

多重対応分析と 幾何学的データ解析

ver2.0 2025/09/04

ver1.0 2023/09/06

計量分析セミナー

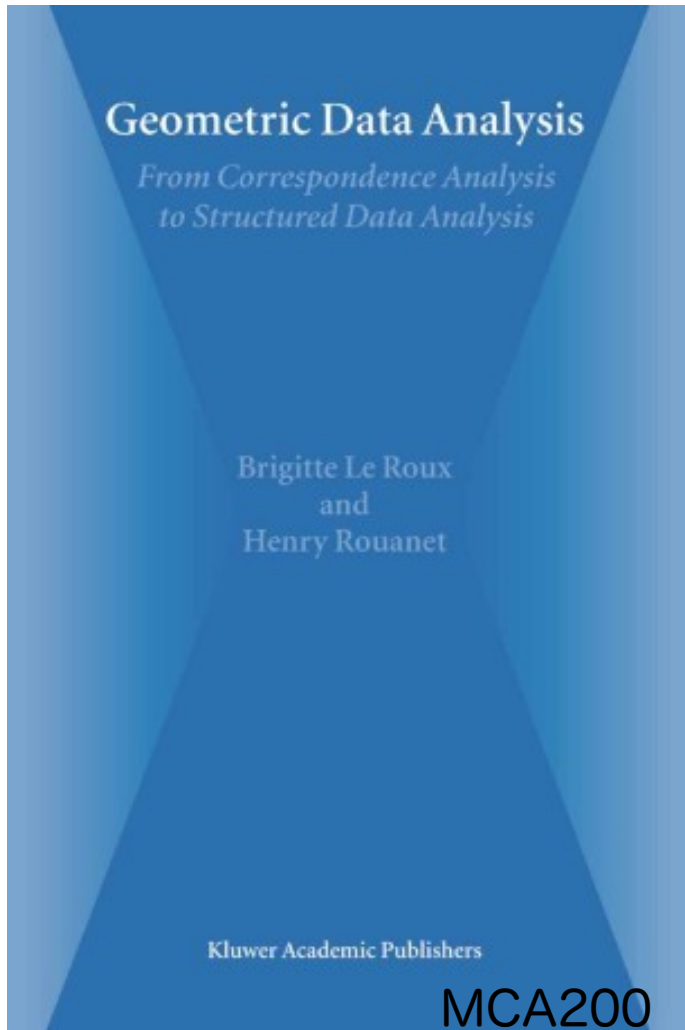
藤本一男

kazuo.fujimoto2007@gmail.com

幾何学的データ解析（GDA）

- MCAを中心においた分析フレームワーク（広義のMCA）
- 実験計画が不可能な、調査データから以下に内部構造（関係性）を抽出できるか。
 - MCAによる空間生成
 - 追加変数による空間分析（構造化モデリング）
 - 記述によって明らかになった差異についての検定（帰納的データ解析）
 - 典型性検定（1標本のt-検定）
 - 同質性検定（2標本のt-検定）

元の本GDA2004の圧縮版



- <https://link.springer.com/book/10.1007/1-4020-2236-0> PDFで読めます。



- <https://helios2.mi.parisdescartes.fr/~lerb/livres/Books.html> では、2004
 - 2005だったり、2010だったりする。
 - でも、Le Roux 先生たちは、2004にしている



MCA2010

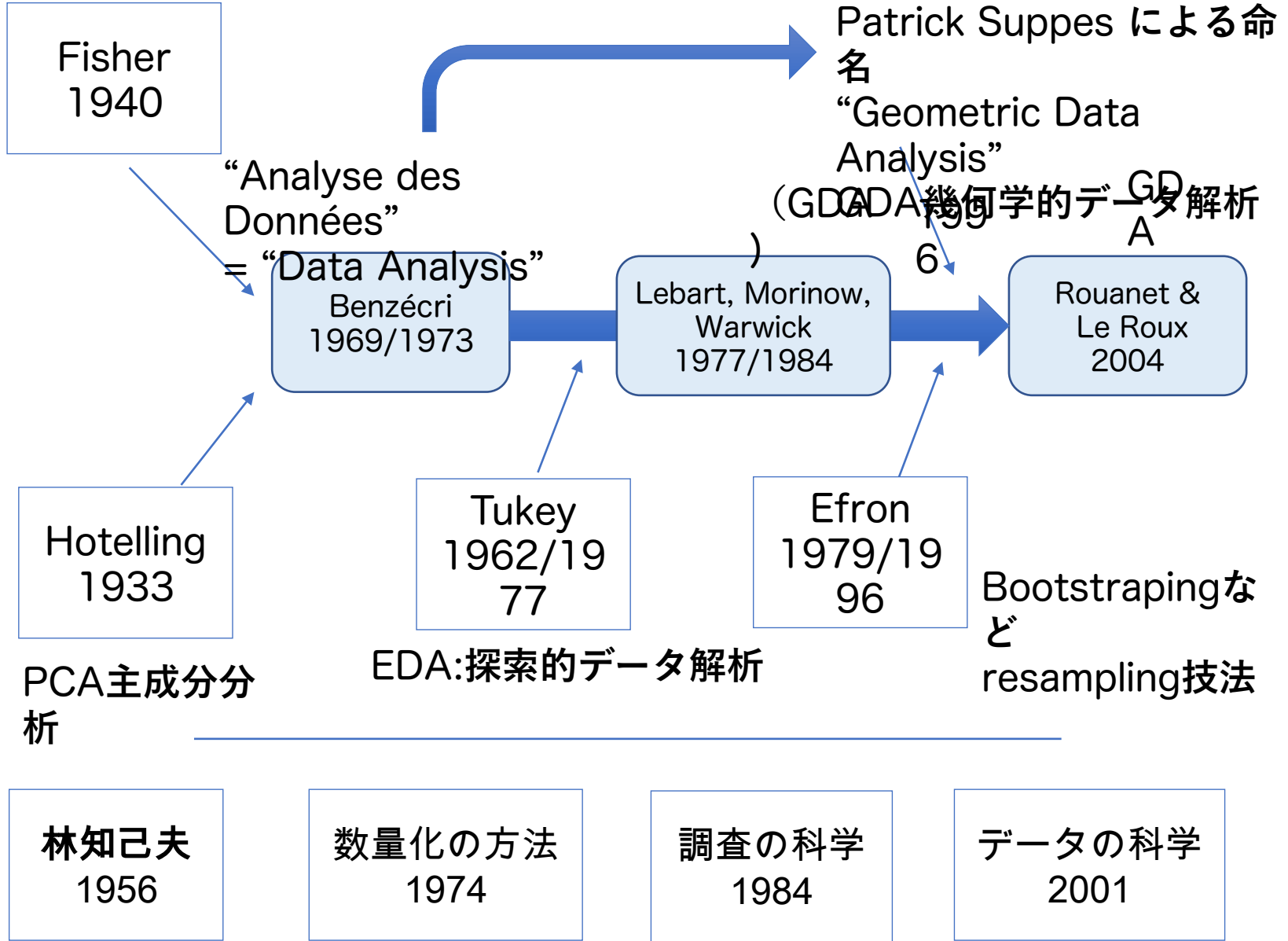


MCA2021

構造化データ解析とANOVA、回帰分析

- MCAp95「伝統的な統計学においても、分散分析ANOVA（およびそれを拡張した多変量分散分析MANOVA）や回帰分析などのいくつかの手法で構造化因子を扱ってきた。こうした手法を幾何学的データ解析に取り入れて統合化することを構造化データ解析（SDA）と呼ぼう。」
- BlueBook 2004のSDAのまとめp268
「GDAのユーザーの中には、個体×変数の表を前に、PCAやMCA をそのまま行い、従来のANOVAや回帰分析を並べ、別々に実施し 解釈する人もいる。構造化データ解析では、データの幾何学的 モデルから始めて、観測データの特別な特性(特に非直交性)を 必要に応じて考慮しながら、ANOVAや回帰手順を「接ぎ木」していく。この方法では、基本的な幾何学的モデル(例えば距離 の定義を変更する)を修正する必要はない。したがって、新しい 構造に出会うたびに「新しい方法を発明する」必要はなく、データの分析がより豊かになる。つまり、固定化された手法の硬直性と分析手順のアドホックな拡散の両方を避けることができるのである。」

最適化尺度



津田塾大学紀要
54号
(2022/3)
用に作成

Analyse des Données からGDAへ

- 1960年代 JPベンゼクリがAnalyse des Données を提唱。直訳すると、Data Analysis、つまり「データ解析」
- これに、構造化データ解析（SDA）や帰納的推論（帰納的データ解析（IDA））を加えて、幾何学的データ解析（GDA）と呼ぶようになった。
 - GDAの命名は、スタンフォード大学のPatrick Suppesによる。原注*1
- ただ、GDAという名称は使用されていないものの、内容的には、Lebart, Morinow, Warwick 1977(仏語)/1984(英語)(日本語訳は1994に大隅らによる『記述的多変量解析』)で展開されている。
 - La Distinction の5章、注2で参照されているのがこれ。
- 林知己夫「データの科学」と同じ発想、アプローチ。

GDAの主要なアイデア、 3つの理論的枠組み

- 三つのアイデア
 - 幾何学的解釈
 - 定式的なアプローチ
 - 記述的であること
(以上の指摘は、CAiP3へのGreenacreの日本語版への序にもある。)
- 3つの理論的枠組み
 - 対応分析
 - 主成分分析
 - 多重対応分析
- 個体×変数
 - 変数カテゴリは、モダリティとも呼ばれる

GDAのステップ

- 構造設計
 - 空間生成する変数：アクティブ変数
 - 空間生成に寄与せず、射影する変数：追加変数
- MCAによる基本分析
 - 変数空間の分析から各座標軸を命名する（新たな変数名）
 - 軸を生成している変数カテゴリを確認
 - 変数空間でのカテゴリの関係の確認
- 個体空間の構造を追加変数を用いて分析する（構造化データ解析：SDA）
- 記述で見えた差異の優位性を検定する（帰納的データ解析：IDA）

『多重対応分析』のデータで例示

- 原著のサポートサイトにあるExcelのデータ
 - https://helios2.mi.parisdescartes.fr/~lerb/Logiciels/Data/Taste_Example.xls
 - これを日本語化したものを使います。ファイルで提供。
- MCAツールは、GDAtools::speMCA
 - <https://cran.r-project.org/web/packages/GDAtools/index.html>
 - <https://cran.r-project.org/web/packages/GDAtools/GDAtools.pdf>
 - https://cran.r-project.org/web/packages/GDAtools/vignettes/GDA_tutorial.pdf
 - <https://github.com/nicolas-robette/GDAtools>

基本的MCA分析

MCA事例:嗜好データ (taste_J*)

	ID	Isup	TV	Film	Art	Eat	Gender	Age	Income
1	1	Active	TV-メロドラマ	映画-アクション	芸術-風景画	外食先-ステーキハウス	女性	55-64	£20-29
2	2	Active	TV-メロドラマ	映画-ホラー	芸術-静物画	外食先-インド料理店	女性	45-54	<£9
3	3	Active	TV-自然	映画-アクション	芸術-風景画	外食先-パブ	女性	55-64	<£9
4	4	Active	TV-メロドラマ	映画-時代劇	芸術-肖像画	外食先-イタリア料理店	女性	65+	£10-19
5	5	Active	TV-コメディ	映画-ホラー	芸術-静物画	外食先-インド料理店	女性	35-44	£10-19
6	6	Active	TV-コメディ	映画-ホラー	芸術-印象派	外食先-インド料理店	女性	18-24	<£9
7	7	Active	TV-ニュース	映画-アクション	芸術-風景画	外食先-インド料理店	女性	25-34	£10-19
8	8	Active	TV-ニュース	映画-ドキュメンタリー	芸術-パフォーマンス・アート	外食先-フィッシュ&チップス	男性	65+	£10-19
9	9	Active	TV-メロドラマ	映画-時代劇	芸術-風景画	外食先-ステーキハウス	女性	65+	<£9
10	10	Active	TV-ニュース	映画-アクション	芸術-風景画	外食先-フィッシュ&チップス	女性	65+	£10-19

Showing 1 to 10 of 1,253 entries

Previous 1 2 3 4 5 ... 126 Next

行：回答者
列：回答設問

*このデータは、
LeRoux&Rouanet2010=2021
で使われているデータを日本語
化したもの。
[https://
helios2.mi.parisdescartes
.fr/~lerb/Logiciels/
Data/Taste_Example.xls](https://helios2.mi.parisdescartes.fr/~lerb/Logiciels/Data/Taste_Example.xls)

指示行列化したもの：データとしては等価

変数TVの回答カテゴリ

変数映画の回答カテゴリ

変数芸術の
回答カテゴリ

変数TVの回答カテゴリ											変数映画の回答カテゴリ									変数芸術の 回答カテゴリ	
Isup.Active	Isup.supp	TV.TV- コメデ ィー	TV.TV- ドラマ	TV.TV- 映画	TV.TV- 自然	TV.TV- ニュー ス	TV.TV- 警察も の	TV.TV- メロド ラマ	TV.TV- スポー ツ		Film. 映 画- アク ショ ン	Film. 映 画- コメ ディ	Film. 映 画- 時代 劇	映 画- ドキ ュメ ンタ リー	Film. 映 画- ホラ ー	Film. 映 画- ミュ ージ カル	Film. 映 画- ロマ ンス	Film. 映 画- SF	Art. 芸 術- 印象 派		
1	1	0	0	0	0	0	0	1	0		1	0	0	0	0	0	0	0	0		
2	1	0	0	0	0	0	0	1	0		0	0	0	0	1	0	0	0	0		
3	1	0	0	0	1	0	0	0	0		1	0	0	0	0	0	0	0	0		
4	1	0	0	0	0	0	0	1	0		0	0	1	0	0	0	0	0	0		
5	1	0	1	0	0	0	0	0	0		0	0	0	0	1	0	0	0	0		
6	1	0	1	0	0	0	0	0	0		0	0	0	0	1	0	0	0	1		
7	1	0	0	0	0	1	0	0	0		1	0	0	0	0	0	0	0	0		
8	1	0	0	0	0	1	0	0	0		0	0	0	1	0	0	0	0	0		
9	1	0	0	0	0	0	0	1	0		0	0	1	0	0	0	0	0	0		
10	1	0	0	0	0	1	0	0	0		1	0	0	0	0	0	0	0	0		
Showing 1 to 10 of 1,215 entries		Previous 1 2 3 4 5 ... 122 Next																			

46カテ
ゴ
リ

1215
行

MCAによる空間生成

次元縮減で生成された空間の座標軸に分解

データ表がもっていた分散

ID	Isup	TV	Film	Art	Eat	Gender	Age	Income	
1	1	Active	TV-メロドラマ	映画-アクション	芸術-風景画	外食先-ステーキハウス	女性	55-64	£20-29
2	2	Active	TV-メロドラマ	映画-ホラー	芸術-静物画	外食先-インド料理店	女性	45-54	<£9
3	3	Active	TV-自然	映画-アクション	芸術-風景画	外食先-パブ	女性	55-64	<£9
4	4	Active	TV-メロドラマ	映画-時代劇	芸術-肖像画	外食先-イタリア料理店	女性	65+	£10-19
5	5	Active	TV-コメディ	映画-ホラー	芸術-静物画	外食先-インド料理店	女性	35-44	£10-19
6	6	Active	TV-コメディ	映画-ホラー	芸術-印象派	外食先-インド料理店	女性	18-24	<£9
7	7	Active	TV-ニュース	映画-アクション	芸術-風景画	外食先-インド料理店	女性	25-34	£10-19
8	8	Active	TV-ニュース	映画-ドキュメンタリー	芸術-パフォーマンス・アート	外食先-フィッシュ&チップス	男性	65+	£10-19
9	9	Active	TV-メロドラマ	映画-時代劇	芸術-風景画	外食先-ステーキハウス	女性	65+	<£9
10	10	Active	TV-ニュース	映画-アクション	芸術-風景画	外食先-フィッシュ&チップス	女性	65+	£10-19

Showing 1 to 10 of 1,253 entries

Previous12345...126Next

	固有値	分散率	累積分散率
dim 1	0.4003554	6.405686	6.405686
dim 2	0.3511658	5.618653	12.024339
dim 3	0.3250090	5.200145	17.224484
dim 4	0.3080672	4.929075	22.153558
dim 5	0.2988670	4.781872	26.935430
dim 6	0.2875943	4.601508	31.536938
dim 7	0.2782293	4.451669	35.988607
dim 8	0.2738541	4.381666	40.370273
dim 9	0.2682125	4.291399	44.661673
dim 10	0.2600118	4.160189	48.821862
dim 11	0.2581629	4.130607	52.952469
dim 12	0.2512422	4.019875	56.972344
dim 13	0.2474271	3.958833	60.931177
dim 14	0.2422577	3.876124	64.807301
dim 15	0.2349924	3.759879	68.567179
dim 16	0.2319159	3.710654	72.277834
dim 17	0.2244237	3.590780	75.868613
dim 18	0.2136934	3.419095	79.287708
dim 19	0.2091020	3.345633	82.633341
dim 20	0.2035845	3.257352	85.890693
dim 21	0.1939230	3.102768	88.993461
dim 22	0.1880943	3.009508	92.002969
dim 23	0.1829606	2.927369	94.930338
dim 24	0.1649057	2.638492	97.568830
dim 25	0.1519482	2.431170	100.000000

