RALSA: Priročnik za uporabo

dr. Plamen Vladkov Mirazchiyski

RALSA: priročnik za uporabo

Avtor: dr. Plamen Vladkov Mirazchiyski
Prevajalca: Igor Peras in dr. Eva Klemenčič Mirazchiyski
Lektor: Davorin Dukič
Tehnična urednica, oblikovanje, prelom: Nina Pertoci

Izdal in založil: *Pedagoški inštitut*
Izdaja: 1. e-izdaja
Zanj: dr. *Igor Ž. Žagar*
Leto izida: 2024

Izdaja dostopna na https://www.pei.si/raziskovalna-dejavnost/projekti/odnos-osmosolcev-v-sloveniji-in-evropi-do-migrantskih-tematik/

Izid pričujočega priročnika je omogočila Javna agencija za raziskovalno in inovacijsko dejavnost (ARIS). Priročnik je nastal v okviru projekta Odnos osmošolcev v Sloveniji in Evropi do migrantskih tematik – podatki raziskave IEA ICCS 2009, 2016 in 2022 (šifra L5-4571). Pedagoški inštitut ni odgovoren za morebitne netočnosti, izpustitve delov ali razlike med tem besedilom in izvirnim besedilom v angleškem jeziku. Izvirno besedilo je dostopno na tej povezavi: https://ralsa.ineri.org/user-guide/. Paket/orodje RALSA in njegovo spletno mesto sta odprtokodna izdelka pod splošno javno licenco GNU (GPL), različica 2.0. Vse njihove dele je mogoče prosto uporabljati, kopirati, deliti, razširjati in spreminjati brez omejitev in v skladu z licenčnimi pogoji, če je naveden prvotni avtor. Vse spremembe morajo biti prav tako na voljo v skladu z enakimi pogoji licence GPL v različici 2.0. Vse spremembe lahko vzajemno uporabi tudi prvotni avtor. Zaradi razvoja programskega orodja, je izvirnik predmet dopolnitev.

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID 220620803
ISBN 978-961-270-402-5 (PDF)

Kazalo

Uvodnik

Ta priročnik je prevod priročnika *RALSA: User Guide* (v angl.), ki je dostopen na povezavi: https://ralsa.ineri.org/user-guide/.

Nastanku orodja/platforme RALSA (The R Analyzer for Large-Scale Assessments) je botrovalo to, da je dr. Plamen Vladkov Mirazchiyski iskal dostopnejše in celovitejše rešitve za analizo podatkov mednarodnih primerjalnih raziskav (znanja) ter mednarodnih primerjalnih študij na področju vzgoje in izobraževanja – zaradi kompleksnosti in dizajna vzorčenja teh raziskav pa tudi zaradi določenih razlik pri kompleksnem dizajnu in vzorčenju teh raziskav. RALSA je tako orodje za analizo podatkov, ki jih zberemo v okviru mednarodnih primerjalnih raziskav in študij na področju vzgoje in izobraževanja. Zasnova teh raziskav je kompleksna, tako v smislu vzorčenja kot v smislu zasnove kognitivnih instrumentov, ki merijo znanja na številnih področjih. Sodelujejo pa ne le učenci oz. dijaki, temveč tudi njihovi učitelji, ravnatelji šol, v nekaterih primerih tudi starši pa tudi IKT-koordinatorji na šolah. Zaradi vsega naštetega klasični programi za statistično izračunavanje niso ustrezni, ker moramo tudi pri analizah upoštevati vse te specifike. Še več, obstajajo tudi določene razlike med posameznimi mednarodnimi primerjalnimi raziskavami in študijami (tako pri zasnovi kognitivnih instrumentov in izračunih na tej podlagi kot pri samem vzorčenju). Prav zato ni na voljo ene rešitve, ki bi bila ustrezna za vsako od teh specifik. RALSA pa vse navedeno upošteva že avtomatično.

RALSA je odprtokodna rešitev in deluje na vseh platformah, na katere je mogoče namestiti programski jezik R, kar ima številne prednosti: (1) uporabo starejše strojne opreme (»zelena« rešitev); (2) poenotenje izračunov ne glede na strojne in programske vire, ki jih imajo uporabniki po svetu; (3) izračuni so preverljivi, saj ima vsakdo dostop do izvorne kode in lahko preveri algoritme; (4) vsakdo lahko prispeva k projektu s predlogi in celo kodo; (5) gradnja skupnosti za medsebojno podporo. V svetu obstaja nekaj rešitev za analize podatkov mednarodnih primerjalnih raziskav in študij, ki pa so bodisi drage, ker zahtevajo plačljivo programsko opremo, ali pa zelo omejene v smislu podatkovnih baz, ki jih je možno analizirati, ali v smislu možnosti specifičnih analiz. Orodje oz. platforma RALSA se pri tern razlikuje, saj uporabniki ne potrebujejo drage programske opreme. V primerjavi z ostalimi rešitvami, ki prav tako temeljijo na odprtokodni in brezplačni rešitvi (programskem jeziku R), RALSA pokrije več podatkovnih baz, več analiz, predvsem pa je pomembno to, da jo lahko uporabljajo tudi tisti, ki niso vešči programiranja v jeziku R, saj je razvit tudi grafični uporabniški vmesnik (GUI), ki ne zahteva nobenega programiranja, temveč le izbor spremenljivk in tipa analiz.

Za prevod uporabniškega priročnika smo se odločili zaradi tega, ker to programsko orodje uporabljamo za izobraževanja v okviru aplikativnega projekta ARIS Odnos osmošolcev v Sloveniji in Evropi do migrantskih tematik – podatki raziskave IEA ICCS 2009, 2016 in 2022. Sicer pa je RALSA uporabna tudi za analize podatkov drugih mednarodnih primerjalnih raziskav in raziskav.

Dr. Eva Klemenčič Mirazchiyski

1. Navodila za namestitev programa

1.1 Uvod

Sledite tem navodilom za namestitev paketa RALSA skupaj z dodatno programsko opremo, ki jo vaš sistem morda zahteva.

1.2 Sistemske zahteve

Računalnik z vsaj 4 GB RAM-a (priporočeno 16 GB ali več).

Namestite najnovejšo različico R (trenutno 4.4.1). RALSA bo delovala na katerem koli operacijskem sistemu (Linux, MacOS, Windows itd.), kjer je mogoče namestiti R. Ostale programske zahteve so odvisne od platforme; glejte navodila za namestitev za različne operacijske sisteme spodaj.

1.3 Namestitev

Linux

Namestite R in RStudio z uporabo upravitelja paketov v vašem operacijskem sistemu. Zaženite RStudio, vnesite naslednji ukaz in sledite navodilom:

```
install.packages("RALSA", dependencies = TRUE)
```

Pripravljeni ste za uporabo paketa RALSA. Naložite paket z vnosom naslednjega ukaza:

```
library(RALSA)
```

Uživajte!

MacOS

Prenesite R in RStudio, dvokliknite na prenesene *.pkg-datoteke in sledite navodilom za namestitev.
Če vaš sistem še nima XQuartza, ga boste morali tudi prenesti in namestiti. Zaženite RStudio, vnesite naslednji ukaz in sledite navodilom:

```
install.packages("RALSA", dependencies = TRUE)
```

Pripravljeni ste za uporabo paketa RALSA. Naložite paket z vnosom naslednjega ukaza:

```
library(RALSA)
```

Uživajte!

Windows

Uporabniki Windowsov bodo morali vložiti nekaj več truda zaradi sistemskih odvisnosti, ki jih operacijski sistem, v nasprotju z Linuxom in MacOS, ne vključuje.

Prenesite najnovejši različici R-ja in RStudia, dvokliknite na preneseni *.exe-datoteki in sledite navodilom za namestitev.
Prenesite najnovejšo različico paketa Rtools (trenutno RTools4.4) s spletnega mesta CRAN, preberite navodila za namestitev, ki so na voljo na spletni strani za prenos, dvokliknite na preneseno *.exe-datoteko in sledite navodilom za namestitev.
Prenesite Strawberry Perl, dvokliknite na preneseno *.msi-datoteko in sledite navodilom za namestitev.
Zaženite RStudio, vnesite naslednji ukaz in sledite navodilom:

```
install.packages("RALSA", dependencies = TRUE)
```

Pripravljeni ste za uporabo paketa RALSA. Naložite paket z vnosom naslednjega ukaza:

```
library(RALSA)
```

Uživajte!

2. Kako začeti z uporabo RALSA (Getting started with RALSA)

2.1 Uvod

Ta razdelek je namenjen uporabnikom, ki še nimajo izkušenj z R-jem ali pa imajo omejene izkušnje z obsežnimi podatki iz mednarodnih raziskav. Če že imate te izkušnje, lahko ta del preskočite.

Potek dela v paketu RALSA je naslednji:

1. Priprava podatkov.

Pretvorite podatke iz SPSS (ali TXT-podatke v primeru PISA-ciklov pred letom 2015). Za vsak raziskovalni cikel morate to storiti le enkrat. Nato lahko podatke uporabite za nadaljnjo obdelavo in/ali analizo, ne samo v trenutni seji, ampak tudi v drugih sejah.

2. Združite podatke iz različnih držav in respondentov, da jih analizirate skupaj.

3. (Neobvezno) Ustvarite slovarje spremenljivk. To je koristno za pregled lastnosti spremenljivk v pretvorjenem in združenem naboru podatkov.

4. (Neobvezno) Rekodirajte spremenljivke. To je lahko uporabno za združevanje kategorij spremenljivk ali za obrat vrstnega reda spremenljivk. To je še posebej koristno pri izvedbi binarne logistične regresijske analize, saj ta vrsta analize lahko uporablja le dihotomno spremenljivko kot odvisno spremenljivko.

5. Izvedite analizo.
    a. Deleži in povprečja; ali
    b. Percentili; ali
    c. Merila (angl. »Benchmarks); ali
    d. Korelacije; ali
    e. Linearna regresija; ali
    f. Binarna logistična regresija.

Priporočamo uporabo RStudia (brezplačno in odprtokodno razvojno okolje za R) zaradi njegovih zmogljivosti, ki olajšajo delo uporabnikov.

RALSA lahko deluje v dveh načinih – preko ukazne vrstice/skripte in grafičnega uporabniškega vmesnika (Graphic User Interface – GUI). Delo z ukazno vrstico, kjer uporabnik ročno vnaša ukaze, je tradicionalen način dela z R-jem in je, vsaj za izkušene uporabnike, najproduktivnejši način za interakcijo z njim.

R deluje s pomočjo paketov. Ko zaženete R/RStudio, je privzeto naloženih več paketov. Vsi drugi paketi morajo biti naloženi, da lahko uporabite funkcionalnost, ki jo dodajo R-ju. Odprite RStudio in v konzoli ali urejevalniku skriptov vnesite naslednje:

```
install.packages("RALSA", dependencies = TRUE)
```

To bo namestilo paket RALSA iz CRAN-repozitorija ali njegovih ogledal. To morate storiti le enkrat. Paketi v R pogosto prejemajo posodobitve. Da preverite, ali so na voljo nove različice paketov, ki so trenutno nameščeni v vašem sistemu, zaženite naslednji ukaz:

```
old.packages()
```

Če vidite, da je paket RALSA prejel posodobitev, ga lahko posodobite skupaj z vsemi drugimi navedenimi paketi z vnosom naslednjega ukaza:

```
update.packages(ask = FALSE)
```

To bo posodobilo vse pakete v vašem sistemu, ki so prejeli posodobitve, ne da bi za vsak paket posebej zahtevali vaše dovoljenje. Zgornje ukaze redno izvajajte, da bodo vaši nameščeni paketi posodobljeni.

Ko ste namestili RALSA, lahko knjižnico naložite z naslednjim ukazom:

```
library(RALSA)
```

To bo paket dodalo v vašo pot in vam omogočilo dostop do vseh njegovih funkcij ter prikazalo nekaj uvodnih sporočil. Pripravljeni ste za uporabo vseh razpoložljivih funkcij.

2.2 Uporaba ukazne vrstice

Vse funkcije v paketu imajo edinstveno ime. Ko začnete vnašati ukaz iz paketa RALSA, bo RStudio po vnosu nekaj znakov prikazal meni, kjer boste videli seznam ukazov, med katerimi lahko izberete želeno funkcijo. Izberite jo z miško (ali s puščičnimi tipkami na tipkovnici in pritisnite **Enter**). Ukaz, ki ste ga začeli vnašati, bo dokončan in dodani bodo okrogli oklepaji na koncu.

Vse (no, skoraj vse) funkcije v R imajo argumente. Argumenti funkciji povedo, kaj želite, da naredi in/ali kako. Vsak argument sledi enačaju in vrednosti. Razmislite o naslednjem primeru. SPSS-podatke bomo pretvorili v .Rdata-datoteke, ki so značilne za obsežne mednarodne raziskave. To je prvi korak, ki ga je treba narediti, ker raziskave in študije ne zagotavljajo svojih naborov podatkov v obliki, ki je značilna za R, temveč v SPSS in SAS. Po pretvorbi jih lahko uporabite za nadaljnjo obdelavo (združevanje, rekodiranje, pregledovanje slovarjev spremenljivk) in izvedbo analiz. Spodaj je koda, ki pretvori podatke PIRLS 2016 iz SPSS v datoteke .RData:

```
lsa.convert.data(inp.folder = "E:/IDB/PIRLS_2016_IDB/Data/PIRLS",
                 ISO = c("aus", "svn"),
                 out.folder = "C:/temp")
```

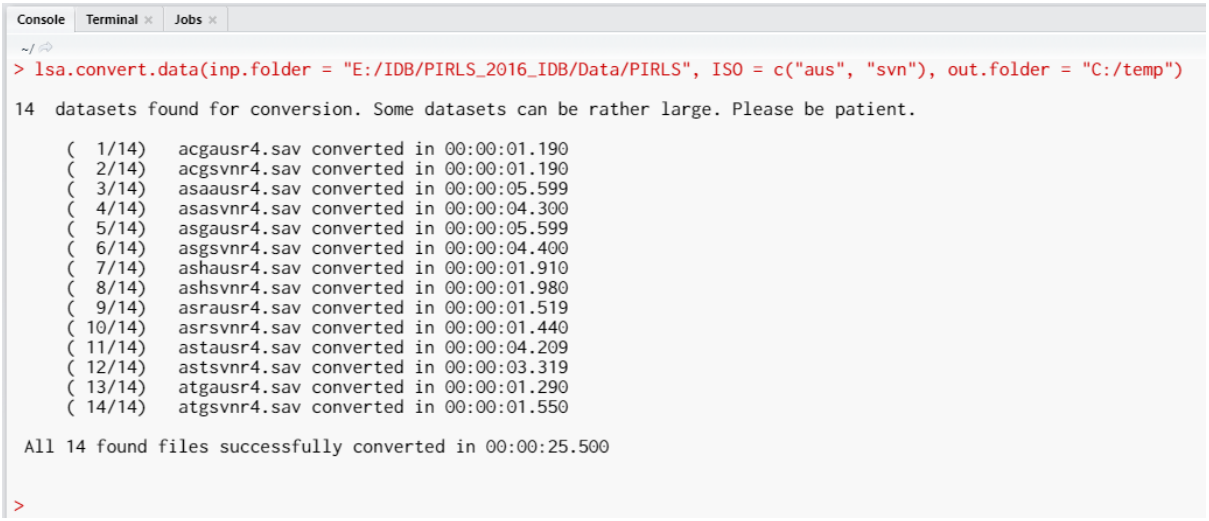
Zgornji del kode kliče funkcijo `lsa.convert.data`. Trije argumenti so ji posredovani:

- **inp.folder** – vnosna mapa, funkciji pove, kje so shranjene SPSS-datoteke, ki jih želite pretvoriti. Funkcija bo uporabila SPSS-datoteke v tej mapi. Prilagodite to pot mapi, kjer so shranjene izvorne SPSS-datoteke.
- **ISO** – trimestne ISO 3166 alfa oznake držav. Ni vam treba poznati ozadja sheme kodiranja ISO 3166. To so edinstvene trimestne kode, ki predstavljajo okrajšave imen držav. To so četrti, peti in šesti znaki v imenu datoteke podatkov. V tem primeru »aus« predstavlja Avstralijo in »svn« Slovenijo (opomba — Slovenija, ne Slovaška!).
- **out.folder** – izhodna mapa, mapa, kamor želite shraniti pretvorjene datoteke. Dobro je, da je ta mapa drugačna od vnosne, sicer bo mapa, kjer so shranjene izvorne SPSS-datoteke, postala zelo nepregledna. Prilagodite to pot mapi, kamor želite shraniti pretvorjene datoteke.

Bodire pozorni na to, kako so poti do datotek posredovane `inp.folder` in `out.folder` – R ne sprejema obratnih poševnic (\) v poteh do datotek, temveč so rezervirani posebni znaki. R uporablja UNIX-ovo konvencijo s poševnicami naprej (/), zato ko prilepite pot, samo zamenjajte vse obratne poševnice s poševnicami naprej. To morate storiti le, ko uporabljate MS Windows, saj Linux in MacOS uporabljata poševnice naprej v poteh do datotek.

Ta funkcija ima tudi druge argumente (glejte referenčni priročnik), ki niso bili posredovani temu klicu. Ti imajo privzete vrednosti, in če vam te privzete vrednosti ustrezajo, jih ni treba posredovati pri klicu funkcije.

Izberite vrstice sintakse in na tipkovnici pritisnite **Ctrl + Enter**. RStudio bo izvedel kodo in pretvoril SPSS-datoteke iz izvorne mape ter shranil pretvorjene datoteke .RData v mapo output. Med izvajanjem bo funkcija natisnila izpis v konzoli RStudia:



Izpis v konzoli prikazuje zaporedno številko datotek iz skupnega števila datotek, čas, potreben za pretvorbo vsake datoteke, in skupen čas, potreben za vse datoteke, ki ste jih izbrali.

Zgornji primer prikazuje običajno delovno rutino za vse funkcije. Seveda se število in vrsta argumentov za vsako funkcijo razlikujeta, vendar je to splošni potek dela.

Posebna opomba je potrebna glede uporabe »verjetnostnih vrednost« (PV, angl. *plausible values*). PV so rezultati dosežkov v preizkušenem predmetu, vendar se, za razliko od običajnih testnih rezultatov, predstavljajo kot več kot en rezultat zaradi zasnove testa. Npr., za PIRLS 2016 bo imel skupni rezultat branja petih PV: »ASRREA01«, »ASRREA02«, »ASRREA03«, »ASRREA04« in »ASRREA05«. V PISA 2018 ima skupna ocena matematike naslednjih deset PV: »PV1MATH«, »PV2MATH«, »PV3MATH«, »PV4MATH«, »PV5MATH«, »PV6MATH«, »PV7MATH«, »PV8MATH«, »PV9MATH« in »PV10MATH«. Pri izračunu statističnih podatkov, povezanih z dosežki, je treba uporabiti vse PV za določen rezultat. PISA 2015 in kasnejše raziskave vključujejo deset PV, medtem ko druge raziskave vključujejo pet PV. Statistični podatki se izračunajo za vsako PV, končne ocene in standardne napake pa se izračunajo z združevanjem vseh posameznih ocen PV z uporabo zapletenih formul. Vmesnik bo vedno prikazoval le osnovni del vsakega nabora PV:

- Za TIMSS (vključno s preTIMSS in eTIMSS), PIRLS (vključno s prePIRLS in ePIRLS), RLII, TIMSS Advanced in TiPi bodo te vrednosti prikazane kot »ASRREA«, »ASRLIT«, »ASMMAT« itd., tj. brez številk na koncu vsake PV v naboru.
- Za PISA, ICCS in ICILS bodo prikazane kot »PV#MATH«, »PV#READ«, »PV#CIV« itd., tj. številke znotraj korena bodo zamenjane z »#«.

Tako bodo te vrednosti izgledale v analitični funkciji (recimo **`lsa.pcts.means`** s podatki učencev in ravnateljev šol PIRLS 2016) za prve navedene primere:

```
lsa.pcts.means(data.file =
"C:/temp/PIRLS_2016_ACG_ASG_merged.RData«,
               split.vars = "ITSEX«, PV.root.avg = "ASRREA«,
               output.file =
"C:/temp/PIRLS_2016_Percentages_and_Means.xlsx«)
```

In tako bo izgledala koda za drugi primer (npr. `lsa.pcts.means` s podatki učencev iz PISA 2018):

```
lsa.pcts.means(data.file = "C:/temp/cy07_msu_stu_qqq.RData«,
               split.vars = "ST004D01T«, PV.root.avg = "PV#MATH«,
               output.file =
"C:/temp/PISA_2018_Percentages_and_Means.xlsx«)
```

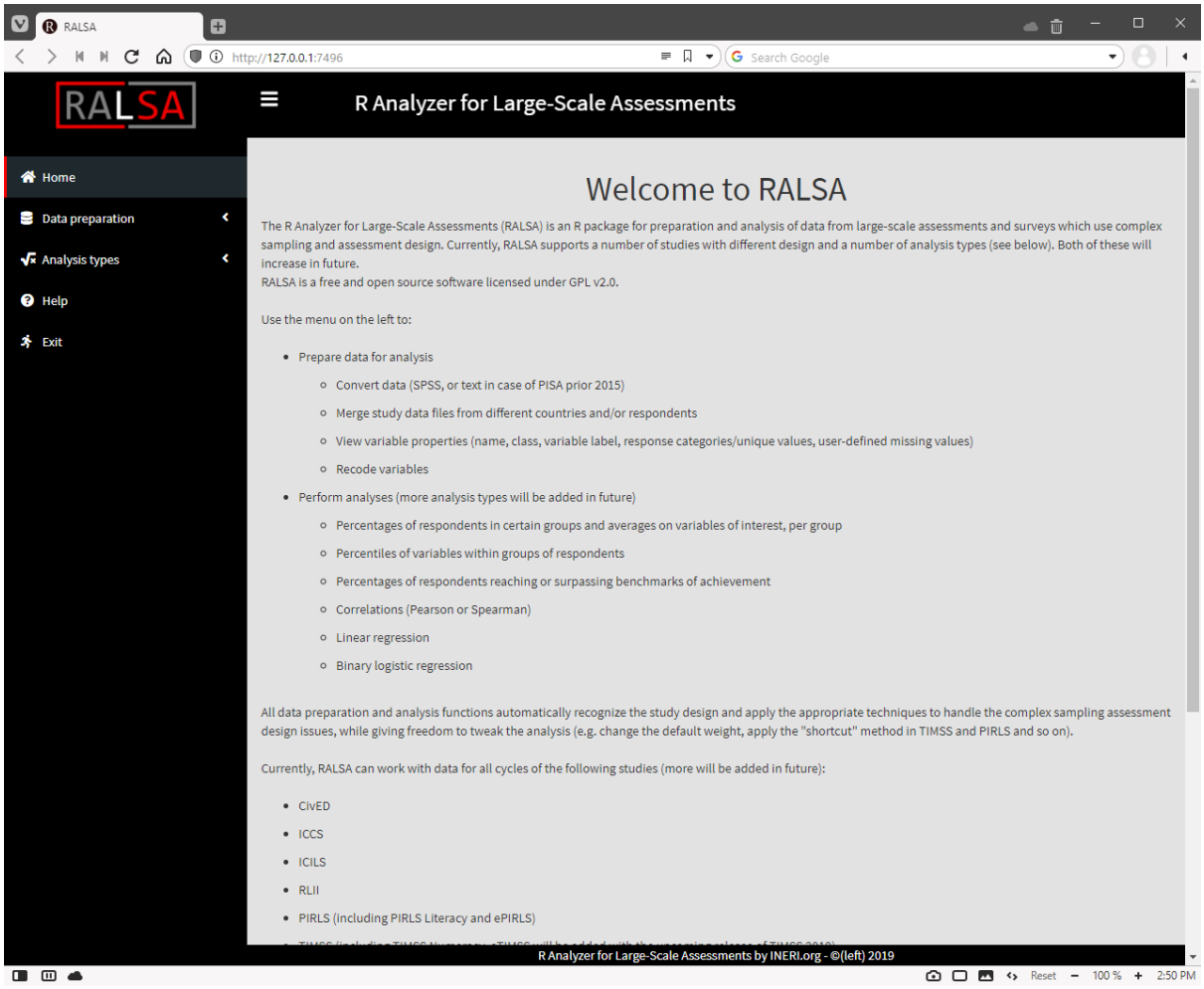
2.3 Uporaba grafičnega vmesnika (GUI)

Kot je bilo že prej omenjeno, R tradicionalno deluje preko ukazne vrstice. Vendar pa R že nekaj let ponuja orodja za ustvarjanje grafičnih uporabniških vmesnikov (GUI). RALSA ponuja tak uporabniški vmesnik. Za hiter uvod v to, kako začeti in uporabljati GUI, si oglejte demo video. Podrobnosti so navedene v nadaljevanju.

Grafični uporabniški vmesnik (GUI) lahko zaženete z naslednjim ukazom v konzoli RStudia:

```
ralsaGUI()
```

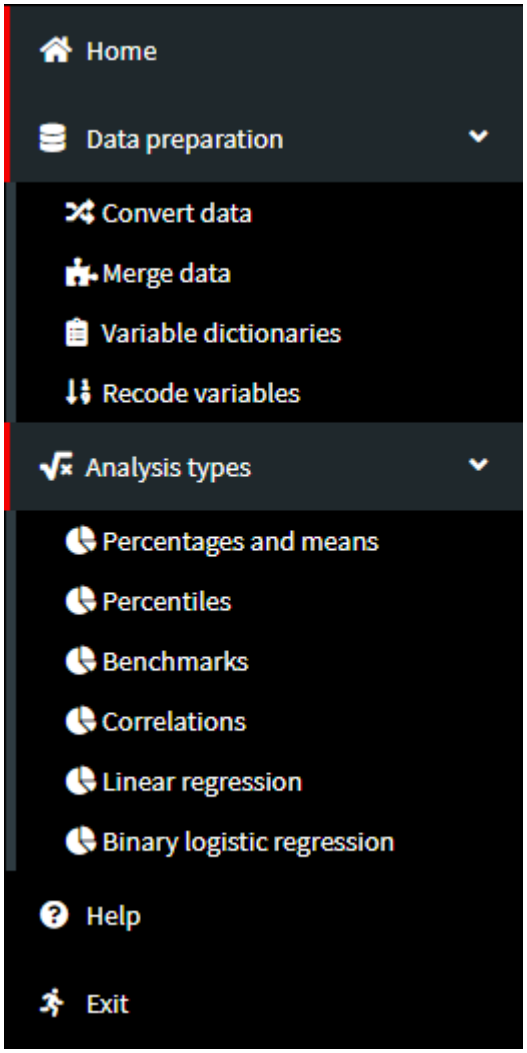
To bo zagnalo GUI v vašem privzetem brskalniku. Začetni zaslon po nalaganju GUI bo videti takole:



Opomba: Z bodočimi posodobitvami se lahko nekateri deli GUI razlikujejo od predstavitve tukaj.

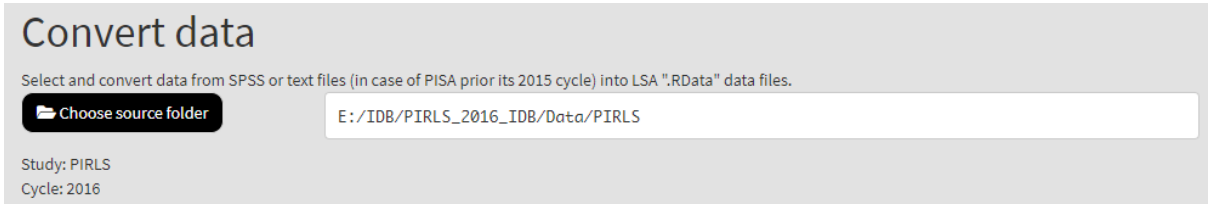
Če morate prilagoditi velikost elementov GUI, držite tipko Ctrl na tipkovnici in se z miško pomikajte gor in dol z miško, dokler ne dobite najboljše slike.

Uporabite meni na levi strani. Ko kliknete na Priprava podatkov (Data preparation) in Vrste analiz (Analysis types), se bodo razširili in vam ponudili različne možnosti:

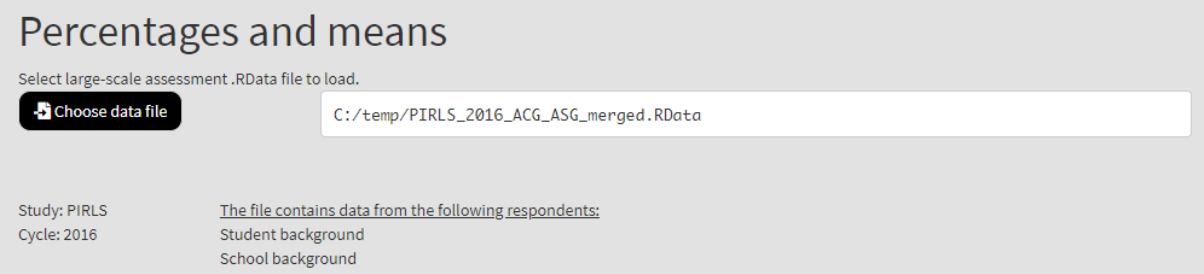


Podpostavke so precej samoumevne. Prva postavka pod Priprava podatkov je Pretvori podatke (Convert data), kar smo že obravnavali v prejšnjem razdelku, ko smo pretvarjali SPSS-podatke v izvorne .Rdata-datoteke. Vse funkcije za pripravo podatkov in vrste analiz delujejo na enak način – izberite mapo ali naložite podatkovno datoteko in sledite navodilom na zaslonu. GUI vas bo na vsakem koraku vodil naprej in vam povedal, kaj storiti. Če pogoj v določenem koraku ni izpolnjen, bo GUI prikazal opozorilo in vam ne bo dovolil nadaljevanja.

Vsi različni razdelki imajo skupne elemente. Prvi skupni element, ki ga boste videli v večini podmenijev pod Priprava podatkov, je gumb Izberi izvorno mapo (Select source folder). Npr. pri pretvarjanju podatkov morate izbrati mapo, ki vsebuje SPSS- (ali TXT- v primeru PISA pred ciklom 2015) datoteke, ki jih želite pretvoriti v izvorne .Rdata-datoteke:

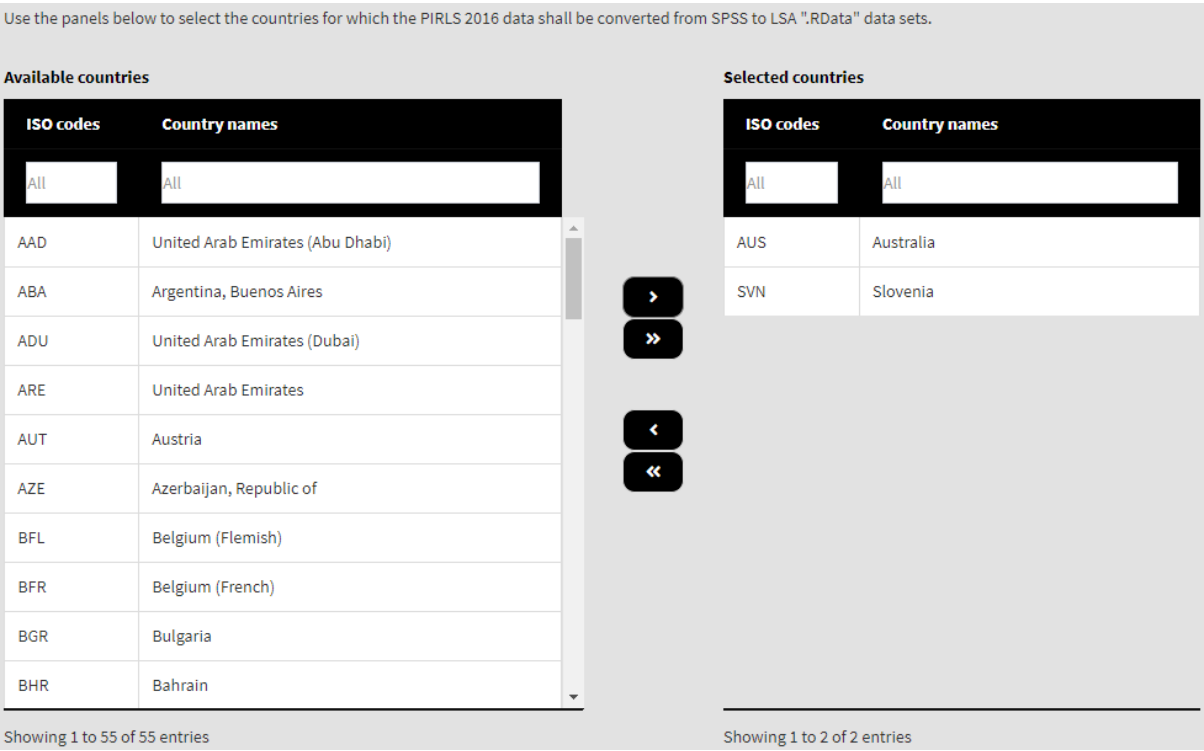


Ali pa: ko izbirate datoteko, ki jo boste uporabili za izračun določene vrste statistike, boste morali izbrati podatkovno datoteko. To je primer za analizo Odstotki in povprečja (Percentages and means), kjer je treba najprej naložiti podatkovno datoteko z uporabo gumba Izberi podatkovno datoteko (Choose data file)



Iz zgornjih primerov lahko opazite, da se raziskava in cikel prepoznata samodejno. V drugem primeru pa vidimo tudi podatke, iz katerih vrst respondentov so podatki združeni v datoteko.

Osrednji skupni elementi GUI so tisti, ki vam omogočajo izbiro postavk, ki jih želite vključiti v izračune. Npr., pri pretvarjanju podatkov, potem ko izberete mapo, ki vsebuje SPSS-datoteke za raziskave, boste morali izbrati države, katerih podatke želite pretvoriti, kot je prikazano na spodnjem posnetku zaslona:



Uporabite desni enojni puščici, da premaknete posamezne države, ki jih izberete, s seznama razpoložljivih držav na levi na seznam izbranih držav na desni strani. In nasprotno: če že imate izbrane države na seznamu izbranih držav, uporabite levo usmerjeno puščico, da jih premaknete nazaj na seznam razpoložljivih držav (tj. jih odznačite). Uporabite levi in desni dvojni puščici, da premaknete vse države med obema seznamoma.

Nekateri drugi skupni elementi v vseh razdelkih GUI so prikazani na spodnji sliki.

Prvi element je potrditveno polje **Use shortcut method for computing SE**. To bo prikazano samo pri različnih vrstah analiz, in sicer le pri analizi podatkov TIMSS (tudi TIMSS Numeracy [tj. preTIMSS] in eTIMSS), PIRLS (tudi PIRLS Literacy [tj. prePIRLS] in ePIRLS) ter pri skupni raziskavi TIMSS in PIRLS iz leta 2011 (tj. TiPi). Pred cikli TIMSS in PIRLS 2011 je izračun napake vzorčenja uporabljal 75 jackknife replikacijskih uteži (tj. eno replicirano utež za vsako jackknife replikacijsko območje). To je znano kot »skrajšana« metoda. Od skupnega cikla 2011 naprej TIMSS in PIRLS uporabljata 150 jackknife replikacij (tj. dve replicirani uteži za vsako jackknife območje), znano kot »polna« metoda. Namen implementirane »skrajšane« metode v RALSA je, da se lahko ponovijo ocene v mednarodnih poročilih TIMSS in PIRLS pred letom 2011. Priporočamo, da vedno uporabljate polno metodo pri izračunih za pridobitev natančnejših ocen standardnih napak. Za pregled vzorčenja, merjenja in skupnih napak ocen si oglejte tehnično dokumentacijo raziskav.

Drugi element je gumb **Define the output file name**. Poleg njega je gumb **Open the output when done**. Ti dve postavki sta skupni za vse vrste analiz. Določitev imena izhodne (output) datoteke je prav tako na voljo pri **Merge data**, **Variable dictionaries** in **Recode data** v razdelku funkcij **Data preparation** v RALSA, medtem ko vas **Convert data** prosi, da določite izhodno mapo za pretvorjene podatke. Ko kliknete na gumb, se bo prikazal pogovorni okvir za shranjevanje datoteke. Pojdite v mapo, kjer želite shraniti izhodno datoteko, in določite ime datoteke za izhod. Potrditveno polje poleg gumba počne to, kar pravi — izhodna/output MS Excel-datoteka se bo odprla, ko bodo vsi izračuni zaključeni.

Pod razdelkom za določitev izhodne datoteke boste videli sintakso za analizo (ali pripravo podatkov). To je ista sintaksa, ki jo običajno vpišete ročno za pripravo podatkov ali izvajanje analize. Ko kliknete na gumb **Execute syntax** spodaj, se bo prikazala konzola, ki bo pokazala napredek vseh operacij.

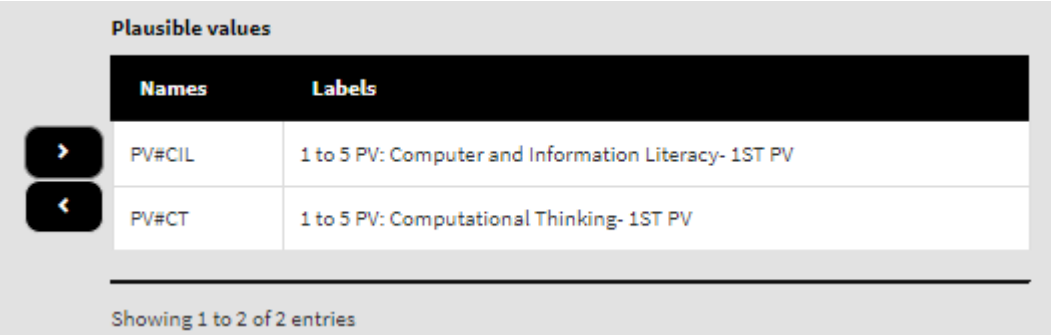
Vsak podrazdelek v razdelkih **Data preparation** ali **Analysis types** bo imel vrsto drugih elementov, ki so z njimi povezani, vendar bi bilo tukaj odveč, da bi jih opisali.

Posebno opombo je treba narediti glede uporabe »plausible values« (PVs oz. verjetnostnih vrednosti). PV-ji dosežki so na testiranem področju, vendar drugačne od običajnih rezultatov testov, saj predstavljajo več kot en rezultat zaradi zasnove testa. Npr., za PIRLS 2016 bo imela skupna ocena branja pet PV: »ASRREA01«, »ASRREA02«, »ASRREA03«, »ASRREA04« in »ASRREA05«. V PISA 2018 bo imel rezultat matematičnega preizkusa naslednjih 10 PV: »PV1MATH«, »PV2MATH«, »PV3MATH«, »PV4MATH«, »PV5MATH«, »PV6MATH«, »PV7MATH«, »PV8MATH«, »PV9MATH« in »PV10MATH«. Pri izračunih statistike, ki vključujejo dosežke, je treba uporabiti vse PV za določen rezultat. PISA 2015 in kasnejše raziskave PISA vključujejo 10 PV, vse druge raziskave pet. Statistike se izračunajo za vsako PV, končne ocene in njihove standardne napake pa se izračunajo z združevanjem vseh posameznih ocen PV z uporabo kompleksnih formul. Vmesnik bo vedno prikazal le koren imena vsakega nabora PV:

- Za TIMSS (tudi preTIMSS in eTIMSS), PIRLS (tudi prePIRLS in ePIRLS), RLII, TIMSS Advanced in TiPi bodo prikazani kot »ASRREA«, »ASRLIT«, »ASMMAT« itd., torej brez številk na koncu vsake PV v naboru PV.
- Za PISA, ICCS in ICILS bodo prikazani kot »PV#MATH«, »PV#READ«, »PV#CIV« itd., torej z zamenjavo številk znotraj korena z »#«.

Tako bodo ti podatki izgledali v vmesniku za prve navedene:

In tako bodo podatki izgledali za druge navedene:



Poskusite se poigrati s programom. Uživajte!

3. Priprava podatkov na analize

3.1 Pretvorba podatkov (Convert data)

3.1.1 Uvod

Podatki iz mednarodnih raziskav so na voljo v formatih datotek SPSS, SAS in TXT (v primeru raziskave PISA pred ciklom 2015). Različni R-paketi, ki uvažajo datoteke SPSS in SAS, kot so `foreign` in `haven`, npr., ter funkcije v privzeto naloženih R-paketih, kot sta `read.delim`, `read.delim2` iz paketa `utils`, ali funkcije v drugih paketih, kot sta `readr` ali `data.table`, lahko berejo TXT ločene datoteke. Vendar imajo vse te funkcije različne implementacije metod za uvoz podatkov in vrste objektov, ki nastanejo pri uvozu, kar vpliva na nadaljnje izračune (priprava podatkov ali analiza). Tako je težko napovedati, kako bo uporabnik uvozil podatke in kakšen bo rezultat. Poleg tega je veliko priročneje imeti podatke v nativnem .Rdata-formatu s standardizirano strukturo. Zato je treba zagotoviti t-funkcionalnost za pretvorbo podatkov v .Rdata-format.

Funkcija `lsa.convert.data` pretvori podatke iz formatov datotek SPSS iz raziskav IEA in PISA (2015 in kasnejši cikli) ter TXT (v primeru PISA pred ciklom 2015) v .Rdata-datoteke. Pri tem dodaja nekatere lastnosti objektom, shranjenim v teh datotekah, tako da druge funkcije (priprava podatkov ali analiza) v RALSA vedo, kaj narediti z njimi, ko so posredovani.

RALSA dodaja novo metodo funkciji `print` iz osnovnega paketa za tiskanje lastnosti `lsa.data` v konzoli (raziskava, cikel, vrsta respondenta, število držav, ključ – ID države, če imajo spremenljivke uporabniško določene manjkajoče vrednosti) ter predogled podatkov. Metoda tiskanja za `lsa.data` je na voljo samo v načinu ukazne vrstice.

Funkcija `lsa.select.countries.PISA` uporabnikom omogoča, da izberejo države za analizo iz pretvorjene PISA-datoteke in jo shranijo kot novo datoteko ali objekt v pomnilniku. PISA-podatkovne datoteke vsebujejo vse države po respondentih. To je lahko precej nevšečno za uporabnika, ki ne potrebuje analizirati podatkov vseh držav v datoteki.

Nekaj pomembnih opomb:

- Ne spreminjajte nobene izmed izvornih SPSS- ali TXT- (.sps-) datotek pred njihovo pretvorbo. Prenesite datoteke, razpakirajte arhive in ne spreminjajte njihovih imen ali vsebine.
- Datotek ni treba ponovno pretvoriti za nadaljnje (priprava podatkov ali analiza) operacije. Pretvorite jih enkrat, nato pa uporabite pretvorjene datoteke.
- V nadaljevanju boste videli frazo »IEA-like« ali »podobno kot IEA«. To pomeni, da so datoteke na voljo po ciklu, državi in vrsti respondenta, v nasprotju s primerom, kjer so datoteke na voljo za vse države skupaj po raziskavi in ciklu (kot to počne PISA). Študiji, ki nista izvajani s strani IEA, vendar uporabljata enako strukturo datotek, sta npr. TALIS in TALIS 3S.

3.1.2 Funkcija za pretvorbo podatkov, metoda tiskanja `lsa.data` in funkcija za izbiro držav PISA ter njihovi argumenti

Funkcija `lsa.convert.data` ima naslednje argumente:

- `inp.folder` – mapa, ki vsebuje podatkovne datoteke SPSS, podobne IEA, ali besedilne ASCII-datoteke ter .sps-datoteke za uvoz podatkov OECD PISA iz ciklov pred letom 2015. Če je prazna, se uporabi delovni imenik (`getwd()`).
- `PISApre15` – Pri pretvorbi PISA-datotek nastavite na TRUE, če so vhodne datoteke iz ciklov PISA pred letom 2015 (besedilni format ASCII z .sps-kontrolnimi datotekami), ali na `FALSE` (privzeto), če so v formatu SPSS .sav, kot v primeru raziskav IEA ipd. ter OECD PISA 2015 ali kasneje. Prezrto, če vhodna mapa vsebuje raziskave, podobne IEA.
- `ISO` – vektor, ki vsebuje znake ISO-kode državnih podatkovnih datotek, ki jih je treba pretvoriti (npr. `ISO = c("aus", "svn")`). Če nobena izmed datotek ne vsebuje navedenih ISO-kod v svojih imenih, se kode prezrejo in prikaže se opozorilo. Prezrto pri pretvorbi PISA-datotek (tako za cikle pred 2015 kot za 2015 in kasneje). Ta argument ni občutljiv na velike in male črke, tj. ISO-kode lahko posredujete v malih ali velikih črkah.
- `missing.to.NA` – Ali naj bodo uporabniško določene manjkajoče vrednosti pretvorjene v `NA`? Če je `TRUE`, bodo vse uporabniško določene manjkajoče vrednosti iz SPSS-datotek (ali določene v uvoznih sintaksah OECD PISA) uvožene kot `NA`. Če je `FALSE` (privzeto), bodo pretvorjene v veljavne vrednosti, manjkajoče kode pa bodo dodeljene atributu »missings« za vsako spremenljivko.
- `out.folder` – pot do mape, kjer bodo shranjene pretvorjene datoteke. Če je prazna, je enaka kot `inp.folder`, in če je tudi `inp.folder` manjkajoča, bo to `getwd()`.

V zvezi s funkcijo `lsa.convert.data` je metoda tiskanja za objekte `lsa.data` naložena v pomnilnik. Ko je ukaz (ali preprosto ime objekta) klican, ukaz `print` v R preveri, ali je objekt razreda `lsa.objects`, in če je, natisne strukturiran izpis, ki vsebuje ime raziskave, cikel, vrsto respondentov, skupno število držav, ključ, ali podatki vsebujejo uporabniško določene manjkajoče vrednosti ali ne, in del podatkov. Slednji argumenti so lahko posredovani funkciji `print`:

- `x` – objekt `lsa.data`.
- `col.nums` – katere stolpce natisniti, pozicije po številu.

Funkcija `lsa.select.countries.PISA` ima naslednje argumente:

- `data.file` – pretvorjena PISA-podatkovna datoteka za izbiro podatkov držav. Bodisi ta bodisi `data.object` mora biti izbran, vendar ne oba. Glejte podrobnosti.
- `data.object` – PISA-objekt v pomnilniku za filtriranje. Bodisi ta bodisi `data.file` mora biti izbran, vendar ne oba. Glejte podrobnosti.
- `cnt.names` – vektor znakov, ki vsebuje imena držav, kakor se pojavljajo v podatkih, ki naj ostanejo v izvoženi PISA-datoteki ali objektu v pomnilniku.
- `output.file` – polna pot do datoteke s filtriranimi podatki držav, ki naj bo zapisana na disk. Če ni navedena, bo PISA-objekt zapisan v pomnilnik.

Opombe:

1. Raziskave IEA pa tudi OECD TALIS in nekatere druge organizacije nudijo svoje podatke v formatu SPSS .sav s podobno strukturo: ena datoteka na državo in vrsto respondenta (npr. ravnatelj, učenec, učitelj itd.) za vsako populacijo. Za raziskave IEA in OECD TALIS uporabite argument `ISO` za določitev tridelnih ISO-kod državam, katerih podatki naj bodo pretvorjeni. Tridelne ISO-kode za posamezno državo lahko najdete v priročniku za raziskavo. Npr., ISO-kode držav, ki so sodelovale v PIRLS 2016, so navedene v priročniku raziskave na straneh 52–54. Za pretvorbo datotek vseh držav v prenesenih podatkih iz raziskav IEA in OECD TALIS preprosto izpustite argument `ISO`. Cikli OECD PISA pred letom 2015 pa ne zagotavljajo SPSS .sav ali drugih binarnih datotek, temveč ASCII-besedilne datoteke, ki jih spremljajo SPSS-sintaktične (.sps) datoteke, ki se uporabljajo za uvoz besedilnih datotek v SPSS. Te datoteke so za vsak tip respondenta in vsebujejo podatke vseh držav. Funkcija `lsa.convert.data` pretvori podatke iz katerega koli vira, pri čemer zagotavlja, da je struktura izhodnih .Rdata-datotek enaka, čeprav je struktura vhodnih datotek različna (SPSS-binarne datoteke proti ASCII-besedilne datoteke in uvoz .sps-datoteke). Podatki iz PISA 2015 in kasneje so na voljo v SPSS-formatu (vse države v eni datoteki na vrsto respondenta). Tako je treba argument `PISApre15` nastaviti na `TRUE` pri pretvorbi podatkov iz PISA pred letom 2015. Privzeti nastavek za argument `PISApre15` je `FALSE`, kar pomeni, da funkcija pričakuje, da bodo v imeniku v `inp.folder` najdene IEA podobne SPSS-binarne datoteke na državo in vrsto respondenta ali OECD PISA 2015 (ali kasneje) SPSS .sav-datoteke. Če je `PISApre15 = TRUE` in so kodne države določene v `ISO`, bodo ignorirane, ker PISA-datoteke podatke vseh držav vsebujejo skupaj.

2. Datoteke, ki jih je treba pretvoriti, morajo biti v lastnem imeniku, iz ene raziskave, enega cikla in ene populacije. V primeru OECD PISA pred letom 2015 mora imenik vsebovati tako ASCII-besedilne datoteke kot SPSS .sps-sintaktične datoteke. Če imenik vsebuje podatkovne nize iz več raziskav ali ciklov, bo operacija prekinjena z napakami.

3. Če pot za argument `inp.folder` ni določena, bo funkcija iskala datoteke v delovnem imeniku (tj. kot ga vrne `getwd()`). Če pot za argument `out.folder` ni določena, bo uporabljen isti imenik kot `inp.folder`, datoteke pa bodo shranjene tja. Če sta argumenta `inp.folder` in `out.folder` oba manjkajoča, bo uporabljen imenik iz `getwd()` za iskanje, pretvorbo in shranjevanje datotek.

4. Če je `missing.to.NA` nastavljeno na `TRUE`, bodo vse uporabniško določene manjkajoče vrednosti iz SPSS uvožene kot `NA`, kar je edina vrsta manjkajoče vrednosti v R. To bo najpogostejši primer pri analizi teh podatkov, saj razlog za manjkajoči odgovor običajno ni pomemben. Vendar, če je treba vedeti razloge za manjkajoče odgovore, kot pri analizi dosežkov (tj. neobdelani proti izpuščeni ali nedoseženi), je treba argument nastaviti na `FALSE` (privzeti za ta argument), kar bo vse uporabniško določene manjkajoče vrednosti pretvorilo v veljavne.

5. Ko prenašate .sps-datoteke (ASCII-besedilne in kontrolne .sps) za OECD PISA-datoteke pred letom 2015 (npr. https://www.oecd.org/pisa/pisaproducts/pisa2009database-downloadabledata.htm), jih shranite brez spreminjanja njihovih imen in brez spreminjanja vsebine datotek. Funkcija bo iskala datoteke s prvotnimi imeni.

6. Različne raziskave in cikli definirajo kategorijo »ne vem« (ali podobno) diskretnih spremenljivk na različne načine – bodisi kot veljavne bodisi kot manjkajoče vrednosti. Funkcija `lsa.convert.data` nastavi vse takšne ali podobne kode na manjkajoče vrednosti. Če je treba to spremeniti, se lahko uporabi tudi `lsa.recode.vars` (prav tako glejte `lsa.vars.dict`).

3.1.3 Pretvorba SPSS-podatkov (angl. *converting SPSS data*)

Funkcija `lsa.convert.data` pretvori podatke SPSS v izvorni format .RData iz naslednjih raziskav:

- CivED;
- ICCS;
- ICILS;
- RLII;
- PIRLS (vključno s PIRLS Literacy in ePIRLS);
- TIMSS (vključno s TIMSS Numeracy, eTIMSS);
- TiPi (skupna raziskava TIMSS in PIRLS);
- TIMSS Advanced;
- SITES;
- TEDS-M;
- PISA (cikli od 2015 dalje);
- TALIS;
- TALIS Starting Strong Survey (imenovana tudi TALIS 3S);
- REDS.

Med raziskavo PISA (od 2015 dalje) in ostalimi raziskavami na zgornjem seznamu (v slogu IEA) obstaja razlika. Vse raziskave razen PISA zagotavljajo datoteke glede na cikel, državo in tip anketiranca (npr. ravnatelj šole, učitelj, dijak ali starš). Npr., to so datoteke SPSS za Slovenijo iz raziskave PIRLS 2016 (za ogled načina poimenovanja datotek in spremenljivk ter opisa podatkovne baze se obrnite na 4. poglavje uporabniškega vodnika PIRLS 2016):

- `acgsvnr4.sav` – odgovori ravnateljev šol na vprašalnik;
- `asasvnr4.sav` – odgovori dijakov na dosežke;
- `asgsvnr4.sav` – odgovori dijakov na vprašalnik, vključuje tudi PV;
- `ashsvnr4.sav` – odgovori staršev na vprašalnik;
- `asrsvnr4.sav` – zanesljivost postavk dosežkov;
- `astsvnr4.sav` – povezovalne datoteke dijak-učitelj;
- `atgsvnr4.sav` – odgovori učiteljev na vprašalnik.

Ta razdelitev na ločene datoteke omogoča zelo priročno delo z datotekami in združevanje prilagojenih naborov podatkov za analizo, tj. združevanje samo potrebnih držav in tipov anketirancev.

Podatkovne datoteke SPSS iz raziskave PISA (od cikla 2015 naprej) so organizirane drugače. OECD zagotavlja podatke po ciklu in tipu anketiranca. Vsaka datoteka po tipu anketiranca vsebuje podatke iz vseh držav, ki sodelujejo v ciklu. Npr., to so datoteke iz cikla PISA 2018:

- `CY07_MSU_SCH_QQQ.sav` – odgovori ravnateljev šol na vprašalnik;
- `CY07_MSU_STU_COG.sav` – odgovori dijakov na postavke dosežkov;
- `CY07_MSU_STU_QQQ.sav` – odgovori dijakov na vprašalnik, vključuje tudi PV;
- `CY07_MSU_STU_TIM.sav` – časovni dnevnik postavk dosežkov dijakov;
- `CY07_MSU_TCH_QQQ.sav` – odgovori učiteljev na vprašalnik;
- `CY07_VNM_STU_COG.sav` – odgovori dijakov na postavke dosežkov, samo za Vietnam;
- `CY07_VNM_STU_PVS.sav` – PV samo za Vietnam.

Pomembno je omeniti, da se vrste datotek in njihovo poimenovanje razlikujejo od enega cikla PISA do drugega. Vendar pa funkcija `lsa.convert.data` te razlike samodejno obravnava in samodejno prepozna ime raziskave ter lastnosti podatkovnih datotek za uspešno pretvorbo.

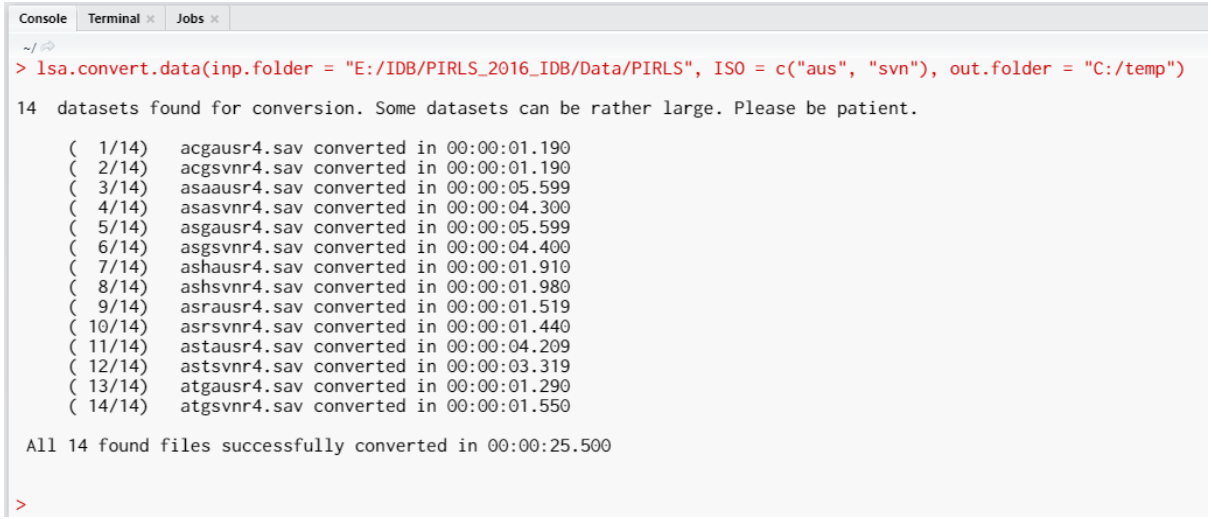
3.1.4 Pretvorba datotek SPSS z ukazno vrstico

Kot primer bomo pretvorili podatke iz raziskave PIRLS 2016. V RStudiu izvedite naslednjo sintakso:

```
lsa.convert.data(inp.folder = "E:/IDB/PIRLS_2016_IDB/Data/PIRLS",
                 ISO = c("aus", "svn"),
                 out.folder = "C:/temp")
```

Ta ukaz bo pretvoril podatke iz PIRLS 2016 za Avstralijo in Slovenijo. Funkcija bo vzela vse razpoložljive datoteke za ti dve državi za ta cikel, ki so v `inp.folder`, in jih shranila v `out.folder`. Če `inp.folder` vsebuje podatke iz več kot ene raziskave in/ali cikla ali več kot ene populacije, se bo funkcija ustavila z napako. Če je `out.folder` enak kot `inp.folder`, bo funkcija vseeno pretvorila podatke, vendar bo na koncu prikazala opozorilo. Ti dve funkcionalnosti sta namenoma vključeni – za ohranjanje urejenosti map in strukture datotek vedno shranjujte podatke SPSS iz ene raziskave/cikla/populacije v eno mapo ter za pretvorjene datoteke .RData vedno izberite drugo mapo kot za izvorne datoteke SPSS.

RStudio bo v konzoli prikazal sporočila o poteku operacije.



Kaj pa, če želimo pretvoriti datoteke iz vseh držav, sodelujočih v PIRLS 2016, v mapi `inp.folder`? Preprosto lahko izpustimo argument `ISO` (torej ne podamo nobenih ISO-kod držav) v klicu funkcije. S tem bomo funkciji `lsa.convert.data` sporočili, naj vzame datoteke SPSS iz vseh držav, shranjene v mapi `inp.folder`, in jih pretvori:

```
lsa.convert.data(inp.folder = "E:/IDB/PIRLS_2016_IDB/Data/PIRLS",
                 out.folder = "C:/temp")
```

Upoštevajte, da bo čas izvedbe odvisen od števila datotek na vrsto respondenta v zbirki podatkov, velikosti posameznih vrst datotek in števila držav. Vendar pa je s podatkovnimi bazami v obliki IEA postopek pretvorbe dokaj hiter. Npr., pretvorba celotne zbirke podatkov PIRLS 2016 (glavni PIRLS, ne ePIRLS ali prePIRLS), ki vsebuje podatke iz 57 držav (sedem datotek na državo), traja približno 11 minut. Če vam ni treba pretvoriti podatkov iz vseh držav, ampak le tiste, ki jih boste potrebovali v svojih analizah, uporabite argument ISO, kot je prikazano zgoraj.

Kaj pa datoteke SPSS iz PISA 2015 in kasnejših ciklov? Postopek je enak. Vendar pa so, kot je omenjeno na začetku, datoteke SPSS za cikel PISA na voljo za vse države v okviru ene vrste datotek. Posledično ne boste mogli pretvoriti podatkov iz teh zbirk podatkov za posamezne države, ampak za celotne datoteke. Če poskusite prenesti argument `ISO` v klic, bo ta prezrt. Sintaksa bo videti tako:

```
lsa.convert.data(inp.folder = "E:/IDB/PISA_2018_IDB/Data",
                 out.folder = "C:/temp")
```

V konzoli bodo natisnjena sporočila, podobna prejšnji pretvorbi zbirke podatkov PIRLS 2016. Upoštevajte, da lahko zaradi kompleksne strukture posameznih datotek, ki vsebujejo podatke iz vseh držav in veliko število spremenljivk, pretvorba datotek SPSS iz PISA traja resnično dolgo časa. Celotna zbirka podatkov PISA 2015 npr. traja približno 5,5 ure na računalniku s procesorjem Intel Core i7-7500U procesorjem in z 16 GB RAM-a. To je lahko precej frustrirajoče. Lahko pustite računalnik, da dela čez noč, in zjutraj bodo datoteke podatkov pretvorjene.

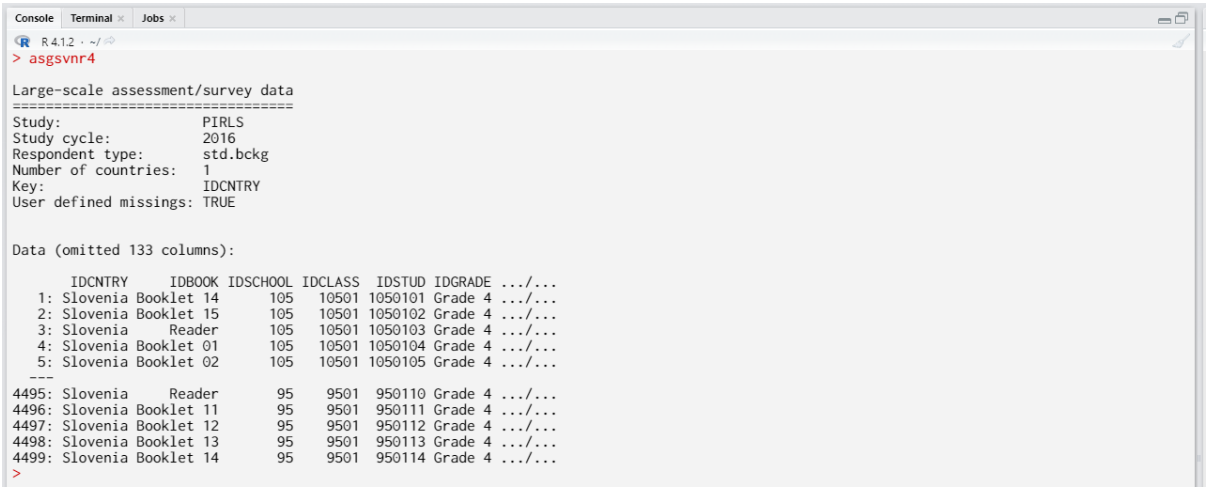
Datoteka se lahko naloži v pomnilnik in natisne v konzolo tako, da preprosto podate ime naloženega objekta. Če želite npr. naložiti pretvorjeno datoteko s podatki, recimo datoteko s podatki učencev za Slovenijo:

```
load("C:/temp/asgsvnr4.RData")
```

Funkcija `load` bo prebrala datoteko in naložila objekt `asgsvnr4` iz nje v pomnilnik. Objekt lahko natisnemo tako, da preprosto podamo njegovo ime:

```
asgsvnr4
```

Funkcija `print` iz osnovnega paketa bo našla ustrezno metodo za `lsa.data` in na zaslonu prikazala naslednji izpis:



Upoštevajte, da se privzeto prikaže prvih šest stolpcev. Če potrebujemo, lahko natisnemo različne stolpce glede na naše zanimanje, tako da podamo številke stolpcev:

```
print(x = asgsvnr4, col.nums = c(20:25))
```

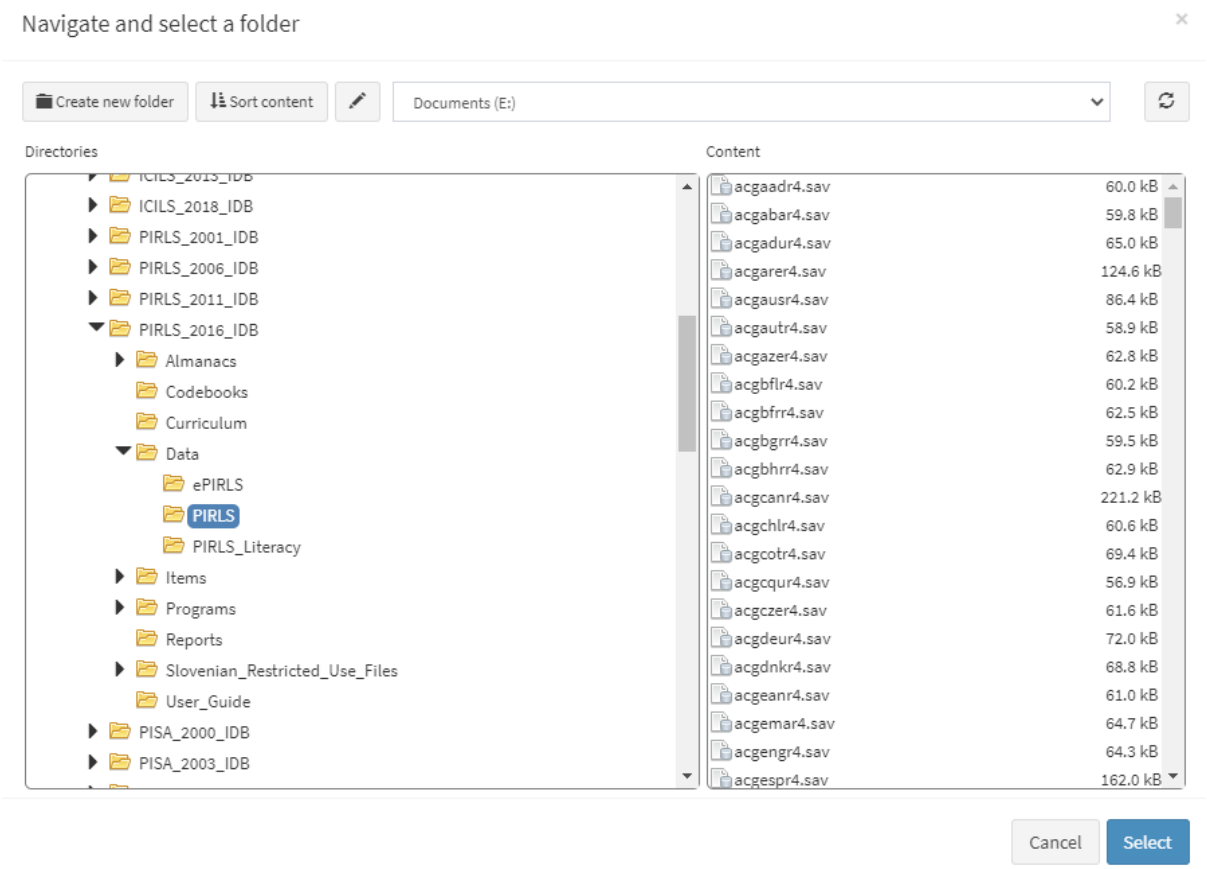
Ustrezen izpis bo videti takole:

3.1.5 Pretvorba SPSS-datotek s pomočjo GUI (angl. *Converting SPSS files using the GUI*)

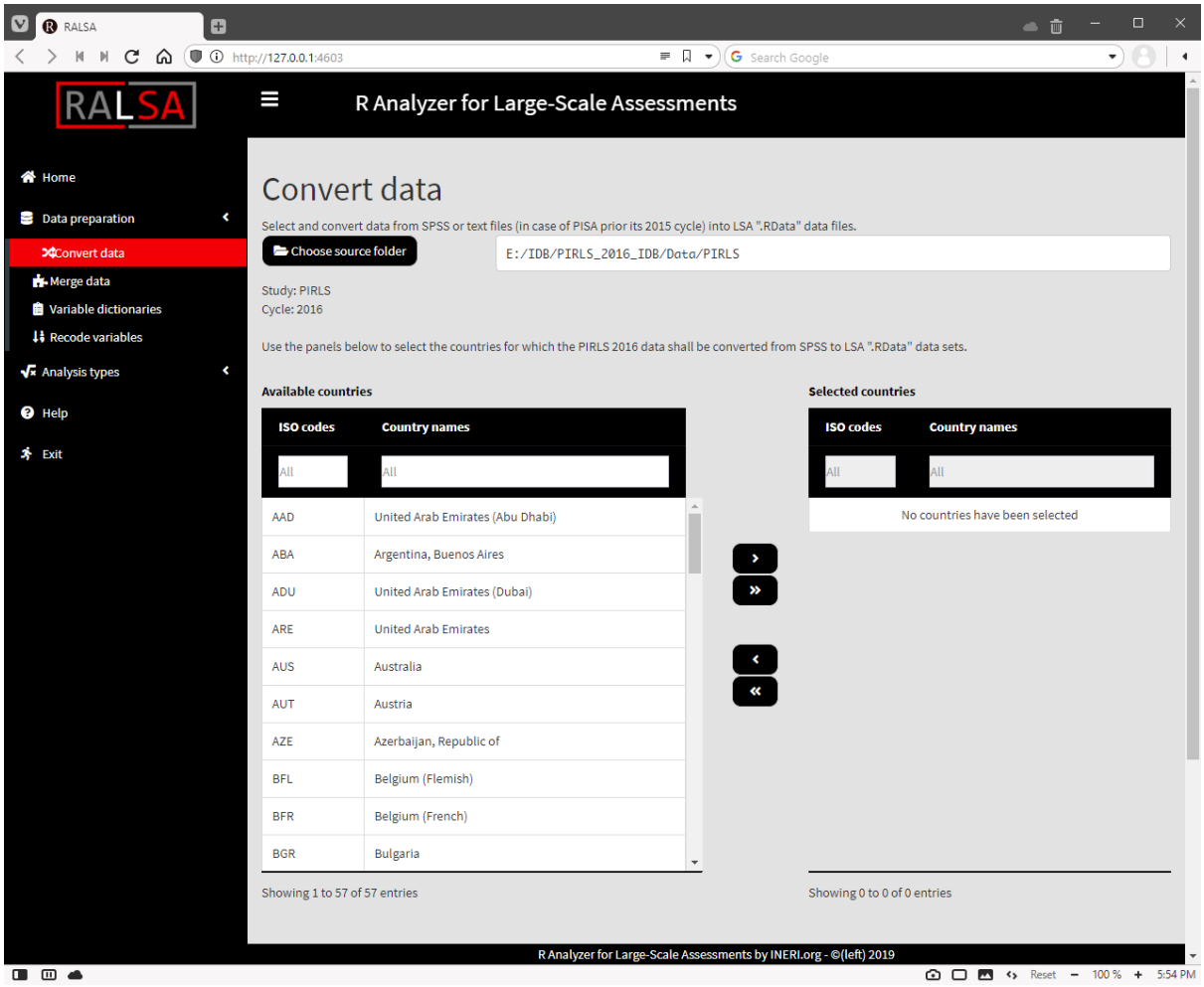
Za zagon uporabniškega vmesnika RALSA v RStudio izvedite naslednji ukaz:

```
ralsaGUI()
```

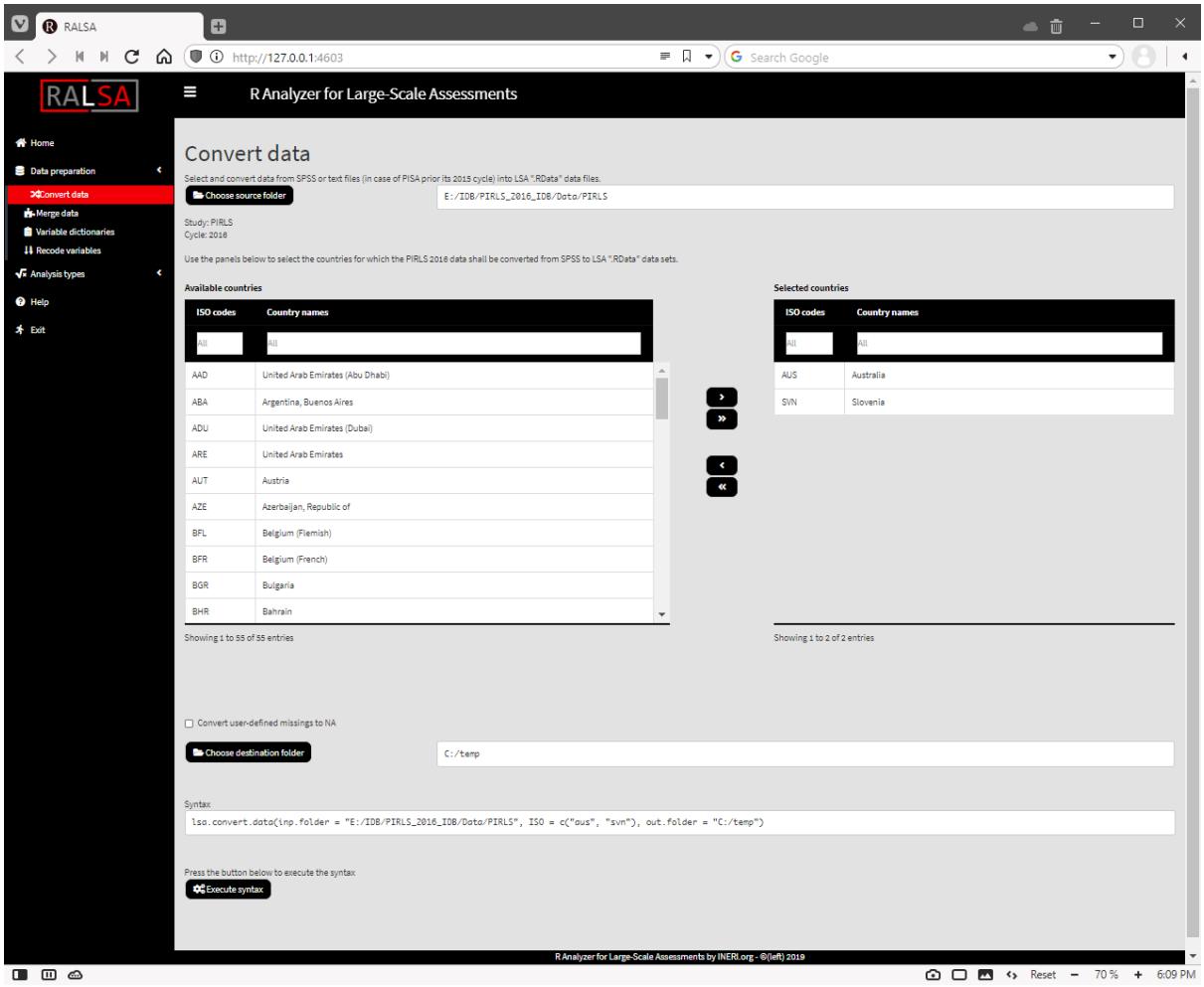
Ko se GUI odpre v vašem brskalniku, izberite **Data preparation** > **Convert data** iz menija na levi. Ko ste navigirani v odsek **Convert data** v GUI, kliknite na **Choose source folder**. Nato se pomaknite do mape, ki vsebuje SPSS-datoteke za raziskavo (npr. PIRLS 2016). V levem podoknu dialognega okna za izbiro mape izberite mapo, ki vsebuje podatke. Razpoložljive SPSS-datoteke v mapi bodo prikazane v desnem podoknu:



Ko potrdite izbiro s pritiskom na **Select**, boste videli naslednji zaslon:



Uporabite miško za izbiro in enojne ter dvojne puščične gumbe za premikanje držav med seznamoma **Available countries** (Razpoložljive države) in **Selected countries** (Izbrane države). Enojne puščične gumbe lahko uporabite za izbiro in premik posameznih držav med dvema paneloma. Dvojne puščične gumbe lahko uporabite za premik vseh držav med paneloma, tudi če nobena ni izbrana. Izberimo Avstralijo in Slovenijo. Pod seznamom se bosta pojavila potrditveno polje in gumb. Če želite, da se vse uporabniško določene manjkajoče vrednosti v SPSS pretvorijo v edino manjkajočo vrednost, ki jo podpira R, označite možnost **Convert user-defined missings to NA**. Če tega ne želite, pustite polje prazno (privzeto), v tem primeru bodo uporabniško določene manjkajoče vrednosti v SPSS dodeljene kot dodatni atribut spremenljivkam, ki jih lahko uporabite kasneje, če bo potrebno. Če pa menite, da jih ne boste potrebovali, preprosto označite polje. Kliknite na gumb **Choose destination folder** (Izberite ciljno mapo) in z uporabo pogovornega okna za izbiro mape navigirajte do mape, v katero želite shraniti pretvorjene datoteke. Ko potrdite, se na dnu zaslona prikažejo pot do ciljne mape, polje za ukazno sintakso in polje za izvajanje sintakse.



