D- interprétation d'une ACP

1- choix du nombre d'axes

3- interprétation des axes

2- représentation graphique

4- Qualité de représentation

5- Synthèse

D-Interprétation d'une ACP

- La décomposition précédente est faite par des logiciels statistiques (Statistica, R , SAS….).

- Non automatique (interprétation des résultats):

  - ✓ choisir le nombre q d'axes factoriels (ou de composantes principales) à retenir pour obtenir un résumé suffisamment précis de l'information contenue dans le tableau initial

  - ✓ construire les graphiques

  - ✓ Donner une signification aux nouvelles variables.

  - ✓ Evaluer la qualité de ce résumé

On dispose de 6 variables représentant les taux de différents délits commis pour 100000 habitants dans 20 Etats des Etats-unis. Ces données peuvent être mises dans un tableau individu*variable

ETAT	Meurtre	Rapt	vol	attaque	viol	larcin
Alabama	14.2	25.2	96.8	278.3	1135.5	1881.9
Alaska	10.8	51.6	96.8	284.0	1331.7	3369.8
Arizona	9.5	34.2	138.2	312.3	2346.1	4467.4
Arkansas	8.8	27.6	83.2	203.4	972.6	1862.1
California	11.5	49.4	287.0	358.0	2139.4	3499.8
Colorado	6.3	42.0	170.7	292.9	1935.2	3903.2
Connecticut	4.2	16.8	129.5	131.8	1346.0	2620.7
Delaware	6.0	24.9	157.0	194.2	1682.6	3678.4
Florida	10.2	39.6	187.9	449.1	1859.9	3840.5
Georgia	11.7	31.1	140.5	256.5	1351.1	2170.2
Hawaii	7.2	25.5	128.0	64.1	1911.5	3920.4
Idaho	5.5	19.4	39.6	172.5	1050.8	2599.6
Illinois	9.9	21.8	211.3	209.0	1085.0	2828.5
Indiana	7.4	26.5	123.2	153.5	1086.2	2498.7
Iowa	2.3	10.6	41.2	89.8	812.5	2685.1
Kansas	6.6	22.0	100.7	180.5	1270.4	2739.3
Kentucky	10.1	19.1	81.1	123.3	872.2	1662.1
Louisiana	15.5	30.9	142.9	335.5	1165.5	2469.9
Maine	2.4	13.5	38.7	170.0	1253.1	2350.7
Maryland	8.0	34.8	292.1	358.9	1400.0	3177.7

D-1 Choix du nombre d’axes à retenir

- **Deux critères empiriques pour sélectionner le nombre d’axes :**

  - Critère du coude : sur l’eboulis des valeurs propres, on observe un décrochement (coude) suivi d’une décroissance régulière. On sélectionne les axes avant le décrochement
  - Critère de Kaiser: on ne retient que les axes dont l’inertie est supérieure à l’inertie moyenne I/p (un peu étroit).

    <u>Kaiser en ACP normée</u>: I/p= 1 : On ne retiendra que les axes associés à des valeurs propre supérieures à 1

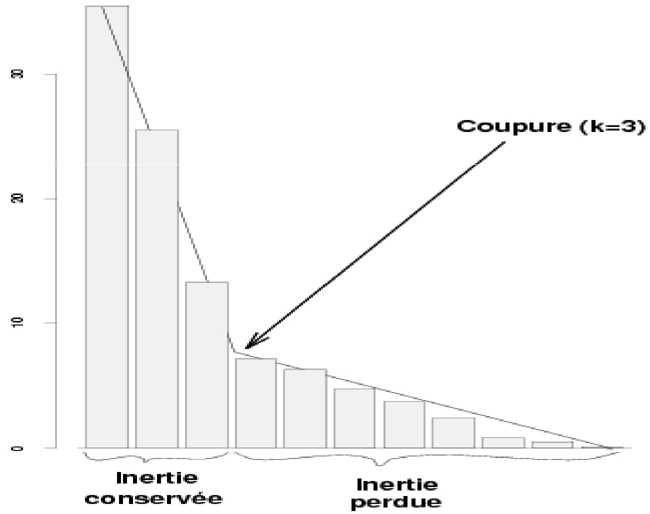
  - **Dans la pratique, on retient en fait les q axes que l’on sait interpréter**

Rq: Critère du Scree-test : on sélectionne les axes correspondant à des différences secondes >0 (un peu large)

Analyse des sorties de l'ACP

Choix du nombre de facteurs

But : obtenir le maximum d'inertie conservée avec le minimum de facteurs.
Règle du "coude" : on coupe l'éboulis des valeurs propres à l'endroit où celui-ci possède un "coude".



Le choix dépend des objectifs de l'analyse :
description des données $\longrightarrow$ en général au plus 4 ou 5 facteurs (difficultés d'interprétation).
Compression ou recodage $\longrightarrow$ ce nombre peut-être très grand.

D-1 Choix du nombre d’axes

- ✓ Critère de Kaiser : nous conduit à retenir 2 axes, expliquant 82% de l’inertie totale.

