

関西SASユーザー会 -SASの便利な使い方について-

SAS Institute Japan Ltd.
Solution Planning Center

東 一成(Kazunari Azuma)
Kazunari.Azuma@sas.com

The Power to Know.

今回のユーザー会の内容

- ◆ 履歴データを用いた顧客ランクの算出
 - ◆ 顧客の簡単な履歴データを用いて、一般的に流通業などで使われている簡単な分析の例を紹介
 - 「デシル分析」
 - 「RFM分析」
 - ◆ 上記の分析をSASでどのように行うのかを紹介する。

デシル分析(decile)とは？

- ◆ 顧客毎に売上金額などを集計し、10に分ける手法。DECILEとは英語で10分位の意味。
- ◆ これにより顧客に10段階のランク付けを行い、ランク別に売上が多い製品を求めたり、本当に「2:8」の法則が当てはまっているかを確認したりする事が可能。
- ◆ 各ランクの顧客の貢献度などを測る。金額だけの単純な手法であるが、簡単に算出が可能。

3

今回のサンプルデータ

顧客番号	買上日	商品区分	金額
1	1991	1991-12-10 雑誌・書籍	500
2	1991	1991-01-05 時計・貴金属	1100
3	1991	1991-02-14 時計・貴金属	2370
4	1991	1991-02-05 航空チケット	7000
5	1991	1991-02-09 洋服	200
6	1991	1991-02-10 航空チケット	1700
7	1991	1991-04-04 時計・貴金属	28400
8	1991	1991-04-09 時計・貴金属	1700
9	1991	1991-04-15 洋服	200
10	1991	1991-05-06 レストラン	7000
11	1991	1991-05-06 時計・貴金属	500
12	1991	1991-05-15 時計・貴金属	18500
13	1991	1991-05-20 自動車	100
14	1991	1991-07-04 時計・貴金属	300
15	1991	1991-07-11 レストラン	600
16	1991	1991-07-14 レストラン	620
17	1991	1991-06-06 レストラン	670
18	1992	1991-12-19 洋服	200
19	1992	1991-12-30 洋服	100
20	1992	1991-02-19 雑誌・書籍	700
21	1992	1991-09-07 ワゴン	1000
22	1992	1991-06-14 洋服	1000
23	1992	1991-07-02 自動車	2700
24	1992	1991-07-20 自動車	500
25	1992	1991-11-11 キャンション	200
26	1992	1991-11-10 自動車	18000
27	1992	1991-12-00 キャンション	500
28	1992	1991-12-05 靴・鞋	4000
29	1992	2000-01-30 レストラン	500
30	1992	1991-01-05 時計・貴金属	1000
31	1992	1991-02-15 雑誌・書籍	8000
32	1992	1991-02-18 時計・書籍	1000

- ◆ 今回は先のようなデータを利用した。
 - 顧客番号 (CUST_ID)
 - 買上日 (DATE)
 - 商品区分 (SYOUHIN)
 - 金額 (PRICE)
- ◆ 上記の変数で50万件のデータを利用してデシル分析を行っていく。

4

デシル分析:ステップ1 データをサマリーする

- 顧客ごとの売上ランクを出すために顧客ごと売上高SUMMARYプロシジャで算出する。

```
/*顧客ごとに売上金額をサマライズする*/
PROC SUMMARY DATA=k_wg.tran NWAY;
    BY cust_id;
    VAR price;
    OUTPUT OUT=dat1(DROP=_TYPE_ _FREQ_) SUM=price;
RUN;
/*BYでは無くCLASSで処理をすれば事前のソート処理が不要。
ただしCUST_IDはメモリ上に記憶されるので複数組み合わせると、メモリ不足になる場合があります。*/
```

DAT1

顧客番号	金額
1	540
2	81570
3	20685
4	10270
5	30250
6	10280
7	25680
8	10520
9	10985
10	10980
11	85430
12	40080
13	11080
14	10130
15	11180
16	21480
17	10080
18	30570
19	10220
20	10680

デシル分析:ステップ2 顧客ごとのランクを算出する

- 顧客ランクをRANKプロシジャで算出し、売上が高い順に9~0のランクをつける。ランクを格納する「RANK」という変数を追加している。

```
/*顧客を金額が多い順に10分割する*/
PROC RANK DATA=dat1 GROUP=10 OUT=dat2;
    RANKS rank;
    VAR price;
RUN;
/*「GROUP=10」の部分を「PERCENT」に置き
なおすと、順位を百分率で表します*/
```

