

**UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA**



Universidad de El Salvador

Hacia la libertad por la cultura

**MÉTODOS ROBUSTOS APLICADOS A LA
CLASIFICACIÓN DEL ESTADO NUTRICIONAL
DE LA NIÑEZ SALVADOREÑA (FESAL 2008).**

**TRABAJO DE GRADUACIÓN PRESENTADO POR:
CARLOS ERNESTO CANIZALES RIVERA**

**PARA OPTAR AL GRADO DE:
MASTER EN ESTADÍSTICA**

**ASESORES:
DR. ADOLFO HERNÁNDEZ ESTRADA.
UNIVERSIDAD COMPLUTENSE DE MADRID
DR. JOSÉ NERYS FUNES TORRES.
UNIVERSIDAD DE EL SALVADOR**

CIUDAD UNIVERSITARIA, SAN SALVADOR, AGOSTO 2012.

UNIVERSIDAD DE EL SALVADOR

RECTOR : ING. MARIO ROBERTO NIETO LOVO

SECRETARIO

GENERAL : DRA. ANA LETICIA ZA VALETA DE AMAYA

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

DECANO : MSC. MARTÍN ENRIQUE GUERRA CÁCERES

SECRETARIO : LIC. NELSON EDILTRUDYS GÓMEZ CEDILLOS

ESCUELA DE MATEMÁTICA

DIRECTOR : DR. JOSÉ NERYS FUNES TORRES

SECRETARIA : ALBA IDALIA CÓRDOVA CUÉLLAR

CIUDAD UNIVERSITARIA, SAN SALVADOR, AGOSTO 2012.

ÍNDICE

INTRODUCCIÓN.....	7
PLANTEAMIENTO DEL PROBLEMA	9
JUSTIFICACIÓN.....	11
OBJETIVOS.....	13
HIPÓTESIS	14
METODOLOGÍA	15
CAPÍTULO I FUNDAMENTO TEÓRICO.....	17
1.1 INTRODUCCIÓN.....	18
1.2 DESNUTRICIÓN	18
1.3 COMPARACIÓN DE POBLACIONES	23
1.3.1 EL CONTRASTE DE KOLMOGOROV-SMIRNOV	24
1.3.2 EL CONTRASTE DE SHAPIRO-WILKS.....	25
1.3.3 EL CONTRASTE DE WILCONXON-MANN-WHITNEY	26
1.3.4 EL CONTRASTE DE KRUSKAL - WALLIS	30
1.3.5 EL CONTRASTE DE LA MEDIANA	32
1.4 TÉCNICAS DE CLASIFICACIÓN	34
1.4.1 ANÁLISIS DISCRIMINANTE.....	34
1.4.2 REGRESIÓN LOGÍSTICA	37
CAPÍTULO II INFORME DESCRIPTIVO.....	41
2.1 INTRODUCCIÓN.....	42
2.2 COMPOSICIÓN DE LA BASE DE DATOS	42
2.2.1 DISEÑO MUESTRAL.....	43
2.2.2 UNIDAD MUESTRAL.....	44
2.3 VARIABLES CONSIDERADAS.....	45
2.3.1 VARIABLES MEDIDAS EN EL NIÑO.....	46

2.3.2	VARIABLES MEDIDAS EN LA MADRE.	47
2.3.3	VARIABLES MEDIDAS EN AMBOS (MADRE Y NIÑO).....	50
2.3.4	VARIABLES ANTROPOMÉTRICAS.....	51
2.4	DESCRIPCIÓN DEMOGRÁFICA	53
2.5	ESTADO NUTRICIONAL DE LOS NIÑOS.....	63
2.6	ESTADO NUTRICIONAL DE LOS NIÑOS EN FUNCIÓN DE LAS VARIABLES DEMOGRÁFICAS.	67
CAPÍTULO III REGRESIÓN LOGÍSTICA Y ANÁLISIS DISCRIMINANTE EN LA DETERMINACIÓN DEL ESTADO NUTRICIONAL DE LOS NIÑOS.		
3.1	INTRODUCCIÓN.....	78
3.2	DIVISIÓN DE LA MUESTRA.	78
3.3	APLICACIÓN DE LA REGRESIÓN LOGÍSTICA.	80
3.4.1	VARIABLES SIGNIFICATIVAS.....	80
3.4.2	ESTIMACIÓN DEL MODELO LOGÍSTICO.	83
3.4.3	VALIDACIÓN DEL MODELO LOGÍSTICO.....	89
3.4.4	INTERPRETACIÓN DEL MODELO LOGÍSTICO.	101
3.4.5	APLICACIÓN ILUSTRATIVA DEL MODELO LOGÍSTICO.	103
3.4	APLICACIÓN DEL ANÁLISIS DISCRIMINANTE.....	104
3.4.1	ESTIMACIÓN FUNCIÓN LINEAL DISCRIMINANTE.....	105
3.4.2	ESTIMACIÓN VALIDACIÓN CRUZADA.....	112
3.4.3	ESTIMACIÓN FUNCIÓN CUADRÁTICA DISCRIMINANTE.....	114
3.4.4	VALIDACIÓN ANÁLISIS DISCRIMINANTE.	117
3.4.5	APLICACIÓN ILUSTRATIVA DEL ANÁLISIS DISCRIMINANTE.	119
3.5	COMPARACIÓN DE LOS MÉTODOS CONSIDERADOS.....	120
3.6	ESTIMACIÓN DE LOS MODELOS USANDO TODOS LOS DATOS.....	121
3.6.1	MODELO DE REGRESIÓN LOGÍSTICO CON TODOS LOS DATOS 121	
3.6.2	ANÁLISIS DISCRIMINANTE CON TODOS LOS DATOS	126
CAPÍTULO IV CONTRASTE DE HIPÓTESIS.		
		133

4.1	INTRODUCCIÓN.....	134
4.2	COMPARACIÓN DEL ESTADO NUTRICIONAL SEGÚN SECTOR DE RESIDENCIA.....	134
4.1	CONSTRASTE DE NORMALIDAD.....	135
4.2	IGUALDAD DE DISTRIBUCIONES.....	135
4.3	PRUEBA DE WILCONXON-MAN-WHITNEY.....	136
4.3	COMPARACIÓN DEL ESTADO NUTRICIONAL SEGÚN NIVEL EDUCATIVO DE LA MADRE.....	137
4.3.1	CONSTRASTE DE NORMALIDAD.....	137
4.3.2	PRUEBA DE KRUSKAL-WALLIS.....	139
4.3.3	PRUEBA DE LA MEDIANA.....	139
4.3.4	COMPARACIÓN DOS A DOS.....	140
4.4	COMPARACIÓN DEL ESTADO NUTRICIONAL SEGÚN EDAD DE LA MADRE.....	142
4.4.1	CONSTRASTE DE NORMALIDAD.....	142
4.4.2	PRUEBA DE KRUSKAL-WALLIS.....	143
4.4.3	PRUEBA DE LA MEDIANA.....	144
4.4.4	COMPARACIÓN DOS A DOS.....	145
4.5	COMPARACIÓN DEL ESTADO NUTRICIONAL SEGÚN ESTADOS CIVILES FAMILIARES.....	147
4.5.1	CONSTRASTE DE NORMALIDAD.....	147
4.5.2	PRUEBA DE KRUSKAL-WALLIS.....	148
4.5.3	PRUEBA DE LA MEDIANA.....	149
4.5.4	COMPARACIÓN DOS A DOS.....	150
4.6	COMPARACIÓN DEL ESTADO NUTRICIONAL SEGÚN DESEO DE CONCEBIRLO.....	152
4.6.1	CONSTRASTE DE NORMALIDAD.....	152
4.6.2	PRUEBA DE KRUSKAL-WALLIS.....	153
4.6.3	PRUEBA DE LA MEDIANA.....	154

4.6.4	COMPARACIÓN DOS A DOS.....	154
4.7	COMPARACIÓN DEL ESTADO NUTRICIONAL SEGÚN NIVEL SOCIOECONÓMICO.....	156
4.7.1	CONSTRASTE DE NORMALIDAD.....	156
4.7.2	PRUEBA DE KRUSKAL-WALLIS.	157
4.7.3	PRUEBA DE LA MEDIANA.....	158
4.7.4	COMPARACIÓN DOS A DOS.....	159
	CAPÍTULO V CONCLUSIONES	160
5.1	CONCLUSIONES REGRESIÓN LOGÍSTICA	161
5.2	CONCLUSIONES ANÁLISIS DISCRIMINANTE	162
5.3	CONCLUSIONES REGRESIÓN LOGÍSTICA Y ANÁLISIS DISCRIMINANTE	163
	ANEXOS	165
	BIBLIOGRAFÍA.	177

INTRODUCCIÓN

En la presente investigación se realiza una aplicación de los modelos de regresión logística y del análisis discriminante con el fin de determinar el conjunto de variables más importantes en la estimación del estado nutricional de la niñez salvadoreña. Dichas técnicas, se utilizan en la base de datos de la Encuesta de Salud Familiar (FESAL) realizada en el año 2008 por la Asociación Demográfica Salvadoreña.

Es de mencionar que el objetivo principal de la investigación es encontrar un modelo estadístico con el cual se pueda predecir, con alta probabilidad, el estado nutricional de la niñez salvadoreña en base a información disponible a priori por las autoridades competentes del Ministerio de Salud; y que en base a dicha información y al modelo estimado, las autoridades puedan identificar a niños que sufren de desnutrición o que se encuentran en un alto riesgo de padecerla (pues la frontera entre los niños que padecen y los que están en un alto riesgo de padecer la desnutrición no está bien definida) y ejecutar programas en ellos que ayuden a minimizar el impacto de la desnutrición.

El documento se encuentra dividido en cuatro capítulos. En el primer capítulo, se presenta el fundamento teórico de la investigación, es decir, una breve y concisa exposición de las principales herramientas estadísticas necesarias en la regresión logística, el análisis discriminante, y en el contraste de hipótesis que se plantean en la investigación. En el segundo se presenta la descripción de la base de datos, las variables consideradas en el estudio; y un breve informe descriptivo de las variables, el capítulo finaliza con un breve informe sobre la relación de las variables con el estado nutricional de los niños.

En el tercero se presenta la aplicación del análisis discriminante y de la regresión logística; se muestra la estimación de los parámetros y la validación de los modelos encontrados. Mientras que en el cuarto capítulo se presentan los contrastes de las hipótesis planteadas en esta investigación. Se ha decidido incorporar un mini capítulo en el cual se presentan las conclusiones de los principales resultados obtenidos.

Es de mencionar que para poder realizar el análisis que se presenta en este documento, se ha utilizado el paquete estadístico SPSS 18 (PASWS Statistics 18).

PLANTEAMIENTO DEL PROBLEMA

Desde hace mucho tiempo que conocemos de la importancia de una buena alimentación y de una dieta balanceada y rica en vitaminas y minerales en los primeros años de vida de cualquier ser humano. Por una parte, una correcta alimentación no solo constituye el buen desarrollo físico y mental de los niños, sino que, además, le ayuda a crear defensas en su sistema inmunológico que pueden salvarle la vida ante las enfermedades de las cuales estará expuesto en toda su vida.

La desnutrición, es hoy en día uno de los principales problemas a los cuales nuestra sociedad debe encontrar una forma rápida y eficiente de darle solución; si bien es cierto que no todo el mundo padece de desnutrición, pero si es cierto que en los lugares en los cuales se padece la sufren en distintas maneras y en algunos lugares inclusive trae consigo la muerte; siendo las principales personas que la padecen niños en general.

Según datos de la Organización Mundial de la Salud, la desnutrición constituye uno de los problemas más devastadores que aquejan a los pobres del mundo, que está asociado a una serie de enfermedades causantes de la mortalidad muy elevada especialmente en niños muy pequeños. Basta con darnos cuenta de la cruel verdad que sufrieron miles de personas en el Norte de África a finales del año pasado.

Pero no nos alejamos demasiado; aún en nuestro país vive gente la cual se va a la cama sin hacer los tres tiempos necesarios de comida, algunos son más privilegiados que otros en el sentido de que si hacen los tres tiempos de comida, sin embargo, la comida que ingieren no tiene los requerimientos nutricionales

necesarios para una dieta balanceada. Según datos del informe de “Índice de Desarrollo Humano 2010” de la O.N.U., cerca del 10 % de la población de nuestro país para el año 2006 sufría de algún nivel de desnutrición, mientras que un 11% de insuficiencia en las necesidades mínimas de energía alimentaria. Según ese mismo informe, para el año 2008 la tasa de mortalidad para niños menores de 5 años (por cada mil habitantes) fue de 18.

Bajo este contexto, se plantea el siguiente problema, dada la base de datos de la Encuesta de Salud Familiar FESAL realizada en el año 2008 por la Asociación Demográfica Salvadoreña y que contiene variables sobre: Planificación familiar, Salud materna, Actividad sexual y reproductiva en mujeres de 15 a 49 años de edad, Salud en la niñez e Indicadores de nutrición materno infantil, todas ellas medidas en 4624 niños menores de cinco años en El Salvador; determinar en base a dicha información, los modelos de regresión logística y técnicas de análisis discriminante, qué variables influyen en el estado nutricional de la niñez salvadoreña.

JUSTIFICACIÓN

El Salvador es uno de los países de Latinoamérica que más sufre de violencia, los hechos de violencia están a la orden del día; pero entre otros de los muchos problemas que sufre se encuentra la desnutrición. La desnutrición es una de las principales causas de muerte para las personas en sus primeros años de vida, debido a que una mala alimentación implica que el organismo o mejor dicho el sistema inmunológico no está preparado para combatir ciertas enfermedades que se sufre en los primeros años vida y que pueden llevar consigo a la muerte.

En nuestro país existen muchas organizaciones sin fines de lucro que han destinado buena cantidad de dinero a ayudar a niños que sufren de desnutrición severa y de enfermedades relacionadas con ésta. El Ministerio de Educación, por ejemplo, cuenta con un programa de alimentación para niños y niñas ubicados principalmente en sectores rurales del país, sin embargo, a pesar de esto el tipo de alimentos no cuenta con los requerimientos nutricionales necesarios de una dieta balanceada, y en el peor de los casos solo lo reciben aquellos que pueden asistir a un Centro Escolar a una edad en la que muchas de las enfermedades mortales ya han pasado. En su mayoría los programas son para calmar el sufrimiento de las personas que los padecen, pero existen muy pocos programas orientados a disminuir dicho fenómeno.

Sin embargo, cabe entonces preguntarnos, ¿Qué sectores de nuestra sociedad son los más afectados?, ¿Cómo hacer para identificar a niños que podría sufrir o padecer de desnutrición?, ¿Saben las madres y padres de familia que alimentos deben de dar a sus hijos en los primeros años de vida para evitar desnutrición?

El objetivo de esta investigación es responder a estas y muchas más preguntas que pueden surgir con relación a esta problemática. Lo que se hará para abordar este problema es utilizar técnicas estadísticas avanzadas para determinar qué variables son importantes en la determinación del estado nutricional de los niños menores de cinco años de edad; contaremos para esto con la base de datos de la Encuesta Nacional de Salud Familiar (FESAL) conducida por la Asociación Demográfica Salvadoreña en el año 2008.

OBJETIVOS

OBJETIVO GENERAL:

Determinar un modelo estadístico para estimar el estado nutricional de los niños y niñas menores de cinco años, y estimar en base al modelo, las variables que más influyen en el estado nutricional de los niños y niñas en El Salvador.

OBJETIVOS ESPECÍFICOS:

- Determinar las causas y tipos de desnutrición infantil en El Salvador.
- Decidir mediante técnicas de análisis estadístico las variables más importantes para estimar del estado nutricional de los niños menores de cinco años de edad.
- Encontrar y estimar los parámetros de un modelo estadístico que permitan la estimación del estado nutricional de los niños y niñas en El Salvador.
- Hacer comparaciones del estado nutricional de los niños y niñas en base a distintas variables de índole demográfico, social, económico y geográfico de nuestro país.
- Hacer una clasificación del estado nutricional de los niños y niñas por zonas geográficas de El Salvador.

HIPÓTESIS

- El estado nutricional de los niños y niñas menores de cinco años de edad no es el mismo para los diferentes niveles educativos de sus madres.
- El comportamiento del estado nutricional de los niños y niñas menores de cinco años de edad es el mismo para los que viven en zonas rurales y urbanas del país.
- El comportamiento del estado nutricional de los niños y niñas menores de cinco años de edad no es el mismo para las madres jóvenes que las madres ya maduras (Edad en la cual quedó embarazada).
- El comportamiento del estado nutricional de los niños y niñas menores de cinco años de edad no es el mismo en los distintos estados civiles de sus padres de familia.
- El comportamiento del estado nutricional de los niños y niñas menores de cinco años de edad no es el mismo para las madres que deseaban el embarazo que de las madres que no lo deseaban.
- El comportamiento del estado nutricional de los niños y niñas menores de cinco años de edad no es el mismo en los distintos niveles socioeconómico de la familia.

METODOLOGÍA

Para la realización de este trabajo se seguirán las siguientes fases: Limpieza y preparación de la base de datos, Selección de las variables, Estimación de parámetros y Diagnóstico, Validación y Análisis de resultados.

- En Limpieza y preparación de la base de datos; se asegura que la base de datos sea consistente, se crearán y codificarán las variables necesarias e importantes para llevar a cabo el análisis. Posteriormente se dividirá la muestra en dos partes; una primera parte con aproximada el 75% de los datos destinados para seleccionar las variables estadísticamente relevantes y para estimar los parámetros del modelo, el restante 25% de los datos, se destinará para evaluar el ajuste del modelo encontrado a los datos. La separación de ambas muestras se hará mediante muestreo aleatorio simple.
- En Selección de variables; mediante pruebas de independencia y diagramas de dispersión se determinará el conjunto óptimo de variables que estén lo más fuertemente relacionados con la variable dependiente que mide el estado nutricional de los niños; asimismo se elegirá dicho conjunto de tal manera que entre ellas no estén fuertemente relacionadas (modelo libre de multicolinealidad).
- En Estimación de parámetros y Diagnóstico; una vez seleccionadas las variables a formar parte del modelo, se procederá a estimar los parámetros del modelo de regresión logística y del análisis discriminante, teniendo en cuenta que el modelo encontrado cumpla con todas las hipótesis que se requieran para tener un buen ajuste. Esta parte se hará con el 75% de los datos.

- Validación y Análisis de los resultados; una vez estimados los parámetros del modelo logístico y del análisis discriminante, se procederá a validarlos, es decir, evaluar el ajuste en el 25% de los datos que no fueron utilizados para estimar los parámetros, y observar qué tan próximo es el modelo para describir los datos. Finalmente se procederá a interpretar y comentar los modelos encontrados.

CAPÍTULO I

FUNDAMENTO

TEÓRICO

1.1 INTRODUCCIÓN

El objetivo de este capítulo es presentar una breve introducción de las técnicas estadísticas que serán utilizadas en el trabajo de investigación. Asimismo, se presentan algunas definiciones que se consideran importantes sobre la desnutrición en niños menores de cinco años de edad.

El capítulo inicia con una introducción sobre desnutrición: se presenta su significado, los diferentes tipos de desnutrición, sus efectos en los niños, y las diferentes medidas antropométricas y sus criterios para medir la desnutrición en niños menores de cinco años de edad. Posteriormente, se presentan herramientas estadísticas de comparación de poblaciones, las cuales serán necesarias para el contraste de las hipótesis planteadas en esta investigación, se presentan pruebas no paramétricas, debido a que la hipótesis de normalidad falla para las variables en estudio (en el capítulo 4 se dan mayores detalles).

Se finaliza con una exposición de las técnicas de clasificación en Estadística, tales como: Regresión Logística y Análisis Discriminante; se presenta para cada una de ellas su fundamento, la forma de estimar los parámetros asociados, y algunos comentarios que deben tenerse en cuenta a la hora de utilizarlos.

1.2 DESNUTRICIÓN

En este apartado exponemos de manera breve la teoría más relevante sobre desnutrición. Iniciamos con la definición de desnutrición, y que presentamos a continuación.

SIGNIFICADO DE DESNUTRICIÓN:

- Significa que el cuerpo de una persona no está obteniendo los nutrientes suficientes. Esta condición puede resultar del consumo de una dieta

inadecuada o mal balanceada, bien sea por trastornos digestivos, problemas de absorción.

- Es la enfermedad provocada por el insuficiente aporte de combustibles (hidratos de carbono, grasas y proteínas).

TIPOS DE DESNUTRICIÓN:

Se puede hablar de varios tipos de desnutrición como desnutrición primaria (deficiencia de micronutrientes, deficiencia de macronutrientes) y desnutrición secundaria.

Desnutrición primaria: cuando los aportes de nutrientes no pueden ser aportados por la situación económica, cultural y/o educativa. Dentro de este tipo se muestra otra clasificación, que tiene que ver con la deficiencia de micronutrientes, se pueden clasificar de la siguiente manera:

- **Deficiencia de Macronutrientes:** La forma más frecuente de desnutrición, denominada marasmo, ocurre cuando el individuo no puede ingerir cantidades suficientes de alimentos en general. El primer tipo de desnutrición es el marasmo son de delgadez extrema y apenas tienen masa muscular en su cuerpo. Suelen parecer ancianos, encogidos y arrugados con costillas y articulaciones prominentes. El segundo tipo de desnutrición es el Kwashiorkor, provocado por una falta de energía proteica y de algunos macronutrientes en la dieta; entre sus principales síntomas pueden mencionarse los siguientes: abdomen hinchado, anemia severa y endema (hinchazón de pies, piernas y brazos que representan el 30% del peso corporal del niño). Ambos tipos de desnutrición reducen la resistencia a las

infecciones y pueden acabar matando a quienes la padecen (por sí solos alteran los niveles de sales y minerales del cuerpo).

- **Deficiencia de Micronutrientes:** La ingesta insuficientes de vitaminas y minerales provoca otras formas de desnutrición. Los micronutrientes son necesarios en pequeñas cantidades para asegurar un funcionamiento metabólico correcto.

Desnutrición secundaria: si los aportes nutricionales son adecuados, pero debido a otras enfermedades, la absorción o utilización de los alimentos no es adecuada.

CAUSAS DE DESNUTRICIÓN:

Las causas de la desnutrición se pueden dar por diferentes factores, a saber:

Factores Medioambientales:

- La información disponible permite sostener que aproximadamente la mitad de los problemas nutricionales ocurren en hogares de zonas rurales localizados en ambientes muy expuestos a riesgos ambientales.
- Las cifras más altas de desnutrición (y sus consecuencias) se observan en países donde la agricultura a menudo es afectada por desastres naturales.

Factores Sociales y económicos:

- La desnutrición se asocia principalmente con extrema pobreza. Entre los diversos aspectos relacionados con la pobreza que inciden en la desnutrición, destacan los siguientes:
- El bajo nivel de ingresos limita el acceso a los alimentos en calidad y/o cantidad.

- El bajo nivel parental, en especial el de la madre, inciden negativamente en la desnutrición de los niños.

Factores Biológicos:

- Un deficiente estado nutricional materno, como consecuencia de una mala nutrición previa, aumenta los riesgos de desnutrición intrauterina y bajo peso al nacer.
- La ausencia o insuficiencia de lactancia materna (primeros seis meses de vida) expone al niño a ingerir alimentos que no satisfacen los requerimientos nutricionales a esa etapa de desarrollo.

EFFECTOS DE LA DESNUTRICIÓN EN LOS NIÑOS:

Los signos físicos, psicológicos y otros cambios que produce la desnutrición son:

- Déficit de peso y la estatura que se espera para la edad.
- Atrofia muscular (se observa un desarrollo inadecuado en los músculos).
- Piel seca, áspera y descamándose.
- Retardo en la pubertad.
- La frecuencia cardíaca está acelerada y son frecuentes las continuas infecciones respiratorias.
- Los niños, paradójicamente presentan anorexia, crecimiento del hígado y alteración en el ritmo de las disposiciones fecales.
- Raquitismo, osteoporosis, debilidad muscular, anemia por falta de hierro.
- La falta de vitamina C conduce a úlceras en la córnea y puede llevar a la ceguera.
- En algunos casos la muerte.

MEDICIÓN DE LA DESNUTRICIÓN:

La antropometría es la técnica más usada en la evaluación del estado nutricional, ya que proporciona información fundamental acerca de la suficiencia del aporte de macronutrientes. Las mediciones más usadas son en función del peso y la talla. La Organización Mundial de la Salud recomienda el uso de curvas de crecimiento elaboradas por el National Center for Health Statistics (NHCS), ya que los pesos y las tallas de niños provenientes de grupos socio-económicos alto y medio de países subdesarrollados son similares a los de niños de países desarrollados con antecedentes comparables.

El peso como parámetro aislado no tiene validez y debe expresarse en función de la edad o de la talla. La relación peso/edad (P/E) es un buen indicador durante el primer año de vida, pero no permite diferenciar a niños constitucionales pequeños. Su uso como parámetro único no es recomendable, se acepta como normal una variación de más o menos el 10% con respecto al peso esperado para la edad. O valores que estén ubicados entre + 1 desviaciones estándar y - desviaciones estándar. Entre - 1 y - 2 desviaciones estándar debe considerarse en riesgo de desnutrir y un peso bajo 2 desviaciones estándar es sugerente de desnutrición. El P/E no debe usarse como parámetro de evaluación nutricional especialmente en mayores de 2 años.

La talla también debe expresarse en función de la edad (T/E). El crecimiento lineal continuo es el mejor indicador de dieta adecuada y de estado nutricional a largo plazo. Es importante considerar que es un parámetro muy susceptible a errores de medición, y que por lo tanto, debe ser repetida. Se acepta como normal una talla entre el 95% y el 105% del estándar para la edad, lo que en las curvas del NCHS corresponde aproximadamente a valores entre percentil 10 y 90 para la edad, al igual que para la relación anterior se considera un niño desnutrido o en riesgo de

desnutrir si su talla se encuentra bajo dos desviaciones estándar de la talla que el correspondería para su edad.

El índice de peso para la talla (P/T) es un buen indicador de estado nutricional actual y no requiere un conocimiento preciso de la edad. Es útil para el diagnóstico, tanto de desnutrición como de sobrepeso y obesidad. Su uso como único parámetro de evaluación puede no diagnosticar como desnutridos a algunos niños que efectivamente lo son (algunos casos de retraso global de crecimiento como por ejemplo). Se considera normales los valores que se ubiquen entre los percentiles 10 y 90, con excepción del primer trimestre de vida en que es deseable que los valores se ubiquen entre los percentiles 25 y 75. Los valores bajo el percentil 10 son indicativos de desnutrición y sobre el percentil 90, indican sobrepeso; al igual que en las relaciones anteriores se considera a un niño desnutrido o en riesgo si su peso para su edad se encuentra bajo dos desviaciones estándar de la que realmente le corresponde.

En la práctica se recomienda el uso combinado de los tres índices lo que permite una evaluación más precisa.

1.3 COMPARACIÓN DE POBLACIONES

A continuación se hace una breve descripción de las técnicas estadísticas que serán utilizadas en este trabajo, con el fin de facilitar la comprensión de los resultados para las personas no expertas en temas de estadística y que puedan verificar la validez de los mismos. Los dos primeros contrastes que se presentan se utilizan para contrastar la bondad de ajuste de una variable aleatoria, es decir, sirven para contrastar la hipótesis de que la variable sigue una determinada distribución de probabilidad, la distribución de contraste es en su mayoría la distribución normal. Las siguientes dos pruebas se utilizan para comparar dos poblaciones que no

tienen distribución normal (equivalente a la prueba t cuando falla la normalidad); mientras que las últimas dos se utilizan para comparar más de dos poblaciones cuando éstas no tienen distribuciones normales (equivalente al ANOVA cuando falla la normalidad).

1.3.1 EL CONTRASTE DE KOLMOGOROV-SMIRNOV

Este contraste compara la función de distribución teórica con la empírica. Es válido únicamente para variables continuas.

La hipótesis nula en este contraste es que la muestra proviene de un modelo continuo con función acumulada $F(x)$. El procedimiento para construir el contraste es:

- Ordenar los valores muestrales, de manera que: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$
- Calcular la función de distribución empírica de la muestra, $F_n(x)$ con:

$$F_n(x) = \begin{cases} 0, & \text{si } x < x_{(1)} \\ \frac{r}{n}, & \text{si } x_{(r)} < x < x_{(r+1)} \\ 1, & \text{si } x \geq x_{(n)} \end{cases}$$

- Calcular la discrepancia máxima entre la función de distribución observada (o empírica) y la teoría con el estadístico: $D_n = \max |F_n(x) - F(x)|$ cuya distribución, cuando $F(x)$ es cierta, se ha tabulado.

- Si la distancia calculada D_n es mayor que la encontrada en las tablas, fijado α , rechazamos el modelo $F(x)$.

Este contraste tiene la ventaja de que no requiere agrupar los datos y el inconveniente de que si calculamos $F(x)$ estimando parámetros de la población mediante la muestra, la distribución de D_n es sólo aproximada: el contraste es conservador, tendiendo a aceptar H_0 .

1.3.2 EL CONTRASTE DE SHAPIRO-WILKS

Este contraste mide el ajuste de la muestra representada en papel probabilístico normal a una recta. Se rechaza la normalidad cuando el ajuste es malo, que corresponde a valores pequeños del estadístico. El estadístico que se utiliza para contrastar la hipótesis de normalidad se expresa en la siguiente ecuación:

$$w = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})} \left[\sum_{j=1}^h a_{j,n} (x_{(n-j+1)} - x_{(j)}) \right]^2$$

Donde: $h = \begin{cases} \frac{n}{2}; & \text{si } n \text{ es par} \\ \frac{n-1}{2}; & \text{si } n \text{ es impar} \end{cases}$; $x_{(j)}$ es el valor ordenado en la muestra que

ocupa la posición j Y los coeficientes $a_{j,n}$ ya se encuentran tabulados.

Existen tablas para la distribución de w y se rechaza la normalidad cuando el valor calculado es menor que el valor crítico dado en las tablas. La razón es que w mide el ajuste a la recta y no la discrepancia con la hipótesis nula.

1.3.3 EL CONTRASTE DE WILCONXON-MANN-WHITNEY

Las siguientes dos pruebas que se presentan a continuación se utilizan cuando se desean comparar dos poblaciones, dichas poblaciones deben ser independientes entre sí; las dos pruebas se utilizan cuando la distribución de la variable de interés (la cual deseamos comparar en ambas poblaciones) no se comporta de manera normal o aproximadamente normal en cada una de las dos poblaciones, es decir, las dos pruebas son equivalentes a la prueba t de Student cuando falta la normalidad.

1.3.3.1 CONTRASTE DE MANN-WHITNEY

Los datos consisten de dos muestras aleatorias. Sean X_1, X_2, \dots, X_n denotando la muestra aleatoria de tamaño n de la primera población, y sea Y_1, Y_2, \dots, Y_m la muestra aleatoria correspondiente a la segunda población. Se trata entonces de contrastar que ambas muestras provienen de la misma población. Unimos las dos muestras en una sola para formar una muestra única, y ordenamos las observaciones de menor a mayor.

SUPOSICIONES CONCERNIENTES A LA PRUEBA:

- Ambas muestras son muestras aleatorias de sus respectivas poblaciones.
- Las muestras son independientes entre sí.
- Ambas muestras provienen de variables continuas.

LAS HIPÓTESIS QUE SE CONTRASTAN EN DICHA PRUEBA SON:

Sean $F(x)$ y $G(x)$ las funciones de distribución correspondientes a la población 1 y 2, respectivamente, de este modo las hipótesis pueden enunciarse de la siguiente manera:

- $H_0: F(x) = G(x)$ para todo valor de x .
- $H_1: F(x) \neq G(x)$ para algún valor de x .

El test se basa en contar, en todos los posibles $n \times m$ pares de valores $(X_i; Y_j)$, en cuántos se verifica que $Y_j < X_i$. Una forma equivalente y mucho más eficiente es considerar

$$D_{ij} = \begin{cases} 1 & Y_j < X_i \\ 0 & Y_j > X_i \end{cases}$$

EL ESTADÍSTICO DE CONTRASTE ES;

$$U = \sum_{i=1}^n \sum_{j=1}^m D_{ij}$$

Es intuitivo que valores muy grandes o muy pequeños de este estadístico, (los más extremos son $U = nm$ y $U = 0$), conducirán al rechazo de la hipótesis nula.

REGLA DE DECISIÓN:

Los cuantiles u_p de la distribución U ya han sido tabulados para diferentes tamaños de muestras para $p = 0.001, 0.005, 0.01, 0.025, 0.05$ y 0.1 , los cuantiles superiores no están dados pero pueden calcularse de la siguiente manera: $u_{1-p} = nm - u_p$. De este modo la regla consiste en:

- Rechazar H_0 si $U < u_{\frac{\alpha}{2}}$ o si $U > u_{1-\frac{\alpha}{2}}$.
- En caso contrario se acepta H_0 .

En caso de que los tamaños de muestras no se encuentren tabulados pueden obtenerse los valores críticos de la siguiente manera; $u_\alpha = \frac{nm}{2} + Z_\alpha \sqrt{\frac{nm(n+m+1)}{12}}$; donde Z_α el cuantil de la distribución normal estándar.

1.3.3.2 CONTRASTE DE WILCONXON

Los datos consisten de dos muestras aleatorias. Sean X_1, X_2, \dots, X_n denotando la muestra aleatoria de tamaño n de la primera población, y sea Y_1, Y_2, \dots, Y_m la muestra aleatoria correspondiente a la segunda población. Se trata entonces de contrastar que ambas muestras provienen de la misma población. Unimos las dos muestras en una sola para formar una muestra única, y ordenamos las observaciones de menor a mayor.

SUPOSICIONES CONCERNIENTES A LA PRUEBA:

- Ambas muestras son muestras aleatorias de sus respectivas poblaciones.
- Las muestras son independientes entre sí.
- Ambas muestras provienen de variables continuas.

LAS HIPÓTESIS QUE SE CONTRASTAN EN DICHA PRUEBA SON:

Sean $F(x)$ y $G(x)$ las funciones de distribución correspondientes a la población 1 y 2, respectivamente, de este modo las hipótesis pueden enunciarse de la siguiente manera:

- $H_0: F(x) = G(x)$ para todo valor de x .
- $H_1: F(x) \neq G(x)$ para algún valor de x .

Llamaremos rango de un dato al orden que ocupa en esta ordenación; a la observación más pequeña se le asigna rango 1 (más pequeña en ambas muestras), a la siguiente más pequeña rango 2, y así sucesivamente hasta que finalmente a la observación más grande se le asigna rango $n + m$ (más grande en ambas muestras).

En el caso de que existan datos con los mismos valores (empates), asignamos a cada una de ellas el rango promedio que le correspondería en caso de que no existieran empates.

EL ESTADÍSTICO DE PRUEBA ES:

Definimos S , como la suma de rangos correspondientes a todas las observaciones de la población 1, es decir, $W = \sum_{i=1}^n R(X_i)$; donde $R(X_i)$ denota el rango asignado a X_i

REGLA DE DECISIÓN:

Los cuantiles w_p de la distribución W ya han sido tabulados para diferentes tamaños de muestras, de este modo la regla consiste en:

- Rechazar H_0 si $W < w_{\frac{\alpha}{2}}$ o si $W > w_{1-\frac{\alpha}{2}}$.
- En caso contrario se acepta H_0 .

