Outils maths pour la gestion en S2.
Régression linéaire - Séries chronologiques

Enoncés des exercices

Université Paul Sabatier - Toulouse 3
IUT de Toulouse
Département GEA PONSAN

Clement Rau
clement.rau@iut-tlse3.fr

1 Ajustement linéaire

Exercice 1

On a trouvé des coefficients de corrélation élevés (positifs ou négatifs) entre les variables suivantes :

1. " âge du mari et âge de la femme au moment du mariage ",
2. "nombre d'abonnés au téléphone dans le département de la Seine et nombre d'étudiants inscrits à la Faculté de Droit de Paris",
3. " consommation de tabac et cancer des poumons ",
4. " prix de la bière et consommation de la bière".

Dans quels cas, selon vous, une corrélation élevée est-elle le signe d'une relation causale ?

Exercice 2

Sur une série de 20 personnes appartenant à une même tranche d'âge, on a étudié les caractères t, taille en mètres et p, pointure des chaussures (en points de Paris). Les résultats obtenus sont les suivants :

$$\sum_{i=1..20} t_i = 34,28 \ ; \ \sum_{i=1..20} p_i = 848$$

$$\sum_{i=1..20} p_i t_i = 1455,18 \ ; \ \sum_{i=1..20} t_i^2 = 58,8614 \quad \sum_{i=1..20} p_i^2 = 35996.$$

1. Déterminer le coefficient de corrélation linéaire entre t et p. Conclusion ?
2. Déterminer une équation de la droite de régression des moindres carrés de p en t.
3. Estimer p lorsque t=1,75.

Exercice 3

Les pourcentages des personnes âgées (65 ans et plus) de la population française entre 1960 et 1980, sont donnés dans le tableau suivant :

Année	1960	1965	1970	1975	1980
Pourcentage	11,6	12	12,8	13,4	14,1

1. Représenter le nuage de points. (mettre les années en abscisse !)
2. Justifier l'opportunité d'un ajustement linéaire par le calcul du coefficient de corrélation.
3. Déterminer une équation de la droite d'ajustement de y (pourcentage des personnes âgées) en t (année) par la méthode des moindres carrés. Tracer cette droite dans le repère.
4. Si l'évolution se poursuit de la même façon donner une estimation du pourcentage des personnes âgées en l'an 2010.
5. Limites du modèle.
   (a) Selon ce modèle, est-il possible de ne pas avoir de personnes âgées. Si oui, en quelle année ?

(b) Selon ce modèle, est-il possible de n'avoir que des personnes âgées. Si oui, en quelle année ?

Exercice 4

Une entreprise livre des produits conditionnés en colis cartonnés. On a observé l'évolution du nombre de colis livrés par l'entreprise entre 1997 et 2004 :

Année	Rang de l'année x_i	Nombre de colis q_i
1997	1	7332
1998	2	8249
1999	3	8838
2000	4	9280
2001	5	9639
2002	6	9943
2003	7	10187
2004	8	10402

1. Représenter le nuage de points $M(x_i; q_i)$. Effectuer l'ajustement linéaire et tracer cette droite.
2. On pose $y_i = \ln(x_i)$ et $z_i = \ln(q_i)$, où ln désigne le logarithme népérien. Déterminer les valeurs de y_i et z_i. Sur un autre graphique, représenter le nuage de points $M(y_i; z_i)$. Effectuer alors l'ajustement linéaire de z en y, puis tracer cette droire.
3. Calculer les coefficients de corrélation linéaire entre x_i et q_i , puis entre y_i et z_i . Que peut on déduire des valeurs obtenues ?
4. A l'aide des questions précédentes, déterminer une expression de q en fonction de x la plus "plausible". Estimer alors le nombre de colis qui seront livrés par cette entreprise en 1997, en 2050. Commenter ces résultats.

Exercice 5

Une entreprise étudie un coefficient de performance x sur un employé fraîchement recruté, en fonction du temps. La méthode d'obtention de x n'est pas l'objet de l'exercice. Les résultats sont rassemblés dans le tableau ci-dessous :

Temps t_i (en jours)	0	30	60	90	120	150	180	210
x_i	0,180	0,392	0,569	0,690	0,784	0,838	0,879	0,908

1. Représenter le nuage de points de coordonnées $(t_i; x_i)$, ainsi que la droite des moindres carrés de x en t (après calcul de l'équation de celle-ci). Préciser le coefficient de corrélation linéaire entre x et t.
2. On pose $y = \ln(1 - x)$. Calculer les valeurs de y_i arrondies à 10^{-3} près et représenter le nuage de points de coordonnées $(t_i; y_i)$. Réaliser la régression linéaire de y en t et la tracer. Préciser le coefficient de corrélation linéaire entre y et t.
3. En s'appuyant sur les questions précédentes, déterminer la formule la mieux adaptée pour exprimer x en fonction de t ?
4. Déterminer alors la "meilleure valeur" de x lorsque $t = 220,\ 240$, puis 250.
5. Déterminer le temps t au bout duquel $x = 0,99$.

