

12.4. Estimation de la variance à l'aide des poids de bootstrap

12.4.1 Méthode du bootstrap

Le plan d'échantillonnage pour l'Enquête nationale sur la santé de la population (ENSP) est complexe. On utilise donc une méthode de rééchantillonnage pour calculer des estimations de la variance. Pour l'ENSP, on a utilisé la méthode du bootstrap.

Cette méthode consiste à diviser les enregistrements sur le fichier de données en sous-groupes (ou répétitions) et à déterminer la variance entre les estimations pour chaque répétition. Les répétitions sont formées de la manière suivante: à l'intérieur de chaque strate, on choisit un échantillon aléatoire simple (ÉAS), avec remise, de $n-1$ grappes parmi les n grappes de la strate. Pour chaque échantillon ÉAS, le facteur de pondération est recalculé pour chaque enregistrement dans la strate. Ce poids recalculé correspond au premier poids bootstrap. Ce processus est répété B fois afin de former un ensemble de B poids bootstrap pour chaque enregistrement sur le fichier de données.

12.4.2 Poids bootstrap pour l'ENSP

Pour l'ENSP, six ensembles de poids bootstrap ont été créés. Le premier ensemble, B5H35, est utilisé avec le fichier de microdonnées H35. Il contient 500 poids bootstrap pour chaque enregistrement. Trois ensembles ont été créés pour le fichier de microdonnées H356 et contiennent chacun 500 poids bootstrap. Le premier de ces trois ensembles (B5H356) est utilisé pour les analyses où on utilise le poids WT66. Pour les analyses réalisées à l'aide du poids WT66_N, deux ensembles de poids bootstrap sont utilisés: on utilise l'ensemble B5H356A pour les analyses sur les questions de la section *Enquête de la promotion sur la santé* et B5H356C pour les analyses sur les questions de la section *Service de santé pour enfants* pour les enfants au Manitoba et en Alberta (voir Section 12.1.3 du *Guide du fichier de microdonnées à grande diffusion de l'ENSP, 1996-1997* pour la liste de ces questions et à qui elles s'appliquent). Finalement, les ensembles B5LONGF et B5LONGP sont utilisés avec les fichiers de microdonnées longitudinales LNGF et LNGP. Ils contiennent 500 poids bootstrap pour chaque enregistrement.

Les poids bootstrap doivent être utilisés en association avec les tableaux de la variabilité d'échantillonnage approximative (voir Section 10 du document ci-haut mentionné pour une description plus complète de ces tableaux). Ces tableaux peuvent être utilisés pour obtenir des coefficients de variation (C.V.) approximatifs pour des estimations catégoriques telles que des totaux et des ratios, pour des domaines spécifiques. Les poids bootstrap devraient être utilisés pour obtenir des estimations de variance plus précises ou pour des domaines qui ne sont pas couverts par les tableaux de la variabilité d'échantillonnage approximative et pour des analyses plus complexes comme des régressions. Les mêmes règles de confidentialité et seuils de diffusion s'appliquent pour les estimations de variance provenant du bootstrap.

Ces poids doivent être utilisés afin de calculer une estimation valide de la variance d'une estimation donnée. Par exemple, si on veut calculer l'estimation d'un total en utilisant le fichier H35, on suit les étapes suivantes:

A) On calcule d'abord une estimation du total en utilisant le poids final WT56.

B) On calcule ensuite une estimation du total voulu en utilisant chacun des 500 poids bootstrap comme pondération de l'échantillon. On obtient alors 500 estimations du total en question.

C) Finalement, on calcule la variance de ces 500 estimations. Cette variance est l'estimation de la variance du total calculé en A).

12.4.3 Programme BOOTVAR.SAS

Le programme SAS BOOTVAR.SAS est fourni avec les ensembles de données comme exemple pour calculer l'estimation de la variance. Il peut être utilisé par les utilisateurs ou ceux-ci peuvent créer leur propre programme. Une copie du programme apparaît à la fin de cette section.

Le programme BOOTVAR.SAS calcule une estimation (par exemple, un total) et une estimation de la variance de cette dernière. Il est toutefois conseillé de calculer l'estimation désirée (dans ce cas-ci, le total) à l'aide d'un autre programme pour vérifier que le programme BOOTVAR.SAS calcule bien l'estimation voulue.

Dans cette section, on décrit comment préparer les données pour le programme BOOTVAR.SAS et comment l'utiliser. Deux exemples sont donnés ainsi que le temps requis pour soumettre le programme (ces programmes ont été exécutés en utilisant un Pentium Pro-S 200MHz, avec 64 MB de RAM).

12.4.3.1 Définition des variables pour l'analyse

Le programme BOOTVAR.SAS permet de calculer des estimations de variance pour des totaux, des rapports, des différences entre des rapports, et des paramètres de régression linéaire et logistique. Dans un premier temps, les données doivent être préparées et sauvegardées dans un fichier SAS pour réaliser les analyses voulues. Pour des estimations de totaux, rapports et différences entre des rapports, les estimations sont calculées en sommant le poids des enregistrements qui ont la caractéristique d'intérêt. Une variable indicatrice doit être créée pour identifier ces enregistrements (voir exemple 1A). Pour les analyses de régression, les variables à analyser doivent être définies de la façon requise par SAS (voir exemple 1B).

Une variable supplémentaire (REGION) peut être nécessaire si les analyses sont réalisées pour des sous-groupes géographiques d'enregistrements sur le nouveau fichier de données créé pour le programme BOOTVAR.SAS. Par exemple, si le nouveau fichier d'analyse contient tous les enregistrements des personnes âgées de 12 ans et plus au Canada, l'ensemble des enregistrements représente un groupe qui correspond au Canada. Si les analyses sont faites au niveau du Canada, la variable REGION n'est pas nécessaire. Toutefois, si les analyses doivent être réalisées au niveau des provinces, alors la variable REGION est nécessaire. Cette variable doit être créée par l'utilisateur et prendra, pour chaque enregistrement, la valeur de la région géographique de cet enregistrement, dans ce cas-ci, la province. Autre exemple, si le nouveau fichier d'analyse contient tous les enregistrements des personnes âgées de moins de 12 ans en Ontario, l'ensemble des enregistrements représente un groupe qui correspond à la province Ontario. Si les analyses sont faites au niveau de la province Ontario, la variable REGION n'est pas nécessaire. Toutefois, si les analyses sont faites au niveau des régions sanitaires, la variable REGION est alors nécessaire et prendra la valeur de la région sanitaire pour chaque enregistrement. De même, si le nouveau fichier d'analyse contient tous les enregistrements des personnes âgées de 12 ans et plus dans les provinces de l'Ontario, du Manitoba et de l'Alberta, la variable REGION sera nécessaire si les analyses sont faites au niveau des provinces ou encore au niveau des régions sanitaires.

L'exemple suivant démontre comment définir les variables qui seront ensuite utilisées pour calculer des estimations de variance à l'aide du programme BOOTVAR.SAS.

Exemple 1: Définition des variables pour l'analyse

Exemple 1A: Totaux et rapports

```

/*****
/*
/*          UTILISE LE FICHIER TRANSVERSAL DE LA COMPOSANTE GÉNÉRALE
/*          DE 1996 (H35.TXT)
/*          POUR CALCULER LE NOMBRE DE PERSONNES DIABÉTIQUES
/*          POUR LES GROUPES D'ÂGES 12-17, 18-24, 25-44, ET 45 ET PLUS
/*          ET POUR LES HOMMES, FEMMES, ET LES DEUX,
/*          PAR PROVINCE
/*
*****/

libname in1 'c:\data';

proc datasets kill;
run;

proc format;
  value fsex      1='hommes'
                  2='femmes';

  value fage      1='12-17'
                  2='18-24'
                  3='25-44'
                  4='45+';

  value fdiab     1='Oui'
                  0='Non';

```

```

run;

data in1.cross;
%let datafid='c:\data\h35.txt';
%include 'c:\data\h35_in.sas';
keep region di ab total dhc6_sex hommes femmes
    agegrp age1217 age1824 age2544 age45 magegrp fagegrp
    mage1217 mage1824 mage2544 mage45 fage1217 fage1824 fage2544 fage45
    mdi ab fdi ab di ab1217 di ab1824 di ab2544 di ab45
    mdi a1217 mdi a1824 mdi a2544 mdi a45 fdi a1217 fdi a1824 fdi a2544 fdi a45
    wt56 dhc6_sex total realukey personid;

region=prc6_cur;

if ccc6_1j=1 then di ab=1;
else if ccc6_1j=2 then di ab=0;
else di ab=.;

if di ab=. then delete;

/*****
/* définit les domaines d'intérêt avec des
/* variable indicatrices (0/1)
*****/

total=1;

/* sexe */

if dhc6_sex=1 then hommes=1;
else hommes=0;

if dhc6_sex=2 then femmes=1;
else femmes=0;

/* groupe d'âge */

if dhc6_age < 18 then agegrp=1;
else if dhc6_age < 25 then agegrp=2;
else if dhc6_age < 45 then agegrp=3;
else agegrp=4;

if agegrp=1 then age1217=1; else age1217=0;
if agegrp=2 then age1824=1; else age1824=0;
if agegrp=3 then age2544=1; else age2544=0;
if agegrp=4 then age45=1; else age45=0;

/* groupe âge-sexe */

if agegrp=1 then do;
    if dhc6_sex=1 then magegrp=1;
    else if dhc6_sex=2 then fagegrp=1;
end;
else if agegrp=2 then do;
    if dhc6_sex=1 then magegrp=2;
    else if dhc6_sex=2 then fagegrp=2;
end;
else if agegrp=3 then do;
    if dhc6_sex=1 then magegrp=3;
    else if dhc6_sex=2 then fagegrp=3;
end;
else if agegrp=4 then do;
    if dhc6_sex=1 then magegrp=4;
    else if dhc6_sex=2 then fagegrp=4;
end;

if magegrp=1 then mage1217=1; else mage1217=0;
if magegrp=2 then mage1824=1; else mage1824=0;
if magegrp=3 then mage2544=1; else mage2544=0;
if magegrp=4 then mage45=1; else mage45=0;

if fagegrp=1 then fage1217=1; else fage1217=0;
if fagegrp=2 then fage1824=1; else fage1824=0;
if fagegrp=3 then fage2544=1; else fage2544=0;
if fagegrp=4 then fage45=1; else fage45=0;

/* diabète*sexe */

if di ab=1 then do;
    if dhc6_sex=1 then mdi ab=1; else mdi ab=0; /* hommes diabétiques */
    if dhc6_sex=2 then fdi ab=1; else fdi ab=0; /* femmes diabétiques */
end;
else if di ab=2 then do;
    mdi ab=0;

