



FACULTÉ DE DROIT
ÉCONOMIE & GESTION

Sujet d'examen des semestres impairs 1^{ère} session
Année universitaire 2023-2024

Intitulé de l'épreuve :	Analyse de données qualitatives
Nom de l'enseignant :	Nils Berglund
Année :	M1 ESA
Mention / Spécialité / Parcours :	2023-2024
Durée de l'épreuve :	2 heures
Documents autorisés :	Résumé manuscrit de 4 pages A4
Matériels autorisés :	Calculatrice non programmable
	10/01/2024

P1/7

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

$$A = \begin{bmatrix} 17 & -6 \\ -6 & 8 \end{bmatrix}.$$

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 $c = \frac{1}{2}$.
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 respectivement 4 et 3 modalités. Celles de la première sont notées A, B, C, D, alors que celles de la seconde sont notées a, b, c.

	a	b	c
A	10	30	20
B	10	10	20
C	20	0	40
D	0	40	0

1. Donner la définition de la propriété d'équivalence distributionnelle. Utiliser cette propriété afin de simplifier le tableau de contingence ci-dessus.
2. Illustrer la propriété d'équivalence distributionnelle dans un cas particulier (pour un choix particulier de lignes ou de colonnes).
3. 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

4. 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]

Le tableau sur la page suivante montre les résultats de l'élection présidentielle américaine de novembre 2020. Il représente, pour les 50 états plus le District of Columbia (DC), le nombre de voix comptées pour les candidats suivants (source : *Washington Post*) :

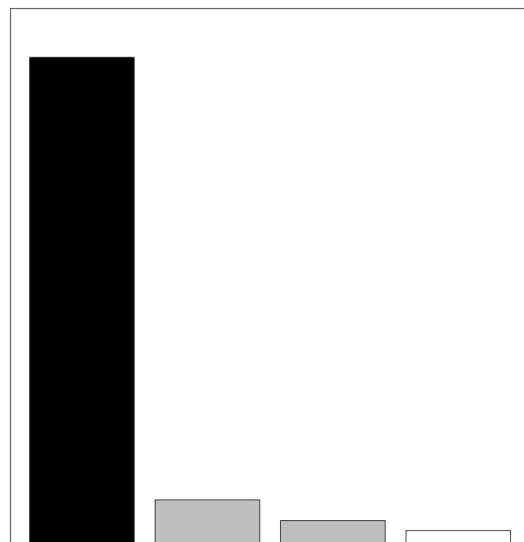
- Joseph Biden (Démocrates)
- Donald Trump (Républicains)
- Jo Jorgensen (Libéraux)
- Autres, dont Kanye West, Howie Hawkins.
- Non listés (noms ajoutés au bulletin électoral).

Ci-dessous on a représenté les résultats d'une AFC faite sur ces données.

Questions :

1. Expliquer les lignes suivantes fournies par le logiciel R:
 Number of cases in table: 157815214
 Number of factors: 2
 Test for independence of all factors:
 Chisq = 6199279, df = 200, p-value = 0
2. Quelle est la proportion de l'inertie contenue dans le premier axe factoriel ? Quelle est la proportion de l'inertie contenue dans le plan défini par les deux premier axes factoriels ?
3. Expliquer le rôle des contributions absolues dans une AFC, ainsi que celui des contributions relatives.
4. Quelles sont les modalités mal représentées sur le plan factoriel ? Expliquez comment vous trouvez ces modalités.
5. Quelles sont les modalités contribuant fortement au premier axe factoriel ? Donner une description résumant la signification de cet axe.
6. Quelles sont les modalités contribuant fortement au second axe factoriel ? Donner une description résumant la signification de cet axe.
7. Quels sont les états les plus mal représentés sur le *premier axe* factoriel ? Que peut-on dire à leur sujet ?
8. Discuter les cas particuliers des états suivants : DC (District of Columbia), Wyoming (WY), Arkansas (AR), Virginie (VA), Nevada (NV).

```
> afcin$TOT
      inertia      cum      cum(%)
Ax1 0.033165191 0.03316519 84.42872
Ax2 0.003194416 0.03635961 92.56075
Ax3 0.001798445 0.03815805 97.13906
Ax4 0.001123830 0.03928188 100.00000
```



