



Universidade de Vigo

Trabajo Fin de Máster

---

# Curvas de crecimiento de la población infantil gallega

---

María Arias Vázquez

Máster en Técnicas Estadísticas

Curso 2018-2019

## Propuesta de Trabajo Fin de Máster

<p><b>Título en galego:</b> Curvas de crecemento da poboación infantil galega</p>
<p><b>Título en español:</b> Curvas de crecimiento de la población infantil gallega</p>
<p><b>English title:</b> Growth curves of the Galician child population</p>
<p><b>Modalidad:</b> Modalidad B</p>
<p><b>Autor/a:</b> María Arias Vázquez, Universidad de Santiago de compostela</p>
<p><b>Director/a:</b> María del Carmen Iglesias Pérez, Universidad de Vigo; ,</p>
<p><b>Tutor/a:</b> M<sup>Â</sup>l Isolina Santiago Pérez, Departameto de epidemoloxía da DXSP; ,</p>
<p><b>Breve resumen del trabajo:</b></p> <p>No ano 2013 a Dirección Xeral de Saúde Pública realizou un estudo transversal de obesidade infantil entre os escolares galegos de 6 a 15 anos, onde se recolleron medidas antropométricas de 7.438 nenos e nenas de 6 a 15 anos. O obxectivo deste TFM é modelizar, con estes datos, curvas percentís de crecemento suavizadas empregando métodos paramétricos e non paramétricos para cada unha destas medidas antropométricas, así como do índice de masa corporal (IMC). Tamén se pretende elaborar táboas de percentís da poboación infantil galega por idade e sexo e comparar os valores obtidos cos estándares nacionais e internacionais.</p>

# Agradecimientos

Agradecer la entrega y las ganas de mis tutoras, Soly y Maca. Sin ellas no habría sido posible hacer un trabajo tan pulido y perfeccionado. Me han motivado mucho para seguir aprendiendo y trabajando. Muchas gracias.

# Índice general

<b>Resumen</b>	<b>v</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Encuesta de obesidad infantil en Galicia 2013</b>	<b>4</b>
2.1. Metodología de la encuesta . . . . .	4
2.2. Análisis exploratorio de los datos . . . . .	6
<b>3. Marco teórico</b>	<b>12</b>
3.1. Preliminares . . . . .	12
3.1.1. Modelos GAMLSS . . . . .	13
3.1.2. Algoritmos . . . . .	16
3.1.3. Técnicas de suavización . . . . .	21
3.1.4. Splines polinómicos o de regresión . . . . .	22
3.1.5. Splines de suavizado . . . . .	23
3.1.6. Splines penalizados . . . . .	23
3.2. Estimación percentil mediante los métodos LMS y BCPE . . . . .	25
3.2.1. Método LMS . . . . .	26
3.2.2. Método BCPE (LMSP) . . . . .	27
3.2.3. Estimación de las curvas $\mu$ , $\sigma$ , $\nu$ y $\tau$ para los modelos BCPE . . . . .	28
<b>4. Software utilizado</b>	<b>30</b>
4.1. Selección de la distribución adecuada . . . . .	30
4.2. Estimación de los grados de libertad . . . . .	31
4.3. Ajuste y diagnosis de los modelos . . . . .	33
4.4. Curvas percentiles . . . . .	37
<b>5. Estimación de los modelos</b>	<b>38</b>
5.1. IMC en niños . . . . .	38
5.2. IMC en niñas . . . . .	49
5.3. Peso en niños . . . . .	61
5.4. Peso en niñas . . . . .	63

5.5. Talla en niños . . . . .	66
5.6. Talla en niñas . . . . .	71
5.7. Perímetro de cintura en niños . . . . .	75
5.8. Perímetro de cintura en niñas . . . . .	77
5.9. Perímetro de cadera en niños . . . . .	80
5.10. Perímetro de cadera en niñas . . . . .	83
<b>6. Curvas y tablas de percentiles</b>	<b>86</b>
6.1. IMC en niños . . . . .	87
6.2. IMC en niñas . . . . .	89
6.3. Peso en niños . . . . .	91
6.4. Peso en niñas . . . . .	93
6.5. Talla en niños . . . . .	95
6.6. Talla en niñas . . . . .	97
6.7. Perímetro de cintura en niños . . . . .	99
6.8. Perímetro de cintura en niñas . . . . .	101
6.9. Perímetro de cadera en niños . . . . .	103
6.10. Perímetro de cadera en niñas . . . . .	105
<b>7. Comparación entre niños y niñas</b>	<b>107</b>
7.1. IMC . . . . .	108
7.2. Peso . . . . .	109
7.3. Talla . . . . .	110
7.4. Perímetro de cintura . . . . .	111
7.5. Perímetro de cadera . . . . .	112
7.6. Comentarios sobre las gráficas . . . . .	113
<b>8. Comparación con otros estudios</b>	<b>114</b>
8.1. Orbegozo en niños . . . . .	116
8.2. Orbegozo en niñas . . . . .	117
8.3. Cole en niños . . . . .	118
8.4. Cole en niñas . . . . .	119
8.5. OMS en niños . . . . .	120
8.6. OMS en niñas . . . . .	121
8.7. Comentarios sobre las gráficas . . . . .	122
<b>9. Conclusiones y limitaciones</b>	<b>123</b>
9.1. Conclusiones . . . . .	123
9.2. Limitaciones . . . . .	124
<b>Bibliografía</b>	<b>125</b>

# Resumen

## Resumen en español

El objetivo de este trabajo fue la construcción de curvas percentiles de crecimiento de la población infantil gallega. Estas curvas son útiles tanto en el ámbito clínico, para evaluar el desarrollo y crecimiento de los niños, como en el ámbito epidemiológico, donde los percentiles se utilizan como valores de referencia para la clasificación de individuos.

En el año 2013, la Dirección Xeral de Saúde Pública realizó un estudio antropométrico de la población escolar gallega de 6 a 15 años, con mediciones de 7.438 alumnos representativos de esa población. Dichos datos se utilizaron en este trabajo para estimar las curvas percentiles de peso, talla, índice de masa corporal (IMC), perímetro de cintura y perímetro de cadera, en función de la edad, y para niños y niñas por separado.

Este trabajo se centra en los modelos aditivos generalizados de localización, escala y forma (GAMLSS), que generalizan los métodos clásicos de estimación de curvas percentiles: el método LMS de Cole y Green (1982), que asume una distribución Box-Cox normal con 3 parámetros, y el método BCPE o LMSP de Rigby y Stasinopoulos (2004), que asume una distribución Box-Cox Power Exponential con 4 parámetros. Se utilizaron splines cúbicos y penalizados para modelizar la dependencia de cada parámetro respecto a la edad y conseguir curvas centiles suaves. La estimación y diagnosis de los modelos se realizó con la librería `gamlss` de R.

## English abstract

The objective of this work was the construction of growth percentiles of the Galician child population. These curves are useful both in the clinical setting and in the epidemiological field, where percentiles are used as reference values for the classification of individuals.

In 2013, the General Directorate of Public Education carried out an anthropometric study of the Galician school population from 6 to 15 years old, with measurements of 7,438 representative students of that population. These data are used in this work to estimate the percentile curves of weight, height, body mass index (BMI), waist circumference and hip perimeter, based on age, and boys and girls separately.

This work focuses on the generalized additive models of location, scaling and shape (GAMLSS), which generalize the estimation methods of percentiles: the LMS method of Cole and Green (1982),

which assumes a normal Box-Cox distribution with 3 parameters, and the BCPE or LMSP method of Rigby and Stasinopoulos (2004), which assumes a Box-Cox Power Exponential distribution with 4 parameters. Cubic and penalized splines were used to model the dependence of each of them. The estimation and diagnosis of the models were carried out with the `gamlss` library of R.

# Capítulo 1

## Introducción

Las curvas percentiles de crecimiento infantil son útiles en el ámbito clínico para evaluar el desarrollo y crecimiento de los niños. En este ámbito se usan como una herramienta de cribado, ya que permiten identificar sujetos con mediciones inusuales, que corresponden a valores extremos de la distribución, y que pueden ser indicativos de un problema de salud o de nutrición. Por otra parte, los percentiles de las curvas también se utilizan como valores de referencia para la clasificación de sujetos en estudios epidemiológicos, por ejemplo, para definir obesidad y estimar sus prevalencias.

Las medidas antropométricas que se utilizan habitualmente para valorar el crecimiento infantil son el peso, la talla y el índice de masa corporal (IMC), que se calcula como  $\text{Peso (kg)} / \text{Talla}^2 \text{ (m)}$ . Algunos estudios incluyen, además, el perímetro de cintura y el perímetro craneal. La talla a la edad adulta es un indicador del estado de bienestar de una población, ya que tiene relación directa con el desarrollo socioeconómico. El IMC es un indicador del estado ponderal, y se utiliza para establecer categorías de bajo peso, peso normal, sobrepeso y obesidad. El perímetro de cintura es un indicador de obesidad central que se asocia con riesgo metabólico.

La distribución de estas medidas va cambiando de forma continua con la edad, por lo que el cálculo de percentiles debe hacerse mediante la estimación de curvas suavizadas a lo largo de la edad.

Cole propuso en 1990 [6] el método LMS basado en asumir que la variable de las mediciones es asimétrica pero sigue una distribución normal tras una transformación Box–Cox. Las mediciones antropométricas, en particular el peso y la talla, suelen seguir ese patrón. El principal problema del método es la posible presencia de curtosis, que no se corrige con la transformación.

En el método de Cole, la distribución de las mediciones, condicionada a los valores de la edad, se puede resumir mediante tres parámetros que dan nombre al método: la potencia de la transformación (L), la media (M), y el coeficiente de variación (S). Para estimar las curvas L, M y S de forma suave en función de la edad Cole y Green propusieron en 1992 [2] aplicar el método de máxima verosimilitud penalizada, y con este método desarrollaron valores de referencia internacionales para la clasificación del sobrepeso y la obesidad (Cole, 2000 [4]) a partir del índice de masa corporal (IMC) en niños de 2 a 18 años; posteriormente los ampliaron al bajo peso (Cole, 2007 [5]). Para ambos trabajos utilizaron datos de 6 estudios transversales de crecimiento con representatividad a nivel nacional en Brasil, Gran Bretaña, Hong-Kong, Países Bajos, Singapur y Estados Unidos ( $n = 192.727$  niños de 0 a 25 años).



Los valores de referencia se obtuvieron a partir de las curvas L, M y S de cada uno de los 6 estudios, calculando primero los percentiles que a los 18 años igualaban los puntos de corte de adultos (IMC de 16, 17, 18,5, 25 y 30 para delgadez de grado 1, 2, 3, sobrepeso y obesidad, respectivamente) y ponderando después los valores obtenidos con las 6 encuestas. En el año 2012 (Cole, 2012 [3]) reformularon los valores de referencia para que pudiesen expresarse en términos de percentiles; para ello, ponderaron primero las 6 curvas L, M y S, y a partir de las curvas resultantes obtuvieron los percentiles correspondientes a los puntos de corte de adultos.

En el año 2007 la OMS [7] publicó las curvas de referencia del crecimiento de niños y niñas entre 5 y 19 años, en edad escolar y adolescentes. Previamente, había desarrollado las curvas estándar de crecimiento desde el nacimiento hasta los 5 años, basándose en un estudio multicéntrico. En el caso de las curvas a partir de 5 años, la OMS reconstruyó las curvas con los datos utilizados para las curvas NCHS de 1977 (unión de 3 bases de datos,  $n = 22917$ ). Además, añadió a la base de datos la información de los niños de 18 a 71 meses del estudio de menores de 5 años, para conseguir una transición suave entre las curvas de ambos estudios, y estimó las nuevas curvas entre los 5 y los 19 años con una muestra de 18 meses a 24 años. Los datos se depuraron para evitar la influencia de pesos muy elevados con respecto a la talla. Se ajustaron modelos BCPE con splines cúbicos. En todos los casos, los modelos se reducen al método LMS, puesto que el valor del parámetro que ajusta la curtosis es 2.

En España, el estudio más importante de crecimiento infantil se inició en 1976 financiado por la Fundación Orbegozo. En el año 1988 [11] se publicaron las curvas y tablas de crecimiento de 0 a 18 años basadas en datos de 3 cohortes (0, 3 y 9 años) que se seleccionaron aleatoriamente entre la población de Bilbao entre 1978 y 1980 y a los que se siguió hasta 1988. Este fue el estudio longitudinal mixto (Hernández, 1988). El estudio longitudinal puro fue el resultado de seguir a la cohorte de 0 años hasta completar su crecimiento (1978-2000), realizando mediciones cada 6 meses a cada niño. En el año 2000, al terminar el estudio longitudinal, se realizó un estudio transversal de validación, y los resultados conjuntos de ambos estudios se publicaron en 2004 [20]. Las curvas de percentiles se calcularon a partir de los z scores ( $z=x-m/s$ ) asumiendo normalidad de las mediciones, y fueron suavizadas en el caso del estudio longitudinal. Dicho estudio permitía realizar el seguimiento del crecimiento de un niño, y el estudio transversal era válido para estudios epidemiológicos. Con los datos del estudio transversal, compuesto por 6.433 niños y niñas de 0 a 18 años representativos de la población de Vizcaya, se publicaron unas nuevas curvas y tablas en el año 2011 [12] resultantes de aplicar el método LMS. Estas curvas son las que se utilizan para monitorizar la evolución del IMC en los controles pediátricos de salud, y se incluyen en las cartillas de salud infantil.

El Estudio Transversal Español de crecimiento (ETEC) [14] combinó los datos de 32.064 sujetos de 0 a 24 años procedentes de 4 estudios transversales realizados entre los años 2000 y 2004 en Andalucía, Barcelona, Bilbao y Zaragoza. Con estos datos estimaron las curvas percentiles de peso, talla e IMC utilizando el método LMS de Cole, y compararon los percentiles obtenidos con los de otros estudios nacionales e internacionales realizados 20 años atrás (Carrascosa 2008). Carrascosa et al concluyeron que en esas dos décadas se produjo en España una aceleración secular del crecimiento, y el incremento fue mucho mayor en el peso que en la talla, sobre todo en los sujetos con talla superior al percentil 50. Esto implica que es necesario actualizar periódicamente los patrones utilizados como referencia para

valorar el crecimiento durante la infancia y adolescencia, y los valores para definir el sobrepeso y la obesidad.

En Galicia hay muy pocos estudios antropométricos en población infantil realizados fuera del ámbito clínico, y se restringen a grupos de edad específicos (10 a 12 años), a una sola provincia (Pontevedra) o no son recientes. Por este motivo, la DXSP decidió realizar en el año 2013 [15] un estudio epidemiológico que permitiese caracterizar de forma detallada, por sexo y edad, el estado ponderal (bajo peso, peso normal, sobrepeso y obesidad) de la población escolar de 6 a 15 años. El objetivo de este trabajo es estimar las curvas de crecimiento de peso, talla, IMC, perímetro de cintura y perímetro de cadera a partir de los datos recogidos en ese estudio.

## Capítulo 2

# Encuesta de obesidad infantil en Galicia 2013

### 2.1. Metodología de la encuesta

En 2013 la Dirección Xeral de Saúde Pública (DXSP) realizó un estudio transversal, con el objetivo de conocer la prevalencia poblacional de bajo peso, peso normal, sobrepeso y obesidad, además de la prevalencia de obesidad abdominal o central en escolares de Galicia de 6 a 15 años [15]. A continuación, se describe el diseño del estudio.

- *Ámbito poblacional*

La población objeto del estudio está constituida por dos grupos de escolares: los de Enseñanza Primaria (EP), que tienen entre 6 y 11 años, y los de Enseñanza Secundaria Obligatoria (ESO), que tienen entre 12 y 15 años. El tamaño de los dos grupos se estimó en unos 133.000 y 87.500 alumnos, respectivamente, según los datos de la Consellería de Cultura, Educación e Ordenación Universitaria correspondientes al curso 2012–2013.

- *Ámbito geográfico y temporal*

El ámbito geográfico del estudio abarcó todo el territorio Gallego y la recogida de datos se realizó durante los meses de octubre y noviembre de 2013, es decir, en el primer trimestre del curso 2013-2014.

- *Unidades de la encuesta*

En este estudio las unidades de análisis, o casos individuales a encuestar, fueron los alumnos matriculados en centros de EP de Galicia en el curso 2013-2014, y los alumnos matriculados en centros de ESO de Galicia en el mismo curso. Las unidades de muestreo fueron los centros de EP y ESO y las aulas de los cursos correspondientes, 1º a 6º de EP y 1º a 4º de ESO.

- *Tipo de muestreo*

Para seleccionar la muestra se hizo, en primer lugar, una clasificación de los centros escolares de Galicia en tres grupos: los que solo imparten EP, los que solo imparten ESO y los que imparten ambos niveles. En cada grupo se realizó un muestreo bietápico estratificado, seleccionando centros educativos en la primera etapa, y aulas en la segunda, e incluyendo en el estudio todos los alumnos de las aulas seleccionadas. El criterio de estratificación de los centros fue el grado de urbanización del municipio al que pertenecen, según una clasificación realizada por el Instituto Galego de Estatística.

El marco final del muestreo se componía de: 812 centros que imparten EP, con 6.692 aulas de 1º a 6º y 131.648 alumnos y 478 centros que imparten ESO con 4.269 aulas de 1º a 4º con 87.307 alumnos. Entre estos centros, hay 238 que imparten los dos niveles, en los que estudian el 35 % de los alumnos de EP y el 37 % de los de ESO.

- *Tamaño de muestra*

El tamaño de la muestra de escolares se determinó, asumiendo muestreo aleatorio simple, para cada curso por separado. En todos los casos se consideró un tamaño medio de 22.000 alumnos por curso, un nivel de confianza del 95 %, una prevalencia esperada de obesidad del 10 %, una precisión del 2.8 % y un efecto de diseño <sup>1</sup> de 1.5. Los tamaños de muestra obtenidos se aumentaron para compensar las pérdidas esperadas por negativas a participar en el estudio, que se estimaron en un 10 % y 15 % para EP y ESO, respectivamente. Finalmente, se establecieron los tamaños de la muestra en 4.500 alumnos de EP y 3.200 de ESO, es decir, un total de 7.700. Los tamaños de muestra resultantes se repartieron proporcionalmente por estrato, considerando por separado los tres grupos de centros según el nivel educativo que imparten (solo EP, solo ESO, EP y ESO). A partir del tamaño medio de las aulas de cada estrato, se calculó el número de aulas a seleccionar en ese estrato, y el número de centros se determinó considerando una muestra de 3 aulas por centro en EP y 2 aulas por centro en ESO. Se seleccionaron 137 centros y 417 aulas, de las que finalmente participaron 416.

- *Recogida de la información*

La recogida de información fue realizada por 9 equipos, de dos miembros cada uno, previamente formados y entrenados. A los centros seleccionados se les envió una carta en la que se explicaban las características del estudio y se solicitaba la colaboración del centro. Posteriormente, se contactó por teléfono con ellos para fijar el día que los equipos irían al centro para realizar las mediciones. Los colegios solicitaron a los padres o tutores de los niños el consentimiento para participar en el estudio.

La información recogida, que se registraba de forma anónima en un dispositivo electrónico, comprendía las mediciones antropométricas y un breve cuestionario de 7 preguntas entre las que se

---

<sup>1</sup> Se define como el cociente de la varianza estimada bajo el diseño utilizado, entre la varianza obtenida considerando un muestreo aleatorio simple para un mismo tamaño de muestra

incluía el sexo y la fecha de nacimiento. También se recogieron estas dos variables de los niños que estaban ausentes ese día o que rechazaron participar. Las mediciones se realizaron de la siguiente forma [artículo ¿cuántos niños hay con exceso de peso en Galicia]: el peso se midió en ropa ligera y sin zapatos usando una báscula Seca homologada y aproximando a 0,1 kilogramos. La estatura se midió en el plano de Frankfurt, con los niños descalzos, con un tallímetro homologado Tanita portable de base fija aproximando a 0,1 centímetros. El perímetro de cintura se midió en un punto equidistante entre la última costilla y la cresta ilíaca con una cinta no elástica, situando al niño de pie con los pies juntos, el abdomen relajado y los brazos colgando junto a los costados, y aproximando la medición a 0.1 centímetros.

- *Tratamiento de los datos*

Como paso previo al análisis estadístico, se realizó un tratamiento de la base de datos que consistió en la depuración y corrección de errores. En cuanto a la edad, se calculó en meses a partir de la fecha de nacimiento y la fecha de la entrevista.

## 2.2. Análisis exploratorio de los datos

En el estudio se obtuvo información de 8.140 escolares con edades entre 6 y 15 años. Se dispone de medidas antropométricas de 7.438 alumnos, lo que supone una tasa de participación del 91.4%. En la Tabla 2.1 se detallan los tamaños de la muestra por edad y sexo. Los resultados detallados del estudio están publicados en el artículo [15] de la Dirección Xeral de Saúde Pública. Estos resultados, abarcan un análisis de las medidas antropométricas a través de indicadores como el IMC, el perímetro de cintura, la baja estatura o el peso. En el artículo se utilizaron estos indicadores y distintos puntos de corte definidos en el, para obtener tablas y gráficos de prevalencias de malnutrición, bajo peso, peso normal, sobrepeso, obesidad y obesidad abdominal en función del sexo, la edad y del ámbito de residencia de los escolares.

Edad	Niños	Niñas	Todos
6	483	426	909
7	366	315	681
8	284	308	592
9	331	315	646
10	401	416	817
11	412	377	789
12	406	384	790
13	393	358	751
14	401	367	768
15	366	329	695
Todas	3843	3595	7438

Tabla 2.1: Tabla de frecuencias de la muestra por edad y sexo

Como paso previo a la estimación de percentiles, se analizó la distribución, para niños y niñas, de cada una de las variables dependientes (IMC, peso, talla, cintura y cadera) en función de la edad en años mediante un box plot. En los gráficos se representan como puntos los valores, llamados valores atípicos, con  $y > p75 + 1.5R$  y  $y < p25 - 1.5R$ , siendo  $R = p75 - p25$  el recorrido intercuartílico y  $p25$  y  $p75$  los percentiles de orden 25 y 75, respectivamente. Entre ellos destacan algunos que parecen “extremadamente” altos (o bajos) en relación al resto. Para identificarlos se usa un criterio más restrictivo:  $y > p75 + 3R$  (o  $y < p25 - 3R$ ). Estos puntos se consideran de aquí en adelante como valores atípicos extremos.

Con este criterio se identificaron los valores atípicos extremos para cada una de las variables de este trabajo. En las Tablas 2.2 y 2.3 se muestran estos valores junto a los percentiles  $p25$  y  $p75$  empíricos por edad de cada variable para niños y niñas, respectivamente.

Los valores atípicos extremos pueden generar problemas a la hora de ajustar los modelos que estiman las curvas percentiles, es por ello que, en este trabajo se eliminaron de la muestra los que se identificaron para cada variable en función de la edad y del sexo.

	ncuest	edad (meses)	edad (años)	<b>IMC</b>	p25	p75
1	3214102	80	6	27.13	15.31	17.59
2	3609106	81	6	27.89	15.31	17.59
3	1525201	116	9	33.22	16.35	20.41
4	1517318	116	9	34.8	16.35	20.41
5	3210311	125	10	36.8	16.74	21.51
6	2717414	157	13	46.89	18.27	22.71
7	1504520	186	15	44.32	19.26	23.72
8	1539416	189	15	40.07	19.26	23.72
9	3643404	182	15	40.3	19.26	23.72
	ncuest	edad (meses)	edad (años)	<b>peso</b>	p25	p75
1	3616211	77	6	43.5	21.3	26.55
2	3609106	81	6	47.5	21.3	26.55
3	1503109	88	7	57.4	24.2	31.35
4	1517318	116	9	78.3	28.85	39.9
5	3210311	125	10	90.7	32.9	44.2
6	1522304	138	11	98	37.05	49.9
7	2717414	157	13	118.1	45.3	61.4
8	2714414	171	14	113.5	52	67.1
9	1504520	186	15	128.7	56	71.65
10	3643404	182	15	123	56	71.65
	ncuest	edad (meses)	edad (años)	<b>cintura</b>	p25	p75
1	3609106	81	6	89.3	51.95	58.1
2	2717414	157	13	122.3	65.2	78
3	1504520	186	15	124.3	69	79.95
4	1539416	189	15	118.7	69	79.95
5	1537407	81	15	116.4	69	79.95
6	3643404	182	15	114	69	79.95
	ncuest	edad (meses)	edad (años)	<b>cadera</b>	p25	p75
1	3609106	81	6	91.1	61.15	68.05
2	1517318	116	9	112.4	69.75	79.95
3	2717414	157	13	136.8	81.6	92.9
4	1504520	186	15	134.5	88.03	97.95
5	3643404	182	15	130.9	88.03	97.95

Tabla 2.2: Valores atípicos extremos de las variables en los niños.

En los niños se identificaron 9 valores atípicos extremos en el IMC, 10 en el peso, 6 en el perímetro de cintura y 5 en el perímetro de cadera, y todos ellos son superiores a  $p75 + 3R$  (atípicos por arriba). Estos valores atípicos extremos suponen el 0.2%, 0.26%, 0.15% y 0.13% de la muestra de los niños,

respectivamente, es decir, un porcentaje muy pequeño de la muestra. No se encontró ningún valor atípico extremo de la talla.

	ncuest	edad (meses)	edad (años)	<b>IMC</b>	p25	p75
1	3601122	75	6	26.03	15.15	17.73
2	2717405	151	12	38.9	17.97	22.95
	ncuest	edad (meses)	edad (años)	<b>peso</b>	p25	p75
1	3640106	81	6	42.8	20.73	25.8
2	3602210	71	6	41.5	20.73	25.8
3	2717405	151	12	94.8	41.60	54.75
4	2722513	175	14	100.5	49.25	62
5	1524510	186	15	105	51.20	61.9
	ncuest	edad (meses)	edad (años)	<b>talla</b>	p25	p75
1	3610501	133	15	189	157.6	165.1
	ncuest	edad (meses)	edad (años)	<b>cintura</b>	p25	p75
1	1509105	77	6	79.1	51.7	58.48
2	2717405	151	12	113.4	62	74.7
	ncuest	edad (meses)	edad (años)	<b>cadera</b>	p25	p75
1	2717405	151	12	124.7	80.70	91.7
2	2722513	175	14	128.4	87.70	97.75
3	1524510	186	15	128	89.6	98.3

Tabla 2.3: Valores atípicos extremos de las variables en las niñas.

En las niñas se identificaron 2 valores atípicos extremos en el IMC, 5 en el peso, 1 en la talla, 2 en el perímetro de cintura y 3 en el perímetro de cadera. En este caso, suponen, respectivamente, el 0.05 %, 0.14 %, 0.03 %, 0.05 % y 0.08 % de la muestra de las niñas. Todos ellos son atípicos extremos por arriba, excepto el de la talla que es inferior a  $p25 - 3R$  (atípico extremo por debajo).

En las Figuras 2.1 y 2.2, se representan los box-plot de cada variable con los atípicos extremos correspondientes.



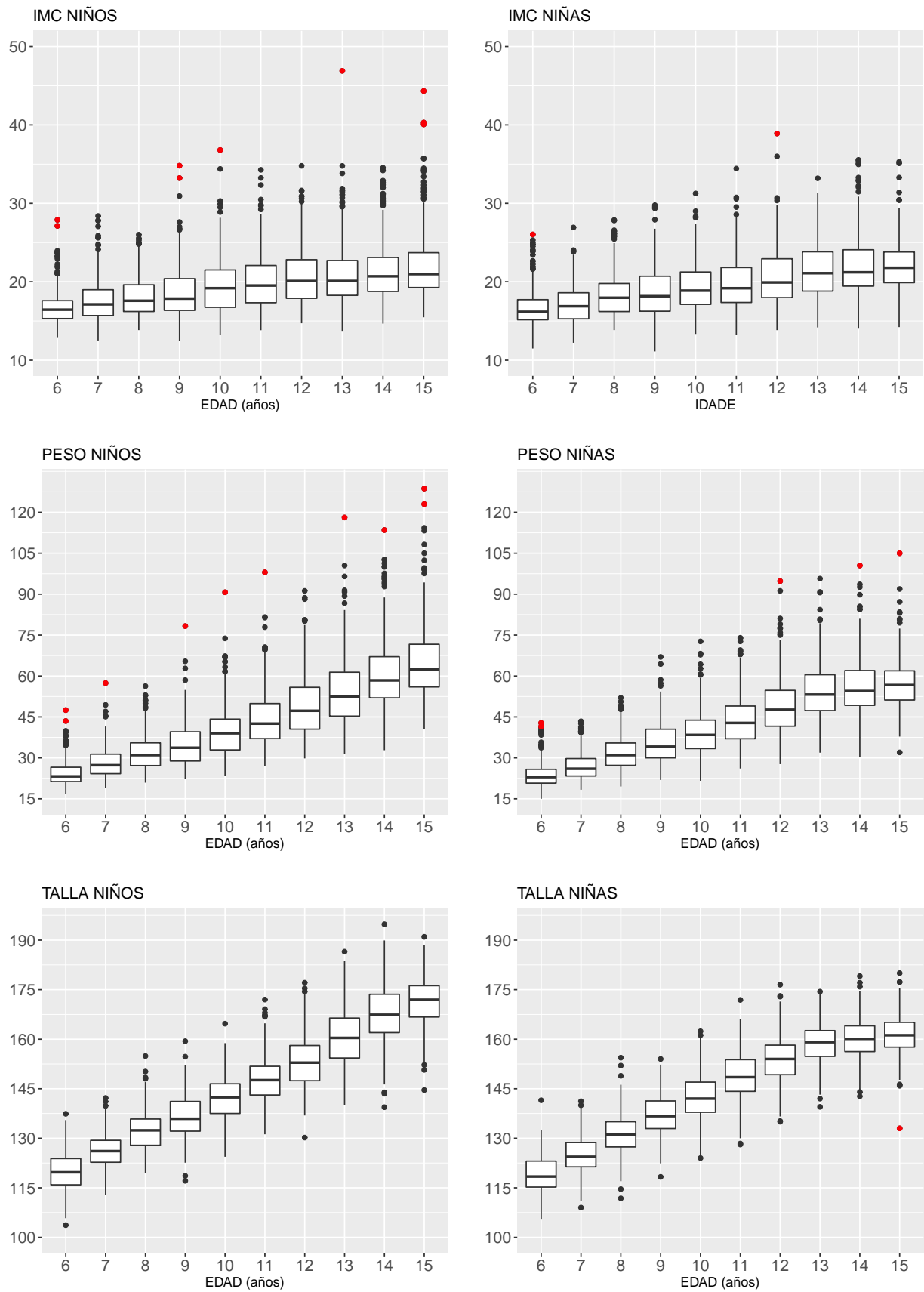


Figura 2.1: Gráficos box plot con los atípicos extremos, en rojo, de IMC ( $\text{kg}/\text{cm}^2$ ), peso (kg) y talla (cm) de niños y niñas.

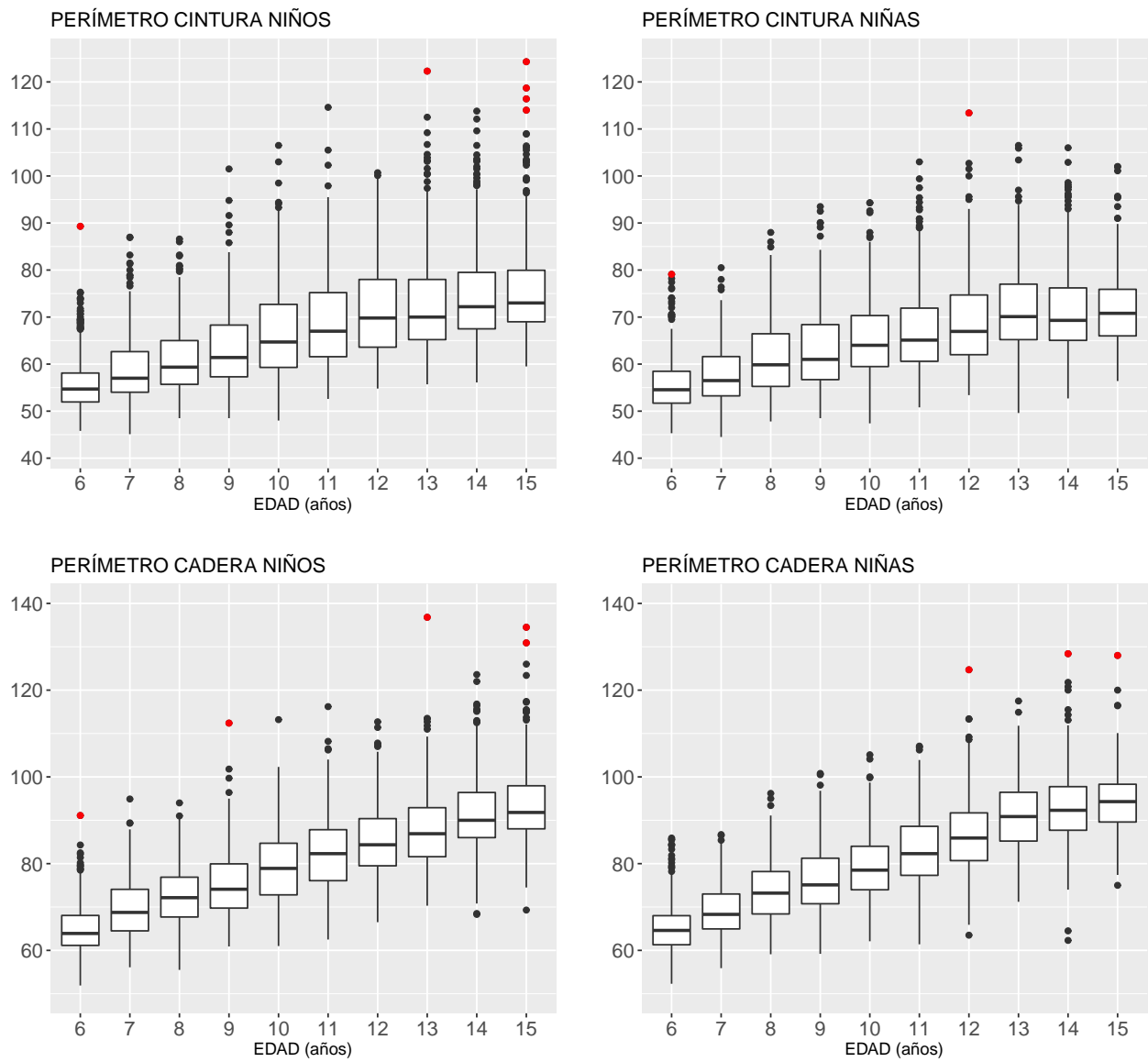


Figura 2.2: Gráficos box plot con los atípicos extremos, en rojo, del perímetro de cintura (cm) y del perímetro de cadera (cm) de niños y niñas.

## Capítulo 3

# Marco teórico

### 3.1. Preliminares

Para la construcción de curvas percentiles de crecimiento existen dos metodologías diferenciadas: (i) el método no paramétrico de estimación de la regresión cuantil (Kroenker, 1978 [13], 2005[?]), y (ii) el método (semi)paramétrico LMS de Cole y Green (1992) [2] y sus extensiones, entre las que destaca el método BCPE utilizado por la OMS [2].

El modelo de regresión cuantil no asume una distribución para la variable respuesta, por lo tanto, es flexible y reduce el sesgo causado al asumir una distribución particular (que puede ser errónea). Pero esta reducción del sesgo suele venir acompañada de un aumento en la variabilidad de las curvas centiles, sobre todo cerca de los extremos, y en general las curvas son menos agradables estéticamente que las producidas por métodos paramétricos. Además, un posible problema de la regresión cuantil es que diferentes curvas de cuantiles para diferentes valores de probabilidad pueden cruzarse, algo que es imposible con el enfoque paramétrico. Finalmente, el modelo de regresión cuantil carece de una fórmula explícita que permita el cálculo del cuantil de la variable respuesta  $y_p(x)$  dados valores conocidos de la probabilidad  $p$  y de la edad  $x$ . Este fue uno de los requisitos establecidos por un comité de expertos de la Organización Mundial de la Salud para la adopción de un método para la construcción de las curvas estándar mundiales.

En el presente trabajo se aplicó la segunda metodología de estimación debido a los argumentos esgrimidos en el párrafo anterior y el interés de la DXSP de utilizar las metodologías de referencia en el ámbito de la salud, tanto para cumplir los requerimientos de la OMS, como para facilitar la comparación de las curvas y valores percentiles obtenidos con los de otros estudios previos realizados con similar metodología.

El método LMS de Cole y Green (1992) y el método BCPE de la OMS son dos métodos semiparamétricos. Son paramétricos porque asumen una distribución paramétrica conocida (específica de cada método) para la variable de interés  $Y$  (peso, IMC,...), y son no paramétricos en el sentido de que cada parámetro (desconocido) de la distribución se modeliza como una función suave de la edad. El método de suavización clásico es el de splines cúbicos de suavizado, aunque recientemente se están imponiendo los splines penalizados.

Los métodos LMS y BCPE pueden verse como casos particulares de los denominados modelos aditivos generalizados de localización, escala y forma (GAMLSS), introducidos por Rigby y Stasinopoulos (2005) [17], de modo que la construcción de las curvas percentiles de este trabajo se realizó dentro del marco de los modelos GAMLSS.

En esta sección se describen los modelos GAMLSS, prestando especial atención al modelo aditivo con suavización de tipo spline utilizado en la estimación de las curvas centiles de crecimiento del presente trabajo.

### 3.1.1. Modelos GAMLSS

Para este apartado de los preliminares, se han utilizado las referencias [17], [19].

Los modelos GAMLSS son una clase general de modelos estadísticos para una variable respuesta unidimensional que incluyen como casos particulares a los bien conocidos modelos lineales generalizados (GLM) y a los modelos aditivos generalizados (GAM). En los modelos GAMLSS se asume una distribución paramétrica para la variable respuesta  $Y$  y se permite modelizar, no solo el parámetro de localización (la media o mediana), sino también los parámetros de dispersión, de escala y de forma, como funciones de una o varias variables explicativas. Estas funciones pueden adoptar una forma paramétrica y/o no paramétrica aditiva (suave) de las variables explicativas, permitiéndose además la presencia de efectos aleatorios.

La distribución para la variable respuesta en el GAMLSS se puede seleccionar de una familia muy general de distribuciones, que incluye las distribuciones de la familia exponencial (Normal, Binomial, Poisson,...) así como distribuciones continuas y discretas altamente sesgadas o kurtóticas no pertenecientes a la familia exponencial. Dos ejemplos especialmente importantes de distribuciones continuas que se utilizaron en este trabajo son la distribución Box–Cox Normal (distribución con tres parámetros) y la distribución Box–Cox Power Exponential (con cuatro parámetros) que se describirán más adelante.

Matemáticamente, sea  $Y$  la variable respuesta de la que se tienen  $n$  observaciones independientes  $y_i$  con  $i = 1, 2, \dots, n$ , que siguen una distribución paramétrica determinada,  $F(y_i|\boldsymbol{\theta}^i)$ , condicionada a  $\boldsymbol{\theta}^{iT}$  un vector de parámetros, cada uno de los cuales puede ser función de variables explicativas. En la mayoría de los casos prácticos se tienen como mucho cuatro parámetros ( $p = 4$ ), normalmente  $\boldsymbol{\theta}^i = (\theta_{1i}, \theta_{2i}, \theta_{3i}, \theta_{4i}) = (\mu_i, \sigma_i, \nu_i, \tau_i)$ . Siendo  $\mu_i$  y  $\sigma_i$  con  $i = 1, 2, \dots, n$  parámetros de localización y escala (media o mediana y coeficiente de variación o varianza) respectivamente y, por otro lado,  $\nu_i$  y  $\tau_i$  con  $i = 1, 2, \dots, n$  son parámetros de forma (por ejemplo, parámetros de asimetría y curtosis).

Las variables explicativas se introducen en el modelo a través de los “predictores”  $\eta_k$ , que se relacionan con los parámetros mediante funciones de enlace o funciones link monótonas, denotadas por  $g_k(\cdot)$ . Estas funciones son conocidas y dependen de la distribución seleccionada para la variable respuesta.

A continuación, se define el modelo GAMLSS [19].

**Definición 3.1.** Sea  $\mathbf{y}^T = (y_1, \dots, y_n)$  el vector  $n$ -dimensional de la variable respuesta, donde  $y_i \sim F(y_i|\boldsymbol{\theta}^i)$ . Para  $k = 1, 2, \dots, p$ , sean  $g_k(\cdot)$  las funciones link conocidas que relacionan los parámetros de

la distribución con las variables explicativas de la siguiente forma:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Z}_{jk} \boldsymbol{\gamma}_{jk}, \quad (3.1)$$

donde  $\boldsymbol{\theta}_k$  y  $\boldsymbol{\eta}_k$  son vectores  $n$ -dimensionales,  $\boldsymbol{\theta}_k^T = (\theta_{1k}, \theta_{2k}, \dots, \theta_{nk})$ , donde  $\boldsymbol{\beta}_k = (\beta_{1k}, \dots, \beta_{J'_k k})$  es un vector  $J'_k$ -dimensional. El número de covariables continuas con efecto no lineal que se utilizan para explicar cada parámetro  $\boldsymbol{\theta}_k$  se representa por  $J_k$ , para  $k = 1, \dots, p$ ,  $\mathbf{X}_k$  es una matriz de diseño fijada de orden  $n \times J'_k$ ,  $\mathbf{Z}_{jk}$  es una matriz fijada de orden  $n \times q_{jk}$  y  $\boldsymbol{\gamma}_{jk}$  una variable aleatoria  $q_{jk}$ -dimensional. A (3.1) le llamaremos GAMLSS.

Esta definición refleja la posibilidad de modelizar cada parámetro como una combinación de términos paramétricos ( $\mathbf{X}_k \boldsymbol{\beta}_k$ ) y términos aditivos ( $\mathbf{Z}_{jk} \boldsymbol{\gamma}_{jk}$ ). Ambos términos están compuestos por una matriz de diseño fijada, que depende de las variables explicativas,  $\mathbf{X}_k$  para los términos paramétricos y  $\mathbf{Z}_{jk}$  para los términos aditivos, con  $k = 1, \dots, 4$  y  $j = 1, \dots, J_k$  y por los vectores de parámetros para cada término,  $\boldsymbol{\beta}_k$  y  $\boldsymbol{\gamma}_{jk}$ , respectivamente.

Los términos paramétricos y aditivos que componen los GAMLSS permiten una formulación general de muchos modelos o submodelos. En particular, los términos aditivos pueden modelizar los suavizados no paramétricos o los efectos aleatorios. A continuación, se describen algunos casos particulares de GAMLSS, que se corresponden con modelos clásicos de la regresión univariante y que permiten entender de donde provienen los GAMLSS [18].

- **Modelos lineales (LM):** son los modelos de regresión univariante más sencillos, cuyas características y, también, limitaciones principales son asumir normalidad para la variable respuesta y linealidad en la relación de las variables explicativas y la variable respuesta. Por lo tanto, para  $i = 1, 2, \dots, n$  se tiene que  $y_i \sim N(\mu_i, \sigma)$  y  $\mu_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}$ . Utilizando notación matricial, se tiene:

$$\mathbf{y} \sim N(\boldsymbol{\mu}, \sigma) \quad , \quad \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}, \quad (3.2)$$

donde  $\boldsymbol{\mu}$  es el vector de dimensión  $n$  formado por los  $\mu_i$ ,  $\mathbf{X}$  es la matriz de diseño de dimensión  $n \times p$  ( $p = r + 1$ ) formada por las variables explicativas y  $\boldsymbol{\beta}$  el vector de parámetros de dimensión  $p$ , que ha de estimarse.

Asumir normalidad en la variable respuesta es una limitación en muchas aplicaciones prácticas. Por ello y con el fin de atender aquellas situaciones en que la variable respuesta no siga una distribución normal, se introdujeron los modelos lineales generalizados (GLM).

- **Modelos lineales generalizados (GLM):** introducidos por Nelder y Wedderburn (1972), incluyen varios modelos de regresión conocidos, incorporando aquellos donde la distribución de la variable respuesta pertenece a la familia exponencial (normal, gamma, binomial, poisson, ...) como, por ejemplo, los modelos de regresión lineal o de regresión logística.

Además, en este modelo se introdujeron las funciones *link* monótonas ( $g(\cdot)$ ) para la modelización de la media,  $\mu$ , de la variable respuesta en función de las variables explicativas. Por lo tanto, para  $\mathbf{y} \sim \text{ExpF}(\boldsymbol{\mu}, \boldsymbol{\phi})$ ,<sup>1</sup> se tiene:

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}. \quad (3.3)$$

Con este modelo se consiguió flexibilizar la restricción de normalidad, pero la de linealidad seguía presente. Con el fin de trabajar sin la suposición de linealidad, se introdujeron los modelos aditivos generalizados (GAM).

- **Modelos aditivos generalizados (GAM):** desarrollados por Hastie y Tibshirani (1990) y Wood (2006), estos modelos son una extensión no paramétrica de los GLM. La idea de los GAM es que sean los propios datos los que determinan la relación entre el predictor lineal  $\eta$  y las variables explicativas. Dado  $\mathbf{y} \sim \text{ExpF}(\boldsymbol{\mu}, \boldsymbol{\phi})$ , se tiene:

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \sum_{j=1}^J h_j(\mathbf{x}_j), \quad (3.4)$$

donde  $h$  son funciones de suavizado no paramétricas, que se aplican sobre las variables explicativas continuas que no presentan una relación lineal con respecto a la variable respuesta. Consiguiéndose de esta forma una modelización conjunta de dependencia suave con las variables explicativas, para el parámetro que regula la media de la distribución.

En resumen, este modelo utiliza funciones suavizadoras para modelizar la relación entre las variables explicativas y las variables respuesta, de una forma no paramétrica.

Con los modelos GAM quedaron solventadas las dos limitaciones principales que presentaban los modelos LM. Sin embargo, los GAM no permiten introducir distribuciones diferentes a las de la familia exponencial ni modelizar parámetros distintos a la media directamente a partir de las variables explicativas.

Tanto en los modelos GLM como en los GAM, la varianza viene dada por  $\text{var}(y) = \phi v(\mu)$ , donde  $\phi$  es un parámetro de dispersión, por lo que, depende de la media. Además, si se considera una distribución de la familia exponencial, tanto la asimetría como la curtosis también dependen de la media. Es decir, en estos modelos, ni el parámetro de escala ni los de forma son modelizados explícitamente a partir de las variables explicativas (como ocurre con la media), sino que se hace a través de la relación de las variables independientes con la media, y de ésta con el resto de parámetros.

Para solventar estos inconvenientes, se introdujeron los modelos GAMLSS definidos en la ecuación (3.1). Concretamente, este trabajo se centró en la *formulación aditiva semi-paramétrica* de los GAMLSS, un submodelo que combina la propiedad de los GAMLSS de modelizar todos los parámetros

<sup>1</sup>  $\text{ExpF}$  denota la familia exponencial, cuya función de densidad para  $y$  univariante se define:

$$f_Y(y; \mu, \sigma) = \exp\left\{\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right\},$$

donde para  $y$  univariante  $E(y) = \mu = b'(\theta)$  y  $\text{Var}(y) = \phi v(\mu)$  ( $v(\mu) = b''[\theta(\mu)]$ ).

de la distribución (no solo la media) con el uso de funciones no paramétricas suaves de las variables explicativas de los modelos GAM.

- La **formulación aditiva semi-paramétrica** de los GAMLSS se define como:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}(\mathbf{x}_{jk}), \quad (3.5)$$

donde  $h_{jk}$  es la función de suavizado que ya aparecía en el modelo GAM (3.4) y es una función desconocida de la covariable  $X_{jk}$ . Esta formulación viene de tomar  $\mathbf{Z}_{jk} \boldsymbol{\gamma}_{jk} = h_{jk}(\mathbf{x}_{jk})$  en el modelo general GAMLSS (3.1)

Para estimar las funciones de suavizado se pueden usar distintos métodos: splines cúbicos, p-splines, loess, etc. Según el método que se utilice, se tiene una formulación diferente para  $\mathbf{Z}$  y  $\boldsymbol{\gamma}$ . Dado que  $\boldsymbol{\gamma}$  representa un efecto aleatorio, en realidad se está escribiendo el suavizado no paramétrico como un modelo de efectos aleatorios.

La estimación del modelo en los GAMLSS se lleva a cabo mediante la maximización del logaritmo de la función de verosimilitud penalizada, cuya definición se presenta a continuación.

**Definición 3.2.** La función de log verosimilitud penalizada,  $\ell_p$ , se define por

$$\ell_p = \ell - \frac{1}{2} \sum_{k=1}^p \sum_{j=1}^{J_k} \lambda_{jk} \boldsymbol{\gamma}_{jk}^t \mathbf{G}_{jk} \boldsymbol{\gamma}_{jk}, \quad (3.6)$$

donde  $\ell = \sum_{i=1}^n \log f(y_i | \boldsymbol{\theta}^i)$  es el logaritmo de la función de verosimilitud,  $\lambda_{jk}$  es un conjunto de parámetros de penalización (o de suavizado) denominados hiperparámetros y  $\mathbf{G}_{jk} = \mathbf{G}_{jk}(\boldsymbol{\lambda}_{jk})$  es una matriz simétrica de dimensión  $q_{jk} \times q_{jk}$ , cuya expresión depende del método de suavizado que se utilice. Es decir, la segunda parte de la ecuación representa a los términos penalizados procedentes de la suavización no paramétrica. Para el efecto aleatorio  $\boldsymbol{\gamma}_{jk}$  se asume normalidad e independencia, de forma que  $\boldsymbol{\gamma}_{jk} \sim N_{q_{jk}}(0, \mathbf{G}_{jk}^{-1})$ .

Por lo tanto, se tiene que, para ajustar un modelo GAMLSS se deben estimar  $\boldsymbol{\beta}$ ,  $\boldsymbol{\gamma}$  y  $\boldsymbol{\lambda}$ . A continuación se muestran los algoritmos para la estimación de estos términos.

### 3.1.2. Algoritmos

Para el ajuste de los modelos GAMLSS, Rigby y Stasinopoulos (2005) introdujeron dos algoritmos básicos: CG (Cole and Green) o RS (Rigby and Stasinopoulos). Estos algoritmos maximizan el logaritmo de la función de verosimilitud penalizada con respecto a  $\boldsymbol{\beta}$  y  $\boldsymbol{\gamma}$  para  $\boldsymbol{\lambda}$  dado.

Los parámetros  $\boldsymbol{\lambda}$  pueden ser fijados manualmente o estimados mediante un proceso automático, lo cual es más aconsejable.

Se ha utilizado la referencia [17].

- El **algoritmo RS** generaliza el utilizado por Rigby y Stasinopulos. En algunas funciones de densidad los parámetros de distribución son información ortogonal (ya que los valores esperados de las derivadas cruzadas de la función de verosimilitud son 0), por ejemplo, modelos de ubicación y escala o modelos de familias de dispersión, para estos casos, el algoritmo RS es más simple y adecuado. Para las distribuciones que se utilizaron en este trabajo, se eligió el método RS, que funcionó de forma exitosa. Es por ello que se explicará con más detalle.

Este algoritmo no utiliza las derivadas cruzadas de la función logaritmo de la verosimilitud (a diferencia del CG) y puede describirse utilizando los siguientes pasos (anidados): el ciclo exterior, el ciclo interior y el algoritmo backfitting (hacia atrás). Estos pasos se detallan a continuación, en primer lugar se motivan intuitivamente y luego se presenta el algoritmo específico.

1. *Ciclo exterior:*

Sea  $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0, \nu_0, \tau_0)$  valores iniciales de los parámetros de distribución. Para cada parámetro  $\boldsymbol{\theta}_k$  se ajusta un modelo usando el *ciclo interior*, a partir de las últimas estimaciones hechas para los parámetros restantes. Por ejemplo, para  $\mu$  se ajusta un modelo a partir de  $\hat{\sigma}, \hat{\nu}$  y  $\hat{\tau}$ . Se calcula la Desvianza Global (GD)<sup>2</sup>, si esta no cambia (no se hace más pequeña), quiere decir que converge y el algoritmo se detiene pero, si por lo contrario, la GD cambia se repite este proceso.

