureki, 学歴
1, 中学
2, 高校
3, 専門学校
4, 大学
5, 大学院
 , 不明

kazoku_kosei, 家族構成
1, 独身同居家族あり
2, 独身単身
3, 既婚子供あり
4, 既婚子供なし
5, 独身子供あり
 , 不明

sei, 性別
1, 男性
2, 女性

nenrei, 年齢
```

編集画面を閉じると、ファイルの内容は更新されます。

注意
・慣れないうちは、新規の名前を指定するようにしてください。既に存在するCSVファイル を開くと、内容が対象データに定義された変数ラベル、文字変数のフォーマット値を参照した ラベル定義CSVファイルに初期化 されます。復元できませんので、既存ファイルのオープンには注意してください。
・既存CSVラベル定義ファイルを開く場合も、改めてその時点でデータセットに定義済みの全変数の { **変数名**, **変数ラベル** }、**カテゴリ上限** 以下の値の種類数を持つ文字タイプ変数の { **値**, **値ラベル** } が表示されます。(編集後、この定義をデータにつけないまま、再編集しようとすると、編集した作業結果がすべて元に戻り、無駄になります。) 一旦作成したラベル定義CSVファイルの再編集は、必ず、**(1) ラベル定義ファイル**(CSV形式)を選択し、編集したラベル定義CSVファイルを選択し実行 し、データに変数ラベル、値ラベルを付けてから行ってください。 また、改めて再編集を行うより、微調整で済む場合は、**(1) ラベル定義ファイル**(CSV形式)を選択し、**編集** ボタンを押して行ってください。
・対象データの全変数が表示されます。変数の数が多い場合は編集画面が表示されるまで時間がかかる場合があります。**データ加工** 画面で分析に用いないことが明らかな変数はあらかじめ削除しておくと効率的です。
・既に定義されている変数ラベル、値ラベルがあれば、それが表示されます。既存定義の無い場合の変数ラベルは変数名、値ラベルは値がデフォルト表示されますので、ラベルを必要に応じて編集してください。
・値ラベルは **カテゴリ上限** 以下の値の種類数を持つ

文字タイプ変数のみ表示されます。

(4)を選択した場合、以下は必須です。

名前

… 新規定義作成で編集保存する変数ラベル･値ラベル定義CSVファイルに名前を付けます。なお、このファイルは分析ルートディレクトリの下のSAMPLEディレクトリに保存されます。既存の名前を指定すると、注意メッセージが出現しますが、上書き保存は可能です。

（任意指定）

カテゴリ数上限 コンボボックス

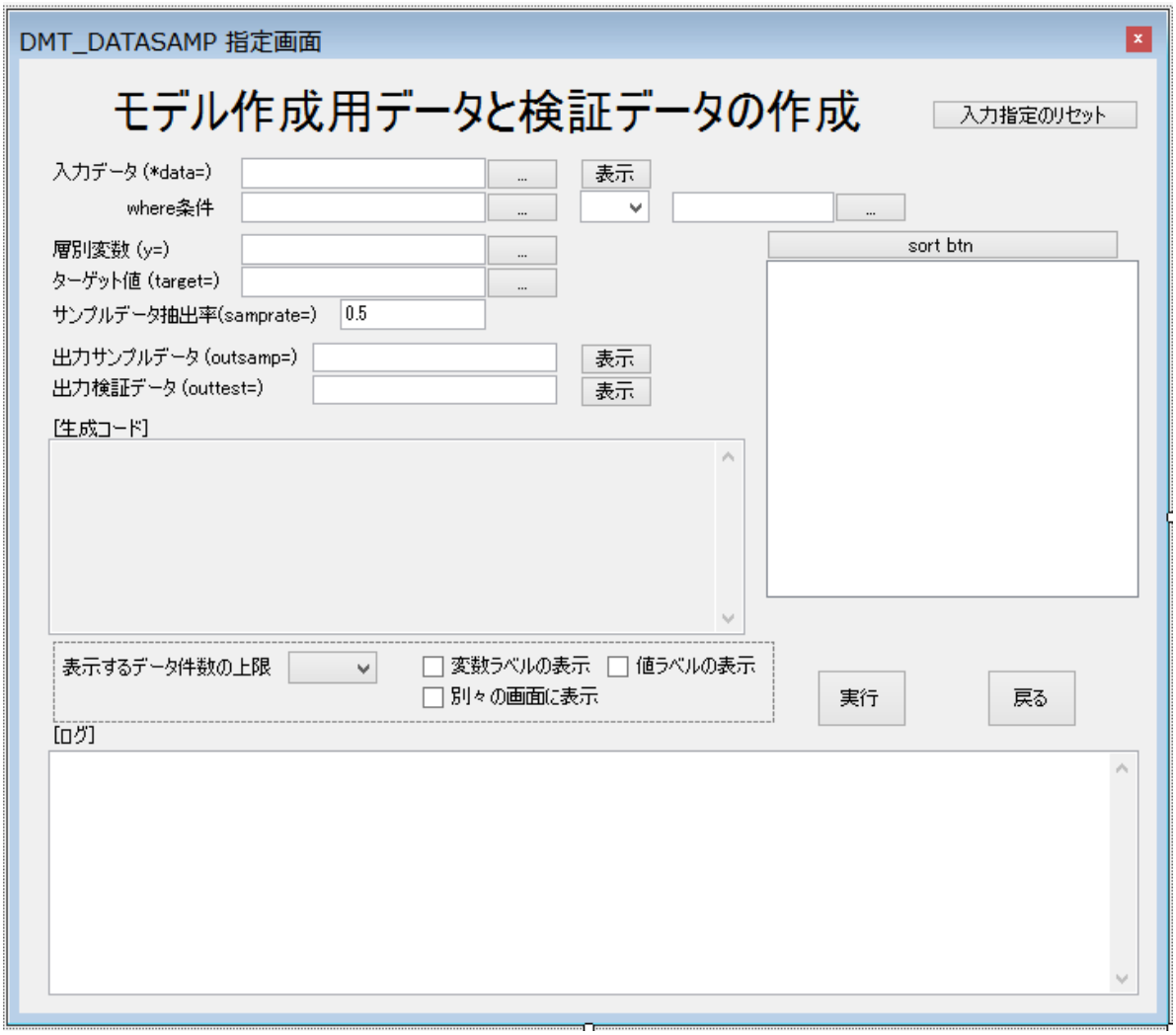
… 文字タイプ変数の ｛ 変数名，変数ラベル ｝ の行に続いて ｛ 値，値ラベル ｝ の行を表示する値の種類数（カテゴリ数）の上限を指定します。たとえば、10 とすると、値の種類数が10を超える文字タイプ変数は ｛ 変数名， 変数ラベル ｝ のみ表示され、｛ 値，値ラベル ｝ は表示されません。

なお、値の種類数には欠損（ブランク）を含みます。

（TIPS）カテゴリ数上限 について：

50近いカテゴリ数を持つ都道府県コードなど、値ラベルを付与した方が良いと思われるカテゴリカル変数の最大カテゴリ数を考慮して設定します。**カテゴリ数上限** の設定は、比較的大きめの値を設定しておくことをお勧めします。

8.4 検証確保（dmt_datasamp）



8.4.1 概要

DMT_DATASAMPマクロを呼び出し、分析データセットからオブザベーションをランダム抽出します。

主な用途は次の2つです。

（1）　モデル作成用データセットとモデル検証用データセットの作成
（2）　データセットから任意の抽出率の単純サンプリングまたは層別サンプリング

（1）　モデル作成用データセットとモデル検証用データセットの作成

モデル作成データへのモデルの過剰適合がお気ているかどうかを確認する目的で、分析に用いることができるデータセットから、分類木の場合はターゲット別の層別サンプリングの方法、回帰木の場合は単純サンプリングの方法で、オブザベーションをランダム抽出し、モデル作成用データでモデルを作成し、その精度をモデル検証用データで確認するモデル作成方法が一般的に採用されています。

DMT_DATASAMPでは層別変数とターゲット値をそれぞれターゲット変数（y=パラメータ）およびターゲット（target=パラメータ）で指定することにより、ターゲット／非ターゲット別に抽出率（samprate=パラメータ）に応じた件数割合でそれぞれランダム抽出を行い、同じ抽出率の方のターゲット／非ターゲットをそれぞれ集めて分類木モデル作成用データとモデル検証用データを作成する機能を持っています。

（2）　データセットから任意の抽出率の単純サンプリングまたは層別サンプリング

ターゲット変数（y=パラメータ）の指定を行わない場合は、抽出率（samprate=パラメータ）に応じた単純サンプリングを行います。サンプリング結果は抽出率で指定した割合の方をサンプル（outsamp=パラメータで名前をつけたデータセット）、残りをテスト（outtest=パラメータで名前をつけたデータセット）

に出力します。

ターゲット変数（y=パラメータ）の指定を行った場合は、ターゲット変数の値別に抽出率（samprate=パラメータ）に応じた層別サンプリングを行います。サンプリング結果は抽出率で指定した割合の方をサンプル（outsamp=パラメータで名前をつけたデータセット)、残りをテスト（outtest=パラメータで名前をつけたデータセット）に出力します。なお、許容するターゲット変数の値の種類はデフォルトで最大100までとしていますが、maxgrp=パラメータで変更可能です。

ターゲット変数（y=パラメータ）およびターゲット（target=パラメータ）の指定を行った場合は、ターゲット値とそれ以外のすべての値を非ターゲットとした2つのカテゴリ別に抽出率（samprate=パラメータ）に応じた層別サンプリングを行います。サンプリング結果は抽出率で指定した割合の方をサンプル（outsamp=パラメータで名前をつけたデータセット)、残りをテスト（outtest=パラメータで名前をつけたデータセット）に出力します。

8.4.2 指定方法

（コマンド実行モードでの指定）

%dmt_datasamp(help,data=,outsamp=_sampdata
,outtest=_testdata,samprate=0.66667,testrate=,y=,target=,maxgrp=STURGES
,seed=1,language=JAPANESE)

（GUI実行モードでの変更点）

・help, testrate=パラメータは使用不可。
・samprate=パラメータのデフォルトは **0.5**
・outsamp=パラメータのデフォルトは SAMP_&data
・outtest=パラメータのデフォルトは TEST_&data
（&data はdata=パラメータの値です。）
・maxgrp=, seed=はオプション画面で指定します。

（必須パラメータ）

以下の1個のパラメータは省略できません。

入力データ(data=)
　... 入力データセット名の指定.
　入力データセット名の後にwhere=データセットオプションを指定できます。

（単純サンプリングの場合のパラメータ）

以下の4個のパラメータは任意指定です。(=の右辺の値はデフォルト値を表しています)

サンプル抽出率の指定 (samprate=0.66667)
テストデータ抽出率の指定 (testrate=)
　... （コマンド実行モードのみ有効）
出力サンプルデータ名の指定
　(outsamp=_sampdata)
　... モデル作成用データセット名の指定
出力検証データの指定 (outtest=_testdata)
　... 出力する残りのモデル検証用データセット名を指定.
乱数シード値の指定 (seed=1)
　... (正の整数を与える)

注意： samprate=, testrate=パラメータはいずれか1つのみを指定します。

（層別サンプリングの場合のパラメータ）

以下のパラメータは層別サンプリングを行う場合の必須指定です。次の target=パラメータを同時指定しない場合は、ターゲット変数のすべての値別に層別サンプリングを行う指定となります。

層別変数の指定 (y=)
　... ターゲット変数名を指定. (単一変数名のみ指定可)

以下のパラメータは任意指定です。target=パラメータの指定があればターゲットとターゲット以外のすべての値の2つのグループ別の層別サンプリング指定となります。

ターゲット値の指定 (target=)
　... （単一値のみ指定可、ただし数値タイプの場合のみ、あるしきい値以上または以下または超または未満を指定可）

許容最大層別数の指定 (maxgrp=100)
　... 層別数がデフォルトより大きい場合は、値を増やしてください。

その他、単純サンプリングの場合に記載した4個のパラメータが任意指定です。

（その他のパラメータ）

help ... パラメータ指定方法をログ画面に表示します。
　このオプションは単独で用います。（このパラメータはコマンド実行モードでのみ有効）
　例：%dmt_datasamp(help)

言語(language=JAPANESE)
　... 言語の選択 他に ENGLISH が指定可能

8.4.3 パラメータの詳細

入力データ (data=)
例：data=a, data=a(where=(DM="1"))

層別変数 (y=)
層別サンプリングを行う場合は必須指定です。

target=パラメータを一緒に指定しない場合は、y=パラメータに指定した変数の値別にsamprate=パラメータで指定した共通の抽出率で各層のランダムサンプリングを行い、outsamp=データセットにまとめて出力されます。（残りはouttest=データセットに出力されます。）y=パラメータに指定する変数は 2以上maxgrp=パラメータの値以下の値の種類数を持つ必要があります。
例：y=flag

(TIPS) 層別変数は1個のみ指定可能です。性別かつ年齢階層別といった2重層別を行いたい場合は、2つの変数の値の組合せを値に持つ変数（クロス変数）を作成し、それをy=パラメータに指定します。

ターゲット値 (target=)
y=パラメータと同時にターゲット値を指定します。指定します。ターゲット値と非ターゲット値の2つのグループ別の層別サンプリングを行います。

ターゲット変数が文字タイプの場合は1種類の値を指定します。特殊な文字（+,-など）を含まない限り引用符で囲む必要はありません。ターゲット変数が数値タイプの場合は1種類の値、もしくはあるしきい値を境とした「以上」、「以下」、「超」、「未満」のいずれかの範囲を指定可能です。数値変数タイプで範囲を指定する場合は引用符で囲んではいけません。

例1：y=flag,target=A　（ターゲット変数が文字タイプ変数で、その値"A"をターゲットに指定する場合）
例2：y=sales,target=1000　（ターゲット変数が数値タイプで、その値1000をターゲットに指定する場合）
例3：y=sales,target=>1000　（ターゲット変数が数値タイプで、その値1000超をターゲットに指定する場合）
例4：y=sales,target=>=1000　（ターゲット変数が数値タイプで、その値1000以上をターゲットに指定する場合。　target==>1000と指定してもかまいません。）
例5：y=sales,target=<1000　（ターゲット変数が数値タイプで、その値1000未満をターゲットに指定する場合）
例6：y=sales,target=<=1000　（ターゲット変数が数値タイプで、その値1000以下をターゲットに指定する場合。　target==<1000と指定してもかまいません。）

注：文字タイプ変数のターゲット値は、大文字、小文字が区別される点に注意してください。（変数名は大文字・小文字の区別はありません。）

乱数シード値 (seed=1)
正の整数値を指定すると、同じシード値に対して常に同じコンピュータ乱数系列が生成されます。一方、値0を指定すると、生成されるコンピュータ乱数系列は実行するたびに異なるものとなります。分析結果の再現性を求める場合は、シード値は0以外に指定してください。

許容最大層別数 (maxgrp=100)
非常にたくさんのカテゴリを持つ層別変数を誤って指定した場合に実行を行わないようにするためのオプションです。指定の値を超える場合はエラーとして分析を中断します。問題がない場合は、値を大きくして再実行してください。

8.4.4 データセット出力

ランダム抽出されたオブザベーションが outsamp=パラメータとouttest=パラメータに指定されたデータセットに出力されます。

8.4.5 欠損値の取り扱い

y=層別変数に文字タイプ、数値タイプいずれの変数を指定した場合も、欠損値は有効な値の1つとみなされます。

数値タイプのターゲット変数の欠損値(.)は、特殊欠損値(._,.A~.Z)と区別して他の数値と同様に取り扱われます。

8.4.6 制限

入力データセットに以下の変数が存在する場合、警告を出して処理を中止します。入力データセットから削除しておくか、変数名を変えてください。

```
_remain_n _got_n _target_n _random _chkrand
_seed _targflg _obsno
```

8.4.7 コマンド実行モードでの注意

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

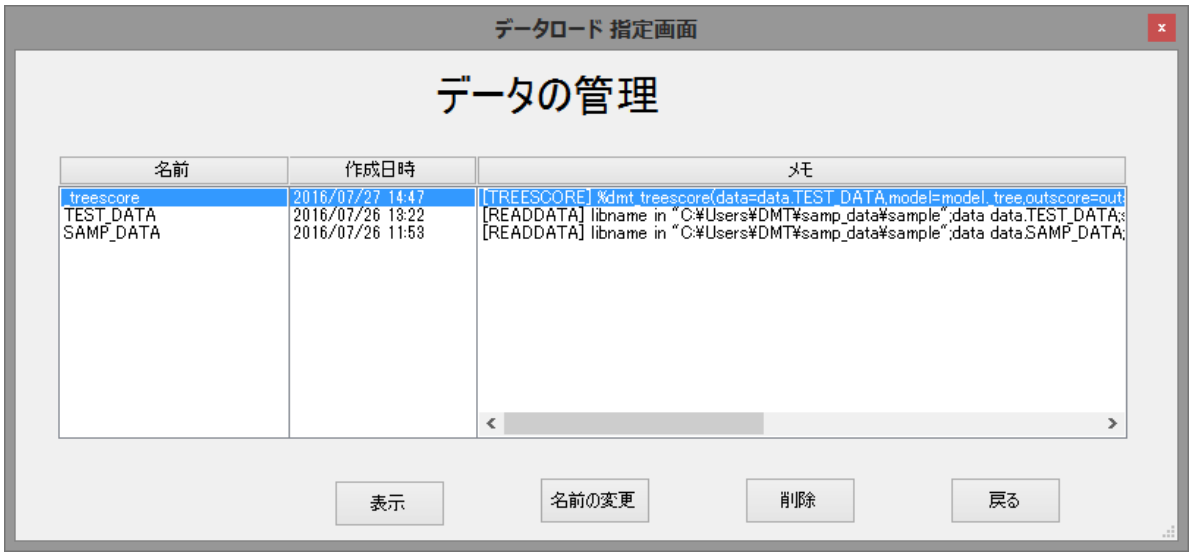
また、以下のユーザ定義フォーマットがWORKライブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
$_item
```

さらに、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

e_name e_type lab&i nobs spc&i typ&i zketa
_speclen _specnum _errormsg

8.5 データ管理



8.5.1 概要

「データ読込」、「データ加工」、「検証確保」、「予測付与」画面で作成した「分析ディレクトリ」の下の**「データセットディレクトリ」**に保存されているデータセットを操作（表示・名前の変更・削除）します。
この機能はマクロモジュールには含まれていません。GUI実行モードでのみ指定可能です。

メモ欄の最初の鍵カッコは以下の画面で作成されたことを表します。

[READDATA] ... データ読込
[CONVDATA] ... データ加工
[SAMPDATA]と[TESTDATA] ... 検証確保
[TREESCORE] ... 予測付与

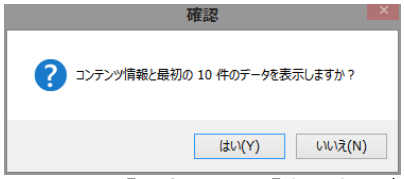
続いてデータを作成したときに実行したプログラムが記述されています。

8.5.2 操作方法

名前,
作成日時,
メモ
と書かれたバーをクリックすると、データセットリストをその項目の昇順・または降順で並べ替えることができます。

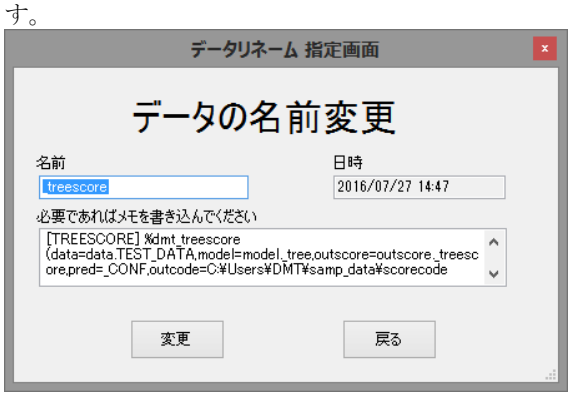
操作したいデータセット名をクリックして選択すると、操作ボタンが表示されますので、表示・名前の変更・削除の操作を行います。

表示　データの内容を表示します。



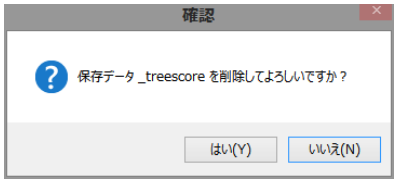
表示件数は**「設定画面」**の**「表示するデータ件数の上限」**で変更可能です。

名前の変更　データの名前とメモ内容を確認・変更します。



名前は本システムが自動で付与する接頭辞（検証確保での SAMP_ と TEST_ ）を考慮して半角英数字で32-5=27文字以内に設定してください。
（先頭はアルファベットまたは_(アンダーバー)）
なお、名前の変更は、元の名前を参照している他の項目（モデル作成画面の入力パラメータ値など）とは自動連動しません。そのため、再指定が必要になるなどの影響があります。

削除　データを削除します。



削除すると、元に戻せません。

合は、**「設定画面」**の**「分析ディレクトリ」**の下の**「データセットディレクトリ」「表示」**ボタンを押し、起動するWindowsエクスプローラで行うと便利です。削除したいデータセット名が書かれたディレクトリをすべて同時選択してから削除します。

(TIPS) 多数のデータセットをまとめて削除したい場

9. 分析画面 ②項目分析

モデル作成を行う前に、分析する各項目の分布とモデルの目的変数との関連を調べます。

9.1 クロス分析（dmt_cross）

9.1.1 概要

クロス分析（DMT_CROSS）はモデル作成の事前分析に用いるためのツールです。主な用途は次の3つです。

（1）　説明変数の選択とツリーモデル作成画面との連携
（2）　説明変数の分布の確認
（3）　ターゲット変数の分布の確認

（１）　説明変数の選択とツリーモデル作成画面との連携
多数の説明変数候補が存在するような場合、その中からあらかじめターゲット分布との関連性が一定以上認められる説明変数のみを選択して、モデル作成に進むと分析効率が良くなります。DMT_CROSS を実行すると、説明変数をAIC値という汎用的な統計的基準により一律に評価し、評価値に従って説明変数の説明力に序列がつけられます。そして、統計的有意性が認められる説明変数のみをツリーモデル作成画面の指定へ引き継ぐことができます。

（２）　説明変数の分布の確認
数値タイプ説明変数については、可能な限り件数が等しくなるように事前に自動カテゴライズを行い、カテゴリごとの数値の範囲が表示されます。このとき、欠損値は欠損値だけのカテゴリが生成されます。分析結果から、外れ値や異常値、また欠損値の割合などをチェックすることが可能です。

文字タイプ説明変数については、各カテゴリの該当件数をそのまま表示していますので、件数が異常に少ない、または多いカテゴリが無いかどうか、カテゴリバランスに違和感が無いかどうか、また存在するはずの無いカテゴリが入っていないかどうか、といったチェックが可能です。

（３）　ターゲット変数の分布の確認
任意の説明変数のカテゴリ別に、ターゲット／非ターゲットの出現件数と出現率、もしくは数値タイプターゲット変数の平均値、標準偏差、最小値、最大値の集計結果をレポートします。例えば、明細データから顧客数や金額（顧客単価）を店舗別、日別、などで集計を行い、グラフ表示することもDMT_CROSSとDMT_CROSSPLOTを用いて行うことができます。

9.1.2 指定方法

（コマンド実行モードでの指定）

```
%dmt_cross(help,data=,control=
,x=,dropx=&y,outcross=_cross,outfmt=_fmt
,outaic=_aic,oaicall=_aicall
,lastcatm=N,nomergen=STURGES
,crossaic=,title=,crosslvl=1,print=,labeldat=&data
,maxcatn=1000,itmunit1=100,itmunit2=10
,order=,pctf=7.2,meanf=best8.,aicf=best8.
,d_label=[D].c_label=[C],dif_label=[D]-[C]
,language=JAPANESE,std_mod_min_n=9
,outhtml=dmt_crosstab.html,outpath=)
```

（GUI実行モードでの変更点）
・help, outhtml=, outpath=パラメータは指定不可。（outhtml=, outpath=指定は自動で行われます。）
・ 常に print=N（ただし、実行終了後に分析結果表の表示が可能です）
・lastcatm=, nomergen=, maxcatn=はオプション画面で指定します。

（必須パラメータ）

以下の5個のパラメータの内、data=, y=, x= の3個は常に必須指定です。control=パラメータは、施策実施効果をAIC評価する場合に対照群データの指定に用います。また、target=パラメータは、ターゲット値の出現率に関するAIC計算を行う場合に指定しなければなりません。

入力データセットの指定（data=)
入力対照データの指定（control=)
ターゲット変数名の指定（y=)
　.... (単一変数名のみ指定可)
説明変数リストの指定（x=)
　.... (例: a b c x1-x4 a--z f_:)
ターゲット値の指定（target=)
　....（ターゲット／非ターゲットの度数分割表におけるAIC計算を行う場合にのみ必須）

（オプションパラメータ）

以下の27個のパラメータは任意指定です。（=の右辺の値はデフォルト値を表しています）

help ... 指定方法のヘルプメッセージの表示.（コマンド実行モードでのみ有効）
クロス出力データセット名の指定（outcross=_cross)
　... DMT_CROSSTAB, DMT_CROSSPLOTで用いる。
AIC値出力データセット名の指定（outaic=_aic)
全変数のAIC値出力データセット名の指定
　(oaicall=_aicall)
説明変数間のクロスレベルの指定（crosslvl=1)
　.... (1 または 2)
最終カテゴリ併合指定（lastcatm=N)
　... 数値説明変数カテゴライズにおいて、最後のカテゴラリ件数が少ない場合１つ前のカテゴリに併合するか否かの選択(Y/N).
非併合数値タイプ説明変数最大カテゴリ数
　(nomergen=STURGES)
　... 指定の値を超える値の種類数（欠損を除く）を持つ数値説明変数はカテゴライズしてから分析に用いる.(デフォルトはスタージェスの式の値)
x=説明変数リストから除外する変数リストの指定
　(dropx=&y)
クロスレベル2のAIC値基準（crossaic=)
　... 2変数のクロス変数のAIC計算結果において表示する最大AIC値の指定.(ブランク、または最大AIC値のいずれか)
分析結果表の表示出力を行うか否かの選択（print=)
　... 画面に結果を表示するか否かの選択(Y/N).明示的な指定が無い場合は分析する説明変数の数が99以下の場合はY,100以上の場合はNに自動設定.（コマンド実行モードでのみ有効. GUI実行

モードでは常にN）
分析結果のカテゴリ表示順序の指定（order=）
... クロス分析結果表における説明変数値の並び順を指定します. (order= /A/D) 値の昇順(ブランク), ターゲット出現率または平均値の昇順(A), ターゲット出現率または平均値の降順(D)（コマンド実行モードでのみ有効. GUI実行モードでは常にブランク）
変数ラベルと値ラベルを使用（labeldat=&data）
... 説明変数のラベルとフォーマットを指定のデータセットのディスクリプタ部を参照して使用.（コマンド実行モードでのみ変更可能. GUI実行モードでは常にデフォルト値 &data に設定）
分析に用いる文字タイプ説明変数の最大カテゴリ数(maxcatn=1000)
... 分析に用いる文字タイプ説明変数の最大カテゴリ数の指定
全体の標準偏差を用いる最小カテゴリ件数の指定(std_mod_min_n=9)
（コマンド実行モードでのみ有効）
itmunit1=100
... 1個の表に出力する説明変数の個数単位の指定.（コマンド実行モードでのみ有効）
itmunit2=10
... クロスレベル=2の表を出力するとき、1個の表に出力する
説明変数１の個数単位の指定.
（コマンド実行モードでのみ有効）
集計フォーマット定義データセットの出力先の指定(outfmt=_fmt)
... DMT_CROSSTAB, DMT_CROSSPLOTで用いる.（コマンド実行モードでのみ有効. GUI実行モードでは自動保存）
画面出力のタイトルの指定（title=）
.... %str,%nrstr,%bquote などの関数で囲んで指定する（コマンド実行モードでのみ有効）
百分率の表示フォーマットの指定（pctf=7.2）
平均値･標準偏差の表示フォーマットの指定(meanf=best8.)
AIC値の表示フォーマットの指定（aicf=best8.）
差分AIC分析結果表における処理群(DATA)を表す　記号（d_label=[D]）
差分AIC分析結果表における対照群(Control)を表す記号（c_label=[C]）
差分AIC分析結果表における処理群−対照群間の差を表す記号（dif_label=[D]-[C]）
言語の選択（language=JAPANESE）
HTML出力ファイル名（outhtml=dmt_crosstab.html）（コマンド実行モードでのみ有効）
HTMLファイル出力ディレクトリの指定（outpath=）（コマンド実行モードでのみ有効）

9.1.3 パラメータの詳細

入力データ（data=）
このパラメータは省略できません。control=の指定がある場合は、処理群を表す入力データセットを指定します。
例：data=a, data=a(where=(DM="1"))

入力対照データ（control=）
処理群と対照群間の応答差を分析するときに、対照群を表す入力データセットを指定します。
例：control=b, contol=a(where=(DM="0"))

ターゲット変数（y=）
ターゲット変数名を指定します。 このパラメータは省略できません。
例：y=flag,　y=sales_amount

ターゲット値（target=）
ターゲット値を指定します。このパラメータは文字タイプターゲット変数の特定の値、もしくは数値タイプターゲット変数の特定の値もしくは範囲をターゲット値とみなして、その出現率（または実施群と非実施群間の出現率の差）を分析したい場合は省略できません。（数値タイプターゲット変数の値そのものの分布の違いを分析したい場合は指定してはいけません。）

ターゲット変数が文字タイプの場合は1種類の値を指定します。特殊な文字（+,-など）を含まない限り引用符で囲む必要はありません。ターゲット変数が数値タイプの場合は1種類の値、もしくはあるしきい値を境とした「以上」、「以下」、「超」、「未満」のいずれかの範囲を指定可能です。数値変数タイプで範囲を指定する場合は引用符で囲んではいけません。

例1：y=flag,target=A　（ターゲット変数が文字タイプ変数で、その値"A"をターゲットに指定する場合）
例2：y=sales,target=1000　（ターゲット変数が数値タイプで、その値1000をターゲットに指定する場合）
例3：y=sales,target=>1000　（ターゲット変数が数値タイプで、その値1000超をターゲットに指定する場合）
例4：y=sales,target=>=1000　（ターゲット変数が数値タイプで、その値1000以上をターゲットに指定する場合。 target==>1000と指定してもかまいません。）
例5：y=sales,target=<1000　（ターゲット変数が数値タイプで、その値1000未満をターゲットに指定する場合）
例6：y=sales,target=<=1000　（ターゲット変数が数値タイプで、その値1000以下をターゲットに指定する場合。 target==<1000と指定してもかまいません。）

注：文字タイプ変数のターゲット値は、大文字、小文字が区別される点に注意してください。（変数名は大文字・小文字の区別はありません。）

説明変数（x=）

説明変数を指定します。このパラメータは省略できません。間に1個以上のスペースを入れて、複数の説明変数を指定可能です。また、3通りの省略指定(-,--,:)と3つの特殊指定（_ALL_,_NUMERIC_,_CHARACTER_） も利用可能です。

例1：x=age　（説明変数1個を指定）
例2：x=age seibetsu　（説明変数2個を指定）
例3：x=abc1-abc100　（変数名がabcで始まり1から100までの数字で終わる100個の説明変数を指定）
例4：data=a,x=nenrei--jukyo　（入力データセットaに含まれる変数を定義された変数順で検索して、nenreiからjukyoの範囲に含まれる全変数を説明変数に指定）
例5：data=a,x=abc:　（入力データセットaに含まれるabcで始まる全説明変数を指定）
例6：x=age x1-x5 q: time--yz1 nenshu　（説明変数指定方法の複合例）
例7；x=_all_　（全変数）
例8；x=_character_　age（全文字タイプ変数とage）

除外する説明変数 (dropx=&y)
x=パラメータと組み合わせて用い、x=パラメータに指定した説明変数の中で分析から除外する説明変数を指定します。
デフォルトは dropx=&y すなはち、ターゲット変数が除外されます。なお、dropx=パラメータに何か指定すると、常にターゲット変数も除外変数に加わります。x=パラメータにターゲット変数を指定し、dropx=, と明示的にブランク指定を行った場合のみターゲット変数は除外されずに分析に加わることになります。
x=パラメータと同じ指定方法が使えます。

例：
x=_all_,dropx=a_:　（a_で始まる変数およびターゲット変数以外のdata=入力データセットの全変数を説明変数に指定）

クロスレベル (crosslvl=1)
説明変数間のクロスをとった説明変数を作成してターゲットとの関連性を分析するか否かを指定します。CROSSLVL=1（デフォルト）の場合は説明変数同士の組み合わせは分析しません。すべての説明変数についてターゲットとの関連性を別々に分析します。CROSSLVL=2を指定すると、CROSSLVL=1の分析に加えて、全説明変数から2つの説明変数を取り出す全組み合わせについて説明変数間のクロス説明変数を作成し、ターゲット変数との関連性を分析し、以下の要件を満たすクロス説明変数について結果を報告します。

（1） crossaic=パラメータに値を指定しなかった場合（デフォルト）
クロス説明変数とターゲット変数とのAIC値をCross_AIC、クロス説明変数を構成する元の2つの説明変数とターゲット変数とのAIC値をそれぞれSubset_AIC1，Subset_AIC2とすると、以下の4個のいずれかのケースに合致するCross_AICを持つクロス説明変数についてのみ結果を表示します。

ケース	Cross_AIC	Subset_AIC1	Subset_AIC2
1	**負**	**非負**	**非負**
2	Subset_AIC2 **未満**	**非負**	**負**
3	Subset_AIC1 **未満**	**負**	**非負**
4	(Subset_AIC1+Subset_AIC2)**未満**	**負**	**負**

ケース１は元の2変数ともに単独ではターゲットと関連がないと認められるのにクロスを取った変数とは関連が認められる場合です。ケース２とケース３は単独では一方の変数がターゲットと関連があるがもう一方は関連が無い場合で、組み合わせると関連性が高くなるケースです。ケース４は元の2変数が共に単独でターゲットと関連があるが、組み合わせた場合の相乗効果が高いと認められるケースです。

（2） crossaic=パラメータに値を指定した場合
crossaic=パラメータに指定した値以下のAIC値を持つクロス説明変数の結果を出力します。
例：crosslvl=2,crossaic=-100　（ターゲット変数との分割表のAIC値が-100以下のクロス説明変数のみ表示します。）

クロスレベル2のAIC値基準 (crossaic=)
説明変数間のクロスをとった説明変数（クロス説明変数）の分析結果表示を制御します。（crosslvl=パラメータの説明を参照）デフォルトはブランク（指定なし）です。

出力クロス集計データ (outcross=_cross)
分析集計結果をデータセットに出力します。指定が無くても _cross という名前でWORKライブラリに出力されます。このデータセットはDMT_CROSSTAB, DMT_CROSSPLOTの入力に用います。

出力AIC統計量データ (outaic=_aic)
説明変数ごとのAIC値をデータセットに出力します。指定が無くても _aic という名前でWORKライブラリに出力されます。

出力全AIC統計量データ (oaicall=_aicall)
crosslvl=2を指定した場合、分析する説明変数の2つの全組合せを含むAIC値をデータセットに出力するよう指定します。 指定が無くても _aicall という名前でWORKライブラリに出力されます。crossslvl=1

の場合はoutaic=出力データセットと全く同じ内容になります。

非併合数値タイプ説明変数最大カテゴリ数 (nomergen=STURGES)
個々の数値タイプ説明変数のカテゴライス方法に関して、欠損値を除いた値の種類数がこの値以下の場合、その数値説明変数は個々の値をカテゴリとみなすように指定します。デフォルトはスタージェスの公式により計算された値です。

CEIL(1+log2(N))

ただし、CEILは整数値への切り上げ関数、log2は2を底とする対数関数、Nは欠損値を除くデータ件数を表します。

分析に用いる文字タイプ説明変数の最大カテゴリ数 (maxcatn=1000)
この指定は文字タイプ変数が単なるオブザベーション識別変数であって分析対象では無いとみなすためのパラメータです。デフォルトは1000です。文字タイプ説明変数のカテゴリ数が指定の数を超える場合、その文字タイプ説明変数は分析対象から除外されます。

最終カテゴリ併合 (lastcatm=N)
数値タイプ説明変数のカテゴライズ方法に関して、最後のカテゴリを最後から2番目のカテゴリに併合するか否かを指定します。デフォルトはN（併合しない）です。
「ノード分割アルゴリズム」の「（１）数値説明変数のカテゴライズ」に記載したように、一般にタイが存在する数値変数（たとえば年齢）の場合、カテゴライズ結果は最後にカテゴリのみ他のカテゴリより件数がかなり少なくなる可能性があります。そのため最後のカテゴリを1つ前のカテゴリと併合する方がモデルの安定性が高まる場合があります。

9.1.4 クロスレベル 2 の既定の数値変数のカテゴライズ

クロスレベル1の数値変数の既定のカテゴライズ方法は、以下のスタージェスの式によってカテゴリ数を決定しています。

階級数=ceil(1+log2(N))

しかし、クロスレベル2においては、クロス変数のカテゴリ数が多くならないようにするため、各数値変数の既定のカテゴリ数を、以下のようにスタージェスの式の値の平方根をとった値とし、その組合せによってクロス変数のカテゴリが生成されるようにしています。

クロスレベル2の個々の数値変数の階級数
=ceil(sqrt(ceil(1+log2(N))))

この仕組みを無効にするには、nomergen=パラメータに数値変数のカテゴリ数を定数で指定してください。

9.1.5 ツリーモデルとの連携機能

GUI実行モードでは、クロス分析実行後  を押すと、クロス分析で指定した入力データ、ターゲット変数、ターゲット値、説明変数などのパラメータを引き継ぎますが、説明力が無いと判断された説明変数は除外する指定（dropx=パラメータに追加指定）を行ったツリーモデル作成画面に移行します。

コマンド実行モードでは、目的変数と関連があると判定された説明変数項目をグローバルマクロ変数 &_XSEL、関連が無いと判定された説明変数項目を &_XDEL にそれぞれ出力します。これらは同じSASまたはWPSセッション内で、続いてツリーモデルを作成するとき説明変数指定を容易にするために用いることができます。

なお、いずれのモードでも、クロスレベル=2を指定した場合は、クロスレベル=2で有意（AIC<0）となったクロス変数を構成する説明変数も説明力があると判断されます。

例：
```
/* (1-1)説明変数ごとの関連分析 */
%dmt_cross(data=samp_data,y=flg,target=1,
x=sei nenrei jukyo kazoku_kosei gakureki shokushu
kinmusaki gyoshu nenshu DM)

/* (1-2)ツリー分析 */
%dmt_tree(data=samp_data,y=flg,target=1,
x=&_XSEL,
mincnt=50,maxlvl=10,outmodel=tree1)
```

9.1.6 コマンド実行モードで有効なパラメータの詳細

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。
例：%dmt_cross(help)

itmunit1=100
１つの表として画面出力する単独説明変数とターゲット変数との関連表に含まれる単独説明変数の最大数を設定します。デフォルトは100。指定の数を1単位として、分析結果は別々の表に出力されます。なお、カテゴリ数が非常に多い文字タイプ説明変数を多数分析するような場合、表出力を行うTABULATEプロシジャがコンピュータ資源不足などの理由でエラーになる可能性があります。そのような場合でもoutcross=パラメータに指定した分析結果データセットとoutfmt=パラメータに指定したフォーマット定義データセットが出力されていることを確認してください。（指定がなくてもデフォルトでWORK._cross、WORK._fmt にそれぞれ出力されます。）これら2つのデータセットが出力されていれば、再度

DMT_CROSSを実行しなくても、DMT_CROSSTABを用いて任意の範囲の結果集計表の画面表示を自由に行うことができます。

itmunit2=10
１つの表として画面出力するクロス説明変数とターゲット変数との関連表に含まれるクロス説明変数の最大数を設定します。デフォルトは10。指定の数を1単位として、分析結果は別々の表に出力されます。なお、カテゴリ数が非常に多い文字タイプ説明変数を多数分析するような場合、表出力を行うTABULATEプロシジャがコンピュータ資源不足などの理由でエラーになる可能性があります。そのような場合でも outcross=パラメータに指定した分析結果データセットとoutfmt=パラメータに指定したフォーマット定義データセットが出力されていることを確認してください。（指定がなくてもデフォルトでWORK._cross、WORK._fmt にそれぞれ出力されます。）これら2つのデータセットが出力されていれば、再度DMT_CROSSを実行しなくても、DMT_CROSSTAB （crosslvl=2パラメータを指定）を用いて任意の範囲の結果集計表の画面表示を自由に行うことができます。

labeldat=&data
ラベルとフォーマットが定義されたデータセットを指定することにより、分析結果の全変数名と文字タイプ変数値に、それぞれ定義された変数ラベルとフォーマットが付加されて表示されるようになります。デフォルトは分析データセットに設定してあり、変数ラベルやフォーマットが定義されていた場合、自動的に用いられます。もしもラベルもフォーマットも定義されていない場合は、変数名、変数値がそのまま表示されます。
数値タイプ説明変数には、フォーマットが定義されていたとしても無視します。なお、フォーマット定義された変数を含むデータセットをアクセスするためには、そのフォーマットライブラリもアクセス可能になっている必要がありますので、ラベル定義されたデータセットを保存して再利用したい場合は、フォーマットライブラリも保存しておく必要があります。

outfmt=_fmt
DMT_CROSSで算出した集計表における数値変数の範囲などを表示するためのフォーマット定義をデータセットに出力します。このデータセットはDMT_CROSSTAB, DMT_CROSSPLOTの入力に用います。

print=
実行結果の画面出力を行うか否かを指定します。デフォルトは分析する（実際に集計を行う）説明変数の数が99個以下の場合はY（画面出力を行う）、100個以上の場合はN（画面出力を行わない）です。print=N（画面出力は行わない）指定は入力データセットの件数が多く、かつ、説明変数の数も多く時間がかかりそうな分析の場合に効果的です。実行後、DMT_CROSSTAB, DMT_CROSSPLOTを用いて、特定の説明変数や、ターゲット変数と関連の強い上位の説明変数に絞って画面表示するといった使い方ができます。

title=
画面出力される表にタイトルを指定できます。指定しない（デフォルト）場合、以下のタイトルが自動的に付与されます。

%bquote(DMT_CROSS 分析結果: データセット:&data ターゲット:&y=%left(&target))

タイトルを指定する場合、上記のように%bquote関数の中に記述してください。

9.1.7 HTML 出力

分析結果の図表はhtmlファイルに出力されます。保存先はデフォルトではSASディスプレイマネージャまたはWPSワークベンチの管理下（ワークスペース内の一時保存ファイル）です。outpath=パラメータを指定すると、保存先を変更できます。（必ずフルパス指定します。引用符で囲んでも囲まなくてもかまいません）同時にouthtml=パラメータを指定すると、保存するhtmlファイルに自由に名前を付けることができます。

outhtml=dmt_crosstab.html
分析結果を保存するHTML出力ファイル名を指定します。
例：outhtml=out1.html,

outpath=
HTML図表出力ファイルの保存ディレクトリを指定します。このパラメータを指定しない場合（デフォルト）、HTMLファイルはSASディスプレイマネージャまたはWPSワークベンチの管理下に作成されます。outpath=指定を行う場合、値は必ずフルパスで指定する必要があります。なお、パス指定全体を引用符で囲んでも囲まなくてもかまいません。
例：outpath='G:¥temp'

9.1.8 実行例

GUI実行モードではprint=Nに設定されています。しかし、実行後に分析結果表示出力を行うかどうかを選択可能です。
コマンド実行モードではprint=パラメータを指定しない場合、分析する説明変数の数が99個以内の場合はprint=Y、100個以上の場合はprint=Nに設定されます。

(1)target=パラメータを指定し、ターゲット出現率の分布との関連を分析する場合

ターゲットの出現率と各説明変数の統計的関連性を

AIC 値で評価し、関連の強い順に説明変数をリストした表を出力します。

%dmt_cross(data=samp_data,y=flg,target=1,x=seinenshu)

DMT_CROSS 分析結果: 分析データセット: samp_data, ターゲット: flg="1"

				トータル件数	ターゲット件数	ターゲット再現率%	ターゲット出現率%
NO	AIC値	説明変数	値				
0	.	{ANY}	{ALL}	2,000	457	100.00	22.85
1	-16.4648	SEI 性別	1 男性	1,291	256	56.02	19.83
			2 女性	709	201	43.98	28.35
2	0.77788	NENSHU 年収	.	555	112	24.51	20.18
			102~255	121	36	7.88	29.75
			256~302	122	24	5.25	19.67
			303~349	124	43	9.41	34.68
			350~400	121	32	7.00	26.45
			401~449	123	34	7.44	27.64
			450~500	121	26	5.69	21.49
			501~552	122	18	3.94	14.75
			553~602	124	30	6.56	24.19
			603~663	122	28	6.13	22.95
			664~736	125	28	6.13	22.40
			737~834	121	26	5.69	21.49
			836~1278	99	20	4.38	20.20

タイトル:
コマンド実行モードでは、**分析データ名** と **ターゲット変数名=ターゲット値** が表示されます。

項目:（**カッコ内は** language=English **を指定した場合の項目表示**）

NO（NO）... 説明変数の関連の強さの順序を表します。 ただし、NO=0 は全体の集計値を意味します。

AIC 値（AIC）... AIC 統計量。値が負で絶対値が大きいほど目的変数との関連が強いことを意味します。上記の例では、全体のターゲット出現率 22.85%に対して、男性のターゲット出現率は 19.83%、女性のターゲット出現率は 28.35%となっており、観測された男女間の出現率の差異は統計的に有意であることを示しています。一方、年収については、ほぼ等しい件数になるようにスタージェスの式（ceil(1+log2(2000-555))）による 12 個＋1 個（欠損値）の順序カテゴリに分けたときのカテゴリ別出現率を集計すると、14.75%～34.68%の範囲に分布しますが、AIC>0 と計算され、ターゲット出現率との関連性は無いという解釈になります。（なお、数値変数の関連性はカテゴリ数を減らす（例えば nomergen=5 と指定する）と有意になりやすくなります。また、数値変数、文字変数ともに、分析データ件数を増やすことによっても有意になりやすくなります。）

説明変数（ITEM）... 説明変数名。変数ラベルが定義されている場合は変数ラベルも表示されます。

値（VALUE）... カテゴリ値。 数値変数は自動的にカテゴライズされます。

トータル件数（TOTAL-(N)）... カテゴリの総件数
ターゲット件数（TARGET-(N)）... カテゴリのターゲット件数
ターゲット再現率%（SUPPORT-(%)）... ターゲット件数／（No=0 のターゲット件数）＊１００
ターゲット出現率%（CONFIDENCE-(%)）... ターゲット件数／トータル件数＊１００

CROSSLVL=2 を指定した場合は、クロス説明変数を作成し、それとターゲットとの関連表を表示します。

%dmt_cross(data=samp_data,y=flg,target=1,x=seinenshu,crosslvl=2)

DMT_CROSS 分析結果: 分析データセット: samp_data, ターゲット: flg="1"

				トータル件数	ターゲット件数	ターゲット再現率%	ターゲット出現率%
NO	AIC値	説明変数	値				
0	.	{ANY}	{ALL}	2,000	457	100.00	22.85
1	-16.4648	SEI 性別	1 男性	1,291	256	56.02	19.83
			2 女性	709	201	43.98	28.35
2	-2.28293	NENSHU 年収	.	555	112	24.51	20.18
			102~348	363	102	22.32	28.10
			349~498	364	92	20.13	25.27
			499~655	364	76	16.63	20.88
			656~1278	354	75	16.41	21.19

DMT_CROSS 分析結果: 分析データセット: samp_data, ターゲット: flg="1"

						トータル件数	ターゲット件数	ターゲット再現率%	ターゲット出現率%
NO	AIC値	説明変数1	値1	説明変数2	値2				
3	-19.7049	NENSHU 年収	.	SEI 性別	1 男性	393	79	17.29	20.10
					2 女性	162	33	7.22	20.37
			102~348	SEI 性別	1 男性	235	53	11.60	22.55
					2 女性	128	49	10.72	38.28
			349~498	SEI 性別	1 男性	223	44	9.63	19.73
					2 女性	141	48	10.50	34.04
			499~655	SEI 性別	1 男性	224	45	9.85	20.09
					2 女性	140	31	6.78	22.14
			656~1278	SEI 性別	1 男性	216	35	7.66	16.20
					2 女性	138	40	8.75	28.99

crosslvl=2 を指定すると、クロスレベル 1 の結果とクロスレベル 2 の結果が別々に表示されます。年収と性別のカテゴリの組合せを新たなカテゴリとして持つクロス変数「年収*性別」の AIC 値-19.7049 は個々の AIC 値-16.4648 と-2.28293 のいずれよりも小さく、クロスレベル 2 のクロス変数分析結果の表示ルール「ケース 4」に該当するため、分析結果表に表示されています。
ただし、数値変数 NENSHU のカテゴリ数が ceil(sqrt(12))=4 に変更されています。（カテゴリ数が減った関係で年収とターゲット出現率とは関連があるという分析結果に変化しています。）

クロス説明変数の項目：（カッコ内は language=English を指定した場合の項目表示）

NO（NO）... 説明変数の関連の強さの順序を表します。 クロスレベル 1 の最後の番号に続いて番号付けされます。ただし、NO=0 は全体の集計値を意味します。

AIC 値（AIC）... AIC 統計量（値が負で絶対値が大きいほど目的変数との関連が強いことを表します）

説明変数 1（ITEM1）... クロス説明変数を構成する説明変数 1 の名前

値 1（VALUE1）... クロス説明変数を構成する説明変数 1 のカテゴリ値

説明変数 2（ITEM2）... クロス説明変数を構成する説明変数 2 の名前

値 2（VALUE2）... クロス説明変数を構成する説明変数 2 のカテゴリ値

以下、CROSSLVL=1 の表と同じ項目が表示されます。

(2)target=パラメータを指定せず、ターゲット変数の分布との関連を分析する場合

ターゲット変数の値の変動と各説明変数の統計的関連性を一元配置分散分析モデルにおける AIC 値で評価し、関連の強い順に説明変数をリストした表を出力します。目的変数が欠損値のオブザベーションは削除されます。

%dmt_cross(data=samp_data,y=nenshu,x=seinenrei)

DMT_CROSS 分析結果: 分析データセット: samp_data, ターゲット: nenshu

				件数	平均値	標準偏差	最小値	最大値
NO	AIC値	説明変数	値					
0	.	{ANY}	{ALL}	1,445	514.0498	202.7175	102	1278
1	-1.42879	SEI 性別	1 男性	898	506.3474	201.0586	102	1249
			2 女性	547	526.6947	204.7845	102	1278
2	7.401576	NENREI 年齢	20~22	126	494.8254	195.3707	108	1070
			23~25	132	524.4924	215.3209	125	1052
			26~28	122	556.6721	245.1042	161	1278
			29~32	149	516.1544	208.4451	166	1245
			33~36	144	489.8958	197.9444	106	1111
			37~40	144	513.7569	200.249	139	1198
			41~44	169	515.0355	194.371	102	1115
			45~48	150	534.9933	192.5019	149	937
			49~52	139	492.0576	181.3424	104	1138
			53~58	126	503.3254	188.1139	102	1217
			59~60	44	517.4773	199.0686	126	861

表示出力リストの項目の説明:（カッコ内は英語設定の場合の表示です。）

NO（NO）... 説明変数の関連の強さの順序を表します。 ただし、NO=0 は全体の集計値を意味します。目的変数NENSHUが欠損のオブザベーションは削除されて、残りの 1,445 件が分析に用いられていることがわかります。

AIC 値（AIC）... AIC 統計量（値が負で絶対値が大きいほど目的変数との関連が強いことを表します）

説明変数（ITEM）... 説明変数名

値（VALUE）... カテゴリ値。数値変数は自動的にカテゴライズされます。なお、NENREI のカテゴリはスタージェスの式（ceil(1+log2(1445))=12）により 12 個のカテゴリに分けることを目標としましたが、タイ値が多く存在したため、11 個のカテゴリに縮小されています。

件数（TOTAL-(N)）... カテゴリ値に該当する件数
平均値（MEAN）... カテゴリのターゲット変数平均値
標準偏差（STD）... カテゴリのターゲット変数標準偏差
最小値（MIN）... カテゴリのターゲット変数最小値
最大値（MAX）... カテゴリのターゲット変数最大値

CROSSLVL=2 を指定した場合は、クロス説明変数とターゲット変数との関連表を表示します。（ここでは省略）

(3)control=パラメータとtarget=パラメータを指定し、処理群と対照群間のターゲット出現率の差を分析する場合

ターゲット出現率の実施群-対照群間の差と各説明変数の統計的関連性を AIC 値で評価し、関連の強い順に説明変数をリストした表を出力します。

例：%dmt_cross(data=samp_data(where=(DM="1"))
,control=samp_data(where=(DM="0"))
,y=flg,target=1,x=sei nenshu)

DMT_CROSS 分析結果: 分析データセット[D]: samp_data(where=(DM="1")), ターゲット: flg="1", 対照データセット[C]: samp_data(where=(DM="0"))

				[D]-[C]出現率の差%	[D]-[C]出現率の差の標準誤差%	[D]トータル件数	[D]ターゲット件数	[D]ターゲット再現率%	[D]ターゲット出現率%	[C]トータル件数	[C]ターゲット件数	[C]ターゲット再現率%	[C]ターゲット出現率%	個別AIC値
NO	AIC値	説明変数	値											
0	.	{ANY}	{ALL}	11.36	2.03	619	190	100.00	30.69	1,381	267	100.00	19.33	.
1	-42.9607	SEI 性別	1 男性	-1.67	2.51	344	64	33.68	18.60	947	192	71.91	20.27	-19.8344
			2 女性	28.54	3.47	275	126	66.32	45.82	434	75	28.09	17.28	-21.1262
2	13.63147	NENSHU 年収	.	12.47	3.75	162	47	24.74	29.01	393	65	24.34	16.54	1.555897
			102~255	14.42	9.17	35	14	7.37	40.00	86	22	8.24	25.58	1.720172
			256~302	15.37	7.67	40	12	6.32	30.00	82	12	4.49	14.63	1.379922
			303~349	15.33	8.89	45	20	10.53	44.44	79	23	8.61	29.11	1.709971
			350~400	7.81	9.18	31	10	5.26	32.26	90	22	8.24	24.44	1.489359
			401~449	15.87	8.50	42	16	8.42	38.10	81	18	6.74	22.22	1.637706
			450~500	13.27	8.47	32	10	5.26	31.25	89	16	5.99	17.98	1.736804
			501~552	5.98	6.98	37	7	3.68	18.92	85	11	4.12	12.94	1.752822
			553~602	-6.69	8.47	36	7	3.68	19.44	88	23	8.61	26.14	-2.72529
			603~663	7.88	8.42	35	10	5.26	28.57	87	18	6.74	20.69	1.607701
			664~736	13.61	7.77	45	14	7.37	31.11	80	14	5.24	17.50	1.710877
			737~834	18.86	7.73	45	15	7.89	33.33	76	11	4.12	14.47	0.62718
			836~1278	5.07	8.50	34	8	4.21	23.53	65	12	4.49	18.46	1.428351

表示出力項目の説明:（カッコ内は英語設定の場合の表示です。）

NO（NO）... 説明変数の関連の強さの順序を表します。 ただし、NO=0 は全体の集計値を意味します。

AIC 値（AIC）... AIC 統計量（値が負で絶対値が大きいほど目的変数との関連が強いことを表します）

説明変数（ITEM）... 説明変数名

値（VALUE）... カテゴリ値。数値変数は自動的にカテゴライズされます。

[D]−[C]出現率の差%
([D]−[C]Dif. of CONFIDENCE(%)) ... [D]（処理群）のターゲット出現率% - [C]（対照群）の出現率%。
[D]−[C]出現率の差の標準誤差%
([D]−[C]StdErr of Dif. of CONFIDENCE(%)) ... 出現率の差%の推計値としてのばらつき（標準偏差）を表します。小さいほど良い推計値であることを意味します。

以下の集計値は[D]（処理群）と[C]（対照群）それぞれについて表示されます。

トータル件数（TOTAL−(N)）... カテゴリの総件数
ターゲット件数（TARGET−(N)）... カテゴリのターゲット件数
ターゲット再現率%（SUPPORT−(%)）... ターゲット件数／（No=0 のターゲット件数）＊１００
ターゲット出現率%（CONFIDENCE−(%)）... ターゲット件数／トータル件数＊１００

最後に以下の項目が表示されます。

個別 AIC 値 (Each AIC) ... カテゴリの AIC 値。カテゴリの[D]-[C]出現率の差が全体平均の[D]-[C]出現率の差と比較して統計的に有意であるかどうかを判定します。（より正確には、カテゴリごとにターゲットの全体出現率で調整後の処理群と対照群間のターゲット出現率の差の有意性を表す AIC 値を計算しています。）負の値で絶対値が大きいほど有意であることを意味します。説明変数の AIC 値はその変数の各カテゴリの個別 AIC 値を合計した値から 2 を差し引いた値で与えています。
上記の例では、NENSHU に関する個別 AIC 値は、553-602 のカテゴリのみ -2.7 と負の値をとっており、処理群と対照群間のターゲット出現率の差 -6.69 が全体平均の 11.36 と有意であることを示しています。

CROSSLVL=2 を指定した場合は、クロス説明変数とターゲットとの関連表を表示します。（ここでは省略）

(4) control=パラメータを指定し、target=パラメータを指定せず、処理群と対照群間のターゲット変数の平均値の差を分析する場合

ターゲット変数の平均値の実施群-対照群間の差と各説明変数の統計的関連性を AIC 値で評価し、関連の強い順に説明変数をリストした表を出力します。

例：%dmt_cross(data=samp_data(where=(DM="1"))
,control=samp_data(where=(DM="0"))
,y=kingaku,x=sei nenshu)

DMT_CROSS 分析結果: 分析データセット[D]: samp_data(where=(DM="1")), ターゲット: kingaku, 対照データセット[C]: samp_data(where=(DM="0"))

				[D]-[C]平均値の差	[D]-[C]平均値の差の標準誤差	[D]件数	[D]平均値	[D]標準偏差	[C]件数	[C]平均値	[C]標準偏差	個別AIC値
NO	AIC値	説明変数	値									
0	.	{ANY}	{ALL}	29.39109	351.9012	619	129.588	257.2506	1,381	100.197	240.1178	.
1	-51.3579	SEI 性別	1 男性	-48.3264	299.0199	344	56.10756	171.8592	947	104.434	244.6984	-27.0971
			2 女性	130.5538	386.6487	275	221.5055	311.1448	434	90.95161	229.5346	-22.2608
2	-4.89021	NENSHU 年収	.	2.784375	81.80334	162	30.06173	47.22545	393	27.27735	66.79479	-18.9139
			102~255	38.5309	220.6036	35	123.2286	160.752	86	84.69767	151.0786	1.91234
			256~302	72.98841	255.4988	40	134.025	207.0467	82	61.03659	149.7041	0.257554
			303~349	17.21266	289.6953	45	149.2	198.9193	79	131.9873	210.6051	1.900268
			350~400	6.210394	313.9653	31	133.3548	216.9979	90	127.1444	226.9055	1.754218
			401~449	36.44444	323.3434	42	155.6667	232.0595	81	119.2222	225.1651	1.973427
			450~500	10.79284	323.2868	32	120.0625	221.0056	89	109.2697	235.9468	1.848951
			501~552	22.00986	323.8484	37	106.6216	235.8422	85	84.61176	221.9376	1.972565
			553~602	-68.7715	427.1905	36	124.9444	273.6443	88	193.7159	328.0404	-0.48423
			603~663	18.93169	447.7116	35	178.8857	317.7384	87	159.954	315.4171	1.972673
			664~736	87.72083	491.3037	45	236.9333	367.892	80	149.2125	325.6299	1.162337
			737~834	127.9012	516.0578	45	260.8222	402.001	76	132.9211	323.5906	-0.15973
			836~1278	1.774661	618.6431	34	215.8824	418.6953	65	214.1077	455.4268	1.913346

表示出力項目の説明:（カッコ内は英語設定の場合の表示です。）

NO（NO）... 説明変数の関連の強さの順序を表します。ただし、NO=0 は全体の集計値を意味します。

AIC 値（AIC） ... AIC 統計量（値が負で絶対値が大きいほど目的変数との関連が強いことを表します）

説明変数（ITEM） ... 説明変数名

値（VALUE） ... カテゴリ値。数値変数は自動的にカテゴライズされます。

[D]−[C]平均値の差%（[D]−[C]Dif. of MEAN） ... [D]（処理群）のターゲット平均値 - [C]（対照群）の平均値。
[D]−[C]出現率の差の標準誤差%（[D]−[C]StdErr of Dif. of MEAN） ... 平均値の差の推計値としてのばらつき（標準偏差）を表します。小さいほど良い推計値であることを意味します。

以下の集計値は[D]（処理群）と[C]（対照群）それぞれについて表示されます。

件数（TOTAL−(N)） ... カテゴリ値に該当する件数
平均値（MEAN） ... カテゴリのターゲット変数平均値
標準偏差（STD） ... カテゴリのターゲット変数標準偏差

最後に以下の項目が表示されます。

個別 AIC 値(Each AIC) ... カテゴリの AIC 値。カテゴリの[D]-[C]平均値の差が全体の[D]-[C]平均値の差と比較して統計的に有意であるかどうかを判定します。（より正確には、カテゴリごとに全体平均値で調整後の処理群と対照群間の平均値の差の有意性を表す AIC 値を計算しています。）負の値で絶対値が大きいほど有意であることを意味します。説明変数の AIC 値はその変数の各カテゴリの個別 AIC 値を合計した値から 2 を差し引いた値で与えています。
上記の例では、NENSHU に関する個別 AIC 値は、553-602 のカテゴリのみ -0.48 と負の値をとっており、処理群と対照群間の KINGAKU 平均値の差 -68.77 が全体平均の 29.39 と比較して有意であることを示しています。

CROSSLVL=2 を指定した場合は、クロス説明変数とターゲット変数との関連表を表示します。（ここでは省略）

9.1.9 層別分析の例

例えば、住居区分別にクロス分析を行うには、以下のように、コマンド実行モードでマクロ言語を使ったプログラムを書いて実行します。

```
proc freq data=samp_data(keep=jukyo);
   tables jukyo/noprint out=jukyo(keep=jukyo);
run;
data _null_;
   if _n_=1 then call symput("n",compress(n));
   set jukyo nobs=n;
   call symput("JUKYO"||left(_n_),compress(jukyo));
run;

(マクロ変数値の確認)
%put &n, &JUKYO1, &JUKYO2, ... , &&JUKYO&n;

(ログ)
8, , 1, ... , 7

(住居区分別にクロス分析を行う)
%macro create_cross;
   %do i=1 %to &n;
      %dmt_cross(data=samp_data(where=(jukyo="&
```

```
&JUKYO&i")),y=flg,target=1,x=sei--DM,outcross=CR
OSS_&&JUKYO&i,outfmt=FMT_&&JUKYO&i,outaic
=AIC_&&JUKYO&i)
   %end;
%mend create_cross;
%create_cross
```

注意：層別変数は文字変数で半角英数字の短い値を仮定しています。数値変数の場合は、where=(変数名=値)の値を引用符で囲うとエラーになります。

9.1.10 データセット出力

出力クロス集計データ(outcross=_cross)
クロス分析結果画面情報イメージをデータセット出力します。(GUI実行モードでは、**分析ルートディレクトリ¥CROSS¥outcrossの出力データセット名¥outcrossの出力データセット名.WPD** の名前で保存されます)

outcross=出力データセット（control=パラメータ指定なしの場合）

変数名	タイプ	長さ	内容	備考
ITEM_NO	数値	8	変数の順序（AIC値の小さい順）	0は全体を意味する
K	数値	8	説明変数クロスレベル数	0は全体を意味する
AIC	数値	8	AIC統計量	負の絶対値が大きいほど関連大を表す
ITEM1	文字	32	説明変数1	
CAT1	文字	可変	説明変数1のカテゴリ値	
ITEM2	文字	32	説明変数2	crosslvl=2指定の場合
CAT2	文字	可変	説明変数2のカテゴリ値	crosslvl=2指定の場合
TOTAL_N	数値	8	カテゴリ件数	
TARGET_N	数値	8	カテゴリ内ターゲット件数	target=パラメータ指定の場合
CONF_PCT	数値	8	カテゴリ内ターゲット出現率	
SUPPORT_PCT	数値	8	カテゴリ内ターゲット再現率	
MEAN	数値	8	カテゴリ内ターゲット変数平均値	target=パラメータ指定なしの場合
STD	数値	8	カテゴリ内ターゲット変数標準偏差	
MIN	数値	8	カテゴリ内ターゲット変数最小値	
MAX	数値	8	カテゴリ内ターゲット変数最大値	
CATEGORY1	数値	8	説明変数1のカテゴリ値の順序を決めるための変数	文字変数カテゴリはオブザベーション番号、数値変数カテゴリは平均値
item_cat1	文字	可変	説明変数1の名前＋カテゴリ番号	フォーマット表示するためユニークな値を持たせてある
CATEGORY2	数値	8	説明変数2のカテゴリ値の順序を決めるための変数	crosslvl=2指定の場合
item_cat2	文字	可変	変数変数2の名前＋カテゴリ番号	crosslvl=2指定の場合

outcross=出力データセット（control=パラメータ指定ありの場合）

変数名	タイプ	長さ	内容	備考
ITEM_NO	数値	8	変数の順序（AIC値の小さい順）	0は全体を意味する
K	数値	8	説明変数クロスレベル数	0は全体を意味する
AIC	数値	8	AIC統計量	負の絶対値が大きいほど関連大を表す
ITEM1	文字	32	説明変数1	
CAT1	文字	可変	説明変数1のカテゴリ値	
ITEM2	文字	32	説明変数2	crosslvl=2指定の場合
CAT2	文字	可変	説明変数2のカテゴリ値	crosslvl=2指定の場合
TOTAL_N1	数値	9	処理群のカテゴリ件数	
TOTAL_N2	数値	9	対照群のカテゴリ件数	
TARGET_N1	数値	8	処理群のカテゴリ内ターゲット件数	target=パラメータ指定の場合
CONF_PCT1	数値	8	処理群のカテゴリ内ターゲット出現率（%表示）	
SUPPORT_PCT1	数値	8	処理群のカテゴリ内ターゲット再現率（%表示）	
TARGET_N2	数値	8	対照群のカテゴリ内ターゲット件数	
CONF_PCT2	数値	8	対照群のカテゴリ内ターゲット出現率（%表示）	
SUPPORT_PCT2	数値	8	対照群のカテゴリ内ターゲット再現率（%表示）	
CONF_PCT3	数値	8	処理群のターゲット出現率と対照群のターゲット出現率の差（%表示）	
CONF_PCT3_SE	数値	8	ターゲット出現率の差の標準誤差（%表示）	
MEAN1	数値	8	処理群のカテゴリ内ターゲット変数平均値	target=パラメータ指定なしの場合
STD1	数値	8	処理群のカテゴリ内ターゲット変数標準偏差	
MEAN2	数値	8	対照群のカテゴリ内ターゲット変数平均値	
STD2	数値	8	対照群のカテゴリ内ターゲット変数標準偏差	
MEAN1	数値	8	処理群のカテゴリ内ターゲット変数平均値	
MEAN3	数値	8	処理群のターゲット出現率と対照群のターゲット平均値の差	
MEAN3_SE	数値	8	ターゲット平均値の差の標準誤差	
EACH_AIC	数値	8	カテゴリの個別AIC値	
CATEGORY1	数値	8	説明変数1のカテゴリ値の順序を決めるための変数	文字変数カテゴリはオブザベーション番号、数値変数カテゴリは平均値
item_cat1	文字	可変	説明変数1の名前＋カテゴリ番号	フォーマット表示するためユニークな値を持たせてある
CATEGORY2	数値	8	説明変数2のカテゴリ値の順序を決めるための変数	crosslvl=2指定の場合
item_cat2	文字	可変	変数変数2の名前＋カテゴリ番号	crosslvl=2指定の場合

outfmt=出力データセット
outcross=データセットの説明変数名、説明変数カテゴリ値の表示フォーマット定義を格納しています。

（GUI実行モードでは、**分析ルートディレクトリ¥CROSS¥outcrossの出力データセット名¥_fmt.WPD** の名前で自動保存されます）

outfmt=出力データセット

変数名	タイプ	長さ	内容	備考
fmtname	文字	32	フォーマット名	値"_item"は変数名、値"_cat"は変数値に関する。
start	文字	320	開始値	
end	文字	320	終了値	
hlo	文字	1	high/low/other 識別フラグ	
type	文字	1	タイプ	
label	文字	289	説明変数2のAIC値	289=変数名(32)+空白(1)+変数ラベル(256)

出力AIC統計量データ(outaic=_aic)
クロス分析結果画面表示された説明変数とターゲット変数とのAIC値をデータセット出力します。
デフォルトは WORK._ AIC という名前で出力されます。（GUI実行モードでは、**分析ルートディレクトリ¥CROSS¥outcrossの出力データセット名¥_aic.WPD** の名前で自動保存されます）

outaic=出力データセット

変数名	タイプ	長さ	内容	備考
K	数値	8	説明変数クロスレベル数	0は全体を意味する
varname1	文字	32	説明変数1	
varname2	文字	32	説明変数2	crosslvl=2指定の場合
aic	数値	8	AIC値	
subset_aic1	数値	8	説明変数1のAIC値	crosslvl=2指定の場合
subset_aic2	数値	8	説明変数2のAIC値	crosslvl=2指定の場合

