

EXPERIENCIA EN EL USO DE R PARA EL AJUSTE DE CURVAS DE CRECIMIENTO

USING R FOR ADJUSTING GROWTHS CURVES

Autores:

Mercedes Rubén Quesada¹, Lilia E. Rodríguez Chávez², Mercedes Esquivel Lauzurique³, Margarita Orue Carmona⁴

¹) CECAM, Cuba, mquesada@cecam.sld.cu

²) CECAM, Cuba, lily@cecam.sld.cu

³) Facultad de Ciencias Médicas "Julio Trigo", Cuba, mesqui@infomed.sld.cu

⁴) CECAM, Cuba, margarita@cecam.sld.cu

RESUMEN:

La construcción de curvas de crecimiento para dimensiones antropométricas es un proceso engorroso que hasta ahora, en el país, se ha realizado manualmente. Con el objetivo de automatizar este proceso se realizó una actualización del estado del arte de métodos para ajuste de curvas de crecimiento basados en programas de computación. Se presenta una experiencia realizada con datos de circunferencia de cintura de varones entre 8 y 19 años de edad, residentes en Ciudad de La Habana en el año 1998, donde se aplicaron los métodos adoptados por la Organización Mundial de la Salud (OMS) para la elaboración de patrones internacionales de crecimiento. El soporte computacional de dichos métodos se basa en el paquete GAMLSS del proyecto R, software de código abierto y completamente libre.

PALABRAS CLAVE: R, software libre, curvas de crecimiento, gamlss.

ABSTRACT:

Child growth curves construction is a tiresome and time consuming process that has been done manually till now in our country. A review of methods for adjusting growth curve was carried out in order to automate this process. Techniques and procedures adopted by the World Health Organization (WHO) for the construction of child growth standards were applied to adjust waist circumference curve for boys from 8 to 19 years old. The methods recommended by WHO rest in the use of the computational package GAMLSS, which was developed in R, free and open source software for statistical computing.

KEYWORDS: R, free software, growth curves, gamlss.

1. INTRODUCCIÓN

Existe una política nacional de migración a software libre. Aparte de cualquier razón económica, dicha política también está sustentada en la necesidad de incorporarnos de manera activa a los retos de la sociedad de la información. Por este motivo es necesario elevar el nivel de desarrollo tecnológico y que nuestros especialistas sean no sólo consumidores de software, sino que sean capaces de producirlo o extender sus funcionalidades. A las universidades corresponde una participación importante en llevar a cabo este proceso.

El Centro de Cibernética Aplicada a la Medicina (CECAM) de la Universidad de Ciencias Médicas de La Habana, está encargado de dirigir metodológicamente el desarrollo de la disciplina Informática Médica que se imparte en todas las carreras de Ciencias Médicas. Los elementos de software que se incluyen para apoyar la estadística que se imparte dentro de esta disciplina se clasifican, en su gran mayoría, dentro del software propietario. En los últimos años uno de los más usados ha sido el SPSS.

El uso de este paquete de programas estadísticos es también muy común dentro de la investigación en Salud por su versatilidad y facilidad de uso. En este contexto, un ejemplo lo constituye el análisis de los datos de las encuestas de crecimiento y desarrollo de niños y adolescentes cubanos que se ejecutan en el país desde el año 1972 por parte de los especialistas del departamento de Crecimiento y Desarrollo (DCD), radicado en la facultad de Ciencias Médicas Julio Trigo [1][2][3][4].

Entre los resultados de estos estudios están las llamadas “curvas de crecimiento” de diferentes dimensiones antropométricas. Las curvas de crecimiento son la representación gráfica, a través de curvas de percentiles seleccionados, de la relación dimensión – edad. En Cuba, hasta la fecha, el procedimiento de construcción de las curvas se ha hecho de forma manual, a partir de resúmenes de estadígrafos obtenidos en el SPSS.