変数空間

その座標軸をもとに
二つの空間が生成
される。

個体空間

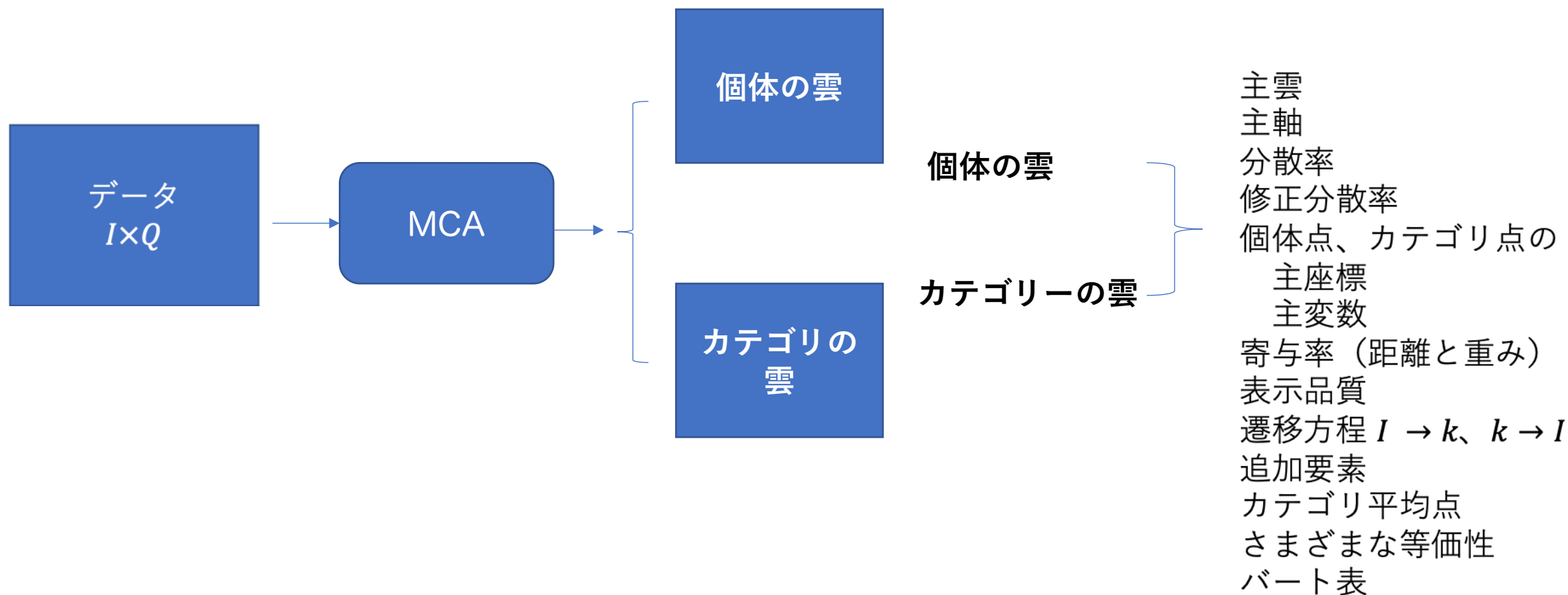
1215 x 4 行列

4 変数 = カテゴリ数29
(8+8+7+6)

29-1次元までとられる

3.1 MCAの原理

データ、MCA、基本統計量、解釈



「変数」空間の生成

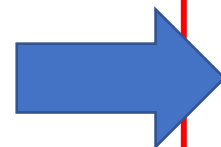
ID	Isup	TV	Film	Art	Eat	Gender	Age	Income	
1	1	Active	TV-メロドラマ	映画-アクション	芸術-風景画	外食先-ステーキハウス	女性	55-64	£20-29
2	2	Active	TV-メロドラマ	映画-ホラー	芸術-静物画	外食先-インド料理店	女性	45-54	<£9
3	3	Active	TV-自然	映画-アクション	芸術-風景画	外食先-パブ	女性	55-64	<£9
4	4	Active	TV-メロドラマ	映画-時代劇	芸術-肖像画	外食先-イタリア料理店	女性	65+	£10-19
5	5	Active	TV-コメディ	映画-ホラー	芸術-静物画	外食先-インド料理店	女性	35-44	£10-19
6	6	Active	TV-コメディ	映画-ホラー	芸術-印象派	外食先-インド料理店	女性	18-24	<£9
7	7	Active	TV-ニュース	映画-アクション	芸術-風景画	外食先-インド料理店	女性	25-34	£10-19
8	8	Active	TV-ニュース	映画-ドキュメンタリー	芸術-パフォーマンス・アート	外食先-フィッシュ&チップス	男性	65+	£10-19
9	9	Active	TV-メロドラマ	映画-時代劇	芸術-風景画	外食先-ステーキハウス	女性	65+	<£9
10	10	Active	TV-ニュース	映画-アクション	芸術-風景画	外食先-フィッシュ&チップス	女性	65+	£10-19

Showing 1 to 10 of 1,253 entries

Previous

1

2345...126Next



Dim28まである

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
TV-コメディ	0.7879974	0.9601999	0.2552989	-0.9357013	-1.1215
TV-ドラマ	-0.4961301	0.0525183	0.9810580	-0.5714160	1.2427
TV-映画	0.5736479	0.6941642	-0.6062750	1.0401520	0.4511
TV-自然	-0.7751191	0.0988366	-0.2344062	1.2460357	-0.7244
TV-ニュース	-0.8810760	0.0028543	0.0867057	-0.2434420	-0.2797
TV-警察もの	0.1917016	-0.4047489	-0.4057086	-0.1879865	0.1150
TV-メロドラマ	0.8701470	-1.0948268	0.7065701	0.2470517	0.0652
TV-スポーツ	-0.0450841	0.1325673	-1.4689864	-0.6262140	0.7679
映画-アクション	-0.0704556	0.1268878	-0.6537612	-0.4409041	0.6459
映画-コメディ	0.7499473	0.3064859	0.3072429	-0.6009223	-0.9784
映画-時代劇	-1.3279042	0.0367022	1.2401304	-0.1279792	0.1746
映画-ドキュメンタリー	-1.0221864	-0.1924633	-0.5218485	0.5416278	-0.6103
映画-ホラー	1.0919312	0.9977497	-0.1028270	-0.0220692	0.2583
映画-ミュージカル	-0.1353606	-1.2859419	0.1092429	-0.1957597	-0.2529
映画-ロマンス	1.0338394	-1.2402387	1.2153113	0.8933137	0.5361
映画-SF	-0.2083792	0.6733172	-0.6455216	2.0263114	-0.3261
芸術-印象派	-0.5593787	0.9865129	0.8244203	0.2089799	0.3114
芸術-風景画	-0.2310199	-0.3902110	-0.3128153	-0.1487870	-0.2938
芸術-現代美術	0.9428482	0.9607713	0.2851469	0.1413773	0.2436
芸術-パフォーマンス・アート	0.0883572	0.0753910	0.0683117	-0.6109836	0.7054
芸術-肖像画	1.0199551	-0.5496364	0.1422850	0.6772620	0.6433

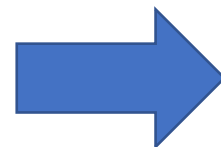
全変数カテゴリ分 (29行)

「個体」空間の生成

	ID	Isup	TV	Film	Art	Eat	Gender	Age	Income
1	1	Active	TV-メロドラマ	映画-アクション	芸術-風景画	外食先-ステーキハウス	女性	55-64	£20-29
2	2	Active	TV-メロドラマ	映画-ホラー	芸術-静物画	外食先-インド料理店	女性	45-54	<£9
3	3	Active	TV-自然	映画-アクション	芸術-風景画	外食先-バブ	女性	55-64	<£9
4	4	Active	TV-メロドラマ	映画-時代劇	芸術-肖像画	外食先-イタリア料理店	女性	65+	£10-19
5	5	Active	TV-コメディ	映画-ホラー	芸術-静物画	外食先-インド料理店	女性	35-44	£10-19
6	6	Active	TV-コメディ	映画-ホラー	芸術-印象派	外食先-インド料理店	女性	18-24	<£9
7	7	Active	TV-ニュース	映画-アクション	芸術-風景画	外食先-インド料理店	女性	25-34	£10-19
8	8	Active	TV-ニュース	映画-ドキュメンタリー	芸術-パフォーマンス・アート	外食先-フィッシュ&チップス	男性	65+	£10-19
9	9	Active	TV-メロドラマ	映画-時代劇	芸術-風景画	外食先-ステーキハウス	女性	65+	<£9
10	10	Active	TV-ニュース	映画-アクション	芸術-風景画	外食先-フィッシュ&チップス	女性	65+	£10-19

Showing 1 to 10 of 1,253 entries

Previous 1 2 3 4 5 ... 126 Next



Dim28まである

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
1	0.1353093622	-0.901984475	-0.4323046887	-0.14561108883	0.5252096484
2	1.2024365556	0.328563329	0.2638075182	0.29726561643	0.0827767400
3	-0.5370106529	-0.333733358	-0.5648763197	0.14990289649	-0.4385325875
4	0.2136401677	-0.451225657	1.1141370254	0.33419672137	0.4981295820
5	1.1699785032	1.195527931	0.0659147224	-0.23546950841	-0.4599627031
6	0.7225735345	1.416187118	0.3761584224	-0.25105762627	-0.2280237670
7	-0.2664249744	0.064106352	-0.4380746986	-0.28904712298	0.0573939729
8	-0.6140031079	-0.380673096	-0.2981257886	-0.09182242795	-0.1839411972
9	-0.3615201602	-0.940031539	0.3982101637	-0.00466359371	0.3096833838
10	-0.3641543297	-0.442372721	-0.5231054790	-0.32619161528	-0.0664244769
11	-0.0658544300	0.699634592	-0.6215774722	-0.30030998482	0.8133104785
12	0.5276892807	0.688774839	-0.8580612260	-0.33075949901	0.7822917479
13	-0.8204602995	0.100068250	1.0349343287	-0.40652582447	0.6080227847
14	0.0005180078	0.213883267	0.6258418116	-0.61954865941	0.0807135171
15	0.1574393936	-0.107851284	-0.6540096582	-0.26406890204	0.2379300882
16	0.5493860228	0.681003601	0.0968928854	-0.10184928126	0.1757035160
17	-1.2423837512	0.088551259	1.1842459654	-0.29713805806	0.2514118540
18	-0.7632544969	0.026059287	0.3924401538	-0.14809962786	-0.1581322917
19	-0.1402361351	0.260532265	-0.2709417632	-0.49722957773	0.5144053487
20	0.5132720478	1.469244066	0.6264586258	-0.36175789971	-0.1580870679
21	0.8580140712	0.089993304	0.6939327727	-0.07416147113	-0.4128633866
22	-0.3542980279	0.685404430	-0.0801857450	0.54298803537	0.1308440100
23	0.3006860820	0.517147707	0.3578473849	-0.47944668894	-1.1883926907
24	0.4273321539	-1.137875804	0.0907502788	0.14154007247	0.1590210294
25	-0.2185982238	-0.242577572	-0.5961629032	0.17848274005	-0.3365980079
26	0.9343765378	0.619690010	-0.2239522125	0.10131588418	0.1055704525
27	0.0874046094	-1.433347895	0.1823581654	-0.18964126523	-0.4884899046
28	0.9716947609	1.166757340	0.5697999232	-0.65293422802	-0.7546826241
29	-0.2808017396	-0.153652511	-1.3827228355	0.57233516084	0.4020108373
30	-1.4659992557	0.058002393	0.7725044023	-0.17602418439	0.0399281665
31	1.2583589746	-1.549471238	0.7679223127	0.86776287599	0.4696837342
32	1.0290410803	0.142215908	0.2059001367	0.55195762411	0.0464802367

全個体分 (1215
行)

修正前分散率だと3軸までの累積は、17.2%

	eigen	rate	cum.rate
dim 1	0.4003554	6.405686	6.405686
dim 2	0.3511658	5.618653	12.024339
dim 3	0.3250090	5.200145	17.224484
dim 4	0.3080672	4.929075	22.153558
dim 5	0.2988670	4.781872	26.935430
dim 6	0.2875943	4.601508	31.536938
dim 7	0.2782293	4.451669	35.988607
dim 8	0.2738541	4.381666	40.370273
dim 9	0.2682125	4.291399	44.661673
dim 10	0.2600118	4.160189	48.821862
dim 11	0.2581629	4.130607	52.952469
dim 12	0.2512422	4.019875	56.972344
dim 13	0.2474271	3.958833	60.931177
dim 14	0.2422577	3.876124	64.807301
dim 15	0.2349924	3.759879	68.567179
dim 16	0.2319159	3.710654	72.277834
dim 17	0.2244237	3.590780	75.868613
dim 18	0.2136934	3.419095	79.287708
dim 19	0.2091020	3.345633	82.633341
dim 20	0.2035845	3.257352	85.890693
dim 21	0.1939230	3.102768	88.993461
dim 22	0.1880943	3.009508	92.002969
dim 23	0.1829606	2.927369	94.930338
dim 24	0.1649057	2.638492	97.568830
dim 25	0.1519482	2.431170	100.000000

0.25

修正分散率の計算

GDAtools のmodif.rate
で計算。
modif.rate(res.MCA)



	mrate	cum.mrate
dim 1	47.5863769	47.58638
dim 2	21.5433076	69.12968
dim 3	11.8432759	80.97296
dim 4	7.0975087	88.07047
dim 5	5.0266188	93.09709
dim 6	2.9750065	96.07209
dim 7	1.6774323	97.74953
dim 8	1.1977665	98.94729
dim 9	0.6982038	99.64550
dim 10	0.2109945	99.85649
dim 11	0.1402603	99.99675
dim 12	0.0032481	100.00000

修正分散率だと3軸までの累積は、81.0%

Variance rates and modified rates. The variance rate of axis ℓ is

$$\tau_{\ell} = \frac{\lambda_{\ell}}{V_{\text{cloud}}} = \frac{\lambda_{\ell}}{\frac{K}{Q} - 1}.$$

The mean of eigenvalues is $\bar{\lambda} = (\frac{K}{Q} - 1)/(K - Q) = 1/Q$.

Owing to the high dimensionality of clouds, the variance rates of the first principal axes are usually quite low. To better appreciate the importance of the first axes, Benzécri (1992, p. 412) proposed to use *modified rates*.

For $\ell = 1, 2, \dots, \ell_{\max}$ such that $\lambda_{\ell} > \bar{\lambda}$, calculate

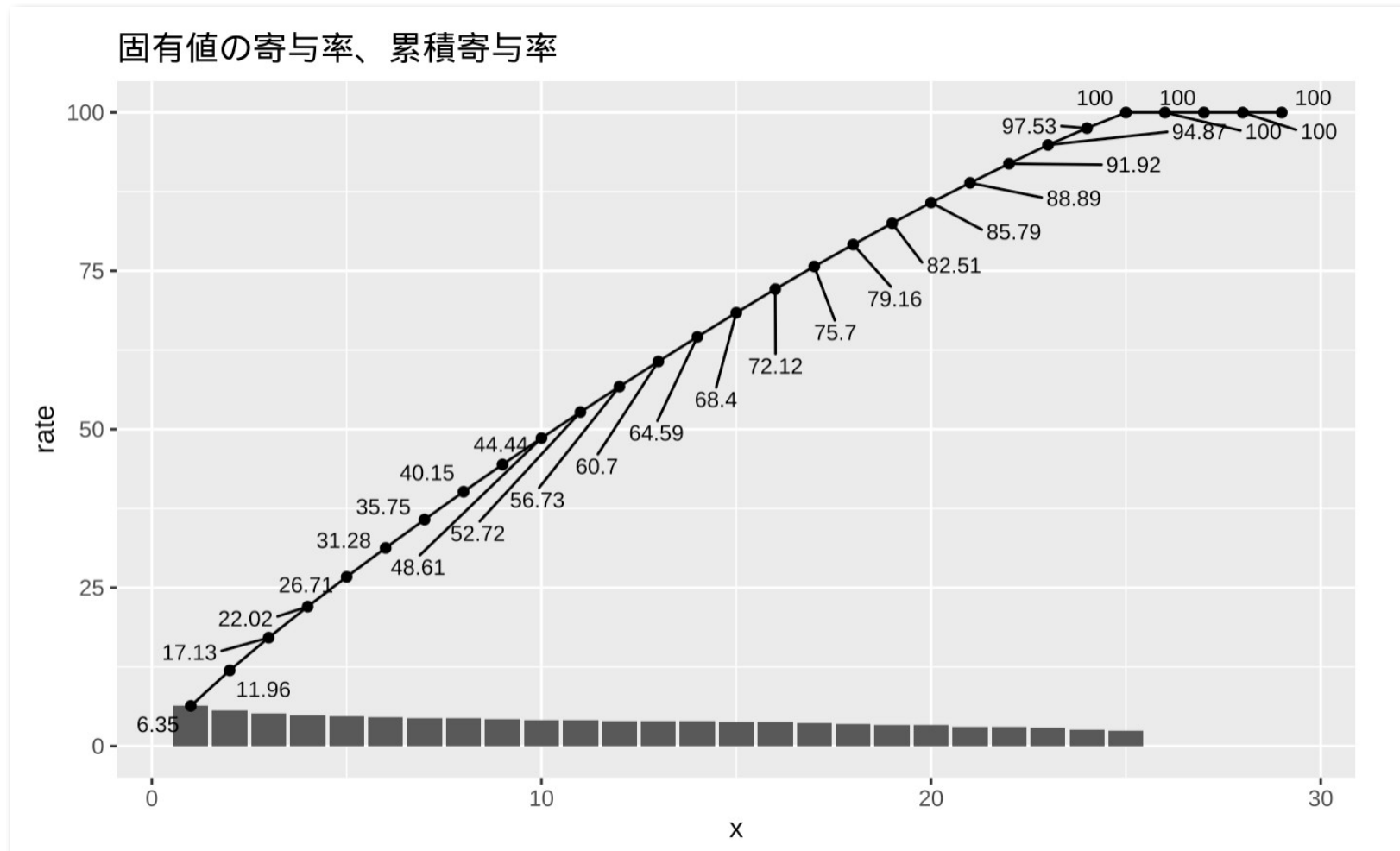
(1) the pseudo-eigenvalue $\lambda'_{\ell} = (\frac{Q}{Q-1})^2 (\lambda_{\ell} - \bar{\lambda})^2$;

(2) the sum $S = \sum_{\ell=1}^{\ell_{\max}} \lambda'_{\ell}$;

then, for $\ell \leq \ell_{\max}$, the modified rates are equal to $\tau'_{\ell} = \lambda'_{\ell}/S$.

