

# **Estimation de la variance à l'aide des poids de bootstrap Guide de l'utilisateur du programme BOOTVARF\_V30.SAS (VERSION 3.0)**

## **1. Introduction**

Ce guide s'adresse aux utilisateurs du programme SAS BOOTVARF\_V30.SAS conçu pour faire l'estimation de variances à l'aide de la méthode du bootstrap.

La section 2 du présent guide explique brièvement la méthode de rééchantillonnage du bootstrap utilisée pour estimer la variance. La section 3 explique en détails les règles d'utilisation du programme BOOTVARF\_V30.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 qui peut être utilisée avec le programme (nom des fichiers, variables à utiliser, etc.) sont disponibles dans le document « AnnexeC\_XYZ », où XYZ identifie l'enquête.

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

Le principal changement apporté au programme est qu'une version unique supporte maintenant toutes les enquêtes de Statistique Canada qui utilisent le programme BOOTVAR. L'utilisateur n'a qu'à spécifier certains paramètres (voir document AnnexeC\_XYZ) à la section 1 du programme BOOTVARF\_V30.SAS.

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 de Statistique Canada sont généralement complexes. Comme le calcul de la variance pour de tels plans ne peut être fait à l'aide de formules simples, on a souvent 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. Ce processus est répété  $B$  fois, créant ainsi  $B$  nouveaux échantillons (ou répétitions). Pour chaque échantillon, le facteur de pondération est recalculé pour chaque enregistrement dans la strate et ces  $B$  facteurs de pondération sont appelés les poids bootstrap. La même estimation est ensuite calculée pour chacun des  $B$  échantillons en utilisant les poids bootstrap et l'estimation de la variance correspond à la simple variance entre les  $B$  estimations.

Les poids bootstrap ont été produits et sont disponibles avec les données. Le programme BOOTVARF\_V30.SAS utilise les poids bootstrap pour obtenir des estimations de variance pour de simples statistiques telles que des totaux et ratios, de même que pour des analyses plus complexes comme des régressions. Ces estimations de variance devraient être utilisées pour

obtenir des indicateurs de qualité et peut pouvoir appliquer les règles et seuils de diffusion définis pour l'enquête.

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

A) On calcule tout d'abord une estimation (total, ratio, 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 ratio, 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\_V30.SAS**

Le programme BOOTVARF\_V30.SAS permet de calculer des estimations de variance pour des totaux, des ratios, des différences entre des ratios et des paramètres de régression linéaire et logistique.

L'estimation de la variance se fait en *deux étapes* et implique l'utilisation de trois programmes SAS. La *première étape* consiste à créer un fichier de données contenant les variables à être utilisées pour faire l'analyse (premier programme). La *deuxième étape* consiste à utiliser BOOTVARF\_V30.SAS (et MACROF\_V30.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 AnnexeC\_XYZ pour le nom des fichiers et variables à utiliser.

2- Création des variables requises pour l'analyse: Des variables dérivées à partir des d'intérêt devraient être créées à cette étape. Il peut être nécessaire de créer des variables dichotomiques (1 ou 0) pour identifier les enregistrements ayant les différentes caractéristiques à l'étude. On doit donc créer une variable dichotomique qui prendra une valeur de 1 pour les enregistrements ayant la caractéristique d'intérêt et une valeur de 0 sinon. Par exemple, pour des estimations de

totaux, de ratios et de différences entre des ratios, ces variables dichotomiques serviront à identifier les enregistrements ayant la caractéristique d'intérêt afin de sommer leurs poids pour obtenir, lors de l'étape 2, les totaux ou ratios. Voir l'exemple à l'annexe B pour plus de détails.

Le fichier d'analyse doit donc contenir:

- Les variables nécessaires pour l'analyse (variables dérivées, incluant les variables dichotomiques, et autres variables qui ne nécessitent pas de modification). Pour réduire le temps d'exécution du programme, NE PAS conserver les variables inutiles.
- 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\_V30.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 moyennes peuvent être obtenues en utilisant la macro ratio à l'étape 2. Une variable dichotomique doit être créée pour le dénominateur, identifiant ainsi 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\_V30.SAS***

Une fois le nouveau fichier SAS d'analyse créé à l'étape 1, l'étape suivante consiste à exécuter le programme BOOTVARF\_V30.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\_V30.SAS. MACROF\_V30.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\_V30.SAS.* Des changements peuvent être nécessaires seulement dans quelques cas particuliers expliqués plus loin.