出力全AIC統計量データ(oaicall=_aicall)
さらに、outaic=出力データセットと同じ形式のデータセットWORK._AICALLが自動的に出力されます。この中には、crosslvl=2を指定した場合に画面出力されないクロス説明変数を含むすべての説明変数のAIC値が含まれています。（GUI実行モードでは、**分析ルートディレクトリ¥CROSS¥outcrossの出力データセット名¥_aicall.WPD** の名前で自動保存されます）

oaicall=出力データセット

変数名	タイプ	長さ	内容	備考
K	数値	8	説明変数クロスレベル数	0は全体を意味する
varname1	文字	32	説明変数1	
varname2	文字	32	説明変数2	crosslvl=2指定の場合
aic	数値	8	AIC値	
subset_aic1	数値	8	説明変数1のAIC値	crosslvl=2指定の場合
subset_aic2	数値	8	説明変数2のAIC値	crosslvl=2指定の場合

9.1.11 欠損値の取り扱い

文字タイプのターゲット変数、説明変数はいずれも有効な値の1つとみなされます。

数値タイプの説明変数に特殊欠損値(.A~.Z)が存在した場合は通常欠損値(.)に変換された上で使用されます。

数値タイプのターゲット変数の欠損値は、target=パラメータを指定しなかった場合、データに存在すると、そのオブザベーションは分析から除外されます。target=パラメータを指定した場合は、数値タイプのターゲット変数の欠損値(.)は、特殊欠損値(._,.A~.Z)と区別して他の数値と同様に取り扱われます。

9.1.12 制限

処理するオブザベーション数に制限はありませんが、コンピュータ資源等の制約により実質的に取扱えるオブザベーション数には限りがあります。

1度に入力できる説明変数の最大数は2,000です。ただし、各変数のカテゴリ数、その他の要因によるコンピュータ資源不足などの理由で1回の分析では2,000未満の説明変数しか取り扱えない場合もあり得ます。そのような場合は、1回の分析において指定する説明変数の数を少なくして実行してください。特に、crosslvl=2 を指定する場合は変数の数を少なめに設定してください。

入力データセットに以下の変数が存在する場合、警告を出して処理を中止します。入力データセットから削除しておくか、変数名を変えてください。(_v&i.cは_V+数字+Cという形式の変数名を表します。)

_id _item _obsno _targflg _v&i.c

9.1.13 コマンド実行モードでの注意

ユーザ定義フォーマットがついた変数を含むデータセットをアクセスするためには、そのフォーマットも利用可能でなければなりません。ユーザ定義フォーマットのついた変数を含む分析データセットを永久保存する場合は、そのフォーマットも永久保存してください。（注：GUI実行モードでは自動的に利用可能にする仕組みが備わっています）

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

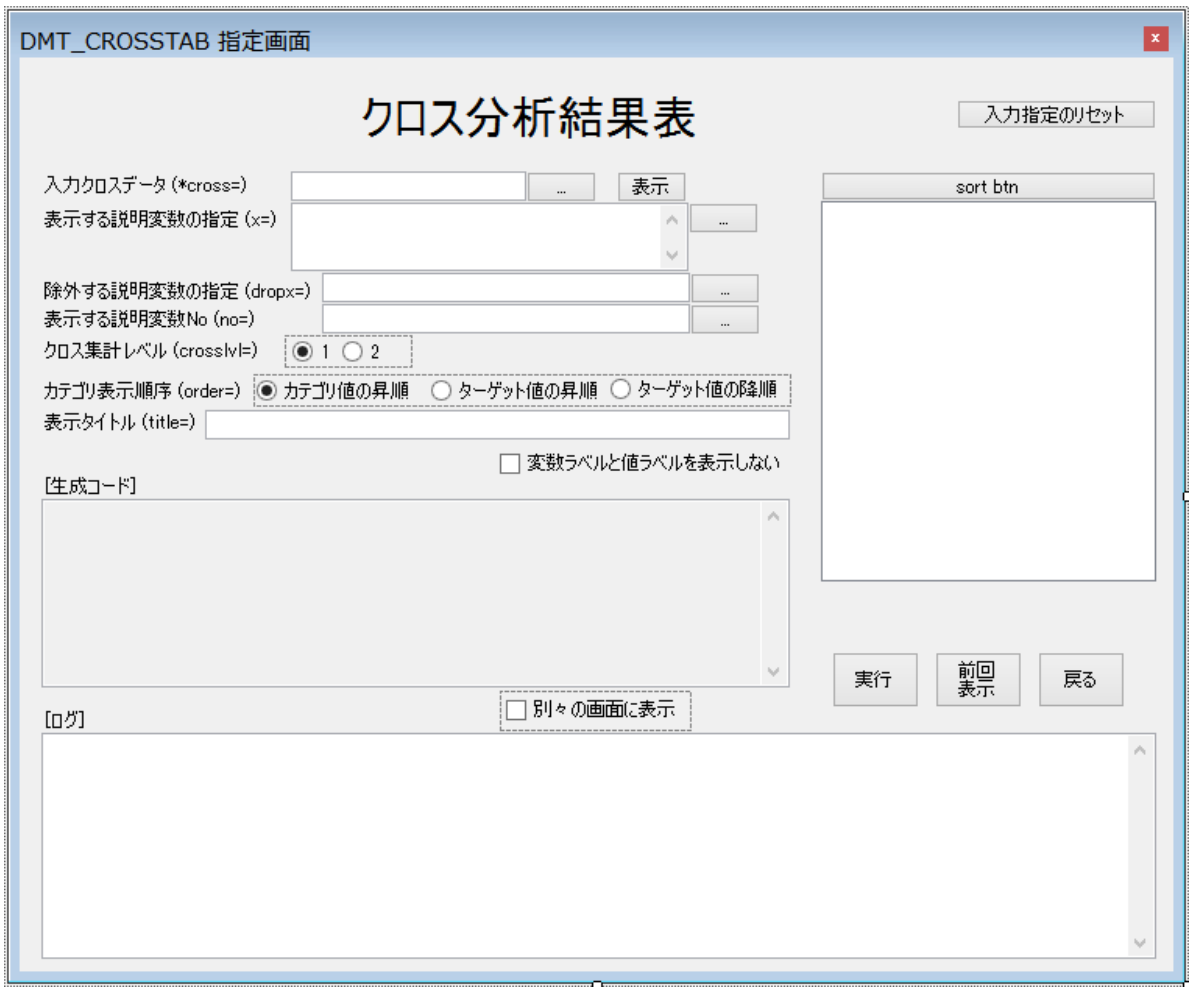
また、以下のユーザ定義フォーマットがWORKライブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
$_AIC $_cat $C&i.V $_DELITM $_DELITM $_item
$_ITMCAT $_VARTYP V&i.C
```

さらに、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
e_name e_type lab&i nobs spc&i typ&i zketa
_speclen _specnum _errormsg _XSEL _XDEL
```

9.2 結果表（dmt_crosstab）



9.2.1 概要

クロス分析結果表（DMT_CROSSTAB） は**クロス分析**（DMT_CROSS） の分析結果出力データセットを入力として、分析結果の全部または指定の一部を画面表示するためのマクロです。AIC値に基づく関連の大きい順に並べた説明変数番号の開始ー終了範囲、もしくは説明変数名を指定することにより表示する範囲を選べます。

9.2.2 指定方法

（コマンド実行モードでの指定）

%dmt_crosstab(help,cross=_cross,fmt=_fmt
,x=,dropx=,no=,crosslvl=1,no0=Y,title=,nolabel=
,order=,pctf=7.2,meanf=best8.,aicf=best8.
,d_label=[D].c_label=[C],dif_label=[D]-[C]
,language=JAPANESE
,outhtml=dmt_crosstab.html,outpath=**)**

（GUI実行モードでの変更点）

・help, fmt=, outhtml=, outpath=パラメータは指定不可。（fmt=, outhtml=, outpath=指定は自動で行われます。）
・no0= はオプション画面で指定します。

（必須パラメータ）

必須パラメータはありません。

（オプションパラメータ）

20個のパラメータはすべて任意指定です。（=の右辺の値はデフォルト値を表しています）
%dmt_crosstab() とパラメータ指定なしでマクロを呼出すと、WORK._cross データセットに含まれる crosslvl=1 の全説明変数について分析結果を画面表示する指定になります。

help ... 指定方法のヘルプメッセージの表示（コマンド実行モードでのみ有効）
クロス集計結果入力データセット名の指定

(cross=_cross)
... DMT_CROSSで作成したクロス分析結果出力データセットを指定します。
集計フォーマット定義入力データセット名の指定
(fmt=_fmt)
... DMT_CROSSで作成した分析結果出力データセットを指定します。（コマンド実行モードでのみ有効。GUI実行モードでは自動的に使用されます。）
表示したい説明変数リストの指定（x=)
... 説明変数を名前で選択表示します(例: x=a b c, x=x1-x4 a--z f_:)
x=説明変数リストから除外する変数リストの指定
(dropx=)
表示する説明変数の番号の指定（no=)
... 説明変数を関連の強さを表す番号により選択表示します(例: no=1:3, no=1 2 5)
表示するクロス集計レベルの指定(crosslvl=1)
全体平均値の表示（no0=Y)
... ターゲットの全体平均出現率もしくはターゲット変数の全体平均値を表の最初の行に表示する（Y）かしない（N）かを指定。
分析結果のカテゴリ表示順序の指定（order=)
... クロス分析結果表における説明変数値の並び順を指定。(order= /A/D) 値の昇順(ブランク), ターゲット出現率または平均値の昇順(A), ターゲット出現率または平均値の降順(D)
（コマンド実行モードでのみ有効。 GUI実行モードでは常にブランク）
変数ラベルと値ラベルを表示しない（nolabel=N)
… 変数ラベルと値ラベルを用いずに変数名、変数値を用いた結果表を作成。
画面出力のタイトルの指定（title=)
.... (%str,%nrstr,%bquote などの関数で囲んで指定すること)
百分率の表示フォーマットの指定（pctf=7.2)
平均値･標準偏差の表示フォーマットの指定
(meanf=best8.)
AIC値の表示フォーマットの指定（aicf=best8.)
差分AIC分析結果表における処理群（DATA）を表す記号（d_label=[D])
差分AIC分析結果表における対照群（Control）を表す記号（c_label=[C])
差分AIC分析結果表における処理群−対照群間の差を表す記号（dif_label=[D]-[C])
言語の選択（language=JAPANESE)
HTML出力ファイル名（outhtml=dmt_crosstab.html)
（コマンド実行モードでのみ有効）
HTMLファイル出力ディレクトリの指定（outpath=)（コマンド実行モードでのみ有効）

9.2.3 パラメータの詳細

表示する説明変数の指定（x=)
表示したい説明変数名を指定します。このパラメータを省略すると、全変数が指定されたものとみなされます。間に1個以上のスペースを入れて、複数の説明変数を指定可能です。また、コロン(:)省略指定とハイフン(-)省略指定と_ALL_特殊指定 も利用可能ですが、,ハイフンハイフン(--)省略指定と_NUMERIC_,_CHARACTER_ 特殊指定は実行結果が目的に合致しなくなるため指定しないでください。（cross=データセットの item1,item2 変数名を参照して変数名を抽出した_data データセットを作成し、ここから有効な名前の説明変数名をチェックしています。）
なお、no=パラメータと同時に指定可能です。指定した場合は、crosslvl=パラメータの条件を満たす中で、x=パラメータとno=パラメータのいずれかの条件を満たす範囲が選択されます。
また、crosslvl=2 指定の場合、x=パラメータに指定された変数が item1, item2 のいずれかに該当するクロス説明変数が抽出されます。

例1：x=age　（説明変数1個を指定）
例2：x=age seibetsu　（説明変数2個を指定）
例3：cross=a,x=abc:　（入力データセットaに含まれるabcで始まる全説明変数を指定）
例4：x=age x1-x5 q: nenshu　（説明変数指定方法の複合例）

除外する説明変数の指定（dropx=)
x=パラメータと共に指定します。x=パラメータに指定した説明変数の中で分析から除外する説明変数を指定します。デフォルトは ブランクです。x=パラメータと同じ指定方法が使えます。
例：
x=_all_,dropx=a_:　（a_で始まる変数およびターゲット変数以外のdata=入力データセットの全変数を説明変数に指定）

表示する説明変数Noの指定（no=)
cross=データセットの 変数 item_no の値に対応して表示する範囲をリスト指定します。デフォルトのブランクはcross=データセットに含まれるすべての番号の説明変数を表示対象とする意味です。no=パラメータの指定方法は、番号を表す数字(例 1)、もしくは､範囲を表す 数字-数字(例 1-5) をブランクで区切って並べて指定します。例えば、以下のような指定が可能です。

例1：no=1　（最初の説明変数1個のみを指定）
例2：no=1 5 7　（説明変数3個を指定）
例3：no=1-7　（1番から7番の連続した範囲を指定）
例4：no=1-7 10　（1番から7番の連続した範囲と10番の8個の説明変数を指定）

存在しない番号が指定された場合はエラーになります。（cross=データセットの item_no変数値を参照して"NO"＋"item_no変数値"を名前とした変数を持つ _item_no データセットを作成し、ここから有効な名前の説明変数名をチェックしています。）なお、x=パラメータと同時に指定可能です。指定した場合は、

両者の条件のいずれかを満たす範囲が選択されます。

クロス集計レベル (crosslvl=1)
説明変数とターゲット値もしくはターゲット変数間の関連性分析結果（crosslvl=1）、2つの説明変数間のクロス説明変数とターゲット値もしくはターゲット変数間の関連性分析結果（crosslvl=2）のいずれかを表示するかを選択します。cross= 入力データセットに crosslvl=2（変数K=2）の分析結果が存在するときは、crosslvl=2 を指定することにより、その部分の分析結果を選択して表示できます。

全体平均値の表示 (no0=Y)
ターゲット値の全体出現率またはターゲット変数の全体平均値の集計結果を表す行を最初の行に表示するか否かを選択します。デフォルトは no0=Y （表示する）です。no0=YまたはNを指定します。

表示タイトル (title=)
画面出力される表にタイトルを指定できます。
タイトルを指定する場合、特殊文字が含まれている場合は、%bquote関数の中に、%と&を文字として認識させたい場合は%nrstr関数の中に記述してください。

言語(language=JAPANESE)
分析実行中のメッセージ出力、結果の表のタイトル、表項目などの表示言語を選択します。ただし、現バージョンでは、日本語か英語の2種類のみ選択可能です。
例：language=ENGLISH

分析結果のカテゴリ表示順序の指定 (order=)
order=パラメータを指定しない場合、各説明変数のカテゴリの並びはカテゴリ値の昇順です。
order=A を指定すると、ターゲット値（出現率、平均値、または処理群と対照群間の出現率または平均値の差）の昇順にカテゴリを並べて表示します。
order=D を指定すると、ターゲット値の降順にカテゴリを並べて表示します。
変数ラベルと値ラベルを表示しない (nolabel=N)
変数ラベルと文字変数値に対する値ラベルのかわりに、本来の変数名、変数値の表示に変わります。元の値を知りたい場合や日本語ラベルを表示したく無い場合に指定します。

9.2.4 コマンド実行モードで有効なパラメータの詳細

fmt=_fmt
DMT_CROSSマクロを実行すると出力されるAIC集計表の説明変数のラベルと文字説明変数カテゴリ値のフォーマット定義データセットを入力として指定します。デフォルトは WORK._fmt です。

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。
例：%dmt_crosstab(help)

9.2.5 HTML 出力

分析結果の図表はhtmlファイルに出力されます。保存先はデフォルトではSASディスプレイマネージャまたはWPSワークベンチの管理下（ワークスペース内の一時保存ファイル）です。outpath=パラメータを指定すると、保存先を変更できます。（必ずフルパス指定します。引用符で囲んでも囲まなくてもかまいません）同時にouthtml=パラメータを指定すると、保存するhtmlファイルに自由に名前を付けることができます。

outhtml=dmt_crosstab.html

分析結果を保存するHTML出力ファイル名を指定します。
例：outhtml=out1.html,

outpath=
HTML図表出力ファイルの保存ディレクトリを指定します。このパラメータを指定しない場合（デフォルト）、HTMLファイルはSASディスプレイマネージャまたはWPSワークベンチの管理下に作成されます。
outpath=指定を行う場合、値は必ずフルパスで指定する必要があります。なお、パス指定全体を引用符で囲んでも囲まなくてもかまいません。
例：outpath='G:¥temp'

9.2.6 実行例

DMT_CROSS マクロの画面からは分析実行後、自動的にクロス分析結果表の表示が行えます。しかし、DMT_CROSSTABマクロを用いると、x=パラメータ、no=パラメータ、crosslvl=パラメータ およびno0=パラメータにより、表示する範囲を選択し、繰り返し画面出力することが可能です。

例１：関連の高い方から 3 個の説明変数のみ表示
```
%dmt_cross(data=samp_data,y=flg,target=1
,x=sei--DM,crosslvl=2,print=N)

%dmt_crosstab(crosslvl=1,no=1-3)
```

DMT_CROSS 分析結果データセット: _cross

				トータル件数	ターゲット件数	ターゲット再現率%	ターゲット出現率%
NO	AIC値	説明変数	値				
0	.	{ANY}	{ALL}	2,000	457	100.00	22.85
1	-423.28	JUKYO 住居	不明	66	25	5.47	37.88
			1 持家(自己所有)	400	15	3.28	3.75
			2 持家(家族所有)	251	9	1.97	3.59
			3 賃貸マンション	285	130	28.45	45.61
			4 借家	390	161	35.23	41.28
			5 アパート	251	95	20.79	37.85
			6 寮	84	4	0.88	4.76
			7 社宅	273	18	3.94	6.59
2	-239.976	GAKUREKI 最終学歴	不明	3	0	0.00	0.00
			1 中学	356	184	40.26	51.69
			2 高校	689	172	37.64	24.96
			3 専門学校	513	48	10.50	9.36
			4 大学	293	25	5.47	8.53
			5 大学院	146	28	6.13	19.18
3	-44.545	KAZOKU_KOSEI 家族構成	不明	48	16	3.50	33.33
			1 独身同居家族あり	697	193	42.23	27.69
			2 独身単身	307	91	19.91	29.64
			3 既婚子供あり	572	86	18.82	15.03
			4 既婚子供なし	349	59	12.91	16.91
			5 独身子供あり	27	12	2.63	44.44

例２：変数を選択して表示
%dmt_crosstab(crosslvl=2,x=DM)

DMT_CROSS 分析結果データセット: _cross

						トータル件数	ターゲット件数	ターゲット再現率%	ターゲット出現率%
NO	AIC値	説明変数1	値1	説明変数2	値2				
0	.	{ANY}	{ALL}	{ANY}	{ALL}	2000.00	457.00	100.00	22.85
11	-478.803	DM プロモーション	0 非実施	JUKYO 住居	不明	44.00	14.00	3.06	31.82
					1 持家(自己所有)	276.00	1.00	0.22	0.36
					2 持家(家族所有)	176.00	1.00	0.22	0.57
					3 賃貸マンション	184.00	80.00	17.51	43.48
					4 借家	269.00	99.00	21.66	36.80
					5 アパート	183.00	65.00	14.22	35.52
					6 寮	61.00	0.00	0.00	0.00
					7 社宅	188.00	7.00	1.53	3.72
			1 実施	JUKYO 住居	不明	22.00	11.00	2.41	50.00
					1 持家(自己所有)	124.00	14.00	3.06	11.29
					2 持家(家族所有)	75.00	8.00	1.75	10.67
					3 賃貸マンション	101.00	50.00	10.94	49.50
					4 借家	121.00	62.00	13.57	51.24
					5 アパート	68.00	30.00	6.56	44.12
					6 寮	23.00	4.00	0.88	17.39
					7 社宅	85.00	11.00	2.41	12.94
15	-290.527	DM プロモーション	0 非実施	GAKUREKI 最終学歴	不明	0.00	0.00	0.00	.
					1 中学	217.00	79.00	17.29	36.41
					2 高校	468.00	114.00	24.95	24.36
					3 専門学校	368.00	29.00	6.35	7.88
					4 大学	220.00	20.00	4.38	9.09
					5 大学院	100.00	25.00	5.47	23.15
			1 実施	GAKUREKI 最終学歴	不明	3.00	0.00	0.00	0.00
					1 中学	139.00	105.00	22.98	75.54
					2 高校	221.00	58.00	12.69	26.24
					3 専門学校	145.00	19.00	4.16	13.10
					4 大学	73.00	5.00	1.09	6.85
					5 大学院	38.00	3.00	0.66	7.89
19	-79.5023	DM プロモーション	0 非実施	SEI 性別	1 男性	947.00	192.00	42.01	20.27
					2 女性	434.00	75.00	16.41	17.28
			1 実施	SEI 性別	1 男性	344.00	64.00	14.00	18.60
					2 女性	275.00	126.00	27.57	45.82

DM がクロス変数の一方に含まれるクロス変数のみを抽出した AIC 分析結果を表示しています。
上記の例は、DM を含む 2 変数のクロス効果（交互作用効果）の中で、購入有無の変動と関連が高い甲後作用効果を抽出しています。なお、DM 送付有無と他の説明変数との交互作用効果はアップリフトモデルで重要です。

例３：分析結果のカテゴリ表示順序の指定

%dmt_cross(data=samp_data,y=flg,target=1,x=seinenshu,crosslvl=2,print=N)

%dmt_crosstab(order=A,crosslvl=1
,title=%nrstr(%dmt_crosstab(order=A,crosslvl=1)))

%dmt_crosstab(order=A,crosslvl=1)

NO	AIC値	説明変数	値	トータル件数	ターゲット件数	ターゲット再現率%	ターゲット出現率%
0	.	{ANY}	{ALL}	2,000	457	100.00	22.85
1	-16.4648	SEI 性別	1 男性	1,291	256	56.02	19.83
		SEI 性別	2 女性	709	201	43.98	28.35
2	-2.28293	NENSHU 年収	.	555	112	24.51	20.18
		NENSHU 年収	499~655	364	76	16.63	20.88
		NENSHU 年収	656~1278	354	75	16.41	21.19
		NENSHU 年収	349~498	364	92	20.13	25.27
		NENSHU 年収	102~348	363	102	22.32	28.10

%dmt_crosstab(order=D,crosslvl=2
,title=%nrstr(%dmt_crosstab(order=D,crosslvl=2)))

%dmt_crosstab(order=D,crosslvl=2)

NO	AIC値	説明変数1	値1	説明変数2	値2	トータル件数	ターゲット件数	ターゲット再現率%	ターゲット出現率%
0	.	{ANY}	{ALL}		{ALL}	2,000	457	100.00	22.85
3	-19.7049	NENSHU 年収	102~348	SEI 性別	2 女性	128	49	10.72	38.28
		NENSHU 年収	349~498	SEI 性別	2 女性	141	48	10.50	34.04
		NENSHU 年収	656~1278	SEI 性別	2 女性	138	40	8.75	28.99
		NENSHU 年収	102~348	SEI 性別	1 男性	235	53	11.60	22.55
		NENSHU 年収	499~655	SEI 性別	2 女性	140	31	6.78	22.14
		NENSHU 年収	.	SEI 性別	2 女性	162	33	7.22	20.37
		NENSHU 年収		SEI 性別	1 男性	393	79	17.29	20.10
		NENSHU 年収	499~655	SEI 性別	1 男性	224	45	9.85	20.09
		NENSHU 年収	349~498	SEI 性別	1 男性	223	44	9.63	19.73
		NENSHU 年収	656~1278	SEI 性別	1 男性	216	35	7.66	16.20

9.2.7 コマンド実行モードでの注意

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

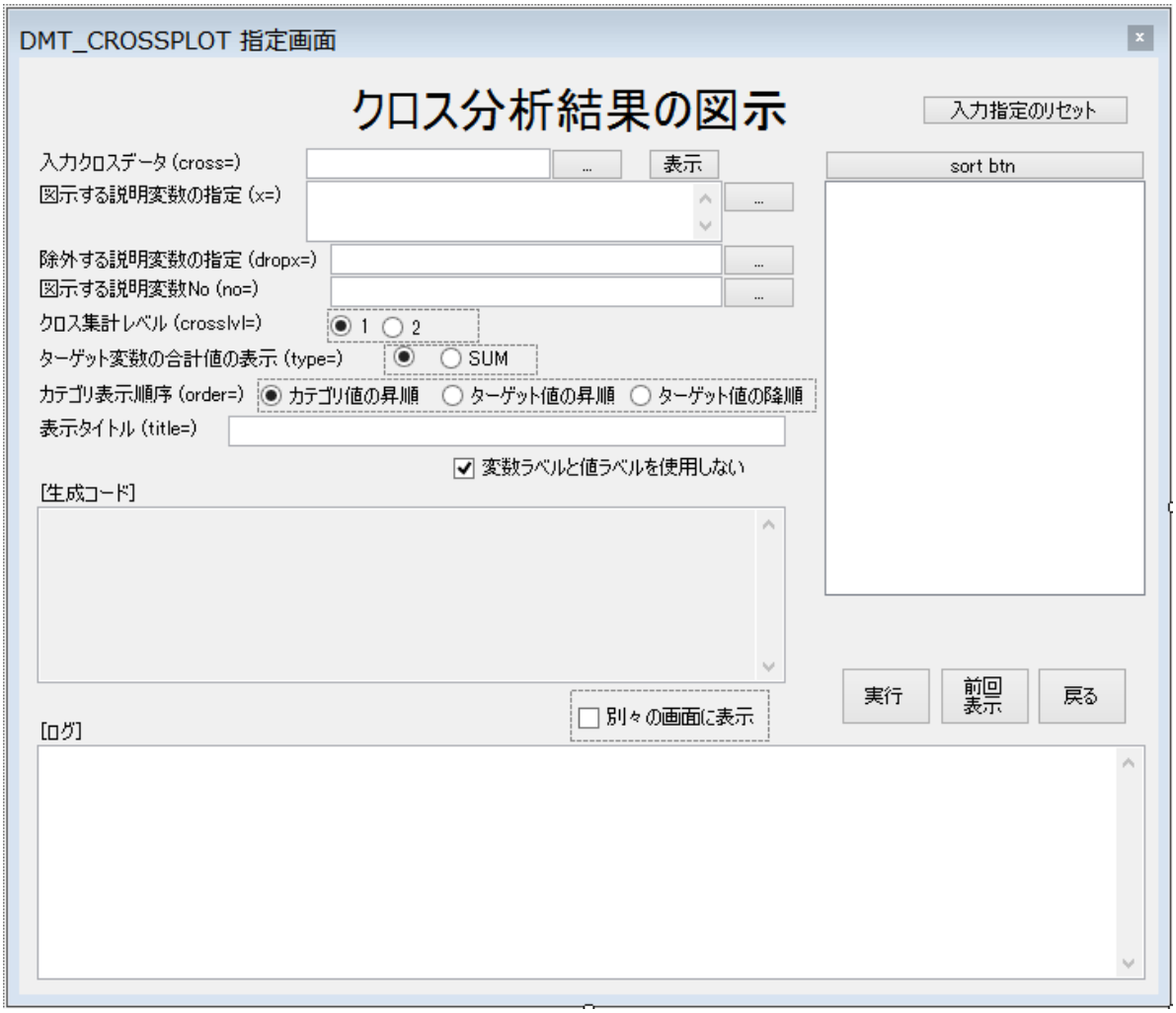
また、以下のユーザ定義フォーマットがWORKライブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
$_AIC $_cat $_item $_ITMCAT $_VARTYP
```

さらに、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
_errormsg
```

9.3 結果図（dmt_crossplot）



9.3.1 概要

クロス分析結果の図示（DMT_CROSSPLOT）は**クロス分析**（DMT_CROSS）で分析した各説明変数カテゴリごとのターゲット出現率、もしくはターゲット変数の分布の違いをグラフ表示するマクロです。AIC値に基づく関連の大きい順に並べた説明変数番号の開始ー終了範囲、もしくは任意の説明変数名を指定することにより図示する範囲を選べます。

9.3.2 指定方法

（コマンド実行モードでの指定）

```
%dmt_crossplot(help,cross=_cross,fmt=_fmt
,x=,dropx=,no=,crosslvl=1,type=,title=,nolabel=
,order=,pctf=7.2,meanf=best8.,aicf=best8.
,d_label=[D].c_label=[C],dif_label=[D]-[C]
,language=JAPANESE,graph_language=ENGLISH
,dev=GIF,outhtml=dmt_crossplot.html,outpath=)
```

（GUI実行モードでの変更点）

・help, fmt=, outhtml=, outpath=パラメータは指定不可。（fmt=, outhtml=, outpath=指定は自動で行われます。）

（必須パラメータ）

必須パラメータはありません。

（オプションパラメータ）

22個のパラメータはすべて任意指定です。（=の右辺の値はデフォルト値を表しています）
なお、%dmt_crossplot() とパラメータ指定なしでマクロを呼出すと、WORK._cross データセットに含まれる crosslvl=1 の全説明変数についてグラフを画面表示する指定になります。

help ... 指定方法のヘルプメッセージの表示（コマンド実行モードでのみ有効）
クロス集計結果入力データセット名の指定

(cross=_cross)
... DMT_CROSSで作成したクロス分析結果出力データセットを指定します。
集計フォーマット定義入力データセット名の指定
(fmt=_fmt)
... DMT_CROSSで作成した分析結果出力データセットを指定します。（コマンド実行モードでのみ有効。GUI実行モードでは自動的に使用されます。）
図示したい説明変数リストの指定（x=)
... 説明変数を名前で選択表示します(例: x=a b c, x=x1-x4 a--z f_:)
x=説明変数リストから除外する変数リストの指定
(dropx=)
図示する説明変数の番号の指定（no=)
... 説明変数を関連の強さを表す番号により選択表示します(例: no=1:3, no=1 2 5)
図示するクロス集計レベルの指定(crosslvl=1)
ターゲット変数の合計値の表示（type=)
... type=SUM 指定で平均値の代わりに合計値を表示します.
分析結果のカテゴリ表示順序の指定（order=)
... クロス分析結果表における説明変数値の並び順を指定。(order= /A/D) 値の昇順(ブランク), ターゲット出現率または平均値の昇順(A), ターゲット出現率または平均値の降順(D)
（コマンド実行モードでのみ有効。 GUI実行モードでは常にブランク）
変数ラベルと値ラベルを表示しない（nolabel=N)
… 変数ラベルと値ラベルを用いずに変数名、変数値を用いた結果表を作成。
画面出力のタイトルの指定（title=)
.... (%str,%nrstr,%bquote などの関数で囲んで指定すること)
百分率の表示フォーマットの指定（pctf=7.2)
平均値・標準偏差の表示フォーマットの指定
(meanf=best8.)
AIC値の表示フォーマットの指定（aicf=best8.)
差分AIC分析結果表における処理群（DATA）を表す記号（d_label=[D])
差分AIC分析結果表における対照群（Control）を表す記号（c_label=[C])
差分AIC分析結果表における処理群−対照群間の差を表す記号（dif_label=[D]-[C])
言語の選択（language=JAPANESE)
グラフ画面表示言語の選択
(graph_language=ENGLISH)
HTML出力ファイル名（outhtml=dmt_crosstab.html)
（コマンド実行モードでのみ有効）
HTMLファイル出力ディレクトリの指定（outpath=)（コマンド実行モードでのみ有効）
グラフデバイスの指定（dev=GIF)

9.3.3 パラメータの詳細

表示する説明変数の指定（x=)
表示したい説明変数名を指定します。このパラメータを省略すると、全変数が指定されたものとみなされます。間に1個以上のスペースを入れて、複数の説明変数を指定可能です。また、コロン(:)省略指定とハイフン(-)省略指定と_ALL_特殊指定 も利用可能ですが、,ハイフンハイフン(--)省略指定と_NUMERIC_,_CHARACTER_ 特殊指定は実行結果が目的に合致しなくなるため指定しないでください。（cross=データセットの item1,item2 変数名を参照して変数名を抽出した_data データセットを作成し、ここから有効な名前の説明変数名をチェックしています。）
なお、no=パラメータと同時に指定可能です。指定した場合は、crosslvl=パラメータの条件を満たす中で、x=パラメータとno=パラメータのいずれかの条件を満たす範囲が選択されます。
また、crosslvl=2 指定の場合、x=パラメータに指定された変数が item1, item2 のいずれかに該当するクロス説明変数が抽出されます。

例1：x=age　（説明変数1個を指定）
例2：x=age seibetsu　（説明変数2個を指定）
例3 : cross=a,x=abc:　（入力データセットaに含まれるabcで始まる全説明変数を指定）
例4：x=age x1-x5 q: nenshu　（説明変数指定方法の複合例）

除外する説明変数の指定（dropx=)
x=パラメータと共に指定します。x=パラメータに指定した説明変数の中で分析から除外する説明変数を指定します。デフォルトは ブランクです。x=パラメータと同じ指定方法が使えます。
例：
x=_all_,dropx=a_:　（a_で始まる変数およびターゲット変数以外のdata=入力データセットの全変数を説明変数に指定）

表示する説明変数Noの指定（no=)
cross=データセットの 変数 item_no の値に対応して表示する範囲をリスト指定します。デフォルトのブランクはcross=データセットに含まれるすべての番号の説明変数を表示対象とする意味です。no=パラメータの指定方法は、番号を表す数字(例 1)、もしくは､範囲を表す 数字-数字(例 1-5) をブランクで区切って並べて指定します。例えば、以下のような指定が可能です。

例1：no=1　（最初の説明変数1個のみを指定）
例2：no=1 5 7　（説明変数3個を指定）
例3：no=1-7　（1番から7番の連続した範囲を指定）
例4：no=1-7 10　（1番から7番の連続した範囲と10番の8個の説明変数を指定）

存在しない番号が指定された場合はエラーになります。（cross=データセットの item_no変数値を参照して"NO"＋"item_no変数値"を名前とした変数を持つ_item_no データセットを作成し、ここから有効な名

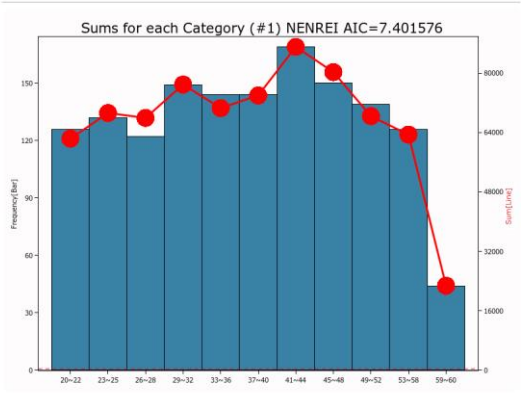
前の説明変数名をチェックしています。）なお、x=パラメータと同時に指定可能です。指定した場合は、両者の条件のいずれかを満たす範囲が選択されます。

クロス集計レベル (crosslvl=1)
説明変数とターゲット値もしくはターゲット変数間の関連性分析結果（crosslvl=1）、2つの説明変数間のクロス説明変数とターゲット値もしくはターゲット変数間の関連性分析結果（crosslvl=2）のいずれかを表示するかを選択します。cross= 入力データセットに crosslvl=2（変数K=2）の分析結果が存在するときは、crosslvl=2 を指定することにより、その部分の分析結果を選択して表示できます。

ターゲット変数の合計値の表示(type=)
数値タイプのターゲット変数の分布を図示するとき、type=SUMを指定すると、平均値の代わりに合計値の折れ線表示に変更されます。
例：
%dmt_cross(data=samp_data,y=nenshu,x=nenrei)
%dmt_crossplot(type=sum,nolabel=Y)

DMT_CROSS 分析結果: 分析データセット: samp_data, ターゲット: nenshu

NO	AIC値	説明変数	値	件数	平均値	標準偏差	最小値	最大値
0	.	(ANY)	(ALL)	1,445	514.0498	202.7175	102	1278
1	7.401576	NENREI 年齢	20~22	126	494.8254	195.3707	108	1070
			23~25	132	524.4924	215.3209	125	1052
			26~28	122	556.6721	245.1042	161	1278
			29~32	149	516.1544	208.4451	166	1245
			33~36	144	489.8958	197.9444	106	1111
			37~40	144	513.7569	200.249	139	1198
			41~44	169	515.0355	194.371	102	1115
			45~48	150	534.9933	192.5019	149	937
			49~52	139	492.0576	181.3424	104	1138
			53~58	126	503.3254	188.1139	102	1217
			59~60	44	517.4773	199.0686	126	861



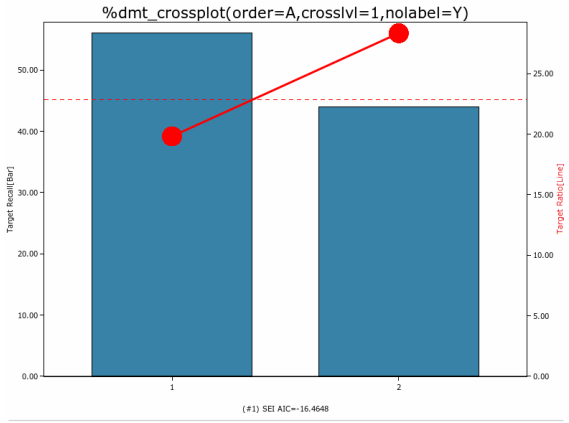
表示タイトル (title=)
画面出力される表にタイトルを指定できます。
タイトルを指定する場合、特殊文字が含まれている場合は、%bquote関数の中に、%と&を文字として認識させたい場合は%nrstr関数の中に記述してください。タイトルを指定すると、アイテム番号、変数名、AIC値がグラフの下部に表示されます。

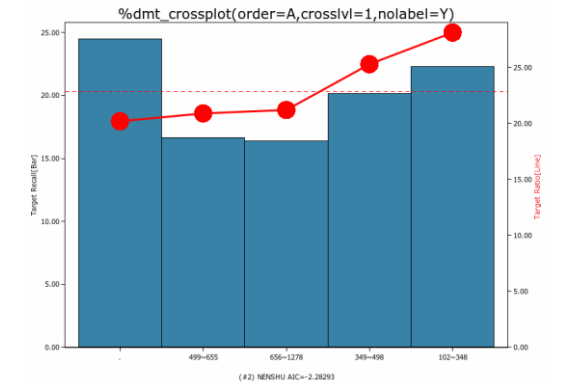
言語(language=JAPANESE)
分析実行中のメッセージ出力、結果の表のタイトル、表項目などの表示言語を選択します。ただし、現バージョンでは、日本語か英語の2種類のみ選択可能です。
例：language=ENGLISH

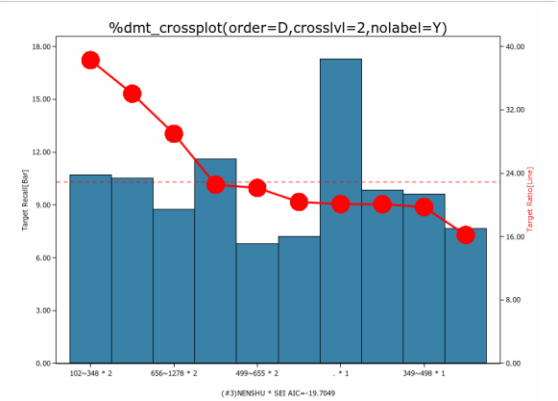
グラフ画面表示言語 (graph_language=ENGLISH)
グラフィック出力画面に表示する既定のタイトルや軸ラベル等に表示する言語を指定します。graph_language=ENGLISH が既定です。※ 現行 WPS ではグラフ上には日本語が表示できませんので、デフォルトの graph_language=ENGLISH を変更しないでください。

分析結果のカテゴリ表示順序の指定 (order=)
order=パラメータを指定しない場合、各説明変数のカテゴリの並びはカテゴリ値の昇順です。
order=A を指定すると、ターゲット値（出現率、平均値、または処理群と対照群間の出現率または平均値の差）の昇順にカテゴリを並べて表示します。
order=D を指定すると、ターゲット値の降順にカテゴリを並べて表示します。

例 ： %dmt_cross(data=samp_data,y=flg,target=1,x=sei nenshu,crosslvl=2,print=N)
%dmt_crossplot(order=A,crosslvl=1,nolabel=Y,title=%nrstr(%dmt_crossplot(order=A,crosslvl=1,nolabel=Y)))
%dmt_crossplot(order=D,crosslvl=2,nolabel=Y,title=%nrstr(%dmt_crossplot(order=D,crosslvl=2,nolabel=Y)))







図の赤色の折れ線がカテゴリごとのターゲット出現率を表します。（赤色の水平な点線は全体平均）

変数ラベルと値ラベルを表示しない（nolabel=N)
変数ラベルと文字変数値に対する値ラベルのかわりに、本来の変数名、変数値の表示に変わります。元の値を知りたい場合や日本語ラベルを表示したく無い場合に指定します。

9.3.4 コマンド実行モードで有効なパラメータの詳細

fmt=_fmt
DMT_CROSSマクロを実行すると出力されるAIC集計表の説明変数のラベルと文字説明変数カテゴリ値のフォーマット定義データセットを入力として指定します。デフォルトは WORK._fmt です。

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。
例：%dmt_cross(help)

std_mod_min_n=9
処理群と対照群間のターゲットの差を分析する場合に、データ件数の少ないターゲット出現率や平均値の標準偏差を修正する基準を与えるパラメータです。そもそも施策実施群の顧客属性と施策非実施群の顧客属性はアンバランスとなることが多いと考えられます。そのため、同一説明変数カテゴリに該当するデータ件数が、処理群と対照群の間で非常にアンバランスとなる場合が起こりえます。そのとき、データ件数が少ない群の方のカテゴリではターゲット出現率や平均値はバラツキ（標準偏差）が大きくなると考えられますが、計算上の標準偏差は0または0に近い不自然な値が得られる場合があります。このような事態を避けるため、std_mod_min_n=パラメータは、指定の値以下のデータ件数から計算されるカテゴリ内のターゲット出現率または目的変数の平均値の標準偏差の計算値が全データの標準偏差より小さい場合に全データの標準偏差に置き換えるよう指示します。

9.3.5 HTML 出力

分析結果の図表はhtmlファイルに出力されます。保存先はデフォルトではSASディスプレイマネージャまたはWPSワークベンチの管理下（ワークスペース内の一時保存ファイル）です。outpath=パラメータを指定すると、保存先を変更できます。（必ずフルパス指定します。引用符で囲んでも囲まなくてもかまいません）同時にouthtml=パラメータを指定すると、保存するhtmlファイルに自由に名前を付けることができます。

outhtml=dmt_crossplot.html
分析結果を保存するHTML出力ファイル名を指定します。
例：outhtml=out1.html,

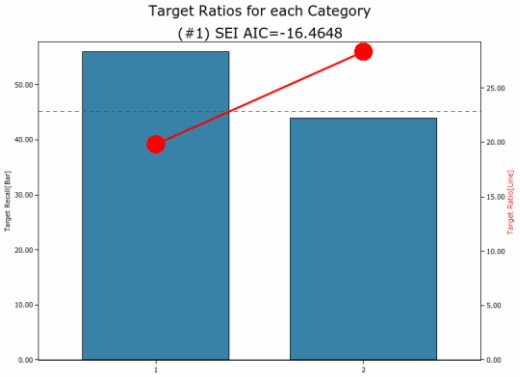
outpath=
HTML図表出力ファイルの保存ディレクトリを指定します。このパラメータを指定しない場合（デフォルト）、HTMLファイルはSASディスプレイマネージャまたはWPSワークベンチの管理下に作成されます。outpath=指定を行う場合、値は必ずフルパスで指定する必要があります。なお、パス指定全体を引用符で囲んでも囲まなくてもかまいません。
例：outpath='G:¥temp'

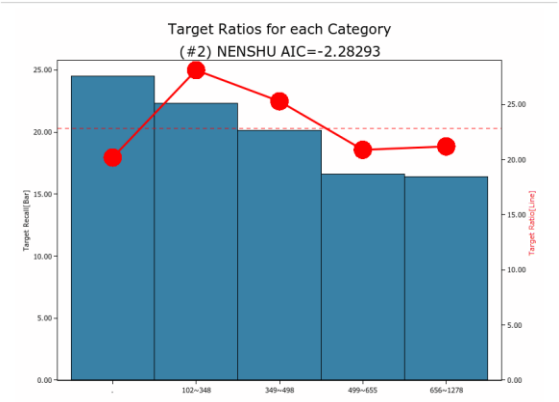
9.3.6 実行例

(1)target=パラメータを指定し、ターゲット出現率の分布との関連を分析した場合

ターゲットの出現率と各説明変数の統計的関連性をAIC 値で評価した分析結果図を出力します。

%dmt_cross(data=samp_data,y=flg,target=1,x=seinenshu,crosslvl=2,print=N)
%dmt_crossplot(nolabel=Y)

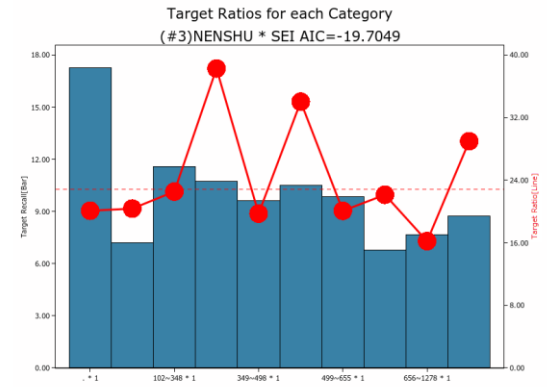




グラフのタイトルには、分析結果の番号(#)、説明変数名とラベル、AIC 値が表示されます。横軸は各説明変数のカテゴリ、縦軸はカテゴリに含まれるターゲット値の再現率（縦棒グラフ表示）と出現率（丸印の折れ線グラフ表示）を表しています。点線の水平線はターゲット出現率の全体平均値を表します。
標準では crosslvl=1 の条件に合致する分析結果のみを表示します。
なお、nolabel=Y オプションは変数ラベルやデータラベルに付けられた日本語の表示を行わないようにするためです。

crosslvl=2 の分析結果を図示するには、crosslvl=2 パラメータを指定します。

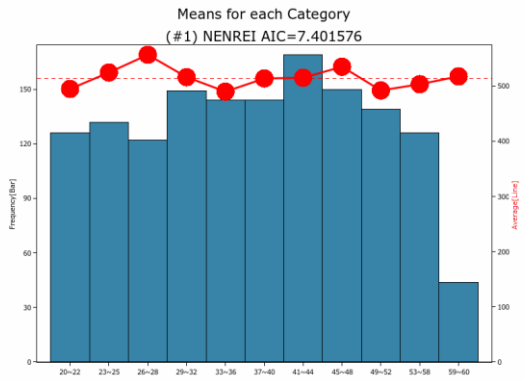
%dmt_crossplot(crosslvl=2,nolabel=Y)



グラフのタイトルには分析結果の番号(#)、説明変数名 1 とそのラベル、"＊" 記号、説明変数名 2 とそのラベル、AIC 値が表示されます。（変数ラベルは nolabel=Y 指定により表示されません）
横軸には説明変数 1 のカテゴリ、"＊" 記号、説明変数 2 のカテゴリが表示されます。縦軸はカテゴリに含まれるターゲット値の再現率（縦棒グラフ表示）と出現率（丸印の折れ線グラフ表示）、そして全体平均出現率（水平な点線）が表示されています。

(2)target=パラメータを指定せず、ターゲット変数の分布との関連を分析した場合

例 ： %dmt_cross(data=samp_data,y=nenshu,x=nenrei,print=N)
%dmt_crossplot(nolabel=Y)

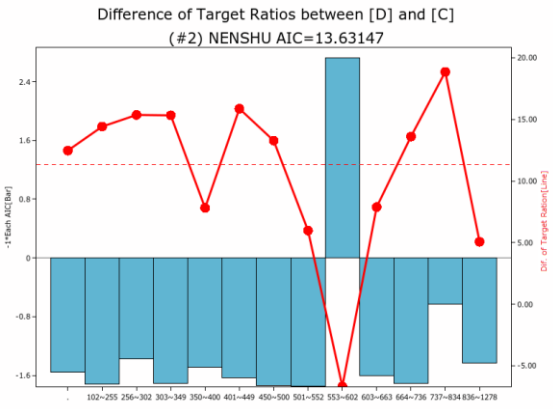


グラフのタイトルは分析結果の番号(#)、説明変数名とそのラベル、AIC 値が表示されます。
横軸は説明変数カテゴリ値とフォーマット値が表示されます。縦軸はカテゴリ該当件数（棒グラフ表示）とターゲット変数の平均値（丸印のドットと折れ線表示）が表示されます。なお、水平な点線は目的変数の全体平均値を表します。

type=SUM パラメータを指定すると、平均値ではなく、カテゴリ別合計値の表示に変わります。

(3)control=パラメータとtarget=パラメータを指定し、処理群と対照群間のターゲット出現率の差を分析した場合

例 ： %dmt_cross(data=samp_data(where=(DM="1"))
,control=samp_data(where=(DM="0"))
,y=flg,target=1,x=sei nenshu,print=N)
%dmt_crossplot(x=NENSHU,nolabel=Y)



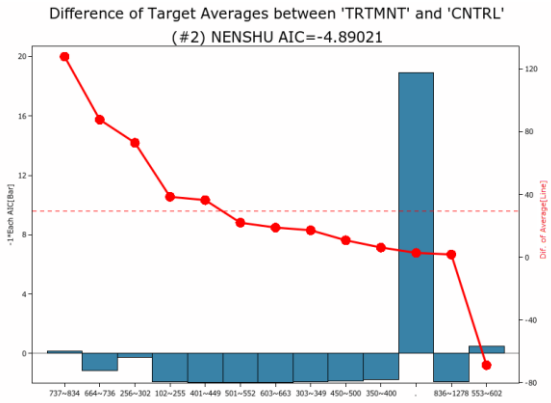
グラフのタイトルは分析結果の番号(#)、説明変数名とそのラベル、AIC 値が表示されます。
横軸は説明変数カテゴリ値とフォーマット値が表示されます。縦軸は個別 AIC 値の符号を逆にした値（棒グラフ表示）と処理群と対照群のターゲット出現率

の差（丸印のドットと折れ線表示）が表示されます。
なお、水平な点線は処理群と対照群のターゲット出現率の全体平均値の差を表します。

棒グラフの値は個別 AIC 値の符号を逆転させた値です。したがって、棒の値が 0 より大きいカテゴリは処理群と対照群の出現率の差が有意であることを意味し、上方向に高いほど有意とみなされます。上図の場合、年収が 553-602 の範囲のカテゴリで処理群と対照群の出現率の差は処理群の方が 5%ポイントほど低くなっていますが、この差は有意です。それ以外のカテゴリでは処理群と対照群の出現率の差は有意ではありません。変数全体の AIC 値は 13.6 となっており、変数 NENSHU のカテゴリ間の処理群と対照群の出現率の差のばらつきは、全体として有意では無いという結論です。

(4) control=パラメータを指定し、target=パラメータを指定せず、処理群と対照群間のターゲット変数の平均値の差を分析した場合

例：%dmt_cross(data=samp_data(where=(DM="1"))
,control=samp_data(where=(DM="0"))
,y=kingaku,x=sei nenshu,print=N)
%dmt_crossplot(x=NENSHU,order=D,d_label='TRTMNT',c_label='CNTRL',nolabel=Y)



グラフのタイトルは分析結果の番号(#)、説明変数名とそのラベル、AIC 値が表示されます。
横軸は説明変数カテゴリ値とフォーマット値が表示されます。縦軸は個別 AIC 値の符号を逆にした値（棒グラフ表示）と処理群と対照群のターゲット平均値の差（丸印のドットと折れ線表示）が表示されます。
なお、水平な点線は処理群と対照群の各全体平均値の差を表します。
order=D オプションを指定しているため、ターゲット平均値の差の降順に横軸のカテゴリが並べられています。また、d_label=,c_label=オプションの指定によって処理群、対照群それぞれを意味する表示テキストがデフォルトから変更されています。

棒グラフの値は個別 AIC 値の符号を逆転させた値です。したがって、棒の値が 0 より大きいカテゴリは処理群と対照群の平均値の差が有意であることを意味し、上方向に高いほど有意とみなされます。上図の場合、年収が 737-834、欠損、553-602 の 3 つのカテゴリで処理群と対照群の平均値の差は有意となっています。変数全体の AIC 値は-4.89 となっており、変数 NENSHU のカテゴリ間の処理群と対照群の平均値の差のばらつきは、全体として有意という結論です。

9.3.7 コマンド実行モードでの注意

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

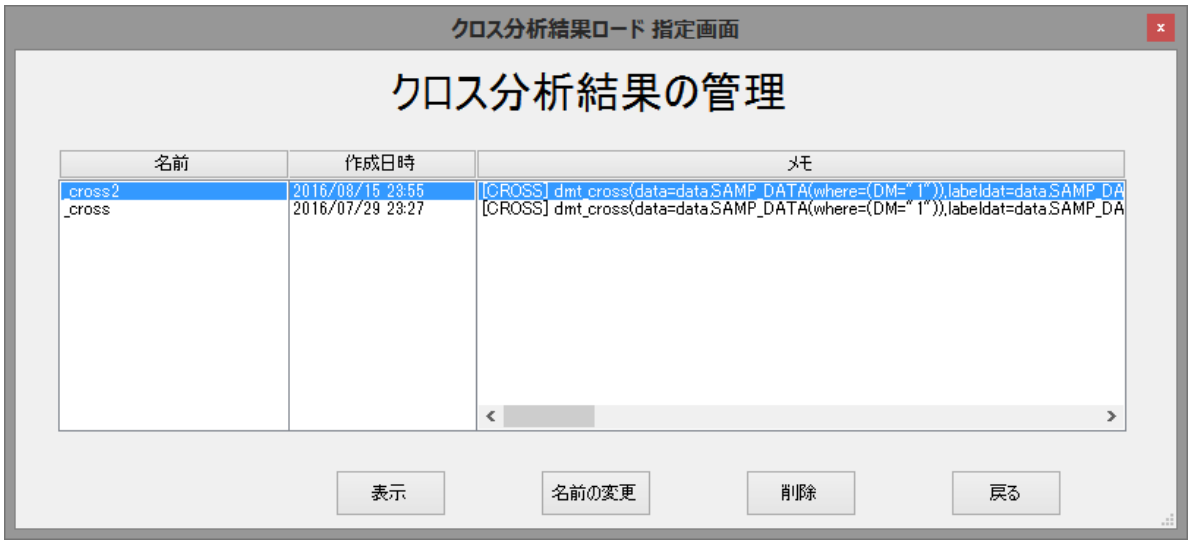
また、以下のユーザ定義フォーマットがWORKライブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
$_AIC $_cat _cat $_item $_ITEMC $_ITMCAT
$_VARTYP
```

さらに、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
_errormsg
```

9.4 結果管理



9.4.1 概要

「クロス分析」画面で作成したクロス分析結果データセットを操作（表示・名前の変更・削除）します。
この機能はマクロモジュールには含まれていません。GUI実行モードでのみ指定可能です。

メモ欄の最初の鍵カッコは以下の画面で作成されたことを表します。

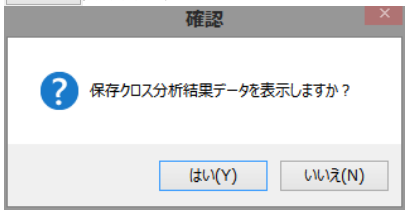
[CROSS] ... クロス分析
続いてデータを作成したときに実行したプログラムが記述されています。

9.4.2 操作方法

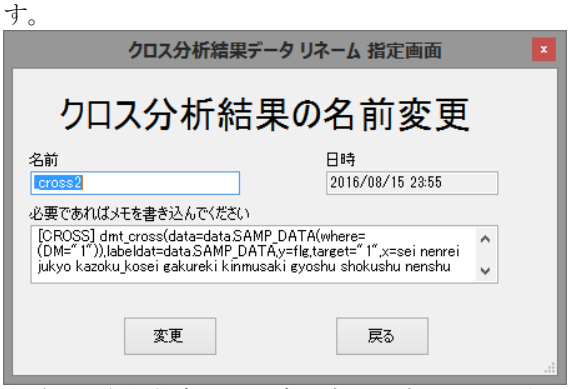
名前,
作成日時,
メモ
リストの上にあるバーをクリックすると、データセットリストを各項目の昇順・または降順で並べ替えることができます。

操作したいクロス分析結果データ名をクリックすると、操作ボタンが表示されますので、表示・名前の変更・削除の操作を行います。

表示 分析結果データの内容を表示します。

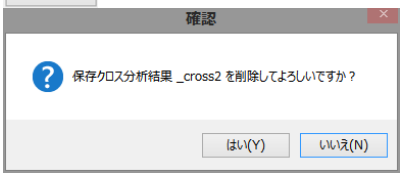


名前の変更 データの名前とメモ内容を確認・変更します。



名前は半角英数字で32文字以内に設定してください。
（先頭はアルファベットまたは_(アンダーバー)）
なお、名前の変更は、元の名前を参照している他の項目（モデル作成画面の入力パラメータ値など）とは自動連動しません。そのため、再指定が必要になるなどの影響があります。

削除 データを削除します。



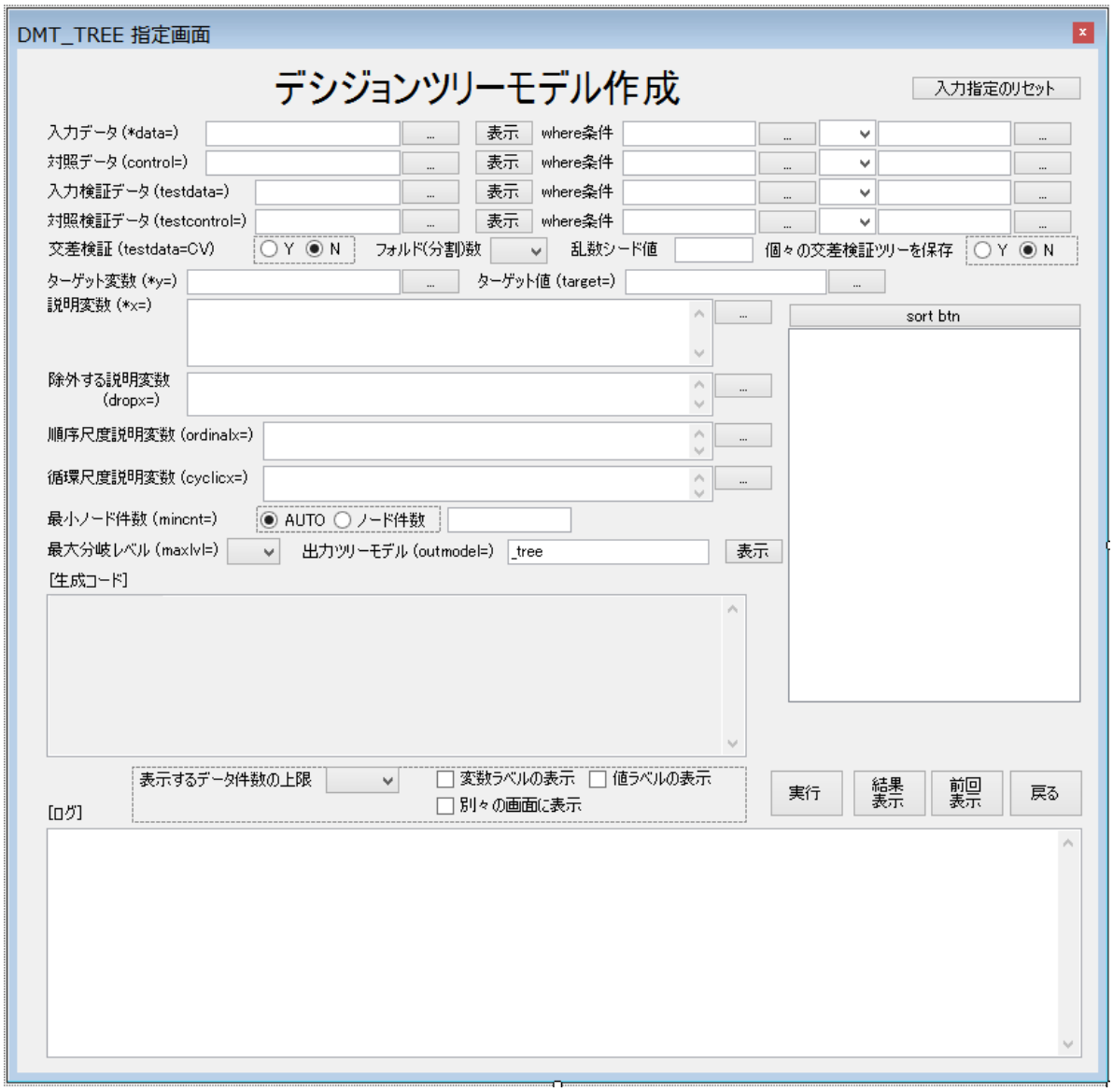
削除すると、元に戻せません。

(TIPS) 多数のデータセットをまとめて削除したい場合は、**「設定画面」**の**「分析ディレクトリ」**の下の**「データセットディレクトリ」「表示」**ボタンを押し、起動するWindowsエクスプローラで行うと便利です。削除したいデータセット名が書かれたディレクトリをすべて同時選択してから削除します。

10. 分析画面 ③モデル作成表示

デシジョンツリーモデルを作成し、モデルの内容を表示します。

10.1 モデル作成（dmt_tree）



10.1.1 概要

デシジョンツリーモデル作成（DMT_TEEE）はデシジョンツリー（決定木）モデルを作成するプログラムです。以下の特徴があります。

（１）　分類木モデル、回帰木モデルの両方に対応
（２）　アップリフトモデルの作成
（３）　交差検証法によるモデル検証
（４）　最大2000の説明変数の指定が可能
（５）　個々の説明変数に尺度指定（名義・順序・循環）が可能
（６）　AIC基準による説明変数選択
（７）　動的なノード最小必要件数の指定が可能
（８）　モデルの表示、検証、新規データへのモデルあてはめ、収益計算等さまざまな機能が利用可能
（９）　SAS言語で開発されていること

（１）分類木もしくは回帰木の作成
DMT_TEEEはターゲット出現率を予測する分類木モデル、ターゲット変数の値を予測する回帰木モデル、

いずれにも対応しています。

（２）アップリフトモデル
本バージョンから処理群（data=入力データ）と対照群（control=入力データ）間のターゲット出現率またはターゲット平均値の差を予測するツリーモデル（アップリフトモデル）の作成が可能です。

（３）交差検証法によるモデル検証
本バージョンからモデル作成データと別に用意した検証用データを用いたモデル検証の他、モデル作成データのみを用いた交差検証法によるモデル検証も可能になりました。(GUI実行モードではtestdata=CV指定、コマンド実行モードでは DMT_CVTREEマクロの実行)

交差検証の流れは以下のとおりです。

（ステップ1：モデル作成）
まず、最初に分析データ全部を使ったモデル（outmodel=_TREE）を作成します。

（ステップ2：交差検証）
次に分析データセットをいくつかの同じサンプルサイズのデータセットにランダムに分割します。分割数を5とすると、分割されたデータには1番から5番までの番号を順につけておきます。
この中から、モデル作成用と検証用データを順に入れ替えながら、5個のモデルの作成と検証データへのモデル適用を行います。最初は、検証用に1番、モデル作成用に2番から5番までを併合したデータを使い、2回目は、検証用に2番、モデル作成用に1番と3番から5番までの併合データ、... 、最後は、検証用に5番、モデル作成用に1番から4番までの併合データを使います。

これらのモデルは、ステップ1で作成したモデルと同じ方法（指定する説明変数、使用する分割基準、ノード終端条件など）で作成します。結果として、相互に微妙に異なる5個の個々の交差検証モデル（_TREE_CV1～_TREE_CV5）と、元の分析データの各オブザベーションに対して、いずれかの個々の交差検証モデルの予測値が付与された検証結果データ（_TREE_CVSC）が得られます。

（ステップ3：予測結果の整理）
検証結果データ（_TREE_CVSC）の交差検証モデル予測値は、集計することにより、元のモデル（_TREE）の予測誤差の評価に用いることができます。
DMT_CVTREEでは、交差検証モデル予測値を元のモデル（_TREE）のノード（中間および終端）別の平均値に集計する方法で、検証データに対するモデル形式データセット（_TREE_CV）として整理しています。テストデータが与えられた場合のモデル評価方法がすべて適用できます。

以上のように、DMT_CVTREEは交差検証結果として、個々の交差検証モデル(_TREE_CV1～_TREE_CV5)、交差検証予測値（_TREE_CVSC)、検証結果モデル形式データセット（_TREE_CV）3種類のデータセットを出力します。

特に、検証結果モデル形式データセット（_TREE_CV）は、あたかも、元のツリーモデルの分割規則を、他の新しいテストデータに適用したときに得られるものと同じ形式ですので、DMT_GAINCHART, DMT_COMPAREPLOT, DMT_UPLIFTCHARTなどのモデル評価方法がすべて適用できます。

なお、ステップ2の分析データセットの分割数は、本アプリケーションではデフォルトを5（2～20の範囲で有効）に設定しています。

（４）最大 2000 の説明変数の指定が可能
指定できる説明変数の数は最大2000までとしています。ただし、コンピュータ資源の制約からそれ以下の説明変数しか用いることができない場合もあり得ます。説明変数の指定には-（ハイフン)、--（ハイフンハイフン)、:（コロン)、_ALL_（全変数)、_NUMERIC_（全数値タイプ変数)、_CHARACTER_（全文字タイプ変数）の各省略形式およびそれらの混合指定が可能です。

（５）個々の説明変数に尺度指定（名義・順序・循環）が可能
文字タイプ説明変数は名義尺度（デフォルト)、順序尺度（ordinalx=パラメータに指定)、循環尺度（cyclicx=パラメータに指定）のいずれかに設定可能です。また、数値タイプ説明変数は順序尺度（ordinalx=パラメータに指定)、循環尺度（cyclicx=パラメータに指定）のいずれかに設定可能です。数値タイプ説明変数のデフォルト尺度は、splitpts=パラメータの値によって変更できます。

なお、順序尺度とはカテゴリの値の間に順序関係があるとみなす尺度（例 変数："A 優"、"B 良"、"C 可"）であり、循環尺度とは順序関係にあるカテゴリの最初と最後のカテゴリが隣り合っているとみなす尺度（例："A 朝"、"B 昼"、"C 夜"）です。名義尺度はカテゴリ間に何ら順序関係が無いとする尺度（例："東京"、"大阪"、"名古屋"、"福岡"）です。

尺度設定と関係するパラメータは以下のとおりです。個々の文字変数は名義、順序、循環いずれかの尺度指定、数値変数は順序もしくは循環のいずれかの尺度指定が可能です。

説明変数タイプ	splitpts= パラメータの値	パラメータに変数名を指定		設定される尺度
		ordinalx=	cyclicx=	
文字	無関係	×	×	名義(nominal)
		○	×	順序(ordinal)
		×	○	循環(cyclic)
数値	1	×	×	順序(ordinal)
		×	○	循環(cyclic)
	×（または2）	×	×	循環(cyclic)
		○	×	順序(ordinal)
			注：×は無指定を意味します	

（６）AIC基準による説明変数選択
各ノードの分岐に用いる説明変数の選択基準としてAIC基準を採用し、その後カテゴリ併合を行っています。（カテゴリ併合を含めて変数選択を一度に検索するより効率的）ただし、ノード最小件数条件を満たすカテゴリ併合法が見つからない場合は、次に説明力が高い説明変数を順次探索しています。
採用された説明変数のカテゴリ値の子ノードへの振り分け方法はエントロピー基準（分類木）、群内平方和最小基準（回帰木）、AIC基準（アップリフトモデル）によるカテゴリ併合法を採用しています。

（７）動的なノード最小必要件数の指定が可能
分類木モデルの場合、各ノードにおけるターゲット予測出現率pの統計的誤差（真の出現率からの観測出現率の誤差）はそのノードに含まれる件数Nの平方根に反比例します。同様に、回帰木モデルの場合も、各ノードにおけるターゲット予測値yの統計的誤差（真の値と観測値の誤差）もそのノードに含まれる件数Nの平方根に反比例します。

この性質を利用して、すべてのノードが予測値の大きさに比例した許容範囲内の誤差に収まることが期待できるモデルを構築する機能を実現しています（mincnt=AUTO指定（デフォルト））。ターゲット出現率pまたはターゲット予測値yのノードがpもしくはyの一定倍数（err_rate=パラメータ）以内の上下許容誤差範囲に収まるだけの件数Nを持っているかどうかをチェックし、これを満たす最適のノード分岐説明変数カテゴリを探索します。この機能を用いてモデル構築用データセットへ過剰適合したモデルの作成を自動的に防ぐことが期待できます。
なお、アップリフトモデルでは、処理群、対照群の両方においてこの条件を満たすノード分岐を行うよう制御しています。

（８）モデルの表示、検証、新規データへのモデルあてはめ、収益計算等さまざまな機能が利用可能
DMT_TREE 実行結果は作成したモデルをデータセットに出力します。このモデルデータセットを入力として、モデルの表示（DMT_TREETAB, DMT_NODETAB）、モデルの精度検証（DMT_GAINCHART, DMT_COMPAREPLOT, DMT_CORRECTTAB）、新規データへのモデルあてはめ（DMT_TREESCORE）、収益計算（DMT_GAINCHART）などのDMTデシジョンツリーアプリケーションに備わっているさまざまな機能を続けて実行することができます。

（９）SAS言語で開発されていること
本アプリケーションは全部がSAS言語で開発されており、コマンド実行モードでは、各マクロモジュールは、SASプログラムによるユーザアプリケーションの中に自由に組み込むことができます。分析結果の大部分はデータセット出力されますので、現バージョンのDMTデシジョンツリーアプリケーションがサポートしていないレポート表示やグラフ表示なども、ユーザプログラミングにより作成することが可能です。

