

SE DÉMÊLER AVEC LA PROCÉDURE MIXED

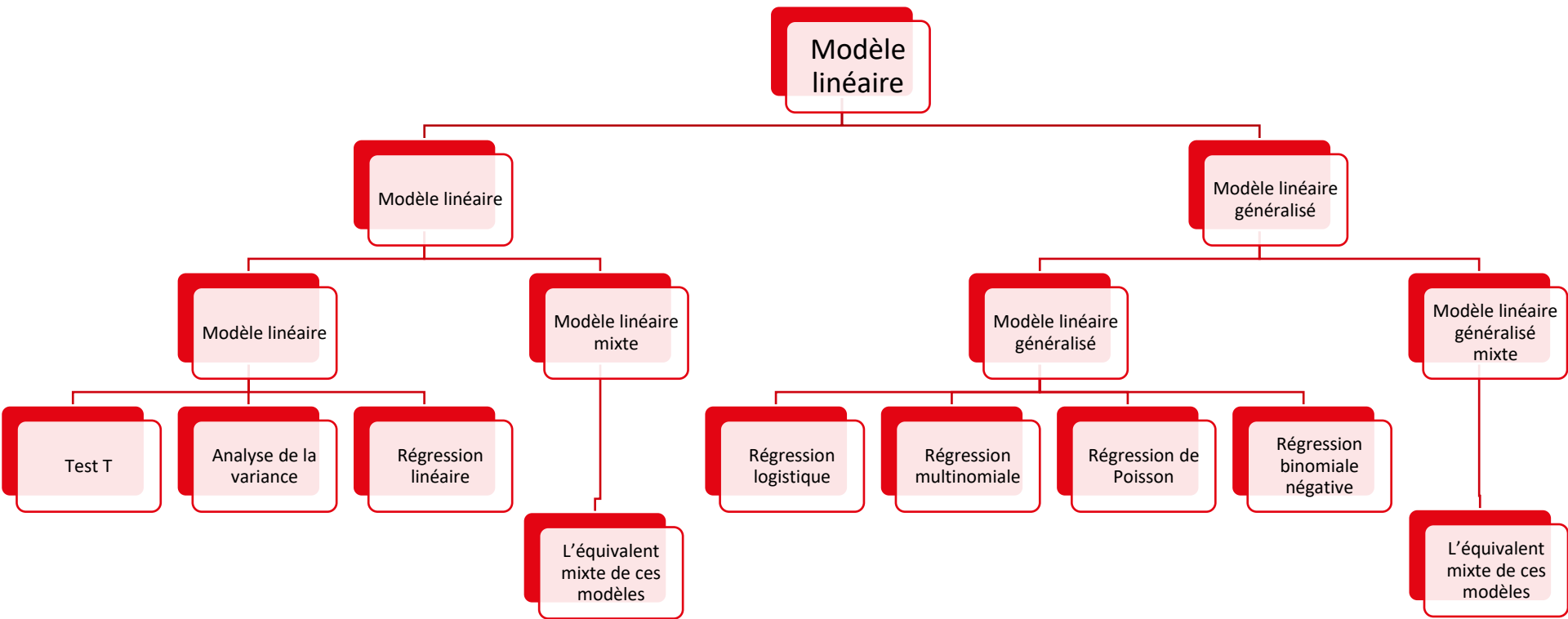
David Emond et Anne-Sophie Julien

Club SAS de Québec
25 octobre 2023

Plan de la présentation

- Modèles linéaires
- Procédures SAS
 - Liste pour modèles linéaires
 - PROC MIXED
 - PROC GLIMMIX
- Exemples de programmes

MODÈLES LINÉAIRES



Modèles linéaires

Permet de modéliser une distribution sous-jacente qui est normale.

Modèles linéaires généralisés

Permet de modéliser une distribution sous-jacente qui n'est pas normale.

- Binomiale
- Ordinale
- Multinomiale
- De Poisson
- Binomiale négative
- Gamma
- Avec inflation de zéros

Modèles linéaires mixtes

Dans les modèles standards, on suppose l'indépendance entre les observations. Les modèles mixtes permettent de tenir compte de la dépendance entre certaines observations (pseudorépétitions, répétitions temporelles ou spatiales, données groupées, etc.) ou de contrôler pour des particularités d'un schéma expérimental (restrictions à la randomisation, création de blocs, etc.).

