



MINISTERE
DU TRAVAIL ET DE LA FORMATION
PROFESSIONNELLE,
*en charge de la fonction publique
et de la recherche*

DIRECTION GENERALE
DES RESSOURCES HUMAINES
.....

CONCOURS EXTERNE POUR LE RECRUTEMENT D'ATTACHES D'ADMINISTRATION

<u>TROISIEME EPREUVE D'ADMISSIBILITE</u>

Une composition de mathématiques et statistiques.

SPECIALITE : STATISTIQUES

Jeudi 10 août 2017

(Durée : 4 heures – coefficient 2)

Le sujet comporte 14 pages (page de garde incluse).

La calculatrice est autorisée.

Important :

Il est interdit aux candidats de signer leur composition ou d'y mettre un signe quelconque pouvant indiquer la provenance de la copie. Les compositions doivent rester anonymes.

Liste des documents :

- Document 1 (10 pages) :
Problème : énoncé et annexe.

- Document 2 (1 page) :
Exercice 1.

- Document 3 (1 page) :
Exercice 2.

DOCUMENT 1

PROBLEME

On considère le tableau de données décrivant pour 9 pays : la consommation pour 6 types d'aliments (en kg, par personne et par an, en 1998), les mortalités par maladie cardiovasculaire et par cancer (pour 100000 personnes, en 1998) ainsi que les durées de longévité homme et femme (en années).

(Ce tableau est donné en annexe, les noms et les libellés des variables figurent aussitôt après ce tableau de données).

1. On donne la moyenne et l'écart-type des 8 premières variables (voir annexe). Calculer ces caractéristiques statistiques pour les variables **longfem** et **longhom**. Commenter leurs coefficients de variation.
2. Un tableau, en annexe, donne les coefficients de corrélations linéaires entre les variables. Compléter ce tableau en calculant le coefficient de corrélation linéaire entre la variable **longfem** et la variable **longhom**. Interpréter ce tableau des corrélations ainsi que les valeurs-tests calculées à partir de ces coefficients.
3. On a effectué une Analyse en composantes principales normée (ACPN) sur le tableau des données (tous les résultats associés figurent en annexe).
 - 3.1 Donner les objectifs de cette méthode et justifier son utilisation.
 - 3.2 Interpréter le tableau des valeurs propres de la matrice des corrélations.
 - 3.3 Etudier la position des variables dans le plan formé par les 2 premières composantes principales et en donner une interprétation. (On pourra s'aider d'une représentation graphique).
Donner une signification des 2 premières composantes principales.
 - 3.4 En annexe, les tableaux fournissent les coordonnées, les contributions ainsi que les cosinus carrés des pays avec les premiers axes factoriels. En s'intéressant uniquement aux 2 premiers axes factoriels, donner une interprétation de l'ensemble de ces résultats.
 - 3.5 Donner une synthèse des résultats obtenus avec cette ACPN.
4. Un découpage en 2 classes de chacune des variables a été effectué (la classe 1 étant codée 1 et la classe 2 étant codée 2). Le tableau de recodage est donné en annexe. Les libellés des variables ne changent pas. Puis une Analyse des Correspondances Multiples (ACM) est faite sur le tableau des données mis sous forme disjonctive complète. Les résultats sont fournis en annexe.
 - 4.1 Donner le tri à plat de la variable « poisson ».
 - 4.2 Combien y a-t-il de pays ayant une mortalité cardiovasculaire supérieure ou égale à 170 et une consommation de fruits et légumes strictement inférieure à 200 ?
Parmi les pays ayant une consommation de fruits et légumes strictement inférieure à 200, combien y en a-t-il, en proportion, qui ont une mortalité cardiovasculaire supérieure ou égale à 170 ?
Parmi les 9 pays, quelle est la proportion de ceux qui ont une mortalité cardiovasculaire supérieure ou égale à 170 ?
 - 4.3 Combien existe-t-il d'axes de l'ACM potentiellement significatifs?
 - 4.4 Quel est le pourcentage d'inertie expliqué par le plan factoriel principal ?
 - 4.5 En utilisant les résultats de l'annexe, étudier les axes 1 et 2. Donner à ces axes une interprétation.
 - 4.6 Donner une synthèse des résultats obtenus avec cette ACM.

