

# **Estimation de la variance à l'aide des poids de bootstrap**

## **Guide de l'utilisateur du programme BOOTVARF\_V20.SAS**

### **(VERSION 2.0)**

## **1. Introduction**

Ce guide s'adresse aux utilisateurs du programme SAS BOOTVARF\_V20.SAS conçu pour faire l'estimation de variances à l'aide de la méthode du bootstrap. Ce programme permet d'estimer des variances pour l'Enquête nationale sur la santé de la population (ENSP) (volets ménages et établissements de santé) et pour l'Enquête sur la santé dans les collectivités canadiennes (ESCC).

La section 2 du présent guide explique brièvement la méthode de rééchantillonnage utilisée pour estimer la variance. La section 3 explique en détails les règles d'utilisation du programme BOOTVARF\_V20.SAS ainsi que les étapes préliminaires à effectuer. Les programmes à utiliser sont à l'annexe A. L'annexe B contient un exemple complet d'utilisation (programmes et résultats). Finalement, les particularités de chaque enquête (nom des fichiers, variables d'identification à utiliser, etc.) sont à l'annexe C.

### ***Changements par rapport à la version précédente:***

Le principal changement apporté au programme est qu'il se veut plus facile d'utilisation pour les utilisateurs. Toutes les commandes qui requièrent un changement ont été regroupées. Les changements requis sont énoncés plus clairement et les commandes qui ne doivent pas être modifiées (les macros par exemple) ont été enlevées du programme. Ces commandes se retrouvent maintenant dans le nouveau programme MACROF\_V20.SAS, que l'utilisateur ne doit pas modifier.

Il est aussi plus facile de tester le programme. Le seul changement à effectuer est de spécifier un nombre plus petit de poids bootstrap (au moins 2) au début du programme.

La macro pour les modèles linéaires généralisés n'est plus disponible. Cette macro nécessitait de nombreuses modifications de la part de l'utilisateur afin de l'adapter aux besoins de chaque analyse. Pour analyser des différences de moyennes, il est plutôt suggéré d'utiliser la macro "Différence de Rapports", qui est plus facile d'utilisation.

Finalement, le programme peut être utilisé aussi bien avec les données de l'ESCC que celles de l'ENSP. Aucune modification n'est requise à cet effet puisque le programme détecte automatiquement l'enquête.

Il est à noter que le programme a été testé et fonctionne avec les versions 6.12 et 8.2 de SAS.

## **2. Méthode du bootstrap**

Les plans d'échantillonnage pour les enquêtes sur la santé sont complexes. Comme le calcul de la variance pour de tels plans ne peut être fait à l'aide de simples formules, on a recours à une méthode de rééchantillonnage pour calculer des estimations de la variance.

La méthode du bootstrap consiste à sous-échantillonner l'échantillon initial. 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. Un certain nombre  $B$  de nouveaux échantillons (ou répétitions) sont ainsi créés. La même estimation est ensuite calculée pour chacun des  $B$  échantillons, ce qui permet d'obtenir  $B$  estimations différentes. L'estimation de la variance correspond à la simple variance entre les  $B$  estimations. Afin d'obtenir chacune des  $B$  estimations, un poids spécifique à chaque échantillon est nécessaire. Pour chaque échantillon ÉAS, le facteur de pondération est donc recalculé pour chaque enregistrement dans la strate. Ces  $B$  poids, les poids bootstrap, ont été produits et sont disponibles avec les données.

Avec le programme BOOTVARF\_V20.SAS, les poids bootstrap sont utilisés pour obtenir des estimations de variance précises pour de simples statistiques telles que des totaux et rapports, de même que 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.

Voici les principales étapes à effectuer pour obtenir une estimation valide de la variance d'une estimation donnée:

- A) On calcule d'abord une estimation (total, rapport, etc...) en utilisant le poids final inclus dans le fichier de données. Cette estimation est l'estimation ponctuelle.
- B) On calcule ensuite cette même estimation en utilisant cette fois chacun des  $B$  poids bootstrap contenus dans le fichier de poids bootstrap. On obtient ainsi  $B$  estimations (du total, du rapport, etc...)
- C) Finalement, on calcule la variance de ces  $B$  estimations. Cette variance est l'estimation de la variance de l'estimation ponctuelle calculée en A.

