

00365

2



**UNIVERSIDAD NACIONAL
AUTONOMA DE MEXICO**

**FACULTAD DE CIENCIAS
DIVISION DE ESTUDIOS DE POSGRADO**

**ANALISIS ESTADISTICO DE LA DESNUTRICION
INFANTIL EN EL MEDIO RURAL EN MEXICO.**

T E S I S

QUE PARA OBTENER EL GRADO ACADEMICO DE:

**MAESTRO EN CIENCIAS
(M A T E M A T I C A S)**

P R E S E N T A:

RUTH SELENE FUENTES GARCIA

DIRECTOR DE TESIS:

DRA. GUILLERMINA ESLAVA GOMEZ

MEXICO. D. F.

2000

282159



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**Análisis estadístico de la desnutrición infantil en el
medio rural en México.**

Ruth Selene Fuentes García

Índice General

1	Introducción	2
§1.1	ENAL '96	3
§1.1.1	Antecedentes	3
§1.1.2	Objetivos	4
§1.1.3	Descripción de la Encuesta	4
2	Análisis Exploratorio	12
§2.1	Indicadores del Estado Nutricional	12
§2.2	Casos Válidos para el Análisis	14
§2.3	Descripción de Variables	15
§2.4	Selección de Variables	20
3	Modelos de Regresión Lineal	38
§3.1	Consideraciones generales del modelo	39
§3.2	Modelos a partir de variables observadas	43
§3.3	Modelos a partir de índices resumen y variables observadas	59
§3.4	Modelos de regresión para un estado en cada región	69
§3.5	Modelos de regresión polinomiales	76
4	Efectos de diseño en los modelos de regresión.	81
§4.1	Inferencias basadas en modelos.	82
§4.2	Inferencias basadas en el diseño.	84
§4.3	Inferencias utilizando pesos muestrales para la ENAL'96	87
5	Conclusiones y Comentarios	97
	Apéndice A	101
	Apéndice B	102
	Apéndice C	104
	Bibliografía	112

Resumen

A partir de la información de la Cuarta Encuesta Nacional de la Nutrición en el Medio Rural en México (ENAL'96), realizada por el Instituto Nacional de la Nutrición Salvador Zubirán en 1996, dirigida a población infantil de niños menores de cinco años de edad, se analizó la información del entorno familiar con relación al peso para la edad, considerado para fines de este trabajo como un indicador del estado de nutrición del menor.

Se presenta un análisis exploratorio de variables referentes al hogar del menor, su nivel socioeconómico, las características de su madre y de los menores entrevistados. Una vez seleccionadas las variables que reflejan cambios con respecto al indicador del estado de nutrición, se construyen índices lineales que resumen aspectos de los rubros antes mencionados. Se exploraron diversos modelos de regresión lineal y polinomial, con el fin de explicar el estado de desnutrición infantil a partir de dicha información. El análisis considera la regionalización del país en tres regiones de acuerdo a grados de desnutrición en el medio rural del país. También se considera el efecto del diseño en los modelos propuestos proponiendo estimadores ponderados a través de los pesos muestrales. Sin embargo, ningún modelo logra predecir dicho estado de desnutrición con un buen grado de precisión aunque si describe relaciones de dependencia. Finalmente se proponen algunos modelos plausibles que auxilian en la descripción y explicación de diversos factores que influyen en los niveles de desnutrición infantil.

Capítulo 1

Introducción

Los países latinoamericanos tienen en la desnutrición infantil un serio problema de salud. En México el interés de instituciones de salud y gubernamentales ha dado origen a emprender estudios que busquen detectar factores de riesgo, así como ubicar geográficamente las zonas de riesgo. Esto permite emprender acciones que mejoren las condiciones de nutrición y evaluar sus resultados. La ENAL'96 forma parte de los estudios que ha realizado el Instituto Nacional de la Nutrición Salvador Zubirán, en áreas rurales en México. Esta encuesta se describe en este capítulo.

Existen numerosos estudios que buscan ubicar los factores de riesgo asociados a la desnutrición infantil valiéndose de análisis estadísticos. Para esto se utilizan, como ha propuesto la Organización Mundial de la Salud, medidas antropométricas como indicadores del estado de nutrición. En los modelos los indicadores se consideran como las variables respuesta y se analiza respecto a variables socioeconómicas y demográficas.

En este trabajo se analizó la información del entorno familiar con relación al peso para la edad, considerado para fines de este trabajo como un indicador del estado de nutrición del menor con el fin de exhibir la relación que guardan o como lo afectan. En el segundo capítulo se presenta un análisis exploratorio de las variables que se han considerado, así como de la relación que guardan con el peso para la edad de los menores.

En un tercer capítulo se analizan y proponen modelos de regresión plausibles que auxilian en la descripción y explicación de diversos factores que influyen en los niveles de desnutrición infantil. El análisis considera la regionalización del país en tres regiones de acuerdo a grados de desnutrición en el

medio rural del país.

Frecuentemente en el área de la salud la información que se obtiene proviene de encuestas, que por diversas razones, tienen diseños complejos. En este caso, los supuestos de los modelos planteados se ven afectados por el diseño estratificado trietápico de la encuesta. Los efectos del diseño se consideran en el cuarto capítulo. Una vez descritos los enfoques para incluirlos se hace uso de los factores de expansión para proponer estimadores ponderados a través de los pesos muestrales.

§1.1 ENAL '96

§1.1.1 Antecedentes

A partir de 1958 en Instituto Nacional de la Nutrición Salvador Zubirán (INNSZ) inició el estudio sistemático de las características, distribución y magnitud de la desnutrición en México. En la década de los 70 se intentó establecer un Programa Nacional de Alimentación dentro de la estructura del Consejo Nacional para la Ciencia y Tecnología con el fin de enfrentar los graves problemas de desnutrición del país. Buscando proveer información actualizada y a su vez confiable, el Instituto Nacional de Nutrición llevó a cabo la primera encuesta Nacional de Alimentación en el Medio Rural en 1974, ENAL 74. A pesar de los obstáculos que constituyeron las crisis políticas y económicas, así como la falta de continuidad en los programas sociales de los nuevos gobiernos, la serie de Encuestas Nacionales de Alimentación en el Medio Rural de 1974, 1979 y 1989 aportan información que permite comparar los niveles de desnutrición, su distribución geográfica y su evolución en años recientes. Por esta razón se considero conveniente llevar a cabo una cuarta encuesta.

La Cuarta Encuesta Nacional de Alimentación en el Medio Rural Mexicano levantada en 1996, publicándose sus resultados en 1997, *ENAL 96*, constituye un importante esfuerzo interinstitucional en el que el DIF, IMSS-Solidaridad, el Instituto Nacional Idigenista, Secretaria de Salud, los gobiernos de los estados y las delegaciones de SEDESOL, apoyaron al INNSZ para la realización de la encuesta, como instrumento de diagnóstico, planeación y evaluación de los programas sociales y de combate a la pobreza.

§1.1.2 Objetivos

Los objetivos generales de la ENAL 96, como lo señalan los coordinadores de esta encuesta en un documento interno¹ fueron fundamentalmente:

- Conocer la situación actual alimentaria y nutricional del medio rural mexicano.
- Comparar las condiciones actuales con las reportadas en las encuestas de 1974, 1979 y 1989.
- Identificar zonas en condiciones críticas de alimentación y nutrición.
- Identificar factores de riesgo asociado con desnutrición en las comunidades y familias estudiadas.
- Coadyuvar al establecimiento y perfeccionamiento de sistemas locales de vigilancia epidemiológica de la nutrición.

§1.1.3 Descripción de la Encuesta

La ENAL 96 recabó información a través de una encuesta que se dividió en 10 cédulas que se ocuparon de los siguientes aspectos:

1. Identificación.
2. **Características de la Vivienda y Saneamiento.**
3. **Recursos para la Alimentación Familiar.**
4. **Composición Familiar.**
5. Migración.
6. Actividades Agrícolas.
7. **Preescolares.**
8. Mujeres de 12 a 49 años.

¹ ÁVILA, A., SHAMAH, T., CHÁVEZ A . *Encuesta Nacional de Alimentación el Medio Rural 1996. (1997)*

9. Consumo Familiar de Alimentos.

10. Alimentación de la Mujer Responsable del Hogar el día anterior.

En estas cédulas se obtuvo información referente a las características socioeconómicas de las familias, indicadores antropométricos de los preescolares, lactancia, ablactación y morbilidad, indicadores de fecundidad y mortalidad infantil y consumo familiar de alimentos.

Las variables que se analizaron para estudiar su relación y posibles efectos sobre el estado de nutrición de los menores corresponden a variables de las cédulas o módulos 2 , 3 , 4 y 7 . Las cédulas del cuestionario se anexan en el apéndice A. Para el Módulo 2 se consideraron todas las variables² :

- Material de Techo (techo1) ,
 1. Palma o Cartón
 2. Madera o Teja
 3. Zinc o Asbesto
 4. Ladrillo o Concreto
 5. Otro
- Material de Paredes (pared1),
 1. Caña, Cartón o Madera Tosca
 2. Adobe
 3. Tabique sin aplanado
 4. Cemento, Yeso u Otro Aplanado
 5. Otro
- Material de Piso (piso1),
 1. Tierra
 2. Madera Tosca

²Los nombres de las variables en la base de datos se muestran entre paréntesis

3. Cemento sin pulir
 4. Mosaico, Madera o Cemento Pulidos
 5. Otro
- Lugar de donde obtiene agua para beber (agual),
 1. Depósito de Agua Estancada
 2. Depósito de Agua Corriente
 3. Hidrante Público
 4. Intradomiciliaria Entubada
 5. Otro
 - Disposición de Excretas de la Vivienda (excre1),
 1. Fecalismo al Ras de Suelo
 2. Letrina sin Arrastre de Agua
 3. Fosa Séptica
 4. Drenaje
 5. Otra
 - Lugar donde calienta sus alimentos (calent1),
 1. Fogón en Piso
 2. Fogón en Alto
 3. Estufa de Gas
 4. Otro
 - Tiene Cocina Separada (cocinal),
 1. Si
 2. No
 - Tiene ventilación (ventil),
 1. Si

- 2. No
- Tiene Animales Adentro (animal1),
 - 1. Si
 - 2. No
- Tiene Energía Eléctrica (elec1),
 - 1. Si
 - 2. No
- Número de cuartos que usan en su casa para dormir (nocuar1),
- Número de personas que duermen habitualmente en casa (duermel),
- Hay en la vivienda Radio (radio1),
 - 1. Si
 - 2. No
- Hay en la vivienda TV Blanco y Negro (TVBL1),
 - 1. Si
 - 2. No
- Hay en la vivienda TV a color (TVcol1),
 - 1. Si
 - 2. No
- Hay en la vivienda Refrigerador (refri1),
 - 1. Si
 - 2. No

En el caso del Módulo 2 las variables consideradas fueron:

- Gasto Semanal en Alimentos (gasto1),

- Cría Animales para Alimentación (criaa1),

1. Si
2. No

- Cultiva Alimentos (cultal1),

1. Si
2. No

Para el Módulo 4, se tomarán las variables que corresponden a la madre del menor entrevistado:

- Edad de la Madre (edadma1),

- Escolaridad (escola1),

1. Analfabeto
2. Sabe Leer y Escribir
3. Cursa Primaria o Primaria Incompleta
4. Primaria Completa
5. Secundaria Completa
6. Bachillerato o Equivalente
7. Carrera Técnica
8. Estudios Profesionales

- Idioma (idioma1),

1. Indígena
2. Español
3. Bilingüe

- Condición Laboral (condi1),

1. Asalariado
2. No asalariado

3. Propietario no Empleador
4. Propietario Empleador
5. Ama de Casa
6. Estudiante
7. Desempleado

Finalmente del Módulo 7 se estudiaron las variables de los menores de 5 años entrevistados³, con un máximo de 3 menores por hogar, vale la pena enfatizar que en algunos de los hogares entrevistados no se encontraron niños de la edad de interés del estudio. Las variables son:

- Edad (edad),
 1. Si
 2. No
- Alimentación al seno materno (seno),
 1. Si
 2. No
- Alimentación regular con biberón en los primeros 12 meses (biberon),
 1. Si
 2. No
- Edad de Ablactación (mesablacta),
- Estuvo enfermo en los últimos 15 días (enfermo),
- Enfermedad que padeció (enferme),
 1. Diarrea
 2. Infección Respiratoria
 3. Otra

³Es importante señalar que al analizar la base se encontraron niños mayores de 5 años y menores de 6 años.

La encuesta se dirigió a las familias residentes en las localidades rurales del país cuya población es de 500 a 2,499 habitantes con población económicamente activa dedicada principalmente al sector primario de acuerdo con la información del X Censo Nacional de Población y Vivienda, 1990 (INEGI). En la Cartografía⁴ de la encuesta, se establece que existen en el país 13,187 localidades rurales con población entre 500 y 2,500 habitantes, donde viven casi 13 millones de individuos. Es importante mencionar que hay una población de alrededor de 10 millones de residentes en localidades con población entre 1 y 499 habitantes.

El diseño muestral corresponde a un estratificado trietápico. Se contemplaron 5 grandes regiones como dominio de estudio para los tres indicadores del estado nutricional, peso para la edad, peso para la talla y talla para la edad. Las regiones⁵ quedaron conformadas como sigue:

1. Región Noroeste: Baja California, Baja California Sur, Chihuahua, Durango, Sinaloa y Sonora.
2. Región Noreste: Aguascalientes, Coahuila, Nuevo León, San Luis Potosí, Tamaulipas y Zacatecas.
3. Región Centro-Occidental: Colima, Guanajuato, Jalisco, Michoacán, Nayarit y Querétaro.
4. Región Centro: Hidalgo, Morelos, México, Puebla, Tlaxcala y Veracruz.
5. Región Sur: Campeche, Chiapas, Guerrero, Oaxaca, Querétaro, Tabasco y Yucatán.

⁴BARRAGÁN L., AMBROCIO, R., ÁVILA, A., ÁVILA, M.C. *Cartografía. Encuesta Nacional de Alimentación el Medio Rural 199.6*

⁵Por la definición de la población objetivo de la encuesta es claro que se excluye al Distrito Federal, por lo que se tienen los 31 estados restantes.

Cuadro 1.1. Número de menores encuestados por hogar.

	No. de Hogares entrevistados	No. de niños menores a 5 años en el hogar
	16679	0
	13061	1 (13061 casos)
	6936	2 (13872 casos)
	1556	3 (4668 casos)
Total	38232	31601 casos

La *unidad primaria de muestreo* la conformaron las localidades, en cada municipio o conjunto de municipios que conformaron los *estratos* (regiones). En cada uno de los 372 estratos se seleccionaron un mínimo de dos localidades. El tamaño de la muestra de localidades seleccionadas aleatoriamente fue 855, ubicadas en 649 municipios. La unidad de aplicación de la encuesta fue la *familia*, entendiendo por ésta a las personas que viven bajo el mismo techo, comparten el gasto y la alimentación doméstica. Se realizaron un total de **38,232** encuestas a nivel familiar y en una tercera etapa se encuestaron hasta tres de los menores de 5 años que hubiese en los hogares, entrevistando a un total de **31,601** menores.

Capítulo 2

Análisis Exploratorio

§2.1 Indicadores del Estado Nutricional

Un aspecto primordial para el desarrollo de este trabajo es la definición de *medidas antropométricas* como *indicadores del estado nutricional* de los menores. La Organización Mundial de la Salud¹ plantea que el problema nutricional más importante actualmente es el de desnutrición proteinoenergética. Por esa razón sugiere elegir las siguientes tres medidas para evaluar el estado de nutrición:

- i) Edad,
- ii) Peso,
- iii) Talla: longitud supina para niños menores de 2 años, talla en pie para niños de 2 años en adelante.

Estas tres mediciones se combinan para formar tres indicadores del estado nutricional:

- *Peso para la edad,*
- *Peso para la talla,*
- *Talla para la edad.*

¹ORGANIZACIÓN MUNDIAL DE LA SALUD. *Medición del Cambio del Estado Nutricional (1983).*

Para calcular estos indicadores se compara al individuo con una población de referencia. La OMS recomienda para este fin los datos recolectados por el Centro Nacional de Estadística Sanitaria de los Estados Unidos de América. Se ubica al individuo de acuerdo a su edad y sexo o sexo y talla, según sea el caso, en la población de referencia. Se construyen estadísticas con un proceso semejante a la estandarización de una variable normal, de ahí que se conozcan como *score* "z" de peso para la edad (**PEDZ**), peso para la talla (**PETZ**) y talla para la edad (**TEDZ**). Estas medidas de *score* z se calculan considerando a la población de referencia **A** como aquella que comprende individuos de la misma edad y sexo que el individuo evaluado y la población de referencia **B** la que está compuesta por individuos de la misma talla y sexo que el individuo a evaluar. Estas poblaciones de referencia se han considerado como las poblaciones de referencia recomendadas por la ONU² que corresponden a datos recabados por el Centro Nacional de Estadística Sanitaria de los Estados Unidos de América³. Se calculan a partir de las siguientes expresiones, considerando la distribución de la desviación estándar alrededor de la mediana:

PEDZ = (Peso del individuo - Mediana del peso en la población de referencia **A**) / Desviación estándar del peso en la población de referencia **A**.

PETZ = (Peso del individuo - Mediana del peso en la población de referencia **B**) / Desviación estándar del peso en la población de referencia **B**.

TEDZ = (Talla del individuo - Mediana de la talla en la población de referencia **A**) / Desviación estándar del peso en la población de referencia **A**.

El cálculo de los indicadores estuvo a cargo del INNSZ. Los valores de los *scores* se incluyeron en la base para preescolares.

²ORGANIZACIÓN MUNDIAL DE LA SALUD. *Medición del Cambio del Estado Nutricional (1983)*.

³PUBLIC HEALTH SERVICE. HEALTH RESOURCES ADMINISTRATION. *NCHS growth charts(1976)*.

La Organización Mundial de la Salud señala que el indicador de desnutrición proteinoenergética más aceptado y utilizado con mayor frecuencia es el peso para la edad. Resaltando que es conveniente y muy importante buscar precisión en el dato de la edad. En el caso de los niños encuestados en la ENAL 96, la edad se calculo a partir de la fecha de nacimiento.

Para el desarrollo del trabajo, apoyados en las observaciones que se discutieron con el Dr. Abelardo Ávila Curiel, investigador del departamento de nutrición del INNSZ, consideramos al *score z* del peso para la edad (**PEDZ**), como el indicador para el estado de nutrición de los niños menores de 5 años. Esto con el objeto de describir la relación entre el peso para la edad y las variables de los modulos considerdos.

§2.2 Casos Válidos para el Análisis

Para el análisis del estado de nutrición de los menores se unieron las bases correspondientes a los modulos de interés antes mencionados. La base de datos tiene un total de **31,601** registros que corresponden a los menores⁴ encuestados. Estos menores provienen de **21,553** hogares, se entrevistaron a lo más tres menores por hogar. En los **16,679** hogares encuestados restantes no se encontraron niños menores de 5 años.

En primer lugar se busco retirar del análisis los casos cuya variable PEDZ no era válida, ya sea como casos faltantes o por posibles errores de captura. Se deteminó como valores válidos para PEDZ aquellos que se encuentran entre -5 y 5 inclusive. Con este criterio se filtro la base, retirando 2,076 casos que corresponden al 6.6% de las observaciones. Se cuidó que estas observaciones no se concentraran en uno o varios estados. De los 2,076 casos eliminados el 16.2% pertenecen al Estado de México, el 10.5% a Guerrero y el 9.1% a Guanajuato. En el resto de los estados no se concentró más del 6% de los casos excluidos.

En un segundo filtro se retiraron 679 casos más, en donde la edad de los menores era mayor o igual 5 años. En este caso el 15.8% de estos casos pertenecía a Puebla, el 11.6% a Veracruz y el 9.7% a Guerrero. Para el resto

⁴En esta base se tienen menores de 6 años, aunque en el documento del INNSZ los reportes se refieren a niños menores de 5 años.

Cuadro 2.1. Número de menores encuestados por hogar.

Menores encuestados	Frecuencia	Porcentaje
1	11841	63.1
2	5820	31.0
3	1113	5.9
Total	26820	100

de los estados no se tienen porcentajes mayores a 5.3%.

Finalmente se excluyeron de la base los menores cuyas madres tenían en el registro de edad menos de 12 años y más de 50 años. En esta etapa de filtrado se retiraron 2,026 casos de los cuales el 11.8% son de Veracruz, el 9.9% de Guerrero y en el resto de los estados no se concentro más del 6.8%.

Con los filtros anteriores se tiene una base final con **26,820** niños encuestados que provienen de **18,774** encuestas a nivel familiar. En estos hogares se entrevistaron uno, dos o tres menores. El cuadro 2.1 muestra las frecuencias del número de menores encuestados.

§2.3 Descripción de Variables

Una vez establecidos los casos válidos a considerar se analizaron las variables agrupadas en cuatro bloques. En esta etapa se definieron los rangos de las variables continuas, construyendo en ocasiones variables discretas a partir de estas. También se crearon variables para resumir información recabada en varias variables de un módulo.

- Bloque A. Vivienda: techo1, piso1, pared1, agua1, excre1, calent1, cocina1, venit1, animal1, eelcel1, hacinamiento, criaail, cultall.
- Bloque B. Bienes materiales en el hogar: TVBL1, TVcoll1, radiol1, refril1,
- Bloque C. Nivel Socioeconómico: Gasto per capita (gastopc), actividad del jefe de hogar (actjf).
- Bloque D. Madre del Menor: edadma1, escola1, idioma1, condi1.

- Bloque E. Menor: edad, enfermo, enferme, mesablacta, uno, tres, seis, doce, intergen.

De las variables anteriores, es importante aclarar que se han incluido las variables hacinamiento, gastopc, uno, tres, seis, doce e intergen. El hacinamiento considera el número promedio de personas que duermen por cuarto en el hogar, el gasto semanal en alimentos se considera per capita (gastopc). Las variables uno, tres, seis, doce, son variables de tres categorías que consideran:

- Uno: Tipo de alimentación que recibió el niño los primeros doce meses. Para niños de 0 a 5 años.
- Tres: Tipo de alimentación que recibió el niño los primeros doce meses. Para niños de 3 meses a 5 años.
- Seis: Tipo de alimentación que recibió el niño los primeros doce meses. Para niños de 6 meses a 5 años.
- Doce: Tipo de alimentación que recibió el niño los primeros doce meses. Para niños de 1 a 5 años.

1. Pecho
2. Biberón
3. Mixto

Finalmente el *intervalo intergenésico* (intergen) corresponde a la distancia en meses de edad al hermano menor o igual a cinco años de edad más cercano. Además de estas variables se consideraron otras, ya sea de las cédulas originales como si tiene ventilación, energía eléctrica, si cría animales o cultiva alimentos o bien creadas para resumir información como, enfermo de diarrea, hijo único, etcétera. Se decidió excluirlos del análisis pues no presentan asociación o cambios respecto a las categorías de peso para la edad, PEDZ.

Es importante enfatizar que las variables consideradas en cada bloque se analizan con el objeto de encontrar las posibles relaciones que guarden con el indicador del estado nutricional, que para fines de este trabajo es el peso

para la edad PEDZ. Por ello, en primer lugar se describe a dicho indicador según los datos recabados en la ENAL 96.