ANNEXE

	huilevege	huilanim	fruitleg	viande	poisson	lait	cardvasc	cancer	longhom	longfem
Japon	12	2,1	166,4	43,2	71,2	68,7	110	106	77	84
France	16,5	18,1	221,6	96,8	28,3	262	108	131	74	82
Royaume-Uni	16	9,4	171,5	73,5	20,2	223	193	137	74	80
Espagne	26,9	4	251,6	99,6	37,3	161	143,8	120,8	74	80
Pays-Bas	17	8,2	224,4	86,4	14,4	340	160,8	136,7	74	80
Norvège	13,2	16,5	178,4	59,4	61,3	265	174,4	121,7	73	80
Etats-Unis	24,4	6,5	237	118,4	21,7	255	188	131	73	79
Mexique	11,5	3,7	145,7	44,9	10,3	99,8	174,7	81,2	69	74
Hongrie	14,3	25,6	164,1	87,7	4,5	161	369,9	191,9	64	73

Noms des variables	Libellés des variables
Huiles végétales	huilevege
Huiles et graisses animales	huilanim
Fruits et légumes	fruitleg
Viandes	viande
Poissons	poisson
Produits laitiers	lait
Mortalité par maladie cardio-vasculaire	cardvasc
Mortalité par cancer	cancer
Longévité homme	longhom
Longévité femme	longfem

Libellés des variables	Moyenne	Écart-type
huilevege	16,867	5,064
huilanim	10,456	7,478
fruitleg	195,633	35,850
viande	78,878	24,228
poisson	29,911	21,551
lait	204,000	82,636
cardvasc	180,289	73,119
cancer	128,589	27,930

Tableau des corrélations

	huilevege	huilanim	fruitleg	viande	poisson	lait	cardvasc	cancer	longhom	longfem
huilevege	1,000									
huilanim	-0,237	1,000								
fruitleg	0,883	-0,152	1,000							
viande	0,824	0,233	0,819	1,000						
poisson	-0,129	-0,269	-0,021	-0,406	1,000					
lait	0,276	0,314	0,531	0,549	-0,227	1,000				
cardvasc	-0,116	0,639	-0,340	0,158	-0,552	-0,058	1,000			
cancer	0,126	0,776	0,101	0,509	-0,391	0,322	0,760	1,000		
longhom	0,210	-0,604	0,394	-0,058	0,655	0,164	-0,888	-0,513	1,000	
longfem	0,148	-0,357	0,387	-0,006	0,724	0,183	-0,792	-0,296		1,000

Matrice des valeurs-tests

	huilevege	huilanim	fruitleg	viande	poisson	lait	cardvasc	cancer	longhom	longfem
huilevege	99,99									
huilanim	-0,73	99,99								
fruitleg	4,17	-0,46	99,99							
viande	3,50	0,71	3,46	99,99						
poisson	-0,39	-0,83	-0,06	-1,29	99,99					
lait	0,85	0,97	1,78	1,85	-0,69	99,99				
cardvasc	-0,35	2,27	-1,06	0,48	-1,87	-0,17	99,99			
cancer	0,38	3,10	0,30	1,69	-1,24	1,00	2,99	99,99		
longhom	0,64	-2,10	1,25	-0,18	2,35	0,50	-4,23	-1,70	99,99	
longfem	0,45	-1,12	1,23	-0,02	2,75	0,56	-3,23	-0,92	5,38	99,99

RESULTATS CONCERNANT L'ANALYSE EN COMPOSANTES PRINCIPALES NORMEE

Tableau des valeurs propres

Numéro	Valeur propre	Pourcentage	Pourcentage cumulé
1	4,256	42,6	42,6
2	3,256	32,6	75,1
3	1,266	12,7	87,8
4	0,711	7,1	94,9
5	0,251	2,5	97,4
6	0,185	1,9	99,2
7	0,049	0,5	99,7
8	0,027	0,3	100,0
9	0,000	0,0	100,0
10	0,000	0,0	100,0

Tableau des corrélations des variables avec les facteurs (composantes principales)

Libellé de la variable	Axe 1	Axe 2	Axe 3
huilevege	0,153	0,849	0,398
huilanim	-0,733	0,115	-0,599
fruitleg	0,300	0,924	0,112
viande	-0,215	0,950	0,097
poisson	0,711	-0,250	-0,413
lait	-0,054	0,682	-0,417
cardvasc	-0,946	-0,048	-0,005
cancer	-0,731	0,400	-0,404
longhom	0,947	0,177	-0,187
longfem	0,850	0,207	-0,429

Tableau des coordonnées des pays

Pays	Axe 1	Axe 2	Axe 3
Japon	2,987	-2,460	-0,529
France	0,598	1,257	-1,100
Royaume-Uni	-0,074	-0,268	-0,273
Espagne	1,362	1,910	1,382
Pays-Bas	0,232	1,435	-0,348
Norvège	0,462	-0,886	-1,653
Etats-Unis	0,127	2,430	0,930
Mexique	-0,537	-3,005	2,007
Hongrie	-5,157	-0,413	-0,415