- Analyse de la variance à mesures répétées
- Analyse de dispositifs expérimentaux: chassé-croisé, plan à blocs aléatoires complets, plan en tiroirs (split-plot), etc.
- Modèles multiniveaux

Modèles linéaires mixtes

Effets fixes

- Les modalités utilisées sont les seules d'intérêt pour lesquelles on désire obtenir une conclusion.
- Ces modalités ne pourraient pas être remplacées par d'autres sans changer l'objectif de recherche.

Effets aléatoires

- Les modalités utilisées sont un échantillon de toutes celles disponibles.
- On ne désire pas comparer les modalités entre elles.
- On veut que les conclusions soient généralisables à la population complète de modalités.
- Les modalités utilisées pourraient donc très bien être remplacées par d'autres.

PROCÉDURES SAS

Procédures SAS

- Modèles linéaires

- GLM
- REG
- TTEST

... mixtes

MIXED

- Modèles linéaires généralisés

- LOGISTIC
- GENMOD

... mixtes

GLIMMIX

Syntaxe générale

```
PROC MIXED DATA = donnees METHOD = REML MAXITER = 100;  
  CLASS id_sujet facteur_fixe facteur_aleatoire facteur_repete;  
  MODEL y = facteur_fixe|facteur_repete x_continu;  
  RANDOM facteur_aleatoire;  
  REPEATED facteur_repete / SUBJECT = id_sujet TYPE = AR(1)  
  LSMEANS facteur_fixe*facteur_repete / CL ADJ = BON DIFFS;  
  SLICE facteur_fixe*facteur_repete/SLICEBY=facteur_fixe;  
  
RUN;
```

Énoncé PROC MIXED

- METHOD =
 - REML : par défaut
 - ML
 - MIVQUE0
 - TYPE3
- MAXITER
 - Permet d'augmenter le nombre d'itérations en cas de problèmes de convergence

Énoncé CLASS

On y spécifie

- Facteurs (fixes et aléatoires), identifiant du sujet
- Modalités de référence pour chaque facteur
 - (REF = FIRST/LAST/"niveau")

Énoncé MODEL

Partie essentielle

- $Y = X_1 X_2 X_3$ -> effets simples
- $Y = X_1 X_2 X_3 X_1 * X_2$ -> effets simples, 1 interaction double spécifique
- $Y = X_1 | X_2 | X_3$ -> effets simples, interactions doubles et triple
- $Y = X_1 | X_2 | X_3 @2$ -> effets simples, interactions doubles

Énoncé MODEL

Options possibles après /

- **NOINT** : retirer l'ordonnée à l'origine
- **SOLUTION** : paramètres du modèle
- Résidus
 - **RESIDUAL**
 - **INFLUENCE**
 - **OUTP**
 - **OUTPM**
 - **VCIRY**
- Degrés de liberté
 - **DDF** = liste de degrés connus
 - **DDFM** = **KR/KR2/SAT**.

Énoncé RANDOM

- On y liste les effets aléatoires
- Option possible après /
 - **SOLUTION** : estimations des paramètres aléatoires

Énoncé REPEATED

- **REPEATED** facteur_repete / **SUBJECT** = id_sujet **TYPE** = AR(1)
 - Mesures répétées
 - Structures de covariance
TYPE = VC, UN, CS, CSH, AR(1), ARH(1), ANTE(1), SP(POW), etc.
- **REPEATED** / **GROUP** = facteur_heterogene;
 - Variance hétérogène pour chaque modalité d'un facteur

Énoncé LSMEANS

LSMEANS facteur_fixe*facteur_repete / CL ADJ = BON DIFFS;

- Moyennes des moindres carrés des facteurs ou combinaisons de facteurs
- CL : Intervalles de confiance
- DIFFS : Comparaisons multiples
- ADJ : Ajustement pour les comparaisons multiples (Bonferroni, Tukey, etc.)