2. *Ciclo interior:*

Dentro de los ajustes hechos en 1, para cada parámetro, es necesario utilizar el *ciclo interior*. Es un algoritmo de puntuación o score (algoritmo de tipo Newton-Raphson para la derivada de la función de log-verosimilitud) que utiliza, para ajustar cada parámetro, la siguiente modificación de la variable respuesta:

$$\mathbf{z}_k = \boldsymbol{\eta}_k + \mathbf{W}_{kk}^{-1} \mathbf{u}_k, \quad (3.7)$$

donde  $\mathbf{z}_k, \boldsymbol{\eta}_k$ , y  $\mathbf{u}_k$  son vectores  $n$ -dimensionales,  $\mathbf{u}_k = \frac{\partial \ell}{\partial \boldsymbol{\eta}_k}$  es la derivada parcial del logaritmo de la función de verosimilitud respecto al predictor lineal (función de puntuación o score) y  $\mathbf{W}_{kk}$  es una matriz de pesos iterativos para  $k = 1, \dots, p$  que se define de una de las siguientes formas:  $-E \left[ \frac{\partial^2 \ell}{\partial \boldsymbol{\eta}_k^2} \right]$ ,  $-\frac{\partial^2 \ell}{\partial \boldsymbol{\eta}_k^2}$  o  $diag \left\{ \left( \frac{\partial \ell_i}{\partial \boldsymbol{\eta}_{ik}} \frac{\partial \ell_i}{\partial \boldsymbol{\eta}_{ik}} \right) \right\}$ , según se utilice el algoritmo de puntuación de Fisher, el de Newton-Raphson o el de cuasi Newton-Raphson, respectivamente.

En esta iteración se repiten los ajustes ponderados (3.7) hasta alcanzar la convergencia, de forma que, mientras se van utilizando distintos pesos, la variable respuesta se va modificando. Entonces, dadas las estimaciones actualizadas de los parámetros de distribución, se calculan  $\mathbf{z}_k$  y  $\mathbf{W}_{kk}$ , con estos valores se ajustan los términos paramétricos y aditivos a partir del

<sup>2</sup> Global Deviance, es uno de los criterios de selección del mejor modelo:

$$GD = 2\ell(\hat{\boldsymbol{\theta}})$$



*algoritmo backfitting* y se recalculan las estimaciones del predictor lineal y de los parámetros de distribución. Este ciclo converge cuando los valores no cambien.

### 3. *Algoritmo backfitting*:

En este paso se calculan las estimaciones de los parámetros  $\beta$  y los términos  $\gamma$ . Utilizando los pesos  $\mathbf{W}_{kk}$  calculados en la *iteración interior*, el *algoritmo backfitting* estima tanto las variables lineales como las no paramétricas (o suavizadas) para  $\mathbf{z}_k$ . El proceso se sigue de la siguiente forma:

Dados  $\mathbf{z}_k$ ,  $\mathbf{W}_{kk}$  y  $\hat{\gamma}_{jk}$  con  $k = 1, \dots, p$  y  $j = 1, \dots, J_k$ , estimados anteriormente:

- o En primer lugar, se calculan los residuos parciales para  $\beta_k$ ,

$$\varepsilon_\beta = \mathbf{z}_k - \sum_{j=1}^{J_k} \mathbf{Z}_{jk} \hat{\gamma}_{jk} \quad (3.8)$$

que actúan como regresandos contra la matriz de diseño  $\mathbf{X}_k$  utilizando los pesos iterativos  $\mathbf{W}_{kk}$  para estimar un nuevo  $\hat{\beta}_k$ .

- o A continuación, se calculan los residuos parciales con respecto a  $\gamma_{jk}$ ,

$$\varepsilon_\gamma = \mathbf{z}_k - \mathbf{X}_k \hat{\beta}_k - \sum_{j=1, t \neq j}^{J_k} \mathbf{Z}_{tk} \hat{\gamma}_{tk} \quad (3.9)$$

y estos residuos se suavizan utilizando la matriz de suavizado  $\mathbf{S}_{jk}$  para actualizar  $\hat{\gamma}_{jk}$ , donde

$$\mathbf{S}_{jk} = \mathbf{Z}_{jk} (\mathbf{Z}_{jk}^T \mathbf{W}_{kk} \mathbf{Z}_{jk} + \mathbf{G}_{jk})^{-1} \mathbf{Z}_{jk}^T \mathbf{W}_{kk}. \quad (3.10)$$

Se repite este proceso hasta que se encuentren estimaciones  $\hat{\beta}_k$  y términos  $\hat{\gamma}_{jk}$  que no cambien.

A continuación, se va a describir esquemáticamente el algoritmo RS, pero antes es necesario aclarar la notación:

- \*  $k$ : índice de parámetros.
- \*  $r$ : índice de iteración del algoritmo exterior.
- \*  $i$ : índice de iteración del algoritmo interior.
- \*  $m$ : índice de iteración del algoritmo backfitting.
- \*  $\gamma_{jk}^{(r,i,m)}$ : valor actualizado de  $\gamma_{jk}$  en la  $r$ -ésima iteración exterior,  $i$ -ésima iteración interior y  $m$ -ésima iteración del backfitting.
- \*  $\gamma_{jk}^{(r,i,\cdot)} = \hat{\gamma}_{jk}$ : valor estimado de  $\gamma_{jk}$  en la convergencia del algoritmo backfitting para la  $r$ -ésima iteración exterior e  $i$ -ésima iteración interior, que es el valor de inicio para la  $(i+1)$ -ésima iteración del algoritmo interior de la  $r$ -ésima iteración del algoritmo externo.

**Representación del algoritmo:**

**PASO 0: (A)** Se inicializa,  $\theta_k^{(1,1)}$  y  $\gamma_{jk}^{(1,1)}$  para  $k = 1, \dots, p$  y  $j = 1, \dots, J_k$  y se evalúan los predictores lineales  $\eta_k^{(1,1)} = g_k(\theta_k^{(1,1)})$ .

**PASO 1: (B)** Se abre el *ciclo exterior*  $r =, \dots$ , hasta convergencia. Se ajustan modelos para  $\theta_k$  dadas las estimaciones de los parámetros restantes.

(i) Se abre el *ciclo interior* (dentro del ciclo exterior)  $i = 1, \dots$ , hasta convergencia. Dadas las estimaciones de los parámetros anteriores, se calculan los valores de  $\mathbf{u}_k^{(r,i)}$ ,  $\mathbf{z}_k^{(r,i)}$  y  $\mathbf{W}_{kk}^{(r,i)}$ .

(ii) Se abre el *algoritmo backfitting* (dentro del ciclo interior)  $m = 1, \dots$ , hasta convergencia:

(ii.1) Se calculan los residuos parciales  $\varepsilon_{\beta}^{(r,i,m)}$  como se definió en la ecuación (3.8) y se obtiene  $\beta_k^{(r,i,m+1)}$ ,

(ii.2) se calculan los residuos parciales  $\varepsilon_{\gamma}^{(r,i,m)}$  como se definió en la ecuación (3.9) y se obtiene la actualización de los términos aditivos  $\mathbf{Z}_{jk}\gamma_{jk}^{(r,i,m+1)}$ .

(ii.3) Si el algoritmo convergió, se tiene  $\beta_k^{(r,i,\cdot)} = \beta_k^{(r,i+1)}$  y  $\gamma_{jk}^{(r,i,\cdot)} = \gamma_{jk}^{(r,i+1)}$  y se cierra el *algoritmo backfitting*, en caso contrario, se actualiza  $m$  y se repite este proceso (desde (ii.1)).

(iii) Se actualizan los valores de  $\eta_k^{(r,i+1)}$  y  $\theta_k^{(r,i+1)}$ .

(iv) Si el algoritmo convergió, se tiene  $\beta_k^{(r,\cdot)} = \beta_k^{(r+1,1)}$  y  $\gamma_{jk}^{(r,\cdot)} = \gamma_{jk}^{(r+1,1)}$  y se recalculan los valores  $\eta_k^{(r+1,1)} = \eta_k^{(r,\cdot)}$  y  $\theta_k^{(r+1,1)} = \theta_k^{(r,\cdot)}$  y se cierra el *ciclo interior*. En caso contrario, se actualiza  $i$  y se repite el proceso (desde (ii)).

**PASO 2: (C)** Se actualiza el valor de  $k$ .

**PASO 3: (D)** Por último, se termina el *ciclo exterior*, si el cambio en la verosimilitud (penalizada) es suficientemente pequeño; de lo contrario se actualiza  $r$  y se repite el proceso (desde PASO 1).

- El **algoritmo CG** es una generalización del método usado por Cole y Green (1992) en la maximización de la verosimilitud penalizada. También es un algoritmo de puntuación o score que, como el algoritmo RS, está formado por tres ciclos anidados: un ciclo exterior, otro interior y el algoritmo backfitting. Las principales diferencias con respecto al método RS son que en las matrices de pesos  $\mathbf{W}_{ks}$  con  $k, s = 1, 2, \dots, p$ , se utilizan derivadas cruzadas de la función de log-verosimilitud, y que la actualización de dichas matrices de pesos se produce después de haber ajustado todos los parámetros  $\theta_k$ ,  $k = 1, 2, \dots, p$ . Se omite la presentación exhaustiva del algoritmo, ya que no fue el método utilizado en el trabajo.

En resumen, la maximización de la función  $\ell_p$  respecto de los parámetros  $\beta_k$  y los términos  $\gamma_{jk}$  con  $k = 1, \dots, p$  y  $j = 1, \dots, J_k$ , para la estimación del modelo, conduce a la utilización de uno de los métodos descritos en esta sección. La idea es calcular primeras y segundas derivadas (y derivadas parciales cruzadas en el caso del CG) de la ecuación (3.6) para dar un paso de uno de los algoritmos de

puntuación mencionados. Cada paso de los algoritmos de puntuación se integra mediante el backfitting que recorre los parámetros y los términos aditivos de los  $k$  predictores lineales.

■ Métodos de estimación del parámetro  $\lambda$ :

La estimación de  $\lambda$  suele hacerse utilizando alguno de los siguientes criterios: validación cruzada generalizada (GCV), criterio akaike generalizado (GAIC) o métodos basados en máxima verosimilitud (ML/REML). Cada uno de estos métodos puede usarse globalmente, cuando se aplica fuera del algoritmo o localmente, dentro de este. Los métodos locales son más rápidos, sin embargo, los globales son más fiables.

Concretamente en este trabajo se utilizaron dos métodos, uno local y otro global. El primero consistió en aplicar máxima verosimilitud local y el segundo en la minimización del criterio GAIC.

- El método de **máxima verosimilitud local** interfiere en la parte del algoritmo backfitting donde se ajusta la variable  $\gamma$ . Para estimar  $\lambda$ , se asume un modelo de efectos aleatorios:

$$\varepsilon = \mathbf{Z}\gamma + \alpha, \quad (3.11)$$

donde  $\mathbf{Z}$  representa la matriz de diseño para la suavización de la variable explicativa correspondiente,  $\varepsilon$  son los residuos parciales dentro del backfitting y se asume que  $\alpha \sim N(0, \sigma_\alpha)$  y  $\gamma \sim N(0, \sigma_\gamma \mathbf{G}^{-1})$ .

Se tiene que el parámetro de suavizado  $\lambda$ , es un ratio entre las varianzas  $\sigma_\alpha^2$  y  $\sigma_\gamma^2$ , es decir,  $\lambda = \frac{\sigma_\alpha^2}{\sigma_\gamma^2}$ .

Para estimar  $\sigma_\alpha$ ,  $\sigma_\gamma$  y  $\gamma$  se sigue:

1. En primer lugar, se calcula  $\lambda$  utilizando mínimos cuadrados penalizados (PLS), en notación matricial (omitiendo los subíndices):

$$\hat{\gamma} = (\mathbf{Z}^t \mathbf{W} \mathbf{Z} + \lambda \mathbf{G}^{-1})^{-1} \mathbf{Z}^t \mathbf{W} \varepsilon. \quad (3.12)$$

2. Con el valor  $\hat{\gamma}$  estimado en el punto anterior, se calcula  $\hat{\varepsilon} = \mathbf{Z} \hat{\gamma} = \mathbf{S} \varepsilon$ , donde  $\mathbf{S}$  es la matriz definida en (3.10), con esto se obtiene:

$$\sigma_\alpha^2 = \frac{(\varepsilon - \hat{\varepsilon})^t (\varepsilon - \hat{\varepsilon})}{(n - \text{tr}(\mathbf{S}))}, \quad (3.13)$$

$$\sigma_\gamma = \frac{\hat{\gamma}^t \hat{\gamma}}{\text{tr}(\mathbf{S})}, \quad (3.14)$$

y, con esto, se tienen el valor que se desea estimar:

$$\lambda = \frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_\gamma^2} \quad (3.15)$$

3. Si  $\lambda$  no cambia el algoritmo converge y se para, en caso contrario, se vuelve al paso 1.

- Método de **minimización del GAIC global con respecto al parámetro  $\lambda$**

Es un criterio introducido por Hastie y Tibshiran (1990) para la estimación de los hiperparámetros en los modelos GAM, donde se utilizaron splines cúbicos (que se desarrollarán en la Sección 3.1.3) para modelizar la dependencia suave respecto de la variable explicativa del predictor. En este caso se utiliza de forma global, ya que se calculan fuera del modelo. La estimación de este  $\lambda$  se puede hacer a través de la minimización del criterio GAIC, que se define a continuación:

**Definición 3.3.** El criterio GAIC se define como,

$$GAIC(\#) = -2\hat{\ell} + \#df, \quad (3.16)$$

donde  $\hat{\ell}$  denota la log verosimilitud del ajuste,  $df$  denota los grados efectivos de libertad usados en el modelo, los cuales representan la dimensión de parámetros de la parte paramétrica y de los efectos aleatorios y  $\#$  una penalización fijada.

Para un solo término de spline suavizado se tiene que el parámetro  $\lambda$  correspondiente está relacionado con los grados de libertad,  $df = tr(\mathbf{S})$ , con los que se ajustó ese término.

En el caso de que haya  $p$  términos de splines de suavizado (correspondiéndose con los  $p$  parámetros),  $\boldsymbol{\lambda}$  es un vector de hiperparámetros tal que  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)$  correspondientes a las  $p$  funciones de suavizado. Estos hiperparámetros pueden ajustarse conjuntamente minimizando el criterio GAIC. Sin embargo, con múltiples splines de suavizado,  $\sum_{j=1}^p tr(\mathbf{S}_j)$  solamente es una aproximación de la complejidad que proporcionan los grados de libertad del modelo. En la Sección 4.2, se explica como se aplicó este método.

### 3.1.3. Técnicas de suavización

Esta sección se centra en describir las técnicas de suavización empleadas para estimar las funciones  $h(x)$  del modelo (3.5), fijándose por tanto en la parte no paramétrica del modelo. Dependiendo de las técnicas que se utilicen, la función de verosimilitud penalizada del modelo tiene una expresión determinada. En este trabajo se utilizaron métodos basados en splines, concretamente los splines cúbicos y los splines penalizados o p-splines. Se ha utilizado la referencia [1].

Respecto a los modelos de suavizado con splines, existen dos grandes enfoques: splines de suavizado que utilizan el mismo número de parámetros y de observaciones, lo cual debilita su eficiencia a medida que aumenta el número de observaciones; y splines polinómicos o de regresión, que se pueden ajustar utilizando mínimos cuadrados una vez que se selecciona el número de nodos (aunque la selección de los nodos exige algoritmos engorrosos). Los p-splines combinan las ventajas de los dos grupos, pues superan la fuerte dependencia del número y localización de los nodos impuestos por los splines polinómicos, fijando un número moderado que garantice la flexibilidad suficiente, y controlan la suavidad agregando un término de penalización para evitar una sobreestimación.

A continuación se desarrollan estas técnicas de suavización. Se empieza estableciendo la definición formal de spline polinómico [10], se continúa introduciendo los splines cúbicos de suavizado y, posteriormente, se definen los p-splines. Los dos últimos métodos son los que se utilizan en este trabajo.

### 3.1.4. Splines polinómicos o de regresión

Los splines polinómicos están estrechamente relacionados con la idea de regresión polinomial, es decir, con suponer que el efecto de la covariable  $x$  sobre  $y$  es un polinomio de grado  $l$ ,

$$h(x_i) = \gamma_0 + x_i\gamma_1 + \cdots + x_i^l\gamma_l,$$

tanto los coeficientes de la regresión  $\gamma_j$  como la función  $h$  se pueden estimar utilizando mínimos cuadrados ordinarios (OLS).

Sin embargo, hay una forma más efectiva de regresión polinómica, que consiste en dividir el dominio de la covariable  $x$  en intervalos y estimar cada polinomio en cada intervalo por separado. Es decir, utilizar polinomios definidos a trozos (en lugar de un modelo global). Evidentemente, esto no conduce a una suavidad general, sobre todo en los extremos de los intervalos. Es por esto que se introducen restricciones de suavidad. Finalmente, la regresión polinómica a trozos junto con las restricciones de suavidad conducen a los splines polinómicos.

**Definición 3.4.** Una función  $h : [a, b] \rightarrow \mathbf{R}$  es un spline polinómico de grado  $l$  si verifica las siguientes condiciones:

1.  $h(x)$  es  $(l - 1)$  veces continuamente diferenciable.
2.  $h(x)$  es un polinomio de grado  $l$  en todo el intervalo  $[k_j, k_j + 1]$  con  $j = 1, \dots, m - 1$ .

Donde  $a = k_1 < k_2 < \cdots < k_{m-1} < k_m = b$  es una partición del intervalo  $[a, b]$ . A la vista de la definición anterior, se puede decir que el espacio de splines polinómicos es un espacio vectorial  $(m + l - 1)$  dimensional, y que a su vez es un subespacio del espacio de funciones  $(l - 1)$  veces continuamente diferenciables. Por lo tanto, cada spline polinómico de grado  $l$  puede ser determinado de forma única como combinación lineal de  $d = m + l - 1$  funciones  $B_1, \dots, B_d$ , que forman una base de splines.

Esto lleva a poder escribir el problema de regresión no paramétrica de la siguiente forma:

$$y_i = \sum_{j=1}^d \gamma_j B_j(x_i) + \varepsilon_i. \quad (3.17)$$

La ventaja de los splines polinómicos, es que permiten entender la regresión no paramétrica como una regresión lineal con muchos parámetros posibles. La matriz de diseño se define de la siguiente forma:

$$\mathbf{Z} = \begin{pmatrix} B_1(x_1) & \cdots & B_d(x_1) \\ \vdots & & \vdots \\ B_1(x_n) & \cdots & B_d(x_n) \end{pmatrix},$$

la regresión se puede expresar en forma matricial:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad (3.18)$$

donde  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_d)'$  son los coeficientes de la regresión que pueden estimarse por mínimos cuadrados ordinarios.

En particular, los splines cúbicos son splines polinómicos de grado 3. La forma más típica son los splines cúbicos naturales, donde la derivada segunda de  $h(x)$  se hace 0 para el primer y último punto sobre el que está definido el conjunto de Splines, lo que conlleva a que la dimensión de la base de splines naturales cúbicos coincida con el número de nodos elegidos, es decir  $d = m$ . Los splines cúbicos están cubiertos extensamente en la literatura [10], [16].

### 3.1.5. Splines de suavizado

Los splines de suavizado son la solución al problema de minimizar la suma de residuos al cuadrado penalizados:

$$RSS(h, \lambda) = \sum_{i=1}^n (y_i - h(x_i))^2 + \lambda \int_{x_1}^{x_n} (h''(x))^2 dx \quad (3.19)$$

la solución de esta minimización son splines cúbicos naturales con los nodos iguales a las observaciones de  $X$ , es decir se tienen  $n$  nodos. Supóngase que se quiere aproximar  $h(x)$  por un spline de suavizado, se puede presentar como una combinación lineal de funciones básicas de splines cúbicos naturales de la forma:  $h(x) = \sum_{j=1}^d \gamma_j B_j(x_i)$ , donde, como ya se ha dicho,  $d = n$  y  $B_j$  funciones básicas de splines naturales. Entonces, se define una penalización cuadrática para  $\boldsymbol{\gamma}$ , a través de la integral del cuadrado de la derivada segunda de  $h$  (es decir, la curvatura) de la siguiente forma:

$$\begin{aligned} \int (h''(x))^2 dx &= \int \left( \sum_{j=1}^d \gamma_j B_j''(x) \right)^2 dx \\ &= \int \left( \sum_{r=1}^d \sum_{j=1}^d \gamma_r \gamma_j B_r''(x) B_j''(x) \right) dx \\ &= \sum_{r=1}^d \sum_{j=1}^d \gamma_r \gamma_j \int B_r''(x) B_j''(x) dx \\ &= \boldsymbol{\gamma}' \mathbf{K} \boldsymbol{\gamma}, \end{aligned} \quad (3.20)$$

donde  $\mathbf{K}[r, j] = \int B_r''(x) B_j''(x) dx$ , es decir, la matriz  $\mathbf{K}$  está formada por las segundas derivadas de las funciones básicas. Esta matriz tiene dimensión  $n \times n$ , lo cual hace que sea una matriz pesada. La parte penalizada del logaritmo de la función de verosimilitud  $\ell_p$ , cuando se utilizan splines de suavizado para modelizar el suavizado, es de la forma  $\lambda \int (h''(x))^2 dx$ .

### 3.1.6. Splines penalizados

Eilers y Marx (1996) introdujeron los splines penalizados o p-splines para estimar los efectos suaves de las covariables con el fin de describir una amplia clase de funciones flexibles utilizando un número moderado de parámetros. Estos splines permiten estimar un gran número de efectos no paramétricos simultáneamente.

La idea principal de los p-splines es: aproximar la función  $h(x)$  de la ecuación (3.5) con un spline polinómico utilizando un número fijo de nodos (20–40 aproximadamente) que garantice la suficiente flexibilidad, introducir un término de penalización adicional para evitar un ajuste excesivo y minimizar el criterio PLS, en lugar de el OLS.

Cuando se utilizan p-splines, los enfoques más comunes para expresar los splines como representación de funciones básicas son los basados en series de tiempo y los basados en B-splines. Este último es el que se utilizó para este trabajo.

Los p-splines basados en B-splines consisten en  $(l + 1)$  trozos polinómicos que se unen  $(l - 1)$  veces de forma continuamente diferenciable, esta fusión se hace de forma suave en los nodos, para lograr la restricción de suavidad. La principal ventaja de las B-splines es su definición local, ya que sus funciones solo son positivas en un intervalo basado en  $l + 2$  nodos. Su definición, de forma recursiva, viene dada por:

$$B_j^l(x) = \begin{cases} I(k_j \leq x < k_{j+1}) & \text{si } l = 0 \\ \frac{x - k_{j-l}}{k_j - k_{j-l}} B_j^{l-1}(x) + \frac{k_{j+1} - x}{k_{j+1} - k_{j+1-l}} B_j^{l-1}(x) & \text{si } l > 0 \end{cases}, \quad (3.21)$$

para  $j = 1, \dots, d - 1$ . Siendo  $I$  la función indicadora. Se puede ver que esta construcción recursiva precisa de  $2l$  nodos externos al dominio  $[a, b]$ , es decir, la expansión a  $k_{l-1}, \dots, k_{m+l-1}, k_{m+l}$  nodos. Nótese que si los nodos interiores son equidistantes, también lo son los nodos externos.

De manera análoga a lo que se ha descrito anteriormente, la estimación de un spline polinomial utilizando una representación B-spline, como hacen los p-splines, se puede ver como una estimación de un modelo lineal con un gran número de parámetros y matriz de diseño:

$$\mathbf{Z} = \begin{pmatrix} B_1^l(x_1) & \cdots & B_d^l(x_1) \\ \vdots & & \vdots \\ B_1^l(x_n) & \cdots & B_d^l(x_n) \end{pmatrix}.$$

Dado que las B-splines se definen de forma local,  $\mathbf{Z}$  se compone principalmente de ceros. Esto da como resultado una matriz de banda para  $\mathbf{Z}'\mathbf{Z}$  con longitud de banda  $l$ .

Por otro lado, para el caso de los p-splines, la parte penalizada de la función  $\ell_p$ , viene dada por la siguiente expresión:

$$\text{pen}(\lambda) = \lambda \sum_{m=r+1}^d (\Delta^r \gamma_j)^2, \quad (3.22)$$

siendo  $\lambda$  el parámetro de suavizado ya conocido que controla el *tradeoff* o balance entre la fidelidad de los datos y la suavización del modelo; y  $\Delta^r$  representa la diferencia de orden  $r$  aplicada a los coeficientes  $\gamma_j$  de las B-splines, que se define de la siguiente forma:

$$\begin{aligned} \Delta^1 \gamma_j &= \gamma_j - \gamma_{j-1} \\ \Delta^2 \gamma_j &= \Delta^1 (\Delta^1 \gamma_j) = \gamma_m - 2\gamma_{j-1} + \gamma_j - 2 \\ &\vdots \\ \Delta^r \gamma_j &= \Delta^1 (\Delta^{j-1} \gamma_j), \end{aligned}$$

Además, este término de penalización se puede escribir en forma matricial. Para ello, se define la matriz de diferencias de dimensión  $(d-1) \times d$  como:

$$\mathbf{D}_1 = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{pmatrix},$$

y de forma recursiva se definen las matrices de orden superior,  $\mathbf{D}_k = \mathbf{D}_1 \mathbf{D}_{r-1}$ . Por lo tanto, se tiene que la expresión para la parte penalizada en forma matricial tiene la siguiente forma:

$$\lambda \sum_{m=r+1}^d (\Delta^r \gamma_j)^2 = \lambda \boldsymbol{\gamma}' \mathbf{D}'_r \mathbf{D}_r \boldsymbol{\gamma} = \lambda \boldsymbol{\gamma}' \mathbf{K}_r \boldsymbol{\gamma},$$

donde la matriz de penalización  $\mathbf{K}_r$  viene dada por  $\mathbf{D}'_r \mathbf{D}_r$ , en este caso la matriz  $\mathbf{K}_r$  tiene dimensión  $r \times r$  lo que implica que es mucho más pequeña que la de los splines de suavizado, disminuyendo la dificultad computacional.

En resumen, Eilers y Marx (1996) sugirieron utilizar un conjunto de funciones B-spline en la variable explicativa  $X$  (cuyas evaluaciones en los valores  $\boldsymbol{x}$  de  $X$  son las columnas de la matriz de diseño). Ellos sugirieron el uso de un número moderadamente grande de nodos equidistantes, a los que se conectan los segmentos de spline, para asegurar una flexibilidad suficiente en las curvas ajustadas, además impusieron penalizaciones directamente sobre los parámetros  $\boldsymbol{\gamma}$  de las funciones B-splines para garantizar una suavidad suficiente de las curvas ajustadas. Se puede decir que son splines de bajo rango en el que el número de nodos es mucho menor que la dimensión de los datos.

### 3.2. Estimación percentil mediante los métodos LMS y BCPE

Una vez explicados los modelos GAMLSS en general, la formulación semi-paramétrica aditiva en particular y la suavización con splines, en esta sección se describen los modelos de estimación de curvas percentiles seleccionados para este trabajo, el método LMS [2] y el BCPE [19].

Con fines introductorios, supongamos que  $Y$  es una variable continua y que se quiere calcular su percentil de orden  $\alpha$ . Si los valores de  $y$  siguen una distribución normal,  $y \sim N(\mu, \sigma)$ , se tiene

$$z = \frac{y - \mu}{\sigma} \sim N(0, 1).$$

Cada valor  $z$ , se denomina “z-score” o puntuación tipificada y se interpreta como el número de desviaciones de la media de un determinado valor  $y$  de la variable. En la  $N(0, 1)$  existe una correspondencia inmediata entre cada z-score y el percentil que representa, a través de la tabla de la  $N(0, 1)$ . Por ejemplo, si  $z = 1.96$  entonces  $z$  es el percentil 97.5<sup>th</sup>, dado que  $\Phi(1.96) = 0.975$  siendo  $\Phi$  la función de distribución normal estándar. Recíprocamente, dado un  $\alpha$  particular, el percentil de orden  $\alpha$  viene dado por la puntuación  $z_\alpha = \Phi^{-1}(\alpha)$ . Además, es inmediato comprobar que el percentil de orden  $\alpha$  de  $Y \sim N(\mu, \sigma)$ ,  $0 < \alpha < 1$  es:

$$Y_\alpha = \mu + \sigma z_\alpha, \tag{3.23}$$



Sin embargo, cuando la distribución de la variable  $Y$  no es normal, el cálculo del  $z$ -score o puntuación tipificada no se corresponde de forma clara con un determinado percentil. En el contexto de distribuciones unimodales y asimétricas Cole definió una transformación particular de  $Y$  dependiente de 3 parámetros ( $L, M$  y  $S$ ) para obtener normalidad. Su transformación proporciona una fórmula sencilla para el percentil  $Y_\alpha$  a partir de los 3 parámetros y del  $z_\alpha$  de una  $N(0, 1)$ . Además, con el objetivo de obtener curvas de percentiles suaves de  $Y$  en función de la variable explicativa edad  $X$ , el método LMS modeliza cada uno de los tres parámetros como una función suave (spline) de la edad. A partir de los valores estimados para cada edad ( $L(x), M(x)$  y  $S(x)$ ) se calculan las curvas percentiles,  $Y_\alpha(x)$ .

Finalmente, dado que el método de Cole no corrige la presencia de curtosis, el método BCPE (o LMSP) generaliza la transformación de Cole introduciendo un cuarto parámetro adicional ( $P = power$ ) que tenga en cuenta la curtosis (a mayores de la asimetría). A continuación, se explica en que consisten estos métodos.

### 3.2.1. Método LMS

En este modelo se asume que  $Y$  sigue una distribución llamada Box-Cox Normal o Box-Cox Cole y Green (BCCG), con tres parámetros ( $\mu, \sigma, \nu$ ). Precisamente Cole y Green proponen una transformación de  $Y$  que da lugar a una  $N(0, 1)$  y proporciona  $z$ -scores (o puntuaciones  $z$ ) adecuadas para el cálculo de los percentiles. Esta transformación se denomina LMS y se define a continuación.

**Definición 3.5.** La transformación LMS de  $Y$  fue definida por Cole y Green de la siguiente forma

$$z_{LMS} = \begin{cases} \frac{1}{\sigma\nu} \left[ \left( \frac{Y}{\mu} \right)^\nu - 1 \right] & \text{si } \nu \neq 0, \\ \frac{1}{\sigma} \log \left( \frac{Y}{\mu} \right) & \text{si } \nu = 0, \end{cases} \quad (3.24)$$

para  $0 < Y < \infty$ , donde  $\mu, \sigma > 0$  y donde se asume que  $z_{LMS}$  sigue una distribución normal estándar. Por otro lado,  $\mu$  es el parámetro  $M$  que representa la media o mediana,  $\sigma$  es el parámetro  $S$  que representa el coeficiente de variación y  $\nu$  es el parámetro  $L$  que representa el coeficiente de asimetría. De aquí vienen las siglas L, M y S del modelo.

Diremos que  $Y \sim BCCG(\mu, \sigma, \nu)$  si al hacerle la transformación (3.24) se obtiene una normal estándar  $N(0, 1)$ . Los parámetros ( $\mu, \sigma, \nu$ ) son los parámetros de la distribución. La función de densidad de  $Y$  viene dada por:

$$f_Y(y) = \frac{y^{\nu-1} \exp\left(-\frac{1}{2}z^2\right)}{\mu^\nu \sigma \sqrt{\pi} \Phi\left(\frac{1}{\sigma|\nu|}\right)},$$

para  $y > 0$ , donde  $z$  viene dado por la ecuación (3.24) y  $\Phi$  es la función de distribución normal estándar.

El método LMS es un caso particular de un modelo GAMLSS donde hay una única variable explicativa  $X$ . Su formulación es la siguiente: Dado  $X = x$ ,  $Y$  sigue una distribución  $BCCG(\mu, \sigma, \nu)$  y cada uno de los parámetros es modelado como una función no paramétrica suave de  $x$ , es decir

$$\begin{aligned}
g_1(\mu) &= h_1(x), \\
g_2(\sigma) &= h_2(x), \\
g_3(\nu) &= h_3(x),
\end{aligned}
\tag{3.25}$$

donde las funciones “link”,  $g_k(\cdot)$  para  $k = 1, 2, 3$ , son la identidad para  $\mu$  y  $\nu$  y el logaritmo para  $\sigma$ .

Puesto que es un caso particular de los modelos GAMLSS, el ajuste se hace mediante el método que maximiza la verosimilitud penalizada, cuya expresión depende del método de suavización que se aplique. En la sección (3.2.3) se detallará este proceso.

Una vez estimados los valores de los parámetros de la distribución  $(\mu, \sigma, \nu)$  en función de los valores de  $X$ , la expresión siguiente muestra que es muy fácil calcular las curvas percentiles para cualquier edad con este modelo.

**Definición 3.6.** El percentil de orden  $\alpha$  de una variable  $Y$  con distribución  $BCCG(\mu, \sigma, \nu)$  viene dado por:

$$Y_\alpha = \begin{cases} \mu [1 + \sigma \nu z_\alpha]^\frac{1}{\nu} & \text{si } \nu \neq 0, \\ \mu \exp(\sigma z_\alpha) & \text{si } \nu = 0, \end{cases}
\tag{3.26}$$

donde  $z_\alpha$  es el percentil de orden  $\alpha$  de una distribución normal estándar.

El método LMS se encarga principalmente de corregir la asimetría, no modeliza la curtosis. Se utiliza para calcular los  $z$ -scores, o equivalentemente las curvas percentiles, cuando la distribución de la muestra esté sesgada pero tenga un pico “normal”.

Nótese que si  $\nu = 1$ , este modelo se reduce a un modelo normal con dos parámetros.

### 3.2.2. Método BCPE (LMSP)

En este modelo se asume que  $Y$  sigue una distribución llamada Box–Cox Power Exponencial (BCPE), con cuatro parámetros. Esta distribución también se relaciona con la transformación  $z_{LMS}$ , definida en (3.24), pero en este caso se asume que  $z_{LMS}$  sigue una distribución power exponencial estándar. A continuación se definen las densidades de  $Z$  y de  $Y$ .

**Definición 3.7.** Una variable  $Z$  sigue una distribución power exponencial estándar de parámetro  $\tau > 0$  si su función de densidad  $Z$  es

$$f_Z(z) = \frac{\tau}{c2^{(1+\frac{1}{\tau})}\Gamma(\frac{1}{\tau})} \exp^{-0.5|\frac{z}{c}|^\tau},
\tag{3.27}$$

para  $-\infty < z < \infty$  y  $\nu > 0$ , donde  $c^2 = 2^{\frac{-2}{\tau}}\Gamma(\frac{1}{\tau}) [\Gamma(\frac{3}{\tau})]^{-1}$ . Esta parametrización asegura que  $Z$  tiene media 0 y desviación estándar 1 para todo  $\tau > 0$ . Además, cuando  $\tau = 2$  se reduce a la distribución normal estándar.  $\tau$  va a ser el cuarto parámetro  $P$  (power) que permitirá ajustar la curtosis.

De manera similar al método anterior, se dice que  $Y \sim BCPE(\mu, \sigma, \nu, \tau)$ , si al hacerle la transformación (3.24) se obtiene una distribución power exponencial estándar de parámetro  $\tau$ . Se puede

demostrar que la función de densidad de  $Y$  toma la siguiente forma:

$$f_Y(y) = f_Z(z) \left| \frac{dz}{dy} \right| = \frac{y^{\nu-1}}{\mu^\nu \sigma} f_Z(z). \quad (3.28)$$

El percentil  $Y_\alpha$  para una variable con distribución BCPE se obtiene mediante el uso de una función de aspecto mucho más complejo que la que se tiene para el método LMS (3.26), y viene dado por la siguiente expresión:

$$Y_\alpha = \begin{cases} \mu [1 + \sigma \nu z_\alpha]^\frac{1}{\nu} & \text{si } \nu \neq 0, \\ \mu \exp(\sigma z_\alpha) & \text{si } \nu = 0, \end{cases} \quad (3.29)$$

donde  $z_\alpha = F_Z^{-1}(\alpha)$  es el percentil de orden  $\alpha$  de una distribución power exponencial estándar con parámetro de potencia  $\tau$ . Asumiendo que la probabilidad de truncamiento es nula,  $z_\alpha$  viene dado por:

$$z_\alpha = \begin{cases} -c \left[ 2F_S^{-1}(1 - 2\alpha)^\frac{1}{\tau} \right] & \text{si } \alpha \leq 0.5, \\ c \left[ 2F_S^{-1}(2\alpha - 1)^\frac{1}{\tau} \right] & \text{si } \alpha > 0.5, \end{cases} \quad (3.30)$$

donde  $F_S^{-1}$  es la inversa de la función de distribución de  $S$ , que sigue una distribución gamma con función de densidad  $f_S(s) = s^{(\frac{1}{\tau})-1} \frac{\exp(-s)}{\Gamma(\frac{1}{\tau})}$ .

Cabe destacar que si  $\tau = 2$  la distribución de BCPE se reduce a LMS, o lo que es lo mismo, el método BCPE generaliza al método LMS. Entonces la curtosis será la de una normal.

Igual que en el método de Cole, la dependencia respecto de la edad se modeliza suponiendo que para cada  $X = x$  dado,  $Y$  sigue una distribución  $BCPE(\mu, \sigma, \nu, \tau)$  donde cada uno de los parámetros es una función no paramétrica suave de  $x$ . La estimación de los parámetros se realiza a través del método de máxima verosimilitud penalizada, que se explicó para los GAMLSS.

El método BCPE también conocido como método LMSP, corrige tanto la asimetría como la curtosis. Por lo tanto, este método se utiliza para estimar las curvas percentiles cuando la distribución sea asimétrica y además presente un pico no “normal”.

### 3.2.3. Estimación de las curvas $\mu$ , $\sigma$ , $\nu$ y $\tau$ para los modelos BCPE

En esta sección, se detalla el proceso de estimación para el método BCPE ya que, como ya se ha dicho, es una generalización del método LMS.

Se ajustan los parámetros de distribución como funciones de variables explicativas como se definió en (3.5). Particularmente, en este trabajo se consideró una única variable explicativa  $X$ , que es la edad. Por lo tanto el ajuste mediante el modelo BCPE, se define de la siguiente forma:

$$\begin{aligned} g_1(\mu) &= h_1(x), \\ g_2(\sigma) &= h_2(x), \\ g_3(\nu) &= h_3(x), \\ g_4(\tau) &= h_4(x), \end{aligned} \quad (3.31)$$

donde las funciones “*link*”,  $g_k(\cdot)$  para  $k = 1, 2, 3, 4$ , en este caso serán la identidad para  $\mu$  y  $\nu$  y logaritmo para  $\sigma$  y  $\tau$ . Recuerdese que  $Y \sim BCPE(\mu, \sigma, \nu, \tau)$ <sup>3</sup>.

La estimación de este modelo se hace por medio de la maximización del logaritmo de la función de máxima verosimilitud, a través del algoritmo RS explicado en la Sección 3.1.2. Para la parte suavizada del modelo se utilizaron tanto p-splines como splines cúbicos. En los splines cúbicos fue conveniente estimar los parámetros de suavización o grados efectivos de libertad, que se denotan por sus siglas en inglés (*df*:degree of freedom), fuera del modelo (ver Sección 3.1.3).

El total de grados efectivos de libertad en los splines combina los usados en las funciones de suavizado  $h_1(x)$ ,  $h_2(x)$ ,  $h_3(x)$  y  $h_4(x)$  para estimar los parámetros  $\mu$ ,  $\sigma$ ,  $\nu$  y  $\tau$ , respectivamente. Cada curva tendrá su propio *df*, dependiendo de la cantidad de suavización que precise. Los grados de libertad efectivos también pueden diferir de niños a niñas en caso de indicadores de crecimiento. Cuando los *df* son altos las curvas son más complejas y por tanto el suavizado no es tan efectivo. El objetivo es encontrar el menor *df* que aún proporcione un buen ajuste a la tendencia observada de los valores de  $L$ ,  $M$ ,  $S$  y  $P$  para cada edad.

Para establecer un criterio de selección de los *df* se utilizó el criterio GAIC, de forma que los *df* para cada parámetro se estimaron para minimizar el criterio GAIC(#) del ajuste. Estos grados de libertad efectivos se denotan por  $df_\mu$ ,  $df_\sigma$ ,  $df_\nu$  y  $df_\tau$ . Se utilizó un procedimiento automático, a través de la función `find.hyper` disponible en la librería GAMLSS de R, que proporciona valores óptimos de los grados de libertad que minimizan el GAIC (3.16). (Esto se explica en profundidad en la Sección 4.2 dentro del capítulo de software). Los valores óptimos resultantes de este procedimiento automático no incluyen la constante y el término lineal del modelo GAMLSS, por lo tanto se denotaron como  $(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$ . De esta forma, los grados efectivos de libertad para cada parámetro son  $df_{(\mu, \sigma, \nu, \tau)} = df'_{(\mu, \sigma, \nu, \tau)} + 2$ .

Una vez que se obtiene la estimación de los *df*, se utilizan para trazar curvas  $\mu$ ,  $\sigma$ ,  $\nu$  y  $\tau$  en función de la edad y encontrar los valores estimados de estos parámetros. Finalmente, estas estimaciones específicas por edad se utilizan para encontrar percentiles diferentes en cada edad usando la distribución correspondiente en cada modelo, como se ha descrito anteriormente. Se denota el modelo elegido como BCPE  $(df_\mu, df_\sigma, df_\nu, df_\tau)$ <sup>4</sup>.

---

<sup>3</sup>Para el método LMS,  $Y \sim BCCG(\mu, \sigma, \nu)$ , siendo BCCG la distribución Box Cox Normal, cuyas funciones *link* son la identidad para  $\mu$  y  $\nu$  y el logaritmo para  $\sigma$

<sup>4</sup>En caso de que la distribución BCPE sea la distribución adecuada.

## Capítulo 4

# Software utilizado

Para estimar las curvas percentiles de las variables antropométricas en niños y niñas gallegas de 6 a 15 años se aplicaron los modelos LMS y BCPE mediante el paquete `GAMLSS` de `R`. En este capítulo se describen las funciones utilizadas y sus argumentos. Se han utilizado las referencias [18], [21] y [22].

### 4.1. Selección de la distribución adecuada

Ya se explicó que en los modelos `GAMLSS` se asume una distribución de la variable respuesta que, puede pertenecer a diversas familias. Por lo tanto, el primer paso en la estimación percentil es seleccionar una distribución adecuada para  $Y$ . En el paquete `gamlss` de `R` están disponibles todas las familias de distribuciones bajo el nombre de `gamlss.family`.

Como paso previo al ajuste de los modelos mediante suavización por splines cúbicos, se ajustaron distintos modelos para cada variable (IMC, peso, talla, perímetro de cintura y perímetro de cadera) y sexo mediante p-splines con distintas familias de distribuciones, a través de las funciones `gamlss` y `pb()` de `R`. Las familias de distribuciones que se utilizaron en este paso fueron: la distribución `BCCG` (Box Cox Normal de Cole y Green) que se corresponde con el modelo LMS, la distribución `BCPE` que se corresponde con el modelo BCPE y la distribución `NO` (la distribución Normal) que se corresponde con un modelo normal. A continuación, en la Tabla 4.1, se hace un resumen de estas funciones.

Distribución	Nombre en R	Funciones <i>link</i>			
		$\mu$	$\sigma$	$\nu$	$\tau$
Normal	<code>NO()</code>	Identidad	Log	–	—
Box–Cox Cole Green	<code>BCCG()</code>	Identidad	Log	Identidad	–
Box–Cox Power Exponential	<code>BCPE()</code>	Identidad	Log	Identidad	Log

Tabla 4.1: Resumen de las distribuciones continuas que se utilizaron en este trabajo.

Este paso previo fue útil para seleccionar el tipo de distribución y, como consecuencia, qué modelo utilizar para cada variable y sexo de este trabajo. Una vez que se ajustaron los modelos mediante p-splines con las distintas distribuciones, se compararon entre sí a través de los criterios AIC y GAIC

con penalización  $k = 2.5$ . De forma que, el modelo adecuado para cada variable y sexo era aquel que minimizaba estos criterios. El código que se utilizó para este proceso se describe a continuación:

```
>gamlss(y~ pb(edad), sigma.fo = ~ pb(edad), data =Y,
family ="NO"),
>gamlss(y~ pb(edad), sigma.fo = ~ pb(edad), nu.fo= ~ pb(edad), data =Y,
family ="BCCG", "BCPE")),
>gamlss(y~ pb(edad), sigma.fo = ~ pb(edad), nu.fo= ~ pb(edad),
tau.fo= ~ pb(edad),data =Y, family = "BCPE").
```

donde  $y$ , de aquí en adelante, representa a cada una de las variables antropométricas que se utilizaron en este trabajo.

La razón por la que se utilizaron p-splines como tratamiento de los términos aditivos para la selección de la distribución adecuada, es que la función `pb()` calcula automáticamente los parámetros de suavización dentro del algoritmo que estima el modelo y, por lo tanto, los grados efectivos de libertad para cada caso, recortando significativamente el tiempo de convergencia de los modelos. Además, los p-splines en sí son mucho más rápidos, ya que la matriz de penalización es más pequeña que en el caso de los splines de suavizado, como se explicó en la Sección 3.1.6.

En los pasos siguientes de estimación de los hiperparámetros y ajuste de los modelos, los términos aditivos se modelizaron mediante splines cúbicos con las funciones `gamlss` y `cs()` de R.

## 4.2. Estimación de los grados de libertad

Una vez que ya se tiene la distribución adecuada para cada variable antropométrica y sexo, se puede ajustar un modelo apropiado para cada caso. Con el fin de obtener curvas confiables y comparables en su procedimiento con las de los estudios de referencia en ciencias de la salud, se decidió tratar los términos aditivos de los modelos mediante splines cúbicos, a través de la función `cs()`. Para esta función, a diferencia de `pb()`, es necesario estimar los grados efectivos de libertad para cada variable y sexo, utilizando la muestra, antes de ajustar el modelo. Este proceso es el que se va a detallar en esta sección.

Recuérdese que, los grados de libertad para el término aditivo suavizado se toman como grados de libertad extra, del ajuste lineal. En la función `cs()` cada valor de  $x$  distinto es un nodo.

La estimación de los parámetros de suavizado se hizo por medio de la función `find.hyper`. Esta función utiliza la función `optim` de R que minimiza el criterio GAIC con una penalización definida por el usuario.

Para el proceso automático de estimación que realiza la función `find.hyper`, se necesita partir de algunos valores iniciales de  $(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$ , es recomendable probar valores iniciales bajos, medios y altos, para asegurarse de que el proceso converge a un mínimo global de  $GAIC(\#)$  y no se queda en un mínimo local. Además, la función se puede aplicar con distintos valores de la penalización  $\#$ .

A continuación, se describe el proceso que automatiza la función `find.hyper` para el modelo BCPE, procedimiento que se sigue en 3 pasos [19]:

1. Valores iniciales.

- (a) Se seleccionan las funciones  $link\ g_k(\cdot)$ , para  $k = 1, 2, 3, 4$  en (3.31)
- (b) Se selecciona la penalización  $\#$  para cada valores de los grados efectivos de libertad utilizados en los modelos para  $\mu, \sigma, \nu$  y  $\tau$ .

## 2. Selección del modelo.

- (a) Partiendo del modelo BCPE(1,1,1,1), es decir, aquel cuyos parámetros de distribución se ajustan por medio de una constante, se aplica una selección hacia adelante de  $df'_\mu$  que proporciona un valor seleccionado,  $df'_{\mu_1}$  que minimiza el criterio GAIC( $\#$ ).
- (b) Ahora, a partir del modelo BCPE( $df'_{\mu_1}, 1, 1, 1$ ), se aplica, otra vez, el procedimiento hacia adelante para encontrar el valor efectivo de los grados de libertad para  $\sigma$ ,  $df'_{\sigma_1}$ , que minimiza GAIC( $\#$ ).
- (c) Partiendo del modelo BCPE( $df'_{\mu_1}, df'_{\sigma_1}, 1, 1$ ), se crea una tabla con valores del criterio GAIC( $\#$ ) para combinaciones de ( $df'_\nu$  y  $df'_\tau$ ) y se busca la combinación ( $df'_\nu$  y  $df'_\tau$ ) que minimice GAIC( $\#$ ).

## 3. Pulir el modelo BCPE( $df'_{\mu_1}, df'_{\sigma_1}, df'_\nu, df'_\tau$ )

- (a) Se afinan los grados de libertad de  $\sigma$ , cambiando  $df'_{\sigma_1}$ , de uno en uno, si el valor de GAIC( $\#$ ) decrece,
- (b) De manera similar se afina  $df'_{\mu_1}$ .

Los pasos 2 y 3 pueden ser automatizados por la función `find.hyper`, que se ha comentado antes, en el proceso de estimación de los hiperparámetros  $df_\mu, df_\sigma, df_\nu$  y  $df_\tau$ .

La función `find.hyper` se formula mediante el siguiente código:

```
find.hyper(model=, par=, lower=, steps=, factr =, parscale=, k=, method=),
```

cuyos argumetos se describen a continuación:

- **model**: este es el modelo GAMLSS que se quiere ajustar, dentro de `quote`. Esto se introduce de la siguiente manera:

```
quote(gamlss(y~cs(edad, df=p[1]), sigma.fo=~cs(edad, df=p[2]),
nu.fo=~cs(edad, df=p[3]))...data=datos_modelo)),
```

donde `p[i]` denota los parámetros de suavización a estimar y `cs()` los splines cúbicos por medio de los cuales se modelizan los términos aditivos.

- **par**: valores iniciales de los parámetros que se van a estimar. Recordamos que es conveniente tomar valores altos, medios y bajos. En este trabajo se han utilizado valores de 15, 6 y 2, para cada hiperparámetro, en tres búsquedas diferentes.
- **lower**: el valor más bajo que pueden alcanzar los hiperparámetros en la estimación. En este caso se utilizó 0.1 para todos los parámetros de suavización.

- **steps**: un vector de tamaños de paso para la aproximación de la diferencia finita al gradiente, en escala par / parscale. En este caso se tomó  $2e9$ .
- **parscale**: es un vector de valores de escala para los parámetros. La optimización se realiza en par / parscale y deben ser comparables en el sentido de que un cambio de unidad en cualquier elemento produce aproximadamente un cambio de unidad en el valor escalado. En este caso se ha utilizado un vector de unos para todos los parámetros.
- **k**: valor específico de la penalización que se utiliza para el criterio GAIC. En este trabajo se utilizaron penalizaciones de 2, 2.5, 3 y  $\log(n)$ . Nótese que, cuando  $\# = 2$  el criterio  $\text{GAIC}(\#)$  se corresponde con el criterio AIC (Criterio de Información de Akaike) y cuando  $\# = \log(n)$  con el criterio BIC (Criterio de Información Bayesiano). En general, el AIC proporcionó modelos con un número de df altos y curvas poco suaves, mientras que con el BIC los df fueron excesivamente bajos con un mal ajuste de los modelos a los datos. Por ello se consideraron las penalizaciones 2.5 y 3, que proporcionaron modelos adecuados, al igual que en los artículos [7] y [19].
- **method**: método de maximización de la log verosimilitud. Ya se han comentado los posibles métodos y el uso adecuado de cada uno de ellos. En este caso se utilizó `RS()`.

En resumen, para cada penalización ( $k = 2, 2.5, 3, \log(n)$ ) se realizaron tres búsquedas diferentes correspondientes a cada uno de los vectores de valores iniciales. Entonces, para cada penalización había tres resultados diferentes, seleccionándose aquel cuyo GAIC fuera inferior.

Por lo tanto, para cada variable antropométrica y sexo se obtenía generalmente, con la función `find.hyper`, cuatro valores diferentes para los hiperparámetros que, posteriormente se redondearon dando lugar a cuatro posibles modelos.

