

カワイイを数学的に考えてみた

芝浦工業大学 数理科学研究会

～多変量解析の基礎知識～

平成 24 年 3 月 28 日

何か不明な点や計算ミス等がありましたら加筆修正しますので指摘をお願いします

制作:若松宏輔
前川健

目次

1	はじめに	1
2	調査項目・調査対象	1
3	調査項目と順位についての相関係数	1
4	ヒストグラム	2
4.1	AKB	2
4.2	女優・タレント	6
4.3	分析結果	9
5	主成分分析	9
5.1	女優・タレント	10
5.2	AKB	12
5.3	分析結果	12
6	因子分析	13
6.1	分析結果	13
7	クラスター分析	14
7.1	AKB	15
7.2	女優・タレント	16
7.3	分析結果	17
8	判別分析	17
8.1	分析結果	19

1 はじめに

大学生生活の最重要事項の一つ、それは「彼女をつくること」である。そして、せっかくならカワイイ人とお付き合いをしたいと誰もが考える。そこで、カワイイとはどのようなことなのかについてある指標をもうけて調べることにした。ここでの指標とは「顔のパーツに表れる比率」をもちいて考えることにした。また、ここでの「顔」とは「髪などに隠されていない、顔のよく見えている部分」とする。

2 調査項目・調査対象

調査項目は七項目をもうけた。うち三つが顔のパーツのバランスに関するもので、残りの四つは目と顔のバランスに関するものである。上記より、「顔」とは「髪などに隠されていない、顔のよく見えている部分」とするので前髪を下ろしている人の場合は「髪の生え際」ではなく、顔の見える前髪の毛先からとして測った。また、それぞれの調査項目には理想比を設置した。

調査対象としては人気絶頂のAKB48のみなさんにした。なぜなら、彼女たちは選挙なるものを行い順位を決める。よって、これには少なからず国民の民意が反映され順位が上位ほど多くの人が「カワイイ」と感じているはずである。さらにあるランキングサイトからアイドルだけではなく、女優・タレント50人についても調査した。

調査項目	略称	理想比
「顔の横幅」：「髪の生え際からあごの先まで」	「顔横」：「顔縦」	1:1.61
「鼻の頭からあごの先まで」：「髪の生え際から鼻の頭までの長さ」	「BD」：「AB」	1:1.61
「鼻の頭から口の中心」：「口の中心からあごの先まで」	「BC」：「CD」	1:1.61
「目と目の間」：「鼻の幅」	「間」：「鼻」	1:1
「目と目の間」：「目の横幅」	「間」：「横」	1:1
「目の縦」：「目の横幅」	「縦」：「横」	1:2~3
「目の横幅」：「顔の横幅」	「目横」：「顔横」	1:4~5

3 調査項目と順位についての相関係数

相関係数は、 r_{xy} であらわし、 x と y の直感的な関連の程度を表わすものである。相関係数の数値は、 $-1 \sim 1$ の間で、1に近いほどに正の相関が強いと呼び x が増加すると、 y が増加する傾向が強くなる。つまり x と y の関係性が強いことを示している。また、数値が -1 に近いほど負の相関が強いと呼び、 y が減少する傾向が強くなり x と y の関係が真逆であることを示している。そして、数値が 0 に近いときは無相関だと考えるが、注意しなければならないのは $r_{xy} \doteq 0$ だからといって x と y に関連がないとは限らないことである。

AKBの相関係数

	順位	ABBD	CDBC	顔縦顔横	間鼻	間横	横縦	差平均	顔横目横	差平均1
順位	1	-0.076	0.112	-0.004	-0.063	-0.002	-0.180	0.036	-0.056	0.288

