



Pisa 2012 Almo-Daten und Analyse-Programme

Kurt Holm

Almo Statistik-System

www.almo-statistik.de

holm@almo-statistik.de

kurt.holm@jku.at

2013

Die Daten zur Pisa-Studie 2012 liegen nunmehr im Almo-Format vor.

Diese Almo-Pisa-Daten und einige Auswertungs-Programme können in Almo herunter geladen werden. Sie befinden sich in folgenden beiden "gezipten" Paketen

Pisa2012
Pisa2012_Math_Read_Science

Das Paket Pisa2012

Es enthält folgenden 2 Ordner

1. **Daten.** Enthalten sind die 2 Dateien

"Int_Stu_A_CH_D_NL.dir". Dies sind die Daten aus der Schülerbefragung für die Länder

A Österreich
CH Schweiz
D Deutschland
NL Niederlande

Die Datei umfasst 646 Variable und 25445 Schüler.

"Int_Stu.nam". Dies ist die Datei der Variablennamen

Enthalten sind auch die 2 Programm-Masken

Erzeuge_fre_Datei.Alm. Erzeugt aus der oben genannten Datendatei eine Datei im lesbaren Format FREI (ASCII)

Erzeuge_A_oder_CH_oder_D_oder_NL.Alm. Erzeugt aus der oben genannten Datendatei eine Datei für ein einzelnes Land im binären Format DIREKT und im lesbaren Format FREI

2. **Programme.** Enthalten sind folgende Beispielprogrammen und ihre Ergebnisse

Land_mit_MeanPV.Alm.

Mit "MeanPV" ist der Mittelwert der "plausible values" aus Mathematik, Lesen und Wissenschaft gemeint. Dies ist der in der Öffentlichkeit bekannte "Pisa-Index". Für die 4 Länder wird dieser Wert ermittelt.

Land_u_Immigrat_mit_MeanPV.Alm.

Ermittelt werden die "MeanPV" je Land und (innerhalb des Landes) Immigrations-Status

Land_u_Region_mit_MeanPV.

Ermittelt werden die "MeanPV" je Land und Region (für Deutschland beispielsweise je Bundesland)

Land_u_LanguageHome_mit_MeanPV.Alm

Ermittelt werden die "MeanPV" je Land und (innerhalb des Landes) Language.spoken.at.home

Generierte Variable

Die "Pisa-Indices" für Mathematik, Lesen, Wissenschaft und für die 7 speziellen Mathematik-Dimensionen bestehen je aus 5 "plausible values" Dies sind die Variablen (siehe die Datei der Variablennamen "Int-Stu.nam")

V501 bis 505 für Mathematik
V541 bis 545 für Lesen
V546 bis 550 für Wissenschaft
und insgesamt V506 bis V540 für die 7 speziellen Mathematik-Dimensionen

Der eigentliche "Pisa-Index" wird als Mittelwert aus diesen 5 "plausible values" gebildet. Sie werden in den Almo-Daten generiert und erhalten die Nummern V636 bis V645 und folgende Variablennamen

Name 636 =Mathematik;
Name 637 =Lesen;
Name 638 =Wissenschaft;

Name 639 =Mathe.Change.and.Relationship;
Name 640 =Mathe.Quantity;
Name 641 =Mathe.Space.and.Shape;
Name 642 =Mathe.Uncertainty.and.Data;
Name 643 =Mathe.Employ;
Name 644 =Mathe.Formulate;
Name 645 =Mathe.Interpret;

Dann wird noch eine weitere Variable V646 generiert - mit dem Namen

Name 646 =Land:
(008) ALB Albania,
(784) ARE United Arab Emirates,
(032) ARG Argentina,
(036) AUS Australia,
(040) AUT Austria,
usw.

Dies ist die Variablen 1 Country umkodiert auf numerische Werte

**Die Datei "Int_Stu_A_CH_D_NL.dir"
und
das Programme "Erzeuge_fre_Datei.Alm"**

1. Die Datei **Int_Stu_A_CH_D_NL.dir** enthält die Datei aus der Schüler-Befragung für die Länder
 - a. A Österreich
 - b. CH Schweiz
 - c. D Deutschland
 - d. NL Niederlande

im Almo-Format "direkt". Die Daten sind also in binärer, nicht lesbarer Form geschrieben.