### 4.3. Ajuste y diagnosis de los modelos

Una vez que se estimaron y redondearon los parámetros de suavización óptimos, el siguiente paso fue ajustar los modelos, a través de la siguiente función<sup>1</sup>:

```
gamlss(y~cs(edad, df=), sigma.fo=~cs(edad, df=), nu.fo=~cs(edad, df=),
tau.fo = ~cs(idade_meses, df=), family = BCPE, data=datos_modelo)
```

Para cada uno de los modelos que se ajustaron, se hizo un análisis de los residuos (diagnosis). En esta sección se describe el proceso de diagnosis que se siguió en el trabajo.

Primeramente, conviene describir como son los residuos de los modelos GAMLSS, también llamados residuos cuantiles randomizados.

**Definición 4.1.** Un residuo cuantil randomizado  $\hat{r}$  (o  $z$ -score) para una observación  $y$  con parámetros ajustados  $(\hat{\mu}, \hat{\sigma}, \hat{\nu}, \hat{\tau})$  viene dado por la siguiente expresión

$$\hat{r} = \Phi^{-1} \left( F(y|\hat{\theta}_k) \right),$$

---

<sup>1</sup>En el caso de que la distribución seleccionada fuese la BCPE. Ya se ha comentado que, en el caso de que sea la distribución BCCG solo habría 3 parámetros y en el caso de la normal 2.



donde  $\Phi$  representa la función de distribución de una normal estándar y  $F(y|\hat{\theta}_k)$  la función de distribución (ajustada) para  $y$ .

La característica principal de estos residuos es que, sea cual sea la distribución elegida para la variable respuesta, siempre siguen una distribución normal estándar, siempre y cuando el modelo sea correcto. Por lo que la finalidad de este proceso de diagnóstico es comprobar que el modelo es correcto y los residuos siguen una distribución normal.

En primer lugar, para cada modelo ajustado fue posible obtener un resumen de los residuos por medio de la función `plot()` que proporcionaba la media, la varianza, el coeficiente de asimetría y el coeficiente de curtosis de estos.

Por otro lado, para comprobar que los modelos se ajustaban bien a los datos y que se cumplía la normalidad de los residuos, se hizo una diagnosis de cada uno de ellos. Por otro lado, además de comprobar la validez de los modelos, también sirvió para comparar los modelos de cada variable y sexo entre sí y seleccionar el mejor.

Para este proceso, se han utilizado las funciones `wp()` (worm plot) y `Q.stats()` de la librería `gamlss` de R. A continuación describe en que consiste cada una de estas funciones [22].

- El *worm plot* (“diagrama de gusano”) es una herramienta de diagnosis, introducida por Buuren y Fredrils (2001), consta de una colección de gráficos QQ-plot sin tendencia<sup>2</sup>, los cuales consisten en visualizar diferencias entre dos distribuciones condicionadas a los valores de una covariable (en este caso, la edad). En este caso, se basan en contrastar la hipótesis nula de normalidad de los residuos para cada modelo.

La Figura 4.1 muestra un ejemplo de un único gráfico que da la salida de la función `wp()`, es decir un único QQ-plot sin tendencia. A través de esta figura, se van a explicar los elementos componen el gráfico:

- En el eje horizontal, se representan los cuantiles teóricos de una normal  $N(0, 1)$  y en el eje vertical la diferencia entre los cuantiles teóricos y empíricos basados en los residuos del modelo. Por lo tanto, los puntos que forman un gusano, representan cuanto difieren los residuos de cada modelo de una distribución normal  $N(0, 1)$ . En resumen, un gusano plano, más o menos distribuido de forma uniforme sobre la línea discontinua roja (horizontal)  $y = 0$ , en un determinado grupo de edad, indica que los datos siguen una distribución normal estándar y por tanto que el modelo es adecuado para ese intervalo de edad.
- Las líneas negras discontinuas representan un intervalo de confianza del 95 % de los cuantiles de la distribución normal estándar. Para un cuantil  $z$  dado con probabilidad asociada  $p$  y un tamaño de muestra  $n$ , el intervalo de confianza que se acaba de mencionar se contruye como:

$$\pm 1.96 f^{-1}(z) \sqrt{\frac{p(1-p)}{n}},$$

---

<sup>2</sup> En inglés “detrended”, que en estadística hace referencia a eliminar tendencia a largo plazo, para enfatizarla a corto plazo.

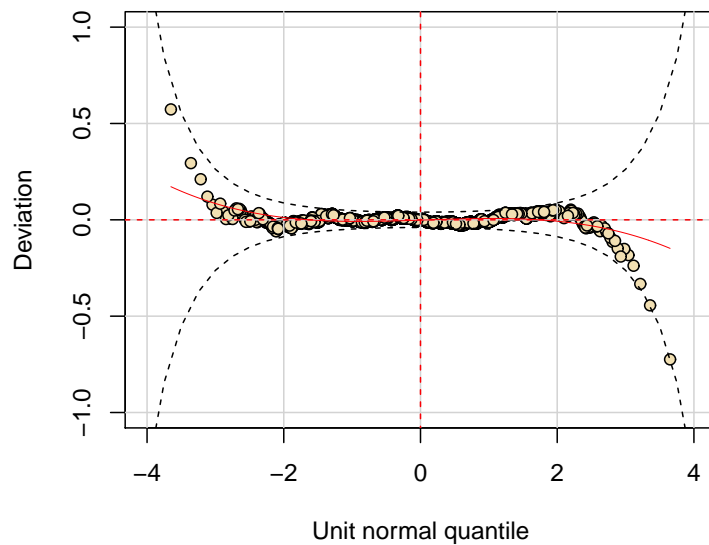


Figura 4.1: Ejemplo de salida obtenida con la función `wp()`.

siendo  $f(z)$  la función de densidad de la  $N(0, 1)$ . El intervalo de confianza da una impresión de la variación del muestreo y delinea la región donde se ubicará el gusano la mayor parte del tiempo si las distribuciones teórica y empírica concuerdan.

- La curva continua roja representa un ajuste cúbico que adaptaron Van Buuren y Fredricks (2002) para cuantificar las características básicas de la forma del gusano. Estos modelos cúbicos se componen de los coeficientes: constante, lineal, cuadrático y cúbico representados, respectivamente, por  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ ,  $\hat{\beta}_2$  y  $\hat{\beta}_3$  que indican diferencias entre media, varianza, asimetría y curtosis de las distribuciones empírica y teórica en función del rango de edad. Para el diagnóstico del modelo se categorizan como valores inadecuados aquellos que verifiquen

$$\left| \hat{\beta}_i \right| > 0.10 \text{ con } i = 1, 2, \quad \left| \hat{\beta}_2 \right| > 0.05 \quad \text{y} \quad \left| \hat{\beta}_3 \right| > 0.03.$$

Estos valores inadecuados indican diferentes problemas que se reflejan en el modelo como se indica en la Tabla 4.2.

El worm plot es un método gráfico, simple y flexible para respaldar las decisiones que se toman sobre el modelo, más concretamente decisiones sobre la suavidad que, como ya se ha mencionado está estrechamente relacionada con los grados de libertad que se escogen.

La función `wp()` puede tener como salida un único gráfico si se deja el argumento `xvar=NULL` que trae la propia función por defecto, o varios gráficos si a ese argumento se le da una de las variables explicativas. En este caso se tiene una gráfica por cada uno de los intervalos que le indiquemos en el argumento `n.iter=""`.

Los QQ-plot sin tendencia se organizan en forma de una cuadrícula, por lo que, por

ejemplo, las opciones  $3 \times 3$  o  $4 \times 4$  producen respectivamente 9 o 16 intervalos. Estos intervalos son contiguos y no se solapan.

Para la elección del número de intervalos, se recomienda que haya al menos 200 o 300 puntos por grupo. Nótese que, a mayor grupos de edad se tiene una gráfica más detallada pero menos estable. En la referencia [wp], se recomienda no utilizar 9 grupos de edad ya que ocultan importantes desviaciones de la normalidad y se establece que 16 grupos es una buena opción para la diagnosis.

Momento	Forma del wp	Residuos	Variable respuesta
Media	Puntos por encima de 0	Media alta	Parámetro de localización bajo
	Puntos por debajo de 0	Media baja	Parámetro de localización alto
Varianza	Pendiente positiva	Varianza alta	Parámetro de escala bajo
	Pendiente negativa	Varianza baja	Parámetro de escala alto
Asimetría	Forma de U	Asimetría positiva	Asimetría baja
	Forma de U inversa	Asimetría negativa	Asimetría alta
Curtosis	Forma de S girada a la izquierda	Leptocúrtica	Curtosis baja
	Forma de S girada a la izquierda y dada la vuelta	Platicúrtica	Curtosis alta

Tabla 4.2: Problemas detectados a través del worm plot.

En este trabajo se realizaron los gráficos worm plot con 16 grupos de edad en meses, cuyo rango va desde 68.5 a 191.5 meses. En los gráficos worm plot, los intervalos se representan de izquierda a derecha y de abajo arriba. Siguiendo esta organización, los diagramas se numeraron del 1 al 16, correspondiéndose con los intervalos de edad, para facilitar así el análisis de estos gráficos.

- El  $Q.stats$  permite investigar más a fondo el ajuste dentro de los grupos de edad, apoyándose en el cálculo de los estadísticos  $Z$  y  $Q$ , que se definen a continuación, para evaluar la normalidad de los residuos para cada intervalo de edad (Royston y Wright (2000)).

En primer lugar, se define  $G$  como el número de grupos de edad y  $\{r_{gi}, i=1, \dots, n_i\}$  los residuos en el intervalo de edad  $g$ , con media  $\bar{r}_g$  y desviación estándar  $s_g$ , para  $g = 1, \dots, G$ .

El estadístico  $Q$  se basa en 4 estadísticos  $Z_{gj}$ , para  $j = 1, 2, 3, 4$  que, en cada grupo de edad  $g$ , contrastan respectivamente las hipótesis nulas:

$$H_{g,1} : \mu = 0$$

$$H_{g,2} : \sigma = 1$$

$$H_{g,3} : \nu = 1$$

$$H_{g,4} : \tau = 2,$$

es decir, se contrasta si los residuos tienen distribución  $N(0, 1)$ .

A partir de los valores de  $Z_{gj}$  se calculan los estadísticos  $Q_j$  de la siguiente forma:

$$Q_j = \sum_{g=1}^G Z_{gj}^2,$$

para  $j = 1, 2, 3, 4$ . Bajo la hipótesis nula de que los residuos siguen una distribución  $N(0, 1)$ , Royston y Wright (2000) sugieren que  $Q_j$  sigue aproximadamente una distribución Ji-cuadrado con grados de libertad

$$\begin{aligned} g_1 &= G - df_\mu \\ g_2 &= G - \frac{(df_\sigma + 1)}{2} \\ g_3 &= G - df_\nu \\ g_4 &= G - df_\tau, \end{aligned}$$

para  $Q_1$ ,  $Q_2$ ,  $Q_3$  y  $Q_4$ , respectivamente. Cabe destacar que, los grados de libertad para el estadístico  $Q_4$  fueron propuestos en la referencia [3].

En resumen, valores significativos de  $Q_1$ ,  $Q_2$ ,  $Q_3$  o  $Q_4$  indican insuficiencias en el modelo para los parámetros  $\mu$ ,  $\sigma$ ,  $\nu$  y  $\tau$ , respectivamente. Estas insuficiencias se pueden corregir aumentando los grados de libertad con los que se ha ajustado el parámetro correspondiente.

El estadístico  $Z_{gj}$  proporciona la contribución del grupo  $g$  al  $Q_j$ , por lo tanto, ayuda a identificar que grupos de edad están haciendo que  $Q_j$  sea significativo y con ello, en que grupos de edad el modelo es inadecuado. Se considera un indicativo de deficiencia del modelo si en un grupo de edad  $g$  se cumple que  $|Z_{gj}| > 2$ , para  $i = 1, 2, 3, 4$ .

La función `Q.stats` de R, nos da el valor de estos estadísticos para cada grupo de edad y el tamaño de la muestra que hay en cada uno de esos grupos. Esta función también nos devuelve el valor del estadístico del test de normalidad de Agostino,  $K^2$ .

En resumen, el apartado de la diagnosis permite seleccionar aquellos modelos que mejor se ajustan a cada variable y descartar los modelos que posean irregularidades. Tras elegir el modelo que mejor funciona, el siguiente paso fue dibujar las curvas percentiles.

## 4.4. Curvas percentiles

Las curvas percentiles de cada modelo para cada variable, se dibujaron mediante la función `centiles` del paquete de R.

Además, es de gran interés la interpretación gráfica de los parámetros  $\mu$ ,  $\sigma$ ,  $\nu$  y  $\tau$  ajustados, de la distribución utilizada frente a la edad. Los parámetros ajustados  $\mu$  y  $\sigma$  dan información sobre como evolucionan la mediana y el coeficiente de variación con la edad. La interpretación de la asimetría para cada edad se basa en el signo de  $\nu$ , es decir, cuando  $\nu < 0$  ( $\nu > 0$ ) se dice que se tienen asimetría positiva (negativa) y la interpretación de la curtosis (en caso de utilizar el método BCPE) se basa en si  $\tau$  está por encima o por debajo de 2, es decir, si  $\tau < 2$  ( $\tau > 2$ ) se dice que la curtosis es la de una distribución leptocúrtica (paticúrtica).

Para esta interpretación gráfica de los parámetros ajustados se utilizó la función `fitted.plot()`.

En la siguiente sección, que muestra los resultados obtenidos para cada variable, se siguió el mismo esquema de 4 pasos de esta sección.

## Capítulo 5

# Estimación de los modelos

En esta sección, se describe el proceso de selección del modelo más adecuado para el IMC, peso, talla, perímetro de cintura y perímetro de cadera frente a la edad, que incluye ajuste y diagnóstico. Se detalla en mayor medida el proceso para el IMC, ya que para el resto se procedió de manera similar.

### 5.1. IMC en niños

Se identificaron 9 atípicos extremos en el IMC en los niños que se eliminaron de la muestra. En la Tabla 5.1 se muestran los resultados del ajuste de los modelos con p-splines y distribuciones BCCG, BCPE y Normal, organizados de menor a mayor GAIC con penalización 2.5. Por lo tanto, para este caso la distribución BCPE es la adecuada.

Distribución	GAIC
BCPE	18899.55
BCCG	18943.77
NO	19654.24

Tabla 5.1: Resumen de los valores de GAIC, con penalización 2.5, para el ajuste mediante p-splines en el IMC en niños.

El siguiente paso fue estimar los grados de libertad para cada parámetro  $(df_\mu, df_\sigma, df_\nu, df_\tau)$ , por medio de la función `find.hyper`, comentada en la Sección 4.2. Los resultados de los valores  $(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$  se muestran en la Tabla 5.2. Como ya se explicó, la función se aplicó con distintas penalizaciones para el criterio GAIC y distintos valores iniciales de los grados de libertad, para asegurar la convergencia a un mínimo global y no local. Una primera comparación de las tablas 5.1 y 5.2 indica valores ligeramente inferiores del GAIC(2.5) usando `cs()` frente a `pb()`. Esta pauta se observó a lo largo del estudio.

Nótese que, en la penalización  $k = \log(n)$  la estimación de los grados de libertad para la media alcanzó el límite inferior (0.1). Cuando ocurre esto, se debe realizar una búsqueda adicional para verificar si el parámetro correspondiente se puede simplificar más a un término constante o simplemente

Penalización(#)	Valores iniciales	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	GAIC(#)
2	(15,15,15,15)	(4.68, 14.52, 1.50, 2.60)	18886.49
	(6,6,6,6)	(4.67, 24.68, 1.51, 2.66)	18885.74
	(2,2,2,2)	(4.72, 23.89, 1.51, 2.73)	18885.71
2.5	(15,15,15,15)	(4.03, 6.28, 1.44, 1.96)	18898.57
	(6,6,6,6)	(4.04, 5.64, 1.37, 1.88)	18898.56
	(2,2,2,2)	(3.90, 2.26, 1.40, 1.81)	18898.44
3	(15,15,15,5)	(3.17, 2.09, 1.30, 1.57)	18906.88
	(6,6,6,6)	(3.19, 2.05, 1.29, 1.58)	18906.88
	(2,2,2,2)	(1.44, 2.12, 1.33, 1.54)	18906.79
$\log(n) = 8.2516$	(15,15,15,5)	(0.1, 1.32, 0.17, 0.81)	18969.67
	(6,6,6,6)	(0.1, 1.32, 0.16, 0.81)	18969.68
	(2,2,2,2)	(0.1, 1.32, 0.16, 0.81)	18969.68

Tabla 5.2: Resultados de la función `find.hyper` para el ajuste del IMC en los niños.

a un término lineal, es decir,  $df = 1$  o  $df = 2$ , respectivamente. Para seleccionar un caso u otro, se utiliza el criterio del mínimo valor de GAIC.

En este caso, fue necesario hacer dos búsquedas adicionales en dos etapas. En primer lugar, se hizo una búsqueda para comprobar si la media se ajustaba como una constante o un término lineal,  $df_\mu = 1$  o  $df_\mu = 2$ , respectivamente. Los resultados indicaron que  $\mu$  debía aproximarse por un término lineal ( $df_\mu = 2$ ), además la estimación de los grados de libertad para  $\nu$  alcanzó el límite inferior. Por ello fue necesario realizar otra búsqueda para comprobar si el coeficiente de asimetría se podía aproximar por una constante o por un término lineal. En la Tabla 5.3, se muestran los resultados.

Primera etapa			
Penalización(#)	$df_\mu$	$(df'_\sigma, df'_\nu, df'_\tau)$	GAIC(#)
8.2516	2	(1.3, 0.1, 0.82)	18969.18
	1	(0.1, 0.1, 0.18)	19901.13
Segunda etapa			
Penalización(#)	$(df_\mu, df_\nu)$	$(df'_\sigma, df'_\tau)$	GAIC(#)
8.2516	(2,2)	(1.29, 0.8)	18969.24
	(2,1)	(1.28, 0.8)	18961.59

Tabla 5.3: Resultados de las búsquedas adicionales para  $\mu$  (primera etapa) y para  $\nu$  (segunda etapa), por haber alcanzado el límite inferior (0.1) en el IMC de los niños. Se subraya en amarillo el valor mínimo del criterio GAIC.

En total se obtuvieron 4 vectores diferentes para los valores de los hiperparámetros, uno para cada penalización. Estos valores se redondearon para ajustar los modelos, en la Tabla 5.4) se pueden ver, acompañados de la penalización de la que provienen y los valores estadísticos que caracterizan la

distribución de los residuos para cada modelo ajustado. Recuérdese que la distribución de los residuos debe aproximarse lo máximo posible a la de una  $N(0,1)$  que, como ya se sabe, tiene coeficiente de asimetría cero y el valor de la curtosis 3.

			Resumen de los residuos			
Penalización	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	$(df_\mu, df_\sigma, df_\nu, df_\tau)$	Media	Varianza	Coef.Asim.	Coef.Curt.
2	(5,24,1,3)	(7,26,3,5)	-0.0062	1.0008	0.0144	2.9714
2.5	(4,2,1,2)	(6,4,3,4)	-0.0045	0.9997	0.0130	2.9823
3	(1,2,1,1)	(3,4,3,3)	-0.0042	1.0004	0.0113	2.9821
8.25	(0,1,0,1)	(2,3,1,3)	-0.0022	0.9987	0.0158	2.9856

Tabla 5.4: Tabla de resumen residual.

El siguiente paso fue hacer un análisis de residuos para cada modelo, para comprobar se ajustaban bien y, además, elegir el más adecuado. Para ello se utilizaron los gráficos `worm plot` y los valores de los estadísticos obtenidos con la función `Q.stats`, comentados en la Sección 4.3.

En las Figuras 5.1, 5.2 y 5.3 se representan los gráficos worm plot para los modelos BCPE (7,26,3,5), BCPE (6,4,3,4) y BCPE (3,4,3,3), respectivamente. Se omitieron los resultados para el modelo BCPE (2,3,1,3), ya que son similares que los del modelo BCPE (3,4,3,3).

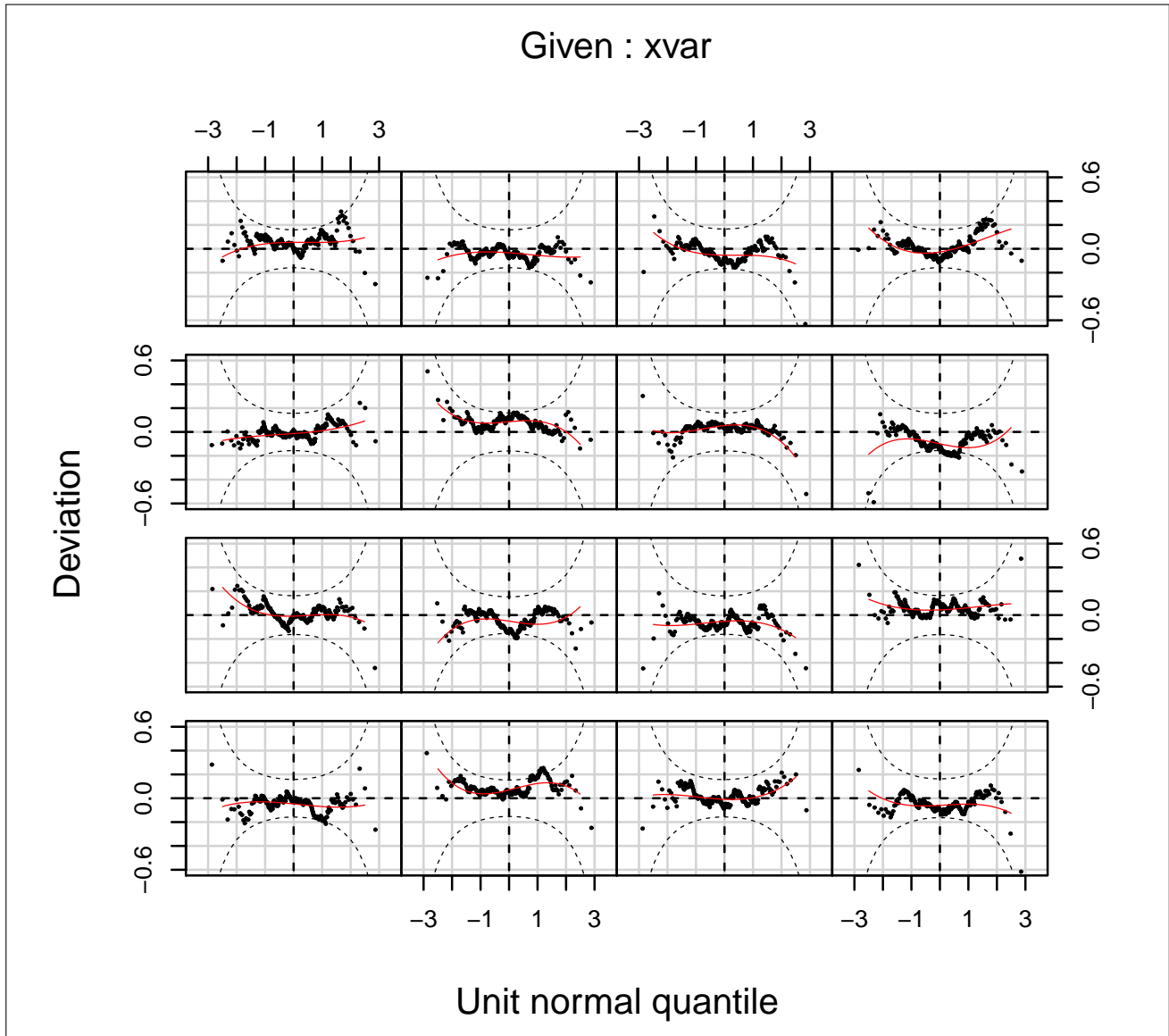


Figura 5.1: Gráfico worm plot para el modelo BCPE (7,26,3,5) que ajusta el IMC en los niños.

Como ya se ha comentado, la forma de la cadena de puntos describe un aspecto diferente (media, varianza, asimetría y curtosis) del ajuste del modelo.

En la Figura 5.1, se perciben posibles desajustes leves para la media y la curtosis. Para la media, se aprecia un desplazamiento por encima de  $y = 0$  en los intervalos 2 y 10 indicando que, quizás para esos grupos de edad la media ajustada es muy pequeña. Para la curtosis, también en los intervalos 2 y 10 se aprecia que quizás la curtosis ajustada es demasiado pequeña, ya que se ve que la forma de S es doblada hacia abajo y para los intervalos 6 y 12 se aprecia forma de S doblada hacia arriba, por lo



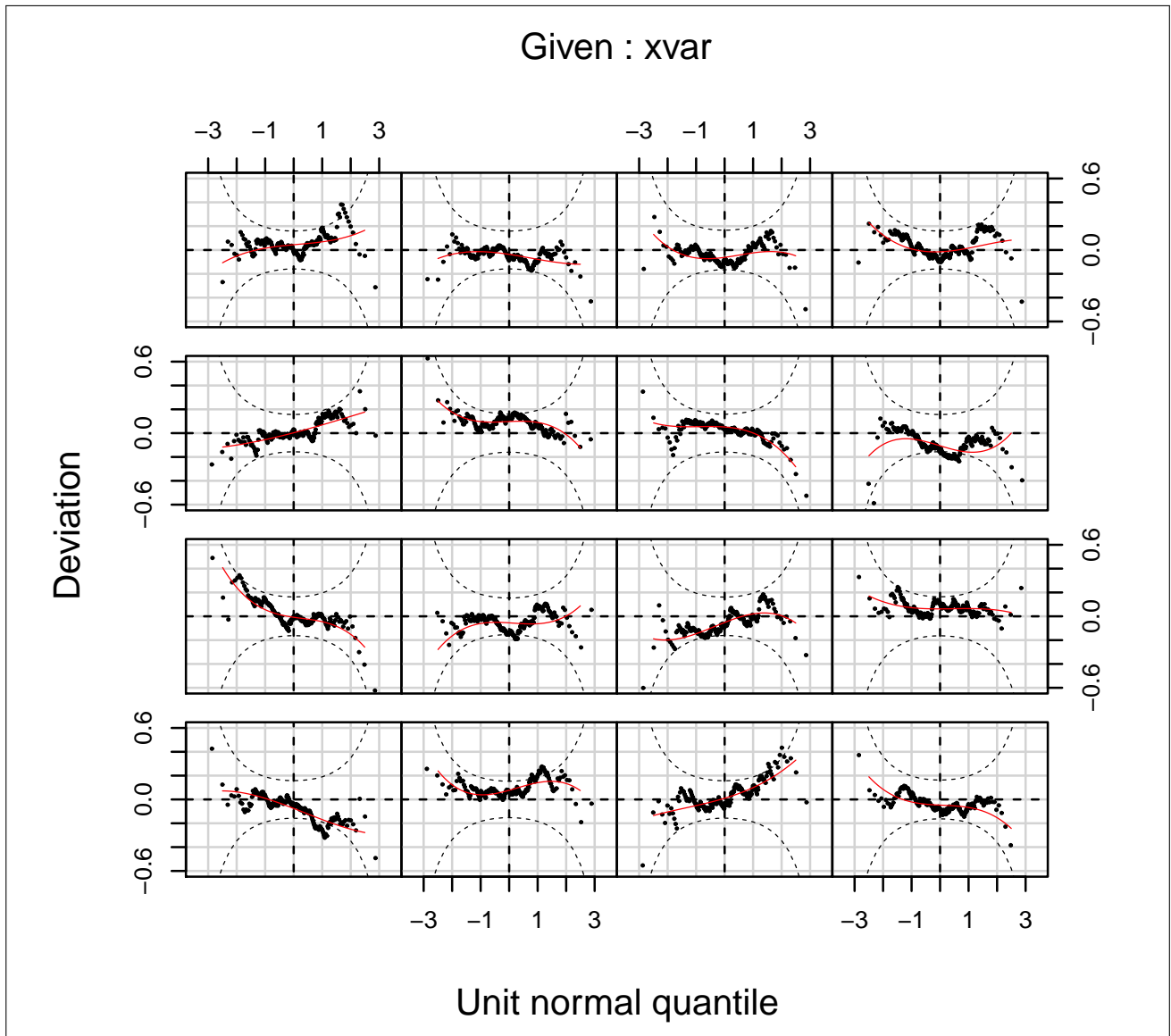


Figura 5.2: Gráfico worm plot para el modelo BCPE (6,4,3,4) que ajuste el IMC en los niños.

tanto, puede que la curtosis sea grande para ellos.

La Figura 5.2, además de para la media y la curtosis, muestra posibilidad de insuficiencia para la varianza del modelo ajustado. Para la media, en los intervalos 1, 2 y 10 hay un pequeño desplazamiento sobre el origen, indicando que a lo mejor la media es muy pequeña para los dos primeros y muy grande para los dos últimos. Para la varianza, se percibe tendencia negativa en los intervalos 1 y 5, y tendencia positiva para el grupo de edad 3, indicando que la varianza ajustada puede que sea demasiado grande para los primeros intervalos y demasiado pequeña para el segundo. Finalmente, para la curtosis los intervalos 2, 6 y 12 exponen forma de ese doblada hacia arriba o hacia abajo respecto al origen.

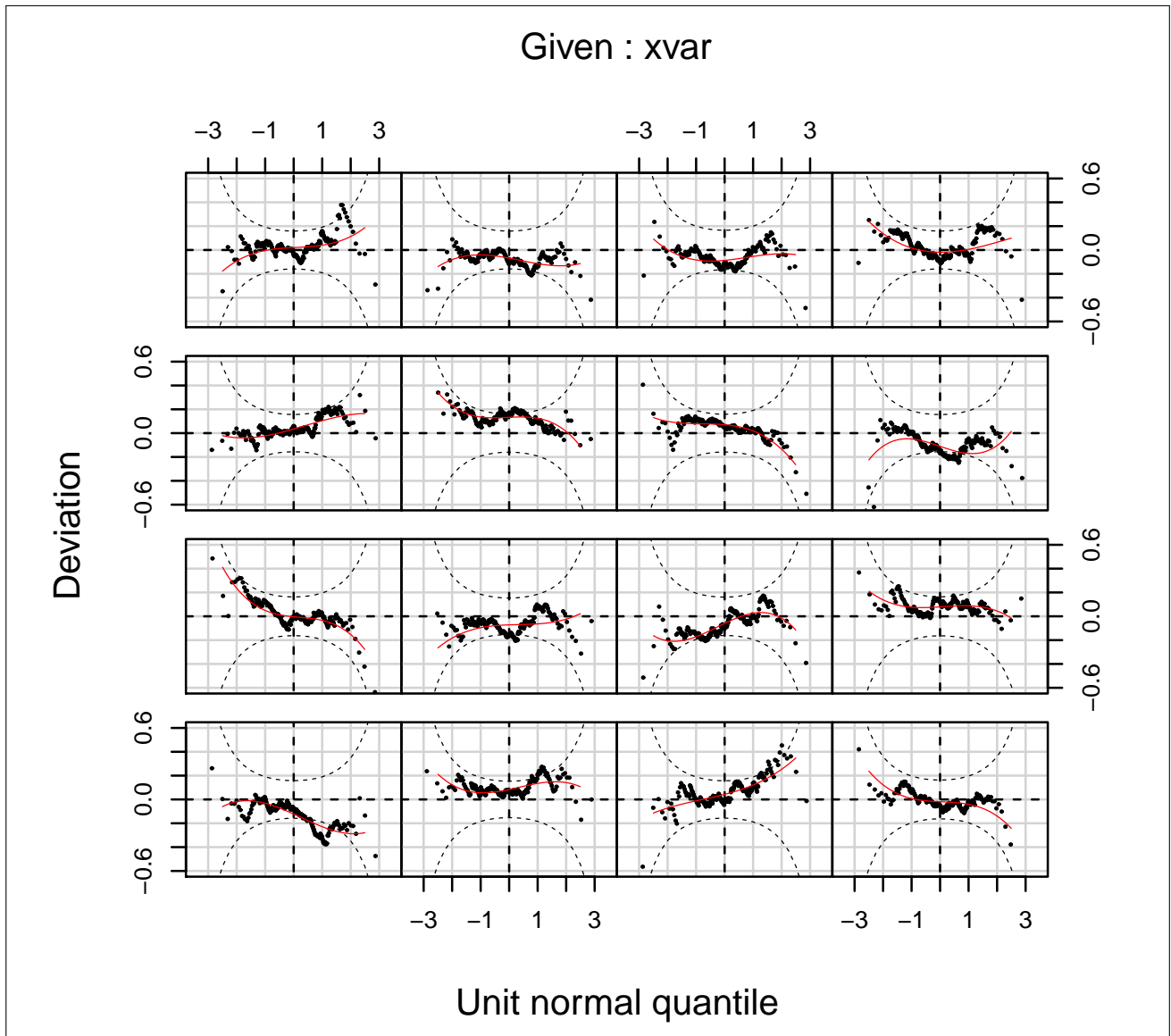


Figura 5.3: Gráfico worm plot para el modelo ajustado BCPE (3,4,3,3) que ajusta el IMC en los niños.

La Figura 5.3, presenta irregularidades más notorias que la Figura 5.1. Para la media, los intervalos 1, 10 y 12 presentan cadenas de puntos notoriamente desplazados por encima o por debajo del origen, indicando que la media es muy pequeña o muy grande, respectivamente, para esos grupos de edad. Se aprecia tendencia negativa en los intervalos de edad 1 y 5 y tendencia positiva en 3 y 7, indicando que la varianza ajustada puede que sea demasiado grande y pequeña, respectivamente. Finalmente, las posibles insuficiencias para la curtosis se manifiestan en los intervalos 5 y 12.

Para profundizar en estas posibles insuficiencias, se utilizó la función `Q.stats`, que como ya se comentó en la Sección 4.3, cuantifica las características básicas de la forma que adopta el “gusano” de los gráficos worm plot, para cada intervalo de edad. En las Tablas 5.5, 5.6 y 5.7 se presentan los valores de estos estadísticos  $Q_j$  y  $Z_{gj}$  para los modelos BCPE(7,26,3,5), BCPE (6,4,3,4) y BCPE (3,4,3,3). Nuevamente, se omiten los resultados para el modelo BCPE (2,3,1,3), ya que son similares al modelo

BCPE (3,4,3,3).

n°	Intervalos de edad	Valores $Z_{gj}$				Agostino( $K^2$ )	N
		$\mu$	$\sigma$	$\nu$	$\tau$		
1	68.5 to 75.5	-0.78	-0.23	-0.08	0.36	0.14	243
2	75.5 to 82.5	1.30	0.29	0.48	-1.17	1.60	259
3	82.5 to 90.5	0.13	0.20	0.65	0.45	0.62	246
4	90.5 to 99.5	-0.83	-0.19	0.24	-0.33	0.17	228
5	99.5 to 108.5	0.10	-0.47	0.58	-0.66	0.78	229
6	108.5 to 117.5	-0.89	0.28	-0.46	1.51	2.48	257
7	117.5 to 124.5	-0.99	-0.02	-0.47	-0.19	0.25	233
8	124.5 to 131.5	0.82	0.05	0.41	-0.02	0.17	226
9	131.5 to 138.5	-0.15	0.54	0.17	0.14	0.05	247
10	138.5 to 145.5	1.27	-0.43	-0.15	-1.05	1.12	237
11	145.5 to 152.5	0.45	-0.15	-0.90	-0.57	1.13	241
12	152.5 to 160.5	-1.51	-0.07	-0.07	1.36	1.84	249
13	160.5 to 167.5	0.73	0.39	-0.32	0.80	0.74	239
14	167.5 to 174.5	-0.64	-0.15	-0.28	0.28	0.16	240
15	174.5 to 181.5	-0.66	-0.52	0.43	-0.32	0.28	220
16	181.5 to 191.5	0.07	0.45	1.17	-0.37	1.50	240
	Estadístico Q	10.96	1.68	4.25	8.79	13.04	3834
	df para Q	9	2.50	13	11	24	
	p-val para Q	0.28	0.54	0.99	0.64	0.97	

Tabla 5.5: Tabla de estadísticos  $Q$  y  $Z$  para el modelo BCPE (7,26,3,5) del IMC de los niños.

En la Tabla 5.5, se confirma que los desvíos que se apreciaron para el modelo BCPE (7,26,3,5), en la Figura 5.1, en determinados grupos de edad para la media y la curtosis no son grandes, ya que los valores de los  $Z_{gj}$  que se muestran en la tabla, son inferiores en valor absoluto a dos. Además, los  $Q_j$  con  $j = 1, 2, 3, 4$ , son todos no significativos, indicando que el modelo es válido.

En la Tabla 5.6, como en el caso anterior, las supuestas insuficiencias que se apreciaron para el modelo BCPE (6,4,3,4), en la Figura 5.2, en determinados grupos de edad para la media, la varianza y la curtosis no son significativas. Los estadísticos  $Q_j$  tampoco son significativos, indicando que el modelo se ajusta bien para cualquier parámetro.

Por último, a diferencia de los modelos anteriores, la Tabla 5.7 verifica que existen irregularidades significativas para la media ( $j = 1$ ) en el modelo BCPE (3,4,3,3). Para los intervalos de edad  $g = 1$  y  $g = 10$  se verifica que  $|Z_{g1}| > 2$ , lo que indica que el modelo no se ajusta bien para la media en esos grupos de edad. Además, el p-valor para el estadístico  $Q_1$  es significativo, por tanto se concluye el modelo no se ajusta bien para la media. Para el resto de parámetros las aparentes insuficiencias no son significativas y, por lo tanto, el modelo se ajusta bien para ellos.

Una vez hecho este análisis, se descartaron los modelo BCPE (3,4,3,3) y BCPE (2,3,1,3) del proceso de selección del mejor ajuste, ya que presentaban resultados similares con insuficiencias para la media.

n°	Intervalos de edad	Valores $Z_{gj}$				Agostino( $K^2$ )	N
		$\mu$	$\sigma$	$\nu$	$\tau$		
1	68.5 to 75.5	-1.24	-1.81	-0.17	0.34	0.14	243
2	75.5 to 82.5	1.44	0.47	0.53	-1.18	1.66	259
3	82.5 to 90.5	0.37	1.81	0.48	0.30	0.32	246
4	90.5 to 99.5	-0.67	-0.97	0.21	-0.62	0.42	228
5	99.5 to 108.5	0.08	-1.71	0.54	-1.08	1.46	229
6	108.5 to 117.5	-0.99	0.67	-0.53	1.52	2.58	257
7	117.5 to 124.5	-0.96	1.36	-0.36	-0.50	0.38	233
8	124.5 to 131.5	1.03	-0.26	0.21	-0.30	0.13	226
9	131.5 to 138.5	0.12	1.33	0.18	-0.05	0.03	247
10	138.5 to 145.5	1.44	-0.66	-0.19	-1.02	1.08	237
11	145.5 to 152.5	0.43	-0.95	-0.93	-0.45	1.07	241
12	152.5 to 160.5	-1.64	-0.39	-0.19	1.56	2.47	249
13	160.5 to 167.5	0.64	0.85	-0.10	0.72	0.54	239
14	167.5 to 174.5	-0.72	-0.51	-0.35	0.37	0.26	240
15	174.5 to 181.5	-0.67	0.04	0.65	-0.57	0.75	220
16	181.5 to 191.5	0.19	-0.13	1.09	-0.26	1.26	240
	Estadístico Q	13.57	17.29	4.04	10.51	14.55	3834
	df para Q	10	13.50	13	12	25	
	p-val para Q	0.19	0.21	0.99	0.57	0.95	

Tabla 5.6: Tabla de estadísticos  $Q$  y  $Z$  para el modelo BCPE (6,4,3,4) del IMC de los niños.

n°	Intervalos de edad	Valores $Z_{gj}$				Agostino( $K^2$ )	N
		$\mu$	$\sigma$	$\nu$	$\tau$		
1	68.5 to 75.5	-2.03	-1.72	-0.29	0.75	0.65	243
2	75.5 to 82.5	1.53	0.40	0.51	-0.82	0.94	259
3	82.5 to 90.5	0.86	1.68	0.40	0.42	0.33	246
4	90.5 to 99.5	-0.25	-1.06	0.19	-0.73	0.56	228
5	99.5 to 108.5	0.11	-1.62	0.47	-1.31	1.93	229
6	108.5 to 117.5	-1.28	0.77	-0.49	1.08	1.42	257
7	117.5 to 124.5	-1.11	1.44	-0.37	-0.95	1.04	233
8	124.5 to 131.5	1.24	-0.26	0.08	-0.70	0.49	226
9	131.5 to 138.5	0.64	1.18	0.24	-0.34	0.18	247
10	138.5 to 145.5	2.00	-0.77	-0.16	-1.18	1.42	237
11	145.5 to 152.5	0.75	-1.05	-0.86	-0.48	0.98	241
12	152.5 to 160.5	-1.73	-0.38	-0.26	1.75	3.14	249
13	160.5 to 167.5	0.27	0.94	-0.14	0.99	0.99	239
14	167.5 to 174.5	-1.21	-0.46	-0.36	0.60	0.50	240
15	174.5 to 181.5	-1.02	0.03	0.70	-0.37	0.63	220
16	181.5 to 191.5	0.17	-0.25	1.19	-0.11	1.44	240
	TOTAL Q stats	22.25	16.84	4.11	12.52	16.63	3834
	df for Q stats	13	13.50	13	13	26	
	p-val for Q stats	0.05	0.24	0.99	0.49	0.92	

Tabla 5.7: Tabla de estadísticos  $Q$  y  $Z$  para el modelo BCPE (3,4,3,3) del IMC de los niños

Por lo tanto, se dibujaron las curvas percentiles del IMC frente a la edad (Figura 5.4) para los modelos BCPE (7,26,3,5) y BCPE (6,4,3,4) que han resultado válidos.

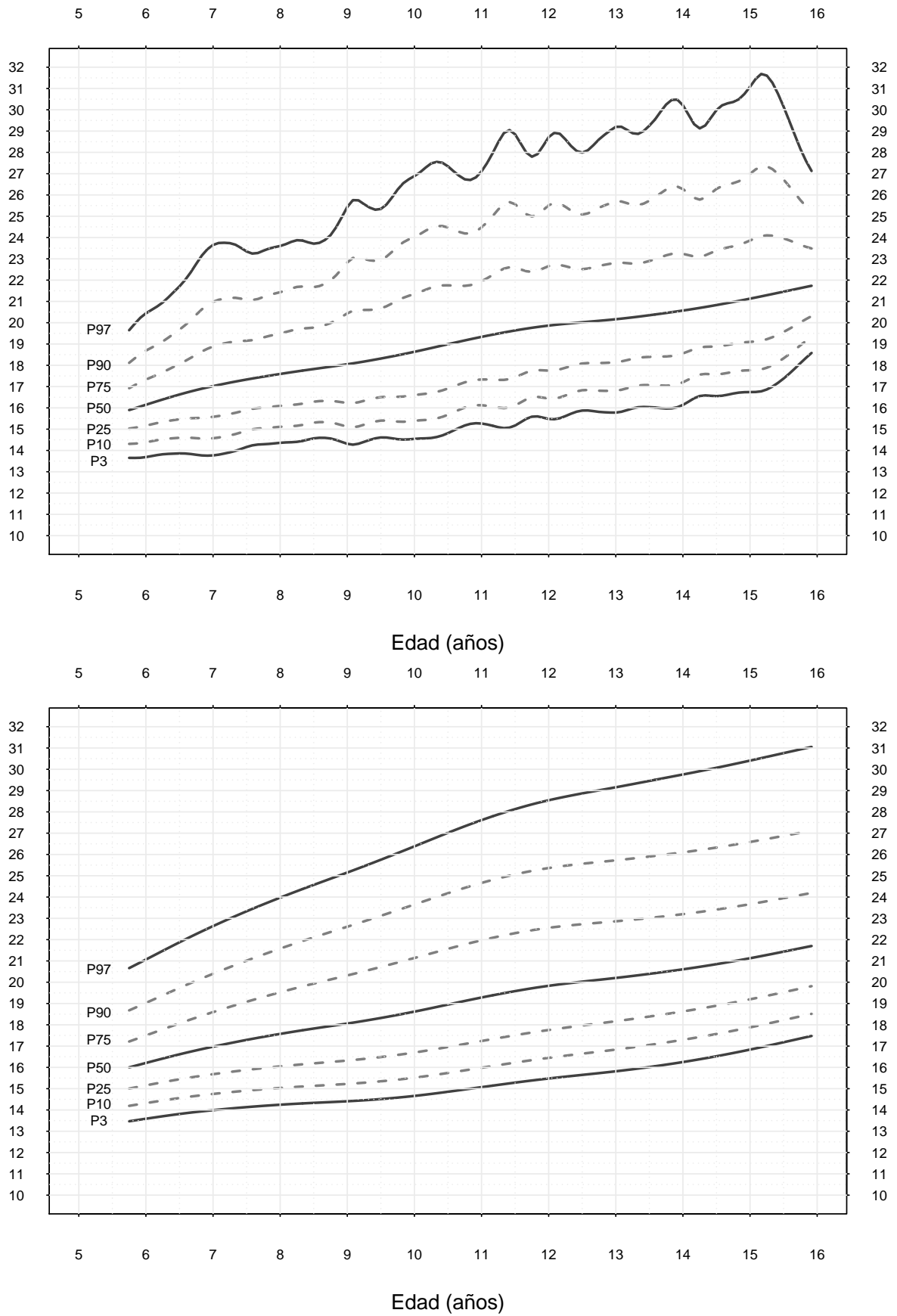


Figura 5.4: Curvas percentiles para el IMC ( $\text{kg}/\text{m}^2$ ) de los niños con el modelo BCPE (7,26,3,5) (arriba) y con el modelo BCPE (6,4,3,4) (abajo).

El modelo BCPE (7,26,3,5) presenta una clara sobreestimación de la varianza que se manifiesta en la gran variabilidad de las curvas de percentiles sobre todo en los extremos. Por tanto, este modelo no parece adecuado. Sin embargo, el modelo BCPE(6,4,3,4) presenta curvas suaves que, como ya se ha mencionado, es el objetivo y, además, presenta buena diagnosis. Por tanto, este modelo es el seleccionado.

Una vez elegido el modelo BCPE(6,4,3,4) para construir las curvas percentiles para el IMC de los niños de 6 a 15 años, se representaron (Figura 5.5) los modelos ajustados para cada parámetro ( $\mu$ ,  $\sigma$ ,  $\nu$ ,  $\tau$ ) en función de la edad.

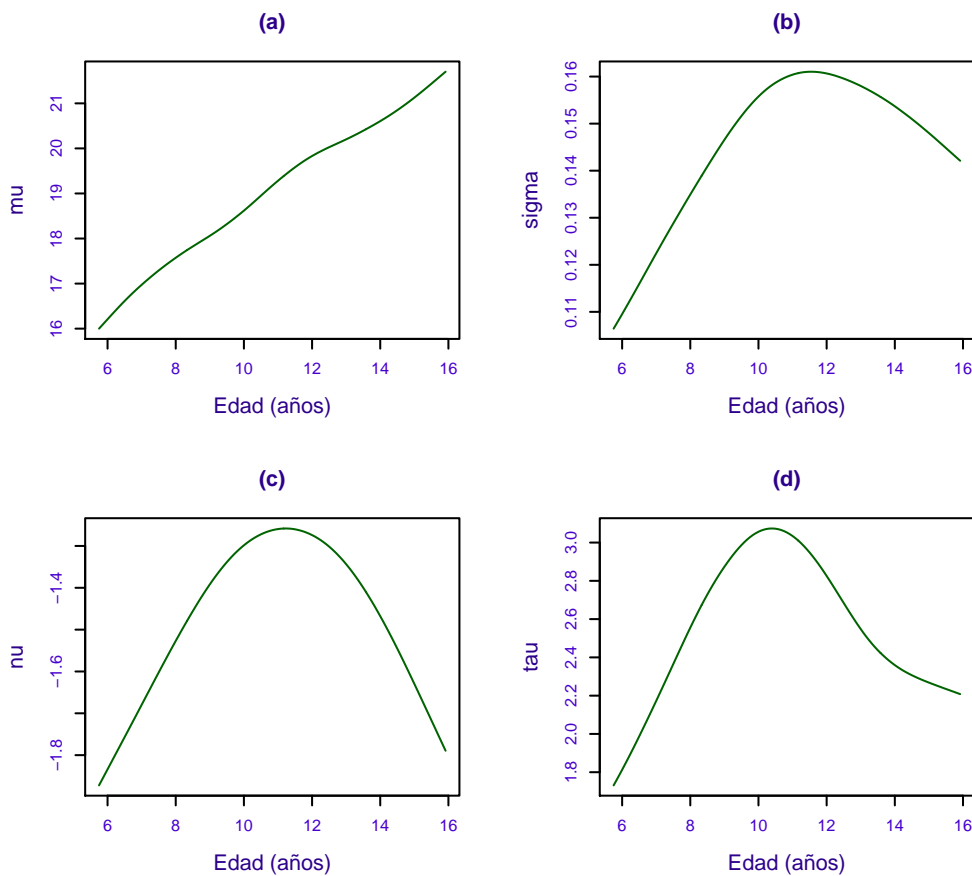


Figura 5.5: Gráfico de los parámetros  $\mu$  (a),  $\sigma$  (b),  $\nu$  (c) y  $\tau$  (d) ajustados con el modelo BCPE (6,4,3,4), en función de la edad, para el IMC de los niños.

En la Figura 5.5 se puede ver que el modelo ajustado para  $\mu$ , indica que la mediana del IMC en los niños crece progresivamente desde los 5 hasta los 16 años. Para  $\sigma$ , el modelo ajustado crece hasta los 12 años, donde se alcanza la mayor variabilidad, y después se produce un pequeño decrecimiento hasta los 16 años. Por otro lado, la distribución del IMC para los niños de 6 a 15 años tiene fuerte asimetría positiva, ya que  $\nu < 0$  para todas las edades, además la asimetría positiva disminuye gradualmente desde los 6 hasta los 11 años y a partir de ahí, aumenta con la edad hasta llegar casi al punto de partida a los 15 años. Por último, el parámetro  $\tau$  indica que la distribución del IMC para los niños

de 6 a 8 años es similar a la de una normal ( $\tau \sim 2$ ), a partir de los 8 años, se produce un aumento progresivo de  $\tau$  hasta los 10 años y después se produce otro descenso hasta los 15 años hacia el valor 2.4, por tanto desde los 8 hasta los 15 años la distribución del IMC es platicúrtica ( $\tau > 2$ ).

## 5.2. IMC en niñas

Para las niñas había 2 valores atípicos extremos en el IMC, que se eliminaron de la muestra. La distribución BCPE resultó ser la adecuada para modelizar el IMC en las niñas, según muestran los resultados del ajuste mediante p-splines (Tabla 5.8).

Distribución	GAIC
BCPE	17760.33
BCCG	17775.49
NO	18222.32

Tabla 5.8: Resumen de los valores de GAIC, con penalización 2.5, para el ajuste mediante p-splines en el IMC en niñas.

La Tabla 5.9 refleja los resultados de la estimación de los hiperparámetros mediante la función `find.hyper`. Nuevamente, en la penalización  $k = \log(n)$  se alcanzó el límite inferior de la búsqueda para la media, por lo que fue necesario una búsqueda adicional con esta penalización, para comprobar si  $df_\mu = 1$  o  $df_\mu = 2$ . Los resultados de esta búsqueda adicional se muestran en la Tabla 5.10.

Penalización(#)	Valores iniciales	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	GAIC(#)
2	(15,15,15,15)	(12.04, 2.38, 1.45, 4.81)	17742.65
	(6,6,6,6)	(6.23, 2.30, 1.45, 4.68)	17742.54
	(2,2,2,2)	(6.62, 2.34, 1.45, 4.62)	17742.54
2.5	(15,15,15,15)	(4.60, 2.07, 1.36, 4.07)	17753.2
	(6,6,6,6)	(4.66, 2.08, 1.35, 4.07)	17753.2
	(2,2,2,2)	(4.68, 2.07, 1.36, 4.04)	17753.21
3	(15,15,15,5)	(3.87, 1.92, 1.28, 3.67)	17762.96
	(6,6,6,6)	(3.86, 1.94, 1.28, 3.66)	17762.96
	(2,2,2,2)	(3.85, 1.92, 1.28, 3.68)	17762.95
$\log(n) =$ 8.1867	(15,15,15,5)	(0.1, 1.34, 0.73, 0.53)	17829.03
	(6,6,6,6)	(0.1, 1.34, 0.73, 0.53)	17829.03
	(2,2,2,2)	(0.1, 1.34, 0.74, 0.53)	17829.04

Tabla 5.9: Resultados de la función `find.hyper` para el ajuste del IMC en las niñas.