El estadístico de Wilcoxon contabiliza, para cada valor de la primera muestra (muestra X), cuántos valores (de cualquiera de las dos muestras) son menores o iguales que él y luego suma estas cantidades recorriendo todos los datos en la muestra X . La única diferencia entre el estadístico de Wilcoxon y el de Mann-Whitney radica en que W contabiliza los valores de la propia muestra X que

preceden a cada dato de la muestra X , mientras que el estadístico U sólo tiene en cuenta, a estos efectos, los datos de la muestra Y .

Como consecuencia ambos estadísticos son equivalentes ya que se puede obtener el valor de uno conocido el del otro, mediante la siguiente relación:

$$W = U + \frac{n(n+1)}{2}$$

1.3.4 EL CONTRASTE DE KRUSKAL - WALLIS

Los datos consisten de k muestras aleatorias de diferentes tamaños. Denotamos a la i -ésima muestra de tamaño n_i por $X_{i1}, X_{i2}, \dots, X_{in_i}$. Entonces podemos ordenar los datos en k columnas de la siguiente manera:

Muestra 1	Muestra 2	...	Muestra k
$X_{(1,1)}$	$X_{(2,1)}$		$X_{(k,1)}$
$X_{(1,2)}$	$X_{(2,2)}$		$X_{(k,2)}$
$X_{(1,n_1)}$	$X_{(2,n_2)}$		$X_{(k,n_k)}$

Note que $X_{(j,n_i)}$ corresponde a la observación que ocupa la posición n_i en la j -ésima muestra ordenada (de menor a mayor), es decir, la n_i -ésima observación más pequeña en la muestra.

Decimos entonces que N es el número total de observaciones $N = \sum_{i=1}^k n_i$.

Asignamos rango 1 a la observación más pequeña de total de observaciones (las k muestras combinadas), rango 2 a la segunda observación más pequeña, y así hasta que asignamos rango N a la observación más grande del total de observaciones.

Sea R_i la suma de los rangos asignados a la muestra i -ésima, es decir,

$$R_i = \sum_{j=1}^{n_i} R(X_{ij}), \quad i = 1, 2, \dots, k$$

SUPOSICIONES CONCERNIENTES A LA PRUEBA:

- Cada muestra es una muestra aleatoria.
- Las muestras son independientes entre sí.
- La característica de interés es continua, es decir, cada X_{ij} es una variable continua.

LAS HIPÓTESIS QUE SE CONTRASTAN EN DICHA PRUEBA SON:

- H_0 : Todas las k poblaciones tienen función de distribución idéntica.
- H_1 : Al menos una de las poblaciones tiene una función de distribución distinta.

EL ESTADÍSTICO DE PRUEBA ES;

$$T = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{[R_i - n_i(N+1)/2]^2}{n_i}$$

Una forma equivalente del estadístico es la siguiente;

$$T = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

REGLA DE DECISIÓN:

Cuando se estudian $k = 3$ poblaciones y todos los tamaños muestrales son mayores que 5, o cuando el número de poblaciones es mayor que 3, el estadístico de prueba T sigue una distribución Chi-Cuadrado con $(k-1)$ grados de libertad. De este modo la regla consiste en:

- Rechazar H_0 sí $T > \chi_{1-\frac{\alpha}{2}}$.
- En caso contrario se acepta H_0 .

1.3.5 EL CONTRASTE DE LA MEDIANA

La Prueba de la Mediana ha sido diseñada para examinar o contrastar si varias muestras provienen de una población con la misma mediana. La Prueba de la Mediana es simplemente una aplicación especial de una Prueba Chi-Cuadrado (con totales marginales fijos). Para contrastar si varias (k) poblaciones tienen la misma mediana, se selecciona una muestra aleatoria de cada una de las (k) poblaciones. Se construye una tabla de contingencia de dimensión $2 \times k$, y en la i -ésima columna aparecen el número de observaciones de la población i -ésima que se encuentran por encima y por debajo de la gran mediana (la mediana obtenida al

combinar las k poblaciones). A dicha tabla de contingencia así obtenida se le aplica posteriormente la prueba Chi-Cuadrado.

Muestra	1	2	...	k	
$<$ Gran mediana	O_{11}	O_{12}	...	O_{1k}	a
\geq Gran mediana	O_{21}	O_{22}	...	O_{2k}	b
Totales	n_1	n_1	...	n_1	N

SUPOSICIONES CONCERNIENTES A LA PRUEBA:

- Cada muestra es una muestra aleatoria.
- Las muestras son independientes entre sí.
- Si todas las poblaciones tienen la misma mediana, entonces todas las poblaciones tienen la misma probabilidad p de que una observación cualquiera en su correspondiente muestra exceda el valor de la gran mediana.

LAS HIPÓTESIS QUE SE CONTRASTAN EN DICHA PRUEBA SON:

- H_0 : Todas las k poblaciones tienen la misma mediana.
- H_1 : Al menos dos de las poblaciones tienen diferente mediana.

EL ESTADÍSTICO DE PRUEBA ES;

$$T = \frac{N^2}{ab} \sum_{i=1}^k \frac{\left(O_{1i} - \frac{n_i a}{N}\right)^2}{n_i} = \frac{N^2}{ab} \sum_{i=1}^k \frac{O_{1i}^2}{n_i} - \frac{Na}{b}$$

En el caso de que el valor de a sea aproximadamente igual al valor de b , la ecuación anterior puede simplificarse a una expresión mucho más sencilla de cálculo:

$$T = \sum_{i=1}^k \frac{(O_{1i} - O_{2i})^2}{n_i}$$

REGLA DE DECISIÓN:

La región crítica de tamaño aproximadamente igual a α corresponde a valores de T mayores que el cuantil $(1-\alpha)$ de la variable aleatoria Chi-cuadrado con $(k-1)$ grados de libertad, denotado por $\chi_{1-\alpha}$. De este modo si $T > \chi_{1-\alpha}$ se rechazará la hipótesis nula de igualdad de medianas.

1.4 TÉCNICAS DE CLASIFICACIÓN

Las siguientes dos técnicas que se describen a continuación son utilizadas ampliamente para clasificar a un individuo en base a ciertas características observables y a una determinada regla de decisión: la procedencia a una y solamente una de las poblaciones consideradas en el estudio. La primera se utiliza cuando la población es normal aunque puede emplearse cuando esta condición falla. La segunda es menos exigente a la distribución de las características y puede emplearse cuando éstas no son numéricas.

1.4.1 ANÁLISIS DISCRIMINANTE

Dado un conjunto de individuos donde cada uno de ellos pertenece a una de varias sub poblaciones (digamos $\Pi_1, \Pi_2, \dots, \Pi_k$), se observa para cada individuo un vector numérico multivariante $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ a partir del cual se intenta predecir a qué sub población pertenece.

DISCRIMINADORES LINEALES

Supongamos que la distribución x_i es normal multivariante con matriz de covarianzas igual a Σ en cada una de las sub poblaciones y en las cuales el vector de medias es μ_i . Se considera la distancia de Mahalanobis de x_i en cada una de las k sub poblaciones.

$$d^2(x_i, \mu_i) = (x_i - \mu_i)^T \Sigma^{-1} (x_i - \mu_i); \quad i = 1, 2, \dots, k$$

Un criterio de clasificación consiste en asignar x_i a la población más próxima:

$$\text{Si } d^2(x_i, \mu_j) = \min \{d^2(x_i, \mu_1), d^2(x_i, \mu_2), \dots, d^2(x_i, \mu_k)\}, \text{ asignamos a } x_i \text{ a } \Pi_j.$$

PROBABILIDADES A POSTERIORI

El grado de confianza al clasificar una observación depende de la probabilidad de acertar. La probabilidad a posteriori se interpreta de la siguiente manera: no es más que la probabilidad de asignar a x_i a la población más probable.

$$P(\Pi_j / x_i) = \frac{f_j(x_i)}{\sum_{j=1}^k f_j(x_i)}$$

Cuando la observación tenga la misma probabilidad de pertenecer a cada sub población y donde $f_j(x_i)$ es la función de distribución normal multivariante en la sub población Π_j .

DISCRIMINADORES CUADRÁTICOS

Si admitiendo la normalidad multivariante de las observaciones la hipótesis de igualdad de matriz de covarianzas no fuese admisible, el procedimiento de

resolver el problema es clasificar la observación en el grupo con máxima probabilidad a posteriori. Esto también equivale a clasificar la observación $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ en el grupo donde se minimice la función:

$$\frac{1}{2} \log |\Sigma_j| + \frac{1}{2} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \text{ para cada } i = 1, 2, \dots, k$$

Donde nuevamente se asume que la probabilidad de pertenecer a cada una de las sub poblaciones es la misma.

DISCRIMINACIÓN EN POBLACIONES NO NORMALES

Para el caso en que no se tenga confianza plenamente de la normalidad de las poblaciones el análisis discriminante aún puede utilizarse y el criterio de clasificación se basa únicamente en las probabilidades a posteriori, es decir, asignar x_i a la población más probable. La ecuación:

$$P(\Pi_j / x_i) = \frac{f_j(x_i)}{\sum_{j=1}^k f_j(x_i)}$$

Sigue siendo válida solamente que en esta ocasión $f_j(x_i)$ representa la función de densidad (que puede ser cualquiera) en la sub población Π_j .

PROBABILIDADES DE CLASIFICACIÓN ERRÓNEA

Al igual que todo modelo estadístico, el Análisis Discriminante no está exento de errores; el error ocurre cuando una observación es clasificada en una sub población a la cual claramente no pertenece. El error de clasificación errónea se define

entonces como el cociente de las observaciones mal clasificadas entre el total de observaciones que han sido clasificadas correctamente.

COMENTARIOS DEL ANÁLISIS DISCRIMINANTE

Hay que tener en cuenta que el número de parámetros a estimar en el caso cuadrático es mucho mayor al del caso lineal. Este gran número de parámetros hace que incluso en muestras grandes, la discriminación cuadrática sea inestable y, aunque las matrices de covarianzas sean muy diferentes se obtienen con frecuencia mejores resultados con la función lineal que con la cuadrática. Si recomienda entonces utilizar ambas reglas utilizando la validación cruzada, y en caso de que las diferencias sean pequeñas, tomar la función lineal.

1.4.2 REGRESIÓN LOGÍSTICA

Se considera el caso en que la variable de respuesta, en un problema de regresión, sólo asume dos valores posibles: 0 y 1; esos números son por lo regular asignaciones arbitrarias a una respuesta cualitativa.

EL MODELO DE REGRESIÓN LOGÍSTICA

Supóngase que el modelo tiene la forma:

$$y_i = X_i' \beta + \varepsilon_i$$

En donde: $X_i' = [1, x_{i1}, x_{i2}, \dots, x_{ip}]$, $\beta' = [\beta_0, \beta_1, \dots, \beta_p]$ y la variable respuesta y_i toma los valores 0 ó 1.

Se supondrá que la variable de respuesta y_i es una variable aleatoria con distribución de Bernoulli, con la siguiente distribución de probabilidad:

y_i	Probabilidad
1	$P(y_i = 1) = \pi_i$
0	$P(y_i = 0) = 1 - \pi_i$

Se llama función de respuesta logística a la función que tiene la siguiente expresión;

$$E[y_i] = \frac{\exp(X_i' \beta)}{1 + \exp(X_i' \beta)} \quad \text{o} \quad E[y_i] = \frac{1}{1 + \exp(-X_i' \beta)}$$

Definimos $\eta = x' \beta$, que representa la porción estructurada del modelo en términos de la media de la función respuesta, o mejor dicho el predictor lineal. Una función de la $\eta = \ln\left(\frac{\pi}{1-\pi}\right)$, el cual representa la transformación logit de la probabilidad π , y la relación $\frac{\pi}{1-\pi}$ en la transformación se llama ventaja (en algunos textos a la transformación logit se llama ventaja logarítmica).

ESTIMACIÓN DE LOS PARÁMETROS

Cada observación de la muestra sigue la distribución de Bernoulli, por lo que la distribución de probabilidad de cada observación es

$$f_i(y_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i}, \quad i = 1, 2, \dots, n$$

Resulta que $E[y_i] = \frac{\exp(X_i'\beta)}{1 + \exp(X_i'\beta)} = \pi_i$; esto significa que el modelo de regresión

logística da una estimación de que la i -ésima observación tome el valor de 1.

Al utilizar el método de máxima verosimilitud para conseguir una estimación de los parámetros, se llega a que el logaritmo de la función de verosimilitud (función soporte) es:

$$\ln L(y, \beta) = \sum_{i=1}^n y_i X_i' \beta - \sum_{i=1}^n \ln(1 + \exp(X_i' \beta))$$

Para obtener los estimadores $\hat{\beta}$ se pueden utilizar varios métodos iterativos de aproximación, uno de los más utilizados es mínimos cuadrados ponderados iterativos.

El valor esperado del modelo de regresión logística se escribe con frecuencia como sigue:

$$\hat{y}_i = \hat{\pi}_i = \frac{\exp(X_i' \hat{\beta})}{1 + \exp(X_i' \hat{\beta})}$$

INTERPRETACIÓN DE LOS PARÁMETROS DEL MODELO

No resulta muy difícil interpretar los parámetros de un modelo de regresión logística. Si para la i -ésima variable aumentamos en una unidad, es decir, pasamos del valor x_i al valor $x_i + 1$, manteniendo constante a las demás variables. El predictor lineal en x_i es $\hat{\eta}(x_i) = x_i' \hat{\beta}$, mientras que en $x_i + 1$ es $\hat{\eta}(x_i + 1) = (x_i') \hat{\beta} + \hat{\beta}_i$,

donde $\hat{\beta}_i$ contiene el valor de 0 en todas las posiciones excepto en la i -ésima que contiene el valor de 1.

La diferencia entre los dos valores predichos es $\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) = \hat{\beta}_i$, puesto que $\hat{\eta}(x_i)$ sólo es el logaritmo de la ventaja cuando la i -ésima variable regresora toma el valor x_i , mientras que $\hat{\eta}(x_i + 1)$ sólo es el logaritmo de la ventaja cuando la i -ésima variable regresora toma el valor $x_i + 1$. Por consiguiente, la diferencia entre los dos valores ajustados es

$$\begin{aligned}\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) &= \ln(\text{ventaja}_{x_i+1}) - \ln(\text{ventaja}_{x_i}) \\ &= \ln\left(\frac{\text{ventaja}_{x_i+1}}{\text{ventaja}_{x_i}}\right) \\ &= \hat{\beta}_i\end{aligned}$$

Al tomar antilogaritmos se obtiene el cociente de ventaja

$$\frac{\text{ventaja}_{x_i+1}}{\text{ventaja}_{x_i}} = e^{\hat{\beta}_i}$$

El cual se puede interpretar como el aumento estimado en la probabilidad de obtener un 1 en la variable de respuesta y asociado con un cambio unitario en el valor de la variable regresora X_i , manteniendo constante el valor de las demás variables en la regresión. En otras palabras lo que esto significa es que la regresión logística nos dice que variables son las más importantes en la determinación de la ocurrencia o no de un suceso.

CAPÍTULO II

INFORME

DESCRIPTIVO

2.1 INTRODUCCIÓN.

En este capítulo se presenta un breve informe descriptivo de las principales variables consideradas en el trabajo de investigación. Primero se describen aspectos de la base de datos con la cual se está trabajando tales como: su composición, tamaño y diseño muestral. Posteriormente se presenta una lista y descripción de las variables que se consideran más importantes en el estudio.

Luego se hace una descripción demográfica de los datos, se analizan variables como: Sector de residencia, Departamento de residencia, Edad de la madre, Edad del niño, Género del niño, Educación de la madre, Estado Civil-Familiar, Nivel socioeconómico, Orden de nacimiento del niño y Situación de trabajo de la madre. Posteriormente se analiza el comportamiento de cada uno de las medidas antropométricas definidas en la sección 1.2 del capítulo I.

Finalmente se analiza el estado nutricional de los niños menores de cinco años en función de las variables demográficas consideradas.

2.2 COMPOSICIÓN DE LA BASE DE DATOS

La Asociación Demográfica Salvadoreña en el año 2008 realizó la Encuesta Nacional de Salud Familiar (FESAL); dicha encuesta se realiza cada cinco años en nuestro país y es la novena que se realiza desde el año 1973. Entre sus principales objetivos se encuentra obtener información de salud de niños y niñas menores de cinco años junto con el de sus respectivas madres en edad fértil de 15 a 49 años con énfasis en la salud reproductiva. La encuesta se realizó entre abril y noviembre del año 2008.

La encuesta consta de las siguientes 12 secciones:

- Identificación y selección de la entrevistada.
- Información general de la entrevistada.
- Historia de embarazos y nacimientos.
- Salud en la niñez.
- Servicios asociados a la salud mujer (Salud materna).
- Planificación familiar.
- Actividad sexual y reproductiva en mujeres de 15 a 24 años.
- Violencia contra la mujer.
- Infecciones de transmisión sexual.
- Asistencia escolar de los hijos de 5 a 19 años de edad.
- Vivienda e Indicadores de nutrición materno infantil.

2.2.1 DISEÑO MUESTRAL

El diseño de la muestra fue probabilístico estratificado y multietápico. La selección de la muestra fue estratificada por departamento y área de residencia urbana y rural, debido a la división administrativa con la cual trabaja el Ministerio de Salud. La selección de la muestra se realizó en tres etapas;

- La primera etapa consistió en seleccionar aleatoriamente segmentos censales de cada estrato, con una probabilidad de selección proporcional al número de viviendas existentes en cada segmento, se utilizó la base cartográfica pre censal que DIGESTYC actualizó en el año 2007; seleccionando en esta etapa un total de 617 segmentos censales.
- Previo a la segunda etapa se realizó la actualización cartográfica de cada uno de los 617 segmentos censales seleccionados en la primera etapa, se

contó en esta fase con el apoyo técnico de la Dirección General de Estadísticas y Censos (DIGESTYC) y del Centro Nacional de Registros (CNR). En esta etapa se seleccionó 35 viviendas dentro de cada segmento censal en forma aleatoria, con el fin de asegurar que todas las viviendas en cada segmento tuvieran igual probabilidad de selección. Tanto en el área urbana como rural, las viviendas deshabitadas se consideraron elegibles para visita, no así las que estuvieran en construcción o destruidas.

- En la tercera etapa se seleccionó aleatoriamente a una única mujer en edad fértil (MEF) de 15 a 49 años de edad por cada vivienda. Primero se registró la información referente al número total de personas que residen habitualmente en la vivienda, luego se listaron todas las MEF, la probabilidad de selección de cada MEF fue inversamente proporcional al número de ellas en la vivienda (se aplica como factor de ponderación en el informe de la Asociación Demográfica Salvadoreña).

2.2.2 UNIDAD MUESTRAL.

Se definió como vivienda elegible, aquella edificación que tuviera acceso independiente y, cuando eran habitadas por más de una persona, que éstas compartieran la misma alimentación, ya sea que las familias fueran nucleares o extensas. Por lo que no se enumeraron edificaciones utilizadas exclusivamente para fines productivos, comerciales u oficinas, hoteles, guarniciones militares, iglesias.

Para las mujeres entrevistadas que tuvieron al menos un niño que nació vivo a partir de 2003, se seleccionó aleatoriamente uno para administrar las preguntas sobre los temas de salud en la niñez (en el informe de la Asociación Demográfica

Salvadoreña se utiliza el respectivo factor de ponderación para compensar la probabilidad desigual junto con el factor de ponderación de la MEF).

En total se seleccionaron a 21,595 viviendas (MEF); de éstas por razones que no se pudieron controlar solamente se logró recabar completamente la información correspondiente a 12,008 (de 13,329 MEF que iniciaron la encuesta), de las restantes lastimosamente no se pudo recabar información en parte porque no se encontraba MEF habitando en la vivienda (el 21.0% del total de viviendas) y en las demás porque se negaron a brindar información (2.3%) o simplemente nadie se encontraba en el hogar a la hora de las entrevistas (se hicieron más de una visita).

De las 12,008 MEF, un total de 5,178 dieron a luz al menos a un niño desde el año 2003; sin embargo, en este informe únicamente se trabaja con 4,624 debido a que no se tiene información vital de todos ellos, debido en parte porque algunos murieron antes que se realizará la encuesta y a otros por causas que aún no comprendo no contaban con medidas antropométricas (su peso y su talla), por lo que han sido descartados del análisis.

En el presente informe se tomará a las 4,624 niños y niñas menores de cinco años (de manera equivalente las 4,624 MEF) como a unidad muestral; es decir, la población objetivo son todos aquellos niños y niñas en El Salvador que al momento de realizar la encuesta tenían una edad menor a los cinco años.

2.3 VARIABLES CONSIDERADAS.

A continuación se hace un breve descripción de las variables consideradas en este trabajo, las cuales se pretenden que ayuden a determinar el estado nutricional de los niños y niñas menores de cinco años de edad, algunas se listan solamente para el efecto de hacer comparaciones entre ellas.

2.3.1 VARIABLES MEDIDAS EN EL NIÑO.

- **Talla:** Representa la altura en centímetros del niño. Para los niños menores de 2 años se mide acostado (longitud), mientras que para los mayores se mide con los niños estando de pie (altura). Toma valores que van desde los 47 hasta los 114.4 centímetros.
- **Peso:** Constituye el peso del niño medido en kilogramos. Se pesan en básculas de preferencia con ropa ligera la cual no afecta las medidas. Toma valores que van desde los 2.7 hasta los 31.2 kilogramos (la O.M.S. presenta sus curvas de referencia en kilogramos).
- **Edad:** Es la edad medida en meses exactos en la cual se tomaron las medidas de peso y talla del niño, esto es así para poder utilizar las tablas de referencia de la Organización Mundial de la Salud.. Toma valores en el rango de 0 a 59 meses.
- **Género:** Define el género de los niños participantes en el estudio. Como ya se sabe los únicos valores que puede tomar son: Masculino o Femenino.
- **Número de línea:** Indica el orden de nacimiento del niño seleccionado. Es una variable discreta que toma valores comprendidos entre 1 (para el primer nacimiento) y 15.
- **Diarrea últimos 15 días:** Es una variable con la cual se mide si el niño seleccionado presentó diarrea durante los 15 días previos a la encuesta. Es una variable dicotómica que puede tomar únicamente dos valores: Sí o No.

- **Vitamina A:** Es una variable con la cual se pretende conocer si el niño ha tomado vitamina A en algún momento de su vida. Es una variable dicotómica que sólo toma dos valores: Sí y No.
- **Jarabe de Hierro:** Es una variable con la cual se pretende conocer si el niño ha tomado en algún momento de su vida Jarabe de Hierro. Es dicotómica al igual que la variable Vitamina A.
- **Jarabe de Cinc:** Es una variable con la que se pretende conocer si el niño ha tomado Jarabe de Cinc. Es dicotómica al igual que las dos anteriores.

2.3.2 VARIABLES MEDIDAS EN LA MADRE.

- **Estado civil familiar:** Es una variable con la cual se pretende determinar el estado civil de la familia (o de la madre). Es una variable cualitativa con las siguientes clasificaciones: Soltera, Separada (por divorcio, viudez o separada) y Acompañada (matrimonio o acompañada).
- **Religión:** Es una variable con la cual se pretende determinar o conocer la religión a la cual pertenece la madre del niño seleccionado. Es una variable cualitativa en la cual se han hecho las siguientes categorías: No tiene, Católica, Protestante (en sus distintas denominaciones), Otras (Testigos de Jehová o Mormones).
- **Edad:** Representa la edad cumplida en años de la madre. Es una variable discreta que toma valores comprendidos entre los 15 y 48 años y es la edad en la cual se considera a las mujeres en edad fértil.

- **Derechohabiente del ISSS:** Se utiliza para identificar si la madre tiene algún beneficio médico con el Instituto Salvadoreño del Seguro Social. Es una variable dicotómica con las siguientes categorías: Sí (para madres cotizantes o beneficiarias) y No.
- **Nivel académico:** Mide el último nivel de estudios que aprobó (de manera satisfactoria) la madre. Es una variable de tipo ordinal en la cual se consideran las siguientes opciones: Ninguno, Primaria/Básica, Bachillerato/Media y Educación Superior (licenciadas, doctoras, ingenieras, etc.).
- **Edad al dejar de estudiar:** Es una variable en la cual se mide la edad en años cumplidos en la que la madre abandonó sus estudios. Toma valores en el rango de 6 hasta 42 años.
- **Razón de abandono de estudio:** Es una variable con la cual se pretende conocer la causa por la cual la madre abandonó sus estudios. Las categorías consideradas en este trabajo son las siguientes: Por trabajo, Problemas familiares, Problemas económicos, Embarazo, No le gustaba.
- **Embarazo actual:** Es una variable con la cual se pretende conocer si la madre se encontraba embarazada al momento de realizar la encuesta. Los valores que puede tomar son los siguientes: Sí, No y No sabe.
- **Deseo embarazo actual:** Es una variable con la cual se pretende conocer si la madre deseaba su actual embarazo (para las que estaban embarazadas). Toma tres categorías posibles: Sí, No y No sabe/No responde.

- **Situación de trabajo:** Es una variable con la cual se pretende conocer si la madre trabaja para conseguir el dinero. Las categorías o valores que se consideran son: No trabaja, Trabaja en casa y Trabaja fuera del hogar.
- **Total hijos en casa:** Es una variable con la cual se pretende conocer la cantidad de hijos que tiene la madre viviendo en el mismo hogar del niño seleccionado. Toma valores en el rango de 1 hasta 11 niños.
- **Total hijos deseados:** Es una variable con la cual se pretende conocer la cantidad de hijos que la madre desearía tener sin considerar en cuenta a los que ya ha dado a luz. Toma valores que van desde 0 (ninguno) hasta los cuatro niños.
- **Remesas:** Es una variable con la cual se pretende conocer si la madre recibe ayuda económica por parte de algún familiar en el extranjero en concepto de remesas. Es una variable dicotómica que toma dos valores: Sí o No.
- **Peso de la mujer:** Representa el peso registrado de la madre medido en libras. Toma valores en el rango de 35.0 hasta 163.6 libras.
- **Talla de la mujer:** Representa la altura en centímetros de la madre. Toma valores que van desde los 96.9 hasta los 186.8 centímetros.
- **Anemia:** Con esta variable lo que se pretende medir es si la madre al momento de realizar la encuesta presentaba o padecía algún grado de Anemia. Los valores que puede tomar son solamente dos: Sí o No.

2.3.3 VARIABLES MEDIDAS EN AMBOS (MADRE Y NIÑO).

- **Departamento:** Es una variable con la cual se pretende conocer el departamento de residencia al momento de realizar la entrevista.
- **Sector:** Es una variable con la cual se pretende conocer el sector de residencia al momento de realizar la encuesta. Puede ser Urbano o Rural.
- **Hijo deseado:** Es una variable con la cual se pretende conocer si el hijo que ha sido seleccionado fue deseado al momento de concebirlo. Las categorías que se han considerado son: Quería entonces, Esperar más tiempo y No lo quería.
- **Servicios básicos:** Es una con la cual se mide el número de servicios básicos que dispone la familia, toma valores entre 0 y 9. Los servicios básicos son los siguientes: Electricidad, Televisor, Choro de agua (propio de la familia, no compartido), Cocina (de gas o eléctrica), Refrigerador, Inodoro, Teléfono, Vehículo propio, Vivienda con cuatro o más habitación.
- **Lactando:** Es una variable con la cual se pretende conocer si el niño ha recibido pecho en algún momento de su vida después de su nacimiento. Solo puede tomar dos valores: Sí o No.
- **Controles Posparto:** Es una variable con la cual se pretende conocer si la madre y el niño han participado de al menos un control después del nacimiento del niño. Es una variable dicotómica con las siguientes clasificaciones: Sí o No.

- **Controles Prenatal:** Es una variable con la cual se pretende conocer si la madre se estuvo participando de controles prenatales. Es una variable dicotómica con las siguientes clasificaciones: Sí o No.

2.3.4 VARIABLES ANTROPOMÉTRICAS.

- **Estado nutricional Peso para la Talla:** Es una variable que compara el peso del niño con el peso que debería corresponderle dada su estatura y su género. Si se encuentra bajo los criterios de la O.M.S. el niño no está desnutrido, en caso contrario si lo estará.
- **Estado nutricional Peso para la Edad:** Es una variable que compara el peso del niño con el peso que debería corresponderle dada su edad y su género. Si se encuentra bajo los criterios de la O.M.S. el niño no está desnutrido, en caso contrario si lo estará.
- **Estado nutricional Talla para la Edad:** Es una variable que compara la talla del niño con la talla que debería corresponderle dada su edad y su género. Si se encuentra bajo los criterios de la O.M.S. el niño no está desnutrido, en caso contrario si lo estará.
- **Estado nutricional Combinado:** Es una variable o mejor dicho un indicador del estado nutricional tomando la información combinada de las tres variables antropométricas mencionadas anteriormente, se toman como punto de partida el Peso para la Talla y luego se van combinando la información para obtener un único índice.

Para la construcción de estos índices se realizó lo siguiente:

- Se digitó, en base a los valores de referencia de la O.M.S., el peso que le corresponde a un niño según su talla y género observado; se digitó también la desviación típica (para el peso) que le corresponde dada su talla y género.
- Se estandarizó el peso del niño para la talla, es decir, se hizo la diferencia entre el peso observado del niño y el peso que le corresponde, finalmente se dividió, dicha diferencia, por la desviación típica (del peso) que le corresponde dada su talla y género.
- La variable Estado nutricional Peso para la Talla se calculó a partir de la anterior: si el peso estandarizado es menor a -2.0 el niño es considerado como desnutrido, en caso contrario es considerado un niño normal.
- Los demás índices fueron creados de forma parecida.

El Estado nutricional combinado se construyó con la siguiente regla:

- Si el Índice Peso para la Talla clasifica a un niño como desnutrido, entonces el niño es considerado como desnutrido.
- Si el Índice Peso para la Talla clasifica a un niño como normal pero no así el Índice Peso para la Edad se clasifica como desnutrido.
- Si tanto el Índice Peso para la Talla como Peso para la Edad clasifican a un niño como normal pero no así el Índice Talla para la Edad se clasifica al niño como desnutrido.

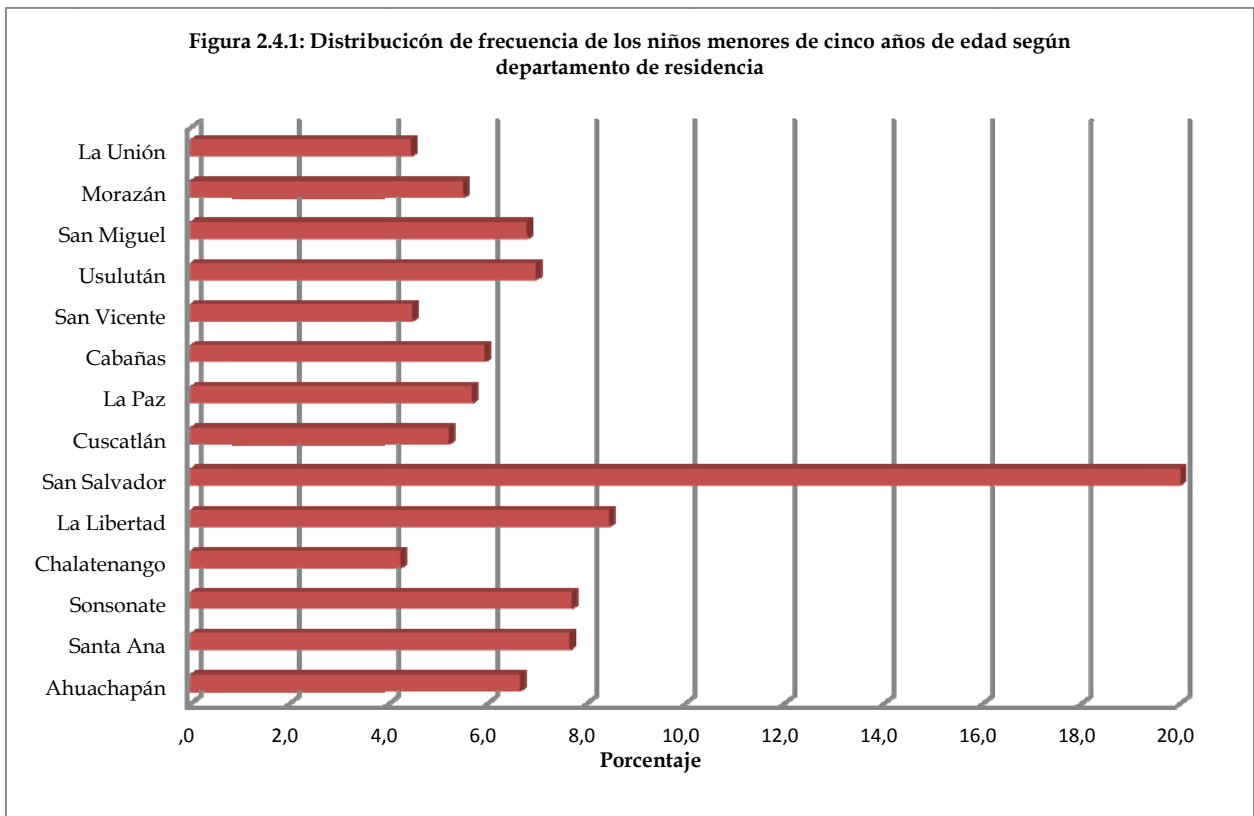
- Finalmente si los tres Índices no clasifican al niño como desnutrido entonces el niño se clasifica como normal.

2.4 DESCRIPCIÓN DEMOGRÁFICA

En esta sección se hace una breve descripción demográfica de los 4,624 niños que al momento de realizar la Encuesta Nacional de Salud (FESAL 2008) tenían una edad no mayor a los cinco años de edad, y los cuales son considerados como la población objetivo. Se toman para esto las principales variables típicas a esta clase de estudios.

DEPARTAMENTO DE RESIDENCIA.

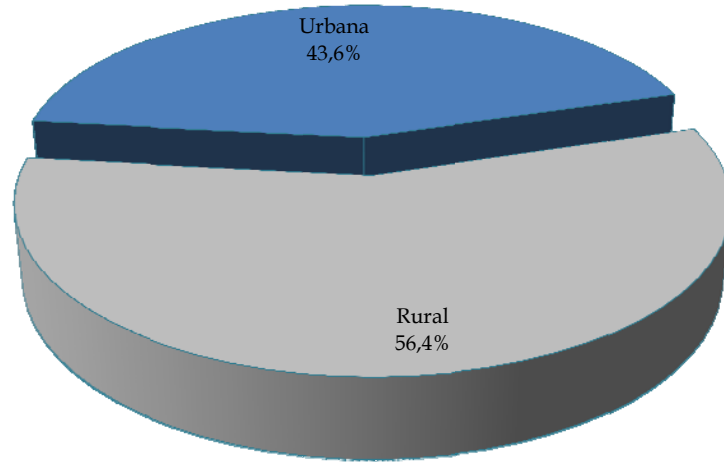
En la figura que se muestra a continuación se presenta la distribución de frecuencia de los niños menores de cinco años según el departamento de residencia. En la figura se puede apreciar que aproximadamente un 20% de los niños pertenecen o mejor dicho residen en el departamento de San Salvador, para ser más concreto son 934 niños; mientras que los restantes 13 departamentos tienen un porcentaje de niños que varía desde el 4% aproximadamente hasta un poco más del 8%, en cantidades exactas estos porcentajes representa a 196 y 356 niños, respectivamente.



