

TD5 : téléphones mobiles.

(Correction)

1 Introduction

La revue « 60 Millions de consommateurs » a présenté dans son numéro de novembre 2001 une étude sur les téléphones mobiles aboutissant à un tableau comportant un nombre important de caractéristiques touchant à l'utilisation de ces appareils. Une part de ces caractéristiques représentent des observations correspondant à des mesures objectives, associées le plus souvent à des jugements de valeur exprimés sur une échelle comportant cinq degrés.

Dans le but de fournir une cartographie de ces appareils pouvant servir à les différencier selon le maximum de ces critères il a été construit un tableau de données ci dessous, retenant les variables jugées les plus pertinentes :

```
> mobiles2
```

| | Commodite | Acoustiq | TpsCharge | AutoVeille | Autoconv | PuisGSM | SensiGSM | SensiDCS |
|--------------|-----------|----------|-----------|------------|----------|---------|----------|----------|
| Philips | 5 | 3 | 5 | 5 | 5 | 5 | 3 | 3 |
| Nokia3310 | 4 | 4 | 2 | 5 | 5 | 4 | 4 | 4 |
| Nokia6210 | 4 | 3 | 3 | 5 | 5 | 4 | 4 | 4 |
| Sony | 5 | 4 | 4 | 4 | 2 | 4 | 4 | 4 |
| SiemensSL45 | 5 | 3 | 5 | 5 | 1 | 4 | 4 | 4 |
| EricssonT20 | 3 | 4 | 3 | 5 | 5 | 3 | 3 | 3 |
| SiemensS35i | 4 | 4 | 3 | 5 | 2 | 4 | 4 | 4 |
| Alcatel304 | 3 | 4 | 4 | 5 | 2 | 5 | 3 | 3 |
| Samsung | 4 | 3 | 4 | 3 | 2 | 5 | 3 | 3 |
| Trium | 3 | 3 | 4 | 5 | 2 | 5 | 4 | 4 |
| Motorola | 3 | 4 | 3 | 5 | 2 | 4 | 4 | 4 |
| Panasonic | 4 | 4 | 5 | 4 | 3 | 3 | 1 | 2 |
| Alcatel501 | 4 | 2 | 4 | 5 | 2 | 5 | 2 | 3 |
| Ericsson2628 | 3 | 3 | 1 | 3 | 3 | 4 | 3 | 4 |
| Sagem | 3 | 3 | 3 | 5 | 3 | 3 | 1 | 2 |

Les appareils testés sont donc au nombre de 15 ; le numéro de modèle n'a été précisé que lorsque plusieurs téléphones d'une même marque sont présents. On ne donne volontairement pas d'explication sur le sens des variables, même si certaines sont simples à deviner : *l'analyse sera faite sur les propriétés numériques des variables, pas leur interprétation.*

Pour pouvoir utiliser l'ACP normée les degrés de l'échelle de jugement ont été transformés en notes allant de 1 à 5. Pour toutes ces données, les valeurs les plus grandes sont considérées comme meilleures.

On donne ci-dessous la matrice de corrélation des 8 variables :

```
> round(cor(mobiles2), digits=2)
```

| | Commodite | Acoustiq | TpsCharge | AutoVeille | Autoconv | PuisGSM | SensiGSM | SensiDCS |
|------------|-----------|----------|-----------|------------|----------|---------|----------|----------|
| Commodite | 1.00 | -0.12 | 0.54 | -0.03 | -0.01 | 0.17 | 0.21 | 0.15 |
| Acoustiq | -0.12 | 1.00 | -0.12 | 0.06 | 0.11 | -0.43 | 0.23 | 0.09 |
| TpsCharge | 0.54 | -0.12 | 1.00 | 0.19 | -0.30 | 0.25 | -0.18 | -0.36 |
| AutoVeille | -0.03 | 0.06 | 0.19 | 1.00 | 0.18 | -0.03 | 0.16 | 0.05 |
| Autoconv | -0.01 | 0.11 | -0.30 | 0.18 | 1.00 | -0.27 | -0.04 | -0.11 |
| PuisGSM | 0.17 | -0.43 | 0.25 | -0.03 | -0.27 | 1.00 | 0.34 | 0.29 |
| SensiGSM | 0.21 | 0.23 | -0.18 | 0.16 | -0.04 | 0.34 | 1.00 | 0.93 |
| SensiDCS | 0.15 | 0.09 | -0.36 | 0.05 | -0.11 | 0.29 | 0.93 | 1.00 |