Penalización(#)	$df_{\mu}$	$(df'_{\sigma}, df'_{\nu}, df'_{\tau})$	GAIC(#)
$\log(n) =$	2	(1.33, 0.71, 0.6)	17828.5
8.1867	1	(0.1, 0.1, 1.05)	18869.14

Tabla 5.10: Resultados de la búsqueda adicional para  $\mu$  en el IMC de las niñas. Se subraya en amarillo el valor mínimo del criterio GAIC.

Siguiendo el proceso de los niños, se seleccionaron los grados de libertad, de cada penalización, correspondientes al mínimo valor del criterio GAIC y se redondearon al entero más próximo, para ajustar los modelos. En este caso existen 5 posibles vectores, dos para la penalización  $k = 2$  y uno para cada una de las otras penalizaciones. En la Tabla 5.11 se pueden ver estos valores, acompañados de la penalización de la que provienen y de los valores estadísticos que caracterizan la distribución de los residuos para cada modelo ajustado.

			Resumen de los residuos			
Penalización	$(df'_{\mu}, df'_{\sigma}, df'_{\nu}, df'_{\tau})$	$(df_{\mu}, df_{\sigma}, df_{\nu}, df_{\tau})$	Media	Varianza	Coef.Asim.	Coef.Curt.
2	(7,2,1,5)	(9,4,3,7)	-0.0014	1.0001	0.0047	3.0789
2	(6,2,1,5)	(8,4,3,7)	-0.0014	1.0002	0.0047	3.0794
2.5	(5,2,1,4)	(7,4,3,6)	-0.0012	1.0001	0.0049	3.082
3	(4,2,1,4)	(6,4,3,6)	-0.0013	1.0002	0.005	3.0815
8.18	(0,1,1,1)	(2,3,3,3)	-0.0012	0.9988	0.0035	3.0862

Tabla 5.11: Tabla de resumen residual.

De entre los modelos BCPE (9,4,3,7) y BCPE (8,4,3,7), puesto que provienen de la misma penalización, se seleccionó el de menor valor del criterio GAIC para la penalización  $k = 2$ . Para BCPE (9,4,3,7),  $GAIC(2)=17743.24$  y para BCPE (8,4,3,7),  $GAIC(2)=17743.27$ . Por lo tanto, aunque la diferencia era mínima, se escogió el modelo BCPE (9,4,3,7).

Como para el caso de los niños, se hizo un análisis de residuos para cada uno de los modelos ajustados para comprobar su buen funcionamiento y seleccionar el mejor. El modelo BCPE (2,3,3,3) se eliminó de este proceso de selección, puesto que los resultados (no mostrados) de la diagnosis que mostraron los worm plots y el **Q.stats** indicaban que el modelo no se ajustaba bien para los coeficientes de variación, asimetría y curtosis, además, el test de de Agostino indicaba que se rechazaba la normalidad.

Por lo tanto, el proceso de selección se centró en los otros tres modelos. En las figuras 5.6, 5.7 y 5.8 se representan los gráficos worm plot para los modelos BCPE (9,4,3,7), BCPE (7,4,3,6) y BCPE (6,4,3,6), respectivamente.

En los tres gráficos worm plot de estos tres modelos se apreciaron algunos posibles desajustes para media, varianza, asimetría y curtosis. A continuación, se especifican estas insuficiencias.

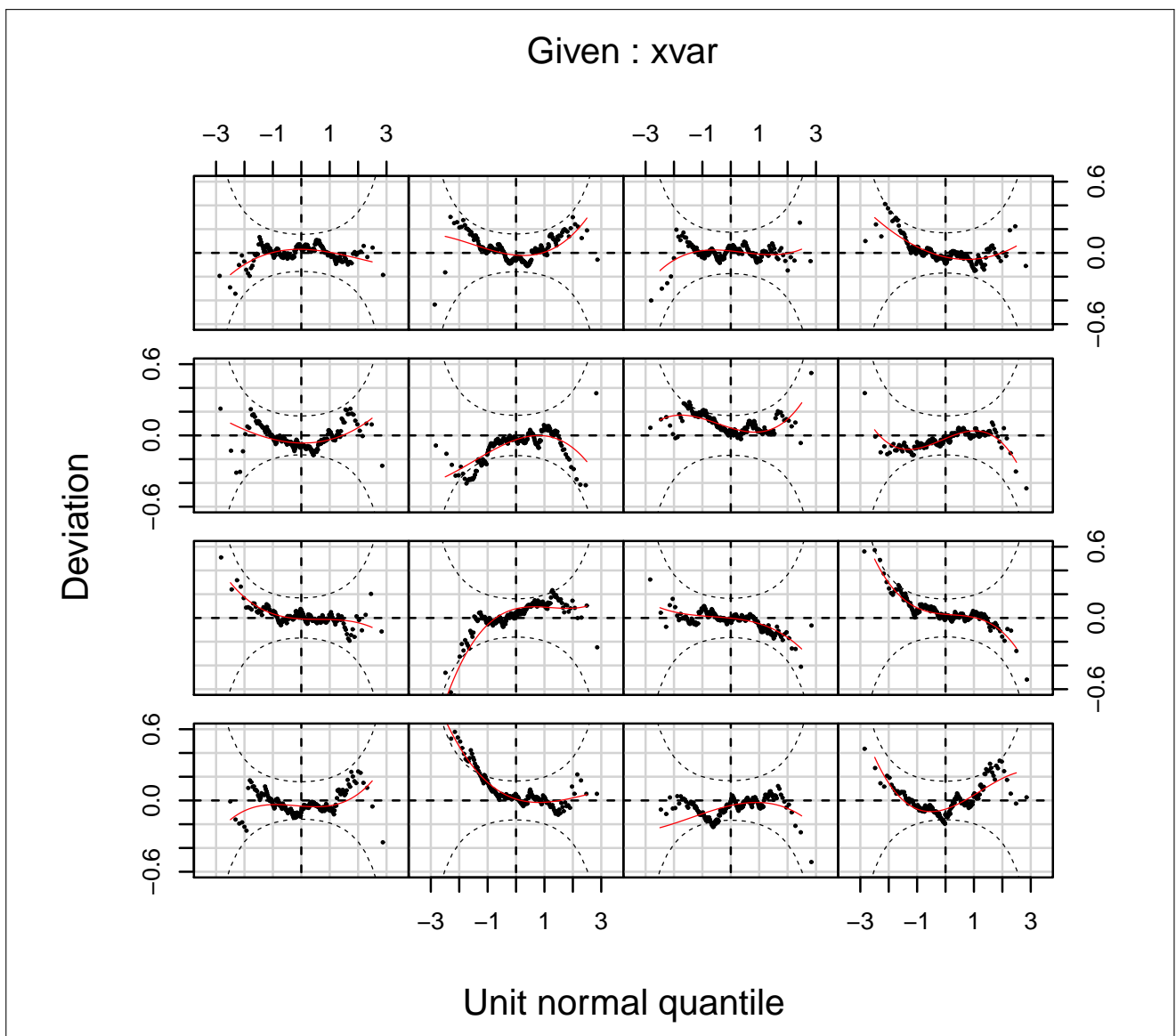


Figura 5.6: Gráfico Worm plot para el modelo ajustado con BCPE (9,4,3,7) del IMC de las niñas

En la Figura 5.6, para la media se identifica un desplazamiento leve por encima de  $y = 0$  en el intervalo 11. Para la varianza, se aprecia tendencia negativa en los intervalos 2 y 8 y tendencia positiva en el intervalo 6, esto puede indicar que la varianza ajustada sea muy grande o muy pequeña, respectivamente. Para la asimetría, se identifica forma convexa en los intervalos 2 y 4 y forma cóncava para los intervalos 6 y 10. Finalmente, la curtosis puede presentar algunas insuficiencias en los intervalos 1, 3, 6 y 12.

En la Figura 5.13, se reflejan las mismas insuficiencias que para el caso anterior, aunque para la media se aprecia que además del intervalo 11, también los intervalos 3 y 10 presentan irregularidades. Que en la media se refleje este comportamiento, tiene que ver con la disminución en una unidad de los grados de libertad del modelo BCPE(6,4,3,6) con respecto al anterior, BCPE(7,4,3,6).

La Figura 5.8, presenta las mismas insuficiencias aparentes que en el caso anterior. Con el mismo

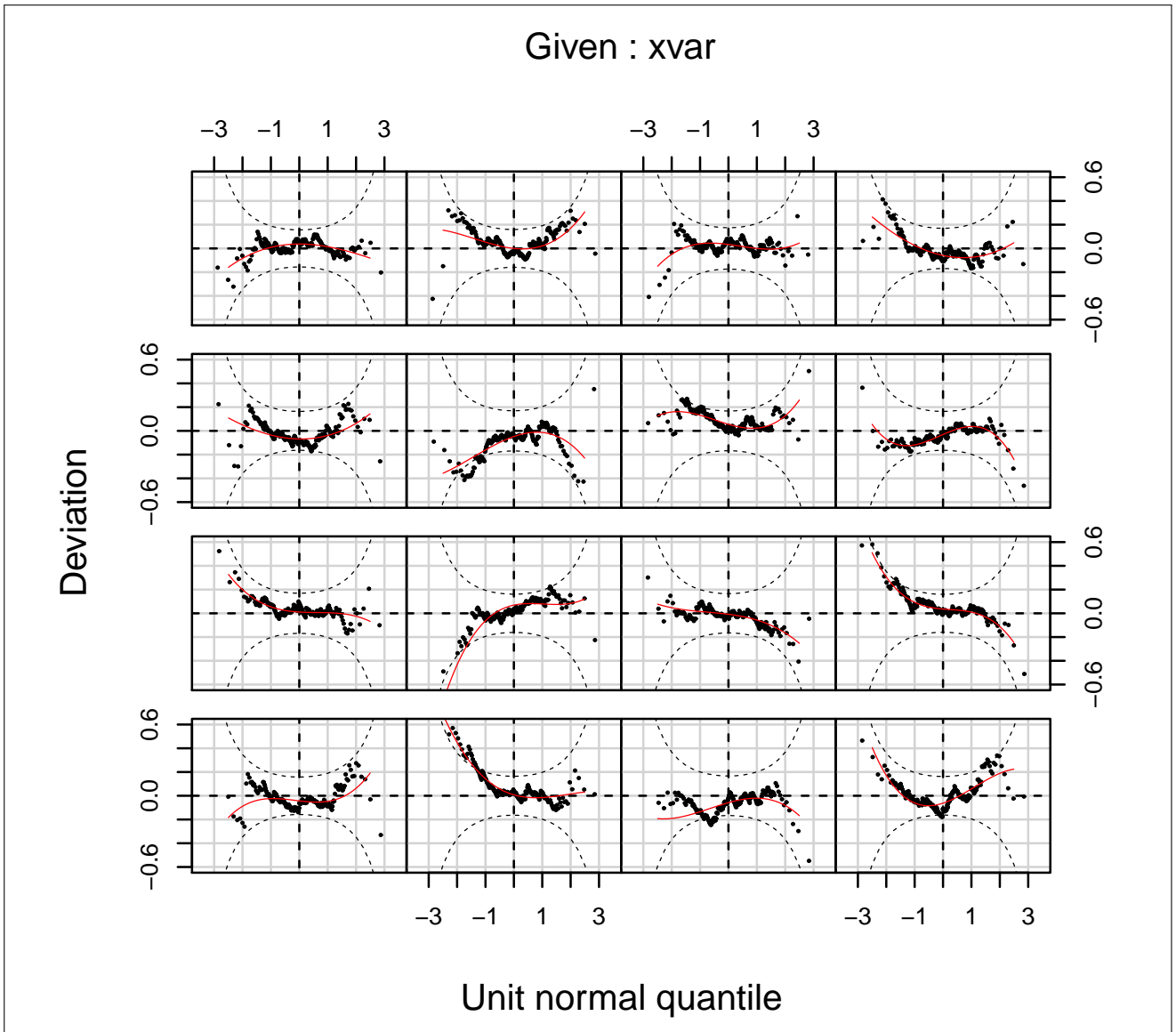


Figura 5.7: Gráfico Worm plot para el modelo BCPE (7,4,3,6) del IMC de las niñas.

argumento que dimos para la Figura 5.7, como los grados de libertad para la media son todavía menores, se reflejan las irregularidades de una forma un poco más contundente.

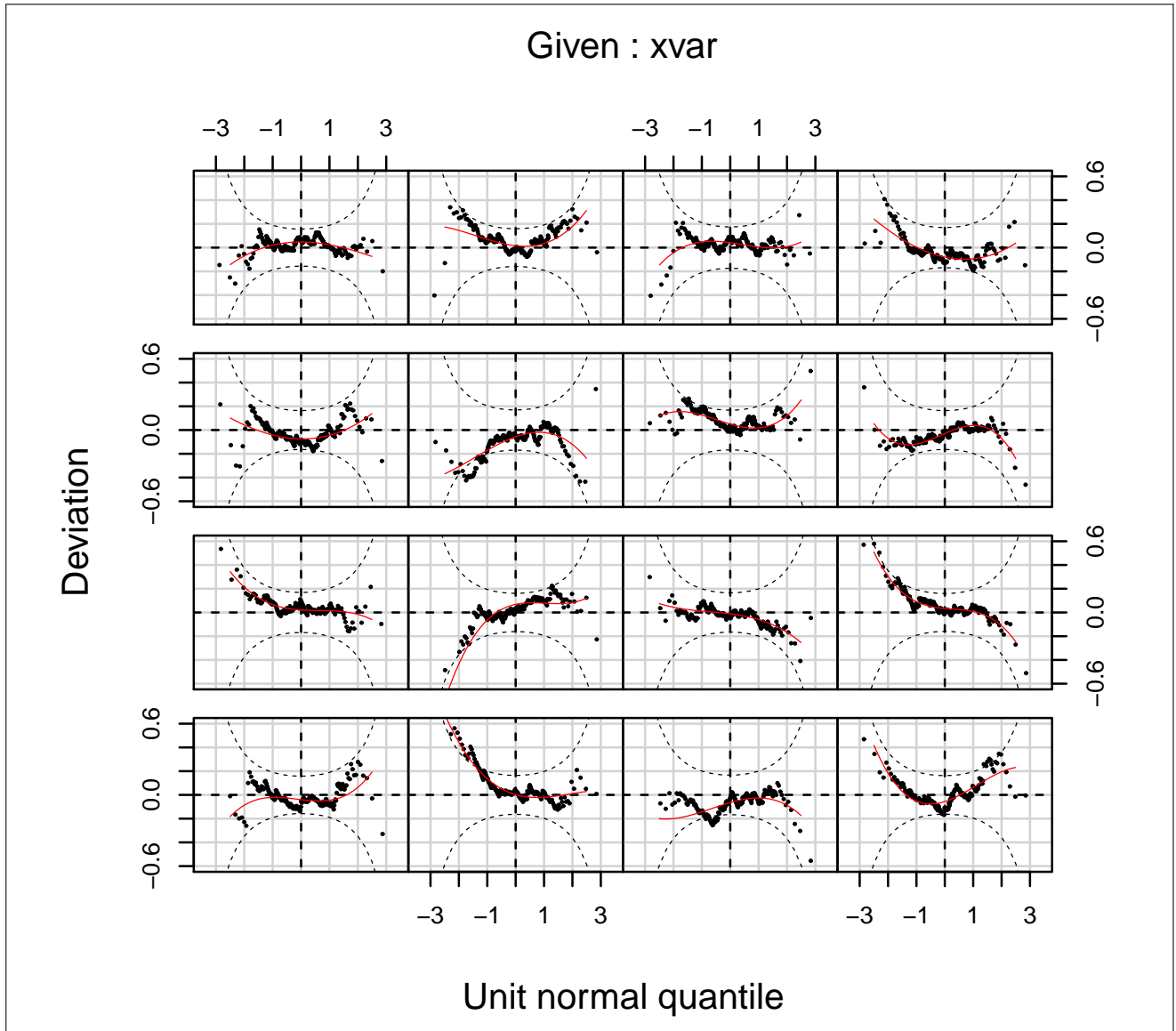


Figura 5.8: Gráfico Worm plot para el modelo ajustado con BCPE (6,4,3,6) del IMC de las niñas.

Para contrastar estas posibles insuficiencias, en las Tablas 5.12, 5.13 y 5.14 se presentan los valores de los estadísticos  $Z$  y  $Q$  para los modelos BCPE (9,4,3,7), BCPE (7,4,3,6) y BCPE (6,4,3,6), respectivamente.

nº	Intervalos de edad	Valores $Z_{gj}$				Agostino( $K^2$ )	N
		$\mu$	$\sigma$	$\nu$	$\tau$		
1	68.5 to 76.5	-0.62	0.52	0.14	1.25	1.59	243
2	76.5 to 83.5	1.02	-2.06	2.11	-0.09	4.46	225
3	83.5 to 91.5	-0.95	0.86	-0.96	0.68	1.38	218
4	91.5 to 100.5	-0.21	0.69	2.00	-0.75	4.57	229
5	100.5 to 108.5	0.15	-0.88	0.65	-0.55	0.73	212
6	108.5 to 117.5	0.26	2.34	-2.32	1.89	8.97	233
7	117.5 to 123.5	-0.23	-0.97	-0.51	-0.45	0.46	221
8	123.5 to 130.5	0.69	-1.75	0.54	-1.61	2.89	237
9	130.5 to 137.5	-0.51	0.09	1.07	0.30	1.24	226
10	137.5 to 144.5	-1.10	1.19	-1.20	-0.20	1.47	211
11	144.5 to 151.5	1.25	-0.50	0.85	1.05	1.82	216
12	151.5 to 158.5	-0.54	0.60	-0.32	-1.56	2.53	225
13	158.5 to 167.5	0.07	0.20	-0.99	0.34	1.09	248
14	167.5 to 174.5	0.28	0.14	1.31	0.74	2.26	239
14	174.5 to 181.5	0.06	0.09	-0.49	0.69	0.72	199
16	181.5 to 191.5	-0.02	-1.02	1.23	0.22	1.56	211
	Estadístico Q	6.42	19.14	23.62	14.13	37.75	3593
	df para Q	7	13.50	13	9	22	
	p-val para Q	0.49	0.14	0.03	0.12	0.02	

Tabla 5.12: Tabla de estadísticos  $Q$  y  $Z$  para el modelo BCPE (9,4,3,7).

nº	Intervalos de edad	Valores $Z_{gj}$				Agostino( $K^2$ )	N
		$\mu$	$\sigma$	$\nu$	$\tau$		
1	68.5 to 76.5	-0.55	0.51	0.11	1.45	2.11	243
2	76.5 to 83.5	0.96	-2.03	2.07	-0.24	4.35	225
3	83.5 to 91.5	-1.14	0.87	-0.81	0.26	0.73	218
4	91.5 to 100.5	-0.03	0.62	2.02	-0.88	4.87	229
5	100.5 to 108.5	0.43	-0.91	0.67	-0.60	0.81	212
6	108.5 to 117.5	0.13	2.40	-2.38	2.07	9.97	233
7	117.5 to 123.5	-0.31	-0.99	-0.47	-0.36	0.35	221
8	123.5 to 130.5	0.79	-1.77	0.58	-1.62	2.96	237
9	130.5 to 137.5	-0.56	0.08	1.11	0.28	1.31	226
10	137.5 to 144.5	-1.26	1.21	-1.18	-0.21	1.43	211
11	144.5 to 151.5	1.14	-0.50	0.84	0.99	1.69	216
12	151.5 to 158.5	-0.61	0.61	-0.32	-1.67	2.90	225
13	158.5 to 167.5	0.19	0.18	-0.97	0.23	0.99	248
14	167.5 to 174.5	0.57	0.09	1.28	0.78	2.24	239
15	174.5 to 181.5	0.28	0.05	-0.55	0.79	0.93	199
16	181.5 to 191.5	-0.36	-0.95	1.24	0.27	1.61	211
	Estadístico Q	7.58	19.26	23.59	15.66	39.26	3593
	df para Q	9.00	13.50	13	10	23	
	p-val para Q	0.58	0.13	0.04	0.11	0.02	

Tabla 5.13: Tabla de estadísticos  $Q$  y  $Z$  para el modelo BCPE (7,4,3,6).

n°	Intervalos de edad	Valores $Z_{gj}$				Agostino( $K^2$ )	N
		$\mu$	$\sigma$	$\nu$	$\tau$		
1	68.5 to 76.5	-0.52	0.50	0.10	1.45	2.12	243
2	76.5 to 83.5	0.87	-2.01	2.08	-0.24	4.36	225
3	83.5 to 91.5	-1.25	0.89	-0.81	0.23	0.71	218
4	91.5 to 100.5	0.08	0.59	2.04	-0.87	4.92	229
5	100.5 to 108.5	0.59	-0.94	0.68	-0.61	0.84	212
6	108.5 to 117.5	0.12	2.40	-2.37	2.06	9.87	230
7	117.5 to 123.5	-0.36	-0.99	-0.46	-0.36	0.34	221
8	123.5 to 130.5	0.76	-1.77	0.59	-1.62	2.96	237
9	130.5 to 137.5	-0.65	0.09	1.11	0.28	1.32	226
10	137.5 to 144.5	-1.39	1.22	-1.18	-0.21	1.45	211
11	144.5 to 151.5	1.06	-0.49	0.83	0.99	1.68	216
12	151.5 to 158.5	-0.59	0.61	-0.32	-1.67	2.89	225
13	58.5 to 167.5	0.35	0.15	-0.96	0.22	0.98	248
14	167.5 to 174.5	0.81	0.05	1.26	0.77	2.18	239
15	174.5 to 181.5	0.38	0.02	-0.59	0.81	1.01	199
16	181.5 to 191.5	-0.66	-0.91	1.25	0.30	1.66	211
	Estadístico Q	8.83	19.17	23.64	15.62	39.26	3593
	df para Q	10.00	13.50	13	10	23	
	p-val para Q	0.55	0.14	0.03	0.11	0.02	

Tabla 5.14: Tabla de estadísticos  $Q$  y  $Z$  para el modelo BCPE (6,4,3,6).

En la Tabla 5.12, se confirma que los desvíos que aparecieron en la Figura 5.6 del modelo BCPE (9,4,3,7) en determinados grupos de edad, para la media y para la curtosis no son grandes, puesto que los valores de  $Z_{gj}$  correspondientes no superan a 2 en valor absoluto. Sin embargo, para los intervalos  $g = 2, 6$  para el coeficiente de variación ( $j = 2$ ) y los intervalos  $g = 2, 6, 4$  para la asimetría ( $j = 3$ ), se confirma que hay irregularidades, ya que presentan valores grandes de  $Z_{gj}$ . Para el coeficiente de variación, el estadístico  $Q_2$  no es significativo, al contrario de lo que ocurre con la asimetría, que si presenta significación. Por lo tanto, el modelo no ajusta bien la asimetría. Además, el p-valor asociado al test de Agostino indica que se rechaza la normalidad.

En las Tablas 5.13 y 5.14, se aprecia el mismo diagnóstico que para el caso anterior. Además, para estos modelos se verifica la irregularidad de la curtosis en el intervalo 6, aún siendo el estadístico  $Q$  de la curtosis, no significativo.

En los 3 casos, se observó significación para la asimetría. Puede que esta esté condicionada por valores residuales “demasiado” grandes (o pequeños). Para identificar estos residuos, se utilizó un criterio que consideraba demasiado grandes (o pequeños) los residuos mayores que 3.5 en valor absoluto. Se identificaron tres residuos inferiores a -3.5, los mismos para los 3 modelos (Tabla 5.15). Se eliminaron de la muestra los datos correspondientes a estos residuos y se ajustaron los modelos BCPE (9,4,3,7), BCPE (7,4,3,6) y BCPE (6,4,3,6) para la nueva muestra.

	ncuest	sexo	edad (meses)	edad (años)	imc	p25	p75
1	2701102	Mujer	73	6	11.49	15.15	17.72
2	3625208	Mujer	91	7	12.22	15.27	18.66
3	3204226	Mujer	115	9	11.11	16.19	20.71

Tabla 5.15: Tabla de los atípicos residuales.

Para hacer la diagnosis de estos modelos, se analizan directamente los valores de los estadísticos  $Z$  y  $Q$ . En La Tabla 5.16 se muestran p-valores de los estadísticos  $Q_j$  para los modelos BCPE (9,4,3,7), BCPE (7,4,3,6) y BCPE (6,4,3,6) con la nueva muestra.

Todos los valores de los estadísticos  $Z_{gj}$  son menores que 2 (datos no mostrados), es decir, se han corregido las insuficiencias de los tres modelos. Además los p-valores de los estadísticos  $Q_j$  pasan a ser todos no significativos. Por tanto, ahora se puede decir que los tres modelos se ajustan bien para los cuatro parámetros de distribución.

Modelos	p-valores de $Q$			
	$\mu$	$\sigma$	$\nu$	$\tau$
BCPE (9,4,3,7)	0.48	0.25	0.18	0.40
BCPE (7,4,3,6)	0.58	0.25	0.18	0.34
BCPE (6,4,3,6)	0.56	0.26	0.18	0.34

Tabla 5.16: p-valores del estadístico  $Q$  de los modelos BCPE (9,4,3,7), BCPE (7,4,3,6) y BCPE (6,4,3,6) para el IMC de los niños con la muestra sin los atípicos residuales.

En la Figura 5.9 se representan las curvas percentiles del IMC frente a la edad obtenidas con los modelos BCPE(9,4,3,7) y BCPE(6,4,3,6) con la muestra corregida, respectivamente. Se omite la representación de las curvas percentiles correspondientes con el modelo BCPE (7,4,3,6), ya que solamente se diferencia de BCPE(6,4,3,6) en un grado de libertad para la media, en consecuencia sus curvas son prácticamente iguales.

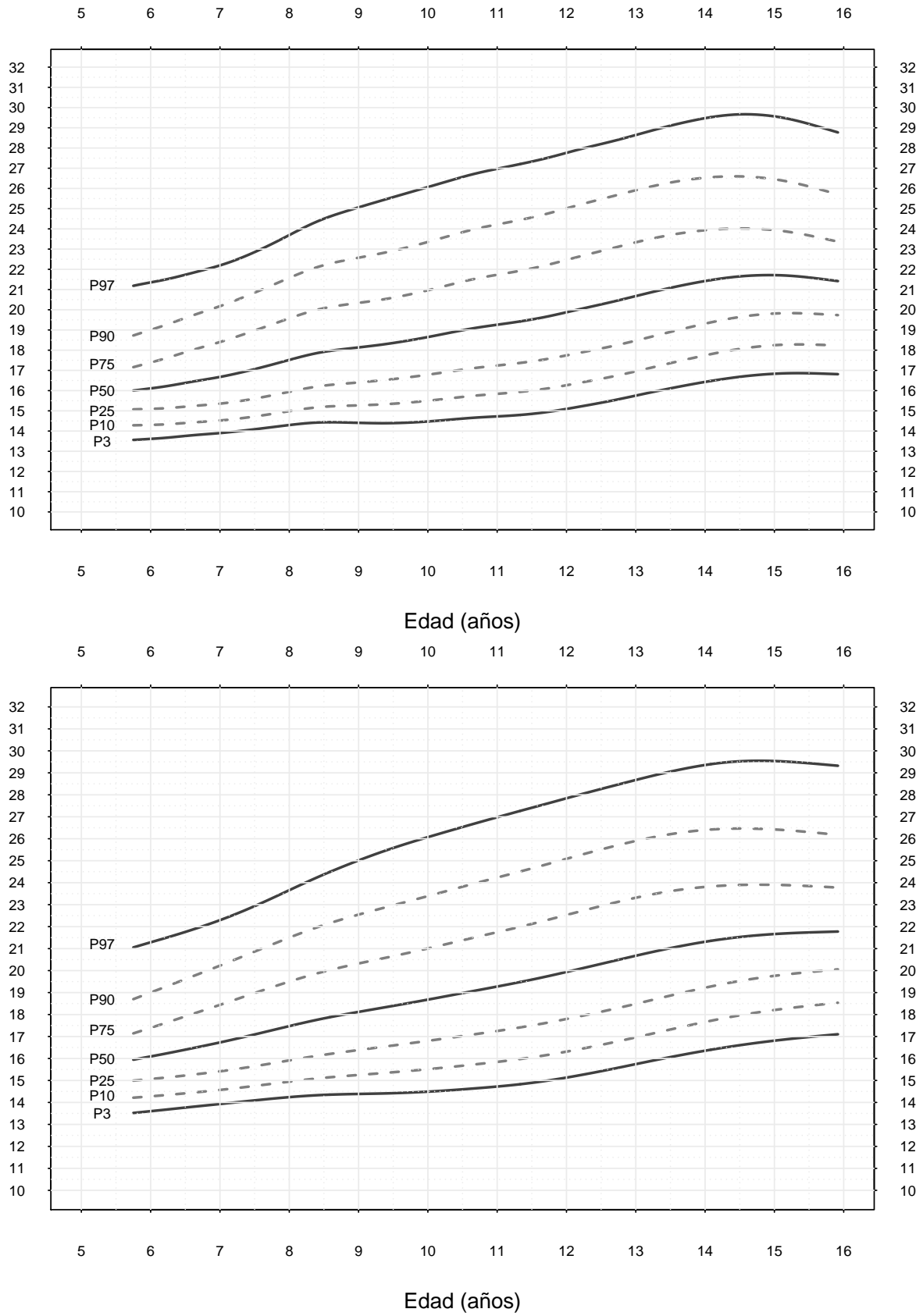


Figura 5.9: Curvas percentiles para el IMC ( $\text{kg}/\text{m}^2$ ) de las niñas con el modelo BCPE (9,4,3,7) (arriba) y con el modelo BCPE(6,4,3,6) (abajo).



Los dos modelos proporcionaron curvas percentiles semejantes, lo que dificultó el proceso de elección. Teniendo en cuenta que el objetivo de los modelos es encontrar el menor número de grados de libertad que proporcionen un buen ajuste para los parámetros de distribución, se podría haber seleccionado directamente el modelo BCPE(6,4,3,6), ya que es el más sencillo. Pero antes de apoyarnos en esto, se realizó una nueva búsqueda de los grados de libertad por medio de la función `find.hyper`; en este caso, se eliminaron de la muestra los 2 atípicos extremos y los 3 datos correspondientes a los residuos demasiado pequeños identificados en los modelos ajustados anteriormente, es decir, en total 5 datos de la muestra.

Los resultados de la búsqueda son muy parecidos a los obtenidos en la Tabla 5.9, por lo que se omitieron y se mostraron directamente los valores redondeados en la Tabla 5.17. Se puede ver que el hecho de haber eliminado 5 datos de la muestra (en vez de 3), solo afecta, ligeramente, a la penalización mas pequeña ( $k = 2$ ).

Penalización	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	$(df_\mu, df_\sigma, df_\nu, df_\tau)$	Resumen de los residuos			
			Media	Varianza	Coef.Asim.	Coef.Curt.
2	(19,2,2,5)	(21,4,4,7)	-0.0009	0.9995	0.0089	2.9886
2.5	(5,2,1,4)	(7,4,3,6)	-0.0010	0.9997	0.0092	2.9930
3	(4,2,1,4)	(6,4,3,6)	-0.0011	0.9997	0.0092	2.9928
8.1859	(0,1,1,1)	(2,3,3,3)	-0.0019	0.9986	0.0086	2.9890

Tabla 5.17: Tabla de resumen residual.

El modelo BCPE(2,3,3,3) se rechazó, nuevamente, ya que los resultados de la diagnosis indicaron que no se ajustaba bien para el coeficiente de variación ni para la curtosis, además, el test de Agostino rechazaba la normalidad de los residuos.

Los valores correspondientes a la penalización 2, produjeron una sobreestimación para la media, lo que nos llevó a rechazar este modelo. En la Figura 5.10, se representan las curvas percentiles del modelo BCPE (21,4,4,7), donde se aprecia esta sobreestimación, y del modelo BCPE(6,4,3,6) a modo comparativo.

Finalmente, se eligió el modelo mas sencillo, como indica el criterio que se mencionó anteriormente. En consecuencia, el modelo BCPE (6,4,3,6) fue el seleccionado para representar las curvas percentiles en función de la edad del IMC de las niñas de 6 a 15 años.

Se representaron graficamente los parámetros del modelo ajustado respecto a la edad, y se pueden ven en la Figura 5.11.

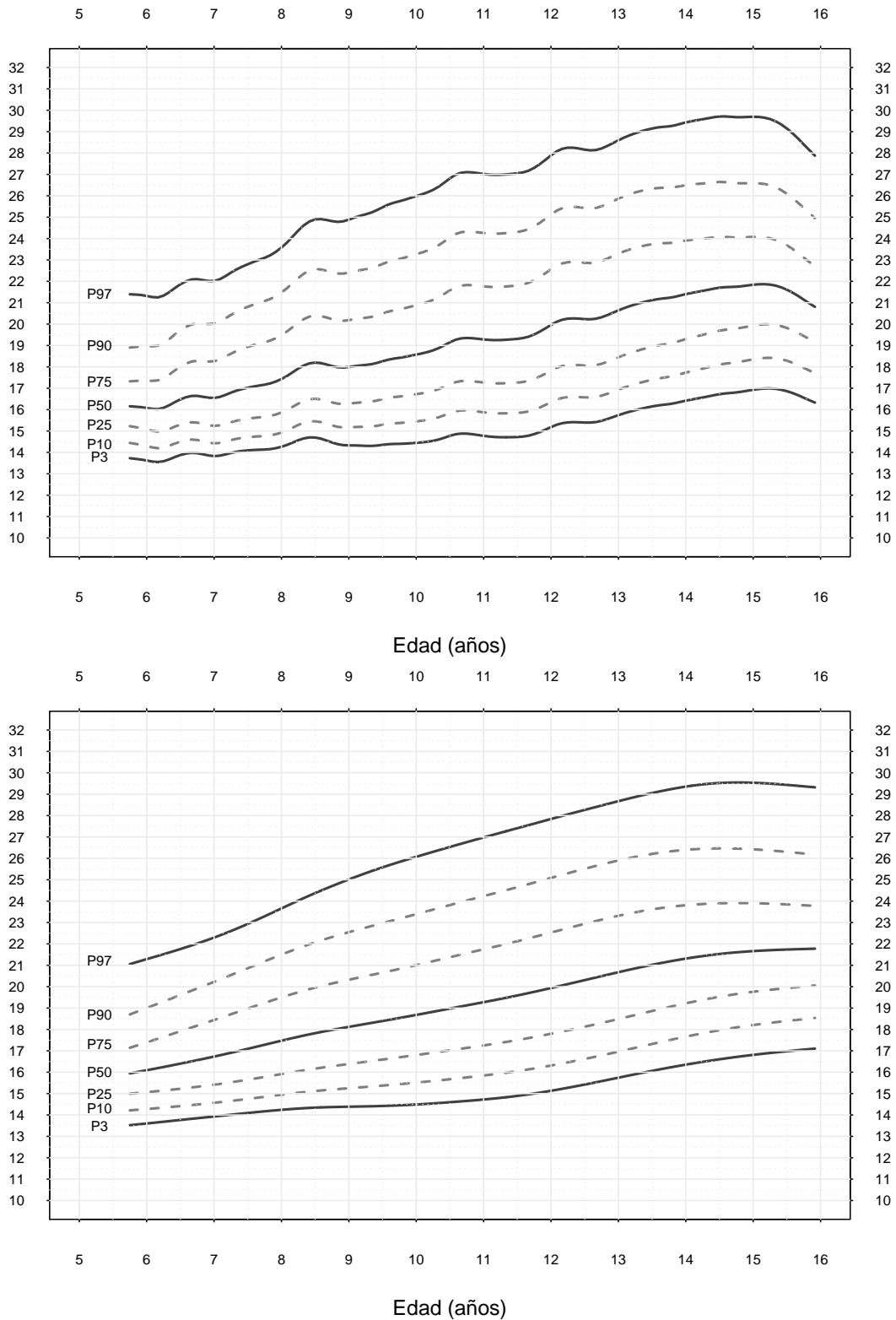


Figura 5.10: Curvas percentiles para el IMC ( $\text{kg}/\text{m}^2$ ) de las niñas con el modelo BCPE (21,4,4,7) (arriba) y el modelo BCPE(6,4,3,6) (abajo).

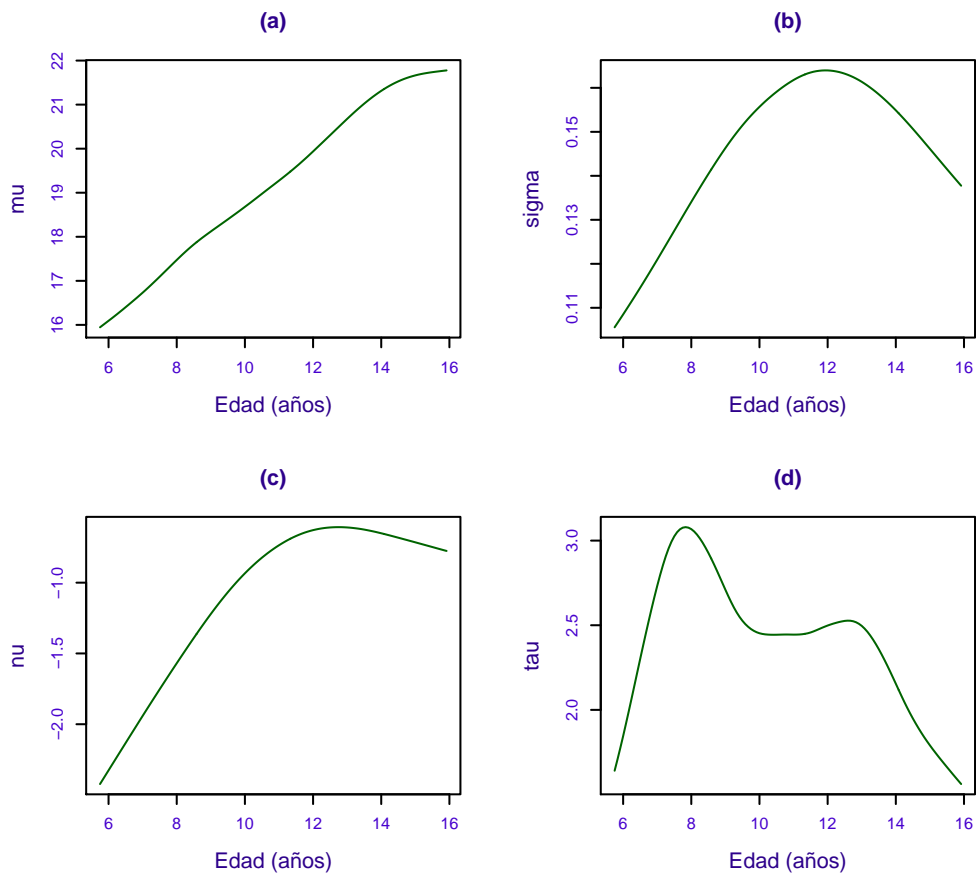


Figura 5.11: Gráfico de los parámetros  $\mu$  (a),  $\sigma$  (b),  $\nu$  (c) y  $\tau$  (d) ajustados para el modelo BC-PE(6,4,3,6), que ajusta el IMC en las niñas en función de la edad.

La evolución de los parámetros  $\mu$  y  $\sigma$  respecto a la edad, para el modelo ajustado para el IMC de las niñas es similar a la del modelo ajustado para el IMC de los niños. La mediana del IMC de las niñas crece progresivamente desde los 6 hasta los 15 años, ralentizándose levemente el aumento a partir de los 13 años. Para  $\sigma$ , el modelo ajustado crece hasta los 12 años, donde se produce un pequeño decrecimiento hasta los 15 años. Como para el caso de los niños, la distribución del IMC en niñas tiene asimetría positiva ( $\nu < 0$ ) en todo el intervalo de edad, en este caso los valores son más pequeños y el decrecimiento a partir de los 12. Por último, el modelo ajustado para  $\tau$ , indica que la distribución del IMC para las niñas de 6 años es prácticamente normal, a partir de ahí la curtosis aumenta hasta 3 a los 8 años, luego vuelve a disminuir hasta los 10 años donde se estabiliza entorno a una curtosis de 2.5 (distribución platicúrtica) hasta los 13 años, con una pequeña oscilación de por medio, por último desde los 13 hasta los 15 se produce un descenso de la curtosis hacia la de una distribución normal ( $\tau = 2$ ).

### 5.3. Peso en niños

Para el peso en niños había 10 atípicos extremos que se eliminaron de la muestra. El ajuste mediante p-splines concluyó que la distribución adecuada para esta variable en niños era la BCPE (Tabla 5.18).

Distribución	GAIC
BCPE	26669.93
BCCG	26691.65
NO	27232.09

Tabla 5.18: Resumen de los valores de GAIC, con penalización 2.5, para el ajuste mediante p-splines en el peso en niño.

Los valores estimados de los grados de libertad para cada parámetro  $(df_\mu, df_\sigma, df_\nu, df_\tau)$ , procedentes de la función `find.hyper` se muestran en la Tabla 5.19.

Penalización(#)	Valores iniciales	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	GAIC(#)
2	(15,15,15,15)	(2.37, 2.07, 1.49, 0.62)	26660.67
	(6,6,6,6)	(2.41, 2.09, 1.53, 0.62)	26660.67
	(2,2,2,2)	(2.40, 2.01, 1.53, 0.62)	26660.67
2.5	(15,15,15,15)	(2.2, 1.91, 1.24, 0.47)	26667.81
	(6,6,6,6)	(2.16, 1.91, 1.23, 0.49)	26667.8
	(2,2,2,2)	(2.14, 1.92, 1.24, 0.51)	26667.79
3	(15,15,15,15)	(2.03, 1.81, 1.09, 0.38)	26674.61
	(6,6,6,6)	(2.02, 1.87, 1.09, 0.37)	26674.61
	(2,2,2,2)	(2.02, 1.81, 1.09, 0.36)	26674.61
$\log(n) = 8.2514$	(15,15,15,15)	(1.53, 1.32, 0.41, 0.1)	26738.12
	(6,6,6,6)	(1.53, 1.32, 0.41, 0.1)	26738.12
	(2,2,2,2)	(1.53, 1.32, 0.40, 0.1)	26738.12

Tabla 5.19: Resultados de la función `find.hyper` para el ajuste del peso en los niños.

En la penalización  $k = \log(n)$  se alcanzó el límite de la búsqueda para el hiperparámetro que regula la curtosis, por lo que fue necesario realizar una búsqueda adicional para comprobar si, para esta penalización, la curtosis se puede simplificar a una constante o a un término lineal. En la Tabla 5.20 se muestran los resultados.

Penalización(#)	$df_\tau$	$(df'_\mu, df'_\sigma, df'_\nu)$	GAIC(#)
$\log(n) =$	2	(1.53, 1.32, 0.33)	26736.08
8.2514	1	(1.55, 1.32, 0.37)	26737.88

Tabla 5.20: Resultados de la búsqueda adicional para  $\tau$  en el peso de los niños. Se subraya en amarillo el valor mínimo del criterio GAIC.

Tras redondear los resultados se obtuvieron 3 posibilidades (Tabla 5.21), para las penalizaciones 2.5 y 3 la opción era la misma.

Penalización	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	$(df_\mu, df_\sigma, df_\nu, df_\tau)$	Resumen de los residuos			
			Media	Varianza	Coef.Asim.	Coef.Curt.
2	(2,2,1,1)	(4,4,3,3)	-0.0006	0.9992	0.0078	2.9178
2.5	(2,2,1,0)	(4,4,3,2)	-0.0014	0.9992	0.0071	2.9213
3	(2,2,1,0)	(4,4,3,2)	-0.0014	0.9992	0.0071	2.9213
8.2514	(1,1,0,0)	(3,3,2,2)	-0.0012	0.9995	0.0130	2.9180

Tabla 5.21: Tabla de resumen residual.

De aquí en adelante, para la diagnosis se comentarán los resultados de la función `Q.stats` y solo se mostrarán los p-valores del estadístico  $Q$  que, como ya se ha dicho, definen si el modelo se ajusta bien para cada uno de los parámetros de distribución

Para el modelo BCPE (3,3,2,2), se observaron varias insuficiencias, una relativa a la media en el sexto grupo de edad y otras dos relativas al coeficiente de variación y el de asimetría para el decimosexto grupo de edad. Sin embargo, para los otros dos modelos, solamente se identifica una insuficiencia para la asimetría en el decimosexto grupo de edad. Todas estas insuficiencias, como ya se vió en el IMC, se manifiestan a través del valor del estadístico  $|Z_{gj}| > 2$ . Como se puede ver en la Tabla 5.22, para el caso de los modelos BCPE (4,4,3,3) y BCPE (4,4,3,2), la insuficiencia no impidió que los modelos fueran válidos, ya que los p-valores para los estadísticos  $Q_j$  con  $j = 1, 2, 3, 4$  son no significativos. Al contrario ocurre con el modelo BCPE (3,3,2,1), que no se ajustaba bien para la media, por lo que se eliminó del proceso de selección.

Modelos	p-valores de $Q$			
	$\mu$	$\sigma$	$\nu$	$\tau$
BCPE(4,4,3,3)	0.45	0.61	0.80	0.87
BCPE(4,4,3,2)	0.46	0.58	0.76	0.84
BCPE(3,3,2,1)	0.05	0.14	0.21	0.59

Tabla 5.22: p-valores del estadístico  $Q$  de los modelos que ajustan el peso en los niños.

Por tanto, de entre los modelos BCPE(4,4,3,3) y BCPE(4,4,3,2) para estimar las curvas percentiles para el peso de niños de 6 a 15 años se eligió el segundo, ya que era el más sencillo.

A continuación, en la figura 5.12, se representan las gráficas de los parámetros del modelo ajustado respecto a la edad.

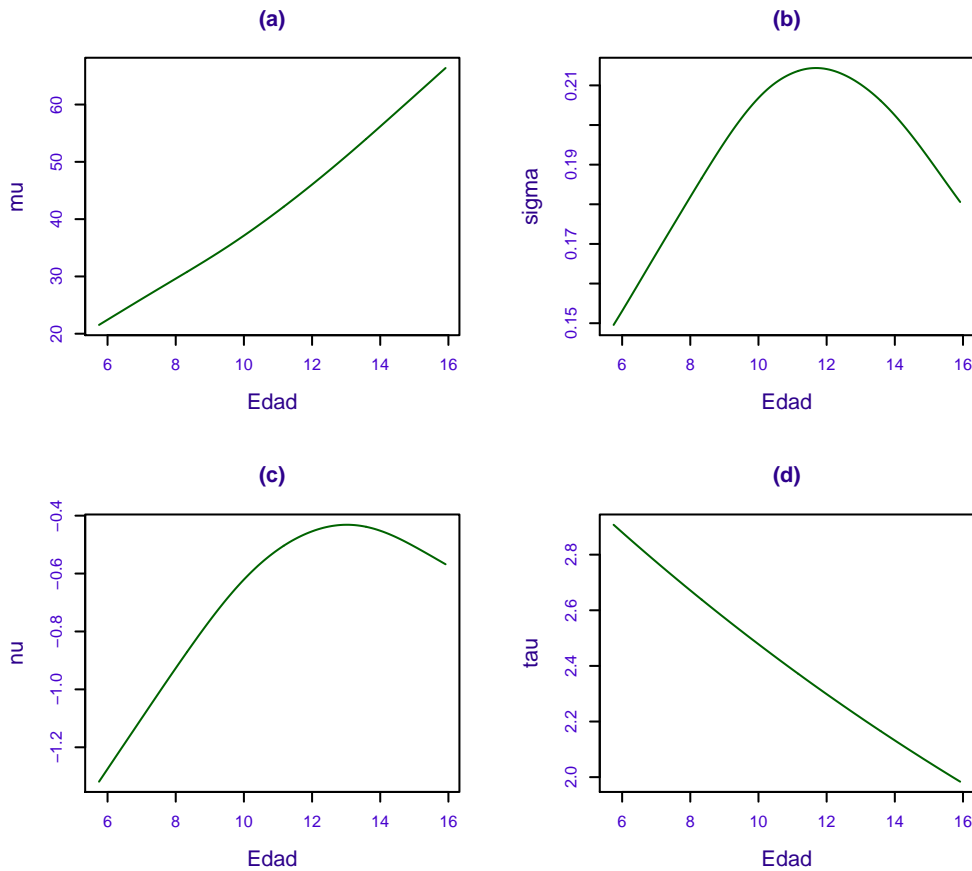


Figura 5.12: Gráfico de los parámetros  $\mu$  (a),  $\sigma$  (b),  $\nu$  (c) y  $\tau$  (d) ajustados para el modelo BCPE (4,4,3,2), que ajusta el peso en los niños en función de la edad.

El modelo ajustado para  $\mu$  indica que la mediana del peso en los niños crece de forma homogénea con la edad. Para  $\sigma$ , el modelo ajustado indica que aumenta progresivamente hasta los 12 años, donde se produce la mayor variabilidad, y luego disminuye lentamente hasta los 15 años. La distribución del peso para los niños de 6 a 15 años presenta asimetría positiva ( $\nu < 0$ ), de los 6 a los 13 años  $\nu$  disminuye progresivamente y después se produce un pequeño aumento hasta los 15 años. En cuanto a la curtosis, el parámetro  $\tau$  disminuye progresivamente desde los 6 hasta los 15 años, de forma que la distribución del peso en niños empieza siendo platicúrtica ( $\tau > 2$ ) y acaba siendo la de una normal ( $\tau = 2$ ), a los 15 años.

## 5.4. Peso en niñas

En el peso de las niñas había 5 valores atípicos extremos, que se eliminaron de la muestra. La distribución adecuada según el ajuste mediante p-splines resultó ser la BCPE, como se muestra en la

Tabla 5.23.

Distribución	GAIC
BCPE	24525.08
BCCG	24528.29
NO	24923.52

Tabla 5.23: Resumen de los valores de GAIC, con penalización 2.5, para el ajuste mediante p-splines en el peso en niñas.

En la Tabla 5.24 se presentan los valores estimados de los grados de libertad para cada parámetro  $(df_{\mu}, df_{\sigma}, df_{\nu}, df_{\tau})$ , procedentes de la función `find.hyper`.

Penalización(#)	Valores iniciales	$(df_{\mu}^l, df_{\sigma}^l, df_{\nu}^l, df_{\tau}^l)$	GAIC(#)
2	(15,15,15,15)	(5.32, 2.13, 0.96, 3.86)	24513.77
	(6,6,6,6)	(5.32, 2.13, 1.0952, 3.82)	24513.77
	(2,2,2,2)	(5.33, 2.13, 0.95, 3.89)	24513.77
2.5	(15,15,15,15)	(5.02, 2.01, 0.84, 3.31)	24523.68
	(6,6,6,6)	(5.03, 2.00, 0.84, 3.31)	24523.68
	(2,2,2,2)	(5.05, 2.00, 0.85, 3.36)	24523.68
3	(15,15,15,15)	(4.74, 1.89, 0.74, 2.75)	24533.06
	(6,6,6,6)	(4.71, 1.89, 0.76, 2.51)	24533.06
	(2,2,2,2)	(4.73, 1.90, 0.75, 2.74)	24533.06
$\log(n) =$ 8.1859	(15,15,15,15)	(3.41, 1.45, 0.1, 0.23)	24609.1
	(6,6,6,6)	(3.40, 1.45, 0.1, 0.23)	24609.1
	(2,2,2,2)	(3.40, 1.45, 0.1, 0.23)	24609.1

Tabla 5.24: Resultados de la función `find.hyper` para el ajuste del peso en las niñas.

