UNIVERSIDAD DE PANAMÁ

VICERRECTORIA DE INVESTIGACIÓN Y POSTGRADO

PROGRAMA CENTROAMERICANO DE MAESTRÍA EN MATEMÁTICA

EL MODELO DE REGRESIÓN LOGÍSTICA Y SU APLICACIÓN EN EL ANÁLISIS DE LA DESNUTRICIÓN INFANTIL

GISELA DEL CARMEN CERVANTES DE LEÓN

TESIS PRESENTADA COMO UNO DE LOS REQUISITOS PARA OPTAR POR EL GRADO DE MAESTRO EN CIENCIAS CON ESPFCIALIZACIÓN EN ESTADÍSTICA MATEMÁTICA

PANAMÁ, REPÚBLICA DE PANAMÁ

2003

APROBADO POR:

M.Sc. GLADYS SEGURA
PRESIDENTE

Edilberto de León

M.Sc. EDILBERTO DE LEON
MIEMBRO

M.Sc. JOSÉ OCHOA
MIEMBRO

REPRESENTANTE DE LA VICERRECTORÍA DE INVESTIGACIÓN Y POSTGRADO

FECHA: 7/IV/03.

AGRADECIMIENTO

Ante todo, le doy gracias a Dios, nuestro señor, por damos la fuerza para vencer los obstáculos y hacer posible que se concretara ésta tan anhelada meta.

También, quiero expresar mi agradecimiento a mis padres por el apoyo que me han dado durante mi vida y por valor que le dieron a mi educación, lo que me ha permitido seguir hacia delante

A mis hermanos por su motivación y entereza.

A la profesora Gladys Segura por colaboración y motivación, al profesor Edilberto De León por su dedicación

A la Dirección de Política Social del Ministerio de Economía y Finanzas por facilitarnos la base de datos de la Encuesta de Niveles de Vida de 1997, utilizada en este estudio. En especial a la Sección de Informática que nos suministró dicha base con las variables analizadas en este trabajo

A todos,

Gracias

DEDICATORIA

A Dios ante todo y a la siempre bienaventurada Virgen María.

A mi sobrinito Sergio, una personita muy especial para mí

A mi futuros colegas, los estudiantes de la Escuela de Estadística.

A ellos, con todo mi cariño, dedico este trabajo de graduación

Gisela

ÍNDICE

INDICE DE CUADRO

INDICE DE TABLA

INDICE DE GRÁFICA

RESUMEN

El presente trabajo tiene por objetivo generalizar el modelo de regresión logística simple a más de una variable independiente A través de este estudio estimaremos e interpretaremos los coeficientes de modelo, veremos las diferentes técnicas de selección y significancía de variables y la adecuación del modelo Se realiza su aplicación a datos provenientes de la Encuesta de Niveles de vida de 1997 donde se analiza las variables asociadas a la desnutrición infantil en menores de cinco años

SUMMARY

The objetive of this work is to generalize the simple logistic model of regression to more than independent variable Through this research we will estimate and interpret the coefficient models, we will see the various technique of seleccion and variable meanings and fitness model Its application is oriented to data coming from a survey about Niveles de Vida from 1997 where variables associated to infantile malnutricion in children under five years are analyzed

INTRODUCCIÓN

El esfuerzo por mejorar la calidad de vida de la población, lograr una distribución de los ingresos y servicios del Estado más equitativa y disminuir progresivamente los niveles de pobreza son los elementos centrales de la política económica y social en la agenda de los gobiernos de Panamá.

Para cumplir las metas de desarrollo social con equidad, el Ministerio de Economía y Finanzas, emprendió la tarea de actualizar, ampliar y mejorar las fuentes de información sobre el bienestar y la pobreza, que permitiera la producción de datos oportunos y confiables, para diseñar, evaluar las políticas sociales y medir su impacto en la población Las Encuesta de Niveles de Vida (ENV 97) es un estudio cuyo propósito es proporcionar información a nivel de hogares y personas, sobre las condiciones de vida y bienestar de la población

Dentro de este contexto, este trabajo de investigación analiza las variables asociadas a la desnutrición en la niñez menor de 5 años, mediante el uso de los datos provenientes de la ENV-97 como fuente secundaria Para este análisis se requirió utilizar el modelo de regresión logística, cuya aplicación es recomendada por la organización Panamericana de la Salud (OPS) en materia de desnutrición

El análisis de regresión lineal trata de establecer la naturaleza de la relación existente entre una variable respuesta (dependiente) y una o más variables independientes El problema del análisis de regresión lineal consiste en que no siempre la variable respuesta esta dada en escala de intervalo. Por lo que surge el modelo de Regresión Logística Multivariado, el cual nos permite expresar mediante un modelo matemático la relación funcional existente entre una variable respuesta discreta y un conjunto de variables independientes En donde la variable respuesta puede tomar dos o más valores posibles y es codificada como cero (cuando el evento no ocurre) o uno (cuando el evento ocurre)

Al desarrollar un modelo de regresión logística debemos utilizar solo aquellas variables que sean útiles para predecir los valores de las variables dependientes Debemos también considerar que cuando las variables independientes son discretas solo pueden ser incluidas en el modelo después de ser codificada dentro de un diseño de variables indicadoras

Utilizando este modelo, podemos obtener una ecuación denominada logit, que nos permitirá predecir probabilidades de un evento de interés, además este modelo nos proporciona información adicional, a través de la construcción de intervalos de confianza, prueba de hipótesis y medidas de asociación, como la razón ods.

En este estudio utilizamos un lenguaje sencillo esperando de esta manera su compresión por cualquier lector, sin ser necesario tener grandes habilidades matemática Se subdivide en cuatro capítulos. El primero describe los aspectos fundamentales del estudio El segundo capítulo se trata aspectos generales sobre la desnutrición infantil y algunos factores que influyen en la misma

El tercer capítulo trata sobre la metodología empleada en nuestro caso, es sobre el Modelo de Regresión Logística y otros aspectos referentes a la base de datos utilizada En el cuarto capítulo se muestra el análisis de la información de ENV 97 a través de cuadros estadísticos y gráficas, el análisis de datos a través del modelo de regresión logística

Finalmente en base a los resultados obtenidos procederemos a presentar las conclusiones y recomendaciones.

Esperamos que este trabajo de graduación sirva de referencia a estudiantes de la escuela de estadística e investigadores que requieran aplicar esta técnica.

CAPÍTULO I

ASPECTOS FUNDAMENTALES DEL ESTUDIO

1.1. PLANTEAMIENTO DEL PROBLEMA

El fenómeno de la desnutrición afecta a nivel nacional al 16% de los niños y niñas panameños menores de cinco años (cerca de 50,000) Sin embargo, existen marcada diferencias en su prevalencia por área En el área indígena, este fenómeno es más alarmante, ya que más de la mitad 50.5% de los niños y niñas sufren de desnutrición, le sigue el área rural donde cerca del 24.4% de los niños y niñas están desnutrido, mientras que en el área urbana los niños y niñas en ésta condición disminuye al 7. 1 %

La desnutrición infantil, tiene serias repercusiones e implicaciones a largo plazo, debido a que se trata de una enfermedad, cuyo poder destructivo se percibe a través de diversas secuelas que van desde la mayor propensión a contraer enfermedades durante toda la vida, el padecer algún tipo de discapacidad (retardo mental entre otras) y la merma de la capacidad intelectual del niño y niña que la padece Este tipo de secuelas, ubica en clara desventaja social y económica a los niños y niñas afectados, ya que esta condición los predispone a incursionar en el mercado laboral en actividades menos lucrativas y ser menos productivos en la vida adulta, perpetuando así el ciclo intergeneracional de la pobreza Además, la desnutrición constituye una flagrante violación de los derechos de los niños y niñas a la supervivencia

En cuanto a la mortalidad, los niños y niñas desnutridos por su deteriorado estado nutricional son susceptibles a las infecciones y otras enfermedades lo que agrava su situación produciéndoles en algunos casos la muerte debido a su baja resistencia a las mismas Los que sobreviven son niños y niñas débiles, decaídos que no pueden aprovechar bien la educación escolar y no alcanzan llegar a adultos, la constitución física que habrían podido tener si hubieran estado bien nutridos En el ámbito internacional las cifras estadísticas indican que cerca de la mitad de la mortalidad infantil ocurrida en el mundo se debe a la desnutrición.

Cabe destacar, que la desnutrición como fenómeno, es consecuencia de una injusta distribución de la riqueza económica, social y cultural, razón por la cual no debe entenderse únicamente desde el contexto clínico y preventivo, pues además intervienen, por lo general, otros factores de tipo demográfico, ambiental, político y reproductivo.

En este sentido, a continuación se presentan algunos datos estadísticos que caracterizan nuestra república y que inciden indudablemente sobre la producción, disponibilidad y consumo de los alimentos, y, en consecuencia, sobre el estado de nutrición del niño y niña panameño.

En el plano nacional existen algunos hechos fundamentales entre los que podemos mencionar.

- El crecimiento demográfico que ha sido acelerado en los últimos 30 años, ya que, de 1,428,082 habitantes en 1970 aumentó a 2,815,644 en 2000, es decir, un 97 1% equivalente a 1,387,562 personas, registrándose un crecimiento anual promedio de 46,252 personas
- El ingreso per cápita relativamente alto en Panamá (B/ 3,080 en 1997), pareciera ser en parte más ficticio que real, puesto que ocurre en una minoría de la población, ya que la pobreza sigue siendo un problema generalizado. Más de un millón de personas, es decir, el 37% de la población viven bajo el nivel de pobreza y de éstos, más de medio millón (19%) vive en condiciones de extrema pobreza, concentrándose la mayor parte (66 7%) de los pobres del país en el área rural Esto nos indica en gran medida el alto índice de desigualdad que existe en el país.
- El aumento anual del costo de la vida

  La canasta básica familiar que incluye sólo productos comestibles necesarios para la alimentación de una familia con un promedio de 5 integrantes, se estima en B/ 224 00 para 2001 Sin embargo, esta suma es comparable con el salario mínimo mensual que gana un gran número de panameños Al cual no se le ha incluido el aumento por concepto en servicios de luz, teléfono, agua, combustibles, transporte y otros servicios

- La grave deficiencia institucional de la agricultura, ya que sólo un tercio de toda la tierra agrícola en manos privadas cuenta con títulos de propiedad completo y el 84% de estas tierras son propiedad de los no pobres Además, el papel que juega la agricultura en Panamá es pequeño, representando sólo el 7% del PIB y el 22% del empleo total
- La desigualdad en la distribución de la tierra, pues a pesar de ser los pobres el segmento poblacional mayoritario (64 9%) en las áreas rurales, sólo son dueños de un tercio de la tierra
- El descenso en el cultivo y producción de arroz, maíz y frijoles productos importantes en la dieta del panameño En el año agrícola 1999-2000 la cosecha de arroz disminuyó en 137,100 quintales respecto al año 1996-1997 que fue de 5,082,900 quintales. Por su parte, la cosecha de maíz y frijol también disminuyeron en 349,200 y 7,800 quíntales respectivamente, con relación al año agrícola 1985-1986 Igual situación se observó en el total de superficie sembrada para los tres rubros

  Como podemos apreciar el cultivo de éstos tres rubros a disminuido y sin embargo, durante esos mismos años ha habido un crecimiento de la población, así se llega inevitablemente a un desequilibrio y a una más difícil disponibilidad de los alimentos
- El factor cultural principalmente de la madre influye también en la desnutrición Así por ejemplo, a causa de una mala educación, la madre introduce tardíamente la carne (proteínas animales) en la dieta del niño o niña, en lugar de hacerlo a los nueve o diez meses, lo hace a los tres años cuando ya los signos de desnutrición son evidentes, o cuando ya no hay niño o niños, aunque sigue habiendo desnutrición familiar

- El grado apreciable en la razón de dependencia, en donde un tercio de nuestra población se hace cargo de los otros dos tercios, lo que es sin duda, una carga pesada Una desproporción tal, se traduce inevitablemente en mayor desnutrición y morbilidad en la población, principalmente en los niños y niñas

Por otra parte, la finalidad de este trabajo es medir el efecto de factores sociodemográficos y su repercusión en el estado nutricional del niño y la niña, así como la relación que existe entre ciertas variables exógenas o del medio ambiente y el estado nutricional del niño o la niña

1.2. JUSTIFICACIÓN

El buen estado nutricional es la base de la salud Una persona bien nutrida es activa, su organismo funciona bien, es más resistente a las enfermedades, alcanza mejor desarrollo físico y tiene probabilidades de alcanzar una edad más avanzada

En los niños y niñas, la desnutrición ataca especialmente aquellos que tienen una ingesta inadecuada de alimentos, así como aquellos que no están protegidos contra las enfermedades frecuentes y no reciben atención médica apropiada Sin embargo, cada tipo de desnutrición es resultado de una compleja interacción de factores que incluye aspectos como el grado de acceso de las familias a los alimentos, así como a los servicios básicos que promueven el desarrollo humano

Cada tipo de desnutrición mina y destruye el organismo humano de manera diferente, la desnutrición leve, provoca el retardo en el crecimiento y maduración del niño o la niña, es inclusive la causa de que gran número de niños y niñas mueran a temprana edad, la carencia de micronutriente como· la vitamina A, causa ceguera y afecta el sistema inmunológico y aumenta la propensión a las diarreas, la carencia de yodo tiene efecto en el retardo del crecimiento y merma la capacidad intelectual, la deficiencia de vitamina D en el organismo es la causa de la presencia en el niño o niña del raquitismo infantil, que se manifiesta primeramente a nivel del sistema nervioso, la anemia por carencia de hierro afecta el desarrollo psicomotor al igual que el desarrollo cognoscitivo que se refleja en la reducción del cociente de inteligencia

Como podemos observar, la desnutrición priva al niño o a la niña no sólo en el aspecto físico, sino también en el intelectual, lo que significa que aquellos niños y niñas desnutridos que superen la infancia enfrentarán un futuro de carencias Esto conlleva, a que la atención al niño y la niña se realice desde el punto de vista de la medicina clínica, preventiva y social; entenderlo en su crecimiento y desarrollo físico, emocional y espiritual.

