Data Science
Factor Correspondence Analysis

Joint latent analysis

Stéphane Marchand-Maillet

Department of Computer Science

Master en Sciences Informatiques - Autumn semester

Table of contents

What is this lecture about?

- ⋆ Linear latent models can be used for representing a population sample wrt its most informative features (PCA)
- ⋆ We use linear latent models as common traits of joint phenomenons to study them together
- ⋆ The data arises from the co-occurrence of the latent factors

Reading: [1] (Parts I and II)

Analyzing categorical data

- ⋆ A D-categorical variable Z takes values in D categories. Provided a population of N items, $z_i \in \mathbb{N}^D$ represents the number of items falling in category $i \in [\![D]\!]$
- ⋆ The joint occurrence of D_1- and D_2-categorical variables form the contingency table $\mathbf{X}$
- ⋆ $\mathbf{X} = \{\mathbf{X}_{ij}\}$ is a $D_1 \times D_2$ (integer) matrix and $\mathbf{X}_{ij}$ $(\in \mathbb{N})$ is the count of items from the population falling in categories i for the first variable and j for the second variable

Q: Given categorical data as contingency table $\mathbf{X} \in \mathbb{N}^{D_1 \times D_2}$, we wish to understand its structure, including for visualization

Note: A categorical variable is conceptually similar to a discrete r.v that takes one discrete value with some probability

Analyzing contingency tables

Example contingency table

X	Belmont	Cheseaux	Crissier	Epalinges	Jouxtens	Lausanne	Le Mont	Paudex	Prilly	Pully	Renens	Romanel	Margin ($\mathbf{r}$)
Aucune Formation	6	26	114	36	3	2126	23	11	251	73	244	15	2928
Scolarité Oblgatoire	344	677	2220	1401	150	40165	994	280	3491	3670	7039	556	60987
Formation Professionnelle	752	1116	1729	2253	252	39941	1486	476	4200	4721	5638	1029	63593
Maturité	163	128	249	554	51	10405	311	81	570	1465	888	126	14991
Formation Professionnelle Supérieure	155	135	211	497	65	5583	298	63	452	989	553	127	9128
Ecole Professionelle Supérieure	62	36	90	147	24	1709	111	21	131	306	195	52	2884
Université Haute école	196	96	169	675	110	9302	380	106	344	2010	437	84	13909
Autre	10	15	31	50	1	990	18	7	90	86	95	23	1416
Margin ($\mathbf{c}^T$)	1688	2229	4813	5613	656	110221	3621	1045	9529	13320	15089	2012	$N = 169836$

- ⋆ Given the $D_1 \times D_2$ data, we can look at marginal distributions
- ⋆ $\mathbf{r} = \mathbf{X}\mathbf{1}_{D_2}$ and $\mathbf{c} = \mathbf{X}^T\mathbf{1}_{D_2}$ are the row and column marginals
- ⋆ $\sum_{i=1}^{D_1} \mathbf{r}(i) = \mathbf{r}^T\mathbf{1}_{D_1} = \langle \mathbf{r}, \mathbf{1}_{D_1} \rangle = N$ so that $\langle \frac{1}{N}\mathbf{r}, \mathbf{1}_{D_1} \rangle = 1$
- ⋆ $\sum_{i=1}^{D_2} \mathbf{c}(i) = \mathbf{c}^T\mathbf{1}_{D_2} = \langle \mathbf{c}, \mathbf{1}_{D_2} \rangle = N$ so that $\langle \frac{1}{N}\mathbf{c}, \mathbf{1}_{D_2} \rangle = 1$

Note: Source : François Micheloud

Column-frequency table

We first consider rows as "features" and columns as "samples"

