

ALLVARTRANSFO

Syntaxe

ALLVARTRANSFO X ;
XSUPPL XS ;
NSIGNIF ALPHA;
BOXCOX LMIN LMAX DL;
JOHNSON DU0;
XTRANS XT1 XT2 XT3 XT4 XT5 XT6 ;
XSTRANS XST1 XST2 XST3 XST4 XST5 XST6;
XCOR COR;
XSCOR CORP MANQP;
OUTPAR PARAM;
NOPRINT.

Sous-commandes

XSUPPL : données supplémentaires à transformer
NSIGNIF : définition du niveau de signification
BOXCOX : définition des valeurs de λ à considérer pour la transformation de BOX et COX
JOHNSON : définition du pas de variation u_0 utilisé dans la transformation de JOHNSON
XTRANS : enregistrement des données initiales transformées
XSTRANS : enregistrement des données supplémentaires transformées
XCOR : enregistrement des corrélations entre données initiales transformées et scores normaux
XSCOR : enregistrement des corrélations entre les données supplémentaires transformées et les scores normaux
OUTPAR : enregistrement des paramètres des transformations
NOPRINT : suppression des impressions.

Résumé

Macro réalisant, de manière systématique, diverses transformations de variables et calculant, pour chacune de celles-ci, la corrélation entre les données transformées et les scores normaux correspondants. Les six séries suivantes sont étudiées.

1. Pas de transformation
2. Transformation logarithmique
3. Transformation de BOX et COX, avec une valeur de lambda égale à -1 , $-0,5$, 0 ou $0,5$, la valeur retenue étant celle qui maximise la vraisemblance
4. Transformation de BOX et COX, avec une valeur de lambda égale à -2 ou 2 , la valeur retenue étant celle qui maximise la vraisemblance
5. Transformation de BOX et COX, avec recherche de la valeur optimale de lambda par le maximum de vraisemblance
6. Transformation de JOHNSON

Pour la transformation numérotée 5, la recherche de lambda se fait en deux temps. La vraisemblance est d'abord calculée pour les valeurs de lambda variant de LMIN à LMAX par pas de DL. Soit L0 la valeur qui, parmi les valeurs testées, maximise la vraisemblance. Une nouvelle recherche est ensuite réalisée en faisant varier lambda de $L0 - DL$ à $L0 + DL$, par pas de $DL/10$. La valeur finalement retenue pour lambda est celle qui maximise la vraisemblance pour ce nouvel ensemble de valeurs lambda testées.

La transformation numérotée 6 correspond à une des trois fonctions particulières de JOHNSON, communément notées S_B , SL et SU :

$$\begin{aligned} S_B & : z = \gamma + \eta \log_e \left(\frac{x - \varepsilon}{\lambda + \varepsilon - x} \right) \\ SL & : z = \gamma + \eta \log_e (x - \varepsilon) \\ SU & : z = \gamma + \eta \sinh^{-1} \left(\frac{x - \varepsilon}{\lambda} \right). \end{aligned}$$

Le choix de l'une des transformations et l'estimation des paramètres dépendent des valeurs estimées de quatre pourcentiles de la série de données à transformer, qu'on note :

$$x(p_1), \quad x(p_2), \quad x(p_3) \quad \text{et} \quad x(p_4),$$

$p_i (i = 1, \dots, 4)$ étant les probabilités associées aux pourcentiles estimés. Ces probabilités p_i correspondent aux quatre valeurs de la fonction de répartition de la distribution normale réduite :

$$p_1 = \Phi(-3u_0), \quad p_2 = \Phi(-u_0), \quad p_3 = \Phi(u_0) \quad \text{et} \quad p_4 = \Phi(3u_0),$$

u_0 étant un nombre fixé.

Le choix de la transformation à retenir et l'estimation des paramètres de la transformation dépendent de la valeur retenue pour u_0 . La macro réalise les calculs pour différentes valeurs de u_0 et calcule chaque fois la corrélation entre les données transformées et les scores normaux correspondants. La valeur de u_0 finalement retenue est celle pour laquelle la corrélation est maximum.

Les données transformées peuvent être enregistrées dans le fichier. D'éventuelles données supplémentaires peuvent également être transformées en utilisant les paramètres des transformations définis à partir de la variable initiale.

Les corrélations entre les données transformées et les scores normaux correspondants peuvent être enregistrées, à la fois pour la variable initiale et pour les données supplémentaires.

Macros utilisées

BOXCOXTRANS : voir BOXCOXTRANS.doc

JOHNSONTRANS : voir JOHNSONTRANS.doc

Arguments d'entrée

X : colonne contenant la variable pour laquelle les transformations sont déterminées

XS : colonne contenant les données supplémentaires qui seront transformées de manière identique à la variable X

ALPHA : niveau de signification du test de RYAN-JOINER, basé sur la corrélation entre une série d'observations et les scores normaux. Seules les valeurs 0.10, 0.05 et 0.01 sont autorisées. Par défaut ALPHA = 0.05

LMIN, LMAX, DL : valeur minimum, valeur maximum et pas de variation de lambda pour la transformation numérotée 5. Par défaut, LMIN = - 3, LMAX = 3 et DL = 0.1

DU0 : pas de variation de u_0 dans la recherche de la transformation de JOHNSON. Par défaut, DU0 = 0.01

Arguments de sortie

XT1 – XT6 : colonnes contenant les valeurs transformées de la variable X pour les transformations numérotées de 1 à 6

XST1 – XST6 : colonnes contenant les valeurs transformées des données supplémentaires (variable XS) pour les transformations numérotées de 1 à 6

COR : colonne contenant les corrélations des données transformées et des scores normaux correspondants pour la variable X.

COR(1) à COR(6) : corrélation pour les transformations 1 à 6

COR(7) : corrélation pour la première transformation de X rencontrée ne conduisant pas au rejet de l'hypothèse de normalité

COR(8) : valeur théorique du coefficient de corrélation en dessous de laquelle l'hypothèse est rejetée

CORP : colonne contenant les corrélations des données transformées et des scores normaux correspondants pour la variable XS.

CORP(1) à CORP(6) : corrélations pour les transformations 1 à 6

CORP(7) : corrélation pour la transformation de XS correspondant à la première transformation de X ayant conduit au non rejet de l'hypothèse de normalité

MANQP : nombre de données manquantes dans les transformations de XS. Ces données manquantes peuvent provenir du fait que le domaine de variation de XS n'est pas identique au domaine de variation de X. Ainsi, si XS contient une valeur nulle alors que X n'en contient pas, la transformation logarithmique de XS conduira à une donnée manquante

MANQP(1) à MANQP(6) : nombre de données manquantes pour les transformations numérotées 1 à 6

MANQP(7) : nombre de données manquantes dans la transformation de XS correspondant à la première transformation de X ayant conduit au non rejet de l'hypothèse de normalité

PARAM : variable contenant les paramètres des transformations

PARAM(1) : valeur de lambda pour la transformation de BOX et COX numérotée 5

PARAM(2) à PARAM(5) : valeurs de eta, gamma, lambda et epsilon pour la transformation de JOHNSON

PARAM(6) : code de la transformation de JOHNSON ($S_B = 1$, $SL = 2$ et $SU = 3$)

PARAM(7) : numéro de la première transformation de X ne conduisant pas au rejet de l'hypothèse de normalité