Énoncé SLICE

SLICE facteur1*facteur2 / SLICEBY=facteur2 LINES ADJ = BON;

- Test de différences entre les niveaux du facteur 1 pour chaque modalité du facteur 2
- LINES : Représentation des différences de façon graphique
- ADJ : Ajustement pour les comparaisons multiples

Autres énoncés d'intérêt

- PARMS

- Valeurs de départ des paramètres de covariance

- CONTRAST

- Test d'hypothèses sur des paramètres ou des moyennes
- Groupe 1 VS Groupes 2 et 3 : `CONTRAST "G1 vs G2-G3" groupe 1 -0.5 -0.5;`

- ESTIMATE

- Estimation de nouveaux paramètres ou de nouvelles moyennes,
- Moyenne de 2 groupes : `CONTRAST "G1 et G2" groupe 0.5 0.5 0;`

- LSMESTIMATE

- Test d'hypothèses sur les moyennes
- Similaire à `ESTIMATE`, mais avec une écriture simplifiée lorsque plusieurs termes sont considérés

Autres énoncés d'intérêt

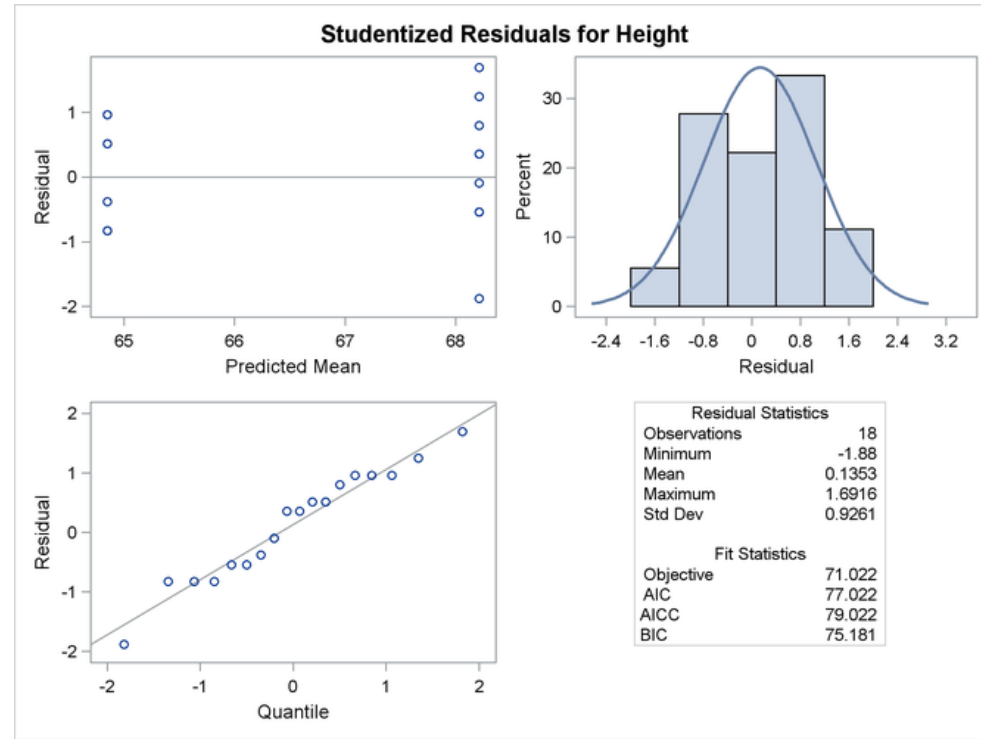
- STORE

- Enregistrement des résultats du modèle dans un objet pour les réutiliser sans exécuter à nouveau le modèle : `STORE OUT = mod1;`
- Utilisation de la procédure PLM pour ajouter des résultats d'intérêt à partir des résultats enregistrés :

```
ODS SELECT lsmeans diffs linesplot;  
PROC PLM RESTORE = mod1;  
    LSMEANS trt1*trt2 / ADJUST = tukey LINES;  
RUN;
```

Graphiques automatiques

- ODS GRAPHICS ON;
- Vérification des postulats et de l'ajustement du modèle



https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.3/statug/statug_mixed_details54.htm