Importance of components:

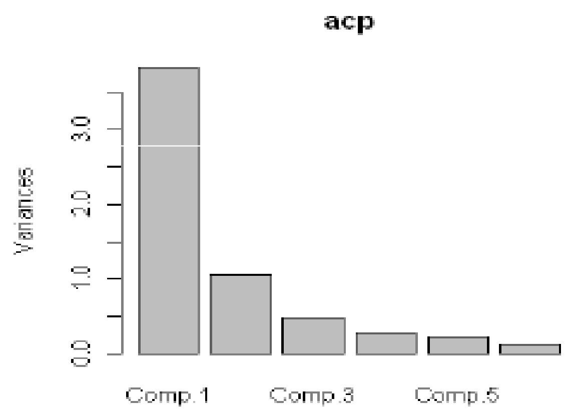
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	1.8670647	1.1924148	0.68759281	0.54252803	0.46761698	0.32623640
Proportion of Variance	0.5809884	0.2369755	0.07879731	0.04905611	0.03644427	0.01773836
Cumulative Proportion	0.5809884	0.8179639	0.89676125	0.94581736	0.98226164	1.00000000

$\sqrt{\lambda_k}$

I_k / I

$I = 6 = \sum \lambda_k$

- ✓ Critère du coude :Décrochement au troisième axe, puis décroissance régulière à partir du troisième axe : seuls les **deux premiers axes** présentent un éventuel intérêt.



D-1 Choix du nombre d’axes à retenir (ou du nombre de composantes principales)

<u>Conclusion :</u>

- ✓ Le premier axe conserve 58% de l’inertie du nuage. Il est peu probable qu’il soit dû au hasard. Il existe une structuration importante des données qui va se manifester sur le premier axe (1/2 au lieu de 1/6).
- ✓ Le second axe conserve une part importante de l’inertie totale, 24%.
- ✓ La chute est importante dès le troisième axe qui ne conserve plus que 7% de l’inertie totale (<seuil moyen 1/6=17%)

On peut décider de ne retenir que les deux premiers axes (le premier plan factoriel) car il compréhensible par l’œil (c’est un plan) et ne déforme pas trop le nuage (il explique 82% de l’inertie du nuage)

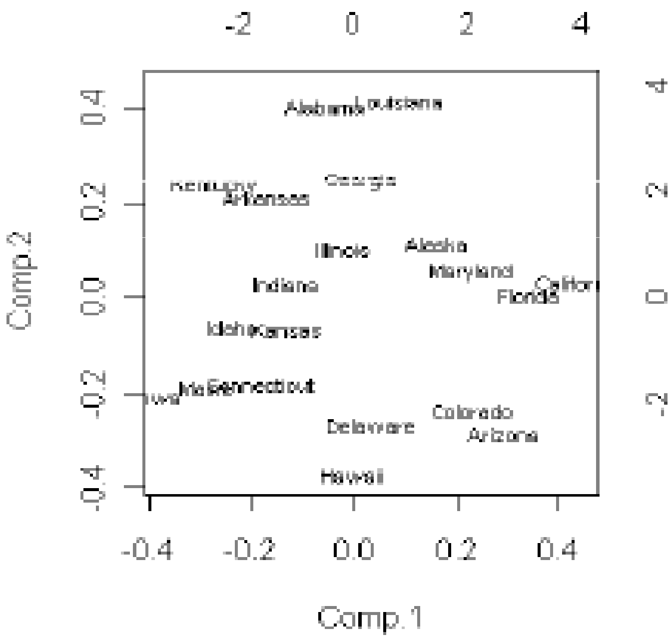
D-2 Construction des nuages de points projetés

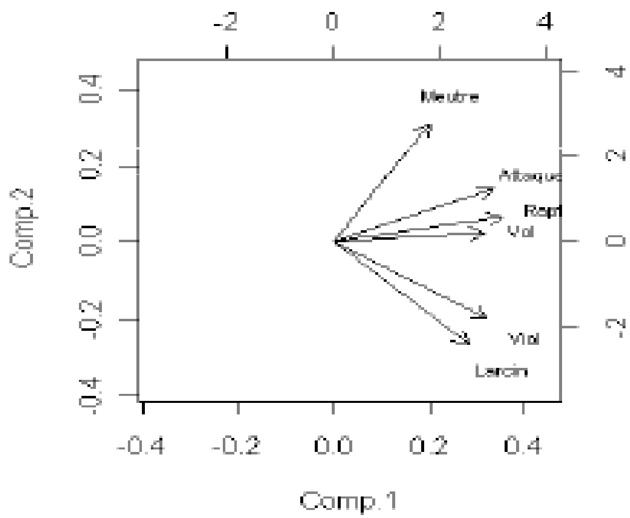
- ✓ Chaque nuage de points (variables et individus) est construit en projection sur les **plans factoriels** : un plan factoriel est un repère du plan défini par deux des q axes factoriels retenus.

  Ex : Si l'on retient 3 axes, on tracera 3 graphiques pour chaque nuage: le nuage projeté sur le plan (axe1, axe2), celui projeté sur le plan (axe1, axe3), celui projeté sur le plan (axe2,axe3).