DAT2

顧客番号	金額	Rank for the data
1	1008	9
2	1007	8
3	1002	7
4	1005	6
5	1004	5
6	1008	4
7	1006	3
8	1007	2
9	1008	1
10	1008	0
11	1010	9
12	1011	8
13	1012	7
14	1013	6
15	1014	5
16	1016	4
17	1018	3
18	1017	2
19	1018	1
20	1018	0

デシル分析:ステップ3

ランク別に売上金額の統計量を算出する

```
/*RANK毎の売上金額の統計量算出*/
PROC MEANS DATA=dat2 N SUM MEAN MAX MIN MAXDEC=2;
  CLASS rank;
  VAR price;
RUN;
```

Variable	price	オブザーベーション数	合計	平均値	最大値	最小値
0		1741	18590563.10	10678.09	11591.00	350.00
1		1742	21892418.30	12567.40	13573.50	11592.00
2		1742	25542256.90	14662.60	15840.00	13574.00
3		1742	30007182.70	17225.71	18740.00	15845.00
4		1742	35528189.50	20395.06	22253.50	18746.00
5		1742	42638689.50	24476.86	26948.00	22255.10
6		1742	52335108.80	30043.12	33620.10	26950.00
7		1742	67711786.90	38870.14	45104.00	33639.00
8		1742	97642285.30	56051.83	71670.00	45110.00
9		1742	295833554.60	169824.08	6278049.30	71676.00

7

デシル分析:ステップ4-1

『2:8』の法則をグラフで確認する

◆ データ加工

```
/*RANK別で売上金額を集計し、高い順にソートする*/
PROC SUMMARY DATA=dat2 NWAY;
  CLASS rank;
  VAR price;
  OUTPUT OUT=dat3(DROP=_FREQ_ _TYPE_)
  SUM=price;
RUN;
PROC SORT DATA=dat3;
  BY DESCENDING price;
RUN;QUIT;
/*ランク別の売上金額を累計する*/
DATA dat4;
  SET dat3;
  RETAIN;
  id=_n_;
  amount+price;
  FORMAT amount 12.0;
RUN;
```

ランク別の売上集計と売上累計

Rank for variable price	変数	id	amount	
1	0	295833554	1	295833554
2	0	97642285	2	30049844
3	0	67711786	3	43161021
4	0	52335108	4	51802279
5	0	42638689	5	58066168
6	0	35528189	6	65889878
7	0	30007182	7	62189671
8	0	25542256	8	64733884
9	0	21892418	9	66931413
10	0	18590563	10	68722086

8

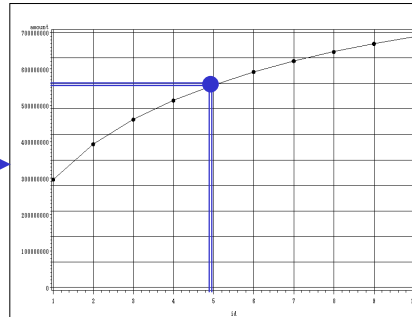
デシル分析:ステップ4-2

『2:8』の法則をグラフで確認する

- ◆ SAS/GRAPHのGCHARTプロシジャでグラフ作成

```

/* 『2:8の法則』をグラフで確認 */
OPTIONS DEVICE=GIF;
SYMBOL I=JOIN V=DOT;
PROC GPLOT DATA=dat4;
    PLOT amount * id/
    VZERO /*軸を0からスタートさせるために必要*/
    HREF=1 to 10 by 1
    VREF=0 to 700000000 by 70000000;
RUN;
QUIT;
    
```



今回のデモデータでは「2:8の法則」ではなく、「5:8の法則」だった。

デシル分析:ステップ4-3

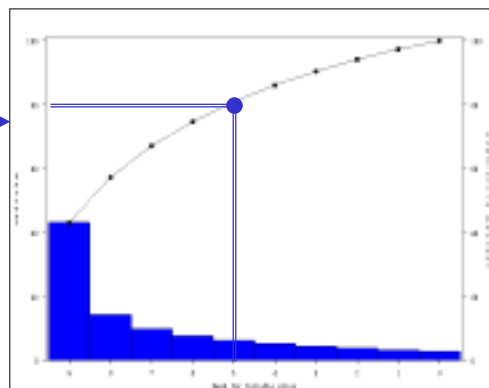
『2:8』の法則をグラフで確認する

- ◆ SAS/QCのPARETOプロシジャでグラフ作成

```

PROC PARETO DATA=dat3;
    VBAR rank /
    ANCHOR=TC
    WEIGHT=price
    SCALE = PERCENT
    CBARS =BLUE;
RUN;
    
```

