

Un exemple de code SAS pour une analyse de variance à deux facteurs croisés :
une réponse rendement (Rdt) en fonction de deux facteurs :un facteur Engrais
(modalités A,B,C) et un facteur Plantation (4 modalités I, II, III, IV) :

```
/* lecture des données */
```

```
data Tab;  
input Engrais $ Plant $ Rdt;  
datalines;  
A I 15  
A I 14  
A I 17  
A II 18  
A II 17  
A III 12  
A III 13  
A III 12  
A IV 14  
A IV 15  
B I 16  
B I 19  
B II 23  
B II 24  
B II 25  
B III 15  
B III 14  
B IV 12  
B IV 11  
C I 18  
C I 17  
C II 20  
C II 21  
C II 21  
C III 17  
C III 19  
C III 17  
C IV 12  
C IV 13  
;  
run;
```

```
proc print data=Tab;  
run;
```

```
proc glm data=Tab; /* glm pour General Linear Model : cette procédure peut  
travailler sur des variables explicatives qui sont des facteurs  
(notre cas ici) et/ou des régresseurs */
```

```
class Engrais Plant; /* Class permet de déclarer les variables qui seront considérées  
comme des facteurs dans le modèle ; en général ces variables  
ont été lues comme qualitatives mais ce n'est pas obligatoire ; les  
différentes valeurs de ces variables permettent de constituer les  
groupes d'individus utilisés dans l'approche d'ANOVA */
```

```

model Rdt = Engrais Plant Engrais*Plant / solutions ;
      /* Model permet de definir le modele sous la forme
Réponse = liste de facteurs ou de régresseurs ;
* est utilisé pour déclarer l'interaction entre deux
facteurs*/
      /* l'option solutions permet d'obtenir les valeurs
d'estimations des paramètres du modèle considéré*/

```

```
run;
```

```

data Tab2;
set Tab;
New = Engrais II Plant;
run;

```

```

proc print data=Tab2;
run;

```

```
*/ 11 premières lignes du tableau crée
```

```
Obs. Engrais Plant Rdt new
```

```

1 A I 15 A I
2 A I 14 A I
3 A I 17 A I
4 A II 18 A II
5 A II 17 A II
6 A III 12 A III
7 A III 13 A III
8 A III 12 A III
9 A IV 14 A IV
10 A IV 15 A IV
11 B I 16 B I

```

```
*/
```

```

proc glm data=Tab2;
class New;
model Rdt= New;

```

```

/* 12 niveaux ranges avec Engrais en premier
A I A II A III A IV B I B II B III B IV C I C II C III C IV
*/

```

```

/*Important de savoir comment sont « rangés » les modalités de New pour pouvoir
définir correctement les contrastes qui nous intéressent (mettre des 0 et des 1 ou -1 là
où il faut)
*/

```

```
contrast 'Engrais A contre B Plant 1' New 1 0 0 0 -1 0 0 0 0 0 0 ;
```

/ Contrast permet de faire le test d'égalité à 0 du contraste défini par les coefficients donnés dans la liste de valeurs 0 /1 ; la chaîne de caractère entre quotes ' ' est obligatoire et vous pouvez mettre ce que vous voulez ; l'idée est de vous permettre de clairement définir le contraste que vous voulez tester sur la variable précisée, ici New ... */*

```
contrast 'Plant I contre II Engrais B' New 0 0 0 0 -1 1 0 0 0 0 0 ;
```

```
contrast 'Plant I+II contre III Engrais B' New 0 0 0 0 0.5 0.5 -1 0 0 0 0 ;
```

```
means New / adjust=bon;
```

/ means permet de faire toutes les comparaisons deux à deux d'un facteur précise (ici New) via une approche comparaisons multiples que vous précisez : bon pour Bonferroni, Snk pour Student-Newman-Keuls etc ; si vous êtes dans une expérience non-orthogonale (non-équilibrée), utilisez lsmeans à la place de means pour comparer les moyennes calculées non pas sur les données brutes mais après prise en compte de la modélisation */*

/ dans l'exemple ici, attention aux interprétations de means car fait toutes les comparaisons !: donc pour parler des effets relatifs à chaque facteur il faut s'intéresser à ceux pour lesquels une modalité de l'un des facteurs est constante*/*

```
run;
```