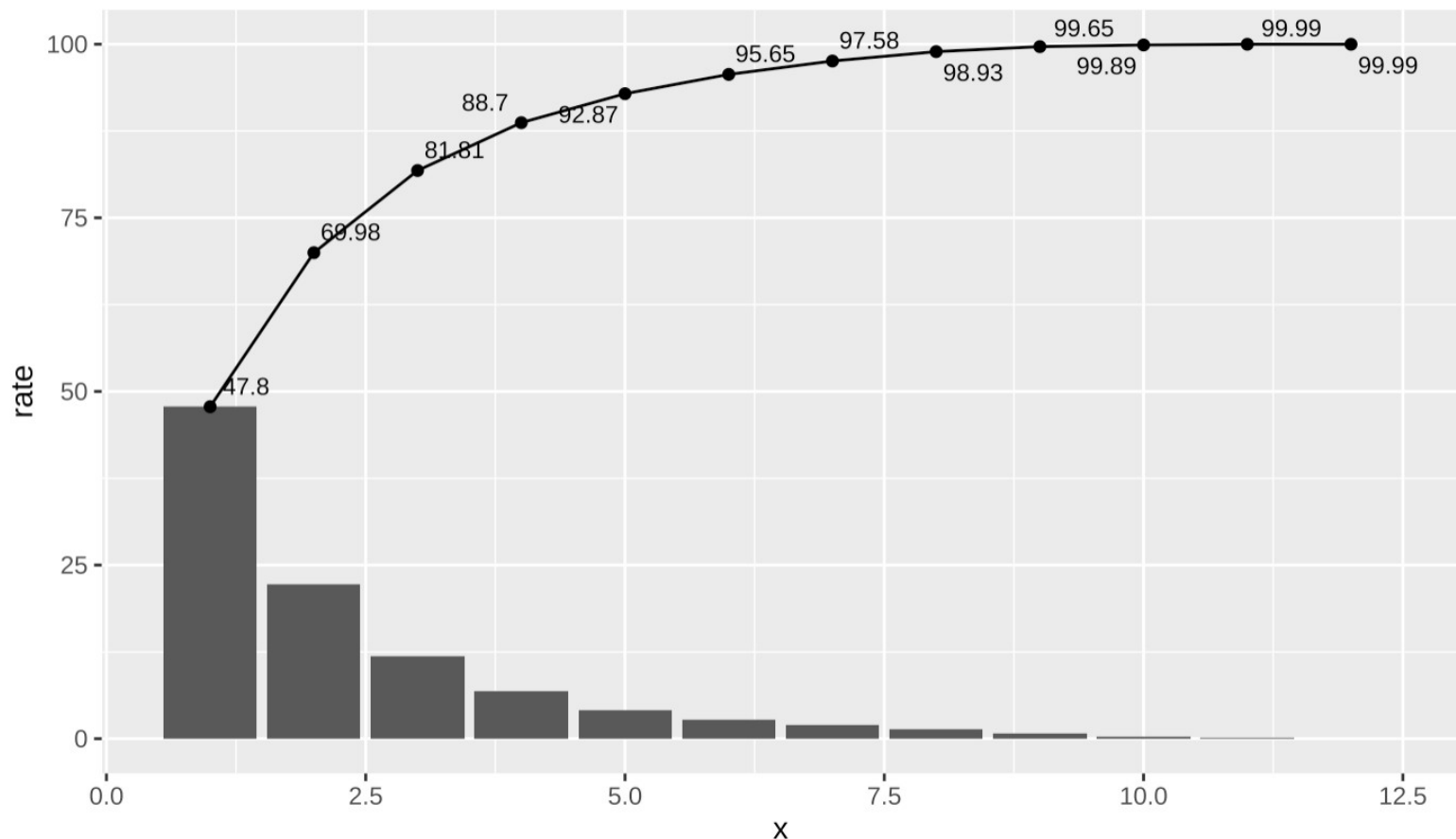
Modified rates can be interpreted as an index of the departure of the cloud from sphericity (i.e., all eigenvalues are equal).

生の固有値、分散率で計算した寄与率、 累積寄与率



修正寄与率、累積修正寄与率

固有値の修正寄与率、累積修正寄与率



x	rate	cum
<int>	<dbl>	<dbl>
1	47.80	47.80
2	22.18	69.98
3	11.83	81.81
4	6.89	88.70
5	4.17	92.87
6	2.78	95.65
7	1.93	97.58
8	1.35	98.93
9	0.72	99.65
10	0.24	99.89

3軸までみれば、全情報の
82%は扱える。
4軸までなら、88.7%。

3軸でいくと決めて、様子
をみて、4軸も検討しようか。

分散の分解：その 1

- 分析対象のデータは、MCAによって、次元縮減され、それは、大きい順に、第 1 軸、第 2 軸、…、に分解される。
- これが、最初の分解。
- そして、多くの場合、1、2 軸という平面、もしくは、3 軸を加えた、立体でデータの分散の分解を考えていくことになる。

MCA模試図的に…

個体l	変数 1	変数 2	…..	変数Q
1				
2				
3				
:				
:				
l				



個体l	cat1 -1	cat1 -	cat1 -k1	cat2 -1	cat2 -	cat2 -k2	….	catQ -1	catQ -2	catQ -kq
1										
2										
3										
:										
:										
l										



次元縮減

	固有値	寄与率	累積寄与率
Dim1			
Dim2			
:			
Dimn			

「変数」雲

変数	Dim1	Dim2	….	Dim n
cat1-1				
cat1-2				
cat1-3		座標値		
:				
:				
CatQ-q				

「個体」雲

個体l	Dim1	Dim2	….	Dim n
1				
2				
3		座標値		
:				
:				
l				

次元縮減

	固有値	寄与率	累積寄与率
Dim1			
Dim2			
:			
Dimn			

Benzécriの修正寄与率

修正寄与率	累積修正寄与率

何軸まで分析対象にするかを累積寄与率を睨んで決める。その時、修正寄与率、累積修正寄与率を見ること。

変数雲

変数	Dim1	Dim2	...	Dim n
cat1-1				
cat1-2				
cat1-3				
:				
:				
CatQ-q				

各軸を解釈（軸に名前をつける）するために、軸に対する変数カテゴリの寄与を確認する。
それをもとに軸に名前をつける。

個体雲の解釈は、軸との関係でみていく。

個体雲

個体I	Dim1	Dim2	...	Dim n
1				
2				
3				
:				
:				
I				

変数雲：

各セルごとに、座標値、度数をもっているので、そこから、分散が計算でき各軸への**寄与率**を計算できる。
そこから、Dim 1、Dim2…の解釈を行う。

この軸の解釈=名称が、あらたな「変数名」

分析のステップ（１）軸の解釈 (Dim1)

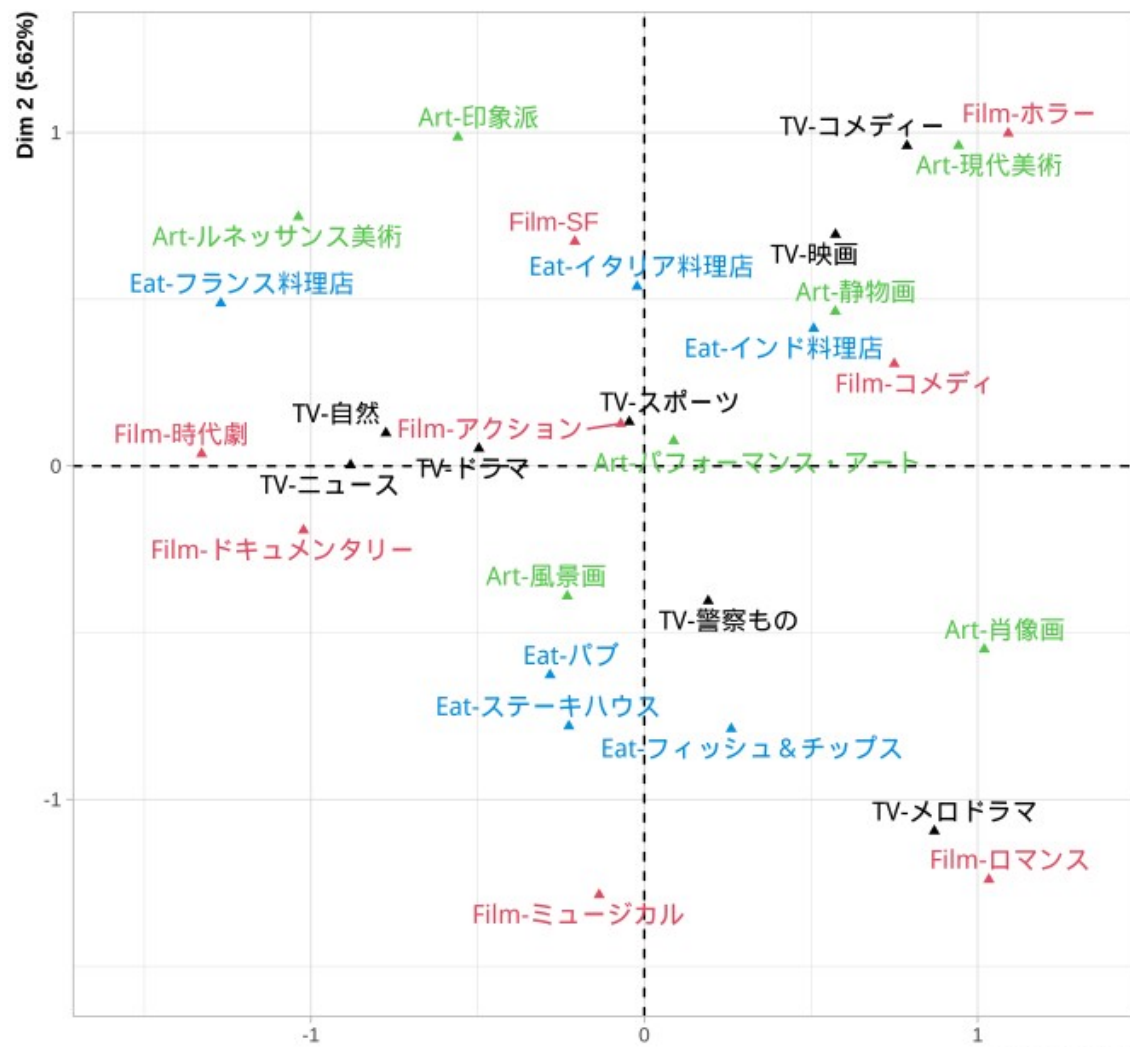
- 変数空間の座標軸の解釈を変数雲をもとに行う
 - 各軸に対して寄与率の大きなものを並べて判定する。

	var	moda	ctr1	ctr2	weight	ctrtot	cumctrtot
5	映画	時代劇	-12.69		140	34.2	34.2
7		ドキュメンタリー	-5.37		100		
6		ホラー		3.8	62		
8		ロマンス		5.55	101		
9		コメディ		6.79	235		
4	TV	ニュース	-8.78		220	26.91	61.11
2		自然	-4.91		159		
1		コメディ		4.85	152		
3		メロドラマ		8.37	215		
10	外食先	フランス料理店	-8.21		99	13.55	74.66
11		インド料理店		5.34	402		
13	芸術	現代美術		5.03	110	11.29	85.95
12		肖像画		6.26	117		

この手順から軸の名前をつける

- この事例では以下の通り。（MCA2010=2021:72-74）から短縮表現。
 - Dim1
 - 事実&伝統的 vs 架空&現代的
 - Dim2
 - 大衆的 vs 洗練
 - Dim3
 - 硬い vs 軟らかい
- この表記は、マップに記入するのがよい。

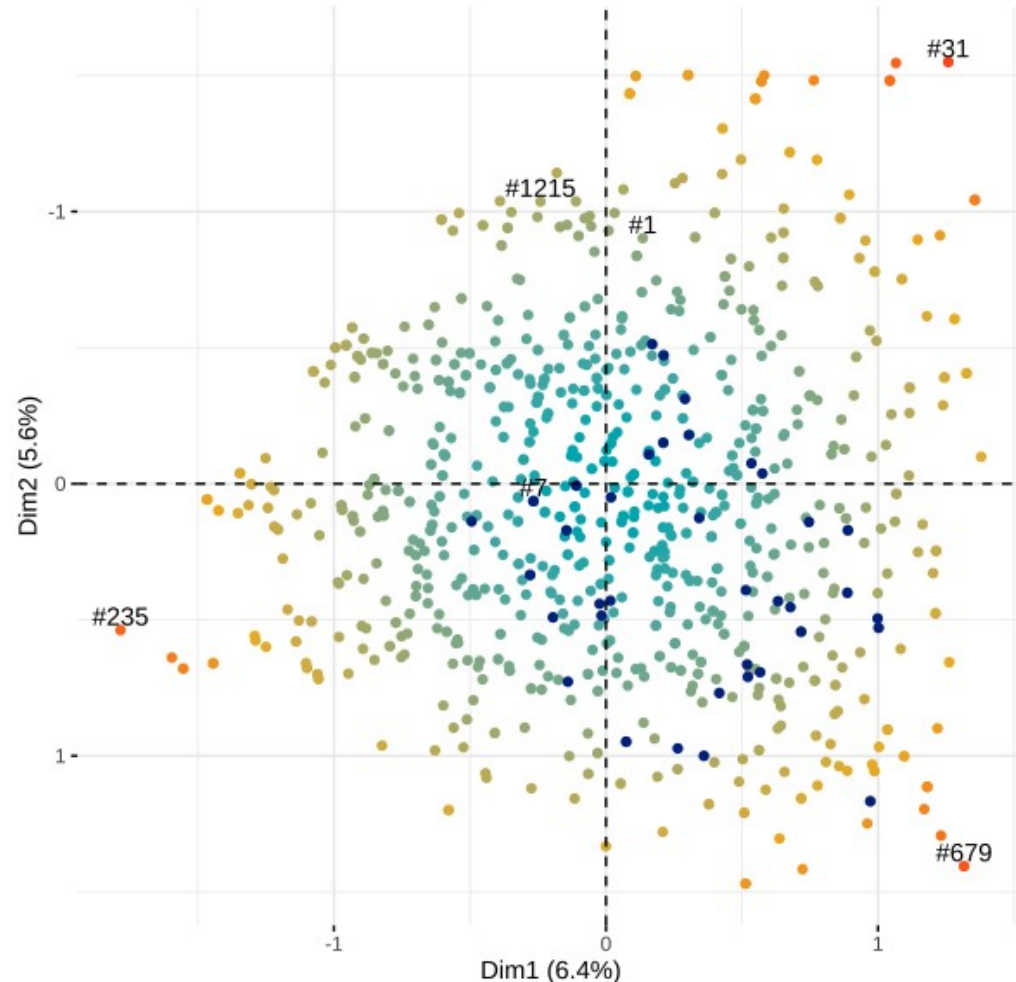
図1.2 嗜好データの例 (変数)



- MCA2021のグラフでのポイントアイコンのサイズは、度数。
- ここでは、変数ごとに色分けを行ってみた。
- 他にも、寄与率、 $\cos^2$ 、ポイント選択などの「フィルタリング」を行なって、解釈を進める。

図1.3 嗜好データの例（個体）

図1.3 嗜好データの例 1-2平面に布置した1,215人の個体雲



- テキストの図は、5つの個体の番号を表示。
- ここでは、寄与率で色分けをしている。
- 寄与率は、重心（原点）に近いほど小さい。慣性が小さい＝剛体が重心を中心に回転する仕組み。

図1.4 個体を年齢グループで選択し表示

図1.3 嗜好データの例 1-2平面に布置した「55歳以上」の個体雲

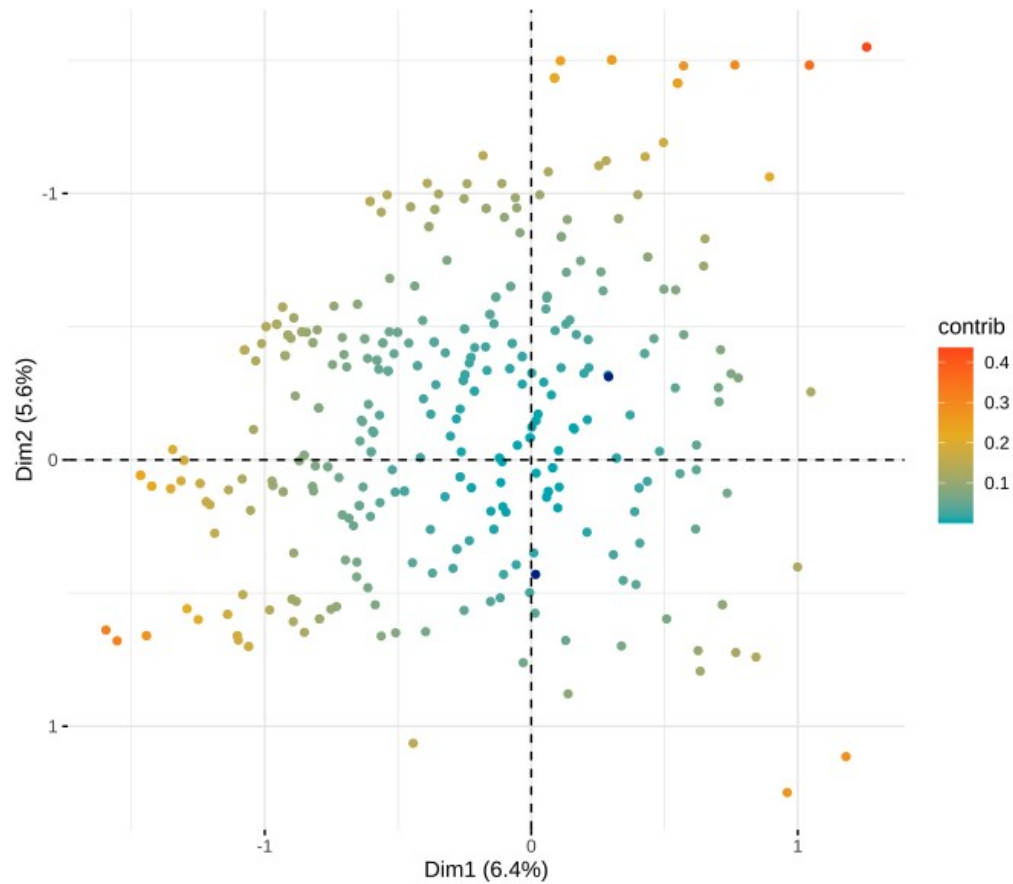
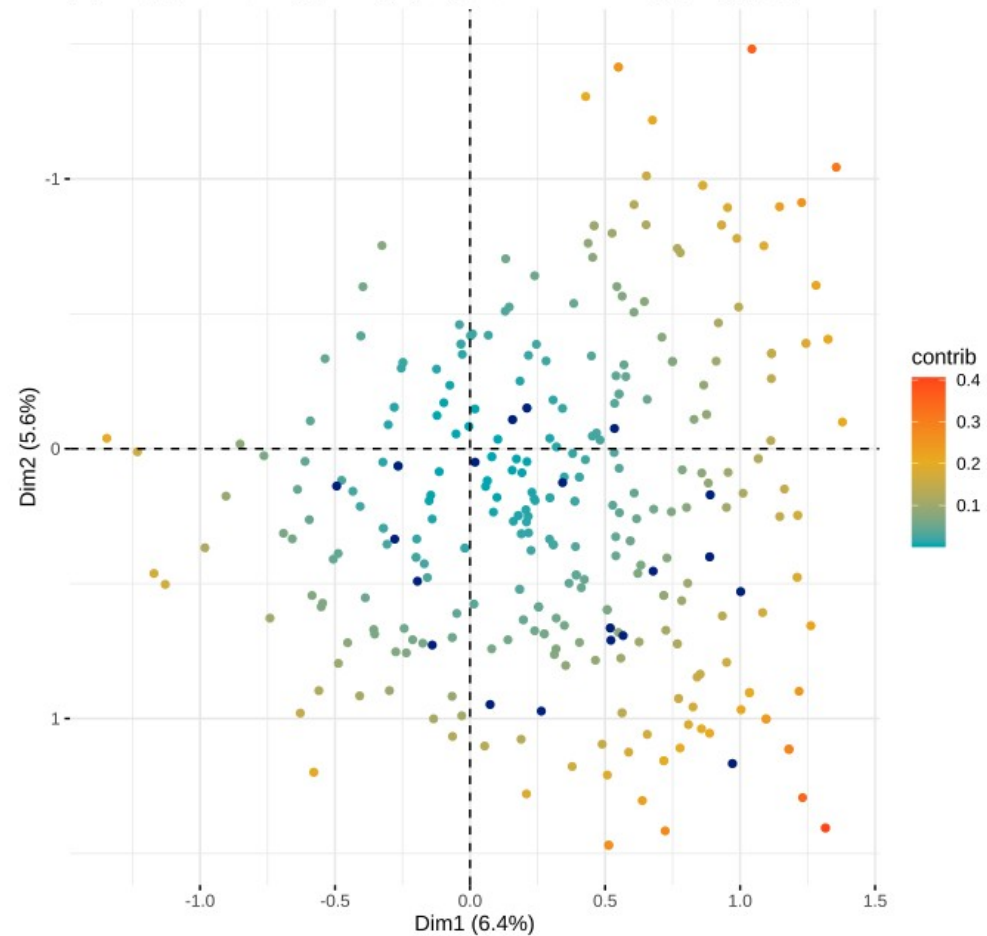