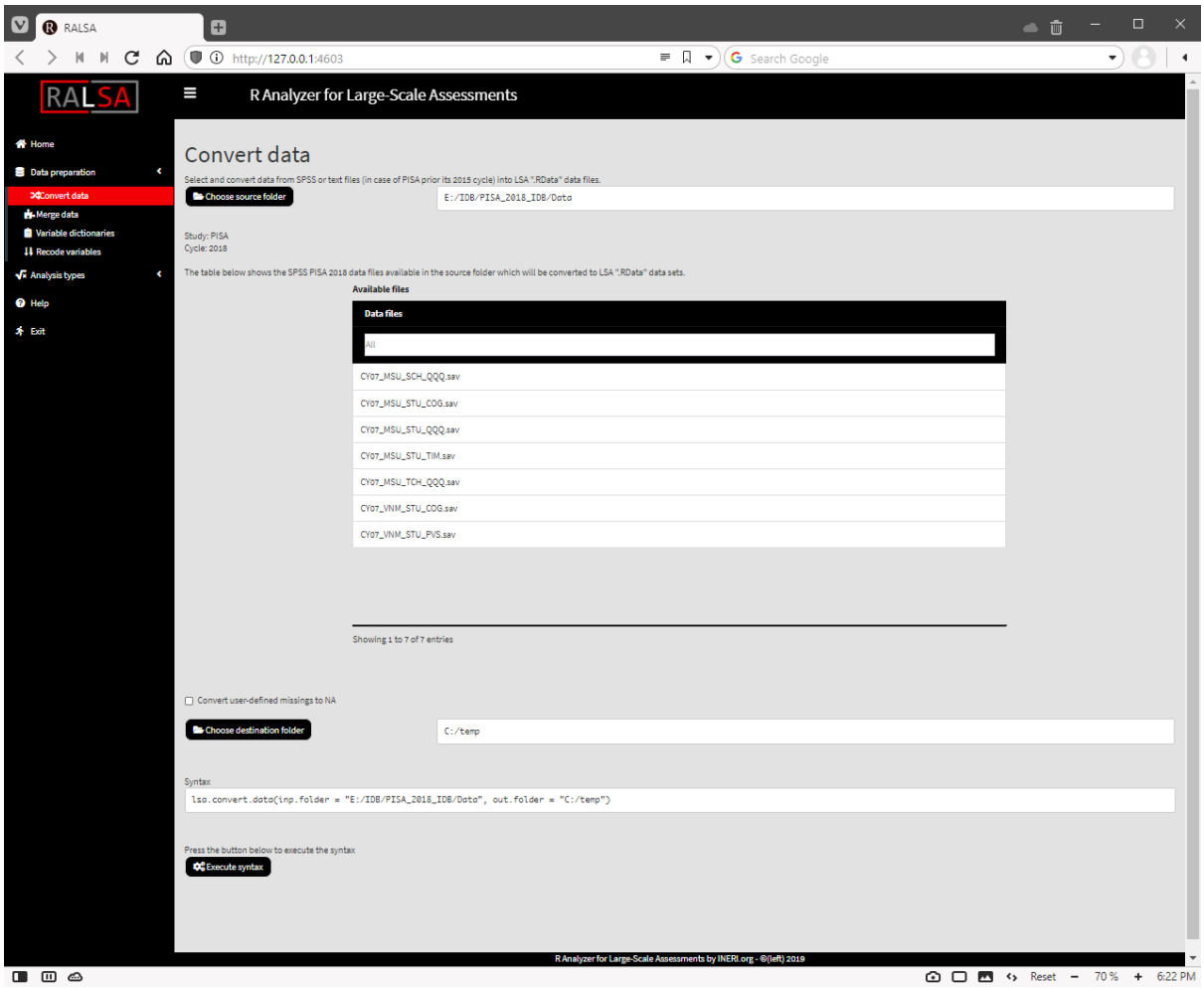
Kliknite na gumb **Execute syntax** (Izvedi sintakso). Prikazalo se bo pojavno sporočilo, ki vas obvesti, da se je pretvorba datotek začela. Na dnu zaslona se bo pojavila konzola, ki se nenehno posodablja in prikazuje potekajoče operacije:

```
14  datasets found for conversion. Some datasets can be rather large. Please be patient.

    (  1/14)   acgausr4.sav converted in 00:00:01.239
    (  2/14)   acgsvnr4.sav converted in 00:00:01.169
    (  3/14)   asaausr4.sav converted in 00:00:05.370
    (  4/14)   asasvnr4.sav converted in 00:00:04.050
    (  5/14)   asgausr4.sav converted in 00:00:05.669
    (  6/14)   asgsvnr4.sav converted in 00:00:04.519
    (  7/14)   ashausr4.sav converted in 00:00:01.900
    (  8/14)   ashsvnr4.sav converted in 00:00:02.029
    (  9/14)   asrausr4.sav converted in 00:00:01.470
    ( 10/14)   asrsvnr4.sav converted in 00:00:01.709
    ( 11/14)   astausr4.sav converted in 00:00:04.339
    ( 12/14)   astsvnr4.sav converted in 00:00:03.669
    ( 13/14)   atgausr4.sav converted in 00:00:01.309
    ( 14/14)   atgsvnr4.sav converted in 00:00:01.250

 All 14 found files successfully converted in 00:00:25.720
```

Pomaknite se navzdol, če konzole ne vidite takoj. Ko so vse operacije končane, se bo na zaslonu prikazalo pojavno sporočilo, ki vas bo obvestilo o končani pretvorbi.

Za pretvorbo SPSS-datotek iz PISA 2015 in poznejših ciklov sledite enakim korakom kot zgoraj. Ne boste mogli izbrati posameznih držav ali datotek, kot je razloženo v prejšnjem

razdelku. V uporabniškem vmesniku se bo prikazala tabela z vsemi datotekami, ki so na voljo v izbranem imeniku, skupaj z gumbom **Convert user-defined missings to NA** (Pretvori uporabniško določene manjkajoče vrednosti v `NA`) in gumbom **Execute syntax** (Izvedi sintakso):



Kliknite gumb **Execute syntax** (Izvedi sintakso). Pretvorba lahko zaradi velikih velikosti datotek PISA traja precej dolgo.

3.1.6 Pretvorba PISA 2012 in starejših TXT-datotek

Kot je bilo omenjeno, ta možnost velja le za podatke PISA pred letom 2015 (2000, 2003, 2006, 2009, 2012). Ti podatki so bili na voljo v TXT-datotekah skupaj z SPSS- in SAS-kontrolnimi sintaksami za uvoz v SPSS ali SAS ter shranitev v njihovih datotečnih formatih. Funkcija `lsa.convert.data` uporablja TXT-datoteke in ustrezne SPSS-sintakse za pretvorbo podatkov v naravni .Rdata-format. Pomembno je, da prenesete TXT-datoteke in njihove ustrezne SPSS-sintakse z OECD-jeve spletne strani, jih razpakirate iz ZIP-datotek in jih vnesete v isto mapo, brez kakršnih koli sprememb (brez sprememb imen datotek ali njihove vsebine). Tukaj bomo prikazali primer s podatki PISA 2012.

Pretvorba TXT-datotek z uporabo ukazne vrstice

Postopek za pretvorbo podatkov PISA iz ciklov pred letom 2015 je enak kot za SPSS datoteke. Edina razlika je, da morate v argumente klicev izrecno dodati `PISApre15 = TRUE`, da funkcija poišče TXT- in SPS-kontrolne sintakse. Tukaj je primer:

```
lsa.convert.data(inp.folder =
"E:/IDB/PISA_2012_IDB/Data/TXT_Data_Files",
                 PISApre15 = TRUE, out.folder = "C:/temp")
```

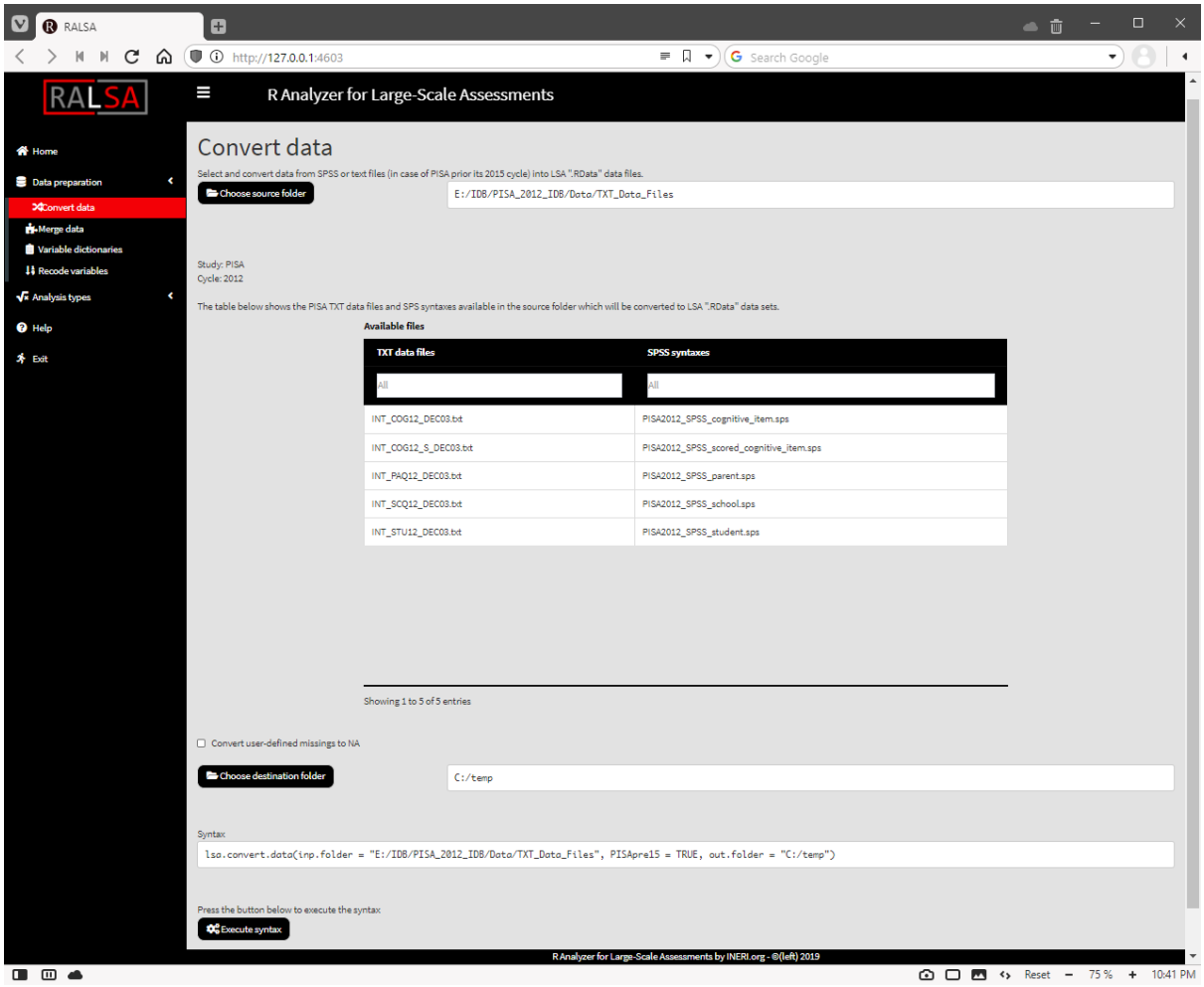
Podobno kot pri SPSS-datotekah bo funkcija med pretvorbo TXT-datotek izpisovala sporočila na zaslon, ko bo pretvorba posamezne datoteke končana. Če funkcija najde SPSS-sintakse in TXT-datoteke brez ustreznih parov, bo pretvorila preostale, za katere so pari najdeni, in ob koncu izpisala opozorilo. Pretvorjene datoteke bodo shranjene v `out.folder`.

Kot je bilo omenjeno, PISA-datoteke vsebujejo podatke vseh držav po vrsti respondentov. Vendar pa analiza morda ne bo zahtevala vključitve vseh držav. RALSA omogoča nalaganje PISA-datoteke, izbiro samo potrebnih držav in shranjevanje rezultatov v novo datoteko. To lahko storite z uporabo funkcije `lsa.select.countries.PISA`. Izvajanje naslednje sintakse bo vzelo pretvorjeno PISA-datoteko, ki vsebuje podatke vseh držav, ki sodelujejo v ciklu 2021, in izbralo samo podatke iz Avstralije in Slovenije ter shranilo podatke teh dveh držav v novo mapo:

```
lsa.select.countries.PISA(data.file =
"C:/temp/pisa2012_spss_student.RData",
                          cnt.names = c("Australia", "Slovenia"),
                          output.file =
"C:/temp/Selected/pisa2012_spss_student.RData")
```

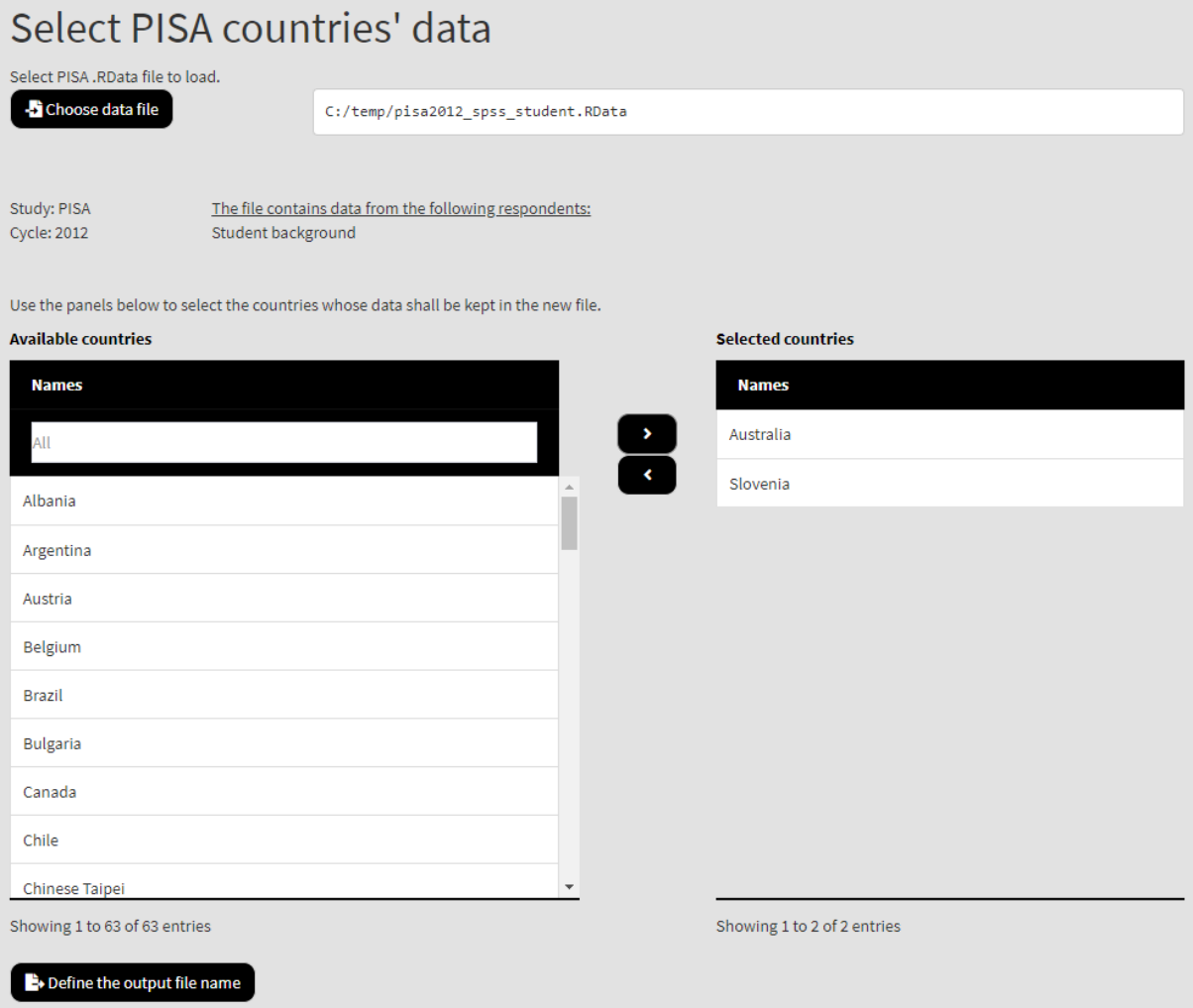
Pretvorba TXT-datotek z uporabo GUI

Postopek za pretvorbo TXT-datotek z uporabo GUI je enak kot pri pretvorbi PISA 2015 in kasnejših SPSS-datotek. Izberite **Data preparation** (Priprava podatkov) > **Convert data** (Pretvori podatke) iz menija na levi. Ko se pomaknete na oddelek za pretvorbo podatkov v GUI, kliknite na **Choose source folder** (Izberi izvorno mapo). Držav ali posameznih datotek ne boste mogli izbrati iz razlogov, pojasnjenih v prejšnjem razdelku. Vmesnik bo prikazal tabelo z vsemi TXT-datotekami in njihovimi ustreznimi kontrolnimi SPSS-datotekami, možnostjo **Convert user-defined missings to NA** (Pretvori uporabniško določene manjkajoče vrednosti v NA) in gumbom **Execute syntax** (Izvedi sintakso).



Kliknite gumb Izvedi sintakso. Sintaksa bo izvedena, vse operacije pa bodo prikazane v oknu konzole na dnu zaslona. To okno se bo posodabljalo pri vsakem koraku. Upoštevajte, da so TXT-datoteke za PISA pred letom 2015 precej velike, saj vsebujejo podatke vseh držav skupaj za vsak tip datoteke, kar lahko povzroči dolge čase pretvorbe.

Kot je bilo omenjeno prej, PISA-datoteke vsebujejo podatke vseh držav po vrsti respondentov. Vendar pa analiza morda ne bo zahtevala vključitve vseh držav. RALSA omogoča nalaganje PISA-datoteke, izbiro samo potrebnih držav in shranjevanje rezultatov v novo datoteko. Da bi to storili, pojdite na **Data Preparation** > **Select PISA countries**. Naložite pretvorjeno PISA-datoteko. GUI bo prikazal seznam **Available Countries** na levi strani. Izberite države, ki vas zanimajo, in uporabite gumb z desno puščico, da jih premaknete na seznam **Selected Countries**.



Kliknite gumb **Define the output file**, pojdite do želenega imenika in vpišite ime datoteke, pod katero bodo shranjeni podatki PISA za izbrane države. GUI bo prikazal gumb **Execute syntax**. Ko ga pritisnete, se bodo izvedle vse operacije. Ko bo datoteka shranjena, bo GUI prikazal sporočilo, da so bile vse operacije zaključene.

3.2 Združevanje podatkovnih datotek raziskav iz različnih držav in/ali respondentov

3.2.1 Uvod

Velike mednarodne raziskave in študije so pogosto mednarodne primerjalne. Vsaka analiza, katere cilj je primerjati države, bi zahtevala večkratno ponavljanje iste analize z uporabo podatkov iz različnih držav ter nato združitev rezultatov vseh držav za primerjavo. Tak postopek morda ni optimalen ali celo primeren. Bolje bi bilo, če bi podatke iz vseh držav združili v enotno zbirko in nato izvedli analizo po državah (ali skupinah znotraj posamezne države, če je treba). To bi zahtevalo najprej združitev podatkov iz vseh držav.

Drug razlog za združevanje podatkov je, da te raziskave zbirajo podatke od različnih respondentov – učencev, njihovih učiteljev, ravnateljev šol, staršev in v nekaterih primerih koordinatorjev IKT v šolah. Raziskovalec je morda zainteresiran za dosežke učencev glede

na starost ali kvalifikacije njihovih učiteljev, ali pa učencev, ki jih nadlegujejo, glede na vrsto območja (urbano ali podeželsko), kjer se nahaja njihova šola. V teh primerih so spremenljivke razporejene v podatke učencev, medtem ko je drugi sklop spremenljivk v podatkih učiteljev ali ravnateljev. To pomeni, da je treba te zbirke podatkov združiti za vsako državo pred združevanjem vseh držav, ki nas zanimajo, in izvedbo analize. Združevanje podatkov različnih respondentov se morda zdi enostavno, vendar lahko postane zelo zapleteno in je odvisno od zasnove raziskave oz. študije. Upoštevajte primer združevanja podatkov učencev in učiteljev z uporabo podatkov TIMSS za 4. razred. TIMSS ne izbere ločenega vzorca učiteljev. Namesto tega vzorči celotne razrede v vsaki izbrani šoli, učitelji v vzorcu pa so le tisti, ki poučujejo izbrane učence. Pri združevanju podatkov učencev in učiteljev moramo zagotoviti, da je med njimi vzpostavljena pravilna povezava, tj. da je vsak učenec povezan le z učiteljem, ki ga poučuje, in z nobenim drugim. Takšno združevanje je lahko še zapletenejše pri podatkih TIMSS za 8. razred, kjer ima vsak učenec ločene učitelje za matematiko in naravoslovje. Da je še zapleteneje, so v nekaterih državah učenci v razredu razdeljeni v skupine, te skupine pa poučuje več različnih učiteljev matematike in naravoslovja, kar še dodatno oteži povezovanje. Tudi zapletenejši scenariji so možni. Pomembno je vzpostaviti pravilno povezavo med učenci in njihovimi starši, učitelji in ravnatelji pri združevanju podatkov več različnih respondentov. Poleg vsega tega nekatere raziskave, razen TIMSS (ICCS in ICILS), ne omogočajo združevanja podatkov učencev in učiteljev zaradi svojega vzorčenja. Vsaka raziskava ima edinstveno zasnovo.

To je naloga funkcije `lsa.merge.data` – upošteva zasnovo vzorčenja raziskave pri združevanju podatkov različnih respondentov. Če raziskovalec zahteva združevanje kombinacije datotek, ki za določeno raziskavo ni mogoča, se funkcija ustavi in prikaže sporočilo o napaki, da prepreči napačne korake. Funkcija prav tako poskrbi za ohranjanje lastnosti spremenljivk, kot so oznake razredov, spremenljivk in vrednosti. Zato za to nalogo močno priporočamo uporabo funkcije `lsa.merge.data`, da se izognete napakam, ki bi lahko vplivale na izračune v analizi. Uporaba drugih metod za združevanje podatkov pomeni, da tveganje prevzemate sami.

Upoštevajte, da funkcija deluje samo v primeru raziskav, kjer so datoteke za vsak cikel na voljo po državah, po tipu respondentov (npr. učenec, učitelj itd.). To vključuje vse podprte raziskave razen PISA. V vsakem ciklu PISA zagotavlja datoteke po tipu respondentov, pri čemer vsaka datoteka vsebuje podatke za vse države v vsaki izmed datotek. To, skupaj z večanjem konvencij o poimenovanju datotek, ki se spreminjajo v vsakem ciklu, ne omogoča iskanja zanesljivega pristopa za združevanje njihovih podatkov.

3.2.2 Funkcija za združevanje podatkov in njeni argumenti

Funkcija `lsa.merge.data` ima naslednje argumente:

- `inp.folder` – mapa, ki vsebuje zbirke podatkov. Zbirke podatkov morajo biti v formatu .RData, ki ga je ustvarila funkcija `lsa.convert.data`.
- `file.types` – katere tipe datotek (tj. respondente) je treba združiti.
- `ISO` – vektor, ki vsebuje ISO-kode državnih datotek, ki jih je treba vključiti v združeno datoteko.

- `out.file` – polna pot do datoteke, v katero bodo shranjeni podatki. Objekt, shranjen v datoteki, bo imel isto ime.

Opombe:

- Funkcija združuje datoteke iz raziskav, kjer so datoteke na voljo po državah in tipih respondentov (npr. učenec, šola, učitelj). To vključuje vse raziskave razen PISA.
- `inp.folder` določa pot do mape, ki vsebuje .Rdata-datoteke, ki jih je ustvarila funkcija `lsa.convert.data`. Mapa mora vsebovati samo datoteke za eno raziskavo, en cikel in eno populacijo (npr. TIMSS 2015 4. razred ali TIMSS 2015 8. razred, ne pa obeh) ali način izvedbe (npr. PIRLS 2016 ali ePIRLS 2016, ne pa obeh; ali TIMSS 2019 ali TIMSS 2019 Bridge, ne pa obeh). Vse datoteke v vhodni mapi morajo biti izvožene z enakimi nastavitvami (TRUE ali FALSE) argumenta `missing.to.NA` funkcije `lsa.convert.data`. Če mapa ni navedena v argumentu, bo uporabljena delovna mapa (`getwd()`).
- Mapa mora vsebovati samo datoteke za eno raziskavo, en cikel in eno populacijo. V nasprotnem primeru bo funkcija ustavljena z napakami. To je izvedeno, da se ohrani čista struktura map in prepreči naključne napake. Prednostna je uporaba `out.file`, ki je drugačna od `inp.folder`, iz istega razloga.
- `file.types` je seznam tipov respondentov s komponentnimi imeni in z njihovimi spremenljivkami, ki jih je treba združiti. Imena tipov datotek so tričrkovne kode, prvi trije znaki ustrezajo imenom datotek. Elementi so vektorji imen spremenljivk z velikimi črkami, `NULL` pomeni vse spremenljivke v ustrezni datoteki. Npr., v TIMSS bo »asg« združil samo podatke na ravni učencev za 4. razred, `c(asg, atg)` bo združil podatke na ravni učencev in učiteljev za 4. razred, `c(bsg, btm)` bo združil podatke na ravni učencev in učiteljev matematike za 8. razred. Če združevanje ni mogoče zaradi zasnove raziskave, bo funkcija ustavljena z napako.
- `ISO` je vektor znakov, ki določa države, katerih podatke je treba združiti. Elementi v vektorju so četrti, peti in šesti znaki v imenih datotek. Npr., `c("aus", "swe", "svn")` bo združil podatke iz Avstralije, Švedske in Slovenije za tipe datotek, določene v `file.types`. Tričrkovne ISO-kode za vsako državo lahko najdete v priročniku raziskave v obravnavi. Npr., ISO-kode držav, ki sodelujejo v PIRLS 2016, so navedene v priročniku na straneh 52–54. Če datoteke za določeno državo v `inp.folder` ne obstajajo, bo izpisano opozorilo. Če argument `ISO` manjka, bodo združeni podatki za vse države v mapi za določene tipe datotek.
- `out.file` mora vsebovati polno pot (vključno z razširitvijo .RData, če manjka, bo dodana) do izhodne datoteke (tj. datoteke, ki vsebuje združene podatke). Datoteka vsebuje objekt z enakim imenom in ima razširitev `lsa.data`. Ima dodaten atribut `file.type`, ki prikazuje podatke, iz katerih respondentov so na voljo po zaključenem združevanju. Npr., združevanje podatkov na ravni učencev s podatki učiteljev v TIMSS-4. razredu bo temu atributu dodelilo »std.bckg.tch.bckg«. Objekt ima dva dodatna atributa: ime raziskave (study) in cikel raziskave (cycle). Objekt v .Rdata-datoteki je indeksiran po ID-spremenljivki države. Če izhodna mapa ni navedena, bo združena datoteka shranjena v delovni mapi (`getwd()`) kot `merged_data.RData`.

3.2.3 Združevanje podatkov preko ukazne vrstice

Kot prvi primer bomo združili datoteke vprašalnikov za v PIRLS 2016 sodelujoče učence iz Avstralije in Slovenije, ki smo jih pretvorili v prejšnjem koraku. V RStudiu izvedite naslednji ukaz:

```
lsa.merge.data(inp.folder = "C:/temp", ISO = c("aus", "svn"),
               file.types = list(asg = NULL),
               out.file =
"C:/temp/merged/PIRLS_2016_ASG_AUS_SVN.RData")
```

Argument `inp.folder` usmerja v mapo, kjer smo v prejšnjem koraku shranili pretvorjene datoteke. Argument `ISO` sprejme vektor znakov z dvomičnimi ISO-kodami držav (to so četrti, peti in šesti znaki v imenih datotek). Če želite združiti datoteke iz vseh držav, ki so na voljo v mapi, preprosto izpustite ta argument. Argument `out.file` določa, kje shraniti združeno datoteko. Tudi tukaj je priporočljivo, da ohranite stvari urejene in shranite datoteko v mapo, ki je drugačna od `inp.folder`.

Argument `file.types` je zapletenejši. Biti mora posredovan kot seznam, kjer je vsaka komponenta tip datoteke z okrajšavo (prvi do tretji znak v imenih datotek) in ima vrednost. V PIRLS 2016 okrajšava »asg« pomeni podatke iz vprašalnika za ozadje učencev. Tako, kot je navedeno v zgornjem klicu (`asg = NULL`), `NULL` pomeni »vzemite vse spremenljivke v datotekah za ta tip datotek«. Kaj pa, če želimo izbrati le določene spremenljivke, ki jih bomo združili, recimo spol učencev (ASBG01), število knjig doma (ASBG04) in koliko učenci uživajo v branju (ASBR06E)? Takrat moramo namesto `NULL` posredovati vektor znakov za `asg` v argumentu. Sintaksa bo videti takole:

```
lsa.merge.data(inp.folder = "C:/temp", ISO = c("aus", "svn"),
               file.types = list(asg = c("ASBG01", "ASBG04",
"ASBR06E")),
               out.file =
"C:/temp/merged/PIRLS_2016_ASG_AUS_SVN.RData")
```

Upoštevajte, da čeprav nobene od identifikacijskih, sledenjskih in utežnih spremenljivk pa tudi imena PV niso vključeni v vektor spremenljivk za `asg`, bodo avtomatsko dodani v združeno datoteko, za to se vam ni treba bati. To velja za vse tipe datotek pri združevanju. Lahko naredimo še zapletenejši primer z združevanjem vseh spremenljivk iz PIRLS 2016-datotek za učence in učitelje iz Avstralije in Slovenije, in sicer takole:

```
lsa.merge.data(inp.folder = "C:/temp", ISO = c("aus", "svn"),
               file.types = list(asg = NULL, atg = NULL),
               out.file =
"C:/temp/merged/PIRLS_2016_ASG_ATG_AUS_SVN.RData")
```

In če želimo združiti le nekatere spremenljivke v obeh tipih datotek, lahko to storimo takole:

```
lsa.merge.data(inp.folder = "C:/temp", ISO = c("aus", "svn"),
```

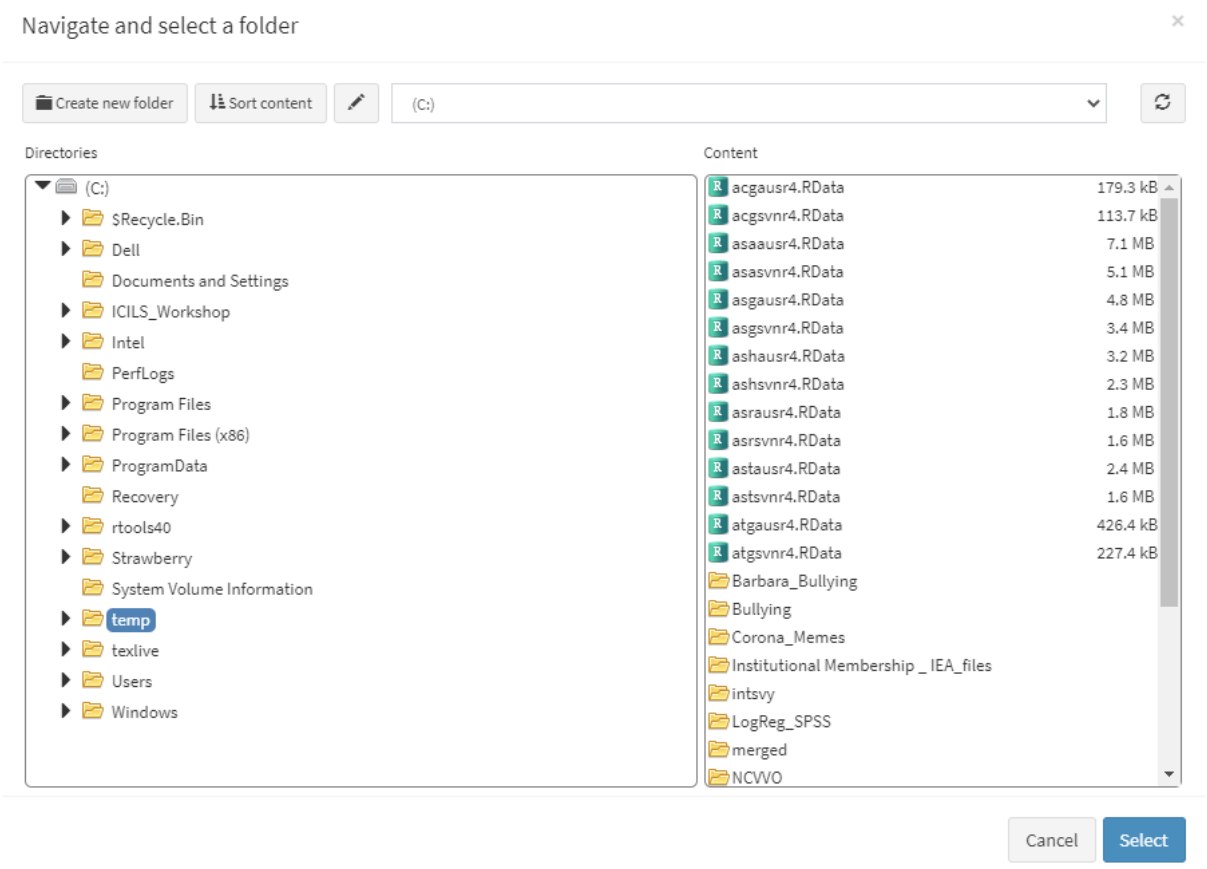
```
                file.types = list(asg = c("ASBG01", "ASBG04",
"ASBR06E"), atg = c("ATBG05BF", "ATBG05BG", "ATBG05BH")),
                out.file =
"C:/temp/merged/PIRLS_2016_ASG_ATG_AUS_SVN.RData")
```

3.2.4 Združevanje podatkov preko GUI

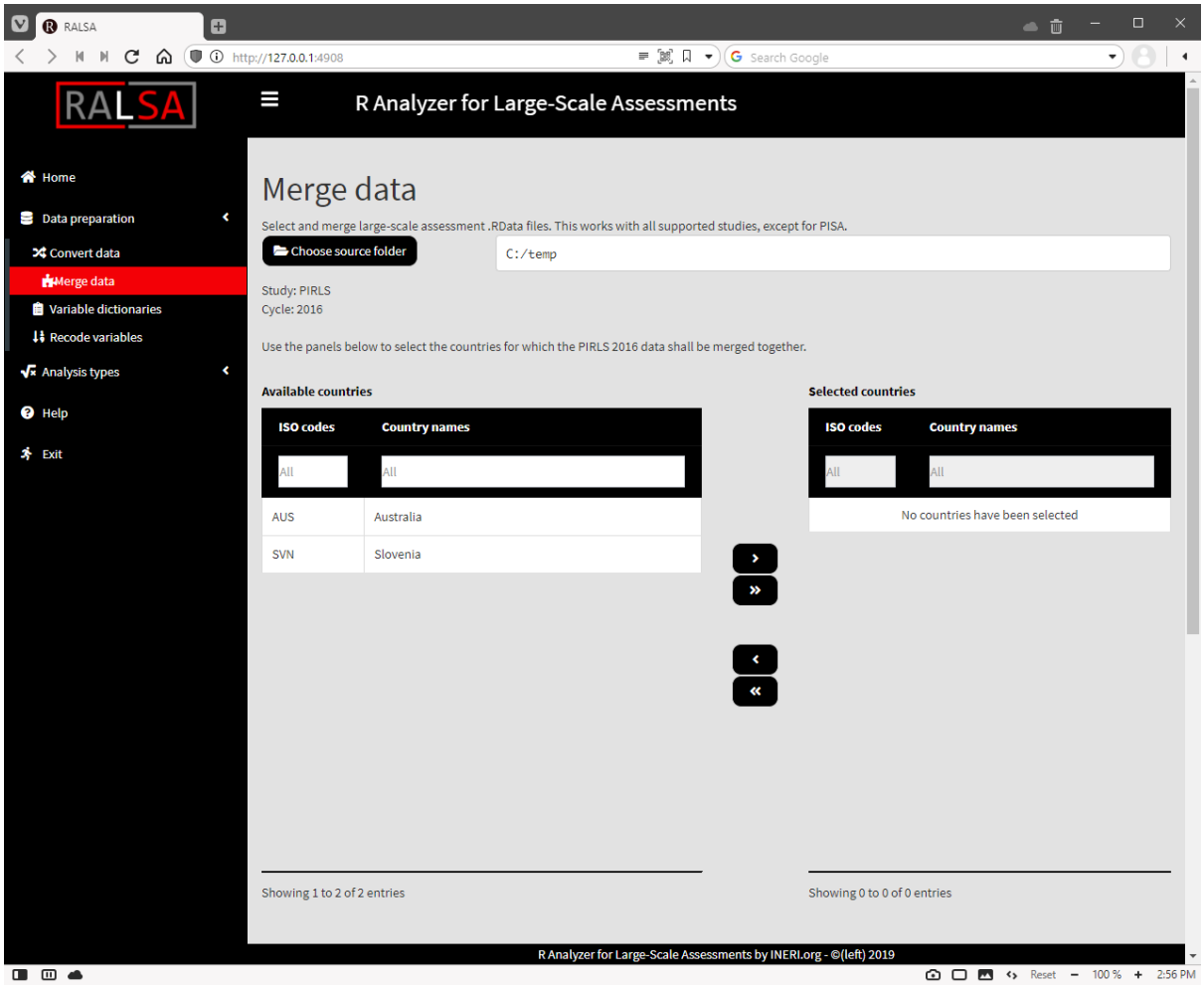
Za zagon uporabniškega vmesnika RALSA v RStudiu izvedite naslednji ukaz:

```
ralsaGUI()
```

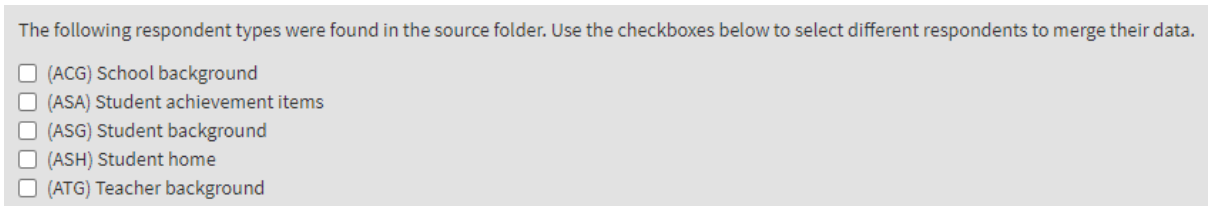
Ko se GUI odpre v vašem brskalniku, izberite **Data preparation** > **Merge data** iz menija na levi strani. Ko se premaknete v razdelek **Merge data** v GUI, kliknite gumb **Choose source folder**. Navigirajte do mape, ki vsebuje pretvorjene .Rdata-datoteke za raziskavo (npr. PIRLS 2016). Na levem panelu okna za izbiro mape izberite mapo, ki vsebuje podatke. Na voljo .Rdata-datoteke v mapi bodo prikazane na desnem panelu:



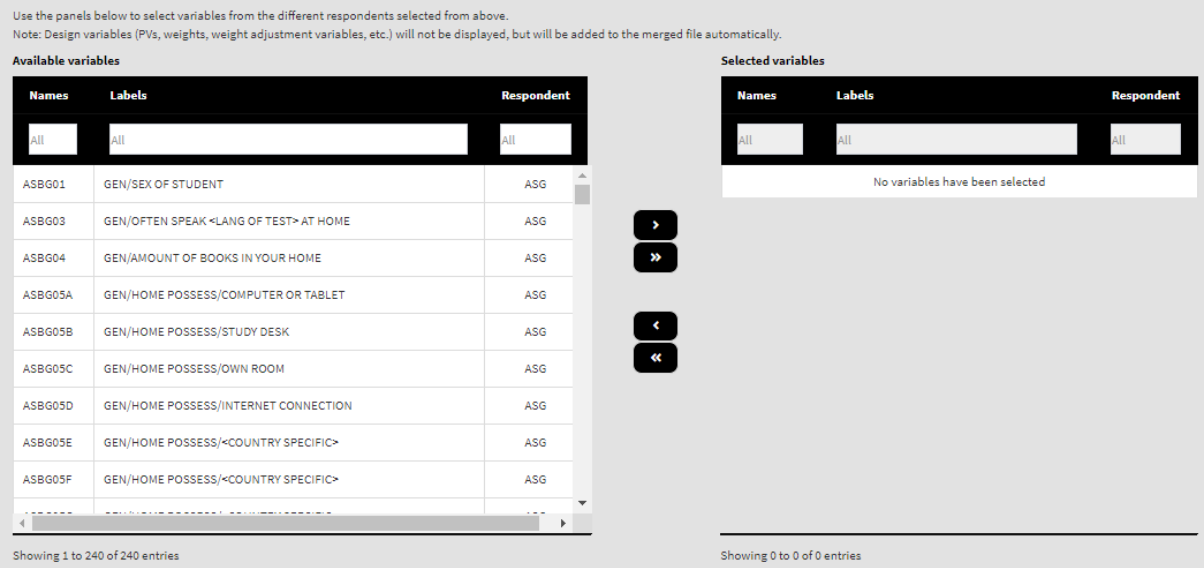
Ko izbiro potrdite z izborom možnosti **Select**, se bo prikazal naslednji zaslon:



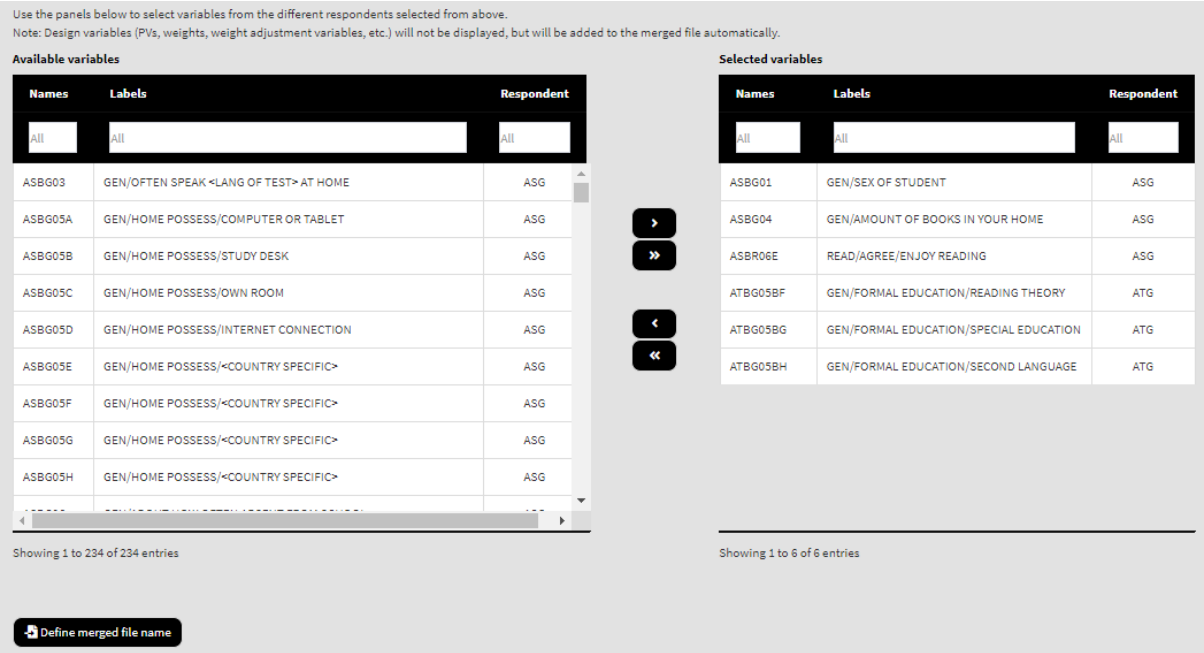
V prejšnjem primeru smo podatke pretvorili le za dve državi – Avstralijo in Slovenijo. Če so v izvorni mapi shranjeni podatki za več držav, bodo ti prav tako prikazani. Z miško izberite in uporabite enojne ter dvojne puščice za premikanje držav med seznamoma **Available countries** in **Selected countries**. Enojne puščice lahko uporabite za izbiro in premikanje posameznih držav med obema seznamoma. Dvojne puščice lahko uporabite za premikanje vseh držav med seznamoma, tudi če nobena ni izbrana. Izberimo obe državi, Avstralijo in Slovenijo. Pojavila se bo skupina več potrditvenih polj za vrsto podatkovnih datotek respondentov, ki so bile najdene v mapi, eno polje za vsako vrsto podatkov:



Lahko izberete več vrst datotek za združevanje. Če združitev ni mogoča (kar bo odvisno od zasnove vzorčenja raziskave, glejte uporabniški priročnik in tehnično poročilo določene raziskave), se bo prikazalo opozorilno sporočilo. V tem primeru bomo združili podatkovne nize za ozadje učencev in učiteljev. Zato označite potrditveni polji »(ASG) Ozadje učencev« in »(ATG) Ozadje učiteljev«. Ko to storite, se spodaj pojavita še dva dodatna panela:

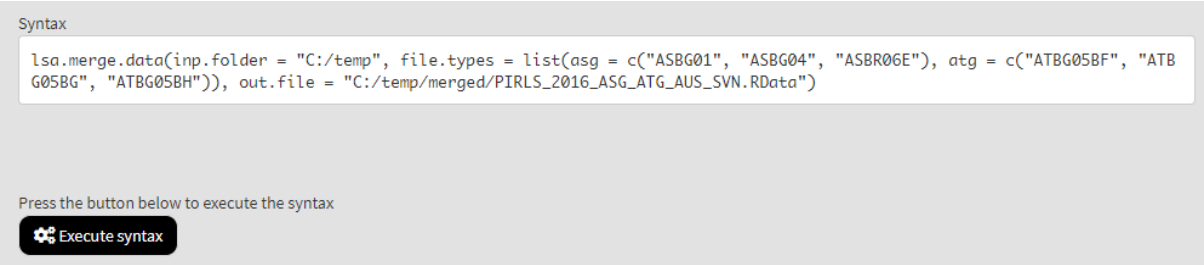


Zadnji stolpec na vsakem panelu prikazuje, kateremu tipu respondentov zgoraj pripada spremenljivka. Uporabite miško za izbiro posameznih spremenljivk in enojne puščice za premik s seznama razpoložljivih spremenljivk na seznam izbranih spremenljivk ter obratno. Dvojne puščice lahko uporabite za izbiro vseh ali nobenih spremenljivk. Filter polja na vrhu panelov lahko uporabite za hitro iskanje potrebnih spremenljivk. Za ta primer izberimo enake spremenljivke, kot smo jih izbrali v prejšnjem primeru pri uporabi možnosti ukazne vrstice (`asg = c("ASBG01", "ASBG04", "ASBR06E"), atg = c("ATBG05BF", "ATBG05BG", "ATBG05BH")`). Prosimo, upoštevajte, da bodo vse identifikacijske, sledenjske in načrtovalne spremenljivke (vzorčenje, uteži in PV) samodejno dodane k združeni datoteki. Ko so na panele izbranih spremenljivk dodane spremenljivke, se prikaže gumb za določitev imena združene datoteke:

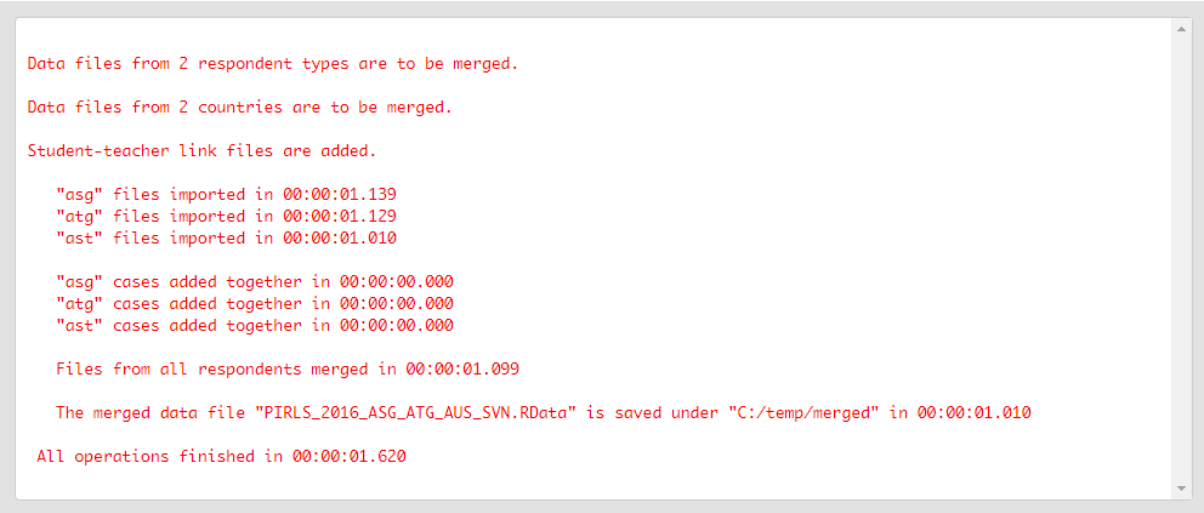


Kliknite gumb za določitev imena združene datoteke, pojdite v mapo, kjer želite shraniti združeno datoteko, in določite želeno ime datoteke. Shranimo jo lahko kot

»C:/temp/merged/PIRLS_2016_ASG_ATG_AUS_SVN.RData«. Ko to storite, se spodaj prikaže gumb za izvedbo sintakse (**Execute syntax**)



Kliknite gumb za izvedbo sintakse. Sintaksa bo izvedena, vse operacije pa bodo prikazane v oknu konzole na dnu zaslona. Posodabljale se bodo pri vsakem koraku:



Ko bodo vse operacije zaključene, se bo na zaslonu pojavilo obvestilno sporočilo.

3.3 Slovarji spremenljivk (angl. *Variable dictionaries*)

3.3.1 Uvod

Pri izvajanju analize je treba vnaprej poznati lastnosti vseh vključenih spremenljivk. Funkcija `lsa.vars.dict` ustvarja slovarje spremenljivk, ki vključujejo informacije o imenih spremenljivk, razredih (numerični, faktorski ali znakovni), oznakah kot tudi o njihovih ravneh (tj. odgovornih kategorijah pri faktorskih spremenljivkah) ali edinstvenih vrednostih (pri numeričnih ali znakovnih spremenljivkah) ter uporabniško določenih manjkajočih vrednostih (če obstajajo). Funkcija v konzoli R/RStudio vedno izpiše slovarje na zaslon in ponuja možnost shranjevanja le-teh v besedilno datoteko. Močno priporočamo uporabo te možnosti, saj shranjevanje slovarjev spremenljivk v datoteko omogoča nadaljnje sklicevanje pri rekodiranju spremenljivk (kar bo obravnavano v naslednjem poglavju) ali pri nastavljanju analize. Izhod funkcije `lsa.vars.dict` je jedrnat, vendar informativen glede lastnosti spremenljivk in zagotavlja dovolj podrobnosti. Uporabljajo se lahko pretvorjene .RData-datoteke mednarodnih raziskav ali datoteke, kjer so združeni različne države in/ali tipi respondentov.

3.3.2 Funkcija za slovarje spremenljivk in njeni argumenti

Funkcija `lsa.vars.dict` ima naslednje argumente:

- `data.file` – celotna pot do .RData-datoteke, ki vsebuje objekt `lsa.data`. Lahko se določi ta argument ali `data.object`, vendar ne oboje.
- `data.object` – objekt v spominu, ki vsebuje objekt `lsa.data`. Lahko se določi ta argument ali `data.file`, vendar ne oboje.
- `var.names` – vektor imen spremenljivk, katerih slovarji bodo ustvarjeni.
- `out.file` – neobvezno; celotna pot do .txt-datoteke, kamor bodo shranjeni slovarji, če je treba.
- `open.out.file` – neobvezno; če je navedena pot do datoteke v `out.file`, ali naj se po pisanju datoteke ta odpre.

Opombe:

- Slovarji za spremenljivke v `var.names` bodo izpisani kot tabele na zaslonu. Za vsako spremenljivko bodo slovarji vsebovali ime spremenljivke, razred spremenljivke, oznako spremenljivke, edinstvene vrednosti spremenljivke (glejte spodaj) in uporabniško določene manjkajoče vrednosti (če obstajajo).
- Predstavitev edinstvenih vrednosti bo odvisna od razreda spremenljivke. Če je spremenljivka faktorska, bodo prikazane ravni faktorjev. Če je spremenljivka numerična ali znakovna, bodo natisnjene edinstvene vrednosti do šeste.
- Uporabniško določene manjkajoče vrednosti za faktorske spremenljivke bodo prikazane kot besedilni nizi. Za numerične spremenljivke bodo to cela števila, ki jim sledijo oznake v oklepajih.
- Če je za `out.file` navedena celotna pot, bo enak izhod zapisan v .txt-datoteko z besedilom na vrhu, ki sporoča, katera datoteka/objekt je bila uporabljena.

3.3.3 Prikaz in shranjevanje slovarjev spremenljivk z uporabo ukazne vrstice

V tem primeru bomo uporabili datoteko, združeno v prejšnjem primeru z uporabo ukazne vrstice, za vse spremenljivke v datoteki (če izpustimo argument `var.names`, bo funkcija ustvarila slovarje za vse spremenljivke v datoteki). V RStudiu izvedite naslednjo sintakso:

```
lsa.vars.dict(data.file =
"C:/temp/merged/PIRLS_2016_ASG_ATG_AUS_SVN.RData")
```

Funkcija naloži datoteko, ustvari slovarje in jih natisne v konzoli RStudia. Celotnega izhoda s slovarji za vseh 63 spremenljivk v datoteki ni mogoče prikazati, spodnji posnetki zaslona prikazujejo prve in zadnje spremenljivke.

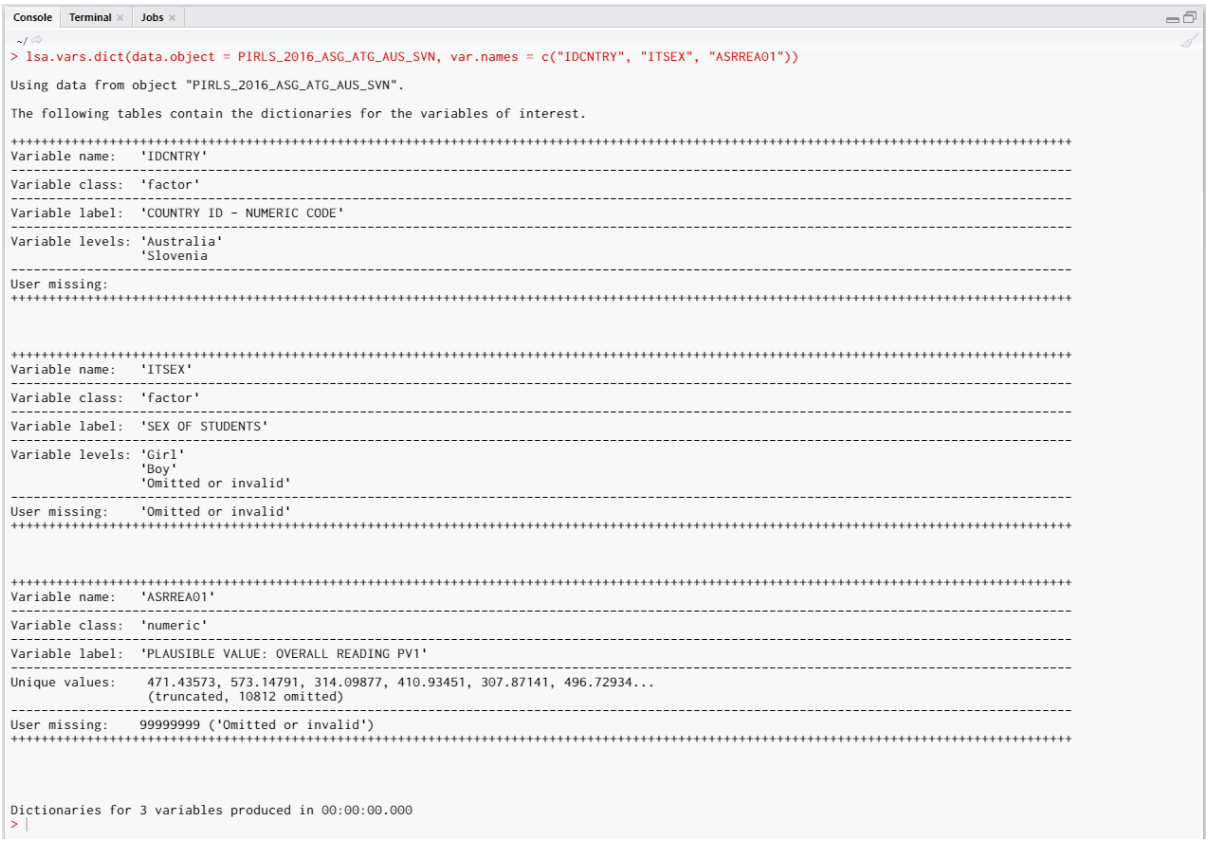
```
Console  Terminal  Jobs
~/
> lsa.vars.dict(data.file = "C:/temp/merged/PIRLS_2016_ASG_ATG_AUS_SVN.RData", var.names = c("IDCNTRY", "IDSTUD", "ASBG01", "ASBG04", "ASBR06
E", "ITSEX", "ASRIBM04", "ASRIBM05", "VERSION", "SCOPE" ))

Data file "C:/temp/merged/PIRLS_2016_ASG_ATG_AUS_SVN.RData" imported in 00:00:01.050

The following tables contain the dictionaries for the variables of interest.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Variable name:   'IDCNTRY'
----------------------------------------------------------------------------------------------------------------------------------------
Variable class:  'factor'
----------------------------------------------------------------------------------------------------------------------------------------
Variable label:  'COUNTRY ID - NUMERIC CODE'
----------------------------------------------------------------------------------------------------------------------------------------
Variable levels: 'Australia'
                 'Slovenia
----------------------------------------------------------------------------------------------------------------------------------------
User missing:
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++



++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Variable name:   'IDSTUD'
----------------------------------------------------------------------------------------------------------------------------------------
Variable class:  'numeric'
----------------------------------------------------------------------------------------------------------------------------------------
Variable label:  'STUDENT ID'
----------------------------------------------------------------------------------------------------------------------------------------
Unique values:    10501, 10502, 10503, 10504, 10505, 10506...
                  (truncated, 9507 omitted)
----------------------------------------------------------------------------------------------------------------------------------------
User missing:
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++



++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Variable name:   'ASBG01'
----------------------------------------------------------------------------------------------------------------------------------------
Variable class:  'factor'
----------------------------------------------------------------------------------------------------------------------------------------
Variable label:  'GEN/SEX OF STUDENT'
----------------------------------------------------------------------------------------------------------------------------------------
Variable levels: 'Girl'
                 'Boy'
                 'Omitted or invalid'
----------------------------------------------------------------------------------------------------------------------------------------
User missing:    'Omitted or invalid'
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
```

```
Console  Terminal  Jobs
~/
Variable name:   'ASRIBM05'
----------------------------------------------------------------------------------------------------------------------------------------
Variable class:  'factor'
----------------------------------------------------------------------------------------------------------------------------------------
Variable label:  'INT. READING SCALE BENCHMARK REACHED'
----------------------------------------------------------------------------------------------------------------------------------------
Variable levels: 'BELOW 400'
                 'AT OR ABOVE 400 BUT BELOW 475'
                 'AT OR ABOVE 475 BUT BELOW 550'
                 'AT OR ABOVE 550 BUT BELOW 625'
                 'AT OR ABOVE 625'
                 'Omitted or invalid'
----------------------------------------------------------------------------------------------------------------------------------------
User missing:    'Omitted or invalid'
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++



++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Variable name:   'VERSION'
----------------------------------------------------------------------------------------------------------------------------------------
Variable class:  'numeric'
----------------------------------------------------------------------------------------------------------------------------------------
Variable label:  'VERSION'
----------------------------------------------------------------------------------------------------------------------------------------
Unique values:    4
----------------------------------------------------------------------------------------------------------------------------------------
User missing:    99999999 ('Omitted or invalid')
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++



++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Variable name:   'SCOPE'
----------------------------------------------------------------------------------------------------------------------------------------
Variable class:  'factor'
----------------------------------------------------------------------------------------------------------------------------------------
Variable label:  'SCOPE OF THIS FILE'
----------------------------------------------------------------------------------------------------------------------------------------
Variable levels: 'Public Use File (PUF)'
                 'Restricted Use File (RUF)'
                 'Omitted or invalid'
----------------------------------------------------------------------------------------------------------------------------------------
User missing:    'Omitted or invalid'
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++



Dictionaries for 63 variables produced in 00:00:01.169
>
```

Output (izhodno datoteko) lahko omejimo samo na tiste spremenljivke, ki nas zanimajo. Ustvarimo slovarje za identifikacijsko številko države (IDCNTRY), spol učenca (ITSEX) in prvo verjetno vrednost za splošno branje (ASRREA01). Tokrat bomo uporabili objekt `lsa.data` v pomnilniku. V ta namen bomo najprej naložili datoteko **»PIRLS_2016_ASG_ATG_AUS_SVN.RData«** in jo uporabili v klicu funkcije prek argumenta `data.object` namesto argumenta `data.file`. Celotna koda izgleda tako:

```
load("C:/temp/merged/PIRLS_2016_ASG_ATG_AUS_SVN.RData")

lsa.vars.dict(data.object = PIRLS_2016_ASG_ATG_AUS_SVN,
              var.names = c("IDCNTRY", "ITSEX", "ASRREA01"))
```

Bodite pozorni na to, da smo najprej naložili združeno datoteko. Datoteka vsebuje objekt z enakim imenom kot ime datoteke, brez pripone »RData«. V vrstici 4 zgoraj uporabljamo objekt, ki je zdaj v RAM-u, in ne več datoteke. Imena spremenljivk, za katere želimo slovarje, posredujemo argumentu `var.names`. Izhod v konzoli RStudia izgleda takole:



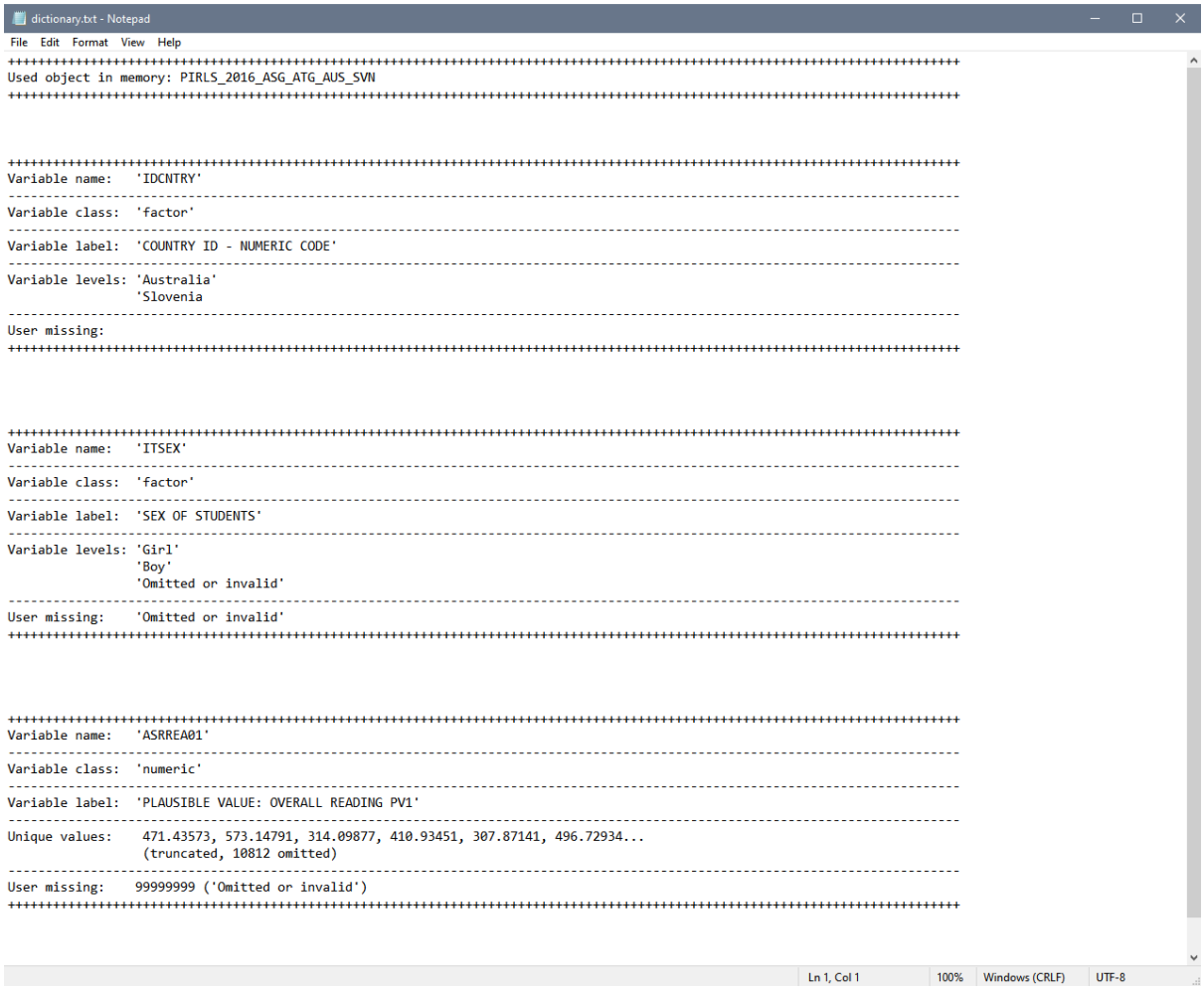
Za vsako spremenljivko je prikazana ločena tabela, ki predstavlja njene lastnosti: ime, razred, oznako, ravni/unikatne vrednosti in uporabniško definirane manjkajoče vrednosti, če obstajajo. Bodite pozorni na unikatne vrednosti za zadnjo spremenljivko, ki je prva verjetna vrednost za splošno branje. To je zvezna spremenljivka, ki lahko zavzame skoraj katero koli vrednost. Output bi bil precej dolg, če bi bile predstavljene vse razpoložljive vrednosti. Zato je prikazanih samo prvih šest, tabela pa nas obvešča, koliko več unikatnih vrednosti je v tej spremenljivki. Prav tako bodite pozorni na razliko v uporabniško definiranih manjkajočih vrednostih med faktorji in številskimi spremenljivkami (ITSEX in ASRREA01). Za faktorje je ena od ravni

dodeljena kot manjkajoča vrednost. Za številčne spremenljivke je manjkajoča vrednost dodeljena kot poimenovana številčna vrednost.

Vse to deluje dobro. Vendar pa bi pogosto morali pregledati slovarje za veliko spremenljivk, celo za vse spremenljivke znotraj podatkovne datoteke. Priročneje bi bilo shraniti slovarje v datoteko in jih kasneje uporabiti kot referenco. Da to dosežemo, dodajmo argumenta `out.file` in `open.out.file` v zgornjo sintakso:

```
lsa.vars.dict(data.object = PIRLS_2016_ASG_ATG_AUS_SVN,
              var.names = c("IDCNTRY", "ITSEX", "ASRREA01"),
              out.file = "C:/temp/merged/dictionary.txt",
              open.out.file = TRUE)
```

Prvi argument (`out.file`) funkciji pove, kam naj shrani output/izhodno datoteko (tekstovna datoteka z vsemi slovarji), tako, da navede pot do nje, drugi (`open.out.file`) pa funkciji naroči, naj odpre datoteko, potem ko so vsi slovarji za zahtevane spremenljivke ustvarjeni. Besedilna datoteka se bo odprla v privzetem programu, povezanem z besedilnimi datotekami.

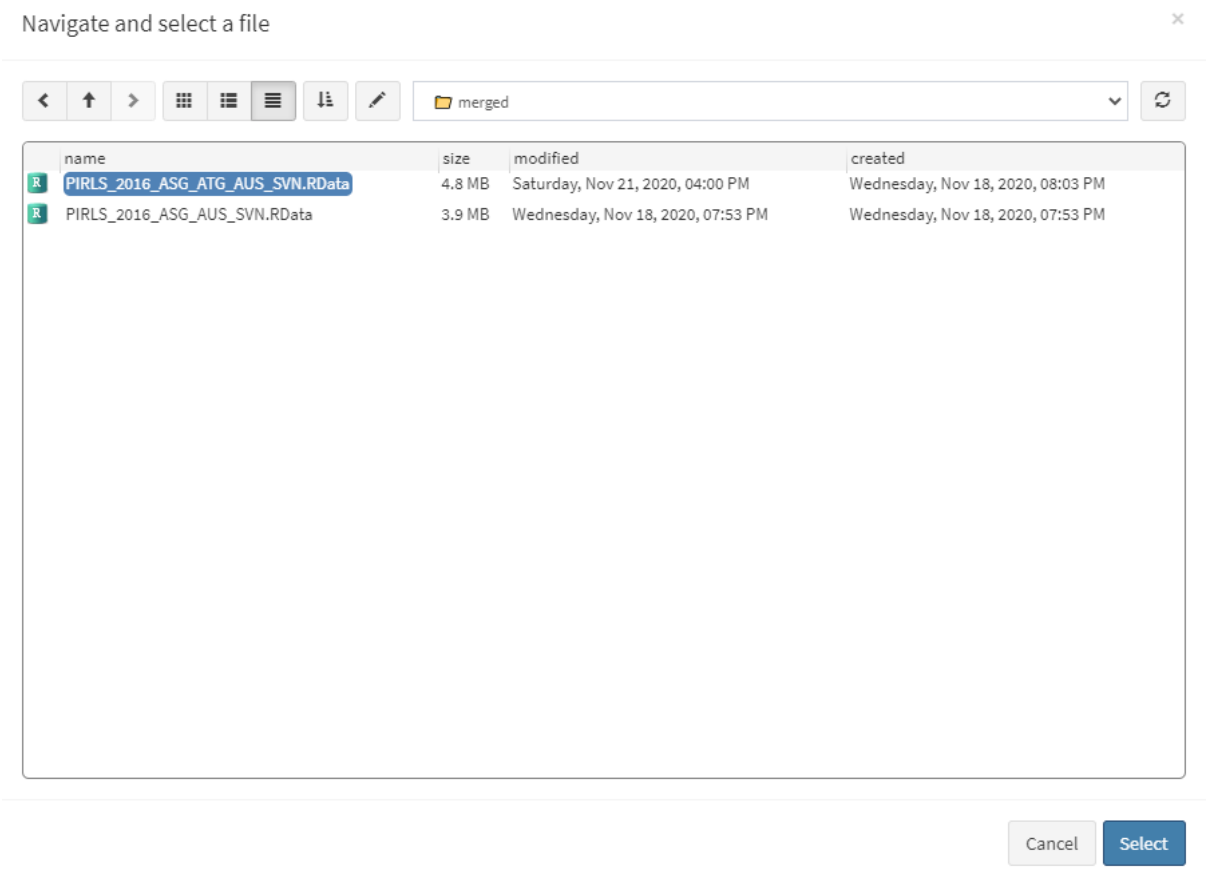


3.3.4 Prikazovanje in shranjevanje slovarjev spremenljivk z uporabo GUI (grafičnega uporabniškega vmesnika)

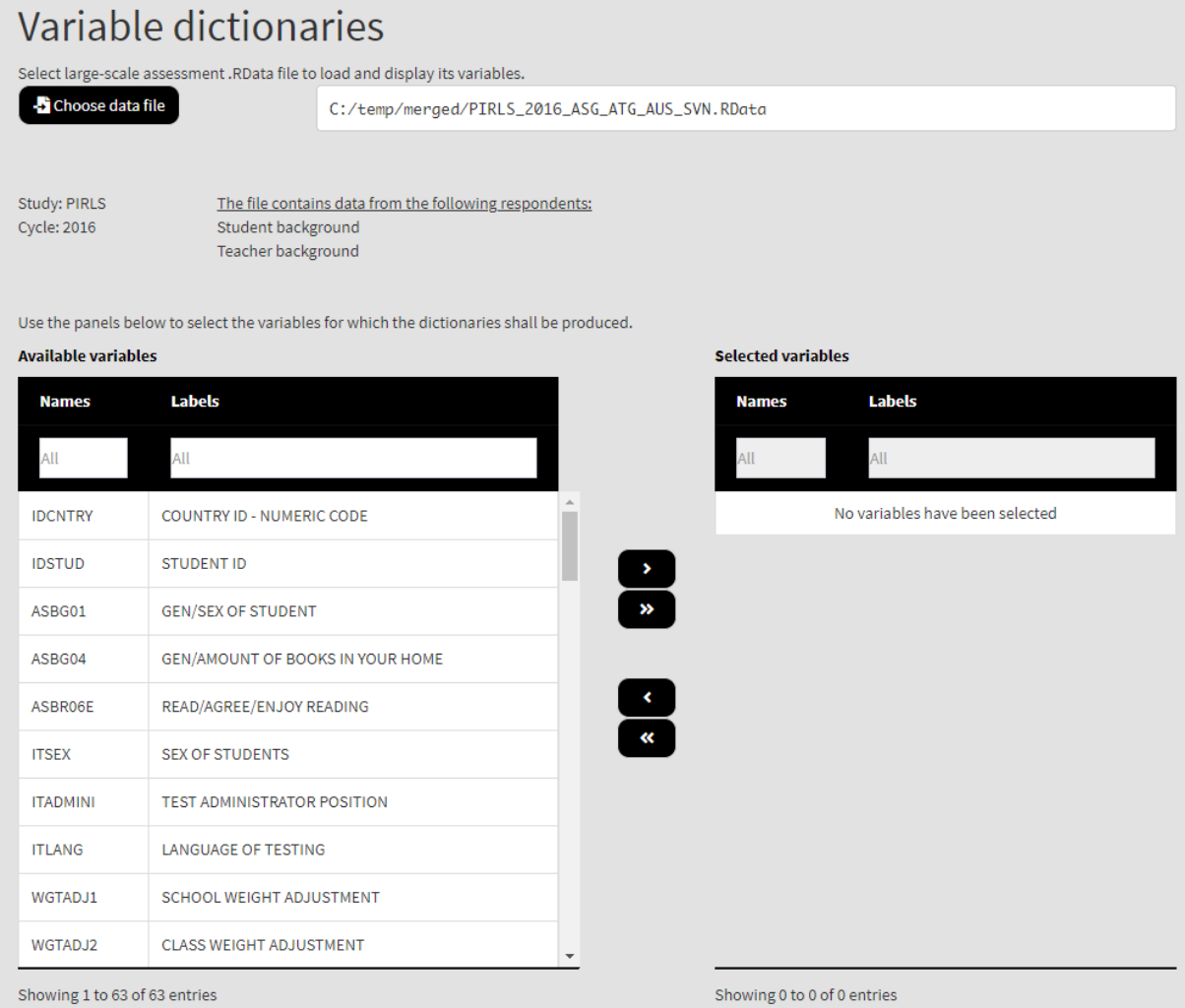
Za začetek uporabniškega vmesnika RALSA v RStudiu izvedite naslednji ukaz:

```
ralsaGUI()
```

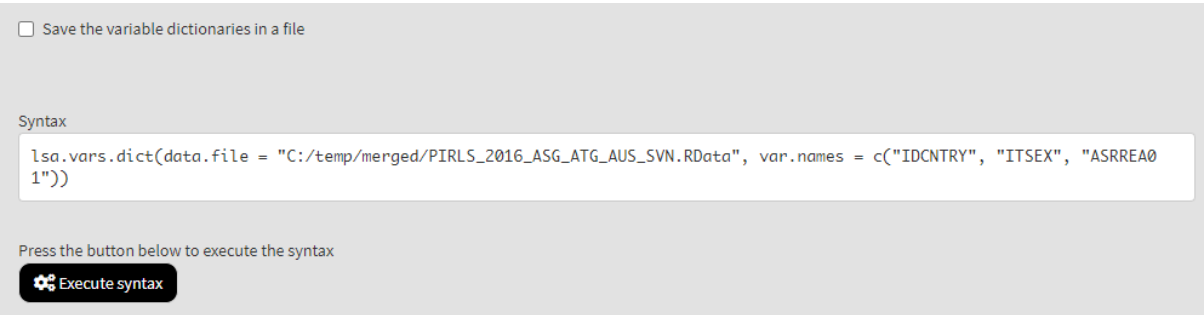
Ko se GUI odpre v vašem brskalniku, izberite **Data preparation** (Priprava podatkov) > **Variable dictionaries** (Slovarji spremenljivk) iz menija na levi. Ko se v GUI premaknete na Variable dictionaries, kliknite gumb **Choose data file** (Izberi podatkovno datoteko). Pomaknite se do mape, ki vsebuje združeno datoteko **PIRLS_2016_ASG_ATG_AUS_SVN.RData**, jo izberite in kliknite gumb **Select** (Izberi).