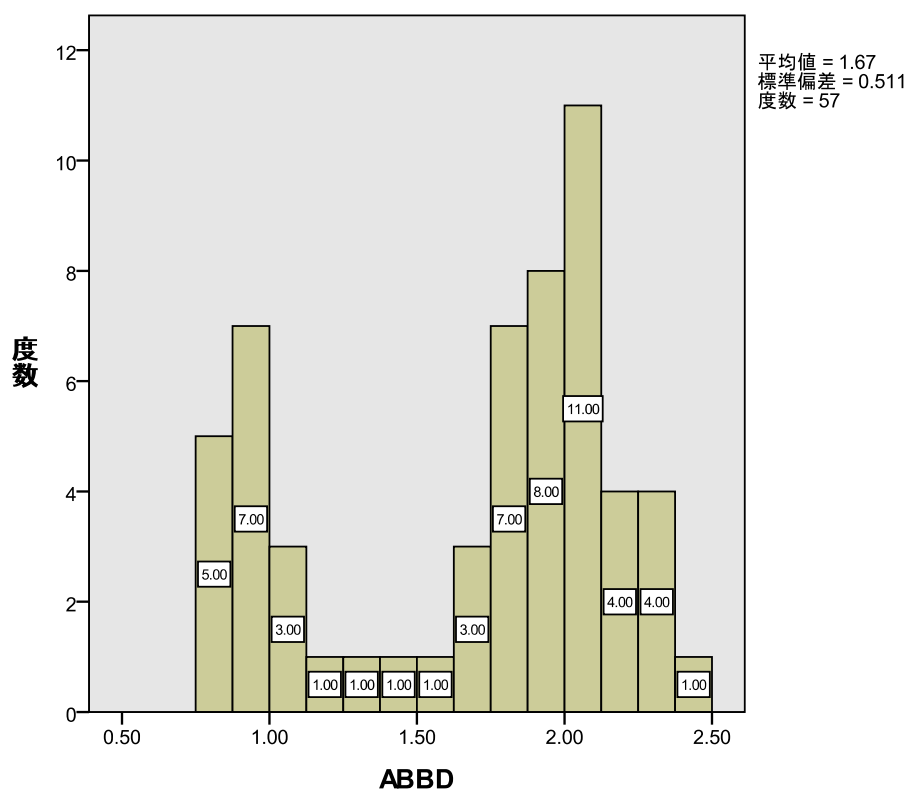
女優・タレントの相関係数

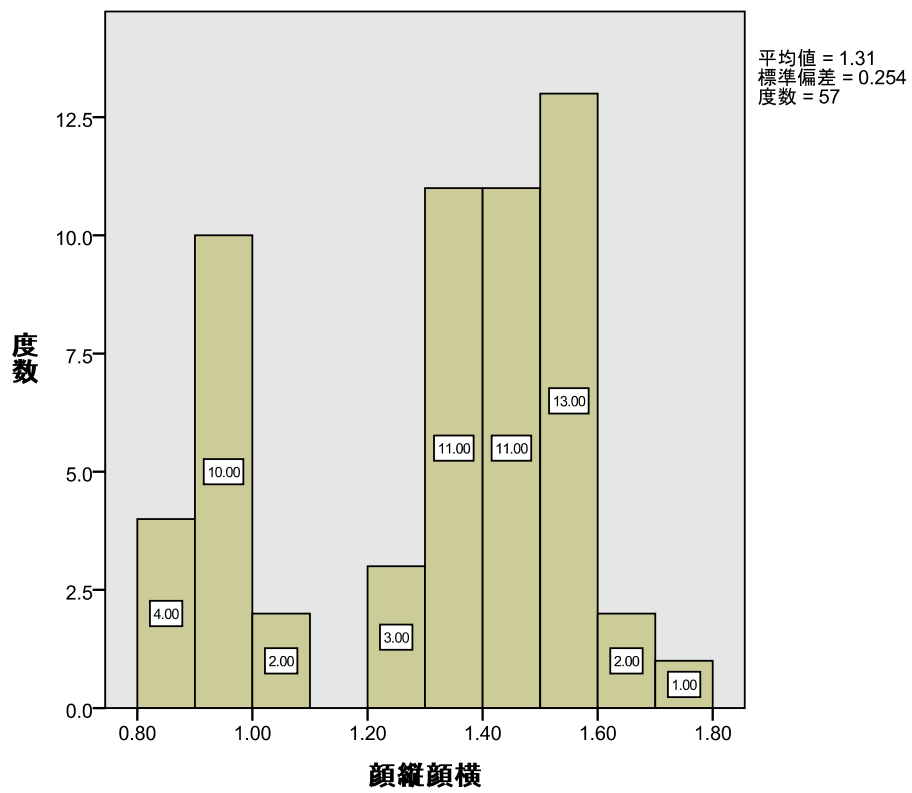
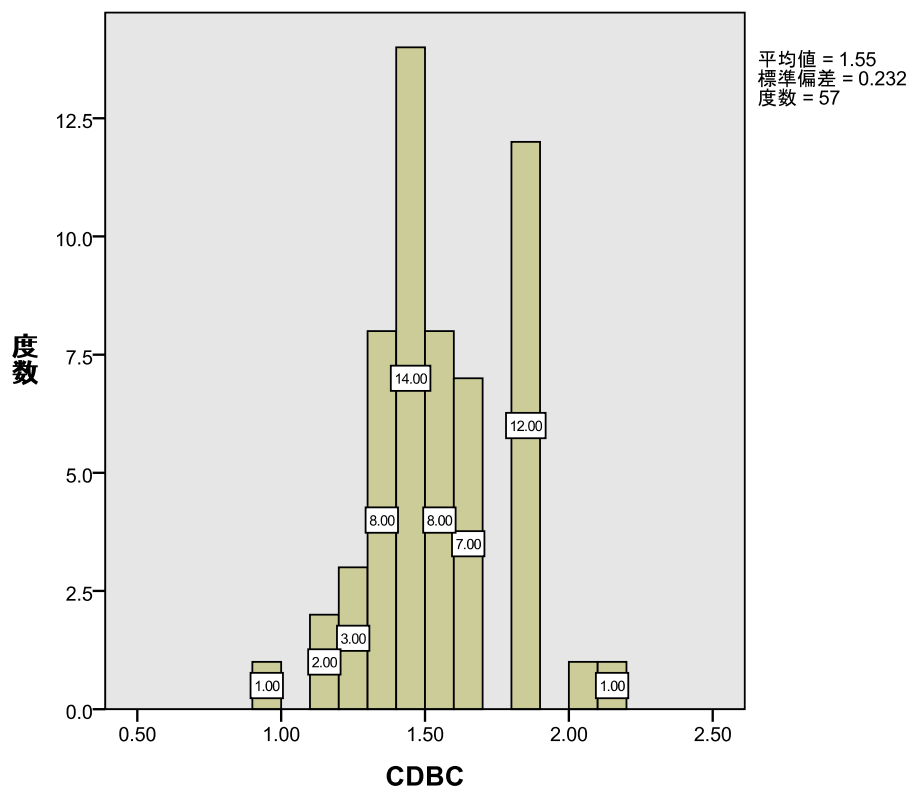
	順位	ABBD	CDBC	顔縦顔横	間鼻	間横	横縦	顔横目横
順位	1	-0.218	-0.112	-0.069	-0.158	-0.149	0.397	-0.004

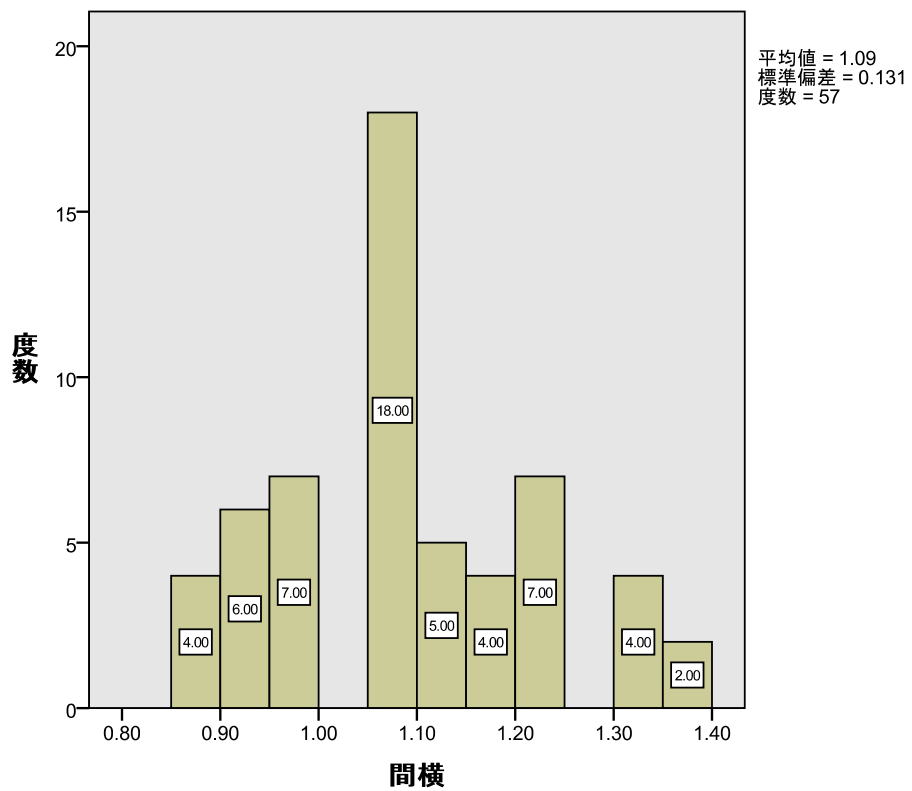
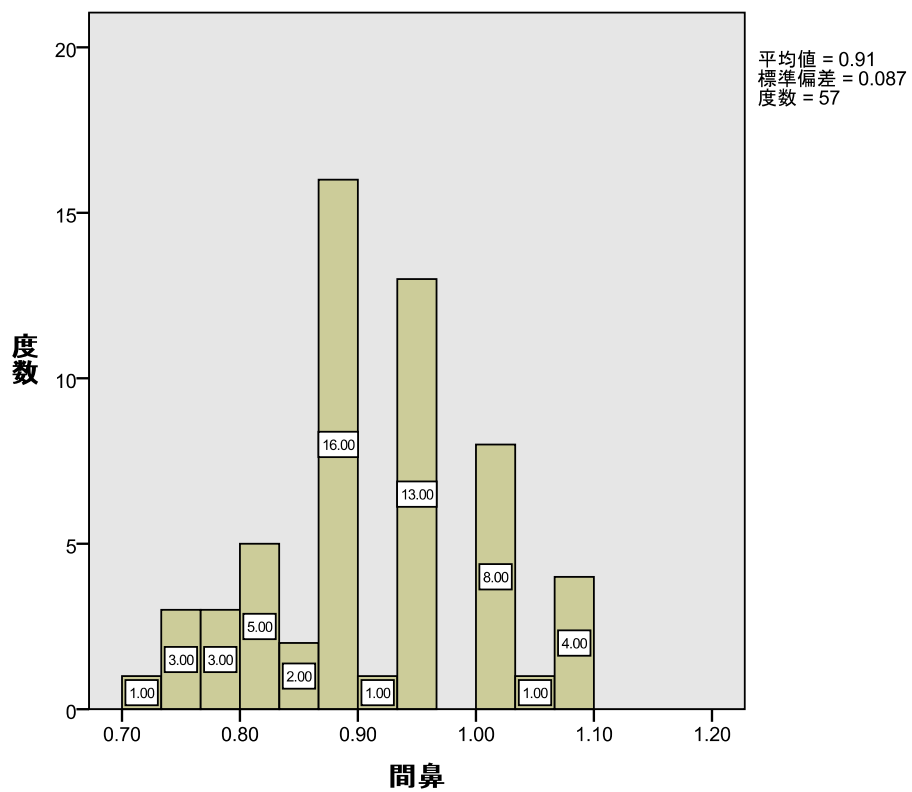
4 ヒストグラム

