

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

Guide de l'utilisateur du programme BOOTVARF_V30.SPS

(VERSION 3.0)

1. Introduction

Ce guide s'adresse aux utilisateurs du programme SPSS BOOTVARF_V30.SPS 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.SPS 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.SPS.

Il est à noter que le programme a été testé et fonctionne avec les versions 10.0 et 11.0 de SPSS.

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.SPS 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 pour 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.SPS

Le programme BOOTVARF_V30.SPS 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 SPSS. 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.SPS (et MACROF_V30.SPS) pour estimer les variances.

Étape 1: Création du fichier d'analyse

On crée d'abord le fichier SPSS 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.SPS` 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 SPSS contenant les variables nécessaires pour l'analyse. Un exemple de programme permettant de créer ce fichier est inclus à l'annexe A (programme `ETAPE1.SPS`).

Étape 2: Calcul des variances à l'aide du programme `BOOTVARF_V30.SPS`

Une fois le nouveau fichier SPSS d'analyse créé à l'étape 1, l'étape suivante consiste à exécuter le programme `BOOTVARF_V30.SPS`. 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.SPS`. `MACROF_V30.SPS` 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.SPS`.* Des changements peuvent être nécessaires seulement dans quelques cas particuliers expliqués plus loin.

Le programme BOOTVARF_V30.SPS 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.SPS

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 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.SPS, selon ses besoins (voir les notes incluses dans le programme BOOTVARF_V30.SPS 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.SPS`.

Résultats obtenus avec `BOOTVARF_V30.SPS`

Les résultats suivants sont obtenus après l'exécution de `BOOTVARF_V30.SPS` 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, 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 %

Les résultats suivants sont obtenus après l'exécution de `BOOTVARF_V30.SPS` 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 %
CIU95:	Borne supérieure de l'intervalle de confiance à 95 %

L'annexe A contient le programme `BOOTVARF_V30.SPS`, précédé d'un exemple de programme à utiliser pour préparer le fichier d'analyse (`ETAPE1.SPS`). 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).

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

```
*****
*                               ÉTAPE1.SPS                               *
*                               *                                         *
* Ce programme permet de créer le fichier de données SPS contenant les variables *
* nécessaires pour utiliser le programme BOOTVARF_V30.SPS                *
*                               *                                         *
* IL N'EST PAS NÉCESSAIRE D'UTILISER CE PROGRAMME. Les variables          *
* dichotomiques peuvent être créées directement dans la fenêtre d'édition des *
* données et le fichier d'analyse peut-être créé à partir de là.         *
*****
```

***** IMPORTANT: Les fichiers de données contiennent des points (.) pour représenter les décimales. *****

***** Si l'utilisateur travaille avec un système d'exploitation en français, il est important de s'assurer que *****

***** le symbole décimal par défaut du système d'exploitation est le point, et non la virgule. Par exemple, *****

***** dans un environnement Windows, le symbole décimal est spécifié dans le Panneau de Configuration, *****

***** sous Options Régionales puis Nombres. *****

* Ne pas modifier le bloc d'instruction suivant.
FILE HANDLE INFILE/NAME = !QUOTE(!DataIn).
INCLUDE FILE=!QUOTE(!Layout).
SET PRI ON.

* Création des variables dichotomiques (-> REMPLACER les instructions suivantes (jusqu'à "SAVE") *

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

execute

SAVE OUTFILE !QUOTE(!DataOut)

/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.SPS. *
 * 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. *

* Remplacer "SAVE" par "XSAVE" dans la commande précédente si des estimations ponctuelles sont effectuées à ce moment.

*** NE PAS OUBLIER D'EXÉCUTER LES DERNIÈRES LIGNES DU PROGRAMME (APRÈS L'EXEMPLE) ***.

* Calcul des estimations ponctuelles *

* Conseillé, mais non obligatoire... *

*** Exemples: ***.

WEIGHT BY **variable_de_poids**.