Ko je datoteka naložena, boste videli dva panela z razpoložljivimi spremenljivkami in izbranimi spremenljivkami (slednji je trenutno prazen):



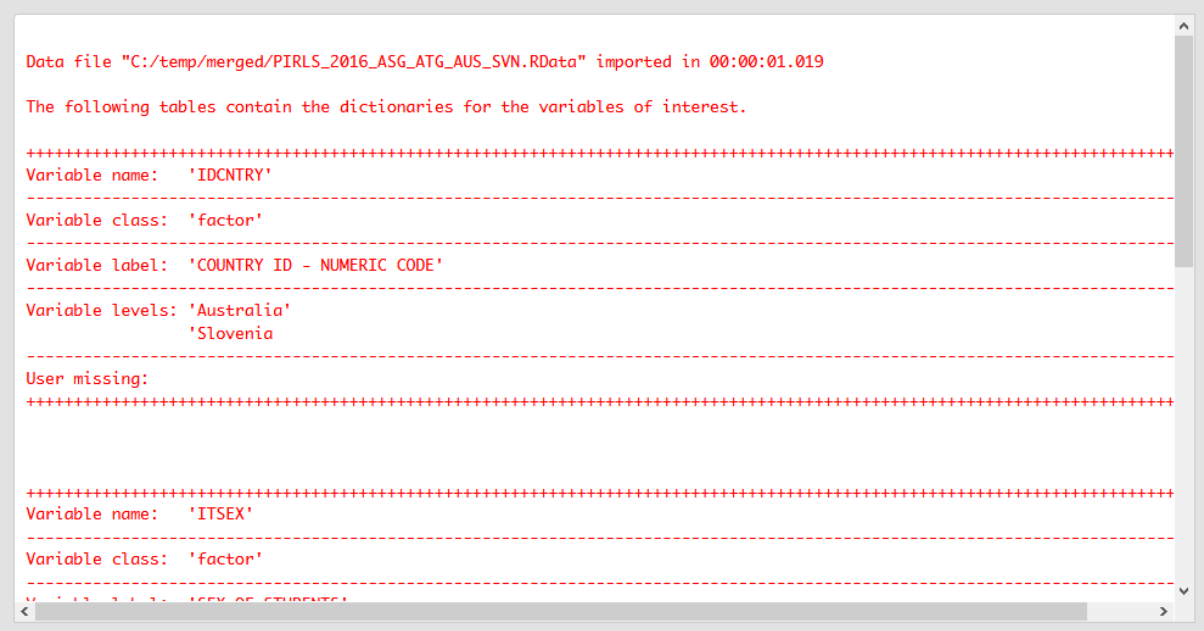
Uporabite miško za izbiro posameznih spremenljivk in enojne puščice za premik s seznama **Available variables** (razpoložljive spremenljivke) na seznam **Selected variables** (izbrane spremenljivke) ter obratno. Dvojne puščice lahko uporabite za izbiro vseh ali nobenih spremenljivk. Filter polja na vrhu panelov lahko uporabite za hitro iskanje potrebnih spremenljivk. Izberimo spremenljivke IDCNTRY (identifikacija države), ITSEX (sledilna spremenljivka za spol učenca) in ASRREA01 (prva verjetnostna vrednost za celoten bralni dosežek učenca). Ko so na panelu **Selected variables** (izbrane spremenljivke) dodane kakršne koli spremenljivke, se bodo prikazali naslednji elementi:



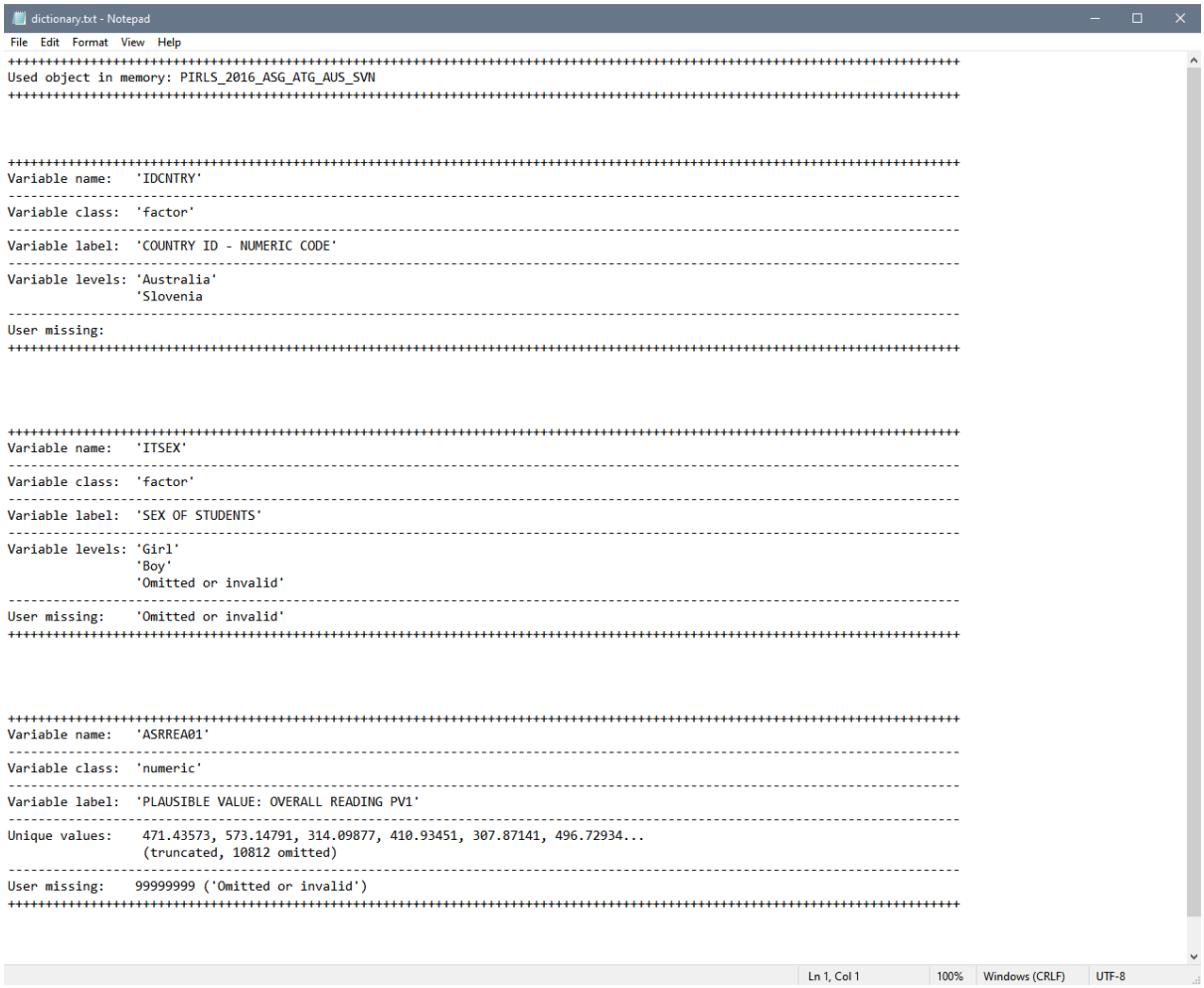
Če morate slovarje shraniti v datoteko, označite potrditveno polje **Save the variable dictionaries in a file** (Shrani slovarje spremenljivk v datoteko). V nasprotnem primeru se bodo

slovarji prikazali samo v konzoli v **GUI** (grafični uporabniški vmesnik), ko pritisnete gumb **Execute syntax** (Izvrši sintakso). Če označite polje, bo vmesnik prikazal še eno potrditveno polje, ki vas vpraša, ali želite, da se datoteka po vseh končanih operacijah samodejno odpre v vašem privzetem urejevalniku besedila. Če je potrditveno polje označeno, bo vmesnik prikazal tudi gumb **Define output file name** (Določi ime output/izhodne datoteke). Kliknite nanj, se pomaknite do mape, kamor želite shraniti datoteko, določite ime datoteke in kliknite **Save** (Shrani) v pogovornem oknu za shranjevanje datoteke. Končne nastavitve bodo videti tako:

Kliknite gumb **Execute syntax.** Sinteza bo izvedena, izhod pa bo prikazan v konzoli, ki se bo pojavila na dnu zaslona.



Ko so vse operacije zaključene, se bo datoteka z besedilnimi slovarji odprla v vašem privzetem urejevalniku besedil:



To datoteko lahko obdržite za nadaljnje sklicevanje pri kodiranju spremenljivk (kar bomo obravnavali v naslednjem razdelku) ali za izvajanje analiz.

3.4 Diagnostične tabele podatkov (angl. *Data diagnostic tables*)

***Opomba**: Ta funkcija je namenjena le kot pripomoček za diagnostične namene, za pregled spremenljivk pred izvedbo dejanske analize. Ni namenjena za dejansko analizo podatkov mednarodnih raziskav. Poročanje statistike iz nje lahko vodi do pristranskih in napačnih zaključkov.*

3.4.1 Uvod

Ko izvajamo analizo, moramo vnaprej poznati lastnosti vseh vključenih spremenljivk. Funkcija **lsa.vars.dict** (obravnavana v prejšnjem poglavju) ustvari slovarje spremenljivk, ki vključujejo informacije o imenih spremenljivk, razredih (numerični, faktorski ali znakovni), oznakah ter njihovih ravneh (tj. kategorije odgovorov v primeru faktorskih spremenljivk) ali edinstvenih vrednostih (v primeru numeričnih ali znakovnih spremenljivk) ter uporabniško določenih manjkajočih vrednostih (če obstajajo). Vendar pa je vedno koristno poznati dejansko vsebino podatkov — kakšna je frekvenca vsake vrednosti v spremenljivki (v primeru kategorijskih spremenljivk) ali kakšna je povprečna vrednost, varianca, standardni odklon itd. (v primeru

zveznih spremenljivk). Funkcija `lsa.data.diag` ustvari frekvenčne (v primeru kategorijskih spremenljivk) in deskriptivne (v primeru zveznih spremenljivk) tabele, ki jih lahko pregledujemo in na podlagi tega sprejmemo odločitve za dejansko analizo ali za ponovno kodiranje spremenljivk. Funkcija izračuna vse statistike po spremenljivkah in postavi tabele v Excelov delovni zvezek, kjer je za vsako spremenljivko ustvarjen ločen list. Dodana je tudi preglednica »Index« za lažjo navigacijo. Vsi rezultati so izračunani po državah. Uporabljene so lahko pretvorjene datoteke .RData mednarodnih raziskav ali datoteke, kjer so združeni podatki iz različnih držav in/ali vrst respondentov. Funkcija je tudi posplošena na podatke, ki niso iz mednarodnih raziskav, in jo lahko uporabimo s katerim koli `data.frame` ali `data.table`.

3.4.2 Funkcija za izdelavo diagnostičnih tabel podatkov in njeni argumenti

Funkcija `lsa.data.diag` ima naslednje argumente:

- `data.file` – datoteka, ki vsebuje objekt `lsa.data`. Določen mora biti bodisi ta argument ali `data.object`, ne pa oba hkrati.
- `data.object` – objekt v pomnilniku, ki vsebuje objekt `lsa.data`. Določen mora biti bodisi ta argument ali `data.file`, ne pa oba hkrati.
- `split.vars` – spremenljivka(e) za razdelitev rezultatov. Če ni podanih razdelitvenih spremenljivk, bodo rezultati izračunani na ravni držav (če se uporabljajo uteži) ali vzorcev (če se ne uporabljajo uteži).
- `variables` – imena spremenljivk, za katere naj se izračunajo statistike. Če so spremenljivke faktorji ali znaki, bodo izračunane frekvence, če so numerične, pa deskriptivi, razen če je `cont.freq` nastavljen na `TRUE`.
- `weight.var` – ime spremenljivke, ki vsebuje uteži, če so potrebne utežene statistike. Če ime utežne spremenljivke ni navedeno, bo funkcija samodejno izbrala privzeto utežno spremenljivko za določen lsa.data, odvisno od vrste respondenta. »none« pomeni, da se uporabljajo neobtežene statistike.
- `cont.freq` – določitev, ali naj se vrednosti numeričnih kategorij obravnavajo kot kategorijske, da se izračunajo frekvence.
- `include.missing` – določitev, ali naj bodo `NA` in uporabniško določene manjkajoče vrednosti (če so na voljo) vključene kot kategorije za razdelitev spremenljivk v `split.vars`. Privzeto je `FALSE`.
- `output.file` – polna pot do output datoteke, vključno z imenom datoteke. Če je ta izpuščen, bo datoteka z privzetim imenom »Analysis.xlsx« zapisana v delovni imenik (`getwd()`).
- `open.output` – logična vrednost; določitev, ali naj se izhod odpre po zapisu. Privzeto (`TRUE`) odpre izhod v privzetem programu za preglednice, nameščenem na računalniku.
- `...` – dodatni argumenti.

Opomba:

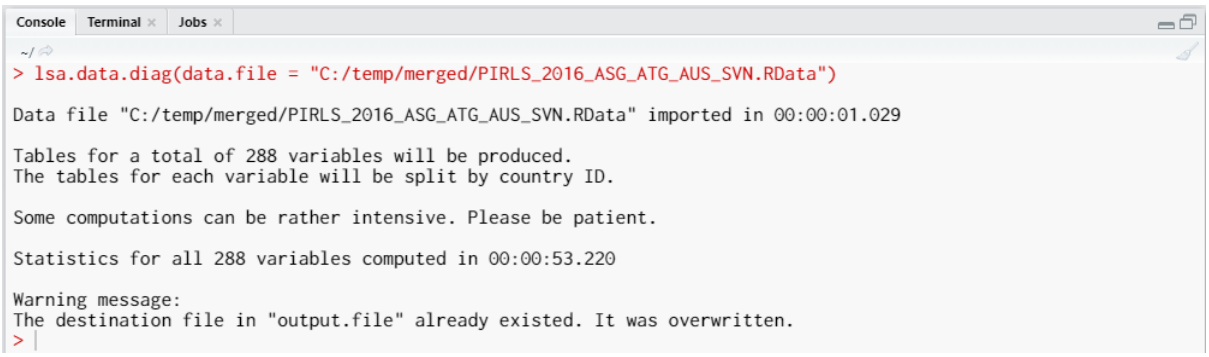
Ta funkcija je namenjena le kot pripomoček za diagnostične namene, da preverimo spremenljivke pred dejansko analizo. Ni namenjena za dejansko analizo podatkov velikih raziskav. Poročanje o statistiki iz nje lahko vodi do pristranskih in napačnih zaključkov.

3.4.3 Izdelava diagnostičnih tabel podatkov z uporabo ukazne vrstice

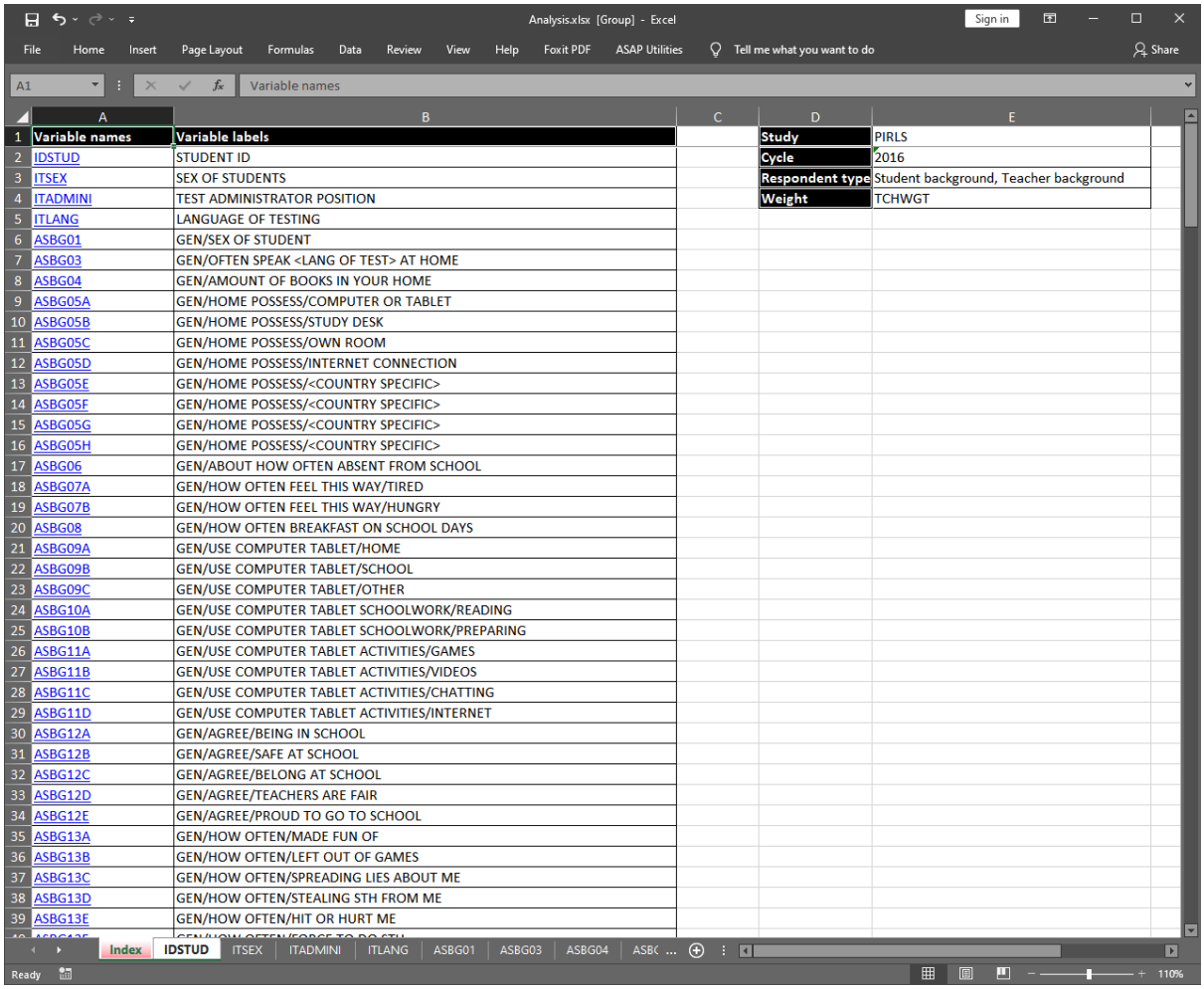
V tem primeru bomo uporabili datoteko, združeno v prejšnjem primeru, z uporabo ukazne vrstice za vse spremenljivke v datoteki (če izpustimo argument `var.names`, bo funkcija ustvarila diagnostične tabele za vse spremenljivke v datoteki). V RStudiou izvedite naslednjo sintakso:

```
lsa.data.diag(data.file =
"C:/temp/merged/PIRLS_2016_ASG_ATG_AUS_SVN.RData")
```

Zgornji klic sintakse bo vrnil naslednji izhod v konzoli:

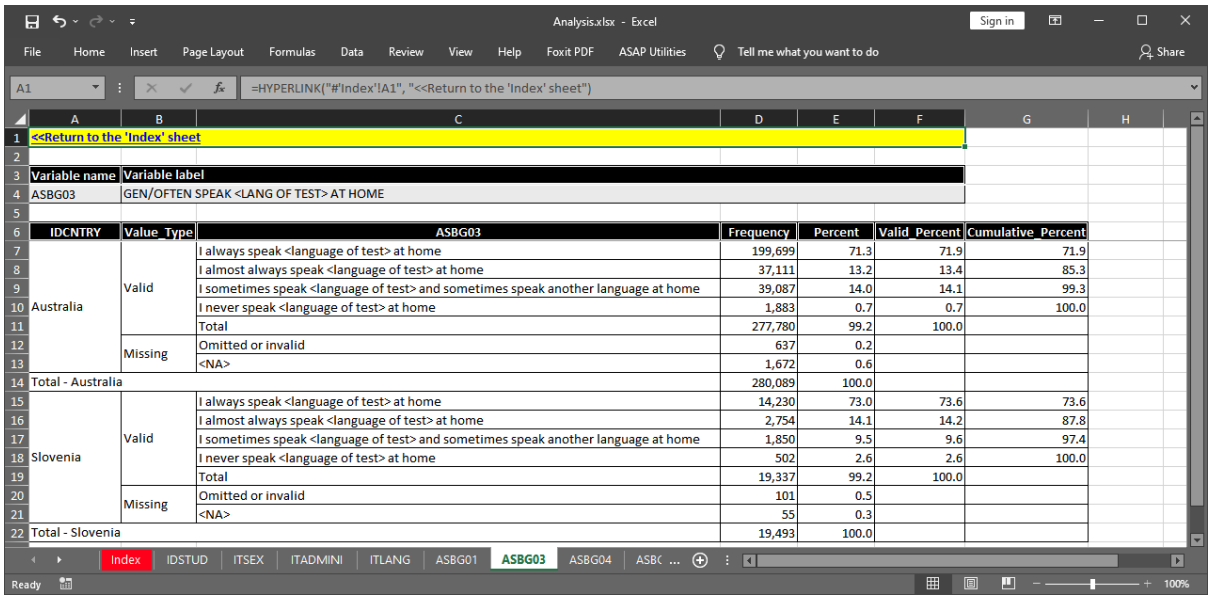


Izvožena datoteka MS Excel z listom »Index« je prikazana na spodnji sliki. List vsebuje informacije o raziskavi, njenem ciklu, vrsti podatkov o respondentih (v tem primeru učencih in učiteljih) in uporabljenih utežeh. V tem primeru utež ni bila izrecno določena, zato je funkcija vzela privzeto utež za to združeno kombinacijo respondentov. Prav tako nismo določili polne poti do imena izhodne datoteke, zato je bila datoteka z imenom »Analysis.xlsx« zapisana v delovni imenik. Prva dva stolpca na listu »Index« vsebujeta imena in oznake spremenljivk. Klik na ime spremenljivke v prvem stolpcu vas bo preusmeril na list, ki vsebuje statistiko za ustrezno spremenljivko.



	A	B	C	D	E
1	**Variable names**	**Variable labels**		**Study**	PIRLS
2	IDSTUD	STUDENT ID		**Cycle**	2016
3	ITSEX	SEX OF STUDENTS		**Respondent type**	Student background, Teacher background
4	ITADMINI	TEST ADMINISTRATOR POSITION		**Weight**	TCHWGT
5	ITLANG	LANGUAGE OF TESTING			
6	ASBG01	GEN/SEX OF STUDENT			
7	ASBG03	GEN/OFTEN SPEAK <LANG OF TEST> AT HOME			
8	ASBG04	GEN/AMOUNT OF BOOKS IN YOUR HOME			
9	ASBG05A	GEN/HOME POSSESS/COMPUTER OR TABLET			
10	ASBG05B	GEN/HOME POSSESS/STUDY DESK			
11	ASBG05C	GEN/HOME POSSESS/OWN ROOM			
12	ASBG05D	GEN/HOME POSSESS/INTERNET CONNECTION			
13	ASBG05E	GEN/HOME POSSESS/<COUNTRY SPECIFIC>			
14	ASBG05F	GEN/HOME POSSESS/<COUNTRY SPECIFIC>			
15	ASBG05G	GEN/HOME POSSESS/<COUNTRY SPECIFIC>			
16	ASBG05H	GEN/HOME POSSESS/<COUNTRY SPECIFIC>			
17	ASBG06	GEN/ABOUT HOW OFTEN ABSENT FROM SCHOOL			
18	ASBG07A	GEN/HOW OFTEN FEEL THIS WAY/TIRED			
19	ASBG07B	GEN/HOW OFTEN FEEL THIS WAY/HUNGRY			
20	ASBG08	GEN/HOW OFTEN BREAKFAST ON SCHOOL DAYS			
21	ASBG09A	GEN/USE COMPUTER TABLET/HOME			
22	ASBG09B	GEN/USE COMPUTER TABLET/SCHOOL			
23	ASBG09C	GEN/USE COMPUTER TABLET/OTHER			
24	ASBG10A	GEN/USE COMPUTER TABLET SCHOOLWORK/READING			
25	ASBG10B	GEN/USE COMPUTER TABLET SCHOOLWORK/PREPARING			
26	ASBG11A	GEN/USE COMPUTER TABLET ACTIVITIES/GAMES			
27	ASBG11B	GEN/USE COMPUTER TABLET ACTIVITIES/VIDEOS			
28	ASBG11C	GEN/USE COMPUTER TABLET ACTIVITIES/CHATTING			
29	ASBG11D	GEN/USE COMPUTER TABLET ACTIVITIES/INTERNET			
30	ASBG12A	GEN/AGREE/BEING IN SCHOOL			
31	ASBG12B	GEN/AGREE/SAFE AT SCHOOL			
32	ASBG12C	GEN/AGREE/BELONG AT SCHOOL			
33	ASBG12D	GEN/AGREE/TEACHERS ARE FAIR			
34	ASBG12E	GEN/AGREE/PROUD TO GO TO SCHOOL			
35	ASBG13A	GEN/HOW OFTEN/MADE FUN OF			
36	ASBG13B	GEN/HOW OFTEN/LEFT OUT OF GAMES			
37	ASBG13C	GEN/HOW OFTEN/SPREADING LIES ABOUT ME			
38	ASBG13D	GEN/HOW OFTEN/STEALING STH FROM ME			
39	ASBG13E	GEN/HOW OFTEN/HIT OR HURT ME			

Kliknimo na povezavo za spremenljivko ASBG03. List za to spremenljivko je prikazan spodaj.



	A	B	C	D	E	F	G	H
1	<<Return to the 'Index' sheet							
2								
3	**Variable name**	**Variable label**						
4	ASBG03	GEN/OFTEN SPEAK <LANG OF TEST> AT HOME						
5								
6	**IDCNTRY**	**Value_Type**	**ASBG03**	**Frequency**	**Percent**	**Valid_Percent**	**Cumulative_Percent**	
7	Australia	Valid	I always speak <language of test> at home	199,699	71.3	71.9	71.9	
8			I almost always speak <language of test> at home	37,111	13.2	13.4	85.3	
9			I sometimes speak <language of test> and sometimes speak another language at home	39,087	14.0	14.1	99.3	
10			I never speak <language of test> at home	1,883	0.7	0.7	100.0	
11			Total	277,780	99.2	100.0		
12		Missing	Omitted or invalid	637	0.2			
13			<NA>	1,672	0.6			
14	Total - Australia			280,089	100.0			
15	Slovenia	Valid	I always speak <language of test> at home	14,230	73.0	73.6	73.6	
16			I almost always speak <language of test> at home	2,754	14.1	14.2	87.8	
17			I sometimes speak <language of test> and sometimes speak another language at home	1,850	9.5	9.6	97.4	
18			I never speak <language of test> at home	502	2.6	2.6	100.0	
19			Total	19,337	99.2	100.0		
20		Missing	Omitted or invalid	101	0.5			
21			<NA>	55	0.3			
22	Total - Slovenia			19,493	100.0			

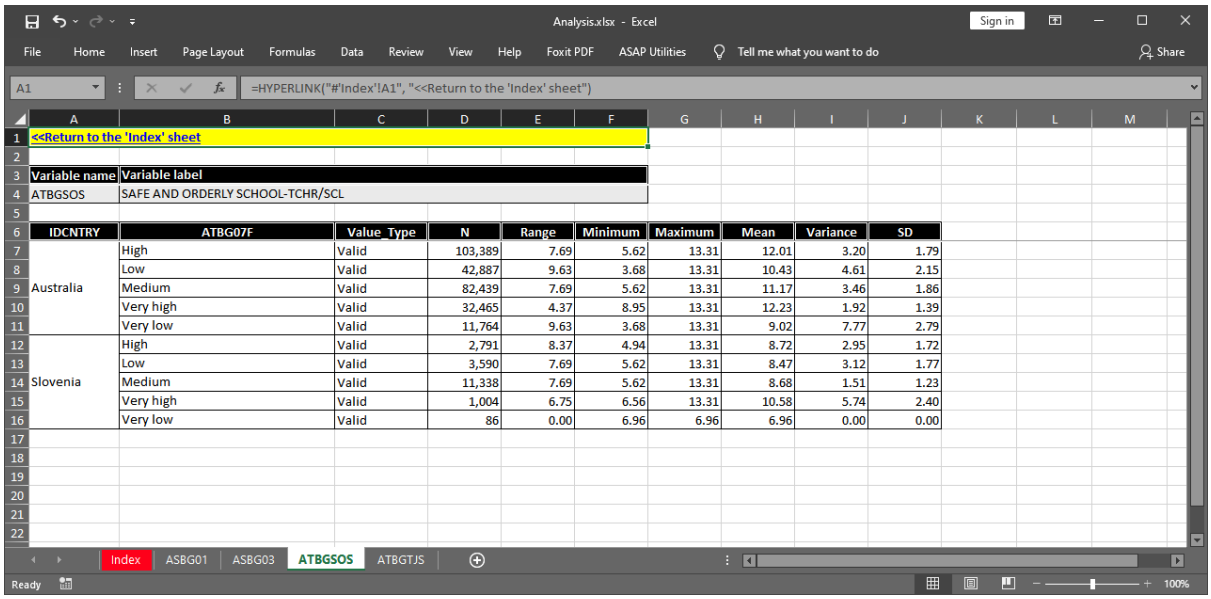
Bodite pozorni na rumeno obarvano celico v zgornjem levem kotu. Vsebuje povezavo, ki omogoča enostavno vrnitev na list »Index«. Prikazana sta tudi ime in oznaka spremenljivke. List prikazuje utežene frekvence, odstotke, veljavne odstotke in kumulativne odstotke. Bodite

pozorni na to, da nismo navedli nobenih spremenljivk za deljenje. Funkcija samodejno deli rezultate po državah, tudi če spremenljivka ID države ni navedena kot spremenljivka za deljenje.

Kot drugi primer vzemimo le nekaj spremenljivk – nekatere kategorijske in nekatere zvezne. Vzemimo spol učenca (ASBG01), pogostost uporabe jezika preizkusa doma (ASBG03), zaznavanje učiteljev o varnem in urejenem šolskem okolju (ATBGSOS) ter lestvico zadovoljstva učiteljev z delom (ATBGTJS). Rezultate bomo tudi razdelili po zaznavanju učiteljev o vključevanju staršev v šolsko življenje (ATBG07F) v vsaki državi.

```
lsa.data.diag(data.file =
"C:/temp/merged/PIRLS_2016_ASG_ATG_AUS_SVN.RData",
              split.vars = "ATBG07F",
              variables = c("ASBG01", "ASBG03", "ATBGSOS",
"ATBGTJS"))
```

Tokrat list »Index« vsebuje vrstice za izbrane spremenljivke. Klik na celico s povezavo ATBGSOS v prvem stolpcu lista »Index« nas pripelje do lista, ki vsebuje statistiko za lestvico učiteljev »Varna in urejena šola« (zvezna spremenljivka):



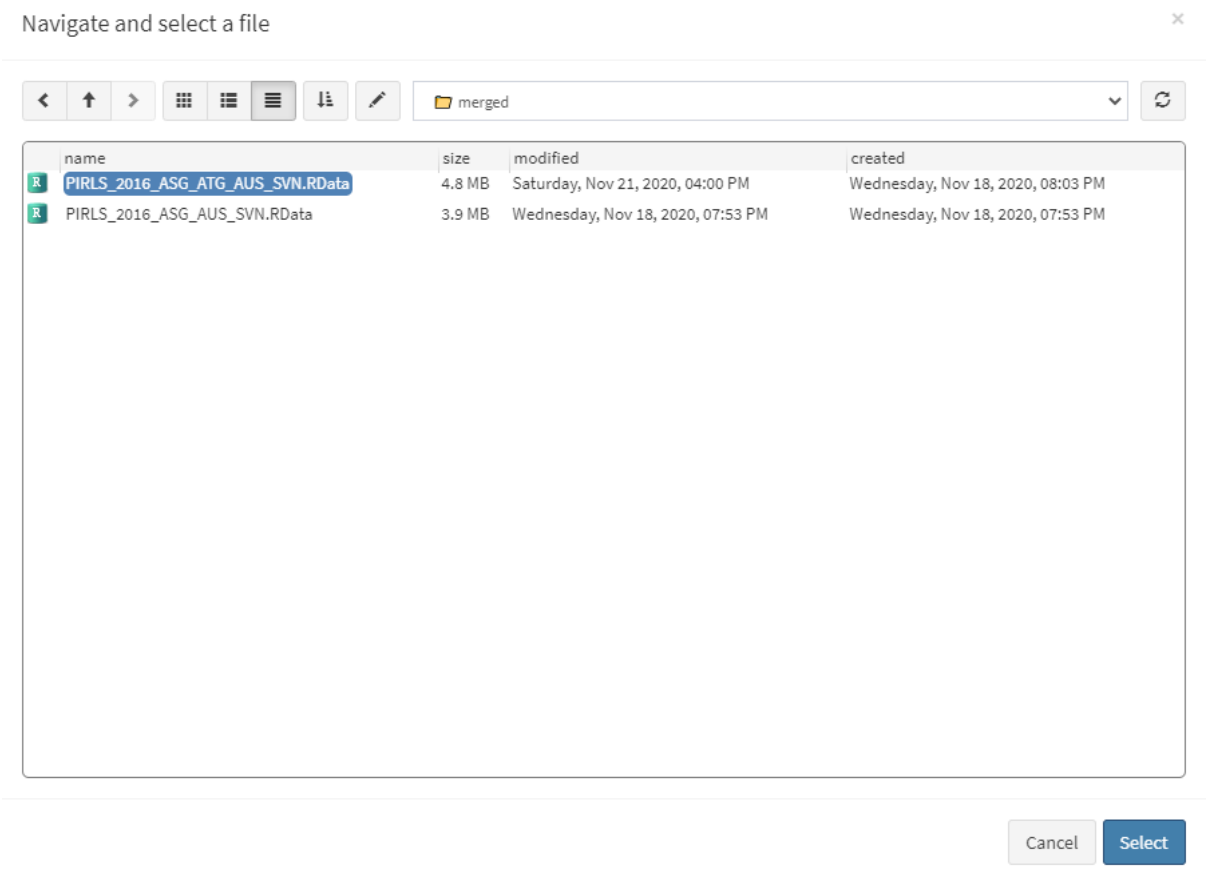
Spremenljivka je zvezna, zato namesto frekvenc in odstotkov tabela prikazuje število uteženih primerov, razpon, minimalne in maksimalne vrednosti, povprečje, varianco in standardni odklon. To je privzeto vedenje funkcije. Lahko ga spremenite z uporabo argumenta `cont.freq`. V tem primeru bo funkcija vse zvezne spremenljivke obravnavala kot kategorijske in izračunala frekvence za vsako vrednost.

3.4.4 Izdelava diagnostičnih tabel podatkov z uporabo GUI (angl. *Producing data diagnostic tables using the GUI*)

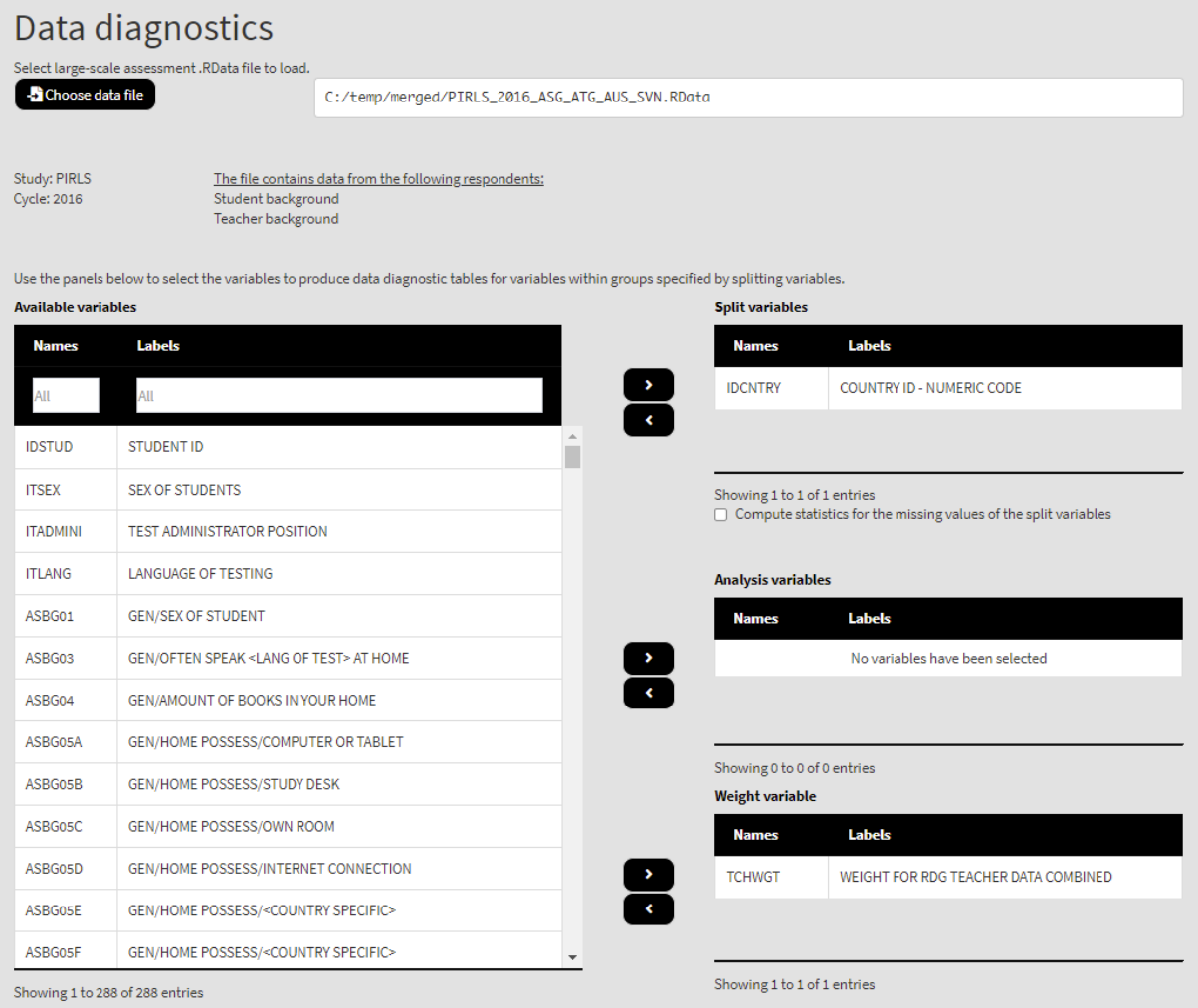
Za začetek uporabniškega vmesnika RALSA izvedite naslednji ukaz v RStudiu:

```
ralsaGUI()
```

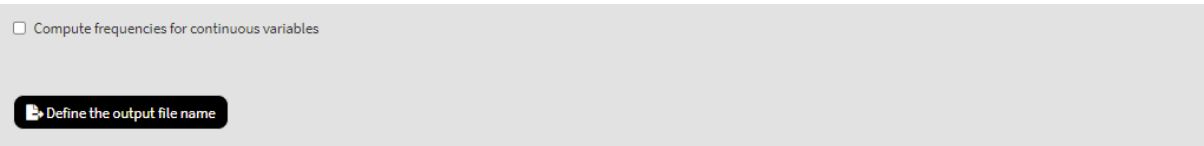
Ko se GUI odpre v vašem brskalniku, izberite **Data preparation** > **Data diagnostics** iz menija na levi strani. Ko se pomaknete do Data diagnostics v GUI, kliknite gumb **Choose data file**. Pojdite v mapo, ki vsebuje združeno datoteko **»PIRLS_2016_ASG_ATG_AUS_SVN.RData«**, izberite jo in kliknite gumb **Select**.



Ko se datoteka naloži, boste videli dva panela z razpoložljivimi spremenljivkami in izbranimi spremenljivkami (slednji je trenutno prazen):

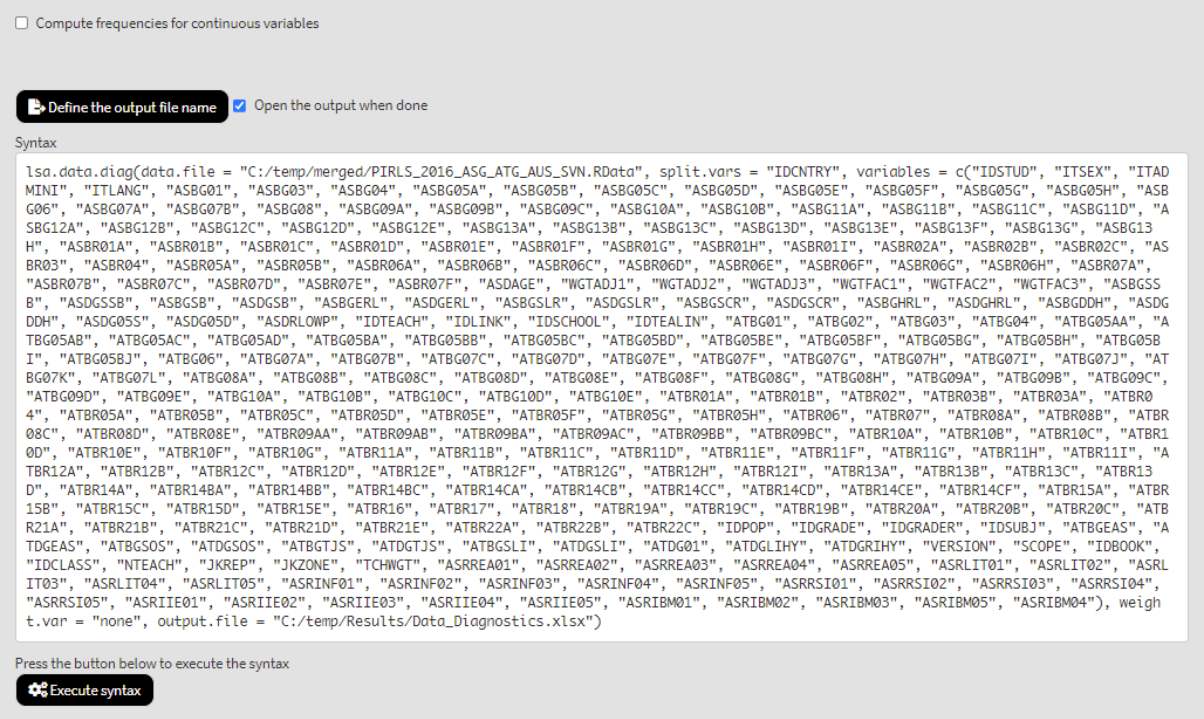


Uporabite miško za izbiro posameznih spremenljivk in enojne puščice za premik s seznama razpoložljivih spremenljivk na seznam spremenljivk za analizo ter obratno. Privzeta utež je izbrana samodejno, vendar jo lahko spremenite z drugo utežjo ali pa izbrišete (statistika bo brez uteži). Uporabite enojne puščice za izbiro vseh ali nobenih spremenljivk. Filter polja na vrhu panelov lahko uporabite za hitro iskanje potrebnih spremenljivk. Izberimo vse razpoložljive spremenljivke v podatkovni datoteki in jih premaknimo na seznam spremenljivk za analizo. Pustili bomo IDCNTRY (spremenljivka ID države) kot edino spremenljivko za delitev (privzeto), tako da bodo vsi rezultati predstavljeni po državah. ID države ni mogoče odstraniti s seznama spremenljivk za delitev. Ko bodo na panelu izbranih spremenljivk prisotne spremenljivke, se bodo pojavili naslednji elementi:

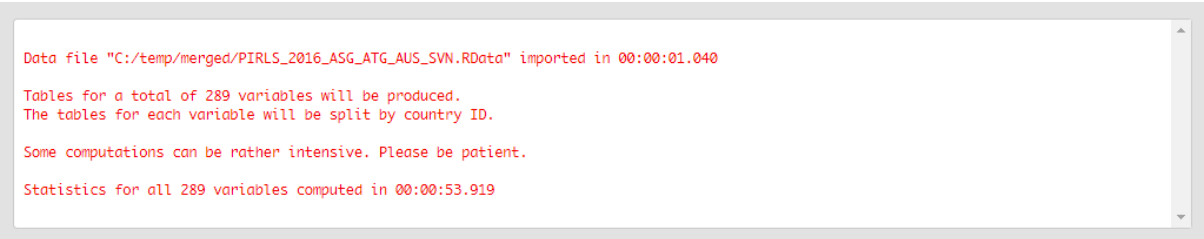


Če je potrditveno polje označeno, bo RALSA obveščena, da želimo zvezne spremenljivke obravnavati kot kategorialne in izračunati frekvence, odstotke, veljavne odstotke in kumulativne odstotke za vsako vrednost v zveznih spremenljivkah namesto njihovega skupnega števila primerov, obsegov, najmanjših in največjih vrednosti, povprečij, varianc in

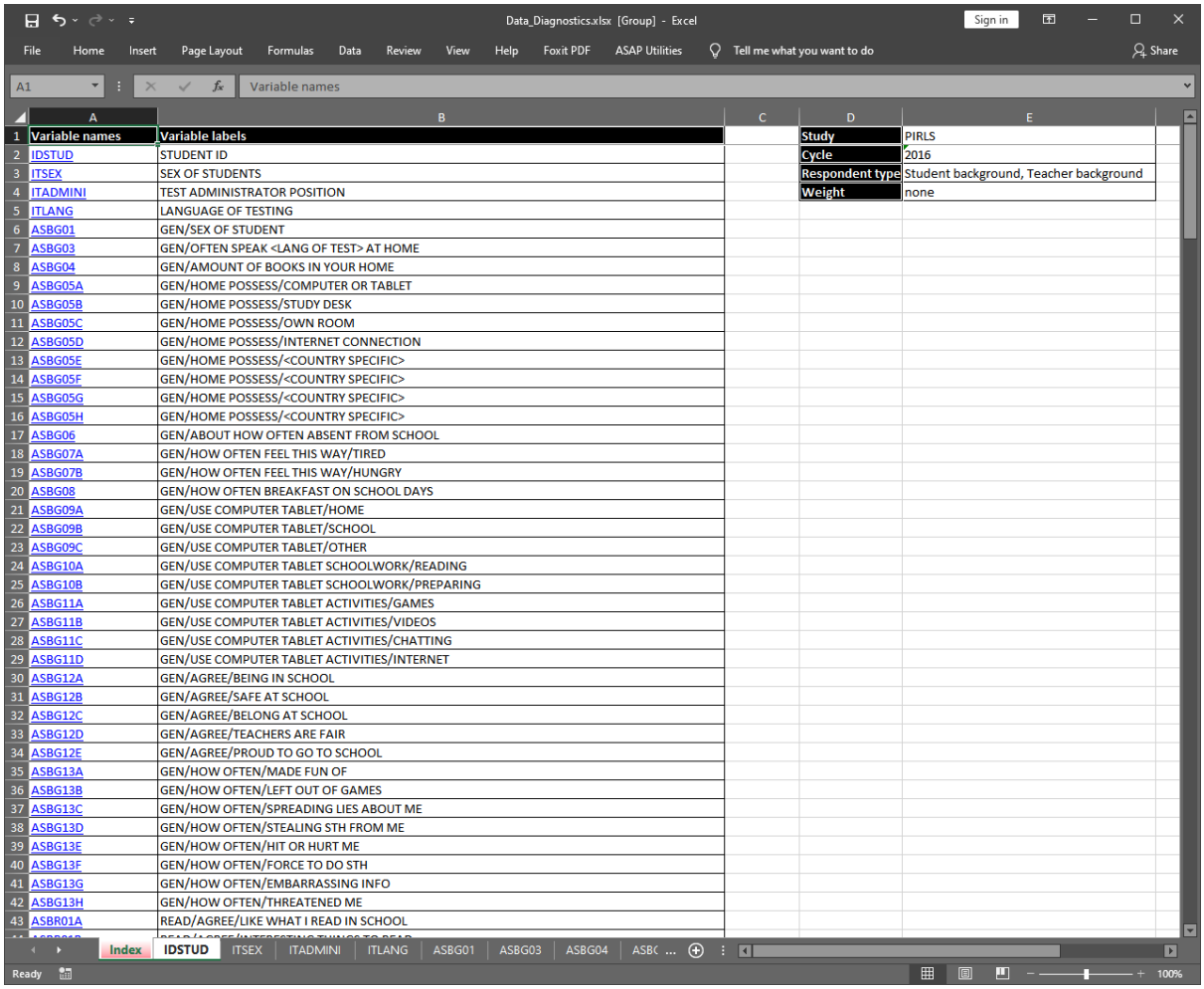
standardnih odklonov. Polje bomo pustili neoznačeno. Pritisnite gumb za določitev imena output datoteke. Pojdite v mapo **»C:/temp/Results«** (ali v mapo, kjer želite shraniti output/izhod) in določite ime output datoteke. Ko to storite, se bo ob gumbu za določitev imena output datoteke pojavilo potrditveno polje. Če je označeno, se bo izhodna datoteka odprla po zaključku vseh izračunov. Pod tem se bo prikazala sintaksa. Pod vsemi temi se bo prikazal gumb za izvajanje sintakse. Končne nastavitve v spodnjem delu zaslona bi morale izgledati takole:



Kliknite gumb za izvajanje sintakse. GUI konzola se bo pojavila na dnu in beležila vse zaključene operacije:



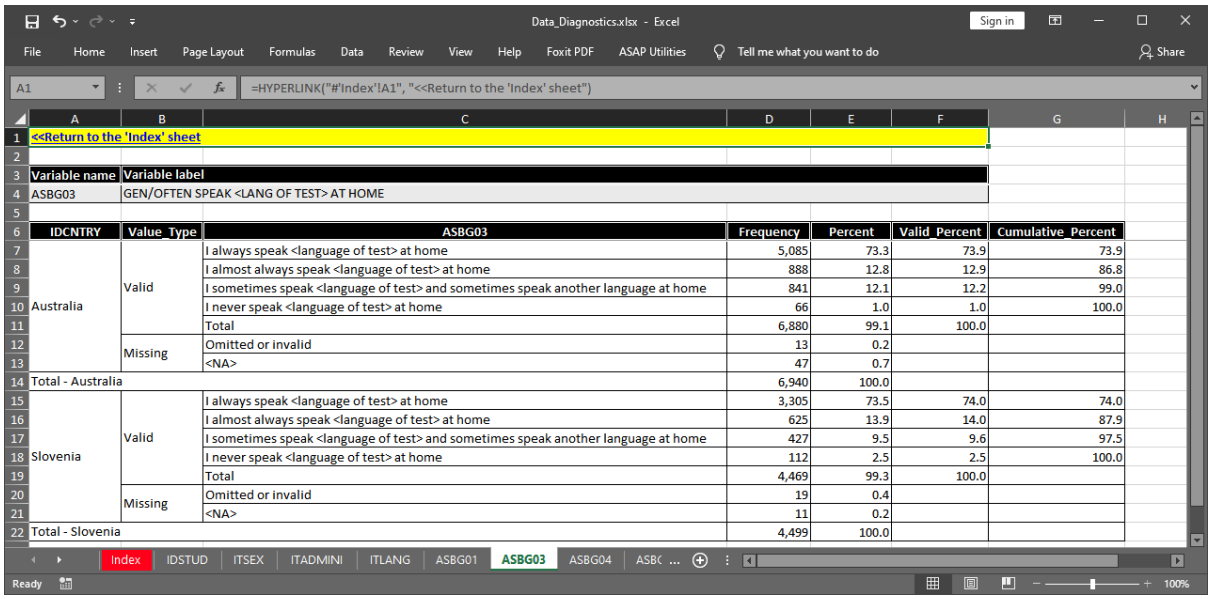
Če je potrditveno polje **Open the output when done** (Odpri output, ko je končano) označeno, se bo output samodejno odprl v privzetem programu za preglednice (običajno MS Excel) po tem, ko bodo vsi izračuni zaključeni. Izvožena datoteka MS Excel s preglednico **Index** (Kazalo) je prikazana na spodnji sliki. Preglednica vsebuje informacije o raziskavi, njenem ciklu, vrsti(-ah) podatkov anketirancev (v tem primeru učenci in učitelji) in uporabljenih utežeh. V tem primeru utež ni bila izrecno določena, zato je funkcija uporabila privzeto utež za to združeno kombinacijo anketirancev. Prva dva stolpca v preglednici **Index** vsebujeta imena in oznake spremenljivk. Klik na ime spremenljivke v prvem stolpcu preklopi na list s statistiko za ustrezno spremenljivko.



Variable names	Variable labels
IDSTUD	STUDENT ID
ITSEX	SEX OF STUDENTS
ITADMINI	TEST ADMINISTRATOR POSITION
ITLANG	LANGUAGE OF TESTING
ASBG01	GEN/SEX OF STUDENT
ASBG03	GEN/OFTEN SPEAK <LANG OF TEST> AT HOME
ASBG04	GEN/AMOUNT OF BOOKS IN YOUR HOME
ASBG05A	GEN/HOME POSSESS/COMPUTER OR TABLET
ASBG05B	GEN/HOME POSSESS/STUDY DESK
ASBG05C	GEN/HOME POSSESS/OWN ROOM
ASBG05D	GEN/HOME POSSESS/INTERNET CONNECTION
ASBG05E	GEN/HOME POSSESS/<COUNTRY SPECIFIC>
ASBG05F	GEN/HOME POSSESS/<COUNTRY SPECIFIC>
ASBG05G	GEN/HOME POSSESS/<COUNTRY SPECIFIC>
ASBG05H	GEN/HOME POSSESS/<COUNTRY SPECIFIC>
ASBG06	GEN/ABOUT HOW OFTEN ABSENT FROM SCHOOL
ASBG07A	GEN/HOW OFTEN FEEL THIS WAY/TIRED
ASBG07B	GEN/HOW OFTEN FEEL THIS WAY/HUNGRY
ASBG08	GEN/HOW OFTEN BREAKFAST ON SCHOOL DAYS
ASBG09A	GEN/USE COMPUTER TABLET/HOME
ASBG09B	GEN/USE COMPUTER TABLET/SCHOOL
ASBG09C	GEN/USE COMPUTER TABLET/OTHER
ASBG10A	GEN/USE COMPUTER TABLET SCHOOLWORK/READING
ASBG10B	GEN/USE COMPUTER TABLET SCHOOLWORK/PREPARING
ASBG11A	GEN/USE COMPUTER TABLET ACTIVITIES/GAMES
ASBG11B	GEN/USE COMPUTER TABLET ACTIVITIES/VIDEOS
ASBG11C	GEN/USE COMPUTER TABLET ACTIVITIES/CHATTING
ASBG11D	GEN/USE COMPUTER TABLET ACTIVITIES/INTERNET
ASBG12A	GEN/AGREE/BEING IN SCHOOL
ASBG12B	GEN/AGREE/SAFE AT SCHOOL
ASBG12C	GEN/AGREE/BELONG AT SCHOOL
ASBG12D	GEN/AGREE/TEACHERS ARE FAIR
ASBG12E	GEN/AGREE/PROUD TO GO TO SCHOOL
ASBG13A	GEN/HOW OFTEN/MADE FUN OF
ASBG13B	GEN/HOW OFTEN/LEFT OUT OF GAMES
ASBG13C	GEN/HOW OFTEN/SPREADING LIES ABOUT ME
ASBG13D	GEN/HOW OFTEN/STEALING STH FROM ME
ASBG13E	GEN/HOW OFTEN/HIT OR HURT ME
ASBG13F	GEN/HOW OFTEN/FORCE TO DO STH
ASBG13G	GEN/HOW OFTEN/EMBARRASSING INFO
ASBG13H	GEN/HOW OFTEN/THREATENED ME
ASBR01A	READ/AGREE/LIKE WHAT I READ IN SCHOOL

Study	PIRLS
Cycle	2016
Respondent type	Student background, Teacher background
Weight	none

Kliknimo na povezavo za spremenljivko ASBG03. List za to spremenljivko je prikazan spodaj.



<<Return to the 'Index' sheet

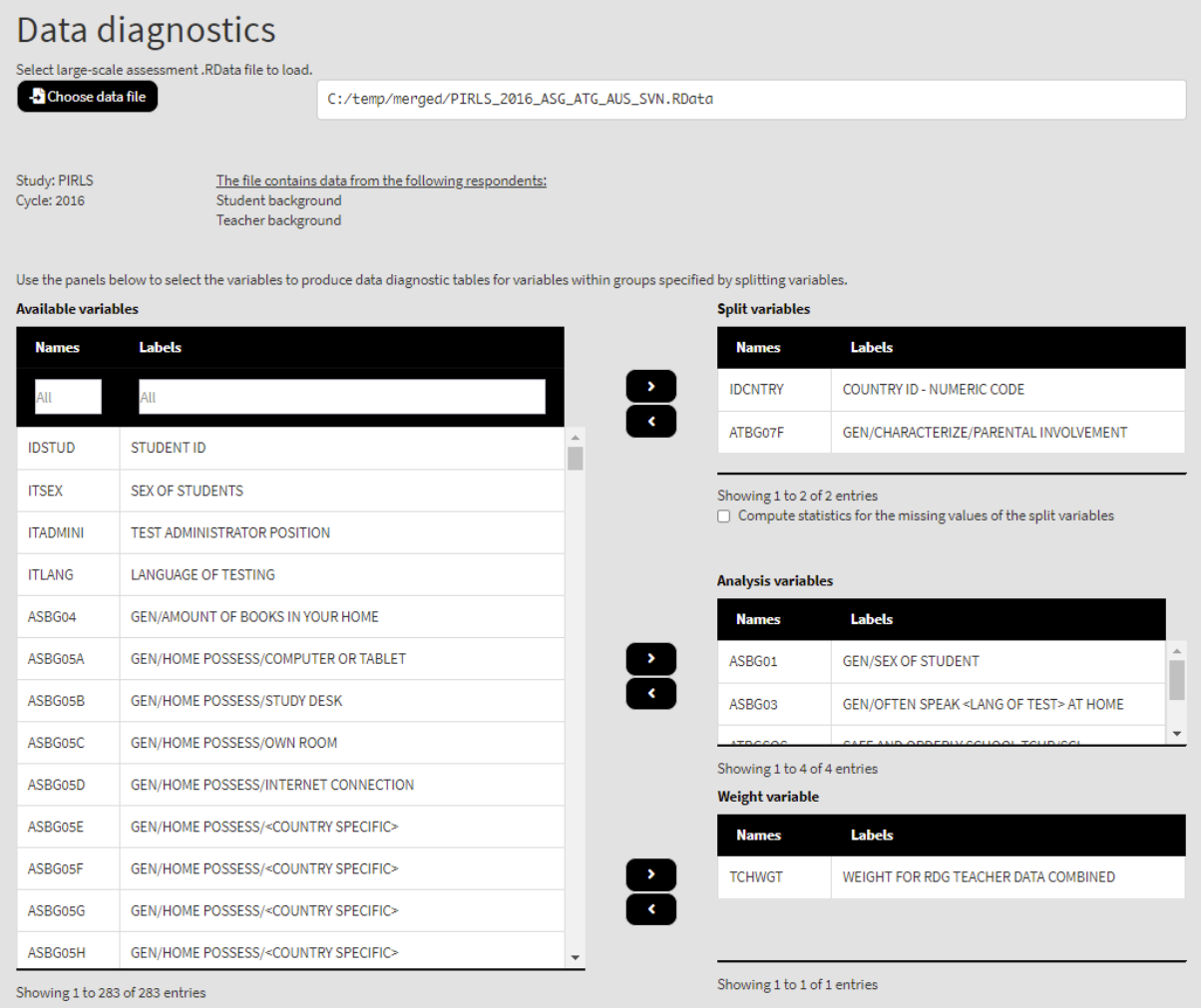
Variable name	Variable label
ASBG03	GEN/OFTEN SPEAK <LANG OF TEST> AT HOME

IDCNTRY	Value_Type	ASBG03	Frequency	Percent	Valid_Percent	Cumulative_Percent
Australia	Valid	I always speak <language of test> at home	5,085	73.3	73.9	73.9
		I almost always speak <language of test> at home	888	12.8	12.9	86.8
		I sometimes speak <language of test> and sometimes speak another language at home	841	12.1	12.2	99.0
		I never speak <language of test> at home	66	1.0	1.0	100.0
		Total	6,880	99.1	100.0	
	Missing	Omitted or invalid	13	0.2		
		<NA>	47	0.7		
Total - Australia			6,940	100.0		
Slovenia	Valid	I always speak <language of test> at home	3,305	73.5	74.0	74.0
		I almost always speak <language of test> at home	625	13.9	14.0	87.9
		I sometimes speak <language of test> and sometimes speak another language at home	427	9.5	9.6	97.5
		I never speak <language of test> at home	112	2.5	2.5	100.0
		Total	4,469	99.3	100.0	
	Missing	Omitted or invalid	19	0.4		
		<NA>	11	0.2		
Total - Slovenia			4,499	100.0		

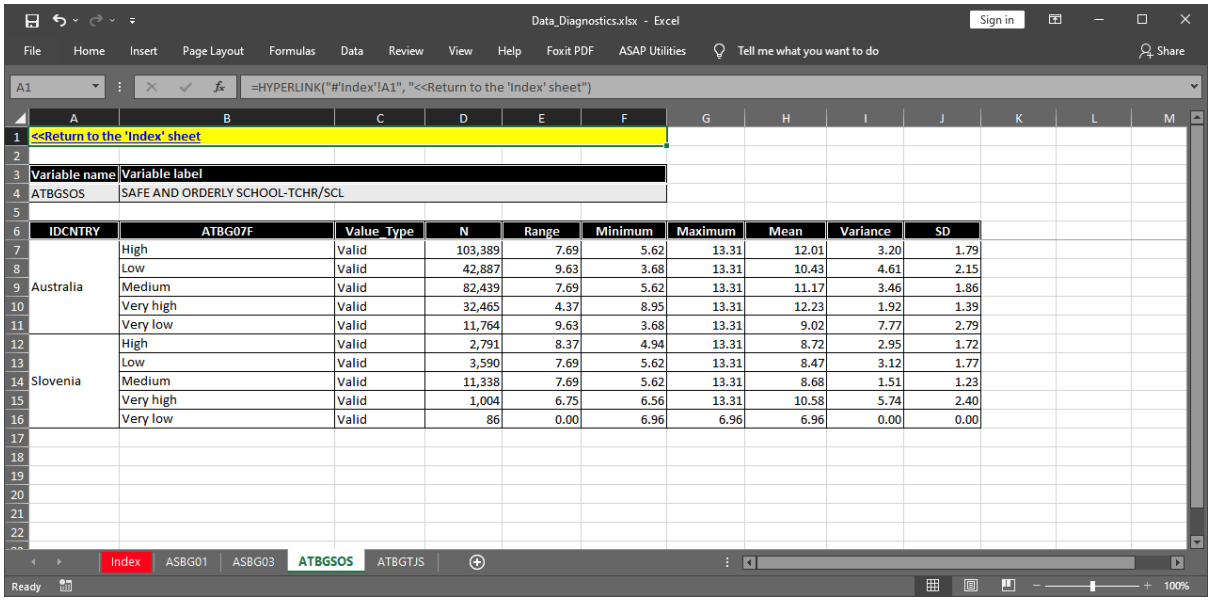
Poglejte rumeno obarvano celico v zgornjem levem kotu. Vsebuje povezavo, ki omogoča enostavno vrnitev na list **Index**. Prikazana sta tudi ime in oznaka za spremenljivko. List prikazuje utežene frekvence, odstotke, veljavne odstotke in kumulativne odstotke. Opozoriti je

treba, da nismo določili nobenih razdelitvenih spremenljivk. Vsi rezultati so predstavljeni po državah.

Kot drugi primer vzemimo le nekaj spremenljivk – nekatere kategorialne in nekatere zvezne. Vzemimo spol učencev (ASBG01), pogostost govorjenja jezika testa doma (ASBG03), lestvico učiteljskih zaznav o varnem in urejenem šolskem okolju (ATBGSOS) ter lestvico zadovoljstva učiteljev z delom (ATBGTJS). Rezultate bomo prav tako razdelili po zaznavah učiteljev o sodelovanju staršev v šolskem življenju (ATBG07F) v vsaki državi (izbrane spremenljivke lahko iz **Analysis variables** prestavite eno po eno ali osvežite brskalnik, kar bo ponastavilo celoten GUI in nato znova naložilo datoteko z rezultati). Izberite spremenljivko ATBG07F s seznama **Available variables** in jo dodajte na seznam **Split variables**. Nato izberite spremenljivke ASBG01, ASBG03, ATBGSOS in ATBGTJS s seznama Available variables (uporabite lahko iskalna polja na vrhu) in jih dodajte v **Analysis variables**. Končne nastavitve bi morale izgledati takole:



Določite ime izhodne datoteke in zaženite sintakso. Po zaključku vseh izračunov se bo Excelov delovni zvezek z rezultati samodejno odprl. Tokrat list **Index** vsebuje vrstice za izbrane spremenljivke. Ko kliknete na celico s povezavo ATBGSOS v prvem stolpcu lista **Index**, vas bo to popeljalo na list, ki vsebuje statistiko za lestvico učiteljskih zaznav o varnem in urejenem šolskem okolju (zvezna spremenljivka):



Data_Diagnostics.xlsx - Excel

=HYPERLINK("#'Index'!A1", "<<Return to the 'Index' sheet")

<<Return to the 'Index' sheet

Variable name	Variable label
ATBGSOS	SAFE AND ORDERLY SCHOOL-TCHR/SCL

IDCNTRY	ATBG07F	Value_Type	N	Range	Minimum	Maximum	Mean	Variance	SD
Australia	High	Valid	103,389	7.69	5.62	13.31	12.01	3.20	1.79
	Low	Valid	42,887	9.63	3.68	13.31	10.43	4.61	2.15
	Medium	Valid	82,439	7.69	5.62	13.31	11.17	3.46	1.86
	Very high	Valid	32,465	4.37	8.95	13.31	12.23	1.92	1.39
	Very low	Valid	11,764	9.63	3.68	13.31	9.02	7.77	2.79
Slovenia	High	Valid	2,791	8.37	4.94	13.31	8.72	2.95	1.72
	Low	Valid	3,590	7.69	5.62	13.31	8.47	3.12	1.77
	Medium	Valid	11,338	7.69	5.62	13.31	8.68	1.51	1.23
	Very high	Valid	1,004	6.75	6.56	13.31	10.58	5.74	2.40
	Very low	Valid	86	0.00	6.96	6.96	6.96	0.00	0.00

Index | ASBG01 | ASBG03 | ATBGSOS | ATBGTJS

Ready

Spremenljivka je **zvezna**, zato namesto frekvenc in odstotkov tabela prikazuje število primerov z utežmi, razpon, minimalne in maksimalne vrednosti, povprečje, varianco in standardni odklon. To je privzeto delovanje funkcije. To se lahko spremeni tako, da obkljukate polje **Izračunaj frekvence za zvezne spremenljivke**. V tem primeru bo funkcija vse zvezne spremenljivke obravnavala kot kategorijske in bo izračunala frekvence za vsako vrednost.

3.5 Rekodiranje spremenljivk (angl. *Recode Variables*)

3.5.1 Uvod

Pri izvajanju analiz pogosto naletimo na potrebo po rekodiranju določenih spremenljivk. Takšna rekodiranja vključujejo obračanje vrednosti, združevanje več kategorij v eno ali celo nastavitev določenih vrednosti kot manjkajočih. Analiza, ki to najpogosteje zahteva, je verjetno **binarna logistična regresija**, saj ta vrsta analize zahteva dihotomno (tj. binarno) odvisno spremenljivko, česar pa v podatkovnih zbirkah pogosto ni veliko. Zato je ta funkcionalnost zelo uporabna. Rekodiranje se lahko izvede za posamezno spremenljivko ali več spremenljivk z enako strukturo hkrati. Funkcionalnost rekodiranja lahko prav tako zelo fleksibilno obravnava uporabniško določene manjkajoče vrednosti. Med rekodiranjem se pojavi veliko opozoril, ki preprečujejo napake. Priporočamo, da vedno rekodirate spremenljivke v nove, namesto da prepišete obstoječe. Tako bodo izvirne spremenljivke vedno na voljo za morebitne druge analize.

Pred rekodiranjem spremenljivk je smiselno preveriti njihove lastnosti z ogledom njihovih slovarjev. To je lahko zelo koristno, saj pomaga razumeti, kaj spremenljivke predstavljajo, in tako preprečiti napake.

3.5.2 Funkcija za rekodiranje spremenljivk in njeni argumenti

Funkcija `lsa.recode.vars` ima naslednje argumente:

- `data.file` – polna pot do .Rdata-datoteke, ki vsebuje objekt lsa.data. Lahko določite to ali `data.object`, ne pa obeh hkrati.
- `data.object` – objekt v spominu, ki vsebuje objekt lsa.data. Lahko določite to ali `data.file`, ne pa obeh hkrati.
- `src.variables` – imena izvornih spremenljivk istega razreda, katerih vrednosti bodo rekodirane.
- `new.variables` – neobvezno, vektor imen novih spremenljivk, kamor bodo shranjene rekodirane vrednosti, enake dolžine kot src.variables. Če ni navedeno, se bodo `src.variables` prepisale.
- `old.new` – niz z navodili za rekodiranje, ki ustreza dolžini ravni faktorjev (ali unikatnih vrednosti v primeru numeričnih ali znakovnih spremenljivk).
- `new.labels` – novi označevalci, če so spremenljivke src.variables tipa faktor, ali označevalci, ki bodo dodeljeni rekodiranim vrednostim (tj. pretvorba numeričnih ali znakovnih spremenljivk v faktorje) z enako dolžino kot nove želene vrednosti.
- `missings.attr` – neobvezno, seznam znakovnih vektorjev za dodelitev uporabniško določenih manjkajočih vrednosti za vsako rekodirano spremenljivko.
- `variable.labels` – neobvezno, niz oznak spremenljivk, ki jih bo dodeljenih novim spremenljivkam.
- `out.file` – polna pot do .Rdata-datoteke, kamor bodo podatki zapisani. Če ni določeno, bo objekt zapisan v spomin.

Opombe:

- Podatki morajo biti razreda lsa.data, kar pomeni, da so bili pretvorjeni iz SPSS (ali TXT v primeru PISA pred ciklom 2015) s funkcijo `lsa.convert.data`.
- Argumenta `data.file` in `data.object` se izključujeta, torej je treba podati enega izmed njiju, ne pa obeh.
- Če polna pot do `out.file` ni navedena, bodo podatki ostali v spominu kot objekt velikih raziskav (angl. *large-scale assessment*). To je uporabno, če želite narediti dodatne spremembe, npr. brisanje stolpcev.
- Imena spremenljivk v `src.variables` morajo biti istega razreda in strukture, torej morajo imeti enako število ravni in oznak v primeru faktorjev ali enake unikatne vrednosti v primeru numeričnih ali znakovnih spremenljivk. Če se razredi razlikujejo, bo funkcija sporočala napako.

