

UTILISATION DE LA REGRESSION LINEAIRE ET DE L'ETUDE DES SERIES CHRONOLOGIQUES APPLIQUEES AUX DONNEES METEOROLOGIQUES

Sébastien BRIDIER, Université Aix-Marseille
Hervé QUENOL, Université Lille 1

Type : TD

Niveau : A partir de la deuxième année de DEUG - Licence

Durée : 3 heures + 3 heures

Thèmes : CLIMATOLOGIE, STATISTIQUES, ANALYSE DE DONNEES

Objectifs :

Ces TD proposent l'utilisation des outils statistiques tels que la régression linéaire et l'étude des séries chronologiques pour étudier les données de températures relevées au cours de la période 1993-1997 dans une station météorologique du réseau du Comité Interprofessionnel du Vin de Champagne (CIVC).

L'application, sur EXCEL, est l'occasion d'utiliser les fonctions d'extraction de données.

Ces deux exercices montrent comment les outils utilisés en statistique permettent d'analyser les variables météorologiques à partir de longues séries de données.

Déroulement :

La première application concerne l'étude comparée des températures sous abri¹ et des températures en indice actinothermique² avec la méthode de la régression linéaire. Elle permet de mettre en évidence les différences de températures entre les deux instruments de mesures suivant l'heure de la journée et les conditions atmosphériques. Durant la journée, le capteur en indice actinothermique (1) reçoit le rayonnement direct du soleil et s'échauffe plus que le capteur sous abri (2). Durant la nuit, le phénomène s'inverse, le capteur (1) se refroidit plus rapidement que le capteur (2). Cette différence est accentuée lorsque le ciel est dégagé et elle est atténuée par temps nuageux (effet de serre).

- Ce TD est l'occasion d'utiliser la fonction « tableau croisé dynamique » pour faire une extraction des données. **Durée 3h.**

La seconde application concerne l'étude d'une série chronologique d'une série de températures sous abri (1993-1997). Elle permet de mettre en évidence le cycle diurne et le cycle mensuel des températures.

- Ce TD est l'occasion d'utiliser la fonction « tableau croisé dynamique » et la fonction « RECHERCHEV » pour l'extraction des données. **Durée 3h. La démarche corrigée est fournie.**

¹ Le capteur mesure la température de l'air à 2m du sol dans un abri ventilé.

² Le capteur mesure la température de l'air à 10 cm du sol (il n'est pas protégé du rayonnement et se comporte comme un bourgeon à l'air libre).

Première application

REGRESSION LINEAIRE : MISE EN RELATION DE DEUX CARACTERES

Relation entre les températures sous abri et les températures en indice actinothermique

Documents : Tableau élémentaire présentant les températures sous abri¹ relevées à 2m du sol et les températures en indice actinothermique² à 10 cm sur une station météorologique en Champagne. Les données sont au pas de temps horaire sur l'année 1994.

A télécharger sur <http://feuillesdegeo.free.fr>

Sources : Données issues du réseau météorologique du CIVC

1- On considère les caractères suivants : « la température sous abri » et « la température en indice actinothermique à 10 cm ». A l'aide d'un graphique, faire apparaître l'éventuelle relation statistique entre ces deux caractères pour l'ensemble de l'année 1994. Si cette relation statistique existe, en préciser le sens, la forme et l'intensité.

2- Si on essaie de déterminer la relation linéaire qui définit la température en indice actinothermique comme une fonction de la température sous abri, quelle en est l'expression ? Tracer la droite de régression qui représente cette relation. Quelle est la signification de cette droite ? Calculer le coefficient de Bravais-Pearson et le coefficient de détermination. Commenter le graphique obtenu et les coefficients.

3- Si cette relation moyenne était strictement vérifiée à chaque instant T de l'année 1994, quelle serait la température en indice actinothermique ? Calculer les résidus (écart entre les valeurs observées et les valeurs estimées par la relation moyenne). Préciser la signification de ces écarts.

Représenter la distribution des résidus, heure par heure pour chaque mois de l'année (utiliser le tableau croisé dynamique), sous la forme d'une surface (option surface dans l'assistant graphique d'Excel). Commenter le graphique.