10.1.2 指定方法

（コマンド実行モードでの指定）

```
%dmt_tree(help
,data=,control=,y=,target=,x=
,dropx=&y,ordinalx=,cyclicx=,outmodel=_tree
,mincnt=AUTO,err_rate=0.1,maxlvl=5,lastcatm=N
,splitpts=2,nomergen=STURGES,maxcatn=1000
,precat=Y,std_mod_min_n=9,keep_node_data=N
,node_data_prefix=
,language=JAPANESE)
```

コマンド実行モードで交差検証機能つきデシジョンツリーモデル作成を行う場合は、dmt_treeマクロではなく、以下のdmt_cvtreeマクロを指定します。

```
%dmt_cvtree(help,fold=5,seed=1
,data=,control=,y=,target=,x=
,dropx=&y,ordinalx=,cyclicx=,outmodel=_tree
,mincnt=AUTO,err_rate=0.1,maxlvl=5,lastcatm=N
,splitpts=2,nomergen=STURGES,maxcatn=1000
,precat=Y,std_mod_min_n=9,keep_node_data=N
,node_data_prefix=
,language=JAPANESE)
```

dmt_cvtreeマクロはfold=パラメータとseed=パラメータが追加指定できる点だけがdmt_treeマクロとの相違点です。

（GUI実行モードでの変更点）

・helpパラメータは指定不可。
・以下のアイテムが入力可能。
　入力検証データ(testdata=)
・testdata=CV に Y を指定することにより交差検証モデルが作成可能。
・実行結果の表示が可能
（ツリー分岐図、ゲインチャート（分類木の場合のみ）、アップリフトチャート（アップリフトモデルのみ）、比較プロットの表示を選択できます）
・seed=, err_rate=, lastcatm=, splitpts=, nomergen=, maxcatn=, precat= **はオプション画面で指定します。**

（必須パラメータ）

以下の5個のパラメータの内、data=, y=, x= の3個は

常に必須指定です。control=パラメータは、施策実施効果を分析するアップリフトモデルを作成する場合に対照群データの指定に用います。また、target=パラメータは、ターゲット値の出現率（または処理群と対照群間のターゲット出現率の差）の大小を分岐基準とするツリーモデルを作成する場合に指定しなければなりません。

入力データの指定 (data=)
入力対照データの指定 (control=)
ターゲット変数の指定 (y=)
　... (単一変数名のみ指定可)
説明変数リストの指定 (x=)
　... (例: a b c x1-x4 a--z f_:)
ターゲット値の指定 (target=)
　...（ターゲット／非ターゲットの度数分割表におけるAIC計算を行う場合にのみ必須）

（オプションパラメータ）

以下の18個のパラメータは任意指定です。(=の右辺の値はデフォルト値を表しています)

help ... 指定方法のヘルプメッセージの表示.（コマンド実行モードでのみ有効）
出力モデルデータセット名の指定 (outmodel=_tree)
ノード最小必要件数の指定 (mincnt=AUTO)
　... AUTOまたは正の整数を指定
推計値の標準誤差に対する許容誤差率
　(err_rate=0.1) ... 0超1未満の値を指定
最大分岐レベル (maxlvl=5)
　... 1～20 までの正の整数を指定
最終カテゴリ併合 (lastcatm=N)
　... 数値説明変数カテゴライズにおいて、最後のカテゴラリ件数が少ない場合１つ前のカテゴリに併合するか否かの選択(Y/N).
非併合数値タイプ説明変数最大カテゴリ数
　(nomergen=STURGES)
　... 指定の値以下の値の種類数を持つ数値説明変数の個々の値を分析に用いる.(デフォルトはスタージェスの式の値)
除外する説明変数 (dropx=&y)
　... x=説明変数リストから除外する変数リストの指定
数値タイプ説明変数の最大しきい値数 (splitpts=2)
　... 1または2(デフォルト)。
順序尺度説明変数の指定 (ordinalx=)
循環尺度説明変数の指定 (cyclicx=)
分析に用いる文字タイプ説明変数の最大カテゴリ数(maxcatn=1000)
　... 2～5000の範囲で指定可能
あらかじめ数値変数をカテゴライズ(precat=Y)
　… 分析開始時にあらかじめ数値タイプ説明変数をまとめてカテゴライズを行うか否かを選択
言語の選択 (language=JAPANESE)
交差検証実行時のデータ分割数 (fold=5)
交差検証実行時のデータ分割に用いる乱数シードの指定 (seed=1)
全体の標準偏差を用いる最小カテゴリ件数の指定
　(std_mod_min_n=9)（コマンド実行モードでのみ有効）
WORKライブラリにノードデータセットを残す
　(keep_node_data=N)（コマンド実行モードでのみ有効）
WORKライブラリに残すノードデータセットの接頭辞をつける (node_data_prefix=)（コマンド実行モードでのみ有効）

（以下はGUI実行モードでのみ指定可能なオプション）

入力検証データ (testdata=)
　… モデル検証用データを指定します。
　testdata=CV に Y を指定すると、交差検証を行います。

10.1.3 パラメータの詳細

入力データ (data=)
入力データセット名を指定します。このパラメータは省略できません。control=パラメータも指定する場合は、data=パラメータには処理群（施策実施群）を表す入力データセットを指定します。
例：data=a, data=a(where=(DM="1"))

入力対照データ (control=)
処理群と対照群間の応答差を分析するときに、対照群を表す入力データセットを指定します。
例：control=b, contol=a(where=(DM="0"))

ターゲット変数 (y=)
ターゲット変数名を指定します。 このパラメータは省略できません。
例：y=flag,　y=sales_amount

ターゲット値 (target=)
ターゲット値を指定します。このパラメータは文字タイプターゲット変数の特定の値、もしくは数値タイプターゲット変数の特定の値もしくは範囲をターゲット値とみなして、その出現率（または実施群と非実施群間の出現率の差）を分析したい場合は省略できません。（数値タイプターゲット変数の値そのものの分布の違いを分析したい場合は指定してはいけません。）

ターゲット変数が文字タイプの場合は1種類の値を指定します。特殊な文字（+,-など）を含まない限り引用符で囲む必要はありません。ターゲット変数が数値タイプの場合は1種類の値、もしくはあるしきい値を境とした「以上」、「以下」、「超」、「未満」のいずれかの範囲を指定可能です。数値変数タイプで範囲を指定する場合は引用符で囲んではいけません。

例1：y=flag,target=A　（ターゲット変数が文字タイプ変数で、その値"A"をターゲットに指定する場合）
例2：y=sales,target=1000　（ターゲット変数が数値タイプで、その値1000をターゲットに指定する場合）
例3：y=sales,target=>1000　（ターゲット変数が数値タイプで、その値1000超をターゲットに指定する場合）
例4：y=sales,target=>=1000　（ターゲット変数が数値タイプで、その値1000以上をターゲットに指定する場合。　target==>1000と指定してもかまいません。）
例5：y=sales,target=<1000　（ターゲット変数が数値タイプで、その値1000未満をターゲットに指定する場合）
例6：y=sales,target=<=1000　（ターゲット変数が数値タイプで、その値1000以下をターゲットに指定する場合。　target==<1000と指定してもかまいません。）

注：文字タイプ変数のターゲット値は、大文字、小文字が区別される点に注意してください。（変数名は大文字・小文字の区別はありません。）

説明変数 (x=)
説明変数を指定します。このパラメータは省略できません。間に1個以上のスペースを入れて、複数の説明変数を指定可能です。また、3通りの省略指定(-,--,:)と3つの特殊指定
（_ALL_,_NUMERIC_,_CHARACTER_）　も利用可能です。

例1：x=age　（説明変数1個を指定）
例2：x=age seibetsu　（説明変数2個を指定）
例3：x=abc1-abc100　（変数名がabcで始まり1から100までの数字で終わる100個の説明変数を指定）
例4：data=a,x=nenrei--jukyo　（入力データセットaに含まれる変数を定義された変数順で検索して、nenreiからjukyoの範囲に含まれる全変数を説明変数に指定）
例5：data=a,x=abc:　（入力データセットaに含まれるabcで始まる全説明変数を指定）
例6：x=age x1-x5 q: time--yz1 nenshu　（説明変数指定方法の複合例）
例7；x=_all_　（全変数）
例8；x=_character_　age（全文字タイプ変数とage）

除外する説明変数 (dropx=&y)
x=パラメータと組み合わせて用い、x=パラメータに指定した説明変数の中で分析から除外する説明変数を指定します。
デフォルトは dropx=&y すなはち、ターゲット変数が除外されます。なお、dropx=パラメータに何か指定すると、常にターゲット変数も除外変数に加わります。x=パラメータにターゲット変数を指定し、dropx=, と明示的にブランク指定を行った場合のみターゲット変数は除外されずに分析に加わることになります。
x=パラメータと同じ指定方法が使えます。

例：
x=_all_,dropx=a_:　（a_で始まる変数およびターゲット変数以外のdata=入力データセットの全変数を説明変数に指定）

数値タイプ説明変数の最大しきい値数 (splitpts=2)
数値説明変数が分岐候補説明変数に選択された場合のカテゴリ併合方法を指定します。1または2を指定できます。(2がデフォルト)。1を指定するとk個のカテゴリを2つに分ける(k-1)通りの併合パターンのみを計算し、採用された場合あるしきい値の前後に分かれることになります。（すべての数値説明変数がデフォルトで順序尺度とみなされます）2(デフォルト)の場合は、2つに分けるパターンと3つに分けて最初と最後を一緒にするパターンの両方を計算し、最適な併合パターンを探索します.（すべての数値説明変数がデフォルトで循環尺度とみなされます）

順序尺度説明変数 (ordinalx=)
カテゴリ併合の際に順序制約を付けたい説明変数名を指定します。間に1個以上のスペースを入れて、複数の説明変数を指定可能です。また、3通りの省略指定（-,--,:）と3つの特殊指定
（_ALL_,_NUMERIC_,_CHARACTER_）　も利用可能です。
x=パラメータと同じ指定方法が使えます。
なお、文字タイプ説明変数の尺度はデフォルトで名義尺度、数値タイプ説明変数の尺度は、splitpts=1を指定した場合は順序尺度、splitpts=2を指定した場合は循環尺度がデフォルトです。デフォルト以外の尺度を指定したい文字タイプ説明変数と数値タイプ説明変数を指定します。

循環尺度説明変数 (cyclicx=)
カテゴリ併合の際に循環制約を付けたい説明変数名を指定します。間に1個以上のスペースを入れて、複数の説明変数を指定可能です。また、3通りの省略指定（-,--,:）と3つの特殊指定
（_ALL_,_NUMERIC_,_CHARACTER_）　も利用可能です。
x=パラメータと同じ指定方法が使えます。
なお、文字タイプ説明変数の尺度はデフォルトで名義尺度、数値タイプ説明変数の尺度はデフォルトで、splitpts=1の場合は順序尺度、splitpts=2の場合はで循環尺度です。デフォルト以外の尺度を指定したい文字タイプ説明変数と数値タイプ説明変数を指定します。

最小ノード件数 (mincnt=AUTO)
生成されるノードの最小件数条件を指定します。
AUTO または 正の整数を指定します。指定が無い場合は AUTO を指定したものとみなされます。

（分類木モデルでAUTO（デフォルト）を指定した場合）
生成されるノードの該当件数をN、ターゲット出現率をp、許容誤差率をerr_rate（ERR_RATE=パラメータで指定します）とすると、以下の条件を満たすノードのみを生成します。

SQRT{p*(1-p)/N}<=err_rate*p

この式の左辺は、N個の抽出データ上で観測されたターゲット出現率pを母集団における真のターゲット出現率の推計値とした場合の標準誤差を表しています。この標準誤差が右辺の観測比率pのerr_rate倍以内に収まるような件数N以上のオブザベーションを持つ条件がノードに課せられます。

上式をNについて解くと、

N>=p*(1-p)/(err_rate*p)*(err_rate*p)

となります。mincnt=AUTO 指定を行うと、ノード必要件数は固定的ではなく、ノードごとのターゲット出現率に応じた一定の誤差許容率を満たすノード必要件数を動的に設定します。

（回帰木モデルでAUTO（デフォルト）を指定した場合）
生成されるノードの該当件数をN、ターゲット平均値と標準偏差をそれぞれm,s、許容誤差率をerr_rate（ERR_RATE=パラメータで指定します）とすると、以下の条件を満たすノードのみを生成します。

s/SQRT(N)<=err_rate*m

この式の左辺は、N個の抽出データ上で観測されたターゲット平均値の標準誤差を表しています。この標準誤差が右辺の観測平均値mのerr_rate倍以内に収まるような件数N以上のオブザベーションを持つ条件がノードに課せられます。

上式をNについて解くと、

N>=s*s/(err_rate*err_rate*m*m)

となります。

しかし、上式では、s=0となるノードは N>=0となってしまうので、

M>=max(N,OYA_N/10,10)

という条件を満たすM をノード必要件数として設定しています。ただし、Nは上記不等式が等式のときのN、OYA_Nは親ノード件数を表します。

mincnt=AUTO 指定を行うと、ノード必要件数は固定的ではなく、ノードごとのターゲット出現率またはターゲット変数平均値に応じた一定の誤差許容率を満たすノード必要件数を動的に設定します。

なお、control=パラメータを指定した場合（アップリフトモデル）では、data=処理群データセット、control=対照群データセットともに、上記条件を満たす要件が課せられます。

（任意の正の整数を指定した場合）
この場合は、各ノードの最小必要件数は固定的になり、分岐後の2つの子ノード件数が共に指定の件数条件を満たすノードのみ生成されます。

推計値の標準誤差に対する許容誤差 (err_rate=0.1)
err_rateは mincnt=AUTO 指定の場合に有効です。0<err_rate<1 の範囲で指定可能です。1に近い値を指定することは、分類木モデルでは許容する誤差範囲（標準誤差）を予測値（0から1の範囲であることに注意）と同じ程度に設定することを意味しますので、予測値のブレが非常に大きなモデルが出来てしまう危険性が高くなります。逆に0に近い値を指定することは、相対的に誤差が小さいノードを生成することにつながりますが、ターゲット出現率の値が0または1に近いノードは非常に多くのノード件数が必要となりますので、そのようなノードは生成されにくくなります。
回帰木モデルの場合も平均値の標準誤差が平均値のerr_rate 倍に収まるために必要な件数を計算してmincntの値を動的に決定します。
入力データセットの件数があまり豊富で無い場合は、このパラメータ値を大きくするか、mincnt=指定に定数値を指定します。

最大分岐レベル (maxlvl=5)
ツリーの最大分岐階層数を指定します。デフォルトは5としていますが、1から最大20までの整数値を指定可能です。（ただし、コンピュータ資源不足などの理由により指定の最大分岐階層までモデル生成でできない場合があり得ます。）モデル生成プロセスは、すべてのノードがこの条件に達するか、mincnt=パラメータ条件を満たす子ノードをそれ以上作成できない場合に終了します。maxlvl=の値をどう指定すれば良いかに関して、mincnt=パラメータのような統計的根拠はありません。最も複雑なルールがこの指定値の数の説明変数の複合によって決定される可能性があること、また生成されるツリーモデルに含まれるルール数（終端セグメント数）が2のmaxlvl乗を超えることはないこと、これらのことと、作成するツリーモデルの用途を考慮してmaxlvlの値を調整してください。

出力ツリーモデル (outmodel=_tree)
生成されたツリーモデルを出力するデータセットに名前をつけます。

入力検証データの指定 (testdata=)
モデル検証用データを指定します。このパラメータ

はGUI実行モードでのみ指定可能です。指定された場合は、「結果表示」ボタンを押した際に作成したモデルが検証データに適用され、ツリー分岐表、ゲインチャート、アップリフトチャートの表示に用いられます。また、比較プロットは検証データの指定が無いと表示されません。

最終カテゴリ併合 (lastcatm=N)
数値タイプ説明変数のカテゴライズ方法に関して、最後のカテゴリを最後から2番目のカテゴリに併合するか否かを指定します。デフォルトはN（併合しない）です。
一般にタイが存在する数値変数（たとえば年齢）の場合、カテゴライズ結果は最後のカテゴリのみ他のカテゴリより件数がかなり少なくなる可能性があります。そのため最後のカテゴリを1つ前のカテゴリと併合する方がモデルの安定性が高まる場合があります。

あらかじめ数値変数をカテゴライズ (precat=Y)
分析開始時にあらかじめ1度だけすべての数値タイプ説明変数をまとめてカテゴライズする(Y)か否(N)かを選択します。precat=Y がデフォルト。
precat=N を指定すると、ノード分割を行うたびに数値説明変数のカテゴライズが行われます。precat=N を指定するとモデルの精度が良くなる可能性がありますが、相対的に実行時間が増加します。

非併合数値タイプ説明変数最大カテゴリ数 (nomergen=STURGES)
個々の数値タイプ説明変数のカテゴライス方法に関して、欠損値を除いた値の種類数がこの値以下の場合、その数値説明変数は個々の値をカテゴリとみなすように指定します。デフォルトはスタージェスの公式で計算された値です。

CEIL(1+log2(N))

ただし、CEILは整数値への切り上げ関数、log2は2を底とする対数関数、Nは欠損値を除くデータ件数を表します。

分析に用いる文字タイプ説明変数の最大カテゴリ数 (maxcatn=1000)
このパラメータは文字タイプ変数が単なるオブザベーション識別変数であって分析対象では無いとみなすためのパラメータです。デフォルトは1000です。
文字タイプ説明変数のカテゴリ数が指定の数を超える場合、その文字タイプ説明変数は分析対象から除外されます。**2〜5000の範囲で指定可能です。**

言語 (language=JAPANESE)
分析実行中のメッセージ出力、結果の表のタイトル、表項目などの表示言語を選択します。ただし、現バージョンでは、日本語か英語の2種類のみ選択可能です。
例：language=ENGLISH

10.1.4 交差検証モデルのパラメータ

以下の指定を行うと、作成するツリーモデルの検証を行う交差検証モデルを作成します

（GUI実行モード）
「交差検証 (testdata=CV)」 の Y にチェックを入れて以下のオプションを指定します。

フォルド(分割)数 (fold=5)
分析データをランダムに指定の数のグループに分割し、同数の交差検証モデルを作成します。fold=2〜20の範囲の整数で指定できます。

乱数シード値 (seed=1)
交差検証実行時のデータ分割に用いる乱数シード値を指定します。正の整数値を指定すると、同じシード値に対して常に同じコンピュータ乱数系列が生成されます。一方、値0を指定すると、生成されるコンピュータ乱数系列は実行するたびに異なるものとなります。分析結果の再現性を求める場合は、シード値は0以外に指定してください。

個々の交差検証ツリーを保存 (Y/N)
交差検証実行時に作成されるfold=パラメータ指定数個の個々の交差検証用ツリーモデルをモデル管理画面に登録して参照可能とするか否かを指定します。Nがデフォルトです。デフォルトではoutmodel=パラメータに指定した分析結果出力モデルと出力モデル名の後に _CV の接尾辞のついた検証用モデル形式データセットの2つのツリーモデルが出力されます。

Yを指定すると、上記2つのツリーモデルの他に、出力モデル名の後に _CV1, _CV2, ... , _CVfold （foldはfoldパラメータの値）の接尾辞が付いた個々の交差検証モデルも出力されます。これらの出力ツリーモデルは、モデル分岐表作成やゲインチャート作成など、他のモデルと同様の操作が可能です。

なお、**個々の交差検証ツリーを保存** (Y/N) の指定に関わらず、outmodel=パラメータに指定した分析結果出力モデルが入ったディレクトリ内に以下のデータセットが保存されます。(**「設定」画面**の**「ツリーモデルディレクトリ」**の**「表示」ボタン**から検索することができます。)

個々の交差検証モデル:
ツリーモデル名_CV1 〜 ツリーモデル名_CVfold
（foldはfoldパラメータの値）
全体交差検証モデル:
ツリーモデル名_CV
個々の交差検証モデルによるモデル予測値を含むデータセット:
ツリーモデル名_CVSC

（コマンド実行モード）

DMT_TREEマクロではなく、DMT_CVTREEマクロを使用します。

outmodel=パラメータに指定した出力モデルデータセット名の後に、_CV1, _CV2, ... , _CVfold （foldはfoldパラメータの値）という接尾辞が付いた個々の交差検証モデルを表すモデル形式データセットと _CVの接尾辞が付いた 全体交差検証モデルを表すモデル形式データセット、さらに、_CVSC の接尾辞がついた 予測スコアデータセットがWORKライブラリに作成されます。

注意： (1) **個々の交差検証モデル**は交差検証時の分割データの状況によって生成されない場合があります。その場合の**個々の交差検証モデルによるモデル予測値**は、その交差検証モデル作成データの全体平均出現率または全体平均値です。これらの予測値に基づいて、**個々の交差検証モデルによるモデル予測値を含むデータセット**を作成しています。
(2) 実行時間は交差検証データ分割回数だけ余分にかかります。分析データ件数が十分と思われる場合は、交差検証ではなく、DMT_DATASAMPを用いて分析データをモデル作成データと検証データに分けてモデル作成とモデル検証を行うことをお勧めします。

10.1.5 コマンド実行モードで有効なパラメータの詳細

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。（GUI 実行モードでは指定できません。）　例：%dmt_tree(help)

std_mod_min_n=9
アップリフトモデルにおけるデータ件数の少ないターゲット出現率や平均値の標準偏差を修正する基準を与えるパラメータです。
そもそも施策実施群の顧客属性と施策非実施群の顧客属性はアンバランスとなることが多いと考えられます。そのため、同一説明変数カテゴリに該当するデータ件数が、処理群と対照群の間で非常にアンバランスとなる場合が起こりえます。そのとき、データ件数が少ない群の方のカテゴリではターゲット出現率や平均値はバラツキ（標準偏差）が大きくなると考えられますが、計算上の標準偏差は0または0に近い不自然な値が得られる場合があります。このような事態を避けるため、std_mod_min_n=パラメータは、指定の値以下のデータ件数から計算されるカテゴリ内のターゲット出現率または目的変数の平均値の標準偏差の計算値が全データの標準偏差より小さい場合に全データの標準偏差に置き換えるよう指示します。

keep_node_data=N
分析終了時にノード分割ごとに生成された中間ノードと終端ノードの所属オブザベーションがそれぞれ含まれるデータセットをWORKライブラリに残すかどうか選択します。
デフォルトは残さない設定です。
例：keep_node_data=Y

node_data_prefix=
keep_node_data=Y を指定した場合に、WORKライブラリに生成される、各ノードの所属オブザベーションを含むデータセットの先頭に付けるプリフィックスワードを指定します。(デフォルトはヌル値、半角で8文字以内)
例:：node_data_prefix=_A_
→ 既定のノード名 _N は _A__N, _C_N10（_Cで始まるノード名は対照群のノードです） は _C_A__N10 に変わります。

keep_node_data=Y を指定したDMT_TREEを実行すると、実行終了後もWORKライブラリに既定のノードデータセット名（_N, _N1, _C_N10 など）が削除されずに残ります。しかし、続いて別のDMT_TREEを、同じく、keep_node_data=Y を指定して実行すると、同じ名前のノードデータセットは新しいものに置き換わってしまいます。

node_data_prefix=パラメータは、WORKライブラリに残っているノードデータはそのまま残しておき、別の名前でノードデータを残したい場合に指定します。

なお、影響するのはWORKライブラリに生成されるノードデータセット名だけです。モデルデータセット内の変数PNODE, CNODE1, CNODE2などの値のノード名（_N, _N11, _C_N001 など）には影響しません。

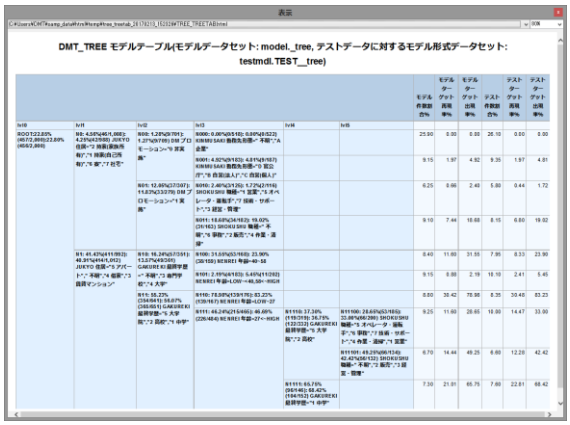
10.1.6 実行例

コマンド実行モードでは表示出力はありません。
モデル作成後、DMT_TREETAB などのモデルデータを入力とする分析結果表示マクロを実行してください。

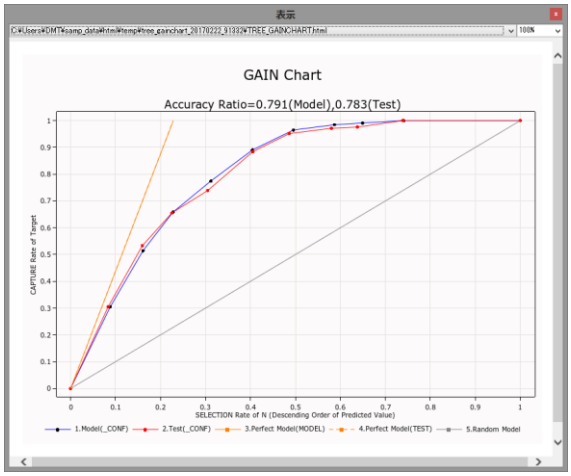
GUI実行モードでは、モデル作成処理終了後、**結果表示 ボタン** を押すと保存された **出力ツリーモデル**（outmodel）と**入力検証データ**（testdata）（もしも指定があれば）または**交差検証モデル**（もしも指定があれば）を用いて、以下の図表を表示します。

［分類木の場合］
(1) ツリー分岐表
検証データの指定、または交差検証モデルの指定がある場合は、検証データにモデルを適用したモデル形式データセット（_TEST+モデル名）を作成、または交差検証モデル（モデル名+_CV）を利用してターゲット予測値を１つのノード内に表示するツリー分岐表を表示します。検証データの指定がない場合は、検証モデル（_TEST+モデル名）は作成されず、モデルのターゲット予測値のみを表示するツリー分岐表

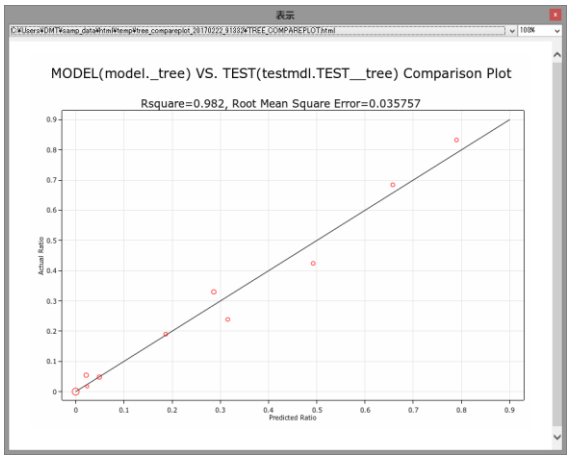
を表示します。



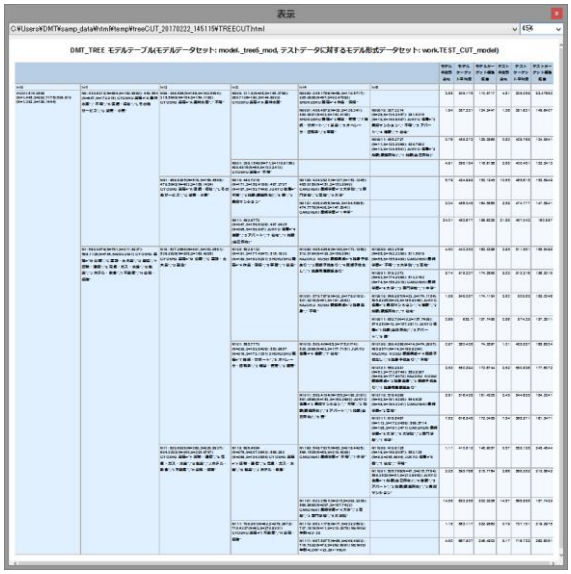
(2) ゲインチャート
検証データが利用可能の場合は、モデルと検証のゲインチャートを１つの図に表示します。
検証データが利用できない場合は、モデルのゲインチャートを表示します。



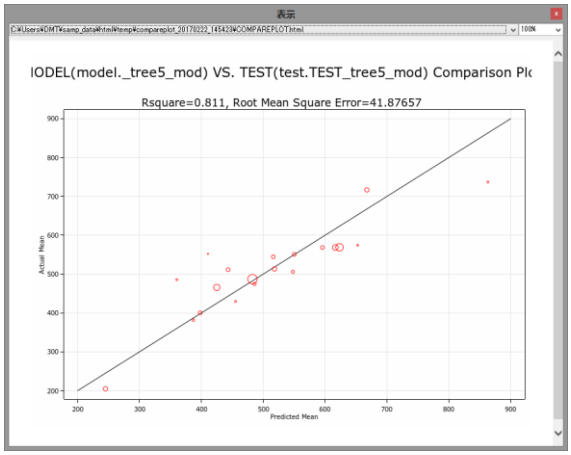
(3) 比較プロット
検証データの指定がある場合のみ表示できます。



[回帰木の場合]
(1) ツリー分岐表
検証データの指定、または交差検証モデルの指定がある場合は、検証データにモデルを適用したモデル形式データセット（_TEST+モデル名）を作成、または交差検証モデルを利用してターゲット予測値を１つのノード内に表示するツリー分岐表を表示します。
検証データの指定がない場合は、検証モデル（_TEST+モデル名）は作成されず、モデルのターゲット予測値のみを表示するツリー分岐表を表示します。

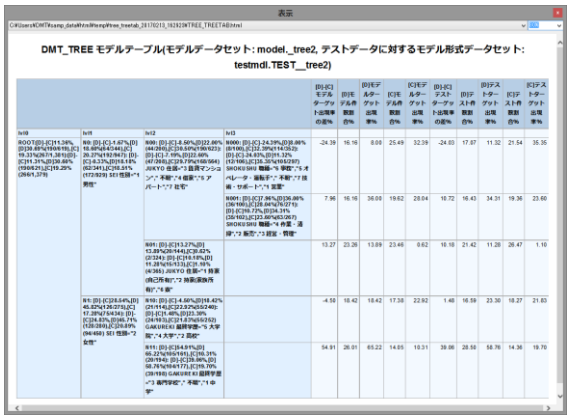


(2) 比較プロット
検証データが利用可能の場合のみ表示します。

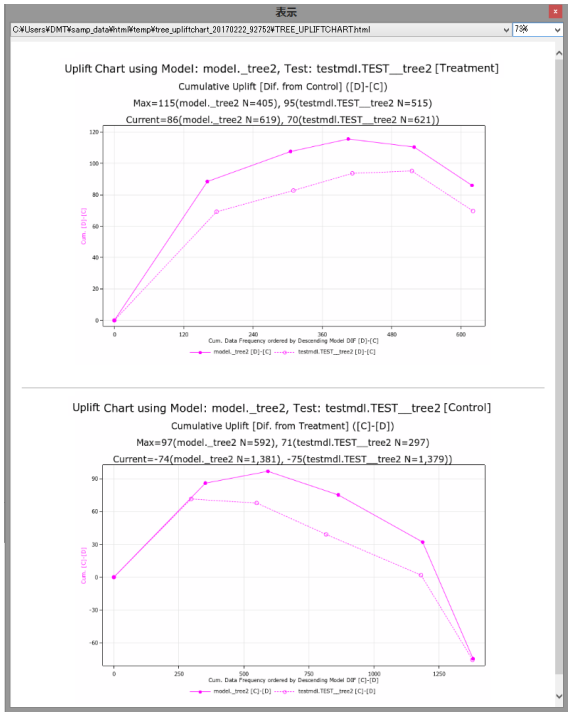


[アップリフトモデルの場合]
(1) ツリー分岐表
検証データの指定、または交差検証モデルの指定がある場合は、検証データにモデルを適用したモデル形式データセット（_TEST+モデル名）を作成、または交差検証モデル（モデル名+_CV）を利用してターゲット予測値を１つのノード内に表示するツリー分

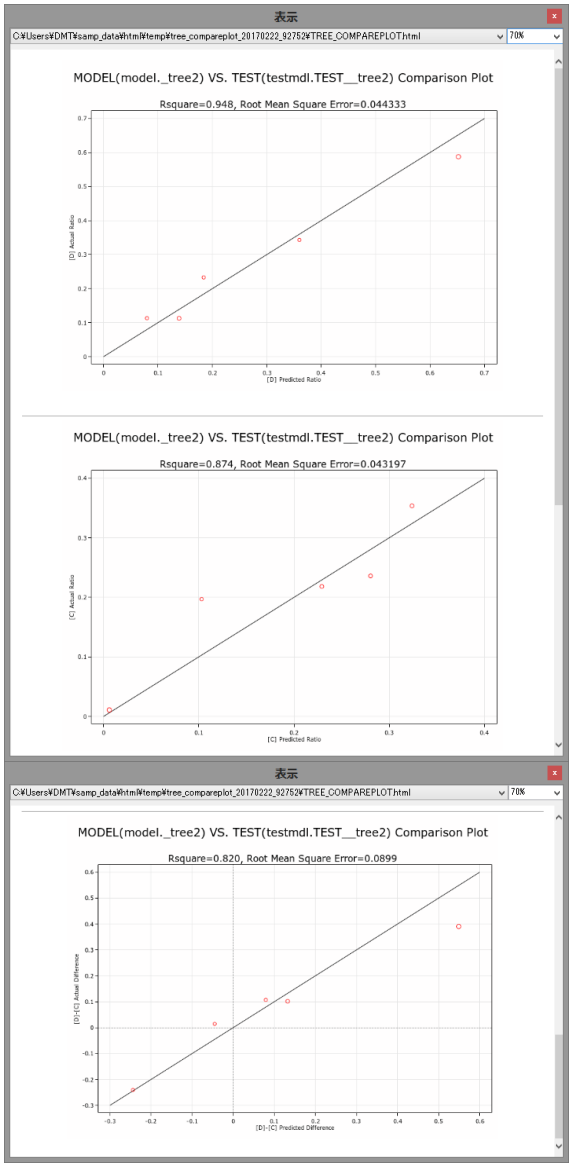
岐表を表示します。検証データの指定がない場合は、検証モデル（_TEST+モデル名）は作成されず、モデルのターゲット予測値のみを表示するツリー分岐表を表示します。



(2) アップリフトチャート
検証データが利用可能の場合は、モデルと検証のアップリフトチャートを１つの図に表示します。
検証データが利用できない場合は、モデルのアップリフトチャートを表示します。



(2) 比較プロット
検証データが利用可能の場合のみ表示します。



10.1.7 層別分析の例

例えば、住居区分別にクロス分析を行うには、以下のように、コマンド実行モードでマクロ言語を使ったプログラムを書いて実行します。

```
proc freq data=samp_data(keep=jukyo);
   tables jukyo/noprint out=jukyo(keep=jukyo);
run;
data _null_;
   if _n_=1 then call symput("n",compress(n));
   set jukyo nobs=n;
   call symput("JUKYO"||left(_n_),compress(jukyo));
run;
```

(マクロ変数値の確認)

```
%put &n, &JUKYO1, &JUKYO2, ... , &&JUKYO&n;
```

(ログ)

```
8, , 1, ... , 7
```

(住居区分別に、クロス分析をまず行い、次にその結果から説明力のある変数のみでツリーモデルを作成)

```
%macro create_model;
  %do i=1 %to &n;
    %dmt_cross(data=samp_data(where=(jukyo="&&JUKYO&i")),y=flg,target=1,x=sei--DM,outcross=CROSS_&&JUKYO&i)
    %dmt_tree(data=samp_data(where=(jukyo="&&JUKYO&i")),y=flg,target=1,x=sei--DM,dropx=&_XDEL,mincnt=50,maxlvl=5,outmodel=JUKYO_&&JUKYO&i..MODEL)
    %if  %sysfunc(exist(JUKYO_&&JUKYO&i..MODEL)) %then %do;
      %dmt_treescore(model=JUKYO_&&JUKYO&i..MODEL,data=test_data(where=(jukyo="&&JUKYO&i")),outscore=JUKYO_&&JUKYO&i..SCORE)
      %dmt_gainchart(data=JUKYO_&&JUKYO&i..SCORE,y=flg,target=1);
    %end;
  %end;
%mend create_model;
%create_model
```

注意：層別変数は文字変数で半角英数字の短い値を仮定しています。数値変数の場合は、where=(変数名=値)の値を引用符で囲うとエラーになります。

10.1.8 データセット出力

生成されたツリーモデルがoutmodel=パラメータに指定されたデータセットに出力されます。
デフォルトは WORK._TREE という名前で出力されます。

以下の項目がデータセットに含まれています。

（分類木モデルの場合）
outmodel=出力データセット

変数名	タイプ	長さ	内容	備考
PNODE	文字	可変	親ノードの名前	"_Nxxxx"の値。ただしxxxxは0/1の文字列
CNODE1	文字	可変	子ノード1の名前	"_Nxxxx"の値。ただしxxxxは0/1の文字列
CNODE2	文字	可変	子ノード2の名前	"_Nxxxx"の値。ただしxxxxは0/1の文字列
TERM1	文字	3	子ノード1の終端識別	"YES"または"NO"
TERM2	文字	3	子ノード2の終端識別	"YES"または"NO"
ITEM	文字	32	分岐に用いる説明変数名	
ITEM_TYPE	文字	2	分岐に用いる説明変数のタイプ	"C"（文字タイプ）または"N"（数値タイプ）
CNODE1_CAT	文字	5000	子ノード1のカテゴリ値	"a","b"（文字変数の場合）、または1~10（数値変数の場合）といった形式
CNODE2_CAT	文字	5000	子ノード2のカテゴリ値	"a","b"（文字変数の場合）、または1~10（数値変数の場合）といった形式
CNODE1_TOT_N	数値	8	子ノード1に含まれる件数	
CNODE2_TOT_N	数値	8	子ノード2に含まれる件数	
CNODE1_TARG_N	数値	8	子ノード1に含まれるターゲット件数	
CNODE2_TARG_N	数値	8	子ノード2に含まれるターゲット件数	
Dif_Entropy	数値	8	親ノードの状態から2つの子ノードに分かれた状態に移行したときのエントロピー値の差	必ず0もしくは負の値（減少を表す）になるが、減少幅が大きいほど分岐後の子ノード間のターゲット出現率の差が大きいことを表す

（回帰木モデルの場合）
outmodel=出力データセット

変数名	タイプ	長さ	内容	備考
PNODE	文字	可変	親ノードの名前	"_Nxxxx"の値。ただしxxxxは0/1の文字列
CNODE1	文字	可変	子ノード1の名前	"_Nxxxx"の値。ただしxxxxは0/1の文字列
CNODE2	文字	可変	子ノード2の名前	"_Nxxxx"の値。ただしxxxxは0/1の文字列
TERM1	文字	3	子ノード1の終端識別	"YES"または"NO"
TERM2	文字	3	子ノード2の終端識別	"YES"または"NO"
ITEM	文字	32	分岐に用いる説明変数名	
ITEM_TYPE	文字	2	分岐に用いる説明変数のタイプ	"C"（文字タイプ）または"N"（数値タイプ）
CNODE1_CAT	文字	5000	子ノード1のカテゴリ値	"a","b"（文字変数の場合）、または1~10（数値変数の場合）といった形式
CNODE2_CAT	文字	5000	子ノード2のカテゴリ値	"a","b"（文字変数の場合）、または1~10（数値変数の場合）といった形式
WSS	数値	8	2つの子ノードにおけるターゲット変数の群内平方和の合計値	
CNODE1_TOT_N	数値	8	子ノード1に含まれる件数	
CNODE2_TOT_N	数値	8	子ノード2に含まれる件数	
CNODE1_MEAN	数値	8	子ノード1のターゲット平均値	
CNODE2_MEAN	数値	8	子ノード2のターゲット平均値	
CNODE1_STD	数値	8	子ノード1のターゲット標準偏差	
CNODE2_STD	数値	8	子ノード2のターゲット標準偏差	
Dif_WSS	数値	8	親ノードの状態から2つの子ノードに分かれた状態に移行したときのターゲット変数の群内平方和の減少分（＝群間平方和）	必ず0もしくは正の値になり、大きいほど分岐後の子ノード間のターゲット平均値の差が大きいことを表す

（分類木アップリフトモデルの場合）
outmodel=出力データセット

変数名	タイプ	長さ	内容	備考
PNODE	文字	可変	親ノードの名前	"_Nxxxx"の値。ただしxxxxは0/1の文字列
CNODE1	文字	可変	子ノード1の名前	"_Nxxxx"の値。ただしxxxxは0/1の文字列
CNODE2	文字	可変	子ノード2の名前	"_Nxxxx"の値。ただしxxxxは0/1の文字列
TERM1	文字	3	子ノード1の終端識別	"YES"または"NO"
TERM2	文字	3	子ノード2の終端識別	"YES"または"NO"
ITEM	文字	32	分岐に用いる説明変数名	
ITEM_TYPE	文字	2	分岐に用いる説明変数のタイプ	"C"（文字タイプ）または"N"（数値タイプ）
CNODE1_CAT	文字	5000	子ノード1のカテゴリ値	"a","b"（文字変数の場合）、または1~10（数値変数の場合）といった形式
CNODE2_CAT	文字	5000	子ノード2のカテゴリ値	"a","b"（文字変数の場合）、または1~10（数値変数の場合）といった形式
AIC	数値	8	AIC値	値が負で絶対値が大きいほど有意な分岐であることを意味する。
D_CNODE1_TOT_N	数値	8	子ノード1の処理群に含まれる件数	子ノード1の統計量
D_CNODE1_TARG_N	数値	8	子ノード1の処理群に含まれるターゲット件数	
C_CNODE1_TOT_N	数値	8	子ノード1の対照群に含まれる件数	
C_CNODE1_TARG_N	数値	8	子ノード1の対照群に含まれるターゲット件数	
DIF_CNODE1_CONF	数値	8	子ノード1の処理群と対照群間のターゲット出現率の差	
DIF_CNODE1_SE	数値	8	子ノード1の処理群と対照群間のターゲット出現率の差の標準誤差	
D_CNODE2_TOT_N	数値	8	子ノード2の処理群に含まれる件数	子ノード2の統計量
D_CNODE2_TARG_N	数値	8	子ノード2の処理群に含まれるターゲット件数	
C_CNODE2_TOT_N	数値	8	子ノード2の対照群に含まれる件数	
C_CNODE2_TARG_N	数値	8	子ノード2の対照群に含まれるターゲット件数	
DIF_CNODE2_CONF	数値	8	子ノード2の処理群と対照群間のターゲット出現率の差	
DIF_CNODE2_SE	数値	8	子ノード2の処理群と対照群間のターゲット出現率の差の標準誤差	

（回帰木アップリフトモデルの場合）
outmodel=出力データセット

変数名	タイプ	長さ	内容	備考
PNODE	文字	可変	親ノードの名前	"_Nxxxx"の値。ただしxxxxは0/1の文字列
CNODE1	文字	可変	子ノード1の名前	"_Nxxxx"の値。ただしxxxxは0/1の文字列
CNODE2	文字	可変	子ノード2の名前	"_Nxxxx"の値。ただしxxxxは0/1の文字列
TERM1	文字	3	子ノード1の終端識別	"YES"または"NO"
TERM2	文字	3	子ノード2の終端識別	"YES"または"NO"
ITEM	文字	32	分岐に用いる説明変数名	
ITEM_TYPE	文字	2	分岐に用いる説明変数のタイプ	"C"（文字タイプ）または"N"（数値タイプ）
CNODE1_CAT	文字	5000	子ノード1のカテゴリ値	"a","b"（文字変数の場合）、または1~10（数値変数の場合）といった形式
CNODE2_CAT	文字	5000	子ノード2のカテゴリ値	"a","b"（文字変数の場合）、または1~10（数値変数の場合）といった形式
AIC	数値	8	AIC値	値が負で絶対値が大きいほど有意な分岐であることを意味する。
D_CNODE1_TOT_N	数値	8	子ノード1の処理群に含まれる件数	子ノード1の統計量
D_CNODE1_MEAN	数値	8	子ノード1の処理群のターゲット平均値	
D_CNODE1_STD	数値	8	子ノード1の処理群のターゲット標準偏差	
C_CNODE1_TOT_N	数値	8	子ノード1の対照群に含まれる件数	
D_CNODE1_MEAN	数値	8	子ノード1の対照群のターゲット平均値	
D_CNODE1_STD	数値	8	子ノード1の対照群のターゲット標準偏差	
DIF_CNODE1_MEAN	数値	8	子ノード1の処理群と対照群間のターゲット平均値の差	
DIF_CNODE1_SE	数値	8	子ノード1の処理群と対照群間のターゲット平均値の差の標準誤差	
D_CNODE2_TOT_N	数値	8	子ノード2の処理群に含まれる件数	子ノード2の統計量
D_CNODE2_MEAN	数値	8	子ノード2の処理群のターゲット平均値	
D_CNODE2_STD	数値	8	子ノード2の処理群のターゲット標準偏差	
C_CNODE2_TOT_N	数値	8	子ノード2の対照群に含まれる件数	
D_CNODE2_MEAN	数値	8	子ノード2の対照群のターゲット平均値	
D_CNODE2_STD	数値	8	子ノード2の対照群のターゲット標準偏差	
DIF_CNODE2_MEAN	数値	8	子ノード2の処理群と対照群間のターゲット平均値の差	
DIF_CNODE2_SE	数値	8	子ノード2の処理群と対照群間のターゲット平均値の差の標準誤差	

もしも交差検証を指定した場合は、個々の交差検証モデル（outmodel=出力モデル名+_CVとoutmodel=出力モデル名+_CV1~_CVfold数 のモデル形式データセット）や交差検証予測値が付けられた分析デー

タ（outmodel=出力モデル名+_CVSC）が出力されます。交差検証モデルについては、個々のモデルも全体モデルも上記と同じ項目が含まれています。

交差検証予測値を含むデータセット（outmodel=出力モデル名+_CVSC）には、以下の項目が追加されます。

交差検証予測値データセット（元のモデル作成データに追加される項目）
データセット名：出力モデル名+_CVSC

変数名	タイプ	長さ	内容	備考
_CV_NO	数値	8	交差検証データ分割番号	1～fold数
_CV_CONF	数値	8	交差検証モデル予測値	分類木モデルの場合
_CONF	数値	8	モデル予測値	
_CV_MEAN	数値	8	交差検証モデル予測値	回帰木モデルの場合
_MEAN	数値	8	モデル予測値	
_CV_DIF_CONF	数値	8	交差検証モデル予測値（処理群予測値-対照群予測値）	分類木アップリフトモデルの場合
_CV_D_CONF	数値	8	交差検証モデル予測値（処理群予測値）	
_CV_C_CONF	数値	8	交差検証モデル予測値（対照群予測値）	
DIF_CONF	数値	8	モデル予測値（処理群予測値-対照群予測値）	
D_CONF	数値	8	モデル予測値（処理群予測値）	
C_CONF	数値	8	モデル予測値（対照群予測値）	
_CV_DIF_MEAN	数値	8	交差検証モデル予測値（処理群予測値-対照群予測値）	回帰木アップリフトモデルの場合
_CV_D_MEAN	数値	8	交差検証モデル予測値（処理群予測値）	
_CV_C_MEAN	数値	8	交差検証モデル予測値（対照群予測値）	
DIF_MEAN	数値	8	モデル予測値（処理群予測値-対照群予測値）	
D_MEAN	数値	8	モデル予測値（処理群予測値）	
C_MEAN	数値	8	モデル予測値（対照群予測値）	
_NODE	文字	可変	モデル所属ノード番号	交差検証モデル予測値の集計用の outmodel=出力モデルの所属ノード番号
_TERM	文字	3	終端ノード判定フラグ	YES/NO
_UNMATCH	文字	3	アンマッチ判定フラグ	YES/NO

さらに、コマンド実行モードで keep_node_data=Y を指定すると、ルートノード(_N)以外のすべての生成された中間ノードおよび終端ノードが、ノード名をデータセット名としてWORKライブラリに生成されたまま消さずに残ります。これらは、ノードごとの詳細な内容を調べたい場合に役に立つと思われます。

例:_N0,_N100, _C_N111 （_Cで始まるノード名は対照データのノードを意味します。）

これらのデータセットには以下の変数が含まれます。

・ターゲット変数
・全説明変数
・_obsno（入力データセットのオブザベーション番号）
・_targflg（ターゲット値(1)、非ターゲット値(0)を識別する変数。分類木モデル、分類木アップリフトモデルの場合のみ）

10.1.9 欠損値の取り扱い

文字タイプのターゲット変数、説明変数はいずれも有効な値の1つとみなされます。

数値タイプの説明変数に特殊欠損値(._,.A~.Z)が存在した場合は通常欠損値(.)に変換した上で使用されます。

数値タイプのターゲット変数の欠損値は、回帰木モデル作成時(target=パラメータを指定しなかった場合)にデータに存在すると、オブザベーション単位で分析から除外します。分類木モデル作成時(target=パラメータを指定した場合)は、数値タイプのターゲット変数の欠損値(.)は、特殊欠損値(._,.A~.Z)と区別して他の数値と同様に取り扱われます。

10.1.10 制限

評価版のDMT_TREEマクロで処理できる入力データセットのオブザベーション数の最大は2,000です。製品版ではこの制限はありませんが、コンピュータ資源等の制約により実質的に取扱えるオブザベーション数には限りがあります。

入力できる説明変数の最大数は2,000です。ただし、各変数のカテゴリ数、その他の要因によるコンピュータ資源不足などの理由で1回の分析では2,000未満の説明変数しか取り扱えない場合もあり得ます。

指定可能な最大階層数は20に設定しています。
ただし、20階層まですべての親ノードが子ノードに分岐するとした場合、2の20乗（=1,048,576）個の終端ノードが生成され、中間ノードを含めるとその倍の数のノードをワーク領域に保持します。コンピュータ資源（メモリ、ポインタその他）の制約、その他の理由から、20階層未満のツリーしか作成できない場合もあり得ます。

注意：1回の分岐において3つ以上のノードに分ける機能、同時に3つ以上のターゲット峻別行う機能は現バージョンのDMT_TREE、DDMT_CVTREEにはありません。

入力データセットに以下の変数が存在する場合、警告を出して処理を中止します。入力データセットから削除しておくか、変数名を変えてください。(_v&i.cは_V+数字+Cという形式の変数名を表します。)

```
_id _item _obsno _targflg _v&i.c
```

10.1.11 コマンド実行モードでの注意

ユーザ定義フォーマットがついた変数を含むデータセットをアクセスするためには、そのフォーマットも利用可能でなければなりません。ユーザ定義フォーマットのついた変数を含む分析データセットを永久保存する場合は、そのフォーマットも永久保存してください。

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

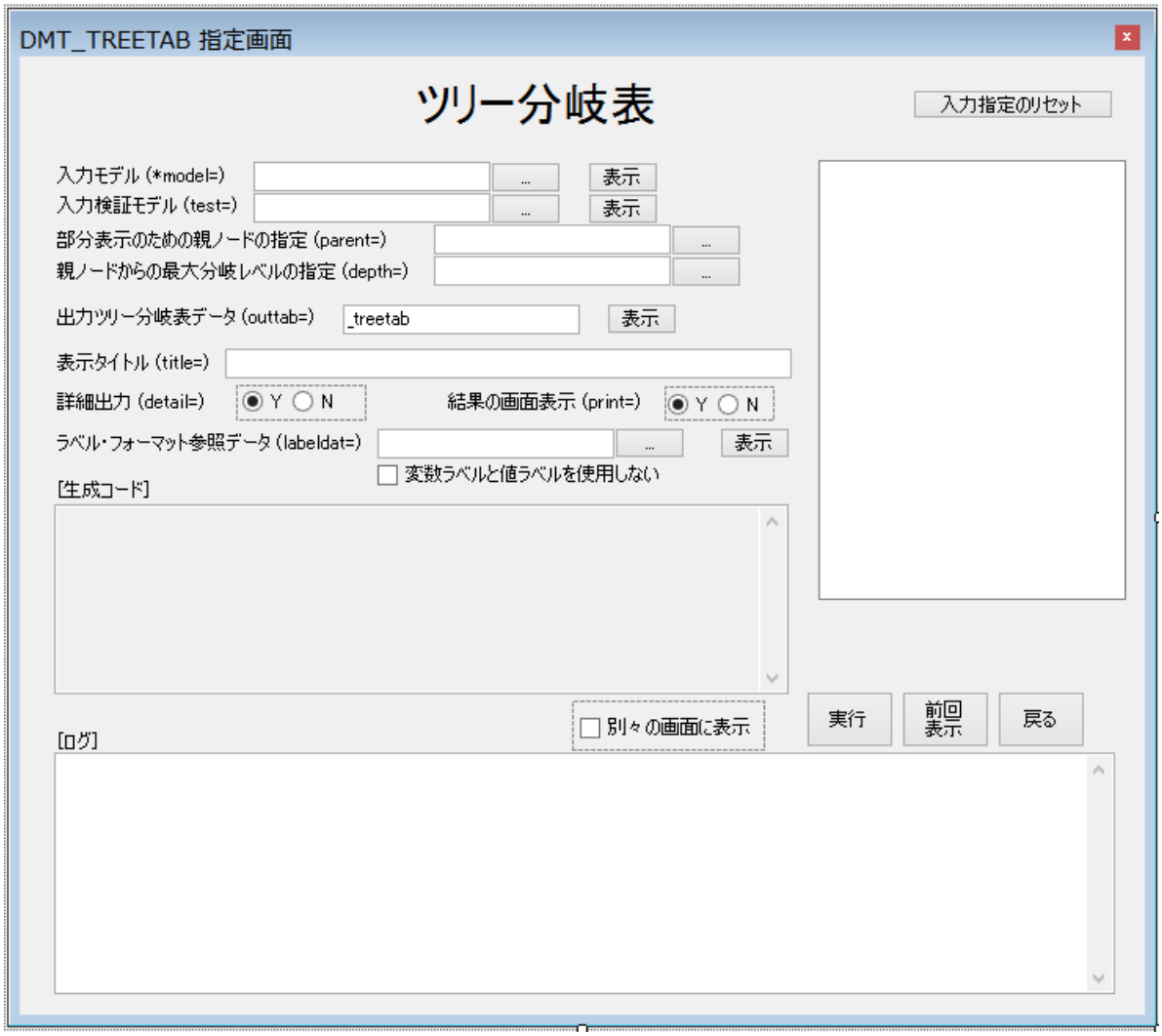
また、以下のユーザ定義フォーマットがWORKライブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
$_item $_VARTYP $_VARSCL
```

さらに、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
e_name e_type lab&i nobs spc&i typ&i zketa
_nofound _speclen _specnum _delnode _errormsg
```

10.2 分岐表（dmt_treetab）



10.2.1 概要

ツリー分岐表（DMT_TREETAB）は**デシジョンツリーモデル作成**（DMT_TREE）を実行して作成されたモデルデータセット、または**新しいデータを基準にモデル予測値修正**（DMT_TREESCORE）を実行して作成されたモデル形式データセットを入力として、ツリーモデルの内容をノード分岐過程がわかる階層形式の表として画面表示します。表示される各ノードの情報は、各ノードにおける件数やターゲット値の分布情報および親ノードからの分岐に用いられた説明変数名と値です。DMT_TREEで作成されたモデルデータセットとそのモデルをDMT_TREESCOREを用いてテストデータに適用して得られたテストデータにおけるモデル形式データセットの**両方**を入力とした場合は、各ノードにおけるモデル作成データ、テストデータそれぞれの該当件数、ターゲット件数およびターゲット出現率またはターゲット平均値と標準偏差を表の各ノードの中に同時表示します。

10.2.2 指定方法

（コマンド実行モードでの指定）

```
%dmt_treetab(help,model=,test=,parent=N
,depth=&_max_lvl,outtab=_treetab,print=Y,labeldat=
,nolabel=N,detail=N,title=,pctf=7.2,meanf=best8.
,d_label=[D].c_label=[C],dif_label=[D]-[C]
,language=JAPANESE
,outhtml=dmt_treetab.html,outpath=)
```

（GUI実行モードでの変更点）

・help, outhtml=, outpath=パラメータは指定不可。（自動で行われます。）
・print=Y に固定。
・labeldat=パラメータは自動入力。

（必須パラメータ）

以下の1個のパラメータは省略できません。

入力モデル（model=）

（オプションパラメータ）

以下の18個のパラメータは任意指定です。（=の右辺の値はデフォルト値を表しています）

help ... 指定方法のヘルプメッセージの表示。（コマンド実行モードでのみ有効）
入力検証モデル（test=）
... モデルをテストデータに適用して得られたモデル形式入力データセット名を指定。
部分表示のための親ノードの指定（parent=N）
... 描きたいツリーのルートノードを指定。
親ノードからの最大分岐レベルの指定
(depth=&_max_lvl)
... 描きたいツリーのルートノードからの深さレベル数を指定。
出力ツリー分岐表データ（outtab=_treetab）
... ツリー表画面出力データセットの名前を指定。
結果の画面表示（print=Y）
... ノードテーブルの画面表示を行うか否かを指定。(Y または N, GUI実行モードではY固定)
変数ラベルと値ラベルを表示しない（nolabel=N）
… 変数ラベルと値ラベルを用いずに変数名、変数値を用いた結果表を作成。
詳細出力（detail=N）
... 詳細な終端ノード統計量を表示。
画面出力のタイトルの指定（title=）
.... %str,%nrstr,%bquote などの関数で囲んで指定する（コマンド実行モードでのみ有効）
百分率の表示フォーマットの指定（pctf=7.2）
平均値・標準偏差の表示フォーマットの指定
(meanf=best8.)
アップリフトモデルにおける処理群（DATA）を表す記号
(d_label=[D])
アップリフトモデルにおける対照群（Control）を表す記号
(c_label=[C])
アップリフトモデルにおける処理群−対照群間の差を表す記号（dif_label=[D]-[C])
言語の選択（language=JAPANESE）
HTML出力ファイル名（outhtml=dmt_crosstab.html）
（コマンド実行モードでのみ有効）
HTMLファイル出力ディレクトリの指定（outpath=）（コマンド実行モードでのみ有効）
ラベル・フォーマット参照データ（labeldat=）
（コマンド実行モードでのみ有効）

10.2.3 パラメータの詳細

入力モデル（model=）
入力モデルデータセット名を指定します。このパラメータは省略できません。
例：model=bunseki1

入力検証モデル（test=）
dmt_treescoreを用いてモデルをモデル検証用データセットに適用したときのモデル形式データセットが得られている場合、そのモデル形式データセット名を指定します。この指定により、各ノードごとに、モデルの集計値に加えて検証データにおける集計値も同時表示されます。
例：test=test1

部分表示のための親ノードの指定（parent=N）
ツリーモデルを部分表示するための指定です。指定のノードをルートノードとみなした場合の部分ツリーを表示します。デフォルトはparent=N、すなはち本来のルートノードです。ただし、ノード名の最初の"_"(アンダースコア,アンダーバー)は省略して指定しなければなりません。ツリーの階層数が多く、一度にツリー全体を表示することができない場合、または表示できたとしても大きな表となってしまいページにフィットしないような場合、このパラメータとdepth=パラメータを用いてモデルの部分表示を行います。たとえば、parent=N0を指定してdmt_treetabを実行すると、ノードN0から分岐しているノードのみが画面表示されます。続いてparent=N1を指定してdmt_treetabをもう一度実行すると、今度はノードN1から分岐しているノードのみが画面表示されます。2つの表示を合わせるとモデル全体の情報が得られます。（さらにN1とN0の関係は paraent=N, depth=1 を指定した部分ツリーで表示することもできます。）

なお、ノード名の規則はルートノードをNとし、第一階層の2つのノードをそれぞれN0,N1としています。第k階層の任意のノードを Nxxx...x（ただし、xは0または１のいずれかの値を持ち、xxx...xの部分はk個のxの列だとします）とすると、その子ノードはNxxx...x0とNxxx...x1と表記されます。
例： parent=N01

親ノードからの最大分岐レベルの指定
(depth=&_max_lvl)
ツリーモデルを部分表示するときに用いる指定です。指定階層数のみを表示します。デフォルトはmodel=に指定されたモデルデータセットの最大階層です。ツリーモデルでは重要な説明変数ほどツリーの浅い階層の分岐に使用されますので、作成されたモデルの最大階層数が大きい場合、最初の階層の分岐を見ることが重要です。なお、この指定は相対的な階層数を意味しています。parent=指定があれば、parent=に指定されたノードをルートノードとみなしてそこからdepth=パラメータの値の階層数までを表示します。

出力ツリー分岐表データ（outtab=_treetab）
ツリー表画面表示用データセットを出力したい場合に指定します。デフォルトは_treetabです。ツリー表

画面表示用データセットとは、最終的な表示を行うproc tabulateへ直接入力できるデータセットの意味です。

表示タイトル (title=)
画面出力される表にタイトルを指定できます。指定しない（デフォルト）場合、以下のようなタイトルが自動的に付与されます。

%bquote(DMT_TREE **モデルデータセット**: &model, **テストデータに対するモデル形式データセット**: &test)

タイトルを指定する場合、上記のように%bquote関数の中に記述してください。

詳細出力 (detail=N, または details=N)
ツリー分岐表の表示項目を制御します。デフォルトはdetail=N。detail=Yを指定すると、表示項目数が増えます。

言語 (language=JAPANESE)
分析実行中のメッセージ出力、結果の表のタイトル、表項目などの表示言語を選択します。ただし、現バージョンでは、日本語か英語の2種類のみ選択可能です。
例：language=ENGLISH

結果の画面表示 (print=Y)
ツリーテーブルを画面表示する（Y）かしない（N）かを指定します。デフォルトは画面表示する（Y）です。print=Nを指定しても、outtab=パラメータに指定したデータセットに画面表示するためのツリー情報が出力されます。

変数ラベルと値ラベルを表示しない (nolabel=N)
Yを指定すると、表示が元の変数名、値に変わります。

ラベル・フォーマット参照データ (labeldat=)
ラベルとフォーマットが定義されたデータセットを指定することにより、分析結果の全変数名と文字タイプ変数値に、それぞれ定義された変数ラベルとフォーマットが付加されて表示されるようになります。この指定が無い場合は、変数名、変数値がそのまま表示されます。数値タイプ説明変数には、フォーマットが定義されていたとしても無視します。なお、フォーマット定義された変数を含むデータセットをアクセスするためには、そのフォーマットライブラリもアクセス可能になっている必要があります。ラベル定義されたデータセットを保存して再利用したい場合は、フォーマットライブラリも保存しておく必要があります。（GUI実行モードではモデルがどのデータから作成されたかを記録しているため、そのデータが存在する場合は自動入力されます）

10.2.4 コマンド実行モードで有効なパラメータの詳細

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。
例：%dmt_treetab(help)

10.2.5 HTML 出力

分析結果の図表はhtmlファイルに出力されます。保存先はデフォルトではSASディスプレイマネージャまたはWPSワークベンチの管理下（ワークスペース内の一時保存ファイル）です。outpath=パラメータを指定すると、保存先を変更できます。（必ずフルパス指定します。引用符で囲んでも囲まなくてもかまいません）同時にouthtml=パラメータを指定すると、保存するhtmlファイルに自由に名前を付けることができます。

outhtml=dmt_treetab.html
分析結果を保存するHTML出力ファイル名を指定します。
例：outhtml=out1.html,

outpath=
HTML図表出力ファイルの保存ディレクトリを指定します。このパラメータを指定しない場合（デフォルト）、HTMLファイルはSASディスプレイマネージャまたはWPSワークベンチの管理下に作成されます。outpath=指定を行う場合、値は必ずフルパスで指定する必要があります。なお、パス指定全体を引用符で囲んでも囲まなくてもかまいません。
例：outpath='G:¥temp'

10.2.6 実行例

例１：分類木（検証結果表示なし、変数ラベル、値ラベルなし）

```
%dmt_tree(data=data.samp_data,y=flg,target=1,x=sei nenrei,outmodel=flg1)
%dmt_treetab(model=flg1)
```

DMT_TREE モデルテーブル (モデルデータセット: flg1)

			件数割合%	ターゲット再現率%	ターゲット出現率%
LVL0	LVL1	LVL2			
ROOT:22.85% (457/2,000)	N0: 41.44%(92/222) NENREI=LOW~23		11.10	20.13	41.44
	N1: 20.53%(365/1,778) NENREI=23<~HIGH	N10: 18.09% (214/1,183) SEI="1"	59.15	46.83	18.09
		N11: 25.38%(151/595) SEI="2"	29.75	33.04	25.38

例２：回帰木（変数ラベル、値ラベルあり、検証結果表示あり、数値の表示フォーマット指定あり）

```
%dmt_tree(data=samp_data,y=nenshu,x=sei
nenrei,outmodel=nenshu1,maxlvl=2)
%dmt_treescore(model=nenshu1,data=test_data,y=
nenshu,outmodel=TEST_nenshu1)
%dmt_treetab(model=nenshu1,test=TEST_nenshu1,
labeldat=samp_data,pctf=3.,meanf=6.1)
```

注：GUI実行モードではlabeldat=パラメータは自動設定されます。

DMT_TREE モデルテーブル(モデルデータセット: nenshu1, テストデータに対するモデル形式データセット: TEST_nenshu1)

			モデル件数割合%	モデルターゲット平均値	モデルターゲット標準偏差	テスト件数割合%	テストターゲット平均値	テストターゲット標準偏差
lvl0	lvl1	lvl2						
ROOT:514.0 (N=1,445,S=202.7):508.8 (N=1,392,S=198.1)	N0: 506.3 (N=898,S=201.1): 504.5 (N=841,S=203.0) SEI 性別="1 男性"	N00: 495.5(N=561,S=190.9): 498.3 (N=537,S=199.5) NENREI 年齢=33~58	39	495.5	190.9	39	498.3	199.5
		N01: 524.4(N=337,S=215.8): 515.3 (N=304,S=208.5) NENREI 年齢=LOW~<33,58<~HIGH	23	524.4	215.8	22	515.3	208.5
	N1: 526.7 (N=547,S=204.8): 515.5 (N=551,S=190.3) SEI 性別="2 女性"	N10: 544.2(N=363,S=214.7): 523.5 (N=371,S=193.9) NENREI 年齢=23~48	25	544.2	214.7	27	523.5	193.9
		N11: 492.1(N=184,S=178.7): 499.0 (N=180,S=181.6) NENREI 年齢=LOW~<23,48<~HIGH	13	492.1	178.7	13	499.0	181.6

例３：ツリーの部分表示。ルートノードから1階層分のみ表示

```
%dmt_treetab(model=nenshu1,test=TEST_nenshu1,
labeldat=samp_data,meanf=6.1,parent=N,depth=1)
```

DMT_TREE 部分モデルテーブル (モデルデータセット: nenshu1, ROOT ノード から 1 階層下までのノードを表示)

		件数割合%	ターゲット平均値	ターゲット標準偏差	テスト件数割合%	テストターゲット平均値	テストターゲット標準偏差
lvl0	lvl1						
ROOT:514.0(N=1,445,S=202.7):508.8 (N=1,392,S=198.1)	N0: 506.3(N=898,S=201.1): 504.5 (N=841,S=203.0) SEI 性別="1 男性"	62.15	506.3	201.1	60.42	504.4	203.0
	N1: 526.7(N=547,S=204.8): 515.5 (N=551,S=190.3) SEI 性別="2 女性"	37.85	526.7	204.8	39.58	515.5	190.3

例４：詳細表示

```
%dmt_tree(data=SAMP_DATA(where=(DM="1")),co
ntrol=SAMP_DATA(where=(DM="0")),y=flg,target=1
,x=sei nenrei jukyo,outmodel=flg_uplift
,mincnt=100,maxlvl=5)
%dmt_treetab(model=flg_uplift,depth=1)
```

```
%dmt_treetab(model=flg_uplift,depth=1,detail=Y)
```

注：DMT_TREETABマクロの detail=Yオプションは、アップリフトモデルの表示の場合のみ有効です。

DMT_TREE 部分モデルテーブル (モデルデータセット: _tree_flg_uplift, ROOT ノード から 1 階層下までのノードを表示)

		[D]-[C]ターゲット出現率の差%	[D]件数割合%	[D]ターゲット出現率%	[C]件数割合%	[C]ターゲット出現率%
LVL0	LVL1					
ROOT:[D]-[C]11.36%(SE=2.14%),[D]30.69%(190/619),[C]19.33%(267/1,381)	N0: [D]-[C]-1.67%(SE=2.47%),[D]18.60%(64/344),[C]20.27%(192/947) SEI="1"	-1.67	55.57	18.61	68.57	20.28
	N1: [D]-[C]28.54%(SE=3.51%),[D]45.82%(126/275),[C]17.28%(75/434) SEI="2"	28.54	44.43	45.82	31.43	17.28

DMT_TREE 部分モデルテーブル (モデルデータセット: _tree_flg_uplift, ROOT ノード から 1 階層下までのノードを表示)

		[D]-[C]ターゲット出現率の差%	[D]-[C]ターゲット出現率の差の標準誤差%	[D]件数割合%	[D]ターゲット再現率%	[D]ターゲット出現率%	[C]件数割合%	[C]ターゲット再現率%	[C]ターゲット出現率%
LVL0	LVL1								
ROOT:[D]-[C]11.36%(SE=2.14%),[D]30.69%(190/619),[C]19.33%(267/1,381)	N0: [D]-[C]-1.67%(SE=2.47%),[D]18.60%(64/344),[C]20.27%(192/947) SEI="1"	-1.67	2.47	55.57	33.68	18.61	68.57	71.91	20.28
	N1: [D]-[C]28.54%(SE=3.51%),[D]45.82%(126/275),[C]17.28%(75/434) SEI="2"	28.54	3.51	44.43	66.32	45.82	31.43	28.09	17.28

