

【データ加工処理詳細】

5. FNファイルから処置有無を追加 (1. ENファイル～のEN_06～08をFNファイルで)

1. FN_06 処置	
1) レセコードリスト作成	
(1) グループ項目	施設コード データ識別番号 入院年月日(西暦) レセコード
(2) 並び替え	施設コード データ識別番号 入院年月日(西暦) レセコード
2) MYDBに保存する	
2. FN_07 処置有無	
1) レセコード(EN_06)から処置一覧対象の有無を判定	
(1) 計算値	例) A型ボツリヌス毒素(ボトックス)_有無=MAX(IF(コード='xxxxxxxx',1,0))⇒「処置一覧」参照
(2) グループ項目	施設コード データ識別番号 入院年月日
(3) 並び替え	施設コード データ識別番号 入院年月日(西暦)
2) MYDBに保存する	
3. FN_08 ALLKEYLISTコピー	
1) 1～4(EN～JOIN)の加工処理で得られたデータのKEYリストをコピーする	
(1) 対象TBL	ALLKEY_LIST
2) MYDBに保存する	
4. FN_09 処置有無出力	
1) FN_08と一致するデータをFN_07から抽出し、CSV化する	
(1) 対象TBL	FN_07 FN_08
(2) 結合条件	施設コード データ識別番号 入院年月日
(3) 計算値	例) A型ボツリヌス毒素(ボトックス)_有無=IF(FN_07.A型ボツリヌス毒素(ボトックス)_有無=1,"有","無")
2) CSVへ保存する	

【データ加工処理詳細】

7. 疾病別データマート作成

1. 疾病別レセコードリストを作成	
1) 入院別レセコードリストを作成する	
(1) 対象TBL	FN(DB2上)
(2) グループ項目	施設コード データ識別番号 入院年月日(西暦) レセプト電算処理コード コードカウント=明細数
(3) 抽出条件	20030701 ≤ 入院年月日 ≤ 20031031
(4) 並び替え	施設コード データ識別番号 入院年月日 レセプト電算処理コード
2) 結果(csv)を、DB2へロード	
2. KEYリスト作成	
1) EN～JOINの加工処理で得られたデータのKEYリストを作成する	
(1) KEY	施設コード データ識別番号 入院年月日 退院年月日 診断群分類番号
(2) 並び替え	施設コード データ識別番号 入院年月日
2) 結果(csv)を、DB2へロード	

【データ加工処理詳細】

3. 各マスタ(薬剤・薬物療法名・材料・診療行為)とレセ電コードリスト、およびKEYリストをマッチングし、各種計算を行う。

1) 各マスタ(薬剤・薬物療法名・材料・診療行為)とレセ電コードリスト、およびKEYリストをマッチングし、各種計算を行う。
 ※マスタごとに処理を行う。

(1) 対象TBL

薬剤/薬物療法名マスタ
材料マスタ
診療行為マスタ
1.レセ電コードリスト
2.KEYリスト

(2) 結合条件

各マスタとレセ電コードリスト
レセプト電算処理コード
レセ電コードリストとKEYリスト
施設コード
データ識別番号
入院年月日

(3) 計算値

合計=Sum(レセ電リスト、コードカウント)
 入院数=明細数
 1入院平均件数=合計/入院数
 最大=Max(レセ電リスト、コードカウント)
 最小=Min(レセ電リスト、コードカウント)
 金額(1入院平均件数×単価)=1入院平均件数×マスタ、単価 ※薬物療法名を調査時には、単価→平均金額
 金額(合計×単価)=合計×マスタ、単価 ※薬物療法名を調査時には、単価→平均金額

(4) 抽出条件

診断群分類番号が*で始まるもの

例)

対象	*
全ての疾病	ろ
MDC=01	01%
MDC6=060050	060050%

(5) グループ項目

レセプト電算処理コード

(6) グループ条件

薬物療法名のみ:薬物療法名≠Null

(7) 並び替え

金額(1入院平均件数×単価) (降順)

2) 各結果(疾病ごと、マスタごと)をエクセルへ保存

4. 上記結果ファイルから、疾病ごとに追加する項目を選択し、追加項目リストを作成

1) 上記結果ファイルから、疾病ごと(MDC6別)に追加する項目を選ぶ

(1) 追加項目 選択基準

金額(1入院平均件数×単価)	入院数	他
薬物療法名	¥1,000以上	10件以上 上位20件
薬剤	¥1,000以上	10件以上 薬物療法名に含まれていない、上位20件
材料	¥1,000以上	10件以上 上位20件
診療行為	500点以上	10件以上 加算等は含まない、上位50件 (診療行為名称に右表を含まない)

外泊	入院時食事療養(1)
加算	標準負担額
減算	特定薬剤治療管理料
基本料	薬剤管理指導料
診断料	処方料(その他)
実施料	診療情報提供料(D)
判断料	

2) 選択した項目のレセ電コード&MDC6のリストをDB2にロード

【データ加工処理詳細】

5. 疾病別データ抽出

1) EN~JOINの加工処理で得られたデータから、疾病ごと(MDC6別)にデータ抽出をする。
 ※外部プログラムで行う。分析対象疾病分。

(1) 抽出条件

MDC6=(各疾病に該当する基本DPC)

(2) 並び替え

施設コード
 データ識別番号
 入院年月日

2) CSVへ保存 ①

6. KEYリストと追加項目リストをマッチングし、疾病ごとに追加項目をKEYに付加する

1) KEYリストと追加項目リストをマッチングし、疾病ごとに追加項目をKEYに付加する
 ※外部プログラムで行う。分析対象疾病分。

(1) 対象TBL

2.KEYリスト(主表)
4.追加項目リスト
1.レセ電コードリスト

(2) 結合条件

疾病別KEYリストとレセ電コードリスト
施設コード
データ識別番号
入院年月日
レセ電コードリストと追加項目リスト
レセ電コード

(3) 抽出条件

MDC6=(各疾病に該当する基本DPC)

(4) 計算値

追加項目、有無=レセ電コードリストに追加対象のコードがあれば「有」、なければ「無」をセット

2) CSVへ保存 ②

7. 5で抽出したデータと6で作成したデータをマージ

1) 5で抽出したデータと6で作成したデータをマージ
 ※①+②を外部プログラムでマージ

3. DPC データ・マイニングの留意点

3. DPCマイニングの留意点

3-1. 欠損値の扱いについて

使用データ……デモ用サンプルデータ

3-2. 病名、手術名など多重回答項目の扱いについて

使用データ……060050(肝疾患)