Figure 79.15: Panel of the Studentized (Marginal) Residuals

GLIMMIX

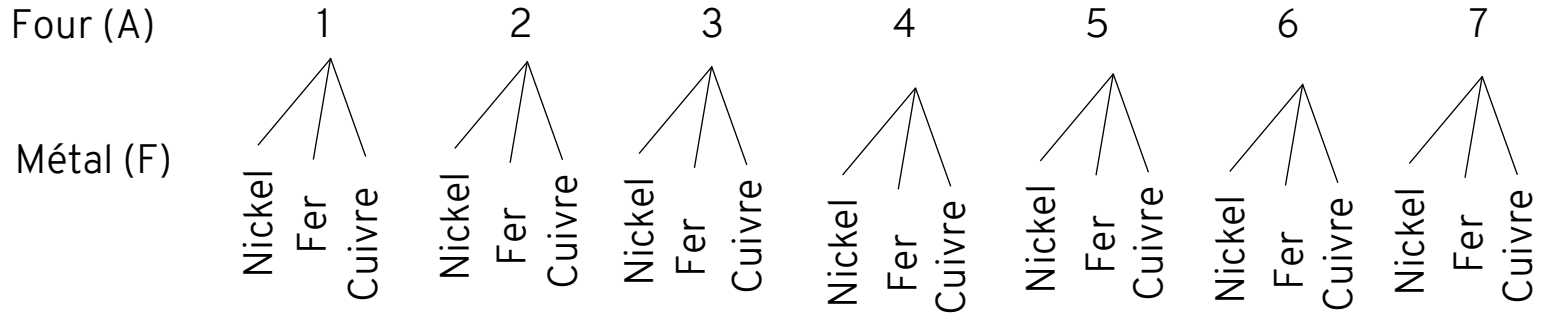
```
PROC GLIMMIX DATA = donnees ;  
    CLASS id_sujet facteur_fixe facteur_aleatoire facteur_repete;  
    MODEL y = facteur_fixe / DIST = binary LINK = logit;  
    RANDOM facteur_aleatoire;  
  
RUN;
```

GLIMMIX : Ses particularités

- **METHOD** : RSPL (par défaut), Laplace, QUAD, etc.
- **DIST** : normal, binomial, poisson, gamma, lognormal, etc.
- **LINK** : identity, logit, log, cumlogit, glogit, etc.
- Absence d'énoncé **REPEATED**
 - **MIXED** : **REPEATED** temps / **SUBJECT** = id **TYPE** = AR(1);
 - **GLIMMIX** : **RANDOM** temps / **SUBJECT** = id **TYPE** = AR(1) **RESIDUAL**;

EXEMPLES DE PROGRAMMES

Exemple 1 : Plan à blocs aléatoires complets



21 morceaux (observations) provenant de 7 fours

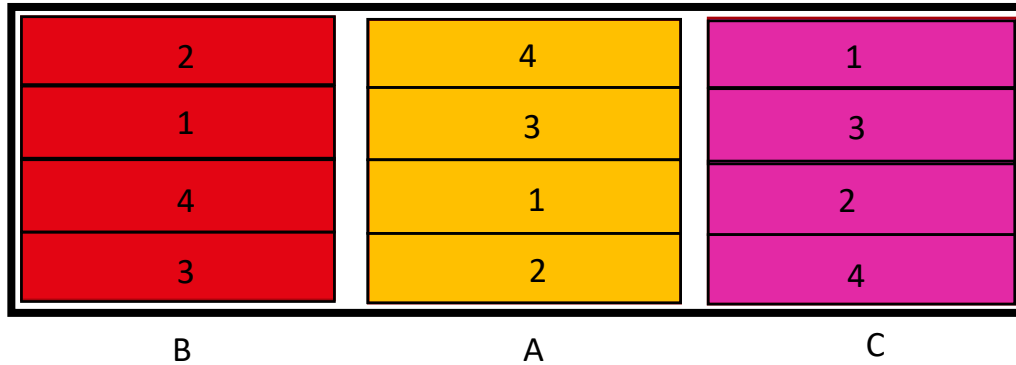
Par opposition à un plan complètement aléatoire où on aurait fondu 21 lingots dans 21 fours distincts

Exemple 1 : Syntaxe

```
PROC MIXED DATA = soudure;  
    CLASS lingot metal;  
    MODEL pression = metal;  
    RANDOM lingot;  
    LSMEANS metal / PDIFF;  
RUN;
```

Exemple 2: Plan en tiroirs (split plot) dans un PBAC

Design pour l'un des trois blocs :



