



FACULTÉ DE DROIT
ÉCONOMIE & GESTION

Sujet d'examen de 1^{ère} session Semestres 1-3-5 Année universitaire 2015-2016

Intitulé de l'épreuve :	Analyse de données qualitatives
Nom de l'enseignant :	Nils Berglund
Mention / Spécialité / Parcours :	M1 ESA
Année :	2015-2016
Durée de l'épreuve :	2 heures
Documents autorisés :	Résumé manuscrit de 4 pages A4
Matériels autorisés :	Calculatrice non programmable
	15/12/2015

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.

Exercice 1 [7 points]

Le tableau suivant résume les résultats d'une enquête portant sur le type de vacances (à la mer, à la montagne, visite d'une ville) choisies par un échantillon de 18 personnes. Les questions portaient sur la classe d'âge, la catégorie socioprofessionnelle, et le montant dépensé par personne pour les vacances.

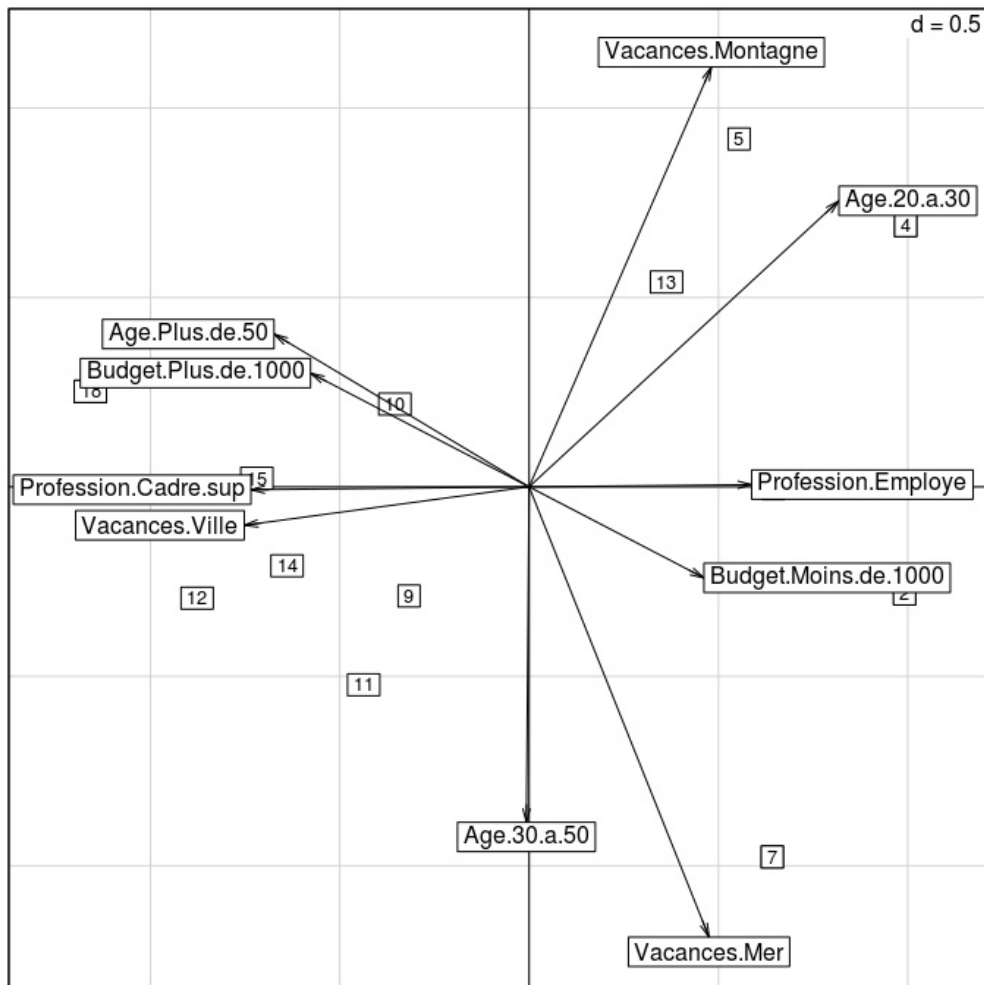
	Age	Profession	Vacances	Budget
1	20-30 ans	Employé	Mer	Moins de 1000 euros
2	20-30 ans	Employé	Mer	Moins de 1000 euros
3	20-30 ans	Employé	Montagne	Moins de 1000 euros
4	20-30 ans	Employé	Montagne	Moins de 1000 euros
5	20-30 ans	Employé	Montagne	Plus de 1000 euros
6	30-50 ans	Employé	Mer	Moins de 1000 euros
7	30-50 ans	Employé	Mer	Moins de 1000 euros
8	30-50 ans	Employé	Montagne	Moins de 1000 euros
9	30-50 ans	Employé	Ville	Plus de 1000 euros
10	30-50 ans	Cadre sup	Montagne	Plus de 1000 euros
11	30-50 ans	Cadre sup	Ville	Moins de 1000 euros
12	30-50 ans	Cadre sup	Ville	Plus de 1000 euros
13	Plus de 50 ans	Employé	Montagne	Moins de 1000 euros
14	Plus de 50 ans	Cadre sup	Mer	Plus de 1000 euros
15	Plus de 50 ans	Cadre sup	Ville	Moins de 1000 euros
16	Plus de 50 ans	Cadre sup	Ville	Plus de 1000 euros
17	Plus de 50 ans	Cadre sup	Ville	Plus de 1000 euros
18	Plus de 50 ans	Cadre sup	Ville	Plus de 1000 euros