3-3. ギャップ分析の指標について

使用データ……060050(肝疾患)

3-1. 欠損値の扱いについて

欠損データの扱い その1 ……データ設定機能の利用

- ・カテゴリ項目は項目単位で無効値扱いの指定可能。(データ設定の有効値タブ)
- ・数値項目は入力データに従って自動的に欠損値扱いされるのみ。

欠損データの扱い その2 ……TreeとRBFでの扱いの違い

☆ Tree Case

- ・カテゴリ欠損は空白値とし、1つのカテゴリ値として扱っている。
- ・数値欠損はゼロ値として扱っている。
- logic的に欠損値を特別に処理している訳ではない。
- ・データ設定による異常値設定(バケット限界)も影響しない。

☆ RBF Case

- ・カテゴリ欠損は空白値とし、1つのカテゴリ値として扱っている。
- “不明”を欠損扱いにした場合、グループ間構造構造には影響しないが、グループ特性解釈時のグループ内変数の並び順に影響するので、不明は欠損扱いすべきである。
- ・数値欠損は欠損値として扱っている。Treeのようにゼロ値扱いではない。(Logicは後日河村より)
- ・1レコード中の欠損値項目が説明変数の8割以上あると計算の対象から除外される。
- ・データ設定による異常値設定(バケット限界)した場合、グループ間構造構造にもグループ内変数の並び順にも影響せず、単に当該項目の表示に異常値の度数表示するのみ。

欠損データの扱い その3...DPCデータマイニングでの考慮点

☆Tree Case

- ・カテゴリー欠損はブランク値、数値欠損はゼロ値として扱われるので、欠損値(NGデータ)を多く含む項目がツリーの分岐項目として現れることがある。NGが分岐条件に現れたら、説明変数からはずす等の対処が必要。

