

BOOTVAR

GUIDE DE L'USAGER (BOOTVAR 3.1 - VERSION SAS)

STATISTIQUE CANADA
AVRIL 2005

TABLE DES MATIÈRES

| | |
|--|-----------|
| 1. Introduction | 3 |
| Changements par rapport à la version précédente:..... | 3 |
| 2. Méthode du bootstrap..... | 4 |
| 3. Description des étapes d'utilisation du BOOTVAR..... | 5 |
| Étape 1: Création du fichier d'analyse | 5 |
| Étape 2: Calcul des variances à l'aide de BOOTVAR.SAS | 6 |
| 4. Conseils d'utilisation et options additionnelles pour le BOOTVAR.SAS | 7 |
| 5. Alternatives au BOOTVAR pour l'utilisation de la méthode bootstrap | 8 |
| Annexe A: Fiches techniques..... | 9 |
| Annexe B: Exemple d'utilisation | 10 |

1. Introduction

Ce guide s'adresse aux utilisateurs de la version SAS du BOOTVAR conçu pour faire l'estimation de variances à l'aide de la méthode du bootstrap. Le BOOTVAR permet d'estimer la variance pour des totaux, des ratios (incluant des proportions), des différences de ratios (ou proportions), des percentiles, des modèles de régression linéaire, des modèles de régression logistique et des tests d'indépendance du Chi-carré. Le BOOTVAR ne génère pas de poids bootstrap, mais utilise plutôt ceux fournis avec les données d'enquêtes.

La section 2 du présent guide explique brièvement la méthode de rééchantillonnage du bootstrap utilisée pour estimer la variance. La section 3 explique en détails le fonctionnement du BOOTVAR ainsi que les étapes préliminaires à son utilisation. Des conseils d'utilisation, de même que certaines autres options possibles avec le BOOTVAR sont présentés dans la section 4. La section 5 discute quant à elle du fait que l'estimation de variances bootstrap peut être accomplie à l'aide d'autres logiciels disponibles commercialement. Finalement, trois annexes accompagnent ce guide d'utilisateur. L'annexe A contient les fiches techniques donnant les détails de chacun des macro-programmes inclus dans le BOOTVAR. L'annexe B contient un exemple complet d'utilisation (programmes et résultats). Finalement, les paramètres spécifiques à chaque enquête de Statistique Canada utilisant la méthode du bootstrap, et qui sont nécessaires à l'utilisation du BOOTVAR sont présentés dans l'annexe C. Cette dernière n'est pas incluse dans le présent document, mais constitue plutôt un document séparé distribué avec le BOOTVAR, ou disponible sur demande via les coordonnées générales laissées dans la documentation de chaque enquête.

Changements par rapport à la version précédente:

Depuis la version 3.0, le BOOTVAR est considéré générique par le fait qu'il tente de supporter toutes les enquêtes de Statistique Canada ayant recours à la méthode du bootstrap pour l'estimation de variances. Cet aspect générique requiert tout simplement que l'utilisateur spécifie quelques paramètres additionnels (par rapport aux versions précédentes) à la section 1 du BOOTVAR. Les données à spécifier pour ces paramètres sont présentées dans l'annexe C.

Depuis sa version 2.0, le BOOTVAR est constitué de deux programmes. Pour la version 3.1, ces deux programmes sont nommés BOOTVARF_V31.SAS et MACROF_V31.SAS, et sont présentés dans la section 3. Afin d'alléger le texte de ce guide d'utilisateur, les deux programmes seront référés via leurs appellations abrégées BOOTVAR.SAS et MACRO.SAS.

Les modifications apportées à la version 3.1 (par rapport à la version 3.0) sont :

- Ajout d'un macro-programme estimant la variance pour des percentiles.
- Ajout d'un macro-programme produisant le test d'indépendance du Chi-carré.
- Ajout d'un paramètre permettant d'utiliser la méthode du bootstrap moyen.
- Ajout des tailles d'échantillons utilisées pour les totaux, ratios, et différences de ratios, dans les résultats présentés en sortie.
- Ajout de la statistique t et de la valeur p (p-value) correspondante pour les résultats de régression linéaire.
- Ajout de la statistique z et de la valeur p (p-value) correspondante pour les résultats de différences de ratios.
- Possibilité de modifier le niveau de signification utilisé par défaut pour le calcul des intervalles de confiance.
- Possibilité d'utiliser des variables contenant des valeurs négatives pour estimer la variance d'un total, d'un ratio et d'une différence de ratios.
- Possibilité de calculer l'effet de plan pour un total ou un ratio.

Il est important de noter que la version SAS 3.1 du BOOTVAR a été testée et fonctionne avec les versions 6.12 à 8.2 de SAS. Son utilisation avec des versions précédentes ou ultérieures de SAS ne garantit pas de résultats exacts.

2. Méthode du bootstrap

Plusieurs enquêtes de Statistique Canada utilisent des plans de sondage complexes afin de sélectionner leurs échantillons. Puisque le calcul de la variance pour de tels plans ne peut être fait à l'aide de formules simples, on doit avoir recours à des méthodes approximatives. Les méthodes par rééchantillonnage figurent parmi celles-ci, dont plus particulièrement la méthode du bootstrap. Celle-ci possède plusieurs propriétés intéressantes et représente la méthode recommandée par plusieurs enquêtes de Statistique Canada.