10.2.7 データセット出力

outtab=パラメータ に指定されたデータセットにツリーテーブルデータセットが出力されます。デフォルトはWORK._TREETABという名前で作成されます。

（分類木モデルの場合）
outtab=出力データセット

変数名	タイプ	長さ	内容	備考
lvl0	文字	可変	ツリーノード階層とノード定義説明変数情報を表す変数	0階層目のノード＝ルートノード
lvl1... lvlk	文字	可変		1階層目～k階層目のノード
N_PCT	数値	8	終端ノードの件数構成比率(%)	ノード件数／総件数
TARG_N_PCT	数値	8	終端ノードのターゲット再現率(%)	ターゲット件数／ターゲット総件数
CONF_PCT	数値	8	終端ノードのターゲット出現率(%)	ターゲット件数／ノード件数
TEST_N_PCT	数値	8	検証データの終端ノードの件数構成比率(%)	test= パラメータを指定した場合に作成される変数
TEST_TARG_N_PCT	数値	8	検証データの終端ノードのターゲット再現率(%)	
TEST_CONF_PCT	数値	8	検証データの終端ノードのターゲット出現率(%)	

（回帰木モデルの場合）
outtab=出力データセット

変数名	タイプ	長さ	内容	備考
lvl0	文字	可変	ツリーノード階層とノード定義説明変数情報を表す変数	0階層目のノード＝ルートノード
lvl1... lvlk	文字	可変		1階層目～k階層目のノード
MEAN	数値	8	終端ノードのターゲット平均値	
STD	数値	8	終端ノードのターゲット標準偏差	
N_PCT	数値	8	終端ノードの件数構成比率(%)	ノード件数／総件数
TEST_MEAN	数値	8	検証データの終端ノードのターゲット平均値	test= パラメータを指定した場合に作成される変数
TEST_STD	数値	8	検証データの終端ノードのターゲット標準偏差	
TEST_N_PCT	数値	8	検証データの終端ノードの件数構成比率(%)	

（分類木アップリフトモデルの場合）
outtab=出力データセット

変数名	タイプ	長さ	内容	備考
lvl0	文字	可変	ツリーノード階層とノード定義説明変数情報を表す変数	0階層目のノード＝ルートノード
lvl1… lvlk	文字	可変		1階層目～k階層目のノード
D_N_PCT	数値	8	終端ノードの件数構成比率(%)[処理群]	
D_TARG_N_PCT	数値	8	終端ノードのターゲット再現率(%)[処理群]	detai=Yの場合のみ出力される
D_CONF_PCT	数値	8	終端ノードのターゲット出現率(%)[処理群]	
C_N_PCT	数値	8	終端ノードの件数構成比率(%)[対照群]	
C_TARG_N_PCT	数値	8	終端ノードのターゲット再現率(%)[対照群]	detai=Yの場合のみ出力される
C_CONF_PCT	数値	8	終端ノードのターゲット出現率(%)[対照群]	
DIF_CONF_PCT	数値	8	終端ノードのターゲット出現率の差([処理群]-[対照群])	
DIF_SE_PCT	数値	8	終端ノードのターゲット出現率の差の標準誤差	detai=Yの場合のみ出力される
TEST_D_N_PCT	数値	8	検証データの終端ノードの件数構成比率(%)[処理群]	test= パラメータを指定した場合に作成される変数(上記と同じ項目名に接頭辞TEST_がつく)(TEST_DIF_CONF_PCT, TEST_D_TARG_N_PCT, TEST_C_TARG_N_PCTは detail=Yの場合のみ出力される)
TEST_D_TARG_N_PCT	数値	8	検証データの終端ノードのターゲット再現率(%)[処理群]	
（途中省略）				
TEST_DIF_SE_PCT	数値	8	検証データの終端ノードのターゲット出現率の差の標準誤差	

（回帰木アップリフトモデルの場合）
outtab=出力データセット

変数名	タイプ	長さ	内容	備考
lvl0	文字	可変	ツリーノード階層とノード定義説明変数情報を表す変数	0階層目のノード＝ルートノード
lvl1… lvlk	文字	可変		1階層目～k階層目のノード
D_N_PCT	数値	8	終端ノードの件数構成比率(%)[処理群]	
D_MEAN	数値	8	終端ノードのターゲット平均値[処理群]	
D_STD	数値	8	終端ノードのターゲット標準偏差[処理群]	detai=Yの場合のみ出力
C_N_PCT	数値	8	終端ノードの件数構成比率(%)[対照群]	
C_MEAN	数値	8	終端ノードのターゲット平均値[対照群]	
C_STD	数値	8	終端ノードのターゲット標準偏差[対照群]	detai=Yの場合のみ出力
DIF_MEAN	数値	8	終端ノードのターゲット平均値の差([処理群]-[対照群])	
DIF_SE	数値	8	終端ノードのターゲット平均値の差の標準誤差	detai=Yの場合のみ出力
TEST_D_N_PCT	数値	8	検証データの終端ノードの件数構成比率(%)[処理群]	test= パラメータを指定した場合に作成される変数(上記と同じ項目名に接頭辞TEST_がつく)(TEST_DIF_SE, TEST_D_STD, TEST_C_STDは detail=Yの場合のみ出力される)
TEST_D_MEAN	数値	8	検証データの終端ノードのターゲット平均値[処理群]	
（途中省略）				
TEST_DIF_SE	数値	8	検証データの終端ノードのターゲット平均値の差の標準誤差	

10.2.8 コマンド実行モードでの注意

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

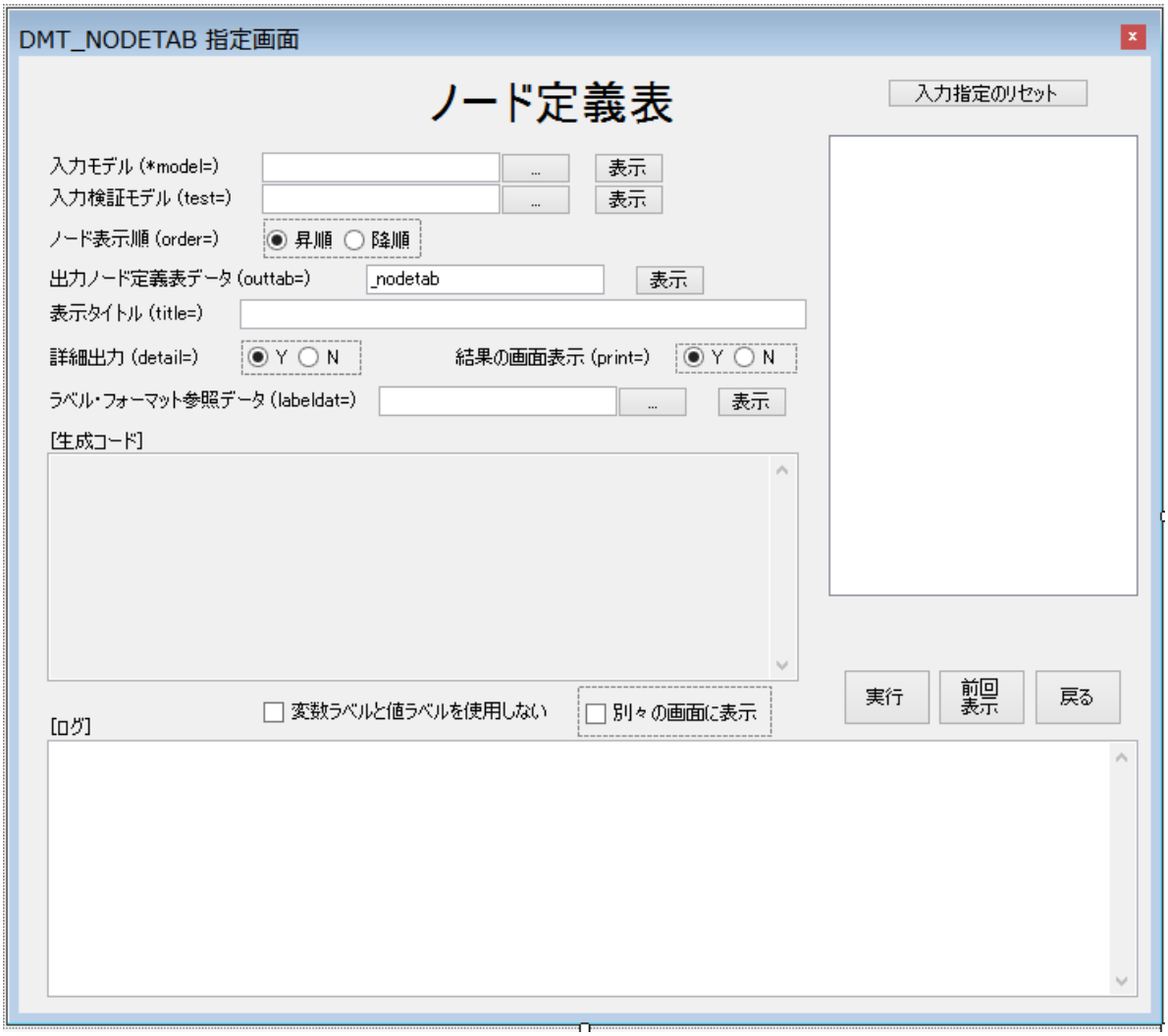
また、以下のユーザ定義フォーマットがWORKライブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
$NODE_C $_item
```

さらに、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
nobs zketa e_name e_type _errormsg
```

10.3 ノード表（dmt_nodetab）



10.3.1 概要

ノード定義表（DMT_NODETAB）は**デシジョンツリーモデル作成**（DMT_TREE）を実行して作成されたモデルデータセット、または**新しいデータを基準にモデル予測値修正**（DMT_TREESCORE）を実行して作成されたモデル形式データセットを入力として、ツリーモデルの各終端ノードを、分類木モデルにおいてはターゲット出現率、回帰木モデルにおいてはターゲット変数の平均値の小さい、または大きい順に並べて、その終端ノードに至るすべての中間ノードを含むノード情報を表形式で画面表示します。

ツリー分岐表（DMT_TREETAB）がモデルのノード分岐過程をそのまま表現するのに対して、DMT_NODETABは終端ノードのターゲット出現率の大きさの順にその終端ノードに至るすべての中間ノードを含む説明変数値の組合せを横一線に見やすい形で表示します。DMT_TREEで作成されたモデルデータセットとそのモデルをDMT_TREESCOREを用いてテストデータに適用して得られたテストデータにおけるモデル形式データセットの**両方**を入力した場合は、各ノードにおけるモデル作成データ、テストデータそれぞれの該当件数、ターゲット件数および、分類木においてはターゲット出現率、回帰木においてはターゲット変数の平均値を各ノードの中に同時表示します。

10.3.2 指定方法

（コマンド実行モードでの指定）

```
%dmt_nodetab(help,model=,test=
,outtab=_nodetab,order=ascending
,print=Y,labeldat=,nolabel=N,detail=N
,title=,pctf=7.2,meanf=best8.
,d_label=[D].c_label=[C],dif_label=[D]-[C]
```

,language=JAPANESE
,outhtml=dmt_nodetab.html,outpath=)

（GUI実行モードでの変更点）

・help, outhtml=, outpath=パラメータは指定不可。(自動で行われます。)
・print=Y に固定。
・labeldat=パラメータは自動入力。

（必須パラメータ）

以下の1個のパラメータは省略できません。

入力モデル (model=)

（オプションパラメータ）

以下の17個のパラメータは任意指定です。(=の右辺の値はデフォルト値を表しています)

help ... 指定方法のヘルプメッセージの表示。（コマンド実行モードでのみ有効）
入力検証モデル (test=)
... モデルをテストデータに適用して得られたモデル形式入力データセット名を指定。
出力ノード定義表データ (outtab=_nodetab)
... ツリーノード定義表画面出力データセットの名前を指定。
ノード表示順の指定 (order=ascending)
... 終端ノードの表示順序をターゲット出現率の小さい順とするか大きい順とするかの選択（ascending/descending）.
結果の画面表示 (print=Y)
... ノードテーブルの画面表示を行うか否かを指定。(Y または N, GUI実行モードではY固定)
変数ラベルと値ラベルを表示しない (nolabel=N)
… 変数ラベルと値ラベルを用いずに変数名、変数値を用いた結果表を作成。
詳細出力 (detail=N)
... 詳細な終端ノード統計量を表示。
画面出力のタイトルの指定 (title=)
.... %str,%nrstr,%bquote などの関数で囲んで指定する（コマンド実行モードでのみ有効）
百分率の表示フォーマットの指定 (pctf=7.2)
平均値・標準偏差の表示フォーマットの指定 (meanf=best8.)
アップリフトモデルにおける処理群(DATA)を表す記号 (d_label=[D])
アップリフトモデルにおける対照群(Control)を表す記号 (c_label=[C])
アップリフトモデルにおける処理群−対照群間の差を表す記号 (dif_label=[D]-[C])
言語の選択 (language=JAPANESE)
HTML出力ファイル名 (outhtml=dmt_nodetab.html)（コマンド実行モードでのみ有効）
HTMLファイル出力ディレクトリの指定 (outpath=)（コマンド実行モードでのみ有効）
ラベル・フォーマット参照データ (labeldat=)（コマンド実行モードでのみ有効）

10.3.3 パラメータの詳細

入力モデル (model=)
入力モデルデータセット名を指定します。このパラメータは省略できません。
例：model=bunseki1

入力検証モデル (test=)
dmt_treescoreを用いてモデルをモデル検証用データセットに適用したときのモデル形式データセットが得られている場合、そのモデル形式データセット名を指定します。この指定により、各ノードごとに、モデルの集計値に加えて検証データにおける集計値も同時表示されます。
例：test=test1

ノード表示順 (order=ascending)
終端ノードの表示順序を指定します。デフォルトはascending（ターゲット出現率またはターゲエット平均値の小さい順）です。descending を指定すると、ターゲット出現率またはターゲット平均値の大きい順に終端ノードが並べられた表になります。

出力ノード定義表データ (outtab=_nodetab)
ノードテーブルデータセットの出力先を指定します。デフォルトは_nodetabです。

表示タイトル(title=)
画面出力される表にタイトルを指定できます。指定しない（デフォルト）場合、以下のようなタイトルが自動的に付与されます。

%bquote(DMT_TREE ノードテーブル(モデルデータセット名: &model) ターゲット出現率の小さい順))

タイトルを指定する場合、上記のように%bquote関数の中に記述してください。

詳細出力 (detail=N, または details=N)
ツリーノード表の表示項目を制御します。デフォルトはdetail=N。detail=Yを指定すると、表示項目数が増えます。

言語 (language=JAPANESE)
分析実行中のメッセージ出力、結果の表のタイトル、表項目などの表示言語を選択します。ただし、現バージョンでは、日本語か英語の2種類のみ選択可能です。
例：language=ENGLISH

結果の画面表示 (print=Y)
ノードテーブルを画面表示する（Y）かしない（N）かを指定します。デフォルトは画面表示する（Y）で

す。print=Nを指定しても、outtab=パラメータに指定したデータセットに画面表示するためのノード情報が出力されます。

変数ラベルと値ラベルを表示しない (nolabel=N)
Yを指定すると、表示が元の変数名、値に変わります。

ラベル・フォーマット参照データ (labeldat=)
ラベルとフォーマットが定義されたデータセットを指定することにより、分析結果の全変数名と文字タイプ変数値に、それぞれ定義された変数ラベルとフォーマットが付加されて表示されます。この指定が無い場合は、変数名、変数値がそのまま表示されます。数値タイプ説明変数には、フォーマットが定義されていたとしても無視されます。なお、フォーマット定義された変数を含むデータセットをアクセスするためには、そのフォーマットライブラリもアクセス可能になっている必要があり、ラベル定義されたデータセットを保存して再利用したい場合は、フォーマットライブラリも保存しておく必要があります。(GUI実行モードではモデルがどのデータから作成されたかを記録しているため、そのデータが存在する場合は自動入力されます)

10.3.4 コマンド実行モードで有効なパラメータの詳細

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。
例：%dmt_nodetab(help)

10.3.5 HTML 出力

分析結果の図表はhtmlファイルに出力されます。保存先はデフォルトではSASディスプレイマネージャまたはWPSワークベンチの管理下（ワークスペース内の一時保存ファイル）です。outpath=パラメータを指定すると、保存先を変更できます。(必ずフルパス指定します。引用符で囲んでも囲まなくてもかまいません）同時にouthtml=パラメータを指定すると、保存するhtmlファイルに自由に名前を付けることができます。

outhtml=dmt_nodetab.html
分析結果を保存するHTML出力ファイル名を指定します。
例：outhtml=out1.html,

outpath=
HTML図表出力ファイルの保存ディレクトリを指定します。このパラメータを指定しない場合（デフォルト)、HTMLファイルはSASディスプレイマネージャまたはWPSワークベンチの管理下に作成されます。outpath=指定を行う場合、値は必ずフルパスで指定する必要があります。なお、パス指定全体を引用符で囲んでも囲まなくてもかまいません。
例：outpath='G:¥temp'

10.3.6 実行例

例１：分類木（検証結果表示なし、変数ラベル、値ラベルなし）

```
%dmt_tree(data=samp_data,y=flg,target=1,x=sei
nenrei,outmodel=flg1)
%dmt_nodetab(model=flg1)
```

DMT_TREE ノードテーブル(モデルデータセット: flg1) ターゲット出現率の小さい順

No.	ノード	LVL1	LVL2	件数割合%	ターゲット再現率%	ターゲット出現率%	累積件数割合%	累積ターゲット再現率%	累積ターゲット出現率%
1	_N10	N1: 20.53%(365/1,778) NENREI=23<~HIGH	N10: 18.09% (214/1,183) SEI="1"	59.15	46.83	18.09	59.15	46.83	18.09
2	_N11	N1: 20.53%(365/1,778) NENREI=23<~HIGH	N11: 25.38% (151/595) SEI="2"	29.75	33.04	25.38	88.90	79.87	20.53
3	_N0	N0: 41.44%(92/222) NENREI=LOW~23		11.10	20.13	41.44	100.00	100.00	22.85

デフォルトでは、出現率の昇順に終端ノードが並べられて表示されます。
LVL1～LVL2の各セルの中には、ツリー分岐表の各ノードと同じく、ツリーノードの識別番号：各ノードのターゲット出現率%（ターゲット件数／件数）そして、 親ノードから分岐する条件を表す 説明変数=値（値の範囲）が表示されます。
ノードテーブルを作成することにより、各終端ノードの特徴（説明変数の値の組合せ）が把握しやすくなります。

例２：回帰木（変数ラベル、値ラベルあり、検証結果表示あり、数値の表示フォーマット指定あり、ノードはモデル予測値（平均値）の降順に並べる）

```
%dmt_tree(data=samp_data,y=nenshu,x=sei
nenrei,outmodel=nenshu1,maxlvl=2)
%dmt_treescore(model=nenshu1,data=test_data,y=
nenshu,outmodel=TEST_nenshu1)
```

```
%dmt_nodetab(model=nenshu1,test=TEST_nenshu
1,labeldat=samp_data,
pctf=3.,meanf=6.1,order=descending)
```

注：GUI実行モードではlabeldat=パラメータは自動設定されます。

DMT_TREE ノードテーブル (モデル: nenshu1, テスト: TEST_nenshu1 の比較)
ターゲット平均値の大きい順

No.	終端ノード	lvl1	lvl2	件数割合%	ターゲット平均値	ターゲット標準偏差	累積件数割合%	累積ターゲット平均値	累積ターゲット標準偏差	テスト件数割合%	テストターゲット平均値	テストターゲット標準偏差	テスト累積件数割合%	テスト累積ターゲット平均値	テスト累積ターゲット標準偏差
1	_N10	N1: 526.7 (N=547,S=204.8): 515.5 (N=551,S=190.3) SEI 性別="2 女性"	N10: 544.2 (N=363,S=214.7): 523.5 (N=371,S=193.9) NENREI 年齢=23~48	25	544.2	214.7	25	544.2	214.7	27	523.5	193.9	27	523.5	193.9
2	_N01	N0: 506.3 (N=898,S=201.1): 504.5 (N=841,S=203.0) SEI 性別="1 男性"	N01: 524.4 (N=337,S=215.8): 515.3 (N=304,S=208.5) NENREI 年齢 =LOW~<33,58<~HIGH	23	524.4	215.8	48	534.7	215.5	22	515.3	208.5	48	519.8	200.6
3	_N00	N0: 506.3 (N=898,S=201.1): 504.5 (N=841,S=203.0) SEI 性別="1 男性"	N00: 495.5 (N=561,S=190.9): 498.3 (N=537,S=199.5) NENREI 年齢=33~58	39	495.5	190.9	87	517.2	205.8	39	498.3	199.5	87	510.3	200.4
4	_N11	N1: 526.7 (N=547,S=204.8): 515.5 (N=551,S=190.3) SEI 性別="2 女性"	N11: 492.1 (N=184,S=178.7): 499.0 (N=180,S=181.6) NENREI 年齢 =LOW~<23,48<~HIGH	13	492.1	178.7	100	514.0	202.7	13	499.0	181.6	100	508.8	198.1

各ノードセル内にはモデル情報（平均値（件数、標準偏差））に続いて：（コロン）の後に検証データにモデルの分岐条件を当てはめたときの情報（平均値（件数、標準偏差））が追加されて表示されます。

例３：詳細表示

```
%dmt_tree(data=SAMP_DATA(where=(DM="1")),co
ntrol=SAMP_DATA(where=(DM="0")),y=kingaku
,x=sei nenrei jukyo,outmodel=kingaku_uplift
,mincnt=100,maxlvl=5)
%dmt_nodetab(model=kingaku_uplift)
```

```
%dmt_nodetab(model=kingaku_uplift,detail=Y)
```

注：DMT_TREETABマクロの detail=Yオプションは、アップリフトモデルの表示の場合のみ有効です。

DMT_TREE ノードテーブル(モデルデータセット: kingaku_uplift) ターゲット平均値の差の小さい順

No.	ノード	LVL1	LVL2	[D]-[C]ターゲット平均値の差	[D]件数割合%	[D]ターゲット平均値	[C]件数割合%	[C]ターゲット平均値	[D]-[C]累積ターゲット平均値の差	[D]累積件数割合%	[D]累積ターゲット平均値	[C]累積件数割合%	[C]累積ターゲット平均値
1	_N00	N0: [D]-[C]-48.3264(SE=299.0199),[D] 56.10756(N=344,S=171.8592),[C] 104.434(N=947,S=244.6984) SEI="1"	N00: [D]-[C]-75.6604(SE=358.6228),[D] 82.765(N=200,S=214.8141),[C]158.4254 (N=623,S=287.1675) JUKYO="3","","4","5","7"	-75.6604	32.31	82.765	45.11	158.4254	-75.6604	32.31	82.765	45.11	158.4254
2	_N01	N0: [D]-[C]-48.3264(SE=299.0199),[D] 56.10756(N=344,S=171.8592),[C] 104.434(N=947,S=244.6984) SEI="1"	N01: [D]-[C]18.46605(SE=64.5786),[D] 19.08333(N=144,S=64.10186),[C]0.617284 (N=324,S=7.832455) JUKYO="2","1","6"	18.46605	23.26	19.08333	23.46	0.617284	-48.3265	55.57	56.10756	68.57	104.434
3	_N10	N1: [D]-[C]130.5538(SE=386.6487),[D] 221.5055(N=275,S=311.1448),[C] 90.95161(N=434,S=229.5346) SEI="2"	N10: [D]-[C]22.55774(SE=99.96399),[D] 24.82857(N=105,S=93.59675),[C]2.270833 (N=240,S=35.10623) JUKYO="1","2","6","7"	22.55774	16.96	24.82857	17.38	2.270833	-34.9847	72.54	48.79287	85.95	83.77761
4	_N11	N1: [D]-[C]130.5538(SE=386.6487),[D] 221.5055(N=275,S=311.1448),[C] 90.95161(N=434,S=229.5346) SEI="2"	N11: [D]-[C]142.3226(SE=455.1069),[D] 342.9824(N=170,S=335.4804),[C]200.6598 (N=194,S=307.5309) JUKYO="5","","3","4"	142.3226	27.46	342.9824	14.05	200.6598	29.39108	100.00	129.5881	100.00	100.197

DMT_TREE ノードテーブル(モデルデータセット: kingaku_uplift) ターゲット平均値の差の小さい順

No.	ノード	LVL1	LVL2	[D]-[C]ターゲット平均値の差	[D]-[C]ターゲット平均値の差の標準誤差	[D]件数割合%	[D]ターゲット平均値	[D]ターゲット標準偏差	[C]件数割合%	[C]ターゲット平均値	[C]ターゲット標準偏差	[D]-[C]累積ターゲット平均値の差	[D]-[C]累積ターゲット平均値の差の標準誤差	[D]累積件数割合%	[D]累積ターゲット平均値	[D]累積ターゲット標準偏差	[C]累積件数割合%	[C]累積ターゲット平均値	[C]累積ターゲット標準偏差
1	_N00	N0: [D]-[C]-48.3264 (SE=299.0199),[D] 56.10756 (N=344,S=171.8592), [C]104.434 (N=947,S=244.6984) SEI="1"	N00: [D]-[C]-75.6604 (SE=358.6228),[D] 82.765 (N=200,S=214.8141), [C]158.4254 (N=623,S=287.1675) JUKYO="3","","4","5","7"	-75.6604	358.6228	32.31	82.765	214.8141	45.11	158.4254	287.1675	-75.6604	358.6228	32.31	82.765	214.8141	45.11	158.4254	287.1675
2	_N01	N0: [D]-[C]-48.3264 (SE=299.0199),[D] 56.10756 (N=344,S=171.8592), [C]104.434 (N=947,S=244.6984) SEI="1"	N01: [D]-[C]18.46605 (SE=64.5786),[D] 19.08333 (N=144,S=64.10186), [C]0.617284 (N=324,S=7.832455) JUKYO="2","1","6"	18.46605	64.5786	23.26	19.08333	64.10186	23.46	0.617284	7.832455	-48.3265	299.0199	55.57	56.10756	171.8592	68.57	104.434	244.6984
3	_N10	N1: [D]-[C]130.5538 (SE=386.6487),[D] 221.5055 (N=275,S=311.1448), [C]90.95161 (N=434,S=229.5346) SEI="2"	N10: [D]-[C]22.55774 (SE=99.96399),[D] 24.82857 (N=105,S=93.59675), [C]2.270833 (N=240,S=35.10623) JUKYO="1","2","6","7"	22.55774	99.96399	16.96	24.82857	93.59675	17.38	2.270833	35.10623	-34.9847	273.0495	72.54	48.79287	157.6468	85.95	83.77761	222.9429
4	_N11	N1: [D]-[C]130.5538 (SE=386.6487),[D] 221.5055 (N=275,S=311.1448), [C]90.95161 (N=434,S=229.5346) SEI="2"	N11: [D]-[C]142.3226 (SE=455.1069),[D] 342.9824 (N=170,S=335.4804), [C]200.6598 (N=194,S=307.5309) JUKYO="5","","3","4"	142.3226	455.1069	27.46	342.9824	335.4804	14.05	200.6598	307.5309	29.39108	351.9012	100.00	129.5881	257.2506	100.00	100.197	240.1178

は WORK._NODETAB という名前です。

10.3.7 データセット出力

outtab=パラメータに指定したデータセットに画面出力イメージをデータセット出力します。デフォルト

（分類木モデルの場合）
outtab=出力データセット

変数名	タイプ	長さ	内容	備考
no	数値	8	順序	ターゲット出現率の小さい／大きい順
termnode	文字	可変	終端ノード番号	″ Nxxxx″の値。ただしxxxxは0/1の文字列
lvl1… lvlk	文字	可変		1階層目～k階層目のノード
N_PCT	数値	8	件数構成比率	
TARG N_PCT	数値	8	ターゲット再現率	
CONF_PCT	数値	8	ターゲット出現率	
CUM_N_PCT	数値	8	累積件数構成比率	
CUM_TARG_PCT	数値	8	累積ターゲット再現率	
CUM_CONF_PCT	数値	8	累積ターゲット出現率	
TEST_N_PCT	数値	8	検証データの件数構成比率	test= パラメータを指定した場合に作成される変数
TEST_TARG_N_PCT	数値	8	検証データのターゲット再現率	
TEST_CONF_PCT	数値	8	検証データのターゲット出現率	
TEST_CUM_N_PCT	数値	8	検証データの累積件数構成比率	
TEST_CUM_TARG_PCT	数値	8	検証データの累積ターゲット再現率	
TEST_CUM_CONF_PCT	数値	8	検証データの累積ターゲット出現率	

（回帰木モデルの場合）
outtab=出力データセット

変数名	タイプ	長さ	内容	備考
no	数値	8	順序	ターゲット出現率の小さい／大きい順
termnode	文字	可変	終端ノード番号	″ Nxxxx″の値。ただしxxxxは0/1の文字列
lvl1… lvlk	文字	可変		1階層目～k階層目のノード
N_PCT	数値	8	件数構成比率	
MEAN	数値	8	ターゲット平均値	
STD	数値	8	ターゲット標準偏差	
CUM_N_PCT	数値	8	累積件数構成比率	
CUM_MEAN	数値	8	累積ターゲット平均値	
CUM_STD	数値	8	累積ターゲット標準偏差	
TEST_N_PCT	数値	8	検証データの件数構成比率	test= パラメータを指定した場合に作成される変数
TEST_MEAN	数値	8	検証データのターゲット平均値	
TEST_STD	数値	8	検証データのターゲット標準偏差	
TEST_CUM_N_PCT	数値	8	検証データの累積件数構成比率	
TEST_CUM_MEAN	数値	8	検証データの累積ターゲット平均値	
TEST_CUM_STD	数値	8	検証データの累積ターゲット標準偏差	

（分類木アップリフトモデルの場合）
outtab=出力データセット

変数名	タイプ	長さ	内容	備考
no	数値	8	順序	ターゲット出現率の小さい／大きい順
termnode	文字	可変	終端ノード番号	"_Nxxxx"の値。ただしxxxxは0/1の文字列
lvl1… lvlk	文字	可変		1階層目～k階層目のノード
D_TOT_N	数値	8	総件数[処理群]	
D_TARG_N	数値	8	総ターゲット件数[処理群]	
C_TOT_N	数値	8	総件数[対照群]	
C_TARG_N	数値	8	総ターゲット件数[対照群]	
D_N	数値	8	件数[処理群]	
D_N_PCT	数値	8	件数構成比率[処理群]	表示項目
D_TARG_N	数値	8	ターゲット件数[処理群]	
D_TARG_N_PCT	数値	8	ターゲット再現率[処理群]	表示項目(detai=Yの場合のみ)
D_CONF_PCT	数値	8	ターゲット出現率[処理群]	表示項目
C_N	数値	8	件数[対照群]	
C_N_PCT	数値	8	件数構成比率[対照群]	表示項目
C_TARG_N_PCT	数値	8	ターゲット再現率[対照群]	表示項目(detai=Yの場合のみ)
C_CONF_PCT	数値	8	ターゲット出現率[対照群]	表示項目
DIF_CONF_PCT	数値	8	ターゲット出現率の差([処理群]-[対照群])	表示項目
DIF_SE_PCT	数値	8	ターゲット出現率の差の標準誤差	表示項目(detai=Yの場合のみ)
D_CUM_N	数値	8	累積件数[処理群]	
D_CUM_N_PCT	数値	8	累積件数構成比率[処理群]	表示項目
D_CUM_TARG_N	数値	8	累積ターゲット件数[処理群]	
D_CUM_TARG_N_PCT	数値	8	累積ターゲット再現率[処理群]	表示項目(detai=Yの場合のみ)
D_CUM_CONF_PCT	数値	8	累積ターゲット出現率[処理群]	表示項目
C_CUM_N	数値	8	累積件数[対照群]	
C_CUM_N_PCT	数値	8	累積件数構成比率[対照群]	表示項目
C_CUM_TARG_N	数値	8	累積ターゲット件数[対照群]	
C_CUM_TARG_N_PCT	数値	8	累積ターゲット再現率[対照群]	表示項目(detai=Yの場合のみ)
C_CUM_CONF_PCT	数値	8	累積ターゲット出現率[対照群]	表示項目
CUM_DIF_CONF_PCT	数値	8	累積ターゲット出現率の差([処理群]-[対照群])	表示項目(detai=Yの場合のみ)
CUM_DIF_SE_PCT	数値	8	累積ターゲット出現率の差の標準誤差	表示項目
TEST_D_TOT_N	数値	8	検証データの総件数[処理群]	test= パラメータを指定した場合に作成される変数(上記と同じ項目名に接頭辞 TEST_がつく)
TEST_D_TARG_N	数値	8	検証データの総ターゲット件数[処理群]	
（途中省略）				
TEST_CUM_DIF_SE_PCT	数値	8	検証データの累積ターゲット出現率の差の標準誤差	

（回帰木アップリフトモデルの場合）
outtab=出力データセット

変数名	タイプ	長さ	内容	備考
no	数値	8	順序	ターゲット出現率の小さい／大きい順
termnode	文字	可変	終端ノード番号	"_Nxxxx"の値。ただしxxxxは0/1の文字列
lvl1… lvlk	文字	可変		1階層目～k階層目のノード
D_TOT_N	数値	8	総件数[処理群]	
C_TOT_N	数値	8	総件数[対照群]	
D_N	数値	8	件数[処理群]	
D_N_PCT	数値	8	件数構成比率[処理群]	表示項目
D_MEAN	数値	8	ターゲット平均値[処理群]	表示項目
D_STD	数値	8	ターゲット標準偏差[処理群]	表示項目(detai=Yの場合のみ)
C_N	数値	8	件数[対照群]	
C_N_PCT	数値	8	件数構成比率[対照群]	表示項目
C_MEAN	数値	8	ターゲット平均値[対照群]	表示項目
C_STD	数値	8	ターゲット標準偏差[対照群]	表示項目(detai=Yの場合のみ)
DIF_MEAN	数値	8	終端ノードのターゲット平均値の差([処理群]-[対照群])	表示項目
DIF_SE	数値	8	終端ノードのターゲット平均値の差の標準誤差	表示項目(detai=Yの場合のみ)
D_CUM_N	数値	8	累積件数[処理群]	
D_CUM_N_PCT	数値	8	累積件数構成比率[処理群]	表示項目
D_CUM_MEAN	数値	8	累積ターゲット平均値[処理群]	表示項目
D_CUM_STD	数値	8	累積ターゲット標準偏差[処理群]	表示項目(detai=Yの場合のみ)
C_CUM_N	数値	8	累積件数[対照群]	
C_CUM_N_PCT	数値	8	累積件数構成比率[対照群]	表示項目
C_CUM_MEAN	数値	8	累積ターゲット平均値[対照群]	表示項目
C_CUM_STD	数値	8	累積ターゲット標準偏差[対照群]	表示項目(detai=Yの場合のみ)
CUM_DIF_MEAN	数値	8	終端ノードのターゲット平均値の差([処理群]-[対照群])	表示項目
CUM_DIF_SE	数値	8	終端ノードのターゲット平均値の差の標準誤差	表示項目(detai=Yの場合のみ)
TEST_D_TOT_N	数値	8	検証データの総件数[処理群]	test= パラメータを指定した場合に作成される変数(上記と同じ項目名に接頭辞 TEST_がつく)
TEST_C_TOT_N	数値	8	検証データの総件数[対照群]	
（途中省略）				
TEST_CUM_DIF_SE	数値	8	検証データの終端ノードのターゲット平均値の差の標準誤差	

10.3.8 コマンド実行モードでの注意

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

また、以下のユーザ定義フォーマットがWORKライ

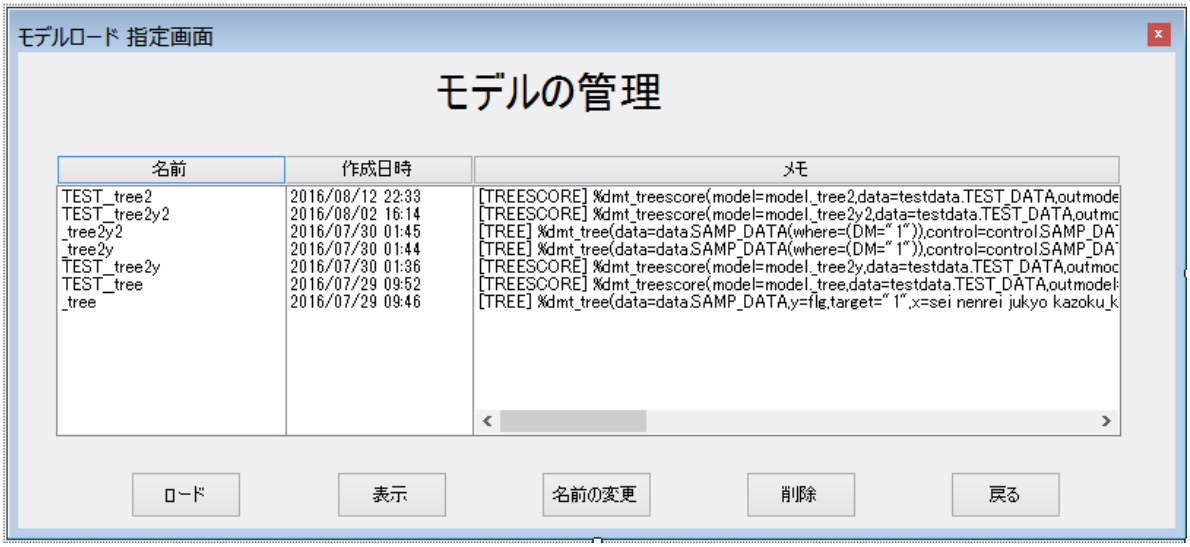
ブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
$NODE_C $NODE_D $_ORDER $_item
```

さらに、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
nobs zketa e_name e_type _errormsg
```

10.4 モデルの管理



10.4.1 概要

「ツリーモデル作成」、「ツリーの枝刈り」、「ツリーの枝接ぎ」、「予測値修正」の各画面で作成したツリーモデルを操作（表示・名前の変更・削除）します。
この機能はマクロモジュールには含まれていません。
GUI実行モードでのみ指定可能です。

メモ欄の最初の鍵カッコは以下の画面で作成されたことを表します。

[TREE] ... ツリー作成
[TREESCORE] ... モデル予測値修正
[TREECUT] ... ツリーの枝刈り
[TREEADD] ,,, ツリーの枝接ぎ

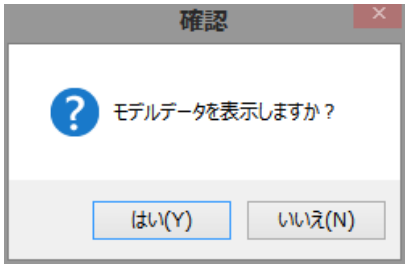
続いてデータを作成したときに実行したプログラムが記述されています。

10.4.2 操作方法

名前,
作成日時,
メモ
リストの上にあるバーをクリックすると、データセットリストを各項目の昇順・または降順で並べ替えることができます。

操作したいモデル名をクリックすると、操作ボタンが表示されますので、表示・名前の変更・削除の操作を行います。

表示 分析結果データの内容を表示します。



名前の変更 データの名前とメモ内容を確認・変更します。

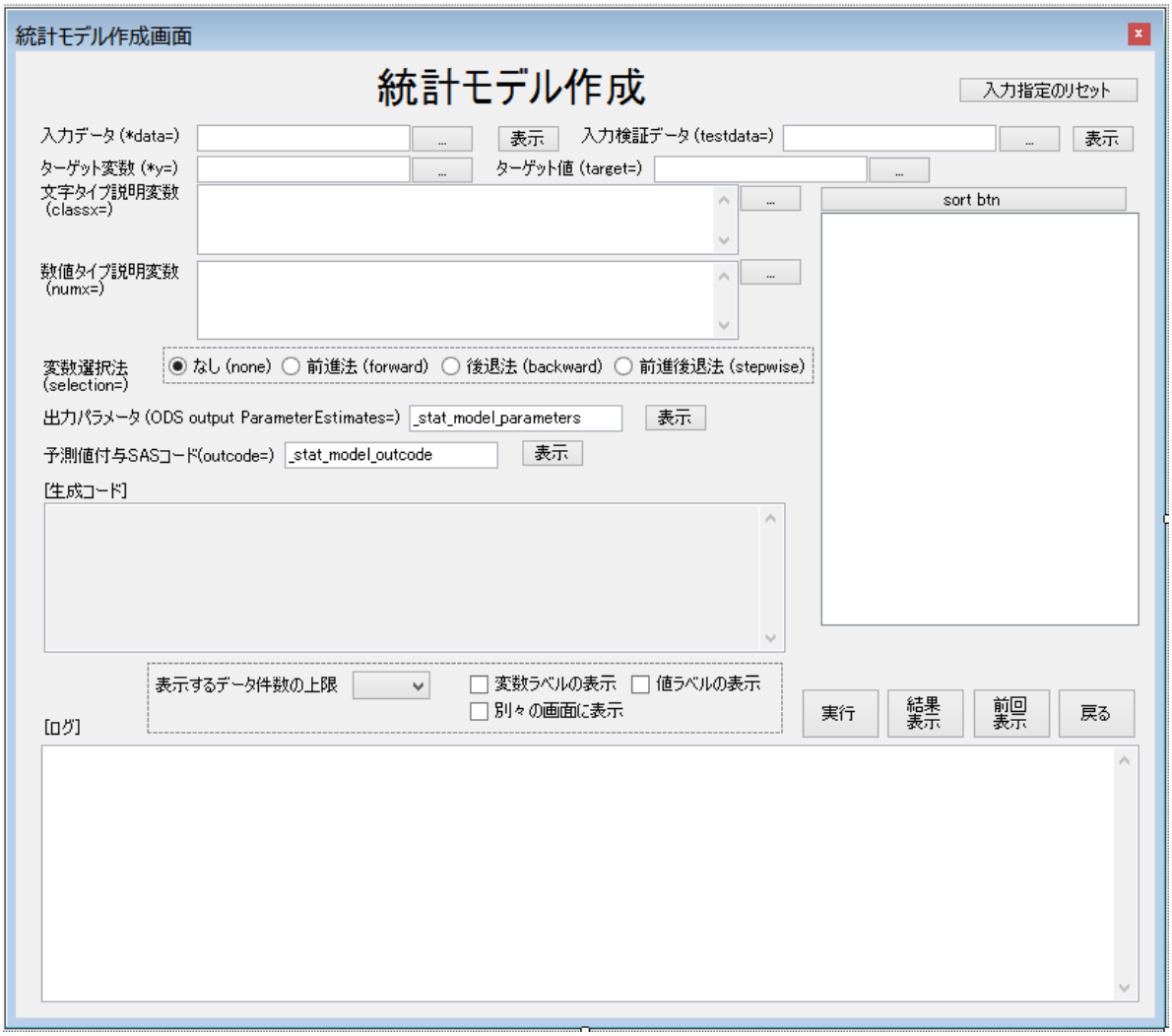
名前は半角英数字で22文字以内（TEST_の接頭辞や_CV10などの接尾辞が自動的に付けられる可能性があるため）に設定してください。（先頭はアルファベットまたは_(アンダーバー)）
なお、名前の変更は、元の名前を参照している他の項目（モデル作成画面の入力パラメータ値など）とは自動連動しません。そのため、再指定が必要になるなどの影響があります。

削除 データを削除します。

削除すると、元に戻せません。

(TIPS) 多数のモデルを関連ファイルと一緒にまとめて削除したい場合は、**「設定画面」**の**「分析ディレクトリ」**の下の**「ツリーモデルディレクトリ」「表示」**ボタンを押し、起動するWindowsエクスプローラで行うと便利です。削除したいデータセット名が書かれたディレクトリをすべて同時選択してから削除します。

10.5 統計モデル（stat_model）



10.5.1 概要

統計モデル作成（STAT_MODEL）はデータに統計モデルを適用するための画面です。ターゲット変数が数値タイプでかつターゲット値が与えられない場合は線形回帰分析、ターゲット変数が文字タイプの場合、もしくは数値タイプであってもターゲット値が与えられた場合は線形ロジスティック分析がモデル構築に用いられます。

いずれの場合も切片項の有無と説明変数選択および変数選択における有意確率基準をオプション指定できます。

パラメータ推計結果データセットと、予測値を付与するためのSASコードファイルが出力されます。このSASコードはコピーして「データ加工」画面の「変数生成・変換・条件抽出SASステートメント」欄に張り付けることにより、予測値を付与することに利用できます。

また、結果表示ボタンを押すと、分析結果リスト、ゲインチャート（ターゲット値が与えられた場合のみ)、比較プロットを表示できます。

統計モデルを採用するのに適切な状況は一般的に以下のとおりです。

・データ件数が数千件以下と少ない場合
・ターゲット変数の変動に関して、説明変数間の交互作用効果が少ないと考えられる場合
・説明変数とターゲット変数の変動の間に強い線形性（比例性）が認められる場合

注意： SASバージョン9.2以降またはWPSバージョン3.01以降で動作します。

回帰分析モデルでは SASではGLMSELECTプロシ

ジャ、WPSはGLMMODプロシジャ＋REGプロシジャを組み合わせて実行します。
ロジスティックモデルでは LOGISTICプロシジャが実行されます。

変数選択法は、回帰分析モデルではSASバージョン9.2以降の場合はSBC基準の変数単位の選択、それ以外は、文字タイプ説明変数をすべて値ごとにダミー変数化してから変数選択しているため、文字タイプ説明変数は値単位、数値説明変数は変数単位の変数選択結果となります。一方、ロジスティックモデルではいずれも変数単位の変数選択が行われます。

10.5.2 指定方法

この機能はマクロモジュールに含まれていません。
GUI実行モードでのみ指定可能です。

（必須パラメータ）

以下の5個のパラメータは省略できません。
ただし、回帰モデルを作成する場合はターゲット値(target=)は指定してはいけません。
また、文字タイプ説明変数(classx=)と数値タイプ説明変数(numx=)の指定は、いずれか一方の指定があれば他方の指定は必須ではありません。

入力データ (data=) ... 入力データセット名の指定.
ターゲット変数(y=) ... ターゲット変数名の指定.
(単一変数名のみ指定可)
ターゲット値 (target=) ... ターゲット値の指定.（単一値のみ指定可、ただし数値タイプの場合のみ、あるしきい値以上または以下または超または未満を指定可）回帰モデルを作成する場合はターゲット値は指定してはいけません。
文字タイプ説明変数 (classx=) ... 文字タイプ説明変数リストの指定. (例: a b c) x1-x4 a--z f_: などの省略指定は指定不可.
数値タイプ説明変数 (numx=) ... 数値タイプ説明変数リストの指定. (例: a b c) x1-x4 a--z f_: などの省略指定は指定不可.

（オプションパラメータ）

以下の8個のパラメータは任意指定です。(=の右辺の値はデフォルト値を表しています)

入力検証データ (testdata=)
... モデル検証用データを指定します。
切片項 (intercept)
... モデルの切片項パラメータの有無を指定します.（「あり」がデフォルト）(※オプション画面で変更可能)
ロジスティックモデルの最大反復計算回数
(maxiter=100) ... ロジスティックモデルの最尤法計算の反復回数を指定します. (※オプション画面で変更可能)
変数選択法 (selection=NONE) ... 変数選択法を選択します.
(selection=NONE/FORWARD/BACKWARD/STEPWISE) ただし、SAS9.2以降の回帰分析モデルでは、SBC（シュワルツの情報量基準）による変数選択法のみ使用できます。
説明変数をモデルに入れるときの有意確率基準
(slentry=0.15)
... 説明変数をモデルに加えるときの有意確率基準. 変数選択法を指定した場合に有効. (※オプション画面で変更可能。SAS9.2以降の回帰分析モデルでは無効)
説明変数をモデルから除くときの有意確率基準
(slstay=0.15)
... 説明変数をモデルから除外するときの有意確率基準. 変数選択法を指定した場合に有効. (※オプション画面で変更可能。SAS9.2以降の回帰分析モデルでは無効)
出力パラメータ (ODS output ParameterEstimates=)
... パラメータ推計結果リストを出力するデータセット名を指定.
予測値付与SASコード(outcode=) ... モデル予測値を計算するSASステートメントを出力するファイル名を指定.

10.5.3 パラメータの詳細

入力データ(data=)
入力データセット名を指定します。このパラメータは省略できません。例：data=a

入力検証データ (testdata=)
モデル検証用データを指定します。指定された場合は、「結果表示」ボタンを押した際に作成したモデルがモデル作成データと検証データ（指定があれば）に適用され、ゲインチャート（ロジスティックモデルの場合のみ）と比較プロット表示に用いられます。

ターゲット変数 (y=)
ターゲット変数名を指定します。 このパラメータは省略できません。例：y=flag

ターゲット値 (target=)
ロジスティックモデルを作成したい場合、ターゲット変数のターゲット値を指定します。ターゲット変数が文字タイプの場合にはこのパラメータは省略できません。ターゲット変数が数値タイプでターゲット変数の大きさを予測する回帰モデルを作成する場合は指定してはいけません。

ターゲット変数が文字タイプの場合は1種類の値を指定します。特殊な文字（+,-など）を含まない限り引用符で囲む必要はありません。(GUI画面からの選択を行うと自動的に複引用符で値が囲まれます) ターゲット変数が数値タイプの場合は1種類の値、もしくはあるしきい値を境とした「以上」、「以下」、「超」、「未満」のいずれかの範囲を指定可能です。数値変

数タイプで範囲を指定する場合は引用符で囲んではいけません。

例1 : y=flag,target=A （ターゲット変数が文字タイプ変数で、その値"A"をターゲットに指定する場合）
例2 : y=sales,target=1000 （ターゲット変数が数値タイプで、その値1000をターゲットに指定する場合）
例3 : y=sales,target=>1000 （ターゲット変数が数値タイプで、その値1000超をターゲットに指定する場合）
例4 : y=sales,target=>=1000 （ターゲット変数が数値タイプで、その値1000以上をターゲットに指定する場合。 target==>1000と指定してもかまいません。）
例5 : y=sales,target=<1000 （ターゲット変数が数値タイプで、その値1000未満をターゲットに指定する場合）
例6 : y=sales,target=<=1000 （ターゲット変数が数値タイプで、その値1000以下をターゲットに指定する場合。 target==<1000と指定してもかまいません。）

注：文字タイプ変数のターゲット値は、大文字、小文字が区別される点に注意してください。（変数名は大文字・小文字の区別はありません。）

文字タイプ説明変数 (classx=)
文字タイプの説明変数を指定します。このパラメータと数値タイプ説明変数の指定のいずれかは省略できません。間に1個以上のスペースを入れて、複数の説明変数を指定可能です。なお、省略指定（-,--,:）と3つの特殊指定
（_ALL_,_NUMERIC_,_CHARACTER_） はサポートされていません。

数値タイプ説明変数 (numx=)
数値タイプの説明変数を指定します。このパラメータと文字タイプ説明変数の指定のいずれかは省略できません。間に1個以上のスペースを入れて、複数の説明変数を指定可能です。なお、省略指定（-,--,:）と3つの特殊指定
（_ALL_,_NUMERIC_,_CHARACTER_） はサポートされていません。

切片項 (intercept)
モデルに切片項パラメータを含むか否かを指定します.（含む（「あり」）がデフォルト）

ロジスティックモデルの最大反復計算回数 (maxiter=100)
最尤法によるパラメータ推計時の最大反復計算回数を指定します。反復回数が十分で無い場合、最尤法によるパラメータ推計は収束に至らない場合があります。変数選択を指定した場合は、
各変数選択段階でのパラメータが収束しないまま、次の変数選択段階に進む場合があります。このような場合、このオプションの値を大きくするとパラメータ推計結果が収束する場合があります。

変数選択法 (selection=NONE)
変数選択は予測誤差の小さいモデルを作る効果があります。回帰モデル、ロジスティックモデル共に、以下の3つの選択法を指定可能です。

前進法（forward）切片項のみ含むモデルから開始し、モデルに含まれていない説明変数の中でモデルに追加するための有意確率基準（slentry）を満たす中から最も説明力の高い説明変数を逐次的にモデルに追加していく方法。
モデルに追加するための有意確率基準（slentry）を満たすとは、slentry値以下の有意確率を意味し、基準を満たす説明変数が1個も存在しないときモデル構築は終了します。

後退法（backward）指定した全説明変数を含むモデルから開始し、モデルに含まれている説明変数の中でモデルに残るための有意確率基準（slstay）を満たさない中から最も説明力の低い説明変数を探して逐次的にモデルから削除していく方法。
モデルに残るための有意確率基準（slstay）を満たさないとは、slstay値超の有意確率を意味し、基準を満たす説明変数が1個も存在しないときモデル構築は終了します。

前進後退法（stepwise）切片項のみ含むモデルから開始し、モデルに含まれていない説明変数でモデルに追加するための有意確率基準（slentry）を満たす中から最も説明力の高い説明変数をモデルに追加し、追加した時点でモデルに含まれている説明変数の中でモデルに残るための有意確率基準（slstay）を満たさない説明力の低い説明変数が存在すればモデルから削除していく方法。追加と削除を交互にチェックすることからこの名前がつけられています。削除と追加が連続して発生しないときモデル構築は終了します。

説明変数をモデルに入れるときの有意確率基準 (slentry=0.15)
モデルに含まれていない説明変数の中からモデルに追加するときの有意確率基準を指定します。

説明変数をモデルから除くときの有意確率基準 (slstay=0.15)
モデルに含まれている説明変数の中でモデルから除くための有意確率基準を指定します。なお、slstayはモデルに残るための基準という意味です。

出力パラメータ(ODS output ParameterEstimates=)
パラメータ推計結果リストを指定の名前でデータセット出力します。

予測値付与SASコード (outcode=)
モデル予測値を計算するSASステートメントを出力

するファイル名を指定します。
このコードはコピーして「データ加工」画面の「変数生成・変換・条件抽出SASステートメント」欄に張り付けることにより、予測値を付与することに利用できます。

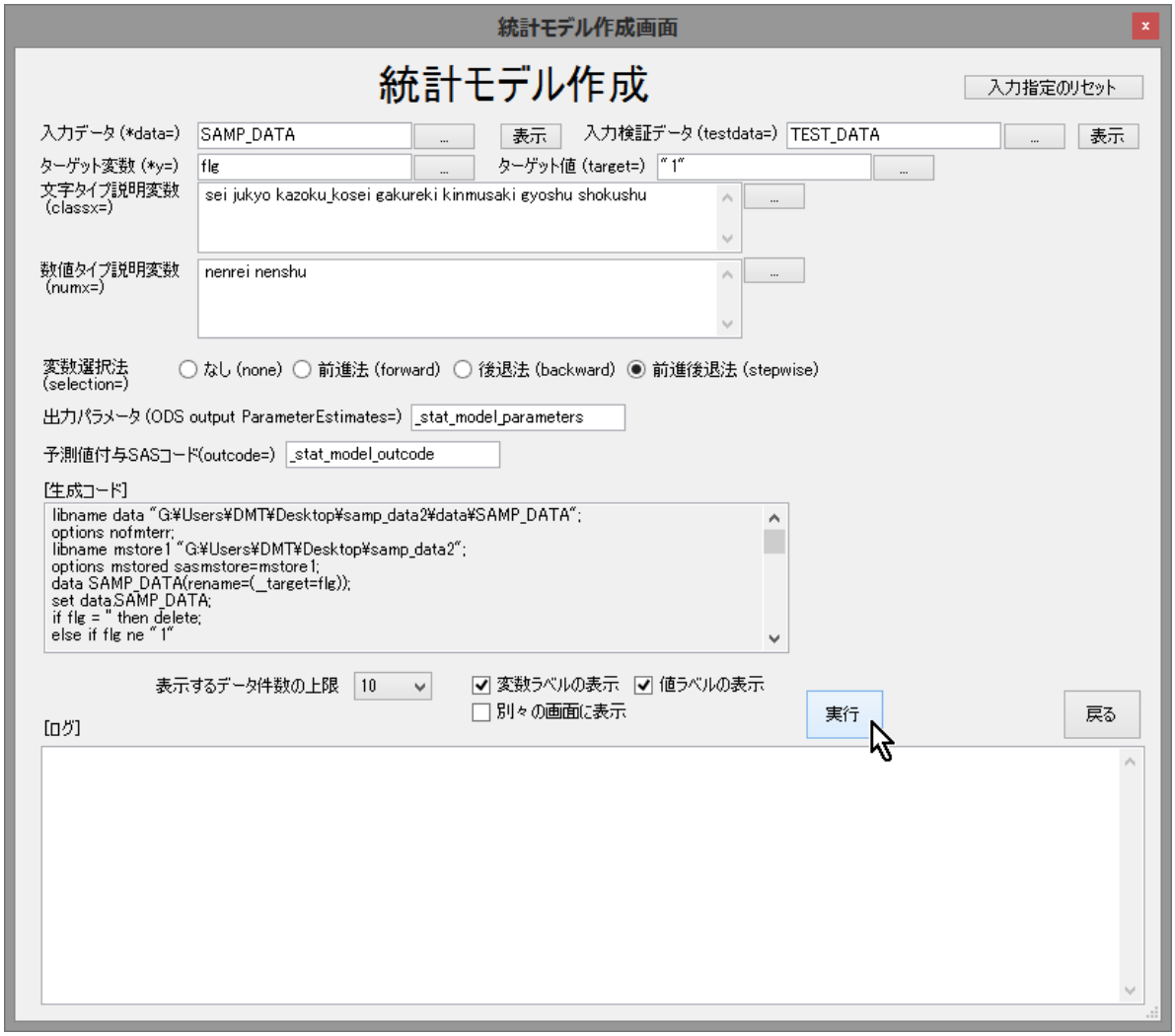
10.5.4 実行例

「結果表示」ボタンを押すと分析結果が表示されます。

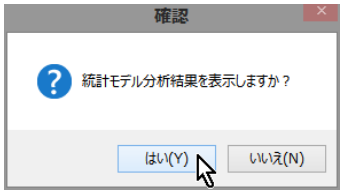
・分析結果アウトプット
SASの分析出力結果を表示します。
・ゲインチャート（ロジスティックモデルの場合のみ）
① モデル作成データのゲインチャート
② モデル検証データのゲインチャート

・比較プロット
実績値（横軸）対 予測値（縦軸）の散布図です。①モデル作成データの散布図（最大5000オブザベーション）
② モデル作成データ 予測値の大きさの順に10グループ化した後の実績平均 対 予測平均 の散布図
③ モデル検証データの散布図（最大5000オブザベーション）
④ モデル検証データ 予測値の大きさの順に10グループ化した後の実績平均 対 予測平均 の散布図

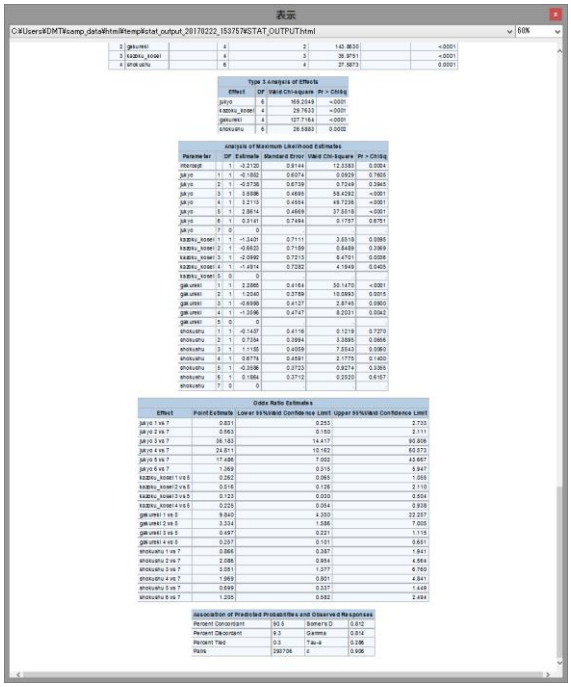
（ロジスティックモデルの実行例）



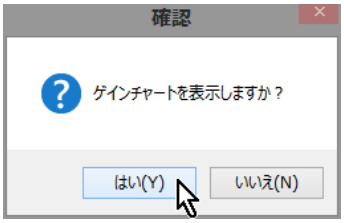
実行終了後、 結果表示 をクリック。

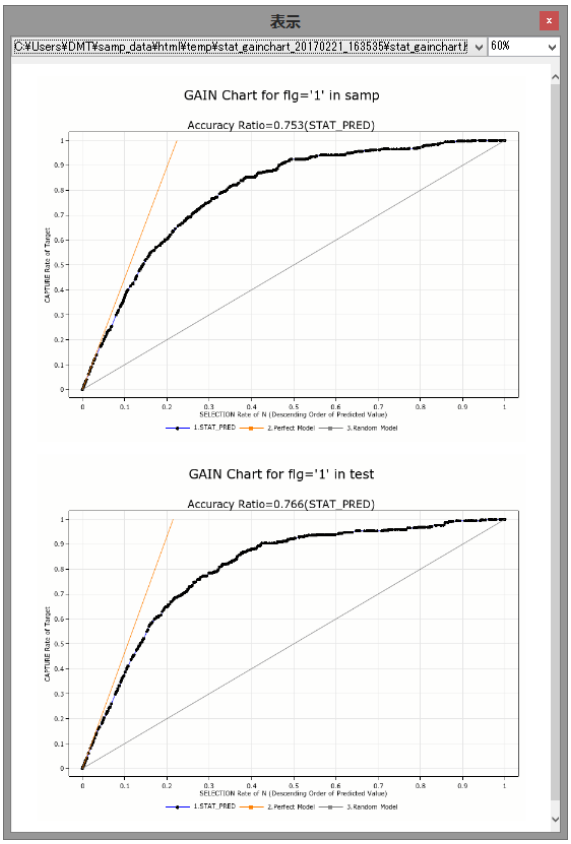


（途中省略）

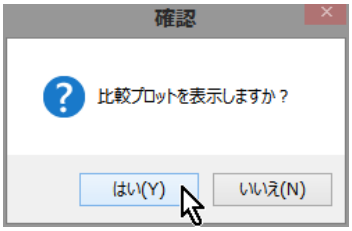


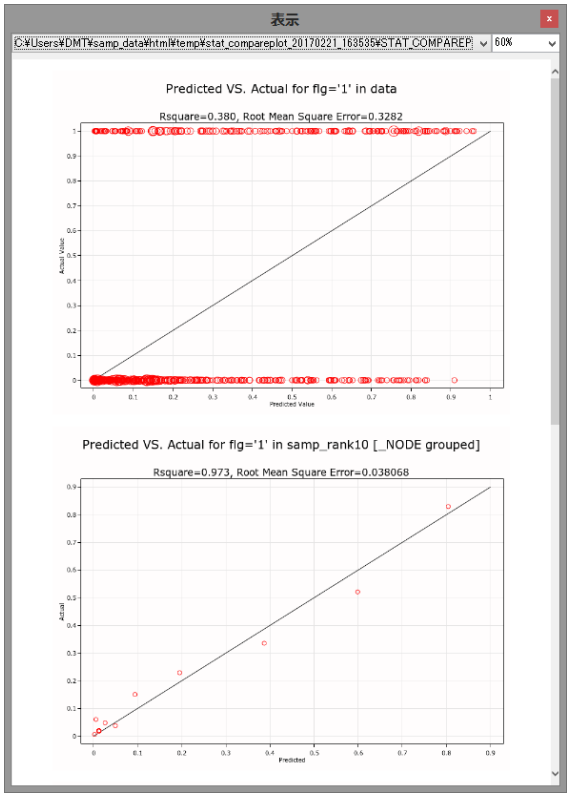
モデル作成過程の情報が表示されます。

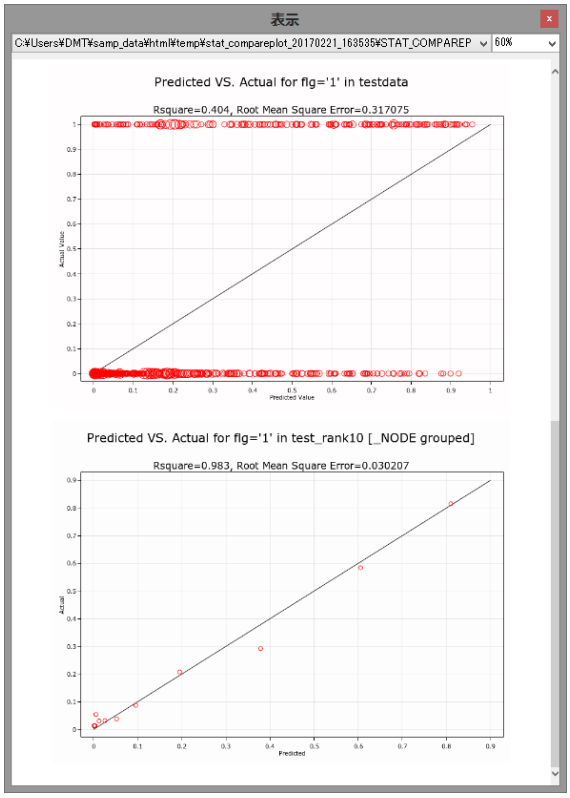




モデル作成データと検証データでのゲインチャートが、別々に表示されます。

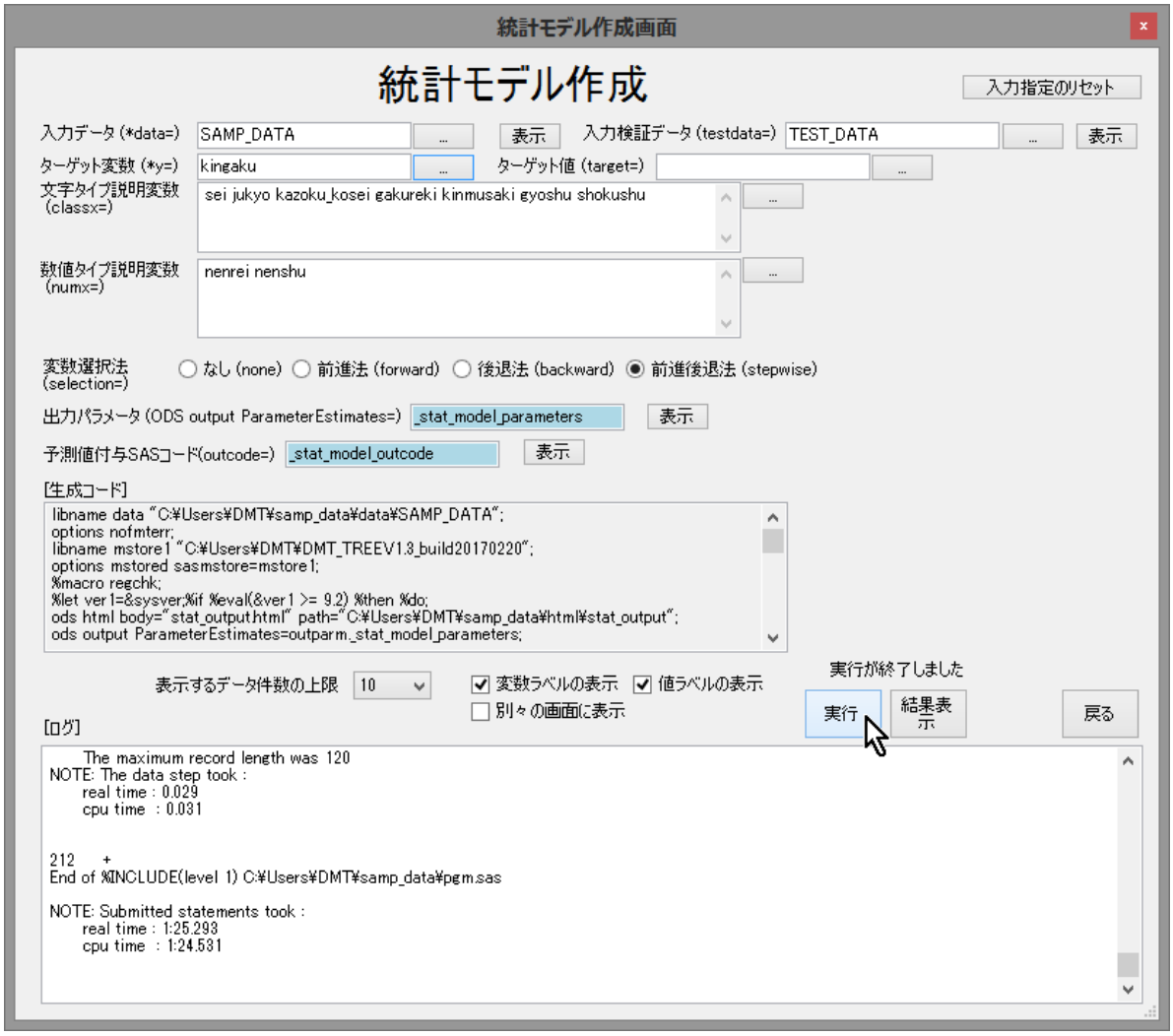




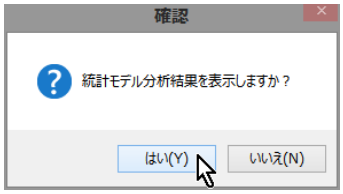


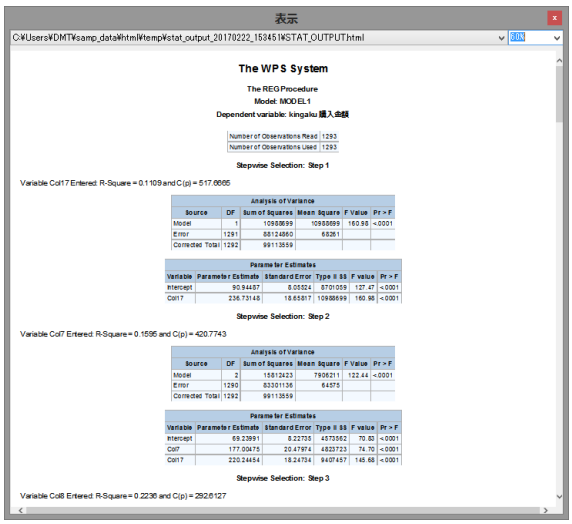
モデル作成データと検証データ、それぞれのデータにおけるモデル予測値（横軸）と実際値（縦軸）の比較プロットが表示されます。最初の図は個々の予測値単位、2番目の図は予測値を10個のランクに分けた場合の集計単位です。

（回帰モデルの実行例）

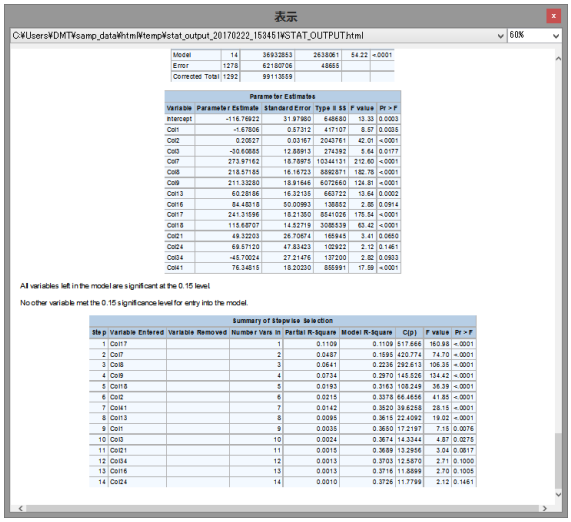


実行終了後、 結果表示 をクリック。

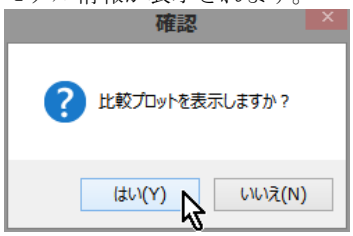


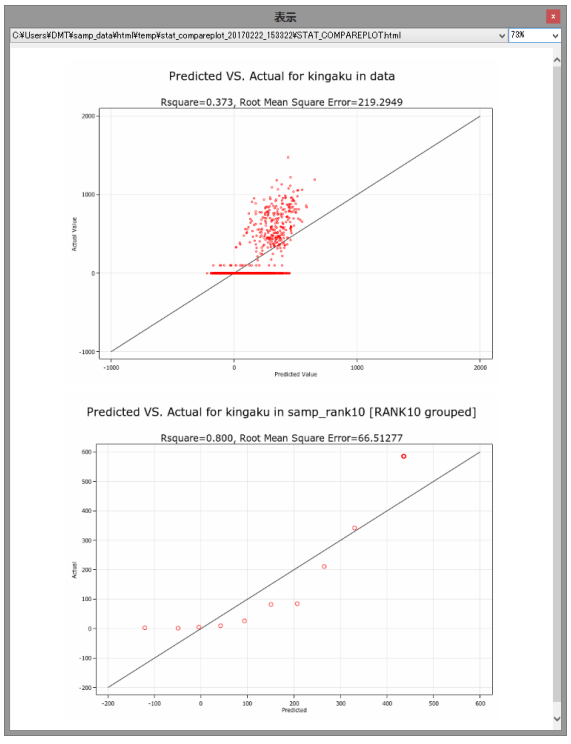


（途中省略）



モデル情報が表示されます。



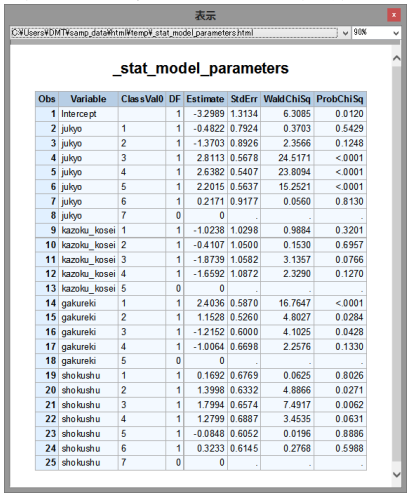


モデル作成データと検証データ、それぞれのデータにおけるモデル予測値（横軸）と実際値（縦軸）の比較プロットが表示されます。最初の図は個々の予測値単位、2番目の図は予測値を10個のランクに分けた場合の集計単位でプロットされています。

10.5.5 データセット出力

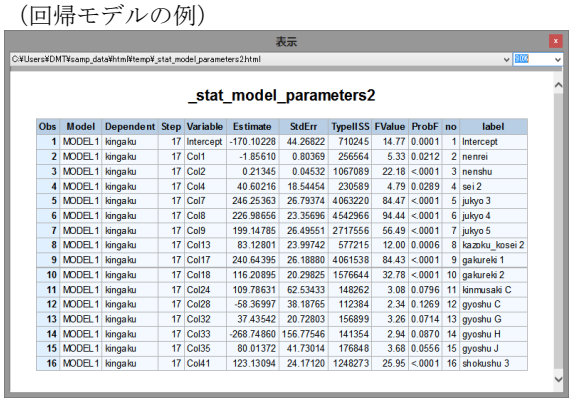
パラメータ推計結果データセットが、**出力パラメータ** (ODS output ParameterEstimates=) に指定した名前で **分析ディレクトリ**の下の**統計モデルディレクトリ**内に出力されます。 表示 を押すと内容が表示されます。

（ロジスティックモデルの場合）



Obs	Variable	ClassVal0	DF	Estimate	StdErr	WaldChiSq	ProbChiSq
1	Intercept		1	-3.2989	1.3134	6.3085	0.0120
2	jukyo	1	1	-0.4822	0.7924	0.3703	0.5429
3	jukyo	2	1	-1.3703	0.8926	2.3566	0.1248
4	jukyo	3	1	2.8113	0.5678	24.5171	<.0001
5	jukyo	4	1	2.6382	0.5407	23.8094	<.0001
6	jukyo	5	1	2.2015	0.5637	15.2521	<.0001
7	jukyo	6	1	0.2171	0.9177	0.0560	0.8130
8	jukyo	7	0	0	.	.	.
9	kazoku_kosei	1	1	-1.0238	1.0298	0.9884	0.3201
10	kazoku_kosei	2	1	-0.4107	1.0500	0.1530	0.6957
11	kazoku_kosei	3	1	-1.8739	1.0582	3.1357	0.0766
12	kazoku_kosei	4	1	-1.6592	1.0872	2.3290	0.1270
13	kazoku_kosei	5	0	0	.	.	.
14	gakureki	1	1	2.4036	0.5870	16.7647	<.0001
15	gakureki	2	1	1.1528	0.5260	4.8027	0.0284
16	gakureki	3	1	-1.2152	0.6000	4.1025	0.0428
17	gakureki	4	1	-1.0064	0.6698	2.2576	0.1330
18	gakureki	5	0	0	.	.	.
19	shokushu	1	1	0.1692	0.6769	0.0625	0.8026
20	shokushu	2	1	1.3998	0.6332	4.8866	0.0271
21	shokushu	3	1	1.7994	0.6574	7.4917	0.0062
22	shokushu	4	1	1.2799	0.6887	3.4535	0.0631
23	shokushu	5	1	-0.0848	0.6052	0.0196	0.8886
24	shokushu	6	1	0.3233	0.6145	0.2768	0.5988
25	shokushu	7	0	0	.	.	.