図1.3 嗜好データの例 1-2平面に布置した「18~34歳」の個体雲



★13
個

参考：p71の「13個」

	Qname	Dim 1	Dim 2	Dim 3
★	TV-ニュース	8.77743335	0.0001050202	0.1047095
★	TV-コメディ	4.85077620	8.2114230962	0.6272052
	TV-警察もの	0.15487580	0.7871122729	0.8544972
★	TV-自然	4.90966786	0.0910088905	0.5530999
	TV-スポーツ	0.01420702	0.1400435868	18.5798744
	TV-映画	1.97876735	3.3034047795	2.7226602
	TV-ドラマ	1.69517322	0.0216559870	8.1651172
★	TV-メロドラマ	8.36646043	15.1001334886	6.7954357

	Qname	Dim 1	Dim 2	Dim 3
	映画-アクション	0.09924282	0.36697885	10.52582632
	映画-コメディ	6.79277860	1.29342231	1.40442949
★	映画-時代劇	12.68760258	0.01105005	13.63110341
★	映画-ドキュメンタリー	5.37004932	0.21704377	1.72407894
★	映画-ホラー	3.79927157	3.61648256	0.04150249
	映画-ミュージカル	0.08192607	8.42972618	0.06573162
★	映画-ロマンス	5.54811664	9.10297613	9.44419051
	映画-SF	0.22539698	2.68294853	2.66447241

	Qname	Dim 1	Dim 2	Dim 3
	芸術-パフォーマンス・アート	0.04213001	0.03496876	0.03102046
	芸術-風景画	1.73353973	5.63855258	3.91526640
★	芸術-ルネッサンス美術	3.04456984	1.79954109	1.11744484
	芸術-静物画	1.19799162	0.89361683	0.06147481
★	芸術-肖像画	6.25555799	2.07103886	0.14995891
★	芸術-現代美術	5.02567375	5.94955153	0.56623693
	芸術-印象派	2.01020307	7.12798995	5.37867937

	Qname	Dim 1	Dim 2	Dim 3
	外食先-フィッシュ&チップス	0.374316601	3.894196	0.6636940
	外食先-パブ	1.152963284	6.464242	0.1351212
★	外食先-インド料理店	5.336903408	4.006515	0.3610517
	外食先-イタリア料理店	0.005409728	3.869514	2.9447721
★	外食先-フランス料理店	8.211261607	1.382224	3.5028733
	外食先-ステーキハウス	0.257733554	3.492535	3.2684714

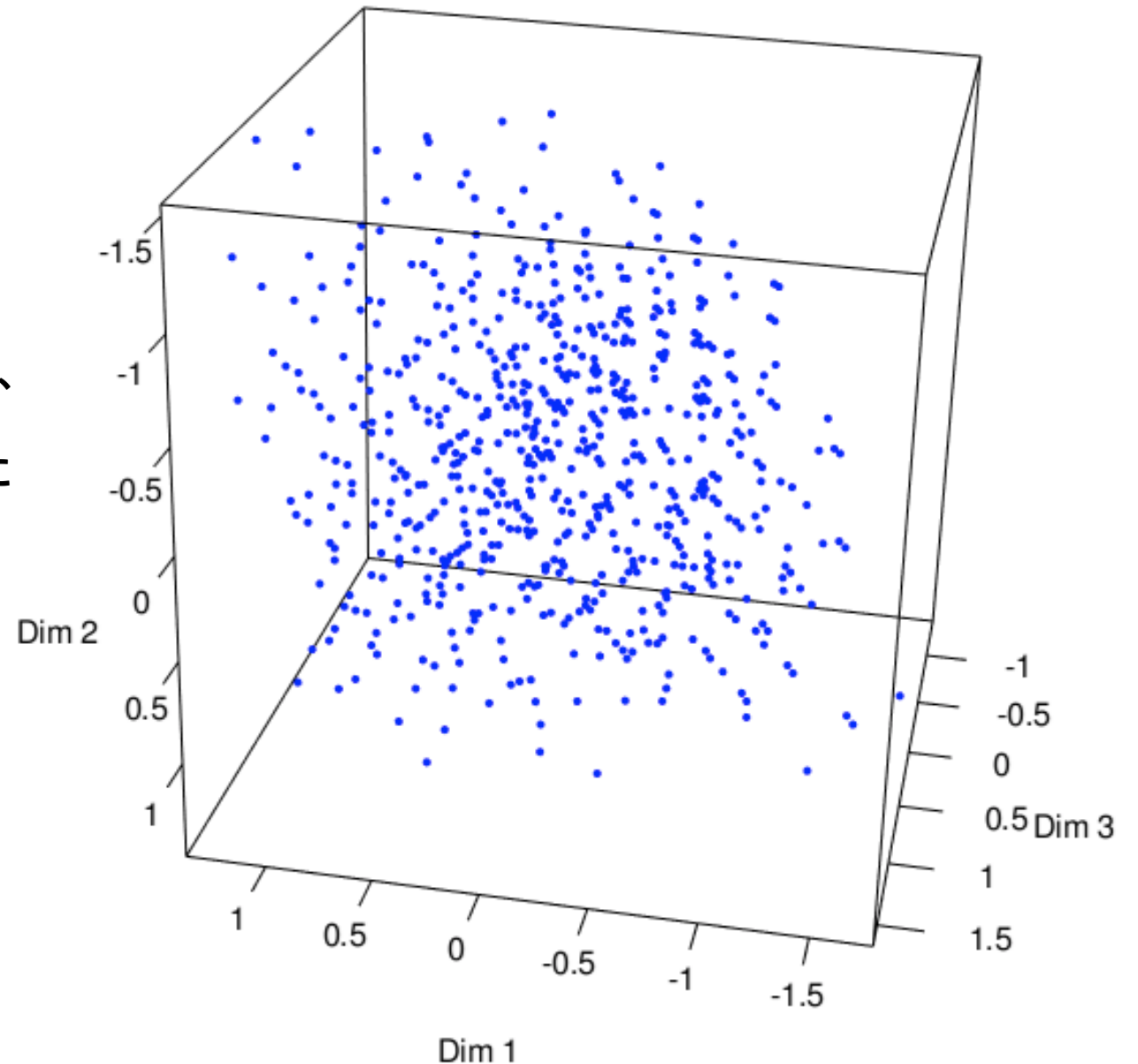
3D散布図で表示

グルグルまわるので、面白くはありますが
わかりやすいかというとそうでもないので、

軸を指定した2D表示と合わせて使うこと
になります。

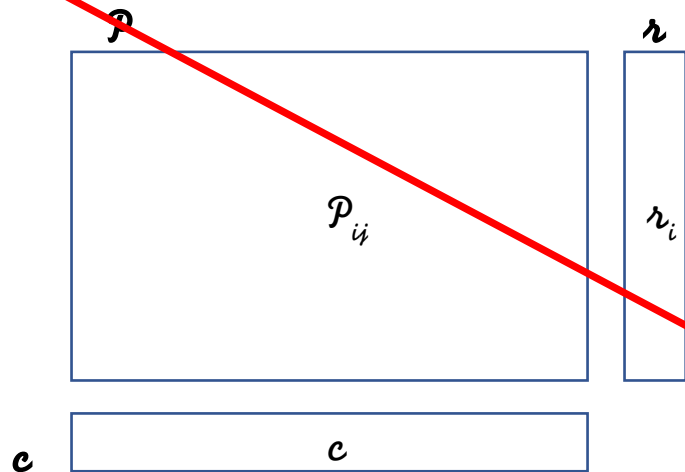
2Dでのフィルタリング（名前を色分け、
Ctr、cos2、など）の手法の方が重要。

- FactoShiny
- explor <- 私の好み



MCAの二つのバリエーション

- Specific MCA (speMCA)
 - MCAする際に、空間生成からはずすカテゴリを選定するカテゴリ特定MCA（大隅他訳では「限定多重対応分析」と呼んでいる）
- Class Specific Analysis (CSA)
 - MCAする対象とする個体を選ぶ、個体特定MCA（大隅他訳では「集団限定多重対応分析」と呼んでいる）
- どちらも、元データ表のサブセットを作ってMCAを行うのではなく、**特定MCA**は、**元のMCA**との関係进行分析可能にするために、周辺度数を維持して特定MCA用の周辺度数（質量）を用いてMCAを行う。次ページにその関係を図示。
 - 『津田塾大学紀要』55号139–140（この最後のページの説明は誤り！）
- Greenacreは、サブセットMCAと呼ぶ。



図A-1 対応行列、行和、列和の基本形

$$S = D_r^{-1/2}(P - rc^t)D_c^{-1/2}$$

$$S = UDV^t$$

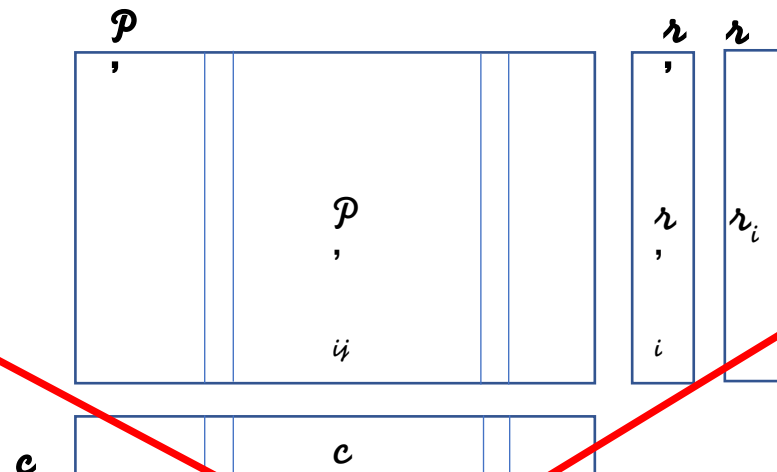
$$\Phi = D_r^{-1/2}U$$

$$\Gamma = D_c^{-1/2}V$$

$$F = D_r^{-1/2}UD_\alpha = \Phi D_\alpha$$

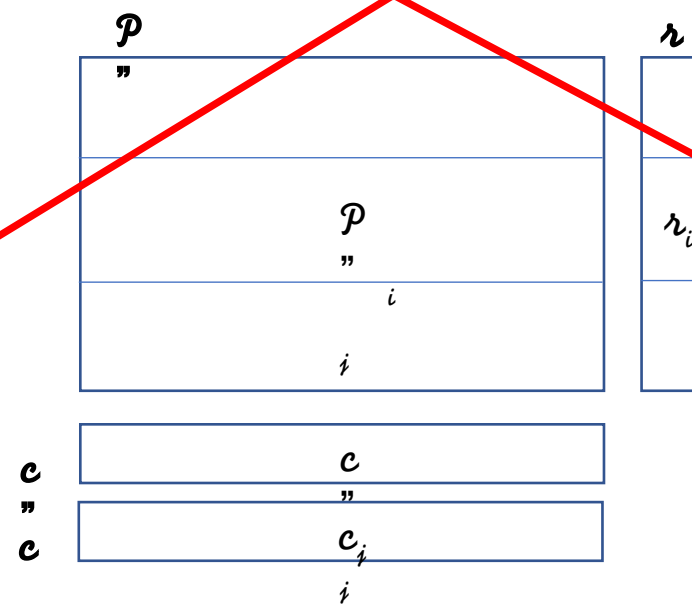
$$G = D_c^{-1/2}VD_\alpha = \Gamma D_\alpha$$

2023/02/28



図A-2 speMCAでの対応行列、行和、列和

$$S' = D_r'^{-1/2}(P' - r'c'^t)D_c'^{-1/2}$$



図A-3 CSAでの対応行列、行和、列和

$$S' = D_r'^{-1/2}(P' - r'c'^t)D_c'^{-1/2}$$

対応分析研究会第17回

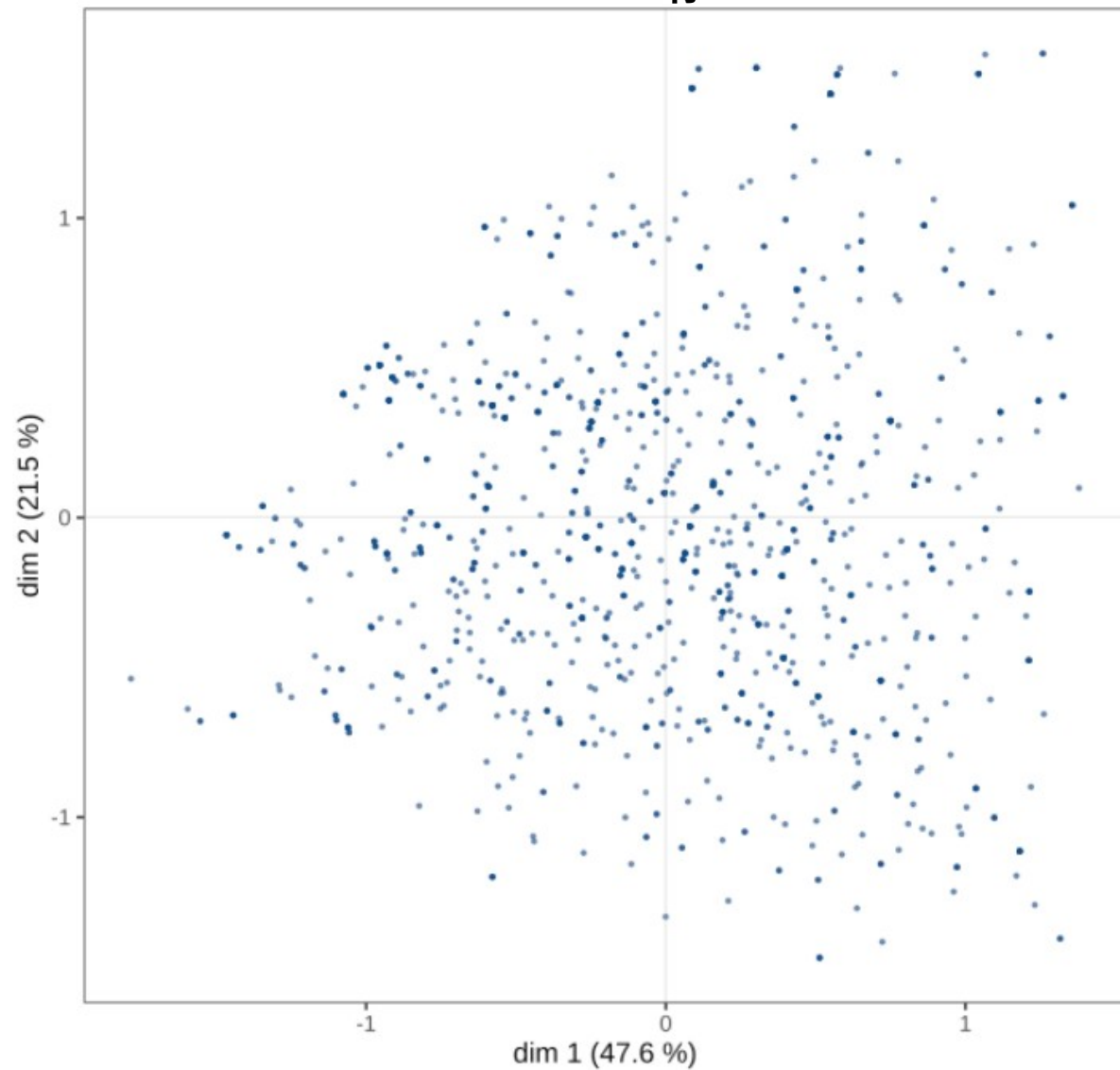
正しい処理

- 標準化残差行列をつくるところまでは同じ。
- そうやってできた残差行列から、junk指定の列（カテゴリ）を除去した、 S' をつくる。
- そして、その S' をSVDする。
- こういう手順です。
- 申し訳ありません！

構造化データ解析 (SDA)