SECTOR DE RESIDENCIA.

En la siguiente figura se presenta la distribución de frecuencia de los niños de acuerdo al sector de su residencia. De los 4,624 niños considerados en este trabajo 2,017 residen en zonas urbanas de nuestro país y representan el 43.6% del total. Los restantes 2,607 niños provienen de las distintas zonas rurales de nuestro país y representan el 56.4% del total de niños.

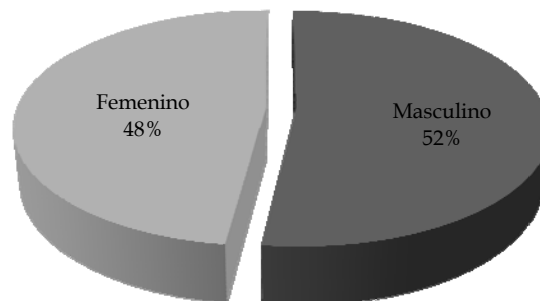
Figura 2.4.2: Distribución de frecuencia de los niños menores de cinco años según sector de residencia



GÉNERO DEL NIÑO.

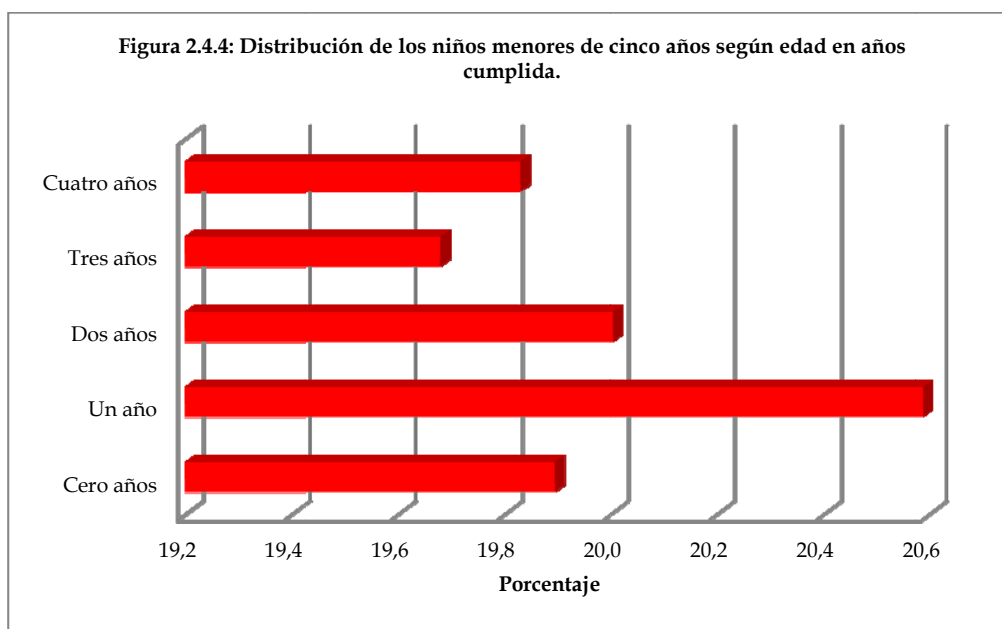
En la siguiente figura se muestra la distribución de frecuencia de los niños considerados en este trabajo según su género. En el se puede apreciar que 2,405 niños quienes representan el 52% del total son del sexo masculino; mientras que los restantes 2,219 son del sexo masculino y representan el 48%.

Figura 2.4.3: Distribución de frecuencia de los niños menores de cinco años de edad según su género.



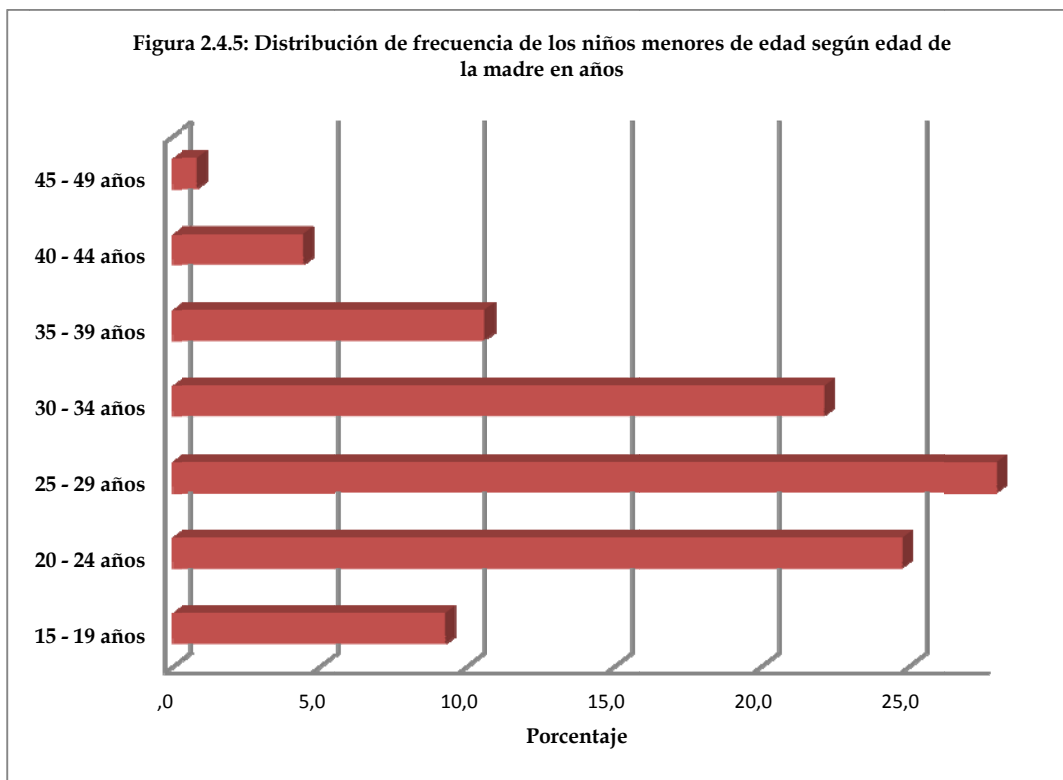
EDAD DEL NIÑO EN AÑOS CUMPLIDOS.

En la siguiente figura se muestra la distribución de frecuencia de los niños según la edad cumplida en años. En términos generales podría decirse que los niños son bastante jóvenes, pues aproximadamente un 60% de ellos tienen una edad en años cumplidos que no sobrepasa los dos años; el restante aproximadamente 40% de ellos tienen edades de tres o cuatros años cumplidos.



EDAD DE LA MADRE EN AÑOS CUMPLIDOS.

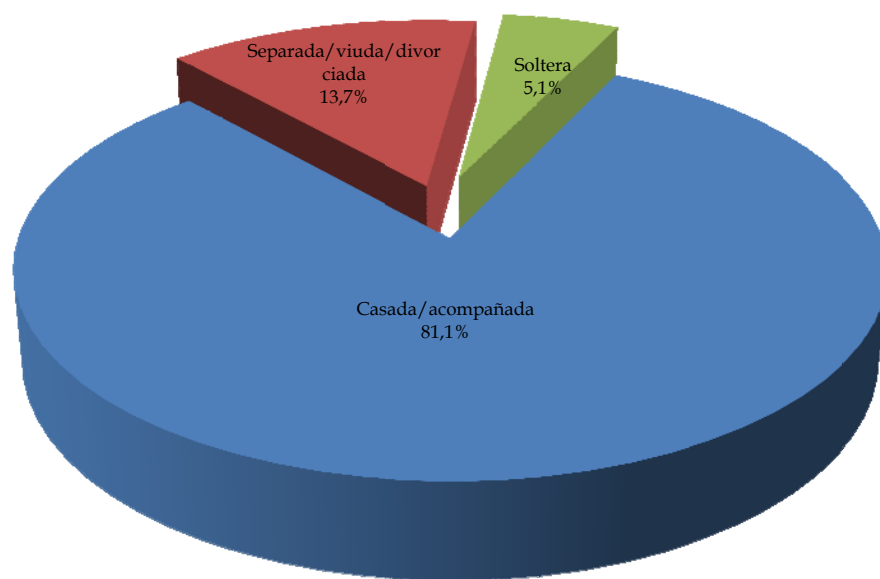
En la siguiente figura se muestra la distribución de frecuencia de los niños menores de cinco años según la edad de la madre. Los niños provienen principalmente de madres jóvenes pues aproximadamente un 34% (una tercera parte) son madres con edades menores que no sobrepasan los 24 años, este porcentaje equivale a 1,573 niños. Un 28% provienen de madres con edades comprendidas entre los 25 y 29 años de edad, los cuales son un total de 1,295 niños. Mientras que los restantes 38% provienen de madres de madres con edades entre los 30 y 45 años de edad.



ESTADO CIVIL/FAMILIAR.

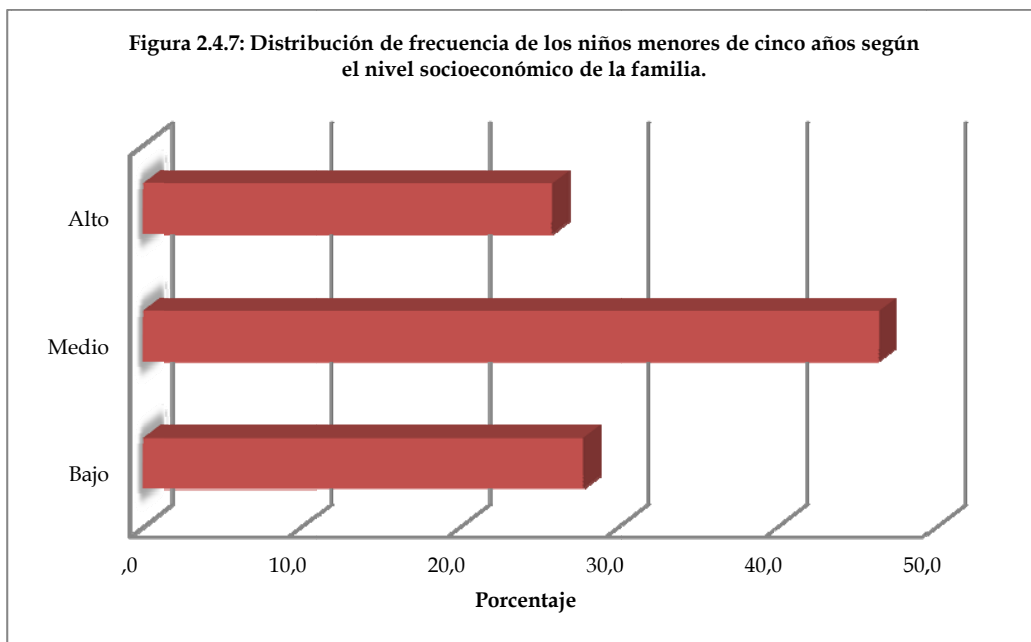
En la figura que se muestra a continuación se muestra la distribución de frecuencia de los niños y niñas menores de cinco años de edad según el estado familiar de sus familias al momento de realizar la encuesta. En su mayoría los niños provienen de hogares de los cuales su madre se encontraba viviendo con un compañero de vida (este compañero de vida no necesariamente es el padre del niño), pues 3,751 de los niños provienen de dicho hogares y representan el 81.1%; 635 niños provienen de hogares en los cuales sus padres se han separado y representa el 13.7%; los restantes 238 niños provienen de hogares en los cuales su madre es soltera y que por lo tanto se las ingenua ella misma para sobrevivir (los padres nunca reconocieron a los hijos) y representa el 5.1% del total de niños.

Figura 2.4.6: Distribución de frecuencias de los niños menores de cinco años de edad según el estado familiar de su madre



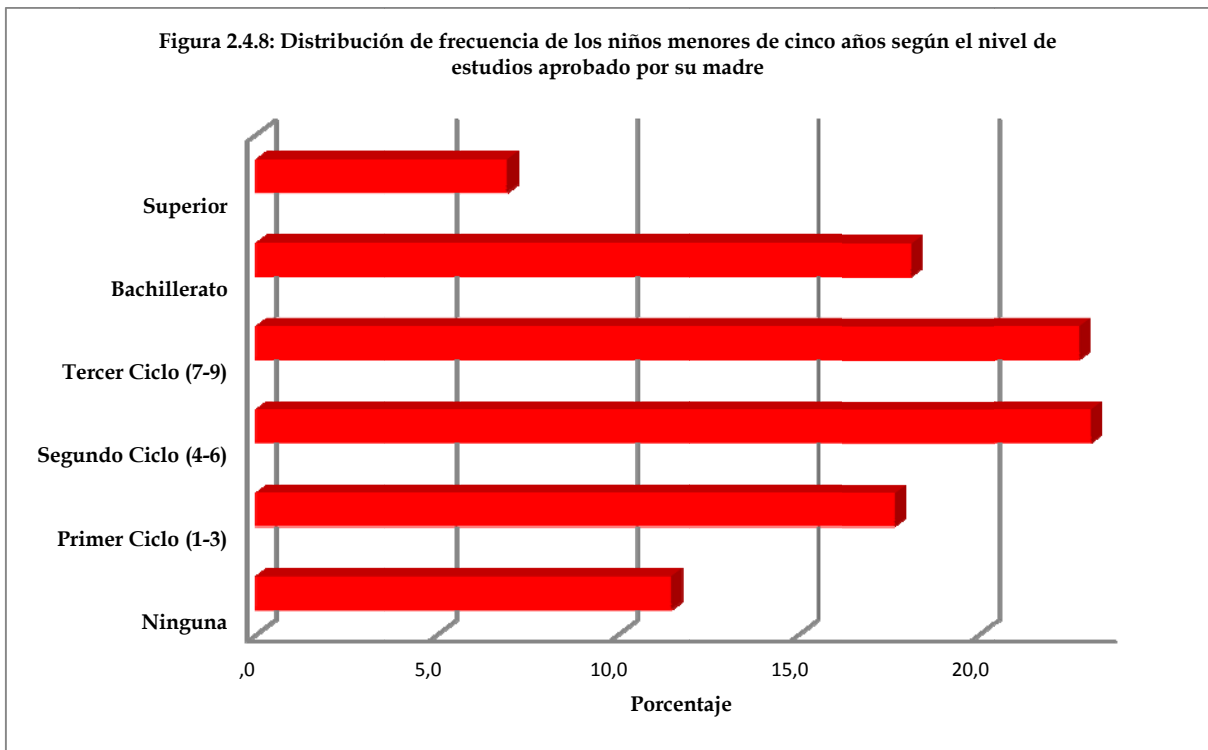
NIVEL SOCIOECONÓMICO.

En la siguiente figura se muestra la distribución de frecuencia de los niños según el nivel socioeconómico de la familia, entiéndase familia como el hogar formado por el niño y quienes residen en su mismo hogar al momento de realizar la encuesta. Aproximadamente la mitad de los niños provienen de hogares con un nivel socioeconómico medio; mientras que aproximadamente la cuarta parte de los niños provienen de hogares con nivel alto y con nivel bajo, respectivamente.



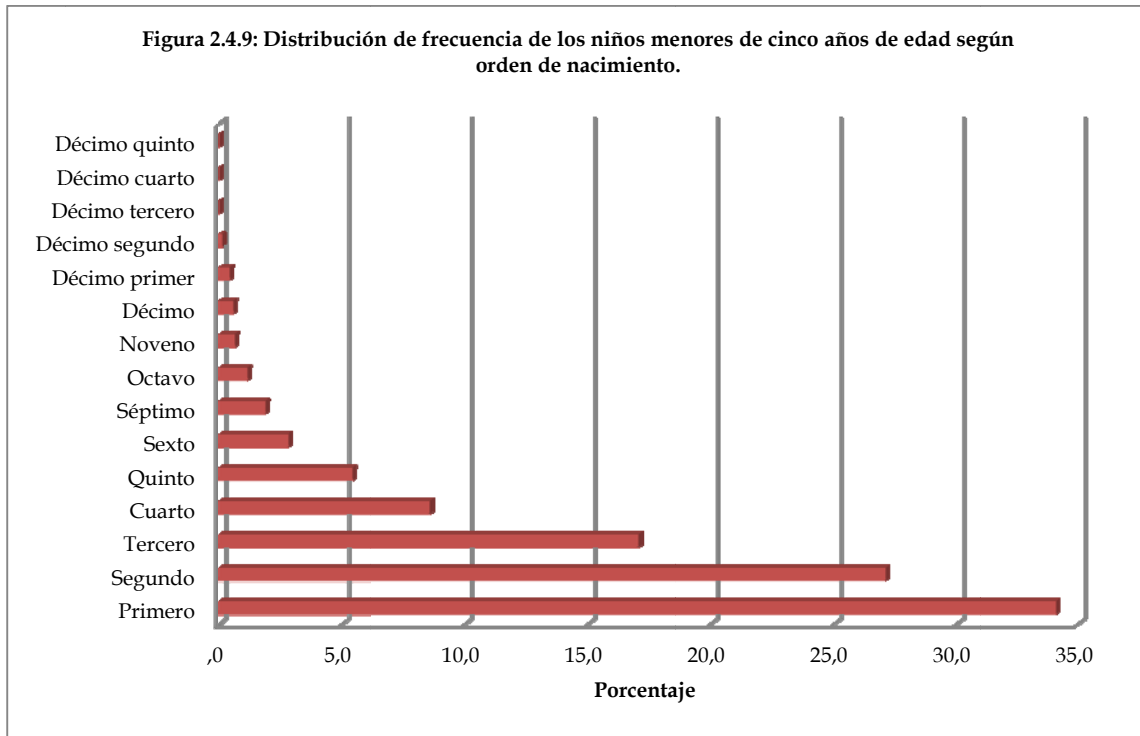
NIVEL EDUCATIVO DE LA MADRE.

En la siguiente figura se muestra la distribución de frecuencia de los niños según el nivel de estudios aprobado por su madre. Hablando en términos generales, las madres presentan un nivel de estudios bastante bajo, pues aproximadamente un 75% (3,466) de ellas obtuvo un nivel que no sobrepasa el noveno grado de educación básica. Un 18.1% (837) logró terminar el bachillerato; y apenas aproximadamente un 7% (321) logró concluir sus estudios universitarios. Es alarmante observar el poco nivel educativo que las madres tienen, pues solo entre el segundo y tercer ciclo hay aproximadamente el 45% (2,119) de las madres.



NÚMERO DE LÍNEA.

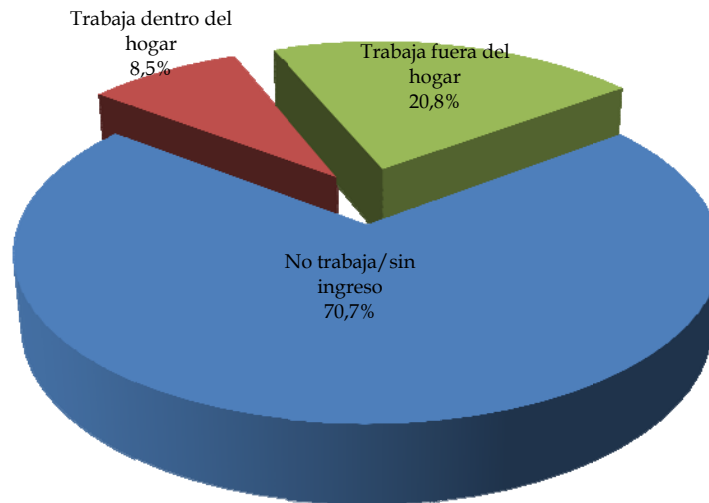
En el siguiente gráfico se muestra la distribución de frecuencia de los niños según su orden de nacimiento (el orden de nacimiento incluye a los hijos previos que nacieron y por diferentes razones ya habían muerto antes de llevar a cabo la encuesta). Se observa que 1,577 son primogénitos y representan a poco más de la tercera parte de los niños (un 34.1%). Un total de 2,043 de los niños tienen un orden de nacimiento del segundo o tercero y representan casi la mitad de la muestra (un 44% aproximadamente). Los niños cuyo orden de nacimiento son del noveno en adelante no representan cada uno ni el 1%.



SITUACIÓN DE TRABAJO.

A continuación se muestra la distribución de frecuencia de los niños según la situación de trabajo de la madre. Al momento de realizar la encuesta 3,269 niños cuyas madres no tienen ninguna fuente de financiamiento y representa respectivamente el 70.7% del total. Los restantes 1,355 trabajan ya sea en sus hogares (probablemente tengan alguna tienda o simplemente su trabajo no requiere la necesidad de salir del hogar) o fuera de éste, con 394 y 961 niños quienes representan el 8.55% y 20.8% respectivamente.

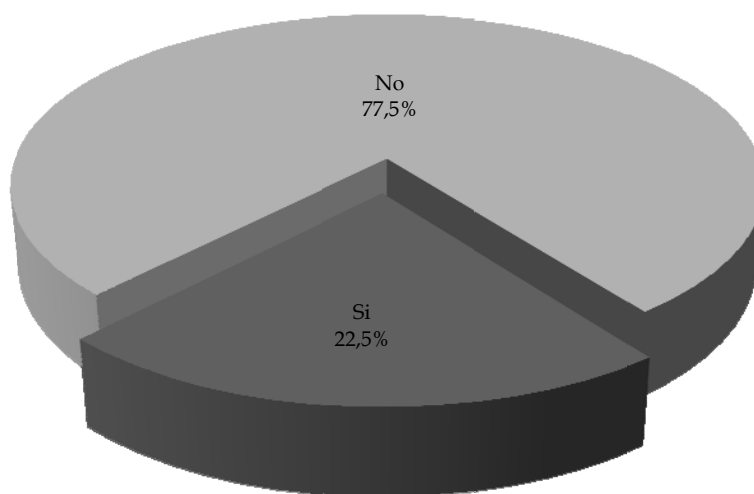
Figura 2.4.10: Distribución de frecuencia de los niños menores de cinco años de edad según situación de trabajo de la madre.



BENEFICIARIA DEL ISSS.

En el siguiente gráfico se muestra la distribución de frecuencia de los niños de acuerdo a si su madre es Beneficiaria del Instituto Salvadoreño del Seguro Social (ISSS). Para un total de 3,585 niños sus madres no tienen ningún beneficio del ISSS y representa el 77.5%; mientras que para los restantes 1,039 niños sus madres sí lo tienen y representan 22.5%. Hay que mencionar que de éstos 1,039 solamente 406 son hijos de madres cotizantes (pagan directamente por recibir el servicio médico) y 633 son beneficiarias (reciben el servicio porque alguna persona en su mayoría su compañera de vida paga dichos servicios médicos).

Figura 2.4.11: Distribución de frecuencia de los niños menores de cinco años según beneficio de la madre del ISSS

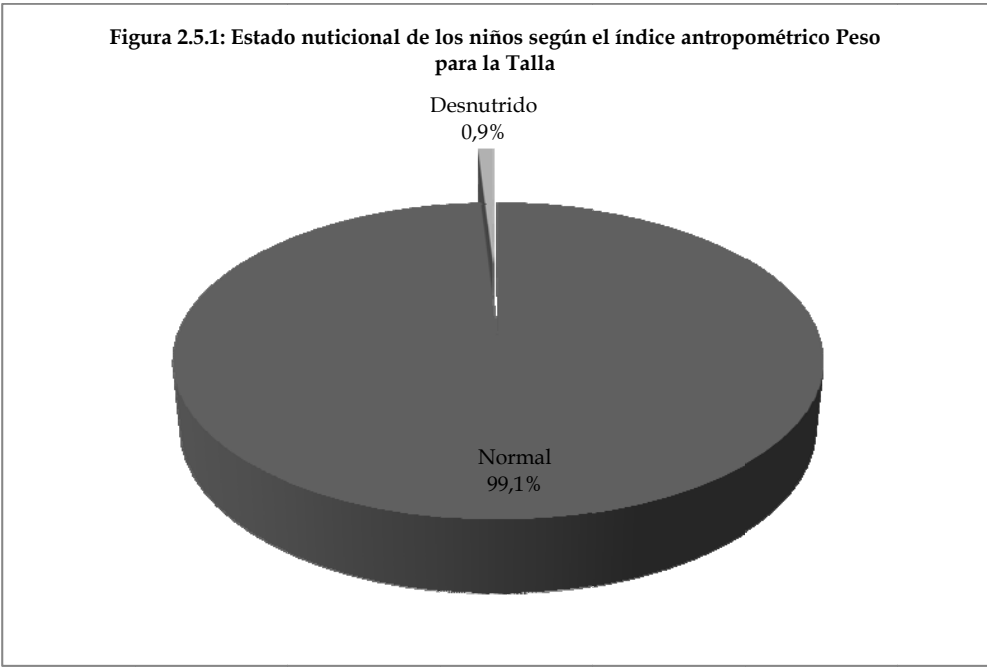


2.5 ESTADO NUTRICIONAL DE LOS NIÑOS.

A continuación se muestra el estado nutricional de los niños según los distintos índices antropométricos considerados; es de mencionar que solamente se distingue a niños desnutridos y no desnutridos sin hacer distinción entre los distintos tipos de desnutrición.

ÍNDICE PESO PARA LA TALLA.

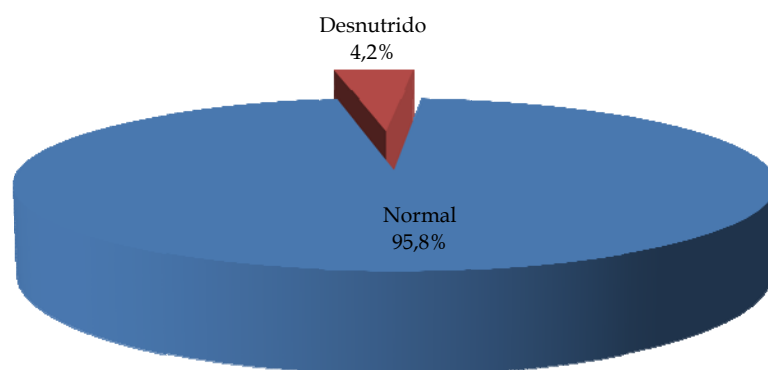
En el siguiente gráfico se muestra la distribución de frecuencia del estado nutricional de los niños utilizando el índice antropométrico Peso para la Talla. En la figura se aprecia que según este índice únicamente el 0.9% de los niños son considerados como desnutridos, dicho porcentaje corresponde a 43 niños. Mientras que los restantes 4,581 niños se consideran normales o, mejor dicho, no desnutridos y representan el 99.1% del total de niños.



ÍNDICE PESO PARA LA EDAD.

En el siguiente gráfico se muestra la distribución de frecuencia del estado nutricional de los niños utilizando el índice antropométrico Peso para la Edad. En la figura se aprecia que según este índice el 4.2% de los niños son considerados como desnutridos, dicho porcentaje corresponde a 193 niños. Mientras que los restantes 4,431 niños se consideran normales y representan el 95.8% del total de niños. Sin embargo, es de tener en cuenta que aproximadamente el 40% de los niños son mayores de dos años y por consiguiente este indicador podría no detectar a posibles niños desnutridos.

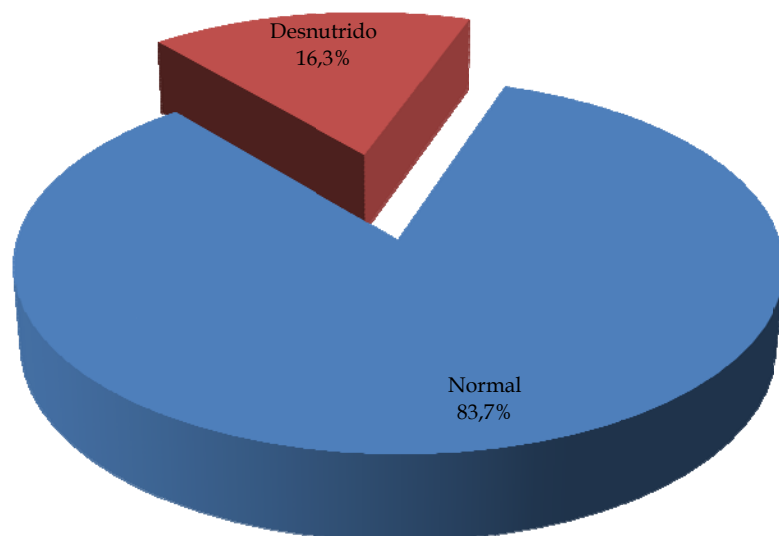
Figura 2.5.2: Estado nutricional de los niños según el índice antropométrico
Peso para la Edad



ÍNDICE TALLA PARA LA EDAD.

En el siguiente gráfico se muestra la distribución de frecuencia del estado nutricional de los niños utilizando el índice antropométrico Talla para la Edad. En la figura se aprecia que según este índice el 16.3% de los niños son considerados como desnutridos; esto indica que según este índice un buen porcentaje de los niños de nuestro país no tienen el tamaño que deberían de tener para su edad, dicho porcentaje corresponde a 755 niños. Mientras que los restantes 3,869 niños se consideran normales y representan el 83.7% del total de niños.

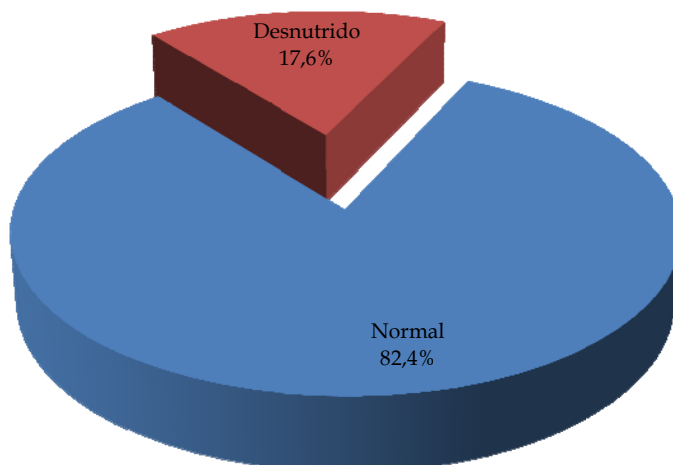
Figura 2.5.3: Estado nutricional de los niños según el índice antropométrico Talla para la Edad.



ÍNDICE COMBINADO.

En el siguiente gráfico se muestra la distribución de frecuencia del estado nutricional de los niños utilizando el índice antropométrico combinado de los tres anteriores. En la figura se aprecia que según este índice el 17.6% de los niños son considerados como desnutridos, dicho porcentaje corresponde a 814 niños. Mientras que los restantes 3,810 niños se consideran normales o mejor dicho no desnutridos y representan el 82.4% del total de niños.

Figura 2.5.4: Estado nutricional de niño según índice antropométrico combinado



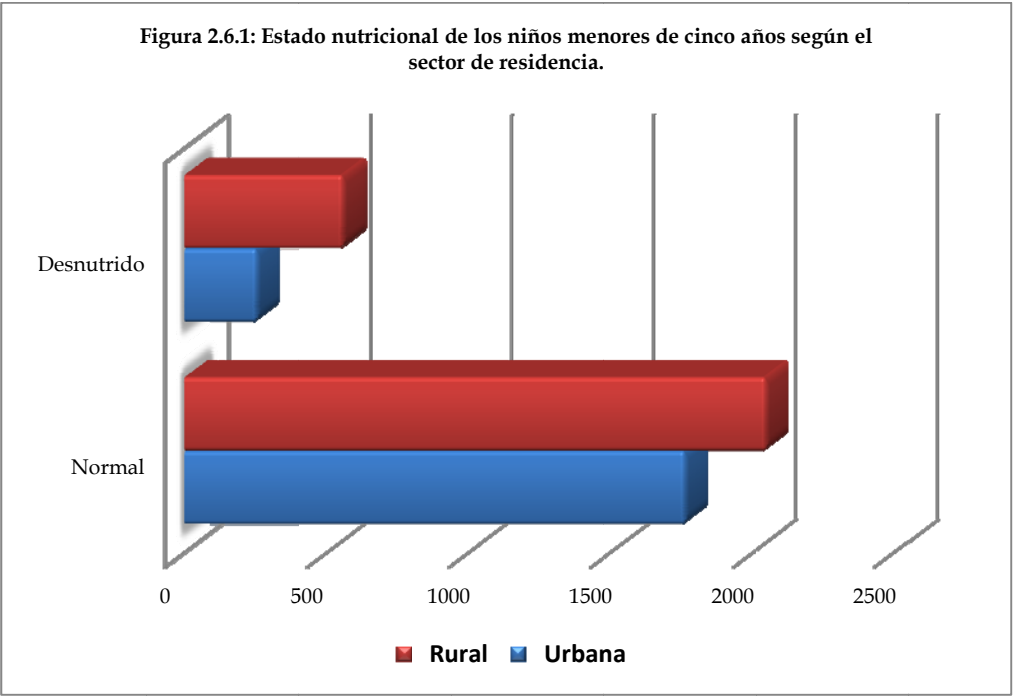
2.6 ESTADO NUTRICIONAL DE LOS NIÑOS EN FUNCIÓN DE LAS VARIABLES DEMOGRÁFICAS.

En esta sección se hace un breve estudio del estado nutricional de los niños en comparación con algunas variables demográficas seleccionadas. No basta solamente con saber cuántos niños desnutridos han sido identificados, sino que también hay que saber cómo están distribuidos en nuestro país, es decir, donde existe mayor riesgo de que los niños sufran de algún tipo de desnutrición. Se toma para esto el Índice de nutrición combinado.

ESTADO NUTRICIONAL SEGÚN SECTOR DE RESIDENCIA.

