

INSTITUTO POLITÉCNICO NACIONAL
ESCUELA SUPERIOR DE FÍSICA Y
MATEMÁTICAS



MODELACIÓN DEL ÍNDICE DE MASA
CORPORAL DE ADULTOS MEXICANOS Y SU
RELACIÓN CON EL PORCENTAJE DE GRASA
CORPORAL

Tesis que, para obtener el título de Licenciado en Física
y Matemáticas, presenta

Leticia Escobar Zaragoza

Director: Dra. Eloisa Díaz-Francés Murguía

Director interno: Dr. Roberto S. Acosta Abreu

México D. F., 25 de Marzo del 2009.

A mis amados padres: Timoteo y Maria
y mis queridas hermanas Lilian y Lya

Agradecimientos

Quiero expresar mi agradecimiento a dos instituciones que han contribuido en mi formación académica: al Instituto Politécnico Nacional y la Escuela Superior de Física y Matemáticas y al Centro de Investigación en Matemáticas, A.C., por el apoyo económico durante mi estancia para realizar la tesis de licenciatura.

Los datos analizados en esta tesis provienen de la Encuesta Nacional de Salud y Nutrición 2006, diseñada y conducida por el Instituto Nacional de Salud Pública.

Extiendo mi agradecimiento al Dr. Ramón Sebastián Salat Figols, Dr. José María Rocha Martínez y al M. en C. Raymundo García Zamudio por haber aceptado ser sinodales para este trabajo de tesis. Al Dr. Roberto Segundo Acosta Abreu por la guía y apoyo brindado durante mi estancia en ESFM además de haber aceptado co-dirigir este trabajo. Al Dr. José Luis Castro Quilantán por la atención y apoyo durante el proceso de titulación. Especialmente quiero agradecer a la Dra. Eloísa Díaz-Francés Murguía, por aceptarme como alumna y tener el privilegio de trabajar a su lado, además por el apoyo y consejos en el ámbito académico y personal, pues gracias a ellos lleve a término este trabajo.

Ante todo quiero expresar mi agradecimiento a las personas más amadas y que han sido la fuente de inspiración más grande: Mis padres Timoteo Escobar y María Zaragoza y mis hermanas Lilian E. Z. y Lya E. Z. Particularmente agradezco a mi querida amiga Blanca Lucia Moreno Ley por su amistad, apoyo y comprensión durante varios años que hicieron fácilmente soportable la tensión y el trabajo de la Licenciatura.

De igual manera agradezco a todos aquellos que han sido una excelente compañía y apoyo durante los periodos de mi estancia en IPN-ESFM y CIMAT, Gto.: Evelia Cortés, Miriam Martínez, Imelda Santillán, Angel Alegría, Miguel Angel Martínez, Hugo D. Sánchez, Joel Iglesias, Selomit Ramírez, Carolina Creoglio, Roxana Góngora y Atanasio Vega.

Prefacio

El propósito principal de este trabajo es modelar estadísticamente el índice de masa corporal (IMC) de mexicanos. Un modelo estadístico adecuado podrá utilizarse para tomar decisiones políticas de salud relevantes. El IMC se ha utilizado como un indicador de riesgos cardiovasculares y también para determinar obesidad en hombres y mujeres. Es barato y fácil de obtener pues sólo requiere medir la talla y peso de la persona. El IMC es el cociente del peso de la persona dividido entre su altura, medida en metros y elevada al cuadrado. Actualmente es el indicador más usado para predecir el porcentaje de grasa corporal en una persona, por su bajo costo y sencillez. Sin embargo, en México no se cuenta con un modelo estadístico general que describa su distribución.

En la comunidad médica el IMC ha sido de gran aceptación por ser de bajo costo y sencillo de calcular. Es por ello que diversos organismos del sector salud como la OMS sugieren utilizarlo como indicador del estado nutricional de un individuo. Para México, la Encuesta Nacional de Salud y Nutrición, levantada en todo el país por el Instituto Nacional de Salud Pública en 2006 (ENSANUT 2006) aporta una base de datos, donde es posible realizar un análisis del IMC para la población de adultos mexicanos. En dicha encuesta fueron obtenidas diversas medidas antropomórficas, como talla y peso, por lo que se cuenta con el IMC.

En esta tesis se desea encontrar un modelo estadístico adecuado para el IMC para hombres y mujeres mexicanos apartir de esta encuesta. Las ventajas de contar con modelos estadísticos adecuados para el IMC son diversas e importantes, pues esto ayuda a cuantificar la probabilidad de presentar problemas de sobrepeso, obesidad y riesgos cardiovasculares en la población mexicana. Por otra parte, se podrán utilizar estos resultados como base para diversas políticas de salud a nivel nacional y estatal. A la fecha no se ha modelado el IMC en México con anterioridad, si bien se han realizado

estudios para describir la talla y peso, sobre todo de niños y adolescentes mexicanos.

En el Capítulo 1 se hace una breve descripción del marco teórico sobre el que se sustentan los resultados producto de esta tesis. En el Capítulo 2 se presenta y se modela estadísticamente los datos considerados. Esto incluye la propuesta de un modelo estadístico, el proceso de estimación de los parámetros de este modelo y la validación del modelo estimado con los datos observados. Además de los de la encuesta mencionada arriba, se consideraron datos de pacientes de la Dra. Yolanda Méndez Romero, del Hospital Aranda de la Parra en León, Guanajuato, a los que se les evaluó a través de pliegues corporales su porcentaje de grasa corporal. Finalmente en el Capítulo 3 se presentan las conclusiones generales de este trabajo.

Los cálculos realizados se efectuaron con Matlab, V. 7, Statistica (de Statsoft Inc.), SPSS y Scientific Word en una computadora personal, durante mi estancia para realizar la tesis de licenciatura en el Centro de Investigación en Matemáticas ubicado en la ciudad de Guanajuato.

Para facilitar la lectura de esta tesis, las figuras y tablas del Capítulo 2 se pasaron a los Apéndices A y B.

Contenido

1	Introducción	1
1.1	Índice de masa corporal	1
1.2	Síndrome Metabólico	6
1.3	Conceptos básicos de modelación estadística	7
1.3.1	Modelación Estadística	8
1.3.2	Función de Verosimilitud	9
1.3.3	Intervalos de Verosimilitud-Confianza	11
1.3.4	Función de verosimilitud perfil	13
1.3.5	Validación del modelo estadístico en consideración.	16
1.3.6	Ejemplo de tiempos de caídas de canicas	22
1.4	Modelos estadísticos para el IMC	27
1.4.1	Modelo Lognormal	28
1.4.2	Los modelos de Box-Cox	30
2	Datos antropométricos de Mexicanos	37
2.1	Introducción	37
2.2	Modelación estadística de datos de ENSANUT 2006	39
2.2.1	Modelos para el IMC por estado	40
2.2.2	Modelos para el IMC a nivel nacional	48
2.2.3	Análisis de resultados	53

2.3	Porcentaje de grasa corporal y su relación con el IMC	54
2.4	Conclusiones	58
3	Conclusiones	60
	Apéndice A. Figuras de Capítulo 2	63
	Apéndice B. Tablas del Capítulo 2	117

Lista de Figuras

1.1	Verosimilitud perfil relativa	15
1.2	Verosimilitud perfil relativa para μ	26
1.3	Verosimilitud perfil relativa para σ	27
2.1	Verosimilitud relativa de λ para hombres por estados	64
2.2	Verosimilitud relativa de λ para hombres por estados	64
2.3	Verosimilitud relativa de λ para hombres por estados	65
2.4	Verosimilitud relativa de λ para hombres por estados	65
2.5	Verosimilitud relativa de λ para mujeres por estados	66
2.6	Verosimilitud relativa de λ para mujeres por estados	66
2.7	Verosimilitud relativa de λ para mujeres por estados	67
2.8	Verosimilitud relativa de λ para mujeres por estados	67
2.9	Hombres de Aguascalientes. Ajuste con el Modelo Lognormal y gráfica Q-Q	68
2.10	Mujeres de Aguascalientes. Ajuste con el Modelo Lognormal y gráfica Q-Q	68
2.11	Hombres de Baja California. Ajuste con el Modelo Lognormal y gráfica Q-Q	69
2.12	Mujeres de Baja California. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ y gráfica Q-Q	69

2.13 Hombres de Baja California Sur. Ajuste con el Modelo Lognormal y gráfica Q-Q	70
2.14 Mujeres de Baja California Sur. Ajuste con el Modelo de Box-Cox $\lambda=0.5$ y gráfica Q-Q	70
2.15 Hombres de Campeche. Ajuste con el Modelo Lognormal y gráfica Q-Q	71
2.16 Mujeres de Campeche. Ajuste con el Modelo Lognormal y gráfica Q-Q	71
2.17 Hombres de Coahuila. Ajuste con el Modelo de Box-Cox $\lambda=0.5$ y gráfica Q-Q	72
2.18 Mujeres de Coahuila. Ajuste con el Modelo Lognormal y gráfica Q-Q .	72
2.19 Hombres de Colima. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	73
2.20 Mujeres de Colima. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	73
2.21 Hombres de Chiapas. Ajuste con el Modelo de Box-Cox $\lambda=-1$ y gráfica Q-Q	74
2.22 Mujeres de Chiapas. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	74
2.23 Hombres de Chihuahua. Ajuste con el Modelo Lognormal y gráfica Q-Q	75
2.24 Mujeres de Chihuahua. Ajuste con el Modelo Lognormal y gráfica Q-Q	75
2.25 Hombres del Distrito Federal. Ajuste con el Modelo de Box-Cox $\lambda=0.5$ y gráfica Q-Q	76
2.26 Mujeres del Distrito Federal. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ y gráfica Q-Q	76
2.27 Hombres de Durango. Ajuste con el Modelo Lognormal y gráfica Q-Q .	77
2.28 Mujeres de Durango. Ajuste con el Modelo Lognormal y gráfica Q-Q .	77
2.29 Hombres de Guanajuato. Ajuste con el Modelo Lognormal y gráfica Q-Q	78
2.30 Mujeres de Guanajuato. Ajuste con el Modelo Lognormal y gráfica Q-Q	78
2.31 Hombres de Guerrero Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ y gráfica Q-Q	79

2.32	Mujeres de Guerrero. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ y gráfica Q-Q	79
2.33	Hombres de Hidalgo. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	80
2.34	Mujeres de Hidalgo. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	80
2.35	Hombres de Jalisco. Ajuste con el Modelo de Box-Cox $\lambda=0.5$ y gráfica Q-Q	81
2.36	Mujeres de Jalisco. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	81
2.37	Hombres de México. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	82
2.38	Mujeres de México. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	82
2.39	Hombres de Michoacán. Ajuste con el Modelo Lognormal y gráfica Q-Q	83
2.40	Mujeres de Michoacán. Ajuste con el Modelo Lognormal y gráfica Q-Q	83
2.41	Hombres de Morelos. Ajuste con el Modelo Lognormal y gráfica Q-Q .	84
2.42	Mujeres de Morelos. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	84
2.43	Hombres de Nayarit . Ajuste con el Modelo Lognormal y gráfica Q-Q .	85
2.44	Mujeres de Nayarit. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	85
2.45	Hombres del Nuevo León. Ajuste con el Modelo de Box-Cox $\lambda=0.5$ y gráfica Q-Q	86
2.46	Mujeres de Nuevo León. Ajuste con el Modelo Lognormal y gráfica Q-Q	86
2.47	Hombres de Oaxaca. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	87
2.48	Mujeres de Oaxaca. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	87
2.49	Hombres de Puebla. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ y gráfica Q-Q	88
2.50	Mujeres de Puebla. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	88
2.51	Hombres de Querétaro. Ajuste con el Modelo Lognormal y gráfica Q-Q	89
2.52	Mujeres de Querétaro. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ y gráfica Q-Q	89

2.53 Hombres de Quintana Roo. Ajuste con el Modelo Lognormal y gráfica Q-Q	90
2.54 Mujeres de Quintana Roo. Ajuste con el Modelo Lognormal y gráfica Q-Q	90
2.55 Hombres de San Luis Potosí. Ajuste con el Modelo de Box-Cox $\lambda=0.5$ y gráfica Q-Q	91
2.56 Mujeres de San Luis Potosí. Ajuste con el Modelo Lognormal y gráfica Q-Q	91
2.57 Hombres de Sinaloa. Ajuste con el Modelo de Box-Cox $\lambda=0.5$ y gráfica Q-Q	92
2.58 Mujeres de Sinaloa. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ y gráfica Q-Q	92
2.59 Hombres de Sonora. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	93
2.60 Mujeres de Sonora. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	93
2.61 Hombres de Tabasco. Ajuste con el Modelo Ajuste con el Modelo de Box-Cox $\lambda=0.5$ y gráfica Q-Q	94
2.62 Mujeres de Tabasco. Ajuste con el Modelo Lognormal y gráfica Q-Q . .	94
2.63 Hombres de Tamaulipas. Ajuste con el Modelo Lognormal y gráfica Q-Q	95
2.64 Mujeres de Tamaulipas. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ y gráfica Q-Q	95
2.65 Hombres de Tlaxcala. Ajuste con el Modelo Lognormal y gráfica Q-Q .	96
2.66 Mujeres de Tlaxcala. Ajuste con el Modelo Lognormal y gráfica Q-Q .	96
2.67 Hombres de Veracruz. Ajuste con el Modelo Lognormal y gráfica Q-Q .	97
2.68 Mujeres de Veracruz. Ajuste con el Modelo Lognormal y gráfica Q-Q .	97
2.69 Hombres de Yucatán. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ y gráfica Q-Q	98
2.70 Mujeres de Yucatán. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ y gráfica Q-Q	98

2.71	Hombres de Zacatecas. Ajuste con el Modelo Lognormal y gráfica Q-Q	99
2.72	Mujeres de Zacatecas. Ajuste con el Modelo Lognormal y gráfica Q-Q	99
2.73	Aguascalientes, gráfica Q-Q, modelo Gumbel(29.1075, 5.2102)	100
2.74	Hombres de Aguascalientes. 1000 repeticiones, con el modelo Lognormal	100
2.75	Hombres de Baja California. 1000 repeticiones, con el modelo Lognormal	101
2.76	Hombres de Campeche. 1000 repeticiones, con el Modelo Lognormal	101
2.77	Hombres de Coahuila. 1000 repeticiones, con el Modelo de Box-Cox $\lambda=0.5$	102
2.78	Hombres de Colima. 1000 repeticiones, con el Modelo Lognormal	102
2.79	Hombres de Chihuahua. 1000 repeticiones, con el Modelo Lognormal	103
2.80	Hombres del Distrito Federal. 1000 repeticiones, con el Modelo de Box-Cox $\lambda=0.5$	103
2.81	Hombres de Sonora. 1000 repeticiones, con el Modelo Lognormal	104
2.82	Hombres de Tabasco. 1000 repeticiones, con el Modelo de Box-Cox $\lambda=0.5$	104
2.83	Mujeres de Chihuahua. 1000 repeticiones, con el Modelo de Box-Cox $\lambda=0.5$	105
2.84	Cuantiles estimados de hombres por estado. Ordenados respecto de Cuantil 0.85	106
2.85	Cuantiles estimados de mujeres por estado. Ordenados respecto del Cuantil 0.85	107
2.86	Verosimilitud relativa de adultos mexicanos	108
2.87	Ajuste con el Modelo de Mezcla y el modelo Lognormal. Hombres	108
2.88	Ajuste con el Modelo de Mezcla y el modelo Lognormal. Mujeres	109
2.89	Gráfica de probabilidades de los modelos propuestos. Hombres Nacional	109
2.90	Gráfica de probabilidades de los modelos propuestos. Mujeres Nacional	110
2.91	Hombres Nacional. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)	110

2.92	Mujeres Nacional. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)	111
2.93	Hombres Nacional. 1000 repeticiones, con el Modelo Lognormal	111
2.94	Mujeres Nacional. 1000 repeticiones, con el Modelo Lognormal	112
2.95	Dispersión del porcentaje de grasa vs. IMC. Hombres y mujeres de León	112
2.96	Dispersión del porcentaje de grasa vs. lnIMC. Hombres de León	113
2.97	Dispersión del porcentaje de grasa vs. lnIMC. Mujeres de León	113
2.98	Hombres adultos con obesidad y colesterol alto	114
2.99	Mujeres adultas con obesidad y colesterol alto	114
2.100.	Hombres adultos con obesidad y obesidad androide	115
2.101.	Mujeres adultas con obesidad y obesidad androide	115
2.102.	Hombres adultos con Síndrome metabólico	116
2.103.	Mujeres adultos con Síndrome metabólico	116

Lista de Tablas

1.1	Clasificación del IMC de diferentes organizaciones.	4
1.2	Clasificación del estado nutricional de mexicanos.	5
1.3	Relaciones entre nivel de confianza y de verosimilitud para un parámetro unidimensional y cuantiles correspondientes de una Ji-cuadrada de un grado de libertad	12
1.4	Ejemplo: Tiempos de caída de canicas	22
1.5	Intervalos de verosimilitud-confianza para los datos de canicas	26
2.1	Elección de una λ con significado. Ejemplos para hombres.	42
2.2	Elección de valores de λ para los 32 estados mexicanos	44
2.3	Modelo Lognormal global para hombres y mujeres mexicanos.	49
2.4	Criterio de Akaike en la elección de un modelo más simple.	50
2.5	Coeficientes de asimetría y curtosis, modelo Lognormal a nivel nacional	51
2.6	Porcentajes estimados de la clasificación nutricional para adultos mexicanos	53
2.7	Coeficientes C y M de Durnin y Womersley (1967)	56
B.1	Cuantiles teóricos estimados para adultos mexicanos	118
B.2	Proporción de individuos por estado. Datos proporcionados por INEGI en el censo del año 2000	118
B.3	Tamaño de muestras del IMC en adultos mexicanos.	119

B.4	Inferencia para λ en la familia de Box-Cox. Hombres	120
B.5	Inferencia para λ en la familia de Box-Cox. Mujeres	121
B.6	Modelo de Box-Cox propuesto de adultos mexicanos.	122
B.7	Coeficientes de asimetría y curtosis de los datos transformados para hombres	123
B.8	Coeficientes de asimetría y curtosis de los datos transformados para mujeres	124
B.9	Porcentaje estimado del estado nutricional de hombres por entidad fe- derativa	125
B.10	Porcentaje estimado del estado nutricional de mujeres por entidad fe- derativa	126
B.11	Porcentaje empírico del estado nutricional de hombres por entidad fe- derativa	127
B.12	Porcentaje empírico del estado nutricional de mujeres por entidad fe- derativa	128

Capítulo 1

Introducción

1.1 Índice de masa corporal

El objetivo principal de esta tesis es proponer un modelo estadístico, que describa bien a la distribución del índice de masa corporal (IMC) en la población de adultos mexicanos. Lambert Adolphe Jacques Quetelet, en 1841 lo definió por primera vez como:

$$\text{IMC} = \frac{w}{h^2}, \quad (1.1)$$

donde w es el peso de la persona en kilogramos y h la talla correspondiente en metros elevada al cuadrado. Quetelet propuso un modelo Gaussiano, acampanado y simétrico, para la distribución del IMC en la población francesa de esa época, donde no prevalecía tan fuertemente la obesidad como hoy en día. Recientemente en el área de la salud, el IMC ha sido utilizado como herramienta para explicar y valorar diversos problemas importantes. Por ejemplo, es el indicador principal para el sobrepeso y la obesidad, ya que posee como atributos ser sencillo de obtener e interpretar. A continuación se presentan ejemplos tomados de la literatura científica donde contar con un modelo estadístico sería de gran utilidad para poder cuantificar la probabilidad de que el IMC tome ciertos valores en intervalos de interés:

1. Rotimi *et al.* (1995), Cox *et al.* (1998) mencionan que el sobrepeso y la obesidad han comenzado a ser tratados de manera simultánea con enfermedades crónicas como la Diabetes mellitus y del tipo cardiovascular, pues además de deteriorar el estado de salud por sí solos, se ha mostrado que son un indicador de riesgo de adquirir dichas enfermedades.
2. Seidell *et al.* (1996) reportan que el IMC y el riesgo de mortalidad siguen una relación no monótona y por tanto es de suma importancia contar con un modelo adecuado para el IMC.
3. Kern *et al.* (1990), Bonadonna R. y Bonora E. (1997) señalaron que en la obesidad hay una menor respuesta a la actividad insulínica, que puede recuperarse tras la pérdida de peso. Por ello, es importante tener un modelo adecuado para el IMC y con ello poder definir mejor situaciones serias de obesidad.

En todos los trabajos anteriores se menciona la importancia de tener un modelo estadístico adecuado para el IMC, ya que ciertos puntos de corte e intervalos específicos resultan ser indicadores para la detección de enfermedades del tipo cardiovascular. Además con dicho modelo y sus cuantiles es posible realizar un análisis más detallado de la relación entre la obesidad y el género, la edad y el nivel socioeconómico o sus hábitos alimenticios. Luego el disponer de un modelo que sea adecuado para representar a la población mexicana es muy útil. Permite entre otras cosas, calcular probabilidades de sucesos como los mencionados en los ejemplos anteriores, de forma más precisa que utilizando la frecuencia de aparición de dicho suceso en la muestra observada.

El grupo internacional de trabajo en obesidad (IOTF) y la Organización Mundial de la Salud (OMS), han definido la obesidad como la epidemia del siglo XXI por las dimensiones adquiridas a lo largo de las últimas décadas, su impacto sobre la mortalidad, la calidad de vida y el gasto sanitario. En los países desarrollados el aumento progresivo de la obesidad se ha descrito como un problema emergente en los últimos

años. Martorell (2002) menciona que de manera similar sucede en determinadas áreas de países en vías de desarrollo.

En México, el Instituto Nacional de Salud Pública realiza una encuesta aleatoria representativa a nivel nacional, recabando información de relevancia en el sector salud: *la encuesta nacional de salud y nutrición* (ENSANUT 2006). Allí se define a la obesidad como el resultado de un desequilibrio entre ingestión y el gasto energético; mientras que la *Organización Mundial de la Salud* (OMS 1995) la define como la situación en la que el almacenamiento de grasa se acompaña de riesgos para la salud. Así se derivan dos problemas: determinar cuándo hay exceso de tejido adiposo y cuándo este exceso es perjudicial para la salud.

En los últimos años se han desarrollado técnicas para medir la composición corporal de una persona, los cuáles según su campo de aplicación pueden dividirse en: técnicas epidemiológicas, clínicas o de investigación, más detalles en Oria E. *et al.* (2002). La más precisa en la medición de masa grasa y masa libre de grasa, es la de activación de neutrones, pero su complejidad y precio hacen que sólo haya dos centros en el mundo que lo utilicen. Por ello, el método de referencia con el cual suelen compararse los demás es el de la densitometría obtenida por inmersión en agua. Aunque recientemente, se está difundiendo la absorciometría con rayos X de doble energía (DXA) como la técnica adecuada en la valoración del tejido adiposo. Aunque éstas y otras técnicas son muy precisas, la dificultad principal radica en el excesivo costo de los instrumentos y el personal que debe ser capacitado para el uso correcto de cada una de las técnicas. Teniendo lo anterior como referencia, dentro de la comunidad médica, siempre se han buscado marcadores de obesidad de fácil obtención, basados en el peso, la altura o la circunferencia de la cintura.

El IMC apartir de las recomendaciones del Colegio Británico de Médicos, se generaliza como una medida estándar para clasificar y determinar pesos corporales saludables

o con sobrepeso, ya que en general se correlaciona bien con la masa grasa. Por ejemplo en Vázquez (1999) se muestran índices de correlación para estas variables de 0.7 a 0.9. La primera encuesta nacional de salud y nutrición norteamericana (National Health and Nutrition Examination Survey I) realizada en el período 1971-1974 buscó el percentil 85 para definir la obesidad (considerando que los valores de la normalidad convenida, estaría entre los percentiles 15 y 85). Así pues valorando el IMC y el grosor de varios pliegues cutáneos, se señalaron como límites de obesidad los valores de 27.8 kg/m² en hombres y 27.3 kg/m² para mujeres. Sin embargo, a través del tiempo distintos trabajos habían mostrado una menor mortalidad para un IMC de 20-25 kg/m² y así se propusieron varias clasificaciones relativas al peso y sobrepeso. En la Tabla (1.1) se presentan las clasificaciones para la obesidad propuestas por diversos organismos como la OMS, la Sociedad Española para el Estudio de la Obesidad (SEEDO), que han ido coincidiendo en el valor de 30 kg/m² para definir a la obesidad.

	Grado	SEEDO	NHANESIII	OMS	SEEDO	OMS
AÑO		1995	1996	1998	2000	2000 y 2004
Normal		18-26.9	18-24.9	18-24.9	18-24.9	18.5-24.99
Sobrepeso	I		25.29.9	25-29.9	25-26.9	25-27.49
	II				27-29.9	27.49-29.99
Obesidad	I	27-29.9	30-34.9	30-34.9	30-34.9	30-34.99
	II	30-34.9	35-39.9	35-39.9	35-39.9	35-39.99
	III	35-39.9	≥40	≥40	40-49.9	≥40
	IV	≥40			≥50	

Tabla 1.1: Clasificación del IMC de diferentes organizaciones.

Sin embargo, en nuestro país, se consideran los lineamientos propuestos en la Norma Oficial Mexicana (1998) y el Manual de Procedimientos (2002), que consideran los límites mostrados en la Tabla (1.2), para definir el estado nutricional de mexicanos

apartir del IMC.

Peso	Normal		Sobrepeso		Obesidad I		Obesidad II		Obesidad III		Obesidad
Bajo	Mín	Máx	Mín	Máx	Mín	Máx	Mín	Máx	Mín	Máx	Mórbida
<18	18	24.9	25	26.9	27	29.9	30	34.9	35	39.9	≥ 40

Tabla 1.2: Clasificación del estado nutricional de mexicanos.

Aunque el IMC mantiene una buena correlación con la cantidad de grasa total del organismo en adultos de países desarrollados, esta relación no es tan buena en ancianos o en personas que hayan sufrido procesos catabolizantes. Las diferencias raciales se muestran como evidencia desde la NHANES I que señalaba menor mortalidad en hombres de raza blanca para IMC de 24.8 kg/m², en tanto que para los de raza negra era de 27.1 kg/m², correspondiendo estos valores en mujeres anglosajonas a 24.3 kg/m² y 26.8 kg/m² en afroamericanas; pero a través del tiempo el valor de menor mortalidad ha ido convergiendo hasta 25kg/m² según lo reportan diversos organismos en el sector salud. En la tesis de Licenciatura de Luis López (2004) se mostró que en grupos afroamericanos de ascendencia nigeriana la distribución de la muestra de Barbados era simétrica con mediana 25.53, cuantil 15 ($Q_{15} = 21.58$) y cuantil 85 ($Q_{85} = 30.21$). Es decir, que aparentemente los grupos de raza negra bajo condiciones de ejercicio y alimentación adecuados, sí tienen una distribución de IMC similar a otros grupos raciales, que hoy en día tienden a seguir una distribución lognormal o incluso con cola más pesada a la derecha. Recientemente, se han observado diferencias entre anglosajonas e hispanoamericanas, teniendo estas últimas más grasa para un IMC similar, incluso en clases socioeconómicas similares como lo menciona CASAS (2001).

Cada vez se dispone de evidencia más uniforme sobre el impacto de la obesidad en la mayor parte de las enfermedades crónicas, no sólo en relación con las cardiovasculares, sino también con otros problemas frecuentes como la Diabetes mellitus, la Artrosis, los problemas psicológicos e incluso algunos tipos de cáncer, tales como el de colon, el de

mama y el de endometrio. El impacto importante sobre las enfermedades crónicas, el costo sanitario, la calidad de vida y sus crecientes dimensiones ubican a la obesidad como un problema importante de salud pública de la vida moderna.

La OMS (2000, 2004) ha extendido a todas las poblaciones en general los puntos de corte mostrados en la Tabla (1.1), como un indicador nutricional para todos los adultos. Sin embargo, los intervalos de interés podrían estimarse a través del análisis de los cuantiles teóricos de un modelo estadístico para la variación del IMC para una población en particular, que en esta tesis será la población mexicana de adultos. A la luz de esto, Durazo-Arvizu *et al.* (1997) realizaron para el IMC un análisis estadístico de mínima mortalidad, obteniendo los intervalos de confianza a través de una regresión logística donde la probabilidad de mortalidad se relacionó con una combinación lineal de variables que incluyeron al IMC y a la edad. Además supusieron que $\frac{1}{\text{IMC}}$ sigue una distribución normal y la utilizaron para estimar el modelo. Desde 1841, cuando Quetelet propuso que la distribución normal describía bien al IMC, se ha continuado con la búsqueda de un modelo estadístico que describa bien a la distribución del IMC. En 1995, Nevill y Holder propusieron la distribución Normal para $\frac{1}{\text{IMC}}$ y recientemente en López (2004) se propone usar tres modelos estadísticos de la familia de Box-Cox incluyendo al lognormal para describir a 12 poblaciones de raza negra. La tarea en esta tesis será sustentar la elección de un modelo estadístico apropiado para la población mexicana.

1.2 Síndrome Metabólico

El Síndrome Metabólico es el conjunto de ciertas enfermedades de riesgo para la salud, se caracteriza por la aparición de problemas metabólicos comunes de forma simultanea o secuencial en un mismo individuo. De acuerdo con la definición de la International

Diabetes Federation (IDF, 2005), para que una persona tenga síndrome metabólico debe presentar:

- Obesidad central (definido como circunferencia de cintura ≥ 94 cm para hombres caucásicos y ≥ 80 cm para mujeres caucásicas, con valores étnicos específicos para otros grupos, en México se han establecido en el Manual de procedimientos 2002 los valores, circunferencia de cintura ≥ 102 cm para hombres y ≥ 80 cm para mujeres).

Más dos de los siguientes 4 factores:

- Nivel de triglicéridos (TG) elevados: ≥ 150 mg/dL (1,7 mmol/L), o tratamiento específico para esta anormalidad lipídica.
- Colesterol HDL reducido: < 40 mg/dL (1,03 mmol/L) en hombres y < 50 mg/dL (1,29 mmol/L) en mujeres, o tratamiento específico para esta anormalidad lipídica.
- Tensión arterial (TA) elevada: TA sistólica ≥ 130 o TA diastólica ≥ 85 mm Hg, o tratamiento de hipertensión previamente diagnosticada
- Glucosa plasmática en ayunas elevada ≥ 100 mg/dL (5,6 mmol/L), o diabetes tipo 2 previamente diagnosticada. Si la glucosa en ayunas es $> 5,6$ mmol/L o 100 mg/dL, la prueba de tolerancia oral a la glucosa (PTOG) es fuertemente recomendada pero no es necesaria para definir la presencia del síndrome.

1.3 Conceptos básicos de modelación estadística

Considerando que el objetivo principal de esta tesis es la modelación del IMC de mexicanos, en este capítulo realizaremos una breve descripción del marco teórico sobre el que se sustenta la elección de dicho modelo estadístico.

1.3.1 Modelación Estadística

En las ciencias naturales se desea explicar, entender y en ocasiones predecir algunos fenómenos aleatorios repetibles. La labor del estadístico comienza desde la planeación para adquirir los datos. Después debe proponerse un modelo estadístico que describa bien al fenómeno aleatorio de interés y validarlo a la luz de los datos obtenidos.

Sprott (2000, p.7) define a un modelo estadístico paramétrico como un conjunto de modelos probabilísticos $f(x; \theta)$ descritos por una función de probabilidad o de densidad que está indicada por el vector de parámetros θ (que puede ser unidimensional). El vector de parámetros $\theta = (\theta_1, \theta_2)$ es generalmente desconocido. En ocasiones resulta de interés solamente un subconjunto de parámetros θ_1 y a los restantes θ_2 se les denomina de ruido o de estorbo.

Box (1980) y Sprott (2000) definen a la modelación estadística en la ciencia, como el proceso que consiste de iteraciones de los siguientes cuatro pasos:

1. Proponer un modelo estadístico $f(x; \theta)$ adecuado para el fenómeno de interés repetible.
2. Revisar que los conjuntos de datos provenientes de muestras o experimentos distintos, sean homogéneos con respecto al parámetro de interés θ_1 en $\theta = (\theta_1, \theta_2)$ para poderlos combinar y aumentar la evidencia para así estimar mejor a θ_1 , con todos ellos.
3. Estimar los parámetros de interés θ_1 en presencia de los parámetros de estorbo θ_2 .
4. Validar con estas estimaciones de los parámetros el modelo $f(x; \theta)$ a la luz de los datos observados.

En el caso que el modelo elegido no describa bien a los datos, entonces habrá que regresar al paso (1) y repetir el proceso. A continuación se presenta la herramienta necesaria para realizar eficientemente la estimación de parámetros de interés del modelo estadístico propuesto $f(x; \theta)$.

1.3.2 Función de Verosimilitud

Un concepto central en la Estadística es la función de verosimilitud, pues es el método más eficiente en general para estimar los parámetros θ de un modelo paramétrico $f(x; \theta)$, para un conjunto de datos observados.

Generalmente se concibe que las observaciones son realizaciones de variables aleatorias independientes e idénticamente distribuidas $X_1, \dots, X_n$, expresándolas como $x_1, \dots, x_n$, es decir, los valores que ya fueron observados de esas variables aleatorias. Sea $x = (x_1, \dots, x_n)$ una muestra observada de variables aleatorias discretas independientes e idénticamente distribuidas con función de probabilidad $f(x; \theta) = P(X = x; \theta)$, que depende de un número finito de parámetros reales desconocidos $\theta = (\theta_1, \dots, \theta_k) \in \Theta \subset \mathbb{R}^k$. El espacio parametral es Θ , la región de valores posibles que puede tener el vector de parámetros θ .

La función de verosimilitud, fue definida por primera vez por Fisher (1921, p. 24), como una función del vector de parámetros θ que es proporcional a la probabilidad de haber observado la muestra $P(X = x; \theta)$. Entonces la función de verosimilitud de θ , $L(\theta; x)$, se define para una muestra de variables aleatorias discretas $x = (x_1, \dots, x_n)$ como una función real valuada del parámetro con dominio en Θ , tal que es proporcional a la probabilidad de la muestra observada:

$$L(\theta; x) \propto P(X = x; \theta). \quad (1.2)$$

Es importante notar que la función de verosimilitud en (1.2) se define como proporcional y no igual, a la función de probabilidad $P(X = x; \theta)$. Por tanto hace sentido considerar razones de verosimilitud para comparar la plausibilidad de dos valores θ_1 y θ_2 , para ver cuál de ellos hace más probable la muestra observada. Así se tiene la razón de verosimilitudes:

$$\frac{L(\theta_1; x)}{L(\theta_2; x)} = k. \quad (1.3)$$

Cuando $k > 1$ se preferirá θ_1 para los datos pues este valor del parámetro hace k veces más probable a la muestra observada que el valor θ_2 . En general, conviene comparar los valores de θ con aquel valor que maximice la probabilidad de la muestra observada. A este valor $\hat{\theta}$, se le llama el estimador de máxima verosimilitud (EMV),

$$\hat{\theta} = \arg[\max L(\theta; x)]. \quad (1.4)$$

Es decir, $L(\hat{\theta}; x) \geq L(\theta; x)$ para todo valor de θ . Sprott en 1965 propuso usar una versión estandarizada de la verosimilitud, que vale uno en el máximo y a la cual llamó "*verosimilitud relativa*",

$$R(\theta; x) = \frac{L(\theta; x)}{L(\hat{\theta}; x)}, \quad (1.5)$$

donde $0 \leq R(\theta; x) \leq 1$. A través de esta función se pueden comparar todos los valores posibles de θ con el EMV y su gráfica es sencilla de interpretar.

En (1.2) se definió la función de verosimilitud para variables discretas. Ahora para una variable continua X se tiene que $P(X = x; \theta) = 0$ para todo x y θ ; luego no es posible hacer uso de (1.2). Bajo este contexto es común utilizar *la aproximación continua a la verosimilitud* definida por Kalbfleisch (1985, pp. 25 y 26) para variables continuas, independientes e idénticamente distribuidas como $f(x; \theta)$ que es :

$$L(\theta; x) \propto \prod_{i=1}^n f(x_i; \theta). \quad (1.6)$$

Muchos libros de texto definen directamente a la verosimilitud en el caso de variables aleatorias continuas como en (1.6), sin resaltar que esto es una mera aproximación que

en la mayoría de los casos resulta ser muy buena; así lo muestra Kalbfleisch (1985, pp. 26-29) para unos datos con distribución exponencial. Para más detalles del caso continuo, se sugieren las referencias de Sprott (2000, pp. 10 y 11) y Kalbfleish (1985, pp. 25-29).

Usualmente es conveniente trabajar con el logaritmo natural de $L(\theta; x)$, la denominada función de Log-verosimilitud,

$$l = l(\theta; x) = \log[L(\theta; x)].$$

Luego la Ecuación (1.6) en términos de la Log-verosimilitud es

$$l(\theta; x) = \sum_{i=1}^n \log[f(x_i; \theta)].$$

1.3.3 Intervalos de Verosimilitud-Confianza

Los intervalos o regiones de verosimilitud muestran los valores más plausibles del parámetro θ a la luz de la muestra observada x . Kalbfleisch (1985, Vol. 2, p. 18) define un intervalo de verosimilitud (IV) de nivel c para θ , como:

$$IV(c) = \{\theta \mid R(\theta; x) \geq c\}, \text{ donde } 0 \leq c \leq 1.$$

Cuando θ es unidimensional, gráficamente el $IV(c)$ se obtiene al trazar a la altura c una línea paralela al eje horizontal sobre $R(\theta; x)$. Luego los valores de θ en $IV(c)$ tendrán verosimilitud relativa mayor o igual que c , y todo valor de θ fuera de $IV(c)$ tendrá verosimilitud relativa menor y hará menos probable a la muestra observada. Es decir, el $IV(c)$ separa los valores plausibles de θ a un nivel c de los que no lo son.

Es conveniente graficar y analizar la función de verosimilitud relativa completa, pues un intervalo de verosimilitud por sí sólo no es muy informativo. Por ello, en muchos casos se sugiere considerar varios intervalos y marcar en ellos al EMV $\hat{\theta}$, pues esto permite indicar el cambio en la plausibilidad de los valores de θ dentro del intervalo

y la posible asimetría de la función de verosimilitud. Nótese que $\hat{\theta}$ está contenido en todos los intervalos de verosimilitud pues $R(\hat{\theta}; x) = 1$.

Las inferencias en términos de verosimilitud, se resumen en proporcionar para varios intervalos de distintos niveles de verosimilitud, lo siguiente:

- El nivel de verosimilitud.
- El extremo inferior y superior del intervalo, A y B , respectivamente.
- El EMV $\hat{\theta}$.
- Un nivel de confianza asociado al intervalo, si es posible darlo.

Así con varios intervalos se puede reconstruir bien la forma de la función de verosimilitud. Es común que se den intervalos de verosimilitud de niveles $c = 0.036, 0.15$ y 0.25 porque en general suelen estar asociados con niveles del 99, 95 y 90% de confianza, en el caso de que el parámetro sea unidimensional. Pero, es necesario mencionar que en la elección del nivel c , deben considerarse factores relevantes al objetivo de estudio y éstos serán dependientes del contexto. Obsérvese en la Tabla (1.3) los $IV(c)$ con $c = 0.258, 0.146$ y 0.036 tienen una probabilidad de cobertura aproximada del 90%, 95% y 99%, respectivamente. Las Figuras (1.2) y (1.3) muestran estos intervalos de verosimilitud-confianza, marcados sobre la verosimilitud relativa del ejemplo de canicas en la Sección 1.2.6.

Nivel de confianza	Nivel de verosimilitud	Cuantiles $\chi^2_{(1)}$
$(1 - \alpha)$	c	$Q_{1-\alpha}$
0.90	0.258	2.706
0.95	0.146	3.841
0.99	0.036	6.635

Tabla 1.3: Relaciones entre nivel de confianza y de verosimilitud para un parámetro unidimensional y cuantiles correspondientes de una Ji-cuadrada de un grado de libertad

Entonces, un intervalo de verosimilitud-confianza es óptimo en el sentido de que contiene información de la plausibilidad del parámetro, a la luz de la muestra observada y porque plasma en el nivel de confianza la incertidumbre de que este intervalo contenga al verdadero valor de parámetro. En Kalbfleisch (1985, Vol. 2, pp. 113-116) se pueden ver con más detalle la obtención de la Tabla (1.3).

1.3.4 Función de verosimilitud perfil

La función de verosimilitud maximizada o perfil es una herramienta en Estadística que permite estimar por separado el parámetro de interés de los restantes parámetros de estorbo. En la literatura científica tiene lugar por primera vez, en el contexto de modelos econométricos cuando Hood y Koopmans (1953, pp. 156 y 157) la denominaron "verosimilitud concentrada", pero la usaron para simplificar el problema numérico de maximización global de la verosimilitud. Primero se obtiene la perfil de θ , y luego se maximiza ésta para encontrar el EMV global. Después en 1964 Box y Cox la utilizaron, pero son Sprott y Kalbfleisch en 1970 quienes la definen formalmente y le dan su nombre como verosimilitud maximizada.

Luego, la función de verosimilitud maximizada o perfil L_p del parámetro de interés θ_1 , se define como:

$$L_p(\theta_1; x) = \max_{\theta_2 | \theta_1} L(\theta_1, \theta_2; x) = L\left(\theta_1, \widehat{\theta}_2(\theta_1); x\right). \quad (1.7)$$

donde $\widehat{\theta}_2(\theta_1)$ es el valor que maximiza la verosimilitud, para ese valor fijo de θ_1 y se le llama el estimador restringido de θ_2 . Entonces, es claro que (1.7) se obtiene maximizando sobre θ_2 a la función de verosimilitud $L(\theta_1, \theta_2; x)$ fijando a θ_1 .

Nótese que la función de verosimilitud perfil en (1.7) se obtiene al reemplazar en la función de verosimilitud global $L(\theta_1, \theta_2; x)$ el parámetro de estorbo θ_2 por el estimador

restringido $\widehat{\theta}_2(\theta_1)$, donde este último depende de θ_1 . El EMV global de θ_2 coincidirá con el estimador evaluado en el EMV global $\widehat{\theta}_1$, $\widehat{\theta}_2 = \widehat{\theta}_2(\theta_1)$. La verosimilitud perfil relativa de θ_1 , se define como:

$$R_p(\theta_1; x) = \frac{L(\theta_1, \widehat{\theta}_2(\theta_1); x)}{L(\widehat{\theta}; x)},$$

donde $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2)$ es el EMV de $\theta = (\theta_1, \theta_2)$.

Montoya (2008) da un ejemplo para el caso de un parámetro bidimensional θ . Menciona que la verosimilitud global de los parámetros θ_1 y θ_2 , $L(\theta_1, \theta_2; x)$, es una superficie en $\mathbb{R}^3$, cuyo dominio es el plano cartesiano asociado al espacio parametral. La verosimilitud es una función real valuada, que está definida para cada pareja $(\theta_1, \theta_2) \in \Theta$. Así, si se toma una posición en un punto lejano sobre el eje designado a θ_2 el parámetro de estorbo, entonces la silueta o perfil observada de la verosimilitud global $L(\theta_1, \theta_2; x)$ es la función de verosimilitud maximizada de θ_1 . Por ello, la verosimilitud maximizada es conocida también con el nombre de verosimilitud perfil. En la Figura 1.1, tomada de Montoya (2008) se presenta la gráfica de la verosimilitud perfil de θ proyectada sobre el plano vertical al fondo. En primer plano, se muestra la superficie completa de la verosimilitud.

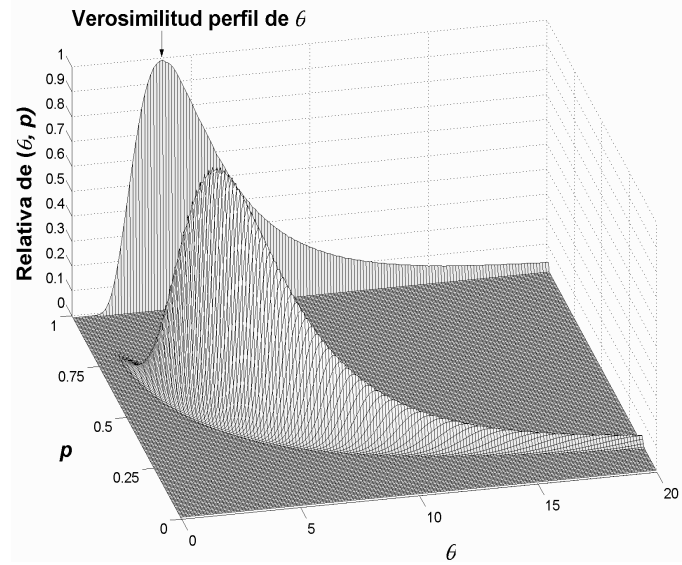


Figura 1.1: Verosimilitud perfil relativa

Algunas características de gran importancia de la verosimilitud perfil son:

1. El EMV perfil de θ_1 es igual a el EMV global $\hat{\theta}_1$.
2. El estadístico de la razón de verosimilitud basado en la verosimilitud perfil,

$$-2 \log \left[L_P(\theta_0; x) - L_P(\hat{\theta}_1; x) \right] = -2 \log R_P(\theta_0; x),$$

equivale al estadístico de razón de verosimilitud global. Para más detalles ver Barndorff-Nielsen y Cox (1994, pp. 88-90).

3. Una región de verosimilitud basada en la verosimilitud perfil $R^k = \{\theta_1 : L_p(\theta_1) \geq k\}$, suele ser también una región de confianza aproximada para el parámetro θ_1 de interés, siempre que la dimensión del parámetro de interés sea menor a 5. Para más detalles, véase Pawitan (2004, Sección 3.5).

1.3.5 Validación del modelo estadístico en consideración.

La etapa final del proceso de modelación estadística, es la validación, esto es confrontar el modelo estadístico propuesto y estimado con la muestra observada. Existen diversas técnicas auxiliares para la validación, dentro de las cuales los métodos gráficos son comúnmente utilizados. Además existen otras pruebas estadísticas analíticas que también son usadas. A continuación, se presentan las herramientas utilizadas en la validación del modelo propuesto para el IMC de mexicanos que se presenta en el Capítulo 2.

Cuantiles

Sea $x = (x_1, \dots, x_n)$ una muestra de observaciones. Los *estadísticos de orden* de la muestra, son los valores $x_{(1)}, \dots, x_{(n)}$ iguales a los x_i , acomodados en orden creciente, Castillo (1988, p. 24):

$$x_{(1)} = \min_{i=1, \dots, n} \{x_i\} \leq x_{(2)} \leq \dots \leq x_{(n)} = \max_{i=1, \dots, n} \{x_i\}.$$

La función de *distribución empírica* denotada por $F_n(x)$ indica la proporción de datos de la muestra que son menores o iguales a x , además es una función real valuada y escalonada, cuyo dominio es $\mathbb{R}$ y su imagen es el $[0, 1]$. Se define así,

$$F_n(x) = \begin{cases} 0, & \text{para } x < x_{(1)}, \\ \frac{i}{n}, & \text{para } x_{(i)} \leq x < x_{(i+1)}, \\ 1, & \text{para } x_{(n)} \leq x. \end{cases}$$

El cuantil Q_α de probabilidad α , es un valor que acumula a la izquierda de él una probabilidad α ,

$$P[X \leq Q_\alpha] = \alpha.$$

Una manera más general de definir a un cuantil, en especial para el caso de funciones de distribución escalonadas es a través de la inversa generalizada de una función de distribución $F(x; \theta)$ a la que se le llama Función de Cuantiles. Sea $F(x; \theta)$ una función de distribución, entonces la Función de Cuantiles se define como:

$$Q(\alpha) = F^{-1}(x; \theta) = \inf \{x : F(x; \theta) \geq \alpha\}, \text{ para } 0 \leq \alpha \leq 1.$$

Así el cuantil Q_α de probabilidad α , es precisamente $Q_\alpha \equiv Q(\alpha)$. Por tanto, nótese que dada la función de distribución empírica $F_n(x)$, cada observación corresponde a varios cuantiles de probabilidad α distintos ya que:

$$x_{(i)} \text{ es el cuantil } Q_\alpha \text{ de } F_n(x) \text{ para } \alpha \in \left(\frac{i-1}{n}, \frac{i}{n} \right].$$

Es decir, Q_α es el valor que deja al $\alpha \times 100\%$ de la muestra por debajo y el $(1 - \alpha) \times 100\%$ restante por encima. Algunos valores de Q_α de especial interés son:

α	Q_α
0.5	mediana
0.25, 0.75	cuartiles
0.1, ..., 0.9	deciles
0.01, ..., 0.99	percentiles

La mediana, el cuantil Q_{50} , se interpreta como el valor central de la muestra, es decir que hay el mismo número de elementos de la muestra inferiores a ella como superiores. Los cuantiles permiten dar una visualización fácil de interpretar de la distribución de la muestra. Ahora para estimar la dispersión de los datos, se suele calcular el recorrido o rango que es la diferencia entre la observación mayor y la menor, aunque este recorrido no da información sobre las regiones donde se concentran más los datos. Se entiende mejor la dispersión de una muestra por los intervalos inter-cuartiles como $[Q_{25}, Q_{75}]$ que

contienen la mitad de los valores centrales de la muestra y los intervalos inter-deciles como $[Q_{10}, Q_{90}]$, el cual contiene al 80% de los valores de la muestra.

Existen diversos métodos que indican si el ajuste de un modelo a un juego de datos es razonable. Dentro de los más comunes encontramos las herramientas gráficas, que permiten diagnosticar de manera visual el ajuste propuesto, como lo son los histogramas y las gráficas cuantil-cuantil que se describen a continuación.

Histogramas

Frecuentemente se utilizan los histogramas de frecuencias relativas, debido a su sencillez y fácil interpretación como una aproximación o estimación de la densidad $f(x; \theta)$ atrás de los datos. Cuando el tamaño de la muestra crece las alturas del histograma convergen a una versión reescalada de la función de densidad para cierto número de clases. Ahora, para la validación del modelo propuesto para el IMC mexicanos, nuestra primera aproximación consiste en superponer en un mismo gráfico un histograma de la muestra con la densidad $f(x; \theta)$ estimada de la densidad considerada. Encima de un cierto intervalo $[a_{h-1}, a_h)$, el histograma asigna un rectángulo o barra cuya superficie iguala la frecuencia empírica de este intervalo. Véanse las Figuras 2.9 a la 2.72. Sin embargo, el inconveniente de un histograma es la arbitrariedad en la elección de las clases o intervalos. Por tanto, se sugiere no limitarse a utilizar sólo los histogramas. Enseguida, se da otra herramienta visual que se usa en el Capítulo 2 como técnica formal para validar el modelo propuesto para el IMC de datos mexicanos.

Gráficas Cuantil-Cuantil

Este tipo de gráficas, que se denotarán aquí para abreviar gráfica-QQ, se utilizarán comúnmente para validar el ajuste de un modelo estadístico propuesto para un juego de datos. Una gráfica-QQ permite comparar dos distribuciones cualesquiera, graficando

los cuantiles de ellas en los ejes cartesianos. Si dos distribuciones son similares, sus cuantiles deben serlo también. En especial si en el eje vertical se acomodan las observaciones ordenadas, que son los cuantiles empíricos correspondientes a la distribución empírica $F_n(x)$ y en el eje horizontal se grafican los cuantiles estimados bajo el modelo propuesto, entonces se obtiene la gráfica-QQ que más se utiliza para la validación del modelo considerado. La gráfica-QQ se obtiene graficando los puntos de coordenadas $(\widehat{Q}(\frac{i}{n}), x_{(i)})$, donde $x_{(i)}$ es el cuantil empírico de probabilidad $\frac{i}{n}$ y $\widehat{Q}(\frac{i}{n}) = F^{-1}(x_{(i)}; \widehat{\theta})$, es el cuantil estimado de la misma probabilidad. Si el ajuste es correcto la función de distribución empírica de las muestras debería estar cerca de la función cuantil estimada y los puntos $(x_{(i)}, Q(\frac{i}{n}))$ estarán cerca de la recta a 45° . Es necesario mencionar, que para designar cuando un valor está demasiado lejos o cerca de la recta y así validar si el ajuste es bueno, se pueden simular muchas muestras bajo el modelo estimado y obtener gráficas-QQ correspondientes para compararlas con la de la muestra observada.

J. S. Marrón de la Universidad de Carolina del Norte de Chapel Hill (2004) comenta que es posible construir bandas de confianza para los cuantiles, usando las gráficas-QQ de muchas muestras simuladas bajo el modelo estimado. Para una muestra observada de tamaño n , se simulan k muestras de tamaño n (usualmente $k = 1000$). De las k simulaciones, para cada cuantil fijo se ordenan los valores simulados de ese cuantil y se extraen de ellos los cuantiles de probabilidad 0.05 y 0.95. Con este par de valores por cuantil observado, se construyen bandas de confianza para la gráfica-QQ. Posteriormente, se gráfica sobre esta banda los puntos correspondientes a la muestra observada. Se dirá que el modelo elegido es adecuado para los datos si la gran mayoría de los datos cae dentro de los límites de la banda.

