

PROC REG

REGRESSION LINEAIRE SIMPLE OU MULTIPLE

Avec la procédure PROC REG, la régression linéaire, simple ou multiple, cherche à expliquer la variable dépendante « y » par une ou plusieurs variables explicatives « x ». Sous Sas, la régression linéaire simple (avec une seule variable explicative x) prend la forme

```
PROC REG DATA=nomtab1 optnum;  
  MODEL y=x / opt;  
  VAR y x var5 var6 var7 ...;  
  FREQ y x;  
  OUTPUT OUT=nomtab2;  
  RESTRICT equal;  
  TEST equa2;  
  MTEST equa3;  
  WEIGHT var5;  
  ID var6;  
  BY var7 ...;
```

alors que la régression linéaire multiple avec plusieurs variables explicatives (x est alors remplacé par x1, x2, x3 ...) prend la forme

```
PROC REG DATA=nomtab1 optnum;  
  MODEL y=x1 x2 x3 ... /opt;  
  VAR y x1 x2 x3 ... var5 var6 var7;  
  FREQ y x1 x2 x3 ...;  
  OUTPUT OUT=nomtab2=y x1 x2 x3 ...;  
  RESTRICT equal;  
  TEST equa2;  
  MTEST equa3;  
  WEIGHT var5;  
  ID var6;  
  BY var7 ...;
```

Panorama des options disponibles

Quel que soit le type de régression (simple ou multiple), la commande PROC REG est suivie éventuellement de l'instruction DATA= le nom du tableau d'entrée (Data) Sas à utiliser (ici, *nomtab1*) et d'une ou de plusieurs options *optnum* séparées par un blanc. Lorsque la régression ne porte pas sur la dernière table Sas mémorisée, la commande PROC REG doit être suivie de l'option DATA=*nomtab1* où *nomtab1* est le nom de la table (Data) Sas contenant les données à analyser. Si l'option DATA=*nomtab1* est absente, l'analyse porte alors sur la dernière table Sas mémorisée. Les options *optnum* figurant dans le tableau ci-après, suivant l'éventuelle option DATA=*nomtab1*, permettent de préciser les conditions techniques de la régression.

Option	Utilité
SIMPLE	fournit les statistiques de base relatives à chaque variable
USSCP	produit la matrice des produits croisés
ALL	fabrique les statistiques de base de chaque variable (SIMPLE) ainsi que la matrice des produits croisés (USSCP)
EPSILON	fixe la précision des calculs (par défaut, 1E-8)
NOPRINT	interdit l'affichage des résultats

Panorama des instructions disponibles

De nombreuses instructions, séparées chacune par un point-virgule, peuvent figurer après la commande PROC REG comme l'indique le tableau ci-dessous qui précise l'utilité de chacune de ces instructions.

Instruction	Utilité																						
MODEL	<p>spécifie le modèle de régression (par défaut, avec une constante). Chaque régression peut ici être précisée en ajoutant le séparateur « / » puis, après un blanc, une ou plusieurs des options <i>opt</i> suivantes.</p> <table border="1"> <tbody> <tr> <td>INT</td> <td>ajoute les tests associés à la constante</td> </tr> <tr> <td>NOINT</td> <td>supprime la constante</td> </tr> <tr> <td>NOUNI</td> <td>n'édite pas les statistiques univariées</td> </tr> <tr> <td>SOLUTION</td> <td>Edite la solution du modèle $Y=a.X+b$</td> </tr> <tr> <td>XPX</td> <td>crée le produit matriciel $X'X$</td> </tr> <tr> <td>INVERSE</td> <td>génère le produit matriciel inverse $(X'.X)^{-1}$</td> </tr> <tr> <td>SS1</td> <td>diminue la somme totale des carrés des résidus pour chaque terme du modèle testé</td> </tr> <tr> <td>SS2</td> <td>augmente la somme partielle des carrés des résidus</td> </tr> <tr> <td>COVB</td> <td>édite la matrice de variances-covariances entre les paramètres estimés</td> </tr> <tr> <td>CORRB</td> <td>édite la matrice de corrélation entre les paramètres estimés</td> </tr> <tr> <td>DW</td> <td>calcule le coefficient de Durbin-Watson</td> </tr> </tbody> </table>	INT	ajoute les tests associés à la constante	NOINT	supprime la constante	NOUNI	n'édite pas les statistiques univariées	SOLUTION	Edite la solution du modèle $Y=a.X+b$	XPX	crée le produit matriciel $X'X$	INVERSE	génère le produit matriciel inverse $(X'.X)^{-1}$	SS1	diminue la somme totale des carrés des résidus pour chaque terme du modèle testé	SS2	augmente la somme partielle des carrés des résidus	COVB	édite la matrice de variances-covariances entre les paramètres estimés	CORRB	édite la matrice de corrélation entre les paramètres estimés	DW	calcule le coefficient de Durbin-Watson
INT	ajoute les tests associés à la constante																						
NOINT	supprime la constante																						
NOUNI	n'édite pas les statistiques univariées																						
SOLUTION	Edite la solution du modèle $Y=a.X+b$																						
XPX	crée le produit matriciel $X'X$																						
INVERSE	génère le produit matriciel inverse $(X'.X)^{-1}$																						
SS1	diminue la somme totale des carrés des résidus pour chaque terme du modèle testé																						
SS2	augmente la somme partielle des carrés des résidus																						
COVB	édite la matrice de variances-covariances entre les paramètres estimés																						
CORRB	édite la matrice de corrélation entre les paramètres estimés																						
DW	calcule le coefficient de Durbin-Watson																						
VAR	fixe la liste des variables retenues (y et x dans la régression simple ; $y, x1, x2, x3$ dans la régression multiple). En l'absence de cette instruction, toutes les variables sont retenues.																						
FREQ	estime en pourcentages, simples et cumulés, les variables citées (ici, y et x dans la régression simple ; $y, x1, x2, x3, \dots$ dans la régression multiple)																						
OUTPUT OUT =	indique, après le premier signe d'égalité, le nom du tableau Sas de sortie (ici, <i>nomtab2</i>) contenant les diverses statistiques données par la régression et, après le second signe d'égalité, l'éventuelle constante et les coefficients de la régression attachés aux variables exogènes (x ou $x1, x2, x3 \dots$).																						
RESTRICT	définit les contraintes linéaires que doivent respecter les coefficients a , contraintes séparées par une virgule (par exemple RESTRICT $a+b=1$;))																						
TEST	teste la contrainte linéaire sur les variables (par exemple, TEST $a=0$;))																						
MTEST	teste une contrainte linéaire sur plusieurs variables endogènes																						
WEIGHT	crée la variable (ici, <i>var5</i>) permettant de pondérer le poids des autres variables																						
ID	retient une variable (ici, <i>var6</i>) comme identificateur (à défaut, l'identificateur est le numéro de ligne N attribué à chaque observation)																						
BY	réalise les calculs par classe de variables (ici, <i>var7, \dots</i>).																						