Préparation à l'agrégation externe de Sciences Sociales

Statistique descriptive multivariée

2023-2024

Exercice 1 (2010)

Les données suivantes concernent l'économie mexicaine, entre 1955 et 1964. On dispose sur cette période du PIB (variable expliquée Q en millions de pesos) et des variables explicatives I (emploi en milliers de personnes) et K (capital en millions de pesos). Par ailleurs, on donne également les valeurs des variables Y, X_1 et X_2 qui sont respectivement les logarithmes népériens des variables Q, L et X ; ainsi que de y, x_1 et x_2 obtenues en centrant les variables Y,X_1 et X_2.

Année	Q	L	K	$Y = \ln Q$	$X_1 = \ln L$	$X_2 = \ln K$	y	x_1	x_2
1955	114043	8310	182113	11,64	9,03	12,11	-0,25	-0,12	-0,24
1956	120410	8529	193749	11,70	9,05	12,17	-0,20	-0,09	-0,18
1957	129487	8738	205192	11,77	9,08	12,23	-0,13	-0,06	-0,12
1958	134705	8952	215130	11,81	9,10	12,28	-0,09	-0,04	-0,07
1959	139960	9171	225021	11,85	9,12	12,32	-0,04	-0,02	-0,03
1960	150511	9569	237026	11,92	9,17	12,38	0,02	0,02	0,02
1961	157897	9527	248897	11,97	9,16	12,42	0,07	0,02	0,07
1962	165286	9662	260661	12,02	9,18	12,47	0,12	0,04	0,12
1963	178491	10334	275466	12,09	9,24	12,53	0,19	0,10	0,18
1964	199457	10981	295378	12,20	9,30	12,60	0,31	0,15	0,25

Le but de l'exercice est de tester un ajustement du type Cobb-Douglas : $Q = AL^{\alpha}K^{\beta}$.

1. Montrer, en utilisant des variables auxiliaires, que cet ajustement se ramène à un ajustement linéaire.
2. On note x la matrice de format 10×2 ayant x_1 et x_2 pour vecteurs colonnes et ${}^t x$ sa transposée.
   (a) Calculer la matrice de variance-covariance de (x_1, x_2) et en déduire le produit matriciel $B = {}^t xx$. On donnera les résultats numériques à 10^{-4} près.
   (b) Justifier l'inversibilité de B et calculer son inverse (on pourra utiliser directement la calculatrice).
3. On rappelle que si $a = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ alors l'estimateur donné par la méthode des moindres carrés ordinaire est : $\widehat{a} = ({}^t xx)^{-1}\, {}^t xy$.
   (a) En déduire les estimations de α, β et de A.
   (b) Donner les interprétations de α et β.
   (c) Que suggèrent les résultats quant aux rendements d'échelle ?

Exercice 2 (2011)

On dispose des données économiques pour 12 supermarchés $(s_1, \ldots, s_{12})$ d'une même entreprise de distribution. Pour chaque supermarché, on dispose des cinq variables suivantes :

CA = chiffre d'affaire ; AM = amortissement ; PTT = poste téléphone et transport ; RES = résultat d'exploitation ; CS = charge salariale.

Les tableaux et graphiques suivants sont les résultats d'une analyse en composantes principales, qui a pour objectif d'étudier d'une part, les liaisons entre les variables économiques, et d'autre part, les proximités entre supermarchés.

1. Calculer les pourcentages d'inertie de chaque axe.
2. Quel pourcentage de l'inertie totale explique le premier plan factoriel ?
3. Que représente la matrice des composantes principales ?
4. Expliquer le graphique cercle des corrélations et faire le lien avec la matrice des corrélations des variables initiales.
5. Calculer la qualité de représentation de s_1 (donnée manquante du dernier tableau).