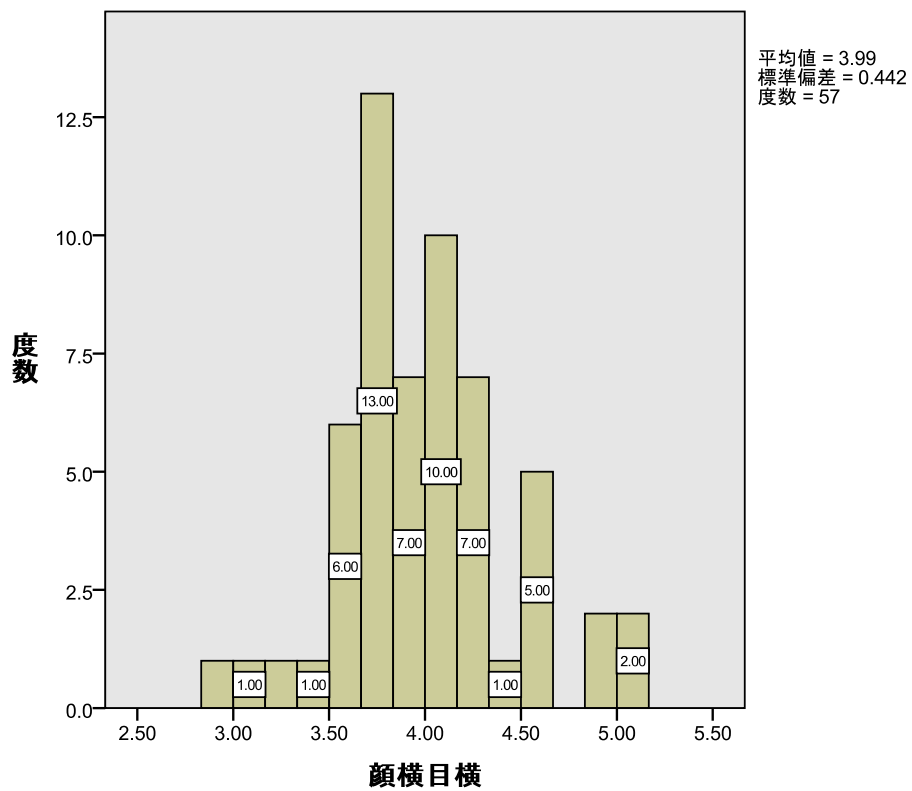
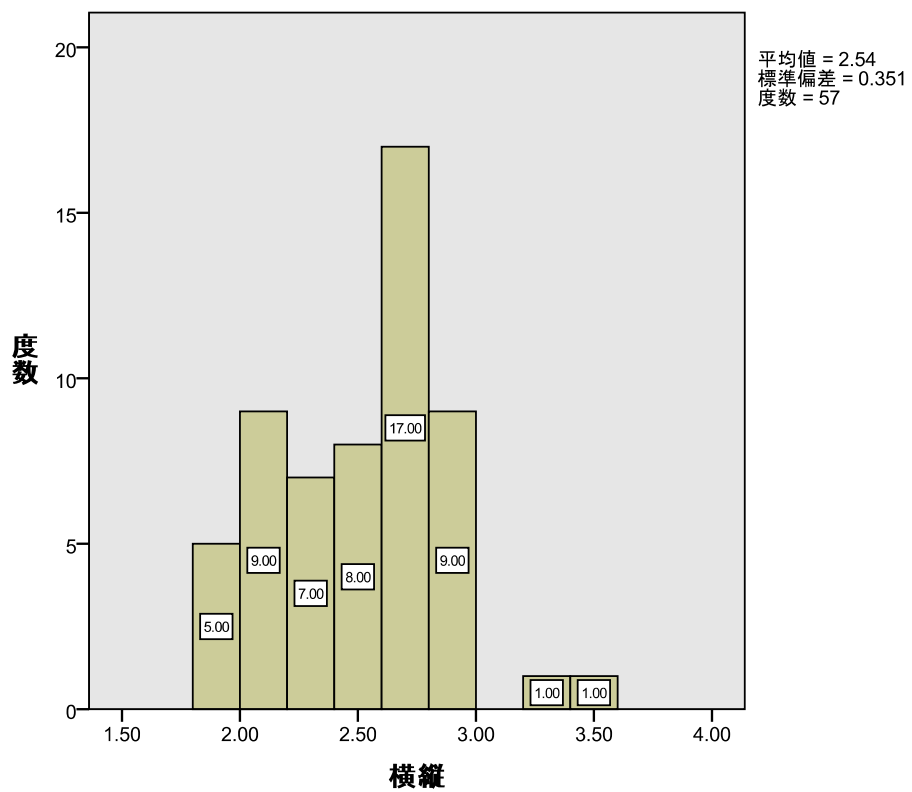
ヒストグラムとは、計量した特性の分布の形を把握するための度数分布のグラフ表示のことを言います。データの存在している範囲を幾つかの区間に分けて、各区間に入るデータの出現の頻度をカウントして度数表を作成しグラフ化したものです。

4.1 AKB

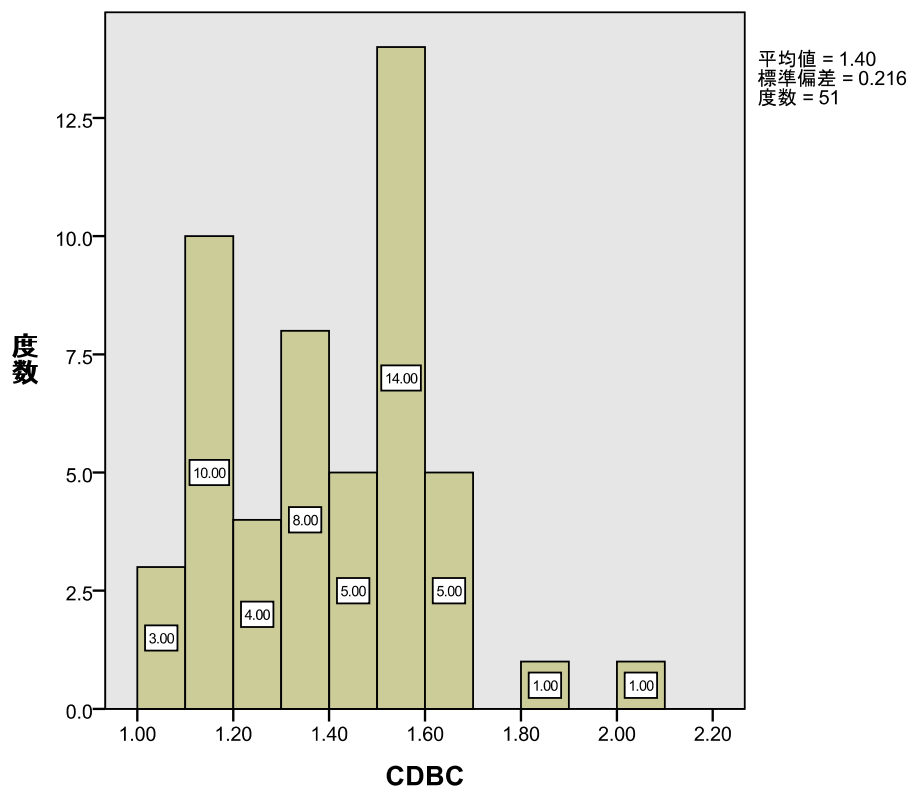
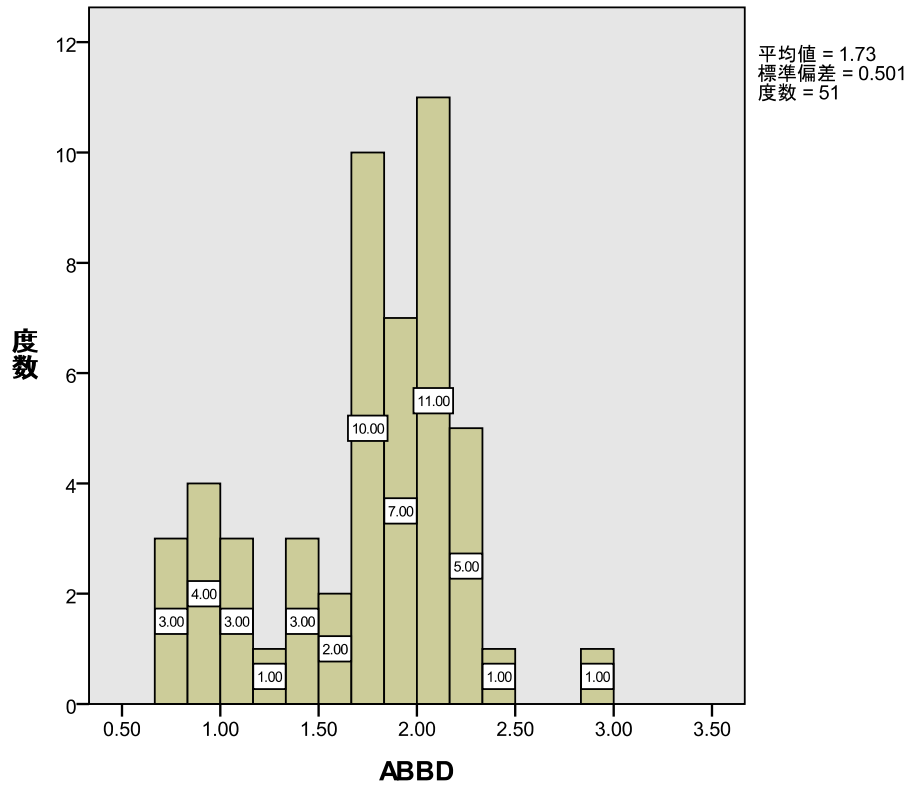