Los modelos estadísticos propuestos para el IMC de mexicanos, se validarán en el Capítulo 2, en ellos se involucra a la distribución normal, por lo que la gráfica-QQ será muy útil. Los métodos gráficos son de gran utilidad sin embargo, no constituyen

la única respuesta matemática del ajuste. Para cuantificar el alejamiento de la distribución empírica con respecto a una distribución ideal teórica empleamos también otras técnicas de validación que en nuestro caso son las pruebas de normalidad que se describen a continuación, ya que en los modelos estadísticos propuestos se involucra a la distribución normal. A continuación se definen dos indicadores de normalidad.

Pruebas de normalidad

Con el propósito de determinar si existen razones para determinar si una muestra de datos se distribuye como una normal, se compara la asimetría y curtosis muestrales contra las teóricas esperadas bajo distribuciones normales. La asimetría y la curtosis son características de la distribución que permiten identificar como se separan o concentran los valores de la muestra, con respecto a la moda. Las comparaciones que se realizan consisten en comprobar que los coeficientes de asimetría y curtosis sean muy similares a los reportados en la literatura para una distribución normal.

Mood *et al.* (1974, pp. 75 y 76) definen para una variable aleatoria X , con media $E(X) = \mu$ y distribución unimodal, al *coeficiente de asimetría de Fisher* (g_1) y al *coeficiente de curtosis* (g_2), como:

$$g_1 = \frac{E(X - \mu)^3}{[E(X - \mu)^2]^{3/2}}, \quad g_2 = \frac{E(X - \mu)^4}{[E(X - \mu)^2]^2}. \quad (1.8)$$

Tenemos que g_1 posee la característica de indicar cuando la muestra es simétrica respecto de la media. Es decir, existe la misma probabilidad del lado izquierdo y derecho de la media. Diremos que la muestra es simétrica con respecto a la media si $g_1 = 0$, de lo contrario para $g_1 > 0$ la distribución es asimétrica positiva (mayor concentración de datos a la derecha) y si $g_1 < 0$ la distribución será asimétrica negativa (mayor concentración de datos a la izquierda).

Por otro lado, el coeficiente de curtosis g_2 analiza el grado de concentración de los datos alrededor de la media de la distribución, además de poder determinar el ancho de las colas de la distribución. En la literatura científica, se tiene la siguiente clasificación para los posibles valores de g_2 . Si $g_2 = 3$ se dice que la distribución es normal, en el caso de $g_2 < 3$ la distribución de los datos estará mas concentrada alrededor de la media, entonces sus colas son menos gruesas que las de la distribución normal. Por último cuando $g_2 > 3$, hay menor concentración de los datos alrededor de la media, entonces las colas de la distribución son más gruesas que las de la normal.

Fisher (1970, pp. 73-75) recomienda utilizar los estimadores de los coeficientes de asimetría y de curtosis para verificar la hipótesis de normalidad de una muestra $x_1, x_2, \dots, x_n$. Se definen como

$$\hat{g}_1 = \frac{k_3}{(k_2)^{3/2}} \quad (1.9)$$

y

$$\hat{g}_2 = \frac{k_4}{(k_2)^2}, \quad (1.10)$$

donde

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad k_l = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^l, \quad l > 1$$

y las varianzas muestrales correspondientes son:

$$var(\hat{g}_1) = \frac{6n(n-1)}{(n-2)(n-1)(n+3)} \quad (1.11)$$

y

$$var(\hat{g}_2) = \frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}. \quad (1.12)$$

Suponiendo, n muy grande, tenemos que $\hat{g}_1 \sim N(g_1, var(\hat{g}_1))$ y $\hat{g}_2 \sim N(g_2, var(\hat{g}_2))$. Enseguida se muestran los intervalos de confianza para un α dado. Para la asimetría, se define como

$$\hat{g}_1 \pm t_{1-\frac{\alpha}{2}, n-1} * \sqrt{\widehat{var}(\hat{g}_1)} \quad (1.13)$$

y para la curtosis

$$\widehat{g}_2 \pm t_{1-\frac{\alpha}{2}, n-1} * \sqrt{\widehat{var}(\widehat{g}_2)} \quad (1.14)$$

donde $t_{1-\frac{\alpha}{2}, n-1}$ denota al cuantil $1 - \frac{\alpha}{2}$ de una distribución t de Student con $n - 1$ grados de libertad. Con estos intervalos podemos verificar la normalidad de la muestra si los valores 0 y 3 están contenidos en los intervalos de asimetría y curtosis, respectivamente. Más detalles de como fueron obtenidos (1.13) y (1.14), se recomienda ver Hogg y Craig (1995, p.261).

1.3.6 Ejemplo de tiempos de caídas de canicas

Se utilizó un conjunto de datos recolectados por alumnos de la Facultad de Matemáticas de Guanajuato, sobre los conteos de caída de canicas, desde la azotea de un edificio del CIMAT (Centro de Investigación en Matemáticas A.C.) al piso. En la Tabla (1.4) se tienen 100 observaciones (tiempos de caída) tomados por los alumnos con ayuda de un cronómetro.

1.44	1.60	1.32	1.55	1.36	1.25	1.54	1.36	1.31	1.38
1.51	1.60	1.31	1.39	1.44	1.39	1.35	1.36	1.33	1.50
1.68	1.46	1.47	1.50	1.32	1.42	1.36	1.48	1.41	1.44
1.45	1.52	1.36	1.32	1.29	1.42	1.39	1.30	1.33	1.51
1.47	1.38	1.65	1.40	1.27	1.32	1.38	1.34	1.62	1.50
1.49	1.50	1.57	1.39	1.32	1.31	1.45	1.46	1.42	1.39
1.30	1.36	1.46	1.47	1.38	1.31	1.32	1.44	1.49	1.31
1.44	1.30	1.25	1.54	1.27	1.32	1.43	1.34	1.34	1.41
1.55	1.35	1.40	1.26	1.32	1.38	1.44	1.39	1.39	1.48
1.41	1.32	1.31	1.40	1.47	1.23	1.53	1.46	1.39	1.47

Tabla 1.4: Ejemplo: Tiempos de caída de canicas

Por hipótesis cada x_i se distribuye como una normal con parámetros μ y σ^2 , entonces su función de densidad es $f(x_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2 \right]$. Utilizando la aproximación continua a la verosimilitud, se tiene:

$$\begin{aligned} L(\mu, \sigma; x) &\propto \prod_{i=1}^{100} f(x_i) = \prod_{i=1}^{100} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2 \right], \\ &= \frac{1}{\sigma^{100}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^{100} (x_i - \mu)^2 \right], \\ &= \frac{1}{\sigma^{100}} \exp \left[-\frac{1}{2\sigma^2} \left(100(\bar{x} - \mu)^2 + \sum_{i=1}^{100} (x_i - \bar{x})^2 \right) \right]. \end{aligned}$$

Las funciones de logverosimilitud y score son respectivamente:

$$\begin{aligned} l(\mu, \sigma; x) &= -100 \ln(\sigma) - \frac{1}{2\sigma^2} \left[100(\bar{x} - \mu)^2 + \sum_{i=1}^{100} (x_i - \bar{x})^2 \right], \\ S(\mu, \sigma; x) &= \left\{ \frac{100}{\sigma^2} (\bar{x} - \mu), -\frac{100}{\sigma} + \frac{1}{\sigma^3} \left[100(\bar{x} - \mu)^2 + \sum_{i=1}^{100} (x_i - \bar{x})^2 \right] \right\}. \end{aligned}$$

Los EMV de μ y σ utilizando la función score, se encuentran resolviendo el sistema de ecuaciones $S(\mu, \sigma; x) = (0, 0)$, se tiene dos ecuaciones, de la primer ecuación:

$$\frac{100}{\sigma^2} (\bar{x} - \mu) = 0,$$

se despeja a μ y se tiene:

$$\hat{\mu} = \bar{x}.$$

Ahora, de la segunda ecuación:

$$-\frac{100}{\sigma} + \frac{1}{\sigma^3} \left(100(\bar{x} - \mu)^2 + \sum_{i=1}^{100} (x_i - \bar{x})^2 \right) = 0,$$

despejando a σ , se sigue que:

$$\hat{\sigma} = \sqrt{\frac{1}{100} \sum_{i=1}^{100} (x_i - \hat{\mu})^2},$$

Finalmente, al sustituir los valores correspondientes se tiene:

$$\hat{\mu} = 1.4078, \quad \hat{\sigma} = 0.094.$$

Enseguida se muestra la matriz de información de Fisher. Para obtenerla se calculan las segundas derivadas parciales de la logverosimilitud y se evalúa en cada estimador $\hat{\mu}$ y $\hat{\sigma}$.

$$\begin{aligned} I_{\hat{\mu}, \hat{\sigma}} &= - \left(\begin{array}{cc} \frac{\partial l(\mu, \sigma; x)}{\partial \mu^2} & \frac{\partial l(\mu, \sigma; x)}{\partial \sigma \partial \mu} \\ \frac{\partial l(\mu, \sigma; x)}{\partial \mu \partial \sigma} & \frac{\partial l(\mu, \sigma; x)}{\partial \sigma^2} \end{array} \right) \bigg|_{\hat{\mu}=\bar{x}, \sigma^2=\hat{\sigma}^2}, \\ &= - \left(\begin{array}{cc} -\frac{100^2}{\sum_{i=1}^{100} (x_i - \bar{x})^2} & -200 \frac{(\bar{x} - \bar{x})}{\hat{\sigma}^3} \\ -200 \frac{(\bar{x} - \bar{x})}{\hat{\sigma}^3} & -\frac{2*100^2}{\sum_{i=1}^{100} (x_i - \bar{x})^2} \end{array} \right) \bigg|_{\hat{\mu}=\bar{x}, \sigma^2=\hat{\sigma}^2}, \\ &= \frac{100}{\hat{\sigma}^2} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}. \end{aligned}$$

Se sabe que $\hat{\sigma}^2 > 0$. Además, la matriz $I_{\hat{\mu}, \hat{\sigma}}$ es definida positiva, pues sus valores propios $\lambda_1 = \frac{100}{\hat{\sigma}^2} = 11.49$ y $\lambda_2 = \frac{200}{\hat{\sigma}^2} = 22.98$, son positivos. Por lo que, efectivamente los EMV propuestos $\hat{\mu} = \bar{x}$ y $\hat{\sigma} = \sqrt{\frac{1}{100} \sum_{i=1}^{100} (x_i - \mu)^2}$ son máximos para la función de verosimilitud.

Ahora se calcula la función de logverosimilitud perfil de μ y σ , como:

$$\begin{aligned} l_p(\mu) &= l(\mu, \hat{\sigma}_\mu; x), \\ &= -100 \ln(\hat{\sigma}) - \frac{1}{2\hat{\sigma}^2} \left[100(\bar{x} - \mu)^2 + \sum_{i=1}^{100} (x_i - \bar{x})^2 \right], \\ &= -100 \ln(0.09) - \frac{1}{2(0.09)^2} \left[100(1.41 - \mu)^2 + \sum_{i=1}^{100} (x_i - 1.41)^2 \right], \\ &= 473.4 - 56.88 [100(1.41 - \mu)^2 + 0.88]. \end{aligned}$$

Para σ , se tiene:

$$\begin{aligned}
l_p(\sigma) &= l(\hat{\mu}_\sigma, \sigma; x), \\
&= -100 \ln(\sigma) - \frac{1}{2\sigma^2} \left(100(\bar{x} - \hat{\mu})^2 + \sum_{i=1}^{100} (x_i - \bar{x})^2 \right), \\
&= -100 \ln(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{100} (x_i - \bar{x})^2 \\
&= -100 \ln(\sigma) - \frac{100}{2\sigma^2} \hat{\sigma}^2, \\
&= 473.4 - 56.88 [100(1.41 - \mu)^2 + 0.88],
\end{aligned}$$

Las expresiones de la verosimilitud perfil para μ y σ son:

$$\begin{aligned}
R_p(\mu) &= \frac{L(\mu, \hat{\sigma}_\mu; x)}{L(\hat{\mu}, \hat{\sigma}; x)} \\
&= \exp [l_p(\mu) - l_p(\hat{\mu})], \\
&= \exp [-5687.6(1.41 - \mu)^2],
\end{aligned}$$

para σ

$$\begin{aligned}
R_p(\sigma) &= \frac{L(\hat{\mu}_\sigma, \sigma; x)}{L(\hat{\mu}, \hat{\sigma}; x)} \\
&= \exp [l_p(\sigma) - l_p(\hat{\sigma})], \\
&= \exp \left[286.7 - 100 \ln(\sigma) - \frac{0.88}{2\sigma^2} \right].
\end{aligned}$$

Los intervalos de verosimilitud confianza se muestran en la Tabla (1.5) y gráficamente en las Figuras 1.2 y 1.3.

Nivel de confianza	Nivel de verosimilitud	Intervalos de verosimilitud confianza	
$(1 - \alpha)$	c	μ	σ
0.90	0.258	[1.393,1.423]	[0.084,0.106]
0.95	0.146	[1.389,1.426]	[0.082,0.108]
0.99	0.036	[1.383,1.432]	[0.079,0.114]

Tabla 1.5: Intervalos de verosimilitud-confianza para los datos de canicas

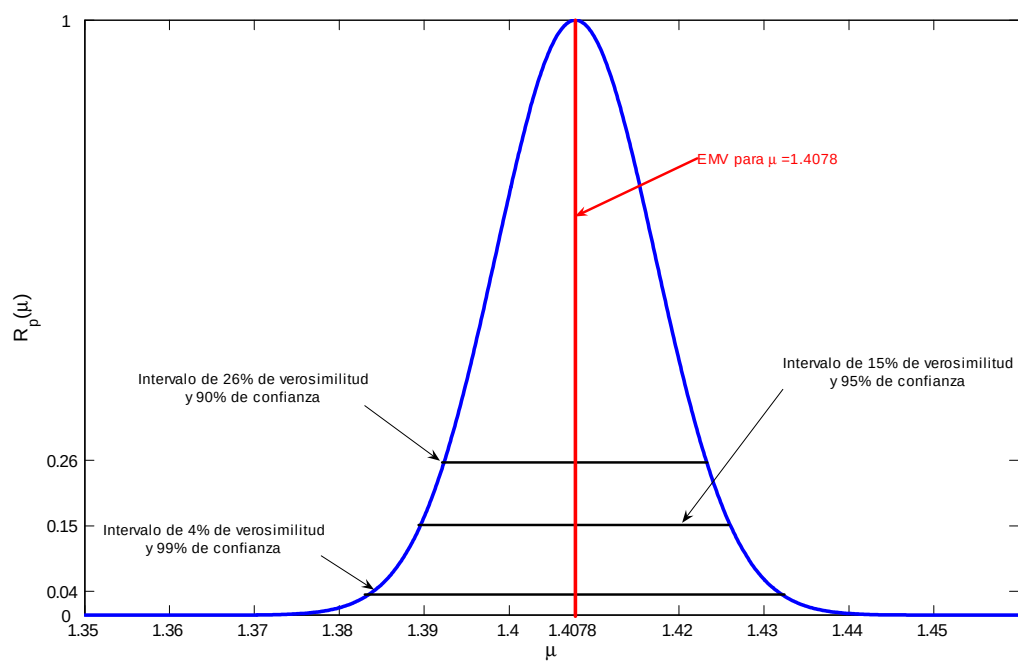


Figura 1.2: Verosimilitud perfil relativa para μ

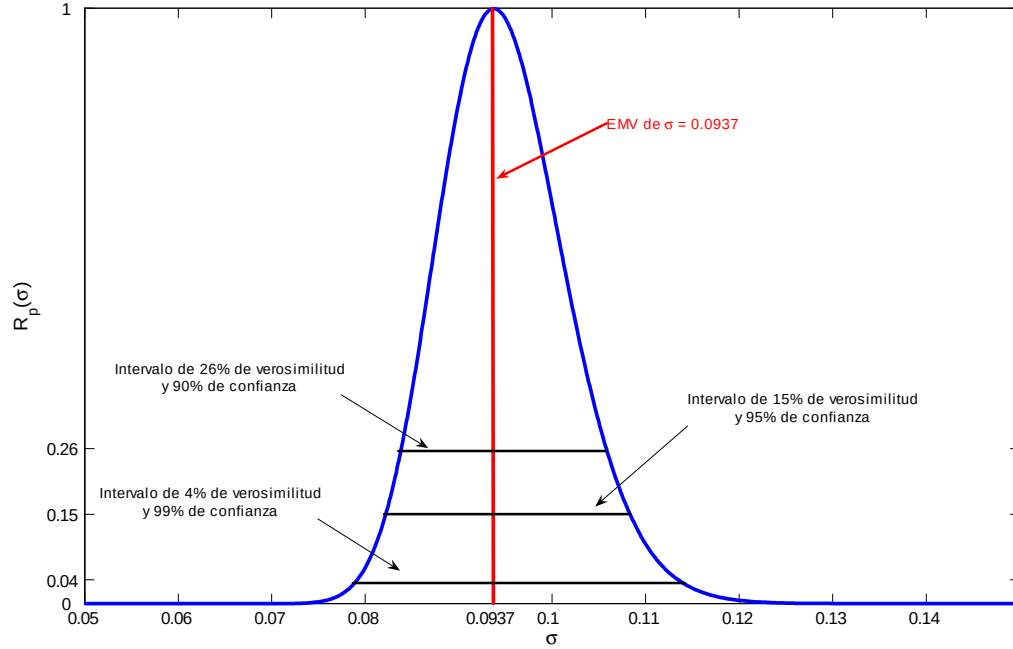


Figura 1.3: Verosimilitud perfil relativa para σ

1.4 Modelos estadísticos para el IMC

En esta sección se propondrán varios modelos estadísticos que describieron adecuadamente al IMC de mexicanos. Esto implica presentar familias de modelos probabilísticos, indicados por el menor número posible de parámetros. En esta sección se argumentan las razones por las que los modelos que se propondrán son razonables. Enseguida se define el modelo lognormal, se comenta su génesis y sus aplicaciones. Bajo el supuesto de que el modelo lognormal suele describir bien cambios proporcionales en el tamaño de un individuo u objeto, es inmediato pensar en que el modelo lognormal es un candidato natural para modelar al IMC. Sin embargo, como se muestra a lo largo del Capítulo 2, no se cumple esto para todas las poblaciones de estudio, por ello es necesaria la

búsqueda de un modelo más general como lo es la familia de modelos de Box-Cox, que incluye como caso particular al lognormal.

1.4.1 Modelo Lognormal

Una variable aleatoria podrá modelarse con la distribución lognormal si puede considerarse como un producto multiplicativo de muchos pequeños factores independientes. Por ejemplo en 1903, Kapteyn formula la llamada "*Teoría del efecto proporcional*". En esta se menciona que el incremento de una variable es función de su valor anterior, multiplicada por una variable aleatoria, esto es:

$$X_t - X_{t-1} = \epsilon_t g(X_{t-1}).$$

Se supone que para distintos tiempos t , las ϵ_t son variables aleatorias independientes y frecuentemente se suponen también con igual distribución. Además que $X_t - X_{t-1}$ es una proporción aleatoria del valor anterior de la variable.

En Ecología y desde que Malthus y Darwin notaron que el crecimiento de plantas y animales es multiplicativo, se ha utilizado el modelo lognormal para describir el crecimiento de poblaciones de individuos y patrones de abundancia de especies en cualquier comunidad. Los ecólogos durante décadas han aplicado herramientas estadísticas, como los modelos de distribución de abundancia, dentro de los cuales la distribución lognormal es aplicada frecuentemente. Se utiliza para analizar las regularidades en diversas entidades biológicas y ecológicas, abstractas o delimitadas en el espacio y tiempo, con el objetivo de detectar similitudes y diferencias entre las entidades. Por ejemplo, Sir Robert May uno de los ecólogos matemáticos más afamados a nivel mundial, demostró entre 1975 y 1981 que la mayor parte de las comunidades inalteradas, suelen ajustarse con la distribución lognormal. Esto se debe a que este tipo de comunidades poseen un número considerable de especies y que sus abundancias

dependen de múltiples factores operando de manera independiente. Además, cuando alguna comunidad ha sido modelada con una distribución lognormal y sucede un cambio después de cierto tiempo a un ajuste, por ejemplo geométrico, dicho cambio es considerado como un indicador indirecto de impactos ambientales de riesgo para la comunidad.

Se dice que una variable aleatoria positiva X de parámetros $\mu, \sigma^2 > 0$ tiene una distribución lognormal, si $Y = \ln(X)$ tiene una distribución normal con media μ y varianza σ^2 , $N(\mu, \sigma^2)$. La función de densidad de $X = \exp(Y)$ es:

$$f(x; \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{[\ln(x) - \mu]^2}{2\sigma^2} \right\} I_{(0,\infty)}(x), \quad (1.15)$$

donde $I_{(0,\infty)}(x)$ es la función indicadora de x en el intervalo $(0, \infty)$.

Usando el Teorema de Cambio de Variable es sencillo encontrar la función de densidad (1.15) apartir de la función de densidad de Y . Además, observemos que la variable X es positiva. En este caso, la distribución lognormal puede ser una representación más realista que la distribución normal para describir características de un individuo como peso, altura o densidad, ya que éstas toman valores positivos alejados del cero y suelen presentar una distribución con cola derecha pesada. Tal es el caso del IMC, como se verá en el Capítulo 2.

Dennis y Patil (1998), mencionan con más detalle el uso de la distribución lognormal en Ecología; aplicaciones geológicas pueden verse en Agterberg (1970) y en las de Koch y Link (1974). Existen aplicaciones en diversas disciplinas como Astronomía, Agricultura, Biología, Estadística de partículas pequeñas, Economía, Sociología, Física y en la industria se considera una gran competidora de la distribución Weibull. Se recomienda al lector ver la referencia (Johnson y Kotz, 1970, pp. 112-136, 1982, V. 5 pp. 134 y 135) donde se discute con más detalle la historia de la génesis y los usos de la distribución lognormal. Como el modelo lognormal, es un caso particular de la familia

de modelos de Box-Cox, se presentará en la siguiente sección el proceso de estimación de los parámetros por máxima verosimilitud.

Por la génesis del modelo lognormal se pensaría que éste sería el mejor candidato para modelar al IMC. Sin embargo, es la familia de Transformaciones de Box-Cox, la que posee la virtud de ser el mejor candidato para modelar a todas las muestras del IMC de mexicanos, además de que contiene a los modelos que se han propuesto con anterioridad, para modelar al IMC; por ejemplo el de Nevill y Holder de (1995) y el de Quetelet.

1.4.2 Los modelos de Box-Cox

En 1964, George P. Box y Sir David R. Cox propusieron la familia de modelos estadísticos de transformaciones de una variable aleatoria X en otra $Y_{(\lambda)}$ definiéndolos así:

$$y_{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \ln x & \text{si } \lambda = 0 \end{cases} \quad (1.16)$$

donde $x > 0$ y $\lambda \in \mathbb{R}$. A la familia de distribuciones de X que depende del parámetro λ se llama "Familia de modelos de Box-Cox".

La transformación $y_{(\lambda)}$ es convexa si $\lambda > 1$ y cóncava si $\lambda < 1$. Carroll y Rupert (1988. pp. 121- 124), mostraron que una transformación de X a $g(X)$ incrementa la asimetría si g es una función convexa y se reduce si g es cóncava. Por tanto, si la distribución tiene cola pesada a la derecha entonces una transformación como en (1.16) con $\lambda < 1$ podría simetrizarla, en cambio si la distribución tiene cola pesada a la izquierda se sugiere una transformación como en (1.16) con $\lambda > 1$ para simetrizarla.

La transformación es monótona creciente para todo $\lambda \in \mathbb{R}$. Dicha propiedad es de gran interés y utilidad, ya que después de la transformación, se preserva el orden de la

escala original en la escala transformada y los cuantiles de X y Y corresponden a la misma observación (pero transformada).

Notemos que la transformación $y_{(\lambda)}$ es continua para todo $\lambda \in \mathbb{R}$, para $\lambda \neq 0$ ya que se tiene que: $\lim_{\lambda \rightarrow 0} y_{(\lambda)} = \ln x$. Aplicando la regla de L'Hôpital tenemos que:

$$\begin{aligned} \lim_{\lambda \rightarrow 0} y_{(\lambda)} &= \lim_{\lambda \rightarrow 0} \frac{x^\lambda - 1}{\lambda}, \\ &= \lim_{\lambda \rightarrow 0} \frac{\exp(\ln x^\lambda) - 1}{\lambda}, \\ &= \lim_{\lambda \rightarrow 0} \frac{\frac{\partial [\exp(\ln x^\lambda) - 1]}{\partial \lambda}}{\frac{\partial \lambda}{\partial \lambda}}, \\ &= \lim_{\lambda \rightarrow 0} [\exp(\lambda \ln x)] \ln x, \\ &= \ln x. \end{aligned}$$

Generalmente, el valor de λ se interpreta como la escala en la que se simetriza la distribución de los datos originales. Otra cualidad de las transformaciones de Box-Cox es que los datos transformados adquieren al mismo tiempo las propiedades de homoscedasticidad y simetría.

Para modelar una variable aleatoria X mediante las transformaciones de Box-Cox, se considera que existen valores para λ tales que $Y_{(\lambda)} \sim N(\mu, \sigma^2)$; después se procede a validar este supuesto con los datos de interés. Por tanto, utilizando el teorema de cambio de variable, la función de densidad de X es:

$$\begin{aligned} f(x; \lambda, \mu, \sigma) &= \left| \frac{\partial y_{(\lambda)}}{\partial x} \right| \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(\frac{y_{(\lambda)} - \mu}{\sigma} \right)^2 \right] I_{(0,\infty)}(x), \quad (1.17) \\ &= \frac{x^{\lambda-1}}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(\frac{x^\lambda - (1 + \lambda\mu)}{\lambda\sigma} \right)^2 \right] I_{(0,\infty)}(x). \end{aligned}$$

Notemos que si $\lambda = 0$ tenemos el modelo Lognormal y para $\lambda = -1$ el modelo de Nevill y Holder. La expresión (1.17) denota un modelo con 3 parámetros. Cabe

mencionar que λ tiene un significado y nivel lógico diferente de μ y σ , los parámetros de localización y escala, ya que determina el modelo.

Como primer paso, dirigiremos nuestra atención a transformaciones $y_{(\lambda)}$ que sean plausibles a la luz de los datos. Se considerará la facilidad de interpretación, la simplicidad y el significado físico o contexto teórico. Se destaca que en el proceso de elección de un miembro de la familia de Box-Cox, la λ no se elige como el EMV de este parámetro, pues como se sabe al tomar una muestra diferente este valor cambia ligeramente. Por ello se usará la verosimilitud perfil de λ para obtener un intervalo de valores razonables sustentados por los datos. Se elegirá un λ con significado especial teórico ($\lambda = 0.5, 0, -0.5, -1$, etc.) que esté contenido en dicho intervalo. El siguiente paso es estimar los parámetros μ y σ del modelo (1.17), condicionados al valor de λ elegido bajo consideraciones físicas o por el contexto teórico.

Estimación de los parámetros de un modelo de Box-Cox

Aquí se presenta como elegir un modelo de Box-Cox apropiado para datos de un fenómeno natural de interés. En 1964, Box y Cox utilizaron la función de verosimilitud perfil en la elección de un estimador de λ , como se describió arriba. Supongamos que existe un λ , tal que para la muestra de variables aleatorias $X_1, \dots, X_n$, i.i.d, se tiene que $Y_{i(\lambda)} \sim N(\mu, \sigma^2)$, para $i = 1, 2, \dots, n$, y $\mu \in \mathbb{R}$ y $\sigma > 0$. La función de verosimilitud perfil apartir de (1.17), a la luz de los datos $x = (x_1, \dots, x_n)'$ es:

$$L(\lambda, \mu, \sigma; x) \propto \prod_{i=1}^n f(x_i; \lambda, \mu, \sigma) = \left(\prod_{i=1}^n x_i \right)^{\lambda-1} \left(\frac{1}{\sigma} \right)^n \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left[\frac{x^\lambda - (1 + \lambda\mu)}{\lambda\sigma} \right]^2 \right\}. \quad (1.18)$$

En este paso se procede a encontrar los valores plausibles de λ a la luz de los datos. Ahora, μ y σ toman el rol de parámetros de estorbo. Entonces se usará la función de verosimilitud perfil de λ , $L_p(\lambda) \equiv L(\lambda, \hat{\mu}_\lambda, \hat{\sigma}_\lambda)$. Usualmente es conveniente trabajar con

el logaritmo natural de $L(\theta; x)$, $l_p(\lambda) = \ln(L_p(\lambda))$ que es la función de log-verosimilitud perfil. De (1.18), la función de *log-verosimilitud perfil* de (λ, μ, σ) es:

$$l(\lambda, \mu, \sigma) = \left[(\lambda - 1) \sum_{i=1}^n \ln(x_i) \right] - n \ln \sigma - \frac{1}{2} \sum_{i=1}^n \left[\frac{x^\lambda - (1 + \lambda\mu)}{\lambda\sigma} \right]^2. \quad (1.19)$$

La función de verosimilitud perfil de λ se obtendrá al maximizar la función (1.19) sobre (μ, σ) para cada valor fijo de λ . Por ello se procede de manera habitual, primero derivando (1.19) con respecto a μ y a σ y así encontrar el máximo asociado a cada valor de λ . Para ello utilizamos la función *Score*, que es el vector de primeras derivadas parciales con respecto a los parámetros de estorbo, en este caso (μ, σ) :

$$S(\mu, \sigma | \lambda) = \left(\frac{\partial l(\lambda, \mu, \sigma)}{\partial \mu}, \frac{\partial l(\lambda, \mu, \sigma)}{\partial \sigma} \right).$$

De aquí que los estimadores de μ y σ , para un valor de λ fijo, sean:

$$\begin{aligned} \hat{\mu}_\lambda &= \frac{1}{n} \sum_{i=1}^n y_{i(\lambda)}, \\ \hat{\sigma}_\lambda &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{i(\lambda)} - \hat{\mu}_\lambda)^2}. \end{aligned} \quad (1.20)$$

Sustituyendo los valores de (1.20) en (1.19), la función de *log-verosimilitud perfil* de λ será:

$$l_p(\lambda) = (\lambda - 1) \sum_{i=1}^n \ln(x_i) - \frac{n}{2} \ln \hat{\sigma}_\lambda^2. \quad (1.21)$$

Para encontrar el EMV perfil de λ , se maximiza a (1.21), los pasos son los siguientes:

1. Se da una rejilla de λ' s en el intervalo de $[-3, 3]$, que es el rango habitual donde se encuentra λ para la mayoría de fenómenos naturales.
2. Para cada valor fijo de λ en la rejilla se calcula la *log-verosimilitud perfil* (1.21).

3. Se encuentra el máximo de la perfil de λ para así hallar el EMV $\hat{\lambda}$ con ayuda del método de Nelder-Mead, que es un algoritmo de búsqueda y optimización numérica.
4. Se da un intervalo del 15% de verosimilitud perfil para λ .

Con la función de verosimilitud perfil relativa de λ , $R_p(\lambda; x)$, se calculan los intervalos de verosimilitud para λ de distintos niveles. Por ejemplo un nivel de 0.15 de verosimilitud ayuda a sugerir modelos de transformaciones que suelen ser razonables a la luz de los datos. Generalmente no se elegirá el EMV perfil de λ , si no que un valor con significado físico o biológico de λ , que esté contenido en este intervalo de verosimilitud. Como se mostró en la Sección 1.2.4 los intervalos de 0.15 de verosimilitud son generalmente además intervalos de confianza de 0.95. Finalmente, después de la elección del λ adecuado, se estiman μ y σ como en (1.20), luego el modelo que describe a la v.a. X es:

$$f(x; \lambda, \hat{\mu}_\lambda, \hat{\sigma}_\lambda).$$

En ocasiones puede ser que existan dos valores λ_1 y λ_2 dentro del intervalo de cierto nivel de verosimilitud que hagan a la muestra muy plausible. Entonces, es necesaria la comparación de ambos valores para elegir alguno de ellos. Para ello, se utiliza la razón de verosimilitud (1.3) para comparar los dos modelos:

$$rv = \frac{L_1(\lambda_1, \hat{\mu}_{\lambda_1}, \hat{\sigma}_{\lambda_1}; x)}{L_2(\lambda_2, \hat{\mu}_{\lambda_2}, \hat{\sigma}_{\lambda_2}; x)},$$

donde $x = (x_1, x_2, \dots, x_n)'$ y L_1, L_2 son los modelos inducidos por λ_1 y λ_2 respectivamente. Como se explicó en la Sección 1.2.4 el valor de rv indicará el número de veces que es más probable un modelo sobre el otro. Por ejemplo, si $rv = 15$ esto indica que el modelo inducido por λ_1 es 15 veces mejor que el de λ_2 . El determinar cuando es mejor un modelo sobre otro depende también de qué modelo haga más sentido al contexto

del problema. Cuando se tienen resultados muy parecidos para diferentes modelos, es entonces indistinguible elegir uno sobre el otro.

En el análisis de plausibilidad para el IMC de mexicanos de la Sección 2.3.1, se muestra con ejemplos la elección del λ adecuado para cada una de las poblaciones (hombres y mujeres por entidad federativa). Es de interés mencionar que un aspecto importante en la modelación es la repetibilidad del fenómeno; es decir, que muchos experimentos repetidos bajo las mismas condiciones deberán concluir a la elección del mismo modelo. Una vez elegido el modelo, se verifica su congruencia, al confrontarlos con los datos reales. Para ello, utilizamos los cuantiles de la familia de Box-Cox; enseguida se muestra este hecho.

Cuantiles de un modelo de Box-Cox

Como se mencionó en la Sección 1.2.5. los cuantiles son una poderosa herramienta de comparación de dos distribuciones; más aún, permiten establecer igualdad o diferencias entre dos distribuciones cualesquiera. Ahora, la variable aleatoria $Y_\lambda \sim N(\mu, \sigma^2)$, donde Y_λ está determinada bajo la transformación de Box-Cox con λ elegida. El cuantil de nivel α para Y_λ lo denotaremos como Q_α^y ; éste puede obtenerse en función de un cuantil normal estándar (z_α), la media μ y la desviación estándar σ , de la siguiente forma:

$$P\left(\frac{Y_\lambda - \mu}{\sigma} \leq z_\alpha\right) = \alpha \quad \Leftrightarrow \quad P(Y_\lambda \leq \sigma z_\alpha + \mu) = \alpha,$$

se sigue:

$$Q_\alpha^y = \sigma z_\alpha + \mu.$$

Los cuantiles en la escala transformada se pueden obtener como

$$Q_\alpha^y = \begin{cases} \frac{(Q_\alpha^x)^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln(Q_\alpha^x), & \lambda = 0 \end{cases}.$$

Es natural pensar, que los cuantiles transformados no ofrecen alguna interpretación para el problema, pues la escala de los datos transformados no es directamente interpretable. Sin embargo, se puede estimar el cuantil de nivel α , Q_α^x en la escala original apartir del cuantil transformado que sí es interpretable:

$$Q_\alpha^x = \begin{cases} (\lambda Q_\alpha^y + 1), & \lambda \neq 0 \\ \exp(Q_\alpha^y), & \lambda = 0 \end{cases}.$$

Los cuantiles determinados de dos distribuciones diferentes se pueden comparar, y valorar así las discrepancias de los resultados. En el siguiente capítulo se presenta una aplicación de la familia de transformaciones de Box-Cox en la modelación del IMC de mexicanos, donde se hará uso de todo el instrumental visto hasta el momento.

Capítulo 2

Datos antropométricos de Mexicanos

2.1 Introducción

La nutrición a lo largo del curso de la vida es una de las principales determinantes de la salud, del desempeño físico, mental y de la productividad. En México la Secretaría de Salud creó el Sistema Nacional de Encuestas de Salud (SNES) el cual ha desarrollado más de 20 encuestas nacionales, cuyo objetivo ha sido obtener información periódica sobre las condiciones de salud y nutrición. Para México de manera reciente existen dos encuestas que aportan información relevante acerca de la prevalencia de enfermedades crónicas. La primera de ellas, la ENSA (1998,1999), y la segunda es la Encuesta Nacional de Salud y Nutrición ENSANUT (2006), muestran las altas cifras de desnutrición, anemia, aumentos sin precedentes en la prevalencia de sobrepeso y obesidad, en niños de edad escolar, adolescentes y adultos. La ENSANUT 2006 unifica el contenido de las encuestas de salud y nutrición, creando una base de datos que permite abordar problemas de desnutrición y de rezago epidemiológico, principales retos de salud pública actualmente. Además amplió su cobertura para lograr, representatividad en

cada una de las entidades federativas de México.

En el capítulo anterior se presentó a las transformaciones de Box-Cox, como una familia de modelos estadísticos que podía describir bien al IMC de mexicanos. El trabajo de este capítulo consiste en encontrar los miembros de esa familia, que describan adecuadamente la distribución del IMC de hombres y mujeres mexicanos, además de hacer un diagnóstico del ajuste de estos modelos a los datos observados. Un modelo estadístico para el IMC es una mera aproximación a la realidad compleja del IMC, pero lo suficientemente útil para ser adoptado como modelo para explicar o predecir. En ese sentido, modelar el IMC de las poblaciones de mexicanos con la familia de Box-Cox ofrece ventajas, como la descripción compacta y resumida de los datos.

El Instituto Nacional de Salud Pública realizó la encuesta ENSANUT 2006, que recabó información relacionada con el estado de salud y nutrición de la población mexicana, con la prevalencia de algunos padecimientos, con la calidad y respuesta de los servicios de salud y con el gasto en salud que realizan los hogares mexicanos. Con fines de realizar una comparación del IMC para adultos entre las poblaciones, aquí se examinarán los individuos mayores de 20 años de ambos sexos de esta encuesta, que es representativa de los estados del país. Se estudiaron 13,229 hombres y 19,798 mujeres. Para hombres y mujeres, el IMC de cada persona, fue calculado con la Ecuación (1.1). Debido a la existencia de un dimorfismo sexual en las mediciones del cuerpo y de los niveles de distribución de grasa en mamíferos, se estudiará a hombres y mujeres por separado.

En la Tabla (B.3) se presentan los tamaños de muestra de ENSANUT 2006 de cada una de las entidades federativas. En esta encuesta, las mediciones antropométricas se tomaron bajo un protocolo estandarizado (entrenamiento personal, uso de la misma tecnología y equipos de medición).

Se incluyeron como datos válidos aquellos en los que se contaba con las variables

peso, altura y cuyos valores de IMC estaban entre 10 y 58 unidades. Se consideraron valores válidos de altura para adultos, los datos comprendidos entre 1.30 y 2 metros. Se excluyeron del análisis los datos que estuvieron fuera de los límites aceptados y las mujeres embarazadas, debido a cambios de peso no relacionados con los factores que se pretenden estudiar. Se hará una comparación entre los resultados obtenidos, considerando los subconjuntos de cada una de las entidades federativas. Los datos permitirán identificar elementos relacionados, con la salud poblacional e interpretar los resultados obtenidos.

En la siguiente sección, se proponen modelos estadísticos de la familia de Box-Cox, para describir la distribución del IMC en las poblaciones de interés. Para facilitar la lectura las figuras y cuadros, que se mencionan a lo largo de este capítulo se concentran en un apéndice al final de la tesis.

2.2 Modelación estadística de datos de ENSANUT 2006

La familia de modelos estadísticos de Box-Cox posee muchas propiedades que la convierten en el candidato idóneo para modelar la distribución de cada una de las poblaciones descritas en la sección anterior. Esta familia tiene como miembros al modelo de Nevill y Holder (1995) así como el modelo lognormal y el normal ($\lambda = -1, 0, 1$), respectivamente. El modelo lognormal describe de manera natural la génesis del IMC y es un miembro de esta familia más general. La familia de Box-Cox es muy rica en formas y abarca distribuciones simétricas y asimétricas con colas pesadas a la derecha e izquierda. Cuando se desea ajustar esta familia a una muestra de datos específica, el procedimiento a seguir es:

1. Graficar la verosimilitud perfil relativa para λ dada en la Sección 1.2.3. Dar un

intervalo de verosimilitud-confianza para λ sustentado por los datos, por ejemplo del 15% de verosimilitud. Se elegirá un λ que tenga significado físico o biológico y que esté contenido en el intervalo mencionado.

2. Encontrar los estimadores de máxima verosimilitud de los parámetros μ_λ y σ_λ condicionando en ese valor de λ elegido para el modelo normal asociado como se mencionó en la Sección 1.3.
3. Validar el modelo con la λ elegida, $\hat{\mu}_\lambda$ y $\hat{\sigma}_\lambda$ a través de los métodos gráficos de diagnóstico descritos en la Sección 1.3.
4. En caso de que el modelo elegido no ajuste bien a los datos, probar con otro valor de λ con significado físico, contenido en el intervalo de 15% de verosimilitud para λ . Si apesar de esto el modelo no ajusta bien, se deberá considerar algún otro modelo distinto a la familia de Box-Cox.

2.2.1 Modelos para el IMC por estado

Hombres

En las Figuras 2.1 a 2.4 se presentan las verosimilitudes perfiles relativas de λ para cada muestra de las entidades federativas de la población de hombres. Obsérvese que todas ellas son simétricas alrededor del máximo. Como se mencionó en la Sección 1.2.4 los intervalos de verosimilitud de λ son además intervalos de confianza aproximados. Nótese que no existe un único valor común de λ plausible, para los 32 estados, aunque hay valores comunes de λ por grupos de estados, ya que las curvas se intersecan mucho. Los valores plausibles de λ estan en un rango de -1.39 a 0.88. Además en las figuras se muestran gráficamente los intervalos de 15% de verosimilitud y 95% de confianza. Observamos que en su mayoría los intervalos están traslapados. Más aún, en algunos casos suelen tener intervalos muy similares o prácticamente iguales como Baja Califor-

nia Sur y Chihuahua o el Distrito Federal y México. Nótese que en 29 de los 32 estados el valor $\lambda = 0$, se encuentra entre los valores de λ plausibles dentro del intervalo de 15% de verosimilitud y 95% de confianza de cada uno de ellos. Recuérdese que $\lambda = 0$ denota al modelo lognormal; sólo los estados de Coahuila, Chiapas y Oaxaca, no contienen al 0 como valor plausible.

A continuación se muestra con ejemplos, la elección de una λ apropiada en la familia de Box-Cox. En la Tabla (B.4) se muestran los intervalos del 15% de verosimilitud y 95% de confianza para λ en cada entidad federativa. Así mismo, se muestra $\hat{\lambda}_p$ el estimador de máxima verosimilitud de λ y la verosimilitud perfil relativa de λ para tres valores diferentes: $\lambda = -0.5, 0, 0.5$. Como se mostró en la Sección 1.2.3, el valor de R_p es de gran interpretación. Por ejemplo, el valor de $R_p(\lambda; x) = 0.99$ del Distrito Federal indica que el modelo asociado a $\hat{\lambda}_p = 0.48$ hace a la muestra casi tan probable como el indicado por el modelo $\lambda = 0$. Sabemos que si bien el EMV $\hat{\lambda}_p$, determina al mejor miembro de la familia de Box-Cox para estas 32 muestras específicas, este valor no hubiese permanecido constante si se hubiesen tomado otras muestras con otras personas. Además un valor con significado biológico como $\lambda = 0$ será preferible sobre $\hat{\lambda}_p$ siempre que sea sustentado por los datos. En el caso del Distrito Federal, y considerando que los valores de $\lambda = -0.5$ y $\lambda = 0.5$ hacen a la muestra poco plausible, no se dudaría en elegir al modelo indicado por $\lambda = 0$, que es prácticamente igual de plausible que tomar a $\hat{\lambda}_p$.

Inferencia para λ en la familia de Box-Cox						
Hombres Nacional						
Muestra	Intervalo de 15% ver. y 95% de confianza.	$\hat{\lambda}_p$	λ_{elegida}	R_p		
				$\lambda = -0.5$	$\lambda = 0$	$\lambda = 0.5$
Baja California	[-0.21,0.70]	0.238	0	0.006	0.596	0.539
Chiapas	[-1.42,-0.45]	-0.932	-1	0.214	7×10^{-4}	2×10^{-8}
Distrito Federal	[0.07,0.90]	0.482	0.5	1×10^{-5}	0.071	0.996

Tabla 2.1: Elección de una λ con significado. Ejemplos para hombres.

Por otro lado, hay estados donde el valor de $R_p(\lambda; x)$ es muy parecido, para dos valores de diferentes de λ . Por ejemplo, la muestra de Baja California donde los modelos indicados por $\lambda = 0.5$ y $\lambda = 0$, hacen a la muestra igualmente probable. Nótese que para $\lambda = 0$, la muestra tiene mayor plausibilidad que para $\lambda = 0.5$. Además $\lambda = 0$ posee mayor significado biológico y como se muestra más adelante el ajuste con el modelo indicado por $\lambda = 0$ es mejor que el indicado por $\lambda = 0.5$. De lo anterior y considerando que el valor de R_p para $\lambda = -0.5$ hace a la muestra poco plausible, se eligió a $\lambda = 0$ para los hombres de Baja California.

Por último, en el caso de hombres de Chiapas, la Figura 2.1 muestra la verosimilitud perfil de λ , donde se observa que los tres valores propuestos para λ hacen a la muestra poco plausible. Sin embargo, el valor de $\lambda = -1$ asociado al modelo de Nevill y Holder tiene $R_p(\lambda = -1; x) = 0.96$. Es decir, que el modelo indicado por $\lambda = -1$, hace a la muestra prácticamente igual de plausible que el indicado por el EMV $\hat{\lambda}_p = -0.93$. Por tanto, se eligió $\lambda = -1$ para los hombres de Chiapas.

Sabemos que los valores de λ con significado relevante en el contexto de estos datos de IMC son $\lambda = 0, -1$, además de valores intermedios como -0.5 y 0.5 . Con excepción de $\lambda = 0.5$, los demás están asociados a datos con distribuciones asimétricas con colas

derechas pesadas. Por ejemplo, para el estado de Coahuila se propone $\lambda = 0.5$ como estimador de λ pues tiene una ligera asimetría a la izquierda. Gráficamente en las Figuras 2.1, 2.2, 2.3 y 2.4; de manera más precisa en la Tabla (B.4) puede observarse que en las muestras de Guanajuato, Oaxaca y Veracruz la plausibilidad para $\lambda = -0.5$ es mayor que bajo $\lambda = 0$. Como se verá más adelante el ajuste con el valor $\lambda = -0.5$, es mucho mejor, por lo que escogió este valor para estos estados.

Mujeres

De manera similar que en las muestras de hombres, para las muestras de mujeres no existe un valor único de λ plausible para las 32 muestras. Sin embargo, el rango en el cual los valores de λ varían es menor comparado con el de los hombres. Los valores plausibles de λ están en un rango de -0.74 a 0.62, que se muestran en las Figuras 2.5 a 2.8 que corresponden a las gráficas de las verosimilitudes perfiles relativas de λ de cada entidad federativa. En la Tabla (B.6) se muestran los parámetros estimados, bajo el mejor modelo elegido de Box-Cox para cada muestra de datos. Se tomaron en cuenta consideraciones de plausibilidad, repetibilidad e interpretación física y biológica.

La siguiente tabla, resume la elección del valor de λ para hombres y mujeres de cada entidad federativa. Se observa que el modelo lognormal inducido por $\lambda = 0$, fue elegido en la mayoría de las ocasiones como el mejor miembro de la familia de Box-Cox. Es decir, el modelo lognormal es adecuado para modelar a la mayoría de las muestras de hombres y mujeres mexicanos.

λ_{elegido}	Hombres	Mujeres	Total
-1	1	0	1
-0.5	3	8	11
0	21	22	43
0.5	7	2	9
Total	32	32	64

Tabla 2.2: Elección de valores de λ para los 32 estados mexicanos

Validación de los modelos propuestos

Aquí, se evalúa el ajuste de los modelos propuestos con los datos observados. Toda la información disponible para las 32 muestras se resume en la Tabla (B.6). Además con estos modelos se obtienen resultados e información relevante para caracterizar a los datos.

Cuando se cuenta con muestras grandes como las que aquí se considerarán, es un reto mayor el proceso de modelación. Se intentará describir adecuadamente a la mayoría de los datos observados como se logro con los modelos propuestos. El tamaño de las poblaciones en estudio, son del orden de cientos para las 64 muestras de hombres y mujeres, se tiene por objetivo verificar si estos datos son lo suficientemente bien aproximados por un miembro de la familia de Box-Cox. En los casos donde se obtienen resultados favorables, se tendrá en consideración que los modelos elegidos son una muy buena aproximación a la realidad.

En las Figuras 2.9 a la 2.72. se muestran los histogramas de frecuencias relativas de los datos en la escala original, junto con la función de densidad de Box-Cox (1.17) ajustada para cada muestra. También se muestran las gráficas-QQ de los datos

transformados según la λ elegida, en la Tabla (B.6). En estas gráficas, las observaciones transformadas ordenadas se comparan con los cuantiles estimados del modelo normal. Alejamientos severos de la recta a 45° indican distanciamientos de la hipótesis de normalidad. En general, para las 64 muestras de hombres y mujeres, los modelos propuestos son muy buenos en la descripción del IMC, pues las gráficas-QQ exhiben en todos los casos la alta plausibilidad de la suposición de normalidad de los datos transformados. Se puede notar que en 62 muestras, la distribución normal parece modelar bien a los datos transformados al menos desde el cuantil empírico $Q_{0.02}$ hasta el $Q_{0.97}$. Las peores situaciones donde las gráficas-QQ no son tan lineales como las anteriores, son:

- Hombres de Aguascalientes, el modelo propuesto describe adecuadamente a los datos desde el cuantil $Q_{0.051}$ hasta el $Q_{0.977}$ (Figura 2.9).
- Hombres de Baja California, donde se modela adecuadamente a los datos desde el cuantil empírico $Q_{0.007}$ hasta el $Q_{0.956}$ (Figura 2.11).

Las discrepancias en los otros cuantiles, son aún razonables bajo los modelos considerados para las 64 muestras, pues estos ajustan bien a la mayoría del umbral de los cuantiles empíricos, y donde no lo hacen, existen algunas explicaciones posibles, que se comentan adelante.

A continuación se verifica que la variable transformada $IMC^{(\lambda)}$ cumpla las propiedades inherentes al modelo normal. Una de las características es simetría respecto de la moda y, la otra, colas ligeras asociadas a este modelo. El coeficiente de simetría muestral y el coeficiente de curtosis usualmente se usan para medir estas características. Cuando el 0 no está contenido en el intervalo de 99% de confianza para la simetría y el 3 no está contenido en el intervalo de 99% de confianza para el coeficiente de curtosis, hay evidencia de que el modelo normal no es tan adecuado para los datos. En las Tablas (B.7)

y (B.8) se muestran los coeficientes de simetría y curtosis por estado para hombres y mujeres, donde se observan algunos detalles que es conveniente explicar.

1. Hombres. En la Tabla (B.7) se observa que el 0, está contenido en cada intervalo del coeficiente de asimetría para cada estado, sin embargo, el 3 no está contenido en el intervalo de 99% de confianza de algunos estados, lo que indica colas más pesadas que las del modelo normal. Las muestras correspondientes son Aguascalientes, Baja California, Campeche, Coahuila, Colima, Chihuahua, Distrito Federal, Sonora y Tasco. Por ejemplo, en Tabasco notamos que el intervalo es $[3.38, 4.62]$ que no contiene al 3, esto se muestra en la Figura 2.59, en esta muestra existe un hombre de 30 años con 1.64 metros y tan sólo 27.8 kilogramos, estas medidas extremas dan como resultado un IMC de 10.273 kg/m^2 . Esto último explica la magnitud de la curtosis al tener cola pesada izquierda. Otro caso es el de Colima donde un hombre de 38 años tiene las medidas de 1.659 metros y 147.5 kilogramos lo que da como resultado un IMC de 53.59 kg/m^2 , explicando la curtosis y la cola pesada a la derecha. De manera similar sucede con los demás estados, donde el 3 no se encuentra en el intervalo de 99% de confianza para la curtosis, pues en cada estado existen individuos con características y medidas tan extremas, que explican las colas pesadas a la derecha o izquierda, según sea el caso. No obstante, en las Figuras 2.9 a 2.72 se muestra que el ajuste normal para cada estado es muy bueno para la gran mayoría de sus cuantiles centrales.
2. Mujeres. En la Tabla (B.8) se observa que a diferencia de los hombres, todos los intervalos de asimetría contienen al 0 y para Chihuahua el intervalo de 99% de confianza para la curtosis no contiene al 3. Lo anterior se justifica ya que existen cinco mujeres con medidas extremas, la primera de las mujeres tiene 25 años con 1.61 metros de altura y 33 kilogramos de peso, por lo que su IMC es 12.73 kg/m^2 , las demás tienen IMC mayor de 50 unidades. Estos cinco integrantes de

la muestra con medidas tan extremas, explican las colas más pesadas que las de la normal.