El peso para la edad que hemos considerado toma valores entre $[-5, 5]$ y en la gráfica 2.1 se puede observar el histograma de esta variable.

A partir de la variable PEDZ se han creado dos variables discretas: *PEDZ2* y *PEDZ3*. Para discretizar a la variable se utilizó el criterio propuesto por la OMS. Se consideran los rangos definidos de acuerdo al cuadro 2.2 del *score z* para definir el estado de nutrición de los menores en la variable *PEDZ2* y de acuerdo al cuadro 2.3 para definir el estado de nutrición de los menores en la variable *PEDZ3*.

Cuadro 2.2. Estado de Nutrición de acuerdo al *score z* de Peso para la edad.

Rangos para PEDZ2	Estado Nutricional
$PEDZ > -2$	Normal
$PEDZ \leq -2$	Desnutrición Severa

Cuadro 2.3. Estado de Nutrición de acuerdo al *score z* de Peso para la edad.

Rangos para PEDZ3	Estado Nutricional
$PEDZ \geq -1$	Normal
$-1 > PEDZ > -2$	Desnutrición Leve
$PEDZ \leq -2$	Desnutrición Severa

Los cuadros 2.4 y 2.5 muestra las frecuencias para los menores en la muestra de casos válidos de las variables *PEDZ* y *PEDZ3*.

Cuadro 2.4. Tabla de Frecuencias de la variable PEDZ2.

	Frecuencia	Porcentaje
1 Desnutrición Severa	4742	17.7
2 Normal	22078	82.3
Total	26820	100.0

Cuadro 2.5. Tabla de Frecuencias de la variable PEDZ3.

	Frecuencia	Porcentaje
1 Desnutrición Severa	4742	17.7
2 Desnutrición Leve	7104	26.5
3 Normal	14974	55.8
Total	26820	100.0

Ahora se presentan los análisis análogos de las variables continuas: Edad de la madre, gasto per cápita en alimentos, hacinamiento, edad de ablactación de los menores e intervalo intergenésico, estas dos últimas medidas en meses.

Para la edad de la madre, se ha establecido como rango válido el intervalo [12, 50]. La gráfica 2.2 muestra el histograma correspondiente a esta variable. Se propone además una clasificación en cuatro categorías: Menos de 18 años, [18, 30) años, [30, 40) años y Más de 40 años.

El *gasto per cápita* semanal en alimentos, cuyo histograma se presenta en la gráfica 2.3, toma valores en el intervalo (0.20, 265). Esta variable se agrupa en tres categorías: Menor a 15 pesos, [15, 30) pesos y 30 pesos ó Más.

En lo que se refiere al *índice de hacinamiento* se encuentra en una rango que va de 0.14 a 14 personas por cuarto. La gráfica 2.4 corresponde a su histograma de frecuencias.

También se considero la *edad de ablactación* en meses, esta variable toma valores de 0 a 40 meses y su histograma se muestra en la gráfica 2.5, para esta variable también se construyo una variable discreta. En este caso se proponen como edades de ablactación tempranas a aquellas menores o iguales a tres

meses, oportunas a aquellas mayores a tres y menores a seis meses y tardías a aquellas de seis meses o más.

Para el *intervalo intergenésico*, que desde el punto de vista médico se considera saludable a partir de 24 meses, se tienen 4,116 niños en los hogares entrevistados que si tienen hermanos menores a seis años. Es decir, el intervalo intergenésico toma valores desde cero, en el caso de ser hermanos gemelos hasta 70 meses. Su histograma de frecuencias se muestra en la gráfica 2.6.

Además de las variables continuas se analizó, inicialmente, la rama de actividad y condición laboral de la madre con el fin de crear una variable para la actividad económica. Sin embargo, más del 90% de ellas son amas de casa. Por ello se construyó una variable categórica que describiera la *actividad* y *estatus del jefe de hogar*, que no necesariamente es el padre de los menores entrevistados. Esta variable se creo tomando en cuenta las variables de rama de actividad y condición laboral del jefe de hogar, variables que forman parte del módulo 4.

Por otra parte, con base en un análisis de conglomerados para el índice del estado nutricional⁵, que consiste en las tres variables PEDZ, TEDZ y PETZ se dividió al territorio nacional en tres regiones. El análisis se llevó a cabo utilizando el método de Ward, con distancia euclidiana. A partir de los primeros dos componentes principales, que explican alrededor del 90% de la varianza, se propuso una división del territorio nacional que presenta continuidad geográfica.

- I. Región Noreste y Norte-occidente: Aguascalientes, Baja California, Baja California Sur, Chihuahua, Coahuila, Colima, Durango, Jalisco, Michoacán, Nayarit, Nuevo León, Sinaloa, Sonora, Tamaulipas y Zacatecas.
- II. Región Centro: Estado de México, Guanajuato, Morelos, Querétaro, San Luis Potosí, Tabasco, Tlaxcala y Veracruz.
- III. Región Centro-sur: Campecha, Chiapas, Guerrero, Hidalgo, Oaxaca, Puebla, Quintana Roo, Yucatán.

⁵ESLAVA GÓMEZ G. ET AL. *Magnitud y distribución geográfica de la desnutrición infantil en el medio rural de México (Somatido)*.

Antes de ocuparse de la relación que existe entre las variables descritas en esta sección y las variables categóricas que corresponden al indicador del estado nutricional, es necesario abordar otro problema referente a la información que se está analizando.

En 11,841 de los hogares solo se entrevistó un menor por hogar, en 5,820 de los hogares se entrevistaron 2 menores en cada hogar y en 1,113 de los hogares se entrevistaron a 3 menores de cinco años. Por ello en 6,933 hogares los menores comparten las variables que se refieren a su vivienda y en caso de ser hermanos también las que se refieren a su madre. Tenemos entonces conglomerados de tamaño muy pequeño entre 1 y 3. Es por ello que para abordar el problema de correlación entre las observaciones se tomó una muestra aleatoria de un menor por hogar. Esta muestra se llevó a cabo con igual probabilidad de selección al interior del conglomerado. Esto es en los hogares con dos menores entrevistados se seleccionó uno con probabilidad un medio y en caso de hogares con tres menores se seleccionó uno con probabilidad un tercio.

Se tiene entonces una base con 18,774 menores que corresponden a un menor por hogar en la muestra de casos válidos.

§2.4 Selección de Variables

El objetivo de esta sección es considerar las variables para los cuatro bloques que se han establecido, con el fin de seleccionar aquellas que mantengan una relación de importancia con respecto al indicador del estado nutricional PEDZ.

Una vez contando con una base que contiene la información de 18,774 menores se obtuvieron tablas de contingencia para las variables discretas que se describen en la sección anterior respecto a la variable PEDZ3, es decir, el indicador del estado nutricional en tres categorías, esto con el fin de ubicar a las variables que presentan cambios importantes en su distribución respecto a PEDZ3. También se incluyen las gráficas de las variables continuas contra el *score z* del peso para la edad.

En el caso de las variables discretas, para algunas de éstas se colapsaron categorías ya que su comportamiento respecto al indicador del estado

nutricional era similar y se encontraron pocos casos en las categorías de la propuesta inicial de categorías. En otras se resumió la información de dos variables en una sola, este es el caso de la variable televisión en la que se consideró la presencia o ausencia del aparato televisor, ya fuese en blanco y negro o en color.

Se presentan ahora las tablas y gráficas para las variables en las que se consideran importantes en relación al estado nutricional de los menores. En las gráficas de las variables continuas contra el peso para la edad, puede observarse que los valores de esta última variable que corresponden a distintas categorías del estado nutricional, no se concentran en alguna zona de valores para la primera variable.

Sin embargo, a partir de las tablas de contingencia si puede pensarse en una asociación que podría considerarse significativa del peso para la edad con variables como gasto per capita semanal, tabla 1.7 y edad de ablactación en rangos, tabla 1.9. No así en el caso de la edad de la madre, tabla 1.5. Para algunas de las variables discretas restantes, también se encuentran asociaciones como sucede con: lugar donde calienta sus alimentos, lugar de donde obtiene agua, escolaridad de la madre, disposición de excretas en la vivienda, idioma de la madre, material de construcción de piso, techo y paredes, región del país (Norte, Norte-Occidente; Centro y Centro-Sur), cuenta con refrigerador y televisión. Algunas de estas variables tienen un comportamiento muy similar en relación a PEDZ3, como las variables referentes a los materiales de construcción, las que describen a si cuenta con televisión y con refrigerador. Por otro lado se tienen las variables como escolaridad de la madre, idioma de la madre, edad de ablactación del menor, tipo de alimentación en la edad de interés, que reflejan cierta asociación con el estado de nutrición.

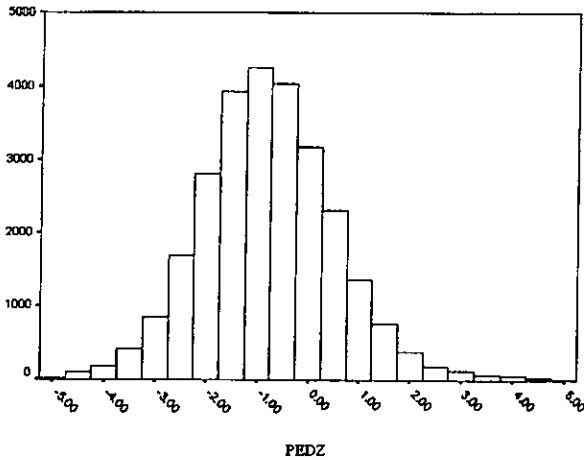
Ahora bien, ya que el estado de nutrición de los niños menores a 12 meses es en general normal, como puede observarse en la gráfica 2.7, que presenta el promedio del peso para la edad trimestral por región y global. Con el fin de que variables como edad de la madre y tipo de alimentación puedan reflejar una asociación, en caso de haberla, se seleccionó una muestra aleatoria de un menor por hogar cuya edad fuese mayor o igual a 12 meses.

Se consideraron ahora 21,755 con edades a partir de 12 meses y se seleccionó un niño por hogar. Esta base tiene un total de 16,604 casos. Pero

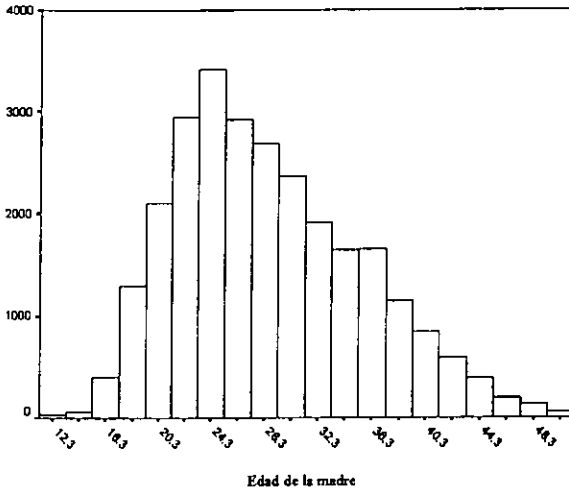
las variables no presentan cambios importantes que indiquen la presencia de una mayor asociación entre éstas y el peso para la edad.

El interés de este trabajo es ahora buscar, a partir de estas variables analizadas, un modelo que pueda describir o identificar, en lo posible, las relaciones de éstas con el valor del *score z* para el peso para la edad o la categoría correspondiente al estado de nutrición del menor.

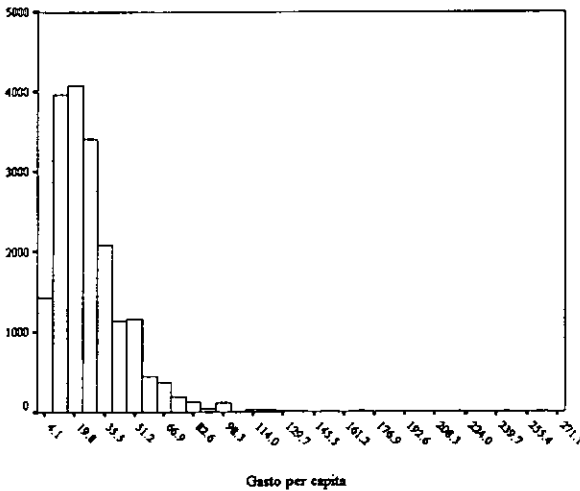
Gráfica 2.1: Histograma de peso para la edad (*Score Z*).



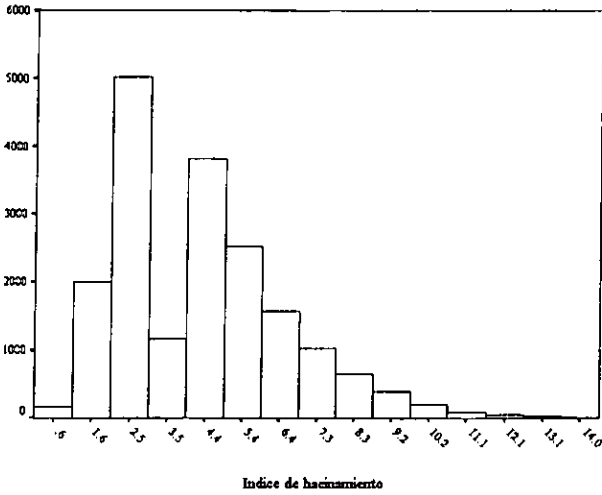
Gráfica 2.2: Histograma de edad de la madre (al momento de la entrevista).



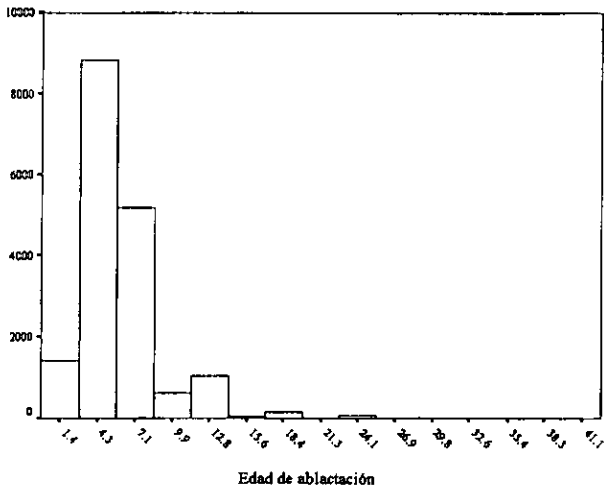
Gráfica 2.3: Histograma para el gasto per cápita semanal en alimentos.



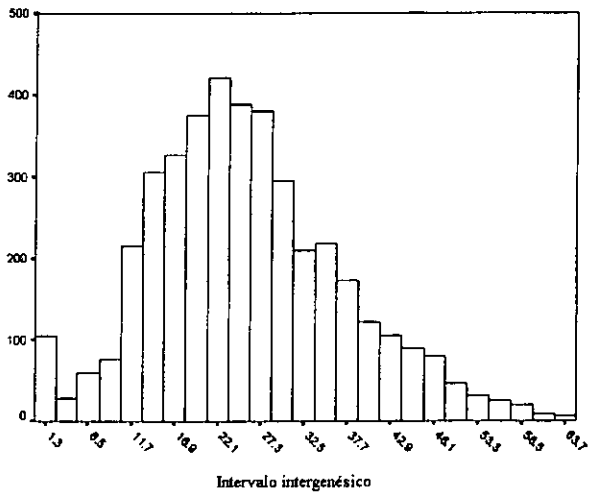
Gráfica 2.4: Histograma para el índice de hacinamiento.



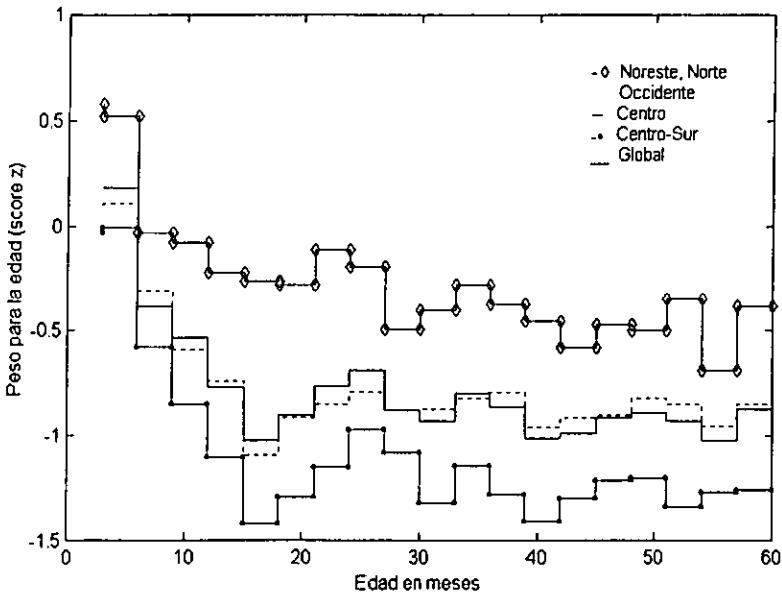
Gráfica 2.5: Histograma para la edad de ablactación en meses.



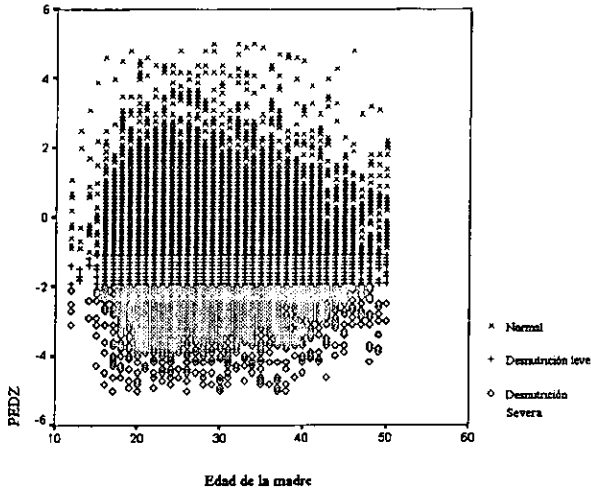
Gráfica 2.6: Histograma para el intervalo intergenésico.



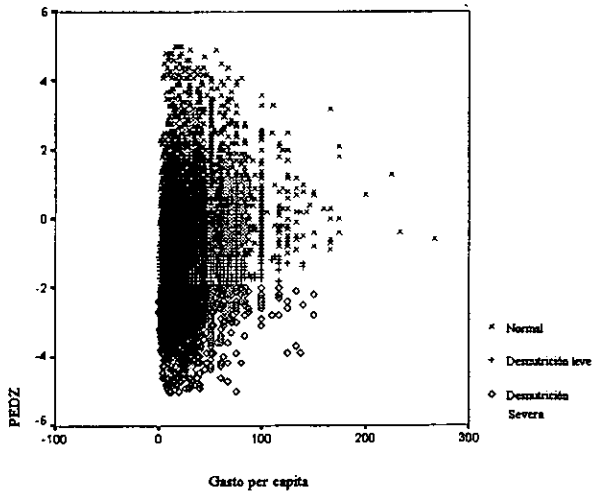
Gráfica 2.7: Promedio de peso para la edad (PEDZ3) por grupo de edad.



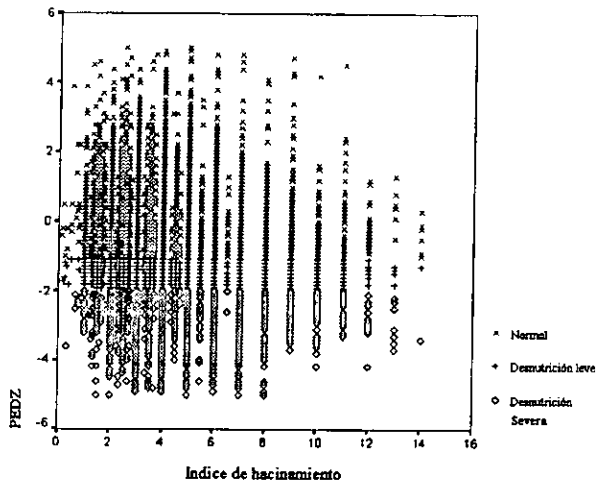
Gráfica 2.8: Peso para la edad (score z) vs edad de la madre.



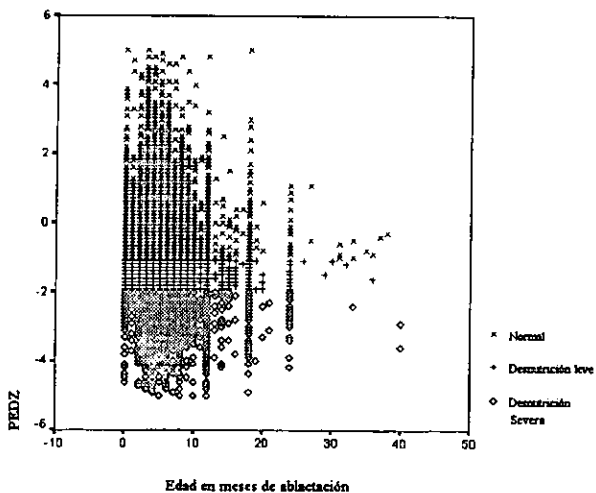
Gráfica 2.9: Peso para la edad (score z) vs gasto per capita semanal en alimentos.



Gráfica 2.10: Peso para la edad (score z) vs índice de hacinamiento.



Gráfica 2.11: Peso para la edad (score z) vs edad de ablactación.



Gráfica 2.12: Peso para la edad (score z) vs intervalo intergenésico.

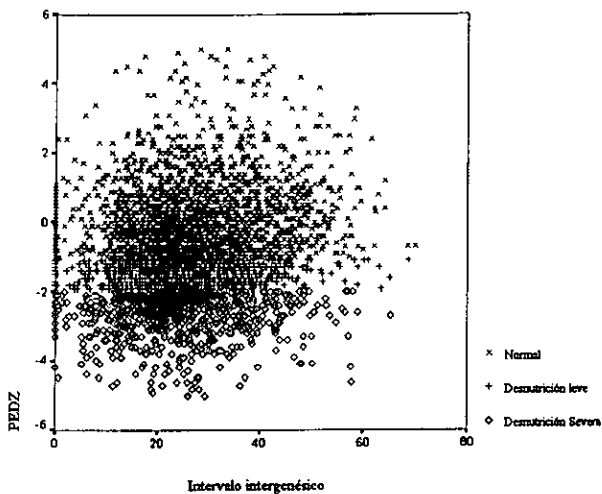


Tabla 2.1. Indicador nutricional por Lugar de donde obtiene el agua para beber.

Peso para la edad tricotómica	Lugar de donde obtiene el agua para beber		Total
	No Intradomiciliaria	Intradomiciliaria entubada	
Desnutrición Severa			
Frecuencia	1496	1538	3034
Porcentaje	18.7%	15.3%	16.8%
Desnutrición Leve			
Frecuencia	2236	2478	4714
Porcentaje	28.0%	24.7%	26.2%
Normal			
Frecuencia	4258	6018	10276
Porcentaje	53.3%	60.0%	57.0%
Total			
Frecuencia	7990	10034	18024
Porcentaje	44.3%	55.7%	100.0%

Tabla 2.2. Indicador Nutricional por Lugar donde calienta sus alimentos.