### **3. Estimation de la variance avec le programme BOOTVARF\_V20.SAS**

Le programme BOOTVARF\_V20.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.

L'estimation de la variance se fait en *deux étapes* et devrait impliquer 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 BOOTVARF\_V20.SAS (et MACROF\_V20.SAS) pour estimer les variances.

### ***Étape 1: Création du fichier d'analyse***

On crée d'abord le fichier SAS de données qui servira, à l'étape 2, de fichier d'entrée pour le programme estimant la variance. 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 doit être lu et le cliché d'article aussi fourni doit être utilisé pour spécifier les variables contenues dans le fichier. Voir l'annexe C pour le nom des fichiers et variables à utiliser.

2- Création des variables requises pour l'analyse: Des variables dichotomiques (1 ou 0) doivent être créées à partir des variables d'intérêt qualitatives contenues dans le fichier de données initial. Ceci a pour but d'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 rapports et de différences entre des rapports, 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 rapports. 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 dichotomiques et autres variables qui ne nécessitent pas de modification). Pour réduire le temps d'exécution du programme, NE PAS conserver les variables inutiles.
- La (Les) variable(s) d'identification des répondants
- Si nécessaire, la(les) variable(s) de ventilation, identifiant les groupes pour lesquels une analyse séparée est désirée (ex: province, sexe, etc...).
- Si des analyses ne sont voulues que pour un certain sous-groupe (une province ou un groupe d'âge par exemple), ne conserver que les enregistrements faisant partie de ce sous-groupe.

### **REMARQUES:**

- Il est suggéré d'effectuer les estimations ponctuelles à cette étape pour s'assurer que l'on calcule bien l'estimation voulue et que le programme BOOTVARF\_V20.SAS calcule bien cette même estimation. Dans ce cas, il est nécessaire de conserver la variable de poids lors de la création du fichier d'analyse.

- Les estimations de pourcentages sont des rapports. Deux variables dichotomiques doivent donc être créées; une pour le numérateur identifiant les individus ayant la caractéristique d'intérêt parmi le groupe d'intérêt et une pour le dénominateur identifiant les individus faisant partie du groupe d'intérêt.
- Les estimations de moyennes peuvent être considérées comme des estimations de rapports. Le numérateur est la variable identifiant la caractéristique d'intérêt parmi le groupe d'intérêt et le dénominateur est une variable dichotomique identifiant les individus faisant partie du groupe d'intérêt.

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 A (programme ETAPE1.SAS).

### ***Étape 2: Calcul des variances à l'aide du programme BOOTVARF\_V20.SAS***

Une fois le nouveau fichier SAS d'analyse créé à l'étape 1, l'étape suivante consiste à exécuter le programme BOOTVARF\_V20.SAS. Avant l'exécution, il faut spécifier les paramètres désirés et les analyses voulues. Il est à noter que ce programme fait appel au programme MACROF\_V20.SAS. MACROF\_V20.SAS contient les différentes macros qui permettent le calcul de l'estimation de la variance. *Pour une utilisation standard du programme d'estimation de la variance, aucune modification par l'utilisateur n'est requise au programme MACROF\_V20.SAS.* Des changements peuvent être nécessaires seulement dans quelques cas particuliers expliqués plus loin.

Le programme BOOTVARF\_V20.SAS est inclus à l'annexe A. Les parties à changer 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:

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
- Le nom du fichier de données à utiliser (fichier d'analyse créé à l'étape 1)
- Le nom du répertoire et du fichier contenant les poids bootstrap
- La(les) variable(s) de ventilation (ie: si l'analyse est faite séparément pour des sous-groupes spécifiques (ex: province, sexe))
- Le nombre de poids bootstrap à utiliser
- Le nom du répertoire où se trouve le programme MACROF\_V20.SAS

N.B.: L'annexe C contient l'information relative aux fichiers dont l'utilisateur a besoin (noms de fichiers, nom de certaines variables, nombre de poids bootstrap).

