



# 顧客の傾向を把握するための アンケートの可視化

東海大学大学院理学研究科

山田実俊 宗像昌平

指導教員 東海大学理学部

山本義郎

# 目次

- アンケートデータについて
- アソシエーション分析と可視化
- 対応分析と可視化
- 対応分析を利用したアソシエーションルールの可視化
- まとめと今後の課題

# 目的

- 通常顧客データで得られる情報は、性別、年齢、出身地など外面的な情報であるが、アンケートデータを解析することによって内面的な情報を得ることができる。
- 本報告ではメディア層と呼ばれる、M1層(20～34歳の男性)・M2層(35～49歳の男性)・M3層(50歳以上の男性)・F1層(20～34歳の女性)・F2層(35～49歳の女性)・F3層(50歳以上の女性)の6つの層について内面的な特徴に違いがあるかを見つけるために、多肢選択アンケート結果の可視化について提案を行う。

# アンケートデータについて

- 平成25年度データ解析コンペティションにおいて提供されたアンケートデータを解析する。
- アンケートにはSA(単一選択)・MA(多肢選択)・N(自由記述)の3パターンがある。
- メディア層ごとにMAのアンケートを集計する。

| メディア層 |        | 性別 |    |
|-------|--------|----|----|
|       |        | 男性 | 女性 |
| 年齢    | 20～34歳 | M1 | F1 |
|       | 35～49歳 | M2 | F2 |
|       | 50歳～   | M3 | F3 |

## Q41 質問内容

あなたがよく見るテレビ番組の「TVジャンル」を教えてください。

5

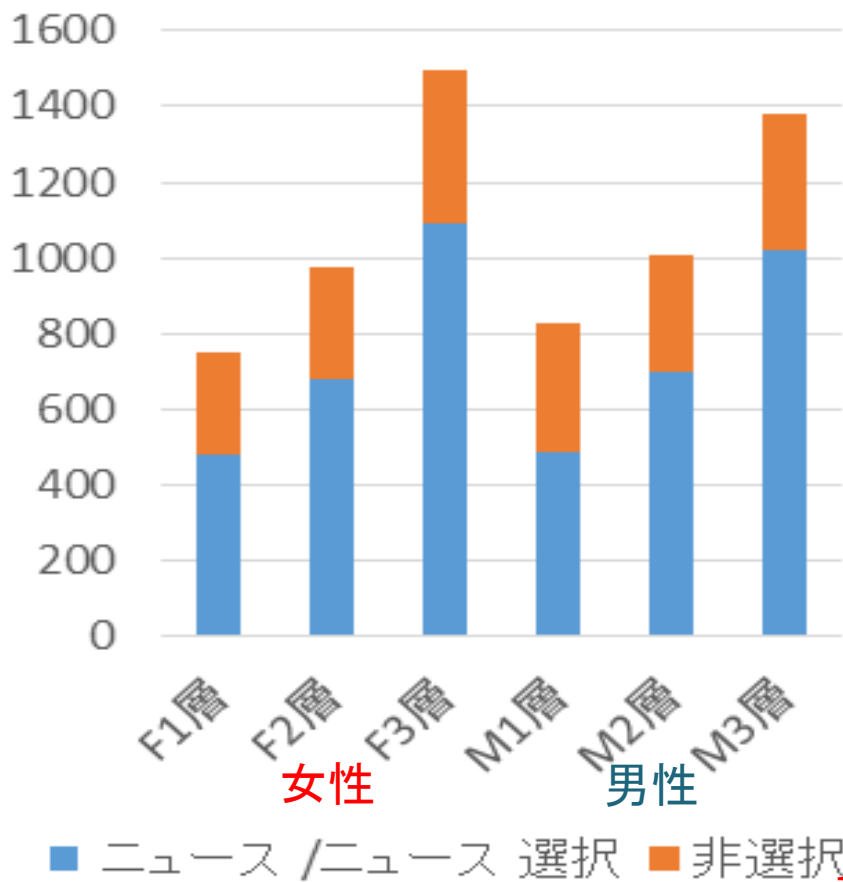
|    |                    |    |                         |    |                                |
|----|--------------------|----|-------------------------|----|--------------------------------|
| 1  | ニュース / ニュース        | 17 | 教育・趣味 / 旅・紀行            | 32 | その他の映画/邦画                      |
| 2  | ニュース / スポーツニュース    | 18 | 教育・趣味/ 料理               | 33 | その他の映画/アジア映画(韓国、中国、台湾 など)      |
| 3  | ニュース / その他         | 19 | 教育・趣味/ 競馬、ギャンブル         | 34 | その他の映画/洋画(アジア映画を除く)            |
| 4  | ワイドショー             | 20 | 教育・趣味/ 幼児・子供向け教育番組      | 35 | 映画/その他                         |
| 5  | ドキュメンタリー           | 21 | 教育・趣味/ その他              | 36 | 音楽/ 歌謡番組(主に、歌手が出演して歌うもの)       |
| 6  | 情報 / その他           | 22 | ドラマ/国内ドラマ               | 37 | 音楽/ 歌謡番組(主に、プロモーションビデオが流されるもの) |
| 7  | バラエティ / コント、漫才、お笑い | 23 | ドラマ/アジアドラマ(韓国、中国、台湾 など) | 38 | 音楽/ クラシック                      |
| 8  | バラエティ / クイズ番組      | 24 | ドラマ/その他の海外ドラマ           | 39 | 音楽 / その他                       |
| 9  | バラエティ / ランキング番組    | 25 | ドラマ / その他               | 40 | 通販                             |
| 10 | バラエティ/情報バラエティ      | 26 | アニメ・特撮 / 国内 ファミリー向け     | 41 | その他【    】                      |
| 11 | バラエティ / アイドル       | 27 | アニメ・特撮/ 国内 青年・大人向け      | 42 | わからない                          |
| 12 | バラエティ / その他        | 28 | アニメ・特撮 / 海外             | 43 | ない                             |
| 13 | スポーツ / 野球          | 29 | アニメ・特撮 / その他            |    |                                |
| 14 | スポーツ / サッカー        | 30 | アニメ・特撮映画/国内             |    |                                |
| 15 | スポーツ / ゴルフ         | 31 | アニメ・特撮映画/海外             |    |                                |
| 16 | スポーツ / その他         |    |                         |    |                                |