Deux unités expérimentales :

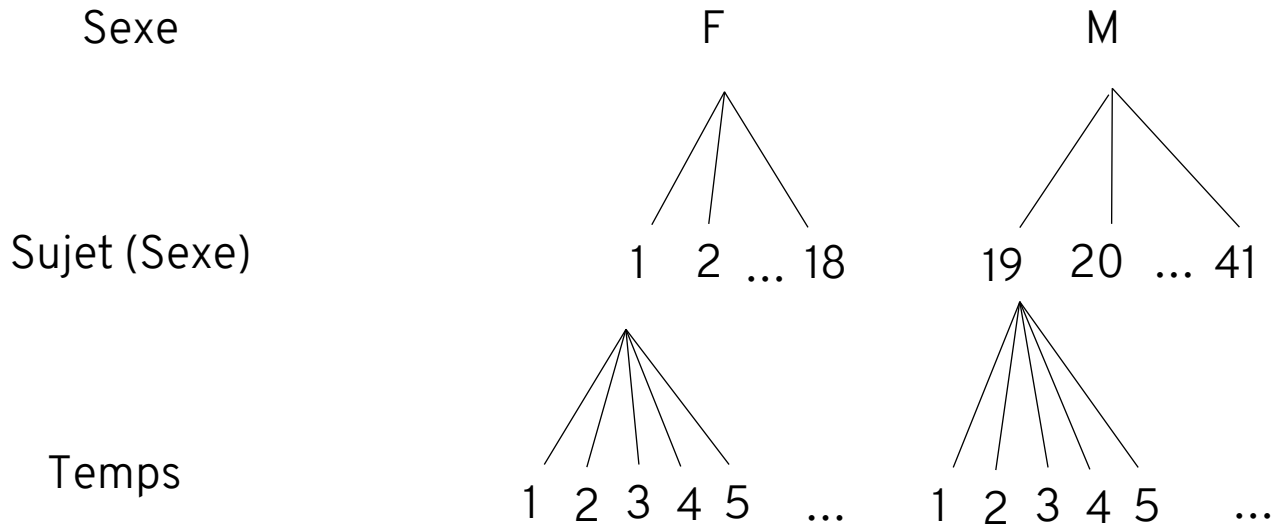
- Parcelle principale (substrat A, B, C)
- Sous-parcelle (champignon 1, 2, 3, 4)

- 3 blocs, 9 parcelles, 36 sous-parcelles

Exemple 2 : Syntaxe

```
PROC MIXED DATA = regeneration;  
  CLASS bloc substrat champignon;  
  MODEL hauteur = substrat|champignon / SOLUTION;  
  RANDOM bloc bloc*substrat / SOLUTION;  
  LSMEANS substrat / PDIFF;  
RUN;
```

Exemple 3 : ANOVA à mesures répétées



Nombre de sujets = 41

Nombre d'observations = $41 * 5 = 205$

Exemple 3 : Syntaxe

```
/*Modèle Mixte : Avec un effet aléatoire pour le sujet*/
```

```
PROC MIXED DATA = donnees;  
  CLASS id temps sexe;  
  MODEL note = sexe | temps / SOLUTION OUTPM = sortie VCIRY;  
  RANDOM id(sexe) / solution;  
  LSMEANS sexe*temps / SLICE=(sexe temps) PDIFF;
```

```
RUN;
```

```
/*Modèle Mixte : Avec la modélisation de la dépendance  
temporelle*/
```

```
PROC MIXED DATA = donnees;  
  CLASS id temps sexe;  
  MODEL note = sexe|temps / SOLUTION OUTPM = sortie VCIRY;  
  REPEATED temps / SUBJECT=id(sexe) TYPE = ar(1);  
  LSMEANS sexe*temps / SLICE=(sexe temps) PDIFF;
```

```
RUN;
```