Peso para la edad tricotómica	Lugar donde calienta sus alimentos			Total
	Fogón en piso	Fogón en alto	Estufa de gas	
Desnutrición Severa				
Frecuencia	771	1525	812	3108
Porcentaje	26.6%	20.1%	9.9%	16.7%
Desnutrición Leve				
Frecuencia	873	2237	1741	4851
Porcentaje	30.1%	29.5%	21.3%	26.0%
Normal				
Frecuencia	1257	3817	5611	10685
Porcentaje	43.3%	50.4%	68.7%	57.3%
Total				
Frecuencia	2901	7579	8164	18644
Porcentaje	15.6%	40.7%	43.8%	100.0%

Tabla 2.3. Indicador nutricional por Cuenta con cocina separada.

Peso para la edad tricotómica	Cocina separada		Total
	Si	No	
Desnutrición Severa			
Frecuencia	2476	645	3121
Porcentaje	16.0%	19.8%	16.7%
Desnutrición Leve			
Frecuencia	4024	845	4869
Porcentaje	26.0%	26.0%	26.0%
Normal			
Frecuencia	8972	1761	10733
Porcentaje	58.0%	54.2%	57.3%
Total			
Frecuencia	15472	3251	18723
Porcentaje	82.6%	17.4%	100.0%

Tabla 2.4. Indicador Nutricional por Grado de escolaridad de la madre.

Peso para la edad tricotómica	Grado de escolaridad de la madre				Total
	Analfabeto	Sabe leer y escribir	Primaria incompleta o cursándola	Primaria completa	
Severa					
Frecuencia	693	487	883	1003	3066
Porcentaje	26.2%	21.1%	16.4%	12.2%	16.5%
Leve					
Frecuencia	822	618	1512	1879	4831
Porcentaje	31.1%	26.8%	28.1%	22.9%	26.1%
Normal					
Frecuencia	1127	1202	2990	5316	10635
Porcentaje	42.7%	52.1%	55.5%	64.8%	57.4%
Total					
Frecuencia	2642	2307	5385	8198	18532
Porcentaje	14.3%	12.4%	29.1%	44.2%	100.0%

Tabla 2.5. Indicador Nutricional por Edad de la Madre.

Peso para la edad tricotómica	Edad de la madre				Total
	Menos de 18	[18-30)	[30-40)	Más de 41	
Severa					
Frecuencia	72	1670	1068	320	3130
Porcentaje	17.2%	16.1%	17.0%	18.8%	16.7%
Leve					
Frecuencia	98	2706	1626	456	4886
Porcentaje	23.4%	26.1%	25.9%	26.8%	26.0%
Normal					
Frecuencia	249	6007	3576	926	10758
Porcentaje	59.4%	57.9%	57.0%	54.4%	57.3%
Total					
Frecuencia	419	10383	6270	1702	18774
Porcentaje	2.2%	55.3%	33.4%	9.1%	100.0%

Tabla 2.6. Indicador Nutricional por Disposición de excretas en la vivienda.

Peso para la edad tricotómica	Disposición de excretas en la vivienda			Total
	Fecalismo al ras del suelo	Letrina sin arrastre de agua	Fosa séptica o drenaje	
Desnutrición Severa				
Frecuencia	1434	1066	615	3115
Porcentaje	21.7%	16.7%	10.8%	16.7%
Desnutrición Leve				
Frecuencia	1891	1751	1220	4862
Porcentaje	28.7%	27.4%	21.5%	26.1%
Normal				
Frecuencia	3271	3571	3843	10685
Porcentaje	49.6%	55.9%	67.7%	57.3%
Total				
Frecuencia	6596	6388	5678	18662
Porcentaje	35.3%	34.2%	30.4%	100.0%

Tabla 2.7. Indicador Nutricional por Gasto per capita semanal en alimentos.

Peso para la edad tricotómica	Gasto per capita			Total
	Menor a 15	[15-30)	30 ó más	
Desnutrición Severa				
Frecuencia	1140	1178	779	3097
Porcentaje	22.8%	17.0%	11.6%	16.6%
Desnutrición Leve				
Frecuencia	1531	1919	1421	4871
Porcentaje	30.6%	27.6%	21.1%	26.1%
Normal				
Frecuencia	2334	3851	4530	10715
Porcentaje	46.6%	55.4%	67.3%	57.4%
Total				
Frecuencia	5005	6948	6730	18683
Porcentaje	28.6%	37.2%	36.0%	100.0%

Tabla 2.8. Indicador nutricional por Idioma de la madre.

Peso para la edad tricotómica	Idioma		Total
	Indígena o bilingüe	Español	
Desnutrición Severa			
Frecuencia	1085	2005	3090
Porcentaje	28.4%	13.7%	16.7%
Desnutrición Leve			
Frecuencia	1163	3650	4813
Porcentaje	30.5%	24.9%	26.0%
Normal			
Frecuencia	1570	9017	10587
Porcentaje	41.1%	61.5%	57.3%
Total			
Frecuencia	3818	14672	18490
Porcentaje	20.6%	79.4%	100.0%

Tabla 2.9. Indicador Nutricional por Edad de ablactación.

Peso para la edad tricotómica	Edad de ablactación			Total
	Temprana	Oportuna	Tardía	
Desnutrición Severa				
Frecuencia	633	788	1574	2995
Porcentaje	12.4%	15.3%	22.0%	17.2%
Desnutrición Leve				
Frecuencia	1210	1328	2136	4674
Porcentaje	23.7%	25.7%	29.9%	26.9%
Normal				
Frecuencia	3253	3042	3431	9726
Porcentaje	63.8%	59.0%	48.0%	55.9%
Total				
Frecuencia	5096	5158	7141	17395
Porcentaje	29.3%	29.7%	41.1%	100.0%

Tabla 2.10. Indicador Nutricional por Material de las paredes.

Peso para la edad tricotómica	Material de las paredes			Total
	Caña, cartón o madera tosca	Adobe tabique sin aplanado	Cemento, yeso u otro aplanado	
Severa				
Frecuencia	960	1605	432	2997
Porcentaje	21.4%	16.5%	10.9%	16.5%
Leve				
Frecuencia	1338	2523	872	4733
Porcentaje	29.8%	26.0%	22.0%	26.0%
Normal				
Frecuencia	2197	5578	2667	10442
Porcentaje	48.9%	57.5%	67.2%	57.5%
Total				
Frecuencia	4495	9706	3971	18172
Porcentaje	24.7%	53.4%	21.9%	100.0%

Tabla 2.11. Indicador Nutricional por Material del piso.

Peso para la edad tricotómica	Material del piso			Total
	Tierra	Madera tosca u otro	Cemento mosaico o madera	
Desnutrición Severa				
Frecuencia	1700	50	1372	3122
Porcentaje	22.5%	20.7%	12.6%	16.7%
Desnutrición Leve				
Frecuencia	2186	62	2621	4869
Porcentaje	28.9%	25.7%	24.0%	26.0%
Normal				
Frecuencia	3679	129	6926	10734
Porcentaje	48.6%	53.5%	63.4%	57.3%
Total				
Frecuencia	7565	241	10919	18725
Porcentaje	40.4%	1.3%	58.3%	100.0%

Tabla 2.12. Indicador nutricional por Cuenta con refrigerador.

Peso para la edad tricotómica	Refrigerador		Total
	Si	No	
Desnutrición Severa			
Frecuencia	503	2604	3107
Porcentaje	9.6%	19.4%	16.7%
Desnutrición Leve			
Frecuencia	1054	3803	4857
Porcentaje	20.2%	28.3%	26.0%
Normal			
Frecuencia	3663	7029	10692
Porcentaje	70.2%	52.3%	57.3%
Total			
Frecuencia	5220	13436	18656
Porcentaje	28.0%	72.0%	100.0%

Tabla 2.13. Indicador Nutricional por Región.

Peso para la edad tricotómica	Región			Total
	Noroeste, Norte-Occidente	Centro	Centro-Sur	
Severa				
Frecuencia	438	898	1794	3130
Porcentaje	8.7%	14.5%	23.7%	16.7%
Leve				
Frecuencia	937	1697	2252	4886
Porcentaje	18.7%	27.4%	29.8%	26.0%
Normal				
Frecuencia	3636	3608	3514	10758
Porcentaje	72.6%	58.2%	46.5%	57.3%
Total				
Frecuencia	5011	6203	7560	18774
Porcentaje	26.7%	33.0%	40.3%	100.0%

Tabla 2.14. Indicador Nutricional por Material del techo.

Peso para la edad tricotómica	Material del techo			Total
	Palma o cartón	Madera, teja, zinc o asbesto	Ladrillo o concreto	
Desnutrición Severa				
Frecuencia	956	1617	470	3043
Porcentaje	21.6%	17.3%	10.5%	16.7%
Desnutrición Leve				
Frecuencia	1269	2483	1010	4762
Porcentaje	28.6%	26.6%	22.6%	26.1%
Normal				
Frecuencia	2205	5242	2988	10455
Porcentaje	49.8%	56.1%	66.9%	57.2%
Total				
Frecuencia	4430	9342	4478	18250
Porcentaje	24.3%	51.2%	24.5%	100.0%

Tabla 1.15. Indicador nutricional por Cuenta con televisión.

Peso para la edad tricotómica	Televisión		Total
	Si	No	
Desnutrición Severa			
Frecuencia	1623	1482	3105
Porcentaje	13.6%	22.2%	16.7%
Desnutrición Leve			
Frecuencia	2811	2024	4835
Porcentaje	23.6%	30.3%	26.0%
Normal			
Frecuencia	7460	3179	10639
Porcentaje	62.7%	47.6%	57.3%
Total			
Frecuencia	11894	6685	18579
Porcentaje	64.0%	36.0%	100.0%

Tabla 1.16. Indicador nutricional por
Actividad del Jefe de Hogar.

Peso para la edad tricotómica	Actividad del Jefe de Hogar					Total
	Peon	Ejida- tario	Obrero, Empleado o Trabajador por su Cuenta	Dueño o Emplea- dor	Otros	
Severa						
Frecuencia	679	928	429	680	191	2908
Porcentaje	16.3%	18.9%	13.2%	19.0%	13.0%	16.7%
Leve						
Frecuencia	1072	1405	751	1039	307	4574
Porcentaje	25.7%	28.5%	23.2%	29.0%	29.0%	26.3%
Normal						
Frecuencia	2421	2593	2058	1868	970	9910
Porcentaje	58.0%	52.6%	63.6%	52.1%	66.1%	57.0%
Total						
Frecuencia	4172	4927	3238	3587	1468	17392
Porcentaje	24.0%	28.3%	18.6%	20.6%	8.4%	100.0%

Tabla 1.17. Indicador nutricional por Tipo de alimentación
en los primeros 12 meses del menor de 1 a 5 años de edad.

Peso para la edad tricotómica	Tipo de alimentación			Total
	Mixta	Pecho	Biberón	
Desnutrición Severa				
Frecuencia	877	1621	240	2738
Porcentaje	14.0%	23.6%	13.4%	18.4%
Desnutrición Leve				
Frecuencia	1615	2147	431	4193
Porcentaje	25.8%	31.3%	24.1%	28.1%
Normal				
Frecuencia	3768	3093	1118	7979
Porcentaje	60.2%	45.1%	62.5%	53.5%
Total				
Frecuencia	6260	6861	1789	14910
Porcentaje	42.0%	46.0%	12.0%	100.0%

Capítulo 3

Modelos de Regresión Lineal

México, como muchos otros países en vías de desarrollo, tiene en la desnutrición infantil un serio problema de salud pública. La desnutrición de los niños en edades preescolares ha sido objeto de varios estudios. Existe interés en el déficit del peso para la edad, la talla para la edad y la emaciación o déficit de peso para la talla, como indicadores del status nutricional del menor y su asociación con diversas variables socioeconómicas y demográficas. Esto se debe a que al identificar las variables que pueden incidir o favorecer la desnutrición infantil, así como determinar si existen zonas geográficas o familias que sean notablemente afectadas en términos de desnutrición infantil, puede ayudar a proponer políticas de alimentación y nutrición, objetivos que han planteado los coordinadores de la ENAL'96. En México, como en otros países, se han tomado acciones relacionadas con la utilización de suplementos alimenticios para combatir la desnutrición. Las variables antes mencionadas pueden ayudar a evaluar y diseñar estos programas de apoyo.

En algunos estudios como son los realizados por Rivera et al (1991), Pelto et al (1994), Rivera-Dommarco et al (1995), se han analizado variables como hacinamiento en los hogares, ingestión de alimentos, aspectos relativos a la vivienda, escolaridad de los padres, etcétera, como variables asociadas o explicativas para algún o algunos indicadores antropométricos de menores preescolares, como posibles indicadores del estado de nutrición de los menores. Para ello se han propuesto algunas metodologías estadísticas como regresiones lineales, regresiones logísticas y correlaciones bivariadas. En estos estudios se ha señalado que el déficit para la talla, a diferencia de la emaciación o el déficit en el peso para la edad, muestra cambios más signifi-

cativos en términos estadísticos en relación con las variables analizadas.

Es importante enfatizar que en estos estudios es posible medir muchas variables socioeconómicas y demográficas, sin embargo, existen algunos factores que no se cuantifican directamente. Algunas de estos pueden ser el cuidado proporcionado al menor, la violencia intrafamiliar, la presencia de alcoholismo, etcétera. Quizá a estos factores de riesgo no medidos pueda atribuirse la presencia en la encuesta de menores con condiciones socioeconómicas y demográficas favorables que presentan déficits importantes en sus medidas antropométricas, así como menores con condiciones aparentemente adversas cuyas medidas antropométricas son óptimas con relación a la población de referencia. Esto como se verá más adelante, repercute en la modelación estadística del score z del peso para la edad.

§3.1 Consideraciones generales del modelo

El interés de este capítulo es básicamente encontrar algunos modelos plausibles para explicar el comportamiento del indicador nutricional, PEDZ, como *variable dependiente o respuesta*. En esta etapa del análisis consideraremos a las variables de los cinco bloques descritos en el capítulo anterior como *variables independientes o regresores*, entendiéndolo por éstas a las variables que se han observado y cuyos cambios puedan provocar o transmitir “cambios” en la variable respuesta.

Dado que nos interesa encontrar como los cambios en las variables independientes afectan al peso para la edad, una relación simple que involucre algunas de las variables independientes sería muy conveniente. El hecho de que cualquier modelo es de alguna manera erróneo, lo que se desea es buscar los mejores, no solo considerando buenos a aquellos que replacen a la variable respuesta observada, PEDZ, que denotaremos y , por los valores ajustados de una variable \hat{y} , que discrepe, bajo algún criterio, muy poco. Sino tomando en cuenta que la simplicidad del modelo, que involucra parsimonia de los parámetros, es una cualidad deseable. No deseamos incluir variables que no sean necesarias para alcanzar los objetivos que se persiguen. El hecho de que las variables dentro de cada rubro pueden significar, en términos de la respuesta, un mismo tipo de características para un cierto bloque quizá permita

que un número menor de variables regresores describan satisfactoriamente la relación de éstas con el peso para la edad.

Para elegir los métodos de análisis es importante considerar los objetivos que se persiguen. Si bien un modelo de regresión no significa que exista una relación de causalidad entre la variable respuesta y los regresores, entendiendo por ésta la presencia de evidencia que nos convenza de que los cambios que para un individuo presente una variable independiente, cuando permanezcan constantes las otras, cambiaran sistemáticamente la respuesta del individuo en cuestión. En el contexto del problema que nos ocupa, un modelo de regresión ayudaría a confirmar las relaciones de causalidad que en la experiencia de los expertos deben presentarse y que en algunos casos han sido evidenciadas en diversos estudios con relación a problemas de desnutrición en menores preescolares.

En este capítulo se propondrán *modelos de regresión lineal* que comenzaremos por describir brevemente. Considerando modelos de regresión lineal de la forma:

$$y_i = \sum_{j=1}^k \beta_j f_j(X_{i1}, \dots, X_{ik}) + \varepsilon_i$$

donde f_j son funciones conocidas, X_{i1}, \dots, X_{ik} corresponden a las variables independientes o regresores en la i -ésima medición que suponemos constantes, $i = 1, \dots, n$ y ε_i es el error aleatorio entre las respuestas observadas y la función ajustada que mide el error en el ajuste. Considerando primero a $f_j(X_{i1}, \dots, X_{ik}) = X_{ij}$, $i = 1, \dots, n$; $j = 1, \dots, k$.

En notación matricial,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

donde

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}; \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n1} & \dots & x_{nk} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}; \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

A los parámetros β_0, \dots, β_k se les conoce generalmente como *coeficientes de regresión*.

A las diferencias entre el valor observado y el valor ajustado $e_i = y_i - \hat{y}_i$ se les conoce como *residuales* o residuos. En notación matricial $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$. Los residuales son una herramienta importante para evaluar que el modelo ajustado sea adecuado.

Recordemos que en estos modelos se asume que el término que corresponde al error tiene $\mathbf{E}(\varepsilon) = \mathbf{0}$ y $\text{Var}(\varepsilon) = \sigma^2 I_n$, además de la no correlación entre los errores lo cual significa que el proceso de medición de \mathbf{X} debe ser el mismo y de manera que una medición no afecte a otra.

En este caso se tienen $k = 18$ variables independientes con $n = 18,774$ mediciones. Si y_i denota la i -ésima observación de la variable respuesta (PEDZ) y x_{ij} el i -ésimo valor para el j -ésimo regresor, $i = 1, \dots, n$, $j = 1, \dots, k$. Los coeficientes de regresión son estimados por el método de mínimos cuadrados, donde la función de mínimos cuadrados es

$$S(\beta_0, \dots, \beta_k) = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2$$

y debe ser minimizada respecto a β_0, \dots, β_k .

Del proceso de minimización¹ se obtienen las siguientes expresiones para los estimadores de los coeficientes de regresión y para los valores ajustados,

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \\ \hat{\mathbf{y}} &= \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \end{aligned}$$

siempre que $(\mathbf{X}'\mathbf{X})^{-1}$ exista, es decir, que ninguna columna de la matriz sea una combinación lineal de las otras o bien que el rango de $r(\mathbf{X}) = k$ sea menor o igual que n . Los estimadores para los coeficientes de regresión son insesgados y su covarianza está dada por $\text{Cov}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Dentro de

¹ DRAPER, N.R., SMITH, H. *Applied Regression Analysis* (1966).

la clase de estimadores insesgados, el teorema de Gauss-Markov establece que los estimadores de mínimos cuadrados tienen varianza mínima bajo el modelo presentado. Más aún, si se supone que los errores siguen una distribución normal, entonces $\hat{\beta}$ es también el estimador máximo verosímil para β . Entonces estaremos suponiendo que $\mathbf{y} \sim N_n(\mathbf{y}; \mathbf{X}\beta, \sigma^2 \mathbf{I}_n)$.

Además de los estimadores de los coeficientes de regresión, un estimador para la varianza σ^2 es necesario para probar hipótesis sobre los coeficientes. En lo que respecta al estimador de la varianza, se puede obtener a partir de la suma de cuadrados de los residuales,

$$\hat{\sigma}^2 = \frac{\mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y}}{n - p} = \frac{\mathbf{e}'\mathbf{e}}{n - p}$$

dado que $\hat{\sigma}^2$ depende de la suma de cuadrados de los residuales, las violaciones de los supuestos del modelo sobre los errores pueden provocar que el estimador no sea útil. En este sentido $\hat{\sigma}^2$ depende del modelo.

Los estimadores para los coeficientes de regresión y la varianza bajo el modelo establecido, con la suposición de normalidad cumplen las propiedades de suficiencia y consistencia.

Otra parte imprescindible del análisis involucra verificar que el modelo sea adecuado y de ser necesario, desarrollar modelos mejorados a partir de ello. Haciendo énfasis en que la falta de ajuste siempre debe reportarse para evitar un uso arbitrario de los modelos propuestos, hay que señalar que la importancia de ésta debe asignarse considerando los objetivos del análisis y la naturaleza de los datos.

En el análisis de regresión lineal, el uso de estadísticas de prueba, así como el estudio gráfico de los residuales entre otros, pueden ayudar a verificar el ajuste y los supuestos de varianza constante y normalidad de los errores.

En lo que respecta a la evaluación del modelo, se define la estadística R^2 , *coeficiente de determinación múltiple* como una medida de la proporción de la variabilidad en \mathbf{y} que ha sido explicada por \mathbf{X} , y está dada por

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{(\hat{\beta}'\mathbf{X}'\mathbf{y} - n\bar{y}^2)}{(\mathbf{y}'\mathbf{y} - n\bar{y}^2)} \in [0, 1]$$

donde valores cercanos a cero de la estadística R^2 significarían que el modelo

no funciona en el sentido de que exista una relación lineal entre y y X y valores cercanos a uno podrían significar que el modelo es bueno.

Dentro de la evaluación del modelo se deben considerar también las hipótesis de significancia global y sobre cada coeficiente de regresión, que mencionaremos más adelante, y el análisis de residuales. El análisis de residuales, que son una medida de la variabilidad no explicada por el modelo y pueden pensarse como realizaciones de los errores aleatorios ϵ , permite verificar las suposiciones que sobre estos se han hecho. Para considerar las diferencias entre las escalas de las variables y se consideraran los *residuales studentizados* r_i , que dividen a los residuales entre su error estándar,

$$r_i = \frac{e_i}{\sqrt{\sum_{i=1}^n e_i^2 \left[1 - \left(\frac{1}{n} - \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \right]}}$$

A partir del análisis de residuales, en caso de que los supuestos no se sostengan por el modelo, es posible por un lado proponer transformaciones con el fin de intentar corregir el modelo y mejorarlo. Además de esto, como se describirá en el siguiente capítulo, se ha propuesto considerar el efecto del diseño utilizado para la obtención de la muestra.

§3.2 Modelos a partir de variables observadas

Para proponer algunos modelos de regresión lineal se consideraron las variables de los cinco bloques etiquetadas como lo indica la tabla 3.1.

Tabla 3.1. Variables consideradas en los modelos de regresión.

Etiqueta	Variables	Etiqueta	Variables
TECHO	Material del Techo	PISO	Material del Piso
AGUA	Lugar de donde obtiene el agua	EXCRETAS	Disposición de Excretas
CALENT	Lugar donde calienta sus alimentos	COCINA	Cocina separada
HACINA	Indice de hacinamiento	TV	Tiene TV
REFRI1	Tiene refrigerador	GASTOPC	Gasto per capita
ACTJF3	Actividad del jefe de Hogar	EDADMA1	Edad de la Madre
ESCOLA	Escolarida de la Madre	IDIOMA	Idioma de la madre
EDAD	Edad del menor	ABLACTAB	Edad de Ablactación
DOCE	Tipo de alimentación	INTEGEN	Intervalo intergenésico

Además, cabe señalar que el comportamiento del peso para la edad es diferente con respecto a las tres regiones que se han considerado. Por esto se han seleccionado algunas variables regresoras por rubro y por región, aquellas que describen “mejor” al peso para la edad. Para ello se utilizaron *métodos de selección de variables* considerando a todas las posibles regresiones desde 2 hasta 18 variables, así como métodos “stepwise”. Algunos modelos de regresión por rubro y por región, reportando el valor de la estadística R^2 en cada caso, se anexan el apéndice B. A partir de las variables seleccionadas por rubro y por región se consideraron todas como regresoras en un segundo análisis para proponer cinco modelos. Como se observará más adelante, los modelos de regresión propuestos cumplen con una función descriptiva, que indica en cierto sentido factores de riesgo relacionados a valores bajos del score z del peso para la edad.