```

```

        fdi ab=0;
    end;

    /* di abète*groupe d' âge */
    if di ab=1 then do;
        if age1217=1 then di ab1217=1; el se di ab1217=0;
        if age1824=1 then di ab1824=1; el se di ab1824=0;
        if age2544=1 then di ab2544=1; el se di ab2544=0;
        if age45=1 then di ab45=1; el se di ab45=0;
    end;
    el se do;
        di ab1217=0;
        di ab1824=0;
        di ab2544=0;
        di ab45=0;
    end;

    /* di abète*sexe*groupe d' âge */
    if mdi ab=1 then do;
        if age1217=1 then mdi a1217=1; el se mdi a1217=0;
        if age1824=1 then mdi a1824=1; el se mdi a1824=0;
        if age2544=1 then mdi a2544=1; el se mdi a2544=0;
        if age45=1 then mdi a45=1; el se mdi a45=0;
    end;
    el se do;
        mdi a1217=0;
        mdi a1824=0;
        mdi a2544=0;
        mdi a45=0;
    end;

    if fdi ab=1 then do;
        if age1217=1 then fdi a1217=1; el se fdi a1217=0;
        if age1824=1 then fdi a1824=1; el se fdi a1824=0;
        if age2544=1 then fdi a2544=1; el se fdi a2544=0;
        if age45=1 then fdi a45=1; el se fdi a45=0;
    end;
    el se do;
        fdi a1217=0;
        fdi a1824=0;
        fdi a2544=0;
        fdi a45=0;
    end;

    proc freq data=in1.cross;
    format agegrp fage. magegrp fage. fagegrp fage. di ab fdi ab. ;
    table region*(di ab di ab2544 mdi a45 fdi a45)
           region*di ab*total region*mdi a45*mage45 region*di ab45*age45 ;
    weight wt56;
    title 'Nombre de diabétiques par groupe d' âge et région';
    run;

```

Exemple 1B: Régression logistique

```

/*****
/*
/*          UTILISE LE FICHIER TRANSVERSAL DE LA COMPOSANTE SANTÉ          */
/*          DE 1996 (H356.TXT)                                              */
/*          POUR ETUDIER LA RELATION ENTRE LA DÉCLARATION DE PROBLÈMES    */
/*          DE SANTÉ CHROMIQUES ET LE TYPE D'INTERVIEW (PAR PROCURATION    */
/*          OU NON).                                                         */
/*
/*****

libname in1 'c:\data';

data in1.h356;
    %let datafid="c:\data\h356.txt";
    %include "c:\data\h356_in.sas";
    keep realukey personid dhc6_age dhc6_sex am56_pxy ccc6dany
    age1217 age1820 age2124 age2534 age3544 age4554 age5564
    age6574 age7584 age85 female nonproxy wt66;
    if dhc6_age >= 12;

/* Variables indicatrices pour les groupes d' âge */

age1217=0;
age1820=0;

```

```

age2124=0;
age2534=0;
age3544=0;
age4554=0;
age5564=0;
age6574=0;
age7584=0;
age85=0;

if 12 <= dhc6_age <=17 then age1217=1;
if 18 <= dhc6_age <=20 then age1820=1;
if 21 <= dhc6_age <=24 then age2124=1;
if 25 <= dhc6_age <=34 then age2534=1;
if 35 <= dhc6_age <=44 then age3544=1;
if 45 <= dhc6_age <=54 then age4554=1;
if 55 <= dhc6_age <=64 then age5564=1;
if 65 <= dhc6_age <=74 then age6574=1;
if 75 <= dhc6_age <=84 then age7584=1;
if dhc6_age>=85 then age85=1;

female=0;
if dhc6_sex=2 then female=1;

/* Valeurs manquantes */

if ccc6dany>2 then ccc6dany=.;
if ccc6dany=2 then ccc6dany=0;

nonproxy=0;
if am56_pxy>2 then nonproxy=.;
if am56_pxy=2 then nonproxy=1;
run;

proc logistic data=in1.h356 descending;
  model ccc6dany = nonproxy age1820 age2124 age2534 age3544
    age4554 age5564 age6574 age7584 age85 female;
  weight wt66;
  title 'Problèmes de santé chroniques';
run;

```

Exemple 1C: Modèle linéaire généralisé

```

/*****
*/
/*      UTILISE LE FICHIER LONGITUDINAL POUR 1996      */
/*      CALCULE LE NOMBRE MOYEN DE CONSULTATIONS AVEC UN MÉDECIN PAR      */
/*      LES FEMMES AYANT DES ENFANTS SELON LEUR ARRANGEMENT FAMILIAL      */
/*      (MONOPARENTAL OU VIVANT AVEC UN CONJOINT, AVEC ENFANTS)      */
/*      */
/*****

libname in1 'c:\data';

data long;
  %let datafid="c:\data\lngf.txt";
  %include "c:\data\lngf_in.sas";
  keep realukey personid dhc4dlvg dhc6dlvg dhc6decf dhc4decf dhc6gage sex mhc6dds
  hcc6_2a hcc6_2c wt66l f;
run;

data in1.long;
  set long;
  if sex=2; /* femmes seulement */
  if dhc6gage >= 5; /* 12 ans et plus */
  if dhc4decf in (5,9); /* femmes vivant en couple avec enfants < 25 ans seulement
                        ou monoparental avec enfants < 25 and seulement
                        (1994) */
  if dhc6decf in (5,9); /* femmes vivant en couple avec enfants < 25 ans seulement
                        ou monoparental avec enfants < 25 and seulement
                        (1996) */
  if dhc4dlvg in (4,5); /* vivant en couple ou célibataire avec enfants seulement
                        (1994) */
  if dhc6dlvg in (4,5); /* vivant en couple ou célibataire avec enfants seulement
                        (1996) */

  /* arrangement familial */

  if dhc4dlvg=4 and dhc6dlvg=4 then living='couple 94-96';
  else if (dhc4dlvg=4 or dhc4dlvg=5) and dhc6dlvg=5 then living='monoparental';
  else delete;

```

```

/* nombre de consultations avec un médecin, 1996 */
if hcc6_2a < 996 and hcc6_2c < 996 then doc=sum(hcc6_2a, hcc6_2c);
else delete;
run;

```

12.4.3.2 Utilisation du programme BOOTVAR.SAS

Une fois le nouveau fichier d'analyse créé, le programme BOOTVAR.SAS peut être exécuté. L'utilisateur doit toutefois compléter le programme en spécifiant des paramètres et en indiquant la liste des variables à analyser. Le programme BOOTVAR.SAS est divisé en quatre sections.

À la **section 1**, l'utilisateur doit spécifier:

- le nombre de poids bootstrap utilisés (500 pour tous les ensembles),
- les répertoires contenant le nouveau fichier d'analyse (créé par l'utilisateur, comme dans l'exemple 1), l'ensemble de poids bootstrap, et le fichier qui contiendra les résultats,
- le nom du nouveau fichier d'analyse,
- le nom du fichier contenant l'ensemble de poids bootstrap, et
- des paramètres qui indiquent si les analyses sont faites à un niveau géographique utilisant la variable REGION.

À la **section 2**, les seuls changements à faire sont à la macro pour le modèle linéaire généralisé. Cette macro doit être modifiée par l'utilisateur selon l'analyse voulue.

À la **section 3**, l'utilisateur doit spécifier la liste des variables à conserver pour les analyses. Le programme BOOTVAR.SAS est exécuté assez rapidement lorsqu'on calcule des estimations de variance pour des totaux, des rapports, et des différences entre des rapports. Cependant, plus de temps est requis dans le cas des estimations de variance pour les paramètres d'une régression. Afin de limiter le temps d'exécution, il est donc recommandé de ne conserver que les variables nécessaires aux fins de l'analyse voulue.

En plus des variables d'analyse, deux variables doivent être conservées sur le fichier: les variables REALUKEY et PERSONID. Ces variables sont nécessaires pour faire l'appariement entre le fichier de données et le fichier contenant l'ensemble de poids bootstrap durant l'exécution du programme. Il est à noter que le programme BOOTVAR.SAS lit deux fichiers: les données et les poids bootstrap, et ces deux fichiers doivent être des fichiers SAS.

Finalement, à la **section 4**, l'utilisateur doit indiquer quelles analyses seront exécutées et pour quelles variables.

Les exemples suivants indiquent en caractères *italiques* ce que l'utilisateur doit spécifier dans le cas d'estimations de variance pour des totaux, des rapports, et des paramètres de régression logistique. Il est à noter que seuls des extraits du programme où des changements sont nécessaires sont présentés. Il faut cependant soumettre le programme au complet.

Exemple 2: Programme BOOTVAR.SAS

Exemple 2A: Totaux et rapports

```
.
.
.

/***** Section 1: Déclaration des variables macros *****/
/***** Section 1: Déclaration des variables macros *****/

%let R=01; /* paramètre du bootstrap moyen (fixe) */

/***** L'UTILISATEUR DOIT INDiquer LE NOMBRE DE POIDS BOOTSTRAP À UTILISER, *****/
/***** DIVISÉ PAR 100 *****/
/***** %let B=5; pour les fichiers de poids bootstrap B5H35, B5H356, *****/
/***** B5H356A, B5H356C, B5LONGF et B5LONGP *****/
/***** *****/

%let B=5; /* nombre de poids de bootstrap Bx100 */

/***** L'UTILISATEUR DOIT INDiquer LE RÉPERTOIRE CONTENANT LE FICHIER *****/
/***** D'ANALYSE (in1) ET LE FICHIER CONTENANT LES POIDS DE BOOTSTRAP *****/
/***** (in2). LES FICHIERs DE SORTIES SERONT SAUVEGARDÉS DANS LE *****/
/***** RÉPERTOIRE SPÉCIFIÉ POUR in3. *****/
/***** *****/

libname in1 "c:\data";
libname in2 "c:\bootstrp";
libname in3 "c:\output";

/***** L'UTILISATEUR DOIT INDiquer LE NOM DU FICHIER SAS CONTENANT LES VARIABLES *****/
/***** D'ANALYSE. *****/
/***** %let Mfile=xxx; *****/
/***** *****/

%let Mfile=in1.cross; /* Nom du fichier SAS principal */

/***** L'UTILISATEUR DOIT INDiquer LE NOM DU FICHIER SAS CONTENANT LES POIDS DE *****/
/***** BOOTSTRAP. *****/
/***** %let bsamp=b5h35; (à utiliser avec le fichier H35.TXT) OU *****/
/***** %let bsamp=b5h356; (à utiliser avec le fichier H356.TXT) OU *****/
/***** %let bsamp=b5h356a; (à utiliser avec le fichier H356.TXT) OU *****/
/***** %let bsamp=b5h356c; (à utiliser avec le fichier H356.TXT) OU *****/
/***** %let bsamp=b5longf; (à utiliser avec le fichier LONGF.TXT) *****/
/***** %let bsamp=b5longp; (à utiliser avec le fichier LONGP.TXT) *****/
/***** *****/

%let bsamp=in2.b5h35; /* fichier SAS des poids bootstrap */

/***** L'UTILISATEUR DOIT SPÉCIFIER LA VALEUR DES PARAMÈTRES by, *****/
/***** wo ET kp. CEUX-CI INDICENT SI L'ANALYSE EST FAITE AU *****/
/***** NIVEAU RÉGIONAL OU POUR L'ENSEMBLE DU FICHIER. *****/
/***** *****/
/***** SI L'ANALYSE EST FAITE POUR L'ENSEMBLE DU FICHIER, ALORS *****/
/***** %let by=*; *****/
/***** %let wo=; *****/
/***** %let kp=; *****/
/***** *****/
/***** SI NON, SI L'ANALYSE EST FAITE AU NIVEAU RÉGIONAL, ALORS *****/
/***** %let by=; *****/
/***** %let wo=; *****/
/***** *****/
```



```

/**** %let kp=region; ****/
/*****/

%let by=;
%let wo=;
%let kp=region;

data time;
  format start datetime16.;
  start=datetime();
  output;
run;

/*****/
/**** Section 2: Déclaration des macros ****/
/*****/

.
.
.

*****/
* Section 3: Programme principal */
*****/

/* Lit le fichier principal */

/*****/
/**** À CAUSE DU GRAND NOMBRE D'OBSERVATIONS, SEULES LES VARIABLES ****/
/**** REQUISES POUR FAIRE LES ANALYSES ET LES VARIABLES, ****/
/**** "REALUKEY" ET "PERSONID" DEVRAIENT ÊTRE CONSERVÉES DÉSORMAIS. ****/
/*****/

data nphs (index=(id=(realukey personid)));
  set &Mfile (keep= realukey personid region di ab total dhc6_sex
hommes femmes agegrp age1217 age1824 age2544 age45
magegrp fagegrp mage1217 mage1824 mage2544 mage45
fage1217 fage1824 fage2544 fage45 mdi ab fdi ab
di ab1217 di ab1824 di ab2544 di ab45 mdi a1217 mdi a1824
mdi a2544 mdi a45 fdi a1217 fdi a1824 fdi a2544 fdi a45
wt56 dhc6_sex total);
run;

/* Lit les poids de bootstrap */
/* FWGT est le même poids que WT56, WT66, WT66_N, WT66LF ou WT66LP selon le
fichier analysé */

data bsamp (index=(id=(realukey personid)));
  set &bsamp;
  keep fwgt realukey personid bsw1-bsw&B.00;
run;

/* Apparie le fichier principal et les poids bootstrap */

&by data bs_nphs (index=(region));
&wo data bs_nphs;
  merge nphs (in=in1) bsamp (in=in2);
  by realukey personid;
  if in1;
run;

proc datasets library=work;
  delete nphs bsamp;
run;