製造業などでよく利用されているが、パレート図を用いると、データ加工処理もそれほど必要なく、ランク別の売上に対する貢献度をグラフで表示する事が出来る。

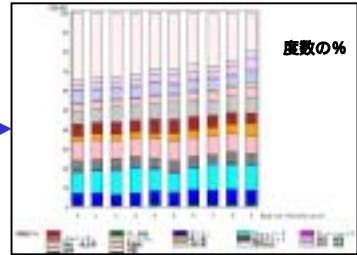
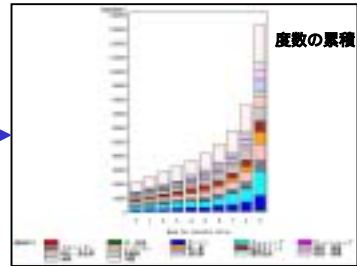


デシル分析:ステップ5-3 ランクと商品の関係を見ている

- ◆ ランクと商品の関係をグラフで確認する。

```

OPTIONS DEVICE=GIF;
/*パターン1*/
PROC GCHART DATA=dat5 ;
    VBAR rank / SUBGROUP=syohuin DISCRETE;
RUN;
/*パターン2*/
AXIS1 VALUE=none LABEL=none;
PROC GCHART DATA=dat5;
    VBAR rank / TYPE=pct SUBGROUP=syohuin G100
    GROUP=rank NOZERO MAXIS=axis1;
RUN;QUIT;
    
```



デシル分析:ステップ5-4 ランクと商品の関係を見ている

- ◆ ランクと商品の関係を度数集計で見ている

```

/*ランクと商品で度数集計*/
PROC SORT DATA=dat6;
    BY rank DESCENDING price;
RUN;
PROC FREQ DATA=dat6 order=freq;
    BY rank;
    WEIGHT freq;
    TABLES syohuin;
RUN;
    
```

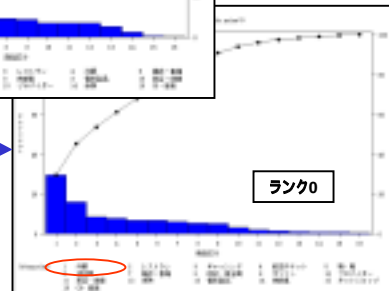
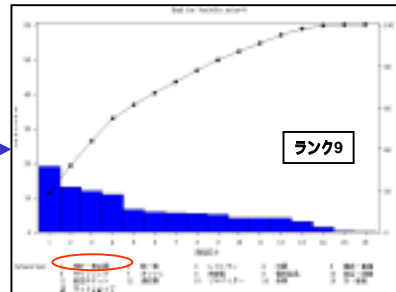
ランク9 ← 比較 !! → ランク0



デシル分析:ステップ5-5 ランクと商品の関係を見定める

- ◆ ランク別に売上金額に貢献している商品をパレート図で見つける。

```
PROC PARETO DATA=dat6;
  BY rank;
  VBAR syouhin /
  ANCHOR=TC
  WEIGHT=price
  SCALE = PERCENT
  CBARS =BLUE;
RUN;
```



15

デシル分析:まとめ

- ◆ 今回は、SASデータステップやSASプロシジャを利用して顧客を10分割し、各グループがどの程度売上に貢献しているかを分かるようにした。
- ◆ この場合、トップの2~3グループに対する依存度が高いが、このような「2:8の法則」はパレート原理とも呼ばれている。
- ◆ 顧客の属性データ等を利用すれば、各グループの特徴を知ることが可能になる。
- ◆ 今回は品質管理のソフトウェアであるSAS/QCに含まれるPARETOプロシジャを利用したが、製造業における品質管理だけではなく、営業活動やマーケティング活動の管理を行う事もできる。

16

RFM分析とは？

- ◆ RFMによる顧客ランキング
 - 流通業やサービス業などで良く利用される分析手法。
 - 優良顧客を定義するために履歴データの以下の指標を利用する
 - ◆ R(Recency)・・・最終利用日
 - ◆ F(Frequency)・・・頻度
 - ◆ M(Monetary)・・・利用金額
 - 上記の指標に対して、各企業で定めた指数を掛けて、ポイントを算出する
 - ◆ EX.2週間前に購入は5ポイント、10回～20回は10ポイント、購入金額が10万円を超えると10ポイント。5ポイント+10ポイント+10ポイント=25ポイント
 - このようなRFMを使う事により、金額だけではなく他の指標を用いて、顧客の分類を行える点でデシル分析より柔軟でもある。