Para los primeros tres modelos que se reportan en la tabla 3.2.a, el análisis se separó para cada región, se utilizaron las variables que para cada rubro se propusieron en el primer análisis y se realizó nuevamente un proceso de selección.

En el caso del modelo para la primera región la hipótesis de significancia global

$$H_0 : \beta_1 = \dots = \beta_k = 0 \quad vs \quad H_1 : \beta_j \neq 0 \quad p.a. \quad j$$

se rechaza, por lo que al menos un regresor contribuye de manera significativa al modelo. En las pruebas parciales para cada coeficiente de regresión

$$H_0 : \beta_j = 0 \quad vs \quad H_1 : \beta_j \neq 0$$

se rechaza para todas las variables, es decir, son significativamente diferentes de cero, exceptuando quizá al hacinamiento que solo puede rechazarse con un 86.51% de confianza.

Para los modelos de la segunda y tercera región, reportados en la tabla 3.2.a, tanto la hipótesis de significancia global como las hipótesis parciales sobre los coeficientes se rechazan.

Los modelos que se presentan en la tabla 3.3.a, son modelos para los que no se consideraron por separado las observaciones de acuerdo a la región. Se consideraron las variables en los modelos por región y se seleccionaron

dos modelos con 10 regresores, ya sea incluyendo la variable región como regresor, que se presenta en segundo término, o no tomándola en cuenta para el análisis. En ambos modelos las hipótesis globales y parciales para los coeficientes se rechazan.

Tabla 3.2.a Modelos de Regresión con variables originales por Región.
Región: 1.Noreste, Norte-Occidente; 2.Centro; 3.Centro-Sur.

REGION=1					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	10	676.13324	67.61332	43.313	0.0001
Error	4761	7432.13847	1.56105		
C Total	4771	8108.27171			
Root MSE	1.24942	R-square	0.0834		
Dep Mean	-0.26920	Adj R-sq	0.0815		
C.V.	-464.13073				
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-1.286456	0.16717036	-7.695	0.0001
EXCRETAS	1	0.066923	0.02438554	2.744	0.0061
CALENT	1	0.106471	0.03574121	2.979	0.0029
HACINA	1	-0.017791	0.01189656	-1.495	0.1349
REFRI_1	1	0.176586	0.04000434	4.414	0.0001
GASTOPC	1	0.004409	0.00086093	5.121	0.0001
ESCOLA	1	0.071184	0.02193063	3.246	0.0012
IDIOMA	1	0.243268	0.10809825	2.250	0.0245
EDAD	1	-0.013357	0.00120940	-11.045	0.0001
ABLACTAB	1	0.004042	0.00083444	4.844	0.0001
INTERGEN	1	0.002725	0.00050767	5.367	0.0001
Durbin-Watson D			1.910		
(For Number of Obs.)			4772		
1st Order Autocorrelation			0.045		

REGION=2
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	10	791.34385	79.13439	57.671	0.0001
Error	5543	7605.90010	1.37216		
C Tota	5553	8397.24395			

Root MSE	1.17139	R-square	0.0942
Dep Mean	-0.74728	Adj R-sq	0.0926
C.V.	-156.75406		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-1.440241	0.10723942	-13.430	0.0001
CALENT	1	0.099708	0.02621503	3.803	0.0001
HACINA	1	-0.035923	0.00897613	-4.002	0.0001
REFRI_1	1	0.165029	0.03978897	4.148	0.0001
GASTOPC	1	0.006144	0.00109361	5.618	0.0001
ACTJF3	1	0.019773	0.00793965	2.490	0.0128
ESCOLA	1	0.053477	0.01637767	3.265	0.0011
IDIOMA	1	0.227929	0.05009114	4.550	0.0001
EDAD	1	-0.011140	0.00108414	-10.275	0.0001
ABLACTAB	1	0.005334	0.00073272	7.280	0.0001
INTERGEN	1	0.002857	0.00042845	6.668	0.0001

Durbin-Watson D	1.912
(For Number of Obs.)	5554
1st Order Autocorrelation	0.044

REGION=3					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	10	1355.60018	135.56002	86.694	0.0001
Error	7014	10967.52373	1.56366		
C Total	7024	12323.12391			
Root MSE		1.25046	R-square	0.1100	
Dep Mean		-1.06011	Adj R-sq	0.1087	
C.V.		-117.95569			

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-1.561892	0.07157942	-21.820	0.0001
PISO	1	0.066586	0.01674865	3.976	0.0001
EXCRETAS	1	0.090803	0.02170466	4.184	0.0001
CALENT	1	0.099101	0.02586311	3.832	0.0001
TV	1	0.131028	0.03295266	3.976	0.0001
GASTOPC	1	0.003289	0.00095858	3.432	0.0006
ESCOLA	1	0.058631	0.01431728	4.095	0.0001
IDIOMA	1	0.229500	0.03421289	6.708	0.0001
EDAD	1	-0.007513	0.00109556	-6.857	0.0001
ABLACTAB	1	0.003912	0.00064137	6.100	0.0001
DOCE	1	-0.185606	0.02078274	-8.931	0.0001

Durbin-Watson D 1.824
 (For Number of Obs.) 7025
 1st Order Autocorrelation 0.088

A partir de éstos modelos es importante enfatizar que la proporción de varianza explicada por el modelo no excede el 11%. Por otra parte, como puede observarse en las gráficas 3.1 del modelo global por región, en este caso para la región 3 y en la gráfica 3.2 que corresponden al modelo globales sin la

variable región incluida, no pueden considerarse como modelos predictivos, los valores de la variable observada aparecen dispersos respecto a los valores ajustados. Los residuales de acuerdo a estas mismas gráficas no presentan una tendencia muy clara para describir la heterogeneidad de las varianzas que, refleja cierto comportamiento creciente y decreciente.

Además, de la gráfica 3.3 para la probabilidad acumulada observada contra la esperada, se tiene que la distribución de los residuales presenta las colas más pesadas que la distribución normal. Los supuestos no son del todo satisfechos lo que puede deberse a que se requiera una transformación en las variables o bien a la falta de independencia. También es posible que se tenga correlación entre las observaciones a nivel localidad. El comportamiento de los residuales de los demás modelos propuestos es en general muy similar. Por ello no se presentan los análisis para todos los modelos.

De acuerdo a las gráficas para los residuales contra las variables independientes continuas, en el caso de la edad del menor se observa en la gráfica 3.4 una tendencia que podría significar una necesidad de transformar la variable. Se propuso entonces utilizar el logaritmo de la edad del menor (LEDAD).

Hecho esto se obtuvieron los modelos que se presentan en las tablas 3.2.b y 3.3.b. Los modelos son básicamente los mismos, excepto que para la región 1, en el modelo la variable edad de ablactación ya no resulta significativa, por ello se retiró del modelo. En estos casos, al graficar los residuales contra las variables independientes ya no se observan tendencias, en lo que se refiere al ajuste del modelo y a los supuestos, no se logran cambios favorables importantes.

Los modelos en las tablas 3.3.b y 3.4.b no se han complicado mucho en lo que respecta a la interpretación de los coeficientes de regresión y se ha utilizado la información que arrojó la verificación de los modelos en las tablas 3.3.a y 3.4.a, mejorando así un poco los modelos a proponer en términos del tipo de relación entre los regresores y la variable respuesta.

Tabla 3.3.a Modelos de Regresión con variables originales para el total Nacional.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	10	4097.16693	409.71669	267.154	0.0001
Error	17816	27323.23368	1.53363		
C Total	17826	31420.40061			
Root MSE	1.23840	R-square	0.1304		
Dep Mean	-0.74287	Adj R-sq	0.1299		
C.V.	-166.70536				
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-1.797115	0.06062036	-29.645	0.0001
EXCRETAS	1	0.081410	0.01270284	6.409	0.0001
CALENT	1	0.156168	0.01605515	9.727	0.0001
HACINA	1	-0.021340	0.00523135	-4.079	0.0001
REFRI_1	1	0.191165	0.02336348	8.182	0.0001
GASTOPC	1	0.005754	0.00054135	10.628	0.0001
ESCOLA	1	0.067085	0.00963601	6.962	0.0001
IDIOMA	1	0.332010	0.02571813	12.910	0.0001
EDAD	1	-0.012634	0.00063558	-19.878	0.0001
ABLACTAB	1	0.004785	0.00041166	11.624	0.0001
INTERGEN	1	0.002724	0.00025237	10.794	0.0001
Durbin-Watson D		1.850			
(For Number of Obs.)		17827			
1st Order Autocorrelation		0.075			

Analysis of Variance

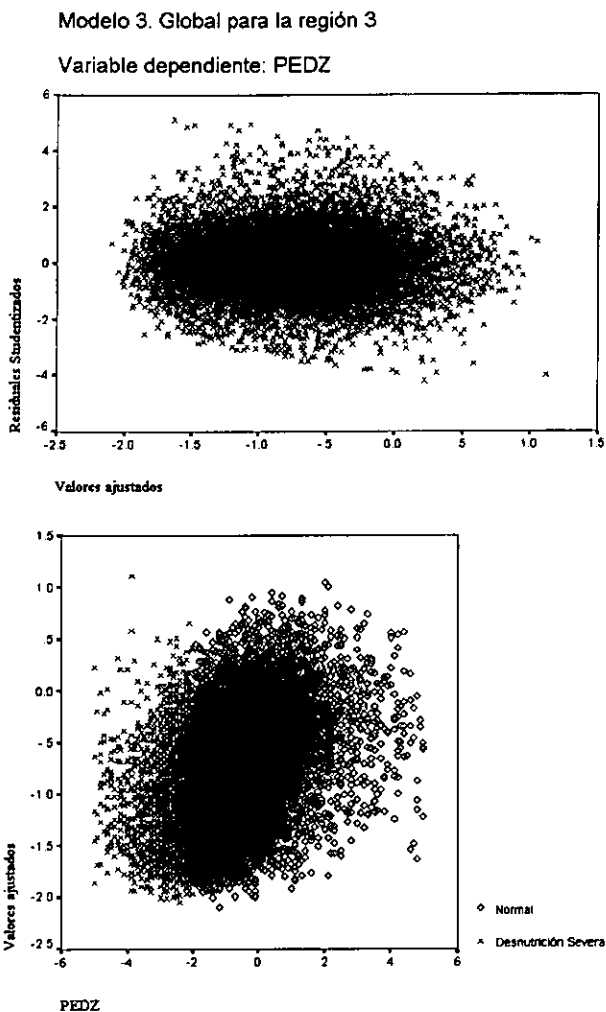
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	11	4492.70177	408.42743	270.210	0.0001
Error	17815	26927.69884	1.51152		
C Total	17826	31420.40061			

Root MSE	1.22944	R-square	0.1430
Dep Mean	-0.74287	Adj R-sq	0.1425
C.V.	-165.49898		

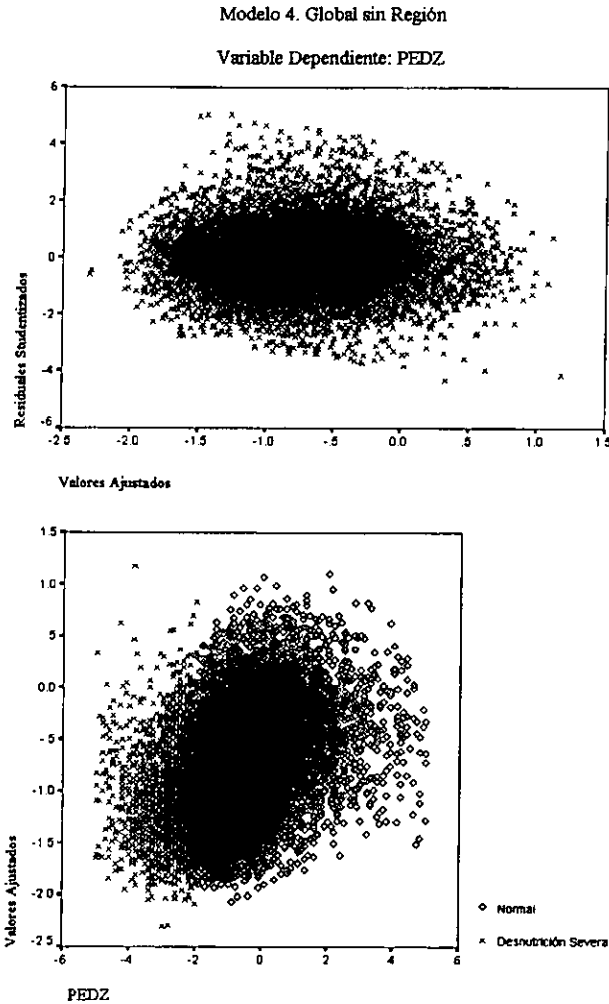
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-1.093878	0.07424085	-14.734	0.0001
EXCRETAS	1	0.069306	0.01263310	5.486	0.0001
CALENT	1	0.108082	0.01621379	6.666	0.0001
HACINA	1	-0.019046	0.00519543	-3.666	0.0002
REFRI_1	1	0.155342	0.02329989	6.667	0.0001
GASTOPC	1	0.004723	0.00054119	8.728	0.0001
ESCOLA	1	0.063056	0.00956952	6.589	0.0001
IDIOMA	1	0.235668	0.02621743	8.989	0.0001
EDAD	1	-0.012784	0.00063105	-20.258	0.0001
ABLACTAB	1	0.004854	0.00040871	11.877	0.0001
INTERGEN	1	0.002778	0.00025056	11.086	0.0001
REGION	1	-0.212842	0.01315744	-16.177	0.0001

Durbin-Watson D	1.867
(For Number of Obs.)	17827
1st Order Autocorrelation	0.066

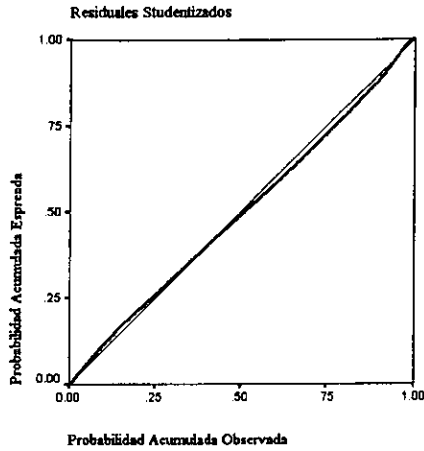
Gráfica 3.1: Valores ajustados contra los residuales studentizados y contra la variable respuesta para el modelo global para la región 3.



Gráfica 3.2: Valores ajustados contra los residuales studentizados y contra la variable respuesta para el modelo global sin región.



Gráfica 3.3: P-P Normal para los residuales studentizados del modelo para la región 3.



Gráfica 3.4: Edad del menor vs residuales modelo para la región 3.

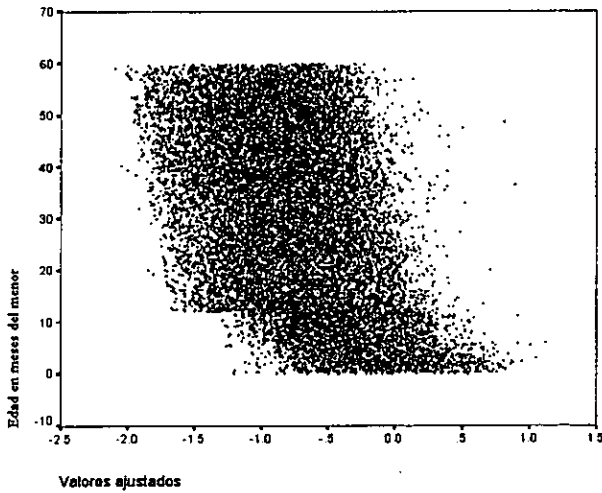


Tabla 3.2.b Modelos de Regresión con variables originales por Región.
Región: 1.Noreste, Norte-Occidente; 2.Centro; 3.Centro-Sur.

REGION=1					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	9	691.68235	76.85359	49.344	0.0001
Error	4760	7413.67811	1.55750		
C Total	4769	8105.36046			
Root MSE	1.24800	R-square	0.0853		
Dep Mean	-0.26966	Adj R-sq	0.0836		
C.V.	-462.79589				
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-0.665409	0.17102351	-3.891	0.0001
EXCRETAS	1	0.064551	0.02436339	2.650	0.0081
CALENT	1	0.101547	0.03571043	2.844	0.0045
HACINA	1	-0.020139	0.01186971	-1.697	0.0898
REFRI_1	1	0.170736	0.03996754	4.272	0.0001
GASTOPC	1	0.004475	0.00085969	5.205	0.0001
ESCOLA	1	0.069219	0.02190226	3.160	0.0016
IDIOMA	1	0.238812	0.10797353	2.212	0.0270
LEDAD	1	-0.296637	0.01955313	-15.171	0.0001
INTERGEN	1	0.002563	0.00050106	5.115	0.0001
Durbin-Watson D		1.907			
(For Number of Obs.)		4770			
1st Order Autocorrelation		0.046			

REGION=2
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	10	866.57047	86.65705	63.930	0.0001
Error	5541	7510.81328	1.35550		
C Total	5551	8377.38374			

Root MSE	1.16426	R-square	0.1034
Dep Mean	-0.74829	Adj R-sq	0.1018
C.V.	-155.58943		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-0.823903	0.12313746	-6.691	0.0001
CALENT	1	0.097221	0.02605664	3.731	0.0002
HACINA	1	-0.035530	0.00891397	-3.986	0.0001
REFRI_1	1	0.164922	0.03954262	4.171	0.0001
GASTOPC	1	0.006136	0.00108662	5.647	0.0001
ACTJF3	1	0.020209	0.00789733	2.559	0.0105
ESCOLA	1	0.052862	0.01627403	3.248	0.0012
IDIOMA	1	0.207363	0.04983902	4.161	0.0001
LEDAD	1	-0.283678	0.02195697	-12.920	0.0001
ABLACTAB	1	0.001882	0.00082490	2.282	0.0225
INTERGEN	1	0.002852	0.00041726	6.835	0.0001

Durbin-Watson D	1.915
(For Number of Obs.)	5552
1st Order Autocorrelation	0.042

REGION=3
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	10	1425.44949	142.54495	91.989	0.0001
Error	7007	10857.97965	1.54959		
C Total	7017	12283.42914			

Root MSE	1.24483	R-square	0.1160
Dep Mean	-1.06191	Adj R-sq	0.1148
C.V.	-117.22489		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-1.087217	0.09329336	-11.654	0.0001
PISO	1	0.066408	0.01668004	3.981	0.0001
EXCRETAS	1	0.089709	0.02161333	4.151	0.0001
CALENT	1	0.098466	0.02575668	3.823	0.0001
TV	1	0.128965	0.03281308	3.930	0.0001
GASTOPC	1	0.003357	0.00095422	3.518	0.0004
ESCOLA	1	0.058969	0.01425483	4.137	0.0001
IDIOMA	1	0.226162	0.03408259	6.636	0.0001
LEDAD	1	-0.236396	0.02373121	-9.961	0.0001
ABLACTAB	1	0.001585	0.00070975	2.233	0.0256
DOCE	1	-0.131982	0.02186085	-6.037	0.0001

Durbin-Watson D	1.821
(For Number of Obs.)	7018
1st Order Autocorrelation	0.089

Tabla 3.3b. Modelos de Regresión con variables originales para el total Nacional.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	10	4355.72121	435.57212	287.161	0.0001
Error	17805	27007.00531	1.51682		
C Total	17815	31362.72652			
Root MSE	1.23159	R-square	0.1389		
Dep Mean	-0.74394	Adj R-sq	0.1384		
C.V.	-165.54924				
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-1.141018	0.07038763	-16.210	0.0001
EXCRETAS	1	0.079789	0.01263584	6.314	0.0001
CALENT	1	0.152453	0.01597118	9.545	0.0001
HACINA	1	-0.021509	0.00519897	-4.137	0.0001
REFRI_1	1	0.187471	0.02324365	8.065	0.0001
GASTOPC	1	0.005797	0.00053832	10.769	0.0001
ESCOLA	1	0.065503	0.00958474	6.834	0.0001
IDIOMA	1	0.318191	0.02560040	12.429	0.0001
LEDAD	1	-0.306550	0.01277075	-24.004	0.0001
ABLACTAB	1	0.001308	0.00046714	2.801	0.0051
INTERGEN	1	0.002648	0.00024671	10.734	0.0001
Durbin-Watson D		1.851			
(For Number of Obs.)		17816			
1st Order Autocorrelation		0.075			

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	11	4739.72031	430.88366	288.151	0.0001
Error	17804	26623.00621	1.49534		
C Total	17815	31362.72652			

Root MSE	1.22284	R-square	0.1511
Dep Mean	-0.74394	Adj R-sq	0.1506
C.V.	-164.37271		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-0.450090	0.08211715	-5.481	0.0001
EXCRETAS	1	0.067909	0.01256793	5.403	0.0001
CALENT	1	0.105006	0.01613172	6.509	0.0001
HACINA	1	-0.019346	0.00516379	-3.746	0.0002
REFRI_1	1	0.152038	0.02318414	6.558	0.0001
GASTOPC	1	0.004787	0.00053821	8.894	0.0001
ESCOLA	1	0.061641	0.00951967	6.475	0.0001
IDIOMA	1	0.223340	0.02609850	8.558	0.0001
LEDAD	1	-0.306889	0.01268001	-24.203	0.0001
ABLACTAB	1	0.001402	0.00046386	3.022	0.0025
INTERGEN	1	0.002685	0.00024496	10.960	0.0001
REGION	1	-0.209758	0.01308953	-16.025	0.0001

Durbin-Watson D	1.868
(For Number of Obs.)	17816
1st Order Autocorrelation	0.066

Estos modelos se han propuesto a partir de las variables seleccionadas, en la siguiente sección se presentaran los modelos a partir de índices resumen, buscando reducir el modelo en términos del número de variables, para buscar describir más fácilmente las relaciones de éstas con la variable respuesta.

§3.3 Modelos a partir de índices resumen y variables observadas

En una segunda etapa, se construyeron índices para cada uno de los cinco bloques, proponiendo como índices a la primera componente principal obtenida a partir de la matriz de correlaciones de las variables de cada bloque. La idea subyacente en un análisis de *componentes principales* consiste en representar las características principales de la estructura de datos multivariados en términos de un número menor de variables. Se busca proyectar a los datos en un espacio de menor, representando la mayor variabilidad posible, esto es, que las características no descritas se puedan atribuir a “ruido” no relevante. Recordando que en un análisis de componentes principales se tiene un vector \mathbf{X} de dimensión k con media μ y matriz de dispersión Σ , en este caso el vector está formado por las variables que se han considerado en cada bloque. En todos los bloques se tienen variables discretas, sin embargo, todas las variables se pueden considerar ordinales puesto que conforme las categorías crecen las condiciones son “mejores”.

El análisis de componentes principales utiliza entonces la descomposición espectral, esto es, se obtiene una matriz ortogonal $\mathbf{T} = (t_1, \dots, t_k)$ tal que

$$\mathbf{T}'\Sigma\mathbf{T} = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_k),$$

donde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0$ son los eigenvalores ordenados de Σ que es una matriz simétrica de $k \times k$. Sea $\mathbf{z} = [(z_j)] = \mathbf{T}'(\mathbf{X} - \mu)$, entonces $z_j = t_j'(\mathbf{X} - \mu)$ para $j = 1, \dots, k$ se conoce como el j -ésimo componente principal de \mathbf{X} . Ahora $w_j = \lambda_j^{-\frac{1}{2}} z_j$ se conocen como los componentes principales estandarizados de \mathbf{X} . Como el eigenvector t_j tiene longitud unitaria se tiene que z_j es la proyección ortogonal de $(\mathbf{X} - \mu)$ en dirección de t_j .