Question 1. *Quel est le couple de variables qui sont les plus corrélées entre elles ? Les moins corrélées entre elles ? Quelles sont les variables les plus opposées ?*

Le couple le plus corrélé ($r \approx 1$) est (SensiDCS, SensiGSM) avec 0,93.

Le couple le moins corrélé ($r \approx 0$) est (AutoConv, Commodité) avec $-0,01$.

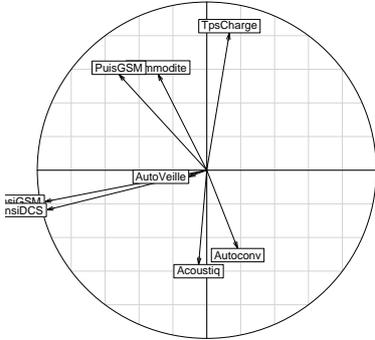
Le couple le plus opposé ($r \approx -1$) est (PuisGSM, Acoustiq) avec $-0,43$.

2 Une première analyse en composantes principales

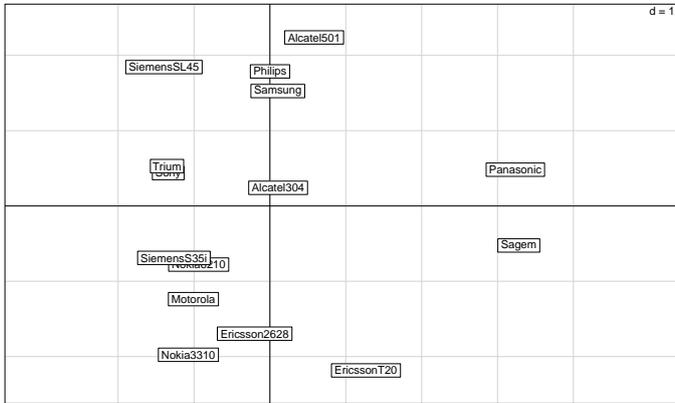
On fait une analyse en composantes principales des données centrées-réduites. On obtient les valeurs propres suivantes :

```
> round(pca2$eig, digits=2)
[1] 2.23 1.94 1.30 1.01 0.89 0.47 0.13 0.02
```

On représente ci-dessous le cercle des corrélations du premier plan principal (axe 1 en abscisse, axe 2 en ordonnée), accompagné du tableau des corrélations entre les variables et les 5 premiers facteurs, la projection des individus sur le premier plan principal et les contributions (en 10000ièmes) de ces individus aux 5 premiers facteurs :