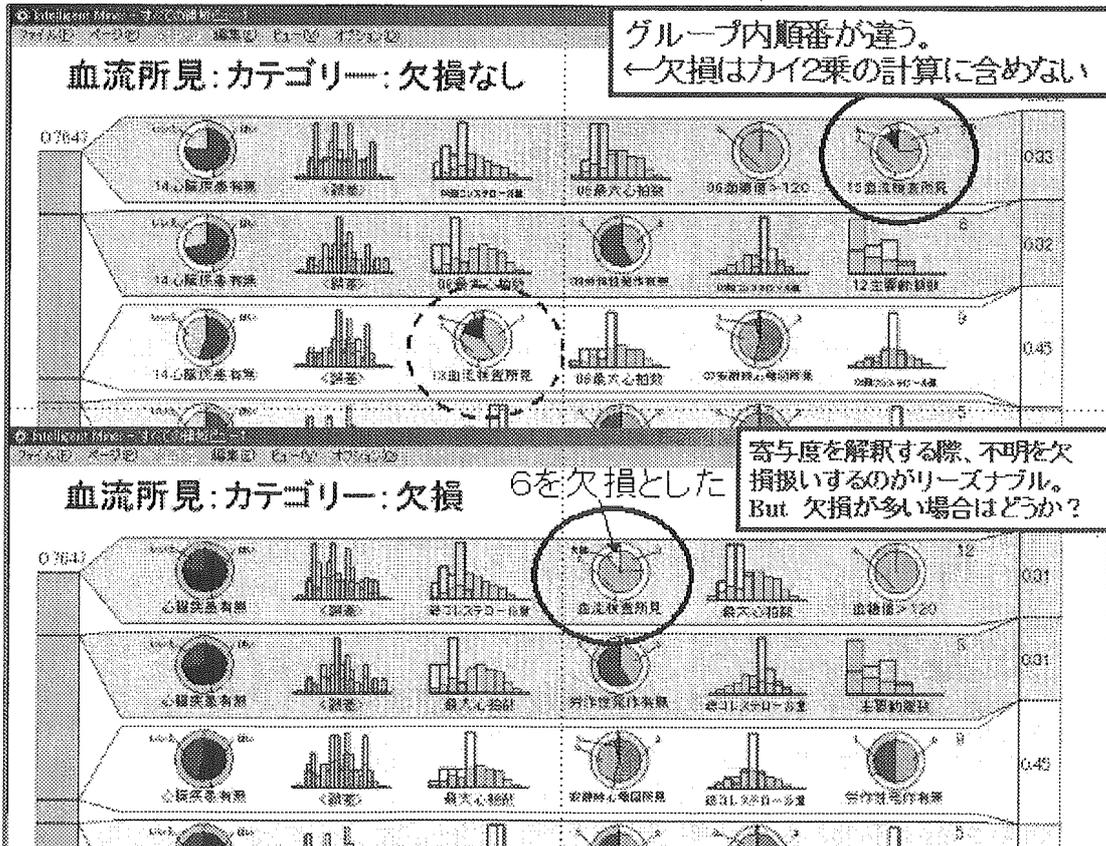
☆RBF Case

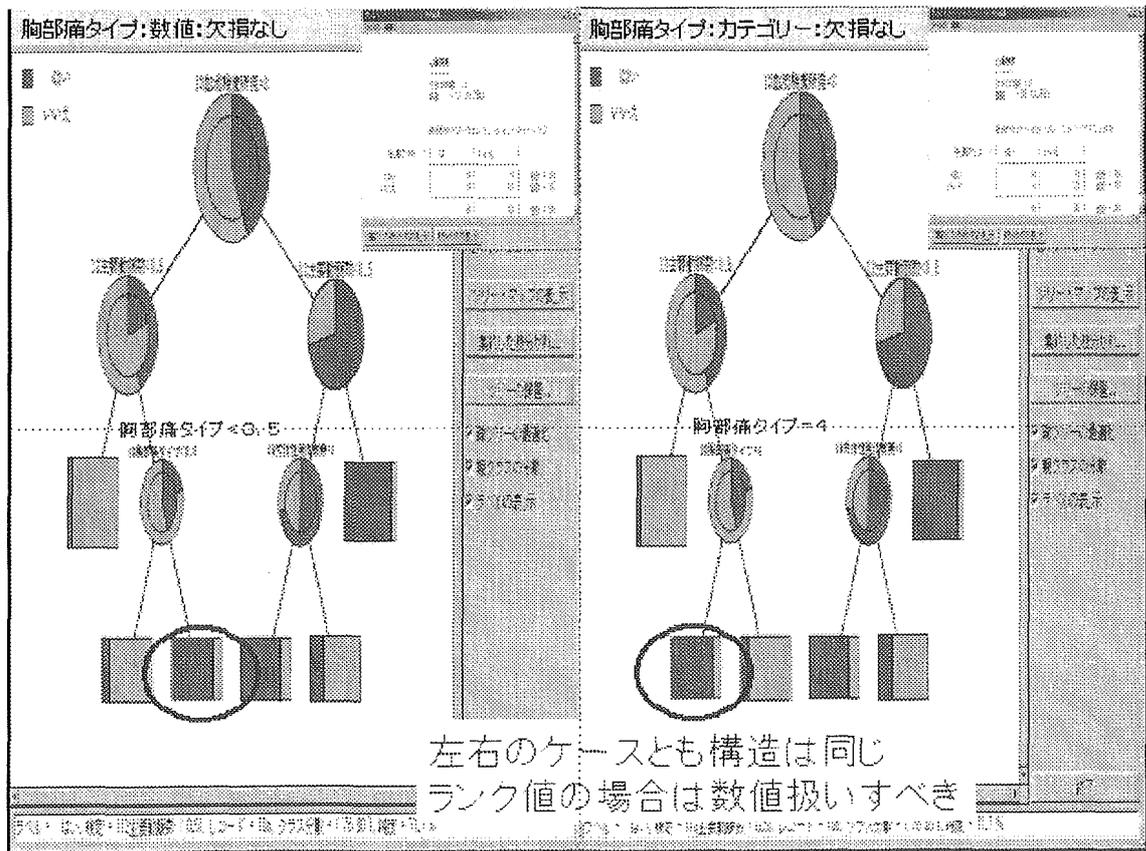
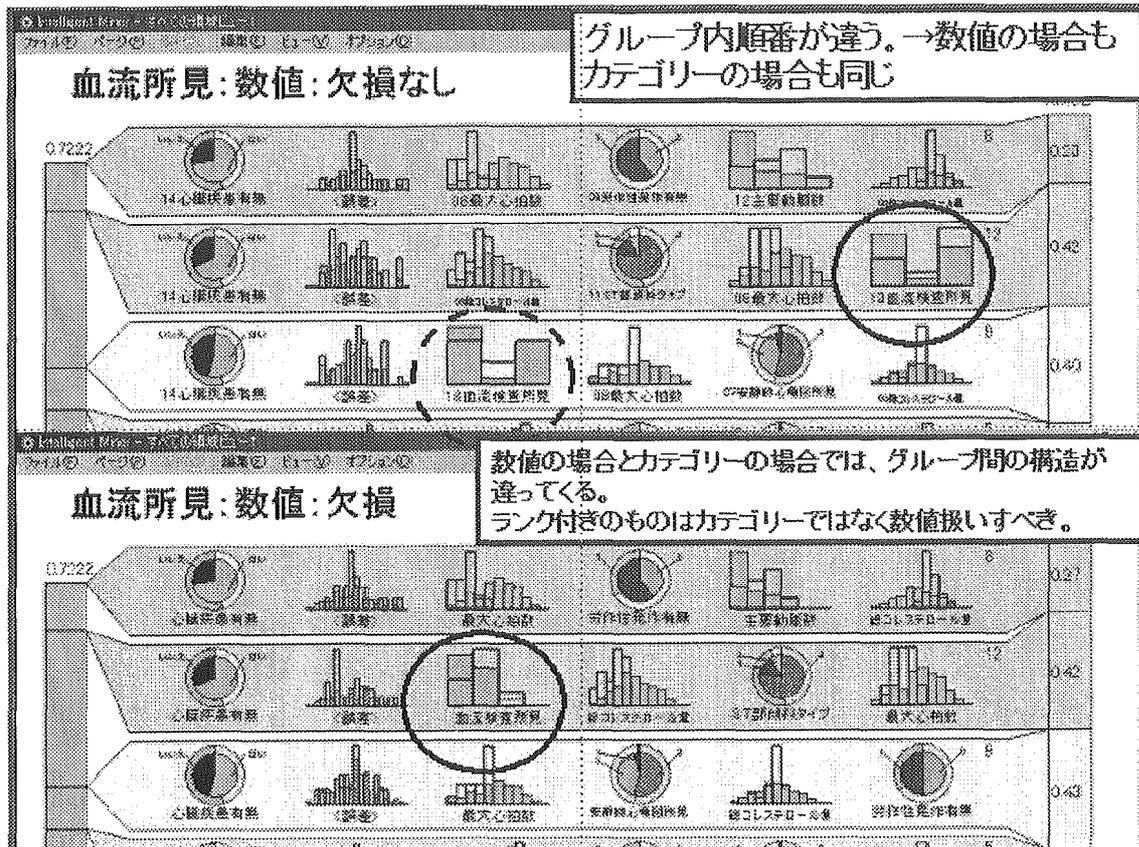
- ・上記と同様に、各グループの寄与度の上位(左から6番目くらいまで)にNGを多く含む項目が現れるようであれば、説明変数からはずす等の対処が必要。