Los resultados de la búsqueda adicional de la penalización  $k = \log(n)$ , por haber alcanzado el límite en el coeficiente de asimetría, se muestran en la Tabla 5.25

Penalización(#)	$df_{\nu}$	$(df_{\mu}^l, df_{\sigma}^l, df_{\tau}^l)$	GAIC(#)
$\log(n) =$	2	(3.41, 1.45, 0.25)	24609.03
8.1859	1	(3.31, 1.41, 0.48)	24615.97

Tabla 5.25: Resultados de la búsqueda adicional para  $\nu$  en el peso de las niñas. Se subraya en amarillo el valor mínimo del criterio GAIC.

En este caso se ajustaron 3 modelos distintos (Tabla 5.26), resultantes del redondeo de los valores anteriores. Para las penalizaciones 2.5 y 3, los modelos ajustados coinciden.

Penalización	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	$(df_\mu, df_\sigma, df_\nu, df_\tau)$	Resumen de los residuos			
			Media	Varianza	Coef.Asim.	Coef.Curt.
2	(5,2,1,4)	(7,4,3,6)	5.7e-5	0.9989	0.0052	2.9541
2.5	(5,2,1,3)	(7,4,3,5)	0.0001	0.9991	0.0054	2.9548
3	(5,2,1,3)	(7,4,3,5)	0.0001	0.9991	0.0054	2.9548
8.1859	(3,1,0,0)	(5,3,2,2)	0.0013	1.0001	0.0067	2.9478

Tabla 5.26: Tabla de resumen residual.

En cuanto a los resultados del **Q-stats**, el modelo BCPE (7,4,3,5) presentaba una insuficiencia para la curtosis en el cuarto grupo de edad. En el modelo BCPE (7,4,3,6) no se detectó ninguna irregularidad. Se puede ver en la Tabla 5.27 que los p-valorés de los estadísticos  $Q_j$  para estos dos modelos son no significativos. Sin embargo, para el modelo BCPE (5,3,2,2) se detectaron varias insuficiencias, dos para el coeficiente de variación ( $g = 2, 10$ ), una para la asimetría ( $g = 7$ ) y cuatro para la curtosis ( $g = 1, 3, 4, 10$ ), además, el estadístico  $Q_4$  para la curtosis es significativo, lo que indica que no se ajustana bien para este parámetro. Debido a esto, se eliminó del proceso de selección.

Modelos	p-valorés de $Q$			
	$\mu$	$\sigma$	$\nu$	$\tau$
BCPE(7,4,3,6)	0.88	0.56	0.90	0.16
BCPE(7,4,3,5)	0.88	0.56	0.89	0.10
BCPE (5,3,2,2)	0.28	0.11	0.67	0.004

Tabla 5.27: p-valorés del estadístico  $Q$  de los modelos que ajustan el peso en las niñas.

Aplicando el criterio del modelo más sencillo, se seleccionó el modelo BCPE (7,4,3,5) para estimar las curvas percentiles del peso en las niñas de 6 a 15 años. Las gráficas de los parámetros del modelo ajustado respecto a la edad se representan en la figura 5.13.



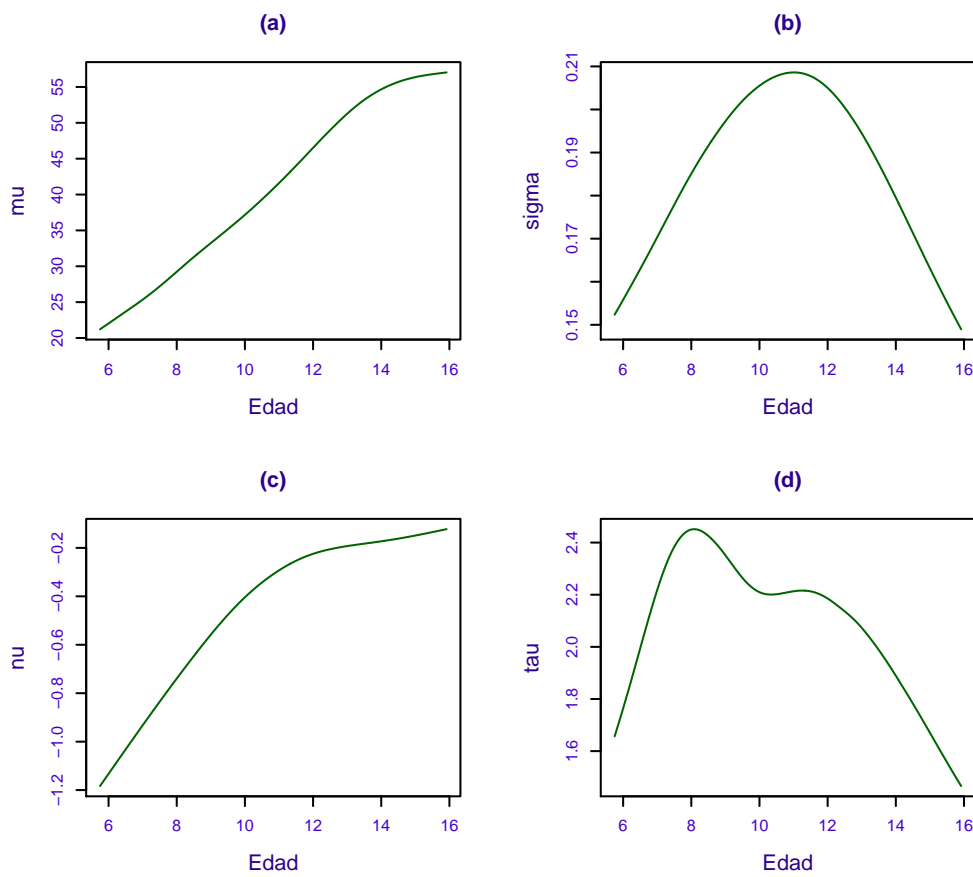


Figura 5.13: Gráfico de los parámetros  $\mu$  (a),  $\sigma$  (b),  $\nu$  (c) y  $\tau$  (d) ajustados para el modelo BCPE (7,4,3,5), que ajusta el peso en las niñas en función de la edad.

Como en el caso de los niños, el modelo ajustado para  $\mu$ , indica que la mediana para el peso aumenta progresivamente con la edad, sin embargo, en este caso el aumento se ralentiza a partir de los 13 años. Para  $\sigma$ , el modelo ajustado indica que aumenta de los 6 a los 11 años, donde alcanza la mayor variabilidad, y luego vuelve a disminuir hasta los 15 años. El modelo para  $\nu$ , indica que la distribución del peso en niñas presenta asimetría positiva ( $\nu < 0$ ), que disminuye progresivamente con la edad, ralentizándose a partir de los 11 años. Para  $\tau$ , se aprecian diferencias significativas con respecto al caso de los niños, en este caso el modelo ajustado indica que la distribución de los 6 a los 7 años pasa de ser leptocúrtica ( $\tau < 2$ ) a mesocúrtica ( $\tau = 2$ ), respectivamente, a partir de los 7 años la  $\tau$  aumenta hasta alcanzar el máximo valor a los 8 años (platicúrtica) luego vuelve a disminuir, con una pequeña oscilación a los 11 años, pasando por una distribución normal a los 14 años y leptocúrtica en los 15 años.

## 5.5. Talla en niños

Para la talla en los niños no había ningún atípico extremo que eliminar. El ajuste con p-splines determinó que la distribución adecuada era la BCCG, es decir, un ajuste con tres parámetros, ( $\mu$ ,  $\sigma$ ,

$\nu$ ). Cabe destacar que el valor del GAIC con distribución BCCG se diferenciaba mínimamente del GAIC con distribución BCPE, como se puede ver en la Tabla 5.28. Recuérdese que, la distribución BCPE con  $\tau = 2$  constante, se reduce a la distribución BCCG.

Distribución	GAIC
BCCG	25433.80
BCPE	25434.07
NO	25450.63

Tabla 5.28: Resumen de los valores de GAIC para el ajuste mediante p-splines en la talla en niños.

En la Tabla 5.29 se presentan los valores estimados de los grados de libertad para cada parámetro  $(df_\mu, df_\sigma, df_\nu, df_\tau)$ , procedentes de la función `find.hyper`. Para esta búsqueda se partió de un modelo GAMLSS con distribución BCPE, como se hizo en el IMC y en el peso.

Penalización(#)	Valores iniciales	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	GAIC(#)
2	(15,15,15,15)	(7.63, 4.93, 5.63, 0.31)	25420.45
	(6,6,6,6)	(7.62, 4.93, 5.54, 0.33)	25420.45
	(2,2,2,2)	(7.62, 4.94, 5.58, 0.31)	25420.45
2.5	(15,15,15,15)	(6.89, 4.57, 4.60, 0.1)	25433.06
	(6,6,6,6)	(6.93, 4.56, 4.59, 0.1)	25433.06
	(2,2,2,2)	(6.91, 4.59, 4.55, 0.1)	25433.06
3	(15,15,15,15)	(6.33, 4.29, 3.70, 0.1)	25444.7
	(6,6,6,6)	(6.29, 4.28, 3.62, 0.1)	25444.7
	(2,2,2,2)	(6.01, 4.25, 0.93, 0.1)	25444.45
$\log(n) =$ 8.254	(15,15,15,15)	(3.08, 2.67, 0.1, 0.1)	25529.24
	(6,6,6,6)	(3.08, 2.67, 0.1, 0.1)	25529.24
	(2,2,2,2)	(3.08, 2.68, 0.1, 0.1)	25529.24

Tabla 5.29: Resultados de la función `find.hyper` para el ajuste de la talla en los niños.

Para las penalizaciones 2.5, 3 y  $\log(n)$  los grados de libertad estimados con el procedimiento automático alcanzaron el límite inferior de la búsqueda en los parámetros que modelizan la curtosis y, en el caso de la penalización  $\log(n)$ , también la asimetría. Por lo tanto, son necesarias varias búsquedas adicionales, cuyos resultados se muestran en la Tabla 5.30.

Penalización(#)	$df_\tau$	$(df'_\mu, df'_\sigma, df'_\nu)$	GAIC(#)
2.5	2	(6.92, 4.57, 4.44)	25433.06
	1	(7, 4.52, 4.61)	25434.35
3	2	(6.05, 4.27, 0.8)	25444.42
	1	(5.93, 4.23, 0.7)	25445.16
Penalización(#)	$(df_\nu, df_\tau)$	$(df'_\mu, df'_\sigma)$	GAIC(#)
8.254	(2,2)	(3.06, 2.56)	25528.34
	(1,2)	(2.63, 2.5)	25533.17
	(2,1)	(2.87, 2.6)	25523.03
	(1,1)	(2.29, 2.55)	25528.24

Tabla 5.30: Resultados de la función `find.hyper` del proceso de búsqueda adicional, para el ajuste de la talla en niños. Se subraya en amarillo el valor mínimo del criterio GAIC.

Siguiendo los resultados del ajuste previo para varias distribuciones con p-splines, la curtosis debía poder simplificarse a una constante. Solamente para la penalización  $\log(n)$  los resultados indicaron que  $df_\nu = 2$  y  $df_\tau = 1$ . Para el caso de las otras penalizaciones, se obtuvo que  $\tau$  se ajustaba mediante un término lineal ( $df_\tau = 2$ ). El GAIC(2.5) del modelo BCPE con `cs()` tuvo un valor ligeramente inferior al GAIC(2.5) de BCCG con `pb()` por lo que se consideró adecuado utilizar la distribución BCPE en el ajuste con `cs()` En la Tabla 5.31 se muestran los grados de libertad redondeados con los que se ajustaron los modelos

Penalización	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	$(df_\mu, df_\sigma, df_\nu, df_\tau)$	Resumen de los residuos			
			Media	Varianza	Coef.Asim.	Coef.Curt.
2	(8,5,6,0)	(10,7,8,2)	0.0002	1.0003	-0.0002	3.0285
2.5	(7,5,4,0)	(9,7,6,2)	0.0002	1.0003	-0.0004	3.0329
3	(6,4,1,0)	(8,6,3,2)	0.0003	1.0003	-0.0007	3.0417
8.254	(3,3,0,0)	(5,5,2,1)	0.0008	1.0003	-0.0001	3.0388

Tabla 5.31: Tabla de resumen residual.

En el caso del modelo BCPE (5,5,2,1), el valor ajustado para  $\tau = 2.07 \sim 2$ , por lo que es equivalente a ajustar un modelo de Cole tal que BCCG(5,5,2).

Analizando los valores del `Q.stats` de este modelo, se observaron insuficiencias para la media ( $g = 15$ ) y la curtosis ( $g = 2$ ). Además, el modelo era inadecuado puesto que no se ajustaba bien para la media ni para el coeficiente de asimetría. Esto se puede comprobar en la significación de los estadísticos  $Q_1$  y  $Q_3$ , respectivamente, de la Tabla 5.32.

Por otro lado, en el modelo BCPE (8,6,3,2), también se observaron varias insuficiencias, para la curtosis ( $g = 2, 6$ ) y para la asimetría ( $g = 8$ ). Además, para la asimetría el estadístico  $Q$  era significativo, lo que indica que el modelo no se ajusta bien para ese parámetro. Por tanto, también se descarta este modelo de la elección.

Finalmente, los modelos BCPE (10,7,8,2) y BCPE (9,7,6,2), presentaban solamente una insuficiencia para la curtosis ( $g = 6$ ), pero todos los estadísticos  $Q$  eran no significativos, por lo tanto ambos modelos son válidos. Siguiendo el criterio de la sencillez, el modelo BCPE (9,7,6,2) se eligió para representar las curvas percentiles para la talla en niños de 6 a 15 años.

Modelos	p-valores de $Q$			
	$\mu$	$\sigma$	$\nu$	$\tau$
BCPE (10,7,8,2)	0.73	0.95	0.15	0.38
BCPE (9,7,6,2)	0.70	0.95	0.12	0.35
BCPE (8,6,3,2)	0.63	0.85	0.04	0.26
BCCG (5,5,2)	0.05	0.55	0.05	0.12

Tabla 5.32: p-valores del estadístico  $Q$  de los modelos que ajustan la talla en los niños.

A continuación, en la figura 5.14, se presentan las gráficas de los parámetros del modelo ajustado respecto a la edad.

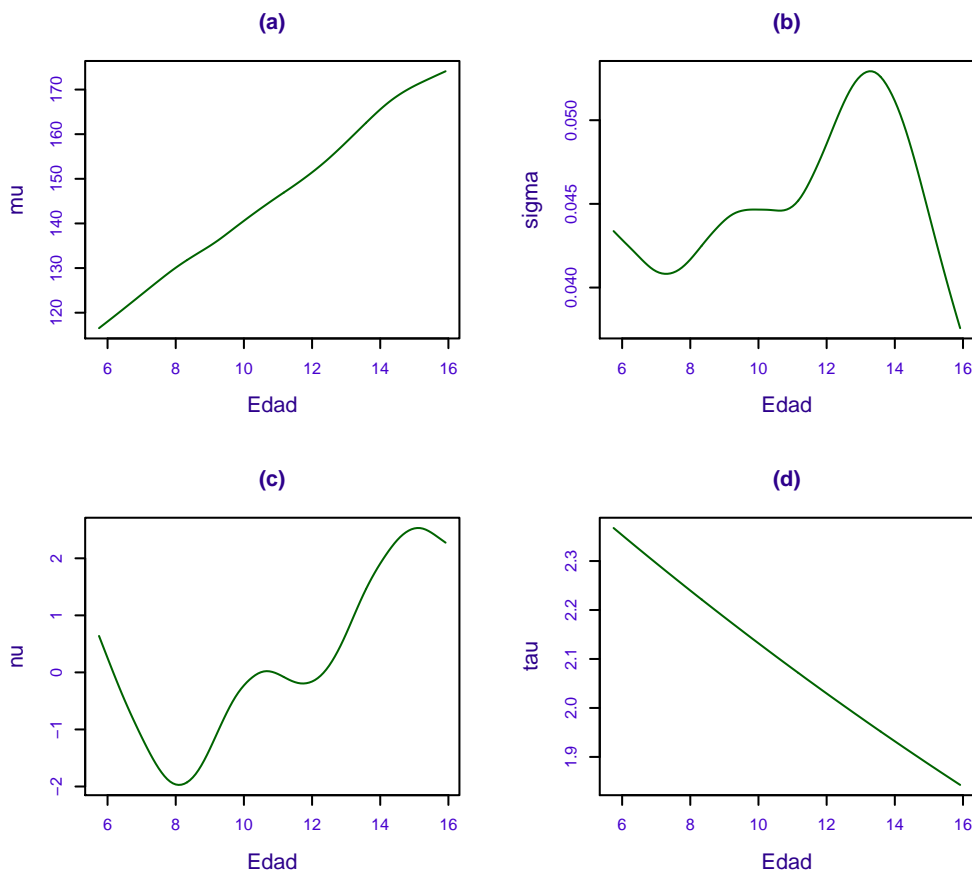


Figura 5.14: Gráfico de los parámetros  $\mu$  (a),  $\sigma$  (b),  $\nu$  (c) y  $\tau$  (d) ajustados para el modelo BCPE (9,7,6,2), que ajusta la talla en los niños en función de la edad.

El modelo ajustado para  $\mu$ , indica que la mediana para la talla en los niños, aumenta progresivamente con la edad desde los 120 cm a los 170 cm aproximadamente. Para  $\sigma$ , el modelo ajustado indica que oscila entre los valores (0.04,0.045) desde los 6 hasta los 11 años, a partir de ahí se produce un pequeño aumento hasta los 14 años, donde se alcanza la mayor variabilidad y, finalmente, se produce un descenso progresivo hasta los 15 años. La distribución para de la talla en niños de 6 años y desde los 11 a los 15 años es negativamente asimétrica ( $\nu > 0$ ), exceptuando a los 13 años que es la asimetría de la de una normal ( $\nu = 0$ ), entre los 7 y los 10 años la asimetría es positiva ( $\nu < 0$ ). El parámetro ajustado  $\tau$ , indica que la curtosis disminuye progresivamente tomando valores cercanos a 2 (distribución normal) en todo el intervalo de edad.

Finalmente, con efectos comparativos se muestran las gráficas de los parámetros de la distribución BCPE con `cs()` y de la BCCG con `pb()`

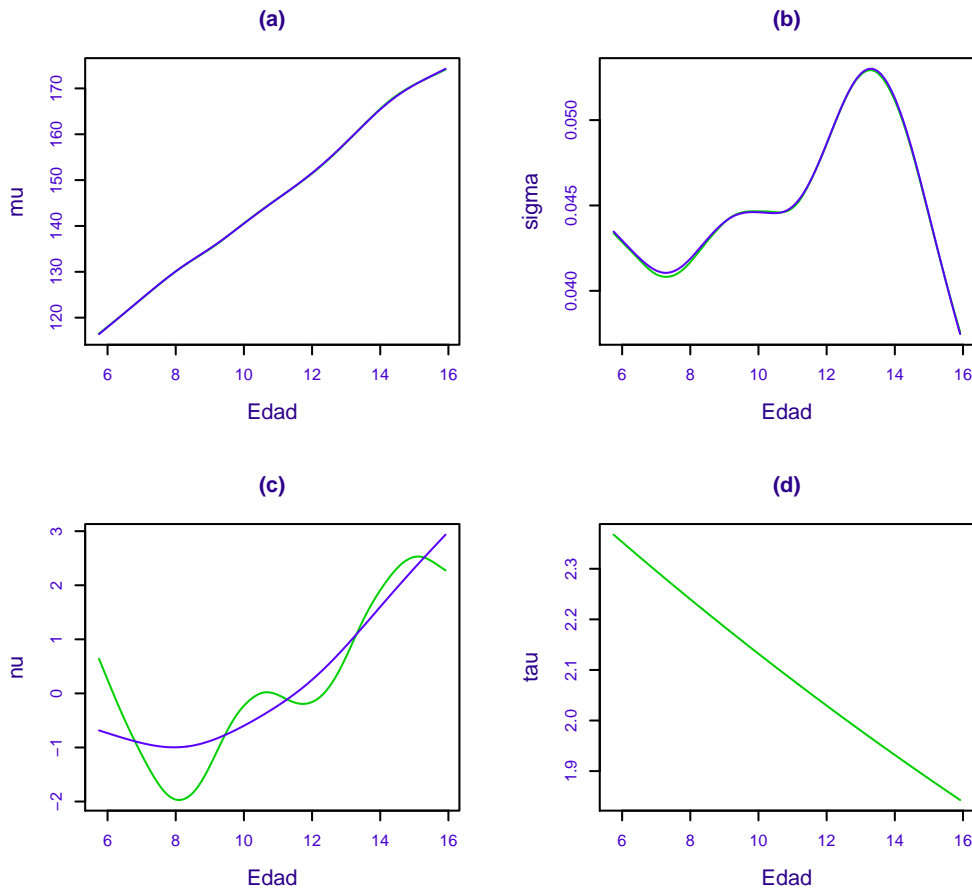


Figura 5.15: Gráfico de los parámetros  $\mu$  (a),  $\sigma$  (b),  $\nu$  (c) y  $\tau$  (d) ajustados para el modelo BCPE (9,7,6,2) en verde y para el modelo BCCG (8,7,3) en azul, que ajusta la talla en los niños en función de la edad.

El modelo BCCG ajustado con p-spines, estimó que los siguientes grados de libertad:  $df_{\mu} = 8$ ,  $df_{\sigma} = 7$  y  $df_{\nu} = 3$ . En la Figura 5.15, se puede ver que los parámetros ajustados  $\mu$  y  $\sigma$  son prácticamente iguales para ambos modelos. Sin embargo, se produce una variación significativa en el caso del parámetro  $\nu$

del modelo BCCG con p-splines con respecto al modelo BCPE con splines cúbicos de suavizado. En el caso del modelo BCCG (8,7,3), los valores de  $\nu$  a lo largo de la edad están dentro del rango en el que se movían los valores de este parámetro en el modelo BCPE (9,7,6,2), pero no se produce tanta oscilación. Los grados de libertad con los que se ajusta el parámetro que regula la asimetría en el caso del modelo BCCG con p-splines son  $df_\nu = 3$ , tres puntos por debajo de los que se utilizan en el otro caso, además los p-splines tienden a proporcionar curvas más suaves, por estos motivos, la curva que dibuja el parámetro ajustado para ese caso es mucho más sencilla.

### 5.6. Talla en niñas

Para la talla de las niñas, se eliminó el único atípico extremo había. El ajuste mediante p-splines determinó que, la distribución adecuada para el ajuste de la talla en las niñas es la BCCG (Tabla 5.33).

Distribución	GAIC
BCCG	23154.48
NO	23155.13
BCPE	23157.55

Tabla 5.33: Resumen de los valores de GAIC, con penalización 2.5, para el ajuste mediante p-splines en la talla en niñas.

En la Tabla 5.34 se presentan los valores estimados de los grados de libertad para cada parámetro  $(df_\mu, df_\sigma, df_\nu, df_\tau)$ , procedentes de la función `find.hyper`, partiendo de la distribución BCPE.

Penalización(#)	Valores iniciales	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	GAIC(#)
2	(15,15,15,15)	(5.21, 3.95, 0.1, 0.1)	23148.88
	(6,6,6,6)	(5.20, 3.94, 0.1, 0.1)	23148.88
	(2,2,2,2)	(5.28, 4.05, 0.1, 0.1)	23148.89
2.5	(15,15,15,15)	(4.73, 2.76, 0.1, 0.1)	23157.16
	(6,6,6,6)	(4.75, 2.75, 0.1, 0.1)	23157.16
	(2,2,2,2)	(4.68, 2.59, 0.1, 0.1)	23157.17
3	(15,15,15,15)	(4.49, 1.55, 0.1, 0.1)	23164.5
	(6,6,6,6)	(4.44, 1.56, 0.1, 0.1)	23164.5
	(2,2,2,2)	(4.44, 1.57, 0.1, 0.1)	23164.5
log(n) = 8.187	(15,15,15,15)	(3.37, 0.88, 0.1, 0.1)	23232.38
	(6,6,6,6)	(3.37, 0.88, 0.1, 0.1)	23232.38
	(2,2,2,2)	(3.36, 0.88, 0.1, 0.1)	23232.38

Tabla 5.34: Resultados de la función `find.hyper` que ajusta la talla para las niñas.

En este caso, los grados de libertad estimados con el procedimiento automático para las 4 penalizaciones, alcanzan el límite inferior de la búsqueda de los grados de libertad para la asimetría y curtosis.

Por lo tanto, para cada penalización se realizó una búsqueda adicional para verificar si el modelo se puede simplificar más para solo un término lineal o simplemente un término constante en la asimetría y/o la curtosis.

En la Tabla 5.35 se muestran los resultados óptimos de todas las combinaciones posibles de  $(df_\nu, df_\tau)=(i, j)$ , con  $i, j = 1, 2$ .

Penalización(#)	$(df_\nu, df_\tau)$	$(df'_\mu, df'_\sigma)$	GAIC(#)
2	(2,2)	(5.17, 3.85)	23148.55
	(1,2)	(5.15, 3.8)	23148.17
	(2,1)	(5.2, 3.85)	23146.36
	(1,1)	(5.17, 3.86)	23146.25
2.5	(2,2)	(4.77, 2.77)	23156.73
	(1,2)	(4.73, 2.76)	23155.84
	(2,1)	(4.73, 2.78)	23154.31
	(1,1)	(4.72, 2.79)	23153.43
3	(2,2)	(4.43, 1.48)	23163.98
	(1,2)	(4.42, 1.49)	23162.9
	(2,1)	(4.43, 1.48)	23161.09
	(1,1)	(4.42, 1.48)	23159.71
log(n) = 8.187	(2,2)	(3.36, 0.78)	23230.89
	(1,2)	(3.39, 0.79)	23224.34
	(2,1)	(3.38, 0.78)	23222.79
	(1,1)	(3.38, 0.78)	23216.25

Tabla 5.35: Resultados de la función `find.hyper` de las búsquedas adicionales para el ajuste de la talla en niñas. Se subraya en amarillo el valor mínimo del criterio GAIC. Se subraya en amarillo el valor mínimo del criterio GAIC.

Con estas búsquedas adicionales se comprobó que tanto la asimetría como la curtosis eran constantes ya que, para cada penalización el modelo con  $(df_\nu, df_\tau)=(1, 1)$  era el que minimizaba el criterio GAIC.

Como ya se ha comentado, un modelo BCPE se puede reducir a un modelo de Cole si  $\tau = 2$  o a un modelo normal NO si además  $\nu = 1$ . Por lo tanto, se comprobó si el modelo BCPE del que partimos se podía reducir a una distribución BCCG o a una distribución normal NO, para ello se ajustaron los modelos con los hiperparámetros redondeados que se muestran en la Tabla 5.36.

Penalización	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	Modelos ajustados	$\nu$	$\tau$
2	(5,4,0,0)	BCPE (7,6,1,1)	0.35	1.91
2.5	(5,3,0,0)	BCPE (7,5,1,1)	0.36	1.904
3	(4,1,0,0)	BCPE (6,3,1,1)	0.37	1.91
8.187	(3,1,0,0)	BCPE (5,3,1,1)	0.37	1.91

Tabla 5.36: Valores ajustados de los parámetros  $\nu$  y  $\tau$  de los modelos ajustados para la talla en las niñas.

Para cada penalización, los valores del parámetro que regula la curtosis de los modelos ajustados son muy próximos a dos, indicando que el modelo BCPE se puede reducir al modelo de Cole, es decir, con distribución BCCG. Sin embargo, el parámetro  $\nu$  toma valores distintos de uno, lo que indica que el modelo no se reduce a un modelo normal. En resumen, se ajustaron cuatro modelos de Cole, con asimetría constante, cuyo resumen residual se muestra en la tabla 5.37.

			Resumen de los residuos			
Penalización	$(df'_\mu, df'_\sigma, df'_\nu)$	$(df_\mu, df_\sigma, df_\nu)$	Media	Varianza	Coef.Asim.	Coef.Curt.
2	(5,4,0)	(7,6,1)	7.8e-5	1.0002	-0.0008	3.0865
2.5	(5,3,0)	(7,5,1)	0.0003	1.0002	-0.0012	3.0873
3	(4,1,0)	(6,3,1)	0.0003	1.0002	-0.0016	3.0911
8.187	(3,1,0)	(5,3,1)	7.8e-05	1.0002	-0.0008	3.0865

Tabla 5.37: Tabla de resumen residual.

Según los resultados de la función `Q.stats`, el modelo más sencillo BCCG(5,3,1) presentaba varias insuficiencias: dos para la media ( $g = 12, 13$ ), una para el coeficiente de variación ( $g = 8$ ) y uno para la curtosis ( $g = 5$ ). Además, como se puede ver en la Tabla 5.38, el modelo no se ajusta bien para la media, ya que el estadístico  $Q_1$  es significativo. Debido a estas irregularidades, el modelo se eliminó de la selección.

Por otro lado, los modelos BCCG (7,6,1), BCCG (7,5,1) y BCCG (6,3,1) presentaban una irregularidad en la curtosis ( $g = 5$ ). Además, el modelo BCCG (6,3,1) poseía otra insuficiencia para el coeficiente de variación ( $g = 8$ ). Sin embargo, todos estos modelos se ajustaban bien para todos los parámetros, como indica la no significación de los estadísticos  $Q_j$  con  $j = 1, 2, 3, 4$ .

Modelos	p-valores de $Q$			
	$\mu$	$\sigma$	$\nu$	$\tau$
BCCG(7,6,1)	0.53	0.68	0.84	0.35
BCCG(7,5,1)	0.53	0.60	0.83	0.34
BCCG(6,3,1)	0.37	0.28	0.84	0.34
BCCG(5,3,1)	0.08	0.28	0.84	0.34

Tabla 5.38: p-valores del estadístico  $Q$  de los modelos que ajustan la talla en niñas.



Por lo tanto, de entre los modelos válidos, se seleccionó el modelo BCCG(6,3,1) para ajustar las curvas percentiles en función de la edad para la talla en las niñas, ya que era el más sencillo, de entre los modelos válidos.

A continuación, en la figura 5.16, se presentan las gráficas de los parámetros del modelo ajustado respecto a la edad.

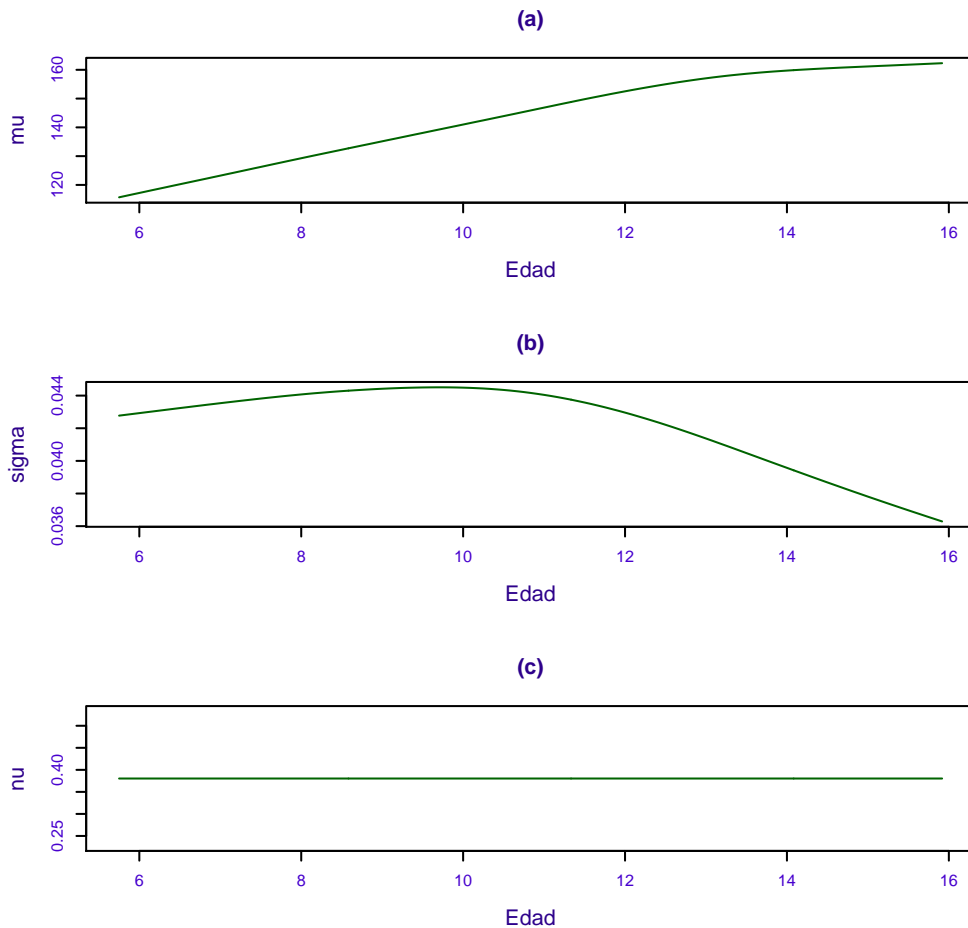


Figura 5.16: Gráfico de los parámetros  $\mu$  (a),  $\sigma$  (b) y  $\nu$  (c) ajustados para el modelo BCCG(6,3,1), que ajusta la talla en las niñas en función de la edad.

El modelo ajustado para  $\mu$ , como para el caso del ajustado para los niños, indica que la mediana de la talla en las niñas aumenta con la edad, ralentizándose este crecimiento a partir de los 13 años. Para  $\sigma$  el modelo ajustado indica que el coeficiente de variación aumenta lentamente desde los 6 hasta los 10 años, donde se alcanza la mayor variabilidad y luego disminuye progresivamente con la edad hasta los 16 años. Por último,  $\nu = 0.37$  constante para todas las edades, es decir, la asimetría es negativa para todo el intervalo de edad.

### 5.7. Perímetro de cintura en niños

Se identificaron 6 valores atípicos extremos en el perímetro de cintura de los niños, que se eliminaron de la muestra. Según el ajuste de los modelos mediante p-splines, la distribución adecuada para esta variable en niños es la distribución BCPE (Tabla 5.39).

Distribución	GAIC
BCPE	26635.54
BCCG	26689.32
NO	27531

Tabla 5.39: Resumen de los valores de GAIC, con penalización 2.5, para el ajuste mediante p-splines en el parímetro de cintura en niños.

En la Tabla 5.40 se presentan los valores estimados de los grados de libertad para cada parámetro  $(df_\mu, df_\sigma, df_\nu, df_\tau)$ , procedentes de la función `find.hyper`.

Penalización(#)	Valores iniciales	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	GAIC(#)
2	(15,15,15,15)	(2.50, 2.62, 1.72, 1.4)	26626.59
	(6,6,6,6)	(2.51, 2.61, 1.73, 1.41)	26626.59
	(2,2,2,2)	(2.52, 2.63, 1.72, 1.43)	26626.58
2.5	(15,15,15,15)	(2.25, 2.43, 1.61, 1.28)	26634.58
	(6,6,6,6)	(2.25, 2.42, 1.61, 1.28)	26634.58
	(2,2,2,2)	(2.25, 2.43, 1.61, 1.28)	26634.58
3	(15,15,15,15)	(2.08, 2.31, 1.53, 1.12)	26642.32
	(6,6,6,6)	(2.09, 2.3, 1.51, 1.17)	26642.29
	(2,2,2,2)	(2.08, 2.30, 1.51, 1.16)	26642.29
log(n) = 8.2624	(15,15,15,15)	(1.23, 1.68, 0.82, 0.51)	26713.75
	(6,6,6,6)	(1.22, 1.69, 0.82, 0.51)	26713.75
	(2,2,2,2)	(1.23, 1.69, 0.82, 0.50)	26713.75

Tabla 5.40: Resultados de la función `find.hyper` que ajusta el perímetro de cintura en los niños.

Estos resultados se redondearon para ajustar los modelos que se muestran en la Tabla 5.41.

Penalización	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	$(df_\mu, df_\sigma, df_\nu, df_\tau)$	Resumen de los residuos			
			Media	Varianza	Coef.Asim.	Coef.Curt.
2	(2,3,2,1)	(4,5,4,3)	-0.0137	1.00006	0.02	3.0074
2.5	(2,2,2,1)	(4,4,4,3)	-0.0143	0.9994	0.02	3.0088
3	(2,2,1,1)	(4,4,3,3)	-0.0135	0.9996	0.0254	3.0109
8.2624	(1,2,1,0)	(3,4,3,2)	-0.0140	0.9996	0.0225	3.0143

Tabla 5.41: Tabla de resumen residual.

La función `Q.stats` indicó que el modelo BCPE (3,4,3,2) poseía dos irregularidades, una para el coeficiente de variación ( $g = 1$ ) y otra en el para la curtosis ( $g = 14$ ). Además, los valores de los estadísticos  $Z_{g,1}$  (para la media) eran grandes aunque no lo suficiente para considerarse insuficiencias, pero esto se reflejó en la significación del estadístico  $Q_1$  de la media, indicando que el modelo no se ajustaba bien para  $\mu$ . El modelo se omitió del proceso de selección.

Por otro lado, los modelos BCPE (4,5,4,3), BCPE (4,4,4,3) y BCPE (4,4,3,3), presentaban un único valor irregular en el octavo grupo de edad, que no impidió que fueran modelos válidos, ya que sus estadísticos  $Q$  son eran significativos. Por lo tanto, el modelo que se seleccionó fue BCPE (4,4,3,3), el más sencillo.

Modelos	p-valores de $Q$			
	$\mu$	$\sigma$	$\nu$	$\tau$
BCPE (4,5,4,3)	0.23	0.53	0.88	0.25
BCPE (4,4,4,3)	0.21	0.44	0.87	0.25
BCPE (4,4,3,3)	0.20	0.42	0.76	0.25
BCPE (3,4,3,2)	0.07	0.39	0.8	0.13

Tabla 5.42: p-valores del estadístico  $Q$  de los modelos modelos que ajustan el perímetro de cintura en niños.

A continuación, en la figura 5.17, se presentan las gráficas de los parámetros del modelo ajustado respecto a la edad.

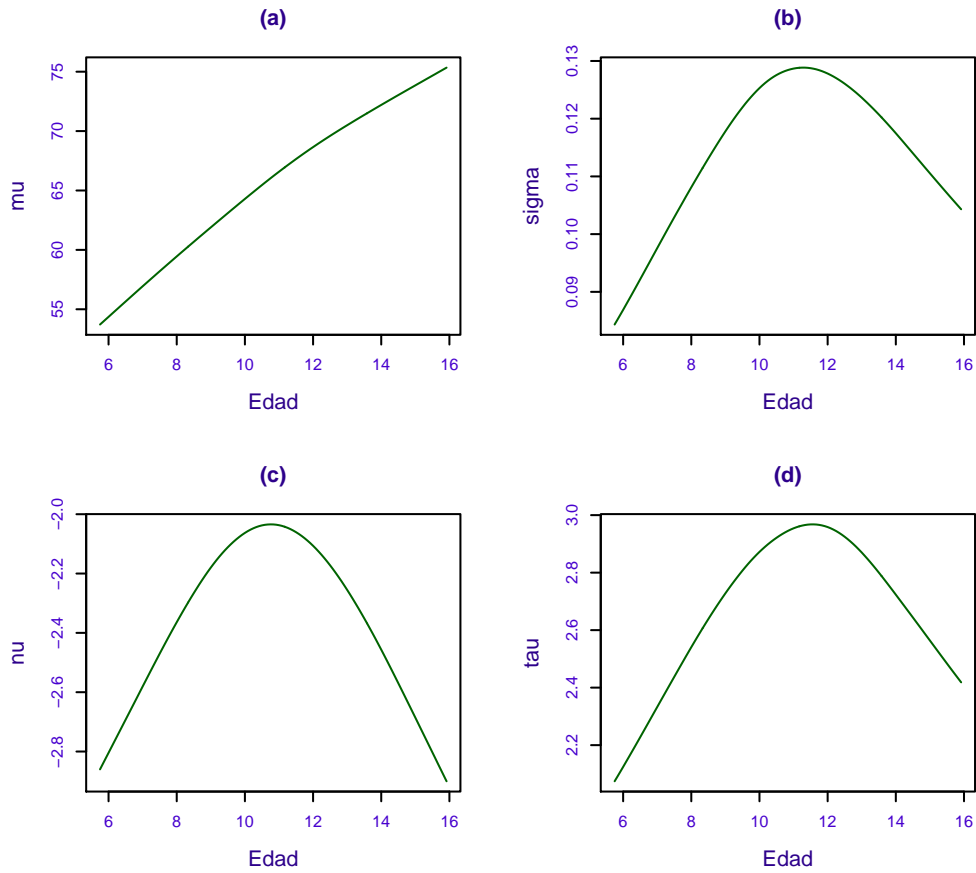


Figura 5.17: Gráfico de los parámetros  $\mu$  (a),  $\sigma$  (b),  $\nu$  (c) y  $\tau$  (d) ajustados para el modelo BCPE (4,4,3,3), que ajusta el perímetro de cintura en los niños en función de la edad.

El modelo ajustado para  $\mu$  indica que la mediana del perímetro de cintura en niños aumenta progresivamente con la edad desde los 55 cm hasta los 75 cm, aproximadamente. Para  $\sigma$ , el modelo ajustado aumenta progresivamente con la edad hasta los 11 años, donde alcanza la mayor variabilidad, luego disminuye lentamente hasta los 15 años, todos los valores son muy cercanos a cero. La distribución del perímetro de cintura en niños es asimétricamente positiva ( $\nu < 0$ ), esta asimetría disminuye progresivamente con la edad desde los 6 hasta los 11 años, luego vuelve a aumentar con la edad hasta los 15 años. Por último, para  $\tau$  el modelo ajustado indica que la distribución es platicúrtica ( $\tau > 2$ ) para todas las edades,  $\tau$  aumenta hasta los 11 años, donde se alcanza el mayor valor y luego vuelve a disminuir lentamente hasta los 15 años.

### 5.8. Perímetro de cintura en niñas

En el perímetro de cintura en niñas se identificaron 2 valores atípicos extremos, que se eliminaron de la muestra. En la Tabla 5.43, se refleja que la distribución adecuada en este caso, nuevamente, la BCPE.

Distribución	GAIC
BCPE	24675.19
BCCG	24694.61
NO	25238.90

Tabla 5.43: Resumen de los valores de GAIC, con penalización 2.5, para el ajuste mediante p-splines en el parámetro de cintura en niñas.

En la Tabla 5.44 se presentan los valores estimados de los grados de libertad para cada parámetro  $(df_\mu, df_\sigma, df_\nu, df_\tau)$ , procedentes de la función `find.hyper`.

Penalización(#)	Valores iniciales	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	GAIC(#)
2	(15,15,15,15)	(3.64, 2.32, 1.04, 1.37)	24666.66
	(6,6,6,6)	(3.65, 2.31, 1.034, 1.36)	24666.66
	(2,2,2,2)	(3.64, 2.31, 1.03, 1.36)	24666.66
2.5	(15,15,15,15)	(3.25, 2.09, 0.91, 1.17)	24674.64
	(6,6,6,6)	(3.29, 2.08, 0.91, 1.16)	24674.64
	(2,2,2,2)	(3.24, 2.09, 0.90, 1.17)	24674.64
3	(15,15,15,15)	(2.94, 1.95, 0.8, 1.04)	24682.21
	(6,6,6,6)	(2.93, 1.94, 0.79, 1.03)	24682.21
	(2,2,2,2)	(2.95, 1.97, 0.74, 1.03)	24682.22
$\log(n) = 8.1867$	(15,15,15,15)	(1.61, 1.35, 0.1, 0.36)	24748.33
	(6,6,6,6)	(1.62, 1.35, 0.1, 0.36)	24748.33
	(2,2,2,2)	(1.62, 1.35, 0.1, 0.36)	24748.33

Tabla 5.44: Resultados de la función `find.hyper` que ajusta el perímetro de cintura en las niñas.

La tabla 5.45 muestra los resultados de la búsqueda adicional que se realizó para comprobar si  $df_\nu = 1$  ó  $df_\nu = 2$  en la penalización  $\log(n)$

Penalización(#)	$(df_\nu)$	$(df'_\mu, df'_\sigma, df'_\tau)$	GAIC(#)
8.254	2	(1.6, 1.34, 0.38)	24748.03
	1	(1.56, 1.33, 0.45)	24741.94

Tabla 5.45: Resultados de la función `find.hyper` del proceso de búsqueda adicional, para el ajuste del perímetro de cintura en niñas. Se subraya en amarillo el valor mínimo del criterio GAIC.

Los valores redondeados con los que se ajustaron los modelos se muestran en la Tabla 5.46. Nótese que para las penalizaciones 2.5 y 3 se obtuvieron los mismos grados de libertad redondeados para el ajuste.

			Resumen de los residuos			
Penalización	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	$(df_\mu, df_\sigma, df_\nu, df_\tau)$	Media	Varianza	Coef.Asim.	Coef.Curt.
2	(4,2,1,1)	(6,4,3,3)	-0.0046	1.00001	0.011	2.9909
2.5	(3,2,1,1)	(5,4,3,3)	-0.0044	1.00001	0.0108	0.9992
3	(3,2,1,1)	(5,4,3,3)	-0.0044	1.00001	0.0108	0.9992
8.254	(2,1,0,0)	(4,3,1,2)	-0.0045	1.00001	0.0124	2.9873

Tabla 5.46: Tabla de resumen residual.

Todos los modelos ajustados presentaron insuficiencias en el segundo, cuarto y séptimo grupo de edad para varianza, curtosis y asimetría, respectivamente. Además para el modelo BCPE(4,3,1,2) se identificó otra insuficiencia en el primer intervalo de edad para la asimetría. En la Tabla 5.51 se ve reflejado que los modelos BCPE(6,4,3,3) y BCPE (5,4,3,3) son válidos, puesto que los estadísticos  $Q_j$  son todos no significativos, sin embargo el modelo BCPE (4,3,1,2) no lo es, ya que no se ajusta bien para la asimetría ni la curtosis. Además el test de Agostino rechazó, para este modelo, la normalidad. Por lo tanto, el modelo BCPE (4,3,1,2) se eliminó del proceso.

El modelo más sencillo de los dos válidos, BCPE(5,4,3,3), es el que se utilizó para estimar las curvas percentiles del perímetro de cintura en las niñas de 6 a 15 años.

Modelos	p-valores de $Q$			
	$\mu$	$\sigma$	$\nu$	$\tau$
BCPE (6,4,3,3)	0.83	0.69	0.12	0.45
BCPE (5,4,3,3)	0.76	0.70	0.13	0.45
BCPE (4,3,1,2)	0.46	0.31	0.04	0.097

Tabla 5.47: p-valores del estadístico  $Q$  de los modelos que ajusta el perímetro de cintura en las niñas.

A continuación, en la figura 5.18, se presentan las gráficas de los parámetros del modelo ajustado respecto a la edad.

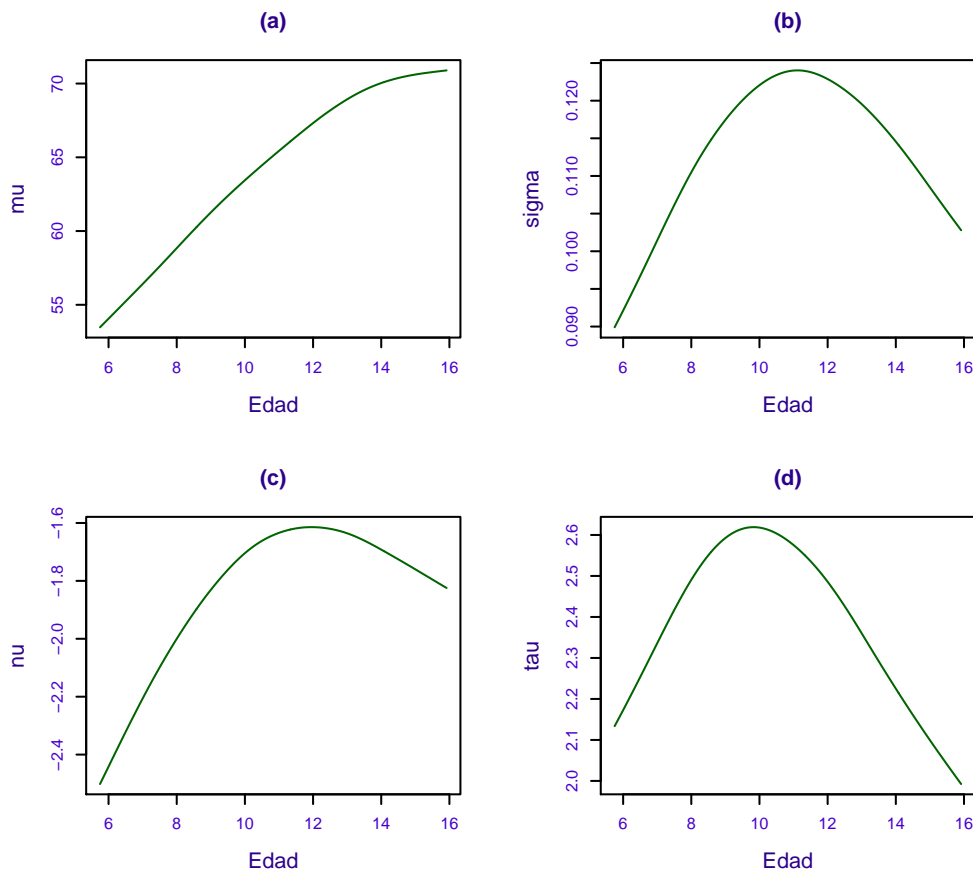


Figura 5.18: Gráfico de los parámetros  $\mu$  (a),  $\sigma$  (b),  $\nu$  (c) y  $\tau$  (d) ajustados para el modelo BCPE (5,4,3,3), que ajusta el perímetro de cintura en las niñas en función de la edad.

Como para el caso de los niños, el modelo ajustado para  $\mu$  indica que la mediana del perímetro de cintura en niñas aumenta progresivamente con la edad, en este caso desde los 55 cm hasta los 70 cm, aproximadamente. Para  $\sigma$  el recorrido es el mismo que para el modelo ajustado para los niños, pero en este caso los valores no son tan cercanos al cero. Para  $\nu$ , el modelo ajustado indica que la distribución del perímetro de cintura en niñas, como en niños, es asimétricamente positiva ( $\nu < 0$ ), disminuye progresivamente con la edad desde los 6 hasta los 11 años, luego vuelve a aumentar con la edad hasta los 15 años. Por último, para  $\tau$  el modelo ajustado indica que la distribución es platicúrtica ( $\tau > 2$ ) desde 6 hasta los 15 años, donde la curtosis aumenta hasta los 11 años, donde se alcanza el mayor valor y luego vuelve a disminuir lentamente hasta los 15 años, donde la distribución es la de una normal ( $\tau = 2$ ).

## 5.9. Perímetro de cadera en niños

Para el perímetro de cadera de los niños se identificaron 5 valores atípicos extremos, que se eliminaron de la muestra. El ajuste mediante p-splines determinó que la distribución adecuada para trabajar con esta variable en niños es la distribución BCPE (Tabla 5.48).

Distribución	GAIC
BCPE	26151.77
BCCG	26172.53
NO	26424.05

Tabla 5.48: Resumen de los valores de GAIC, con penalización 2.5, para el ajuste mediante p-splines en el perímetro de cadera en niños.

En la Tabla 5.49 se presentan los valores estimados de los grados de libertad para cada parámetro  $(df_\mu, df_\sigma, df_\nu, df_\tau)$ , procedentes de la función `find.hyper`.