FREQ VAR= **liste de variables**.

LOGISTIC REG VAR = **var** WITH **liste de variables** /PRI SUM
 /CRI BCO(0.0001) LCO(0) ITE(25) EPS(10E-12).

WEIGHT OFF.

* NE PAS effacer les instructions suivantes

SET PRI NO.

!ENDDFINE /* !Go.

SET MPR ON.

!Go.

SET MPR NO.

Programme BOOTVARF V30.SPS

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...) ***
***                               ***
*****

DEFINE !Bootvar()

PRESERVE.
SET COM NO HEA ON MES ON ERR ON RES ON JOU ON LEN NONE WID 132 PRI NO MPR NO MITERATE=3000.

!Let !Tot=!Null
!Let !Reg=!Null

*****
* VEUILLEZ INDiquer LE NOM DU FICHIER D'ANALYSE (CRÉÉ À L'ÉTAPE 1) (Répertoire et nom du fichier) *
*****

!LET !Mfile=Répertoire_et_nom_du_fichier_d'analyse_étape_1' /* <- (ex: c:\data\analyse.sav).

*
*****
* VEUILLEZ INDiquer LE RÉPERTOIRE ET 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 SPSS (enlever l' " * " ) *
*****

*!LET !Bwsav=Répertoire_et_nom_du_fichier_de_poids_bootstrap_en_format_SPSS_(avec_extension)'

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

*!LET !BWTXT=Répertoire_et_nom_du_fichier_de_poids_bootstrap_en_format_ASCII_(avec_extension)'
*!LET !Bwlay=Répertoire_et_nom_du_cliché_d'article_en_syntaxe_SPSS_(avec_extension)'
*!LET !Bwsav=Répertoire_et_nom_du_fichier_de_poids_bootstrap_en_format_SPSS_(créés_dans_cette_section)
(avec_extension)'

*FILE HANDLE BOOTLAY/NAME = !QUOTE(!BWTXT).
*INCLUDE FILE=!QUOTE(!Bwlay).
*SAVE OUT=!QUOTE(!Bwsav) /UNC.

*****
* VEUILLEZ SPÉCIFIER LA (LES) VARIABLE(S) DE VENTILATION, SI DÉSIRÉ (EX: PROVINCE, SEXE, ETC...): *
* Écrire le nom de la(les) variable de ventilation ci-dessous. *
* *
* - Si l'analyse est faite pour l'ensemble des données incluses dans le fichier créé à l'étape 1, mettre deux *
* apostrophes ( %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_vide_(garder_les_apostrophes)''

```



```
*****
* VEUILLEZ INDiquer L'INFORMATION SUIVANTE (SPÉCIFIQUE À L'ENQUÊTE QUE VOUS UTILISEZ): *
* Vous devez spécifier: *
* 1- La(les) variable(s) d'identification unique (entre apostrophes, 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) identification unique '
!LET !fwgt = poids_final
!LET !prefixe = 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.SPS SI L'USAGER N'A PAS FAIT DE MODIFICATIONS) *
*****
```

```
INCLUDE FILE='nom_du_répertoire_de_macrof_v30.sps\MACROF_V30.SPS'.
```

```
* NE PAS MODIFIER LA LIGNE SUIVANTE. (ALLER À LA PARTIE 2)
```

```
!Prepares ident=!ident /fwgt=!fwgt /prefixe=!prefixe /B=!B /Classes=!Classes /mfile=!Mfile /bwsav=!bwsav.
```

```
*****
*** 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 !B !Classes/ Var = nom_de_la_variable.
*!Let !Tot=1.
```

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

```
*!Rapport !B !Classes/ Varlist = variable_au_numérateur variable_au_dénominateur .
*!Let !Tot=1.
```

```
* POUR OBTENIR L'ESTIMATION DE LA VARIANCE D'UNE DIFFÉRENCE DE RAPPORTS, EXÉCUTER:
```

```
**** ATTENTION: Voir la remarque au début de la partie 2... ****
```

```
*!Diffrat !B !Classes/ Varlist = variable1 variable2 variable3 variable4 .
*!Let !Tot=1.
```