Tomando en consideración la necesidad de automatizar el procedimiento de obtención de las curvas de crecimiento, especialistas del departamento de Crecimiento y Desarrollo (DCD) y del Centro de Cibernética Aplicada a la Medicina (CECAM) se propusieron actualizar el estado del arte en relación a los procedimientos computacionales para la construcción y el ajuste de curvas de crecimiento, adoptar y aplicar los universalmente más aceptados, priorizando los elaborados en software libre.

2. DESARROLLO

La publicación más completa encontrada relacionada con el tema es la que resume el trabajo de un grupo de expertos convocados por la OMS en el año 2006 con el objetivo de seleccionar métodos apropiados para la construcción de curvas de crecimiento [5]. La metodología finalmente aplicada y recomendada por esta organización, puede consultarse en distintos artículos [6],[7] y tiene varias ventajas: se basa en el criterio unificado de un grupo de expertos de distintas áreas del conocimiento con una larga y fructífera trayectoria científica; describe detalladamente los procedimientos usados de tal forma que su reproducibilidad es factible, todo el procesamiento puede

realizarse a través del paquete GAMLSS, desarrollado en R, un entorno y lenguaje para el cálculo estadístico y generación de gráficos, libre y totalmente gratuito.

Por estas razones, se decidió valorar la factibilidad del uso de R, como alternativa de software libre para el procesamiento de los resultados de estos estudios y aplicar este grupo de recomendaciones en una primera experiencia, utilizando los datos de circunferencia de cintura recolectados en sendas investigaciones, llevadas a cabo en Ciudad de La Habana en los años 1998 y 2005. En este trabajo se presentan los resultados correspondientes al sexo masculino para el año 1998.

2.1 Características generales de R

R, también conocido como "GNU S", es una implementación en software libre del premiado lenguaje S, desarrollado en los Laboratorios AT&T Bell; se dice que es un dialecto de S. Actualmente, S y R son dos lenguajes muy utilizados para el análisis estadístico de investigaciones. Se distribuye con licencia GNU GPL o General Public. La GPL no pone ninguna restricción al uso de R pero su distribución ha de ser GPL [8][9].

R puede ser descargado de forma gratuita desde <http://cran.es.r-project.org>

Entre los atractivos de R están: la capacidad de combinar, sin fisuras, análisis "preempaquetados" (ej., regresión logística) con análisis ad-hoc, específicos para una situación; capacidad de manipular, modificar datos y funciones, y producir gráficos de alta calidad. Algo a destacar es el hecho de que la comunidad de R está formada por estadísticos de gran renombre, que mantienen actualizado el sistema añadiendo nuevos paquetes y revisando los ya existentes. Existen ya extensiones específicas a nuevas áreas como es la bioinformática [8][10].

Para usuarios avanzados, R posee también un lenguaje de programación completo con el que añadir nuevas técnicas mediante la definición de funciones [11][12].

A pesar de estas características que hacen a R muy atractivo, tiene una aparente desventaja con respecto a paquetes de programas como el SPSS y es la falta de una interfaz gráfica completa, amigable, que permita un mayor grado de flexibilidad, fluidez y productividad sin tener que aprender un lenguaje de programación. Sin embargo, R provee un acceso relativamente sencillo a una gran variedad de técnicas estadísticas y gráficas, aunque al principio parezca algo complicado. Además, ya existen algunas interfaces gráficas como Brodgar, SciViews, ObveRsive, ViSta y, quizás la más utilizada, R commander que se puede obtener desde el mismo sitio de R [10].