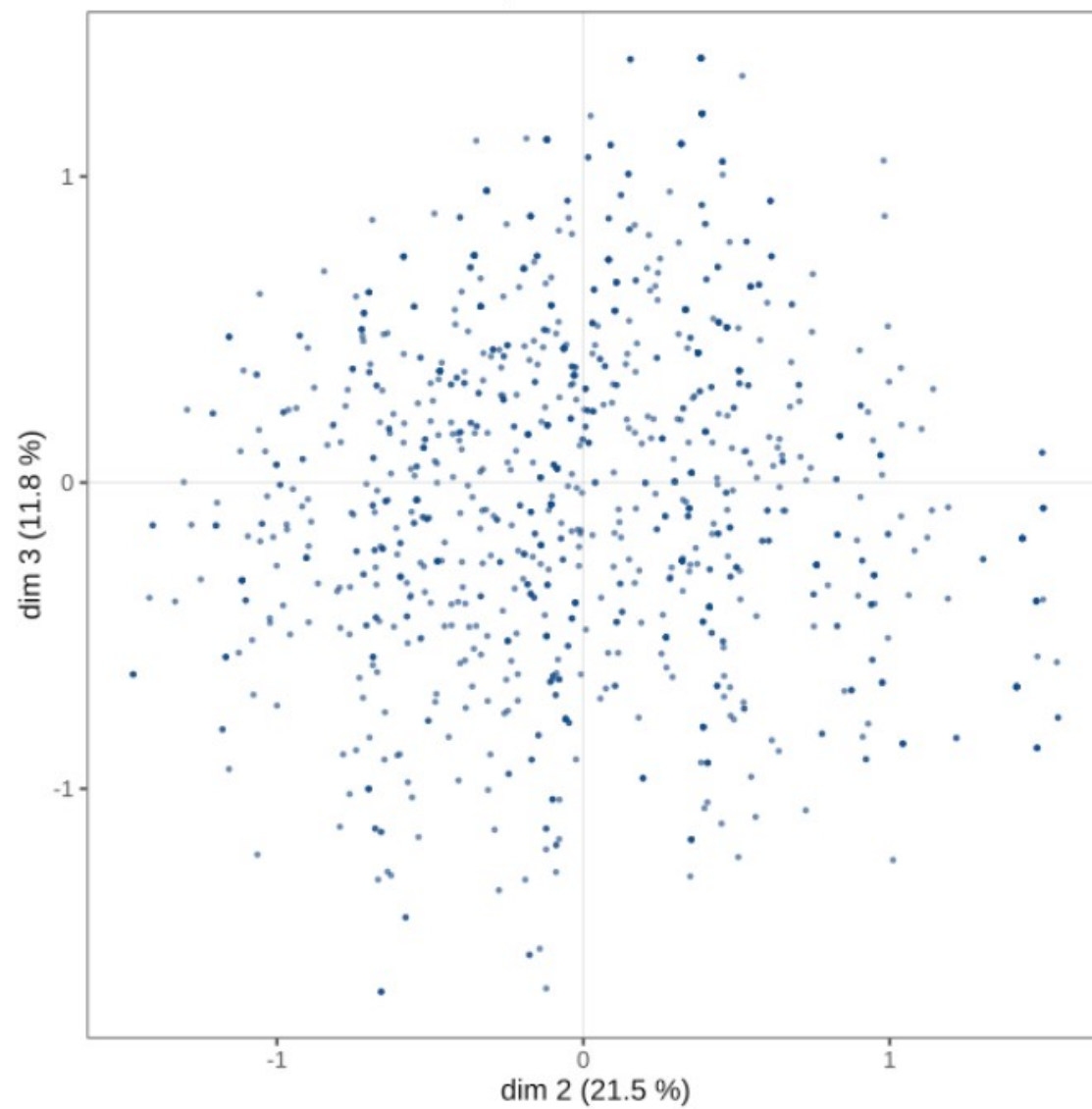
嗜好データの個体雲1-2軸

p12として保
存



嗜好データの個体雲2-3軸 p67 図3.2

p23として保
存



10 性別の分析

```
varsup(resmca = res.speMCA,var = .d_a$Gender) -> res.varsup_Gender
```

10.1 重み

```
res.varsup_Gender$weight
```

```
## 男性 女性  
## 513 702
```

10.2 平均点の座標（これはstd）

```
res.varsup_Gender$coord[,1:3]
```

```
##          dim.1      dim.2      dim.3  
## 男性 -0.177703 -0.266200  0.525805  
## 女性  0.129860  0.194531 -0.384242
```

10.3 分散、性別内（群内分散）、性別間（群間分散）、合計、 η^2

```
res.varsup_Gender$var[,1:3]
```

```
##          dim.1      dim.2      dim.3  
## 男性  0.291503  0.252755  0.256650  
## 女性  0.463911  0.391608  0.261315  
## within 0.391117  0.332981  0.259345  
## between 0.009239  0.018185  0.065664  
## total  0.400355  0.351166  0.325009  
## eta2    0.023076  0.051784  0.202037
```

図4.2左

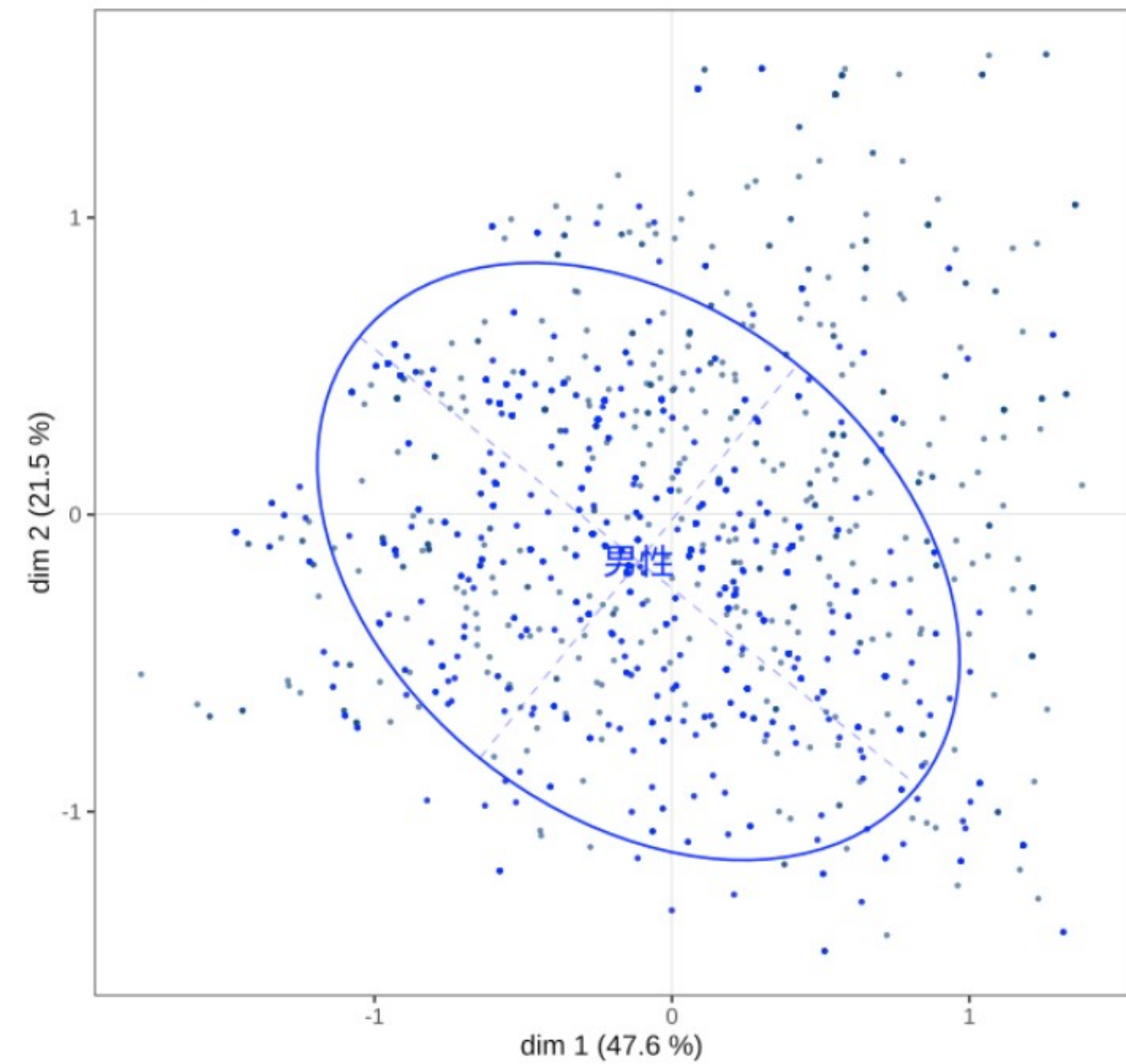
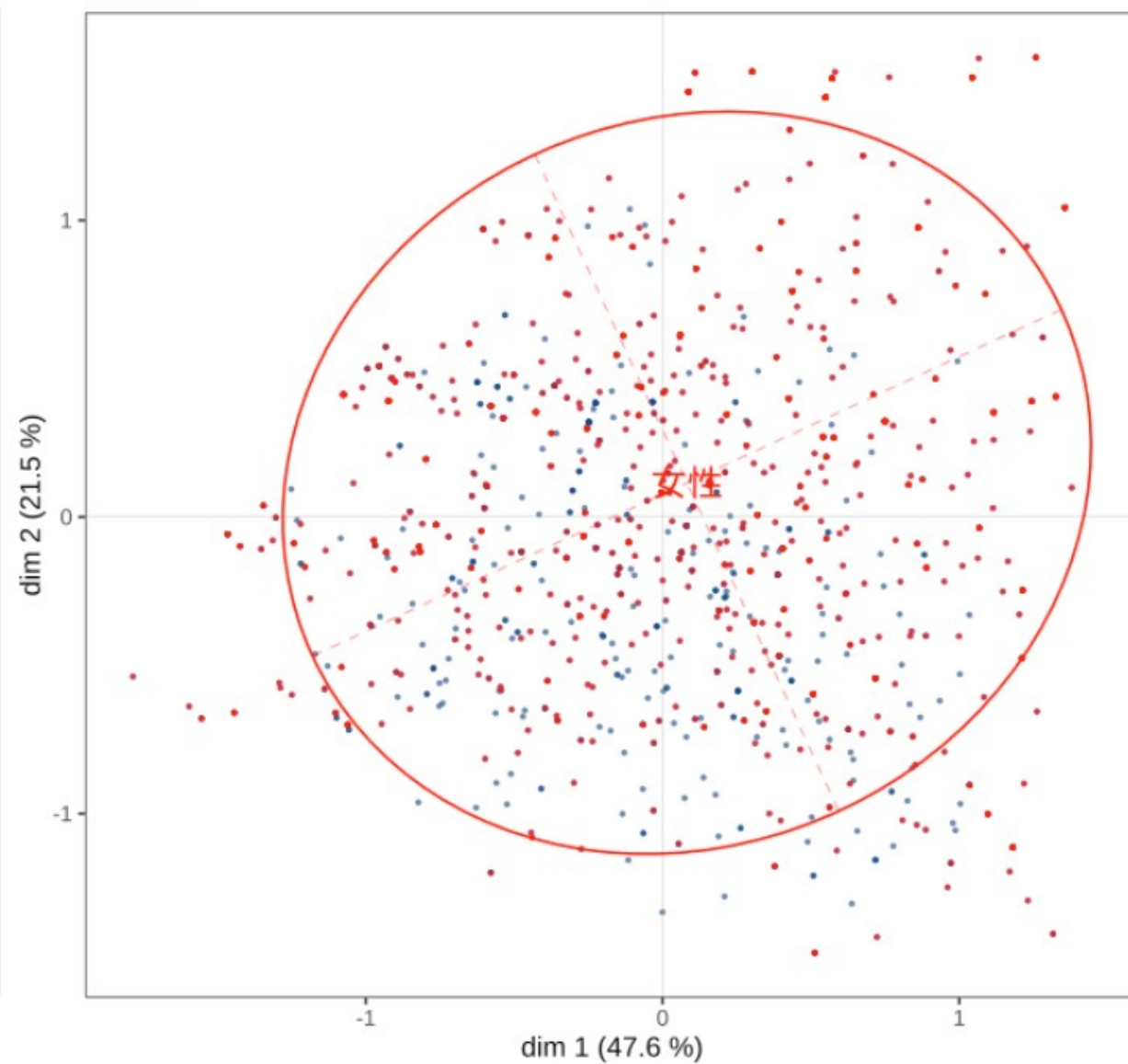
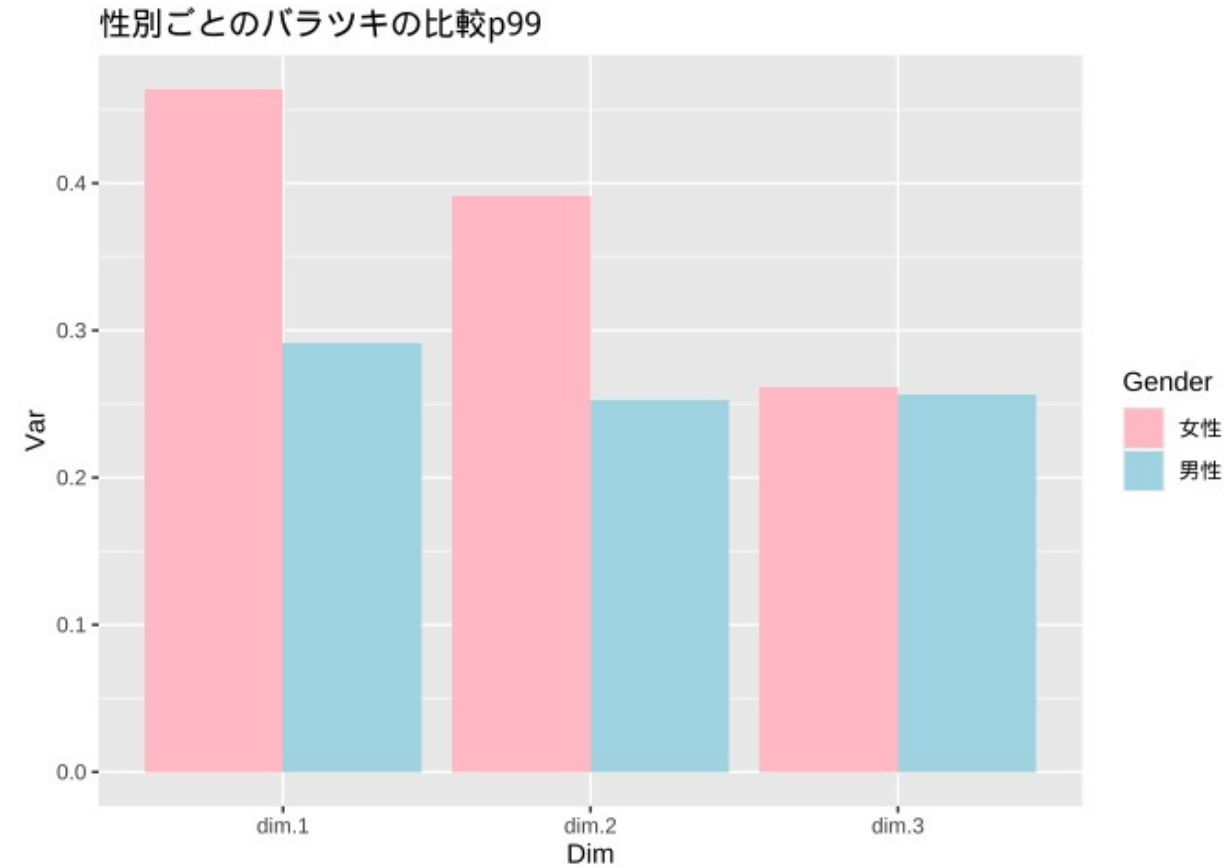
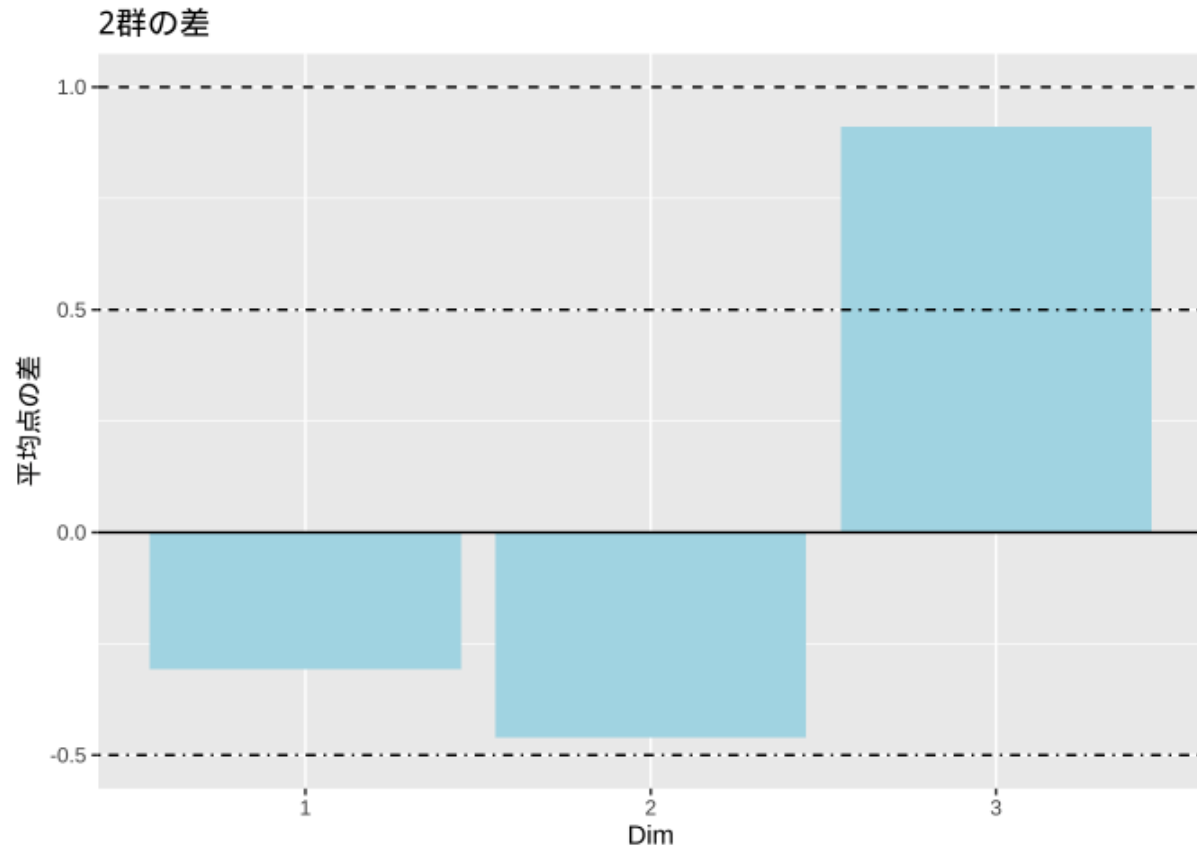


