Chapter 1 デー	-タソースの結合	1-1
1.1.1	データソースの結合の概要	1-2
1.2.1	レコード追加の概要	1-3
1.2.2	レコード追加ノード -データソースの読込み-	1-4
1.2.3	レコード追加ノード -メインデータセット-	1-8
1.2.4	レコード追加ノード -すべてのデータセット-	1-14
1.2.5	レコード追加ノード -フィールド名の調整-	1-17
1.3.1	レコード結合の概要	1-21
1.3.2	レコード結合ノード -データソースの読込み-	1-22
1.3.3	レコード結合ノード -データソースの結合-	1-24

Chapter 2 フィ	'ールドの作成	2-1
2.1.1	フィールド作成の概要	2-2
2.2.1	CLEM式オプションによるフィールド作成	2-3
2.3.1	フラグ型オプションによるフィールド作成	2-10
2.4.1	名義型オプションによるフィールド作成	2-15
2.5.1	(参考)ステート型オプションによるフィールド作成	2-20
2.6.1	(参考)カウント型オプションによるフィールド作成	2-21
2.7.1	条件付きオプションによるフィールド作成	2-22
2.8.1	データ分類ノードによるフィールド作成	2-27
2.9.1	スーパーノードによるカプセル化	2-33
Chapter 3 フィ	'ールドの変換	3-1
3.1.1	フィールド変換の概要	3-2
3.1.2	ストリームの確認	3-3
3.2.1	フラグ設定ノードの概要	3-4
3.2.2	フラグ設定ノード	3-5
3.3.1	再構成ノードの概要	3-11
3.3.2	再構成ノード	
3.4.1	置換ノードの概要	3-19
3.4.2	置換ノード	

Chapter 4	レコード設定	4-1
4.1.1	レコード設定の概要	4-2
4.1.2	ストリームの確認	4-3
4.2.1	条件抽出ノード	4-4
4.3.1	サンプルノード	4-13
4.4.1	レコード集計ノードの概要	4-18
4.4.2	レコード集計ノード	4-19
Chapter 5	日付データの加工	5-1
5.1.1	日付データの概要	
5.1.2	ストリームの確認	5-3
5.2.1	ストレージと入力形式の確認	5-5
5.2.2	レコードのソート	5-6
5.3.1	インデックスフィールドの作成	5-9
5.4.1	日付フィールドの計算-経過日数の算出	5-16
5.4.2	日付フィールドの計算-日付や時刻の一部を抽出	
5.5.1	レコード集計	
Chapter 6	ノードの生成	6-1
6.1.1	ノードの生成の概要	6-2
6.1.2	ストリームの確認	6-3
6.2.1	棒グラフによるデータ分類ノードの生成	6-4
6.3.1	ヒストグラムによるフィールド作成ノードの生成	6-11
6.4.1	棒グラフによるバランスノードの生成	
6.5.1	テーブルノードによる条件抽出ノードの生成	
6.6.1	散布図ノードによる領域の条件抽出ノードの生成	

本書では、IBM SPSS Modeler 18.2Jを使用しています。

IBM およびSPSSは、International Business Machines Corp.の登録商標です。

本書を無断で複写複製(コピー)することは、著作権法上の例外を除き、禁じられています。



IBM SPSS Modelerでは、目的や用途に応じたさまざまなノードを利用することで、フィールド形 式を変換することができます。例えば、カテゴリ型フィールドのフラグ型フィールドへの変換や、 欠損値を特定の値に置換、ストレージの変更などを行うことができます。

Contents



§3.1.1 フィールド変換の概要

フィールドの変換は、データの準備フェーズで良く利用されるデータ加工の機能です。例 えば、カテゴリ型フィールドに基づいて2値のフラグ型フィールドを作成したり、フィールド 名の変更や除外を行うことなどができます。

フィールド設定パレットには以下のようなさまざまなノードが含まれています。

	フラグ設定	名義型フィールドに基づいて、複数のフラグ型フィー ルドを派生させるノードです。
E	再構成	名義型またはフラグ型フィールドについて、連続型フ ィールドの値で再構成させるノードです。
	行列入替	レコードとフィールドを入れ替える場合に使用するノ ードです。
	匿名化	フィールド名や値の下流の表示方法を変換し、元のデ ータを隠す場合に使用するメードです。
	置換	フィールド値の置換やストレージの変換を行う場合に 使用するメードです。
×	フィルター	下流の渡すフィールドの絞り込みや、フィールド名の 変更を行う場合に使用するノードです。
	フィールド順序	下流のノードに表示するフィールドの順番を定義する ためのノードです。
RFM	RFM分析	リーセンシ、フリクエンシ、マネタリー(RFM)の分析 を行うノードです。

Table3.1.1 フィールド設定パレットに含まれる主要なノード

ここでは、フラグ設定、再構成、置換ノードの基本的な使用方法を確認します。

§3.3.1 再構成ノードの概要

再構成ノードは、名義型フィールドとして定義されたカテゴリ値を元にして、複数の連続 型フィールドを派生する場合に使用します。フラグ設定ノードでは、派生されたフィールド には**T**または**F**が保持されますが、再構成ノードでは連続型フィールドの値を保持させる場合 に用います。