$\mathbf{C}^{\mathsf{T}}$	Belmont	Cheseaux	Crissier	Epalinges	Jouxtens	Lausanne	Le Mont	Paudex	Prilly	Pully	Renens	Romanel	$\mathbf{f}_c$
Aucune Formation	0.36%	1.17%	2.37%	0.64%	0.46%	1.93%	0.64%	1.05%	2.63%	0.55%	1.62%	0.75%	1.72%
Scolarité Oblgatoire	20.38%	30.37%	46.13%	24.96%	22.87%	36.44%	27.45%	26.79%	36.64%	27.55%	46.65%	27.63%	35.91%
Formation Professionnelle	44.55%	50.07%	35.92%	40.14%	38.41%	36.24%	41.04%	45.55%	44.08%	35.44%	37.36%	51.14%	37.44%
Maturité	9.66%	5.74%	5.17%	9.87%	7.77%	9.44%	8.59%	7.75%	5.98%	11.00%	5.89%	6.26%	8.83%
Formation Professionnelle Supérieure	9.18%	6.06%	4.38%	8.85%	9.91%	5.07%	8.23%	6.03%	4.74%	7.42%	3.66%	6.31%	5.37%
Ecole Professionnelle Supérieure	3.67%	1.62%	1.87%	2.62%	3.66%	1.55%	3.07%	2.01%	1.37%	2.30%	1.29%	2.58%	1.70%
Université Haute école	11.61%	4.31%	3.51%	12.03%	16.77%	8.44%	10.49%	10.14%	3.61%	15.09%	2.90%	4.17%	8.19%
Autre	0.59%	0.67%	0.64%	0.89%	0.15%	0.90%	0.50%	0.67%	0.94%	0.65%	0.63%	1.14%	0.83%
Margin	1	1	1	1	1	1	1	1	1	1	1	1	

- ⋆ We normalize $\mathbf{X}$ column-wise into $\mathbf{C}$ so that each row of $\mathbf{C}$ is a categorical distribution
- ⋆ If $\mathbf{D}_c = \mathrm{diag}(\mathbf{c})$ then $\mathbf{C} = \mathbf{D}_c^{-1}\mathbf{X}^{\mathsf{T}}$

⇒ so that $\mathbf{C}\mathbf{1}_{D_1} = \mathbf{1}\mathbf{1}_{D_2}$

Column-frequency table

We first consider rows as "features" and columns as "samples"

$\mathbf{C}^T$	Belmont	Cheseaux	Crissier	Epalinges	Jouxtens	Lausanne	Le Mont	Paudex	Prilly	Pully	Renens	Romanel	$\mathbf{f}_c$
Aucune Formation	0.36%	1.17%	2.37%	0.64%	0.46%	1.93%	0.64%	1.05%	2.63%	0.55%	1.62%	0.75%	1.72%
Scolarité Oblgatoire	20.38%	30.37%	46.13%	24.96%	22.87%	36.44%	27.45%	26.79%	36.64%	27.55%	46.65%	27.63%	35.91%
Formation Professionnelle	44.55%	50.07%	35.92%	40.14%	38.41%	36.24%	41.04%	45.55%	44.08%	35.44%	37.36%	51.14%	37.44%
Maturité	9.66%	5.74%	5.17%	9.87%	7.77%	9.44%	8.59%	7.75%	5.98%	11.00%	5.89%	6.26%	8.83%
Formation Professionnelle Supérieure	9.18%	6.06%	4.38%	8.85%	9.91%	5.07%	8.23%	6.03%	4.74%	7.42%	3.66%	6.31%	5.37%
École Professionelle Supérieure	3.67%	1.62%	1.87%	2.62%	3.66%	1.55%	3.07%	2.01%	1.37%	2.30%	1.29%	2.58%	1.70%
Université Haute école	11.61%	4.31%	3.51%	12.03%	16.77%	8.44%	10.49%	10.14%	3.61%	15.09%	2.90%	4.17%	8.19%
Autre	0.59%	0.67%	0.64%	0.89%	0.15%	0.90%	0.50%	0.67%	0.94%	0.65%	0.63%	1.14%	0.83%
Margin	1	1	1	1	1	1	1	1	1	1	1	1	

- ⋆ The column-marginal profile is $\mathbf{f}_c = \frac{1}{N}\mathbf{r}$ so that $\mathbf{f}_c^T \mathbf{1}_{D_1} = 1$

  The column-marginal profile provides the relative importance (density) of each (row) category (feature)

⇒ $\mathbf{M}^c_{\chi^2} = \mathrm{diag}(\mathbf{f}_c)^{-1}$ can be used as χ^2 metric for normalizing column frequencies [1]