2. Die Datei umfasst 646 Variable und 25445 Schüler.
3. Mit dieser Datei werden die Auswertungsprogramme gerechnet. Dabei sollten in der 1. Eingabe-Box 660 Variable (oder mehr) vereinbart werden. Die Datei umfasst zwar nur 646 Variable. Dem Benutzer stehen so jedoch noch 14 freie Variable zur Verfügung. Benötigt er mehr, dann braucht er nur eine höhere Variablenzahl zu vereinbaren.
4. Das Programm **Erzeuge_fre_Datei.Alm** erzeugt aus dieser Direkt-Datei eine

Datei **Int_Stu_A_CH_D_NL.fre** im gewohnten lesbaren Format. Die Datei umfasst 646 Variable. Die Variablen V1, V635 sind Zeichenvariable (Worte) V1 ist das Land - kodiert mit AUT, DEU, CHE, NLD (siehe die Datei der Variablennamen **Int_Stu.nam**)

Ein Datensatz, also ein Schüler wird mit 500 Zeichen je Zeile in insgesamt 6 bis 7 Zeilen geschrieben. Wird die Datei geöffnet, dann belegt sie im Almo-Fenster 159089 Zeilen. Wenn der Computer zu knapp mit Speicher ausgestattet ist, dann kann Almo die Datei nicht laden. 8 GB genügen aber auf jeden Fall.

3. Will der Benutzer die Datei in Almo öffnen und anschauen, dann muss er zuvor durch Klick auf das Menü "Allerlei/Einstellungen" im Eingabefeld "wieviel Zeichen je Zeile" mindestens 502 eintragen. Dann Klick auf den Knopf "Speichern" und Almo neu starten.

Das Programm "Erzeuge_Stu_A_oder_CH_oder_D_oder_NL.Alm"

1. Dieses Programm erzeugt aus der Datei "Int_Stu_A_CH_D_NL.dir" eine Ein-Land-Datei für entweder A oder CH oder D oder NL. Dabei wird eine "dir"-Datei im binären, nicht lesbaren Format DIREKT und eine "fre"-Datei im lesbaren Format FREI erzeugt. Mit der "dir"-Datei sollten dann die Auswertungsprogramme gerechnet werden.

Die Datei der Variablennamen "Int_Stu.nam"

1. In Almo sind die Variablennamen in der Datei "Int_Stu.nam" enthalten. Diese Datei kann in ein Fenster geladen werden und beliebig verändert werden. So kann der Benutzer z.B. deutschsprachige Variablennamen für die englischen einsetzen.
2. Die Variablen erhalten in der Pisa-Studie kurze Variablennamen. So erhält beispielsweise die 376. Variable den Namen EC05Q01. Da diese Variablennamen nichtssagend sind, werden sie durch einen "label" (eine kurze inhaltliche Erklärung) ergänzt. In Almo wird der Kurz-Namen und der "label" zu einem Variablennamen zusammengefasst. Der Vorteil ist, dass man so sofort weiß, was sich inhaltlich hinter dem Variablennamen verbirgt. Almo verwendet jedoch nur die ersten 80 Zeichen. Der Rest wird abgeschnitten. In der Dialogbox zur Variablen-Auswahl werden davon nur die ersten 32 gezeigt.

Beispiel: Die 376. Variable EC05Q01 hat in Almo den Variablennamen

Name 376 = EC05Q01.First.language.learned

Zwischen Kurzname und label wird ein Punkt eingefügt. Blanks sind nicht erlaubt. An ihrer Stelle wird ein Punkt geschrieben.

Auf den Variablennamen folgen die Ausprägungs-Namen

Beispiel:

Name 376 = EC05Q01.First.language.learned:

- (1)test language.or.other official national language(s) or dialect,
- (2)test language.or.other official national language(s) or dialect,
- (3)other than.test language.or.other official national language,
- (7)N/A,(8)Invalid,(9)Missing;

Nach dem Variablennamen und einem Doppelpunkt folgen die Ausprägungsnamen. In diesen darf ein Blank vorkommen.