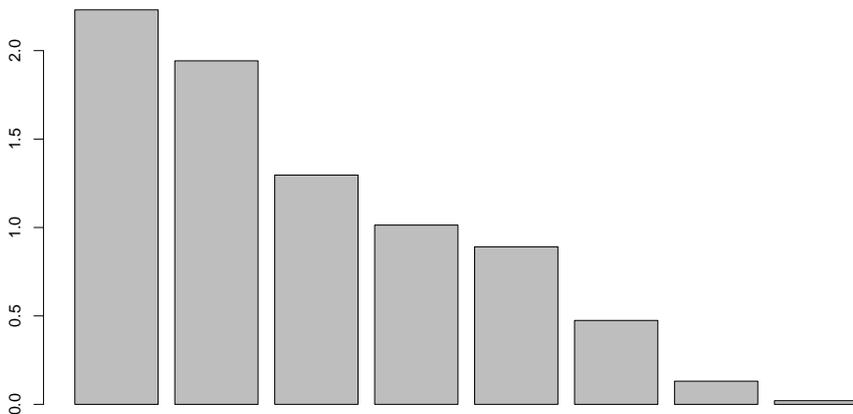
```
> round(pca2$co, digits=2)
      Comp1 Comp2 Comp3 Comp4 Comp5
Commodite -0.29  0.57 -0.46  0.23  0.52
Acoustiq  -0.05 -0.56 -0.44  0.58 -0.19
TpsCharge  0.13  0.82 -0.45  0.17 -0.14
AutoVeille -0.11 -0.04 -0.66 -0.53 -0.49
Autoconv   0.18 -0.47 -0.38 -0.47  0.56
PlusGSM    -0.52  0.57  0.30 -0.31 -0.07
SensiGSM   -0.96 -0.19 -0.14  0.04 -0.01
SensiDCS   -0.95 -0.24  0.07  0.02  0.04
```



```
> inertia.dudi(pca2, ro=T)$row.abs
      Axis1 Axis2 Axis3 Axis4 Axis5
Philips      0 1096 1115  905 1342
Nokia3310   345 1340  389  178  519
Nokia6210   263  208  188 1118  793
Sony        536   65  161 2330  346
SiemensSL45 583 1168  381  327   65
EricssonT20 479 1637  549   66    0
SiemensS35i 478  163  118  302  311
Alcatel304    3   20    4    1 2090
Samsung      3  800 1687  513  543
Trium       549   96  243  461 1234
Motorola    303  526    2   90 1429
Panasonic   3127   80  356 1753   49
Alcatel501  102 1720  436 1634   49
Ericsson2628 12  987 4341   29  965
Sagem      3215   95   29  295  266
```

Question 2. Faire une représentation en histogramme des valeurs propres. Combien de composantes principales faut-il retenir ? Quel est le pourcentage d'inertie totale expliquée par le sous-espace principal correspondant ? Si on ne retient que le premier plan principal, quel est le ratio expliqué ?

On obtient l'histogramme suivant pour les valeurs propres :



Le critère de Kaiser ($\lambda > 1$) amène à considérer les 3 premières valeurs propres (on pourrait en conserver 4, mais la quatrième valeur est vraiment proche de 1). L'inertie totale est 8 et l'inertie portée par les 3 premiers axes est $2.23 + 1.94 + 1.30 = 5.47$. Cela représente 68% de l'inertie totale.

Si on ne considère que les deux premières valeurs propres, on n'a que 52% de l'inertie totale. On voit que les valeurs propres ne décroissent pas très vite : la qualité de l'ACP ne sera pas très bonne.

Question 3. Quelles sont les variables qui déterminent la première composante principale ? La proximité des ces variables sur le cercle est-elle cohérente avec ce que l'on sait des données d'origine ?

La première composante principale est déterminée négativement par **SensiGSM** et **SensiDCS**. On le voit sur le cercle, mais aussi par les corrélations, respectivement $-0,96$ et $-0,95$. Il n'y a pas de variable fortement corrélée positivement avec cet axe.

Comme ces variables sont très près du cercle, on peut déduire de leur proximité dans le cercle de corrélation qu'elles sont très corrélées l'une à l'autre. En effet, ce sont les deux variables notées à la question 1 comme ayant une corrélation de $0,93$.

Question 4. Quels sont les individus qui déterminent le plus la première composante principale ? Quelle est leur contribution ?

Les individus qui déterminent la première composante principale sont **Panasonic** et **Sagem**. On le voit sur la projection des individus mais surtout par leur contribution aux axes (qui sont données ici en 10000ièmes) :

- **Panasonic** : 0,3127
- **Sagem** : 0,3215

Ces valeurs sont à comparer avec le poids de chaque individu, qui est ici $1/15 \approx 0,07$. Elles sont très au dessus des valeurs normales, puisque les contributions des individus somment à 1 et que ces deux éléments à eux seuls en représentent près des deux tiers.