Table 1 – Matrice des corrélations

	CA	AM	PTT	RES	CS
CA	1.000				
AM	-0.186	1.000			
PTT	0.007	-0.222	1.000		
RES	0.970	-0.105	-0.174	1.000	
CS	0.865	-0.330	0.322	0.728	1.000

Table 2 – Valeurs propres

1	2	3	4	5
2.793	1.306	0.769	0.129	0.003

Table 3 – Composantes principales

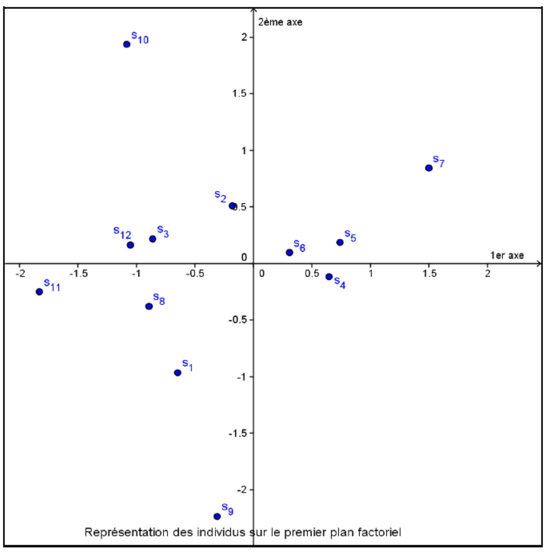
	1	2	3	4	5
s_1	-0.648	-0.967	0.090	-0.426	0.434
s_2	-0.180	0.508	-0.877	0.913	0.993
s_3	-0.858	0.214	-1.021	-0.317	0.587
s_4	0.648	-0.119	-0.114	-2.380	-0.245
s_5	0.742	0.185	-0.219	-0.224	-1.986
s_6	0.307	0.096	-0.358	-0.360	2. 010
s_7	1.499	0.844	-0.341	1. 338	-0.482
s_8	-0.892	-0.380	-0.532	-0.512	-0.239
s_9	-0.309	-2.237	1.407	1.005	0.137
s_{10}	-1.083	1.936	2.269	-0.115	0.000
s_{11}	-1.827	-0.253	0.689	0.099	0.559
s_{12}	-1.049	0.162	-0.994	0.978	-0.902

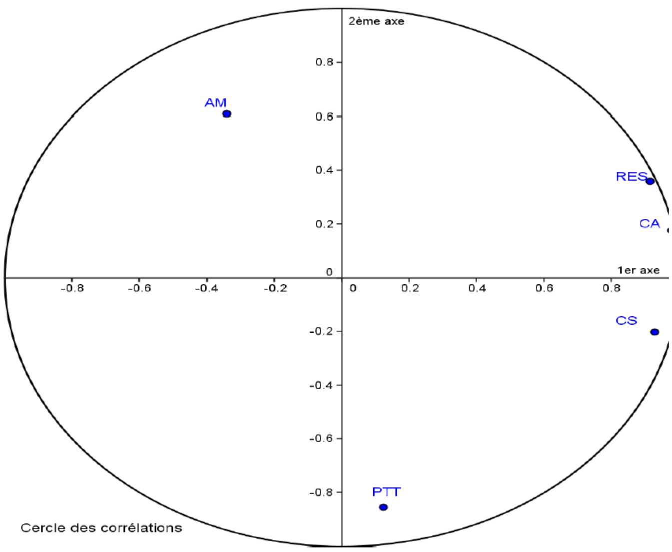
Table 4 – Corrélations entre variables initiales et composantes principales

	1	2	3	4	5
CA	0.979	0.177	0.060	0.063	-0.041
AM	-0.341	0.610	0.715	-0.036	0.000
PTT	0.124	-0.856	0.485	0.128	0.002
RES	0.915	0.359	-0.007	0.180	0.031
CS	0.929	-0.202	0.141	-0.274	0.012

Table 5 – Norme au carré et qualité de représentation sur le premier plan factoriel

s_1	1,73	……..%
s_2	2,88	10,09 %
s_3	2,27	34,46%
s_4	6,17	7,03%
s_5	4,63	12,64%
s_6	4,40	2,35%
s_7	5,10	58,05%
s_8	1,54	60,95%
s_9	8,11	62,90%
s_{10}	10,08	48,81%
s_{11}	4,20	81,02%
s_{12}	3,88	29,00%





Exercice 3 (Calculatrice)

On considère les donées suivantes concernant la consommation d'eau chaude en litres et la température sur une période de 5 jours.

Jour	1	2	3	4	5
X : température (en °C)	-6	-4	5	0	2
Y : consommation (en L)	40	36	23	32	28

1. Déterminer la moyenne de chaque série.
2. Répresenter le nuage de points associé à la série statistique et tracer la droite de régression de X en Y.