## Partie 2:

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

- les totaux
- les rapports
- les différences entre rapports
- les modèles de régression (linéaire ou logistique)

Pour les différences de rapports: Si l'utilisateur désire effectuer une différence de rapports, il doit modifier *si nécessaire* la macro *diff\_rat* du programme MACROF\_V20.SAS, selon ses besoins (voir les notes incluses dans le programme BOOTVARF\_V20.SAS pour plus de détails).

Pour les régressions: Les variables catégoriques seront traitées comme des variables continues. Il faut donc créer à l'étape 1 des variables dichotomiques pour chacune des valeurs possibles (sauf une) de la variable catégorique afin que cette variable soit traitée correctement.

Modification du programme pour effectuer des tests: L'exécution du programme peut prendre un certain temps (surtout pour l'analyse de modèles complexes). Il est possible de réduire le nombre de poids bootstrap utilisés afin de tester le programme (**Cependant, pour obtenir les estimations de variance finales, il est indispensable d'utiliser l'ensemble complet des poids bootstrap fournis**). Pour tester le programme, il suffit de modifier le nombre de poids à utiliser, dans la première partie du programme BOOTVARF\_V20.SAS.

### ***Résultats obtenus avec BOOTVARF\_V20.SAS***

Les résultats suivants sont obtenus après l'exécution de BOOTVARF\_V20.SAS pour des totaux, rapports et différences de rapports. (Voir l'exemple de l'annexe B pour l'interprétation des résultats.):

TYPE:	Type d'estimation (total, rapport,diff_rapport)
VAR1 & VAR2:	Variables utilisées pour calculer les estimations. Pour un total, VAR2=Aucune
VAR3 & VAR4:	Variables utilisées pour calculer les différences de rapports
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%

Note: Dans le cas des moyennes (considérées comme des rapports), l'estimation du paramètre et les bornes de l'intervalle de confiance doivent être divisées par 100 puisque les résultats sont sous forme de pourcentages.

Les résultats suivants sont obtenus après l'exécution de BOOTVARF\_V20.SAS pour des régressions linéaires et logistiques. (Voir l'exemple de l'annexe B pour l'interprétation des résultats.):

BETA:	Paramètre à estimer
BHAT:	Estimation du paramètre
ODDS:	Rapport de cotes (régression logistique seulement)
WALD:	Statistique de Wald (régression logistique seulement)
PVALUE:	Valeur p de la statistique de Wald (régression logistique seulement)
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
CIL95:	Borne inférieure de l'intervalle de confiance à 95% (du rapport de cotes pour la régression logistique)
CIU95:	Borne supérieure de l'intervalle de confiance à 95% (du rapport de cotes pour la régression logistique)

L'annexe A contient le programme BOOTVARF\_V20.SAS, précédé d'un exemple de programme à utiliser pour préparer le fichier d'analyse (ETAPE1.SAS). L'annexe B contient un exemple complet (programmes et résultats). Finalement, l'annexe C contient l'information relative aux fichiers dont l'utilisateur a besoin (noms de fichiers, nom de certaines variables, nombre de poids bootstrap).

## Annexe A - Programmes à exécuter

### Programme ETAPE1.SAS

(donné à titre d'exemple, l'utilisateur pouvant utiliser son propre programme)

Les parties en *caractère gras* doivent être modifiées.

```
*****
*                               *
*           ÉTAPE1.SAS          *
*                               *
* Ce programme permet de créer le fichier de données      *
* SAS contenant les variables nécessaires pour utiliser   *
* le programme BOOTVARF_V20.SAS                          *
*****;

LIBNAME in1 'nom_du_répertoire_où_sauvegarder_le_fichier';

*** 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.nom_du_fichier_d''analyse; /* Fichier qui sera utilisé avec BOOTVARF_V20.SAS */
%let datafid="location_et_nom_du_fichier_source";
%include "location_et_nom_du_cliché_d'article";

*** Création des variables dichotomiques ***
*** (des exemples sont présentés ci-dessous, utilisant les variables de l'ENSP, cycle 3) ***;

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

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

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

keep liste_des_variables_à_conserver;

* Il est recommandé de ne conserver que les variables nécessaire *
* afin de diminuer le temps d'exécution du programme BOOTVARF_V20.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;

*****
* calcul des estimations ponctuelles *
* Conseillé, mais non obligatoire... *
*****;

proc freq data=in1.nom_du_fichier_d''analyse;
table variables_désirées;
weight variable_de_poids;
run;

proc logistic data=in1.nom_du_fichier_d''analyse;
model variable_dépendante = variables_indépendantes;
weight variable_de_poids;
run;
```