Considerando lo anterior supondríamos que el ajuste normal de los datos transformados no es adecuado, en los estados mencionados en los puntos 1 y 2 anteriores para hombres y mujeres. En el caso hombres de Aguascalientes, se observa en la Figura 2.9 que la cola derecha es más pesada que la descrita por el modelo normal, así que pudiera ser que un modelo con colas más pesadas describa bien a los datos. Se ajustó así a los datos de hombres de Aguascalientes con el modelo Gumbel para máximos. Se obtienen los estimadores de los parámetros de la distribución Gumbel, el parámetro de escala $a = 29.1075$ y el parámetro de forma $b = 5.2102$, se ajusta el modelo con los estimadores propuestos, la Figura 2.73 muestra la gráfica-QQ, nótese que los valores de IMC extremos de la cola derecha son bien modelados, sin embargo datos centrales y los de la cola izquierda muestran severos alejamientos de la recta con pendiente 1. Contrastando las Figuras 2.9 y 2.73, el modelo Lognormal describe mucho mejor a los datos que el Gumbel que debe descartarse. Además como se muestra en la Figura 2.74 la muestra observada yace dentro de las bandas de confianza de la gráfica-QQ. Por tanto este modelo lognormal si es razonable para los hombres de Aguascalientes.

En las Figuras 2.74 a 2.83 se muestran las gráficas-QQ de los estados mencionados en los puntos 1 y 2. Se simularon 1000 muestras para cada estado considerando los valores estimados de la Tabla (B.6). Se observa en las figuras que en su gran mayoría los puntos graficados se encuentran dentro de la banda envolvente y muy pocos quedan fuera y estos corresponden a una o a ambas colas. Aunado a esto el ajuste con el modelo elegido de la familia de Box-Cox es mejor. Cabe mencionar que como se observó, los datos cuyo IMC es muy grande fueron bien modelados, pudiera ser que las colas pertenecieran a distribuciones distintas. De ser así, se estaría observando una mezcla de poblaciones, de cualquier manera se ha mostrado que los miembros elegidos de la familia de Box-Cox que se muestran en la Tabla (B.6), describen bien

a las 64 muestras de hombres y mujeres mexicanos, pues describen bien a los datos mayoría de sus cuantiles. Apartir de este hecho, se considerará a estos modelos como la aproximación suficientemente buena que se estaba buscando.

2.2.2 Modelos para el IMC a nivel nacional

Se observa que los datos de las 32 entidades federativas, se ajustan razonablemente bien con cada modelo propuesto y a la mayoría de ellos los describió bien un modelo lognormal. Con el objetivo de encontrar relación del IMC con alguna otra variable de tipo físico, geográfico o social, se realizó un análisis más detallado para cada muestra, pero no se encontró relación alguna que permitiera tener grupos de estados que compartieran características y más aún que sus parámetros fueran parecidos. Es de interés modelar también las poblaciones de hombres y mujeres conjuntamente a nivel nacional. Aquí se propone un modelo que integre los resultados obtenidos para cada entidad federativa, para así modelar el IMC de manera global en todo el país para hombres y por separado para mujeres.

Modelo de mezcla de distribuciones

En la Tabla (B.6) se encuentran los parámetros con los cuáles fue posible modelar a cada entidad federativa de México. Para la modelación de hombres y mujeres a nivel nacional se propone un modelo general que es una mezcla de los modelos propuestos para las entidades federativas. Así el modelo general estimado que se propone para hombres (y uno similar para mujeres), es:

$$g(X) = \sum_{i=1}^{32} p_i f(x_i; \hat{\mu}_i, \hat{\sigma}_i), \quad (2.1)$$

donde p_i es el porcentaje de la población nacional de hombres que pertenece al estado i –ésimo. La información proporcionada por el INEGI en el censo 2000, permite

asignar el valor de cada p_i , valores que se concentran en la Tabla (B.2). Nótese que $f(x_i; \hat{\mu}_i, \hat{\sigma}_i^2)$ es la función de densidad estimada del i –ésimo estado para hombres con los parámetros descritos en la Tabla (B.6). De igual manera se propone un modelo similar de mezcla pero para las mujeres.

Modelo Lognormal

Nótese que el modelo (2.1), tiene 64 parámetros desconocidos, suponiendo que los λ 's ya son fijas y elegidas. En la Tabla (2.2), es notable que el valor de $\lambda = 0$ se elige como el modelo adecuado para la mayoría de las muestras. Bajo este contexto y considerando que un modelo exacto no existe, puesto que todos constituyen simplificaciones de la realidad, serán preferibles modelos con menos variables que expliquen bien a los datos, ya que son más sencillos y estables. La Figura 2.86 muestra la verosimilitud relativa para hombres y mujeres, los EMV para λ son: 0.0399 y -0.1211, para hombres y mujeres, respectivamente. Nótese que para mujeres el cero no está dentro del intervalo de 15% de verosimilitud y 95% de confianza pero el λ es muy cercano a cero; para hombres es razonable proponer al modelo Lognormal. Como se muestra más adelante, el ajuste con el modelo Lognormal para ambos sexos resulta modelar razonablemente bien a la mayoría de los datos.

Comparación y selección de modelos

En las Figuras 2.87 y 2.88, se presenta el histograma de los datos de mujeres y hombres, respectivamente y superpuestas las curvas de los ajustes realizados con la densidad (2.1) y la densidad Lognormal. De las figuras, se puede pensar que el modelo Lognormal es un buen candidato para modelar las poblaciones en estudio, ya que es mucho más simple que el de la mezcla de densidades y tiene sólo dos parámetros para cada población, mostrados en el siguiente cuadro.

	N	λ elegida	EMV de los parámetros	
			$\hat{\mu}_\lambda$	$\hat{\sigma}_\lambda^2$
Hombres	13,229	0	3.2911	0.0275
Mujeres	19,798	0	3.3337	0.0363

Tabla 2.3: Modelo Lognormal global para hombres y mujeres mexicanos.

La razón de verosimilitud de estos dos modelos evaluados en los EMV, para cada muestra es una herramienta útil para compararlos. Sin embargo, como un modelo tiene muchos más parámetros que el otro, se utiliza el criterio de Akaike mostrado en Pawitan, (2004), este criterio permite comparar modelos de distinto número de parámetros, mediante el siguiente cálculo:

$$AIC = -2 \left[l \left(\hat{\theta}; x \right) - n^\circ \text{ de parámetros} \right].$$

Así, se calculó el Criterio de Akaike para hombres y mujeres. El criterio de selección será escoger modelos con valores más bajos de AIC. Considerando los valores de la Tabla (B.6) tenemos como resultado los siguientes resultados:

	Criterio de Akaike	
	<i>Mezcla</i>	<i>Lognormal</i>
Hombres	7.7181e+004	7.7084e+004
Mujeres	1.2265e+005	1.2253e+005

Tabla 2.4: Criterio de Akaike en la elección de un modelo más simple.

Nótese de la tabla anterior, el valor del criterio de Akaike en ambas poblaciones es menor para el modelo Lognormal. Más aún, las Figuras 2.89 y 2.90 se gráfcaron las probabilidades estimadas con los modelos propuestos. Notamos que a simple vista los datos forman una recta a 45 grados, lo que indica que ambos modelos describen igualmente bien a las poblaciones. Apartir de estos resultados es razonable proponer

el modelo Lognormal, que es mucho más simple para modelar a las poblaciones de hombres y mujeres mexicanos.

En las Figuras 2.91 y 2.92 se muestra el ajuste y la gráfica-QQ para cada población, se observa que el ajuste con el modelo Lognormal es sumamente bueno, además en las gráficas-QQ se observa que este modelo describe bien a la mayoría de los datos. Los cuantiles empíricos discrepantes para las poblaciones de hombres y mujeres del modelo estimado son señalados las Figuras 2.91 y 2.92. Nótese que para mujeres es mucho mejor el ajuste, se modela bien al menos desde el cuantil empírico $Q_{0.00091}$ hasta el $Q_{0.9979}$ mientras que para hombres desde $Q_{0.00106}$ hasta el $Q_{0.9979}$, por lo tanto utilizando el modelo Lognormal se garantiza que al menos el 99% de ambas poblaciones estará bien modelado. Los datos que no se modelan bien son de personas con obesidad mórbida o con marcada desnutrición. Es razonable pensar que estos datos provienen de poblaciones distintas al resto que describe bien el modelo elegido.

El IMC transformado deberá presentar características de la distribución normal, como la simetría alrededor de la moda y colas ligeras. En la siguiente tabla, se muestran los coeficientes de simetría y curtosis para las poblaciones de hombres y mujeres mexicanos, que son una herramienta para determinar simetría y colas ligeras de la distribución.

Coeficientes de simetría y curtosis de los datos transformados				
Modelo Lognormal				
Muestra Nacional	Asimetría	Intervalo de 99% de confianza para la asimetría	Curtosis	Intervalo de 99% de confianza para la curtosis
Hombres	-0.0237	[-0.065, 0.018]	3.5026	[3.419, 3.586]
Mujeres	0.0723	[0.038, 0.106]	3.122	[3.054, 3.190]

Tabla 2.5: Coeficientes de asimetría y curtosis, modelo Lognormal a nivel nacional

En la tabla anterior, se observa que el intervalo de curtosis para la población de hombres no contiene al 3, pero como se menciono antes, en la población existen casos muy extremos como un hombre con IMC de 10.273 kg/m^2 y 3 casos más donde el valor de IMC es menor a 13 kg/m^2 . Por otro lado existen también casos extremos con IMC superior a 50 kg/m^2 , todos estos individuos que presentan medidas tan extremas, explican que las colas sean más pesadas que las del modelo normal.

En el caso de hombres, de la gráfica-QQ mostrada en la Figura 2.91, parecería razonable pensar que las personas con desnutrición o con obesidad mórbida, pudieran seguir una distribución diferente al contar con características físicas muy distintas del resto de la población. De hecho, se nota una ligera separación de estos grupos en la gráfica. De manera similar, para la población de mujeres se observa en la Tabla (2.5) que el 0 y el 3 no se encuentran dentro de los intervalos de asimetría y curtosis, respectivamente también debido a pocos casos extremos que podrían pertenecer a poblaciones distintas. Como se observa en la Figura 2.92 la pesadez de la cola derecha es mayor, además de considerar que también existen casos muy extremos, más aún, con problemas severos de obesidad en cualquiera de su clasificaciones, teniendo 42 casos con IMC superior a 50 kg/m^2 .

Las Figuras 2.91 y 2.92 muestran los ajustes sobre el histograma de cada población, estos indican que es razonable modelar a las dos poblaciones con el modelo Lognormal, además en las Figuras 2.93 y 2.94 se han graficado las bandas que propone J.S. Marron (2004) Nótese que sólo algunos datos caen fuera del envolvente que ha considerado el cuantil $Q_{0.05}$ y $Q_{0.95}$ de 1000 simulaciones; se observa así que se está modelando bien a la gran mayoría de los datos.

2.2.3 Análisis de resultados

En la Tabla (2.2), se observó que para la mayoría de los estados el modelo Lognormal fue el adecuado, modelando bien a la mayoría de los datos, además el modelo Lognormal se elige para modelar conjuntamente a hombres y mujeres mexicanos. Los criterios de la Norma oficial mexicana para la clasificación de la obesidad, se han basado en los en los valores del IMC de 25 y 27, mientras que la OMS se basa en los cuantiles $Q_{0.5}$ y $Q_{0.85}$ de las poblaciones caucásicas. Para todas muestras de la población mexicana, estos valores corresponden a cuantiles mucho menores que $Q_{0.5}$. Lo que es más notable, el cuantil $Q_{0.85}$ para la población mexicana, está por encima del valor 30 que indica Obesidad de tipo II.

La Tabla (2.6) muestra el porcentaje considerando la clasificación de la Norma oficial mexicana, bajo el modelo Lognormal se estima que aproximadamente el 18.41% de la población adulta de hombres tiene sobrepeso, mientras que en mujeres el 15% de la población presenta sobrepeso. Sin embargo, las mujeres tienen aproximadamente 57.64% de su población con algún grado de obesidad, mientras que los hombres tienen 49.26% de su población en esta situación. Es decir, para mujeres, aproximadamente 72% de la población esta por encima de un IMC de 25 unidades. El porcentaje es ligeramente superior al de los hombres, lo que indica que la mayor parte de la población de mujeres, se encuentra en riesgo de adquirir algún tipo de enfermedad cardiovascular pues presenta sobrepeso y obesidad.

Población	Infrapeso	Normal	Sobrepeso	Obesidad I	Obesidad II	Obesidad III	Mórbida
Hombres	0.94	31.39	18.41	24.92	18.97	4.38	0.99
Mujeres	0.87	26.42	15.07	21.89	23.65	8.64	3.46

Tabla 2.6: Porcentajes estimados de la clasificación nutricional para adultos mexicanos

Según los criterios propuestos en la Norma Oficial Mexicana, es recomendable para las personas tener un valor de IMC menor a 27, pues después de este valor las personas entran en riesgo de adquirir enfermedades que afectan gravemente su estado de salud, esto se mencionó con más detalle en la Sección 1.1. A nivel nacional las mujeres tienen más problemas de sobrepeso y obesidad que los hombres. En las Tablas (B.9) y (B.10) se muestran los cuantiles estimados para mujeres y hombres de los 32 estados. En las Figuras 2.84 y 2.85 se muestran los cuantiles empíricos de cada estado donde se muestra la proporción de personas adultas con problemas severos nutricionales. Es notable que las mujeres de Sonora y Nuevo León bajo el modelo propuesto se encuentran en mayor riesgo respecto de los demás estados pues aproximadamente el 19% de su población se ubica en obesidad III y obesidad mórbida que como se mencionó en la Sección 1.1, es un problema nutricional muy severo. Por otro lado, las mujeres de Oaxaca presentan mayor prevalencia de desnutrición, con 1.6% de su población. Similarmente sucede en los hombres de Tabasco que tienen aproximadamente 2% de su población con desnutrición.

2.3 Porcentaje de grasa corporal y su relación con el IMC

El objetivo de esta sección, es analizar la relación del IMC y el porcentaje de grasa corporal (%GC) para dos muestras de adultos hombres y mujeres, que acudieron a revisión médica al Hospital Aranda de la Parra en León, Guanajuato. Se realizó un estudio descriptivo y observacional en un grupo de individuos mayores de 18 años de ambos sexos, pacientes de la Dra. Yolanda Méndez Romero, que acudieron a consulta para evaluación nutricional, como parte de una revisión médica especializada del año 1998 a 2004. En los expedientes se recabó información sobre la edad, antecedentes

familiares hereditarios, actividad física, grado de escolaridad, dieta habitual, actividad física y se midió explícitamente el porcentaje de grasa corporal a través de pliegues corporales. Se consideraron estos datos para valorar la relación entre el IMC y el porcentaje de grasa corporal, se estudiaron 715 hombres y 338 mujeres. Esta muestra tiene como característica particular, que los individuos son en su mayoría profesionistas a los cuales el servicio médico, es otorgado por la empresa o institución donde laboran.

Como se comentó al principio del Capítulo 1, existen diversas técnicas para determinar el estado nutricional de una persona, dentro de las cuales el %GC, es el método más fiable para identificar el grado de obesidad de una persona. Sin embargo, es el más difícil y complicado de medir. Existen diversos métodos que estiman el %GC, dentro de los cuales se considera a la densitometría, como aquella que se aproxima más al valor real. A pesar de su aparente sencillez, la densitometría por inmersión implica disponer de unas instalaciones voluminosas y caras. Debido a esto, varios autores han intentado obtener fórmulas para conocer la densidad orgánica de una persona, apartir de pliegues adiposo cutáneos, siendo de fácil determinación por personal previamente entrenado, usando para ello lipocalibradores o plicómetros (pinzas metálicas que sirven para medir el grueso de pliegues en varias partes del cuerpo). Las fórmulas más utilizadas para estimar la densidad corporal, son las de Durnin-Womersley (1967) y las de Jackson-Pollock (1958). Para la estimación del %GC, usualmente se utiliza la Ecuación (2.2) propuesta por SIRI (1961), pues en diversos estudios se ha mostrado que los valores estimados del %GC con la ecuación de SIRI, se aproximan a los valores obtenidos mediante densitometría.

En la muestra de León, la médico internista Yolanda Méndez tomó mediciones de pliegues cutáneos con un plicómetro y utilizó como guía para determinar el porcentaje de grasa corporal de un paciente, tablas basadas en la ecuación de Durnin-Womersley (1967). Dicha ecuación tiene como base la medición de los pliegues bicipital, tricipital,

supraíliaco y subescapular. La expresión matemática de la ecuación de SIRI (1961) es:

$$\%GC = \left(\frac{4.95}{DC} - 4.5 \right) * 100, \quad (2.2)$$

donde la Densidad Corporal (DC) se obtuvo mediante la ecuación propuesta por Durnin y Womersley (1967):

$$DC = C - M * \log(S), \quad (2.3)$$

donde C y M son números que se eligen de una tabla mostrada más adelante y S es la suma de los pliegues bicipital, tricipital, subescapular y supraíliaco. El siguiente cuadro muestra los coeficientes C y M para la suma de los cuatro pliegues recogidos en las tablas de Durnin y Womersley.

Hombres							
Suma de los cuatro pliegues	Edad→	17-20	20-29	30-39	40-49	50+	17-72
	C	1.1620	1.1631	1.1422	1.1620	1.1715	1.1765
	M	0.0630	0.0632	0.0544	0.0700	0.0779	0.0744
Mujeres							
Suma de los cuatro pliegues	Edad→	16-19	20-29	30-39	40-49	50+	16-68
	C	1.1549	1.1599	1.1423	1.1333	1.1339	1.1567
	M	0.0678	0.0678	0.0632	0.0612	0.0645	0.0717

Tabla 2.7: Coeficientes C y M de Durnin y Womersley (1967)

Los coeficientes de C y M se obtienen para cada sexo (densidad corporal global) y de forma específica para cada grupo de edad y sexo (densidad corporal específica). Según sea la manera de calcular la densidad corporal DC y sustituirla en la ecuación general de SIRI (2.2), será el nombre que adquiera dicha ecuación. Al utilizar la densidad global obtenemos la ecuación de Siri-global (SIRIG) y al utilizar la densidad específica la ecuación de Siri-específica (SIRIE).

SIRIE (DC depende de pliegues, sexo y edad) se calcula utilizando la Ecuación (2.2) y DC definida en (2.3). Para SIRIG (DC depende de pliegues y sexo) se calcula utilizando la ecuación (2.2) y DC apartir de las ecuaciones de Durnin-Womersley (1967).

$$DC = \begin{cases} 1.17 - 0.07 * \log S, & \text{hombres entre 10 y 72 años} \\ 1.16 - 0.07 * \log S, & \text{mujeres entre 16 y 78 años} \end{cases}.$$

Nótese que las ecuaciones propuestas por diversos autores tienen por objetivo el medir un doble pliegue de grasa celular subcutánea excluyendo al tejido muscular.

Para la muestra de León, el %GC se determinó según la ecuación elaborada por Durnin y Womersley, y se usó la ecuación de SIRIE. Como se mencionó al principio del capítulo se considerarán hombres y mujeres por separado. A continuación se valora la relación entre el IMC y el porcentaje de grasa corporal.

En la Figura 2.95 se muestra el %GC contra el IMC para hombres y mujeres. Se observa que en general, el %GC de las mujeres es mayor que el de los hombres. Nótese, que la forma curvada como están dispersos los datos, sugiere la transformación logarítmica del IMC. En las Figuras 2.96 y 2.97 se transforma el IMC de hombres y mujeres; se observa que existe una mejor relación lineal del $\ln(\text{IMC})$ con el %GC. Sin embargo, en el caso de los hombres, la relación no es tan buena después de la transformación pues el valor $R^2=0.3763$ es menor respecto del obtenido para mujeres cuyo valor es $R^2=0.611$, como se aprecia en las Figuras 2.96 y 2.97. Los hombres se encuentran más dispersos alrededor de la recta de regresión que las mujeres. Además las relaciones para hombres y mujeres no son tan fuertes, como la deducida en otras investigaciones. En Vázquez (1999) se realiza un estudio, donde se muestran índices de correlación para el IMC y el %GC de 0.7 a 0.9. Nótese por ejemplo que para dos valores muy diferentes de %GC se tiene el mismo valor de IMC, en las Figuras 2.96 y 2.97 son indicados 4 casos, dos para hombres *A*, *B* y dos para mujeres *C* y *D*, con estas

características. Esto indica que no se puede predecir de manera adecuada al %GC, con sólo el IMC, sobretodo para hombres.

Nótese en la Figura 2.98 que los hombres que presentan algún tipo de obesidad y tienen más del 16 % de grasa corporal tienen un nivel de colesterol superior al recomendado (200 mg/dL), y en la Figura 2.99 para las mujeres con algún tipo de obesidad y 25% de grasa corporal presentan niveles de colesterol superiores al recomendado. En las Figuras 2.100 y 2.101 se indican hombres y mujeres que tienen obesidad androide que se definió en la Sección 1.1.1, Nótese que en su mayoría los hombres que presentan IMC superior a 27 unidades también tienen obesidad androide y un %GC superior a 22; de manera similar para mujeres que tienen más de 25% de grasa corporal y tienen algún tipo de obesidad también presentan obesidad androide. Finalmente en las Figuras 2.102 y 2.103 se observa que la obesidad está relacionada con diversos problemas de salud, en particular aquellos que son considerados dentro del conjunto de enfermedades que definen al Síndrome Metabólico

2.4 Conclusiones

La familia de modelos de transformaciones de Box-Cox, surge de manera natural al incluir los modelos de Nevill y Holder y el Lognormal, que fueron propuestos con anterioridad. Uno de los resultados de esta tesis, es que con sólo 4 miembros de la familia de transformaciones de Box-Cox $\lambda = -1, -0.5, 0$ y 0.5 , se logró modelar a las 64 muestras de hombres y mujeres de cada estado. En la mayoría de los estados, el modelo lognormal fue adecuado. Sólo la muestra de hombres de Chiapas, fue modelada mediante la transformación propuesta por Nevill y Holder asociado al valor $\lambda = -1$. Lo anterior deriva en proponer a la familia de modelos de Box-Cox para modelar a cada estado pues como se mostro en secciones anteriores estos modelos ajustan bien a

los datos observados, además con estos modelos se podrían establecer normas o tomar decisiones en el ámbito de salud a nivel estatal.

El resultado más importante de este capítulo, es que a nivel nacional tanto para hombres como para mujeres el modelo Lognormal fue razonable. Se proponen dos modelos uno de mezcla de los modelos de Box-Cox para los estados y uno mucho más sencillo que es el modelo Lognormal. Se elige el modelo Lognormal para modelar conjuntamente a hombres y mujeres, ya que como se mostró en las Figuras 2.91 y 2.92 describió bien a la mayoría de los datos. En resumen, se propone modelar a las poblaciones de hombres y mujeres mexicanos con los dos modelos lognormales mencionados. Se observó que la gráfica de densidad estimada para las mujeres tiene cola más pesada a la derecha que la de los hombres, lo que indica que hay más mujeres que hombres con problemas de sobrepeso y obesidad.

El IMC se ha utilizado como indicador del estado nutricional de un individuo, prediciendo al %GC, ya que el IMC es fácil de obtener y tiene bajos costos, con los datos de León se sustentó que el IMC se relaciona bien con el %GC en el caso de las mujeres, sin embargo para hombres esta relación es muy débil. Además por ejemplo, para dos casos con diferente %GC se tiene el mismo IMC, lo que indica que el IMC por sí solo no puede describir bien al %GC. Se sugiere utilizar otras variables como edad y pliegues cutáneos para poder estimar mejor al %GC de hombres.

Capítulo 3

Conclusiones

En esta tesis se modeló estadísticamente de manera exitosa el IMC para mexicanos adultos. Se proponen aquí dos modelos lognormales para describir el IMC de hombres y de mujeres en México. La ventaja de contar con un modelo probabilístico adecuado radica en que se puede calcular de manera precisa la probabilidad de que el IMC caiga en algún intervalo de interés y que además se puede cuantificar la incertidumbre en dichas afirmaciones. Por otra parte, a partir de estos modelos propuestos también se pueden estimar acertadamente los porcentajes poblacionales de mexicanos que presentan problemas de sobrepeso y de los distintos grados de obesidad. Al comparar el problema de obesidad por género, se resalta que es mayor éste en mujeres que en hombres. Posiblemente esto se deba al hecho de que las mujeres sufren cambios fisiológicos fuertes durante los embarazos y que de manera natural sufren un aumento de peso en estos periodos que es difícil eliminar tras el nacimiento de los hijos. A nivel nacional, hay un 15% de mujeres con problemas de sobrepeso y un 57% de obesas; para hombres, el 18% tiene problemas de sobrepeso y el 49% de obesidad.

A nivel estatal también se encontraron modelos adecuados para cada estado, dentro de la familia de modelos de Box y Cox. Se encontraron también discrepancias importantes en la distribución del IMC en los diversos estados mexicanos. Resultan

relevantes los hallazgos sobre los estados del país que presentan porcentajes más altos de obesidad. En el caso de mujeres, los estados con mayores problemas de obesidad y donde al menos había un 65% de personas con sobrepeso y obesas fueron al norte, Sonora, Nuevo León, Coahuila, y al sur Campeche, Yucatán. En el caso de hombres, los estados con mayores problemas fueron Campeche, Yucatán, Sonora, Colima y Tamaulipas. En estos estados, al menos el 60% de la población presentó problemas de sobrepeso y obesidad. También es importante notar que en cuanto a desnutrición, los porcentajes en todos los estados son menores al 3% y que tan sólo los estados de Jalisco, Sonora y Zacatecas, presentaron problemas en esta dirección, se observa que el problema es ligeramente mayor en el caso de mujeres que en el de hombres adultos mexicanos. La proporción estimada de mujeres con desnutrición es del 1.4% y en el caso de hombres, del 1.1%. Al parecer, un problema que requiere inmediata atención es la alta proporción de mexicanos con problemas de sobrepeso y obesidad en comparación a situaciones anteriores, años atrás de desnutrición.

Además se observó que en el grupo de pacientes de León que se analizó, el IMC muestra un mayor poder de predicción del porcentaje de grasa corporal en mujeres, y que por tanto allí puede ser un buen indicador de problemas de obesidad por sí solo. En contraste, para hombres, la correlación entre el IMC y el porcentaje de grasa corporal es mucho menor y será necesario usar además otras medidas antropométricas o fisiológicas para poder predecir adecuadamente riesgos cardiovasculares e identificar el grado de obesidad real de los varones. En esta dirección, hay más por explorar en el futuro.

Se resalta aquí que si el uso de la información presentada en este trabajo fuera modificar políticas de salud a nivel estatal, entonces se recomienda considerar el modelo de Box y Cox ajustado al estado en consideración. En contraste, si lo que se desea es contar con un resultado nacional, pues entonces los modelos lognormales propuestos aquí para hombres y mujeres son simples y adecuados a analizar, considerar y usar.

Finalmente, se destaca, que si bien este trabajo ofrece algunos resultados a partir de los cuáles se pueden establecer políticas de en el ámbito de la salud de los mexicanos, el trabajo en este tema no está terminado. Por el contrario consideramos que esta primer aproximación es un buen punto de partida para futuras investigaciones médicas.

Apéndice A. Figuras del Capítulo 2

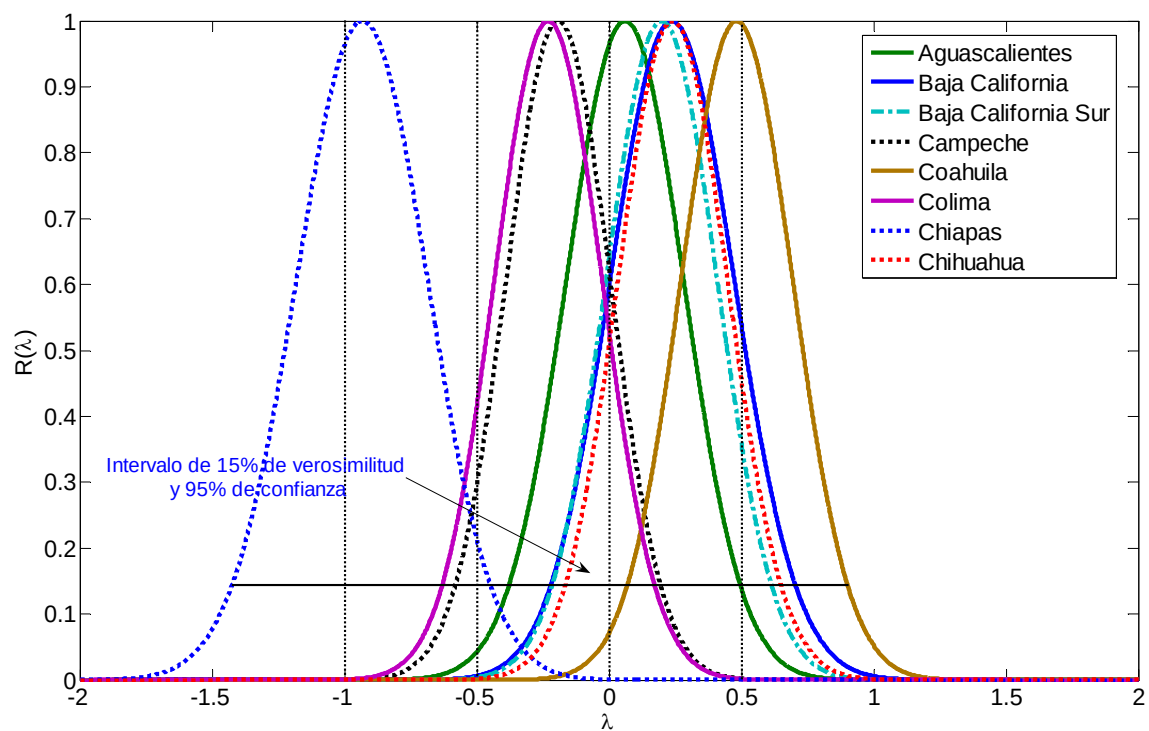


Figura 2.1. Verosimilitud relativa de λ para hombres por estados.

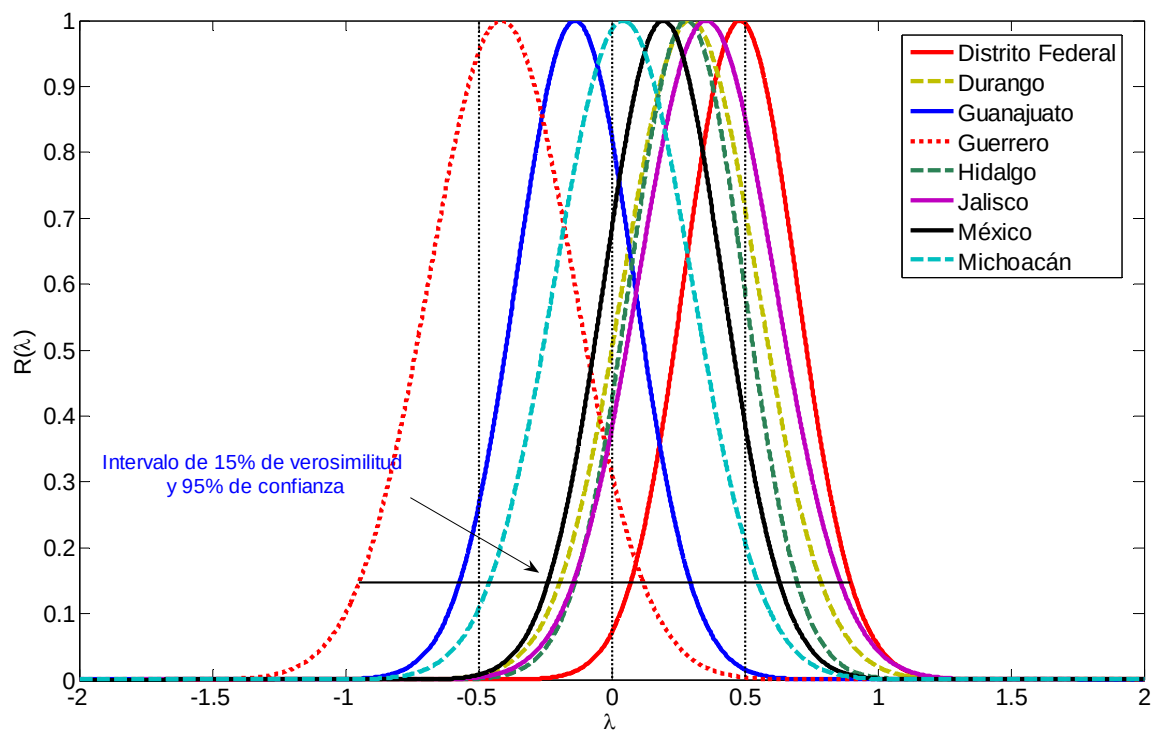


Figura 2.2. Verosimilitud relativa de λ para hombres por estados.

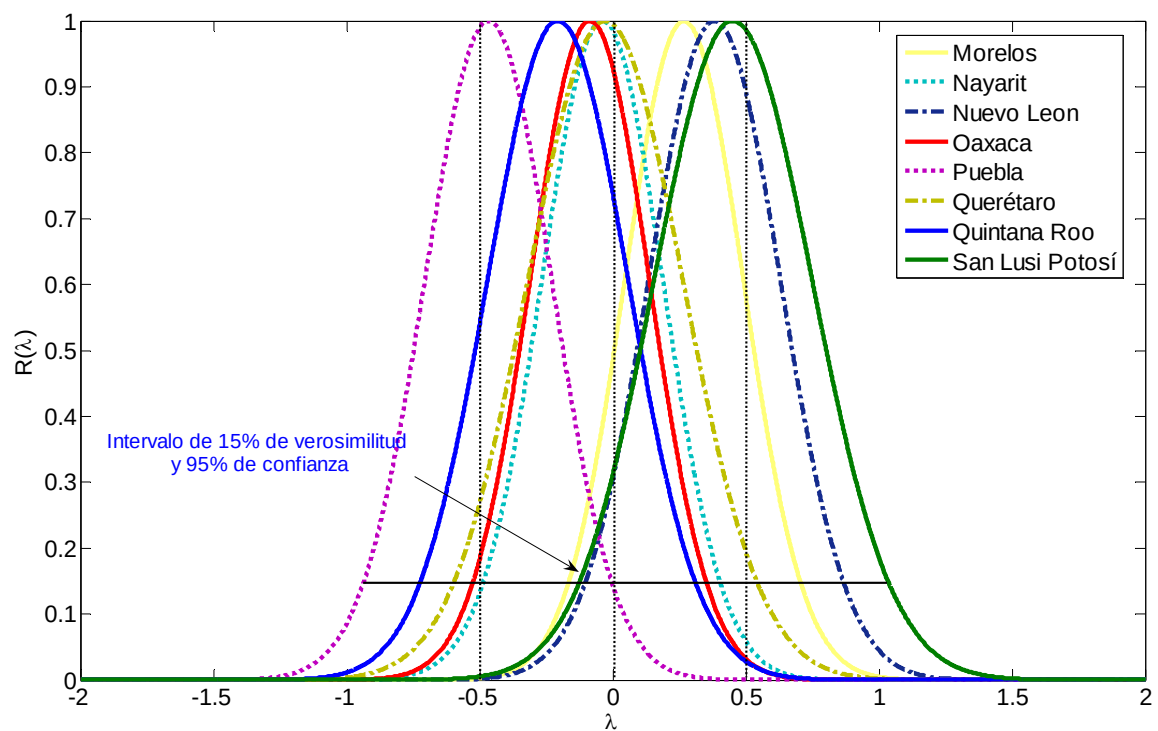


Figura 2.3. Verosimilitud relativa de λ para hombres por estados.

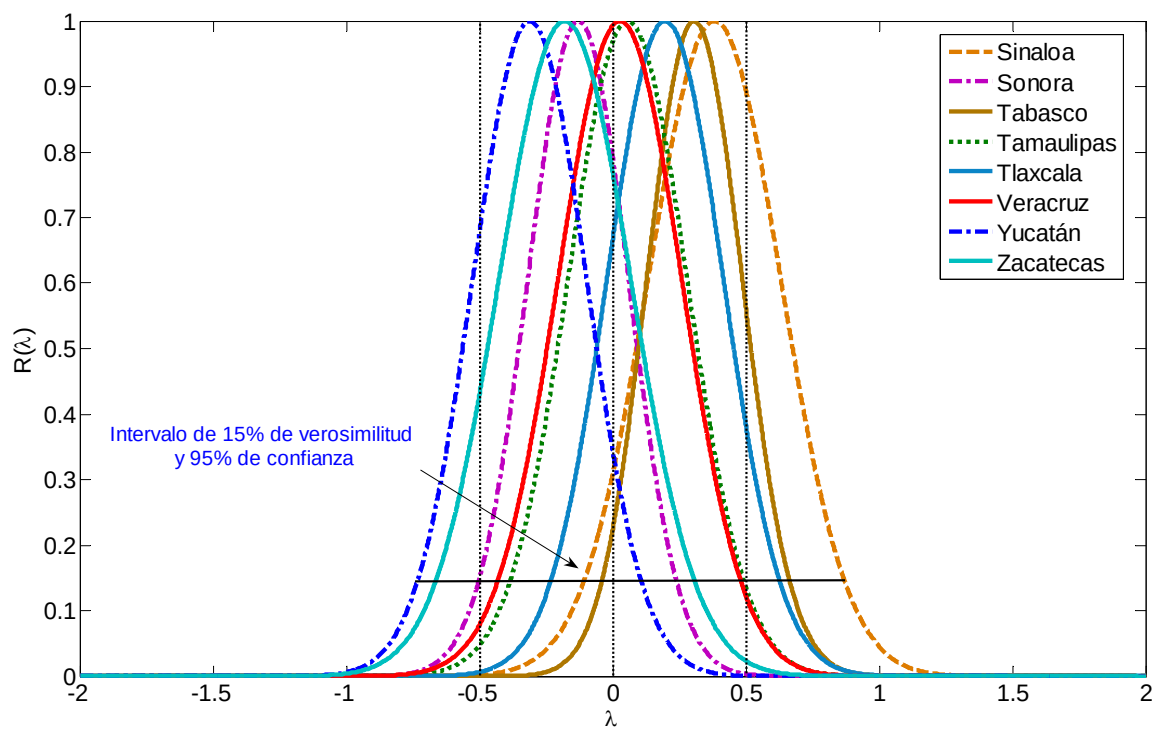


Figura 2.4. Verosimilitud relativa de λ para hombres por estados.

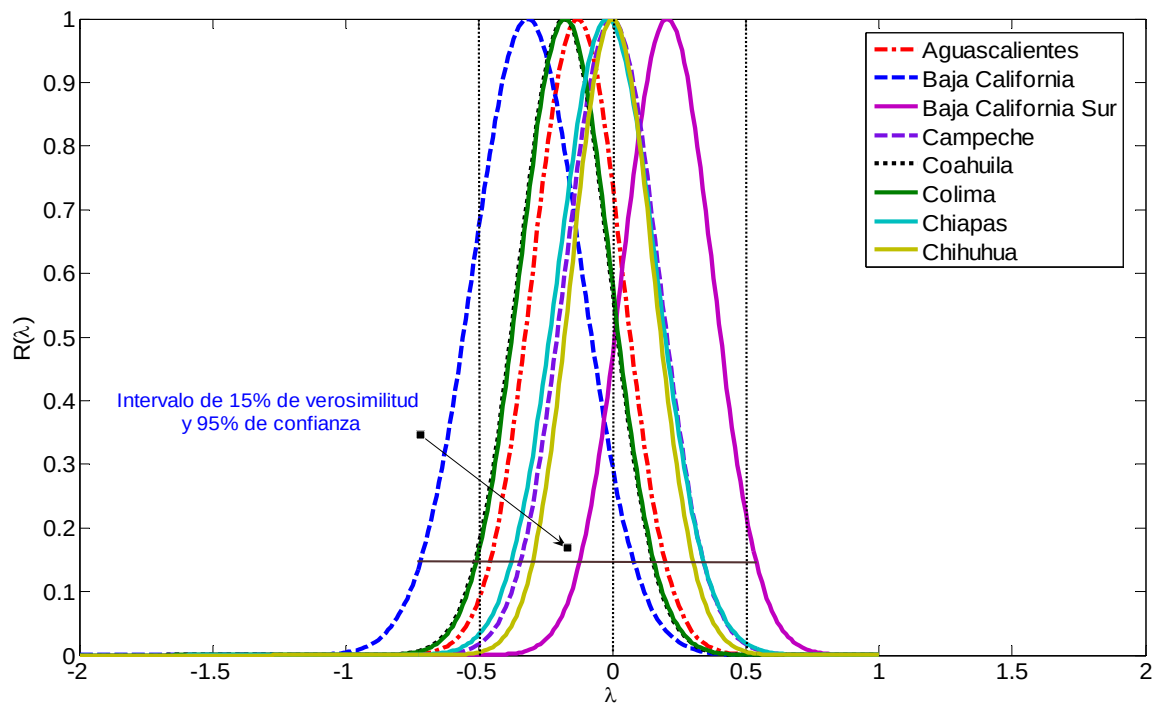


Figura 2.5. Verosimilitud relativa de λ para mujeres por estados.

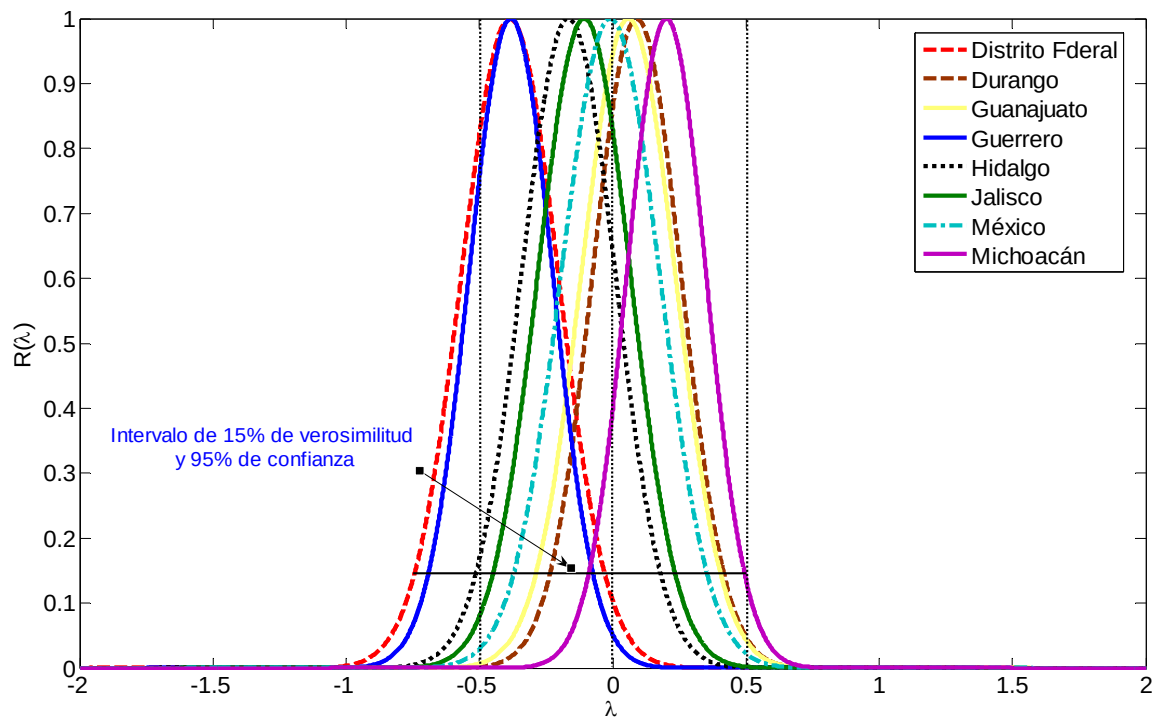


Figura 2.6. Verosimilitud relativa de λ para mujeres por estados.

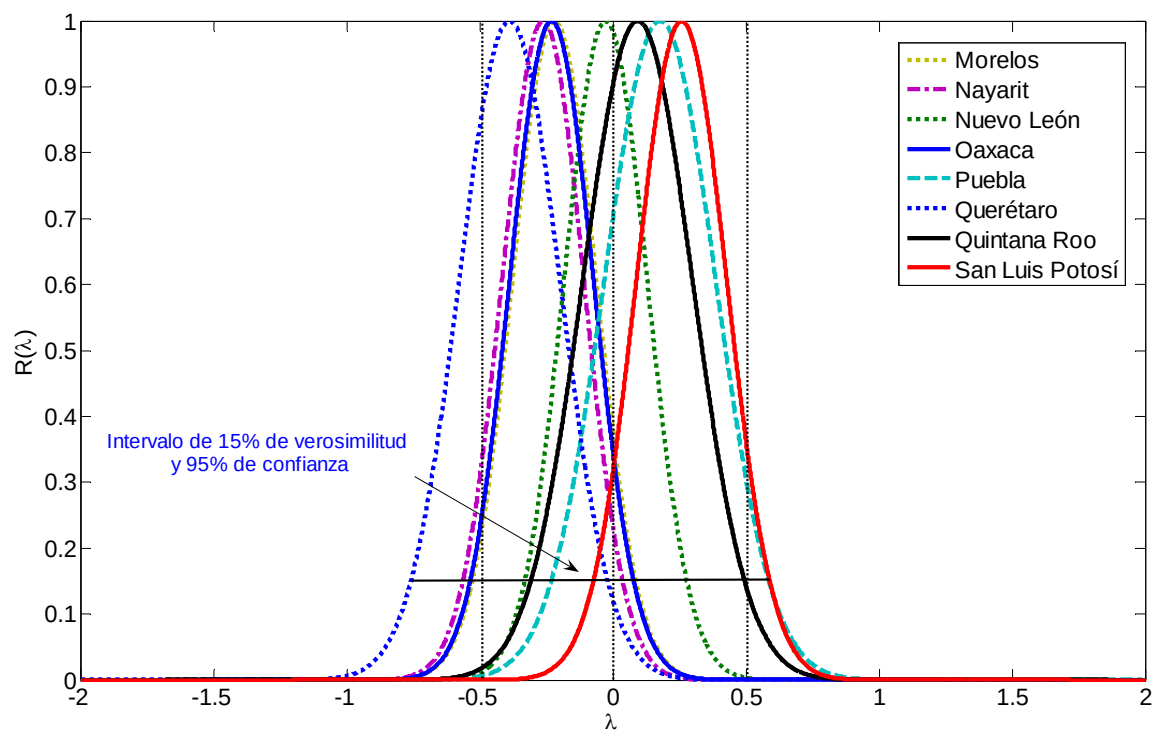


Figura 2.7. Verosimilitud relativa de λ para mujeres por estados.

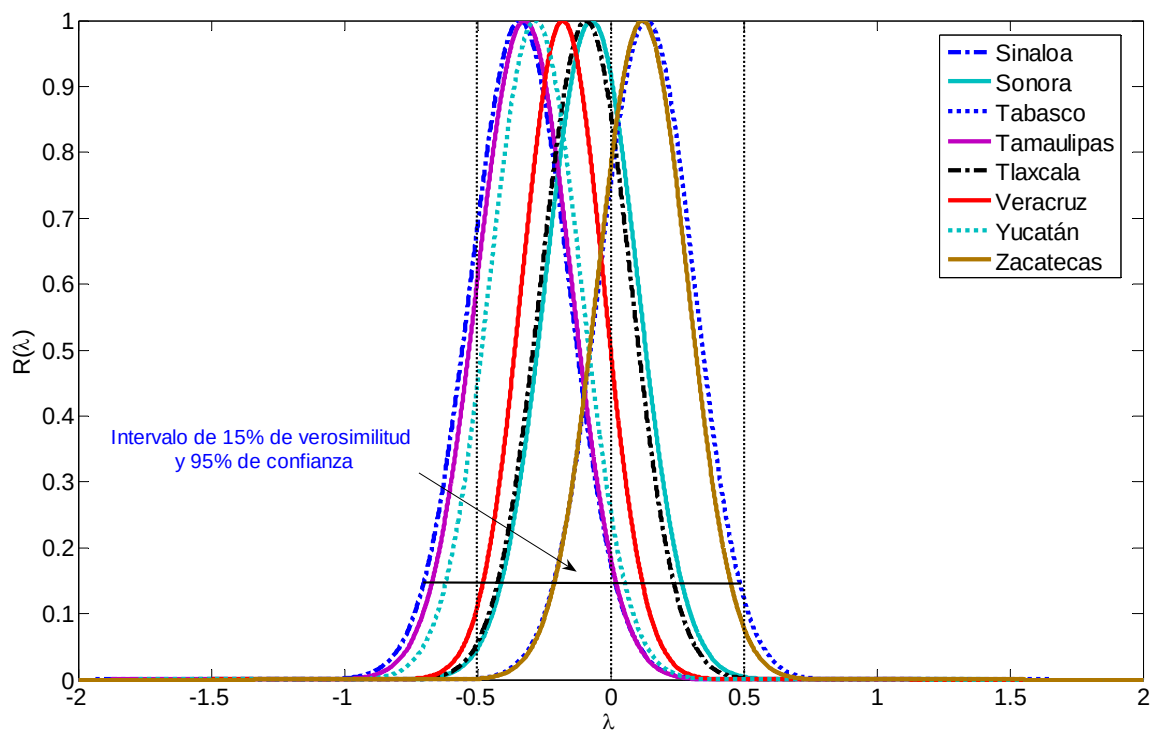


Figura 2.8. Verosimilitud relativa de λ para mujeres por estados.