# メディア層による回答数の集計

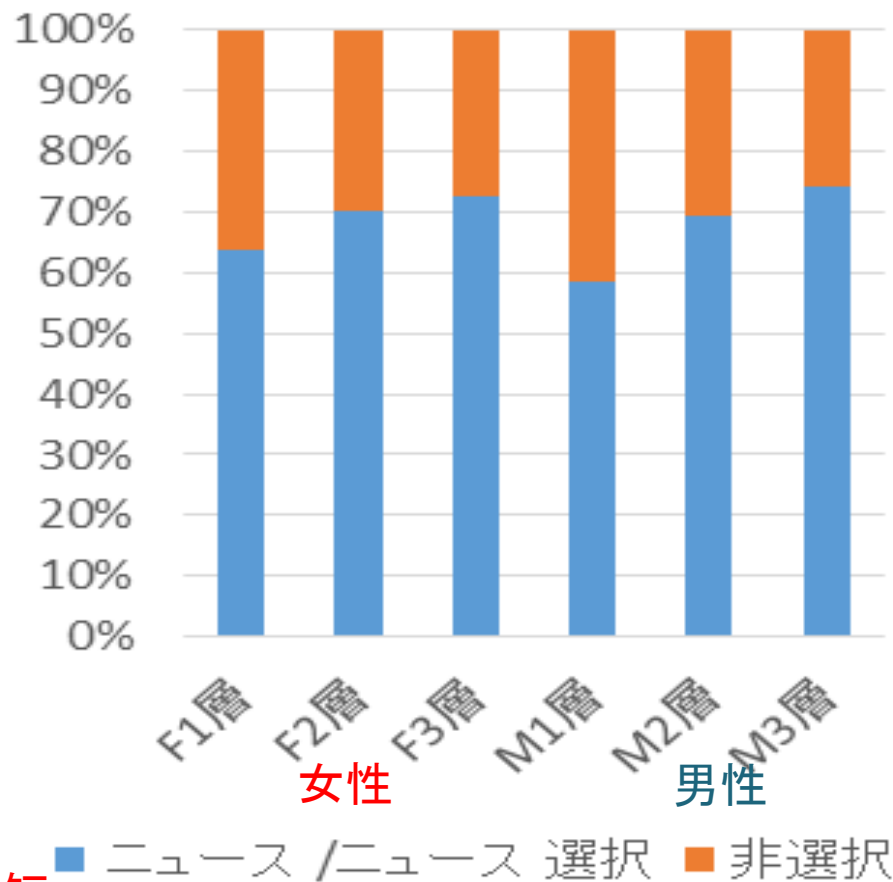
積み上げ棒グラフ

帯グラフ

## ニュース / ニュース



## ニュース / ニュース

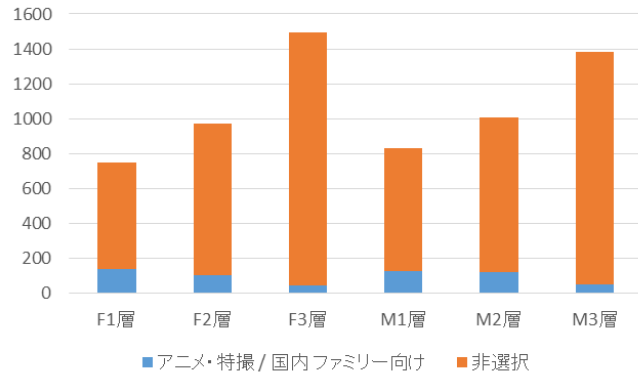


人数を把握しやすい

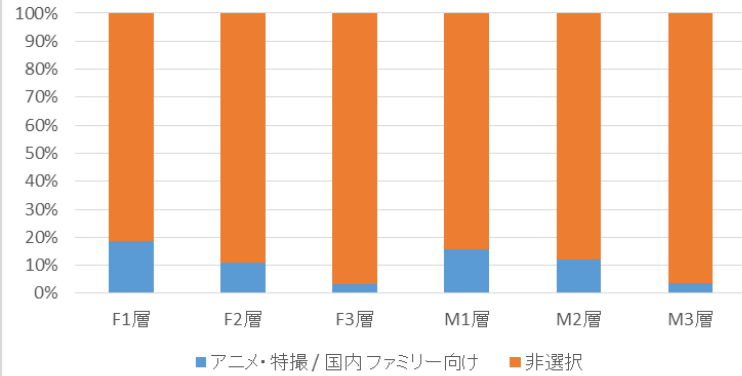
1長1短

割合を把握しやすい

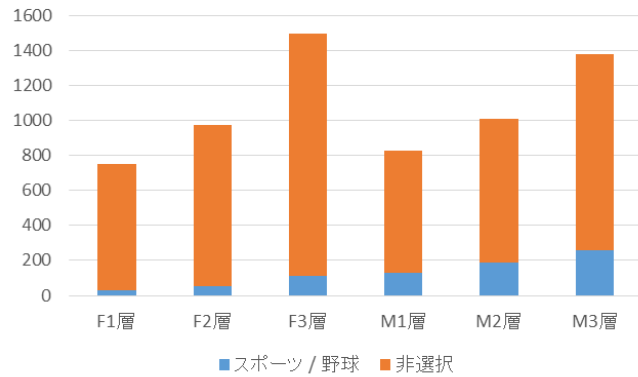
アニメ・特撮 / 国内 ファミリー向け



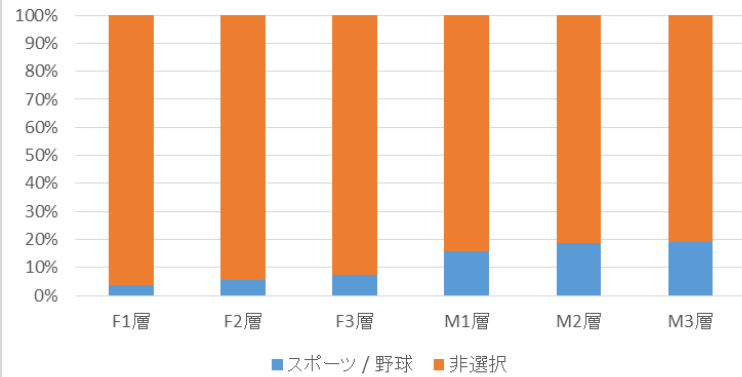
アニメ・特撮 / 国内 ファミリー向け



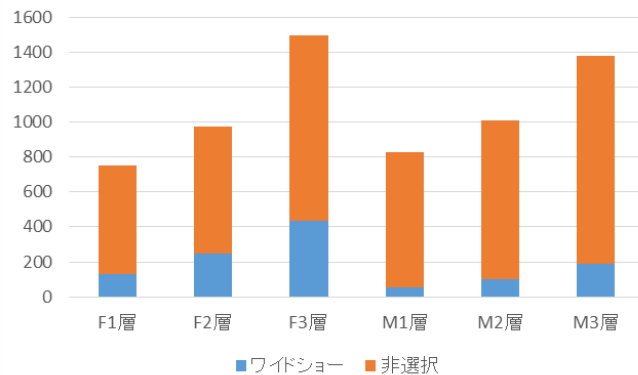
スポーツ / 野球



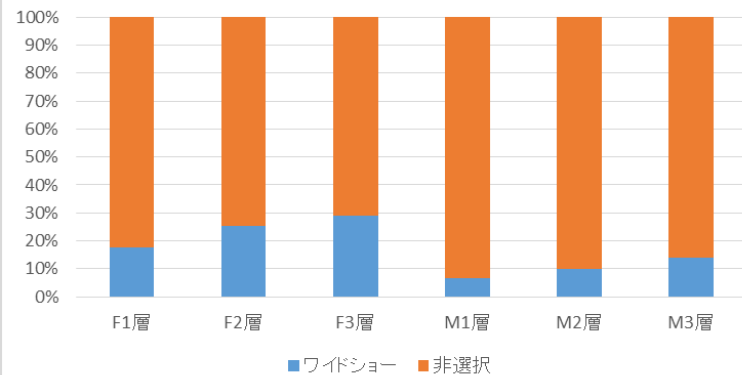
スポーツ / 野球



ワイドショー



ワイドショー



# 集計の結果から

- アニメ系以外のTVジャンルのほとんどがF3層・M3層が多く視聴されていることがわかったが、他のメディア層の特徴はあまり見られなかった。
- このような集計方法ではここで一部しか載せていないように、全ての回答項目についてメディア層に差がある分類を見つけるのは難しい。