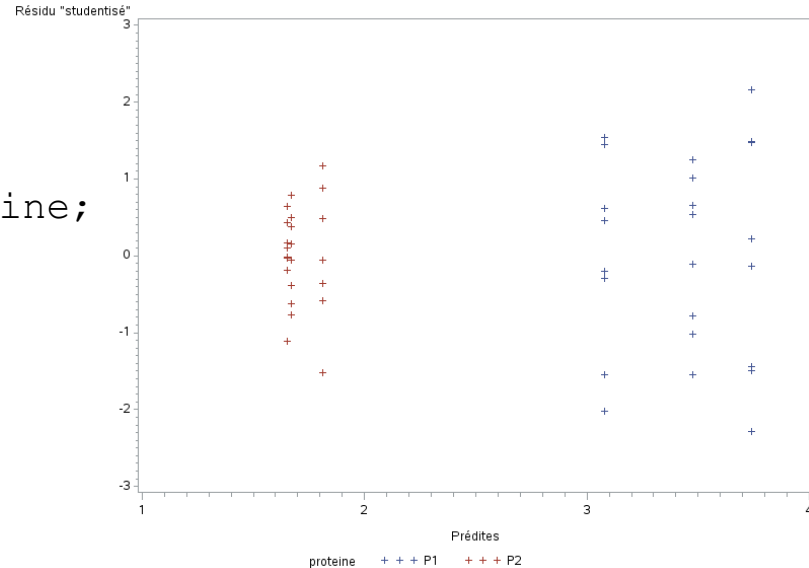
Exemple 4 : Hétéroscédasticité (X catégorique)

```
PROC MIXED DATA = donnees;  
  CLASS proteine cystamine;  
  MODEL concentration = proteine|cystamine  
    /OUTP = sortie RESIDUAL;
```

RUN;

```
PROC GPLOT DATA = sortie;  
  PLOT studentresid*pred=proteine;
```

RUN; QUIT;



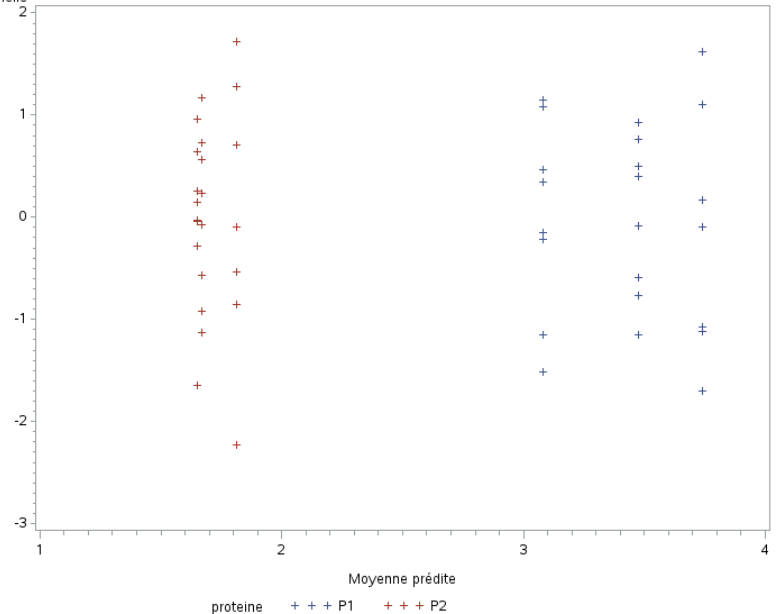
Exemple 4 : Syntaxe

```
PROC MIXED DATA = donnees;  
  CLASS proteine cystamine;  
  MODEL concentration =  proteine|cystamine  
    / OUTPM = sortie VCIRY; Résidu mis à l'échelle  
  REPEATED / GROUP =  proteine;
```

```
RUN;
```