État	Abrév.	Biden	Trump	Jorgensen	Autres	Non listé
Alabama	AL	849624	1441170	25176	0	7312
Alaska	AK	153778	189951	8897	2673	1961
Arizona	AZ	1672143	1661686	51465	0	1928
Arkansas	AR	423932	760647	13133	21358	0
California	CA	11098676	5994674	187885	192216	80
Colorado	CO	1804196	1364471	52454	34449	0
Connecticut	CT	1080680	715291	20227	7541	580
Delaware	DE	296268	200603	5000	2139	0
DC	DC	317323	18586	2036	3274	3137
Florida	FL	5297045	5668731	70324	30301	1055
Georgia	GA	2474507	2461837	62138	0	84
Hawaii	HI	366130	196864	5539	5005	0
Idaho	ID	287021	554118	16404	9817	745
Illinois	IL	3461763	2438696	65991	48000	5547
Indiana	IN	1242495	1729852	58900	0	1951
Iowa	IA	758881	897467	19726	10766	1470
Kansas	KS	551144	752903	29466	0	291
Kentucky	KY	772474	1326646	26234	10082	0
Louisiana	LA	856034	1255776	21645	14607	0
Maine	ME	431277	360446	14036	9452	0
Maryland	MD	1985023	976414	33488	21683	0
Massachusetts	MA	2382202	1167202	47009	18658	16337
Michigan	MI	2804040	2649852	60381	23939	1090
Minnesota	MN	1717077	1484065	34976	10033	9965
Mississippi	MS	539494	756731	8024	9568	1423
Missouri	MO	1253561	1717912	41165	8307	4723
Montana	MT	244786	343602	15252	0	2110
Nebraska	NE	374583	556846	20283	0	559
Nevada	NV	703486	669890	14738	18217	0
New Hampshire	NH	424921	365654	13235	0	620
New Jersey	NJ	2608271	1883140	31673	26065	14879
New Mexico	NM	501614	401894	12585	6232	0
New York	NY	4943074	3148139	57464	52416	0
North Carolina	NC	2684306	2758779	48678	12194	13315
North Dakota	ND	114902	235595	9393	0	1929
Ohio	OH	2679165	3154734	67569	18812	2225
Oklahoma	OK	503890	1020280	24731	9251	0
Oregon	OR	1340291	958411	41581	11832	14078
Pennsylvania	PA	3459998	3378362	79400	0	6423
Rhode Island	RI	306210	199837	5047	2533	0
South Carolina	SC	1091541	1385103	27016	8769	0
South Dakota	SD	150471	261043	11095	0	0
Tennessee	TN	1139376	1849820	29842	26898	2044
Texas	TX	5259126	5890347	126243	33396	8799
Utah	UT	560282	865140	38477	23808	612
Vermont	VT	242820	112704	3608	5742	1942
Virginia	VA	2413568	1962430	64761	0	19765
Washington	WA	2369437	1584588	80489	25614	166
West Virginia	WV	235984	545382	10687	2599	0
Wisconsin	WI	1630716	1610151	38333	5266	7825
Wyoming	WY	73491	193559	5768	2208	1739
TOTAL		80933097	74078021	1859667	785720	158709

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

```
> round(afcin$col.abs*100)      > round(afcin$col.rel*100)
      Axis1  Axis2                Axis1 Axis2
Biden    4641    77      Biden    9966   16
Trump    5170    19      Trump   -9988   -4
Jorgensen  20    13      Jorgensen -513   31
Autres   165   6263      Autres   1742 -6376
Non liste  4   3629      Non liste  61  5240
```

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

```
> round(afcin$row.abs*100)      > round(afcin$row.rel*100)
      Axis1  Axis2                Axis1 Axis2
AL      415   217      AL -9309  469
AK       14    16      AK -3506  391
AZ       17   136      AZ -3352 2584
AR      205   874      AR -6437 -2644
CA     2252  1415      CA  9426 -570
CO       62   337      CO  5658 -2964
CT       84    4      CT  9353   46
DE       20    0      DE  9057    2
DC      465   216      DC  9228  412
FL      135    0      FL -6063   -1
GA       25   112      GA -3378 1460
HI       73    17      HI  9735 -215
ID      194   188      ID -8490 -792
IL      202    67      IL  9621 -310
IN      272    63      IN -8929  200
IA       49    30      IA -9187 -546
KS      111    18      KS -8161  127
KY      378    83      KY -9787 -206
LA      210   161      LA -9195 -677
ME        4   125      ME  1962 -5356
MD      501    17      MD  9828  -33
MA      598   608      MA  8714   854
MI        3    10      MI -2746 -881
MN        4   253      MN  1196 6737
MS      103    54      MS -8474 -428
MO      235    19      MO -9864   78
MT       56    93      MT -6887 1104
NE      112    16      NE -8912  122
NV        0   343      NV  -17 -8587
NH        1    50      NH  465  3975
NJ      126   174      NJ  6967   926
NM        8    19      NM  7137 -1697
NY      503    66      NY  9258 -116
NC       38   292      NC -4748 3510
ND      104    82      ND -8151  620
OH      184    1      OH -9716   -7
OK      420   125      OK -9664 -276
```

OR	64	530	OR	3423	2714
PA	22	364	PA	-2760	4371
RI	26	0	RI	9359	-5
SC	128	15	SC	-9604	-111
SD	83	1	SD	-8666	14
TN	421	348	TN	-8887	-708
TX	229	18	TX	-9649	75
UT	155	701	UT	-5213	-2272
VT	79	1	VT	7966	-8
VA	21	1430	VA	1197	7814
WA	172	23	WA	7413	-96
WV	291	25	WV	-9918	-81
WI	11	239	WI	-3237	6552
WY	118	3	WY	-8285	23