En bref, la méthode du bootstrap consiste à tirer plusieurs sous-échantillons à partir de l'échantillon complet. Ces sous-échantillons sont le résultat d'un échantillonnage aléatoire simple (ÉAS) avec remise de $n-1$ grappes, parmi les n grappes présentes à l'intérieur de chaque strate. Le nombre de sous-échantillons, noté B , varie d'une enquête à l'autre selon les besoins et objectifs de celles-ci. Un facteur de pondération propre à chaque sous-échantillon (aussi appelé répétition) est assigné à chaque enregistrement faisant partie du sous-échantillon bootstrap. On réfère à ce facteur de pondération comme étant un poids bootstrap. Pour obtenir l'estimation de variance d'une estimation ponctuelle (statistique calculée à partir du poids de sondage), il suffit de recalculer cette même estimation ponctuelle B fois en utilisant les B poids bootstrap. La variabilité observée entre les B résultats représente l'estimation de la variance.

Les poids bootstrap sont produits et fournis par l'enquête. Le BOOTVAR utilise ces poids bootstrap pour procéder à l'estimation de variances, de même qu'à d'autres mesures de variabilité telles que l'erreur-type, l'intervalle de confiance, et le coefficient de variation. Ces mesures devraient être utilisées pour déterminer si une estimation ponctuelle doit ou non être publiée (vérifier les lignes directrices de l'enquête à cet effet), ou encore pour procéder au calcul de tests statistiques.

En résumé, voici les principales étapes que le BOOTVAR effectue pour calculer la variance d'une estimation ponctuelle donnée:

- a) L'estimation ponctuelle (total, ratio, etc...) est d'abord calculée en utilisant le poids de sondage inclus dans le fichier de données.
- b) La même estimation ponctuelle est calculée en utilisant cette fois chacun des B poids bootstrap contenus dans le fichier de poids bootstrap. B estimations bootstrap (du total, du ratio, etc...) sont ainsi obtenues.
- c) Finalement, la variance (selon la formule utilisée pour un plan aléatoire simple) de ces B estimations bootstrap est calculée. Cette variance correspond à l'estimation de la variance de l'estimation ponctuelle calculée en a).

La littérature contient un bon nombre d'articles décrivant ou évaluant la méthode du bootstrap dans un contexte d'utilisation avec des données provenant d'enquêtes à plans complexes. Deux d'entre-eux sont laissés ici en référence aux utilisateurs désireux d'approfondir leurs connaissances sur la méthode du bootstrap. La première référence (Rao et Wu, 1988) représente un recueil plutôt technique sur la théorie entourant l'utilisation du bootstrap, alors que la seconde (Yeo, Mantel et Liu, 1999) présente plutôt un exemple d'application du bootstrap dans un contexte d'enquête.

- Rao, J.N.K. and Wu, C.F.J. (1988). *Resampling Inference with Complex Survey Data*. Journal of the American Statistical Association. Vol. 83, No. 401, 231-241.
- Yeo, D., Mantel, H. and Liu, T-P. (1999). *Bootstrap Variance Estimation for the National Population Health Survey*. 1999 Proceedings of the Survey Research Methods Section, American Statistical Association, pp. 778-783.

3. Description des étapes d'utilisation du BOOTVAR

Le BOOTVAR constitue un ensemble de macro-programmes où chacun d'entre eux permet l'estimation de variances pour un type de statistique bien précis. Les fiches techniques contenues dans l'annexe A présentent la variété de macro-programmes disponibles dans cette version du BOOTVAR, et fournis l'information nécessaire à leurs utilisations. Il est primordial de consulter ces fiches techniques afin de prendre connaissance des contraintes et limites de ces macro-programmes.

Le processus se déroule en *deux étapes* et implique l'utilisation de trois programmes SAS. La *première étape* consiste à créer un fichier de données contenant les variables à être utilisées pour faire l'analyse (premier programme). La *deuxième étape* consiste à utiliser BOOTVAR.SAS (et MACRO.SAS) pour estimer les variances.

Étape 1: Création du fichier d'analyse

Il faut d'abord créer le fichier SAS de données qui servira, à l'étape 2, de fichier d'entrée au BOOTVAR. Les tâches suivantes doivent être faites à cette étape:

1. Lecture du fichier de données
2. Création des variables requises pour l'analyse

1- Lecture du fichier de données: Le fichier d'analyse est créé à partir du fichier contenant les données d'enquête. Ce fichier est habituellement lu à l'aide du cliché d'enregistrement fourni avec les données. L'annexe C présente le nom du fichier de données associé à chaque enquête supportant le BOOTVAR.

2- Création des variables requises pour l'analyse: Des variables dérivées à partir des variables d'intérêt devraient être créées à cette étape. Il peut être nécessaire de créer des variables dichotomiques (1 ou 0) pour identifier les enregistrements ayant les différentes caractéristiques à l'étude. On doit donc créer une variable dichotomique qui prendra une valeur de 1 pour les enregistrements ayant la caractéristique d'intérêt et une valeur de 0 sinon. Par exemple, pour des estimations de totaux, de ratios et de différences entre des ratios, ces variables dichotomiques serviront à identifier les enregistrements ayant la caractéristique d'intérêt afin de sommer leurs poids pour obtenir, lors de l'étape 2, les totaux ou ratios. Voir l'exemple à l'annexe B pour plus de détails.

