



Sujet d'examen - 1^{ère} session Semestres 1-3-5 Année universitaire 2016-2017

Intitulé de l'épreuve :	Analyse de données qualitatives
Nom de l'enseignant :	Nils Berglund
Mention / Spécialité / Parcours :	M1 ESA
Année :	2016-2017
Durée de l'épreuve :	2 heures
Documents autorisés :	Résumé manuscrit de 4 pages A4
Matériels autorisés :	Calculatrice non programmable
	20/12/2016

P1/7

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 [8 points]

1. Dans chacun des cas suivants, on a indiqué les valeurs propres différentes de 0 et 1 de la matrice S décrivant une AFC. Que peut-on dire sur le nombre n et p de modalités des deux variables ? Esquisser dans chaque cas la forme du nuage de points.

- (a) $\lambda_1 = 0.9, \lambda_2 = 0.5$.
 (b) $\lambda_1 = 0.5, \lambda_2 = 0.5$.
 (c) $\lambda_1 = 0.9, \lambda_2 = 0.1$.

2. On considère les deux matrices

$$X = \frac{1}{10} \begin{pmatrix} 8 & 2 \\ 2 & 8 \\ 10 & 0 \end{pmatrix} \quad Y = \frac{1}{10} \begin{pmatrix} 4 & 3 & 3 \\ 0 & 5 & 5 \end{pmatrix}$$

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, c et d représentent trois modalités d'une première variable qualitative, alors que A, B et C représentent quatre modalités d'une seconde variable qualitative.

	a	b	c	d	
A	10	10	20	60	100
B	20	20	40	20	100
C	10	20	20	50	100
	40	50	80	130	300

4. On effectue un test du chi-deux sur un tableau de contingence de 5 lignes et 5 colonnes, dont l'effectif total est de 100.
- (a) Comment calcule-t-on la valeur χ^2 du chi-deux associée à ce tableau ? Que doit-on prendre pour le nombre de degrés de liberté ?
- (b) Quelle est la valeur minimale que peut prendre χ^2 ? Pour quel type de tableau de contingence obtient-on cette valeur ? Qu'est-ce que cela signifie si l'on obtient cette valeur ?
- (c) Quelle est la valeur maximale que peut prendre χ^2 ? Pour quel type de tableau de contingence obtient-on cette valeur ? Qu'est-ce que cela signifie si l'on obtient cette valeur ?

Exercice 1 [6 points]

Le tableau suivant résume les résultats d'une enquête portant sur 20 étudiants de l'Université d'Orléans, nous informant sur la faculté dans laquelle ils sont inscrits, le moyen de transport utilisé pour se rendre à l'université, et leur genre.

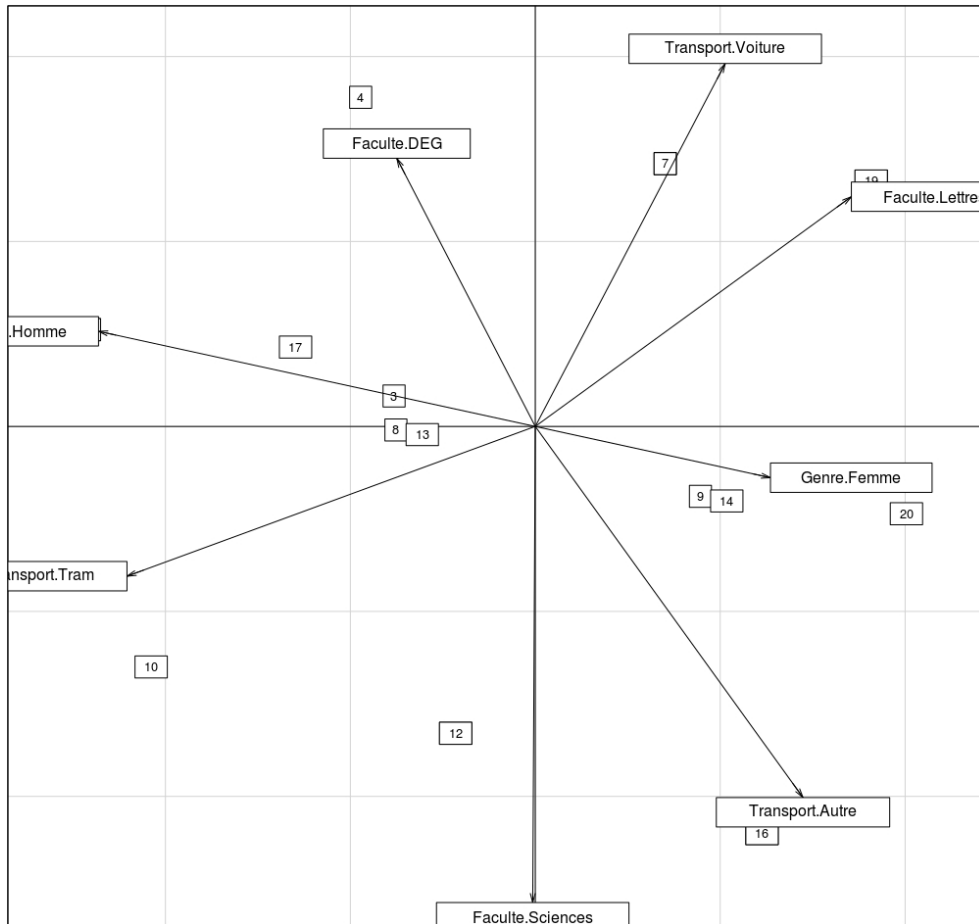
	Faculté	Transport	Genre
1	DEG	Tram	Homme
2	DEG	Tram	Homme
3	DEG	Tram	Femme
4	DEG	Voiture	Homme
5	DEG	Voiture	Femme
6	DEG	Voiture	Femme
7	DEG	Voiture	Femme
8	DEG	Autre	Homme
9	DEG	Autre	Femme
10	Sciences	Tram	Homme
11	Sciences	Tram	Femme
12	Sciences	Tram	Femme
13	Sciences	Voiture	Homme
14	Sciences	Voiture	Femme
15	Sciences	Autre	Femme
16	Sciences	Autre	Femme
17	Lettres	Tram	Homme
18	Lettres	Voiture	Femme
19	Lettres	Voiture	Femme
20	Lettres	Autre	Femme