# アソシエーション分析

- アソシエーション分析とはデータの要素間の相関をルールの形で抽出することが可能で、データ内に存在する価値のある項目間の関係を把握できる分析である。

$$\{A \Rightarrow B\}$$

(A は条件部、B は結論部)

- アソシエーション分析は支持度・確信度・リフトを求めることによって、関係性の強さを調べる。

# アソシエーション分析に使われる 評価指標

- 支持度

$$\left. \begin{array}{l} \text{Supp}(A \Rightarrow B) \\ \text{Supp}(B \Rightarrow A) \end{array} \right\} = \frac{n(A \cap B)}{n(\Omega)} = P(A \cap B)$$

- 確信度

$$\text{Conf}(A \Rightarrow B) = \frac{\text{Supp}(A \Rightarrow B)}{\text{Supp}(A)} = \frac{n(A \cap B)}{n(A)} = P(B|A)$$

- リフト

$$\text{Lift}(A \Rightarrow B) = \frac{\text{Conf}(A \Rightarrow B)}{P(B)} = \frac{P(B|A)}{P(B)}$$

$n(A)$ : Aのデータの個数 10

# アソシエーション分析で メディア層の違いを見つける

- アソシエーション分析がよく使われるのは同時購買データであるが、MAアンケートデータにおいてもチェックされた項目を購入商品と考えると、ある回答者が同時にチェックした項目についてマーケットバスケット分析と同様に扱うことができる。
- メディア層も1つの項目として扱った。
- アソシエーション分析を用いて、

「**〇〇層だから××である。**」

というメディア層とアンケート結果の関係性を見つける。

# 「メディア層⇒TVジャンル」の アソシエーションルール

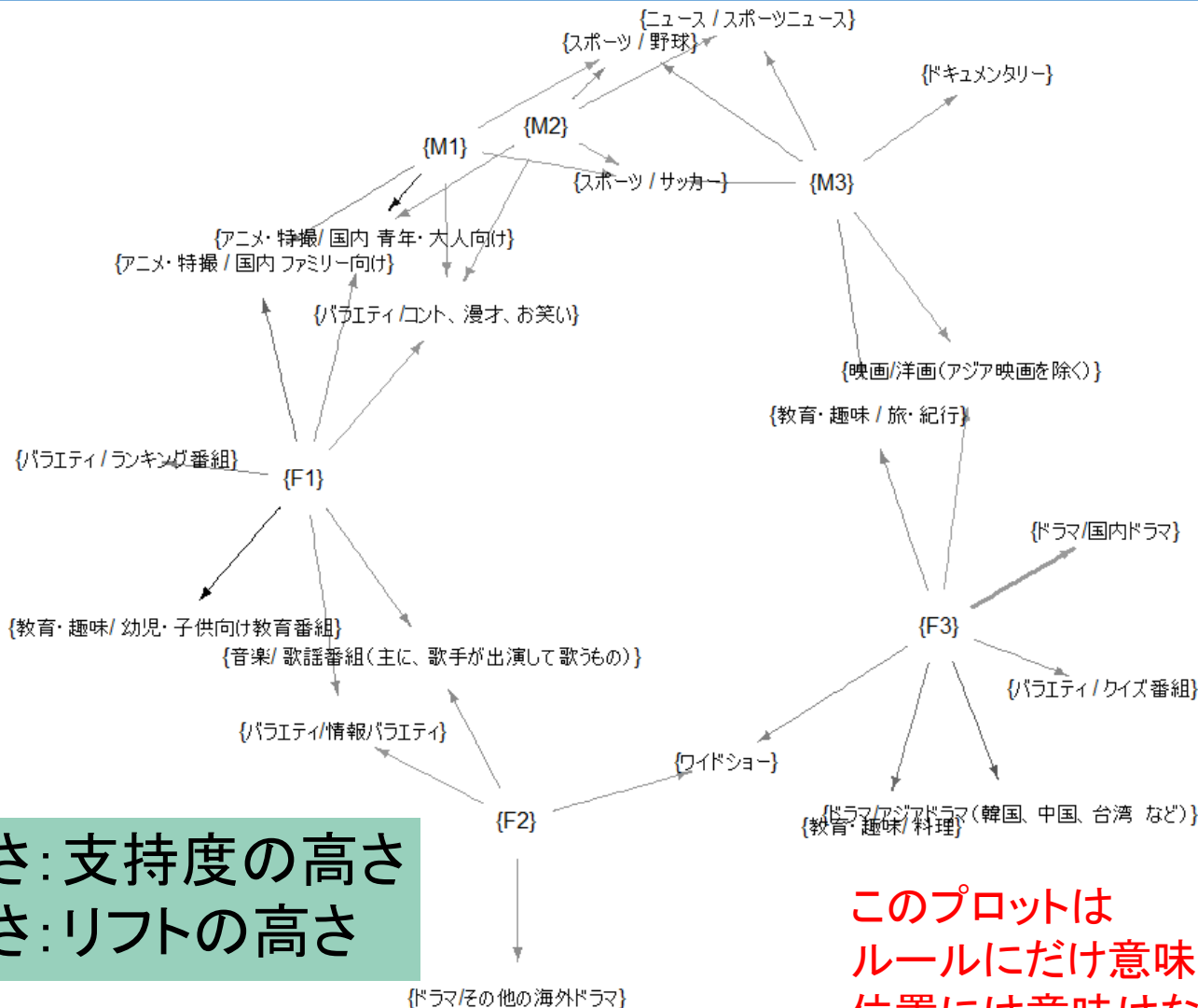
支持度:0.01以上  
確信度:0.1以上  
リフト:1.3以上

|    | lhs | rhs                      | support | confidence | lift  |
|----|-----|--------------------------|---------|------------|-------|
| 1  | F1層 | 教育・趣味/ 幼児・子供向け教育番組       | 0.012   | 0.104      | 3.250 |
| 2  | M1層 | アニメ・特撮/ 国内 青年・大人向け       | 0.028   | 0.216      | 3.048 |
| 3  | F3層 | ドラマ/アジアドラマ(韓国、中国、台湾 など)  | 0.045   | 0.193      | 2.194 |
| 4  | F1層 | アニメ・特撮 / 国内 ファミリー向け      | 0.022   | 0.187      | 2.023 |
| 5  | F3層 | 教育・趣味/ 料理                | 0.029   | 0.124      | 1.920 |
| 6  | M3層 | 教育・趣味/ その他               | 0.074   | 0.345      | 1.693 |
| 7  | M1層 | アニメ・特撮 / 国内 ファミリー向け      | 0.020   | 0.156      | 1.687 |
| 8  | F3層 | ワイドショー                   | 0.068   | 0.291      | 1.615 |
| 9  | F1層 | バラエティ / ランキング番組          | 0.013   | 0.111      | 1.612 |
| 10 | M3層 | スポーツ / 野球                | 0.041   | 0.189      | 1.575 |
| 11 | M2層 | スポーツ / 野球                | 0.029   | 0.186      | 1.550 |
| 12 | F3層 | 教育・趣味/ その他               | 0.073   | 0.313      | 1.535 |
| 13 | F3層 | バラエティ / クイズ番組            | 0.062   | 0.268      | 1.522 |
| 14 | M3層 | スポーツ / サッカー              | 0.049   | 0.230      | 1.508 |
| 15 | F1層 | 音楽/ 歌謡番組(主に、歌手が出演して歌うもの) | 0.022   | 0.189      | 1.490 |