Le fichier d'analyse doit donc contenir:

- Les variables nécessaires pour l'analyse (variables dérivées, incluant les variables dichotomiques, et autres variables qui ne nécessitent pas de modification). Pour réduire le temps d'exécution du programme, NE PAS conserver les variables inutiles.
- Les identificateurs uniques des enregistrements présents dans l'échantillon. Voir l'annexe C pour connaître le nom de la(des) variable(s) d'identification unique.
- Si nécessaire, la(les) variable(s) de ventilation, identifiant les groupes pour lesquels une analyse séparée est désirée (ex.: province, sexe, etc...).
- Si les analyses ne sont faites que pour un certain sous-groupe de l'échantillon total (une province ou un groupe d'âge par exemple), ne conserver que les enregistrements faisant partie de ce sous-groupe afin de réduire le temps d'exécution du programme.

REMARQUE:

- Il est suggéré de produire les estimations ponctuelles lors de cette première étape, et ce à l'aide du poids sondage contenu dans le fichier de données de l'enquête. Puisque le BOOTVAR produit également l'estimation ponctuelle, il sera alors possible pour l'utilisateur de valider son travail en s'assurant que l'estimation ponctuelle produite par le BOOTVAR correspond bien à celle calculée lors cette première étape. Une divergence entre les deux résultats indiquerait alors que les paramètres spécifiés dans le BOOTVAR ne réplique pas le concept mesuré à la première étape. Noter que pour produire les estimations ponctuelles lors de la première étape, l'utilisateur doit s'assurer de conserver la variable de poids lors de la création du fichier d'analyse.

L'utilisateur doit créer son propre programme pour préparer le fichier d'analyse SAS contenant les variables nécessaires pour l'analyse. Un exemple de programme permettant de créer ce fichier est inclus à l'annexe B (programme ETAPE1.SAS).

Étape 2: Calcul des variances à l'aide de BOOTVAR.SAS

Une fois le nouveau fichier d'analyse créé à l'étape 1, l'étape suivante consiste à exécuter le programme BOOTVAR.SAS. Avant son exécution, il faut spécifier les paramètres nécessaires et les analyses voulues. Il est à noter que ce programme fait appel au programme MACRO.SAS. MACRO.SAS contient les différents macro-codes calculant les estimations de variances. *Pour une utilisation standard du programme d'estimation de la variance, aucune modification par l'utilisateur n'est requise au programme MACRO.SAS.* Le BOOTVAR étant distribué sous forme de codes accessibles, il est toutefois possible pour l'utilisateur plus expérimenté en programmation SAS de modifier les macro-codes afin de satisfaire ses besoins non comblés par le BOOTVAR. Dans le cas échéant, il est recommandé de renommer le programme MACRO.SAS afin d'éviter toute confusion avec la version originale du programme.

Un exemple d'utilisation du programme BOOTVAR.SAS est inclus à l'annexe B. Les parties à éditer par l'utilisateur sont en ***caractère gras***. Le reste du programme ne doit pas être modifié. Ce programme est divisé en ***deux parties***. La ***première partie*** permet à l'utilisateur de définir les différents paramètres, et la ***deuxième partie*** permet de spécifier les analyses voulues.

Partie 1 du BOOTVAR.SAS:

Dans cette partie, l'utilisateur doit spécifier:

- Le nom du répertoire où est situé le fichier d'analyse créé à l'étape 1 et où seront sauvegardés les résultats produit par le BOOTVAR
- Le nom du fichier d'analyse créé à l'étape 1
- Le nom du répertoire et du fichier contenant le fichier de poids bootstrap
- Le nom de la(des) variable(s) de ventilation (i.e.: si l'analyse est faite séparément pour des sous-groupes spécifiques (ex.: province, sexe))
- Les paramètres spécifiques à l'enquête utilisé, c'est-à-dire: le nom de la(des) variable(s) d'identification unique des enregistrements contenus dans l'échantillon, le nom de la variable de poids de sondage, le préfixe utilisé pour nommer les variables de poids bootstrap, le paramètre du bootstrap moyen, puis le nombre de poids bootstrap à utiliser. L'annexe C contient l'information nécessaire pour initialiser correctement les paramètres, selon l'enquête utilisée.
- Le nom du répertoire où se trouve le programme MACRO.SAS

Partie 2 du BOOTVAR.SAS:

Cette partie permet de spécifier les analyses voulues et d'obtenir des estimations de variance pour:

- les totaux
- les ratios (proportions ou moyennes)
- les différences entre ratios
- les percentiles
- les modèles de régression (linéaire ou logistique)
- les tests d'indépendance du Chi-carré

Se référer à l'annexe A pour une description détaillée de ces différents types d'analyse et pour une description des différents résultats produits par le macro-programme correspondant.