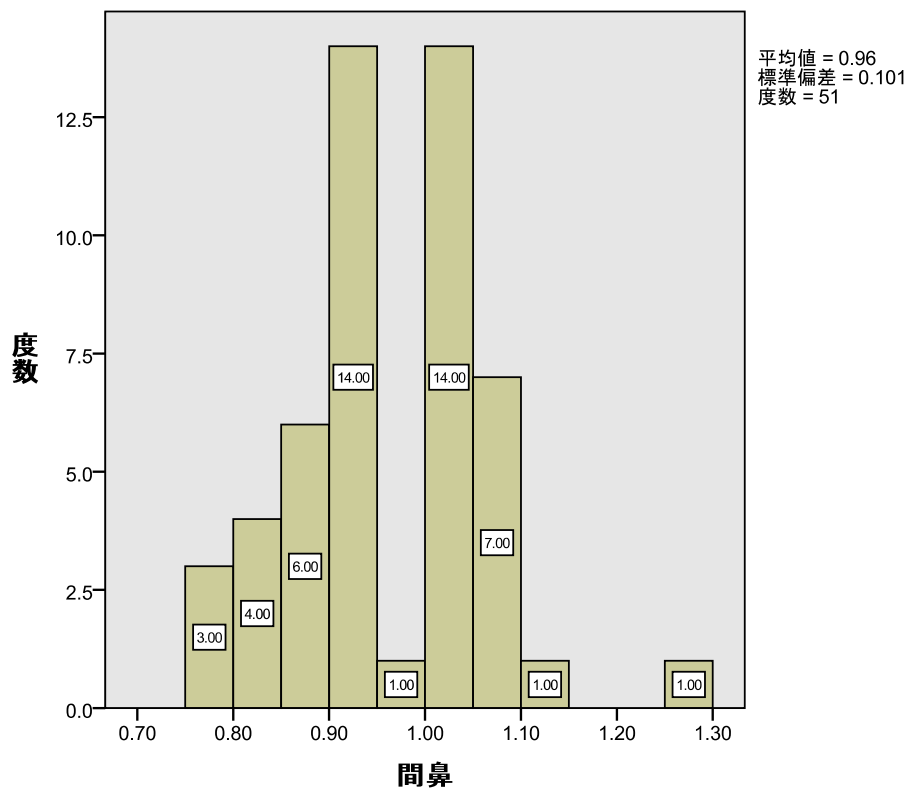
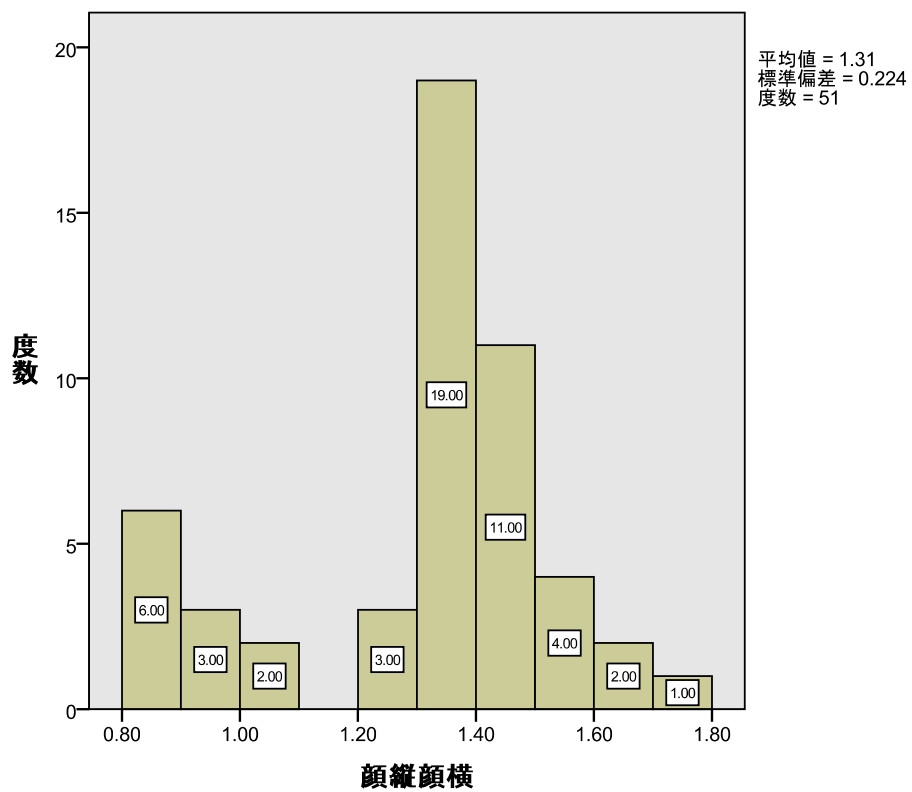


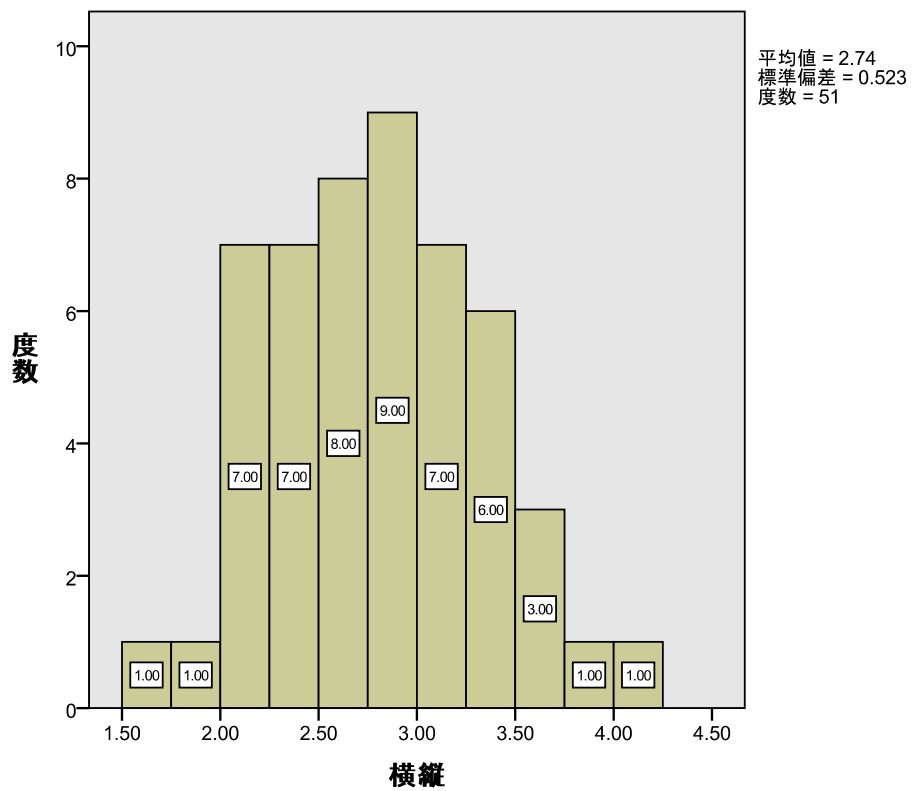
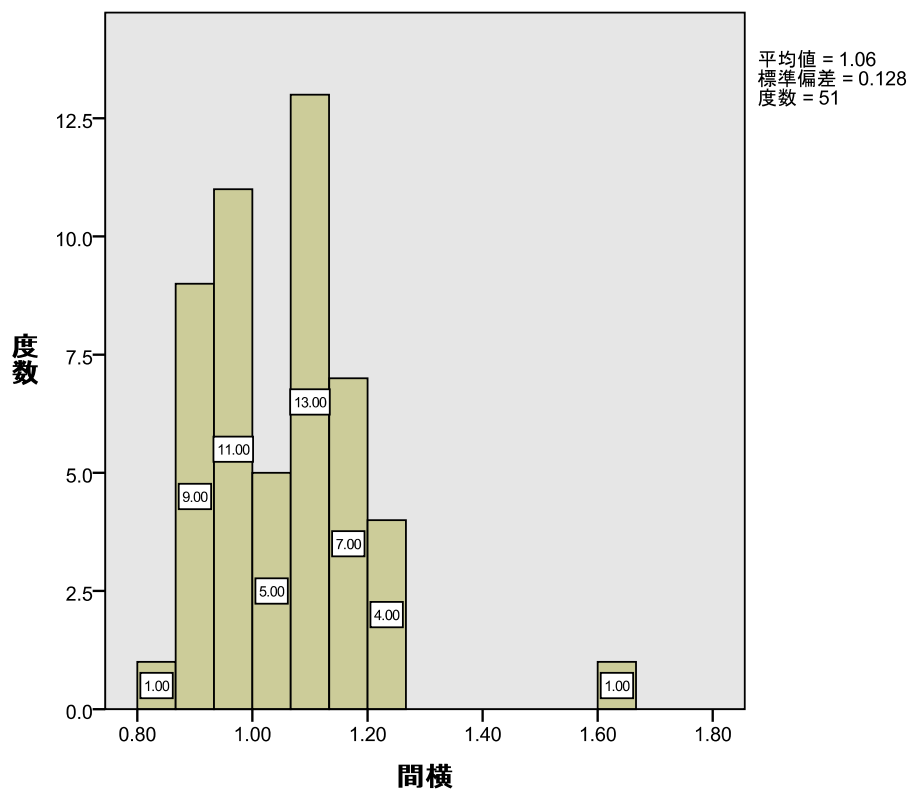