Tratándose la desnutrición de un estado susceptible de prevención no es comprensible que nuestra sociedad no destine los recursos necesarios para reducir su prevalencia, sobre todo en los niños y niñas en etapa de crecimiento quienes tienen derecho a vivir sanamente y quienes son el futuro del país

Ante esta panorámica, no cabe duda de la transcendencia e importancia de este estudio sobre la desnutrición infantil y del aporte que este trabajo de investigación brindará a la comunidad en general Ya que el hambre no es una condición transitoria, sino que es crónica, debilita y a veces es mortal, malogra la vida de los afectados, y tiene sus efectos negativos en la economía

Diversos estudios han tratado este tema lo que demuestra la gran relevancia del mismo, pero ninguno hasta la fecha, en nuestro país, ha utilizado como técnica estadística de análisis el modelo de regresión logística, cuya aplicación es recomendada por la Organización

Panamericana de la Salud (OPS) en materia de desnutrición Por lo que consideramos que el presente estudio servirá de referencia a investigadores que trabajan con esta problemática, estudiantes y profesionales de la estadística que deseen aplicar esta técnica de regresión

1.3. OBJETIVO GENERAL

Conocer la situación nutricional de la población infantil menor de 5 años en el ámbito nacional, mediante la medición de su magnitud y determinación de los factores asociados e influyentes con este fenómeno, usando como técnica estadística el análisis de regresión logística.

1.3.1. Objetivos Específicos

- Conocer algunas características de tipo social y económica de la población infantil menor de 5 años.
- Medir la magnitud de los factores de riesgos asociados a la desnutrición
- Construir un modelo de regresión logística multivariado con las variables que mejor explican la situación nutricional de la niñez menor de 5 años con el fin de realizar inferencias acerca de la población

1.4. HIPÓTESIS

H_1 · Niños y Niñas mayores de 1 año de edad tienen más probabilidad de presentar un deterioro en su estado nutricional

H_2 Niños y niñas que residen en áreas rurales e indígena es más probable que presente deficiencia en su estado nutricional.

H_3 Niños y niñas afiliados a un sistema de seguridad social es más probable que presente un mejor estado nutricional.

H_4 Es más probable que las niñas presente un estado nutricional deficiente con relación a los niños

1.5. LIMITACIONES

El presente estudio ha significado un gran esfuerzo, sobre todo en el ordenamiento y depuración de los datos utilizados, nos fue difícil manipular la base de datos por la forma como fue diseñada la misma Esto no permitió que se incluyeran variables de interés como los factores maternos que hubieran dada una mejor explicación de este fenómeno

Por otra parte, la escasa o nula bibliografía sobre el modelo de regresión logística en nuestro país, nos limito significativamente en la compresión del mismo, lo cual incidió en una demora en la realización de este trabajo

Pese a estas limitaciones consideramos que los resultados son significativos desde el punto de vista académico y de la realidad nacional, ya que en el análisis se incluye información de la población infantil menor de 5 años de todo el país

CAPÍTULO II

LA DESNUTRICIÓN INFANTIL

2.1. ANTECEDENTES HISTÓRICOS DE LA DESNUTRICIÓN

La alimentación ha sido una de las necesidades y preocupaciones fundamentales del hombre y uno de los factores determinantes de la formación y progreso de las sociedades Los hombres primitivos dependían para su alimentación de la caza, de la pesca y de la recolección de productos vegetales silvestres, vivían en forma nómada y organizada en pequeños grupos para poder efectuar más eficazmente esas actividades Las primeras organizaciones sociales sedentarias fueron posibles cuando el hombre aprendió a domesticar animales y a cultivar plantas para la obtención de sus alimentos, de allí se ha progresado hasta la constitución de las naciones y grandes ciudades de millones de habitantes, tales como las conocemos ahora, gracias a los adelantos en los sistemas de producción, conservación y distribución de alimentos.

La selección de los alimentos se hizo primero con el propósito inicial de satisfacer el hambre y estuvo condicionada por la existencia de ellos en cada renglón y por la experiencia acumulada en relación con el gusto y la inocuidad de los productos

La importancia de la alimentación ha sido reconocida en la medicina desde el origen de esta ciencia, ya Aristóteles advertía que no todos los alimentos son adecuados para todas las personas lo que dependía de su estado de salud, pero siempre se dio mucha mayor importancia a la relación entre alimentación y salud en sus aspectos negativos.

Entre los pueblos antiguos como los hebreos, griegos, romanos, chinos e hindúes ha existido la preocupación por proteger a sus habitantes Se tienen evidencias de relaciones y derechos internacionales al respecto, así como de la preocupación por dar al niño y a la niña una garantía de subsistencia y educación para lograr un desarrollo adecuado, aunque los pueblos fueran guerreros

Respecto a la normativa y derecho internacional, los países de la antiguedad, los contemplaron en sus respectivos códigos y los establecían en alianzas, tratados, acuerdos y convenios

Según el Dr Marco Antonio Jaime Vela, en los pueblos prehispánicos de América las culturas Maya, Azteca e Inca, mostraron en sus códigos, toda una normativa de protección a la infancia, tanto en educación como en alimentación, y aunque esta última, actualmente nos pudiera parecer deficiente, la desnutrición en niños, niñas sólo se presentaba por enfermedad, no por falta de alimentos

Es en la revolución Francesa (1789) con la declaración de los derechos del hombre y del ciudadano, que se habla de garantías universales, convirtiéndose en pionera del Derecho Internacional Humanitario, se empieza a presentar principalmente en los países africanos y asiáticos con su lema ***Igualdad Libertad y Fraternidad para todos los hombres y pueblos del mundo***. Promoviendo la cooperación y ayuda de las naciones pudientes, a los pueblos más necesitados, para librarlos de la opresión y miseria

En los países del resto del mundo, desde el inicio del siglo hasta la fecha, son la sobrepoblación, las guerras, el desempleo, los desastres naturales y los desequilibrios económicos y políticos los causantes de que se incremente el problema de la desnutrición, haciendo crisis en la primera y segunda guerra mundial

En pleno siglo XXI, hay países en que la desnutrición infantil alcanza cifras alarmantes, y la muerte por hambre es una realidad

2.2. CONCEPTO

La desnutrición normalmente es el resultado de la combinación de una ingesta alimentaría inadecuada La desnutrición no debe atribuirse a la subalimentación, sino que es el resultado palpable de la capacidad individual de utilización de los alimentos ingeridos Además de la cantidad de alimentos ingeridos, se considera el estado sanitario global del individuo En este sentido, la desnutrición es un indicador de salud alimentaría [1]

La desnutrición es una enfermedad caracterizada en la mayoría de los casos por carencia alimentaría, acompañada por ausencia de estimulación psicoafectiva, siendo la pobreza la causa estructural Se manifiesta con peso y talla inferiores a los valores esperados para la edad

El grupo mas expuesto, es el de los lactantes y el de los niños en edad preescolar, ya que este periodo de la vida se caracteriza por un rápido crecimiento, que exige un consumo mayor de calorías y de nutrientes

Las familias indigentes sufren la escasez en la disponibilidad de alimentos como consecuencia del desempleo o de su inserción en actividades informales de baja productividad económica Estos hogares presentan grandes carencias de agua potable y saneamiento

En los niños la desnutrición es sinónimo de deficiencias en el crecimiento, ya que los niños desnutridos tienen una estatura y un peso menores de lo que deberían tener atendiendo a su edad. Para conseguir una medida rápida de la desnutrición en una población, debe medirse y pesarse a los niños y compararse después los resultados con los de la "población de referencia", de la que se sabe que ha crecido correctamente. Pesar y medir la estatura son las formas más comunes de evaluar la desnutrición en la población

La desnutrición en los niños y las niñas preescolares y escolares les afectan directamente en el crecimiento físico, intelectual y desarrollo motor retardado

---

1 JAIME Marco Antonio, Desnutrición, Fundación Mexicana para la Salud, 1997, 2 pág

2.3. TIPOS CLÍNICOS

Se distinguen los siguientes tipos clínicos de desnutrición desnutrición calórica (marasmo), proteica (Kwashiorkor o pluricarencial) y mixta

__ Desnutrición Calórica (Marasmo)

Se puede observar en el lactante menor, pero su incidencia es más alta entre el año los dos años de edad El déficit es principalmente energético y se instala lentamente, de modo progresivo.

Existe enflaquecimiento, por disminución del tejido adiposo y, luego, del muscular. Clínicamente se observa disminución de peso, luego de la talla y del perímetro craneano La piel aparece como arrugada, por disminución del tejido celular subcutáneo En casos avanzados se puede llegar al marasmo propiamente tal, con mayor grado de enflaquecimiento, adquiriendo el niño "cara de viejo", sobresaliendo los ojos y pómulos

El niño puede observarse irritable, con llanto monótono, o apático No obstante contrasta el déficit nutricional, que puede ser intenso, con una mirada vivaz En general no hay disminución de la albúmina y, salvo casos extremos, los índices bioquímicos son relativamente normales

El niño, con esta desnutrición como se indico anteriormente, se desnutre por déficit en el aporte de nutrientes, en su cantidad y/o calidad, respecto a sus necesidades Aquí la causa es socio génica El menor nace sano pero, de alguna manera, por influencia de un medio adverso familiar y social, no recibe la alimentación y la estimulación sensoriomotriz que requiere para su normal crecimiento y desarrollo

El médico se orienta hacia desnutrición de causa calórica cuando encuentra que el niño tiene un entorno deprimente Habitualmente se trata de una familia en condición de extrema miseria, con serios problemas económicos y, especialmente, culturales Puede obtenerse el antecedente de existir hermanos con desnutrición, de tratarse de una familia con unión inestable o sin unión, haber irresponsabilidad o apatía en la conducta de los padres respecto de su(s) hijo(s); consumo excesivo de alcohol y/o drogas, baja escolaridad y edad materna (adolescencia), trabajo inestable o cesantía del padre, problemas serios en el tipo de vivienda, en saneamiento básico, en higiene dentro y fuera del hogar, presencia de hacinamiento.

En la anamnesis alimentaría destaca la lactancia materna de escasa duración, ausente o excesivamente prolongada, introducción inoportuna o inadecuada de alimentación láctea artificial y alimentos sólidos

__ <u>Desnutrición Proteica (Kwashiorkor o Pluricarencial)</u>

Se origina por déficit de proteínas en la dieta, puede ser también consecuencia de una enfermedad celíaca u otra patología que determine diarrea crónica La presencia de edema es cardinal para el diagnóstico y se observa con albuminemias bajo 3%. El edema es distal, en las extremidades y cara, puede también generalizarse

Existen alteraciones en la piel y faneríos que son importantes. El pelo es seco, fino decolorado en bandas (signo de la bandera) y se cae fácilmente Toma habitualmente un aspecto pajizo Se puede encontrar hiperqueratosis y descamación de la piel, ocasionalmente hay alteraciones en la pigmentación, atrofia de las papilas de la lengua, ulceraciones en los ángulos de los labios y de los ojos, fisuras y sangramiento de las encías

Comúnmente hay hepatomegalia (hígado graso) A diferencia de la desnutrición calórica, en el kwashiorkor el niño está muy apático, indiferente al medio, siendo difícil interesarlo, incluso con la alimentación. Los índices bioquímicos se alteran gravemente (hiponatremia, hipokalemia, hipocalcemia, disminución del tiempo de protrombina, anemia y otros)

En el trabajo médico es fundamental determinar la(s) causa(s) de la desnutrición del niño Ya que ésta puede deberse a afecciones que interfieran con la ingesta de los alimentos como estenosis pilórica, reflujo gastro esofágico, hernia diafragmática y otras causas de vómitos, patologías que afectan la absorción de los alimentos como, por ejemplo, enfermedad celíaca, fibrosis quística y otras causas de mala absorción intestinal; enfermedades que alteran el transporte o la utilización de los alimentos a nivel tisular, (enfermedades metabólicas, hipotiroidismo), patologías que aumentan los requerimientos nutricionales y/o determinan hipercatabolismo, (enfermedades infecciosas graves o prolongadas, ciertas cardiopatías, enfermedades malignas, nefropatías y otras) En ocasiones una afección puede desnutrir a un niño por mecanismos múltiples, por ejemplo, si determina diarrea, vómitos, temperatura elevada, hipercatabolismo, etc

. __ Desnutrición Mixta

Los tipos clínicos someramente descritos están interrelacionados entre sí y corresponden a los extremos de formas intermedias En la desnutrición mixta se observa edema y paralelamente, enflaquecimiento del niño, por aporte insuficiente de calorías y de proteínas, de modo paralelo o sucesivo

2.4. ESTRATEGIAS, PREVENCIÓN Y TRATAMIENTO

Las acciones de prevención y tratamiento de la desnutrición infantil han sido prioritarios para la salud pública Las medidas de protección materno infantil, de amplia cobertura, y focalizadas a los grupos de mayor riesgo, resultan fundamentales en la prevención de la desnutrición infantil La mejoría del saneamiento ambiental, de la calidad de vivienda, trabajo, educación, ingresos, y otros, contribuyen importantemente a disminuir su incidencia

El tratamiento de los casos individuales requiere de un enfoque interdisciplinario, que incluye la mejoría del entorno familiar y social, conexión a redes de apoyo comunitarias; educación, prescripción y suplementación alimentaría, estimulación sensoriomotriz y afectiva, apoyo psicológico a la familia, así como un adecuado manejo médico de los trastornos metabólicos producidos por la desnutrición, patologías asociadas o determinantes de ella

El niño con desnutrición severa requiere ser hospitalizado en muchos casos para su estudio y tratamiento inicial Cuando las condiciones del hogar son extremadamente adversas, es necesario sacar al niño en forma transitoria de su medio e incorporarlo a un hogar sustituto (colocación familiar) o ingresarlo a un centro de recuperación nutricional, mientras el menor se recupera en las áreas físicas y psicomotoras, y se logran condiciones familiares y ambientales más favorables

El tratamiento de la desnutrición calórica requiere aumentar gradualmente los aportes calóricos y proteicos del paciente, lo más rápido que la tolerancia digestiva del niño lo permita, hasta llegar a un 150 a 200% de las recomendaciones por kilo de peso. Los requerimientos de agua, sodio y potasio pueden estar inicialmente aumentados Debe efectuarse suplementación con multivitamínicos, hierro y zinc para evitar carencias de estos elementos en la etapa de recuperación (crecimiento rápido).

La realimentación inicial debe ser cautelosa, y de preferencia con fórmulas sin lactosa Si la vía digestiva no permite aportar nutrientes adecuadamente, se puede considerar la posibilidad de indicar alimentación parenteral. Ante cualquier sospecha de infección bacteriana deben tomarse hemocultivos e iniciar tratamiento con antibióticos de amplio espectro.