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. Quelle modalité est la moins bien représentée sur le plan des deux premiers axes factoriels ?
Que devrait-on faire afin de mieux tenir compte de cette modalité ?
6. Interpréter les trois premiers axes factoriels.
7. Quels sont les étudiants venant majoritairement en voiture ?

```
> acmin$TOT
      inertia      cum      ratio
1 0.4672862 0.4672862 0.2803717
2 0.3864798 0.8537661 0.5122597
3 0.3206345 1.1744006 0.7046404
4 0.2955585 1.4699591 0.8819755
5 0.1967076 1.6666667 1.0000000
```



Contributions absolues et relatives:

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

	Comp1	Comp2	Comp3
Faculte.DEG	389	1457	2959
Faculte.Lettres	900	475	4165
Faculte.Sciences	0	3572	166
Transport.Autre	807	1550	1757
Transport.Tram	2628	353	887
Transport.Voiture	650	2373	28
Genre.Femme	1619	77	13
Genre.Homme	3007	143	25

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

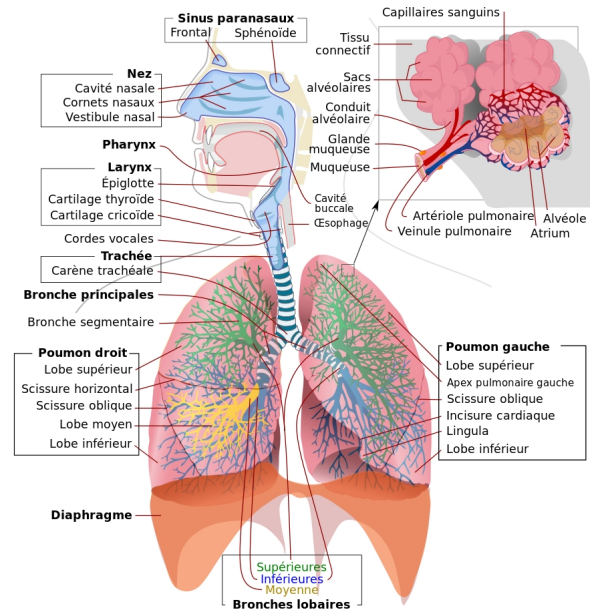
	Comp1	Comp2	Comp3	con.tra
Faculte.DEG	-991	3072	5175	1100
Faculte.Lettres	1577	689	-5008	1600
Faculte.Sciences	0	-6371	-246	1300
Transport.Autre	1508	-2396	2253	1500
Transport.Tram	-5668	-630	-1312	1300
Transport.Voiture	1519	4585	-45	1200
Genre.Femme	6485	-254	37	700
Genre.Homme	-6485	254	-37	1300

Exercice 2 [6 points]

Une étude du marché européen dans le segment des antibiotiques, compilée à partir de 45 millions de prescriptions émises par 40 000 docteurs, a permis d'établir un tableau de contingence comprenant 300 maladies et 60 médicaments.