- ⋆ Note: call $\mathbf{D}_r = \mathrm{diag}(\mathbf{r})$ then $\mathrm{diag}(\mathbf{f}_c) = \frac{1}{N}\mathbf{D}_r$ and $\mathbf{M}^c_{\chi^2} = N\mathbf{D}_r^{-1}$

Column-frequency table

We first consider rows as "features" and columns as "samples"

$\mathbf{C}^\mathsf{T}$	Belmont	Cheseaux	Crissier	Epalinges	Jouxtens	Lausanne	Le Mont	Paudex	Prilly	Pully	Renens	Romanel	$\mathbf{f}_\mathsf{c}$
Aucune Formation	0.36%	1.17%	2.37%	0.64%	0.46%	1.93%	0.64%	1.05%	2.63%	0.55%	1.62%	0.75%	1.72%
Scolarité Oblgatoire	20.38%	30.37%	46.13%	24.96%	22.87%	36.44%	27.45%	26.79%	36.64%	27.55%	46.65%	27.63%	35.91%
Formation Professionnelle	44.55%	50.07%	35.92%	40.14%	38.41%	36.24%	41.04%	45.55%	44.08%	35.44%	37.36%	51.14%	37.44%
Maturité	9.66%	5.74%	5.17%	9.87%	7.77%	9.44%	8.59%	7.75%	5.98%	11.00%	5.89%	6.26%	8.83%
Formation Professionnelle Supérieure	9.18%	6.06%	4.38%	8.85%	9.91%	5.07%	8.23%	6.03%	4.74%	7.42%	3.66%	6.31%	5.37%
Ecole Professionelle Supérieure	3.67%	1.62%	1.87%	2.62%	3.66%	1.55%	3.07%	2.01%	1.37%	2.30%	1.29%	2.58%	1.70%
Université Haute école	11.61%	4.31%	3.51%	12.03%	16.77%	8.44%	10.49%	10.14%	3.61%	15.09%	2.90%	4.17%	8.19%
Autre	0.59%	0.67%	0.64%	0.89%	0.15%	0.90%	0.50%	0.67%	0.94%	0.65%	0.63%	1.14%	0.83%
Margin	1	1	1	1	1	1	1	1	1	1	1	1	

- ⋆ The row-marginal profile is $\mathbf{f}_\mathsf{r} = \frac{1}{N}\mathbf{c}$ so that $\mathbf{f}_\mathsf{r}^\mathsf{T}\mathbf{1}_{D_2} = 1$

  The row-marginal profile provides the relative importance (density) of each (column) category (sample)

⇒ We use $\boldsymbol{W}^\mathsf{c} = \mathrm{diag}(\mathbf{f}_\mathsf{r}) = \mathrm{diag}(\frac{1}{N}\mathbf{c}) = \frac{1}{N}\mathbf{D}_\mathsf{c}$ to preserve population weighting

Column-frequency table

We first consider rows as “features” and columns as “samples”

$\mathbf{C}^T$	Belmont	Cheseaux	Crissier	Epalinges	Jouxtens	Lausanne	Le Mont	Paudex	Prilly	Pully	Renens	Romanel	$\mathbf{f}_c$
Aucune Formation	0.36%	1.17%	2.37%	0.64%	0.46%	1.93%	0.64%	1.05%	2.63%	0.55%	1.62%	0.75%	1.72%
Scolarité Oblgatoire	20.38%	30.37%	46.13%	24.96%	22.87%	36.44%	27.45%	26.79%	36.64%	27.55%	46.65%	27.63%	35.91%
Formation Professionnelle	44.55%	50.07%	35.92%	40.14%	38.41%	36.24%	41.04%	45.55%	44.08%	35.44%	37.36%	51.14%	37.44%
Maturité	9.66%	5.74%	5.17%	9.87%	7.77%	9.44%	8.59%	7.75%	5.98%	11.00%	5.89%	6.26%	8.83%
Formation Professionnelle Supérieure	9.18%	6.06%	4.38%	8.85%	9.91%	5.07%	8.23%	6.03%	4.74%	7.42%	3.66%	6.31%	5.37%
Ecole Professionelle Supérieure	3.67%	1.62%	1.87%	2.62%	3.66%	1.55%	3.07%	2.01%	1.37%	2.30%	1.29%	2.58%	1.70%
Université Haute école	11.61%	4.31%	3.51%	12.03%	16.77%	8.44%	10.49%	10.14%	3.61%	15.09%	2.90%	4.17%	8.19%
Autre	0.59%	0.67%	0.64%	0.89%	0.15%	0.90%	0.50%	0.67%	0.94%	0.65%	0.63%	1.14%	0.83%
Margin	1	1	1	1	1	1	1	1	1	1	1	1	

