



**Sujet d'examen de 1^{ère} session
Semestres 1-3-5
Année universitaire 2014-2015**

Intitulé de l'épreuve : Analyse de données

Nom de l'enseignant : Nils Berglund

Mention / Spécialité / Parcours : M1 ESA

Année : 2014-2015

Durée de l'épreuve : 2 heures

Documents autorisés : Résumé manuscrit de 4 pages A4

Matériels autorisés : Calculatrice non programmable

16/12/2014

Amphi Jousse

P1/6

SUJET

Les téléphones portables doivent être éteints durant l'examen.
Il sera tenu compte de la qualité de la rédaction.
Les points sont donnés à titre indicatif.

Questions de cours [6 points]

1. Quelle est la signification de la p -value?
2. Dans une AFC, pourquoi la matrice $S = X^T Y^T$ admet-elle toujours la valeur propre 1? Pourquoi ignore-t-on cette valeur propre dans l'analyse?
3. Expliquer, sans donner de preuve détaillée, le lien entre les valeurs propres des matrices $S = X^T Y^T$ et $T = Y^T X^T$ dans une AFC.
4. Expliquer la différence entre contributions absolues et relatives dans une AFC.
5. Pourquoi une ACM est-elle souvent plus difficile à interpréter qu'une AFC?
6. Donner un exemple de piège à éviter dans l'analyse d'une ACM.

Exercice 1 [7 points]

Le tableau suivant montre le nombre de brevets en télécommunication déposés dans 9 pays, de 1980 à 1986.

	1986	1985	1984	1983	1982	1981	1980
USA	986	774	711	591	467	404	258
Japon	653	552	361	307	195	129	43
Allemagne	405	357	347	254	294	313	208
France	189	158	200	171	153	184	147
GB	204	182	158	137	92	86	67
Italie	31	28	28	21	22	29	15
Pays-Bas	64	59	61	64	33	30	11
Suède	25	19	31	25	15	12	13
Suisse	23	34	19	30	17	30	15

1. Quels sont les pays qui ont connu la plus forte croissance de dépôts de brevets? La plus faible croissance? Une décroissance?
2. Interpréter le résultat du test du chi-deux ci-dessous.
3. Discuter les inerties.
4. Y a-t-il des points mal représentés sur le plan des deux premiers axes factoriels? Et dans l'espace des trois premiers axes factoriels?
5. Interpréter, si possible, les trois premiers axes factoriels.

Inerties:

```
> summary(as.table(tab))
Number of cases in table: 11541
Number of factors: 2
Test for independence of all factors:
Chisq = 516.4, df = 48, p-value = 9.223e-80
> afc <- dudi.coa(tab, scan = FALSE, nf=4)
> afcin <- inertia.dudi(afc,col.inertia=T,row.inertia=T)
> afcin$TOT
      inertia      cum      ratio
1 3.981921e-02 0.03981921 0.8898905
2 2.643138e-03 0.04246234 0.9489601
3 1.500220e-03 0.04396256 0.9824874
4 6.113650e-04 0.04457393 0.9961504
5 1.279688e-04 0.04470190 0.9990103
6 4.428695e-05 0.04474619 1.0000000
```