RFM分析を工夫する

- ◆ RFM分析は一般に広く知られている

- Recency	直近の購買年月日	2000/9/1
- Frequency	購買回数	20回
- Monetary	購買金額	10万円



- ◆ 優良顧客は誰？
 - ◆ 顧客ランク = $\sqrt{Rm/R * F/Fm * M/Mm}$
- あるターゲットの日から「何日前」にと考える

分析者が重み付け！！

今回のサンプルデータ

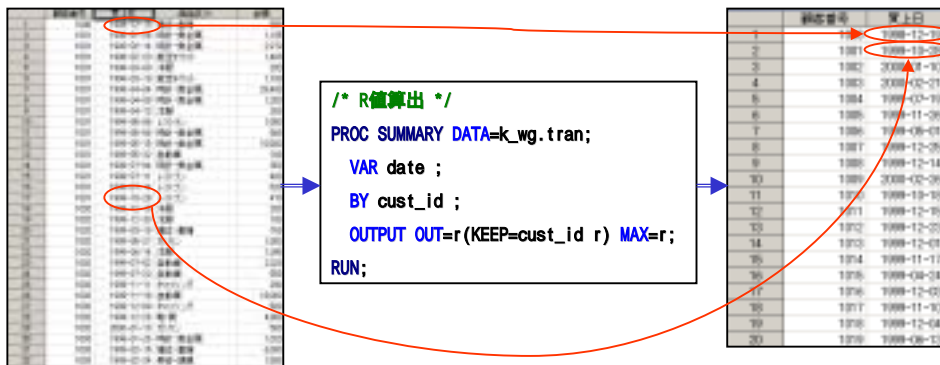
顧客番号	買上日	商品区分	金額	
1	1991	1991-12-19	雑誌・書籍	500
2	1991	1991-01-05	時計・金金庫	1,100
3	1991	1991-02-14	時計・金金庫	2,500
4	1991	1991-02-03	靴・シューズ	1,400
5	1991	1991-05-09	洋服	200
6	1991	1991-03-10	靴・シューズ	1,100
7	1991	1991-04-04	時計・金金庫	24,400
8	1991	1991-04-09	時計・金金庫	1,200
9	1991	1991-04-12	洋服	200
10	1991	1991-05-06	化粧品	1,000
11	1991	1991-05-06	時計・金金庫	500
12	1991	1991-05-15	時計・金金庫	18,000
13	1991	1991-05-22	靴・シューズ	1,100
14	1991	1991-07-04	時計・金金庫	300
15	1991	1991-07-11	化粧品	400
16	1991	1991-07-14	化粧品	820
17	1991	1991-10-06	化粧品	400
18	1992	1991-12-19	洋服	200
19	1992	1991-12-01	洋服	1,000
20	1992	1991-02-19	雑誌・書籍	1,000
21	1992	1991-08-27	化粧品	1,000
22	1992	1991-06-14	洋服	1,000
23	1992	1991-07-03	靴・シューズ	2,500
24	1992	1991-07-20	靴・シューズ	500
25	1992	1991-11-11	靴・シューズ	200
26	1992	1991-11-19	靴・シューズ	18,000
27	1992	1991-12-30	靴・シューズ	500
28	1992	1991-12-05	靴・シューズ	4,000
29	1992	2000-07-30	靴・シューズ	500
30	1992	1991-07-08	時計・金金庫	1,000
31	1992	1991-02-15	雑誌・書籍	8,000
32	1992	1991-02-04	雑誌・書籍	1,000

- ◆ 今回は先のようなデータを利用した。
 - 顧客番号(CUST_ID)
 - 買上日(DATE)
 - 商品区分(SYOUHIN)
 - 金額(PRICE)
- ◆ 上記の変数で50万件のデータを利用してRFM分析を行っていく。

19

RFM分析:ステップ1 「R」の算出

- ◆ 履歴データからR値を算出する。日付の列で最も大きい値のものを出力するようにする。



20

RFM分析:ステップ2 「F」の算出

- 履歴データから、「F」値を算出する。今回は出現頻度をそのまま利用しているが、本来であれば同一日であれば何回買い物をして1回とみなした方がよいと思われる。その場合はSORTプロシジャ等を利用して重複行を削除する必要などがある。

