



PISA 2003

Echantillon de classes de 9^e année en Suisse

Introduction

Ce document est une brève introduction à l'utilisation de l'échantillon de classes de 9^e année dans PISA 2003.

Vous avez téléchargé les fichiers suivants :

1. Brève introduction (word; introduction_PUS_f.doc).
2. Manual for the PISA 2003 Database (pdf; Data_analysis_SPSS.pdf).
3. Questionnaire destiné aux élèves en français ou allemand (pdf; StdQ_CH_FR.pdf ou StdQ_CH_DE.pdf).
4. Série de données sur les élèves – échantillon de classes de 9^e année (SPSS; Students_CH.sav).
5. Livre de codage (codebook) pour la série de données sur les élèves – échantillon de classes de 9^e année (PUS) (excel; Student_codebook.xls).

Nous attirons votre attention sur le fait que la série de données des élèves est un échantillon à usage public (PUS, Public Use Sample). Cet échantillon PUS est un fichier réduit, qui a été traité avec des méthodes statistiques pour minimiser les risques d'identification des individus. Un PUS est nettement plus petit et moins riche en informations qu'une série de données initiale, car plusieurs variables ont été supprimées, décodées ou traitées.

Ne sont pas contenues les données des jeunes de 15 ans qui sont utilisées pour les comparaisons internationales. Ces données peuvent être téléchargées ci-après :

http://pisaweb.acer.edu.au/oe.cd_2003/oe.cd_pisa_data_s1.html

Les fichiers sont décrits brièvement ci-après :

1. Brève introduction

Il s'agit du présent document.

Echantillon de classes de 9^e année en Suisse, PISA 2003

2. Manual for the PISA 2003 Database

La structure des données de PISA est complexe. Il faut notamment en tenir compte lors du calcul des erreurs standard. On recourra par conséquent à ce manuel qui, en plus de fournir les bases théoriques, décrit la syntaxe SPSS pour de nombreux calculs.

3. Questionnaire destiné aux élèves

Un livre de codage est proposé dans le manuel PISA Data Manual pour le questionnaire international.

4. Série de données des élèves

INFO : la série de données initiale des élèves a été utilisée pour les analyses des deux premiers rapports nationaux sur PISA 2003 (premier rapport : chapitre 4).

L'échantillon PUS en question contient 156 variables et 20'972 observations. Avant d'être traitée pour produire l'échantillon PUS, la série de données avait déjà subi d'importantes modifications, à savoir :

- *l'exclusion des élèves des classes spéciales*

Les parts pondérées d'échantillons et de populations d'élèves des classes spéciales variaient selon les cantons. Pour que la comparaison entre les cantons soit la plus correcte possible, on a par conséquent exclu de l'échantillon tous les élèves des classes spéciales. La direction nationale de programme fournira sur demande des informations plus détaillées sur la répartition des élèves entre les cantons, sur le nombre des cas exclus, etc. Pour de plus amples informations : OFS / CDIP (2004): PISA 2003 : Des compétences pour l'avenir, deuxième rapport national, chapitre « Informations techniques ».

- *les variables ajoutées*

Les variables ajoutées dont la dénomination est explicite ne sont pas mentionnées ici. Les variables conservées dans le PUS pour lesquelles des explications sont nécessaires sont expliquées brièvement dans l'ordre de leur position dans la série de données :

natprog_rec	Programme scolaire national : construit à partir de variables exclues du PUS. Leurs valeurs ont été groupées dans le PUS selon le niveau d'exigences.
classid	Variable numérique présentant une valeur distincte pour chaque classe d'école.
langreg	Variable de la région linguistique (allemand, français, italien).
ses	Statut social, économique et culturel. La variable internationale « escs » n'a pas été livrée pour la série de données des classes de neuvième année en Suisse. Un indice « ses » quelque peu différent a par conséquent été créé. La variable a une moyenne de 0 et un écart standard de 1. La syntaxe peut être obtenue auprès de la direction nationale de programme.

En plus des modifications susmentionnées, d'autres changements ont été apportés pour atteindre le niveau de sécurité requis pour un PUS.

- Les variables « studid », « schoolid » et « classid » ont été anonymisées à nouveau par un renumérotage, puis classées de manière aléatoire. L'ensemble de la série de données a été classé selon « schoolid » et ensuite selon « studid ».
- La méthode PRAM (PostRandomisationMethod) est une des méthodes utilisées pour établir un échantillon PUS à partir de la série de données des élèves. La PRAM consiste en une fausse classification volontaire des données : les variables classées peuvent être attribuées à une autre catégorie selon un mécanisme de probabilité prédéfini¹. Ce processus rend encore plus difficile la réidentification des individus. Il permet en revanche de faire des estimations correctes à partir des tableaux de contingence, car les probabilités de passage sont connues. Les informations dont a besoin l'utilisateur final pour pouvoir utiliser correctement une série de données traitées avec la méthode PRAM ou les matrices présentant la probabilité de changement de catégorie figurent dans le livre de codage (cf. point 6). Signalons toutefois que des méthodes compliquées sont nécessaires pour corriger les biais provoqués par la méthode PRAM lors d'analyses complexes (par ex. modèles log-linéaires). Ces méthodes sont trop compliquées pour être présentées ici. Des références bibliographiques concernant la méthode PRAM sont fournies aux personnes intéressées.

¹ Statistics Netherlands, "μ ARGUS version 4.0, User's Manual "

Echantillon de classes de 9^e année en Suisse, PISA 2003

5. Livre de codage pour la série de données sur les élèves – échantillon de classes de 9^e année (PUS)

Ce document comprend des informations et explications détaillées sur la série de données PUS et sur ses variables, dont quelques statistiques descriptives, les matrices PRAM (cf. point 5) de la probabilité de changement de catégorie et quelques tabulations croisées qui montrent les différences de fréquence des cellules avant et après l'utilisation de la méthode PRAM.