```



```

run;

&by proc sort data=alltots;
&by by region;
&by run;

/***** Les résultats des macros total et ratio peuvent être
*** être sauvegardés dans un fichier permanent *****/
/*****/

data in3.ZZZ;
set alltots;
run;

%prnttot;

.
.
.

data time;
set time;
format stop datetime16.;
stop=datetime();
output;
run;

proc print data=time;
title 'Temps requis pour exécuter le programme';
run;

/* Fin du programme SAS Bootvar */

```

(L'exécution de ce programme a pris environ 3 minutes.)

Exemple 2B: Régression logistique

```

.
.
.

/*****/
*** Section 1: Déclaration des variables macros *****/
/*****/

%let R=01; /* paramètre du bootstrap moyen (fixe) */

/*****/
*** L'UTILISATEUR DOIT INDICER LE NOMBRE DE POIDS BOOTSTRAP À UTILISER, ***
*** DIVISÉ PAR 100 *****/
*** %let B=5; pour les fichiers de poids bootstrap B5H35, B5H356, ***
*** B5H356A, B5H356C, B5LONGF et B5LONGP *****/
/*****/

%let B=5; /* nombre de poids de bootstrap Bx100 */

/*****/
*** L'UTILISATEUR DOIT INDICER LE RÉPERTOIRE CONTENANT LE FICHIER ***
*** D'ANALYSE (in1) ET LE FICHIER CONTENANT LES POIDS DE BOOTSTRAP ***
*** (in2). LES FICHIERS DE SORTIES SERONT SAUVEGARDES DANS LE ***
*** RÉPERTOIRE SPÉCIFIÉ POUR in3. *****/
/*****/

libname in1 "c:\data";
libname in2 "c:\bootstrp";
libname in3 "c:\output";

/*****/
*** L'UTILISATEUR DOIT INDICER LE NOM DU FICHIER SAS CONTENANT LES VARIABLES ***
*** D'ANALYSE. *****/
*** %let Mfile=xxx; *****/
/*****/

%let Mfile=in1.h356; /* Nom du fichier SAS principal */

```

```

/*****
*** L'UTILISATEUR DOIT INDiquer LE NOM DU FICHIER SAS CONTENANT LES POIDS DE
*** BOOTSTRAP.
*** %let bsamp=b5h35; (à utiliser avec le fichier H35.TXT) OU
*** %let bsamp=b5h356; (à utiliser avec le fichier H356.TXT) OU
*** %let bsamp=b5h356a; (à utiliser avec le fichier H356.TXT) OU
*** %let bsamp=b5h356c; (à utiliser avec le fichier H356.TXT) OU
*** %let bsamp=b5longf; (à utiliser avec le fichier LONGF.TXT)
*** %let bsamp=b5longp; (à utiliser avec le fichier LONGP.TXT)
*****/

%let bsamp=i n2. b5h356; /* fichier SAS des poids bootstrap */

/*****
*** L'UTILISATEUR DOIT SPÉCIFIER LA VALEUR DES PARAMÈTRES by,
*** wo ET kp. CEUX-CI INDICENT SI L'ANALYSE EST FAITE AU
*** NIVEAU RÉGIONAL OU POUR L'ENSEMBLE DU FICHIER.
*** SI L'ANALYSE EST FAITE POUR L'ENSEMBLE DU FICHIER, ALORS
*** %let by=*;
*** %let wo=;
*** %let kp=;
*** SINON, SI L'ANALYSE EST FAITE AU NIVEAU RÉGIONAL, ALORS
*** %let by=;
*** %let wo=*;
*** %let kp=regi on;
*****/

%let by=*;
%let wo=;
%let kp=;

data time;
format start datetime16.;
start=datetime();
output;
run;

/*****
*** Section 2: Déclaration des macros
*****/
.
.
.

*****
* Section 3: Programme principal
*****
;

/* Lit le fichier principal */

/*****
*** À CAUSE DU GRAND NOMBRE D'OBSERVATIONS, SEULES LES VARIABLES
*** REQUISES POUR FAIRE LES ANALYSES ET LES VARIABLES,
*** "REALUKEY" ET "PERSONID" DEVRAIENT ÊTRE CONSERVÉES DÉSORMAIS.
*****/

data nphs (index=(id=(realukey personid)));
set &Mfile (keep= realukey personid ccc6dany nonproxy age1820 age2124 age2534
age3544 age4554 age5564 age6574 age7584 age85 female);
run;

/* Lit les poids de bootstrap */
/* FWGT est le même poids que WT56, WT66, WT66_N, WT66LF ou WT66LP selon le
fichier analysé */

data bsamp (index=(id=(realukey personid)));
set &bsamp;
keep fwgt realukey personid bsw1-bsw&B.00;
run;

/* Apparie le fichier principal et les poids bootstrap */

&by data bs_nphs (index=(region));
&wo data bs_nphs;
merge nphs (in=in1) bsamp (in=in2);
by realukey personid;
if in1;
run;

proc datasets library=work;
delete nphs bsamp;
run;

```

```

/*****
*** Section 4: Énoncés pour soumettre les macros ****
*****/

.
.
.

/*****/
/*
*/
/* Régression Logistique */
/*
*/
/*****/

proc datasets library=work; /* initialise bs_reglg */
delete bs_reglg;
run;

/*****/
*** Pour soumettre la macro logreg, l'énoncé suivant est utilisé: ****
*** %logreg(infile, yvar, xvar, bssz=, mul ti=); ****
*** où      infile : bs_nphs ****
***      yvar   : la variable dépendante ****
***      xvar   : les variableS indépendantes. Les variables doivent ****
***              être inscrites comme elles le seraient dans l'énon- ****
***              cé d'un modèle (sans virgules entre elles). ****
***      bssz=  : bssz=&B.00 ****
***      mul ti = : mul ti=&R ****
/*****/

title "Estimation de la variance à l'aide du bootstrap &B.00:&R pour des";
title2 "Régressions Logistiques";

* %logreg(bs_nphs, YVAR, XVAR, bssz=&B.00, mul ti=&R);
%logreg(bs_nphs, ccc6dany, nonproxy age1820 age2124 age2534 age3544
age4554 age5564 age6574 age7584 age85 female, bssz=&B.00, mul ti=&R);

/* Détruit les fichiers SAS de la macro logreg */
proc datasets library=work;
delete betas betat bsbeta origest;
run;

.
.
.

%prntlog;

data time;
set time;
format stop datetime16.;
stop=datetime();
output;
run;

proc print data=time;
title 'Temps requis pour exécuter le programme';
run;

/* Fin du programme SAS Bootvar */

```

(L'exécution de ce programme a pris environ 20 minutes.)

Exemple 2C: Modèle linéaire généralisé

```

.
.
.

/*****/
*** Section 1: Déclaration des variables macros ****
*****/

%let R=01;                                /* paramètre du bootstrap moyen (fixe) */

```

```

/***** L'UTILISATEUR DOIT INDIQUER LE NOMBRE DE POIDS BOOTSTRAP À UTILISER, ****/
/**** DIVISÉ PAR 100 ****/
/**** %let B=5; pour les fichiers de poids bootstrap B5H35, B5H356, B5H356A, ****/
/**** B5H356C, B5LONGF et B5LONGP ****/
/**** */

%let B=5; /* nombre de poids bootstrap Bx100 */

/***** L'UTILISATEUR DOIT INDIQUER LE RÉPERTOIRE CONTENANT LE FICHIER ****/
/**** D'ANALYSE (in1) ET LE FICHIER CONTENANT LES POIDS BOOTSTRAP ****/
/**** (in2). LES FICHIERS DE SORTIES SERONT SAUVEGARDES DANS LE ****/
/**** RÉPERTOIRE SPÉCIFIÉ POUR in3. ****/
/**** */

libname in1 "c:\data";
libname in2 "c:\bootstrp";
libname in3 "c:\output";

/***** L'UTILISATEUR DOIT INDIQUER LE NOM DU FICHIER SAS CONTENANT LES VARIABLES ****/
/**** D'ANALYSE. ****/
/**** %let Mfile=xxx; ****/
/**** */

%let Mfile=in1.long; /* Nom du fichier SAS principal */

/***** L'UTILISATEUR DOIT INDIQUER LE NOM DU FICHIER SAS CONTENANT LES POIDS ****/
/**** BOOTSTRAP. ****/
/**** %let bsamp=b5h35; (à utiliser avec le fichier H35.TXT) OU ****/
/**** %let bsamp=b5h356; (à utiliser avec le fichier H356.TXT) OU ****/
/**** %let bsamp=b5h356a; (à utiliser avec le fichier H356.TXT) OU ****/
/**** %let bsamp=b5h356c; (à utiliser avec le fichier H356.TXT) OU ****/
/**** %let bsamp=b5longf; (à utiliser avec le fichier LONGF.TXT) ****/
/**** %let bsamp=b5longp; (à utiliser avec le fichier LONGP.TXT) ****/
/**** */

%let bsamp=in2.b5longf; /* fichier SAS des poids bootstrap */

/***** L'UTILISATEUR DOIT SPÉCIFIER LA VALEUR DES PARAMÈTRES by, ****/
/**** wo ET kp. CEUX-CI INDIQUENT SI L'ANALYSE EST FAITE AU ****/
/**** NIVEAU RÉGIONAL OU POUR L'ENSEMBLE DU FICHIER. ****/
/**** */
/**** SI L'ANALYSE EST FAITE POUR L'ENSEMBLE DU FICHIER, ALORS ****/
/**** %let by=*; ****/
/**** %let wo=; ****/
/**** %let kp=; ****/
/**** */
/**** SINON, SI L'ANALYSE EST FAITE AU NIVEAU RÉGIONAL, ALORS ****/
/**** %let by=; ****/
/**** %let wo=; ****/
/**** %let kp=region; ****/
/**** */

%let by=*;
%let wo=;
%let kp=;

data time;
format start datetime16.;
start=datetime();
output;
run;

/***** Section 2: Déclaration des macros ****/
/**** */
.
.
.

*****;

%macro regglm(infile, yvar, xvar, bssz=, mul ti=);
*****;
proc glm data=&infile noprint;
class &xvar;