Contributions absolues et relatives des lignes:

```
> afcin$row.abs
      Axis1 Axis2 Axis3 Axis4
USA      121   186   935   545
Japon   4867   613   190   299
Allemagne 1865  2773    6   673
France  2697   407   20  1320
GB        10   436    0  2961
Italie   177   216   289   613
Pays-Bas  47  3268  1077  3182
Suede    34  2092   648   48
Suisse   181    9  6834   359

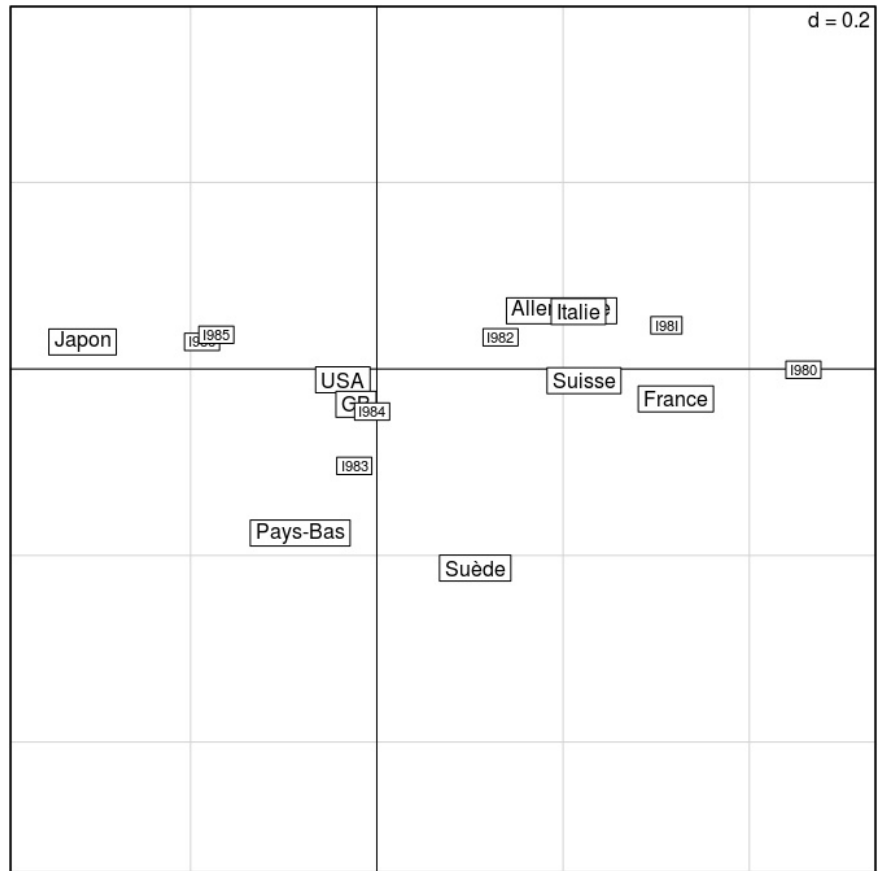
> afcin$row.rel
      Axis1 Axis2 Axis3 Axis4
USA     -6510  -663  1889   449
Japon   -9892   83   -15   -9
Allemagne 9024   891    1   50
France   9812  -98    3  -74
GB     -1112 -3215   -1 -5053
Italie   8047   651  -494   427
Pays-Bas -1335 -6103 -1142  1374
Suede   1668 -6818  1199  -36
Suisse  4057   -13 -5774  -124
```

Contributions absolues et relatives des colonnes:

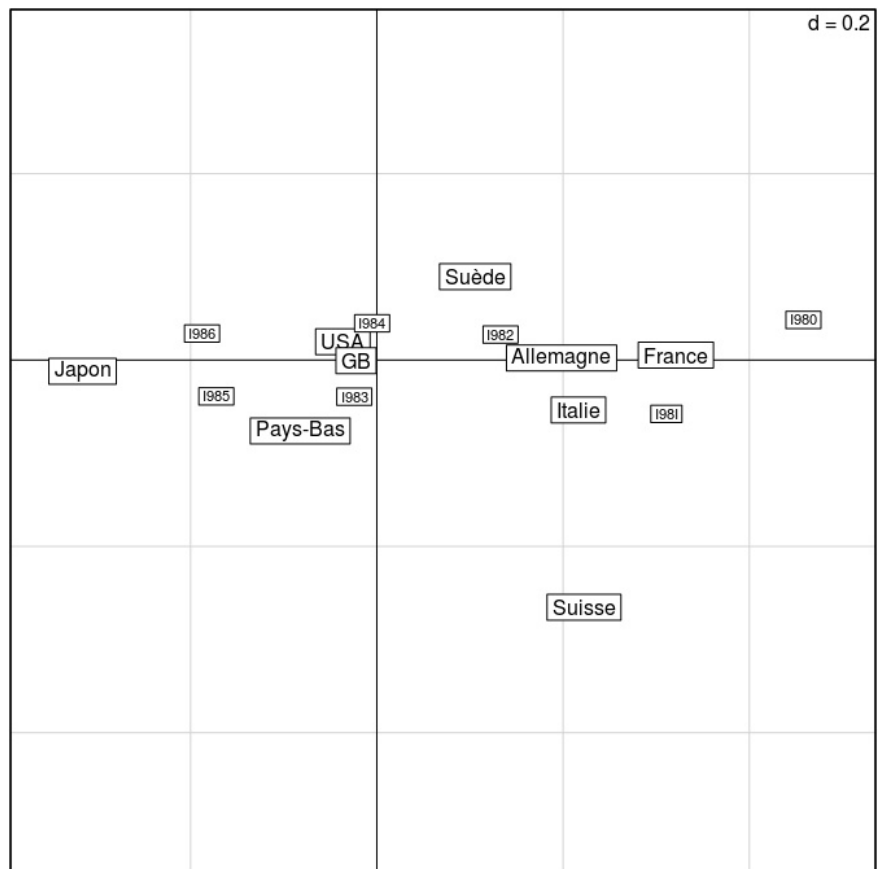
```
> afcin$col.abs
      Comp1 Comp2 Comp3 Comp4
1986  1978   710  1209   57
1985  1398   950  1903  1081
1984    1  1312  1720   482
1983   20  5667  1428    0
1982  495   487   558  2828
1981  2560   873  2352   763
1980  3548    0   830  4789

> afcin$col.rel
      Comp1 Comp2 Comp3 Comp4
1986  -9495   226   219   -4
1985  -8975   405  -460  -107
1984    -62 -5085  3782   432
1983   -450 -8287 -1185    0
1982   8292   542   352   727
1981   9406   213  -326   43
1980   9713    0    86  -201
```