欠損データの扱い その4 ...対処法

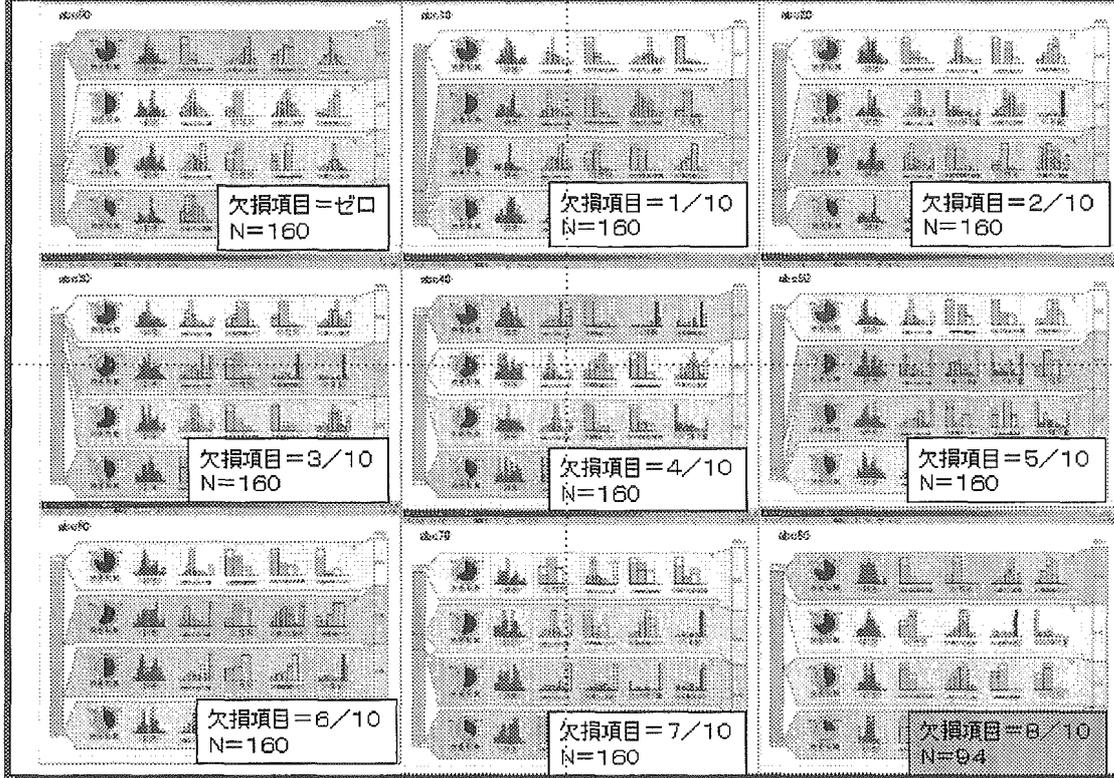
- ・事前に分析の切り口に従って、全項目のデータ充足度(%)一覧を作成しておく。
(・全体 ・MDC 6桁別 ・施設別 ...)
- ・分析切り口でのデータ充足度 ≤ 80% の項目は説明変数には使わないなど。

- ・当該病名診断群分析に関係の無い項目は説明変数から除外する。
(心疾患分析での肝硬変child区分は除外するなど)
- ・肝疾患(O60050)でのTree/RBF分析基本パターンが確定時点で先生方に説明変数候補を教えていただく。





グループ間、グループ内の構造が変化する。→項目ごとの欠損割合をみて許容欠損率設定必要(たとえば20%)



カテゴリデータの無効値化(欠損値)と数値データの異常値化

カテゴリデータの無効値化

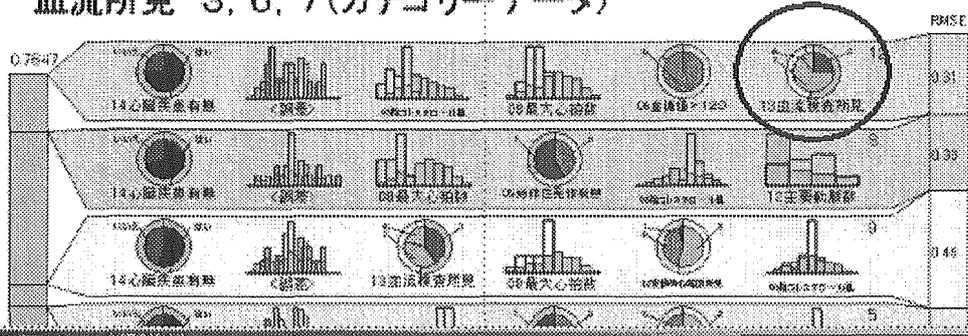
“3”, “6”, “7”
の内、6を無効
(欠損扱い)
としたいケース

**数値データの
有効表示範囲**

10刻みで100~170
100以下、170以上は異常値
として表示したいケース
異常値は表示のみで、計算上
は特別な扱いをしていない。
(カイ2乗も含めて)

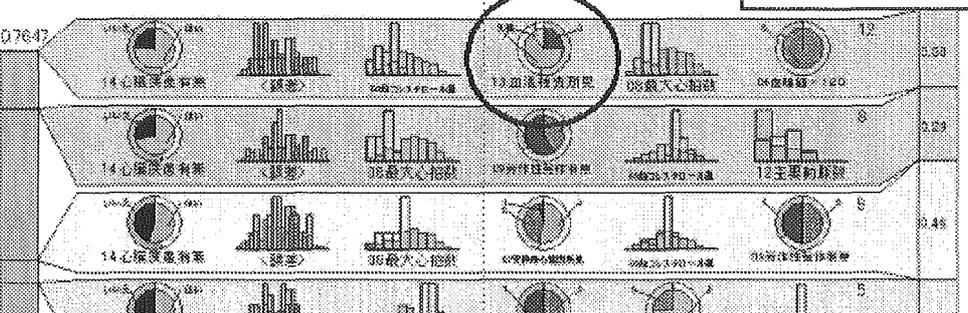
グループ内順番(カイ2乗値)が違っただけ←カテゴリーデータの場合

血流所見 3, 6, 7(カテゴリーデータ)