```

```

        model &yvar= &xvar;
        LSMEANS &xvar / out=orig;
        weight fwgt;
&by by region;
run;

data orig2;
set orig;
drop _name_ stderr;
array meanss(2);
do i=1 to 2;
    if _n_ = i then do;
        meanss(i)=lsmean;
        retain;
    end;
end;

if _n_=2 then do;
    dff1_2= meanss1-meanss2;
    drop lsmean i meanss1- meanss2;
    output;
end;
run;

proc transpose data=orig2 out=origest prefix=bhat name=beta;
var dff1_2;
run;

%let L=%eval (&bssz/10);

%do k=1 %to 10;
    %let j=%eval (1+((&k-1)*&L));
    %let kL=%eval (&k*&L);

    data poid;
        set &infile;
        keep bsw&j -bsw&kL &yvar &xvar &kp;
    run;

    %do i=&j %to &kL;

proc glm data=poid noprint;
class &xvar;
model &yvar= &xvar;
LSMEANS &xvar / out=betas;
weight bsw&i;
&by by region;
run;

data betas2;
set betas;
drop _name_ stderr;
array meanss(2);
do i=1 to 2;
    if _n_ = i then do;
        meanss(i)=lsmean;
        retain;
    end;
end;

if _n_=2 then do;
    dff1_2= meanss1-meanss2;
    drop lsmean i meanss1-meanss2;
    output;
end;
run;

proc transpose data=betas2 out=betat prefix=best name=beta;
var dff1_2;
run;

data betat;
set betat;
rename best1=best&i;
run;

%if (&i =1) %then %do;

data bsbeta;
set betat;
run;

%end;
%else %do;

data bsbeta;
merge bsbeta betat;

```

```

        &by region;
run;

%end;
%end;
%end;

data est;
merge origest bsbeta;
rename bhat1=bhat;
bs_var=&multi*((&bssz-1)*(var(of best1-best&bssz)))/&bssz;
bs_sd=sqrt(bs_var);
bs_cv=abs(round((bs_sd/bhat1)*100,.01));
ci l 95=bhat1-1.96*bs_sd;
ci u 95=bhat1+1.96*bs_sd;
drop best1-best&bssz;
run;

proc append data=est base=bs_reg;
run;

%let printreg=1;
%let dep=&yvar;

%mend regglm;

.
.
.

*****
* Section 3: Programme principal *
*****

/* Lit le fichier principal */

/*****
*** À CAUSE DU GRAND NOMBRE D'OBSERVATIONS, SEULES LES VARIABLES ***
*** REQUISES POUR FAIRE LES ANALYSES ET LES VARIABLES ***
*** "REALUKEY" ET "PERSONID" DEVRAIENT ÊTRE CONSERVÉES DÉSORMAIS. ***
*****/

data nphs (index=(id=(realukey personid)));
set &Mfile (keep= realukey personid living doc);
run;

/* Lit les poids bootstrap */
/* FWGT correspond à WT56, WT66, WT66_N, WT66LF ou WT66LP selon le fichier à
analyser */

data bsamp (index=(id=(realukey personid)));
set &bsamp;
keep fwgt realukey personid bsw1-bsw&B.00;
run;

/* Apparie le fichier principal et les poids bootstrap */

&by data bs_nphs (index=(region));
&wo data bs_nphs;
merge nphs (in=in1) bsamp (in=in2);
by realukey personid;
if in1;
run;

proc datasets library=work;
delete nphs bsamp;
run;

```



```

/*****
*** Section 4: Énoncés pour soumettre les macros
*****/

.
.

/*****/
/*
*/
/* Modèle Linéaire Généralisé */
/*
*/
/*****/

/*****/
*** N.B. La macro regglm doit être modifiée pour répondre aux besoins de
*** l'utilisateur. La macro regglm définie dans ce programme est un
*** exemple d'une analyse qui a été faite en utilisant ce programme.
***
*** Pour soumettre la macro regglm, l'énoncé suivant est utilisé:
*** %regglm(infile, yvar, xvar, bssz=, mul ti=);
*** ou      infile : bs_nphs
***          yvar  : la variable dépendante
***          xvar  : les variables indépendantes. Les variable doivent
***                  être inscrites comme elles le seraient dans l'énon-
***                  cé d'un modèle (sans virgules entre elles).
***          bssz= : bssz=&B.00
***          mul ti= : mul ti=&R
***
*** Nota: Si la macro regglm est adaptée pour calculer des différences entre
*** des moyennes, la façon de calculer l'intervalle de confiance doit
*** tenir compte des comparaisons multiples. La valeur de la loi
*** normale (Z) utilisée dans le calcul de l'intervalle de confiance
*** doit être corrigée à la section 2, Déclaration des macros, en
*** utilisant, par exemple, l'approche de Bonferroni pour les
*** comparaisons multiples.
*****/

* %regglm(bs_nphs, YVAR, XVAR, bssz=&B.00, mul ti=&R);
* %regglm(bs_nphs, doc, l i v i n g, bssz=&B.00, mul ti=&R);

/* Détruit les fichiers SAS de la macro regglm */
proc datasets library=work;
  delete betas betat bsbeta origest;
run;

%prntreg;

.
.

data time;
  set time;
  format stop datetime16.;
  stop=datetime();
  output;
run;

proc print data=time;
  title 'Temps requis pour exécuter le programme';
run;

/* Fin du programme SAS Bootvar */

```

(L'exécution de ce programme a pris environ 16 minutes.)

Les exemples suivant présentent les résultats des analyses réalisées à l'aide des programmes de l'exemple 2. L'exemple 3A présente les résultats des totaux et des rapports. Par exemple, si on s'intéresse au rapport du nombre d'hommes diabétiques âgés de 45 ans et plus au nombre total d'hommes âgés de 45 ans et plus, en Ontario, on se réfère à l'observation 41. La région 35 correspond à la province de l'Ontario (voir le dictionnaire de données du *Guide du fichier de microdonnées à grande diffusion de l'ENSP, 1996-1997* pour les codes associés à chacune des provinces) et la variable Type indique le type d'analyse, dans ce cas-

ci, un rapport. On retrouve au numérateur du rapport la variable *mdia45* (VAR1) et au dénominateur la variable *mage45* (VAR2). L'estimation du rapport est de 7,46% (YHAT) avec un écart-type de 1,04 (BS_SD) et un coefficient de variation de 14,01 (BS_CV). L'intervalle de confiance à 95% pour cette estimation est (5,41%, 9,50%) (CIL95, CIU95). À la suite des résultats, on peut voir que l'exécution de ce programme a débuté à 14:53:11 et s'est terminé à 14:56:03.

Les résultats de la régression logistique sont présentés à l'exemple 3B. Par exemple, le paramètre estimé pour la variable AGE85 est 2,88 (BHAT) et le rapport de cotes est de 17,89 (ODDS). La statistique de Wald et son seuil p associé pour ce paramètre (BHAT) sont de 81,26 (WALD) et $p=0,0000$ (PVALUE) respectivement. L'estimation de la variance et de l'écart-type du paramètre estimé sont de 0,102 (BS_VAR) et 0,320 (BS_SD) et le coefficient de variation est de 11,09 (BS_CV). Finalement, l'intervalle de confiance pour le rapport de cotes est (9,56, 33,50) (CIL95, CIU95). L'exécution de ce programme a débuté à 15:01:17 et s'est terminé à 15:21:01.

Les résultats du calcul de la différence entre les moyennes, calculée à l'aide de la procédure de modèle linéaire généralisé, sont présentés à l'exemple 3C. BETA indique le nom donné à la différence calculée entre le nombre moyen de visites chez le médecin pour les femmes vivant en couple, avec des enfants, et pour les femmes en famille monoparentale (DIFF1_2). La valeur de cette différence est de 0.85 (BHAT). La variance et l'écart-type du paramètre estimé sont de 1,46 (BS_VAR) et 1,21 (BS_SD) respectivement. Le coefficient de variation associé est de 142 (BS_CV). L'intervalle de confiance est donnée par CIL95 et CIU95 (-1,52, 3,22). L'exécution de ce programme a débuté à 12:14:29 et s'est terminé à 12:30:28.

Exemple 3: Résultats du programme BOOTVAR.SAS et temps requis pour l'exécution du programme Exemple 3A: Totaux et rapports

Estimation de la variance à l'aide du bootstrap 500: 01 pour des
Totaux et des Rapports