Tableau des contributions des pays

Pays	Axe 1	Axe 2	Axe 3
Japon	23,291	20,655	2,461
France	0,933	5,395	10,625
Royaume-Uni	0,014	0,245	0,655
Espagne	4,844	12,452	16,765
Pays-Bas	0,141	7,023	1,061
Norvège	0,557	2,681	23,988
Etats-Unis	0,042	20,152	7,590
Mexique	0,753	30,814	35,345
Hongrie	69,426	0,581	1,511

Tableau des cosinus carrés des pays

Pays	Axe 1	Axe 2	Axe 3
Japon	0,533	0,361	0,017
France	0,078	0,347	0,265
Royaume-Uni	0,005	0,060	0,063
Espagne	0,211	0,414	0,217
Pays-Bas	0,012	0,460	0,027
Norvège	0,044	0,161	0,559
Etats-Unis	0,002	0,837	0,123
Mexique	0,020	0,624	0,278
Hongrie	0,956	0,006	0,006

TABLEAU DE RECODAGE DES DONNEES

Libellés des variables	Classe 1	Classe2
huilevege	<15	≥ 15
huilanim	<8	≥ 8
fruitleg	<200	≥ 200
viande	<80	≥ 80
poisson	<25	≥ 25
lait	<200	≥ 200
cardvasc	<170	≥ 170
cancer	<130	≥ 130
longhom	<74	≥ 74
longfem	<80	≥ 80

RESULTATS CONCERNANT L'ANALYSE DES CORRESPONDANCES MULTIPLES

Tri à plat des variables

Libellés des variables	Libellés des modalités	Effectifs
huilevege	1	4
	2	5
huilanim	1	4
	2	5
fruitleg	1	5
	2	4
viande	1	4
	2	5
poisson	1	5
	2	4
lait	1	4
	2	5
cardvasc	1	4
	2	5
cancer	1	4
	2	5
longhom	1	4
	2	5
longfem	1	3
	2	6

Tableau de Burt

	huilevege		huilanim		fruitleg		viande		poisson		lait		cardvasc		cancer		longhom		longfem	
	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
1	4	0																		
2	0	5																		
1	2	2	4	0																
2	2	3	0	5																
1	4	1	2	3	5	0														
2	0	4	2	2	0	4														
1	3	1	2	2	4	0	4	0												
2	1	4	2	3	1	4	0	5												
1	2	3	2	3	3	2	2	3	5	0										
2	2	2	2	2	2	2	2	2	0	4										
1	3	1	3	1	3	1	2	2	2	2	4	0								
2	1	4	1	4	2	3	2	3	3	2	0	5								
1	1	3	2	2	1	3	1	3	1	3	2	2	4	0						
2	3	2	2	3	4	1	3	2	4	1	2	3	0	5						
1	3	1	3	1	3	1	3	1	1	3	3	1	2	2	4	0				
2	1	4	1	4	2	3	1	4	4	1	1	4	2	3	0	5				
1	3	1	2	2	3	1	2	2	3	1	2	2	0	4	2	2	4	0		
2	1	4	2	3	2	3	2	3	2	3	2	3	4	1	2	3	0	5		
1	2	1	2	1	2	1	1	2	3	0	2	1	0	3	1	2	3	0	3	0
2	2	4	2	4	3	3	3	3	2	4	2	4	4	2	3	3	1	5	0	6

Tableau de valeurs propres

Numéro	Valeur propre	Pourcentage	Pourcentage cumulé
1	0,376	37,61	37,6
2	0,278	27,77	65,4
3	0,173	17,29	82,7
4	0,071	7,13	89,8
5	0,069	6,92	96,7
6	0,016	1,62	98,4
7	0,014	1,41	99,8
8	0,002	0,24	100,0
9	0,000	0,00	100,0
10	0,000	0,00	100,0

Tableau des coordonnées des pays

Pays	Axe 1	Axe 2	Axe 3
Japon	0,270	0,960	0,034
France	-0,890	0,040	-0,111
Royaume-Uni	-0,124	-0,279	-0,679
Espagne	-0,476	0,644	0,580
Pays-Bas	-0,817	-0,268	-0,077
Norvège	0,483	0,257	-0,634
Etats-Unis	-0,046	-0,777	0,497
Mexique	1,022	0,007	0,280
Hongrie	0,577	-0,584	0,111