3.5.3 Rekodiranje spremenljivk z ukazno vrstico

V naslednjih primerih bomo združili novo datoteko s podatki učencev in podatki ravnateljev, iz PIRLS 2016 (Avstralija in Slovenija), pri čemer bomo vzeli vse spremenljivke iz obeh vrst datotek:

```
lsa.merge.data(inp.folder = "C:/temp",
               file.types = list(acg = NULL, asg = NULL),
               ISO = c("aus", "svn"),
```

```
                out.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData")
```

Za začetek obrnimo odgovore učencev na vprašanje, kako pogosto se v šoli počutijo utrujeni. Uporabili bomo spremenljivko ASBG07A. Dobro bi bilo, da najprej preverimo slovar za to spremenljivko. To lahko naredimo z naslednjim ukazom v RStudiu:

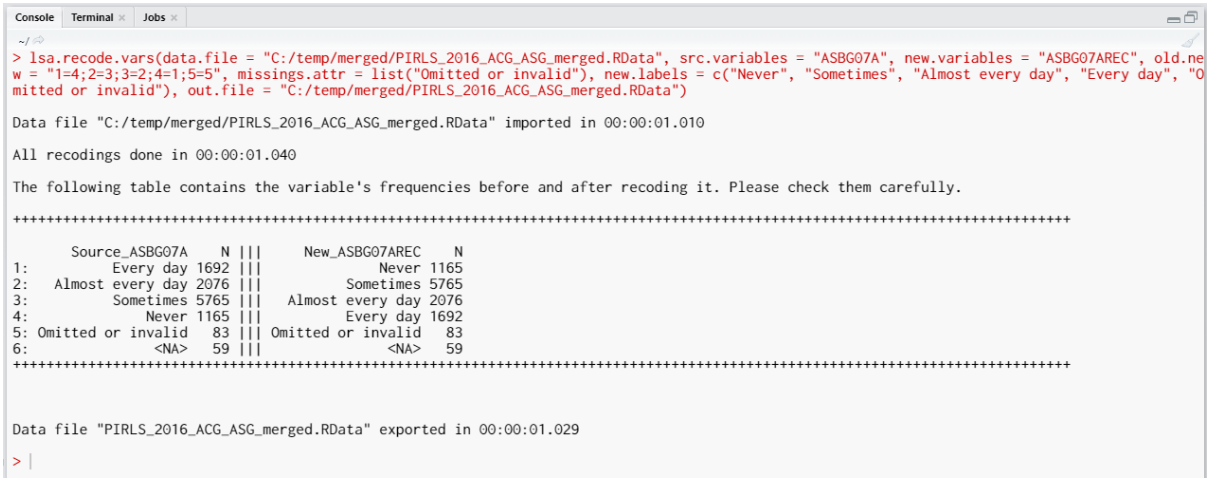
```
lsa.vars.dict(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_AUS_SVN.RData",
              var.names = "ASBG07A")
```

Izvedba zgornje sintakse bo v konzoli vrnila naslednji izpis:

Spremenljivka je faktor (tj. kategorična spremenljivka) z veljavnimi vrednostmi: Vsak dan, Skoraj vsak dan, Včasih in Nikoli. Obstaja še ena vrednost, Izpuščeno ali neveljavno, ki je uporabniško določena manjkajoča vrednost, kot je razvidno iz zgornjega izpisa. Rekodirajmo spremenljivko tako, da se vrstni red odgovorov obrne, in sicer v naslednjem vrstnem redu: Nikoli, Včasih, Skoraj vsak dan in Vsak dan. Uporabniško določeno manjkajočo vrednost Izpuščeno ali neveljavno bomo ohranili nespremenjeno, tako da bo ostala kot zadnja kategorija:

```
lsa.recode.vars(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
                src.variables = "ASBG07A", new.variables =
"ASBG07AREC",
                old.new = "1=4;2=3;3=2;4=1;5=5",
                missings.attr = list("Omitted or invalid"),
                new.labels = c("Nikoli", "Včasih", "Skoraj vsak
dan",
                "Vsak dan", "Izpuščeno ali neveljavno"),
                out.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData")
```

Ko izvedemo zgornjo sintakso, bodo rekodiranja izvedena in shranjena v novi spremenljivki (ASBG07AREC), manjkajoče vrednosti bodo dodeljene kot atribut nove spremenljivke, datoteka pa bo shranjena, s čimer se bo prepisala izvirna datoteka. V konzoli bo vrnjen naslednji izpis, ki bo potrdil uspešno rekodiranje spremenljivke.



Ko v konzoli zaženete zgornjo sintakso, bo tabela prikazala informacije o izvedenem rekodiranju. Na levi strani (Source) bodo prikazane originalne kategorije in njihove frekvence za spremenljivko, ki smo jo rekodirali. Na desni strani (New) bodo prikazane nove kategorije rekodirane spremenljivke in njihove frekvence. Preverite, ali se oznake in frekvence na obeh straneh ujemajo, saj je to ključno za pravilnost analiz, ki vključujejo to spremenljivko.

Pomembne točke:
1. Pot datoteke: Sintaksa bere datoteko podatkov, pri čemer je treba navesti celotno pot do datoteke .RData. Namesto tega lahko uporabite tudi objekt `lsa.data`, ki je že naložen v spomin.

2. Rekodiranje v novo spremenljivko: Originalna spremenljivka (`src.variables = "ASBG07A"`) je rekodirana v novo spremenljivko (`new.variables = "ASBG07AREC"`). Argument `new.variables` bi lahko izpustili, vendar bi to prepisalo originalno spremenljivko, kar ni priporočljivo, saj jo boste morda potrebovali kasneje.

3. Stara in nova vrednost: Stare in nove vrednosti se podajajo z argumentom `old.new`. Vrednosti ločite z enačajem, kjer so na levi stare, na desni pa nove vrednosti. Vsak par starih in novih vrednosti ločite s podpičjem.

4. Preskok vrednosti: Če izpustite katero koli vrednost v `old.new`, se bo ta pretvorila v `NA`. Bodite previdni, če to ni vaš namen.

5. Število novih oznak: Število novih oznak (`new.labels`) za faktor mora biti enako številu parov v `old.new`.

6. Manjkajoče vrednosti: Če v argument `missings.attr` ne vključite oznake »Izpuščeno ali neveljavno«, bo ta postala veljavna vrednost v rekodirani spremenljivki. Če jo izpustite iz argumenta `old.new`, bo rekodirana kot `NA`.

7. Shranjevanje datoteke: Če izpustite argument `out.file`, podatki ne bodo shranjeni na trdi disk. Namesto tega bodo ostali kot objekt `lsa.data` v pomnilniku, kar je lahko koristno, če želite z njimi nadalje delati. Priporočljivo pa je, da svoje delo shranite.

Primer združevanja kategorij za več spremenljivk:
Če želite rekodirati več spremenljivk hkrati, morajo biti vse istega razreda (faktor, numerična ali znakovna spremenljivka), imeti enake stopnje/unikatne vrednosti in uporabniško določene manjkajoče vrednosti (če obstajajo). Tukaj je primer, kako združiti kategorije več spremenljivk hkrati:

```
lsa.recode.vars(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
                src.variables = c("ASBG11A", "ASBG11B", "ASBG11C",
"ASBG11D"),
                new.variables = c("ASBG11AREC", "ASBG11BREC",
"ASBG11CREC", "ASBG11DREC"),
                old.new = "1=1;2=1;3=1;4=2;5=2;6=3",
                new.labels = c("Up to one hour", "More than one
hour", "Omitted or invalid"),
                variable.labels = c("Recoded GEN/USE COMPUTER TABLET
ACTIVITIES/GAMES",
                "Recoded GEN/USE COMPUTER TABLET ACTIVITIES/VIDEOS",
                "Recoded GEN/USE COMPUTER TABLET
ACTIVITIES/CHATTING",
                "Recoded GEN/USE COMPUTER TABLET
ACTIVITIES/INTERNET"),
                missings.attr = list("Omitted or invalid"),
                out.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData")
```

Pomembne opombe za združevanje:

- Število novih spremenljivk mora biti enako številu starih.
- Nove vrednosti v `old.new` se morajo začeti z 1 in biti zaporedne, saj R ne podpira arbitrarnih števil za faktorje.
- Število oznak v `new.labels` mora biti enako številu novih stopenj.
- Novim spremenljivkam lahko dodelite nove oznake z argumentom `variable.labels`. Število oznak mora biti enako številu starih in novih spremenljivk.

Po uspešni izvedbi bo v konzoli prikazan diagnostični izpis, ki bo prikazal uspešnost rekodiranja, kar je pomembno za pravilno nadaljevanje analiz.

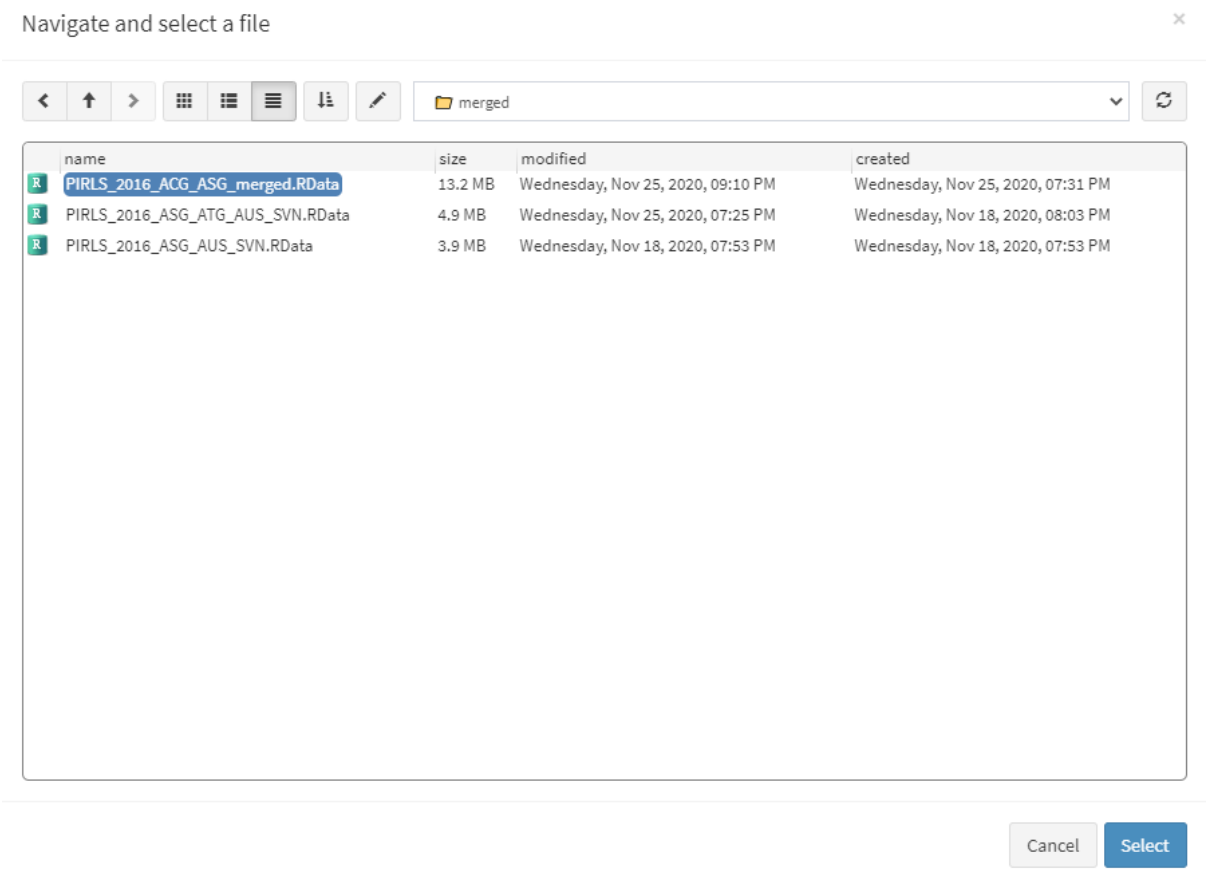
3.5.4 Rekodiranje spremenljivk z uporabo GUI (grafičnega uporabniškega vmesnika)

Za zagon uporabniškega vmesnika RALSA izvedite naslednji ukaz v RStudiu:

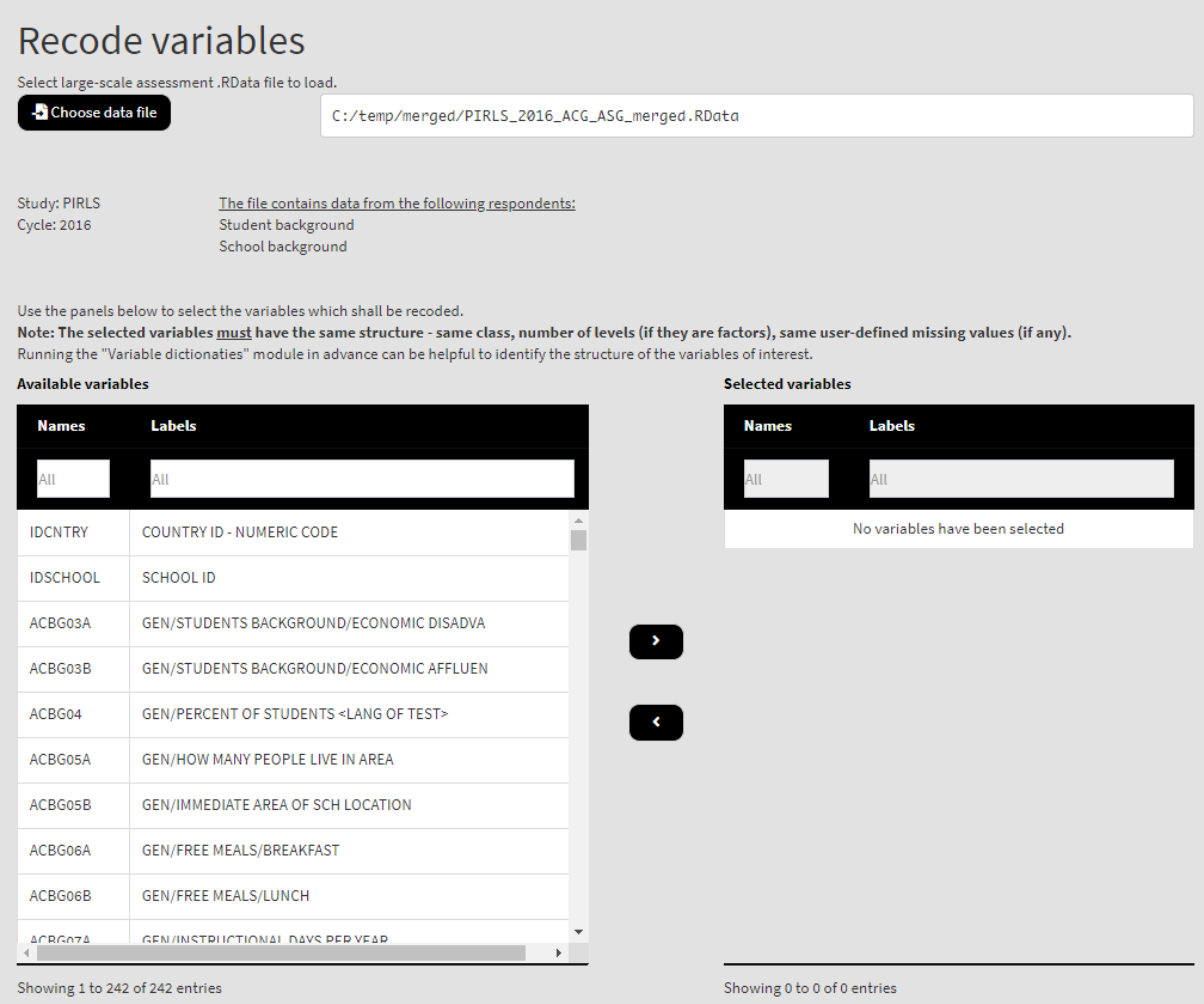
```
ralsaGUI()
```

Za naslednje primere združite novo datoteko s podatki PIRLS 2016 za Avstralijo in Slovenijo ter vključite vse spremenljivke učencev in ravnateljev. Združeno datoteko lahko poimenujete **PIRLS_2016_ACG_ASG_merged.RData.**

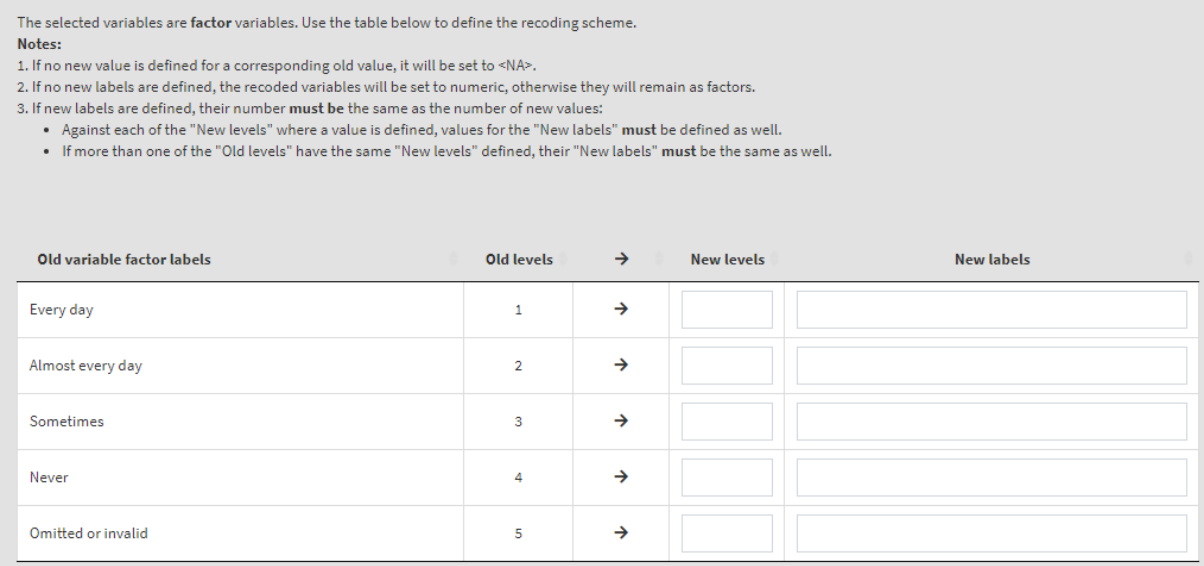
Ko končate združevanje podatkov, v meniju na levi strani izberite **Data preparation** > **Recode variables**. Ko ste v **GUI** navigirani do **Recode variables**, kliknite gumb **Choose data file**. Pomaknite se do mape, ki vsebuje združeno datoteko **»PIRLS_2016_ACG_ASG_merged.RData«**, jo izberite in kliknite gumb **Select**.



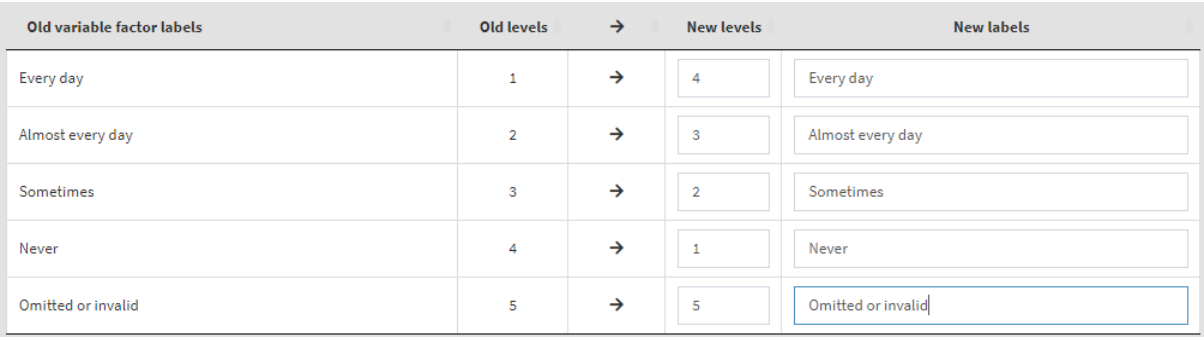
Ko je datoteka naložena, boste videli dva panela: enega z razpoložljivimi spremenljivkami in drugega s trenutno praznimi izbranimi spremenljivkami:



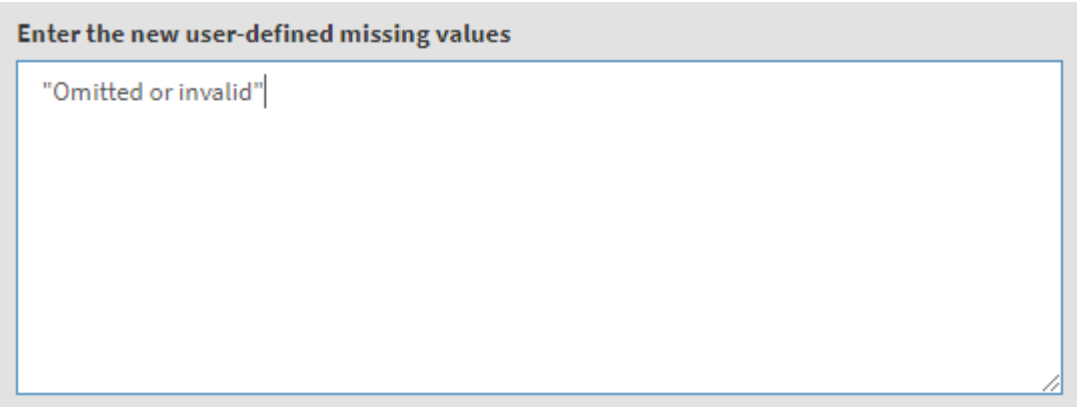
Uporabite miško za izbiro posameznih spremenljivk in jih s pomočjo enojnih puščičnih gumbov premaknite s seznama razpoložljivih spremenljivk na seznam izbranih spremenljivk ter obratno. Na vrhu obeh panelov lahko uporabite polja za filtriranje, da hitro najdete potrebne spremenljivke. Upoštevajte, da morajo imeti izbrane spremenljivke enako strukturo – razred, število ravni/unikatnih vrednosti, oznake (če obstajajo) – in uporabniško določene manjkajoče vrednosti (če obstajajo). Če kateri od teh pogojev ni izpolnjeni, GUI ne bo dovolil nadaljevanja in prikazala se bodo opozorila. Za začetek obrnimo odgovore učencev na vprašanje, kako pogosto se počutijo utrujeni v šoli. Torej uporabimo ASBG07 (najprej se izplača preveriti slovar te spremenljivke). Poiščite spremenljivko na seznamu razpoložljivih spremenljivk na desni strani (za hitrejše iskanje lahko uporabite filter na vrhu) in jo premaknite na seznam izbranih spremenljivk. Priporočljivo bi bilo najprej preveriti slovar te spremenljivke. Ko so na panelu **Selected variables** prisotne kakršne koli spremenljivke, se bodo pojavili naslednji elementi:



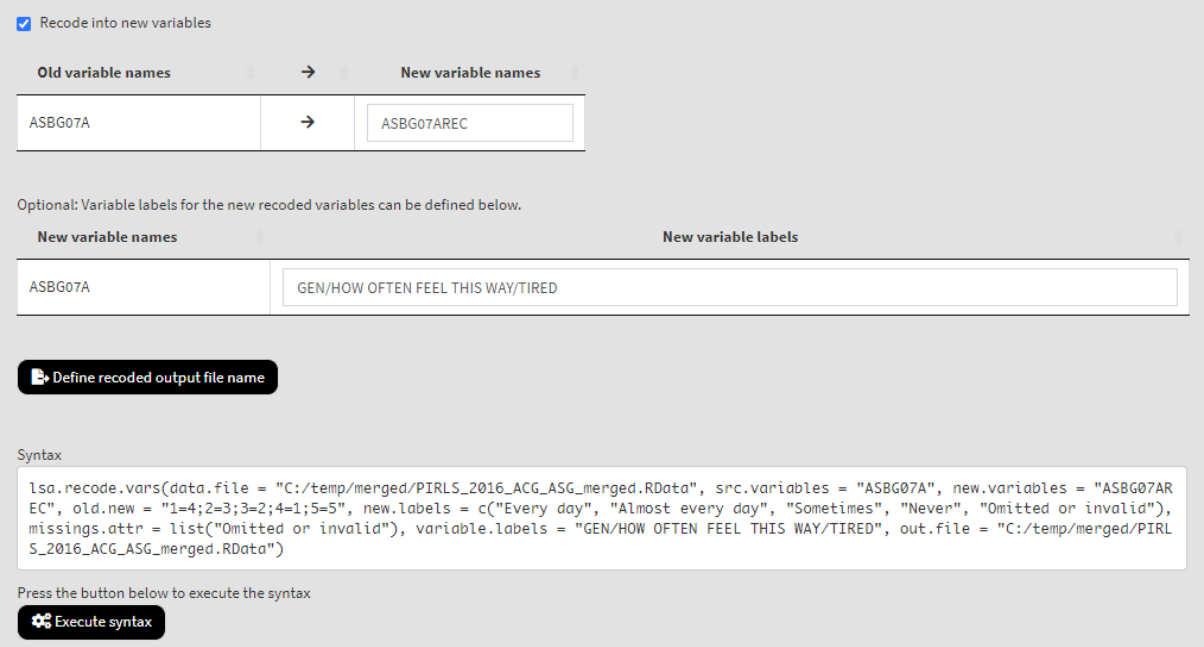
Preberite opombe na vrhu, saj predstavljajo pomembna razmišljanja za nadaljnje korake, ki jih morate izvesti. Za določitev starih in novih vrednosti jih vpišite v stolpec **New levels** v tabeli zgoraj. Upoštevajte, da med določanjem navodil za rekodiranje GUI prikaže različna opozorila v rdeči barvi nad tabelo. Prepričajte se, da jih preberete, saj so pomembna. Vsako opozorilo ne predstavlja napake; nekatera vas samo opozarjajo, da bo izvedeno določeno dejanje. Če želite, da spremenljivka ostane **factor** (kategorična) in ima oznake, vnesite nove oznake v ustrezen stolpec. Če jih izpustite, bo rekodirana spremenljivka postala **numeric** (številska). Poleg tega bo uporabniško določena manjkajoča kategorija v tabeli (»**Omitted or invalid**«) ponovno definirana kot zadnja kategorija (5). Če jo izpustite, bo rekodirana v NA. Nastavitve bodo izgledale takole:



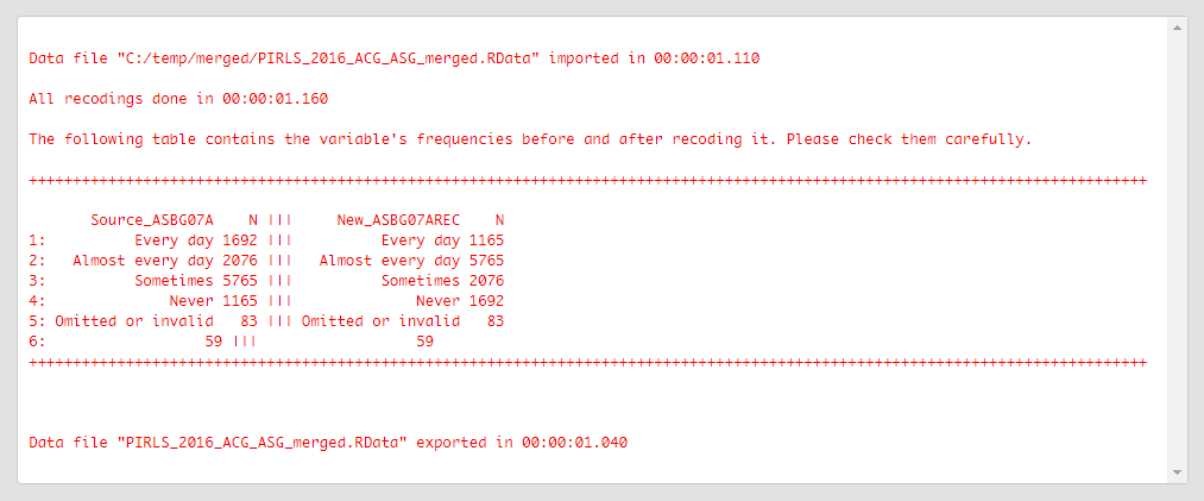
Nad tabelo boste videli besedilno polje za določitev uporabniško določenih manjkajočih vrednosti. Tja bomo dodali **Omitted or invalid**, da bo ostala uporabniško določena manjkajoča vrednost. Lahko tudi dodate druge vrednosti iz zgornje tabele (tj. morajo obstajati kot nove vrednosti) in bodo nastavljene kot uporabniško določene manjkajoče vrednosti, če bo to potrebno. Besedilno polje bo izgledalo takole:



Nad tabelami boste videli potrditveno polje **Recode into new variables**. Privzeto bo izbrano, kar pomeni, da bodo rekodirane vrednosti shranjene v novo spremenljivko. V nasprotnem primeru bo izvirna spremenljivka prepisana. Močno vam svetujemo, da to možnost pustite označeno in rekodiranja shranite v novo spremenljivko. V tabelah spodaj dodajte nova imena spremenljivk in njihove oznake. Ko vnesete novo ime spremenljivke, se bo pojavilo polje **Define recoded output file name**. Kliknite nanj, pojdite v mapo, kjer se nahaja izvorna datoteka, kliknite nanjo in potrdite, da bo prepisana. Ko to storite, se bosta pojavila sintaksa za izvedbo in gumb **Execute syntax**. Končne nastavitve za ta del bodo izgledale takole:



Kliknite gumb **Execute syntax**. Konzola GUI se bo pojavila na dnu in beležila vse zaključene operacije:



Preverite stare (izvorne) in nove vrednosti, da se prepričate, da so bila vsa rekodiranja opravljena, kot je bilo načrtovano.

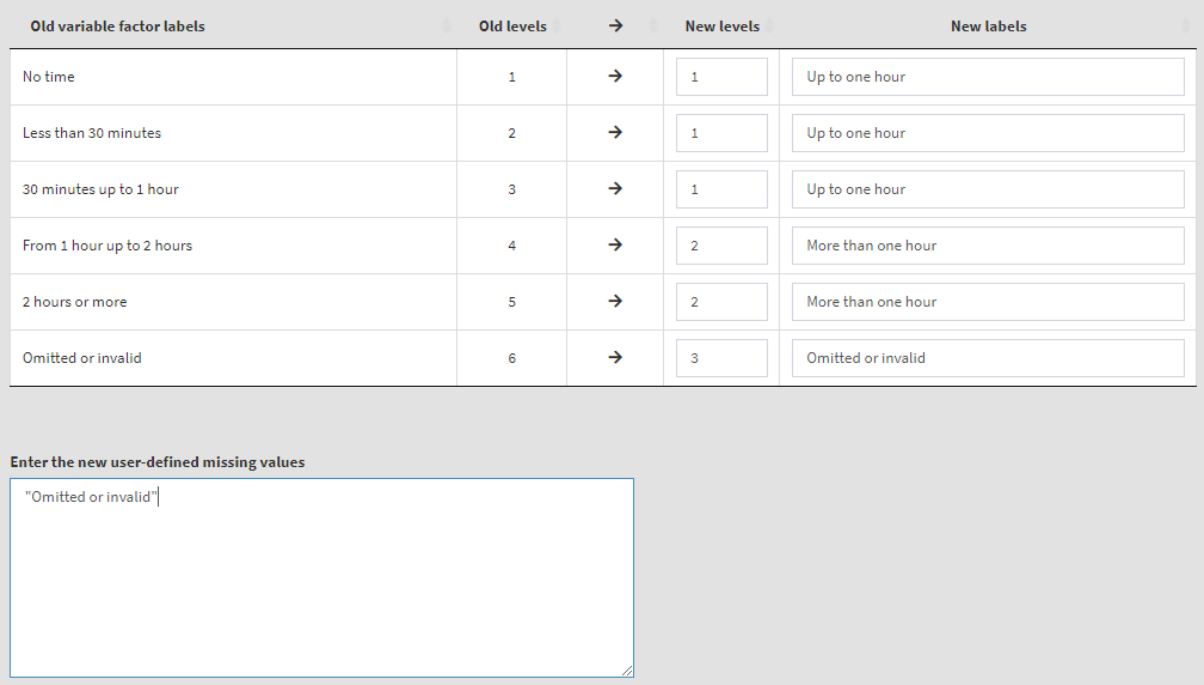
Sedaj bomo združili kategorije več spremenljivk hkrati. Upoštevajte, da morajo biti pri rekodiranju več spremenljivk naenkrat vse spremenljivke istega razreda (faktor, numerična ali besedilna), imeti enake ravni/edinstvene vrednosti in uporabniške manjkajoče vrednosti (če obstajajo). Če kateri od teh pogojev ni izpolnjen, bo funkcija ustavila delo z napako. V naslednjem primeru bomo rekodirali štiri spremenljivke ASBG11A, ASBG11B, ASBG11C in ASBG11D. Te spremenljivke vsebujejo podatke o tem, kako dolgo učenci opravljajo določene dejavnosti na dan – igranje računalniških iger, gledanje videov, klepetanje ali uporaba interneta. Vse spremenljivke imajo enako strukturo – vse so faktorji in njihove kategorije so naslednje:

- Noben čas;
- Manj kot 30 minut;
- 30 minut do 1 ure;
- Od 1 ure do 2 ur;
- 2 uri ali več;
- Izpuščeno ali neveljavno.

Zadnja kategorija je uporabniško določena manjkajoča vrednost. Združili bomo veljavne kategorije odgovorov v dve:

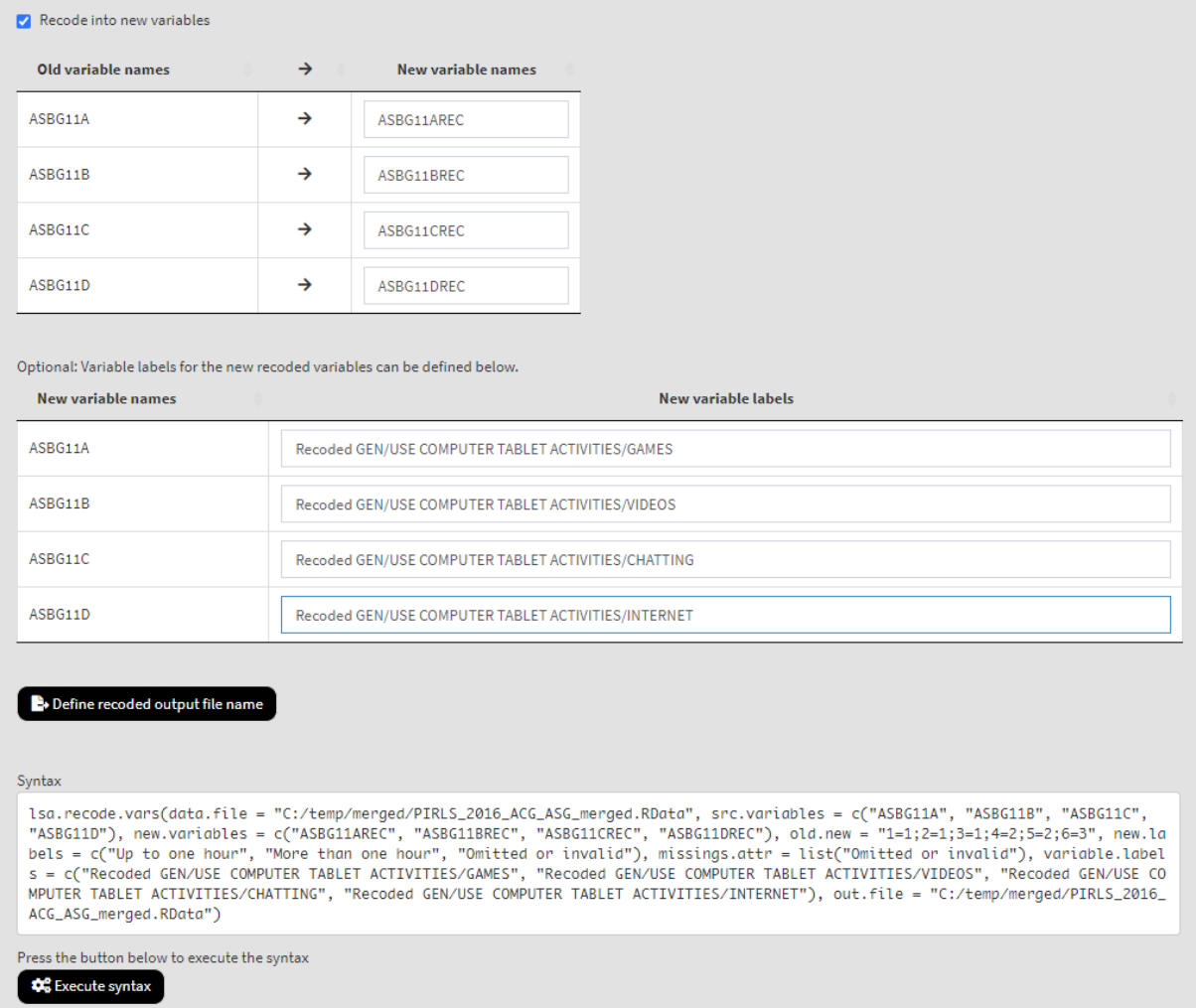
Do ene ure (tj. združimo kategorije 1, 2 in 3); in
Več kot ena ura (tj. združimo kategorije 4 in 5).

Za začetek odstranite ASBG07A s seznama **Selected variables**. S seznama **Available variables** izberite spremenljivke ASBG11A, ASBG11B, ASBG11C in ASBG11D (lahko uporabite filter pod imenom) in jih premaknite na seznam **Selected variables.** Tabela, ki se bo pojavila spodaj, bo uporabljena za nastavljanje rekodiranj za vse štiri spremenljivke naenkrat. Ohranili bomo **Omitted or invalid** kot uporabniško določeno manjkajočo vrednost. Glede na zgoraj opisano rekodiranje bodo nastavitve izgledale takole:



Old variable factor labels	Old levels	→	New levels	New labels
No time	1	→	1	Up to one hour
Less than 30 minutes	2	→	1	Up to one hour
30 minutes up to 1 hour	3	→	1	Up to one hour
From 1 hour up to 2 hours	4	→	2	More than one hour
2 hours or more	5	→	2	More than one hour
Omitted or invalid	6	→	3	Omitted or invalid

Zdaj imamo več kot eno spremenljivko, ki jo je treba rekodirati, in za vsako od starih (izvornih) spremenljivk moramo določiti imena za nove spremenljivke, v katere bodo le-te rekodirane, ter njihove oznake:



Če je treba, spremenite ime izhodne datoteke tako, da kliknete na gumb **Define recoded output file name**. Kliknite gumb **Execute syntax**. Začelo se bo izvajanje, konzola se bo prikazala na dnu GUI. Videli boste sporočilo, ki vas obvesti, ko so vse operacije zaključene. Tokrat ima konzola precej več izpisa. Pomaknite se navzdol in preverite, ali so bila vsa rekodiranja izvedena tako, kot je bilo predvideno.

```
Data file "C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData" imported in 00:00:01.029

All recodings done in 00:00:01.180

The following tables contain the variables' frequencies before and after recoding them. Please check them carefully.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

              Source_ASBG11A    N |||     New_ASBG11AREC    N
1:                   No time 1299 |||     Up to one hour 7962
2:       Less than 30 minutes 4101 ||| More than one hour 2678
3:    30 minutes up to 1 hour 2562 ||| Omitted or invalid  122
4: From 1 hour up to 2 hours 1148 |||                   78
5:            2 hours or more 1530 |||                NaN  NaN
6:         Omitted or invalid  122 |||                NaN  NaN
7:                         78 |||               NaN  NaN
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

              Source_ASBG11B    N |||     New_ASBG11BREC    N
1:                   No time 1913 |||     Up to one hour 8096
2:       Less than 30 minutes 3940 ||| More than one hour 2492
3:    30 minutes up to 1 hour 2243 ||| Omitted or invalid  177
4: From 1 hour up to 2 hours 1293 |||                   75
5:            2 hours or more 1199 |||                NaN  NaN
```

4. Izvedba analiz

4.1 Odstotki in srednje vrednosti (angl. *Percentages and means*)

4.1.1 Uvod

Funkcija `lsa.pcts.means` izračuna odstotke respondentov znotraj skupin respondentov, definiranih s kategorijami razdelitvenih spremenljivk, in srednje vrednosti (aritmetična sredina, mediana ali modus) zveznih spremenljivk znotraj teh skupin. Tako razdelitvene spremenljivke kot tudi spremenljivke za izračun srednjih vrednosti so opcijske. Če razdeliltvene spremenljivke niso podane, se rezultati izračunajo le na ravni države. Če so razdelitvene spremenljivke podane, se podatki znotraj vsake države razdelijo v skupine glede na vse razdelitvene spremenljivke, odstotki anketirancev pa se izračunajo za zadnjo razdelitveno spremenljivko. Če so podane spremenljivke za izračun srednjih vrednosti, se njihove srednje vrednosti znotraj skupin, definiranih z vsemi razdelitvenimi spremenljivkami, izračunajo. Upoštevajte, da je mogoče srednje vrednosti izračunati tako za ozadje/kontekstne spremenljivke kot tudi za nize PV, pri čemer se upoštevata kompleksen vzorec in zasnova ocenjevanja raziskave, ki nas zanima. V slednjem primeru se srednje vrednosti izračunajo za vsako PV v nizu, nato pa se ocene za vse PV v nizu povprečijo, standardna napaka pa se izračuna z uporabo zapletenih formul, ki so odvisne od konkretne raziskave. Ne glede na oceno se standardna napaka izračuna ob upoštevanju kompleksne zasnove vzorčenja in dizajna raziskav. Če vas zanimajo podrobnejši podatki o kompleksni zasnovi vzorčenja in ocenjevanja določene raziskave ter o tem, kako se izračunajo ocene in njihove standardne napake, se obrnite na njeno tehnično dokumentacijo in uporabniški priročnik.

Kot katera koli druga funkcija v paketu RALSA lahko tudi funkcija `lsa.pcts.means` prepozna podatke iz raziskave in uporabi ustrezne tehnike ocenjevanja glede na izvedbo vzorčenja ter zasnovo raziskave brez dodatne skrbi.

4.1.2 Funkcija za odstotke in srednje vrednosti ter njeni argumenti

Funkcija `lsa.pcts.means` ima naslednje argumente:

- `data.file` – datoteka, ki vsebuje objekt lsa.data. Določiti je treba bodisi to bodisi `data.object`, ne pa obeh.
- `data.object` – objekt v spominu, ki vsebuje objekt `lsa.data`. Določiti je treba bodisi tega bodisi `data.file`, ne pa obeh.
- `split.vars` – kategorijska spremenljivka za razdelitev rezultatov. Če niso podane nobene razdelitvene spremenljivke, bodo prikazani rezultati za celotno populacijo držav. Če je podana ena ali več spremenljivk, bodo rezultati razdeljeni po vseh spremenljivkah razen po zadnji, odstotki anketirancev pa bodo izračunani po edinstvenih vrednostih zadnje razdelitvene spremenljivke.
- `bckg.avg.vars` – ime(-na) zveznene(-ih) kontekstualne(-ih) spremenljivke(-k) oz. ozadja(-ij) za izračun srednjih vrednosti. Rezultati bodo izračunani po vseh skupinah, ki jih določajo razdelitvene spremenljivke.
- `PV.root.avg` – koren ime(-na) za nize možnih vrednosti (PV).

- `central.tendency` – katero mero osrednje tendence je treba izračunati – srednja vrednost (privzeto), mediana ali modus.
- `weight.var` – ime spremenljivke, ki vsebuje uteži. Če ime spremenljivke z utežmi ni navedeno, bo funkcija samodejno izbrala privzeto utež za navedene podatke, odvisno od vrste anketiranca.
- `include.missing` – eli naj bodo manjkajoče vrednosti razdelitvenih spremenljivk vključene kot kategorije za razdelitev, pri čemer bodo vse statistike ustvarjene tudi zanje. Privzeto (`FALSE`) vzame vse primere brez manjkajočih vrednosti v razdelitvenih spremenljivkah, preden izračuna statistiko.
- `shortcut` – določitev, ali naj se uporabi »bližnjica« za IEA TIMSS, TIMSS Advanced, TIMSS Numeracy, eTIMSS, PIRLS, ePIRLS, PIRLS Literacy in RLII. Privzeto (`FALSE`) uporablja »polno« zasnovo pri izračunu komponent variance in standardnih napak ocen.
- `graphs` – določitev, ali naj bodo ustvarjeni grafi. Privzeto je `FALSE`.
- `perc.x.label` – niz, prilagojena oznaka za vodoravno os na grafih odstotkov. Ignorirano, če je `graphs = FALSE`.
- `perc.y.label` – niz, prilagojena oznaka za navpično os na grafih odstotkov. Ignorirano, če je `graphs = FALSE`.
- `mean.x.labels` – seznam nizov, prilagojene oznake za vodoravno os na grafih srednjih vrednosti. Ignorirano, če je `graphs = FALSE`.
- `mean.y.labels` – seznam nizov, prilagojene oznake za navpično os na grafih srednjih vrednosti. Ignorirano, če je `graphs = FALSE`.
- `save.output` – določitev, ali naj se izhod shrani v MS Excel-datoteko (privzeto) ali ne (natisnjeno v konzolo ali dodeljeno objektu).
- `output.file` – celotna pot do izhodne datoteke, vključno z imenom datoteke. Če je izpuščeno, bo datoteka z privzetim imenom »Analysis.xlsx« zapisana v delovni imenik (`getwd()`).
- `open.output` – določitev, ali naj se izhod odpre po tem, ko je bil napisan. Privzeto (`TRUE`) se izhod odpre v privzetem programu za preglednice, nameščenem na računalniku.

Opombe:

1. Bodisi `data.file` bodisi `data.object` mora biti naveden kot vir podatkov. Če sta podana oba, se bo funkcija ustavila z napako. Podatki morajo biti razreda `lsa.data` in pretvorjeni iz SPSS (ali TXT v primeru PISA pred letom 2015) z uporabo funkcije `lsa.convert.data`.
2. Funkcija izračuna odstotke anketirancev, določenih po kategorijah razdelitvenih spremenljivk. Odstotki se izračunajo znotraj skupin, določenih z zadnjo razdelitveno spremenljivko. Če so podane zvezne spremenljivke (ozadje ali nizi možnih vrednosti – PV), se njihove srednje vrednosti izračunajo po skupinah, definiranih z eno ali več razdelitvenimi spremenljivkami. Če ni dodanih razdelitvenih spremenljivk, se rezultati izračunajo le po državi.
3. Možno je zagotoviti več zveznih ozadenjskih spremenljivk za izračun njihovih srednjih vrednosti. Upoštevajte, da bodo v tem primeru rezultati nekoliko drugačni v primerjavi z uporabo posameznih ozadnih zveznih spremenljivk v ločenih analizah. To je zato,

ker se primeri z manjkajočimi vrednostmi na `bckg.avg.vars` odstranijo vnaprej, in več kot je spremenljivk, podanih v `bckg.avg.vars`, več primerov bo verjetno odstranjenih.

4. Izračun srednjih vrednosti, ki vključuje možne vrednosti (PV), zahteva navedbo korena imen možnih vrednosti v PV.root.avg. Vse raziskave (razen CivED, TEDS-M, SITES, TALIS in TALIS Starting Strong Survey) imajo niz PV na konstruktu (npr., v TIMSS pet za celotno matematiko, pet za algebro, pet za geometrijo itd.). V nekaterih raziskavah (kot sta TIMSS in PIRLS) se imena PV v nizu vedno začnejo z naborom znakov in končajo s številko PV. Npr., imena niza PV za celotno matematiko v TIMSS so BSMMAT01, BSMMAT02, BSMMAT03, BSMMAT04 in BSMMAT05. Koren PV za ta niz, ki ga je treba dodati v `PV.root.avg`, bo BSMMAT. Funkcija bo samodejno našla vse spremenljivke v tem nizu PV in jih vključila v analizo. V drugih raziskavah, kot so OECD PISA in IEA ICCS ter ICILS, je zaporedna številka vsakega PV vključena v sredino imena. Npr., v ICCS so imena niza PV PV1CIV, PV2CIV, PV3CIV, PV4CIV in PV5CIV. Koren imena PV, ki ga je treba določiti v `PV.root.avg`, je »PV#CIV«. Mogoče je dodati več nizov PV. Vendar upoštevajte, da bo zagotavljanje zveznih spremenljivk za argument `bckg.avg.vars` in koren PV za argument `PV.root.avg` vplivalo na rezultate za PV, saj bodo primeri z manjkajočimi vrednostmi na `bckg.avg.vars` odstranjeni, kar bo vplivalo tudi na rezultate PV. Po drugi strani uporaba več kot enega niza PV hkrati ne bi smela vplivati na rezultate PV, ker te ne bi smele imeti nobenih manjkajočih vrednosti.
5. Če za `bckg.avg.vars` in koren imena PV za `PV.root.avg` niso določene nobene spremenljivke, bo izhod vseboval le odstotke primerov v skupinah, določenih z razdelitvenimi spremenljivkami, če te sploh obstajajo.
6. Privzeta mera osrednje tendence je aritmetična sredina (povprečje). To je mogoče spremeniti v mediano ali modus.
7. Če je `include.missing = FALSE` (privzeto), bodo odstranjeni vsi primeri z manjkajočimi vrednostmi na razdelitvenih spremenljivkah in v statistike bodo vključeni samo primeri z veljavnimi vrednostmi. Upoštevajte, da je podatke iz raziskav mogoče izvoziti na dva različna načina: (1) z nastavitvijo vseh uporabniško določenih manjkajočih vrednosti na `NA`; (2) z uvozom vseh uporabniško določenih manjkajočih vrednosti kot veljavnih in dodajanjem njihovih kod kot dodatnega atributa vsaki spremenljivki. Če je `include.missing` nastavljeno na `FALSE` (privzeto) in so uporabljeni podatki izvoženi z možnostjo (2), bo izhod iz spremenljivke odstranil vse vrednosti, ki ustrezajo vrednostim v njenem atributu missings. V nasprotnem primeru jih bo vključil kot veljavne vrednosti in zanje izračunal statistike.
8. Argument `shortcut` je veljaven le za TIMSS, TIMSS Advanced, TIMSS Numeracy, PIRLS, ePIRLS, PIRLS Literacy in RLII. Prej so te raziskave pri izračunu standardnih napak uporabljale 75 replikacij (angl. *replicates*), ker je bila ena izmed šol v 75 območjih JK podvojena, druga pa izločena. Od TIMSS 2015 in PIRLS 2016 naprej raziskave uporabljajo 150 replikacij in v vsakem območju JK je enkrat šola podvojena, enkrat izločena, torej se izračuni izvedejo dvakrat za vsako območje. Za več podrobnosti glejte tehnično dokumentacijo in uporabniške priročnike TIMSS 2015 in PIRLS 2016. Če je treba tabele in slike replicirati, je treba argument `shortcut` nastaviti na `TRUE`.
9. Če je `graphs = TRUE`, funkcija ustvari grafe, stolpčne grafe za odstotke in napake v srednjih vrednostih za skupine, določene s `split.vars`. Vsi grafi so izdelani po

državah. Če ni določenih `split.vars`, bodo na koncu grafi odstotkov in napak za vse države skupaj. Po potrebi lahko prilagojene oznake za vodoravno in navpično os na stolpčnih grafih ter grafih napak določite z argumenti `perc.x.label`, `perc.y.label`, `mean.x.labels` in `mean.y.labels`.

Če je `save.output = FALSE`, bo vrnjen seznam, ki vsebuje ocene in informacije o analizi. Če je `graphs = TRUE`, bodo grafi dodani na seznam ocen.

Če je `save.output = TRUE` (privzeto), bo shranjena datoteka MS Excel (.xlsx) (ki jo lahko odprete v katerem koli programu za preglednice), kot je določeno s polno potjo v `output.file` (izhodna datoteka). Če argument manjka, bo datoteka Excel z generičnim imenom »Analysis.xlsx« shranjena v delovni imenik (`getwd()`). Delovni zvezek vsebuje tri preglednice. Prva (»Estimates« (ocene)) vsebuje tabelo z rezultati po državah, zadnji del tabele pa vsebuje povprečne rezultate vseh statističnih podatkov držav. Glede na specifikacijo analize lahko v tabeli najdete naslednje stolpce:

- **<Country ID>** – stolpec, ki vsebuje imena držav v datoteki, za katere so izračunane statistike. Natančen naslov stolpca bo odvisen od identifikatorja države, uporabljenega v določeni raziskavi.
- **<Split variable 1>, <Split variable 2>…** – stolpci, ki vsebujejo kategorije, po katerih so bile statistike razdeljene. Natančna imena bodo odvisna od spremenljivk v `split.vars` (razdelitvene spremenljivke).
- **n_Cases** – število primerov v vzorcu, uporabljenem za izračun statističnih podatkov.
- **Sum_<Weight variable>** – ocenjena populacija elementov na skupino po uporabi uteži. Dejanski naziv spremenljivke uteži bo odvisen od uporabljene uteži v analizi.
- **Sum_<Weight variable>_SE** – standardna napaka ocenjene populacije elementov na skupino. Dejanski naziv spremenljivke uteži bo odvisen od uporabljene uteži v analizi.
- **Percentages_<Last split variable>** – odstotki respondentov (ocenjena populacija) na skupine, definirane z razdelitvenimi spremenljivkami v `split.vars`. Odstotki bodo prikazani za zadnjo razdelitveno spremenljivko, ki določa končne skupine.
- **Percentages_<Last split variable>_SE** – standardne napake zgoraj navedenih odstotkov.
- **Mean_<Background variable>** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), povprečje zvezne <Background variable> (ozadenska spremenljivka), navedene v `bckg.avg.vars` (`ozadna.spre.avg`). Za vsako spremenljivko, navedeno v `bckg.avg.vars`, bo en stolpec s povprečnim ocenjenim podatkom.
- **Mean_<Background variable>_SE** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), standardna napaka povprečja zvezne <Background variable>, navedene v `bckg.avg.vars`. Za vsako spremenljivko bo en stolpec s standardno napako povprečja.
- **Variance_<Background variable>** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), varianca za zvezno <Background variable>, navedeno v `bckg.avg.vars`. Za vsako spremenljivko bo en stolpec z ocenjeno varianco.

- **Variance_<Background variable>_SE** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), napaka variance za zvezno <Background variable>, navedeno v `bckg.avg.vars`. Za vsako spremenljivko bo en stolpec z napako variance.
- **SD_<Background variable>** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), standardni odklon za zvezno <Background variable>, navedeno v `bckg.avg.vars`. Za vsako spremenljivko bo en stolpec s standardnim odklonom.
- **SD_<Background variable>_SE** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), napaka standardnega odklona za zvezno <Background variable>, navedeno v `bckg.avg.vars`. Za vsako spremenljivko bo en stolpec z napako standardnega odklona.
- **Median_<Background variable>** – vrnjeno, če je `central.tendency = "median"` (`centralna.tendenca = "mediana"`), mediana za zvezno <Background variable>, navedeno v `bckg.avg.vars`. Za vsako spremenljivko bo en stolpec z mediano.
- **Median_<Background variable>_SE** – vrnjeno, če je `central.tendency = "median"` (`centralna.tendenca = "mediana"`), standardna napaka mediane za zvezno <Background variable>, navedeno v `bckg.avg.vars`. Za vsako spremenljivko bo en stolpec z napako mediane.
- **MAD_<Background variable>** – vrnjeno, če je `central.tendency = "median"` (`centralna.tendenca = "mediana"`), Median Absolute Deviation (MAD) za zvezno <Background variable>, navedeno v `bckg.avg.vars`. Za vsako spremenljivko bo en stolpec z oceno MAD.
- **MAD_<Background variable>_SE** – vrnjeno, če je `central.tendency = "median"` (`centralna.tendenca = "mediana"`), standardna napaka MAD za zvezno <Background variable>, navedeno v `bckg.avg.vars`. Za vsako spremenljivko bo en stolpec z napako MAD.
- **Mode_<Background variable>** – vrnjeno, če je `central.tendency = "mode"` (`centralna.tendenca = "modus"`), modus za ozadenjsko spremenljivko <Background variable>, navedeno v `bckg.avg.vars`. Za vsako spremenljivko bo en stolpec z oceno modusa.
- **Mode_<Background variable>_SE** – vrnjeno, če je `central.tendency = "mode"` (`centralna.tendenca = "modus"`), standardna napaka modusa za zvezno <Background variable>, navedeno v `bckg.avg.vars`. Za vsako spremenljivko bo en stolpec z napako modusa.
- **Percent_Missings_<Background variable>** – odstotek manjkajočih vrednosti za <Background variable>, navedeno v `bckg.avg.vars`. Za vsako spremenljivko bo en stolpec z odstotkom manjkajočih vrednosti.
- **Mean_<root PV>** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), povprečje PV z istim <root PV>, navedenih v `PV.root.avg` (`PV.koren.avg`). Za vsak niz PV, naveden v `PV.root.avg`, bo en stolpec s povprečno oceno.
- **Mean_<root PV>_SE** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), standardna napaka povprečja PV z istim

<root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec s standardno napako povprečja.

- **Mean_<root PV>_SVR** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), komponenta varianc v vzorcu za povprečje PV z istim <root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec s komponento varianc v vzorcu.
- **Mean_<root PV>_MVR** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), komponenta merjenja varianc za povprečje PV z istim <root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec s komponento merjenja varianc.
- **Variance_<root PV>** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), skupna varianca PV z istim <root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec s skupno varianco.
- **Variance_<root PV>_SE** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), standardna napaka skupne variance PV z istim <root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec z napako skupne variance.
- **Variance_<root PV>_SVR** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), komponenta vzorca variance PV z istim <root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec s komponento vzorca variance.
- **Variance_<root PV>_MVR** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), komponenta merjenja variance PV z istim <root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec s komponento merjenja variance.
- **SD_<root PV>** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), standardni odklon PV z istim <root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec s standardnim odklonom.
- **SD_<root PV>_SE** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), standardna napaka standardnega odklona PV z istim <root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec z napako standardnega odklona.
- **SD_<root PV>_SVR** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), komponenta vzorca standardnega odklona PV z istim <root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec s komponento vzorca standardnega odklona.
- **SD_<root PV>_MVR** – vrnjeno, če je `central.tendency = "mean"` (`centralna.tendenca = "povprečje"`), komponenta merjenja standardnega odklona PV z istim <root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec s komponento merjenja standardnega odklona.
- **Median_<root PV>** – vrnjeno, če je `central.tendency = "median"` (`centralna.tendenca = "mediana"`), mediana PV z istim <root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec z mediano.
- **Median_<root PV>_SE** – vrnjeno, če je `central.tendency = "median"` (`centralna.tendenca = "mediana"`), standardna napaka mediane PV z istim

<root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec z napako mediane.

- **MAD_<root PV>** – vrnjeno, če je `central.tendency = "median"` (`centralna.tendenca = "mediana"`), Median Absolute Deviation (MAD) za niz PV, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec z MAD oceno.
- **MAD_<root PV>_SE** – vrnjeno, če je `central.tendency = "median"` (`centralna.tendenca = "mediana"`), standardna napaka MAD za niz PV, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec z napako MAD.
- **Mode_<root PV>** – vrnjeno, če je `central.tendency = "mode"` (`centralna.tendenca = "modus"`), modus PV z istim <root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec z oceno modus.
- **Mode_<root PV>_SE** – vrnjeno, če je `central.tendency = "mode"` (`centralna.tendenca = "modus"`), standardna napaka modusa za PV z istim <root PV>, navedenih v `PV.root.avg`. Za vsak niz PV bo en stolpec z napako modusa.
- **Percent_Missings_<root PV>** – odstotek manjkajočih vrednosti za <root PV>, navedeno v `PV.root.avg`. Za vsak niz PV bo en stolpec z odstotkom manjkajočih vrednosti.

Drugi list (»Analysis information«) vsebuje nekaj dodatnih informacij, povezanih z analizo po državah v naslednjih stolpcih:

- **DATA** (podatkovna datoteka ali podatkovni objekt, uporabljen v analizi) – uporabljena `data.file` ali `data.object`.
- **STUDY** (študija oz. raziskava) – iz katere raziskave izhajajo podatki.
- **CYCLE** (cikel) – iz katerega cikla raziskave izhajajo podatki.
- **WEIGHT** (utež) – katera spremenljivka za utež je bila uporabljena.
- **DESIGN** (načrt vzorčenja) – katera tehnika vzorčenja je bila uporabljena (JRR ali BRR).
- **SHORTCUT** (bližnjica) – logična vrednost, ali je bila uporabljena bližnjica.
- **NREPS** (število replikacijskih uteži) – koliko replikacijskih uteži je bilo uporabljenih.
- **ANALYSIS_DATE** (datum analize) – na kateri datum je bila izvedena analiza.
- **START_TIME** (čas začetka) – kdaj se je analiza začela.
- **END_TIME** (čas konca) – kdaj se je analiza zaključila.
- **DURATION** (trajanje) – kako dolgo je analiza trajala v urah, minutah, sekundah in milisekundah.

Tretji list (»Calling syntax«) vsebuje klic funkcije z vrednostmi za vse parametre, kot je bila izvedena. To je koristno, če je treba analizo kasneje ponoviti.

Če je `graphs = TRUE` (`grafi = TRUE`), bo dodan še en list »Graphs«, ki bo vseboval vse grafe.

Če med izračuni pride do opozoril, bodo ta vključena v dodaten list »Warnings« v delovni zvezek.

4.1.3 Izračun odstotkov in povprečij z uporabo ukazne vrstice

V naslednjih primerih bomo združili novo podatkovno datoteko s podatki učencev in ravnateljev iz PIRLS 2016 (Avstralija in Slovenija), pri čemer bomo vzeli vse spremenljivke iz obeh tipov datotek:

```
lsa.merge.data(inp.folder = "C:/temp",
               file.types = list(acg = NULL, asg = NULL),
               ISO = c("aus", "svn"),
               out.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData")
```

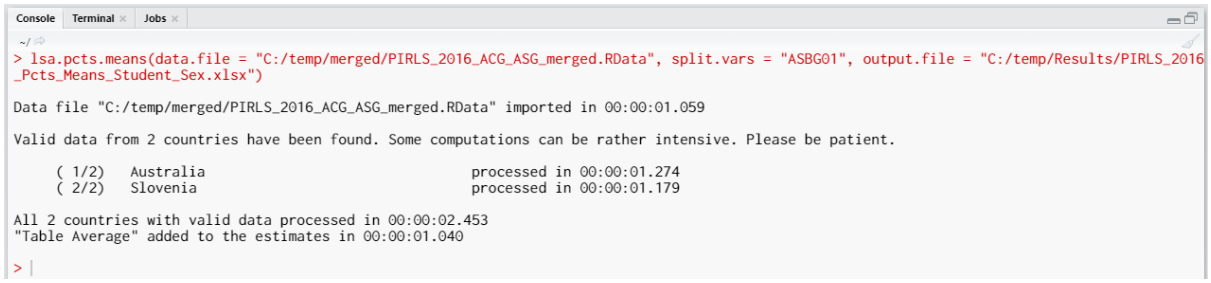
Za začetek izračunajmo odstotke ženskih in moških učencev v Avstraliji in Sloveniji:

```
lsa.pcts.means(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
               split.vars = "ASBG01",
               output.file =
"C:/temp/Results/PIRLS_2016_Pcts_Means_Student_Sex.xlsx")
```

Nekaj stvari, ki jih je treba imeti v mislih:

1. Obstajata dve spremenljivki, ki vsebujeta informacije o spolu učencev:
   - ITSEX, spremenljivka, ki vsebuje informacije o sledenju učencev;
   - ASBG01 (uporabljena v tej analizi), spremenljivka, ki vsebuje spol učencev, kot so ga učenci navedli v vprašalniku.
2. Pri mednarodnih raziskavah morajo biti vse analize izvedene ločeno po državah. Vendar pa ni treba dodajati spremenljivke za ID države (IDCNTRY ali CNT v PISA) kot spremenljivke za razdelitev. Funkcija jo bo samodejno prepoznala in dodala v vektor `split.vars`.
3. Merilo centralne tendence ni bilo izrecno določeno. V tem primeru je bila izračunana aritmetična sredina (povprečje). Za spremembo merila centralne tendence nastavite argument central.tendency na median (mediana) ali mode (modus).
4. Ni treba izrecno določiti spremenljivke za utež. Če spremenljivka za utež ni izrecno določena, bo privzeta utež (v tem primeru celotna utež učenca), uporabljena glede na združene podatke respondentov, identificirana samodejno. Če imate dober razlog za spremembo spremenljivke za utež, to lahko storite z dodajanjem `weight.var = "SENWGT"`.
5. Če izhodna datoteka ni določena, bo izhod shranjen z imenom »Analysis.xlsx« v delovnem imeniku (lahko ga dobite z `getwd()`).
6. Razen če izrecno dodate `save.output = FALSE`, bo izhod zapisan v MS Excel na disk. V nasprotnem primeru bo izhod natisnjen v konzolo.
7. Razen če izrecno dodate `open.output = FALSE` v sintakso, se bo izhodna/output datoteka odprla po končanih izračunih. To je uporabno, kadar se izvede več sintaks za različne analize in ni potrebna takojšnja preglednost izhoda.

Izvedba zgornje kode bo natisnila naslednji output v konzoli RStudia:



Ko so vse operacije končane, bo izhod zapisan na disk kot Excelov delovni zvezek. Če je `open.output = TRUE` (privzeto), se bo datoteka odprla v privzetem programu za preglednice (običajno MS Excel).

Izračunajmo povprečja spremenljivke za dekleta in fante. Uporabili bomo kompleksno lestvico o tem, koliko učenci radi berejo (ASBGSLR; za informacije o tem, kako je ta lestvica sestavljena in kakšne so njene lastnosti, preverite tehnično dokumentacijo PIRLS 2016):

```
lsa.pcts.means(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
               split.vars = "ASBG01",
               bckg.avg.vars = "ASBGSLR",
               output.file =
"C:/temp/Results/PIRLS_2016_Pcts_Means_Student_Sex.xlsx")
```

Klicna sintaksa iz zgoraj dodaja bckg.avg.vars in njeno vrednost, ime spremenljivke ASBGSLR. Output ima podobno strukturo kot prejšnji, le da so tokrat dodani stolpci, ki se nanašajo na povprečje za lestvico.

Funkcija lahko izračuna tudi povprečje za niz (ali nize) verjetnostnih vrednosti (PV). Spodnja klicna sintaksa izračuna povprečje splošnih dosežkov pri branju za fante in dekleta:

```
lsa.pcts.means(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
               split.vars = "ASBG01", PV.root.avg = "ASRREA",
               output.file =
"C:/temp/Results/PIRLS_2016_Pcts_Means_Student_Sex.xlsx")
```

Bodite pozorni na to, kako so določene PV. Pet verjetnostnih vrednosti za splošne dosežke pri branju so ASRREA01, ASRREA02, ASRREA03, ASRREA04 in ASRREA05. V argumentu `PV.root.avg` moramo določiti samo osnovo PV, »ASRREA«. Funkcija bo uporabila to osnovo/skupno ime za izbiro vseh petih PV in jih vključila v izračune.

Izvedba zgornje sintakse bo prepisala prejšnji izhod, ker ima enako ime datoteke (opozorilo bo prikazano v konzoli). Stolpci na listu »Estimates« bodo zdaj drugačni.

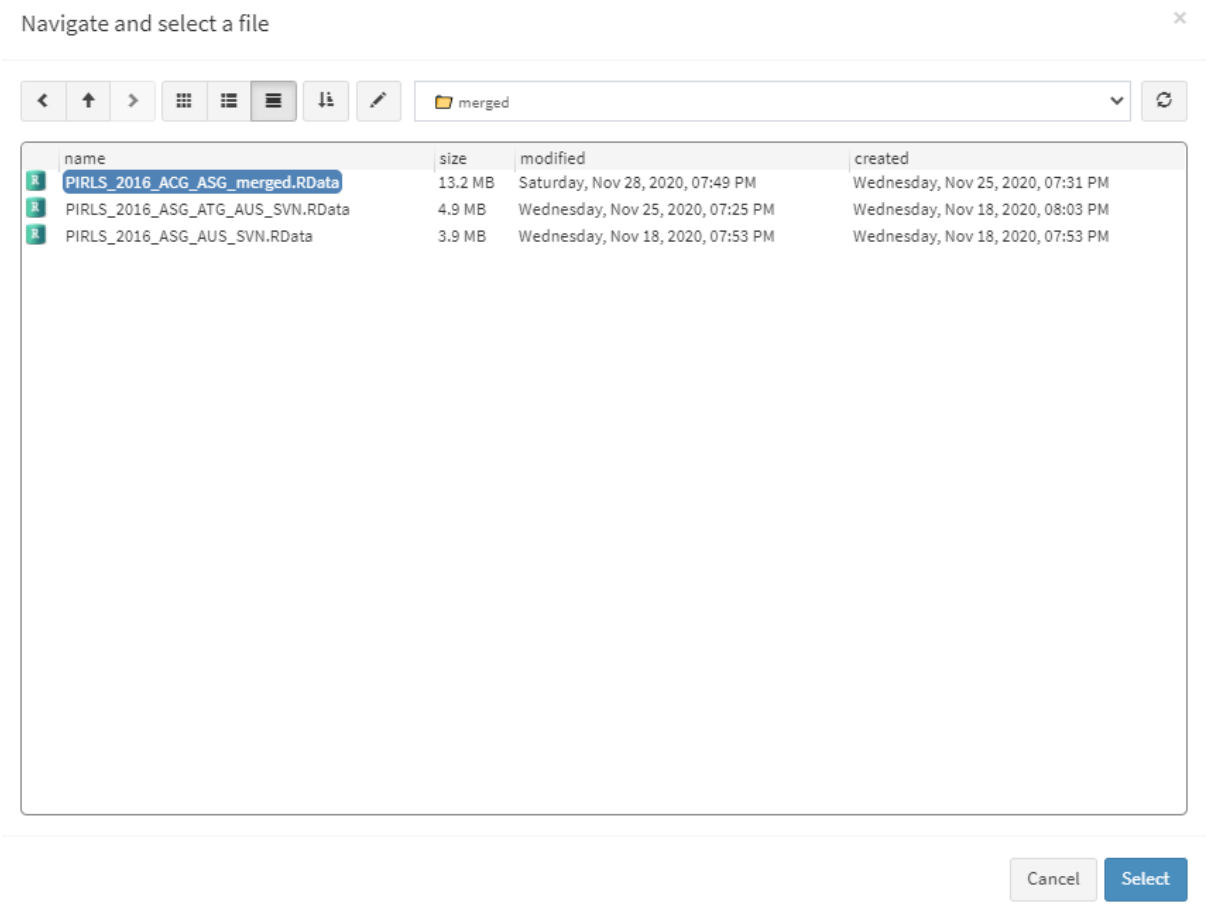
4.1.4 Računanje odstotkov in povprečij z uporabo grafičnega vmesnika (GUI)

Za zagon uporabniškega vmesnika RALSA zaženite naslednji ukaz v RStudiu:

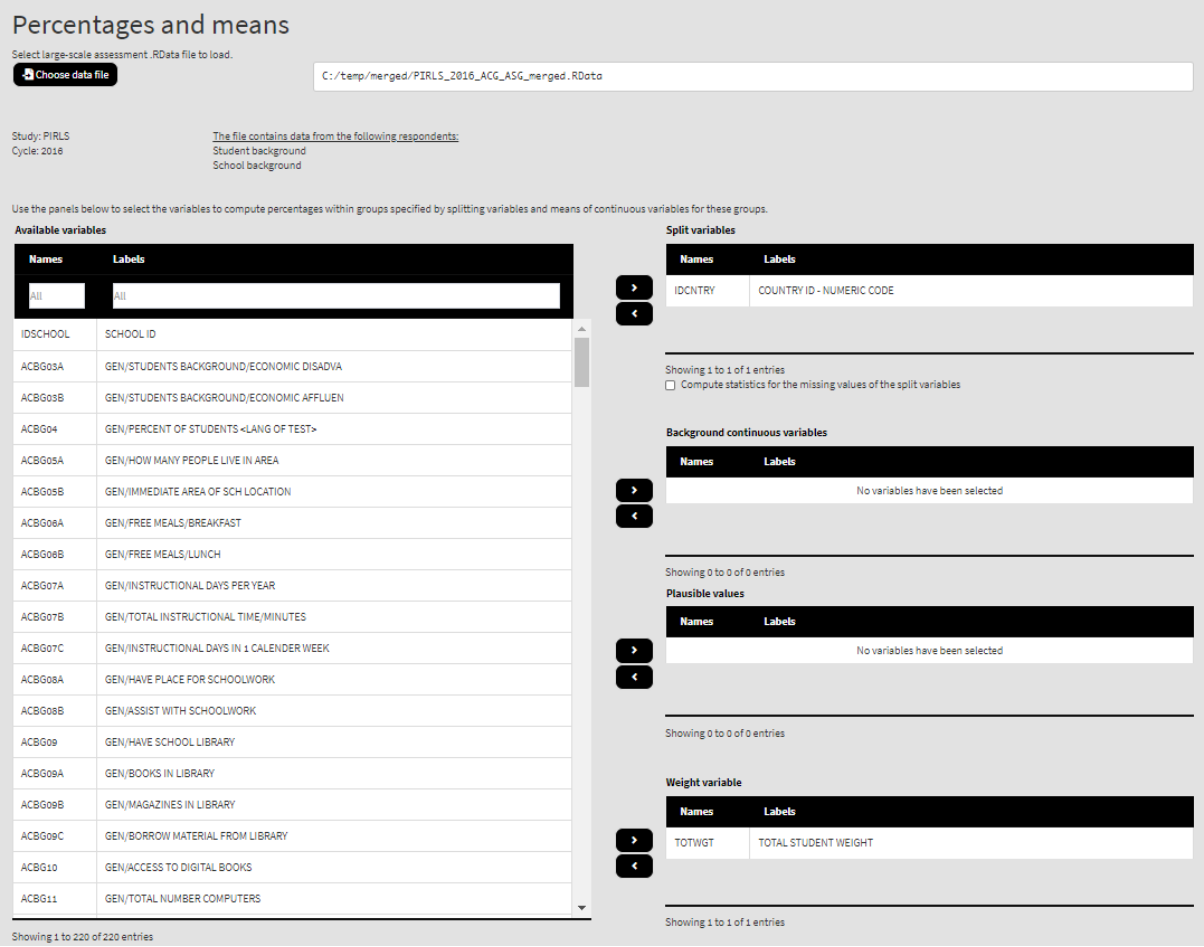
```
ralsaGUI()
```

Za naslednje primere združite novo datoteko z podatki PIRLS 2016 za Avstralijo in Slovenijo, pri čemer upoštevajte vse spremenljivke o učencih in ravnateljih. Kako združiti podatkovne datoteke, si oglejte tukaj. Združeno datoteko lahko poimenujete **»PIRLS_2016_ACG_ASG_merged«**.RData.