例えば、下記のように顧客IDごとに購入した商品が記録されている場合、商品フィールド は名義型フィールドになります。このフィールドに基づいて、カテゴリ値ごとに新規フィー ルドを構成し、それぞれの商品を購入した数量を表す連続型フィールドを構成する場合に使 用できます。



再構成ノードでは、フラク設定ノードと異なり連続型フィールドがデーセットに追加される ことになります。なお、該当する値がないセルは欠損値になります。

TIPS

フラグ設定ノードと異なり、再構成ノードでは集計キーによる集約を行うことはできません。再構成の結果について顧客IDなどによる集計を行う場合は、レコード設定パレットに含まれるレコード集計ノードを使用してください。

§3.3.2 再構成ノード

再構成ノードは、名義型フィールドとして定義されたカテゴリ値を元にして、複数の連続 型フィールドを派生する場合に使用します。

ここでは、支払方法として、現金、カード、ポイントの3カテゴリを持つ名義型フィールド に基づいて、これらの支払金額を保持するフィールドに変換します。

操作手順

1. フィールド設定パレットを開きます。



Figure3.3.3 データ型ノードからリンクした再構成ノード

4. 再構成ノードをダブルクリックして編集画面を開きます。



Figure3.3.4 フラグ設定ノードの編集画面

利用可能なフィールドとして、再構成の対象となる名義型フィールドを指定すると、利用 可能なセット値にカテゴリ値が表示されます。再構成フィールドとして作成したいカテゴリ 値を右側の作成リストに移動します。

5. 利用可能なフィールドのリストから支払方法を選択します。



利用可能な値に、支払方法フィールドに含まれるカテゴリ値が表示されますので、再構成 が必要なカテゴリ値を、再構成フィールドの作成リストに移動します。

6. **カード、ポイント、現金**を再構成フィールドの作成リストに移動します。



Figure3.3.6 値の挿入ダイアログボックス

カテゴリ値がそれぞれの新規フィールドとして作成されます。次に、再構成の結果として 保持する連続型フィールドを**値フィールド**に指定します。この例では、**支払金額**を指定しま す。

7. 値フィールドのリストから支払金額を指定します。



Figure3.3.7 値フィールドの指定

以上の設定で、再構成されたフィールドには支払金額が保持されます。

操作手順

8. OKボタンをクリックします。

9. 出力パレットからテーブルノードを挿入し、再構成ノードにリンクします。



Figure3.3.8 テーブルノードの挿入

操作手順

頭 テーブル	レ(20 フィール	レド、 8,632 レコ−ド) #1		– – ×
	(II (D)			
ידר 😡	1 JV(<u>F</u>)	編集(上) (***********************************	勝(G) ■ 「 1・ 「 1・ 「 1・ 「 1・ 「 1・ 「 1・ 「 1・	ч у Х
テーブル	注釈			
	製約 支	ム方法 カード 支払金額	↓ ■ 支払方法_ポイント_支払金額	支払方法_現金_支払金額
1	0	\$nul	\$ 3179	\$null\$
2	0	\$nul	\$ \$null\$	9888
3	1	\$nul	\$ \$null\$	9173
4	0	\$nul	\$ 4279	\$null\$
5	1	572	0 \$null\$	\$null\$
6	0	616	4 \$null\$	\$null\$
7	0	\$nul	\$ 4365	\$null\$
8	0	\$nul	\$ \$null\$	14901
9	1	\$nul	\$ \$null\$	9440
10	0	\$nul	\$ \$null\$	9049
11	0	\$nul	\$ 4319	\$null\$
12	0	\$nul	\$ \$null\$	6518
13	0	\$nul	\$ \$null\$	14281
14	1	1063	9 \$null\$	\$null\$
15	1	1276	2 \$null\$	\$null\$
16	1	\$nul	\$	4644
17	0	\$nul	\$ \$null\$	5427
18	0	2006	5 \$null\$	\$null\$
19	0	\$nul	\$ 6470	\$null\$
20	0	\$nul	\$ \$null\$	11751

Figure3.3.9 再構成ノードによって作成されたフィールド(フィールド順序変更済)

該当する値がない場合は、欠損値として\$null\$が含まれます。

操作手順

11. **OK**ボタンをクリックして、**テーブル**ノードを閉じます。

POINT

再構成ノードを使用することで、名義型フィールドのカテゴリ値に基づいて、その他の連 続型フィールドの値を保持する新しいフィールドを作成することができます。



6 ノードの生成

新しいフィールドの作成やカテゴリの分類を行う場合、IBM SPSS Modelerではフィールド作成ノ ードやデータ分類ノード、データ分割ノードなどを使用することで、目的のフィールドを作成する ことができます。その他、出力結果のウィンドウに含まれるノードの生成メニュー機能を用いるこ とでも、新しいフィールド作成を行うためのノードを、ストリームに追加することができます。

Contents



§6.1.1 ノードの生成の概要

IBM SPSS Modelerのテーブルやグラフなどの出力には、**生成**メニューが用意されており、 さまざまなノードをストリームに追加することができます。生成メニューを利用することで、 データ分析を行いながら効率的なストリーム作成がサポートされます。

