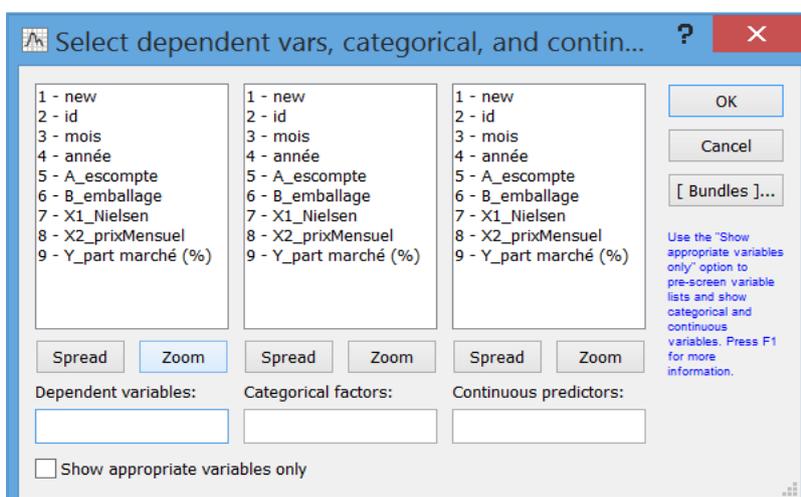
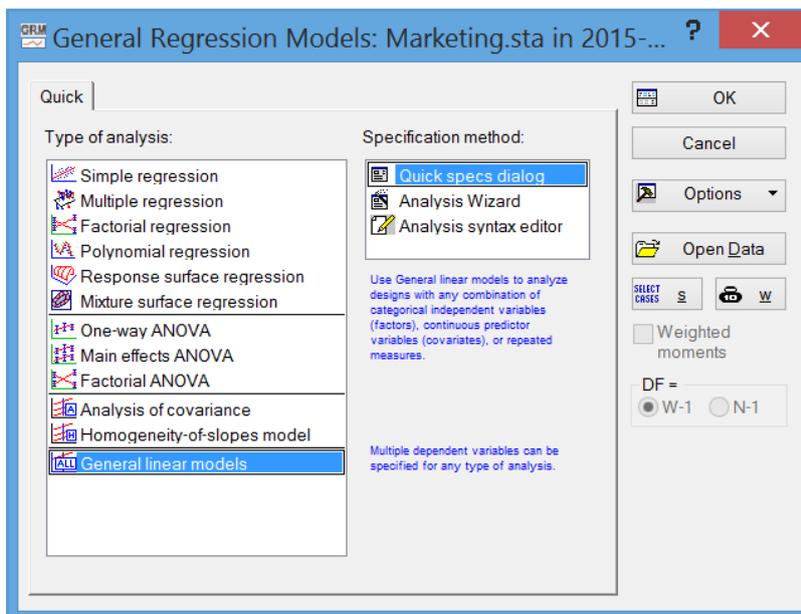
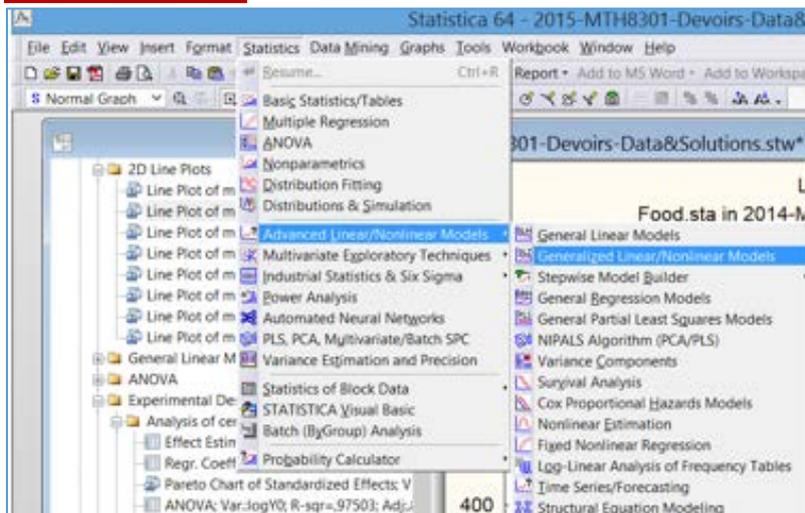


# Modélisation d'une réponse continue Y avec des prédicteurs continus X1, X2,... et des facteurs catégoriques A, B,...

## Utilisation procédure GRM de Statistica

### Procédure GRM



## Codification des variables catégoriques

Les variables catégoriques A, B, C,... sont transformées avec un **codage à effets**.

