

Année universitaire 2015-2016

Devoir Surveillé
Matière : Data Mining

Enseignant(es)	: W.Barhoumi, S.Zouaoui	Date	: 04/10/2015
Filière / Classe	: 3 Ing SI	Durée	: 1h30
Barème indicatif	:	Documents	: aut. / non aut.
Nbre. de pages	: 2	Calculatrice	: aut. / non aut.

Exercice1 (10 points)

On s'intéresse à expliquer l'investissement des entreprises(Y) en fonction des 2 variables : stock de capital (X1) et profit (X2). A partir des données statistiques relatives à 27 entreprises industrielles pour une année donnée. On se propose en premier lieu d'étudier la liaison entre l'investissement et le stock de capital en mettant en œuvre une régression linéaire simple.

$$Y = \beta_0 + \beta_1 * X1 + \varepsilon$$

1. Donner les hypothèses du modèle. (2 points)
2. Donner la formule des deux paramètres β_0 et β_1 du modèle linéaire simple en utilisant la méthode du moindre carrée ordinaire. A quoi consiste cette méthode ? (2 points)

$$AN : \begin{cases} \hat{\beta}_0 = -261.04 \\ \hat{\beta}_1 = 3.55 \end{cases}$$

3. Interpréter des deux paramètres β_0 et β_1 du modèle linéaire simple.(1 point)
4. Donner la formule du coefficient de détermination R^2 . Expliquer chaque terme de la formule.(2 points)
5. Interpréter la valeur de R^2 AN : $R^2 = 0.364$ (1 point)

Pour mieux améliorer le modèle ainsi que le coefficient de détermination, on introduit une deuxième variable explicative relative au profit (X2). On a obtenu les résultats suivants

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$$\begin{cases} \hat{\beta}_0 = -228.9 \\ \hat{\beta}_1 = 2.85 \\ \hat{\beta}_2 = 0.4 \end{cases}$$

6. Déterminer la forme matricielle du modèle linéaire multiple (1 point)
7. Interpréter les nouveaux paramètres. (1 point)

Exercice 2 (10 points)

Une étude est réalisée sur la base des données recueillies auprès des entreprises clientes d'une banque populaire. Les données se composent de deux sous-ensembles : 23 entreprises jugées comme défaillantes et 23 saines. Nous avons retenu l'analyse discriminante pour cette étude. Le tableau ci-dessous résume les ratios retenus dans cette recherche.

Tableau des Ratios

Aspect	Ratio	Intitulé	Formule
Ratios de structure	R1	Ratio d'autonomie financière	Capitaux propres/ capitaux permanents
	R2	Ratio de trésorerie immédiate	Disponibilité/ dettes à court terme
	R3	Ratio d'équilibre financier	Capitaux permanents/ actif immobilisé net
Ratios d'activité	R4	Part des frais financiers dans la valeur ajoutée	Charges financières//valeur ajoutée
	R5	Ratio crédit fournisseurs en mois	(Dettes fournisseurs/ achats TTC+ autres charges externes TTT)*12
	R6	Ratio crédit clients en mois	(Créances clients/CA TTC)*12
Ratio de rentabilité	R7	Rentabilité financière	Résultat net de l'exercice/ capitaux propres

1. Rappeler l'objectif de l'analyse discriminante et ces différentes méthodes d'analyses. (2 points)
2. Combien d'axes discriminants peut-on obtenir ici ? Justifier la réponse. (1 points)
3. Déterminer le théorème de Huygens en expliquant chaque terme du théorème. (2 points)
4. Dédurre le critère utilisé pour déterminer l'axe discriminant (2 point)

Une analyse prédictive en utilisant les fonctions discriminantes linéaires de Fisher a donné les résultats suivants.

Ratios	Appartenance	
	0 (entreprises défaillantes)	1 (entreprises saines)
Ratio 1	42,085	47,649
Ratio 2	-3,028	-3,125
Ratio 3	0,400	0,588
Ratio 4	-7,797	-3,331
Ratio 5	-2,726	-0,830
Ratio 6	3,805	0,531
Ratio 7	10,495	32,589
Constant	-17,437	-24,884

5. Donner les deux fonctions (entreprises défaillantes ou saines) de classement des entreprises (1 points)
6. Prédire la classe d'une entreprise dont tous ces ratios valent 0.5. (2 points)

Bon travail.

Une analyse prédictive en utilisant les fonctions discriminantes linéaires de Fisher a donné les résultats suivants.

Ratios	Appartenance	
	0 (entreprises défaillantes)	1 (entreprises saines)
Ratio 1	42,085	47,649
Ratio 2	-3,028	-3,125
Ratio 3	0,400	0,588
Ratio 4	-7,797	-3,331
Ratio 5	-2,726	-0,830
Ratio 6	3,805	0,531
Ratio 7	10,495	32,589
Constant	-17,437	-24,884

5. Donner les deux fonctions (entreprises défaillantes ou saines) de classement des entreprises (1 points)

$$E_{\text{défaillante}} = -17.437 + 42.085 * \text{Ratio 1} - 3.028 * \text{Ratio 2} + 0.4 * \text{Ratio 3} - 7.797 * \text{Ratio 4} - 2.726 * \text{Ratio 5} + 3.805 * \text{Ratio 6} + 10.495 * \text{Ratio 7}$$

$$E_{\text{saine}} = -24.884 + 47.649 * \text{Ratio 1} - 3.125 * \text{Ratio 2} + 0.588 * \text{Ratio 3} - 3.331 * \text{Ratio 4} - 0.830 * \text{Ratio 5} + 0.531 * \text{Ratio 6} + 32.589 * \text{Ratio 7}$$

6. Prédire la classe d'une entreprise dont tous ces ratios valent 0.5.(2 points)

$$E_{\text{défaillante}} = 4.18$$

$$E_{\text{saine}} = 12.15$$

Entreprises saine puisque $12.15 > 4.18$

Bon travail.