Ce tableau a ensuite été réduit au tableau ci-dessous, comprenant

- 11 infections bactérielles ou virales des voies respiratoires (en ligne) :
 - Sinusite et Sinusite chronique : inflammations des sinus;
 - Pharyngite, Angine : inflammations du pharynx;
 - Asthme, Bronchite, Bronchite accrue, Bronchite chronique : inflammations des bronches;
 - Inflammation des voies respiratoires;
 - Inflammation de la gorge;
 - Pneumonie : inflammation des poumons.
- 10 marques d'antibiotiques (en colonne): AUGME, CEDAX, CEFIX, CLAVU, KLACID, MACLA, PANAC, VECLA, ZIMMOX, ZINNAT.

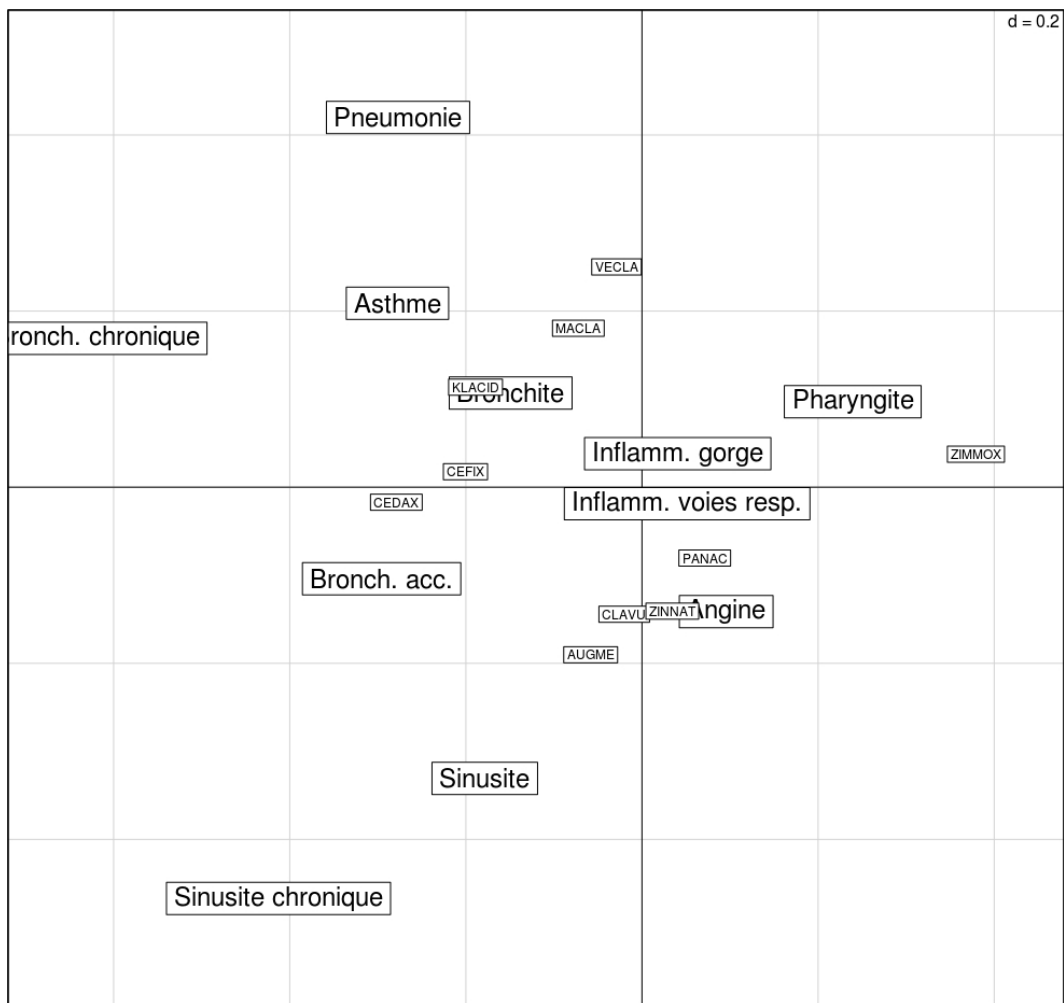
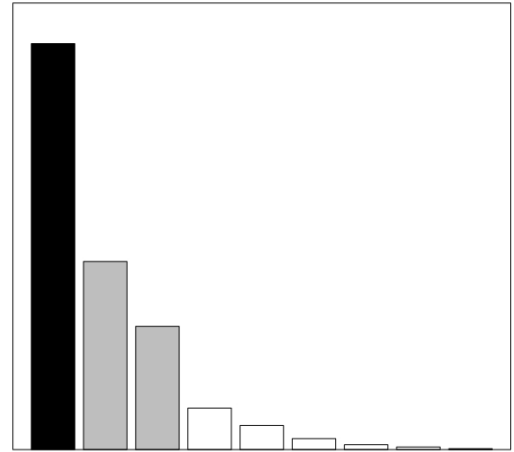


Les nombres sont en milliers de prescriptions.