（回帰モデルの例）

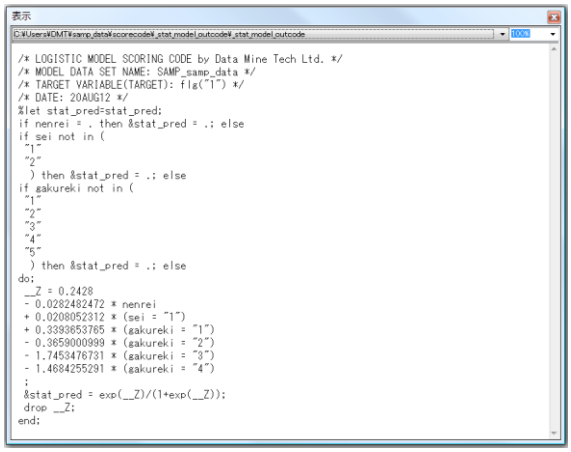


Obs	Model	Dependent	Step	Variable	Estimate	StdErr	TypeIISS	FValue	ProbF	no	label
1	MODEL1	kingaku	17	Intercept	-170.10228	44.26822	710245	14.77	0.0001	1	Intercept
2	MODEL1	kingaku	17	Col1	-1.85610	0.80369	256564	5.33	0.0212	2	nenrei
3	MODEL1	kingaku	17	Col2	0.21345	0.04532	1067089	22.18	<.0001	3	nenshu
4	MODEL1	kingaku	17	Col4	40.60216	18.54454	230589	4.79	0.0289	4	sei 2
5	MODEL1	kingaku	17	Col7	246.25363	26.79374	4063220	84.47	<.0001	5	jukyo 3
6	MODEL1	kingaku	17	Col8	226.98656	23.35696	4542966	94.44	<.0001	6	jukyo 4
7	MODEL1	kingaku	17	Col9	199.14785	26.49551	2717556	56.49	<.0001	7	jukyo 5
8	MODEL1	kingaku	17	Col13	83.12801	23.99742	577215	12.00	0.0006	8	kazoku_kosei 2
9	MODEL1	kingaku	17	Col17	240.64395	26.18880	4061538	84.43	<.0001	9	gakureki 1
10	MODEL1	kingaku	17	Col18	116.20895	20.29825	1576644	32.78	<.0001	10	gakureki 2
11	MODEL1	kingaku	17	Col24	109.78631	62.53433	148262	3.08	0.0796	11	kinmusaki C
12	MODEL1	kingaku	17	Col28	-58.36997	38.18765	112384	2.34	0.1269	12	gyoshu C
13	MODEL1	kingaku	17	Col32	37.43542	20.72803	156899	3.26	0.0714	13	gyoshu G
14	MODEL1	kingaku	17	Col33	-268.74860	156.77546	141354	2.94	0.0870	14	gyoshu H
15	MODEL1	kingaku	17	Col35	80.01372	41.73014	176848	3.68	0.0556	15	gyoshu J
16	MODEL1	kingaku	17	Col41	123.13094	24.17120	1248273	25.95	<.0001	16	shokushu 3

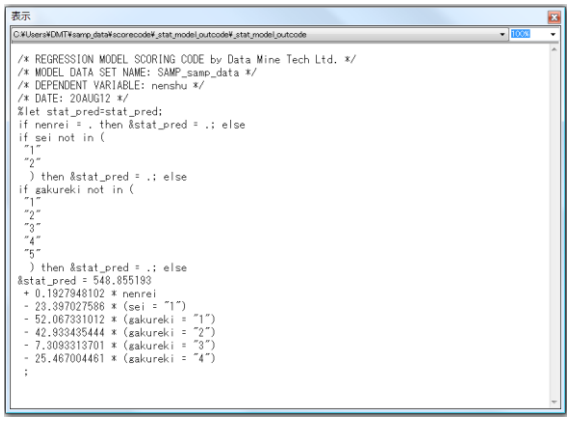
10.5.6 スコアリング用 SAS コード出力

予測値付与SASコード(outcode=) には予測値を付与するために必要なSASコードが**分析ディレクトリ**の下の**スコアコードディレクトリ**内に出力されます。

（ロジスティックモデルの場合）



（回帰モデルの例）



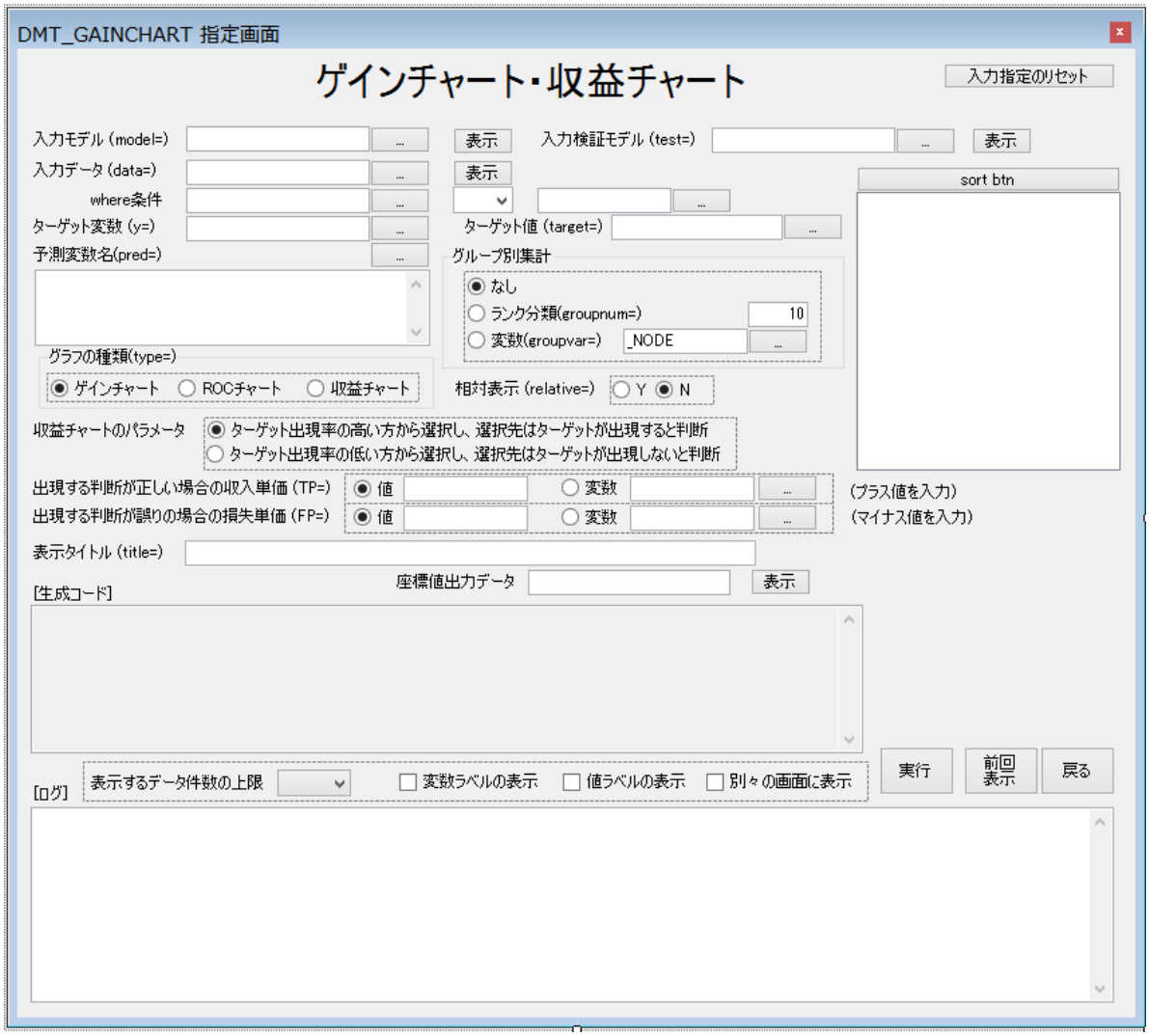
このコードを**「データ加工」**画面に貼り付ける方法で利用することにより、新たなデータに予測値をつけることができます。（ツリーモデルの予測値をデータにつけるための**「予測付与」**画面では利用できません）

なお、ここで作成したコードは**「コード管理」**画面で操作（表示、名前の変更、削除）することができます。

11. 分析画面 ④モデル検証

作成したツリーモデルの予測精度を確認します。

11.1 ゲイン・収益（dmt_gainchart）



11.1.1 概要

ゲインチャート・収益チャート（DMT_GAINCHART）は、分類木モデルの予測ターゲット出現率と実績値(出現または非出現のいずれか)が与えられたデータセットを入力として、モデルの精度を図示する**ゲインチャート**（CAP曲線とも呼ばれる）または**ROCチャート**を描き、精度評価値である AR（Accutacy Ratio）値または ROCエリア（ROC曲線下側面積）を表示するマクロです。さらに、モデルを業務施策に用いたときの対象選択件数と損益額の関係を表す**収益チャート**を描き、最大収益をもたらす選択件数と最大収益額を表示することもできます。ゲインチャート、ROCチャート、収益チャートは type=パラメータで gain(デフォルト),roc,profit をそれぞれ与えることにより切り替えます。

（収益チャート（type=profit 指定）について）

モデル予測ターゲット出現率の大きさ順にオブザベーションを並べたとき、各オブザベーションに対して、ある予測ターゲット出現率の大きさをしきい値として、以下のいずれかの意思決定を行うことを仮定します。

(1)（モデル予測ターゲット出現率＞＝しきい値を満たすオブザベーション）
ターゲットは出現するものとみなします。（これを

「**正予測**」（Positive Prediction）と呼びます。）

(2)（モデル予測ターゲット出現率＜しきい値を満たすオブザベーション）
ターゲットは出現しないものとみなします。（これを「**負予測**」（Negative Prediction）と呼びます。）

さて、モデル分析データやモデル検証データでは、これらの予測が正しかったか誤っていたかが事例として判明しており、個々のオブザベーションについて、以下の正誤表の4個のセル、A 正予測真(TP)、B 正予測偽(FP)、C 負予測偽(FN)、D 負予測真(TN) のいずれに該当していいるかを判断できます。

（正誤表(Confusion Matrix)）

予測	実際		計
	正事例	負事例	
正予測 （ターゲット出現と予測）	A 正予測真 （True Positive）	B 正予測偽 （False Positive）	正予測総件数 A+B
負予測 （ターゲット非出現と予測）	C 負予測偽 （False Negative）	D 負予測真 （True Negative）	負予測総件数 C+D
計	正事例総件数 A+C	負事例総件数 B+D	全体件数 N

しきい値の大きさを変化させると、すべてのオブザベーションの所属先が上記4つのセルの中で変化することになります。 このとき、各セルの件数にそれぞれのセルに対応する損益単価（TP=,FP=,TN=,FN=パラメータで与えます）を掛け合わせて合計すると、そのしきい値を採用したときの施策選択対象件数（正予測総件数）と期待収益が得られます。

model=,test=パラメータを指定し、モデルデータセットを入力とする場合は、TP=,FP=,TN=,FN=パラメータには、それぞれ定数を与えなければなりませんが、data=パラメータを指定し、予測スコアと実際値が入ったデータセットを入力とする場合は、TP=,FP=,TN=,FN=パラメータにはそれぞれの場合に対応する個々のオブザベーションの損益値を値に持つ変数名も指定することができます。

さて、何の出現率を予測するモデルを作成したか、そしてどのような業務施策にモデルを適用するのかによって、その業務施策の選択対象は出現率の大きい方か、小さい方かが決まります。たとえば、ネットショップにおける購買率予測モデルを**クロスセルやアップセルのコンタクト先を見つけることに用いる場合は予測購買率の高い方を選択**することになります。また、ローンの貸し倒れ率を予測するモデルを新たな申込客の**与信判断に用いる場合は、貸し倒れ率の小さい方を選択**することになります。

出現率の高い方を選択する場合は、TP=パラメータとFP=パラメータに値を与えます。TP=パラメータには、例えば、購買するという判断（正予測）が正しかった場合に得られる1人当たりの収入金額を正の値で与えます。そしてFP＝パラメータには、購買するという判断（正予測）が誤りであった場合に失う1人当たりの損失額を負の値で与えます。

一方、**出現率の低い方を選択する場合**は、TN=パラメータとFN=パラメータに値を与えます。TN=パラメータには、例えば、貸し倒れしないという判断（負予測）が正しかった場合に得られる1人当たりの収入金額を正の値で与えます。そしてFN＝パラメータには、貸し倒れしないという判断が誤りであった場合に失う1人当たりの損失額を負の値で与えます。

11.1.2 指定方法

（コマンド実行モードでの指定）

```
%dmt_gainchart(help,data=,y=,target=
,pred=_CONF,count=1,model=,test=,type=GAIN
,TP=0,FP=0,TN=0,FN=0
,groupvar=,groupnum=,relative=N
,ar_rocf=5.3,amountf=comma16.,pctf=7.2
,dev=GIF,title=,language=JAPANESE
,graph_language=ENGLISH
,outhtml=dmt_gainchart.html,outpath=)
```

（GUI実行モードでの変更点）

・help は指定不可。
・count=1 に固定。
・TYPE=PROFIT指定の場合は、TP=,FP=を一緒に指定するか、またはTN=,FN=を一緒に指定するか、いずれかの指定のみが許されます。（それ以外の組合せの2つ、または3つ以上を同時に指定できません）
・座標値出力データに名前を付けることができます。（デフォルトはグラフタイプによって、それぞれ_GAIN,_ROC,_PROFIT）

（入力データセットの個々のオブザベーションに付与された予測値の精度を評価する場合）

以下の3個のパラメータは必須指定です。

入力データ (data=) ... 入力データセット名の指定
(where=(条件式)などのデータセットオプションを指定可能).
ターゲット変数 (y=) ... ターゲット変数名の指定.
ターゲット値 (target=) ... ターゲット値の指定.

以下の4個のパラメータはオプション指定です。(=の右辺の値はデフォルト値を表しています)

予測変数名 (pred=_CONF) ... 予測値変数名の指定.
count=1 ... 入力データセットのオブザベーションが
集計データである場合の重み変数の指定
グループ単位の表示 (groupvar=)
予測値のランク単位の集計表示 (groupnum=)

（1つのツリーモデルを、モデル作成データのみ、またはモデルデータとテストデータ、それぞれに適用した場合の精度を比較評価する場合）

以下の2個のパラメータを指定します。ただし、test=

パラメータは単独指定できません。

入力モデル (model=) ... 入力モデルデータセット名の指定.
入力検証モデル (test=) ... テストデータに対してモデルを適用したときのモデル形式データ

（グラフの種類を選択するパラメータ）

グラフの種類 (type=GAIN) ...
ゲインチャート(type=GAIN)、ROCチャート(type=ROC)、
収益チャート(type=PROFIIT)
の切り替えを指定します。デフォルトはゲインチャート（CAP曲線）です.

（収益チャートのパラメータ）

以下の5個のパラメータは type=PROFIT の場合にのみ有効です。(=の右辺の値はデフォルト値を表しています）なお、最初の4個のすべてのパラメータをデフォルト0のままにして収益チャートを描いても無意味です。1個以上のパラメータの値を0以外に指定して収益チャートを描いてください。(GUI実行モードでは、TP=,FP=のペア、またはTN=,FN=のペアのいずれかのみ指定できます。)

出現する判断が正しい場合の収入単価 (TP=0)
出現する判断が誤りの場合の損失単価 (FP=0)
出現しない判断が正しい場合の収入単価 (TN=0)
出現しない判断が誤りの場合の損失単価 (FN=0) ...
収益チャートの相対表示 (relative=N)

（その他のパラメータ）

以下の10個のパラメータは任意指定です。(=の右辺の値はデフォルト値を表しています)

help ... 指定方法のヘルプメッセージの表示.（コマンド実行モードでのみ有効）
AR値、ROCエリア（ROC値）の表示フォーマットの指定 (ar_rocf=5.3)
収益値の表示フォーマットの指定 (amountf=comma16.)
百分率の表示フォーマットの指定 (pctf=7.2)
表示タイトル (title=) ... 画面出力のタイトルの指定.
(%str,%nrstr,%bquote などの関数で囲んで指定すること)
言語（language=JAPANESE) ... ログやメッセージを表示する言語の選択
グラフ表示言語 (graph_language=ENGLISH) ... ログやメッセージを表示する言語の選択
グラフデバイスの指定 (dev=GIF) ... グラフィックデバイスの指定.
HTML出力ファイル名 (outhtml=dmt_gainchart.html)（コマンド実行モードでのみ有効）
HTMLファイル出力ディレクトリの指定 (outpath=)（コマンド実行モードでのみ有効）

座標値出力データ … 図の座標値をデータ出力します。GUI実行環境では名前を指定できますが、コマンド実行モードでは ゲインチャートの場合 _gain, ROCチャートの場合 _roc, 収益チャートの場合 _profit という固定の名前でWORKライブラリに自動出力されます。

11.1.3 パラメータの詳細

入力モデル (model=)
入力モデルデータセット名を指定します。
例：model=bunseki1

入力検証モデル (test=)
入力モデル形式データセット名を指定します。この指定はmodel=パラメータと一緒に指定する必要があります。
例：test=kensho1

入力データ (data=)
入力データセット名を指定します。データセットオプションを指定できます。data=を指定する場合は、同時に、y=, target=(必要であれば), pred=の指定が必須です。
例：data=a, data=a(where=(DM="1"))

ターゲット変数 (y=)
data= 入力データセットに含まれるターゲット変数名を指定します。 例：y=flag, y=revenue

ターゲット値 (target=)
分類木モデルの予測値と実績値を比較検証する場合、y= ターゲット変数のターゲット値を指定します。回帰木モデルの検証を行う場合は指定してはいけません。
例：target="1"
なお、引用符で囲まなくても構いません。(自動判断します)

予測変数名 (pred=_CONF)
入力データセットに含まれる予測ターゲット出現率を表す変数名を1個～9個まで指定します。
なお、_CONF は分類木モデルの場合の予測変数名デフォルトとなっています。回帰木モデルの検証の場合は、回帰木モデルの予測変数名（デフォルトは _MEAN） を指定してください。
例：pred=Treepred1 Treepred2 STAT_pred

グループ単位の表示 (groupvar=)
data=指定の場合に、入力データに含まれる変数を1個だけ指定します。指定すると、チャートのプロット点が個々のオブザベーション単位から指定変数値が同じグループ単位の表示に変更されます。(注意：DMTデシジョンツリーV1.2の GROUPNODE=Y パラメータ指定は無効になりました。

GROUPVAR=_NODE に置き換えてください。）

予測値のランク単位の集計表示 (groupnum=)
data=指定の場合に、正の整数値を指定します。オブザベーションを予測値の大きさに基づくランクにグループ化（ビニングとも呼ばれる）し、ランクグループ単位の表示に変更します。

AR値、ROCエリア（ROC値）の表示フォーマットの指定
(ar_rocf=5.3)
ゲインチャート、ROC チャートの上部に表示される AR 値や ROC 面積値の表示フォーマットを指定します。

収益値の表示フォーマットの指定
(amountf=comma16.)
収益チャートの上部に表示される収益値の表示フォーマットを指定します。

百分率の表示フォーマットの指定 (pctf=7.2)
relative=Y を指定した収益チャートの上部に表示される件数比率の表示フォーマットを指定します。

グラフ画面表示言語 (graph_language=ENGLISH)
グラフィック出力画面に表示する既定のタイトルや軸ラベル等に表示する言語を指定します。graph_language=ENGLISH が既定です。※ 現行 WPS ではグラフ上には日本語が表示できませんので、デフォルトの graph_language=ENGLISH を変更しないでください。

11.1.4 収益チャートのパラメータの詳細

出現する判断が正しい場合の収入単価
(TP=0, GUI実行モードでは TP=1)
収益チャートにおいて、そのオブザベーションが正予測真(True Positive)の場合の収入単価を指定します。デフォルトは0です。任意の正の数値を指定します。ただし、data=入力データセットと共に指定する場合は、data=データセットに含まれる収入額を表す変数名も指定できます。正予測真とは「ターゲットは出現すると予測して実際も出現した」という正しい予測状況を意味しています。

出現する判断が誤りの場合の損失単価
(FP=0, GUI実行モードでは FP=-1)
収益チャートにおいて、そのオブザベーションが正予測偽(False Positive)の場合の損失単価を指定します。デフォルトは0です。任意の負の数値を指定します。ただし、data=入力データセットと共に指定する場合は、data=データセットに含まれる損失額を表す変数名も指定できます。正予測偽とは「ターゲットが出現すると予測したのに実際は出現しなかった」という予測が誤った状況を意味しています。

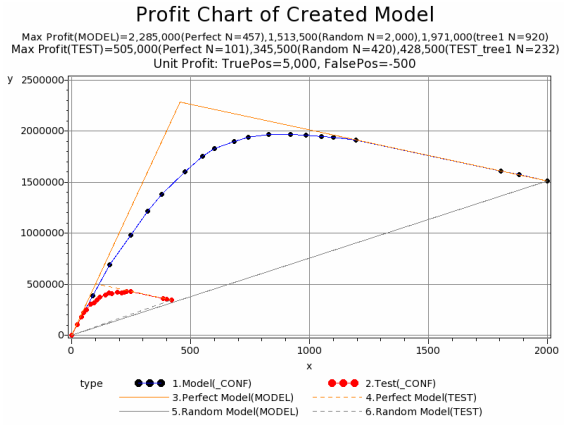
出現しない判断が正しい場合の収入単価
(TN=0, GUI実行モードでは TN=1)
収益チャートにおいて、そのオブザベーションが負予測真(True Negative)の場合の収入単価を指定します。デフォルトは0です。任意の正の数値を指定します。ただし、data=入力データセットと共に指定する場合は、data=データセットに含まれる収入額を表す変数名も指定できます。負予測真とは「ターゲットは出現しないと予測して実際も出現しなかった」という正しい予測状況を意味しています。

出現しない判断が誤りの場合の損失単価
(FN=0, GUI実行モードでは FN=-1)
収益チャートにおいて、そのオブザベーションが負予測偽(False Negative)の場合の損失単価を指定します。デフォルトは0です。任意の負の数値を指定します。ただし、data=入力データセットと共に指定する場合は、data=データセットに含まれる損失額を表す変数名も指定できます。負予測偽とは「ターゲットは出現しないと予測したのに実際は出現した」という状況を意味しています。

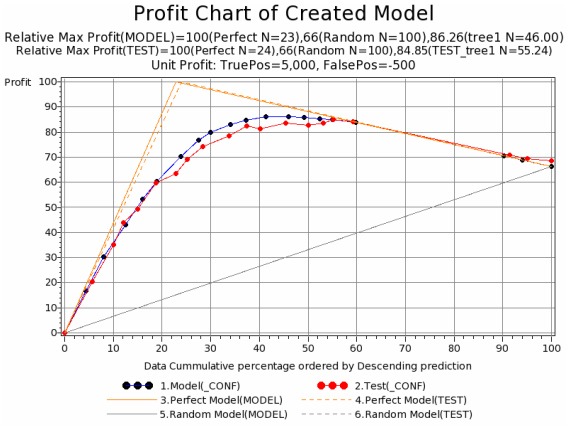
収益チャートの相対表示 (relative=N)
relative=Y を指定すると、収益チャートの縦軸、横軸を、絶対値の最大値が±100（符号は絶対値の最大値の符号）になるように比例変換して表示します。
model=モデルデータと検証データの件数が異なる場合に指定するとモデルと検証を比較しやすい表示になります。

例： %dmt_tree(data=samp_data,y=flg,target=1,x=sei--DM,mincnt=50,maxlvl=10,outmodel=tree1)
%dmt_treescore(model=tree1,data=test_data(where=(uniform(1)<0.2)),y=flg,target=1,outmodel=TEST_tree1)
%dmt_gainchart(model=tree1,test=TEST_tree1,type=PROFIT,TP=5000,FP=-500)
%dmt_gainchart(model=tree1,test=TEST_tree1,type=PROFIT,TP=5000,FP=-500,relative=Y)

(relative= 指定なし)



(relative=Y 指定あり)



11.1.5 GUI 実行モードで有効なパラメータの詳細

座標値出力データ
図の座標値を出力するデータセットに名前をつけます。（コマンド実行モードでは、WORKライブラリに決まった名前(type=指定によって、_gain, _roc,_profitのいずれか）で自動出力されます。）

11.1.6 コマンド実行モードで有効なパラメータの詳細

count=1
data= 入力データセットのオブザベーションが集計データである場合の重み変数名を指定します. 集計データで無い場合はデフォルトcount=1のままにしておきます. なお、重み変数名をこのパラメータで指定する場合、pred= パラメータに指定可能な予測値の数は1個のみになります。また、収益チャートをdata=入力データセットから作成する場合もcount=1にしておかなくてはなりません。（GUI実行モードでは1に固定）例：count=freq

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。（GUI 実行モードでは指定できません。）　例：%dmt_gainchart(help)

11.1.7 HTML 出力

分析結果の図表はhtmlファイルに出力されます。保存先はデフォルトではSASディスプレイマネージャまたはWPSワークベンチの管理下（ワークスペース内の一時保存ファイル）です。outpath=パラメータを指定すると、保存先を変更できます。（必ずフルパス指定します。引用符で囲んでも囲まなくてもかまいません）同時にouthtml=パラメータを指定すると、保存するhtmlファイルに自由に名前を付けることができます。

outhtml=dmt_gainchart.html
分析結果を保存するHTML出力ファイル名を指定します。
例：outhtml=out1.html,

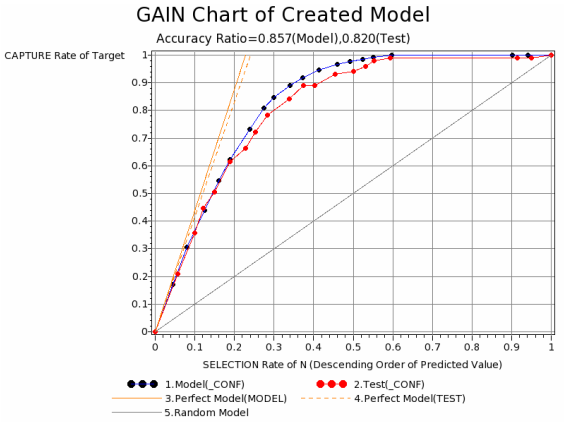
outpath=
HTML図表出力ファイルの保存ディレクトリを指定します。このパラメータを指定しない場合（デフォルト）、HTMLファイルはSASディスプレイマネージャまたはWPSワークベンチの管理下に作成されます。outpath=指定を行う場合、値は必ずフルパスで指定する必要があります。なお、パス指定全体を引用符で囲んでも囲まなくてもかまいません。
例：outpath='G:¥temp'

11.1.8 実行例

以下のように、samp_dataの変数flg=1の出現率を基準とするツリーモデル（tree1）を作成し、test_dataにモデルを当てはめたときのモデル形式データセット（TEST_tree1）を作成します。
ただし、例示のため、test_dataについては、where=(uniform(1)<0.2) をデータセットオプションで指定し、20%ランダム抽出したオブザベーションに対して検証用モデル形式データセットを作成しています。

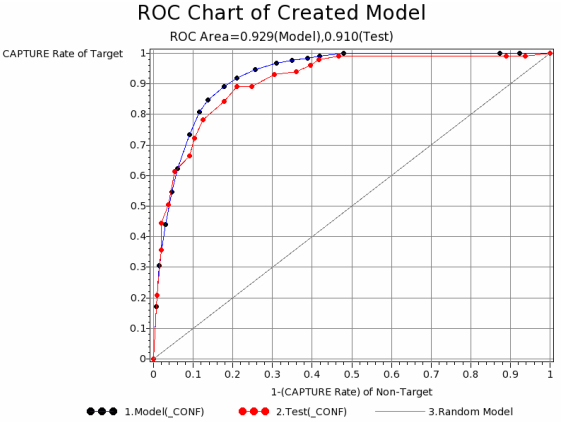
```
%dmt_tree(data=samp_data,y=flg,target=1,x=sei--DM,mincnt=50,maxlvl=10,outmodel=tree1)
%dmt_treescore(model=tree1,data=test_data(where=(uniform(1)<0.2)),y=flg,target=1,outmodel=TEST_tree1)
```

例１：ゲインチャート(CAP曲線)
```
%dmt_gainchart(model=tree1,test=TEST_tree1)
```



ゲインチャート（CAP曲線）における図の縦軸はターゲット再現率（ターゲット累積件数／全ターゲット件数)、横軸はターゲット予測出現率の大きい順にオブザベーションを選択した累積件数率（選択件数／全件数）を表します。model=,test=両パラメータを指定した場合の完全モデルは検証データ(test=) における完全モデルを表示しています。
タイトルにAR値が表示されます。

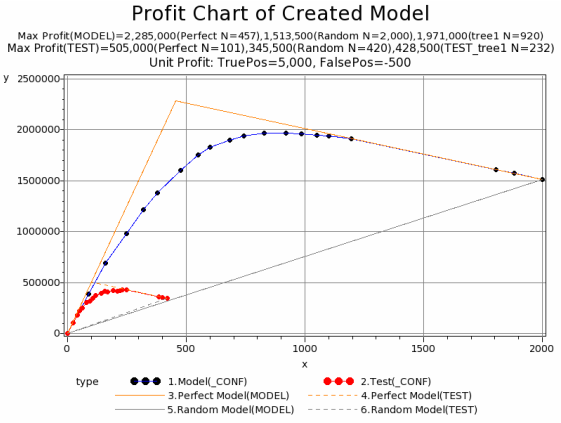
例２：ROCチャート(ROC曲線)
```
%dmt_gainchart(model=tree1,test=TEST_tree1,type=ROC)
```



ROCチャートにおける図の縦軸はターゲット再現率（ターゲット累積件数／全ターゲット件数）、横軸はターゲット予測出現率の大きい順に並べた累積データ上の(1-非ターゲット再現率)を表します。ROC曲線の用語では、ターゲット再現率のことを感度(Sensitivity),
非ターゲット再現率のことを特異度(Specificity)と呼ぶことが多いようです。
タイトルにはROCエリア値が表示されます。

なお、ゲインチャート、ROCチャートいずれを描いた場合でも、実行ログにはAR値、ROCエリア値が共通に表示されます。

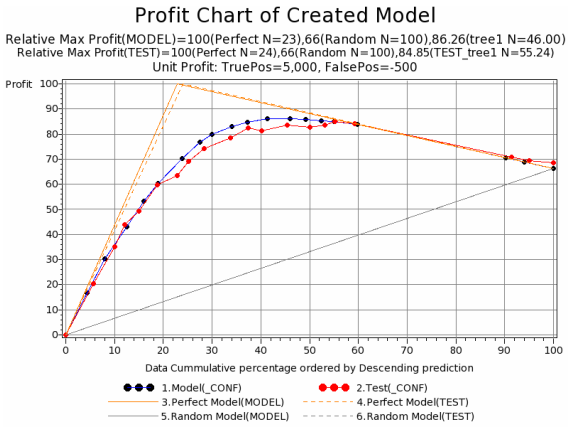
例３：収益チャート
%dmt_gainchart(model=tree1,test=TEST_tree1,type=PROFIT,TP=5000,FP=-500)



収益チャートにおける図の縦軸は収益、横軸はターゲット予測出現率の大きい順、または小さい順にオブザベーションを選択した選択件数です。完全モデル、ランダムモデルを含み、指定したモデルの最大収益とそのときの選択件数がタイトルに表示されます。

上記の例では、flg=1の出現率が高い方からデータを選択し、選択が正しかった（応答があった）場合は5,000の収益が得られ、選択が誤っていた（応答が無かった）場合は500の損失がかかるものとして計算した収益が縦軸に表示されています。モデル作成データでは、完全モデルの場合、実際に応答があったオブザベーションのみ（457件）をすべて選択する明らかに収益最大となり、2,285,000となることが示されています。ランダムモデルでは全件（2,000件）をすべて選択すると収益は1,513,500で最大となります。一方、モデル（tree1）を用いた場合は、モデル予測値から大きい順に920件を抽出したときが1,971,000で収益最大となり、検証データ（20%サンプリング）では232件を抽出したときの428,500が収益最大となることがわかります。

例４：relative=Y パラメータの指定
%dmt_gainchart(model=tree1,test=TEST_tree1,type=PROFIT,TP=5000,FP=-500,relative=Y)



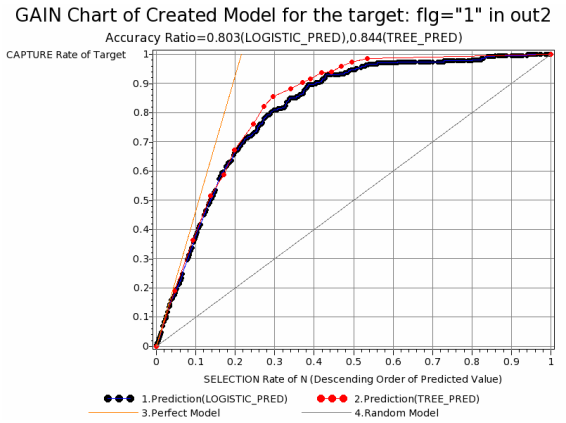
relative=Y パラメータを指定すると、縦軸の収益、横軸の件数ともに、相対表示になります。
これにより、データ件数がアンバランスな場合の収益の変化が相互に比較しやすくなります。

上記の場合、tree1での収益計算結果は検証データへのあてはめ結果とほぼ同じ傾向であることが分かります。

例５：
同じ分析データ（samp_data）で作成したロジスティックモデルとツリーモデルの精度を同じ検証データ（test_data）で比較します。

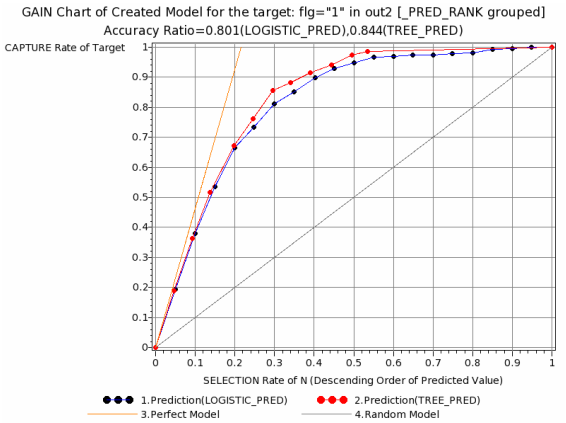
(ロジスティックモデル)
```
proc logistic data=samp_data outmodel=logistic1;
   class sei jukyo kazoku_kosei gakureki shokushu kinmusaki gyoshu DM;
   model flg(event="1")= sei jukyo kazoku_kosei gakureki shokushu kinmusaki gyoshu DM nenrei nenshu/selection=stepwise;
run;
```
(ロジスティックモデルの予測値をLOGISTIC_PREDという変数名でout1データに作成)
```
proc logistic inmodel=logistic1;
```

```
score data=test_data out=out1(rename=('P_あり'n=LOGISTIC_PRED));
run;
(ツリーモデルの作成)
%dmt_tree(data=samp_data,y=flg,target=1,
x=sei nenrei jukyo kazoku_kosei gakureki shokushu kinmusaki gyoshu nenshu DM,
mincnt=50,maxlvl=10,outmodel=tree1)
(out1データに、ツリーモデル予測値をTREE_PREDという変数名で追加したout2データを作成)
%dmt_treescore(model=tree1,data=out1,outscore=out2,pred=TREE_PRED)
(ゲインチャートの作成)
%dmt_gainchart(data=out2,y=flg,target=1,pred=LOGISTIC_PRED TREE_PRED)
```



例６： groupnum=パラメータの指定

```
%dmt_gainchart(data=out2,y=flg,target=1,pred=LOGISTIC_PRED TREE_PRED,groupnum=20)
```

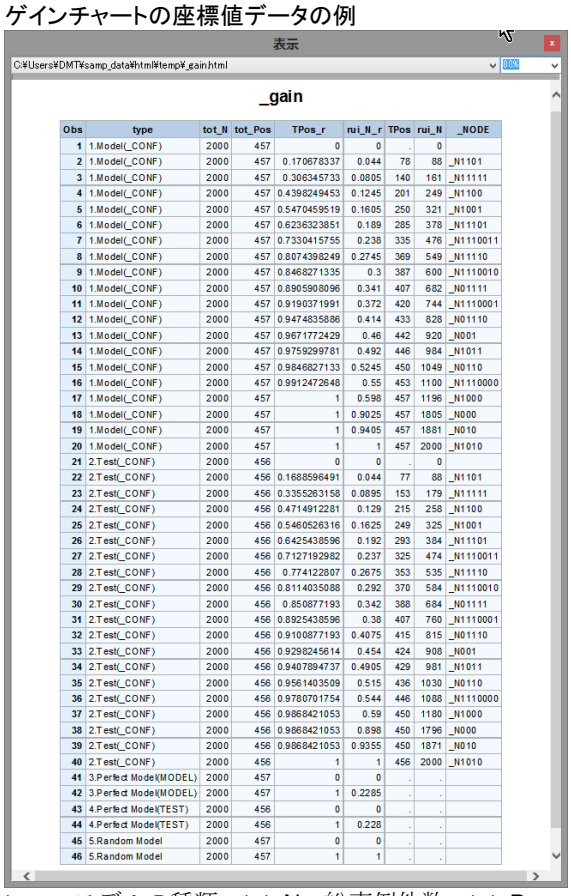


予測値の大きさでランキングした20グループに分けたオブザベーション同士をゲインチャートで比較すると予測値の種類数が異なるモデルが比較しやくすなります。

11.1.9 データセット出力

WORK._GAIN、WORK._ROC、WORK._PROFIT にそれぞれゲインチャート、ROCチャート、収益チャートの座標値（累積構成比率とターゲット再現率など）を格納したデータセットを自動出力します。（GUI実行モードの場合は座標出力データに名前を付けることができます。）

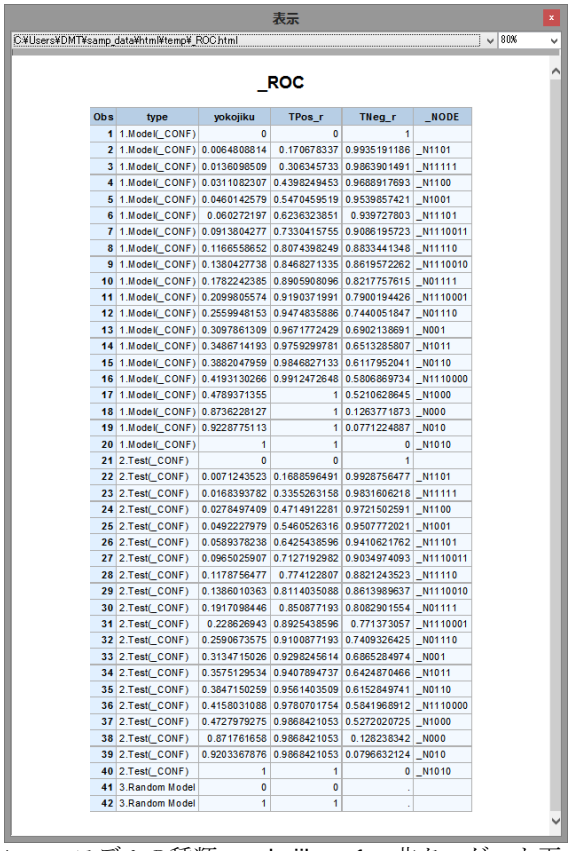
ゲインチャートの座標値データの例



Obs	type	tot_N	tot_Pos	TPos_r	rui_N_r	TPos	rui_N	_NODE
1	1.Model(_CONF)	2000	457	0	0	.	0	
2	1.Model(_CONF)	2000	457	0.170678337	0.044	78	88	_N1101
3	1.Model(_CONF)	2000	457	0.306345733	0.0805	140	161	_N11111
4	1.Model(_CONF)	2000	457	0.4398249453	0.1245	201	249	_N1100
5	1.Model(_CONF)	2000	457	0.5470459519	0.1605	250	321	_N1001
6	1.Model(_CONF)	2000	457	0.6236323851	0.189	285	378	_N11101
7	1.Model(_CONF)	2000	457	0.7330415755	0.238	335	476	_N1110011
8	1.Model(_CONF)	2000	457	0.8074398249	0.2745	369	549	_N11110
9	1.Model(_CONF)	2000	457	0.8468271335	0.3	387	600	_N1110010
10	1.Model(_CONF)	2000	457	0.8905908096	0.341	407	682	_N01111
11	1.Model(_CONF)	2000	457	0.9190371991	0.372	420	744	_N1110001
12	1.Model(_CONF)	2000	457	0.9474835886	0.414	433	828	_N01110
13	1.Model(_CONF)	2000	457	0.9671772429	0.46	442	920	_N001
14	1.Model(_CONF)	2000	457	0.9759299781	0.492	446	984	_N1011
15	1.Model(_CONF)	2000	457	0.9846827133	0.5245	450	1049	_N0110
16	1.Model(_CONF)	2000	457	0.9912472648	0.55	453	1100	_N1110000
17	1.Model(_CONF)	2000	457	1	0.598	457	1196	_N1000
18	1.Model(_CONF)	2000	457	1	0.9025	457	1805	_N000
19	1.Model(_CONF)	2000	457	1	0.9405	457	1881	_N010
20	1.Model(_CONF)	2000	457	1	1	457	2000	_N1010
21	2.Test(_CONF)	2000	456	0	0	.	0	
22	2.Test(_CONF)	2000	456	0.1688596491	0.044	77	88	_N1101
23	2.Test(_CONF)	2000	456	0.3355263158	0.0895	153	179	_N11111
24	2.Test(_CONF)	2000	456	0.4714912281	0.129	215	258	_N1100
25	2.Test(_CONF)	2000	456	0.5460526316	0.1625	249	325	_N1001
26	2.Test(_CONF)	2000	456	0.6425438596	0.192	293	384	_N11101
27	2.Test(_CONF)	2000	456	0.7127192982	0.237	325	474	_N1110011
28	2.Test(_CONF)	2000	456	0.774122807	0.2675	353	535	_N11110
29	2.Test(_CONF)	2000	456	0.8114035088	0.292	370	584	_N1110010
30	2.Test(_CONF)	2000	456	0.850877193	0.342	388	684	_N01111
31	2.Test(_CONF)	2000	456	0.8925438596	0.38	407	760	_N1110001
32	2.Test(_CONF)	2000	456	0.9100877193	0.4075	415	815	_N01110
33	2.Test(_CONF)	2000	456	0.9298245614	0.454	424	908	_N001
34	2.Test(_CONF)	2000	456	0.9407894737	0.4905	429	981	_N1011
35	2.Test(_CONF)	2000	456	0.9561403509	0.515	436	1030	_N0110
36	2.Test(_CONF)	2000	456	0.9780701754	0.544	446	1088	_N1110000
37	2.Test(_CONF)	2000	456	0.9868421053	0.59	450	1180	_N1000
38	2.Test(_CONF)	2000	456	0.9868421053	0.898	450	1796	_N000
39	2.Test(_CONF)	2000	456	0.9868421053	0.9355	450	1871	_N010
40	2.Test(_CONF)	2000	456	1	1	456	2000	_N1010
41	3.Perfect Model(MODEL)	2000	457	0	0	.	.	
42	3.Perfect Model(MODEL)	2000	457	1	0.2285	.	.	
43	4.Perfect Model(TEST)	2000	456	0	0	.	.	
44	4.Perfect Model(TEST)	2000	456	1	0.228	.	.	
45	5.Random Model	2000	457	0	0	.	.	
46	5.Random Model	2000	457	1	1	.	.	

type：モデルの種類 , tot_N：総事例件数 , tot_Pos：ターゲット事例総件数 , TPos_r：ターゲット再現率（縦軸）, rui_N_r：累積選択率（横軸）, TPos：正事例累積件数 , rui_N：累積件数 , _NODE：ノード番号

ROCチャートの座標値データの例

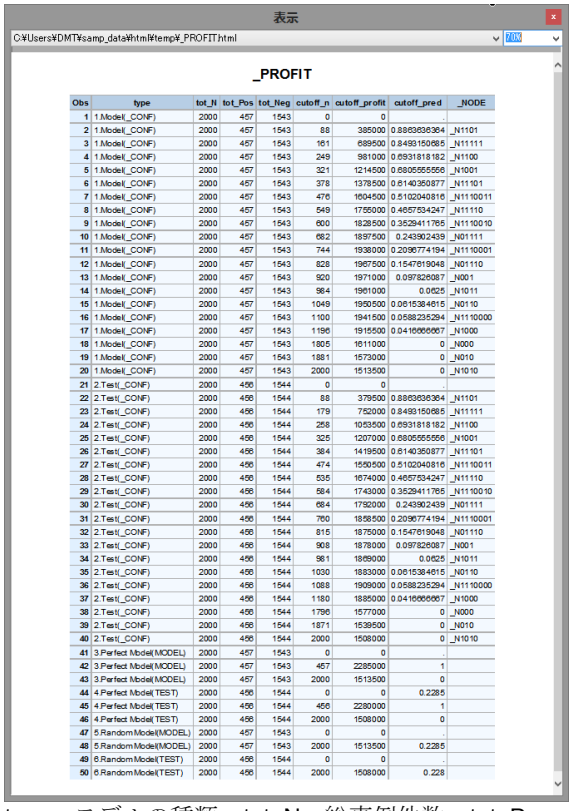


_ROC

Obs	type	yokojiku	TPos_r	TNeg_r	_NODE
1	1.Model(_CONF)	0	0	1	
2	1.Model(_CONF)	0.0064808814	0.170678337	0.9935191186	_N1101
3	1.Model(_CONF)	0.0136098509	0.306345733	0.9863901491	_N11111
4	1.Model(_CONF)	0.0311082307	0.4398249453	0.9688917693	_N1100
5	1.Model(_CONF)	0.0460142579	0.5470459519	0.9539857421	_N1001
6	1.Model(_CONF)	0.060272197	0.6236323851	0.939727803	_N11101
7	1.Model(_CONF)	0.0913804277	0.7330415755	0.9086195723	_N1110011
8	1.Model(_CONF)	0.1166558652	0.8074398249	0.8833441348	_N11110
9	1.Model(_CONF)	0.1380427738	0.8468271335	0.8619572262	_N1110010
10	1.Model(_CONF)	0.1782242385	0.8905908096	0.8217757615	_N01111
11	1.Model(_CONF)	0.2099805574	0.9190371991	0.7900194426	_N1110001
12	1.Model(_CONF)	0.2559948153	0.9474835886	0.7440051847	_N01110
13	1.Model(_CONF)	0.3097861309	0.9671772429	0.6902138691	_N001
14	1.Model(_CONF)	0.3486714193	0.9759299781	0.6513285807	_N1011
15	1.Model(_CONF)	0.3882047959	0.9846827133	0.6117952041	_N0110
16	1.Model(_CONF)	0.4193130266	0.9912472648	0.5806869734	_N1110000
17	1.Model(_CONF)	0.4789371355	1	0.5210628645	_N1000
18	1.Model(_CONF)	0.8736228127	1	0.1263771873	_N000
19	1.Model(_CONF)	0.9228775113	1	0.0771224887	_N010
20	1.Model(_CONF)	1	1	0	_N1010
21	2.Test(_CONF)	0	0	1	
22	2.Test(_CONF)	0.0071243523	0.1688596491	0.9928756477	_N1101
23	2.Test(_CONF)	0.0168393782	0.3355263158	0.9831606218	_N11111
24	2.Test(_CONF)	0.0278497409	0.4714912281	0.9721502591	_N1100
25	2.Test(_CONF)	0.0492227979	0.5460526316	0.9507772021	_N1001
26	2.Test(_CONF)	0.0589378238	0.6425438596	0.9410621762	_N11101
27	2.Test(_CONF)	0.0965025907	0.7127192982	0.9034974093	_N1110011
28	2.Test(_CONF)	0.1178756477	0.774122807	0.8821243523	_N11110
29	2.Test(_CONF)	0.1386010363	0.8114035088	0.8613989637	_N1110010
30	2.Test(_CONF)	0.1917098446	0.850877193	0.8082901554	_N01111
31	2.Test(_CONF)	0.228626943	0.8925438596	0.771373057	_N1110001
32	2.Test(_CONF)	0.2590673575	0.9100877193	0.7409326425	_N01110
33	2.Test(_CONF)	0.3134715026	0.9298245614	0.6865284974	_N001
34	2.Test(_CONF)	0.3575129534	0.9407894737	0.6424870466	_N1011
35	2.Test(_CONF)	0.3847150259	0.9561403509	0.6152849741	_N0110
36	2.Test(_CONF)	0.4158031088	0.9780701754	0.5841968912	_N1110000
37	2.Test(_CONF)	0.4727979275	0.9868421053	0.5272020725	_N1000
38	2.Test(_CONF)	0.871761658	0.9868421053	0.128238342	_N000
39	2.Test(_CONF)	0.9203367876	0.9868421053	0.0796632124	_N010
40	2.Test(_CONF)	1	1	0	_N1010
41	3.Random Model	0	0	.	
42	3.Random Model	1	1	.	

type：モデルの種類 , yokojiku：1 - 非ターゲット再現率（横軸） , TPos_r：ターゲット再現率（縦軸）, TNeg_r：非ターゲット再現率 , _NODE：ノード番号

収益チャートの座標値データの例



_PROFIT

Obs	type	tot_N	tot_Pos	tot_Neg	cutoff_n	cutoff_profit	cutoff_pred	_NODE
1	1.Model(_CONF)	2000	457	1543	0	0		
2	1.Model(_CONF)	2000	457	1543	88	385000	0.8863636364	_N1101
3	1.Model(_CONF)	2000	457	1543	161	689500	0.8493150685	_N11111
4	1.Model(_CONF)	2000	457	1543	249	981000	0.6931818182	_N1100
5	1.Model(_CONF)	2000	457	1543	321	1214500	0.6805555556	_N1001
6	1.Model(_CONF)	2000	457	1543	378	1378500	0.6140350877	_N11101
7	1.Model(_CONF)	2000	457	1543	476	1604500	0.5102040816	_N1110011
8	1.Model(_CONF)	2000	457	1543	549	1755000	0.4657534247	_N11110
9	1.Model(_CONF)	2000	457	1543	600	1828500	0.3529411765	_N1110010
10	1.Model(_CONF)	2000	457	1543	682	1897500	0.243902439	_N01111
11	1.Model(_CONF)	2000	457	1543	744	1938000	0.2096774194	_N1110001
12	1.Model(_CONF)	2000	457	1543	828	1967500	0.1547619048	_N01110
13	1.Model(_CONF)	2000	457	1543	920	1971000	0.097826087	_N001
14	1.Model(_CONF)	2000	457	1543	984	1961000	0.0625	_N1011
15	1.Model(_CONF)	2000	457	1543	1049	1950500	0.0615384615	_N0110
16	1.Model(_CONF)	2000	457	1543	1100	1941500	0.0588235294	_N1110000
17	1.Model(_CONF)	2000	457	1543	1196	1915500	0.0416666667	_N1000
18	1.Model(_CONF)	2000	457	1543	1805	1611000	0	_N000
19	1.Model(_CONF)	2000	457	1543	1881	1573000	0	_N010
20	1.Model(_CONF)	2000	457	1543	2000	1513500	0	_N1010
21	2.Test(_CONF)	2000	456	1544	0	0		
22	2.Test(_CONF)	2000	456	1544	88	379500	0.8863636364	_N1101
23	2.Test(_CONF)	2000	456	1544	179	752000	0.8493150685	_N11111
24	2.Test(_CONF)	2000	456	1544	258	1053500	0.6931818182	_N1100
25	2.Test(_CONF)	2000	456	1544	325	1207000	0.6805555556	_N1001
26	2.Test(_CONF)	2000	456	1544	384	1419500	0.6140350877	_N11101
27	2.Test(_CONF)	2000	456	1544	474	1550500	0.5102040816	_N1110011
28	2.Test(_CONF)	2000	456	1544	535	1674000	0.4657534247	_N11110
29	2.Test(_CONF)	2000	456	1544	584	1743000	0.3529411765	_N1110010
30	2.Test(_CONF)	2000	456	1544	684	1792000	0.243902439	_N01111
31	2.Test(_CONF)	2000	456	1544	760	1858500	0.2096774194	_N1110001
32	2.Test(_CONF)	2000	456	1544	815	1875000	0.1547619048	_N01110
33	2.Test(_CONF)	2000	456	1544	908	1878000	0.097826087	_N001
34	2.Test(_CONF)	2000	456	1544	981	1869000	0.0625	_N1011
35	2.Test(_CONF)	2000	456	1544	1030	1863000	0.0615384615	_N0110
36	2.Test(_CONF)	2000	456	1544	1088	1909000	0.0588235294	_N1110000
37	2.Test(_CONF)	2000	456	1544	1180	1885000	0.0416666667	_N1000
38	2.Test(_CONF)	2000	456	1544	1796	1577000	0	_N000
39	2.Test(_CONF)	2000	456	1544	1871	1539500	0	_N010
40	2.Test(_CONF)	2000	456	1544	2000	1508000	0	_N1010
41	3.Perfect Model(MODEL)	2000	457	1543	0	0	.	
42	3.Perfect Model(MODEL)	2000	457	1543	457	2285000	1	
43	3.Perfect Model(MODEL)	2000	457	1543	2000	1513500	0	
44	4.Perfect Model(TEST)	2000	456	1544	0	0	0.2285	
45	4.Perfect Model(TEST)	2000	456	1544	456	2280000	1	
46	4.Perfect Model(TEST)	2000	456	1544	2000	1508000	0	
47	5.Random Model(MODEL)	2000	457	1543	0	0		
48	5.Random Model(MODEL)	2000	457	1543	2000	1513500	0.2285	
49	6.Random Model(TEST)	2000	456	1544	0	0		
50	6.Random Model(TEST)	2000	456	1544	2000	1508000	0.228	

type：モデルの種類 , tot_N：総事例件数 , tot_Pos：ターゲット事例総件数 , tot_Neg：非ターゲット事例総件数 , cutoff_n：累積選択件数（横軸）, cutoff_profit：選択したオブザベーションから得られる収益（縦軸）, cutoff_pred：ターゲット出現率のしきい値（そのノードのターゲット出現率） , _NODE：ノード番号

なお、data=入力データを指定したときは、変数_NODEは以下のように変更されます。

・GROUPVAR=パラメータを指定した場合はその変数
・GROUPNUM=パラメータを指定した場合は_PRED_RANK
・その他の場合は変数は追加されません。

11.1.10 欠損値の取り扱い

data=入力の場合、いずれかの予測値に欠損が存在するオブザベーションは計算から除外されます。

収益チャートにおいてTP=,FP=,TN=,FN=変数値に欠損値があれば、エラーメッセージを出して分析を中断します。(すべて0はOKですが、意味がありません)

11.1.11 制限

data=入力データセットを指定し、オブザベーションに対する複数のモデルによる予測値を比較する場合にpred=パラメータに指定可能な予測変数の数の上

限は9個です。

```
e_name e_type nobs lab&i spc&i typ&i zketa
_speclen _specnum _errormsg
```

11.1.12 コマンド実行モードでの注意

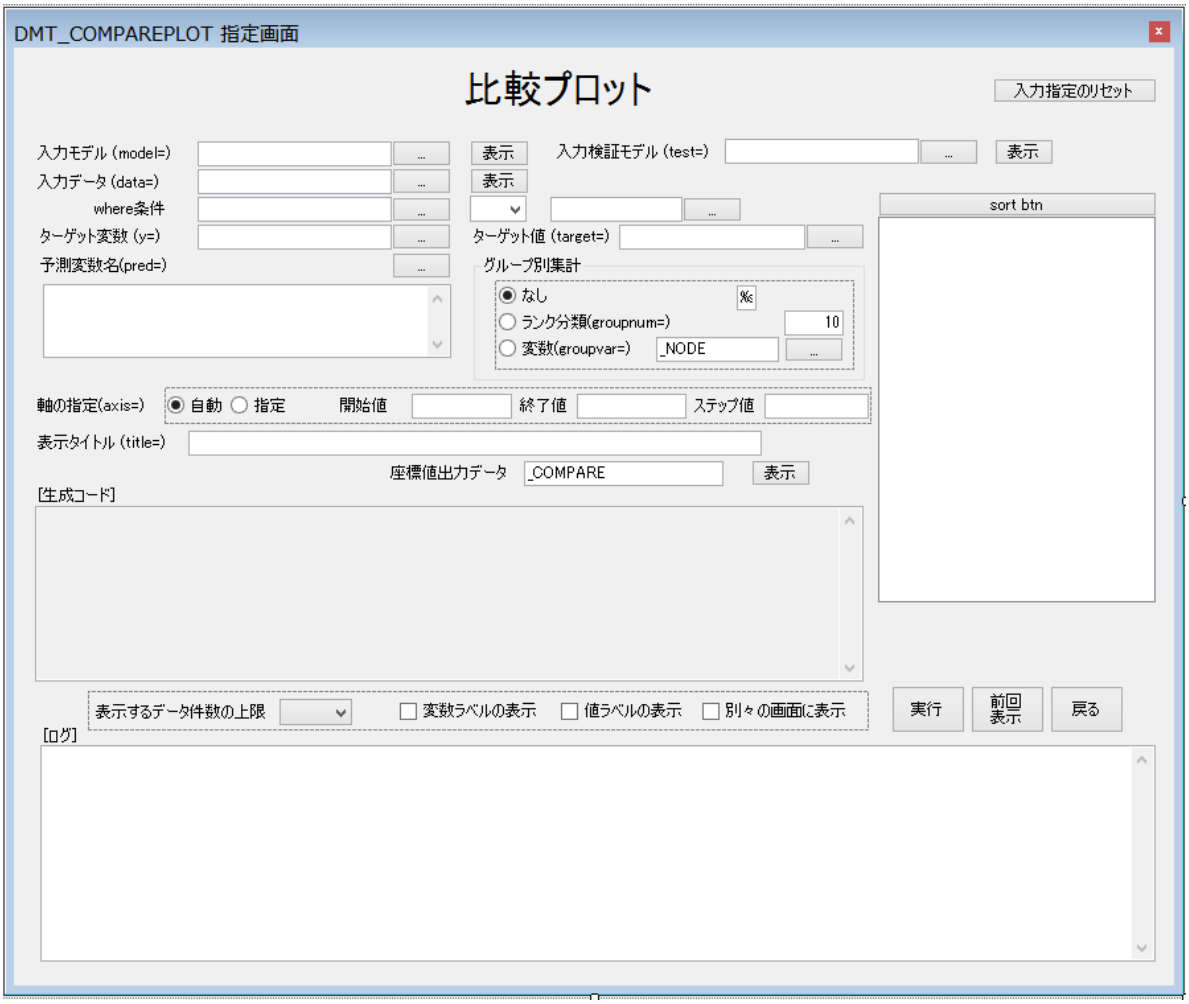
実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