Tableau des contributions des pays

Pays	Axe 1	Axe 2	Axe 3
Japon	2,148	36,847	0,074
France	23,391	0,065	0,795
Royaume-Uni	0,452	3,124	29,619
Espagne	6,686	16,617	21,597
Pays-Bas	19,697	2,872	0,381
Norvège	6,901	2,651	25,844
Etats-Unis	0,061	24,181	15,862
Mexique	30,840	0,002	5,042
Hongrie	9,825	13,641	0,786

Tableau des cosinus carrés des pays

Pays	Axe 1	Axe 2	Axe 3
Japon	0,067	0,849	0,001
France	0,875	0,002	0,014
Royaume-Uni	0,019	0,096	0,566
Espagne	0,218	0,399	0,323
Pays-Bas	0,775	0,083	0,007
Norvège	0,235	0,067	0,404
Etats-Unis	0,002	0,573	0,234
Mexique	0,877	0,000	0,066
Hongrie	0,315	0,323	0,012

Tableau des coordonnées des modalités des variables

Libellé	Axe 1	Axe 2	Axe 3
huilevege			
1	0,959	0,304	-0,126
2	-0,767	-0,243	0,101
huilanim			
1	0,314	0,395	0,836
2	-0,251	-0,316	-0,669
fruitleg			
1	0,726	0,137	-0,427
2	-0,908	-0,171	0,534
viande			
1	0,673	0,448	-0,601
2	-0,538	-0,358	0,480
poisson			
1	0,200	-0,722	0,063
2	-0,250	0,902	-0,079
lait			
1	0,568	0,487	0,604
2	-0,454	-0,390	-0,483
cardvasc			
1	-0,780	0,653	0,256
2	0,624	-0,522	-0,205
cancer			
1	0,530	0,886	0,156
2	-0,424	-0,709	-0,125
longhom			
1	0,830	-0,520	0,152
2	-0,664	0,416	-0,122
longfem			
1	0,844	-0,857	0,711
2	-0,422	0,428	-0,356

Tableau des contributions des modalités des variables

Libellé	Axe 1	Axe 2	Axe 3
huilevege			
1	10,856	1,476	0,408
2	8,685	1,181	0,326
TOTAL	19,541	2,656	0,734
huilanim			
1	1,164	2,503	17,970
2	0,931	2,002	14,376
TOTAL	2,096	4,505	32,346
fruitleg			
1	7,795	0,375	5,866
2	9,744	0,469	7,333
TOTAL	17,539	0,843	13,199
viande			
1	5,352	3,213	9,272
2	4,282	2,571	7,418
TOTAL	9,634	5,784	16,690
poisson			
1	0,589	10,420	0,129
2	0,737	13,025	0,161
TOTAL	1,326	23,445	0,290
lait			
1	3,806	3,800	9,374
2	3,045	3,040	7,499
TOTAL	6,852	6,840	16,873
cardvasc			
1	7,181	6,823	1,683
2	5,745	5,458	1,346
TOTAL	12,927	12,281	3,029
cancer			
1	3,313	12,574	0,627
2	2,650	10,059	0,501
TOTAL	5,964	22,633	1,128
longhom			
1	8,141	4,334	0,596
2	6,512	3,467	0,477
TOTAL	14,653	7,800	1,074
longfem			
1	6,313	8,808	9,759
2	3,156	4,404	4,879
TOTAL	9,469	13,212	14,638

Tableau des cosinus carrés des modalités des variables

Libellé	Axe 1	Axe 2	Axe 3
huilevege			
1	0,735	0,074	0,013
2	0,735	0,074	0,013
huilanim			
1	0,079	0,125	0,559
2	0,079	0,125	0,559
fruitleg			
1	0,660	0,023	0,228
2	0,660	0,023	0,228
viande			
1	0,362	0,161	0,289
2	0,362	0,161	0,289
poisson			
1	0,050	0,651	0,005
2	0,050	0,651	0,005
lait			
1	0,258	0,190	0,292
2	0,258	0,190	0,292
cardvasc			
1	0,486	0,341	0,052
2	0,486	0,341	0,052
cancer			
1	0,224	0,629	0,020
2	0,224	0,629	0,020
longhom			
1	0,551	0,217	0,019
2	0,551	0,217	0,019
longfem			
1	0,356	0,367	0,253
2	0,356	0,367	0,253

DOCUMENT 2

EXERCICE 1

α étant un nombre réel strictement positif, on considère la variable aléatoire U suivant la loi $\gamma(1, \alpha)$ dont la densité de probabilité est donnée par : $f_U(x) = \begin{cases} \alpha e^{-\alpha x} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$

1. Exprimer en fonction α , la covariance entre la variable U et la variable logarithme népérien de U (notée $\ln U$)

2. β étant un réel strictement positif, on définit la variable aléatoire X par $X = U^{1/\beta}$

On notera $W(\alpha, \beta)$ la loi de X (α et β étant des paramètres)

Calculer la densité de probabilité de la loi de X . Quels sont ses moments?