Una vez recuperado el estado general y digestivo del niño éste puede realimentarse con aportes similares a los de la desnutrición calórica

UNICEF propone lo que denomina las seis oportunidades concretas para proteger la vida y el desarrollo normal de la infancia en prácticamente todos los países en desarrollo durante la próxima década

_ **Inmunización universal de la infancia** La cobertura de esta acción ya abarca 2/3 de la población infantil del mundo en desarrollo, lo que permite salvar aproximadamente 2 millones de vidas cada año, sin embargo, cerca de tres millones de niños siguen muriendo cada año por no estar vacunados y porque las enfermedades, la desnutrición y la mortalidad son más frecuentes entre los niños aún no inmunizados, por ello es esencial mantener el esfuerzo y alcanzar una cobertura de inmunización del 80% .

_ **Terapia de rehidratación oral (TRO)**. Según un informe publicado por la OMS (Organización Mundial de la Salud), basado en datos procedentes de 46 países, sólo un 14% de los médicos, un 4% de las enfermeras y un 8% del personal paramédico y un 9% de los agentes comunitarios de salud están capacitados para aplicar el tratamiento de rehidratación oral, pese a ser calificarlo como "el avance médico potencialmente más importante de nuestro siglo".

- **Las infecciones respiratorias agudas**. La cuestión de si debe facilitarse la distribución de antibióticos al nivel de la atención primaria a través de los agentes de salud, es todavía objeto de debate; sin embargo, la OMS y el UNICEF considera que ya se han acumulado datos suficientes que demuestran la utilidad de estos medicamentos

_ **Lactancia materna** La lactancia materna parece estar en retroceso en muchos países en desarrollo, por la influencia de presiones comerciales, el uso de leche en polvo en los hospitales y la creciente incorporación de la mujer al mercado de trabajo, que contribuyen a presentar la alimentación con biberón como una alternativa atractiva Se ha demostrado que los niños alimentados con biberón contraen muchas más enfermedades y tienen una probabilidad hasta 26 veces superior de morir en la infancia, que los niños alimentados exclusivamente con leche materna durante los primeros 6 meses de vida

_ **Espaciamiento de los nacimientos**. Los datos indican que la mayor parte de la mortalidad corresponde a casos en que el número de nacimientos es superior a cuatro, o entre uno y otro media un espacio de menos de 2 años, o el embarazo se produce antes de los 18 o después de los 35

_ **Ataque contra la desnutrición**. Se argumenta que la desnutrición tiene raíces tan profundas en el terreno de la pobreza que sólo el desarrollo económico podría arrancarlas

CAPITULO III

METODOLOGÍA

3.1. ESTIMACIÓN DEL ODS Y RAZÓN ODS

__ **Definición de Ods:**

La ods consiste en el cociente entre dos frecuencia de una categoría de una misma variable Representa la proporción de poder pertenecer a una u otra posibilidad de la misma variable Es decir, si la probabilidad de un suceso es p, la ods a favor de que ocurra es p/(1-p) De esta manera el ods expresa el cociente entre la probabilidad de que algo ocurra (p) y la probabilidad de que no ocurra (1-p), o lo que es lo mismo, el cociente entre el número de las posibilidades "éxito" (así denominamos el resultado que nos interesa), si todos los resultados son igualmente posible·

$$n = e + f$$

Si llamamos a los éxitos, e y f a los fracasos, n al total de posibles resultados del evento, es evidente que

$$P = e/n = e/ e + f \qquad (1)$$

y también

$$ods = P / (1 - P) = e / f \qquad (2)$$

Ejemplo 1

Si de un grupo de 100 personas que asisten a un banquete, 85 presentan al día siguiente vómitos y diarrea, la probabilidad de gastroenteritis será 85/100=0 85, mientras que las ods serán 85 a 15, o sea, 85/15 = 5 7

- Estimación de la Ods (Tabla 2 x2).

A continuación haremos una aplicación para el ods usando de ejemplo los datos de la Tabla 3 1 que consiste en los resultados obtenidos luego de utilizar dos tratamientos A y B en la curación de una enfermedad

Ejemplo 2:

TABLA 3.1.
RESULTADOS DEL TRATAMIENTOS A Y B

Curación	Tratamiento A (X = 0)	Tratamiento B (X = 1)
Si	45	36
No	15	18
Total	60	54

Se define la variable tratamiento como X = 0 para tratamiento A y X = 1 para el tratamiento B, a partir de la Tabla 3.1. podemos estimar la probabilidad de curación para ambos tratamiento.

Para el tratamiento A la probabilidad de curación esta dada por:

$$P\,(X = 0) = 45/60 = 0.75$$

Para el tratamiento B la probabilidad seria:

$$p\,(\,X = 1) = 36/54 = 0.67$$

Como ambas probabilidades son distintas, parece que la probabilidad de curación depende del tratamiento. Luego el ods de curación del tratamiento A es igual a:

$$Ods_{trat.A} = 45/15 = 3$$

El ods de curación para tratamiento B es:

$$Odds_{trat.\,B} = 36/18 = 2$$

— **Razón Ods (RÔ)**

El cociente de el ods de los dos grupos es lo que se denomina razón ods y constituye otra forma de cuantificar la asociación entre dos variables dicotómicas.

Un valor de RÔ> 1 indica la existencia de una relación positiva o directa entre dos variables mientras que una RÔ < 1 señala la presencia de una relación negativa o inversa. También estaríamos ante la presencia de un factor de protección que en las ciencias de la salud se refiere a un factor que protege al organismo de sufrir alguna enfermedad, por ejemplo, el ejercicio disminuye las posibilidades de sufrir un infarto por lo que se considera un factor de protección. Un RÔ = 1 es indicativo de una ausencia de relación entre los dos variables. En los últimos años la razón ods ha sido utilizada ampliamente en informes médicos. Existe tres razones para esto:

Primero, ellas dan un estimado (con intervalos de confianza) para la relación entre dos variables binarias.

Segundo, nos permiten examinar los efectos de otras variables sobre esa relación, utilizando la regresión logística.

Tercero, tiene un modo de interpretación especial y muy conveniente en estudios de casos-control.

El calculo de la razón ods a partir de datos de frecuencia de una tabla 2 x 2 es muy sencillo:

	Buen control	Mal control
Si Suceso	a	c
No Suceso	b	d

$$R\hat{O} = \frac{axd}{bxc} \tag{3}$$

Para los datos de la Tabla 3.1, la razón ods esta dada por:

$$R\hat{O} = \frac{axd}{bxc} = \frac{45 \ x\,18}{15 \ x\,36} = \frac{810}{540} = 1.5$$

También se podría estimar la RÔ mediante el cociente de la ods para el tratamiento A y B obtenidos en el ejemplo 2 Así tenemos

RÔ= 3/2 = 1 5

Podemos calcular el error estándar para el logaritmo de la razón ods y el intervalo de confianza. La fórmula para el cálculo del error estándar del log de la razón ods se basa simplemente en la raíz cuadrada de la suma de los recíprocos de las cuatro frecuencias

$$\hat{E}E(\ln R\hat{O}) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \tag{4}$$

Cuando calculamos la razón ods debemos conocer si dicho valor es diferente de 1 Para tal efecto se construyen los intervalos de confianza, si al nivel $\alpha = 0\ 05$ el intervalo no incluye el valor 1, concluimos que el riesgo es estadísticamente significativo De obtener igual resultado para $\alpha = 0\ 01$ concluimos que es significativo

El calculo del intervalo de confianza para la razón se obtiene a través de la siguiente fórmula·

$$\exp^{[\ln R\hat{O} \pm Z_{\alpha/2} \hat{E}E(\ln R\hat{O})]} \tag{5}$$

RÔ es la estimación puntual de razón ods,

Exp es la base del logaritmo natural elevado a la cantidad entre paréntesis, y

$\hat{E}$ E (ln RÔ) = error estándar del ln de la razón ods

3.2. MODELOS LINEALES GENERALIZADOS

El modelo de regresión logística es un caso particular de los modelos lineales generalizados Estos modelos se componen de tres elementos los cuales se analizaran posteriormente

Los modelos estadísticos se pueden expresar de la siguiente forma·

Observación = Componente sistemático + componente aleatorio En notación matemática

$$Y_i = \mu_i + \varepsilon_i$$

donde

Y_i valor observado de la variable,

μ_i valor esperado de la variable y partes sistemática de la observación;

ε_i · parte aleatoria y cuyo valor esperado es cero

En el modelo lineal general, se supone que μ_i es una función lineal de las variables independientes, es decir, $\mu_i = X_i^t \beta$, donde X_i es un vector de variables independiente para la observación i-ésima y β es un vector de parámetro

Los errores se suponen variables independiente con una distribución normal con media cero y varianza σ^2

En el modelo lineal generalizado la distribución del error puede ser distinta a la normal En estos modelo se supone que existe una función g de las μ_i que es una combinación lineal de los parámetros β, es decir, $g(\mu_i) = X_i^t \beta$ En otras palabras, los modelos lineales generalizados están definidos por tres componente a saber·

Una función de probabilidad f(y) para la variable dependiente Y, que depende de la media y posiblemente de otros parámetros. Esta función pertenece a la familia exponencial (gama, poisson, binomial, normal).

- Una función lineal o predictor lineal en p variables independientes.

$$X_i^T \beta = \beta_0 X_0 + \beta_1 X_1 + \qquad + \beta_p X_p \text{ con } X_0 = 1$$

- Una transformación paramétrica o función de enlace $g(\mu)$ (denominada "link function") que relaciona el predictor lineal $X_i^T\beta$ con media μ, es decir: $g(\mu_i) = X_i^t\beta$ Esta función debe ser monótona y diferenciable

El ajuste de un modelo lineal generalizado se efectúa con el método de máxima verosimilitud

3.3. MODELO DE REGRESIÓN LOGÍSTICA

__ **Planteamiento del Problema Teórico**

Dentro de los denominados modelos de regresión se incluyen un conjunto de técnicas estadísticas que tratan de explicar como varia el valor de la variable dependiente, cuando cambian una o más variables independientes incluidas en el modelo Lo que caracteriza a los distintos modelos de regresión es la naturaleza de la variable dependiente, así, con variables continuas o categórica, el modelo de regresión lineal es la más utilizada, con variable dicotómicas lo es el modelo de regresión logística

La aplicación del modelo de regresión lineal mediante la técnica de mínimo cuadrado, funciona fluidamente desde el punto de vista matemático. Pero cuando la variable a explicar sólo puede tomar dos valores, es decir, la ocurrencia o no de un evento, al aplicar esta técnica para los valores específicos de las variables independientes se obtendrá un número que será diferente de 1 y 0 (los únicos valores posibles de la variable dependiente), valores que no es posible que se presenten En este caso, la regresión lineal debe ser descartada, en cambio la regresión logística se ajusta adecuadamente a esta situación.

Un planteamiento alternativo, la regresión logística, ha sido desarrollada para permitirnos utilizar los modelos de regresión en la predicción de la probabilidad De tener una respuesta categórica particular para un conjunto dado de variables explicativas (que pueden ser numéricas o categóricas)

Por tanto, la regresión logística, consiste en obtener una función logística de las variables independientes que permita clasificar los individuos en una de los dos subpoblaciones o grupos establecidos por los valores de la variable dependiente

La función logística es aquella que halla para cada individuo según los valores de una serie de variables (X_i) la probabilidad (p) de que presente el efecto estudiado. Una transformación logarítmica de dicha ecuación, a la que se llama logit, consiste en convertir la probabilidad (p) en ods De aquí surge la ecuación de la regresión logística, que es parecida a la ecuación de la regresión lineal múltiple

El concepto de proporciones es importante en el análisis de regresión logística, porque permite establecer las primeras diferencias entre los modelos de regresión lineal y de regresión logística En el modelo de regresión lineal se predice el valor medio de la variable dependiente a partir de una o más variables independientes (X_i). En cambio, los modelos de regresión logística permiten predecir la proporción de una de los dos categorías de la variable dependiente (Y = dicotómica) en función de una o más variables independientes

Otra limitación de la regresión lineal en el análisis de la variable dependiente dicotómica es en consecuencia la restringidad de la distribución de los valores de las dos únicas opciones posibles (1,0). La distribución de las variables dicotómicas difiere de la denominadas distribución normal estándar con la desviación típica constante El seguimiento de un patrón de distribución normal de las variables constituye una de las asunciones de la regresión lineal, por lo que no se cumple cuando la variable dependiente es dicotómica. Además, el hecho de que para cada (X_i) sólo existen dos posibles valores de los residuales o errores, viola también la denominada asunción de homostecidad, necesaria en la construcción de un modelo de regresión lineal

Ante estas limitaciones del modelo de regresión lineal para predecir la característica de interés cuando la variable dependiente es dicotómica, intentaremos su ajuste a un modelo de regresión logística multivariado que permita el análisis de la desnutrición infantil; donde la variable dependiente respuesta Y = "Estado Nutricional" puede tomar dos valores posibles Y = 1 "Desnutrido" o Y = 0 "No desnutrido"

Matemáticamente pudiéramos expresar el planteamiento del problema de la siguiente forma

Sean $X_1, \ldots, X_p$ un conjunto de variables independientes observadas con el fin de explicar y/o predecir el valor de Y

Nuestro objetivo consiste en determinar

$$P[\, Y = 1/X_1, \ldots, X_p]$$

y por lo tanto

$$P[\, Y = 0/X_1, \ldots, X_p] = 1 - P[\, Y = 1/X_1, \ldots, X_p]$$

Para ello construiremos un modelo de la forma

$$P[\, Y = 1/X_1, \ldots, X_p] = p\,(X_1, \ldots, X_p, \beta\,)$$

donde $p\,(X_1, \ldots, X_p, \beta\,) : R^k \rightarrow [\,0,1\,]$ es una función que recibe el nombre de función de enlace cuyo valor depende de un vector de parámetros $\beta = (\beta_1, \ldots, \beta_p\,)$

3.3.1. Componentes del Modelo

Sea Y la variable respuesta definida como Y = 1 cuando la característica de interés esta presente y Y = 0 cuando esta característica esta ausente Intentaremos construir un modelo de regresión logística que permita el análisis de la desnutrición infantil.