Das Paket "Pisa2012_Math_Read_Science"

Es enthält die Daten der Testfragen zu Mathematik, Lesen, Wissenschaft für die Länder

A Österreich
CH Schweiz
D Deutschland
NL Niederlande

In den originalen OECD-Daten wird diese Datei "Int_Cog12_S_Dec03.txt" genannt. Die Datei umfasst 121 MB. Wir haben aus ihr die oben angegebenen 4 europäischen Länder herausgeschnitten, und dabei die Daten in das Almo-Format übertragen.
Die Datei erhielt den Namen:

Int_CogS_A_CH_D_NL.dir

Für Almo wurde auch eine Datei der Variablennamen erzeugt. Sie erhielt den Namen

Int_CogS_A_CH_D_NL.nam

Unser Ziel ist es nun, ein Rasch-Modell für die Mathe-Items des Booklet10 für Deutschland mit echten (nicht imputierten) Daten zu rechnen.

Es wird folgende Vorgehensweise empfohlen

1. Mit dem Almo-Programm "Erzeuge_CogS_A_CH_D_NL_in_fre.Alm" kann (aber muss nicht) eine parallele Datei im lesbaren Format ".fre", d.h. FREI (=ASCII) erzeugt werden. Dazu wird die Programm-Maske Prog00mp verwendet. Das ist nur notwendig, wenn der Benutzer die Daten anschauen möchte. Mit diesem Programm sollte aber nicht gerechnet werden, da die Nummerierung der Items eine andere ist.

2. Mit dem Almo-Programm "Haeufig_D_Booklet10.Alm" wird für Deutschland (oder nach Wahl des Benutzers eines der anderen 3 Länder) die Häufigkeit der Mathe-Items für Booklet 10 ausgezählt. Dazu wird die Programm-Maske Prog05m4 verwendet (erreichbar über Knopf "Verfahren", dann Häufigkeitsverteilung) Beispielsweise könnte der Benutzer auch die Daten aus Deutschland und Österreich zusammengefasst analysieren. in der Eingabe-Box "Ein- Und Ausschliessen von Untersuchungsdaten" wird dann geschrieben

Country=276 #=Germany# und Country=40 #Austria# und BOOKID.Booklet=10

Diejenigen Mathe-Items, die als Wert ausschliesslich "Kein-Wert" besitzen, gehören nicht zum Booklet 10. Alle die in den Codes 0 oder 1 oder 2 Besetzungen aufweisen gehören dazu. Es sind dies

V11:17,19,23:26,28:34,37,42:46,51,52,56,57,101:104,110:112

Soll ein anderes Booklet analysiert werden, dann muss entsprechend verfahren werden, um festzustellen, welche Items zu diesem Booklet gehören - es sei denn, der Benutzer

findet in der Dokumentation der Pisa-Studie eine Aufteilung der Items in Booklets.

3. Nun können wir ein Rasch-Modell rechnen. Das geschieht mit dem Algo-Programm "Rasch_Mathe_Booklet10.Alm". Dazu wird die Programm-Maske Prog14m4 verwendet (erreichbar über Knopf "Verfahren", dann Rasch-Modell) Wie so oft erweist es sich, dass die Mathe-Items mit 3 Ausprägungen "verdrehte" Schwellenwerte besitzen. Es empfiehlt sich diese Items zu dichotomisieren, also die Kategorien 1 und 2 zusammenzufassen. Es sind dies V15,16,32,52. In der Umkodierungs-Box wird geschrieben
V15,16,32,52(2=1)

Wichtige Internet-Adressen:

Die Originaldaten können bei der OECD unter der Adresse <http://pisa2012.acer.edu.au/downloads.php> herunter geladen werden.

Ein "Technical Report" für Pisa 2009 findet man unter: <http://www.oecd.org/pisa/pisaproducts/pisa2009technicalreport.htm>

und eine Beschreibung der statistischen Methoden der Pisa-Studie

<http://www.oecd.org/pisa/pisaproducts/pisadataanalysismanualspssandsassecondedition.htm>

Eine sehr übersichtliche deutschsprachige Darstellung von Ergebnissen und Verfahren findet man bei

Manfred Prenzel, Christine Sälzer, Eckhard Klieme, Olaf Köller (Hrsg.):
Pisa 2012

bei

<http://www.pisa.tum.de/>
unter "Nationaler Bericht"

Angekündigt ist von

Schwantner, U. / Schreiner, C. (Hrsg.): PISA 2012. Internationaler Vergleich von Schülerleistungen. Technischer Bericht

zu finden unter

<https://www.bifie.at/buch/2431>