🔄 支払方法 の棒グラフ				_		×
違 ファイル(E) 🛛 📄 編集(E)	[₺] 生成(<u>G</u>)	💰 表示(⊻)		1 ∰	?	×
テーブル ガラっ 注釈	バンドのフィ	ールド作成ノー	ዞ <u>" (D)</u>			
	領域のフィー	・ルド作成ノード	(フラグ	'型)(E)		
值/	領域のフィー	・ルド作成ノード	(セット	型)(<u>S</u>)	度	<u>教</u>
ポイント	領域の条件抽	記ノード(<u>R</u>)			в 1	2662
現金	マークした裏	素の条件抽出ノ	<u>س الا (M</u>)		1	4345
	マークした裏	素のフィールド	作成ノ・	۴(E)		
	テーブルのノ	ード選択				
	テーブルのフ	ィールド作成人	- 15			
	バランス 🌽	- ド (増加)(B)				
	バランスノー	-ド(減少)(R)				
	データ分類ノ	ード (グループ)(<u>C</u>)	N		
	データ分類ノ	'ード(値)		7		
<u> </u>						ок

Figure6.1.1 ノードの生成メニューの例

ノードの生成メニューは、IBM SPSS Modelerの出力に用意されており、代表的な使用方法 として、テーブル結果からの**条件抽出**ノードの作成、棒グラフからの**データ分類**ノードの作 成、ヒストグラムからの**データ分割**ノードの作成、予測モデルからの**フィルター**ノードの作 成、散布図からの**領域の条件抽出**ノードの作成などが挙げられます。

POINT

出力結果に含まれるノードの**生成**メニューを使用することで、さまざまなノードを簡単に 取り出すことができます。

§6.6.1 散布図による領域の条件抽出ノードの生成

領域の条件抽出ノードは、グラフから抽出できる条件抽出ノードです。例えば、散布図を 作成して外れ値と思われる値をレコードから除外する場合などに使用することができます。

ここでは、**年齢**と**支払金額**の**散布図**を作成し、特定の領域に含まれるレコードを条件抽出 ノードとして選択できるようにします。

操作手順

- 1. グラフ作成パレットから散布図ノードをストリームキャンバスに挿入します。
- 2. データ型ノードから散布図ノードをリンクします。
- 3. 散布図ノードをダブルクリックして編集画面を表示します。
- 4. Xフィールドに年齢を選択します。

🛞 ? v. ?		\mathbf{U}	>
<u>X</u> :年齢	Y支払金額	0	
プロット オブ	ション 外観 出力 注釈		
L Xフィール ビ	ド: 🗗 年齢 🛛 🚽 Y	フィールド: 🔗 支払金額	-
オーバーレイ ――	↓ サイズ:		_
パネル:	→ アニメーション:	→ 通過度:	-
オーバーレイ タイナ	注 💿 なし 〇 平滑化		
	○関数 y =		
ОК	▶ 実行(1) キャンセル	適用(A)	Utz vs k (R)

Figure6.6.1 散布図のフィールドの指定

6. 実行ボタンをクリックします。



Figure6.6.2 年齢と支払金額の散布図

散布図の特定の領域を選択するためには、ツールバーの選択範囲をアクティブにしますボ タンを使用します。ボタンが表示されていない場合は、**表示**メニューから**インタラクティブ** を選択してください。

7. 選択範囲をアクティブにしますボタンをクリックします。



Figure6.6.3 選択範囲をアクティブにしますボタンの使用

8. 年齢の50歳以上から70歳まで、支払金額の20000から30000の範囲をドラッグして選 択します。



Figure6.6.4 領域の選択

TIPS

選択範囲が複数ある場合は、そのまま続けて範囲を選択します。また、一度選択した範囲 を取り消す場合はその領域の上で右クリックし、**四角形の削除**または**領域の削除**を選択し ます。



9. 生成メニュー > 領域の条件抽出ノードを選択します。

Figure6.6.5 散布図のノードの生成メニュー

領域のフィールド作成ノード(フラグ型)を使用すると、領域内のデータと領域外のデータ を2つに分割するフラグ型のフィールド作成ノードが作成されます。**領域のフィールド作成ノ** ード(セット型)を使用すると、領域を複数指定した場合に、領域ごとに分割するフィールド 作成ノードが作成されます。**領域の条件抽出ノード**を使用すると、選択した領域のレコード を選択するための条件抽出ノードを作成することができます。

操作手順

10. 領域の条件抽出ノードを選択します。



Figure6.6.6 生成された条件抽出ノード

ストリームキャンバスに(生成)条件抽出ノードが作成されます。

操作手順

11. (生成)条件抽出ノードをダブルクリックして編集画面を表示します。



Figure6.6.7 生成された条件抽出ノードの条件式

散布図で選択した領域内の値が、条件式として自動的にセットされていることが確認でき ます。条件の値は、選択した値に基づいて詳細に設定されているため、目的に応じて適時修 正を行ってご使用ください。

操作手順

12. OKボタンをクリックして、条件抽出ノードを閉じます。