- ✓ L'examen des plans factoriels permettra de visualiser les corrélations entre les variables et d'identifier les groupes d'individus ayant pris des valeurs proches sur certaines variables. Mais il faut avant de lire directement les graphiques interpréter les axes et s'assurer que la projection est fidele a la realite (voir d-4)

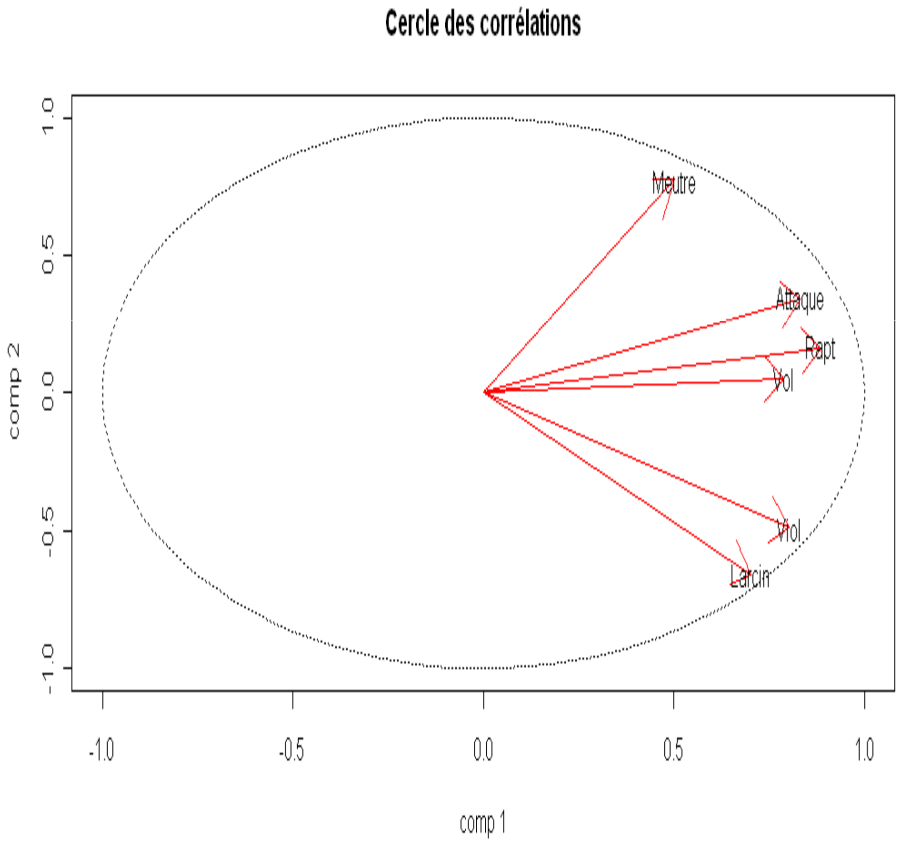
D-2 Construction des nuages de points projetés (SOUS R)



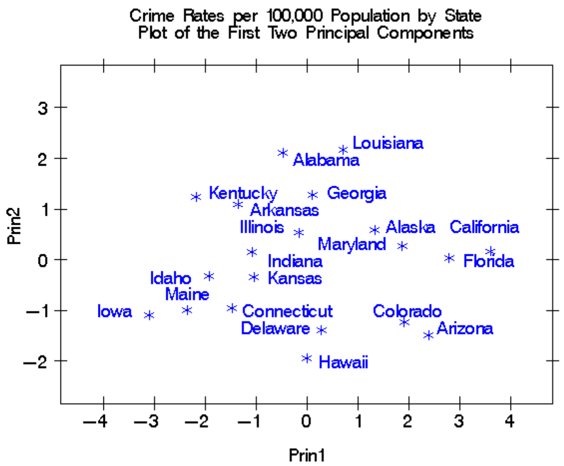


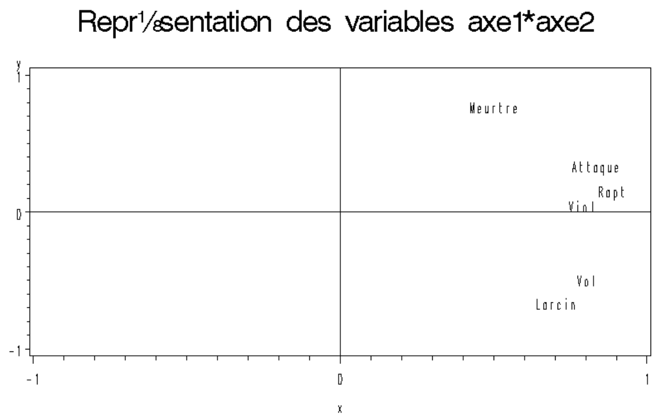
D-2 Construction des nuages de points projetés (SOUS R)

```
a=seq(0,2*pi,length=100)
plot( cos(a), sin(a), type='l',
    lty=3,xlab='comp 1', ylab='comp 2',
    main="Cercle des corrélations" )
v =t(acp$loadings)[1:2,]
arrows(0,0, acp$sdev[1]*v[1,],
    acp$sdev[2]*v[2,], col='red')
text(acp$sdev[1]*v[1,],
    acp$sdev[2]*v[2,],labels=colnames(v)
    )
```



D-2 Construction des nuages de points projetés
(SOUS SAS)





D-2 Construction des nuages de points projetés

✓ Deux types de facteurs :

- Effet taille : les variables sont toutes du même coté de l'axe. (i.e. elles contribuent toutes dans le même sens à la formation de l'axe)

- Effet forme : Deux groupes de variables opposées : celles qui contribuent positivement à l'axe, celles qui contribuent négativement.