• • •

• • •

• • •

8



```

/*****
/****                                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(nom_de_la_variable);

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

    * %ratio(variable_au_numérateur,variable_au_dénominateur);

* POUR OBTENIR L'ESTIMATION DE LA VARIANCE D'UNE DIFFÉRENCE DE RAPPORTS, EXÉCUTER:
-----;
    * ATTENTION: Voir la remarque au début de la partie 2...
    * %diff_rat(VAR1,VAR2,VAR3,VAR4);

        * où: 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   *;

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

    * %regress(var_dépendante,variables_indépendantes_(sans virgule));

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

    * %logreg(var_dépendante,variables_indépendantes_(sans virgule));

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

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

    * data out.non_du_fichier_de_résultats;
    * set &result ;
    * run;

/* Fin du programme SAS BOOTVARF_V20.SAS */

```

## ANNEXE B

Voici un exemple complet d'utilisation du programme BOOTVARF\_V20.SAS. Dans un premier temps, le fichier d'analyse est créé (étape 1). Ensuite, BOOTVARF\_V20.SAS est adapté selon les besoins de l'analyse. Les résultats obtenus suivent après les programmes.

### Mise en situation:

Cet exemple utilise le fichier transversal de la composante générale du cycle 3 (1998) de l'ENSP. On cherche à:

- 1- Calculer le nombre et la proportion de diabétiques dans la population générale et parmi les hommes seulement, par province (seulement quatre provinces seront conservées).
- 2- Étudier la relation entre le diabète, le sexe et le type d'entrevue (par procuration ou non), par province.

### Étape 1:

```
*****
*                               *
*           ÉTAPE1.SAS          *
*                               *
* Ce programme permet de créer le fichier de données *
* SAS contenant les variables nécessaires pour utiliser *
* le programme BOOTVARF_V20.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_V20.SAS */
  %let datafid='D:\Data\h35.txt';
  %include 'D:\Layout\h35_i.sas';

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

*** Création des variables dichotomiques ***
*** (des exemples sont présentés ci-dessous, utilisant les variables de l'ENSP, cycle 3) ***;

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

```

/* Variable indicatrice (0/1) pour le type d'entrevue */
nonproxy=0;
if am58_pxy>2 then nonproxy=.;
if am58_pxy=2 then nonproxy=1;

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

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

keep diab total hommes femmes hdiab fdiab nonproxy wt58 realukey personid prc8_cur;

* Il est recommandé de ne conserver que les variables nécessaire *
* afin de diminuer le temps d'exécution du programme BOOTVARF_V20.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;

*****
* Calcul des estimations ponctuelles *
* Conseillé, mais non obligatoire... *
*****;

PROC SORT DATA=in1.diabete; BY prc8_cur; RUN;

proc freq data=in1.diabete;
table diab hdiab fdiab;
by prc8_cur;
weight wt58;
run;

proc logistic data=in1.diabete;
model diab = nonproxy femmes;
by prc8_cur;
weight wt58;
TITLE "Relation entre diabète, sexe et type d'entrevue";
run;

```

## Etape 2 - Programme BOOTVARF V20.SAS:

```

*****
/*****
/****          PARTIE 1          ****
/*****
/**** Cette partie permet à l'utilisateur de spécifier les différents ****
/**** paramètres désirés (noms de variables, noms de répertoires, noms ****
/**** de fichiers, etc...) ****
/**** ****
/*****;/

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

libname in1 "C:\BOOTVAR";           /* (ex: c:\data) */
libname out "C:\BOOTVAR";          /* (ex: c:\output)*/

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

%let Mfile = in1.diabete;

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

* EXÉCUTEZ CETTE PARTIE SI LES POIDS BOOTSTRAP SONT EN FORMAT SAS (enlever les " * ");
-----;

libname in2 "D:\bootstrp\DATA";
%let bsamp = in2.b5h35;