```
* où: variable1 : la variable au numérateur du 1er rapport *
* variable2 : la variable au dénominateur du 1er rapport *
* variable3 : la variable au numérateur du 2e rapport *
* variable4 : la variable au dénominateur du 2e rapport *
```

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

```
*****
```

```
*!Regress !B !Classes/ Dep = variable_dépendante Indep = variable(s)_indépendante(s)
*!Let !Reg=1.
```

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

```
*!Log_Reg !B !Classes/ Cri=SAS Dep = variable_dépendante Indep = variable(s)_indépendante(s)
*!Let !Reg=1.
```

```
* NOTE: Le critère d'itération par défaut est différent entre SAS et SPSS.
*       Utiliser Cri=SAS pour utiliser les valeurs par défaut de SAS.
*       Indep= doit être spécifié à la fin de l'énoncé puisque le nombre de variables indépendantes est variable.
```

```
*****
*       NE PAS MODIFIER LES LIGNES SUIVANTES       *
*****
```

```
SET PRI NO MPR NO.
```

```
**** Affichage des résultats****.
!!F (!Tot=1) !THEN
+ !Print_T !Classes.      /* Affichage des résultats pour des Totaux, Rapports et Différences de Rapports
!!FEND
!!F (!Reg=1) !THEN
+ !Print_R !Classes.      /* Affichage des résultats pour des Régressions Linéaires et Logistiques
!!FEND
*****
```

```
!Stop !NBlocks.
!ENDDDEFINE /* !Bootvar.
```

```
SET MPR ON.
!Bootvar.
SET MPR NO.
```

```
*** Fin du programme BOOTVARF_V30.SPS ***.
```

ANNEXE B

Voici un exemple complet d'utilisation du programme BOOTVARF_V30.SPS. Dans un premier temps, le fichier d'analyse est créé (étape 1). Ensuite, BOOTVARF_V30.SPS 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'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	Variables d'identification	Variables de poids (dans le fichier de données)	Variables de poids (dans le fichier de poids boot)	Préfixe des poids bootstrap	# de poids
Fichier Général :							
Cycle 3	H35.txt	B5H35	REALUKEY PERSONID	WT58 (M) WT58_S (S)	FWGT	BSW	500

Étape 1:

```

*****
*
*                               ÉTAPE1.SPS
*
* Ce programme permet de créer le fichier de données SPS contenant les variables
* nécessaires pour utiliser le programme BOOTVARF_V30.SPS
*
* IL N'EST PAS NÉCESSAIRE D'UTILISER CE PROGRAMME. Les variables
* dichotomiques peuvent être créées directement dans la fenêtre d'édition des
* données et le fichier d'analyse peut-être créé à partir de là.
*****

```

SET COM NO HEA ON MES ON ERR ON RES ON JOU ON LEN NONE WID 132 PRI NO MPR NO.

* Création du fichier de données SPSS 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.

***** IMPORTANT: Les fichiers de données contiennent des points (.) pour représenter les décimales. *****
 ***** Si l'utilisateur travaille avec un système d'exploitation en français, il est important de s'assurer que *****
 ***** le symbole décimal par défaut du système d'exploitation est le point, et non la virgule. Par exemple, *****
 ***** dans un environnement Windows, le symbole décimal est spécifié dans le Panneau de Configuration, *****
 ***** sous Options Régionales puis Nombres. *****

```

DEFINE !Go()
!LET !Layout = 'D:\LAYOUT\h35_i.sps'
!LET !DataIn = ' D:\Data\h35.txt'
!LET !DataOut = ' C:\BOOTVARF\diabete.sav' /* <- (fichier d'analyse utilisé avec BOOTVARF_V30.SPS)

* Ne pas modifier le bloc d'instruction suivant.
FILE HANDLE INFILE/NAME = !QUOTE(!DataIn).
INCLUDE FILE=!QUOTE(!Layout).
SET PRI ON.