Sea el modelo.

$$\pi = P = \beta_{0,} + \beta_1 X_1 + \beta_2 X_2 + . \ + \beta_p X_p + e \qquad (6)$$

en la que β_0 representa el término independiente o constante asociado a la variable independiente

β_p representa el coeficiente de regresión asociado a la variable independiente

$X_1, X_2, \ , X_p =$ son las variables independiente

$e =$ es el componente del error aleatorio

$P =$ es la probabilidad estimada de que un individuo seleccionado al azar presente la característica de interés

Para el caso nuestro, se analizan datos de una muestra y los coeficientes de regresión de la muestra $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2 \ . \ \hat{\beta}_p)$ se usan como estimaciones de los parámetros verdaderos $(\beta_0, \beta_1, \beta_2 \quad \beta_p)$. Por consiguiente, la ecuación de regresión para el modelo de regresión lineal múltiple con p variables independiente sería

$$\hat{\pi} = P = \hat{\beta}_{0,} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + . \ + \hat{\beta}_p X_p \qquad (7)$$

donde

$$X_{ij} = \left(X_{i0}, X_{i1}, \ldots, X_{ij}\right)$$

y en la que $\hat{\beta}_0$ representa el término independiente o constante asociado a la variable independiente.

$\hat{\beta}_p$ representa el coeficiente de regresión asociado a las variables independientes

Como mencionamos anteriormente la variable respuesta es dicotómica, varia entre 0 y 1 Luego si utilizamos un modelo de regresión lineal las estimaciones resultante de π_i podrían resultar negativas o mayores que uno, lo que sería inaceptable Debido a la limitaciones mencionadas de la ecuación (7) para estimar los valores de P(Y=1), se intentara ajustar el siguiente modelo

$$\pi = P = e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \quad + \hat{\beta}_p X_p} \tag{8}$$

El modelo anterior sólo permite estimar valores de P(Y =1) > 0, pero también mayores de 1, por lo que tampoco constituye el modelo apropiado para predecir la probabilidad de la presencia de la característica de interés (desnutrido P (Y = 1)) El modelo que mejor estima tal probabilidad debido a que restringe los valores predicho es

$$\pi = P = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \quad + \hat{\beta}_p X_p}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \quad + \hat{\beta}_p X_p}} \qquad 0 \leq \pi_i \leq 1 \tag{9}$$

__ **<u>Función de Enlace (Logit)</u>**

La ecuación descrita en (9) se denomina función logística y la misma se procede a transformar Esta transformación es llamada logit y se basa en el cálculo del ods que se expresa de la siguiente forma

$$\eta_i = \ln\left[\frac{\pi_i}{1-\pi_i}\right] \quad (10) \qquad \text{ó} \qquad \eta_i = \ln\left[\frac{P_i}{1-P_i}\right] \quad (11)$$

Así tenemos,

$$\frac{p}{1-p} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \quad + \hat{\beta}_p X_p / \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 \quad + \hat{\beta}_p X_p}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \quad + \hat{\beta}_p X_p}} = e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \quad + \hat{\beta}_p X_p} \tag{12}$$

Si transformamos de forma logarítmica para linealizar la ecuación anterior, el resultado es la transformación logit

$$\eta_i = \ln\left[\frac{p}{1-p}\right] = \log it(p_i) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \quad + \hat{\beta}_p X_p \qquad (13)$$

Esta transformación tiene propiedades deseables del modelo de regresión lineal, ya que (a) se puede asumir que es lineal en los parámetros β; (b) es continua, (c) varía entre menos infinito y más infinito Esta transformación es también la función de enlace η_i (link function) del modelo lineal generalizado que estima el componente sistemático del modelo de regresión logística.

$$\eta_i = \text{logit}(p_i) = X_I^T \beta = \hat{\beta}_0 X_0 + \hat{\beta}_1 X_1 + .. + \hat{\beta}_p X_p$$

o igualmente·

$$\ln\left(\frac{p(X_1, \ ,X_p,\beta)}{1-p(X_1, \ ,X_p)}\right) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \quad + \hat{\beta}_p X_p \qquad (14)$$

La ecuación de regresión logística también puede expresarse de la siguiente forma

$$P(Y=1) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + .. + \hat{\beta}_p X_p}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + .. + \hat{\beta}_p X_p}} \qquad (15)$$

Equivalentemente

$$P(Y=1) = \frac{1}{1 + e^{-(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \quad + \hat{\beta}_p X_p)}} \qquad (16)$$

Ambas ecuaciones (15) y (16) permiten estimar la probabilidad del evento de interés cuando se reemplaza los valores estimados $\hat{\beta}_p$ en la ecuación, según los valores que asuma la variable independiente X_{ji}

Esta última ecuación (16) si son conocidos los coeficientes, es posible calcular directamente la probabilidad del proceso binomial para los distintos valores de la variable X

El ods vendrá dada por:

$$\frac{P(X_1, \quad , X_p, \beta)}{1 - P(X_1, \quad , X_p, \beta)} = e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \ldots + \hat{\beta}_p X_p} \tag{17}$$

3.3.2. Componente Aleatorio

Como en todo modelo lineal generalizado, en el modelo logístico, Y_i tiene un componente sistemático (μ_i) y uno aleatorio (ε_i) El componente aleatorio se distribuye como una binomial con media cero y varianza $\pi_i (1 - \pi_i)$ No se cumple, por tanto, la condición de homosedastecidad de un modelo lineal general, por lo que el método de mínimo cuadrado ordinario no es apropiado para estimar los parámetros. Para obtener estimaciones de los parámetros β se utiliza el método interactivo de máxima verosimilitud

El valor de la variable repuesta Y puede expresar como

$$Y = \pi(x) + \varepsilon \tag{18}$$

El componente aleatorio ε puede asumir uno de los dos posibles valores Si $Y = 1$, entonces $\varepsilon = 1 - \pi(x)$ con probabilidad $\pi(x)$, y si $Y = 0$, entonces $\varepsilon = -\pi(x)$ con probabilidad $1 - \pi(x)$ Así, el componente aleatorio se distribuye como una binomial con media 0 y varianza $\pi_i (1 - \pi_i)$ Este resultado se demuestra a continuación·

Demostración:

La función de probabilidad del error esta dada por:

$$f(\varepsilon) = \pi_i^{(1-\pi_i)}(1-\pi_i)^{\pi_i}$$

Luego la función generatriz de momento esta dada por:

$$\Psi(t) = E(e^{t\varepsilon}) = e^{t(1-\pi i)}\pi_i + e^{-t\pi i}(1-\pi_i)$$

$$\Psi(t) = \pi_i\left[1 + t(1-\pi_i) + \frac{t^2(1-\pi_i)^2}{2!} + \ldots + \frac{t^3(1-\pi_i)^3}{3!} + \ldots\right] + (1-\pi_i)\left[1 - \pi_i + \frac{t^2\pi_i^2}{2!} - \frac{t^3\pi_i^3}{3!} + \ldots\right]$$

Derivando $\Psi'(t)$ con respecto a t:

$$\Psi'(t) = \pi_i\left[(1-\pi_i) + t(1-\pi_{ji})^2 + 3\frac{t^2(1-\pi_i)^3}{2!} + 4\frac{t^3(1-\pi_i)^4}{3!} + \ldots\right] + (1-\pi_j)\left[-\pi_i + \frac{t\,\pi_i^2}{2!} - 3\frac{t^2\pi_i^3}{3!} + 4\frac{t^3\pi_i^4}{4!} + \ldots\right]$$

Evaluando la primera derivada $\Psi'(t)$ en cero:

$$\Psi'(t) = \pi_i(1-\pi_i) + (1-\pi_i)(-\pi_i)$$

$$= (1-\pi_i)(-\pi_i + \pi_i)$$

$$\therefore\ \mu(\in) = 0$$

La segunda derivada de Ψ (t) con respecto a t:

$$\Psi''(t) = \pi_i\left[(1-\pi_i)^2 + t(1-\pi_i)^3 + \frac{t^2(1-\pi_i)^4}{2} + \ldots\right] + (1-\pi_i)\left[\pi_i^2 - t\,\pi_i^3 + \frac{t\pi_i^4}{2} + \ldots\right]$$

Evaluando $\Psi''(t)$ en cero:

$$\Psi''(t) = \pi_i(1-\pi_i)^2 + \pi_i^2(1-\pi_i)$$

$$= (1-\pi_i)[\pi_i(1-\pi_i) + \pi_i^2]$$

Luego la varianza esta dada por.

$$V(\varepsilon) = E(\varepsilon^2) - [E(\varepsilon)]^2$$
$$= (1 - \pi_1)\pi_1 - o^2$$
$$\therefore \quad V(\epsilon) = (1 - \pi_1)\pi_1$$

3.3.3. Variable Dependiente

El modelo logístico es un caso particular de los modelos lineales generalizados, como ya se mencionó Este modelo se usa cuando se tiene una variable dependiente dicotómica o binaria de la forma

$$Y_i = \begin{cases} 1 \text{ el evento ocurre, con probabilidad } \pi_1 \\ 0 \text{ el evento no ocurre, con probabilidad } 1 - \pi_1 \end{cases}$$

La variable respuesta Y se distribuye como una binomial con función de probabilidad

$$f(y) = \pi_i^{y_i}[1 - \pi_i]^{1-y_i}$$

con media y varianza igual a

$$E(Y_1) = \pi_1 \qquad \text{y} \qquad \sigma^2 = \pi_1(1 - \pi_1)$$

<u>Demostración.</u>

Puesto que la variable Y puede asumir dos valores 1 y 0 con probabilidad π_1 y $(1 - \pi_1)$ respectivamente La función generatriz Y es

$$\Psi(t) = E(e^{ty}) = e^t\pi_1 + e^0(1 - \pi_1)$$
$$= e^t\pi_1 + (1 - \pi_1)$$
$$= \pi_1\left(1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \quad\right) + (1 - \pi_1)$$
$$= \pi_1 + \pi_1 t + \pi_1\frac{t^2}{2!} + \pi_1\frac{t^3}{3!} + \quad + (1 - \pi_1)$$

El primer momento alrededor del origen se define como $\Psi'(0) = E(Y)$ derivaremos $\Psi(t)$ con respecto a t.

$$\psi'(t) = \pi_i + \frac{2\pi_i t}{2!} + \frac{3\pi_i t^2}{3!} + \ldots +$$

Evaluando la primera derivada $\Psi'(t)$ en cero

$$\Psi'(0) = \pi_1$$

$$\therefore \quad E(Y_1) = \pi_1$$

El segundo momento alrededor del origen se define como $\Psi''(0) = E(Y^2)$ Luego la segunda derivada esta dada por

$$\Psi''(t) = \pi_1 + \pi_1 t + 3\pi_1 t^2 + \ldots$$

Evaluando $\Psi''(t)$ en cero

$$\Psi''(t) = \pi_1$$

Por definición de varianza tenemos

$$V(y) = E(Y^2) - [E(Y)]^2$$

$$= \pi_1 - \pi_1^2$$

$$= \pi_1 (1 - \pi_1)$$

$$\therefore \quad \sigma^2 = \pi_1 (1 - \pi_1)$$

3.3.4. Estimación de los Parámetros

Los parámetros β_i se estiman mediante el método de máxima verosimilitud, calculando las derivadas parciales respecto a cada una de las β_i e igualándola después a 0 El sistema de ecuaciones resultantes se puede resolver por método numéricos para obtener los estimadores máximo verosímiles de los β_i

Si Y_1 , Y_n son variables aleatorias binomiales independientes con $E(Y_1) = n_1 \pi (X_1)$ donde $n = n_1 + \quad + n_1$ La probabilidad conjunta de la función de $(Y_1, \quad , Y_n)$ es

$$\prod^{I} \pi(x_i)^{Y_i} (1 - \pi(x_i))^{n_i - y_i} \tag{19}$$

$$\prod_{j=1}^{I} \left[\frac{e^{x_i \beta}}{1 + e^{x_i \beta}} \right]^{y_i} \left[\frac{1}{1 + e^{x_i \beta}} \right]^{n_i - y_i} \tag{20}$$

Aplicando logaritmo natural tenemos

$$L(\beta) = \sum_{j=1}^{I} [n_i - y_i) \log(1 - \pi_i(x_i) + y_i \log \pi(x_i)]$$

$$L(\beta) = \sum_{j=1}^{I} [n_i \log(1 - \pi(x_i) + Y_i(\log \pi(x_i) - \log(1 - \pi(x_i))]$$

Reemplazando $\pi (X_1)$ por su valor

$$\pi(X_i) = \frac{e^{\sum_{ij} \beta_{ij} X_{ij}}}{1 + e^{\sum_{ij} \beta_i X_{ij}}}$$

Y aplicando la propiedad de logaritmo

$$L(\beta) = \sum_i [(-n_i \log(1 + e^{\sum_{ij} \beta_{ij} X_{ij}}) + y_i \log(1 + e^{\sum_{ij} \beta_{ij} X_{ij}}) + y_i \beta_{ij} X_{ij} - y_i \log(1 - e^{\sum_{ij} \beta_{ij} X_{ij}})]$$

$$= \sum_i [(-n_i \log(1 + e^{\sum_{ij} \beta_{ij} X_{ij}}) + y_i \sum_{ij} \beta_{ij} X_{ij}]$$

$$= \sum_i y_i \sum_j \beta_{ij} X_{ij} - \sum_j n_i \log(1 + e^{\sum_{ij} n_{ij} X_{ij}})$$

$$= \sum_j (y_i x_{ij}) \beta_{ij} - \sum_i n_i \log(1 + e^{\sum_{ij} \beta_{ij} X_{ij}})$$

Derivando la ecuación de verosimilitud por diferenciación de L con respecto a los elementos de β y igualando el resultado a cero

$$\frac{\partial L}{\partial \beta_a} = \sum_i y_i x_{ia} - \sum_i n_i x_{ia} \left(\frac{\exp(\sum_i \beta_j x_{ij})}{1 + \exp(\sum_i \beta_j x_{ij})} \right)$$

La ecuación de verosimilitud es.

$$\sum_i y_i x_{ia} - \sum_i n_i \pi_i x_{ij} \qquad a = 0, \ldots, p \tag{21}$$

Si X denota una matriz de I x (p + 1) de valores de (X_{ij}) La ecuación de verosimilitud anterior tiene la forma

$$X'y = X'm \qquad y \qquad m_i = n_i \pi_i$$

La matriz de información obtenida de la segunda derivada parcial de log verosimilitud es negativa Bajo condiciones regulares, los estimadores de los parámetros matriz de covarianza es igual a la matriz de información inversa para el modelo de regresión logística

$$\frac{\partial^2 L}{\partial \beta_a \partial \beta_b} = -\sum_i \frac{x_{ia} x_{ib} n_i \exp(\sum_j \beta_j x_{ij})}{\left[1 + \exp(\sum_j \beta_j x_{ij}) \right]^2} = -\sum_i x_{ia} x_{ib} n_i \pi_i (1 - \pi_i) \tag{22}$$

Nosotros estimamos la covarianza por sustitución de β dentro de la matriz que contiene elementos iguales a los negativos (la anterior) e invirtiéndola La estimación de la matriz de covarianza tiene la forma

$$Cov(\beta) = [X' Diag(n_i \pi_i (1 - \pi_i)) X]^{-1} \tag{23}$$

donde Diag [n_i π_i (1 - π_i)] denotada una matriz diagonal I x I teniendo elementos

$\{n_i\ \pi_i(1 - \pi_i\ )\}$ en la diagonal principal La raíz cuadrada de los elementos de la diagonal principal son los errores estándar de los parámetros del modelo

__ **Distribución Muestral de los Estimadores**

Los parámetros estimados tienden según el teorema del límite central en la estimación por máxima verosimilitud sugiere que estos estimadores son asintóticamente normales y su matriz de varianza- covarianza es

$$\Sigma = -\ J^{-1} = -\ (X'IX)^{-1} \tag{24}$$

y su estimación se calcula, particularizando Σ para los coeficientes estimados Recordar que las varianzas de los coeficientes están en la diagonal de esta matriz.