AGUASCALIENTES

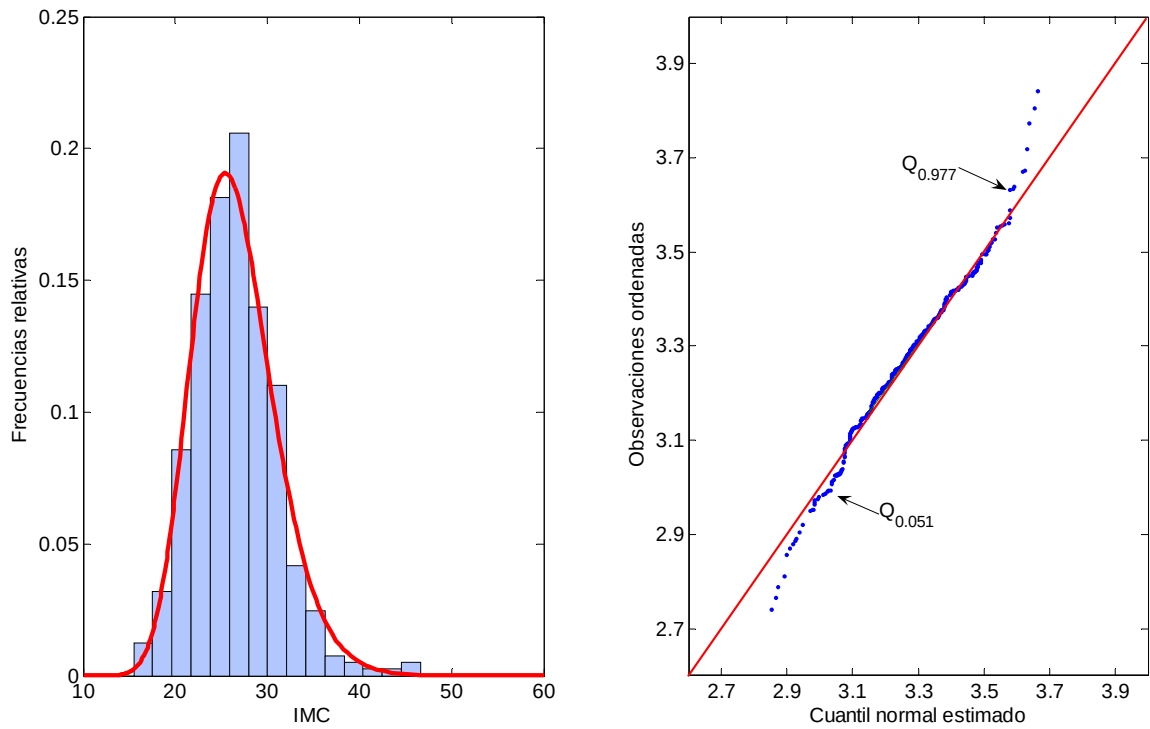


Figura 2.9. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

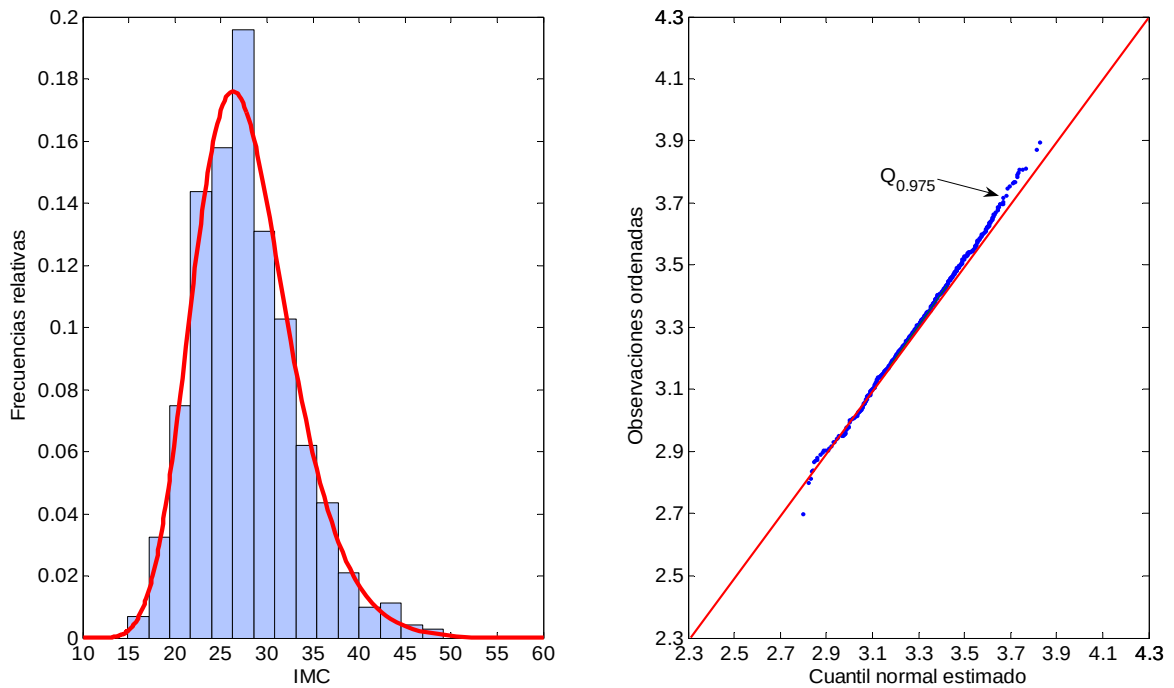


Figura 2.10. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

BAJA CALIFORNIA

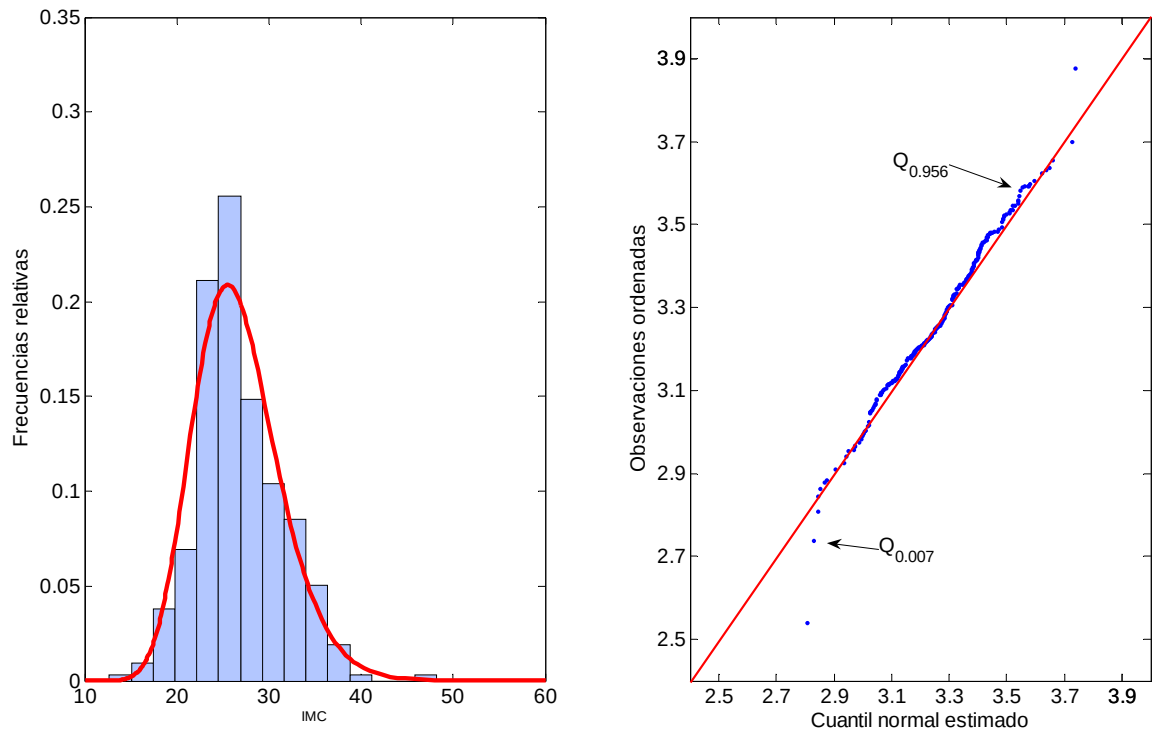


Figura 2.11. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

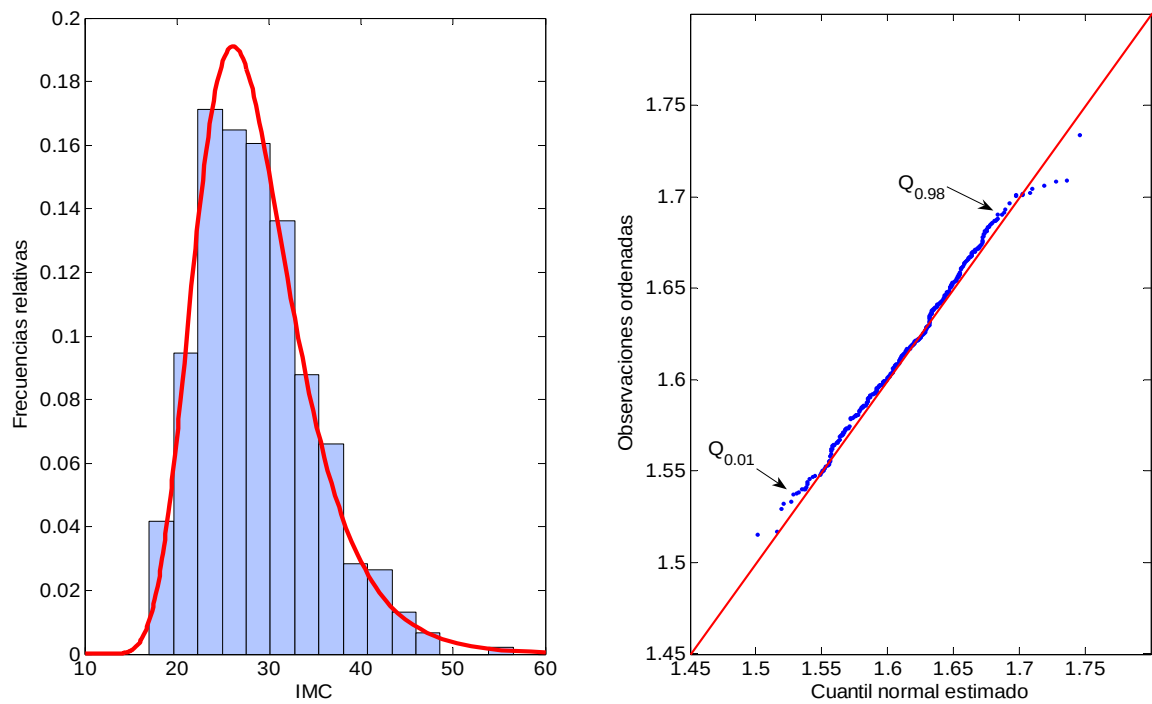


Figura 2.12. Mujeres Ajuste con el Modelo de Box-Cox $\lambda = -0.5$ (izquierda) y gráfica QQ (derecha)

BAJA CALIFORNIA SUR

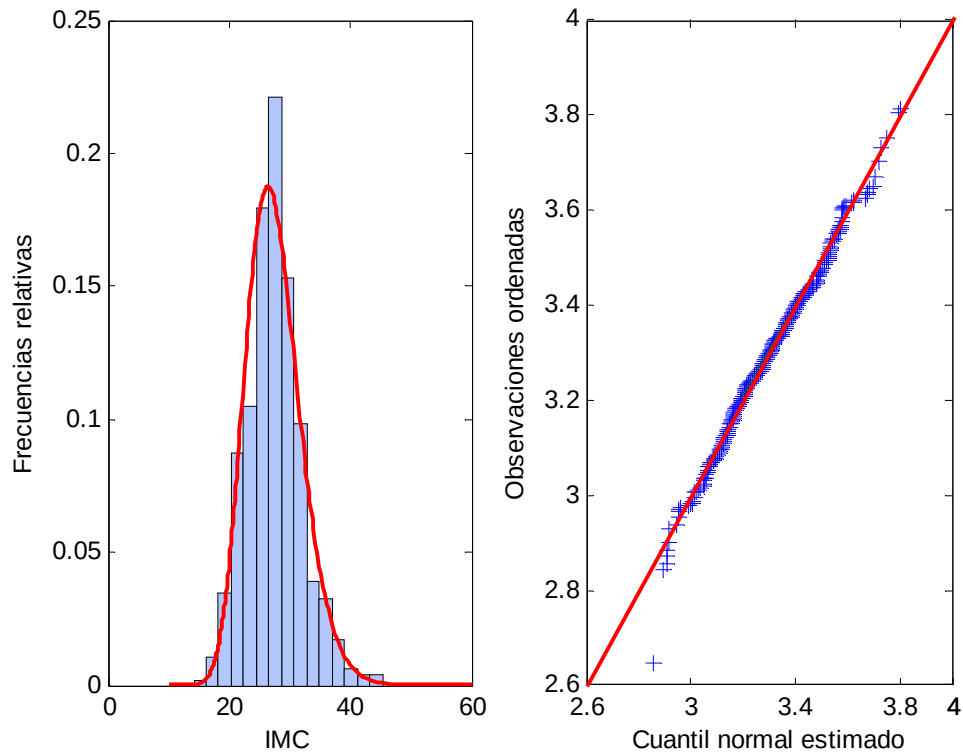


Figura 2.13. Hombres Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

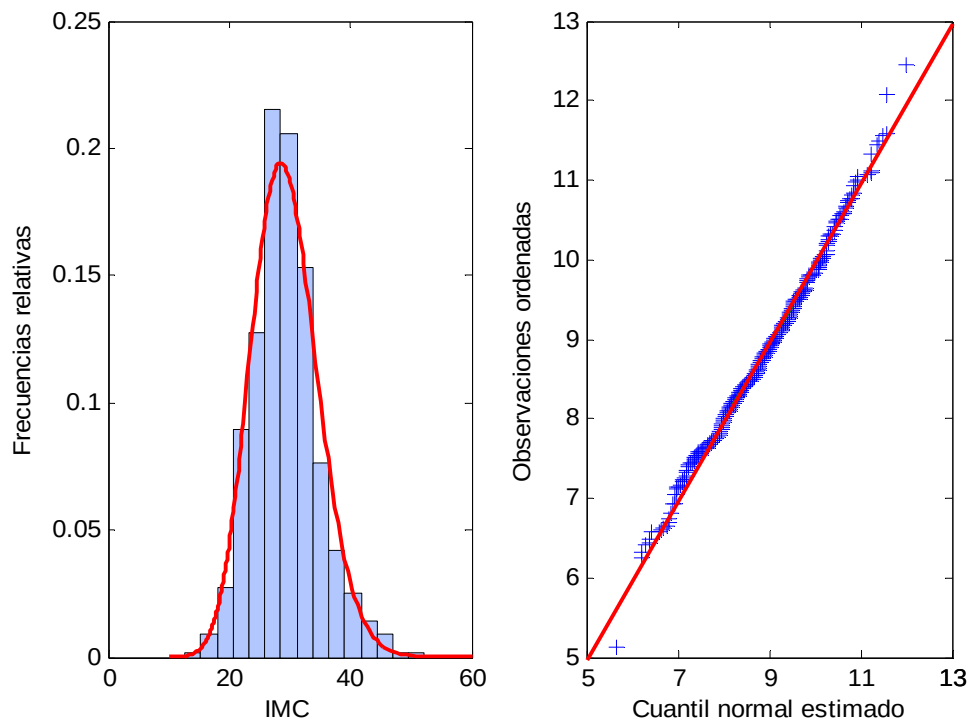


Figura 2.14. Mujeres. Ajuste con el Modelo de Box-Cox $\lambda=0.5$ (izquierda) y gráfica QQ (derecha)

CAMPECHE

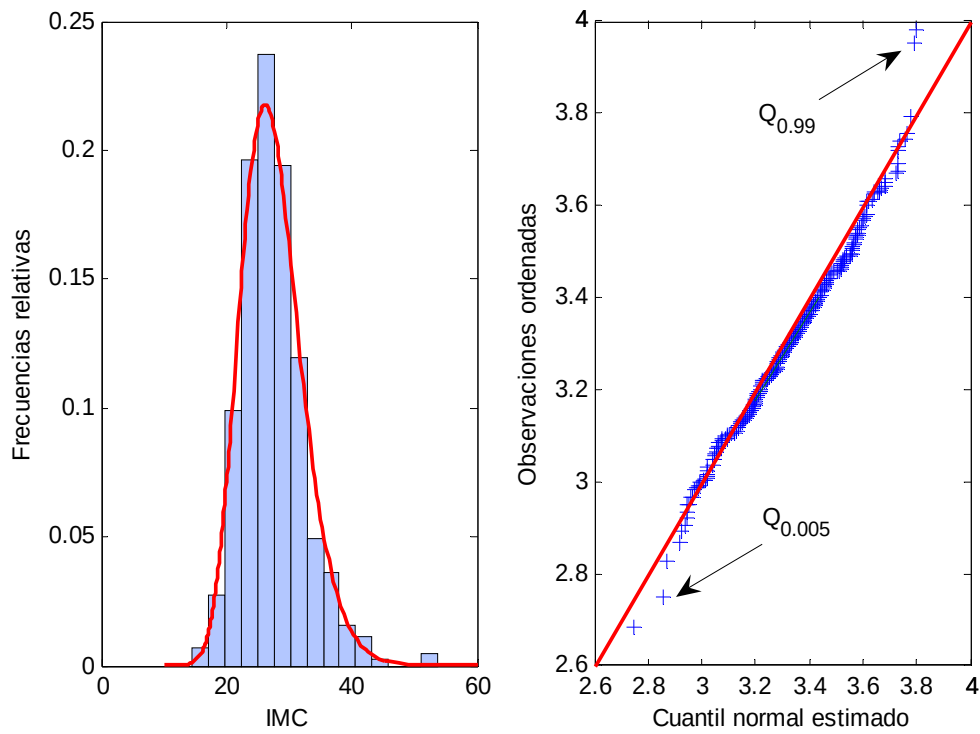


Figura 2.15. Hombres Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

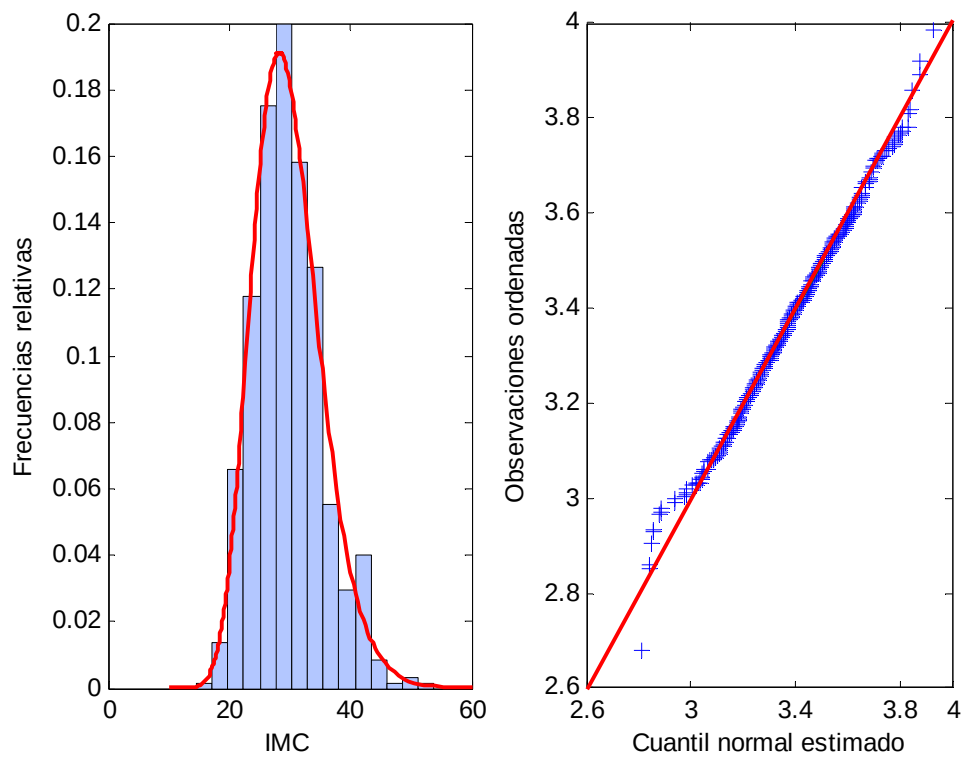


Figura 2.16. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

COAHUILA

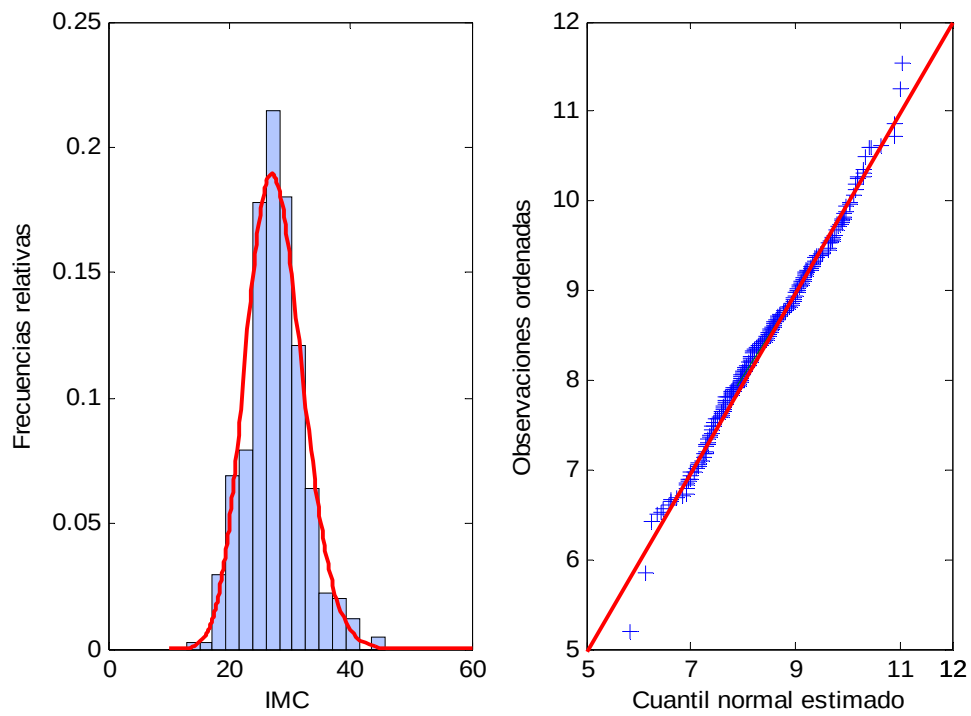


Figura 2.17. Hombres. Ajuste con el Modelo de Box-Cox $\lambda=0.5$ (izquierda) y gráfica QQ (derecha)

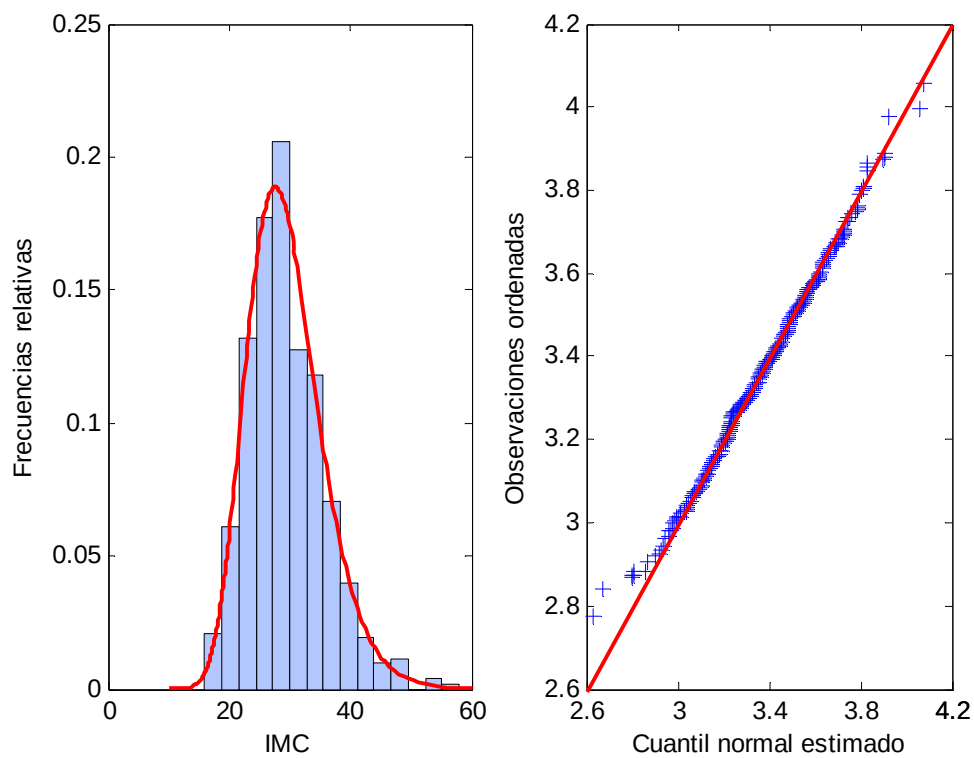


Figura 2.18 Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

COLIMA

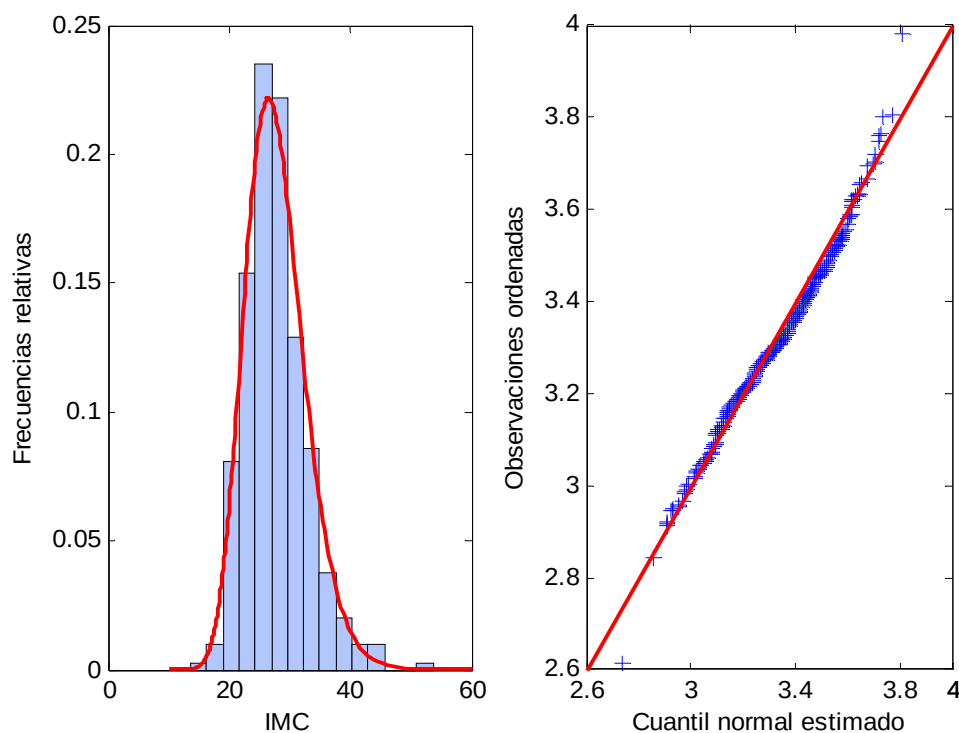


Figura 2.19. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

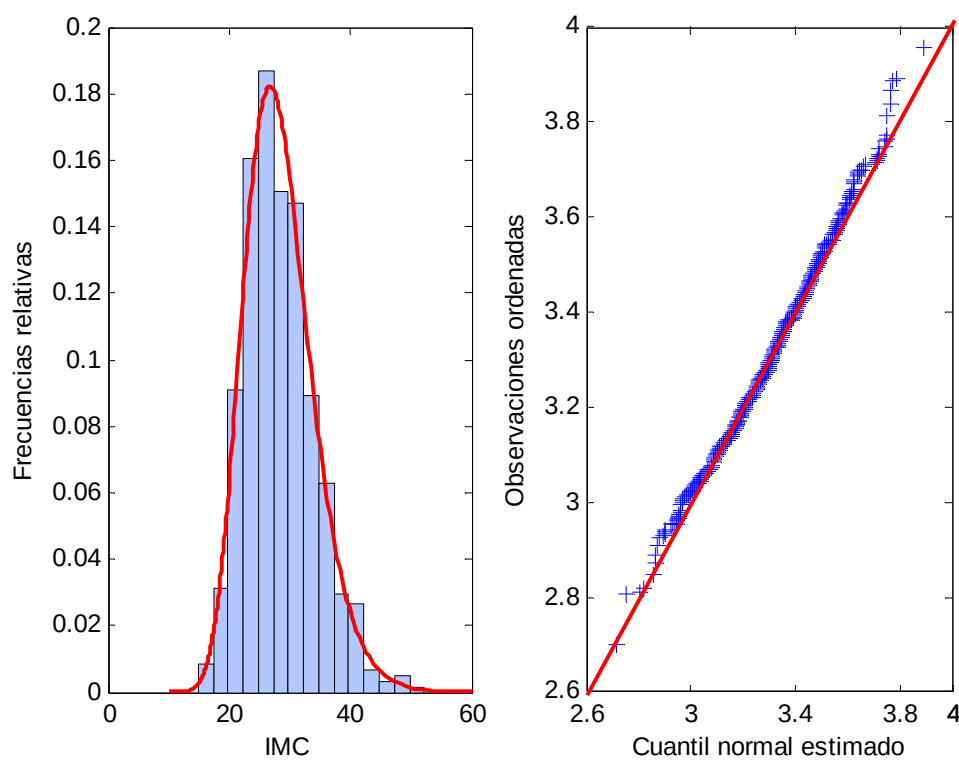


Figura 2.20 Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

CHIAPAS

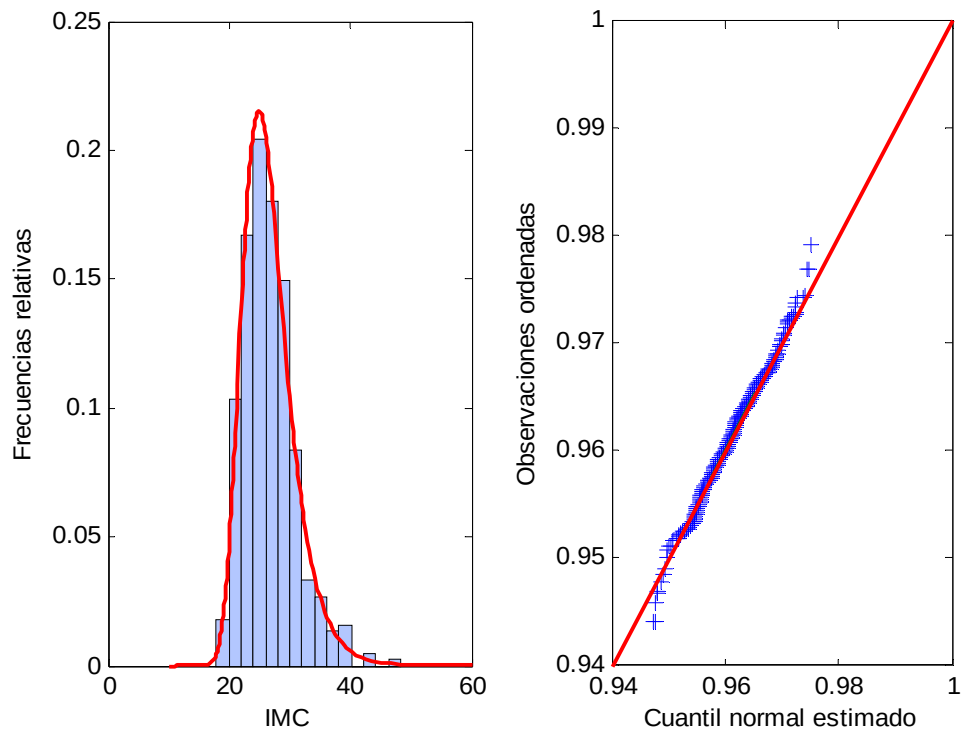


Figura 2.21. Hombres. Ajuste con el Modelo de Box-Cox $\lambda=-1$ (izquierda) y gráfica QQ (derecha)

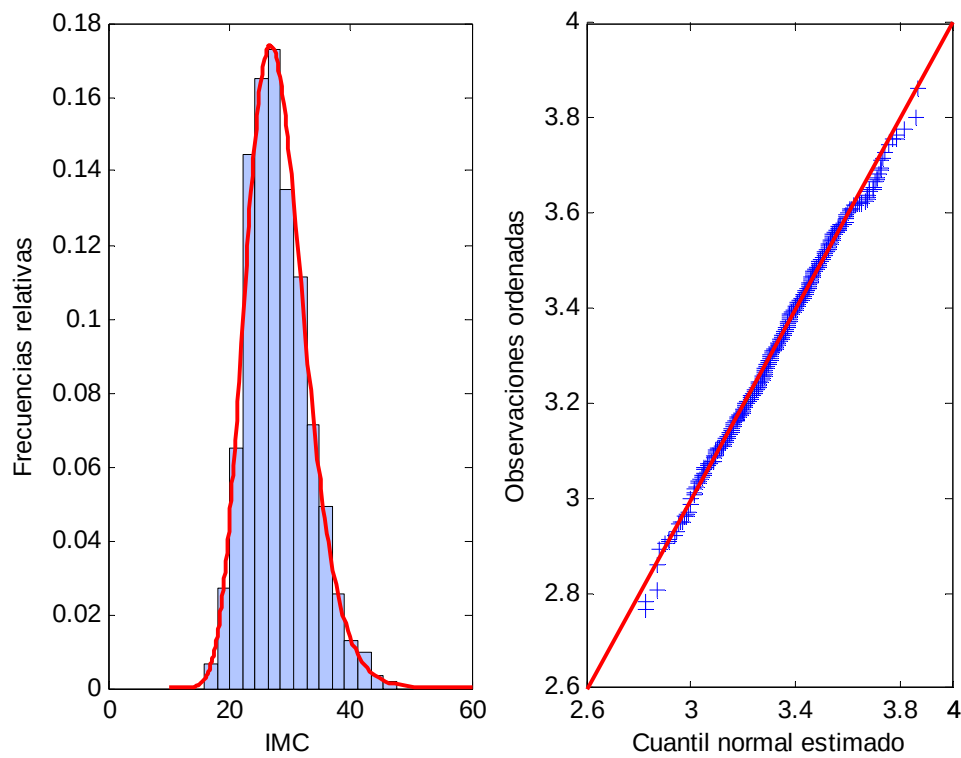


Figura 2.22. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

CHIHUAHUA

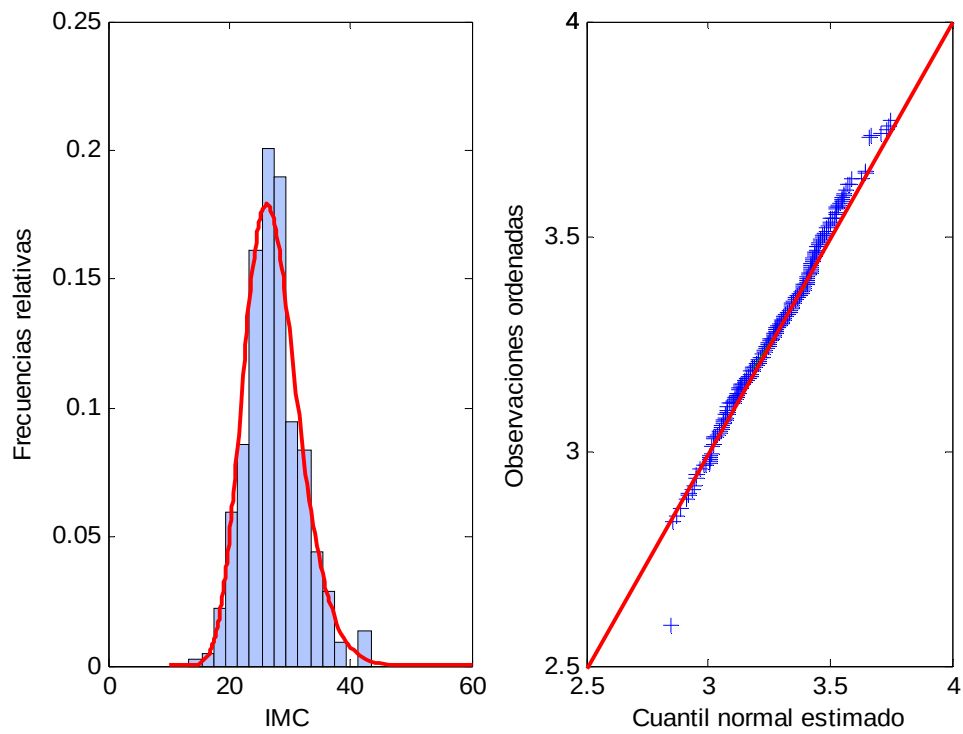


Figura 2.23. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

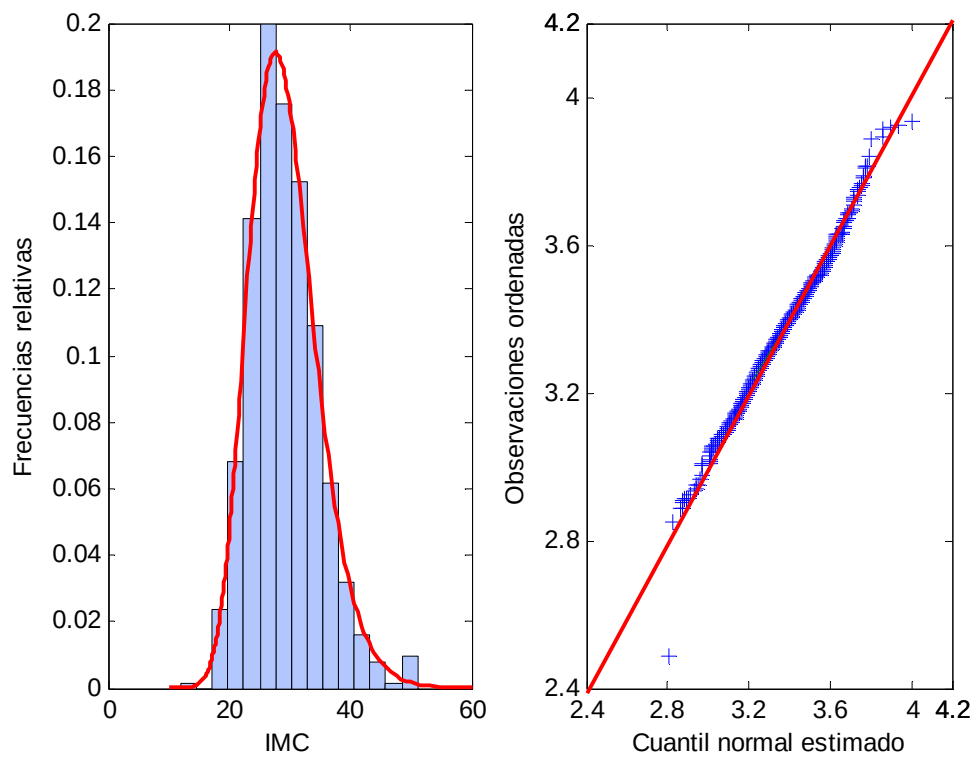


Figura 2.24. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

DISTRITO FEDERAL

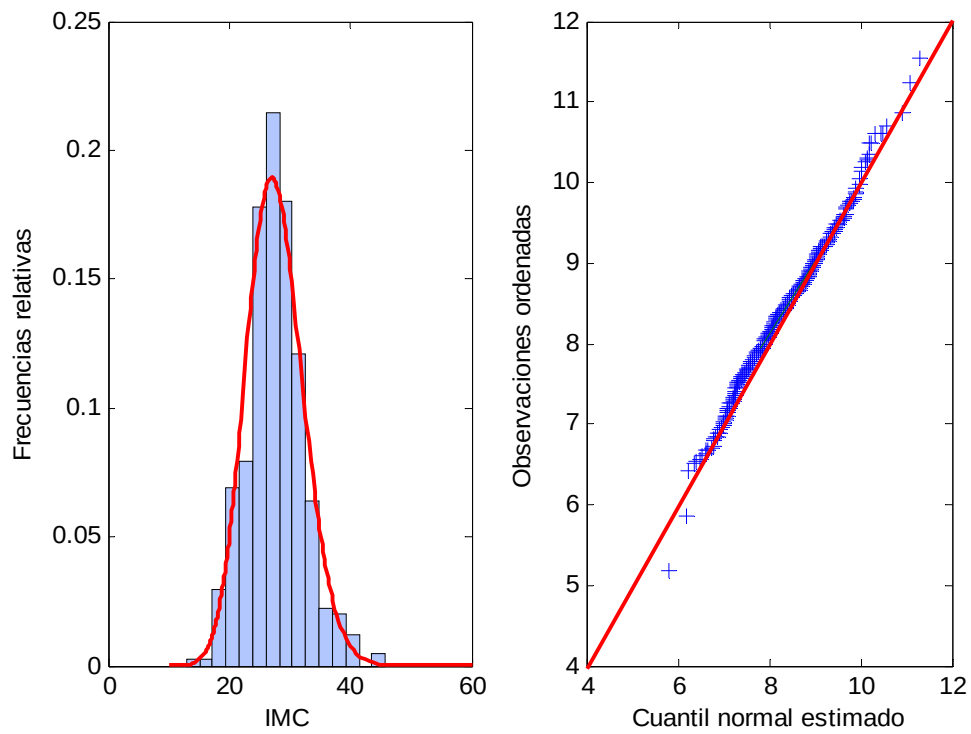


Figura 2.25. Hombres Ajuste con el Modelo de Box-Cox $\lambda=0.5$ (izquierda) y gráfica QQ (derecha)

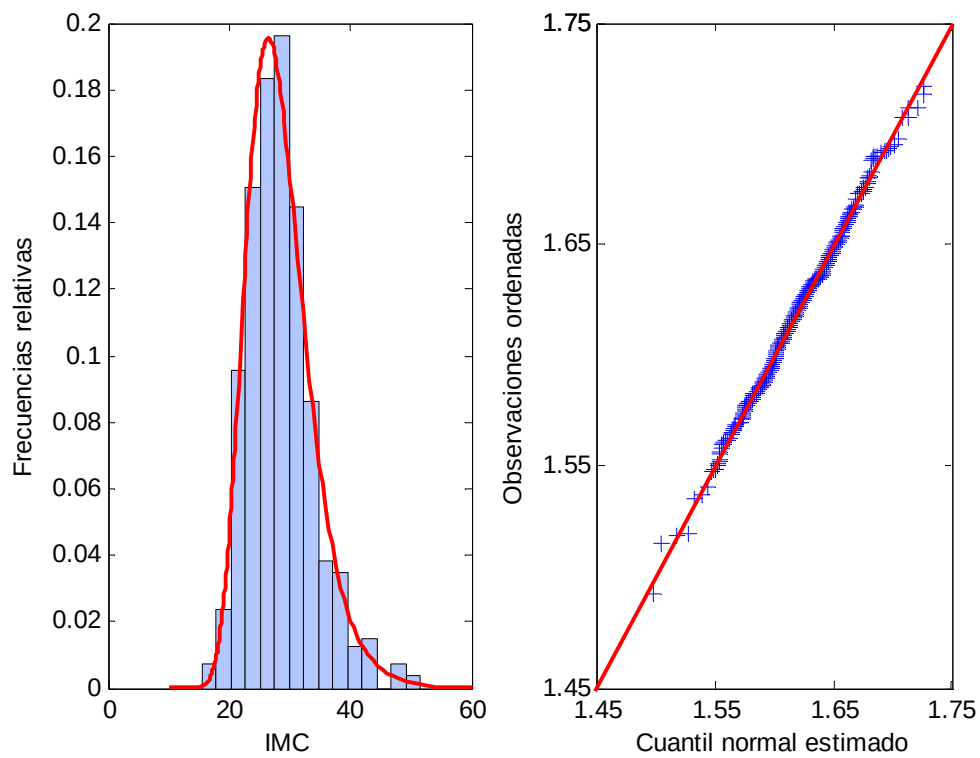


Figura 2.26 Mujeres. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ (izquierda) y gráfica QQ (derecha)

DURANGO

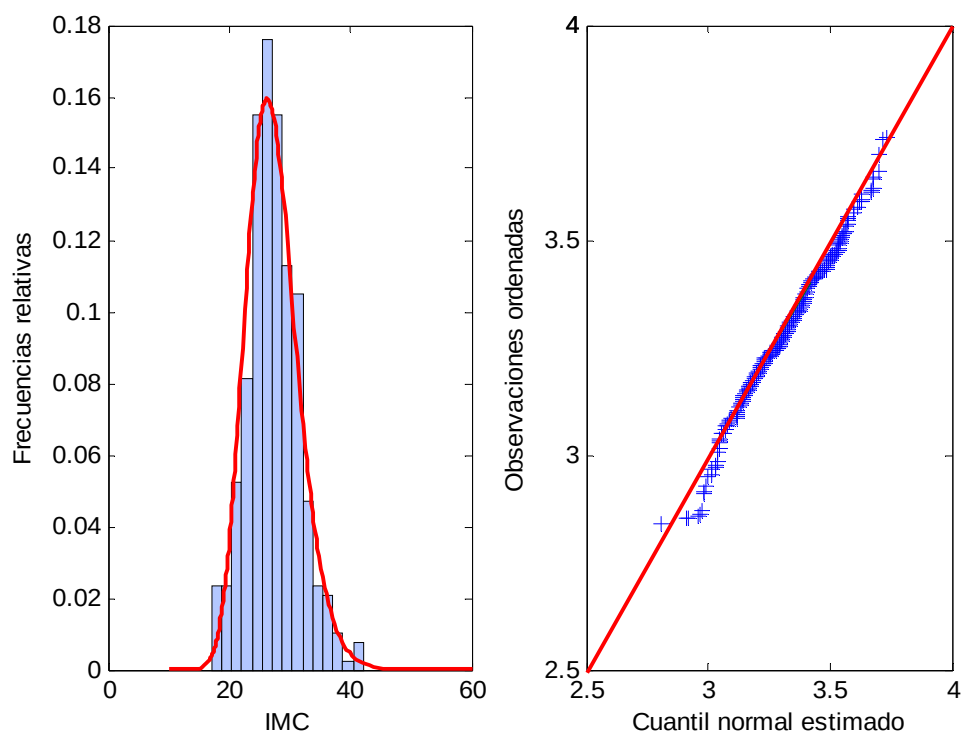


Figura 2.27. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

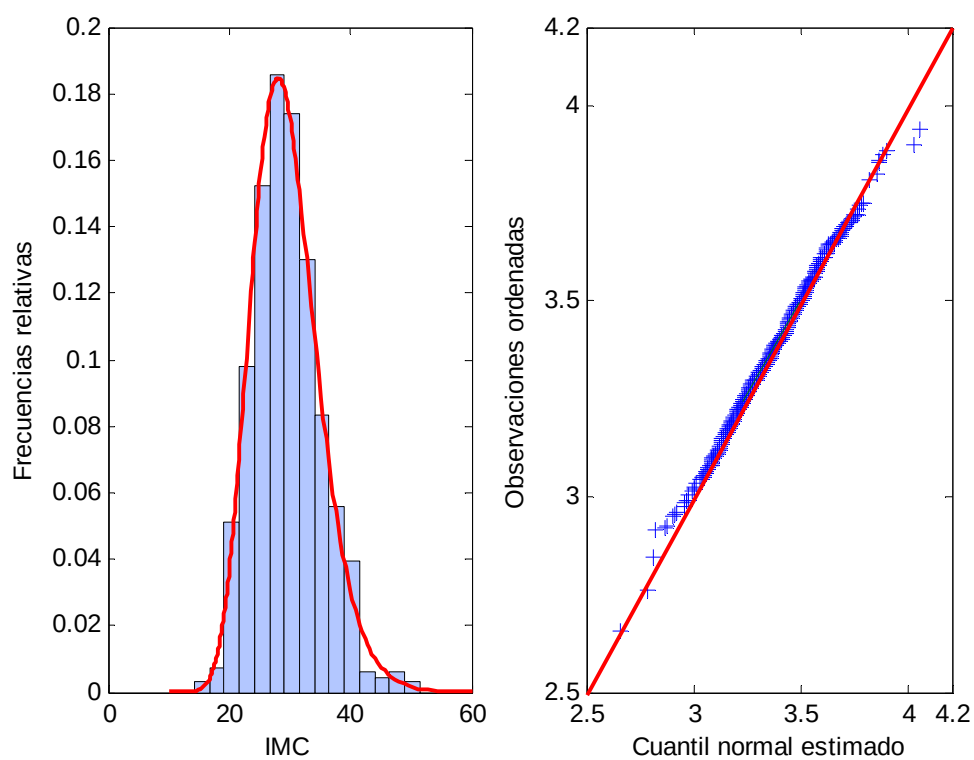


Figura 2.28 Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

GUANAJUATO

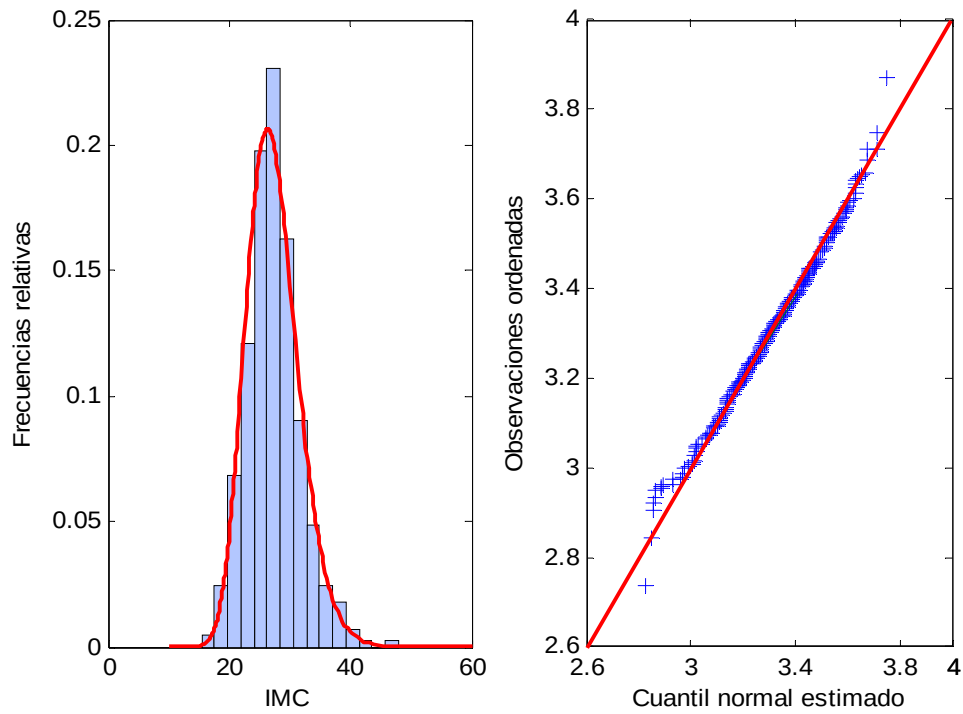


Figura 2.29. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

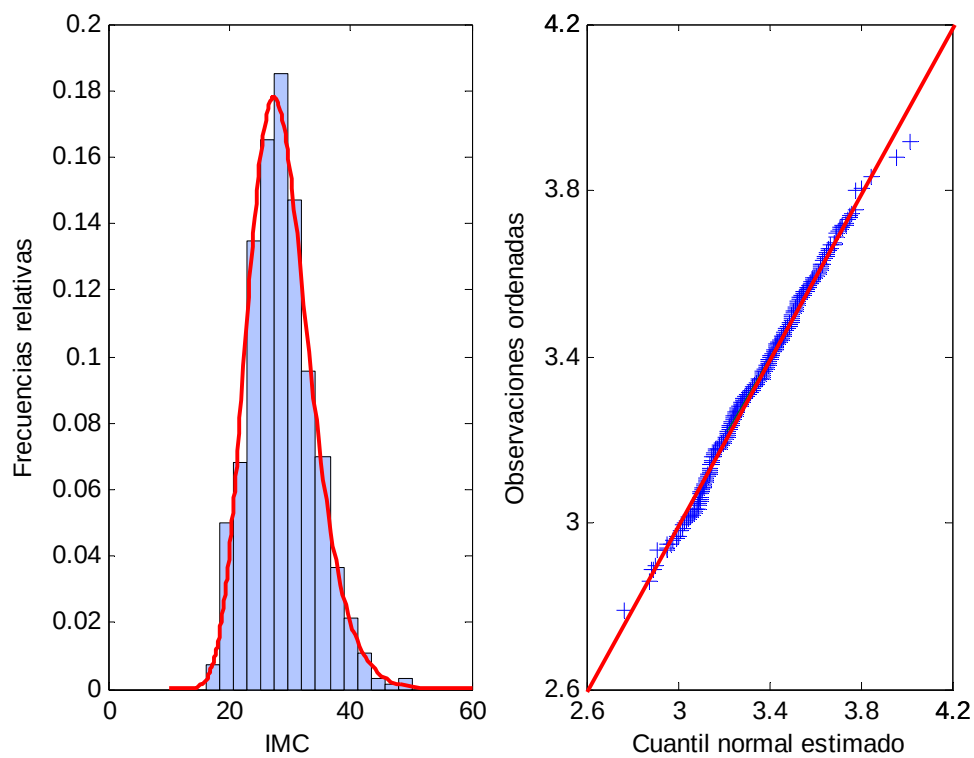


Figura 2.30. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

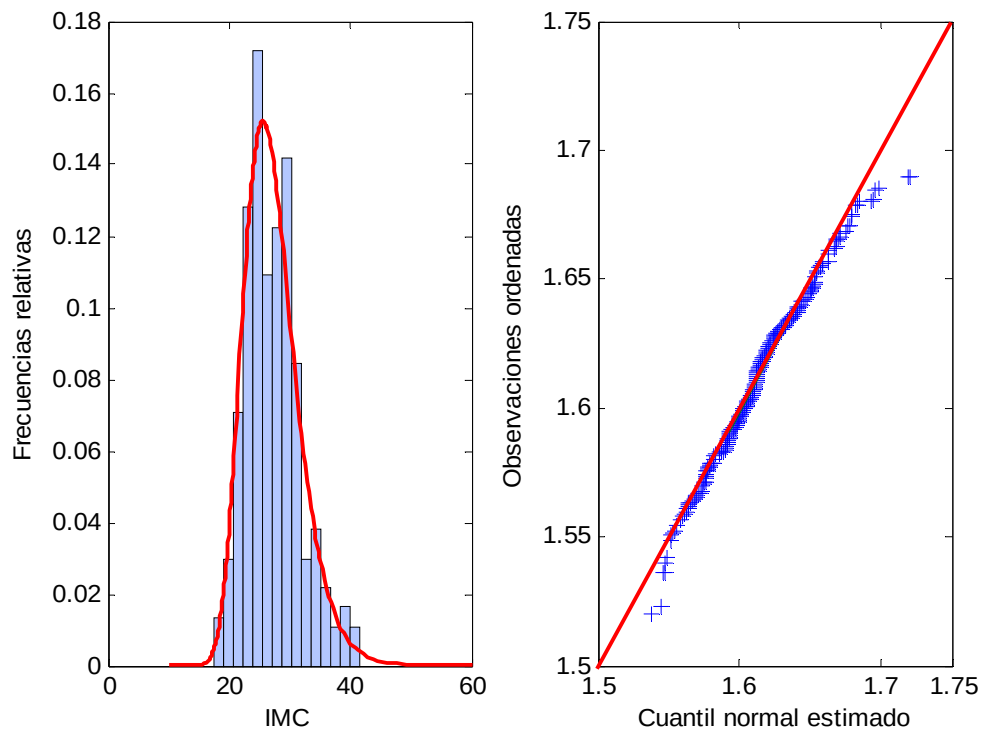


Figura 2.31. Hombres. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ (izquierda) y gráfica QQ (derecha)