D-3 Interprétation des axes

✓ **Interprétation des axes**

Pour chaque axe retenu et chaque nuage, on regarde

- Quelles sont les variables qui participent le plus à la formation de l'axe
- Quels sont les individus qui participent le plus à la formation de l'axe

Outil de mesure : contributions des points (individus si non anonymes et variables) à l'inertie de cet axe.

Ce sont les points dont la contribution est supérieure à la moyenne qui permettent de donner un sens à l'axe.

D-3 Interprétation des axes : nuage des points individus

Contribution de l'individu i à l'inertie de l'axe k :

$$\boxed{CTR_k(e_i) = \frac{p_i c_{ik}^2}{\lambda_k}}$$

Somme des contributions des individus = 100%.

- En pratique: On retient pour l'interprétation les individus dont la contribution est > à la contribution moyenne (>1/n), le sens de la contribution dépend du signe de cik.

- CP (poids égaux): les individus contribuent d'autant + que cik grand en v.a. Contribution importante :

$$|c_{ik}| > \sqrt{\lambda_k}$$

D-3 Interprétation des axes : nuage des points variables

Contribution de la variable j à l'inertie de l'axe k :

Somme=100%

$$\boxed{CTR_k(X_j) = \frac{d^2_{jk}}{\lambda_k} = u_{jk}^{\;2}}$$

•En pratique: On retient pour l'interprétation les variables dont la contribution est > à la contribution moyenne (>1/p), $|u_{jk}| > 1/\sqrt{p}$ le sens de la contribution dépend du signe de ujk.

•CP en ACP normée, ce sont les variables qui sont proches du bord du cercle qui contribuent le plus : $\boxed{d^2_{jk} = r^2(C_k, X_j)}$

D-3 Interprétation des axes : synthèse

A noter

- ✓ Une contribution trop importante d’un des points à un axe doit être regardé avec prudence (~25% d’inertie) .

- ✓ Il faut s’assurer que les points contribuant le plus à l’axe sont bien représentés sur l’axe (sinon il faut les mettre en éléments supplémentaires.)

- ✓ La contribution est juste une aide à l’interprétation :
  - La contribution de certains points peuvent être très légèrement inférieures au seuil et mais conforter l’interprétation de l’axe qua l’on aurait faite sans eux. On les inclut alors dans l’interpr étation.
  - Inversement, lorsqu’une contribution est très forte par rapport à d’autre qui sont pourtant en dessus du seuil, le point détermine l’axe presque exclusivement

D-3 Interprétation des axes : synthèse

- L’interprétation des nouvelles variables (des axes factoriel) se fera à l’aide des individus et variables contribuant le plus à l’axe avec la règle suivante : si une variable a une forte contribution positive à l’axe, les individus ayant une forte contribution positive à l’axe sont caractérisés par une valeur élevée de la variable.

D-3 Interprétation des axes : exemple

- **Interprétation de l'axe 1 : Contribution des individus**

$$\left|c_{i1}\right| > \sqrt{\lambda_1} = 1{,}86$$

Etat	Prin1	Prin2
Iowa	-3.08934	-1.08465
Maine	-2.34364	-0.98693
Kentucky	-2.17767	1.24149
Idaho	-1.91969	-0.31927
Connecticut	-1.47135	-0.94414
Arkansas	-1.35301	1.09825
Indiana	-1.07052	0.15404
Kansas	-1.04117	-0.32997
Alabama	-0.46221	2.11791
Illinois	-0.15469	0.54064
Hawaii	0.00534	-1.93088
Georgia	0.11528	1.28686
Delaware	0.29111	-1.38315
Louisiana	0.71862	2.17636
Alaska	1.33963	0.59409
Maryland	1.87662	0.28145
Colorado	1.91888	-1.22840
Arizona	2.39884	-1.48610
Florida	2.79934	0.03244
California	3.61964	0.16998

D-3 Interprétation des axes : exemple

- acp$scores[,1]

```
    Alabama      Alaska     Arizona    Arkansas  California    Colorado
-0.47421533  1.37443010  2.46115288 -1.38815961  3.71367458  1.96872562
Connecticut    Delaware     Florida     Georgia      Hawaii       Idaho
-1.50957496  0.29867735  2.87206179  0.11827435  0.00547835 -1.96955979
   Illinois     Indiana        Iowa      Kansas    Kentucky   Louisiana
-0.15871175 -1.09833052 -3.16959921 -1.06821737 -2.23424734  0.73729346
      Maine    Maryland
-2.40452228  1.92536969
```

D-3 Interprétation des axes : exemple

Interprétation de l'axe 1 : Contribution des variables

Eigenvectors

	Prin1	Prin2
Meurtre	0.268358	0.648880
Rapt	0.474074	0.134920
Viol	0.421853	0.045097
Attaque	0.445704	0.287959
Vol	0.429817	-.411955
Larcin	0.376675	-.553255

$$\left|u_{jk}\right| > 1/\sqrt{p} = 0,408$$

D-3 Interprétation des axes : exemple

- loadings(acp)[,1]

```
  Meutre       Rapt        Vol   Attaque      Viol    Larcin
0.2683577 0.4740738 0.4218529 0.4457038 0.4298167 0.3766750
```

D-3 Interprétation des axes : exemple

Interprétation axe 1

- Individus:

-	+
Iowa, Maine, Kentuky, Idaho	Californie, Floride Arizona, Maryland, Colorado

Variables :

-	+
	rapt, attaque, vol et viol

Conclusion : L'axe 1 isole les délits rapt, attaque, vol et viol. En réalité, isole l'ensemble des variables sur sa partie positive (effet taille). C'est un axe taux de délits en tout genre. : il oppose les états de Iowa, Maine, Kentuky, Idaho aux états de Californie, Floride Arizona, Maryland, Colorado, marqués par une forte proportion de délits.