Question 5. Quel est le lien entre les variables de la question 3 et les individus de la question 4 ?

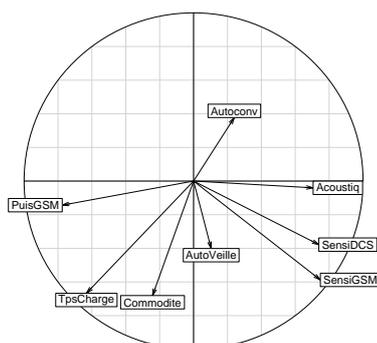
D'une part le premier axe est déterminé par **SensiGSM** et **SensiDCS** (question 3). D'autre part il est déterminé par les deux individus **Panasonic** et **Sagem** (question 4).

Il est en fait facile de vérifier que ces deux portables ont une note hors norme pour ces deux variables, ce qui explique les caractéristiques de la première composante principale.

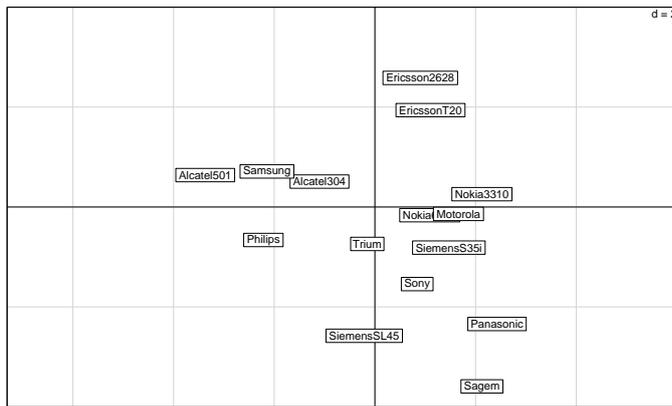
3 Une nouvelle ACP

On décide de faire une nouvelle analyse en composantes principales sur les variables centrées-réduites en passant les téléphones mobiles « perturbateurs » en éléments supplémentaires. On donne ci-dessous les mêmes données que dans la section précédente (valeurs propres, corrélations entre variables et facteurs, projection des individus sur le premier plan principal et contributions des individus aux facteurs) :

```
> round(pca3$eig, digits=2)
[1] 2.73 1.72 1.32 0.93 0.68 0.43 0.17 0.02
```



```
> round(pca3$co, digits=2)
      Comp1 Comp2 Comp3 Comp4 Comp5
Commodite -0.24 -0.68  0.12 -0.63  0.19
Acoustiq  0.71 -0.04  0.21  0.32  0.55
TpsCharge -0.63 -0.67  0.22  0.13  0.18
AutoVeille 0.10 -0.40  0.70  0.41 -0.36
Autoconv  0.24  0.38  0.70 -0.45 -0.16
PuisGSM   -0.78 -0.14 -0.22  0.20 -0.19
SensiGSM  0.75 -0.59 -0.11  0.02 -0.06
SensiDCS  0.74 -0.38 -0.42 -0.12 -0.34
```



```
> inertia.dudi(pca3,ro=T)$row.abs
```

| | Axis1 | Axis2 | Axis3 | Axis4 | Axis5 |
|--------------|-------|-------|-------|-------|-------|
| Philips | 1382 | 195 | 1920 | 922 | 3 |
| Nokia3310 | 1273 | 29 | 431 | 318 | 166 |
| Nokia6210 | 334 | 12 | 324 | 865 | 1712 |
| Sony | 199 | 1064 | 320 | 563 | 2008 |
| SiemensSL45 | 13 | 2968 | 118 | 93 | 1 |
| EricssonT20 | 344 | 1679 | 2626 | 40 | 837 |
| SiemensS35i | 608 | 295 | 29 | 181 | 60 |
| Alcatel304 | 341 | 116 | 96 | 3310 | 557 |
| Samsung | 1295 | 226 | 1160 | 217 | 1473 |
| Trium | 6 | 244 | 358 | 1306 | 1922 |
| Motorola | 776 | 9 | 73 | 1425 | 6 |
| Alcatel501 | 3200 | 182 | 6 | 4 | 1104 |
| Ericsson2628 | 229 | 2981 | 2539 | 758 | 150 |