3.4. ESTIMACIÓN DE LOS INTERVALOS DE CONFIANZAS

Un intervalo de confianza al (1 - α)% para el coeficiente β_i es

$$\hat{\beta}_i \pm Z_{\alpha/2} E\hat{E}(\hat{\beta}_i) \tag{25}$$

Para evaluar la significancia de la razón ods se utiliza el intervalo de confianza, que para muestras grandes es evidente que dichos intervalos están dado por

$$e^{\hat{\beta}_i \pm Z_{\alpha/2} E\hat{E}(\hat{\beta}_i)} \tag{26}$$

Una razón ods es significativamente diferente de 1 a un nivel de significancia del 5%, si el intervalo calculado no incluye el valor 1, es decir si los límites exceden a 1 o son menores que 1

3.5. ANÁLISIS DEL MODELO ESTIMADO

Después de la estimación de los coeficiente del modelo debemos asegurar la significancia de la variable en el modelo. Generalmente involucra la formulación de hipótesis estadística que determina si la variable independiente en el modelo son "significativa" prediciendo la variable respuesta Entre estas pruebas estadística se encuentra la Prueba G o de razón de verosimilitud y la prueba de Wald siendo la primera con la que iniciaremos su descripción

— **Razón de Verosimilitud**

Este estadístico sirve para determinar si las variable incluidas en el modelo de manera global son estadísticamente diferentes de cero En este caso se compara la verosimilitud del modelo sin la variable y con la variable y se establece una prueba basada en la diferencia entre los coeficientes de verosimilitud de los dos modelos

En el caso multivariando bajo la hipótesis de H_0 $\beta_1 = \beta_2 = \quad = \beta_p = 0$, la distribución de G será una chi-cuadrado con grados de libertad igual a la diferencia en número de parámetros de los modelos comparados estadísticamente esta dada por

$$G = -2\ln\left[\frac{(\text{Verosimilitud del modelo sin la variable})}{(\text{Verosimilitud del modelo con la variable})}\right] \qquad (27)$$

Para el caso específico del modelo simple es fácil demostrar que cuando la variable no están en el modelo, la estimación de $\hat{\beta}_0$ es $\ln(n_1 / n_0)$ donde $n_1 = \Sigma Y_i$ y $n_0 = \Sigma(1 - Y_i)$ y que el valor predicho es constante, n_1 / n En este caso el valor de G es como sigue:

$$G = -2\ln\left[\frac{(n_1/n)^{n_1}(n_0/n)^{n_0}}{\prod_{i=1}^{n}\pi_i^{y_i}(1-\pi_i)^{(1-y_i)}}\right] \qquad (28)$$

O su equivalente

$$G = 2\left\{\sum_{i=1}^{n}\left[y_i \ln(\hat{p}_i) + (1-y_i)\ln(1-\hat{p}_i)\right] - \left[n_1 \ln(n_1) + n_0 \ln(n_0) - n\ln(n)\right]\right\} \quad (29)$$

Demostración

Sea

$$\pi(x_i)^{y_i}[1-\pi(x_i)]^{1-y_i} \qquad y_i = 0,1 \quad (30)$$

Luego la función de log verosimilitud del modelo sin la variable (β_1=0) esta dada

$$L(\beta) = \sum_{i}^{n} y_i \ln \pi(x_i) + (n - \sum_{i}^{n} y_i)\ln(1-\pi(x_i) \qquad \text{donde} \qquad \pi(x_i) = \frac{e^{\beta_0}}{1+e^{\beta_0}}$$

por

Luego

$$L(\beta) = \sum_{i}^{n} y_i \beta_0 - \ln(1+e^{\beta_0}) + (n - \sum_{i}^{n} y_i)(-\ln(1+e^{\beta}\ )$$

$$= \sum_{i}^{n} y_i \ln\left[\frac{e^{\beta_0}}{1+e^{\beta_0}}\right] + (n - \sum_{i}^{n} y_i)\ln\left[\frac{1}{1+e^{\beta_0}}\right]$$

$$= \sum_{i}^{n} y_i \beta_0 - n\ln(1+e^{\beta_0})$$

Puesto que $\beta_0 = \ln({}^{n_1}\!/_{n_0})$ *y* $n_i = \sum_{i=1}^{n} y_i$, *sustituyendo éstos valores*

$$L(\beta) = \ln(n_1/n_0)^{\ n_1} - \ln(1+e^{\ln(n_1/n_0)})^n$$

$$L(\beta_0) = n_1 \ln n_1 / n_o - n \ln(1 + e^{\ln n_1/n_o})$$

Por propiedad de la exponencial $e^{\ln n1/n0} = n1/no$ y sustituyendo $n = n_0 + n_1$

$$= \ln\left(\frac{(n_1/n_0)^{n_1}}{(n_0 + n_1/n_0)^n}\right)$$

$$= \ln\left[\left(\frac{n_1}{n_0}\right)^{n_1} * \left(\frac{n_0}{n}\right)^n\right]$$

$$= \ln\left(\frac{n_1}{n_0}\right)^{n_1}\left(\frac{n_0}{n}\right)^{n_1+n_0}$$

$$= \ln\left(\frac{n_1}{n_0} x \frac{n_0}{n}\right)^{n_1}\left(\frac{n_0}{n}\right)^{n_0}$$

$$= \ln\left(\frac{n_1}{n_0}\right)^{n_1}\left(\frac{n_0}{n}\right)^{n_0}$$

Con esto se demuestra el resultado del numerador para el estadístico G para la verosimilitud del término y donde el denominador es la verosimilitud del modelo con la variable

A continuación demostraremos que $\ln[\Sigma Y_i / \Sigma (1 - Y_i)]$ es un estimador de máxima verosimilitud para β_0

La función de máxima verosimilitud del modelo con el término independiente:

$$\ln(L\beta) = \Sigma[Y_i \ln \pi(X_i) + (1 - Y_i) \ln(1 - \pi(X_i)]$$

$$= \Sigma Y_i \ln \pi(X_i) + \Sigma(1 - Y_i) \ln(1 - \pi(X_i))$$

Derivando parcialmente por βo y reemplazando $\pi(X_i) = \dfrac{e^{\beta_0}}{1 + e^{\beta_0}}$

$$\frac{\partial L(\beta_0)}{\partial L\beta_0} = \sum Y_i\left(\frac{1+e^{\beta_0}}{e^{\beta_0}}\right)\left[\frac{e^{\beta_0}(1+e^{\beta_0}) - e^{\beta_0}e^{\beta_0}}{(1+e^{\beta_0})^2}\right] + \sum(1-y_i)\left[(1+e^{\beta_0})\left(\frac{(-e^{\beta_0})}{(1+e^{\beta_0})^2}\right)\right]$$

Igualando a cero,

$$\sum y_i \frac{1}{1+e^{\beta_0}} - \sum (1-y_i)\frac{e^{\beta_0}}{1+e^{\beta_0}} = 0$$

$$\sum y_i - \sum (1-y_i)e^{\beta_0} = 0$$

$$e^{\beta_0} = \frac{\sum y_i}{\sum (1-y_i)}$$

$$\beta_0 = \ln\left(\frac{\sum y_i}{\sum (1-y_i)}\right) \tag{31}$$

Puesto que $n_1 = \Sigma\, y_i$ y $n_0 = \Sigma(1-y_i)$, luego entonces

$$\beta_0 = \ln\left(\frac{n_1}{n_0}\right)$$

__ **Prueba de Wald**

Otro estadístico importante es el de Wald (W), que se usa para probar de manera individual si la hipótesis nula de que la pendiente de la ecuación de regresión para la población es cero Si una ecuación de regresión tiene pendiente 0, un cambio en X no afecta a Y La hipótesis nula y la hipótesis alternativa de dos colas para probar la pendiente son

$$H_0 \quad \beta_1 = 0$$

$$H_1 \cdot \beta_1 \neq 0$$

El estadístico de Wald para el contraste se define como·

$$W = \frac{\beta_1}{\widehat{EE}\left(\widehat{B}_1\right)} \tag{32}$$

Donde W sigue una distribución normal con media cero y varianza 1 Sin embargo, esta ecuación pude ser generalizada para la regresión múltiple de la forma siguiente

$$W = \frac{\beta_p}{\hat{E}E\left(\hat{\beta}_p\right)} \tag{33}$$

en la que p = número de variables independientes del modelo de regresión.

$\hat{E}E(\hat{\beta}_i) =$ errores están dados de los parámetros del modelo

El estadístico de Wald en algunas ocasiones es elevado al cuadrado (como ocurre en la salida de programas como SPPS), distribuyéndose de esta forma como una chi-cuadrado con grados de libertad igual al número de parámetros estimado Su fórmula es .

$$W = \frac{\beta_i^{\,2}}{\hat{E}E(\hat{\beta}_i)^2} \tag{34}$$

Regla de Decisión

Rechazar H_0 si $W > X^2_{\alpha\,(1)}$ en cualquier otro caso, no rechazar H_0 Para variables categóricas el estadístico de Wald tiene grados de libertad igual al número de categorías menos uno Sin embargo, el estadístico de Wald no es confiable cuando el coeficiente de regresión es muy grande En este caso se acostumbra usar la estadística G.

Bondad de Ajuste del Modelo

La técnica del análisis de regresión logística se basa en la elaboración de modelos matemáticos que describen la relación entre un conjunto de variables independientes (Xi) y una variable debe ser teóricamente relevante Para ello se debe garantizar que el modelo construido sea idóneo y/o refleje de forma adecuada los datos utilizados La idoneidad del modelo de regresión logística utilizados se puede valorar siguiendo los siguientes criterios

_ Medidas Análogas a R^2 en Regresión Logística

Entre las medidas análogas a R^2 de regresión lineal se encuentran el Pseudo - R^2 y medidas basadas en el logaritmo de verosimilitud El Pseudo - R^2 o R cuadrado de Cox y Snell se calcula de la siguiente forma

$$\text{Pseudo } R^2 = 1 - \frac{\text{Deviancia modelo propuesto}}{\text{Deviancia modelo nulo}} \qquad (35)$$

Este indicador al igual que el análisis de deviancia no es totalmente adecuado para datos no agrupados Es normal que su valor sea bajo para datos individuales, pues un Pseudo- R^2 de 100% significa que el modelo saturado se reproduce perfectamente, lo que es difícil de lograr

La segunda medida R^2_L asume asunciones propias de la regresión lineal y se estima mediante la siguiente fórmula

$$R^2_L = [\, -2\ln L\, m_0 - (2\ln Lm)] \,/\, -2 \ln Lm_0 \qquad (36)$$

donde $\ln Lm_0$ es el valor del log verosimilitud del modelo que sólo incluye el término independiente y ln lm es el valor de log verosimilitud del modelo seleccionado El valor estimado de la prueba del Pseudo - R^2 y R^2_L produce valores entre 0 y 1, de manera que como más se acerquen al valor de 1 mejor será la bondad de ajuste

La diferentes medidas análogas al coeficiente de determinación múltiple R^2 presentan ventajas y desventajas, pero ninguna de ellas tiene una capacidad de discriminación y explicación de la variabilidad tan precisa como tiene esa medida en el caso de la regresión lineal, por lo que constituye meras aproximaciones

Existen otros indicadores que valoran la bondad de ajuste son a) Ji-cuadrado de Pearson b) prueba de Hosmer-Lemeshow y c) desviación, las tres son medidas resumen de bondad de ajuste que se basan en patrones de covariables, es decir el número de combinaciones posibles para las variables incluidas en el modelo para estudiar un fenómeno en particular

_ Ji-Cuadrado de Pearson

Como se mencionó anteriormente esta estadística es una medida resumen de las diferencias entre los valores observados y los estimados para y=1 calculado para cada patrón de covariable Por lo tanto, los valores ajustados dependen de la probabilidad estimada para cada patrón de covariable. Entonces y_j el valor ajustado para un patrón covariable j tiene la siguiente notación matemática

$$m_j \Pi_j = m_j \left[\frac{e^{\eta_j}}{1+e^{\eta_j}} \right] \tag{37}$$

donde n_j es la logit estimado y m_j es el número de casos con $x=x_j$, $j=1,2,3,\ ,j$ siendo J, el número de patrones covariables En base a este resultado, la medida resumen de la diferencia entre el valor observado y el estimado, el residuo de Pearson para un patrón covariable en particular, se define como

$$r(y_j, \hat{\pi}_j) = \frac{(y_j - m_j \hat{\pi}_j)}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}} \tag{38}$$

El estadístico resumen basado en estos residuos es el estadístico Ji-cuadrado de Pearson cuya notación matemática es

$$\chi^2 = \sum_{j=1}^{J} (y_i, \hat{\pi}_j)^2 \qquad (39)$$

Bajo la hipótesis nula de que modelo ajusta, la distribución de este estadístico es un ji-cuadrado con J-(p+1) grados de libertad. Una ji-cuadrado no significativa indica que el modelo propuesto se acerca al modelo saturado.

El problema de este estadístico es que cuando J=n, es decir el número de patrones covariables se aproxima al número de observaciones, el valor de p calculado para este estadístico, usando la distribución ji-cuadrado con J-(p+1) grados de libertad es cuestionable. Como alternativa a este problema, Hosmer y Lemeshow proponen otro estadístico basado también en patrones de covariables para deciles, denominados deciles de riesgo. A continuación se explica en que consiste este estadístico.