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

Ci-dessous et sur la page suivante on montre le résultat d'une ACM effectuée sur ce jeu de données.

4. Discuter les inerties.
5. Interpréter les deux premiers axes factoriels.
6. Quelle modalité est la moins bien représentée sur le plan des deux premiers axes factoriels?
7. Quelles corrélations peut-on dégager entre variables caractérisant les personnes et leur choix de vacances?

```
> acmin$TOT
      inertia      cum      ratio
1 0.6534920 0.6534920 0.4356613
2 0.2710705 0.9245625 0.6163750
3 0.2414289 1.1659914 0.7773276
4 0.1531277 1.3191191 0.8794127
5 0.1128553 1.4319744 0.9546496
6 0.0680256 1.5000000 1.0000000
```



Contributions absolues et relatives:

```
> acmin$col.abs
```

	Comp1	Comp2	Comp3	Comp4
Age.20.a.30	1333	1135	593	1017
Age.30.a.50	0	2184	3833	12
Age.Plus.de.50	1084	389	1993	1076
Profession.Cadre.sup	1719	0	21	12
Profession.Employe	1375	0	17	10
Vacances.Mer	449	2811	2114	937
Vacances.Montagne	554	2935	1253	113
Vacances.Ville	1577	29	37	257
Budget.Moins.de.1000	849	230	62	2918
Budget.Plus.de.1000	1061	287	77	3648

```
> acmin$col.rel
```

	Comp1	Comp2	Comp3	Comp4
Age.20.a.30	4824	1704	-794	-863
Age.30.a.50	-1	-3874	6058	-12
Age.Plus.de.50	-4249	632	-2886	988
Profession.Cadre.sup	-8087	0	-36	14
Profession.Employe	8087	0	36	-14
Vacances.Mer	1627	-4221	-2827	-795
Vacances.Montagne	2172	4774	1815	104
Vacances.Ville	-6744	-51	59	258
Budget.Moins.de.1000	4992	-560	-134	4022
Budget.Plus.de.1000	-4992	560	134	-4022