Putting all together, we construct the column-covariance matrix

$$\boldsymbol{\Sigma}^c = \underbrace{\mathbf{M}^c_{\chi^2}}_{\text{frequency normalization}} \underbrace{\mathbf{C}^T \mathbf{W}^c \mathbf{C}}_{\text{weighted covariance}} = \mathbf{D}_r^{-1} \mathbf{X} \mathbf{D}_c^{-1} \mathbf{X}^T$$

Row-frequency table

We then transpose the table: columns as “features” and rows as “samples”

$\mathbf{R}$	Belmont	Cheseaux	Crissier	Epalinges	Jouxtens	Lausanne	Le Mont	Paudex	Prilly	Pully	Renens	Romanel	Margin
Aucune Formation	0.20%	0.89%	3.89%	1.23%	0.10%	72.61%	0.79%	0.38%	8.57%	2.49%	8.33%	0.51%	1
Scolarité Oblgatoire	0.56%	1.11%	3.64%	2.30%	0.25%	65.86%	1.63%	0.46%	5.72%	6.02%	11.54%	0.91%	1
Formation Professionnelle	1.18%	1.75%	2.72%	3.54%	0.40%	62.81%	2.34%	0.75%	6.60%	7.42%	8.87%	1.62%	1
Maturité	1.09%	0.85%	1.66%	3.70%	0.34%	69.41%	2.07%	0.54%	3.80%	9.77%	5.92%	0.84%	1
Formation Professionnelle Supérieure	1.70%	1.48%	2.31%	5.44%	0.71%	61.16%	3.26%	0.69%	4.95%	10.83%	6.06%	1.39%	1
Ecole Professionnelle Supérieure	2.15%	1.25%	3.12%	5.10%	0.83%	59.26%	3.85%	0.73%	4.54%	10.61%	6.76%	1.80%	1
Université Haute école	1.41%	0.69%	1.22%	4.85%	0.79%	66.88%	2.73%	0.76%	2.47%	14.45%	3.14%	0.60%	1
Autre	0.71%	1.06%	2.19%	3.53%	0.07%	69.92%	1.27%	0.49%	6.36%	6.07%	6.71%	1.62%	1
$\mathbf{f}_r^{\top}$	0.99%	1.31%	2.83%	3.30%	0.39%	64.90%	2.13%	0.62%	5.61%	7.84%	8.88%	1.18%	

- ⋆ We normalize $\mathbf{X}$ row-wise into $\mathbf{R}$ so that each row of $\mathbf{R}$ is a categorical distribution
- ⋆ If $\mathbf{D}_r = \mathrm{diag}(\mathbf{r})$ then $\mathbf{R} = \mathbf{D}_r^{-1}\mathbf{X}$

⇒ so that $\mathbf{R}\mathbf{1}_{D_2} = 1\mathbf{1}_{D_1}$

Row-frequency table

We then transpose the table: columns as "features" and rows as "samples"