_ Prueba de Hosmer-Lemeshow

Esta prueba al igual que la ji-cuadrado de Pearson es una medida resumen de las diferencias entre los valores y los estimados, para y=1, calculado para cada patrón de covariable, por lo tanto los valores ajustados también dependen de la probabilidad estimada para cada patrón de covariable. La diferencia entre estos dos estadísticos es que la prueba de Hosmer-Lemeshow propone un agrupamiento basado en las probabilidades estimadas. Estas probabilidades están ordenadas ascendentemente, de probabilidades estimadas más pequeñas a las más altas.

La forma propuesta de agrupar estas probabilidades es utilizar deciles para agruparlas probabilidades estimadas. Este procedimiento usa g=10 grupos resultando que el primer grupo tiene n_1=n/10 sujetos que tienen las probabilidades estimadas más pequeñas y el último grupo las probabilidades estimadas más grandes.

Una vez que las probabilidades estimadas son agrupadas, se calcula el estadístico de Hosmer-Lemeshow el cual tiene la siguiente notación matemática

$$\hat{C} = \sum_{k=1}^{g} \frac{(O_k - n'_k \bar{\pi})^2}{\left(n'_k \bar{\pi}_k (1 - \bar{\pi}_k)\right)} \tag{40}$$

donde n_k es el número de patrones covariables en el k-ésimo grupo que contienen información,

$$O_k = \sum_{j=1}^{n'_k} y_j \tag{41}$$

el número de individuos observados con la característica de interés dentro del n'_k patrón covariable, y

$$\bar{\pi}_k = \sum_{j=1}^{n^1 k}{}'_k \frac{m_j \bar{\pi}_j}{n'_k} \tag{42}$$

es la probabilidad promedio estimada en el k-ésimo grupo Cuando J=n y el modelo de regresión logística ajustado es el modelo correcto, la distribución del estadístico Ĉ es aproximadamente una ji-cuadrado con g-2 grados de libertad Una ji-cuadrado no significativa indica que el modelo se aproxima al modelo saturado

_ Desviación D

En el modelo lineal clásico, una medida del grado de desajuste entre lo valores observado y lo predicho por el modelo viene dada por la suma de cuadrados residuales

En regresión logística se compara la verosimilitud del modelo ajustado con la del modelo "perfecto", el llamado modelo saturado, éste es el modelo que reproduce exactamente las observaciones realizadas, pero siempre es un modelo demasiado complicado pues tiene tantos parámetros como sujetos de estudio, sin embargo, lo podemos utilizar como referencia con el que comparar nuestro modelo ajustado Si acaso el punto de vista estadístico, al modelo saturado, podríamos pensar que el modelo ajustado explica nuestros datos suficientemente bien, con la gran ventaja de tener muchos menos parámetros y, por tanto, mucho mas sencillo, si por el contrario, nuestro modelo no da una explicación suficiente a los datos de nuestro estudio

La Desviación es una medida que como poco vale o cuanto mayor sea su valor peor explicara nuestro modelo los datos observados Para el caso de la regresión logística la Desviación viene dada por

$$D = 2\left\{\sum_{i=1}^{k} y_j \ln\left(\frac{y_j}{m_j \hat{p}_j}\right) + (m_j - y_j)\ln\left(\frac{m_j - y_j}{m_j - m_j(1-\hat{p}_j)}\right)\right\} \tag{43}$$

Donde k es el número de grupos de sujetos de estudio en relación a la predictora, m_j es el número de individuos en cada grupo, $m_j\, p_j$ son los valores que predice el modelo ajustado Bajo ciertas ocasiones la desviación no sigue una distribución chi-cuadrado y esto es más la regla que la excepción Para el caso de datos agrupados, es decir, cuando la o las predictoras sean todas categóricas y siempre que en cada casilla haya número suficiente de observaciones, la desviación sigue una distribución χ^2 con k-v grados de libertad, siendo k el número de patrones distintos de la o las predictoras y v el número de parámetros que se estima en el modelo

Así, siempre que la desviación sea menor que el valor de la chi-cuadrado correspondiente no hay evidencia del mal ajuste y diremos que el modelo ajustado reproduce nuestros resultados suficientemente bien Sin embargo, en el caso de datos no agrupados $n_i = 1$, es decir, en las casillas haya pocos individuos o cuando la predictora sea una variable continua, casi cada individuo tendrá un patrón de predictora distinto, pues bien, para este caso no se conoce la distribución de la observación por lo que no podemos saber si un valor concreto de ella implica necesariamente evidencia de un mal ajuste

_ Medida de Asociación Ordinal

Entre las medidas de asociación de tipo ordinal que se pueden utilizar para valorar la idoneidad de un modelo de regresión logística destacan gamma ® y la D de Somer

_ Medida de Eficacia Predictiva

Para medir la eficacia predictiva del modelo y el cálculo de la probabilidad de corte más adecuado se utiliza la curva ROC (Reciver Operating Characteristics)

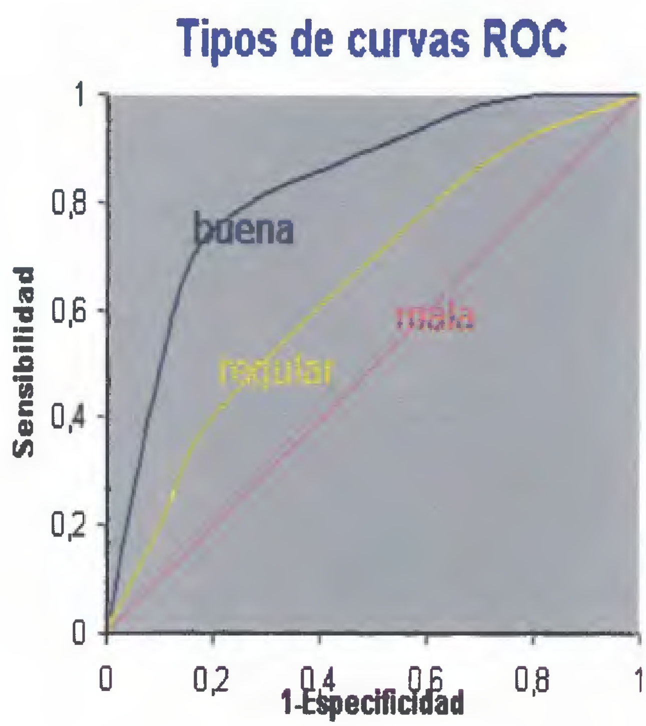
Información contenida en la Curva ROC

– Si la prueba fuera perfecta, es decir, sin solapamiento, hay una región en la que cualquier punto de corte tiene sensibilidad y especifidad iguales a 1, la curva sólo tiene el punto (0,1)

- Si la prueba fuera inútil ambas *fdp*´s coinciden y la sensibilidad (verdaderos positivos) es igual a la proporción de falsos positivos, la curva sería la diagonal de (0,0) a (1,1)

- Las pruebas habituales tienen curvas intermedias

Un parámetro para evaluar la bondad de la prueba es el área bajo la curva que tomará valores entre 1 (prueba perfecta) y 0,5 (prueba inútil). Puede demostrarse, que este área puede interpretarse como la probabilidad de que ante un par de individuos, uno enfermo y el otro sano, la prueba los clasifique correctamente.



__ **Diagnóstico del Modelo**

Evalúan la bondad de ajuste caso por caso mediante el análisis de los residuos del modelo y de influencia en la estimación del vector de parámetros del mismo.

Residuo del modelo:

La clave en la cuantificación en el diagnóstico de regresión logística, como en la regresión lineal. Son los componentes de la " suma de cuadrados de los residuos". E n regresión lineal la suposición clave es que al error de la varianza no depende de la media condicional E |y_j / x_j | Sin embargo, en regresión logística nosotros tenemos errores binomiales, como resultado, el error de varianza es una función de la media condicional

$$V\left(y_j / x_j\right) = m_j E\left(y_j / x_j\right) x \left[1 - E\left(y_j / x_j\right)\right] = m_j \pi(x_j)[1 - \pi(x)] \tag{44}$$

Así el residuo de Pearson toma la forma

$$r_j = r(y_j, \hat{\pi}_j) = \frac{y_j - m_j \hat{\pi}_j}{\sqrt{m_j \hat{\pi}_i (1 - \hat{\pi}_j)}} \tag{45}$$

Y el residuo de la desviación

$$d_j = d(y_j, \hat{\pi}_j) = \sqrt{2\left[y_j \ln \frac{y_j}{m_j \hat{\pi}_j} + (m_j - y_j) \ln \frac{m_j - y_j}{m_j(1 - \hat{\pi}_j)}\right]} \tag{46}$$

Cada residuo va dividido por una estimación aproximada del error estándar, por lo que se espera que si el modelo es adecuado, estas cantidades tengan aproximadamente, media igual a cero y varianza igual a 1

_ Medidas de Apalancamiento (Leverage)

Junto a los residuales para cada patrón covariable otra cantidad esencial de la interpretación del diagnóstico de regresión lineal es la matriz hat y la influencia de valores del mismo En regresión lineal la matriz hat, es la matriz que provee los valores ajustados como una proyección de la variable respuesta en el espacio de las covariables La matriz x es llamada a menudo matriz diseño En regresión lineal la matriz hat es H= x $(x'x)^{-1}$ x' por ejemplo, y = Hy En regresión lineal los residuales (y- ŷ) expresada en término de la matriz hat es (I-H) donde I es la j x j matriz identidad Usando un modelo de regresión lineal, Pregibon (1981) encontró una aproximación lineal de los valores ajustados que produce una hat matriz para regresión logística La matriz es

$$H = v^{1/2} x(x' vx)^{-1} x' v^{1/2} \tag{47}$$

donde V es una matriz j x j diagonal con elementos

$$v_j = m_j \pi(x_j)[1 - \pi(x_j)]. \tag{48}$$

En regresión lineal, los elementos diagonales de la matriz "hat" se denominan "Leverage" (Medidas de apalancamiento) y son proporcionales a la distancia entre x_j y la media de los valores de x La extensión de este concepto a regresión logística requiere algunas consideraciones

En primer lugar, sea hj el j-ésimo elemento diagonal de la martiz H que toma la forma

$$h_i = m_j \hat{\pi}(x_j)(1 - \hat{\pi}(x_j))(1, x'_j)(x' vx')^{-1}(1, x'_j)' \tag{49}$$

o lo que es lo mismo

$$h_j = v_j b_j$$

$$b_j = (1, x_j') (X' v X')^{-1} (1, x'_j)' \tag{50}$$

Como ocurre en regresión lineal, resulta $\sum h_j = p + 1$ (número de parámetros del modelo) Si se observa la matriz " hat " como una matriz n x n , cada elemento diagonal tendrá como límite superior $1/m_j$, siendo m_j , el total de unidades con el mismo patrón Cuando la matriz "hat" se basa en los datos agrupados por patrón de covariables, el límite superior para cualquier elemento diagonal será 1, así por ejemplo, se considera un patrón de covariable con m_j unidades $y_j = 0$ y probabilidad logística Π_j, el residuo de Pearson para cada unidad individual con ese patrón es

$$r_j = \frac{0 - \hat{\pi}_j}{\sqrt{\hat{\pi}_j\left(1 - \hat{\pi}_j\right)}} = -\sqrt{\frac{\hat{\pi}_j}{1 - \hat{\pi}_j}} \tag{51}$$

Mientras que el residuo de Pearson basado en todas las unidades del patrón es

$$r_j = \frac{0 - m_j \hat{\pi}_j}{\sqrt{m_j \hat{\pi}_j\left(1 - \hat{\pi}_j\right)}} = -\sqrt{m_j}\sqrt{\frac{\hat{\pi}_j}{1 - \hat{\pi}_j}} \tag{52}$$

Expresión que toma valores negativos cada vez mayores a medida que m_j crece, y que no serían detectados si las medidas de diagnóstico se realizan en base a los datos desagregados

Pregibon llama "bloques de construcción básico " a las estadísticas de diagnóstico r_j, d_j, y h_j ya que son de interés por ellas mismas y también porque se emplean para el cálculo de otras estadísticas que se verán a continuación

_ <u>Residuo de Pearson estandarizado</u>

Si se considera el residuo para j-ésimo patrón de covariables como

$$y_j - m_j \Pi'(x_j) = (1 - h_j)\, y_j$$

La varianza del residuo es

$$m_j \Pi'(x_j)\,[(1 - \Pi'(x_j)](1 - h_j) \tag{53}$$

lo que sugiere que los residuos de Pearson no tendrán varianza igual a 1 al menos que sean estandarizados El residuo de Pearson estandarizado para el patrón de covariables x_j es

$$r_{sj} = \frac{r_j}{\sqrt{1 - h_j}} \tag{54}$$

donde r_{sj} es el residuo de Pearson ya definido

Otras estadísticas de diagnóstico son las que examinan el efecto de eliminar todas las unidades con un patrón de covariables particulares sobre el valor de los coeficientes estimados y sobre las medidas resumen de ajuste, x^2 y D Donde x^2 esta defina por

$$x^2 = \sum_{j=1}^{J} r\left(y_j, \hat{\pi}_j\right)^2 \tag{55}$$

Esta medida resumen de los residuos de Pearson es una estadística chi-cuadrado y donde la estadística D (Desviación) se define como en la ecuación (43)

_ <u>Medidas derivadas del efecto de cada patrón de covariables sobre el ajuste del modelo</u>

Se puede mostrar, usando aproximadamente lineales, que la disminución en el valor de la estadística X^2 de Pearson debida a la eliminación de las unidades con el patrón x_j es

$$\Delta X_j^2 = r_{sj}^2 \tag{56}$$

Una medida similar puede obtenerse para el cambio en la "Desviación" debido a la eliminación de las unidades con el patrón x_j

$$\Delta D_j = \frac{d_j^2}{1 - h_j} \tag{57}$$

_ <u>Medidas derivadas del efecto de cada patrón de covariables sobre la estimación de los parámetros</u>

Se trata de una medida análoga a la propuesta por Cook para regresión lineal El cambio en el valor de los coeficientes estimados se obtiene como la diferencia estandarizada entre β y $\beta_{(-j)}$, las que representan las estimaciones máximo verosímiles calculadas usando todos los patrones de covariables y excluyendo las m_j unidades con patrón el x_j respectivamente, luego estandarizando vía la matriz de covarianza estimada de $\hat{\beta}$

Pregibón (1981) mostró, mediante una aproximación lineal que esta medida para regresión logística es .