4. Conseils d'utilisation et options additionnelles pour le BOOTVAR.SAS

- Modification du BOOTVAR pour effectuer des tests : L'exécution de certains macro-programmes peut prendre un certain temps (surtout pour les régressions et les tests du Chi-carré). Il est suggéré de tester son programme avec un petit nombre de poids bootstrap (par exemple 10) avant de lancer l'exécution finale. **Il est impératif d'utiliser tous les poids bootstrap fournis par l'enquête pour obtenir les estimations de variances finales.** Le nombre de poids bootstrap utilisé peut être modifié dans la première partie du programme BOOTVAR.SAS, en changeant la valeur donnée à la macro-variable *B*.
- Modification du niveau de signification pour les intervalles de confiance : Le niveau de signification utilisé par défaut dans les sorties produites est de 5 % ($\alpha=0.05$). Pour modifier ce niveau pour le calcul des intervalles de confiance, il suffit de changer la valeur par défaut du α en ajoutant l'énoncé `%let alpha=niveau_voulu;` au tout début de la section 2 du programme BOOTVAR.SAS (avant même le premier appel au macro-programme `%total`). Par exemple, pour produire des intervalles de confiance de 90 %, l'énoncé `%let alpha=0.10;` serait spécifié.
- Calcul de l'effet de plan pour un total ou un ratio : Il est possible d'ajouter le calcul de l'effet de plan lors du calcul d'estimation de variances pour des totaux et des ratios. Ceci est rendu possible en ajoutant l'énoncé `%let deff=1;` au tout début de la section 2 du programme BOOTVAR.SAS (avant même le premier appel au macro-programme `%total`). Noter que l'utilisation de cette option allongera légèrement le temps d'exécution.

5. Alternatives au BOOTVAR pour l'utilisation de la méthode bootstrap

Le Bootvar n'est pas le seul outil capable d'estimer les variances et d'effectuer les tests d'hypothèses à l'aide des poids bootstrap. Bien que les logiciels commerciaux tels que SUDAAN et WesVar ne supportent pas explicitement l'utilisation de poids bootstrap, on peut, en misant sur la similarité des techniques bootstrap et BRR (répliques répétées équilibrées), utiliser ces logiciels pour estimer les variances par bootstrap (voir Phillips 2004). En outre, pour calculer la variance par la méthode bootstrap, on peut utiliser de manière récursive tout logiciel qui offre une commande ou une procédure analytique pouvant produire des estimations pondérées des paramètres d'intérêt et qui offre également la souplesse d'un langage de programmation.

En se basant sur ce principe, des programmes semblables au Bootvar ont été construits pour différents logiciels. Par exemple, la commande Bswreg développé en Stata, peut être utilisée pour calculer la variance selon la méthode bootstrap pour bon nombre de commandes de régressions disponibles dans Stata. Les avantages de ce programme sont présentés dans Piérard et coll. 2004.

- Phillips, O. (2004) Comment utiliser les poids bootstrap avec WesVar et SUDAAN. *Le Bulletin technique et d'information des Centres de données de recherche* 1(2): 6-15.
- Piérard, E., Buckley, N., Chowhan, J. Pour une utilisation plus conviviale de la méthode bootstrap: fichier ADO dans Stata. *Le Bulletin technique et d'information des Centres de données de recherche* 1(1): 20-36.

Annexe A: Fiches techniques

Cette section est encore en développement. Veuillez envoyer un courriel à owen.phillips@statcan.ca indiquant que vous désiriez la documentation mise à jour lorsqu'elle sera prête.

Annexe B: Exemple d'utilisation

Voici un exemple complet d'utilisation du programme BOOTVAR.SAS. Dans un premier temps, le fichier d'analyse est créé (étape 1). Ensuite, BOOTVAR.SAS est adapté selon les besoins de l'analyse. Les résultats obtenus sont présentés après les programmes.

Mise en situation:

Cet exemple utilise le fichier transversal de la composante générale du cycle 3 (1998) de l'Enquête nationale sur la santé de la population (ENSP). Les analyses sont effectuées séparément pour chaque province (variable de ventilation) et seulement quatre provinces seront conservées. On cherche à:

- 1- Calculer le nombre de diabétiques, par sexe
- 2- Calculer la proportion de diabétiques chez les hommes et chez les femmes
- 3- Comparer le taux de diabète chez les hommes avec celui chez les femmes
- 4- Calculer le 75^e percentile de l'âge
- 5- Étudier la relation entre l'âge, le diabète et le sexe (régression linéaire)
- 6- Étudier la relation entre le diabète, le sexe et l'âge (régression logistique)
- 7- Étudier la relation de dépendance entre le diabète et le sexe (test d'indépendance du Chi-2)

Les paramètres nécessaires au BOOTVAR sont (selon l'annexe C):

| Enquête nationale sur la santé de la population (ENSP) - Volet ménages | | | | | | | |
|--|--------------------|----------------------------|--------------------------------------|--------------------------|-----------------------------------|----------------|---------------------|
| | Fichier de données | Fichier de poids bootstrap | Variable(s) d'identification (ident) | Variable de poids (fwgt) | Préfixe des poids bootstrap (bsw) | # de poids (B) | Bootstrap moyen (R) |
| Transversal - Composante générale : | | | | | | | |
| Cycle 3 | H35 | B5H35 | REALUKEY PERSONID | FWGT | BSW | 500 | 1 |