Ko je združevanje podatkov končano, izberite **Analysis types > Percentages and means** iz menija na levi strani. Ko se premaknete na Percentages and means v grafičnem vmesniku, kliknite gumb **Choose data file**. Pomaknite se do mape, ki vsebuje združeno datoteko **»PIRLS_2016_ACG_ASG_merged.RData«**, jo izberite in kliknite gumb Select.



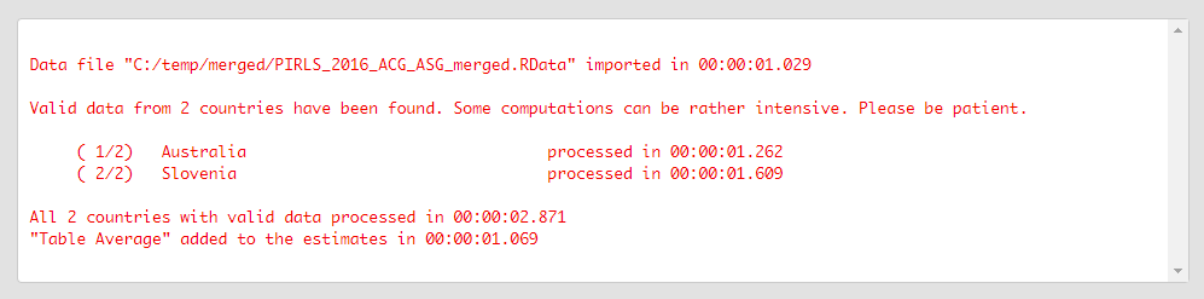
Ko je datoteka naložena, boste na levi strani videli ploščo z razpoložljivimi spremenljivkami (**Available variables**) in nabor panelov na desni, kjer lahko spremenljivke s seznama razpoložljivih spremenljivk dodate v analizo.



Za izbiro spremenljivk s seznama Razpoložljive spremenljivke (**Available variables**) uporabite miško, nato pa s pomočjo puščic na sredini zaslona dodajte (ali odstranite) spremenljivke v različna polja za nastavitev analize. Uporabite iskalna polja na vrhu plošč za hitro iskanje potrebnih spremenljivk.

Za začetek izračunajmo odstotke dijakov in dijakinj v Avstraliji ter Sloveniji. S seznama Razpoložljive spremenljivke (**Available variables**) izberite spremenljivko ASBG01 in jo s pomočjo desne puščice dodajte na seznam Ločene spremenljivke (**Split variables**). To je vse, kar je treba narediti. Pomaknite se navzdol in kliknite gumb **Define output file name** (določi ime izhodne/output datoteke). Pomaknite se do mape **»C:/temp/Results«** (ali mape, kamor želite shraniti output datoteko) in določite ime output datoteke. Ko to storite, se bo poleg določi ime output datoteke (**Define the output file name**) pojavilo potrditveno polje. Če ga obkljukate, se bo output odprl po končanih izračunih. Pod tem se bo prikazala klicna sintaksa. Pod vsem tem bo prikazan gumb Izvedi sintakso (**Execute syntax**). Končne nastavitve v spodnjem delu zaslona bi morale izgledati takole:

Kliknite gumb Izvedi sintakso (**Execute syntax**). Na dnu se bo prikazala konzola GUI, ki bo beležila vse izvedene operacije:

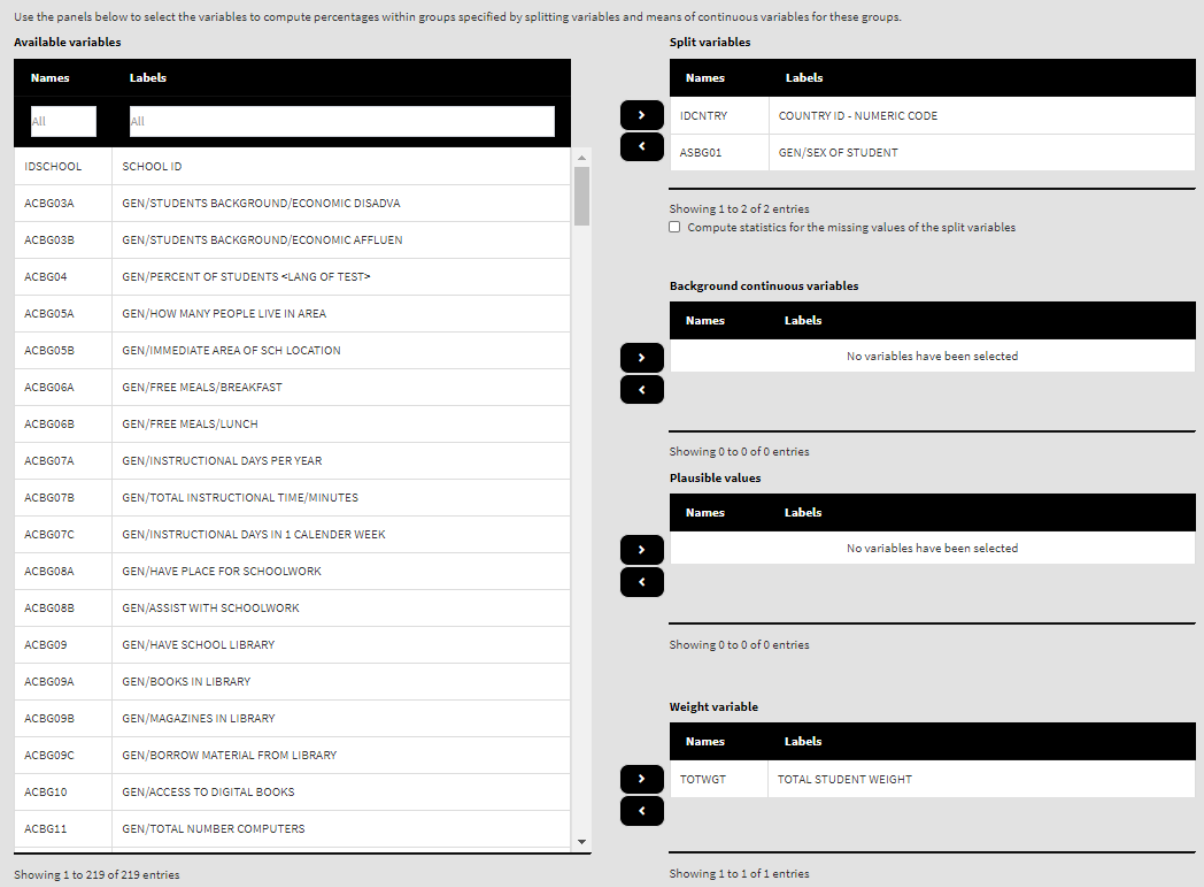


Nekaj stvari, ki jih morate upoštevati:

1. Obstajata dve spremenljivki, ki vsebujeta informacije o spolu učencev:
   - ITSEX, spremenljivka, ki vsebuje informacije o sledenju učencev;
   - ASBG01 (uporabljena v tej analizi), spremenljivka, ki vsebuje podatke o spolu učencev, kot so jih učenci navedli v vprašalniku.
2. Pri velikih mednarodnih raziskavah (angl. *international large-scale assessments*) morajo biti vse analize izvedene ločeno za vsako državo. Spremenljivka z identifikacijsko oznako države (IDCNTRY, ali CNT v PISA) je vedno izbrana kot prva razdelitvena spremenljivka in je ni mogoče odstraniti s plošče Razdelitvene spremenljivke (**Split variables**).
3. Privzeta utežna spremenljivka (angl. *weight variable*) je izbrana in samodejno dodana v ploščo Utežna spremenljivka (**Weight variable**). Spremeniti jo je mogoče z drugo utežno spremenljivko, ki je na voljo v podatkovnem naboru. Če je izbrana privzeta utežna spremenljivka, se ta ne bo prikazala v oknu sintakse. Če na plošči Utežna spremenljivka ni izbrana nobena utežna spremenljivka, bo samodejno uporabljena privzeta.
4. Privzeto polje Uporabi bližnjico za izračun SE (**Use shortcut method for computing SE**) ni označeno. To pomeni, da bo funkcija izračunala standardno napako (SE) z uporabo »polne« metode za komponento vzorčne variance.

Če je označeno polje Odpri output, ko končaš (**Open the output when done**), se bo output samodejno odprl v privzetem programu za preglednice (ponavadi MS Excelu) po zaključku vseh izračunov.

Izračunajmo povprečja za spremenljivko tako za dekleta kot za fante. Uporabili bomo kompleksno lestvico, ki prikazuje, koliko učenci uživajo v branju (ASBGSLR; preverite tehnično dokumentacijo PIRLS 2016 za podrobnosti o konstrukciji te lestvice in njenih značilnostih). Ko ste sprejeli nastavitve za prejšnjo analizo, s seznama Razpoložljive spremenljivke (**Available variables**) izberite spremenljivko ASBGSLR in uporabite puščične gumbe, da jo dodate na ploščo Ozadenjske zvezne spremenljivke (**Background continuous variables**). Nastavitve naj zdaj kot razdelitveni spremenljivki (**Split variables**) vključujejo IDCNTRY in ASBG01 ter kot neprekinjeno osnovno spremenljivko (**Background continuous variables**) ASBGSLR.

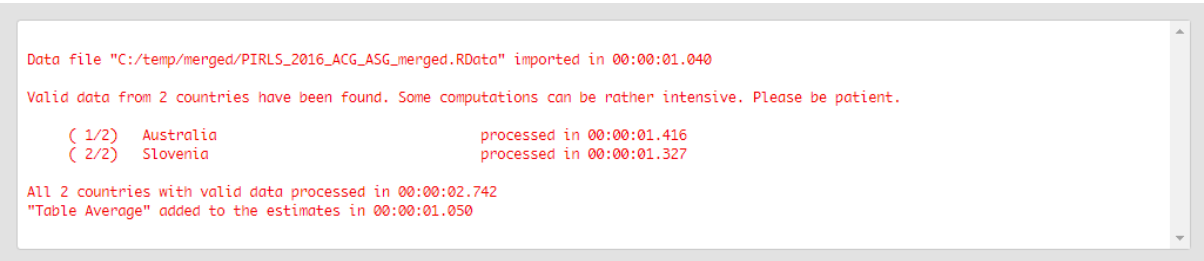


Ker aplikacijo za Odstotke in povprečja (**Percentages and means**) uporabljamo neposredno po izvedbi prejšnje analize, so ostale nastavitve iz prejšnje analize še vedno shranjene. Ni potrebe, da spreminjate katero koli od preostalih nastavitev, razen če to želite. Lahko pa spremenite ime izhodne datoteke, sicer bo prepisana. Upoštevajte, da se bo prikazana sintaksa spremenila, kar bo odražalo vključitev ASBGSLR kot neprekinjene osnovne spremenljivke za izračun povprečja:

```
lsa.pcts.means(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
               split.vars = c("IDCNTRY", "ASBG01"),
               bckg.avg.vars = "ASBGSLR",
               output.file =
"C:/temp/Results/PIRLS_2016_Pcts_Means_Student_Sex.xlsx")
```

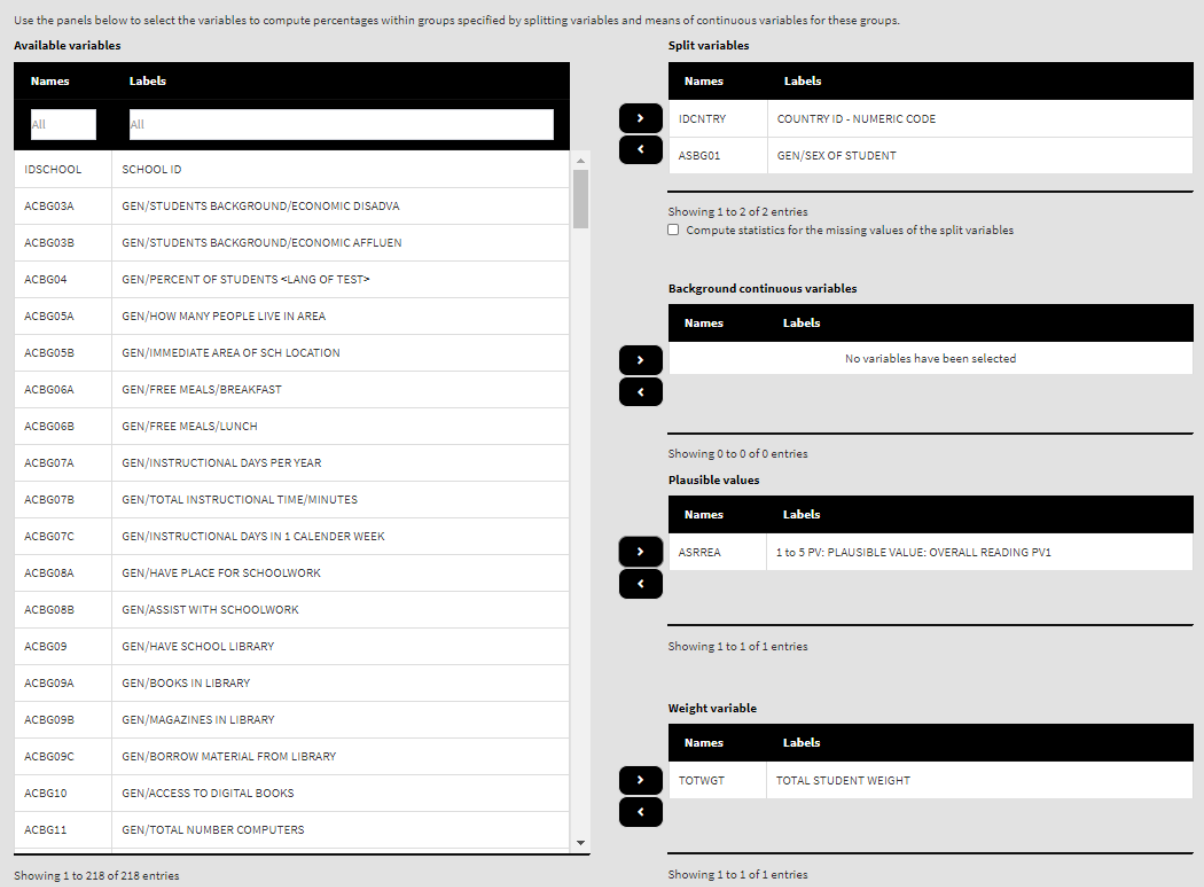
V tej sintaksi je bckg.avg.vars parameter, ki vključuje ASBGSLR kot neprekinjeno osnovno spremenljivko za izračun povprečja.

Pritisnite gumb **Execute syntax** (Izvedi sintakso). Konzola uporabniškega vmesnika (GUI) se bo posodobila in zabeležila vse izvedene operacije:



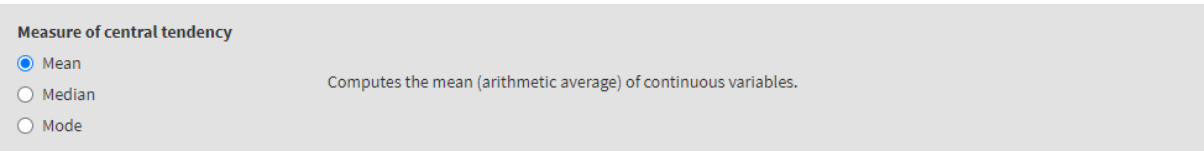
Po zaključku vseh izračunov se output samodejno odpre.

Funkcija lahko tudi izračuna povprečje za nabor (ali nabore) možnih vrednosti. Da nadaljujete, s panela **Background continuous** variables izberite spremenljivko ASBGSLR in jo vrnite na seznam **Available variables** s klikom na levi puščici. S seznama **Available variables** poiščite koren možnih vrednosti za rezultat na bralnem preizkusu (ASRREA). Za iskanje po imenu ali oznaki lahko uporabite filter škatle na vrhu panela. Izberite koren in ga dodajte na panel **Plausible values** z uporabo puščic. Ta del vmesnika bi moral izgledati tako:

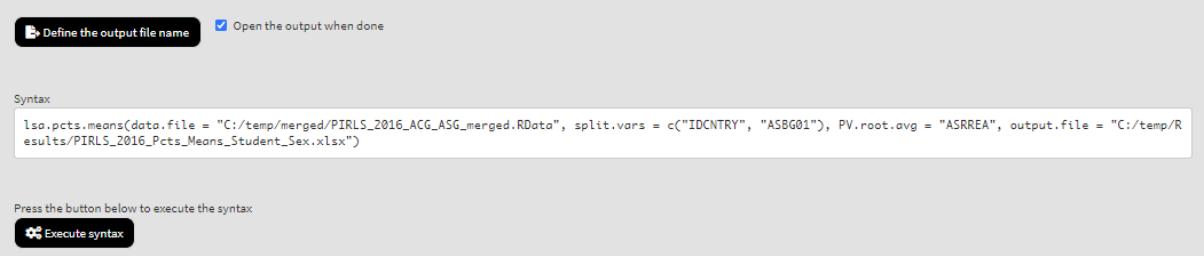


Bodite pozorni na to, kako so določene verjetne vrednosti (PV). Pet verjetnostnih vrednosti za rezultat na preizkusu branja: ASRREA01, ASRREA02, ASRREA03, ASRREA04 in ASRREA05. Seznama **Available variables** in **Plausible values** ne bosta prikazala petih ločenih PV, temveč le njihov koren/skupno ime – ASRREA, brez številk na koncu. Funkcija v ozadju vzame vseh pet PV in jih vključi v izračune.

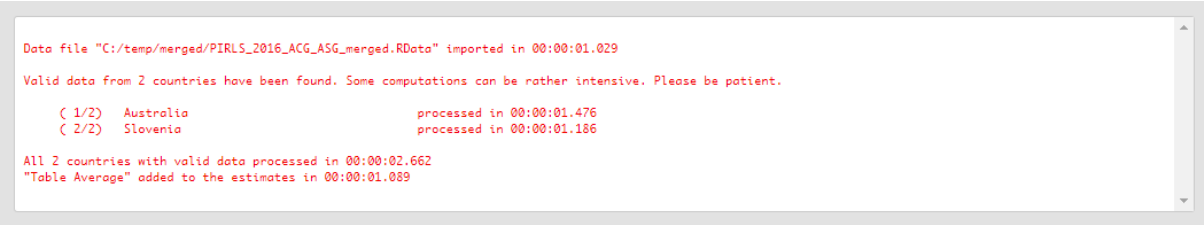
Prav tako bodite pozorni na to, da je privzeta mera osrednje tendence aritmetično povprečje (angl. *mean*). Spremenite ga lahko v mediano ali modus z gumbi neposredno pod paneli za izbiro spremenljivk:



Ker aplikacijo **Percentages and means** uporabljamo neposredno po prejšnji analizi, imamo še vedno preostale nastavitve iz prejšnje analize. Ni potrebe po spreminjanju preostalih nastavitev, razen če to želite. Lahko pa spremenite ime izhodne datoteke, sicer bo prepisana. Bodite pozorni na to, da se prikazani sintaktični ukaz za izračun povprečja na preizkusu iz branja spremeni, saj vključuje koren petih PV, ASRREA:



Pritisnite gumb **Execute syntax**. Konzola GUI se bo posodobila in zabeležila vse zaključene operacije:



Če je odkljukan ukaz **Open the output when done**, se bo output samodejno odprl v privzetem programu za preglednice (po navadi MS Excelu), ko bodo vse izračuni zaključeni.

4.2 Percentili (angl. *Percentiles*)

4.2.1 Uvod

Funkcija `lsa.prctls` izračuna percentile za zvezne spremenljivke. Percentili so točke rezultatov, ki ustrezajo določenemu deležu porazdelitve rezultatov v zvezni spremenljivki. Percentili se lahko izračunajo znotraj skupin, ki so opredeljene s kategorijami razdelitvenih spremenljivk (angl. *split variables*). Te razdelitvene spremenljivke so neobvezne. Če razdelitvenene spremenljivke niso določene, se rezultati izračunajo na ravni države. Če so razdelitvene spremenljivke določene, se podatki znotraj vsake države razdelijo v skupine glede na vse razdelitvene spremenljivke, percentili za zvezne spremenljivke pa se izračunajo za zadnjo razdelitveno spremenljivko. Upoštevajte, da se percentili lahko izračunajo tako za ozadje/kontekstne spremenljivke (angl. *background variables*) kot za nize PV (verjetnostnih vrednosti), pri čemer se upoštevata zapleteno vzorčenje in zasnova ocenjevanja raziskave, ki vas zanima. V slednjem primeru se percentili izračunajo za vsako PV v nizu, nato se povprečijo ocene za vse PV v nizu, standardna napaka pa se izračuna z zapletenimi formulami, ki so odvisne od raziskave. Standardna napaka bo izračunana ob upoštevanju zapletenih zasnov vzorčenja in ocenjevanja v raziskavah. Če vas zanimajo podrobnejši podatki o zapletenih zasnovah vzorčenja in ocenjevanja določene raziskave ter kako se izračunajo ocene in njihove standardne napake, glejte tehnično dokumentacijo in uporabniški priročnik konkretne raziskave.

Kot vsaka druga funkcija v paketu RALSA lahko funkcija `lsa.prctls` prepozna podatke raziskave in uporabi pravilne metode ocenjevanja glede na implementacijo zasnove vzorčenja in ocenjevanja raziskave brez dodatnih prilagoditev.

4.2.2 Funkcija za izračun percentilov in njeni argumenti

Funkcija `lsa.prctls` ima naslednje argumente:

- `data.file` – datoteka, ki vsebuje objekt lsa.data. Treba je določiti bodisi to ali `data.object`, vendar ne obeh hkrati.
- `data.object` – objekt v spominu, ki vsebuje objekt `lsa.data`. Treba je določiti bodisi tega ali `data.file`, vendar ne obeh hkrati.
- `split.vars` – kategorijske spremenljivke, s katerimi se rezultati razdelijo. Če razdelitvene spremenljivke niso določene, bodo prikazani rezultati za splošne populacije držav. Če je določena ena ali več spremenljivk, se rezultati razdelijo glede na vse razen zadnje spremenljivke, odstotki anketirancev pa se izračunajo glede na edinstvene vrednosti zadnje razdelitvene spremenljivke.
- `bckg.prctls.vars` – ime(-na) neprekinjenih ozadnih ali kontekstualnih spremenljivk, za katere se izračunajo percentili. Rezultati bodo izračunani po vseh skupinah, ki jih določajo razdeltvene spremenljivke.
- `PV.root.prctls` – koren skupin verjetnostnih vrednosti (PV) za izračun percentilov.
- prctls – vektor celih števil, ki določa, kateri percentili se izračunajo. Privzeto: `c(5, 25, 50, 75, 95)`.
- `weight.var` – ime spremenljivke, ki vsebuje uteži. Če ni navedeno ime spremenljivke z utežmi, funkcija samodejno izbere privzeto spremenljivko z utežmi glede na podatke, odvisno od vrste anketiranca.
- `include.missing` – logična vrednost, ki določa, ali naj se manjkajoče vrednosti v razdelitvenih spremenljivkah vključijo kot kategorije, po katerih se razdelijo rezultati in izračunajo vse statistike. Privzeto (`FALSE`) vzame vse primere v razdelitvenih spremenljivkah brez manjkajočih vrednosti, preden izračuna statistike.
- `shortcut` – logična vrednost, ki določa, ali naj se pri uporabi PV uporabi metoda »bližnjice« za IEA TIMSS, TIMSS Advanced, TIMSS Numeracy, eTIMSS, PIRLS, ePIRLS, PIRLS Literacy in RLII. Privzeto (`FALSE`) uporabi »polno« zasnovo pri izračunu komponent variance in standardnih napak ocen PV.
- `graphs` – vrednost, ki določa, ali naj se ustvarijo grafikoni. Privzeto: `FALSE`.
- `perc.x.label` – niz, ki določa prilagojeno oznako za horizontalno os v grafikonih odstotkov. Prezrto, če je `graphs = FALSE`.
- `perc.y.label` – niz, ki določa prilagojeno oznako za vertikalno os v grafikonih odstotkov. Prezrto, če je `graphs = FALSE`.
- `prctl.x.labels` – seznam nizov, ki določa prilagojene oznake za horizontalno os v grafikonih percentilov. Prezrto, če je `graphs = FALSE`.
- `prctl.y.labels` – seznam nizov, ki določa prilagojene oznake za vertikalno os v grafikonih percentilov. Prezrto, če je `graphs = FALSE`.
- `output.file` – polna pot do izhodne/output datoteke, vključno z imenom datoteke. Če je izpuščeno, bo datoteka z privzetim imenom »Analysis.xlsx« zapisana v delovni imenik (`getwd()`).
- `open.output` – vrednost, ki določa, ali naj se izhod odpre po zapisu. Privzeto (`TRUE`) odpre izhod v privzetem programu za preglednice, nameščenem na računalniku.

Opombe:

1. Podana morata biti `data.file` ali data.object kot vir podatkov. Če sta podana oba, se bo funkcija ustavila z napako.
2. Funkcija izračuna percentile spremenljivk (ozadnjskih/kontekstualnih spremenljivk ali nizov verjetnostnih vrednosti) po skupinah, določenih z eno ali več kategorijskimi spremenljivkami (razdelitvenimi spremenljivke). Možno je dodati več razdelitvenih spremenljivk, funkcija pa bo izračunala odstotke za vse nastale skupine in njihove percentile na zveznih spremenljivkah. Če razdelitvene spremenljivke niso dodane, bodo rezultati izračunani samo po državah.
3. Možno je podati več zveznih ozadnih spremenljivk, za katere se izračunajo specifični percentili. Upoštevajte, da se rezultati nekoliko razlikujejo v primerjavi z uporabo istih ozadnih spremenljivk v ločenih analizah. To je posledica dejstva, da se primeri z manjkajočimi vrednostmi pri bckg.prctls.vars odstranijo vnaprej, in več kot je spremenljivk, več primerov bo verjetno odstranjenih.
4. Izračun percentilov, ki vključuje verjetnostne vrednosti (PV), zahteva, da se v `PV.root.prctls` poda koren imen verjetnostnih vrednosti. Vse raziskave (razen CivED, TEDS-M, SITES, TALIS in TALIS Starting Strong Survey) imajo nize PV za vsak konstrukt (npr., pri TIMSS pet za celotno matematiko, pet za algebro, pet za geometrijo itd.). V nekaterih raziskavah, kot sta TIMSS in PIRLS, se imena PV v nizu vedno začnejo z znakovnim nizom in končajo z zaporedno številko PV. Npr., imena PV za celotno matematiko pri TIMSS so BSMMAT01, BSMMAT02, BSMMAT03, BSMMAT04 in BSMMAT05. Koren imena PV, ki ga je treba dodati v `PV.root.prctls`, je torej »BSMMAT«. Funkcija bo samodejno našla vse spremenljivke v tem nizu PV in jih vključila v analizo. V drugih raziskavah, kot so OECD PISA in IEA ICCS ter ICILS, se zaporedna številka PV nahaja sredi imena. Npr., pri ICCS so imena PV PV1CIV, PV2CIV, PV3CIV, PV4CIV in PV5CIV. Koren imena PV mora biti določen kot »PV#CIV«. Dodati je mogoče več nizov PV. Vendar pa bo hkratno dodajanje zveznih spremenljivk v argument `bckg.prctls.vars` in korena PV v `PV.root.prctls` vplivalo na rezultate PV, saj bodo primeri z manjkajočimi vrednostmi pri `bckg.prctls.vars` odstranjeni, kar bo vplivalo tudi na rezultate PV. Po drugi strani pa uporaba več kot enega niza PV hkrati ne bi smela vplivati na rezultate ocen PV, saj PV ne smejo imeti manjkajočih vrednosti.
5. Če je `include.missing = FALSE` (privzeto), bodo odstranjeni vsi primeri z manjkajočimi vrednostmi pri razdelitvenih spremenljivkah, obdržani pa bodo samo primeri z veljavnimi vrednostmi v statistiki. Upoštevajte, da se podatki iz raziskav lahko izvozijo na dva različna načina z uporabo `lsa.convert.data`: (1) z nastavitvijo vseh uporabniško določenih manjkajočih vrednosti na `NA`; (2) z uvozom vseh uporabniško določenih manjkajočih vrednosti kot veljavnih ter dodajanjem njihovih kod v dodatni atribut vsaki spremenljivki. Če je include.missing v `lsa.prctls` nastavljeno na `FALSE` (privzeto) in so podatki izvoženi z možnostjo (2), bo output odstranil vse vrednosti iz spremenljivke, ki se ujemajo z vrednostmi v atributu missings. V nasprotnem primeru jih bo funkcija vključila kot veljavne vrednosti in zanje izračunala statistike.
6. Argument `shortcut` je veljaven samo za TIMSS, TIMSS Advanced, TIMSS Numeracy, PIRLS, ePIRLS, PIRLS Literacy in RLII. Prej so te raziskave za izračun standardnih napak uporabljale 75 replikacij, ker je imela ena izmed šol v 75 območjih JK podvojene uteži, druga pa je bila izključena. Od TIMSS 2015 in PIRLS 2016 naprej te raziskave uporabljajo 150 replikatov, in sicer tako, da enkrat podvojijo uteži za šolo in enkrat izpustijo šolo v vsakem območju JK, kar pomeni, da se izračuni izvedejo

dvakrat za vsako območje. Za več podrobnosti glejte tehnično dokumentacijo in uporabniške priročnike TIMSS 2015 in PIRLS 2016. Če je potrebna ponovitev tabel in grafikonov, je treba argument `shortcut` spremeniti na `TRUE`.

7. Če je `graphs = TRUE`, funkcija ustvari grafikone: stolpčne grafikone za odstotke in črtne grafikone za percentile po skupinah, določenih s `split.vars`. Vsi grafikoni so ustvarjeni po državah. Če na koncu ni določenih razdelitvenih spremenljivk, bodo prikazani odstotki in grafikoni napak za vse države skupaj. Po potrebi je mogoče določiti prilagojene oznake za horizontalno in vertikalno os v stolpčnih ter črtnih grafikonih z uporabo argumentov `perc.x.label`, `perc.y.label`, `prctl.x.labels` in `prctl.y.labels`.

Če je `save.output = FALSE`, bo funkcija vrnila seznam z ocenami in informacijami o analizi. Če je `graphs = TRUE`, bodo grafikoni dodani na seznam ocen.

Če je `save.output = TRUE` (privzeto), bo datoteka MS Excel (.xlsx) shranjena, kot je določeno s polno potjo v argumentu `output.file`. Če argument manjka, bo Excelova datoteka z generičnim imenom »Analysis.xlsx« shranjena v delovni imenik (`getwd()`). Datoteka vsebuje tri preglednice. Prva (»Estimates«) vsebuje tabelo z rezultati po državah, zadnji del tabele pa vsebuje povprečne rezultate statistike vseh držav. Naslednji stolpci so lahko v tabeli, odvisno od specifikacije analize:

- **<Country ID>** – stolpec, ki vsebuje imena držav v datoteki, za katere so izračunane statistike. Natančen naslov stolpca bo odvisen od identifikatorja držav, uporabljenega v določeni raziskavi.
- **<Split variable 1>**, **<Split variable 2>**… – stolpci, ki vsebujejo kategorije, po katerih so bile statistike razdeljene. Natančna imena bodo odvisna od spremenljivk v `split.vars`.
- **n_Cases** – število primerov v vzorcu, uporabljenih za izračun statistike.
- **Sum_<Weight variable>** – ocenjeno število elementov v populaciji na skupino po uporabi uteži. Dejanski naziv utežne spremenljivke bo odvisen od spremenljivke, uporabljene v analizi.
- **Sum_<Weight variable>_SE** – standardna napaka ocene števila elementov v populaciji na skupino. Dejanski naziv utežne spremenljivke bo odvisen od spremenljivke, uporabljene v analizi.
- **Percentages_<Last split variable>** – odstotki anketirancev (populacijske ocene) po skupinah, definiranih z razdelitvenimi spremenljivkami v `split.vars`. Odstotki bodo za zadnjo razdelitveno spremenljivko, ki definira končne skupine.
- **Percentages_<Last split variable>_SE** – standardne napake zgoraj navedenih odstotkov.
- **Prctl_<Percentile value>_<Background variable>** – percentil za zveznono spremenljivko <Background variable>, določeno v `bckg.prctls.vars`. Za vsako oceno percentila za vsako spremenljivko, določeno v `bckg.prctls.vars`, bo en stolpec.
- **Prctl_<Percentile value>_<Background variable>_SE** – standardna napaka za percentil zvezne spremenljivke <Background variable>, določene v `bckg.prctls.vars`. Za vsako oceno percentila za vsako spremenljivko, določeno v `bckg.prctls.vars`, bo en stolpec s standardno napako.

- **Percent_Missings_<Background variable>** – odstotek manjkajočih vrednosti za <Background variable>, določeno v `bckg.prctls.vars`. Za vsako spremenljivko, določeno v `bckg.prctls.vars`, bo en stolpec z odstotkom manjkajočih vrednosti.
- **Prctl_<Percentile value>_<root PV>** – percentil za PV z enakim <root PV>, določenim v `PV.root.prctls`. Za vsako oceno percentila za vsak niz PV, določen v `PV.root.prctls`, bo en stolpec.
- **Prctl_<Percentile value>_<root PV>_SE** – standardna napaka na percentil za vsak niz PV z enakim <root PV>, določenim v `PV.root.prctls`. Za vsak niz PV bo en stolpec s standardno napako ocene na percentil.
- **Prctl_<Percentile value>_<root PV>_SVR** – komponenta vzorčne variance za vsak percentil na niz PV z enakim <root PV>, določenim v `PV.root.prctls`. Za vsak niz PV bo en stolpec z oceno vzorčne variance za percentil.
- **Prctl_<Percentile value>_<root PV>_MVR** – komponenta merske variance na percentil za vsak niz PV z enakim <root PV>, določenim v `PV.root.prctls`. Za vsak niz PV bo en stolpec z oceno merske variance na percentil.
- **Percent_Missings_<root PV>** – odstotek manjkajočih vrednosti za <root PV>, določen v `PV.root.prctls`. Za vsak niz PV bo en stolpec z odstotkom manjkajočih vrednosti.

Drugi list (»Analiza informacij«) vsebuje dodatne informacije, povezane z analizo za vsako državo, v naslednjih stolpcih:

- **DATA** – uporabljena spremenljivka `data.file` ali `data.object`.
- **STUDY** – iz katere raziskave prihajajo podatki.
- **CYCLE** – iz katerega cikla raziskave prihajajo podatki.
- **WEIGHT** – katera utežna spremenljivka je bila uporabljena.
- **DESIGN** – katera tehnika ponovnega vzorčenja je bila uporabljena (JRR ali BRR).
- **SHORTCUT** – ali je bila uporabljena metoda bližnjice.
- **NREPS** – koliko uteži za replikacijo je bilo uporabljenih.
- **ANALYSIS_DATE** – datum, na katerega je bila analiza izvedena.
- **START_TIME** – ura začetka analize.
- **END_TIME** – ura zaključka analize.
- **DURATION** – kako dolgo je trajala analiza (v urah, minutah, sekundah in milisekundah).

Tretji list (»Calling syntax«) vsebuje klic funkcije z vrednostmi vseh parametrov, kot je bil izvršen. To je uporabno, če je treba analizo kasneje ponoviti.

Če je `graphs = TRUE`, bo dodan še dodaten list »Grafi«, ki vsebuje vse grafe.

Če se pojavijo kakršna koli opozorila zaradi izračunov, bodo ta vključena na dodatnem listu »Warnings« v delovni knjigi.

4.2.3 Računanje percentilov z uporabo ukazne vrstice

V spodnjem primeru združimo nov podatkovni niz s podatki o učencih in ravnateljih iz PIRLS 2016 (Avstralija in Slovenija), pri čemer vzamemo vse spremenljivke iz obeh vrst datotek:

```
lsa.merge.data(inp.folder = "C:/temp",
               file.types = list(acg = NULL, asg = NULL),
               ISO = c("aus", "svn"),
               out.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData")
```

Za začetek izračunajmo 5., 25., 50., 75. in 95. percentil kompleksne lestvice o tem, koliko učenci radi berejo (ASBGSLR; za informacije o tem, kako je ta lestvica sestavljena in kakšne so njene lastnosti ,preverite tehnično dokumentacijo PIRLS 2016), za dekleta in fante v Avstraliji in Sloveniji:

```
lsa.prctls(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
           split.vars = "ASBG01", bckg.prctls.vars = "ASBGSLR",
           output.file =
"C:/temp/Results/PIRLS_2016_Percentiles_by_Student_Sex.xlsx")
```

Opombe:

1. Obstajata dve spremenljivki, ki vsebujeta informacije o spolu učencev:
   - ITSEX, spremenljivka z informacijami o spremljanju učencev, in
   - ASBG01 (uporabljena v tej analizi), spremenljivka o spolu učencev, kot so ga navedli v vprašalniku.

2. V mednarodnih raziskavah je treba vse analize izvajati ločeno po državah. Vendar ni potrebe po tem, da bi se dodala spremenljivka ID države (IDCNTRY ali CNT v PISA), saj jo bo funkcija samodejno prepoznala in dodala v vektor `split.vars`.

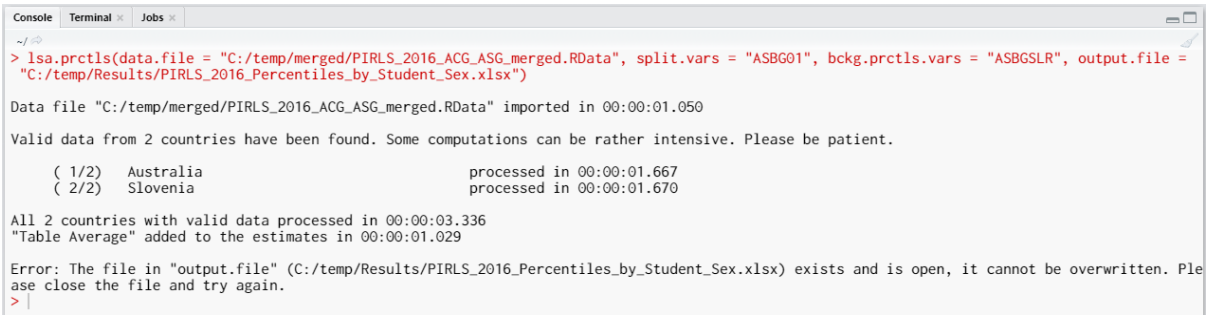
3. Ni potrebe po tem, da se eksplicitno določi utežna spremenljivka. Če utežna spremenljivka ni izrecno določena, bo privzeta utež (v tem primeru skupna utež učencev) uporabljena glede na združene podatke anketirancev, ki jih funkcija samodejno prepozna. Če imate dober razlog za spremembo utežne spremenljivke, lahko to storite z dodajanjem argumenta `weight.var = "SENWGT"`.

4. Vrednosti za argument prctls niso bile določene. V tem primeru bo funkcija samodejno dodala privzete percentile kot `prctls = c(5, 25, 50, 75, 95)`. Če potrebujete drugačne točke v distribuciji, jih preprosto dodajte in upoštevajte pravila za ta argument.

5. Če izhodna datoteka ni določena, se bo output shranil pod imenom »Analysis.xlsx« v delovnem imeniku (prikličete ga lahko z `getwd()`).

6. Če se v klicno sintakso eksplicitno ne doda `open.output = FALSE`, se bo izhodna datoteka odprla po zaključku vseh izračunov. To je uporabno, ko se izvede več klicnih sintaks za različne analize in ni potrebna takojšnja preglednost izhoda.

Izvajanje zgornje kode bo v konzoli RStudia izpisalo naslednji izhod:



Ko so vse operacije zaključene, se izhod shrani na disk kot Excelova delovna knjiga. Če je argument `open.output = TRUE` (privzeto), se datoteka odpre v privzetem pregledniškem programu (običajno MS Excel).

Funkcija lahko izračuna tudi povprečje za niz (ali nize) verjetnostnih vrednosti (angl. *plausible values*). Poglejmo, kako izračunamo 25., 50. in 75. percentil verjetnostnih vrednosti za skupni dosežek branja za dekleta in fante. To omogoča naslednja sintaksa:

```
lsa.prctls(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
           split.vars = "ASBG01", PV.root.prctls = "ASRREA", prctls
= c(25, 50, 75),
           output.file =
"C:/temp/Results/PIRLS_2016_Percentiles_by_Student_Sex.xlsx")
```

Bodite pozorni na način določanja PV. Pet PV za skupni dosežek branja: ASRREA01, ASRREA02, ASRREA03, ASRREA04 in ASRREA05. V argumentu `PV.root.avg` moramo določiti le osnovo PV, torej »ASRREA«. Funkcija bo uporabila to skupno ime, da izbere vseh pet PV in jih vključi v izračune.

Izvajanje zgornje sintakse bo prepisalo prejšnji output, ker ima isto ime datoteke (v konzoli bo prikazano opozorilo). Stolpci na listu Estimates bodo zdaj drugačni.

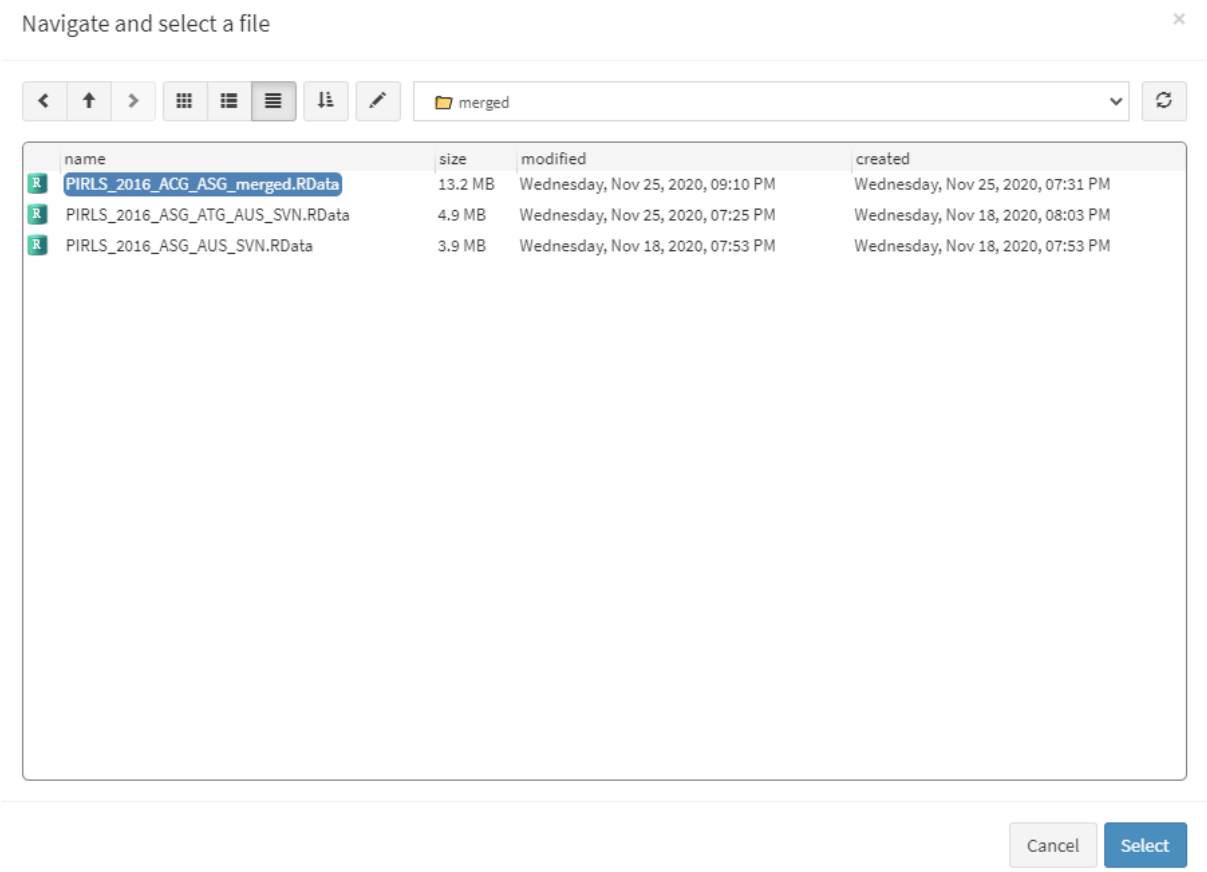
4.2.4 Računanje percentilov z uporabo GUI

Za zagon uporabniškega vmesnika RALSA v RStudiu izvedite naslednji ukaz:

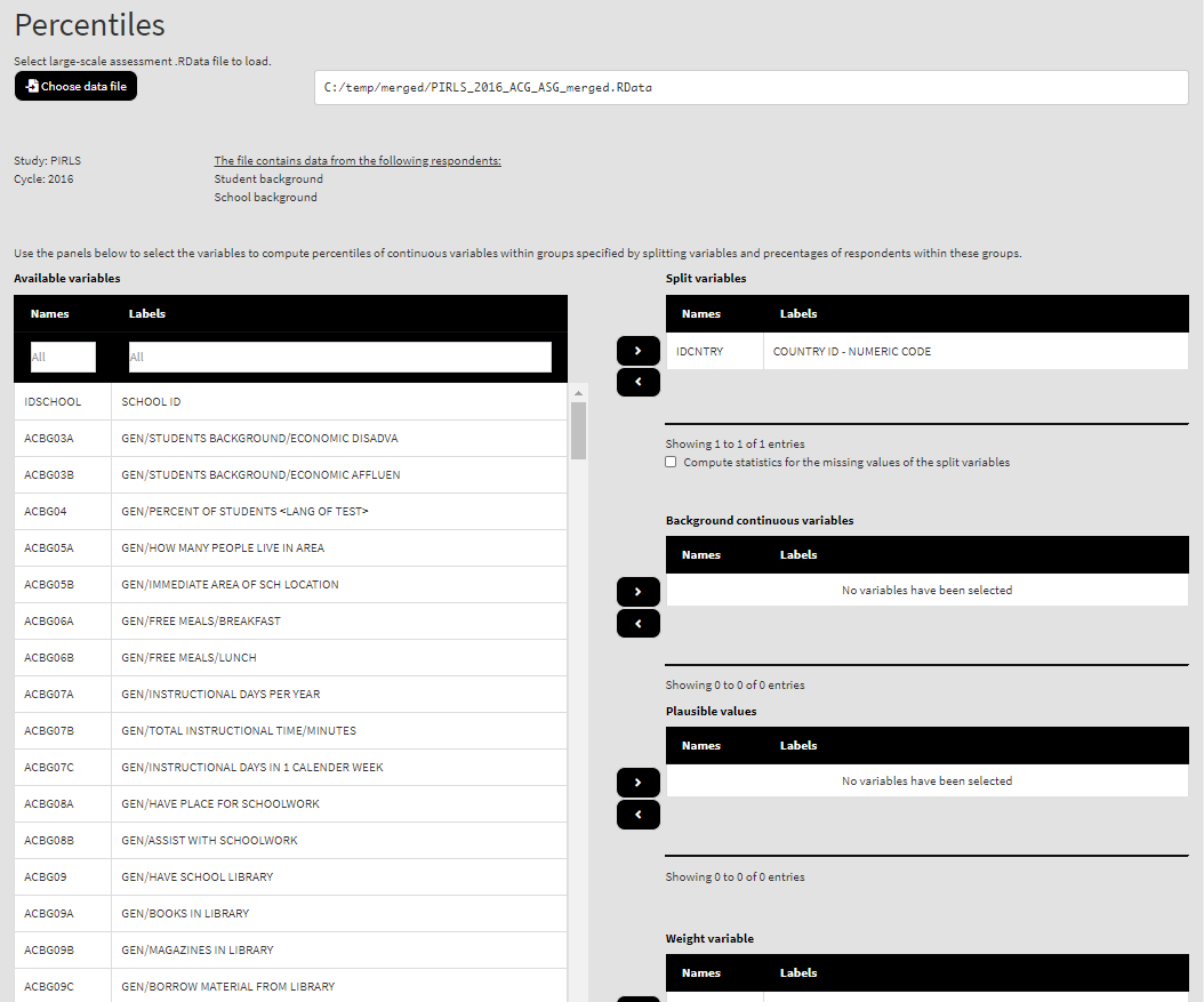
```
ralsaGUI()
```

Za naslednje primere združite novo datoteko s podatki PIRLS 2016 za Avstralijo in Slovenijo, pri čemer vzamete vse spremenljivke o učencih in ravnateljih. Združeno datoteko lahko poimenujete **»PIRLS_2016_ACG_ASG_merged.RData«**.

Ko je združevanje podatkov končano, v meniju na levi izberite **Analysis types > Percentiles**. Ko ste v GUI navigirani do možnosti **Percentiles**, kliknite gumb **Choose data file**. Pojdite do mape, ki vsebuje združeno datoteko **»PIRLS_2016_ACG_ASG_merged.RData«**, jo izberite in kliknite gumb **Select**.



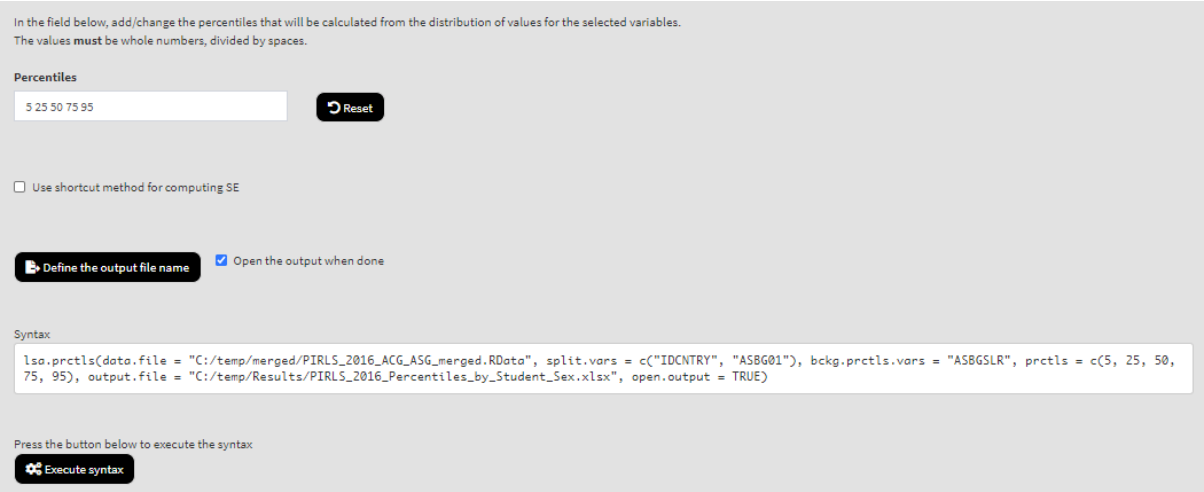
Ko je datoteka naložena, se vam prikaže plošča na levi (razpoložljive spremenljivke) in nabor plošč na desni, kamor lahko dodajate spremenljivke s seznama razpoložljivih.



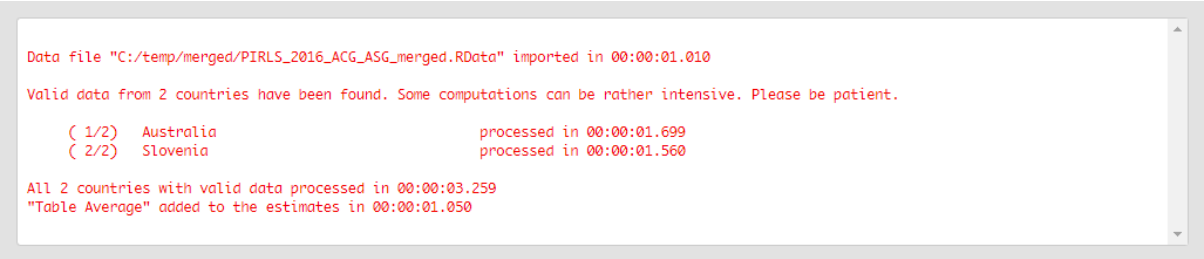
Uporabite miško, da izberete spremenljivke s seznama **Available variables** (Spremenljivke na voljo) in jih s puščičnimi gumbi na sredini zaslona dodate v različna polja (ali odstranite) za nastavitev analize. Uporabite lahko filtre na vrhu panelov, da hitro najdete potrebne spremenljivke.

Za začetek izračunajmo 5., 25., 50., 75. in 95. percentil kompleksne lestvice, ki meri, koliko učenci radi berejo (ASBGSLR; za informacije o tem, kako je ta lestvica sestavljena in kakšne so njene lastnosti, preverite tehnično dokumentacijo PIRLS 2016) za deklice in fante v Avstraliji in Sloveniji. Dodajte spremenljivko ASBG01 na seznam **Split variables** (Spremenljivke za delitev). Poiščite spremenljivko ASBGSLR na seznamu **Available variables** in uporabite desno puščno tipko, da jo dodate v panel **Background continuous variables** (Ozadnjske zvezne spremenljivke). To je vse, kar morate narediti. Pod paneli boste videli besedilno polje s privzetimi percentilnimi vrednostmi (5, 25, 50, 75 in 95). Te vrednosti lahko spremenite v katere koli druge, v skladu z navodili nad poljem, vendar jih za zdaj pustimo takšne, kot so. Poleg polja boste videli gumb **Reset** (Ponastavi). Če ste spremenili vrednosti v polju in želite obnoviti privzete vrednosti, preprosto kliknite nanj. Pomaknite se navzdol in kliknite **Define output file name** (Določite ime output datoteke). Pojdite v mapo **»C:/temp/Results«** (ali v mapo, kjer želite shraniti output) in določite ime izhodne datoteke. Ko to storite, se bo ob Define the output file name (Določitev imena izhodne datoteke) pojavilo potrditveno polje. Če je potrjeno, se bo output odprl po zaključku vseh izračunov. Pod tem bo

prikazana sintaksa. Pod vsemi temi bo prikazan gumb **Execute syntax** (Izvedi sintakso). Končne nastavitve v spodnjem delu zaslona bi morale izgledati tako:



Kliknite gumb **Execute syntax** (Izvedi sintakso). V spodnjem delu se bo pojavila konzola GUI, ki bo beležila vse zaključene operacije.

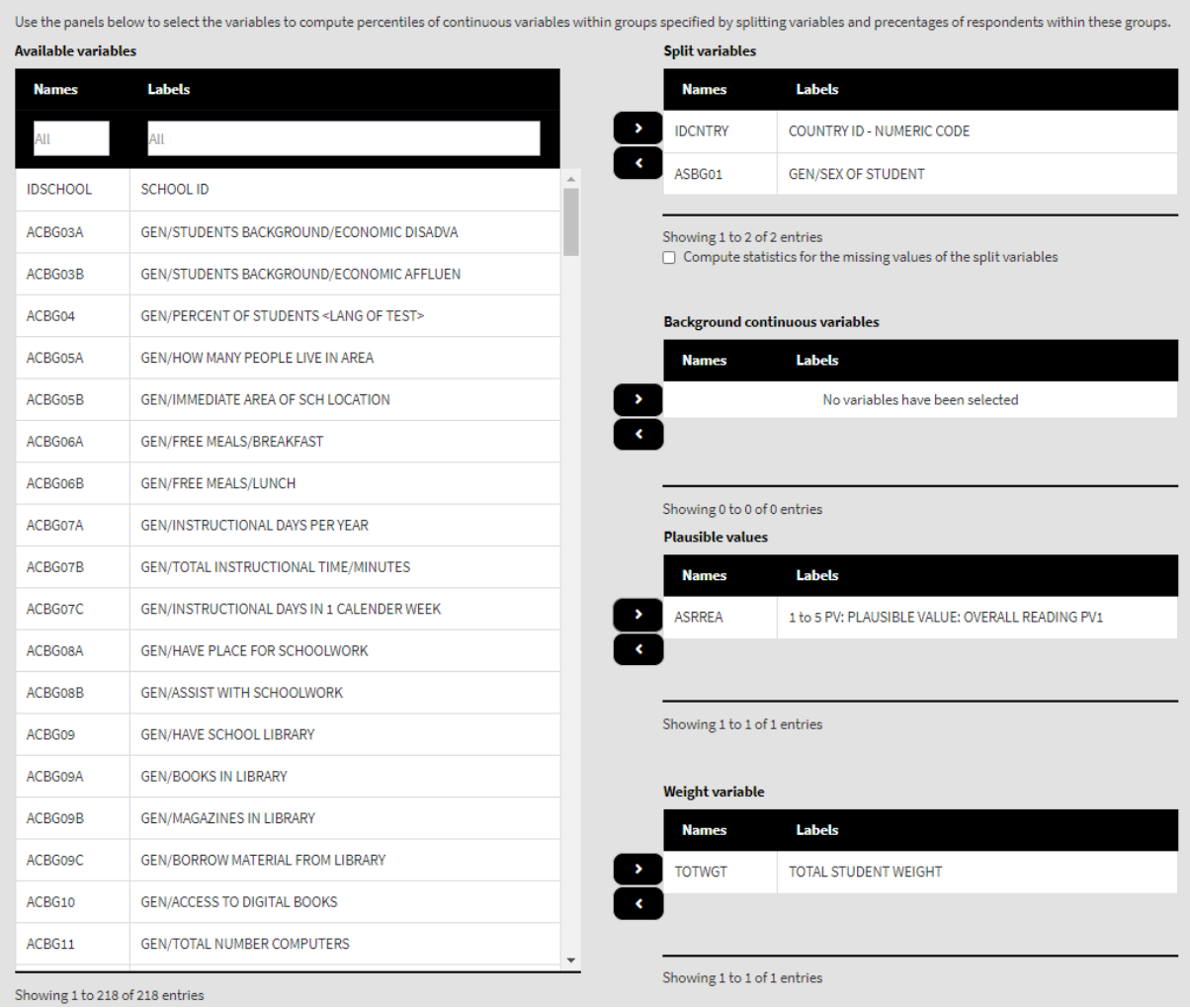


Nekaj stvari, ki jih je treba upoštevati:

1. Obstajata dve spremenljivki, ki vsebujeta informacije o spolu učencev:
   - ITSEX, spremenljivka, ki vsebuje informacije o sledenju učencem, ter
   - ASBG01 (uporabljena v tej analizi), spremenljivka, ki vsebuje spol učencev, kot so ga navedli v vprašalniku.
2. Pri mednarodnih raziskavah je treba vse analize izvajati ločeno po državah. Spremenljivka ID države (IDCNTRY, ali CNT v PISI) je vedno izbrana kot prva razdelitvena spremenljivka in je ne morete odstraniti s panela razdelitvenih spremenljivk.
3. Privzeta utež je izbrana in dodana samodejno na panelu za spremenljivke uteži. Lahko jo spremenite z drugo utežjo, ki je na voljo v naboru podatkov. Če je izbrana privzeta utež, ne bo prikazana v oknu sintakse. Če na panelu spremenljivk uteži ni izbrana nobena utež, se bo privzeta uporabljala samodejno.
4. Polje Use shortcut method for computing SE (Uporabi bližnjico za izračun SE) ni privzeto označeno. To pomeni, da bo funkcija izračunala standardni odklon z uporabo »polne« metode za komponento vzorčne variance.

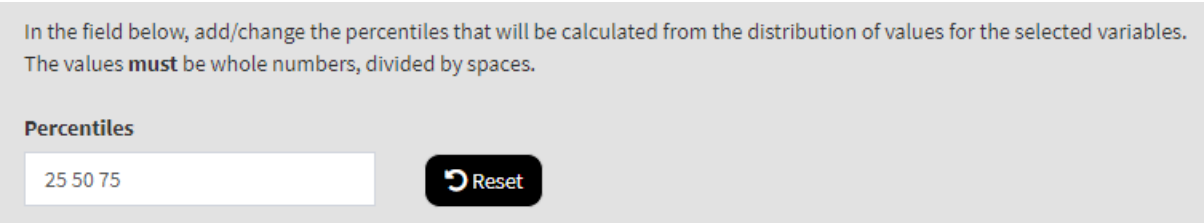
Če je polje **Open the output when done** (Odpri output, ko zaključiš) označeno, se bo output samodejno odprl v privzetem programu za preglednice (običajno MS Excel), ko bodo vse izračuni zaključeni. Glejte razlage o strukturi delovnega zvezka, njegovih listih in stolpcih tukaj.

Funkcija lahko tudi izračuna povprečje za niz (ali nize) verjetnostnih vrednosti. Izračunajmo 25., 50. in 75. percentil verjetnostnih vrednosti za rezultat na bralnem preizkusu za deklice in dečke. Za nadaljevanje odstranite spremenljivko ASBGSLR s panela **Bakground continous variables**. S seznama **Available variables** (Spremenljivke na voljo) poiščite koren splošnih verjetnostnih vrednosti za branje (ASRREA). Za iskanje po imenu ali oznaki lahko uporabite filtre na vrhu panela. Izberite koren in ga z uporabo puščic dodajte na panel **Plausible values** (Verjetne vrednosti). Ta del vmesnika naj bi izgledal tako:

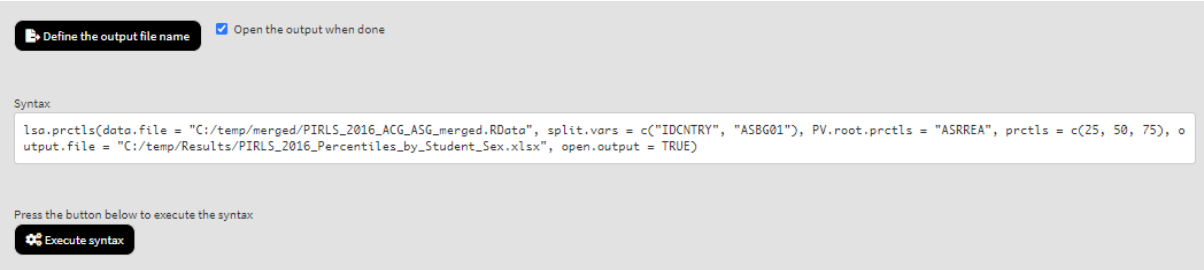


Bodite pozorni na to, kako so določene PV (verjetne vrednosti). Pet PV za rezultat na preizkusu branja: ASRREA01, ASRREA02, ASRREA03, ASRREA04 in ASRREA05. Seznama **Available variables** (Spremenljivke na voljo) in **Plausible values** (Verjetne vrednosti) ne bosta prikazala petih ločenih PV, temveč samo njihov koren/skupno ime – ASRREA, brez številk na koncu. V ozadju bo funkcija vzela vseh pet PV in jih vključila v izračune.

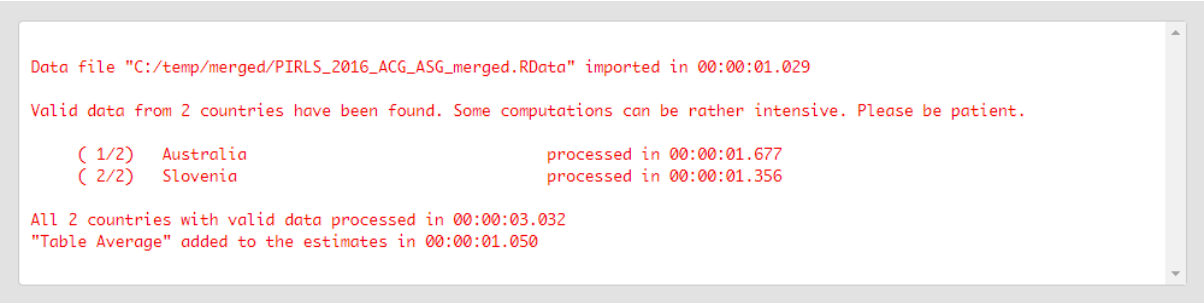
V polju v vmesniku je treba spremeniti vrednosti percentilov, saj nočemo uporabljati privzetih vrednosti. Preprosto kliknite v polje in pustite samo 25, 50 in 75, preostale vrednosti izbrišite. Vrednosti lahko vtipkate ročno. Po spremembi vrednosti naj polje izgleda tako:



Ker aplikacijo **Percentiles** (Percentili) uporabljamo neposredno po izvedbi prejšnje analize, so še vedno prisotne nastavitve iz prejšnje analize. Ni treba spreminjati nobenih preostalih nastavitev, razen če to želite. Lahko pa spremenite ime output datoteke, sicer bo prepisana. Upoštevajte, da se prikazana sintaksa spremeni in odraža vključitev korena petih PV, ASRREA, za izračun povprečja na preizkusu iz branja:



Kliknite gumb **Execute syntax** (Izvedi sintakso). Konzola GUI se bo posodobila in zabeležila vse zaključene operacije:



Ko bodo vsi izračuni zaključeni, se bo output samodejno odprl.

4.3 Deleži respondentov, ki so dosegli ali presegli mejnike (angl. *benchmarks*)

4.3.1 Uvod

Funkcija `lsa.bench` izračuna deleže respondetov v populaciji, ki dosegajo ali presegajo določene ravni uspešnosti. Te se imenujejo »benchmarki« oz. mejniki. Mejniki se lahko določijo le za verjetne vrednosti (angl. *plausible values*). Deleže anketirancev v populaciji

lahko izračunamo znotraj skupin, določenih s kategorijami spremenljivk za deljenje. Spremenljivke za deljenje so neobvezne. Če teh spremenljivk ni, se rezultati izračunajo le na ravni države. Če so te spremenljivke zagotovljene, se bodo podatki znotraj vsake države razdelili v skupine glede na vse spremenljivke za deljenje. Deleži anketirancev v populaciji, ki dosežejo ali presegajo določen mejnik, pa se bodo izračunali za zadnjo spremenljivko za deljenje. Upoštevajte, da se mejniki lahko izračunajo le za en komplet verjetnostnih vrednosti naenkrat, ob upoštevanju kompleksnega vzorčenja in ocenjevanja v raziskavi. Deleži anketirancev bodo izračunani za vsako verjetno vrednost v kompletu, nato pa se bodo ocene za vse verjetne vrednosti v kompletu povprečile, standardna napaka pa se bo izračunala z uporabo kompleksnih formul, ki bodo odvisne od raziskave. Standardna napaka bo izračunana ob upoštevanju kompleksnega vzorčenja in načrtovanja raziskav. Če vas zanimajo podrobnejši podatki o kompleksnem vzorčenju in načrtovanju ocenjevanja določene raziskave ter o tem, kako se izračunajo ocene in njihove standardne napake, si oglejte tehnično dokumentacijo in uporabniški priročnik.

Kot katera koli druga funkcija v paketu RALSA funkcija `lsa.bench` prepozna podatke raziskave in uporabi pravilne tehnike ocenjevanja glede na izvedbo vzorčenja ter načrtovanja ocenjevanja raziskave.

4.3.2 Funkcija benchmarks in njeni argumenti

Funkcija `lsa.bench` ima naslednje argumente:

- `data.file` – datoteka, ki vsebuje objekt `lsa.data`. Določena mora biti ta ali `data.object`, vendar ne oba.
- `data.object` – objekt v pomnilniku, ki vsebuje objekt `lsa.data`. Določen mora biti ta ali `data.file`, vendar ne oba.
- `split.vars` – kategorijska spremenljivka(-e) za deljenje rezultatov. Če spremenljivk za deljenje ni, bodo rezultati za celotno populacijo držav. Če je določena ena ali več spremenljivk, se bodo rezultati razdelili glede na vse razen zadnje spremenljivke, deleži anketirancev pa bodo izračunani glede na edinstvene vrednosti zadnje spremenljivke za deljenje.
- `PV.root.bench` – osnovna imena za komplet(-e) verjetnostnih vrednosti, ki bodo uporabljena za izračun deležev anketirancev, ki dosegajo ali presegajo določeno mejno točko.
- `bench.vals` – vektor celih števil, ki predstavlja mejne točke.
- `bench.type` – niz znakov, ki predstavlja način izračuna deležev anketirancev.
- `pcts.within` – določa, ali naj se deleži izračunajo znotraj skupin, določenih s `split.vars` (`TRUE`) ali ne (`FALSE`, privzeto).
- `bckg.var` – ime zvezne kontekstualne spremenljivke (angl. *background variable*) za izračun povprečja. Rezultati bodo izračunani za vse skupine, določene s spremenljivkami za deljenje, in za skupino uspešnosti (angl. *performance group*).
- `weight.var` – ime spremenljivke, ki vsebuje uteži. Če ime spremenljivke uteži ni določeno, bo funkcija samodejno izbrala privzeto utež za dane podatke, odvisno od vrste anketirancev.

- include.missing – določitev, ali naj se manjkajoče vrednosti spremenljivk za deljenje vključijo kot kategorije za deljenje in naj se za njih izračunajo vse statistike. Privzeto (FALSE) upošteva vse primere spremenljivk za deljenje brez manjkajočih vrednosti pred izračunom statistike.
- shortcut – določitev, ali naj se uporabi metoda »skrajšave« za IEA TIMSS, TIMSS Advanced, TIMSS Numeracy, eTIMSS, PIRLS, ePIRLS, PIRLS Literacy in RLII. Privzeto (FALSE) uporablja celoten načrt pri izračunavanju komponent variance in standardnih napak ocen.
- graphs – določitev, ali naj se izdelajo grafi? Privzeto je FALSE.
- perc.x.label – niz, prilagojen oznaki za horizontalno os v grafih z odstotki. Ignorirano, če je graphs = FALSE.
- perc.y.label – Niz, prilagojen oznaki za vertikalno os v grafih s procenti. Ignorirano, če je graphs = FALSE.
- mean.x.label – seznam nizov, prilagojenih oznakam za horizontalno os v grafih s povprečji. Ignorirano, če je bckg.var izpuščen in/ali graphs = FALSE.
- mean.y.label – seznam nizov, prilagojenih oznakam za vertikalno os v grafih s povprečji. Ignorirano, če je bckg.var izpuščen in/ali graphs = FALSE.
- save.output – določitev, ali naj se output shrani v MS Excel-datoteko (privzeto) ali ne (izpisano na konzolo ali dodeljeno objektu).
- output.file – polna pot do output datoteke, vključno z imenom datoteke. Če je izpuščena, bo datoteka s privzetim imenom »Analysis.xlsx« shranjena v delovni imenik (getwd()).
- open.output – določitev, ali naj se output odpre po tem, ko je bil zapisan. Privzeto (TRUE) odpre output v privzetem programu za preglednice, nameščenem na računalniku.

Opombe:

1. Podana morata biti bodisi data.file bodisi data.object kot vir podatkov. Če sta podana oba, bo funkcija prenehala delovati in prikazala sporočilo o napaki.
2. Funkcija izračuna odstotke respondentov, ki dosegajo ali presegajo določene mejne vrednosti (benchmarki/performance ravni). Ti odstotki se izračunajo z uporabo niza PV (verjetnostnih vrednosti), ki je določen v PV.root.bench. Naenkrat je mogoče dodati samo en niz PV v PV.root.bench. Vse raziskave (razen CivED, TEDS-M, SITES, TALIS in TALIS Starting Strong Survey) imajo niz PV za vsako vsebinsko domeno (npr., v TIMSS pet za splošno matematiko, pet za algebro, pet za geometrijo itd.) in kognitivno domeno (tj. poznavanje, uporaba in razmišljanje). V nekaterih raziskavah (npr. TIMSS in PIRLS) se imena PV v nizu vedno začnejo z nizom znakov in končajo z zaporedno številko PV. Npr., imena spremenljivk PV v nizu za splošno matematiko v TIMSS: BSMMAT01, BSMMAT02, BSMMAT03, BSMMAT04 in BSMMAT05. Korenski naziv PV za ta niz, ki ga je treba dodati v PV.root.avg, bo »BSMMAT«. Funkcija bo samodejno poiskala vse spremenljivke v tem nizu PV in jih vključila v analizo. V drugih raziskavah, kot so OECD PISA in IEA ICCS ter ICILS, je zaporedna številka vsake verjetne vrednosti (PV) vključena v sredino imena. Npr., v ICCS so

imena niza PV: PV1CIV, PV2CIV, PV3CIV, PV4CIV in PV5CIV. Korensko ime PV je treba določiti v `PV.root.bench` kot »PV#CIV«.

3. Več razdelitvenih spremenljivk se lahko doda v `split.vars`. Funkcija bo izračunala odstotke respondentov, ki dosegajo ali presegajo mejne vrednosti za vse oblikovane skupine, in njihove povprečne vrednosti na zveznih spremenljivkah. Če niso dodane nobene razdelitvene spremenljivke, bodo rezultati le za državo.
4. Če je zvezna kontekstualna/ozadenjska spremenljivka podana v `bckg.var`, bo povprečje te spremenljivke izračunano za vsako skupino, ki jo oblikujejo razdelitvene spremenljivke in skupine uspešnosti. V analizo je mogoče dodati samo eno kontekstualno/ozadno spremenljivko. Ta argument je ignoriran, če je `bench.type = "cumulative"`.
5. Mejne vrednosti so podane kot vektor celih števil (npr. `c(475, 500)`) v `bench.vals`. Če mejne vrednosti niso podane, bo funkcija samodejno izbrala vse mejne vrednosti za ustrezno raziskavo in, v nekaterih primerih, za podatke iz specifičnega cikla. To velja za ICCS in PISA, kjer se zahtevnostne ravni (angl. *proficiency levels*) razlikujejo od cikla do cikla.
6. Argument `bench.type` ima dve različni možnosti: »discrete« (privzeto) in »cumulative«. Z uporabo prve se izračunajo odstotki respondentov znotraj meja, določenih z mejno vrednostjo v `bench.vals`. Z uporabo druge bo funkcija izračunala odstotke respondentov, ki so na ali nad mejno vrednostjo v `bench.vals`.
7. Če je `pcts.within = FALSE` (privzeto), bo funkcija izračunala odstotke respondentov, ki dosegajo ali presegajo vsako od mejnih vrednosti, določenih z `bench.vals`. V tem primeru se bodo odstotki vseh respondentov po ravneh uspešnosti sešteli do 100 v vsaki skupini, ki jo določajo razdelitvene spremenljivke. In nasprotno: če je `pcts.within = TRUE`, bo funkcija izračunala odstotke respondentov z dano značilnostjo na vsaki ravni uspešnosti, npr., če želimo vedeti, kolikšni odstotki respondentov imajo ali nimajo računalnika doma na vsaki ravni uspešnosti. Ta argument je ignoriran, če je `bench.type = "cumulative"`.
8. Če niso določene nobene spremenljivke za `bckg.vars`, bo output vseboval le odstotke primerov v skupinah, določenih z razdelitvenimi spremenljivkami in mejno vrednostjo.
9. Če je `include.missing = FALSE` (privzeto), bodo vsi primeri z manjkajočimi vrednostmi na razdelitvenih spremenljivkah odstranjeni in bodo obdržani le primeri z veljavnimi vrednostmi v statistiki. Upoštevajte, da so podatki iz raziskav lahko izvoženi na dva različna načina: (1) nastavitev vseh uporabniško določenih manjkajočih vrednosti na `NA`; (2) uvoz vseh uporabniško določenih manjkajočih vrednosti kot veljavnih in dodajanje njihovih kod v dodatni atribut za vsako spremenljivko. Če je `include.missing` nastavljeno na FALSE (privzeto) in so podatki izvoženi z možnostjo (2), bo izhod odstranil vse vrednosti iz spremenljivke, ki ustreza vrednostim v njenem atributu manjkajočih vrednosti. V nasprotnem primeru jih bo vključil kot veljavne vrednosti in zanje izračunal statistiko.
10. Argument `shortcut` velja samo za TIMSS, TIMSS Advanced, TIMSS Numeracy, PIRLS, ePIRLS, PIRLS Literacy in RLII. Prej so te raziskave pri izračunu standardnih napak uporabljale 75 replikacij, ker je bila ena izmed šol v 75 JK območjih s podvojenimi težami, druga pa odstranjena. Od TIMSS 2015 in PIRLS 2016 naprej raziskave uporabljajo 150 replikacij, pri čemer je imela v vsakem JK območju ena šola

imela podvojene teže, druga pa je bila odstranjena, torej se izračuni izvedejo dvakrat za vsako območje. Za več podrobnosti glejte Foy in LaRoche (2016) ter Foy in LaRoche (2017). Za več podrobnosti si oglejte tehnično dokumentacijo in uporabniške priročnike TIMSS 2015 in PIRLS 2016. Če je potrebna replikacija tabel in grafikonov, je treba argument `shortcut` nastaviti na `TRUE`.