$$\Delta \hat{\beta} = \frac{r^2 h_j}{1 - h_j} \tag{58}$$

Estas estadísticas de diagnóstico, derivadas de los " bloques de construcción básicos", son conceptualmente muy interesantes, ya que permite identificar aquellos patrones de covariables que están pobremente ajustados (grandes valores de ΔX^2 y/ ó ΔD_j y aquellos que tienen una gran influencia sobre los valores de los parámetros estimados (grandes valores de $\Delta\hat{\beta}_j$)

Para interpretar el valor de los parámetro en regresión lineal, se emplea el recurso gráfico y la distribución de los diagnósticos bajo el supuesto de que el modelo es correcto En regresión logística, el recurso fundamental es el gráfico, ya que la distribución de los diagnósticos bajo la hipótesis de que el modelo ajusta, sólo se conoce bajo ciertas condiciones Así, por ejemplo, se dice frecuentemente que la distribución del residuo de Pearson es aproximadamente normal cuando el modelo es correcto Esto es sólo cierto cuando m_j es suficiente grande para justificar que distribución binomial se aproxima a la normal Por ejemplo si $m_j = 1$, r_j tiene sólo dos valores posibles, y obviamente no se cumpliera la normalidad (Jenning, 1986)

Las medidas de diagnóstico se evalúan por patrón de covariables (con datos agrupados); entonces cualquier aproximación de su distribución basada en la normal, dependerá bajo errores binomiales, del número de unidades de los patrones, m_j Cuando el modelo ajustado contiene covariables continuas, de modo que el número de patrones de covariables, J es del mismo orden que n, no se cumplirá la condición que m_j sea suficientemente grande. Es así que, el empleo de la distribución normal estandarizada, o la x^2_1, dara sólo una idea de cuán grandes son los valores de diagnóstico

Se han sugerido distintos tipos de gráficos, dirigidos cada uno a un aspecto particular del ajuste En la práctica es aconsejable restringir la atención en unos pocos de fácil cálculo y significativos para el análisis de regresión logística Ellos pueden ser

- gráfico de ΔX_j^2 versus $\hat{\Pi}_j$
- gráfico de ΔD_j versus $\hat{\Pi}_j$
- gráfico de $\Delta\hat{B}_j$ versus $\hat{\Pi}_j$

Identificados los casos influyente, se debe deducir qué hacer con ellos Un caso extremo influyente no se descarta automáticamente, debido a que puede ser enteramente correcto y simplemente representar un evento poco probable. Si, en cambio, el conocimiento del problema provee de una explicación a ese caso inusual, es decir un caso que indique una situación excepcional que no tiene por qué cubrir el modelo, puede entonces que sea apropiado descartarlo

3.6. EL PROBLEMA DE LA COLINEALIDAD EN LOS MODELOS DE REGRESIÓN LOGÍSTICA

Es uno de los problemas que se puede encontrar en un análisis de regresión Esto ocurre cuando en la estimación de los coeficientes, de un modelo de regresión multivariado donde alguna variable independiente es combinación lineal de otras, el modelo es irresoluble, debido a que, en ese caso, la matriz $X'X$ es singular, es decir, su determinante es cero y no se puede invertir

A este fenómeno se le denomina colinealidad Que una variable X_1 sea combinación lineal de otra X_2, significa que ambas están relacionadas por la expresión $X_1 = \hat{\beta}_1 + \hat{\beta}_2 X_2$ siendo $\hat{\beta}_1$ y $\hat{\beta}_2$ constantes, por lo tanto el coeficiente de correlación entre ambas variables será 1

Otro modo, por tanto, de definir la colinealidad es decir que existe colinealidad cuando alguno de los coeficientes de correlación simple o múltiple entre algunas de las variables independientes es 1, es decir, cuando algunas variables independientes están correlacionadas entre sí.

En la práctica, esta colinealidad exacta raras veces ocurre, pero sí surge con cierta frecuencia la llamada casi-colinealidad, o por extensión, simplemente colinealidad en que alguna variable es "casi" combinación lineal de otra u otras, o dicho de otro modo, algunos coeficientes de correlación simple o múltiple entre las variables independientes están cercanos a 1, aunque no llegan a dicho valor

3.7. CONFUSIÓN E INTERACCIÓN EN LA REGRESIÓN LOGÍSTICA

El concepto de variable de confusión se refiere a una variable independiente (X_1) que está asociada o explica la variable dependiente (Y) y también esta relacionada con la variable independiente (X_2) objeto de estudio, no siendo un factor intermedio o explicativo de la relación entre esas dos variables (X_2 , Y), en otras palabras, la variable de confusión no explica la posible relación entre la variable independiente y la dependiente, aunque su ausencia en el modelo de regresión logística produzca una estimación sesgada de la asociación entre las dos variables objetos de estudio

Contrastar la existencia de confusión requiere, por lo tanto, comparar los coeficientes de regresión obtenidos en dos modelos diferentes y si hay diferencia, existe la confusión, en cuyo caso la mejor estimación es la ajustada Lo habitual es considerar que existe confusión cuando la exponencial del coeficiente (la RÔ) cambia en más del 10%

Las variables modificadas de efecto o interacción indican que la asociación entre una variable independiente (X_1) y la dependiente (Y) no es constante a lo largo de los diferentes valores que puede tener la variable independiente (X_2)

El modelo más sencillo que hace explícita la interacción entre dos variables X_1 y X_2 es

$$\ln (p/q) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_3 \quad (59)$$

En este modelo, el logaritmo del ods para unos valores determinados X_1 , X_2 de X_1 , X_2 es

$$\ln (p/q) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 \quad (60)$$

y para los valores $x_1 + 1$ y x_2

$$\ln (p/q) = \beta_0 + \beta_1 (x_1 + 1) + \beta_2 x_2 + \beta_3 (x_1 + 1)x_2 \quad (61)$$

$$= \beta_0 + \beta_1 x_1 + \beta_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_3 x_2 \quad (62)$$

restando ambas (60) y (61) se encuentra el cambio en ln (p/q) por una unidad de cambio en X_1 manteniendo fijo X_2

$$\beta_1 + \beta_3 x_2 \quad (63)$$

o dicho de otra manera, la razón ods por una unidad de cambio en X_1 manteniendo fijo X_2 es

$$e^{\beta_1 + \beta_3 X_1} = e^{\beta_1} e^{\beta_3 X_2} = e^{\beta_1} (e^{\beta_3})^{x_2} \quad (64)$$

que es diferente para cada valor x_2 de X_2 Del mismo modo, el cambio en In (p/q) por una unidad de cambio en X_2 manteniendo fijo X_1 es:

$$\beta_2 + \beta_3 x_1$$

o en términos de RO, la razón ods por unidad de cambio X_2 manteniendo x_1 es

$$e^{\beta_2+\beta_3 x_1} = e^{2\beta} e^{\beta_3 x_1} = e^{\beta_2}\left(e^{\beta_3}\right)^{x_1} \tag{65}$$

Por lo tanto, contrastar la existencia de interacción entre X_1 y X_2 es contrastar si el coeficiente β_3 es cero (no hay interacción), o distinto de cero (existe interacción) Nótese que para poder interpretar así este contraste es necesario que en el modelo figuren las variables X_1 , X_2 y $X_1 X_2$

En caso de que exista interacción los coeficientes los exponenciales de β_1 y β_2 por sí solo no significan nada y la asociación de las variables X_1 y X_2 con la binomial estará cuantificada por las expresiones anteriores

3.8. LAS VARIABLES CATEGÓRICAS EN EL MODELO LOGÍSTICO

Puesto que la metodología empleada para la estimación del modelo logístico se basa en la utilización de variables cuantitativas, al igual que en cualquier otro procedimiento de regresión, es incorrecto que en el intervengan variables cualitativas, ya sean nominales o ordinales

La asignación de un número a cada categoría no resuelve el problema ya que si tenemos, por ejemplo, la variable realiza ejercicio físico con tres posibles respuestas nunca, esporádicamente, frecuentemente, y le asignamos los valores 0, 1, 2, significa a efectos del modelo, que ejercicio físico frecuentemente es dos veces mayor que solo hacerlo esporádicamente, lo cual no tienen ningún sentido. Más absurdo sería si se trata, a diferencia de ésta, de una variable nominal, sin ninguna relación de orden entre las respuestas, como puede ser el estado civil.

La solución a este problema es crear tantas variables dicotómicas como número de respuestas menos una categoría. Estas nuevas variables, artificialmente creadas, reciben en la literatura anglosajona el nombre de *'dummy'*, traduciéndose en español con diferentes denominaciones como pueden ser variables internas, indicadoras, variables diseño o ficticias.

__ **Codificación Tipo Indicador con Referencia a la Primera Categoría de X_0**

Si una variable explicativa es categórica, con c valores posibles, se crean c-1 variables dicotómicas como variables explicativas. Estas variables se pueden construir de diversas forma y seguir la forma utilizada cambia la interpretación de sus coeficientes $\hat{\beta}_i$ que en general, cuantifican el efecto de el valor de dichas variables con respecto a un valor de referencia. Ejemplo: Suponer que $K = 1$ y X_1 es categórica con $c = 3$.

En este caso la variables categórica puede tomar valores como el el ejemplo anterior del ejercicio físico.

La forma de representar los $c = 3$ posibles valores de la variable X_1 es:

X_1	Ind o	Ind 1	Ind 2
1	1	0	0
2	0	1	0
3	0	0	1

Una vez codificada las variables indicadoras, se sustituye en el modelo de regresión logística la variable X_1 por c-1 = 2 de esas variables. Nunca deben incluirse las c variables indicadoras en el modelo, dado que al estar todas las categorías de X_1 en el modelo por los valores "1" de cada una de las variables indicadoras, no habría ninguna categoría con la que contrastar. Además, que categoría de X_1 estaría representada por los tres ceros lo que permite que exista una relación lineal perfecta entre las variables en cuestión, produciéndose una multicolinealidad perfecta.

Por tanto las dos formas posibles de codificar X_1 es:

X_1	Ind 1	Ind 2
0	0	0
1	1	0
2	0	1

En este caso la ecuación del modelo ajustado viene dada por:

$$\ln\left(\frac{p(Ind_1, Ind_2; \beta)}{1 - p(Ind_1, Ind_2; \beta)}\right) = \hat{\beta}_0 + \hat{\beta}_1 Ind_1 + \hat{\beta}_2 Ind_2$$

El ods de la variable de referencia (la primera respuesta) se obtiene reemplazando el valor cero en las variables indicadoras de la ecuación de regresión logística:

Luego,

$$\ln\left(\frac{P(Ind_1, Ind_2; \beta)}{1 - P(Ind_1, Ind_2; \beta)}\right) = \hat{\beta}_0 + \hat{\beta}_1(0) + \hat{\beta}_2(0) = \hat{\beta}_0$$

$$\frac{P_0}{1 - P_0} = e^{\hat{\beta}_0} \qquad (66)$$

El ods para las variables indicadoras Ind 1 y Ind 2 se obtienen de igual manera reemplazando los valores establecidos en la codificación Los resultados son los siguientes ods para la variable

Variable Ind_1 Variable Ind_2

$$\frac{P_1}{1-P_1} = e^{\hat{\beta}_0+\hat{\beta}_1} \tag{67}$$

$$\frac{P_2}{1-P_2} = e^{\hat{\beta}_0+\hat{\beta}_2} \tag{68}$$

La estimación de la razón ods esta por el cociente del ods de variables y el ods de la variable de referencia

$$R\hat{O}_{Ind\ 1} = \frac{\dfrac{p_1}{1-p_1}}{\dfrac{p_0}{1-p_0}} = \frac{e^{\beta_1+\beta_0}}{e^{\beta_0}} = e^{\beta_1} \tag{69}$$

$$R\hat{O}_{Ind\ 2} = \frac{\dfrac{p_2}{1-p_2}}{\dfrac{p_0}{1-p_0}} = \frac{e^{\beta_0+\beta_2}}{e^{\beta_0}} = e^{\beta_2} \tag{70}$$

Por tanto, e^{β_1} $1 = 1, 2$ compara la razón ods correspondiente $X_1 = 1,2$ frente a la categoría de referencia $X_1 = 0$

— **<u>Codificación Desviación con Referencia a la Primera Categoría de X_0</u>**

En el caso que una categoría no pueda ser considerada de forma natural como nivel de referencia, por ejemplo grupo sanguíneo, para este tipo de variable un posible sistema de clasificación es el de desviación.

Ejemplo

Suponer que $K = 1$ y X_1 es categoría con $c = 3$

El método consiste en crear variables Desv 1 y Desv 2 dada por

X_1	$Desv_1$	$Desv_2$
0	-1	-1
1	1	0
2	0	1

En este caso la ecuación del modelo ajustado viene dada por:

$$\log\left(\frac{p(Ind_1, Ind_2;\beta)}{1-p(Ind_1, Ind_2;\beta)}\right) = \beta_0 + \beta_1 Desv_1 + \beta_2 Desv_2$$

y se tiene que la ods para las variables indicadoras de X_1 se obtiene reemplazando los valores de las codificación realizada, por tanto el ods para categoría X_0 de referencia es:

$$\frac{P_0}{1-P_0} = e^{\hat{\beta}_0-\hat{\beta}_1-\hat{\beta}_2} \tag{71}$$

El ods para las variables indicadoras $Desv_1$ y $Desv_2$ esta dada por:

Variable Desv 1

$$\frac{P_1}{1-P_1} = e^{\hat{\beta}_0+\hat{\beta}_1} \tag{72}$$

Variable Desv 2

$$\frac{P_2}{1-P_2} = e^{\hat{\beta}_0+\hat{\beta}_2} \tag{73}$$

luego multiplicando las tres ods obtenemos:

$$\frac{P_0}{1-P_0}.\frac{P_1}{1-P_1}.\frac{P_2}{1-P_2} = e^{\hat{\beta}_0-\hat{\beta}_1-\hat{\beta}_2}.e^{\hat{\beta}_0+\hat{\beta}_1}.e^{\hat{\beta}_0+\hat{\beta}_2} = e^{3\hat{\beta}_0} \tag{74}$$

Por lo tanto, $e^{\beta i}$ $i = 1,2$ compara las razones del riesgo correspondiente $X_2 = 1, 2$ frente a las de la categoría media cuya razón de riesgo viene dada por:

$$\frac{P_m}{1-P_m} = \sqrt[3]{\frac{P_0}{1-P_0} x \frac{P_1}{1-P_1} x \frac{P_2}{1-P_2}} \tag{75}$$

3.9. POBLACIÓN OBJETIVO

La población objeto de estudio la integran los niños y niñas en edades entre 0 y 4 años de edad en la república En el período de tiempo comprendido entre el 8 de junio al 8 de octubre de 1997.