Además las componentes principales son no correlacionadas y la $\text{Var}(z_j) = \lambda_j$, esto puede verse fácilmente pues,

$$\mathbf{D}[\mathbf{z}] = \mathbf{D}[\mathbf{T}'(\mathbf{X} - \mu)] = \mathbf{T}'\mathbf{D}[\mathbf{X}]\mathbf{T} = \mathbf{T}'\Sigma\mathbf{T} = \Lambda$$

y en el caso de considerarse las componentes estandarizadas

$$\mathbf{D}[\mathbf{w}] = \mathbf{D}[\Lambda^{-\frac{1}{2}}\mathbf{z}] = \Lambda^{-\frac{1}{2}}\mathbf{T}'\mathbf{D}[\mathbf{X}]\mathbf{T}\Lambda^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}}\mathbf{T}'\Sigma\mathbf{T}\Lambda^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}}\Lambda\Lambda^{-\frac{1}{2}} = \mathbf{I}_k.$$

Ahora bien, en la práctica, μ y Σ no se conocen y deben estimarse a partir de una muestra $(\mathbf{X}_1, \dots, \mathbf{X}_n)$. Sea $\hat{\mu} = \bar{\mathbf{X}}$ y sea $\hat{\Sigma} = \sum_i (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})' / n$, con eigenvalores $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_k$ y una matriz ortogonal $\hat{\mathbf{T}} = (\hat{t}_1, \dots, \hat{t}_k)$ de sus eigenvectores correspondientes. Entonces para cada observación \mathbf{X}_i se puede definir un componente principal muestral o estimado como $z_i = (\mathbf{X}_i - \bar{\mathbf{X}})$ obteniendo una nueva matriz de datos

$$\mathbf{Z}' = (z_1, \dots, z_n) = \hat{\mathbf{T}}'(\mathbf{X}_1 - \bar{\mathbf{X}}, \dots, \mathbf{X}_n - \bar{\mathbf{X}}).$$

Seber² plantea que un enfoque poblacional o muestral de componentes principales no tienen diferencia esenciales.

Una desventaja de los componentes principales es que no es invariante ante cambios de escala y los resultados del análisis dependen de las unidades de medición. Entonces se propone utilizar la matriz muestral de correlación, que no depende de las escalas utilizadas para medir las variables originales. $\mathbf{R} = [r_{jk}]$ donde $r_{jj} = 1$ y

$$r_{jk} = \frac{\sum_i (\mathbf{X}_{ij} - \bar{\mathbf{X}}_{.j})(\mathbf{X}_{ik} - \bar{\mathbf{X}}_{.k})}{[\sum_i (\mathbf{X}_{ij} - \bar{\mathbf{X}}_{.j})^2 \sum_i (\mathbf{X}_{ik} - \bar{\mathbf{X}}_{.k})^2]^{1/2}} \quad (j \neq k).$$

Se definieron entonces, a partir de la matriz de correlación muestral y utilizando los componentes principale estandarizados, 5 índices: para el bloque A el índice de vivienda, para el bloque D un índice para la madre, para el E un índice del menor y para los bloques B y C un índice resumen de bienes y un índice de condición socioeconómica respectivamente.

Para el *índice de vivienda*, el primer componente principal explica el 37.95% de la varianza y el eigenvector correspondiente está dado por un promedio ponderado de las característica de la vivienda que se han cosiderado en los modelos de la sección anterior:

- 0.423342 Material del techo
- 0.475085 Material del piso
- 0.232243 Lugar de donde obtiene agua
- 0.406498 Disposición de excretas en la vivienda
- 0.371836 (-) Hacinamiento
- 0.487480 Lugar donde calienta sus alimentos.

²SEBER, G.A.F. *Multivariate Observations (1984)*.

El *índice de bienes* que se ha propuesto explica el 67.11% de la varianza, en este caso el promedio ponderado está dado por:

- 0.707107 Cuenta con televisor
- 0.707107 Cuenta con refrigerador.

Para el bloque C, el *índice de condición socioeconómica* que explica el 55.89% de la varianza está compuesto por un promedio como sigue:

- 0.707107 Gasto per capita en alimentos
- 0.707107 Actividad del jefe del hogar.

El *índice para la madre* explica el 47.08% de la varianza, pero en este caso su interpretación no es muy transparente, como se observó en algunos modelos durante la selección de variables y modelos, valores altos de este índice pueden dar origen a valores altos de la variable respuesta. Entonces por como está constituido el índice, sus valores altos equivalen a madres con mayor escolaridad y bilingües o que hablan español, y en general jóvenes. El eigenvector tiene la forma

- 0.687264 Escolaridad de la madre
- 0.518545 Idioma de la madre
- 0.508704 Edad de la madre.

Finalmente el *índice del menor* que explica el 47.92% de la varianza, para éste tampoco es clara su interpretación, es necesario describirlo con más detalle. En algunos modelos basados en estos índices se observó que, valores pequeños de este índice es posible relacionarlos con valores altos del indicador del estado nutricional, en este sentido, se tendría que los menores con edades de ablactación no muy tempranas, intervalos intergenésicos relativamente grandes y categorías bajas de tipo de alimentación asociadas a alimentación al seno materno o mixta, significarían valores favorables para la variable respuesta.

- 0.414896 Intervalo intergenésico
- 0.640966 Edad de ablactación
- 0.645774 Tipo de alimentación.

Estos índices propuestos se consideraron como regresores para modelos de regresión, donde la variable respuesta es nuevamente el *score* z del peso para la edad. Sin embargo, estos modelos disminuían el poder predictivo del modelo (R^2 menores a los modelos con las variables originales) y tampoco se ganaba en interpretabilidad de los coeficientes respecto al las variables originales. En un segundo análisis se considero el índice de hogar, que fue el índice más estable en diversos modelos considerando apareciendo con coeficientes similares y significativos, que se denoto como *CPB11* y las variables seleccionadas en la sección anterior cuya notación se explica en la tabla 3.1, y siguiendo los mismos procesos de selección. Se proponen tres modelos considerando el índice de hogar y algunas variables de los bloques restantes por región que se presentan en la tabla 3.4 y dos modelos sin separar por región, incluyendo la región como variable independiente y sin incluirla, que se encuentran en la tabla 3.5. Los modelos que solo consideran los índices no resultan muy útiles para detectar las variables que favorecen la presencia de un déficit en el peso para la edad y la medida en que lo hacen.

El comportamiento de los residuales de estos modelos se ejemplifica con la gráfica 3.4 y se observa que del análisis de residuales, concluir que se sostienen los supuestos que se han considerado en el modelo no es muy claro. El modelo continua siendo un modelo que explica en el mejor de los casos un 13.3% de la variabilidad. Aunque en estos modelos se tiene un número menor de variables y esto facilita la interpretación de los coeficientes de regresión.

Tabla 3.4. Modelos de Regresión con índice de hogar y variables por Región. Región: 1.Noreste, Norte-Occidente; 2.Centro; 3.Centro-Sur.

REGION=1					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	6	544.32848	90.72141	57.663	0.0001
Error	4786	7529.81685	1.57330		
C Total	4792	8074.14532			
Root MSE		1.25431	R-square	0.0674	
Dep Mean		-0.26580	Adj R-sq	0.0662	
C.V.		-471.89335			
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-0.772240	0.08336768	-9.263	0.0001
CPB11	1	0.082297	0.01670186	4.927	0.0001
REFRI_1	1	0.185428	0.04029304	4.602	0.0001
GASTOPC	1	0.004563	0.00084365	5.409	0.0001
ESCOLA	1	0.090718	0.02179528	4.162	0.0001
ABLACTAB	1	0.005302	0.00081097	6.538	0.0001
DOCE	1	-0.145505	0.02024354	-7.188	0.0001
Durbin-Watson D			1.912		
(For Number of Obs.)			4793		
1st Order Autocorrelation			0.044		

REGION=2
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	6	756.67192	126.11199	90.283	0.0001
Error	5923	8273.54944	1.39685		
C Total	5929	9030.22136			

Root MSE	1.18188	R-square	0.0838
Dep Mean	-0.74444	Adj R-sq	0.0829
C.V.	-158.76261		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-1.121448	0.07013040	-15.991	0.0001
CPB11	1	0.116778	0.01186840	9.839	0.0001
GASTOPC	1	0.007378	0.00105113	7.019	0.0001
IDIOMA	1	0.188227	0.04865803	3.868	0.0001
ABLACTAB	1	0.005271	0.00070841	7.440	0.0001
DOCE	1	-0.170490	0.01792882	-9.509	0.0001
INTERGEN	1	0.001961	0.00039429	4.973	0.0001

Durbin-Watson D	1.886
(For Number of Obs.)	5930
1st Order Autocorrelation	0.057

REGION=3
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	6	1277.03742	212.83957	133.834	0.0001
Error	7055	11219.75708	1.59033		
C Total	7061	12496.79449			

Root MSE	1.26108	R-square	0.1022
Dep Mean	-1.07264	Adj R-sq	0.1014
C.V.	-117.56777		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-1.126900	0.05806783	-19.407	0.0001
CPB11	1	0.111679	0.01198212	9.320	0.0001
TV	1	0.149031	0.03262957	4.567	0.0001
IDIOMA	1	0.294470	0.03277380	8.985	0.0001
ABLACTAB	1	0.005212	0.00062385	8.354	0.0001
DOCE	1	-0.265387	0.01887168	-14.063	0.0001
INTERGEN	1	0.001701	0.00038090	4.467	0.0001

Durbin-Watson D	1.803
(For Number of Obs.)	7062
1st Order Autocorrelation	0.099

Tabla 3.5. Modelos de Regresión para el total Nacional basados en índice de hogar y variables.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	6	3793.65863	632.27644	406.788	0.0001
Error	17777	27631.01635	1.55431		
C Total	17783	31424.67498			
Root MSE	1.24672	R-square	0.1207		
Dep Mean	-0.74426	Adj R-sq	0.1204		
C.V.	-167.51044				
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-1.133193	0.05080401	-22.305	0.0001
CPB11	1	0.150767	0.00716174	21.052	0.0001
GASTOPC	1	0.007255	0.00053814	13.481	0.0001
EDADMA_1	1	0.003255	0.00132321	2.460	0.0139
IDIOMA	1	0.369359	0.02494736	14.806	0.0001
ABLACTAB	1	0.005128	0.00040721	12.594	0.0001
DOCE	1	-0.191767	0.01105982	-17.339	0.0001
Durbin-Watson D		1.845			
(For Number of Obs.)		17784			
1st Order Autocorrelation		0.077			

Analysis of Variance

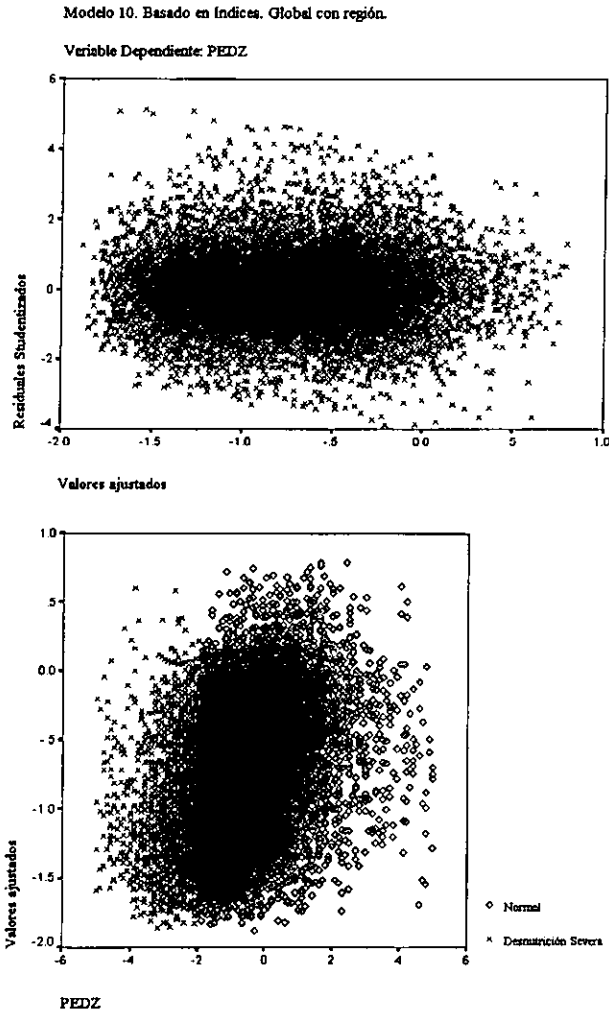
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	7	4199.31991	599.90284	391.689	0.0001
Error	17776	27225.35507	1.53158		
C Total	17783	31424.67498			

Root MSE	1.23757	R-square	0.1336
Dep Mean	-0.74426	Adj R-sq	0.1333
C.V.	-166.28093		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-0.512808	0.06321712	-8.112	0.0001
CPB11	1	0.122019	0.00732534	16.657	0.0001
GASTOPC	1	0.005997	0.00053975	11.111	0.0001
EDADMA_1	1	0.001717	0.00131689	1.304	0.1923
IDIOMA	1	0.258663	0.02568135	10.072	0.0001
ABLACTAB	1	0.005232	0.00040427	12.942	0.0001
DOCE	1	-0.189580	0.01097946	-17.267	0.0001
REGION	1	-0.214855	0.01320182	-16.275	0.0001

Durbin-Watson D	1.864
(For Number of Obs.)	17784
1st Order Autocorrelation	0.068

Gráfica 2.5: Valores ajustados contra los residuales studentizados y contra la variable respuesta para el modelo de índice de hogar y variables global con región.



§3.4 Modelos de regresión para un estado en cada región

Para cada una de las tres regiones en las que se ha dividido el país a lo largo de este trabajo, se seleccionó un estado tipo. Esto con el fin de observar si su comportamiento presenta notables diferencias respecto a la región en la que se encuentran. La tabla 3.6 resume los porcentajes de menores entrevistados con desnutrición por estado y por región.

Tomando en cuenta las mayores diferencias en cuanto a los porcentajes de desnutrición entre la región y los estados que la conforman se tomaron los tres que además, como se detallará más adelante, tuvieron también diferencias en porcentajes de la población rural que representan. Se consideraron para la región Noreste y Norte-Occidente a Michoacán, para la región Centro al Estado de México y para la región Centro-Sur a Chiapas.

A partir de los métodos de selección de variables que se han utilizado, se obtuvieron modelos para los tres estados tipo. En general no presentan grandes discrepancias con respecto a la región a la que pertenecen. En la tabla 3.7 se resumen los modelos para los estados. En el caso de Chiapas, que es uno de los estados con mayor desnutrición en áreas rurales en el país, un 19.7% de los menores entrevistados presentan desnutrición de acuerdo con la partición dicotómica de la variable respuesta *PEDZ2*, la ordenada al origen no resulta significativa. Al quitarla del análisis el ajuste mejora únicamente en cuanto a la estadística R^2 . Recordemos que ésta estadística representa la proporción de variabilidad respecto a \bar{y} explicada por la regresión, entonces en el caso del modelo sin ordenada

$$R_0^2 = \frac{\sum_{i=1}^n \hat{y}_i^2}{\sum_{i=1}^n y_i^2}$$

esto es la proporción de variabilidad alrededor del origen explicada por regresión. Por ello es frecuente que la estadística R_0^2 sea mayor que R^2 aún cuando la suma de cuadrados medios de los residuales, que puede considerarse una medida de calidad de ajuste del modelo, de la regresión con ordenada al origen sea menor que la de aquella sin ésta. Se podrían considerar definiciones alternativas de R_0^2 , sin embargo, en el contexto del problema que nos ocupa, una regresión a través del origen no resulta útil. Lo que sucede con

Tabla 3.6. Porcentajes de desnutrición por estado y región.

Estado	Desnutrición Severa	Normal	Total	Porcentaje de desnutrición de desnutrición
Aguascalientes	27	243	270	10.0%
Baja California	12	105	117	10.3%
Baja California Sur	6	103	109	5.5%
Campeche	135	597	732	18.4%
Coahuila	25	276	301	8.3%
Colima	34	183	217	15.7%
Chiapas	286	1163	1449	19.7%
Chihuahua	17	152	169	10.1%
Durango	26	365	391	6.6%
Guanajuato	102	460	562	18.1%
Guerrero	373	836	1209	30.9%
Hidalgo	126	590	716	17.6%
Jalisco	45	608	653	6.9%
México	111	886	997	11.1%
Michoacán	122	910	1032	11.8%
Morelos	55	433	488	11.3%
Nayarit	14	152	166	8.4%
Nuevo León	16	149	165	9.7%
Oaxaca	268	891	1159	23.1%
Puebla	431	1219	1650	26.1%
Querétaro	65	360	425	15.3%
Quintana Roo	38	123	161	23.6%
San Luis Potosí	111	669	780	14.2%
Sinaloa	30	461	491	6.1%
Sonora	7	176	183	3.8%
Tabasco	118	789	907	13.0%
Tamaulipas	11	226	237	4.6%
Tlaxcala	39	181	220	17.7%
Veracruz	297	1527	1824	16.3%
Yucatán	137	347	484	28.3%
Zacatecas	46	464	510	9.0%
Región 1	438	4573	5011	8.7%
Región 2	898	5305	6203	14.5%
Región 3	1794	5766	7560	23.7%
Nacional	3130	15644	18774	16.7%

el estado de Chiapas, en cuanto a las hipótesis parciales para los coeficientes de regresión, no se repite en ningún otro de la región Centro-Sur. En cuanto a los supuestos y el comportamiento de los residuales, no existe mucha diferencia de aquellos de los estados con respecto a aquellos de las regiones, como ejemplifica la gráfica 2.6.

Tabla 3.7. Modelos de Regresión lineal para un estado en cada región.

MICHUACAN (REGION 1)					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	8	250.77219	31.34652	18.451	0.0001
Error	951	1615.63181	1.69888		
C Total	959	1866.40400			
Root MSE		1.30341	R-square	0.1344	
Dep Mean		-0.39000	Adj R-sq	0.1271	
C.V.		-334.20761			
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-0.750357	0.31549167	-2.378	0.0176
PISO	1	0.136843	0.04658919	2.937	0.0034
COCINA_1	1	-0.263763	0.12247733	-2.154	0.0315
TV	1	0.216454	0.12089239	1.790	0.0737
GASTOPC	1	0.007144	0.00224159	3.187	0.0015
IDIOMA	1	0.488558	0.16651281	2.934	0.0034
LEDAD	1	-0.328831	0.05809383	-5.660	0.0001
ABLACTAB	1	0.006924	0.00206077	3.360	0.0008
INTERGEN	1	0.003473	0.00112200	3.095	0.0020
Durbin-Watson D			1.876		
(For Number of Obs.)			960		
1st Order Autocorrelation			0.062		

ESTADO DE MEXICO (REGION 2)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	6	110.44395	18.40733	16.187	0.0001
Error	962	1093.92343	1.13713		
C Total	968	1204.36739			

Root MSE	1.06637	R-square	0.0917
Dep Mean	-0.66419	Adj R-sq	0.0860
C.V.	-160.55124		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-1.411573	0.20437793	-6.907	0.0001
CALENT	1	0.101707	0.04728340	2.151	0.0317
GASTOPC	1	0.008134	0.00245774	3.309	0.0010
ESCOLA	1	0.134359	0.03790085	3.545	0.0004
LEDAD	1	-0.133739	0.04825149	-2.772	0.0057
ABLACTAB	1	0.004872	0.00208790	2.333	0.0198
INTERGEN	1	0.002360	0.00090636	2.604	0.0094

Durbin-Watson D	1.930
(For Number of Obs.)	969
1st Order Autocorrelation	0.034

CHIAPAS (REGION 3)
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	7	270.17775	38.59682	26.174	0.0001
Error	1413	2083.62302	1.47461		
C Total	1420	2353.80077			

Root MSE	1.21433	R-square	0.1148
Dep Mean	-0.96383	Adj R-sq	0.1104
C.V.	-125.99080		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-0.327025	0.19988258	-1.636	0.1020
TECHO	1	-0.198582	0.06120885	-3.244	0.0012
CALENT	1	0.196102	0.06136063	3.196	0.0014
GASTOPC	1	0.004830	0.00211267	2.286	0.0224
IDIOMA	1	0.236987	0.07272583	3.259	0.0011
LEDAD	1	-0.338009	0.04660619	-7.252	0.0001
DOCE	1	-0.101527	0.04669016	-2.174	0.0298
INTERGEN	1	0.003326	0.00085016	3.912	0.0001

Durbin-Watson D	1.796
(For Number of Obs.)	1421
1st Order Autocorrelation	0.100

CHIAPAS (REGION 3)
Analysis of Variance

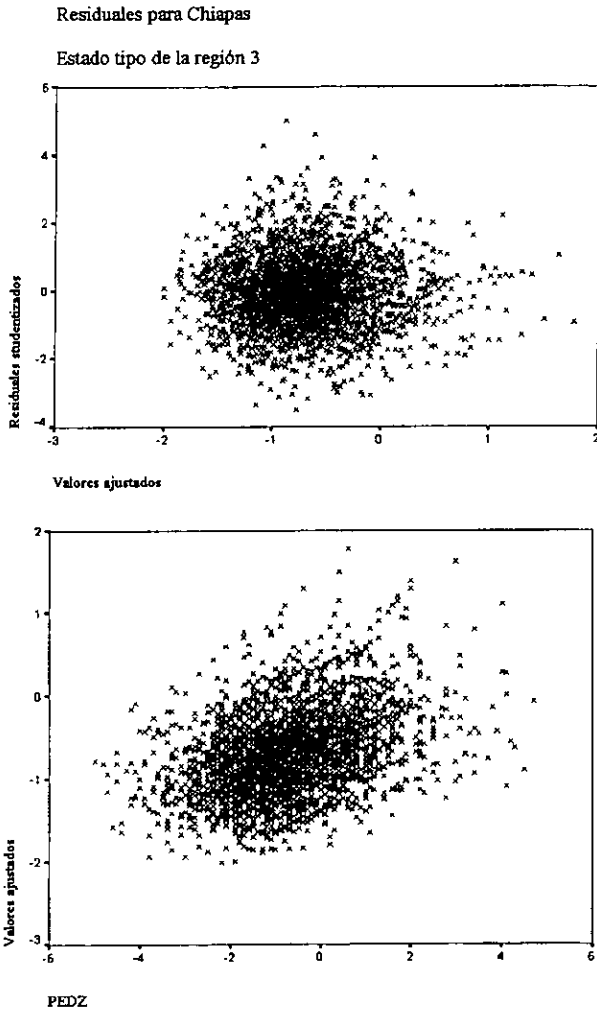
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	7	1586.28977	226.61282	153.494	0.0001
Error	1414	2087.57023	1.47636		
U Total	1421	3673.86000			

Root MSE 1.21505 R-square 0.4318
 Dep Mean -0.96383 Adj R-sq 0.4290
 C.V. -126.06548

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
TECHO	1	-0.252593	0.05157373	-4.898	0.0001
CALENT	1	0.153294	0.05553558	2.760	0.0058
GASTOPC	1	0.004683	0.00211203	2.217	0.0268
IDIOMA	1	0.233243	0.07273290	3.207	0.0014
LEDAD	1	-0.371727	0.04182682	-8.887	0.0001
DOCE	1	-0.093204	0.04643968	-2.007	0.0449
INTERGEN	1	0.003005	0.00082764	3.630	0.0003

Durbin-Watson D 1.795
 (For Number of Obs.) 1421
 1st Order Autocorrelation 0.101

Gráfica 3.6: Valores ajustados contra los residuales studentizados y contra la variable respuesta para el modelo en Chiapas con ordenada al origen.



