



FACULTÉ DE DROIT
ÉCONOMIE & GESTION

Sujet d'examen – 2^{ème} session
Semestres pairs et impairs
Année universitaire 2024-2025

Intitulé de l'épreuve : Analyse de données qualitatives

Nom de l'enseignant : Nils Berglund

Mention / Spécialité / Parcours : M1 ESA

Année d'études : 2024-2025 **Semestre :** 1

Durée de l'épreuve : 2 heures

Documents autorisés : Résumé manuscrit de 4 pages A4

Matériels autorisés : Calculatrice non programmable

6/01/2025

P1/ 5

SUJET

Les téléphones portables, ordinateurs et tablettes doivent être éteints durant l'examen.

Il sera tenu compte de la qualité de la rédaction.

Les points sont donnés à titre indicatif.

Problème 1 [6 points]

On considère la matrice

$$A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix}.$$

1. Déterminer les valeurs propres de A .
2. Déterminer, pour chaque valeur propre de A , un vecteur propre de norme 1.
3. Soit \mathcal{E}_c l'ellipse d'équation

$$\begin{bmatrix} x & y \end{bmatrix} A^{-1} \begin{bmatrix} x \\ y \end{bmatrix} = c^2 \quad (1)$$

Que vaut son demi-grand axe ? Son demi-petit axe ?

4. Calculer les points d'intersection de l'ellipse avec les axes de coordonnées, c'est-à-dire les points de l'ellipse de la forme $(x, 0)$ et $(0, y)$ avec $x, y \in \mathbb{R}$.
5. Esquisser l'ellipse \mathcal{E}_c dans le plan (x, y) pour $c = 1$ et $c = 4$.
6. Si A est la matrice de covariance d'un jeu de données, que représentent les ellipses \mathcal{E}_c ?

Problème 2 [6 points]

Le tableau de contingence suivant montre les effectifs pour un jeu de données croisant deux variables qualitatives. Les variables ont 4, respectivement 3 modalités. Celles de la première sont notées A, B, C, D, alors que les celles de la seconde sont notées a, b, c.

	a	b	c
A	10	170	20
B	100	100	200
C	30	110	60
D	60	20	120

1. Expliquer la propriété d'équivalence distributionnelle. Utiliser cette propriété afin de simplifier le tableau de contingence ci-dessus.
2. Effectuer un test d'indépendance du chi-deux *sur le tableau simplifié*, en expliquant les étapes principales du calcul. Peut-on rejeter l'hypothèse d'indépendance à un ou plusieurs seuils indiqués ci-dessous ?

Tableau de valeurs de la loi du χ^2 :

Degrés de liberté	90%	95%	97.5%	99%	99.9%
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.125
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588
11	17.275	19.675	21.920	24.725	31.264
12	18.549	21.026	23.337	26.217	32.910

3. Calculer, *pour le tableau simplifié*, les matrices X et Y des profils-ligne et des profils-colonne. Il est conseillé d'écrire ces matrices sous forme de fractions.
Déterminer, avec un minimum de calculs, les valeurs propres des matrices $S = X^T Y^T$ et $T = Y^T X^T$.

Problème 3 [8 points]

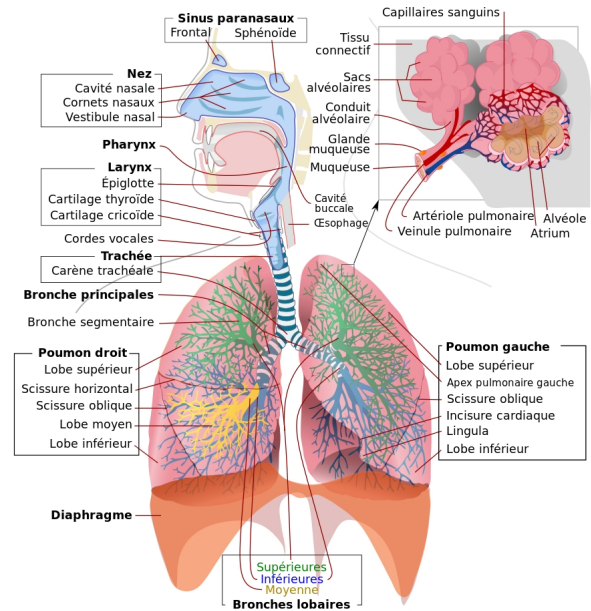
Une étude du marché européen dans le segment des antibiotiques, compilée à partir de 45 millions de prescriptions émises par 40 000 docteurs, a permis d'établir un tableau de contingence comprenant 300 maladies et 60 médicaments.

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.

