Technical report - H Scoring

Stef van Buuren

20221027

## D-score and DAZ

Suppose that you have administered SF, LF or both to one or more children. A next step is to calculate the developmental score (-score) and the age-adjested equivalent (DAZ) for each child. This step is known as **scoring**. The present section provides recipes on how to calculate these score. We may pick one of the following two methods:

1. Online calculator. The online Shiny app <https://tnochildhealthstatistics.shinyapps.io/dcalculator/> is a convenient option for users not familiar with R. The app contains online documentation and instructions and will not be further discussed here.
2. R package dscore. The R package [dscore]<https://CRAN.R-project.org/package=dscore> is a flexible option with all the tools needed to calculate the -score. It is an excellent choice for users familiar with R and users who like to incorporate -score calculations into a workflow.

### Preliminaries

* We use the R language. If you are new to R consult the [Getting Started with R](https://support.rstudio.com/hc/en-us/articles/201141096-Getting-Started-with-R) site;
* You need to install the R package dscore on your local machine;
* The child data need to be stored as a data.frame, a standard R tabular structure;
* You need to run the dscore() function to calculate the -score and DAZ. The function returns a table with six columns with the estimates with the same number of rows as your data.

### Install the dscore package

The dscore package contains tools to

* Map your item names to the GSED convention
* Calculate *D*-scores from item level responses
* Transform the *D*-scores into DAZ, age-standardised Z-scores

The required input consists of *item level* responses on milestones from widely used instruments for measuring child development, including the GSED LF and GSED SF instruments.

There are two versions of the dscore package. The curated and published version that reside on [CRAN](https://CRAN.R-project.org/package=dscore) is recommended for general use. Install the package as

install.packages("dscore")

In some cases, there may be a newer version that includes extensions and bug fixes that are not yet available on CRAN. You can install the development version from [GitHub](https://github.com/) with:

install.packages("remotes")  
remotes::install\_github("d-score/dscore")

Note that the second route requires that you have a local C++ compiler available for building the package from source. We use here dscore 1.6.5.

### Organise your data

The dscore() function accepts item names that follow the GSED 9-position schema. A name with a length of nine characters identifies every milestone. The following table shows the construction of names.

| Position | Description | Example |
| --- | --- | --- |
| 1-3 | instrument | by3 |
| 4-5 | developmental domain | cg |
| 6 | administration mode | d |
| 7-9 | item number | 018 |

Thus, item by3cgd018 refers to the 18th item in the cognitive scale of the Bayley-III. The label of the item can be obtained by

library(dscore)  
get\_labels("by3cgd018")

## by3cgd018   
## "Inspects own hand"

The instrument codes for the GSED LF and GSED SF are gto and gpa, respectively. The first three items in the SF (2020 version) are:

items\_gpa <- get\_itemnames(instrument = "gpa", order = "indm")  
head(items\_gpa, 3)

## [1] "gpalac001" "gpacgc002" "gpafmc003"

The order argument is needed to sort items according to sequence number 1 to 139 as in the SF form (2020 version). For the LF (2020 version) we rely on the default order that provides items numbered within streams fine motor (fm), gross motor (gm) and language (lg), respectively.

items\_gto <- get\_itemnames(instrument = "gto")  
head(items\_gto, 3)

## [1] "gtofmd001" "gtofmd002" "gtofmd003"

The labels from these items can be obtained as

labels\_gto <- get\_labels(items\_gto)

[*NOTE: For user convenience: We need to generalise get\_itemnames() to different versions of SF and LF (that use other item sequences).*]

### Response data format

The dscore package contains methods for binary items only. Each child at a given age occupies a separate row. The gsample dataset in the dscore package is an example of data in the proper format. It contains 10 rows with responses of 10 children measured on the GSED SF and GSED LF at various ages. The dataset has 295 columns. The first seven columns of the dataset look like:

gsample[, 1:7]

## subjid agedays gpalac001 gpacgc002 gpafmc003 gpasec004 gpamoc005  
## 1 1 811 NA NA NA NA NA  
## 2 2 898 NA NA NA NA NA  
## 3 3 203 NA NA NA NA NA  
## 4 4 966 NA NA NA NA NA  
## 5 8 770 NA NA NA NA NA  
## 6 9 306 NA NA NA NA NA  
## 7 10 1214 NA NA NA NA NA  
## 8 11 722 NA NA NA NA NA  
## 9 13 131 1 1 1 1 1  
## 10 14 50 1 1 1 1 1

Each row corresponds to a visit. Columns starting with gpa hold the responses on items from GSED-SF items. Item level data should be coded as 0 (FAIL), 1 (PASS) or NA (not administered, not answered). Note that the easiest items are not administered to children older then 6 months, so we see a large block of NA’s for the early SF items.