Exercice 6

Une étude statistique porte sur les poids en kg respectifs des pères p_i et ceux de leur fils ainés f_i pour $i = 1...12$. Voilà les résultats numériques que nous avons obtenus :

$$\sum_{i=1...12} p_i = 800, \quad \sum_{i=1...12} p_i^2 = 53418, \quad \sum_{i=1...12} p_i f_i = 54107, \quad \sum_{i=1...12} f_i = 811, \quad \sum_{i=1...12} f_i^2 = 54849.$$

1. Déterminer la droite des moindres carrés du poids des fils en fonction du poids des pères.
2. Déterminer la droite des moindres carrés du poids des pères en fonction du poids des fils
3. En quel point se coupent ces deux droites ? Que vaut le produit des pentes des deux droites ?

Exercice 7

Une grande entreprise décide de réaliser un test sur un échantillon de ses employés, pour savoir si la tension artérielle est liée à l'âge ou à un quelconque stress. Pour cela, sur chacun des employés de l'échantillon, on mesure la tension artérielle Y et l'âge X. Après calculs, on a obtenu les résultats suivants :
$\bar{X} = 35$ et $\bar{Y} = 13.5$, $Var(X) = 4$, $Var(Y) = 64$, et $Cov(X, Y) = 10$.
Calculer le coefficient de corrélation linéaire entre X et Y, puis conclure.

Exercice 8

Une étude théorique de l'évolution d'une population en extinction conduit à penser que le nombre d'individus N de cette population varie avec le temps suivant une loi du type $N(t) = aexp(-kt)$, où a et k sont des constantes strictement positives. On veut déterminer expérimentalement la valeur de la constante k. Pour cela, on observe pendant 8 mois un échantillon composé initialement de 200 individus et on note à la fin de chaque mois le nombre de survivants. Les résultats sont les suivants :

t (mois)	1	2	3	4	5	6	7	8
Survivants après le t ieme mois	180	154	140	120	112	97	84	76

1. Quelles variables doit on considérer pour obtenir une valeur approchée de k, par un ajustement linéaire.
2. Réaliser cet ajustement.
3. Quel sera à votre avis, le nombre de survivants de cet échantillon
   (a) à la fin de l'année en cours ?
   (b) à la fin de l'année suivante ?

Exercice 9

On donne deux séries chronologiques relative à la grande Bretagne.

Années	Récepteurs de radio en service X (en centaine de milliers)	Nombre de maladies mentales déclarées Y (pour 1000 habitants)
1924	13	8
1925	20	8
1926	23	9
1927	25	10
1928	27	11
1929	31	11
1930	36	12
1931	46	16
1932	55	18
1933	63	19
1934	70	20
1935	76	21
1936	81	22
1937	85	23

1. Calculer le coefficient de corrélation des deux variables x et y.
2. Commenter le résultat obtenu. Peut on induire de ce résultat que l'écoute des programmes de radiodiffusion altère la santé mentale des auditeurs ? ou que les anglais achètent des récepteurs radio lorsqu'ils deviennent fous ?

2 Séries chronologiques

Exercice 1

On a relevé les chiffres d'affaires trimestriels, exprimés en K€, d'une entreprise au cours des années 2012 à 2015. Ceux-ci sont présentés dans le tableau suivant.

Année \ Trimestre	1	2	3	4
2012	100	250	200	500
2013	91	245	190	485
2014	79	233	183	478
2015	72	222	162	469

1. Représenter les données.
2. Peut-on utiliser un modèle multiplicatif ou additif ? Pourquoi ?
3. Estimer la tendance générale g_t en utilisant des moyennes mobiles.
4. Estimer les coefficients saisonniers.
5. Prédire les chiffres d'affaires trimestriels de l'entreprise pour l'année 2016.

Exercice 2

Un restaurateur souhaite prévoir, au début de chaque semaine, le nombre de couverts qu'il aura à servir chaque jours de cette semaine. Le tableau suivant donne le nombre de couverts servis, chaque jour où son établissement est ouvert, lors des 4 premières semaines de l'année.

Semaine n° \ Jour	Lundi	Mercredi	Vendredi	Samedi	Dimanche
1	20	40	70	100	60
2	18	35	74	110	65
3	15	38	76	122	66
4	15	40	80	131	70

1. Représenter les données.
2. Peut-on utiliser un modèle multiplicatif ou additif ? Pourquoi ?
3. Estimer la tendance générale g_t en utilisant des moyennes mobiles.
4. Estimer les coefficients saisonniers.
5. Prédire le nombre de couverts servis par le restaurateur chaque jour de la 5$^{\text{e}}$ semaine.