4- Reprendre les opérations précédentes en ne considérant que la période du printemps 1994 (mars à mai). Comparer notamment les coefficients de détermination des deux périodes (sur l'année et sur le printemps). Calculer la droite de régression et les résidus. Représenter la distribution des résidus heure par heure pour chaque jour du printemps sous la forme d'une surface. Commenter le graphique.

5- Observer les périodes du 20 au 22 mars et du 28 avril au 1er mai. Faire un graphique pour chacune des périodes représentant la température sous abri, la température en indice actinothermique et la différence entre les deux valeurs. Commenter les deux graphiques.

¹ Le capteur mesure la température de l'air à 2m du sol dans un abri ventilé.

² Le capteur mesure la température de l'air à 10 cm du sol (il n'est pas protégé du rayonnement et se comporte comme un bourgeon à l'air libre).

Deuxième étape

SERIES CHRONOLOGIQUES: ETUDE D'UNE SERIE DE TEMPERATURES

Documents : Tableau élémentaire présentant les températures sous abri relevées à 2m du sol sur une station météorologique en Champagne. Les données sont au pas de temps horaire sur l'année 1994.

A télécharger sur <http://feuillesdegeo.free.fr>

Sources : Données issues du réseau météorologique du CIVC

L'exercice traitera successivement les chroniques au pas temps horaire et au pas de temps journalier.

PARTIE A : Etude des températures de l'année 1994 au pas de temps horaire

Construire la courbe des températures au pas de temps horaire, et commenter.

1- La recherche de l'autocorrélation temporelle : le test de Von Neumann

Si la variable est stationnaire, la moyenne des carrés des écarts entre valeurs successives de Y_t est voisine du double de la variance de Y_t

$$\text{Formule } w = \frac{\sum_{i=1}^{n-1} (y_{i+1} - y_i)^2}{2 \sum_{i=1}^n (y_i - \bar{y})^2}$$

Si w s'écarte de 1, la chronique n'est pas aléatoire.

2- Les composantes d'une chronique organisée

Les valeurs d'une chronique peuvent s'organiser selon le temps de trois façons : tendance, cycle et persistance.

$$\begin{aligned} Y_{ti} &= Y'_{ti} + \text{moyenne mensuelle des résidus } (Y'_{ti} - Y_{ti}) + \text{résidus} \\ &= \text{tendance} + \text{cycle} + \text{bruit} \end{aligned}$$

On peut rendre compte de la tendance par une droite de régression.

Le cycle est traduit par la moyenne horaire des résidus.

Le bruit correspond aux résidus.

2-1 - Trouver l'équation de la droite de régression (correspondant à la droite de tendance).

2-2 - Calculer les résidus. Construire le graphe des résidus (écarts à la tendance).

Calculer les moyennes horaires des résidus et construire le graphe.

2-3 - Calculer les résidus finaux (bruit) et construire le graphe correspondant. Commenter.

Partie B : Etude des températures pour la période 1993-1997 au pas de temps mensuel

Les données de températures mensuelles ont été obtenues à partir des données horaires en utilisant le tableau croisé dynamique. Elles correspondent à la moyenne mensuelle des données horaires.

1- Suivre la démarche présentée dans les points 1 à 2-3. Comparer les résultats obtenus suivant les deux pas de temps.

2- Construire les corrélogrammes au pas de temps horaire et mensuel et commenter.

Corrigé de l'exercice

SERIES CHRONOLOGIQUES: ETUDE D'UNE SERIE DE TEMPERATURES

Documents : Tableau élémentaire présentant les températures sous abri relevées à 2m du sol sur une station météorologique en Champagne. Les données sont au pas de temps horaire sur l'année 1994.