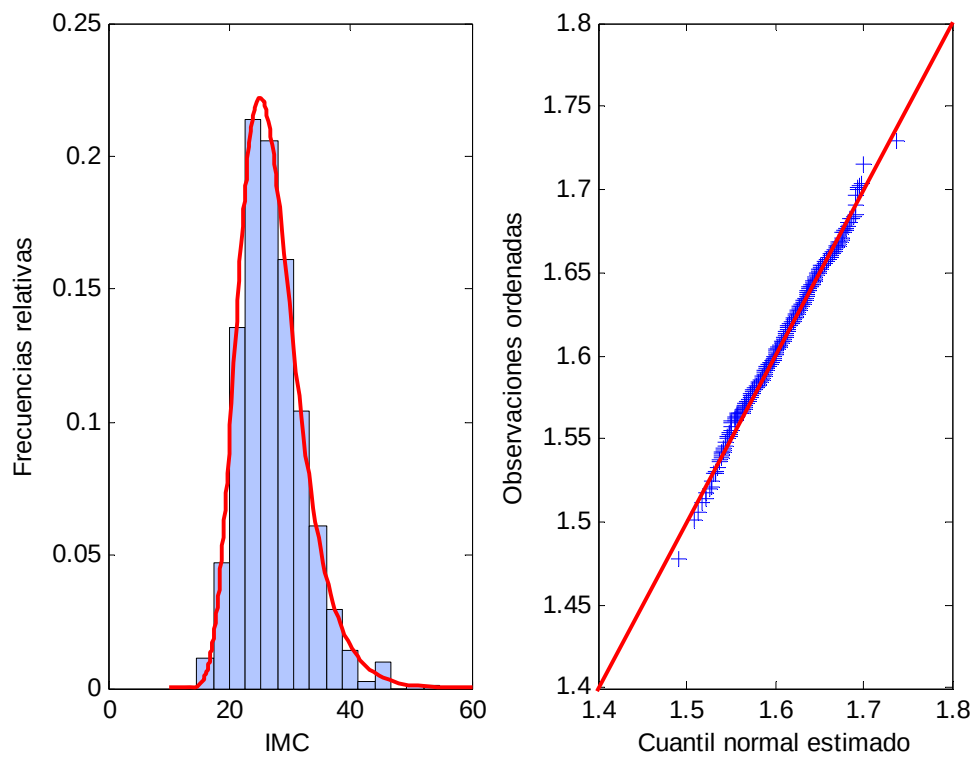


Figura 2.32. Mujeres. Ajuste con el Modelo Box-Cox $\lambda=-0.5$ (izquierda) y gráfica QQ (derecha)

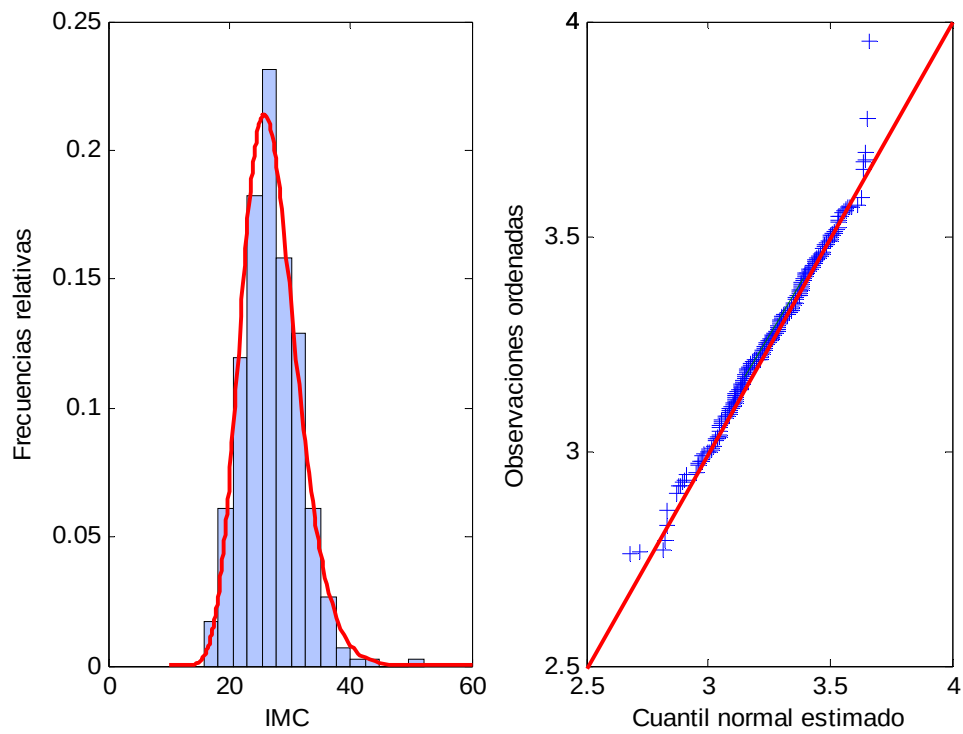


Figura 2.33. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

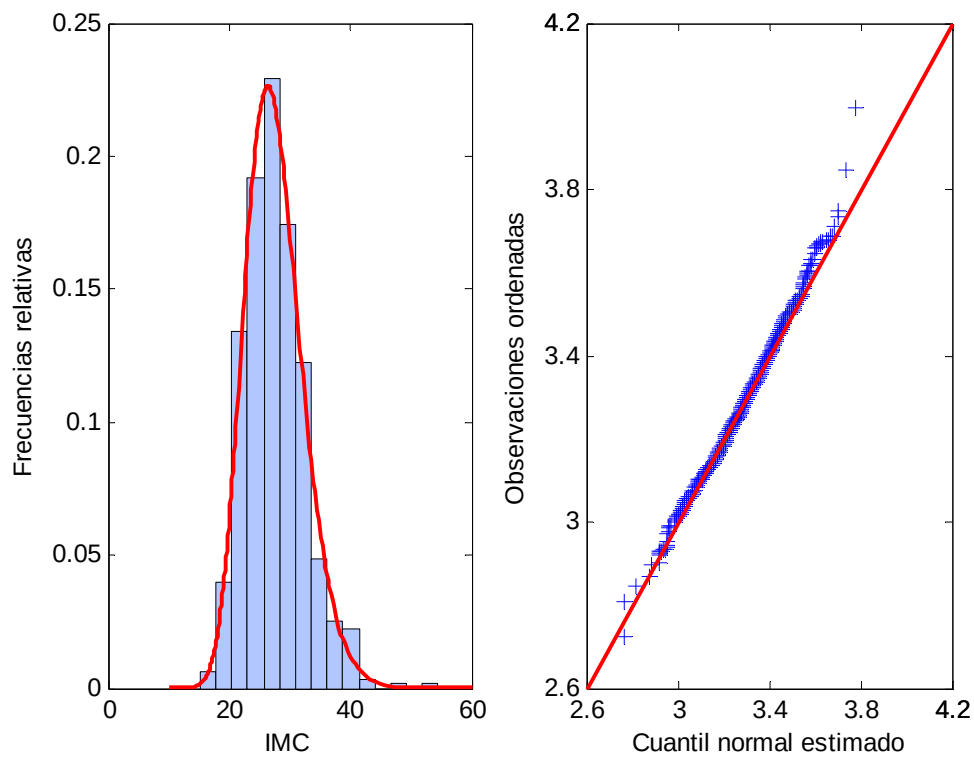


Figura 2.34. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

JALISCO

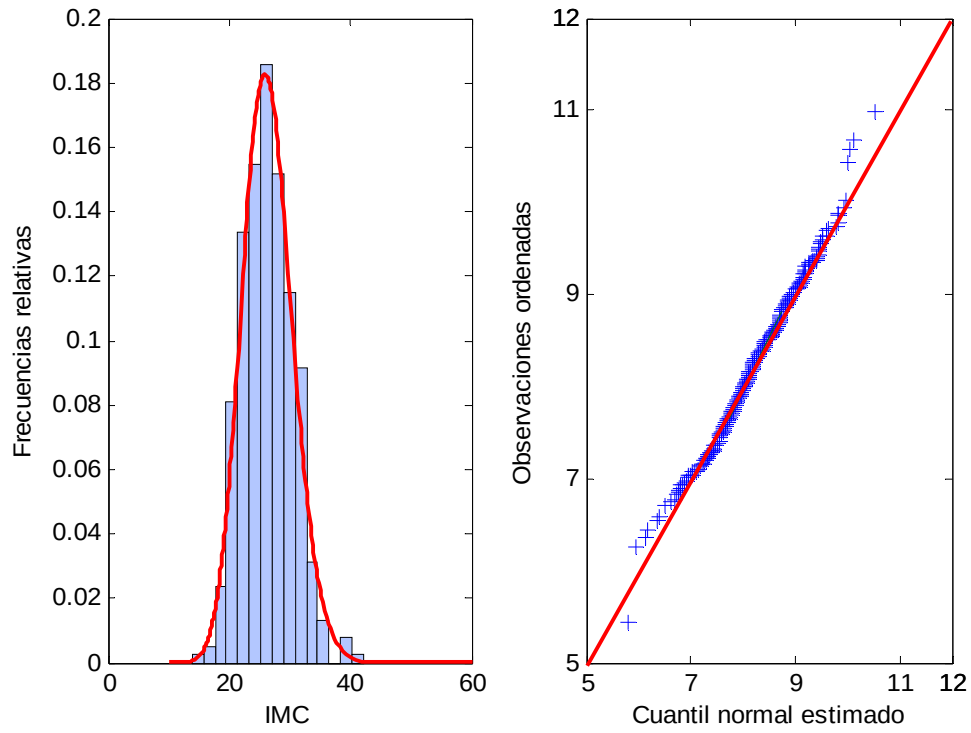


Figura 2.35. Hombres Ajuste con el Modelo de Box-Cox $\lambda=0.5$ (izquierda) y gráfica QQ (derecha)

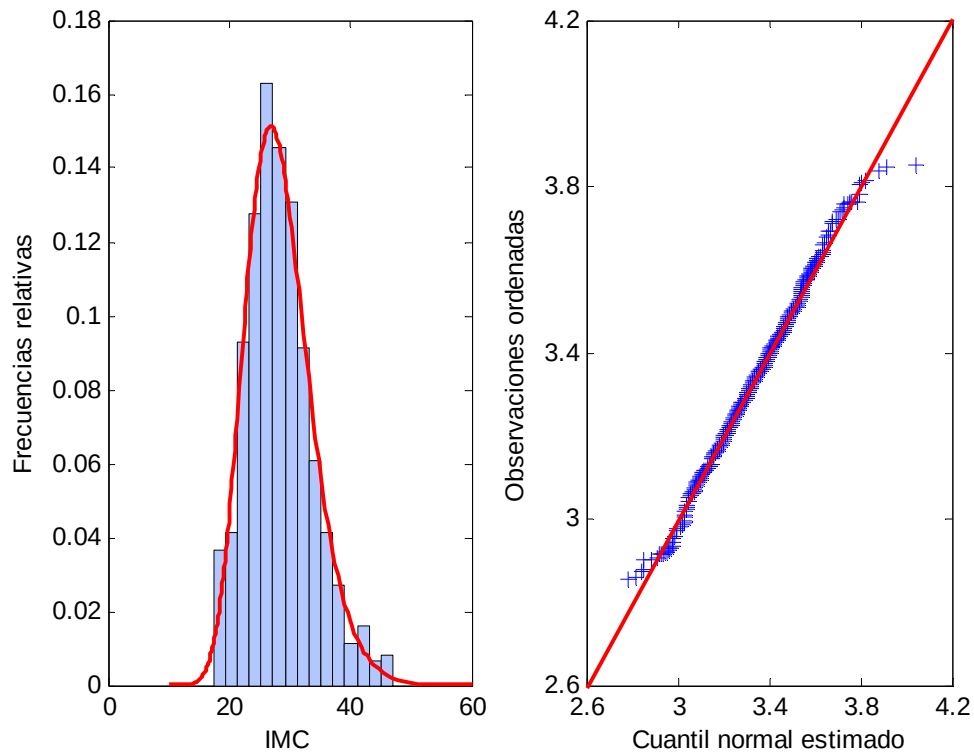


Figura 2.36. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

MÉXICO

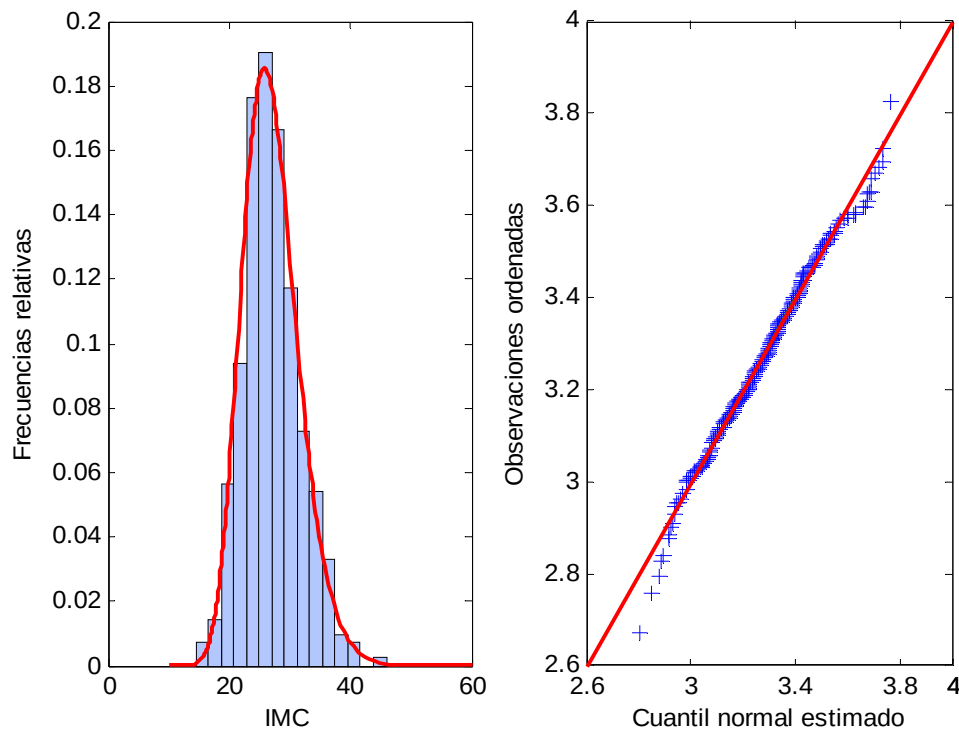


Figura 2.37. Hombres Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

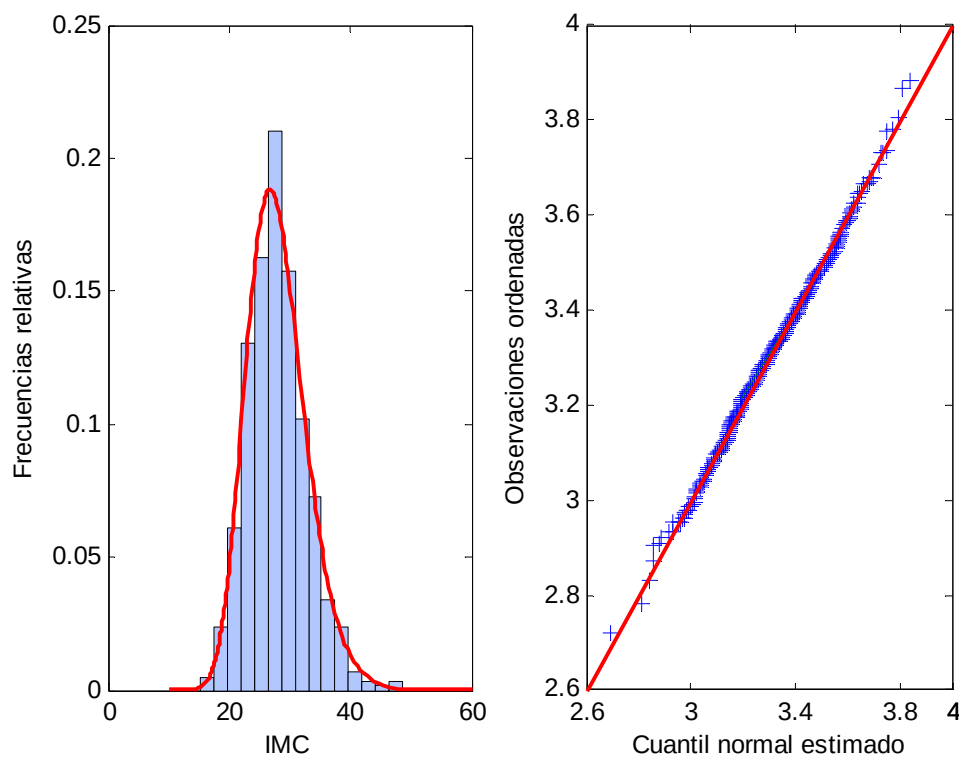


Figura 2.38. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

MICHOACÁN

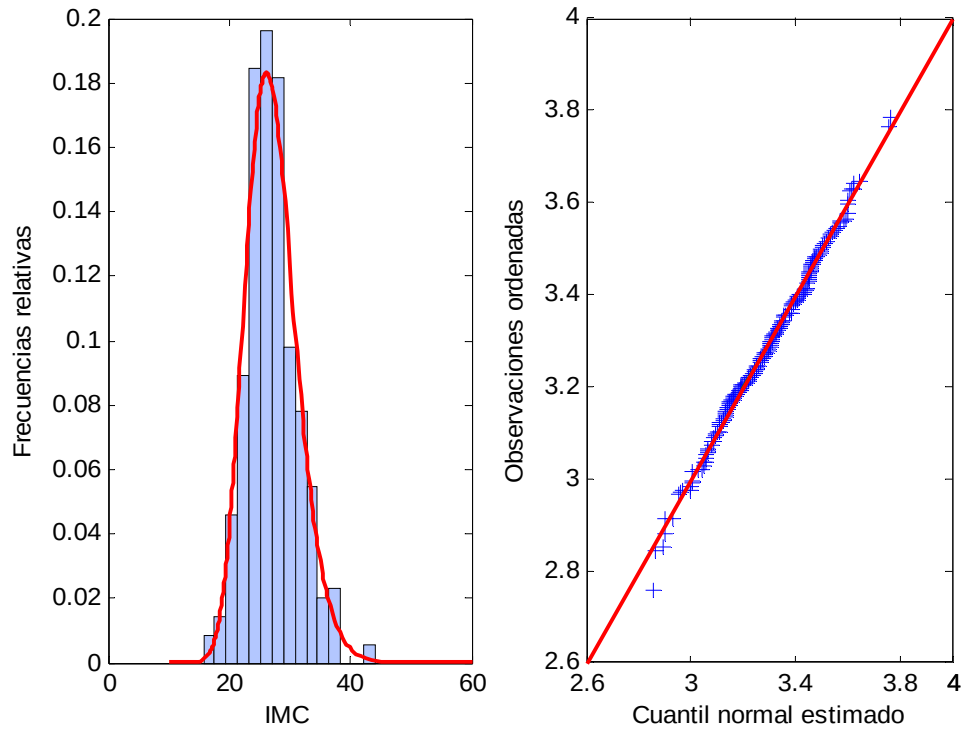


Figura 2.39. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

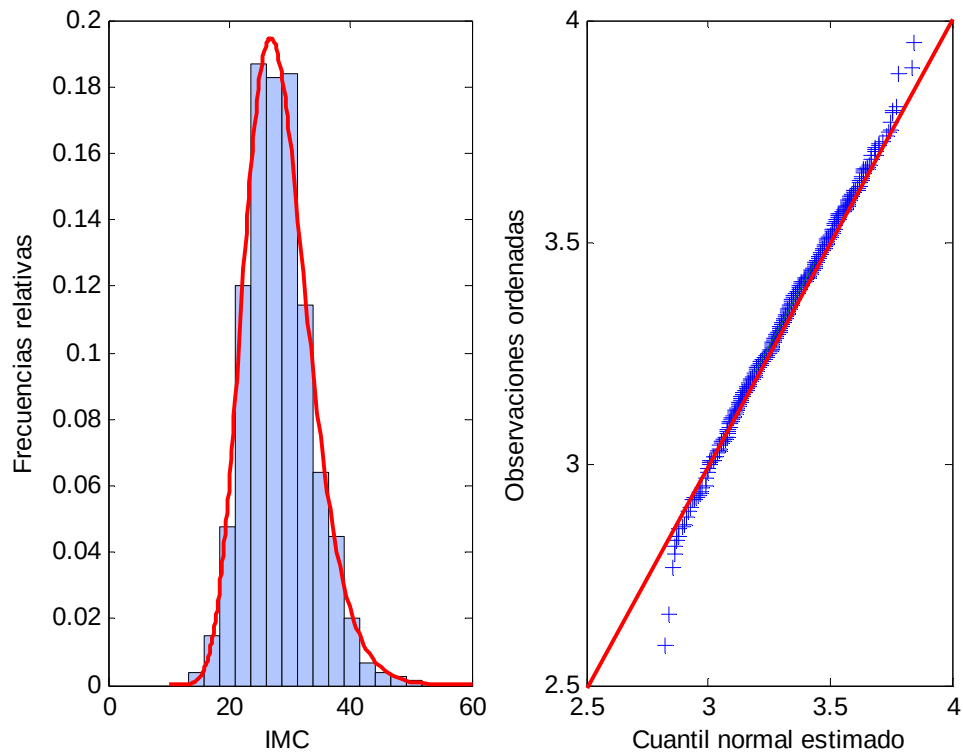


Figura 2.40 Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

MORELOS

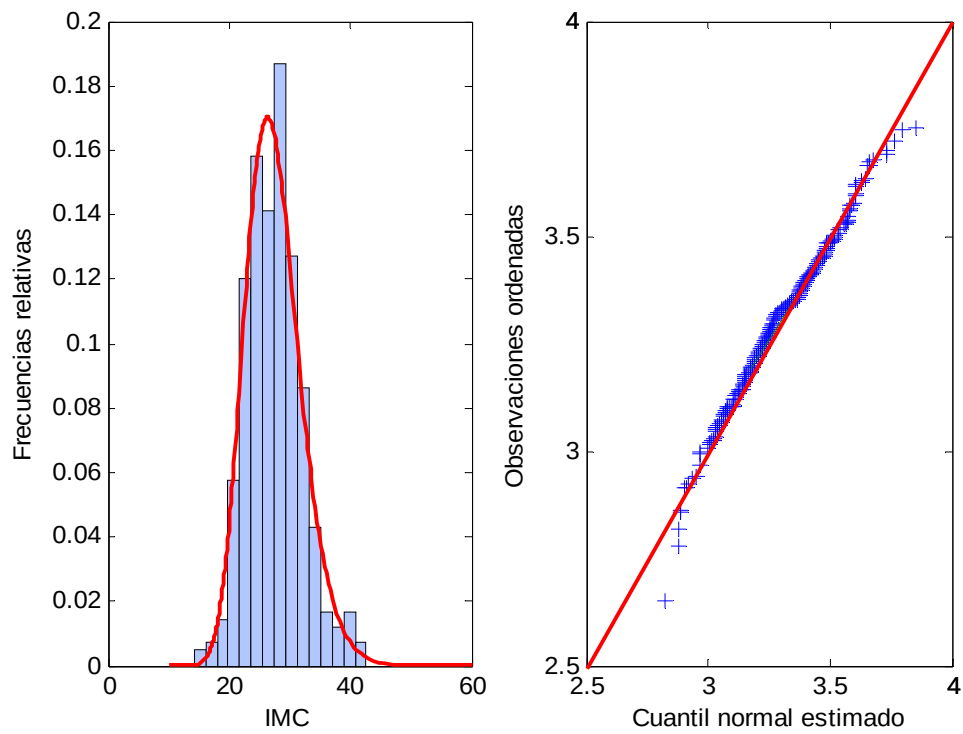


Figura 2.41. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

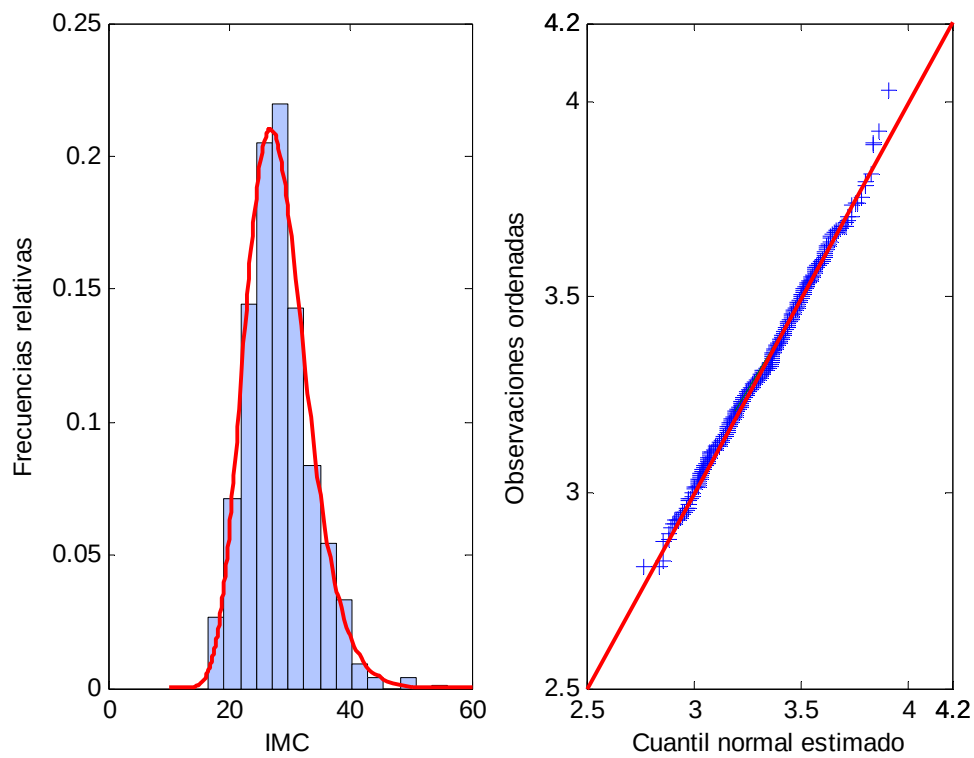


Figura 2.42 Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

NAYARIT

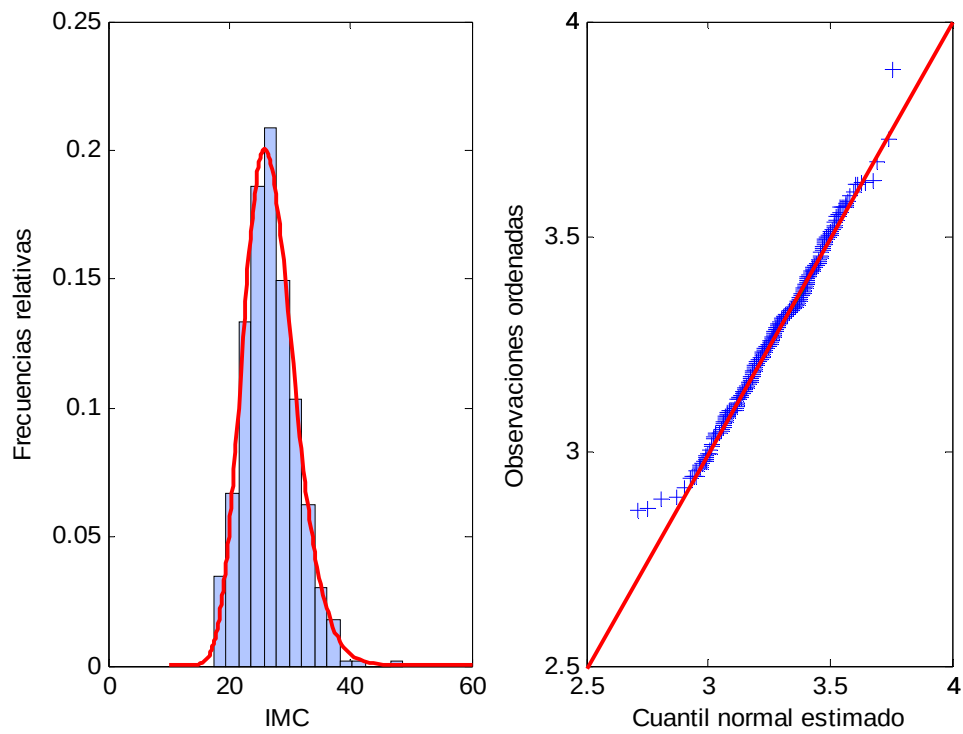


Figura 2.43. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

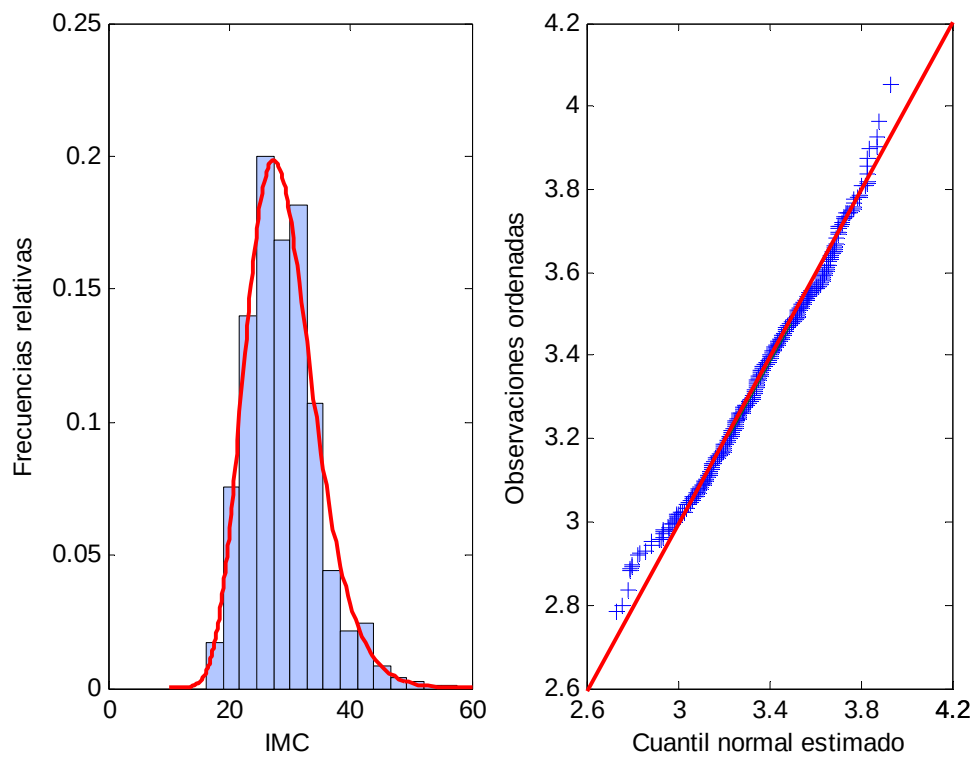


Figura 2.44. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha).

NUEVO LEÓN

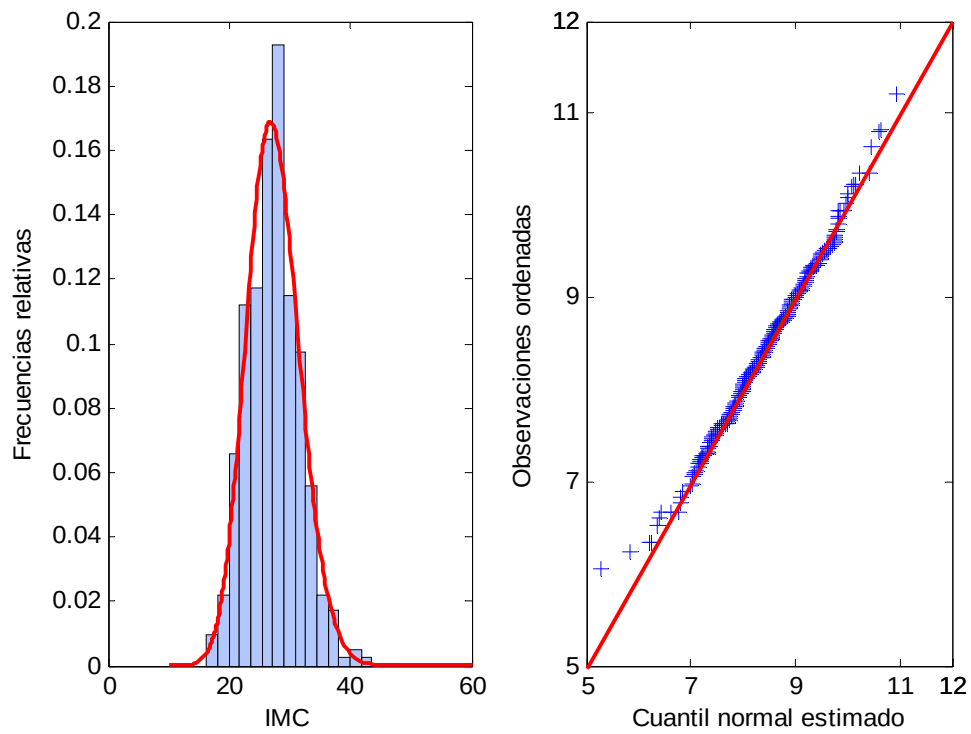


Figura 2.45. Hombres. Ajuste con el Modelo de Box-Cox $\lambda=0.5$ (izquierda) y gráfica QQ (derecha)

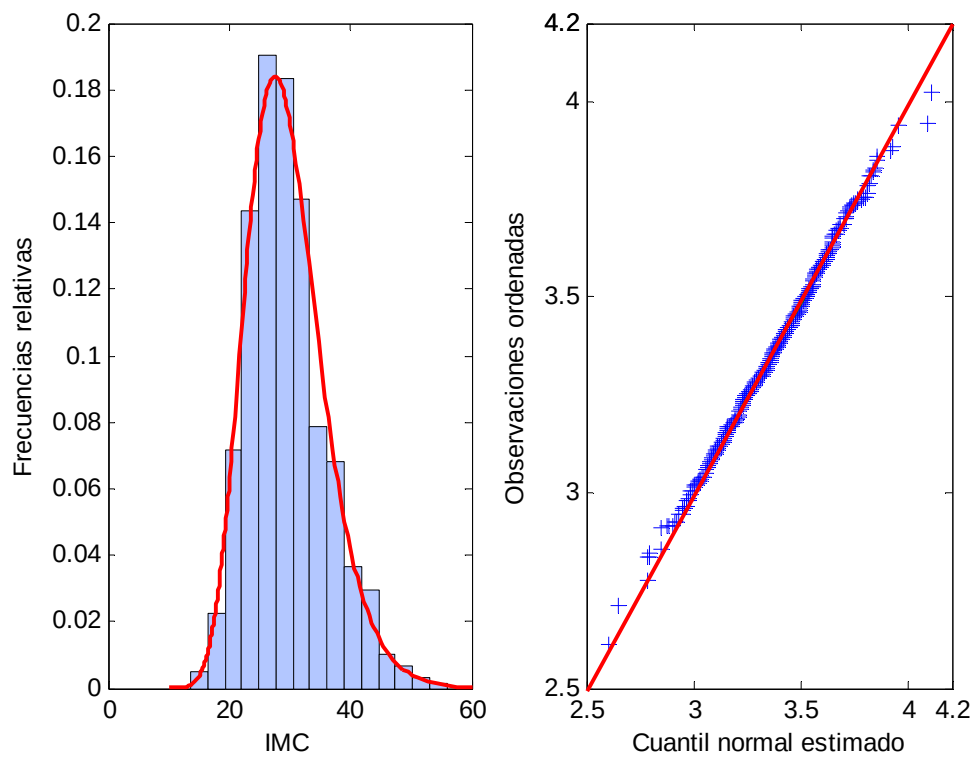


Figura 2.46. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

OAXACA

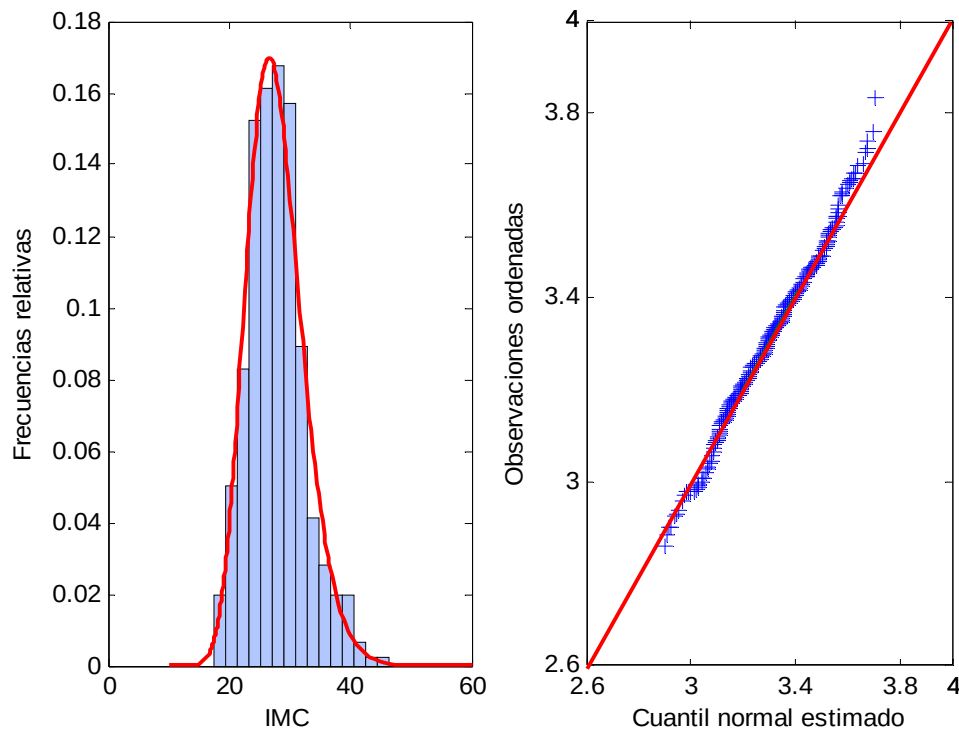


Figura 2.47. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

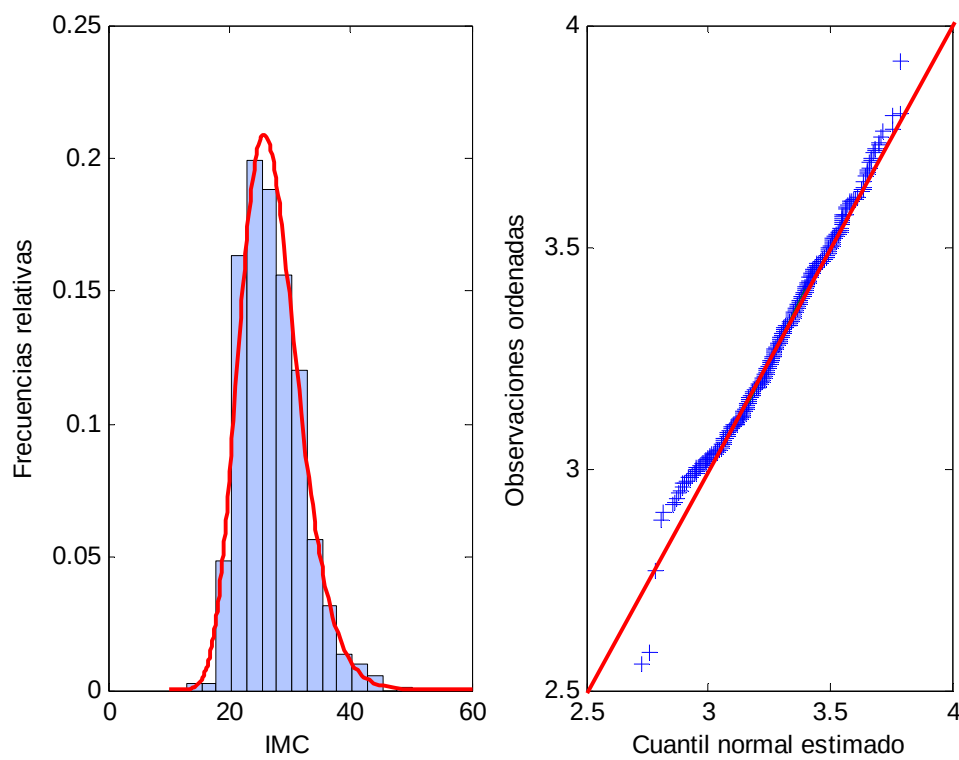


Figura 2.48. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

PUEBLA

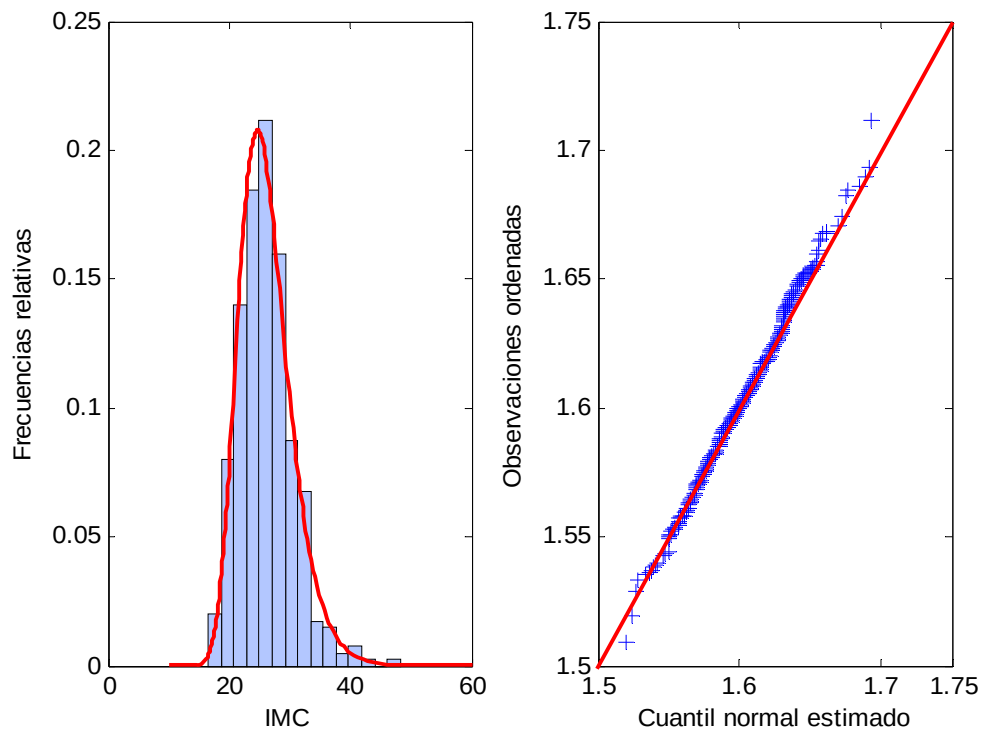


Figura 2.49. Hombres. Ajuste con el Modelo de Box-Cox $\lambda=-0.51$ (izquierda) y gráfica QQ (derecha)

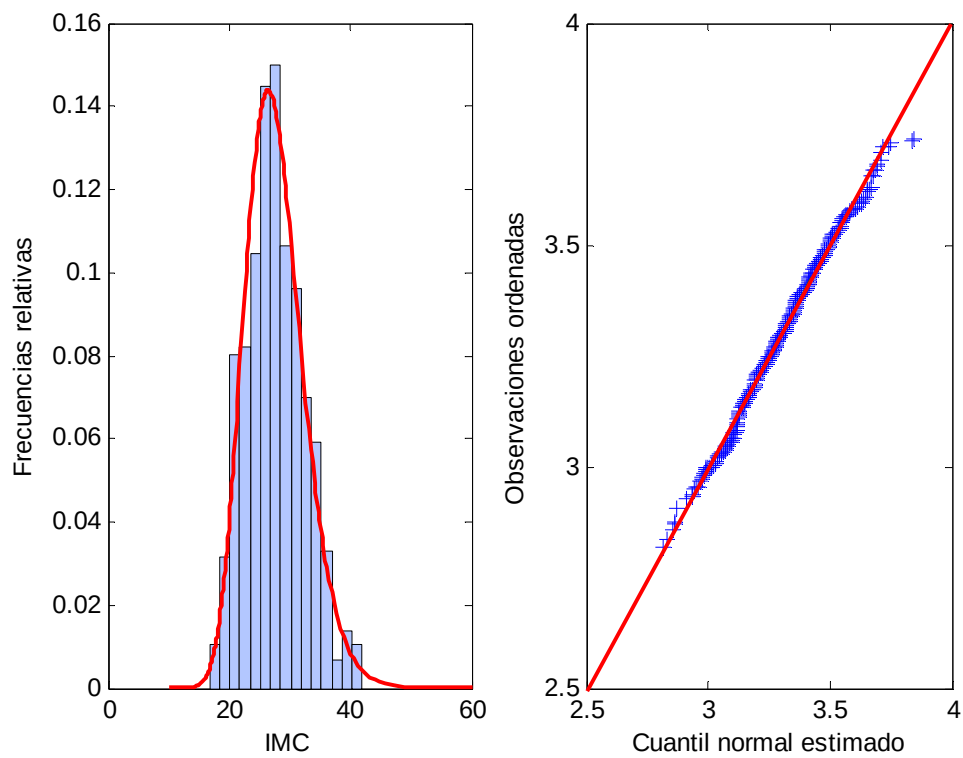


Figura 2.50 Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

QUERÉTARO

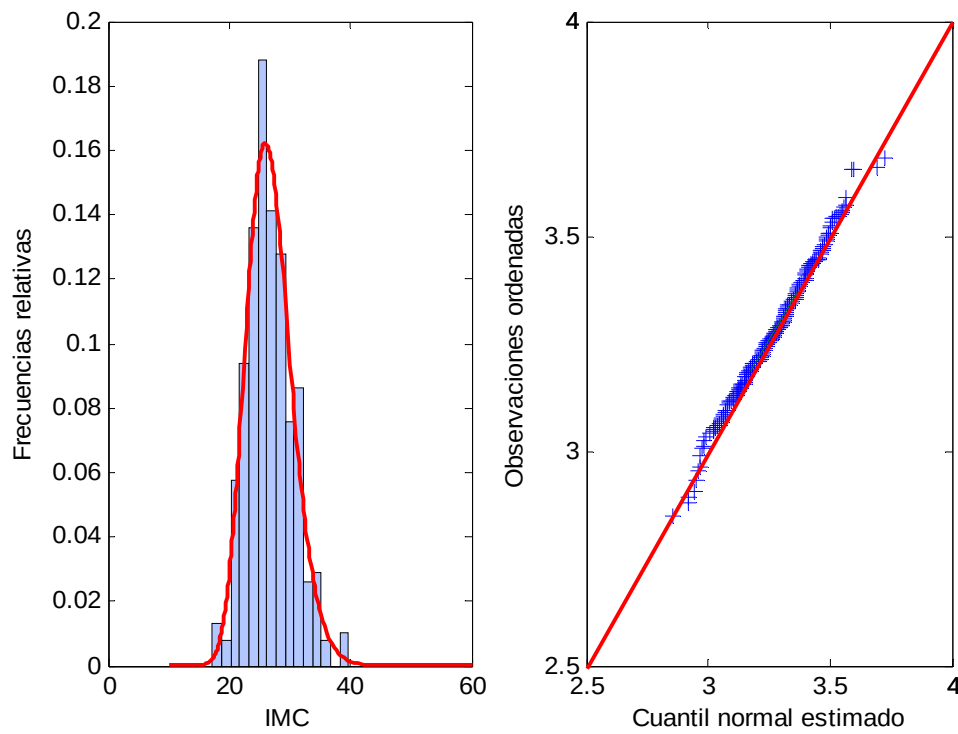


Figura 2.51. Hombres Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

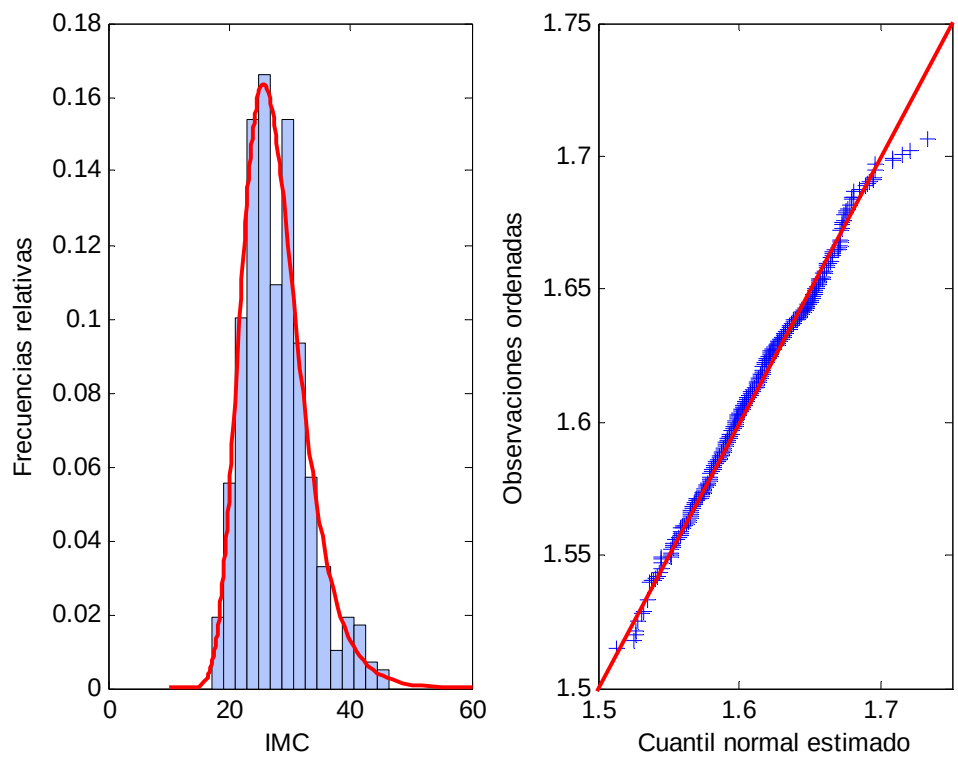


Figura 2.52 Mujeres. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ (izquierda) y gráfica QQ (derecha)

QUINTANA ROO

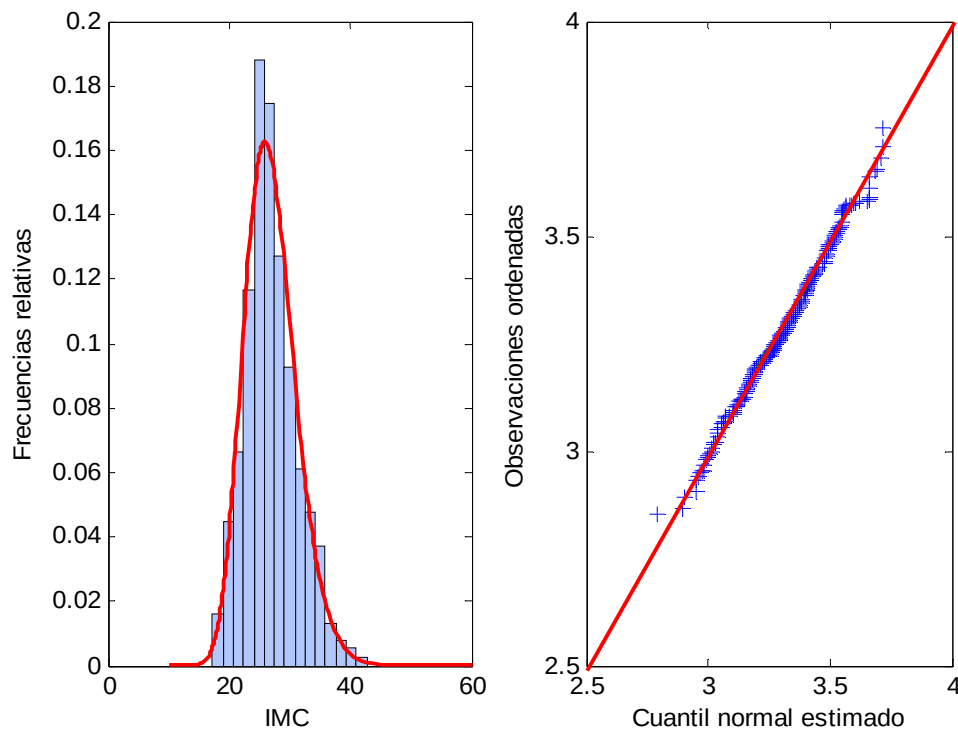


Figura 2.53. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

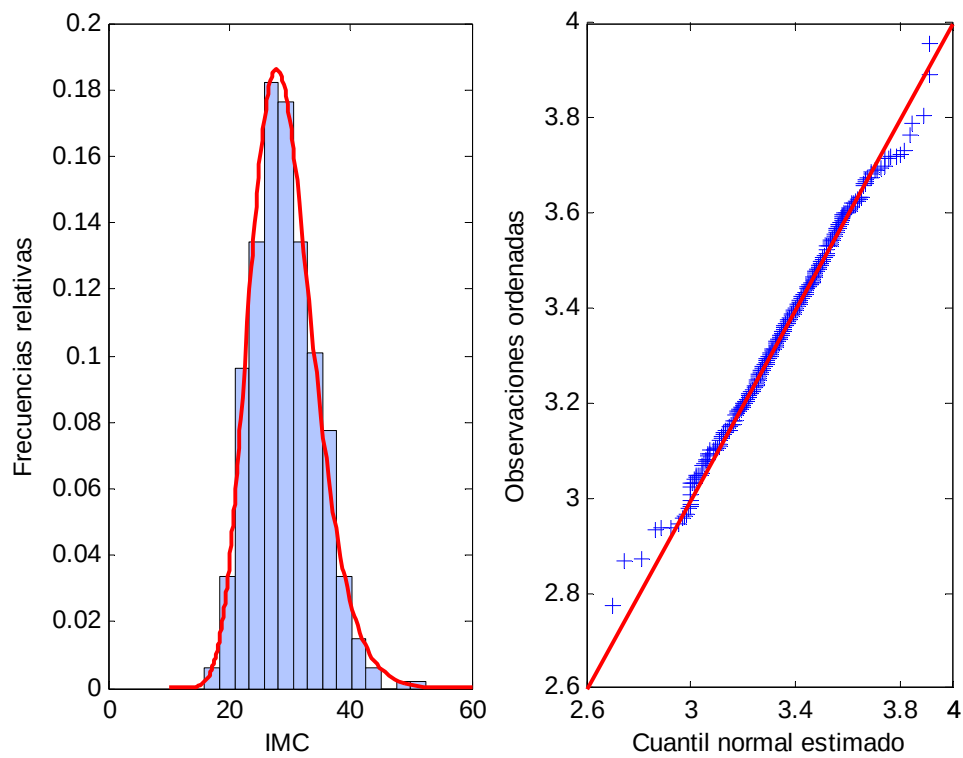


Figura 2.54 Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

SAN LUIS POTOSI

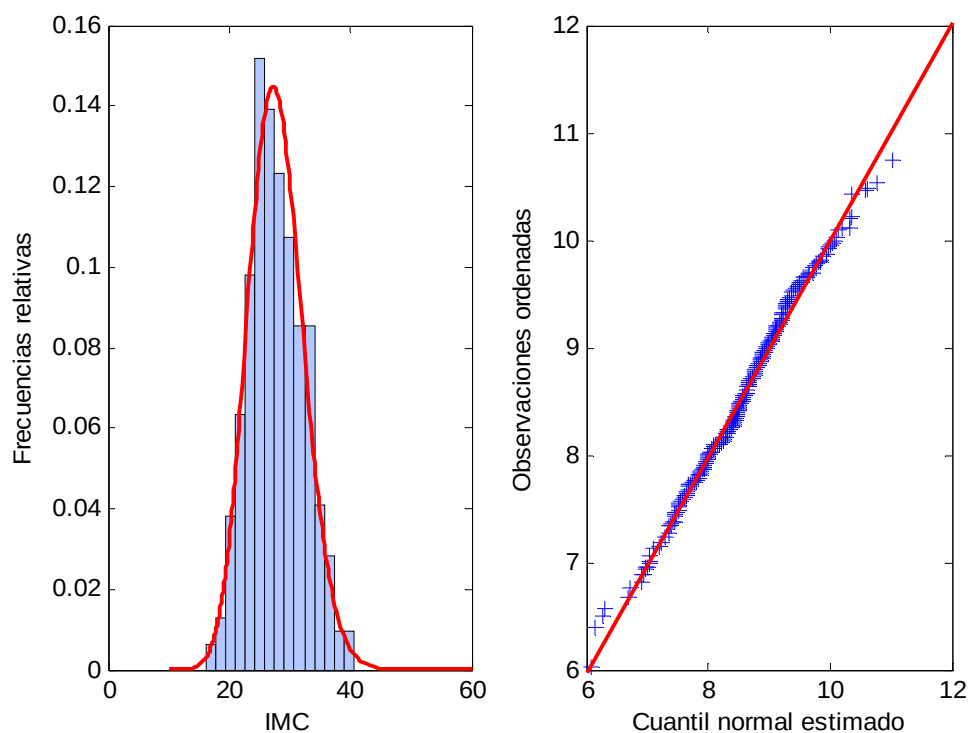


Figura 2.55. Hombres Ajuste con el Modelo de Box-Cox $\lambda=0.5$ (izquierda) y gráfica QQ (derecha)

