

Formation calage sur marges - L'utilisation de la macro Calmar 2

Emmanuel Gros, Antoine Rebecq
emmanuel.gros@insee.fr, antoine.rebecq@insee.fr

INSEE - Division Sondages

29 avril 2015



Sommaire I

- 1 Paramètres de la macro
 - Un peu d'histoire
 - Mise en œuvre de la macro
 - Paramètres relatifs à l'input
 - Paramétrage du calage
 - Paramètres relatifs à l'output
 - Paramètres relatifs aux sorties imprimées
 - Les contrôles

- 2 Un petit exemple commenté de calage sur marges

Chapitre 1

Paramètres de la macro

Partie 1

Un peu d'histoire

Historique Calmar

CALMAR = CALage sur MARges

- CALMAR 1 = macro SAS, écrite en 1992-1993 à l'INSEE.
Objectif : mettre en œuvre les méthodes de calage proposées par Deville et Särndal (JASA, 1992)
- CALMAR 2 = macro SAS, écrite en 2001-2003 à l'INSEE.
Objectif : mettre en œuvre la méthode de traitement de la non-réponse totale par calage généralisé (Deville, 1998)

Historique Calmar

Spécificités de la macro Calmar 2 :

- Calages simultanés à 2 ou 3 niveaux
- Traitement de la non-réponse totale par calage généralisé
- Traitement des colinéarités entre variables auxiliaires
- Une 5ème fonction de calage : sinus hyperbolique

Partie 2

Mise en œuvre de la macro

Paragraphe 1

Paramètres relatifs à l'input

Paramètres relatifs à l'input

DATAMEN = nom de la table SAS contenant les données de l'échantillon

- observations : unités de l'échantillon
- variables : variables de calage, identifiant, pondération initiale

Paramètres relatifs à l'input

**MARMEN = nom de la table SAS contenant l'information
auxiliaire**

- observations : variables de calage
- variables : nom de la variable, nombre de modalités, les marges associées

Paramètres relatifs à l'input

POIDS = variable

variable numérique contenant les pondérations initiales des observations de l'échantillon

PONDQK = variable

variable numérique de pondération des observations de l'échantillon, non liée à la variable spécifiée dans le paramètre POIDS.

IDENT = variable

variable identifiant les observations.

Paramètres relatifs à l'input

PCT= OUI ou NON

si PCT = OUI, les marges des variables catégorielles dans la table MARMEN sont données en pourcentages.

POPMEN = valeur

effectif total de la population (à renseigner si PCT = OUI).

Paramètres relatifs à l'input

Les variables de calage catégorielles peuvent être caractères ou numériques (au sens de SAS).

Les variables de calage numériques doivent être numériques (au sens de SAS).

Observations éliminées : Toute observation ayant une valeur manquante sur l'une des variables du calage ou l'une des variables de pondération, ou prenant une valeur négative ou nulle sur l'une des variables de pondération, est éliminée.

Paragraphe 2

Paramétrage du calage

Paramétrage du calage

M = 1, 2, 3 ou 4

Numéro de la méthode :

- ① méthode linéaire
- ② méthode raking ratio
- ③ méthode logit
- ④ méthode linéaire tronquée.

Paramétrage du calage

LO = valeur

borne inférieure des rapports de poids (si $M = 3$ ou 4).

UP = valeur

borne supérieure des rapports de poids (si $M = 3$ ou 4).

Paramétrage du calage

SEUIL = valeur

seuil pour le test d'arrêt de l'algorithme de Newton (0.0001 par défaut).

MAXITER = n

nombre maximum d'itérations au cours de l'algorithme de Newton (15 par défaut).

Paragraphe 3

Paramètres relatifs à l'output

Paramètres relatifs à l'output

DATAPOI = nom de la table SAS contenant les pondérations finales

- observations : observations de la table en entrée non éliminées
- variables : identifiant, pondération finale.

Paramètres relatifs à l'output

MISAJOUR = OUI ou NON

spécifie le traitement de la table en sortie lorsqu'elle existe déjà :

- si MISAJOUR = OUI, la variable de pondération et l'identifiant sont ajoutés à la table
- si MISAJOUR = NON, la macro crée une nouvelle table, contenant la variable de pondération et l'identifiant, l'ancienne table portant le même nom étant détruite.

Par défaut : MISAJOUR = OUI.

Paramètres relatifs à l'output

POIDSFIN = variable

variable contenant les pondérations finales.

LABELPOI = label

label (éventuel) attribué à la variable de pondération finale.

OBSELI = OUI ou NON

si OBSELI = OUI, la macro crée une table SAS, de nom `__OBSELI`, contenant, pour les observations éliminées, l'identifiant, les variables du calage et les variables de pondération (par défaut : NON).

Paragraphe 4

Paramètres relatifs aux sorties imprimées

Paramètres relatifs aux sorties imprimées

CONT = OUI ou NON

si CONT = OUI (valeur par défaut), des contrôles sont réalisés sur les paramètres de la macro et sur les données figurant dans les tables en entrée.

EDITPOI = OUI ou NON

si EDITPOI = OUI, la macro édite les valeurs des rapports de poids obtenus pour chaque combinaison des valeurs des variables de calage (par défaut : NON).

Paramètres relatifs aux sorties imprimées

STAT = OUI ou NON

si STAT = OUI (valeur par défaut), la macro édite des statistiques (moyenne, écart-type, quantiles, valeurs extrêmes...) et des graphiques¹ relatifs aux distributions des variables " rapport de poids" et " pondération finale".

1. Il s'agit des sorties d'une *proc univariate*

Paramètres relatifs aux sorties imprimées

CONTPOI = OUI ou NON

si **CONTPOI = OUI** (valeur par défaut), la macro édite le contenu de la table **DATAPOI**²

NOTES = OUI ou NON

si **NOTES = NON** (valeur par défaut), les notes produites par SAS durant l'exécution de la macro ne sont pas éditées.

2. Il s'agit des sorties d'une *proc contents*

Partie 3

Les contrôles

Paramètres relatifs aux sorties imprimées

Si **CONT = OUI**, les contrôles suivants sont effectués :

- sur les paramètres de la macro : présence des paramètres obligatoires, cohérence des paramètres, existence des tables SAS ...
- sur le contenu de la table des marges :
 - les variables de calage nommées figurent dans la table-échantillon ;
 - les valeurs des marges sont bien renseignées ;
 - les totaux des marges pour les variables catégorielles sont tous égaux.
- sur les modalités des variables catégorielles : chacune de ces modalités est prise par au moins une observation.
- sur la table en sortie : existence, accès en écriture.

Chapitre 2

Un petit exemple commenté de calage sur marges

Exemple

Exemple du manuel de Calmar2, donné dans les fichiers de documentation (section XIV, p.75).