§3.5 Modelos de regresión polinomiales

Además de las transformaciones de los regresores que se han sugerido a partir de el análisis gráfico de los residuales en los modelos hasta ahora propuestos, se consideraron *regresiones polinomiales*, hasta tercer grado. Utilizando los métodos de selección de variables se obtienen algunos modelos plausibles. En estos casos el ajuste no es mucho mejor que el de los modelos propuestos en la sección 3.2, pero la interpretación de los coeficientes de regresión de las variables si se complica un poco. Aparecen los cuadrados de algunos regresores y algunas interacciones. El comportamiento de los residuales tampoco es muy diferente a las gráficas ya presentadas correspondientes a modelos anteriormente propuestos.

La tabla 3.9 resume los modelos polinomiales por región y la tabla 3.10 el modelo para el total nacional incluyendo a la región como variable explicativa. En cuanto a la notación de las variables transformadas que aparecen, se resumen en la siguiente tabla, las variables que se consideraron en los modelos sin transformaciones se han descrito en la tabla 3.1.

Tabla 3.8. Variables transformadas consideradas en los modelos de regresión polinomiales.

Etiqueta	Variables
CALENSQ	Cuadrado de Lugar donde calienta sus alimentos
EDAMASQ	Cuadrado de Edad de la Madre
LEDADSQ	Cuadrado de Edad del menor
ABLACBSQ	Cuadrado de Edad de Ablactación
DOCESQ	Cuadrado de Tipo de alimentación
EDASESC	Edad * Escolaridad de la madre

Los modelos de regresión lineal y polinomial propuestos en este capítulo describen porcentajes bajos de la variabilidad de la variable respuesta a partir de los regresores. Esta situación se encontró en algunos casos de la literatura al modelar la desnutrición con base en indicadores antropométricos a partir de variables socioeconómicas. Es probable que los factores que no fueron medidos, que ya se han mencionado, justifiquen y describan mejor la variabilidad del peso para la edad de menores cuyas condiciones no concuerdan con el estatus de su peso para la edad. Además de esto, la heterogeneidad de varianzas que se debe al diseño puede repercutir en la calidad del ajuste.

te, así como en la evaluación del mismo. Esto último pues las pruebas de hipótesis dependen de la estimación de la varianza.

Tabla 3.9. Modelos de Regresión polinomial por región.
Región: 1.Noreste, Norte-Occidente; 2.Centro; 3.Centro-Sur.

REGION=1					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	10	722.66253	72.26625	46.572	0.0001
Error	4899	7601.83887	1.55171		
C Total	4909	8324.50140			
Root MSE		1.24568	R-square	0.0868	
Dep Mean		-0.26943	Adj R-sq	0.0849	
C.V.		-462.33853			
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-2.225378	0.47639082	-4.671	0.0001
REFRI_1	1	0.205208	0.03828387	5.360	0.0001
GASTOPC	1	0.004700	0.00082913	5.668	0.0001
EDADMA_1	1	0.076816	0.02394717	3.208	0.0013
ESCOLA	1	0.363125	0.09046626	4.014	0.0001
LEDAD	1	-0.155595	0.05529630	-2.814	0.0049
INTERGEN	1	0.002920	0.00049682	5.878	0.0001
LEDADSQ	1	-0.034694	0.01171113	-2.962	0.0031
CALENSQ	1	0.027744	0.00765724	3.623	0.0003
EDAMASQ	1	-0.000650	0.00031353	-2.074	0.0382
EDADESC	1	-0.008843	0.00288174	-3.069	0.0022
Durbin-Watson D			1.910		
(For Number of Obs.)			4910		
1st Order Autocorrelation			0.045		

REGION=2
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	11	969.36491	88.12408	64.798	0.0001
Error	6020	8187.11455	1.35999		
C Total	6031	9156.47946			

Root MSE	1.16618	R-square	0.1059
Dep Mean	-0.74453	Adj R-sq	0.1042
C.V.	-156.63379		

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > T
INTERCEP	1	-0.699794	0.12316675	-5.682	0.0001
HACINA	1	-0.041449	0.00853088	-4.859	0.0001
GASTOPC	1	0.006653	0.00102930	6.464	0.0001
EDADMA_1	1	0.007747	0.00231617	3.345	0.0008
ESCOLA	1	0.064740	0.01616915	4.004	0.0001
LEDAD	1	-0.161051	0.02707076	-5.949	0.0001
ABLACTAB	1	-0.026600	0.00600207	-4.432	0.0001
DOCE	1	-0.400769	0.06568533	-6.101	0.0001
INTERGEN	1	0.002696	0.00040147	6.714	0.0001
CALENSQ	1	0.028779	0.00565292	5.091	0.0001
ABLACSQ	1	0.000275	0.00005931	4.642	0.0001
DOCESQ	1	0.102963	0.01987210	5.181	0.0001

Durbin-Watson D	1.892
(For Number of Obs.)	6032
1st Order Autocorrelation	0.054

§3.5 Modelos de regresión polinomiales

REGION=3
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	12	1560.25274	130.02106	84.937	0.0001
Error	7005	10723.17640	1.53079		
C Total	7017	12283.42914			

Root MSE	1.23725	R-square	0.1270
Dep Mean	-1.06191	Adj R-sq	0.1255
C.V.	-116.51156		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-1.194886	0.09765587	-12.236	0.0001
PISO	1	0.060339	0.01673747	3.605	0.0003
EXCRETAS	1	0.080446	0.02144860	3.751	0.0002
TV	1	0.119017	0.03275730	3.633	0.0003
GASTOPC	1	0.003175	0.00095727	3.316	0.0009
IDIOMA	1	0.216837	0.03349876	6.473	0.0001
EDADMA_1	1	0.005515	0.00214622	2.570	0.0102
ESCOLA	1	0.067308	0.01476496	4.559	0.0001
LEDAD	1	-0.236588	0.02411999	-9.809	0.0001
DOCE	1	-0.511732	0.06618727	-7.732	0.0001
INTERGEN	1	0.002577	0.00038405	6.711	0.0001
CALENSQ	1	0.022373	0.00640649	3.492	0.0005
DOCESQ	1	0.134721	0.02207045	6.104	0.0001

Durbin-Watson D	1.820
(For Number of Obs.)	7018
1st Order Autocorrelation	0.090

Tabla 3.10. Modelos de Regresión polinomial para el total Nacional con región.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	12	4935.21775	411.26815	276.563	0.0001
Error	17761	26411.84734	1.48707		
C Total	17773	31347.06510			
Root MSE		1.21945	R-square	0.1574	
Dep Mean		-0.74464	Adj R-sq	0.1569	
C.V.		-163.76348			
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-0.265559	0.06993245	-3.797	0.0001
PISO	1	0.046732	0.01068481	4.374	0.0001
EXCRETAS	1	0.065523	0.01253302	5.228	0.0001
HACINA	1	-0.020042	0.00514200	-3.898	0.0001
REFRI_1	1	0.138275	0.02352701	5.877	0.0001
GASTOPC	1	0.004860	0.00053841	9.026	0.0001
IDIOMA	1	0.222827	0.02557573	8.712	0.0001
LEDAD	1	-0.216753	0.01459810	-14.848	0.0001
DOCE	1	-0.399595	0.03977880	-10.045	0.0001
INTERGEN	1	0.002643	0.00024458	10.806	0.0001
REGION	1	-0.201139	0.01316628	-15.277	0.0001
CALENSQ	1	0.025534	0.00389119	6.562	0.0001
DOCESQ	1	0.100269	0.01228380	8.163	0.0001
Durbin-Watson D			1.869		
(For Number of Obs.)			17774		
1st Order Autocorrelation			0.065		

Capítulo 4

Efectos de diseño en los modelos de regresión.

El análisis de regresión, no solo en el ámbito de estudios de salud, se basa con frecuencia en información obtenida a partir de encuestas. Los diseños de estas encuestas por diversas razones, que van desde precisión hasta costos y restricciones geográficas, en general consisten en varias etapas de muestreo que involucran la selección de conglomerados anidados dentro de conglomerados o estratos en etapas anteriores, exceptuando la última etapa. Los conglomerados y estratos seleccionados en las etapas son grupos homogéneos como hogares, hospitales, localidades, etcétera. Estos diseños rara vez satisfacen los supuestos de independencia que requieren los procedimientos para hacer inferencias, en particular los modelos de regresión.

Para considerar los *efectos del diseño* en los modelos algunos autores como, Sugden & Smith (1984), Pfefferman & Smith (1985), Pfefferman & LaVange (Skinner et al, capítulo 12, 1989), Pfefferman (1996), sugieren varias formas de hacerlo. Escencialmente consideran inferencias para encuestas basadas en modelos e inferencias basadas en el diseño, utilizando métodos aleatorizados que incorporen *ponderadores* o pesos muestrales (*sampling weights*), estos ponderadores corresponden al inverso de la probabilidad de inclusión de un elemento en la muestra y se conocen a menudo como *factores de expansión*. Los pesos muestrales asociados a cada individuo representan el número de individuos de la población en estudio que cada uno representa. La forma de incorporar los ponderadores depende de los objetivos y de la metodología estadística que se este utilizando para abordar el problema.

Pfefferman (1993) señala que el uso de pesos muestrales en inferencias sobre los parámetros de un modelo es aún muy controvertido, puesto que algunos autores consideran a los pesos muestrales irrelevantes, mientras que otros los incluyen en cualquier análisis.

En particular, al ajustar modelos de regresión a los datos que provienen de encuestas complejas, el agrupamiento de la población en conglomerados debe tomarse en cuenta, a menos que la relación entre la variable dependiente y los regresores sea la misma a lo largo de los grupos.

§4.1 Inferencias basadas en modelos.

Bajo el enfoque de inferencias basadas en modelos, se considera la idea de *variables de diseño*. Para estas variables se conocen los valores para todos los elementos de la población antes de obtener la muestra. Para una población finita de tamaño N se definen las variables de diseño como $z = (z_1, \dots, z_N)'$ que pueden incluir información que etiquete al estrato o conglomerado correspondiente como variables indicadoras que determinen la pertenencia a los grupos o bien variables para los grupos como medidas de tamaño.

En este caso la variable respuesta $y = (y_1, \dots, y_N)'$, para la que solo se conocen s valores que están en la muestra, se considera una realización de la matriz aleatoria Y . El caso degenerado en donde $Y = y$ es fija se utiliza en el enfoque clásico de inferencias aleatorizadas. Las variables de diseño z pueden considerarse ya sea como realizaciones aleatorias de una matriz o bien como fijas en la distribución condicional de Y dado z .

Considerando la partición de la población respecto a la extracción de una muestra, tenemos dos conjuntos s y s^c de donde se pueden definir y_s, y_{s^c} y z_s, z_{s^c} . La selección de la muestra observada s es la parte final de la encuesta y en diseños probabilísticos se lleva a cabo de acuerdo a un conjunto de probabilidades de selección asignadas

$$[s, p(s), s \in \mathcal{F}],$$

donde \mathcal{F} es el conjunto de todas las muestras posibles. Suponiendo que el proceso de selección de la muestra puede modelarse a través de las variables z y que puede depender también de la variable respuesta y y de un vector

de parámetros ψ de la forma:

$$p(s | y, z : \psi) \quad s \in \mathcal{F}$$

Esta es la distribución discreta de la variable indicadora I que toma valor uno si el elemento está en la muestra y cero en otro caso. En estos diseños se asume que I es independiente de $Y | Z$, esto es, la variable indicadora S y la variable respuesta Y son condicionalmente independientes dados los valores de la matriz Z para las N unidades de la población, $\mathbf{P}(S) = \mathbf{P}(I | Y, Z; \psi)$. Ahora bien, $f(Y | Z; \theta)$ denota la densidad de probabilidad condicional, que es la distribución poblacional antes de obtener la muestra, de $Y | Z$ indexada por el parámetro θ . Sea $g(Z; \phi)$ la distribución marginal de Z . La distribución conjunta de Y y Z en la población es

$$f(Y, Z; \theta, \phi) = f(Y | Z; \theta)g(Z; \phi)$$

donde se asume que θ y ϕ son parámetros distintos. En la inferencia analítica el parámetro objetivo es θ ó una función de θ y ϕ .

Suponiendo ahora que las variables de diseño se conocen para cada elemento de la población y que las variables respuesta solo se conocen para los elementos en la muestra. La distribución conjunta de Y_s , Z e I se obtiene integrado la distribución conjunta de Y , Z e I sobre Y_{s^c} ,

$$f(Y_s, I, Z; \theta, \phi, \psi) = \int f(Y_s, Y_{s^c} | Z; \theta)g(Z; \phi)\mathbf{P}(I | Y_s, Y_{s^c}, Z; \psi)dY_{s^c} \quad (4.1)$$

Ignorar el diseño muestral en las inferencias significa que éstas se basen en la distribución conjunta de Y_s y Z obtenida integrando sobre Y_{s^c} , ignorando $\mathbf{P}(I | Y_s, Y_{s^c}, Z; \psi)$,

$$f(Y_s, Z; \theta, \phi) = \int f(Y_s, Y_{s^c} | Z; \theta)g(Z; \phi)dY_{s^c} \quad (4.2)$$

El hecho de ignorar el diseño significa ignorar la información que provee el esquema de selección más allá de la información que sobre esto ha dado la variable Z . Esto se satisface si el esquema de selección de la muestra depende únicamente de las variables de diseño, $\mathbf{P}(I | Y_s, Y_{s^c}, Z; \psi) = \mathbf{P}(I | Z; \phi)$. Esto se debe a que al condicionar con Z , $\mathbf{P}(I | Y_s, Y_{s^c}, Z; \psi)$ es una constante y las inferencias con (4.1) y (4.2) son equivalentes.

En muchas ocasiones la información sobre Z no esta disponible, existe únicamente información parcial sobre el diseño como pueden ser las probabilidades de selección, Sudgen & Smith(1984) introducen el concepto de *información parcial*. Si la información disponible la denotamos como $D_s(Z) = d_s$, entonces si $P(I | Z = z) = P(I | D_s(Z) = d_s)$, para todo z tal que $D_s(z) = d_s$, lo que implica que $f(Y | D_s, I) = f(Y | D_s)$, el efecto de diseño puede ser ignorado en la infrencia.

Lo anterior se ha formulado en términos de la distribución conjunta de Y_s y Z , pero se puede traducir al caso donde los modelos para hacer inferencias especifican la distribución condicional de algunas de las variables respuesta dadas otras variables de la encuesta, como es el caso del modelos de regresión. Entoces los efectos del diseño son ignorables para estima la regresión de Y sobre X si

$$f(Y_s | X_s, D_s, I) = f(Y_s | X_s, D_s).dY_s$$

Adicionalmente si, $f(Y_s | X_s, D_s) = f(Y_s | X_s)$ la información sobre el diseño que aporta D_s puede también ignorarse y aplicar el análisis de regresión clásico.

Sin embargo, condicionar para los valores de las variables de diseño no siempre es posible o conveniente. Ya sea por la disponibilidad de la información o porque extender el modelo, en este caso de regresión, con la inclusión de variables de diseño como regresores adicionales, puede no ser deseable.

Los métodos de estimación para modelos de regresión para un modelo de varios niveles son complejos y requieren métodos iterativos para obtener los componentes de varianza como un paso necesario para la estimación global. Si se considera que el proceso incluye la estimación de muchos modelos dentro de un proceso de selección de variables por ejemplo, esto representa una inversión de tiempo.

§4.2 Inferencias basadas en el diseño.

En las inferencias basadas en el diseño se pueden considerar los pesos de cada observación. Los estimadores ponderados por la probabilidad de selección, a diferencia de los basados en modelos, son fáciles de obtener, para

el caso de regresión lineal, a través de los programas que implementan los procedimientos de mínimos cuadrados ponderados como SAS.

Por otra parte, si consideramos a toda la población de tamaño N , el estimador $\hat{\beta}$ para β bajo el modelo de regresión $\mathbf{y} = \mathbf{X}\beta + \varepsilon$; $\text{Var}(\varepsilon) = \sigma^2 I$, obtenido utilizando mínimos cuadrados ordinarios, es $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. El estimador ponderado obtenido a través de los mínimos cuadrados ponderados es $\hat{\beta}_w = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$, donde \mathbf{W} es una matriz diagonal con los pesos $w_i = 1/\pi_i$ ordenados en la diagonal principal. Donde $\mathbf{P}(i \in S) = \pi_i$ corresponden a las probabilidades de inclusión en la muestra.

Al considerar que los parámetros que se desean estimar corresponden a modelos que consideran a toda la población de la que se extrae la muestra, ha dado origen a preguntarse si el hecho de que las inferencias provengan de la información de una muestra cambia los parámetros de interés. En cuanto a esto Pfefferman (1993), señala que las distintas posturas de los autores en torno a esta pregunta abarcan desde aquellos que no consideran necesario tomar en cuenta que la información proviene de una muestra mientras los modelos consideran a toda la población, hasta aquellos que piensan que toda la inferencia debe hacerse en torno a parámetros que consideran las variables muestrales. Una posición intermedia de algunos involucra considerar los parámetros de los modelos como objetivo final, pero se concentra en los parámetros que se basan en la muestra, asegurando en cierta forma la robustez de la inferencia. Esta situación da origen al concepto de *consistencia en diseño*. Si la idea de consistencia involucra un comportamiento límite cuando el tamaño de la muestra tiende a infinito, la consistencia en diseño considera que la muestra y la población crecen conservando su estructura de manera que los parámetros obtenidos tiendan a los parámetros objetivo en la población. Entonces

$$plim_{n \rightarrow \infty, N \rightarrow \infty} [t_s(n) - T(N)] = 0,$$

donde $t_s(n)$ es la estadística muestral y $T(N)$ es la estadística definida para toda la población y que proviene de algún proceso de minimización de una función de pérdida o de maximización de la verosimilitud, n es el tamaño de la muestra, N es el tamaño de la población y *plim* es la probabilidad límite bajo la distribución de la aleatorización. Esta consistencia de estima-

dores consistentes en diseño (DC) para estimadores basados en el modelo correspondientes a los parámetros de éste puede establecerse como:

$$(t_s - \theta) = (t_s - T) + (T - \theta) = O_p(n^{1/2}) + O_p(N^{1/2}) = O_p(n^{1/2}),$$

donde la medida de probabilidad $O_p(n^{1/2})$ es con respecto a la distribución de la aleatorización y $O_p(N^{1/2})$ es la medida de probabilidad bajo el modelo. Estas dos medidas de probabilidad definen las tasas de convergencia a cero de cada diferencia. El tamaño n de la muestra y N de la población se refieren a las unidades muestrales que definen a estas tasas. En una muestra de conglomerados, por ejemplo, se refieren al número de conglomerados en la población y en la muestra, más que en los individuos seleccionados.

En cuanto a la varianza del estimador basado en la información de la muestra alrededor de θ , es decir, la varianza sobre todas las posibles muestras de la población finita y todas las posibles realizaciones de la población bajo el modelo, se descompone como:

$$\begin{aligned} \text{Var}(t_s) &= \mathbf{E}_M[\text{Var}_D(t_s | Y)] + \text{Var}_M[\mathbf{E}_D(t_s | Y)] \\ &= \mathbf{E}_M[\text{Var}_D(t_s | Y)] + O_p(N^{-1}), \end{aligned}$$

donde $Y = (Y_1, \dots, Y_N)$ denota los valores poblacionales y \mathbf{E}_M y Var_M definen la esperanza y la varianza bajo el modelo. Entonces en la práctica, cuando la población es mucho mayor que la muestra, la varianza de t_s alrededor de θ es aproximadamente igual a la esperanza bajo el modelo de la varianza de la aleatorización y puede estimarse al estimar la varianza de la aleatorización.

El estimador ponderado $\hat{\beta}_w$ es aproximadamente insesgado en diseño y consistente para $\hat{\beta}$, que es como ya se ha mencionado un estimador insesgado y consistente para los coeficientes del modelo. Lo anterior justifica el uso de estimadores ponderados con pesos muestrales para ajustar modelos bajo efectos de diseño no ignorables.

§4.3 Inferencias utilizando pesos muestrales para la ENAL'96

Los pesos muestrales que se utilizaran para considerar y evaluar el efecto de diseño en los modelos que se han propuesto corresponden a los factores de expansión de los menores entrevistados. Los factores de expansión en cada una de las etapas del muestreo se calcularon la siguiente forma:

Tabla 4.1. Factores de expansión por etapa de muestreo.

Factor de expansión por localidad Primera etapa de muestreo	Total de localidades en el estrato h / Localidades encuestadas en el estrato h
Factor de expansión por vivienda Segunda etapa de muestreo	Total de viviendas en la localidad i -ésima de estrato h / Viviendas encuestadas en la localidad i -ésima de estrato h
Factor de expansión por hogar Tercera etapa de muestreo	Total de menores de 6 años en la j -ésima familia de la localidad i -ésima del estrato h / Menores de 6 años encuestados en la j -ésima familia de la localidad i -ésima del estrato h

El peso muestral que se asignó a cada uno de los menores entrevistados w_i corresponde entonces al producto de los factores de expansión en cada etapa del muestreo. En la gráfica 4.1 se puede observar la distribución de los pesos en un histograma. La tabla 4.2 resume por estado a la población rural que vive en localidades de 500 a 2499 habitantes de acuerdo al X Censo Nacional de Población y Vivienda de 1990, al Conteo Rápido de 1995 y la población a la que expanden los pesos w_i . La población tanto en el censo como en el conteo rápido de 1995 es la población rural total, mientras que la expandida a partir de la muestra se refiere a los menores de 5 años que viven en áreas rurales. Sin embargo, se espera que los porcentajes respecto a la población total se mantengan por estado entre estos.

Algunos estados como Guanajuato, Jalisco, Quintana Roo y Sinaloa están bajorepresentados con una diferencia del orden de un 1%. EL Estado de México se encuentra por debajo del total en el conteo rápido en un 2% que es el más notable. En contraparte, los estados de Puebla y Michoacán tienen en la muestra porcentajes de población que exceden el total del conteo en un 2.9% y 2.2% respectivamente. Los estados de Chiapas y Oaxaca también están ligeramente sobrerrepresentados.

Ahora bien, la matriz W es una matriz positiva definida que tiene en la diagonal los factores de expansión para cada observación. Para el modelo $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ la varianza del término de error esta dada como $\text{Var}(\varepsilon) = \sigma^2 W$.