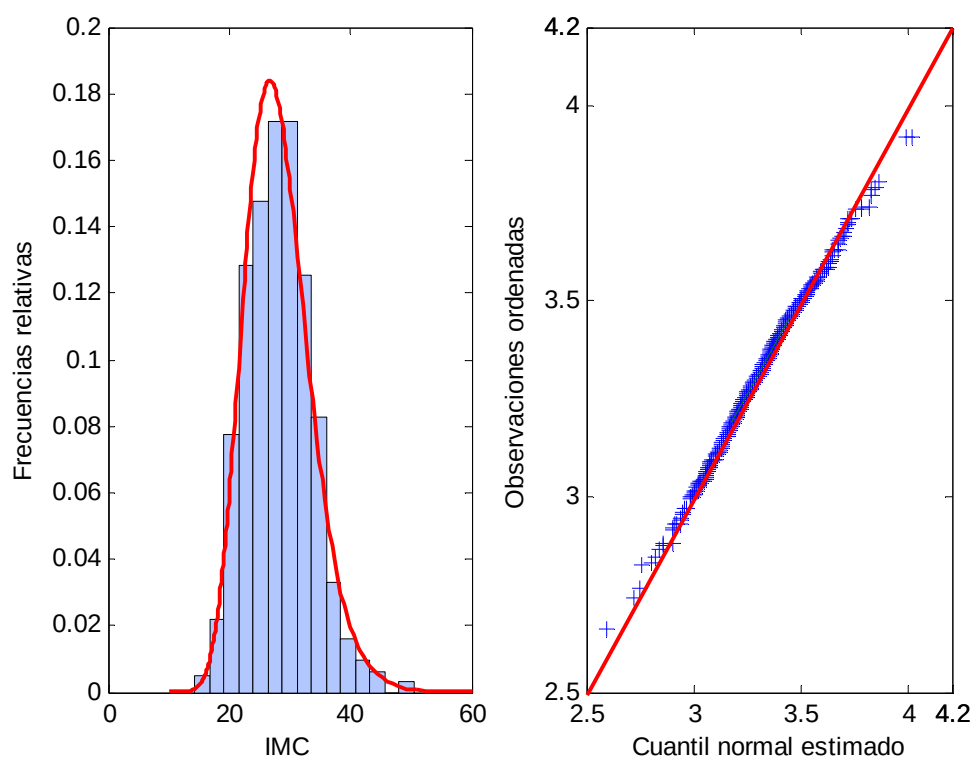


Figura 2.56 Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

SINALOA

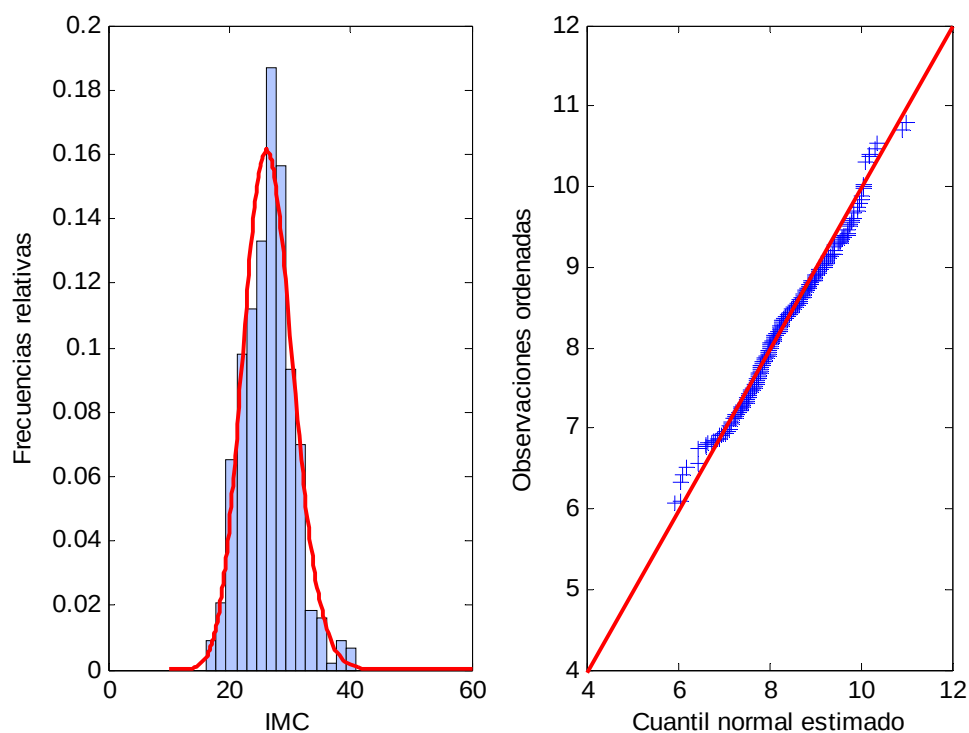


Figura 2.57. Hombres. Ajuste con el Modelo de Box-Cox $\lambda=0.5$ (izquierda) y gráfica QQ (derecha)

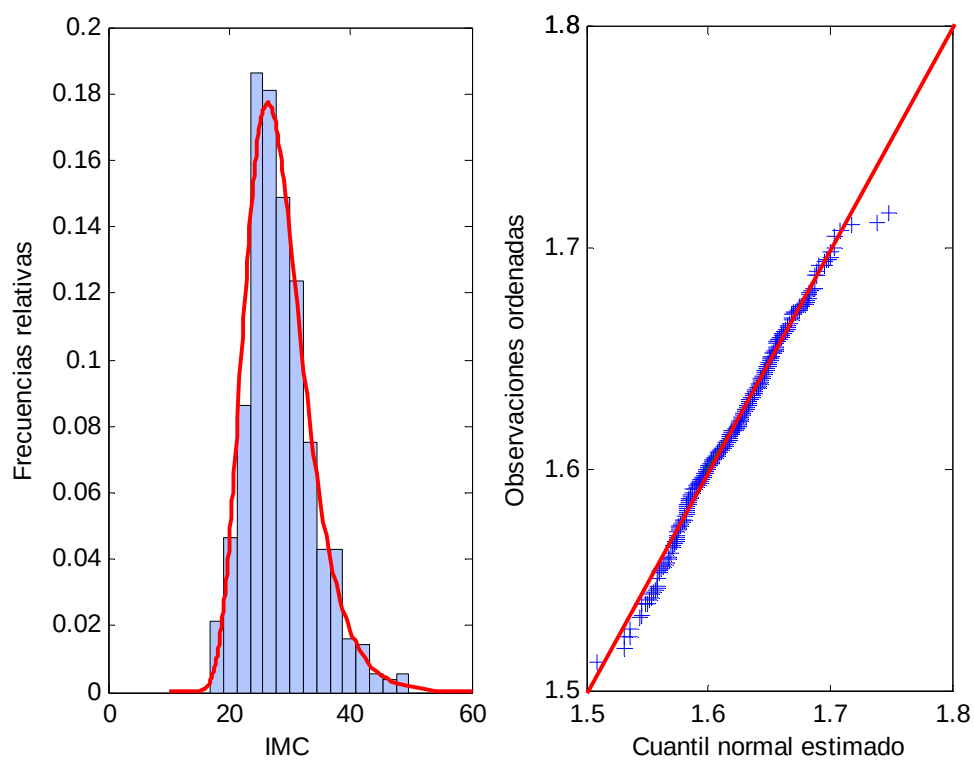


Figura 2.58. Mujeres. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ (izquierda) y gráfica QQ (derecha)

SONORA

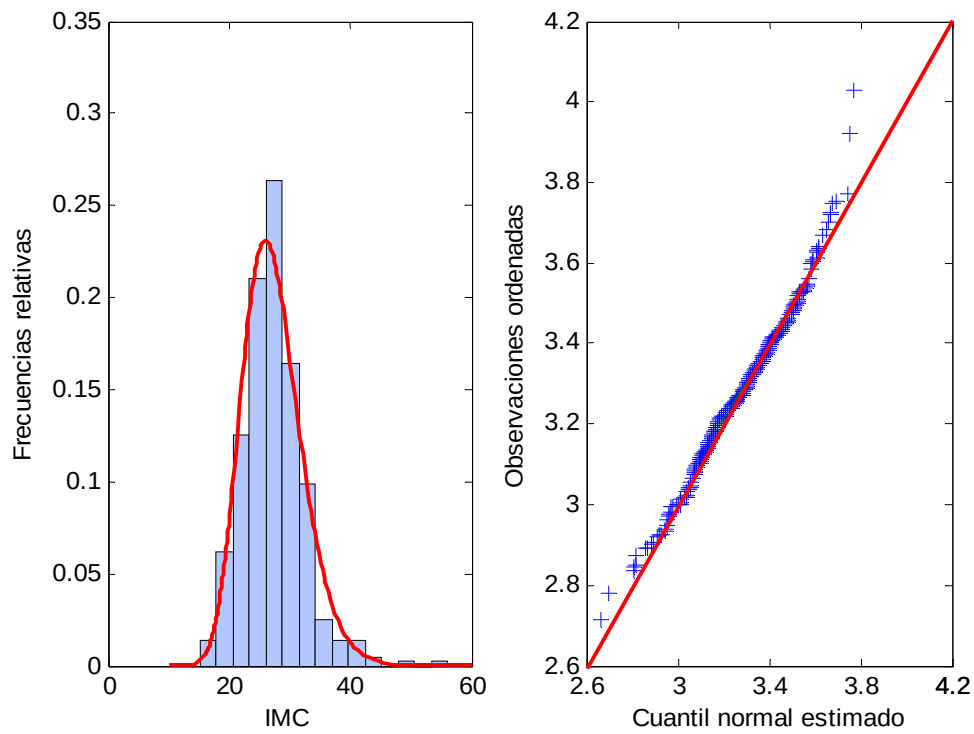


Figura 2.59. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

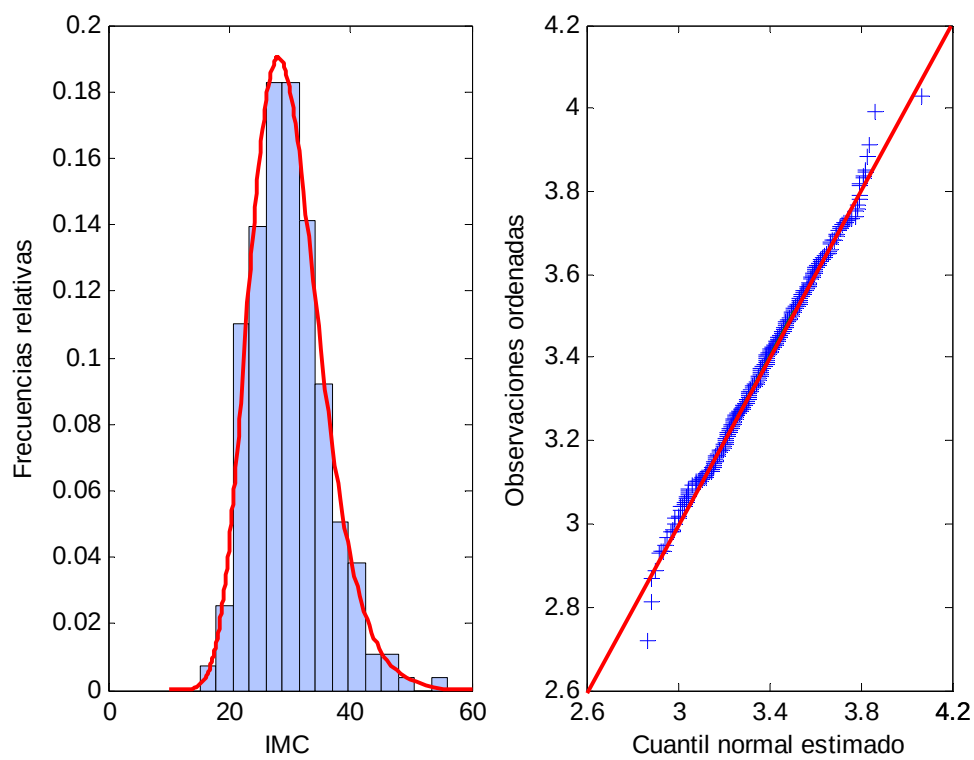


Figura 2.60. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

TABASCO

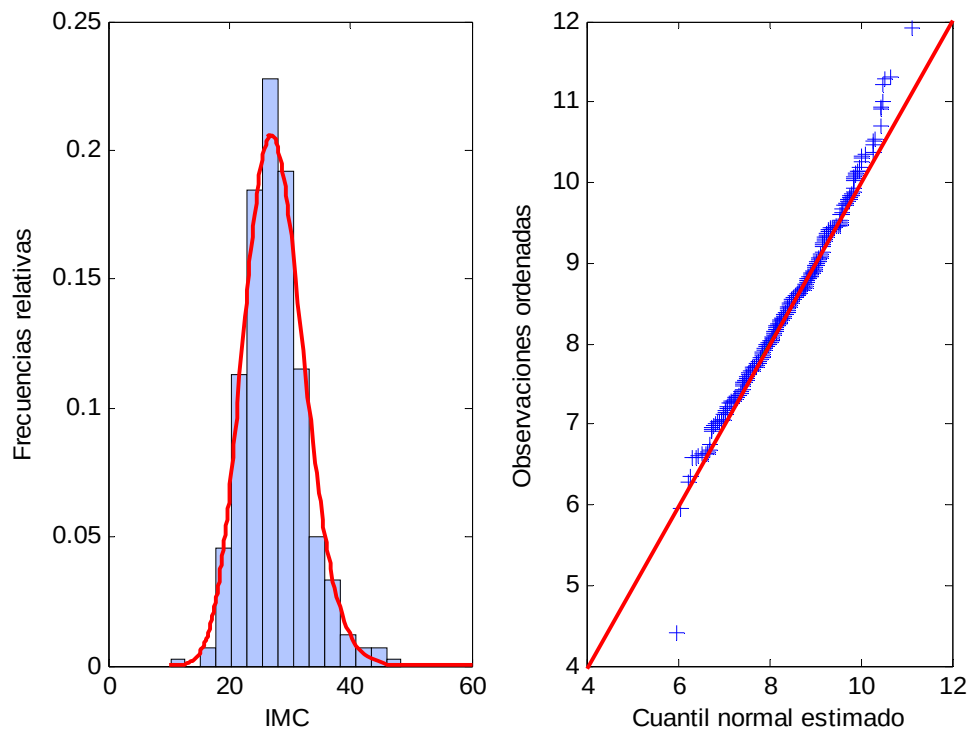


Figura 2.61. Hombres Ajuste con el Modelo de Box-Cox $\lambda=0.5$ (izquierda) y gráfica QQ (derecha)

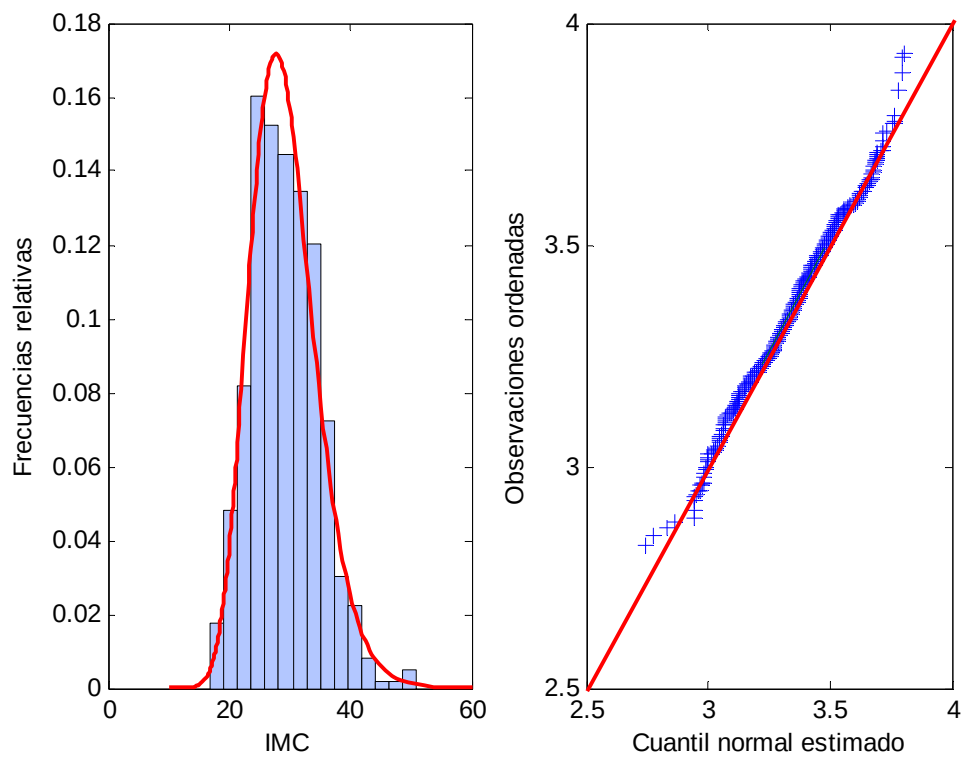


Figura 2.62. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

TAMAULIPAS

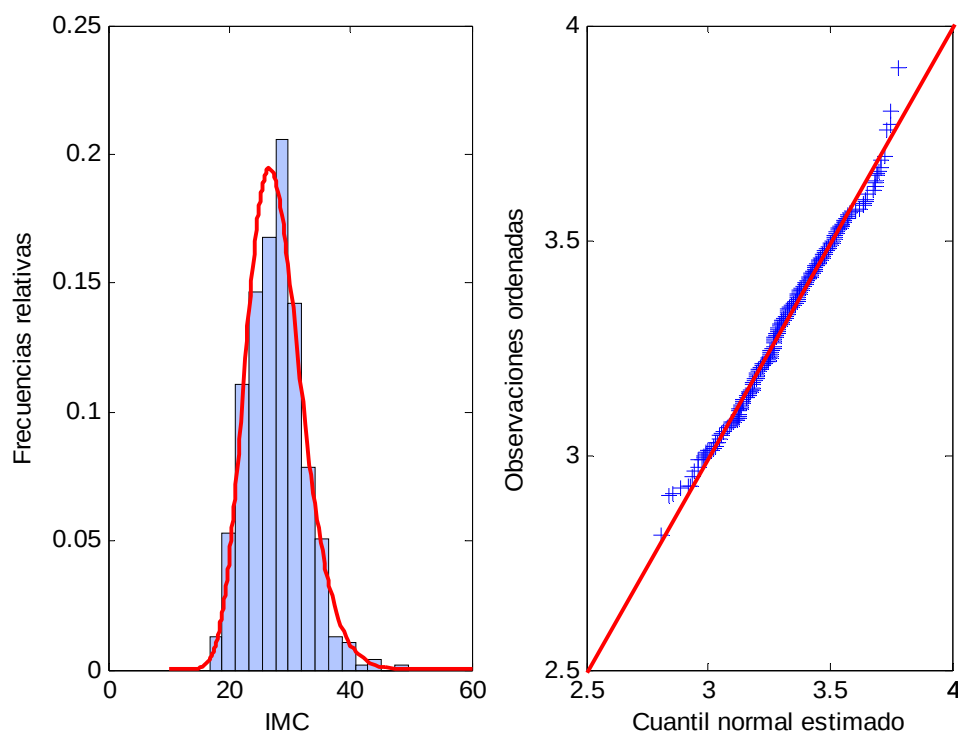


Figura 2.63. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

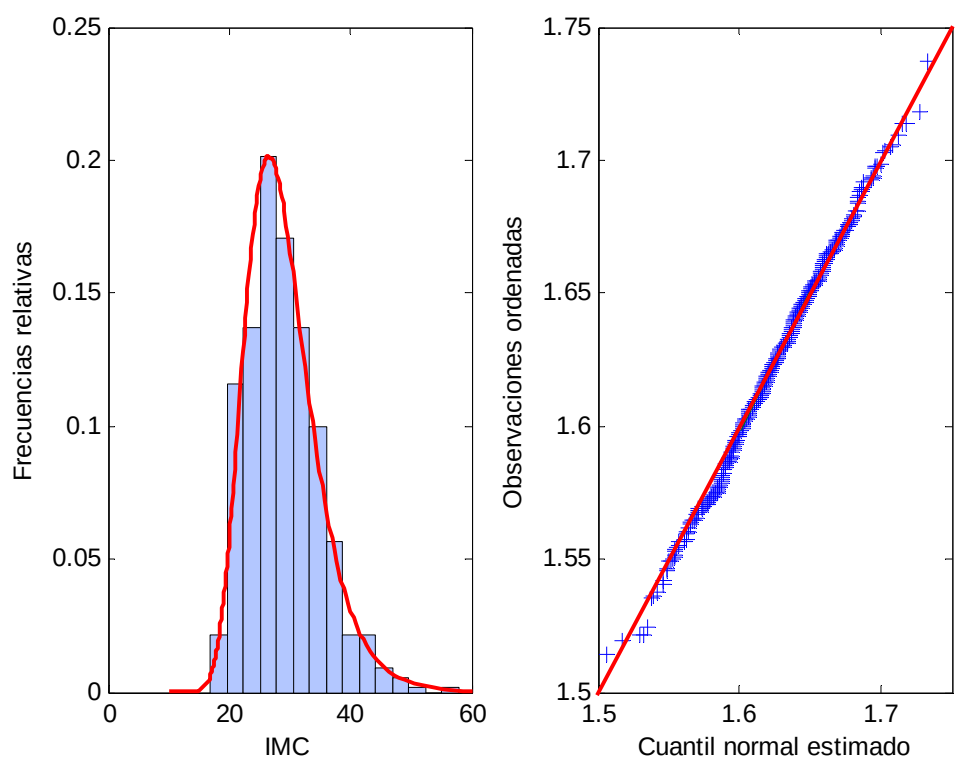


Figura 2.64. Mujeres. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ (izquierda) y gráfica QQ (derecha)

TLAXCALA

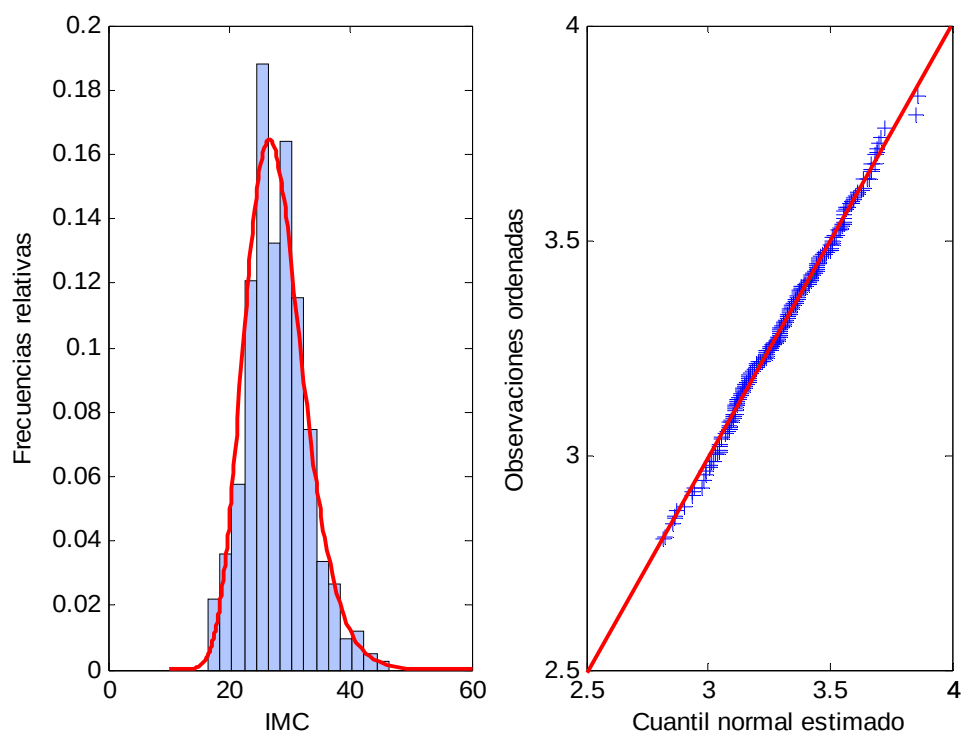


Figura 2.65. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

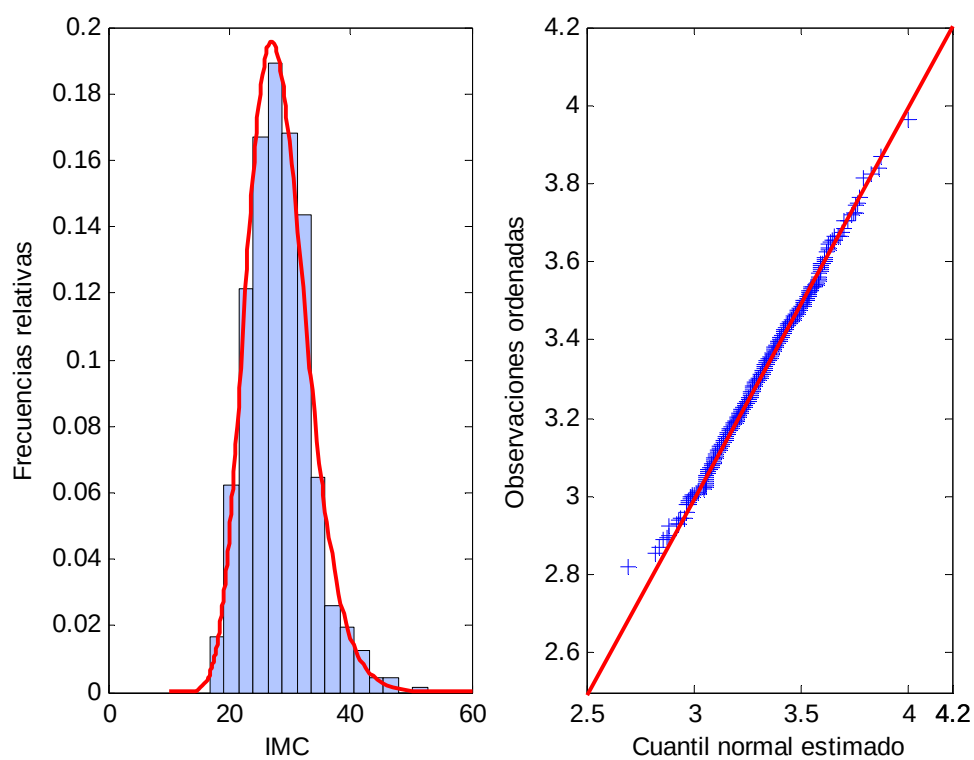


Figura 2.66. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

VERACRUZ

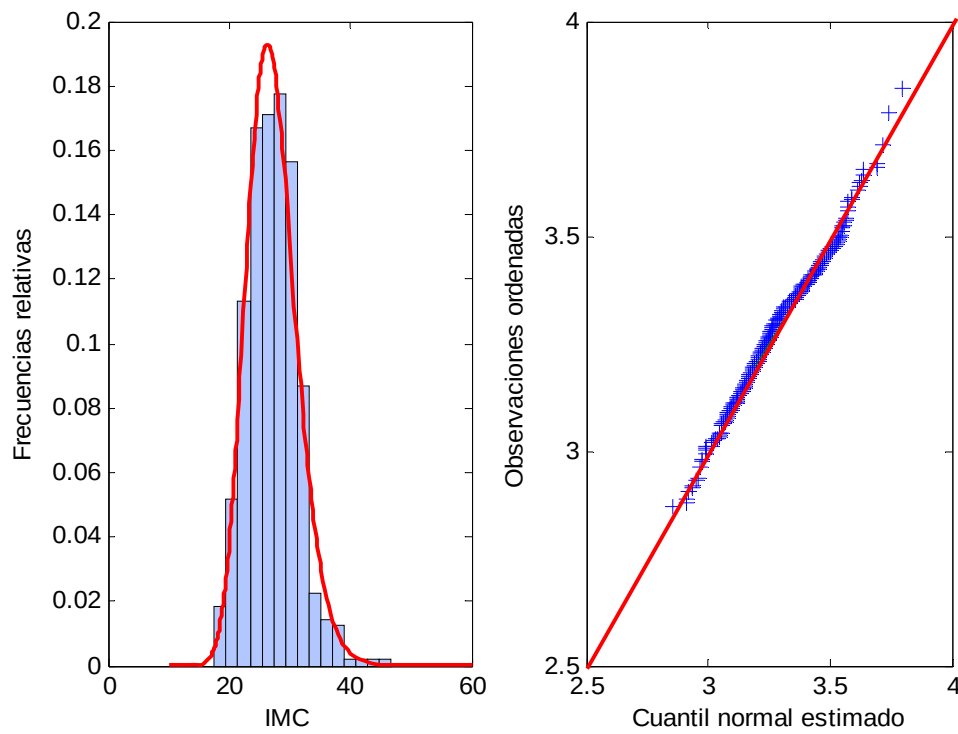


Figura 2.67. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

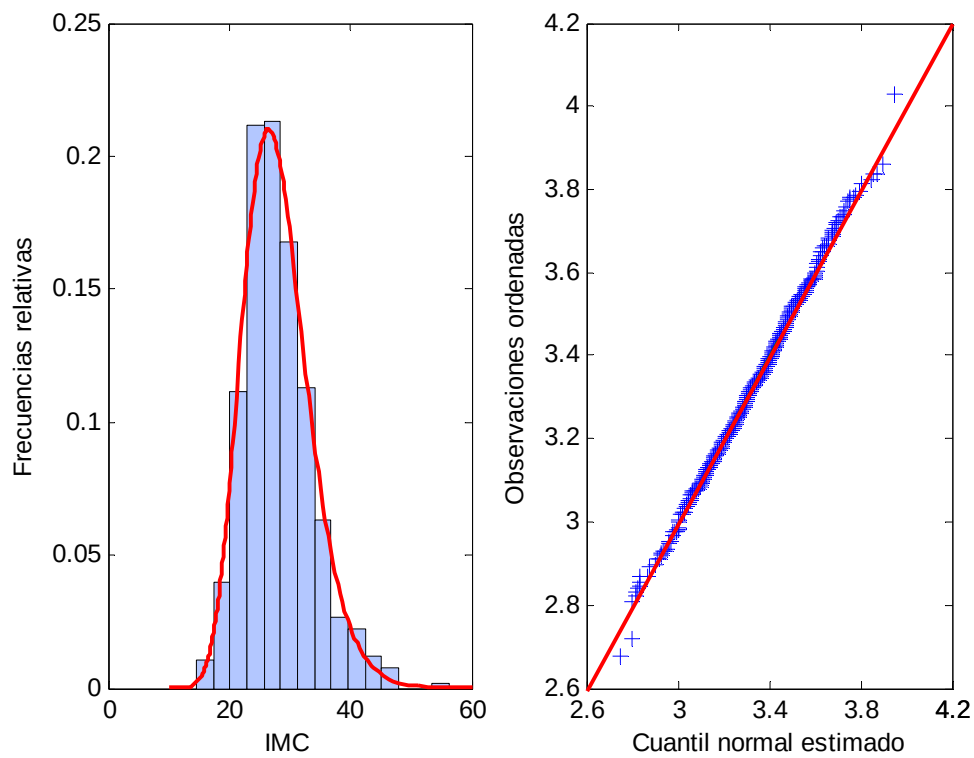


Figura 2.68. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

YUCATÁN

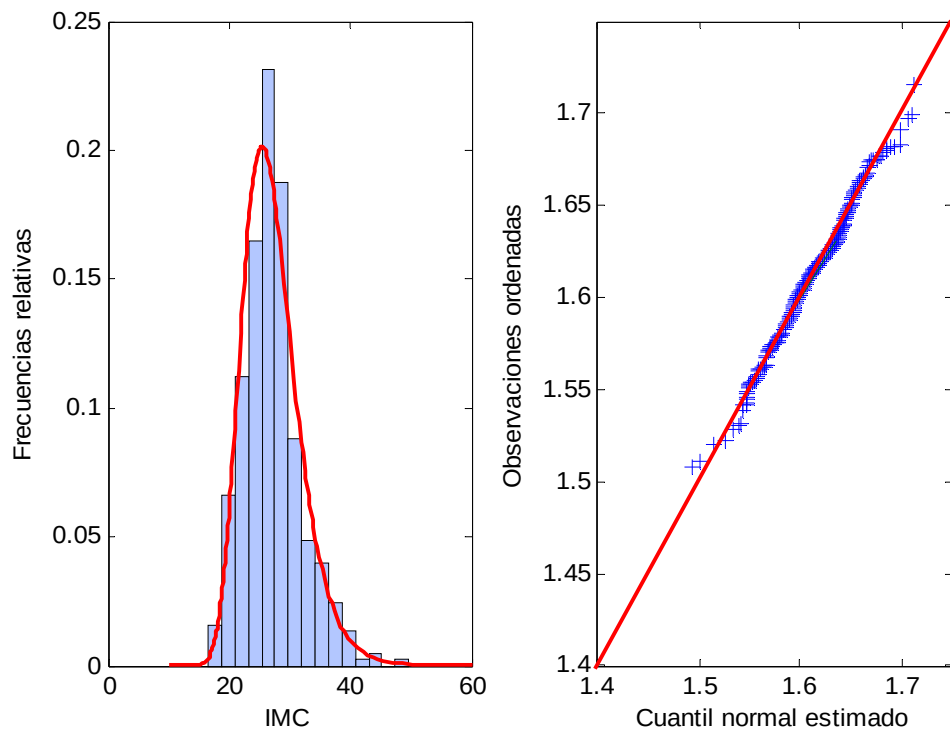


Figura 2.69. Hombres. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ (izquierda) y gráfica QQ (derecha)

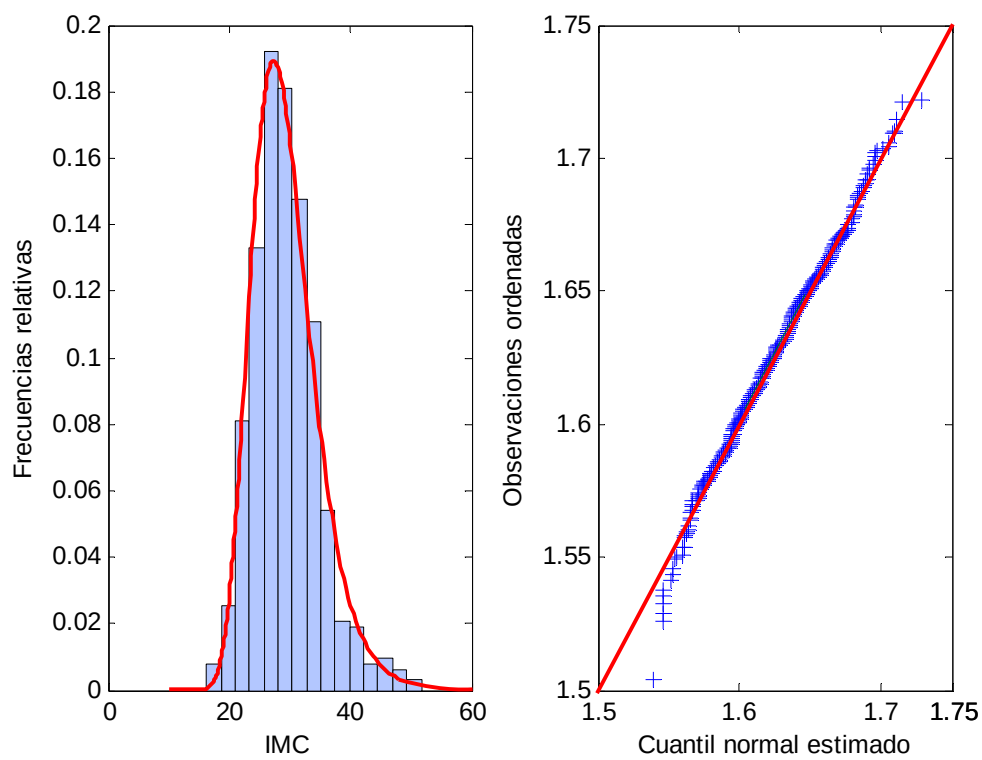


Figura 2.70. Mujeres. Ajuste con el Modelo de Box-Cox $\lambda=-0.5$ (izquierda) y gráfica QQ (derecha)

ZACATECAS

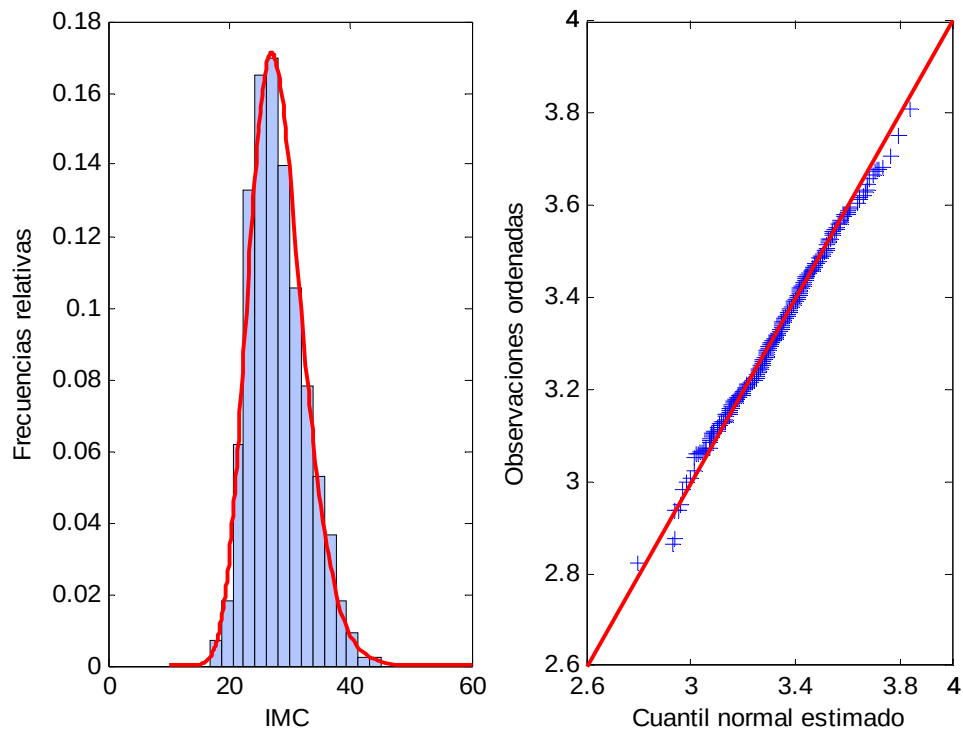


Figura 2.71. Hombres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

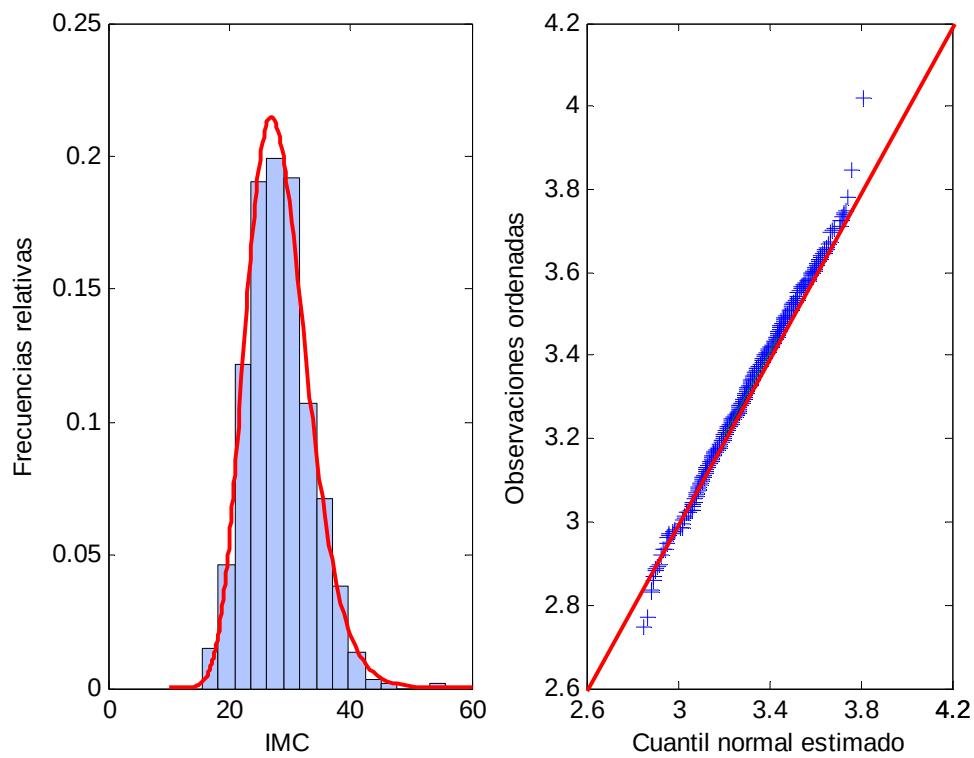


Figura 2.72. Mujeres. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

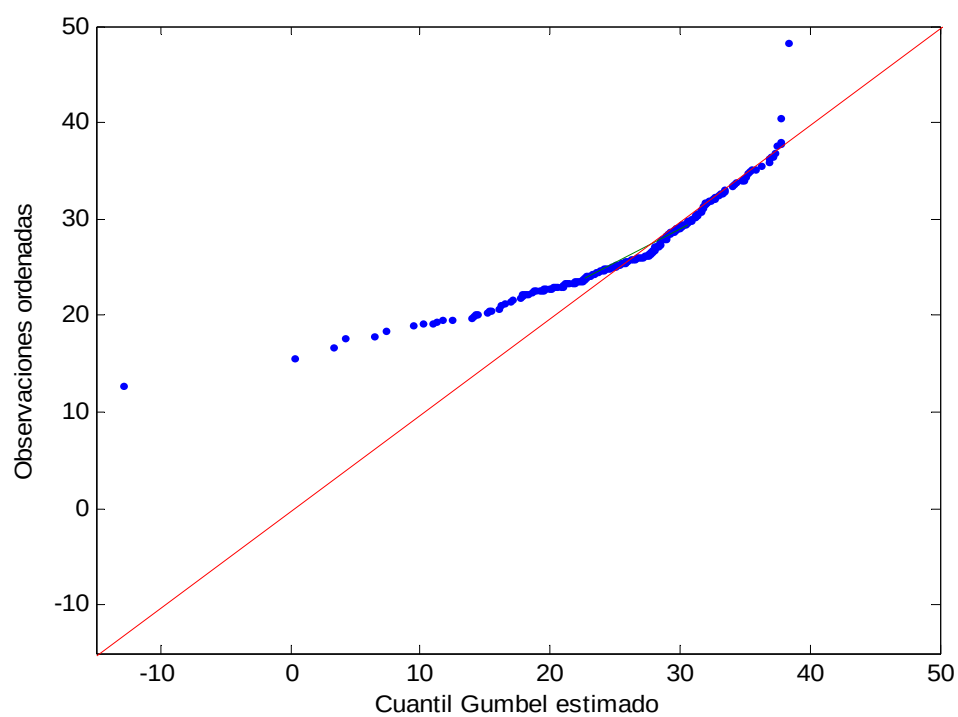


Figura 2.73. Hombres de Aguascalientes.
Ajuste con el modelo Gumbel de parámetros (29.11 5.21)