* EXÉCUTEZ CETTE PARTIE SI LES POIDS BOOTSTRAP SONT EN FORMAT ASCII (.TXT) (enlever les " * ");
-----;

*      data poidboot;
*      %let datafid="location(répertoire)_et_nom_du_fichier_de_poids_(avec_extension)";
*      %include "location(répertoire)_et_nom_du_cliche_d'article_(layout)_(avec_extension)";
*      run;
*      %let bsamp=poidboot;

*****
** VEUILLEZ SPÉCIFIER LA(LES) VARIABLE(S) DE VENTILATION, SI DÉSIÉ (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. **
*****;

%let classes = prc8_cur ;

*****
** VEUILLEZ INDIQUER LE NOMBRE DE POIDS BOOTSTRAP À UTILISER: **
** IMPORTANT: IL EST NÉCESSAIRE D'UTILISER TOUS LES POIDS BOOTSTRAP AU MOMENT **
** D'EFFECTUER LES ANALYSES FINALES. LE PROGRAMME BOOTVARF_V20.SAS **
** DOIT ENSUITE ÊTRE EXÉCUTÉ AU COMPLET. **
** **
** - Se référer à l'annexe C du document pour savoir le nombre de poids **
**   que contient le fichier de poids. **
** - Pour les tests, il faut que B >= 2. **
*****;

%let B = 500;

*****
** VEUILLEZ SPÉCIFIER LE RÉPERTOIRE ET LE NOM DU FICHIER DU PROGRAMME DE MACRO **
** (LE PROGRAMME MACROF_V20.SAS SI L'USAGER N'A PAS FAIT DE MODIFICATIONS) **
*****;

%include "C:\BOOTVAR\MACROF_V20.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(nom_de_la_variable);

    %total(diab);
    %total(hdiab);

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

    * %ratio(variable_au_numérateur,variable_au_dénominateur);

    %ratio(diab,total);
    %ratio(hdiab,hommes);

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

    * %diff_rat(VAR1,VAR2,VAR3,VAR4);

        * où: 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 *;

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

    * %regress(var_dépendante,variables_indépendantes_(sans virgule));

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

    * %logreg(var_dépendante,variables_indépendantes_(sans virgule));

    %logreg(diab,nonproxy femmes);

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

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

    data out.resultats;
      set &result ;
    run;