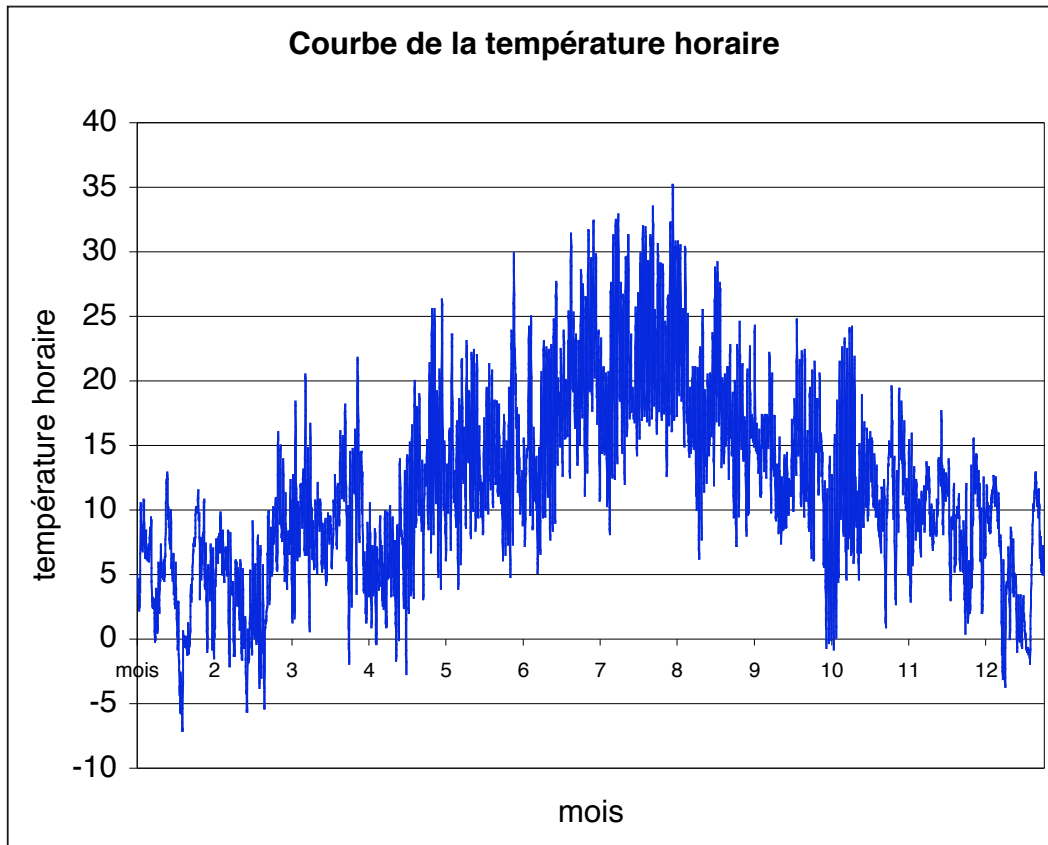
A télécharger sur <http://feuillesdegeo.free.fr>

Sources : Données issues du réseau météorologique du CIVC

L'exercice traitera successivement les chroniques au pas temps horaire et au pas de temps journalier.

PARTIE A : Etude des températures de l'année 1994 au pas de temps horaire

Construire la courbe des températures au pas de temps horaire, et commenter.



La courbe ci-joint est aussi présentée sur la feuille courbe_H.
On constate un cycle saisonnier d'élévation et de diminution des températures.

1- La recherche de l'autocorrélation temporelle : le test de Von Neumann

Si la variable est stationnaire, la moyenne des carrés des écarts entre valeurs successives de Y_t est voisine du double de la variance de Y_t

$$\text{Formule } w = \frac{\sum_{i=1}^{n-1} (y_{i+1} - y_i)^2}{2 \sum_{i=1}^n (y_i - \bar{y})^2}$$

Si w s'écarte de 1, la chronique n'est pas aléatoire.

A partir de la feuille Données_H (données horaires), les calculs doivent être effectués pas à pas.