Vous pouvez maintenant faire le code SAS pour toutes les ANOVA vues en cours, en particulier celle permettant d'obtenir les résultats de l'exemple de la ritaline et des enfants hyperactifs dont les sorties SAS ont été vues en cours de Modèle Linéaire Données rappelées ci-dessous.

```
data Ritalin;  
input Etat $ Trait $ Mes;  
datalines;  
N P 50  
N P 45  
N P 55  
N P 52  
N R 67  
N R 60  
N R 58  
N R 65  
H P 68  
H P 72  
H P 70  
H P 75  
H R 51  
H R 57  
H R 48  
H R 55  
;  
run;
```

Un exemple de code SAS pour une régression d'une variable réponse Y sur plusieurs régresseurs dont une interaction :

(exemple « jouet » juste pour introduire les commandes SAS !)

```
data dat1;
input x z y;
x2=x**2; /* le carré de x*/
z2=z**2;
xz=x*z;
datalines;
1 5 13
2 6.5 14.6
3 10.2 22.5
4 13.4 30.5
5 15.7 37
;
```

```
proc reg data =dat1;
model y= x x2 z xz;
```

/ modèle de régression avec 4 régresseurs dont un représente l'interaction entre les régresseurs x et z ; notez bien que la variable représentant l'interaction doit être calculée avant la procédure Reg : il n'est pas possible par exemple de mettre x*z dans le modèle pour dire de prendre en compte l'interaction entre les régresseurs x et z */*

```
*model y= x z x2 xz / selection = Rsquare cp best=2;
```

/ modèle de régression avec 4 régresseurs dont un représente l'interaction est un modèle considéré comme le plus complet et on cherche à voir si de sous-modèles ne faisant intervenir qu'un ou deux régresseurs ne seraient pas aussi bons, en utilisant les critères Rcarré et Cp de Mallows ; best =2 précise que l'on veut que SAS nous sorte les deux meilleurs modèles possibles selon ces critères correspondants pour un régresseur puis deux régresseurs puis 3 régresseurs.*/*

```
*model y= x x2 z xz / selection = stepwise ;
```

*/*méthode d'approche automatique selon la méthode stepwise (ou encore backward ou forward) pour rechercher un modèle d'intérêt selon ce que l'on a vu en cours */*

```
*model y= x x2 z xz /p clm cli ;
```

/ pour le modèle précisé, obtention des valeurs prédites et des intervalles de confiance de prédiction pour les prédictions d'une valeurs moyenne (clm) ou d'une valeur particulière (cli).*/*

```
*model y= x X2 z xz / R Influence Partial vif;  
    /* pour le modèle précisé, obtention des quantités présentées  
    en cours pour repérer d'éventuels individus trop influents sur  
    certaines caractéristiques du modèle considéré, pour  
    visualiser les corrélations partielles et pour  
    calculer le Variance Influence Facteur qui caractérise les  
    variables trop linéairement dépendantes dans le modèle (voir  
    cours) */
```

```
plot r.*p. ;  
    /* obtenir directement un graphe (résidus, prédits) selon le modèle considéré pour  
    juger de la non corrélation et de la variabilité constante des résidus */
```

```
output out=Sorti p=predi r=resi ;  
    /*idem approche GLM : création d'un tableau de  
    sortie nommé Sorti dans lequel les valeurs de résidus  
    (res) et des prédits (predi) sont stockées. */
```

```
Run ;
```

Un exemple de code SAS pour une analyse de covariance : un facteur, un régresseur et possible interaction facteur/régresseur

→ pas vraiment de nouveautés question approche SAS car on retrouve la proc GLM avec le (ou les) facteur(s) déclaré(s) dans CLASS et le (ou les) régresseur(s) directement mis dans model ; l'interaction facteur/regresseur se déclare dans l'ordre model avec facteur*regresseur. Toutes les options qui se rattachent à GLM sont exploitables.

Exemple

```
data dat1;  
input X S $;  
datalines;  
1 F  
2 F  
3 H  
4 F  
5 H  
etc  
;
```

```
Run;
```

```
Proc GLM data=dat1;  
Class S ;  
Model Y= X S X*S;  
Run;
```