$\mathbb{R}$	Belmont	Cheseaux	Crissier	Epalinges	Jouxtens	Lausanne	Le Mont	Paudex	Prilly	Pully	Renens	Romanel	Margin
Aucune Formation	0.20%	0.89%	3.89%	1.23%	0.10%	72.61%	0.79%	0.38%	8.57%	2.49%	8.33%	0.51%	1
Scolarité Oblgatoire	0.56%	1.11%	3.64%	2.30%	0.25%	65.86%	1.63%	0.46%	5.72%	6.02%	11.54%	0.91%	1
Formation Professionnelle	1.18%	1.75%	2.72%	3.54%	0.40%	62.81%	2.34%	0.75%	6.60%	7.42%	8.87%	1.62%	1
Maturité	1.09%	0.85%	1.66%	3.70%	0.34%	69.41%	2.07%	0.54%	3.80%	9.77%	5.92%	0.84%	1
Formation Professionnelle Supérieure	1.70%	1.48%	2.31%	5.44%	0.71%	61.16%	3.26%	0.69%	4.95%	10.83%	6.06%	1.39%	1
Ecole Professionnelle Supérieure	2.15%	1.25%	3.12%	5.10%	0.83%	59.26%	3.85%	0.73%	4.54%	10.61%	6.76%	1.80%	1
Université Haute école	1.41%	0.69%	1.22%	4.85%	0.79%	66.88%	2.73%	0.76%	2.47%	14.45%	3.14%	0.60%	1
Autre	0.71%	1.06%	2.19%	3.53%	0.07%	69.92%	1.27%	0.49%	6.36%	6.07%	6.71%	1.62%	1
$\mathbf{f}_r^T$	0.99%	1.31%	2.83%	3.30%	0.39%	64.90%	2.13%	0.62%	5.61%	7.84%	8.88%	1.18%	

- ⋆ The row-marginal profile is $\mathbf{f}_r = \frac{1}{N}\mathbf{c}$ so that $\mathbf{f}_r^T \mathbf{1}_{D_2} = 1$

  The row-marginal profile provides the relative importance (density) of each (column) category

⇒ $\mathbf{M}^r_{\chi^2} = \text{diag}(\mathbf{f}_r)^{-1}$ can be used as χ^2 metric for normalizing frequencies

- ⋆ Note: call $\mathbf{D}_c = \text{diag}(\mathbf{c})$ then $\text{diag}(\mathbf{f}_r) = \frac{1}{N}\mathbf{D}_c$ and $\mathbf{M}^r_{\chi^2} = N\mathbf{D}_c^{-1}$

Row-frequency table

We then transpose the table: columns as “features” and rows as “samples”

$\mathbb{R}$	Belmont	Cheseaux	Crissier	Epalinges	Jouxtens	Lausanne	Le Mont	Paudex	Prilly	Pully	Renens	Romanel	Margin
Aucune Formation	0.20%	0.89%	3.89%	1.23%	0.10%	72.61%	0.79%	0.38%	8.57%	2.49%	8.33%	0.51%	1
Scolarité Oblgatoire	0.56%	1.11%	3.64%	2.30%	0.25%	65.86%	1.63%	0.46%	5.72%	6.02%	11.54%	0.91%	1
Formation Professionnelle	1.18%	1.75%	2.72%	3.54%	0.40%	62.81%	2.34%	0.75%	6.60%	7.42%	8.87%	1.62%	1
Maturité	1.09%	0.85%	1.66%	3.70%	0.34%	69.41%	2.07%	0.54%	3.80%	9.77%	5.92%	0.84%	1
Formation Professionnelle Supérieure	1.70%	1.48%	2.31%	5.44%	0.71%	61.16%	3.26%	0.69%	4.95%	10.83%	6.06%	1.39%	1
Ecole Professionelle Supérieure	2.15%	1.25%	3.12%	5.10%	0.83%	59.26%	3.85%	0.73%	4.54%	10.61%	6.76%	1.80%	1
Université Haute école	1.41%	0.69%	1.22%	4.85%	0.79%	66.88%	2.73%	0.76%	2.47%	14.45%	3.14%	0.60%	1
Autre	0.71%	1.06%	2.19%	3.53%	0.07%	69.92%	1.27%	0.49%	6.36%	6.07%	6.71%	1.62%	1
f_r^T	0.99%	1.31%	2.83%	3.30%	0.39%	64.90%	2.13%	0.62%	5.61%	7.84%	8.88%	1.18%	

- ⋆ The column-marginal profile is $\mathbf{f}_c = \frac{1}{N}\mathbf{r}$ so that $\mathbf{f}_c^T \mathbf{1}_{D_1} = 1$

  The column-marginal profile provides the relative importance (density) of each (row) category (sample)

⇒ We use $\boldsymbol{W}^r = \mathrm{diag}(\mathbf{f}_c) = \mathrm{diag}(\frac{1}{N}\mathbf{r}) = \frac{1}{N}\mathbf{D}_r$ to preserve population weighting