Penalización(#)	Valores iniciales	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	GAIC(#)
2	(15,15,15,15)	(2.42, 5.47, 2.00, 6.27)	26142.48
	(6,6,6,6)	(2.3, 5.72, 2.03, 6.06)	26142.47
	(2,2,2,2)	(2.81 5.26, 1.64, 1.22)	26142.9
2.5	(15,15,15,15)	(2.67, 1.99, 1.10, 1.23)	26151.75
	(6,6,6,6)	(2.01, 4.55, 1.50, 4.54)	26153.71
	(2,2,2,2)	(2.05, 2.07, 1.49, 1.13)	26151.31
3	(15,15,15,15)	(1.82, 1.9, 1.32, 1.02)	26158.54
	(6,6,6,6)	(1.85, 1.92, 1.32, 1.03)	26158.54
	(2,2,2,2)	(1.84, 1.90, 1.33, 1.02)	26158.54
$\log(n) = 8.2527$	(15,15,15,15)	(1.04, 1.31, 0.62, 0.44)	26225.11
	(6,6,6,6)	(1.04, 1.31, 0.62, 0.44)	26225.1
	(2,2,2,2)	(1.04, 1.31 0.62, 0.43)	26225.11

Tabla 5.49: Resultados de la función `find.hyper` que ajusta el perímetro de cadera de los niños.

La tabla 5.50 muestra los valores redondeados de los hiperparámetros, con los que se ajustaron los modelos. Nótese que los modelos para las penalizaciones 2.5 y 3 coinciden.

			Resumen de los residuos			
Penalización	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	$(df_\mu, df_\sigma, df_\nu, df_\tau)$	Media	Varianza	Coef.Asim.	Coef.Curt.
2	(2,6,2,6)	(4,8,4,8)	0.0014	1.0014	0.0014	3.0088
2.5	(2,2,1,1)	(4,4,3,3)	0.0022	1.0004	0.0034	3.0163
3	(2,2,1,1)	(4,4,3,3)	0.0022	1.0004	0.0034	3.0163
8.2527	(1,1,1,0)	(3,3,3,2)	-0.0016	1.0004	0.0015	3.034

Tabla 5.50: Tabla de resumen residual.

Como se refleja en la Tabla 5.51, el modelo BCPE (3,3,3,2) no es válido, ya que solamente ajustaba bien la media. Por lo tanto se elimina del proceso de selección.

Los modelos BCPE (4,8,4,8) y BCPE (4,4,3,3) poseían una irregularidad para el quinto grupo de



edad en la curtosis y, a mayores, el BCPE (4,4,3,3) otra en el último grupo de edad para la asimetría. Ambos modelos eran válidos, ya que los estadísticos  $Q_j$  con  $j = 1, 2, 3, 4$  eran no significativos. El modelo más complejo presentaba una pequeña sobreestimación de la varianza, impidiendo que las curvas se vieran como curvas suaves, por lo que se seleccionó el modelo BCPE (4,4,3,3) para representar las curvas percentiles del perímetro de cadera en niños de 6 a 15 años.

Modelos	p-valores de $Q$			
	$\mu$	$\sigma$	$\nu$	$\tau$
BCPE (4,8,4,8)	0.48	0.80	0.18	0.38
BCPE (4,4,3,3)	0.45	0.31	0.12	0.34
BCPE (3,3,3,2)	0.21	0.07	0.07	0.08

Tabla 5.51: p-valores del estadístico  $Q$  de los modelos que ajustan el perímetro de cadera en los niños.

A continuación, en la figura 5.19, se presentan las gráficas de los parámetros del modelo ajustado respecto a la edad.

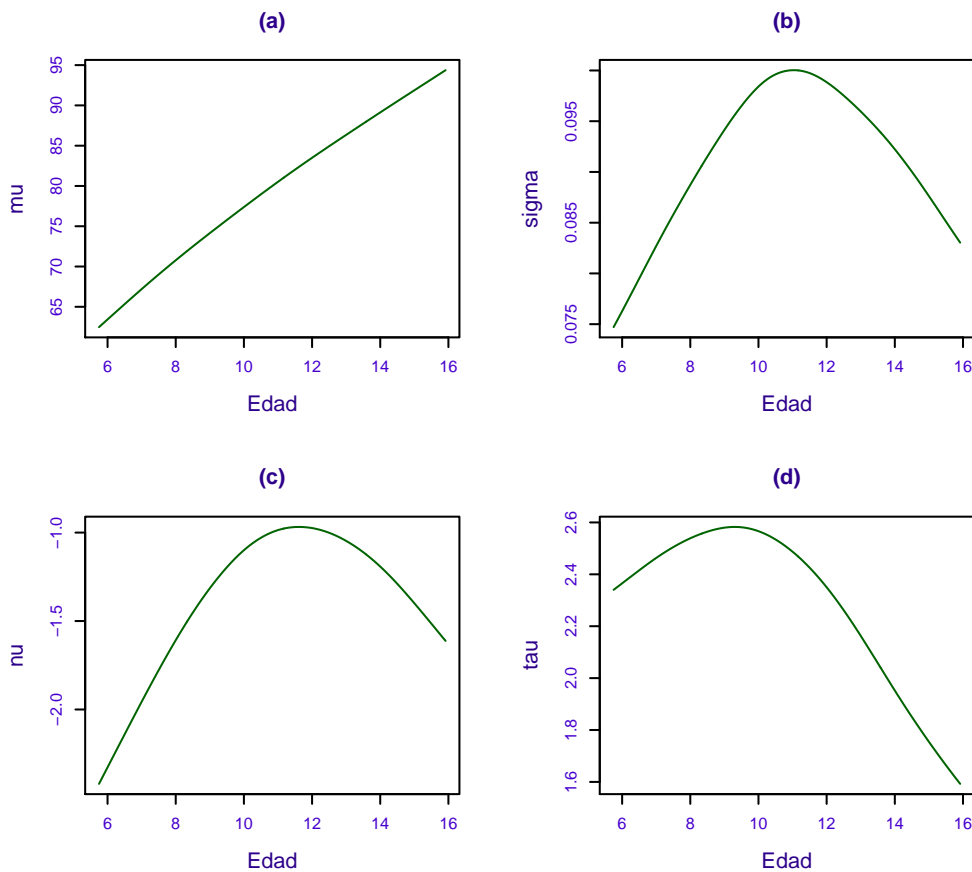


Figura 5.19: Gráfico de los parámetros  $\mu$  (a),  $\sigma$  (b),  $\nu$  (c) y  $\tau$  (d) ajustados para el modelo BCPE (4,4,3,3), que ajusta el perímetro de cadera en los niños en función de la edad.

El modelo ajustado para  $\mu$  indica que la mediana del perímetro de cadera en niños aumenta de forma uniforme con la edad desde los 65 cm hasta los 95 cm, aproximadamente. Para  $\sigma$ , el modelo ajustado aumenta progresivamente con la edad hasta los 11 años, donde alcanza la mayor variabilidad, luego disminuye lentamente hasta los 16 años. La distribución del perímetro de cadera en niñas es asimétricamente positiva ( $\nu < 0$ ), esta asimetría disminuye progresivamente con la edad desde los 6 hasta los 11 años, luego vuelve a aumentar con la edad hasta los 15 años. El valor de  $\tau$  aumenta desde los 6 hasta los 10 años y luego disminuye progresivamente con la edad hasta el final, indicando que la distribución del perímetro de cadera es platicúrtica ( $\tau > 2$ ) desde 6 hasta los 13 años, a los 14 años la distribución es la de una normal ( $\tau = 2$ ) y a los 15 años es leptocúrtica ( $\tau < 2$ ).

### 5.10. Perímetro de cadera en niñas

En esta variable en las niñas se habían encontrado 3 valores atípicos extremos, que se eliminaron de la muestra. Nuevamente, el ajuste mediante p-splines determinó que la distribución adecuada es la distribución BCPE (Tabla 5.52).

Distribución	GAIC
BCPE	24423.22
BCCG	24431.19
NO	24570.78

Tabla 5.52: Resumen de los valores de GAIC, con penalización 2.5, para el ajuste mediante p-splines en el perímetro de cadera en niñas.

En la Tabla 5.53 se presentan los valores estimados de los grados de libertad para cada parámetro ( $df_\mu, df_\sigma, df_\nu, df_\tau$ ), procedentes de la función `find.hyper`.

Penalización(#)	Valores iniciales	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	GAIC(#)
2	(15,15,15,15)	(4.81, 2.53, 0.55, 1.82)	24414.25
	(6,6,6,6)	(4.88, 2.56, 0.53, 1.82)	24414.25
	(2,2,2,2)	(4.84, 2.57, 0.5, 1.82)	24414.25
2.5	(15,15,15,15)	(4.42, 2.27, 0.38, 1.54)	24422.89
	(6,6,6,6)	(4.4, 2.27, 0.38, 1.53)	24422.89
	(2,2,2,2)	(4.43, 2.27, 0.38, 1.53)	24422.89
3	(15,15,15,15)	(4.17, 2.08, 0.23, 1.36)	24431.04
	(6,6,6,6)	(4.17, 2.08, 0.23, 1.34)	24431.04
	(2,2,2,2)	(4.18, 2.08, 0.25, 1.35)	24431.04
$\log(n) = 8.1864$	(15,15,15,15)	(2.69, 1.36, 0.1, 0.63)	24504.09
	(6,6,6,6)	(2.69, 1.36, 0.1, 0.63)	24504.09
	(2,2,2,2)	(2.70, 1.35, 0.1, 0.63)	24504.09

Tabla 5.53: Resultados de la función `find.hyper` que ajusta el perímetro de cadera en las niñas.

Los resultados de la búsqueda adicional necesaria para el hiperparámetro  $df_\nu$  se muestran en la Tabla 5.54.

Penalización(#)	$(df_\nu)$	$(df'_\mu, df'_\sigma, df'_\tau)$	GAIC(#)
8.1864	2	(2.72, 1.36, 0.63)	24503.71
	1	(2.55, 1.31, 0.78)	24513.83

Tabla 5.54: Resultados de la función `find.hyper` del proceso de búsqueda adicional, para el ajuste del perímetro de cadera en niñas. Se subraya en amarillo el valor mínimo del criterio GAIC.

Una vez más, el resultado de redondear es el mismo para las penaliaciones 2.5 y 3 y se muestra en la Tabla 5.55..

			Resumen de los residuos			
Penalización	$(df'_\mu, df'_\sigma, df'_\nu, df'_\tau)$	$(df_\mu, df_\sigma, df_\nu, df_\tau)$	Media	Varianza	Coef.Asim.	Coef.Curt.
2	(5,2,1,2)	(7,4,3,4)	0.0003	0.9997	0.0061	2.9928
2.5	(4,2,0,1)	(6,4,2,3)	0.0018	1.0001	0.0089	2.9926
3	(4,2,0,1)	(6,4,2,3)	0.0018	1.0001	0.0089	2.9926
8.1864	(2,1,0,1)	(4,3,2,3)	0.0034	0.9985	0.0061	2.9858

Tabla 5.55: Tabla de resumen residual.

El modelo BCPE (7,4,3,4) presentaba insuficiencias para la varianza ( $g = 10$ ) y la curtosis ( $g = 14$ ). Además de estas irregularidades, el modelo BCPE (6,4,2,3) presenta otra para la curtosis ( $g = 4$ ). Por otro lado, el modelo BCPE (4,3,2,3) también poseía todas estas irregularidades y a mayores otra en el último grupo de edad para la media. Sin embargo, los estadísticos  $Q_j$  para los modelos BCPE (7,4,3,4) y BCPE (6,4,2,3) era no significativos para  $j = 1, 2, 3, 4$ , como se refleja en la Tabla 5.56, al contrario que para el modelo BCPE (4,3,2,3), que se eliminó del proceso de selección.

Se presenta una situación en la que dos modelos similares sin válidos, por tanto, se eligió el modelos más sencillo.

Modelos	p-valores de $Q$			
	$\mu$	$\sigma$	$\nu$	$\tau$
BCPE (7,4,3,4)	0.88	0.25	0.44	0.16
BCPE (6,4,2,3)	0.81	0.24	0.37	0.11
BCPE (4,3,2,3)	0.18	0.05	0.32	0.10

Tabla 5.56: p-valores del estadístico  $Q$  de los modelos que ajustan el perímetro de cadera en las niñas.

A continuación, en la figura 5.20, se presentan las gráficas de los parámetros del modelo ajustado respecto a la edad.

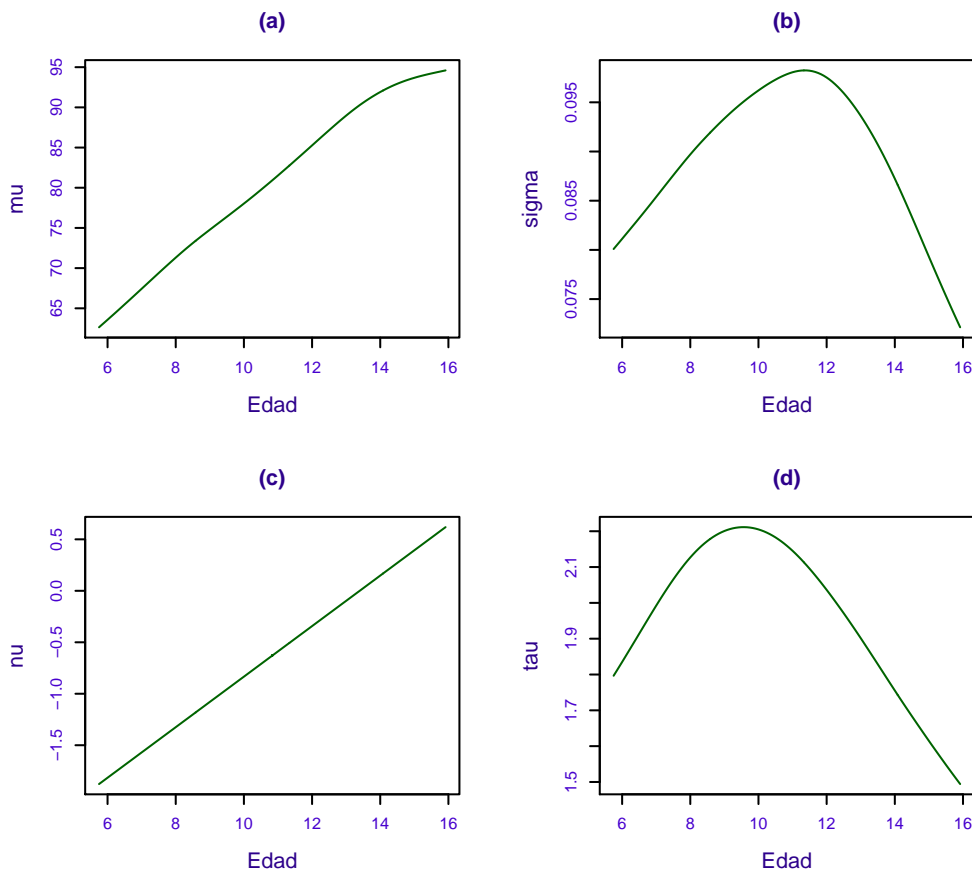


Figura 5.20: Gráfico de los parámetros  $\mu$  (a),  $\sigma$  (b),  $\nu$  (c) y  $\tau$  (d) ajustados para el modelo BCPE (6,4,2,3), que ajusta el perímetro de cadera en las niñas en función de la edad.

El modelo ajustado para  $\mu$ , como en el caso del modelo ajustado en los niños, indica que la mediana del perímetro de cadera en niñas aumenta progresivamente con la edad, en este caso a partir de los 13 años hay una leve ralentización de este aumento. Para  $\sigma$ , el modelo ajustado aumenta progresivamente con la edad hasta los 11 años, donde alcanza la mayor variabilidad, luego disminuye rápidamente hasta los 16 años. Para  $\nu$  se apreciaron diferencias con respecto al modelo de los niños, en este caso el modelo ajustado indica que  $\nu$  homogéneamente con la edad y la distribución del perímetro de cadera en niñas es asimétricamente positiva ( $\nu < 0$ ) desde los 6 hasta los 13 años y asimétricamente positiva desde los 14 hasta los 16 años ( $\nu \geq 0$ ). Con respecto al parámetro ajustado que regula la curtosis,  $\tau$ , indica que distribución es la de una normal para los 7 años y los 13 años ( $\tau = 2$ ), a los 6 años y desde los 14 hasta los 16 años es leptocúrtica ( $\tau < 2$ ) y desde los 8 hasta los 12 años es platicúrtica ( $\tau > 2$ ), la curtosis aumenta lentamente desde los 6 hasta los 10 años y luego disminuye progresivamente con la edad hasta los 16 años.

## Capítulo 6

# Curvas y tablas de percentiles

Con los modelos seleccionados en el capítulo anterior, para cada variable, se representaron las curvas percentiles y se construyeron las tablas de percentiles. Tanto en las curvas como en en las tablas se han recogido los valores medios, representados por el percentil 50 (P50), el resto de cuartiles (P25 y P75), los límites del rango de variación “normal” más utilizados en la clínica (P3 y P97) y los percentiles 10 y 90 (P10 y P90), además, en las tablas percentiles se añadieron los valores correspondientes a los percentiles P5 y P95. En las tablas se pueden ver los valores de los parámetros estimados (L, M, S y P) para cada edad (de medio año en medio año) y los valores de 1, 2 y 3 desviaciones estándar (SD) por debajo y por encima de la media.

6.1. IMC en niños

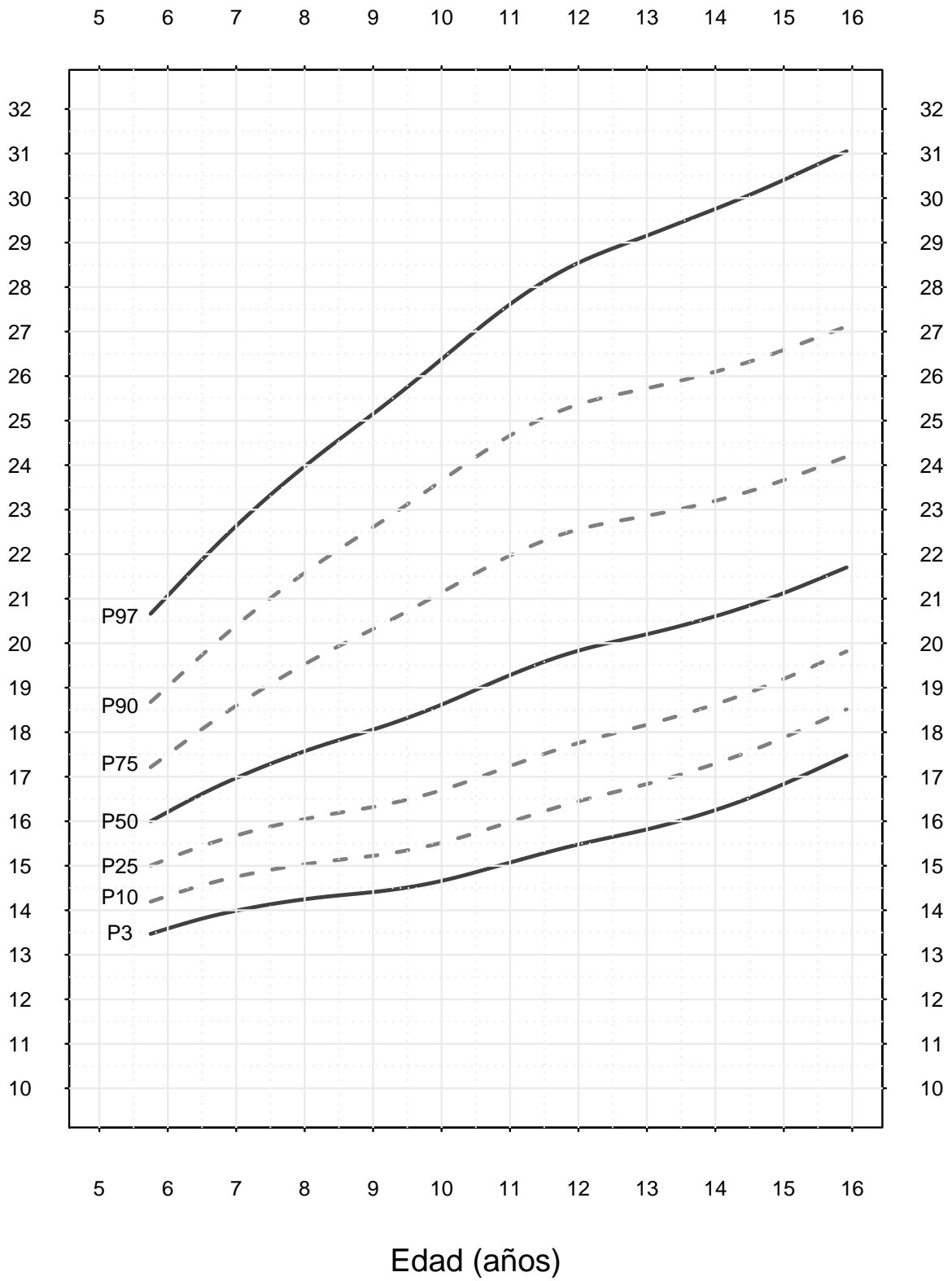


Figura 6.1: Curvas percentiles para el IMC (en kg/m<sup>2</sup>) de los niños con el modelo BCPE (6,4,3,4).

Edad	S	M	L	P	-3SD	-2SD	-1SD	1SD	2SD	3SD
6	0.110	16.208	-1.834	1.813	12.4	13.5	14.7	18.3	21.6	27.8
6.5	0.116	16.613	-1.757	1.987	12.7	13.7	15.0	18.9	22.4	28.5
7	0.122	16.972	-1.680	2.173	12.8	13.9	15.2	19.5	23.2	29.1
7.5	0.129	17.291	-1.602	2.365	13.0	14.0	15.3	20.1	23.9	29.7
8	0.135	17.573	-1.527	2.555	13.1	14.1	15.5	20.6	24.5	30.1
8.5	0.141	17.824	-1.456	2.729	13.2	14.2	15.6	21.1	25.1	30.6
9	0.147	18.060	-1.392	2.874	13.3	14.3	15.7	21.5	25.7	31.2
9.5	0.152	18.321	-1.339	2.987	13.4	14.4	15.8	22.0	26.3	31.9
10	0.156	18.619	-1.298	3.057	13.5	14.5	16.0	22.5	27.0	32.6
10.5	0.159	18.951	-1.272	3.072	13.7	14.7	16.3	22.9	27.6	33.5
11	0.160	19.282	-1.260	3.034	13.8	14.9	16.5	23.4	28.2	34.4
11.5	0.161	19.582	-1.260	2.948	14.0	15.1	16.8	23.7	28.8	35.4
12	0.161	19.831	-1.274	2.825	14.1	15.3	17.0	24.0	29.2	36.3
12.5	0.160	20.026	-1.301	2.684	14.3	15.5	17.2	24.2	29.6	37.3
13	0.158	20.201	-1.344	2.548	14.4	15.6	17.4	24.3	29.9	38.5
13.5	0.156	20.393	-1.400	2.438	14.5	15.8	17.6	24.5	30.3	39.6
14	0.154	20.605	-1.468	2.360	14.8	16.1	17.9	24.7	30.6	40.8
14.5	0.151	20.848	-1.545	2.307	15.0	16.3	18.1	24.9	31.0	42.0
15	0.148	21.127	-1.630	2.268	15.3	16.7	18.5	25.1	31.3	43.1
Edad	P3	P5	P10	P25	P50	P75	P90	P95	P97	
6	13.6	13.9	14.3	15.2	16.2	17.5	19.0	20.2	21.1	
6.5	13.8	14.1	14.6	15.4	16.6	18.1	19.7	21.0	21.9	
7	14.0	14.3	14.8	15.7	17.0	18.6	20.4	21.7	22.6	
7.5	14.1	14.4	14.9	15.9	17.3	19.1	21.0	22.4	23.3	
8	14.2	14.5	15.0	16.0	17.6	19.5	21.6	23.0	24.0	
8.5	14.3	14.6	15.1	16.2	17.8	19.9	22.1	23.6	24.6	
9	14.4	14.7	15.2	16.3	18.1	20.3	22.6	24.1	25.2	
9.5	14.5	14.8	15.3	16.5	18.3	20.7	23.1	24.7	25.8	
10	14.7	15.0	15.5	16.7	18.6	21.1	23.7	25.3	26.4	
10.5	14.9	15.2	15.7	17.0	19.0	21.6	24.2	25.9	27.0	
11	15.1	15.4	16.0	17.2	19.3	22.0	24.7	26.4	27.6	
11.5	15.3	15.6	16.2	17.5	19.6	22.3	25.1	26.9	28.1	
12	15.5	15.8	16.4	17.8	19.8	22.6	25.4	27.2	28.5	
12.5	15.6	16.0	16.6	18.0	20.0	22.7	25.6	27.5	28.9	
13	15.8	16.2	16.8	18.2	20.2	22.9	25.7	27.7	29.2	
13.5	16.0	16.4	17.1	18.4	20.4	23.0	25.9	28.0	29.5	
14	16.3	16.6	17.3	18.6	20.6	23.2	26.1	28.2	29.8	
14.5	16.5	16.9	17.6	18.9	20.8	23.4	26.3	28.5	30.1	
15	16.8	17.2	17.9	19.2	21.1	23.7	26.6	28.8	30.4	

Tabla 6.1: Valores de los percentiles para el IMC (en  $\text{kg}/\text{m}^2$ ) de los niños con el modelo BCPE(6,4,3,4)

6.2. IMC en niñas

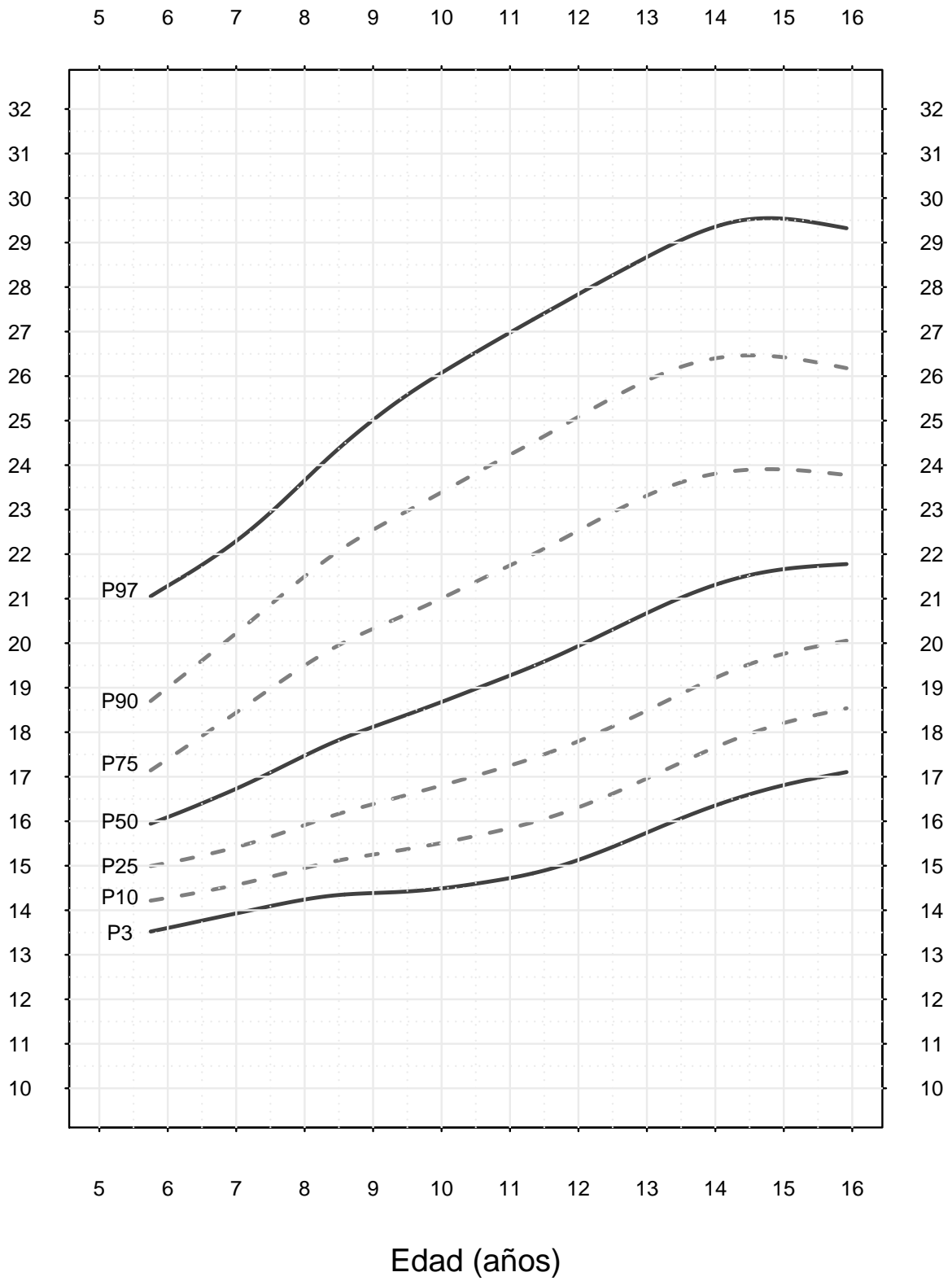


Figura 6.2: Curvas percentiles para el IMC (en  $\text{kg}/\text{m}^2$ ) de las niñas con el modelo BCPE (6,4,3,6).



Edad	S	M	L	P	-3SD	-2SD	-1SD	1SD	2SD	3SD
6	0.108	16.093	-2.326	1.845	12.5	13.5	14.6	18.2	21.9	30.8
6.5	0.115	16.401	-2.133	2.300	12.8	13.7	14.8	18.8	22.3	28.5
7	0.131	17.284	-1.661	3.076	13.0	13.8	14.9	19.4	22.8	27.7
7.5	0.129	17.157	-1.722	3.050	13.2	14.0	15.1	20.0	23.4	27.8
8	0.140	17.825	-1.391	2.928	13.3	14.1	15.4	20.6	24.1	28.6
8.5	0.141	17.878	-1.362	2.894	13.3	14.2	15.6	21.1	24.9	29.6
9	0.150	18.304	-1.117	2.575	13.2	14.2	15.7	21.5	25.5	30.8
9.5	0.151	18.349	-1.092	2.549	13.1	14.3	15.9	21.9	26.1	31.8
10	0.160	19.076	-0.792	2.445	13.0	14.3	16.1	22.2	26.7	32.4
10.5	0.159	18.976	-0.824	2.444	13.0	14.4	16.3	22.7	27.1	32.8
11	0.163	19.427	-0.700	2.447	13.1	14.5	16.5	23.0	27.6	33.2
11.5	0.164	19.758	-0.647	2.478	13.2	14.7	16.7	23.4	28.0	33.6
12	0.164	19.937	-0.629	2.498	13.4	14.9	17.0	23.9	28.4	33.9
12.5	0.164	20.183	-0.614	2.519	13.7	15.2	17.3	24.3	28.8	34.2
13	0.160	20.912	-0.619	2.415	14.0	15.5	17.6	24.7	29.3	34.7
13.5	0.157	21.176	-0.637	2.262	14.2	15.8	18.0	25.0	29.7	35.4
14	0.151	21.498	-0.677	1.980	14.4	16.1	18.4	25.2	30.0	36.3
14.5	0.150	21.556	-0.688	1.920	14.5	16.4	18.7	25.2	30.2	37.2
15	0.140	21.753	-0.759	1.621	14.6	16.6	18.9	25.2	30.2	37.8
Edad	P3	P5	P10	P25	P50	P75	P90	P95	P97	
6	13.6	13.9	14.3	15.1	16.1	17.4	19.0	20.3	21.3	
6.5	13.8	14.0	14.4	15.2	16.4	17.9	19.6	20.8	21.8	
7	13.9	14.2	14.6	15.4	16.7	18.4	20.2	21.4	22.3	
7.5	14.1	14.3	14.8	15.6	17.1	19.0	20.9	22.1	22.9	
8	14.2	14.5	14.9	15.9	17.5	19.5	21.5	22.8	23.7	
8.5	14.3	14.6	15.1	16.2	17.8	20.0	22.1	23.4	24.4	
9	14.4	14.7	15.3	16.4	18.1	20.3	22.5	24.0	25.0	
9.5	14.4	14.8	15.4	16.6	18.4	20.7	23.0	24.5	25.6	
10	14.5	14.9	15.5	16.8	18.7	21.0	23.4	25.0	26.1	
10.5	14.6	15.0	15.7	17.0	19.0	21.4	23.8	25.4	26.5	
11	14.7	15.1	15.8	17.3	19.3	21.7	24.2	25.9	27.0	
11.5	14.9	15.3	16.1	17.5	19.6	22.1	24.7	26.3	27.4	
12	15.1	15.6	16.3	17.8	19.9	22.5	25.1	26.7	27.8	
12.5	15.4	15.9	16.6	18.1	20.3	22.9	25.5	27.2	28.3	
13	15.7	16.2	17.0	18.5	20.7	23.3	25.9	27.6	28.7	
13.5	16.1	16.5	17.3	18.9	21.0	23.6	26.2	27.9	29.1	
14	16.4	16.8	17.7	19.2	21.3	23.8	26.4	28.1	29.4	
14.5	16.6	17.1	18.0	19.5	21.5	23.9	26.5	28.2	29.5	
15	16.8	17.3	18.2	19.8	21.7	23.9	26.4	28.2	29.5	

Tabla 6.2: Valores de los percentiles para el IMC (en kg/m<sup>2</sup>) de las niñas con el modelo BCPE(6,4,3,6)

6.3. Peso en niños

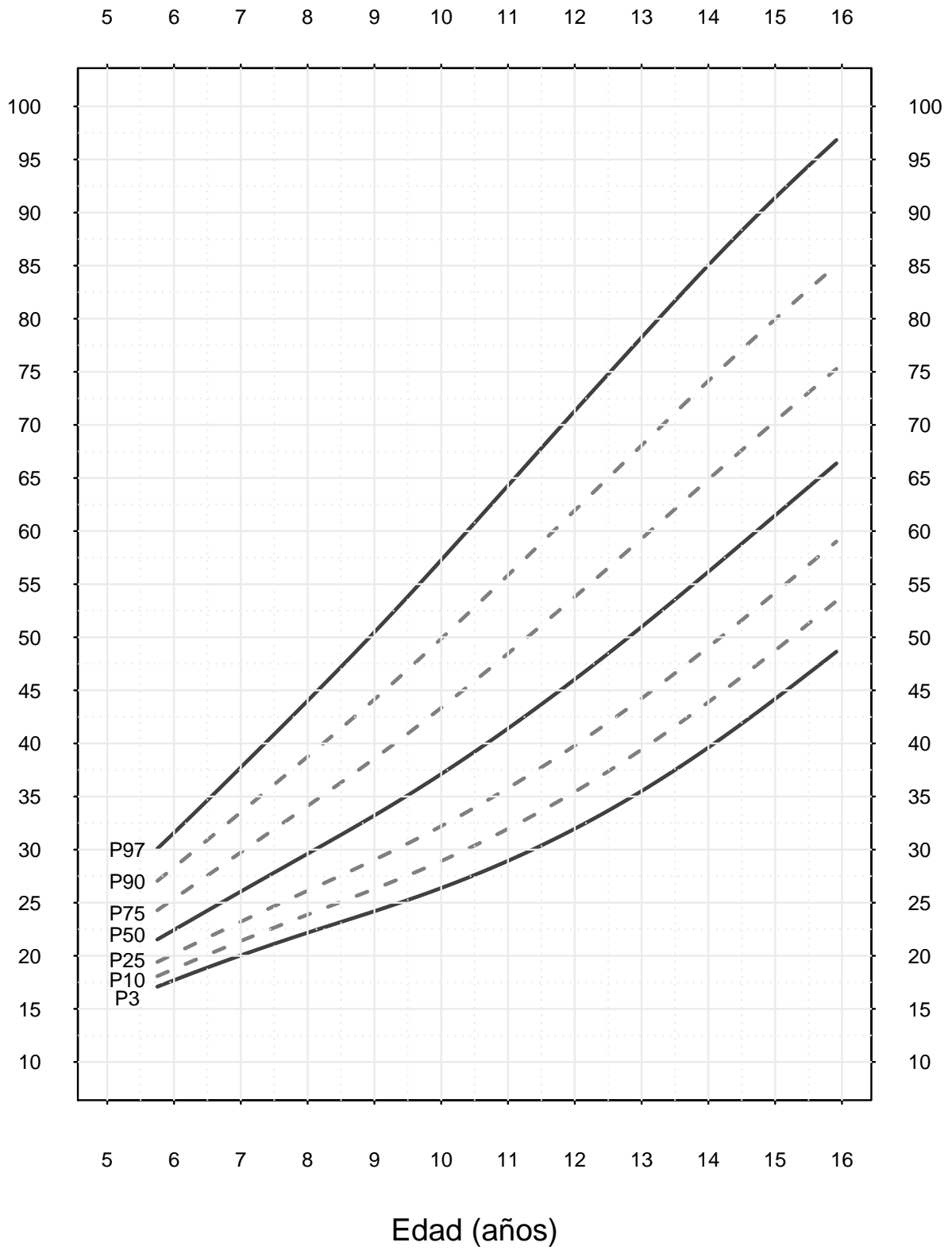


Figura 6.3: Curvas percentiles para el peso (en kg) de los niños con el modelo BCPE (4,4,3,2).

Edad	S	M	L	P	-3SD	-2SD	-1SD	1SD	2SD	3SD
6	0.153	22.443	-1.275	2.880	14.9	16.8	19.2	26.0	32.6	45.4
6.5	0.160	24.255	-1.187	2.827	15.9	17.9	20.4	28.3	35.2	46.8
7	0.168	26.054	-1.100	2.774	16.9	18.9	21.6	30.6	37.9	48.8
7.5	0.175	27.837	-1.013	2.722	17.9	20.1	22.9	33.2	41.0	51.7
8	0.182	29.611	-0.927	2.672	18.9	21.2	24.4	36.0	44.4	55.6
8.5	0.189	31.394	-0.843	2.622	19.8	22.4	26.0	38.7	47.9	60.1
9	0.196	33.218	-0.762	2.573	20.5	23.4	27.4	41.2	51.3	64.5
9.5	0.202	35.120	-0.687	2.525	21.1	24.4	28.9	43.7	54.6	69.0
10	0.207	37.119	-0.621	2.478	21.9	25.5	30.4	46.3	57.9	73.1
10.5	0.211	39.216	-0.564	2.432	22.8	26.8	32.1	49.0	61.2	76.9
11	0.213	41.407	-0.517	2.387	23.9	28.2	33.8	51.8	64.5	80.5
11.5	0.214	43.683	-0.481	2.342	25.2	29.8	35.8	54.7	67.8	84.2
12	0.214	46.040	-0.454	2.298	26.7	31.6	37.9	57.6	71.1	87.9
12.5	0.213	48.471	-0.438	2.256	28.2	33.4	40.1	60.3	74.1	91.4
13	0.210	50.977	-0.432	2.214	29.7	35.3	42.3	62.6	76.6	94.3
13.5	0.207	53.549	-0.437	2.172	31.1	37.0	44.3	64.4	78.5	96.4
14	0.202	56.169	-0.452	2.132	32.3	38.5	45.9	65.5	79.4	97.6
14.5	0.197	58.817	-0.476	2.092	33.2	39.7	47.2	66.0	79.7	98.0
15	0.192	61.479	-0.507	2.053	33.9	40.7	48.3	66.1	79.5	97.8
Edad	P3	P5	P10	P25	P50	P75	P90	P95	P97	
6	17.7	18.1	18.8	20.2	22.4	25.4	28.3	30.3	31.6	
6.5	18.9	19.3	20.1	21.7	24.3	27.6	30.9	33.1	34.6	
7	20.0	20.5	21.4	23.2	26.1	29.7	33.5	36.0	37.7	
7.5	21.1	21.7	22.7	24.7	27.8	31.9	36.1	38.9	40.9	
8	22.2	22.8	23.9	26.1	29.6	34.1	38.8	41.9	44.0	
8.5	23.2	23.9	25.1	27.6	31.4	36.3	41.4	44.8	47.2	
9	24.2	25.0	26.3	29.0	33.2	38.6	44.1	47.9	50.5	
9.5	25.3	26.1	27.6	30.6	35.1	40.9	47.0	51.0	53.8	
10	26.4	27.3	28.9	32.2	37.1	43.4	49.9	54.2	57.3	
10.5	27.6	28.6	30.4	34.0	39.2	45.9	52.8	57.5	60.8	
11	28.9	30.1	32.0	35.8	41.4	48.5	55.8	60.8	64.3	
11.5	30.4	31.6	33.6	37.7	43.7	51.1	58.9	64.1	67.8	
12	32.0	33.3	35.4	39.8	46.0	53.8	61.9	67.4	71.3	
12.5	33.7	35.0	37.4	42.0	48.5	56.5	65.0	70.7	74.8	
13	35.5	37.0	39.4	44.2	51.0	59.3	68.1	74.0	78.2	
13.5	37.5	39.0	41.6	46.6	53.5	62.1	71.1	77.3	81.7	
14	39.6	41.2	43.9	49.1	56.2	64.9	74.1	80.5	85.0	
14.5	41.9	43.5	46.3	51.6	58.8	67.6	77.1	83.6	88.3	
15	44.2	45.9	48.8	54.2	61.5	70.3	80.0	86.6	91.4	

Tabla 6.3: Valores de los percentiles para el peso (kg) de los niños con el modelo BCPE(4,4,3,2)

6.4. Peso en niñas

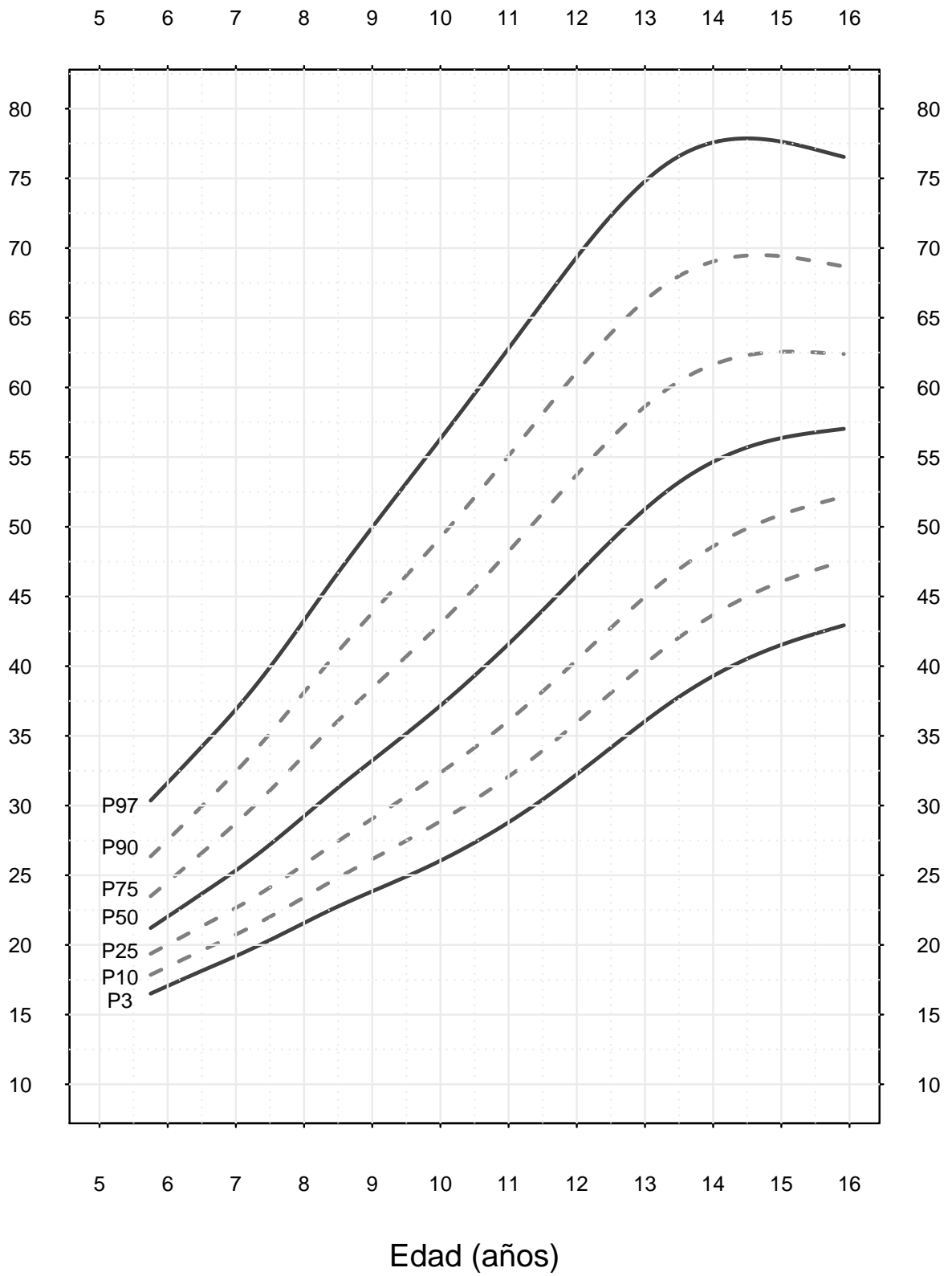


Figura 6.4: Curvas percentiles para el peso (en kg) de las niñas con el modelo BCPE(7,4,3,5).

Edad	S	M	L	P	-3SD	-2SD	-1SD	1SD	2SD	3SD
6	0.156	22.031	-1.133	1.764	14.9	16.8	19.2	26.0	32.6	45.4
6.5	0.163	23.675	-1.032	1.994	15.9	17.9	20.4	28.3	35.2	46.8
7	0.170	25.346	-0.933	2.218	16.9	18.9	21.6	30.6	37.9	48.8
7.5	0.178	27.203	-0.835	2.384	17.9	20.1	22.9	33.2	41.0	51.7
8	0.185	29.234	-0.740	2.450	18.9	21.2	24.4	36.0	44.4	55.6
8.5	0.192	31.289	-0.647	2.426	19.8	22.4	26.0	38.7	47.9	60.1
9	0.197	33.232	-0.558	2.351	20.5	23.4	27.4	41.2	51.3	64.5
9.5	0.201	34.839	-0.489	2.278	21.1	24.4	28.9	43.7	54.6	69.0
10	0.206	37.170	-0.404	2.210	21.9	25.5	30.4	46.3	57.9	73.1
10.5	0.208	39.316	-0.343	2.202	22.8	26.8	32.1	49.0	61.2	76.9
11	0.209	41.594	-0.293	2.213	23.9	28.2	33.8	51.8	64.5	80.5
11.5	0.208	44.011	-0.253	2.213	25.2	29.8	35.8	54.7	67.8	84.2
12	0.205	46.515	-0.225	2.185	26.7	31.6	37.9	57.6	71.1	87.9
12.5	0.200	48.977	-0.206	2.136	28.2	33.4	40.1	60.3	74.1	91.4
13	0.194	51.264	-0.193	2.073	29.7	35.3	42.3	62.6	76.6	94.3
13.5	0.187	53.198	-0.183	1.988	31.1	37.0	44.3	64.4	78.5	96.4
14	0.180	54.665	-0.173	1.889	32.3	38.5	45.9	65.5	79.4	97.6
14.5	0.171	55.701	-0.162	1.782	33.2	39.7	47.2	66.0	79.7	98.0
15	0.163	56.369	-0.149	1.671	33.9	40.7	48.3	66.1	79.5	97.8
Edad	P3	P5	P10	P25	P50	P75	P90	P95	P97	
6	17.1	17.6	18.4	20.0	22.0	24.5	27.5	29.9	31.6	
6.5	18.1	18.7	19.6	21.3	23.7	26.6	29.9	32.4	34.2	
7	19.2	19.8	20.7	22.7	25.3	28.7	32.4	35.0	36.9	
7.5	20.4	21.0	22.0	24.1	27.2	31.1	35.2	38.0	39.9	
8	21.6	22.3	23.4	25.8	29.2	33.6	38.1	41.2	43.3	
8.5	22.8	23.5	24.8	27.5	31.3	36.1	41.1	44.4	46.7	
9	23.8	24.7	26.2	29.1	33.2	38.4	43.8	47.4	50.0	
9.5	24.9	25.9	27.5	30.7	35.2	40.7	46.5	50.4	53.1	
10	26.0	27.1	28.9	32.4	37.2	43.1	49.2	53.4	56.3	
10.5	27.3	28.5	30.4	34.2	39.3	45.6	52.1	56.5	59.5	
11	28.8	30.0	32.1	36.1	41.6	48.2	55.1	59.6	62.8	
11.5	30.4	31.7	33.9	38.2	44.0	51.0	58.1	62.8	66.1	
12	32.2	33.6	36.0	40.4	46.5	53.7	61.1	66.0	69.3	
12.5	34.1	35.6	38.1	42.7	49.0	56.3	63.9	68.9	72.3	
13	36.0	37.6	40.2	45.0	51.3	58.6	66.2	71.3	74.8	
13.5	37.8	39.4	42.1	47.0	53.2	60.4	68.0	73.1	76.6	
14	39.3	41.0	43.7	48.6	54.7	61.6	69.0	74.1	77.6	
14.5	40.5	42.3	45.0	49.9	55.7	62.3	69.5	74.4	77.9	
15	41.5	43.3	46.1	50.9	56.4	62.6	69.4	74.2	77.6	

Tabla 6.4: Valores de los percentiles para el peso (kg) de las niñas con el modelo BCPE(7,4,3,5)

6.5. Talla en niños

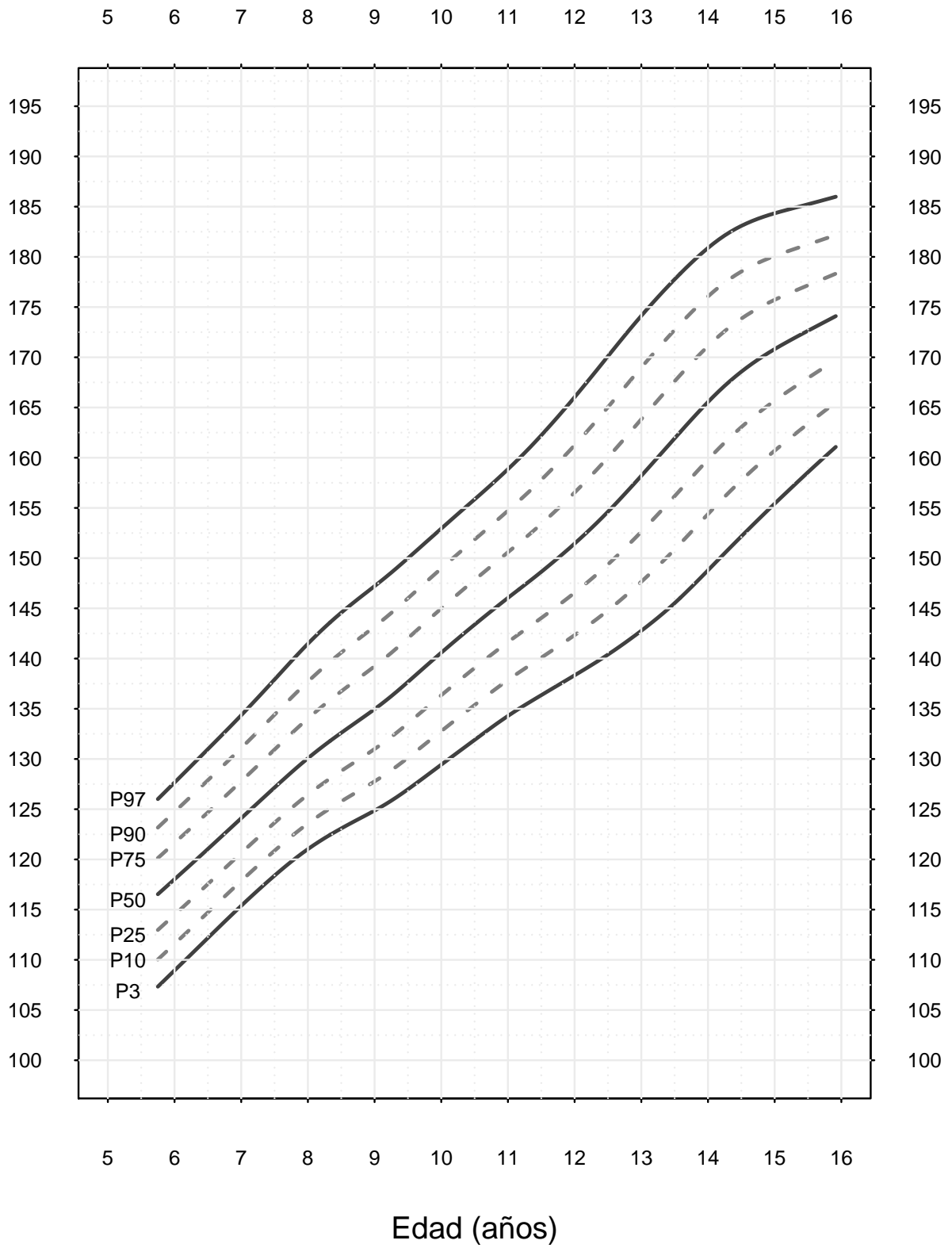


Figura 6.5: Curvas percentiles para la talla (cm) de los niños con el modelo BCPE(9,7,6,2).