11. Če je `graphs = TRUE`, bo funkcija ustvarila grafikon, stolpične diagrame za odstotke in napake za povprečja (če je `bckg.var` podana) na skupinah, določenih s `split.vars`. Vsi grafi bodo izdelani po državah. Če ni določenih `split.vars`, bodo na koncu prikazani odstotki in napake za vse države skupaj. Po potrebi lahko uporabite lastne oznake za horizontalno in vertikalno os za stolpične diagrame ter diagrame napak z argumentoma `perc.x.label`, `perc.y.label`, `mean.x.label` in `mean.y.label`.
12. Če ne določite `save.output = FALSE`, bo output zapisan v MS Excel na disk. V nasprotnem primeru bo output natisnjen na konzolo.
13. Če ni določena nobena pot do output datoteke, bo output shranjen z imenom datoteke »Analysis.xlsx« v delovnem imeniku (kar lahko pridobite z `getwd()`).

Če je `save.output = FALSE`, bo funkcija vrnila seznam, ki vsebuje ocene in informacije o analizi. Če je `graphs = TRUE`, bodo grafi dodani seznamu ocen.

Če je `save.output = TRUE` (privzeto), bo ustvarjena datoteka MS Excel (.xlsx), ki jo je mogoče odpreti v katerem koli programu za preglednice, kot je določeno s popolno potjo v `output.file`. Če argument manjka, bo Excelova datoteka shranjena z generičnim imenom datoteke »Analysis.xlsx« v delovnem imeniku (`getwd()`). Delovni zvezek vsebuje tri preglednice. Prva (»Estimates«) vsebuje tabelo z rezultati po državah, končni del tabele pa vsebuje povprečne rezultate iz statistike vseh držav. V tabeli so naslednji stolpci, odvisno od specifikacije analize:

- **<Country ID>** – stolpec, ki vsebuje imena držav v datoteki, za katere so bile izračunane statistike. Natančen naslov stolpca bo odvisen od identifikatorja države, uporabljenega v posamezni raziskavi.
- **<Split variable 1>**, **<Split variable 2>**… – stolpci, ki vsebujejo kategorije, po katerih so bile statistike razdeljene. Natančna imena bodo odvisna od spremenljivk v `split.vars`.
- **n_Cases** – število primerov, ki dosegajo ali presegajo vsako od mejnih vrednosti z uporabo niza PV. Upoštevajte, da to morda niso celotna števila, saj se izračunajo z uporabo vsakega PV in nato povprečijo.
- **Sum_<Weight variable>** – ocenjena populacija elementov na skupino po uporabi uteži.
- **Sum_<Weight variable>_SE** – standardna napaka ocenjenega števila elementov na skupino.
- **Performance_Group** – oznake za skupine uspešnosti, določene z `bench.vals`.
- **Percentages_<PVs' root name>** – odstotki respondentov (ocenjene populacije), ki dosegajo ali presegajo vsako mejno vrednost (v primeru `bench.type = "discrete"`), ali odstotki respondentov (ocenjene populacije), ki so na ali nad vsako mejno vrednostjo (v primeru `bench.type`

= "cumulative") po skupinah, določenih z razdelitvenimi spremenljivkami v split.vars.

- **Percentages_<PVs' root name>_SE** – standardne napake odstotkov zgoraj.
- **Mean_<Background variable>** – povprečje ozadenjske spremenljivke <Background variable>, določene v bckg.var.
- **Mean_<Background variable>_SE** – standardna napaka povprečja ozadenjske spremenljivke <Background variable>, določene v bckg.var.
- **Variance_<Background variable>** – varianca za ozadensjko spremenljivko <Background variable>, določeno v bckg.var.
- **Variance_<Background variable>_SE** – napaka variance za ozadensjko spremenljivko <Background variable>, določeno v bckg.var.
- **SD_<Background variable>** – standardni odklon za ozadensjko spremenljivko <Background variable>, določeno v bckg.vars.
- **SD_<Background variable>_SE** – napaka standardnega odklona za ozadenjsko spremenljivko <Background variable>, določeno v bckg.avg.var.
- **Percent_Missings_<Background variable>** – odstotek manjkajočih vrednosti za ozadenjsko spremenljivko <Background variable>, določeno v bckg.var.

Druga preglednica (»Analysis information«) vsebuje dodatne informacije, povezane z analizo po državah v naslednjih stolpcih:

- **DATA** – uporabljena data.file ali data.object.
- **STUDY** – iz katere raziskave prihajajo podatki.
- **CYCLE** – iz katerega cikla raziskave so podatki.
- **WEIGHT** – katera utež je bila uporabljena.
- **DESIGN** – katera tehnika vzorčenja je bila uporabljena (JRR ali BRR).
- **SHORTCUT** – ali je bila uporabljena metoda bližnjice.
- **NREPS** – koliko replikacijskih uteži je bilo uporabljenih.
- **ANALYSIS_DATE** – na kateri datum je bila izvedena analiza.
- **START_TIME** – ob katerem času se je analiza začela.
- **END_TIME** – ob katerem času se je analiza končala.
- **DURATION** – koliko časa je trajala analiza v urah, minutah, sekundah in milisekundah.

Tretja preglednica (»Calling syntax«) vsebuje klic funkcije z vrednostmi za vse parametre, kot je bila izvedena. To je koristno, če je treba analizo kasneje ponoviti.

Če je graphs = TRUE, bo dodana dodatna preglednica »Graphs«, ki vsebuje vse grafe.

Če se pojavijo kakršne koli opozorila zaradi izračunov, bodo vključena v dodatno preglednico »Warnings« v delovnem zvezku.

4.3.3 Računanje mejnikov z uporabo ukazne vrstice

V naslednjih primerih bomo združili novo podatkovno datoteko s podatki o učencih in ravnateljih šol iz PIRLS 2016 (Avstralija in Slovenija), pri čemer bomo vzeli vse spremenljivke iz obeh vrst datotek:

```
lsa.merge.data(inp.folder = "C:/temp",
               file.types = list(acg = NULL, asg = NULL),
               ISO = c("aus", "svn"),
               out.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData")
```

Za začetek bomo izračunali odstotke deklet in fantov v Avstraliji in Sloveniji, ki dosegajo ali presegajo vse mejne vrednosti skupnega branja:

```
lsa.bench(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
          split.vars = "ASBG01", PV.root.bench = "ASRREA")
```

Opombe:

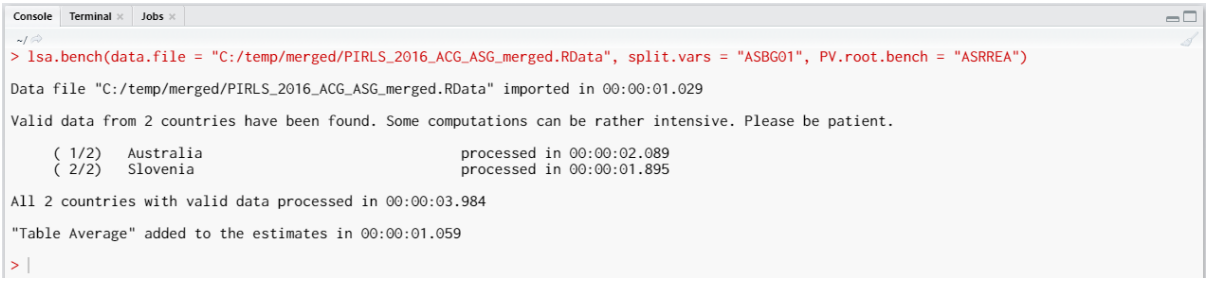
1. Obstajata dve spremenljivki, ki vsebujeta informacije o spolu učencev:
   - ITSEX, spremenljivka, ki vsebuje podatke o spremljanju učencev;
   - ASBG01 (uporabljena v tej analizi), spremenljivka, ki vsebuje spol učencev, kot so ga navedli v vprašalniku.
2. V mednarodnih raziskavah je treba vse analize opraviti ločeno po državah. Vendar pa ni potrebe po dodajanju spremenljivke ID države (IDCNTRY, ali CNT v PISA) kot spremenljivke za delitev. Funkcija jo bo samodejno prepoznala in dodala v vektor `split.vars`.
3. Pet PV za bralne dosežke: ASRREA01, ASRREA02, ASRREA03, ASRREA04, in ASRREA05. V argument `PV.root.bench` moramo navesti le koren PV – »ASRREA«. Funkcija bo uporabila ta koren za izbiro vseh petih PV in jih vključila v izračune.
4. Ni potrebe po natančnem določitvi mejnih vrednosti z argumentom `bench.vals`. Če ga izpustite, bo funkcija samodejno dodala vse mejne vrednosti za raziskavo (in v nekaterih raziskavah celo za ustrezni cikel), za PIRLS glejte spodaj, za vse druge raziskave poiščite informacije v njihovih uporabniških priročnikih in tehnični dokumentaciji.
5. Ni potrebe po izrecnem določanju uteži. Če ni izrecno določena nobena utežna spremenljivka, se bo za podatkovni niz uporabila privzeta utež (v tem primeru skupna utež učencev), ki se prepozna samodejno. Če imate dober razlog za spremembo uteži, to lahko storite z dodajanjem `weight.var = "SENWGT"`.
6. Če ni določena output datoteka, bo izhod shranjen z imenom datoteke »Analysis.xlsx« v delovnem imeniku (pridobljen z `getwd()`).
7. Razen če h klicni sintaksi izrecno dodate `open.output = FALSE`, bo output datoteka odprta po končanju vseh izračunov. To je uporabno, ko se izvajajo številni klici funkcij za različne analize in ni treba imeti takojšnjega pregleda ouputa.

Tabele za PV za vsakega od mejnikov (angl. *benchmark*) v PIRLS so predstavljene v spodnji tabeli.

Score points	Level name
400	Low
475	Intermediate
550	High
625	Advanced

Vir: Mullis, I. V. S. in Prendergast, C. O. (2017). Using Scale Anchoring to Interpret the PIRLS and ePIRLS 2016 Schievement Scales. V: M. O. Martin, I. V. S. Mullis, & M. Hooper (ur.), Methods and Procedures in PIRLS 2016 (str. 13.1–13.23). Lynch School of Education, Boston College.

Z izvajanjem zgornje kode se bo v konzoli RStudia prikazal naslednji output:



Ko bodo vse operacije končane, bo output zapisan na disk kot Excelova datoteka. Če je `open.output = TRUE` (privzeto), se bo datoteka odprla v privzetem programu za preglednice (običajno MS Excel).

Izračunajmo odstotke deklet in fantov, ki dosegajo ali presegajo srednji in višji mejnik (475 in 550 točk), ter za vsako skupino izračunajmo povprečje ene spremenljivke. Uporabili bomo kompleksno lestvico, kako zelo učenci radi berejo (ASBGSLR; za podrobnosti o konstrukciji te lestvice in njenih lastnostih glejte tehnično dokumentacijo PIRLS 2016):

```
lsa.bench(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
          split.vars = "ASBG01", PV.root.bench = "ASRREA",
          bench.vals = c(475, 550), bckg.var = "ASBGSLR")
```

Zgornja klicna sintaksa doda argument bench.vals in njegovo vrednost, vektor mejnih vrednosti za izračun odstotkov učencev, ki dosežejo ali presežejo določene mejne vrednosti v distribuciji – `c(475, 550)`. Doda tudi argument bckg.var in njegovo vrednost, ime spremenljivke ASBGSLR. Output bo imel za vsako skupino podobno strukturo kot prejšnji, vendar bodo tokrat dodani stolpci, ki se nanašajo na povprečje lestvice.

Izvajanje zgornje sintakse bo prepisalo prejšnji output, saj ima datoteka enako ime. Stolpci na listu »Estimates« bodo zdaj različni.

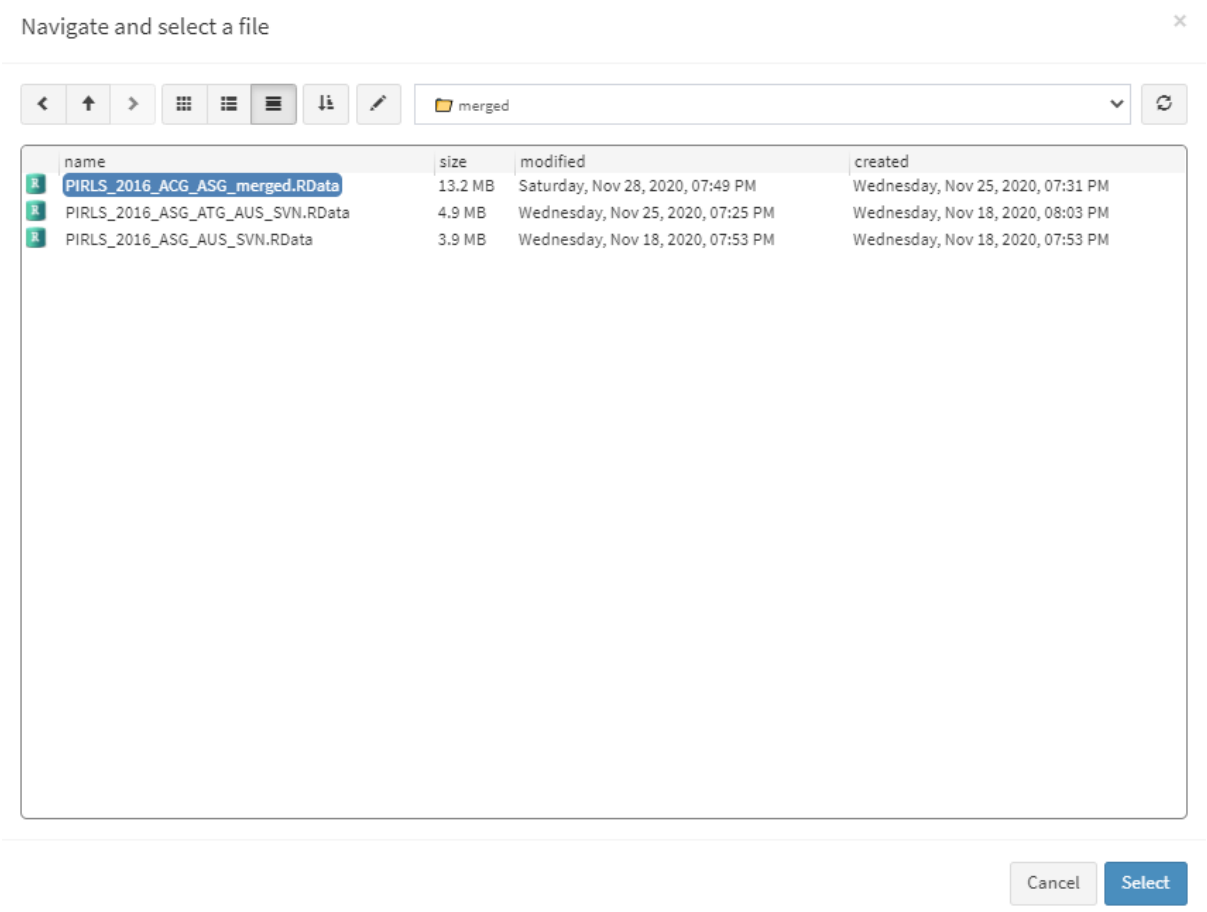
4.3.4 Računanje mejnikov (angl. *benchmarks*) z uporabo grafičnega uporabniškega vmesnika (GUI)

Za zagon RALSA uporabniškega vmesnika izvedite naslednji ukaz v RStudiu:

```
ralsaGUI()
```

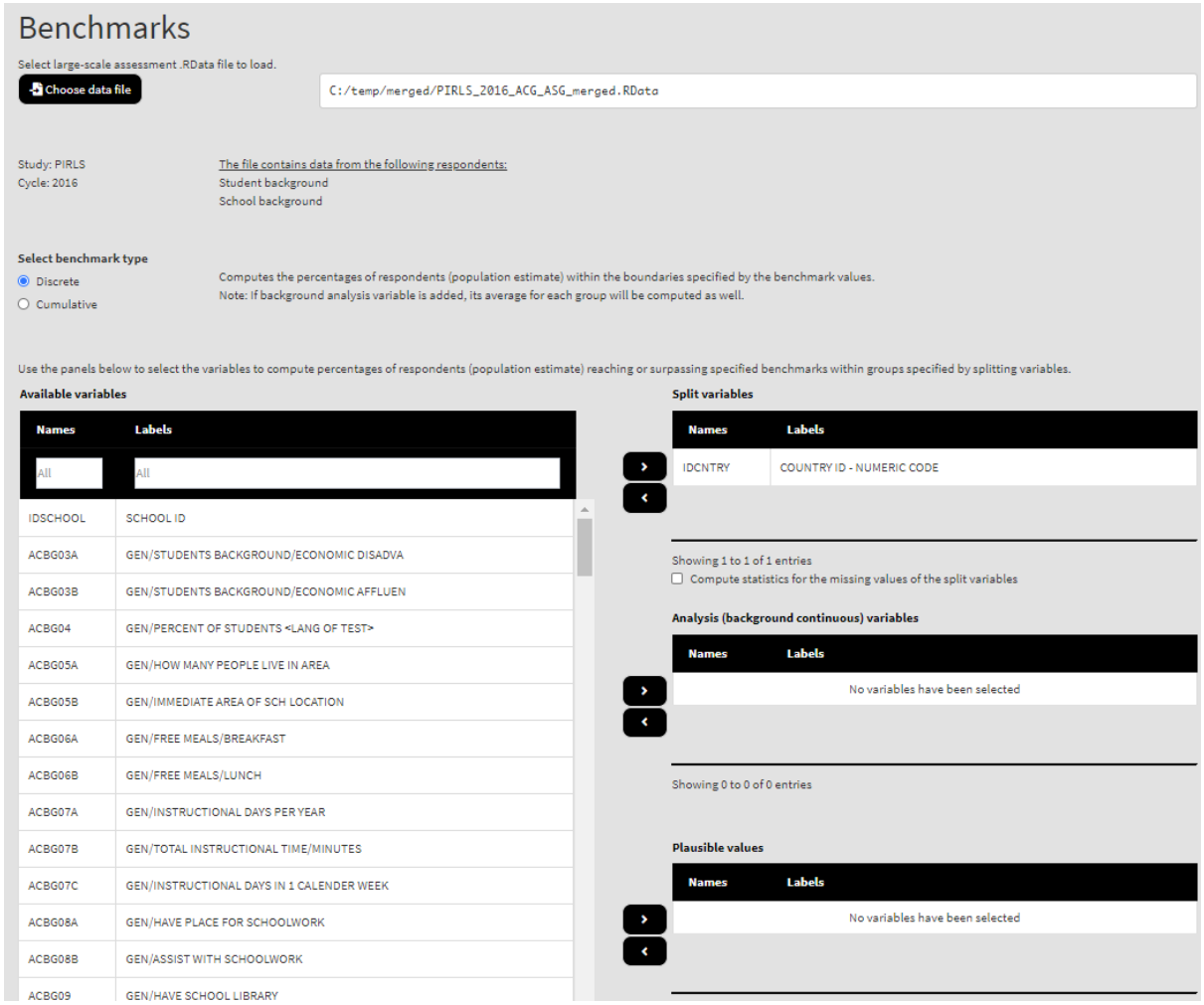
Za primere, ki sledijo, združite nov podatkovni niz s PIRLS 2016-podatki za Avstralijo in Slovenijo ter vključite vse spremenljivke o učencih in ravnateljih. Združeno datoteko lahko poimenujete **»PIRLS_2016_ACG_ASG_merged.RData«**.

Ko končate z združevanjem podatkov, v meniju na levi izberite **Analysis types > Benchmarks**. Ko se vmesnik GUI premakne na Benchmarks, kliknite gumb **Choose data file**. Nato se pomaknite do mape, ki vsebuje združeno datoteko **»PIRLS_2016_ACG_ASG_merged.RData«**, jo izberite in kliknite gumb Select.



Ko je datoteka naložena, boste na levi strani videli ploščo z razpoložljivimi spremenljivkami (Available variables) in niz plošč na desni, kamor lahko dodate spremenljivke s seznama razpoložljivih. Nad ploščami boste prav tako videli informacije o naloženi datoteki in gumbe, s

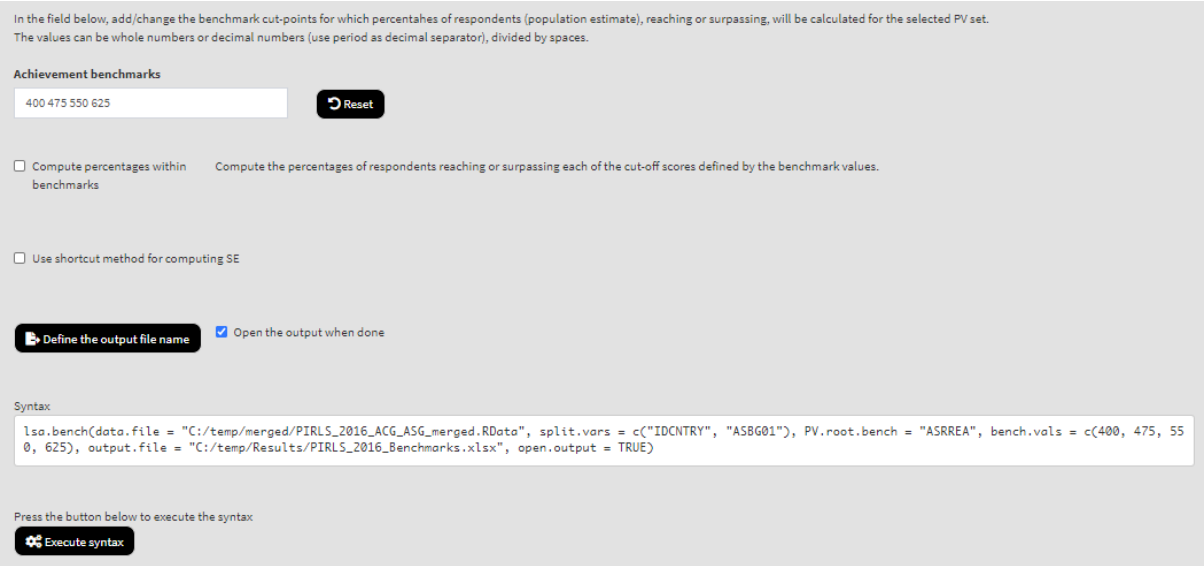
katerimi izberete vrsto mejnih vrednosti – **Discrete** ali **Cumulative**. Pustite privzeto izbrano možnost Discrete.



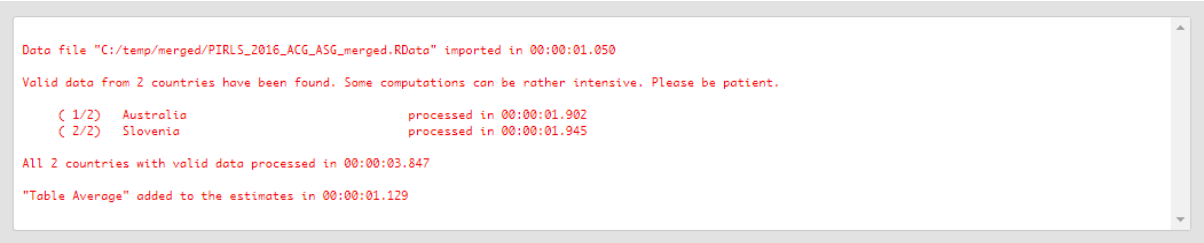
Z miško izberite spremenljivke s seznama **Available variables** (Razpoložljive spremenljivke) in uporabite puščične gumbe na sredini zaslona, da jih dodate v različna polja (ali jih odstranite), s čimer nastavite analizo. Za hitro iskanje želenih spremenljivk lahko uporabite iskalna polja na vrhu plošč.

Za začetek izračunajmo odstotke učencev v Avstraliji in Sloveniji, ki dosežejo ali presežejo vse mejnike skupnega bralnega dosežka. Dodajte spremenljivko ASBG01 na seznam **Split variables** (Delitvene spremenljivke). Na seznamu Available variables poiščite koren verjetnostnih vrednosti (PV) za dosežek iz branja (ASRREA). Uporabite iskalna polja na vrhu plošče za hitro iskanje, bodisi po imenu bodisi po oznaki. Izberite koren in ga s pomočjo puščičnih gumbov dodajte na ploščo **Plausible values** (Verjetne vrednosti). Bodite pozorni na to, kako so določene PV. Pet PV za bralne dosežke: ASRREA01, ASRREA02, ASRREA03, ASRREA04, in ASRREA05. Seznama **Available variables** in **Plausible values** ne bosta pokazala teh petih ločenih PV, temveč samo njihov skupni koren ASRREA, brez številk na koncu. Funkcija bo v ozadju v izračune samodejno vključila vseh pet PV.

Ko to storite, se pomaknite navzdol in kliknite na **Define output file name** (Določite ime output datoteke). Pomaknite se do mape C:/temp/Results (ali do mape, kamor želite shraniti output) in določite ime izhodne datoteke. Ko to storite, se bo ob polju Define output file name prikazal potrditveni okvir. Če ga označite, se bo output odprl po končanih izračunih. Pod tem bo prikazana klicna sintaksa. Spodaj bo prikazan gumb **Execute syntax** (Izvedi sintakso). Končne nastavitve v spodnjem delu zaslona bi morale izgledati takole:



Kliknite gumb **Execute syntax** (Izvedi sintakso). Na dnu zaslona se bo prikazala konzola GUI, kjer bodo zabeležene vse opravljene operacije.



1. Obstajata dve spremenljivki, ki vsebujeta informacije o spolu učencev:
   - ITSEX: spremenljivka, ki vsebuje podatke o sledenju učencev;
   - ASBG01: spremenljivka, ki prikazuje spol učencev, kot so ga navedli v vprašalniku (uporabljena v tej analizi).
2. Pri mednarodnih raziskavah je treba vse analize izvajati ločeno za vsako državo. Spremenljivka z identifikacijsko oznako države (IDCNTRY, ali CNT v PISA) je vedno izbrana kot prva razdelitvena spremenljivka in je ni mogoče odstraniti s plošče Split variables.
3. Za skupno bralno uspešnost (tj. dosežke na preizkusu branja) je samodejno izbranih vseh 5 PV (ASRREA01, ASRREA02, ASRREA03, ASRREA04 in ASRREA05). Vse, kar je treba storiti, je, da na seznam **Plausible values** (verjetnostne vrednosti) dodate koren PV (ASRREA). Za več podrobnosti o korenih PV (tudi za raziskave, ki niso TIMSS in PIRLS) si oglejte dokumentacijo posamezne raziskave.
4. Vrednosti za merila uspešnosti so dodane v besedilno polje **Achievement benchmarks**. Vsa merila uspešnosti za določeno raziskavo (in v nekaterih raziskavah

celo za ustrezen cikel) bodo dodana samodejno. Če jih želite urediti, poskrbite, da bodo vrednosti ločene s presledki. Če želite obnoviti privzete vrednosti, kliknite gumb **Reset**.

5. Privzeta utež je samodejno izbrana in dodana na ploščo **Weight variable**. Lahko jo zamenjate z drugo utežjo, ki je na voljo v podatkovnem naboru. Če je izbrana privzeta utež, se v oknu sintakse ne bo prikazala. Če v plošči Weight variable ni izbrane uteži, bo privzeta uporabljena samodejno.
6. Če je potrjeno okence **Compute percentages within benchmarks**, bodo izračunani odstotki učencev, ki dosežejo ali presežejo vsako od mejnih vrednosti, določenih z vrednostmi meril.
7. Okence **Use shortcut method for computing SE** ni označeno, kar pomeni, da bo funkcija izračunala standardno napako z uporabo »polne« metode za komponento varianc vzorčenja.

Točkovanje za PV pri vsakem mejniku v PIRLS je predstavljeno v spodnji tabeli.

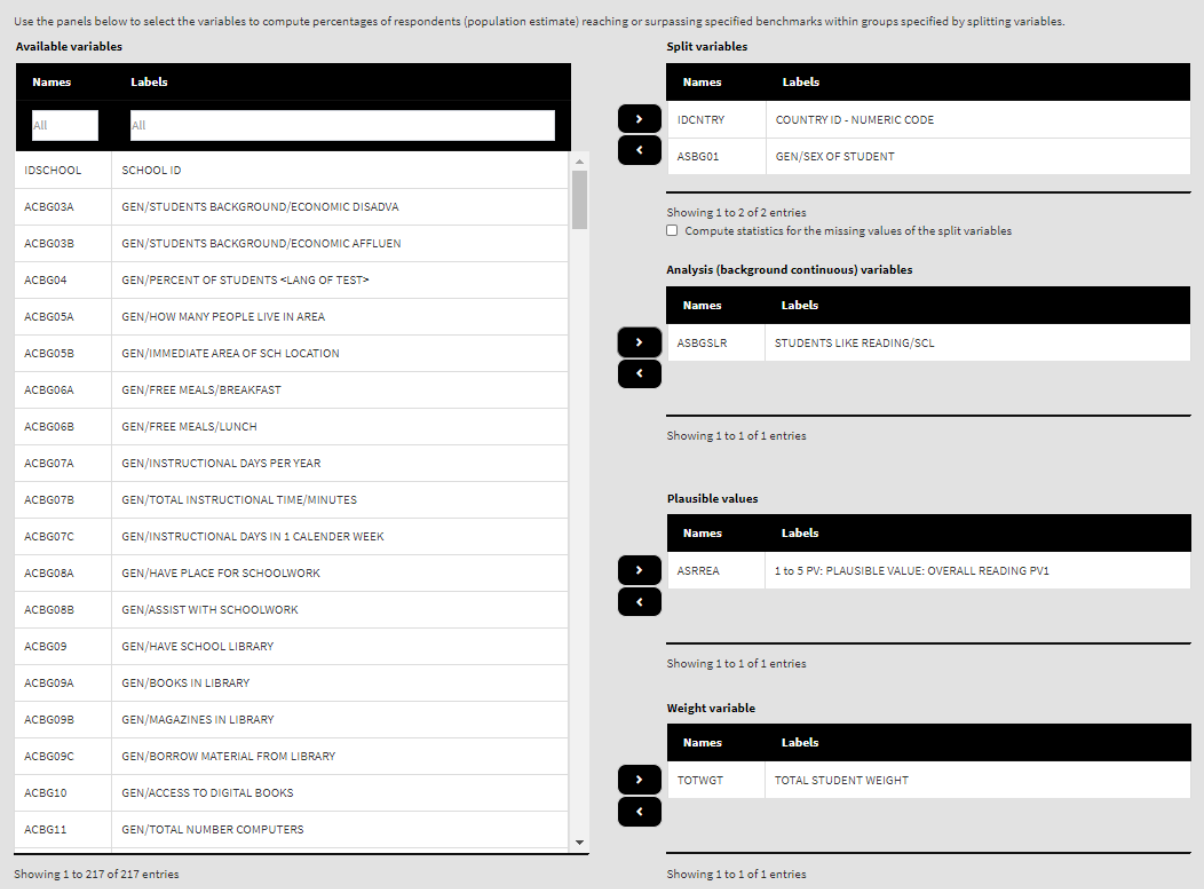
Score points	Level name
400	Low
475	Intermediate
550	High
625	Advanced

Vir: Mullis, I. V. S. in Prendergast, C. O. (2017). Using Scale Anchoring to Interpret the PIRLS and ePIRLS 2016 Schievement Scales. V: M. O. Martin, I. V. S. Mullis, & M. Hooper (ur.), Methods and Procedures in PIRLS 2016 (str. 13.1–13.23). Lynch School of Education, Boston College.

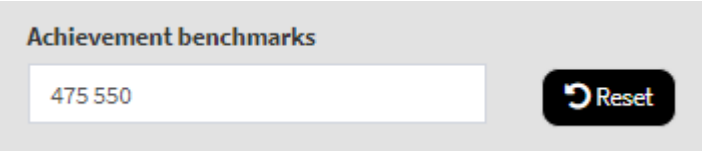
Če je označeno polje **Open the output when done**, se bo output datoteka po končanih izračunih samodejno odprla v privzetem programu za preglednice (običajno MS Excel).

Izračunajmo odstotke učencev in učenk, ki dosegajo ali presegajo srednji in višji mejnik (475 in 550 točk), in za vsako skupino izračunajmo povprečje ene spremenljivke. Uporabili bomo zapleteno lestvico za merjenje, koliko učenci radi berejo (ASBGSLR; za več informacij o oblikovanju in lastnostih te lestvice preverite tehnično dokumentacijo PIRLS 2016).

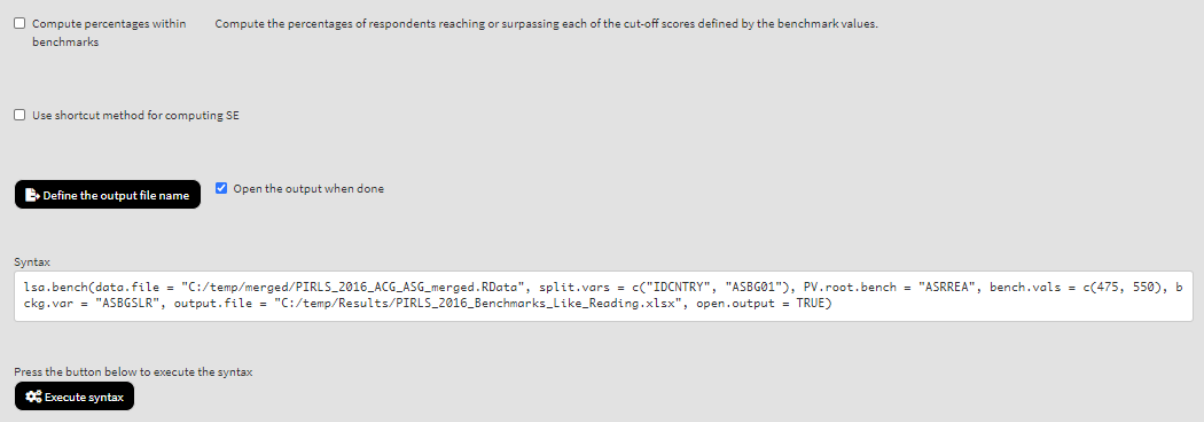
Po nastavitvah prejšnje analize izberite spremenljivko ASBGSLR s seznama **Available variables** in jo s pomočjo puščic dodajte na ploščo **Background continuous variables.** Nastavitve zdaj vključujejo IDCNTRY in ASBG01 kot **Split variables** (Razdelitvene spremenljivke) ter ASBGSLR kot Background continuous variables (Zvezne kontekstualne spremenljivke).



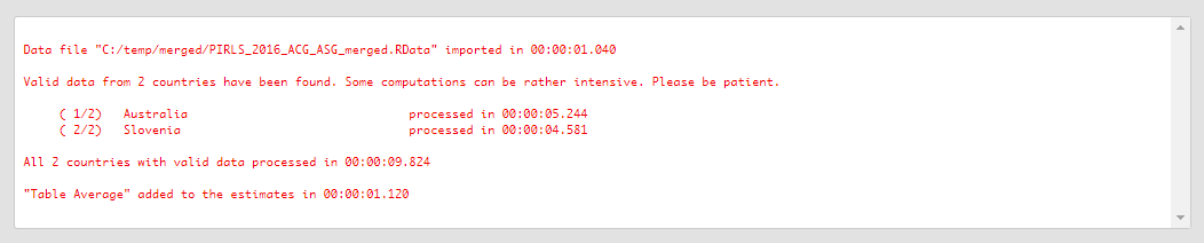
Ker uporabljamo aplikacijo z nastavitvami **Benchmarks** neposredno po izvedbi prejšnje analize, so vse preostale nastavitve enake kot pri prejšnji analizi. Edina sprememba je v polju **Achievement benchmarks values** (Vrednosti mejnikov dosežkov). Uredite vrednosti in jih spremenite v **475** in **550**:



Ni potrebe po spreminjanju preostalih nastavitev, razen če želite. Lahko pa spremenite ime output datoteke, sicer bo prepisana. Upoštevajte, da se prikazana sintaksa spremeni, kar odraža vključitev ASBGSLR kot zvezne ozadenjske spremenljivke (angl. *continuous background variable*) za izračun povprečja:



Kliknite gumb **Execute syntax**. Konzola GUI se bo posodobila in zabeležila vse zaključene operacije:



4.4 Navzkrižne tabele (angl. *Crosstabulations*)

4.4.1 Uvod

Funkcija `lsa.crosstabs` izračuna navzkrižne tabele (angl. Crosstabs) med dvema kategorialnima spremenljivkama znotraj skupin anketirancev, ki so opredeljene z delitvenimi spremenljivkami (angl. *split variables*). Delitvene spremenljivke so opcijske. Če delitvene spremenljivke niso podane, bodo rezultati izračunani le na ravni držav. Če so podane, bodo podatki znotraj vsake države razdeljeni v skupine po vseh delitvenih spremenljivkah, ocene pa bodo izračunane po kategorijah zadnje delitvene spremenljivke. Ocene bodo izračunane z uporabo polne uteži in vseh njenih replikatov. Na koncu bo njihov standardni odklon izračunan z uporabo zapletenih formul, ki bodo odvisne od preučevane raziskave. Ne glede na oceno bo standardni odklon izračunan ob upoštevanju zapletenih vzorčnih zasnov raziskav.

Če vas zanimajo podrobnejše informacije o zapletenih vzorčnih zasnovah posamezne raziskave ter kako so ocene in njihovi standardni odkloni izračunani, si oglejte tehnično dokumentacijo in uporabniški priročnik raziskave.

Funkcija prav tako izračuna **hi-kvadrat** test neodvisnosti med vrstičnimi in stolpičnimi spremenljivkami z Rao-Scottovim popravkom. Ko se uporabljajo podatki iz zapletenih vzorcev

(kot v mednarodnih raziskavah), je tradicionalna **hi-kvadrat** statistika pristranska. **Rao-Scottova** prilagoditev zagotavlja nepristranske ocene.

Kot vsaka druga funkcija v paketu RALSA funkcija `lsa.crosstabs` prepozna podatke iz raziskave in uporabi pravilne tehnike ocenjevanja glede na vzorčno in preizkusno zasnovo raziskave, brez dodatnega prilagajanja.

4.4.2 Funkcija crosstabulations in njeni argumenti

Funkcija `lsa.crosstabs` ima naslednje argumente:

- `data.file` – datoteka, ki vsebuje objekt `lsa.data`. Podati je treba ali to ali `data.object`, vendar ne obeh.
- `data.object` – objekt v pomnilniku, ki vsebuje `lsa.data`. Podati je treba ali tega ali `data.file`, vendar ne obeh.
- `split.vars` – kategorialna spremenljivka(e) za razdelitev rezultatov. Če delitvene spremenljivke niso podane, bodo prikazani rezultati za splošno populacijo držav. Če je podana ena ali več spremenljivk, bodo rezultati razdeljeni po vseh razen po zadnji spremenljivki, odstotki anketirancev pa bodo izračunani po edinstvenih vrednostih zadnje delitvene spremenljivke.
- `bckg.row.var` – ime vrstične kategorialne ozadenske spremenljivke. Rezultati bodo izračunani za vse skupine, določene z delitvenimi spremenljivkami.
- `bckg.col.var` – ime stolpične kategorialne ozadenjske spremenljivke. Rezultati bodo izračunani za vse skupine, določene z delitvenimi spremenljivkami.
- `expected.cnts` – določitev, ali naj bodo izračunane tudi pričakovane frekvence. Privzeta vrednost (`TRUE`) bo izračunala pričakovane frekvence. Če je nastavljena na `FALSE`, bodo v output vključene le opazovane frekvence.
- `row.pcts` – določitev, ali naj se izračunajo vrstični odstotki. Privzeta vrednost (`FALSE`) preskoči izračun vrstičnih odstotkov.
- `column.pcts` – določitev, ali naj se izračunajo stolpični odstotki. Privzeta vrednost (`FALSE`) preskoči izračun stolpičnih odstotkov.
- `total.pcts` – logična vrednost, ali naj se izračunajo skupni odstotki. Privzeta vrednost (`FALSE`) preskoči izračun skupnih odstotkov.
- `weight.var` – ime spremenljivke, ki vsebuje uteži. Če ime utežne spremenljivke ni podano, bo funkcija samodejno izbrala privzeto utežno spremenljivko za podane podatke, odvisno od tipa anketiranca.
- `include.missing` – logična vrednost, ali naj bodo manjkajoče vrednosti delitvenih spremenljivk vključene kot kategorije za razdelitev in zanje izdelani vsi statistični podatki. Privzeta vrednost (`FALSE`) upošteva vse primere delitvenih spremenljivk brez manjkajočih vrednosti pred izračunom statističnih podatkov. Glejte podrobnosti.
- `shortcut` – določitev, ali naj se uporabi metoda »bližnjica« za IEA TIMSS, TIMSS Advanced, TIMSS Numeracy, eTIMSS, PIRLS, ePIRLS, PIRLS Literacy in RLII. Privzeta vrednost (`FALSE`) uporabi »polno« zasnovo pri izračunavanju variance in standardnih napak ocen.
- `graphs` – določitev, ali naj se izdelajo grafi. Privzeta vrednost je `FALSE`.

- `graph.row.label` – niz, ki predstavlja prilagojeno oznako za vrstično spremenljivko v grafih. Ignorirano, če `graphs = FALSE`.
- `graph.col.label` – niz, ki predstavlja prilagojeno oznako za stolpično spremenljivko v grafih. Ignorirano, če `graphs = FALSE`.
- `save.output` – logična vrednost, ali naj se izhod shrani v MS Excel-datoteko (privzeto) ali ne (prikazano v konzoli ali dodeljeno objektu).
- `output.file` – celotna pot do izhodne datoteke, vključno z imenom datoteke. Če je izpuščeno, bo datoteka z privzetim imenom »Analysis.xlsx« shranjena v delovni imenik (`getwd()`).
- `open.output` – določitev, ali naj se output odpre po tem, ko je bil shranjen. Privzeta vrednost (`TRUE`) odpre izhod v privzetem program za preglednice, nameščenem na računalniku.

Opombe:

- `data.file` ali `data.object` morata biti navedena kot vir podatkov. Če sta oba navedena, bo funkcija povzročila napako.
- Funkcija izračuna navzkrižno tabelo po kategorijah delitvenih spremenljivk. Odstotki anketirancev v vsaki skupini se izračunajo znotraj skupin, določenih z zadnjo delitveno spremenljivko. Če delitvene spremenljivke niso dodane, bodo rezultati izračunani le po državah.
- Če je `include.missing = FALSE` (privzeto), bodo vsi primeri z manjkajočimi vrednostmi na delitvenih spremenljivkah odstranjeni, v statistiko pa bodo vključeni le primeri z veljavnimi vrednostmi. Upoštevajte, da se lahko podatki iz raziskav izvozijo na dva različna načina: (1) nastavitev vseh uporabniško določenih manjkajočih vrednosti na `NA`; (2) uvoz vseh uporabniško določenih manjkajočih vrednosti kot veljavnih in dodajanje njihovih kod v dodatni atribut vsake spremenljivke. Če je `include.missing` nastavljeno na `FALSE` (privzeto) in so podatki izvoženi z uporabo možnosti (2), bo izhod odstranil vse vrednosti iz spremenljivke, ki ustreza vrednostim v njenem atributu manjkajočih vrednosti. V nasprotnem primeru jih bo vključil kot veljavne vrednosti in zanje izračunal statistiko.
- Argument `shortcut` je veljaven le za TIMSS, TIMSS Advanced, TIMSS Numeracy, PIRLS, ePIRLS, PIRLS Literacy in RLII. Prej so te raziskave pri izračunu standardnih napak uporabljale 75 replikatov, ker je imela ena izmed šol v 75 JK območjih podvojene uteži, druga pa je bila izključena. Od TIMSS 2015 in PIRLS 2016 naprej raziskave uporabljajo 150 replikatov in v vsakem JK območju šoli enkrat podvojijo uteži ter jo enkrat izključijo, torej se izračuni izvedejo dvakrat za vsako območje. Za več podrobnosti glejte tehnično dokumentacijo in uporabniške priročnike TIMSS 2015 in PIRLS 2016. Če je potrebno ponavljanje tabel in grafikonov, je treba argument shortcut nastaviti na `TRUE`.
- Če je `graphs = TRUE`, bo funkcija ustvarila grafe po kombinaciji kategorij `bckg.row.var` in `bckg.col.var` (ocene populacije) po skupini, določeni s `split.vars`. Vse prikaze bo ustvarila za posamezne države. Če ni `split.vars`, bo na koncu na voljo graf za vse države skupaj. Če je treba, lahko prilagojene oznake za horizontalno in vertikalno os določite z argumentoma `graph.row.label` in `graph.col.label`.

- Če je `save.output = FALSE`, bo output vsebina seznam, ki vsebuje ocene in informacije o analizi. Če je `graphs = TRUE`, bodo grafi dodani na seznam ocen. Če je `save.output = TRUE` (privzeto), bo ustvarjena datoteka v MS Excelu (.xlsx) (ki jo lahko odprete v katerem koli programu za preglednice), kot je določeno s celotno potjo v `output.file`. Če je argument izpuščen, bo Excelova datoteka z generičnim imenom »Analysis.xlsx« shranjena v delovni imenik (`getwd()`).

Delovni zvezek vsebuje štiri preglednice. Prva (»Estimates«) vsebuje tabelo z rezultati po državah, končni del tabele pa vsebuje povprečene rezultate vseh statističnih podatkov po državah. V tabeli lahko najdete naslednje stolpce, odvisno od specifikacije analize:

- **<Country ID>** – stolpec, ki vsebuje imena držav v datoteki, za katere so izračunane statistike. Natančen naslov stolpca bo odvisen od identifikatorja države, uporabljenega v posamezni raziskavi.
- **<Split variable 1>**, **<Split variable 2>**… – stolpci, ki vsebujejo kategorije, po katerih so bile statistike razdeljene. Natančna imena bodo odvisna od spremenljivk v `split.vars`.
- **n_Cases** – število primerov v vzorcu, uporabljenem za izračun statistike za vsako kombinacijo razdelitev, določenih s `split.vars`, če so kakšne, in `bckg.row.var`.
- **Sum_<Weight variable>** – ocenjen številčni podatek populacije po uporabi uteži. Dejanski naziv utežne spremenljivke bo odvisen od utežne spremenljivke, uporabljene v analizi.
- **Sum_<Weight variable>_SE** – standardni odklon ocenjenega številčnega podatka populacije po skupinah. Dejanski naziv utežne spremenljivke bo odvisen od utežne spremenljivke, uporabljene v analizi.
- **Percentages_<Row variable>** – odstotki anketirancev (ocene populacije) po skupinah, določenih z delitvenimi spremenljivkami v `split.vars`, če so kakšne, in vrstično spremenljivko v `bckg.row.var`. Prikazani bodo odstotki za kombinacijo kategorij v zadnji delitveni spremenljivki in vrstično spremenljivko, ki določata končne skupine.
- **Percentages_<Row variable>_SE** – standardni odkloni odstotkov zgoraj.
- **Type** – vrsta izračunanih vrednosti, odvisno od logičnih vrednosti, posredovanih argumentom `expected.cnts`, `row.pcts`, `column.pcts` in `total.pcts`: »Observed count«, »Expected count«, »Row percent«, »Column percent« in »Percent of total«.
- **<Column variable name Category 1>**, **<Column variable name Category 1>**,… – ocenjene vrednosti za vse kombinacije med vrstičnimi in stolpičnimi spremenljivkami, posredovanimi v `bckg.row.var` in `bckg.col.var`. Za vsako kategorijo stolpične spremenljivke bo prikazan en stolpec.
- **<Column variable name Category 1, 2,… n>_SE** – standardni odkloni ocenjenih vrednosti zgoraj.
- **Total** – skupni seštevki za vsake vrste ocenjenih vrednosti (»Observed count«, »Expected count«, »Row percent«, »Column percent« in »Percent of total«), odvisno od logičnih vrednosti (`TRUE`, `FALSE`), posredovanih argumentom `expected.cnts`, `row.pcts`, `column.pcts` in `total.pcts`.
- **Total_SE** – standardni odkloni ocenjenih vrednosti zgoraj.

Druga preglednica vsebuje dodatne informacije, povezane z analizo po državah v naslednjih stolpcih:

- **<Country ID>** – stolpec, ki vsebuje imena držav v datoteki, za katere so izračunane statistike. Natančen naslov stolpca bo odvisen od identifikatorja države, uporabljenega v posamezni raziskavi.
- **<Split variable 1>, <Split variable 2>…** – stolpci, ki vsebujejo kategorije, po katerih so bile statistike razdeljene. Natančna imena bodo odvisna od spremenljivk v `split.vars`.
- **Statistics** – vsebuje imena različnih vrst statistik: hi-kvadrat, stopnje svobode in p-vrednosti.
- **Value** – ocenjene vrednosti zgoraj omenjenih statistik.

Tretja preglednica vsebuje dodatne informacije, povezane z analizo po državah v naslednjih stolpcih:

- **DATA** – uporabljena `data.file` ali `data.object`.
- **STUDY** – iz katere raziskave izvirajo podatki.
- **CYCLE** – iz katerega cikla raziskave so podatki.
- **WEIGHT** – katera utežna spremenljivka je bila uporabljena.
- **DESIGN** – katera tehnika vzorčenja je bila uporabljena (JRR ali BRR).
- **SHORTCUT** – ali je bila uporabljena metoda »shortcut«.
- **NREPS** – koliko replikacijskih uteži je bilo uporabljenih.
- **ANALYSIS_DATE** – na kateri datum je bila analiza opravljena.
- **START_TIME** – ob katerem času se je analiza začela.
- **END_TIME** – ob katerem času se je analiza končala.
- **DURATION** – koliko časa je analiza trajala v urah, minutah, sekundah in milisekundah.

Četrta preglednica vsebuje klic funkcije z vrednostmi za vse parametre, kot je bila izvedena. To je koristno, če bo treba kasneje ponoviti analizo.

Če je `graphs = TRUE`, bo dodana preglednica »Graphs«, ki vsebuje vse prikaze.
Če se pojavijo kakršna koli opozorila, ki izhajajo iz izračunov, bodo vključena v dodatno preglednico »Warnings« v delovnem zvezku.

4.4.3 Računanje navzkrižnih tabel z uporabo ukazne vrstice

V naslednjih primerih bomo združili novo podatkovno datoteko s podatki o učencih in ravnateljih iz PIRLS 2016 (Avstralija in Slovenija), pri čemer bomo vzeli vse spremenljivke iz obeh vrst datotek:

```
lsa.merge.data(inp.folder = "C:/temp",
               file.types = list(acg = NULL, asg = NULL),
               ISO = c("aus", "svn"),
               out.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData")
```

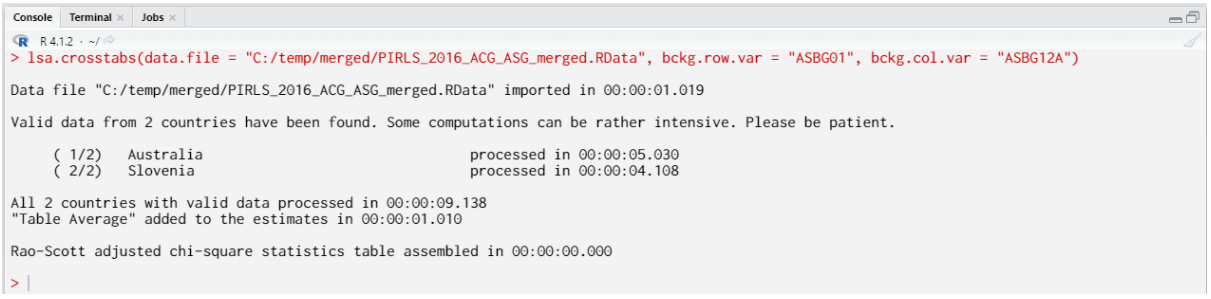
Za začetek izračunajmo navzkrižno tabelo med dvema kategorialnima spremenljivkama – spolom učencev (ASBG01) in tem, koliko se učenci v Avstraliji in Sloveniji strinjajo, da imajo radi šolo (ASBG12A). Spremenljivka ASBG01 ima dve veljavni kategoriji: (1) »Dekle«; (2) »Fant«. Spremenljivka ASBG12A ima štiri veljavne kategorije: (1) Zelo se strinjam; (2) Strinjam se; (3) Ne strinjam se; (4) Sploh se ne strinjam. Spol učencev (ASBG01) bo vrstična spremenljivka, pričakovanje učencev o tem, koliko imajo radi šolo (ASBG12A), pa bo stolpična spremenljivka v tabeli:

```
lsa.crosstabs(data.file = "C:/temp/PIRLS_2016_ACG_ASG_merged.RData",
              bckg.row.var = "ASBG01",
              bckg.col.var = "ASBG12A")
```

Nekaj pomembnih opomb:

- V mednarodnih raziskavah je treba vse analize izvajati ločeno po državah. Vendar ni treba dodajati spremenljivke ID države (IDCNTRY ali CNT v PISA) kot spremenljivke za razdelitev. Funkcija jo bo samodejno prepoznala in dodala v vektor `split.vars`.
- Ni treba posebej določiti utežne spremenljivke. Če utež ni posebej določena, bo uporabljena privzeta utež (v tem primeru skupna utež učencev v tem primeru) za podatkovni niz, ki je samodejno identificirana. Če imate dober razlog za spremembo spremenljivke uteži, lahko to storite tako, da dodate `weight.var = "SENWGT"`.
- Nobenih dodatnih argumentov za izračunavanje odstotkov vrstic, stolpcev in skupnih odstotkov ni bilo določenih. Ti se lahko izračunajo, vendar bo output postal odvečen in težje berljiv.
- Razen če izrecno dodate `save.output = FALSE`, bo output shranjen v MS Excel-datoteko na disku. V nasprotnem primeru bo output natisnjen v konzoli.
- Če ni določena datoteka za output, bo output shranjen z imenom datoteke »Analysis.xlsx« v delovnem imeniku (pridobljen s `getwd()`).
- Razen če h klicu izrecno dodate `open.output = FALSE`, bo output datoteka po končanem izračunu odprta. To je uporabno, ko se izvaja več klicev za različne analize in takojšnji ogledi izhoda niso potrebni.

Izvajanje zgornje kode bo v konzoli RStudia natisnilo naslednji output:



Ko so vsi postopki zaključeni, bo output shranjen na disk kot MS Excel-delovna knjiga. Če je `open.output = TRUE` (privzeto), bo datoteka odprta v privzetem programu za preglednice (običajno MS Excel).

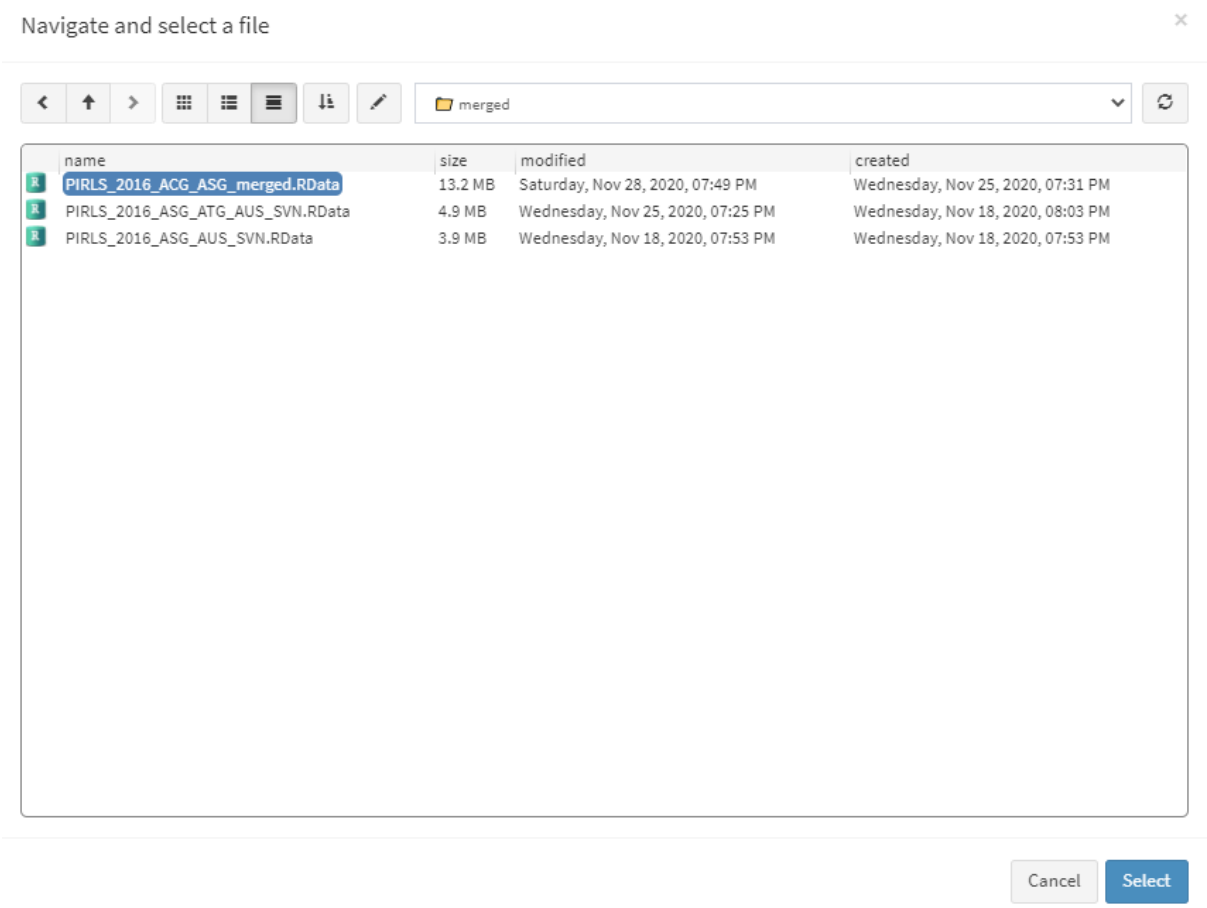
4.4.4 Računanje navkrižnih tabel z uporabo GUI

Za zagon uporabniškega vmesnika RALSA v RStudiu izvedite naslednji ukaz:

```
ralsaGUI()
```

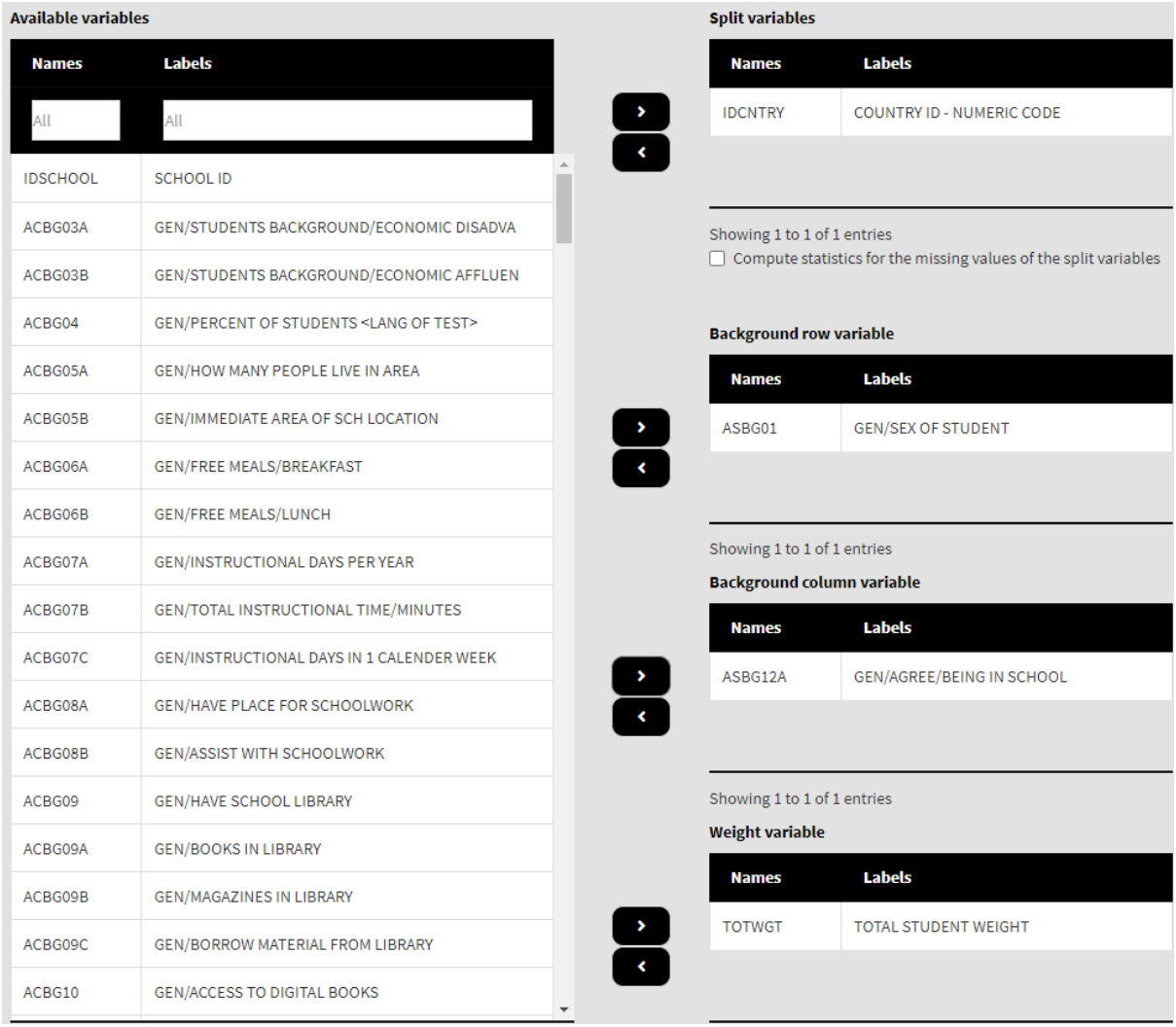
Za naslednje primere združite novo datoteko s podatki PIRLS 2016 za Avstralijo in Slovenijo z vsemi spremenljivkami o učencih in ravnateljih. Združeno datoteko lahko poimenujete **»PIRLS_2016_ACG_ASG_merged.RData«**.

Ko končate z združevanjem podatkov, izberite **Analysis types > Crosstabulations** iz menija na levi strani. Ko pridete do **Crosstabulations** v GUI, kliknite gumb **Choose data file**. Pomaknite se do mape, ki vsebuje združeno datoteko **»PIRLS_2016_ACG_ASG_merged.RData«**, jo izberite in kliknite gumb **Select**.

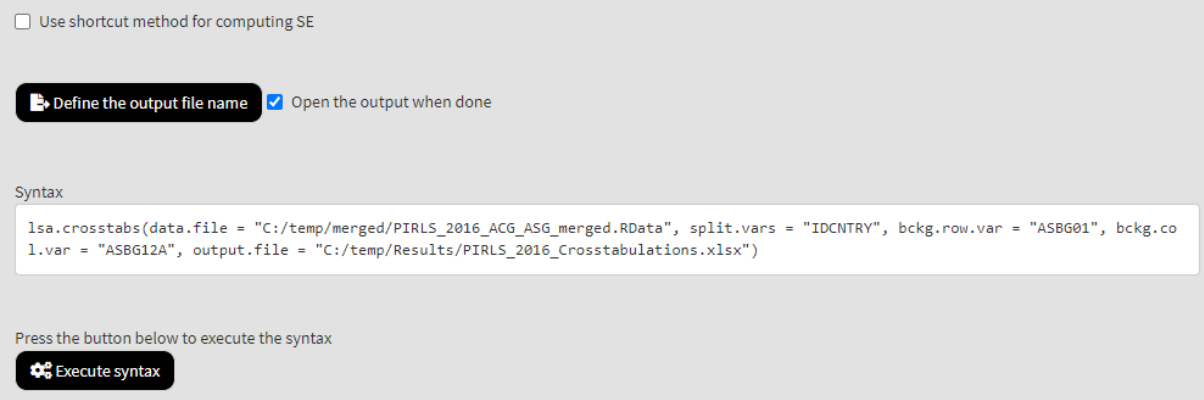


Ko je datoteka naložena, boste na levi strani videli ploščo (razpoložljive spremenljivke) in niz plošč na desni strani, kjer lahko spremenljivke s seznama razpoložljivih dodate. Izračunali bomo navzkrižno tabelo med dvema kategorialnima spremenljivkama – spol učencev (ASBG01) in tem. koliko učenci v Avstraliji in Sloveniji soglašajo z izjavo, da imajo radi šolo (ASBG12A) v Avstraliji in Sloveniji. Spremenljivka ASBG01 ima dve veljavni kategoriji: (1) »Dekle«; (2) »Fant«. Spremenljivka ASBG12A ima štiri veljavne kategorije: (1) Zelo se strinjam; (2) Strinjam se; (3) Ne strinjam se; (4) Sploh se ne strinjam. Spol učencev (ASBG01) bo vrstična spremenljivka, strinjanje učencev glede tega, koliko imajo radi šolo (ASBG12A),

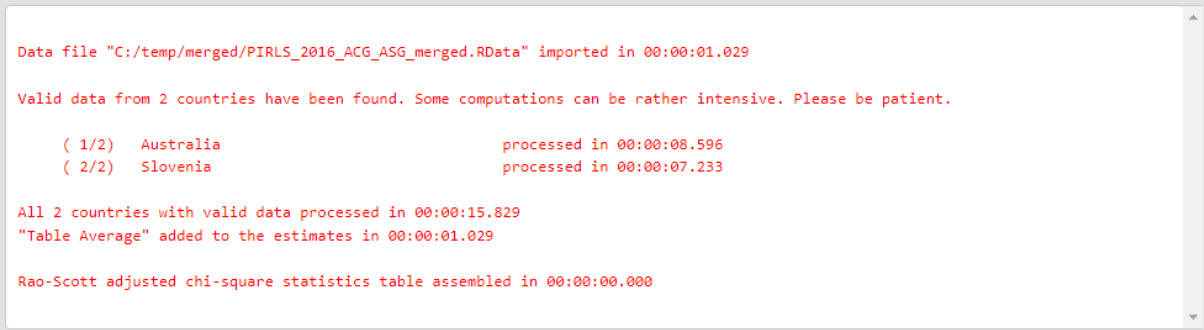
pa bo stolpčna spremenljivka v tabeli. S seznama razpoložljivih spremenljivk izberite spremenljivko ASBG01 in jo dodajte na seznam spremenljivk za vrstice ozadja z desnim gumbom s puščico. S seznama razpoložljivih spremenljivk izberite spremenljivko ASBG12A in jo dodajte na seznam spremenljivk za stolpce ozadja z desnim puščico. To je vse, kar je treba narediti.



Pomaknite se navzdol in kliknite na gumb za določanje output datoteke. Pojdite v mapo **»C:/temp/Results«** (ali v mapo, kjer želite shraniti output) in določite ime output datoteke. Ko to storite, se bo poleg **Define the output file name** prikazalo potrditveno polje. Če je označeno, se bo output odprl po končanih izračunih. Pod tem bo prikazana klicna sintaksa. Pod vsemi temi bo prikazan gumb **Execute syntax.** Končne nastavitve v spodnjem delu zaslona naj izgledajo takole:



Kliknite gumb **Execute syntax**. V spodnjem delu GUI se bo pojavila konzola, ki bo beležila vse zaključene operacije:



Nekaj pomembnih točk:

- Pri mednarodnih raziskavah je treba vse analize izvajati ločeno za vsako državo. Ni pa treba dodati spremenljivke ID države (IDCNTRY ali CNT v PISA) kot spremenljivke za deljenje. Funkcija jo bo samodejno prepoznala in dodala v vektor `split.vars`.
- Ni treba izrecno določiti utežne spremenljivke. Če utež ni navedena, se bo za nabor podatkov samodejno uporabila privzeta utež (v tem primeru utež za učence). Če imate dober razlog za spremembo spremenljivke teže, jo lahko določite z `weight.var = "SENWGT"`.
- Za izračun odstotkov vrstic, odstotkov stolpcev in skupnih odstotkov nista bila določena nobena dodatna argumenta. Te lahko prav tako izračunate, vendar postane out odveč in težje berljiv.
- Če ne odkljukate polja **Open the output when done,** se bo izhodna datoteka odprla po zaključku vseh izračunov.

4.5 Korelacije

4.5.1 Uvod

Funkcija `lsa.corr` izračuna Pearsonove ali Spearmanove korelacijske koeficiente med spremenljivkami znotraj skupin anketirancev, ki jih določajo razdelitvene spremenljivke (angl.

split variables). Razdelitvene spremenljivke so neobvezne. Če niso navedene nobene razdelitvene spremenljivke, se rezultati izračunajo le na ravni države. Če so navedene razdelitvene spremenljivke, se podatki znotraj vsake države razdelijo v skupine po vseh razdelitvenih spremenljivkah, korelacijski koeficienti pa se izračunajo po kategorijah zadnje razdelitvene spremenljivke. Upoštevati je treba, da se korelacijski koeficienti lahko izračunajo med ozadenjskimi/kontekstualnimi spremenljivkami, med ozadenjskimi/kontekstualnimi spremenljivkami in sklopi PV (angl. *plausible value,* verjetna vrednost) ali med PV, ob upoštevanju kompleksnega vzorčenja in oblikovanja ocenjevanja raziskave. Ko je ozadje/kontekstualna spremenljivka korelirana s sklopom PV, se korelacijski koeficienti izračunajo med ozadjem/kontekstualno spremenljivko in vsako PV v sklopu, nato pa se ocenjene vrednosti vseh PV v sklopu povprečijo in njihova standardna napaka izračuna z uporabo kompleksnih formul, ki bodo odvisne od raziskave. Ko sta dva sklopa PV korelirana, prva PV v prvem sklopu korelira s prvo PV v drugem sklopu, nato druga PV v prvem sklopu korelira z drugo PV v drugem sklopu in tako naprej. Na koncu se ocenjene vrednosti povprečijo in njihova standardna napaka izračuna z uporabo kompleksnih formul, ki bodo odvisne od raziskave. Kakor koli že, standardna napaka se vedno izračuna ob upoštevanju kompleksnega vzorčenja in zasnove raziskave. Če želite podrobnejše informacije o kompleksnem vzorčenju in oblikovanju ocenjevanja določene raziskave ter o tem, kako se ocenjene vrednosti in njihove standardne napake izračunajo, jih poiščite v tehnični dokumentaciji in uporabniškim priročnikom specifične raziskave.

Kot pri drugih funkcijah v paketu RALSA lahko funkcija `lsa.corr` prepozna podatke raziskave in uporabi pravilne tehnike ocenjevanja glede na izvedbo vzorčenja in oblikovanja ocenjevanja raziskave brez dodatnih opravil uporabnika.

4.5.2 Funkcija `lsa.corr` in njeni argumenti

Funkcija `lsa.corr` ima naslednje argumente:

- `data.file` – datoteka, ki vsebuje objekt `lsa.data`. Navedena naj bo bodisi ta bodisi `data.object`, vendar ne oba.
- `data.object` – objekt v pomnilniku, ki vsebuje objekt `lsa.data`. Naveden naj bo bodisi ta bodisi `data.file`, vendar ne oba.
- `split.vars` – lategorijske spremenljivke za deljenje rezultatov. Če niso navedene nobene spremenljivke za deljenje, bodo prikazani rezultati za celotno populacijo posamezne države. Če so navedene ena ali več spremenljivk, bodo rezultati razdeljeni po vseh razen zadnje spremenljivke in odstotki anketirancev bodo izračunani po edinstvenih vrednostih zadnje razdelitvene spremenljivke.
- `bckg.corr.vars` – imena zveznih kontekstualnih spremenljivk za izračun korelacijskih koeficientov. Rezultati bodo izračunani za vse skupine in določeni z razdelitvenimi spremenljivkami.
- `PV.root.corr` – korensko ime za nize možnih vrednosti, za katere je treba izračunati korelacijske koeficiente.
- `corr.type` – niz dolžine ena, ki določa vrsto korelacij, ki jih je treba izračunati, bodisi »Pearson« (privzeto) bodisi »Spearman«.