Inf./Antib.	AUG	CED	CEF	CLA	KLA	MAC	PAN	VEC	ZIM	ZIN	Total
Asthme	57	30	29	19	81	48	27	34	38	9	372
Bronchite	190	100	151	96	152	136	119	139	171	37	1291
Bronch. acc.	232	114	134	92	155	80	68	71	102	49	1097
Pharyngite	311	116	168	138	221	188	169	202	638	83	2234
Infl. resp.	265	58	117	101	140	143	124	119	265	79	1411
Sinusite	39	12	13	16	23	9	14	0	18	7	151
Angine	495	142	136	200	230	167	217	141	489	104	2321
Infl. gorge	95	44	74	44	53	60	36	52	134	34	626
Bronch. chr.	23	30	33	9	23	20	8	21	0	8	175
Pneumonie	34	10	0	10	51	35	5	31	17	4	197
Sinusite chr.	41	8	11	10	11	7	8	5	0	6	107
Total	1782	664	866	735	1140	893	795	815	1872	420	9982

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

```
> afcin$TOT
      inertia      cum      ratio
1 4.134181e-02 0.04134181 0.5060017
2 1.915754e-02 0.06049936 0.7404798
3 1.255074e-02 0.07305010 0.8940941
4 4.226228e-03 0.07727633 0.9458209
5 2.454407e-03 0.07973074 0.9758615
6 1.116574e-03 0.08084731 0.9895278
7 4.960694e-04 0.08134338 0.9955994
8 2.651532e-04 0.08160853 0.9988447
9 9.438896e-05 0.08170292 1.0000000
```



Contributions absolues et relatives des lignes:

```
> afcin$row.abs
                Axis1 Axis2
Asthme           695   848
Bronchite         696   772
Bronch. acc.     2324   621
Pharyngite       3101  1100
Inflamm. voies resp.  90    26
Sinusite         117   864
Angine           514  2425
Inflamm. gorge    25    47
Bronch. chronique 1630   261
Pneumonie        366  1816
Sinusite chronique 442  1221

> afcin$row.rel
                Axis1 Axis2 con.tra
Asthme        -4431  2506    794
Bronchite     -4320  2220    815
Bronch. acc.  -8117 -1004   1449
Pharyngite     8109  1333   1935
Inflamm. voies resp. 1347  -181    337
Sinusite      -1730 -5940    341
Angine         2636 -5759    987
Inflamm. gorge   441   388    287
Bronch. chronique -7200   535   1146
Pneumonie     -1418  3257   1307
Sinusite chronique -3712 -4753    602
```

Contributions absolues et relatives des colonnes:

```
> afcin$col.abs
      Comp1 Comp2
AUGME   147 3377
CEDAX  1252   10
CEFIX   843   14
CLAVU    7  801
KLACID  987  769
MACLA   113 1521
PANAC    98  270
VECLA   17 2667
ZIMMOX 6524  137
ZINNAT  12  435

> afcin$col.rel
      Comp1 Comp2 con.tra
AUGME  -703 -7505  1055
CEDAX -6569  -25   965
CEFIX -3476   27  1227
CLAVU  -159 -8132  231
KLACID -4178 1508  1195
MACLA -1032 6433   554
PANAC  1425 -1819  347
VECLA  -113 8422   743
ZIMMOX 9666   94  3416
ZINNAT  229 -3822  267
```

1. Expliquer les lignes suivantes fournies par le logiciel R:
 Number of cases in table: 9982
 Number of factors: 2
 Test for independence of all factors:
 Chisq = 815.6, df = 90, p-value = 2.439e-117
2. Discuter les valeurs des inerties. Une étude restreinte aux deux premiers axes factoriels vous semble-t-elle pertinente ?
3. Quelles sont les modalités mal représentées sur le plan factoriel ?
4. Quelles sont les infections contribuant fortement au premier axe factoriel ? Interpréter cet axe.
5. Quelles sont les trois marques d'antibiotiques ayant le plus de parts de marché ?
6. Quelles sont les marques d'antibiotiques contribuant le plus au second axe factoriel ?
7. Quelles sont les infections pour lesquelles ces antibiotiques sont prescrits en majorité ? Peut-on en déduire une classification des antibiotiques en termes de leur usage ?