On ajoute à ces données la table des cosinus carrés (en 10000ièmes) de l'angle entre les individus et les 5 premiers axes principaux :

```
> abs(inertia.dudi(pca3,ro=T)$row.rel)
```

| | Axis1 | Axis2 | Axis3 | Axis4 | Axis5 | con.tra |
|--------------|-------|-------|-------|-------|-------|---------|
| Philips | 4699 | 418 | 3162 | 1069 | 2 | 1004 |
| Nokia3310 | 6658 | 97 | 1091 | 566 | 218 | 653 |
| Nokia6210 | 2605 | 59 | 1224 | 2295 | 3342 | 438 |
| Sony | 1144 | 3855 | 891 | 1102 | 2893 | 594 |
| SiemensSL45 | 54 | 7997 | 243 | 135 | 1 | 799 |
| EricssonT20 | 1061 | 3267 | 3923 | 42 | 647 | 1106 |
| SiemensS35i | 6209 | 1901 | 144 | 630 | 153 | 334 |
| Alcatel304 | 1850 | 395 | 252 | 6119 | 757 | 629 |
| Samsung | 4940 | 544 | 2143 | 282 | 1409 | 895 |
| Trium | 44 | 1058 | 1191 | 3056 | 3307 | 497 |
| Motorola | 5929 | 41 | 270 | 3709 | 11 | 447 |
| Alcatel501 | 8060 | 289 | 7 | 4 | 697 | 1355 |
| Ericsson2628 | 627 | 5147 | 3366 | 707 | 103 | 1247 |

Question 6. Expliquer en quoi la nouvelle analyse est différente de la précédente et pourquoi elle est utile. Est-ce qu'il sera toujours possible d'interpréter les données des éléments supprimés ?

Cette nouvelle ACP est différente car les téléphones **Panasonic** et **Sagem** ont été retirés de l'analyse. On espère donc que les axes principaux ne seront plus « tirés » vers ces deux éléments, qui avaient une trop forte influence.

Il est toujours possible d'interpréter ces deux éléments (et ils sont d'ailleurs représentés sur la projection des individus sur le premier axe principal). Pour cela, il suffit de calculer leurs coordonnées sur chaque axe principal à partir des coordonnées de la composante principale.

Pour résumer, ces éléments n'ont pas servi à l'analyse, mais on peut voir comment ils se comportent vis-à-vis des résultats de cette analyse.

Question 7. Combien de valeurs propres faut-il retenir, et quelles proportion d'inertie totale cela représente-t-il ?

Le critère de Kaiser conduit à ne retenir que 3 valeurs propres. Cela correspond à une inertie de $2.73 + 1.72 + 1.32 = 5,77$, soit 72% de l'inertie totale.

On a donc une légère augmentation de la qualité globale de la représentation.

Question 8. Pour chacune des trois premières composantes principales, donner les variables qui les caractérisent le plus.

On trouve ces variables en regardant les coefficients de corrélation. On fixe une limite de 0,60.

Le premier axe est caractérisé par **SensiGSM** (corrélation 0,75), **SensiDCS** (0,74) et **Acoustiq** (0,71) d'une part ; **PuisGSM** (-0,78) et **TpsCharge** (-0,63) d'autre part.

Le deuxième axe est plus lié à **Commodite** (-0,68), **TpsCharge** (-0,67) et **SensiGSM** (-0,59).

Le troisième axe correspond à **AutoVeille** (0,70) et **AutoConv** (0,70).