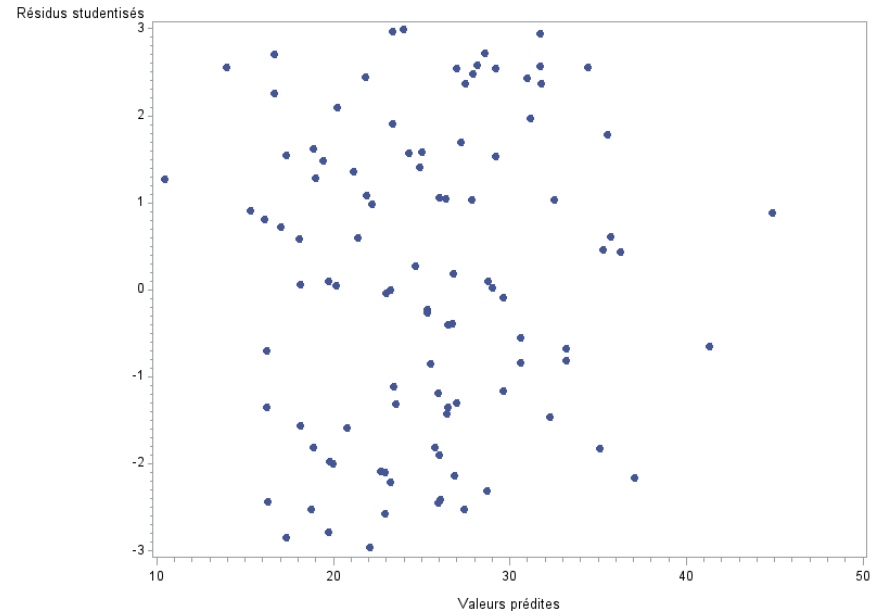
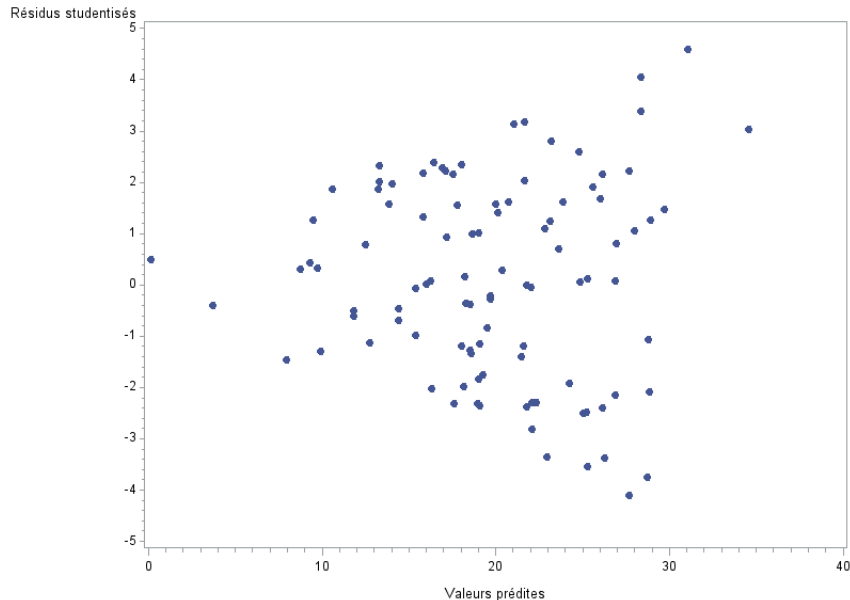
```
PROC GPLOT DATA = sortie;  
  PLOT scaledresid*pred=proteine;
```

```
RUN;QUIT;
```



Exemple 5 : Hétéroscédasticité (X continu)

En présence d'hétéroscédasticité, on peut modéliser la variance comme une fonction de la moyenne, avec la structure « Power of the mean »



Exemple 5 : Hétéroscédasticité (X continu)

- 1 - Ajuster le modèle traditionnel pour extraire les paramètres des effets fixes (/ SOLUTION)
- 2 - Ajuster le modèle où la variance est calculée en fonction de la moyenne (obtenue grâce aux paramètres estimés à l'étape précédente)
- 3 - Les étapes 1 et 2 peuvent être effectuées de façon itérative jusqu'à convergence
- 4 - Vérifier l'amélioration avec les résidus mis à l'échelle et les critères d'ajustement

Exemple 5 : Syntaxe

```
ODS OUTPUT SOLUTIONF=sf;  
PROC MIXED DATA =donnees;  
    MODEL y = x1 x2/ SOLUTION OUTP=sortie RESIDUAL;  
RUN;  
PROC GPLOT DATA = sortie;  
    PLOT studentresid*pred;  
RUN;QUIT;
```

```
PROC MIXED DATA = donnees;  
    MODEL y = x1 x2 / OUTPM=sortie VCIRY;  
    REPEATED / LOCAL=POM(sf);  
RUN;  
PROC GPLOT DATA = sortie;  
    PLOT scaledresid*pred;  
RUN;QUIT;
```

Merci!