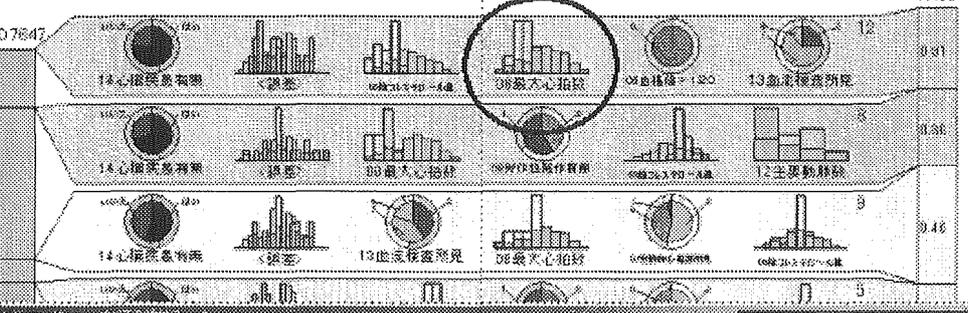
血流所見 3, 6, 7の内6は無効値としたケース

通常
の欠損と
全く同じ扱い



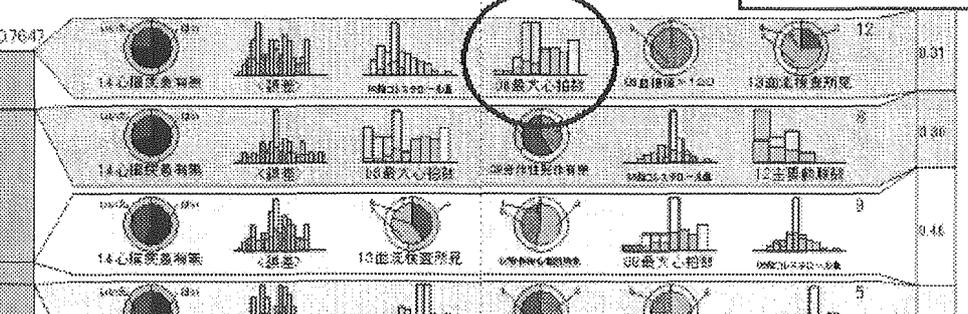
全く同じ(カイ2乗値も)←数値データの場合 表示法がちがうだけ

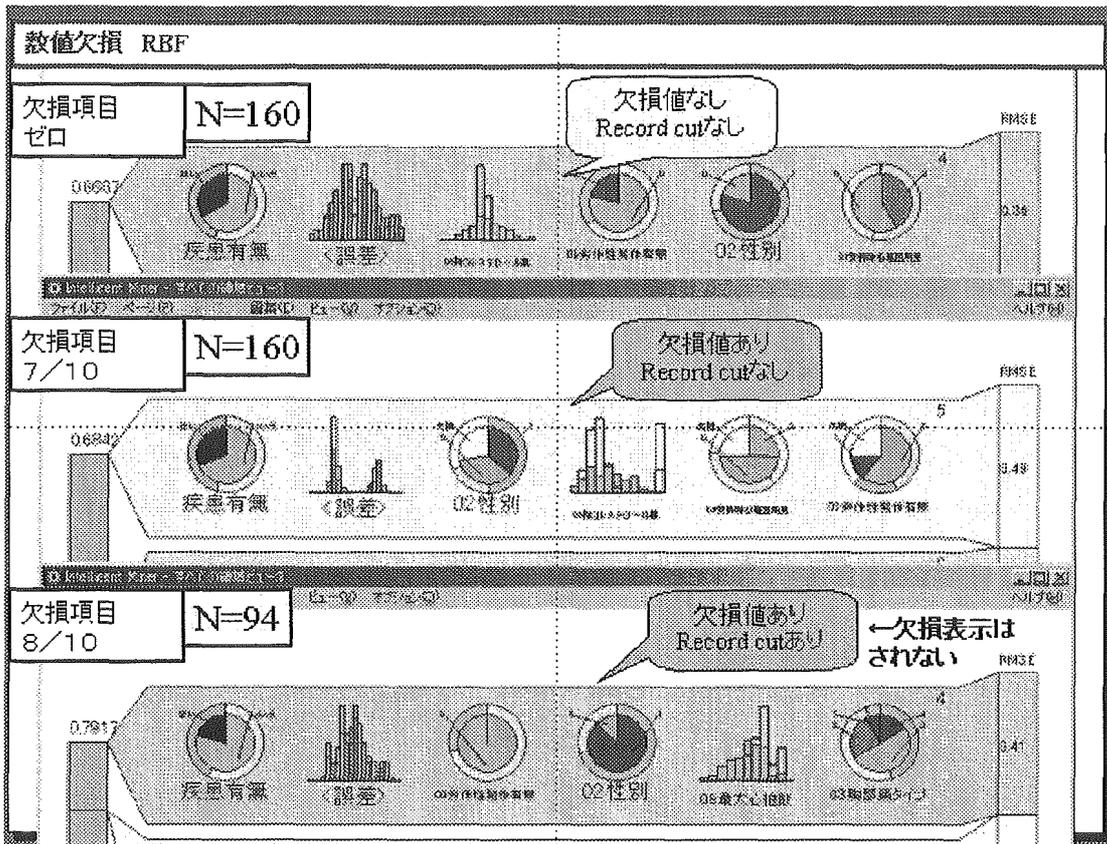
最大心拍数 バスケット(範囲SD 10; -2; 2)



最大心拍数 バスケット(幅Abs 10, 100, 170)