D-3 Interprétation des axes : exemple

- **Interprétation de l'axe 2 : Contribution des individus**

$$\left|c_{i2}\right| > \sqrt{\lambda_2} = 1{,}19$$

Etat	Prin1	Prin2
Iowa	-3.08934	-1.08465
Maine	-2.34364	-0.98693
Kentucky	-2.17767	1.24149
Idaho	-1.91969	-0.31927
Connecticut	-1.47135	-0.94414
Arkansas	-1.35301	1.09825
Indiana	-1.07052	0.15404
Kansas	-1.04117	-0.32997
Alabama	-0.46221	2.11791
Illinois	-0.15469	0.54064
Hawaii	0.00534	-1.93088
Georgia	0.11528	1.28686
Delaware	0.29111	-1.38315
Louisiana	0.71862	2.17636
Alaska	1.33963	0.59409
Maryland	1.87662	0.28145
Colorado	1.91888	-1.22840
Arizona	2.39884	-1.48610
Florida	2.79934	0.03244
California	3.61964	0.16998

➢ acp$scores[,2]

```
    Alabama      Alaska     Arizona    Arkansas  California    Colorado
 2.17292554  0.60952764 -1.52470179  1.12678123  0.17439369 -1.26030699
Connecticut    Delaware     Florida     Georgia      Hawaii       Idaho
-0.96866341 -1.41908466  0.03328554  1.32028630 -1.98104269 -0.32756870
   Illinois     Indiana        Iowa      Kansas    Kentucky   Louisiana
 0.55468133  0.15803896 -1.11283017 -0.33854617  1.27374136  2.23289720
      Maine    Maryland
-1.01257392  0.28875974
```

D-3 Interprétation des axes : exemple

Interprétation de l’axe 2 : Contribution des variables

Eigenvectors

	Prin1	Prin2
Meurtre	0.268358	0.648880
Rapt	0.474074	0.134920
Viol	0.421853	0.045097
Attaque	0.445704	0.287959
Vol	0.429817	-.411955
Larcin	0.376675	-.553255

$$\left|u_{jk}\right| > 1/\sqrt{p} = 0,408$$

```
> loadings(acp)[,2]
   Meutre      Rapt       Vol    Attaque       Viol     Larcin
 0.6488797  0.1349197  0.0450971  0.2879586 -0.4119546 -0.5532552
```

D-3 Interprétation des axes : exemple

Interprétation de l'axe 2

- Individus

-	+
Hawaii, Delaware, Colorado, Arizona	Kentuky, Alabama, Géorgie, Louisiane

- Variable

-	+
Vol larcin	meurtre

Conclusion : L'axe 2 est un axe de gravité des délits: il oppose les états d'Hawaii, Delaware Colorado et Arizona, caractérisés par un taux élevé de délits mineurs aux états de Kentuky, Alabama Géorgie et Louisiane, marqués par un taux relativement élevé de meutres.

D-4 Etude des proximités entre points

- ✓ **Qualité de représentation des points**

Une fois les axes interprétés, on peut regarder les graphiques et analyser plus finement les proximités entre points.

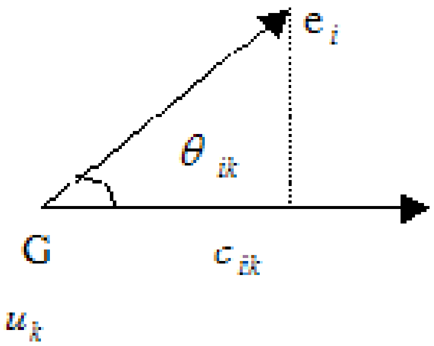
- ➢ Les proximités entre points observées sur un axe ou un plan factoriel doivent correspondent à la réalité (et non être artificiellement créées par l'opération de projection).

- ➢ Pour pouvoir interpréter les proximités entre points, il faut qu'ils soient bien représentés sur l'axe ou le plan en question

Un point est dit bien représenté sur un axe ou un plan factoriel si il est proche de sa projection sur l'axe ou le plan. S'il est éloigné, on dit qu'il est mal représenté. Indicateur =angle formé entre le point et sa projection sur l'axe

D-4 Etude des proximités entre points

- ✓ Qualité de représentation de l'individu i sur l'axe k:



$$\boxed{qlt_k(e_i) = \cos^2(\theta_{ik}) = \frac{c_{ik}{}^2}{\|e_i\|^2}}$$

$$\boxed{\|e_i\|^2 = \sum_{k=1}^{p} c_{ik}{}^2}$$

➢ Lorsque l'angle est proche de 0, c'est-à-dire que le cosinus est proche de 1, l'individu est bien représenté. Dans le cas inverse, l'angle est proche de 90° et le cosinus est proche de 0.