On remplace une variable catégorique A ayant k modalités par k-1 variables prenant une combinaison des valeurs -1 0 selon le codage expliqué dans l'exemple suivant. Ce codage est fait automatique par Statistica sans que l'utilisateur intervienne.

Supposons que la variable catégorique A prend 4 valeurs selon les modalités a / b / c / d. On choisit une modalité de référence, disons d dans le cas ici.

Le choix de d est arbitraire.

La variable A est remplacée par 3 variables, disons Z1, Z2, Z3 définies de la manière suivante :

$$Z1 = 1 \text{ si } A = a \quad Z1 = 0 \text{ si } A=b \quad Z1 = 0 \text{ si } A = c \quad Z1 = -1 \text{ si } A = d$$

$$Z2 = 0 \text{ si } A = a \quad Z2 = 1 \text{ si } A=b \quad Z2 = 0 \text{ si } A = c \quad Z2 = -1 \text{ si } A = d$$

$$Z3 = 1 \text{ si } A = a \quad Z3 = 0 \text{ si } A=b \quad Z3 = 1 \text{ si } A = c \quad Z3 = -1 \text{ si } A = d$$

### Interprétation

Z1 mesure le changement produit sur Y lorsque la modalité de A change de a à d.

Z2 mesure le changement produit sur Y lorsque la modalité de A change de b à d.

Z3 mesure le changement produit sur Y lorsque la modalité de A change de c à d.

### remarques

- ▶ si la variable catégorique prend que 2 valeurs a et b, alors on a une seule variable Z avec seulement 2 valeurs :  $Z = 1$  si  $A = a$  et  $Z = -1$  si  $A = b$
- ▶ avec plusieurs variables catégoriques, on aura des variables Z pour chacune d'entre elles : ZA1, ZA2, ZA3,...pour la variable A et ZB1, ZB2, ZB3 pour la variable B, ainsi de suite.
- ▶ le nombre de variables Z croit vite si la variable catégorique possède plusieurs dizaines de catégories; il est utile de regrouper certaines catégories afin d'avoir un nombre raisonnable de variables Z
- ▶ les variables Z peuvent intervenir selon 3 types d'effets dans les modèles statistiques :
  - effet linéaire d'ordre 1 pour les facteurs catégoriques : A, B, C...
  - effet d'interaction entre les facteurs catégoriques : AB, AC,.. BC...
  - effet d'interaction entre les facteurs catégoriques et les facteurs continus AX1, BX2,...
  - les variables catégoriques ne peuvent pas avoir d'effet quadratique; ainsi  $Z1^2$  ne fait pas de sens; de même que  $Z1Z2$  etc.

## MODÈLES

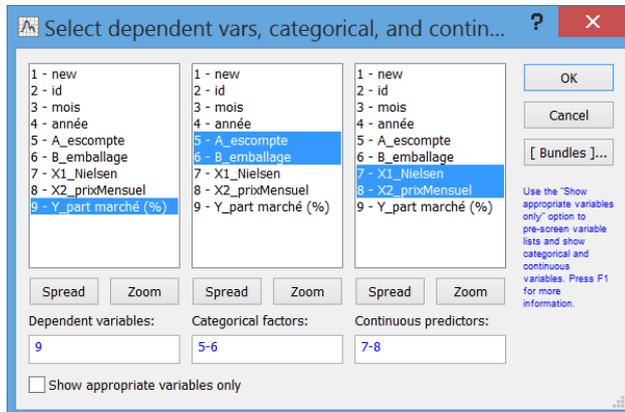
- ▶ M1 : modèle à effet principaux seulement : A, B, C,..., X1, X2, ..  
Il sera représenté par : Z1, Z2, ... X1, X2,... qui sont toutes des variables quantitatives et on peut penser en termes d'un modèle de régression classique
- ▶ M2 : modèle avec des effets principaux et des interactions : comme M1 auquel on ajoute des termes AB, AC,..AX1, AX2,..  
mais les variables catégoriques seront représentées par des produits entre les Z et les X : Z1X1, Z2X2,...; bien entendu il n'y a pas de terme Z1Z2, Z1Z3,...car il n'y a pas d'interaction possible entre 2 catégories
- ▶ M3 : modèles appelé modèle d'analyse de covariance  
comme M2 mais on exclut les interactions entre les facteurs catégoriques et les facteurs continus; on a donc des effets principaux de l'ensemble des facteurs continus et de facteurs catégoriques et des interactions possibles entre les facteurs continus.

remarque : on peut envisager des modèles avec des effets d'interaction triples, etc mais ceux-ci sont rarement nécessaires et compliqués à interpréter.