また、以下のユーザ定義フォーマットがWORKライブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
$_item
```

さらに、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

11.2 比較プロット（dmt_compareplot）



11.2.1 概要

このアプリケーションで、**比較プロット**（DMT_COMPAREPLOT）と呼ぶ分析グラフは、検証用データにおける終端ノード別のターゲット出現率、もしくはターゲット変数値の予測－実績比較プロットを描き、R2乗（R-Square）値を表示するマクロです。
ゲインチャート・収益チャート（DMT_GAINCHART）が表示するゲインチャート、ROCチャートやAR値、ROCエリア値は、ターゲット出現率やターゲット変数値の大きさそのものではなく、予測値と実績値と順序関係における整合性を評価します。それに対して比較プロットやR2乗値は誤差（＝実績値－予測値）の大きさそのものを評価します。

ここで、誤差の定義は、以下のように、2通り選択できます。

(1) model=,test=パラメータを指定した場合
ノード単位の平均誤差の大きさに基づいた 実際値と予測値の比較プロットおよびR2乗値を表示します。

Error(n)=Actual(n)-Pred(n)

ただし、
Error(n) は n 番目のノードの誤差
Actual(n) は n 番目のノードの実績値の平均値
Pred(n) は n 番目のノードの予測値の平均値

R2乗=1-誤差平方和/偏差平方和
=1-Σ{W(n)*Error(n)}/Σ[W(n)*{Actual(n)-Actual_bar}]

ただし、
W(n) は n 番目のノードに含まれるオブザベーション数
Error(n) は n 番目のノードの誤差（上記）
Actual(n) は n 番目のノードの実績値の平均値
Actual_bar は実績値の全体平均値

(2) data=パラメータを指定した場合
標準（groupvar=パラメータ指定なし）ではオブザベ

ーション単位の誤差の大きさに基づいた実際値と予測値の比較プロットおよびR2乗値を表示します。(ただし、groupnode=Yを指定すると、(1)と同じくノード単位の誤差を計算します。)

Error(i)=Actual(i)-Pred(i)

ただし、
Error(i) は i 番目のオブザベーションの誤差
Actual(i) は i 番目のオブザベーションの実績値
Pred(i) は i 番目のオブザベションの予測値

R2乗=1-誤差平方和/偏差平方和
=1-Σ{ Error(i)}/Σ{Actual(i)-Actual_bar}

ただし、
Error(i) は i 番目のオブザベーションの誤差（上記）
Actual(i) は i 番目のオブザベーションの実績値
Actual_bar は実績値の全体平均値

11.2.2 指定方法

（コマンド実行モードでの指定）

%dmt_compareplot(help,data=,y=,target=
,pred=_CONF,plotobs=2000
,groupvar=,groupnum=,model=,test=,axis=,title=
,r2f=5.3,rmsef=best8.
,d_label=[D].c_label=[C],dif_label=[D]-[C]
,dev=GIF,title=,language=JAPANESE
,graph_language=ENGLISH
,outhtml=dmt_compareplot.html,outpath=**)**

（GUI実行モードでの変更点）

・help は指定不可。
・座標出力データに名前を付けることができます。（コマンド実行モードでは _COMPARE 固定）
・plotobs= はオプション画面で指定

（入力データセットの個々のオブザベーションに付与された予測値と実際値を比較する場合）

以下の3個のパラメータは必須指定です。ただし、回帰木モデルの場合はtarget=パラメータは指定してはいけません。

入力データ (data=) ... 入力データセット名の指定.
ターゲット変数 (y=) ... ターゲット変数名の指定.
ターゲット値 (target=) … ターゲット値の指定
（分類木モデルの場合のみ必須）

以下の5個のパラメータはオプション指定です。(=の右辺の値はデフォルト値を表しています)

予測変数名 (pred=_CONF) … 予測変数名の指定.(単一変数のみ指定可)
軸の指定 (axis=) … グラフの縦軸、横軸の値の範囲指定. デフォルトは自動設定です.
図に表示する上限オブザベーション数
(plotobs=2000) … 図に表示する上限オブザベーション数を指定します
グループ単位の表示 (groupvar=)
予測値のランク単位の集計表示 (groupnum=)

（1つのツリーモデルをテストデータに適用した場合の予測値と実際値の誤差を比較する場合）

以下の2個のパラメータを同時に指定します。
（model=データセットは予測値、test=データセットは実績値をそれぞれ計算するために用いられます.）

入力モデル (model=) … 入力モデルデータセット名の指定.
入力検証モデル (test=) … テストデータに対してモデルを適用したときのモデル形式データセット名の指定

（その他のパラメータ）

以下の13個のパラメータはオプション指定です。(=の右辺の値はデフォルト値を表しています)

help … 指定方法のヘルプメッセージの表示.（コマンド実行モードでのみ有効）
R2乗値の表示フォーマットの指定(r2f=5.3)
誤差平均平方値の表示フォーマットの指定
(rmsef=best8.)
アップリフトモデルにおける処理群（DATA）を表す記号
(d_label=[D])
アップリフトモデルにおける対照群（Control）を表す記号
(c_label=[C])
アップリフトモデルにおける処理群−対照群間の差を表す記号 (dif_label=[D]-[C])
表示タイトル (title=) ... 画面出力のタイトルの指定.
(%str,%nrstr,%bquote などの関数で囲んで指定すること)
言語 (language=JAPANESE)
グラフ表示言語 (graph_language=ENGLISH)
グラフデバイスの指定 (dev=GIF) ... グラフィックデバイスの指定.
HTML出力ファイル名
(outhtml=dmt_compareplot.html)（コマンド実行モードでのみ有効）
HTMLファイル出力ディレクトリの指定 (outpath=)（コマンド実行モードでのみ有効）

座標値出力データ … 図の座標値をデータ出力します。GUI実行環境では名前を指定できますが、コマンド実行モードでは _comparet という固定の名前でWORKライブラリに自動出力されます。

11.2.3 パラメータの詳細

入力モデル (model=)
入力モデルデータセット名を指定します。この指定

はtest=パラメータと一緒に指定する必要があります。
例：model=bunseki1

入力検証モデル (test=)
入力モデル形式データセット名を指定します。この指定はmodel=パラメータと一緒に指定する必要があります。
例：test=kensho1

入力データ (data=)
入力データセット名を指定します。データセットオプションを指定できます。data=を指定する場合は、同時に、y=, target=(必要であれば), pred=の指定が必須です。
例：data=a, data=a(where=(DM="1"))

ターゲット変数 (y=)
data= 入力データセットに含まれるターゲット変数名を指定します。 例：y=flag, y=revenue

ターゲット値(target=)
分類木モデルの予測値と実績値を比較検証する場合、y= ターゲット変数のターゲット値を指定します。回帰木モデルの検証を行う場合は指定してはいけません。

予測変数名 (pred=_CONF)
入力データセットに含まれる予測ターゲット出現率を表す変数名を１つだけ指定します。
なお、_CONF は分類木モデルの場合の予測変数名デフォルトとなっています。回帰木モデルの検証の場合は、回帰木モデルの予測変数名（デフォルトは _MEAN） を指定してください。

図に表示する上限オブザベーション数 (plotobs=2000)
data= 入力データセットに含まれるデータから図に表示する上限オブザベーション数を正の整数値で指定します。デフォルトは5000です。入力データセットのオブザベーション数がこの上限を超える場合はランダム抽出を行い上限数のデータのみプロットの対象にしています。なお、R2乗値の計算は全オブザベーションから計算しています。

グループ単位の表示 (groupvar=)
data=指定の場合に、入力データに含まれる変数を1個だけ指定します。指定すると、チャートのプロット点が個々のオブザベーション単位から指定変数値が同じグループ単位の表示に変更されます。（注意：DMTデシジョンツリーV1.2の GROUPNODE=Y パラメータ指定は無効になりました。
GROUPVAR=_NODE に置き換えてください。）

予測値のランク単位の集計表示 (groupnum=)
data=指定の場合に、正の整数値を指定します。オブザベーションを予測値の大きさに基づくランクにグループ化（ビニングとも呼ばれる）し、ランクグループ単位の表示に変更します。

軸の指定 (axis=)
グラフの両軸の範囲を axis=開始値 to 終了値 by 増分値 の形式で指定します。デフォルトはデータから自動計算します。縦軸、横軸とも共通の範囲が用いられます。自動計算結果が見づらい場合は、実際の分布範囲に合わせた範囲を指定することにより見やすくなる場合があります。
例：axis=0 to 0.5 by 0.05

グラフデバイスの指定 (dev=GIF)
グラフ描画に用いるグラフィックデバイス名を指定します。デフォルトは dev=GIF です。
例: dev=JPEG

表示タイトル (title=)
画面出力される表にタイトルを指定できます。指定しない（デフォルト）場合、以下のようなタイトルが自動的に付与されます。

%bquote(&data におけるモデル予測値 対 実績値 (ターゲット:&y="&target"))

タイトルを指定する場合、上記のように%bquote関数の中に記述してください。

言語(language=JAPANESE)
分析実行中のメッセージ出力、結果の表のタイトル、表項目などの表示言語を選択します。ただし、現バージョンでは、日本語か英語の2種類のみ選択可能です。
例：language=ENGLISH

グラフ画面表示言語 (graph_language=ENGLISH)
グラフィック出力画面に表示する既定のタイトルや軸ラベル等に表示する言語を指定します。graph_language=ENGLISH が既定です。※ 現行WPS ではグラフ上には日本語が表示できませんので、デフォルトの graph_language=ENGLISH を変更しないでください。

R2乗値の表示フォーマットの指定 (r2f=5.3)
比較プロットの上部に表示される R2 乗値の表示フォーマットを指定します。

誤差平均平方値の表示フォーマットの指定 (rmsef=best8.)
比較プロットの上部に表示される誤差平均平方（RMSE）の表示フォーマットを指定します。

アップリフトモデルにおける処理群（DATA）を表す記号 (d_label=[D])
model=と test=を指定したモデルがアップリフトモデルの場合に有効。処理群を表す記号を指定します。

アップリフトモデルにおける対照群（Control）を表す記号（c_label=[C])
model=と test=を指定したモデルがアップリフトモデルの場合に有効。対照群を表す記号を指定します。

アップリフトモデルにおける処理群−対照群間の差を表す記号（dif_label=[D]-[C])
model=と test=を指定したモデルがアップリフトモデルの場合に有効。処理群と対照群のターゲット値の差を表す記号を指定します。

11.2.4 GUI 実行モードで有効なパラメータの詳細

座標値出力データ
図の座標値を出力するデータセットに名前をつけます。（コマンド実行モードでは、WORKライブラリに_compare という決まった名前で自動出力されます。）

11.2.5 コマンド実行モードで有効なパラメータの詳細

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。（GUI 実行モードでは指定できません。） 例：%dmt_compareplot(help)

11.2.6 HTML 出力

分析結果の図表はhtmlファイルに出力されます。保存先はデフォルトではSASディスプレイマネージャまたはWPSワークベンチの管理下（ワークスペース内の一時保存ファイル）です。outpath=パラメータを指定すると、保存先を変更できます。（必ずフルパス指定します。引用符で囲んでも囲まなくてもかまいません）同時にouthtml=パラメータを指定すると、保存するhtmlファイルに自由に名前を付けることができます。

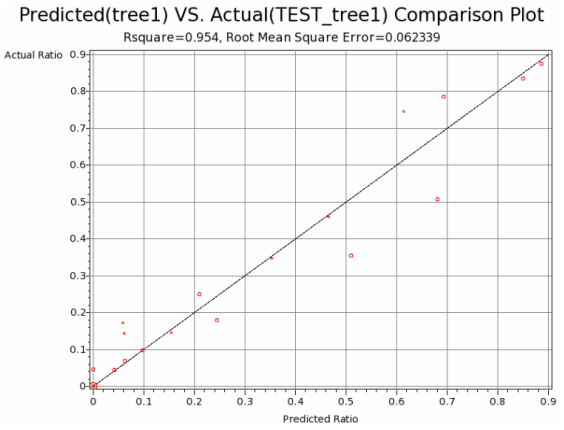
outhtml=dmt_compareplot.html
分析結果図を保存するHTML出力ファイル名を指定します。
例：outhtml=out1.html,

outpath=
HTML図表出力ファイルの保存ディレクトリを指定します。このパラメータを指定しない場合（デフォルト）、HTMLファイルはSASディスプレイマネージャまたはWPSワークベンチの管理下に作成されます。outpath=指定を行う場合、値は必ずフルパスで指定する必要があります。なお、パス指定全体を引用符で囲んでも囲まなくてもかまいません。
例：outpath='G:¥temp'

11.2.7 実行例

例１：分類木の比較プロット

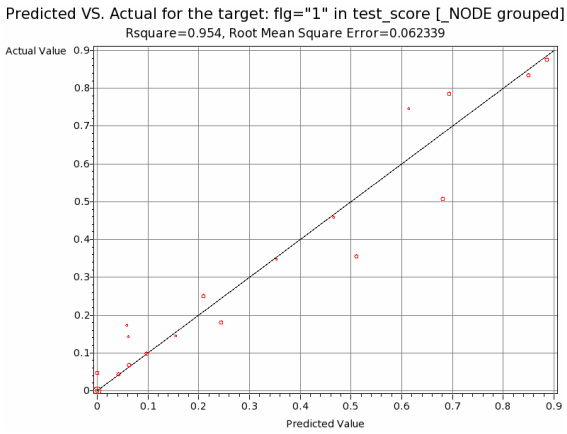
```
%dmt_tree(data=samp_data,y=flg,target=1,x=sei--DM,mincnt=50,maxlvl=10,outmodel=tree1)
%dmt_treescore(model=tree1,data=test_data,y=flg,target=1,outmodel=TEST_tree1)
%dmt_compareplot(model=tree1,test=TEST_tree1)
```



上図のようなプロット図を画面表示します。図の縦軸はターゲット実績値、横軸はモデルによる予測値を表します。プロット点は各終端ノードもしくはオブザベーションを表し、対角線上に乗っている場合、予測と実際が一致していることを表します。ノード単位の比較図におけるプロット点の円の大きさはノード件数の大きさを反映しています。

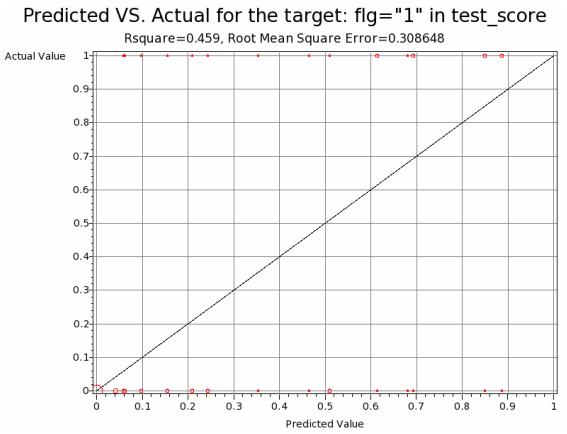
(**TIPS**) 検証データにモデルを適用した場合のモデル形式データセットの作成は、単純に予測値をつけるより、予測値をつけた後の処理（モデルのすべての中間ノードを含むノード別に予測値を集計する処理）が必要になるため、時間がかかります。以下のように予測値をつけたデータを作成し、そのデータを入力して予測値と実際値をノードグループ別(groupvar=_NODE)に集計した比較プロットを描く方が速く実行できます。（全く同じ比較プロットが表示されます）

```
%dmt_treescore(model=tree1,data=test_data,outscore=test_score,pred=PRED1)
%dmt_compareplot(data=test_score,y=flg,target=1,pred=PRED1,groupvar=_NODE)
```



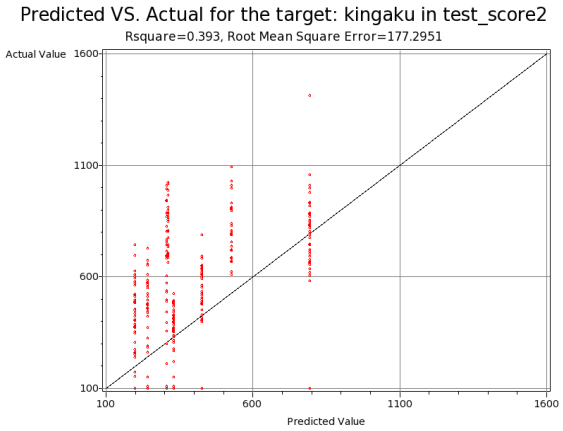
なお、上の例で groupvar=_NODEを指定しない場合、flgの実際値はflg=1（ターゲット出現）またはflg=0（非出現）のいずれかの2値しかとりません。グラフは以下のようになり、オブザベーション単位で誤差を計算し、集計しますので、R2乗値や誤差平均平方の値も変化します。

%dmt_compareplot(data=test_score,y=flg,target=1,pred=PRED1)



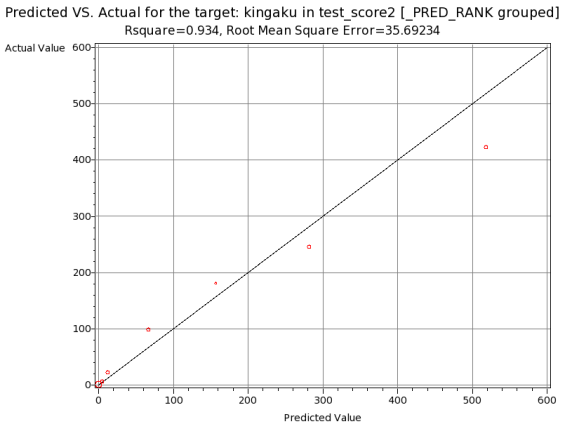
例２：回帰木の比較プロット

%dmt_tree(data=samp_data,y=kingaku,x=sei--DM,mincnt=50,maxlvl=10,outmodel=tree2)
%dmt_treescore(model=tree2,data=test_data,outscore=test_score2,pred=PRED2)
%dmt_compareplot(data=test_score2,y=kingaku,pred=PRED2,axis=100 to 1600 by 500)



ランキング指定を行うと以下のようになります。

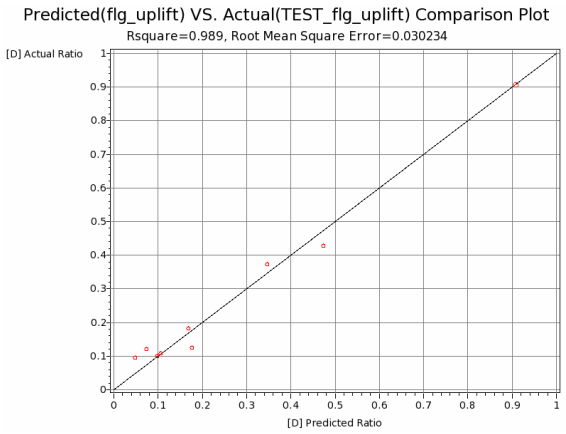
%dmt_compareplot(data=test_score2,y=kingaku,pred=PRED2,groupnum=10)

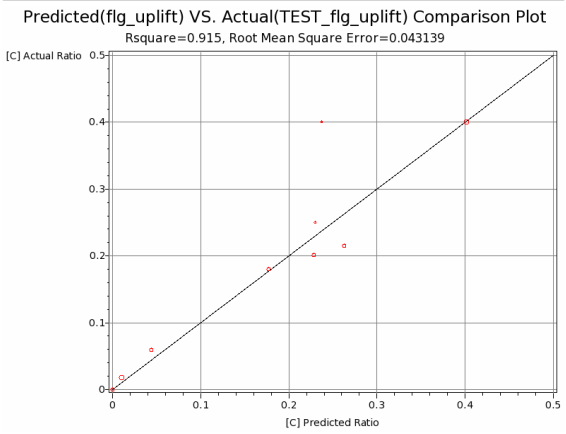


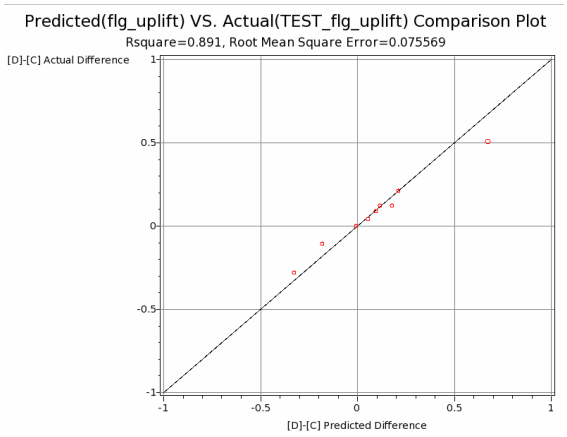
例３：アップリフトモデルの比較プロット

%dmt_tree(data=SAMP_DATA(where=(DM="1")),control=SAMP_DATA(where=(DM="0")),y=flg,target=1,x= sei nenrei jukyo kazoku_kosei gakureki shokushu kinmusaki gyoshu nenshu ,outmodel=flg_uplift,mincnt=100,maxlvl=5)
%dmt_treescore(data=TEST_DATA(where=(DM="1")),control=TEST_DATA(where=(DM="0")),model=flg_uplift,y=flg,target=1,outmodel=TEST_flg_uplift)

%dmt_compareplot(model=flg_uplift,test=TEST_flg_uplift)







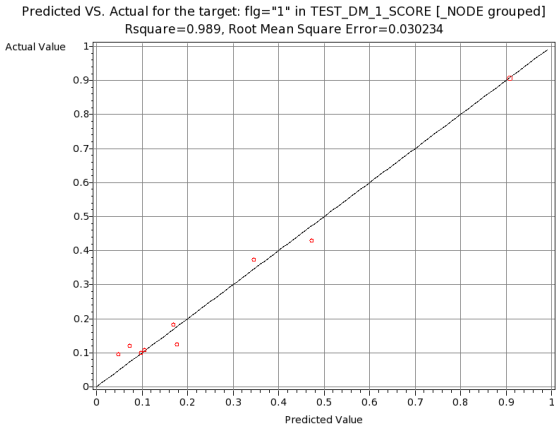
model=とtest=を指定すると、検証データにおける、処理した場合の予測値と実際の値のノード集計結果のプロット図がまず最初に表示されます。2番目のプロット図は対照群に対する予測と実際の比較です。そして、3番目の図は、処理した場合と処理しなかった場合の予測値の差と検証データにおける2群のノード実績値の差を各ツリーノードごとに比較したプロットとなっています。

(**TIP**) 同様の図を検証データに対するアップリフトモデルの処理した場合としなかった場合の2つの予測値をつけたデータセットを作成し、これを入力として作成することもできます。以下のように、

DMT_TREESCOREマクロでdata=,control=,pred=の各パラメータを変えながら2回実施し、その座標出力データセットを用いて、ノードごとのターゲット予測値および実績値の差分を計算し、GPLOTプロシジャで直接プロット図を作成します。

（処理群の検証データに対する比較プロット）

```
%dmt_treescore(data=TEST_DATA(where=(DM="1")
),control=TEST_DATA(where=(DM="0")),model=flg_
uplift,outscore=TEST_DM_1_SCORE,data_pred=dat
a_pred)
%dmt_compareplot(data=TEST_DM_1_SCORE,y=fl
g,target=1,pred=data_pred,groupvar=_NODE)
```

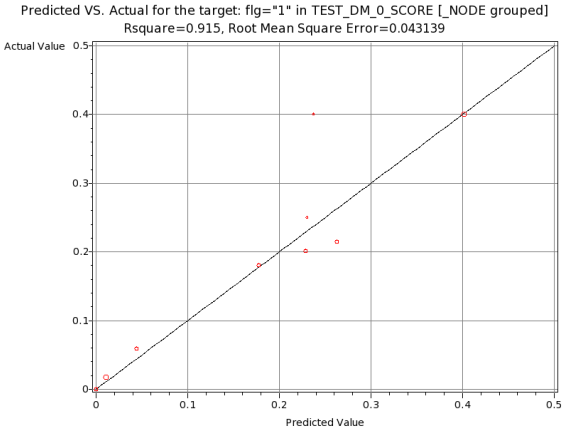


model=,test=指定の処理群の検証データの比較プロットと同じ図が得られます。
コマンド実行モードでは、座標データは _compare という固定の名前で出力されます。次のDMT_COMPAREPLOTの実行によって上書きされてしまわないように、_compare1にコピーしておきます。

```
data _compare1;set _compare;run;
```

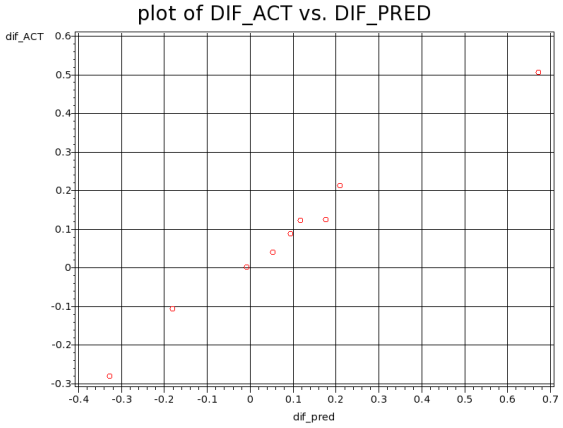
（対照群の検証データに対する比較プロット）

```
%dmt_treescore(data=TEST_DATA(where=(DM="0")
),control=TEST_DATA(where=(DM="1")),model=flg_
uplift,outscore=TEST_DM_0_SCORE,control_pred=
control_pred)
%dmt_compareplot(data=TEST_DM_0_SCORE,y=fl
g,target=1,pred=control_pred,groupvar=_NODE)
```



（処理群と対照群の差に対する比較プロット）
これは処理群、対照群それぞれの比較プロット座標出力データを利用してSAS言語で座標値を計算し、プロットします。

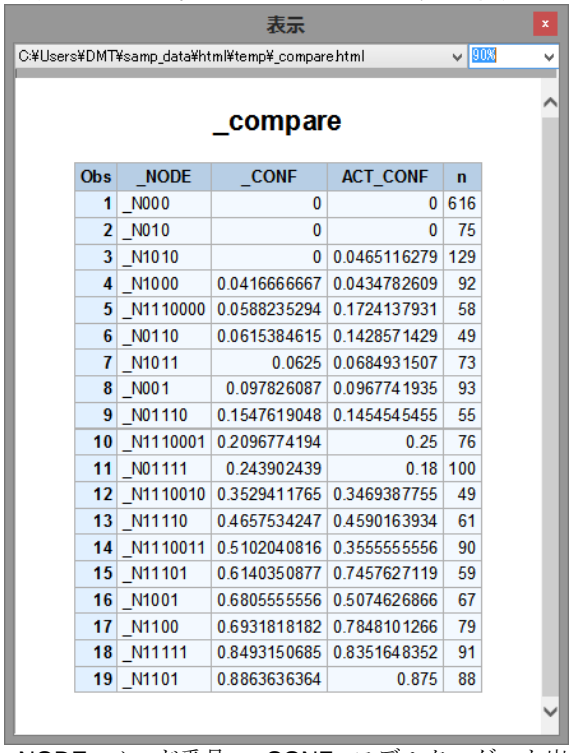
```
data _compare2;
   set _compare1(in=DM_1
rename=(ACT=data_ACT)) _compare(in=DM_0
rename=(ACT=control_ACT));
   if DM_1 then DM="1";
   else DM="0";
run;
proc means data=_compare2 nway noprint;
   class _NODE;
   var data_pred control_pred data_ACT
control_ACT;
   output out=out mean=;
run;
data uplift;
   set out;
   dif_pred=data_pred-control_pred;
   dif_ACT=data_ACT-control_ACT;
run;
title "plot of DIF_ACT vs. DIF_PRED";
symbol1 c=red v=circle;
proc gplot data=uplift;
   plot dif_ACT*dif_pred/autovref autohref;
run;
```



11.2.8 データセット出力

WORK._COMPAREにプロット点の座標値をデータセット出力します。（GUI実行モードでは座標値出力データに名前をつけられます）

比較プロットの座標値データの例（model=,test=指定）



_compare

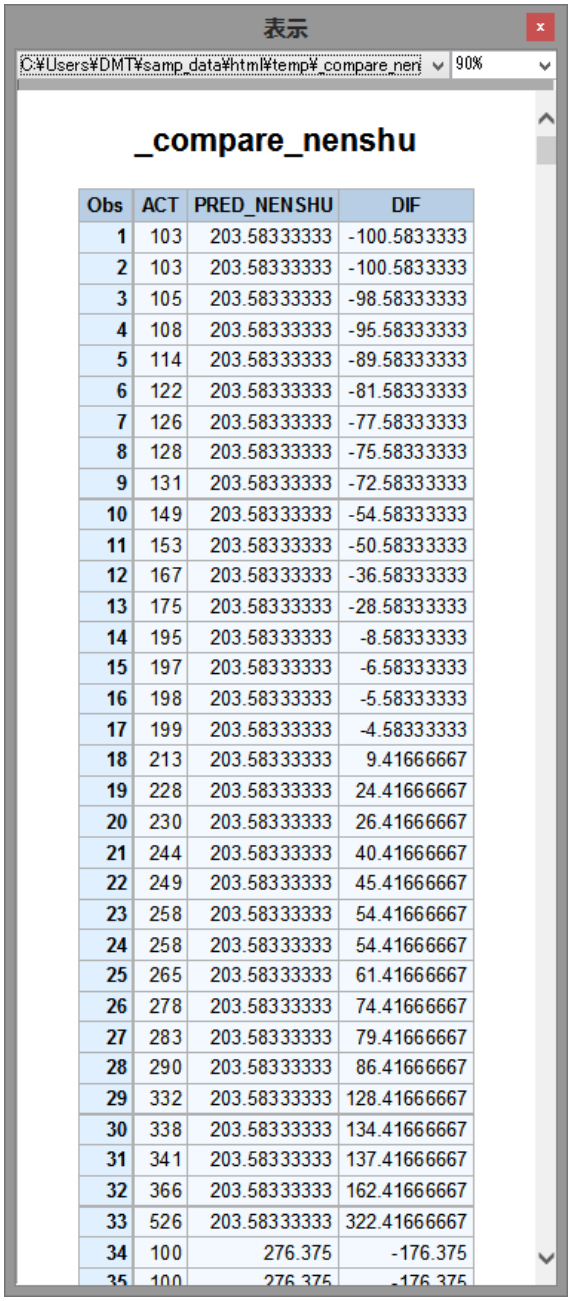
Obs	_NODE	_CONF	ACT_CONF	n
1	_N000	0	0	616
2	_N010	0	0	75
3	_N1010	0	0.0465116279	129
4	_N1000	0.0416666667	0.0434782609	92
5	_N1110000	0.0588235294	0.1724137931	58
6	_N0110	0.0615384615	0.1428571429	49
7	_N1011	0.0625	0.0684931507	73
8	_N001	0.097826087	0.0967741935	93
9	_N01110	0.1547619048	0.1454545455	55
10	_N1110001	0.2096774194	0.25	76
11	_N01111	0.243902439	0.18	100
12	_N1110010	0.3529411765	0.3469387755	49
13	_N11110	0.4657534247	0.4590163934	61
14	_N1110011	0.5102040816	0.3555555556	90
15	_N11101	0.6140350877	0.7457627119	59
16	_N1001	0.6805555556	0.5074626866	67
17	_N1100	0.6931818182	0.7848101266	79
18	_N11111	0.8493150685	0.8351648352	91
19	_N1101	0.8863636364	0.875	88

_NODE：ノード番号 ，_CONF：モデルターゲット出現率 ，ACT：実績ターゲット出現率 ，n：件数

data=パラメータを指定した場合は、オブザベーション単位のデータセットが出力されます。
GROUPVAR=、またはGROUPNUM=パラメータが指定されるとグループ単位の集計結果になり、識別変数（GROUPVARに指定した変数、GROUPNUMを指定した場合は、_PRED_RANK）が追加されます。

ただし、分類木の場合は、実績値は1か0しか存在しませんので、予測値と実績値の存在する組合せを集計した形で出力されます。（Nが追加されます）

比較プロットの座標値データの例（data=指定で回帰木でgroupvar=,groupnum=パラメータを指定しない場合）



_compare_nenshu

Obs	ACT	PRED_NENSHU	DIF
1	103	203.58333333	-100.5833333
2	103	203.58333333	-100.5833333
3	105	203.58333333	-98.58333333
4	108	203.58333333	-95.58333333
5	114	203.58333333	-89.58333333
6	122	203.58333333	-81.58333333
7	126	203.58333333	-77.58333333
8	128	203.58333333	-75.58333333
9	131	203.58333333	-72.58333333
10	149	203.58333333	-54.58333333
11	153	203.58333333	-50.58333333
12	167	203.58333333	-36.58333333
13	175	203.58333333	-28.58333333
14	195	203.58333333	-8.58333333
15	197	203.58333333	-6.58333333
16	198	203.58333333	-5.58333333
17	199	203.58333333	-4.58333333
18	213	203.58333333	9.41666667
19	228	203.58333333	24.41666667
20	230	203.58333333	26.41666667
21	244	203.58333333	40.41666667
22	249	203.58333333	45.41666667
23	258	203.58333333	54.41666667
24	258	203.58333333	54.41666667
25	265	203.58333333	61.41666667
26	278	203.58333333	74.41666667
27	283	203.58333333	79.41666667
28	290	203.58333333	86.41666667
29	332	203.58333333	128.41666667
30	338	203.58333333	134.41666667
31	341	203.58333333	137.41666667
32	366	203.58333333	162.41666667
33	526	203.58333333	322.41666667
34	100	276.375	-176.375
35	100	276.375	-176.375

11.2.9 欠損値の取り扱い

data=入力の場合、指定する予測変数の値のいずれかに欠損が存在するオブザベーションは計算から除外されます。
回帰モデルまたは回帰アップリフトモデルの場合は実績値に欠損があるオブザベーションも計算から除外されます。

11.2.10 制限

プロットする点の数は最大5000です。これを超える場合は、5000件のデータをランダム抽出して表示します。
ただし、R2乗値、誤差平均平方の値は全データから計算された値が表示されます。

11.2.11 コマンド実行モードでの注意

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

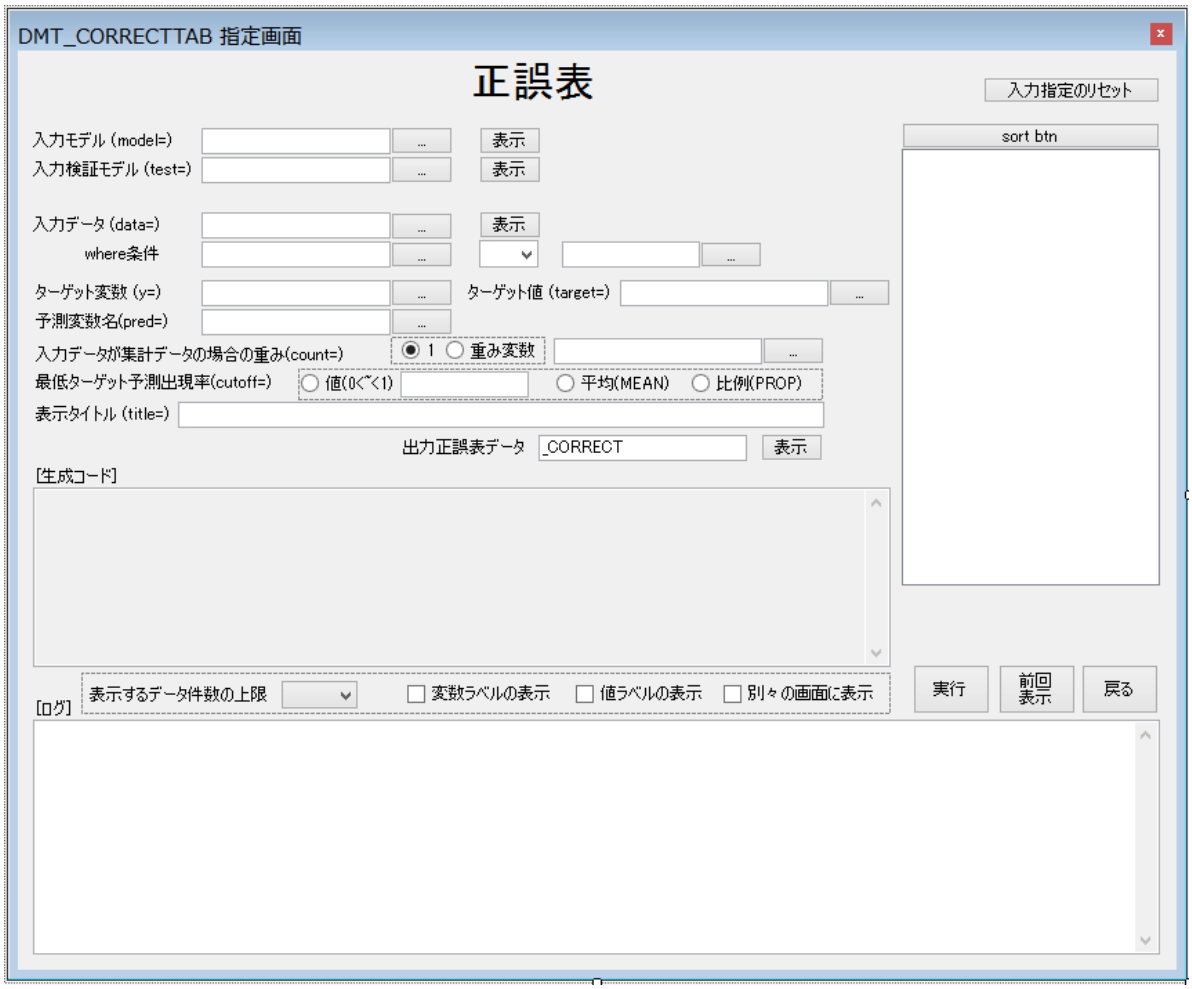
また、以下のユーザ定義フォーマットがWORKライブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
$_item
```

さらに、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
e_name e_type nobs lab&i spc&i typ&i zketa
_speclen _specnum _errormsg
```

11.3 正誤表(dmt_correcttab)



11.3.1 概要

正誤表（DMT_CORRECTTAB）はモデルの予測ターゲット出現率の大きさによってターゲットが出現するか否かを予測する場合の予測と実績の正誤表を作成し、正答率を表示します。
このマクロは分類木モデルの精度を検証する場合のみ有効です。

正誤表とは、予測ターゲット出現率の大きい順にならべたオブザベーションを、ある予測ターゲット出現率をしきい値として、しきい値以上の予測ターゲット出現率のオブザベーションはすべて「正予測(Positive Prediction)」（ターゲット出現）と予測し、しきい値未満の予測ターゲット出現率のオブザベーションはすべて「負予測(Negative Prediction)」（ターゲット非出現）と予測したときに、予測の正負別と実際のターゲット出現有無（正事例、負事例と呼びます）の件数をクロス集計した表のことです。

（正誤表(Confusion Matrix)）

予測	実際		計
	正事例	負事例	
正予測 （ターゲット出現と予測）	A 正予測真 （True Positive）	B 正予測偽 （False Positive）	正予測総件数 A+B
負予測 （ターゲット非出現と予測）	C 負予測偽 （False Negative）	D 負予測真 （True Negative）	負予測総件数 C+D
計	正事例総件数 A+C	負事例総件数 B+D	全体件数 N

上記正誤表において、予測ターゲット出現率のしきい値を変化させることにより、正予測総件数と負予測総件数を増減させることができます。しかし、正事例総件数と負事例総件数はしきい値とは無関係です。4つのセル（A 正予測真、B 正予測偽、C 負予測偽、D 負予測真）の件数（A,B,C,D）を用いて、正答率、ターゲット出現率、ターゲット再現率は以下のように定義されます。

正答率＝(A+D)/N
ターゲット出現率=A/(A+B)
ターゲット再現率=A/(A+C)

11.3.2 指定方法

（コマンド実行モードでの指定）

```
%dmt_correcttab(help,data=,y=,target=
,pred=_CONF,cutoff=0.5,count=1,model=,test=
,title=,language=JAPANESE
,outhtml=dmt_correcttab.html,outpath=)
```

（GUI実行モードでの変更点）

・help は指定不可。
・count=1 に固定。
・出力正誤表データに名前を付けることができます。（コマンド実行モードでは _CORRECT 固定）

（入力データセットの個々のオブザベーションに付与された予測値の精度を評価する場合）

以下の3個のパラメータは必須指定です。

入力データ (data=) ... 入力データセット名の指定.
ターゲット変数 (y=) ... ターゲット変数名の指定.
ターゲット値 (target=) ... ターゲット値の指定.

以下の2個のパラメータはオプション指定です。(=の右辺の値はデフォルト値を表しています)

予測変数名 (pred=_CONF) ... 予測変数名の指定.(単一変数のみ指定可)
count=1 ... 入力データセットのオブザベーションが集計データである場合の重み変数の指定.集計データで無い場合は1を指定します.(GUI実行モードでは1固定)

（1つのツリーモデルをテストデータに適用した場合の予測値の精度を評価する場合）

以下の2個のパラメータを同時に指定します。
（model=データセットは予測値、test=データセットは実績値をそれぞれ計算するために用いられます.）

入力モデル (model=) ... 入力モデルデータセット名の指定.
入力検証モデル (test=) ... テストデータに対してモデルを適用したときのモデル形式データ

（その他のパラメータ）

help ... 指定方法のヘルプメッセージの表示.(コマンド実行モードでのみ有効)
最低ターゲット予測出現率 (cutoff=0.5) ... ターゲットが出現すると予測する予測出現率下限値の指定.0～1の値、または MEAN, PROP が指定できます。
表示タイトル (title=) ... 画面出力のタイトルの指定.
(%str,%nrstr,%bquote などの関数で囲んで指定すること)
言語 (language= %sysfunc(getoption(LOCALE)))
... 言語の選択
出力正誤表データ … 正誤表をデータ出力します。GUI実行環境では名前を指定できますが、コマンド実行モードでは _correct という固定の名前でWORKライブラリに自動出力されます。

11.3.3 パラメータの詳細

入力モデル (model=)
入力モデルデータセット名を指定します。この指定は単独でもtest=パラメータと一緒に指定することも可能です。単独に指定した場合はモデルの精度を検証します。test=と共に指定した場合はモデル予測出現率の参照のために用いられ、test=モデル形式データセットにおけるモデルの精度を検証します。
例：model=bunseki1

入力検証モデル (test=)
精度を検証する対象の入力モデル形式データセット名を指定します。この指定はmodel=パラメータと一緒に指定する必要があります。
例：test=kensho1

入力データ (data=)
入力データセット名を指定します。例：data=a

ターゲット変数 (y=)
data= 入力データセットに含まれるターゲット変数名を指定します。 例：y=flag

ターゲット値 (target=)
data= 入力データセットに含まれるターゲット変数のターゲット値を指定します。
ターゲット変数が文字タイプの場合は1種類の値を指定します。特殊な文字（+,-など）を含まない限り引用符で囲む必要はありません。ターゲット変数が数値タイプの場合は1種類の値、もしくはあるしきい値を境とした「以上」、「以下」、「超」、「未満」のいずれかの範囲を指定可能です。数値変数タイプで範囲を指定する場合は引用符で囲んではいけません。

予測変数名 (pred=_CONF)
入力データセットに含まれる予測ターゲット出現率を表す変数名を指定します。デフォルトは _CONF です。

最低ターゲット予測出現率 (cutoff=0.5)
指定の値以上のターゲット予測出現率を持つ終端ノードまたはオブザベーションはすべてターゲット出現、それ以外の終端ノードまたはオブザベーションはターゲット非出現とみなした正誤表を作成します。デフォルトは0.5に設定しています。model=および test= パラメータと共に指定する場合、cutoff値は 0～1 の範囲の数値、または cutoff=MEAN（モデルの全体出現率（ルートノードのターゲット出現率）をしきい値に設定） または cutoff=PROP（モデルの全体出現率（ルートノードのターゲット出現率）と同

じ割合に近い正予測件数が得られるターゲット出現率をしきい値に設定） が指定できます.
例：cutoff=0.1

表示タイトル (title=)
画面出力される表にタイトルを指定できます。指定しない（デフォルト）場合、以下のようなタイトルが自動的に付与されます。

%bquote(&data **におけるモデルの正誤表(ターゲット:"&target", 予測出現率の下限=&cutoff.)**)

タイトルを指定する場合、上記のように%bquote関数の中に記述してください。

言語(language=JAPANESE)
分析実行中のメッセージ出力、結果の表のタイトル、表項目などの表示言語を選択します。ただし、現バージョンでは、日本語か英語の2種類のみ選択可能です。
例：language=ENGLISH

11.3.4 GUI 実行モードで有効なパラメータの詳細

出力正誤表データ
正誤表を出力するデータセットに名前をつけます。（コマンド実行モードでは、WORKライブラリに _correct という決まった名前で自動出力されます。）

11.3.5 コマンド実行モードで有効なパラメータの詳細

count=1
data= 入力データセットのオブザベーションが集計データである場合の重み変数名を指定します. 集計データではない通常の場合はデフォルトcount=1のままにしておきます. なお、重み変数名をこのパラメータで指定する場合、pred= パラメータに指定可能な予測値の数は1個のみになります。（**GUI実行モードでは指定不可**）例：count=freq

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。（GUI 実行モードでは指定できません。） 例：%dmt_correcttab(help)

11.3.6 HTML 出力

分析結果の図表はhtmlファイルに出力されます。保存先はデフォルトではSASディスプレイマネージャまたはWPSワークベンチの管理下（ワークスペース内の一時保存ファイル）です。outpath=パラメータを指定すると、保存先を変更できます。（必ずフルパス指定します。引用符で囲んでも囲まなくてもかまいません）同時にouthtml=パラメータを指定すると、保存するhtmlファイルに自由に名前を付けることができます。

outhtml=dmt_correcttab.html

分析結果図を保存するHTML出力ファイル名を指定します。
例：outhtml=out1.html,

outpath=
HTML図表出力ファイルの保存ディレクトリを指定します。このパラメータを指定しない場合（デフォルト）、HTMLファイルはSASディスプレイマネージャまたはWPSワークベンチの管理下に作成されます。outpath=指定を行う場合、値は必ずフルパスで指定する必要があります。なお、パス指定全体を引用符で囲んでも囲まなくてもかまいません。
例：outpath='G:¥temp'

11.3.7 実行例

例１：cutoff=0.5（デフォルト）を指定した場合

```
%dmt_tree(data=samp_data,y=flg,target=1,x=sei nenrei jukyo kazoku_kosei gakureki kinmusaki gyoshu shokushu nenshu DM,mincnt=50,maxlvl=10,outmodel=tree1)
%dmt_treescore(model=tree1,data=test_data,y=flg,target=1,outmodel=TEST_tree1)
%dmt_correcttab(model=tree1,test=TEST_tree1,cutoff=0.5)
```

モデル tree1 の テスト TEST_tree1 における正誤表, 予測出現率の下限=0.5
正答率= 86.00%

	実績フラグ					
	1.ターゲット		2.非ターゲット		計	
	件数	%	件数	%	件数	%
予測フラグ						
1.ターゲット	325	16.25	149	7.45	474	23.70
2.非ターゲット	131	6.55	1,395	69.75	1,526	76.30
計	456	22.80	1,544	77.20	2,000	100.00

上図のような画面出力を行います。

例２：cutoff=MEANを指定した場合

```
%dmt_correcttab(model=tree1,test=TEST_tree1,cutoff=mean)
```

モデル tree1 の テスト TEST_tree1 における正誤表, 予測出現率の下限=0.2285
(モデル予測出現率の平均値)
正答率= 81.80%

	実績フラグ					
	1.ターゲット		2.非ターゲット		計	
	件数	%	件数	%	件数	%
予測フラグ						
1.ターゲット	388	19.40	296	14.80	684	34.20
2.非ターゲット	68	3.40	1,248	62.40	1,316	65.80
計	456	22.80	1,544	77.20	2,000	100.00

例３：cutoff=PROPを指定した場合

```
%dmt_correcttab(model=tree1,test=TEST_tree1,cutoff=prop)
```

モデル tree1 の テスト TEST_tree1 における正誤表, 予測出現率の下限=0.5102040816(モデルのターゲット出現率 0.2285 に最も近いしきい値)
正答率= 86.00%

	実績フラグ					
	1.ターゲット		2.非ターゲット		計	
	件数	%	件数	%	件数	%
予測フラグ						
1.ターゲット	325	16.25	149	7.45	474	23.70
2.非ターゲット	131	6.55	1,395	69.75	1,526	76.30
計	456	22.80	1,544	77.20	2,000	100.00

例４：data=,y=,target=,pred=を指定する場合

```
%dmt_treescore(model=tree1,data=test_data
,outscore=TEST_score1,pred=SCORE1)
%dmt_correcttab(data=TEST_score1,y=flg,ta
rget=1,pred=SCORE1)
```

TEST_score1 におけるモデルの正誤表(ターゲット: "1"), 予測出現率の下限=0.5
正答率= 86.00%

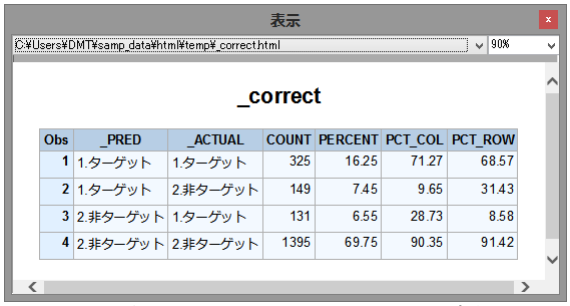
	実績フラグ					
	1.ターゲット		2.非ターゲット		計	
	件数	%	件数	%	件数	%
予測フラグ						
1.ターゲット	325	16.25	149	7.45	474	23.70
2.非ターゲット	131	6.55	1,395	69.75	1,526	76.30
計	456	22.80	1,544	77.20	2,000	100.00

注意：data=指定の場合は、モデルのターゲット出現率に関する情報が利用できないため、cutoff=MEAN, cutoff=PROP は指定できません。

11.3.8 データセット出力

WORK._CORRECTという名前のデータセットに予測と事例（実績）、ターゲット出現とターゲット非出現の2カテゴリ*2カテゴリのクロス集計結果をデータセットに出力します。（GUI実行モードでは名前を変更できます。）

正誤表出力データの例



_PRED：予測のカテゴリ ，_ACUTUAL：事例のカテゴリ ，COUNT：件数 ，PCT_ROW：行百分率 ，PCT_COL：列百分率

11.3.1 欠損値の取り扱い

data=入力の場合、実績値または予測値のいずれかに欠損が存在するオブザベーションは計算から除外されます。

11.3.2 コマンド実行モードでの注意

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

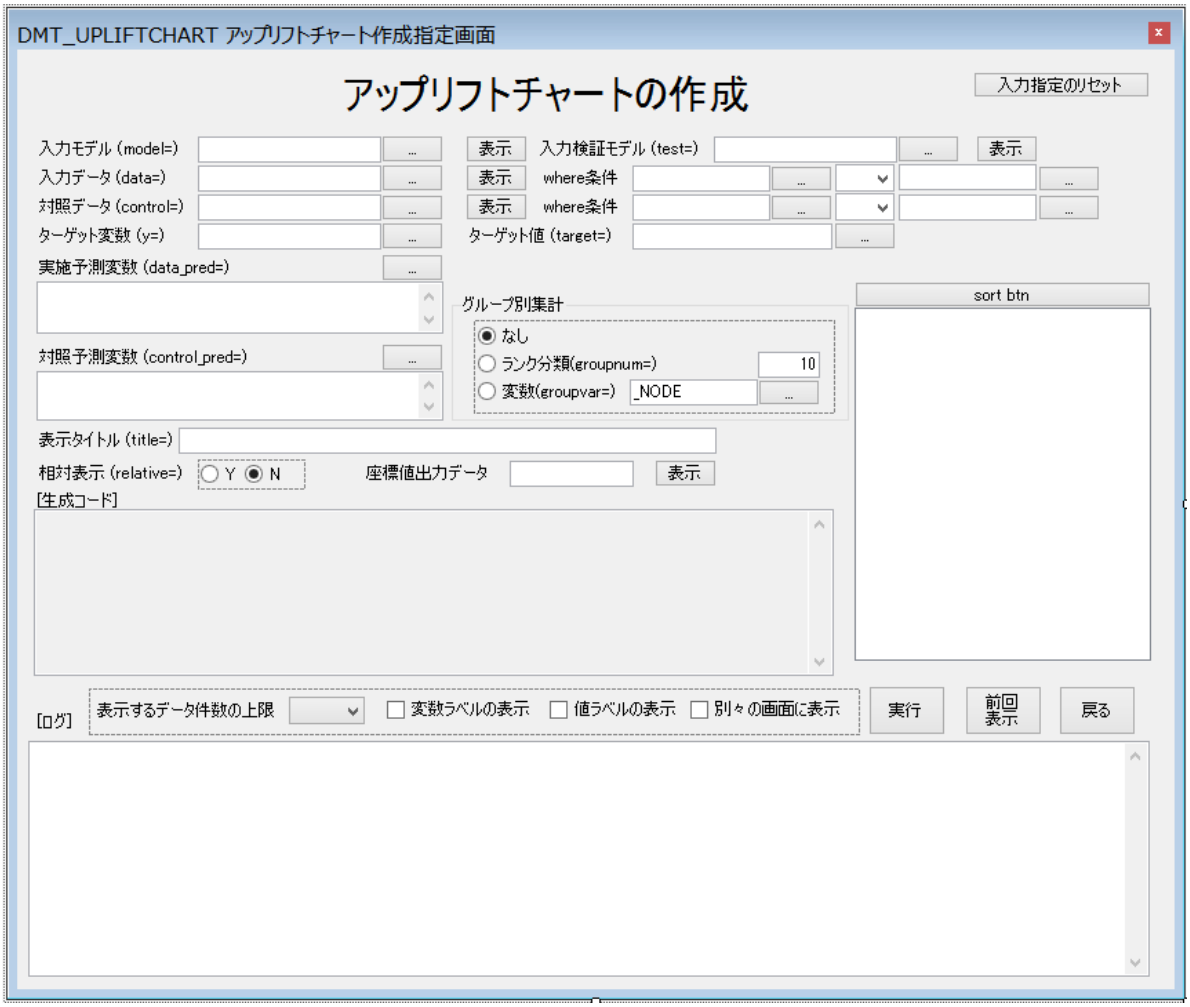
また、以下のユーザ定義フォーマットがWORKライブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
$_item
```

さらに、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
e_name e_type nobs lab&i spc&i typ&i zketa
_speclen _specnum _errormsg
```

11.4 アップリフト図（dmt_upliftchart）



11.4.1 概要

アップリフトチャート（DMT_UPLIFTCHART）は、施策実施効果（ターゲット出現率またはターゲット平均値の施策実施による増加量）を基準として作成したツリーモデル（分類木アップリフトモデル、または回帰木アップリフトモデル）の判別結果を図示します。施策実施群（処理群）、施策非実施群（対照群）それぞれについて、施策実施効果の大きい順（対照群においては逆順）にノードを並べておき、横軸を累積件数、縦軸を累積増加応答量（累積アップリフト）にとった点を結んだグラフがアップリフトチャートです。グラフの上部に累積アップリフト最大となる各モデルの選択件数と期待累積アップリフトの値を表示します。

施策実施群（処理群）のアップリフトチャートでは、アップリフトを「施策を実施した場合の期待応答量－実施しなかった場合の期待応答量」として定義しています。全データに対する累積アップリフト値は、現行の施策実施群全体のアップリフトを表します。もしも、このときの累積アップリフト値が最大となっているときは、現行の施策実施対象選択基準を次回も変更せずに利用すれば良いことを意味します。しかし、最後の方のいくつかのノードが累積アップリフトを押し下げている場合は、これらのノードに該当する顧客は次回の施策実施対象から除外すべきという判断になります。

施策非実施群（対照群）のアップリフトチャートでは、アップリフトを「施策を実施しなかった場合の期待応答量－実施した場合の期待応答量」として定義しています。したがって、累積アップリフト値が正の値の範囲に属するノードは、次回も同様に対照群に属させるべきです。逆に、累積アップリフト値が負の値のノードは、次回は施策対象に含めるべきという判断になります。

11.4.2 指定方法

（コマンド実行モードでの指定）

%dmt_upliftchart(help,data=,control=,y=,target=
,data_pred=,control_pred=
,model=,test=
,groupvar=,groupnum=,relative=N
,amountf=comma16.,pctf=7.2
,d_label=[D].c_label=[C],dif_label=[D]-[C]
,dev=GIF,title=,language=JAPANESE
,graph language=ENGLISH
,outhtml=dmt_upliftchart.html,outpath=)

（GUI実行モードでの変更点）

・help は指定不可。
・座標出力データに名前を付けることができる。（コマンド実行モードでは _UPLIFT 固定）

（入力データセットの個々のオブザベーションに付与された予測値に基づいてグラフを描く場合）

以下の2個のパラメータは常に必須指定です。

実施時の予測変数 (data_pred=) ... 処理群予測変数名の指定.
非実施時の予測変数 (control_pred=) ... 対照群予測変数名の指定.

実際値との比較を行わないで予測値のみを表示する場合は、以下の2つのうち、少なくとも1つの入力データ指定が必須です。いずれも where=(条件式) などのデータセットオプションを指定可能

入力データ (data=) … 処理群の入力データセット名の指定
入力対照データ (control=) … 対照群の入力データセット名の指定

予測値と実際値との比較グラフを描く場合は、上記の data=パラメータとcontrol=パラメータは両方必須です。さらに、分類木アップリフトもしくは回帰木アップリフトのいずれかによって、以下の1個または2個のターゲットに関するパラメータも必須です。ただし、回帰木アップリフトモデルの場合は target=パラメータは指定してはいけません。

ターゲット変数 (y=) ... ターゲット変数名の指定.
ターゲット値 (target=) ... ターゲット値の指定.

以下の3個のパラメータは予測値と実際値との比較グラフを描く場合に指定できるオプション指定です。（=の右辺の値はデフォルト値を表しています）

グループ単位の表示 (groupvar=)
予測値のランク単位の集計表示 (groupnum=)
相対表示 (relative=N)

（1つのツリーモデルを、モデル作成データのみ、またはモデルデータとテストデータ、それぞれに適用した場合を比較する場合）

以下の2個のパラメータを指定します。ただし、test=パラメータは単独指定できません。

入力モデル (model=) ... 入力モデルデータセット名の指定.
入力検証モデル (test=) ... テストデータに対してモデルを適用したときのモデル形式データ

（その他のパラメータ）

以下の10個のパラメータは任意指定です。（=の右辺の値はデフォルト値を表しています）

help ... 指定方法のヘルプメッセージの表示.（コマンド実行モードでのみ有効）
アップリフト値の表示フォーマットの指定 (amountf=comma16.)
百分率の表示フォーマットの指定 (pctf=7.2)
アップリフトモデルにおける処理群（DATA）を表す記号 (d_label=[D])
アップリフトモデルにおける対照群（Control）を表す記号 (c_label=[C])
アップリフトモデルにおける処理群−対照群間の差を表す記号 (dif_label=[D]-[C])
表示タイトル (title=) ... 画面出力のタイトルの指定.(%str,%nrstr,%bquote などの関数で囲んで指定すること)
言語 (language=JAPANESE) ... ログやメッセージを表示する言語の選択
グラフ表示言語 (graph_language=ENGLISH) ... ログやメッセージを表示する言語の選択
グラフデバイスの指定 (dev=GIF) ... グラフィックデバイスの指定.
HTML出力ファイル名 (outhtml=dmt_upliftchart.html)（コマンド実行モードでのみ有効）
HTMLファイル出力ディレクトリの指定 (outpath=)（コマンド実行モードでのみ有効）

座標値出力データ … 図の座標値をデータ出力します。GUI実行環境では名前を指定できますが、コマンド実行モードでは ゲインチャートの場合 _gain, ROCチャートの場合 _roc, 収益チャートの場合 _profit という固定の名前でWORKライブラリに自動出力されます。

11.4.3 パラメータの詳細

入力モデル (model=)
入力モデルデータセット名を指定します。
例：model=bunseki1

入力検証モデル (test=)
入力モデル形式データセット名を指定します。この指定はmodel=パラメータと一緒に指定する必要があります。
例：test=kensho1

入力データ (data=)
処理群として施策実施を行った（または行う予定の）入力データセット名を指定します。データセットオプションを指定できます。data=を指定する場合は、同時に、y=, target=(必要であれば), pred=の指定が必須です。
例：data=a, data=a(where=(DM="1"))

入力対照データ (control=)
対照群として施策実施を行わなかった（または行わない予定の）入力データセット名を指定します。データセットオプションを指定できます。data=を指定する場合は、同時に、y=, target=(必要であれば), pred=の指定が必須です。
例：data=a, data=a(where=(DM="1"))

ターゲット変数 (y=)
data= 入力データセットに含まれるターゲット変数名を指定します。 例：y=flag, y=revenue

ターゲット値 (target=)
分類木モデルの予測値と実績値を比較検証する場合、y= ターゲット変数のターゲット値を指定します。回帰木モデルの検証を行う場合は指定してはいけません。
例：target="1"
なお、引用符で囲まなくても構いません。（自動判断します）

実施時の予測変数 (data_pred=)
施策を実施した場合のターゲット予測値を表す変数名を指定します。このパラメータは常に必須です。（1個～4個までの別々のモデルによる予測変数名をスペースで区切って指定可）
なお、GUI実行モードでは、D_CONF が分類木アップリフトモデルの場合の、D_MEANが回帰木アップリフトモデルの場合のデフォルトとなっています。
例：data_pred=m1_go_pred m2_go_pred

非実施時の予測変数 (control_pred=)
施策を実施しなかった場合のターゲット予測値を表す変数名を指定します。このパラメータは常に必須です。（1個～4個までの別々のモデルによる予測変数名をスペースで区切って指定可。複数のモデルの予測変数を指定する場合は、モデルの指定をdata_pred=の指定順とcontrol=の指定順を対応させてください）

なお、GUI実行モードでは、D_CONF が分類木アップリフトモデルの場合の、D_MEANが回帰木アップリフトモデルの場合のデフォルトとなっています。
例：data_pred=m1_go_pred m2_go_pred

グループ単位の表示 (groupvar=)
data=、または control=指定の場合に、入力データに含まれる変数を1個だけ指定します。指定すると、チャートのプロット点が個々のオブザベーション単位から指定変数値が同じグループ単位の表示に変更されます。

予測値のランク単位の集計表示 (groupnum=)
data=、または control=指定の場合に、正の整数値を指定します。オブザベーションを予測値の大きさに基づくランクにグループ化（ビニングとも呼ばれる）し、ランクグループ単位の表示に変更します。

アップリフト値の表示フォーマットの指定
(amountf=comma16.)
チャートの上部に表示される累積アップリフト値の表示フォーマットを指定します。

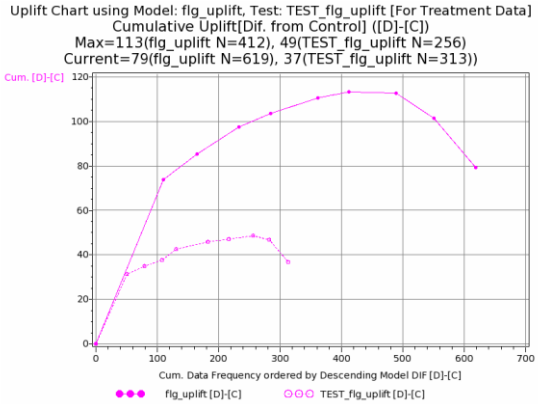
百分率の表示フォーマットの指定 (pctf=7.2)
relative=Y を指定した場合のチャートの上部に表示される件数比率などの表示フォーマットを指定します。

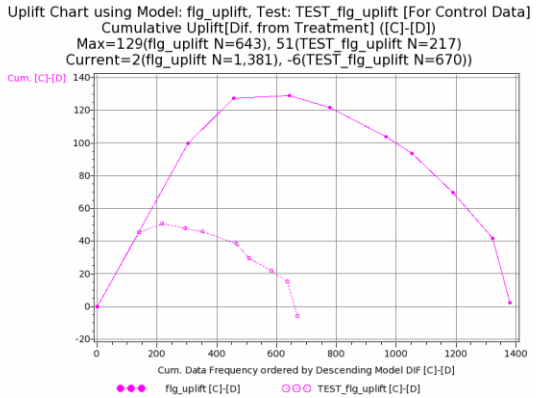
グラフ画面表示言語 (graph_language=ENGLISH)
グラフィック出力画面に表示する既定のタイトルや軸ラベル等に表示する言語を指定します。graph_language=ENGLISH が既定です。※ 現行WPSではグラフ上には日本語が表示できませんので、デフォルトの graph_language=ENGLISH を変更しないでください。

相対表示 (relative=N)
relative=Y を指定すると、チャートの縦軸、横軸を、絶対値の最大値が±100（符号は絶対値の最大値の符号）になるように比例変換して表示します。モデルデータと検証データの件数が異なる場合に指定するとモデルと検証を比較しやすい表示になります。

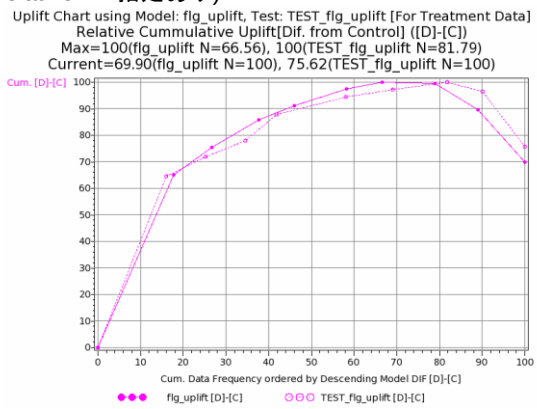
例：
```
%dmt_tree(data=SAMP_DATA(where=(DM="1")),control=SAMP_DATA(where=(DM="0")),y=flg,target=1,x=sei nenrei jukyo kazoku_kosei gakureki shokushu kinmusaki gyoshu nenshu ,mincnt=50,maxlvl=5,outmodel=flg_uplift)

%dmt_treescore(
data=TEST_DATA(where=(DM="1" and uniform(1)<0.5))
,control=TEST_DATA(where=(DM="0" and uniform(1)<0.5))
,model=flg_uplift,y=flg,target=1,outmodel=TEST_flg_uplift)

%dmt_upliftchart(model=flg_uplift,test=TEST_flg_uplift)

%dmt_upliftchart(model=flg_uplift,test=TEST_flg_uplift,relative=Y)
```

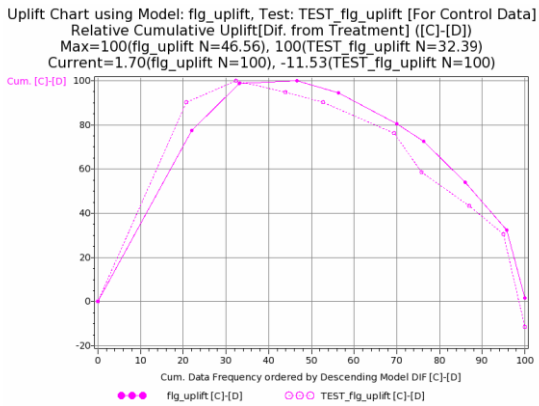
(relative= 指定なし)





(relative=Y 指定あり)





11.4.4 GUI 実行モードで有効なパラメータの詳細

座標値出力データ
図の座標値を出力するデータセットに名前をつけます。(コマンド実行モードでは、WORKライブラリに決まった名前(type=指定によって、_gain, _roc,_profit のいずれか) で自動出力されます。)

11.4.5 コマンド実行モードで有効なパラメータの詳細

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。(GUI 実行モードでは指定できません。) 例：%dmt_gainchart(help)

11.4.6 HTML 出力

分析結果の図表はhtmlファイルに出力されます。保存先はデフォルトではSASディスプレイマネージャまたはWPSワークベンチの管理下（ワークスペース内の一時保存ファイル）です。outpath=パラメータを指定すると、保存先を変更できます。(必ずフルパス指定します。引用符で囲んでも囲まなくてもかまいません) 同時にouthtml=パラメータを指定すると、保存するhtmlファイルに自由に名前を付けることができます。

outhtml=dmt_upliftchart.html
分析結果を保存するHTML出力ファイル名を指定します。
例：outhtml=out1.html,

outpath=
HTML図表出力ファイルの保存ディレクトリを指定します。このパラメータを指定しない場合（デフォルト)、HTMLファイルはSASディスプレイマネージャまたはWPSワークベンチの管理下に作成されます。outpath=指定を行う場合、値は必ずフルパスで指定する必要があります。なお、パス指定全体を引用符で囲んでも囲まなくてもかまいません。
例：outpath='G:¥temp'

11.4.7 実行例

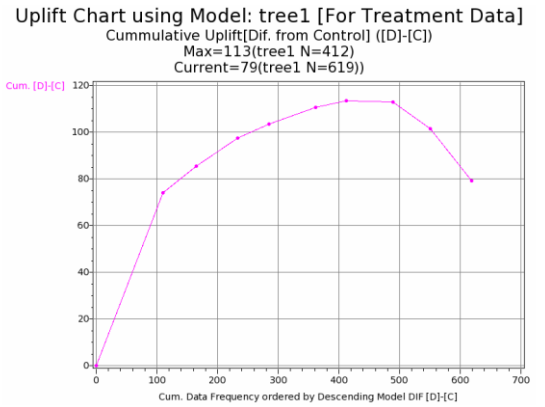
以下のように、DM送付効果を分析するために、samp_dataの中のDM送付先(DM="1")と非送付先(DM="0")における変数flg=1の出現率の差を基準とするアップリフトツリーモデル（tree1）を作成し、モデルを 検証用にとっておいた test_dataにあてはめて、モデル検証用のモデル形式データセット（TEST_tree1）を作成します。

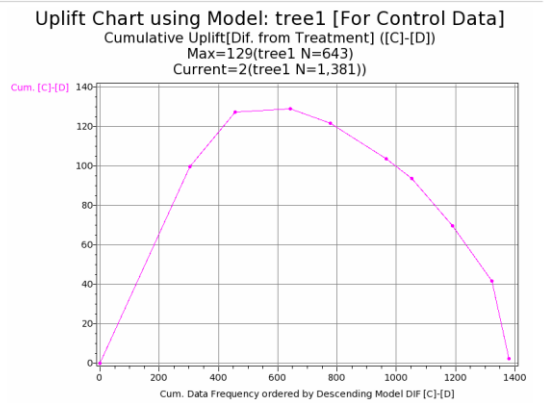
```
%dmt_tree(data=samp_data(where=(DM="1")),contr
ol=SAMP_DATA(where=(DM="0")),y=flg,target=1,x=
sei--nenshu,mincnt=50,maxlvl=10,outmodel=tree1)