Étape 1:

```
*****
*                               *
*      ÉTAPE1.SAS                *
*                               *
* Ce programme permet de créer le fichier de données *
* SAS contenant les variables nécessaires pour utiliser *
* le programme BOOTVARF_V31.SAS *
*****;

LIBNAME in1 'c:\bootvar';

*** Création du fichier de données SAS contenant les variables et enregistrements requis
*** pour l'analyse. Ce fichier devrait être le plus petit possible (contenant seulement
*** les variables et enregistrements nécessaires) afin de réduire le temps d'exécution et
*** la mémoire requise, spécialement dans le cas des régressions. ;

data in1.diabete; /* Fichier qui sera utilisé avec BOOTVARF_V31.SAS */
  %let datafid="D:\Data\h35.txt";
  %include "D:\Layout\h35_1.sas";

*** Création des variables dichotomiques ***

*** (des exemples sont présentés ci-dessous, utilisant les variables du cycle 3 de l'Enquête ***
*** nationale sur la santé de la population) ***;

*** Conserver seulement 4 provinces;
    if prc8_cur in (10 24 35 59);

/* diabète */
    if ccc8_1j=1 then diab=1;
    else if ccc8_1j=2 then diab=0;
    else diab=.;

/* sexe */
    if dhc8_sex=1 then hommes=1;
    else if dhc8_sex=2 then hommes=0;
    else hommes=.;
    if dhc8_sex=2 then femmes=1;
    else if dhc8_sex=1 then femmes=0;
    else femmes=0;

/* diabète*sexe */
    hdiab = diab * hommes; /* hommes diabétiques */
    fdiab = diab * femmes; /* femmes diabétiques */

/* age */
    if DHC8_AGE > 200 then DHC8_AGE=.;

keep realukey personid wt58 prc8_cur diab hdiab fdiab hommes femmes DHC8_AGE;

* Il est recommandé de ne conserver que les variables nécessaires *
* afin de diminuer le temps d'exécution du programme BOOTVARF_V31.SAS. *
* IMPORTANT: les variables d'identification et, si nécessaire, la(les) *
* variable(s) de ventilation (ex: provinces, sexe) doivent être *
* conservées. La variable de poids doit aussi être conservée si les *
* estimations ponctuelles sont calculées à cette étape. *;
```

run;

 ** VEUILLEZ INDIQUER LE NOM DES 2 RÉPERTOIRES SUIVANTS (répertoires seulement): **

```
*****
** VEUILLEZ INDIQUER LE NOM DU FICHIER D'ANALYSE (CRÉÉ À L'ÉTAPE 1) (sans extension): **
*****
```

```
*****
** VEUILLEZ INDiquer LE NOM DU FICHIER CONTENANT LES POIDS BOOTSTRAP: **
** NB: N'exécuter qu'une seule des deux séries de commandes suivantes **
** (mettre l'autre en commentaire, ou l'effacer): **
*****
```

```
*****
** VEUILLEZ SPÉCIFIER LA(LES) VARIABLE(S) DE VENTILATION, SI DÉSIRÉ (EX: PROVINCE, SEXE, ETC...): **
** Écrire le nom de la(des) variable(s) de ventilation ci-dessous. **
** **
** - Si l'analyse est faite pour l'ensemble des données incluses dans le fichier créé à l'étape 1, **
** mettre un point (%let classes =. ) **
** - Si plus d'une variable, laisser un espace entre chacune (%let classes = var1 var2) **
** - NE PAS EFFACER OU METTRE EN COMMENTAIRE CETTE COMMANDE. **
*****
```

```
*****
** VEUILLEZ INDIQUER L'INFORMATION SUIVANTE (SPÉCIFIQUE À L'ENQUETE QUE VOUS UTILISEZ): **
** Vous devez spécifier: **
** 1- La(les) variable(s) d'identification unique (séparées par un espace) **
** 2- La variable de poids final, incluse dans le fichier de poids bootstrap **
** 3- Le préfixe des variables de poids bootstrap **
** 4- Le paramètre du bootstrap moyen **
** 5- Le nombre de poids bootstrap à utiliser (note: Pour les tests, il faut **
** que B >= 2. IL EST NÉCESSAIRE D'UTILISER TOUS LES POIDS BOOTSTRAP AU **
** MOMENT D'EFFECTUER LES ANALYSES FINALES.) **
** - Se référer Annexe C pour obtenir ces informations **
*****
```

```
%let ident = realkey personid;
%let fwgt = fwgt;
%let bsw = bsw;
%let R = 1;
%let B = 500 ;
```

```
*****
** VEUILLEZ SPÉCIFIER LE RÉPERTOIRE ET LE NOM DU FICHIER DU PROGRAMME DE MACRO **
** (LE PROGRAMME MACROF_V31.SAS SI L'USAGER N'A PAS FAIT DE MODIFICATIONS) **
*****;
```

```
%include "D:\Bootstrp\pgm\SAS\MACROF_V31.SAS";
```

```

/*****
/****          PARTIE 2          ****
/*****
/****
/**** Cette partie permet à l'utilisateur de spécifier les différentes ****
/**** analyses désirées ****
/**** ****
/**** ****
/*****/
```

```
* POUR OBTENIR L'ESTIMATION DE LA VARIANCE D'UN TOTAL, EXÉCUTER:
-----;
```

```
    %total(hdiab);
    %total(fdiab);
```

```
* POUR OBTENIR L'ESTIMATION DE LA VARIANCE D'UN RATIO, EXÉCUTER:
-----;
```

```
    %ratio(hdiab,hommes);
    %ratio(fdiab,femmes);
```

```
* POUR OBTENIR L'ESTIMATION DE LA VARIANCE D'UNE DIFFÉRENCE DE RATIOS, EXÉCUTER:
-----;
    * ATTENTION: Voir la remarque au début de la partie 2...;
```

```
    %diff_rat(hdiab,hommes,fdiab,femmes);
```

```

    * où:  VAR1   : la variable au numérateur du 1er ratio      *
    *      VAR2   : la variable au dénominateur du 1er ratio   *
    *      VAR3   : la variable au numérateur du 2e ratio      *
    *      VAR4   : la variable au dénominateur du 2e ratio    *;
```

```
* POUR OBTENIR L'ESTIMATION DE LA VARIANCE D'UN PERCENTILE p (p entre 1 et 99), EXÉCUTER:
-----;
```

```
    %prcnt1e(dhc8_age,75);
```

```
* POUR OBTENIR L'ESTIMATION DE LA VARIANCE DES PARAMÈTRES D'UNE RÉGRESSION, EXÉCUTER:
-----;
```

```
    %regress(dhc8_age,diab femmes);
```

```
* POUR OBTENIR L'ESTIMATION DE LA VARIANCE DES PARAMÈTRES D'UNE RÉGRESSION LOGISTIQUE, EXÉCUTER:
-----;
```

```
    %logreg(diab,femmes dhc8_age);
```

```
* POUR EXÉCUTER UN TEST D'INDÉPENDENCE DU CHI-DEUX, EXÉCUTER:
-----;
```

```
    %chi2(diab,femmes);
```

```
%output; /* Affiche les résultats à l'écran. Ne pas modifier... */
```

```
* POUR SAUVEGARDER LES RÉSULTATS DANS UN FICHIER, EXÉCUTER: (enlever les "")
-----;
```

```
    data out.résultats;
    set &result ;
    run;
```

```
/* Fin du programme SAS BOOTVARF_V31.SAS */
```