Row-frequency table

We then transpose the table: columns as "features" and rows as "samples"

$\mathbb{R}$	Belmont	Cheseaux	Crissier	Epalinges	Jouxtens	Lausanne	Le Mont	Paudex	Prilly	Pully	Renens	Romanel	Margin
Aucune Formation	0.20%	0.89%	3.89%	1.23%	0.10%	72.61%	0.79%	0.38%	8.57%	2.49%	8.33%	0.51%	1
Scolarité Oblgatoire	0.56%	1.11%	3.64%	2.30%	0.25%	65.86%	1.63%	0.46%	5.72%	6.02%	11.54%	0.91%	1
Formation Professionnelle	1.18%	1.75%	2.72%	3.54%	0.40%	62.81%	2.34%	0.75%	6.60%	7.42%	8.87%	1.62%	1
Maturité	1.09%	0.85%	1.66%	3.70%	0.34%	69.41%	2.07%	0.54%	3.80%	9.77%	5.92%	0.84%	1
Formation Professionnelle Supérieure	1.70%	1.48%	2.31%	5.44%	0.71%	61.16%	3.26%	0.69%	4.95%	10.83%	6.06%	1.39%	1
Ecole Professionnelle Supérieure	2.15%	1.25%	3.12%	5.10%	0.83%	59.26%	3.85%	0.73%	4.54%	10.61%	6.76%	1.80%	1
Université Haute école	1.41%	0.69%	1.22%	4.85%	0.79%	66.88%	2.73%	0.76%	2.47%	14.45%	3.14%	0.60%	1
Autre	0.71%	1.06%	2.19%	3.53%	0.07%	69.92%	1.27%	0.49%	6.36%	6.07%	6.71%	1.62%	1
$\mathbf{f}_r^T$	0.99%	1.31%	2.83%	3.30%	0.39%	64.90%	2.13%	0.62%	5.61%	7.84%	8.88%	1.18%	

Putting all together, we construct the row-covariance matrix

$$\boldsymbol{\Sigma}^r = \underbrace{\mathbf{M}^r_{\chi^2}}_{\text{frequency normalization}} \underbrace{\mathbf{R}^T \mathbf{W}^r \mathbf{R}}_{\text{weighted covariance}} = \mathbf{D}_c^{-1} \mathbf{X}^T \mathbf{D}_r^{-1} \mathbf{X}$$

Factor Correspondence Analysis (FCA)

- ⋆ Factor Correspondence Analysis performs PCA on row-wise and column-wise normalized matrices $\mathbf{R}$ and $\mathbf{C}$ respectively

⇒ Decomposition of row- and column covariance matrices $\mathbf{\Sigma}^{\mathsf{r}}$ and $\mathbf{\Sigma}^{\mathsf{c}}$

$$\mathbf{\Sigma}^{\mathsf{r}} = \mathbf{D}_{\mathsf{c}}^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{D}_{\mathsf{r}}^{-1}\mathbf{X} \qquad \mathbf{\Sigma}^{\mathsf{c}} = \mathbf{D}_{\mathsf{r}}^{-1}\mathbf{X}\mathbf{D}_{\mathsf{c}}^{-1}\mathbf{X}^{\mathsf{T}}$$

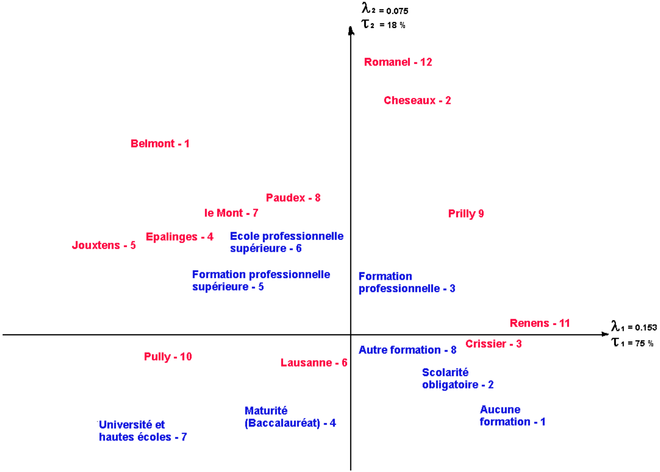
$$\Rightarrow \mathbf{\Sigma}^{\mathsf{r}} = \mathbf{U}_{\mathsf{r}}^{\mathsf{T}}\mathbf{\Lambda}_{\mathsf{r}}^{2}\mathbf{U}_{\mathsf{r}} \qquad \mathbf{\Sigma}^{\mathsf{c}} = \mathbf{U}_{\mathsf{c}}^{\mathsf{T}}\mathbf{\Lambda}_{\mathsf{c}}^{2}\mathbf{U}_{\mathsf{c}}$$