Le programme BOOTVARF\_V30.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 (i.e.: si l'analyse est faite séparément pour des sous-groupes spécifiques (ex.: province, sexe))
- La (Les) variable(s) d'identification des répondants, les variables de poids et le nombre de poids bootstrap à utiliser
- Le nom du répertoire où se trouve le programme MACROF\_V30.SAS

N.B.: Le fichier AnnexeC\_XYZ 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 ratios (moyennes)
- les différences entre ratios
- les modèles de régression (linéaire ou logistique)

Pour les moyennes : Afin d'estimer la variance d'une moyenne, la macro pour les ratios doit être utilisée. Le numérateur représente la variable d'intérêt et le dénominateur est une variable dichotomique qui identifie la population d'intérêt.

Pour les totaux, les ratios et les différences de ratios: Les variables d'intérêts doivent prendre uniquement des valeurs positives ou nulles.

Pour les différences de ratios: Si l'utilisateur désire effectuer une différence de ratios, il doit modifier *si nécessaire* la macro *diff\_rat* du programme MACROF\_V30.SAS, selon ses besoins (voir les notes incluses dans le programme BOOTVARF\_V30.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\_V30.SAS.

### ***Résultats obtenus avec BOOTVARF\_V30.SAS***

Les résultats suivants sont obtenus après l'exécution de BOOTVARF\_V30.SAS pour des totaux, ratios et différences de ratios. (Voir l'exemple de l'annexe B pour l'interprétation des résultats.):

TYPE: Type d'estimation (total, ratio,dif\_rat)  
VAR1 à VAR4: Variables utilisées pour calculer les estimations.  
n, n1, n3: Tailles d'échantillons pour les totaux (n) et les ratios (n1 et n3)  
Estimation: Estimation du paramètre  
BS\_ET: Écart type  
BS\_CV: Coefficient de variation  
IC\_i95: Borne inférieure de l'intervalle de confiance à 95 %  
IC\_s95: Borne supérieure de l'intervalle de confiance à 95 %

Les résultats suivants sont obtenus après l'exécution de BOOTVARF\_V30.SAS pour des régressions linéaires et logistiques. (Voir l'exemple de l'annexe B pour l'interprétation des résultats.):

PARAM: Paramètre à estimer  
BETA : Estimation du paramètre  
COTES: Rapport de cotes (régression logistique seulement)  
WALD: Statistique de Wald (régression logistique seulement)  
VALEURP: Valeur p de la statistique de Wald (régression logistique seulement)  
BSVAR: Variance de l'estimation du paramètre  
BS\_ET: Écart type de l'estimation du paramètre  
BS\_CV: Coefficient de variation de l'estimation du paramètre  
IC\_i95: Borne inférieure de l'intervalle de confiance à 95 % (du rapport de cotes pour la régression logistique)  
IC\_s95: Borne supérieure de l'intervalle de confiance à 95 % (du rapport de cotes pour la régression logistique)

L'annexe A contient le programme BOOTVARF\_V30.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, le fichier AnnexeC\_XYZ 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_V30.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_V30.SAS */
%let datafid="location_et_nom_du_fichier_source";
%include "location_et_nom_du_fichier_source";

*** Création des variables dichotomiques ***

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

/* 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 */
hdlab = diab * hommes; /* hommes diabétiques */
fdlab = diab * femmes; /* femmes diabétiques */

keep liste des variables à conserver;

* Il est recommandé de ne conserver que les variables nécessaires *
* afin de diminuer le temps d'exécution du programme BOOTVARF_V30.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 logstic data=in1. nom_du_fichier_d'analyse;
model variable_dépendante = variables indépendantes;
weight variable_de_poids;
run;

```

**Programme BOOTVARF V30.SAS**

Les parties en *caractère gras* doivent être modifiées.