*****
* Création des variables dichotomiques ( -> REMPLACER les instructions suivantes (jusqu'à "SAVE" ) *
* (des exemples sont présentés ci-dessous, utilisant les variables de l'ENSP, cycle 3) *
*****

```

NUMERIC diab nonproxy total hommes femmes hdiab fdiab.

```

IF (ccc8_1j=1) diab=1.
IF (ccc8_1j=1) diab=0.

compute nonproxy = 0.
IF (am58_pxy>2) nonproxy=$SYSMIS.
IF (am58_pxy=2) nonproxy=1.

compute total = 1.

IF (dhc8_sex=1) hommes=1.
IF (dhc8_sex=1) hommes=0.
IF (dhc8_sex=2) femmes=1.
IF (dhc8_sex=2) femmes=0.

compute hdiab = hommes*diab.
compute fdiab = femmes*diab.

SELECT IF (prc8_cur = 10 or prc8_cur = 24 or prc8_cur = 35 or prc8_cur = 59).

execute.

*****
SAVE OUTFILE !QUOTE(!DataOut)
/KEEP = diab total hommes femmes hdiab fdiab nonproxy wt58 realukey personid prc8_cur .

* Il est recommandé de ne conserver que les variables nécessaires afin
* de diminuer le temps d'exécution du programme BOOTVARF_V30.SPS.
* 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.

```

Annexe B

* Remplacer "SAVE" par "XSAVE" dans la commande précédente si des estimations ponctuelles sont effectuées à ce moment.

*** NE PAS OUBLIER D'EXÉCUTER LES DERNIÈRES LIGNES DU PROGRAMME (APRÈS L'EXEMPLE) ***.

* Calcul des estimations ponctuelles *

* Conseillé, mais non obligatoire... *

*** Exemples: ***.

SORT CASES BY prc8_cur.

SPLIT FILE BY prc8_cur.

WEIGHT BY wt58.

FREQ VAR=diab hdiab fdia.

LOGISTIC REG VAR=diab WITH nonproxy femmes /PRI SUM
/CRI BCO(0.0001) LCO(0) ITE(25) EPS(10E-12).

WEIGHT OFF.

SPLIT FILE OFF.

* NE PAS effacer les instructions suivantes

SET PRI NO.

!ENDDFINE /* !Go.

SET MPR ON.

!Go.

SET MPR NO.

Etape 2 - Programme BOOTVARF V30.SPS:

```

. . .
. . .
. . .

```

```

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

DEFINE !Bootvar()

PRESERVE.
SET COM NO HEA ON MES ON ERR ON RES ON JOU ON LEN NONE WID 132 PRI NO MPR NO MITERATE=3000.

!Let !Tot=!Null
!Let !Reg=!Null

*****
* VEUILLEZ INDIQUER LE NOM DU FICHIER D'ANALYSE (CRÉÉ À L'ÉTAPE 1) (Répertoire et nom du fichier) *
*****

!LET !file='C:\BOOTVAR\diabete.sav'                /* <- (ex: c:\data\analyse.sav).

*****
* VEUILLEZ INDIQUER LE RÉPERTOIRE ET 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 SPSS (enlever l' " * ") *
*****

*!LET !BWsav='D:\BOOTSTRP\DATA\b5h35.sav'

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

*!LET !BWTXT=' Répertoire_et_nom_du_fichier_de_poids_bootstrap_en_format_SPSS_(avec_extension) '
*!LET !BWay=' Répertoire_et_nom_du_cliclé_d'article_en_syntaxe_SPSS_(avec_extension) '
*!LET !Bwsav=' Répertoire_et_nom_du_fichier_de_poids_bootstrap_en_format_SPSS_(créés_dans_cette_section)_(avec_extension) '

*FILE HANDLE BOOTLAY/NAME = !QUOTE(!BWTXT).
*INCLUDE FILE=!QUOTE(!BWay).
*SAVE OUT=!QUOTE(!BWsav) /UNC.

