

## Supplemental Material S1. Examples of creating the smoothed percentile growth curves.

Most suggestions for creating the growth percentile norms refer to the LMS method proposed by Cole and Green (see references in Rigby & Stasinopoulos, 2005, and Stasinopoulos et al., 2017). This method takes into account changes in the median, as well as variability and skewness of the normed quantities with age. It should be noted that the method has been developed for physiological measures of growth, i.e., continuously varying variables. The use for test scores may require some adjustments.

Alternatives to the parametric LMS method include nonparametric estimates. The R package *gamlss* (generalized additive models for location, scale and shape; Rigby & Stasinopoulos, 2005) provides functions that can fit parametric or nonparametric estimates of percentile growth curves. A somewhat different, semi-parametric approach is implemented in a developmental package *cNorm* (for details see Lenhard et al., 2019).

Below are some examples of using the *gamlss* package to derive centile growth curves. All use two variables with data that are needed to fit the models, one specifying the scores (e.g., number of CDI words), the other age. The examples use a data set *normative\_data* with two variables, score and age. The structure of the file is:

```
score      age
127        23.42
254        28.17
...
```

Note that age in this example is in months but with decimal points, i.e., derived from ages in days. This is not necessary but the more precise the age information is, the more accurate fit can be expected, especially in younger children.

The first example is the use of the function `lms()`.

```
growth_model<-lms(pg, vek, data=datt)
```

This function estimates the growth curves using three underlying distribution, the original method proposed by Cole and Green, and two extensions, Box-Cox power exponential distribution and Box-Cox *t*-distribution. If the estimates for all three fail, it provides an estimate using the normal distribution. This is rarely appropriate; it is often desirable to fit the models assuming a specific distribution. The following example shows a model fitting the Box-Cox model according to Cole and Green:

```
growth_model<-gamlss(score~pbm(age, lambda=1),
  sigma.formula=~pb(age, lambda=1),
  nu.formula = pb(age, lambda=1),
  family=BCCGo(), data=normative_data)
```

The parameters `sigma` and `nu` correspond to variance and skewness in the data, and in the above code, the formulas make these parameters dependent on age. This is needed because the variability in scores and their skew changes with age. Age is entered as a smoothed variable using *p-splines*, hence the code *pb(age)* or *pbm(age)*. The *pbm()* version of the smoothing *spline* function means a monotonic *spline*, i.e., includes the assumption that the variable scores only change in one direction. This is assumed for growth: scores should only increase with age. The monotonic version of the spline is not

used for variability and skew: these can change in both directions with age, e.g., when the norming range includes floor as well as ceiling effects on the opposite ends of the age range. The parameter *lambda* is the smoothing parameter. As the parameter value increases, the *spline* becomes more linear. If *lambda* is not specified, it is estimated by the function, but this often results in growth curves that are too wiggly.

The *gamlss* package provides extensions of the Cole and Green model that also allow specifying a model for kurtosis (the tau parameter in the following code):

```
growth_model<-gamlss(score~pbm(age,lambda=1),
  sigma.formula=~pb(age,lambda=1),
  nu.formula = pb(age,lambda=1),
  tau.formula = pb(age,lambda=1),
  family=BCPEo(), data=normative_data)
```

Note that the BCCG family of distributions does not allow zero values, which might be a problem especially when fitting data from younger children. If the data include zeros, it is recommended to use the *sin-arcsinh* distribution, which has four parameters controlling the mean, variance and the weight of the left and right tail, respectively.

```
growth_model<-gamlss(score~pbm(age,lambda=1),
  sigma.formula=~pb(age,lambda=1),
  nu.formula = pb(age,lambda=1),
  tau.formula = pb(age,lambda=1),
  family=SHASHo(), data=normative_data)
```

The *gamlss* package includes a number of additional distributions that can be used to fit the data. For example, the Wordbank percentile (Frank et al., 2016) estimates are generated using the following code, using the *beta* distribution and high values of the smoothing parameters for monotonous splines of age for mean and variability.

```
gam_mod <- gamlss(vocab ~ pbm(age, lambda = 10000),
  sigma.formula = ~ pbm(age, lambda = 10000),
  family = BE, control = gamlss.control(c.crit =
  .1), data = normative_data)
```

Fitting all the above functions provides a figure showing the growth curves with some predetermined percentiles. To obtain a plot of specific percentiles, and adjust various properties of the plot, the following code can be used, here requesting the growth curves for the 15th, 50th, and 85th percentile:

```
centiles(growth_model, cent=c(.15, .5, .85))
```

To create a table of expected raw scores that correspond to given percentiles at a given age, the following code can be used, here providing a table for ages 16 to 30 months and the percentiles 5, 15, 50, 85, and 95.

```
centile_table<-centiles.pred(growth_model, xname="age",
  xvalues=c(16:30), cent=c(5,15,50,85,95))
round(centile_table)
```

As a nonparametric alternative to using the *gamlss* models, it is possible to use the function *quantSheets()*. Here is an example of fitting the function, which generates a plot of specified percentile curves. The *predict()* function is then used to generate a table of expected raw scores for specified percentiles and a given age range.

```
x<-quantSheets(pg, vek, data=datt,  
  cent=c(5, 15, 50, 85, 95), x.lambda=1)  
centile_table<-predict(x, newdata=c(36:60))  
round(centile_table)
```

## References

- Lenhard, A., Lenhard, W., & Gary, S. (2019). Continuous norming of psychometric tests: A simulation study of parametric and semi-parametric approaches. *PLoS ONE* 14(9), e0222279. <https://doi.org/10.1371/journal.pone.0222279>
- Rigby R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). *Journal of Applied Statistics*, 54(3), 507–554.
- Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V., & De Bastiani, F. (2017). *Flexible regression and smoothing: Using GAMLSS in R*. CRC Press.