抽出された全34ルールのうち、  
リフトの高い順で15ルールを表示している

# アソシエーションルールのプロット

支持度: 0.01以上  
確信度: 0.1以上  
リフト: 1.3以上



線の太さ: 支持度の高さ  
線の濃さ: リフトの高さ

このプロットは  
ルールにだけ意味があり、  
位置には意味はない

# 対応分析

- 対応分析はクロス集計結果を用いて、行の要素と列の要素の相関関係が最大になるように数量化して、行の要素と列の要素を多次元空間(今回は2次元)に表現する分析である。
- 対応分析の基礎となるコレスポネンス行列 $Z$

$$z_{ij} = \frac{f_{ij} - f_{i.} \times f_{.j} / n}{\sqrt{f_{i.} \times f_{.j}}} \quad \begin{matrix} (i = 1, \dots, r) \\ (j = 1, \dots, c) \end{matrix}$$

$f_{ij}$ : クロス集計の各成分,       $f_{i.}$ : クロス集計の各行和  
 $f_{.j}$ : クロス集計の各列和,       $n$ : クロス集計の総和

# 対応分析の成分得点

- 行得点

$$X = D_r^{-\frac{1}{2}}V$$

$D_r$ :  $p_{i\cdot}$ を要素とした対角行列 ( $p_{i\cdot} = \frac{f_{i\cdot}}{n}$ )

$V$ :  $ZZ^t$ の固有ベクトル

- 列得点

$$Y = D_c^{-\frac{1}{2}}U$$

$D_c$ :  $p_{\cdot j}$ を要素とした対角行列 ( $p_{\cdot j} = \frac{f_{\cdot j}}{n}$ )

$U$ :  $Z^tZ$ の固有ベクトル

# 対応分析の結果

クロス集計

|                     | F1層 | F2層 | F3層  | M1層 | M2層 | M3層  |
|---------------------|-----|-----|------|-----|-----|------|
| ニュース / ニュース         | 478 | 681 | 1089 | 484 | 700 | 1022 |
| ニュース / スポーツニュース     | 65  | 121 | 266  | 211 | 314 | 442  |
| ワイドショー              | 132 | 248 | 436  | 55  | 100 | 191  |
| ドキュメンタリー            | 91  | 192 | 477  | 116 | 225 | 487  |
| バラエティ / コント、漫才、お笑い  | 216 | 239 | 187  | 241 | 290 | 199  |
| バラエティ / クイズ番組       | 127 | 176 | 401  | 80  | 134 | 216  |
| スポーツ / 野球           | 27  | 54  | 112  | 131 | 188 | 261  |
| スポーツ / サッカー         | 48  | 89  | 162  | 166 | 201 | 318  |
| 教育・趣味 / 料理          | 41  | 76  | 186  | 17  | 32  | 65   |
| 教育・趣味 / 幼児・子供向け教育番組 | 78  | 66  | 8    | 23  | 29  | 2    |
| ドラマ / 国内ドラマ         | 357 | 501 | 805  | 178 | 295 | 519  |
| アニメ・特撮 / 国内ファミリー向け  | 140 | 105 | 48   | 129 | 121 | 51   |
| アニメ・特撮 / 国内青年・大人向け  | 79  | 39  | 11   | 179 | 102 | 46   |
| 映画 / 邦画             | 72  | 79  | 152  | 57  | 65  | 152  |
| 音楽 / クラシック          | 6   | 18  | 74   | 10  | 9   | 72   |



行得点

|                     | X軸     | Y軸     |
|---------------------|--------|--------|
| ニュース / ニュース         | 0.039  | 0.040  |
| ニュース / スポーツニュース     | 0.329  | 1.549  |
| ワイドショー              | -0.910 | -1.184 |
| ドキュメンタリー            | -0.765 | 0.669  |
| バラエティ / コント、漫才、お笑い  | 1.370  | -0.297 |
| バラエティ / クイズ番組       | -0.659 | -0.565 |
| スポーツ / 野球           | 0.602  | 2.005  |
| スポーツ / サッカー         | 0.473  | 1.603  |
| 教育・趣味 / 料理          | -1.308 | -1.115 |
| 教育・趣味 / 幼児・子供向け教育番組 | 2.190  | -3.352 |
| ドラマ / 国内ドラマ         | -0.424 | -0.812 |
| アニメ・特撮 / 国内ファミリー向け  | 2.181  | -0.920 |
| アニメ・特撮 / 国内青年・大人向け  | 3.240  | 0.657  |
| 映画 / 邦画             | -0.250 | -0.008 |
| 音楽 / クラシック          | -1.781 | 0.868  |