*****
* VEUILLEZ SPÉCIFIER LA (LES) VARIABLE(S) DE VENTILATION, SI DÉSIRÉ (EX: PROVINCE, SEXE, ETC...): *
* Écrire le nom de la(les) variable de ventilation ci-dessous. *
* *
* - Si l'analyse est faite pour l'ensemble des données incluses dans le fichier créé à l'étape 1, mettre deux *
* apostrophes ( %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'ENQUETE QUE VOUS UTILISEZ): *
* Vous devez spécifier: *
* 1- La(les) variable(s) d'identification unique (entre apostrophes, 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 !prefixe  = bsw
!LET !B        = 500

```

```

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

```

```

INCLUDE FILE='C:\BOOTVAR\MACROF_V30.SPS'.

```

```

* NE PAS MODIFIER LA LIGNE SUIVANTE. (ALLER À LA PARTIE 2)
*****

```

```

!Prepare ident=!ident /fwgt=!fwgt /prefixe=!prefixe /B=!B /Classes=!Classes /mfile=!Mfile /bwsav=!bwsav.

```

```

*****
***                                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 !B !Classes/ Var = nom_de_la_variable.
*!Let !Tot=1.

```

```

!Total !B !Classes/ Var = diab.
!Total !B !Classes/ Var = hdiab.
!Let !Tot=1.

```

```

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

```

```

*!Rapport !B !Classes/ Varlist = variable_au_numérateur variable_au_dénominateur .
*!Let !Tot=1.

```

```

!Rapport !B !Classes/ Varlist = diab total.
!Rapport !B !Classes/ Varlist = hdiab hommes.
!Let !Tot=1.

```

```

* POUR OBTENIR L'ESTIMATION DE LA VARIANCE D'UNE DIFFÉRENCE DE RAPPORTS, EXÉCUTER:
*****

```

```

**** ATTENTION: Voir la remarque au début de la partie 2... ****

```

```

*!Diffrat !B !Classes/ Varlist = variable1 variable2 variable3 variable4 .
*!Let !Tot=1.

```

```

*      où:  variable1 : la variable au numérateur du 1er rapport *
*           variable2 : la variable au dénominateur du 1er rapport *
*           variable3 : la variable au numérateur du 2e rapport *
*           variable4 : la variable au dénominateur du 2e rapport *

```

```

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

```

```

*!Regress !B !Classes/ Dep = variable_dépendante Indep = variable(s)_indépendante(s)
*!Let !Reg=1.

```

```

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

```

```

*!Log_Reg !B !Classes/ Cri=SAS Dep = variable_dépendante Indep = variable(s)_indépendante(s)
*!Let !Reg=1.

```

```

!Log_Reg !B !Classes/ Cri=SAS Dep = diab Indep = nonproxy femmes
!Let !Reg=1.

```

- * NOTE: Le critère d'itération par défaut est différent entre SAS et SPSS.
- * Utiliser Cri=SAS pour utiliser les valeurs par défaut de SAS.
- * Indep= doit être spécifié à la fin de l'énoncé puisque le nombre de variables indépendantes est variable.

```
*****
*   NE PAS MODIFIER LES LIGNES SUIVANTES   *
*****
SET PRI NO MPR NO.

**** Affichage des résultats****.
!IF (!Tot=1) !THEN
+ !Print_T !Classes. /* Affichage des résultats pour des Totaux, Rapports et Différences de Rapports
!IFEND
!IF (!Reg=1) !THEN
+ !Print_R !Classes. /* Affichage des résultats pour des Régressions Linéaires et Logistiques
!IFEND
*****

!Stop !NBlocks.
!ENDDEFINE /* !Bootvar.

SET MPR ON.
!Bootvar.
SET MPR NO.