### Calculate -score

Once the data are in proper shape, calculation of the -score is straightforward. The gssample dataset has properly named columns that identify each item.

Since one SF item (gpamoc008) is not present in the data, we need to remove it from the vector with item names. You may calculate the -score from the remaining 138 SF items as

items\_gpa <- setdiff(items\_gpa, "gpamoc008")  
sf <- dscore(gsample, items = items\_gpa, xname = "agedays", xunit = "days")  
head(sf)

## a n p d sem daz  
## 1 2.220 29 0.759 65.2 0.734 -0.416  
## 2 2.459 39 0.692 68.3 0.735 -0.173  
## 3 0.556 50 0.660 38.4 0.943 1.055  
## 4 2.645 36 0.750 73.4 0.661 0.895  
## 5 2.108 50 0.500 65.5 0.616 -0.009  
## 6 0.838 50 0.700 41.6 1.008 -0.600

The table below provides the interpretation of the output:

| Name | Interpretation |
| --- | --- |
| a | Decimal age in years |
| n | Number of items used to calculate the -score |
| p | Percentage of passed milestones |
| d | -score estimate, EAP estimate |
| sem | Standard error of measurement, standard deviation of the posterior |
| daz | -score corrected for age |

Calculation of the -score for the GSED LF is similar:

# calculate D-score for LF  
lf <- dscore(gsample, items = items\_gto, xname = "agedays", xunit = "days")  
head(lf)

## a n p d sem daz  
## 1 2.220 45 0.556 67.0 0.627 0.097  
## 2 2.459 53 0.623 70.8 0.569 0.552  
## 3 0.556 34 0.559 34.1 0.840 -0.153  
## 4 2.645 54 0.518 70.6 0.545 0.053  
## 5 2.108 58 0.172 37.5 1.194 -4.448  
## 6 0.838 32 0.562 44.9 0.764 0.315

In general, the -score from SF and LF may differ by a few points, but in exceptional cases (e.g. for row 5) the difference can be large.

It is possible to calculate -score for item subsets by tweaking the items argument. If you do not provide the items argument, then the -score is calculated from all items available (here GSED SF and GSED LF). For example,

# calculate D-score from all available items from all instruments  
head(dscore(gsample, xname = "agedays", xunit = "days"))

## a n p d sem daz  
## 1 2.220 74 0.635 66.3 0.51223 -0.121  
## 2 2.459 92 0.652 70.0 0.38267 0.296  
## 3 0.556 84 0.619 36.0 0.58301 0.365  
## 4 2.645 90 0.611 71.7 0.46082 0.383  
## 5 2.108 108 0.324 45.0 0.00006 -3.837  
## 6 0.838 82 0.646 43.6 0.64544 -0.061

### Phase 1 references and DAZ

We used the GSED Phase I data to calculate age-conditional reference scores for the -score. The references are based on about 12,000 administration of the GSED SF and GSED LF from Bangla Desh, Pakistan and Tanzania. We may extract the references as

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

ref <- builtin\_references %>%   
 filter(pop == "phase1") %>%   
 select(pop, age, mu, sigma, nu, tau, SDM2, SD0, SDP2)  
head(ref)

## pop age mu sigma nu tau SDM2 SD0 SDP2  
## 1 phase1 0.0383 13.7 0.246 1.17 15.4 6.04 13.7 20.7  
## 2 phase1 0.0575 14.4 0.232 1.21 15.5 6.70 14.4 21.3  
## 3 phase1 0.0767 15.0 0.221 1.24 15.7 7.35 15.0 21.9  
## 4 phase1 0.0958 15.7 0.210 1.27 15.8 8.01 15.7 22.5  
## 5 phase1 0.1150 16.4 0.201 1.30 15.9 8.68 16.4 23.1  
## 6 phase1 0.1342 17.0 0.192 1.32 16.0 9.36 17.0 23.7

The columns mu, sigma, nu and tau are the age-varying parameters of a Box-Cox (BCT) distribution.

The figure below plots the -2SD, 0SD and +2SD centiles plus 20 -scores (10 LF and 10 SF) for the gsample data.