Composantes 1 et 2:



Composantes 1 et 3:



Exercice 2 [7 points]

Le tableau suivant résume les résultats d'une enquête d'une maison vendant des produits audiovisuels. L'enquête comprenait des questions sur l'âge, la profession, le style de musique préférée, et la somme investie par mois dans l'achat de leurs produits.

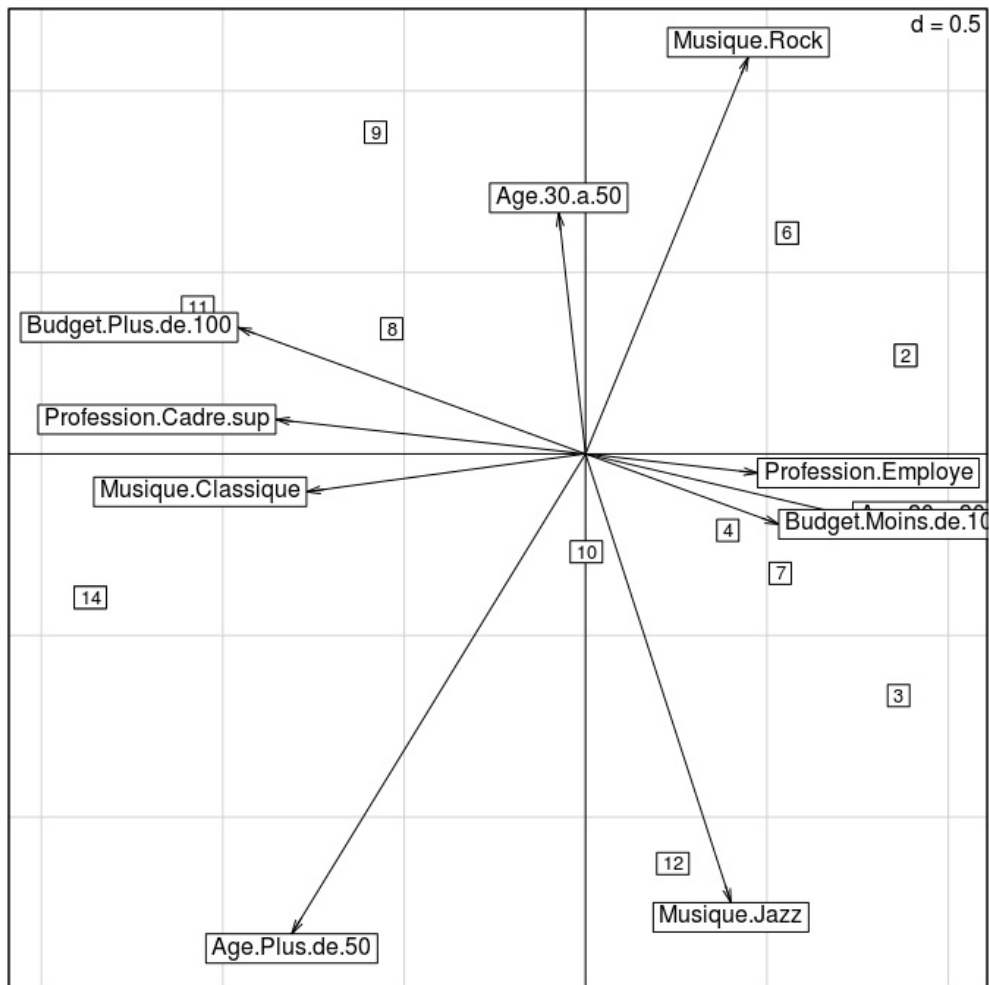
	Age	Profession	Musique	Budget
1	20-30 ans	Employé	Rock	Moins de 100 euros
2	20-30 ans	Employé	Rock	Moins de 100 euros
3	20-30 ans	Employé	Jazz	Moins de 100 euros
4	20-30 ans	Employé	Classique	Moins de 100 euros
5	30-50 ans	Employé	Rock	Moins de 100 euros
6	30-50 ans	Employé	Rock	Moins de 100 euros
7	30-50 ans	Employé	Jazz	Moins de 100 euros
8	30-50 ans	Employé	Classique	Plus de 100 euros
9	30-50 ans	Cadre sup	Rock	Plus de 100 euros
10	30-50 ans	Cadre sup	Jazz	Moins de 100 euros
11	30-50 ans	Cadre sup	Classique	Plus de 100 euros
12	Plus de 50 ans	Employé	Jazz	Moins de 100 euros
13	Plus de 50 ans	Cadre sup	Classique	Plus de 100 euros
14	Plus de 50 ans	Cadre sup	Classique	Plus de 100 euros