Edad	S	M	L	P	-3SD	-2SD	-1SD	1SD	2SD	3SD	
6	0.043	118.023	0.254	2.353	104.4	108.4	112.9	123.3	128.2	132.9	
6.5	0.042	121.038	-0.485	2.324	107.9	111.7	116.0	126.4	131.6	136.7	
7	0.041	124.099	-1.132	2.296	111.3	114.9	119.1	129.6	135.0	140.6	
7.5	0.041	127.156	-1.681	2.268	114.3	117.9	122.1	132.8	138.6	144.8	
8	0.042	130.073	-1.964	2.240	116.8	120.5	124.9	136.0	142.2	149.0	
8.5	0.043	132.606	-1.842	2.212	118.6	122.6	127.1	138.8	145.4	152.6	
9	0.044	134.964	-1.344	2.185	120.0	124.3	129.2	141.3	148.0	155.2	
9.5	0.045	137.671	-0.703	2.158	121.6	126.4	131.7	144.1	150.8	157.7	
10	0.045	140.585	-0.228	2.132	123.6	128.8	134.4	147.1	153.7	160.5	
10.5	0.045	143.377	0.003	2.106	125.8	131.2	137.1	150.0	156.7	163.4	
11	0.045	146.041	-0.029	2.080	128.0	133.6	139.6	152.8	159.7	166.8	
11.5	0.046	148.634	-0.171	2.055	129.8	135.6	141.9	155.7	163.1	170.8	
12	0.049	151.449	-0.160	2.030	131.2	137.6	144.3	159.0	167.0	175.4	
12.5	0.051	154.623	0.118	2.005	132.5	139.5	146.9	162.7	171.1	179.9	
13	0.053	158.198	0.678	1.980	133.8	141.8	150.0	166.6	175.2	183.9	
13.5	0.053	161.928	1.358	1.956	135.2	144.4	153.3	170.4	178.8	187.1	
14	0.051	165.545	1.908	1.932	137.6	147.6	156.9	173.8	181.8	189.7	
14.5	0.048	168.563	2.320	1.908	140.7	151.0	160.2	176.4	184.0	191.4	
19	15	0.044	170.815	2.524	1.885	144.5	154.3	163.1	178.1	185.2	192.1
Edad	C3	C5	C10	C25	C50	C75	C90	C95	C97		
6	108.9	110.0	111.6	114.5	118.0	121.6	124.7	126.5	127.7		
6.5	112.2	113.2	114.7	117.5	121.0	124.7	127.9	129.8	131.0		
7	115.4	116.3	117.9	120.6	124.1	127.8	131.1	133.1	134.4		
7.5	118.4	119.3	120.9	123.7	127.2	130.9	134.4	136.5	137.9		
8	121.0	122.0	123.6	126.4	130.1	134.0	137.7	140.0	141.5		
8.5	123.1	124.1	125.8	128.8	132.6	136.7	140.6	143.0	144.6		
9	124.9	126.0	127.8	131.0	135.0	139.2	143.2	145.6	147.2		
9.5	127.0	128.2	130.1	133.5	137.7	142.0	146.0	148.4	150		
10	129.4	130.7	132.8	136.3	140.6	145.0	149.0	151.4	152.9		
10.5	131.9	133.2	135.4	139.1	143.4	147.8	151.9	154.3	155.9		
11	134.3	135.7	137.9	141.6	146.0	150.6	154.7	157.2	158.9		
11.5	136.4	137.8	140.1	144.0	148.6	153.4	157.8	160.5	162.2		
12	138.3	139.9	142.3	146.6	151.4	156.5	161.2	164.1	166.0		
12.5	140.4	142.1	144.8	149.4	154.6	160.0	165.0	168.1	170.1		
13	142.8	144.7	147.6	152.6	158.2	163.8	169.0	172.1	174.1		
13.5	145.5	147.6	150.9	156.2	161.9	167.6	172.7	175.8	177.8		
14	148.8	151.0	154.4	159.8	165.5	171.1	176.1	179.0	180.9		
14.5	152.2	154.4	157.7	163.0	168.6	173.8	178.5	181.3	183.1		
19	15	155.4	157.5	160.7	165.7	170.8	175.7	180.1	182.7	184.3	

Tabla 6.5: Valores de los percentiles para la talla (cm) de los niños con el modelo BCPE(9,7,6,2).

6.6. Talla en niñas

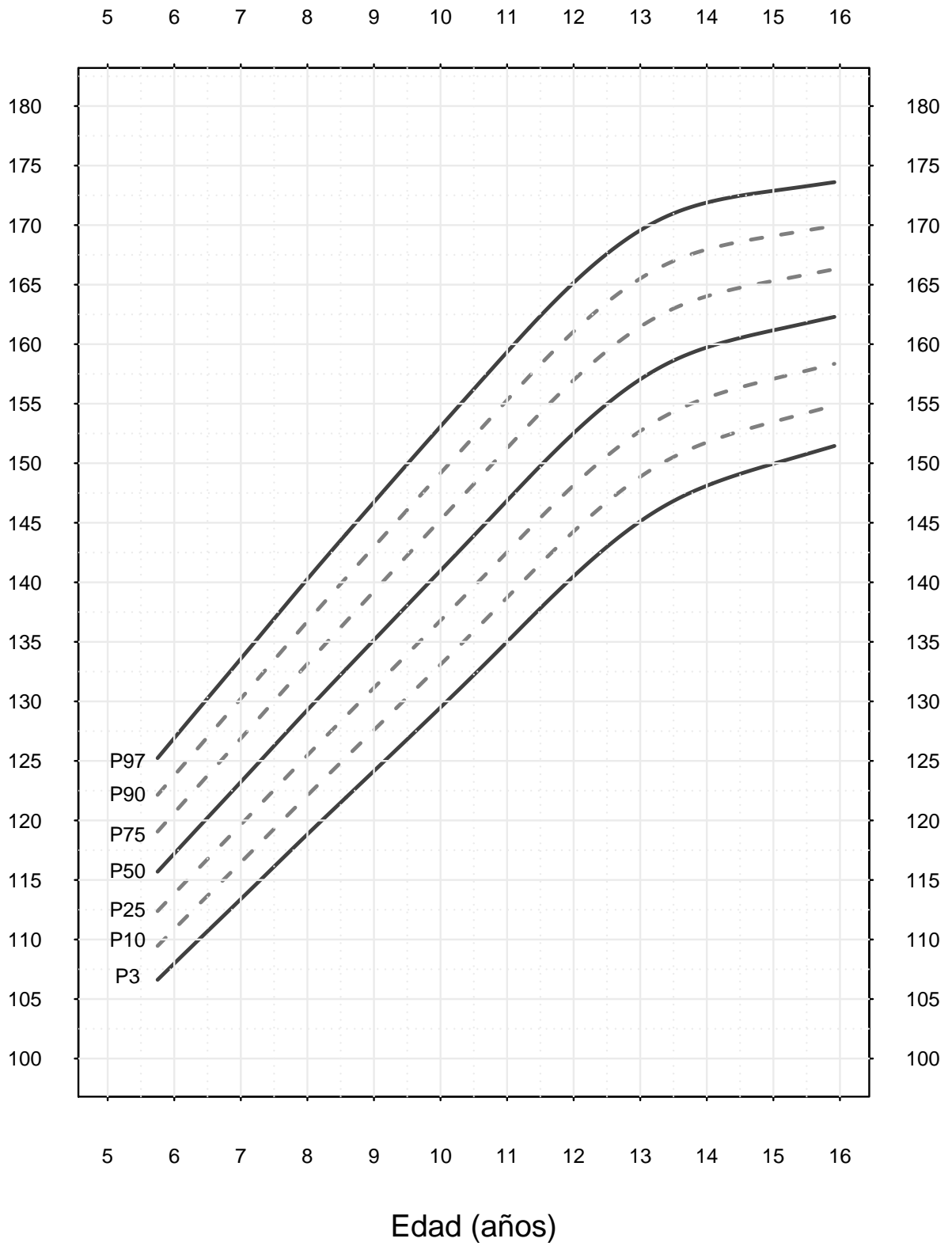


Figura 6.6: Curvas percentilares para la talla (cm) de las niñas con el modelo BCPE (6,3,1).



Edad	S	M	L	-3SD	-2SD	-1SD	1SD	2SD	3SD
6	0.043	117.213	0.380	102.714	107.415	112.2	122.3	127.5	132.9
6.5	0.043	120.228	0.380	105.253	110.108	115.1	125.5	130.9	136.5
7	0.044	123.230	0.380	107.780	112.788	117.9	128.7	134.3	140.0
7.5	0.044	126.261	0.380	110.332	115.494	120.8	131.9	137.6	143.5
8	0.044	129.277	0.380	112.877	118.191	123.7	135.1	141.0	147.1
8.5	0.044	132.230	0.380	115.382	120.840	126.5	138.2	144.3	150.5
9	0.044	135.153	0.380	117.878	123.474	129.2	141.2	147.5	153.9
9.5	0.044	138.072	0.380	120.394	126.120	132.0	144.3	150.7	157.3
10	0.044	140.985	0.380	122.938	128.784	134.8	147.3	153.9	160.6
10.5	0.044	143.910	0.380	125.545	131.495	137.6	150.4	157.0	163.9
11	0.044	146.862	0.380	128.243	134.276	140.5	153.4	160.2	167.1
11.5	0.044	149.783	0.380	130.986	137.079	143.3	156.4	163.2	170.2
12	0.043	152.526	0.380	133.646	139.768	146.1	159.2	166.0	173.0
12.5	0.042	154.982	0.380	136.118	142.237	148.5	161.6	168.4	175.4
13	0.041	157.047	0.380	138.298	144.384	150.6	163.6	170.4	177.3
13.5	0.040	158.624	0.380	140.079	146.102	152.3	165.1	171.8	178.6
14	0.041	158.396	0.380	141.456	147.394	153.5	166.1	172.7	179.4
14.5	0.039	160.525	0.380	142.561	148.402	154.4	166.8	173.2	179.8
15	0.038	161.169	0.380	143.524	149.265	155.1	167.3	173.6	180.1
Edad	P3	P5	P10	P25	P50	P75	P90	P95	P97
6	108	109.1	110.9	113.8	117.2	120.6	123.8	125.7	126.9
6.5	110.7	111.9	113.7	116.8	120.2	123.8	127.0	129.0	130.3
7	113.4	114.6	116.5	119.6	123.2	126.9	130.2	132.3	133.6
7.5	116.1	117.4	119.3	122.6	126.3	130.0	133.5	135.6	136.9
8	118.8	120.1	122.1	125.5	129.3	133.2	136.7	138.9	140.3
8.5	121.5	122.8	124.9	128.3	132.2	136.2	139.9	142.1	143.5
9	124.2	125.5	127.6	131.1	135.2	139.2	143.0	145.3	146.7
9.5	126.8	128.2	130.3	134.0	138.1	142.3	146.1	148.4	149.9
10	129.5	130.9	133.1	136.8	141.0	145.3	149.2	151.5	153.1
10.5	132.2	133.6	135.9	139.6	143.9	148.3	152.2	154.6	156.2
11	135.0	136.5	138.7	142.5	146.9	151.3	155.3	157.7	159.3
11.5	137.8	139.3	141.6	145.4	149.8	154.2	158.3	160.8	162.4
12	140.5	142.0	144.3	148.1	152.5	157.0	161.1	163.5	165.2
12.5	143.0	144.5	146.7	150.6	155.0	159.4	163.5	166.0	167.6
13	145.1	146.6	148.9	152.7	157.0	161.5	165.5	168.0	169.6
13.5	146.8	148.3	150.5	154.3	158.6	163.0	167.0	169.4	171.0
14	148.1	149.5	151.8	155.5	159.7	164.0	168.0	170.3	171.9
14.5	149.1	150.5	152.7	156.4	160.5	164.7	168.6	170.9	172.5
15	150.0	151.3	153.5	157.1	161.2	165.3	169.1	171.4	172.9

Tabla 6.6: Valores de los percentiles para la talla (cm) de las niñas con el modelo BCCG (6,3,1).

6.7. Perímetro de cintura en niños

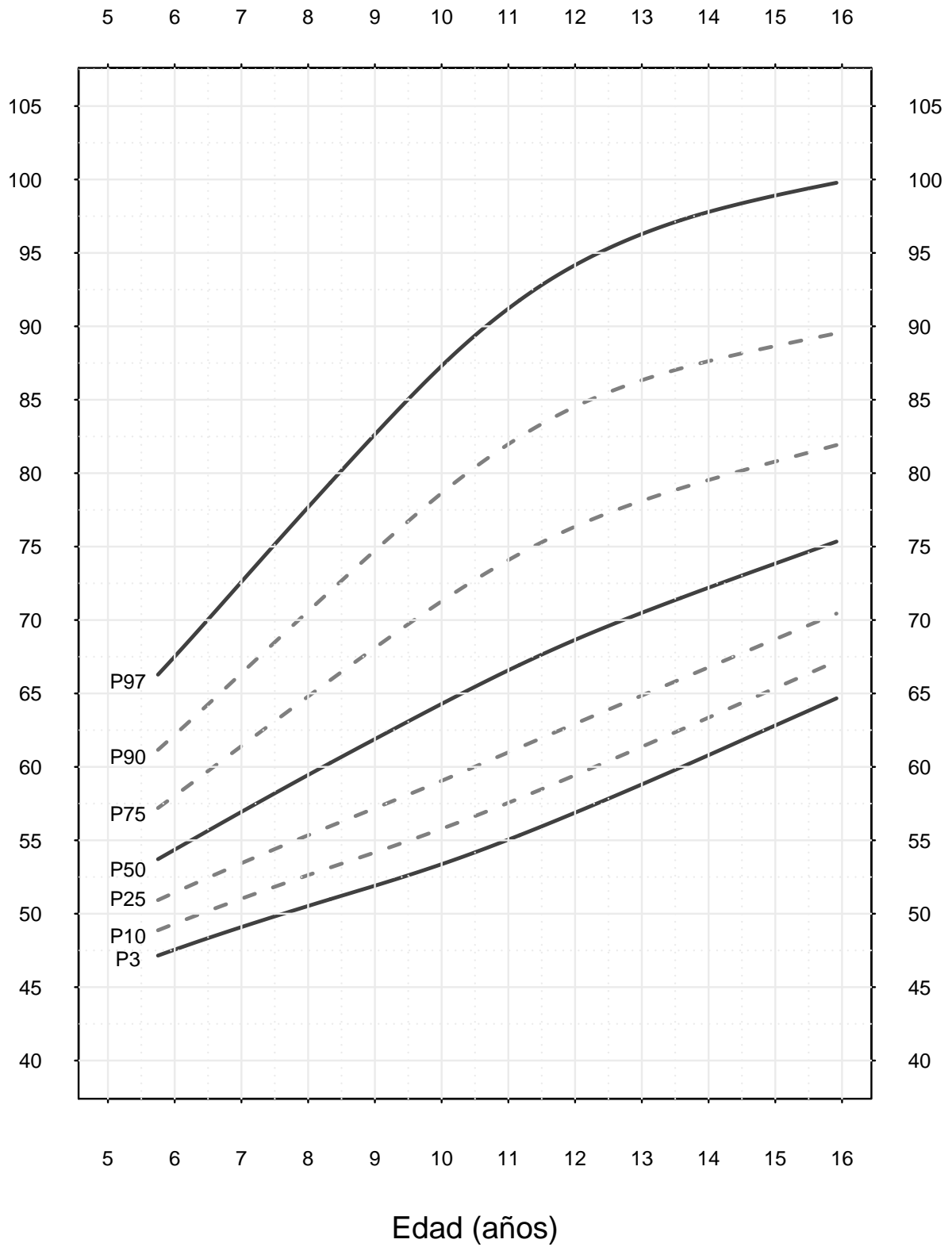


Figura 6.7: Curvas percentiles para el perímetro de cintura (cm) de los niños con el modelo BCPE (4,4,3,3).

Edad	S	M	L	P	-3SD	-2SD	-1SD	1SD	2SD	3SD
6	0.087	54.362	-2.804	2.124	44.9	47.2	50.3	60.1	68.8	84.8
6.5	0.092	55.653	-2.693	2.227	45.6	48.0	51.2	62.0	71.4	88.4
7	0.098	56.934	-2.581	2.333	46.2	48.7	52.1	63.9	74.1	91.9
7.5	0.103	58.199	-2.470	2.438	46.9	49.5	52.9	65.9	76.7	95.4
8	0.108	59.449	-2.364	2.541	47.5	50.2	53.8	67.8	79.3	98.6
8.5	0.113	60.680	-2.266	2.639	48.1	50.8	54.6	69.7	81.9	101.7
9	0.118	61.893	-2.180	2.728	48.7	51.5	55.5	71.5	84.4	104.9
9.5	0.122	63.099	-2.111	2.807	49.4	52.2	56.3	73.3	86.9	107.9
10	0.125	64.290	-2.063	2.872	50.1	53.0	57.2	75.1	89.3	110.9
10.5	0.128	65.455	-2.038	2.921	50.8	53.8	58.1	76.7	91.4	113.7
11	0.129	66.579	-2.037	2.954	51.6	54.6	59.0	78.2	93.3	116.2
11.5	0.129	67.646	-2.061	2.968	52.5	55.5	60.0	79.5	94.9	118.6
12	0.128	68.652	-2.106	2.958	53.4	56.4	60.9	80.6	96.4	120.9
12.5	0.126	69.597	-2.172	2.925	54.4	57.4	61.9	81.5	97.6	123.2
13	0.124	70.494	-2.255	2.870	55.3	58.4	62.8	82.3	98.6	125.7
13.5	0.121	71.359	-2.351	2.800	56.3	59.4	63.8	83.0	99.5	128.2
14	0.117	72.202	-2.457	2.723	57.3	60.4	64.8	83.7	100.2	130.8
14.5	0.114	73.031	-2.570	2.644	58.2	61.4	65.8	84.2	100.8	133.3
15	0.111	73.849	-2.686	2.564	59.2	62.4	66.8	84.8	101.4	135.8
Edad	P3	P5	P10	P25	P50	P75	P90	P95	P97	
6	47.6	48.2	49.3	51.4	54.4	58.0	62.2	65.2	67.5	
6.5	48.3	49.0	50.2	52.5	55.7	59.7	64.3	67.6	70.0	
7	49.1	49.8	51.0	53.4	56.9	61.4	66.4	69.9	72.6	
7.5	49.8	50.6	51.8	54.4	58.2	63.1	68.5	72.3	75.2	
8	50.5	51.3	52.6	55.3	59.4	64.8	70.6	74.7	77.7	
8.5	51.2	52.0	53.4	56.3	60.7	66.5	72.7	77.0	80.2	
9	51.9	52.7	54.2	57.2	61.9	68.1	74.7	79.3	82.6	
9.5	52.6	53.5	55.0	58.1	63.1	69.7	76.7	81.6	85.0	
10	53.4	54.2	55.8	59.1	64.3	71.3	78.6	83.7	87.3	
10.5	54.2	55.1	56.6	60.0	65.5	72.7	80.4	85.6	89.4	
11	55.0	56.0	57.6	61.0	66.6	74.1	82.0	87.4	91.2	
11.5	55.9	56.9	58.5	62.0	67.6	75.3	83.4	88.9	92.8	
12	56.9	57.8	59.4	62.9	68.7	76.4	84.5	90.1	94.2	
12.5	57.8	58.8	60.4	63.9	69.6	77.3	85.5	91.2	95.3	
13	58.8	59.7	61.4	64.9	70.5	78.1	86.3	92.1	96.3	
13.5	59.8	60.7	62.4	65.8	71.4	78.9	87.0	92.8	97.1	
14	60.8	61.7	63.4	66.8	72.2	79.5	87.6	93.4	97.8	
14.5	61.8	62.7	64.4	67.7	73.0	80.2	88.2	94.0	98.4	
15	62.8	63.8	65.4	68.7	73.8	80.8	88.7	94.5	98.9	

Tabla 6.7: Valores de los percentiles para el perímetro de cintura (cm) de los niños con el modelo BCPE (4,4,3,3).

6.8. Perímetro de cintura en niñas

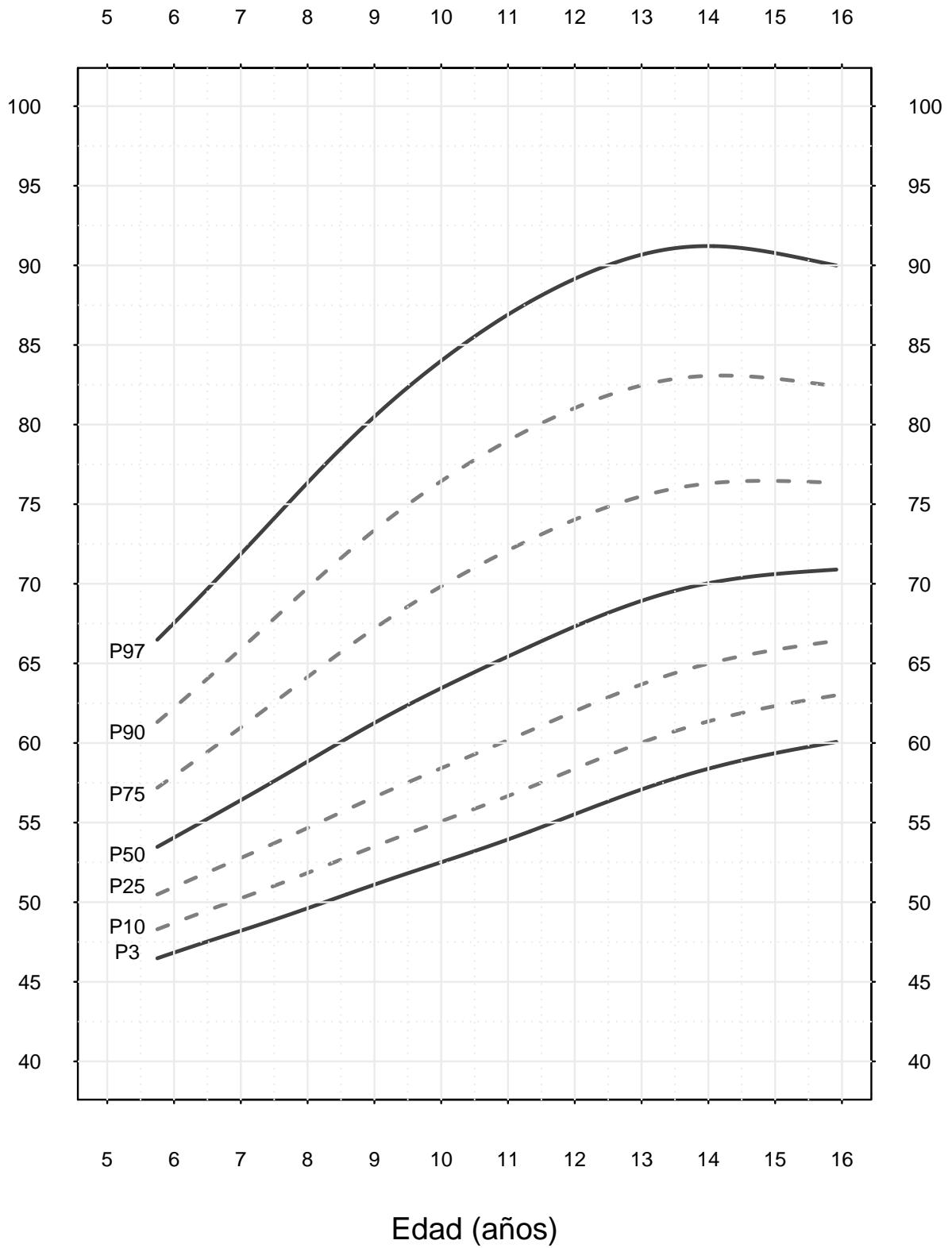


Figura 6.8: Curvas percentiles para el perímetro de cintura (cm) de las niñas con el modelo BCPE (5,4,3,3).

Edad	S	M	L	P	-3SD	-2SD	-1SD	1SD	2SD	3SD
6	0.094	54.643	-2.381	2.212	44.0	46.5	49.7	60.1	68.8	83.3
6.5	0.094	54.447	-2.401	2.199	44.6	47.2	50.5	61.8	71.0	85.9
7	0.102	56.604	-2.188	2.350	45.2	47.8	51.4	63.5	73.2	88.6
7.5	0.108	58.016	-2.065	2.443	45.8	48.5	52.2	65.3	75.5	91.3
8	0.112	59.252	-1.969	2.512	46.4	49.2	53.1	67.0	77.8	93.9
8.5	0.108	58.221	-2.048	2.455	47.0	50.0	54.0	68.8	80.0	96.4
9	0.120	62.383	-1.762	2.615	47.7	50.7	54.8	70.4	82.1	98.7
9.5	0.120	62.564	-1.751	2.617	48.3	51.4	55.7	71.9	84.0	100.9
10	0.124	64.619	-1.656	2.600	48.8	52.1	56.5	73.3	85.7	102.9
10.5	0.122	63.442	-1.704	2.618	49.4	52.7	57.3	74.5	87.2	104.9
11	0.124	65.110	-1.642	2.586	50.0	53.5	58.2	75.7	88.7	106.7
11.5	0.124	65.756	-1.628	2.563	50.7	54.2	59.1	76.7	89.9	108.3
12	0.122	67.616	-1.615	2.467	51.4	55.0	59.9	77.7	91.0	109.8
12.5	0.122	67.902	-1.617	2.447	52.1	55.8	60.8	78.5	91.8	111.1
13	0.118	69.463	-1.656	2.303	52.8	56.6	61.6	79.1	92.5	112.1
13.5	0.120	68.813	-1.632	2.371	53.4	57.2	62.3	79.5	92.9	112.9
14	0.115	69.965	-1.687	2.236	53.9	57.8	63.0	79.8	93.1	113.3
14.5	0.110	70.505	-1.743	2.130	54.4	58.4	63.5	79.8	92.9	113.2
15	0.107	70.701	-1.778	2.070	54.8	58.8	63.9	79.7	92.6	112.8
Edad	P3	P5	P10	P25	P50	P75	P90	P95	P97	
6	46.8	47.5	48.7	50.9	54.1	57.9	62.2	65.3	67.5	
6.5	47.5	48.3	49.5	51.9	55.2	59.4	64.0	67.3	69.7	
7	48.2	49.0	50.2	52.8	56.4	61.0	65.9	69.4	71.9	
7.5	48.9	49.7	51.0	53.7	57.6	62.5	67.8	71.5	74.1	
8	49.6	50.4	51.8	54.7	58.8	64.1	69.7	73.6	76.3	
8.5	50.4	51.2	52.7	55.6	60.1	65.7	71.6	75.6	78.5	
9	51.1	52.0	53.5	56.6	61.3	67.2	73.4	77.5	80.5	
9.5	51.8	52.7	54.3	57.5	62.4	68.6	75.0	79.3	82.3	
10	52.5	53.5	55.1	58.4	63.4	69.8	76.4	80.9	84.0	
10.5	53.2	54.2	55.9	59.3	64.5	71.0	77.8	82.3	85.5	
11	53.9	54.9	56.7	60.2	65.4	72.1	79.0	83.6	86.9	
11.5	54.7	55.7	57.5	61.1	66.4	73.1	80.1	84.8	88.1	
12	55.5	56.6	58.4	62.0	67.3	74.0	81.0	85.8	89.2	
12.5	56.3	57.4	59.2	62.9	68.2	74.8	81.8	86.6	90.0	
13	57.1	58.2	60.0	63.7	68.9	75.5	82.5	87.2	90.7	
13.5	57.8	58.9	60.7	64.4	69.6	76.0	82.9	87.6	91.1	
14	58.4	59.5	61.4	65.0	70.0	76.3	83.1	87.8	91.2	
14.5	58.9	60.0	61.9	65.5	70.4	76.4	83.0	87.7	91.1	
15	59.4	60.5	62.3	65.9	70.6	76.5	82.9	87.4	90.8	

Tabla 6.8: Valores de los percentiles para el perímetro de cintura (cm) de las niñas con el modelo BCPE (5,4,3,3).

6.9. Perímetro de cadera en niños

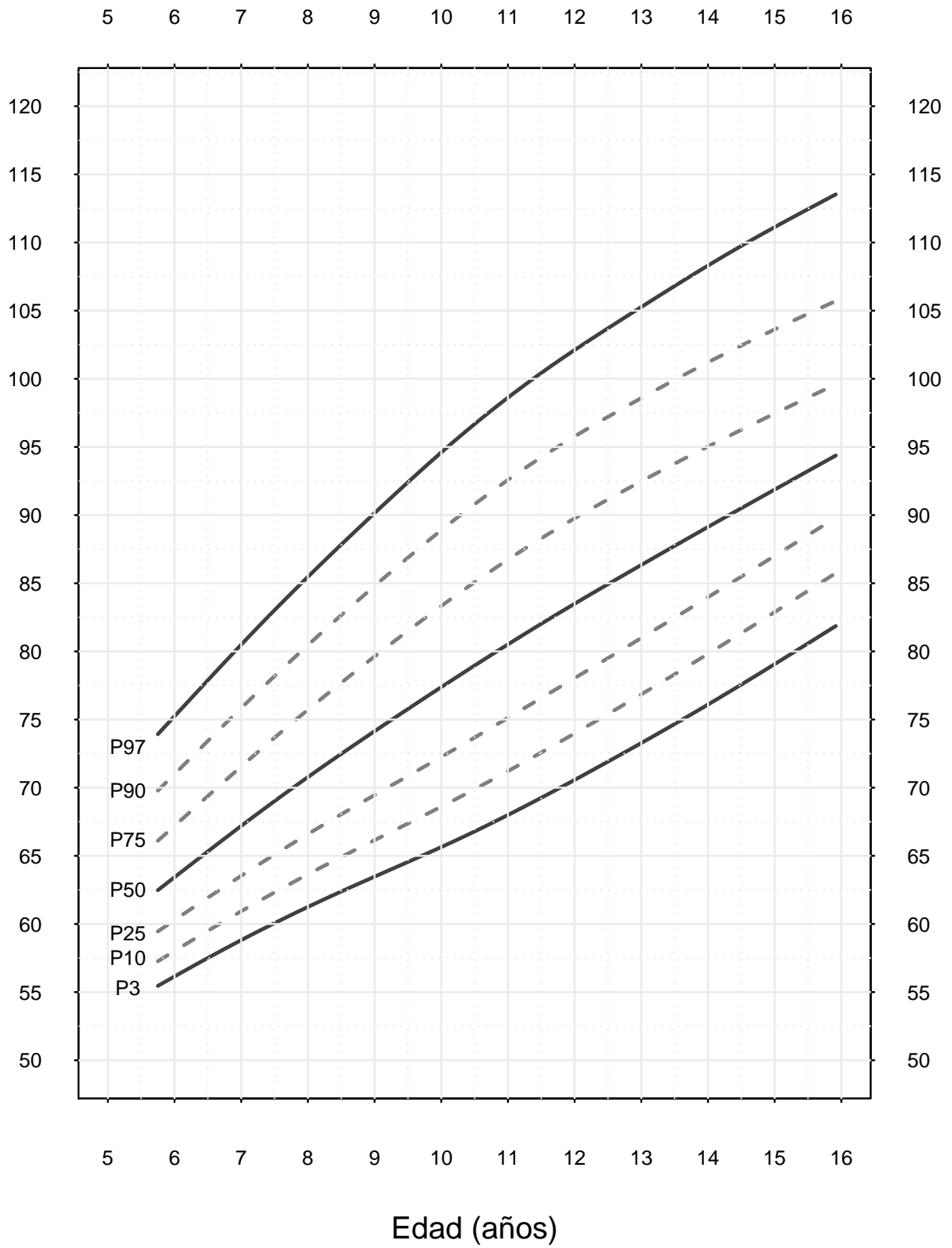


Figura 6.9: Curvas percentiles para perímetro de cadera (cm) de los niños con el modelo BCPE (4,4,3,3).

Edad	S	M	L	P	-3SD	-2SD	-1SD	1SD	2SD	3SD
6	0.076	63.429	-2.327	2.365	53.3	55.8	59.0	69.2	76.2	85.3
6.5	0.079	65.325	-2.141	2.415	54.5	57.1	60.6	71.5	78.8	88.1
7	0.083	67.193	-1.957	2.463	55.6	58.4	62.1	73.8	81.5	90.9
7.5	0.086	69.013	-1.778	2.505	56.7	59.7	63.5	76.1	84.0	93.6
8	0.089	70.775	-1.608	2.539	57.7	60.8	64.9	78.2	86.5	96.2
8.5	0.092	72.481	-1.453	2.564	58.6	61.9	66.3	80.3	88.9	98.7
9	0.094	74.142	-1.315	2.580	59.5	63.0	67.6	82.4	91.3	101.2
9.5	0.097	75.772	-1.196	2.581	60.3	64.0	68.9	84.4	93.5	103.7
10	0.098	77.378	-1.100	2.566	61.2	65.1	70.2	86.3	95.8	106.2
10.5	0.100	78.961	-1.029	2.534	62.1	66.2	71.5	88.1	97.9	108.7
11	0.100	80.514	-0.986	2.487	63.1	67.4	72.9	89.8	99.9	111.0
11.5	0.100	82.026	-0.969	2.426	64.2	68.7	74.3	91.4	101.7	113.3
12	0.099	83.497	-0.975	2.350	65.3	69.9	75.8	92.9	103.4	115.5
12.5	0.098	84.930	-1.001	2.262	66.4	71.3	77.2	94.3	105.1	117.7
13	0.096	86.340	-1.047	2.163	67.6	72.6	78.7	95.7	106.7	120.1
13.5	0.094	87.739	-1.111	2.058	68.8	74.0	80.2	97.0	108.3	122.6
14	0.092	89.129	-1.192	1.952	70.0	75.4	81.7	98.2	109.9	125.2
14.5	0.090	90.506	-1.290	1.851	71.3	76.9	83.2	99.4	111.4	127.9
15	0.088	91.872	-1.399	1.755	72.6	78.3	84.8	100.6	112.8	130.7
Edad	P3	P5	P10	P25	P50	P75	P90	P95	P97	
6	56.2	56.8	58.0	60.3	63.4	67.2	71.0	73.5	75.2	
6.5	57.5	58.3	59.5	61.9	65.3	69.4	73.4	76.1	77.9	
7	58.8	59.6	60.9	63.5	67.2	71.6	75.8	78.6	80.5	
7.5	60.1	60.9	62.3	65.1	69.0	73.7	78.2	81.1	83.0	
8	61.3	62.1	63.6	66.6	70.8	75.7	80.5	83.4	85.5	
8.5	62.4	63.3	64.9	68.0	72.5	77.7	82.7	85.8	87.8	
9	63.5	64.5	66.1	69.5	74.1	79.6	84.8	88.0	90.1	
9.5	64.6	65.6	67.4	70.9	75.8	81.5	86.9	90.2	92.4	
10	65.6	66.7	68.6	72.3	77.4	83.3	88.9	92.3	94.6	
10.5	66.8	67.9	69.9	73.7	79.0	85.1	90.8	94.3	96.7	
11	68.0	69.2	71.2	75.1	80.5	86.7	92.6	96.2	98.6	
11.5	69.3	70.5	72.6	76.6	82.0	88.3	94.2	97.9	100.4	
12	70.6	71.9	74.0	78.0	83.5	89.8	95.8	99.6	102.1	
12.5	71.9	73.2	75.4	79.5	84.9	91.1	97.2	101.1	103.7	
13	73.3	74.6	76.8	81.0	86.3	92.5	98.6	102.6	105.3	
13.5	74.7	76.1	78.3	82.5	87.7	93.8	99.9	104.0	106.8	
14	76.1	77.5	79.8	84.0	89.1	95.0	101.2	105.4	108.3	
14.5	77.6	79.0	81.3	85.5	90.5	96.3	102.4	106.7	109.8	
15	79.1	80.5	82.9	87.0	91.9	97.4	103.6	108.0	111.1	

Tabla 6.9: Valores de los percentiles para perímetro de cadera (cm) de los niños con el modelo BCPE (4,4,3,3).

6.10. Perímetro de cadera en niñas

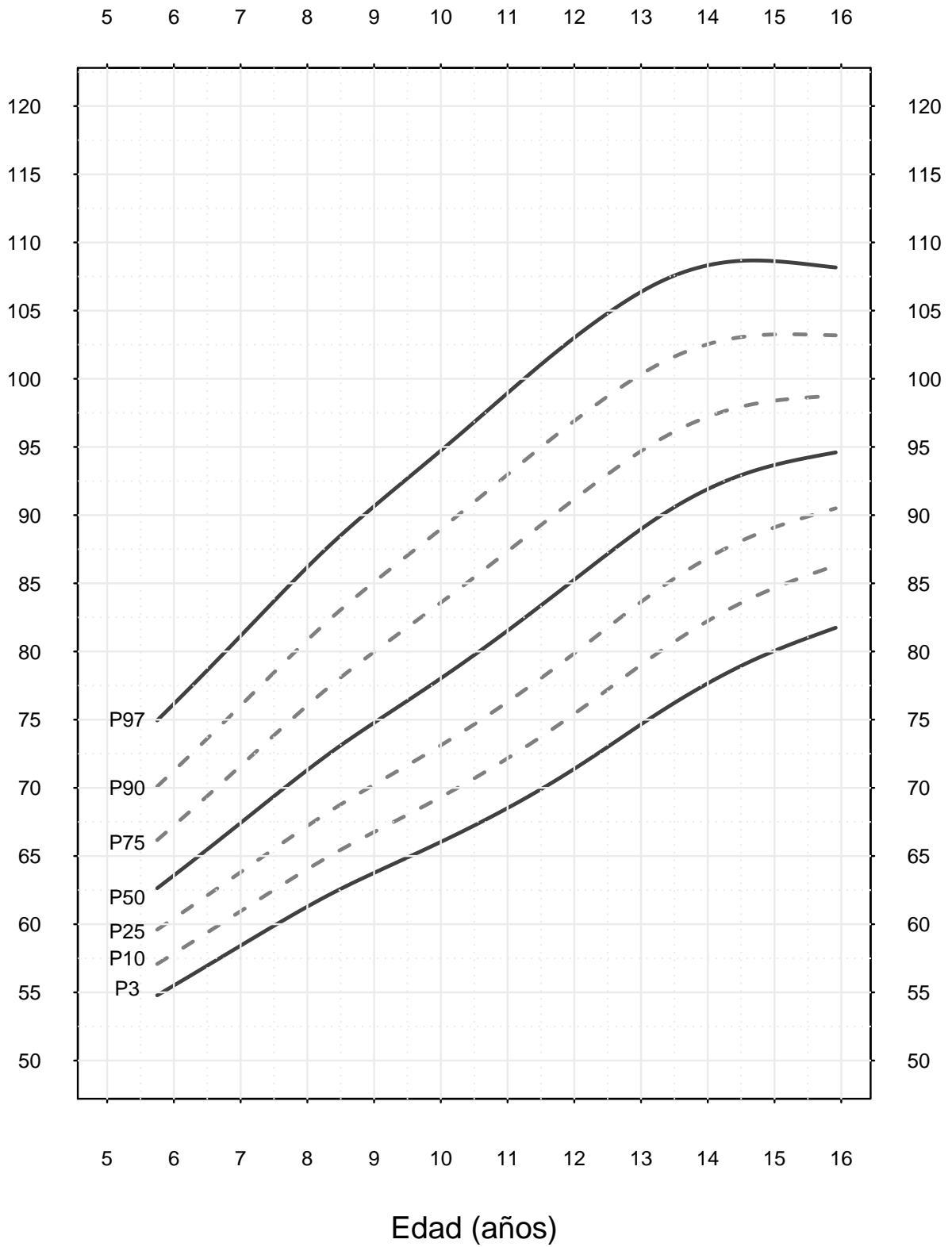


Figura 6.10: Curvas percentiles para el perímetro de cadera (cm) de las niñas con el modelo BCPE (6,4,2,3).



Edad	S	M	L	P	-3SD	-2SD	-1SD	1SD	2SD	3SD
6	0.081	63.577	-1.816	1.835	51.6	55.1	59.0	69.3	77.3	89.1
6.5	0.083	65.484	-1.693	1.914	53.0	56.5	60.6	71.6	79.7	91.3
7	0.085	67.427	-1.570	1.992	54.4	58.0	62.2	73.9	82.3	93.6
7.5	0.088	69.386	-1.448	2.065	55.7	59.4	63.9	76.2	84.8	96.0
8	0.090	71.307	-1.325	2.127	56.9	60.8	65.5	78.5	87.3	98.4
8.5	0.093	73.956	-1.141	2.189	58.1	62.1	66.9	80.7	89.7	100.7
9	0.093	74.776	-1.079	2.201	59.0	63.2	68.3	82.7	91.8	102.8
9.5	0.095	76.392	-0.957	2.211	59.9	64.3	69.6	84.5	93.9	104.9
10	0.096	78.030	-0.834	2.205	60.8	65.4	71.0	86.4	95.9	107.0
10.5	0.097	79.740	-0.711	2.183	61.7	66.6	72.5	88.4	98.0	109.2
11	0.098	81.527	-0.588	2.145	62.6	67.8	74.0	90.3	100.2	111.5
11.5	0.098	83.378	-0.466	2.095	63.6	69.2	75.7	92.3	102.3	113.7
12	0.098	85.267	-0.343	2.036	64.6	70.6	77.4	94.2	104.3	115.8
12.5	0.096	87.157	-0.220	1.971	65.8	72.2	79.3	96.0	106.1	117.6
13	0.094	88.977	-0.097	1.901	67.0	73.8	81.1	97.7	107.6	119.0
13.5	0.091	90.587	0.025	1.829	68.2	75.3	82.9	99.0	108.8	120.1
14	0.087	91.910	0.148	1.755	69.3	76.8	84.4	100.0	109.5	120.6
14.5	0.083	92.936	0.271	1.684	70.2	78.0	85.7	100.6	109.8	120.6
15	0.079	93.678	0.394	1.615	71.1	79.1	86.8	100.9	109.8	120.1
Edad	P3	P5	P10	P25	P50	P75	P90	P95	P97	
6	55.5	56.4	57.8	60.4	63.6	67.2	71.3	74.1	76.2	
6.5	57.0	57.9	59.4	62.1	65.5	69.4	73.6	76.5	78.6	
7	58.4	59.4	61.0	63.8	67.4	71.6	76.0	79.0	81.1	
7.5	59.9	60.9	62.5	65.5	69.4	73.9	78.5	81.5	83.7	
8	61.3	62.3	64.0	67.2	71.3	76.0	80.8	84.0	86.2	
8.5	62.6	63.7	65.5	68.8	73.1	78.1	83.1	86.3	88.5	
9	63.8	64.9	66.8	70.2	74.8	80.0	85.1	88.4	90.6	
9.5	64.9	66.1	68.0	71.7	76.4	81.8	87.0	90.4	92.7	
10	66.0	67.3	69.3	73.1	78.0	83.6	89.0	92.4	94.7	
10	67.2	68.6	70.7	74.7	79.7	85.4	91.0	94.5	96.8	
11	68.5	69.9	72.2	76.3	81.5	87.3	93.0	96.5	99.0	
11.5	69.9	71.4	73.7	78.0	83.4	89.3	95.0	98.6	101.0	
12	71.4	72.9	75.4	79.9	85.3	91.2	96.9	100.6	103.0	
12.5	73.0	74.6	77.2	81.8	87.2	93.0	98.7	102.3	104.8	
13	74.7	76.3	79.0	83.6	89.0	94.7	100.3	103.9	106.4	
13.5	76.2	78.0	80.7	85.4	90.6	96.1	101.6	105.2	107.6	
14	77.7	79.5	82.2	86.9	91.9	97.2	102.5	106.0	108.3	
14.5	78.9	80.8	83.6	88.1	92.9	98.0	103.1	106.4	108.7	
15	80.0	81.9	84.7	89.1	93.7	98.4	103.3	106.4	108.6	

Tabla 6.10: Valores de los percentiles para el perímetro de cadera (cm) de las niñas con el modelo BCPE (6,4,2,3).

## Capítulo 7

# Comparación entre niños y niñas

En este capítulo se representan las gráficas que comparan las curvas percentiles de niños y niñas que se obtuvieron para cada variable.

7.1. IMC

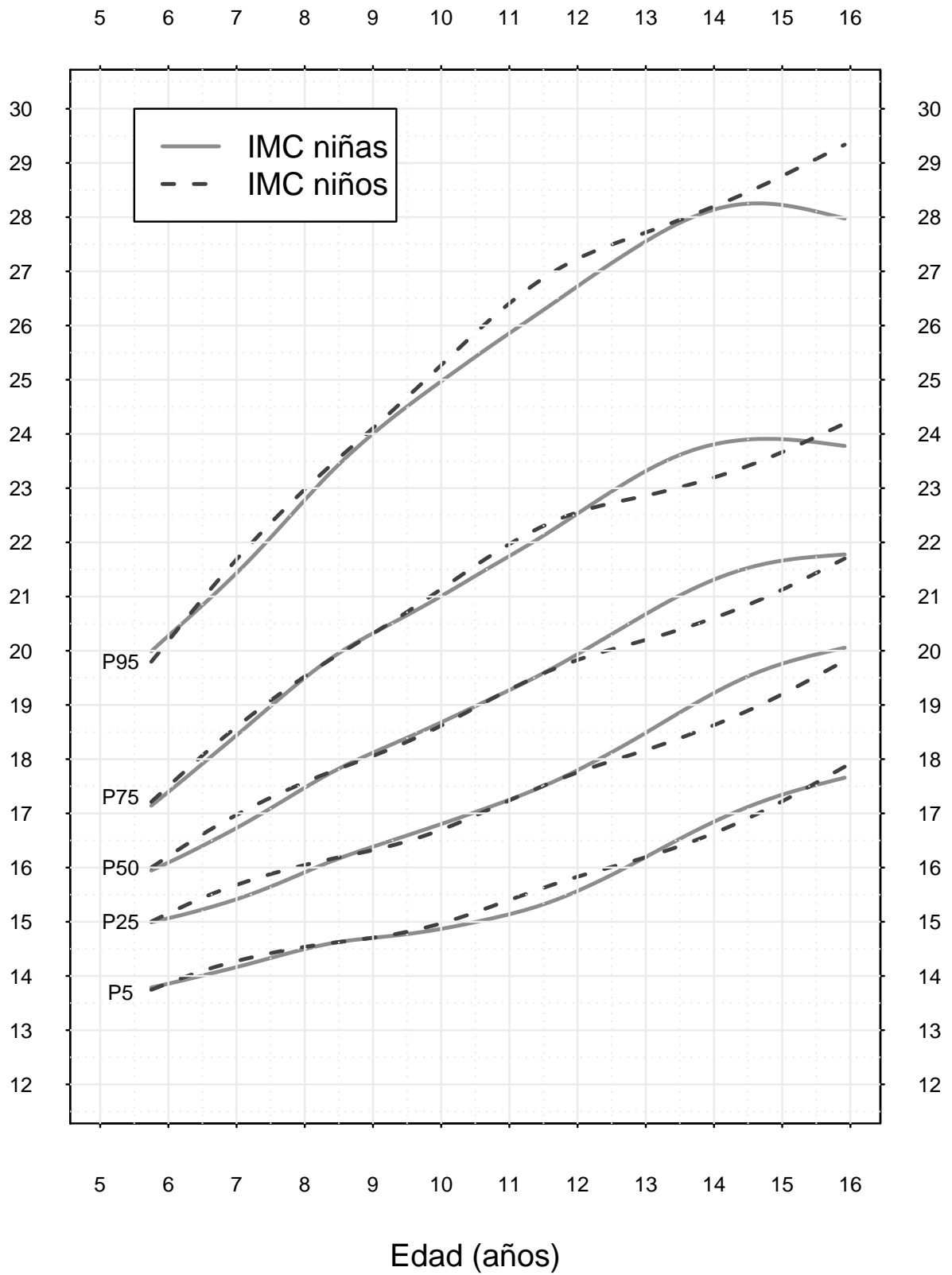


Figura 7.1: Comparación del IMC en niñas (sólido negro) y niños (discontinuo gris).

7.2. Peso

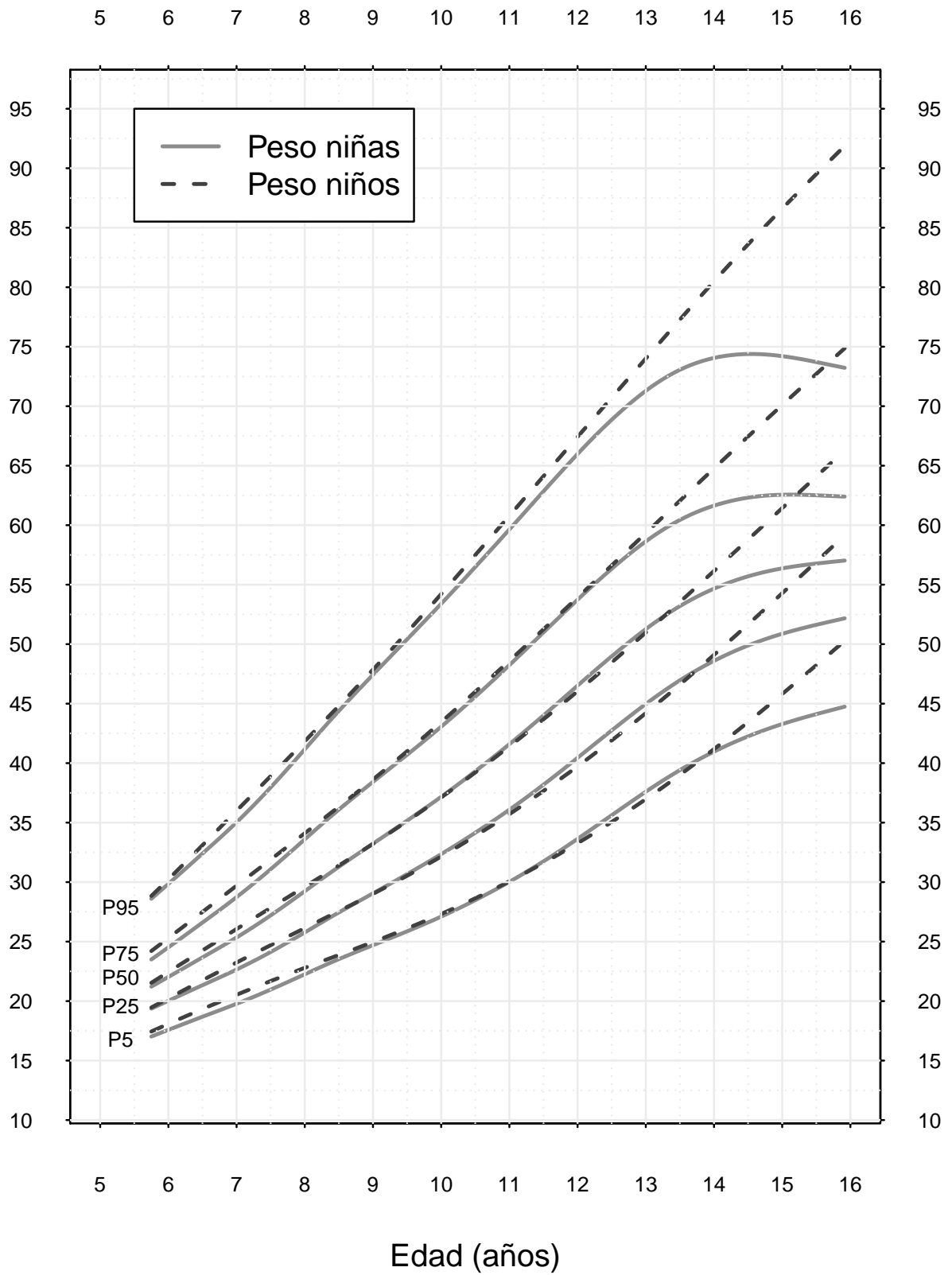


Figura 7.2: Comparación del peso en niñas (sólido gris) y niños (discontinuo negro).

