

## <演習問題>

銀行の架空の顧客データを用いて分析を行います。今回の目的は「ネットバンク」の契約に関する分析です。

### 1. ファイルの読み込み

Excel ファイル [練習問題データ.xlsx] のシート [Data] を読み込み、データセット「EX1」を作成してください。列の属性などは、シート[定義表]を参考にしてください。

### 2. 要約統計量の算出

データセット「EX1」を使用し、要約統計量タスクを用いて、以下 2 つのレポート表示してください。

- A) ネットバンク契約有無別の普通預金額に関する要約統計量・及びヒストグラム
- B) 年齢別のネットバンク契約有無状況（度数集計表）

### 3. データクレンジング

クエリビルダを用いて、データセット「EX1」の普通預金額と定期預金額とローン残高を対数変換してください。

- ・預金額や、残高が 0 の場合、対数変換ができないので 1 を足して対数変換をしてください。  
(X に対して、 $\log(X+1)$ という変換をする)
- ・それぞれ、変数名を Log\_Yokin、Log\_Teiki、Log\_Loan とします。
- ・クレンジング後のデータセットを「EX2」という名前で作成してください。

### 4. SAS Enterprise Guide を用いたロジスティック回帰

クレンジング後のデータセット「EX2」を用いて、ネットバンク契約状況を予測するためのロジスティック回帰モデルを作成してください。モデル選択方法として Stepwise 法を用いて、有意水準は 5%とします。使用する変数は下記の通りとします。

性別、	年代、	普通預金（対数変換後）、
定期預金（対数変換後）、	住宅ローン残高（対数変換後）、	
ATM 利用回数、	給与振込口座、	投資信託利用有無

### 5. SAS Enterprise Guide による新規データの予測

前の演習問題で推定したモデルと、新規データ「FORPRED」を用いて、ネットバンク契約に関する予測確率を算出してください。

## 6. SAS Enterprise Miner を用いたロジスティック回帰

- ・SAS Enterprise Miner を起動し、プロジェクト、ダイアグラムを定義してください。その後、SAS データセット「EXERCISE」を SAS Enterprise Miner に定義してください。
- ・データノードをダイアグラムに配置してください。その後、データ分割ノードを配置します。プロパティパネルで「学習データ：検証データ」=「75%：25%」と設定し、データ分割ノードを実行してください。
- ・回帰ノードをダイアグラムに配置し、ステップワイズ法（有意水準 5%）によるモデル選択を行ってください。使用する変数は、4 と同じとします。さらに、ステップワイズ法のモデル選択の過程の中から、「平均平方誤差」の最小となるモデルを特定してください。

## 7. SAS Enterprise Miner を用いたツリー分析

ツリーノードをダイアグラムに配置してください。プロパティパネルの「最大の深さ」を「5」、「葉のサイズ」を「200」にそれぞれ設定し、平均平方誤差の最小となるツリーモデルを作成してください。

（自己学習）「入力を一度だけ使用する」を「いいえ」とした場合のツリーと比較してみてください。また、対話モードによるツリーも作成してみてください。

※自己学習の回答は割愛いたします。

## 8. SAS Enterprise Miner による新規データの予測

SAS データセット「FORPRED」を SAS Enterprise Miner に定義してください。その際、予測データとして使用するために、役割を「スコア」としてください。

6.の回帰モデルで構築したモデルによる予測確率を算出してください。（データノードとスコアノードを用います。回帰ノードと接続させます。）