列得点

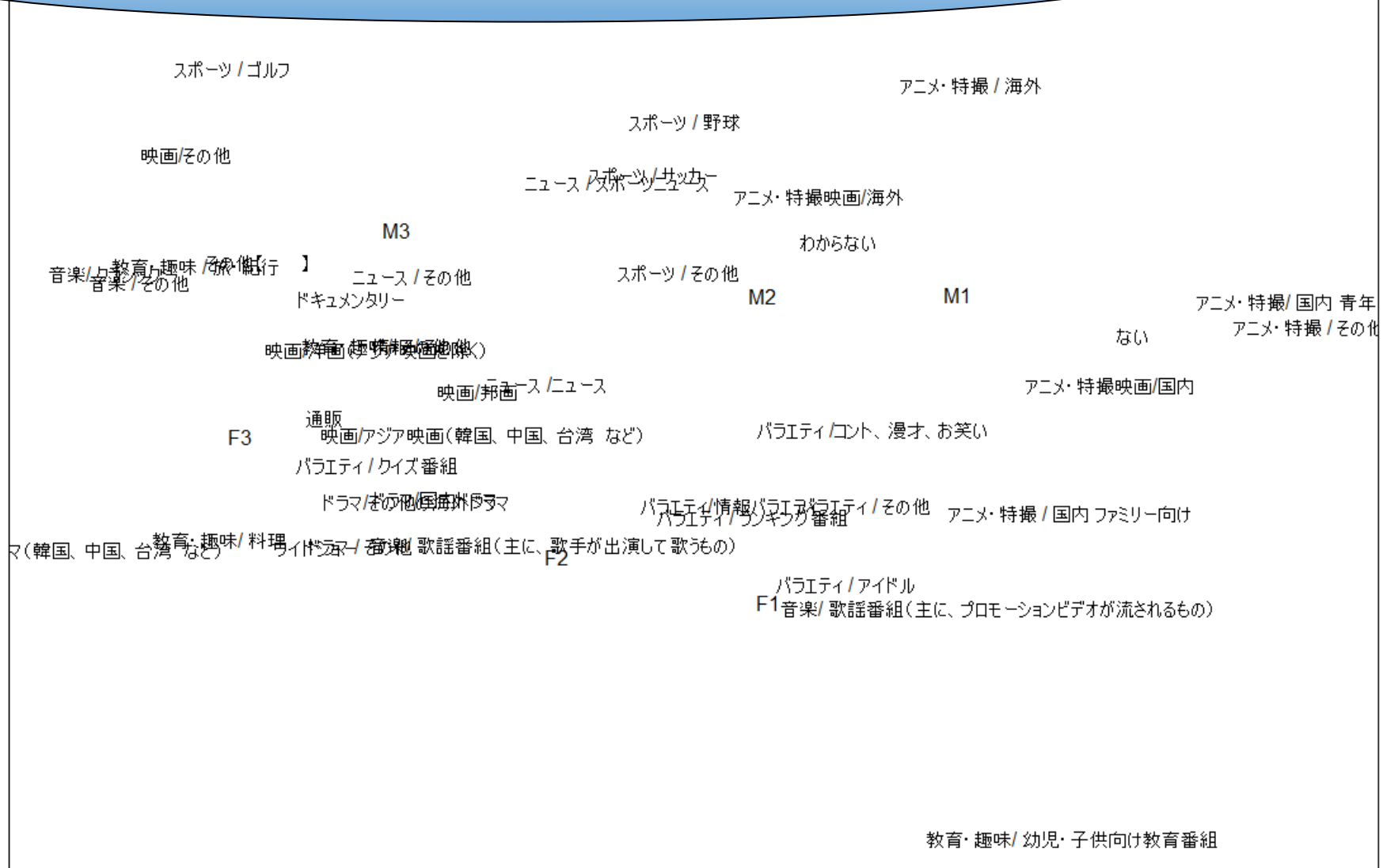
|     | X軸     | Y軸     |
|-----|--------|--------|
| F1層 | 0.945  | -1.592 |
| F2層 | 0.076  | -1.255 |
| F3層 | -1.222 | -0.356 |
| M1層 | 1.720  | 0.710  |
| M2層 | 0.919  | 0.694  |
| M3層 | -0.585 | 1.191  |

注: 実際は43項目全てに対しクロス集計を行い、行得点・列得点を算出している



あなたがよく見るテレビ番組の「TVジャンル」を教えてください。

## 対応分析によるアンケート回答の傾向の可視化





あなたがよく見るテレビ番組の「TVジャンル」を教えてください。

# メディア層を色別に表示



円の大きさ: 質問の回答数  
 ひし形の大きさ: メディア層の割合

あなたがよく見るテレビ番組の「TVジャンル」を教えてください。

# TVジャンルの大分類を色別に表示

教育・趣味/ 競馬、ギャンブル

アニメ・特撮 / 海外

スポーツ/ ゴルフ

スポーツ/ 野球

映画/ その他

ニュース/ スポーツ/ サッカー

アニメ・特撮映画/ 海外

M3

わからない

音楽/ 教育/ 趣味/ その他/ 旅行  
音楽/ その他

ニュース/ その他  
ドキュメンタリー

スポーツ/ その他

M2

M1

アニメ・特撮/ 国内 青年  
アニメ・特撮 / その他

映画/ 邦画 (連綿報道番組を除く)

映画/ 邦画 / ニュース

アニメ・特撮映画/ 国内

通販  
映画/ アジア映画 (韓国、中国、台湾 など)

バラエティ/ コント、漫才、お笑い

バラエティ/ クイズ番組

ドラマ/ その他/ 国内/ 海外/ ドラマ

バラエティ/ 情報/ バラエティ/ その他  
バラエティ/ キング番組

アニメ・特撮 / 国内 ファミリー向け

教育・趣味/ 料理 / ドラマ / 音楽 / 歌謡番組 (主に、歌手が出演して歌うもの)

バラエティ/ アイドル  
音楽 / 歌謡番組 (主に、プロモーションビデオが流されるもの)