顧客番号	購入年月	購入金額	購入回数
1001	1999-01-18	1000	1
1001	1999-02-18	1000	2
1001	1999-03-18	1000	3
1001	1999-04-18	1000	4
1001	1999-05-18	1000	5
1001	1999-06-18	1000	6
1001	1999-07-18	1000	7
1001	1999-08-18	1000	8
1001	1999-09-18	1000	9
1001	1999-10-18	1000	10
1001	1999-11-18	1000	11
1001	1999-12-18	1000	12
1001	2000-01-18	1000	13
1001	2000-02-18	1000	14
1001	2000-03-18	1000	15
1001	2000-04-18	1000	16
1001	2000-05-18	1000	17
1001	2000-06-18	1000	18
1001	2000-07-18	1000	19
1001	2000-08-18	1000	20
1001	2000-09-18	1000	21
1001	2000-10-18	1000	22
1001	2000-11-18	1000	23
1001	2000-12-18	1000	24
1001	2001-01-18	1000	25
1001	2001-02-18	1000	26
1001	2001-03-18	1000	27
1001	2001-04-18	1000	28
1001	2001-05-18	1000	29
1001	2001-06-18	1000	30
1001	2001-07-18	1000	31
1001	2001-08-18	1000	32
1001	2001-09-18	1000	33
1001	2001-10-18	1000	34
1001	2001-11-18	1000	35
1001	2001-12-18	1000	36
1001	2002-01-18	1000	37
1001	2002-02-18	1000	38
1001	2002-03-18	1000	39
1001	2002-04-18	1000	40
1001	2002-05-18	1000	41
1001	2002-06-18	1000	42
1001	2002-07-18	1000	43
1001	2002-08-18	1000	44
1001	2002-09-18	1000	45
1001	2002-10-18	1000	46
1001	2002-11-18	1000	47
1001	2002-12-18	1000	48
1001	2003-01-18	1000	49
1001	2003-02-18	1000	50
1001	2003-03-18	1000	51
1001	2003-04-18	1000	52
1001	2003-05-18	1000	53
1001	2003-06-18	1000	54
1001	2003-07-18	1000	55
1001	2003-08-18	1000	56
1001	2003-09-18	1000	57
1001	2003-10-18	1000	58
1001	2003-11-18	1000	59
1001	2003-12-18	1000	60
1001	2004-01-18	1000	61
1001	2004-02-18	1000	62
1001	2004-03-18	1000	63
1001	2004-04-18	1000	64
1001	2004-05-18	1000	65
1001	2004-06-18	1000	66
1001	2004-07-18	1000	67
1001	2004-08-18	1000	68
1001	2004-09-18	1000	69
1001	2004-10-18	1000	70
1001	2004-11-18	1000	71
1001	2004-12-18	1000	72
1001	2005-01-18	1000	73
1001	2005-02-18	1000	74
1001	2005-03-18	1000	75
1001	2005-04-18	1000	76
1001	2005-05-18	1000	77
1001	2005-06-18	1000	78
1001	2005-07-18	1000	79
1001	2005-08-18	1000	80
1001	2005-09-18	1000	81
1001	2005-10-18	1000	82
1001	2005-11-18	1000	83
1001	2005-12-18	1000	84
1001	2006-01-18	1000	85
1001	2006-02-18	1000	86
1001	2006-03-18	1000	87
1001	2006-04-18	1000	88
1001	2006-05-18	1000	89
1001	2006-06-18	1000	90
1001	2006-07-18	1000	91
1001	2006-08-18	1000	92
1001	2006-09-18	1000	93
1001	2006-10-18	1000	94
1001	2006-11-18	1000	95
1001	2006-12-18	1000	96
1001	2007-01-18	1000	97
1001	2007-02-18	1000	98
1001	2007-03-18	1000	99
1001	2007-04-18	1000	100

```

/* F値算出 */
PROC SUMMARY DATA=k_wg.tran;
  BY cust_id ;
  OUTPUT OUT=f(KEEP=cust_id_freq_
                RENAME=(freq=f_value));
RUN;

```

顧客番号	freq
1	1000
2	1001
3	1002
4	1003
5	1004
6	1005
7	1006
8	1007
9	1008
10	1009
11	1010
12	1011
13	1012
14	1013
15	1014
16	1015
17	1016
18	1017
19	1018
20	1019