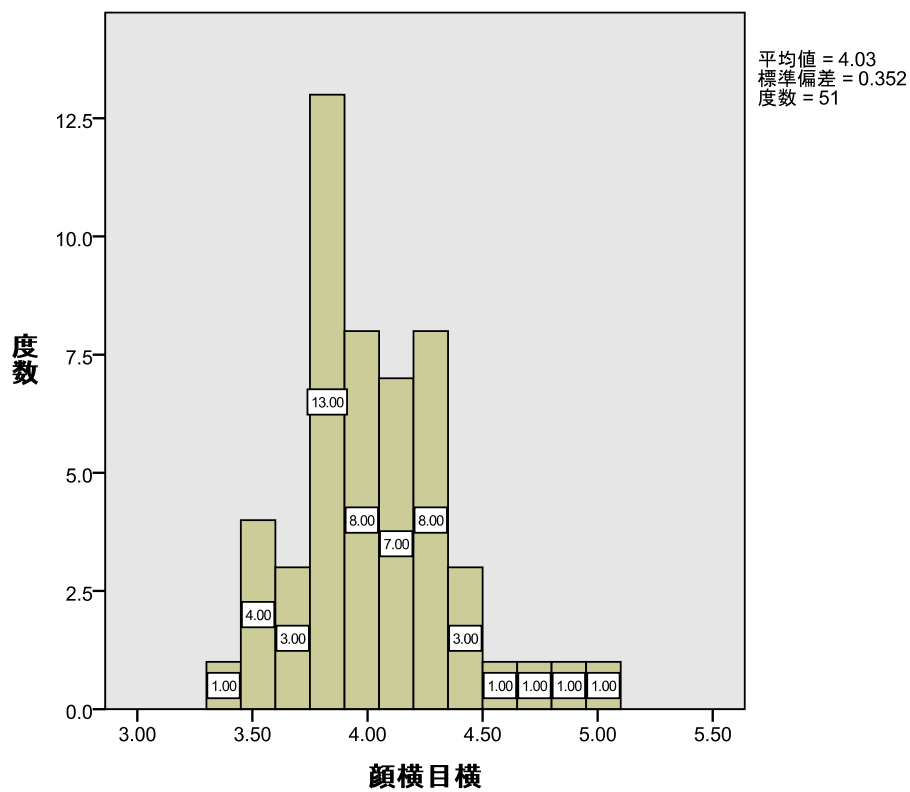


4.2 女優・タレント









4.3 分析結果

平均で見ると黄金比に近づいているもの、理想的な比率になっているものもあるが実際分布として見るとそのようではないことがわかる。

特に女優・タレントでは顔の大きさの比率（ $\frac{\text{顔の縦}}{\text{顔の横}}$ ）に関して白銀比付近がかなり多く出現している。つまり、小顔に関して適切な大きさとは白銀比付近に近づけることが理想的なのではないだろうか!?

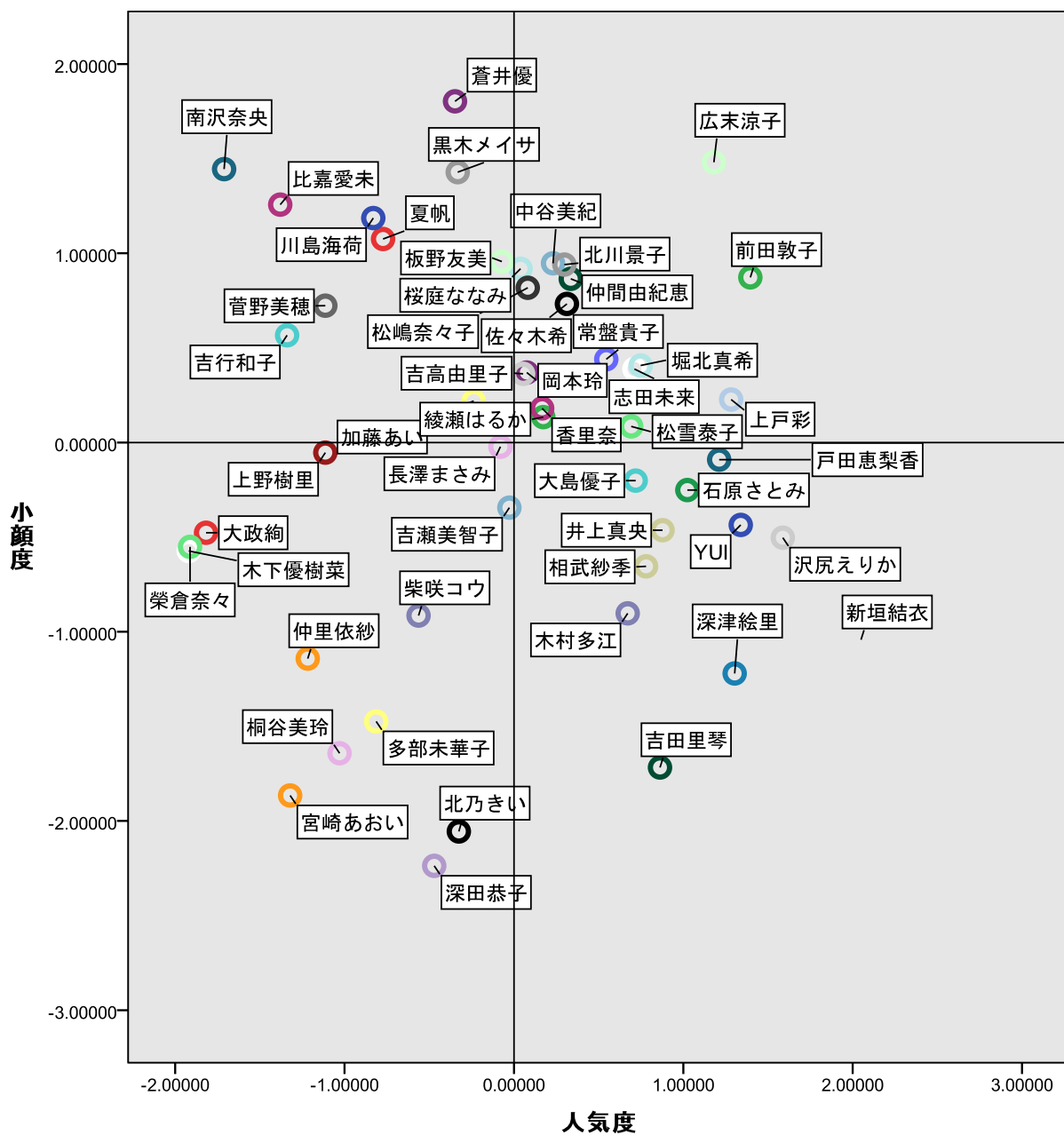
さらに、AKB に比べて女優・タレントの方が平均値が高い理由として女優・タレントの方が年齢層が上なので、成長仕切っているからである。