En ese caso la suma de cuadrados de los residuales que se desea minimizar es la función

$$(\mathbf{y} - \mathbf{X}\beta)'W^{-1}(\mathbf{y} - \mathbf{X}\beta)$$

es decir esta suma de cuadrados generalizada da más peso a algunos de los residuales. La solución de la expresión anterior puede obtenerse a través de la solución ya conocida de mínimos cuadrados ordinarios. Si se toma en cuenta que el modelo de regresión que se ha enunciado varía del anterior únicamente en que la varianza del error es $\text{Var}(\varepsilon) = \sigma^2 W$. Entonces si existe una matriz C simétrica de $n \times n$ tal que, $C'C = CC' = W^{-1}$, conocida como la raíz cuadrada de W^{-1} . La $\text{Var}(C\varepsilon)$ se puede escribir como

$$\begin{aligned}\text{Var}(C\varepsilon) &= C(\sigma^2 W)C' \\ &= \sigma^2 C(C'C)^{-1}C' \\ &= \sigma^2 I_n.\end{aligned}$$

Por otra parte si multiplicamos ambos lados de la ecuación del modelo de regresión por C se obtiene un nuevo modelo de regresión

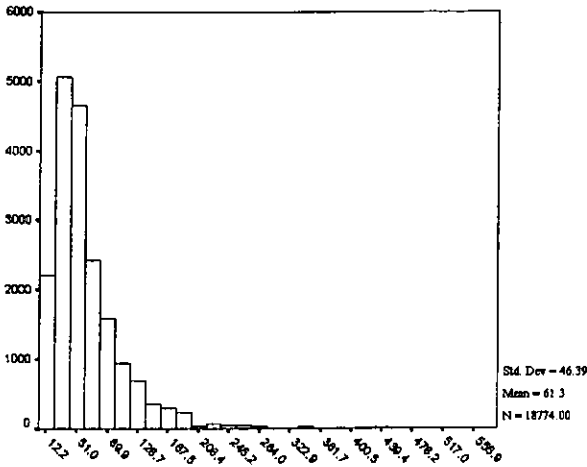
$$C\mathbf{y} = C\mathbf{X}\beta + C\varepsilon.$$

Entonces si se definen $K = C\mathbf{y}$, $M = C\mathbf{X}$ y $d = C\varepsilon$ se tiene un modelo de regresión con varianza constante y

$$\begin{aligned}\hat{\beta} &= [(M'M)^{-1}M'K] \\ &= [(C\mathbf{X})'(C\mathbf{X})]^{-1}(C\mathbf{X})'C\mathbf{y} \\ &= [\mathbf{X}'C'C\mathbf{X}]^{-1}(\mathbf{X}'C'C\mathbf{y}) \\ &= [\mathbf{X}'W^{-1}\mathbf{X}]^{-1}(\mathbf{X}'W^{-1}\mathbf{y}).\end{aligned}$$

Entonces se pueden obtener las soluciones para los estimadores ponderados $\hat{\beta}_w$ a través de mínimos cuadrados ordinarios, multiplicando las matrices \mathbf{X}

Gráfica 4.3: Histograma para los Factores de expansión.



y y por C que en este caso es una matriz diagonal con la raíz cuadrada de los w_i en la diagonal¹.

Algunos autores enfatizan que el uso de estimadores ponderados con los pesos muestrales remueven el sesgo pero la varianza de estos estimadores se incrementa. En el caso de los coeficientes de regresión se esperaría que la varianza de los estimadores ponderados se incrementara.

Las tablas 4.3 y 4.4 contienen los resultados de regresiones ponderadas con los factores de expansión, es importante enfatizar que los coeficientes de regresión no cambian de manera importante. Además de esto, las varianzas estimadas no crecen de manera notable, como se esperaba al utilizar estimadores ponderados para los coeficientes de regresión.

¹WEISBERG SANFORD. *Applied linear regression (1985)*

Es posible que la forma de incorporar los pesos no este considerando todas las complicaciones del diseño. En la literatura se proponen distintas maneras de incorporar la información de los factores de expansión para considerar el efecto del diseño. Se proponen por ejemplo modelos que consideren los conglomerados, sin embargo, los tamaños de los conglomerados son muy pequeños, un 44.1% de los hogares en la muestra tienen solo un menor cuya edad sea de interés entrevistado.

A partir de estos modelos que se presentan como ejemplo, se concluye que pueden ignorarse los efectos de diseño para los modelos de regresión.

Tabla 4.2. Población y porcentaje de población en localidades de 500 a 2499 habitantes por estado.

Estado	Censo 90		Conteo Rápido 95		Factores de Expansión	
	Población	%	Población	%	Población	%
Aguascalientes	98195	0.7	122472	0.9	9549	0.8
Baja California	88359	0.7	102654	0.8	13123	1.1
Baja California Sur	29579	0.2	36881	0.3	704	0.1
Campeche	109826	0.8	101667	0.7	11003	1.0
Coahuila	139885	1.0	127904	0.9	12158	1.1
Colima	40628	0.3	41010	0.3	2774	0.2
Chiapas	950879	7.0	985399	7.3	97572	8.5
Chihuahua	183117	1.4	191262	1.4	16615	1.4
Distrito Federal	0	0.0	9536	0.1	0	0.0
Durango	289059	2.1	278968	2.1	19596	1.7
Guanajuato	781991	5.8	863718	6.4	57227	5.0
Guerrero	726509	5.4	776275	5.7	59650	5.2
Hidalgo	601492	4.5	637508	4.7	50668	4.4
Jalisco	460792	3.4	482912	3.6	29218	2.5
México	1010980	7.5	1221262	9.0	71752	6.2
Michoacán	818361	6.1	775775	5.7	90341	7.9
Morelos	134771	1.0	136512	1.0	10446	0.9
Nayarit	199846	1.5	219683	1.6	14611	1.3
Nuevo León	66424	0.5	77784	0.6	4111	0.4
Oaxaca	1534533	11.4	985160	7.3	97803	8.5
Puebla	1017802	7.5	1039301	7.7	121764	10.6
Querétaro	270295	2.0	277692	2.0	23491	2.0
Quintana Roo	78412	0.6	91227	0.7	5139	0.4
San Luis Potosí	426411	3.2	437645	3.2	39256	3.4
Sinaloa	439349	3.3	447643	3.3	25388	2.2
Sonora	182858	1.4	186268	1.4	11415	1.0
Tabasco	562408	4.2	581356	4.3	52991	4.6
Tamaulipas	156922	1.2	159011	1.2	12029	1.0
Tlaxcala	135852	1.0	137648	1.0	10339	0.9
Veracruz	1375557	10.2	1435678	10.6	125924	10.9
Yucatán	196386	1.5	216496	1.6	18365	1.6
Zacatecas	384670	2.9	372014	2.7	35453	3.1
Nacional	13,492,148	100.0	13,556,321	100.0	1,150,475	100.0

Tabla 4.3 Modelos de Regresión por Región.
Región: 1.Noreste, Norte-Occidente; 2.Centro; 3.Centro-Sur.

REGION=1					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	9	53615.68204	5957.29800	66.006	0.0001
Error	4760	429609.43695	90.25408		
C Total	4769	483225.11900			
Root MSE		9.50021	R-square	0.1110	
Dep Mean		-0.27939	Adj R-sq	0.1093	
C.V.		-3400.37894			
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-0.676519	0.15464580	-4.375	0.0001
EXCRETAS	1	0.066137	0.02488912	2.657	0.0079
CALENT	1	0.080737	0.03549155	2.275	0.0230
HACINA	1	-0.007088	0.01209837	-0.586	0.5580
REFRI_1	1	0.210277	0.04045517	5.198	0.0001
GASTOPC	1	0.005940	0.00080339	7.393	0.0001
ESCOLA	1	0.064554	0.02120373	3.044	0.0023
IDIOMA	1	0.291118	0.08721858	3.338	0.0009
LEDAD	1	-0.340736	0.01995004	-17.079	0.0001
INTERGEN	1	0.002976	0.00049328	6.032	0.0001
Durbin-Watson D			1.888		
(For Number of Obs.)			4770		
1st Order Autocorrelation			0.056		

REGION=2
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	10	54694.63771	5469.46377	62.878	0.0001
Error	5541	481985.81054	86.98535		
C Total	5551	536680.44825			

Root MSE	9.32659	R-square	0.1019
Dep Mean	-0.72498	Adj R-sq	0.1003
C.V.	-1286.46784		

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > T
INTERCEP	1	-0.784594	0.12538432	-6.258	0.0001
CALENT	1	0.098336	0.02604207	3.776	0.0002
HACINA	1	-0.042104	0.00913666	-4.608	0.0001
REFRI_1	1	0.146095	0.03953795	3.695	0.0002
GASTOPC	1	0.006396	0.00110378	5.795	0.0001
ACTJF3	1	0.016922	0.00794791	2.129	0.0333
ESCOLA	1	0.060573	0.01645075	3.682	0.0002
IDIOMA	1	0.153434	0.05425165	2.828	0.0047
LEDAD	1	-0.287254	0.02246541	-12.787	0.0001
ABLACTAB	1	0.001355	0.00083552	1.622	0.1050
INTERGEN	1	0.003341	0.00041968	7.961	0.0001

Durbin-Watson D	1.903
(For Number of Obs.)	5552
1st Order Autocorrelation	0.047

REGION=3
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	10	89687.43987	8968.74399	91.802	0.0001
Error	7007	684560.22043	97.69662		
C Total	7017	774247.66030			

Root MSE	9.88416	R-square	0.1158
Dep Mean	-1.04224	Adj R-sq	0.1146
C.V.	-948.35303		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-0.989367	0.09491600	-10.424	0.0001
PISO	1	0.032869	0.01747797	1.881	0.0601
EXCRETAS	1	0.089947	0.02116622	4.250	0.0001
CALENT	1	0.111696	0.02696894	4.142	0.0001
TV	1	0.128356	0.03375035	3.803	0.0001
GASTOPC	1	0.002394	0.00092259	2.595	0.0095
ESCOLA	1	0.074172	0.01432263	5.179	0.0001
IDIOMA	1	0.233196	0.03457886	6.744	0.0001
LEDAD	1	-0.253280	0.02412298	-10.500	0.0001
ABLACTAB	1	0.001024	0.00071522	1.432	0.1521
DOCE	1	-0.143063	0.02214226	-6.461	0.0001

Durbin-Watson D	1.828
(For Number of Obs.)	7018
1st Order Autocorrelation	0.086

Tabla 4.4 Modelos de Regresión a partir de variables e índice de hogar con región para el total Nacional.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	11	296003.02557	26909.36596	289.355	0.0001
Error	17804	1655733.3167	92.997827268		
C Total	17815	1951736.3422			
Root MSE	9.64354	R-square	0.1517		
Dep Mean	-0.73523	Adj R-sq	0.1511		
C.V.	-1311.62989				
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-0.485155	0.08223578	-5.900	0.0001
EXCRETAS	1	0.070382	0.01259722	5.587	0.0001
CALENT	1	0.107415	0.01629566	6.592	0.0001
HACINA	1	-0.017634	0.00525675	-3.354	0.0008
REFRI_1	1	0.149413	0.02346075	6.369	0.0001
GASTOPC	1	0.005125	0.00051989	9.857	0.0001
ESCOLA	1	0.069575	0.00950733	7.318	0.0001
IDIOMA	1	0.223840	0.02621224	8.540	0.0001
LEDAD	1	-0.321748	0.01281733	-25.103	0.0001
ABLACTAB	1	0.001496	0.00046577	3.211	0.0013
INTERGEN	1	0.002870	0.00024613	11.660	0.0001
REGION	1	-0.201801	0.01314257	-15.355	0.0001
Durbin-Watson D		1.868			
(For Number of Obs.)		17816			
1st Order Autocorrelation		0.066			

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	7	257839.54072	36834.22010	384.774	0.0001
Error	17776	1701688.2785	95.729538618		
C Total	17783	1959527.8192			

Root MSE	9.78415	R-square	0.1316
Dep Mean	-0.73457	Adj R-sq	0.1312
C.V.	-1331.95549		

Variable	DF	Parameter		T for H0:	
		Estimate	Standard Error	Parameter=0	Prob > T
INTERCEP	1	-0.545414	0.06405161	-8.515	0.0001
CPB11	1	0.118050	0.00726043	16.259	0.0001
GASTOPC	1	0.006408	0.00051696	12.396	0.0001
EDADMA_1	1	0.002537	0.00132857	1.909	0.0562
IDIOMA	1	0.261729	0.02576356	10.159	0.0001
ABLACTAB	1	0.005330	0.00040807	13.063	0.0001
DOCE	1	-0.203044	0.01094040	-18.559	0.0001
REGION	1	-0.208624	0.01330593	-15.679	0.0001

Durbin-Watson D	1.860
(For Number of Obs.)	17784
1st Order Autocorrelation	0.070

Capítulo 5

Conclusiones y Comentarios

Las encuestas de salud relacionadas a la desnutrición infantil en menores pre-escolares aportan información que puede ser importante para definir medidas que ayuden a combatir este problema. El alto costo que representa el levantamiento de una encuesta de la magnitud de la ENAL'96 merece explorar diferentes enfoques en los análisis de la información. Los hallazgos que esto ha permitido a lo largo de numerosos estudios han hecho posible reconocer los factores que inciden en la desnutrición infantil.

A lo largo del presente trabajo se han propuesto algunos modelos de regresión que describan la relación de variables socioeconómicas con en el *score z* del peso para la edad, medida antropométrica que puede considerarse como un indicador del estado de nutrición de los menores entrevistados en el medio rural de México. En estos modelos se han incluido las variables disponibles a partir de las cuales se explican las mayores proporciones de variabilidad de las respuesta, que como hemos resaltado, son relativamente bajos. Los modelos no pueden predecir al peso para la edad. Este hecho no sorprende a los expertos en estudios de nutrición. La presencia de muchos factores que no han sido medidos, por la dificultad que hacerlo representa, puede explicar esto en cierto sentido. Los niveles de alcoholismo de los padres y la violencia intrafamiliar, que afectan el cuidado que se proporciona al menor se encuentran entre estos factores. Además de otros, como aspectos migratorios, que nos se analizaron.

Entre las variables que aparecen en la mayoría de los modelos que se han propuesto, se encuentran la disposición de excretas de la vivienda, el lugar donde calientan los alimetos, la escolaridad de la madre, la edad del menor,

la edad de ablactación del menor, el intervalo intergenésico y el gasto per capita en alimentos.

Para la variable lugar donde calienta sus alimentos, se considera que el fogón en piso representa un riesgo de contaminación de los alimentos. Y ya desde el análisis exploratorio se observó que esta categoría presentaba mayor porcentaje de desnutrición que las otras. En cuanto a la disposición de excretas de la vivienda, los sistemas con arrastre de agua son más higiénicos y presentaron un menor porcentaje de desnutrición. Por otra parte, las madres indígenas, reflejadas por su idioma, y las madres con menor escolaridad son características de porcentajes mayores de desnutrición. Al intervalo intergenésico se le puede considerar como indicador del cuidado proporcionado al menor, mientras que el gasto per capita en alimentos es la variable asociada al nivel económico que se presenta en casi todas las regresiones.

También es interesante resaltar las diferencias regionales en el comportamiento del peso para la edad y de los regresores. La región Centro (2) es la que más se asemeja al comportamiento Nacional, de los casos ubicados en esta región el 14.5% presentan desnutrición. La región Centro-Sur (3) es la región con mayores porcentajes de desnutrición (23.7%) y quizá donde las variables consideradas mejor reflejan sus condiciones socioeconómicas. La región Noreste, Norte-Occidente (1) es la región donde solo el 8.7% de los menores entrevistados presentaron desnutrición. En esta región es donde las variables analizadas describen menor variabilidad del peso para la edad.

Otro aspecto considerado fue el diseño de la muestra, en relación a como se ven afectados los supuestos de los modelos de regresión. La información disponible relativa al diseño se tiene en los factores de expansión. Se ha abordado entonces uno de los enfoques que se ha planteado en la literatura que involucra considerar el efecto del diseño haciendo uso de estimadores ponderados por los pesos muestrales para estimar los coeficientes de regresión que son el parámetro objetivo que considera el modelo basado en la población. Sin embargo, se proponen muchos otros para considerar el efecto del diseño. Para los datos de esta encuesta y bajo un modelo de regresión ya se observó que los efectos del diseño pueden ignorarse. Una posibilidad en la utilización de los factores al considerar el diseño muestral es la calibración de éstos para que las muestras expandan al total censal de la población,

reportado por el conteo rápido de 1995.

Los menores entrevistados en los hogares conforman conglomerados cuya correlación habría sido conveniente considerar en los modelos, pero en un 44.1% de los hogares considerados se entrevistó solamente a un menor. En un futuro la utilización de modelos mixtos o jerárquicos se plantea como una posibilidad en donde puede considerarse el diseño.

Adicionalmente, una cuestión que es interesante mencionar es el comportamiento de la población considerada como rural de algunos estados de la república. En estados como Nuevo León, más del 50% de la población que vive en localidades de poblaciones menores a 2500 habitantes, lo hacen el poblaciones de 1 a 499 habitantes. En Chiapas la población está dividida en partes aproximadamente iguales entre localidades de 1 a 499 habitantes y de 500 a 2499 habitantes. Incluir a estas localidades en estudios de nutrición sería interesante, se sabe que entrevistar a niños que viven en localidades con menos de 500 habitantes resulta costoso, pero podrían considerarse como un estrato con una tasa menor de muestreo.

De acuerdo al análisis estadístico se puede concluir que a partir de la información de la ENAL'96, los modelos de regresión lineal y polinomial no logran predecir la desnutrición de los menores, utilizando el peso para la edad para medirla. Aunque si hay modelos que auxilian en la descripción de algunos factores de riesgo asociados a bajos niveles de peso para la edad.

También fue posible constatar que al considerarse el efecto del diseño, para los modelos de regresión sugeridos no se observan cambios significativos en los estimadores puntuales y por intervalo de los coeficientes de regresión. Por lo tanto, para el caso particular de modelos de regresión con el peso para la edad como variable respuesta, puede ignorarse el efecto de diseño.

Además de los modelos de regresión, se hicieron algunas pruebas con modelos logísticos, aunque no se incluyó un reporte de estas. Para ello se propuso como variable respuesta al peso para la edad dicotomizado, PEDZ2. Para las variables continuas, se utilizaron las categorizaciones descritas en el análisis exploratorio y se incluyeron como covariables junto con las variables categóricas. A partir de estos modelos se observó que, al igual que para las regresiones lineales y polinomiales, los ajustes pueden ayudar a describir factores de riesgo, pero no tienen capacidad predictiva. Todos los casos que

presentan desnutrición resultan mal clasificados.

El problema de predicción de la desnutrición presenta tres aspectos que deben reconsiderarse:

- a) Como medir la desnutrición infantil de los menores preescolares, las diferencias entre las medidas antropométricas en el caso de México.
- b) Considerar modelos más complejos que incorporen información de los conglomerados.
- c) Considerar información adicional que ayude a reflejar los aspectos relacionados con la desnutrición que no han sido medidos, quizá con esto los modelos aumenten su capacidad predictiva.

Apéndice A

Se anexan las cédulas que corresponden a los módulos:

- 2. Características de la vivienda y saneamiento,**
- 3. Recursos para la alimentación familiar,**
- 4. Composición familiar,**
- 7. Preescolares.**

Algunos aspectos sobre las variables pueden aclararse a través del cuestionario utilizado para obtener la información. Los módulos que se anexan son únicamente aquellos de los que se tomaron las variables analizadas en este trabajo.

Apéndice B

Variables consideradas en los modelos de regresión.

Etiqueta	Variables	Etiqueta	Variables
X_1	Material del Techo	X_2	Material del Piso
X_3	Lugar de donde obtiene el agua	X_4	Disposición de Excretas
X_5	Lugar donde calienta sus alimentos	X_6	Cocina separada
X_7	Indice de hacinamiento	X_8	Tiene TV
X_9	Tiene refrigerador	X_{10}	Gasto per capita
X_{11}	Actividad del jefe de Hogar	X_{12}	Edad de la Madre
X_{13}	Escolarida de la Madre	X_{14}	Idioma de la madre
X_{15}	Edad del menor	X_{16}	Edad de Ablactación
X_{17}	Tipo de alimentación	X_{18}	Intervalo intergenésico

Las tablas B1.1 y B1.2 resumen los modelos que se consideraron mejores de acuerdo a los criterios de selección de variables utilizados en el capítulo 2 dentro cada rubro y separando los casos por región. En cada caso la variable respuesta corresponde al score z del peso para la edad, pedz. Las variables de cada rubro que se seleccionaron en cada región se consideraron en un análisis posterior para proponer algunos modelos regionales y globales utilizando a las variables observadas como regresores.

Tabla B1.1 Modelos de Regresión por Bloque y por Región.
Región: 1.Noreste, Norte-Occidente; 2.Centro; 3.Centro-Sur.