RFM分析:ステップ3 「M」の算出

- 履歴データから「M」値を算出する。顧客別に利用金額を集計している。本来であれば「一定期間内」「6ヶ月以内」等の制限がつくと思われる。

顧客番号	購入年月	購入金額	購入回数
1001	1999-01-18	1000	1
1001	1999-02-18	1000	2
1001	1999-03-18	1000	3
1001	1999-04-18	1000	4
1001	1999-05-18	1000	5
1001	1999-06-18	1000	6
1001	1999-07-18	1000	7
1001	1999-08-18	1000	8
1001	1999-09-18	1000	9
1001	1999-10-18	1000	10
1001	1999-11-18	1000	11
1001	1999-12-18	1000	12
1001	2000-01-18	1000	13
1001	2000-02-18	1000	14
1001	2000-03-18	1000	15
1001	2000-04-18	1000	16
1001	2000-05-18	1000	17
1001	2000-06-18	1000	18
1001	2000-07-18	1000	19
1001	2000-08-18	1000	20
1001	2000-09-18	1000	21
1001	2000-10-18	1000	22
1001	2000-11-18	1000	23
1001	2000-12-18	1000	24
1001	2001-01-18	1000	25
1001	2001-02-18	1000	26
1001	2001-03-18	1000	27
1001	2001-04-18	1000	28
1001	2001-05-18	1000	29
1001	2001-06-18	1000	30
1001	2001-07-18	1000	31
1001	2001-08-18	1000	32
1001	2001-09-18	1000	33
1001	2001-10-18	1000	34
1001	2001-11-18	1000	35
1001	2001-12-18	1000	36
1001	2002-01-18	1000	37
1001	2002-02-18	1000	38
1001	2002-03-18	1000	39
1001	2002-04-18	1000	40
1001	2002-05-18	1000	41
1001	2002-06-18	1000	42
1001	2002-07-18	1000	43
1001	2002-08-18	1000	44
1001	2002-09-18	1000	45
1001	2002-10-18	1000	46
1001	2002-11-18	1000	47
1001	2002-12-18	1000	48
1001	2003-01-18	1000	49
1001	2003-02-18	1000	50
1001	2003-03-18	1000	51
1001	2003-04-18	1000	52
1001	2003-05-18	1000	53
1001	2003-06-18	1000	54
1001	2003-07-18	1000	55
1001	2003-08-18	1000	56
1001	2003-09-18	1000	57
1001	2003-10-18	1000	58
1001	2003-11-18	1000	59
1001	2003-12-18	1000	60
1001	2004-01-18	1000	61
1001	2004-02-18	1000	62
1001	2004-03-18	1000	63
1001	2004-04-18	1000	64
1001	2004-05-18	1000	65
1001	2004-06-18	1000	66
1001	2004-07-18	1000	67
1001	2004-08-18	1000	68
1001	2004-09-18	1000	69
1001	2004-10-18	1000	70
1001	2004-11-18	1000	71
1001	2004-12-18	1000	72
1001	2005-01-18	1000	73
1001	2005-02-18	1000	74
1001	2005-03-18	1000	75
1001	2005-04-18	1000	76
1001	2005-05-18	1000	77
1001	2005-06-18	1000	78
1001	2005-07-18	1000	79
1001	2005-08-18	1000	80
1001	2005-09-18	1000	81
1001	2005-10-18	1000	82
1001	2005-11-18	1000	83
1001	2005-12-18	1000	84
1001	2006-01-18	1000	85
1001	2006-02-18	1000	86
1001	2006-03-18	1000	87
1001	2006-04-18	1000	88
1001	2006-05-18	1000	89
1001	2006-06-18	1000	90
1001	2006-07-18	1000	91
1001	2006-08-18	1000	92
1001	2006-09-18	1000	93
1001	2006-10-18	1000	94
1001	2006-11-18	1000	95
1001	2006-12-18	1000	96
1001	2007-01-18	1000	97
1001	2007-02-18	1000	98
1001	2007-03-18	1000	99
1001	2007-04-18	1000	100

```

/* M値算出 */
PROC SUMMARY DATA=k_wg.tran;
  VAR price ;
  BY cust_id ;
  OUTPUT OUT=m(KEEP=cust_id m_value)
  SUM=m_value ;
RUN;

```

顧客番号	金額
1	1000
2	1001
3	1002
4	1003
5	1004
6	1005
7	1006
8	1007
9	1008
10	1009
11	1010
12	1011
13	1012
14	1013
15	1014
16	1015
17	1016
18	1017
19	1018
20	1019

RFM分析:ステップ4

「R」「F」「M」の値の統合

顧客番号	購入日
1	1800 1999-12-19
2	1801 1999-10-28
3	1802 2000-01-19
4	1803 2000-03-21
5	1804 1999-01-19
6	1805 1999-11-28
7	1806 1999-08-08
8	1807 1999-12-25
9	1808 1999-12-14