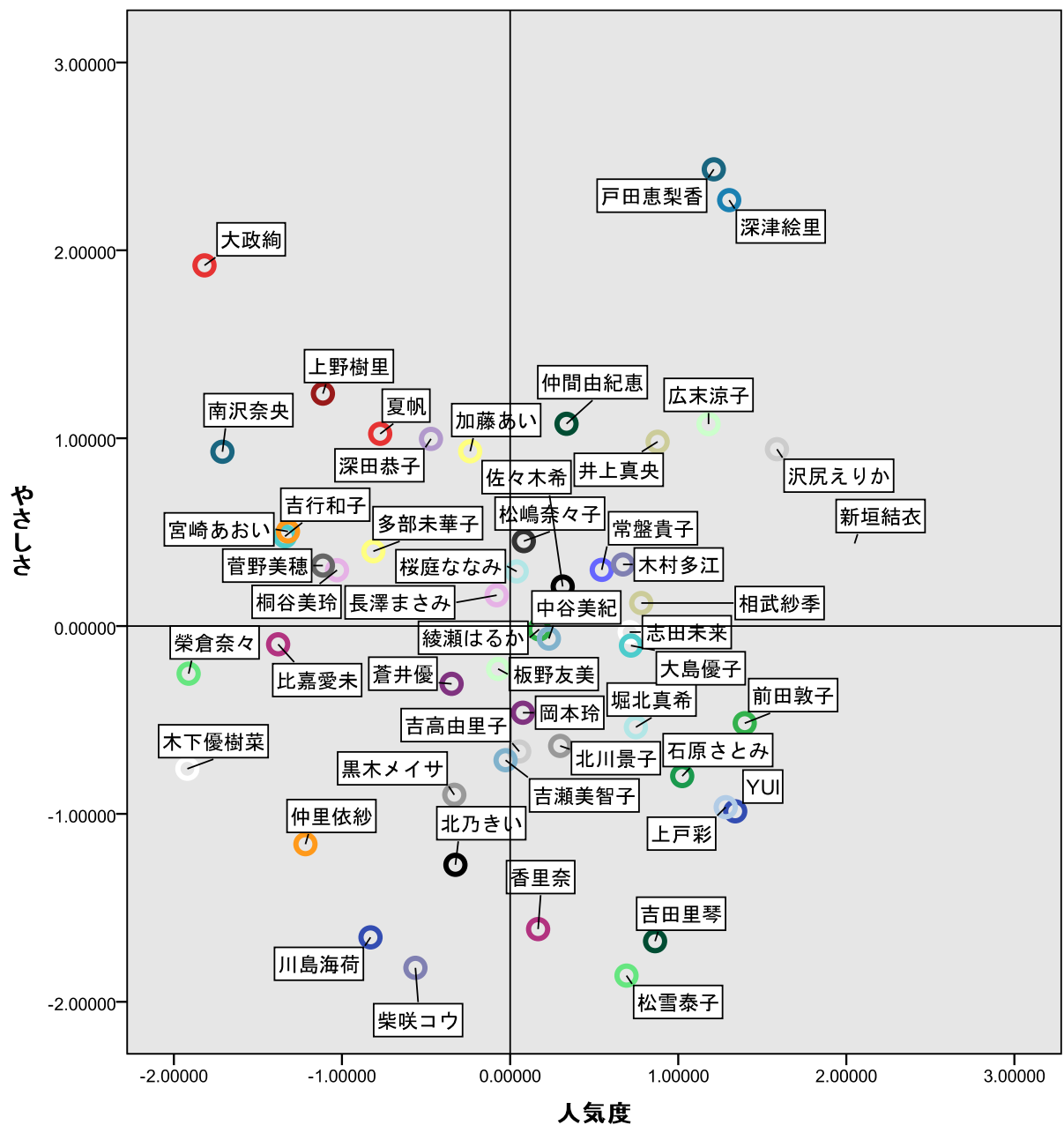
5 主成分分析

主成分分析は、経済学の分野などが発祥の統計手法。複数の変数間の共分散（相関）を少数の合成変数で説明する手法。

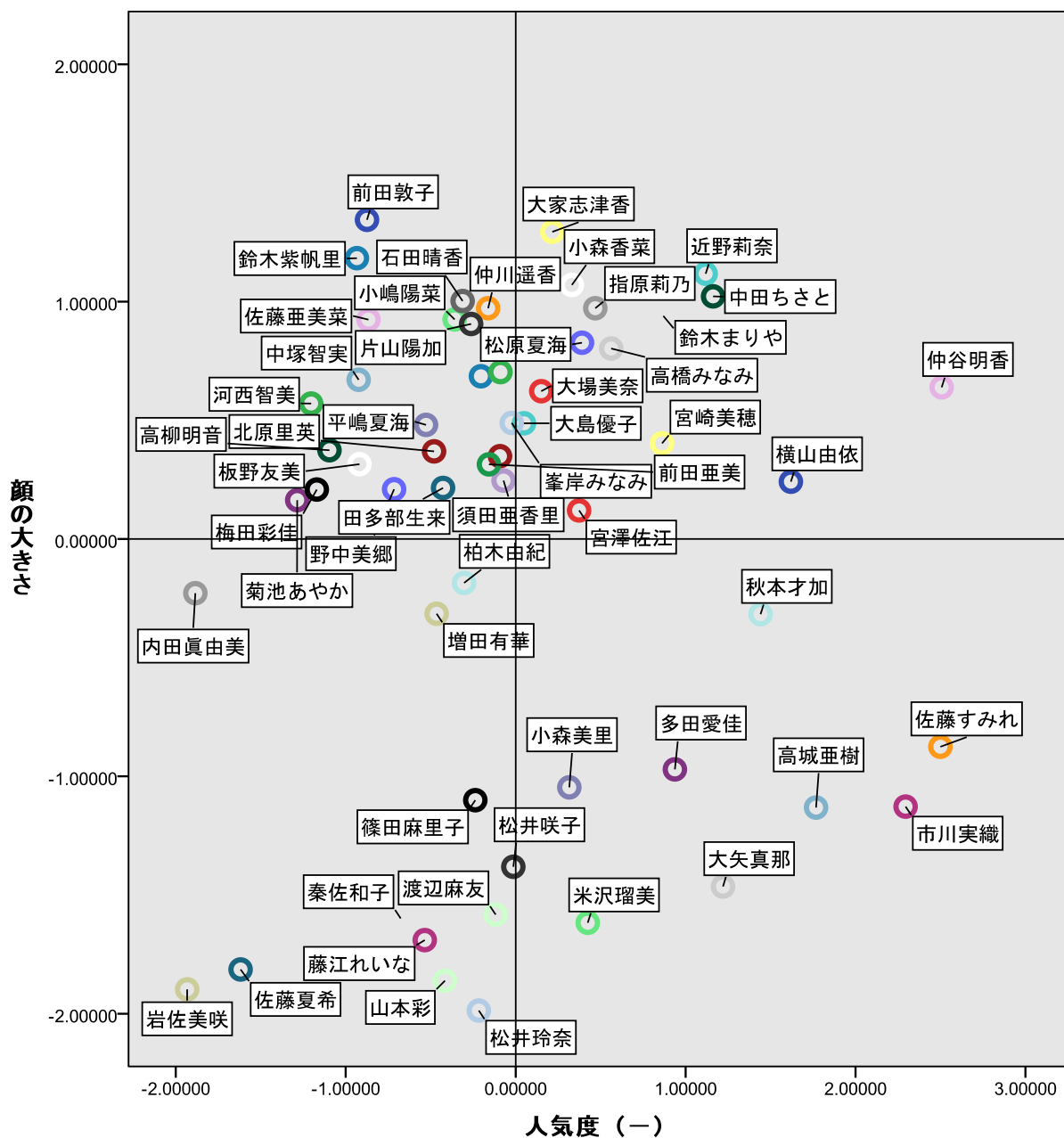
具体的に考えると、市町村の人口・学校数・商業施設数・進学率・第一次産業の就業率・水道の普及率などには通常高い相関がある。この相関を1つの変数が作った偽相関と仮定し、数学的に算出する。それを第一主成分と呼ぶ。第一主成分は数学的な計算結果に過ぎない。この意味を分析者が解釈し、たとえば「都市化指数」などと想定し、都市化の指標とする。その残差に対して同じ計算を適用して、主成分は第二、第三、と作れる。