Calculer la moyenne des températures (02)

Colonne H, calcul de $(F3-F2)^2$ qui correspond à $(Y_{i+1} - Y_i)^2$

Colonne I, calcul de $(F2-\$O\$2)^2$ qui correspond à $(Y_i - Y_{\text{moy}})^2$

La somme en colonne de H et I

*Calculer $w = (\text{somme de H} / (2 * \text{somme de I}))$;*

$w = 0.012$ donc la chronique n'est pas aléatoire

2- Les composantes d'une chronique organisée

Les valeurs d'une chronique peuvent s'organiser selon le temps de trois façons : tendance, cycle et persistance.

$$\begin{aligned} Y_{ti} &= Y'_{ti} + \text{moyenne mensuelle des résidus } (Y'_{ti} - Y_{ti}) + \text{résidus} \\ &= \text{tendance} + \text{cycle} + \text{bruit} \end{aligned}$$

On peut rendre compte de la tendance par une droite de régression.

Le cycle est traduit par la moyenne horaire des résidus

Le bruit correspond aux résidus.

2-1 - Trouver l'équation de la droite de régression (correspondant à la droite de tendance).

Pour calculer l'équation de la droite de tendance, on ajoute une droite de tendance sur le graphe de la feuille Courbe_H, en demandant en option l'équation de cette droite.

Dans la feuille Données_H :

*Les valeurs théoriques sont calculées dans la colonne J (= 0,0006*D2 + 9,0159).*

*Les résidus horaires sont calculés dans la colonne K (= F2 - (0,0006*D2 + 9,0159)).*

2-2 - Calculer les résidus. Construire le graphe des résidus (écarts à la tendance).
Calculer les moyennes horaires des résidus et construire le graphe.

La moyenne horaire des résidus est obtenue avec le rapport de tableau croisé dynamique. On choisit une plage de données comprenant les colonnes C à K (chaque colonne doit avoir obligatoirement un titre). Ensuite on choisit « heure » en ligne et « résidus _heure » en données (moyenne des « résidus-heure »). Le tableau obtenu est placé dans une nouvelle feuille.

Pour apporter les résidus moyens horaires à chaque ligne il faut utiliser la fonction RECHERCHEV Cette fonction permet d'attribuer pour chaque heure le résidu moyen horaire calculé sur l'ensemble de l'année :

RECHERCHEV (valeur_cherchée; table_matrice; No_index)

*Il faut sélectionner les valeurs du tableau croisé dynamique pour créer la matrice mat1.
Il s'agit de chercher une valeur (**valeur cherchée**) dans une cellule (ici l'heure en colonne E), puis de chercher la valeur correspondante dans une matrice (**table_matrice**), en spécifiant la colonne contenant les données à importer (**No_index**).*

RECHERCHEV(E2 ; mat1 ; 2)

2-3 - Calculer les résidus finaux (bruit) et construire le graphe correspondant. Commenter.

Une fois que les résidus moyens horaires sont attribués en chaque ligne, il est possible de calculer le bruit (= K2-L2).

Construire le graphique des résidus finaux (bruits) comme dans une feuille nommée Histo_res_H. Les résidus sont négatifs en hiver et positif en été.

Partie B : Etude des températures pour la période 1993-1997 au pas de temps mensuel

Les données de températures mensuelles ont été obtenues à partir des données horaires en utilisant le tableau croisé dynamique. Elles correspondent à la moyenne mensuelle des données horaires.

1- Suivre la démarche présentée dans les points 1 à 2-3. Comparer les résultats obtenus suivant les deux pas de temps.

2- Construire les corrélogrammes au pas de temps horaire et mensuel et commenter.

Pour construire le corrélogramme au pas de temps horaire, on utilise les résidus de la colonne K. La colonne est copiée dans une nouvelle feuille (Correl_H). Le principe du corrélogramme est d'étudier la corrélation entre la série et elle-même avec un décalage temporel croissant.

On calcul le coefficient de corrélation linéaire de Bravais-Pearson pour chaque couple E_i et E_{i+k} ,

k étant le décalage, on peut travailler ici avec $k = 1$ à 50.

Exemple de calcul : = PEARSON (\$A3: \$A8762; C3: C8762).

On peut ensuite construire le graphe des 50 coefficients de corrélation linéaire de Bravais-Pearson et le commenter.

Il fait apparaître un cycle de 24 h. La même opération est réalisée au pas mensuel et fait apparaître un cycle annuel.