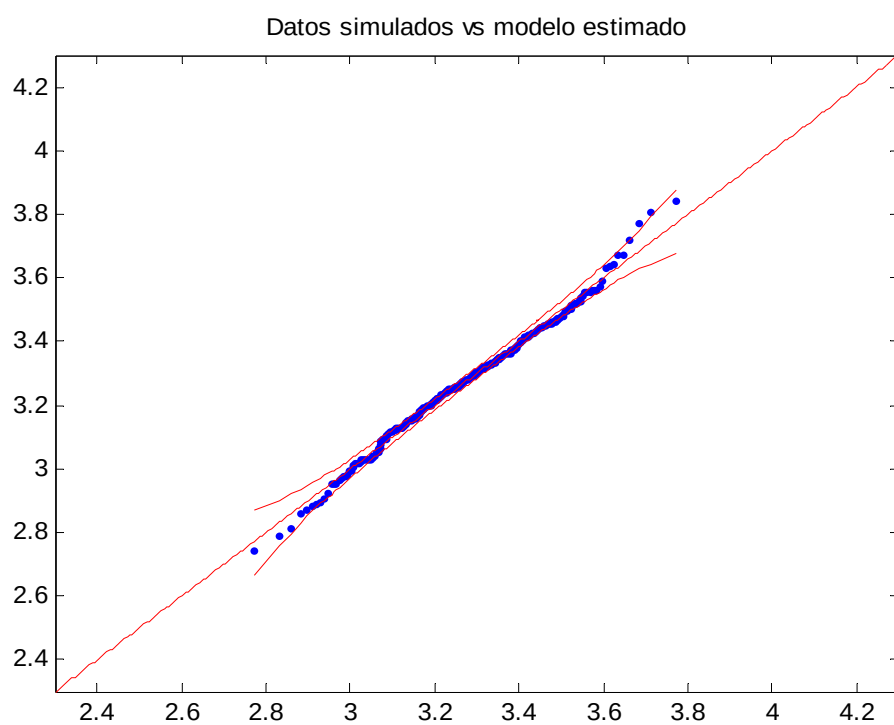


Figura 2.74. Aguascalientes hombres. Modelo Lognormal

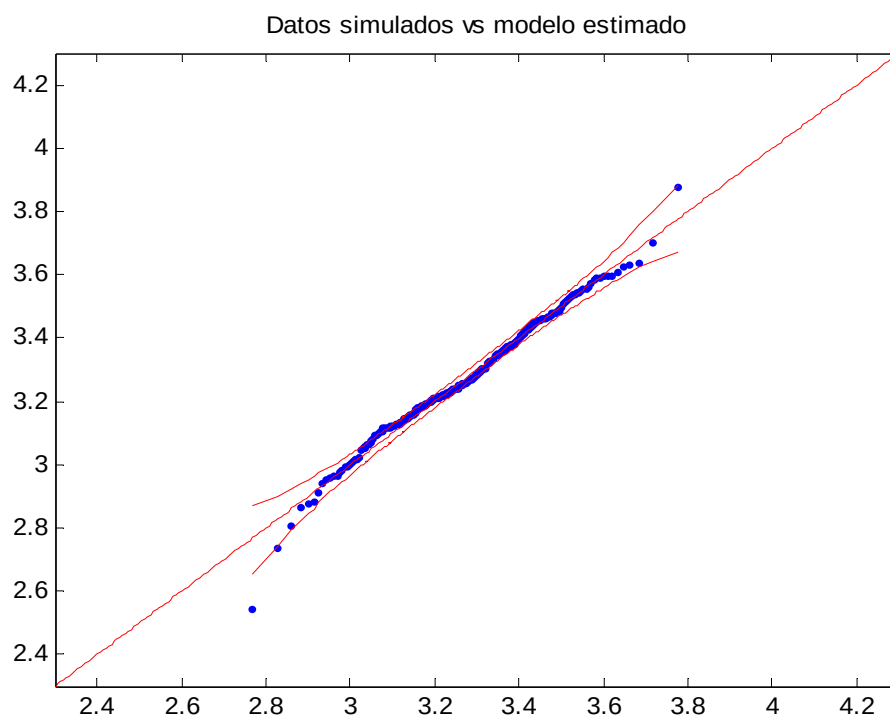


Figura 2.75. Baja California. Hombres. Modelo Lognormal

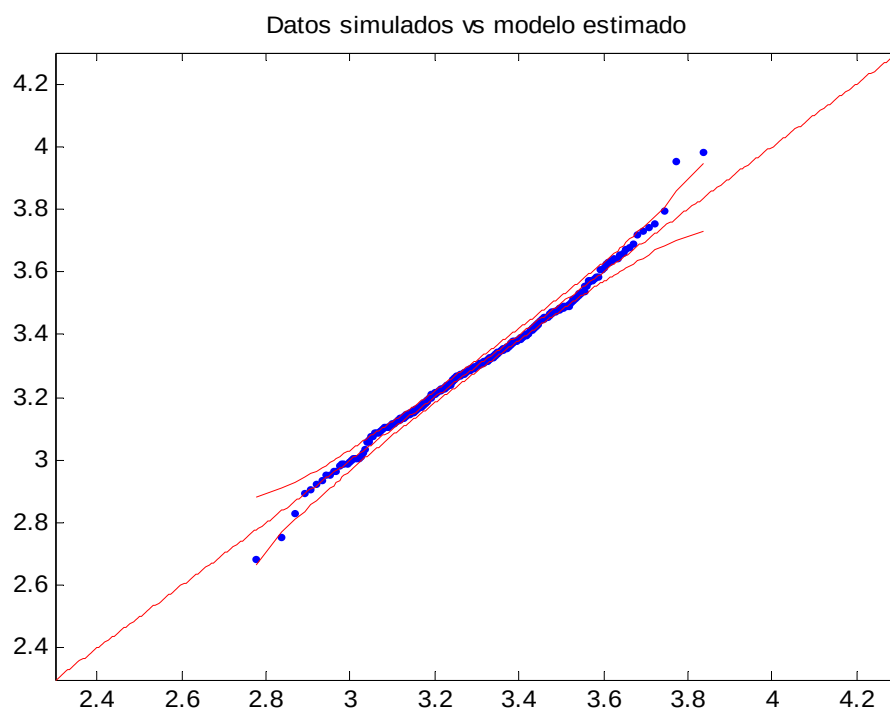


Figura 2.76. Campeche. Hombres. Modelo Lognormal

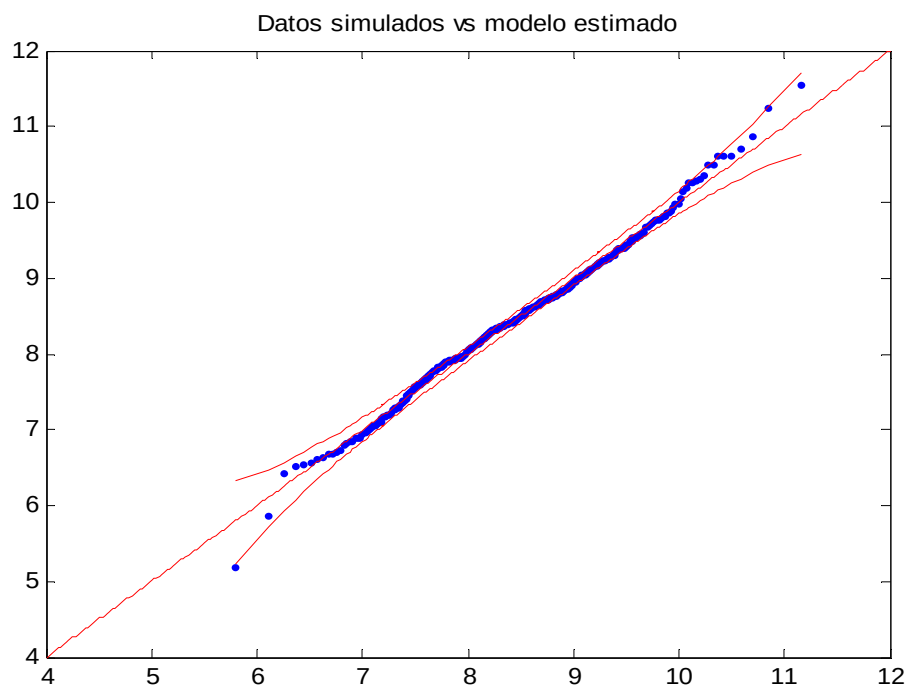


Figura 2.77. Coahuila. Hombres. Modelo de Box-Cox con $\lambda=0.5$

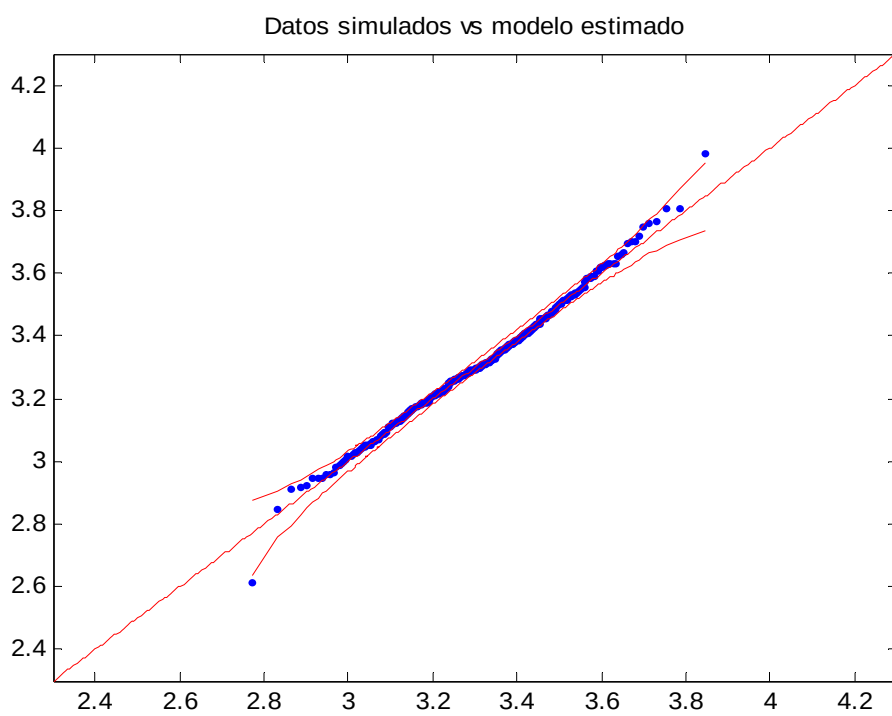


Figura 2.78. Colima. Hombres. Modelo Lognormal

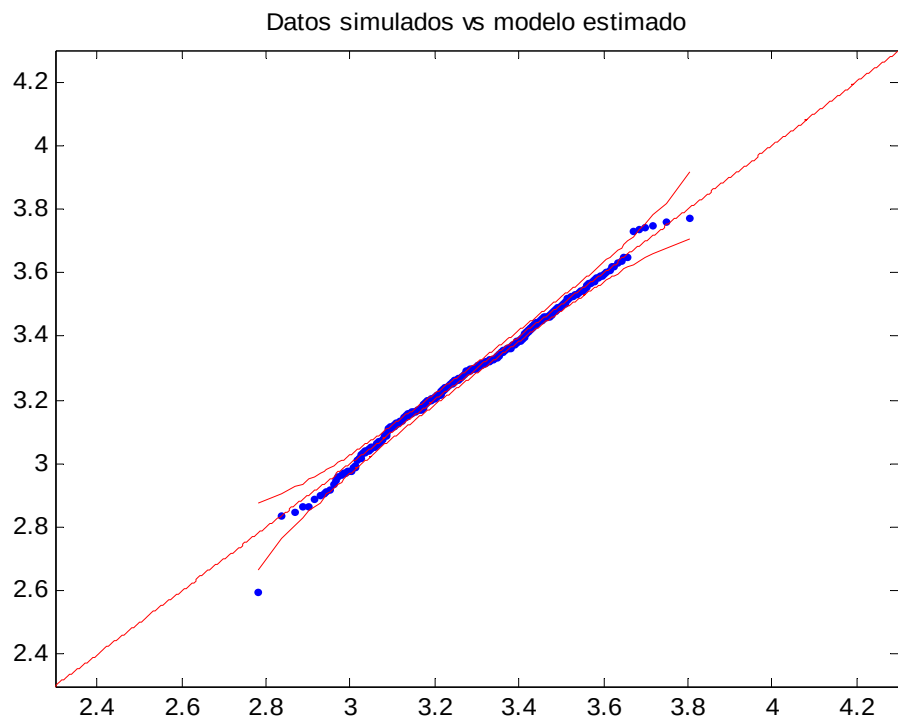


Figura 2.79. Chihuahua. Hombres. Modelo Lognormal

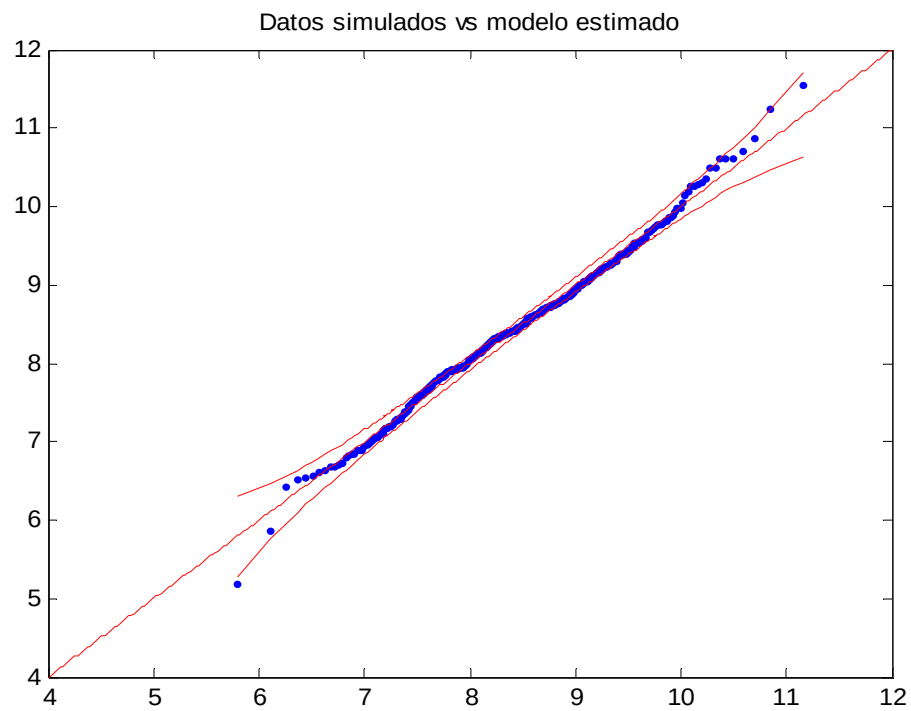


Figura 2.80. Distrito Federal. Hombres. Modelo de Box-Cox con $\lambda=0.5$

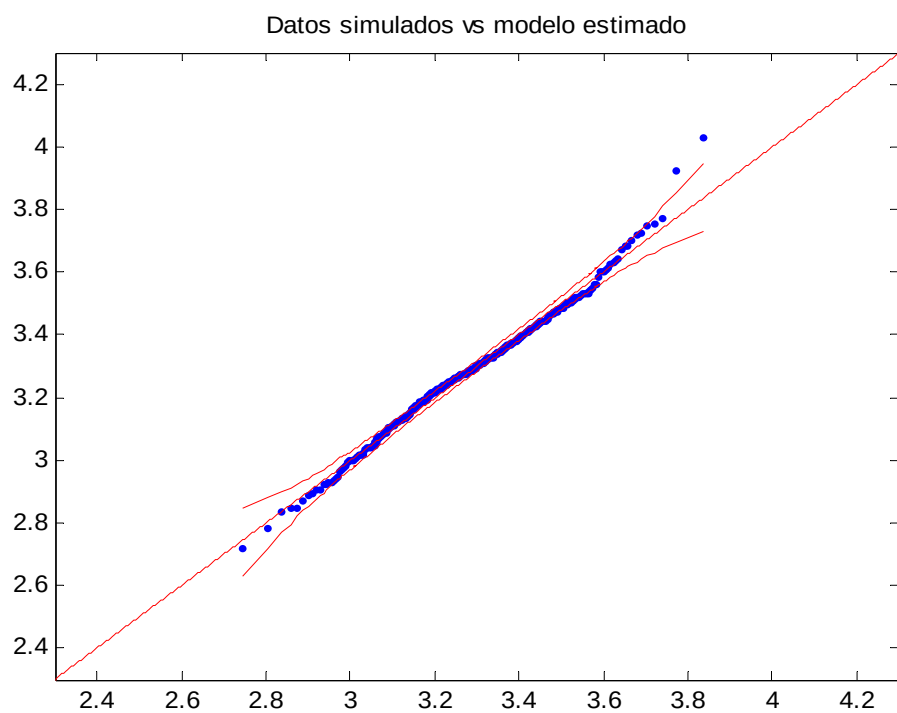


Figura 2.81. Sonora. Hombres. Modelo Lognormal

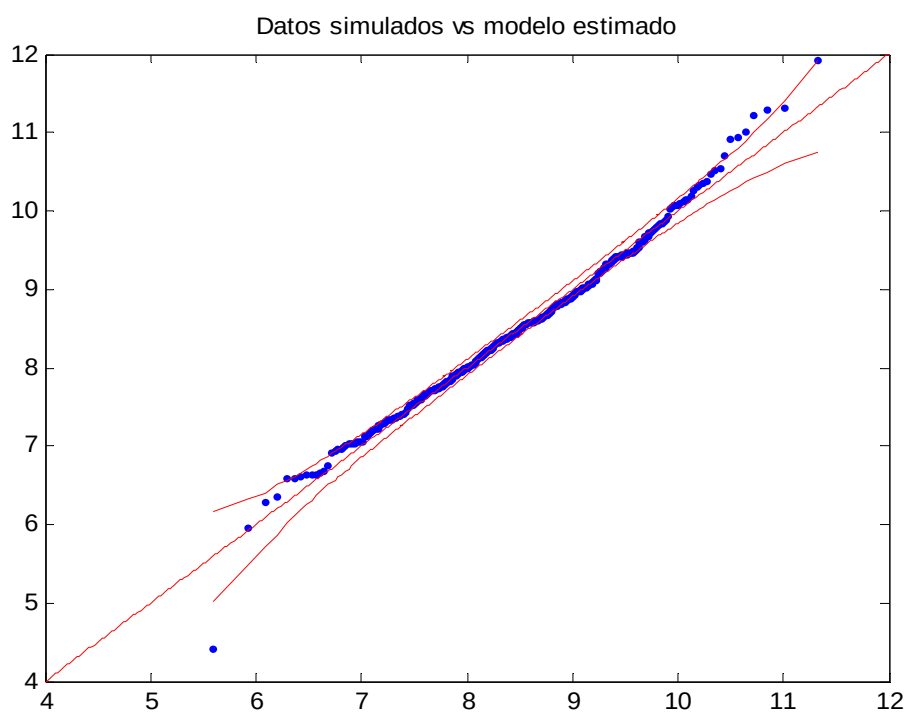


Figura 2.82. Tabasco. Hombres. Modelo de Box-Cox con $\lambda=0.5$

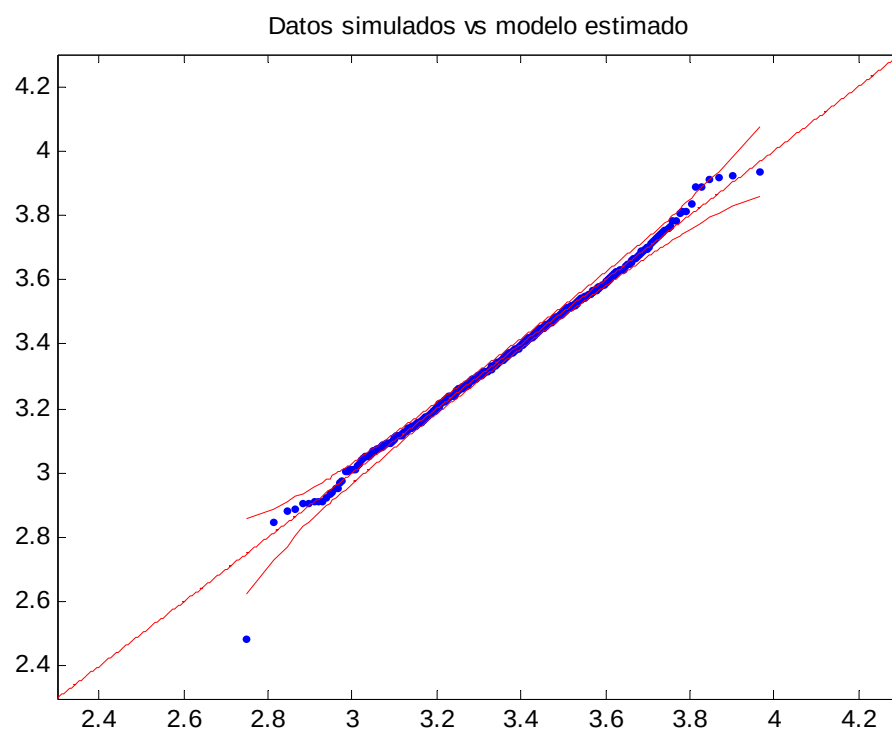


Figura 2.83. Chihuahua. Mujeres. Modelo Lognormal

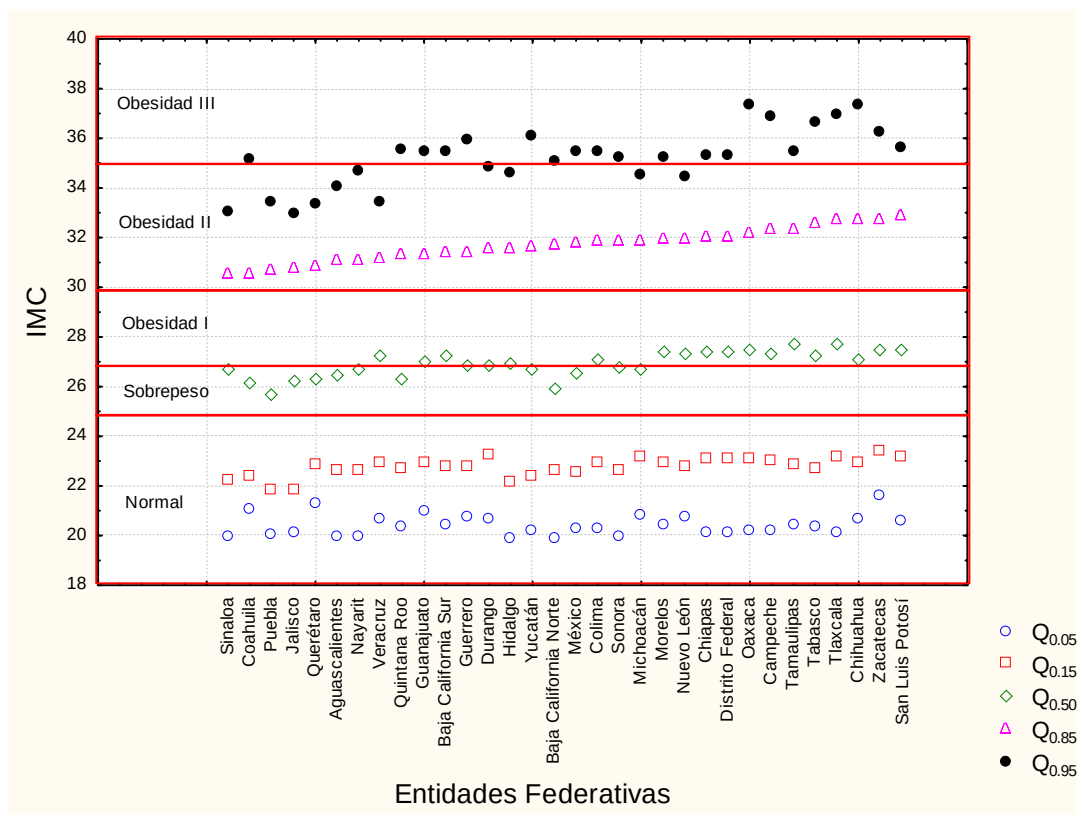


Figura 2.84. Cuantiles estimados de hombres por estado. Ordenados respecto de $Q_{0.85}$

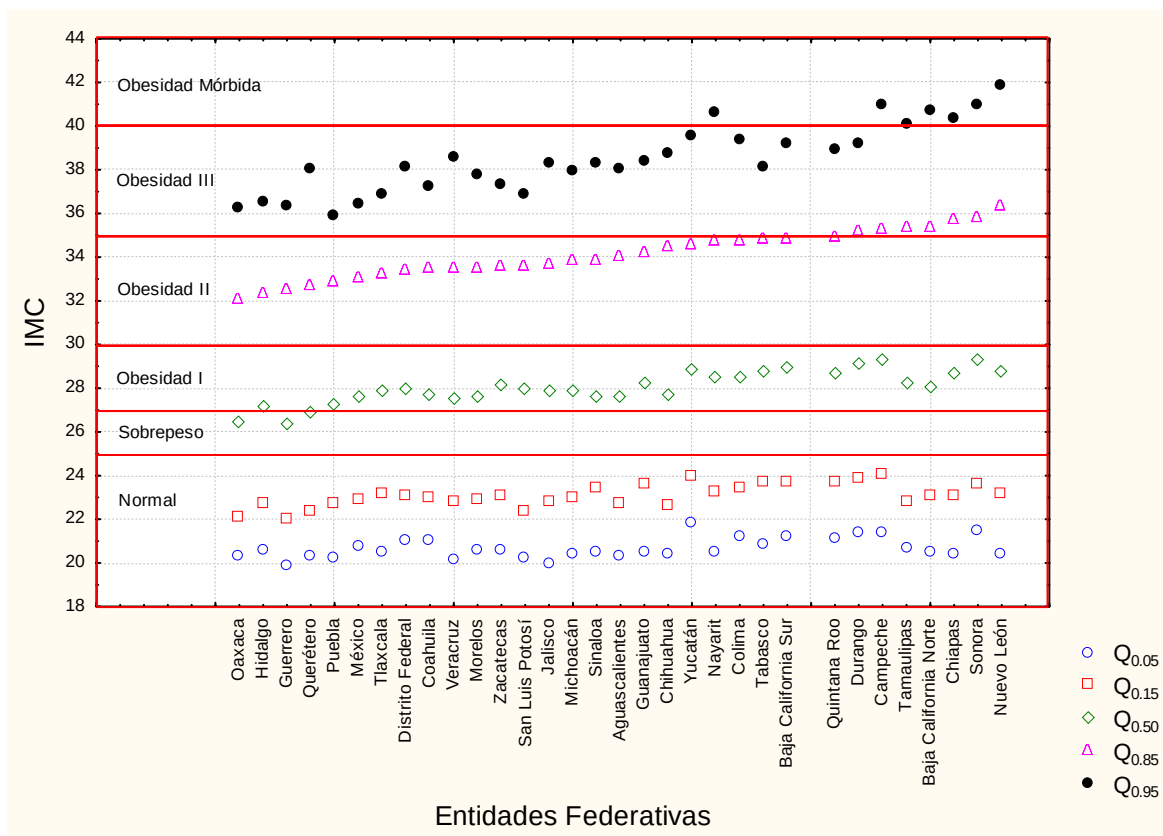


Figura 2.85. Cuantiles estimados de mujeres por estado. Ordenados respecto de Q_{0.85}

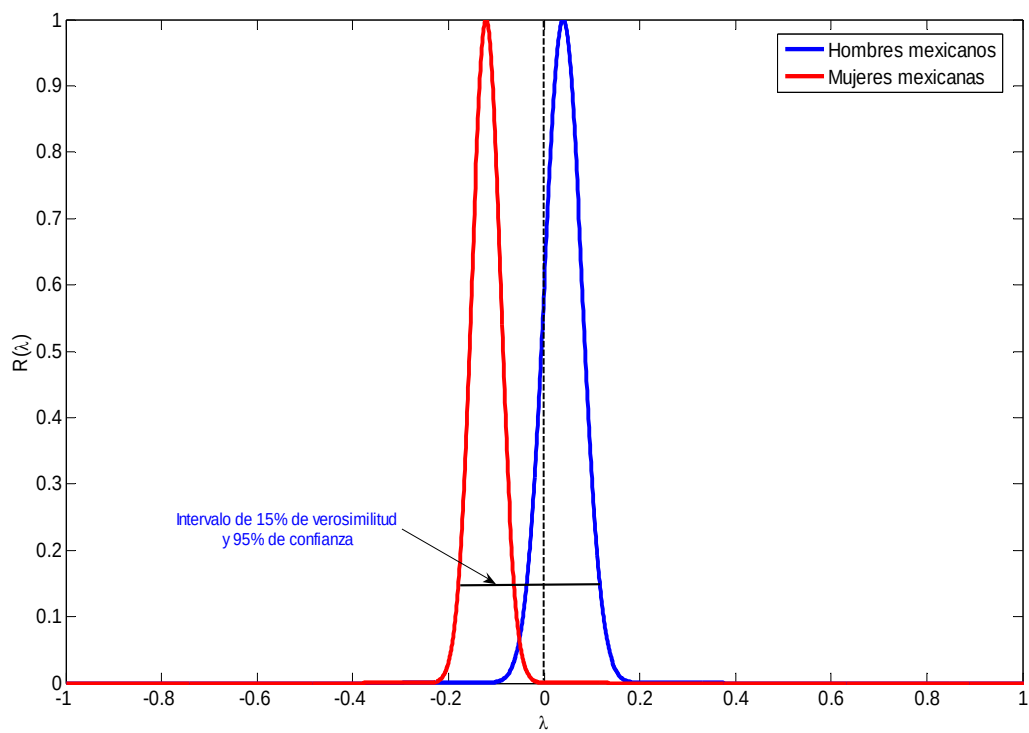


Figura 2.86. Verosimilitud relativa de adultos mexicanos

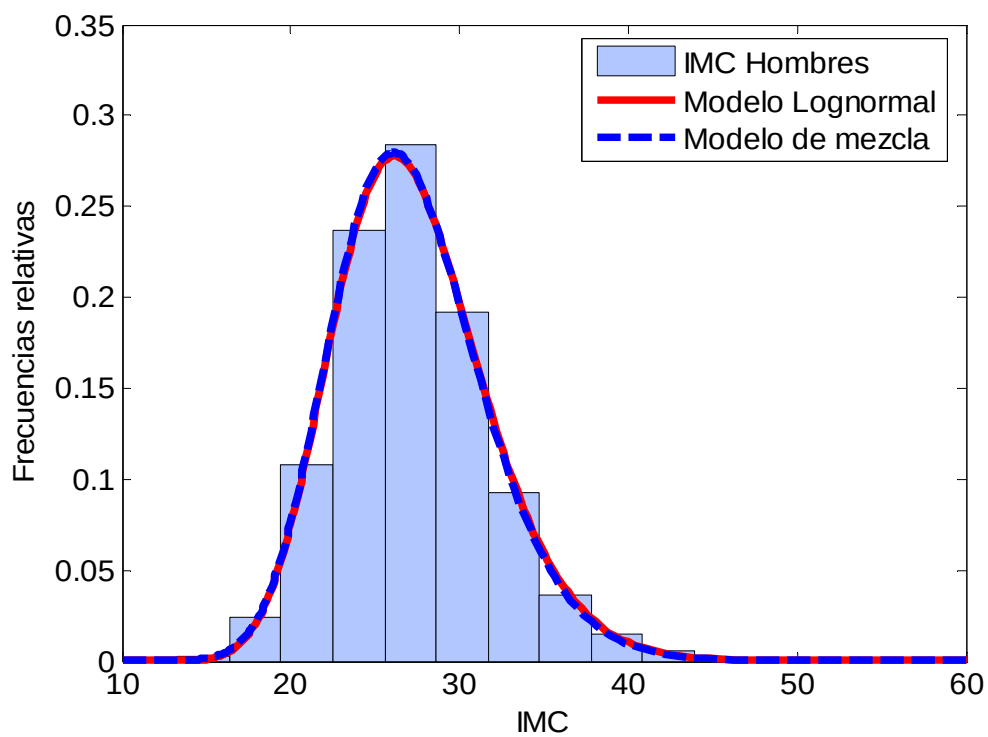


Figura 2.87. Ajuste con el Modelo de Mezcla y el modelo Lognormal. Hombres

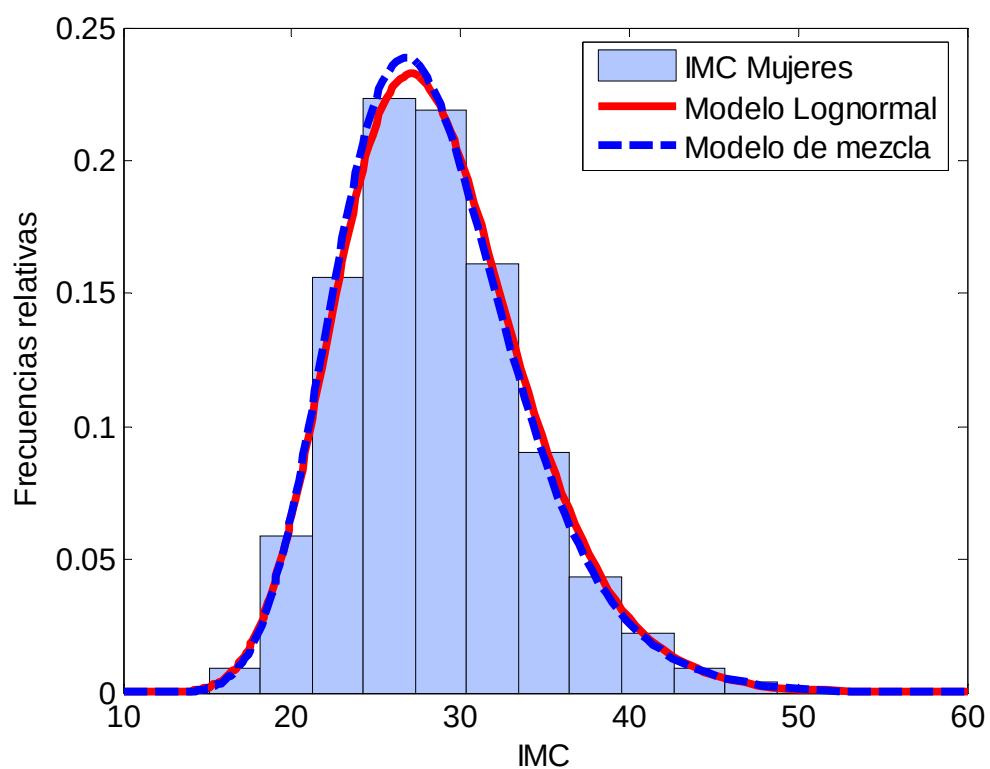


Figura 2.88. Ajuste con el Modelo de Mezcla y el modelo Lognormal. Mujeres

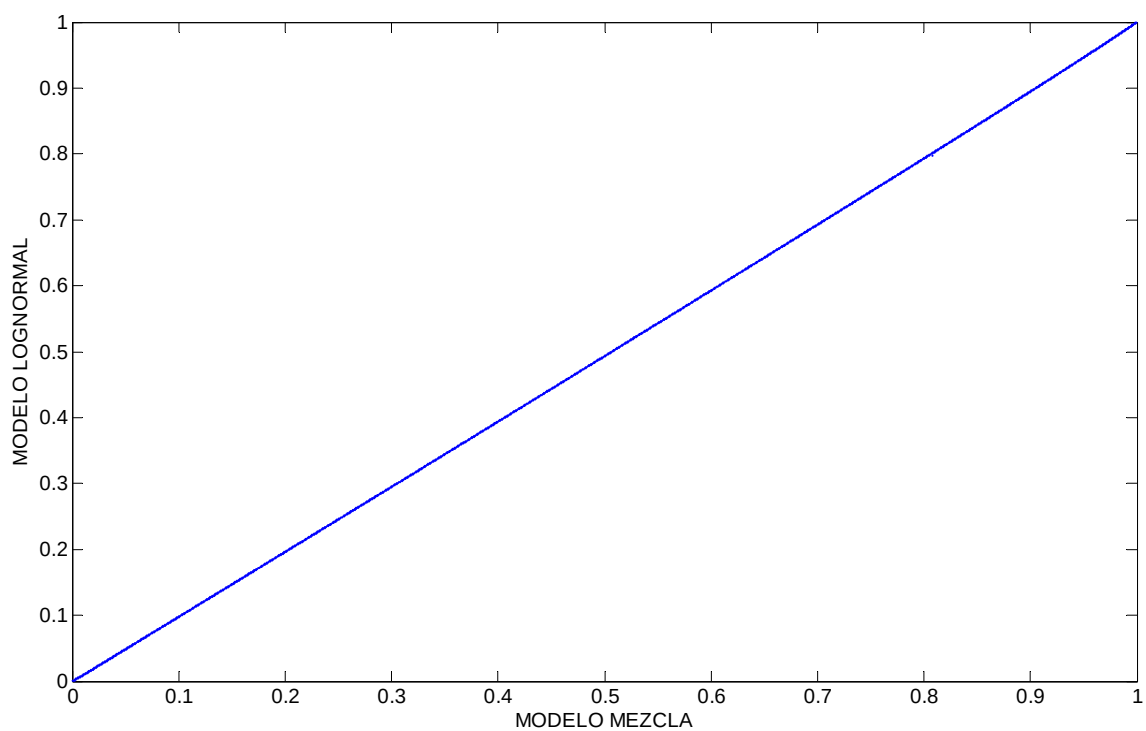


Figura 2.89. Gráfica de probabilidades de los modelos propuestos. Hombres Nacional

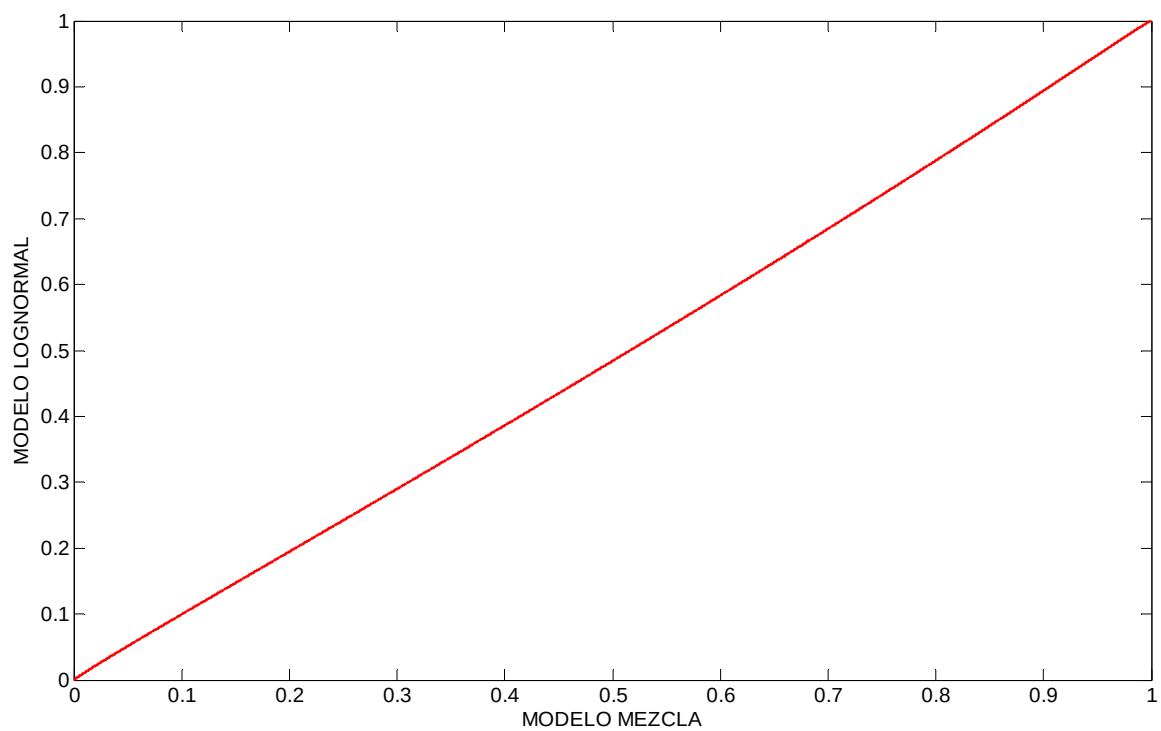


Figura 2.90. Gráfica de probabilidades de los modelos propuestos. Mujeres Nacional

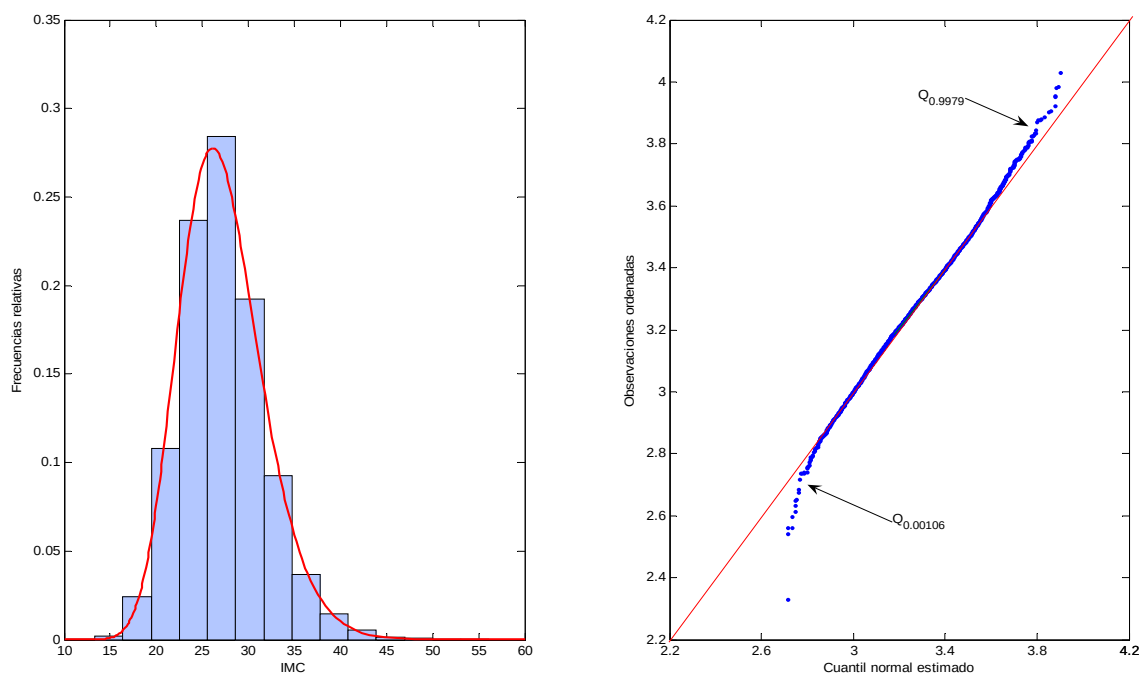


Figura 2.91. Hombres Nacional. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

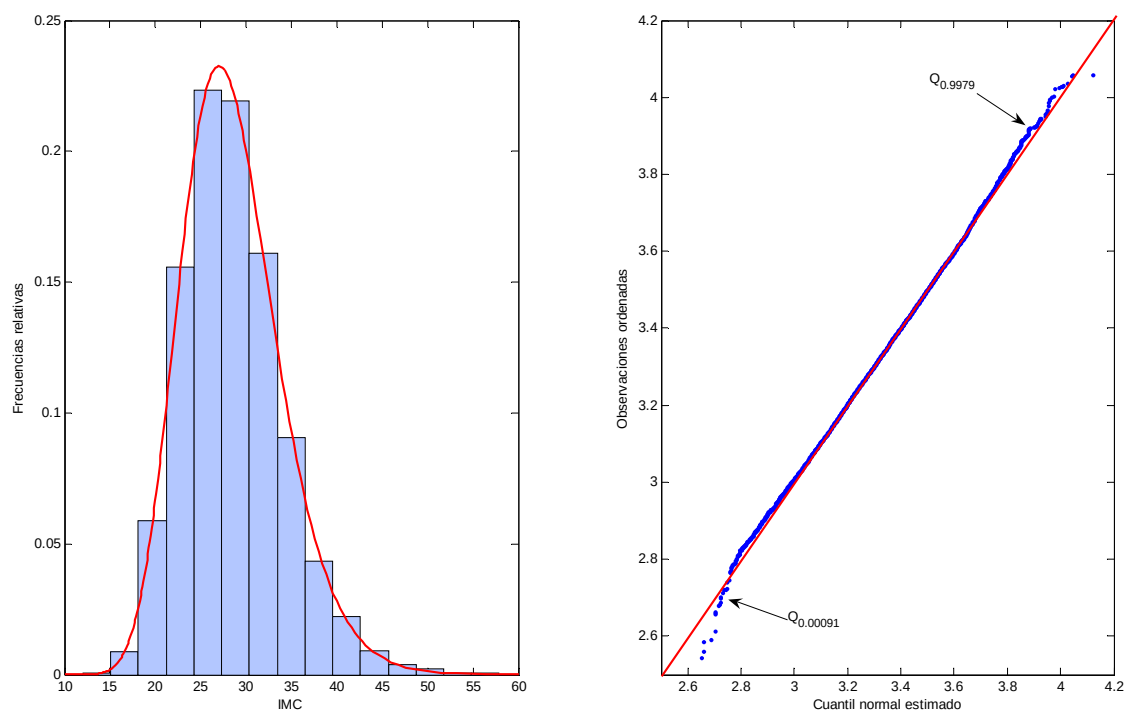


Figura 2.92. Mujeres Nacional. Ajuste con el Modelo Lognormal (izquierda) y gráfica QQ (derecha)

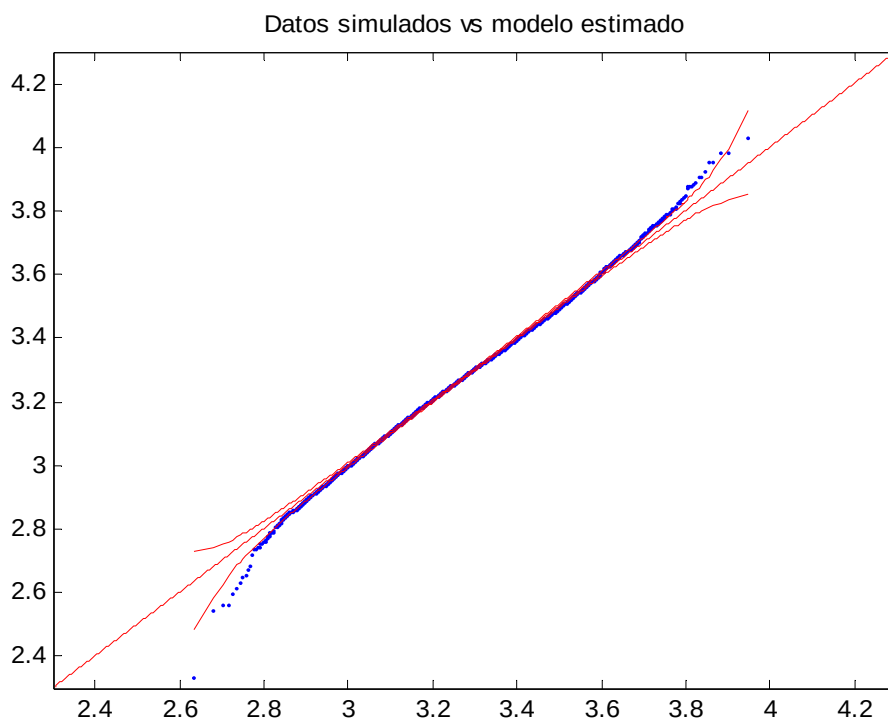


Figura 2.93. Hombres 1000 repeticiones

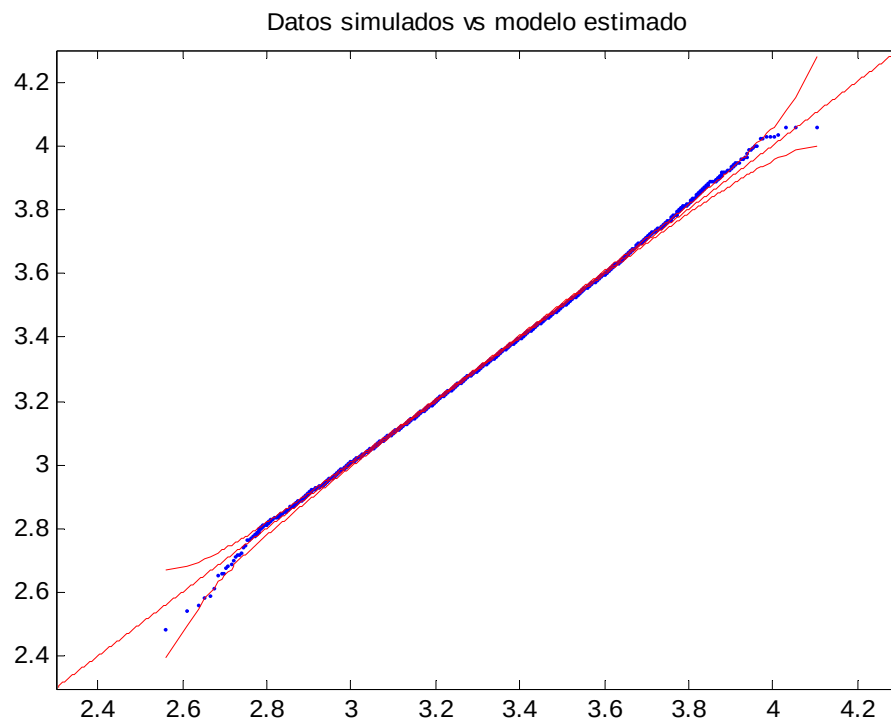


Figura 2.94. Mujeres 1000 repeticiones

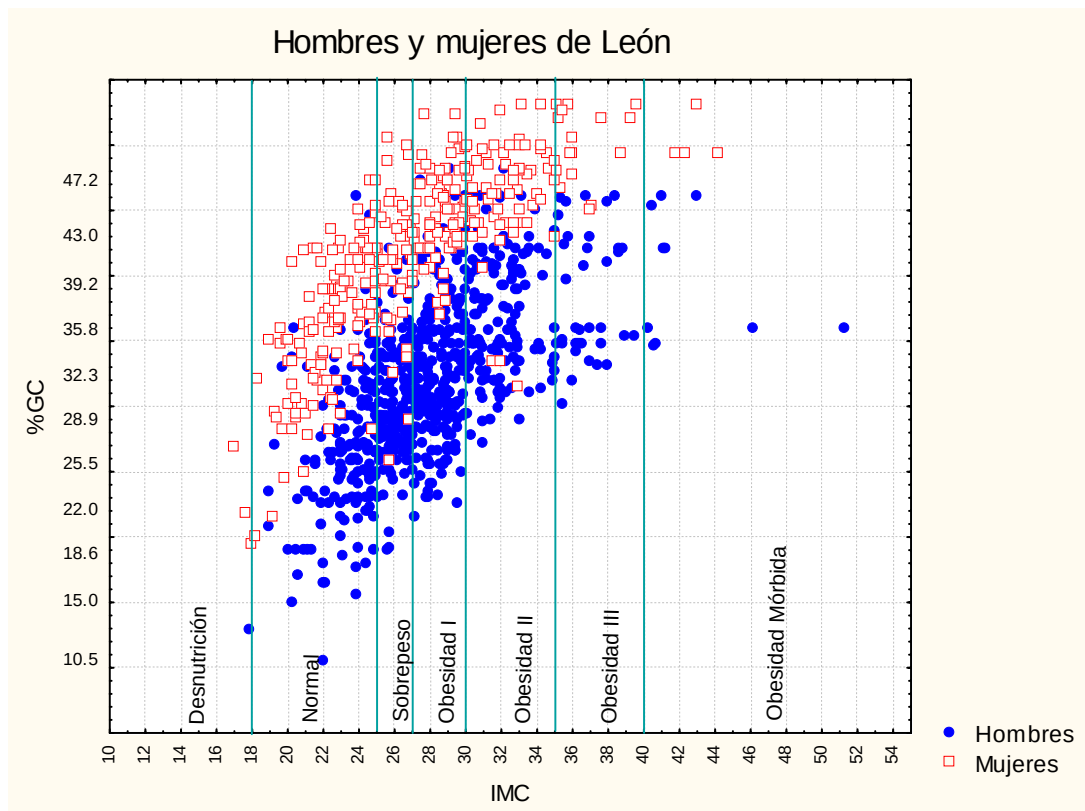


Figura 2.95. Dispersión del %GC vs. IMC

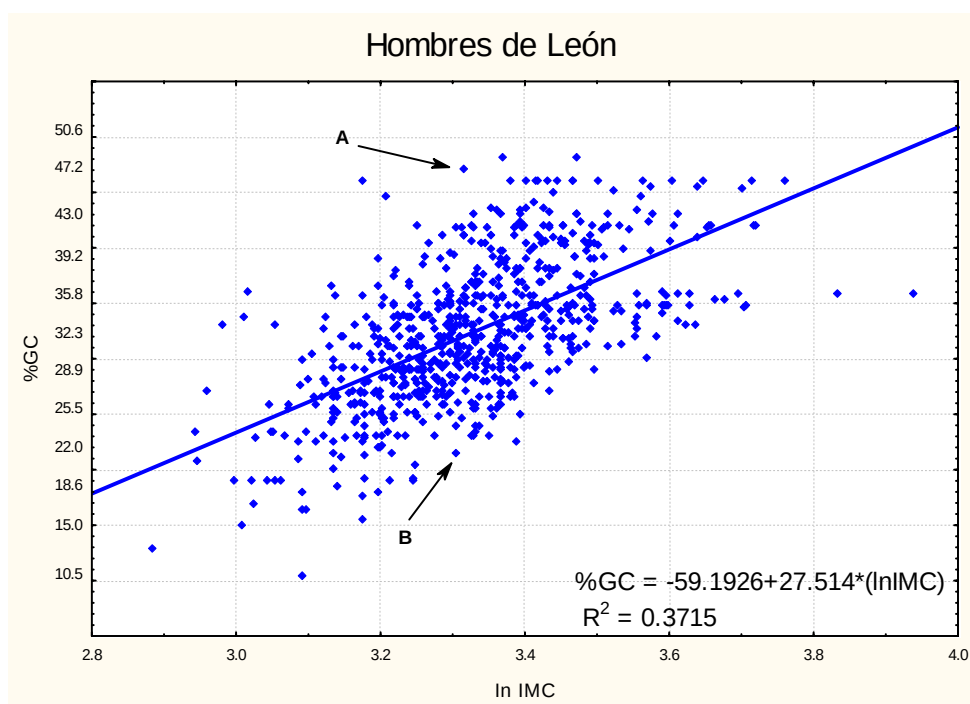


Figura 2.96. Dispersión del %GC vs. $\ln(IMC)$. Hombres de León.

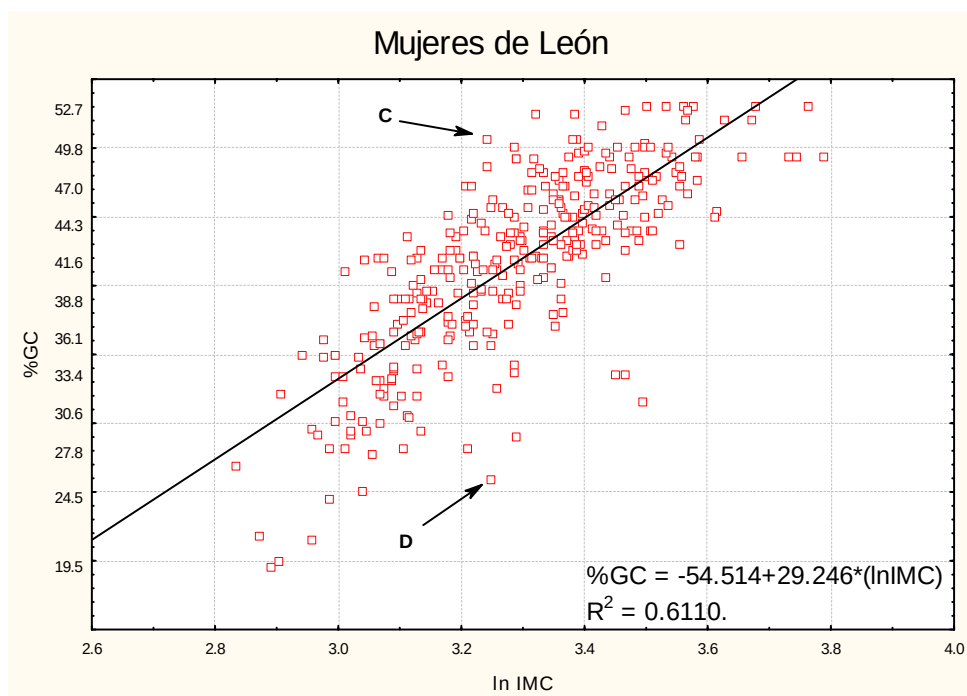


Figura 2.97. Dispersión del %GC vs. $\ln(IMC)$. Mujeres de León

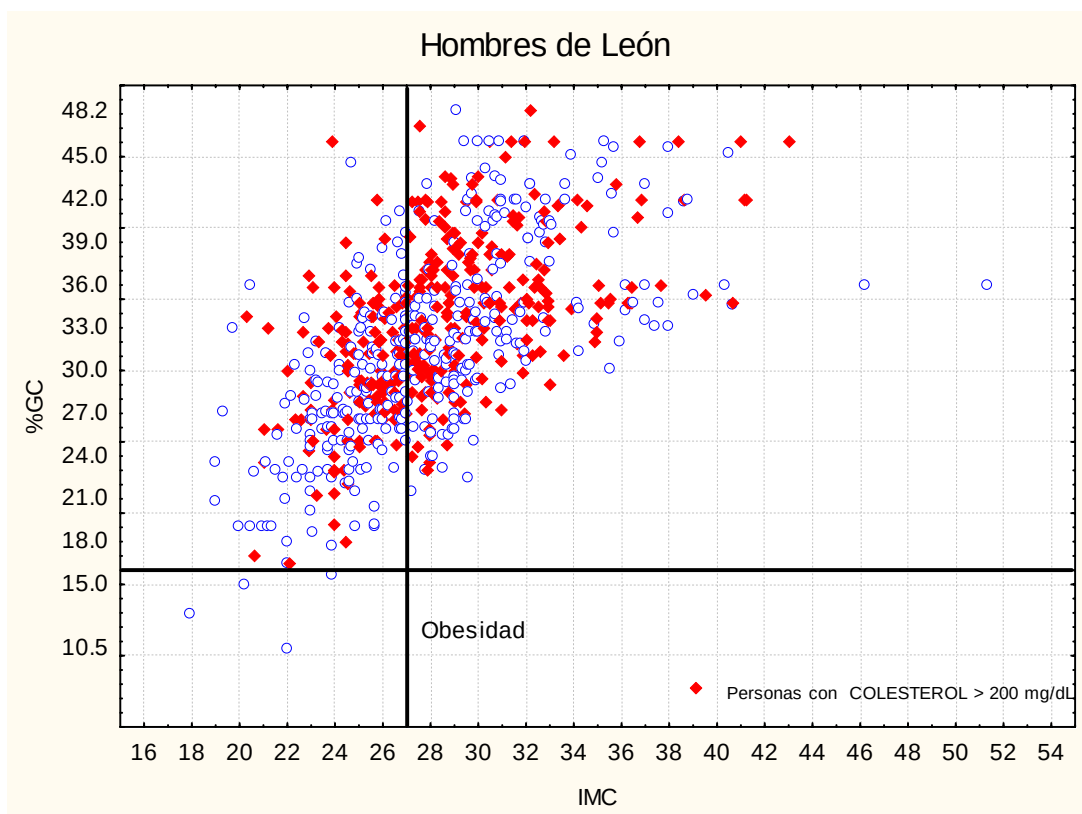


Figura 2.98. Hombres adultos con obesidad y colesterol alto

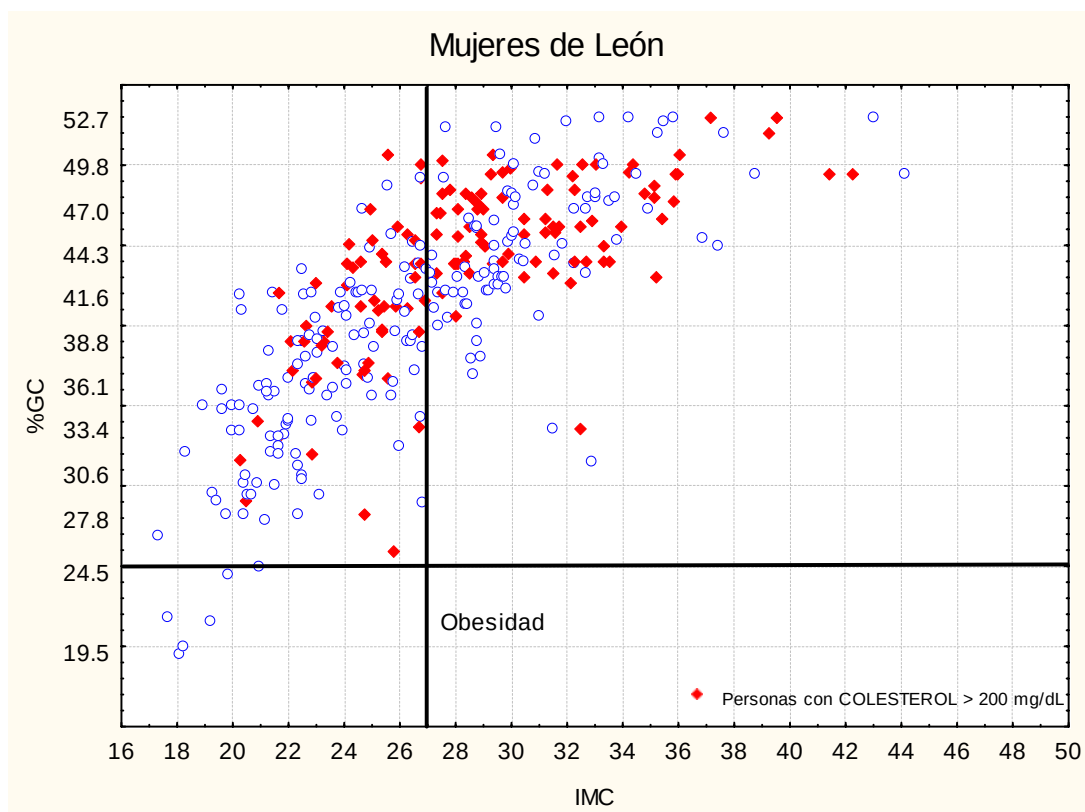


Figura 2.99. Mujeres adultas con obesidad y colesterol alto.