7.3. Talla

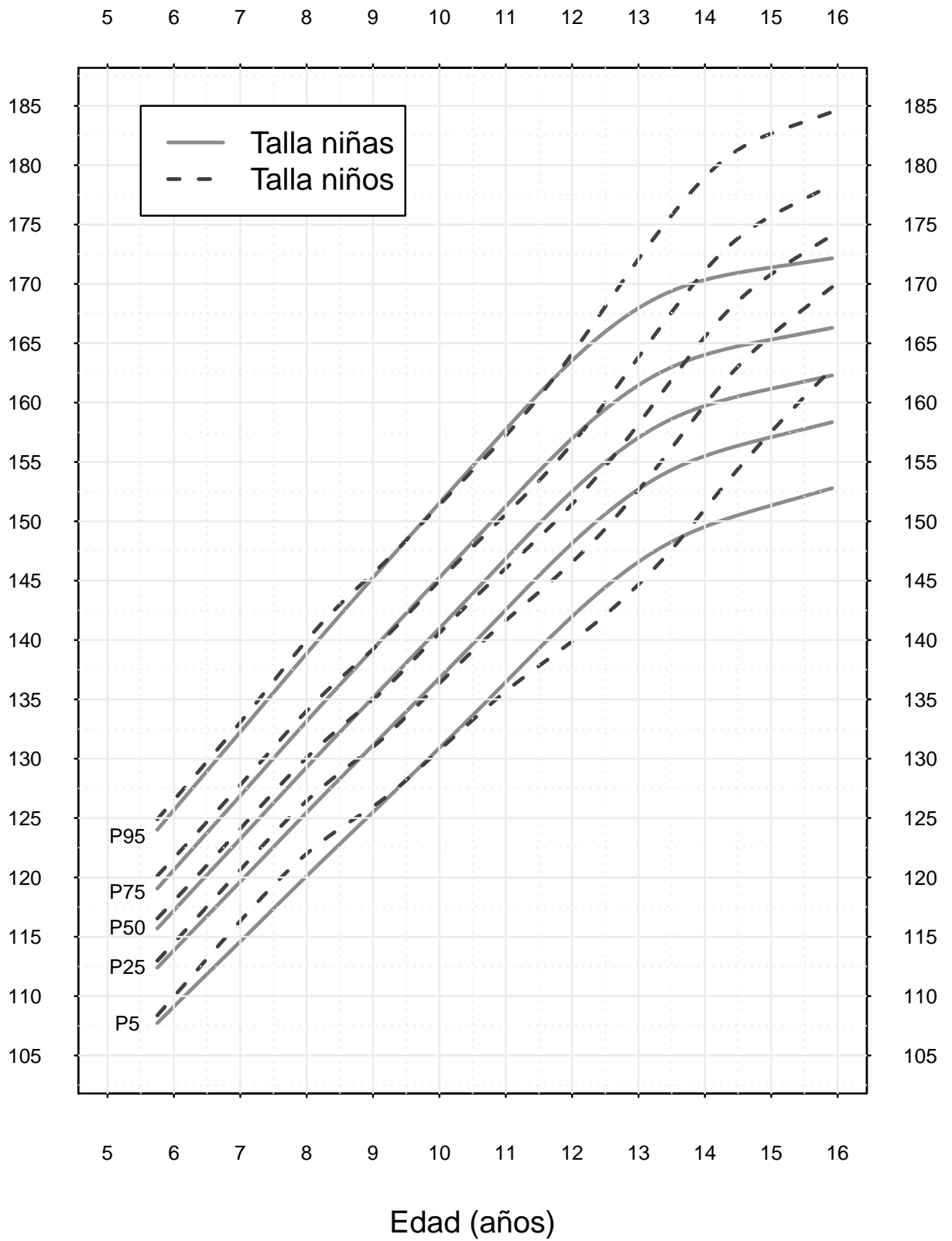


Figura 7.3: Comparación de la talla en niñas (sólido gris) y niños (discontinuo negro).

7.4. Perímetro de cintura

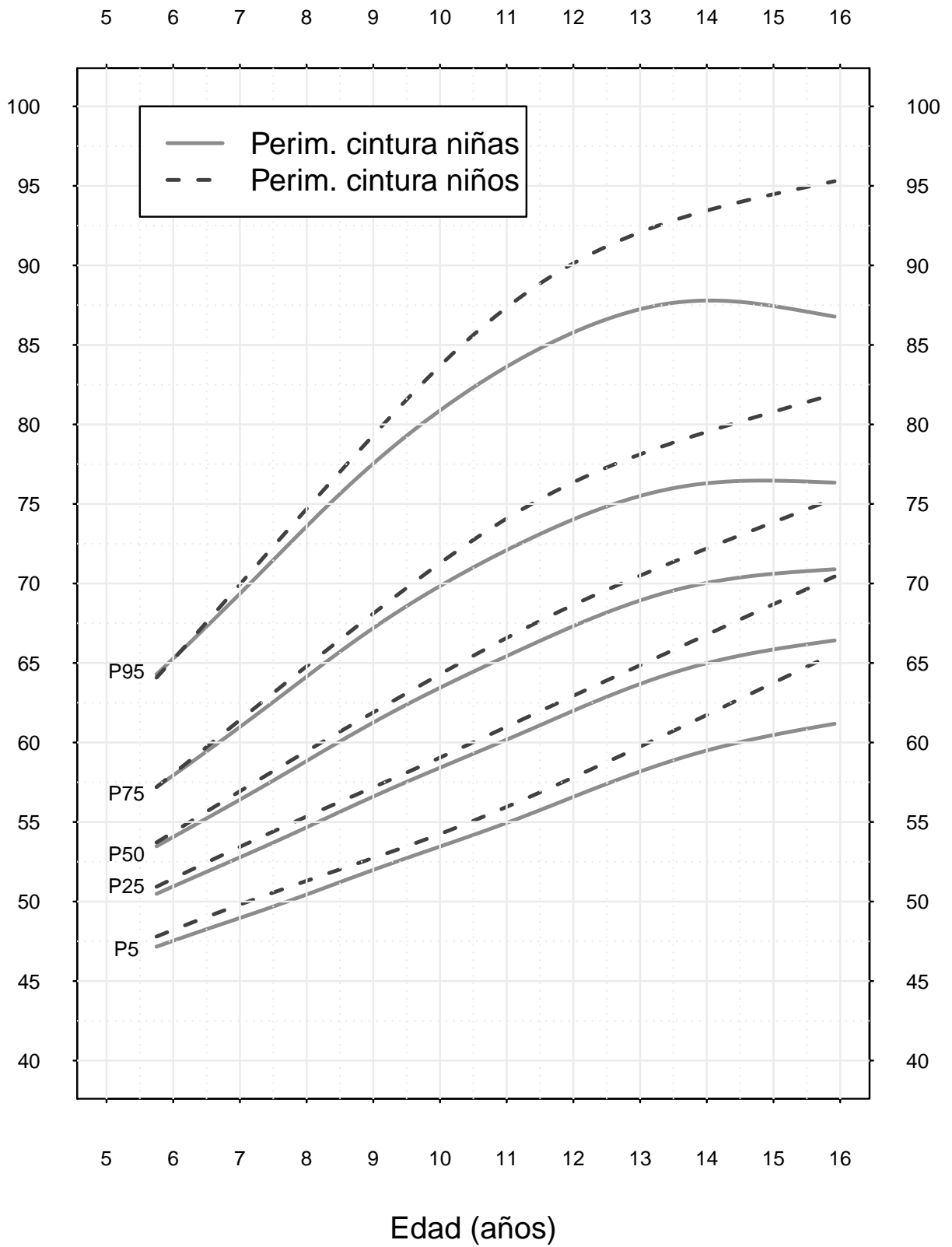


Figura 7.4: Comparación del perímetro de cintura en niñas (sólido gris) y niños (discontinuo negro).

7.5. Perímetro de cadera

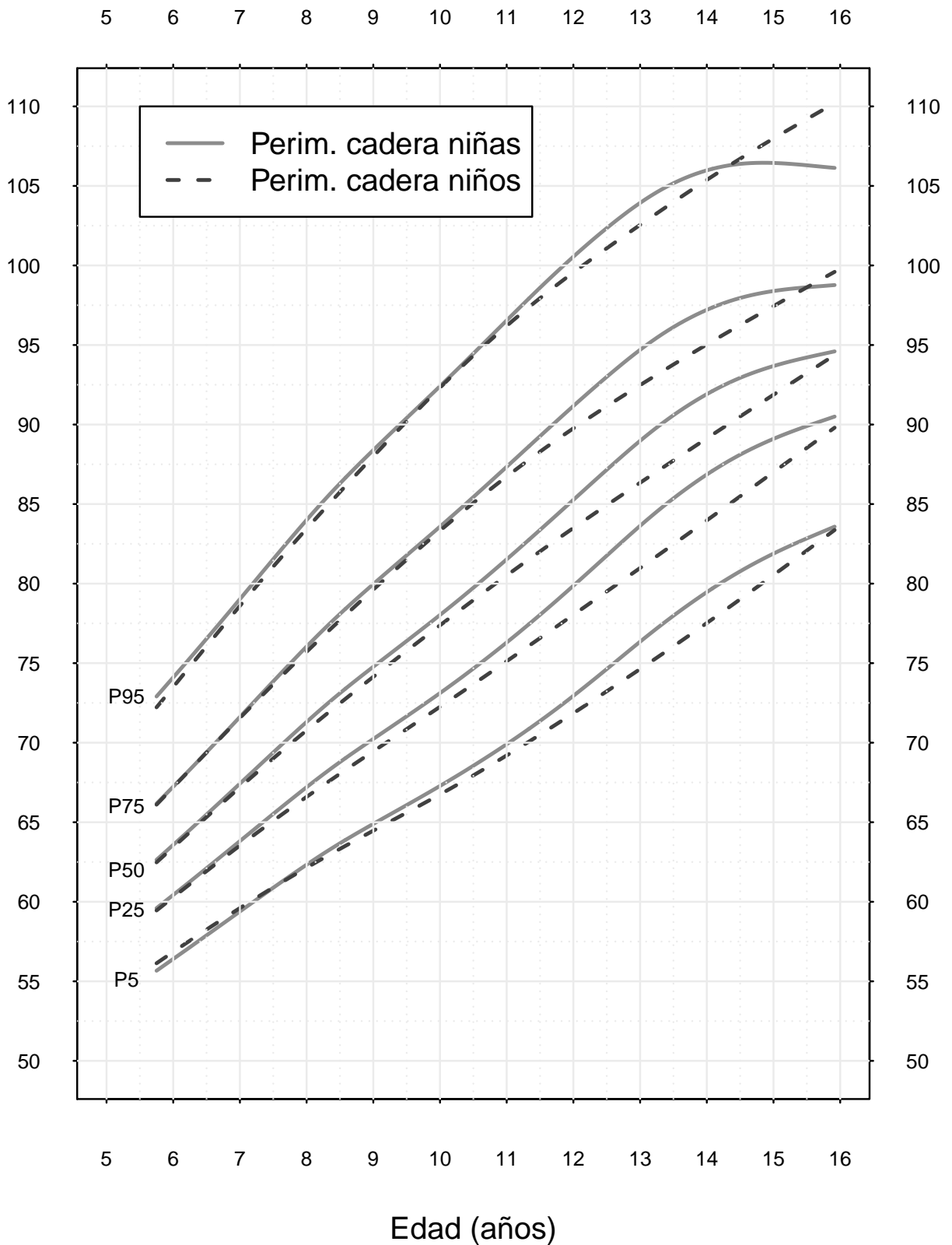


Figura 7.5: Comparación del perímetro de cadera en niñas (sólido gris) y niños (discontinuo negro).

## 7.6. Comentarios sobre las gráficas

1. **IMC:** la Figura 7.1 que compara las curvas percentiles del IMC en niños y niñas, muestra que el IMC, en cualquiera de los percentiles analizados, sigue una evolución similar en niños y niñas hasta que, a partir de los 9 años, los dos percentiles extremos (5 y 95) se separan ligeramente; en los cuartiles, esta diferenciación se observa aproximadamente a los 12 años.
2. **Peso:** la Figura 7.2 que compara las curvas percentiles del peso en niñas y niños, refleja que el aumento en niños y niñas transcurre parejo hasta aproximadamente los 13 años. En este momento, en cualquiera de los percentiles analizados excepto en el 95, el peso en niños continúa con un crecimiento lineal y en las niñas se produce una estabilización. Así, la mediana del peso está en torno a los 50 kg a los 13 años en ambos sexos, pero a los 15 años, su valor está por encima de los 60 kg en los niños superando en algo más de 5 kg a la mediana en las niñas.

Las curvas del percentil 95 en niños y niñas empiezan a separarse, aunque de forma ligera, a partir de los 10 años, haciéndose más evidente la diferencia a los 13; además, la estabilización del peso que se observaba en los otros percentiles, en este caso se transforma en un descenso a partir de los 14 años.

3. **Talla:** en la Figura 7.3 que compara las curvas percentiles de la talla en niñas y niños se observa, al igual que con el peso, un aumento que transcurre parejo en niños y niñas hasta que llegados a los 12 años (percentil 95 y 75) o los 13 (percentiles 50, 25 y 5) en las niñas se observa un aumento menos pronunciado en el crecimiento. Mientras tanto, con independencia del percentil analizado, el crecimiento en los niños continúa. La mediana del peso es de unos 146 cm a los 11 años, tanto en niños como en niñas, y llega a 170 cm a los 15 años en niños, pero en niñas es aproximadamente 162 cm.
4. **Perímetro de cintura:** la Figura 7.4 que compara las curvas percentiles del perímetro de cintura en niñas y niños muestra que cualquiera de los percentiles analizados, presenta valores más altos en los niños. Las diferencias en el perímetro de cintura de niños y niñas situados en los percentiles 5, 25 y 50 se hacen más evidentes a partir de los 14 años, momento en el que en las niñas el valor del perímetro de cintura se estabiliza con una mediana en torno a 70 cm, mientras que en los niños sigue aumentando.
4. **Perímetro de cadera:** en la Figura 7.5 que compara las curvas percentiles del perímetro de cadera en niñas y niños se puede ver que cualquiera de los percentiles analizados, toma un valor similar en los niños y niñas que están clasificados en el percentil 75 y 95 hasta los 11 años, a partir de entonces los valores son más altos en niñas; además en estas últimas al llegar a los 14 años el valor del perímetro de cadera se estabiliza. En relación con los percentiles 5, 25 y 50 la diferenciación en el valor del perímetro entre niños y niñas sucede aproximadamente a los 8 años y a partir de entonces son las niñas las que presentan un valor más alto en el perímetro.



## Capítulo 8

# Comparación con otros estudios

En este capítulo se muestran las curvas comparativas de los percentiles del IMC que se obtuvieron en este trabajo con los datos del estudio realizado por la DXSP y los obtenidos en los estudios de Orbegozo, de la OMS y de Cole. Ya se comentó que el primero es un estudio a nivel nacional y los otros dos son estudios internacionales. A continuación, se hace un pequeño resumen de los datos que recogen cada uno de estos estudios.

- **Fundación Orbegozo** [12]: los datos de el estudio transversal de Orbefozo fueron recogidos en el periodo desde noviembre de 2000 hasta el 31 de octubre de 2001 en niños de diferentes centros de Vizcaya, en una muestra de 6.443 sujetos (3.496 niñas y 2.947 niños) de edades comprendidas entre 0 y 18 años.
- **OMS** [7]: los datos fueron recogidos de 3 estudios: Health Examination Survey (HES) Cycle II (6-11 años), HES Clycle III (12-17 años) y Nutrition Examination Survey (HANES) Cycle I (0-74 años) en una muestra de tamaño total de 11.410 niños y 11.507 niñas.
- **International Obesity Task Force (IOTF)-Cole**[5]: este estudio reúne datos de 6 grandes estudios transversales de crecimiento con representatividad nacional: Brasil, Gran Bretaña, Hong Kong, Países Bajos, Singapur y Estados Unidos, en una muestra de 97.876 niños y 94.851 niñas de 0 a 25 años.

En estas figuras se representan los percentiles correspondientes al bajo peso, sobrepeso y obesidad según los puntos de corte utilizados en cada estudio para niños y niñas por separado. En la Tabla 8 se muestran los percentiles que se corresponden con esos puntos de corte en cada uno de los estudios.

Sexo	Rangos de peso	Orbegozo	Cole	OMS
Niños	Bajo peso	P4	P15.5	P2.3
	Sobrepeso	P79	P90.5	P84.1
	Obesidad	P97.5	P98.9	P97.7
Niñas	Bajo peso	P10	P16.5	P2.3
	Sobrepeso	P89	P89.3	P84.1
	Obesidad	P99	P98.6	P97.7

Tabla 8.1: Orden de los percentiles correspondientes a los puntos de corte utilizados en los estudios de Orbegozo, Cole y OMS, que definen el bajo peso, el sobrepeso y la obesidad.

8.1. Orbeago en niños

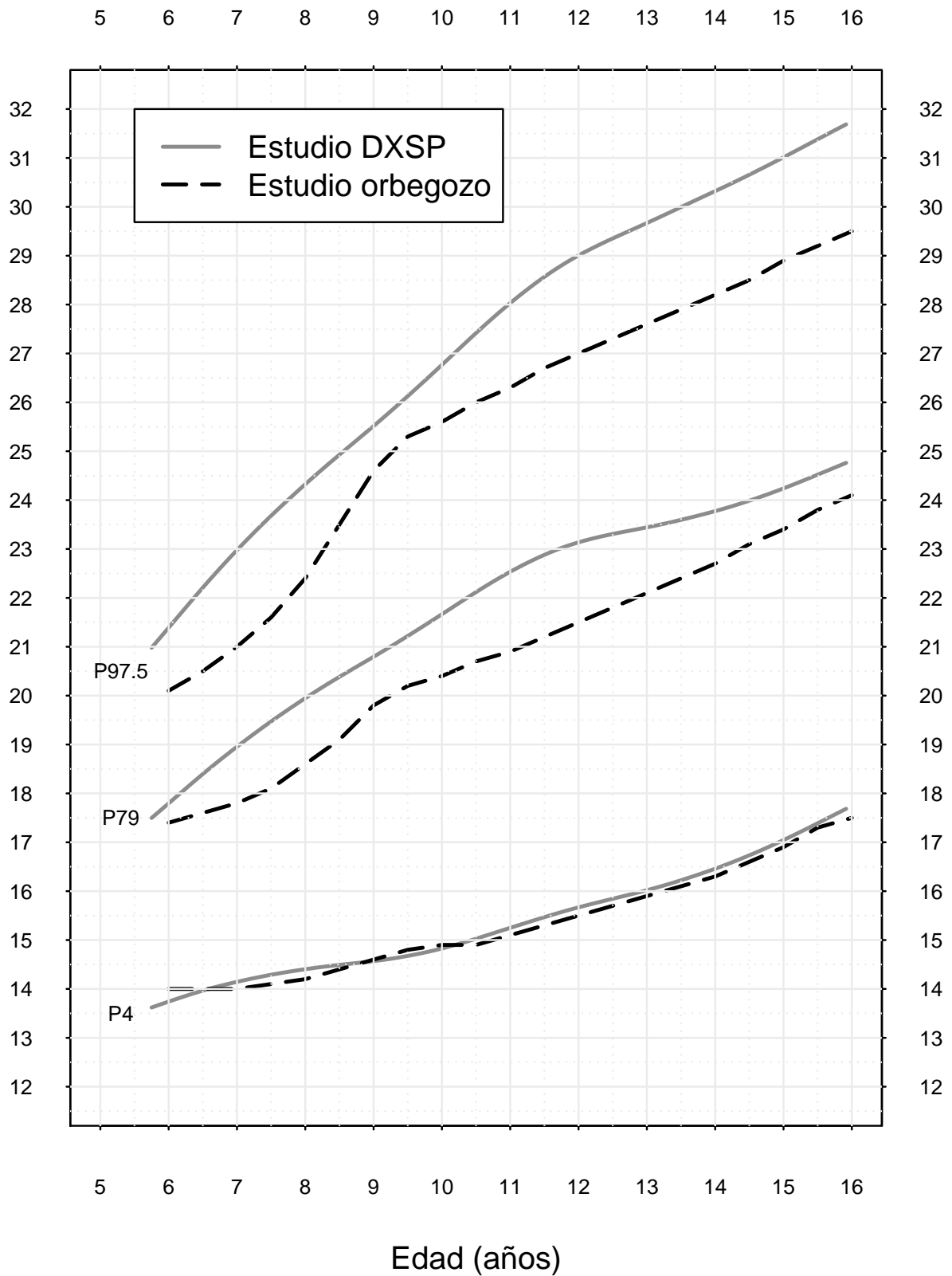


Figura 8.1: Comparación de los percentiles del IMC que definen el bajo peso, sobrepeso y obesidad de los niños con los datos del estudio de DXSP (sólido negro) y con los del estudio de orbeago (discontinuo gris).

8.2. Orbegozo en niñas

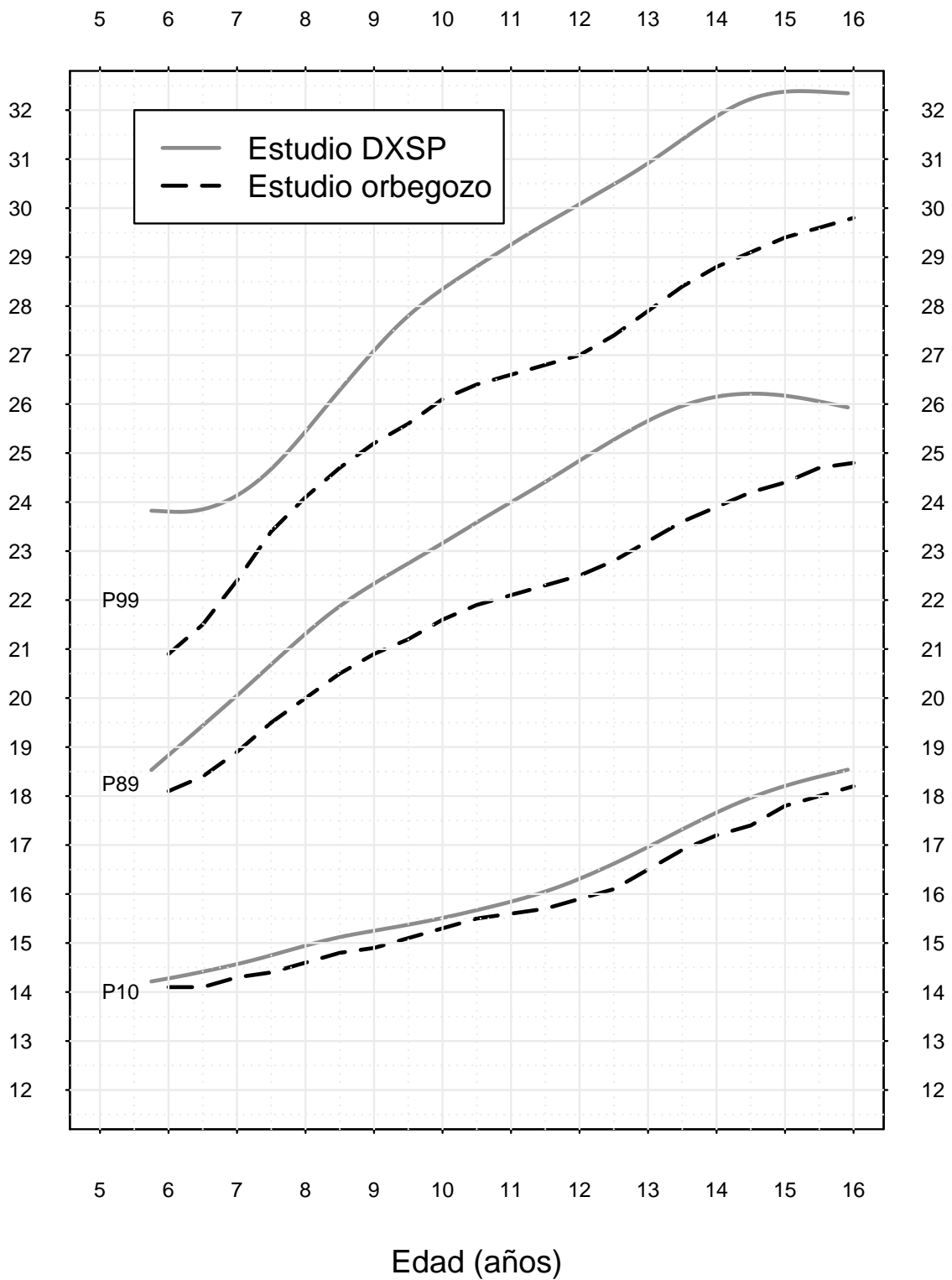


Figura 8.2: Comparación de los percentiles del IMC que definen el bajo peso, sobrepeso y obesidad de las niñas con los datos del estudio de DXSP (sólido negro) y con los del orbegozo (discontinuo gris).

8.3. Cole en niños

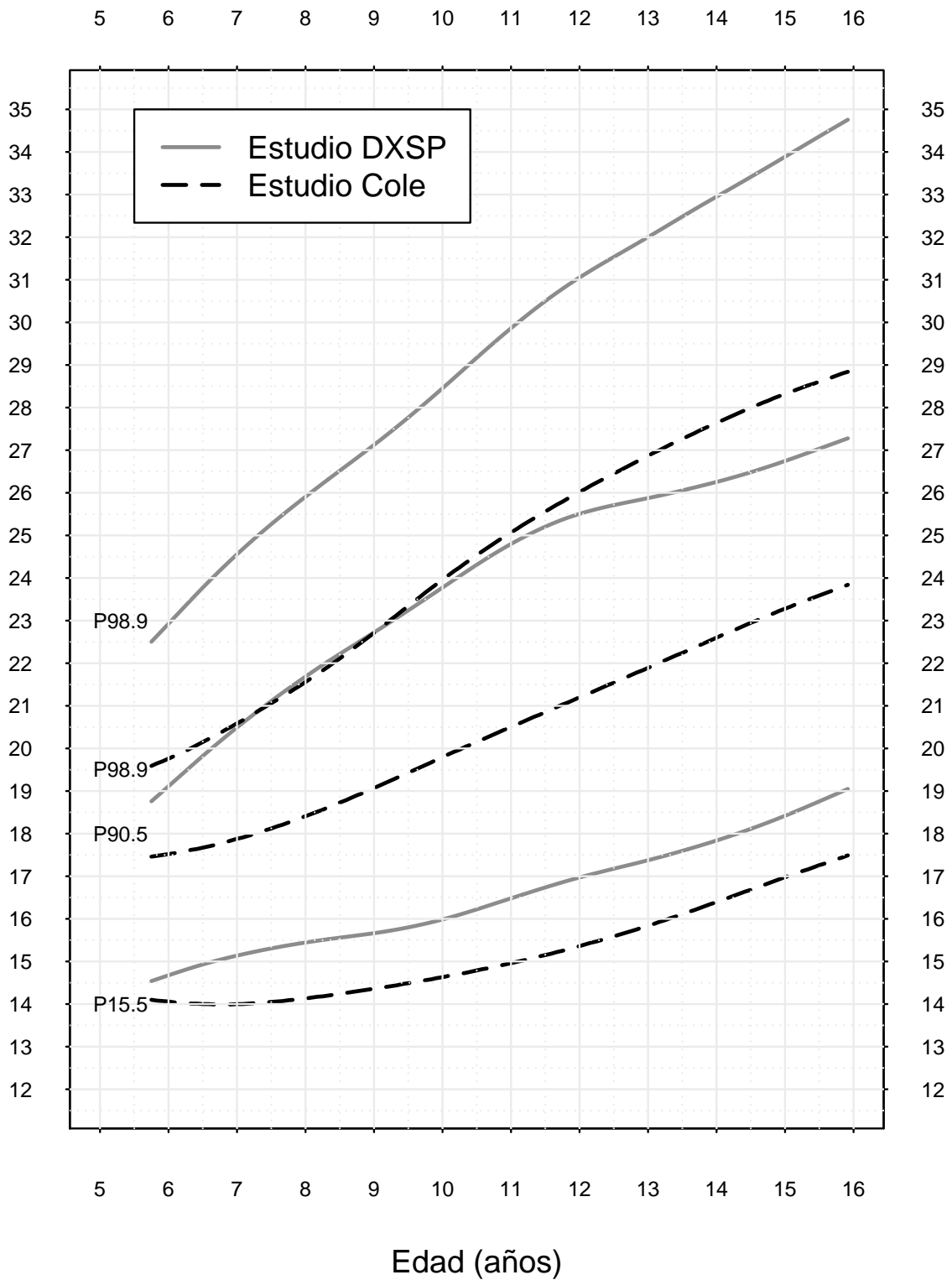


Figura 8.3: Comparación de los percentiles del IMC que definen el bajo peso, sobrepeso y obesidad de los niños con los datos del estudio de DXSP (sólido negro) y con los del estudio de Cole (discontinuo gris).

8.4. Cole en niñas

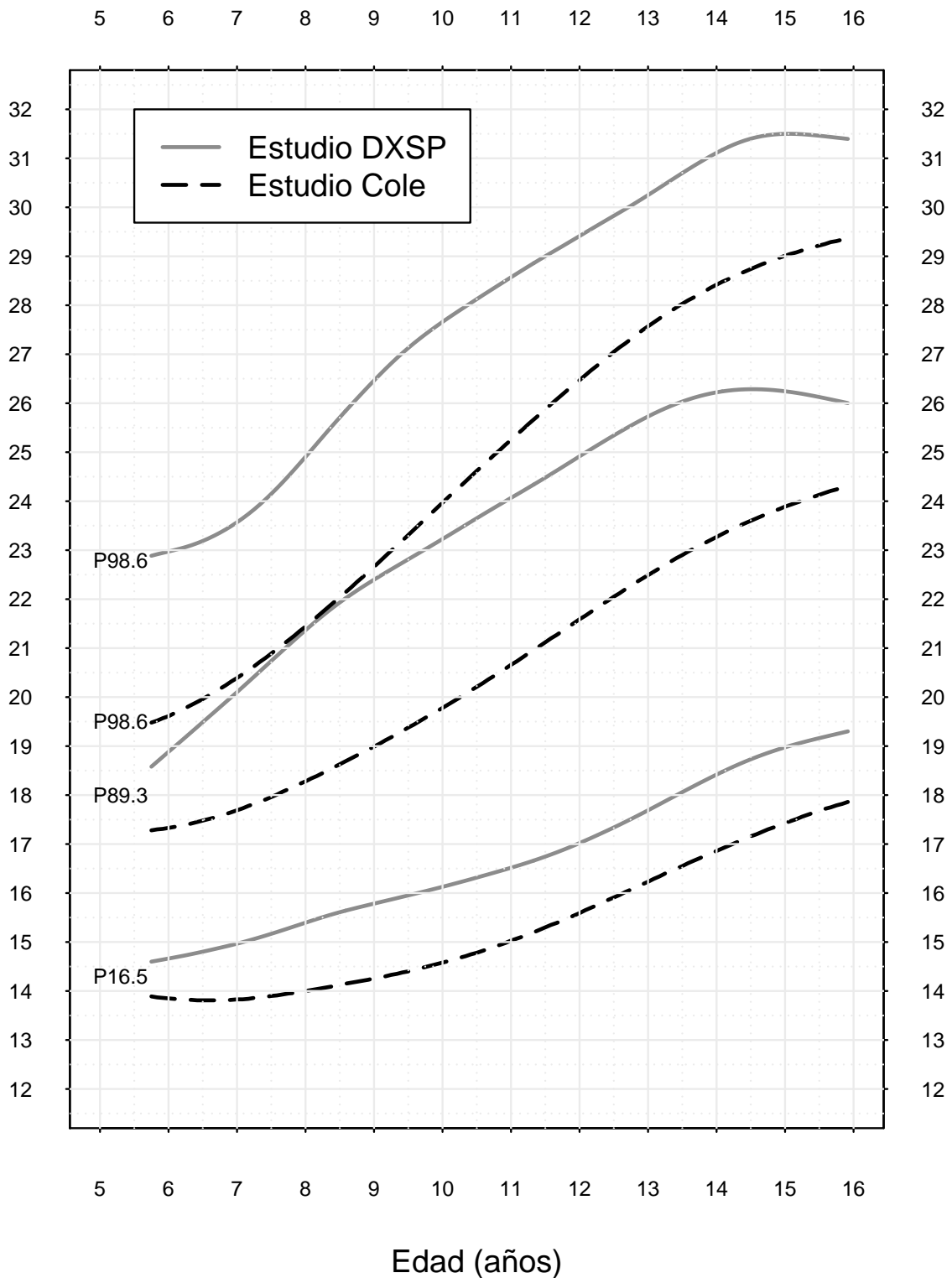


Figura 8.4: Comparación de los percentiles del IMC que definen el bajo peso, sobrepeso y obesidad de las niñas con los datos del estudio de DXSP (sólido negro) y con los del estudio de Cole (discontinuo gris).

8.5. OMS en niños

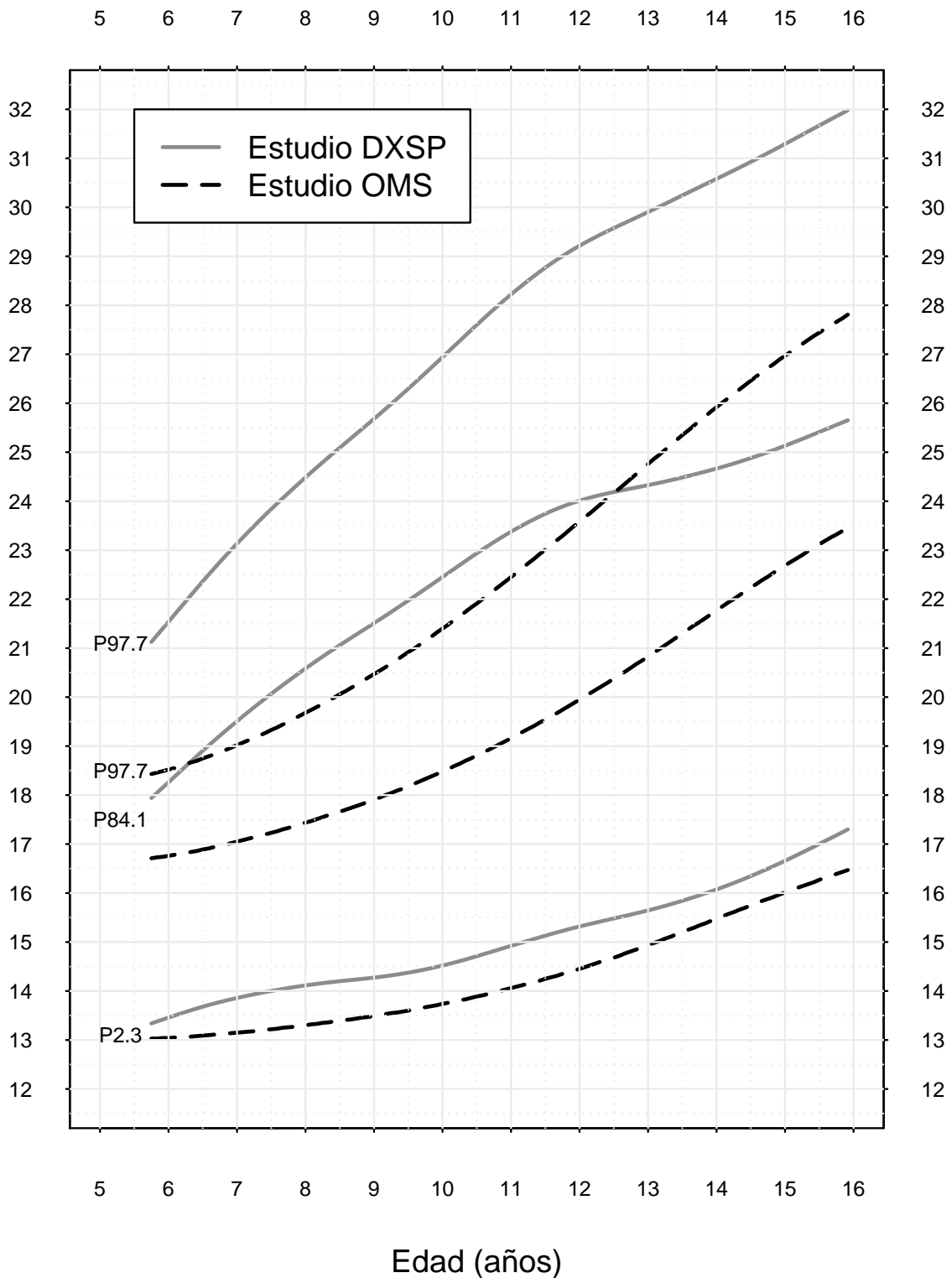


Figura 8.5: Comparación de los percentiles del IMC que definen el bajo peso, sobrepeso y obesidad de los niños con los datos del estudio de DXSP (sólido negro) y los del estudio de la OMS (discontinuo gris).

8.6. OMS en niñas

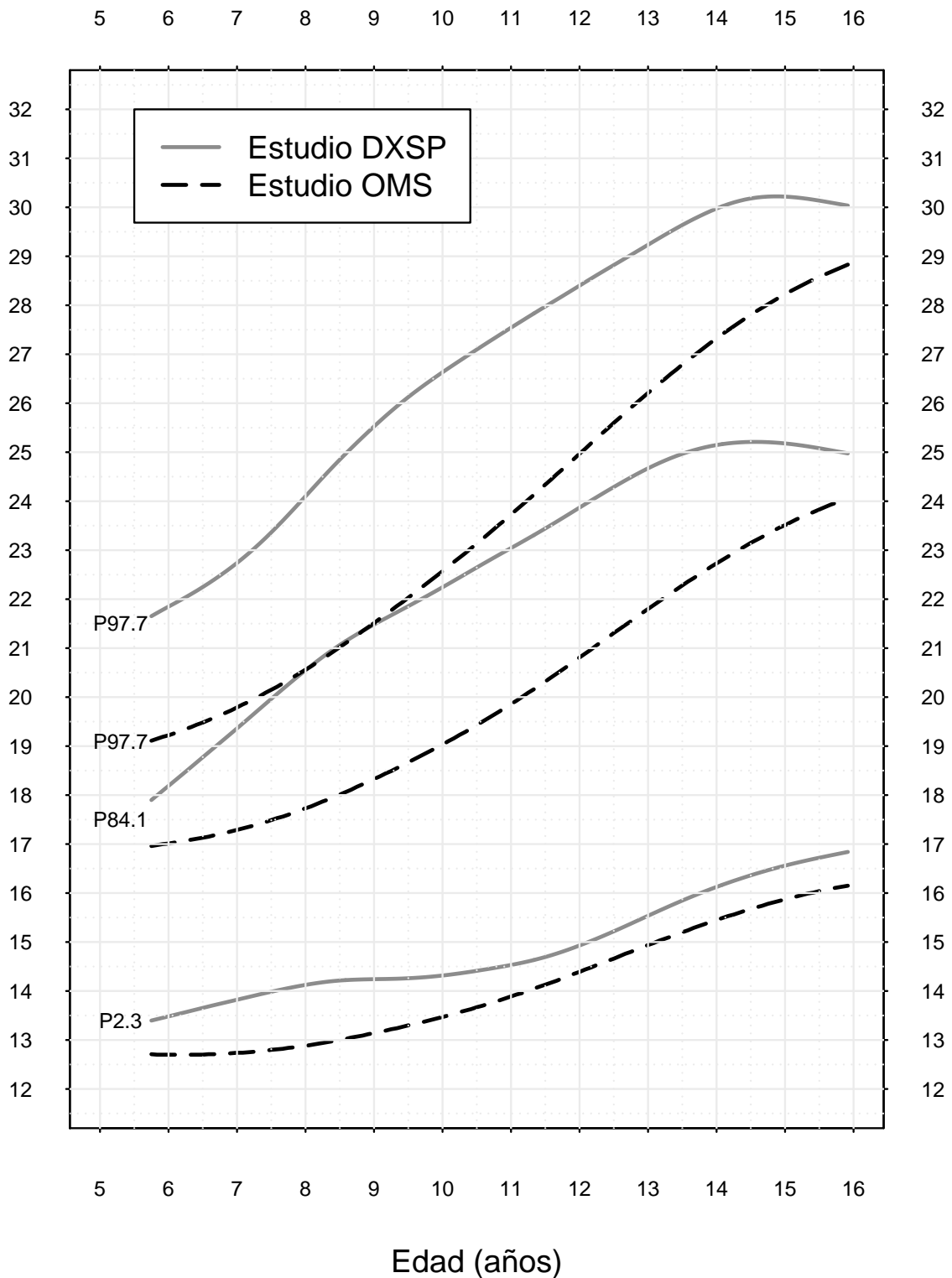


Figura 8.6: Comparación de los percentiles del IMC que definen el bajo peso, sobrepeso y obesidad de las niñas con los datos del estudio de DXSP (sólido negro) y los del estudio de la OMS (discontinuo gris).



## 8.7. Comentarios sobre las gráficas

1. **Orbegozo:** en la Figura 8.1 se puede ver que si se comparan los valores de IMC de los niños obtenidos en el estudio Orbegozo y en el estudio de la DXSP se observa que en el percentil 3 los valores de IMC son similares en ambos estudios. Sin embargo, al comparar los percentiles que definen sobrepeso y obesidad se observa que los valores del IMC obtenidos en el estudio de Galicia son más altos, si bien la evolución de las curvas es similar.

En las niñas (Figura 8.2) se observa algo similar a lo observado en niños. Así, en el percentil 10 no se observan diferencias en la evolución del IMC en función del estudio. Pero las diferencias si se observan en los percentiles más altos, en este caso en el 89 y en el 99, donde los valores del estudio de Galicia son más altos que los derivados del estudio de Orbegozo, y las diferencias son mayores a partir de los 10 años.

2. **Estudios internacionales (OMS y Cole):** Las diferencias entre los valores del IMC en función del estudio utilizado se hacen máximas al comparar los datos de la DXSP con los obtenidos en estudios internacionales (OMS y Cole).

Los valores del IMC de los niños en los 3 percentiles mostrados en las Figuras 8.3 y 8.5 son más altos en el estudio de la DXSP. La evolución del percentil de obesidad del IMC (97,7 en OMS y 98,9 en Cole) es similar en los tres estudios, pero en cambio se observa una diferente evolución en el percentil medio, el de sobrepeso, (84,1 en OMS y 90,5 en Cole). Así, en el estudio de Cole este percentil sigue una tendencia prácticamente lineal, y en el de la OMS es una curva convexa, mientras que en el estudio de Galicia se observa que a los 12 años la curva del percentil de sobrepeso, tanto el P84,1 como el P90,5, pasa de cóncava a convexa.

Por otra parte, en la Figura 8.3 se observa que el percentil 90,5 de la DXSP coincide entre los 7 y los 10 años con el percentil de obesidad del estudio de Cole, por lo que aproximadamente el 10% de los niños gallegos de estas edades tienen obesidad según el criterio de Cole. Sin embargo, la prevalencia de obesos es menor a partir de los 12 años, pues el P90,5 de la DXSP está por debajo del percentil de obesidad de Cole.

Al comparar los valores de la evolución del IMC en niñas en función de la edad (Figuras 8.4 y 8.6) en los percentiles 2,3, 84,1 y 97,7 (OMS) o en los 16,5, 89,3 o 98,6 (Cole) con el estudio de la DXSP se observa que los valores son diferentes y más altos en el estudio de la DXSP a cualquier edad. Las diferencias son máximas entre los 8 y los 14 años en los percentiles de sobrepeso y obesidad, tanto en Cole como en OMS.

## Capítulo 9

# Conclusiones y limitaciones

### 9.1. Conclusiones

Este trabajo ha permitido estimar las curvas percentiles de IMC, peso, talla, perímetro de cintura y perímetro de cadera en niños y niñas gallegos de 6 a 15 años utilizando la metodología clásica para este tipo de curvas: los métodos LMS o BCPE.

Ya se comentaron, a lo largo del trabajo, algunas de las ventajas de esta metodología, entre las que destacan la existencia de fórmulas explícitas para el cálculo de cualquier percentil deseado y la posibilidad de comparar los resultados de este estudio con otros estudios nacionales e internacionales. Por otro lado, estos modelos mostraron cierta sensibilidad a valores atípicos extremos, sobre todo, en penalizaciones pequeñas. Esto se manifestó con valores grandes de los grados de libertad estimados, que generaban curvas percentiles complejas y poco suaves. Por ello se optó por eliminar los atípicos extremos de la muestra.

La comparación entre las curvas percentiles de niños y niñas evidencia que el aumento de peso es similar en ambos sexos hasta los 13-14 años (un poco antes en percentiles altos), y a partir de la pubertad, las niñas tienden a estabilizar su peso y los niños continúan creciendo. El crecimiento en la talla también es similar hasta una edad que oscila entre los 12 y los 14 años, dependiendo del percentil, y después ocurre lo mismo que con el peso.

Todas las curvas percentiles del perímetro de cintura en los niños están por encima de las curvas de las niñas, aunque la diferencia es pequeña hasta los 8-9 años. Con el perímetro de cadera sucede al revés, en general las curvas de las niñas están por encima, aunque los percentiles superiores a la mediana son muy similares hasta los 11 años.

Tanto en niñas como en niños, los percentiles de IMC que definen bajo peso en el estudio Orbeagoz son similares a los percentiles del mismo orden obtenidos con el estudio de la DXSP. Sin embargo, las curvas de la DXSP están por encima de las correspondientes al sobrepeso y a la obesidad en el estudio Orbeagoz, con diferencias que van desde 1 hasta 3  $kg/m^2$  aproximadamente.

Al comparar los percentiles de IMC del estudio de la DXSP con los valores correspondientes a los puntos de corte de bajo peso, sobrepeso y obesidad de los dos estudios internacionales, OMS y Cole, las curvas de la DXSP están siempre por encima, con diferencias importantes en los dos percentiles

superiores.

## 9.2. Limitaciones

Los datos de este estudio proceden de una muestra representativa de la población escolar de Galicia de 6 a 15 años, y el tamaño de muestra, superior a 7.000 niños y niñas, es muy importante.

Las curvas percentiles de crecimiento infantil se construyen habitualmente para niños y niñas de 2 a 18 años, y es aconsejable tener datos de niños con edades inferiores a los 2 años y superiores a los 18 (por ejemplo, desde 1 año hasta los 20 años) para aproximar bien las estimaciones en las edades extremas. En este trabajo el rango de edad de los niños estaba limitado a los 6-15 años, lo que no permitió obtener las curvas completas de crecimiento. Al no poder estimar los percentiles hasta los 18 años, no fue posible identificar los percentiles de IMC que, a los 18 años, pasan por los puntos de corte que definen bajo peso (18,5), sobrepeso (25) y obesidad (30) en adultos. Esto hubiese permitido establecer unos valores de referencia para la clasificación del estado ponderal de los niños.

# Bibliografía

- [1] Castro Rocha Duarte E. M. (2017) Statistical contributions in modeling breast cancer data through structured additive regression (*star*). Tesis, Universidade de Santiago de Compostela. [MT7]
- [2] Cole T. J. and Green P. J. (1992) Smoothing reference centile curves: The lms method and penalized likelihood. *Stat. Med.* (10), 1305–1319.
- [3] Cole T. J. and Lobstein T. (2012) Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. *Pediatric Obesity*, 7, 284–294.
- [4] Cole T. J., Bellizzi M. C., Flegal K. M. and Dietz W. H. (2000) Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ*, 320, 1240–45.
- [5] Cole T. J., Flegal K. M., Nicholls D. and Jackson A. A. (2007) Body mass index cut offs to define thinness in children and adolescents: international survey. *BMJ*.
- [6] Cole T. J., Freeman J. V., and Preece M. A. (1998) British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Stat Med.* 17(4), 407–429
- [7] De Onis M., Onyango A. W., Borghi E., Siyam A., Nishida C., Siekmanna J. (2007) Development of a WHO growth reference for school-aged children and adolescents. *Bulletin of the World Health Organization* (85), 660–7.
- [8] Fernández C, Lorenzo H, Vrotsou K, Aresti U, Rica I, Sánchez E. (2011) Estudio de crecimiento de Bilbao. Curvas y tablas de crecimiento. Estudio transversal. Instituto de Investigación sobre Crecimiento y Desarrollo, Fundación Faustino Orbegozo.
- [9] Fernández C., Lorenzo H., Vrotsou K., Aresti U., Rica I. y Sánchez E. (2011) Estudio de crecimiento de bilbao curvas y tablas de crecimiento (Estudio transversal).
- [10] Green P. J. and Silverman B. W. (1994) *Nonparametric Regression and Generalized Linear Models*. London: Chapman and Hall.
- [11] Hernández M., Castellet J., Narvaiza J. L. , Rincón J. M., Ruiz I, Sánchez E, et al. Curvas y tablas de crecimiento de 0 a 18 años. Instituto de Investigación sobre Crecimiento y Desarrollo, Fundación Faustino Orbegozo. Madrid: Editorial Garsi. 1988

- [12] <https://www.fundacionorbegozo.com/el-instituto-de-investigacion-del-crecimiento-y-desarrollo/graficas-y-tablas/>
- [MT2] oenker R. (2005) Quantile Regression. Cambridge University Press, Cambridge.
- [13] Koenker R. and Bassett G. Jr. (1978) Regression Quantiles. *Econometrica*, 46(1), 33–50.
- [14] Lezcano A. C., García J. M. F., Ramos C. F., Longás A. F., Sigüero J. P. L., González E. S., Ruiz B. S., Fernández D. Y. y Grupo Colaborador Español (2008). Estudio transversal español de crecimiento 2008. Parte II: valores de talla, peso e índice de masa corporal desde el nacimiento a la talla adulta. *An Pediatr (Barc)*. 68(6), 552–69.
- [15] Obesidade infantil en Galicia no ano 2013 outubro de 2014. *BEG* 2014 26(5). ([https://www.sergas.es/gal/documentacionTecnica/docs/SaudePublica/begs/BEG\\_XXVI-5.pdf](https://www.sergas.es/gal/documentacionTecnica/docs/SaudePublica/begs/BEG_XXVI-5.pdf))
- [16] Reinsch, C. (1967). Smoothing by spline functions. *Numerische Mathematik*, 10, 177–183.
- [17] Rigby R. A. and Stasinopoulos D. M. (2005) Generalized additive models for location, scale and shape. *Appl. Statist.* 54, 507–554.
- [18] Rigby R. A. and Stasinopoulos D. M. (2007) Generalized Additive Models for Location Scale and Shape (GAMLSS) in R, (23).
- [19] Rigby R. A., Stasinopoulos D. M. (2004) Smooth centile curves for skew and kurtotic data modelled using the Box–Cox power exponential distribution. *Stat Med.* (23). 3053–76.
- [20] Sobradillo B., Aguirre A., Aresti U., Bilbao A., Fernández-Ramos C., Lizárraga A., Lorenzo H., Madariaga L., Rica I., Ruiz I., Sánchez E., Santamaría C., Serrano J. M., Zabala A., Zurimendi B. and Hernández M. (2004) Curvas y tablas de crecimiento (Estudios longitudinal y transversal). Instituto de Investigación sobre Crecimiento y Desarrollo, Fundación Faustino Orbegozo.
- [21] Stasinopoulos M., Rigby B. and Akantziliotou C. (2008) Instructions on how to use the gamlss package in R Second Edition.
- [22] van Buuren S. and Fredriks M. (2001) Worm plot: a simple diagnostic device for modelling growth reference curves. *Statist. Med.* (20) 1259–1277.