%dmt_treescore(model=tree1
,data=TEST_DATA(where=(DM="1"))
,control=TEST_DATA(where=(DM="0"))
```

,y=flg,target=1,outmodel=TEST_tree1)

例１：モデルのアップリフトチャート
model=パラメータのみ指定します。

%dmt_upliftchart(model=tree1)





モデルに保存されている各ノードにおける処理群の場合の予測値と対照群の場合の予測値に基づき、処理群、対照群それぞれについて、他方の群との差の累積値（累積アップリフト＝Σ(差の期待値*ノード件数)）をプロットした図を表示します。

グラフの上部には、以下の情報が表示されます。

最大（Max）：累積アップリフト最大値とそのときの累積ノード件数
現行（Current）、全データの累積アップリフト値

上記の例では、以下のように結果を読み取ります。

処理群（全619件）のグラフでは、データはすべて実際に処理群に属しています。したがって、グラフの一番右端の累積アップリフト値79は、今回のDM送付先全部を送付したことの効果としてのターゲットの追加応答数の推定値を表しています。
しかし、tree1モデルを用いると、412件のデータに対してのみDM送付を行うと、累積アップリフトが113と最大になることが期待できることがわかります。
（つまり、113-79=34だけ応答数が増える）

一方、対照群（全1,381件）のグラフでは、データはすべて対照群に属しています。したがって、グラフの一番右端の累積アップリフト値2は、今回のDM非送付先全体を非送付としたことの効果としてのターゲットの追加応答数の推定値を表しています。
しかし、1,381件のうち非送付のままとすべき643件を除く738件を非送付ではなく送付としていたなら、129の追加応答が得られていたことがわかります。
（つまり、129-2=127応答数が増える）

なお、応答数の実績を集計すると、以下のとおり。

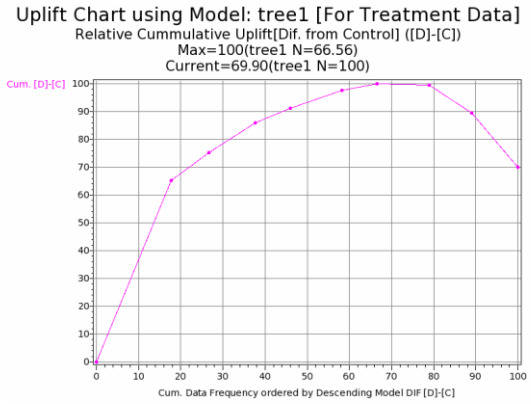
Table of DM by flg			
DM (プロモーション)	flg (購入有無)		
Frequency Row Pct	なし	あり	Total
非実施	1114 80.67	267 19.33	1381
実施	429 69.31	190 30.69	619
Total	1543	457	2000

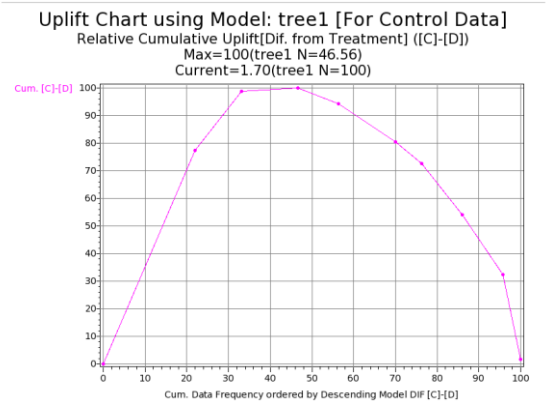
処理群、対照群を合わせると、現行とモデルを用いてDM送付先を最適化した場合を比較すると、以下のようになります。

	DM送付先			DM非送付先			合計		
	件数	応答数	応答率	件数	応答数	応答率	件数	応答数	応答率
現行	619	190	31%	1381	267	19%	2000	457	23%
モデル	1150	383	33%	850	164	19%	2000	547	27%
差	+ 531	+ 193	+ 3%	- 531	- 103	± 0%	±0	+ 90	+ 4%

例２：相対表示

%dmt_upliftchart(model=tree1,relative=Y)



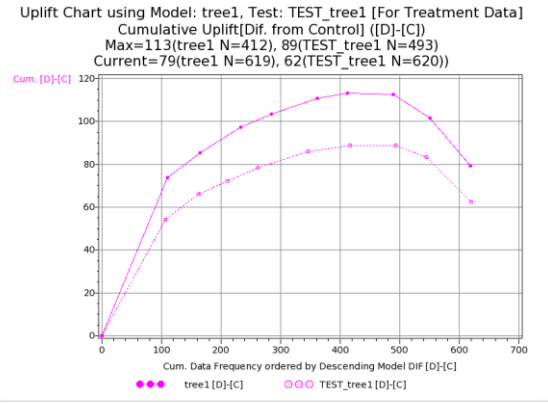


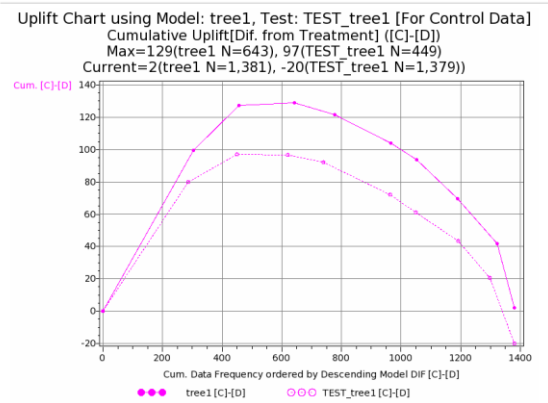
relative=Yを指定すると、プロット点は、絶対値で最大値の値が100または-100になるように、比例変換されます。相対図からは、アップリフト値や件数を絶対数ではなく、割合（百分率で表示）で読み取ることができます。

例えば、上記の処理群のグラフからは、最大アップリフト値を100として、これは66.56%の件数を選択した場合となり、現行の累積アップリフトは最大の69.9%の大きさであることを示しています。

例３：検証結果を加えたモデルのアップリフトチャート

%dmt_upliftchart(model=tree1,test=TEST_tree1)



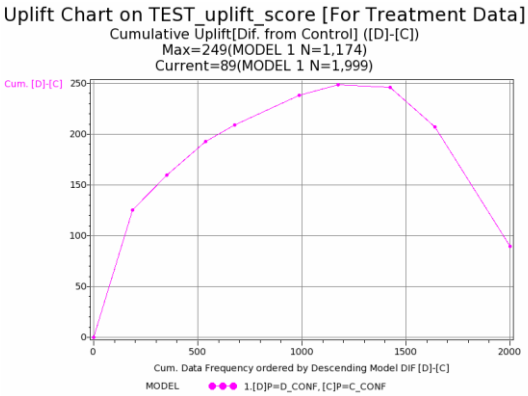


例４：個々のオブザベーションごとに予測スコアがつけられたデータセットを入力し、予測アップリフトを表示

%dmt_treescore(model=flg_uplift,data=TEST_DATA,outscore=TEST_uplift_score)

・すべてDM送付先として入力する場合

%dmt_upliftchart(data=TEST_uplift_score,data_pred=D_CONF,control_pred=C_CONF)

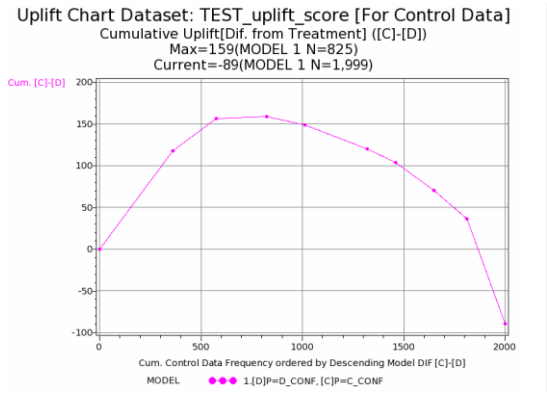


1,174件に対してのみ送付すると249の応答増加となることがわかります。最後の2つの終端ノードに該当する顧客には送付しない方が良いことが分かります。

注意：モデルをTEST_DATAにあてはめたときに1件予測できない欠損データが存在するため、N=1,999となっています。

・すべてDM非送付先とみなして入力する場合

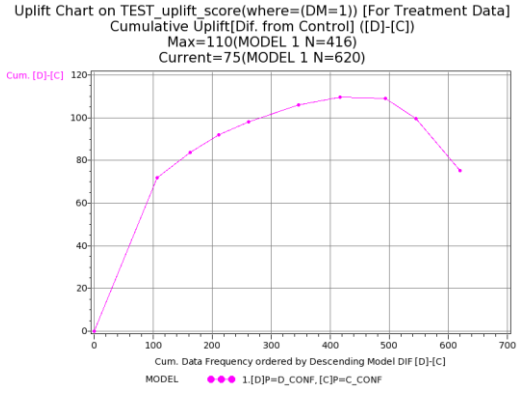
%dmt_upliftchart(control=TEST_uplift_score,data_pred=D_CONF,control_pred=C_CONF)

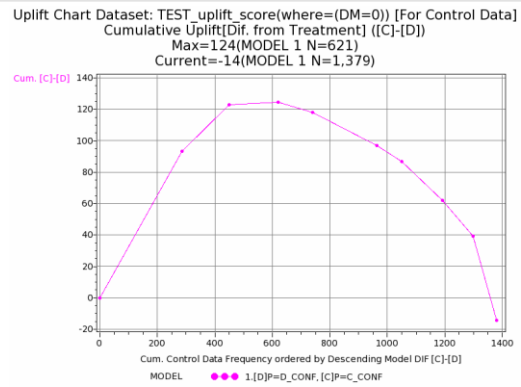


最初の3つのノードは非送付で良いが、残りは送付すべきです。すべて非送付の場合のアップリフト-89に対して最大アップリフト159となっており、その差は248となり、上記のすべて送付したとしたときのアップリフト計算結果と一致します。

・送付先データ、非送付先データの両方を入力する場合

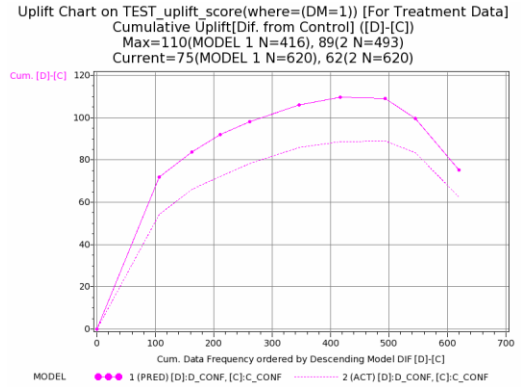
```
%dmt_upliftchart(data=TEST_uplift_score(where=(DM="1")),
control=TEST_uplift_score(where=(DM="0"))
,data_pred=D_CONF,control_pred=C_CONF)
```

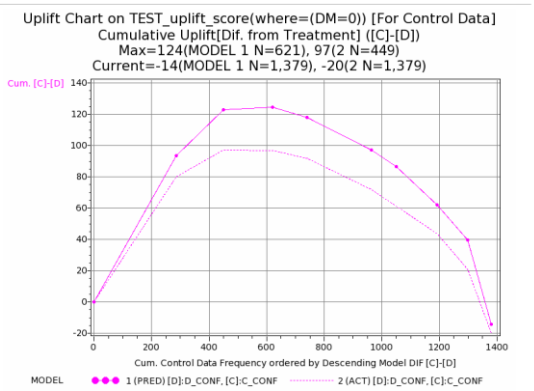




例５：個々のオブザベーションごとに予測スコアと実績値が付与されたデータセットを入力しての予測アップリフトの検証

```
%dmt_upliftchart(data=TEST_uplift_score(where=(DM="1")),
control=TEST_uplift_score(where=(DM="0"))
,data_pred=D_CONF,control_pred=C_CONF
,y=flg,target=1)
```





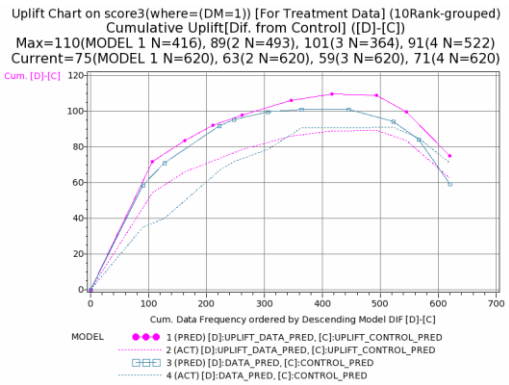
注意：実績値(y=,target=)との比較を行う場合は、処理群と対照群の実績応答差の計算のため、data=,control=の両方を指定しなければなりません。

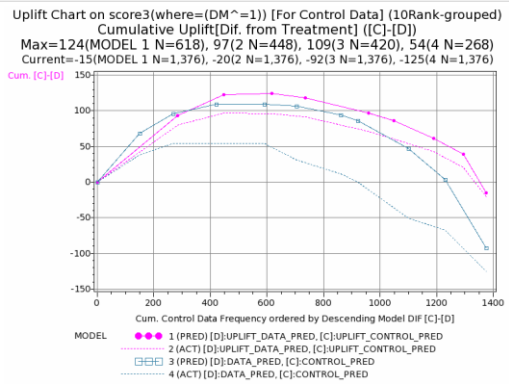
例６：複数のアップリフトモデルの予測と実績の比較

古典的なアップリフトモデルは、処理群と対照群の各データから別々にモデルを作成し、個人ごとに2種類のモデルの予測値を与え、それらの差をアップリフトの推計値とします。この方法で得られたアップリフトモデルと本アプリケーションのアップリフトモデルを比較します。

```
/*（古典的モデル）*/
/* (処理群応答率予測モデル) */
%dmt_tree(data=samp_data(where=(DM="1")),y=flg,
target=1,x=sei--nenshu,maxlvl=10,mincnt=100,outm
odel=D_model)

/* (対照群応答率予測モデル) */
%dmt_tree(data=samp_data(where=(DM="0")),y=flg,
target=1,x=sei--nenshu,maxlvl=10,mincnt=100,outm
odel=C_model)

/* (各モデルの予測値を検証データの各オブザベーションに付与) */
%dmt_treescore(data=test_data,model=D_model,pr
ed=data_pred,outscore=score1,)
%dmt_treescore(data=score1,model=C_model,pred
=control_pred,outscore=score2)

/*（新しいアップリフトモデル）*/
%dmt_tree(data=samp_data(where=(DM="1")),contr
ol=samp_data(where=(DM="0")),y=flg,target=1,x=sei
--nenshu,maxlvl=10,mincnt=50,outmodel=Uplift_mod
el)

/* (アップリフトモデルの予測値も検証データに付与) */
%dmt_treescore(data=score2,model=Uplift_model,d
ata_pred=uplift_data_pred,control_pred=uplift_contr
ol_pred,outscore=score3)
```

```
/* (古典モデルと新しいアップリフトモデルの比較)
*/
%dmt_upliftchart(data=score3(where=(DM="1")),con
trol=score3(where=(DM^="1")),data_pred=uplift_data
_pred data_pred,control_pred=uplift_control_pred
control_pred
,y=flg,target=1,groupnum=10)
```



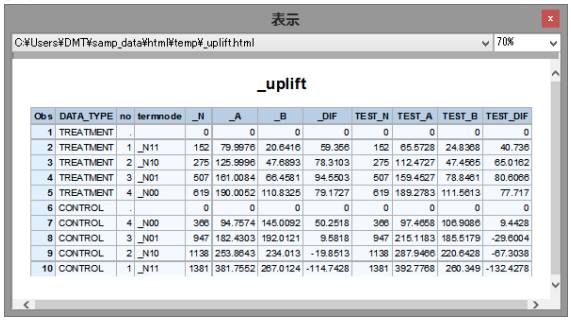


この結果では、新しいアップリフトモデルの方が最大アップリフトが大きく、検証結果との差が小さくなっています。（常にそうとは限りません）

11.4.8 データセット出力

WORK._UPLIFT にアップリフトチャートの座標値を格納したデータセットを自動出力します。
（GUI実行モードの場合は座標出力データに名前を付けることができます。）

アップリフトチャートの座標値データの例（model=,test=指定）



表示

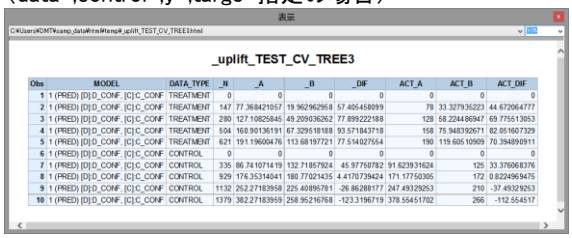
C:¥Users¥DMT¥samp_data¥html¥temp¥_uplift.html 70%

_uplift

Obs	DATA_TYPE	no	termnode	_N	_A	_B	_DIF	TEST_N	TEST_A	TEST_B	TEST_DIF
1	TREATMENT	.		0	0	0	0	0	0	0	0
2	TREATMENT	1	_N11	152	79.9976	20.6416	59.356	152	65.5728	24.8368	40.736
3	TREATMENT	2	_N10	275	125.9996	47.6893	78.3103	275	112.4727	47.4565	65.0162
4	TREATMENT	3	_N01	507	161.0084	66.4581	94.5503	507	159.4527	78.8461	80.6066
5	TREATMENT	4	_N00	619	190.0052	110.8325	79.1727	619	189.2783	111.5613	77.717
6	CONTROL	.		0	0	0	0	0	0	0	0
7	CONTROL	4	_N00	366	94.7574	145.0092	50.2518	366	97.4658	106.9086	9.4428
8	CONTROL	3	_N01	947	182.4303	192.0121	9.5818	947	215.1183	185.5179	-29.6004
9	CONTROL	2	_N10	1138	253.8643	234.013	-19.8513	1138	287.9466	220.6428	-67.3038
10	CONTROL	1	_N11	1381	381.7552	267.0124	-114.7428	1381	392.7768	260.349	-132.4278

DATA_TYPE：処理群、対照群の区別, no：ノードの予測応答差が大きい方からの順番, termnode：ノード番号, _N：累積件数, _A：処理群の場合の累積予測アップリフト, _B：対照群の場合の累積予測アップリフト, _DIF：累積アップリフト（処理群の場合は _A-_B, 対照群の場合は _B-_A）, TEST_N, TEST_A, TEST_B, TEST_DIF：test=指定の場合の検証データにおける数値

アップリフトチャートの座標値データの例
（data=,control=,y=,targe=指定の場合）



表示

C:¥Users¥DMT¥samp_data¥html¥temp¥_uplift_TEST_CV_TREE3.html

_uplift_TEST_CV_TREE3

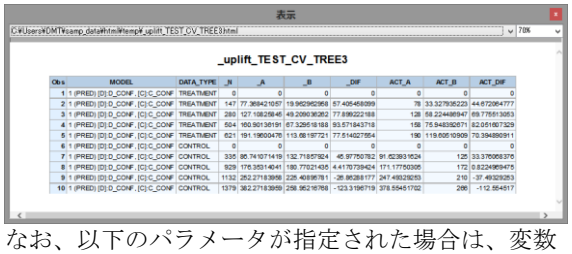
Obs	MODEL	DATA_TYPE	_N	_A	_B	_DIF	ACT_A	ACT_B	ACT_DIF
1	1 (PRED) [D]:D_CONF, [C]:C_CONF	TREATMENT	0	0	0	0	0	0	0
2	1 (PRED) [D]:D_CONF, [C]:C_CONF	TREATMENT	147	77.368421057	19.962962958	57.405458099	78	33.327936223	44.672064777
3	1 (PRED) [D]:D_CONF, [C]:C_CONF	TREATMENT	280	127.10825845	49.209636262	77.899222188	128	58.224486947	69.775513053
4	1 (PRED) [D]:D_CONF, [C]:C_CONF	TREATMENT	594	168.90136191	67.329518188	93.571843718	168	75.948392671	92.051607329
5	1 (PRED) [D]:D_CONF, [C]:C_CONF	TREATMENT	621	191.19600476	113.68197721	77.514027554	190	119.6051090	70.394890911
6	1 (PRED) [D]:D_CONF, [C]:C_CONF	CONTROL	0	0	0	0	0	0	0
7	1 (PRED) [D]:D_CONF, [C]:C_CONF	CONTROL	335	86.741071419	132.71857924	45.97750782	91.623931624	125	33.376068376
8	1 (PRED) [D]:D_CONF, [C]:C_CONF	CONTROL	929	176.35314041	180.77021435	4.4170739424	171.1750305	172	0.8224969475
9	1 (PRED) [D]:D_CONF, [C]:C_CONF	CONTROL	1132	252.27183958	225.40896781	-26.86288177	247.49329253	210	-37.49329253
10	1 (PRED) [D]:D_CONF, [C]:C_CONF	CONTROL	1379	382.27183959	258.36216768	-123.3196719	378.55451702	266	-112.554517

MODEL：モデル名（値にはモデル番号と指定された処理群、対照群の予測変数名のペア名がはいっています。）さらに、groupvar=を

data=入力データを指定したときは、変数 termnode は存在しません。その他、以下のように変更されます。

・モデル名が追加される
・GROUPVAR=パラメータを指定した場合はその変数が追加される
・GROUPNUM=パラメータを指定した場合は、変数 _RANK_NUMが追加される

オブザベーションは回帰アップリフトモデルの場合は入力オブザベーションごと。分類木アップリフトモデルの場合は、実績値が1か0



表示

C:¥Users¥DMT¥samp_data¥html¥temp¥_uplift_TEST_CV_TREE3.html 70%

_uplift_TEST_CV_TREE3

Obs	MODEL	DATA_TYPE	_N	_A	_B	_DIF	ACT_A	ACT_B	ACT_DIF
1	1 (PRED) [D]:D_CONF, [C]:C_CONF	TREATMENT	0	0	0	0	0	0	0
2	1 (PRED) [D]:D_CONF, [C]:C_CONF	TREATMENT	147	77.368421057	19.962962958	57.405458099	78	33.327936223	44.672064777
3	1 (PRED) [D]:D_CONF, [C]:C_CONF	TREATMENT	280	127.10825846	49.209636262	77.899222188	128	58.224486947	69.775513053
4	1 (PRED) [D]:D_CONF, [C]:C_CONF	TREATMENT	594	168.90136191	67.329518188	93.571843718	158	75.948392671	82.051607329
5	1 (PRED) [D]:D_CONF, [C]:C_CONF	TREATMENT	621	191.19600476	113.68197721	77.514027554	190	119.6051090	70.394890911
6	1 (PRED) [D]:D_CONF, [C]:C_CONF	CONTROL	0	0	0	0	0	0	0
7	1 (PRED) [D]:D_CONF, [C]:C_CONF	CONTROL	335	86.741071419	132.71857924	45.97750782	91.623931624	125	33.376068376
8	1 (PRED) [D]:D_CONF, [C]:C_CONF	CONTROL	929	176.35314041	180.77021435	4.4170739424	171.1750305	172	0.8224969475
9	1 (PRED) [D]:D_CONF, [C]:C_CONF	CONTROL	1132	252.27183958	225.40896781	-26.86288177	247.49329253	210	-37.49329253
10	1 (PRED) [D]:D_CONF, [C]:C_CONF	CONTROL	1379	382.27183959	258.36216768	-123.3196719	378.55451702	266	-112.554517

なお、以下のパラメータが指定された場合は、変数が追加されます。
・GROUPVAR=パラメータを指定した場合は指定した変数
・GROUPNUM=パラメータを指定した場合は _PRED_RANK

11.4.9 欠損値の取り扱い

data=入力の場合、いずれかの予測値に欠損が存在するオブザベーションは計算から除外されます。
回帰アップリフトモデルの場合は実績値が欠損のオブザベーションも計算から除外されます。

11.4.10 制限

data=入力データセットを指定し、オブザベーションに対する複数のモデルによる予測値を比較する場合にdata_pred=パラメータとcontrol_predに指定する予測変数ペア数の上限は4個までです。

11.4.11 コマンド実行モードでの注意

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

また、以下のユーザ定義フォーマットがWORKライブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
$_item
```

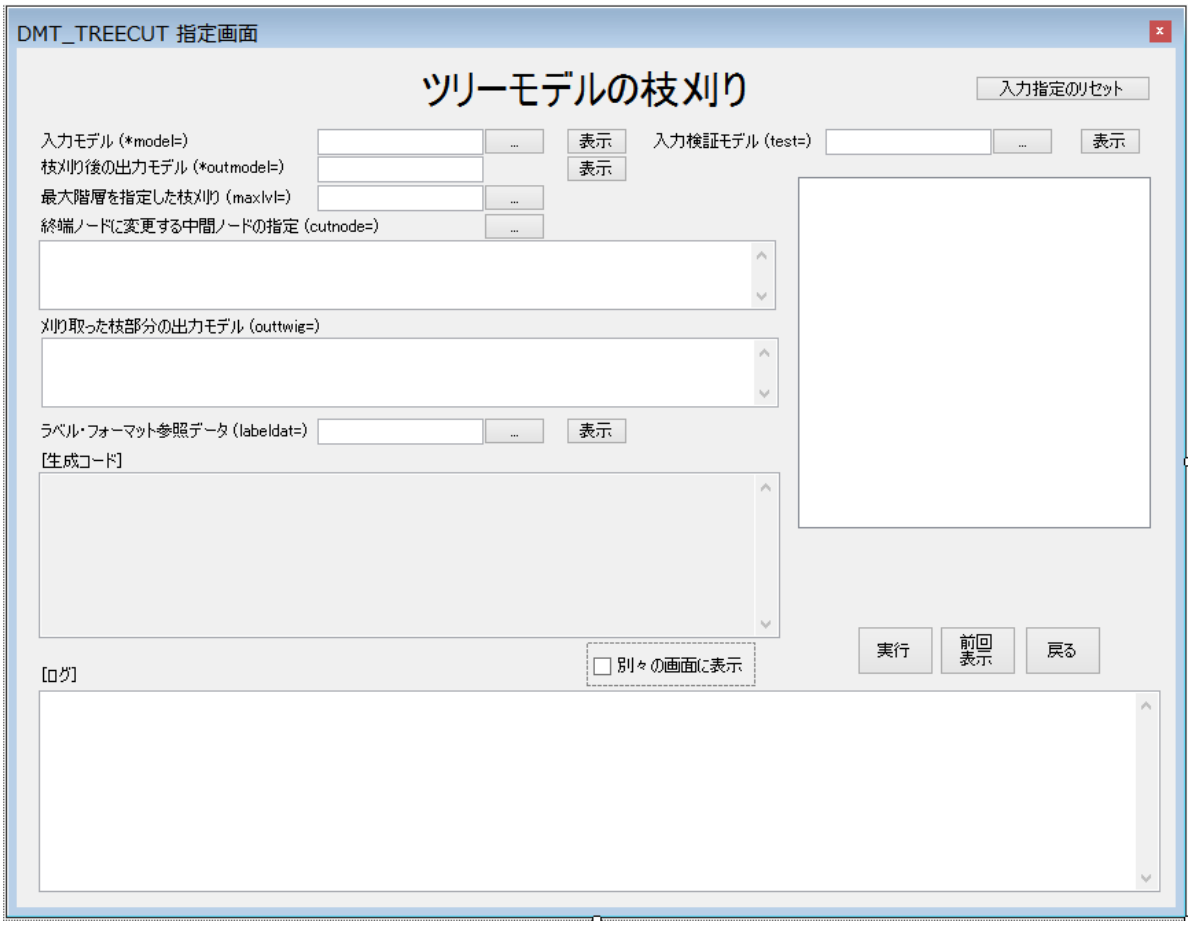
さらに、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
e_name e_type nobs lab&i spc&i typ&i zketa
_speclen _specnum _errormsg
```

12. 分析画面 ⑤モデル調整

既存ツリーモデルの構造の一部変更（枝刈り、枝接ぎ）や、モデル構造を変えずに検証データに基づくモデル予測値の修正を行います。

12.1 枝刈り（dmt_treecut）



12.1.1 概要

ツリーモデルの枝刈り（DMT_TREECUT）はツリーモデルの一部を切り取り（枝刈りと呼ばれます）、モデルをより単純にします。一般に、ツリーモデルでは、末端に近いノードほど該当件数が少なくなるため、モデル作成用データへの過剰適合が起こりやすくなります。そのため、モデルの枝刈りは、過剰適合を避ける効果があると考えられています。

DMT_TREECUTを用いた枝刈りは、特定の中間ノードの名前を指定する、もしくは枝刈りをおこなう階層数を指定する、もしくは検証モデルを指定することにより行います。指定条件に合致する中間ノードから分岐している下位ノードはすべて削除され、その中間ノードは終端ノードに変更されます。1回の実行で同時に最大100個までの中間ノードの枝刈りが可能です。

12.1.2 指定方法

（コマンド実行モードでの指定）

```
%dmt_treecut(help,model=,test=,cutnode=
,outmodel=,outtwig=,maxlvl=,
,pctf=7.2,meanf=best8.,aicf=best8.
,d  label=[D],c_label=[C],dif_label=[D]-[C]
language=JAPANESE))
```

（GUI実行モードでの変更点）

・help は指定不可。
・枝刈り後のツリーモデルの分岐表を表示するときに用いられるラベル・フォーマット参照データ

(labeldat=)を指定可能。

（必須パラメータ）

以下の2個のパラメータは常に必須です。

入力モデル (model=)
　… 入力モデルデータセット名の指定.
枝刈り後の出力モデル (outmodel=)
　… 枝刈り後の出力モデルデータセット名の指定.

（枝刈り指定パラメータ）

以下の3個のパラメータは枝刈り方法を指定します。3個の中の1つのパラメータは必須です。ただし、cutnode=とmaxlvl=については両方を指定することが可能です。両方指定した場合は両者いずれかの条件を満たすノードが枝刈りされます。test=パラメータは他の2つのいずれとも同時に指定できません。単独で指定します。

検証モデルによる枝刈り (test=)
　… 各ノードから分岐する2つの子ノード間のターゲット応答の大小関係を入力モデルと検証モデル間で比較し、矛盾する親ノードを枝刈りします。この指定は単独で指定します。
終端ノードに変更する中間ノードの指定 (cutnode=)
　… 枝刈りを行う中間ノード名の指定.（例 N001 N0111）
最大階層を指定した枝刈り (maxlvl=)
　… 指定の階層がツリーの最大階層となるようまとめて枝刈りを行う.

（オプションパラメータ）

以下の10個のパラメータは任意指定です。（=の右辺の値はデフォルト値を表しています）なお、pctf=, meanf=, aicf=, d_label=, c_label=, dif_label=パラメータは実行ログ画面に出力される枝刈り後のモデルの概要表示にのみ用いられます。

help … 指定方法のヘルプメッセージの表示.（コマンド実行モードでのみ有効）
刈り取った枝部分の出力モデル (outtwig=)
　… 刈りとった枝部分のモデル形式出力データセット名の指定(cutnode=指定に対応)
百分率の表示フォーマットの指定 (pctf=7.2)
平均値・標準偏差の表示フォーマットの指定 (meanf=best8.)
AIC値の表示フォーマットの指定 (aicf= best8.)
アップリフトモデルにおける処理群(DATA)を表す記号 (d_label=[D])
アップリフトモデルにおける対照群(Control)を表す記号 (c_label=[C])
アップリフトモデルにおける処理群−対照群間の差を表す記号 (dif_label=[D]-[C])
言語の選択 (language=JAPANESE)
ラベル・フォーマット参照データ (labeldat=)（GUI実行モードでのみ有効）

12.1.3 パラメータの詳細

入力モデル (model=)
入力モデルデータセット名を指定します。このパラメータは省略できません。
例：model=bunseki1

検証モデルによる枝刈り (test=)
検証データにモデルを適用したモデル形式データセット（検証モデル）名を入力します。検証モデルの中で子ノードのターゲット出現率もしくはターゲット平均値またはそれらの処理群実施群間の差が逆転している親ノードを探して、まとめて枝刈りします。（注意：V1.2ではGUI実行モードでのみ、test=を指定し、かつ、cutnodeボタンを押して出現する「逆転ノード」アイテムを選択することにより、この機能を実行していましたが、V1.3ではtest=を指定するだけで実行するように変更しました。）

枝刈り後の出力モデル (outmodel=)
枝刈り後のモデルデータセットの出力先の名前をつけます。このパラメータは省略できません。例：outmodel=new_model

最大階層を指定した枝刈り (maxlvl=)
1-19 の範囲の整数を指定します。指定の階層に該当する中間ノードをまとめて枝刈りします。このパラメータとcutnode=パラメータのいずれか1つ、もしくは両方指定できます。両方指定した場合は、いずれかの条件を満たすノードが枝刈りされます。なお、maxlvl=指定による枝データセット（outtwig=データセット）は生成されません。例：maxlvl=3

終端ノードに変更する中間ノードの指定 (cutnode=)
枝刈りを行う中間ノード名を指定します。複数の中間ノードの枝刈り指定を行う場合は、ノード名をブランクで区切って指定します。ただし、ノード名の最初の"_"(アンダースコア,アンダーバー)は省略して指定しなければなりません。最大100個のノード名が指定可能です。
例：cutnode=N01 N101 N100001

刈り取った枝部分の出力モデル (outtwig=)
cutnode=指定による枝刈り操作で、元のモデルから除去された枝部分の各部分モデルをモデル形式データセットとして出力します。元の中間モデルをルートノードとみなしたノード番号が新たに割り振られます。枝刈りした部分モデルは使う必要が無いかもしれませんが、DMT_TREEADDを用いて元に戻したい場合、接ぐ枝として使えます。デフォルトは _TWIG_xxxx1 _TWIG_xxxx2 ... _TWIG_xxxxk ただし、xxxx1 xxxx2 ... xxxxk はcutnode=パラメータで指定した枝刈り先中間ノード名を意味します。
名前を付ける場合は、outnode=パラメータに指定し

た枝刈り先中間ノードの指定順に対応して同じ数の名前を付ける必要があります。
例：outtwig =twig1 twig2

言語 (language=JAPANESE)
分析実行中のメッセージ出力、結果の表のタイトル、表項目などの表示言語を選択します。ただし、現バージョンでは、日本語か英語の2種類のみ選択可能です。
例：language=ENGLISH

12.1.4 GUI 実行モードで有効なパラメータの詳細

ラベル･フォーマット参照データ (labeldat=)
枝刈り後のツリー分岐表に変数ラベルと値ラベルの表示を行うために指定します。（**GUI 実行モードでのみ指定できます。**モデル作成時の入力データを記録していますので、存在する場合は、自動入力されます。）

12.1.5 コマンド実行モードで有効なパラメータの詳細

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。（**GUI 実行モードでは指定できません。**）　例：%dmt_treecut(help)

12.1.6 実行例

mincnt=10を指定してノード数の多いツリーモデル tree1 と tree1 を検証データにあてはめたモデル形式データセット（検証ツリー） TEST_tree1を作成します。

```
%dmt_tree(data=samp_data,y=flg,target=1,x=sei--DM,mincnt=10,maxlvl=10,outmodel=tree1)
```

(ログ)

```
生成されたモデルの概要 ...
 出力モデルデータセット: tree1
 ターゲット変数 変数ラベル: FLG 購入有無
 ターゲット値: "1"
 最大分割レベル: 10
 生成された終端ノードの数: 48
 分析データ全体の平均ターゲット出現率: 22.85%
 48 個の終端ノードのターゲット出現率範囲: 0.00% から 100.00%
 分割前の分析データの全体エントロピー: 0.7753872882
 48 個の終端ノード分割後の全体エントロピー: 0.3050956015
```

```
%dmt_treescore(model=tree1,data=test_data,y=flg,target=1,outmodel=TEST_tree1)
```

(ログ)

```
生成されたモデル形式データセットの概要 ...
 出力モデル形式データセット: TEST_tree1
 ターゲット変数 変数ラベル: FLG 購入有無
 ターゲット値: "1"
 最大分割レベル: 10
 生成された終端ノードの数: 48
 分析データ全体の平均ターゲット出現率: 22.80%
 48 個の終端ノードのターゲット出現率範囲: 0.00% から 100.00%
 分割前の分析データの全体エントロピー: 0.774509
 48 個の終端ノード分割後の全体エントロピー: 0.355666
```

例１：検証モデルを指定した枝刈り

```
%dmt_treecut(model=tree1,test=TEST_tree1
,outmodel=tree1_cut1)
```

(ログ)

```
ノート: tree1 と TEST_tree1 を比較した結果、5 個の逆転親ノードが存在します.

_N10001 _N110101 _N11100011 _N111001 _N111100

 ノート: 逆転親ノードを枝刈りし、終端ノードに変換したモデルデータセット tree1_cut1
を作成します.

(途中省略)

枝刈り後のモデルの概要 ...
 出力モデルデータセット: tree1_cut1
 最大分割レベル: 8
 生成された終端ノードの数: 38
 分析データ全体の平均ターゲット出現率: 22.85%
 38 個の終端ノードのターゲット出現率範囲: 0.00% から 100.00%
 分割前の分析データの全体エントロピー: 0.775387
 38 個の終端ノード分割後の全体エントロピー: 0.328765
```

この場合は、5個の中間ノードを終端ノードに変更したため、終端ノード数が48個から38個に減っています。

例２：最大階層数を指定して枝刈り

```
%dmt_treecut(model=tree1,maxlvl=3
,outmodel=tree1_cut2)
```

(ログ)

```
枝刈り後のモデルの概要 ...
 出力モデルデータセット: tree1_cut2
 最大分割レベル: 3
 生成された終端ノードの数: 8
 分析データ全体の平均ターゲット出現率: 22.85%
 8 個の終端ノードのターゲット出現率範囲: 0.00% から 78.98%
 分割前の分析データの全体エントロピー: 0.775387
 8 個の終端ノード分割後の全体エントロピー: 0.480837
```

例３：特定の中間ノードを指定して枝刈り

```
%dmt_treecut(model=tree1
,cutnode=N1 N001,outmodel=tree1_cut3)
```

(ログ)

```
枝刈り後モデルデータセット: tree1_cut3 が生成されました.
 枝データセット: _TWIG_N1 が生成されました.
 枝データセット: _TWIG_N001 が生成されました.

 ... DMT_TREECUT 実行が終わりました.


 枝刈り後のモデルの概要 ...
 出力モデルデータセット: tree1_cut3
 最大分割レベル: 6
 生成された終端ノードの数: 9
 分析データ全体の平均ターゲット出現率: 22.85%
 9 個の終端ノードのターゲット出現率範囲: 0.00% から 52.94%
 分割前の分析データの全体エントロピー: 0.775387
 9 個の終端ノード分割後の全体エントロピー: 0.571113
```

12.1.7 画面出力

コマンド実行モードでは画面出力はありません。
GUI実行モードでは、実行後、枝刈り後のモデルをツリー分岐表で表示する機能があります。

12.1.8 データセット出力

outmodel=パラメータに指定されたデータセットに枝刈り後のモデルデータセットが出力され、outtwig=パラメータに指定されたデータセットに枝刈りによって除去されたモデル部分が出力されます。

モデルデータセットと同じ変数項目が含まれますが、outtwig=データセットのノード番号はルートノードを表す_Nから番号が振り直されて出力されます。

コマンド実行モードで、test=パラメータを指定した場合は、モデルと検証モデルの逆転ノード名を_PROBLEM_PNODE という名前のデータセットをworkライブラリに生成します。この中には、枝刈りすしたノード名が含まれています。

12.1.9 逆転ノードに関するレポート

test=パラメータを指定すると、すべての親ノードについて、ノード分岐後の2つの子ノード間のモデル応答の大小関係がモデルと検証モデル間で逆転していないかどうかをまずチェックします。もしも逆転ノードが見つかれば、上位ノードにまとめた上で、以下のようにログにレポートします。

```
ノート: model.tree10 と test.TEST_tree10 を比較した結果、6 個の逆転親ノードが存在します.

_N10101 _N10110 _N110101 _N111000 _N1110010 _N111100

 ノート: 逆転親ノードを枝刈りし、終端ノードに変換したモデルデータセット outmodel.N0_edakari
を作成します.
```

一方、もしも逆転ノードが見つからなかった場合は、以下のメッセージをログに書き出して処理を終了します。

```
ノート: model.tree10 と test.tree10 を比較しましたが、逆転ノードは存在しません.
 ノート: DMT_TREECUTを終了します. 出力データセットは作成されませんでした.
```

12.1.10 制限

cutnode=指定で1度に指定できる枝刈り中間ノード数は最大100です。101個以上の枝刈り対象ノード指定を行う場合は、繰り返し実行してください。なお、maxlvl=指定およびtest=指定の枝刈りにはこの制限はありません。

12.1.11 コマンド実行モードでの注意

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

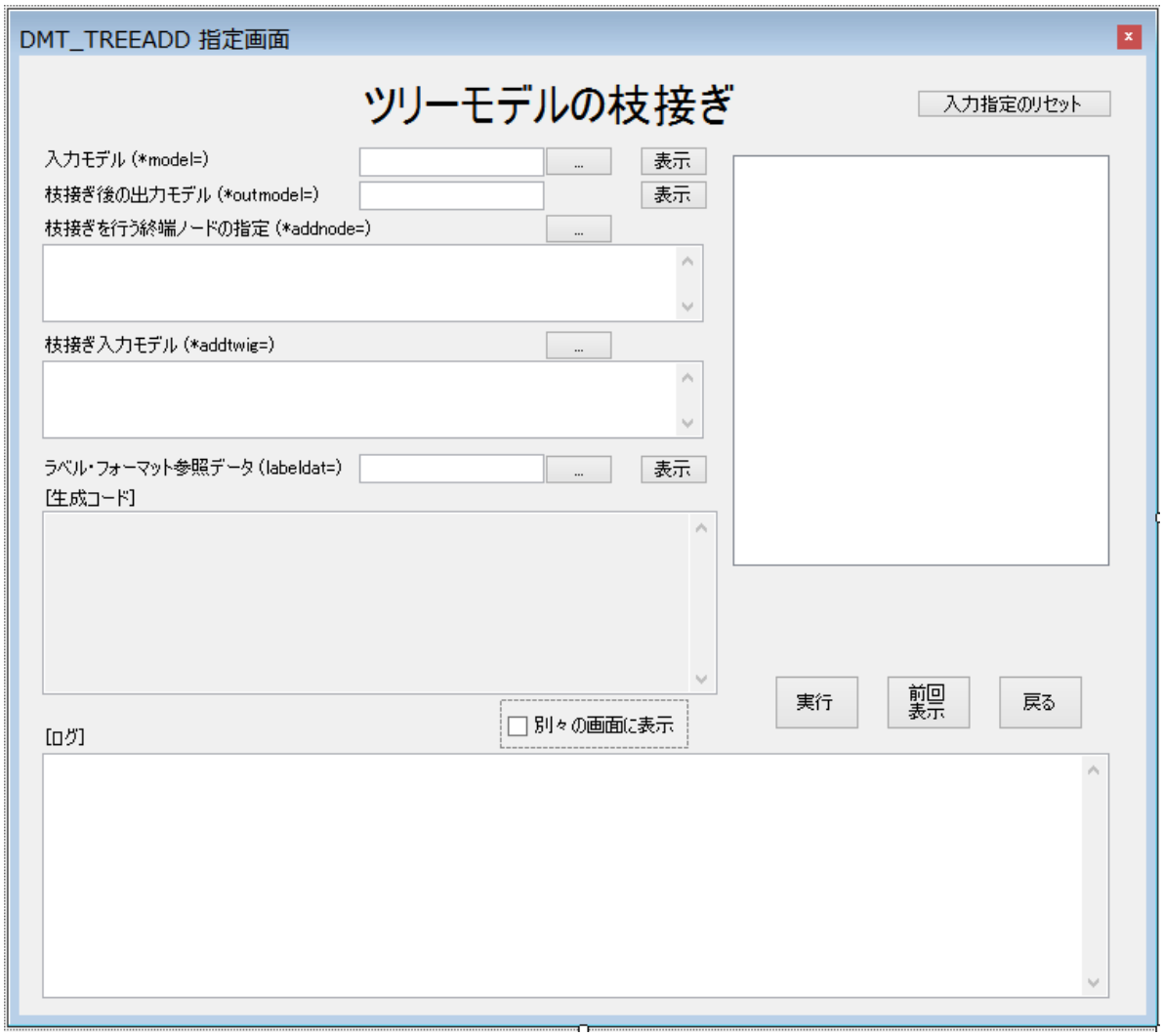
また、以下のユーザ定義フォーマットがWORKライブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
$NODE_C $NODE_D $_ORDER $_item
```

さらに、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
nobs zketa e_name e_type  _errormsg
```

12.2 枝接ぎ（dmt_treeadd）



12.2.1 概要

ツリーモデルの枝接ぎ（DMT_TREEADD）はツリーモデルの終端ノードに別のツリーモデルを接ぎ足し、より大きなモデルにします。
大局的に見てターゲット変数と関連が強い説明変数グループを用い、mincnt=パラメータをAUTOもしくは比較的大きな値に設定して作成したモデルに、局地的に説明力があると思われる特定の説明変数グループをで作成した小さなモデルを接ぎ足すことにより、精度と納得性を両立させたモデルに修正できるかもしれません。

DMT_TREEADDを用いた枝接ぎは、特定の終端ノードの名前を指定することにより行います。指定された終端ノードごとに枝接ぎを行う小さなモデルの名前を指定します。枝接ぎされた終端ノードは中間ノードに変更され、枝接ぎしたモデルのノード番号は自動的に全体のモデルの中で統一されたノード番号に変更されます。1回の実行で同時に最大100個までの終端ノードを指定した枝接ぎが可能です。

12.2.2 指定方法

（コマンド実行モードでの指定）

%dmt_treeadd(help,model=,addnode=,addtwig=
,outmodel=
,language=JAPANESE**)**

（GUI実行モードでの変更点）

・help は指定不可。
・枝刈り後のツリーモデルの分岐表を表示するとき

に用いられるラベル・フォーマット参照データ(labeldat=)を指定可能。

（必須パラメータ）

以下の4個のパラメータは省略できません。

入力モデル（model=）
… 入力モデルデータセット名の指定.
枝接ぎを行う終端ノードの指定（addnode=）
… 枝接ぎを行う終端ノード名の指定.（例 N001 N0111）
枝接ぎ入力モデル（addtwig=）
… 枝接ぎする入力モデルデータセット名の指定（addnode=指定に対応）
枝刈り後の出力モデル（outmodel=）
… 枝刈り後の出力モデルデータセット名の指定.

（オプションパラメータ）

以下の3個のパラメータは任意指定です。(=の右辺の値はデフォルト値を表しています)

help … 指定方法のヘルプメッセージの表示.(コマンド実行モードでのみ有効)
言語の選択（language=JAPANESE）
ラベル・フォーマット参照データ（labeldat=）
（GUI実行モードでのみ有効）

12.2.3 パラメータの詳細

入力モデル（model=）
入力モデルデータセット名を指定します。このパラメータは省略できません。
例：model=bunseki1

枝接ぎ後の出力モデル（outmodel=）
枝接ぎ後の出力モデルデータセットの名前をつけます。このパラメータは省略できません。例：outmodel=new_model

枝接ぎを行う終端ノードの指定（addnode=）
枝接ぎを行う終端ノード名を指定します。このパラメータは省略できません。複数の中間ノードの枝接ぎ指定を行う場合は、ノード名をブランクで区切って指定します。ただし、ノード名の最初の "_" (アンダースコア,アンダーバー) は省略して指定しなければなりません。最大100個のノード名が指定可能です。
例：addnode=N000 N001

枝接ぎ入力モデル（addtwig=）
枝接ぎ操作によりモデルの終端ノードに接ぎ足そうとする枝部分のモデルデータセットを入力します。このパラメータは省略できません。addnode=パラメータに指定した枝接ぎ先終端ノードと同数のモデル形式データセットをaddnode=パラメータ指定順に対応してaddtwig=パラメータに指定する必要があります。
例：addtwig =twig1 twig2

言語（language=JAPANESE）
分析実行中のメッセージ出力、結果の表のタイトル、表項目などの表示言語を選択します。ただし、現バージョンでは、日本語か英語の2種類のみ選択可能です。
例：language=ENGLISH

12.2.4 GUI 実行モードで有効なパラメータの詳細

ラベル・フォーマット参照データ(labeldat=)
枝接ぎ後のツリー分岐表に変数ラベルと値ラベルの表示を行うために指定します。（GUI 実行モードでのみ指定できます。）モデル作成時の入力データを記録していますので、存在する場合は、自動入力されます。

12.2.5 コマンド実行モードで有効なパラメータの詳細

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。（GUI実行モードでは指定できません。） 例：%dmt_treeadd(help)

12.2.6 実行例

例１：特定の終端ノードに別のツリーを枝接ぎする

(例示用のモデルtree1を作成)
```
%dmt_tree(data=samp_data,y=flg,target=1
,x=sei nenrei jukyo kazoku_kosei gakureki shokushu
kinmusaki gyoshu nenshu
,mincnt=100,maxlvl=2,outmodel=tree1)
```

(tree1のモデル分岐表)
```
%dmt_treetab(model=tree1,labeldata=samp_data)
```

DMT_TREE モデルテーブル (モデルデータセット: tree1)

			件数割合%	ターゲット再現率%	ターゲット出現率%
LVL0	LVL1	LVL2			
ROOT:22.85% (457/2,000)	N0: 4.56%(46/1,008) JUKYO 住居="2 持家(家族所有)","1 持家(自己所有)","6 寮","7 社宅"	N00: 2.74%(24/877) KINMUSAKI 勤務先形態="A 企業","D 官公庁","不明"	43.85	5.25	2.74
		N01: 16.79%(22/131) KINMUSAKI 勤務先形態="C 自営(個人)","B 自営(法人)"	6.55	4.81	16.79
	N1: 41.43%(411/992) JUKYO 住居="5 アパート","不明","4 借家","3 賃貸マンション"	N10: 16.24%(57/351) GAKUREKI 最終学歴="不明","3 専門学校","4 大学"	17.55	12.47	16.24
		N11: 55.23%(354/641) GAKUREKI 最終学歴="5 大学院","2 高校","1 中学"	32.05	77.46	55.23

終端ノードN00に性別と年齢を説明変数として分岐させるツリーを作成し、枝接ぎする。

(所属ノード番号をつける)
```
%dmt_treescore(data=samp_data,model=tree1
```

,outscore=score1)

(ノードN00の所属データについて変数DMで分岐するツリーモデルN00_tree1を作成)
%dmt_tree(data=score1(where=(_NODE="_N00")),y=flg,target=1
,x=DM,mincnt=50,maxlvl=1,outmodel=N00_tree1)

(N00_tree1のモデル分岐表)
%dmt_treetab(model=N00_tree1,labeldata=samp_data)

DMT_TREE モデルテーブル (モデルデータセット: N00_tree1)

		件数割合%	ターゲット再現率%	ターゲット出現率%
LVL0	LVL1			
ROOT:2.74% (24/877)	N0: 0.00%(0/609) DM プロモーション="0 非実施"	69.44	0.00	0.00
	N1: 8.96%(24/268) DM プロモーション="1 実施"	30.56	100.00	8.96

(tree1のノードN00にツリーモデルN00_tree1を枝接ぎ)
%dmt_treeadd(model=tree1,addnode=N00,addtwig=N00_tree1,outmodel=tree1_add)

(ログ)

```
枝接ぎ後モデルデータセット: tree1_add が生成されました.
 枝データセット: N00_TREE1 が 入力モデルデータセットノード _N00 に接ぎ木されました.

 ... DMT_TREEADD 実行が終わりました.

 ... 正しいモデル予測値やエントロピ減少値を得るために
は、%DMT_TREESCORE
を用い、このモデルを分析データセットに適用してください.
 (例) %DMT_TREESCORE(model=今回作成したモデルデータセット,data=分析データ,y=ターゲット変数名,
target=ターゲット値,outmodel=正しい予測値やエントロピー減少値を持つ出力モデルデータセット)
```

(tree1_addモデルのツリー分岐表)
%dmt_treetab(model=tree1_add,labeldata=samp_data)

DMT_TREE モデルテーブル (モデルデータセット: tree1_add)

				件数割合%	ターゲット再現率%	ターゲット出現率%
LVL0	LVL1	LVL2	LVL3			
ROOT:22.85% (457/2,000)	N0: 4.56%(46/1,008) JUKYO 住居="2 持家(寮族所有)","1 持家(自己所有)","6 寮","7 社宅"	N00: 2.74%(24/877) KINMUSAKI 勤務先形態="A 企業","D 官公庁","不明"	N000: 0.00% (0/609) DM プロモーション="0 非実施"	30.45	0.00	0.00
			N001: 8.96% (24/268) DM プロモーション="1 実施"	13.40	5.25	8.96
		N01: 16.79%(22/131) KINMUSAKI 勤務先形態="C 自営(個人)","B 自営(法人)"		6.55	4.81	16.79
	N1: 41.43%(411/992) JUKYO 住居="5 アパート","不明","4 借家","3 賃貸マンション"	N10: 16.24%(57/351) GAKUREKI 最終学歴="不明","3 専門学校","4 大学"		17.55	12.47	16.24
		N11: 55.23%(354/641) GAKUREKI 最終学歴="5 大学院","2 高校","1 中学"		32.05	77.46	55.23

注意：この場合は、分析データに基づいて枝接ぎ用モデルを作成していますので、生成されたモデルに分析データを再度適用して予測値を調整する必要はありません。

例２：すべての終端ノードに対して、有効なツリーモデルを枝接ぎする（**注意**）この例示はGUI実行モードでは直接サポートされていません。コマンド実行モードで実行してください。

(例１のscore1作成までは同じ処理を行います)

(score1から終端ノードの値をすべて抽出し、終端ノードの種類数をマクロ変数 &n に、終端ノードの名前を &term1, &term2, ... , &&term&n に格納する)

```
proc freq data=score1;
   tables _NODE/noprint
out=termnodes(keep=_NODE);
run;
data _null_;
   if _n_=1 then call symput("n",compress(n));
   set termnodes nobs=n;
   call symput("term"||left(_n_),compress(_NODE));
run;
```

(マクロ変数値の確認)
%put &n, &term1, &term2, ... , &&term&n;

(ログ)

```
4, _N00, _N01, ... , _N11
```

(終端ノード別にDMの値によって分岐する1階層のツリーモデルを作成するマクロプログラム)

```
%macro create_twigs;
   %do i=1 %to &n;
      %dmt_tree(data=score1(where=(_NODE="&&term&i")),y=flg,target=1
,x=DM,mincnt=50,maxlvl=1,outmodel=&&term&i._tree1)
   %end;
%mend create_twigs;
%create_twigs
```

(N00のログ)

```
入力分析データセット: score1(where=(_NODE="_N00"))
 オブザベーション数: 877
 最小必要ノード件数: 50
 最大分割レベル: 1
 ターゲット変数 変数ラベル: FLG 購入有無
 ターゲット値: "1"
 ターゲット値の出現率を基準としたツリー分析を行います ...

 数値変数の個々の値をカテゴリ値に設定する最大種類数: CEIL(1+log2(N)), Nは非欠損件数
 数値変数のカテゴリ生成方法(最後のカテゴラリ件数が少ない場合１つ前のカテゴリに併合するか否か): N
 説明変数 (尺度) 変数ラベル:
(1) DM (名義) プロモーション

 親ノード _N(N=877,P=2.74%) を分割中です.
  子ノード _N0(N=609,P=0.00%) と _N1(N=268,P=8.96%) が DM によって生成されました.

 ... DMT_TREE 実行が終わりました.

 生成されたモデルの概要 ...
 出力モデルデータセット: _N00_tree1
 ターゲット変数 変数ラベル: FLG 購入有無
 ターゲット値: "1"
 最大分割レベル: 1
 生成された終端ノードの数: 2
 分析データ全体の平均ターゲット出現率: 2.74%
 2 個の終端ノードのターゲット出現率範囲: 0.00% から 8.96%
 分割前の分析データの全体エントロピー: 0.1810054983
```

```
 2 個の終端ノード分割後の全体エントロピー: 0.1329223405
```

(N01のログ) （有効なツリーモデルは生成されず）

```
入力分析データセット: score1(where=(_NODE="_N01"))
 オブザベーション数: 131
 最小必要ノード件数: 50
 最大分割レベル: 1
 ターゲット変数 変数ラベル: FLG 購入有無
 ターゲット値: "1"
 ターゲット値の出現率を基準としたツリー分析を行います ...

 数値変数の個々の値をカテゴリ値に設定する最大種類数: CEIL(1+log2(N)), N は非欠損件数
 数値変数のカテゴリ生成方法(最後のカテゴラリ件数が少ない場合 1 つ前のカテゴリに併合するか否か): N
 説明変数 (尺度) 変数ラベル:
(1) DM (名義) プロモーション

 親ノード _N(N=131,P=16.79%) を分割中です.
  有効な説明変数の併合パターンが存在しません.
  親ノード _N は分割されませんでした.
 ... ツリーモデルは生成されませんでした.
```

(N10のログ) （有効なツリーモデルは生成されず）

```
入力分析データセット: score1(where=(_NODE="_N10"))
 オブザベーション数: 351
 最小必要ノード件数: 50
 最大分割レベル: 1
 ターゲット変数 変数ラベル: FLG 購入有無
 ターゲット値: "1"
 ターゲット値の出現率を基準としたツリー分析を行います ...

 数値変数の個々の値をカテゴリ値に設定する最大種類数: CEIL(1+log2(N)), N は非欠損件数
 数値変数のカテゴリ生成方法(最後のカテゴラリ件数が少ない場合 1 つ前のカテゴリに併合するか否か): N
 説明変数 (尺度) 変数ラベル:
(1) DM (名義) プロモーション

 親ノード _N(N=351,P=16.24%) を分割中です.
  有効な説明変数が存在しません.
  親ノード _N は分割されませんでした.
 ... ツリーモデルは生成されませんでした.
```

(N11のログ)

```
入力分析データセット: score1(where=(_NODE="_N11"))
 オブザベーション数: 641
 最小必要ノード件数: 50
 最大分割レベル: 1
 ターゲット変数 変数ラベル: FLG 購入有無
 ターゲット値: "1"
 ターゲット値の出現率を基準としたツリー分析を行います ...

 数値変数の個々の値をカテゴリ値に設定する最大種類数: CEIL(1+log2(N)), N は非欠損件数
 数値変数のカテゴリ生成方法(最後のカテゴラリ件数が少ない場合 1 つ前のカテゴリに併合するか否か): N
 説明変数 (尺度) 変数ラベル:
(1) DM (名義) プロモーション

 親ノード _N(N=641,P=55.23%) を分割中です.
  子ノード _N0(N=415,P=50.84%) と _N1(N=226,P=63.27%) が DM によって生成されました.

 ... DMT_TREE 実行が終わりました.

 生成されたモデルの概要 ...
 出力モデルデータセット: _N11_tree1
 ターゲット変数 変数ラベル: FLG 購入有無
 ターゲット値: "1"
 最大分割レベル: 1
 生成された終端ノードの数: 2
 分析データ全体の平均ターゲット出現率: 55.23%
 2 個の終端ノードのターゲット出現率範囲: 50.84% から 63.27%
 分割前の分析データの全体エントロピー: 0.9921046454
 2 個の終端ノード分割後の全体エントロピー: 0.9817244873
```

(有効なツリーモデルを各終端ノードに枝接ぎ)

```
%macro do_add;
  %dmt_treeadd(model=tree1,addnode=
  %do i=1 %to &n;
    %if %sysfunc(exist(&&term&i._tree1)) %then %do;
      %str( &&term&i )
    %end;
  %end;
  ,addtwig=
  %do i=1 %to &n;
    %if %sysfunc(exist(&&term&i._tree1)) %then %do;
      %str( &&term&i._tree1 )
    %end;
  %end;
  ,outmodel=tree1_add_allterms)
%mend do_add;
%do_add
```

(ログ)

```
枝接ぎ後モデルデータセット: tree1_add_allterms が生成されました.
 枝データセット: _N00_TREE1 が 入力モデルデータセットノード _N00 に接ぎ木されました.
 枝データセット: _N11_TREE1 が 入力モデルデータセットノード _N11 に接ぎ木されました.
```

(tree1_add_allttermsモデルのツリー分岐表)

```
%dmt_treetab(model=tree1_add_allterms,labeldata=samp_data)
```

DMT_TREE モデルテーブル (モデルデータセット: tree1_add_allterms)

				件数割合%	ターゲット再現率%	ターゲット出現率%
LVL0	LVL1	LVL2	LVL3			
ROOT:22.85% (457/2,000)	N0: 4.56%(46/1,008) JUKYO 住居="2 持家(家族所有)","1 持家(自己所有)","6 寮","7 社宅"	N00: 2.74%(24/877) KINMUSAKI 勤務先形態="A 企業","D 官公庁","" 不明"	N000: 0.00%(0/609) DM プロモーション="0 非実施"	30.45	0.00	0.00
			N001: 8.96%(24/268) DM プロモーション="1 実施"	13.40	5.25	8.96
		N01: 16.79%(22/131) KINMUSAKI 勤務先形態="C 自営(個人)","B 自営(法人)"		6.55	4.81	16.79
	N1: 41.43%(411/992) JUKYO 住居="5 アパート","" 不明","4 借家","3 賃貸マンション"	N10: 16.24%(57/351) GAKUREKI 最終学歴="" 不明","3 専門学校","4 大学"		17.55	12.47	16.24
		N11: 55.23%(354/641) GAKUREKI 最終学歴="5 大学院","2 高校","1 中学"	N110: 50.84% (211/415) DM プロモーション="0 非実施"	20.75	46.17	50.84
			N111: 63.27% (143/226) DM プロモーション="1 実施"	11.30	31.29	63.27

12.2.7 データセット出力

outmodel=パラメータに指定されたデータセットに枝接ぎ後のモデルデータセットが出力されます。

12.2.8 制限

同時に枝接ぎできる終端ノード数は最大100です。100を超える場合は繰り返して実行します。

12.2.9 枝接ぎ後の注意

枝接ぎ部分のモデルが、枝接ぎ前のモデルと共通の分析データセットの枝接ぎ先の終端ノードに対して作成したものである場合を除いて、ノード件数、ターゲット件数、ターゲット平均値、ターゲット標準偏差、エントロピー減少量、群内平方和減少量などの統計量は、枝接ぎ後の全体モデルにおいて正しい値を保持しているとは限りません。その場合は、DMT_TREESCOREを用い、元の入力データセットに

対し枝接ぎ後のモデルを適用し、正しい統計量を再計算する必要があります。DMT_TREEADDを実行すると、その意味のメッセージを常にログに書き出します。必要がある場合は再計算を行ってください。

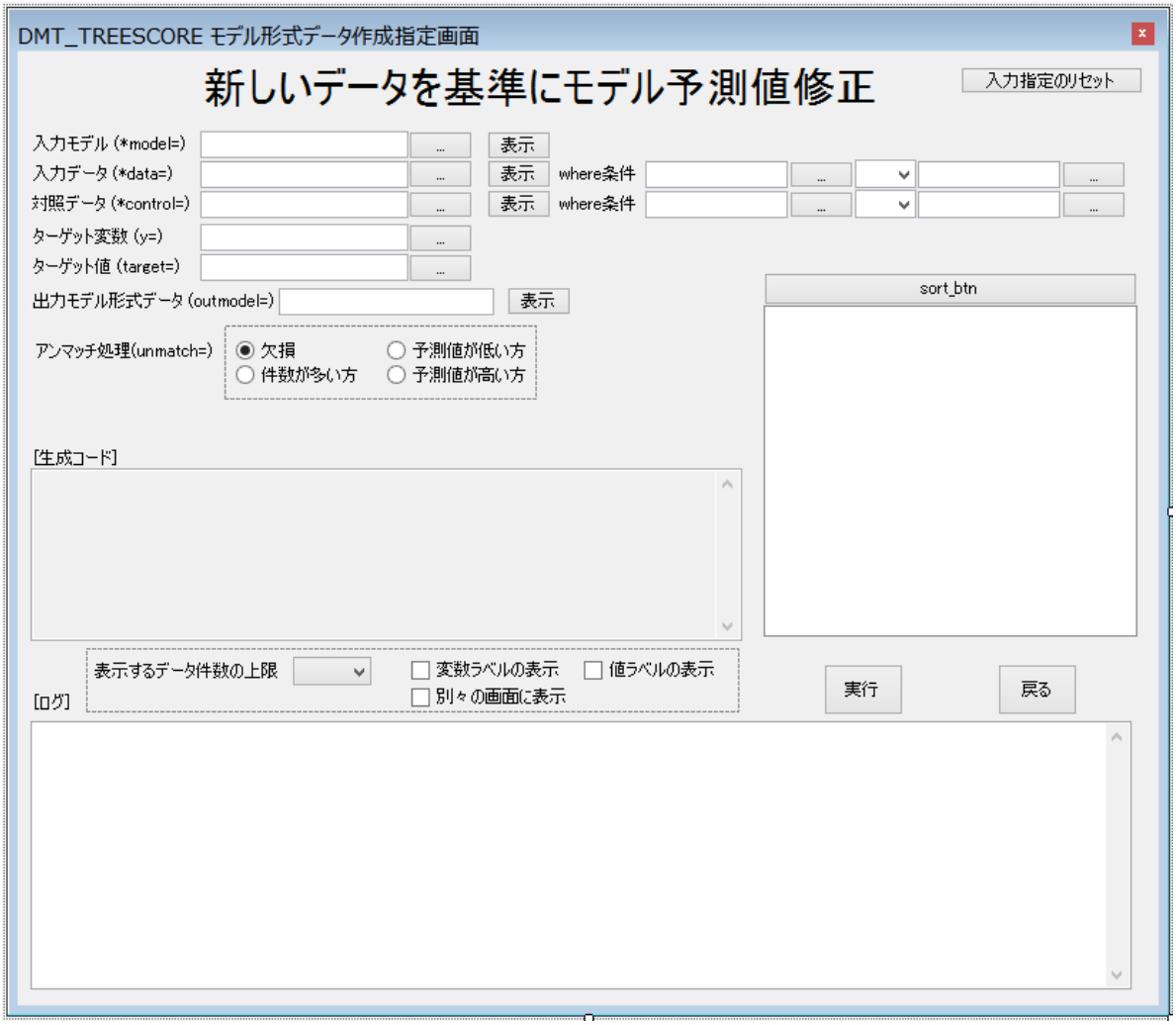
12.2.10 コマンド実行モードでの注意

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

また、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。なお、&iは数字を表し、たいていの場合、説明変数に指定した変数の数だけ存在する可能性があることを表します。

```
nobs zketa e_name e_type _errormsg
```

12.3 予測値修正（dmt_treescore outmodel=）



12.3.1 概要

新しいデータを基準にモデル予測値修正（DMT_TREESCORE）はモデルに採用されている説明変数すべてとターゲット変数を含む入力データセットに対してツリーモデルを適用し、入力データセットにおけるノード別件数、および分類木の場合はターゲット件数、回帰木の場合はターゲット変数の平均値と標準偏差、さらにアップリフトでは関連する他の統計量を集計し、モデルデータセットと同じ形式のデータセット（モデル形式データセット）を出力します。

検証用データセットにDMT_TREESCOREを適用したモデル形式データセットを作成しておくと、以下の各マクロのmodel=パラメータとtest=パラメータに同時指定することにより、モデル作成データとモデル検証データにおける統計量を同時表示することができ、予測と実績との比較検証などが効率良く行えます。

ツリー分岐表（DMT_TREETAB）
ツリーノード定義表（DMT_NODETAB）
ゲインチャート・収益チャート（DMT_GAINCHART）、アップリフトチャート（DMT_UPLIFTCHART）
比較プロット（DMT_COMPAREPLOT）
正誤表（DMT_CORRECTTAB）
モデルの枝刈り（DMT_TREECUT）

ツリーモデルをデータに適用する場合、予測値が付けられないケースが発生することがあります。DMT_TREESCOREはこの問題に unmatch=パラメータで対処しています。デフォルトはアンマッチのまま（予測値は欠損）に設定されますが、強制的にノード分岐を行う3通りのオプション（いずれも予測値を算出します）を選択可能です。

（予測値が付けられないケースが発生する理由）

予測値を付ける方法は非常に単純です。各オブザベーションごとに、モデルの分岐説明変数値を参照しながら、所属中間ノードを逐次的に辿っていき、最終的に達した終端ノードのターゲット出現率を予測値とします。
しかし、ツリーモデルの各ノード分岐規則は、**モデル作成データに実際に存在した**説明変数値に基づいて定義されます。階層が深くなるほどノード件数は少なくなりますので、分岐変数に選ばれた説明変数の値の分布は必ずしも本来存在する可能性のある範囲をすべてカバーしているとは限りません。このため、予測値をつけたいデータの文字タイプ説明変数が分岐変数に採用されているノードにおいて、分岐の途中で分岐先不明（アンマッチ）となるカテゴリを持つオブザベーションが入力される可能性が生じます。数値タイプ変数の場合は、常にあるしきい値に基づく範囲でノード分岐先の定義がなされていますので、分岐先不明となる心配はほとんどありませんが、唯一、モデル作成時には存在しなかった欠損値が入力オブザベーションに存在したときにアンマッチとなり得ます。このような理由で、ツリーモデルをデータに適用する場合、予測値が付けられないケースが発生することがあります。

DMT_TREESCOREはこの問題に unmatch=パラメータで対処しています。デフォルトはアンマッチのまま（予測値は欠損）に設定されますが、強制的にノード分岐を行う3通りのオプション（いずれも予測値を算出します）を選択可能です。

12.3.2 指定方法

（コマンド実行モードでの指定）

%dmt_treescore(help,data=,control=,model=,
outmodel=,y=,target=,unmatch=MISSING,
language=JAPANESE**)**

（GUI実行モードでの変更点）

・help は指定不可。

（必須パラメータ）

以下の2個のパラメータは常に省略できません。

入力データ (data=) … **入力データセット名の指定**
入力モデル (model=) … **入力モデルデータセット名の指定.**

以下の1個のパラメータはモデルがアップリフトモデルの場合は省略できません。

入力対照データ (control=) … **対照群の入力データセット名の指定**

（モデル形式データセットを出力するためのパラメータ）

以下の 3個のパラメータ outmodel=, y=, target= はdata=入力データセットに（アップリフトモデルの場合はcontrol=入力対照データセットにも）model=モデルを適用した場合、ノード別該当件数やターゲット集計値を計算し outmodel= データセットに出力するために用います。モデル形式データセットを出力する場合、分類木および分類木アップリフトモデルの場合、これら3つはすべて必須です。回帰木および回帰木アップリフトモデルの場合はtarget=パラメータを除く2つが必須指定です。

出力モデル形式データ (outmodel=)
… **モデルを入力データセットに適用した場合のノード別実績件数とターゲット件数を出力するモデル形式データセット名の指定.**
ターゲット変数 (y=) … **ターゲット変数名の指定.**
ターゲット値 (target=)
… **ターゲット値の指定.** **（回帰木モデル適用の場合は指定してはいけません）**

（アンマッチ処理のためのパラメータ）

アンマッチ処理 (unmatch=MISSING)
… **アンマッチデータ（モデルのノード分割変数カテゴリに該当しないカテゴリを持つオブザベーション）への対処方法の選択.**

（その他のパラメータ）

以下の2個のパラメータは任意指定です。(=の右辺の値はデフォルト値を表しています)

help … **指定方法のヘルプメッセージの表示.（コマンド実行モードでのみ有効）**
言語（language=JAPANESE) … **ログやメッセージを表示する言語の選択**

12.3.3 パラメータの詳細

入力モデル (model=)
入力モデルデータセット名を指定します。このパラメータは省略できません。
例：model=bunseki1

入力データ (data=)
入力データセット名を指定します。このパラメータは省略できません。例：data=a

ターゲット変数 (y=)
モデルをデータに適用するとき、データに含まれるターゲット変数を指定します。分類木の場合は **ターゲット値**(target=) を同時に指定しなければなりません。例：y=flag

ターゲット値 (target=)

分類木モデルをデータに適用しターゲット出現率に関するノード別集計値を計算するために、データに含まれる y=ターゲット変数のターゲット値を指定します。回帰木モデルの検証を行う場合は指定してはいけません。
例：target="1"
なお、引用符で囲まなくても構いません。（自動判断します）

出力モデル形式データ (outmodel=)
モデルを入力データセットに適用したときのノード別実績件数やその他の集計結果を出力するモデル形式データセット名を指定します。例：
outmodel=new_model

アンマッチ処理 (unmatch=MISSING)
入力データセットの各オブザベーションをツリーモデルのノード分岐規則に従って、分岐先ノードを逐次決定していく過程において、そのオブザベーションの持つ分岐説明変数値がモデルの2つの分岐先ノードのいずれにも該当しないとき（これをアンマッチと呼びます）の対処方法を指定します。一般に、アンバランスなカテゴリを持つ説明変数が分岐に用いられたり、ツリー階層が深くなるほど、アンマッチの発生確率が高くなります。

デフォルト値 **欠損**(MISSING) の場合は outmodel=データセットにはマッチしたデータのみを用いたノード別件数とターゲット件数の集計結果が保存されます。

その他 **件数が多い方**(FREQ)/**予測値が高い方**(HIGH)/**予測値が低い方**(LOW)のいずれかを指定可能です。 これらの場合は、アンマッチが発生した場合、次のように分岐先ノードを決定して終端ノードまで辿る処理を継続し、予測値を付与します。

FREQはモデル上で該当件数の多い方の分岐先ノード，HIGHはモデル上でターゲット出現率や平均値、処理群と対照群間のターゲット値の差分が高い（大きい）方の分岐先ノード，LOWは逆にモデル上で低い（小さい）方の分岐先ノードに強制的に振り分けを行います。

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。（**GUI 実行モードでは指定できません。**） 例：%dmt_treescore(help)

言語 (language=JAPANESE)
分析実行中のメッセージ出力、結果の表のタイトル、表項目などの表示言語を選択します。ただし、現バージョンでは、日本語か英語の2種類のみ選択可能です。
例：language=ENGLISH

12.3.4 実行例

例１：分類木モデルをテストデータに適用し、モデル形式データセット（検証ツリー）を作成。