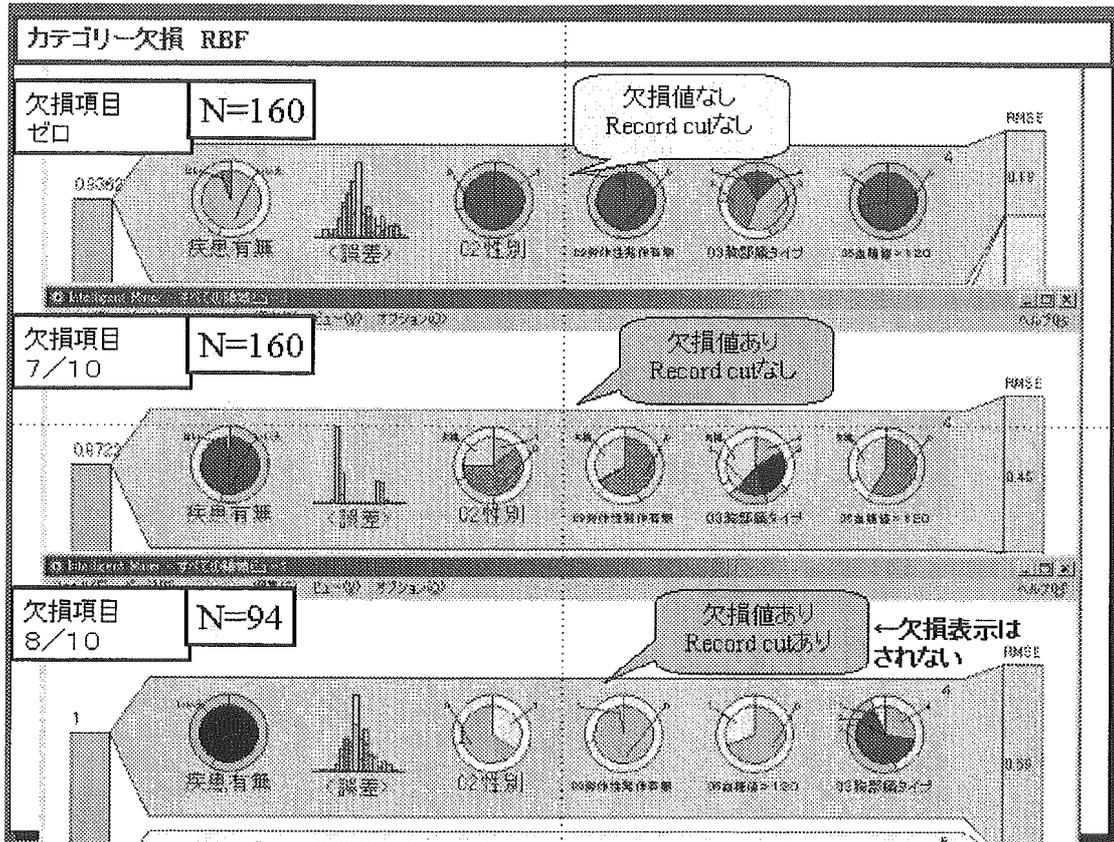
グループ内順番
も変わらず。





数値欠損 RBF

欠損項目	欠損割合	平均欠損率	標準偏差	RMSE	最大誤差	ABS 平均誤差	正規化誤差
ゼロ	0.31	0.68	19.76				
N=160							
欠損項目	欠損割合	平均欠損率	標準偏差	RMSE	最大誤差	ABS 平均誤差	正規化誤差
4	48	39.80	0.666667	0.473405	0.36	0.29	19.87
5	38	29.75	0.605283	0.496754	0.32	0.31	19.27
6	36	19.75	0.469667	0.498888	0.38	0.33	19.65
7	44	27.50	0.431818	0.430379	0.45	0.32	20.01
N=160							
欠損項目	欠損割合	平均欠損率	標準偏差	RMSE	最大誤差	ABS 平均誤差	正規化誤差
6	38	29.75	0.609211	0.48483	0.49	0.48	24.59
6	32	29.80	0.5625	0.496759	0.47	0.48	23.54
7	61	38.13	0.52459	0.499395	0.49	0.49	24.59
4	29	19.19	0.37921	0.480215	0.45	0.44	22.74
N=94							
欠損項目	欠損割合	平均欠損率	標準偏差	RMSE	最大誤差	ABS 平均誤差	正規化誤差
4	24	25.62	0.791667	0.406115	0.41	0.29	20.79
6	16	29.79	0.714286	0.431754	0.51	0.44	28.53
6	21	22.34	0.47619	0.499433	0.59	0.34	17.78
7	21	22.34	0.298995	0.426918	0.52	0.47	26.37



カテゴリー欠損 RBF

平均絶対誤差 0.28
平均最大誤差 1.12
正規化誤差 18.80

欠損項目 ゼロ N=160

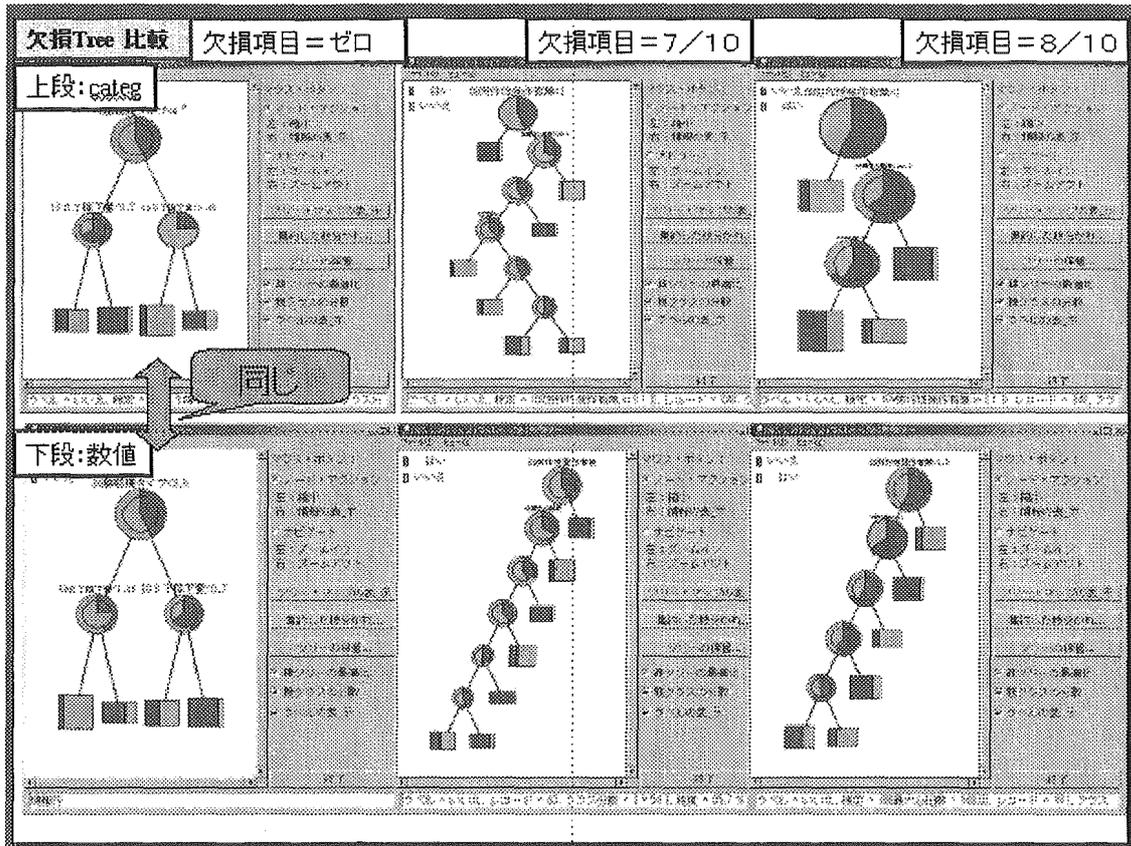
id	種類	相対数	平均	標準偏差	RMSE	最大誤差	ABS 平均誤差	正規化誤差
4	43	29.39	0.93817	0.284439	0.19	0.09	0.17	9.68
6	36	22.50	0.71111	0.497496	0.40	0.08	0.68	29.85
5	45	26.25	0.47619	0.490499	0.52	1.09	0.45	25.54
7	25	21.50	0.5571429	0.221125	0.22	0.09	0.25	15.55