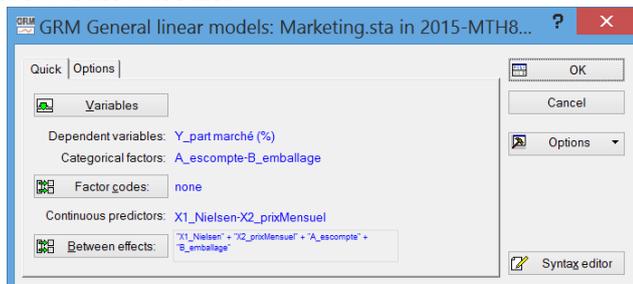
## Spécification du modèle

On peut utiliser les fonctionnalités de GRM pour spécifier le modèle à utiliser.

On identifie la variable dépendante (Y), les facteurs catégoriques (A, B, C,..) et les prédicteurs continus (X1, X2,...). Ici on suppose le cas hypothétique de 2 facteurs catégoriques A et B et de 2 prédicteurs continus X1 et X2



on obtient l'écran suivant

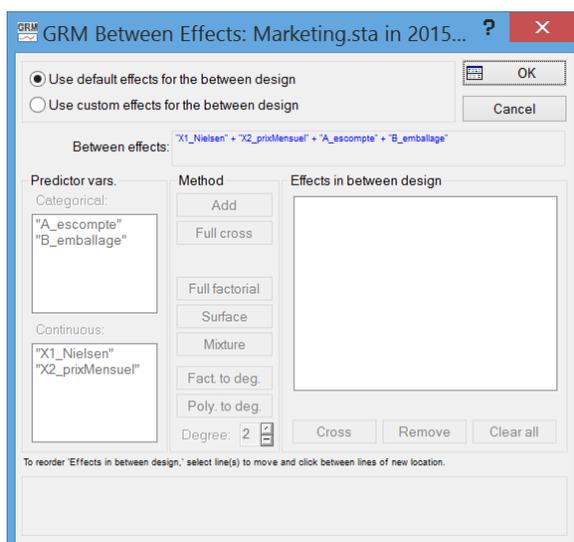


Le modèle est défini dans l'onglet **Between effects**

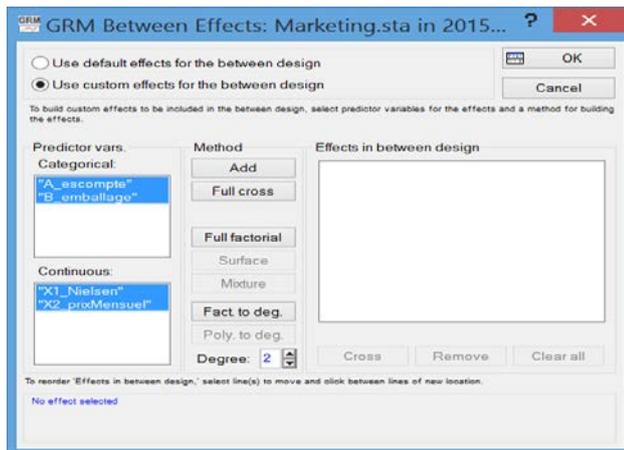
Initialement le modèle par défaut proposé est le modèle M1 : effets principaux seulement.