3. Pour une variable aléatoire Y positive dont la loi est à densité de probabilité f et la fonction de répartition F (on a $F(0) = 0$ et on supposera $F(t) < 1$ pour tout t positif), on définit :

- La « fonction de survie », notée S , telle que : $S(t) = 1 - F(t)$ dont on supposera par la suite que : $\lim_{t \rightarrow \infty} S(t) = 0$
- La « fonction de hasard », notée h , définie par $h(t) = \frac{f(t)}{S(t)}$ si t positif
- La « durée de vie moyenne restante », notée r , définie pour tout t positif par l'espérance conditionnelle de la variable aléatoire $Y - t$ conditionnée par $Y > t$ (notée $r(t) = E(Y - t / Y > t)$)

3.1 Exprimer S en fonction de h

3.2 Caractériser la loi $W(\alpha, \beta)$ par sa fonction de hasard.

3.3 Quelles sont les lois de probabilité dont la fonction de hasard est constante ?

3.4 Exprimer r en fonction de S puis calculer r pour la loi $\gamma(1, \alpha)$. Conclusion.

4. La variable aléatoire Y représente une durée de vie dont la fonction de hasard associée est ici définie par : $h(y/z, \theta) = \exp(-z\theta) \cdot h_0(y)$ où θ représente un paramètre réel et z une variable explicative de la durée de vie.

h_0 s'appelle « fonction de hasard de base » et $H_0(t) = \int_0^t h_0(u) du$ est la « fonction de hasard cumulé de base » que l'on supposera strictement monotone.

4.1 Exprimer la fonction de survie $S(y/z, \theta)$ en fonction de $H_0(y)$

4.2 On considère la variable aléatoire $T = \ln H_0(y) - z\theta$

Etudier la loi de la variable T . Quelle est la loi de la variable $\exp(T)$?

Préciser son espérance mathématique et sa variance. Conclusion.

Montrer que l'on peut écrire $H_0(Y)$ sous la forme $H_0(Y) = \exp(z\theta) + V$ où V est une variable aléatoire dont on précisera son espérance mathématique et sa variance.

DOCUMENT 3

EXERCICE 2

θ étant un paramètre réel strictement positif, on considère la variable aléatoire X suivant une loi normale ayant comme moyenne et comme variance le paramètre θ .

Soit un n -échantillon théorique $(X_1, \dots, X_n)$ de cette loi et $(x_1, \dots, x_n)$ un n -échantillon observé associé.

1. Montrer que $T = \sum_{i=1}^n X_i^2$ est une statistique exhaustive.

Dans la suite du problème, on considèrera $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ variable aléatoire strictement positive.

2. Estimer θ par la méthode du maximum de vraisemblance, on notera $\hat{\theta}_n$ cet estimateur. Est-il un estimateur convergent de θ ?
3. Calculer la quantité d'information (au sens de Fisher) apportée par l'échantillon sur le paramètre θ . En déduire la loi asymptotique de $\hat{\theta}_n$.
4. A partir de l'espérance mathématique et de la variance de X , définir deux estimateurs sans biais de θ (méthode des moments).

En utilisant une fonction de coût quadratique, comparer ces deux estimateurs dans les cas suivants :

- 4.1 En étudiant les risques associés aux estimateurs.
- 4.2 En utilisant le critère minimax et en supposant que θ est borné par un réel A suffisamment grand.
- 4.3 Dans le cadre Bayésien, en supposant que la loi a priori de la variable aléatoire θ est une loi uniforme sur $[0, a]$ avec $a > 0$

5. On se situe ici dans le cas classique (non Bayésien).

On considère $C(\lambda)$, avec $0 \leq \lambda \leq 1$, la classe des estimateurs de θ de la forme :

$$T(\lambda) = \lambda \bar{X}_n + (1 - \lambda) S_n^2 \quad \text{avec} \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Déterminer λ^* pour que $T(\lambda^*)$ soit un estimateur de θ sans biais, à variance minimale.

Comment procéder pour obtenir une valeur numérique de $T(\lambda^*)$?