図4.2右

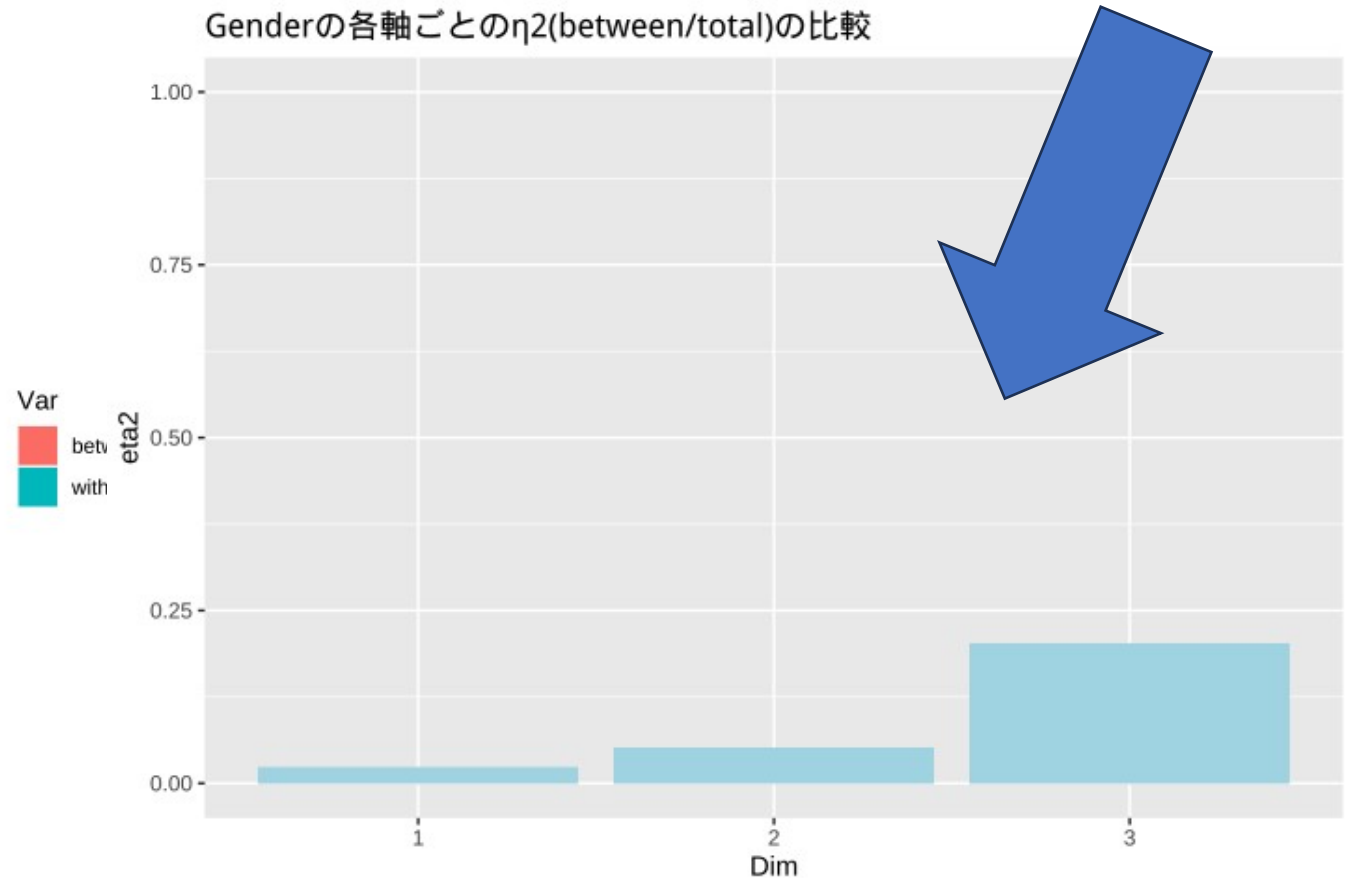
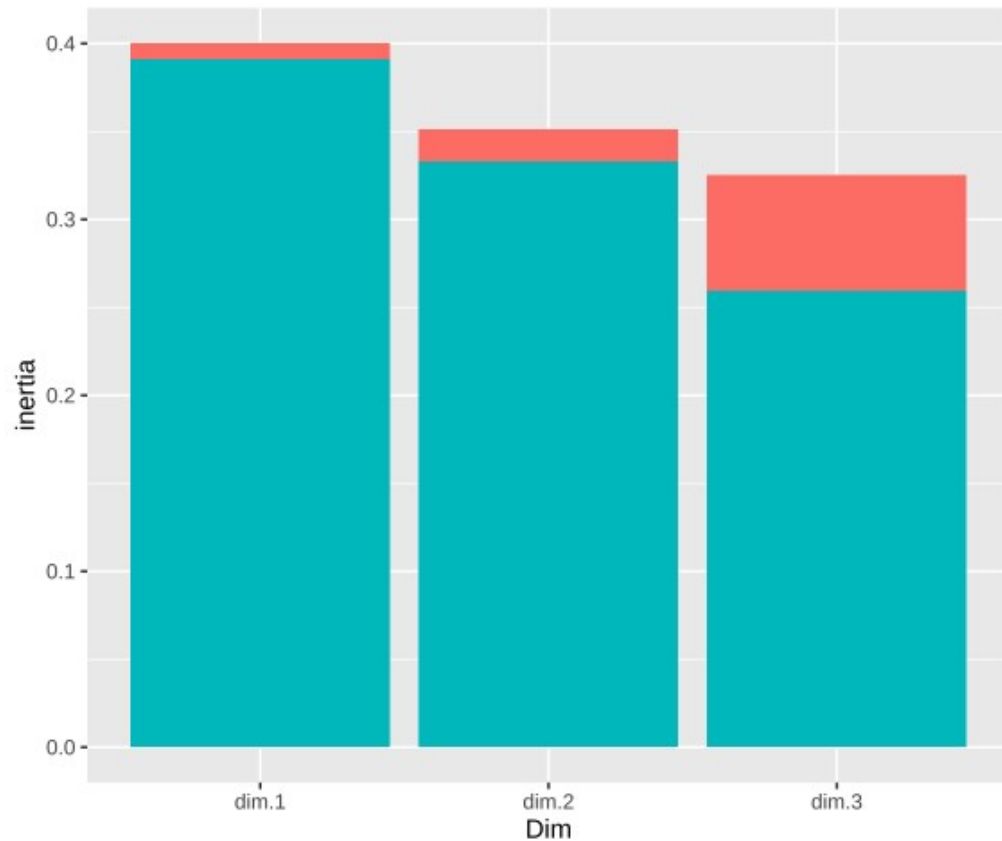


平均点の差と性別ごとのバラツキ比較



Dimごとのbetween、 withinと η^2

分散の比を
見えています



以上から言えることは！

- η^2 が大きい
- 嗜好の男女間の違いは、主に第3主軸での違い「硬い」vs「柔らかい」の違いである。p100

年齢（Age）の分析 p100～

- グラフ
 - 個体のグラフをつくり、pxxというオブジェクトにする
 - そこに、性別による集中楕円をオーバーレイする
- MCA2021の表4.2の解釈をなぞる
 - varsupを使って平均点座標、分散、Vbetween、Vwithin、 η^2 を取得
 - それをもとに、テキストの解釈をなぞってみる。

性別×年齢の 分析 p101～

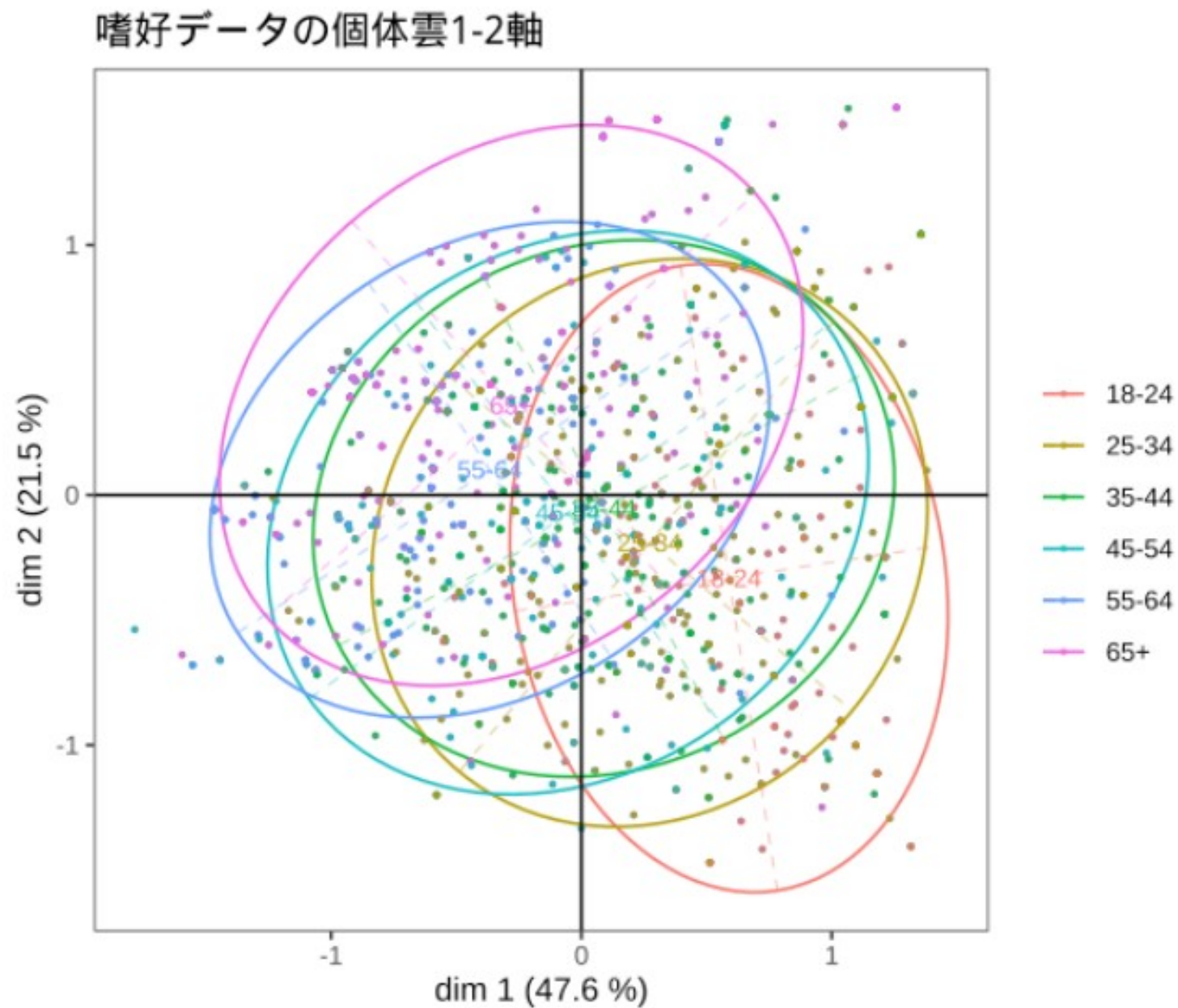


图4.4左

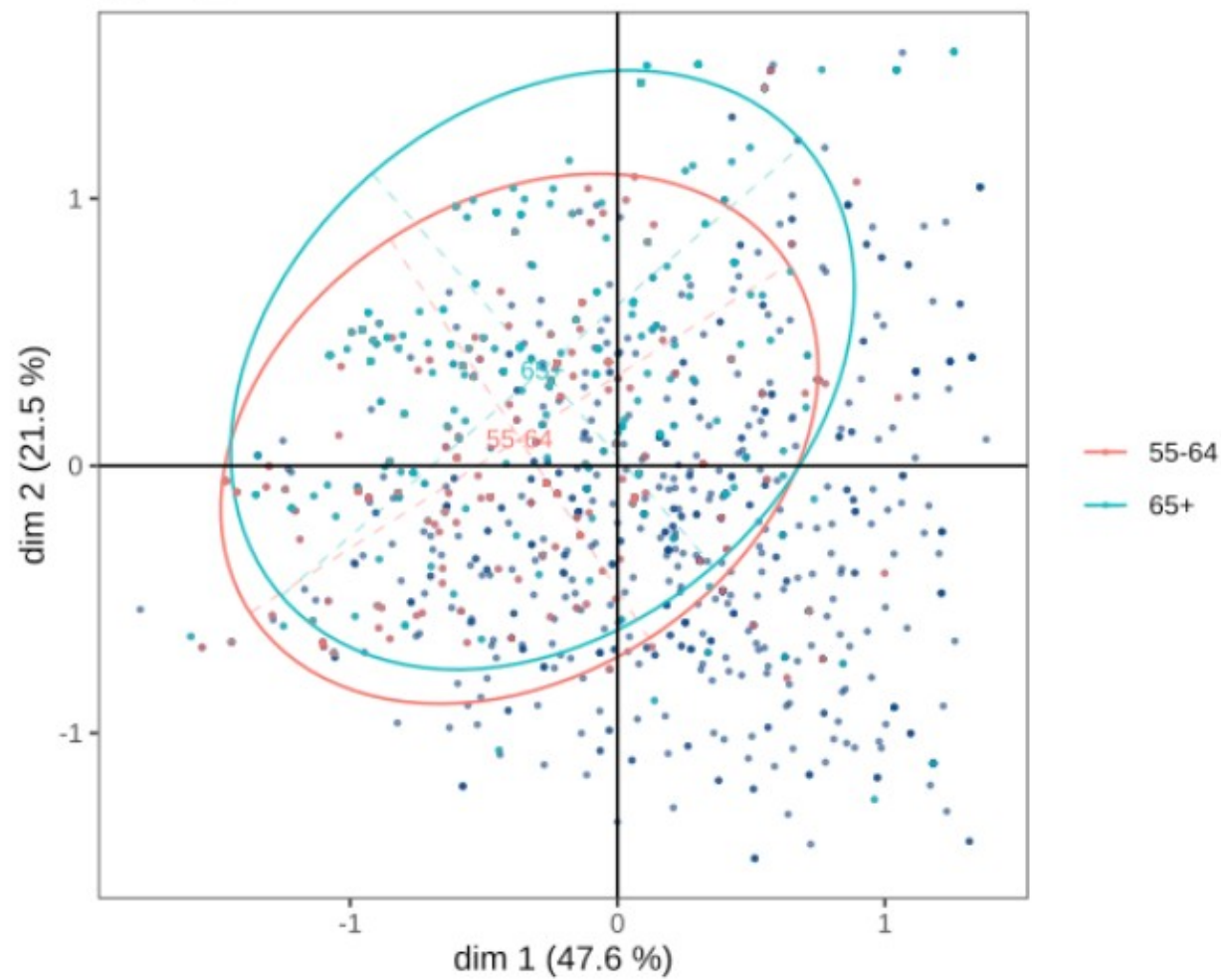
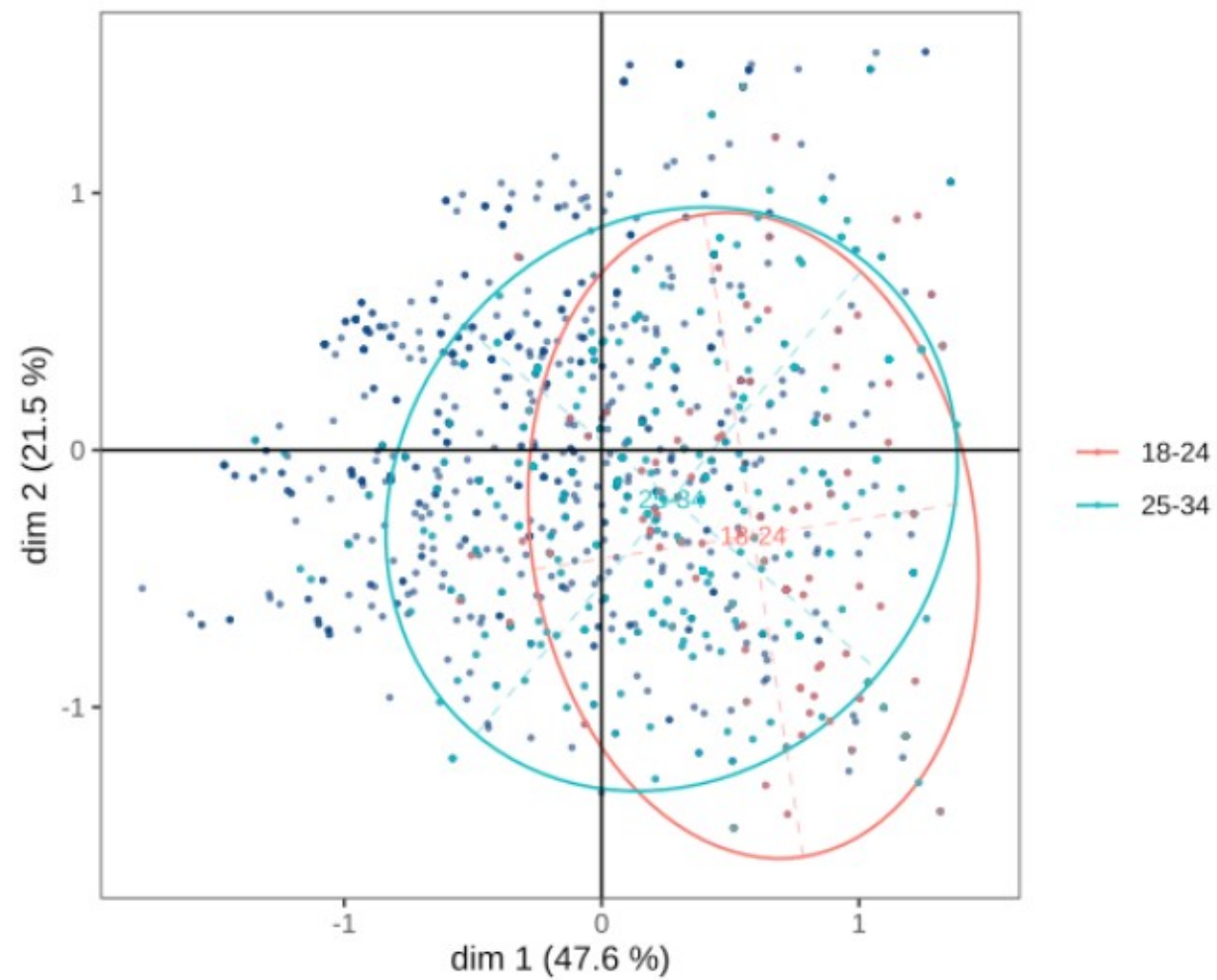
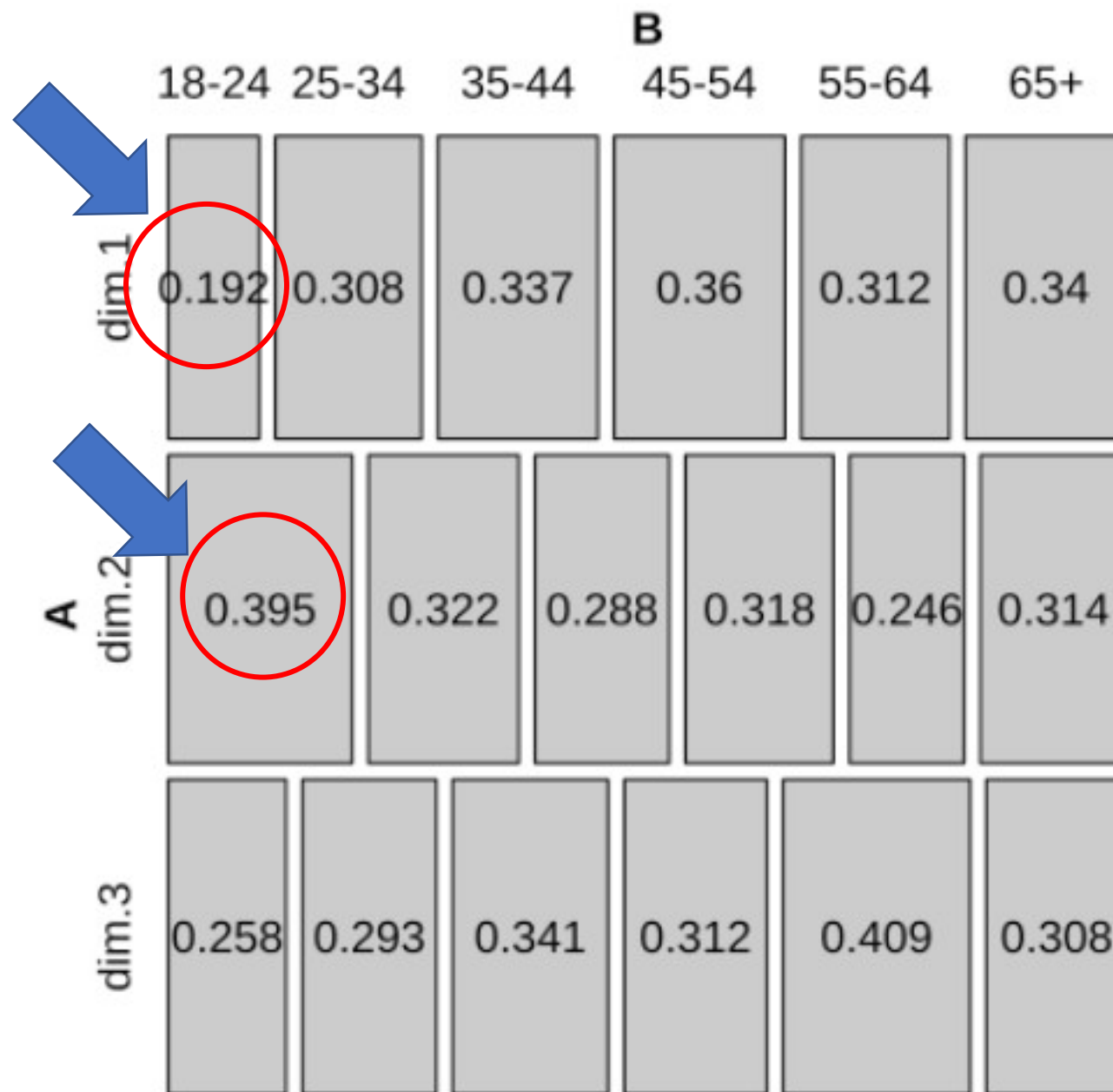