- ⋆ Duality between the two representations : same e.v ($\mathbf{\Lambda}_{\mathsf{r}} = \mathbf{\Lambda}_{\mathsf{c}}$) and the projection of rows (resp. columns) along column (resp. row) axes $\mathbf{u}_k$ is identical modulo factor $\sqrt{\lambda_k}$.
- ⋆ The first e.v $\lambda_1 = 1$ is ignored
- ⋆ FCA searches for a space for quantifying symbolic data and respect their correlation as much as possible

Note: The data is not centered

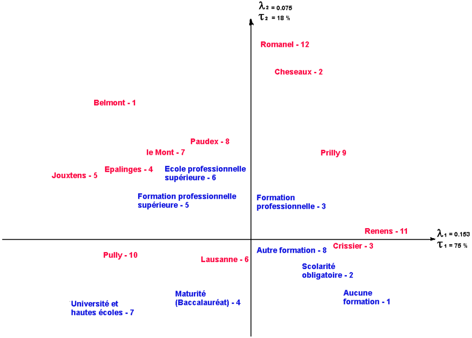
Visualization

Scatter plot



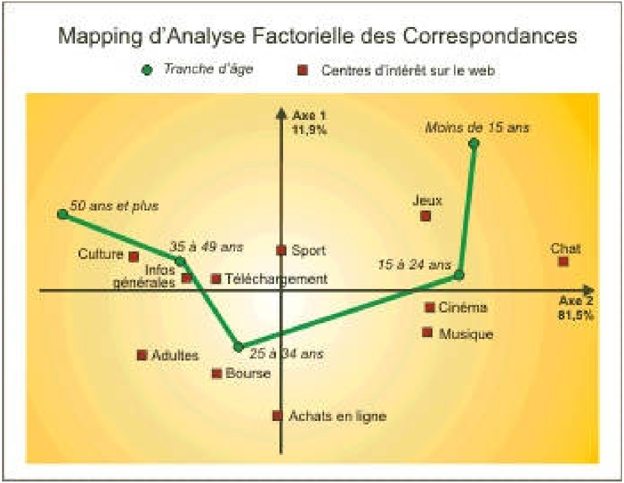
Source : Francois Micheloud

Interpretations



- ⋆ Close points from the same profile indicate similar profiles
- ⋆ Close points from different profiles should be carefully analyzed
- ⋆ Angles between different profiles indicate factor correlation (attractive if $< 90°$, repulsive if $> 90°$)
- ⋆ Angles between points and axes indicate their correlation

Web usage *vs* age



Summary

- ⋆ Contingency tables arise from sensing joint phenomenons
- ⋆ They can be studied via (joint) linear latent factor analysis
- ⋆ These joint views can be superimposed (modulo factors)
- ⋆ The aggregation of their joint views is not direct and requires careful interpretation

⇒ Can be generalized to multi-way correspondence analysis

⇒ Used in poll (questionnaire) analysis

Example questions [mostly require formal – mathematical – answers]

- ⋆ What is a contingency table? Look for an example other than those given here
- ⋆ What are the specificities of such data?
- ⋆ How to compute row- and column-profiles?
- ⋆ How/why do we change metric in linear analysis (eg χ^2)?
- ⋆ Can we directly superimpose dual latent analysis?
- ⋆ Is the role of latent factors the same is in PCA?

References I

[1] Eric J. Beh and Rosaria Lombardo. *Correspondence Analysis: Theory, Practice and New Strategies*. Wiley, 2014.

[2] Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2004.