円の大きさ: 質問の回答数  
ひし形の大きさ: メディア層の割合

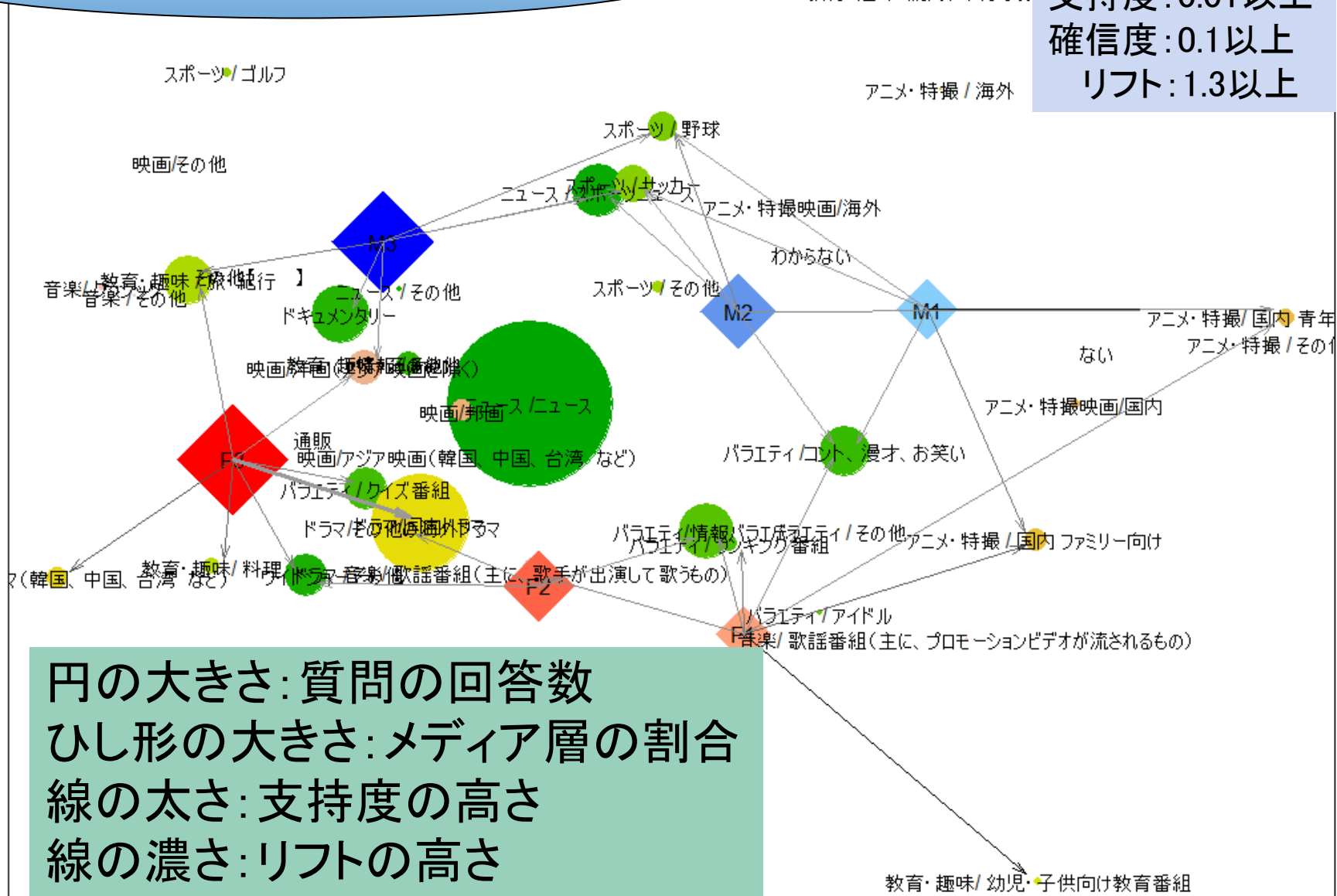
あなたがよく見るテレビ番組の「TVジャンル」を教えてください。

# アソシエーションルールの追加

教育・趣味/ 競馬、ギャンブル

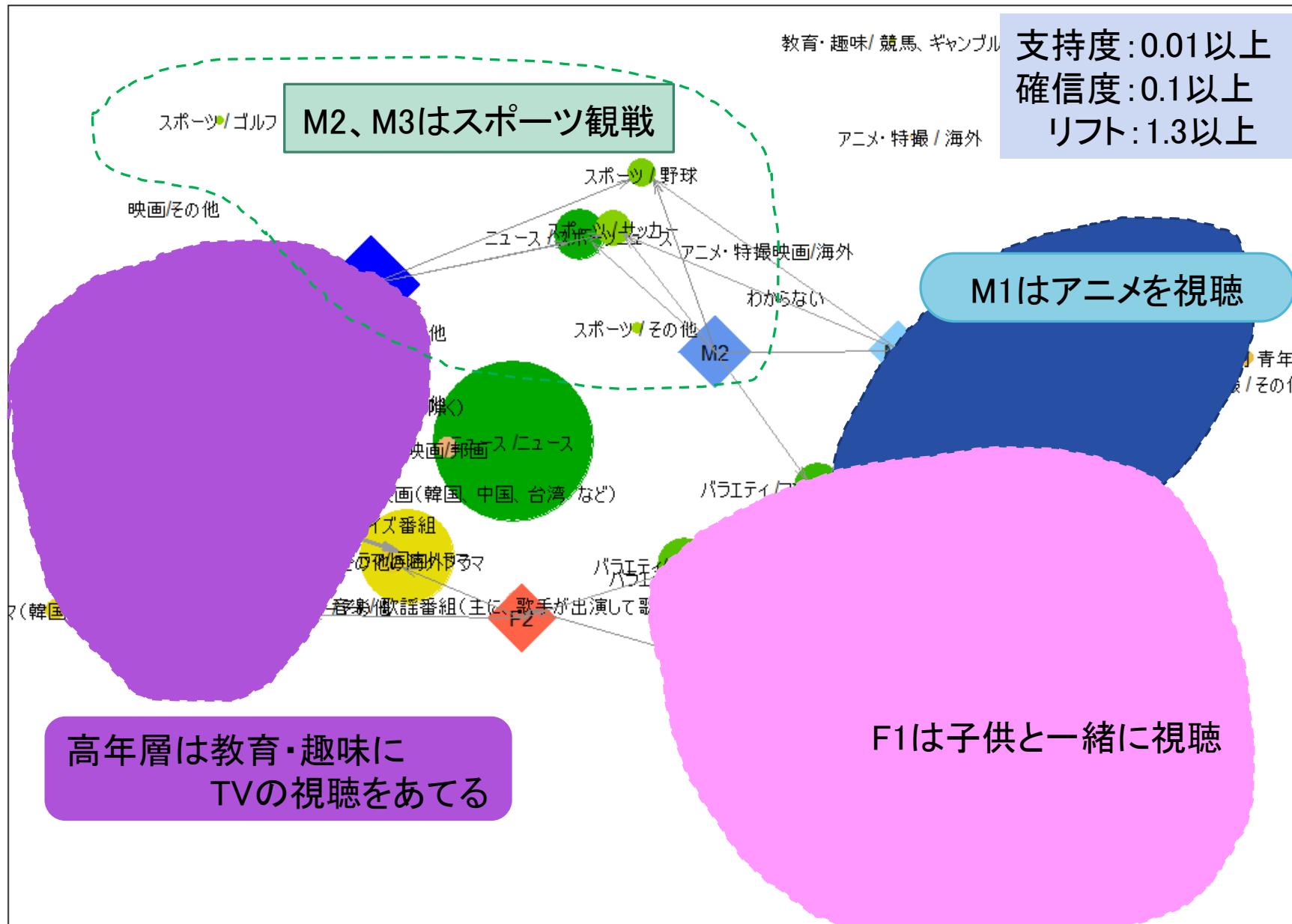
支持度: 0.01以上  
確信度: 0.1以上  
リフト: 1.3以上

アニメ・特撮 / 海外



円の大きさ: 質問の回答数  
 ひし形の大きさ: メディア層の割合  
 線の太さ: 支持度の高さ  
 線の濃さ: リフトの高さ

あなたがよく見るテレビ番組の「TVジャンル」を教えてください。



# 可視化したいアンケート項目の選択

VRP データハンドリング

対象データ: 6,438 行  
1,088 列

行選択 列選択 列追加 集計 日付時刻処理

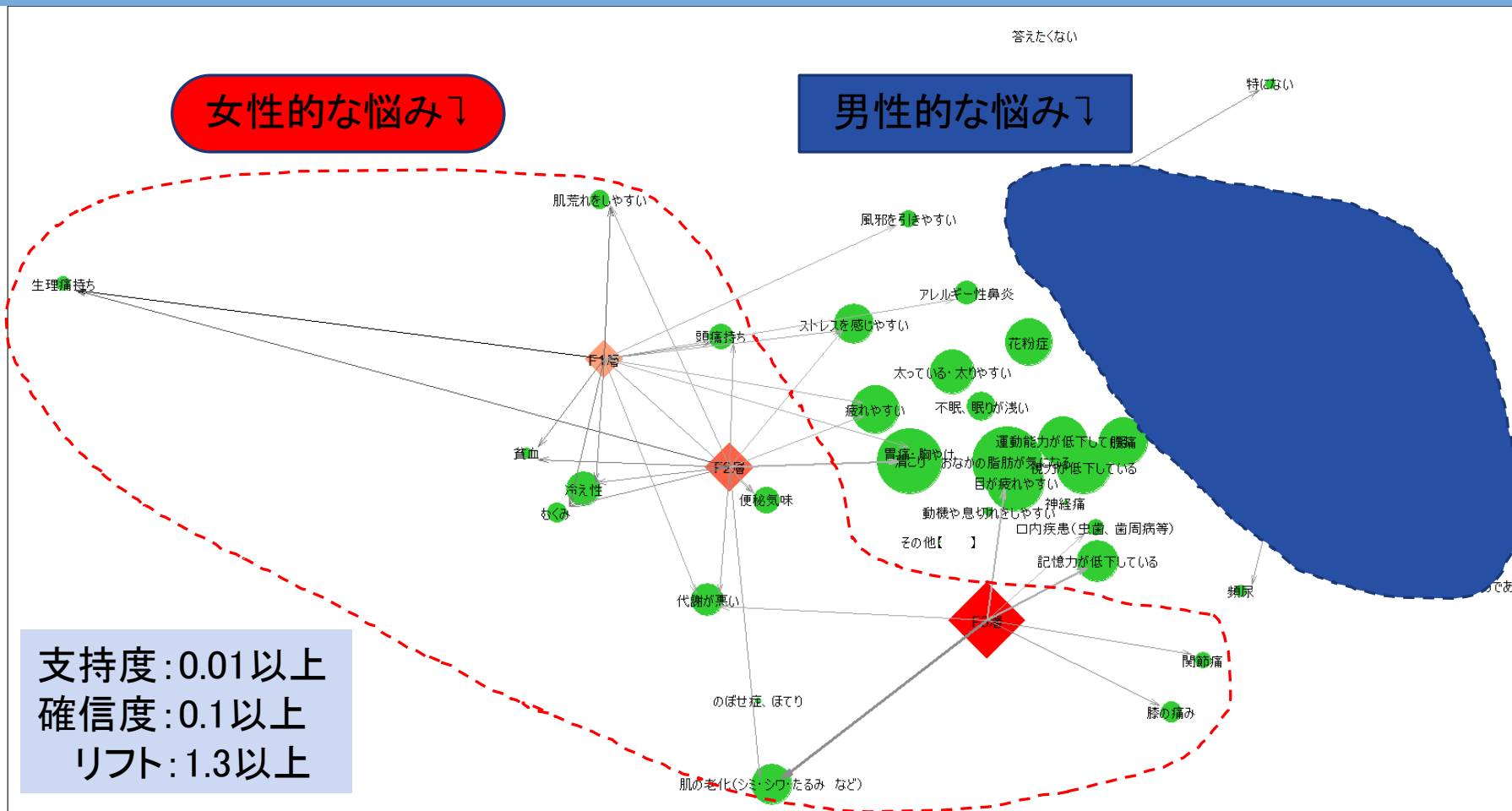
結果名: Q14

| 選択                                  | 列名       | 列属性 |
|-------------------------------------|----------|-----|
| <input checked="" type="checkbox"/> | SAMPLEID | 整数  |
| <input type="checkbox"/>            | Q1       | 整数  |
| <input type="checkbox"/>            | Q2S1N    | 整数  |
| <input type="checkbox"/>            | Q2S2N    | 整数  |
| <input type="checkbox"/>            | Q4S1     | 整数  |
| <input type="checkbox"/>            | Q4S2     | 整数  |
| <input type="checkbox"/>            | Q4S3     | 整数  |