```

...
...
...
/*****
/****          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 "nom_du_répertoire_contenant_fichier_d'analyse_étape_1"; /* (ex: c:\data) */
libname out "nom_du_répertoire_où_sauvegarder_résultats"; /* (ex: c:\output) */

*****
** VEUILLEZ INDIQUER LE NOM DU FICHIER D'ANALYSE (CRÉÉ À L'ÉTAPE 1) (sans extension): **
*****
%let Mfile = in1.nom_du_fichier_d'analyse;

*****
** VEUILLEZ INDIQUER LE NOM DU FICHIER CONTENANT LES POIDS BOOTSTRAP: **
** NB: N'exécuter qu'une seule des deux séries de commandes suivantes **
** (mettre l'autre en commentaire, ou l'effacer): **
*****
* EXÉCUTEZ CETTE PARTIE SI LES POIDS BOOTSTRAP SONT EN FORMAT SAS (enlever les " * ")
-----;
* libname in2 "nom_du_répertoire_contenant_fichier_poids_bootstrap"; /* (ex: c:\bootstrp) */
* %let bsamp = in2.nom_du_fichier_SAS_contenant_les_poids_(sans_extension);

* EXÉCUTEZ CETTE PARTIE SI LES POIDS BOOTSTRAP SONT EN FORMAT ASCII (.TXT) (enlever les " * ")
-----;
* data poi dboot;
* %let datafid="location(répertoire)_et_nom_du_fichier_de_poids_(avec_extension)";
* %include "location(répertoire)_et_nom_du_fichier_de_poids_(avec_extension)";
* run;
* %let bsamp=poi dboot;

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

*****
** VEUILLEZ INDIQUER L'INFORMATION SUIVANTE (SPÉCIFIQUE À L'ENQUÊTE QUE VOUS UTILISEZ): **
** Vous devez spécifier: **
** 1- La(les) variable(s) d'identification unique (séparées par un espace) **
** 2- La variable de poids final, incluse dans le fichier de poids bootstrap **
** 3- Le préfixe des variables de poids bootstrap **
** 4- Le nombre de poids bootstrap à utiliser (note: Pour les tests, il faut **
** que B >= 2. IL EST NÉCESSAIRE D'UTILISER TOUS LES POIDS BOOTSTRAP AU **
** MOMENT D'EFFECTUER LES ANALYSES FINALES.) **
** **
** - Se référer AnnexeC_XYZ pour obtenir ces informations **
*****
%let ident = variable(s)_d'identification_unique;
%let fwgt = poids_final;
%let bsw = préfixe_des_variables_de_poids_bootstrap;
%let B = nombre_de_poids_à_utiliser;

*****
** VEUILLEZ SPÉCIFIER LE RÉPERTOIRE ET LE NOM DU FICHIER DU PROGRAMME DE MACRO **
** (LE PROGRAMME MACROF_V30.SAS SI L'USAGER N'A PAS FAIT DE MODIFICATIONS) **
*****
%include "nom_du_répertoire_de_macrof_v30.sas\MACROF_V30.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);

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

    * %ratio(variable_au_numérateur, variable_au_dénominateur);

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

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

        * où: var1   : la variable au numérateur du 1er ratio      *
          *   var2   : la variable au dénominateur du 1er ratio   *
          *   var3   : la variable au numérateur du 2e ratio      *
          *   var4   : la variable au dénominateur du 2e ratio    *