- Déterminer les tableaux de contingence des variables croisées deux à deux.
- Représenter les données sous forme de tableau disjonctif Z et calculer ses marges.
- Calculer le tableau de Burt B .

La page suivante montre le résultat d'une ACM effectuée sur ce jeu de données.

- Discuter les inerties.
- Interpréter les deux premiers axes factoriels.
- Peut-on dégager des corrélations entre variables caractérisant les consommateurs et leur comportement en termes d'achat?

```
> acmin$TOT
      inertia      cum      ratio
1 0.63652546 0.6365255 0.4243503
2 0.30068794 0.9372134 0.6248089
3 0.26553709 1.2027505 0.8018337
4 0.14857563 1.3513261 0.9008841
5 0.11210485 1.4634310 0.9756206
6 0.03656903 1.5000000 1.0000000
```



> acmin\$col.abs

	Comp1	Comp2	Comp3	Comp4
Age.20.a.30	1005	55	2994	292
Age.30.a.50	18	1432	2579	595
Age.Plus.de.50	910	2426	207	3249
Profession.Cadre.sup	1688	21	249	302
Profession.Empley	938	12	138	168
Musique.Classique	1369	25	1211	2148
Musique.Jazz	297	2827	2431	101
Musique.Rock	466	2765	87	3060
Budget.Moins.de.100	1182	156	37	31
Budget.Plus.de.100	2127	281	67	55

> acmin\$col.rel

	Comp1	Comp2	Comp3	Comp4
Age.20.a.30	3583	-93	-4452	-243
Age.30.a.50	-90	3445	5479	-707
Age.Plus.de.50	-2950	-3713	-280	2458
Profession.Cadre.sup	-6685	39	412	279
Profession.Empley	6685	-39	-412	-279
Musique.Classique	-5423	-47	-2001	-1985
Musique.Jazz	1060	-4761	3615	-84
Musique.Rock	1844	5173	-143	2829
Budget.Moins.de.100	8425	-526	110	51
Budget.Plus.de.100	-8425	526	-110	-51