| <input type="checkbox"/>            | Q4S4     | 整数  |
| <input type="checkbox"/>            | Q4S5     | 整数  |
| <input type="checkbox"/>            | Q4S6     | 整数  |
| <input type="checkbox"/>            | Q4S7     | 整数  |
| <input type="checkbox"/>            | Q4S8     | 整数  |
| <input type="checkbox"/>            | Q4S9     | 整数  |
| <input checked="" type="checkbox"/> | Q14_1    | 整数  |
| <input checked="" type="checkbox"/> | Q14_2    | 整数  |
| <input checked="" type="checkbox"/> | Q14_3    | 整数  |
| <input checked="" type="checkbox"/> | Q14_4    | 整数  |
| <input checked="" type="checkbox"/> | Q14_5    | 整数  |

- Visual R Platformのデータハンドリングによって可視化したいアンケートを選択する。
- Shiftキーを押しながらチェックをすれば、挟まれた区間のチェックが全て変わる(オセロのように)

# 他のアンケート結果（健康の悩み）

24





# アソシエーションルールの パラメータの変更

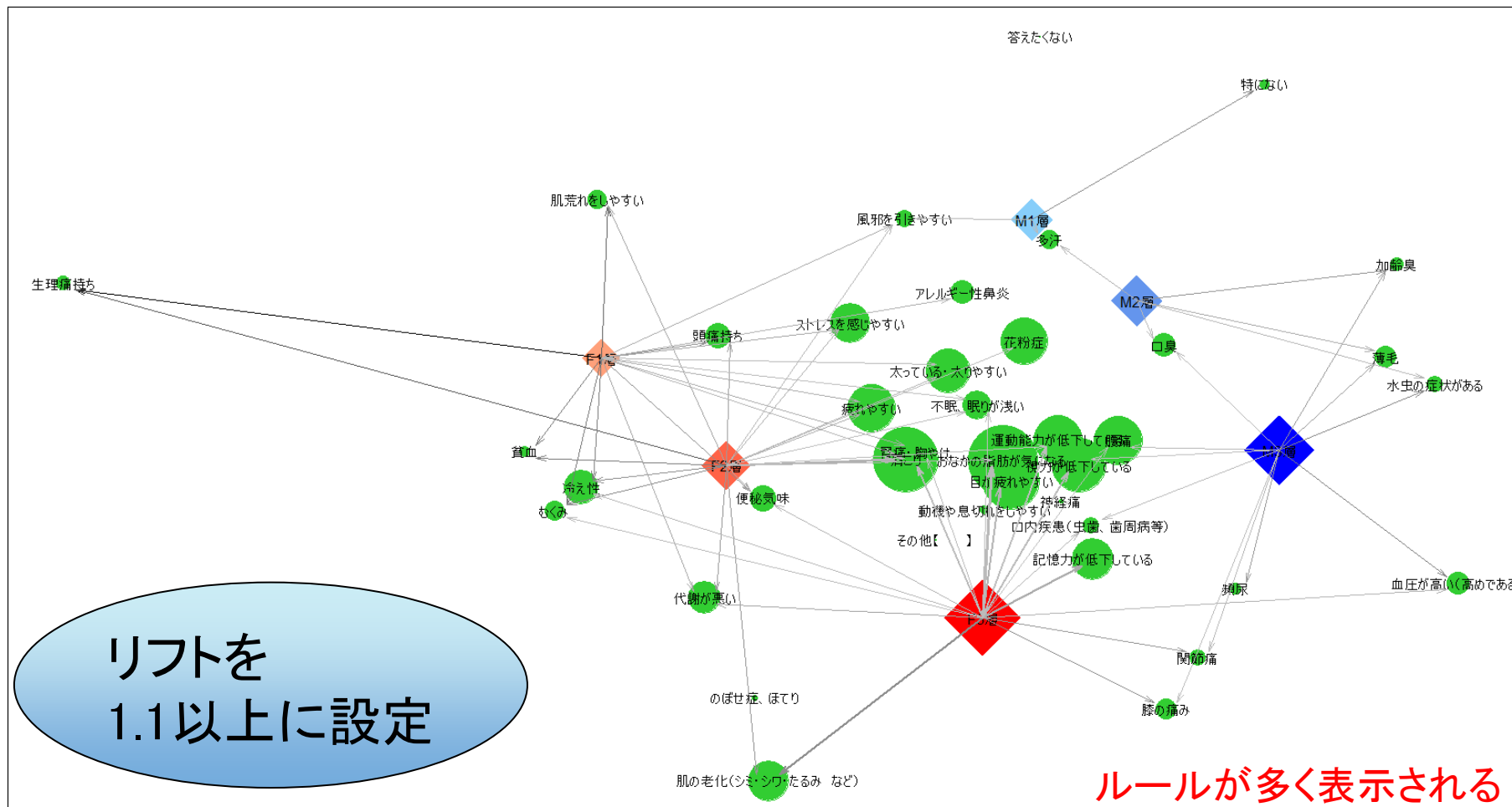
- アソシエーションルールは支持度、確信度、リフトの下限の設定により抽出されるルールが変わるため、得られるプロットに差異が見られる。
- メディア層での特徴付けをしやすいルールを表示するために、それらのパラメータを調整する必要がある。

