

MÉTHODE DES MOINDRES CARRÉS

Le tableau ci-après présente le classement des équipes jouant dans le championnat de Hand Ball Pro D2 en 2011-2012. Il y a 13 équipes et 8 variables ayant des valeurs entières.

Clas	Équipes	Points	Jouées	Gagnées	Nulles	Perdues	Buts+	Buts-	Diff
1	PAYS AIX UNIV HANDBALL	69	26	20	3	3	800	695	105
2	BILLERE HANDBALL	67	26	18	5	3	694	623	71
3	MULHOUSE HANDBALL SUD ALSACE	63	26	17	3	6	792	746	46
4	DIJON BOURGOGNE H.B.	61	26	16	3	7	707	672	35
5	SMV VERNON SAINT MARCEL	51	26	11	3	12	734	727	7
6	H.B.C. SEMUR-EN-AUXOIS	50	26	12	0	14	751	739	12
7	MAINVILLIERS-CHARTRES HANDBALL	50	26	9	6	11	721	752	-31
8	GRAND NANCY ASPPT HB	49	26	10	3	13	702	733	-31
9	ES BESANCON MASCULINE	49	26	11	1	14	713	739	-26
10	SAINTE US	48	26	10	2	14	724	750	-26
11	UMS PONTAULT-COMBAULT HANDBALL	48	26	9	4	13	689	696	-7
12	ANGERS NOYANT HBC	44	26	9	0	17	705	713	-8
13	MASSY HB ESSONNE	44	26	8	2	16	712	767	-55
14	ES NANTERRE HANDBALL 92	35	26	3	3	20	699	791	-92

On suppose que le classement est établi à partir des points de chaque équipe, ces points étant calculés selon une combinaison linéaire des autres données relatives à cette équipe.

Première partie

Question 1 : Déterminer par moindres carrés les coefficients de la combinaison linéaire.

Tracer les résidus et vérifier visuellement leur caractère aléatoire.

Question 2 : En analysant les coefficients des moindres carrés, pourriez-vous indiquer s'il y a des variables non pertinentes qui peuvent être considérées comme du bruit? Justifier votre réponse et procéder à une nouvelle application des moindres carrés en éliminant les variables identifiées comme non pertinentes.

Question 3 : En comparant les résultats de deux questions précédentes, établissez vos conclusions.

Question 4 : Refaire les moindres carrés de la question 1 en normalisant les valeurs des variables afin que l'intervalle de variation soit le même pour toutes les variables et éviter ainsi des distortions éventuelles dues à l'utilisation de la métrique euclidienne. Pour ce faire on utilisera la formule

$$z_{ij} = (x_{ij} - m_j) \frac{B - A}{M_j - m_j} + A$$

avec x_{ij} la valeur de l'élément (i, j) du tableau d'origine et z_{ij} la valeur correspondante normalisée, $M_j = \max_i x_{ij}$ et $m_j = \min_i x_{ij}$. $B - A$ est l'intervalle de variation des valeurs normalisées. On prendra $A = 1$ et $B = 100$.

Question 5 : En comparant les résultats obtenus avec ceux de la question 1, préciser en le justifiant, s'il est judicieux de procéder à une normalisation des valeurs des variables.

Question 6 : Répéter la question 2 pour les valeurs normalisées.

Question 7 : Que constatez-vous par rapport aux résultats de la question précédente.

Deuxième partie

On voudrait étudier selon les modèles établis à la première partie le classement de l'année précédente. Suite à un problème technique, le disque d'archivage des classements a été endommagé. Après récupération, le classement de l'année précédente (2010-2011) est donné dans la table suivante :

Clas	Équipes	Points	Jouées	Gagnées	Nulles	Perdues	Buts+	Buts-	Diff
1	SELESTAT ALSACE HANDBALL	67	26	20	1	5	764	655	109
2	US CRETEIL HANDBALL	66	26	19	2	5	744	670	74
3	MULHOUSE HANDBALL SUD ALSACE	60	26	17	1	8	754	712	42
4	SMV VERNON SAINT MARCEL	57	26	14	3	9	695	678	17
5	PAYS AIX UNIV CLUB HANDBAL	56	26	13	4	9	708	680	28
6	GRAND NANCY ASPPT HB	55	26	14	2	10	691	662	29
7	SAINTE US	51	26	11	3	12	698	723	-25
8	UMS PONTAULT-COMBAULT HANDBALL	49	26	10	3	13	696	701	-5
9	H.B.C. SEMUR-EN-AUXOIS	47	26	8	3	11	735	768	-33
10	BILLERE HANDBALL	46	26	10	0	16	642	694	-52
11	ES NANTERRE	44	26	7	4	15	689	727	-38
12	ES BESANCON MASCULINE	44	26	7	4	15	636	656	-20
13	VILLEURBANNE HANDBALL ASSOCIATION	43	26	7	3	16	647	714	-67
14	ANGERS NOYANT HBC	41	26	6	3	17	704	763	-59

Question 8 : On cherche à savoir si le nombre de points obtenus par chaque équipe et qui se trouve sur le fichier récupéré est correct. Indiquer une méthode, en la justifiant, qui permet de faire cette évaluation et trouver les équipes pour lesquelles ce nombre de points paraît faux.

Question 9 : Donner les valeurs exactes du nombre de points par équipe, pour les équipes dont les valeurs ne paraissent pas exactes selon la réponse à la question précédente.

Détails pratiques

1. Les données sont stockées dans les fichiers `handball2012.data` et `handball2011.data`. Pour les lire en Scilab, on utilisera les commandes

```
- X = read('handball2012.data', -1, 8);  
- Y = read('handball2011.data', -1, 8);
```

2. Pour calculer la pseudo-inverse en Scilab il faut utiliser la fonction `pinv`.

3. Consigner dans un rapport

- (a) le cahier d'analyse et de programmation de votre programme ;
- (b) les résultats numériques ;
- (c) les graphiques des résidus, et
- (d) vos commentaires.

Joindre le source de votre programme et envoyer le tout dans un fichier compacté (zip ou rar).

Délai de la remise du rapport : **le mardi 5 juin.**

N.B. Pour la remise des rapports il faut à Cergy utiliser exclusivement le bouton correspondant à votre groupe de TD sur le site de `sifoci`. Toute autre méthode d'envoi des rapports ne sera pas prise en compte.