Résultats:

Les tableaux suivants représentent les sorties produites par BOOTVAR.SAS pour l'exemple présenté. Se référer aux fiches techniques de l'annexe A pour une description plus complète des résultats.

Estimation de la variance à l'aide du bootstrap pour des TOTAUX (500 répliques bootstrap)

| prc8_cur | Variable | Taille d'échantillon | Total | Erreur-type | Coeff. de variation | Borne inf. intervalle de confiance 95% | Borne sup. intervalle de confiance 95% |
|----------|----------|-------------------------|-----------|-------------|------------------------|---|---|
| Tous | hdiab | 392 | 378528.24 | 20925.33 | 5.53 | 337515.35 | 419541.13 |
| Tous | fdiab | 393 | 320320.34 | 18493.92 | 5.77 | 284072.92 | 356567.76 |
| 10 | hdiab | 35 | 7029.13 | 1380.61 | 19.64 | 4323.19 | 9735.07 |
| 10 | fdiab | 64 | 13712.19 | 1604.37 | 11.70 | 10567.68 | 16856.70 |
| 24 | hdiab | 104 | 110452.80 | 10944.25 | 9.91 | 89002.46 | 131903.14 |
| 24 | fdiab | 95 | 94839.47 | 11271.72 | 11.89 | 72747.31 | 116931.63 |
| 35 | hdiab | 190 | 198237.67 | 15854.33 | 8.00 | 167163.75 | 229311.59 |
| 35 | fdiab | 184 | 164201.93 | 13524.96 | 8.24 | 137693.50 | 190710.36 |
| 59 | hdiab | 63 | 62808.64 | 8568.12 | 13.64 | 46015.43 | 79601.85 |
| 59 | fdiab | 50 | 47566.75 | 7277.83 | 15.30 | 33302.46 | 61831.04 |

Estimation de la variance à l'aide du bootstrap pour des RATIOS (500 répliques bootstrap)

| prc8_cur | Numérateur | Dénominateur | Taille numérateur | Ratio | Erreur-type | Coeff. de variation | Borne inf. intervalle de confiance 95% | Borne sup. intervalle de confiance 95% |
|----------|------------|--------------|----------------------|--------|-------------|------------------------|---|---|
| Tous | hdiab | hommes | 392 | 0.0335 | 0.0019 | 5.53 | 0.0299 | 0.0371 |
| Tous | fdiab | femmes | 393 | 0.0277 | 0.0016 | 5.77 | 0.0246 | 0.0309 |
| 10 | hdiab | hommes | 35 | 0.0263 | 0.0052 | 19.64 | 0.0162 | 0.0365 |
| 10 | fdiab | femmes | 64 | 0.0506 | 0.0059 | 11.70 | 0.0390 | 0.0622 |
| 24 | hdiab | hommes | 104 | 0.0312 | 0.0031 | 9.91 | 0.0252 | 0.0373 |
| 24 | fdiab | femmes | 95 | 0.0262 | 0.0031 | 11.89 | 0.0201 | 0.0323 |
| 35 | hdiab | hommes | 190 | 0.0357 | 0.0029 | 8.00 | 0.0301 | 0.0413 |
| 35 | fdiab | femmes | 184 | 0.0288 | 0.0024 | 8.24 | 0.0241 | 0.0334 |
| 59 | hdiab | hommes | 63 | 0.0324 | 0.0044 | 13.64 | 0.0237 | 0.0411 |
| 59 | fdiab | femmes | 50 | 0.0242 | 0.0037 | 15.30 | 0.0170 | 0.0315 |