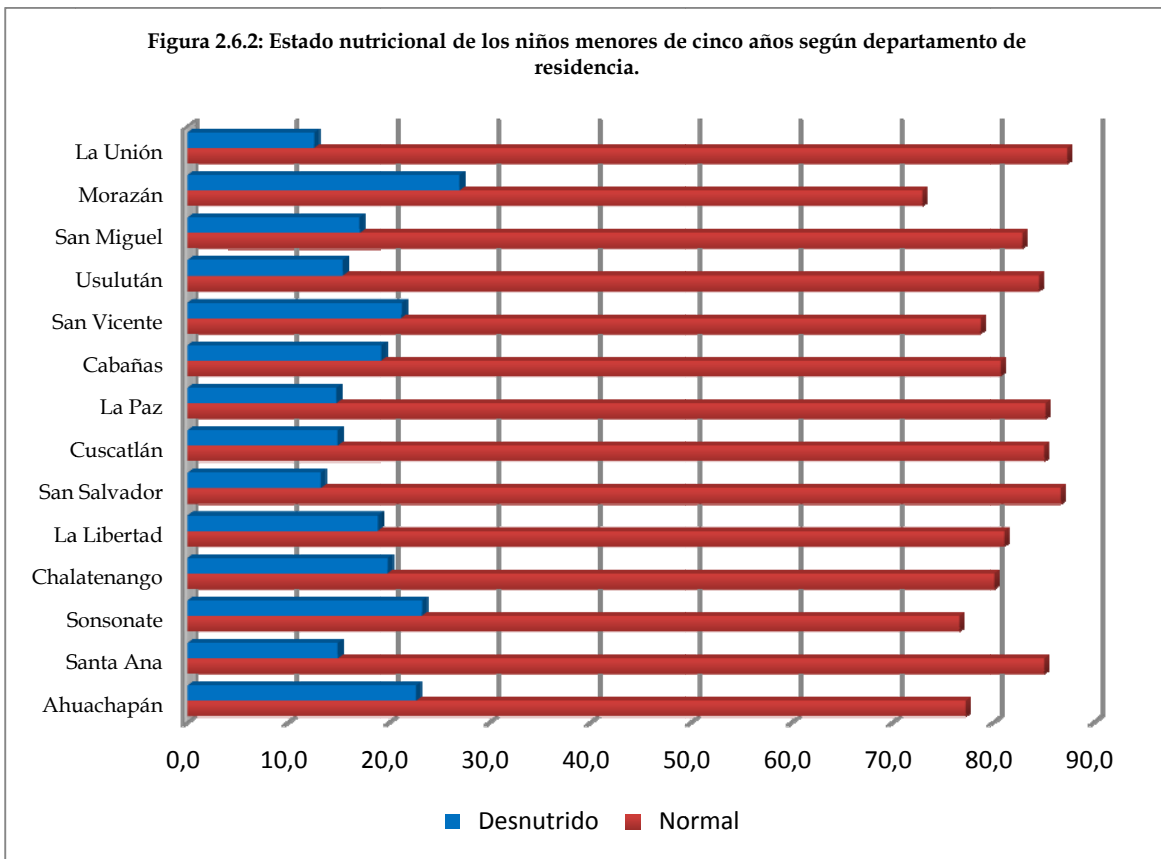
En la figura que se muestra a continuación se ilustra cómo se comporta el estado nutricional de los niños según el sector de residencia de los mismos. Los niños desnutridos son más frecuentes en el sector rural que en el urbano,

aproximadamente dos de cada tres niños desnutridos residen en las distintas zonas rurales de nuestro país. Mientras que para los niños normales no parece existir diferencia entre el sector urbano y el rural. Se puede concluir que el estado nutricional depende en cierto modo del sector de residencia. Más adelante cuando se utilice la regresión logística se verá en qué grado sucede dicha dependencia.



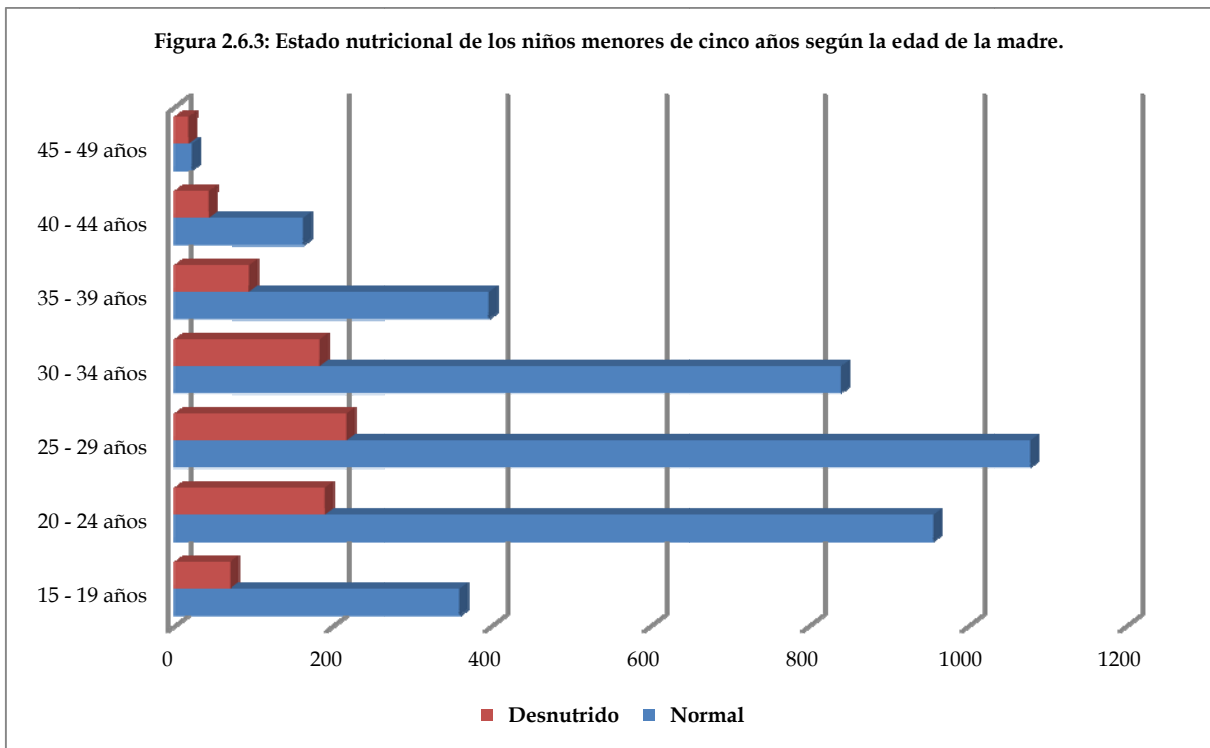
ESTADO NUTRICIONAL SEGÚN DEPARTAMENTO DE RESIDENCIA.

En la figura que se muestra a continuación se ilustra cómo se comporta el estado nutricional de los niños según el departamento de residencia de los mismos. Podemos apreciar que el porcentaje de niños desnutridos observados en cada uno de los departamentos se encuentra entre un 11% y un 22%, con excepción del departamento de Morazán en que se encuentran aproximadamente un 27%. Por lo que en primera instancia, todo parece indicar que esta variable podría resultar ser importante en los modelos que estimaremos más adelante.



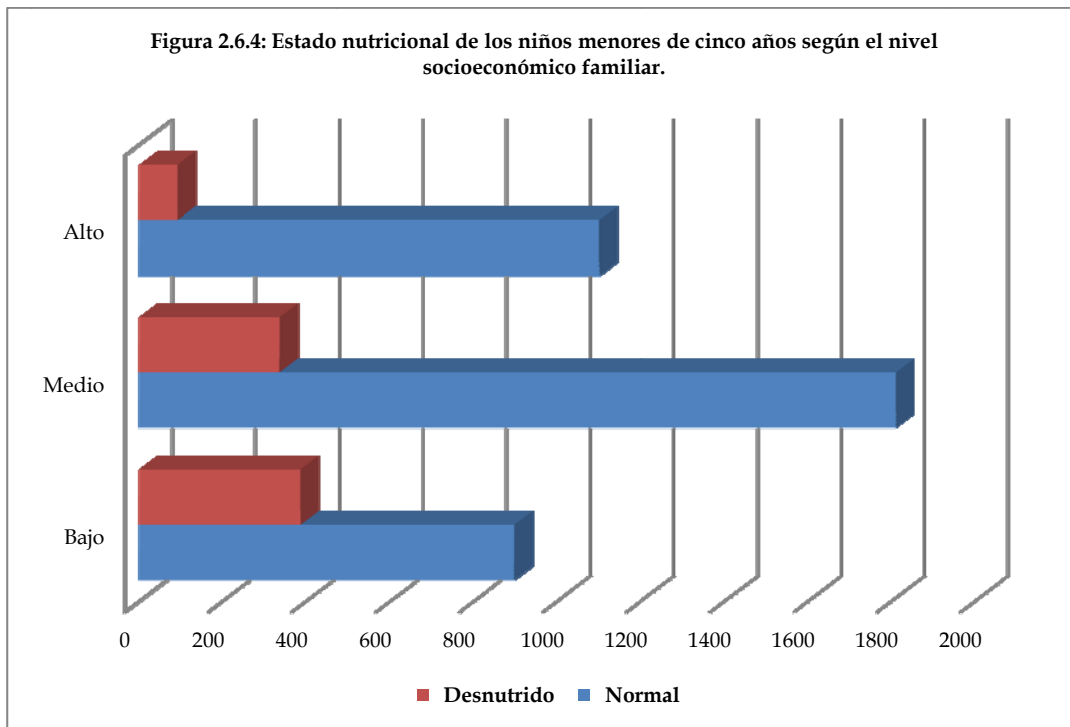
ESTADO NUTRICIONAL SEGÚN EDAD DE LA MADRE.

En la figura que se muestra a continuación se ilustra cómo se comporta el estado nutricional de los niños según la edad de la madre. De la figura se puede apreciar que el estado nutricional de los niños depende en cierto grado de la edad de la madre. Como cabría esperar para las madres jóvenes, digamos no mayores a 30 años, es mucho más probable que su hijo sufra de desnutrición (tal vez se deba a la falta de experiencia, a motivos de trabajo o simplemente a negligencia o descuido), con excepción del grupo de 15 a 19 años, que probablemente se debe a que por ser madres muy jóvenes las madres de éstas (abuelas de los niños) están más pendientes del cuidado del niño. En las madres mayores por su parte, claramente hay menos niños desnutridos que en las jóvenes.



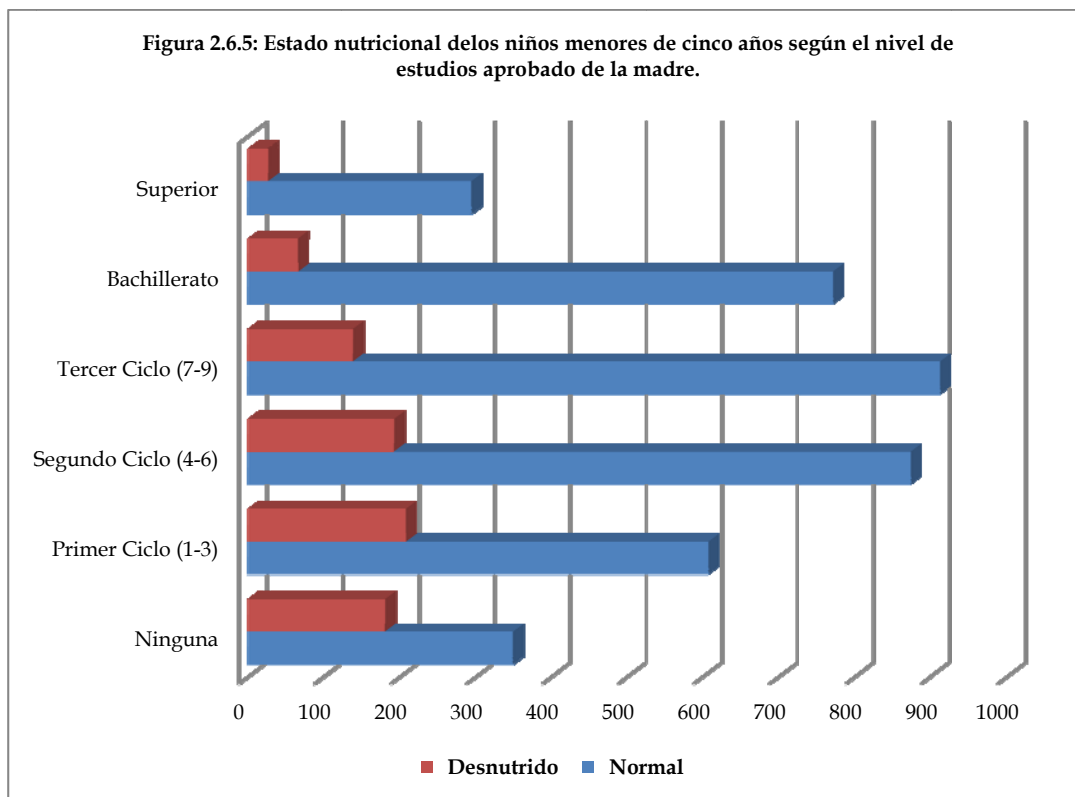
ESTADO NUTRICIONAL SEGÚN NIVEL SOCIOECONÓMICO.

En la figura que se muestra a continuación se ilustra cómo se comporta el estado nutricional de los niños según el nivel socioeconómico de la familia. Claramente como era de esperar las familias con un nivel alto de ingresos son menos propensas a presentar niños con desnutrición. Es decir, a mayor nivel de ingresos menor riesgo de desnutrición y viceversa. Esto indica que esta variable es de vital importancia a la hora de identificar a niños desnutridos.



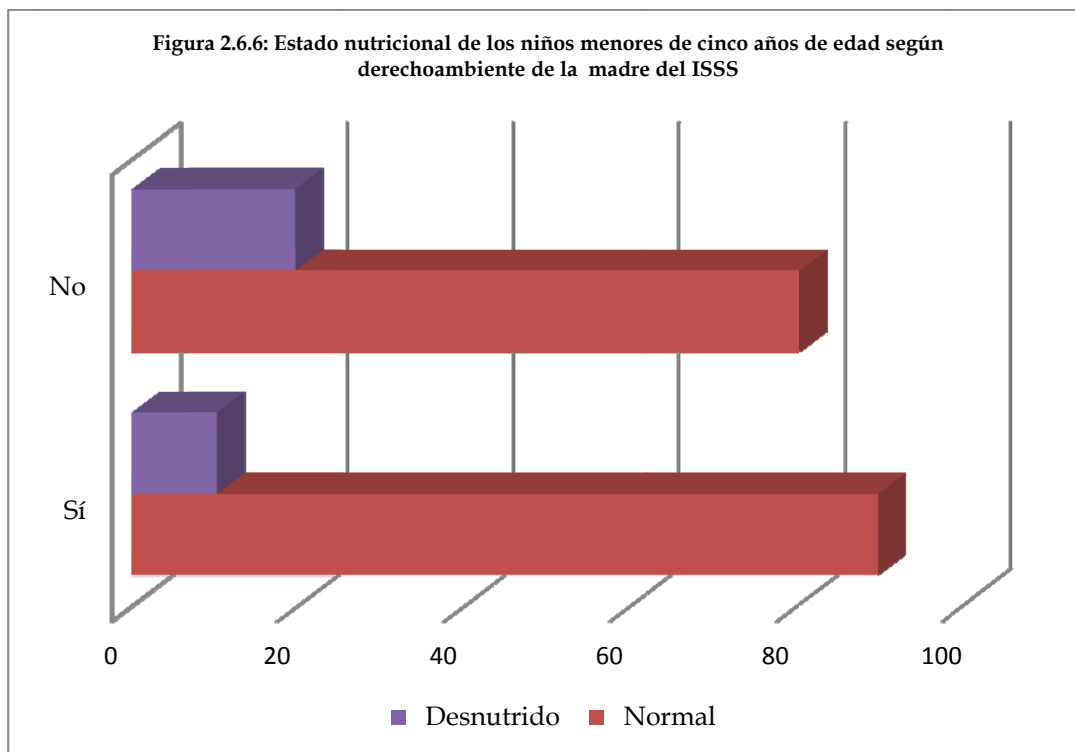
ESTADO NUTRICIONAL SEGÚN NIVEL EDUCATIVO APROBADO DE LA MADRE.

En la figura que se muestra a continuación se ilustra cómo se comporta el estado nutricional de los niños según el nivel de estudios aprobado de la madre. Como también era de esperar el estado nutricional de los niños está fuertemente relacionado con el nivel de estudios aprobado de la madre, pues cuanto mayor sea el número de años aprobado menor es el riesgo de que el niño sufra de desnutrición. De las madres que no lograron aprobar ni siquiera el primer ciclo de estudios se obtiene menor número de niños desnutridos que las que aprobaron el primer ciclo. Es de mencionar que el porcentaje de niños desnutridos en cada uno niveles de estudios disminuye a medida que estos niveles aumentan.



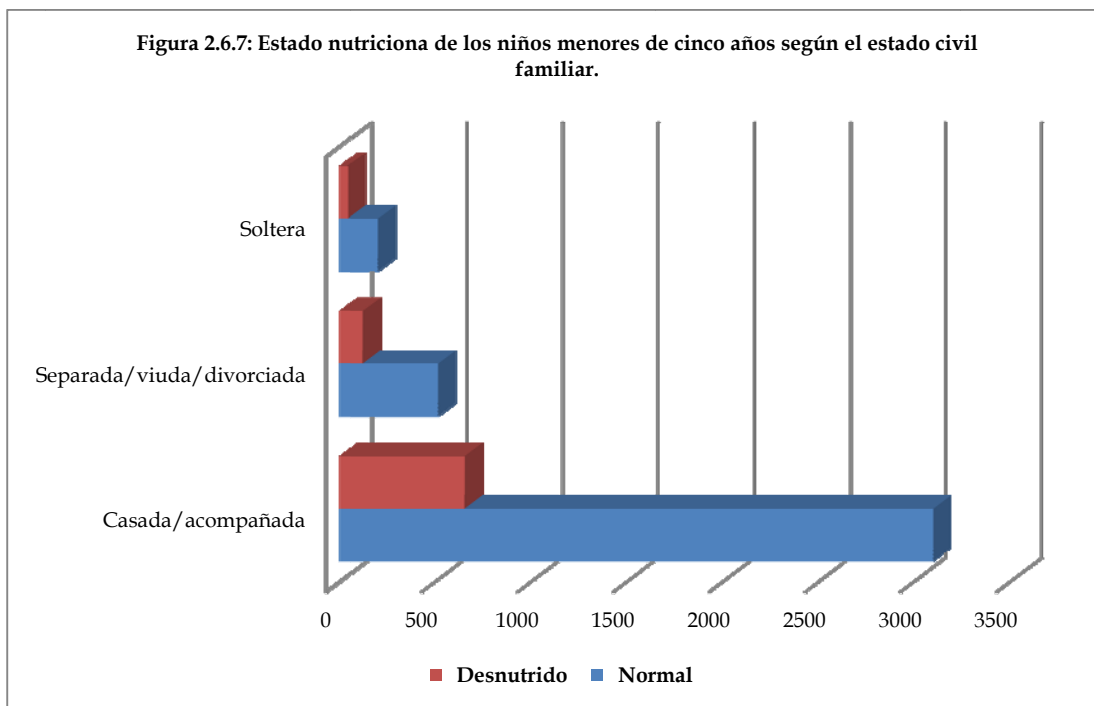
ESTADO NUTRICIONAL SEGÚN DERECHOHABIENTE DEL ISSS.

En la figura que se muestra a continuación se ilustra cómo se comporta el porcentaje de niños desnutridos y no desnutridos según beneficio médico del ISSS de la madre. De la figura mostrada, se puede decir que el comportamiento del estado nutricional es bastante similar en las madres que tienen beneficio médico de las que no lo tienen. Por lo que todo parece indicar que esta variable no será relevante en los modelos que estudiaremos más adelante.



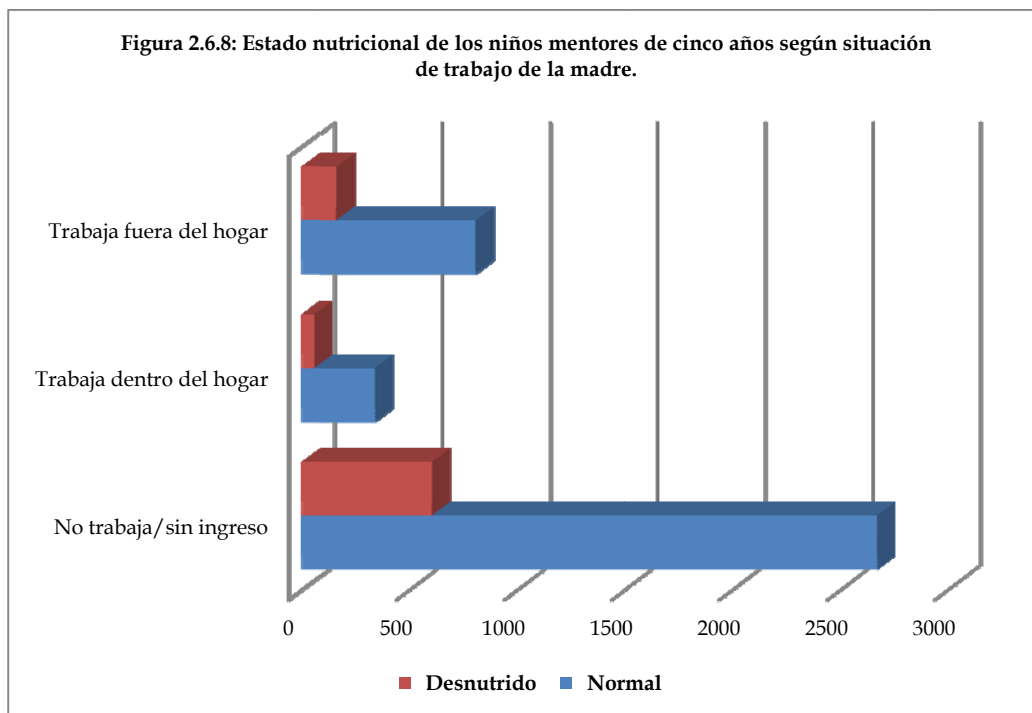
ESTADO NUTRICIONAL SEGÚN ESTADO CIVIL FAMILIAR.

En la figura que se muestra a continuación se ilustra cómo se comporta el estado nutricional de los niños según el estado civil familiar. Del gráfico se puede concluir que el estado nutricional de los niños no está relacionado con el estado familiar de sus familias; pues, si bien es cierto que el número de niños no desnutridos aumenta a medida que lo hace su situación familiar (pasar de autofinanciarse la madre sola a recibir ayuda de su pareja) también lo hacen los niños desnutridos, resultado contradictorio a nuestra intuición de la vida, pues esperaríamos que disminuyeran y no que aumentarían.



ESTADO NUTRICIONAL SEGÚN SITUACIÓN DE TRABAJO DE LA MADRE.

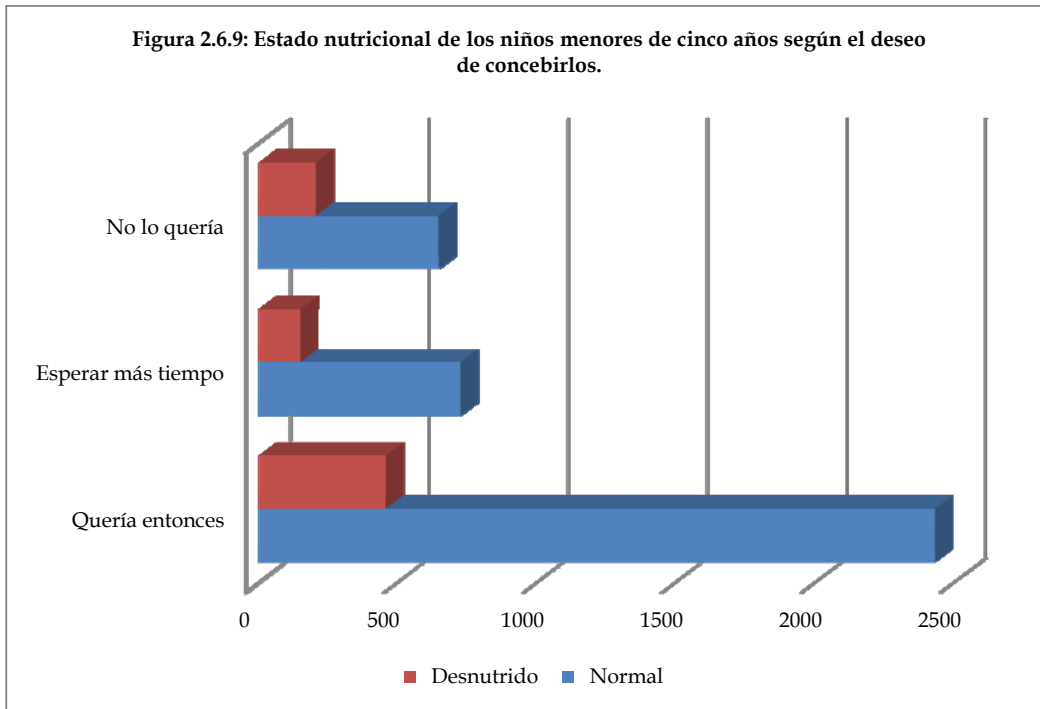
En la figura que se muestra a continuación se ilustra cómo se comporta el estado nutricional de los niños según la situación de trabajo de la madre. Al igual que en el caso anterior, no parece haber indicios que el estado nutricional de los niños dependa de la situación de trabajo de la madre. Si bien es cierto que la mayor cantidad de niños desnutridos corresponden a madres que no trabajan, también es cierto que los niños no desnutridos corresponden en su mayoría a las madres que no trabajan, probablemente se debe a que tengan un mayor cuidado de la alimentación al pasar más tiempo con sus hijos. Si calculamos el porcentaje de niños desnutridos en cada una de las situaciones de trabajo, podremos darnos cuenta que los porcentajes son bastante parecidos.



ESTADO NUTRICIONAL SEGÚN DESEO DE CONCEBIRLO.

En la figura que se muestra a continuación se ilustra cómo se comporta el estado nutricional de los niños según el deseo de concebirlo. Claramente los niños no desnutridos se presentan más para las parejas que desean concebirlos que las parejas que no los querían o que desean esperar más tiempo; sin embargo, lo mismo podría decirse para los niños no desnutridos. Lo cual estaría indicando que el estado nutricional de los niños no parece estar relacionado con el deseo de concebirlos de sus padres. Más adelante, cuando consideremos la regresión logística y el análisis discriminante veremos si esta variable es importante o no en el estado nutricional de los niños.

Figura 2.6.9: Estado nutricional de los niños menores de cinco años según el deseo de concebirlos.



CAPÍTULO III
REGRESIÓN LOGÍSTICA Y
ANÁLISIS
DISCRIMINANTE EN LA
DETERMINACIÓN DEL
ESTADO NUTRICIONAL
DE LOS NIÑOS.

3.1 INTRODUCCIÓN.

En este capítulo se presenta la aplicación de los modelos de regresión logística y el análisis discriminante para determinar cuáles son las variables más importantes que influyen en el estado nutricional de la niñez salvadoreña.

Se comienza con una descripción del procedimiento que se utilizará para estimar y validar los modelos. Posteriormente se presenta la aplicación del modelo logístico; con una parte de los datos se estiman sus parámetros, se evalúa la bondad de ajuste del modelo a los datos, se analiza el resumen de clasificación, finalmente se valida el modelo encontrado con la parte de los datos destinada para este fin.

Para el análisis discriminante, se procede de una forma muy similar al de la regresión logística. Es decir, con una parte de los datos se estiman los parámetros de la función lineal y de la función cuadrática discriminante, se evalúa entre estas dos funciones cuál es la más adecuada al conjunto de datos; finalmente se valida el modelo discriminante con la parte de datos destinada para este fin.

Se hace una comparación y selección entre ambas técnicas de clasificación, eligiendo la más óptima al problema de estimación del estado nutricional de la niñez salvadoreña. Finalmente se estiman los modelos de regresión logística y análisis discriminante utilizando el 100% de los datos y comparando con los resultados obtenidos al utilizar solamente una parte de ellos.

3.2 DIVISIÓN DE LA MUESTRA.

Para la aplicación del Análisis Discriminante y la Regresión Logística se ha decidido dividir la muestra en dos partes: con la primera parte de la muestra que consiste de un total de 3,478 niños menores de cinco años, quienes representan aproximadamente el 75% total, se estimarán los parámetros de ambos modelos.

Con los restantes 1,146, y que representan aproximadamente el 25% restante, se validarán los modelos encontrados. Pues esta es la manera más usual y adecuada de proceder, el tamaño de la muestra es lo suficientemente grande que permite hacerlo.

Para seleccionar ambas partes se utilizó Muestreo Aleatorio Simple, el cual viene implementado en SPSS (solamente hay que indicarle la cantidad de datos que deben seleccionarse para trabajar). Con este tipo de muestreo se tiene en cuenta el tamaño de las diferentes variables consideradas en el estudio, se puede evitar por ejemplo, que en el conjunto de validación exista mayor proporción de niños del sector urbano que en el conjunto de estimación.

En resumen, se verificará si las funciones encontradas se ajustan adecuadamente a la parte de los datos no tomados en cuenta en la estimación de las funciones discriminantes y del modelo de regresión logística.

En el siguiente cuadro se muestra la distribución de frecuencia de los datos en la fase de estimación según el estado nutricional tomando como variable dependiente el índice de nutrición combinado definido en el segundo capítulo de este trabajo de investigación. El 82.4% de los niños son desnutridos, mientras que el restante 17.6% son no desnutridos (porcentajes muy parecidos a los de toda la muestra). Intentaremos encontrar las funciones con las cuales clasifiquemos correctamente a la mayor cantidad de niños posibles.

Tabla 3.2.1: Estado nutricional de los niños. Fase de estimación.

Indice de Nutrición Combinado		Frecuencia	Porcentaje
Válidos	Normal	2867	82,4
	Desnutrido	611	17,6
	Total	3478	100,0

Además, la muestra consiste del 42.6% de niños residentes en el área urbana y el restante 57.4% residen en el área rural (porcentajes muy parecidos a los de toda la muestra).

Tabla 3.2.2: Sector de residencia de los niños. Fase de estimación.

Sector de residencia		Frecuencia	Porcentaje
Válidos	Rural	1997	57,4
	Urbana	1481	42,6
	Total	3478	100,0

3.3 APLICACIÓN DE LA REGRESIÓN LOGÍSTICA.

A continuación procedemos a efectuar el análisis de regresión logística para determinar el estado nutricional de los niños menores de cinco años en el momento en que se llevó a cabo la Encuesta Nacional de Salud Familiar FESAL 2008. En este apartado presentamos los resultados finales junto con los comentarios más sobresalientes después de determinar el mejor conjunto de variables óptimo.

3.4.1 VARIABLES SIGNIFICATIVAS.

La base de datos consta de 519 variables y 4624 registros. En su mayoría las 519 variables son de tipo cualitativo, muchas de ellas medidas en escala dicotómica, pues miden la ausencia o presencia de determinadas características.

Como un primer criterio para empezar a tomar variables en el modelo logístico se consideró conveniente descartar todas aquellas variables que presentaban una alta tasa de no respuesta (más de un 10%), pues esto haría que el tamaño de la muestra decreciera considerablemente y que además claramente no guardaban relación con mi variable de estudio (variables asociadas con la identificación de la unidad muestral y variables relacionadas con el desarrollo de la entrevista). Se realizaron contraste de independencia para poder descartarlas definitivamente.

Por otra parte, es de mencionar que existen variables que se encuentran en la base de datos más de una vez, pero medidas en escalas diferentes; así por ejemplo, el nivel educativo de las madres se encuentra medido por una variable donde se hace referencia a los años de estudio aprobado, y así mismo por otra variable que mide el último nivel de estudios aprobado; en estos caso se optó por tomar aquellas variables que están medidas en escala numérica.

Considerando esto, el número de variables candidatas a formar parte del modelo logístico se reduce a poco menos de treinta. De éstas, para las variables cualitativas se realizaron tablas de contingencia y pruebas de independencia con la variable dependiente índice de nutrición combinado. Para las variables numéricas, se realizó un diagrama de dispersión simple para medir el poder de clasificación que podía apreciarse (se considera que tiene poder de clasificación si se observa un comportamiento que pudiera sugerir que la variable se comporta de manera diferente para niños desnutridos y no desnutridos). Estos análisis se toman en cuenta a la hora incorporar variables en el modelo logístico; en otras palabras lo que se hizo fue tomar a aquellas variables que están claramente relacionadas con la variable dependiente.

También, a la hora de seleccionar las variables se tomó en cuenta la relación que pudiera existir entre ellas; en caso de existir conjuntos de variables relacionadas se optó por tomar únicamente las variables de las cuales se tenga la mayor tasa de respuesta, y a la vez seleccionar a aquellas variables de las que en principio, a la hora de estimar el estado nutricional de los niños, se tenga información a priori, con el fin de no tener que recolectar esa información en futuras encuestas.

Se utilizaron también los métodos de selección hacia adelante, selección hacia atrás y selección por pasos para seleccionar variables en SPSS, obteniéndose en todos ellos el mismo conjunto final de cinco variables.

En el siguiente cuadro se muestran las variables que resultan ser importantes en el modelo de regresión logística, y por consiguiente son las variables más útiles en la determinación del estado nutricional de los niños menores de cinco años.

Tabla 3.3.1: Variables finales en el modelo de regresión logística. Fase de estimación

NOMBRE	SIGNIFICADO
Gindice	Es la variable dependiente con la cual se identifica el estado nutricional del niño: vale 1 si es desnutrido y 0 en caso contrario.
Edupromef	Representa el número de años completos de estudios aprobado por la madre del niño.
P324A	Representa el número de hijos que viven en el mismo hogar del niño en estudio.
ServBasicos	Representa el número de servicios básicos con que cuenta la familia.

El cuadro que se presenta a continuación muestra la codificación que se ha dado a la variable dependiente, índice de nutrición combinado.

Tabla 3.3.2: Codificación de variables en Regresión Logística. Fase de estimación.

Codificación de la variable dependiente	
Valor original	Valor interno
Normal	0
Desnutrido	1

3.4.2 ESTIMACIÓN DEL MODELO LOGÍSTICO.

En el cuadro que se presenta a continuación se puede visualizar el orden en el cual las variables se han incorporando al modelo logístico junto con sus respectivas estimaciones de los parámetros (presentamos únicamente los resultados de la inclusión de variables con el método de selección hacia adelante). La primera variable en incorporarse corresponde al nivel socioeconómico de la familia, mientras que la última, y no por eso deja de ser la de menor importancia, corresponde a si la familia recibe ayuda económica en conceptos de remesas por parte de algún familiar residente en el extranjero.

El modelo logit de la probabilidad de sufrir desnutrición es;

$$\ln\left(\frac{P}{1-P}\right) = -0.893 - 0.197(\text{ServBasi cos}) + 0.142(P324A) - 0.054(\text{edupro m ef})$$

(3.1)

Por lo que el modelo de regresión logística toma la siguiente forma:

$$P(Gindice = 1) = \frac{1}{1 + \exp[0.893 + 0.197(ServBasicos) - 0.142(P324A) + 0.054(edupromef)]}$$

(3.2)

Lo que podemos decir es lo siguiente:

- El número de servicios básicos con que cuenta la familia puede considerarse como la variable más importante en la determinación del estado nutricional de los niños
- El número de hijos que viven en el mismo hogar también influyen positivamente en el estado nutricional de los niños, un resultado bastante acorde a nuestra realidad, pues las familias no podrán brindarles cantidad pero sobre todo calidad en alimentación a todos los niños.
- Como se esperaba el número de años de educación de la madre contribuye significativamente en la disminución del estado nutricional de los niños.