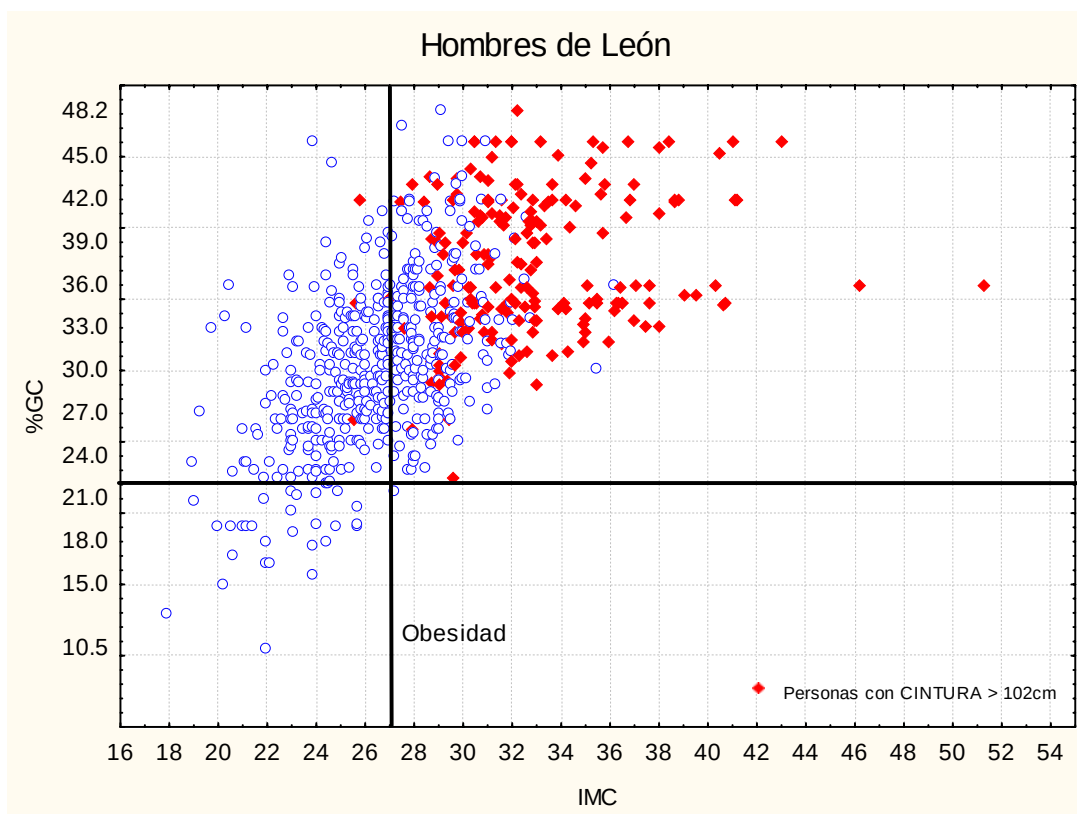


Figura 2.100. Hombres adultos con obesidad y obesidad androide

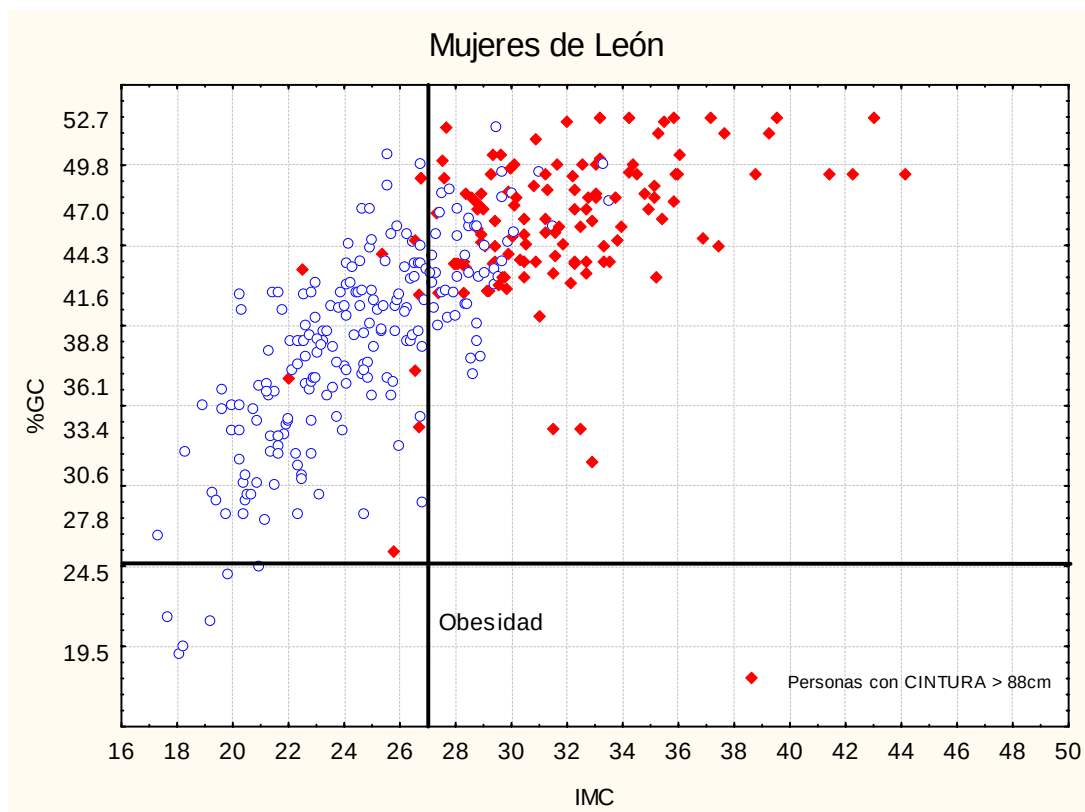


Figura 2.101. Mujeres adultas con obesidad y obesidad androide

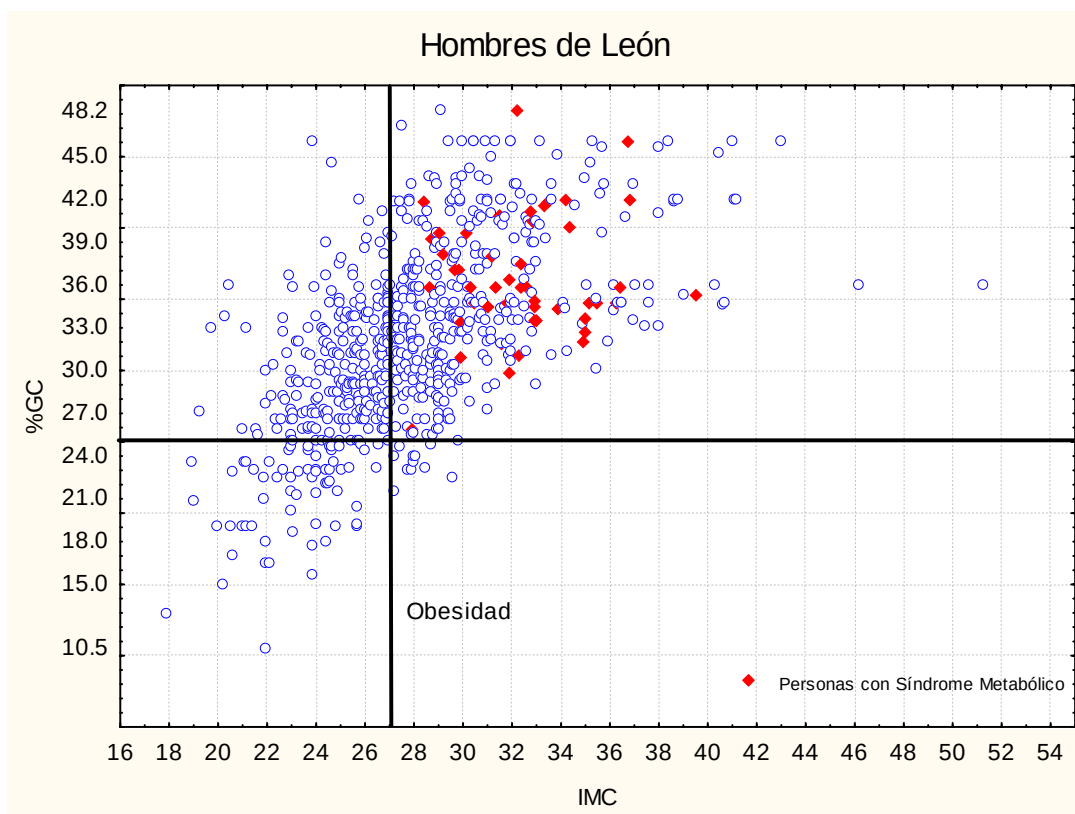


Figura 2.102. Hombres adultos con obesidad y síndrome metabólico

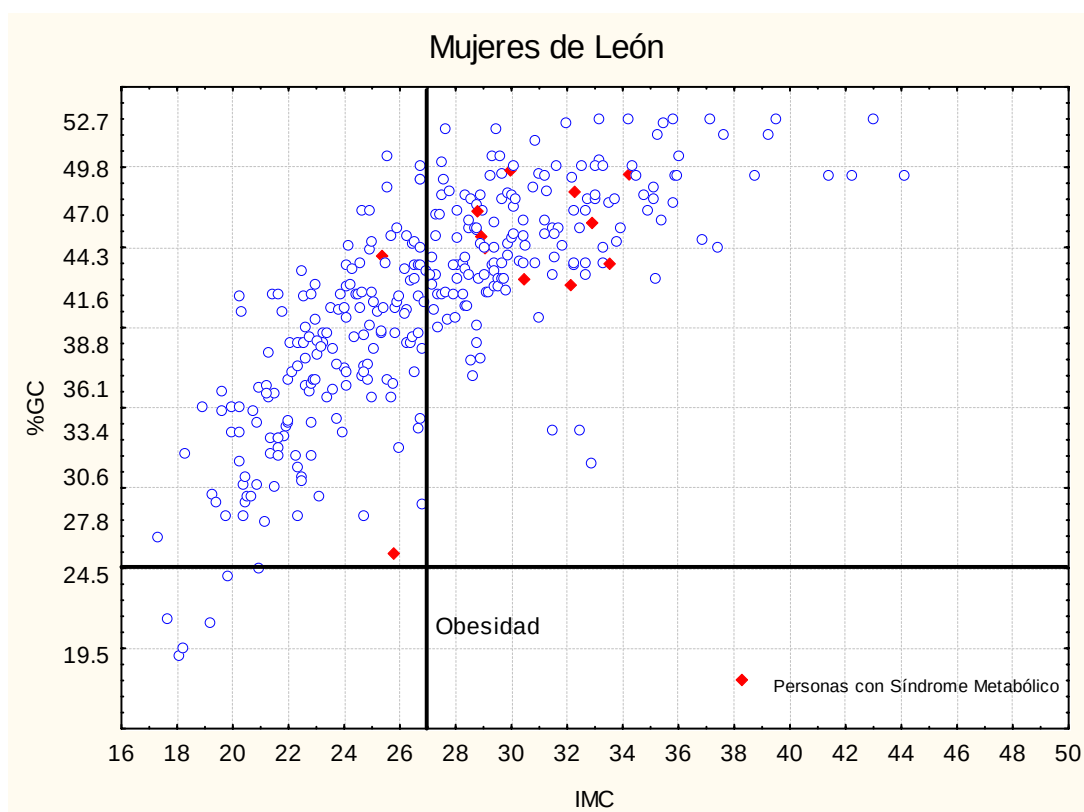


Figura 2.103. Mujeres adultas con obesidad y síndrome metabólico

Apéndice B. Tablas del Capítulo 2

Apéndice B. Tablas del Capítulo 2

Modelo Lognormal.					
Cuantil	Cuantil $\widehat{Q}_\alpha$		Cuantil	Cuantil $\widehat{Q}_\alpha$	
Q_α	hombres	mujeres	Q_α	hombres	mujeres
$Q_{0.05}$	20.06	20.07	$Q_{0.55}$	27.15	28.37
$Q_{0.10}$	21.34	21.55	$Q_{0.60}$	27.74	29.09
$Q_{0.15}$	22.26	22.60	$Q_{0.65}$	28.38	29.85
$Q_{0.20}$	23.01	23.48	$Q_{0.70}$	29.06	30.67
$Q_{0.25}$	23.68	24.26	$Q_{0.75}$	29.81	31.58
$Q_{0.30}$	24.29	24.98	$Q_{0.80}$	30.68	32.63
$Q_{0.35}$	24.88	25.67	$Q_{0.85}$	31.72	33.9
$Q_{0.40}$	25.44	26.34	$Q_{0.90}$	33.07	35.56
$Q_{0.45}$	26	27.01	$Q_{0.95}$	35.19	38.18
$Q_{0.50}$	26.57	27.68			

Tabla B.1: Cuantiles teóricos estimados para adultos mexicanos

Estado	p		Estado	p	
	hombres	mujeres		hombres	mujeres
Aguascalientes	0.0096	0.0098	Nayarit	0.0096	0.0093
Baja California	0.0263	0.0248	Nuevo León	0.0401	0.0386
Baja California Sur	0.0045	0.0042	Oaxaca	0.0348	0.0357
Campeche	0.0072	0.0069	Puebla	0.0515	0.0527
Coahuila	0.0239	0.0232	Querétaro	0.0143	0.0145
Colima	0.0056	0.0055	Quintana Roo	0.0094	0.0085
Chiapas	0.0408	0.0397	San Luis Potosí	0.0236	0.0236
Chihuahua	0.0319	0.0307	Sinaloa	0.0266	0.0255
Distrito Federal	0.0864	0.0901	Sonora	0.0233	0.0222
Durango	0.0149	0.0148	Tabasco	0.0196	0.0192
Guanajuato	0.0469	0.0487	Tamaulipas	0.0286	0.0279
Guerrero	0.0313	0.0319	Tlaxcala	0.0099	0.0098
Hidalgo	0.0227	0.0231	Veracruz	0.0705	0.0712
Jalisco	0.0645	0.0652	Yucatán	0.0172	0.0168
México	0.1346	0.1341	Zacatecas	0.0137	0.0140
Michoacán	0.0402	0.0415			
Morelos	0.0158	0.0161			

Tabla B.2: Proporción de individuos por estado. Datos proporcionados por INEGI en el censo del año 2000

Apéndice B. Tablas del Capítulo 2

Tamaño de muestras del IMC en adultos mexicanos. ENSANUT 2006. Entidades federativas.					
Estado	Tamaño de muestra		Estado	Tamaño de muestra	
	hombres	mujeres		hombres	mujeres
Aguascalientes	388	657	Nayarit	494	699
Baja California	300	423	Nuevo León	410	572
Baja California Sur	456	549	Oaxaca	459	723
Campeche	402	576	Puebla	401	573
Coahuila	405	524	Querétaro	383	578
Colima	396	605	Quintana Roo	378	477
Chiapas	455	630	San Luis Potosí	316	630
Chihuahua	453	631	Sinaloa	428	558
Distrito Federal	405	545	Sonora	437	552
Durango	380	662	Tabasco	417	624
Guanajuato	455	659	Tamaulipas	471	562
Guerrero	367	701	Tlaxcala	415	725
Hidalgo	411	677	Veracruz	485	680
Jalisco	382	626	Yucatán	454	630
México	426	589	Zacatecas	436	673
Michoacán	347	733			
Morelos	417	755	TOTAL	13, 229	19, 798

Tabla B.3: Tamaño de muestras del IMC en adultos mexicanos.

Apéndice B. Tablas del Capítulo 2

Inferencia para λ en la familia de Box-Cox Hombres Nacional						
Muestra	IVC	$\hat{\lambda}_p$	λ_{elegida}	R_P		
				$\lambda = -0.5$	$\lambda = 0$	$\lambda = 0.5$
Aguascalientes	[-0.375, 0.49]	0.0607	0	0.03335	0.95312	0.13269
Baja California	[-0.26, 0.64]	0.183	0	0.01029	0.72638	0.39017
Baja California Sur	[-0.17, 0.62]	0.224	0	0.00156	0.53834	0.39297
Campeche	[-0.67, 0.08]	-0.290	0	0.55383	0.31896	0.00019
Coahuila	[0.04, 0.83]	0.436	0.5	1.99e-5	0.09846	0.95191
Colima	[-0.63, 0.15]	-0.239	0	0.41691	0.48235	0.00092
Chiapas	[-1.39, -0.45]	-0.909	-1	0.23093	0.00062	1.36e-8
Chihuahua	[-0.19, 0.6]	0.205	0	0.00190	0.59283	0.34025
Distrito Federal	[-0.2, 0.72]	0.4824	0.5	0.0063	0.0709	0.9965
Durango	[-0.51, 0.33]	-0.089	0	0.15820	0.91677	0.02162
Guanajuato	[-0.89, 0.12]	-0.137	0	0.26291	0.82554	0.01588
Guerrero	[-0.28, 0.55]	-0.414	-0.5	0.95125	0.31455	0.00350
Hidalgo	[-0.25, 0.73]	0.238	0	0.01180	0.63293	0.57671
Jalisco	[-0.39, 0.36]	0.356	0.5	0.00346	0.37958	0.85492
México	[-0.18, 0.75]	0.280	0	0.00372	0.48916	0.64329
Michoacán	[-0.05, 0.77]	0.044	0	0.10679	0.98583	0.21048
Morelos	[-0.53, 0.33]	-0.094	0	0.18618	0.91292	0.02484
Nayarit	[-0.05, 0.87]	-0.0387	0	0.13176	0.98568	0.0585
Nuevo León	[-0.37, 0.42]	0.3822	0.5	0.00176	0.30467	0.8934
Oaxaca	[-1.05, -0.13]	-0.0086	0	0.18350	0.9289	0.0325
Puebla	[-0.62, 0.45]	-0.4664	-0.5	0.99010	0.1436	0.0002
Querétaro	[-0.69, 0.21]	-0.233	0	0.51687	0.60198	0.00628
Quintana Roo	[-0.25, 0.88]	0.320	0	0.01590	0.53564	0.82166
San Luis Potosí	[-0.3, 0.6]	0.4509	0.5	0.0052	0.3118	0.986
Sinaloa	[-0.53, 0.2]	0.3821	0.5	0.0019	0.3087	0.8939
Sonora	[-0.11, 0.54]	0.214	0	6.721e-5	0.44038	0.24220
Tabasco	[-0.37, 0.45]	0.305	0.5	1.57e-5	0.224	0.5617
Tamaulipas	[-0.31, 0.48]	0.092	0	0.01420	0.90180	0.13187
Tlaxcala	[-0.63, 0.22]	-0.195	0	0.37856	0.66868	0.00534
Veracruz	[-0.73, 0.07]	0.0275	0	0.0784	0.9929	0.1227
Yucatán	[-0.65, 0.28]	-0.183	-0.5	0.67801	0.3443	0.0007
Zacatecas	[-0.18, 55]	0.189	0	0.00119	0.60538	0.26154

Tabla B.4: Inferencia para λ en la familia de Box-Cox. Hombres

Apéndice B. Tablas del Capítulo 2

Inferencia para λ en la familia de Box-Cox. Mujeres, entidades federativas.						
Muestra	IVC.	$\hat{\lambda}_p$	λ_{elegida}	R_p		
				$\lambda = -0.5$	$\lambda = 0$	$\lambda = 0.5$
Aguascalientes	[-0.45,0.17]	-0.138	0	0.0771	0.6870	0.0003
Baja California Norte	[-0.72,0.05]	-0.325	-0.5	0.6724	0.2524	0.0001
Baja California Sur	[-0.01,0.62]	0.307	0.5	2.59e-6	0.1528	0.4775
Campeche	[-0.32,0.32]	0.001	0	0.0103	1.0000	0.0110
Coahuila	[-0.42,0.21]	-0.101	0	0.0470	0.8212	0.0008
Colima	[-0.49,0.16]	-0.168	0	0.1316	0.5956	0.0003
Chiapas	[-0.30,0.35]	0.028	0	0.0068	0.9863	0.0197
Chihuahua	[-0.31,0.26]	-0.025	0	0.0051	0.9861	0.0019
Distrito Federal	[-0.71,-0.03]	-0.367	-0.5	0.7468	0.1071	3.44e-6
Durango	[-0.17,0.46]	0.139	0	0.0002	0.6748	0.0713
Guanajuato	[-0.32,0.34]	0.018	0	0.0095	0.9942	0.0172
Guerrero	[-0.72,-0.16]	-0.430	-0.5	0.8936	0.0131	1.07e-9
Hidalgo	[-0.58,0.1]	-0.239	0	0.3175	0.3758	7.78e-5
Jalisco	[-0.36,0.28]	-0.039	0	0.0198	0.9728	0.0045
México	[-0.32,0.34]	0.014	0	0.0099	0.9965	0.0157
Michoacán	[-0.11,0.46]	0.172	0	1.52e-5	0.4858	0.0746
Morelos	[-0.51,0.07]	-0.220	0	0.1714	0.3341	6.58e-6
Nayarit	[-0.52,0.05]	-0.230	0	0.1866	0.2926	3.46e-6
Nuevo León	[-0.36,0.23]	-0.064	0	0.0149	0.9126	0.0009
Oaxaca	[-0.46,0.11]	-0.169	0	0.0789	0.5258	4.84e-5
Puebla	[-0.38,0.42]	0.024	0	0.0355	0.9930	0.0632
Querétaro	[-0.74,-0.03]	-0.377	-0.5	0.7946	0.1115	6.33e-6
Quintana Roo	[-0.29,0.46]	0.088	0	0.0093	0.9012	0.1014
San Luis Potosí	[-0.06,0.58]	0.259	0.5	1.33e-5	0.2721	0.3259
Sinaloa	[-0.63,0.02]	-0.341	-0.5	0.6880	0.177	0.0002
Sonora	[-0.31,0.33]	0.014	0	0.0073	0.9964	0.0118
Tabasco	[-0.23,0.43]	0.103	0	0.0017	0.8304	0.0614
Tamaulipas	[-0.68,-0.02]	-0.341	-0.5	0.6489	0.1320	3.74e-6
Tlaxcala	[-0.43,0.2]	-0.114	0	0.0610	0.7824	0.0007
Veracruz	[-0.55,0.03]	-0.259	0	0.2659	0.2178	1.95e-6
Yucatán	[-0.62,0.02]	-0.293	-0.5	0.4477	0.1985	6.39e-6
Zacatecas	[-0.24,0.39]	0.079	0	0.0018	0.8871	0.0337

Tabla B.5: Inferencia para λ en la familia de Box-Cox. Mujeres

Apéndice B. Tablas del Capítulo 2

Modelo de Box-Cox propuesto						
Muestra	Hombres			Mujeres		
	λ_{elegida}	EMV		λ_{elegida}	EMV	
		$\hat{\mu}_{\lambda}$	$\hat{\sigma}_{\lambda}^2$		$\hat{\mu}_{\lambda}$	$\hat{\sigma}_{\lambda}^2$
Aguascalientes	0	3.2729	0.0273	0	3.3082	0.1933
Baja California Norte	0	3.2726	0.0296	-0.5	1.6205	0.0394
Baja California Sur	0	3.2997	0.0272	0.5	8.6938	1.0467
Campeche	0	3.3063	0.0306	0	3.36	0.1981
Coahuila	0.5	8.4761	0.7845	0	3.3469	0.2155
Colima	0	3.2980	0.1827	0	3.3168	0.2030
Chiapas	-1	0.9613	0.0059	0	3.3031	0.1856
Chihuahua	0	3.2828	0.1716	0	3.3453	0.2010
Distrito Federal	0.5	8.4081	0.925	-0.5	1.6182	0.0356
Durango	0	3.2827	0.1627	0	3.3536	0.1945
Guanajuato	0	3.8273	0.1748	0	3.3321	0.1850
Guerrero	-0.5	1.6109	0.0326	-0.5	1.6096	0.0374
Hidalgo	0	3.2519	0.15887	0	3.2889	0.1753
Jalisco	0.5	8.3489	0.8871	0	3.3144	0.1986
México	0	3.275	0.1639	0	3.3078	0.1830
Michoacán	0	3.739	0.1614	0	3.3156	0.1964
Morelos	0	3.2710	0.1607	0	3.3069	0.1888
Nayarit	0	3.797	0.180	0	3.3348	0.2032
Nuevo León	0.5	8.2948	0.8753	0	3.3478	0.2210
Oaxaca	0	3.315	0.1631	0	3.2730	0.1838
Puebla	-0.5	1.6103	0.0011	0	3.2943	0.1765
Querétaro	0	3.2724	0.1672	-0.5	1.6122	0.0363
Quintana Roo	0	3.3012	0.1666	0	3.3446	0.1883
San Luis Potosí	0.5	8.2595	0.8618	0	3.3064	0.1986
Sinaloa	0.5	8.2802	0.8807	-0.5	1.624	0.0383
Sonora	0	3.2901	0.1902	0	3.363	0.2071
Tabasco	0.5	8.2986	0.8718	0	3.3400	0.1928
Tamaulipas	0	3.3007	0.1843	-0.5	1.6203	0.0389
Tlaxcala	0	3.2809	0.1570	0	3.3153	0.1827
Veracruz	0	3.6103	0.187	0	3.3048	0.1956
Yucatán	-0.5	1.6031	0.0322	-0.5	1.6242	0.0350
Zacatecas	0	3.2531	0.1826	0	3.3159	0.1876

Tabla B.6: Modelo de Box-Cox propuesto de adultos mexicanos.

Entidad Federativa	Asimetría	Intervalo de 99% de confianza para la asimetría		Curtosis	Intervalo de 99% de confianza para la curtosis	
		Límite inferior	Límite superior		Límite inferior	Límite superior
Aguascalientes	-0.038	-0.360	0.283	3.699	3.060	4.339
Baja California Norte	-0.170	-0.536	0.196	4.081	3.354	4.809
Baja California Sur	-0.120	-0.416	0.177	3.585	2.995	4.175
Campeche	0.141	-0.175	0.457	4.062	3.434	4.691
Coahuila	0.030	-0.284	0.345	3.665	3.039	4.291
Colima	0.156	-0.162	0.475	3.711	3.077	4.344
Chiapas	-0.030	-0.327	0.266	2.983	2.393	3.574
Chihuahua	-0.149	-0.446	0.148	3.662	3.070	4.254
Distrito Federal	0.030	-0.284	0.345	3.665	3.039	4.291
Durango	-0.158	-0.483	0.167	3.382	2.736	4.029
Guanajuato	0.078	-0.218	0.375	3.550	2.959	4.141
Guerrero	-0.032	-0.362	0.299	2.762	2.104	3.419
Hidalgo	-0.173	-0.485	0.140	3.470	2.848	4.091
Jalisco	0.076	-0.248	0.400	3.161	2.516	3.805
México	-0.108	-0.415	0.199	3.232	2.622	3.843
Michoacán	-0.024	-0.364	0.316	3.513	2.837	4.190
Morelos	-0.157	-0.467	0.154	3.499	2.882	4.116
Nayarit	0.019	-0.266	0.304	3.120	2.553	3.687
Nuevo León	0.061	-0.252	0.374	3.121	2.498	3.743
Oaxaca	0.045	-0.251	0.340	3.090	2.502	3.678
Puebla	-0.024	-0.340	0.292	3.202	2.573	3.831
Querétaro	0.012	-0.311	0.336	3.119	2.475	3.763
Quintana Roo	0.099	-0.227	0.424	3.034	2.386	3.682
San Luis Potosí	0.013	-0.343	0.370	2.683	1.974	3.392
Sinaloa	0.061	-0.245	0.367	3.191	2.582	3.800
Sonora	0.093	-0.210	0.396	3.931	3.328	4.534
Tabasco	0.188	-0.122	0.498	4.005	3.388	4.623
Tamaulipas	-0.029	-0.321	0.263	3.068	2.487	3.648
Tlaxcala	-0.112	-0.422	0.199	3.170	2.551	3.789
Veracruz	-0.014	-0.301	0.273	3.293	2.721	3.866
Yucatán	-0.113	-0.410	0.184	3.260	2.668	3.851
Zacatecas	0.084	-0.220	0.387	2.904	2.300	3.507

Tabla B.7. Coeficientes de Asimetría y Curtosis de los datos transformados para hombres.

Entidad Federativa	Asimetría	Intervalo de 99% de confianza para la asimetría		Curtosis	Intervalo de 99% de confianza para la curtosis	
		Límite inferior	Límite superior		Límite inferior	Límite superior
Agascalientes	0.075	-0.172	0.321	3.023	2.531	3.515
Baja California Norte	-0.091	-0.399	0.217	2.743	2.130	3.356
Baja California Sur	0.212	-0.058	0.482	3.499	2.961	4.037
Campeche	0.001	-0.262	0.265	3.120	2.595	3.645
Coahuila	0.114	-0.162	0.390	3.040	2.489	3.591
Colima	0.101	-0.156	0.358	2.893	2.381	3.406
Chiapas	0.010	-0.242	0.262	2.974	2.472	3.476
Chihuahua	0.000	-0.252	0.251	3.513	3.011	4.015
Distrito Federal	-0.076	-0.347	0.195	3.249	2.709	3.789
Durango	-0.053	-0.299	0.193	3.110	2.620	3.600
Guanajuato	-0.032	-0.278	0.215	2.949	2.457	3.440
Guerrero	-0.077	-0.316	0.161	3.197	2.721	3.673
Hidalgo	0.092	-0.151	0.335	3.197	2.712	3.681
Jalisco	0.059	-0.194	0.312	2.924	2.420	3.428
México	0.005	-0.256	0.265	3.236	2.717	3.756
Michoacán	-0.130	-0.363	0.104	3.324	2.858	3.789
Morelos	0.129	-0.101	0.359	3.221	2.762	3.680
Nayarit	0.160	-0.079	0.399	3.114	2.638	3.591
Nuevo León	0.018	-0.247	0.282	3.042	2.515	3.569
Oaxaca	0.140	-0.096	0.375	3.280	2.811	3.749
Puebla	-0.081	-0.345	0.183	2.677	2.150	3.203
Querétaro	-0.053	-0.316	0.210	2.843	2.319	3.367
Quintana Roo	-0.050	-0.340	0.239	2.942	2.365	3.520
San Luis Potosí	-0.148	-0.400	0.104	3.002	2.499	3.504
Sinaloa	-0.093	-0.361	0.175	3.095	2.561	3.629
Sonora	0.043	-0.226	0.312	3.035	2.499	3.572
Tabasco	-0.073	-0.326	0.180	2.877	2.373	3.382
Tamaulipas	-0.092	-0.359	0.174	2.802	2.271	3.334
Tlaxcala	0.050	-0.184	0.285	3.087	2.619	3.556
Veracruz	0.116	-0.127	0.358	3.268	2.785	3.752
Yucatán	-0.138	-0.390	0.114	3.297	2.795	3.800
Zacatecas	-0.069	-0.313	0.175	3.128	2.642	3.614

Tabla B.8. Coeficientes de Asimetría y Curtosis de los datos transformados para mujeres.

Entidad federativa	N	Infrapeso	Normal	Sobrepeso	Obesidad I	Obesidad II	Obesidad III	Mórbida
Aguascalientes	388	1.040	36.156	18.328	22.598	17.498	3.780	0.600
Baja California	300	1.320	36.406	17.638	21.898	17.728	4.230	0.780
Baja California Sur	456	0.660	30.537	17.878	24.028	20.848	5.140	0.910
Campeche	402	0.880	29.997	16.738	22.988	21.648	6.299	1.450
Coahuila	405	1.220	28.317	16.688	24.308	23.168	5.579	0.720
Colima	396	0.950	29.777	16.428	22.708	21.848	6.649	1.640
Chiapas	455	0.170	39.236	20.198	21.538	14.459	3.390	1.010
Chihuahua	453	0.810	32.057	17.778	23.418	20.068	4.960	0.910
Distrito Federal	405	1.220	28.317	16.688	24.308	23.168	5.579	0.720
Durango	380	0.520	31.367	18.888	24.728	19.768	4.140	0.590
Guanajuato	455	0.500	30.777	18.818	24.878	20.138	4.280	0.610
Guerrero	367	0.420	34.227	18.568	22.678	18.118	4.750	1.240
Hidalgo	411	1.160	34.817	17.558	22.328	18.648	4.610	0.880
Jalisco	382	1.440	36.636	19.208	23.818	16.448	2.290	0.160
México	426	1.140	34.617	17.558	22.388	18.758	4.650	0.890
Michoacán	347	0.520	32.107	19.178	24.728	19.158	3.800	0.510
Morelos	417	0.730	31.087	17.748	23.718	20.618	5.150	0.950
Nayarit	494	0.690	34.377	19.068	23.828	18.008	3.550	0.480
Nuevo León	410	0.960	29.357	17.978	25.407	21.748	4.160	0.390
Oaxaca	459	0.610	28.767	17.348	24.088	22.098	5.919	1.170
Puebla	401	0.870	42.806	18.548	20.078	13.879	3.110	0.710
Querétaro	383	0.290	33.597	21.468	25.907	16.398	2.170	0.170
Quintana Roo	378	0.690	34.247	19.048	23.858	18.088	3.580	0.490
San Luis Potosí	316	0.900	26.677	16.948	25.267	24.048	5.519	0.640
Sinaloa	428	1.150	34.757	19.428	24.718	17.388	2.400	0.160
Sonora	437	1.290	33.377	16.768	21.968	19.748	5.569	1.280
Tabasco	417	1.810	29.347	15.788	22.838	22.808	6.349	1.060
Tamaulipas	471	0.560	28.337	17.418	24.308	22.318	5.909	1.150
Tlaxcala	415	0.890	29.017	16.308	22.808	22.318	6.919	1.740
Veracruz	485	0.400	31.057	19.598	25.477	19.398	3.630	0.440
Yucatán	454	0.600	35.456	17.948	21.768	17.798	4.970	1.460
Zacatecas	436	0.370	25.827	17.498	25.317	23.698	6.169	1.120
TOTAL	13,229	0.880	32.830	18.120	23.550	19.440	4.430	0.750

Tabla B.9. Porcentaje estimado del estado nutricional de hombres por entidad federativa.

Entidad federativa	N	Infrapeso	Normal	Sobrepeso	Obesidad I	Obesidad II	Obesidad III	Mórbida
Aguascalientes	657	1.110	27.987	15.208	21.758	22.998	8.289	2.650
Baja California	423	0.700	26.747	14.419	20.318	22.378	9.819	5.619
Baja California Sur	549	1.140	20.838	13.069	21.838	28.277	11.639	3.200
Campeche	576	0.520	19.658	13.209	21.648	27.597	12.359	5.010
Coahuila	524	1.220	23.578	12.919	19.838	24.758	11.809	5.879
Colima	605	1.470	28.107	14.399	20.658	22.788	9.099	3.480
Chiapas	630	0.770	27.567	16.138	23.068	23.188	7.379	1.890
Chihuahua	631	0.790	22.988	13.879	21.508	25.707	10.879	4.250
Distrito Federal	545	0.430	27.307	16.038	22.178	22.348	8.249	3.450
Durango	662	0.540	20.588	13.639	21.988	27.187	11.639	4.420
Guanajuato	659	0.660	24.248	15.028	22.748	25.207	9.249	2.860
Guerrero	701	1.240	36.996	16.268	19.798	17.518	5.849	2.330
Hidalgo	677	0.860	30.747	17.098	23.148	21.098	5.809	1.240
Jalisco	626	1.100	27.377	14.959	21.638	23.328	8.699	2.900
México	589	0.720	27.907	16.508	23.408	22.898	6.919	1.640
Michoacán	733	1.180	27.607	14.859	21.418	23.188	8.749	3.000
Morelos	755	0.910	27.677	15.658	22.438	23.178	7.879	2.260
Nayarit	699	1.090	24.828	13.869	20.878	24.548	10.479	4.310
Nuevo León	572	1.450	23.828	12.549	19.198	24.288	12.089	6.599
Oaxaca	723	1.550	34.887	16.658	21.398	18.938	5.330	1.240
Puebla	573	0.870	30.137	16.838	23.058	21.538	6.169	1.390
Querétaro	578	0.750	32.917	16.558	21.148	19.458	6.599	2.570
Quintana Roo	477	0.540	21.728	14.289	22.568	26.667	10.599	3.610
San Luis Potosí	630	1.240	28.427	15.048	21.418	22.738	8.349	2.780
Sinaloa	558	0.480	28.127	16.058	21.948	21.928	8.059	3.400
Sonora	552	0.740	20.668	12.819	20.638	26.607	12.649	5.879
Tabasco	624	0.650	22.428	14.149	22.098	26.177	10.659	3.840
Tamaulipas	562	0.640	26.547	14.579	20.568	22.528	9.749	5.390
Tlaxcala	725	0.690	26.327	15.908	23.198	23.938	7.859	2.080
Veracruz	680	1.490	29.567	14.939	20.948	22.058	8.189	2.810
Yucatán	630	0.190	21.238	15.198	23.118	25.667	10.179	4.410
Zacatecas	673	0.800	26.577	15.588	22.698	23.818	8.179	2.340
TOTAL	19,798	0.890	27.370	15.330	21.880	22.990	8.460	3.080

Tabla B.10. Porcentaje estimado del estado nutricional de mujeres por entidad federativa.

Entidad federativa	N	Infrapeso	Normal	Sobrepeso	Obesidad I	Obesidad II	Obesidad III	Mórbida
Aguascalientes	388	1.804	32.732	20.619	24.227	17.010	2.577	1.031
Baja California	300	2.000	35.667	20.667	19.000	17.333	4.667	0.667
Baja California Sur	456	1.097	27.193	18.640	26.316	21.053	4.605	1.097
Campeche	402	0.746	28.607	18.159	26.617	18.408	5.473	1.990
Chiapas	455	0.440	38.901	18.022	24.615	12.967	4.396	0.659
Chihuahua	453	1.325	30.464	16.998	27.594	18.102	4.194	1.325
Coahuila	405	0.741	28.148	16.543	27.654	21.235	4.691	0.988
Colima	396	0.505	30.051	18.687	23.485	19.192	5.556	2.525
Distrito Federal	405	0.741	28.148	16.543	27.654	21.235	4.691	0.988
Durango	380	1.579	27.895	20.789	24.211	20.789	3.947	0.789
Guanajuato	455	0.440	29.670	20.440	27.033	17.143	4.396	0.879
Guerrero	367	0.545	36.512	14.714	23.706	17.439	5.995	1.090
Hidalgo	411	1.460	33.577	15.328	24.818	20.438	3.650	0.730
Jalisco	382	1.047	37.435	18.586	24.346	16.492	1.571	0.524
México	426	1.643	33.333	18.779	22.066	18.779	4.695	0.704
Michoacán	347	1.153	30.836	21.326	25.648	17.003	3.458	0.576
Morelos	417	1.199	29.017	18.225	25.180	21.103	4.317	0.959
Nayarit	494	0.607	34.008	18.826	24.494	18.219	3.441	0.405
Nuevo León	410	0.976	29.512	16.341	27.317	21.463	3.659	0.732
Oaxaca	459	0.436	28.976	17.211	25.054	21.569	5.447	1.307
Puebla	401	0.748	41.397	20.698	19.451	14.464	1.995	1.247
Querétaro	383	0.522	33.943	22.454	24.282	16.710	2.089	0.000
Quintana Roo	378	0.529	34.656	21.693	21.164	15.873	5.556	0.529
San Luis Potosí	316	0.633	26.899	19.304	21.835	25.316	5.696	0.316
Sinaloa	428	0.935	33.645	18.925	28.037	15.421	2.570	0.467
Sonora	437	1.373	29.519	21.281	22.883	19.451	3.661	1.831
Tabasco	417	0.959	30.456	16.307	24.460	20.384	5.516	1.919
Tamaulipas	471	0.212	29.087	14.013	27.176	23.779	4.671	1.062
Tlaxcala	415	1.687	27.229	16.386	24.578	21.928	6.265	1.928
Veracruz	485	0.619	30.103	18.144	28.041	19.381	3.093	0.619
Yucatán	454	1.101	31.278	21.145	24.449	14.978	6.167	0.881
Zacatecas	436	0.688	27.294	16.972	24.541	22.248	7.569	0.688
TOTAL	13,229	0.937	31.386	18.414	24.923	18.966	4.384	0.990

Tabla B.11. Porcentaje empírico del estado nutricional de hombres por entidad federativa.

Entidad federativa	N	Infrapeso	Normal	Sobrepeso	Obesidad I	Obesidad II	Obesidad III	Mórbida
Aguascalientes	657	0.913	27.549	15.830	22.222	22.527	7.915	3.044
Baja California	423	0.473	27.896	12.057	20.095	23.168	10.402	5.910
Baja California Sur	549	1.093	18.944	14.936	24.408	26.776	9.836	4.007
Campeche	576	0.521	19.271	13.368	22.222	28.299	10.069	6.250
Chiapas	630	0.635	28.571	15.238	22.857	22.857	7.778	2.064
Chihuahua	631	0.634	22.345	14.105	22.662	25.990	10.143	4.120
Coahuila	524	1.336	23.282	12.786	22.328	22.328	12.405	5.534
Colima	605	1.157	29.256	15.372	18.347	22.975	8.760	4.132
Distrito Federal	545	0.734	26.606	15.046	23.670	23.303	6.973	3.670
Durango	662	0.453	20.393	13.293	22.205	27.644	12.085	3.928
Guanajuato	659	0.607	24.583	12.747	24.886	25.038	8.953	3.187
Guerrero	701	1.427	37.518	15.121	21.113	17.118	5.706	1.997
Hidalgo	677	0.591	30.281	17.430	24.372	20.384	6.204	0.739
Jalisco	626	0.639	26.837	16.773	21.246	23.163	7.508	3.834
México	589	0.679	26.486	17.997	24.448	22.071	6.791	1.528
Michoacán	733	1.637	26.057	15.416	21.419	22.920	9.823	2.729
Morelos	755	0.662	27.285	16.159	24.371	20.927	8.609	1.987
Nayarit	699	0.715	26.323	14.163	18.598	26.466	8.441	5.293
Nuevo León	572	1.224	23.776	13.462	19.580	22.028	12.762	7.168
Oaxaca	723	0.553	38.036	14.938	20.747	18.672	4.979	2.075
Puebla	573	0.873	28.621	17.103	23.560	22.513	6.108	1.222
Querétaro	578	1.211	32.699	16.955	19.031	21.626	4.844	3.633
Quintana Roo	477	0.629	20.964	14.675	23.899	25.157	11.530	3.145
San Luis Potosí	630	1.587	28.413	13.492	19.048	28.095	6.984	2.381
Sinaloa	558	0.717	25.448	18.996	21.864	21.326	8.781	2.867
Sonora	552	0.906	21.196	13.043	18.659	26.812	13.225	6.159
Tabasco	624	0.801	22.115	14.904	18.750	29.327	10.577	3.526
Tamaulipas	562	0.890	26.157	14.591	19.395	22.776	11.032	5.160
Tlaxcala	725	0.552	25.931	15.586	24.000	25.655	6.069	2.207
Veracruz	680	1.471	30.735	14.412	21.912	20.735	7.059	3.677
Yucatán	630	0.317	20.000	15.714	24.127	27.460	7.778	4.603
Zacatecas	673	1.189	25.854	15.156	23.180	22.883	9.807	1.932
TOTAL	19,798	0.874	26.422	15.072	21.886	23.654	8.632	3.460

Tabla B.12. Porcentaje empírico del estado nutricional de mujeres por entidad federativa.

Bibliografía

- [1] Alberti KG, Zimmet P, Shaw J; IDF Epidemiology Task Force Consensus Group. *The metabolic syndrome—a new worldwide definition*. Lancet. 2005 Sep 24-30; 366(9491):1059-62.
- [2] Arancete J. & col. y Grupo Colaborativo SEEDO (2005). Prevalencia de obesidad en España. *Med Clin (Barc)*, V. 125 (12), pp. 460-466.
- [3] Barnard G. A. and Sprott D. A. (1983). The generalized problem of the Nile: Robust confidence sets for parametric functions. *Annals of Statistics*, V. 11 pp. 104–113.
- [4] Black W. (1983). Obesity: a Report of the Royal College of Physicians. *Journal Royal Coll Phys Lon*. V. 17, pp. 5-64.
- [5] Box, G. E.P.& Tiao G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison Wesley.
- [6] Bray, G.A. (2003). Evaluation of Obesity: Who Are the Obese?. *Postgraduate Medicine*, V. 114, pp. 19-27.
- [7] Casas YG, Schiller BC, De Souza CA, Seals Dr. (2001). Total and regional body composition across ageing in healthy Hispanic and white women of similar socioeconomic status. *Am J Clin Nutrition*, V. 73, pp. 13-18.

- [8] Durazo-Arvizu, R., McGee, D., Li, Z. y Cooper, R.S. (1997). Establishing the Nadir of the Body Mas Index-Mortality Relationship: A Case Study. *Journal of the American Statistical Association*. V. 92, pp. 1312-1319.
- [9] Encuesta nacional de salud y nutrición 2006.(ENSANUT), Instituto Nacional de Salud Pública de México.
- [10] Fisher, R. (1921). *On the "probable error" of a coefficient of correlation reduced from a small sample*. *Metron*, V.1, pp. 3-32.
- [11] Hogg, R.V. y Craig, A. (1995). *Introduction to Mathematical Statistics*. Fifth Edition. Prentice Hall: Nueva Jersey.
- [12] Hu G, Tuomilehto J, Silventoinen K, Barengo NC, Peltonen M, Jousilahti P. (2005). The Effects of Phisical Activity and Body Mass Index on Cardiovascular, Cancer and All-cause Mortality Among 47 212 Middle-aged Finnish Men and Women. *Int J. Obes. Relat. Metab Disord*. V. 29, pp. 894-902.
- [13] Key TJ, Schatzkin A, Willet WC, Allen NE, Spencer EA, Travis RC. (2004) Diet, Nutrition and the Prevention of Cáncer. *Public Health Nutrition*. V. 7, pp. 187-200.
- [14] Kotz, S. y Johnson, N.L. (1982). *Encyclopedia of Statistical Science*. John Wiley and Sons: Nueva York.
- [15] Martorell R. (2002). Obesity in the Developing World. En: Caballero B, Popkin BM, (editors). The nutrition transition. *Diet and disease in the developing world*. San Diego : Academic Press; . pp. 147-164.
- [16] Montoya, J. A. (2008). *La Verosimilitud Perfil en la Inferencia Estadística*. Tesis de Doctorado en Ciencias con orientación en Probabilidad y Estadística, CIMAT, Guanajuato, Gto., México.

- [17] Mood, A. M., Graybill, F.A. y Boes, D.C. (1974). *Introduction to the Theory of Statistics*, Third Edition. McGraw-Hill.
- [18] Marron, S. & cols. (2004). *Variable Heavy Tails in Internet Traffic*. Performance Evaluation, 58, pp 261-264.
- [19] Nevill, A.M. y Holder, R.L. (1995). Body mass index: a measure of fatness or leanness?. *British Journal of Nutrition*. V. 73, pp. 507-516.
- [20] Norma Oficial Mexicana NOM-174-SSA1-1998. Para el manejo integral de la Obesidad, RESPYN, V. 4 No.1 Enero-Marzo 2003, pp. 1-9.
- [21] Oria E., Lafita J., Petrina E., Argüelles I. (2002). Composición Corporal y Obesidad. *Anales Sis San Navarra*, V. 25 (Supl. 1), pp. 91-102.
- [22] Pawitan, Y. (2004). *In all Likelihood: Statistical Modelling and Inference using Likelihood*. Oxford: Clarendon Press.
- [23] Quetelet, L.A.J. (1841). *A Treatise on Man and the Development of his Faculties*, Burt Franklin: Nueva York.
- [24] Rotimi, C. N., Cooper , R.S., Ataman, S.A., Osotimehin, B., Kadiri, S., Muna, W., Kingue, S., Fraser, H. y McGee, D. (1995). Distribution of Anthropometric Variables and the Prevalence of Obesity in Populations of West African Origin: The International Collaborative Study on Hypertension in Blacks (ICSHIB). *Obesity Research*. V. 3, pp. 95-105.
- [25] Seidell, J.C. Verschuren, W.M., Van Leer, E. M. y Kronhout, D. (1996). Overweight, underweight and mortality. A prospective study of 48,287 men and women. *Archives of Internal Medicine*. V. 156, pp. 958-963.
- [26] Sprott, D.A. (2000). *Statistical Inference in Science*. Nueva York: Springer-Verlag.

- [27] Sweeney ME. (2001) Composición corporal. Evaluación: instrumentos epidemiológicos, clínicos y de investigación. En: Obesidad: impacto en la enfermedad cardiovascular. Eds. American Heart Association, Futura Publishing Company (Ed española, Medical Trends). Barcelona. 129-137.
- [28] Valtueña S. Kehayias J. (2001). Determinación de la grasa corporal en vivo: de las técnicas bicompartimentales al análisis de la activación de neutrones y la absorciometría de rayos X de doble energía (DXA). *Med Clin (Barc)* V. 116, pp. 590-597.
- [29] Vázquez C. (1999) Epidemiología de la obesidad: estado actual en los países desarrollados. *Endocrinología*. V. 9, pp. 302-318.
- [30] Proceedings of the 2nd Fogarty International Centre (1975). Conference on Obesity. En Bray GA, ed. Obesity in perspective. Washington DC, US Department of Health, Education and Welfare.
- [31] Secretaría de Salud (SSA), México (2002). Manual de Procedimientos: Toma de medidas clínicas y antropométricas, en el adulto y el adulto mayor. Abril, pp. 30, 31, 50 y 51.
- [32] Sociedad Española para el Estudio de la Obesidad (SEEDO) (1996). Consenso español para la evaluación de la obesidad y estudios epidemiológicos. *Med Clin (Barc)*, V. 107, pp. 782-787.
- [33] Sociedad Española para el Estudio de la Obesidad (SEEDO) (2000). Consenso SEEDO para la evaluación del sobrepeso y la obesidad y el establecimiento de criterios de intervención terapéutica. *Med Clin (Barc)*, V. 115, pp. 587-597.
- [34] World Health Organization (WHO) (1995). El Estado Físico: Uso e interpretación de la Antropometría, *Informe de un Comité de Expertos de la OMS*. 854, Organización Mundial de la Salud. Ginebra.

- [35] World Health Organization (WHO) (1998), Programme of Nutrition, Family and Reproductive Health. Obesity: Preventing and managing the global epidemic. *Report of a WHO consultation on obesity*. (Geneva, 3-5 June, 1997). Ginebra.
- [36] World Health Organization (WHO) (2000), Obesity: preventing and managing the global epidemic. *Report of a WHO consultation*. Technical report series 894. Geneva.
- [37] World Health Organization (WHO) (2003), Diet, nutrition and the prevention of chronic diseases. *Report of a Joint FAO/WHO Expert consultation*. Technical report series 916. Geneva. American Petroleum Institute, Technical Data Book - Petroleum Refining, 5th edition, 1992.