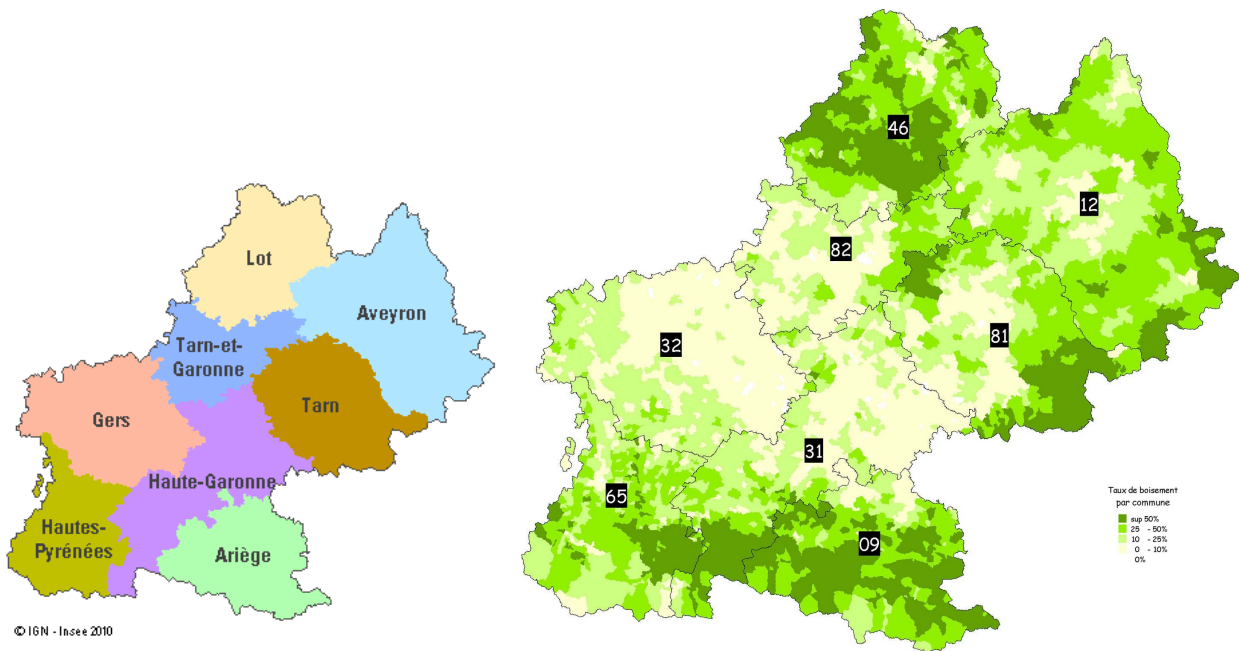
Exercice 2 [7 points]

A la suite d'une étude de 1993, 73000 exploitations agricoles de la région Midi-Pyrénées ont été ventilées dans le tableau de contingence suivant selon le département (en lignes, 8 modalités) et la S.A.U. (Surface Agricole Utilisée, en colonnes, 6 classes).

Les S.A.U. comprennent les classes suivantes : INF05 = moins de 5 hectares ; S0510 = entre 5 et 10 hectares... ; SUP50 = plus de 50 hectares.

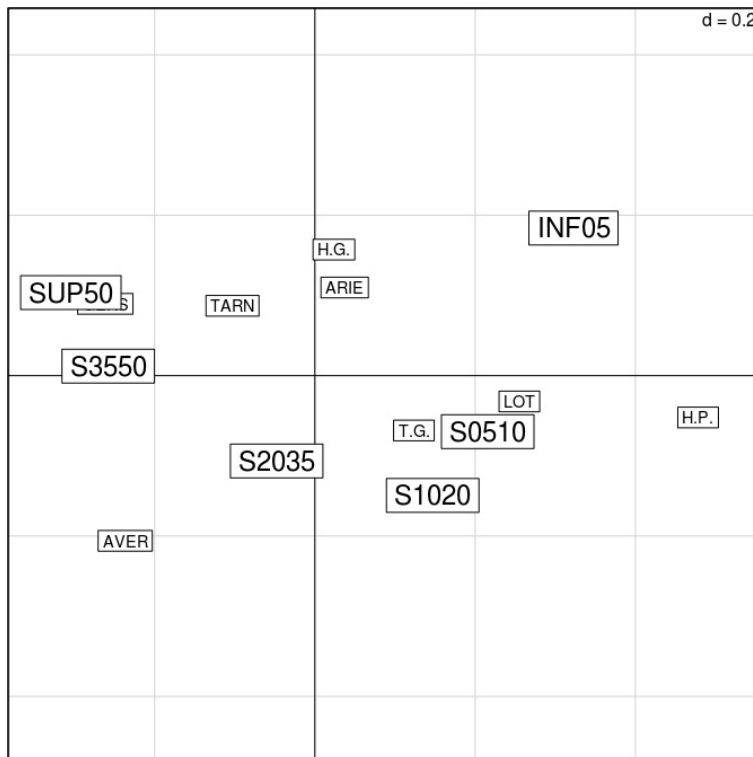
Département/Surface	INF05	S0510	S1020	S2035	S3550	SUP50	Total
09. Ariège (ARIE)	870	330	730	680	470	890	3970
12. Aveyron (AVER)	820	1260	2460	3330	2170	2960	13000
31. Haute-Garonne (H.G.)	2290	1070	1420	1830	1260	2330	10200
32. Gers (GERS)	1650	890	1350	2540	2090	3230	11750
46. Lot (LOT)	1940	1130	1750	1660	770	1140	8390
65. Haute-Pyrénées (H.P.)	2110	1170	1640	1500	550	430	7400
81. Tarn (TARN)	1770	820	1260	2010	1680	2090	9630
82. Tarn-et-Garonne(T.G.)	1740	920	1560	2210	990	1240	8660
Total	13190	7590	12170	15760	9980	14310	73000