顧客番号	購入数
1	1008
2	1007
3	1002
4	1003
5	1004
6	1006
7	1008
8	1007
9	1008

顧客番号	金額
1	1800
2	1801
3	1802
4	1800
5	1804
6	1805
7	1806
8	1807
9	1808

```

/* RFM値算出 */
DATA rfm ;
MERGE r f m ;
/* 全変数のラベル削除 */
ATTRIB _all_ LABEL="";
BY cust_id;
r_value=INTCK('DAY',r,'28FEB00'd);
DROP r;
RUN;
    
```

顧客番号	購入日	購入数	金額
1	1999-12-19	8	1800
2	1999-10-28	7	1801
3	2000-01-19	2	1802
4	2000-03-21	3	1800
5	1999-01-19	4	1804
6	1999-11-28	6	1805
7	1999-08-08	8	1806
8	1999-12-25	7	1807
9	1999-12-14	8	1808

◆ 「R」の値には最終購入日が入っているが、「XXから何日前？」という基準で見るために、INTCK関数を使い、ある日付から何日前かを算出している。この場合は「F」「M」と違い、値が小さい方が価値があることに注意が必要。

RFM分析:ステップ5

「R」「F」「M」の中央値を算出しマクロ変数に格納

- ◆ 中央値を算出するのにUNIVARIATEプロシジャを利用している。V8からはMEANSプロシジャでも算出可能。

顧客番号	購入日	購入数	金額
1	1999-12-19	8	1800
2	1999-10-28	7	1801
3	2000-01-19	2	1802
4	2000-03-21	3	1800
5	1999-01-19	4	1804
6	1999-11-28	6	1805
7	1999-08-08	8	1806
8	1999-12-25	7	1807
9	1999-12-14	8	1808

```

/* R値の中央値算出 */
/* F値の中央値算出 */
/* M値の中央値算出 */
PROC UNIVARIATE DATA=rfm NOPRINT;
VAR r_value f_value m_value ;
OUTPUT OUT=tmp MEDIAN=m_r m_f m_m;
RUN;
QUIT;
    
```

顧客番号	m_r	m_f	m_m
1	8.0	19	2228.0

```

/* R値中央値マクロに格納 */
/* F値中央値マクロに格納 */
/* M値中央値マクロに格納 */
DATA _NULL_ ;
SET tmp ;
CALL SYMPUT('m_r',put(m_r,best.)) ;
CALL SYMPUT('m_f',put(m_f,best.)) ;
CALL SYMPUT('m_m',put(m_m,best.)) ;
RUN;
    
```

- ◆ マクロ変数に中央値に格納する事により、データセットとマージする必要も無く、この後の処理ですぐに活用できる。

RFM分析:ステップ6

「R」「F」「M」の値より顧客ランクを算出

- 最終的に算出する顧客ランクは大きい値の方が良いので、「R」の使い方
に注意が必要。また重視したい指標に重みを加えても良い。

顧客番号	R値	F値	M値	RFMポイント
1	1000	1	54	4.91
2	1001	14	5405	11.2
3	1002	12	2000	4.4
4	1003	14	1400	5.4
5	1004	13	3000	5.4
6	1005	11	1000	3.4
7	1007	15	1800	5.5
8	1008	13	1900	5.4
9	1009	23	1900	6.4
10	1010	15	5500	6.4
11	1011	8	4000	4.4
12	1012	12	1300	4.4
13	1013	26	1300	5.4
14	1014	26	1300	5.4
15	1015	11	11150	6.3
16	1016	7	2400	3.4
17	1017	13	1500	4.4
18	1018	13	3000	4.4
19	1019	12	1000	4.4
20	1020	8	1500	3.4

```

/* R値、F値、M値、RFMポイントのデータセット完成 */
DATA rfm_point ;
SET rfm ;
rfm_value=(&m/r_value * f_value/&m_f * m_value/&m_m)**(1/3) ;
LABEL   cust_id='顧客番号'  r_value='R値'
        f_value='F値'      m_value='M値'
        rfm_value='RFMポイント';
FORMAT rfm_value 8.2;
RUN ;

```

r_rank	f_rank	m_rank	rfm_rank
1	90	19	22250

顧客番号	R値	F値	M値	RFMポイント
92	1091	27	49948	32.039
93	1092	23	19776	18.819
94	1093	23	12138	16.827
95	1094	7	18008	4.7187
96	1095	11	19018	5.0008
97	1096	95	125143	199.871
98	1097	9	18258	5.5111
99	1098	40	15188	10.7168