- `weight.var` – ime spremenljivke, ki vsebuje uteži. Če ime utežne spremenljivke ni navedeno, bo funkcija samodejno izbrala privzeto utež za zagotovljene podatke, odvisno od vrste respondentov.
- `include.missing` – določitev, ali naj se manjkajoče vrednosti razdelitvenih spremenljivk vključijo kot kategorije za deljenje in naj se zanje pripravijo vsi statistični podatki. Privzeto (`FALSE`) obravnava vse primere za razdelitvene spremenljivke brez manjkajočih vrednosti pred izračunom kakršnih koli statističnih podatkov.
- `shortcut` – določitev, ali naj se uporabi metoda »hitro« za IEA TIMSS, TIMSS Advanced, TIMSS Numeracy, eTIMSS, PIRLS, ePIRLS, PIRLS Literacy in RLII. Privzeto (`FALSE`) se uporablja »polna« zasnova pri izračunu komponent variance in standardnih napak ocen.
- `save.output` – določitev, ali naj se output shrani v MS Excel-datoteko (privzeto) ali ne (izpiše v konzolo ali dodeli objektu).
- `output.file` – polna pot do output datoteke, vključno z imenom datoteke. Če je izpuščena, bo datoteka s privzetim imenom »Analysis.xlsx« zapisana v delovni imenik (`getwd()`).
- `open.output` – določitev, ali naj se output odpre po tem, ko je zapisan. Privzeto (TRUE) odpre output v privzetem programu za preglednice, nameščenem v računalniku.

Opombe:

1. Kot vir podatkov mora biti podana bodisi spremenljivka `data.file` bodisi `data.object`. Če sta podani obe, se bo funkcija ustavila z napako.
2. Funkcija izračuna korelacijske koeficiente po kategorijah razdelitvenih spremenljivk. Odstotki anketirancev v vsaki skupini so izračunani znotraj skupin, ki jih določa zadnja razdelitvena spremenljivka. Če ni dodanih nobenih razdelitvenih spremenljivk, se rezultati izračunajo le za državo.
3. Na voljo je lahko več zveznih ozadenjskih spremenljivk in/ali nizov verjetnostnih vrednosti (PV), za katere se izračunajo korelacijski koeficienti. Pomembno je, da v tem primeru rezultati nekoliko odstopajo v primerjavi z uporabo vsakega para ozadenjskih zveznih spremenljivk ali PV v ločenih analizah. To je zato, ker se primeri z manjkajočimi vrednostmi odstranijo vnaprej, in več spremenljivk, kot jih je navedenih za izračun korelacij, več primerov je verjetno odstranjenih. To pomeni, da funkcija podpira samo **listwise** odstranitev.
4. Izračun korelacijskih koeficientov, ki vključujejo možne vrednosti (PV), zahteva navedbo korena imen PV v spremenljivki `PV.root.corr`. Vse raziskave (razen CivED, TEDS-M, SITES, TALIS in TALIS Starting Strong Survey) imajo niz PV v konstruktu (npr., v TIMSS pet za splošno matematiko, pet za algebro, pet za geometrijo itd.). V nekaterih raziskavah (npr. TIMSS in PIRLS) se imena PV v nizu vedno začnejo z znakovnim nizom in končajo s zaporedno številko PV. Npr., imena niza PV za splošno matematiko v TIMSS so: BSMMAT01, BSMMAT02, BSMMAT03, BSMMAT04 in BSMMAT05. Koren PV za ta niz, ki ga je treba dodati v `PV.root.corr`, bo »BSMMAT«. Funkcija bo samodejno poiskala vse spremenljivke v tem nizu in jih vključila v analizo. V drugih raziskavah, kot so OECD PISA, IEA ICCS in ICILS, je zaporedna številka vsake PV vključena na sredino imena. Npr., v ICCS so imena niza PV: PV1CIV, PV2CIV, PV3CIV, PV4CIV in PV5CIV. Koren imena PV je treba določiti

v `PV.root.corr` kot »PV#CIV«. Dodanih je lahko več nizov PV. Upoštevajte pa, da bo določitev več zveznih spremenljivk za argument `bckg.avg.corr` in več korenov PV za argument `PV.root.corr` vplivala na rezultate korelacijskih koeficientov za PV, ker bodo primeri z manjkajočimi vrednostmi v `bckg.corr.vars` odstranjeni, kar bo vplivalo tudi na rezultate iz PV (tj. **listwise** odstranitev). Po drugi strani pa uporaba samo nizov PV za koreliranje ne bi smela vplivati na rezultate za katere koli ocene PV, ker slednje ne bi smele imeti manjkajočih vrednosti.

5. Podano mora biti zadostno število imen spremenljivk (ozadnih/kontekstualnih) ali korenov PV – bodisi dve ozadni spremenljivki bodisi dva korena PV ali mešanica z dolžino dva (tj. ena ozadenska/kontekstualna spremenljivka in en koren PV).
6. Če je `include.missing = FALSE` (privzeto), bodo odstranjeni vsi primeri z manjkajočimi vrednostmi na razdelitvenih spremenljivkah in bodo upoštevani samo primeri z veljavnimi vrednostmi za statistiko. Upoštevajte, da se lahko podatki iz raziskav izvozijo na dva načina: (1) z nastavitvijo vseh uporabniško določenih manjkajočih vrednosti na `NA` ter (2) z uvozom vseh uporabniško določenih manjkajočih vrednosti kot veljavnih in dodajanjem njihovih kod dodatnega atributa vsaki spremenljivki. Če je `include.missing` nastavljena na `FALSE` (privzeto) in so uporabljeni podatki izvoženi z uporabo možnosti (2), bo output odstranil vse vrednosti spremenljivke, ki ustrezajo vrednostim v njenih manjkajočih vrednosti. Sicer pa jih bo vključil kot veljavne vrednosti in zanje izračunal statistiko.
7. Argument `shortcut` velja samo za TIMSS, TIMSS Advanced, TIMSS Numeracy, PIRLS, ePIRLS, PIRLS Literacy in RLII. Prej so te raziskave uporabljale 75 replikatov za izračun standardnih napak, saj je bila ena šola v 75 območjih JK podvojena v svojih utežeh, druga pa izpuščena. Od TIMSS 2015 in PIRLS 2016 naprej uporabljajo te raziskave 150 replikatov, pri čemer ima v vsakem območju JK ena šola enkrat podvojene uteži in enkrat izpuščene, izračuni se torej izvedejo dvakrat za vsako cono. Za več podrobnosti si oglejte tehnično dokumentacijo in uporabniške priročnike TIMSS 2015 in PIRLS 2016. Če je treba ponoviti tabele in slike, je treba argument `shortcut` spremeniti v `TRUE`.
8. Funkcija omogoča **dvosmerni t-test** in **p-vrednosti** za korelacijske koeficiente.

Output, ki ga ustvari funkcija, je shranjen v Excelovi delovni knjigi, ki ima tri liste. Prvi list (»Estimates«) bo vseboval naslednje stolpce, odvisno od tega, katere vrste spremenljivk so bile vključene v analizo:

- **<Country ID>** – stolpec z imeni držav, za katere so izračunane statistike. Natančen naslov stolpca bo odvisen od identifikatorja države, uporabljenega v določeni raziskavi.
- **<Split variable 1>, <Split variable 2>…** – stolpci, ki vsebujejo kategorije, po katerih so bile statistike razdeljene. Natančna imena bodo odvisna od spremenljivk v `split.vars`.
- **n_Cases** – število primerov v vzorcu, uporabljenih za izračun statistike.
- **Sum_<Weight variable>** – ocenjena populacija števila elementov na skupino po uporabi uteži. Dejanski naziv spremenljivke uteži bo odvisen od uporabljene uteži spremenljivke v analizi.
- **Sum_<Weight variable>_SE** – standardna napaka ocenjenega števila elementov populacije na skupino. Dejanski naziv spremenljivke uteži bo odvisen od uporabljene uteži v analizi.

- **Percentages_<Last split variable>** – odstotki anketirancev (ocena populacije) na skupine, določene z razdelitvenimi spremenljivkami v `split.vars`. Odstotki bodo prikazani za zadnjo razdelitveno spremenljivko, ki določa končne skupine.
- **Percentages_<Last split variable>_SE** – standardne napake odstotkov iz zgornjega podatka.
- **Variable** – imena spremenljivk (ozadnih/kontekstualnih ali korenov PV), ki jih je treba primerjati z vrsticami v naslednjih stolpcih, da se tvorijo korelacijske matrike.
- **Correlation_<Background variable>** – korelacijski koeficient za vsako ozadenjsko spremenljivko, določeno v `bckg.corr.vars`, proti sami sebi in vsaki od spremenljivk v stolpcu »Variable«. Za vsako spremenljivko, določeno v `bckg.corr.vars` in/ali nizu PV, določenem v `PV.root.corr`, bo en stolpec z oceno korelacijskega koeficienta.
- **Correlation_<Background variable>_SE** – standardna napaka korelacije za vsako ozadenjsko spremenljivko, določeno v `bckg.corr.vars`. Za vsako spremenljivko, določeno v `bckg.corr.vars` in/ali nizu PV, določenem v `PV.root.corr`, bo en stolpec s standardno napako ocene korelacijskega koeficienta.
- **Correlation_<root PV>** – korelacijski koeficient za vsak niz PV, določen kot koren PV v `PV.root.corr`, proti samemu sebi in vsaki spremenljivki v stolpcu „Variable«. Za vsak niz PV, določen v `PV.root.corr`, in vsako drugo ozadno spremenljivko ali niz PV bo en stolpec z oceno korelacijskega koeficienta.
- **Correlation_<root PV>_SE** – standardna napaka korelacije za vsak niz PV, določen kot koren PV v `PV.root.corr`. Za vsak niz PV bo en stolpec s standardno napako ocene korelacije s katero koli drugo spremenljivko ali nizom PV.
- **Correlation_<root PV>_SVR** – komponenta vzorčne variance za korelacijo med PV z istim korenom in PV, določeno v `PV.root.corr`. Za vsako oceno korelacije bo en stolpec s komponento vzorčne variance.
- **Mean_<root PV>_MVR** – komponenta merilne variance za korelacijo PV z istim korenom PV, določenim v `PV.root.corr`. Za vsako oceno korelacije bo en stolpec s komponento merilne variance.
- **Correlation_<Background variable>_SVR** – komponenta vzorčne variance za korelacijo določene ozadne spremenljivke z nizom PV, s katerimi je korelirana. Za povprečno oceno za vsako ozadenjsko/kontekstualno spremenljivko bo en stolpec s komponento vzorčne variance.
- **Correlation_<Background variable>_MVR** – komponenta merilne variance za korelacijo določene ozadenjske spremenljivke z nizom PV, s katerimi korelira. Za vsako ozadno spremenljivko bo en stolpec s komponento merilne variance.
- **t_<root PV>** – t-testna vrednost za korelacijske koeficiente niza PV, ko jih koreliramo z drugimi spremenljivkami (ozadenjskimi/kontekstualnimi ali drugimi nizi PV).
- **t_<Background variable>** – t-testna vrednost za korelacijske koeficiente ozadenskih spremenljivk, ko jih koreliramo z drugimi spremenljivkami (ozadenskimi/kontekstualnimi ali drugimi nizi PV).
- **p_<root PV>** – p-vrednost za korelacijske koeficiente niza PV, ko jih koreliramo z drugimi spremenljivkami (ozadenjskimi/kontekstualnimi ali drugimi nizi PV).
- **p_<Background variable>** – p-vrednost za korelacijske koeficiente ozadenjskih spremenljivk, ko jih koreliramo z drugimi spremenljivkami (ozadenjskimi/kontekstualnimi ali drugimi nizi PV).

Drugi list (»Analysis information«) vsebuje dodatne informacije, povezane z analizo po državah, v naslednjih stolpcih:

- **DATA** – uporabljen `data.file` ali `data.object`.
- **STUDY** – iz katere raziskave izhajajo podatki.
- **CYCLE** – kateri cikel raziskave podatki predstavljajo.
- **WEIGHT** – katera utežna spremenljivka je bila uporabljena.
- **DESIGN** – katera tehnika ponovnega vzorčenja je bila uporabljena (JRR ali BRR).
- **SHORTCUT** – ali je bila uporabljena bližnjica.
- **NREPS** – koliko uteži za replikacijo je bilo uporabljenih.
- **ANALYSIS_DATE** – datum, ko je bila analiza izvedena.
- **START_TIME** – čas, kdaj se je analiza začela.
- **END_TIME** – čas, kdaj se je analiza končala.
- **DURATION** – trajanje analize v urah, minutah, sekundah in milisekundah.

Tretji list (»Calling syntax«) vsebuje klic funkcije z vrednostmi za vse parametre, kot so bili izvedeni. To je uporabno, če je treba analizo kasneje ponoviti.

4.5.3 Izračun koeficientov korelacije s pomočjo ukazne vrstice

V spodnjih primerih bomo združili novo podatkovno datoteko s podatki o učencih in ravnateljih šol iz PIRLS 2016 (Avstralija in Slovenija), pri čemer bomo uporabili vse spremenljivke iz obeh vrst datotek:

```
lsa.merge.data(inp.folder = "C:/temp",
               file.types = list(acg = NULL, asg = NULL),
               ISO = c("aus", "svn"),
               out.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData")
```

Za začetek bomo izračunali koeficient korelacije med dvema lestvicama v ozadju – Občutek pripadnosti šoli (ASBGSSB) in Medvrstniško nasilje (ASBGSB) v Avstraliji in Sloveniji (za informacije o konstrukciji teh lestvic in njihovih lastnostih preverite tehnično dokumentacijo PIRLS 2016):

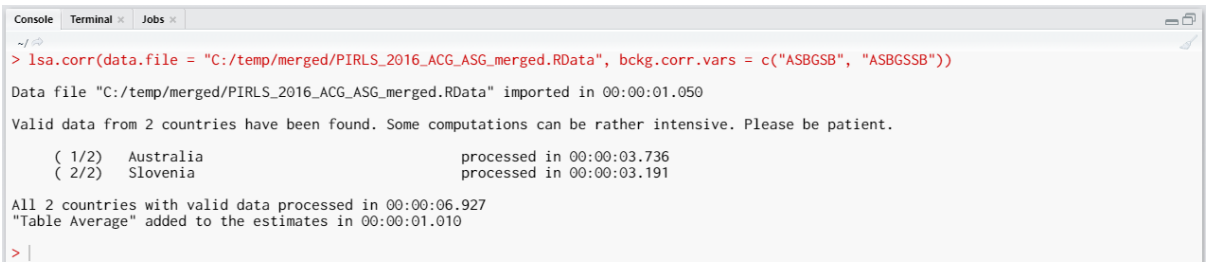
```
lsa.corr(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
         bckg.corr.vars = c("ASBGSB", "ASBGSSB"))
```

Nekaj pomembnih stvari:

1. Naenkrat lahko korelirate več kot dve spremenljivki v ozadju in/ali nize PV. Koeficienti korelacije bodo izračunani za vsak par posebej. Vendar pa bodo vsi primeri z manjkajočimi vrednostmi na vseh spremenljivkah odstranjeni vnaprej. To lahko spremeni rezultate v primerjavi z izvedbo analiz za vsak par posebej.

2. Pri mednarodnih raziskavah morajo biti vse analize izvedene ločeno po državah. Vendar ni treba dodajati identifikatorja države (IDCNTRY ali CNT v PISA) kot razdelitvene spremenljivke. Funkcija jo bo samodejno prepoznala in dodala v vektor `split.vars`.
3. Privzeta metoda korelacije je Pearsonov koeficient korelacije, ki bo uporabljen, če argument `corr.type` ni določen.
4. Ni potrebe po izrecnem določanju utežne spremenljivke. Če utežna spremenljivka ni določena, bo uporabljena privzeta utež (v tem primeru skupna utež za učence), odvisno od združenih podatkov o anketiranih, prepoznana pa je samodejno. Če imate dober razlog za spremembo utežne spremenljivke, lahko to storite z dodajanjem argumenta `weight.var = "SENWGT"`.
5. Če ne dodate izrecno `save.output = FALSE`, bo output shranjen na disk v obliki datoteke MS Excel. V nasprotnem primeru bo rezultat izpisan v konzolo.
6. Če output datoteka ni določena, bo izhod shranjen pod imenom »Analysis.xlsx« v delovnem imeniku (ki ga lahko prikličete z ukazom `getwd()`).
7. Če v sintakso izrecno ne dodate `open.output = FALSE`, se bo output datoteka odprla po končanih izračunih. To je koristno, ko se izvaja več sintaks za različne analize in ni potrebna takojšnja preučitev rezultatov.

Zagon zgornje kode bo v konzoli RStudia izpisal naslednji izhod:



Ko bodo vse operacije zaključene, bo izhod shranjen na disk kot MS Excel-delovni zvezek. Če je `open.output = TRUE` (privzeto), se bo datoteka odprla v privzetem programu za preglednice (običajno MS Excel).

Izračunajmo korelacijo med zapleteno lestvico v ozadju »Medvrstniško nasilje« (ASBGSB; za informacije o konstrukciji te lestvice in njenih lastnostih preverite tehnično dokumentacijo PIRLS 2016) ter nizom PV za rezultat iz branja:

```
lsa.corr(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
         bckg.corr.vars = "ASBGSB", PV.root.corr = "ASRREA")
```

Sintaksa zgornjega klica je podobna prejšnji, vendar vključuje le eno spremenljivko v ozadju (ASBGSB; učenci, ki so tarče nasilja) ter doda argument `PV.root.corr` in njegovo vrednost, PV za skupni dosežek pri branju. Bodite pozorni na to, kako so PV določene. Pet PV za bralni dosežek: ASRREA01, ASRREA02, ASRREA03, ASRREA04 in ASRREA05. V argumentu `PV.root.corr` moramo določiti le koren PV, »ASRREA«. Funkcija bo uporabila ta skupni koren za izbiro vseh petih PV in jih vključila v izračune.

Zagon zgornje sintakse bo prepisal prejšnji output, saj ima definirano isto ime datoteke (v konzoli bo prikazano opozorilo). Stolpci na listu »Estimates« bodo zdaj drugačni.

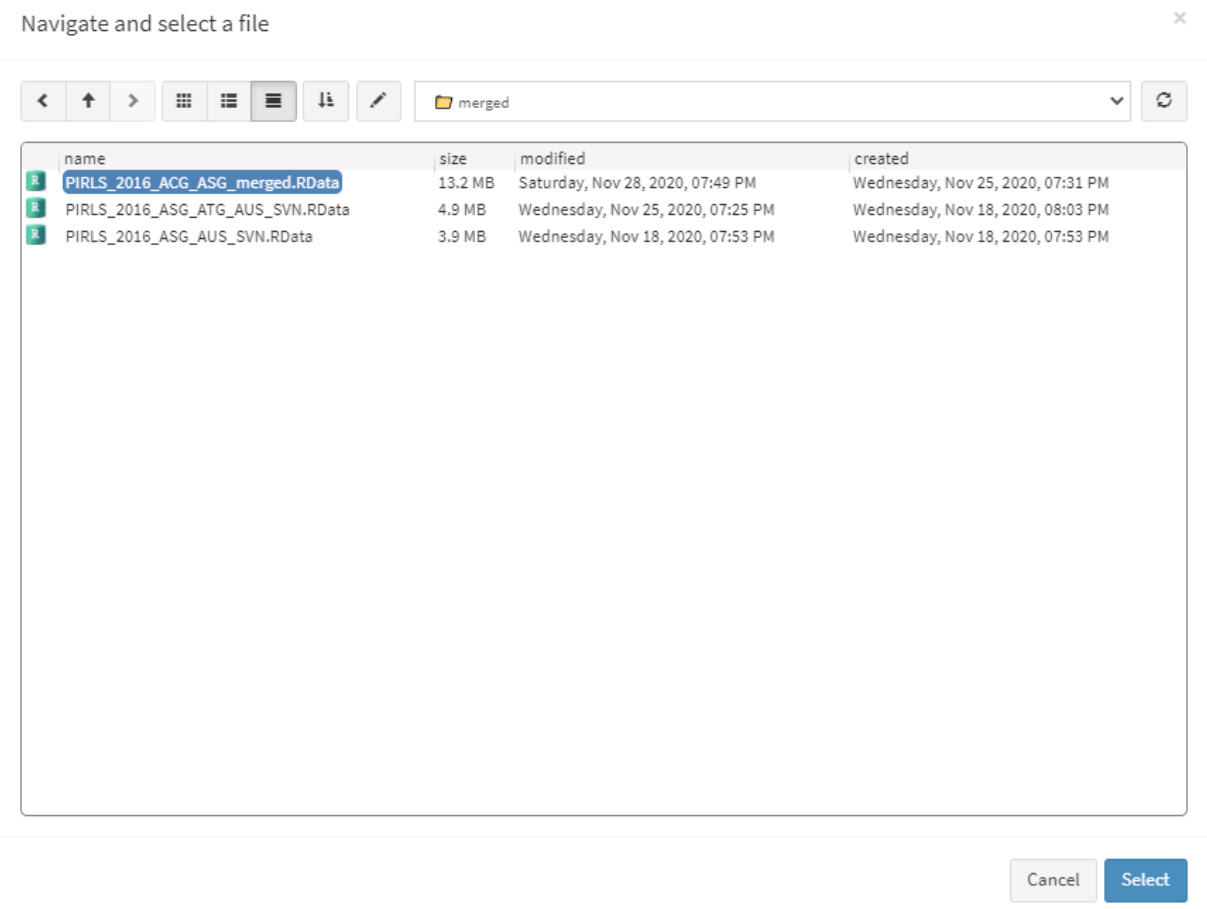
4.5.4 Izračun korelacijskih koeficientov z uporabo grafičnega vmesnika (GUI)

Za zagon uporabniškega vmesnika RALSA v RStudiu izvedite naslednji ukaz:

```
ralsaGUI()
```

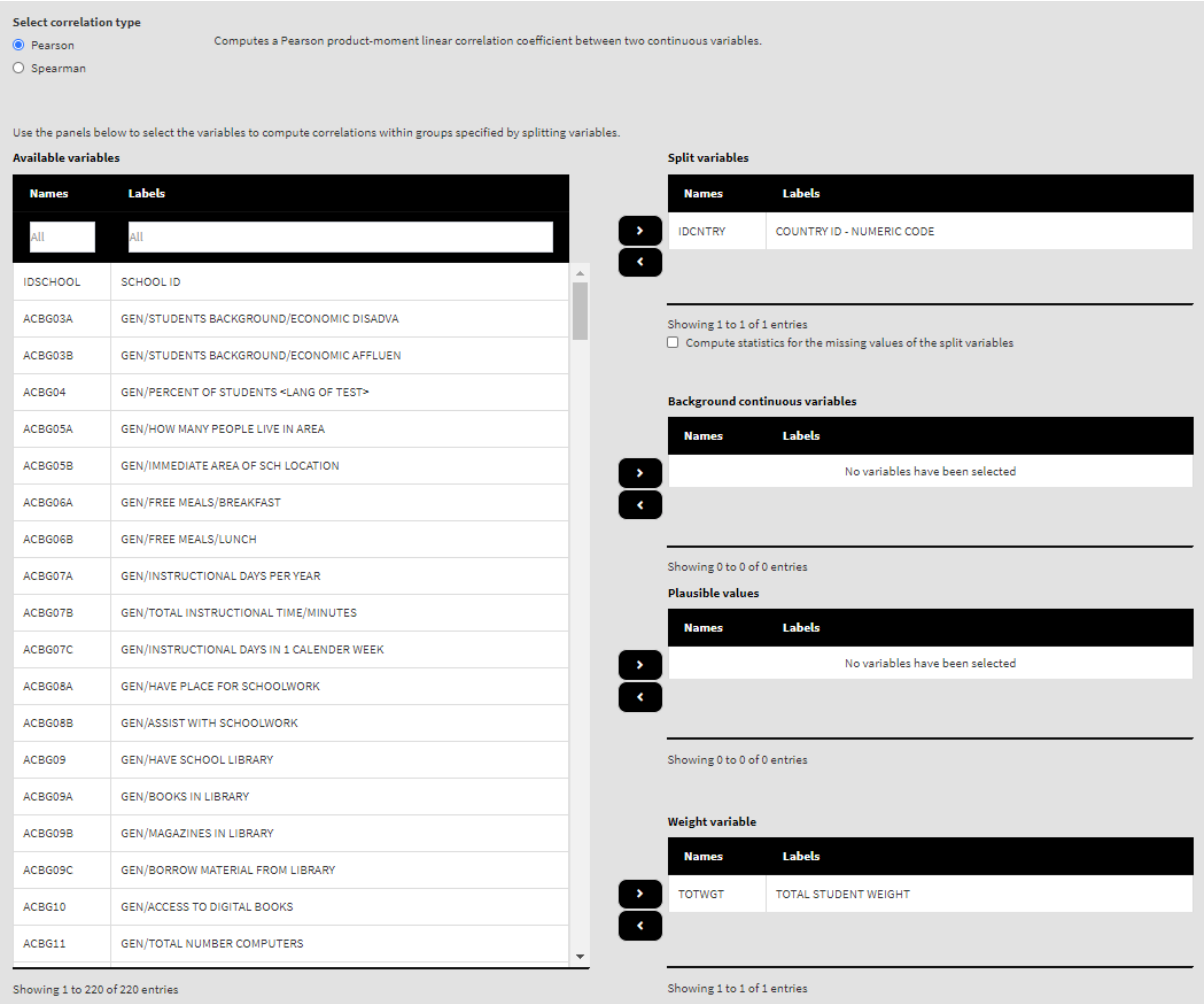
Za primere, ki sledijo, združite novo datoteko s podatki PIRLS 2016 za Avstralijo in Slovenijo ter vključite vse spremenljivke učencev in ravnateljev. Združeno datoteko lahko poimenujete **»PIRLS_2016_ACG_ASG_merged.RData«**.

Ko zaključite združevanje podatkov, v meniju na levi izberite **Analysis types > Correlations**. Ko se pomaknete na stran **Correlations** v GUI, kliknite gumb **Choose data file**. Nato poiščite mapo, ki vsebuje združeno datoteko **»PIRLS_2016_ACG_ASG_merged.RData«**, jo izberite in kliknite gumb **Select**.



Ko je datoteka naložena, boste na levi strani videli panel s seznamom razpoložljivih spremenljivk in niz panelov na desni strani, kamor lahko dodate spremenljivke s seznama razpoložljivih spremenljivk. Nad paneli boste prav tako videli informacije o naloženi datoteki

ter gumbe za izbiro vrste korelacij – Pearson ali Spearman. Ker bomo izračunali korelacijske koeficiente med zveznimi spremenljivkami, pustimo privzeto izbrano možnost, **Pearson**.

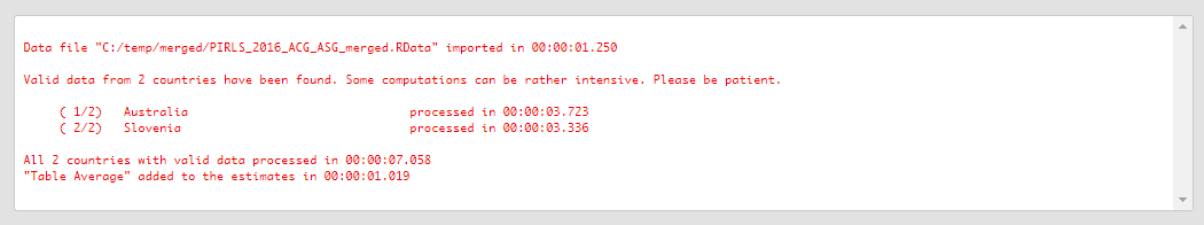


Z miško izberite spremenljivke s seznama **Razpoložljive spremenljivke (Available variables)** in jih s pomočjo puščičnih gumbov na sredini zaslona dodajte v različna polja (ali jih odstranite), da nastavite parametre za analizo. Za hitrejše iskanje spremenljivk lahko uporabite filtre, ki so na vrhu panelov.

Za začetek bomo izračunali korelacijski koeficient med dvema ozadnima lestvicama – »Občutek pripadnosti šoli učencev« (ASBGSSB) in »Medvrstniško nasilje« (ASBGSB) v Avstraliji in Sloveniji (za informacije o tem, kako so te lestvice oblikovane in kakšne so njihove lastnosti, preverite tehnično dokumentacijo PIRLS 2016). Izberite spremenljivki ASBGSSB in ASBGSB s seznama **Razpoložljive spremenljivke** ter ju s pomočjo desne puščice dodajte na seznam Neprekinjene ozadne spremenljivke (**Background continuous variables**).

Ko ste to storili, se pomaknite navzdol in kliknite gumb Določi ime output datoteke (**Define output file name**). Pomaknite se do mape **»C:/temp/Results«** (ali do mape, kamor želite shraniti output) in določite ime output datoteke. Ko to storite, se bo poleg gumba pojavilo potrditveno polje, kjer lahko označite, ali naj se output odpre po zaključku izračunov. Pod tem bo prikazana sintaksa, še nižje pa gumb Izvedi sintakso (**Execute syntax**). Končne nastavitve v spodnjem delu zaslona bodo izgledale takole:

Kliknite na gumb **Execute syntax**. Konzola GUI se bo pojavila na dnu in beležila vse dokončane operacije.

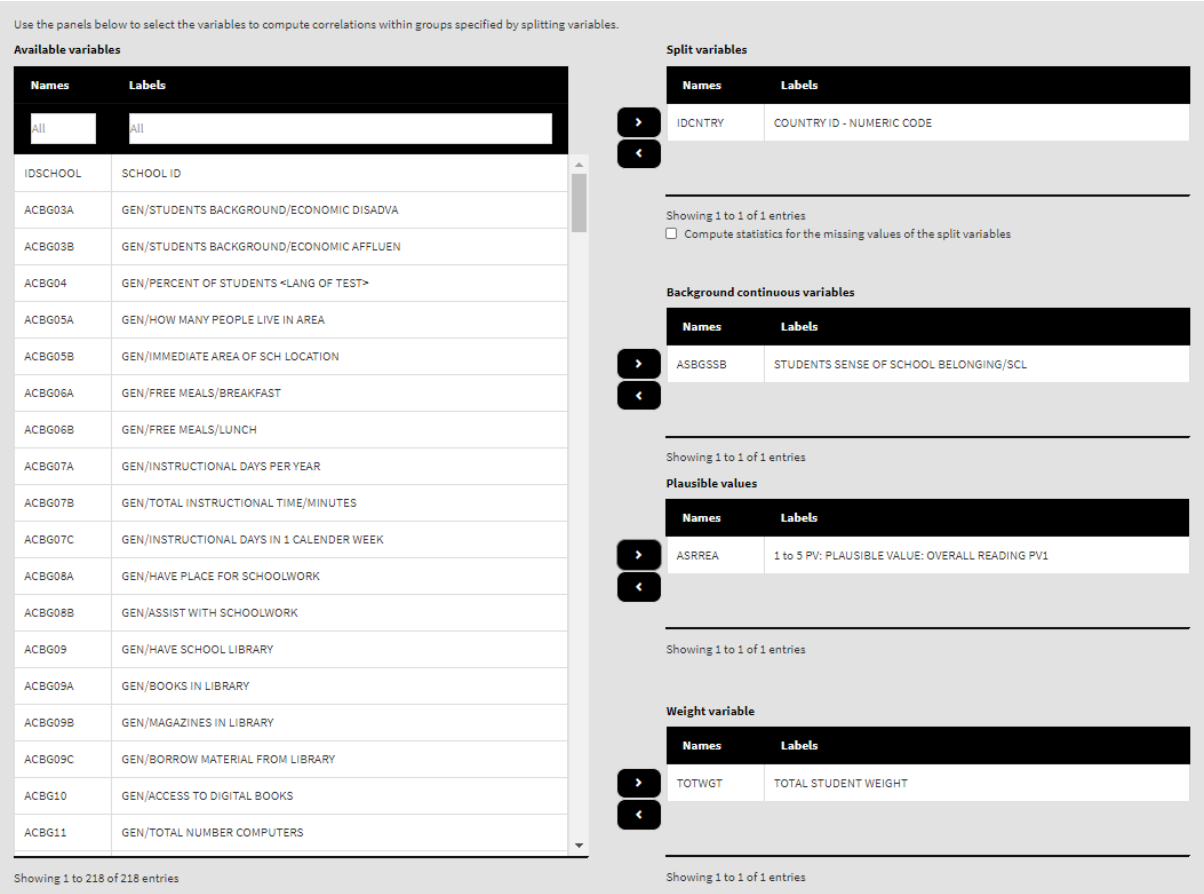


Nekaj stvari, ki jih je treba upoštevati:

- Več kot dve spremenljivki ozadja in/ali nabori PV se lahko povežejo hkrati. Korelacijski koeficienti bodo izračunani za vsak par posebej. Vendar pa bodo vsi primeri z manjkajočimi vrednostmi na vseh spremenljivkah odstranjeni vnaprej. To lahko spremeni rezultate v primerjavi z izvajanjem analiz za vsak par posebej.
- V mednarodnih raziskavah je treba vse analize izvajati ločeno po državah. Spremenljivka ID države (IDCNTRY ali CNT v PISA) je vedno izbrana kot prva spremenljivka za razdelitev in je ne morete odstraniti s panela za razdelitev spremenljivk.
- Privzeta metoda korelacije je Pearsonov koeficient korelacije. Če je treba korelirati kategorialne (ordinalne) spremenljivke, je priporočljivo uporabiti Spearmanovo korelacijo.
- Privzeta utežna spremenljivka je samodejno izbrana in dodana v panel za uteži. Lahko jo spremenite z drugo utežno spremenljivko, ki je na voljo v naboru podatkov. Če je izbrana privzeta utež, ne bo prikazana v oknu za sintakso. Če v panelu za uteži ni izbrana nobena spremenljivka, bo samodejno uporabljena privzeta utež.
- Možnost **Use shortcut method for computing SE** ni privzeto označena. To bo funkciji omogočilo, da izračuna stadardno napako z uporabo metode »full« za komponento variabilnosti vzorca.
- Če je označen okvirček **Open the output when done**, se bo output samodejno odprl v privzetem programu za preglednice (ponavadi MS Excel), ko bodo vsi izračuni zaključeni. Oglejte si razlage o strukturi delovnega zvezka, njenih listih in stolpcih tukaj.

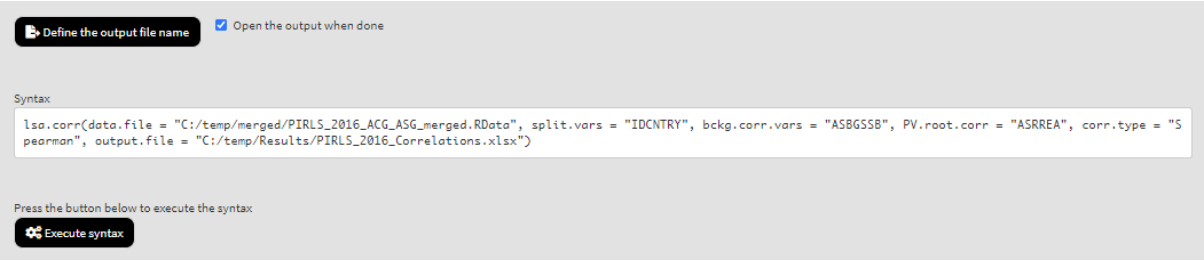
Izračunajmo korelacijo med spremenljivkami medvrstniško nasilje (ASBGSB) in naborom PV za skupne rezultate branja učencev. Izberite ASBGSSB (Občutek pripadnosti šoli) na seznamu **Background continuous variables** in ga premaknite nazaj na seznam **Available variables** z uporabo levega puščičnega gumba. Na seznamu **Available variables** poiščite koren skupnih bralnih PV (ASRREA). Za iskanje lahko uporabite filtre na vrhu panela, bodisi

po imenu bodisi po oznaki. Izberite koren in ga dodajte na panel **Plausible values** z uporabo puščic. Ta del vmesnika naj izgleda takole:

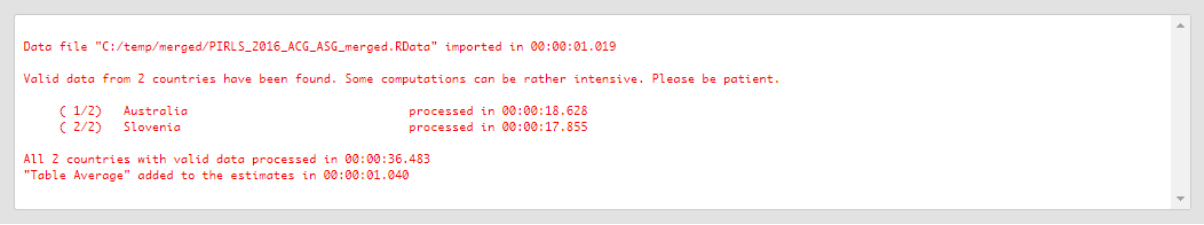


Poglejte, kako so navedene PV. Pet PV za skupni bralni dosežek je ASRREA01, ASRREA02, ASRREA03, ASRREA04 in ASRREA05. Seznama **Available variables** in **Plausible values** ne bosta prikazovala petih ločenih PV, temveč le njihov koren/skupno ime – ASRREA, brez številk na koncu. V ozadju bo funkcija vzela vseh pet PV in jih vključila v izračune.

Ker aplikacijo za **Correlations** uporabljamo neposredno po izvedbi prejšnje analize, imamo še vedno v rabi nastavitve iz prejšnje analize. Ni treba spreminjati nobenih preostalih nastavitev, razen če želite. Lahko pa spremenite ime output datoteke, sicer bo prepisana. Bodite pozorni na to, da se bo prikazana sintaksa spremenila, kar odraža odstranitev spremenljivke ASBGSSB in vključitev korena petih PV, ASRREA, za skupni dosežek iz branja, za izračun korelacij.



Kliknite na gumb **Execute syntax**. Konzola GUI se bo posodobila in zabeležila vse zaključene operacije.



Če je potrjen okvirček **Open the output when done**, se bo output samodejno odprl v privzetem pregledovalniku preglednic (ponavadi MS Excel), ko bodo vsi izračuni zaključeni.

4.6 Linearna regresija

4.6.1 Uvod

Funkcija `lsa.lin.reg` izračuna koeficiente linearne regresije (Ordinary Least Squares – OLS) znotraj skupin respondentov, ki jih določajo spremenljivke za razdelitev (angl. *split variables*). Spremenljivke za razdelitev so neobvezne. Če niso določene nobene spremenljivke za razdelitev, se rezultati izračunajo le na ravni države. Če so te določene, se podatki znotraj vsake države razdelijo v skupine po vseh spremenljivkah za razdelitev, koeficienti linearne regresije pa se izračunajo za zadnjo spremenljivko za razdelitev. Upoštevajte, da se koeficienti linearne regresije lahko izračunajo s kontekstualnimi spremenljivkami (angl. *background variables*) kot odvisnimi ali s PV (angl. *plausible values,* verjetne vrednosti) kot odvisnimi spremenljivkami. Neodvisne spremenljivke so lahko tako kontekstualne spremenljivke kot tudi nizi PV. Vse analize bodo upoštevale kompleksno vzorčenje in zasnovo dotične raziskave.

Ko se nizi PV uporabljajo bodisi kot odvisne bodisi kot neodvisne spremenljivke, se koeficienti regresije izračunajo med kontekstualnimi spremenljivkami in vsako PV v nizu, nato pa se ocenjeni rezultati za vse PV v nizu povprečijo in njihova standardna napaka izračuna s kompleksnimi formulami, ki so odvisne od raziskave. Ko se nizi PV uporabljajo tako kot odvisne kot tudi neodvisne spremenljivke, se prva PV v nizu odvisnih PV regresira na prvo PV v nizu(n) neodvisnih PV, nato se druga PV v nizu odvisnih PV regresira na drugo PV v nizu(n) neodvisnih PV itd. Na koncu se ocenjeni rezultati povprečijo in njihova standardna napaka izračuna s kompleksnimi formulami, ki so odvisne od raziskave.

Kakor koli že, standardna napaka bo izračunana ob upoštevanju kompleksnih vzorčnih in ocenjevalnih zasnov raziskav. Če vas zanimajo podrobnosti o kompleksnih vzorčnih in ocenjevalnih zasnovah določene raziskave ter o tem, kako so ocenjeni rezultati in njihova standardna napaka izračunani, preverite tehnično dokumentacijo in uporabniški priročnik raziskave.

Kot katera koli druga funkcija v paketu RALSA funkcija `lsa.lin.reg` prepozna podatke raziskave in uporabi pravilne tehnike ocenjevanja glede na vzorčno in ocenjevalno zasnovo raziskave brez dodatne pozornosti.

4.6.2 Funkcija linearne regresije in njeni argumenti

Funkcija lsa.lin.reg ima naslednje argumente:

- `data.file` – datoteka, ki vsebuje objekt `lsa.data`. Določiti je treba bodisi to bodisi `data.object`, vendar ne oba.
- `data.object` – objekt v pomnilniku, ki vsebuje objekt `lsa.data`. Določiti je treba bodisi tega bodisi `data.file`, vendar ne oba.
- `split.vars` –kategorijske spremenljivke za razdelitev rezultatov. Če spremenljivke za razdelitev niso določene, bodo rezultati prikazani za celotno populacijo držav. Če je določena ena ali več spremenljivk, bodo rezultati razdeljeni po vseh razen zadnji spremenljivki in odstotki respondentov bodo izračunani glede na edinstvene vrednosti zadnje spremenljivke za razdelitev.
- `bckg.dep.var` – ime ozadenjske ali kontekstualne spremenljivke, uporabljene kot odvisna spremenljivka v modelu.
- `PV.root.dep` – korensko ime niza verjetnostnih vrednosti, uporabljenih kot odvisna spremenljivka v modelu.
- `bckg.indep.cont.vars` – imena zveznih neodvisnih ozadenjskih/kontekstualnih spremenljivk, uporabljenih kot napovedniki v modelu.
- `bckg.indep.cat.vars` – imena kategorijskih neodvisnih ozadenjskih/kontekstualnih spremenljivk, uporabljenih kot napovednikih v modelu za izračun kontrastov (glejte `bckg.cat.contrasts` in `bckg.ref.cats`).
- `bckg.cat.contrasts` – niz vektorjev z dolžino, enako dolžini `bckg.indep.cat.vars`, ki določa vrsto kontrastov za izračun v primeru, da so `bckg.indep.cat.vars` določene.
- `bckg.ref.cats` – vektor celega števila, enak dolžini `bckg.indep.cat.vars` in `bckg.cat.contrasts`, ki določa referenčne kategorije za kontraste za izračun v primeru, da so `bckg.indep.cat.vars` določene.
- `PV.root.indep` – korenska imena niza verjetnostnih vrednosti, uporabljenih kot neodvisne spremenljivke v modelu.
- `interactions` – interakcijski termini – seznam, ki vsebuje vektorje dolžine dva.
- `standardize` – določitev, ali naj bodo odvisne in neodvisne spremenljivke standardizirane za pridobitev beta-koeficientov. Privzeto je `FALSE`.
- `weight.var` – ime spremenljivke, ki vsebuje uteži. Če ime utežne spremenljivke ni določeno, bo funkcija samodejno izbrala privzeto utež za zagotovljene podatke, odvisno od vrste respondentov.
- `include.missing` – določitev, ali naj se manjkajoče vrednosti spremenljivk za razdelitev vključijo kot kategorije za razdelitev in zanje izdelajo vse statistike. Privzeto (`FALSE`) upošteva vse primere spremenljivk za razdelitev brez manjkajočih vrednosti pred izračunom kakršnih koli statistik.
- `shortcut` – določitev, ali naj se za izračun SE za IEA TIMSS, TIMSS Advanced, TIMSS Numeracy, PIRLS, ePIRLS, PIRLS Literacy in RLII uporabi metoda »kratka pot«. Privzeto (`FALSE`) uporablja »polno« zasnovo pri izračunu komponent varianc in standardnih napak ocen.
- `save.output` – določitev, ali naj se output shrani v MS Excel-datoteko (privzeto) ali ne (izpis v konzoli ali dodelitev objektu).

- `output.file` – polna pot do output datoteke, vključno z imenom datoteke. Če je izpuščena, bo datoteka z privzetim imenom »Analysis.xlsx« shranjena v delovni imenik (`getwd()`).
- `open.output` – določitev, ali naj se output odpre po tem, ko je bil zapisan. Privzeto (`TRUE`) odpre output v privzetem programu za preglednice, nameščenem na računalniku.

Opombe:

1. Uporabiti je treba bodisi `data.file` bodisi `data.object` kot vir podatkov. Če sta navedena oba, se bo funkcija ustavila z napako.
2. Funkcija koeficiente linearne regresije izračuna po kategorijah spremenljivk za razdelitev. Odstotki respondentov v vsaki skupini se izračunajo znotraj skupin, ki jih določa zadnja spremenljivka za razdelitev. Če spremenljivke za razdelitev niso dodane, bodo rezultati izračunani le po državah.
3. Če je `standardize = TRUE`, se bodo spremenljivke standardizirale pred izračunom statistik za pridobitev beta-koeficientov regresije.
4. Kot odvisna spremenljivka se lahko poda bodisi kontekstualna spremenljivka (`bckg.dep.var`) bodisi korensko ime niza verjetnostnih vrednosti (`PV.root.dep`), vendar ne oboje.
5. Kontekstualne spremenljivke, poslane v `bckg.indep.cont.vars`, bodo obravnavane kot numerične spremenljivke v modelu. Spremenljivke z diskretnim številom kategorij (tj. faktorjev), poslane v `bckg.indep.cat.vars`, bodo uporabljene za izračun kontrastov. V tem primeru je treba vrsto kontrasta posredovati v `bckg.cat.contrasts` in število referenčnih kategorij za vsako od `bckg.indep.cat.vars`. Število vrst kontrastov in referenčne kategorije morajo biti enake številu `bckg.indep.cat.vars`. Trenutno podprti načini kodiranja kontrastov so:
    a. `dummy` – presečišče je povprečje odvisne spremenljivke za respondente, ki izberejo referenčno kategorijo, nakloni so razlike med presečiščem in povprečjem respondentov na odvisni spremenljivki, ki izbirajo druge kategorije;
    b. `deviation` – presečišče je skupno povprečje na odvisni spremenljivki ne glede na skupino in nakloni so razlike med presečiščem ter povprečjem respondentov na odvisni spremenljivki, ki izbirajo druge kategorije razen referenčne.
    c. `simple` – enako kot pri dummy kodiranju kontrastov, razen presečišča, ki je v tem primeru skupno povprečje.
6. Ko se uporablja `standardize = TRUE`, kodiranje kontrastov `bckg.indep.cat.vars` ni standardizirano. Tako koeficienti regresije morda ne bodo primerljivi z drugimi rešitvami programske opreme za analizo podatkov mednarodnih raziskav, ki npr. uporabljajo SPSS ali SAS, kjer se kodiranje kontrastov kategorijskih spremenljivk (npr. dummy kodiranje) izvaja privzeto. Vendar bodo statistike modela identične.
7. Več zveznih ali kategorijskih ozadenjskih spremenljivk in/ali nizi verjetnostnih vrednosti se lahko predložijo za izračun koeficientov regresije. Upoštevajte, da se bodo rezultati v tem primeru nekoliko razlikovali v primerjavi z uporabo vsakega para enakih ozadenjskih zveznih spremenljivk ali PV v ločeni analizi. To je zaradi tega, ker se

primeri z manjkajočimi vrednostmi odstranijo vnaprej. Več spremenljivk, kot se poda, več primerov je verjetno odstranjenih. Funkcija podpira le **listwise** odstranitev.

8. Izračun koeficientov regresije, ki vključujejo verjetne vrednosti, zahteva zagotavljanje korena imen verjetnostnih vrednosti v `PV.root.dep` in/ali `PV.root.indep`. Vse raziskave (razen CivED, TEDS-M, SITES, TALIS in TALIS Starting Strong Survey) imajo niz PV za vsak konstrukt (npr., v TIMSS pet za splošno matematiko, pet za algebro, pet za geometrijo itd.). V nekaterih raziskavah (npr. TIMSS in PIRLS) se imena PV v nizu vedno začnejo z znakom in končajo z zaporedno številko PV. Npr., imena niza PV za splošno matematiko v TIMSS so BSMMAT01, BSMMAT02, BSMMAT03, BSMMAT04 in BSMMAT05. Koren PV za ta niz, ki ga je treba dodati `PV.root.dep` ali `PV.root.indep`, bo »BSMMAT«. Funkcija bo samodejno našla vse spremenljivke v tem nizu PV in jih vključila v analizo. V drugih raziskavah, kot so OECD PISA in IEA ICCS ter ICILS, je zaporedna številka vsake PV vključena v sredini imena. Npr., v ICCS so imena niza PV PV1CIV, PV2CIV, PV3CIV, PV4CIV in PV5CIV. Koren imena PV mora biti določen v `PV.root.dep` ali `PV.root.indep` kot »PV#CIV«. Več kot en niz PV se lahko doda v `PV.root.indep`.
9. Funkcija lahko prav tako izračuna dvočlenske interakcijske učinke med neodvisnimi spremenljivkami z dostavo seznama v argument `interactions`. Seznam mora vsebovati vektorje dolžine dva in vse spremenljivke v teh vektorjih morajo biti prav tako posredovane kot neodvisne spremenljivke. Upoštevajte naslednje:
    a. Ko je interakcija med dvema zveznima ozadenjskima spremenljivkama (oboje je torej poslano v `bckg.indep.cont.vars`), bo učinek interakcije izračunan med njima takšen, kakršen je.
    b. Ko je interakcija med dvema kategorijskima spremenljivkama (oboje je poslano v `bckg.indep.cat.vars`), bo učinek interakcije izračunan med vsakim možnim parom kategorij obeh spremenljivk razen referenčnih kategorij.
    c. Ko je interakcija med eno zvezno (tj. poslano v `bckg.indep.cont.vars`) in eno kategorijsko (tj. poslano v `bckg.indep.cat.vars`), bo učinek interakcije izračunan med zvezno spremenljivko in vsako kategorijo kategorijske spremenljivke razen referenčne kategorije.
    d. Ko je interakcija med zvezno spremenljivko (tj. poslano v `bckg.indep.cont.vars`) in nizom PV (tj. poslanim v `PV.root.indep`), bo učinek interakcije izračunan med zvezno spremenljivko in vsako PV v nizu, rezultati pa bodo združeni.
    e. Ko je interakcija med kategorijsko spremenljivko (tj. poslano v `bckg.indep.cat.vars`) in nizom PV (tj. poslanim v `PV.root.indep`), bo učinek interakcije izračunan med vsako kategorijo kategorijske spremenljivke (razen referenčne kategorije) in vsako PV v nizu. Rezultati bodo združeni za vsako kategorijo kategorijskih spremenljivk in niz PV.
    f. Ko je interakcija med dvema nizoma PV (tj. poslanima v `PV.root.indep`), bo učinek interakcije izračunan med prvo PV v prvem nizu in prvo PV v drugem nizu, drugo PV v prvem nizu in drugo PV v drugem nizu itd. Rezultati bodo nato združeni.
10. Razen če izrecno dodate `save.output = FALSE`, bo output zapisan v MS Excel na disk. V nasprotnem primeru bo izpisan v konzolo.
11. Če ni navedena output datoteka, bo izhod shranjen z imenom »Analysis.xlsx« v delovnem imeniku (kar lahko prikličete z `getwd()`).

12. Če je `include.missing = FALSE` (privzeto), bodo vsi primeri z manjkajočimi vrednostmi na spremenljivkah za razdelitev odstranjeni in v statistikah ohranjeni le primeri z veljavnimi vrednostmi. Upoštevajte, da se podatki iz raziskav lahko izvozijo na dva načina: (1) nastavljanje vseh uporabniško določenih manjkajočih vrednosti na NA; (2) uvoz vseh uporabniško določenih manjkajočih vrednosti kot veljavnih in dodajanje njihovih kod v dodatni atribut vsake spremenljivke. Če je `include.missing` nastavljeno na `FALSE` (privzeto) in se uporabljajo podatki, izvoženi z možnostjo (2), bo output odstranil vse vrednosti iz spremenljivke, ki ustreza vrednostim v njenem atributu manjkajočih vrednosti. V nasprotnem primeru bodo vključene kot veljavne vrednosti in zanje bodo izračunane statistike.
13. Argument `shortcut` je veljaven samo za TIMSS, TIMSS Advanced, TIMSS Numeracy, PIRLS, ePIRLS, PIRLS Literacy in RLII. Prej so te raziskave pri izračunu standardnih napak uporabljale 75 replikacij, ker je imela ena šola v 75 JK območjih svoje uteži podvojene, druga pa je bila odstranjena. Od TIMSS 2015 in PIRLS 2016 naprej pa raziskave uporabljajo 150 replikacij in v vsakem JK območju ima šola svoje uteži podvojene in enkrat odstranjene, izračuni torej se izvajajo dvakrat za vsako cono. Za več podrobnosti si oglejte tehnično dokumentacijo in uporabniške priročnike TIMSS 2015 in PIRLS 2016. Če je potrebna replikacija tabel in številk, je treba argument `shortcut` spremeniti v `TRUE`.
14. Funkcija `lsa.lin.reg` prav tako zagotavlja model Waldove F-statistike, saj je to ustrezna statistika pri kompleksnih zasnovah vzorčenja. Glejte Bate (2004) ter Rao in Scott (1984). Waldova F-statistika se izračuna z uporabo hi-kvadrat-porazdelitve in testira le ničelno hipotezo. Funkcija zagotavlja dvostranski t-test in p-vrednosti za koeficiente regresije.

Output, ki ga funkcija ustvari, je shranjen v MS Excel-delovnem zvezku. Delovni zvezek ima tri zavihke. Prvi (»Estimates«) bo imel naslednje stolpce, odvisno od vrste spremenljivk, vključenih v analizo:

- **<Country ID>** – stolpec, ki vsebuje imena držav v datoteki, za katere so bili izračunani statistični podatki. Natančen naslov stolpca bo odvisen od identifikatorja države, uporabljenega v raziskavi.
- **<Split variable 1>**, **<Split variable 2>**… – stolpci, ki vsebujejo kategorije, po katerih so bili statistični podatki razdeljeni. Natančna imena bodo odvisna od spremenljivk v `split.vars`.
- **n_Cases** – število primerov v vzorcu, uporabljenem za izračun statističnih podatkov.
- **Sum_<Weight variable>** – ocenjena številka elementov v populaciji na skupino po uporabi uteži. Dejanski naziv uteži bo odvisen od uteži, uporabljene v analizi.
- **Sum_<Weight variable>SE** – standardna napaka ocenjene številke elementov v populaciji na skupino. Dejanski naziv uteži bo odvisen od uteži, uporabljene v analizi.
- ***Percentages*<Last split variable>** – odstotki anketirancev (ocenjene populacije) na skupine, ki jih določajo spremenljivke za razdelitev v split.vars. Odstotki bodo prikazani za zadnjo spremenljivko za razdelitev, ki določa končne skupine.
- **Percentages_<Last split variable>SE** – standardne napake odstotkov zgoraj.
- **Variable** – imena spremenljivk (kontekstualne ali PV ali imena spremenljivk s kontrastnim kodiranjem). Opomba: ko so vključeni interakcijski izrazi, bodo celice z interakcijami v stolpcu Variables vsebovale imena dveh spremenljivk v vsakem izmed interakcijskih izrazov, ločena s količnikom, npr. ASBGSSB.

- **Coefficients** – regresijski koeficienti (presečišča in nakloni).
- **Coefficients_SE** – standardna napaka regresijskih koeficientov (preseki in nakloni) za vsako neodvisno spremenljivko (kontekstualne ali PV ali imena spremenljivk s kontrastnim kodiranjem) v modelu.
- **Coefficients_SVR** – komponenta variacije vzorčenja za regresijske koeficiente, če so PV določene bodisi kot odvisne bodisi kot neodvisne spremenljivke.
- **Coefficients<root PV>_MVR** – komponenta merilne variacije za regresijske koeficiente, če so PV določene bodisi kot odvisne bodisi kot neodvisne spremenljivke.
- **t_value** – t-testna vrednost za regresijske koeficiente.
- **p_value** – p-vrednost za regresijske koeficiente.

Drugi zavihek (»Model statistics«) vsebuje statistiko, povezano z modelom linearne regresije v naslednjih stolpcih:

- **<Country ID>** – stolpec, ki vsebuje imena držav v datoteki, za katere so bili izračunani statistični podatki. Natančen naslov stolpca bo odvisen od identifikatorja države, uporabljenega v raziskavi.
- **<Split variable 1>**, **<Split variable 2>**… – stolpci, ki vsebujejo kategorije, po katerih so bili statistični podatki razdeljeni. Natančna imena bodo odvisna od spremenljivk v `split.vars`.
- **Statistic** – stolpec, ki vsebuje R-kvadrat, prilagojeni R-kvadrat, Waldovo F-statistiko in ocene prostih članov.
- **Estimate** – numerične ocene za vsako od zgoraj navedenih.
- **Estimate_SE** – standardne napake ocen zgoraj.
- **Estimate_SVR** – komponenta variacije vzorčenja, če so PV vključeni v model.
- **Estimate_MVR** – komponenta merilne variacije, če so PV vključeni v model.
- **t_value** – t-testna vrednost za regresijske koeficiente, vrednost za Wald F-statistiko je zagotovljena.
- **p_value** – p-vrednost za regresijske koeficiente, vrednost za Wald F-statistiko je zagotovljena.

Tretji zavihek (»Analysis information«) vsebuje dodatne informacije, povezane z analizo po državi v naslednjih stolpcih:

- **DATA** – uporabljena `data.file` ali `data.object`.
- **STUDY** – iz katere raziskave so podatki.
- **CYCLE** – iz katerega cikla raziskave so podatki.
- **WEIGHT** – katera utež je bila uporabljena.
- **DESIGN** – katera tehnika ponovnega vzorčenja je bila uporabljena (JRR ali BRR).
- **SHORTCUT** – ali je bila uporabljena metoda skrajšave.
- **NREPS** – koliko ponovnih uteži je bilo uporabljenih.
- **ANALYSIS_DATE** – na kateri datum je bila izvedena analiza.
- **START_TIME** – ob katerem času se je analiza začela.
- **END_TIME** – ob katerem času se je analiza končala.
- **DURATION** – koliko časa je analiza trajala v urah, minutah, sekundah in milisekundah.

Četrti zavihek (Calling syntax) vsebuje klic funkcije z vrednostmi za vse parametre, kot je bila izvedena. To je uporabno, če je treba analizo pozneje ponoviti.

4.6.3 Izračun regresijskih koeficientov z uporabo ukazne vrstice

V naslednjih primerih bomo združili novo podatkovno datoteko s podatki o učencih in ravnateljih iz PIRLS 2016 (Avstralija in Slovenija), pri čemer bomo vzeli vse spremenljivke iz obeh vrst datotek:

```
lsa.merge.data(inp.folder = "C:/temp",
               file.types = list(acg = NULL, asg = NULL),
               ISO = c("aus", "svn"),
               out.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData")
```

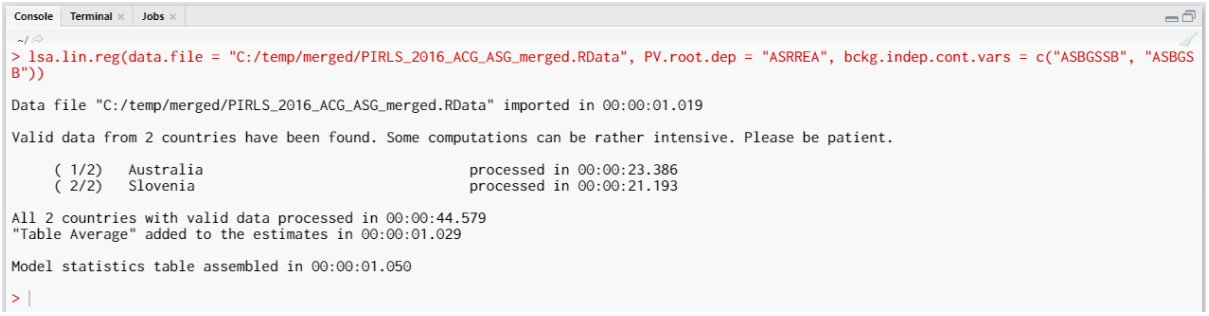
Za začetek bomo izračunali regresijske koeficiente za model, kjer je odvisna spremenljivka niz PV za bralne dosežke, neodvisni spremenljivki pa sta dve kontekstualni lestvici – občutek pripadnosti šoli (ASBGSSB) in medvrstniško nasilje (ASBGSB) v Avstraliji in Sloveniji (za informacijo o tem, kako so te lestvice sestavljene in kakšne so njihove lastnosti, preverite tehnično dokumentacijo PIRLS 2016):

```
lsa.lin.reg(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
            PV.root.dep = "ASRREA",
            bckg.indep.cont.vars = c("ASBGSSB", "ASBGSB"))
```

Nekaj stvari, ki jih je treba upoštevati:

1. Funkcija lahko vzame eno kontekstualno spremenljivko ali niz PV kot odvisno spremenljivko. Neodvisne spremenljivke so lahko več kontekstualnih spremenljivk in/ali nizi PV.
2. Pet PV za bralne dosežke: ASRREA01, ASRREA02, ASRREA03, ASRREA04 in ASRREA05. V argumentu `PV.root.corr` moramo navesti le koren PV, »ASRREA«. Funkcija bo uporabila ta koren, da izbere vseh pet PV in jih vključi v izračune.
3. V mednarodnih raziskavah morajo biti vse analize izvedene ločeno po državah. Vendar ni treba dodati spremenljivke ID države (IDCNTRY ali CNT v PISA) kot spremenljivke za razdelitev. Funkcija jo bo samodejno prepoznala in dodala v vektor `split.vars`.
4. Ni treba izrecno navesti utežne spremenljivke. Če utežna spremenljivka ni izrecno navedena, se bo kot privzeta utež (v tem primeru skupna utež učencev) uporabila za podatkovni niz, samodejno se prepozna glede na združene podatke o respondentih. Če imate dober razlog za spremembo utežne spremenljivke, to lahko storite z dodajanjem `weight.var = "SENWGT"`.
5. Če ni navedena datoteka za output, bo output shranjen z imenom »Analysis.xlsx« v delovnem imeniku (prikličete ga z `getwd()`).
6. Če v klicno sintakso izrecno ne dodate `open.output = FALSE`, bo datoteka z outputom odprta po zaključku vseh izračunov. To je koristno, kadar se izvaja več klicnih sintaks za različne analize in ni treba takoj pregledovati outputoc.

Izvajanje zgoraj navedene kode bo v konzoli RStudia izpisalo naslednji output:



Ko so vse operacije končane, bo izhod shranjen na disk v obliki MS Excel delovnega zvezka. Če je `open.output = TRUE` (privzeto), bo datoteka odprta v privzetem programu za preglednice (ponavadi MS Excel).

Kategorijske spremenljivke lahko dodamo kot kontrastno kodirane spremenljivke in preverimo pomen razlik med kategorijami v odvisni spremenljivki (kontekstne spremenljivke ali PV). Zaenkrat funkcija deluje z naslednjimi kontrastnimi shemami: dummy, deviation in simple. Preverimo razlike v skupnem dosežku branja učencev za tiste, ki imajo različno število knjig doma (ASBG04). Upoštevamo tudi, koliko učenci radi berejo (ASBGSLR; za informacijo o tem, kako je ta lestvica sestavljena in kakšne so njene lastnosti, preverite tehnično dokumentacijo PIRLS 2016). Število knjig doma (ASBG04) ima naslednje veljavne vrednosti:

- Nobene ali zelo malo (0–10 knjig);
- Dovolj za eno knjižno polico (11–25 knjig);
- Dovolj za eno knjižno omaro (26–100 knjig);
- Dovolj za dve knjižni omari (101–200 knjig);
- Dovolj za tri ali več knjižnih omar (več kot 200).

Dodati moramo ASBG04 kot vrednost argumenta `bckg.indep.cat.vars` funkcije `lsa.lin.reg`. Funkcija bo samodejno določila veljavne vrednosti, vendar moramo določiti referenčno kategorijo kot vrednost za `bckg.cat.contrasts` (vrsta kontrastnega kodiranja) in `bckg.ref.cats` (referenčna kategorija). Če ne določimo vrednosti za `bckg.cat.contrasts` (kar bomo storili), bo funkcija samodejno izračunala regresijske koeficiente z dummy kodiranjem (presečišče bo povprečje dosežka branja za učence, ki so izbrali kategorijo, ki jo nastavimo kot referenčno – glejte nadaljevanje) in regresijske koeficiente za preostale kategorije, ki niso referenčne. Če je potrebna katerakoli druga kontrastna shema, jo je treba izrecno določiti z uporabo argumenta `bckg.cat.contrasts`. Določili bomo prvo kategorijo (Nobena ali zelo malo (0–10 knjig) doma). Dodali bomo lestvico, koliko učenci radi berejo (ASBGSLR; za informacije o tem, kako je ta lestvica sestavljena in kakšne so njene lastnosti, preverite tehnično dokumentacijo PIRLS 2016) kot kontrolno spremenljivko `bckg.indep.cont.vars`. Klicna sintaksa izgleda takole:

```
lsa.lin.reg(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
            PV.root.dep = "ASRREA", bckg.indep.cat.vars = "ASBG04",
            bckg.ref.cats = 1, bckg.indep.cont.vars = "ASBGSLR")
```

Izvajanje zgornje sintakse bo prepisalo prejšnji izhod, ker ima enako ime datoteke (v konzoli bo prikazana opozorilo). Stolpci v listu »Estimates« bodo zdaj različni.

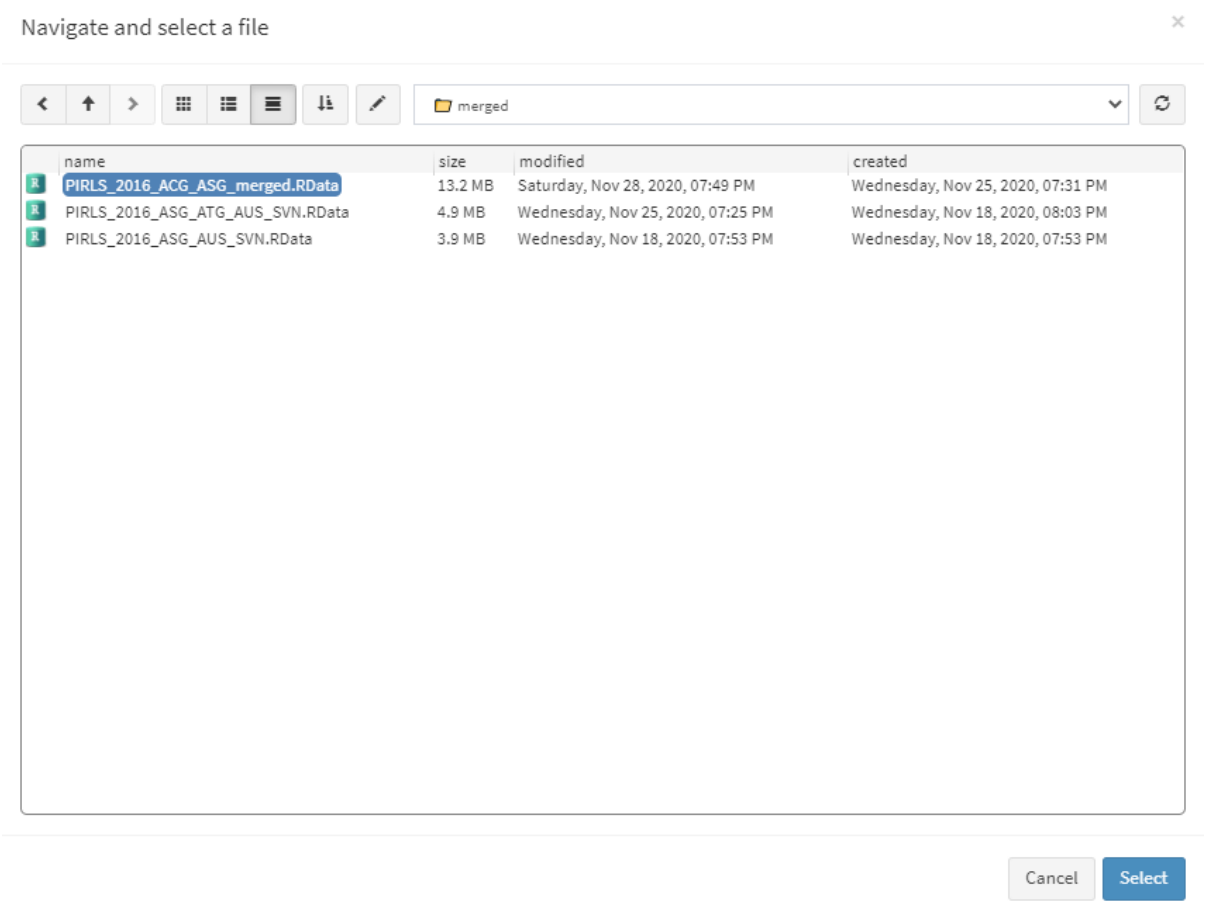
4.6.4 Izračun regresijskih koeficientov z uporabo GUI

Za začetek uporabniškega vmesnika RALSA izvedite naslednji ukaz v RStudiu:

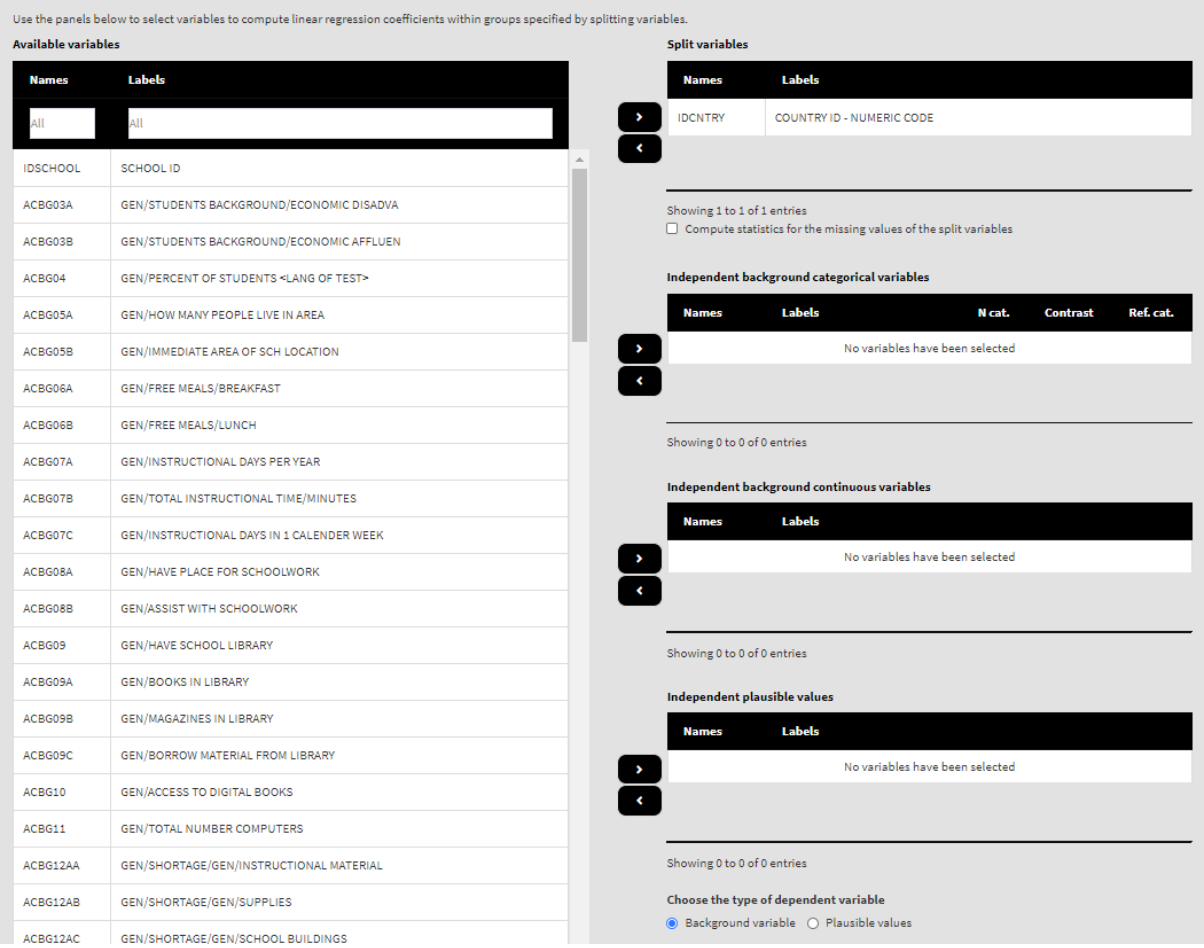
```
ralsaGUI()
```

Za naslednje primere združite novo datoteko s podatki PIRLS 2016 za Avstralijo in Slovenijo, pri čemer vključite vse spremenljivke za učence in ravnatelje. Ime združene datoteke lahko določite kot PIRLS_2016_ACG_ASG_merged.RData.

Ko končate z združevanjem podatkov, izberite **Analysis types > Linear regression** iz menija na levi strani. Ko ste v razdelku Linear regression v GUI-ju, kliknite na gumb **Choose data file**. Pojdite do mape, ki vsebuje datoteko **»PIRLS_2016_ACG_ASG_merged.RData«**, izberite jo in kliknite gumb **Select**.



Ko je datoteka naložena, boste videli ploščo na levi (razpoložljive spremenljivke) in niz plošč na desni, kjer lahko spremenljivke s seznama razpoložljivih dodate. Nad ploščami boste videli tudi informacije o naloženi datoteki.

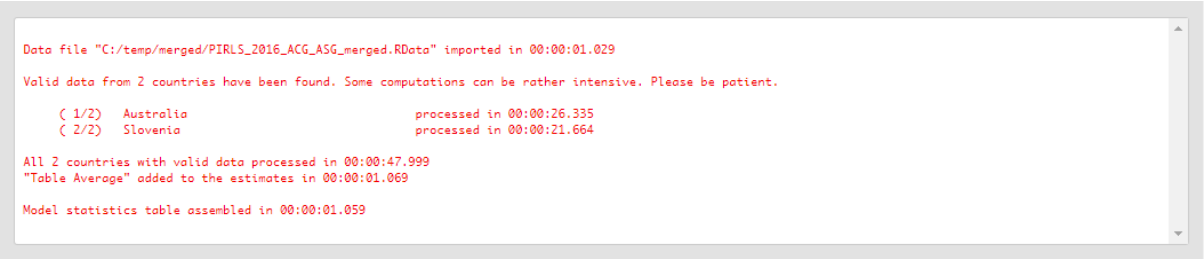


Uporabite miško za izbiro spremenljivk s seznama Razpoložljive spremenljivke (**Available variables**) in puščice na sredini zaslona, da jih dodate v različna polja (ali odstranite) za nastavitev analize. Za hitro iskanje potrebnih spremenljivk lahko uporabite filtre na vrhu plošč.