/* Fin du programme SAS BOOTVARF_V20.SAS */

```

**Résultats et interprétations:**

Les tableaux à la page suivante présentent les résultats des analyses réalisées à l'aide des programmes de l'exemple. Le premier tableau présente les résultats des totaux et des rapports. Par exemple, si on s'intéresse au rapport entre le nombre d'hommes diabétiques et le nombre total d'hommes, en Ontario, on se réfère à l'observation 12. La région 35 correspond à la province de l'Ontario (voir le dictionnaire de données inclus sur le CD-ROM pour connaître 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 *hdiab* (VAR1) et au dénominateur la variable *hommes* (VAR2). L'estimation du rapport est de 3,57% (YHAT) avec un écart-type de 0,29 (BS\_SD) et un coefficient de variation de 8,26 (BS\_CV). L'intervalle de confiance à 95% pour cette estimation est (2,99%, 4,14%) (CIL95, CIU95).

Les résultats de la régression logistique sont présentés dans le deuxième tableau. Par exemple, le paramètre estimé pour la variable *femmes*, pour l'Ontario (observation 9) est -0,34507 (BHAT) et le rapport de cotes est de 0,70817 (ODDS). La statistique de Wald et son seuil p associé pour ce paramètre sont de 7,996 (WALD) et  $p=0,004687$  (PVALUE) respectivement. L'estimation de la variance et de l'écart-type du paramètre estimé sont de 0,014891 (BS\_VAR) et 0,12203 (BS\_SD) et le coefficient de variation est de 35,36 (BS\_CV). Finalement, l'intervalle de confiance pour le rapport de cotes est (0,55753, 0,89952) (CIL95, CIU95).

Finalement, le temps d'exécution qu'à pris le programme est donné à la fin.

## Annexe B

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

Obs	PRC8_CUR	type	var1	var2	yhat	bs_sd	bs_cv	cil95	ciu95
1	10	Total	diab	aucune	20741.31	1778.61	8.58	17255.23	24227.39
2	10	Total	hdiab	aucune	7029.11	1356.31	19.30	4370.73	9687.48
3	10	Rapport	diab	total	3.85	0.33	8.58	3.21	4.50
4	10	Rapport	hdiab	hommes	2.63	0.51	19.30	1.64	3.63
5	24	Total	diab	aucune	205292.21	16330.49	7.95	173284.44	237299.98
6	24	Total	hdiab	aucune	110452.77	10818.85	9.80	89247.83	131657.72
7	24	Rapport	diab	total	2.87	0.23	7.95	2.42	3.32
8	24	Rapport	hdiab	hommes	3.12	0.31	9.80	2.52	3.72
9	35	Total	diab	aucune	362439.56	20692.51	5.71	321882.24	402996.88
10	35	Total	hdiab	aucune	198237.67	16369.18	8.26	166154.07	230321.26
11	35	Rapport	diab	total	3.22	0.18	5.71	2.86	3.58
12	35	Rapport	hdiab	hommes	3.57	0.29	8.26	2.99	4.14
13	59	Total	diab	aucune	110375.38	10661.17	9.66	89479.48	131271.27
14	59	Total	hdiab	aucune	62808.64	8301.60	13.22	46537.50	79079.77
15	59	Rapport	diab	total	2.83	0.27	9.66	2.29	3.37
16	59	Rapport	hdiab	hommes	3.24	0.43	13.22	2.40	4.08

Estimation de la variance à l'aide du bootstrap 500 pour des  
Régressions logistiques  
Variable dépendante: diab

Obs	PRC8_CUR	beta	bhat	Odds	wald	pvalue	bs_var	bs_sd	bs_cv	cil95	ciu95
1	10	Intercept	-4.00372	0.01825	255.389	0.000000	0.062766	0.25053	6.26	0.01117	0.02982
2	10	nonproxy	0.85783	2.35803	11.819	0.000586	0.062260	0.24952	29.09	1.44595	3.84543
3	10	femmes	0.46625	1.59401	2.859	0.090860	0.076036	0.27575	59.14	0.92848	2.73660
4	24	Intercept	-3.89892	0.02026	852.643	0.000000	0.017829	0.13352	3.42	0.01560	0.02633
5	24	nonproxy	0.89121	2.43807	34.116	0.000000	0.023281	0.15258	17.12	1.80787	3.28795
6	24	femmes	-0.36991	0.69080	5.783	0.016180	0.023660	0.15382	41.58	0.51100	0.93387
7	35	Intercept	-3.57575	0.02799	961.191	0.000000	0.013302	0.11534	3.23	0.02233	0.03510
8	35	nonproxy	0.60946	1.83943	23.520	0.000001	0.015793	0.12567	20.62	1.43785	2.35318