* POUR OBTENIR L'ESTIMATION DE LA VARIANCE 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_V30.SAS */

```

## ANNEXE B

Voici un exemple complet d'utilisation du programme BOOTVARF\_V30.SAS. Dans un premier temps, le fichier d'analyse est créé (étape 1). Ensuite, BOOTVARF\_V30.SAS est adapté selon les besoins de l'analyse. Les résultats obtenus et leur interprétation suivent après les programmes.

### Mise en situation:

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

- 1- Calculer le nombre et la proportion de diabétiques dans la population et chez les hommes, 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.

Les paramètres nécessaires au programme (AnnexeC\_Santé) sont :

ENSP - Volet ménages							
	Fichier de données	Fichier de poids bootstrap (Format ASCII: .txt Format SAS: .sd2 ou .sas7bdat)	Variabes d'identification	Variabes de poids (dans le fichier de données)	Variabes de poids (dans le fichier de poids boot)	Préfixe des poids bootstrap	# de poids
<b>Fichier Général :</b>							
Cycle 3	H35.txt	B5H35	REALUKEY PERSONID	WT58 (M) WT58_S (S)	FWGT	BSW	500



**Etape 2 - Programme BOOTVARE V30.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'exécutez 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 poi dboot;
* %let datafid="location(répertoire)_et_nom_du_fichier_de_poids_(avec_extension)";
* %include "location(répertoire)_et_nom_du_fichier_de_poids_(avec_extension)";
* run;
* %let bsamp=poi dboot;

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

%let classes = prc8_cur ;

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

%let ident = REALUKEY PERSONID;
%let fwgt = FWGT;
%let bsw = BSW;
%let B = 500;

*****
** VEUILLEZ SPÉCIFIER LE RÉPERTOIRE ET LE NOM DU FICHIER DU PROGRAMME DE MACRO **
** (LE PROGRAMME MACROF_V30.SAS SI L'USAGER N'A PAS FAIT DE MODIFICATIONS) **
*****

%i nclude "C:\BOOTVAR\MACROF_V30.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_variabl e);
    %total(di ab);
    %total(hdi ab);

* POUR OBTENIR L' ESTIMATION DE LA VARIANCE D' UN RATIO, EXÉCUTER:
-----;
    * %ratio(variabl e_au_numérateur, variabl e_au_dénominateur);
    %ratio(di ab, total);
    %ratio(hdi ab, hommes);

* POUR OBTENIR L' ESTIMATION DE LA VARIANCE D' UNE DIFFÉRENCE DE RATIOS, EXÉCUTER:
-----;
    * ATTENTION: Voir la remarque au début de la partie 2... ;
    * %diff_rat(VAR1, VAR2, VAR3, VAR4);
        * où: var1 : la variable au numérateur du 1er ratio *
        *       var2 : la variable au dénominateur du 1er ratio *
        *       var3 : la variable au numérateur du 2e ratio *
        *       var4 : la variable au dénominateur du 2e ratio *;

* POUR OBTENIR L' ESTIMATION DE LA VARIANCE DES PARAMÈTRES D' UNE RÉGRESSION, EXÉCUTER:
-----;
    * %regress(var_dépendante, variabl es_i ndépendantes_(sans vi rgul e));

* POUR OBTENIR L' ESTIMATION DE LA VARIANCE DES PARAMÈTRES D' UNE RÉGRESSION LOGISTIQUE, EXÉCUTER:
-----;
    * %logreg(var_dépendante, variabl es_i ndépendantes_(sans vi rgul e));
    %logreg(di ab, nonproxy femmes);

%output; /* Affi che les résul tats à l' écran. Ne pas modi fi er... */

* POUR SAUVEGARDER LES RÉSULTATS DANS UN FICHER, EXÉCUTER: (enl ever les "**")
-----;
    data out.resul tats;
    set &resul t ;
    run;

/* Fin du programme SAS BOOTVARF_V30.SAS */

```

**Résultats et interprétation:**

Les tableaux de la page suivante présentent les résultats des analyses réalisées à l'aide des programmes de l'exemple. Les premier et deuxième tableaux présentent les résultats des totaux et des ratios.

Pour les totaux et les ratios, les tableaux donnent, dans un premier temps, les résultats pour l'ensemble du fichier (TOUS). Les résultats pour chaque catégorie de(s) la(les) variable(s) de ventilation suivent ensuite. Par exemple, si on s'intéresse au ratio entre le nombre d'hommes diabétiques et le nombre total d'hommes, en Ontario, on se réfère à l'observation 8 du deuxième tableau. La région 35 correspond à la province de l'Ontario et la variable Type indique le type d'analyse, dans ce cas-ci, un ratio. On retrouve au numérateur du ratio la variable *hdiab* (VAR1) et au dénominateur la variable *hommes* (VAR2). La colonne n1 indique que le fichier de données contient 190 hommes diabétiques en Ontario (taille de l'échantillon pour le numérateur). L'estimation du ratio est de 3,57 % (Estimation) avec un écart type de 0,29 (BS\_ET) et un coefficient de variation de 8,00 % (BS\_CV). L'intervalle de confiance à 95 % pour cette estimation est (3,01 %, 4,13 %) (IC\_i95, IC\_s95).

Les résultats de la régression logistique sont présentés au troisième tableau. Par exemple, le paramètre estimé pour la variable *femmes*, pour l'Ontario (observation 9) est -0,34507 (Beta) et le rapport de cotes est de 0,70817 (COTES). La statistique de Wald et son seuil p associé pour ce paramètre sont de 7,30 (WALD) et p=0,0069 (VALEURP) respectivement. L'estimation de la variance et de l'écart type du paramètre estimé sont de 0,016319 (BS\_VAR) et 0,12775 (BS\_ET) et le coefficient de variation est de 37,02 % (BS\_CV). Finalement, l'intervalle de confiance pour le rapport de cotes est (0,55131; 0,90966) (IC\_i95, IC\_s95).