Estimation de la variance à l'aide du bootstrap pour des DIFFÉRENCES ENTRE RATIOS (500 répliques bootstrap)

| prc8_cur | Num1 | Dén1 | Num2 | Dén2 | Taille Num1 | Taille Num2 | Diff. de ratios | z | Valeur p | E.-t. | C.V. | Borne inf. intervalle de confiance 95% | Borne sup. intervalle de confiance 95% |
|----------|-------|--------|-------|--------|----------------|----------------|--------------------|-------|-------------|--------|-------|---|---|
| Tous | hdiab | hommes | fdiab | femmes | 392 | 393 | 0.0058 | 2.31 | 0.0210 | 0.0025 | 43.34 | 0.0009 | 0.0107 |
| 10 | hdiab | hommes | fdiab | femmes | 35 | 64 | -0.0243 | -2.77 | 0.0055 | 0.0088 | 36.05 | -0.0414 | -0.0071 |
| 24 | hdiab | hommes | fdiab | femmes | 104 | 95 | 0.0050 | 1.16 | 0.2440 | 0.0043 | 85.84 | -0.0034 | 0.0135 |
| 35 | hdiab | hommes | fdiab | femmes | 190 | 184 | 0.0069 | 1.77 | 0.0775 | 0.0039 | 56.64 | -0.0008 | 0.0145 |
| 59 | hdiab | hommes | fdiab | femmes | 63 | 50 | 0.0082 | 1.37 | 0.1716 | 0.0060 | 73.15 | -0.0035 | 0.0198 |

**Estimation de la variance à l'aide du bootstrap pour des PERCENTILES
(500 répliques bootstrap)**

| PRC8_CUR | Variable | Taille d'échantillon | Percentile (1-99) | Valeur du percentile | Erreur-type | Coeff. de variation | Borne inf. intervalle de confiance 95% | Borne sup. intervalle de confiance 95% |
|----------|----------|-------------------------|----------------------|-------------------------|-------------|------------------------|---|---|
| 10 | dhc8_age | 2844 | 75 | 50.00 | 0.46 | 0.91 | 49.10 | 50.90 |
| 24 | dhc8_age | 8221 | 75 | 51.00 | 0.50 | 0.98 | 50.02 | 51.98 |
| 35 | dhc8_age | 13334 | 75 | 51.00 | 0.09 | 0.17 | 50.83 | 51.17 |
| 59 | dhc8_age | 4602 | 75 | 52.00 | 0.51 | 0.98 | 51.00 | 53.00 |

**Estimation de la variance à l'aide du bootstrap pour des RÉGRESSIONS LINÉAIRES
(500 répliques bootstrap)**

----- Modele=1: Variable dépendante = dhc8_age -----

| PRC8_CUR | Variables indépendantes | Beta | Erreur-type | t | Valeur p |
|----------|----------------------------|---------|-------------|--------|----------|
| 10 | Intercept | 39.6002 | 0.1939 | 204.21 | 0.0000 |
| 10 | diab | 17.8948 | 1.9821 | 9.03 | 0.0000 |
| 10 | femmes | 0.5451 | 0.2936 | 1.86 | 0.0634 |
| 24 | Intercept | 40.5369 | 0.1687 | 240.24 | 0.0000 |
| 24 | diab | 18.6715 | 1.0860 | 17.19 | 0.0000 |
| 24 | femmes | 1.9133 | 0.1896 | 10.09 | 0.0000 |
| 35 | Intercept | 40.2896 | 0.1159 | 347.69 | 0.0000 |
| 35 | diab | 19.6350 | 0.9179 | 21.39 | 0.0000 |
| 35 | femmes | 1.7693 | 0.1472 | 12.02 | 0.0000 |
| 59 | Intercept | 40.8565 | 0.2043 | 199.99 | 0.0000 |
| 59 | diab | 19.2997 | 1.5238 | 12.67 | 0.0000 |
| 59 | femmes | 1.5162 | 0.2189 | 6.93 | 0.0000 |

**Estimation de la variance à l'aide du bootstrap pour des RÉGRESSIONS LOGISTIQUES
(500 répliques bootstrap)**

----- Modele=1: Variable dépendante = diab -----

| PRC8_CUR | Variables indépendantes | Beta | Rapport de cotes | Erreur-type | Wald | Valeur p | Borne inf. int. de conf. (rap. de cotes) 95% | Borne sup. int. de conf. (rap. de cotes) 95% |
|----------|----------------------------|---------|---------------------|-------------|--------|----------|--|--|
| 10 | Intercept | -5.9903 | 0.00 | 0.4321 | 192.16 | 0.0000 | 0.0011 | 0.0058 |
| 10 | femmes | 0.6332 | 1.88 | 0.2830 | 5.01 | 0.0252 | 1.0818 | 3.2802 |
| 10 | DHC8_AGE | 0.0525 | 1.05 | 0.0064 | 67.17 | 0.0000 | 1.0408 | 1.0672 |
| 24 | Intercept | -6.0119 | 0.00 | 0.2334 | 663.39 | 0.0000 | 0.0016 | 0.0039 |
| 24 | femmes | -0.3221 | 0.72 | 0.1655 | 3.79 | 0.0517 | 0.5239 | 1.0024 |
| 24 | DHC8_AGE | 0.0553 | 1.06 | 0.0036 | 231.04 | 0.0000 | 1.0494 | 1.0645 |
| 35 | Intercept | -5.9343 | 0.00 | 0.2038 | 847.62 | 0.0000 | 0.0018 | 0.0039 |
| 35 | femmes | -0.3885 | 0.68 | 0.1332 | 8.51 | 0.0035 | 0.5223 | 0.8803 |
| 35 | DHC8_AGE | 0.0568 | 1.06 | 0.0032 | 321.10 | 0.0000 | 1.0518 | 1.0650 |
| 59 | Intercept | -5.9866 | 0.00 | 0.3175 | 355.62 | 0.0000 | 0.0013 | 0.0047 |
| 59 | femmes | -0.4421 | 0.64 | 0.2385 | 3.44 | 0.0637 | 0.4027 | 1.0256 |
| 59 | DHC8_AGE | 0.0550 | 1.06 | 0.0049 | 124.78 | 0.0000 | 1.0464 | 1.0668 |