OBS	REGION	TYPE	VAR1	VAR2	YHAT	BS_SD	BS_CV	CI L95	CI U95
1	10	Total	di ab	None	1227.40	622.52	50.72	7.26	2447.55
2	10	Total	di ab2544	None	257.65	258.76	100.43	-249.51	764.81
3	10	Total	mdi a45	None	969.76	579.77	59.79	-166.60	2106.11
4	10	Total	fdi a45	None	0.00	0.00	.	0.00	0.00
5	10	Ratio	di ab	total	5.14	2.61	50.72	0.03	10.24
6	10	Ratio	mdi a45	mage45	22.32	13.34	59.79	-3.83	48.48
7	10	Ratio	di ab45	age45	10.91	6.52	59.79	-1.87	23.69
8	11	Total	di ab	None	164.33	84.29	51.29	-0.88	329.55
9	11	Total	di ab2544	None	58.45	55.86	95.56	-51.02	167.93
10	11	Total	mdi a45	None	50.64	33.75	66.64	-15.51	116.79
11	11	Total	fdi a45	None	55.24	48.64	88.06	-40.10	150.58
12	11	Ratio	di ab	total	2.92	1.50	51.29	-0.02	5.85
13	11	Ratio	mdi a45	mage45	4.68	3.12	66.64	-1.43	10.80
14	11	Ratio	di ab45	age45	4.69	2.70	57.52	-0.60	9.97
15	12	Total	di ab	None	1895.08	804.13	42.43	318.99	3471.16
16	12	Total	di ab2544	None	0.00	0.00	.	0.00	0.00
17	12	Total	mdi a45	None	1227.79	705.20	57.44	-154.39	2609.98
18	12	Total	fdi a45	None	348.10	213.05	61.20	-69.47	765.68
19	12	Ratio	di ab	total	4.89	2.08	42.43	0.82	8.96
20	12	Ratio	mdi a45	mage45	16.46	9.45	57.44	-2.07	34.98
21	12	Ratio	di ab45	age45	9.95	4.61	46.38	0.90	18.99
22	13	Total	di ab	None	1955.13	777.12	39.75	431.98	3478.28
23	13	Total	di ab2544	None	776.57	423.85	54.58	-54.17	1607.32
24	13	Total	mdi a45	None	0.00	0.00	.	0.00	0.00
25	13	Total	fdi a45	None	1178.56	658.46	55.87	-112.03	2469.15
26	13	Ratio	di ab	total	6.19	2.46	39.75	1.37	11.01
27	13	Ratio	mdi a45	mage45	0.00	0.00	.	0.00	0.00
28	13	Ratio	di ab45	age45	9.33	5.21	55.87	-0.89	19.55
29	24	Total	di ab	None	19276.19	5435.85	28.20	8621.92	29930.45
30	24	Total	di ab2544	None	2911.08	2764.33	94.96	-2507.00	8329.16
31	24	Total	mdi a45	None	7159.76	2759.29	38.54	1751.56	12567.97
32	24	Total	fdi a45	None	9205.35	3502.50	38.05	2340.45	16070.24
33	24	Ratio	di ab	total	6.29	1.77	28.20	2.81	9.76
34	24	Ratio	mdi a45	mage45	12.27	4.73	38.54	3.00	21.53
35	24	Ratio	di ab45	age45	13.12	3.68	28.04	5.91	20.33
36	35	Total	di ab	None	13836.27	1356.42	9.80	11177.70	16494.85
37	35	Total	di ab2544	None	1484.06	482.71	32.53	537.96	2430.17
38	35	Total	mdi a45	None	6483.70	908.26	14.01	4703.50	8263.89
39	35	Total	fdi a45	None	5699.69	877.77	15.40	3979.27	7420.11
40	35	Ratio	di ab	total	2.97	0.29	9.80	2.40	3.54
41	35	Ratio	mdi a45	mage45	7.46	1.04	14.01	5.41	9.50
42	35	Ratio	di ab45	age45	6.62	0.69	10.40	5.27	7.97
43	46	Total	di ab	None	2674.96	926.30	34.63	859.40	4490.51
44	46	Total	di ab2544	None	933.12	573.19	61.43	-190.34	2056.57
45	46	Total	mdi a45	None	409.21	166.98	40.81	81.92	736.50
46	46	Total	fdi a45	None	1271.23	689.66	54.25	-80.51	2622.97
47	46	Ratio	di ab	total	5.94	2.06	34.61	1.91	9.97
48	46	Ratio	mdi a45	mage45	4.74	1.93	40.81	0.95	8.53
49	46	Ratio	di ab45	age45	9.18	3.87	42.13	1.60	16.76
50	47	Total	di ab	None	1326.21	453.64	34.21	437.08	2215.34
51	47	Total	di ab2544	None	248.73	249.03	100.12	-239.37	736.84
52	47	Total	mdi a45	None	237.72	330.29	138.94	-409.64	885.09
53	47	Total	fdi a45	None	839.76	179.77	21.41	487.41	1192.10
54	47	Ratio	di ab	total	3.31	1.13	34.21	1.09	5.53
55	47	Ratio	mdi a45	mage45	3.04	4.22	138.94	-5.24	11.31
56	47	Ratio	di ab45	age45	6.58	2.32	35.27	2.03	11.13
57	48	Total	di ab	None	2929.76	579.13	19.77	1794.66	4064.85
58	48	Total	di ab2544	None	306.84	132.56	43.20	47.03	566.65
59	48	Total	mdi a45	None	1253.02	359.26	28.67	548.88	1957.17
60	48	Total	fdi a45	None	1333.66	389.57	29.21	570.09	2097.22
61	48	Ratio	di ab	total	2.61	0.52	19.76	1.60	3.63
62	48	Ratio	mdi a45	mage45	6.48	1.86	28.67	2.84	10.13
63	48	Ratio	di ab45	age45	6.54	1.42	21.73	3.75	9.32
64	59	Total	di ab	None	1958.76	1220.79	62.32	-433.99	4351.51
65	59	Total	di ab2544	None	0.00	0.00	.	0.00	0.00
66	59	Total	mdi a45	None	1041.07	771.51	74.11	-471.09	2553.23
67	59	Total	fdi a45	None	917.69	919.66	100.21	-884.84	2720.21
68	59	Ratio	di ab	total	1.23	0.76	62.32	-0.27	2.72
69	59	Ratio	mdi a45	mage45	3.34	2.47	74.11	-1.51	8.18
70	59	Ratio	di ab45	age45	3.04	1.90	62.32	-0.67	6.76

Temps requis pour exécuter le programme

OBS	START	STOP
1	20MAY99: 14: 53: 11	20MAY99: 14: 56: 03

Exemple 3B: Régression logistique

```

Estimation de la variance à l'aide du bootstrap 500:01 pour des
Régressions Logistiques
Variable dépendante: cccodany

OBS   BETA      BHAT      ODDS      WALD      PVALUE      BS_VAR      BS_SD      BS_CV      CI L95      CI U95
1     INTERCEP  -0.64096    0.5268    69.945    0.00000    0.00587    0.07664    11.96    0.45331    0.6122
2     NONPROXY   0.05279    1.0542     0.920    0.33746    0.00303    0.05504    104.25    0.94641    1.1743
3     AGE1820    0.39194    1.4798     6.856    0.00883    0.02241    0.14968    38.19    1.10358    1.9844
4     AGE2124    0.25685    1.2929     4.997    0.02540    0.01320    0.11491    44.74    1.03214    1.6194
5     AGE2534    0.62725    1.8725    40.812    0.00000    0.00964    0.09819    15.65    1.54467    2.2698
6     AGE3544    0.38282    1.4664    11.888    0.00056    0.01233    0.11103    29.00    1.17963    1.8229
7     AGE4554    1.07347    2.9255    91.912    0.00000    0.01254    0.11197    10.43    2.34905    3.6435
8     AGE5564    1.11707    3.0559    113.558    0.00000    0.01099    0.10483     9.38    2.48832    3.7529
9     AGE6574    1.76754    5.8564    222.052    0.00000    0.01407    0.11862     6.71    4.64156    7.3892
10    AGE7584    1.86193    6.4361    116.850    0.00000    0.02967    0.17225     9.25    4.59204    9.0207
11    AGE85     2.88430    17.8910    81.261    0.00000    0.10238    0.31996    11.09    9.55603    33.4959
12    FEMALE     0.51649    1.6761    99.007    0.00000    0.00269    0.05191    10.05    1.51399    1.8556

Temps requis pour exécuter le programme

OBS      START      STOP
1      20MAY99: 15: 01: 17      20MAY99: 15: 21: 01

```

Exemple 3C: Modèle linéaire généralisé

```

Estimation de la variance à l'aide du bootstrap 500:01 pour une
Régression
Variable dépendante: doc

OBS   BETA      BHAT      BS_VAR      BS_SD      BS_CV      CI L95      CI U95
1     DIFF1_2    0.85002    1.45689    1.20702    142      -1.51573    3.21578

Temps requis pour exécuter le programme

OBS      START      STOP
1      01JUN99: 12: 14: 29      01JUN99: 12: 30: 28

```

12.4.3.3 Modifications au programme aux fins de test

Comme il a été mentionné précédemment, l'exécution du programme peut prendre un certain temps dans le cas des analyses de régression. Il est possible de réduire le nombre de poids bootstrap utilisés afin de *tester* le programme (**Cependant, pour obtenir les estimations de variance, il est indispensable d'utiliser l'ensemble complet des 500 poids bootstrap fournis**). Par exemple, dans le cas d'une régression logistique, il suffit de spécifier le nombre de fois où la régression sera calculée à l'aide de poids bootstrap (par exemple, 2 fois au lieu de 500 fois) dans la routine "logreg" à la section 2 du programme. L'utilisateur doit ajouter trois énoncés comme montré en *italiques* dans l'exemple 4. La valeur de kL représente le nombre de fois où la régression sera calculée (dans ce cas-ci, 2 fois). Les deux autres paramètres, k et j, doivent égaier 10 et 1 respectivement pour le test. À noter que ces modifications s'ajoutent aux paramètres à spécifier comme montré dans l'exemple 2B. Encore une fois, l'estimation de la variance

On doit également spécifier le nombre de poids bootstrap à lire à la section 3. L'exemple suivant indique où le changement doit être fait. La partie en *italique* doit remplacer la partie entre parenthèses. Le même changement peut être fait pour les routines “regress” et “regglm”.

[illegible]

```

%end;
%else %do;

    data bsbeta;
        merge bsbeta betat;
        &by by region;
    run;

%end;
%end;
%end;

data bs_reglg;
    merge origest bsbeta;
    rename bhat1=bhat;
    bs_var=&multi*((&bssz-1)*(var(of best1-best&bssz)))/&bssz;
    bs_sd=sqrt(bs_var);
    bs_cv=abs(round((bs_sd/bhat1)*100,.01));
    wald=(bhat1/bs_sd)*(bhat1/bs_sd);
    pvalue=1-probchi(wald,1);
    lo95=bhat1-1.96*bs_sd;
    hi95=bhat1+1.96*bs_sd;
    odds=exp(bhat1);
    ci195=exp(lo95);
    ciu95=exp(hi95);
    ydep="&yvar";
    drop best1-best&bssz;
run;

%let printlog=1;
%let dep=&yvar;

%mend logreg;

.
.
.

*****
* Section 3: Programme principal *
*****

/* Lit le fichier principal */

/*****
*** A CAUSE DU GRAND NOMBRE D'OBSERVATIONS, SEULES LES VARIABLES ***
*** REQUISES POUR FAIRE LES ANALYSES ET LES VARIABLES. ***
*** "REALUKEY" ET "PERSONID" DEVRAIENT ÊTRE CONSERVÉES DÉSORMAIS. ***
*****/

data nphs (index=(id=(realukey personid)));
    set &Mfile (keep= realukey personid ccc6dany nonproxy age1820 age2124 age2534
        age3544 age4554 age5564 age6574 age7584 age85 female);
run;

/* Lit les poids de bootstrap */
/* FWGT est le même poids que WT56, WT66, WT66_N, WT66LF ou WT66LP selon le
fichier analysé */

data bsamp (index=(id=(realukey personid)));
    set &bsamp;
    keep fwgt realukey personid bsw1-bsw2; (au lieu de bsw1-bsw&B.00;)
run;

```

(Le programme de l'exemple 2B a été testé de cette façon et l'exécution a pris environ 1 minute.)