Question 9. Si on se réfère à la projection des individus sur le premier plan principal, où pense-t-on trouver ceux qui ont des notes supérieures à la moyenne pour **PuisGSM** ? et pour **TpsCharge** ?

Ce sont plutôt ceux qui sont sur la gauche du diagramme qui ont une note élevée à **PuisGSM**, puisque le premier axe est lié négativement à cette caractéristiques. Toutefois la corrélation de **PuisGSM** avec le premier axe n'est que de 0,78, et il faut s'attendre à des surprises. En effet, si on regarde les données d'origine, on voit que les bonnes notes sont à gauche, sauf **Trium**.

Pour **TpsCharge**, un raisonnement analogue prédit que les individus seront en bas à gauche de la projection mais, là encore, les corrélations ne sont pas très importantes.

Question 10. *Comment mesure-t-on la qualité de la représentation d'un individu par un axe principal ? Quels sont les deux individus les plus mal représentés sur le premier axe principal ?*

La qualité de la représentation d'un individu par une axe se mesure par le cosinus carré de l'angle entre l'individu et l'axe. En se reportant à la table ci-dessus, on voit que les deux individus vraiment mal représentés sont **Trium** (0,0044) et **SiemensSL45** (0,0054).

Question 11. *Comment mesure-t-on la qualité de la représentation d'un individu par un sous-espace principal ? Quels sont les deux individus les plus mal représentés par les 3 premiers axes principaux ?*

Pour trouver ces individus, il faut calculer le cosinus carré des individus avec le sous espace formé par les trois premiers axes principaux. Ce cosinus carré est en fait la somme des valeurs données dans les trois premières colonnes, puisque les cosinus carrés sont additifs dans ce cas. En additionnant les trois colonnes, on trouve :

```
> inertia.dudi(pca3,ro=T)$row.cum[,3]
Philips Nokia3310 Nokia6210 Sony SiemensSL45 EricssonT20 SiemensS35i Alcatel1304 Samsung Trium Motorola Alcatel1501 Eric-
son2628
8279      7846      3888 5889      8295      8251      8255      2498      7627 2293      6240      8356      9140
```

Les deux téléphones les moins bien représentés sont donc **Trium** (0,2293) et **Alcatel1304** (0,2498).

Question 12. *Que peut-on dire de la position des deux éléments supplémentaires sur la projection des individus ? Peut-on en déduire qu'ils sont bien représentés par le sous espace F_3 correspondant aux 3 premiers axes ? On s'aidera des données suivantes, qui récapitulent leurs coordonnées sur les axes principaux :*

```
> round(pca3sup$lisup, digits=2)
      Axis1 Axis2 Axis3 Axis4 Axis5 Axis6 Axis7 Axis8
Panasonic  2.43 -2.34  1.99  0.12  3.52  2.43  0.03  0.92
Sagem      2.13 -3.59  2.04 -0.52  1.15  3.08 -0.36 -0.20
```

Les coordonnées des deux éléments supplémentaires sur la projection des individus sur le premier plan principal sont comparables aux autres.

Pour savoir si les éléments sont bien représentés par les 3 premiers axes principaux, il faut calculer leur cosinus carré par rapport au sous espace. Pour cela, on élève les données ci-dessus au carré et on divise chaque ligne par sa somme, ce qui donne les cosinus carrés des angles par rapport aux 3 premiers axes (en 10000ièmes) :

```
> round(pca3cos2sup[,1 :3]*10000)
      Axis1 Axis2 Axis3
Panasonic 1708 1592 1151
Sagem     1379 3934 1265
```

Les données que nous recherchons sont obtenues en sommant les deux lignes ci-dessus, donnant respectivement 4451 et 6578.

Les deux téléphones sont donc plutôt mal représentés (respectivement 0,4451 et 0,6578 sur les trois premiers axes), mais ils sont quand même dans le milieu du peloton par rapport aux autres cosinus carrés.