D-4 Etude des proximités entre points

- ✓ Qualité de représentation de la variable j sur l'axe k:

$$\boxed{qlt_k(X_j) = \cos^2(\theta_{kj}) = \frac{d_{jk}^{\ 2}}{\|X_j\|^2}}$$

- ✓ <u>En ACP normée</u>,

$$\boxed{qlt_k(X_j) = d^2_{\ jk} = r^2(C_k, X_j)}$$

  - ➢ **une variable est d'autant mieux représentée sur un axe qu'elle est proche du bord du cercle des corrélations et de l'axe, d'autant plus mal représentée qu'elle est proche de l'origine**.

  - ➢ les variables qui contribuent le plus à l'axe sont aussi celles qui sont le mieux représentées et inversement, donc pas besoin d'étude spécifique de la représentativité.

D-4 Etude des proximités entre points

- ✓ Qualité s de représentation sur un plan factoriel

- Individus : Le cosinus carré est est additif sur des sous-espaces orthogonaux. La qualité de représentation sur le plan défini par les axes k et l est égale à

$$qlt_{kl}(e_i) = qlt_k(e_i) + qlt_l(e_i)$$

- Variables: idem. En ACP normée, on interprète les proximités de variables bien représentées sur le plan i.e. proches du bord du cercle de corrélations

D-4 Etude des proximités entre points

✓ **Analyse des proximités**

L’examen des qualités de représentation

- permet de mettre en évidence des proximités éventuelles que l’on n’a pas remarquées Lors de l’interprétation des axes. On interprète les proximités d’éléments bien représentés sur le plan factoriel

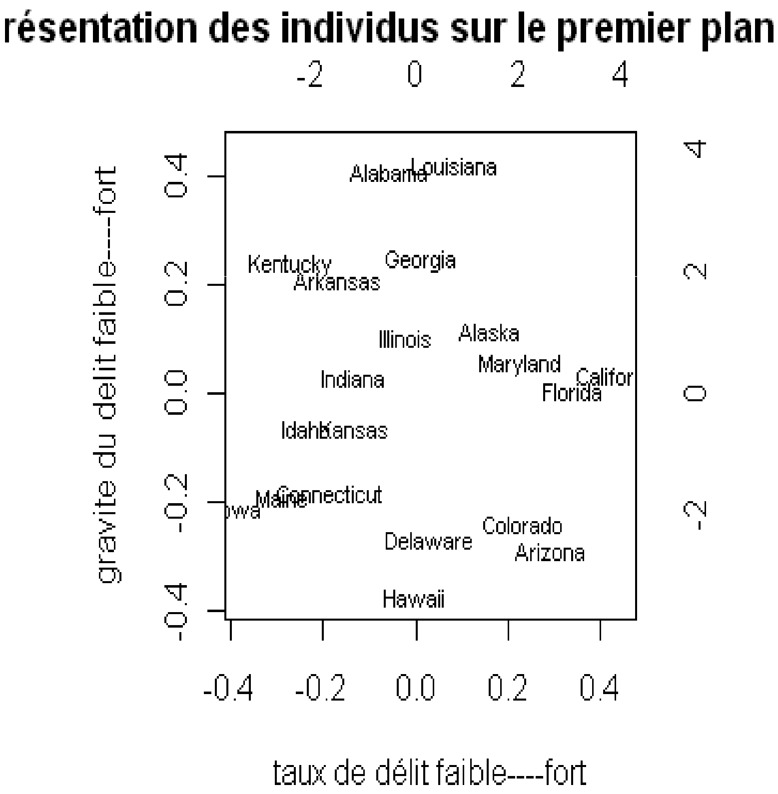
- Permet de repérer les points qui ne contribuent pas fortement à l’inertie de l’axe, mais qui sont bien représentés par cet axe, c'est-à-dire qui présentent des caractéristiques propres à l’axe.