Programme BOOTVAR.SAS

AVERTISSEMENT

* Le gouvernement du Canada (Statistique Canada) possède tous les droits de propriété
* intellectuelle (y compris le droit d'auteur) relatifs à ce logiciel. Sous réserve
* des modalités décrites ci-dessous, vous obtenez une licence non exclusive et non
* transmissible d'utilisation de ce logiciel.
*
* Ce logiciel vous est fourni « tel quel », et le titulaire ne donne aucune garantie
* expresse ou implicite, ce qui comprend les garanties de qualité marchande et de
* convenance à une fin particulière mais ne se limite pas à ces garanties. En aucune
* circonstance le titulaire ne sera responsable de dommages indirects, particuliers ou
* consécutifs, ou d'autres dommages semblables. Ce contrat sera automatiquement résilié,
* sans préavis, si vous omettez de respecter une quelconque des modalités du contrat.;

```

/*****
/* Date: 1er décembre 1997
/* Modi fi é: 30 octobre 1998
*****/

```

```

*****
/****
/****                               Programme SAS BOOTVAR
/****
/**** Ce programme calcule des estimations de variance pour différents types ****/
/**** d'estimations à l'aide des poids bootstrap. A l'aide de macros SAS, ****/
/**** ce programme peut calculer des estimations de variance pour des totaux, ****/
/**** des rapports, et des différences de rapports. Il peut aussi calculer des ****/
/**** estimations de variance pour des paramètres de régressions linéaires, ****/
/**** logistiques ou d'un modèle linéaire généralisé. Ce programme peut aussi ****/
/**** être modifié pour s'adapter à d'autres types d'analyse. ****/
/****
/**** Ce programme est divisé en 4 sections, décrites plus bas. Tout au long ****/
/**** du programme, les changements que l'utilisateur doit apporter sont indi- ****/
/**** qués. ****/
/****
/**** Section 1: Déclaration des variables macros ****/
/**** C'est à cet endroit qu'on définit les variables qui seront ****/
/**** utilisées tout au long du programme. Certains changements ****/
/**** doivent être apportés par l'utilisateur. ****/
/****
/**** Section 2: Déclaration des macros ****/
/**** C'est à cet endroit qu'est définie la portion du programme qui ****/
/**** calcule les estimations de variance. Aucun changement n'est ****/
/**** requis de la part de l'utilisateur dans cette section (à l'ex- ****/
/**** ception de la macro pour un modèle linéaire généralisé (REGGLM). ***/
/**** Cette macro doit être modifiée selon les besoins de ****/
/**** l'utilisateur). ****/
/****
/**** Section 3: Programme principal ****/
/**** C'est à cet endroit qu'on lit les fichiers contenant les ****/
/**** variables à analyser et les poids bootstrap. Les variables ****/
/**** à utiliser dans les analyses doivent être définies par l'utili- ****/
/**** sateur avant d'utiliser ce programme. ****/
/****
/**** Section 4: Énoncés pour soumettre les macros ****/
/**** C'est à cet endroit que l'utilisateur doit indiqué quelle(s) ****/
/**** macro(s) il/elle désire soumettre. À ce stade, l'utilisateur ****/
/**** peut soumettre une macro une fois, une macro plusieurs fois, ****/
/**** ou plusieurs macros. Étant donné le grand nombre d'observa- ****/
/**** tions à analyser, il est recommandé de limiter le ****/
/**** nombre de macros qui sont soumises, en particulier dans le cas ****/
/**** de régressions. ****/
/****
/**** Avant d'utiliser ce programme, l'utilisateur doit préparer un fichier ****/
/**** d'analyse qui contiendra les variables à analyser. Ce fichier ne devrait ****/
/**** contenir que les enregistrements pour lesquels l'analyse est nécessaire ****/
/**** (p.e. si l'analyse est faite pour le groupe d'âge 12 ans et plus, seuls ****/
/**** les enregistrements des individus de 12 ans et plus devraient être inclus ****/
/**** dans le fichier). Une variable appelée REGION doit aussi être créée si ****/
/**** les analyses sont faites à un niveau régional. Par exemple, si l'analyse ****/
/**** est faite au niveau provincial, la variable REGION prendra la même valeur ****/
/**** que la variable qui contient la province sur le fichier de microdonnées ****/
/**** (p.e. PRC6_CUR). Si l'analyse est faite au niveau des régions de santé, ****/
/**** la variable REGION prendra la valeur de la région de santé pour chaque ****/
/**** enregistrement. ****/
/****

```

```

/**** N.B. Pour calculer un total, un rapport, ou une différence entre deux ****/
/**** rapports, les estimations sont obtenues en sommant le poids des enregis- ****/
/**** trements qui ont la caractéristique d'intérêt. Donc, une variable indi- ****/
/**** catrice doit être créée pour chaque variable à analyser. Elle prendra la ****/
/**** valeur 1 lorsque l'enregistrement a la caractéristique d'intérêt et 0 ****/
/**** sinon. ****/
/**** ****/

```

options ps=48 ls=120 ;

```

/**** ****/
/**** Section 1: Déclaration des variables macros ****/
/**** ****/

```

```
%let R=01; /* paramètre du bootstrap moyen (fixe) */
```

```

/**** ****/
/**** L'UTILISATEUR DOIT INDiquer LE NOMBRE DE POIDS BOOTSTRAP À UTILISER, ****/
/**** DIVISÉ PAR 100 ****/
/**** %let B=5; pour les fichiers de poids bootstrap B1H35, B5H356, B5H356A, ****/
/**** B5H356C, B5LONGF et B5LONGP ****/
/**** ****/

```

```
%let B=5; /* nombre de poids Bootstrap Bx100 */
```

```

/**** ****/
/**** L'UTILISATEUR DOIT INDiquer LE RÉPERTOIRE CONTENANT LE FICHIER ****/
/**** D'ANALYSE (in1) ET LE FICHIER CONTENANT LES POIDS BOOTSTRAP ****/
/**** (in2). LES FICHIERS DE SORTIES SERONT SAUVEGARDES DANS LE ****/
/**** RÉPERTOIRE SPÉCIFIÉ POUR in3. ****/
/**** ****/

```

```

libname in1 "c:\data";
libname in2 "c:\bootstrp";
libname in3 "c:\output";

```

```

/**** ****/
/**** L'UTILISATEUR DOIT INDiquer LE NOM DU FICHIER SAS CONTENANT LES VARIABLES ****/
/**** D'ANALYSE. ****/
/**** %let Mfile=xxx; ****/
/**** ****/

```

```
%let Mfile=in1.h356; /* Nom du fichier SAS principal */
```

```

/**** ****/
/**** L'UTILISATEUR DOIT INDiquer LE NOM DU FICHIER SAS CONTENANT LES POIDS ****/
/**** BOOTSTRAP. ****/
/**** %let bsamp=b5h35; (à utiliser avec le fichier H35.TXT) OU ****/
/**** %let bsamp=b5h356; (à utiliser avec le fichier H356.TXT) OU ****/
/**** %let bsamp=b5h356a; (à utiliser avec le fichier H356.TXT) OU ****/
/**** %let bsamp=b5h356c; (à utiliser avec le fichier H356.TXT) OU ****/
/**** %let bsamp=b5longp; (à utiliser avec le fichier LONGP.TXT) ****/
/**** %let bsamp=b5longf; (à utiliser avec le fichier LONGF.TXT) ****/
/**** ****/

```

```
%let bsamp=in2.b5h356; /* fichier SAS des poids bootstrap */
```

```

/**** ****/
/**** L'UTILISATEUR DOIT SPÉCIFIER LA VALEUR DES PARAMÈTRES by, ****/
/**** wo ET kp. CEUX-CI INDiquENT SI L'ANALYSE EST FAITE AU ****/
/**** NIVEAU RÉGIONAL OU POUR L'ENSEMBLE DU FICHIER. ****/
/**** ****/
/**** SI L'ANALYSE EST FAITE POUR L'ENSEMBLE DU FICHIER, ALORS ****/
/**** %let by=; ****/
/**** %let wo=; ****/
/**** %let kp=; ****/
/**** ****/
/**** SINON, SI L'ANALYSE EST FAITE AU NIVEAU RÉGIONAL, ALORS ****/
/**** %let by=; ****/
/**** %let wo=; ****/
/**** %let kp=regi on; ****/
/**** ****/

```



```

%let by=*;
%let wo=;
%let kp=;

data time;
  format start datetime16.;
  start=datetime();
  output;
run;

/*****
*** Section 2: Déclaration des macros ***
*****/

%let printtot=0;
%let printdif=0;
%let printreg=0;
%let printlog=0;

%global dep;

*****,

%macro total(infile, var, bssz=, multi=);
*****,

proc means data=&infile noprint;
  var fwgt bsw1-bsw&bssz;
  weight &var;
  &by by region;
  output out=ytot
         sum=yhat ybs1-ybs&bssz;
run;

data est;
  set ytot;
  length var1 $ 8;
  length var2 $ 8;
  length type $ 8;
  bs_var=&multi*((&bssz-1)*(var(of ybs1-ybs&bssz)))/&bssz;
  bs_sd=sqrt(bs_var);
  bs_cv=round((bs_sd/yhat)*100, .01);
  cil95=yhat-1.96*bs_sd;
  ciu95=yhat+1.96*bs_sd;
  var1="&var";
  var2="Aucune";
  type="Total ";
  drop ybs1-ybs&bssz _type_ _freq_;
run;

proc append data=est base=alltots;
run;

%let printtot=1;

%mend total;

*****,

%macro ratio(infile, var1, var2, bssz=, multi=);
*****,

proc means data=&infile noprint;
  var fwgt bsw1-bsw&bssz;
  weight &var1;
  &by by region;
  output out=ytot
         sum=yhat ybs1-ybs&bssz;
run;

proc means data=&infile noprint;
  var fwgt bsw1-bsw&bssz;

```

```

weight &var2;
&by by region;
output out=xtot
      sum=xhat xbs1-xbs&bssz;
run;

data est;
merge ytot xtot;
array ybs{&bssz};
array xbs{&bssz};
array rbs{&bssz};
length var1 $ 8;
length var2 $ 8;
length type $ 8;
yhat=((yhat/xhat)*100);
do i=1 to &bssz;
  rbs{i}=((ybs{i}/xbs{i})*100);
end;
bs_var=&mul ti *((&bssz-1)*(var(of rbs1-rbs&bssz)))/&bssz;
bs_sd=sqrt(bs_var);
bs_cv=round((bs_sd/yhat)*100, .01);
ci l95=yhat-1.96*bs_sd;
ci u95=yhat+1.96*bs_sd;
var1="&var1";
var2="&var2";
type="Rapport";
drop ybs1-ybs&bssz xbs1-xbs&bssz rbs1-rbs&bssz xhat i _type_ _freq_;
run;

proc append data=est base=all tots;
run;

%let printtot=1;

%mend ratio;

*****.

%macro diff_rat(infile, var1, var2, var3, var4, bssz=, mul ti =);
*****.

proc means data=&infile noprint;
var fwgt bsw1-bsw&bssz;
weight &var1;
&by by region;
output out=ytot
      sum=yhat ybs1-ybs&bssz;
run;

proc means data=&infile noprint;
var fwgt bsw1-bsw&bssz;
weight &var2;
&by by region;
output out=xtot
      sum=xhat xbs1-xbs&bssz;
run;

proc means data=&infile noprint;
var fwgt bsw1-bsw&bssz;
weight &var3;
&by by region;
output out=yytot
      sum=yyhat yybs1-yybs&bssz;
run;

proc means data=&infile noprint;
var fwgt bsw1-bsw&bssz;
weight &var4;
&by by region;
output out=xttot
      sum=xxhat xxbs1-xxbs&bssz;
run;

data est;
merge ytot xtot yytot xttot;