图4.4右



各因子の分散を 比較してみた



性別×年齢変数をつくりそれを分析

12 追加2変数の交互作用

```
.d_a %>% mutate(Gender_Age= str_c(Gender, Age) %>% factor) -> .d_aGA
varsup(resmca = res.speMCA, var = .d_aGA$Gender_Age) -> res.varsup_Gender_Age
res.varsup_Gender_Age$weight
```

```
## 女性18-24 女性25-34 女性35-44 女性45-54 女性55-64 女性65+ 男性18-24 男性25-34
##      53      142      141      117      99      150      40      106
## 男性35-44 男性45-54 男性55-64 男性65+
##      117      74      84      92
```

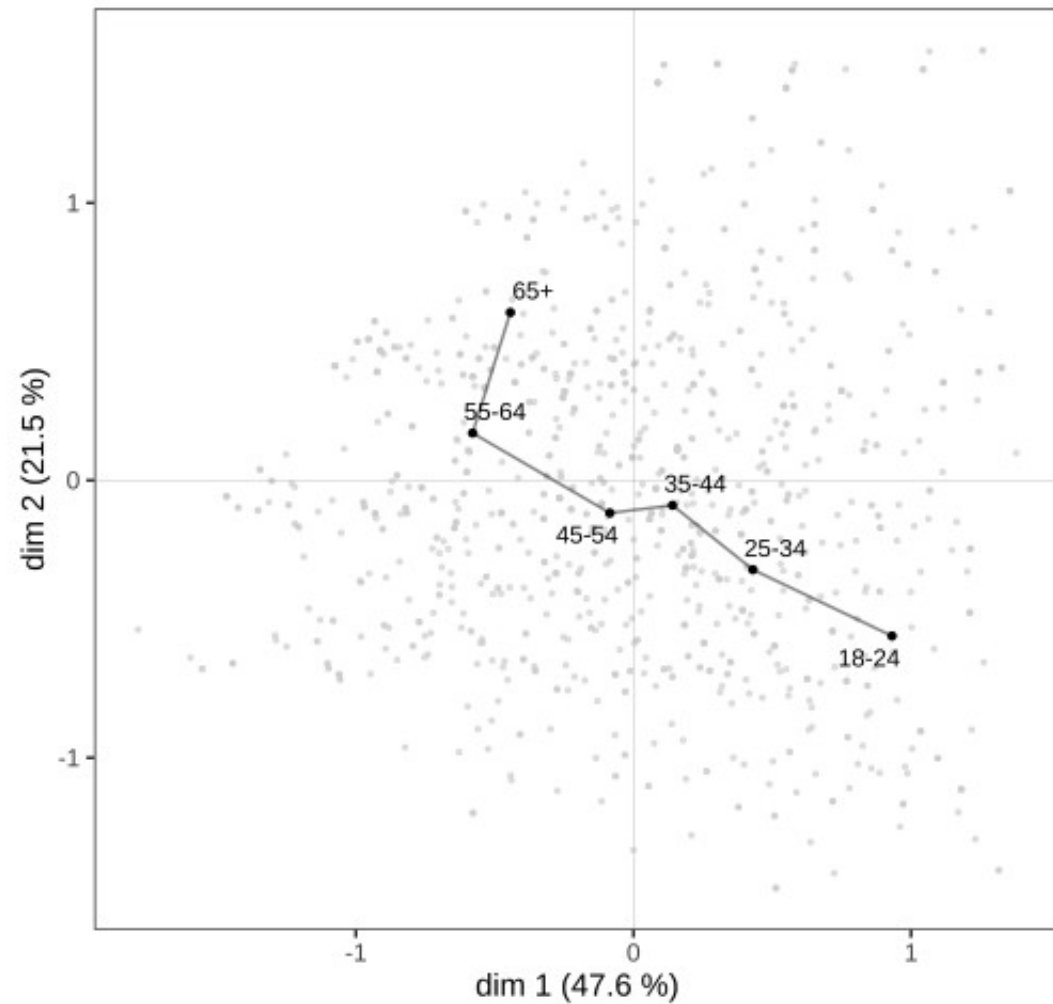
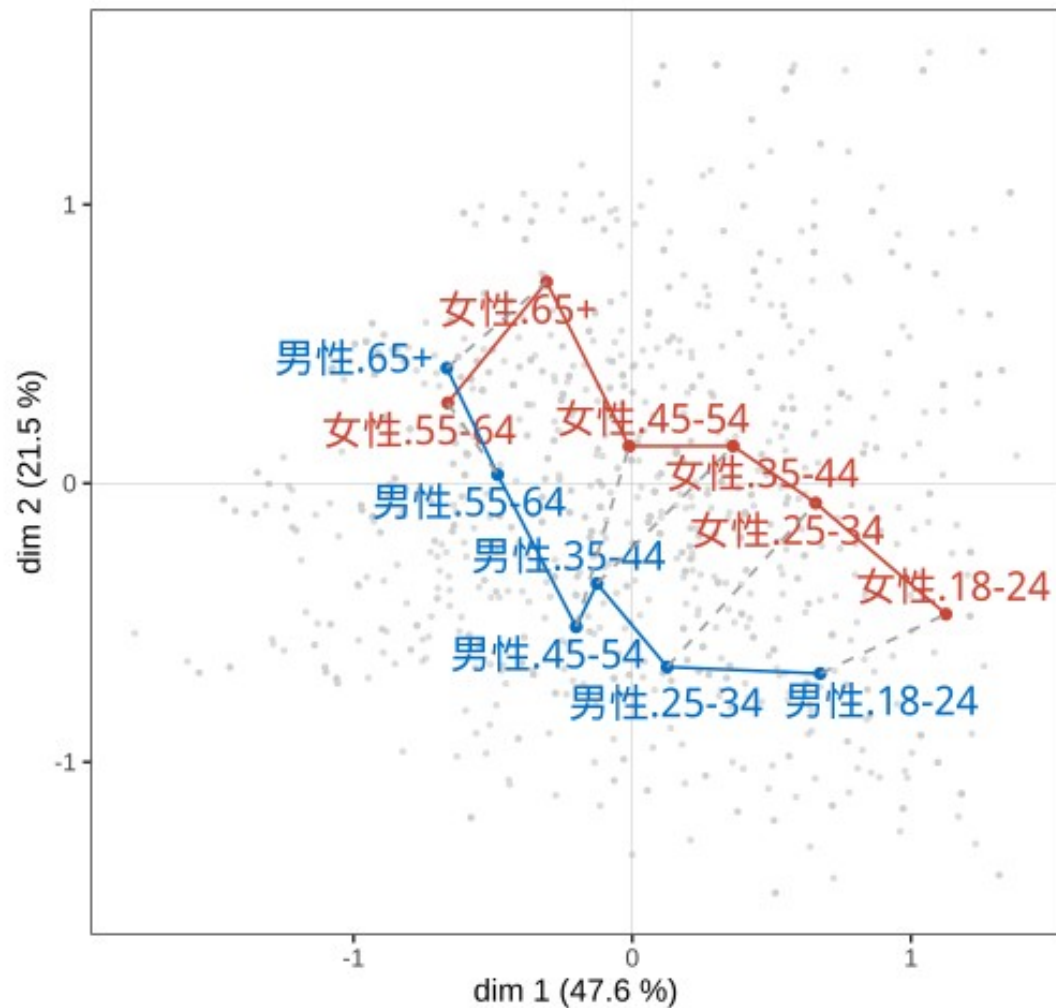
```
res.varsup_Gender_Age$coord[,1:3]
```

```
##      dim.1      dim.2      dim.3
## 女性18-24  1.125357 -0.469211 -0.418175
## 女性25-34  0.657875 -0.070409 -0.417832
## 女性35-44  0.362218  0.133671 -0.341202
## 女性45-54 -0.011532  0.133440 -0.380531
## 女性55-64 -0.661864  0.288703 -0.535772
## 女性65+   -0.307331  0.722569 -0.283797
## 男性18-24  0.674363 -0.682249  0.613144
## 男性25-34  0.125227 -0.658825  0.501056
## 男性35-44 -0.126279 -0.359435  0.613188
## 男性45-54 -0.201166 -0.514517  0.391094
## 男性55-64 -0.482818  0.031290  0.581759
## 男性65+   -0.665134  0.413744  0.462487
```

```
res.varsup_Gender_Age$var[,1:3]
```

```
##      dim.1      dim.2      dim.3
## 女性18-24  0.149859  0.483565  0.135040
## 女性25-34  0.301969  0.357654  0.222110
## 女性35-44  0.359025  0.315968  0.245139
## 女性45-54  0.422335  0.340349  0.311525
## 女性55-64  0.408657  0.298884  0.346152
## 女性65+   0.381104  0.383327  0.253841
## 男性18-24  0.200515  0.267742  0.224111
## 男性25-34  0.251672  0.205778  0.231808
## 男性35-44  0.258494  0.207700  0.293802
## 男性45-54  0.253736  0.191399  0.194294
## 男性55-64  0.191292  0.170871  0.264567
## 男性65+   0.241431  0.180951  0.283661
## within    0.304921  0.287954  0.257337
## between   0.095435  0.063212  0.067672
## total     0.400355  0.351166  0.325009
## eta2      0.238375  0.180006  0.208217
```

交互作用plot



帰納的データ解析 IDA

構造化データ解析（SDA）までは記述統計。IDAで検定が行われる。
SDAで確認された差異は、有意なのかどうか。

典型性検定

同質性検定

関連する「用語集」

- 集中楕円、慣性楕円、指示楕円、信頼楕円p174
- 準拋母集団p176
- 信頼領域、信頼楕円p176
- 典型性検定p179
- 同質性検定p179
- 並び替え検定p180
 - ここで言及されているFisher1935（『実験計画法』）の第3章は、p36～の21「さらに後半な仮説の検定」の部分。Fisher1936は未だ確認してません。

注目するのは個体空間の座標

- MCAの結果の個体座標のデータセット（1～1215）に（追加変数である）性別（Gender）、年齢（Age）、収入（Income）の列を追加する。

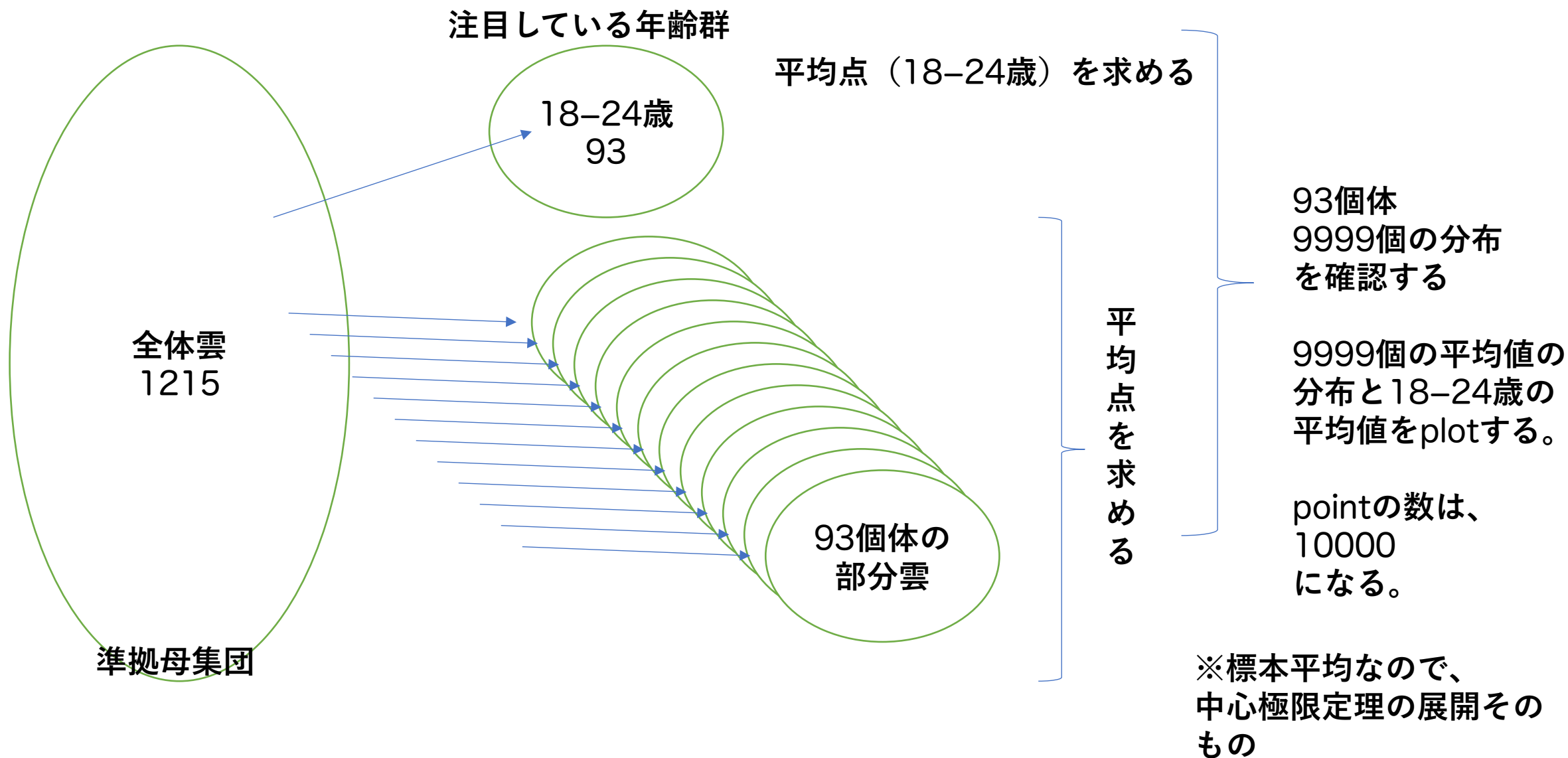
	coord.dim.1	coord.dim.2	coord.dim.3	Gender	Age	Income
1	0.1353094	0.9019845	0.43230469	女性	55-64	£20-29
2	1.2024366	-0.3285633	-0.26380752	女性	45-54	<£9
3	-0.5370107	0.3337334	0.56487632	女性	55-64	<£9
4	0.2136402	0.4512257	-1.11413703	女性	65+	£10-19
5	1.1699785	-1.1955279	-0.06591472	女性	35-44	£10-19
6	0.7225735	-1.4161871	-0.37615842	女性	18-24	<£9
⋮			⋮			
⋮			⋮			
⋮			⋮			

追加変数のカテゴリで
dim.nの部分空間をつくり
その関係进行分析します。

ということは、分散分析
してもいいのでは。
その結果とIDAの関係を比較
することもできそう。
(追って…)

MCAはなにをしているのか

- 嗜好データの変数カテゴリは29個
 - つまり29次元
- これが、次元縮減によって3次元で81%の情報を扱える
 - 2軸までで70%、3軸までで81%
- こうして生成された空間（個体空間、変数空間）の座標軸が新たな「変数」として位置付けられる。
 - この軸の+方向、-方向の解釈は、新たな変数になる。
- この空間を目的「変数」（被説明「空間」）を追加変数によって分析していく、という段取り。
 - active変数、追加変数の設定が構造化モデリング。
- ここで明らかになった、追加変数カテゴリの位置が検定される。