5.1 女優・タレント





5.2 AKB



5.3 分析結果

女優・タレントは3主成分が生成された。それぞれ固有値, 主成分得点による散布図, 画像をもとに第1主成分を「人気度」, 第2主成分を「小顔度」, 第3主成分を「やさしい顔度」とした。注目すべきは第1主成分の「人気度」である。これは順位との相関係数が「約0.4」もある。なおここで言う順位というものは、私たちの主観ではなく第3者が客観的につけたものであるのでこの相関係数はかなりの信頼を置けるだろう。

AKB に関して見ると女優・タレントと同様に名前をつけた。第 1 主成分を「前田ポイント」、第 2 主成分を「人気度 (-)」とした。第 1 主成分について、これは主成分得点が前田敦子が 1 位であったためそのよう命名した。つまり、第 1 主成分の固有値を見ると、顔の大きさに関する値が大きくなっているからである。また、前田敦子が総選挙 No.1 につながっているのかもしれない。

6 因子分析

因子分析は、多変量解析の手法の 1 つで、心理学におけるパーソナリティの特性論的研究など、心理尺度の研究手法として使用される。モデル式の形状などから主成分分析と混同されることもあるが、主成分分析は観測データから合成スコアを構築することが目的であるのに対し、因子分析は観測データが合成量であると仮定し、個々の構成要素を得ようとするのが目的であり、両者は因果関係を異にする。適の例として「器用さ」の個人差の検討が考えられる。A,B,C の 3 人はそれぞれ「ジグソーパズル」「彫刻」「時計」の分解をある速度で器用にこなすことができるとしたときに A,B,C の器用さをどのように評価すればよいかを考える場合、3 人が 3 つのテストにかかった時間に対して因子分析を適用することで、3 つの課題に共通する潜在的な「器用さ」の導出を試みることができる。

また、因子分析は因子の回転が可能である。この回転というのは、「因子の解釈を容易にする」ものである。例えば中華料理店のテーブルの上に料理(要素)が並べてあり、席には飲食者(因子)がいるとする。そして料理(要素)をとるために飲食者(因子)の近くにテーブルを「回転」させ自分の元に近づける。つまり要素と因子の解釈を容易にすることになるイメージである。

6.1 分析結果

女優・タレントの場合

回転後の因子行列 a

	因子 (1)	因子 (2)	因子 (3)
ABBD	-0.090	0.897	-0.50
CDBC	0.034	0.045	0.811
顔縦顔横	-0.310	0.698	-0.017
間鼻	0.671	-0.187	-0.391
間横	0.771	-0.290	-0.129
横縦	-0.244	-0.130	0.548
顔横目横	0.557	-0.064	0.015

3 つの共通因子が生成された。第 1 因子「目力」、第 2 因子「顔の大きさ」、第 3 因子「大人な顔」と命名した。つまり、女優・タレントの顔の印象として共通的にわかることは、「目力」「顔の大きさ」「大人な顔」である。特に「大人な顔」とは意外にも「口の位置」と「目の大きさ」に依存していると見れる。

AKB の場合

回転後の因子行列 a

	因子 (1)	因子 (2)	因子 (3)	因子 (4)
ABBD	0.872	-0.113	0.180	-0.068
CDBC	-0.008	-0.028	-0.001	0.626
顔縦顔横	0.978	0.029	-0.028	0.039
間鼻	-0.076	0.821	-0.075	-0.132
間横	0.016	0.799	0.530	0.218
横縦	-0.165	0.116	-0.615	0.218
顔横目横	-0.027	0.278	0.699	0.227

4つの共通因子が生成された。第1因子「顔の大きさ」、第2因子「はっきりした顔」、第3因子「やさしい目(印象)」、第4因子「童顔度」と命名した。つまり、AKBの顔の印象として共通的にわかることは、「顔の大きさ」「はっきりした顔」「やさしい目(印象)」「童顔度」である。特に「童顔度」を見たときに「口の位置」「目の大きさ」と「目の位置」に依存していると思われる。

7 クラスタ分析

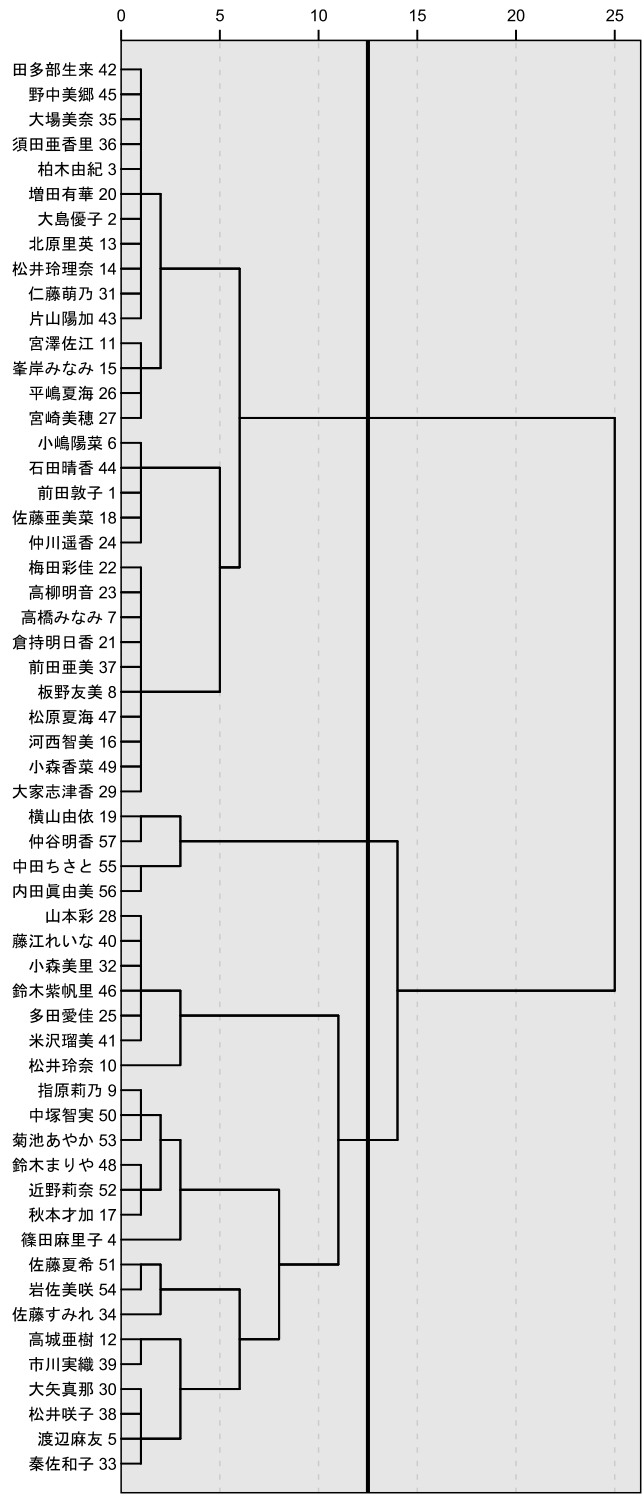
クラスタ分析は、似通った個体あるいは変数のグループ化を行うための分析手法である。