R consta de un sistema base y paquetes que extienden su funcionalidad. Uno de ellos es GAMLSS que es la implementación de los modelos aditivos generalizados para medidas de tendencia, escala y forma (GAMLSS por sus siglas en inglés) utilizado para el ajuste de curvas de crecimiento en el estudio multicéntrico de la OMS. [13]

Este paquete permite el ajuste a los datos de diferentes modelos, que pueden utilizar gran variedad de familias de distribución tanto continuas como discretas, incluyendo distribuciones altamente sesgadas. El paquete comprende: ajuste al modelo, herramientas para la selección del modelo y para el diagnóstico estadístico así como la extracción de información a partir del

modelo ajustado (estimación de percentiles, media, desviación y parámetros del modelo). Para todas estas funciones incorpora facilidades gráficas de alta calidad.

2.2 Construcción de las curvas de crecimiento

Para construir las curvas de crecimiento con el método recomendado por la OMS, se ajustan los datos a una distribución del tipo Box-Cox-Potencia – Exponencial (BCPE) en la edad, con 4 parámetros: μ (mediana), σ (coeficiente de variación), ν (relacionado con la asimetría) y τ (relacionado con la curtosis), suavizados a través de la técnica de splines cúbicos.

El modelo ajustado se puede resumir usando la notación:

$BCPE(X=x, gl(\mu)=n_1, gl(\sigma)=n_2, gl(\nu)=n_3, gl(\tau)=n_4)$ donde n_i ; $i=1,2,3,4$; son los grados de libertad correspondientes a cada uno de los parámetros y x es la edad o una transformación de la misma.

Para seleccionar el modelo, se compararon diferentes ajustes con familias de distribución distintas utilizando la función que le da nombre al paquete, `gamlss()` [14]. Se realizaron alrededor de 20 corridas. El modelo seleccionado resultó el BCPE con $\mu=4$, $\sigma=2$, $\nu=2$, $\tau=2$ grados de libertad y la variable independiente edad elevada a la potencia 1.7

```
gamlss(formula = y ~ cs(nx, df = 4), sigma.formula = ~cs(nx,
df = 2), nu.formula = ~cs(nx, df = 2), tau.formula = ~cs(nx,
df = 2), family = BCPE, data = ncint98v, method = RS())
```

donde nx es la edad transformada.

Para el diagnóstico se usaron:

- La función `plot()` da un resumen de estadísticas univariadas para evaluar si la distribución ajustada se adecua a la distribución normal [14].