```
%dmt_tree(data=samp_data,y=flg,target=1,x=sei--DM,mincnt=50,maxlvl=10,outmodel=tree1)

%dmt_treescore(model=tree1,data=test_data,y=flg,target=1,outmodel=TEST_tree1)
```

例２：分類木アップリフトモデルをテストデータに適用し、モデル形式データセット（検証ツリー）を作成。

```
%dmt_tree(data=samp_data(where=(DM="1")),control=SAMP_DATA(where=(DM="0")),y=flg,target=1,x=sei--nenshu,mincnt=50,maxlvl=10,outmodel=tree1)

%dmt_treescore(model=tree1
,data=TEST_DATA(where=(DM="1"))
,control=TEST_DATA(where=(DM="0"))
,y=flg,target=1,outmodel=TEST_tree1)
```

12.3.5 データセット出力

outmodel=パラメータに指定されたデータセットにテストデータに適用したモデルの各ノードの件数その他の統計量を集計したモデル形式データセットが出力されます。

12.3.6 欠損値の取り扱い

data=,model=, y=,（さらに、必要に応じて、control=, target=）を指定してモデル形式出力データセットを作成する場合、data=入力データセット（および、control=データセット）の y=ターゲット変数に含まれる欠損値は以下のように取り扱われます。

分類木モデルの場合の文字タイプのターゲット変数、数値タイプのターゲット変数はいずれも有効な値の1つとみなされます。

回帰木モデルの場合は、ターゲット変数に欠損値を持つオブザベーションは除外してから処理が行われます。

12.3.7 コマンド実行モードでの注意

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

また、以下のユーザ定義フォーマットがWORKライブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。

```
$_item
```

さらに、以下のグローバルマクロ変数が作成されま

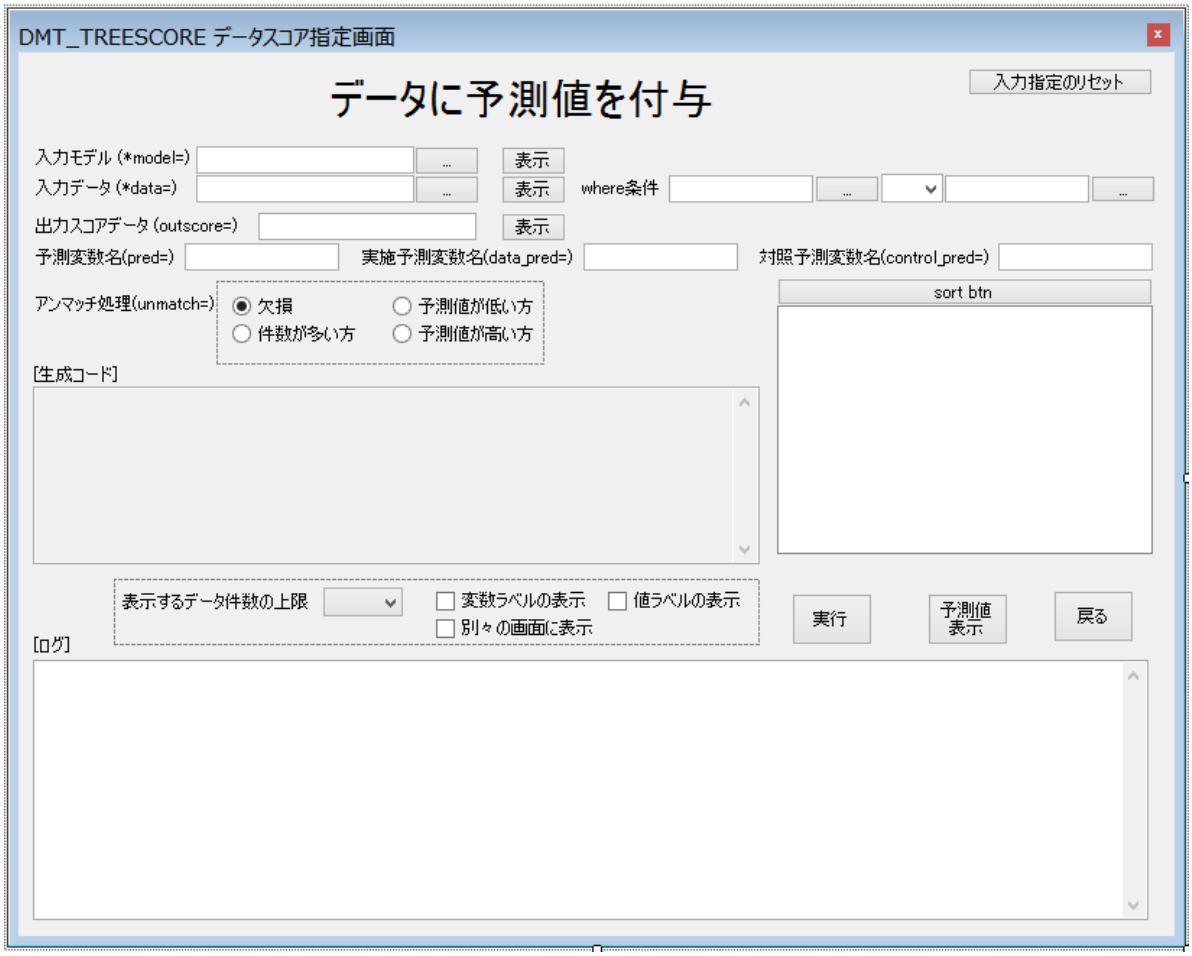
す。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。

```
e_name e_type lab&i nobs spc&i typ&i zketa
_speclen _specnum _errormsg
```

13. 分析画面 ⑥モデル適用

モデルの予測値をデータセットに含まれる全オブザベーションに付与します。

13.1 予測付与（dmt_treescore outscore=）



13.1.1 概要

データに予測値を付与（DMT_TREESCORE）はモデルに採用されている説明変数をすべて含む入力データセットに対してツリーモデルを適用し、モデルの予測値を付与したデータセットを出力します

ツリーモデルをデータに適用する場合、予測値が付けられないケースが発生することがあります。（理由は前項の**新しいデータを基準にモデル予測値修正**を参照）DMT_TREESCOREはこの問題に unmatch=パラメータで対処しています。デフォルトはアンマッチのまま（予測値は欠損）に設定されますが、強制的にノード分岐を行う3通りのオプション（いずれも予測値を算出します）を選択可能です。

13.1.2 指定方法

（コマンド実行モードでの指定）

%dmt_treescore(help,data=,model=,
outscore=_treescore,
pred=,data_pred=,control_pred=,
unmatch=MISSING,
language=JAPANESE**)**

（GUI実行モードでの変更点）

・help は指定不可。

（必須パラメータ）

以下の2個のパラメータは常に必須です。

入力データ (data=) … 入力データセット名の指定
入力モデル (model=) … 入力モデルデータセット名の指定.

（予測値を入力データセットの各オブザベーションにつけるためのパラメータ）

以下の2個パラメータはdata=入力データセットの各オブザベーションに予測値を付与する場合に指定します。(=の右辺の値はデフォルト値を表しています)

出力スコアデータ (outscore=_treescore)
… 予測値を含む出力スコアデータセット名の指定.
予測変数名 (pred=_CONF(または_MEAN,または DIF_CONF,またはDIF_MEAN))
… 予測値を表す変数名の指定.

アップリフトモデルの場合は、さらに、以下の2個の予測変数を指定可能です。

処理群の予測変数名 (data_pred=D_CONF,または D_MEAN))
… 処理した場合の予測値を表す変数名の指定.
対照群の予測変数名 (control_pred=C_CONF,または C_MEAN))
… 対照群に残した場合の予測値を表す変数名の指定.

（アンマッチ処理のためのパラメータ）

アンマッチ処理 (unmatch=MISSING)
… アンマッチデータ（モデルのノード分割変数カテゴリに該当しないカテゴリを持つオブザベーション）への対処方法の選択.

（その他のパラメータ）

以下の2個のパラメータは任意指定です。(=の右辺の値はデフォルト値を表しています)

help … 指定方法のヘルプメッセージの表示.（コマンド実行モードでのみ有効）
言語の選択 (language=JAPANESE)

13.1.3 パラメータの詳細

入力モデル (model=)
入力モデルデータセット名を指定します。このパラメータは省略できません。
例：model=bunseki1

入力データ (data=)
入力データセット名を指定します。このパラメータは省略できません。例：data=a

出力スコアデータ (outscore=_treescore)
入力データセットの各オブザベーションに対するモデル予測ターゲット出現率、もしくはモデル予測ターゲット値をデータセットに出力します。

予測変数名 (pred=_CONF（または_MEAN,または DIF_CONF,またはDIF_MEAN）)
outscore=出力データセットに加えるモデル予測変数名を表す変数名を指定します。デフォルトは分類木モデルの場合は_CONF、回帰木モデルの場合は_MEAN、分類木アップリフトモデルではDIF_CONF、回帰木アップリフトモデルではDIF_MEANです。

処理群の予測変数名 (data_pred=D_CONF,または D_MEAN))
分類木アップリフトモデル、または回帰木アップリフトモデルの場合に、そのオブザベーションを処理群に設定した場合の予測値を表す変数名を指定します。

対照群の予測変数名 (control_pred=C_CONF,または C_MEAN))
分類木アップリフトモデル、または回帰木アップリフトモデルの場合に、そのオブザベーションを対照群に設定した場合の予測値を表す変数名を指定します。

アンマッチ処理(unmatch=MISSING)
入力データセットの各オブザベーションにおいて、分岐説明変数値がモデルの2つの分岐先ノードのいずれにも該当しないとき（これをアンマッチと呼びます）の対処方法を指定します。

デフォルト値 欠損(MISSING) は outscore= データセットの予測値（pred=パラメータに指定した変数の値）に欠損値を与え、自動変数_NODEにはマッチした最後のノード名を与えます。

その他 件数が多い方(FREQ)/予測値が高い方(HIGH)/予測値が低い方(LOW) のいずれかを指定可能です。 これらの場合は、アンマッチが発生した場合、次のように分岐先ノードを決定して終端ノードまで辿る処理を継続し、予測値を付与します。

FREQはモデル上で該当件数の多い方の分岐先ノード，HIGHはモデル上でターゲット出現率が高い方の分岐先ノード，LOWは低い方の分岐先ノードに強制的に振り分けを行います。

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。（GUI 実行モードでは指定できません。） 例：%dmt_treescore(help)

言語 (language=JAPANESE)
分析実行中のメッセージ出力、結果の表のタイトル、表項目などの表示言語を選択します。ただし、現バージョンでは、日本語か英語の2種類のみ選択可能で

す。
例：language=ENGLISH

13.1.4 実行例

例１：分類木モデルの予測値をデータにつける

%dmt_tree(data=samp_data,y=flg,target=1,x=sei--DM,mincnt=50,maxlvl=10,outmodel=tree1)

%dmt_treescore(model=tree1,data=test_data, outscore=test_score1)

	nenshu	DM	flg	kingaku	_NODE	_TERM	_UNMATCH	_CONF
1	376	実施	なし	0	_N010	YES	NO	0
2	.	実施	なし	0	_N11110	YES	NO	0.4657534247
3	913	実施	なし	0	_N01111	YES	NO	0.243902439
4	.	実施	あり	100	_N11111	YES	NO	0.8493150685
5	305	実施	なし	0	_N1000	YES	NO	0.0416666667
6	.	実施	なし	0	_N1010	YES	NO	0
7	.	実施	なし	0	_N01111	YES	NO	0.243902439
8	327	実施	なし	0	_N0110	YES	NO	0.0615384615
9	346	実施	なし	0	_N1011	YES	NO	0.0625
10	713	実施	なし	0	_N11110	YES	NO	0.4657534247
11	.	実施	なし	0	_N01111	YES	NO	0.243902439
12	831	実施	なし	0	_N01111	YES	NO	0.243902439

変数 _NODE, _TERM, _UNMATCH および予測変数名（ここではpred=パラメータ無指定なので _CONF）が追加されます。

例２：分類木アップリフトモデルの処理群の場合の予測値と対照群の場合の予測値をデータにつける

%dmt_tree(data=samp_data(where=(DM="1")),control=SAMP_DATA(where=(DM="0")),y=flg,target=1,x=sei--nenshu,mincnt=50,maxlvl=10,outmodel=uplift_tree1)

%dmt_treescore(model=uplift_tree1 ,data=TEST_DATA,outscore=test_score2)

	flg	kingaku	_NODE	_TERM	_UNMATCH	DIF_CONF	D_CONF	C_CONF
1	なし	0	_N110	YES	NO	0.0535947712	0.0980392157	0.0444444444
2	なし	0	_N0001	YES	NO	-0.008588186	0.1688311688	0.1774193548
3	なし	0	_N110	YES	NO	0.0535947712	0.0980392157	0.0444444444
4	あり	100	_N111	YES	NO	0.6718027735	0.9090909091	0.2372881356
5	なし	0	_N0001	YES	NO	-0.008588186	0.1688311688	0.1774193548
6	なし	0	_N110	YES	NO	0.0535947712	0.0980392157	0.0444444444
7	なし	0	_N010	YES	NO	0.09462486	0.1052631579	0.0106382979
8	なし	0	_N100	YES	NO	-0.180371073	0.0483870968	0.2287581699
9	なし	0	_N101	YES	NO	0.1162687887	0.3461538462	0.2298850575
10	なし	0	_N0001	YES	NO	-0.008588186	0.1688311688	0.1774193548
11	なし	0	_N010	YES	NO	0.09462486	0.1052631579	0.0106382979
12	なし	0	_N111	YES	NO	0.6718027735	0.9090909091	0.2372881356

変数 _NODE, _TERM, _UNMATCH および予測変数名（ここではpred=,data_pred=,control_pred=パラメータがすべて無指定なので それぞれ、DIF_CONF, D_CONF, C_CONF）が追加されます。

13.1.5 データセット出力

出力スコアデータ(outscore=)データセット
data=入力データセットの各オブザベーションに対して、model=入力モデルを適用し、以下の4変数（アップリフトモデルでは6変数）を追加したデータセットを出力します。

_NODE, _TERM, _UNMATCH, &pred

ただし、&predは　予測変数名(pred=)に指定した名前です。

outscore=出力データセットに追加される変数

変数名	タイプ	長さ	内容	備考
_NODE	文字	可変	所属ノード名	unmatch=MISSING 指定かつ _UNMATCH="YES"の場合は最後にマッチした中間ノード名が入る。それ以外の場合は辿りついた終端ノード名が入る
_TERM	文字	3	終端ノード識別変数	unmatch=MISSING 指定かつ _UNMATCH="YES"の場合"NO"となる。それ以外は"YES"が入る
_UNMATCH	文字	3	どこかでアンマッチが発生したかどうかの判定	unmatch=パラメータ指定の如何に関わらず、アンマッチが発生した場合は常に"YES"が入る。それ以外は"NO"
&pred	数値	8	予測ターゲット出現率	unmatch=MISSING 指定かつ _UNMATCH="YES"の場合欠損値となる

注：&pred は pred=パラメータに指定した名前。デフォルトは _CONFまたは_MEAN

アップリフトモデルでは、加えて、data_pred=パラメータに指定した変数名（指定が無ければ、D_CONFまたはD_MEAN）、control_pred=パラメータに指定した変数（指定が無ければ、C_CONFまたはC_MEAN）が出力されます。

13.1.6 欠損値の取り扱い

data=入力データセットに含まれる数値タイプの説明変数に特殊欠損値(._,.A~.Z)が存在した場合は通常欠損値(.)に変換された上で使用されます。

文字タイプのターゲット変数、説明変数はいずれも有効な値の1つとみなされます。

13.1.7 コマンド実行モードでの注意

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

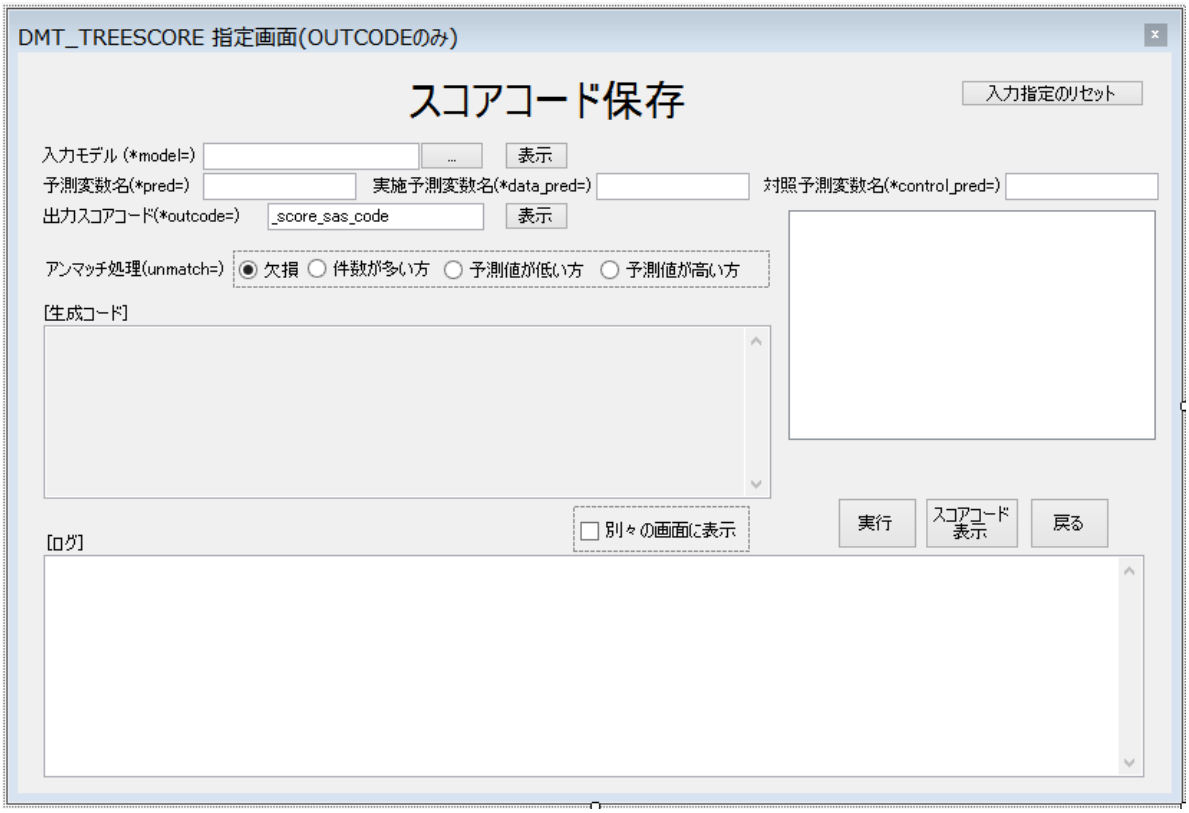
また、以下のユーザ定義フォーマットがWORKライブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。

```
$_item
```

さらに、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。

```
e_name e_type lab&i nobs spc&i typ&i zketa
_speclen _specnum _errormsg
```

13.2 コード保存（dmt_treescore outcode=）



13.2.1 概要

予測値付与SASコード（DMT_TREESCORE）は予測値を付与するためのSASプログラムコードを外部ファイルに出力します。

13.2.2 指定方法

（コマンド実行モードでの指定）

%dmt_treescore(help,model=,
outcode=_score_sas_code,
pred=,data_pred=,control_pred=
unmatch=MISSING,
language=JAPANESE**)**

（GUI実行モードでの変更点）

・help は指定不可。

（必須パラメータ）

以下の1個のパラメータは省略できません。

入力モデル（model=）... 入力モデルデータセット名の指定.

（予測値付与SASコードを出力するためのパラメータ）

以下の2個もしくは3個のパラメータはdata=入力データセットの各オブザベーションに予測値を付与する場合に必須指定です。（=の右辺の値はデフォルト値を表しています）

出力スコアコード（outcode=_score_sas_code）
... 予測値を付与するSASコードを書き出す外部ファイル名の指定.
予測変数名（pred=_CONF(または_MEAN,またはDIF_CONF,またはDIF_MEAN)）
... 予測値を表す変数名の指定.
処理群の予測変数名（data_pred=D_CONF,またはD_MEAN)）
... 処理した場合の予測値を表す変数名の指定.
対照群の予測変数名（control_pred=C_CONF,またはC_MEAN)）
... 対照群に残した場合の予測値を表す変数名の指定.

（アンマッチ処理のためのパラメータ）

アンマッチ処理（unmatch=MISSING）
... アンマッチデータ（モデルのノード分割変数カテゴリに該当しないカテゴリを持つオブザベ

ーション）への対処方法の選択.

（その他のパラメータ）

以下の2個のパラメータは任意指定です。(=の右辺の値はデフォルト値を表しています)

help ... 指定方法のヘルプメッセージの表示.（コマンド実行モードでのみ有効）
言語の選択 (language=JAPANESE)

13.2.3 パラメータの詳細

入力モデル (model=)
入力モデルデータセット名を指定します。このパラメータは省略できません。
例：model=bunseki1

出力スコアコード (outcode=_score_sas_code)
入力データセットの各オブザベーションに予測値を付与するためのSASプログラムコードを外部ファイルに出力します。**デフォルトはSASではSAS起動用ショートカットに定義された作業フォルダー、WPSでは現行のWPSワークスペースの下**に _score_sas_code という名前のファイルに保存されます。保存先ファイル名の物理パスを省略なしで指定する場合も含めて、outcode=c:¥temp¥saspgm1.sas というように常に**引用符なし**で指定します。

アンマッチ処理 (unmatch=MISSING)
出力スコアコードを用いてデータに予測値を付与する過程において、分岐説明変数値がモデルの2つの分岐先ノードのいずれにも該当しないとき（これをアンマッチと呼びます）の対処方法を指定します。

デフォルト値 欠損(MISSING) は予測値（pred=パラメータに指定した変数の値）に欠損値を与え、自動変数_NODEにはマッチした最後のノード名を与えます。

その他 **件数が多い方(FREQ)/予測値が高い方(HIGH)/予測値が低い方(LOW)** のいずれかを指定可能です。 これらの場合は、アンマッチが発生した場合、次のように分岐先ノードを決定して終端ノードまで辿る処理を継続し、予測値を付与します。

FREQはモデル上で該当件数の多い方の分岐先ノード，HIGHはモデル上でターゲット出現率が高い方の分岐先ノード，LOWは低い方の分岐先ノードに強制的に振り分けを行います。

help
パラメータ指定方法をログ画面に表示します。このオプションは単独で用います。（GUI **実行モードでは指定できません。**） 例：%dmt_treescore(help)

言語 (language=JAPANESE)
分析実行中のメッセージ出力、結果の表のタイトル、表項目などの表示言語を選択します。ただし、現バージョンでは、日本語か英語の2種類のみ選択可能です。
例：language=ENGLISH

13.2.4 出力 SAS コードの使用方法

出力したSASコードファイルを用いると、入力データセットに対して、以下のSASステートメントを用いて予測値をつけることができます。

```
data pred_data;
   set input_data;
   %inc "_score_sas_code";
run;
```

ただし、以下の点に注意して用いてください。

（１） input_data にはモデルに採用されたすべての説明変数を含み、かつ、以下の4個の変数が存在しないこと。
_NODE,_TERM,_UNMATCH,&pred（&predはpred=パラメータで指定した名前(デフォルトはモデルによって異なる。_CONFまたは_MEANまたはDIF_CONFまたはDIF_MEAN）

これら4個の変数は%incステートメントで呼び出すコードの中の冒頭において、LENGTHステートメントで変数の型と長さを宣言しています。そのため、input_data にこれらの変数が存在していると %incステートメントの指定が無効になり、エラーが発生します。エラーが発生する場合は、これらの変数をinput_data から削除（drop）した後、用いてください。

（２） %incステートメントとrunステートメントの間には何も書かないこと。
出力されるSASプログラムコードには returnステートメントが存在しますので、%incステートメントの後に追加処理を行うプログラムステートメントを付加しても実行されません。別のDATAステップで追加処理を行うようプログラミングしてください。

```
data pred_data;
   set input_data;
   %inc "_score_sas_code ";
   other statements ...      /* 実行されない */
run;

data pred_data;
   set input_data;
   %inc "_score_sas_code ";
run;
data pred_data2;
   set pred_data;
   other statements ...   /* 実行される */
run;-
```

13.2.5 実行例

例：%dmt_tree(data=samp_data,y=flg,target=1,x=sei--DM,mincnt=50,maxlvl=2,outmodel=tree1)

%dmt_treescore(model=tree1,outcode="C:¥temp¥score_pgm1.sas")

データにモデル予測値をつけるためのSASプログラムコードをテキストファイルに出力します。

OUTCODE=ファイルの例

13.2.6 コマンド実行モードでの注意

実行中にWORKライブラリに _tmp_ で始まる一時データセットがいくつか生成され、実行終了後にすべて削除されます。

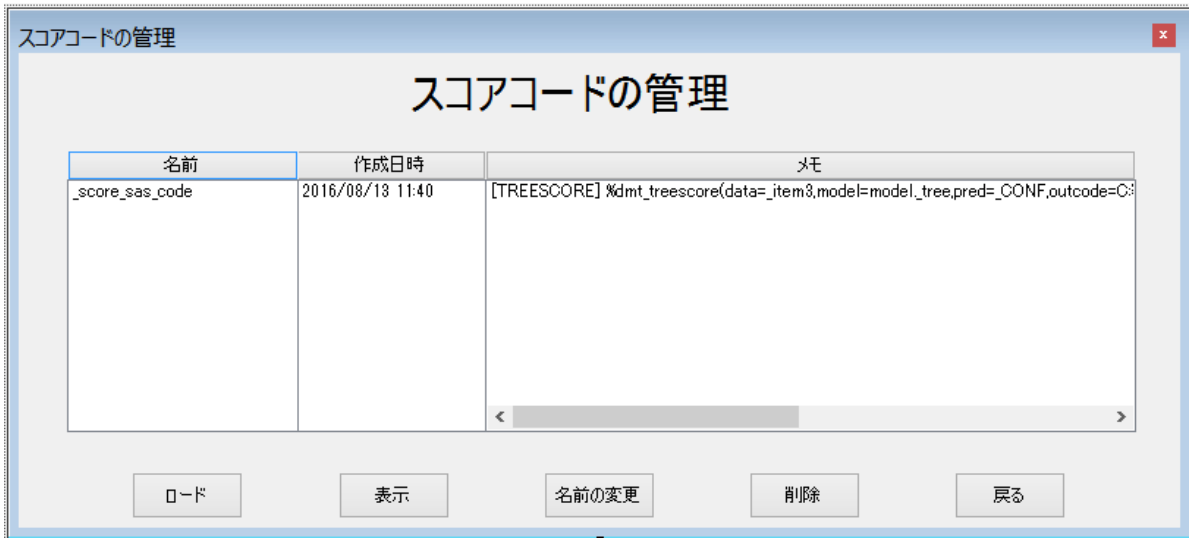
また、以下のユーザ定義フォーマットがWORKライブラリに作成されます。これらは実行後も削除されません。同じ名前のユーザ定義フォーマットは上書きされますので注意してください。

```
$_item
```

さらに、以下のグローバルマクロ変数が作成されます。これらは実行後も削除されません。同じ名前のグローバルマクロ変数は上書きされますので注意してください。

```
e_name e_type lab&i nobs spc&i typ&i zketa
_speclen _specnum _errormsg
```

13.3 コード管理


スコアコードの管理
スコアコードの管理
名前
作成日時
メモ
_score_sas_code
2016/08/13 11:40
[TREESCORE] %dmt_treescore(data=_item3,model=model_tree,pred=_CONF,outcode=C
ロード
表示
名前の変更
削除
戻る

13.3.1 概要

「スコアコード保存」画面で作成したツリーモデル予測値付与プログラムコードファイルを操作（表示・名前の変更・削除）します。
この機能はマクロモジュールには含まれていません。GUI実行モードでのみ指定可能です。

メモ欄の最初の鍵カッコは以下の画面で作成されたことを表します。

[TREESCORE] ... スコアコード保存

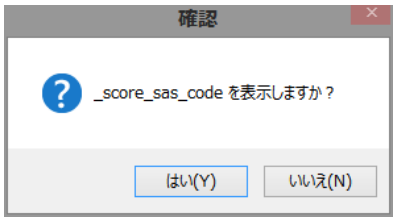
続いてデータを作成したときに実行したプログラムが記述されています。

13.3.2 操作方法

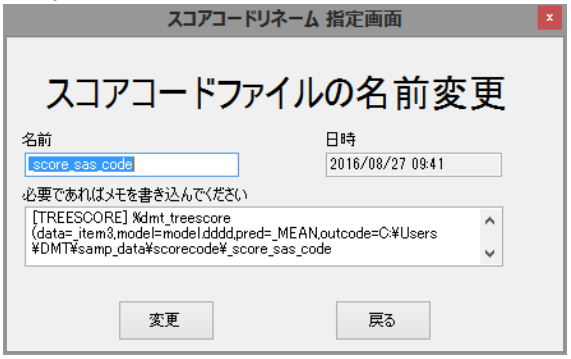
名前,
作成日時,
メモ
リストの上にあるバーをクリックすると、データセットリストを各項目の昇順・または降順で並べ替えることができます。

操作したいコードファイル名をクリックすると、操作ボタンが表示されますので、表示・名前の変更・削除の操作を行います。

表示 プログラムの内容を表示します。

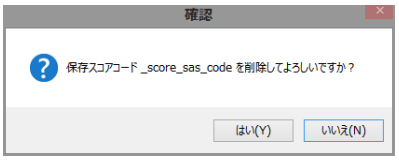


名前の変更 データの名前とメモ内容を確認・変更します。



名前は半角英数字で22文字以内（TEST_の接頭辞や_CV10などの接尾辞が自動的に付けられる可能性があるため）に設定してください。（先頭はアルファベットまたは_(アンダーバー)）

削除 データを削除します。



削除すると、元に戻せません。

(TIPS) 多数のファイルを関連ファイルと一緒にまとめて削除したい場合は、**「設定画面」**の**「分析ディレクトリ」**の下の**「スコアコードディレクトリ」「表示」**ボタンを押し、起動するWindowsエクスプローラで行うと便利です。削除したいデータセット名が書かれたディレクトリをすべて同時選択してから削除します。

14. エラーへの対処方法など

14.1.1 SAS 言語マクロプロセサからのエラーメッセージ(コマンド実行モード)

マクロパラメータの入力間違い等によるエラーはSASマクロプロセッサーからエラーメッセージが出されます。

```
%dmt_trea(data=samp,y=income,target=>50K,mincnt=200,maxlvl=5,x=_all_,dropx=fnlwgt,outmodel=samp_model)
```

```
3710       %dmt_trea(data=samp,y=income,target=>50K,mincnt=200,maxlvl=5,x=_all_,dropx=fnlwgt,outmod
                   -
                   180
3710   ! el=samp_model)
ERROR 180-322: ステートメントが無効か、または順序が正しくありません。
```

```
%dmt_tree(tata=samp,y=income,target=>50K,mincnt=200,maxlvl=5,x=_all_,dropx=fnlwgt,outmodel=samp_model)
```

```
ERROR: キーワードパラメータ TATA はマクロ定義されていません
```

このエラーに対しては、エラー内容を確認し、入力パラメータを訂正して再実行します。

14.1.2 DMT_TREE アプリケーションからのエラーメッセージ(コマンド実行モード)

DMT_TREEアプリケーションは指定できるパラメータ値をチェックし、不適切な値が入力された場合は、エラーメッセージを出して処理を中断します。

```
%dmt_tree(data=samp1,y=income,target=>50K,mincnt=200,maxlvl=5,x=_all_,dropx=fnlwgt,outmodel=samp_model)
```

```
エラー: 指定した入力データセット samp1 が見つかりません.
```

```
%dmt_tree(data=samp,y=income,target=>50K,mincnt=200,maxlvl=5,x=_all_,dropx=flnwgt,outmodel=samp_model)
```

```
エラー: パラメータ DROPX=flnwgt に問題があります. FLNWGT が データセット samp
の中に見つかりません.
```

これに対しても、エラー内容を確認し、入力パラメータを訂正して再実行します。

14.1.3 強制終了後の処置(コマンド実行モード)

dmt_treeやdmt_crossの実行中にユーザーがSASやWPSの処理を強制的に中断した場合は、NONOTESオプションが有効になっている可能性が高く、また、いくつかのデータセットがオープンされたまま残っている可能性もあります。以下のステートメントを最初に入力してください。

```
options notes;
%dmt_release_dsid()
```

その後、セッションが有効かどうかを確認します。確認するには、以下のような簡単なSASプログラムを入力し実行するのが良いでしょう。

```
data a;a=1;run;
```

データセットaが作成されたとの通常のメッセージがログに出れば続けて別の処理を行うことができます。

もしもSASログに通常のメッセージもエラーメッセージも返ってこない場合は、以下の「おまじない」を入力します。

```
;*';*";*/;quit;run;
```

この入力に対して何らかのエラーメッセージが出れば、メッセージを良く読んでから、通常のメッセージが出るようになるまで、簡単なプログラムを入力したりしてセッションの回復状態にします。万一、セッションがどうしても通常状態に戻らない場合は、保存可能なファイルなどを保存した上で、一旦SASまたはWPSを終了し、新たにSASまたはWPSセッションを開始してください。

14.1.4 ライブラリの割り当てを解除する方法(コマンド実行モード)

コンパイル済みマクロカタログライブラリは、DMT_TREEV1.3_SAMPLERUN.sas を実行すると、プログラムの冒頭にある、以下の指定により、mstore というライブラリ名で割当てられた状態にあります。

```
libname mstore "%sysfunc(pathname(sasuser))";
options mstored sasmstore=mstore;
```

mstoreライブラリに存在するマクロを呼び出した後、以下のように通常の方法によりmstoreライブラリの解除を試みても、エラーとなり解除できません。

```
libname mstore;
```

```
ERROR: ライブラリ MSTORE は使用中のため、
クリアまたは再割り当てはできません。
```

ERROR: LIBNAME ステートメントのエラーです。

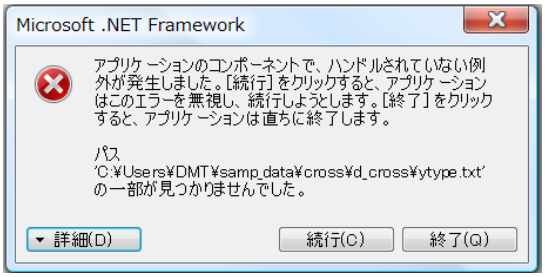
これはマクロカタログライブラリが、割り当てられると占有モードでオープンされる特殊なライブラリであるためです。解除の必要がある場合は、以下のように行います。

```
libname dummy "dummy:¥dummy";   /* 存在しない物理ディレクトリをライブラリ指定する */
options sasmstore=dummy;        /* sasmstore=オプション指定をこのライブラリに変更する */
%dummy ;                        /* 存在しないマクロ呼出を行う (最後の;は必要) */
libname mstore;                 /* mstoreライブラリを解除する */
```

最初の3行（コメントを無視して）を実行すると、それぞれエラーが発生しますが無視します。4行目を実行したとき、以下のようにmstoreライブラリの割り当てが取り消された旨のメッセージが現れると成功です。

NOTE: ライブラリ参照名 MSTORE の割り当てを取り消しました。

14.1.5 Microsoft .NET Framework からの エラーメッセージ(GUI 実行モード)



上記のようなエラーは.NET Frameworkからのエラーメッセージです。エラー内容を確認（通常はファイルが期待された場所に存在しないなどの内容です。）して、一旦終了してから、対応可能であれば対応した後、もう一度GUI実行アプリケーションを起動します。
対応方法が不明の場合は、以下を試してください。

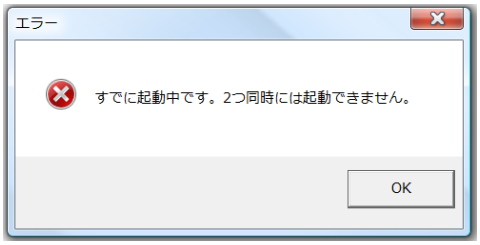
・Windowsのログオフやシャットダウンを行った後GUI画面を再起動する。
・設定画面で新しい分析ルートディレクトリを作成しする。
・データ抽出から順に、分析処理を再実行する。

それでも解決しない場合は、エラー出現箇所とエラー内容をメモしておいて開発元に問合せしてください。

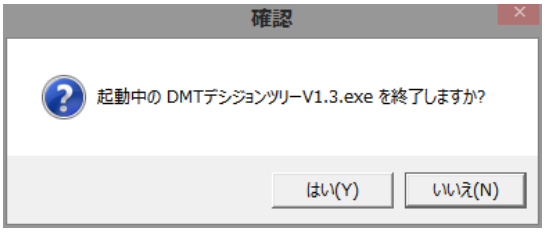
なお、SASで設定した分析ディレクトリをWPSで使用（またはその逆の使用）を行うと、分析ディレクトリ内にファイルタイプが異なるデータセットが混在してしまい、上記のようなエラーが出現する場合がありますので、共用しないでください。（※ これを防止するよう設定画面で警告を出すようにしていますが、設定そのものは可能になっています。）

14.1.6 GUI 実行メニューを 2 つ同時に起動できないというエラー(GUI 実行モード)

GUI画面を１つ起動中にもう１つ起動しようとすると、以下のエラーが現れます。



このエラーメッセージはアプリケーションの複数同時起動を避ける目的でGUI実行アプリケーションが表示しています。しかし、GUI実行アプリケーションが異常終了後に、再度アプリケーションを立ち上げようとした場合、プロセスが残っているために、このエラーメッセージが出現する場合もあります。この場合は、以下のメッセージに「はい」と答えて（DMTデシジョンツリーV1.3.exeのプロセスをすべて終了させます）から再立ち上げを行ってください。（もしも、それでもプロセスが終了しない場合は、一旦ログオフを行ってプロセスを終了させるか、Windowsのタスクマネージャを起動し（Ctrlキー＋Altキー＋Deleteキー同時押し）、プロセスタブを開いて、DMTデシジョンツリーV1.3.exe を終了させてください。）



本アプリケーションからプロセスの終了を指定（「はい」を選択します）

14.1.7 突然 GUI 画面が終了する場合(GUI 実行モード)

WindowsやセキュリティソフトがDMTデシジョンツリーV1.3.exeを有害なソフトウェアと判断して、実行を強制的に停止するため、何かボタンを押すと突然画面が消えてしまう場合があります。このような場合は、実行を許可する設定を行ってください。(通常は許可するかどうかを質問するメニュー画面がポップアップします)

14.1.8 画面から入力データ、クロス分析結果、作成したモデルを選択するボタンで選択画面が開かなくなった場合（GUI 実行モード）

GUI画面実行メニューでは、入力したデータの変数名や型などの情報、作成したモデルの目的変数名や型、採用された説明変数名や型、モデル生成手法などの情報を分析ディレクトリの下の該当するサブディレクトリ内に個別に保存しています。

何等かの理由で、情報の一部が欠けて出力されたデータやモデルのディレクトリがサブディレクトリ内に含まれている場合、そのサブディレクトリのメンバーリストを作成する選択画面が開かなくなります。

このような場合は、設定画面の分析ディレクトリの下の該当するサブディレクトリを開いて、内容を確認してください。
ディレクトリは存在するが必要なファイルが含まれていない等の問題が見つかれば、そのディレクトリは不完全ですので、削除するか再作成して完全なものにしてください。

サブディレクトリ内の全情報が完全であれば選択画面は開くはずです。

15. 付録

15.1 用語の説明

本アプリケーションで用いている主要な用語を説明します。

15.1.1 データ、データセット、変数、オブザベーション

モデルを作成したり、モデルによる予測値をあてはめたりする１まとまりのデータのことをデータセットと呼びます。元来、データは単一の値を意味し、データセットは複数のデータを組織的に集めた単一のファイルを意味するものと考えられます。しかしながら、データセットという呼称が長いこともあり、また使う場面でデータとデータセットを区別できることが多いため、データセットをデータと呼ぶ場合も多くあります。本アプリケーションでもデータセットとデータを区別しないで呼ぶことがあります。
また、データの集合を集合の仕方によって変数（カラム、列または項目）、オブザベーション（インスタンス、行）、データセットと呼びます。変数は同じ属性（たとえば、年齢や性別）を表すデータを集めたものであり、オブザベーションは１つの個体（たとえば、A さん）について複数の変数（年齢、性別、所属など）を集めたものです。

15.1.2 数値タイプ、文字タイプ

変数の持つ特性（プロパティ）の１つ。数値タイプは足し算などの四則演算ができるタイプ、文字タイプはできないタイプのこと。本アプリケーションでは、その変数が含まれる SAS または WPS データセットに定義された変数タイプによって分析変数（ターゲット変数、説明変数）のタイプが決定されます。

15.1.3 ターゲット変数、ターゲット

本アプリケーションでは、モデルの目的変数をターゲット変数と呼び、ターゲット変数の値の中で出現率を予測したい値をターゲットと呼びます。なお、数値タイプターゲット変数の場合は、ある値をしきい値とした上下範囲をターゲットとすることができます。

15.1.4 説明変数

ターゲット出現率の予測に役立つと本アプリケーションのユーザが考え、モデル作成時に指定する変数のことを候補説明変数、または単に説明変数と呼びます。実際にモデルに採用された説明変数のみを指す場合もあります。

15.1.5 モデル、ツリーモデル、ツリー

モデルとは現実の世界の一部を模した仕組みやシステムのこと。ここで扱うモデルは統計モデルの１つで、予測モデルと呼ばれます。これは、（目的変数）＝（説明変数の関数）＋（誤差）の形式で表現され、目的変数の変動を説明変数の値だけを用いてできるだけ近似しようとするモデルです。（説明変数の関数）部分はモデルの種類によってさまざまな形がありますが、ツリーモデルでは、説明変数の値によって逐次的に分岐するノードの集合形式となっています。ツリーモデルのことを単にツリーと呼ぶこともあります。

15.1.6 ノード、親ノード、子ノード、ルートノード、中間ノード、終端ノード

ツリーモデルの用語。ツリーモデルは全体が木構造の形をしており、節点（ノード）と節点間を結ぶ有向結線（アローまたはリンク）の 2 つの要素の組合せで表現されます。出発点のノードは特にルートノード（根ノード）と呼び、下位のノードに向かう結線（出力結線）のみを持ちます。中間ノードは 1 つの入力結線と複数の出力結線を持つノードのことです。そして終端ノード（ターミナルノード）は 1 つの入力結線のみを持ち、他のノードに向かう出力結線を持たないノードを指します。また、親ノード、子ノードは、相対的なノードの位置関係を表す呼称です。たとえば、ノード A からノード B とノード C が直接分岐しているとすれば、ノード A はノード B、ノード C に対する親ノードですが、逆にノード B とノード C はいずれもノード A の子ノードです。しかし、ノード B に別の子ノード D が存在すれば、ノード B はノード D に対する親ノードでもあります。

15.1.7 枝、枝刈り、枝接ぎ

ツリーモデルにおける枝（Branch）とは、特定の中間ノードとその中間ノードにつながっている下位ノードをすべて含む部分木を意味します。枝刈り（Pruning）とはツリーモデルから、部分木を取り除き、ツリーを簡素化する操作のことを意味します。逆に、枝接ぎとは、特定の終端ノードに別のツリーモデルを枝として接ぎ足し、より豊富な

枝を持つツリーにする操作のことを意味します。なお、本アプリケーションでは英単語の長さの関係から枝のことをトウィグ（Twig=小枝）と呼んでいます。また枝刈りをプルーン（Prune）ではなくカット（Cut）、枝接ぎをアッド（Add）と呼んでいます。

15.1.8 AIC 値

AIC（Akaike's Information Criterion=赤池の情報量基準）は広く用いられている統計モデル選択基準の 1 つ。本アプリケーションでは親ノードの分岐に採用する説明変数の優先選択順位を AIC の値によって決定しています。
なお、AIC 値の計算式については、下記の文献の第 6 章分割表解析モデル（P.92～P.106）と第 9 章分散分析モデル（P.155～P.170）を参照してください。

情報量統計学（1983）坂元・石黒・北川　共立出版

なお、DMT_CROSS で表示している AIC の値は以下のように、関連があるとした場合（AIC(モデル)）と関連が全く無いとした場合（AIC(0)）の差をとった値を計算して表示してます。

AIC=AIC(モデル)-AIC(0)

（分割表モデルの場合）
AIC(モデル)= (-2)*(cell-n*log(n))+2*(cat_n*2-1)
AIC(0)= (-2)*(marginal-2*n*log(n))+2*(cat_n+2-2)
ただし、cell はすべての分割表のセルについて Σ{セル件数*log(セル件数)}をとった値、n はデータ件数、cat_n は説明変数のカテゴリ数、marginal はすべての周辺度数について Σ{周辺度数件数*log(周辺度数件数)}をとった値を表します。

この計算式は、説明変数とターゲットの出現有無との関連性を測定する AIC 値と、処理群と対照群間のカテゴリ別の出現率の差の有意性を測定する個別 AIC 値の計算に用いています。

（分散分析モデルの場合）
AIC(モデル)=
n*log(2*3.1415)+n*log(ESS)+n+2*(DF-1+2)
AIC(0)= n*log(2*3.1415)+n*log(WSS0)+n+4
ただし、n はデータ件数、ESS は分散分析モデルにおける誤差平方和、DF はモデルの自由度、WSS0 は全体の修正済平方和、log()は自然対数関数を表します。なお、ESS=0 の場合 AIC= -1E308 としています。

この計算式は、説明変数とターゲット変数との関連性を表す AIC 値と、処理群と対照群間のカテゴリ別の平均値の差の有意性を測定する個別 AIC 値の計算に用いています。

15.1.9 エントロピー

エントロピーはターゲットオブザベーションと非ターゲットオブザベーションの混在度合いを表す量です。ノード内でターゲットと非ターゲットが同じ割合で混ざっているとき最大値をとり、ターゲット出現率が 0 か 1、つまりターゲットと非ターゲットいずれか一方のみが存在するときに最小値をとります。1 つの親ノードに一緒に含まれているときから 2 つの子ノードに分かれた後のエントロピーは、2 つの子ノードのエントロピー値の件数の重み付き平均値として計算されます。本アプリケーションでは分岐後の 2 つの子ノードの重み付き平均エントロピーが最小となるように分岐に用いる説明変数のカテゴリ値を 2 つの子ノードへ振り分けています。

エントロピー計算式は、以下のとおり。

Entropy=-p*log2(p)-(1-p)*log2(1-p)

ただし、p はターゲット出現率、log2()は 2 を底とする対数関数、*は乗算演算子を表します。
p の値は 0 から 1 の範囲ですので、上式からエントロピーの値も 0 から 1 の範囲をとることがわかります。（p=0 または 1 のときエントロピーは 0、p=0.5 のときエントロピーは 1 になります。）

たとえば、親ノードが N=100,p=0.1 とすると、この親ノードのターゲットと非ターゲットの混ざり具合に関するエントロピーは、以下のように求まります。

Entropy(親)=-0.1*log2(0.1)-0.9*log2(0.9)=0.33219

+0.13680=0.46899

この親ノードに含まれるオブザベーションを 2 つの子ノード（N1=40,p1=0.175 と N2=60,p2=0.05）に分けたとすれば、分岐後のエントロピーは、以下のように計算します。

Entropy(子１)=-0.175*log2(0.175)-0.825*log2(0.825)=0.44005+0.22897=0.66902
Entropy(子２)=-0.05*log2(0.05)-0.995*log2(0.995)=0.21610+0.00719=0.22329

Entropy(分岐後)=(N1* Entropy(子１)+N2* Entropy(子２))/(N1+N2)
=(40*0.66902+60*0.22329)/100
=0.40158

元の親ノードのエントロピーは 0.46899 でしたが、2 つの子ノードに分かれた後のエントロピーは 0.40158 と小さくなっています。このように、2 つの子ノードに分かれた後のエントロピーは、分かれる前のエントロピーと比較して、常に等しいか減少します。（p1=p2=p の場合のみ等しくなります。）減少量が大きいほど、件数の重みを考慮したターゲットの出現率の差異が 2 つの子ノード間で大きいことを意味します。

15.1.10 分割レベル、最大分割レベル

分割レベルとは、各ノードのルートノードからの分岐回数を表します。1 回の分岐ごとに説明変数値によって分析データを 1 回分割して 2 つの子ノードを生成するためこのように呼んでいます。最大分割レベルはツリーモデルの生成終了条件の 1 つ。この条件に達したノードは終端ノードになります。

15.1.11 ノード件数、最小ノード件数

各ノードに含まれるオブザベーション件数のことをノード件数と呼びます。最小ノード件数はツリーモデルの生成終了条件の 1 つ。最小ノード件数を満たす 2 つの子ノードを生成できない親ノードは終端ノードになります。

15.1.12 観測比率の標準誤差

データから観測されたターゲット出現率から母集団における真のターゲット出現率との誤差を推計する統計量の 1 つです。データ件数の平方根に反比例します。たとえば、同じターゲット出現率が観測された 2 つのノード（100 件のノード件数を持つノード A と 400 件のノード件数を持つノード B）を比較すると、ノード B はノード A の 4 倍のデータ件数を持つため、観測されたターゲット出現率の真のターゲット出現率との誤差はノード A の半分とみなせます。

計算式は、以下のとおり。

観測比率の標準誤差=SQRT((p*(1-p))/N)

ただし、p はターゲット出現率、N はデータ件数、SQRT()は平方根をとる関数を表します。

15.1.13 2つの観測比率の差の標準誤差

独立した２つの集団 1、集団 2 の観測比率を p1=t1/N1, p2=t2/N2（ただし、N1,N2 は各集団の総件数、t1,t2 は各集団のターゲット件数とします）。集団 1 と集団 2 を併合した集団の観測比率を p=(t1+t2)/(N1+N2) とすると、p1-p2 の標準誤差 は、下記に式により計算されます。

２つの観測比率の差の標準誤差
=SQRT(p(1-p)(1/N1+1/N2))

15.1.14 2つの観測平均値の差の標準誤差

独立した２つの集団 1、集団 2 のそれぞれの件数を N1,N2、観測平均値を m1,m2、観測標準偏差を std1,std2 とすると、m1-m2 の標準誤差は、下記に式により計算されます。

２つの観測平均値の差の標準誤差
=SQRT(((N1-1)*std1^2+(N2-1)*std2^2)/(N1+N2-2)*(1/N1+1/N2))

本アプリケーションでは、上式で N1-1、N2-1 をそれぞれ N1、N2 に置き換えて得られる、以下の式を用いています。

２つの観測平均値の差の標準誤差
=SQRT(std1^2+std2^2)

15.1.15 スタージェスの公式

数値タイプ変数の分布図（ヒストグラム）を作成する場合に推奨されている階級数の計算式。本アプリケーションでは、数値タイプ説明変数のカテゴライズを行うアルゴリズムの中で用いています。

階級数＝1+log2(N)

ただし、N はデータ件数、log2()は 2 を底とする対数関数、本アプリケーションでは CEIL()関数を用いて計算結果を切り下げて整数化しています。

たとえば、N=100 の場合、スタージェスの公式による階級数は、

CEIL(1+log2(100))=CEIL(7.6438...)=8

となります。
なお、本来スタージェスの公式により得られた階級数は、その数値タイプ変数の分布範囲（最大値－最小値）を等間隔に区切り、その区切った範囲に含まれるオブザベーション件数をヒストグラム表示するために用いられます。しかし、本アプリケーションでは、各階級に含まれるオブザベーションが等しくなるような階級のしきい値を求める目的で用いています。

15.1.16 サンプリング、層別サンプリング

一般に、実際の分析対象を選択する際に対象母集団件数が非常に大きい、または既に得られている分析対象データセットの件数が十分大きい場合、その中からランダムに分析対象データを部分抽出することをサンプリング（標本抽出）といいます。サンプリング件数の元の全体件数に対する割合を抽出率と呼び、10%サンプリングを行うといった言い方をします。また、特定の単一カテゴリカル変数の値別または複数カテゴリカル変数の値の組合せ別にサンプリングを行うことを層別サンプリングと呼び、層別サンプリングで無いサンプリングを単純サンプリングと呼びます。たとえば、1000 人の顧客全体から 10%サンプリングを行うと 100 人の顧客が抽出されるが、単純サンプリングの場合は、その中の男女比率は元の顧客全体の男女比率が維持されるとは限りません。一方、性別に 10%層別サンプリングを行うと、男女別にそれぞれ 10%サンプリングを行った結果を結合するため、得られたサンプリングデータにおける男女比率は元のデータセットにおける男女比率に一致します。（ただし、抽出率と層別したときのデータ件数との関係で、完全に等しい比率にできない場合もあり得ます。）

15.1.17 モデル作成用データとモデル検証用データ

分析に用いることができるデータセットをすべてモデル作成に用いると、モデル予測値の精度を検証するデータが残らない点で不都合になります。そこでターゲット別に層別サンプリングを行い、モデル作成用データセットとモデル検証用データセットに分け、モデル作成とモデル検証を別々のデータセットで行うことが一般に行われています。本アプリケーションにも層別サンプリングによりモデル作成用データとモデル検証用データを作成する機能を持っています。

15.1.18 ゲインチャート

予測値の順位がターゲット出現率の順位を反映しているかどうかを判定するための図。CAP（Cumulative Accuracy Profiles）曲線とも呼ばれます。横軸は予測値の大きい順にオブザベーションを並べたときの件数累積百分率を表し、縦軸はターゲット捕捉率（再現率）を表します。縦軸、横軸ともに 0 から 1 の値の範囲をとり、座標（0,0）と（1,1）の 2 つの点を通る曲線を描きます。予測値の順位がターゲット出現率の順位と完全に一致している仮想のモデルは完全モデル、または理想モデルと呼ばれ、そのゲインチャートは座標（0,0）の点と（p,1）（p は全体の平均ターゲット出現率）の点と（1,1）の 3 つの点を直線で結んだ折れ線で表示されます。また、座標（0.0）と（1,1）を結んだ直線（対角線）はランダムな値を予測値とした場合のゲインチャートを表し、ランダムモデル（またはあてずっぽうモデル）と呼ばれます。作成するモデルのゲインチャートは完全モデルとランダムモデルのゲインチャートの中間に位置し、完全モデルに近いほど良いモデルと判断できます。しかし、もしも作成したモデルが完全モデルに非常に近い場合は、むしろ、モデル作成過程（特に用いている説明変数）に問題がある可能性を疑うべきです。

15.1.19 AR 値

AR（Accuracy Ratio）値はモデルの精度評価値の 1

つです。ゲインチャートにおける完全モデルとランダムモデルに挟まれた領域の面積を分母、作成したモデルとランダムモデルの間に挟まれた領域の面積を分子とした比率。作成したモデル予測値の順位とターゲット出現率の順位との対応度合いを数値で表現したもので 0 から 1 の範囲をとります。

AR=分子/分母

ただし、分子はモデルのゲインチャートと対角線に挟まれた領域の面積、分母は完全モデルのゲインチャートと対角線に挟まれた領域の面積です。この面積は台形の面積を求める式を用いて比較的簡単に計算することができます。なお、ROC 曲線の下側領域面積（ROC エリア）と AR は以下の関係があります。

ROC エリア=AR/2+0.5

15.1.20 比較プロット

横軸に予測値、縦軸に実際値をとった散布図のこと。予測値が実際値に一致する点は図の対角線上に並びます。ツリーモデルでは終端ノード単位に散布図の点がプロットされます。ゲインチャートと異なり、予測値の順位ではなく、予測値そのものと実際値との差異を確認できる点で有益です。

15.1.21 R2 乗値と誤差平均平方の平方根

R2 乗値､誤差平均平方の平方根はいずれも誤差(＝実際値－予測値）の観点から見たモデルの精度評価値。R2 乗値は誤差平方和を分子、実際値の偏差平方和（偏差とは各実際値から実際値全体の平均値を差し引いた値のこと）を分母とした比率を 1 から引いた値｡誤差が 0 の場合 R2 乗値は 1 となります。誤差が大きいほど小さな値をとりますが、誤差が大きいと R2 乗値はマイナス値をとる場合もあり得ます。予測値が実際値とずれているような場合、AR 値が 1 であっても R2 乗値は 1 にはなりません。誤差平均平方の平方根（平均 2 乗誤差の平方根）の値は、推計値の平均的な誤差の大きさをターゲット出現率の尺度で表したものです。

R2 乗値=1-誤差平方和／偏差平方和=1-Σ{(y-y_pred)**2}/Σ{(y-y_mean)**2}

ただし、y は実績値、y_pred は予測値、y_mean は実績値の平均値、Σ{}は{}内の式をオブザベーションごとに計算し、それらの合計をとる演算記号、**は累乗演算子です。

15.1.22 正誤表と正答率

正誤表（Confusion Matrix）および正答率（Accuracy）はターゲット予測出現率の値からターゲットが出現するか否かの2つのクラスの予測に変換した上で、実際の状態と比較した場合のモデル精度を評価します。ターゲット予測出現率にあるしきい値を与え、しきい値以上の予測出現率を持つ対象はターゲット出現、しきい値未満はターゲット非出現とみなした2つのカテゴリを持つ予測変数に変換した上で、実際の状態（こちらもターゲット出現もしくはターゲット非出現の2つのカテゴリを持つ）を表す変数とクロス集計を行ったものが正誤表です。正誤表の2*2=4個のセルの内、予測値と実際値が一致している2つのセルが正しく予測できたセル、その他の2つのセルは誤った予測を行ったセルを意味します。正答率は予測が正しかったセルの合計件数を全件数で割った値です。
なお、ターゲット件数が非常に少ない、例えば100件中1件のみがターゲットで残り99件は非ターゲットである場合、100件すべてを非ターゲットと予測しても正答率は0.99と計算されます。このように、状況によっては、みかけ上非常に正答率が高いモデルを安易に作ることが出来る場合があるため、正答率の取り扱いには注意が必要です。

15.1.23 群内平方和と群間平方和

集団のバラツキの大きさを表す統計量。回帰木モデルのノード分岐条件(AIC 条件およびカテゴリ併合方法探索)に用いています。親ノードのターゲット変数 Y のバラツキ(変動)の大きさは、Y の集団内の平均を Ybar とすると、群内修正済平方和 WSS=Σ(Y-Ybar)^2 と表されます。Σ()はデータ件数 n についてすべて足しこむことを表します。WSS をデータ件数で割れば、分散 VAR=WSS/n と呼ばれ、さらに VAR の平方根をとると、標準偏差 SD=sqrt(VAR)と呼ばれます。さて、親ノードが 2 つの子ノードに分かれると、ターゲット変数 Y の変動は、以下のように表されます。

(親ノードの群内平方和 WSS)=(子ノード 1 の群内

平方和 WSS1)+(子ノード 2 の群内平方和 WSS2)+(群間平方和 BSS)

そして、2 つの子ノードの群内平方和の合計(WSS1+WSS2)ができるだけ小さくなる基準で分割に用いる説明変数を探索します。上の式から、群内平方和の合計(WSS1+WSS2)を小さすることは、群間平方和 BSS を大きくすることに他なりません。群間平方和はモデル平方和とも呼ばれ、説明変数のカテゴリに分けることによって元の集団にあった Y の大きな変動を吸収します。

15.1.24 ROC 曲線

ROC 曲線(Receiver Operating Characteristic Curv)は医薬や測定機器分野で良く使われる、診断精度評価図です。これらの分野で用いる場合の ROC 曲線の用語では、正予測の判定を「陽性」、負予測の判定を「陰性」と呼びます。判定結果が正しかったか間違っていたかによって、真陽性、偽陽性(「擬陽性」ではありません)、真陰性、偽陰性に分かれます。また、ターゲット再現率のことを「感度(Sensitivity)」、非ターゲット再現率のことを「特異度(Specificity)」と呼びます。データをモデル予測値の大きい順にならべておいて、縦軸は「感度」(=True Positive Rate)、横軸は 1-「特異度」(正誤表からただちに正予測偽割合(=False Positive rate)に一致することがわかります)をとった点を結んだ曲線が ROC 曲線です。ゲインチャートの完全モデルに対応する ROC 曲線は原点(0,0)と左上の点(0.1)と右上の点(1,1)を結んだ直角の折れ線になります。

15.1.25 ROC エリア

ROC エリアは ROC 曲線の下側面積 AUC(Area Under roc Curv)とも呼ばれ、分類木を含む分類モデルの一般的なモデル精度評価値の 1 つです。ROC エリアの計算式は単純で、ROC 曲線上で原点座標(0.0)と ROC 曲線と右下座標(1,0)に囲まれた部分の面積を計算します。完全モデルの場合 ROC エリア=1、ランダムモデルの場合 ROC エリア=0.5 となり、1 に近いほど精度が高いことを意味します

なお、AR 値と以下の関係式が成立します。

ROC エリア=AR/2+0.5

15.1.26 名義尺度・順序尺度・循環尺度

名義尺度・順序尺度・循環尺度は一般に文字タイプの変数の尺度の分類です。文字変数の値(カテゴリ)に順序関係(特定のカテゴリ同士が隣接するという関係)が全く無いとみなす場合、その文字変数は名義尺度と呼ばれ、特定の 2 つのカテゴリ同士が隣接しあって全カテゴリが決まった順番に並ぶとみなす場合には順序尺度と呼ばれ、さらに順序尺度の最初のカテゴリと最後のカテゴリが相互に隣接していると仮定する場合は循環尺度と呼ばれます。これらの尺度は分析者によって自由に決めることができるものです。例えば、"A","B","C"という 3 つの値を持つ文字変数は分析のときに名義尺度、順序尺度、循環尺度いずれにも取扱えます。

一方、数値タイプの変数の場合は、とびとびの値をとるのでなければ、間隔尺度(差の大きさに意味がある尺度。例えば速度)もしくは比尺度(正の値をとり、倍数に意味がある場合の尺度。例えば身長や体重)となります。しかしながら、ツリーモデルに用いる場合は、数値タイプ変数は変動範囲の中のある値を境としてカテゴライズが行われるため、間隔尺度あるいは比尺度の性質は失われ、カテゴリ間の順序関係だけが残されて、順序尺度もしくは循環尺度の取扱いになります。なお、数値タイプの説明変数をカテゴライズした後に名義尺度として取扱うことも考えられます。しかしながら、そのような取扱いが必要なのは、一般に連続変数ではなく、離散的な値をとる数値変数のではないかと推察します。もしもそうであれば、数値変数ではなく、文字変数として入力することによって実現可能なので、DMT_TREE では数値タイプ説明変数を名義尺度として取扱うことはできないようにしています。

15.1.27 線形回帰モデル

数値変数の予測や数値変数の変動要因の分析を行う代表的な統計モデル。以下の式で表現されます。

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

ただし、y はターゲット変数、$x_1, x_2 \ldots x_k$ は説明変数、$\beta_0, \beta_1, \beta_2 \ldots \beta_k$ は未知の係数(パラメータ)、ε は誤差項で、平均0分散一定の正規分布を仮定します。

ここで、β_0 は特に切片項（定数項とも）と呼ばれ、全説明変数値=0のときのターゲット変数の期待値を表します。もしも全ての説明変数が原点0を持つ比例尺度変数であり、すべての説明変数値=0のときターゲット変数値も0になるべきと考えられる場合は、切片項=0（切片項なし）という制約を与えるべきです。

文字タイプ説明変数は各値ごとにダミー変数（該当すれば1,非該当の場合0の値をとる2値変数）に変換された上で上式の説明変数に加えられます。

パラメータ推計値は誤差 ε の2乗和が最小になる基準で決定されます（最小2乗法）

線形回帰モデルの予測値 $\hat{y}$ は以下の式で与えられます。

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

15.1.28 線形ロジスティックモデル

ターゲット値の出現確率の予測やターゲット値の出現確率に影響を与える要因分析を行う代表的な統計モデル。以下の式で表現されます。

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

ただし、$x_1, x_2 \ldots x_k$ は説明変数、
$\beta_0, \beta_1, \beta_2 \ldots \beta_k$ は未知の係数（パラメータ）、
$\hat{p}$ はパラメータ $\beta_0, \beta_1, \beta_2 \ldots \beta_k$ のセットと説明変数 $x_1, x_2 \ldots x_k$ のセットが与えられた場合のターゲット予測出現率を表します。

なお、$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$
とおいて、上式を $\hat{p}$ について解くと、

$$\hat{p} = \frac{\exp(Z)}{1+\exp(Z)}$$

と表現できます。（予測式）

さて、パラメータ $\beta_0, \beta_1, \beta_2 \ldots \beta_k$ の仮説下で実際にターゲット値が出現したオブザベーションは説明変数 $x_1, x_2 \ldots x_k$ に対するターゲット予測出現率 $\hat{p}$ の確率、ターゲット非出現のオブザベーションは $(1-\hat{p})$ の確率で事例が発生すると予測したとみなすことができます。（仮説と事例が関連しているほど、この確率は大きな値をとることがわかります。）

仮説下でのすべてのオブザベーションのターゲット事例の出現確率（同時確率）は、各オブザベーションの出現確率をすべて掛け合わせることによって得られ、**尤度**（L）と呼ばれます。

$$L = \prod \{(\text{target} = 1) * \hat{p} + (\text{target} = 0) * (1-\hat{p})\}$$

ただし、target=1 はターゲット出現事例のオブザベーション、target=0はターゲット非出現事例のオブザベーションを意味します。

パラメータ推計値は、尤度 L 最大（対数をとった**対数尤度** $\ln(L)$ 最大と同等）基準で決定されます。（最尤法）

15.1.29 アップリフトモデル

アップリフトモデルはマーケティング分野から発展したデータマイニングモデルの１つです。

一般に、マーケティング施策を実施すると購入が増える顧客とそうでない顧客がいると考えられます。DMに反応して購入金額が増える顧客（A）もいれば、逆にDMに反発して購入を取りやめる顧客（B）もいるかもしれません。また、放置しておいてもたくさん買ってくれる顧客（C）もいるかもしれませんし、DMを出しても出さなくても全く買う気が起きない顧客（D）もいるかもしれません。

アップリフトモデルは、売上そのものではなく、施策実施効果（売上の増加分）が高い／低い顧客集団を見分けることに関心があり、顧客を上記の（A）から（D）のいずれかに分類し、施策実施先の最適化に用いるためのモデルです。

モデルの説明変数の統計的有意性を評価する方法は、さまざまな方法が考えられますが、一つの方法は以下のとおりです。

(1) 施策実施群のデータセット（data）と対照群のデータセット（control）を縦に連結したデータセットを作成し分析データとします。

(2) データから、説明変数ごとに以下のモデルを構築し、AとBの交互作用効果の有意性をAICその他の統計量で評価します。

モデル: 目的変数= Aの主効果＋Bの主効果
＋AとBの交互作用効果

ただし、
Aは当該説明変数、
Bは施策実施群と対照群を識別するダミー変数。

上記の方法による説明変数ごとのアップリフト効果の推定は、ロジスティック回帰モデル、または分散共分散分析モデルを用いて比較的簡単に行えます。しかし、ロジスティック回帰モデルで発生の可能性のあるエラー（「準完全分離」）を回避するため、本アプリケーションでは上記と同じ考え方を独自のアルゴリズムで実現しています。

15.2 お問合せ先

本マニュアルに関するご質問、その他のお問合せは以下の宛先までお願いします。

データマインテック株式会社
分析ツール開発部マニュアル担当

〒201-0004 東京都狛江市岩戸北 3-3-6-405
info@dataminetech.co.jp

なお、本マニュアルは予告なく改訂される場合があります。下記のホームページで公開する最新のマニュアルをご参照ください。

http://www.dataminetech.co.jp/