```

```

array ybs{&bssz};
array xbs{&bssz};
array yybs{&bssz};
array xxbs{&bssz};
array drbs{&bssz};
length var1 $ 8;
length var2 $ 8;
length var3 $ 8;
length var4 $ 8;
length type $ 10;
yhat=((yhat/xhat)-(yyhat/xxhat))*100;
do i=1 to &bssz;
  drbs{i}={((ybs{i}/xbs{i})-(yybs{i}/xxbs{i}))*100};
end;
bs_var=(&mul ti *((&bssz-1)*(var(of drbs1-drbs&bssz)))/&bssz);
bs_sd=sqrt(bs_var);
bs_cv=abs(round((bs_sd/yhat)*100,.01));
ci l95=yhat-1.96*bs_sd;
ci u95=yhat+1.96*bs_sd;
var1="&var1";
var2="&var2";
var3="&var3";
var4="&var4";
type="Di ff_Rapport";
drop ybs1-ybs&bssz xbs1-xbs&bssz yybs1-yybs&bssz xxbs1-xxbs&bssz drbs1-drbs&bssz
      xhat xxhat yyhat i _type_ _freq_;
run;

proc append data=est base=di ffrat;
run;

%let printdf=1;

%mend di ff_rat;

*****;

%macro regress(infile,yvar,xvar,bssz,multi=);
*****;

proc reg data=&infile outest=orig(keep=&kp intercept &xvar) noprint;
  model &yvar=&xvar;
  weight fwgt;
&by by region;
run;

proc transpose data=orig out=origest(drop=_label_) prefix=bhat name=beta;
  var intercept &xvar;
&by by region;
run;

%let L=%eval (&bssz/10);

%do k=1 %to 10;
  %let j=%eval (1+((&k-1)*&L));
  %let kL=%eval (&k*&L);

  data poid;
    set &infile;
    keep bsw&j -bsw&kL &yvar &xvar &kp;
  run;

  %do i=&j %to &kL;

    proc reg data=poid outest=betas(keep=&kp intercept &xvar) noprint;
      model &yvar=&xvar;
      weight bsw&i;
&by by region;
    run;

    proc transpose data=betas out=betat prefix=best name=beta;
      var intercept &xvar;
&by by region;
    run;

```

```

data betat;
  set betat;
  drop _label_;
  rename best1=best&i;
run;

%if (&i =1) %then %do;

  data bsbeta;
    set betat;
  run;

%end;
%else %do;

  data bsbeta;
    merge bsbeta betat;
&by    by region;
  run;

%end;
%end;
%end;

data est;
  merge orig est bsbeta;
  rename bhat1=bhat;
  bs_var=&mul ti *((&bssz-1)*(var(of best1-best&bssz)))/&bssz;
  bs_sd=sqrt(bs_var);
  bs_cv=abs(round((bs_sd/bhat1)*100, .01));
  ci l95=bhat1-1.96*bs_sd;
  ci u95=bhat1+1.96*bs_sd;
  ydep=&yvar;
  drop best1-best&bssz;
run;

proc append data=est base=bs_reg;
run;

%let printreg=1;
%let dep=&yvar;

%mend regress;

*****;

%macro logreg(infile, yvar, xvar, bssz=, mul ti=);

*****;
proc logistic data=&infile outest=orig(keep=&kp intercep &xvar) descending noprint;
  model &yvar=&xvar;
  &by by region;
  weight fwgt;
run;

proc transpose data=orig out=origest prefix=bhat name=beta;
  var intercep &xvar;
  &by by region;
run;

%let L=%eval (&bssz/10);

%do k=1 %to 10;
  %let j=%eval (1+((&k-1)*&L));
  %let kL=%eval (&k*&L);

  data poid;
    set &infile;
    keep bsw&j -bsw&kL &yvar &xvar &kp;
  run;

  %do i=&j %to &kL;

    proc logistic data=poid outest=betas(keep=&kp intercep &xvar) noprint descending;
      model &yvar=&xvar;
      &by by region;

```

```

weight bsw&i;
run;

proc transpose data=betas out=betat prefix=best name=beta;
var intercep &xvar;
&by by region;
run;

data betat;
set betat;
rename best1=best&i;
run;

%if (&i =1) %then %do;

data bsbeta;
set betat;
run;

%end;
%else %do;

data bsbeta;
merge bsbeta betat;
&by by region;
run;

%end;
%end;
%end;

data est;
merge origest bsbeta;
rename bhat1=bhat;
bs_var=&mul ti *((&bssz-1)*(var(of best1-best&bssz)))/&bssz;
bs_sd=sqrt(bs_var);
bs_cv=abs(round((bs_sd/bhat1)*100,.01));
wald=(bhat1/bs_sd)*(bhat1/bs_sd);
pvalue=1-probchi(wald,1);
lo95=bhat1-1.96*bs_sd;
hi95=bhat1+1.96*bs_sd;
odds=exp(bhat1);
ci lo95=exp(lo95);
ci u95=exp(hi95);
ydep="&yvar";
drop best1-best&bssz;
run;

proc append data=est base=bs_reglg;
run;

%let printlog=1;
%let dep=&yvar;

%mend logreg;

*****

%macro regglm(infile, yvar, xvar, bssz=, mul ti =);

*****
proc glm data=infile noprint;
class &xvar;
model &yvar= &xvar;
LSMEANS &xvar / out=orig;
weight fwgt;
&by by region;
run;

data orig2;
set orig;
drop _name_ stderr;
array meanss{3};
do i=1 to 3;
if _n_ = i then do;
meanss(i)=lsmear;

```

```

        retain;
    end;
end;

if _n_=3 then do;
    di ff1_2= meanss1-meanss2;
    di ff1_3= meanss1-meanss3;
    di ff2_3= meanss2-meanss3;
    drop smc6_2 lsmean i meanss1- meanss3;
    output;
end;
run;

proc transpose data=orig2 out=origest prefix=bhat name=beta;
var di ff1_2 di ff1_3 di ff2_3;
run;

%let L=%eval (&bssz/10);

%do k=1 %to 10;
    %let j=%eval (1+((&k-1)*&L));
    %let kL=%eval (&k*&L);

    data poids;
        set &infile;
        keep bsw&j -bsw&kL &yvar &xvar &kp;
    run;

    %do i=&j %to &kL;

    proc glm data=poids noprint;
        class &xvar;
        model &yvar= &xvar;
        LSMEANS &xvar / out=betas;
        weight bsw&i;
        &by by region;
    run;

    data betas2;
        set betas;
        drop _name_ stderr;
        array meanss{3};
        do i=1 to 3;
            if _n_ = i then do;
                meanss(i)=lsmean;
                retain;
            end;
        end;

        if _n_=3 then do;
            di ff1_2= meanss1-meanss2;
            di ff1_3= meanss1-meanss3;
            di ff2_3= meanss2-meanss3;
            drop smc6_2 lsmean i meanss1- meanss3;
            output;
        end;
    run;

    proc transpose data=betas2 out=betat prefix=best name=beta;
        var di ff1_2 di ff1_3 di ff2_3;
    run;

    data betat;
        set betat;
        rename best1=best&i;
    run;

    %if (&i =1) %then %do;

        data bsbeta;
            set betat;
        run;

    %end;
    %else %do;

        data bsbeta;

```

```

        merge bsbeta betat;
        &by by region;
run;

%end;
%end;
%end;

data est;
    merge origest bsbeta;
    rename bhat1=bhat;
    bs_var=&multi*((&bssz-1)*(var(of best1-best&bssz)))/&bssz;
    bs_sd=sqrt(bs_var);
    bs_cv=abs(round((bs_sd/bhat1)*100,.01));
    cil95=bhat1-2.39406*bs_sd;
    ciu95=bhat1+2.39406*bs_sd;
    drop best1-best&bssz;
run;

proc append data=est base=bs_reg;
run;

%let printreg=1;
%let dep=&yvar;

%mend regglm;

*****;

%macro prnttot;

*****;
%i f &prnttot=1 %then %do;

&by    proc sort data=alltots;
&by    by region;
&by    run;

/*****/
/* Imprime les résultats des macro total et ratio */
/*****/

proc print data=alltots;
    title "Estimation de la variance à l'aide du bootstrap &B.00: &R pour des";
    title2 "Totaux et des Rapports";
    var &kp type var1 var2 yhat bs_sd bs_cv cil95 ciu95;
    format yhat bs_sd cil95 ciu95 11.2;
run;
%end;

/*****/
/**** Oû: ****/
/**** type      : type d'estimation (Total ou Rapport) ****/
/**** var1 et var2 : variables utilisées pour calculer ****/
/****              les estimations. Pour un total, ****/
/****              var2=Aucune. ****/
/**** yhat      : estimation du paramètre (en % pour ****/
/****              un rapport) ****/
/**** bs_sd     : écart-type ****/
/**** bs_cv     : coefficient de variation ****/
/**** cil95     : borne inférieure de l'intervalle de ****/
/****              confiance à 95% ****/
/**** ciu95     : borne supérieure de l'intervalle de ****/
/****              confiance à 95% ****/
/*****/

%mend prnttot;

*****;

%macro prntdiff;

*****;
%i f &prntdiff=1 %then %do;

```

```

&by proc sort data=di ffrat;
&by by region;
&by run;

/*****
/* Imprime les résultats de la macro di ff_rat */
*****/

proc print data=di ffrat;
  title "Estimation de la variance à l'aide du bootstrap &B.00:&R pour des";
  title2 "Différences entre des Rapports";
  var &kp type var1 var2 var3 var4 yhat bs_sd bs_cv ci l95 ci u95;
  format yhat bs_sd ci l95 ci u95 11.2;
run;
%end;

/*****
*** Où: ***
*** type : type d'estimation (Di ff_Rapport) ***
*** var1, var2, ***
*** var3 et var4 : variables utilisées pour calculer ***
*** l'estimation ***
*** yhat : estimation (différence en %) ***
*** bs_sd : écart-type ***
*** bs_cv : coefficient de variation ***
*** ci l95 : borne inférieure de l'intervalle de ***
*** confiance à 95% ***
*** ci u95 : borne supérieure de l'intervalle de ***
*** confiance à 95% ***
*****/

%mend prntdi ff;

*****,

%macro prntreg;

*****,
%if &prntreg=1 %then %do;

/*****
/* Imprime les résultats des macros regress et regglm */
*****/

proc print data=bs_reg;
  title "Estimation de la variance à l'aide du bootstrap &B.00:&R pour une";
  title2 "Régression";
  title3 "Variable dépendante: &dep";
  var &kp beta bhat bs_var bs_sd bs_cv ci l95 ci u95;
run;
%end;

/*****
*** Où: ***
*** beta : paramètre à estimer ***
*** bhat : estimation du paramètre ***
*** bsvar : variance de l'estimation du paramètre ***
*** bs_sd : écart-type de l'estimation du ***
*** paramètre ***
*** bs_cv : coefficient de variation de l'esti- ***
*** mation du paramètre ***
*** ci l95 : borne inférieure de l'intervalle de ***
*** confiance à 95% ***
*** ci u95 : borne supérieure de l'intervalle de ***
*** confiance à 95% ***
*****/

%mend prntreg;

*****,

%macro prntlog;

*****,