欠損項目 7/10 N=160

id	種類	相対数	平均	標準偏差	RMSE	最大誤差	ABS 平均誤差	正規化誤差
4	38	22.50	0.97222	0.164439	0.45	0.09	0.45	22.74
6	44	27.50	0.68097	0.474918	0.45	0.08	0.44	22.89
5	15	11.88	0.61579	0.492376	0.49	0.08	0.48	24.19
7	61	38.10	0.588235	0.29444	0.48	0.09	0.43	23.85

欠損項目 8/10 N=94

id	種類	相対数	平均	標準偏差	RMSE	最大誤差	ABS 平均誤差	正規化誤差
4	25	30.25	0	0	0.60	1.05	0.42	20.12
5	22	29.75	0.678571	0.447025	0.44	0.08	0.29	22.23
6	14	14.35	0.428571	0.494372	0.51	0.08	0.09	29.89
7	23	24.47	0	0	0.49	0.08	0.23	19.89



欠損Tree 比較 **左側:数値** **右側:categ**

欠損項目 ゼロ

tree#0
欠損の数 = 0
頻度 = 7 回 (21.21%)

tree#1
欠損の数 = 0
頻度 = 7 回 (21.21%)

同じ

総称がワザワザのコンフィュージョン・マトリックス

手廻りクラス	誤り	正しく	合計
誤り	88	84	合計 = 172
正しく	18	121	合計 = 139
	106	205	合計 = 311

総称がワザワザのコンフィュージョン・マトリックス

手廻りクラス	誤り	正しく	合計
誤り	88	84	合計 = 172
正しく	18	121	合計 = 139
	106	205	合計 = 311

欠損項目 7/10

tree#0
欠損の数 = 7
頻度 = 1 回 (3.21%)

tree#1
欠損の数 = 7
頻度 = 1 回 (3.21%)

同じ

総称がワザワザのコンフィュージョン・マトリックス

手廻りクラス	誤り	正しく	合計
誤り	84	80	合計 = 164
正しく	18	116	合計 = 134
	102	196	合計 = 298

総称がワザワザのコンフィュージョン・マトリックス

手廻りクラス	誤り	正しく	合計
誤り	84	80	合計 = 164
正しく	18	116	合計 = 134
	102	196	合計 = 298

欠損項目 8/10

tree#0
欠損の数 = 8
頻度 = 7 回 (22.26%)

tree#1
欠損の数 = 8
頻度 = 7 回 (22.26%)