RFM分析:ステップ7

「R」「F」「M」の値に主観的な指標をつける

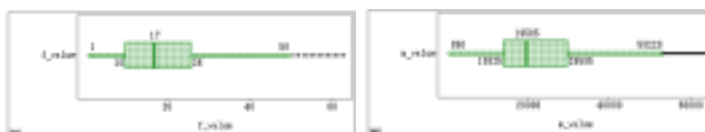
- SASのフォーマットの機能を利用して、「F」と「M」に
対して、ランク付けを行う。

```

/*良く行われるランク付け*/
PROC FORMAT;
VALUE f_rank LOW=10='F_RANK1' 11-18='F_RANK2' 19-30='F_RANK3' 31-60='F_RANK4' 61-HIGH='F_RANK5';
VALUE m_rank LOW=14487='M_RANK1' 14488-21637='M_RANK2' 21638-36260='M_RANK3' 36261-68905='M_RANK4'
68906-HIGH='M_RANK5';
RUN;
QUIT;

```

ポイントの切り口は、下記のような分布のグラフを書いて決定していく事も考えられる



RFM分析:ステップ8-1

フォーマットを利用して「F」「M」のランクの分布を見る

```
/*ランク付けのクロス集計表*/
PROC FREQ DATA=rfm_point;
  FORMAT f_value f_rank. m_value m_rank.;
  TABLES f_value * m_value;
RUN;quit;
```

- ◆ 当然のことながら、「f_rank5」「m_rank5」の組み合わせの人が最も良い顧客と想定できる。

表: F_value * m_value						
m_value(横)						
F_value(縦)	M_RANK1	M_RANK2	M_RANK3	M_RANK4	M_RANK5	合計
F_RANK1	1894	1147	793	388	144	4366
F_RANK2	1939	712	439	220	93	2993
F_RANK3	8125	3839	1762	934	317	14977
F_RANK4	8179	2939	1836	1136	742	14832
F_RANK5	1448	1106	902	493	180	4129
F_RANK6	831	748	500	259	99	2437
F_RANK7	1171	637	708	371	137	3024
F_RANK8	1929	2771	1981	1078	500	8259
F_RANK9	1827	2817	2047	1181	639	8511
F_RANK10	180	480	1028	1119	576	3383
F_RANK11	1391	274	730	384	139	2678
F_RANK12	371	1808	2161	1140	585	6065
F_RANK13	428	1137	2433	1546	742	5886
F_RANK14	22	92	288	496	217	1015
F_RANK15	879	838	179	100	431	2425
F_RANK16	1491	628	1384	727	408	4638
F_RANK17	932	278	482	1275	417	3174
合計	4071	4212	4029	2804	1842	19958
	14.28	14.14	14.05	10.07	6.37	68.91

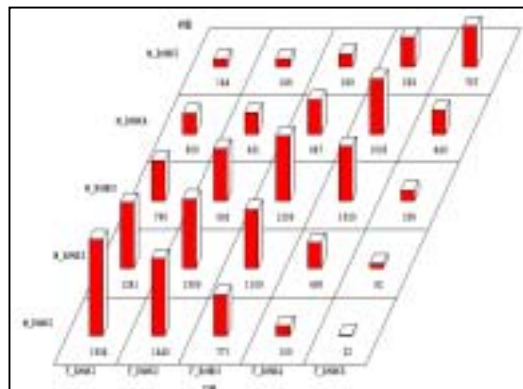
27

RFM分析:ステップ8-2

フォーマットを利用して「F」「M」のランクの分布を見る

```
OPTIONS DEVICE=GIF;
PROC gchart DATA=rfm_point;
  where m_value ne .;
  BLOCK f_value/DISCRETE
  GROUP=m_value
  OUTLINE=BLACK;
  FORMAT f_value f_rank. m_value m_rank.;
RUN;
QUIT;
```

- ◆ 「F」「M」の各ランクをグラフで表現している。他にも「R」や年齢などの属性と組み合わせることにより、色々な事が分かってくると思われる。



28

RFM分析まとめ

- ◆ RFM分析自体は過去から使われてきているので、手法自体は最新のものではない。
- ◆ ただ、今回のサンプルデータのように「顧客ID」「日付」「商品」「値段」程度の情報からでも、何とかして顧客をセグメントするための値として活用する事が出来る。
- ◆ 履歴データとは今回のようなPOSデータだけではなく、Webログ関連のようなデータでも、SASデータセットに取り込めれば活用が可能。



The Power to Know.™

**本日の資料・サンプルプログラム・サンプルデータ(小さく縮小してあります)は
以下のURLから取得できます。(曜日の昼くらいから閲覧可能)**

<http://www.sas.com/japan/usergroups/wg/archive/index.html>