```



```

%if &printlog=1 %then %do;

/*****
/* Imprime les résultats de la macro logreg */
*****/

proc print data=bs_reglg;
  title "Estimation de la variance à l'aide du bootstrap &B.00:&R pour des";
  title2 "Régressions logistiques";
  title3 "Variable dépendante: &dep";
  var &kp beta bhat odds wald pvalue bs_var bs_sd bs_cv ci l95 ci u95;
run;
%end;

/*****
*** Où: ***
*** beta : paramètre à estimer ***
*** bhat : estimation du paramètre ***
*** odds : rapport de cotes ***
*** wald : statistique de Wald ***
*** pvalue : valeur p de la statistique de Wald ***
*** bsvar : variance de l'estimation du paramètre ***
*** bs_sd : écart-type de l'estimation du paramètre ***
*** bs_cv : coefficient de variation de l'estimation du paramètre ***
*** ci l95 : borne inférieure de l'intervalle de confiance à 95% du rapport de cotes ***
*** ci u95 : borne supérieure de l'intervalle de confiance à 95% du rapport de cotes ***
*****/

%mend prntlog;

*****
* Section 3: Programme principal
*****

/* Lire le fichier principal */

/*****
*** À CAUSE DU GRAND NOMBRE D'OBSERVATIONS, SEULES LES VARIABLES ***
*** REQUISES POUR FAIRE LES ANALYSES ET LES VARIABLES ***
*** "REALUKEY" ET "PERSONID" DEVRAIENT ÊTRE CONSERVÉES DÉSORMAIS. ***
*****/

data nphs (index=(id=(realukey personid)));
  set &Mfile (keep= realukey personid + VARIABLES À ANALYSER);
run;

/* Lire les poids bootstrap */
/* FWGT correspond à WT56, WT66, WT66_N, WT66LF ou WT66LP selon le fichier à analyser */

data bsamp (index=(id=(realukey personid)));
  set &bsamp;
  keep fwgt realukey personid bsw1-bsw&B.00;
run;

/* Appareiller le fichier principal et les poids bootstrap */

&by data bs_nphs (index=(region));
&wo data bs_nphs;
  merge nphs (in=in1) bsamp (in=in2);
  by realukey personid;
  if in1;
run;

proc datasets library=work;
  delete nphs bsamp;
run;

```

```

/***** Section 4: Énoncés pour soumettre les macros *****/
/*****

/***** Les estimations de variance sont calculées à l'aide des macros SAS définies *****/
/***** plus tôt dans le programme. Ces macros SAS peuvent être soumises durant *****/
/***** l'exécution du programme, selon les besoins de l'utilisateur. Pour *****/
/***** soumettre une macro, il suffit de spécifier un énoncé indiquant le nom de *****/
/***** la macro à soumettre et les paramètres à utiliser. Dans ce programme, les *****/
/***** paramètres indiquent quel fichier doit être lu, quelles variables seront *****/
/***** utilisées, combien de poids bootstrap seront utilisés, et combien *****/
/***** d'échantillons de bootstrap ont été choisis. Le nom du fichier à lire est *****/
/***** toujours le même (BS_NPHS) et les deux derniers paramètres ont déjà été *****/
/***** définis au début du programme. L'utilisateur n'a donc qu'à spécifier les *****/
/***** variables à analyser. *****/
/*****
/***** Chaque soumission d'une macro donnera une estimation de la variance d'un *****/
/***** seul paramètre à estimer. Si plus d'un paramètre et sa variance doivent *****/
/***** être calculés, il faut alors soumettre la macro à plusieurs reprises. *****/
/*****
/***** Un exemple d'un énoncé pour soumettre une macro apparaît dans le programme *****/
/***** à l'endroit où l'utilisateur doit spécifier cet énoncé. L'exemple est mis *****/
/***** en commentaire dans le programme. L'énoncé utilisé par l'utilisateur sera *****/
/***** identique à l'exemple, à l'exception des variables à analyser. *****/
/*****

/*****/
/* */
/* Total */
/* */
/*****/

proc datasets library=work; /* initialise all tots */
delete all tots;
run;

/*****/
/***** Pour soumettre la macro total, l'énoncé suivant est utilisé: *****/
/***** %total(infile, var, bssz=, mul ti=); *****/
/***** ou      infile : bs_nphs *****/
/***** var      : la variable pour laquelle un total doit être *****/
/***** calculé *****/
/***** bssz=    : bssz=&B.00 *****/
/***** mul ti=  : mul ti=&R *****/
/*****/

* %total(bs_nphs, VAR, bssz=&B.00, mul ti=&R);

/* Détruit les fichiers SAS de la macro total */

proc datasets library=work;
delete ytot est;
run;

/*****/
/* */
/* Rapport */
/* */
/*****/

/*****/
/***** Pour soumettre la macro ratio, l'énoncé suivant est utilisé: *****/
/***** %ratio(infile, var1, var2, bssz=, mul ti=); *****/
/***** ou      infile : bs_nphs *****/
/***** var1    : la variable au numérateur du rapport *****/
/***** var2    : la variable au dénominateur du rapport *****/
/***** bssz=    : bssz=&B.00 *****/
/***** mul ti=  : mul ti=&R *****/
/*****/

* %ratio(bs_nphs, VAR1, VAR2, bssz=&B.00, mul ti=&R);

```

```

/* Détruit les fichiers SAS de la macro ratio */
proc datasets library=work;
  delete ytot xtot est;
run;

%prnttot;

/*****
*** Les résultats des macro total et ratio ***
*** peuvent être sauvegardés dans un fichier permanent ***
*****/

*   data in3.ZZZ;
*   set alltots;
*   run;

/*****
/*
/* Différence entre des Rapports
/*
*****/

proc datasets library=work; /* initialise di ffrat */
  delete di ffrat;
run;

/*****
*** Pour soumettre la macro di ff_rat, l'énoncé suivant est utilisé: ***
*** %di ff_rat(infile, var1, var2, var3, var4, bssz=, mul ti=); ***
*** ou      infile : bs_nphs ***
***      var1 : la variable au numérateur du 1er rapport ***
***      var2 : la variable au dénominateur du 1er rapport ***
***      var3 : la variable au numérateur du 2e rapport ***
***      var4 : la variable au dénominateur du 2e rapport ***
***      bssz= : bssz=&B.00 ***
***      mul ti= : mul ti=&R ***
***
*** Nota: L'intervalle de confiance est calculé pour une analyse faisant ***
*** une seule comparaison de rapports. Si des comparaisons multiples sont ***
*** faites, la façon de calculer l'intervalle de confiance doit tenir ***
*** compte de cette situation. Pour cette raison, dans le cas de compari- ***
*** sons multiples, la valeur de la loi normale (Z) utilisée dans le ***
*** calcul de l'intervalle de confiance doit être corrigée à la section 2, ***
*** Déclaration des macros, en utilisant, par exemple, l'approche de ***
*** Bonferroni pour les comparaisons multiples. ***
*****/

*   %di ff_rat(bs_nphs, VAR1, VAR2, VAR3, VAR4, bssz=&B.00, mul ti=&R);

/* Détruit les fichiers SAS de la macro di ff_rat */
proc datasets library=work;
  delete ytot xtot yytot xxtot est;
run;

%prntdi ff;

/*****
*** Les résultats de la macro di ff_rat peuvent ***
*** être sauvegardés dans un fichier permanent ***
*****/

*   data in3.ZZZ;
*   set di ffrat;
*   run;

/*****
/*
/* Régression
/*
*****/

proc datasets library=work; /* initialise bs_reg */
  delete bs_reg;

```

```

run;

/*****
*** Pour soumettre la macro regress, l'énoncé suivant est utilisé:
*** %regress(infile, yvar, xvar, bssz=, mul ti=);
*** ou      infile : bs_nphs
*** yvar    : la variable dépendante
*** xvar    : les variables indépendantes. Les variables doivent
***         être inscrites comme elles le seraient dans l'énon-
***         cé d'un modèle (sans virgules entre elles).
*** bssz=   : bssz=&B.00
*** mul ti= : mul ti=&R
***
*****/

* %regress(bs_nphs, YVAR, XVAR, bssz=&B.00, mul ti=&R);

/* Détruit les fichiers SAS de la macro regress */
proc datasets library=work;
delete betas betat bsbeta origest;
run;

/*****
*/
/* Régression Logistique */
/*
*****/

proc datasets library=work; /* initialise bs_reglg */
delete bs_reglg;
run;

/*****
*** Pour soumettre la macro logreg, l'énoncé suivant est utilisé:
*** %logreg(infile, yvar, xvar, bssz=, mul ti=);
*** ou      infile : bs_nphs
*** yvar    : la variable dépendante
*** xvar    : les variables indépendantes. Les variables doivent
***         être inscrites comme elles le seraient dans l'énon-
***         cé d'un modèle (sans virgules entre elles).
*** bssz=   : bssz=&B.00
*** mul ti= : mul ti=&R
***
*****/

* %logreg(bs_nphs, YVAR, XVAR, bssz=&B.00, mul ti=&R);

/* Détruit les fichiers SAS de la macro logreg */
proc datasets library=work;
delete betas betat bsbeta origest;
run;

/*****
*/
/* Modèle Linéaire Généralisé */
/*
*****/

/*****
*** N.B. La macro regglm doit être modifiée pour répondre aux besoins de
*** l'utilisateur. La macro regglm définie dans ce programme est un
*** exemple d'une analyse qui a été faite en utilisant ce programme.
***
*** Pour soumettre la macro regglm, l'énoncé suivant est utilisé:
*** %regglm(infile, yvar, xvar, bssz=, mul ti=);
*** ou      infile : bs_nphs
*** yvar    : la variable dépendante
*** xvar    : les variables indépendantes. Les variables doivent
***         être inscrites comme elles le seraient dans l'énon-
***         cé d'un modèle (sans virgules entre elles).
*** bssz=   : bssz=&B.00
*** mul ti= : mul ti=&R
***
*****/

/* Nota: Si la macro regglm est adaptée pour calculer des différences entre */

```

```

/**** des moyennes, la façon de calculer l'intervalle de confiance doit ****/
/**** tenir compte des comparaisons multiples. La valeur de la loi ****/
/**** normale (Z) utilisée dans le calcul de l'intervalle de confiance ****/
/**** doit être corrigée à la section 2, Déclaration des macros, en ****/
/**** utilisant, par exemple, l'approche de Bonferroni pour les ****/
/**** comparaisons multiples. ****/
/**** ****/
/*****

* %regglm(bs_nphs, YVAR, XVAR, bssz=&B.00, mul ti =&R);

/* Détruit les fichiers SAS de la macro regglm */
proc datasets library=work;
  delete betas betat bsbeta origest;
run;

%prntreg;

/*****
/**** Les résultats des macros regress et regglm ****/
/**** peuvent être sauvegardés dans un fichier permanent ****/
/**** ****/

* data in3.ZZZ;
* set bs_reg;
* run;

%prntlog;

/*****
/**** Les résultats de la macro logreg peuvent ****/
/**** être sauvegardés dans un fichier permanent ****/
/**** ****/

* data in3.ZZZ;
* set bs_reglg;
* run;

data time;
  set time;
  format stop datetime16.;
  stop=datetime();
  output;
run;

proc print data=time;
  title 'Temps requis pour exécuter le programme';
run;

/* Fin du programme SAS Bootvar */

```