Summary of the Quantile Residuals

mean = 0.001452842

variance = 0.9991229

coef. of skewness = -0.004582074

coef. of kurtosis = 2.967186

Filliben correlation coefficient = 0.9991348

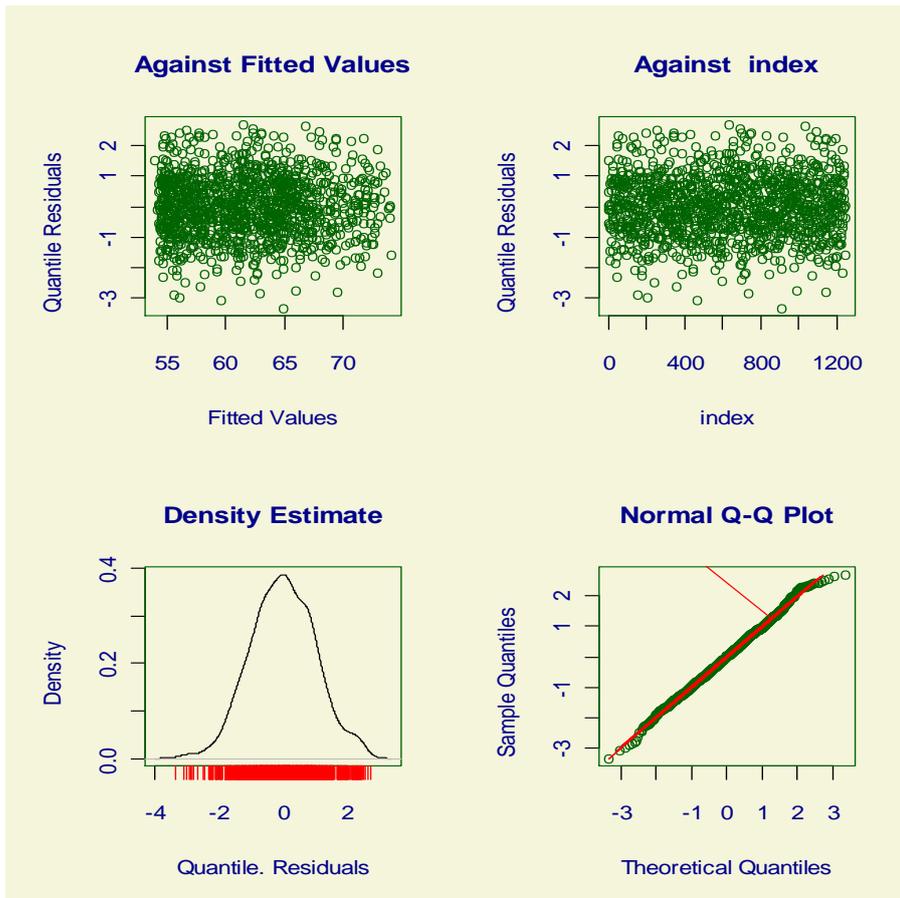


Fig. 1: Resultados de la función plot()

- La función `wp()` muestra gráficamente como se ajustan los parámetros del modelo a la distribución normal, dada la edad. Se pueden obtener resultados para el intervalo en su conjunto y para subintervalos . [14].

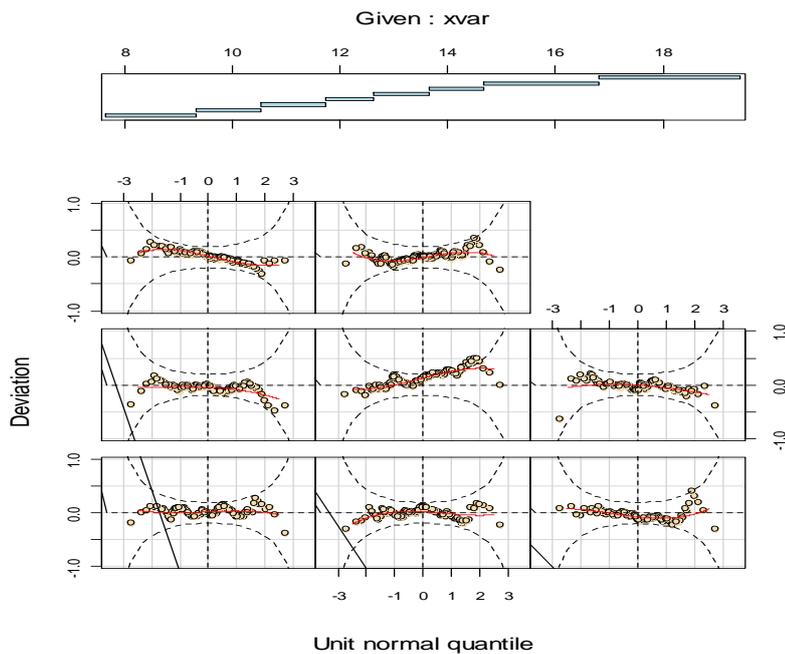


Fig. 2: Resultados de la función wp()

- La función `Q.stats()` resume los estadísticos Q de Royston y Wright que permiten contrastar las hipótesis de adecuación de cada uno de los parámetros del modelo [14].