例えば、高校のテストの点数で、文系科目が得意な生徒と、理系科目が得意な生徒をグルーピングできる。つまり、クラスタ分析は対象間の距離を定義して、距離の近さによって対象を分類する方法の総称である。なのでサンプルの値があれば「似た能力を持った生徒をグルーピングできないか」「いくつかのグループに分けることができるのか」「あるグループにはどのような特徴をもったものが集まるのか」「グループ間の違いは何か」を検討することができる。ここではもっとも明確なクラスタを作る(分類感度が高い)ウォード法を用いる。

7.1 AKB

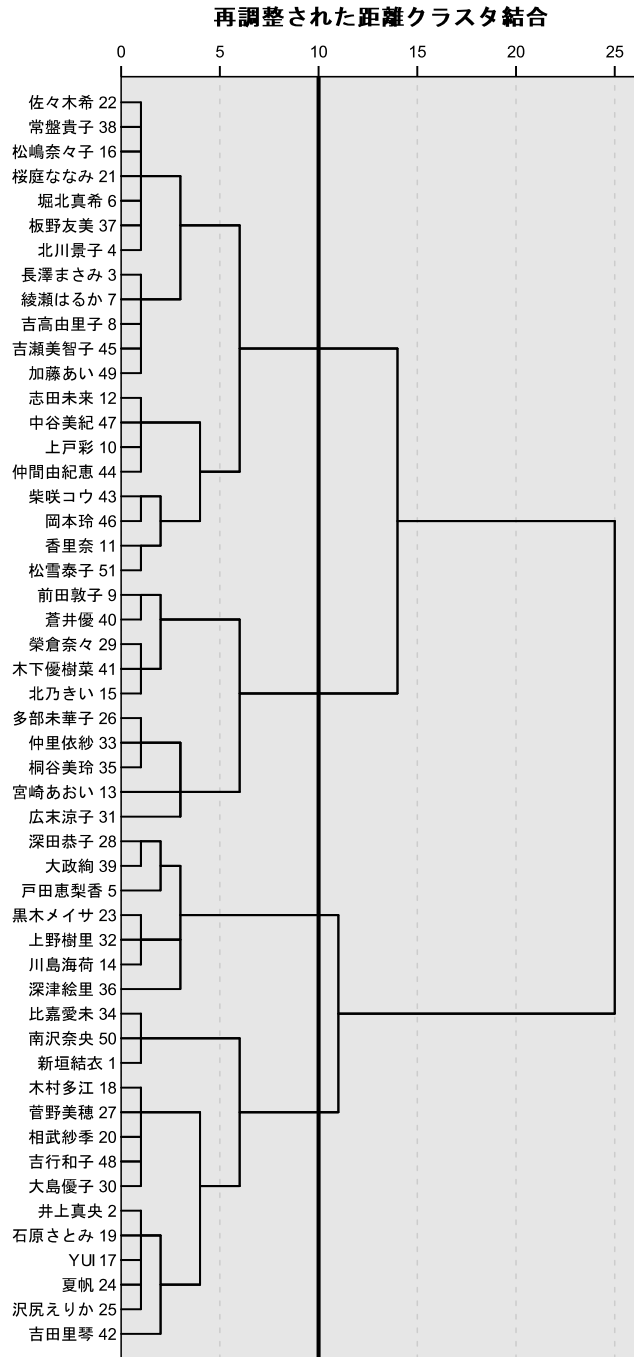
Ward 法を使用するデンドログラム

再調整された距離クラスタ結合



7.2 女優・タレント

Ward 法を使用するデンドログラム



7.3 分析結果

クラスター分析は横軸の距離の値によってグループ分けをする手法である。

AKB では、12.5 という値でグループ全体を 4 つに分けて見ることにした。それぞれのグループに画像、データの傾向から名前をつけた。

第 1 グループ「普通系カワイイ」は AKB の図より、「田多部生来」から「大家志津香」までの大きなグループである。特徴としては一般的である。

第 2 グループ「ぼっちゃり系カワイイ」は、「横山由依」から「内田真由美」までの少数派である。特徴はそのままである (画像より)。

第 3 グループ「お嬢様系カワイイ」は、「山本彩」から「松井玲奈」の少数派。特徴はお嬢様気質な感じがしたので (画像より)。

第 4 グループ「小顔系カワイイ」は、「指原莉乃」から「秦佐和子」までである。特徴としては前髪がそろっている (パツツン状態)。髪型からの印象としては「真のカワイイ系」といえるだろう。

女優・タレントでは 10.0 という値でグループ全体を 4 つに分けた見ることにした。AKB と同じく名前をつけた。

第 1 グループ「バランス系カワイイ」は女優・タレントの図より、「佐々木希」から「松雪泰子」までのグループである。特徴としては万能でそしてサイコー (制作者より) である。

第 2 グループ「妹系カワイイ」は、「前田敦子」から「広末涼子」までである。特徴は全体的に顔の印象が幼い感じである (画像より)。

第 3 グループ「キレイ系カワイイ」は、「深田恭子」から「深津絵里」の少数派。特徴として「真のキレイ系」といってもいいだろう (画像より)。

第 4 グループ「お姉さん系カワイイ」は、「比嘉愛未」から「吉田里琴」までである。特徴としては、清楚で大人の女な感じである。(注; オネエ系ではありません)

8 判別分析

判別分析 (はんべつぶんせき) は、事前に与えられているデータが異なるグループに分かれることが明らかかな場合、新しいデータが得られた際に、どちらのグループに入るのかを判別するための基準 (判別関数) を得るための手法。

「健常者」と「ある疾患にかかっている患者」に対する検査値があったとする。この値より判別方程式というものをつくり、ある人の検査値を代入して「正の値になれば健常者」、「負の値になれば患者」と判別できる。

ここでは時間の関係上女優・タレントのみを判別分析を用いて、判別することにした。判別する項目として、グループを容姿だけを見て「カワイイ系」「キレイ系」のグループに分けた。「カワイイ系」を「1」、「キレイ系」を「2」としたとき以下の図のように分けることができた。

予想グループ番号	
カワイイ系 (1)	キレイ系 (2)
新垣結衣	戸田恵梨香
井上真央	吉高由里子
長澤まさみ	香里奈
堀北真希	松嶋奈々子
綾瀬はるか	YUI
前田敦子	木村多江系
上戸彩	黒木メイサ
志田未来	深田恭子
北乃きい	桐谷美玲
石原さとみ	比嘉愛未
相武紗季	深津絵里
桜庭ななみ	板野友美
佐々木希	常磐貴子
夏帆	大政絢
多部未華子	柴咲コウ
菅野美穂	仲間由紀恵
榮倉奈々	吉瀬美智子
大島優子	中谷美紀
広末涼子	加藤あい
仲里依紗	松雪泰子
上野樹里	北川景子
蒼井優	
木下優樹菜	
吉田里琴	
岡本玲	
吉行和子	
南沢奈央	
川島海荷	
宮崎あおい	

8.1 分析結果

判別変数は以下のとおりである.

$ABBD, CDBC$, 顔縦顔横, 間鼻, 間横, 横縦, 顔横目横

ここでの判別関数式は $z = (-0.84 * ABBD) + (-3.2720 * CDBC) + (3.247 * \text{顔縦顔横}) + (0.916 * \text{間鼻}) + (-4.447 * \text{間横}) + (0.359 * \text{横縦}) + (2.251 * \text{顔横目横}) + (-3.786 * \text{定数})$ というもので決定できる.

判別後

	カワイイ系 (1)	キレイ系 (2)	合計
カワイイ系 (1)	20 人	9 人	29 人
キレイ系 (2)	5 人	17 人	22 人
合計	25 人	26 人	51 人

表: 判別表

元のグループ化されたケースのうち, 72.5 % が正しく分類された. つまり全体の 72.5 % はそのままの判断基準でよい.

判別後	
2 から 1 に判別	1 から 2 に判別
黒木メイサ	井上真央
北川景子	上戸彩
板野友美	志田未来
常盤貴子	石原さとみ
柴咲コウ	夏帆
	上野樹里
	吉田里琴
	南沢奈央

参考文献

- [1] 喜田 安哲:データ分析と SPSS(2), 2003 年, 北樹出版
- [2] 牟田 淳:デザインのための数学, 2010 年, オーム社
- [3] 永田 靖:多変量解析入門, 2001 年, サイエンス社
- [4] 石村 貞夫:SPSS でやさしく学ぶ多変量解析, 1999 年, 東京図書