ただし、この分布はシンプル！

- 先に確認したように、抽出した93個体の部分雲の平均点の分布（標本分布）は、中心極限定理によって、**正規分布で近似できる**。
- 平均はゼロ。
- 分散は、
$$V = \frac{1}{n} \frac{N-n}{N-1} \lambda$$
- ここで $\frac{N-n}{N-1}$ は、有限母集団修正。

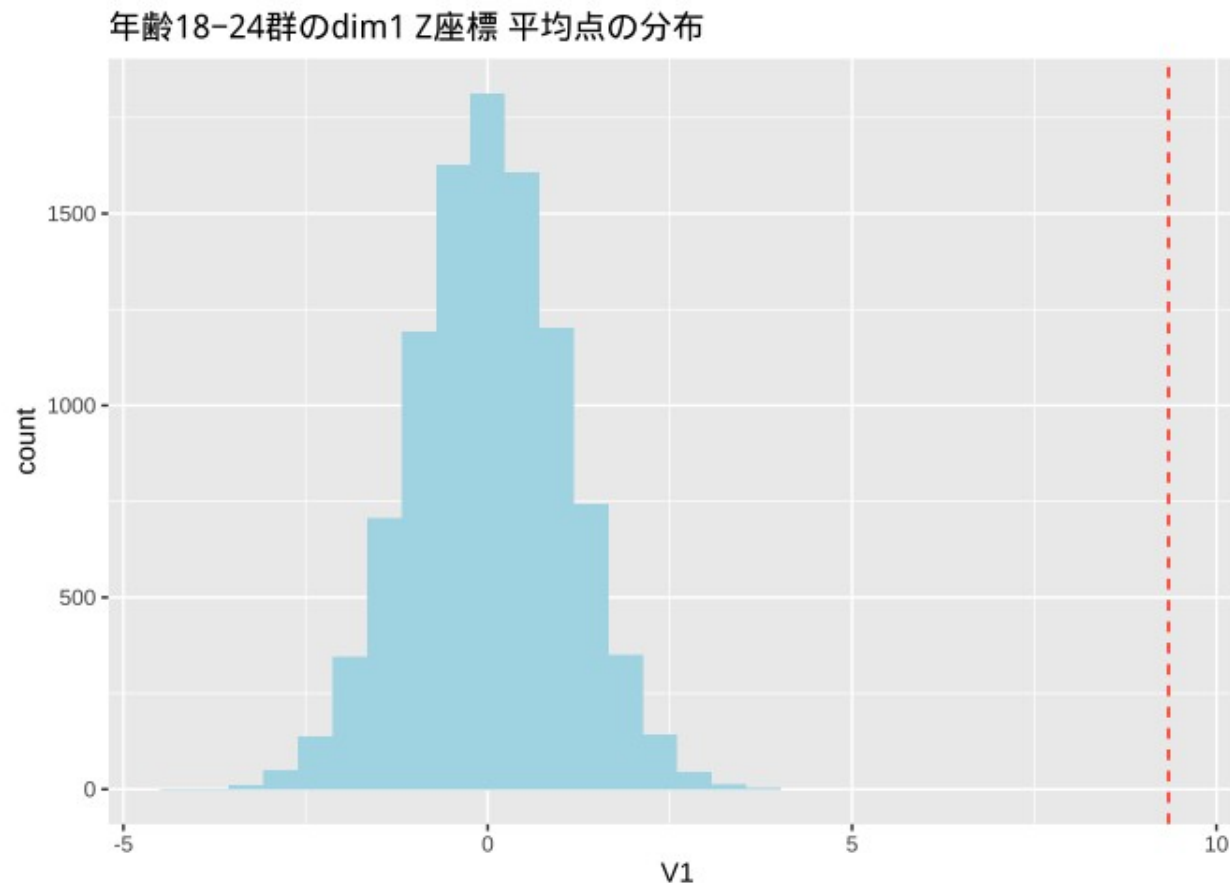
東大基礎統計学I『統計学入門』p189～「9.4 有限母集団と有限母集団修正」

典型性検定

- 前回報告したスライドを再掲します。
- ここでやっていることは、いわゆるリサンプリング。
- 並べ替え検定でのリサンプリングは、非復元抽出
 - 二つの部分集合 ($n1$, $n2$) を統合して（プールして、というらしい）、 $n1$ 個を抽出するので、非復元抽出でいいかなとは思いますが、
- 典型性検定では、参照母集団を考えるときに、（たとえば） $1/100$ の杉並区のサンプルが得られているなら、それを100倍して「みなし母集団」として、リサンプリングすればいいように思うので、そのときは、復元抽出でもいいではないのか。

こういう分布に検定統計量を位置付ける

- 平均ゼロ、分散は、 N と n と λ （軸の固有値）から計算される。
- この正規分布に注目している部分集合の平均点の標準座標を位置付け、分散と比べてどのくらい離れているかを確認する。



分散はいくつになるのか

- $N=1215$ 、 $n=93$ 、 $\lambda=0.4004$ (Dim1)
- これを、 $V = \frac{1}{n} \frac{N-n}{N-1} \lambda$
- に代入すると
 - $((1/93) (1215 - 93) / (1215 - 1)) * 0.4004 = 0.4004 * (1/93) * 0.924$
 - $= 0.00397866$ (有限母集団修正あり)
- という値が得られる。
 λ/n が標本分散の値。それに有限母集団補正 (0.924) をかけている。
- 「18-24歳」の軸1の平均点の標準座標 (Z値なので) は、+9.34なので、これは、非常に有意ということになる。9.34 σ

で、「組合せ論枠組み」とはなんなのか

- 確率ではなく、割合でp値を出す。
- p値は、Neyman=Pearson体系のように判定基準ではなく、レベルを表すものとして解釈していく。
 - だから、非典型性の検定、非同質性の検定、ではなく、典型性検定、同質性検定、なのだろうか..
（シャピロウィクスの正規性検定のように、帰無仮説が「正規分布している」なので、正規性を（積極的には）確認できないので、組合せ論でできないか考えてみたが、そもそも正規性を仮定する必要がなかった..）
- Fisher派としては、p値によって「有意」が確認されたら、その先に検討に入る。（p値で有意が確認されたら、対立仮説が「正しい」ではないぞ、ということ。）

t-検定のalternativeということでしょうか

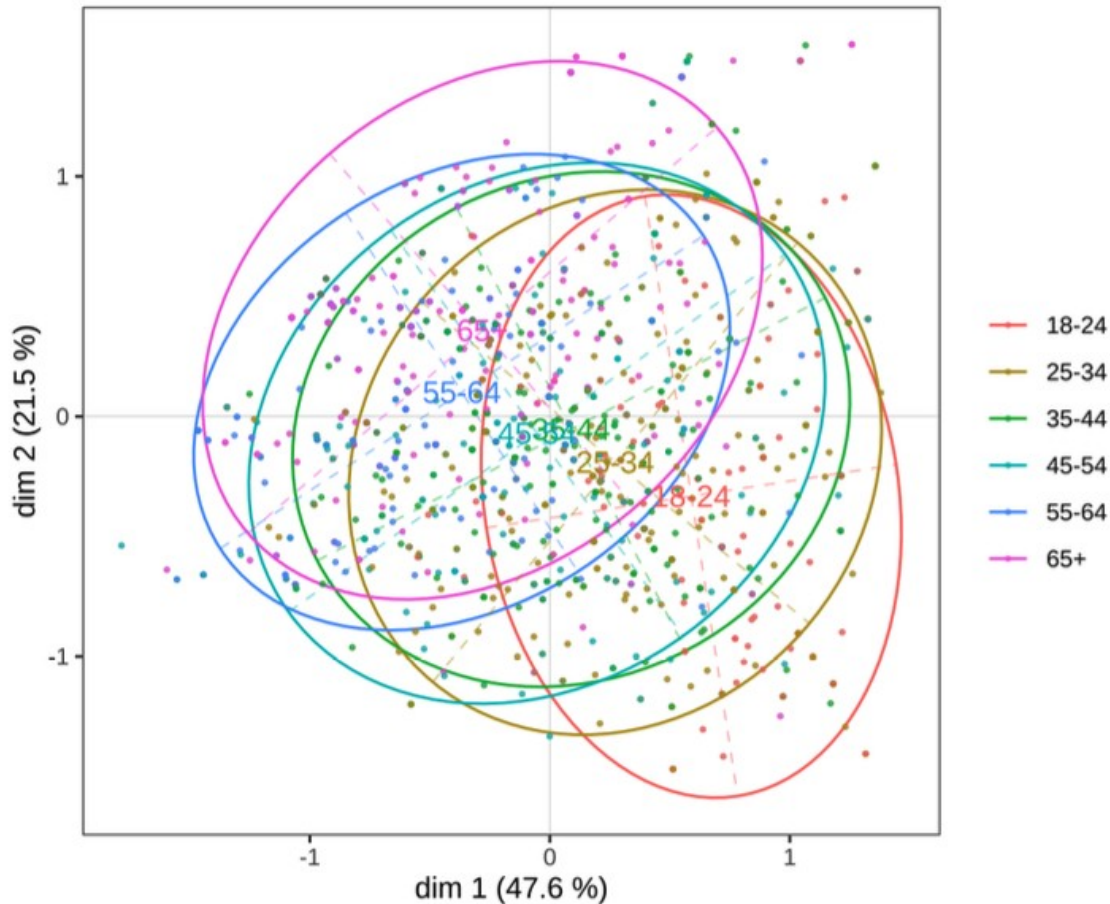
- 典型性検定
 - 参照母集団の平均とのズレ
- 同質性検定
 - 二つの部分集合の平均のズレ
- これを（もろもろの仮定を必要とする）「確率論」の枠組みではなく、記述統計のresultの割合で解釈していく。

典型性レベルを表す指標としての p 値？

- 第4章のデータで、典型性検定を有意（非典型）/非有意（典型ではないとは言えない）の「判定」ではなく、典型性レベルでみるとどう見えるかを見してみる。
- データは、嗜好データ（taste example）
- GDAtools2.0の`dimtypicality()`を使う。
 - （前回やっている…）

GDAtoolsV2.0で加わったfunction dimtypicality (前回のスライドから)

Ageの回答カテゴリの集中楕円



```
GDAtools::dimtypicality(res.speMCA,.d0[, "Age"], dim=c(1,2))
```

\$dim.1		weight	test.stat	p.value
Age.18-24	93	9.342901	0.00000	
Age.25-34	248	7.591076	0.00000	
Age.35-44	258	2.545227	0.01092	
Age.45-54	191	-1.279119	0.20086	
Age.65+	242	-7.703930	0.00000	
Age.55-64	183	-8.505159	0.00000	

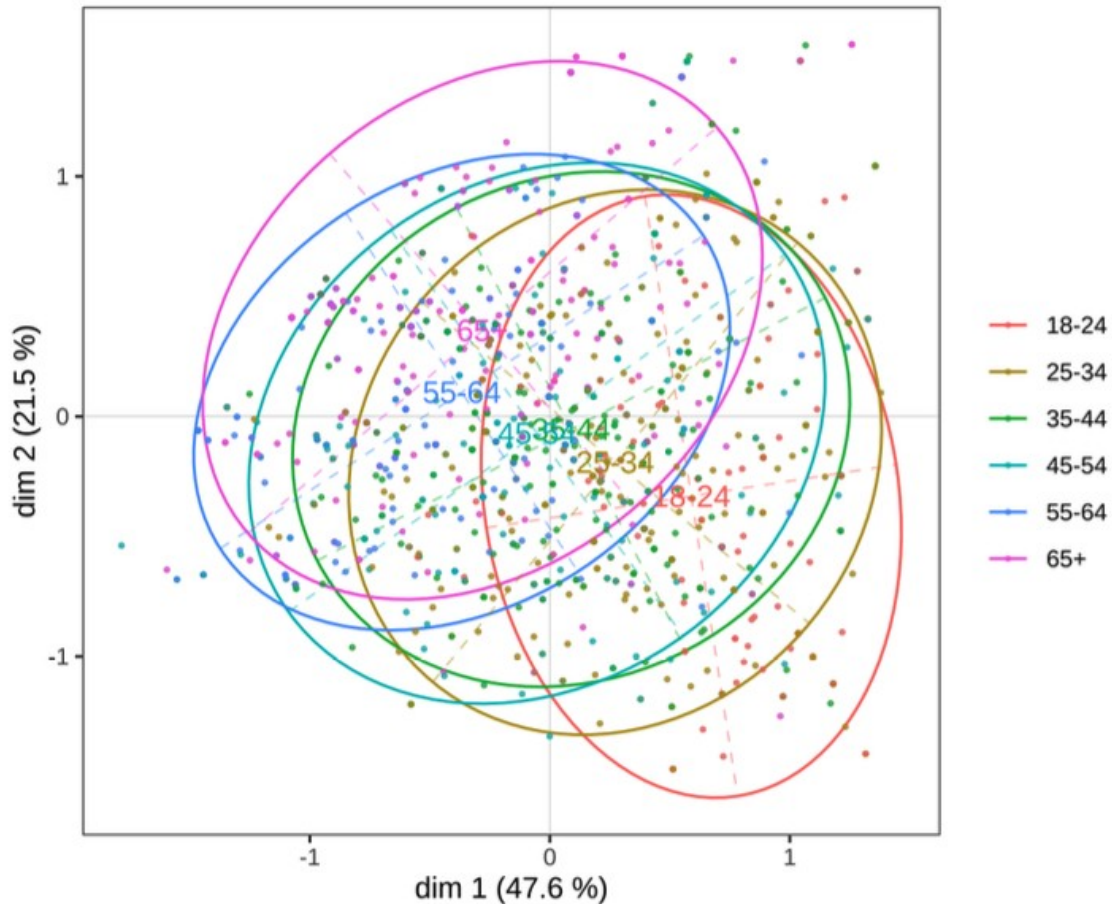
\$dim.2		weight	test.stat	p.value
Age.65+	242	10.515601	0.00000	
Age.55-64	183	2.502285	0.01234	
Age.35-44	258	-1.627232	0.10369	
Age.45-54	191	-1.769648	0.07679	
Age.18-24	93	-5.625920	0.00000	
Age.25-34	248	-5.680095	0.00000	

p117の+9.34はこの
test.stat：検定統計
量
です。
p値はゼロ。

コードを読んでみましたが、
使われているのは「近似計
算」つまり正規分布近似で
計算してました。
並べ替え計算をやるなら、
繰り返し数の設定などが必
要。それに、時間がかかり
ます！

GDAtoolsV2.0で加わったfunction dimtypicality (見どころ変更)

Ageの回答カテゴリの集中楕円



```
GDAtools::dimtypicality(res.speMCA,.d0[,"Age"],dim=c(1,2))
```

\$dim.1

	weight	test.stat	p.value
Age.18-24	93	9.342901	0.00000
Age.25-34	248	7.591076	0.00000
Age.35-44	258	2.545227	0.01092
Age.45-54	191	-1.279119	0.20086
Age.65+	242	-7.703930	0.00000
Age.55-64	183	-8.505159	0.00000

\$dim.2

	weight	test.stat	p.value
Age.65+	242	10.515601	0.00000
Age.55-64	183	2.502285	0.01234
Age.35-44	258	-1.627232	0.10369
Age.45-54	191	-1.769648	0.07679
Age.18-24	93	-5.625920	0.00000
Age.25-34	248	-5.680095	0.00000

統計検定量
(test.stat)
は、典型からのズレの
方向。
p値は、典型性レベル。
小さいほど、非典型、
つまり特徴あり。
大きいのは典型水準
大。

GDAtools::dimtypicalityはANOVA?

- `dimtypicality(resmca, vars, dim=c(1,2), max.pval=1)`

resmca MCAの

result

dim1	dim2	dim...

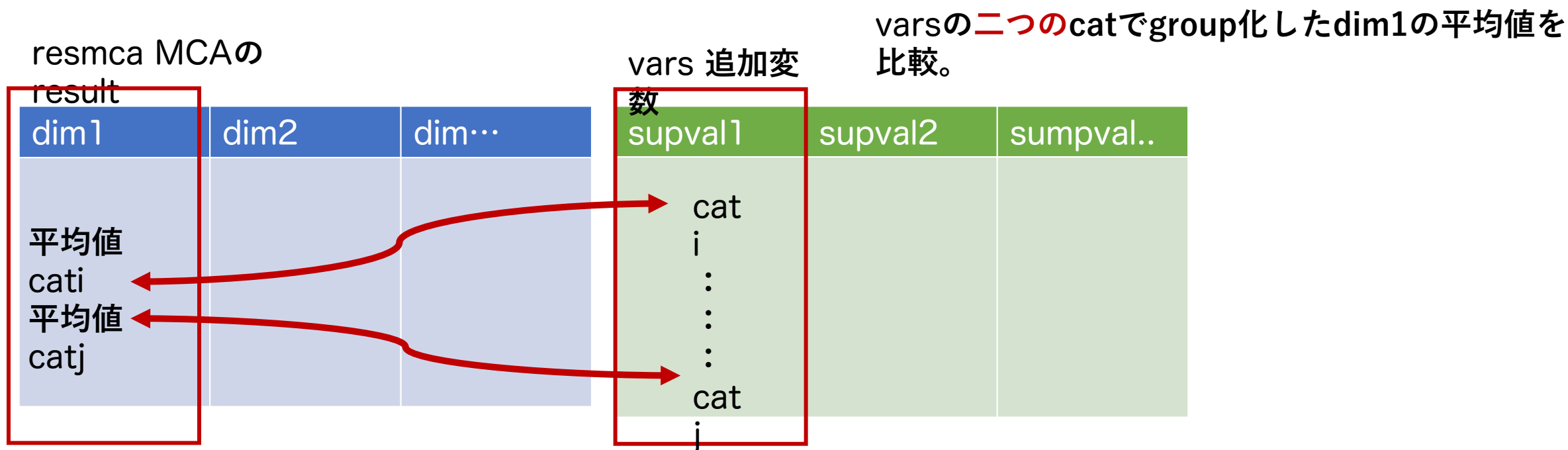
vars 追加変

supval1	supval2	sumpval..
cat 1		
cat 2		
cat 3		
:		

varsのcatでgroup化したdim1の平均値をdim1全体の平均値と比較。

同質性検定 `homog.test` はMANOVA？

- `homog.test(resmca, vars, dim=c(1, 2))`



まとめ

- t-検定、分散分析の非確率論的置き換え、と考えたら、使える場面がいくらでもありそうです。
- 典型性、同質性のレベル指標としp値を使う（仮説の判定、判断ではなく）ということが、「統計的推測を現在よりも自由に用いることができるし、また用いるべきである」p113、にいう「自由に」の意味として理解できそうです。