Intervalos de edad	Z1	Z2	Z3	Z4	AgostinoK2
7.655 to 9.335	0.193284906	0.0932808	0.006739252	0.21388590	0.0457926
9.335 to 10.535	-0.058518787	0.0512599	0.637695943	0.53716553	0.6952029
10.535 to 11.745	-0.673115539	-0.5092409	0.904557540	0.64290264	12315482
11.745 to 12.625	-0.752581122	-0.5701007	0.497923241	0.05982185	0.2515062
12.625 to 13.645	1566353891	18950852	0.027774193	0.43627205	0.1911047
13.645 to 14.675	-0.280411633	-0.5096902	0.655588153	0.49652803	0.6763359
14.675 to 16.795	0.141753816	-15090976	0.290162521	0.69336568	0.5649503
16.795 to 19.425	-0.005770999	0.7502737	0.322802926	0.73335222	0.6420072
TOTAL Q stats	3.612423868	7.2870884	2.181648	2.20662128	4.298448
dj for Q stats	1.999079799	5.4997537	1.999477056	1.99939955	7.9988766
ppval for Q stats	0.164157432	0.2459516	0.718801955	0.69772824	0.8191613

Tabla I: Estadísticos Q de Royston y Wright

- Finalmente la curva de percentiles suavizados se obtuvo a través de la función `centiles()` [14].

Percentiles suavizados por BCPE(4,2,2,2) edad^{1.7}. Varones

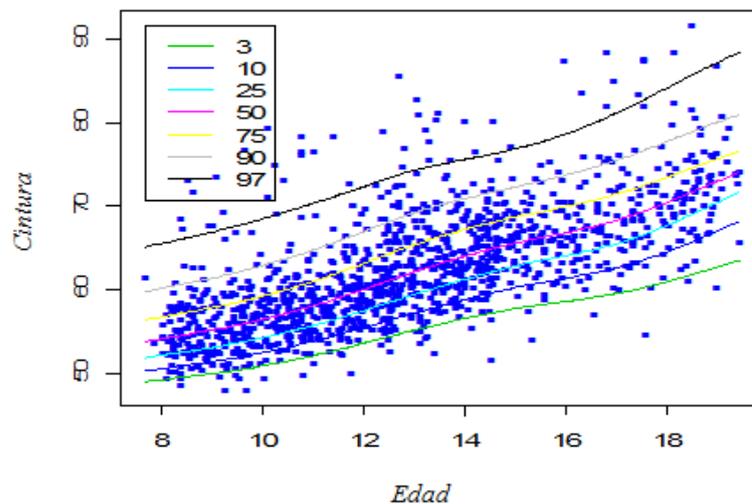


Fig. 3: Gráfico de percentiles suavizados

- La tabla de percentiles a partir de valores específicos de la edad se obtuvo con la función `centiles.pred()` [14].

Tabla II: Tabla de percentiles

ix	C3	C10	C25	C50	C75	C90	C97
8	49,271728	50,63198	52,195275	54,204164	56,80018	60,209284	65,561975
9	50,005214	51,466265	53,12784	55,22853	57,917713	61,443836	66,940049
10	50,956383	52,532159	54,305699	56,512749	59,307276	62,949633	68,548488
11	52,182362	53,883555	55,785746	58,131172	61,061816	64,806232	70,380483
12	53,655677	55,50715	57,568207	60,096534	63,197457	67,014598	72,391555
13	55,216402	57,255095	59,496637	62,202215	65,434628	69,25749	74,330495
14	56,622917	58,891618	61,301	64,075769	67,271108	70,95301	75,62462
15	57,811888	60,332531	62,890337	65,65724	68,756203	72,365923	76,932638
16	58,703895	61,452146	64,149828	66,925903	70,017386	73,778729	78,713695
17	59,666551	62,726076	65,620511	68,427359	71,549505	75,600639	81,227875
18	61,05003	64,632024	67,788299	70,528604	73,53651	77,821634	84,225102
19	62,710742	67,035143	70,447573	72,95561	75,646254	80,00552	87,129722

Los valores pronósticos de cada parámetro se obtienen a partir de la función `predict()` [14].