支持度:0.01以上 確信度:0.1以上 **リフト:1.1以上**

26

(Rスクリプトで実現)

普段感じている健康上の悩みについて、あなたに当てはまるものを全てお知らせください。

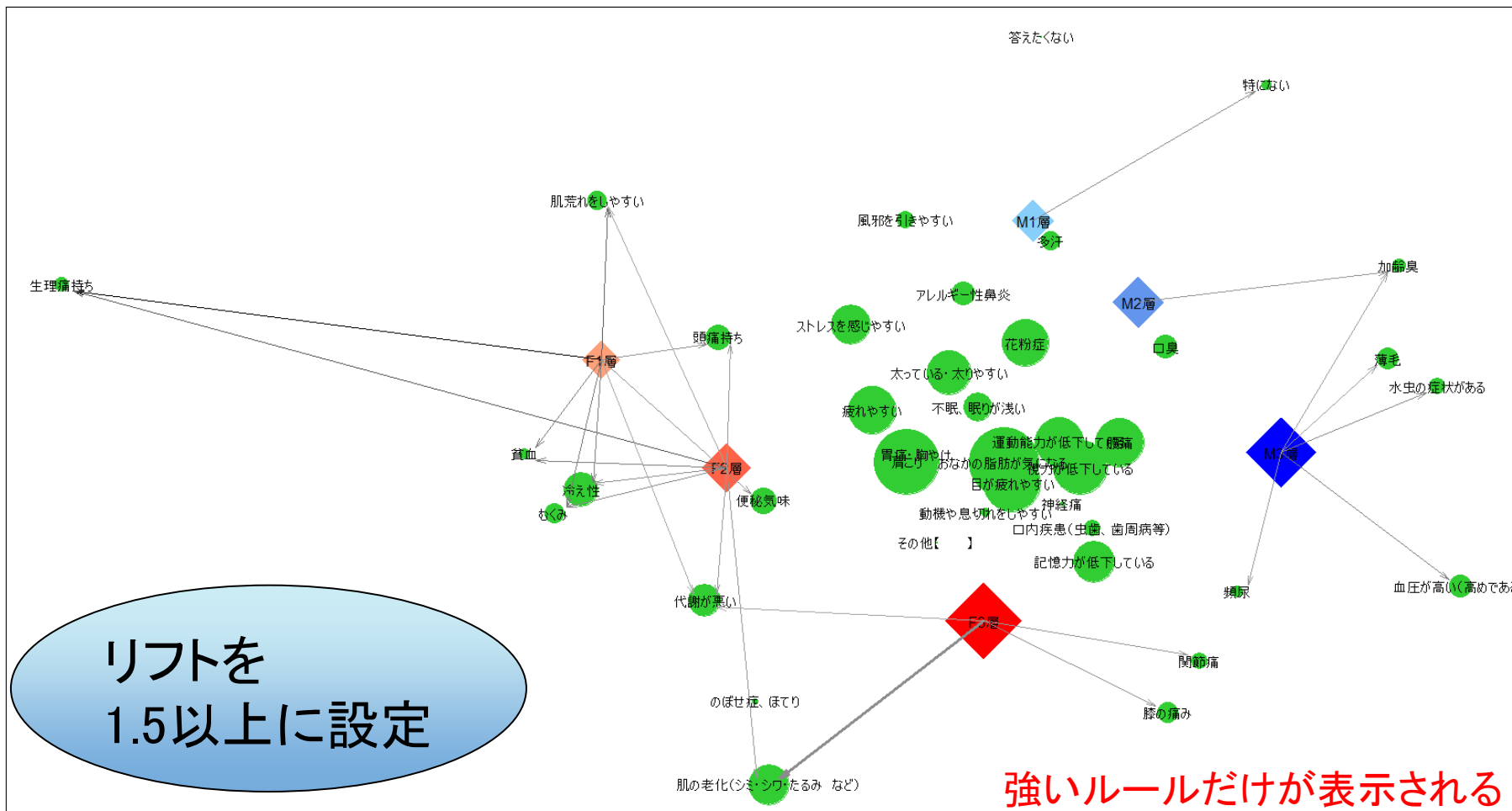


支持度:0.01以上 確信度:0.1以上 **リフト:1.5以上**

27

(Rスクリプトで実現)

普段感じている健康上の悩みについて、あなたに当てはまるものを全てお知らせください。



# まとめ

- 今回提案した対応分析とアソシエーション分析を組み合わせた可視化によってMAアンケートについてメディア層の反応を把握しやすくなった。
- 複数の質問にも対応させ、可視化を自分の見やすい形で操作することができた。
- 2項目以上のアソシエーションルールに対する可視化についても考えたい。

# 参考文献

- 山田実俊, 山本義郎 (2014), 多肢選択アンケートのメディア層の反応の可視化. 日本計算機統計学会第28回大会 論文集, pp.19-22
- 伊藤晃, 吉川大弘, 古橋武, 池田龍二, 加藤孝浩 (2010) . アソシエーション分析における可視化を用いた興味深いルールの探索. 名古屋大学, トッパン・フォームズ株式会社, pp.684-689.
- Rによるアソシエーション分析,  
<http://qiita.com/hidetarou2013@github/items/dc3b448542c5fce7a6ce>
- 山口和範, 高橋淳一, 竹内光悦, 「図解入門 よくわかる多変量解析の基本と仕組み」, 株式会社 秀和システム, 2004.
- 中山慶一郎(2009), 対応分析によるデータ解析. pp.133-145
- 「Rと対応分析」, フリーソフトによるデータ解析・マイニング第26回-同志社大学情報公開用サーバ, <http://www1.doshisha.ac.jp/~mjn/R/26/26.html>