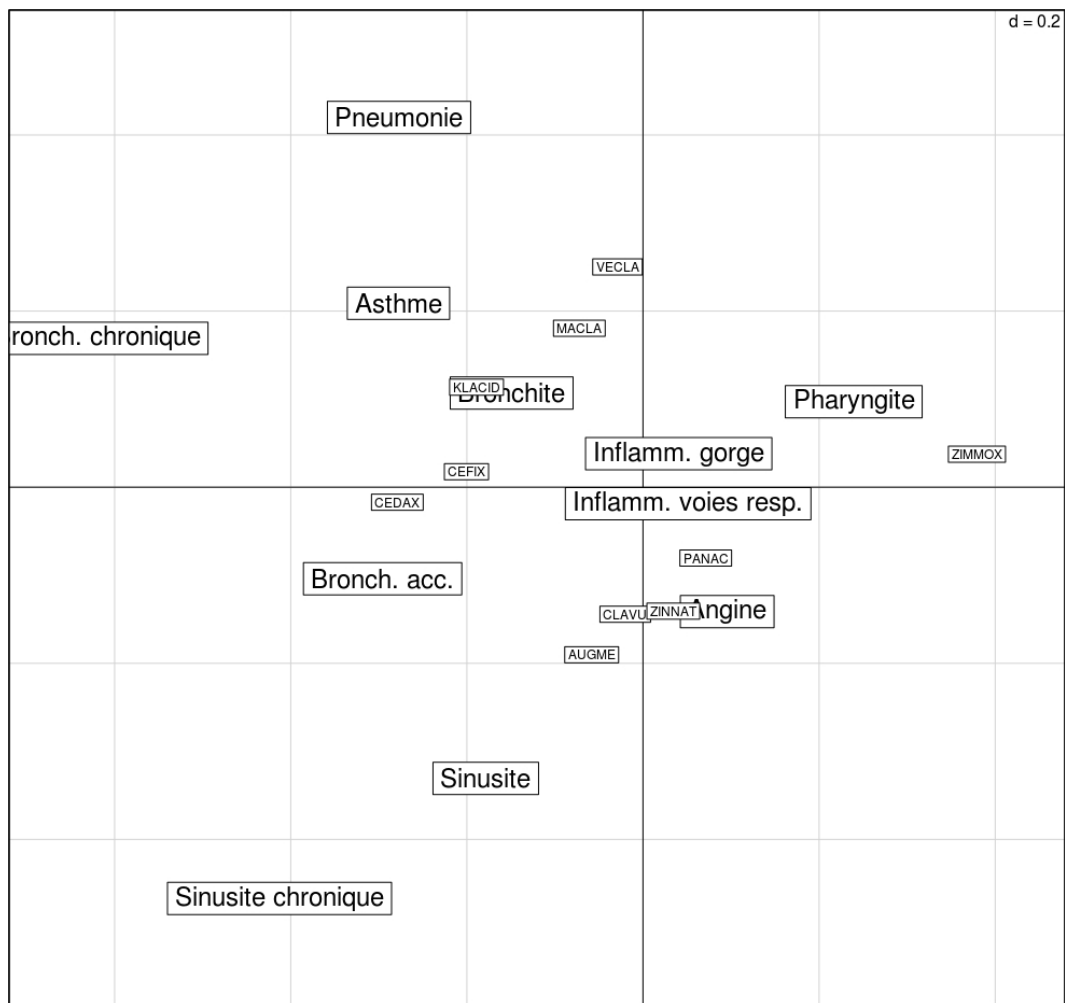
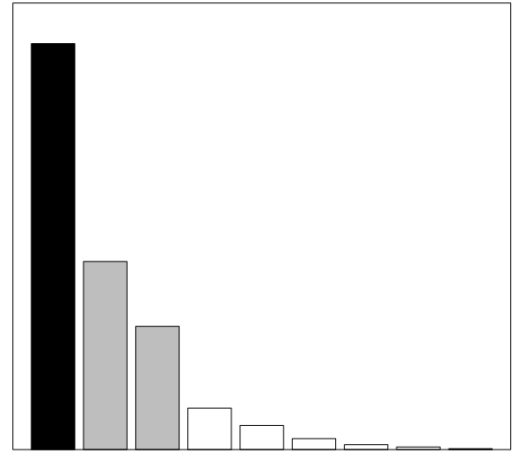


Les nombres sont en milliers de prescriptions.

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 (multipliées par 10000) :

> afcin\$row.abs			> afcin\$row.rel		
	Axis1	Axis2		Axis1	Axis2 con.tra
Asthme	695	848	Asthme	-4431	2506 794
Bronchite	696	772	Bronchite	-4320	2220 815
Bronch. acc.	2324	621	Bronch. acc.	-8117	-1004 1449
Pharyngite	3101	1100	Pharyngite	8109	1333 1935
Inflamm. voies resp.	90	26	Inflamm. voies resp.	1347	-181 337
Sinusite	117	864	Sinusite	-1730	-5940 341
Angine	514	2425	Angine	2636	-5759 987
Inflamm. gorge	25	47	Inflamm. gorge	441	388 287
Bronch. chronique	1630	261	Bronch. chronique	-7200	535 1146
Pneumonie	366	1816	Pneumonie	-1418	3257 1307
Sinusite chronique	442	1221	Sinusite chronique	-3712	-4753 602

Contributions absolues et relatives des colonnes (multipliées par 10000) :

> afcin\$col.abs			> afcin\$col.rel		
	Comp1	Comp2		Comp1	Comp2 con.tra
AUGME	147	3377	AUGME	-703	-7505 1055
CEDAX	1252	10	CEDAX	-6569	-25 965
CEFIX	843	14	CEFIX	-3476	27 1227
CLAVU	7	801	CLAVU	-159	-8132 231
KLACID	987	769	KLACID	-4178	1508 1195
MACLA	113	1521	MACLA	-1032	6433 554
PANAC	98	270	PANAC	1425	-1819 347
VECLA	17	2667	VECLA	-113	8422 743
ZIMMOX	6524	137	ZIMMOX	9666	94 3416
ZINNAT	12	435	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 ?