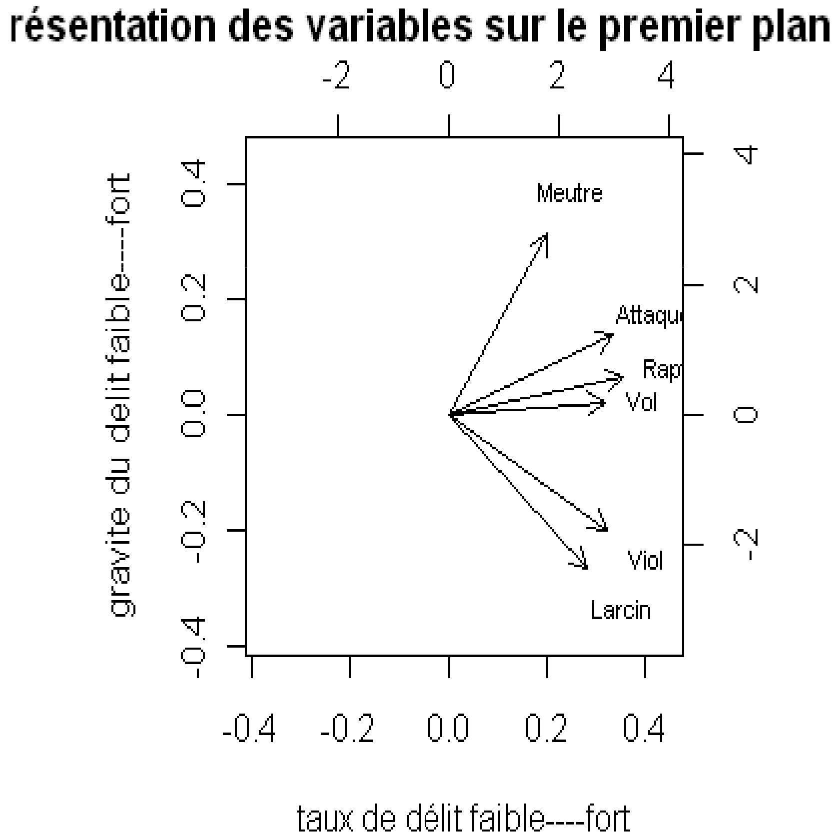
D-4 Etude des proximités entre points

- La proximité dans l'espace entre deux individus bien représentés traduit la ressemblance réelle de ces deux individus du point de vue des valeurs prises par les variables. (Lorsque la qualité de représentation de deux individus est bonne, leur proximité observée retrace leur proximité réelle dans l'espace)

  Rappel : La lecture directe des proximités sur le graphique peut donc s'avérer erronée (pas d'interprétation des proximités entre individus mal représentés).

- La proximité entre deux variables sur un axe donne, si les deux variables sont bien représentées sur l'axe (proches de l'axe et du bord du cercle) , une approximation de leur corrélation.
  - **Deux variables proches sont corrélées positivement**
  - **Deux variables qui s'opposent sont corrélées négativement**
  - **Deux variables orthogonales sont non corrélées**.


résentation des individus sur le premier plan
gravite du delit faible---fort
taux de délit faible----fort
Alabama
Louisiana
Kentucky
Georgia
Arkansas
Illinois
Alaska
Maryland
Califor
Indiana
Florida
Idaho
Kansas
Connecticut
owa
Maine
Colorado
Delaware
Arizona
Hawaii


résentation des variables sur le premier plan
gravite du delit faible---fort
taux de délit faible----fort
Meutre
Attaqu
Rap
Vol
Viol
Larcin

D-5 Synthèse

- ✓ Définition des composantes principales
- ✓ Synthèse globale des proximités des points sur les plans factoriels.
- ✓ Construction éventuelle du tableau « réduit » C de dimension n*q : ses lignes sont les valeurs prises par les n individus sur les q composantes principales retenues. La k° composante principale aura la même signification que le k° axe.

D-6 Exemple

ETAT	Prin1	Prin2	QLT1	QLT2
Iowa	-3.08934	-1.08465	0.85597	0.10551
Maine	-2.34364	-0.98693	0.73178	0.12977
Kentucky	-2.17767	1.24149	0.71555	0.23256
Idaho	-1.91969	-0.31927	0.86602	0.02395
Connecti	-1.47135	-0.94414	0.59764	0.24608
Arkansas	-1.35301	1.09825	0.54374	0.35826
Indiana	-1.07052	0.15404	0.75524	0.01564
Kansas	-1.04117	-0.32997	0.90431	0.09083
Alabama	-0.46221	2.11791	0.03970	0.83364
Illinois	-0.15469	0.54064	0.01044	0.12747
Hawaii	0.00534	-1.93088	0.00000	0.62744
Georgia	0.11528	1.28686	0.00694	0.86522
Delaware	0.29111	-1.38315	0.03901	0.88054
Louisian	0.71862	2.17636	0.08889	0.81527
Alaska	1.33963	0.59409	0.31012	0.06099
Maryland	1.87662	0.28145	0.50093	0.01127
Colorado	1.91888	-1.22840	0.65410	0.26806
Arizona	2.39884	-1.48610	0.59333	0.22771
Florida	2.79934	0.03244	0.86745	0.00012
Californ	3.61964	0.16998	0.93149	0.00205

D-6 Exemple

```
>x=acp$loadings
>QLT=x^2/matrix(rep(apply(x^2,1,sum),ncol(x^2)),dim(x^2));QLT[,1:2]
```

	Comp.1	Comp.2
Alabama	3.970450e-02	0.8336393299
Alaska	3.101184e-01	0.0609914688
Arizona	5.933328e-01	0.2277146491
Arkansas	5.437433e-01	0.3582567065
California	9.314876e-01	0.0020541441
Colorado	6.541030e-01	0.2680578029
Connecticut	5.976402e-01	0.2460799462
Delaware	3.900634e-02	0.8805363057
Florida	8.674495e-01	0.0001165112
Georgia	6.943376e-03	0.8652185301

D-6 Exemple

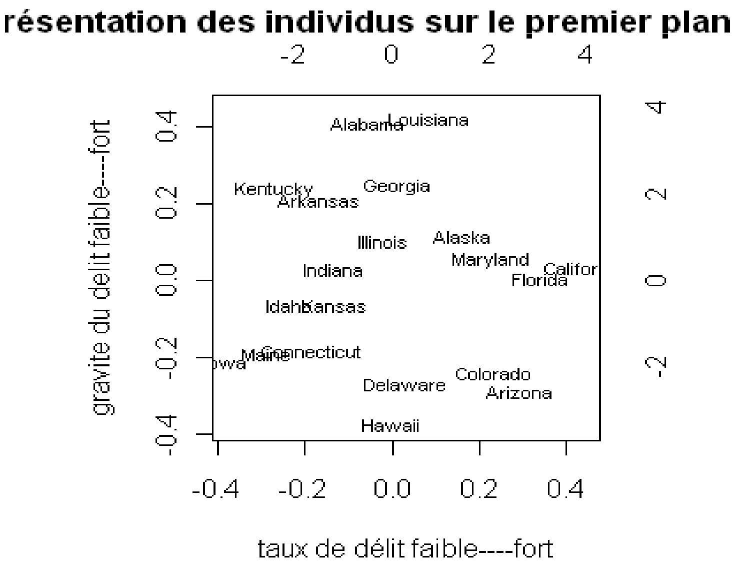
Qualité de représentation des individus sur le plan principal