Basándonos en los resultados anteriores, podemos decir que: para determinar el estado nutricional de los niños según el modelo logístico basta considerar variables que traten sobre la composición o tamaño de la familia, el nivel educativo de la madre y el número de servicios básicos con que se cuenta en el hogar. Variables tales como: la edad de la madre, sector de residencia, estado civil familiar, entre otras; que en principio se creen vitales resultaron ser irrelevantes en la determinación del estado nutricional.

Tabla 3.3.3: Regresión logística, historial de interacciones. Fase de estimación.

Historial de iteraciones^{a, b, c, d}

Iteración		-2 log de la verosimilitud	Coeficientes			
			Constant	ServBasicos	P324A	edupromef
Paso 1	1	3116,500	-,647	-,157		
	2	3037,539	-,552	-,255		
	3	3034,975	-,514	-,280		
	4	3034,970	-,512	-,281		
	5	3034,970	-,512	-,281		
Paso 2	1	3081,045	-1,034	-,135	,120	
	2	2996,443	-1,113	-,223	,167	
	3	2993,637	-1,101	-,246	,173	
	4	2993,631	-1,100	-,248	,174	
	5	2993,631	-1,100	-,248	,174	
Paso 3	1	3073,527	-,951	-,112	,106	-,022
	2	2983,073	-,945	-,179	,140	-,044
	3	2979,378	-,897	-,196	,142	-,053
	4	2979,367	-,893	-,197	,142	-,054
	5	2979,367	-,893	-,197	,142	-,054

a. Método: Por pasos hacia adelante (Condicional)

b. En el modelo se incluye una constante.

c. -2 log de la verosimilitud inicial: 3232,958

d. La estimación ha finalizado en el número de iteración 5 porque las estimaciones de los parámetros han cambiado en menos de ,001.

En el siguiente cuadro se muestra la Prueba de Hosmer y Lemeshow. Dicha prueba se utiliza para medir la correspondencia entre los valores observados y los pronosticados con el modelo logístico, (utilizando para medir dicha correspondencia la prueba Chi-Cuadrado). Un valor de Chi-Cuadrado no significativo indica un buen ajuste del modelo. Como puede apreciarse en el cuadro, el modelo al cual hemos llegado presenta una significancia bastante alta, y significa que no existe evidencia estadística para decir que la distribución de los

valores de la variable dependiente y el de sus pronósticos con el modelo de regresión logística sea distinta, es decir, que no existe evidencia estadística para decir que el modelo de regresión logística encontrado se ajuste mal a los datos.

Tabla 3.3.4: Regresión logística, prueba de Hosmer y Lemeshow. Fase de estimación.

Prueba de Hosmer y Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	10,880	7	,144
2	9,992	8	,266
3	6,072	8	,639

En el siguiente cuadro se muestran tres principales medidas de resumen del modelo. La primera de ellas corresponde a la lejanía, y mide la importancia de incorporar una variable más al modelo; la segunda el R^2 de Cox y Snell; mientras que el tercero el R^2 de Nagelkerke, estas dos medidas juegan un papel similar al coeficiente de determinación en regresión lineal múltiple, es decir, se utilizan para medir el ajuste del modelo a los datos. Sin embargo, es de mencionar que el R^2 de Cox y Snell nunca puede tomar valores de 1; mientras que el R^2 de Nagelkerke es una modificación y se encuentra definido en el intervalo de 0 a 1. Para que un modelo sea considerado exitoso debe contener una lejanía pequeña y los mayores valores posibles de R^2 tanto de Cox y Snell como de Nagelkerke.

En base a esto, se puede decir que se pueden tomar todas las variables anteriormente mencionadas y con esto el modelo se ajustaría a los datos; sin embargo, es de mencionar que los valores de R^2 son bastante bajos para ambos, pero es lo mejor que se puede alcanzar con los datos disponibles.

Tabla 3.3.5: Regresión logística, resumen del modelo. Fase de estimación.

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	3034,970 ^a	,055	,091
2	2993,631 ^a	,066	,110
3	2979,367 ^a	,070	,116

a. La estimación ha finalizado en el número de iteración 5 porque las estimaciones de los parámetros han cambiado en menos de ,001.

En el siguiente cuadro se muestran los resultados de la clasificación de los niños en base al modelo de regresión logística. En este momento se ha considerado conveniente tomar como punto de corte el valor de 0.17, pues existe un porcentaje similar de niños desnutridos en los datos. Se observa, por tanto que el modelo clasifica correctamente a aproximadamente las dos terceras partes de niños desnutridos y clasifica mal a la restante tercera parte; algo similar ocurre con los niños no desnutridos. Más adelante cuando nos dediquemos a validar el modelo encontrado, definiremos apropiadamente el punto de corte para minimizar el porcentaje de niños clasificados incorrectamente.

Tabla 3.3.6: Regresión logística. Clasificación final. Fase de estimación.

Tabla de clasificación^a

Observado			Pronosticado		
			Indice de Nutrición Combinado		Porcentaje correcto
			Normal	Desnutrido	
Paso 1	Indice de Nutrición Combinado	Normal	1895	972	66,1
		Desnutrido	244	367	60,1
	Porcentaje global				65,0
Paso 2	Indice de Nutrición Combinado	Normal	1830	1037	63,8
		Desnutrido	229	382	62,5
	Porcentaje global				63,6
Paso 3	Indice de Nutrición Combinado	Normal	1815	1052	63,3
		Desnutrido	207	404	66,1
	Porcentaje global				63,8

a. El valor de corte es ,170

En el siguiente cuadro se muestra la significancia de cada uno de los parámetros del modelo de regresión logística encontrado; todos los parámetros son significativos y por consiguiente se trabajará con todos ellos. Es de mencionar, que si eliminamos el término independiente se encuentra un modelo, que si bien clasifica un poco mejor a los niños desnutridos, clasifica un poco mal a los niños no desnutridos; mas sin embargo, el ajuste mediante la prueba de Hosmer nos indica que es malo, por consiguiente se toma en cuenta (consúltese los anexos).

Por lo que el modelo de regresión logística es:

$$P(G_{\text{indice}} = 1) = \frac{1}{1 + \exp\left[0.893 + 0.197(\text{ServBasi cos}) - 0.142(P324A) + 0.054(\text{edupro m ef})\right]}$$

(3.3)

Tabla 3.3.7: Regresión logística, significancia de los parámetros. Fase es estimación.

		Variables en la ecuación						I.C. 95% para EXP(B)	
		B	E.T.	Wald	gl	Sig.	Exp(B)	Inferior	Superior
Paso 1 ^a	ServBasicos	-,281	,021	183,031	1	,000	,755	,725	,787
	Constante	-,512	,081	39,547	1	,000	,599		
Paso 2 ^b	P324A	,174	,027	42,400	1	,000	1,190	1,129	1,254
	ServBasicos	-,248	,021	133,789	1	,000	,781	,749	,814
	Constante	-1,100	,124	78,785	1	,000	,333		
Paso 3 ^c	P324A	,142	,028	25,889	1	,000	1,153	1,091	1,217
	edupromef	-,054	,014	14,135	1	,000	,947	,921	,975
	ServBasicos	-,197	,025	61,434	1	,000	,821	,782	,863
	Constante	-,893	,135	43,466	1	,000	,410		

a. Variable(s) introducida(s) en el paso 1: ServBasicos.

b. Variable(s) introducida(s) en el paso 2: P324A.

c. Variable(s) introducida(s) en el paso 3: edupromef.

3.4.3 VALIDACIÓN DEL MODELO LOGÍSTICO.

Como se dijo anteriormente la muestra considerada en este trabajo ha sido dividida en dos partes, una parte de ella ya fue utilizada para la estimación de los parámetros del modelo de regresión logística. En esta sección procedemos a validar el modelo de regresión logística encontrado previamente con la otra parte de datos almacenados para este fin.

En el siguiente cuadro se muestra la distribución de los datos que se usarán para validar el modelo de regresión logística según el estado nutricional de los niños. Se observa que el 17.7% de la muestra corresponde a niños desnutridos, mientras que el restante 82.3% corresponde a niños no desnutridos. Estos porcentajes son bastante similares a los que se usaron en la fase de estimación.

Tabla 3.3.8: Estado nutricional de los niños. Fase de validación.

Índice de Nutrición Combinado		Frecuencia	Porcentaje
Válidos	Normal	943	82,3
	Desnutrido	203	17,7
	Total	1146	100,0

La expresión (3.3) corresponde al modelo logístico con el cual se determinará la probabilidad de que un niño sufra de desnutrición.

El procedimiento para validar el modelo se hizo en lo siguiente:

- Calculamos la probabilidad de que el niño sufra de desnutrición con ayuda de la ecuación (3.3).
- Luego utilizamos diferentes puntos de corte, de tal manera que podamos elegir aquel modelo en el cual el porcentaje de clasificación de niños desnutridos como no desnutridos, y de los niños no desnutridos clasificados como desnutridos sea tolerable.
- Para poder elegir el punto de corte se procedió a calcular la clasificación del modelo logístico para diferentes valores que van desde 0.00 hasta 0.20 en incrementos de 0.005 en 0.005, y de 0.20 hasta 0.30 en incrementos de 0.1 en 0.1.
- Una vez obtenidos los porcentajes de niños clasificados incorrectamente para normales y desnutridos (cuyo cuadro resumen se presenta en los anexos); se procedió a fijar el máximo porcentaje de niños desnutridos

clasificados incorrectamente como no desnutridos. Dicho porcentaje se fijó en un 10%, pues el modelo en principio está considerado para identificar a niños desnutridos y ver qué se puede hacer con ellos. Pues es más grave dejar a un niño desnutrido sin ninguna ayuda. Y luego se trató de minimizar el porcentaje de niños no desnutridos como desnutridos.

- El punto de corte elegido es de 0.085; esto significa que para un niño cuya probabilidad de desnutrir con el modelo logístico es superior o igual a dicho valor, entonces se considera como desnutrido, de lo contrario será considerado como normal.
- Aplicaremos la prueba de Hosmer Lemeshow para evaluar la bondad de ajuste del modelo a los datos, enseguida explicamos un detalle este paso.
- Finalmente calcularemos el error de estimación del modelo, $(G_{indice = 1}) - P(G_{indice = 1})$, y realizaremos un análisis de los residuos.

Para validar el modelo se han creado 10 grupos y poder así utilizar la prueba de Hosmer y Lemeshow. Los grupos se han formado de la siguiente manera: en el primer grupo se encuentran todos aquellos niños cuya probabilidad de ser desnutridos con la regresión logística resulta ser menor o igual al 10%; en el segundo grupo se encuentran aquellos niños cuya probabilidad de ser desnutridos es mayor al 10% y menor o igual al 20%, y así sucesivamente, hasta que el décimo grupo corresponde a todos aquellos niños cuya probabilidad de sufrir desnutrición es mayor al 90% y menor o igual al 100%.

La variable con la cual se identifica dichos grupos ha sido llamada DEC_LOGISTICA, y toma el valor de 1 si la probabilidad de desnutrido no sobrepasa el 10%; 2 si es mayor del 10% pero menor o igual al 20%, y así sucesivamente hasta el valor de 10.

En el siguiente cuadro se muestra la suma de las probabilidades de desnutrir para cada uno de los grupos antes mencionados, estas sumas son necesarias para poder realizar la prueba Hosmer y Lemeshow, y representan el número esperado de niños desnutridos en cada uno de los grupos.

Tabla 3.3.9: Frecuencias esperadas modelo de regresión logística. Fase de validación.

Resúmenes de casos

Suma

DEC_LOGISTICA	Probabilidad de desnutrir
1,00	26,76
2,00	60,66
3,00	47,81
4,00	37,15
5,00	22,33
Total	194,71

Es de mencionar que en el último grupo se han incorporado todos aquellos que tienen probabilidad de desnutrir superior al 0,4, para asegurarnos que las frecuencias esperadas sean mayores a 5.

En el cuadro que se muestra a continuación se muestra el número de niños desnutridos y no desnutridos en cada uno de los grupos formados previamente.

Debe tomarse el número de niños desnutridos en cada grupo como frecuencias observadas en la prueba de Hosmer y Lemeshow y así poder calcular el estadístico de Chi-Cuadrado (y compararlo con el valor crítico de una Chi-Cuadrado con 9 grados de libertad), con el cual contrastaremos la bondad del ajuste del modelo de regresión logística a los datos.

Tabla 3.3.10: Estado nutricional de los niños en los grupos creados. Fase de validación.

Tabla de contingencia DEC_LOGISTICA * Indice de Nutrición Combinado

Recuento

		Indice de Nutrición Combinado		Total
		Normal	Desnutrido	
DEC_LOGISTICA	1,00	339	29	368
	2,00	359	66	425
	3,00	150	44	194
	4,00	72	37	109
	5,00	23	27	50
Total		943	203	1146

- El estadístico de prueba se define como:

$$\chi_0^2 = \sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i}$$

- O_i representa las frecuencias observadas de niños desnutridos con el modelo de regresión logística en cada uno de los grupos.
- E_i representa las frecuencias esperadas de niños desnutridos con el modelo de regresión logística en cada uno de los grupos (las sumas de las probabilidades en cada grupo).

- El cual sigue una distribución Chi-Cuadrado con 4 grados de libertad (pues existen cuatro grupos con frecuencias esperadas iguales a cero y un grupo con frecuencias esperadas menor a 5).

Los cálculos asociados con dicha prueba se muestran en el siguiente cuadro.

Tabla 3.3.11: Regresión logística, prueba de Hosmer y Lemeshow. Fase de estimación.

DEC_LOGISTICA	PRUEBA CHI-CUADRADO				
	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1,00	29	26,76	2,24	5,009	0,187
2,00	66	60,66	5,34	28,508	0,470
3,00	44	47,81	-3,81	14,525	0,304
4,00	37	37,15	-0,15	0,022	0,001
5,00	27	22,33	4,67	21,846	0,978
				SUMA	1,940

De los cálculos anteriores se obtiene que $\chi_0^2 = 1.94$. El valor crítico para un nivel de confianza del 95% y con 4 grados de libertad es de $\chi_{4,0.95}^2 = 9.49$, con estos resultados se concluye que el modelo de regresión logística presenta una buena bondad de ajuste a los datos. De hecho la probabilidad de observar en la distribución Chi-Cuadrado un valor mayor a 1.94 con 4 grados de libertad es igual a 0.7467 (significancia de la prueba).

En el siguiente cuadro se muestra la comparación del estado nutricional de los niños y la clasificación generada mediante el modelo de regresión logística. Se observa que con el punto de corte elegido el modelo clasifica bastante bien a los niños desnutridos, pero clasifica pésimamente a los niños no desnutridos, además el modelo sobreestima el número de niños desnutridos. En términos generales el modelo de regresión logística encontrado clasifica correctamente al 429 de los niños quienes representan el 37.4%; los restantes 717 niños han sido clasificado erróneamente y representan el 62.6% del total.

Tabla 3.3.12: Regresión logística, clasificación del modelo. Fase de validación.

Tabla de contingencia Índice de Nutrición Combinado * Clasificación con el modelo de Regresión Logística

			Clasificación con el modelo de Regresión Logística		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	247	696	943
		% dentro de Índice de Nutrición Combinado	26,2%	73,8%	100,0%
	Desnutrido	Recuento	21	182	203
		% dentro de Índice de Nutrición Combinado	10,3%	89,7%	100,0%
Total	Recuento		268	878	1146
	% dentro de Índice de Nutrición Combinado		23,4%	76,6%	100,0%

Validez; es el grado en que las predicciones coinciden con las observaciones y tienen dos componentes: Calibración y Discriminación. La calibración compara el estado nutricional de los niños predichos con su verdadero estado nutricional en base al índice de nutrición combinado; la prueba de Hosmer y Lemeshow sirve para conocer la calibración del modelo, que como ya se vio antes se tiene un valor de Chi-Cuadrado $\chi_0^2 = 1.94$ con un valor de p igual a 0.7467, que como ya se mencionó antes es indicativo del buen ajuste del modelo logístico a los datos. Mientras que la discriminación evalúa el grado en que el modelo distingue entre niños desnutridos y no desnutridos.

Además se pueden definir otros criterios para determinar la capacidad predictiva del modelo.

- La sensibilidad (S) o proporción de verdaderos positivos del modelo, mide el porcentaje de niños desnutridos que han sido clasificados con el modelo logístico como niños desnutridos.
- La especificidad (E) o proporción de verdaderos negativos del modelo, mide el porcentaje de niños no desnutridos que han sido clasificados con el modelo logístico como niños no desnutridos.
- Proporción de falsos positivos (PFP), mide el porcentaje de niños no desnutridos que han sido clasificados con el modelo logístico como niños desnutridos.
- Proporción de falsos negativos (PFN), mide el porcentaje de niños desnutridos que han sido clasificados con el modelo logístico como niños no desnutridos.

A continuación se muestran estos criterios utilizando la misma muestra con la cual se obtuvieron las estimaciones de los parámetros del modelo. En el siguiente cuadro se presenta el resumen de clasificación del modelo.

Tabla 3.3.13: Clasificación final modelo de regresión logística, datos utilizados en la estimación del modelo. Fase de validación.

Tabla de contingencia Índice de Nutrición Combinado * Clasificación con el modelo de Regresión Logística

			Clasificación con el modelo de Regresión Logística		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	683	2184	2867
		% dentro de Índice de Nutrición Combinado	23,8%	76,2%	100,0%
	Desnutrido	Recuento	43	568	611
		% dentro de Índice de Nutrición Combinado	7,0%	93,0%	100,0%
Total		Recuento	726	2752	3478
		% dentro de Índice de Nutrición Combinado	20,9%	79,1%	100,0%

Del cuadro anterior podemos obtener las medidas anteriormente definidas para la valoración de la eficacia del modelo.

Tabla 3.3.14: Eficacia productiva del modelo, datos usados en la estimación del modelo. Fase de validación.

		Porcentaje
Sensibilidad del modelo	$S = \frac{568}{611}$	93,0
Especificidad del modelo	$E = \frac{683}{2867}$	23,8
Proporción de falsos positivos	$FPF = \frac{2184}{2867}$	76,2
Proporción de falsos negativos	$PFN = \frac{43}{611}$	7,0

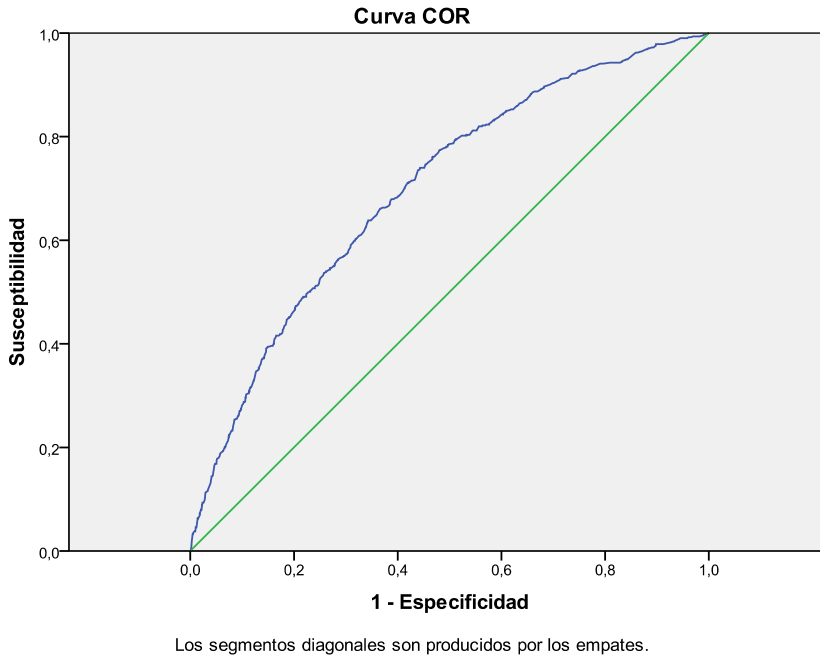
Y como ya se indicó anteriormente, el modelo clasifica correctamente en su mayoría a los niños desnutridos pero no lo hace tan bien con los niños no desnutridos.

La eficiencia productiva del modelo y el cálculo del punto de probabilidad de corte más adecuado se puede realizar a partir de las denominadas curvas ROC. La curva ROC evalúa gráficamente la capacidad del modelo para discriminar entre niños desnutridos y no desnutridos en función de los diferentes valores de la probabilidad umbral. Para ello se calculan los diversos valores de sensibilidad asociados a las diferentes probabilidades de corte o umbral. También se calculan los valores de especificidad o de su complementario (Proporción de falsos negativos) asociados a los mismos niveles de probabilidad de corte. En el eje de ordenadas se señalan los valores de sensibilidad mientras que en el eje de las abscisas los valores correspondientes a la proporción de falsos positivos que están asociados a las diferentes probabilidades de corte.

El área bajo la curva es una medida conjunta de la eficiencia productiva del modelo, un área igual a 0.5 indica que el modelo carece de poder predictivo, el valor máximo que puede alcanzar es 1 indicando un potente poder predictivo del modelo. El punto de curvatura de la curva ROC indica aquella combinación de valores de sensibilidad y especificidad que determina la probabilidad de punto de corte más apropiada.

En el siguiente gráfico que se muestra a continuación se presenta la curva ROC, para el modelo de regresión logística utilizando el mismo conjunto de datos con los cuales se estimamos sus parámetros.

Figura 3.3.15: Curva COR del modelo de regresión logístico. Fase de validación.
Datos usados para estimación del modelo.



De la figura anterior, puede tomarse que el valor de corte de la probabilidad más adecuada que se puede tomar es un valor cercano a 0.17; sin embargo, este punto de corte le da igual importancia a los dos tipos de errores de clasificación (clasificar a niños desnutridos como no desnutridos y viceversa). Por lo que utilizaremos el punto de corte anteriormente elegido de 0.085

El área bajo la curva ROC se muestra en el siguiente cuadro, en dicho cuadro podemos apreciar que el ajuste del modelo a los datos es bastante adecuado, pues el área es igual 0.699 rechazando la hipótesis nula de que el área entre la curva es igual a 0.5.

Tabla 3.3.16: Curva ROC, medidas de resumen, datos usados para estimación del modelo. Fase de validación.

Área bajo la curva

Variables resultado de contraste: Probabilidad de desnutrir

Área	Error típ. ^a	Sig. asintótica ^b	Intervalo de confianza asintótico al 95%	
			Límite inferior	Límite superior
,699	,011	,000	,677	,722

La variable (o variables) de resultado de contraste: Probabilidad de desnutrir tiene al menos un empate entre el grupo de estado real positivo y el grupo de estado real negativo. Los estadísticos pueden estar sesgados .

a. Bajo el supuesto no paramétrico

b. Hipótesis nula: área verdadera = 0,5

Al tomar como punto de corte una probabilidad de 0.085, se obtienen los siguientes resultados de clasificación del modelo para el 25% de los datos destinados para la validación del modelo.

Tabla 3.3.17: Clasificación final modelo de regresión logística. Fase de validación.

Tabla de contingencia Índice de Nutrición Combinado * Clasificación con el modelo de Regresión Logística

			Clasificación con el modelo de Regresión Logística		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	247	696	943
		% dentro de Índice de Nutrición Combinado	26,2%	73,8%	100,0%
	Desnutrido	Recuento	21	182	203
		% dentro de Índice de Nutrición Combinado	10,3%	89,7%	100,0%
Total		Recuento	268	878	1146
		% dentro de Índice de Nutrición Combinado	23,4%	76,6%	100,0%

En el siguiente cuadro se obtienen las medidas anteriormente definidas para la valoración de la eficacia del modelo.

Tabla 3.3.18: Eficacia productiva del modelo. Fase de validación.

		Porcentaje
Sensibilidad del modelo	$S = \frac{182}{203}$	89,7
Especificidad del modelo	$E = \frac{247}{943}$	26,2
Proporción de falsos positivos	$PFP = \frac{696}{943}$	73,8
Proporción de falsos negativos	$PFN = \frac{21}{203}$	10,3

Y como se comentó anteriormente el modelo de regresión logística encontrado, clasifica casi perfectamente a los niños desnutridos, más no así a los niños no desnutridos. Sin embargo, el modelo sobreestima el número de niños con problemas de nutrición.

3.4.4 INTERPRETACIÓN DEL MODELO LOGÍSTICO.

Una vez que hemos determinado el modelo de regresión logística y lo hemos validado y visto que presenta un ajuste razonable a los datos, procedemos a interpretar los parámetros del modelo, teniendo en cuenta que no existe multicolinealidad en modelo encontrado.

Como se recordará el modelo resultante es el siguiente:

$$P(Gindice = 1) = \frac{1}{1 + \exp[0.893 + 0.197(ServBasi\ cos) - 0.142(P324A) + 0.054(edupro\ mef)]}$$

(3.4)

Al escribir el modelo en oddsratio, se tendrá;

$$\frac{P(G_{indice} = 1)}{P(G_{indice} = 0)} = \exp[-0.893 - 0.197(ServBasi\ cos) + 0.142(P324A) - 0.054(edupro\ mef)]$$

(3.5)

De la ecuación anterior podemos obtener los siguientes comentarios:

- $\exp(-0.197) = 0.821$ No se puede interpretar directamente este número; pues la relación entre las variables no es de causa y efecto, es decir, incrementar el número de Servicios básicos con que cuenta la familia no reducirá siempre el riesgo de sufrir de desnutrición, en algunos casos sí pero en otros no. Las variables tienen una relación, sin embargo, ésta no es de causa y efecto, por lo que es difícil interpretar el efecto directo de los Servicios básicos en el estado nutricional de los niños.
- $\exp(0.142) = 1.153$ Significa que incrementar en una unidad el número de niños que viven el mismo hogar del niño entrevistado incrementa en un 15.3% en el riesgo de sufrir desnutrición del niño (nuevamente asumiendo que mantenemos constante al resto de variables).
- $\exp(-0.054) = 0.947$ Significa que incrementar en una unidad el número de años de estudio de la madre reduce en apenas un 5.3% en el riesgo de sufrir desnutrición del niño (nuevamente asumiendo que mantenemos constante al resto de variables).

Los comentarios anteriores son válidos pues no existe multicolinealidad entre las variables, en realidad existe multicolinealidad pero ésta no es grave (Consúltense los anexos).

En el siguiente cuadro, se muestra el resumen de las probabilidades de padecer desnutrición calculada a partir del modelo logístico encontrado. En él se aprecia que el número esperado de niños desnutridos es de 194.71, mientras que el verdadero número de niños desnutridos es de 203, es decir, con el modelo se comete un error de 8.29 niños desnutridos estimados. Por otra parte, el porcentaje de niños desnutridos observados es del 17.7%, y con el modelo logístico observamos una probabilidad media (porcentaje medio) de niños desnutridos igual al 16.99%, teniendo un error de aproximadamente el 0.71% en el porcentaje de niños desnutridos estimados.

Tabla 3.3.19: Cercanía del modelo logístico. Fase de validación.

Estadísticos		
Probabilidad de desnutrir		
N	Válidos	1146
	Perdidos	0
Media		,1699
Suma		194,71

3.4.5 APLICACIÓN ILUSTRATIVA DEL MODELO LOGÍSTICO.

A continuación se muestran los cuadros del número de niños desnutridos y no desnutridos observados en comparación con la clasificación realizada con el modelo logístico encontrado, usando un punto de corte igual a 0.085 en cada una de las zonas del país identificadas por el Ministerio de Salud Pública, se utilizan

todos los datos (datos usados para validación y para estimación del modelo logístico).

Tabla 3.3.20: Aplicación ilustrativa del modelo logístico en zonas del país.

			Clasificación con el modelo de Regresión Logística		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	171	641	812
		% dentro de Índice de Nutrición Combinado	21,1%	78,9%	100,0%
	Desnutrido	Recuento	16	190	206
		% dentro de Índice de Nutrición Combinado	7,8%	92,2%	100,0%
Total	Recuento	187	831	1018	
	% dentro de Índice de Nutrición Combinado	18,4%	81,6%	100,0%	

			Clasificación con el modelo de Regresión Logística		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	104	370	474
		% dentro de Índice de Nutrición Combinado	21,9%	78,1%	100,0%
	Desnutrido	Recuento	4	109	113
		% dentro de Índice de Nutrición Combinado	3,5%	96,5%	100,0%
Total	Recuento	108	479	587	
	% dentro de Índice de Nutrición Combinado	18,4%	81,6%	100,0%	

A) Occidente

B) Central

			Clasificación con el modelo de Regresión Logística		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	371	439	810
		% dentro de Índice de Nutrición Combinado	45,8%	54,2%	100,0%
	Desnutrido	Recuento	27	97	124
		% dentro de Índice de Nutrición Combinado	21,8%	78,2%	100,0%
Total	Recuento	398	536	934	
	% dentro de Índice de Nutrición Combinado	42,8%	57,4%	100,0%	

			Clasificación con el modelo de Regresión Logística		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	144	670	814
		% dentro de Índice de Nutrición Combinado	17,7%	82,3%	100,0%
	Desnutrido	Recuento	11	161	172
		% dentro de Índice de Nutrición Combinado	6,4%	93,6%	100,0%
Total	Recuento	155	831	986	
	% dentro de Índice de Nutrición Combinado	15,7%	84,3%	100,0%	

C) Metropolitana

D) Paracentral

			Clasificación con el modelo de Regresión Logística		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	140	760	900
		% dentro de Índice de Nutrición Combinado	15,6%	84,4%	100,0%
	Desnutrido	Recuento	6	193	199
		% dentro de Índice de Nutrición Combinado	3,0%	97,0%	100,0%
Total	Recuento	146	953	1099	
	% dentro de Índice de Nutrición Combinado	13,3%	86,7%	100,0%	

E) Oriental

3.4 APLICACIÓN DEL ANÁLISIS DISCRIMINANTE

Para la selección de variables se han utilizado los mismos criterios mencionados en la regresión logística; es decir: tasa de no respuesta, variables medidas en escala numérica, relación con la variable dependiente “índice de nutrición combinado”, multicolinealidad. En base a eso las variables más relevantes en el Análisis Discriminante se muestran en el siguiente cuadro.

Tabla 3.4.1: Variables finales en el Análisis Discriminante. Fase de estimación

Nombre	Significado
Gindice	Es la variable dependiente con la cual se identifica el estado nutricional del niño: vale 1 si es desnutrido y 0 en caso contrario.
Edupromef	Representa el número de años completos de estudios aprobados por la madre del niño.
P324A	Representa el número de hijos que viven en el mismo hogar del niño en estudio.
ServBasicos	Representa el número de servicios básicos con que cuenta la familia.
medad	Es la edad en años cumplidos de la madre al momento de dar a luz al niño en estudio.

3.4.1 ESTIMACIÓN FUNCIÓN LINEAL DISCRIMINANTE.

En el cuadro que se presenta a continuación se muestran las variables que según el Análisis Discriminante son útiles para ayudar a determinar el estado nutricional de los niños. Se ha utilizado el criterio de inclusión de variables por pasos como criterio de entrada y el método lambda de Wilks. En dicho cuadro se aprecia que la primera variable en considerarse es la educación de la madre; mientras que la última en ingresar es la del nivel socioeconómico medio de la familia.

Tabla 3.4.2: Orden de inclusión de variables, Análisis Discriminante. Fase de estimación.

Variables introducidas/excluidas^{a, b, c, d}

Paso	Introducidas	Lambda de Wilks							
		Estadístico	gl1	gl2	gl3	F exacta			
						Estadístico	gl1	gl2	Sig.
1	Servicios básicos	,944	1	1	3476,000	206,747	1	3476,000	,000
2	P324. A) Total hijos que viven en casa	,930	2	1	3476,000	131,627	2	3475,000	,000
3	Educación promedio (años)	,927	3	1	3476,000	91,128	3	3474,000	,000
4	edad de madre al nac (anos)	,926	4	1	3476,000	69,812	4	3473,000	,000

En cada paso se introduce la variable que minimiza la lambda de Wilks global.

- a. El número máximo de pasos es 8.
- b. La F parcial mínima para entrar es 3.84.
- c. La F parcial máxima para salir es 2.71
- d. El nivel de F, la tolerancia o el VIN son insuficientes para continuar los cálculos.

Se aprecia que con el Análisis Discriminante, las variables más importantes para determinar el estado nutricional de los niños corresponde a: el número de servicios básicos con que cuenta la familia, el número de hijos que también viven en el mismo hogar (tamaño de la familia), a la educación de la madre, y a la edad en la cual dio a luz la madre al niño estudiado.

El Análisis Discriminante se ha utilizado con probabilidades previas de pertenencia a cada grupo iguales (en cada estado nutricional), las cuales se muestran en el siguiente cuadro, pues de lo contrario el cuadro resumen de clasificación es pésimo; es decir, se obtienen mejores resultados asumiendo que un niño puede o no ser desnutrido, lo cual tiene lógica pues a la hora de estimar el estado nutricional únicamente contaremos con información a priori de las variables consideradas solamente, por lo que tiene sentido asignarles igual probabilidad.

En el cuadro que se muestra a continuación se presenta el autovalor asociado a la única función discriminante calculada, y nos da una idea de la importancia de esta función. El cuadrado de la correlación canónica (0.273), nos indica el grado de explicación de la varianza total de la variable dependiente y cuyo valor es de apenas el 7.45%; indicando que el 92.55% de la variabilidad no es explicada por la función discriminante.

Tabla 3.4.3: Coeficientes funciones lineales discriminantes. Fase de estimación.

Autovalores				
Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	,080 ^a	100,0	100,0	,273

a. Se han empleado las 1 primeras funciones discriminantes canónicas en el análisis.

En el cuadro que se presenta a continuación se muestra la prueba M de Box, con la cual contrastamos la hipótesis de igualdad de matrices de covarianzas en ambos estados nutricionales. En dicho cuadro podemos apreciar que a la luz de la evidencia muestral, los datos nos sugieren que las matrices de covarianzas son distintas, y que por consiguiente sería recomendable una discriminación cuadrática, sin embargo, continuaremos con la función discriminante lineal y luego compararemos la diferencia entre ambas funciones (recordemos que esta prueba es muy sensible al tamaño de la muestra y a desviaciones de la hipótesis de normalidad).

Tabla 3.4.4: Coeficientes funciones lineales discriminantes. Fase de estimación.

M de Box		73,717
F	Aprox.	7,354
	gl1	10
	gl2	5362113,413
	Sig.	,000

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

Las funciones de clasificación con el Análisis Discriminante son las siguientes (obtenidas del cuadro 3.3.5):

Para los niños desnutridos:

$$(G_{\text{indice}} = 1): 0.375(\text{ServBásicos}) + 0.039(P324A) + 0.175(\text{edupro m ef}) + 0.638(\text{medad}) - 9.979$$

(3.6)

Para los niños normales:

$$(G_{\text{indice}} = 0): 0.577(\text{ServBásicos}) - 0.216(P324A) + 0.212(\text{edupro m ef}) + 0.660(\text{medad}) - 10.801$$

(3.7)

Tabla 3.4.5: Coeficientes funciones lineales discriminantes de Fisher y de la función canónica discriminante. Fase de estimación.