Les cartes suivantes montrent la situation géographique des départements de la région ainsi que le taux de boisement.



Voici les résultats d'une AFC faite sur ces données.

```
> afcin$TOT
      inertia      cum      ratio
1 0.0550125256 0.05501253 0.7470790
2 0.0149081227 0.06992065 0.9495337
3 0.0023949245 0.07231557 0.9820572
4 0.0007794898 0.07309506 0.9926428
5 0.0005417623 0.07363682 1.0000000
```



Contributions absolues et relatives des lignes:

> afcin\$row.abs			> afcin\$row.rel			
	Axis1	Axis2		Axis1	Axis2	con.tra
ARIE	14	440	ARIE	463	4042	220
AVER	1813	5072	AVER	-5637	-4273	2403
H.G.	14	2314	H.G.	202	8829	531
GERS	2001	864	GERS	-8898	1042	1680
LOT	1360	80	LOT	9512	-152	1069
H.P.	4214	185	H.P.	9817	-117	3207
TARN	253	671	TARN	-4388	3147	432
T.G.	329	373	T.G.	5364	-1646	459

Contributions absolues et relatives des colonnes:

> afcin\$col.abs			> afcin\$col.rel			
	Comp1	Comp2		Comp1	Comp2	con.tra
INF05	3420	4102	INF05	7517	2444	3399
S0510	879	341	S0510	8195	-860	802
S1020	655	2495	S1020	4475	-4620	1094
S2035	89	1641	S2035	-1281	-6377	521
S3550	1653	13	S3550	-9195	19	1343
SUP50	3304	1408	SUP50	-8683	1003	2842

1. Expliquer les lignes suivantes fournies par le logiciel R:
 Number of cases in table: 73000
 Number of factors: 2
 Test for independence of all factors:
 Chisq = 5375, df = 35, p-value = 0
2. Discuter les valeurs des inerties. Une étude restreinte aux deux premiers axes factoriels vous semble-t-elle pertinente?
3. Quelles sont les modalités contribuant fortement au premier axe factoriel? Interpréter cet axe.
4. Peut-on donner une interprétation du second axe factoriel?
5. Y a-t-il des départements mal représentés sur le plan factoriel?
6. Quels départements sont caractérisés par de petites exploitations? lesquels par de grandes exploitations?
7. Qu'est-ce qui distingue l'Aveyron des autres départements?

Questions de cours [6 points]

1. Comment est définie la loi du chi-deux? Donner un exemple d'utilisation de cette loi.
2. On considère les deux matrices

$$X = \frac{1}{10} \begin{pmatrix} 5 & 5 \\ 4 & 6 \\ 3 & 7 \end{pmatrix} \quad Y = \frac{1}{10} \begin{pmatrix} 6 & 2 & 2 \\ 5 & 4 & 1 \end{pmatrix}$$

Que peut-on dire sur les valeurs propres des matrices $S = X^T Y^T$ et $T = Y^T X^T$, sans calculer ni S , ni T ?

Calculer S et déterminer ses valeurs propres. En déduire sans calculs les valeurs propres de la matrice T .

3. Illustrer la propriété d'équivalence distributionnelle à l'aide du tableau de contingence suivant. Ici a, b, et c représentent trois modalités d'une première variable qualitative, alors que A, B, C et D représentent quatre modalités d'une seconde variable qualitative.

	a	b	c	
A	10	40	50	100
B	20	50	50	120
C	10	25	45	80
D	10	25	25	60
	50	140	170	360

4. Pourquoi est-il préférable d'éviter les modalités à très faible effectif dans une ACM?