Iowa	0,85597	0,10551	0,96148
Maine	0,73178	0,12977	0,86155
Kentucky	0,71555	0,23256	0,94811
Idaho	0,86602	0,02395	0,88997
Connecti	0,59764	0,24608	0,84372
Arkansas	0,54374	0,35826	0,902
Indiana	0,75524	0,01564	0,77088
Kansas	0,90431	0,09083	0,99514
Alabama	0,0397	0,83364	0,87334
Illinois	0,01044	0,12747	0,13791
Hawaii	0	0,62744	0,62744
Georgia	0,00694	0,86522	0,87216
Delaware	0,03901	0,88054	0,91955
Louisian	0,08889	0,81527	0,90416
Alaska	0,31012	0,06099	0,37111
Maryland	0,50093	0,01127	0,5122
Colorado	0,6541	0,26806	0,92216
Arizona	0,59333	0,22771	0,82104
Florida	0,86745	0,00012	0,86757
Californ	0,93149	0,00205	0,93354

D-6 Exemple

certains groupes de pays se détachent quant à leur comportement de délit :

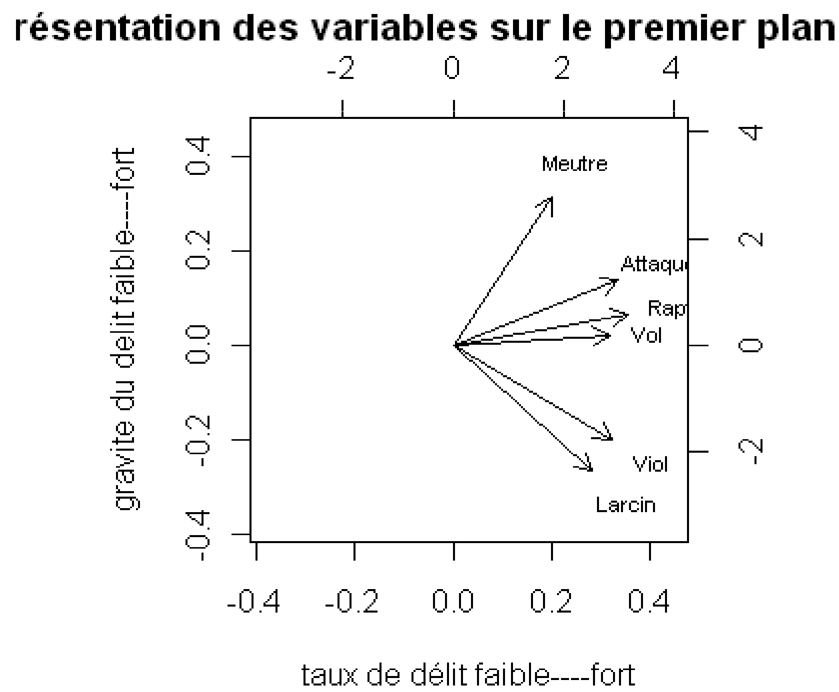
➢Californie, Floride : caractérisé par un fort taux de délits en tous genre, mais pas très différencié en ce qui concerne leur gravité



➢ Colorado, Arizona: fort taux de délits en tous genre, et particulièrement ceux de faible gravité

➢ Louisiane, Alabama Georgie, kentuky: taux de délit moyen en tout genre mais forte représentativité de meurtres

➢Iowa, maine : peu de délites et de faible gravité



D-6 Exemple

- Tableau C

Etat	Prin1	Prin2
Iowa	-3.08934	-1.08465
Maine	-2.34364	-0.98693
Kentucky	-2.17767	1.24149
Idaho	-1.91969	-0.31927
Connecticut	-1.47135	-0.94414
Arkansas	-1.35301	1.09825
Indiana	-1.07052	0.15404
Kansas	-1.04117	-0.32997
Alabama	-0.46221	2.11791
Illinois	-0.15469	0.54064
Hawaii	0.00534	-1.93088
Georgia	0.11528	1.28686
Delaware	0.29111	-1.38315
Louisiana	0.71862	2.17636
Alaska	1.33963	0.59409
Maryland	1.87662	0.28145
Colorado	1.91888	-1.22840
Arizona	2.39884	-1.48610
Florida	2.79934	0.03244
California	3.61964	0.16998

E- Limites

- ✓ Principale faiblesse de l'ACP:

- ✓ sensibilité aux points extrêmes. Ce manque de robustesse est notamment lie au rôle central qu'y joue le coefficient de corrélation : les points extrêmes, en perturbant les moyennes et corrélations, polluent fortement l'analyse - on peut cependant envisager de les déplacer en point supplémentaire.

- ✓ l'ACP est inadaptée aux phénomènes non linéaires qui plus est en grande dimension. Pour ce genre de problème, d'autres méthodes ont été développées, comme l'ACPN (Analyse en Composantes Principales par Noyau).