*** Fin du programme BOOTVARF_V30.SPS ***.
```


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 la douzième ligne dans le tableau. 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* et au dénominateur la variable *hommes* (VARIABLES). 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 (neuvième ligne) 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 7,9964 (WALD) et $p=0,004687$ (PVALUE) respectivement. L'estimation de l'écart-type du paramètre estimé est 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.

21 Mar 02 Estimation de la Variance [500 poids bootstrap]

PRC8_CUR	TYPE.	VARIABLE	YHAT	BS_SD	BS_CV	CIL95	CIU95
10	Total	DIAB	20741.32	1778.61	8.58	17255.24	24227.40
10	Total	HDIAB	7029.13	1356.31	19.30	4370.75	9687.51
10	Rapport	DIAB TOTAL	3.85	.33	8.58	3.21	4.50
10	Rapport	HDIAB HOMMES	2.63	.51	19.30	1.64	3.63
24	Total	DIAB	205292.27	16330.49	7.95	173284.50	237300.04
24	Total	HDIAB	110452.80	10818.85	9.79	89247.85	131657.75
24	Rapport	DIAB TOTAL	2.87	.23	7.95	2.42	3.32
24	Rapport	HDIAB HOMMES	3.12	.31	9.80	2.52	3.72
35	Total	DIAB	362439.60	20692.51	5.71	321882.28	402996.92
35	Total	HDIAB	198237.67	16369.18	8.26	166154.08	230321.26
35	Rapport	DIAB TOTAL	3.22	.18	5.71	2.86	3.58
35	Rapport	HDIAB HOMMES	3.57	.29	8.26	2.99	4.14
59	Total	DIAB	110375.39	10661.17	9.66	89479.50	131271.28
59	Total	HDIAB	62808.64	8301.60	13.22	46537.50	79079.78
59	Rapport	DIAB TOTAL	2.83	.27	9.66	2.29	3.37
59	Rapport	HDIAB HOMMES	3.24	.43	13.22	2.40	4.08

21 Mar 02 Estimation de la Variance [500 poids bootstrap]

PRC8_CUR	TYPE	BETA	BHAT	ODDS	WALD	PVALUE	BS_SD	BS_CV	CIL95	CIU95
10	Regression Logistique	Const_	-4.00379	.01825	255.3510	.00000000	.25055	6.26	.01117	.02982
10	Regression Logistique	NONPROXY	.85788	2.35815	11.8177	.00058671	.24955	29.09	1.44594	3.84587
10	Regression Logistique	FEMMES	.46627	1.59403	2.8588	.09087612	.27577	59.14	.92845	2.73674
24	Regression Logistique	Const_	-3.89894	.02026	852.2376	.00000000	.13356	3.43	.01560	.02633
24	Regression Logistique	NONPROXY	.89123	2.43812	34.0970	.00000001	.15263	17.13	1.80774	3.28831
24	Regression Logistique	FEMMES	-.36991	.69080	5.7831	.01618064	.15382	41.58	.51100	.93387
35	Regression Logistique	Const_	-3.57575	.02799	961.1974	.00000000	.11533	3.23	.02233	.03510
35	Regression Logistique	NONPROXY	.60946	1.83943	23.5198	.00000124	.12567	20.62	1.43785	2.35318
35	Regression Logistique	FEMMES	-.34507	.70817	7.9964	.00468716	.12203	35.36	.55753	.89952
59	Regression Logistique	Const_	-3.99966	.01832	301.8469	.00000000	.23021	5.76	.01167	.02877
59	Regression Logistique	NONPROXY	1.05766	2.87963	18.8915	.00001384	.24334	23.01	1.78732	4.63951
59	Regression Logistique	FEMMES	-.46337	.62916	4.5912	.03213705	.21625	46.67	.41180	.96126

```

-----
Temps de départ:      21-MAR-2002 13:45:22.048
Temps de fin :       21-MAR-2002 14:35:20.700
Temps pour exécuter le programme:      0:49:58.652
-----

```