Ce modèle peut être changé à volonté selon le choix que l'on fait : M2 ou M3

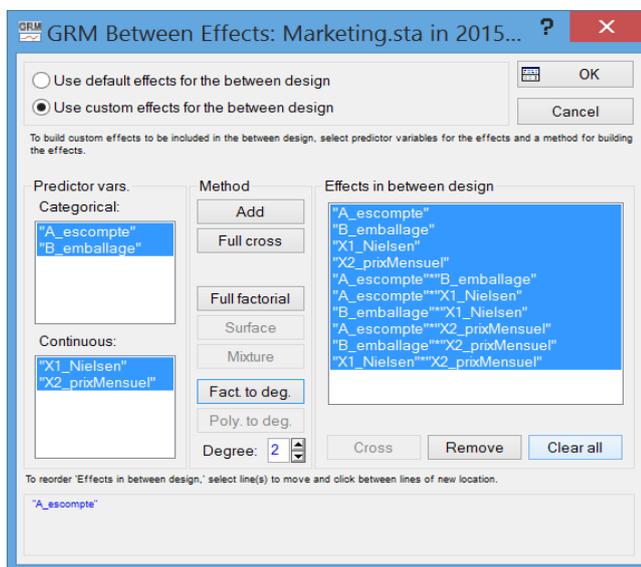
En cliquant sur l'onglet **Between effects** on obtient l'écran suivant :



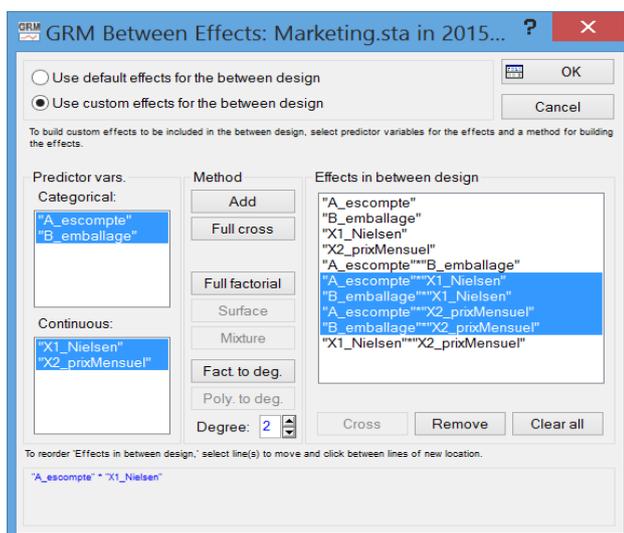
Et on peut changer le modèle en utilisant **Use custom effects**. On obtient alors

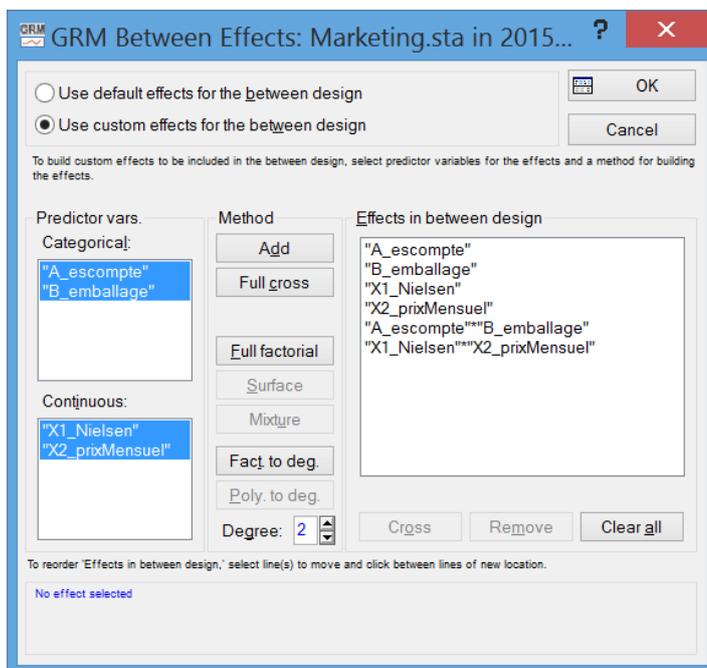


À l'aide des onglets sous **Method** on construit le modèle qui apparaîtra dans la zone **Effects in between design**. Par exemple, si on choisit **Fact. to deg.** et **Degree 2**, on obtient



On peut alors éditer la zone **Effects in between design**. Par exemple, on peut enlever (**Remove**) les interactions entre les facteurs catégoriques et les facteurs continus





Le modèle spécifié ici est un modèle classique dit **d'analyse de covariance**.

Un autre avantage du module GRM est la représentation des effets en ordre d'importance dans le système à l'étude : c'est le graphique de Pareto. Ce diagramme est d'une très grande utilité pour comprendre l'influence des facteurs catégoriques et des prédicteurs continus sur une variable de réponse. Il est le complément idéal pour accompagner un tableau d'analyse de la variance. Voici un exemple.