No. de variables	Región	Bloque	Modelo	R ²
4	1	Vivienda	$\hat{Y} = -0.8648 + 0.0443X_2 + 0.1803X_5 + 0.0995X_4 - 0.0557X_7$	0.0241
4	2	Vivienda	$\hat{Y} = -1.1979 + 0.0769X_2 + 0.1576X_5 + 0.0822X_4 - 0.0586X_7$	0.0407
4	3	Vivienda	$\hat{Y} = -1.6693 + 0.1039X_2 + 0.1942X_5 + 0.1061X_4 - 0.0361X_7$	0.0400
2	1	Bienes, Materiales	$\hat{Y} = 0.3466 - 0.1343X_8 - 0.3056X_9$	0.0181
2	2	Bienes, Materiales	$\hat{Y} = -0.0317 - 0.1846X_8 - 0.2655X_9$	0.0198
2	3	Bienes, Materiales	$\hat{Y} = -0.3002 - 0.2538X_8 - 0.2187X_9$	0.0194
2	1	Ingreso, Actividad	$\hat{Y} = -0.5337 + 0.0072X_{10} - 0.0030X_{11}$	0.0155
2	2	Ingreso, Actividad	$\hat{Y} = -1.1258 + 0.0124X_{10} + 0.0242X_{11}$	0.0312
2	3	Ingreso, Actividad	$\hat{Y} = -1.2759 + 0.0085X_{10} + 0.0034X_{11}$	0.0115
2	1	Madre del menor	$\hat{Y} = -1.2444 + 0.1569X_{13} + 0.2251X_{14}$	0.0126
2	2	Madre del menor	$\hat{Y} = -1.7310 + 0.1298X_{13} + 0.3132X_{14}$	0.0220
2	3	Madre del menor	$\hat{Y} = -1.9885 + 0.1253X_{13} + 0.3384X_{14}$	0.0337
2	1	Menor	$\hat{Y} = -0.1155 - 0.0168X_{15} + 0.034X_{18}$	0.0455
2	2	Menor	$\hat{Y} = -0.6843 - 0.0146X_{15} + 0.0038X_{18}$	0.0373
2	3	Menor	$\hat{Y} = -0.4693 - 0.0086X_{15} - 0.2474X_{17}$	0.0572

Apéndice C

Programas en SAS

Los modelos de regresión lineal se obtuvieron con SAS System para Windows versión 6.12. Se presentan ejemplos de las rutinas que se utilizaron para seleccionar las variables y los modelos, los programas para los modelos que se sugieren.

```
/**/ Calculo de variables y algunos /**/
**/ejemplos de métodos de selección /**/
**/ variables. /**/
proc reg data=TESIS.POLIREG;
  model pedz= TECHO PISO AGUA EXCRETAS
  CALENT COCINA_1
  HACINA TVREFRI_1 GASTOPC ACTJF3
  EDADMA_1 ESCOLA IDIOMA
  ABLACTAB DOCE INTERGEN REGION
  edad ledadsq escolsq docesq hacinsq excrec
  calensq techosq pisosq
  aguasq gastsq actjfsq
  edamasq ablacsq intersq
  ledadcu escolcu docecu hacincu
  excrecu calencu techocu pisocu
  aguacu gastcu actjfcu edamacu ablaccu intercu
  reggas edadesc hacgast pisexc/
```

```
selection= stepwise dw B;
run;
proc reg data=TESIS.POLIREG;
  model pedz= TECHO PISO AGUA
  EXCRETAS CALENT COCINA_1
  HACINA TVREFRI_1 GASTOPC
  ACTJF3 EDADMA_1 ESCOLA IDIOMA
  ABLACTAB DOCE INTERGEN REGION
  edad ledadsq escolsq docesq hacinsq excresq
  calensq techosq pisosq
  aguasq gastsq actjfsq edamasq ablacsq intersq
  ledadcu escolcu docecu hacincu
  excrecu calencu techocu pisocu aguacu gastcu
  actjfcu edamacu ablaccu intercu
  reggas edadesc hacgast pisexc/
  selection=rsquare dw B;
  run;
proc reg data=TESIS.POLIREG;
  model pedz= TECHO PISO AGUA EXCRETAS
  CALENT COCINA_1
  HACINA TVREFRI_1 GASTOPC ACTJF3
  EDADMA_1 ESCOLA IDIOMA
  ABLACTAB DOCE INTERGEN REGION
  edad ledadsq escolsq docesq hacinsq excresq
  calensq techosq pisosq
  aguasq gastsq actjfsq edamasq ablacsq intersq
  ledadcu escolcu docecu hacincu
  excrecu calencu techocu pisocu aguacu gastcu
  actjfcu edamacu ablaccu intercu
  reggas edadesc hacgast pisexc/
  selection=adjsquare dw B;
  run;
proc reg data=TESIS.POLIREG;
by REGION;
```

```
model PEDZ= TECHO PISO AGUA EXCRETAS
CALENT COCINA_1
HACINA TV REFRI_1 GASTOPC ACTJF3 EDADMA_1
ESCOLA IDIOMA
ABLACTAB DOCE INTERGEN
ledad ledadsq escolsq docesq hacinsq excresq
calensq techosq pisosq
aguasq gastsq actjfsq edamasq ablacsq intersq
edadesc hacgast pisexc/
selection= stepwise B;
run;
**/**/**/**/**MODELOS PLAUSIBLES **/**/**/**/
proc reg data=TESIS.POLIREG;
model PEDZ = EXCRETAS CALENT HACINA REFRI_1
GASTOPC ESCOLA IDIOMA LEDAD ABLACTAB
INTERGEN REGION
/DW B;
run;
proc reg data=TESIS.POLIREG;
model PEDZ = CPB11 CPB21 CPB31 CPB41 CPB51
REGION
/DW B;
run;
proc reg data=TESIS.POLIREG;
BY REGION;
model PEDZ = EXCRETAS CALENT HACINA REFRI_1
GASTOPC
ESCOLA IDIOMA LEDAD INTERGEN/DW B;
run;
proc reg data=TESIS.POLIREG;
BY REGION;
model PEDZ = PISO EXCRETAS CALENT TV GASTOPC
ESCOLA
IDIOMA LEDAD ABLACTAB DOCE/DW B;
```

```
run;
proc reg data=TESIS.POLIREG;
BY REGION;
model PEDZ = CALENT HACINA REFRI_1 GASTOPC
ACTJF3
  ESCOLA IDIOMA LEDAD ABLACTAB I
  NTERGEN/DW B;
run;
proc reg data=TESIS.POLIREG;
BY REGION;
model PEDZ = CPB11 CPB21 CPB31 CPB41
CPB51 /DW B;

run;
proc reg data=TESIS.POLIREG;
model PEDZ = CPB11 GASTOPC EDADMA_1
IDIOMA
ABLACTAB DOCE REGION /DW B;
run;
proc reg data=TESIS.POLIREG;
BY REGION;
model PEDZ = CPB11 REFRI_1 GASTOPC
ESCOLA
ABLACTAB DOCE/DW B;
run;
proc reg data=TESIS.POLIREG;
BY REGION;
model PEDZ = CPB11 GASTOPC IDIOMA
ABLACTAB DOCE INTERGEN/DW B;
run;
proc reg data=TESIS.POLIREG;
BY REGION;
model PEDZ = CPB11 TV IDIOMA
ABLACTAB DOCE INTERGEN/DW B;
```

```
run;
**/**/**/**/ MODELOS PLAUSIBLES
CON WLS**/**/**/**/**/
proc reg data=TESIS.POLIREG;
model PEDZ = EXCRETAS CALENT HACINA REFRI_1
GASTOPC ESCOLA IDIOMA LEDAD ABLACTAB INTERGEN
REGION
/DW B;
WEIGHT F3B;
run;
proc reg data=TESIS.POLIREG;
model PEDZ = CPB11 CPB21 CPB31 CPB41
CPB51 REGION
/DW B;
WEIGHT F3B;
run;
proc reg data=TESIS.POLIREG;
BY REGION;
model PEDZ = EXCRETAS CALENT HACINA
REFRI_1 GASTOPC
ESCOLA IDIOMA LEDAD INTERGEN/DW B;
WEIGHT F3B;
run;
proc reg data=TESIS.POLIREG;
BY REGION;
model PEDZ = PISO EXCRETAS CALENT TV GASTOPC ESCOLA
IDIOMA LEDAD ABLACTAB DOCE/DW B;
WEIGHT F3B;
run;
proc reg data=TESIS.POLIREG;
BY REGION;
model PEDZ = CALENT HACINA REFRI_1 GASTOPC ACTJF3
ESCOLA IDIOMA LEDAD ABLACTAB INTERGEN/DW B;
WEIGHT F3B;
```



```
run;
proc reg data=TESIS.POLIREG;
BY REGION;
model PEDZ = CPB11 CPB21 CPB31 CPB41 CPB51 /DW B;
WEIGHT F3B;
run;
**/**/**/ MODELOS CON FACTOR COMO VARIABLE
ADICIONAL**/**/
**/**/**/NO SIGNIFICATIVOS**/**/
proc reg data=TESIS.POLIREG;
model PEDZ = EXCRETAS CALENT HACINA REFRI_1
GASTOPC ESCOLA IDIOMA LEDAD ABLACTAB INTERGEN
REGION F3B
/ DW B;
run;
proc reg data=TESIS.POLIREG;
model PEDZ = CPB11 CPB21 CPB31 CPB41 CPB51
REGION F3B
/DW B;
run;
proc reg data=TESIS.POLIREG;
BY REGION;
model PEDZ = EXCRETAS CALENT HACINA REFRI_1
GASTOPC
ESCOLA IDIOMA LEDAD INTERGEN F3B/DW B;
run;
proc reg data=TESIS.POLIREG;
BY REGION;
model PEDZ = PISO EXCRETAS CALENT TV GASTOPC
ESCOLA
IDIOMA LEDAD ABLACTAB DOCE F3B/DW B;
run;
proc reg data=TESIS.POLIREG;
BY REGION;
```

```
model PEDZ = CALENT HACINA REFRI_1 GASTOPC ACTJF3
  ESCOLA
IDIOMA LEDAD ABLACTAB INTERGEN F3B/DW B;
  run;
proc reg data=TESIS.POLIREG;
BY REGION;
model PEDZ = CPB11 CPB21 CPB31 CPB41 CPB51
  F3B/DW B;
  run;
/*ANALISIS PARA LOS TRES ESTADOS POR REGION*/**/**/
proc reg data=TESIS.EDOTIPOB;
BY CLAVEEDO;
model PEDZ = PISO COCINA_1 TV GASTOPC IDIOMA LEDAD
  ABLACTAB
  INTERGEN/DW B;
  run;
proc reg data=TESIS.EDOTIPOB;
BY CLAVEEDO;
model PEDZ = CALENT GASTOPC ESCOLA LEDAD ABLACTAB
  INTERGEN/DW B;
  run;
proc reg data=TESIS.EDOTIPOB;
BY CLAVEEDO;
model PEDZ = TECHO CALENT GASTOPC IDIOMA LEDAD DOCE
  INTERGEN/DW B;
  run;
proc reg data=TESIS.EDOTIPOB;
BY CLAVEEDO;
model PEDZ = TECHO CALENT GASTOPC IDIOMA LEDAD DOCE
  INTERGEN/DW B NOINT;
  run;
proc reg data=TESIS.EDOTIPOB;
BY CLAVEEDO;
model PEDZ = TECHO CALENT IDIOMA LEDAD DOCE
```

```
INTERGEN/DW B NOINT;
WEIGHT F3B;
run;
proc reg data=TESIS.EDOTIPOB;
BY CLAVEEDO;
model PEDZ = CPB11 GASTOPC IDIOMA
ABLACTAB DOCE INTERGEN/ DW B;
run;
proc reg data=TESIS.EDOTIPOB;
BY CLAVEEDO;
model PEDZ = CPB11 GASTOPC ESCOLA IDIOMA
ABLACTAB INTERGEN/DW B;
run;
proc reg data=TESIS.EDOTIPOB;
BY CLAVEEDO;
model PEDZ = CPB11 REFRI_1 GASTOPC IDIOMA
ABLACTAB DOCE INTERGEN/DW B;
run;
proc reg data=TESIS.EDOTIPOB;
BY CLAVEEDO;
model PEDZ = CPB11 CPB21 CPB31 CPB41 CPB51 /DW B;
run;
```




3.- RECURSOS PARA LA ALIMENTACION FAMILIAR

1

¿CUANTO GASTA A LA SEMANA EN ALIMENTOS?

N\$

¿LA FAMILIA, O ALGUNO DE SUS MIEMBROS RECIBEN ALGUN TIPO DE AYUDA ALIMENTARIA?

2 SI NO

¿CUAL?

- 3 TORTILLAS
 4 LECHE
 5 DESPENSA
 6 COCINAS POPULARES
 7 DESAYUNOS ESCOLARES
 8 OTRO _____

¿CRIA ANIMALES PARA ALIMENTACION?

9 SI NO

SI LA RESPUESTA FUE SI, MARCAR EL RECUADRO DE TIPO Y DESTINO

¿DE QUE TIPO?

¿A QUE LOS DESTINA?

AUTOCONSUMO VENTA AMBOS

- 10 GANADO MENOR
 11 GANADO MAYOR
 12 OTROS _____

¿CULTIVA ALIMENTOS EN CASA?

13 SI NO

SI LA RESPUESTA FUE SI, MARCAR EL RECUADRO DE TIPO Y DESTINO

¿DE QUE TIPO?

¿A QUE LOS DESTINA?

AUTOCONSUMO VENTA AMBOS

- 14 FRUTALES
 15 HORTALIZAS
 16 OTROS _____



4.- COMPOSICION FAMILIAR

ANOTAR LA INFORMACION SOLICITADA DE TODOS LOS MIEMBROS DE LA FAMILIA

CLAVE	M/F	EDAD		MAYORES DE 5 AÑOS		MAYORES DE 12 AÑOS	
		SEXO, PAREN. (AÑOS)	ESTADO (AÑOS)	ESCOLA- (AÑOS)	IDIOMA	OCUPACION PRINCIPAL	CONDICION
01		1					
02							
03							
04							
05							
06							
07							
08							
09							
10							
11							
12							
13							
14							

PARENTESCO CON EL JEFE DE FAMILIA

- 1 JEFE DE FAMILIA
 2 CONYUGE
 3 HIJO(A)
 4 NIETO(A)
 5 PADRE O MADRE
 6 HERMANO(A)
 7 SOBRINO(A)
 8 FAMILIAR POLITICO
 9 OTROS
 0 NINGUNO

RAMA DE ACTIVIDAD

- 0 NO PROCEDE
 1 AGRICULTURA
 2 GANADERIA
 3 SILVICULTURA
 4 PESCA
 5 ARTESANIA
 6 MANUFACTURA
 7 COMERCIO FORMAL
 8 COMERCIO INFORMAL
 9 CONSTRUCCION
 10 SERVICIOS
 11 OTROS

ESTADO FISIOLOGICO

- 1 EMBARAZO
 2 LACTANCIA
 3 DISCAPACIDAD
 SEXO
 M = MASCULINO
 F = FEMENINO

ESCOLARIDAD

- 0 NO PROCEDE
 1 ANADEFABETO
 2 SABE LEER Y ESCRIBIR
 3 CURSA PRIMARIA O PRIMARIA INCOMPLETA
 4 PRIMARIA COMPLETA
 5 SECUNDARIA COMPLETA
 6 BACHILLERATO O EQUIVALENTE
 7 CARRERA TECNICA
 8 ESTUDIOS PROFESIONALES

CONDICION LABORAL

- 0 NO PROCEDE
 1 ASALARIADO
 2 NO ASALARIADO
 3 PROPIETARIO NO EMPLEADOR
 4 PROPIETARIO EMPLEADOR
 5 AMA DE CASA
 6 ESTUDIANTE
 7 DESEMPLEADO

IDIOMA

- 1 INDIGENA
 2 ESPANOL
 3 BILINGUE



7.- PREESCOLARES

¿CUANTOS NIÑOS MENORES DE 5 AÑOS HAY EN ESTA FAMILIA?

(SI HAY MÁS DE 3 NIÑOS ANOTE LA INFORMACIÓN REQUERIDA EN ESTE MÓDULO, SOLO DE LOS 3 MÁS PEQUEÑOS)

¿QUIÉN ES LA MADRE DEL NIÑO?

ANOTAR LA CLAVE DE LA MADRE ASIGNADA EN EL MÓDULO 4 (COMPOSICIÓN FAMILIAR); SI LA MADRE DEL NIÑO NO VIVE CON EL ANOTAR '99'

NIÑO 1: _____ NIÑO 2: _____ NIÑO 3: _____

NIÑO 1

NOMBRE DEL NIÑO

APELLIDO PATERNO

FECHA DE NACIMIENTO

DA: _____ MES: _____ AÑO: _____

PESO

_____ K.G.

TALLA

_____ CMS

SEXO

M = MASCULINO
F = FEMELEINO

LACTANCIA MATERNA

¿FUE ALIMENTADO REGULARMENTE CON LECHE EN BIBERÓN LOS PRIMEROS 12 MESES?

SI NO

¿DURANTE CUANTOS MESES?

_____ MESES

SI LA RESPUESTA FUE SI ¿A QUE EDAD INICIÓ?

_____ MESES

SI ACTUALMENTE AUN TOMA, ANOTAR '99'

ABLACTACION

¿A QUE EDAD RECIBO REGULARMENTE OTROS ALIMENTOS DISTINTOS A LA LECHE O LIQUIDOS?

_____ MESES

SI NO HA SIDO AUN ABLACTADO, ANOTAR '99'

ENFERMEDADES RESPIRATORIAS EN LOS ÚLTIMOS 12 DIAS?

SI NO

SI LA RESPUESTA FUE SI ¿DE QUE SE TRATA?

1 - OJARRA 2 - INFECCION RESPIRATORIA 3 - OTRA

NIÑO 3

NOMBRE DEL NIÑO

APELLIDO PATERNO

FECHA DE NACIMIENTO

DA: _____ MES: _____ AÑO: _____

PESO

_____ K.G.

TALLA

_____ CMS

SEXO

M = MASCULINO
F = FEMELEINO

LACTANCIA MATERNA

¿FUE ALIMENTADO REGULARMENTE CON LECHE EN BIBERÓN LOS PRIMEROS 12 MESES?

SI NO

¿DURANTE CUANTOS MESES?

_____ MESES

SI LA RESPUESTA FUE SI ¿A QUE EDAD INICIÓ?

_____ MESES

SI ACTUALMENTE AUN TOMA, ANOTAR '99'

ABLACTACION

¿A QUE EDAD RECIBO REGULARMENTE OTROS ALIMENTOS DISTINTOS A LA LECHE O LIQUIDOS?

_____ MESES

SI NO HA SIDO AUN ABLACTADO, ANOTAR '99'

ENFERMEDADES RESPIRATORIAS EN LOS ÚLTIMOS 12 DIAS?

SI NO

SI LA RESPUESTA FUE SI ¿DE QUE SE TRATA?

1 - OJARRA 2 - INFECCION RESPIRATORIA 3 - OTRA

NIÑO 2

NOMBRE DEL NIÑO

APELLIDO PATERNO

FECHA DE NACIMIENTO

DA: _____ MES: _____ AÑO: _____

PESO

_____ K.G.

TALLA

_____ CMS

SEXO

M = MASCULINO
F = FEMELEINO

LACTANCIA MATERNA

¿FUE ALIMENTADO REGULARMENTE CON LECHE EN BIBERÓN LOS PRIMEROS 12 MESES?

SI NO

¿DURANTE CUANTOS MESES?

_____ MESES

SI LA RESPUESTA FUE SI ¿A QUE EDAD INICIÓ?

_____ MESES

SI ACTUALMENTE AUN TOMA, ANOTAR '99'

ABLACTACION

¿A QUE EDAD RECIBO REGULARMENTE OTROS ALIMENTOS DISTINTOS A LA LECHE O LIQUIDOS?

_____ MESES

SI NO HA SIDO AUN ABLACTADO, ANOTAR '99'

ENFERMEDADES RESPIRATORIAS EN LOS ÚLTIMOS 12 DIAS?

SI NO

SI LA RESPUESTA FUE SI ¿DE QUE SE TRATA?

1 - OJARRA 2 - INFECCION RESPIRATORIA 3 - OTRA

BIBLIOGRAFÍA

- [1] ÁVILA, A., SHAMAH, T., CHÁVEZ, A. *Encuesta Nacional de Alimentación y Nutrición en el Medio Rural 1996. Instituto Nacional de la Nutrición Salvador Zubirán.* Documento Interno del INNSZ, 1997.
- [2] BARRAGÁN, L., AMBROCIO, R., ÁVILA, A., ÁVILA, M.C. *Cartografía. Encuesta Nacional de Alimentación el Medio Rural 1996.* Documento Interno del INNSZ, 1997.
- [3] COX, D.R., WERMUTH, NANNY. *Multivariate Dependencies. Models, analysis and interpretation.* Monographs on Statistics and Applied Probability. Chapman & Hall, 1996.
- [4] DRAPER, N.R., SMITH, H. *Applied Regression Analysis.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., 1966.
- [5] ESLAVA GÓMEZ, G., PALMER ARRACHE, C., MÉNDEZ RAMÍREZ, I., ÁVILA, A., SHAMAH, T. Magnitud y distribución geográfica de la desnutrición infantil en el medio rural de México. Manuscrito sometido.
- [6] KRZANOWSKI, W.J. , MARRIOT, F.C.H. *Multivariate Analysis. Part 1. Distributions, Ordination and Inference.* Kendall's Library of Statistics 1, 1994.
- [7] KRZANOWSKI, W.J. , MARRIOT, F.C.H. *Multivariate Analysis. Part 2. Clasification, Covariance Structures and Repeated Measurements.* Kendall's Library of Statistics 2, 1995.
- [8] MCCULLAGH, P., NELDER, J.A. *Generalized Linear Models.* Monographs on Statistics and Applied Probability. Chapman & Hall, 1983.

- [9] MONTGOMERY, D.C., PECK, E.A. *Introduction to linear regression analysis. Second edition.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., 1992.
- [10] ORGANIZACIÓN MUNDIAL DE LA SALUD. *Medición del Cambio del Estado Nutricional.* Ginebra, Suiza, 1983.
- [11] PALMER ARRACHE, CATALINA. *El Cálculo de Varianza en Muestras Complejos. Aplicación en la Encuesta de Nutrición 1996.* Tesis de maestría, IIMAS, UNAM, 1999.
- [12] PELTO, GRETEL H. , URGELLO, JOCELYN, ALLEN LINDSAY H., CHÁVEZ, ADOLFO, MARTÍNEZ, HOMERO, MENESES, LUZMARIA, CAPACCIONE, CONSTANCE, BACKSTRAND, JEFFERY. Household size, food intake and anthropometric status of school-age children in a highland mexican area. *Soc Set Med, 1994; 33,19; 1135-1140.*
- [13] PFEFFERMANN, D. The use of sample weights for survey data analysis. *Statistical Methods in Medical Research, 1996; 5; 239-261.*
- [14] PFEFFERMANN, D., SMITH T.M.F. Regression Models for Grouped Populations in Cross-Section Surveys. *International Statistical Review, 1985; 54,1; 37-59.*
- [15] PFEFFERMANN, D., SMITH T.M.F. The role of samplin weights when modeling surevey data. *International Statistical Review, 1993; 61,2; 317-337.*
- [16] PUBLIC HEALTH SERVICE, HEALTH RESOURCES ADMINISTATION. NCHS growth charts. *Rockville, MD, 1976; HRA 76; 1120, 25, 3.*
- [17] RIVERA, JUAN A., HABICHT, JEAN-PIERRE, ROBSON, DOUGLAS S. Effect of supplementary feeding on recovery from mild to maoderate wasting in prescholl children. *Am J Clin Nutr, 1991; 54; 62-68.*
- [18] RIVERA-DOMMARCO, JUAN , GONZÁLEZ-COSSÍO, TERESA, FLORES, MARIO, HERNÁNDEZ-AVILA, MAURICIO, LEZANA , MIGUEL ANGEL, SEPÚLVEDA-AMOR, JAIME. Déficit de talla y emaciación en menores de

- cinco años en distintas regiones de México. *Salud Pública Mex*, 1995; 37,2; 95-107.
- [19] RODRÍGUEZ, GERMÁN, GOLDMAN, NOREEN. An assesment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society. A* , 1995; 158, Part 1, 73-89.
- [20] SEBER, G.A.F. *Multivariate Observations*. John Wiley & Sons, 1984.
- [21] SKINNER, C.J., HOLT, D., SMITH, T.M.F. *Analysis of complex surveys*. John Wiley & Sons, 1989.
- [22] SUDGEN, R.A., SMITH, T.M.F. Ignorable and informative design in survey sampling inference. *Biometrika*, 1984; 71, 3; 495-506.
- [23] WEISBERG SANFORD. *Applied linear regression. Second Edition*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1985.

ÍNDICE DE MATERIAS

- R^2 , coeficiente de determinación múltiple, 42
- índice
- de bienes, 61
 - de condición socioeconómica, 61
 - del menor, 61
 - para la madre, 61
- índice de hacinamiento, 18
- índice de vivienda, 60
- ENAL 96, 3
- actividad del jefe de hogar, 19
- coeficientes de regresión, 41
- componentes principales, 59
- consistencia en diseño, 85
- edad de ablactación, 18
- efectos del diseño, 81
- estratos, 11
- factores de expansión, 81
- familia, 11
- gasto per capita, 18
- indicadores del estado nutricional, 12
- información parcial, 84
- intervalo intergenésico, 16, 19
- métodos de selección de variables, 44
- medidas antropométricas, 12
- modelos de regresión lineal, 40
- PEDZ, 13
- PEDZ2, 17
- PEDZ3, 17
- Peso para la edad, 12
- Peso para la talla, 12
- PETZ, 13
- ponderadores, 81
- regresiones polinomiales, 76
- residuales, 41
- residuales studentizados, 43
- score z, 13
- talla para la edad, 12
- TEDZ, 13
- unidad primaria de muestreo, 11
- variable dependiente o respuesta, 39
- variables de diseño, 82

variables independientes o regresores, 39