9	35	femmes	-0.34507	0.70817	7.996	0.004687	0.014891	0.12203	35.36	0.55753	0.89952
10	59	Intercept	-3.99955	0.01832	301.873	0.000000	0.052991	0.23020	5.76	0.01167	0.02877
11	59	nonproxy	1.05756	2.87934	18.892	0.000014	0.059203	0.24332	23.01	1.78722	4.63883
12	59	femmes	-0.46337	0.62916	4.591	0.032133	0.046763	0.21625	46.67	0.41180	0.96125

Temps requis pour exécuter le programme

Obs	debut	fin
1	20MAR02:10:16:17	20MAR02:10:23:51

## ANNEXE C

Noms des variables et des fichiers à utiliser avec le programme

Note: Pour une liste complète des variables disponibles, consulter le dictionnaire des données, disponible dans la documentation fournie.

ENSP - Volet ménages					
	Nom du fichier de données	Nom du fichier de poids bootstrap (Format ASCII: .txt Format SAS: .sd2 ou .sas7bdat)	Variables d'identification	Variables de poids	# de poids
<b>Fichier Général :</b>					
<b>Cycle 1</b>	H3H5.txt	B5H35	REALUKEY PERSONID	WT54 (M) SHRWT5 (S)	500
<b>Cycle 2</b>	H35.txt	B5H35	REALUKEY PERSONID	WT56 (M) WT56_S (S)	100
<b>Cycle 3</b>	H35.txt	B5H35	REALUKEY PERSONID	WT58 (M) WT58_S (S)	500
<b>Fichier Santé :</b>					
<b>Cycle 1</b>	H3H5H6.txt	B5H356	REALUKEY PERSONID	WT64 (M) SHRWT6 (S)	500
<b>Cycle 2</b>	H356.txt	B5H356	REALUKEY PERSONID	WT66 (M) WT66_S (S)	500
<b>Cycle 2</b> <i>questions de la section Enquête de la promotion sur la santé</i>	H356.txt	B5H356A	REALUKEY PERSONID	WT66_N (M) WT66_SN (S)	500
<b>Cycle 2</b> <i>questions de la section Service de santé pour enfants pour les enfants au Man. et en Alb.</i>	H356.txt	B5H356C	REALUKEY PERSONID	WT66_N (M) WT66_SN (S)	500
<b>Cycle 3</b>	H356.txt	B5H356	REALUKEY PERSONID	WT68 (M) WT68_S (S)	500
<b>Fichier Longitudinal :</b>					
<b>Cycle 2 - Complet</b>	LNGF.txt	B5LONGF	REALUKEY PERSONID	WT66LF (M) WT66SLF (S)	500
<b>Cycle 2 - Partiel</b> (Maître seulement)	LNGP.txt	B5LONGP	REALUKEY PERSONID	WT66LP (M)	500

(M) - Fichiers Maîtres

(S) - Fichiers de Partage



## Annexe C

<b>Cycle 2 - Carré (Maître seulement)</b>	LONG.txt	B5LONGS	REALUKEY PERSONID	WT66LF (M)	500
<b>Cycle 3 – Complet</b>	LNGF.txt	B5LNGF	REALUKEY PERSONID	WT68LF (M) WT68_SLF(S)	500
<b>Cycle 4 – Complet</b>	LONG.txt (M) LNGF.txt (S)	B5LNGF	REALUKEY PERSONID	WT60LF (M) WT60SLF (S)	500
<b>Cycle 4 - Carré (Maître seulement)</b>	LONG.txt	B5LONG	REALUKEY PERSONID	WT64LS (M)	500
<b>Cycle 4 - Complet Cycle 1 &amp; Cycle 4 (Maître seulement)</b>	LONG.txt	B5LNGFE	REALUKEY PERSONID	WT60LFE (M)	500

<b>ENSP - Volet Établissements de Santé</b>					
	<b>Nom du fichier de données</b>	<b>Nom du fichier de poids bootstrap</b> <i>(Format ASCII: .txt Format SAS: .sd2 ou .sas7bdat)</i>	<b>Variables d'identification</b>	<b>Variable de poids</b>	<b># de poids</b>
<b>Fichier Longitudinal :</b>					
<b>Cycle 3</b>	LNGF.txt	B2LNGF	UNIQUEID	WTI8LF (M) WTI8_S (S)	2000
<b>Cycle 4</b>	LNGF.txt	B2LNGF	UNIQUEID	WTI0LF (M) WTI0SLF (S)	2000

<b>ESCC</b>					
	<b>Nom du fichier de données</b>	<b>Nom du fichier de poids bootstrap</b> <i>(Format ASCII: .txt Format SAS: .sd2 ou .sas7bdat)</i>	<b>Variables d'identification</b>	<b>Variable de poids</b>	<b># de poids</b>
<b>Cycle 1.1</b>	HS.txt	B5	SAMPLEID PERSONID	WTSA_M (M) WTSA_S (S)	500
<b>Cycle 1.1</b> <i>Î.P.É - Échantillon supplémentaire acheté</i>	HS.txt	B5_PEI	SAMPLEID PERSONID	WTSA_PEM (M) WTSA_PES (S)	500
<b>Cycle 1.1</b> <i>Quatrième trimestre</i>	HS.txt	B5_Q4	SAMPLEID PERSONID	WTSA_Q4M (M) WTSA_Q4S (S)	500

(M) - Fichiers Maîtres

(S) - Fichiers de Partage