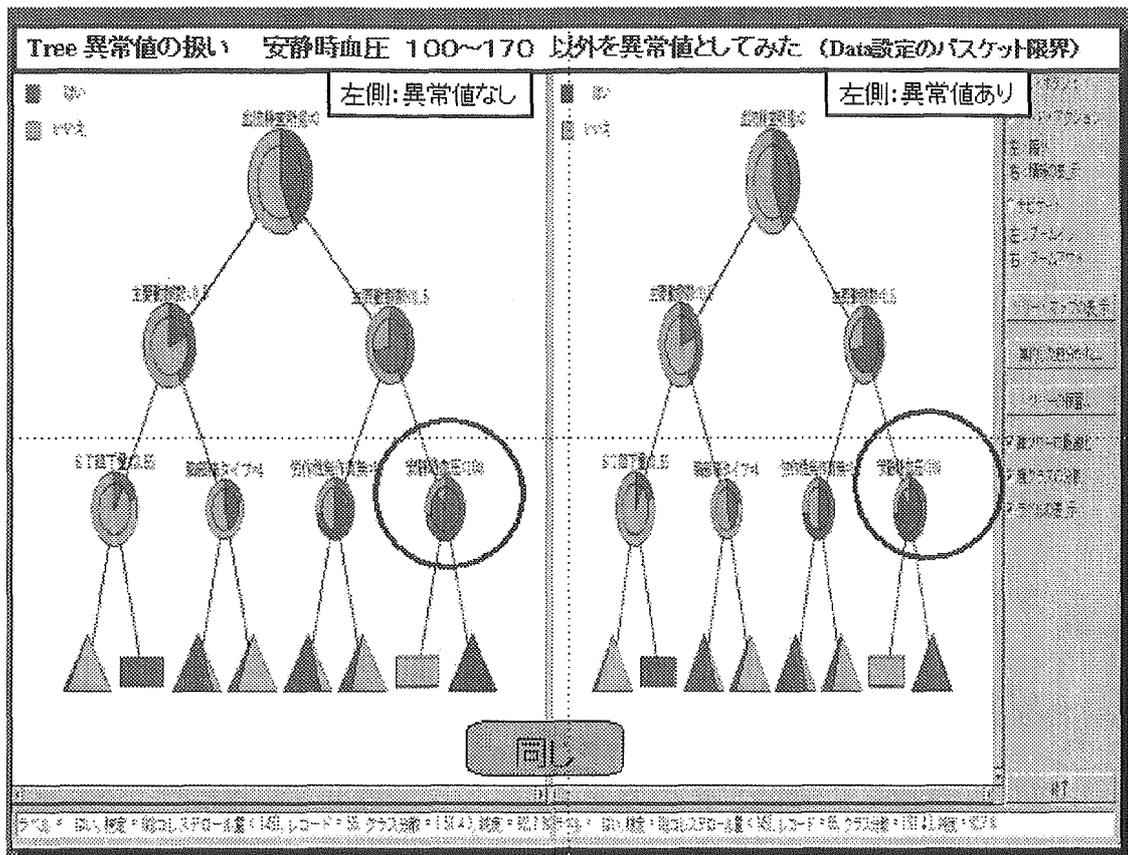
同じ

総称がワザワザのコンフィュージョン・マトリックス

手廻りクラス	誤り	正しく	合計
誤り	80	76	合計 = 156
正しく	18	112	合計 = 130
	98	188	合計 = 286

総称がワザワザのコンフィュージョン・マトリックス

手廻りクラス	誤り	正しく	合計
誤り	80	76	合計 = 156
正しく	18	112	合計 = 130
	98	188	合計 = 286



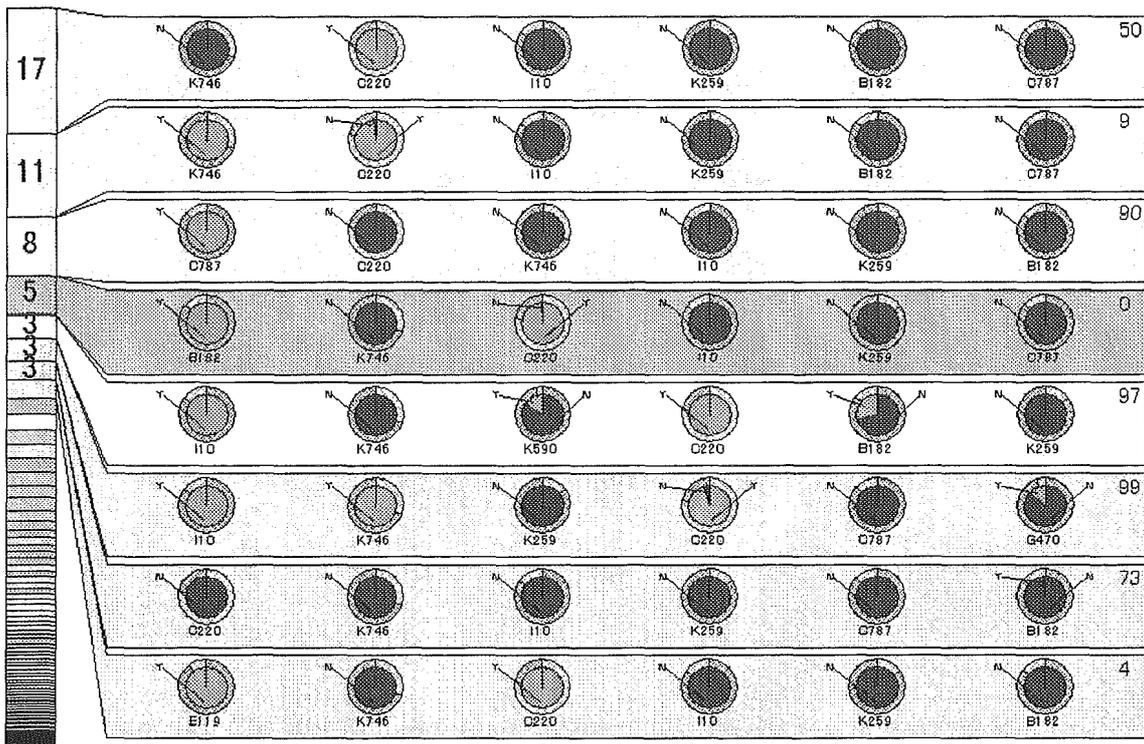
3-2. 病名, 手術名など多重回答項目の扱いについて

- ①入院病名+併存病名+発症病名の7項目に出現するすべての病名のうち、出現頻度の高い下記30病名の有無を新たにデータ項目として追加し、クラスタリングを試みた。

B171 B181 B182 C189 C220 C221 C780 C787 C795 E119 E722 E880 G470 I10 I209
I859 K210 K259 K295 K590 K729 K746 K769 K830 M5456M8199N40 R11 R18 R609

- ②結果、下記のような複数病名を持つ患者群が得られた。
肝細胞癌、肝細胞癌+肝硬変、転移性肝癌、肝細胞癌+C型肝炎、
肝細胞癌+肝硬変+食道静脈瘤……
- ③肝疾患特有の上位10病名の有無フィールドを追加して決定木を作成してみたが、それらを追加してもツリーの構造に変化は無かった。
- ④病名を上位16の複数病名に分類したフィールドを追加して決定木を作成してみたが、それらを追加してもツリーへの影響は微小であった。
- ⑤病名を上位16の複数病名に分類したフィールドを用いてアソシエーション分析を試みた。
結果、肝硬変患者の92%は肝細胞癌であるが、肝細胞癌患者の33%が肝硬変であるなどのアソシエーション・ルール図が得られた。
- ⑥同様に、複数手術のアソシエーションを試みた。
結果、K9202であれば79%の確率で出来高>定額であるなどのルールが得られた。

NN(日 Δ%)



肝細胞癌

NN(日 Δ%) クラスター 50 母集団 16.93%

Code	Value
B171	108
B181	116
B182	283
C787	31
C788	30
C789	308
C790	40
C791	236
C792	87
C793	114
C794	140
C795	334
C796	70
C797	230
C798	74
C799	191
C800	90
C801	100
C802	100
C803	116
C804	90
C805	90
C806	116
C807	116
C808	116
C809	116
C810	116
C811	116
C812	116
C813	116
C814	116
C815	116
C816	116
C817	116
C818	116
C819	116
C820	116
C821	116
C822	116
C823	116
C824	116
C825	116
C826	116
C827	116
C828	116
C829	116
C830	116
C831	116
C832	116
C833	116
C834	116
C835	116
C836	116
C837	116
C838	116
C839	116
C840	116
C841	116
C842	116
C843	116
C844	116
C845	116
C846	116
C847	116
C848	116
C849	116
C850	116
C851	116
C852	116
C853	116
C854	116
C855	116
C856	116
C857	116
C858	116
C859	116
C860	116
C861	116
C862	116
C863	116
C864	116
C865	116
C866	116
C867	116
C868	116
C869	116
C870	116
C871	116
C872	116
C873	116
C874	116
C875	116
C876	116
C877	116
C878	116
C879	116
C880	116
C881	116
C882	116
C883	116
C884	116
C885	116
C886	116
C887	116
C888	116
C889	116
C890	116
C891	116
C892	116
C893	116
C894	116
C895	116
C896	116
C897	116
C898	116
C899	116
C900	116