Za začetek izračunajmo regresijske koeficiente za model, kjer je odvisna spremenljivka niz PV za bralni dosežek, neodvisni spremenljivki pa sta dve kontekstualni lestvici – »Občutek pripadnosti šoli« (ASBGSSB) in »Medvrstniško nasilje« (ASBGSB) v Avstraliji in Sloveniji (za informacije o tem, kako je ta lestvica sestavljena in kakšne so njene lastnosti, preverite tehnično dokumentacijo PIRLS 2016). S seznama Razpoložljive spremenljivke (**Available variables**) Izberite spremenljivki ASBGSSB in ASBGSB in ju premaknite na seznam Neodvisne zvezne spremenljivke (**Independent continuous variables**) s pomočjo desne puščice. Pomaknite se navzdol in iz para gumbov pod Izberite vrsto odvisne spremenljivke izberite Verjetne vrednosti (**Plausible values**). Na seznamu Razpoložljive spremenljivke (**Available variables**) poiščite koren verjetnostnih vrednosti za rezultat na bralnem preizkusu (ASRREA). Uporabite filtre na vrhu plošče za iskanje po imenu ali oznaki. Izberite koren in ga dodajte na panele odvisnih verjetnostnih vrednosti s pomočjo desne puščice. To je vse, kar je treba storiti. Pomaknite se navzdol in kliknite na Določi ime output datoteke (**Define the output file name**). Pojdite v mapo **»C:/temp/Results«** (ali v mapo, kjer želite shraniti output) in določite ime output datoteke. Ko to storite, se bo poleg Določite ime output datoteke (**Define the output file name**) pojavilo potrditveno polje. Če je označeno, se bo output odprl po zaključku vseh izračunov. Pod tem boste videli potrditveno polje Standardizirani koeficienti (**Standardized coefficients**). Če je označeno, bodo spremenljivke standardizirane pred

izračunom statistik. Pod tem bo prikazan zapis sintakse. Pod vsemi temi bo prikazan gumb za Izvedbo sintakse. Končne nastavitve v spodnjem delu zaslona bi morale izgledati takole:

Kliknite gumb Izvedi sintakso. Konzola GUI se bo pojavila na dnu in beležila vse dokončane operacije:



Nekaj opomb:

1. Funkcija lahko vzame eno kontekstualno spremenljivko ali niz PV kot odvisno spremenljivko. Neodvisnih spremenljivk pa je lahko več: več ozadinskih/kontekstualnih spremenljivk in/ali nizi PV.
2. Pet PV za bralne dosežke: ASRREA01, ASRREA02, ASRREA03, ASRREA04 in ASRREA05. Seznami razpoložljivih spremenljivk in odvisnih verjetnostnih vrednosti ne bodo prikazali petih ločenih PV, temveč le njihov osnovni/skupni naziv – ASRREA, brez številk na koncu. Funkcija bo v ozadju vzela vseh pet PV in jih vključila v izračune.
3. V mednarodnih raziskavah je treba vse analize izvajati ločeno po državah. Spremenljivka ID države (IDCNTRY ali CNT v PISA) je vedno izbrana kot prva razdelitvena spremenljivka in je ne morete odstraniti s panela Razdelitvene spremenljivke (**Split variables**).
4. Privzeta utežna spremenljivka je samodejno izbrana in dodana v panel utežnih spremenljivk. Lahko jo spremenite z drugo utežjo, ki je na voljo v podatkovnem nizu. Če je izbrana privzeta utežna spremenljivka, ne bo prikazana v oknu sintakse. Če v panelu Utežne spremenljivke ni izbrana nobena spremenljivka, bo privzeta utež samodejno uporabljena.
5. Če je označen okvirček Standardizirani koeficienti (**Standardized coefficients**), bodo regresijski koeficienti izračunani na standardiziranih spremenljivkah in bodo v outputu vključeni beta-koeficienti.

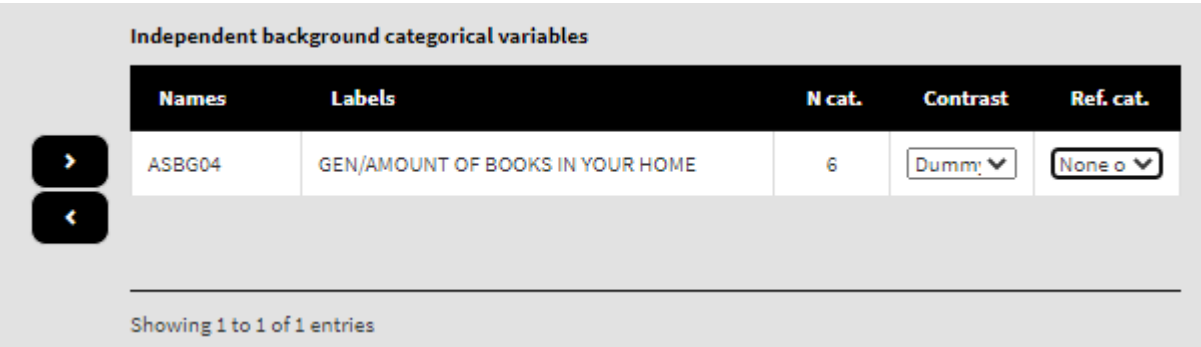
6. Okvirček Uporabi metodo bližnjice za izračun SE ni privzeto odkljukan. To bo funkciji omogočilo, da izračuna standardno napako z uporabo »polne« metode za komponento vzorčne variance.

Če je označen okvirček Odpri output, ko končano (**Open the output when done**), se bo output samodejno odprl v privzetem programu za preglednice (ponavadi MS Excel), ko bodo vsi izračuni zaključeni.

Kategorijske spremenljivke je mogoče dodati kot spremenljivke z nasprotnimi kodami in lahko preizkusimo pomembnost razlik med kategorijami v odvisni spremenljivki. Za zdaj funkcija lahko deluje z naslednjimi kontrastnimi shemami: dummy, deviation in simple. Preizkusimo razlike v splošnih bralnih dosežkih učencev glede na število knjig doma (ASBG04) ob kontrolni spremenljivki, kako učenci uživajo v branju (ASBGSLR; za informacije o tem, kako je ta lestvica sestavljena in kakšne so njene lastnosti, preverite tehnično dokumentacijo PIRLS 2016)). Spremenljivka števila knjig doma (ASBG04) ima naslednje veljavne vrednosti:

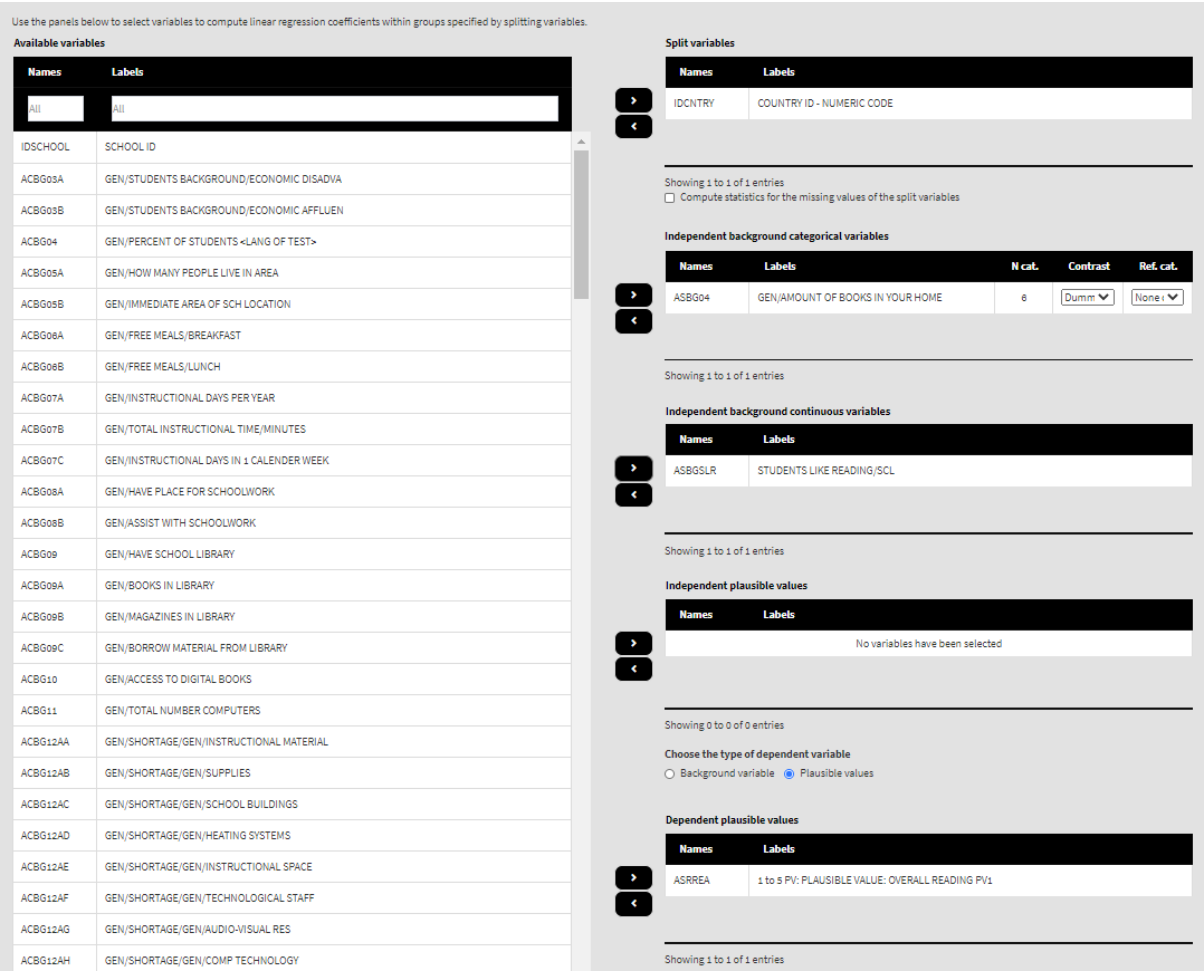
- Nobene ali zelo malo (0–10 knjig)
- Dovolj za eno polico (11–25 knjig)
- Dovolj za en regal (26–100 knjig)
- Dovolj za dva regala (101–200 knjig)
- Dovolj za tri ali več regalov (več kot 200)

Odstranite spremenljivke s seznama **Independent continous background variables** in dodajte spremenljivko ASBG04 (število knjig doma) na seznam kategoričnih spremenljivk (**Independent background categorical variables**). Videli boste, da bo seznam samodejno prikazal število kategorij za spremenljivko, spustni seznam z različnimi shemami kodiranja (dummy, deviation in simple, glejte stolpec N cat.) in spustni seznam s kategorijami spremenljivke za izbiro:

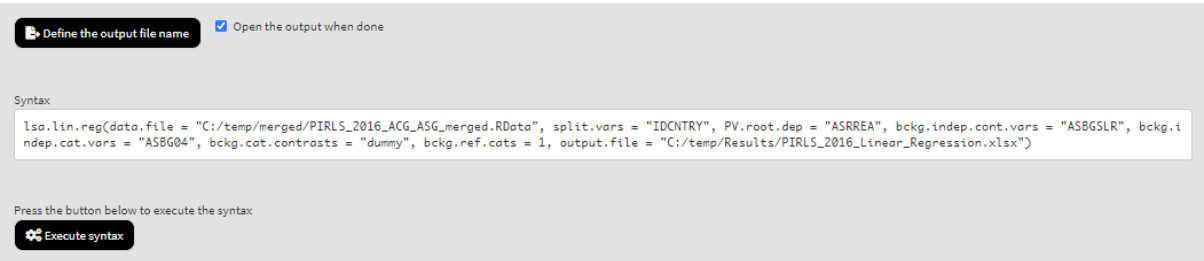


Privzeto sta izbrana **dummy contrast coding scheme** in prva razpoložljiva kategorija kot referenca. Pustimo privzete nastavitve. Funkcija bo samodejno izračunala regresijske koeficiente z **dummy coding** (presečišče bo povprečen dosežek pri branju za učence, ki so izbrali kategorijo, ki jo nastavimo kot referenco – glejte v nadaljevanju) in regresijske koeficiente za preostale kategorije bodo razlike v dosežku za učence, ki so izbrali katero koli drugo kategorijo razen referenčne. Če je potrebna drugačna shema kontrasta, jo lahko spremenite tako, da kliknete na spustni seznam in izberete bodisi **deviation** bodisi **simple**. Kot referenco bomo pustili bomo prvo kategorijo (Nobene ali zelo malo (0–10 knjig)). Na seznam **Independent background continuous** bomo dodali **spremenljivko koliko učenci**

radi berejo (ASBGSLR; za informacije o tem, kako je ta lestvica sestavljena in kakšne so njene lastnosti, preverite tehnično dokumentacijo PIRLS 2016) kot kontrolno spremenljivko. Na seznamu **Dependent plausible values** bomo pustili niz petih **PV** o dosežku pri branju .



Ker aplikacijo za **Linearno regresijo** uporabljamo takoj po izvedbi prejšnje analize, bomo še vedno imeli preostale nastavitve iz prejšnje analize. Ni treba spreminjati nobenih preostalih nastavitev, razen če želite. Lahko pa spremenite ime output datoteke, sicer bo prepisano. Opazili boste, da se bo prikazana sintaksa spremenila, kar odraža odstranitev spremenljivk iz prejšnje analize ter vključitev ASBG04 kot **Independent categorical variable** in ASBGSLR kot **Independent background continuous variable**.



Pritisnite gumb **Execute syntax**. Konzola GUI se bo posodobila in zabeležila vse izvedene operacije:

```
Data file "C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData" imported in 00:00:01.029

Valid data from 2 countries have been found. Some computations can be rather intensive. Please be patient.

     ( 1/2)   Australia                          processed in 00:00:18.515
     ( 2/2)   Slovenia                           processed in 00:00:17.394

All 2 countries with valid data processed in 00:00:35.908
"Table Average" added to the estimates in 00:00:01.059

Model statistics table assembled in 00:00:01.050
```

Če je potrjeno polje **Open the output when done**, se bo rezultat samodejno odprl v privzetem programu za preglednice (običajno MS Excel) po končanju vseh izračunov.

4.7 Binarna logistična regresija

4.7.1 Uvod

Funkcija `lsa.bin.log.reg` izračuna koeficiente logistične regresije v skupinah anketirancev, ki so določene z razdelitvenimi spremenljivkami (angl. *split variables*), kjer je odvisna spremenljivka binarna (tj. dihotomna, z dvema različnima vrednostma). Razdelitvene spremenljivke so opcijske. Če razdelitvene spremenljivke niso navedene, se rezultati izračunajo samo na ravni države. Če so te navedene, se podatki v vsaki državi razdelijo v skupine glede na vse razdelitvene spremenljivke. Koeficienti logistične regresije pa se izračunajo za zadnjo razdelitveno spremenljivko. Neodvisne spremenljivke so lahko tako kontekstualne spremenljivke (angl. *background contextual variables*) kot tudi nizi PV (angl. *plausible values,* verjetne vrednosti). Vsi izračuni bodo upoštevali kompleksen vzorec in zasnovo ocenjevanja v raziskavi. Ko se kot neodvisne spremenljivke uporabijo nizi PV, se koeficienti logistične regresije izračunajo med odvisno spremenljivko in vsako PV v nizu, nato pa se ocene za vse PV v nizu povprečijo, standardna napaka pa se izračuna s kompleksnimi formulami, ki so odvisne od raziskave.

Če vas zanimajo podrobnejše informacije o kompleksnih vzorcih in zasnovah ocenjevanja določene raziskave ter kako se izračunajo ocene in njihove standardne napake, si oglejte tehnično dokumentacijo in uporabniški priročnik dotične raziskave.

Kot katerakoli druga funkcija v paketu RALSA funkcija `lsa.bin.log.reg` samodejno prepozna podatke raziskave in uporabi pravilne tehnike ocenjevanja glede na vzorec ter izvedbo zasnove ocenjevanja raziskave.

4.7.2 Funkcija binarne logistične regresije in njeni argumenti

Funkcija `lsa.bin.log.reg` ima naslednje argumente:

- `data.file` – datoteka, ki vsebuje `lsa.data` objekt. Navedemo lahko bodisi to bodisi `data.object`, vendar ne obeh.
- `data.object` – objekt v pomnilniku, ki vsebuje `lsa.data` objekt. Navedemo lahko bodisi tega bodisi `data.file`, vendar ne obeh.
- `split.vars` – kategorijske spremenljivke, po katerih se rezultati razdelijo. Če razdelitvene spremenljivke niso podane, bodo prikazani rezultati za celotno populacijo držav. Če je podana ena ali več spremenljivk, se rezultati razdelijo po vseh razen po zadnji spremenljivki, odstotki anketirancev pa se izračunajo po edinstvenih vrednostih zadnje razdelitvene spremenljivke.
- `bin.dep.var` – ime binarne (tj. le dve različni vrednosti) ozadnje ali kontekstualne spremenljivke, ki se uporablja kot odvisna spremenljivka v modelu.
- `bckg.indep.cont.vars` – imena zveznih neodvisnih ozadnjih ali kontekstualnih spremenljivk, ki se uporabljajo kot napovedne spremenljivke v modelu.
- `bckg.indep.cat.vars` – imena kategorijskih neodvisnih ozadnjih ali kontekstualnih spremenljivk, ki se uporabljajo kot napovedne spremenljivke v modelu za izračun kontrastov (glej `bckg.cat.contrasts` in `bckg.ref.cats`).
- `bckg.cat.contrasts` – vektor celih števil z enako dolžino kot dolžina `bckg.indep.cat.vars`, ki določa vrste kontrastov za izračun, če so podane `bckg.indep.cat.vars`.
- `bckg.ref.cats` – niz znakov z enako dolžino kot dolžina `bckg.indep.cat.vars` in `bckg.cat.contrasts`, ki določa referenčne kategorije za izračun kontrastov, če so podane `bckg.indep.cat.vars`.
- `PV.root.indep` – koren imena za nabor PV, ki se uporabljajo kot neodvisne spremenljivke v modelu.
- `interactions` – interakcijski členi – seznam, ki vsebuje vektorje dolžine dva.
- `standardize` – določitev, ali naj bodo odvisne in neodvisne spremenljivke standardizirane za proizvodnjo beta-koeficientov. Privzeta vrednost je `FALSE`.
- `weight.var` – ime spremenljivke, ki vsebuje uteži. Če ime utežne spremenljivke ni navedeno, bo funkcija samodejno izbrala privzeto utežno spremenljivko za podane podatke, odvisno od vrste anketiranca.
- `norm.weight` – določitev, ali naj bodo uteži normalizirane, preden jih uporabimo. Privzeto je `FALSE`.
- `include.missing` – določitev, ali naj se manjkajoče vrednosti razdelitvenih spremenljivk vključijo kot kategorije, po katerih se razdelijo in za katere se izračunajo vse statistike, Privzeto (`FALSE`) se upoštevajo vsi primeri razdelitvenih spremenljivk brez manjkajočih vrednosti, preden se izračuna statistika.
- `shortcut` – določitev, ali naj se za IEA TIMSS, TIMSS Advanced, TIMSS Numeracy, eTIMSS, PIRLS, ePIRLS, PIRLS Literacy in RLII uporabi »bližnjica«. Privzeto (`FALSE`) se pri izračunu komponent variance in standardnih napak ocen uporabi »polna« metoda.
- `save.output` – določitev, ali naj se rezultat shrani v MS Excel-datoteko (privzeto) ali ne (v tem primeru se natisne na konzolo ali dodeli objektu).

- `output.file` – polna pot do izhodne datoteke, vključno z imenom datoteke. Če je izpuščeno, se v delovni imenik (`getwd()`) zapiše datoteka s privzetim imenom »Analysis.xlsx«.
- `open.output` – določitev, ali naj se rezultat odpre po tem, ko je zapisan. Privzeta vrednost (`TRUE`) odpre rezultat v privzetem programu za preglednice, nameščenem na računalniku.

Opombe:

1. Bodisi `data.file` bodisi `data.object` mora biti zagotovljen kot vir podatkov. Če sta zagotovljena oba, se bo funkcija ustavila z napako. Funkcija izračuna koeficiente binarne logistične regresije po kategorijah razdelitvenih spremenljivk. Odstotki anketirancev v vsaki skupini so izračunani znotraj skupin, določenih z zadnjo razdelitveno spremenljivko. Če razdelitvene spremenljivke niso dodane, bodo rezultati izračunani samo na ravni države.
2. Če je `standardize = TRUE`, bodo spremenljivke standardizirane pred izračunom statistike, da se zagotovijo beta-regresijski koeficienti.
3. Binarna (tj. dihotomna) ozadna/kontekstualna spremenljivka mora biti podana v bin.dep.var (številčna ali faktor). Če ima spremenljivka več kot dve kategoriji, se bo funkcija ustavila z napako. Funkcija samodejno prekodira dve kategoriji bin.dep.var v `0` in `1`, če ti dve nista že taki (npr. kot 1 in 2 v faktorjih). Če ima izbrana spremenljivka več kot dve različni vrednosti (uporabite `lsa.var.dict` za pregled le-teh), jih je mogoče združiti z uporabo `lsa.recode.vars`.
4. Ozadne/kontekstualne spremenljivke, poslane v `bckg.indep.cont.vars`, bodo v modelu obravnavane kot številske spremenljivke. Spremenljivke z diskretnim številom kategorij (tj. faktorji), poslane v `bckg.indep.cat.vars`, bodo uporabljene za izračun kontrastov. V tem primeru je treba vrsto kontrasta podati v `bckg.cat.contrasts`, skupaj s številom referenčnih kategorij za vsako izmed `bckg.indep.cat.vars`. Število vrst kontrastov in referenčnih kategorij mora biti enako številu `bckg.indep.cat.vars`. Trenutno podprte sheme kontrastnega kodiranja so:
    - `dummy` (imenovano tudi »indikator« v logistični regresiji) – razmerja verjetnosti prikazujejo, kakšna je verjetnost za pozitiven (tj. `1`) izid v binarni odvisni spremenljivki v primerjavi z negativnim izidom (tj. `0`) za kategorijo spremenljivke v `bckg.indep.cat.cats` v primerjavi z referenčno kategorijo te dummy kodirane spremenljivke. Presečišče (angl. *intercept*) prikazuje logaritem verjetnosti za referenčno kategorijo, ko so vse druge ravni `0`.
    - `deviation` (v logistični regresiji imenovano tudi »učinek«) – primerjava učinka vsake kategorije spremenljivke (razen referenčne) s kodiranjem `deviation` z vsemi učinki skupaj (kar je intercept).
    - `simple` – enako kot pri `dummy` kontrastnem kodiranju, razen da je intercept v tem primeru skupni učinek.
5. Upoštevajte, da pri uporabi `standardize = TRUE` kontrastno kodiranje `bckg.indep.cat.vars` ni standardizirano. Zato regresijski koeficienti morda ne bodo primerljivi z drugimi programski rešitvami za analizo podatkov iz mednarodnih raziskav, kot sta npr. SPSS ali SAS, kjer se kontrastno kodiranje kategorijskih

spremenljivk (npr. `dummy` kodiranje) izvaja privzeto. Vendar pa bodo statistike modela enake.

6. Za izračun regresijskih koeficientov je mogoče vključiti več zveznih ali kategorijskih spremenljivk ter/ali naborov PV. Upoštevajte, da se bodo v tem primeru rezultati rahlo razlikovali od tistih, ki bi jih dobili z uporabo vsakega para istih zveznih spremenljivk ali PV v ločenih analizah. To je zaradi tega, ker se primeri z manjkajočimi vrednostmi odstranijo vnaprej, in več kot je spremenljivk, več primerov bo verjetno odstranjenih. Funkcija podpira le **listwise** deletion**.**
7. Izračun regresijskih koeficientov, ki vključujejo verjetne vrednosti, zahteva določitev korena imen PV v `PV.root.dep` in/ali `PV.root.indep`. Vse raziskave (razen CivED, TEDS-M, SITES, TALIS in TALIS Starting Strong Survey) imajo niz možnih vrednosti za vsak konstrukt (npr., v TIMSS pet za splošno matematiko, pet za algebro, pet za geometrijo itd.). V nekaterih raziskavah (npr. TIMSS in PIRLS) se imena možnih vrednosti v nizu vedno začnejo z naborom znakov in končajo z zaporednim številom PV. Npr., imena niza možnih vrednosti za splošno matematiko v TIMSS so BSMMAT01, BSMMAT02, BSMMAT03, BSMMAT04 in BSMMAT05. Koren možnih vrednosti za ta niz, ki se doda v `PV.root.dep` ali `PV.root.indep`, bo BSMMAT. Funkcija bo samodejno našla vse spremenljivke v tem nizu možnih vrednosti in jih vključila v analizo. V drugih raziskavah, kot so OECD PISA in IEA ICCS ter ICILS, je zaporedna številka vsake PV vključena v sredini imena. Npr., v ICCS so imena niza možnih vrednosti PV1CIV, PV2CIV, PV3CIV, PV4CIV in PV5CIV. Koren imena PV mora biti naveden v `PV.root.dep` ali `PV.root.indep` kot PV#CIV. Več nizov možnih vrednosti je mogoče dodati v `PV.root.indep`.
8. Funkcija lahko prav tako izračuna dvoučinkovne interakcije med neodvisnimi spremenljivkami z uporabo seznama v argumentu `interactions`. Seznam mora vsebovati vektorje dolžine dva, vse spremenljivke v teh vektorjih pa morajo biti tudi navedene kot neodvisne spremenljivke. Bodite pozorni na naslednje:

- Ko je interakcija med dvema neodvisnima zveznima spremenljivkama (obe sta torej navedeni v `bckg.indep.cont.vars`), bo učinek interakcije med njima izračunan v taki obliki, kot sta.
- Ko je interakcija med dvema kategorialnima spremenljivkama (obe sta torej navedeni v `bckg.indep.cat.vars`), bo učinek interakcije izračunan med vsakim možnim parom kategorij obeh spremenljivk razen referenčnih kategorij.
- Ko je interakcija med eno zvezno (tj. navedeno v `bckg.indep.cont.vars`) in eno kategorialno spremenljivko (tj. navedeno v `bckg.indep.cat.vars`), bo učinek interakcije izračunan med zvezno spremenljivko in vsako kategorijo kategorialne spremenljivke razen referenčne kategorije.
- Ko je interakcija med zvezno spremenljivko (tj. navedeno v `bckg.indep.cont.vars`) in nizom PV (tj. navedenim v `PV.root.indep`), bo učinek interakcije izračunan med zvezno spremenljivko in vsako PV v nizu, rezultati pa bodo agregirani.
- Ko je interakcija med kategorialno spremenljivko (tj. navedeno v `bckg.indep.cat.vars`) in nizom PV (tj. navedenim v `PV.root.indep`), bo učinek interakcije izračunan med vsako kategorijo kategorialne spremenljivke (razen referenčne kategorije) in vsako PV v nizu. Rezultati bodo agregirani za vsako kategorijo kategorialnih spremenljivk in niz PV.

- Ko je interakcija med dvema nizoma PV (tj. navedenima v `PV.root.indep`), bo učinek interakcije izračunan med prvo PV v prvem nizu in prvo PV v drugem nizu, drugo PV v prvem nizu in drugo PV v drugem nizu itd. Rezultati bodo nato agregirani.

9. Če je `norm.weight = TRUE`, bodo uteži normalizirane pred uporabo v modelu. To je lahko potrebno v nekaterih državah, kjer lahko ekstremne uteži za nekatere primere povzročijo prenapihnjene ocene zaradi popolne ločitve modela. Posledica normalizacije uteži je, da se bo število elementov v populaciji seštelo na število primerov v vzorcu. Uporabljajte previdno.
10. Če je `include.missing = FALSE` (privzeto), bodo vsi primeri z manjkajočimi vrednostmi na spremenljivkah za delitev odstranjeni in v statistiko bodo vključeni samo primeri z veljavnimi vrednostmi. Upoštevajte, da se podatki iz raziskav lahko izvozijo na dva različna načina: (1) nastavljanje vseh uporabniško določenih manjkajočih vrednosti na `NA`; (2) uvoz vseh uporabniško določenih manjkajočih vrednosti kot veljavnih in dodajanje njihovih kod v dodatni atribut vsake spremenljivke. Če je `include.missing` nastavljeno na `FALSE` (privzeto) in se uporabljajo podatki, izvoženi z možnostjo (2), bo output odstranil iz spremenljivke odstranil vse vrednosti, ki se ujemajo z vrednostmi v njenem atributu manjkajočih vrednosti. V nasprotnem primeru jih bo vključil kot veljavne vrednosti in zanje izračunal statistiko.
11. Argument shortcut je veljaven le za TIMSS, TIMSS Advanced, TIMSS Numeracy, PIRLS, ePIRLS, PIRLS Literacy in RLII. Prej so te raziskave pri izračunu standardnih napak uporabljale 75 replikatov, ker je imela ena od šol v 75 JK območjih podvojene uteži, druga pa je bila odstranjena. Od TIMSS 2015 in PIRLS 2016 naprej raziskave uporabljajo 150 replikatov in v vsakem JK območju je imela ena šola podvojene teže, druga pa je bila odstranjena, torej se izračuni izvedejo dvakrat za vsako območje. Za več podrobnosti glejte tehnično dokumentacijo in uporabniške priročnike TIMSS 2015 in PIRLS 2016. Če je potrebna replikacija tabel in slik, je treba argument `shortcut` nastaviti na `TRUE`. Funkcija nudi dvostranske t-teste in p-vrednosti za regresijske koeficiente.
12. Razen če izrecno nastavite `save.output = FALSE`, bo output shranjen v MS Excel na disku. V nasprotnem primeru bo output natisnjen na konzolo.
13. Če ni določene output datoteke, bo output shranjen z imenom datoteke »Analysis.xlsx« v delovnem imeniku (ki ga lahko pridobite z `getwd()`).

Output, ki ga ustvari funkcija, je shranjen v MS Excel-delovni zvezek. Delovni zvezek ima tri liste. Prvi list (»Estimates«) bo imel naslednje stolpce, odvisno od tega, katere vrste spremenljivk so bile vključene v analizo:

- **<Country ID>** – stolpec, ki vsebuje imena držav v datoteki, za katere so izračunane statistike. Točen naslov stolpca bo odvisen od identifikatorja države, ki se uporablja v določeni raziskavi.
- **<Split variable 1>**, **<Split variable 2>**... – stolpci, ki vsebujejo kategorije, po katerih so statistike razdeljene. Točna imena bodo odvisna od spremenljivk v `split.vars`.
- **n_Cases** – število primerov v vzorcu, ki je bil uporabljen za izračun statistike.
- **Sum_<Weight variable>** – ocenjena populacija števila elementov na skupino po uporabi uteži. Dejansko ime utežne spremenljivke bo odvisno od utežne spremenljivke, uporabljene v analizi.

- **Sum_<Weight variable>SE** – standardna napaka ocenjene populacije števila elementov na skupino. Dejansko ime utežne spremenljivke bo odvisno od utežne spremenljivke, uporabljene v analizi.
- **Percentages<Last split variable>** – odstotki anketirancev (populacijske ocene) na skupino, opredeljeno s spremenljivkami v `split.vars`. Odstotki bodo za zadnjo razdelitveno spremenljivko, ki opredeljuje končne skupine.
- **Percentages_<Last split variable>SE** – standardne napake odstotkov zgoraj.
- **Variable** – imena spremenljivk (ozadje/kontekstne spremenljivke ali PV ali kontrastno kodirana imena spremenljivk). Opomba: če so vključeni interakcijski izrazi, bodo celice z interakcijami v stolpcu Spremenljivke vsebovale imena dveh spremenljivk v vsakem interakcijskem izrazu, ločena z dvopičjem, npr. ASBGSSB.
- **Coefficients** – logistični regresijski koeficienti (presečišče in nakloni).
- **Coefficients_SE** – standardna napaka logističnih regresijskih koeficientov (presečišče in nakloni) za vsako neodvisno spremenljivko (ozadje/kontekstne ali PV ali kontrastno kodirana imena spremenljivk) v modelu.
- **Coefficients_SVR** – komponenta vzorčne variance za logistične regresijske koeficiente, če so osnovne PV določene kot odvisne ali neodvisne spremenljivke.
- **Coefficients<root PV>_MVR** – komponenta merilne variance za logistične regresijske koeficiente, če so osnovne PV določene kot odvisne ali neodvisne spremenljivke.
- **Wald_Statistic** – Waldova (z) statistika.
- **p_value** – p-vrednost za regresijske koeficiente.
- **Odds_Ratio** – razmerja verjetnosti pri logistični regresiji.
- **Odds_Ratio_SE** – standardne napake za razmerja verjetnosti pri logistični regresiji.
- **Wald_L95CI** – spodnji 95-odstotni interval zaupanja za modelno osnovane koeficiente logistične regresije.
- **Wald_U95CI** – zgornji 95-odstotni interval zaupanja za modelno osnovane koeficiente logistične regresije.
- **Odds_L95CI** – spodnji 95-odstotni interval zaupanja za razmerja verjetnosti.
- **Odds_U95CI** – zgornji 95-odstotni interval zaupanja za razmerja verjetnosti.

Drugi list (»Model statistics«) vsebuje statistike, povezane z binarno logistično regresijo, v naslednjih stolpcih:

- **<Country ID>** – stolpec, ki vsebuje imena držav v datoteki, za katere so izračunane statistike. Točen naslov stolpca bo odvisen od identifikatorja države, uporabljenega v raziskavi.
- **<Split variable 1>**, **<Split variable 2>**... – stolpci, ki vsebujejo kategorije, po katerih so statistike razdeljene. Točna imena bodo odvisna od spremenljivk v `split.vars`.
- **Statistic** – stolpec, ki vsebuje Null Deviance (-2LL, brez napovednikov v modelu, samo konstanta, imenovana tudi »baseline«), Deviance (-2LL, po dodajanju napovednikov, preostala devianca, imenovana tudi »nova«), DF Null (stopnje svobode za ničelno devianco), DF Residual (stopnje svobode za preostalo devianco), Akaikejev informacijski kriterij (AIC), Bayesov informacijski kriterij (BIC), hi-kvadrat modela, različne R-kvadrat-statistike (Hosmer in Lemeshow – HS, Cox in Snell – CS, in Nagelkerke – N).
- **Estimate** – numerične ocene za zgoraj navedene statistike.
- **Estimate_SE** – standardne napake ocen zgoraj.

- **Estimate_SVR** – komponenta vzorčne variance, če so PV vključene v model.

- **Estimate_MVR** – komponenta merilne variance, če so PV vključene v model.

Tretji list (»Analysis information«) vsebuje dodatne informacije, povezane z analizo za vsako državo, v naslednjih stolpcih:

- **DATA** – uporabljena `data.file` ali `data.object`.
- **STUDY** – iz katere raziskave so podatki.
- **CYCLE** – iz katerega cikla raziskave so podatki.
- **WEIGHT** – katera utežna spremenljivka je bila uporabljena.
- **DESIGN** – katera metoda ponovnega vzorčenja je bila uporabljena (JRR ali BRR).
- **SHORTCUT** – ali je bila uporabljena bližnjica.
- **NREPS** – koliko ponovitvenih uteži je bilo uporabljenih.
- **ANALYSIS_DATE** – datum, ko je bila analiza izvedena.
- **START_TIME** – čas začetka analize.
- **END_TIME** – čas konca analize.
- **DURATION** – kako dolgo je trajala analiza v urah, minutah, sekundah in milisekundah.

Četrti list (»Calling syntax«) vsebuje klic funkcije z vrednostmi za vse parametre, kot je bila izvedena. To je uporabno, če je treba analizo kasneje ponoviti.

4.7.3 Izračun binarnih logističnih regresijskih koeficientov z uporabo ukazne vrstice

V naslednjih primerih bomo združili novo datoteko s podatki o učencih in ravnateljih iz PIRLS 2016 (Avstralija in Slovenija), pri čemer bomo vzeli vse spremenljivke iz obeh vrst datotek:

```
lsa.merge.data(inp.folder = "C:/temp",
               file.types = list(acg = NULL, asg = NULL),
               ISO = c("aus", "svn"),
               out.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData")
```

Za začetek izračunajmo binarne logistične regresijske koeficiente za model, ki napoveduje, ali se učenci v Avstraliji in Sloveniji strinjajo ali ne, da jih učitelji obravnavajo pošteno (spremenljivka ASBG12D), glede na njihov občutek pripadnosti šoli (ASBGSSB; za informacije o tem, kako je ta lestvica zgrajena in kakšne so njene lastnosti, preverite tehnično dokumentacijo PIRLS 2016). Funkcija `lsa.bin.log.reg` sprejme samo binarne (tj. dihotomne) spremenljivke kot odvisne, medtem ko so odgovori na vprašanje, koliko se učenci strinjajo ali ne, da jih učitelji obravnavajo pošteno (ASBG12D), razdeljeni v štiri različne kategorije:

- Zelo se strinjam;
- Strinjam se;
- Ne strinjam se;
- Sploh se ne strinjam.

Zato moramo najprej spremeniti ASBG12D tako, da združimo kategorije v dve, kjer sta »Sploh se ne strinjam« in »Ne strinjam se« združeni v prvo kategorijo, »Zelo se strinjam« in »Strinjam se« pa v drugo kategorijo:

```
lsa.recode.vars(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
                src.variables = "ASBG12D",
                old.new = "1=2;2=2;3=1;4=1;5=3",
                new.variables = "ASBG12Dr",
                new.labels = c("Ne strinjam se", "Strinjam se",
"Izpuščeno ali neveljavno"),
                missings.attr = "Izpuščeno ali neveljavno",
                variable.labels = "GEN/STRINJANJE/UČITELJI SO
POŠTENI - REKODIRANO",
                out.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData")
```

Upoštevajte, da ASBG12D prekodiramo v novo spremenljivko (ASBG12Dr). To je priporočljivo, ker bomo originalno spremenljivko ASBG12D ohranili nespremenjeno. Novoustvarjeni spremenljivki prav tako dodelimo oznako. Izračunajmo logistične regresijske koeficiente z uporabo nove spremenljivke kot odvisne in ASBGSSB kot neodvisne:

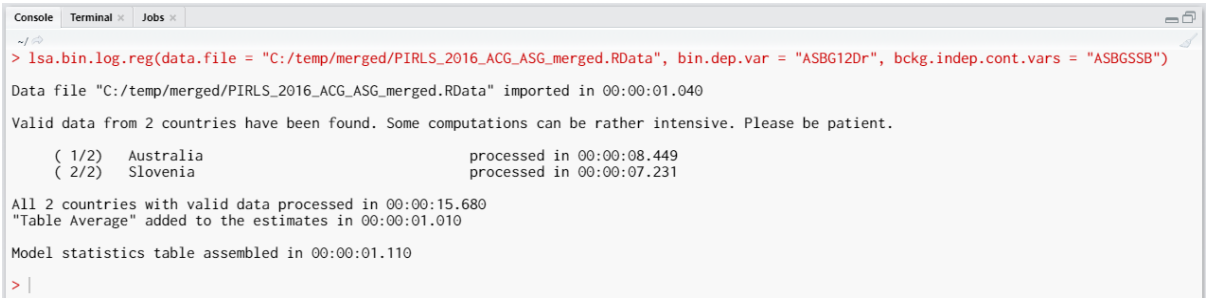
```
lsa.bin.log.reg(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
                bin.dep.var = "ASBG12Dr",
                bckg.indep.cont.vars = "ASBGSSB")
```

Nekaj stvari, ki jih je treba upoštevati:

- Funkcija lahko kot odvisno sprejme eno binarno spremenljivko. Neodvisne spremenljivke so lahko več ozadnih/kontekstnih spremenljivk in/ali nabori PV. Če so PV vključene kot neodvisne spremenljivke, bo vsak nabor PV predstavljen z njihovim osnovnim imenom. Npr., pet PV za rezultat na preizkusu branja: ASRREA01, ASRREA02, ASRREA03, ASRREA04 in ASRREA05. V argumentu `PV.root.corr` moramo določiti le osnovo PV, »ASRREA«. Funkcija bo uporabila to osnovo/ime in izbrala vseh pet PV ter jih vključila v izračune.
- Pri mednarodnih raziskavah je treba vse analize izvajati ločeno po državah. Ni pa treba dodajati spremenljivke ID države (IDCNTRY ali CNT v PISA) kot razdelitvene spremenljivke. Funkcija jo bo samodejno prepoznala in dodala v vektor `split.vars`.
- Ni treba izrecno določiti utežne spremenljivke. Če utežna spremenljivka ni izrecno določena, bo uporabljena privzeta utež (v tem primeru skupna utež učenca), odvisno od združenih podatkov o anketirancih, ki je samodejno prepoznana. Če imate dober razlog za spremembo utežne spremenljivke, lahko to storite z dodajanjem `weight.var = "SENWGT"`.
- Če output datoteka ni določena, bo output shranjen z imenom »Analysis.xlsx« v delovni imenik (prikličete ga lahko z `getwd()`).

- Če v sintakso sklica izrecno ne dodate `open.output = FALSE`, se bo output datoteka odprla po vseh končanih izračunih. To je uporabno, kadar se izvede več klicev sintaks za različne analize in ni potrebna takojšnja preglednost outputa.

Izvajanje zgornje kode bo natisnilo naslednji izhod v konzoli RStudia:



Ko bodo vse operacije končane, bo output zapisan na disk kot MS Excel-delovni zvezek. Če je `open.output` nastavljeno na `TRUE` (privzeto), se bo datoteka odprla v privzetem programu za preglednice (običajno MS Excel).

Kategorne spremenljivke lahko dodate kot kontrastno kodirane spremenljivke, kjer lahko testirate pomembnost razlik med kategorijami v odvisni spremenljivki. Trenutno funkcija podpira naslednje kontrastne sheme: `dummy`, `deviation` in `simple`. Preverimo razlike v logaritemski verjetnosti med učenkami in učenci, pri čemer upoštevamo občutek pripadnosti šoli pri učencih (ASBGSSB; za informacije o tem, kako je ta lestvica zgrajena in kakšne so njene lastnosti, preverite tehnično dokumentacijo PIRLS 2016). Ta analiza nadgrajuje prejšnjo, saj dodaja spol učencev (ASBG01) kot kategorno ozadno spremenljivko v argument funkcije `bckg.indep.cat.vars`. Spremenljivka spola učencev (ASBG01) ima dve veljavni vrednosti:

- Dekle;
- Fant.

Dodati moramo ASBG01 kot vrednost argumenta `bckg.indep.cat.vars` funkcije `lsa.bin.log.reg`. Funkcija bo samodejno določila veljavne vrednosti, vendar moramo določiti referenčno kategorijo kot vrednost argumentov `bckg.cat.contrasts` (vrsta kontrastnega kodiranja) in `bckg.ref.cat` (referenčna kategorija). Če ne določimo vrednosti za `bckg.cat.contrasts` (kar bomo tudi storili), bo funkcija samodejno izračunala regresijske koeficiente z `dummy` kodiranjem (presečišče bo logaritem verjetnosti za odvisno spremenljivko za učence, ki spadajo v kategorijo, ki smo jo izbrali kot referenco). Regresijski koeficienti za preostale kategorije (v tem primeru le eno, ker imamo dva spola) bodo razlike v logaritemski verjetnosti za učence, ki spadajo v katero koli drugo kategorijo razen referenčne. Če je potrebna katera koli druga kontrastna shema, jo je treba izrecno določiti z argumentom `bckg.cat.contrasts`. Kot referenco bomo določili prvo kategorijo (»Dekle«). Kot kontrolno spremenljivko z vrednostjo `bckg.indep.cont.vars` bomo dodali občutek pripadnosti šoli (ASBGSSB; za informacije o tem, kako je ta lestvica zgrajena in kakšne so njene lastnosti, preverite tehnično dokumentacijo PIRLS 2016) kot kontrolno spremenljivko z vrednostjo `bckg.indep.cont.vars`. Sintaksa klica izgleda takole:

```
lsa.bin.log.reg(data.file =
"C:/temp/merged/PIRLS_2016_ACG_ASG_merged.RData",
                bin.dep.var = "ASBG12Dr",
                bckg.indep.cont.vars = "ASBGSSB",
                bckg.indep.cat.vars = "ASBG01",
                bckg.ref.cats = 1)
```

Izvedba zgornje sintakse bo prepisala prejšnji izhod, ker ima definirano isto ime datoteke (v konzoli bo prikazano opozorilo). Stolpci na listu »Estimates« bodo zdaj drugačni.

4.7.4 Izračun binarnih logističnih regresijskih koeficientov z uporabo grafičnega vmesnika (GUI)

Za zagon uporabniškega vmesnika RALSA v RStudiu izvedite naslednji ukaz:

```
ralsaGUI()
```

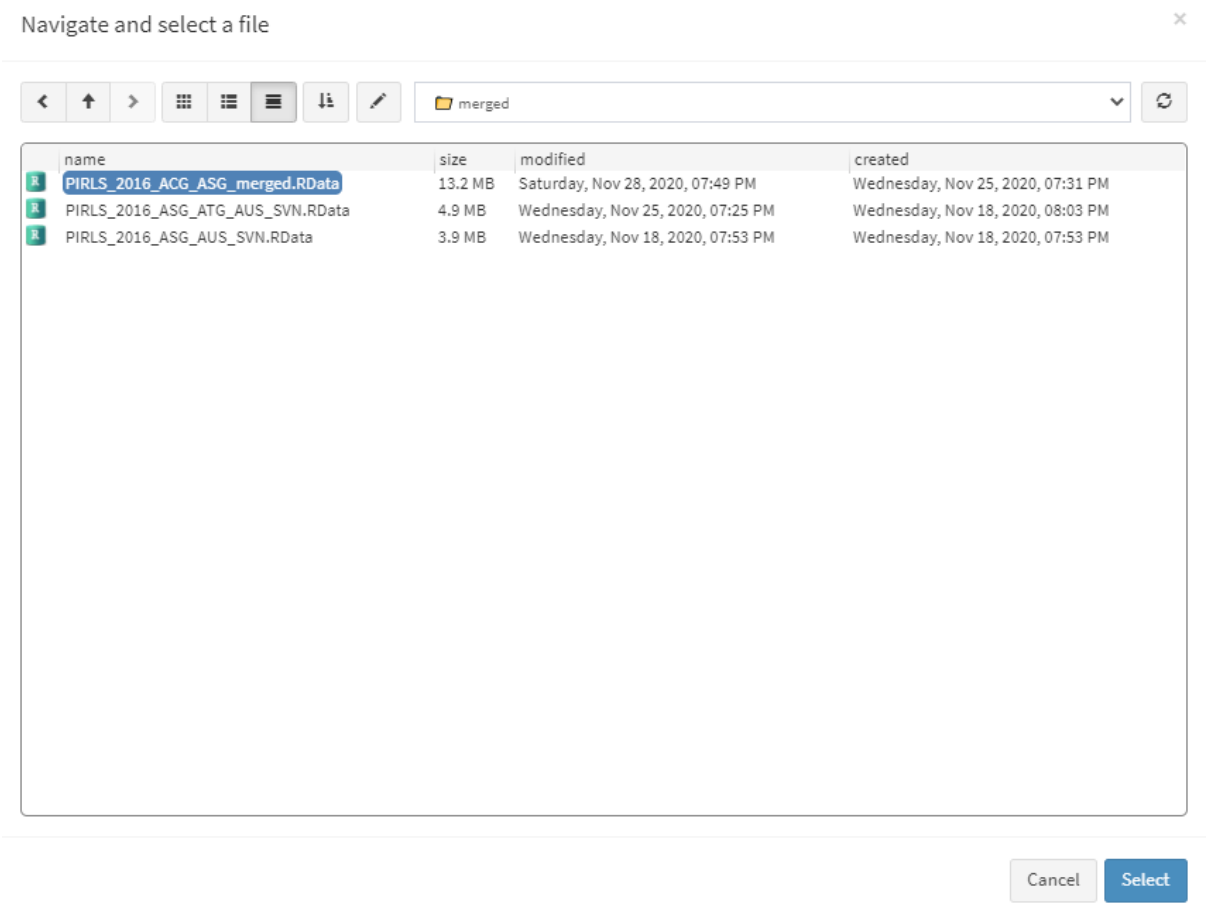
V primerih, ki sledijo, združite nov datotečni niz s podatki PIRLS 2016 za Avstralijo in Slovenijo, pri čemer vzamete vse spremenljivke, ki se nanašajo na učence in ravnatelje. Združeno datoteko lahko poimenujete **»PIRLS_2016_ACG_ASG_merged.RData«**.

Za začetek izračunajmo binarne logistične regresijske koeficiente za model, ki napoveduje, ali se bodo učenci v Avstraliji in Sloveniji strinjali ali ne strinjali, da jih učitelji obravnavajo pravično (spremenljivka ASBG12D), glede na njihov občutek pripadnosti šoli (ASBGSSB; za informacije o tem, kako je ta lestvica zgrajena in kakšne so njene lastnosti, preverite tehnično dokumentacijo PIRLS 2016). Logistična regresija kot odvisne spremenljivke sprejema samo binarne (torej dihotomne) spremenljivke, medtem ko so odgovori na vprašanje, koliko se učenci strinjajo ali ne strinjajo, da jih učitelji obravnavajo pravično (ASBG12D), organizirani v štiri ločene kategorije:

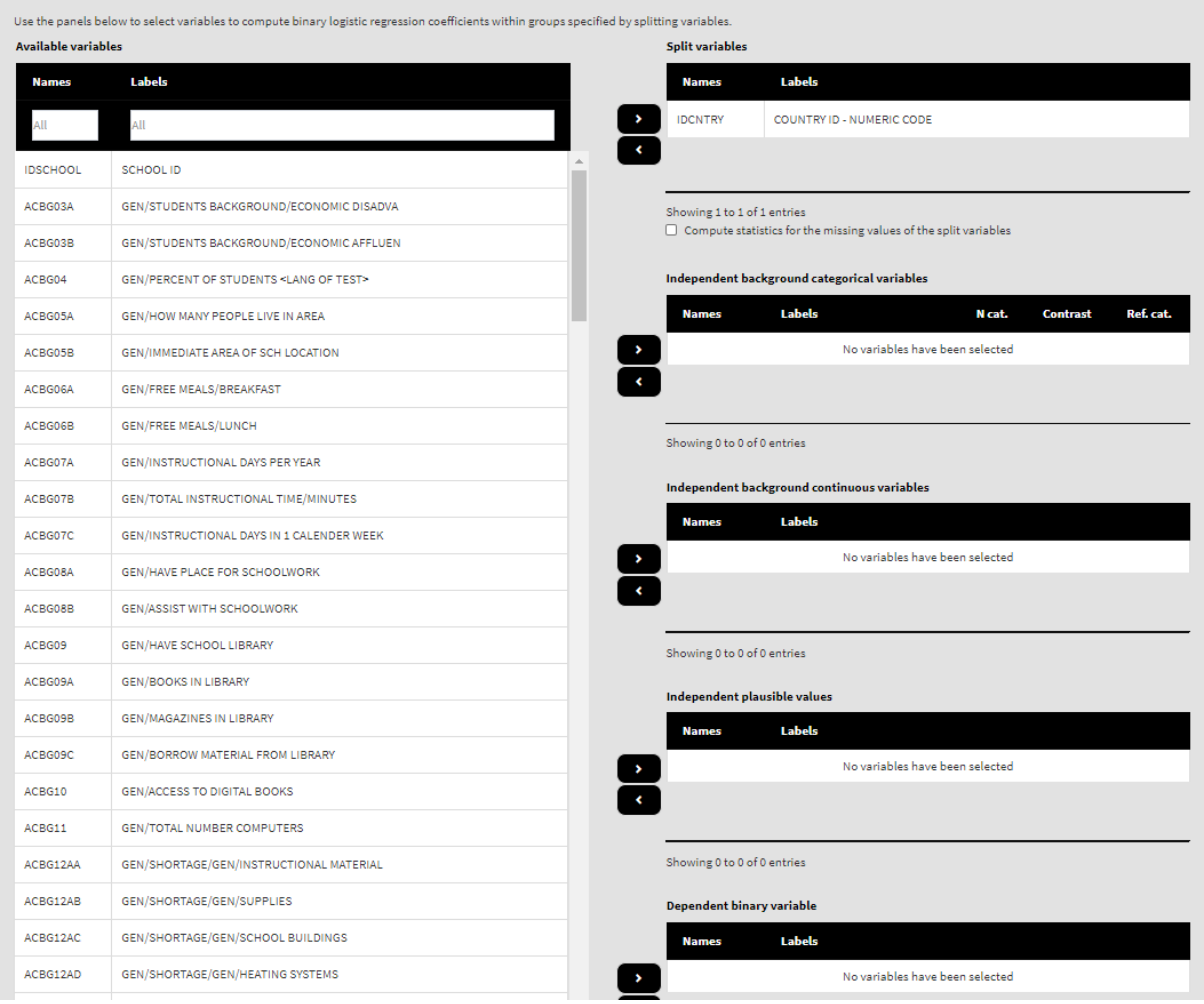
- Zelo se strinjam;
- Strinjam se;
- Ne strinjam se;
- Sploh se ne strinjam.

Zato moramo najprej ponovno kodirati ASBG12D tako, da združimo kategorije v dve: »Sploh se ne strinjam« in »Ne strinjam se« bomo združili v prvo kategorijo, medtem ko bosta »Zelo se strinjam« in »Strinjam se« združeni v drugo kategorijo. Da izvedete ta ponovna kodiranja, morate spremenljivko ASBG12D ponovno kodirati v novo spremenljivko. Poimenujmo jo kot ASBG12Dr ali z imenom, ki vam ustreza. Pojdite na **Data preparation > Recode variables**, naložite združeno datoteko **»PIRLS_2016_ACG_ASG_merged.RData«** in ponovno kodirajte ASBG12D v ASBG12Dr, tako da združite »Sploh se ne strinjam« in »Ne strinjam se« v eno kategorijo (1 – »Ne strinjam se«) ter »Zelo se strinjam« in »Strinjam se« v drugo kategorijo (2 – »Strinjam se«). Ne pozabite dodeliti vrednosti »Izpuščeno ali neveljavno« kot manjkajočo vrednost.

Ko je ponovno kodiranje končano, iz menija na levi izberite **Analysis types > Binary logistic regression**. Ko ste v meniju **Binary logistic regression** v GUI, kliknite gumb **Choose data file**. Pojdite v mapo, ki vsebuje združeno datoteko **»PIRLS_2016_ACG_ASG_merged.RData«**, jo izberite in kliknite **Select**.



Ko je datoteka naložena, boste na levi strani videli panel z razpoložljivimi spremenljivkami (angl. *available variables*) ter nabor panelov na desni, kamor lahko dodate spremenljivke s seznama razpoložljivih. Nad paneli boste prav tako videli informacije o naloženi datoteki.



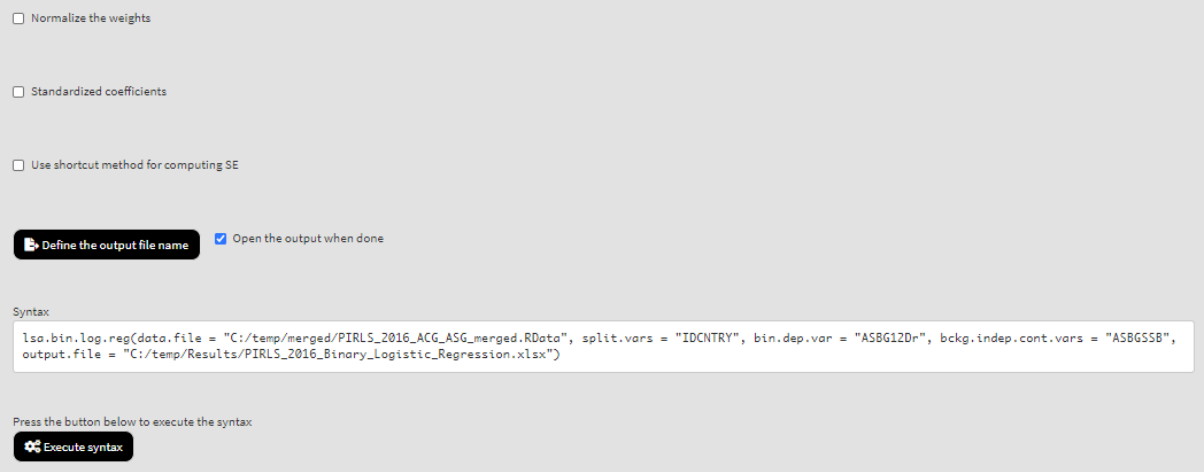
Uporabite miško, da izberete spremenljivke s seznama **Available variables** in jih s pomočjo puščic na sredini zaslona dodate v različna polja (ali jih odstranite), da nastavite analizo. Za hitro iskanje potrebnih spremenljivk lahko uporabite iskalna polja na vrhu panelov.

Izračunajmo logistične regresijske koeficiente, pri čemer bo nova spremenljivka odvisna, ASBGSSB pa neodvisna. Izberite novo rekodirano spremenljivko ASBG12Dr s seznama **Available variables** in jo s pomočjo puščice na desno premaknite na seznam **Dependent binary variable**. Izberite spremenljivko ASBGSSB s seznama **Available variables** in jo s pomočjo puščice na desno premaknite na seznam **Independent background continuous variables.**

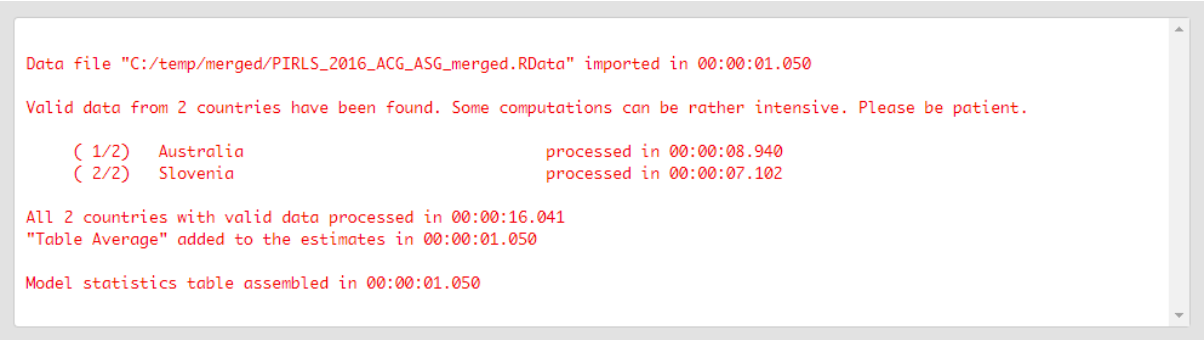
To je vse, kar je treba narediti. Pomaknite se navzdol in kliknite na **Define output file name**. Pomaknite se do mape **»C:/temp/Results«** (ali do mape, kamor želite shraniti output datoteko) in določite ime izhodne datoteke. Ko to storite, se bo poleg možnosti **Define the output file name** pojavilo potrditveno polje. Če ga označite, se bo output odprl, ko bodo vsi izračuni končani.

Pod tem boste videli potrditveno polje **Normalize the weights**. Če je označeno, se uteži pred uporabo v izračunih normalizirajo.

Videli boste tudi potrditveno polje **Standardized coefficients**. Če ga označite, se bodo spremenljivke standardizirale, preden bodo izračuni zaključeni.



Kliknite na gumb **Execute syntax**. Na dnu zaslona se bo pojavila konzola GUI, ki bo beležila vse izvedene operacije:



Nekaj stvari, ki jih je treba upoštevati:

1. Funkcija lahko kot odvisno spremenljivko vzame eno binarno spremenljivko. Neodvisne spremenljivke so lahko več ozadnih/kontekstualnih spremenljivk in/ali nizi PV. Če so PV vključene kot neodvisne spremenljivke, bo vsak niz PV predstavljen z osnovnim imenom. Npr., pet PV za splošno bralno uspešnost: ASRREA01, ASRREA02, ASRREA03, ASRREA04 in ASRREA05. Osnovno ime vseh PV v nizu bo prikazano na seznamu **Available variables** (ali na seznamu **Independent plausible values**, če je izbran tam), v primeru splošne bralne uspešnosti pa bo to ASRREA. Funkcija bo uporabila to osnovno ime, da bo izbrala vseh pet PV in jih vključila v izračune.
2. Pri mednarodnih raziskavah se morajo vse analize izvajati ločeno po državah. Spremenljivka za ID države (IDCNTRY, ali CNT v PISA) je vedno izbrana kot prva razdelitvena spremenljivka in je ni mogoče odstraniti s panela **Split variables**.
3. Privzeta spremenljivka za uteži je izbrana in samodejno dodana na panel **Weight variable**. Zamenjate jo lahko z drugo utežno spremenljivko, ki je na voljo v naboru podatkov. Če je izbrana privzeta utež, ta ne bo prikazana v oknu s sintakso. Če na panelu **Weight variable** ni izbrane nobene uteži, se bo samodejno uporabila privzeta.
4. Če je odkljukano polje **Standardized coefficients**, bodo regresijski koeficienti izračunani na standardiziranih spremenljivkah, beta-koeficienti pa bodo vključeni v izhodne podatke.

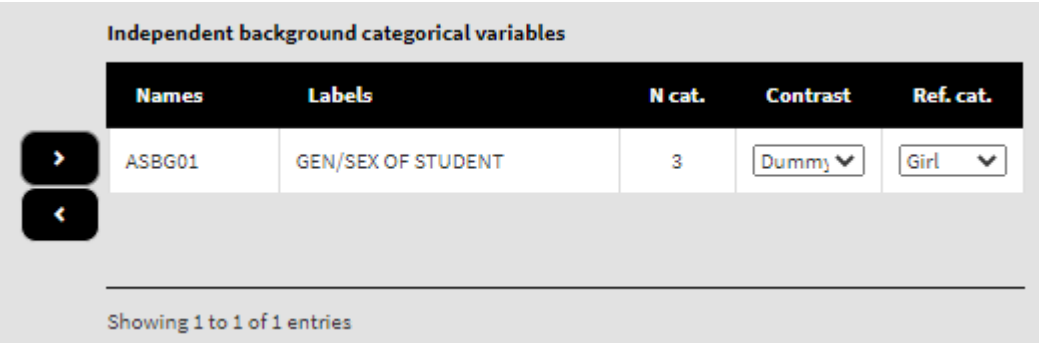
5. Polje **Use shortcut method for computing SE** ni privzeto odkljukano. To bo omogočilo, da funkcija izračuna standardno napako z uporabo »polne« metode za komponento vzorčne variance.

Če je potrjeno polje **Open the output when done**, se bo izhodni rezultat samodejno odprl v privzeti program za preglednice (običajno MS Excel), ko bodo vsi izračuni zaključeni.

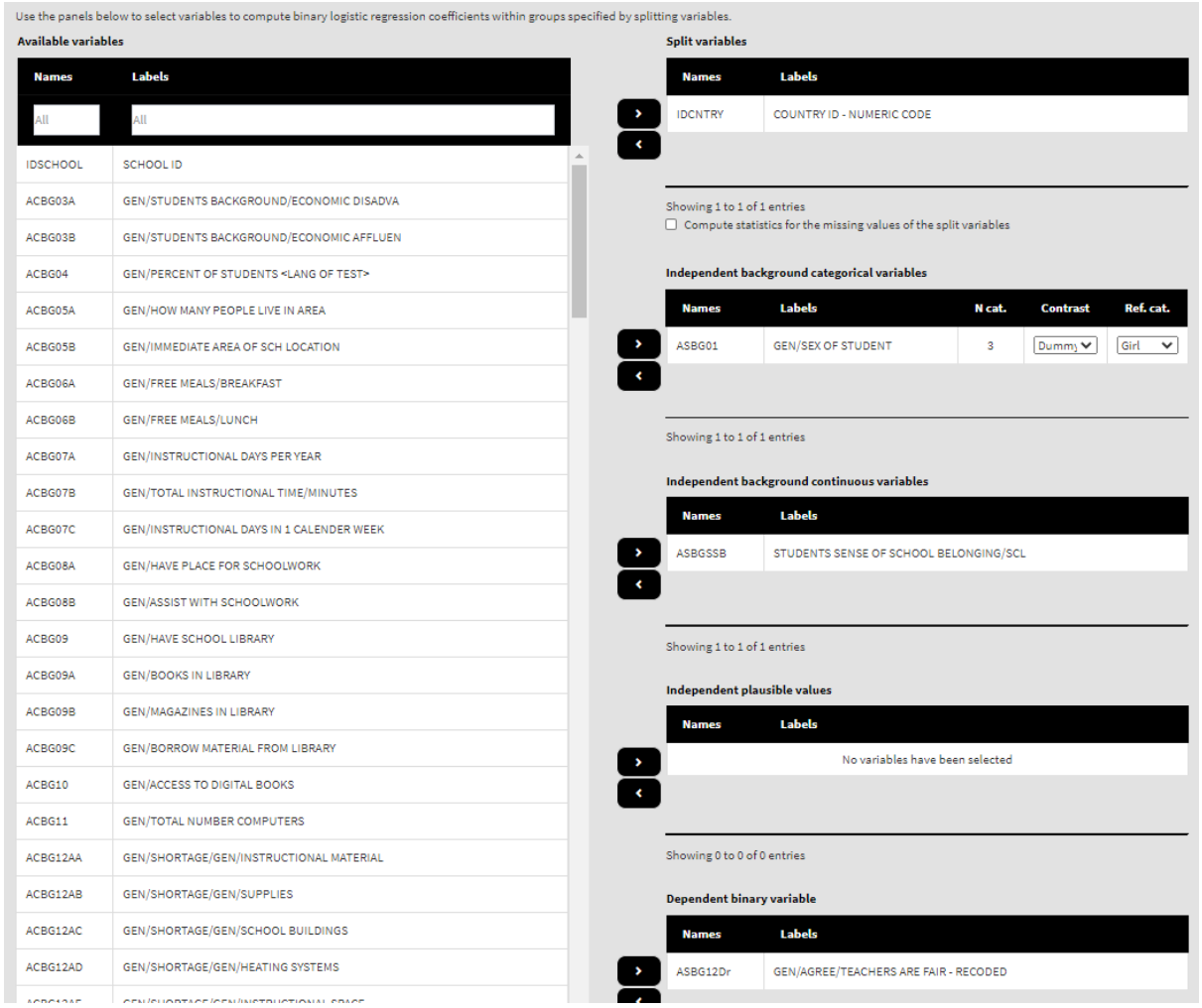
Kategorijske spremenljivke je mogoče dodati kot spremenljivke s kodiranjem kontrastov, pri čemer se lahko preizkusi pomembnost razlik med kategorijami v odvisni spremenljivki. Trenutno lahko funkcija deluje z naslednjimi shemami kontrastov: **dummy**, **deviation** in **simple**. Preizkusimo razlike v logaritmu kvot med dekleti in fanti, ko nadziramo občutek pripadnosti šoli učencev (ASBGSSB; za informacije o tem, kako je ta lestvica zgrajena in kakšne so njene lastnosti, preverite tehnično dokumentacijo PIRLS 2016). Ta analiza razširja prejšnjo, tako da doda spol učencev (ASBG01) kot kategorialno ozadno spremenljivko na seznam **Independent background categorical variables**. Spremenljivka za spol učencev (ASBG01) ima dve veljavni vrednosti:

- Dekle;
- Fant.

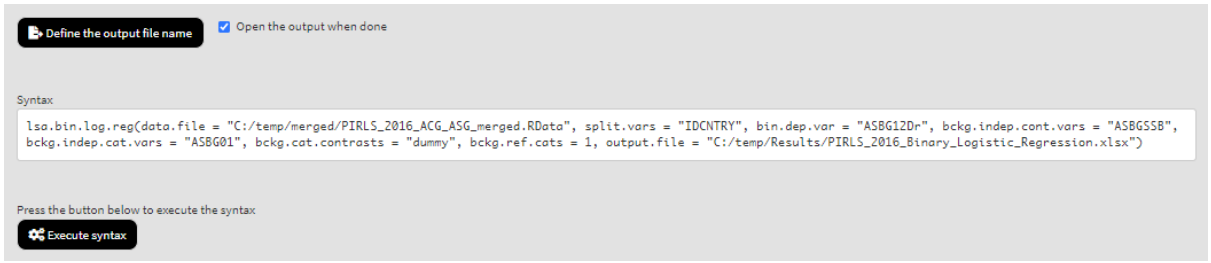
Poiščite spremenljivko ASBG01 na seznamu **Available variables** (uporabite lahko filter na vrhu), jo izberite in dodajte na seznam **Independent background categorical variables** s pomočjo desnega puščičnega gumba. Videli boste, da bo seznam samodejno prikazal število kategorij za spremenljivko, spustni seznam z različnimi kodirnimi shemami (**dummy**, **deviation**, **simple**, glejte stolpec **N cat.**) ter spustni seznam s kategorijami spremenljivke, med katerimi lahko izbirate:



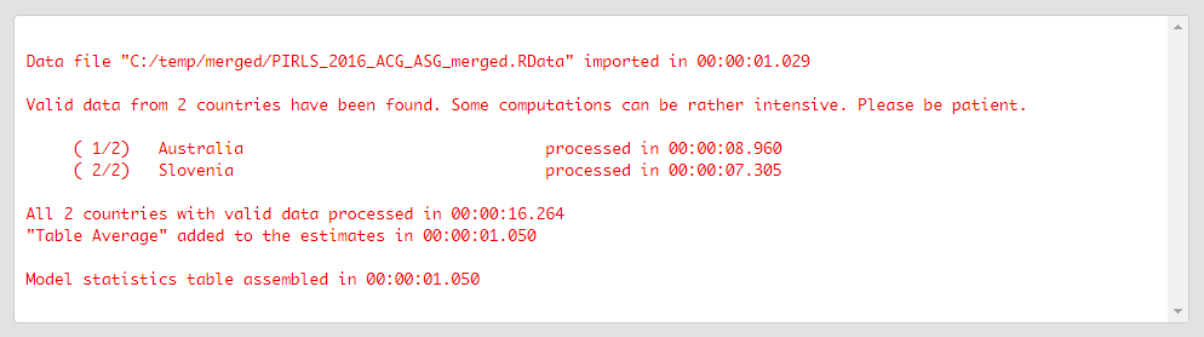
Privzeto je izbrana shema kodiranja kontrastov **dummy** in prva razpoložljiva kategorija kot referenca. Pustimo privzete nastavitve. Funkcija bo samodejno izračunala regresijske koeficiente s kodiranjem **dummy** (presečišče bo logaritem kvot za odvisno spremenljivko za učence, ki spadajo v kategorijo, ki smo jo izbrali za referenco – glejte nadaljevanje) in regresijske koeficiente za preostale kategorije (v tem primeru bo to le ena, saj imamo dva spola) kot razlike v logaritmu kvot za učence, ki spadajo v katero koli drugo kategorijo razen referenčne. Če je potrebna druga shema kontrasta, jo lahko spremenimo s klikom na spustni meni in izbiro med **deviation** ali **simple**. Kot referenco bomo pustili bomo prvo kategorijo (**Dekle**). Spremenljivko **ASBGSSB** pustimo kot kontrolno spremenljivko na seznamu **Independent background continuous variables**:



Ker uporabljamo aplikacijo z **Binary logistic regression** neposredno po izvedbi prejšnje analize, imamo še vedno shranjene vse preostale nastavitve iz prejšnje analize. Ni treba spreminjati nobene od teh nastavitev, razen če želite. Lahko pa spremenite ime output datoteke, sicer bo prepisana. Upoštevajte, da se bo prikazana sintaksa spremenila, kar bo odražalo vključitev spremenljivke ASBG01 kot neodvisne kategorialne spremenljivke:



Pritisnite gumb **Execute syntax**. Konzola GUI se bo posodobila in beležila vse izvedene operacije:



Če je polje **Open the output when done** označeno, se bo output po zaključku vseh izračunov samodejno odprl v privzetem programu za preglednice (običajno MS Excel).

5. Reference

Bate, S. M. (2004). *Generalized Linear Models for Large Dependent Data Sets* [Doctoral Thesis]. University of London.

Foy, P. (ur.) (2018). *PIRLS 2016 User Guide for the International Database*. TIMSS & PIRLS International Study Center.

Foy, P., & LaRoche, S. (2017). Estimating Standard Errors in the PIRLS 2016 Results. V: M. O. Martin, I. V. S. Mullis, & M. Hooper (ur.), *Methods and Procedures in PIRLS 2016* (str. 4.1–4.22). Lynch School of Education, Boston College.

Foy, P., & Yin, L. (2016). TIMSS 2015 Achievement Scaling Methodology. V: M. O. Martin, I. V. S. Mullis, & M. Hooper (ur.), *Methods and Procedures in TIMSS 2015* (str. 13.1–13.62). TIMSS & PIRLS International Study Center.

Hilbe, J. M. (2015). *Practical Guide to Logistic Regression*. CRC Press.

LaRoche, S., Joncas, M., & Foy, P. (2016). Sample Design in TIMSS 2015. V: M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (str. 3.1–3.37). TIMSS & PIRLS International Study Center.

LaRoche, S., Joncas, M., & Foy, P. (2017). Sample Design in PIRLS 2016. In M. O. Martin, I. V. S. Mullis, & M. Hooper (ur.), *Methods and Procedures in PIRLS 2016* (str. 3.1–3.34). Chestnut Hill, MA: Lynch School of Education, Boston College.

Mirazchiyski, P.V. (2021). RALSA: The R analyzer for large-scale assessments. *Large-scale Assess Educ 9*(21), str. 1–24. https://doi.org/10.1186/s40536-021-00114-4

Mirazchiyski, P.V. (2021). RALSA: Design and Implementation. *Psych, 3*(2), str. 233–248. https://doi.org/10.3390/psych3020018

OECD. (in press). *PISA 2018 Technical Report*. OECD.

Rao, J. N. K., & Scott, A. J. (1984). On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data. *The Annals of Statistics*, *12*(1). https://doi.org/10.1214/aos/1176346391

Rao, J. N. K., & Scott, A. J. (1987). On Simple Adjustments to Chi-Square Tests with Sample Survey Data. *The Annals of Statistics*, *15*(1), str. 385–397.

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting. *Educational Researcher, 39*(2), str. 142–151.

Rutkowski, L., Rutkowski, D., & von Davier, M. (2014). A Brief Introduction to Modern International Large-Scale Assessment. V: L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of International Large-Scale Assessments: Background, Technical Issues, and Methods of Data Analysis* (str. 3–10). CRC Press.

Skinner, C. (2019). Analysis of Categorical Data for Complex Surveys. *International Statistical Review*, *87*(S1), str. S64–S78. https://doi.org/10.1111/insr.12285

UCLA: Statistical Consulting Group (2020). "R LIBRARY CONTRAST CODING SYSTEMS FOR CATEGORICAL VARIABLES." *IDRE Stats - Statistical Consulting Web Resources*. Dostop: 16, junij, 2020 (https://stats.idre.ucla.edu/r/library/r-library-contrast-coding-systems-for-categorical-variables/).