Coeficientes de la función de clasificación			Coeficientes de las funciones canónicas discriminantes	
	Índice de Nutrición ...		Función	
	Normal	Desnutrido	1	
Servicios básicos	,577	,375	Servicios básicos	,271
Educación promedio (años)	,212	,175	Educación promedio (años)	,051
P324. A) Total hijos que viven en casa	-,216	,039	P324. A) Total hijos que viven en casa	-,343
edad de madre al nac (años)	,660	,638	edad de madre al nac (años)	,029
(Constante)	-10,801	-9,979	(Constante)	-1,345

Funciones discriminantes lineales de Fisher

Coefficientes no tipificados

De este modo la función canónica discriminante es:

$$L(\text{ServBásicos}; P324A; \text{edupro m ef}) = 0.271(\text{ServBásicos}) + 0.051(P324A) - 0.343(\text{edupro m ef}) + 0.029(\text{medad}) - 1.345 \quad (3.8)$$

En el siguiente cuadro se muestra la matriz estructura, en ella podemos apreciar el orden de importancia de las variables en el Análisis Discriminante (correlación entre las variables y la función discriminante). Se puede apreciar que la variable más importante es el número de servicios básicos con que cuenta la familia, seguido del número de años promedio de la madre, el número de hijos que viven en el mismo hogar, y finalmente la edad de la madre al momento del nacimiento del niño entrevistado.

Tabla 3.4.6: Matriz de estructura. Fase de estimación.

Matriz de estructura

	Función
	1
Servicios básicos	,860
Educación promedio (años)	,747
P324. A) Total hijos que viven en casa	-,634
edad de madre al nac (anos)	-,129

Correlaciones intra-grupo combinadas entre las variables discriminantes y las funciones discriminantes canónicas tipificadas
Variables ordenadas por el tamaño de la correlación con la función.

En el siguiente cuadro se muestra la tabla resumen del Análisis Discriminante; puede apreciarse que únicamente se clasifica correctamente al 66.2% de los niños. En los niños no desnutridos se clasifican correctamente al 66.9%; mientras que para los niños desnutridos se clasifican correctamente el 62.5%.

Tabla 3.4.7: Clasificación Análisis Discriminante. Fase de estimación.

Resultados de la clasificación^a

Indice de Nutrición Combinado			Grupo de pertenencia pronosticado		Total
			Normal	Desnutrido	
Original	Recuento	Normal	1919	948	2867
		Desnutrido	229	382	611
	%	Normal	66,9	33,1	100,0
		Desnutrido	37,5	62,5	100,0

a. Clasificados correctamente el 66,2% de los casos agrupados originales.

En el cuadro que se muestra a continuación se presenta la prueba de igualdad de medias para cada una de las variables consideradas en los dos grupos (dos estados nutricionales). En base a ella observamos que las medias son significativamente diferentes en ambos estados nutricionales.

Tabla 3.4.8: Prueba de igualdad de medias de variables significativas en el análisis discriminante. Fase de estimación.

Pruebas de igualdad de las medias de los grupos

	Lambda de Wilks	F	gl1	gl2	Sig.
Servicios básicos	,944	206,747	1	3476	,000
Educación promedio (años)	,957	156,005	1	3476	,000
P324. A) Total hijos que viven en casa	,969	112,459	1	3476	,000
edad de madre al nac (anos)	,999	4,662	1	3476	,031

A continuación se muestra el comportamiento promedio de cada una de las variables discriminantes según el estado nutricional de los niños. El número promedio de servicios básicos y el número promedio de años de educación de la madre es mayor en los niños no desnutridos que en los niños desnutridos. En promedio, hay un niño más viviendo en los hogares en donde han sido identificados niños como desnutridos que en los hogares en donde no han sido identificados.

Tabla 3.4.9: Medidas de resumen de variables consideradas en análisis discriminante según estado nutricional. Fase de estimación.

Indice de Nutrición Combinado		Estadísticos de grupo			
		Media	Desv. típ.	N válido (según lista)	
				No ponderados	Ponderados
Normal	Servicios básicos	4,4029	2,24383	2867	2867,000
	Educación promedio (años)	7,0115	4,45872	2867	2867,000
	P324. A) Total hijos que viven en casa	2,3561	1,47902	2867	2867,000
	edad de madre al nac (anos)	25,3122	6,25498	2867	2867,000
Desnutrido	Servicios básicos	2,9689	2,21115	611	611,000
	Educación promedio (años)	4,5712	4,01915	611	611,000
	P324. A) Total hijos que viven en casa	3,0851	1,81233	611	611,000
	edad de madre al nac (anos)	25,9247	6,86623	611	611,000
Total	Servicios básicos	4,1509	2,30340	3478	3478,000
	Educación promedio (años)	6,5828	4,48144	3478	3478,000
	P324. A) Total hijos que viven en casa	2,4842	1,56727	3478	3478,000
	edad de madre al nac (anos)	25,4198	6,36985	3478	3478,000

3.4.2 ESTIMACIÓN VALIDACIÓN CRUZADA.

Al utilizar la validación cruzada en el análisis discriminante se obtienen resultados bastante similares a los obtenidos anteriormente. En el siguiente cuadro se muestran las variables que resultan ser más relevantes al problema de clasificación del estado nutricional de los niños, se obtiene el mismo conjunto de variables discriminantes.

Tabla 3.4.10: Variables introducidas, validación cruzada. Fase de estimación.

Variables introducidas/excluidas^{a, b, c, d}

Paso	Introducidas	Lambda de Wilks							
		Estadístico	gl1	gl2	gl3	F exacta			
						Estadístico	gl1	gl2	Sig.
1	Servicios básicos	,944	1	1	3476,000	206,747	1	3476,000	,000
2	P324. A) Total hijos que viven en casa	,930	2	1	3476,000	131,627	2	3475,000	,000
3	Educación promedio (años)	,927	3	1	3476,000	91,128	3	3474,000	,000
4	edad de madre al nac (anos)	,926	4	1	3476,000	69,812	4	3473,000	,000

En cada paso se introduce la variable que minimiza la lambda de Wilks global.

- a. El número máximo de pasos es 8.
- b. La F parcial mínima para entrar es 3.84.
- c. La F parcial máxima para salir es 2.71
- d. El nivel de F, la tolerancia o el VIN son insuficientes para continuar los cálculos.

Las funciones lineales discriminantes no cambian pues en la tabla que se muestra a continuación se visualizan los coeficientes; dichos coeficientes no resultan ser diferentes a los encontrados previamente.

Tabla 3.4.11: Funciones lineales discriminante de Fisher y función canónica discriminante, validación cruzada. Fase de estimación.

Coeficientes de la función de clasificación			Coeficientes de las funciones canónicas discriminantes	
	Indice de Nutrición ...			Función
	Normal	Desnutrido		1
Servicios básicos	,577	,375	Servicios básicos	,271
Educación promedio (años)	,212	,175	Educación promedio (años)	,051
P324. A) Total hijos que viven en casa	-,216	,039	P324. A) Total hijos que viven en casa	-,343
edad de madre al nac (anos)	,660	,638	edad de madre al nac (anos)	,029
(Constante)	-10,801	-9,979	(Constante)	-1,345

Funciones discriminantes lineales de Fisher

Coefficientes no tipificados

Con la validación cruzada no mejoran mucho los resultados de clasificación, puesto que las funciones discriminantes son aproximadamente las mismas. En el cuadro se puede apreciar que se pierde un 0.5% de clasificación correcta para los niños desnutridos, el porcentaje global se reduce en apenas un 0.1%. Claramente no se aprecian diferencias marcadas

Tabla 3.4.12: Clasificación análisis discriminante, validación cruzada. Fase de estimación.

Resultados de la clasificación^{b, c}

		Índice de Nutrición Combinado	Grupo de pertenencia pronosticado		Total
			Normal	Desnutrido	
Original	Recuento	Normal	1919	948	2867
		Desnutrido	229	382	611
	%	Normal	66,9	33,1	100,0
		Desnutrido	37,5	62,5	100,0
Validación cruzada ^a	Recuento	Normal	1919	948	2867
		Desnutrido	232	379	611
	%	Normal	66,9	33,1	100,0
		Desnutrido	38,0	62,0	100,0

a. La validación cruzada sólo se aplica a los casos del análisis. En la validación cruzada, cada caso se clasifica mediante las funciones derivadas a partir del resto de los casos.

b. Clasificados correctamente el 66,2% de los casos agrupados originales.

c. Clasificados correctamente el 66,1% de los casos agrupados validados mediante validación cruzada.

3.4.3 ESTIMACIÓN FUNCIÓN CUADRÁTICA DISCRIMINANTE.

En el siguiente cuadro se muestran las variables que resultan ser relevantes en el Análisis Discriminante, se observa que se tiene el mismo conjunto de variables que obtuvimos anteriormente.

Tabla 3.4.13: Variables introducidas, función cuadrática discriminante. Fase de estimación

Variables introducidas/excluidas^{a, b, c, d}

Paso	Introducidas	Lambda de Wilks							
		Estadístico	gl1	gl2	gl3	F exacta			
						Estadístico	gl1	gl2	Sig.
1	Servicios básicos	,944	1	1	3476,000	206,747	1	3476,000	,000
2	P324. A) Total hijos que viven en casa	,930	2	1	3476,000	131,627	2	3475,000	,000
3	Educación promedio (años)	,927	3	1	3476,000	91,128	3	3474,000	,000
4	edad de madre al nac (anos)	,926	4	1	3476,000	69,812	4	3473,000	,000

En cada paso se introduce la variable que minimiza la lambda de Wilks global.

- a. El número máximo de pasos es 8.
- b. La F parcial mínima para entrar es 3.84.
- c. La F parcial máxima para salir es 2.71
- d. El nivel de F, la tolerancia o el VIN son insuficientes para continuar los cálculos.

En los cuadros que se muestran enseguida, se presentan los resultados de la prueba M de Box, con la cual se contrasta la hipótesis de igualdad de matriz de covarianzas de las variables discriminantes en los niños desnutridos y no desnutridos; puesto que la significancia es alta se concluye que las matrices de covarianzas pueden considerarse las mismas en ambos estados nutricionales.

Tabla 3.4.14: Prueba de igualdad matriz de covarianzas, funciones canónicas discriminantes. Fase de estimación

Logaritmo de los determinantes			Resultados de la prueba		
Indice de Nutrición Combinado	Rango	Logaritmo del determinante	M de Box	1,510	
Normal	1	-,014	F Aprox.	1,509	
Desnutrido	1	,063	gl1	1	
(matriz identidad)	1	,000	gl2	9336236,996	
			Sig.	,219	

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos de las funciones canónicas discriminantes.

Contrasta la hipótesis nula de que las matrices de covarianzas pertenecen a poblaciones iguales para las funciones canónicas discriminantes.

Por todo lo que hemos encontrado anteriormente concluimos que el análisis discriminante lineal podría ser el más adecuado, sin embargo, como ya lo mencionamos solo es adecuado para clasificar a los niños no desnutridos. En el siguiente cuadro se muestra los resultados de clasificación con la discriminación cuadrática. En él podemos apreciar que no existen grandes diferencias a la discriminación lineal, y que por consiguiente se tomará una función lineal para tratar de estimar el estado nutricional de la niñez salvadoreña.

Tabla 3.4.15: Clasificación análisis discriminante cuadrático. Fase de estimación.

Resultados de la clasificación ^a					
Indice de Nutrición Combinado			Grupo de pertenencia pronosticado		Total
			Normal	Desnutrido	
Original	Recuento	Normal	1970	897	2867
		Desnutrido	242	369	611
	%	Normal	68,7	31,3	100,0
		Desnutrido	39,6	60,4	100,0

a. Clasificados correctamente el 67,3% de los casos agrupados originales.

3.4.4 VALIDACIÓN ANÁLISIS DISCRIMINANTE.

En base a los comentarios previos, para validar el modelo encontrado utilizaremos la discriminación lineal (asumiendo igualdad de matriz de covarianzas en ambos estados nutricionales); pues parece ser la más adecuada y sencilla a nuestro conjunto de datos. Utilizamos dicha función en la parte de datos de la muestra que almacenamos para llevar a cabo tal fin, solamente para recordar las funciones lineales discriminantes de Fisher son:

Para los niños desnutridos:

$$(G_{\text{indice}} = 1): 0.375(\text{ServBásicos}) + 0.039(P324A) + 0.175(\text{edupro m ef}) + 0.638(\text{edad}) - 9.979$$

(3.6)

Para los niños normales:

$$(G_{\text{indice}} = 0): 0.577(\text{ServBásicos}) - 0.216(P324A) + 0.212(\text{edupro m ef}) + 0.660(\text{edad}) - 10.801$$

(3.7)

Calculamos ambas funciones discriminantes de Fisher en cada uno de los elementos de la muestra y utilizamos la siguiente regla de asignación:

- Si se cumple $(G_{\text{indice}} = 1) \geq (G_{\text{indice}} = 0)$, clasificamos al niño como desnutrido.
- Si por el contrario $(G_{\text{indice}} = 1) < (G_{\text{indice}} = 0)$, clasificamos al niño como no desnutrido.

Puesto que los coeficientes de ambas funciones son muy parecidos, en el SPSS no se logra apreciar la diferencia en cuanto a clasificación se refiere; es por eso que se utilizará la función canónica discriminante siguiente:

$$L(\text{ServBasi cos}; P324A; \text{edupro m ef}) = 0.271(\text{ServBasi cos}) + 0.051(P324A) - 0.343(\text{edupro m ef}) + 0.029(\text{medad}) - 1.345$$

La regla de clasificación se convierte entonces en:

- Si se cumple $L(\text{ServBasi cos}; P324A; \text{edupro m ef}) < 0$, clasificamos al niño como desnutrido.
- Si por el contrario $L(\text{ServBasi cos}; P324A; \text{edupro m ef}) \geq 0$, clasificamos al niño como no desnutrido.

Siguiendo dicha regla de clasificación y de generación, obtenemos el siguiente cuadro resumen del modelo.

Tabla 3.4.16: Clasificación del estado nutricional, Análisis Discriminante. Fase de validación.

Tabla de contingencia Índice de Nutrición Combinado * Clasificación con el Análisis Discriminante

			Clasificación con el Análisis Discriminante		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	588	355	943
		% dentro de Índice de Nutrición Combinado	62,4%	37,6%	100,0%
	Desnutrido	Recuento	64	139	203
		% dentro de Índice de Nutrición Combinado	31,5%	68,5%	100,0%
Total		Recuento	652	494	1146
		% dentro de Índice de Nutrición Combinado	56,9%	43,1%	100,0%

Puede apreciarse que las funciones encontradas tienen aproximadamente el mismo poder de clasificación tanto en niños desnutridos como no desnutridos, pues los porcentajes de clasificación correcta son más o menos parecidos. En términos generales, con el análisis discriminante lineal se clasifican correctamente al 63.4% de los niños.

3.4.5 APLICACIÓN ILUSTRATIVA DEL ANÁLISIS DISCRIMINANTE.

En los siguientes cuadros se muestran los cuadros del número de niños desnutridos y no desnutridos observados en comparación con la clasificación realizada con el análisis discriminante usándolo en cada una de las zonas del país identificadas por el Ministerio de Salud Pública, se utilizan todos los datos (datos usados para validación y para estimación).

Tabla 3.4.17: Aplicación ilustrativa del análisis discriminante en zonas del país.

Tabla de contingencia Índice de Nutrición Combinado * Clasificación con el Análisis Discriminante

			Clasificación con el Análisis Discriminante		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	417	395	812
		% dentro de Índice de Nutrición Combinado	51,4%	48,6%	100,0%
	Desnutrido	Recuento	60	146	206
		% dentro de Índice de Nutrición Combinado	29,1%	70,9%	100,0%
Total	Recuento	477	541	1018	
	% dentro de Índice de Nutrición Combinado	46,9%	53,1%	100,0%	

A) Occidente

Tabla de contingencia Índice de Nutrición Combinado * Clasificación con el Análisis Discriminante

			Clasificación con el Análisis Discriminante		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	284	190	474
		% dentro de Índice de Nutrición Combinado	59,9%	40,1%	100,0%
	Desnutrido	Recuento	27	86	113
		% dentro de Índice de Nutrición Combinado	23,9%	76,1%	100,0%
Total	Recuento	311	276	587	
	% dentro de Índice de Nutrición Combinado	53,0%	47,0%	100,0%	

B) Central

Tabla de contingencia Índice de Nutrición Combinado * Clasificación con el Análisis Discriminante

			Clasificación con el Análisis Discriminante		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	707	103	810
		% dentro de Índice de Nutrición Combinado	87,3%	12,7%	100,0%
	Desnutrido	Recuento	76	48	124
		% dentro de Índice de Nutrición Combinado	61,3%	38,7%	100,0%
Total	Recuento	783	151	934	
	% dentro de Índice de Nutrición Combinado	83,8%	16,2%	100,0%	

C) Metropolitana

Tabla de contingencia Índice de Nutrición Combinado * Clasificación con el Análisis Discriminante

			Clasificación con el Análisis Discriminante		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	423	391	814
		% dentro de Índice de Nutrición Combinado	52,0%	48,0%	100,0%
	Desnutrido	Recuento	43	129	172
		% dentro de Índice de Nutrición Combinado	25,0%	75,0%	100,0%
Total	Recuento	466	520	986	
	% dentro de Índice de Nutrición Combinado	47,3%	52,7%	100,0%	

D) Paracentral

Tabla de contingencia Índice de Nutrición Combinado * Clasificación con el Análisis Discriminante

			Clasificación con el Análisis Discriminante		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	438	462	900
		% dentro de Índice de Nutrición Combinado	48,7%	51,3%	100,0%
	Desnutrido	Recuento	38	161	199
		% dentro de Índice de Nutrición Combinado	19,1%	80,9%	100,0%
Total		Recuento	476	623	1099
		% dentro de Índice de Nutrición Combinado	43,3%	56,7%	100,0%

E) Oriental

3.5 COMPARACIÓN DE LOS MÉTODOS CONSIDERADOS.

En los dos cuadros que se presentan a continuación se muestran el resumen de clasificación para el modelo de regresión logística y para el análisis discriminante con el 25% de los datos destinados para la validación. En ambos cuadros podemos apreciar que, con el modelo logístico sobreestimamos mucho más el número de niños desnutridos que con el discriminante, más sin embargo, el porcentaje de niños desnutridos clasificados incorrectamente con el modelo logístico es casi la tercera parte de los clasificados con discriminante. Mientras que el porcentaje de niños no desnutridos clasificados incorrectamente en el modelo logístico es casi el doble al cometido en análisis discriminante.

Tabla 3.5.1: Resumen de clasificación Regresión Logística.

Tabla de contingencia Índice de Nutrición Combinado * Clasificación con el modelo de Regresión Logística

			Clasificación con el modelo de Regresión Logística		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	247	696	943
		% dentro de Índice de Nutrición Combinado	26,2%	73,8%	100,0%
	Desnutrido	Recuento	21	182	203
		% dentro de Índice de Nutrición Combinado	10,3%	89,7%	100,0%
Total		Recuento	268	878	1146
		% dentro de Índice de Nutrición Combinado	23,4%	76,6%	100,0%

Tabla 3.5.2: Resumen de clasificación Análisis Discriminante.

Tabla de contingencia Índice de Nutrición Combinado * Clasificación con el Análisis Discriminante

			Clasificación con el Análisis Discriminante		Total
			Normal	Desnutrido	
Índice de Nutrición Combinado	Normal	Recuento	588	355	943
		% dentro de Índice de Nutrición Combinado	62,4%	37,6%	100,0%
	Desnutrido	Recuento	64	139	203
		% dentro de Índice de Nutrición Combinado	31,5%	68,5%	100,0%
Total	Recuento		652	494	1146
	% dentro de Índice de Nutrición Combinado		56,9%	43,1%	100,0%

Como se comentó antes, nuestro modelo está pensado para identificar a niños desnutridos y ver qué es lo que se puede hacer con ellos. Teniendo en cuenta esto y en base a los resultados mostrados previamente podemos decir que la regresión logística parece ser la técnica de clasificación más adecuada para estimar el estado nutricional de la niñez salvadoreña en comparación con el Análisis Discriminante, al menos con las variables que disponemos.

3.6 ESTIMACIÓN DE LOS MODELOS USANDO TODOS LOS DATOS.

En esta sección se ha considerado conveniente realizar el ajuste del modelo logístico y del análisis discriminante a todo el conjunto de datos, para luego observar si existen diferencias cuando utilizábamos únicamente el 75% de los datos.

3.6.1 MODELO DE REGRESIÓN LOGÍSTICO CON TODOS LOS DATOS

En el cuadro que se presenta a continuación se muestra la estimación de parámetros del modelo logístico cuando se utilizan todos los datos. En comparación con los resultados observados cuando utilizamos el 75% de los datos podemos decir que: el valor de la constante pasa de ser -0.893 a -0.82, observando una pequeña diferencia de 0.073. En cuanto al parámetro estimado para los

servicios básicos, pasamos de tener -0.197 a -0.206, es decir, un error de apenas 0.01. Para la variable número de hijos, pasamos de tener 0.142 a 0.129, una diferencia de estimación de 0.013. Mientras que para la variable Educación promedio de la madre, se pasa de tener un valor estimado de -0.054 a -0.052, es decir, un error de estimación de apenas 0.002.

Cuadro 3.6.1. Estimación de los parámetros del modelo de regresión logístico con el 100% de los datos.

Historial de iteraciones^{a, b, c, d}

Iteración		-2 log de la verosimilitud	Coeficientes			
			Constant	ServBasicos	P324A	edupromef
Paso 1	1	4138,586	-,624	-,160		
	2	4031,357	-,520	-,260		
	3	4027,796	-,481	-,285		
	4	4027,789	-,479	-,287		
	5	4027,789	-,479	-,287		
Paso 2	1	4098,912	-,982	-,140	,111	
	2	3985,437	-1,038	-,230	,153	
	3	3981,603	-1,022	-,254	,160	
	4	3981,595	-1,021	-,255	,160	
	5	3981,595	-1,021	-,255	,160	
Paso 3	1	4089,672	-,901	-,117	,097	-,022
	2	3969,001	-,874	-,188	,127	-,043
	3	3964,079	-,824	-,206	,129	-,051
	4	3964,064	-,820	-,206	,129	-,052
	5	3964,064	-,820	-,206	,129	-,052

a. Método: Por pasos hacia adelante (Condicional)

b. En el modelo se incluye una constante.

c. -2 log de la verosimilitud inicial: 4303,393

d. La estimación ha finalizado en el número de iteración 5 porque las estimaciones de los parámetros han cambiado en menos de ,001.

El modelo logit de la probabilidad de sufrir desnutrición es;

$$\ln\left(\frac{P}{1-P}\right) = -0.820 - 0.206(\text{ServBasi cos}) + 0.129(P324A) - 0.052(\text{edupro m ef})$$

Por lo que el modelo de regresión logística toma la siguiente forma:

$$P(\text{Gindice} = 1) = \frac{1}{1 + \exp[0.820 + 0.206(\text{ServBasi cos}) - 0.129(P324A) + 0.052(\text{edupro m ef})]}$$

En el siguiente cuadro se muestra la Prueba de Hosmer y Lemeshow, la cual nos sirve para conocer la bondad de ajuste del modelo encontrado a los datos. En el cuadro podemos apreciar que el modelo parece ajustar bien a los datos, sin embargo, el ajuste no es tan bueno al observado cuando utilizamos el 75%, pues la significancia de la prueba en el tercer paso usando el 75% es aproximadamente 1.7 veces mayor cuando se utiliza el 100% de los datos; además la significancia con el 75% se incrementa en cada uno de los pasos (incorporar variables en el modelo), mientras que con el 100% se reduce del primer al tercer paso.

Cuadro 3.6.2. Prueba de bondad de ajuste de Hosmer y Lemeshow, modelo de regresión logístico con el 100% de los datos

Prueba de Hosmer y Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	6,170	7	,520
2	16,991	8	,030
3	8,552	8	,381

En el cuadro que se presenta enseguida se muestra las mismas medidas de resúmenes mencionados cuando se utilizó el 75% de los datos. Observemos que la lejanía del modelo cuando se utiliza el 100% de los datos es aproximadamente 1000 unidades mayor al del 75%; Sin embargo, la diferencia entre el R^2 de Cox y Snell y el R^2 de Nagelkerke son prácticamente despreciables.

Cuadro 3.6.3. Resumen del modelo de regresión logístico con el 100% de los datos.

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Coxy Snell	R cuadrado de Nagelkerke
1	4027,789 ^a	,058	,096
2	3981,595 ^a	,067	,111
3	3964,064 ^a	,071	,117

a. La estimación ha finalizado en el número de iteración 5 porque las estimaciones de los parámetros han cambiado en menos de ,001.

En el siguiente cuadro se muestra la matriz de correlaciones de las variables que han resultado ser significativas en el modelo de regresión logística cuando se utiliza el 100% de los datos. Cuando comparamos las matriz de correlación de las variables utilizando únicamente el 75% de los datos y el 100% de los datos, nos damos cuenta que la diferencia es mínima es prácticamente despreciable.

Cuadro 3.6.4. Matriz de correlaciones, variables significativas en el modelo logístico con el 100% de los datos.

Matriz de correlaciones

		Constant	ServBasicos	P324A	edupromef
Paso 1	Constant	1,000	-,827		
	ServBasicos	-,827	1,000		
Paso 2	Constant	1,000	-,711	-,749	
	ServBasicos	-,711	1,000	,238	
	P324A	-,749	,238	1,000	
Paso 3	Constant	1,000	-,353	-,773	-,399
	ServBasicos	-,353	1,000	,040	-,515
	edupromef	-,399	-,515	,296	1,000
	P324A	-,773	,040	1,000	,296

En el siguiente cuadro se muestra el resumen de clasificación del modelo logístico utilizando el 100% de los datos y tomando un punto de corte de 0.085 para la probabilidad de desnutrición. En el podemos apreciar que el porcentaje global de niños clasificados correctamente es de apenas el 35.4%, un porcentaje similar al obtenido con el 75% de los datos. Sin embargo, el porcentaje correcto de niños desnutridos clasificados correctamente es del 93.1%, 3.4% mayor al obtenido con el 75% de los datos; mientras que el porcentaje niños normales clasificados correctamente es del 23.1%, un 3.1% menos obtenido en el 75% de los datos.

Cuadro 3.6.5. Variables introducidas en el modelo de regresión logístico con toda la muestra.

Tabla de clasificación^a

Observado			Pronosticado		
			Indice de Nutrición Combinado		Porcentaje correcto
			Normal	Desnutrido	
Paso 1	Indice de Nutrición Combinado	Normal	766	3044	20,1
		Desnutrido	51	763	93,7
	Porcentaje global				33,1
Paso 2	Indice de Nutrición Combinado	Normal	855	2955	22,4
		Desnutrido	62	752	92,4
	Porcentaje global				34,8
Paso 3	Indice de Nutrición Combinado	Normal	881	2929	23,1
		Desnutrido	56	758	93,1
	Porcentaje global				35,4

a. El valor de corte es ,085

3.6.2 ANÁLISIS DISCRIMINANTE CON TODOS LOS DATOS

En el siguiente cuadro se muestra el orden de inclusión de las variables que resultan ser significativas en el modelo discriminante lineal al utilizar todos los datos de la muestra. Al comparar los resultados a los obtenidos cuando se utilizó únicamente el 75% de los datos, nos damos cuenta que los resultados son prácticamente los mismos; las variables se introducen en el mismo orden.

Cuadro 3.6.6. Variables introducidas en modelo discriminante lineal utilizando el 100% de los datos.

Variables introducidas/excluidas^{a,b,c,d}

Paso	Introducidas	Lambda de Wilks							
		Estadístico	gl1	gl2	gl3	F exacta			
						Estadístico	gl1	gl2	Sig.
1	Servicios básicos	,941	1	1	4622,000	288,964	1	4622,000	,000
2	P324. A) Total hijos que viven en casa	,929	2	1	4622,000	176,193	2	4621,000	,000
3	Educación promedio (años)	,927	3	1	4622,000	121,611	3	4620,000	,000
4	edad de madre al nac (anos)	,925	4	1	4622,000	93,407	4	4619,000	,000

En cada paso se introduce la variable que minimiza la lambda de Wilks global.

- a. El número máximo de pasos es 8.
- b. La F parcial mínima para entrar es 3.84.
- c. La F parcial máxima para salir es 2.71
- d. El nivel de F, la tolerancia o el VIN son insuficientes para continuar los cálculos.

Al igual que como se hizo con el 75% de los datos. Se considera conveniente realizar la prueba de igualdad de matriz de covarianzas en ambos estados nutricionales. Y tal como ocurrió en ese caso, rechazamos la hipótesis de igualdad de matrices pues la significancia de la Prueba M de Box es demasiado pequeña. Obtenemos las mismas conclusiones utilizando el 100% o el 75% de los datos de la muestra.

Cuadro 3.6.7. Prueba de igualdad de matriz de covarianzas en modelo discriminante lineal con el 100% de los datos.

Resultados de la prueba

M de Box		90,321
F	Aprox.	9,016
	gl1	10
	gl2	9530368,301
	Sig.	,000

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

En el siguiente cuadro se muestran los coeficientes de las funciones lineales discriminantes de Fisher al utilizar el 100% de los datos. En comparación de los resultados obtenidos con el 75% de los datos; podemos apreciar que para la variable Servicios básicos, las diferencias en las funciones de niños normales y desnutridos son respectivamente de 0.006 y 0.004. Para la variable Educación de la madre estas mismas diferencias son de 0.014 y 0.014, respectivamente. Para la variable Número de hijos, las diferencias son respectivamente de 0.06 y 0.045. Finalmente para la variable Edad de la madre al nacimiento, estas mismas diferencias son respectivamente de 0.007 y 0.008. En general, como puede apreciarse las diferencias son mínimas, casi despreciables.

Cuadro 3.6.8. Funciones lineales discriminantes con el 100% de los datos.

Coeficientes de la función de clasificación			Coeficientes de las funciones canónicas discriminantes	
	Índice de Nutrición ...			Función
	Normal	Desnutrido		1
Servicios básicos	,583	,371	Servicios básicos	,284
edad de madre al nac (anos)	,653	,630	edad de madre al nac (anos)	,031
Educación promedio (años)	,226	,189	Educación promedio (años)	,049
P324. A) Total hijos que viven en casa	-,156	,084	P324. A) Total hijos que viven en casa	-,322
(Constante)	-10,869	-9,938	(Constante)	-1,488

Funciones discriminantes lineales de Fisher

Coefficientes no tipificados

Las funciones lineales discriminantes son entonces;

Para los niños desnutridos:

$$(G_{\text{índice}} = 1): 0.371(\text{ServBásicos}) + 0.084(P324A) + 0.189(\text{edupro mef}) + 0.630(\text{medad}) - 9.938$$

Para los niños no desnutridos:

$$(G_{\text{indice}} = 0): 0.583(\text{ServBásicos}) - 0.156(P324A) + 0.226(\text{edupro m ef}) + 0.653(\text{medad}) - 10.869$$

En el siguiente cuadro se muestra el resumen de clasificación del análisis discriminante lineal utilizando el 100% de los datos. Puede apreciarse que los resultados son muy similares a los obtenidos al utilizar el 75% de los datos; el porcentaje global de clasificación correcta se aumenta apenas un 0.2% debido únicamente a que el porcentaje de niños normales clasificados correctamente aumenta en esa misma cantidad, mientras que el porcentaje de niños desnutridos clasificados correctamente se mantiene sin cambio.

Cuadro 3.6.9. Resumen de clasificación modelo lineal discriminante con el 100% de los datos.

Resultados de la clasificación^a

		Índice de Nutrición Combinado	Grupo de pertenencia pronosticado		Total
			Normal	Desnutrido	
Original	Recuento	Normal	2556	1254	3810
		Desnutrido	306	508	814
	%	Normal	67,1	32,9	100,0
		Desnutrido	37,6	62,4	100,0

a. Clasificados correctamente el 66,3% de los casos agrupados originales.

En el siguiente cuadro se muestra el resumen de clasificación con la función lineal discriminante utilizando el 100% de los datos pero utilizando también la validación cruzada. Puede observarse que los resultados son muy parecidos a los obtenidos con el 75% de los datos, el porcentaje global de clasificación correcta es el mismo en ambos casos. Mientras que los porcentajes de clasificación correcta para cada uno de los estados nutricionales son muy similares.

Cuadro 3.6.10. Resumen de clasificación, validación cruzada, modelo lineal discriminante con el 100% de los datos.

Resultados de la clasificación^{b, c}

Índice de Nutrición Combinado			Grupo de pertenencia pronosticado		Total
			Normal	Desnutrido	
Original	Recuento	Normal	2556	1254	3810
		Desnutrido	306	508	814
	%	Normal	67,1	32,9	100,0
		Desnutrido	37,6	62,4	100,0
Validación cruzada ^a	Recuento	Normal	2554	1256	3810
		Desnutrido	310	504	814
	%	Normal	67,0	33,0	100,0
		Desnutrido	38,1	61,9	100,0

a. La validación cruzada sólo se aplica a los casos del análisis. En la validación cruzada, cada caso se clasifica mediante las funciones derivadas a partir del resto de los casos.

b. Clasificados correctamente el 66,3% de los casos agrupados originales.

c. Clasificados correctamente el 66,1% de los casos agrupados validados mediante validación cruzada.

En el siguiente cuadro se muestra el orden de introducción de las variables significativas en el modelo de discriminación cuadrática utilizando el 100% de los datos. Podemos apreciar que el orden es el mismo cuando se utilizó el 75% de los datos por lo que no existe diferencia entre ambos tamaños de datos.

Cuadro 3.6.11. Variables introducidas en el modelo cuadrático discriminante utilizando el 100% de los datos.

Variables introducidas/excluidas^{a, b, c, d}

Paso	Introducidas	Lambda de Wilks							
		Estadístico	gl1	gl2	gl3	F exacta			
						Estadístico	gl1	gl2	Sig.
1	Servicios básicos	,941	1	1	4622,000	288,964	1	4622,000	,000
2	P324. A) Total hijos que viven en casa	,929	2	1	4622,000	176,193	2	4621,000	,000
3	Educación promedio (años)	,927	3	1	4622,000	121,611	3	4620,000	,000
4	edad de madre al nac (anos)	,925	4	1	4622,000	93,407	4	4619,000	,000

En cada paso se introduce la variable que minimiza la lambda de Wilks global.

a. El número máximo de pasos es 8.

b. La F parcial mínima para entrar es 3.84.

c. La F parcial máxima para salir es 2.71

d. El nivel de F, la tolerancia o el VIN son insuficientes para continuar los cálculos.

En el siguiente cuadro se muestra la prueba de igualdad de matrices de covarianzas para las funciones canónicas discriminantes al utilizar el 100% de los datos. En base a los resultados se puede concluir, siempre y cuando se trabaje con un nivel de confianza del 95%, que las matrices son iguales. Llegamos a la misma conclusión que cuando utilizábamos el 75%, sin embargo, la aceptación de la hipótesis era mucho más evidente.

Cuadro 3.6.12. Contraste de igualdad de matrices de funciones discriminantes utilizando el 100% de los datos.

Resultados de la prueba

M de Box		3,231
F	Aprox.	3,229
	gl1	1
	gl2	1,658E7
	Sig.	,072

Contrasta la hipótesis nula de que las matrices de covarianzas pertenecen a poblaciones iguales para las funciones canónicas discriminantes.