Ejemplo para el caso del parámetro μ

```
> (pnu <- predict(modcv_4222t, what="nu", newdata=newcintv, data=ncint98v))
```

```
[1] -5.1460402 -4.9700412 -4.7855612 -4.5947244 -4.4021609 -4.1947473
[7] -3.9538925 -3.6790944 -3.3854421 -3.0764976 -2.7438545 -2.3898041
[13] -2.0382488 -1.7376386 -1.5423063 -1.4605768 -1.4593658 -1.4742955
[19] -1.4616981 -1.3987127 -1.2896430 -1.1438239 -0.9742183
```

3. CONCLUSIONES

El SPSS ha constituido por muchos años la vía fundamental para el análisis de datos de crecimiento y desarrollo. Sin embargo, no ha sido posible con este software darle solución al ajuste de curvas de crecimiento de forma automatizada. Por otra parte, la OMS recomienda el uso del paquete GAMLSS desarrollado en R. En este trabajo se corroboró la efectividad de dicho programa en la elaboración de las curvas de crecimiento de una dimensión antropométrica en particular: la circunferencia de cintura. Paralelamente se muestra la factibilidad del uso de R como una alternativa de software libre para el análisis estadístico de datos de investigaciones. La desventaja que se le señala a R con respecto a la interfaz lo es cada vez menos, ya que en su continuo perfeccionamiento se han confeccionado varias interfaces gráficas que mejoran la interactividad del usuario con las funciones del paquete, como por ejemplo el R COMMANDER.

En el ámbito de la Universidad de Ciencias Médicas, sería muy conveniente

trabajar en la inserción de R para sustituir el uso de software propietario tanto en la investigación como en la docencia.

4. REFERENCIAS BIBLIOGRAFICAS

1. Jordán J et al. The 1972 Cuban National Child Growth Study as an example of population health monitory: design and methods. *Annals of Human Biology*. 1975; 2(2).
2. Jordán J y colaboradores. *Desarrollo Humano en Cuba*. La Habana: Editorial científico técnica; 1979.
3. Gutiérrez JA y colaboradores. The 1982 Cuban national growth and development study: Secular changes in weight and height. *Acta Médica Auxológica*. 1993; 25: 115-127.
4. Esquivel M, Romero JM, Berdasco A, Gutiérrez JA, Jiménez JM, Posada E y Rubén M. Estado nutricional de preescolares de Ciudad de La Habana entre 1972 y 1993. *Rev. Panam. Salud Pública*. 1997; 1(5): 349-354.
5. Borghi E et al. Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. *Statist. Med*. 2006; 25: 247-265.
6. WHO Child Growth Standards. Methods and Development. Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age. [citado el 20 Ago. 2011]. Disponible en: http://www.who.int/childgrowth/publications/technical_report_pub/en/index.html
7. WHO Child Growth Standards. Methods and Development. Head circumference-for-age, arm circumference-for-age, triceps skinfold-for-age and subscapular skinfold-for-age. [citado el 20 Ago. 2011]. Disponible en: http://www.who.int/childgrowth/publications/technical_report_2/en/index.html
8. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. [citado el 20 Ago. 2011]. Disponible en: <http://www.R-project.org>
9. Paradis E. R para Principiantes. [citado el 20 Ago. 2011]. Disponible en: http://cran.r-project.org/doc/contrib/rdebuts_es.pdf
10. Carmona Francesc. *Curso básico de R*. 2007.
11. Short T. R Reference Card. [citado el 20 Ago. 2011]. Disponible en: <http://cran.r-project.org/doc/contrib/Short-refcard.pdf>
12. Owen WJ. *The R Guide*. [citado el 20 Ago. 2011]. Disponible en: <http://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf>
13. Stasinopoulos M, Rigby B y Akantziliotou C. *gamlss: Generalized Additive Models for Location Scale and Shape*. R package version 1.9-9. [citado el 20 Ago. 2011]. Disponible en: <http://www.gamlss.com>

14. Stasinopoulos M, Rigby B y Akantziliotou C. Instructions on how to use the gamlss package in R. Second edition. 2008. [citado el 20 Ago. 2011]. Disponible en: <http://www.gamlss.com>