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

Obs	prc8_cur	type	var	n	Estimation	bs_et	bs_cv	IC_i95	IC_s95
1	Tous	Total	diab	785	698848.46	27250.05	3.90	645438.37	752258.55
2	Tous	Total	hdiab	392	378528.18	20925.33	5.53	337514.54	419541.83
3	10	Total	diab	99	20741.31	1845.88	8.90	17123.38	24359.23
4	10	Total	hdiab	35	7029.11	1380.61	19.64	4323.11	9735.10
5	24	Total	diab	199	205292.21	15960.30	7.77	174010.03	236574.40
6	24	Total	hdiab	104	110452.77	10944.25	9.91	89002.04	131903.51
7	35	Total	diab	374	362439.56	19721.07	5.44	323786.27	401092.86
8	35	Total	hdiab	190	198237.67	15854.33	8.00	167163.17	229312.16
9	59	Total	diab	113	110375.38	10847.03	9.83	89115.19	131635.56
10	59	Total	hdiab	63	62808.64	8568.12	13.64	46015.12	79602.15

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

Obs	prc8_cur	type	var1	var2	n1	Estimation	bs_et	bs_cv	IC_i95	IC_s95
1	Tous	Ratio	diab	total	785	0.0306	0.0012	3.90	0.0282	0.0329
2	Tous	Ratio	hdiab	hommes	392	0.0335	0.0019	5.53	0.0299	0.0371
3	10	Ratio	diab	total	99	0.0385	0.0034	8.90	0.0318	0.0453
4	10	Ratio	hdiab	hommes	35	0.0263	0.0052	19.64	0.0162	0.0365
5	24	Ratio	diab	total	199	0.0287	0.0022	7.77	0.0243	0.0331
6	24	Ratio	hdiab	hommes	104	0.0312	0.0031	9.91	0.0252	0.0373
7	35	Ratio	diab	total	374	0.0322	0.0018	5.44	0.0288	0.0356
8	35	Ratio	hdiab	hommes	190	0.0357	0.0029	8.00	0.0301	0.0413
9	59	Ratio	diab	total	113	0.0283	0.0028	9.83	0.0228	0.0337
10	59	Ratio	hdiab	hommes	63	0.0324	0.0044	13.64	0.0237	0.0411

Estimation de la variance à l'aide du bootstrap pour une Régression Logistique  
Variable dépendante: diab  
(500 répliques bootstrap)

Obs	PRC8_CUR	param	beta	cotes	wald	valeurp	bs_var	bs_et	bs_cv	IC_i95	IC_s95
1	10	Intercept	-4.00372	0.01825	266.88	0.000000	0.060064	0.24508	6.12	0.01129	0.02950
2	10	nonproxy	0.85783	2.35803	12.84	0.000339	0.057293	0.23936	27.90	1.47504	3.76961
3	10	females	0.46625	1.59401	2.72	0.098902	0.079831	0.28254	60.60	0.91619	2.77330
4	24	Intercept	-3.89892	0.02026	866.89	0.000000	0.017536	0.13242	3.40	0.01563	0.02627
5	24	nonproxy	0.89121	2.43807	31.64	0.000000	0.025100	0.15843	17.78	1.78726	3.32587
6	24	females	-0.36991	0.69080	5.31	0.021237	0.025782	0.16057	43.41	0.50428	0.94630
7	35	Intercept	-3.57575	0.02799	1044.82	0.000000	0.012238	0.11062	3.09	0.02254	0.03477
8	35	nonproxy	0.60946	1.83943	21.99	0.000003	0.016888	0.12995	21.32	1.42582	2.37303
9	35	females	-0.34507	0.70817	7.30	0.006909	0.016319	0.12775	37.02	0.55131	0.90966
10	59	Intercept	-3.99955	0.01832	300.47	0.000000	0.053238	0.23073	5.77	0.01166	0.02880
11	59	nonproxy	1.05756	2.87934	16.65	0.000045	0.067175	0.25918	24.51	1.73250	4.78534
12	59	females	-0.46337	0.62916	4.22	0.039865	0.050835	0.22547	48.66	0.40443	0.97878