En el siguiente cuadro se muestra el resumen de clasificación del modelo cuadrático discriminante al utilizar el 100% de los datos. El porcentaje global de clasificación correcta se reduce un 0.2% en comparación cuando se utilizó el 75% de los datos, esta disminución debida principalmente a que el porcentaje de niños desnutridos clasificados correctamente se reduce en esa misma cantidad.

Cuadro 3.6.13. Resumen de clasificación, modelo cuadrático discriminante utilizando el 100% de los datos.

Resultados de la clasificación^a

Indice de Nutrición Combinado			Grupo de pertenencia pronosticado		Total
			Normal	Desnutrido	
Original	Recuento	Normal	2613	1197	3810
		Desnutrido	324	490	814
	%	Normal	68,6	31,4	100,0
		Desnutrido	39,8	60,2	100,0

a. Clasificados correctamente el 67,1% de los casos agrupados originales.

Al igual que cuando se utilizaba únicamente el 75% de los datos no parece existir mucha diferencia entre la discriminación lineal con la cuadrática, por lo que se puede considerar más conveniente una discriminación lineal.

CAPÍTULO IV
CONTRASTE DE
HIPÓTESIS.

4.1 INTRODUCCIÓN.

En este capítulo se procede a realizar los contrastes de las hipótesis que han sido planteadas en este trabajo de investigación. Consideraremos como variable dependiente el valor estandarizado del Peso para la Edad, Peso para la Talla y Talla para la Edad de los niños definidas previamente en el capítulo 2 sección 2.3.4.

En total son seis hipótesis que se contrastan en este capítulo. La teoría estadística de todos los contrastes que aquí se utilizan ha sido presentada en el primer capítulo. Primero se realiza el contraste de normalidad de cada una de las variables dependientes en los diferentes niveles de las variables de contraste, y como resulta que la normalidad falla, se utilizan contrastes no paramétricos sobre igualdad de distribuciones e igualdad de valores medianos (equivalentes a la prueba t y ANOVA).

Cada hipótesis es contrastada en una sección y al final de cada sección siempre se realizan comparación dos a dos entre los niveles de cada variable de contraste considerada, pues resulta que la hipótesis nula de igualdad de distribuciones se rechaza en todas.

4.2 COMPARACIÓN DEL ESTADO NUTRICIONAL SEGÚN SECTOR DE RESIDENCIA.

La hipótesis a contrastar es la siguiente:

- El comportamiento del estado nutricional de los niños y niñas menores de cinco años de edad es el mismo para los que viven en zonas rurales y urbanas del país.

4.1 CONTRASTE DE NORMALIDAD.

Antes que nada verificaremos la normalidad de cada una de las variables antropométricas en cada uno de los sectores de residencia. Si se cumple la hipótesis de normalidad en ambos sectores procedemos a realizar el contraste paramétrico de comparación de medias; en caso contrario realizamos su equivalente prueba no paramétrica.

Tabla 4.2.1: Contraste de normalidad, variables antropométricas. Prueba de Kolmogorov-Smirnov y Prueba de Shapiro Willk.

Pruebas de normalidad							
Sector de residencia		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
Valor Z de Peso correspondiente a la edad y sexo del niño	Rural	,071	2607	,000	,937	2607	,000
	Urbana	,066	2017	,000	,949	2017	,000
Valor Z de Peso correspondiente a la talla y sexo del niño	Rural	,073	2607	,000	,903	2607	,000
	Urbana	,090	2017	,000	,786	2017	,000
Valor Z de Talla correspondiente a la edad y sexo del niño	Rural	,035	2607	,000	,979	2607	,000
	Urbana	,038	2017	,000	,966	2017	,000

a. Corrección de la significación de Lilliefors

En base a los resultados del cuadro anterior podemos concluir que ninguna de las variables antropométricas sigue una distribución normal en cada uno de los sectores de residencia. Esto implica que debemos realizar una prueba no paramétrica para contrastar la hipótesis planteada.

4.2 IGUALDAD DE DISTRIBUCIONES.

La primera prueba que realizamos es la de Kolmogorov-Smirnov; con ella la hipótesis nula que estamos verificando es: *“La distribución de cada variable antropométrica es la misma en ambos sectores de residencia”*. En el cuadro que se muestra a continuación se muestran los resultados de realizar tal contraste.

Tabla 4.2.2: Contraste de igualdad de distribuciones, variables antropométricas.

Prueba Kolmogorov-Smirnov.

Estadísticos de contraste^a

		Valor Z de Peso correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la talla y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño
Diferencias más extremas	Absoluta	,115	,055	,123
	Positiva	,115	,055	,123
	Negativa	,000	-,001	-,001
Z de Kolmogorov-Smirnov		3,887	1,862	4,140
Sig. asintót. (bilateral)		,000	,002	,000

a. Variable de agrupación: Sector de residencia

De los resultados mostrados en el cuadro anterior podemos apreciar que las variables antropométricas se comportan de manera distinta en ambos sectores de residencia; indicándonos de este modo que el estado nutricional de los niños menores de cinco años se comporta de manera diferente en ambos sectores de residencia.

4.3 PRUEBA DE WILCONXON-MAN-WHITNEY.

La prueba que desarrollamos a continuación es equivalente a la anterior, en el sentido que se puede utilizar para verificar la hipótesis de igualdad de distribuciones, sin embargo, la prueba es equivalente a la prueba de comparación de medias cuando falla la hipótesis de normalidad dicha prueba es la de Wilconxon-Mann-Whitney; y los resultados de su aplicación se muestran en el siguiente cuadro.

Claramente se llega a rechazar la hipótesis de igualdad de medias, con lo que se obtienen conclusiones similares a los de la prueba anterior.

Tabla 4.2.3: Prueba no paramétrica de comparación de medias, variables antropométricas. Prueba U de Mann-Whitney y W de Wilcoxon.

Estadísticos de contraste ^a			
	Valor Z de Peso correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la talla y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño
U de Mann-Whitney	2251426,500	2471868,500	2193116,500
W de Wilcoxon	5650954,500	5871396,500	5592644,500
Z	-8,391	-3,494	-9,686
Sig. asintót. (bilateral)	,000	,000	,000

a. Variable de agrupación: Sector de residencia

4.3 COMPARACIÓN DEL ESTADO NUTRICIONAL SEGÚN NIVEL EDUCATIVO DE LA MADRE.

La hipótesis a contrastar es la siguiente:

- El estado nutricional de los niños y niñas menores de cinco años de edad no es el mismo para los diferentes niveles educativos de sus respectivas madres.

4.3.1 CONTRASTE DE NORMALIDAD.

Al igual que se hizo para contrastar la primera hipótesis planteada en esta investigación, necesitaremos verificar que la variable de interés, en nuestro caso, las variables antropométricas se comportan de manera normal o aproximadamente normal en los distintos niveles educativos considerados de la madre. Antes de continuar merece la pena realizar el siguiente comentario; las madres identificadas con un nivel Básico de estudios corresponden a madres que a lo sumo lograron aprobar el nivel Básico, por lo que no necesariamente implica madres con un nivel mínimo de educación Básica. En cambio las madres con un nivel de educación

Superior efectivamente corresponden a madres que lograron culminar sus estudios universitarios. Mientras que las madres con un nivel de Bachillerato corresponden a madres que culminaron la educación Básica, iniciaron sus estudios de Bachillerato pero que no necesariamente lograron terminarlos (algunas inclusive iniciaron estudios universitarios pero no los completaron).

En caso de aceptarse la normalidad se procederá a realizar un Análisis de Varianza, de lo contrario se utilizará su equivalente no paramétrico, la prueba de Kruskal - Wallis.

Tabla 4.3.1: Contraste de normalidad, variables antropométricas. Prueba de Kolmogorov-Smirnov y Prueba de Shapiro Wilk.

Pruebas de normalidad							
Nivel educativo de la madre		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
Valor Z de Peso correspondiente a la edad y sexo del niño	Ninguno	,050	531	,003	,963	531	,000
	Básica (1 - 9)	,074	2935	,000	,934	2935	,000
	Bachillerato	,079	1037	,000	,939	1037	,000
	Superior	,064	121	,200*	,981	121	,093
Valor Z de Peso correspondiente a la talla y sexo del niño	Ninguno	,058	531	,000	,970	531	,000
	Básica (1 - 9)	,090	2935	,000	,796	2935	,000
	Bachillerato	,072	1037	,000	,933	1037	,000
	Superior	,115	121	,000	,841	121	,000
Valor Z de Talla correspondiente a la edad y sexo del niño	Ninguno	,043	531	,022	,966	531	,000
	Básica (1 - 9)	,040	2935	,000	,966	2935	,000
	Bachillerato	,045	1037	,000	,984	1037	,000
	Superior	,064	121	,200*	,965	121	,003

a. Corrección de la significación de Lilliefors

*. Este es un límite inferior de la significación verdadera.

Del cuadro anterior se aprecia que las variables antropométricas medidas en los niños no se comportan de manera normal en cada uno de los niveles educativos de sus respectivas madres.

4.3.2 PRUEBA DE KRUSKAL-WALLIS.

En vista que la prueba de normalidad ha sido rechazada, utilizaremos la prueba de Kruskal-Wallis, la cual es una generalización de la prueba de Wilconxon-Mann-Whitney al caso de más de dos poblaciones (ANOVA en poblaciones normales). La hipótesis que se contrasta con dicha prueba es la siguiente: *“La distribución de cada variable antropométrica de los niños menores de cinco años es idéntica para los distintos niveles educativos de la madre”*. Los resultados de dicha prueba se muestran a continuación.

Tabla 4.3.2: Prueba de igualdad de distribuciones, variables antropométricas.

Prueba de Kruskal-Wallis.

Estadísticos de contraste^{a, b}

	Valor Z de Peso correspondie nte a la edad y sexo del niñ o	Valor Z de Peso correspondie nte a la talla y sexo del niño	Valor Z de Talla correspondie nte a la edad y sexo del niñ o
Chi-cuadrado	211,115	54,677	238,425
gl	3	3	3
Sig. asintót.	,000	,000	,000

a. Prueba de Kruskal-Wallis

b. Variable de agrupación: Nivel educativo de la madre

De los resultados mostrados en el cuadro anterior, podemos concluir que el estado nutricional de los niños menores de cinco años se comporta de manera diferente en los distintos niveles educativos de su respectiva madre.

4.3.3 PRUEBA DE LA MEDIANA.

Se considera importante en esta etapa de la investigación realizar el contraste de la Mediana. Con dicha prueba la hipótesis a contrastar es la siguiente: *“Cada variable*

antropométrica tiene el mismo valor mediano en cada uno de los niveles educativos de la madre". Los resultados de dicha prueba se muestran en el siguiente cuadro.

Tabla 4.3.3: Prueba de igualdad de distribuciones, variables antropométricas.

Prueba de la Mediana.

Estadísticos de contraste^b

	Valor Z de Peso correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la talla y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño
N	4624	4624	4624
Mediana	-,3125	,4392	-,8873
Chi-cuadrado	142,494 ^a	31,529 ^a	158,896 ^a
gl	3	3	3
Sig. asintót.	,000	,000	,000

a. 0 casillas (,0%) tienen frecuencias esperadas menores que 5. La frecuencia de casilla esperada mínima es 60,5.

b. Variable de agrupación: Nivel educativo de la madre

De los resultados del cuadro anterior se concluye que cada variable antropométrica tiene diferentes valores medianos en cada uno de los niveles educativos de la madre, y por consiguiente el estado nutricional de los niños menores de cinco años es diferente.

4.3.4 COMPARACIÓN DOS A DOS.

En vista que a la luz de los resultados anteriores se observa que cada una de las variables antropométricas se comporta de manera diferente para los distintos niveles educativos de la madre, se considera pertinente realizar la prueba de Kolmogorov-Smirnov para determinar cuáles son los niveles educativos que difieren entre sí, es decir, una comparación dos a dos. Los resultados de realizar las comparaciones dos a dos para cada una de las variables antropométricas se muestran en el cuadro 4.3.4.

Tabla 4.3.4: Prueba de Kolmogorov-Smirnov de variables antropométricas.

Comparación en parejas del Nivel educativo de la madre.

			EDUCACIÓN MEDIA			BACHILLERATO			SUPERIOR		
			Valor Z de Peso correspondie nte a la edad y sexo del niño	Valor Z de Peso correspondie nte a la talla y sexo del niño	Valor Z de Talla correspondie nte a la edad y sexo del niño	Valor Z de Peso correspondie nte a la edad y sexo del niño	Valor Z de Peso correspondie nte a la talla y sexo del niño	Valor Z de Talla correspondie nte a la edad y sexo del niño	Valor Z de Peso correspondie nte a la edad y sexo del niño	Valor Z de Peso correspondie nte a la talla y sexo del niño	Valor Z de Talla correspondie nte a la edad y sexo del niño
NINGUN A	Diferencias más extremas	Absoluta	,172	,105	,173	,301	,152	,337	,367	,232	,326
		Positiva	,002	,001	,003	,001	,001	,004	,367	,232	,326
		Negativa	-,172	-,105	-,173	-,301	-,152	-,337	-,007	,000	-,008
	Z de Kolmogorov-Smirnov		3,655	2,232	3,674	5,650	2,842	6,322	3,642	2,302	3,241
	Sig. asintót. (bilateral)		,000	,000	,000	,000	,000	,000	,000	,000	,000
EDUCACI ÓN MEDIA	Diferencias más extremas	Absoluta				,156	,065	,181	,218	,133	,257
		Positiva				,156	,065	,181	,218	,133	,257
		Negativa				,000	-,001	-,001	-,011	-,007	-,007
	Z de Kolmogorov-Smirnov				4,319	1,810	5,001	2,355	1,437	2,766	
	Sig. asintót. (bilateral)				,000	,003	,000	,000	,032	,000	
BACHILL ERATO	Diferencias más extremas	Absoluta							,115	,097	,135
		Positiva							,115	,097	,135
		Negativa							-,030	-,022	-,065
	Z de Kolmogorov-Smirnov							1,198	1,012	1,408	
	Sig. asintót. (bilateral)							,113	,258	,038	

En dicha tabla podemos apreciar que el estado nutricional de los niños se comporta de manera idéntica para las madres con educación superior y las madres con bachillerato, pues para dos de los tres índices antropométricos se acepta la hipótesis de igualdad de distribuciones en ambos niveles educativos. Para el resto de parejas de niveles educativos que puede tomarse se observa mediante el test que la distribución del estado nutricional es significativamente diferente.

Es de mencionar que cuando se compara el nivel educativo de Ninguno con el resto de los niveles, la igualdad de distribuciones se rechaza debido a valores negativos extremos en la diferencias entre ambas distribuciones, esto nos estaría indicando que la distribución del nivel Ninguno en educación se encuentra por debajo de las distribuciones del resto de los niveles educativos. Cuando comparamos las madres con nivel de educación Medio con el resto de los niveles

educativos, el rechazo de la igualdad de las distribuciones se debe a valores positivos extremos en la diferencia de las distribuciones, indicando que la distribución del nivel educativo Medio se encuentra por debajo del resto de los niveles educativos. En resumen estos resultados podría interpretarse de la siguiente manera; a menor nivel educativo de la madre la distribución de cada uno de los índices antropométricos toma valores más pequeños y por consiguiente es más probable que sus respectivos hijos padezcan de desnutrición.

4.4 COMPARACIÓN DEL ESTADO NUTRICIONAL SEGÚN EDAD DE LA MADRE.

La hipótesis a contrastar es la siguiente:

- El comportamiento del estado nutricional de los niños y niñas menores de cinco años de edad no es el mismo para las madres jóvenes que las madres ya maduras (Edad cumplida).

Se ha considerado que los grupos de edad con la cual se someterá a prueba la hipótesis anteriormente son los siguientes: el primer grupo lo conforman madres con edades entre 15 y 19 años cumplidos (madres muy jóvenes); el segundo madres con edades entre 20 y 24 años cumplidos (madres con edad intermedia); mientras que el tercer grupo lo conforman madres con edades entre 30 y 49 años cumplidos (madres muy maduras).

4.4.1 CONTRASTE DE NORMALIDAD.

Al igual que se ha hecho anteriormente, lo primero que debemos hacer es realizar el contraste de normalidad de las variables antropométricas en cada uno de los

distintos grupos de edades considerados, esto para determinar si se debe realizar un Análisis de Varianza o su equivalente no paramétrico. Los resultados de llevar a cabo dicho contraste se muestran en el cuadro que se presenta a continuación.

Tabla 4.4.1: Contraste de normalidad, variables antropométricas. Prueba de Kolmogorov-Smirnov y Prueba de Shapiro Willk.

Pruebas de normalidad							
Edad actual (años)		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
Valor Z de Peso correspondiente a la edad y sexo del niño	15-19	,071	429	,000	,951	429	,000
	20-24	,061	1145	,000	,966	1145	,000
	25-49	,074	3050	,000	,932	3050	,000
Valor Z de Peso correspondiente a la talla y sexo del niño	15-19	,087	429	,000	,870	429	,000
	20-24	,068	1145	,000	,946	1145	,000
	25-49	,085	3050	,000	,811	3050	,000
Valor Z de Talla correspondiente a la edad y sexo del niño	15-19	,032	429	,200 [*]	,978	429	,000
	20-24	,042	1145	,000	,969	1145	,000
	25-49	,033	3050	,000	,975	3050	,000

a. Corrección de la significación de Lilliefors

*. Este es un límite inferior de la significación verdadera.

Del cuadro anterior, podemos apreciar que las variables no se distribuyen de manera aproximadamente normal en cada uno de los grupos de edades, por lo que al igual que como se hizo en la hipótesis anterior se realizará el contraste de Kruskal-Wallis.

4.4.2 PRUEBA DE KRUSKAL-WALLIS.

La hipótesis que realmente estamos contrastando es la siguiente: *“La distribución de cada variable antropométrica de los niños menores de cinco años es idéntica para la distintas edades de la madre”*. Los resultados de dicha prueba se muestran a continuación. En base a ellos podemos concluir que no existe evidencia significativa o estadística que nos permita decir que el estado nutricional de los niños se comporte de manera diferente en cada uno de los grupos de edad

considerados, nótese que para uno de los tres indicadores antropométricos se acepta dicha hipótesis siempre y cuando el nivel de significancia sea menor al 7%.

Tabla 4.4.2: Prueba de igualdad de distribuciones, variables antropométricas.

Prueba de Kruskal-Wallis.

Estadísticos de contraste^{a, b}

	Valor Z de Peso correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la talla y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño
Chi-cuadrado	3,690	1,964	5,405
gl	2	2	2
Sig. asintót.	,158	,375	,067

a. Prueba de Kruskal-Wallis

b. Variable de agrupación: Edad actual (años)

4.4.3 PRUEBA DE LA MEDIANA.

Al igual que antes hemos considerado conveniente realizar la Prueba de la Mediana para contrastar la hipótesis siguiente: *“El valor mediano de cada variable antropométrica es el mismo en cada uno de los grupos de edades de la madre”*. Los resultados se muestran en el siguiente cuadro. Se observa que con excepción de la variable antropométrica Talla para la Edad podemos concluir que el estado nutricional de los niños es el mismo (hablando de valores medianos) para cada uno de los distintos grupos de edad considerados de la madre.

Tabla 4.4.3: Prueba de igualdad de distribuciones, variables antropométricas.

Prueba de la Mediana.

Estadísticos de contraste^b

	Valor Z de Peso correspondie nte a la edad y sexo del niñ o	Valor Z de Peso correspondie nte a la talla y sexo del niño	Valor Z de Talla correspondie nte a la edad y sexo del niñ o
N	4624	4624	4624
Mediana	-,3125	,4392	-,8873
Chi-cuadrado	2,935 ^a	1,554 ^a	6,635 ^a
gl	2	2	2
Sig. asintót.	,230	,460	,036

a. 0 casillas (,0%) tienen frecuencias esperadas menores que 5. La frecuencia de casilla esperada mínima es 214,5.

b. Variable de agrupación: Edad actual (años)

4.4.4 COMPARACIÓN DOS A DOS.

Se ha considerado conveniente realizar una comparación entre cada par de parejas de grupos de edades para el índice antropométrico Talla para la edad pues es el único del cual se concluyó anteriormente que se comporta de manera diferente. Los resultados de llevar a cabo dichas comparaciones se presentan el cuadro que se muestra a continuación.

En base a los resultados obtenidos en el cuadro, podemos apreciar que el estado nutricional de los niños menores de cinco años, según el índice antropométrico Talla para la edad, se comporta de manera idéntica para cada uno de los grupos de edades excepto para los grupos de 15 a 19 años en comparación con el grupo de 25 a 49 años, madres muy jóvenes versus madres muy maduras.

Tabla 4.4.4: Prueba de Kolmogorov-Smirnov de variables antropométricas.

Comparación por pares, Edad de la madre.

			20 - 24 AÑOS	25 - 49 AÑOS
			Valor Z de Talla correspondiente a la edad y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño
15 - 19 AÑOS	Diferencias más extremas	Absoluta	,060	,074
		Positiva	,060	,074
		Negativa	-,014	-,002
	Z de Kolmogorov-Smirnov	1,054	1,442	
	Sig. asintót. (bilateral)	,216	,031	
20 - 24 AÑOS	Diferencias más extremas	Absoluta		,038
		Positiva		,038
		Negativa		-,005
	Z de Kolmogorov-Smirnov		1,105	
	Sig. asintót. (bilateral)		,174	

En resumen podemos concluir a la luz de todos los resultados obtenidos anteriormente que el estado nutricional de los niños parece comportarse de manera idéntica en cada uno de los grupos de edades considerados, lo que significa que la edad de la madre no parece influir en el padecimiento de desnutrición de sus hijos, es decir, que no hay un grupo de edad en el cual sea más probable de observar niños desnutridos.

4.5 COMPARACIÓN DEL ESTADO NUTRICIONAL SEGÚN ESTADOS CIVILES FAMILIARES.

La hipótesis a contrastar es la siguiente:

- El comportamiento del estado nutricional de los niños y niñas menores de cinco años de edad no es el mismo en los distintos estados civiles de los padres de familia.

Para someter a prueba esta hipótesis se ha considerado conveniente realizar la siguiente clasificación de los estados civiles familiares: el primer grupo lo conforman madres casadas o acompañadas (madres en unión); el segundo lo conforman las madres separadas, divorciadas o viudas (madres que ya no están con su pareja); y el tercer grupo lo conforman las madres solteras. Como se dijo antes las madres en unión no necesariamente están unidas con el padre de familia del niño en estudio.

4.5.1 CONTRASTE DE NORMALIDAD.

Al igual que en todos los contrastes anteriores primero verificaremos si las variables antropométricas presentan una distribución normal en cada uno de los distintos estados civiles familiares. Los resultados se muestran en el siguiente cuadro.

Tabla 4.5.1: Contraste de normalidad, variables antropométricas. Prueba de Kolmogorov-Smirnov y Prueba de Shapiro Willk.

Pruebas de normalidad							
Estado civil/familiar		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
Valor Z de Peso correspondiente a la edad y sexo del niño	Casada/acompañada	,069	3751	,000	,943	3751	,000
	Separada/viuda/divorciada	,088	635	,000	,939	635	,000
	Soltera	,064	238	,020	,952	238	,000
Valor Z de Peso correspondiente a la talla y sexo del niño	Casada/acompañada	,064	3751	,000	,926	3751	,000
	Separada/viuda/divorciada	,154	635	,000	,606	635	,000
	Soltera	,142	238	,000	,790	238	,000
Valor Z de Talla correspondiente a la edad y sexo del niño	Casada/acompañada	,034	3751	,000	,978	3751	,000
	Separada/viuda/divorciada	,049	635	,001	,942	635	,000
	Soltera	,043	238	,200*	,977	238	,001

a. Corrección de la significación de Lilliefors

*. Este es un límite inferior de la significación verdadera.

En base a los resultados mostrados en el cuadro anterior podemos concluir que las variables antropométricas no se distribuyen de manera normal en cada uno de los estados civiles familiares; y por consiguiente debemos realizar un contraste no paramétrico para someter a prueba la hipótesis planteada.

4.5.2 PRUEBA DE KRUSKAL-WALLIS.

La hipótesis que estamos contrastando con esta prueba es la siguiente: *“La distribución de cada variable antropométrica de los niños menores de cinco años es idéntica para los distintos estados civiles familiares”*. Los resultados de dicha prueba se muestran a continuación.

Tabla 4.5.2: Prueba de igualdad de distribuciones, variables antropométricas.

Prueba de Kruskal-Wallis.

Estadísticos de contraste^{a, b}

	Valor Z de Peso correspondie nte a la edad y sexo del niñ o	Valor Z de Peso correspondie nte a la talla y sexo del niño	Valor Z de Talla correspondie nte a la edad y sexo del niñ o
Chi-cuadrado	5,006	,139	10,452
gl	2	2	2
Sig. asintót.	,082	,933	,005

a. Prueba de Kruskal-Wallis

b. Variable de agrupación: Estado civil/familiar

De los resultados de la tabla anterior si tomamos un nivel de significancia del 5% podemos concluir que para dos de las tres variables antropométricas no existe evidencia suficiente como para decir que el estado nutricional de los niños se comporte de manera diferente en cada uno de los estados civiles familiares, sin embargo, para uno de los indicadores antropométricos el no rechazo de la hipótesis de igualdad de distribuciones se acepta por un “*pelito*”. De manera general podemos decir que estado nutricional de los niños no se comporta de la misma manera en cada uno de los estados civiles familiares.

4.5.3 PRUEBA DE LA MEDIANA.

Al igual que como se ha hecho antes, consideramos conveniente realizar la Prueba de la Mediana. La hipótesis a contrastar mediante esta prueba es la siguiente: “*Para cada una de las variables antropométricas su valor mediano es el mismo en cada uno de los estados civiles familiares*”. Los resultados de dicha prueba se muestran en el siguiente cuadro.

Tabla 4.5.3: Prueba de igualdad de distribuciones, variables antropométricas.

Prueba de la Mediana.

Estadísticos de contraste^b

	Valor Z de Peso correspondie nte a la edad y sexo del niñ o	Valor Z de Peso correspondie nte a la talla y sexo del niño	Valor Z de Talla correspondie nte a la edad y sexo del niñ o
N	4624	4624	4624
Mediana	-,3125	,4392	-,8873
Chi-cuadrado	5,548 ^a	,168 ^a	12,752 ^a
gl	2	2	2
Sig. asintót.	,062	,920	,002

a. 0 casillas (,0%) tienen frecuencias esperadas menores que 5. La frecuencia de casilla esperada mínima es 119,0.

b. Variable de agrupación: Estado civil/familiar

En dicho cuadro podemos apreciar que únicamente para el índice antropométrico Peso para la Talla se acepta dicha hipótesis. Mientras que para las dos variables antropométricas restantes se rechaza (para la variable Peso para la Edad se aceptaría siempre y cuando el nivel de significancia sea menor al 6%)

4.5.4 COMPARACIÓN DOS A DOS.

Se ha considerado conveniente realizar una comparación entre cada par de parejas de estado civil familiar y los índices antropométricos en los cuales se rechazó anteriormente la igualdad de distribuciones y la igualdad de valores medios. Se ha tomado además el índice antropométrico Peso para la edad pues en la Prueba de la Mediana la aceptación de la hipótesis se da con una probabilidad muy baja, de hecho casi en el límite. Los resultados se muestran en el siguiente cuadro.

Tabla 4.5.4: Prueba de Kolmogorov-Smirnov de variables antropométricas.

Comparación por parejas del estado civil familiar.

			SEPARADA/VIUDA/DIVORCIADA		SOLTERA	
			Valor Z de Peso correspondiente a la edad y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la edad y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño
CASADA ACOMPAÑADA	Diferencias más extremas	Absoluta	,053	,087	,066	,035
		Positiva	,008	,004	,066	,035
		Negativa	-,053	-,087	-,015	-,034
	Z de Kolmogorov-Smirnov		1,228	2,023	,990	,522
	Sig. asintót. (bilateral)		,098	,001	,281	,948
SEPARADA VIUDA DIVORCIADA	Diferencias más extremas	Absoluta			,108	,108
		Positiva			,108	,108
		Negativa			-,010	-,006
	Z de Kolmogorov-Smirnov				1,419	1,420
	Sig. asintót. (bilateral)				,036	,035

Puede apreciarse que utilizando el índice Talla para la edad, este se comporta de manera diferente si se comparan las madres en unión (casadas o acompañadas) con las madres separadas (separadas, viudas o divorciadas); lo mismo se dice cuando se comparan madres solteras con madres separadas, y únicamente se concluye que para las madres solteras se comporta de manera idéntica que las madres en unión. Un comportamiento bastante parecido puede decir del índice antropométrico Peso para la edad.

A la luz de todos los resultados obtenidos se puede concluir que el estado nutricional de los niños se comporta de manera diferente en cada uno de los estados civiles de sus respectivas familias.

4.6 COMPARACIÓN DEL ESTADO NUTRICIONAL SEGÚN DESEO DE CONCEBIRLO.

La hipótesis a contrastar es la siguiente:

- El comportamiento del estado nutricional de los niños y niñas menores de cinco años de edad no es el mismo para las madres que deseaban el embarazo y de las que no lo deseaban.

4.6.1 CONTRASTE DE NORMALIDAD.

Al igual que en todos los contrastes anteriores, lo primero que verificaremos es la normalidad de las variables antropométricas en cada uno de los distintos deseos de concebirlo de los padres. Los resultados se muestran en el siguiente cuadro.

Tabla 4.6.1: Contraste de normalidad, variables antropométricas. Prueba de Kolmogorov-Smirnov y Prueba de Shapiro Willk.

deseo del hijo		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
Valor Z de Peso correspondiente a la edad y sexo del niño	Quería entonces	,070	2890	,000	,959	2890	,000
	Esperar mas tiempo	,081	878	,000	,869	878	,000
	No lo quería	,063	856	,000	,954	856	,000
Valor Z de Peso correspondiente a la talla y sexo del niño	Quería entonces	,061	2890	,000	,940	2890	,000
	Esperar mas tiempo	,095	878	,000	,824	878	,000
	No lo quería	,121	856	,000	,667	856	,000
Valor Z de Talla correspondiente a la edad y sexo del niño	Quería entonces	,036	2890	,000	,977	2890	,000
	Esperar mas tiempo	,040	878	,002	,992	878	,000
	No lo quería	,046	856	,000	,945	856	,000

a. Corrección de la significación de Lilliefors

En base a los resultados anteriores podemos apreciar que ninguna de las variables antropométricas sigue una distribución normal, por lo que se procederá como en los casos anteriores con procedimientos no paramétricos para poder contrastar dicha hipótesis.

4.6.2 PRUEBA DE KRUSKAL-WALLIS.

La hipótesis que estamos contrastando con esta prueba es la siguiente: “La distribución de cada variable antropométrica de los niños menores de cinco años es idéntica para los distintos deseos de concebirlos de sus padres”. Los resultados de dicha prueba se muestran a continuación.

Tabla 4.6.2: Prueba de igualdad de distribuciones, variables antropométricas.

Prueba de Kruskal-Wallis.

Estadísticos de contraste^{a, b}

	Valor Z de Peso correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la talla y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño
Chi-cuadrado	23,402	5,616	28,699
gl	2	2	2
Sig. asintót.	,000	,060	,000

a. Prueba de Kruskal-Wallis

b. Variable de agrupación: deseo del hijo

De los resultados de la tabla anterior, si tomamos un nivel de significancia del 5% podemos concluir que para dos de las tres variables antropométricas existe evidencia suficiente como para decir que el estado nutricional de los niños se comporta de manera diferente en cada uno de los deseos de concebirlo de sus padres; mientras que para el restante indicador se concluye que su distribución es idéntica en cada uno de los deseos de concebirlo, aunque dicha aceptación ocurre con una probabilidad baja. Lo que realmente nos damos cuenta es que el estado nutricional se comporta de manera diferente en cada uno de los deseos de concebirlo de sus padres.

4.6.3 PRUEBA DE LA MEDIANA.

Al igual que como se ha hecho antes, consideramos conveniente realizar la Prueba de la Mediana. La hipótesis a contrastar mediante en esta prueba es la siguiente: “Para cada una de las variables antropométricas su valor mediano es el mismo en cada uno de los deseos de concebirlos de sus padres”. Los resultados de dicha prueba se muestran en el siguiente cuadro.

Tabla 4.6.3: Prueba de igualdad de distribuciones, variables antropométricas.

Prueba de la Mediana.

Estadísticos de contraste^b

	Valor Z de Peso correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la talla y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño
N	4624	4624	4624
Mediana	-,3125	,4392	-,8873
Chi-cuadrado	9,228 ^a	4,046 ^a	12,265 ^a
gl	2	2	2
Sig. asintót.	,010	,132	,002

a. 0 casillas (,0%) tienen frecuencias esperadas menores que 5. La frecuencia de casilla esperada mínima es 428,0.

b. Variable de agrupación: deseo del hijo

En base a los resultados anteriores podemos concluir que para dos de los tres índices antropométricos considerados su valor mediano es distinto en cada uno de los deseos de concebirlo de sus padres. Se observa que la hipótesis se acepta para el índice Peso para la Talla tal y como sucedió en la prueba de Kruskal-Wallis, solo que en esta prueba la aceptación se da con una mayor probabilidad.

4.6.4 COMPARACIÓN DOS A DOS.

Se ha considerado conveniente realizar comparaciones entre cada par de parejas para los distintos deseos de concepción de las madres y los índices

antropométricos en los cuales se rechazó anteriormente la igualdad de distribuciones y la igualdad de valores medios. Se ha tomado todos los índices pues para Peso para la Talla se aceptó la igualdad de la distribución con una probabilidad muy baja. Los resultados se muestran a continuación.

Tabla 4.6.4: Prueba de Kolmogorov-Smirnov de variables antropométricas.
Comparación por pares del Deseo de concebir al hijo.

		ESPERAR MÁS TIEMPO			NO LO QUERÍA			
		Valor Z de Peso correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la talla y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la talla y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño	
QUERÍA ENTONCES	Diferencias más extremas	Absoluta	,060	,049	,052	,094	,052	,099
		Positiva	,013	,033	,003	,005	,006	,003
		Negativa	-,060	-,049	-,052	-,094	-,052	-,099
	Z de Kolmogorov-Smirnov	1,546	1,281	1,352	2,406	1,333	2,553	
	Sig. asintót. (bilateral)	,017	,075	,052	,000	,057	,000	
ESPERAR MÁS TIEMPO	Diferencias más extremas	Absoluta				,080	,057	,081
		Positiva				,010	,036	,009
		Negativa				-,080	-,057	-,081
	Z de Kolmogorov-Smirnov				1,660	1,197	1,694	
	Sig. asintót. (bilateral)				,008	,114	,006	

De los resultados que se muestran en dicho cuadro se concluye que el estado nutricional de los niños es diferente dependiendo del deseo que tenían sus padres al momento de concebirlos. Los hijos de las madres que deseaban en ese entonces el embarazo presentan un comportamiento diferente en los tres índices antropométricos a las que querían esperar más tiempo y a las que no los querían. Para las madres que querían esperar más tiempo y las que no los querían dos de los tres índices antropométricos se comportan de manera distinta.

