



PISA 2003

Neuntklassstichprobe der Schweiz

Anleitung

Dieses Dokument ist eine kurze Anleitung zur Verwendung der Stichproben der neunten Klassen bei PISA-2003.

Sie haben die folgenden Dateien heruntergeladen:

1. Kurze Anleitung (word; Anleitung_PUS_d.doc).
2. Manual for the PISA 2003 Database (pdf; Data_analysis_SPSS.pdf).
3. Fragebogen für die Schülerinnen und Schüler auf Deutsch oder Französisch (pdf; StdQ_CH_DE.pdf oder StdQ_CH_FR.pdf).
4. Datensatz der Schülerinnen und Schüler - Neuntklassstichprobe (SPSS; Students_CH.sav).
5. „Codebook“ zum Datensatz der Schülerinnen und Schüler – Neuntklassstichprobe (PUS) (excel; Student_codebook.xls).

Wir weisen darauf hin, dass der Datensatz der Schülerinnen und Schüler ein „Public Use Sample“ (PUS) ist. Unter PUS versteht man ein reduziertes File, das durch statistische Methoden überarbeitet worden ist, um das Risiko der Entschlüsselung der Identität der Einzelnen zu minimieren. Im Vergleich zum ursprünglichen Datensatz ist ein PUS wesentlich kleiner und weniger reich an Informationen, weil viele Variablen entfernt, umkodiert oder behandelt wurden.

Nicht enthalten sind die Daten der 15-jährigen Jugendlichen, die für den internationalen Vergleich verwendet werden. Diese Daten stehen zum Download unter:

http://pisaweb.acer.edu.au/oeod_2003/oeod_pisa_data_s1.html

Im Folgenden werden die einzelnen Dateien kurz beschrieben:

1. Kurze Anleitung

Dieses Dokument lesen Sie gerade.

2. Manual for the PISA 2003 Database

PISA weist eine komplexe Datenstruktur auf. Insbesondere bei der Berechnung von Standardfehlern ist die Rechnung zu tragen. Es wird deshalb sehr empfohlen, dieses Handbuch zu verwenden. Es liefert nicht nur die theoretischen Grundlagen, sondern beschreibt auch die SPSS-Syntax für zahlreiche Berechnungen.

3. Fragebogen für die Schülerinnen und Schüler

Im PISA Data Manual findet sich ein Codebook für den internationalen Fragebogen.

4. Datensatz der Schülerinnen und Schüler

INFO: Der ursprüngliche Datensatz der Schülerinnen und Schüler wurde für die Analysen des ersten (Kapitel 4) und zweiten nationalen Berichts zu PISA 2003 verwendet.

Der mitgelieferte PUS enthält 156 Variablen und 20'972 Beobachtungen. Schon bevor er zum PUS überarbeitet wurde, erfuhr der Datensatz einige wichtige Modifikationen:

- Ausschluss der Schülerinnen und Schüler in Sonderklassen

Die gewichteten Stichprobenanteile und Populationsanteile von Schülerinnen und Schülern aus Sonderklassen wiesen nach Kantonen Diskrepanzen auf. Um den Vergleich zwischen den Kantonen möglichst „fair“ zu gestalten wurden deshalb alle Schülerinnen und Schüler aus Sonderklassen aus der Stichprobe entfernt. Details zur Verteilung dieser Schülerinnen und Schüler nach Kantonen, die Anzahl der ausgeschlossenen Fälle, etc. können auf Wunsch bei der nationalen Programmleitung bezogen werden. Für ausführlichere Informationen siehe: BFS/EDK (2004): „PISA 2003: Kompetenzen für die Zukunft“, zweiter nationaler Bericht, Kapitel „Technische Informationen“.

- angefügte Variablen

Falls die angefügten Variablen durch die Labels selbsterklärend sind, werden sie hier nicht speziell erwähnt. Die Variablen, welche im PUS geblieben sind und Erklärungen benötigen, werden in der Reihenfolge der Positionen im Datensatz kurz erläutert:

natprog_rec	Nationales Schulprogramm: wurde aus Variablen gebaut, die vom PUS entfernt worden sind. Im PUS sind deren Werte nach Niveau der Ansprüche gruppiert worden.
classid	Numerische Variable, die für jede Schulklasse einen distinkten Wert aufweist.
langreg	sprachregionale Variable (Deutsch, Französisch, Italienisch)
ses	sozialer, ökonomischer und kultureller Status. Die internationale Variable „escs“ wurde für den schweizerischen Datensatz der neunten Klassen nicht mitgeliefert. Es wurde daher ein etwas anderer Index „ses“ gebildet. Die Variable hat einen Mittelwert von 0 und eine Standardabweichung von 1. Die Syntax kann bei der nationalen Programmleitung bezogen werden.

Ausser den oben aufgeführten Änderungen wurden zusätzliche Modifikationen vorgenommen, die nötig waren, um das Sicherheitsniveau zu erreichen, das für ein PUS verlangt wird.

- Die Variablen „studid“, „schoolid“ und „classid“ wurden mittels Durchnummerierung weiter anonymisiert und zufällig geordnet. Der ganze Datensatz ist nach „schoolid“ und anschliessend nach „studid“ geordnet worden.

- Zu den Methoden, die verwendet wurden, um den Datensatz der Schülerinnen und Schüler in ein PUS zu verwandeln, zählt unter anderen die PRAM (PostRAndomisationMethod). PRAM besteht in einer absichtlichen Miss-Klassifizierung: kategorische Variablen können von einer Kategorie zu einer anderen übergehen, nach einem vorherbestimmten Wahrscheinlichkeitsmechanismus¹. Die Umklassifizierung macht es schwieriger, die Einzelnen zu re-identifizieren. Gleichzeitig bleibt es möglich, korrekte Schätzungen der Kontingenztabellen zu bekommen, weil die Übergangswahrscheinlichkeiten bekannt sind. Die Informationen, die der Endbenutzer braucht, um einen durch PRAM behandelten Datensatz richtig verwenden zu können bzw. die Matrizen der Wahrscheinlichkeit des Kategorienwechsels sind im „Codebook“ enthalten (siehe Punkt 6). Man muss trotzdem darauf hinweisen, dass komplizierte Methoden nötig sind, um die von PRAM verursachten Verzerrungen in komplexen Analysen (z.B. loglinear Models) zu korrigieren. Diese Methoden werden aufgrund Ihres Schwierigkeitsgrades nicht illustriert. Für die Interessierten wird eine Bibliographie bezüglich PRAM mitgeliefert.

5. Codebook“ zum Datensatz der Schülerinnen und Schüler – Neuntklassstichprobe (PUS)

Dieses Dokument enthält ausführlichere Informationen und Erklärungen zum PUS Datensatz und dessen Variablen, darunter einige Deskriptivstatistiken, die PRAM (siehe Punkt 5) Matrizen der Wahrscheinlichkeit des Kategorienwechsels und einige Crosstabulationen, welche die Unterschiede in der Häufigkeit der Zellen vor und nach dem PRAM aufzeigen.

¹ Statistics Netherlands, “μ ARGUS version 4.0, User’s Manual ”