**Tests d'indépendance du CHI-DEUX ajusté (ajustement de Rao-Scott du second degré)
(500 répliques bootstrap)**

----- Test=1: Variables: diab VS femmes -----

| PRC8_CUR | Statistique | Deg. de liberté | Valeur du Chi-deux | Valeur p |
|----------|---|-----------------|--------------------|----------|
| 10 | Chi-2 ajusté (ajust. R.-S. du sec. degré) | 1 | 6.41 | 0.0113 |
| 24 | Chi-2 ajusté (ajust. R.-S. du sec. degré) | 1 | 1.29 | 0.2564 |
| 35 | Chi-2 ajusté (ajust. R.-S. du sec. degré) | 1 | 2.97 | 0.0849 |
| 59 | Chi-2 ajusté (ajust. R.-S. du sec. degré) | 1 | 1.70 | 0.1919 |

Note: Un astérisque (*) dans la colonne Statistique indique que le tableau contient une ou des cellule(s) avec 5 observations (non-pondérées) ou moins. Le Chi-deux peut ne pas être un test valide dans de tels cas.

**Tests d'indépendance du CHI-DEUX: Statistiques variées provenant des tableaux croisés
(500 répliques bootstrap)**

----- Test=1: Variables: diab VS femmes -----

| PRC8_CUR | diab | Statistique | femmes_0 | femmes_1 | total |
|----------|------|-------------|----------|----------|--------|
| 10 | . | Pourcentage | 0.4938 | 0.5062 | . |
| 10 | . | Variance | 0.0000 | 0.0000 | . |
| 10 | . | _C.V._ | 0.0000 | 0.0000 | . |
| 10 | 0 | Pourcentage | 0.4787 | 0.4765 | 0.9552 |
| 10 | 0 | Variance | 0.0000 | 0.0000 | 0.0000 |
| 10 | 0 | _C.V._ | 0.0062 | 0.0073 | 0.0042 |
| 10 | 1 | Pourcentage | 0.0152 | 0.0296 | 0.0448 |
| 10 | 1 | Variance | 0.0000 | 0.0000 | 0.0000 |
| 10 | 1 | _C.V._ | 0.1965 | 0.1170 | 0.0890 |
| 24 | . | Pourcentage | 0.4915 | 0.5085 | . |
| 24 | . | Variance | 0.0000 | 0.0000 | . |
| 24 | . | _C.V._ | 0.0009 | 0.0008 | . |
| 24 | 0 | Pourcentage | 0.4734 | 0.4929 | 0.9663 |
| 24 | 0 | Variance | 0.0000 | 0.0000 | 0.0000 |
| 24 | 0 | _C.V._ | 0.0039 | 0.0039 | 0.0027 |
| 24 | 1 | Pourcentage | 0.0182 | 0.0156 | 0.0337 |
| 24 | 1 | Variance | 0.0000 | 0.0000 | 0.0000 |
| 24 | 1 | _C.V._ | 0.0998 | 0.1188 | 0.0779 |
| 35 | . | Pourcentage | 0.4899 | 0.5101 | . |
| 35 | . | Variance | 0.0000 | 0.0000 | . |
| 35 | . | _C.V._ | 0.0001 | 0.0001 | . |
| 35 | 0 | Pourcentage | 0.4690 | 0.4927 | 0.9617 |
| 35 | 0 | Variance | 0.0000 | 0.0000 | 0.0000 |
| 35 | 0 | _C.V._ | 0.0036 | 0.0029 | 0.0022 |
| 35 | 1 | Pourcentage | 0.0210 | 0.0174 | 0.0383 |
| 35 | 1 | Variance | 0.0000 | 0.0000 | 0.0000 |
| 35 | 1 | _C.V._ | 0.0800 | 0.0824 | 0.0544 |
| 59 | . | Pourcentage | 0.4943 | 0.5057 | . |
| 59 | . | Variance | 0.0000 | 0.0000 | . |
| 59 | . | _C.V._ | 0.0003 | 0.0003 | . |
| 59 | 0 | Pourcentage | 0.4754 | 0.4914 | 0.9668 |
| 59 | 0 | Variance | 0.0000 | 0.0000 | 0.0000 |
| 59 | 0 | _C.V._ | 0.0054 | 0.0044 | 0.0034 |
| 59 | 1 | Pourcentage | 0.0189 | 0.0143 | 0.0332 |
| 59 | 1 | Variance | 0.0000 | 0.0000 | 0.0000 |
| 59 | 1 | _C.V._ | 0.1365 | 0.1532 | 0.0983 |