Todo parece indicar que entre menos era el deseo de embarazarse y de tener al niño más es el riesgo de que éste padezca de desnutrición, pues las hipótesis de

igualdad de distribución se rechazan todas por valores extremos negativos en las diferencias extremas de distribuciones.

4.7 COMPARACIÓN DEL ESTADO NUTRICIONAL SEGÚN NIVEL SOCIOECONÓMICO.

La hipótesis a contrastar es la siguiente:

- El comportamiento del estado nutricional de los niños y niñas menores de cinco años de edad no es el mismo para los distintos niveles socioeconómicos de la familia.

4.7.1 CONTRASTE DE NORMALIDAD.

Al igual que en los contrastes anteriores, lo primero que verificamos es la normalidad de las variables antropométricas en cada uno de los niveles socioeconómicos de la familia. Los resultados se muestran en el siguiente cuadro.

Al igual que como ha ocurrido en todas las hipótesis anteriores se descarta la normalidad de las variables antropométricas, por lo que nuevamente utilizaremos la prueba de Kruskal -Wallis y la Prueba de la Mediana para contrastar esta hipótesis.

Tabla 4.7.1: Contraste de normalidad, variables antropométricas. Prueba de Kolmogorov-Smirnov y Prueba de Shapiro Willk.

Nivel socioeconómico		Pruebas de normalidad			Pruebas de normalidad		
		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
Valor Z de Peso correspondiente a la edad y sexo del niño	Bajo	,052	1285	,000	,983	1285	,000
	Medio	,076	2146	,000	,921	2146	,000
	Alto	,074	1193	,000	,941	1193	,000
Valor Z de Peso correspondiente a la talla y sexo del niño	Bajo	,066	1285	,000	,894	1285	,000
	Medio	,086	2146	,000	,788	2146	,000
	Alto	,088	1193	,000	,900	1193	,000
Valor Z de Talla correspondiente a la edad y sexo del niño	Bajo	,035	1285	,001	,974	1285	,000
	Medio	,042	2146	,000	,965	2146	,000
	Alto	,038	1193	,000	,973	1193	,000

a. Corrección de la significación de Lilliefors

4.7.2 PRUEBA DE KRUSKAL-WALLIS.

La hipótesis que estamos contrastando es la siguiente: *“La distribución de cada variable antropométrica de los niños menores de cinco años es idéntica para los distintos niveles socioeconómicos de la familia”*. Los resultados de dicha prueba se muestran a continuación.

Tabla 4.7.2: Prueba de igualdad de distribuciones, variables antropométricas. Prueba de Kruskal-Wallis.

Estadísticos de contraste ^{a, b}			
	Valor Z de Peso correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la talla y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño
Chi-cuadrado	245,654	47,519	324,970
gl	2	2	2
Sig. asintót.	,000	,000	,000

a. Prueba de Kruskal-Wallis

b. Variable de agrupación: Nivel socioeconómico

En base a los resultados anteriores podemos concluir que el estado nutricional de los niños menores de cinco años se comporta de manera claramente diferente en cada uno de los distintos niveles socioeconómicos de su familia.

4.7.3 PRUEBA DE LA MEDIANA.

Al igual que como se ha hecho antes, consideramos conveniente realizar la Prueba de la Mediana. La hipótesis a contrastar mediante en esta prueba es la siguiente: *“Para cada una de las variables antropométricas su valor mediano es el mismo en los distintos niveles socioeconómicos de la familia”*. Los resultados de dicha prueba se muestran en el siguiente cuadro.

Tabla 4.7.3: Prueba de igualdad de distribuciones, variables antropométricas.

Prueba de la Mediana.

Estadísticos de contraste ^b			
	Valor Z de Peso correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la talla y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño
N	4624	4624	4624
Mediana	-,3125	,4392	-,8873
Chi-cuadrado	170,204 ^a	31,251 ^a	222,989 ^a
gl	2	2	2
Sig. asintót.	,000	,000	,000

a. 0 casillas (.0%) tienen frecuencias esperadas menores que 5. La frecuencia de casilla esperada mínima es 596,5.

b. Variable de agrupación: Nivel socioeconómico

En base a los resultados anteriores podemos concluir que el estado nutricional de los niños menores de cinco años se comporta de manera diferente en cada uno de los niveles socioeconómicos de su familia. Resultado que concuerda con el obtenido en la prueba de Kruskal-Wallis.

4.7.4 COMPARACIÓN DOS A DOS.

Nuevamente se ha considerado conveniente realizar comparaciones entre cada par de parejas de los tres índices antropométricos para los distintos niveles socioeconómicos de la familia. Los resultados se muestran a continuación.

En base a los resultados mostrados, se observa que todos los índices se comportan de manera diferente en cada uno de los niveles socioeconómicos de la familia. Todo parece indicar que a mayor nivel socioeconómico la distribución de los índices antropométricos se encuentra por encima del nivel precedente; esto indica que es más probable que las familias con bajo nivel socioeconómico tengan hijos que padezcan de desnutrición (pues la distribución para este nivel se encuentra a la izquierda de los otros dos niveles).

Tabla 4.7.4: Prueba de Kolmogorov-Smirnov de variables antropométricas.
Comparación por pares del Nivel socioeconómico de la familia.

		NIVEL MEDIO			NIVEL ALTO			
		Valor Z de Peso correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la talla y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la edad y sexo del niño	Valor Z de Peso correspondiente a la talla y sexo del niño	Valor Z de Talla correspondiente a la edad y sexo del niño	
NIVEL BAJO	Diferencias más extremas	Absoluta	,148	,055	,170	,267	,121	,307
		Positiva	,000	,002	,001	,267	,121	,307
		Negativa	-,148	-,055	-,170	,000	-,002	-,001
	Z de Kolmogorov-Smirnov	4,190	1,563	4,818	6,643	2,998	7,634	
	Sig. asintót. (bilateral)	,000	,015	,000	,000	,000	,000	
NIVEL MEDIO	Diferencias más extremas	Absoluta				,133	,077	,144
		Positiva				,133	,077	,144
		Negativa				-,001	-,001	-,003
	Z de Kolmogorov-Smirnov				3,676	2,139	3,993	
	Sig. asintót. (bilateral)				,000	,000	,000	

CAPÍTULO V

CONCLUSIONES

5.1 CONCLUSIONES REGRESIÓN LOGÍSTICA

Entre los resultados de importancia en la regresión logística se pueden mencionar los siguientes:

- El modelo de regresión logístico permite relacionar la variable dependiente Índice de nutrición combinado y las variables independientes: Número de años completos de educación de la madre, Número de servicios básicos y El número de hijos que viven en el mismo hogar en la Encuesta de Salud Familiar (FESAL) realizada por la Asociación Demográfica Salvadoreña en el año 2008.
- Las variables que están relacionadas de manera negativa con el estado nutricional de la niñez son la Educación de la madre y el Número de servicios básicos con que cuenta la familia, esto significa que al incrementar de manera individual o de manera conjunta a dichas variables, en promedio se espera que la probabilidad de padecer desnutrición de los niños se reduzca.
- El Número de hijos que viven en el hogar influye positivamente en el estado nutricional, es decir, a mayor número de niños viviendo bajo un mismo techo mayor será el riesgo que padezcan de desnutrición (mayor probabilidad de padecerla).
- Con las variables consideradas en este estudio solamente es posible clasificar correctamente el estado nutricional de la niñez salvadoreña para aproximadamente las dos terceras partes (tomando como punto de corte el valor de 0.085); teniendo en cuenta que se sobreestima el número de niños

desnutridos reales, sin embargo, el porcentaje de niños desnutridos clasificados correctamente por el modelo se encuentra cercano al 90%.

- Con ayuda del modelo de regresión logístico encontrado, puede apreciarse que las variables independientes con las que se cuenta en este momento son insuficientes para estimar el estado nutricional, pues la variabilidad de la variable dependiente es explicada en apenas un 11.6% (R^2 de Nagelkerke) en base a dichas variables.

5.2 CONCLUSIONES ANÁLISIS DISCRIMINANTE

Entre los resultados de importancia del análisis discriminante se pueden mencionar los siguientes:

- El análisis discriminante permite relacionar la variable dependiente Índice de nutrición combinado y las variables independientes: Número de años completos de educación de la madre, Número de servicios básicos, Edad de la madre al nacimiento del niño y Número de hijos que viven en el mismo hogar en la Encuesta de Salud Familiar (FESAL) realizada por la Asociación Demográfica Salvadoreña en el año 2008.
- Según el análisis discriminante la variable más importante en la estimación del estado nutricional de los niños salvadoreños es; el Número de servicios básicos con que cuenta la familia. Mientras que la variable menos importante es la Edad de la madre al nacimiento del niño.
- Con las variables consideradas en este estudio solamente es posible clasificar correctamente el estado nutricional de la niñez salvadoreña para

aproximadamente las dos terceras partes de los niños; un cantidad similar sucede en los niños desnutridos y no desnutridos.

- Al igual que el modelo de regresión logístico, el análisis discriminante sobreestima el número de niños desnutridos.
- Con ayuda de la función canónica discriminante puede apreciarse que las variables independientes con las que se cuenta son insuficientes para estimar el estado nutricional, pues la variabilidad de la variable dependiente es explicada en apenas un 7.45% en base a dichas variables.
- La función lineal discriminante proporciona resultados muy parecidos a la función cuadrática a pesar que la hipótesis de igualdad de matrices covarianzas es rechazada, esto es debido principalmente al tamaño de la muestra.

5.3 CONCLUSIONES REGRESIÓN LOGÍSTICA Y ANÁLISIS DISCRIMINANTE

Entre los resultados de mayor importancia se pueden mencionar los siguientes:

- Los modelos ajustados de regresión logística y de análisis discriminante a la muestra ampliada (100% de los datos) clasifican de forma similar a niños desnutridos y no desnutridos como lo hacen en la muestra de estimación (75% de los datos).

- El modelo estimado podría ser extendido a toda la población salvadoreña de niños menores de cinco años para el año 2008 esperando un porcentaje global de clasificación alrededor de las dos terceras partes.
- Los modelos estimados de regresión logística y análisis discriminante para el año 2008 podrían utilizarse como guía para hacer estudios comparativos en otros períodos pasados o futuros.
- La única diferencia entre el modelo logístico y el análisis discriminante radica en la variable Edad de la madre al nacimiento del niño, la cual es significativa en el análisis discriminante pero no así en el modelo logístico. Además el porcentaje de variabilidad de la variable dependiente por los modelos es muy bajo, muy malo.
- En términos generales puede decirse que la regresión logística es modelo más adecuado para tratar de estimar el estado nutricional de la niñez salvadoreña en comparación con el análisis discriminante.

ANEXOS

A. Comportamiento de variables antropométricas.

Cuadro A.1: Tablas de contingencia, Índice de nutrición combinado vs. Estado nutricional Peso para la Edad, Estado nutricional Peso para la Talla y Estado nutricional Talla para la Edad.

Tabla de contingencia Estado Nutricional Peso para la Edad * Índice de Nutrición Combinado

Recuento

		Índice de Nutrición Combinado		Total
		Normal	Desnutrido	
Estado Nutricional Peso para la Edad	Normal	3810	621	4431
	Desnutrido	0	193	193
Total		3810	814	4624

Tabla de contingencia Estado Nutricional Peso para la Talla * Índice de Nutrición Combinado

Recuento

		Índice de Nutrición Combinado		Total
		Normal	Desnutrido	
Estado Nutricional Peso para la Talla	Normal	3810	771	4581
	Desnutrido	0	43	43
Total		3810	814	4624

Tabla de contingencia Estado Nutricional Talla para la Edad * Índice de Nutrición Combinado

Recuento

		Índice de Nutrición Combinado		Total
		Normal	Desnutrido	
Estado Nutricional Talla para la Edad	Normal	3810	59	3869
	Desnutrido	0	755	755
Total		3810	814	4624

Cuadro A.2: Tablas de contingencia, entre las variables antropométricas: Estado nutricional Peso para la Edad, Estado nutricional Peso para la Talla y Estado nutricional Talla para la Edad.

Tabla de contingencia Estado Nutricional Talla para la Edad * Estado Nutricional Peso para la Edad

Recuento

		Estado Nutricional Peso para la Edad		Total
		Normal	Desnutrido	
Estado Nutricional Talla para la Edad	Normal	3833	36	3869
	Desnutrido	598	157	755
Total		4431	193	4624

Tabla de contingencia Estado Nutricional Talla para la Edad * Estado Nutricional Peso para la Talla

Recuento

		Estado Nutricional Peso para la Talla		Total
		Normal	Desnutrido	
Estado Nutricional Talla para la Edad	Normal	3837	32	3869
	Desnutrido	744	11	755
Total		4581	43	4624

Tabla de contingencia Estado Nutricional Peso para la Edad * Estado Nutricional Peso para la Talla

Recuento

		Estado Nutricional Peso para la Talla		Total
		Normal	Desnutrido	
Estado Nutricional Peso para la Edad	Normal	4408	23	4431
	Desnutrido	173	20	193
Total		4581	43	4624

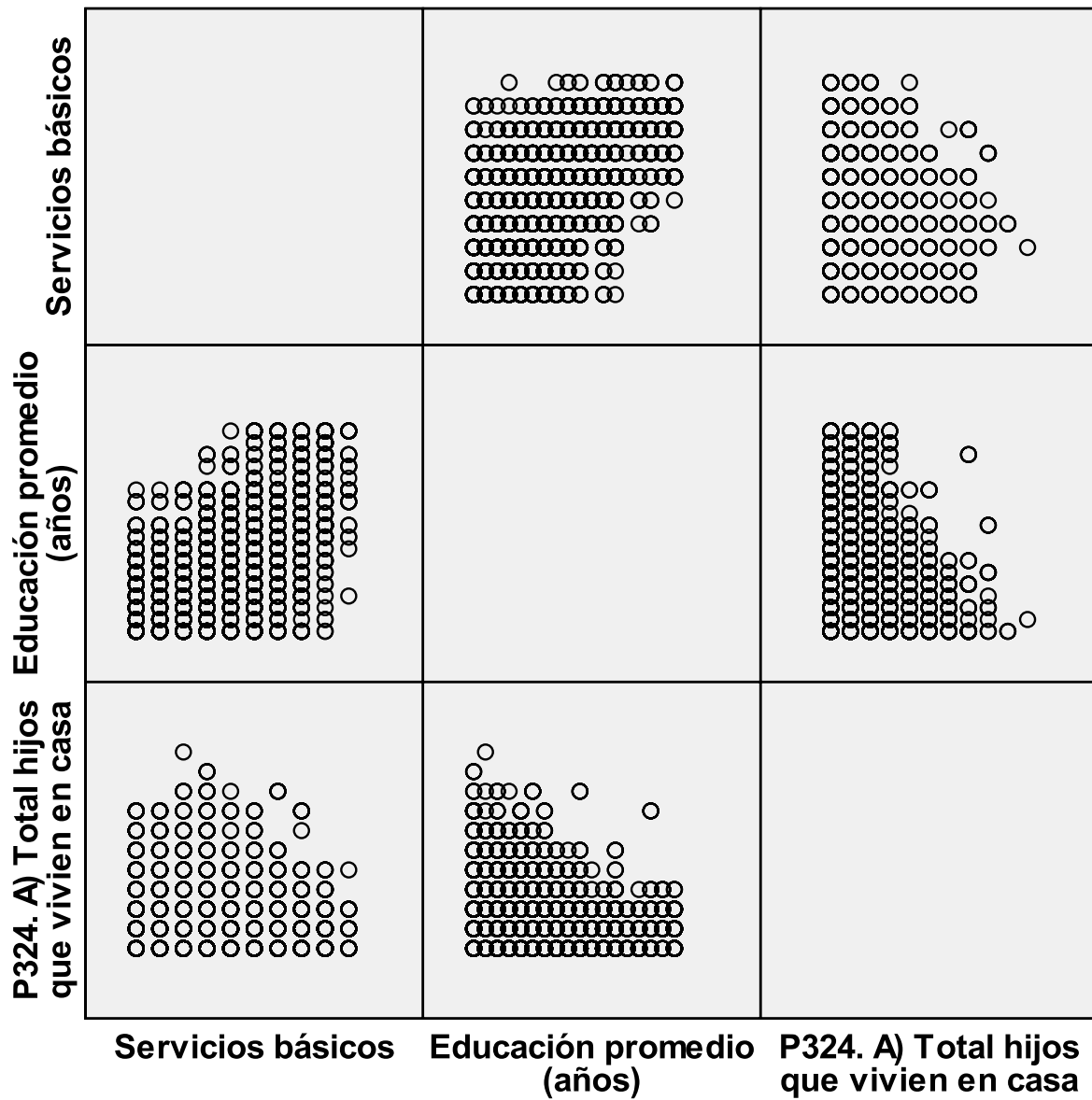
Tabla de contingencia Estado Nutricional Peso para la Talla * Estado Nutricional Talla para la Edad * Estado Nutricional Peso para la Edad

Recuento

Estado Nutricional Peso para la Edad			Estado Nutricional Talla para la Edad		Total
			Normal	Desnutrido	
Normal	Estado Nutricional Peso para la Talla	Normal	3810	598	4408
		Desnutrido	23	0	23
	Total		3833	598	4431
Desnutrido	Estado Nutricional Peso para la Talla	Normal	27	146	173
		Desnutrido	9	11	20
	Total		36	157	193

B. Multicolinealidad de las variables explicativas, Regresión Logística.

Figura B.1: Matriz de diagrama de dispersión. Variables independientes (Educación promedio de la madre, Número de servicios básicos y Número de hijos en casa), modelo de Regresión Logística.



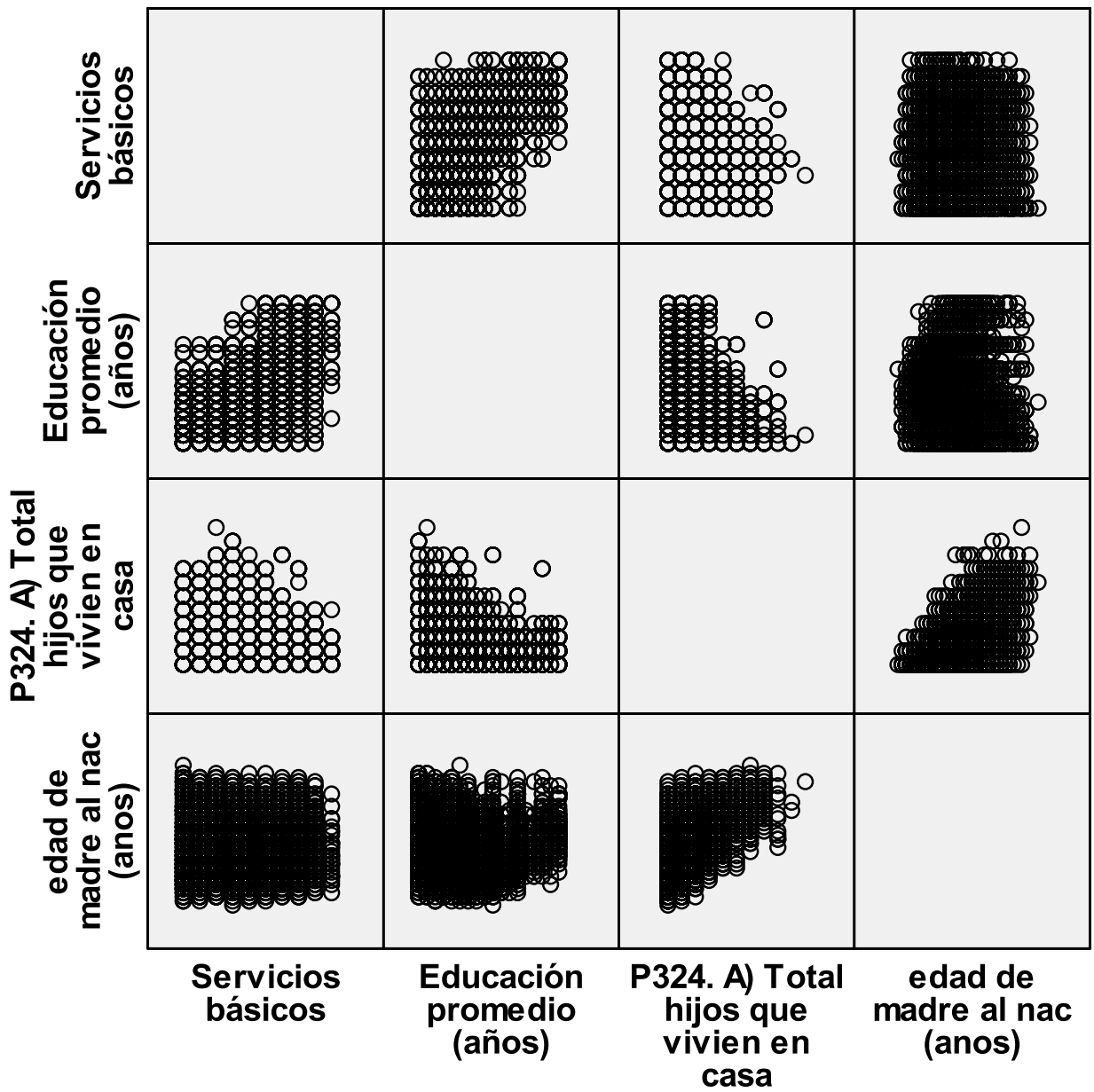
Cuadro B.2: Matriz de correlaciones. Variables independientes (Educación promedio de la madre, Número de servicios básicos y Número de hijos en casa), modelo de Regresión Logística.

Matriz de correlaciones

		Constant	ServBasicos	P324A	edupromef
Paso 1	Constant	1,000	-,826		
	ServBasicos	-,826	1,000		
Paso 2	Constant	1,000	-,707	-,748	
	P324A	-,748	,232	1,000	
	ServBasicos	-,707	1,000	,232	
Paso 3	Constant	1,000	-,355	-,772	-,396
	P324A	-,772	,039	1,000	,292
	edupromef	-,396	-,511	,292	1,000
	ServBasicos	-,355	1,000	,039	-,511

C. Multicolinealidad de las variables explicativas, Análisis Discriminante.

Figura C.1: Matriz de diagrama de dispersión. Variables independientes (Educación promedio de la madre, Número de servicios básicos, Edad de la madre al nacimiento del niño y Número de hijos en casa), Análisis Discriminante.



D. Modelo de Regresión Logístico si se elimina el termino constante.

Cuadro D.1: Resumen de clasificación del modelo logístico si se elimina la constante del modelo. Fase de estimación.

Tabla de clasificación^a

Observado			Pronosticado		
			Indice de Nutrición Combinado		Porcentaje correcto
			Normal	Desnutrido	
Paso 1	Indice de Nutrición Combinado	Normal	1444	1423	50,4
		Desnutrido	174	437	71,5
	Porcentaje global				54,1
Paso 2	Indice de Nutrición Combinado	Normal	1699	1168	59,3
		Desnutrido	193	418	68,4
	Porcentaje global				60,9

a. El valor de corte es ,170

Cuadro D.2: Prueba de bondad de ajuste de Hosmer Lemeshow de modelo logístico si se elimina la constante. Fase de estimación.

Prueba de Hosmer y Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	50,779	7	,000
2	23,656	8	,003

E. CÓDIGO EN SPSS PARA VALIDAR EL MODELO DE REGRESIÓN LOGÍSTICO.

***CALCULANDO EL LOGIT EN REGRESIÓN LOGÍSTICA.**

```
COMPUTE LOGITO =-0.8927831533863975 -0.1969402911024095*(ServBasicos)
+0.14203306091590315*(P324A) -0.053942248492328905*(edupromef).
VARIABLE LABELS LOGITO 'Cálculo del logit'.
EXECUTE.
```

***CALCULANDO LA PROBABILIDAD DE PADECER DESNUTRICIÓN.**

```
COMPUTE PROB_DESNUTRIDO = 1/(1+EXP(-1*LOGITO)).
VARIABLE LABELS PROB_DESNUTRIDO 'Probabilidad de desnutrir'.
EXECUTE.
```

***CLASIFICACIÓN CON EL MODELO LOGÍSTICO PUNTO DE CORTE 0.085 (8.5%).**

*** NIÑOS DESNUTRIDOS**

```
IF (PROB_DESNUTRIDO >= 0.085) CLASIF_LOGISTICA=1.
VARIABLE LABELS CLASIF_LOGISTICA 'Clasificación con el modelo de
Regresión Logística'.
EXECUTE.
```

*** NIÑOS NO DESNUTRIDOS**

```
IF (PROB_DESNUTRIDO < 0.085) CLASIF_LOGISTICA=0.
VARIABLE LABELS CLASIF_LOGISTICA 'Clasificación con el modelo de
Regresión Logística'.
EXECUTE.
```

*** DEFINIENDO LOS PUNTOS DE CORTES (DECILES) PARA USAR LA PRUEBA DE HOSMER Y LEMESHOW.**

IF (PROB_DESNUTRIDO <= 0.1) DEC_LOGISTICA=1.
 IF (PROB_DESNUTRIDO > 0.1 & PROB_DESNUTRIDO <= 0.2)
 DEC_LOGISTICA=2.
 IF (PROB_DESNUTRIDO > 0.2 & PROB_DESNUTRIDO <= 0.3)
 DEC_LOGISTICA=3.
 IF (PROB_DESNUTRIDO > 0.3 & PROB_DESNUTRIDO <= 0.4)
 DEC_LOGISTICA=4.
 IF (PROB_DESNUTRIDO > 0.4) DEC_LOGISTICA=5.
 EXECUTE.

F. Proporción de Falsos Positivos y Proporción de Falsos Negativos para el modelo de Regresión Logístico en diferentes puntos de corte.

Cuadro F.1: Proporción de Falsos Positivos y Falsos Negativos en el modelo de Regresión Logístico para diferentes puntos de corte. Fase de validación.

PUNTO DE CORTE	FALSO S "+"	FALSO S "-"	PUNTO DE CORTE	FALSO S "+"	FALSO S "-"	PUNTO DE CORTE	FALSO S "+"	FALSO S "-"
0,000	1,00000	0,00000	0,085	0,73807	0,10345	0,170	0,33192	0,38916
0,005	1,00000	0,00000	0,090	0,71050	0,10345	0,175	0,31495	0,40887
0,010	1,00000	0,00000	0,095	0,67126	0,11330	0,180	0,29480	0,42857
0,015	1,00000	0,00000	0,100	0,64051	0,14286	0,185	0,28526	0,43842
0,020	1,00000	0,00000	0,105	0,62036	0,17241	0,190	0,27147	0,44828
0,025	1,00000	0,00000	0,110	0,59279	0,17241	0,195	0,26617	0,45813
0,030	1,00000	0,00000	0,115	0,56946	0,17734	0,200	0,25981	0,46798
0,035	0,99788	0,00000	0,120	0,54295	0,20197	0,210	0,23436	0,48276

PUNTO DE CORTE	FALSO S "+"	FALSO S "-"	PUNTO DE CORTE	FALSO S "+"	FALSO S "-"	PUNTO DE CORTE	FALSO S "+"	FALSO S "-"
0,040	0,98940	0,00493	0,125	0,52386	0,21182	0,220	0,21845	0,51232
0,045	0,97985	0,01478	0,130	0,49629	0,21675	0,230	0,20679	0,51724
0,050	0,96288	0,01478	0,135	0,46766	0,25616	0,240	0,19300	0,55665
0,055	0,94486	0,02956	0,140	0,44751	0,26108	0,250	0,17179	0,58621
0,060	0,90456	0,04433	0,145	0,42842	0,28079	0,260	0,15164	0,59113
0,065	0,88335	0,04926	0,150	0,40085	0,31527	0,270	0,13256	0,62069
0,070	0,84942	0,06404	0,155	0,37434	0,33990	0,280	0,12195	0,64532
0,075	0,80912	0,06404	0,160	0,35207	0,35961	0,290	0,11029	0,67488
0,080	0,78261	0,07389	0,165	0,33828	0,37931	0,300	0,10074	0,68473

G. CÓDIGO EN SPSS PARA VALIDAR EL ANÁLISIS DISCRIMINANTE.

*FUNCIÓN LINEAL DISCRIMINA DE FISHER, NIÑOS DESNUTRIDOS.

COMPUTE FisherDesnu = 0.3748184756349974*(ServBasicos) +
0.1746434773415824*(edupromef) + 0.03933109373130916*(P324A) +
0.6379812856725077*(medad) -- 9.979123142982573.

VARIABLE LABELS FisherDesnu 'Función Lineal de Fisher niños desnutridos'.
EXECUTE.

*FUNCIÓN LINEAL DISCRIMINATE DE FISHER, NIÑOS NO DESNUTRIDOS.

COMPUTE FisherNormal =0.5768084720791927*(ServBasicos) +
0.21239230781393523*(edupromef) - 0.21636357585152358*(P324A) +
0.6596130252793944*(medad) - 10.80077620310987.

VARIABLE LABELS FisherNormal 'Función Lineal de Fisher niños normales'.
EXECUTE.

***CLASIFICACIÓN CON LAS FUNCIONES DE FISHER.**

*** NIÑOS DESNUTRIDOS.**

IF (FisherDesnu >= FisherNormal) CLAS_DISC_LI=1.

VARIABLE LABELS CLAS_DISC_LI 'Clasificación con el Análisis Discriminante'.

EXECUTE.

*** NIÑOS NO DESNUTRIDOS**

IF (FisherDesnu < FisherNormal) CLAS_DISC_LI=0.

VARIABLE LABELS CLAS_DISC_LI 'Clasificación con el Análisis Discriminante'.

EXECUTE.

***FUNCION CANÓNICA DISCRIMINANTE LINEAL.**

COMPUTE FCanonica = 0.2711553054116229*(ServBasicos) +
0.050674765264872153*(edupromef) - 0.3432495046456959*(P324A) +
0.029038868572436732*(medad) - 1.3445999267731934.

VARIABLE LABELS FisherNormal 'Función Canónica Discriminante'.

EXECUTE.

***CLASIFICACIÓN CON LA FUNCION CANONICA.**

*** NIÑOS NO DESNUTRIDOS**

IF (FCanonica >= 0) CLAS_DISC_LI=0.

VARIABLE LABELS CLAS_DISC_LI 'Clasificación con el Análisis Discriminante'.

EXECUTE.

*** NIÑOS DESNUTRIDOS**

IF (FCanonica < 0) CLAS_DISC_LI=1.

**VARIABLE LABELS CLAS_DISC_LI 'Clasificación con el Análisis
Discriminante'.**

EXECUTE.

BIBLIOGRAFÍA.

LIBROS SOBRE ESTADÍSTICA

PEÑA, D. (2002): Regresión y Diseños de Experimentos. Editorial Alianza.

PEÑA, D. (2005): Fundamentos de Estadística. Editorial Alianza.

PEÑA, D. (2002): Análisis de datos multivariantes. Editorial Mc Graw Hill.

MONGOMERY, D.; PECK, E. y VINING, G. (2002): Introducción al análisis de regresión lineal. Editorial CECSA.

HAIR, J.; ANDERSON, R.; TATHAM, R. y BLACK, W. (2008): Análisis Multivariante, 5ª Edición. Editorial Prentice Hall.

SILVA AYÇAGUER, L. C. y BARROSO UTRA, I. M. (2004): Regresión Logística. Editorial LA MURALLA.

LEMUS GÓMEZ, R. y LEMUS GÓMEZ, O. (2005): Construcción de un modelo de regresión logístico sobre la oferta laboral a jefes(as) de hogares en El Salvador, Tesis de Maestría en Estadística. <http://virtual.ues.edu.sv>.

GARCÍA CABRERA, R. Y. y SORTO VÁSQUEZ, S. E. (2007): Estadística no paramétrica con aplicaciones, Tesis de Licenciatura en Estadística. <http://virtual.ues.edu.sv>.

KLEINBAUM, D. G. and KLEIN, M. (2010): Logistic Regression, Third Edition. Springer.

PAMBEL, F. C. (2000): Logistic Regression, A Primer. Sage Publications.

AGRESTI, A. (2002): Categorical Data Analysis, Second Edition. John Wiley & Sons, Inc. Publication.

CUADRAS, C. M. (2010): Nuevos métodos de análisis multivariante. CMC Editions.

DOBSON, A. J. (2002): Introduction to generalized linear models, Second Edition. Chapman & Hall/CRC.

CAO ABAD, R.; FERNÁNDEZ, M.; FERNÁNDEZ, S. N.; PRESSEDO QUINDIMIL, M. A.; VÁSQUEZ BRAGE, M.; VILAR FERNÁNDEZ, J. A. y VILAR FERNÁNDEZ, J. M. (2006): Introducción a la Estadística y sus aplicaciones. Editorial La Pirámide.

CONOVER, W.J. (1971): Practical Nonparametric Statistic. John Wiley & Sons Inc. Publication.

GIL PASCUAL, J.A. (____): Métodos de investigación en educación, Volumen III, Análisis multivariante. Madrid. UNED.

ARTICULOS SOBRE NUTRICIÓN

SANTISTEBAN, J. (2001): Curso de Nutrición - Evaluación del estado nutricional. www.ehas.org.

HODGUSON, M. I. (____): Evaluación del estado nutricional, www.escuela.med.puc.cl

ESTRADA JORGE, G. I. y MATOS, A. (2011): Análisis nutricional de niños en etapa materna e inicial en el Colegio "Divina Santa Marta" y Fundecresa. Ministerio del poder popular para la educación U.E.C. "Divina Santa Marta".

_____. (2011): Exploración Física y medidas antropométricas. www.infonutricion.com.

ÁLVAREZ MARTA, C.; LÓPEZ, A. y ESTRADA, A. (2009): Estado nutricional de niños de Antioquia, Colombia, según dos sistemas de referencia. Rev Panam Salud Publica. 2009; 25(3):196-203.

CANCELA, M. P. (____): Evaluación antropométrica en la desnutrición infantil. www.innatia.com.

DE BELAUSTEGUI, E. A.; CASSE, A. C. y Otros. (2008): Evaluación del estado nutricional de niños de 2 a 6 años en la localidad de San Roque. Revista de posgrado de la VIa Catedr  de Medicina, N  183, Julio 2008.

HORTON, S.; SHEKAR, M. and OTHER. (2010): Scaling Up Nutrition. The International Bank for Reconstruction and Development.

RADRIGAN, M. E.; VALLEJOS, E. y OTROS. (1973): Clasificación del estado nutricional. Rev. Chilena de Pediatría, Vol. 44, N° 5, 1973.

DEPARTAMENTO DE NUTRICIÓN PARA LA SALUD Y DESARROLLO.
(_____): Patrones de crecimiento infantil de la OMS. www.oms.org.

UNICEF. (2010): Informe anual 2010. www.unicef.org/publications.

UNICEF. (2006): Progreso para la infancia, Un balance sobre la nutrición.
www.unicef.org/publications.

PNUD. (2010): Informe sobre Desarrollo Humano 2010. Ediciones Mundi-
Prensa México.