3.10. ENCUESTA DE NIVELES DE VIDA

La Encuesta de Niveles de Vida recoge la información necesaria para obtener una medida lo más completa posible de las principales dimensiones de las condiciones y niveles de vida de los hogares y las personas, abordando los siguientes aspectos

Nivel de Hogares, Viviendas y Personas

- Características Físicas de la vivienda, tamaño, edad, tenencia y titulación, pagos de arriendo, acceso a servicios de agua, luz, teléfono, niveles de consumo, costos y calidad, gastos en mantenimiento y reformas. Recreación de las personas del hogar
- Estructura, composición y tamaño de los hogares y los núcleos familiares asociados Niveles educativos y ocupaciones de los padres no residentes en el hogar
- Cuidado del niño, lactancia materna, subsidios de alimentación, inmunización, presencia de diarrea y enfermedades respiratorias, demanda y acceso a los servicios de salud, gastos por enfermedad, accidentes y tratamientos, seguros, prácticas y hábitos de salud y desnutrición
- Demanda por servicios de educación preescolar, subsidios, gastos anuales y mensuales Alfabetismo y lengua materna, demanda por servicios de educación formal (matrícula, repitencia, etc), subsidios, gastos anuales y mensuales, no-asistencia escolar, etc.
- Lugar de nacimiento y residencia anterior, razones y tiempo de migración.
- Actividades de trabajo (empleados, desempleados, inactivos), condición de inactividad

Ocupación, Rama de Actividad, tiempo de trabajo, tamaño de la empresa, categoría ocupacional, ingresos del trabajo independiente, ingresos de los asalariados, etc

- Control del embarazo, gastos del embarazo, nacidos vivos, atención al parto, peso de los niños al nacer y gastos del parto.

A Nivel Comunidad de la comunidad se recoge valiosa información entre ellas Características de la Comunidad, servicios básicos, servicios comunitarios etc

3.11. DISEÑO Y SELECCIÓN DE LA MUESTRA

Para los propósitos y objetivos de la ENV97, la Dirección de Estadística y Censo (DEC) de la Contraloría General de la República diseñó una muestra probabilística, multietápica e independiente en cada dominio de estudio

La estratificación geográfica utilizada para efecto de selección, permite que cada subconjunto esté representado en la muestra y poder alcanzar adecuadamente los objetivos de la investigación

- Región Metropolitana

a) Ciudad de Panamá y San Miguelito (Urbano)

Ciudad de Panamá, Distrito de San Miguelito, Áreas en Desarrollo de la Ciudad de Panamá y Distrito de San Miguelito, Edificios nuevos de 12 y más viviendas de la Ciudad de Panamá y el Distrito de San Miguelito y Asentamientos Espontáneos de la Ciudad de Panamá y el Distrito de San Miguelito

b) Resto de la Región Metropolitana

- **Area Urbana**

Ciudad de Colón, Resto del Distrito de Panamá y el Distrito de Colón (incluye Ancón y Cristóbal), Áreas en Desarrollo del Resto del Distrito de Panamá, Distrito de Arraiján, Distrito de la Chorrera y del Distrito de Colón, Edificios nuevos de 12 y más viviendas del

Resto del Distrito de Panamá y del Distrito de Colón, Asentamientos espontáneos del Resto del Distrito de Panamá, Distrito de Arraiján y del Distrito de Colón y Área Urbana del Resto de los Distritos de la Provincia de Panamá y Colón.

- **Area Rural No Indígena**

  Área Rural del Resto del Distrito de Panamá y Colón, Área Rural del Resto de los Distritos de la Provincia de Panamá y Colón.

- Resto del País (excluye las provincias de Panamá y Colón)

Para cada una de las Provincias del Resto del País se definieron los siguientes estratos **Area Urbana, b) Area Rural y c) Área Rural Indígena y Área Rural de Difícil Acceso.**

Se creó un estrato por cada área respectivamente, es decir, se agruparon las UPM's bajo las categoría respectiva para todas las provincias y se creó un solo estrato nacional

Los estratos están dados en función de los niveles de ponderación de los resultados Es decir, que son 20 estratos que corresponden a

- Colón Urbano, Colón Rural, Ciudad de Panamá y San Miguelito (Urbano), Resto de la Provincia de Panamá (Urbano), Resto de la Provincia de Panamá (Rural), Coclé Urbano, Coclé Rural, Herrera Urbano, Herrera Rural, Los Santos Urbano, Los Santos Rural, Veraguas Urbano, Veraguas Rural, Bocas del Toro Urbano, Bocas del Toro Rural, Chiriquí Urbano, Chiriquí Rural, Darién Rural, Indígena y Difícil Acceso

__ **Tamaño de la Muestra:**

La ENV trabaja con una muestra sin reemplazo. El tamaño muestral estimado ascendió a 5,591 viviendas, se calculó una tasa de respuesta total del 90% lo que permite obtener al final una muestra efectiva de 5,000 viviendas Este tamaño permite lograr estimaciones aceptables para cada uno de los dominios de estudio considerados.

CAPITULO IV

ANÁLISIS Y PRESENTACIÓN DE LA INFORMACIÓN

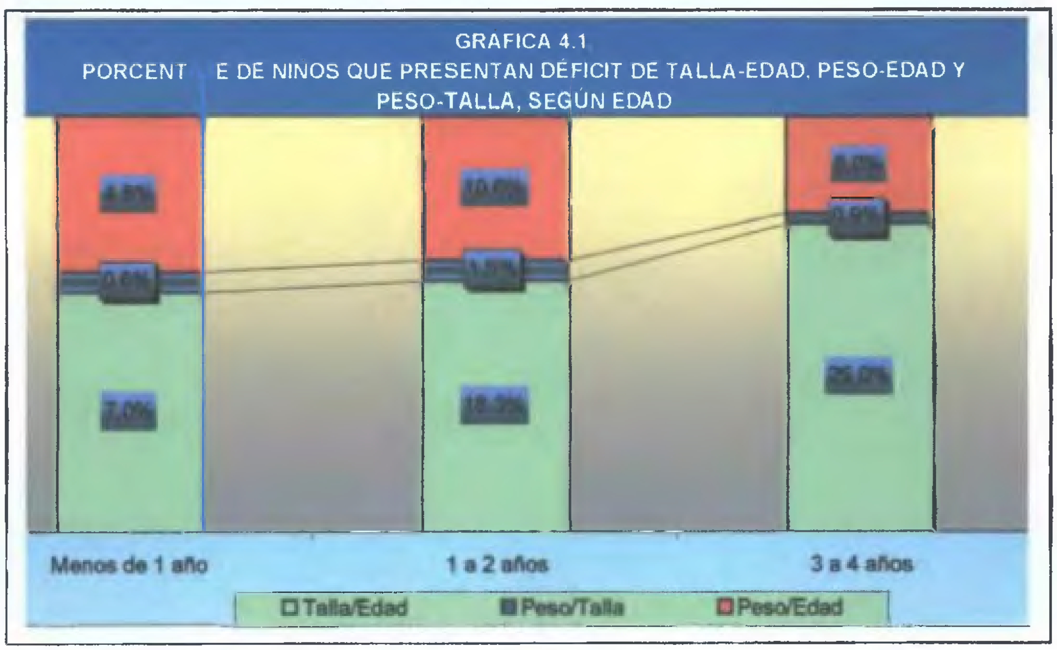
4.1. ANALISIS DESCRIPTIVO DE LA INFORMACIÓN

En esta sección haremos una breve descripción de la información para posteriormente aplicar el modelo de regresión logística y extraer nuestra conclusiones acerca del fenómeno de la desnutrición infantil.

CUADRO 4.1.
PREVALENCIA DE DÉFICIT NUTRICIONAL EN NIÑOS Y NIÑAS POR EDAD, SEGÚN MEDIDAS ANTROPOMÉTRICAS

MEDIDAS ANTROPOMÉTRICAS		EDAD EN AÑOS					TOTAL
		Menos de 1	1	2	3	4	
FALTA DE PESO (PESO/EDAD)	Número	23	60	37	42	32	194
	Porcentaje	1.0	2.6	1.6	1.8	1.3	8.5
AGUDA (PESO/TALLA)	Número	3	12	2	6	2	25
	Porcentaje	0.1	0.5	0.08	0.3	0.08	1.1
CRÓNICA (TALLA/EDAD)	Número	35	87	82	110	109	423
	Porcentaje	1.5		3.6	4.8	4.7	18.4
TOTAL	Prevalencia	44	105	88	116	111	464
	Porcentaje	1.9	4.6	3.8	5.1	4.8	20.2
	TOTAL	499	478	441	418	458	2294
	Porcentaje	21.8	20.8	19.2	18.2	20.0	100.0

FUENTE: ENCUESTA DE NIVELES DE VIDA. MINISTERIO DE ECONÓMICA Y FINANZAS. 1997



En este cuadro se puede observar la prevalencia de la desnutrición infantil, según los indicadores antropométricos Las medidas antropométricas clasifican a un individuo dentro de una escala de puntaje Z para indicar su estado de nutricional Esta situación plantea que del total de niños y niñas seleccionados en el estudio aproximadamente el 20 2% presentaban algún problema de desnutrición Donde el retardo en el crecimiento, es decir, la desnutrición crónica es la más relevante ya que del total de niños el 18 4% que equivale a 423 casos estaban en esta condición, un 8 5% de estos tenían bajo peso para su edad y un 1 1% de esta misma muestra padecen de desnutrición actual (aguda)

Un factor demográfico de gran influencia en los niveles de desnutrición, es la edad En la Gráfica 4 1 podemos apreciar con mayor claridad la incidencia de la desnutrición infantil según este factor

La desnutrición crónica afecta en mayor grado a los niños y niñas mayores de un año, encontrándose su mayor prevalencia, 9 5%, en los niños y niñas en edades de 3 y 4 años con un porcentaje del 4 8% y 4 7% para ambas edades respectivamente La menor incidencia se observa en los niños menores de un año con el 1 5%, lo que puede deberse a que en esas edades los niños aún son lactantes y su ingesta diaria no requiere de grandes cantidades de alimentos sólidos

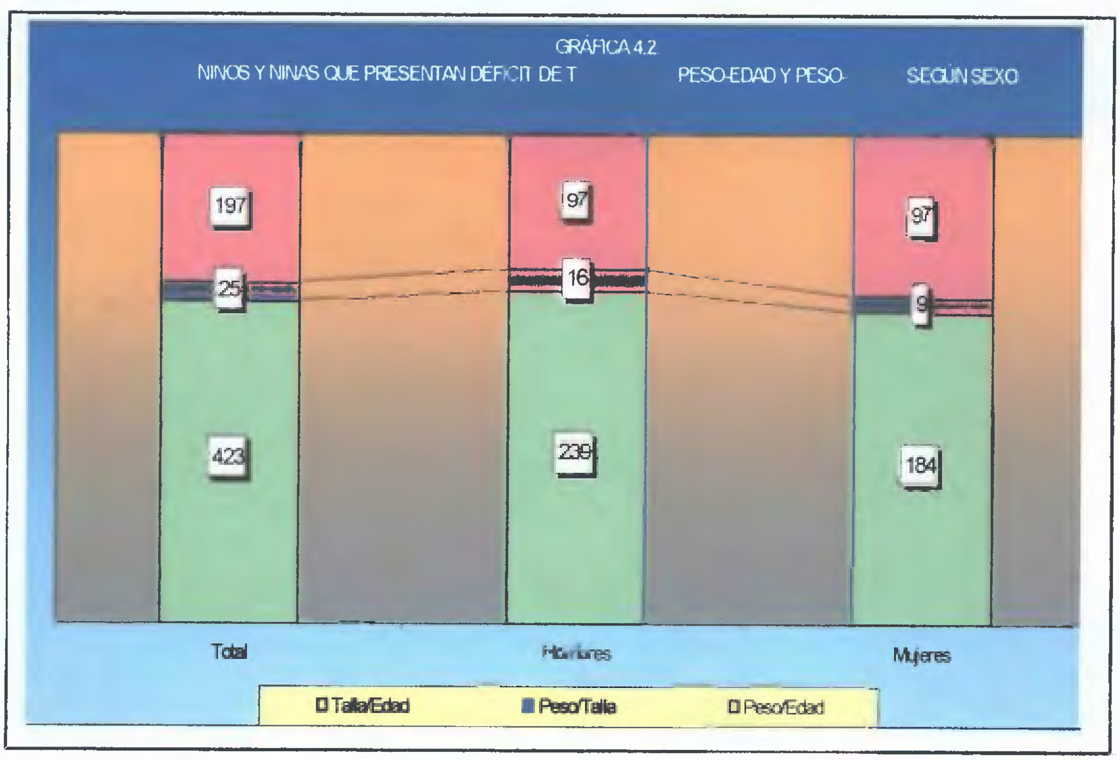
Al utilizar el indicador peso edad para evaluar el estado nutricional de los niños y niñas este presenta una situación diferente al indicador anterior, siendo su prevalencia de 8 5% Este indicador refleja un mayor porcentaje (2 6%) en los niños con un año de edad, siguiéndoles con un 1 8% los menores con 3 años de edad, en los niños con 2 y 4 años de edad, se observa un porcentaje casi similar con aproximadamente un 1.6% y 1 3%

1.3% respectivamente; la menor frecuencia se encuentra en los menores de un año.

CUADRO 4.2.
PREVALENCIA DE DÉFICIT NUTRICIONAL EN NIÑOS Y NIÑAS POR SEXO, SEGÚN MEDIDAS ANTROPOMETRICAS

MEDIDAS ANTROPOMÉTRICAS		SEXO				TOTAL	
		HOMBRES		MUJERES			
		NUMERO	PORCENTAJE	NÚMERO	PORCENTAJE	NÚMERO	PORCENTAJE
FALTA DE PESO (PESO/EDAD)		97	4.2	97	4.2	194	8.5
AGUDA (PESO/TALLA)		16	0.7	9	0.4	25	1.1
CRÓNICA (TALLA/EDAD)		239	10.4	184	8.0	423	18.4
TOTAL	Prevalencia	263	11.5	201	8.8	464	20.2
	TOTAL	1,171	51.0	1,123	48.9	2,294	100.0

FUENTE: ENCUESTA DE NIVELES DE VIDA. MINISTERIO DE ECONOMÍA Y FINANZAS 1997



Con respecto a las frecuencias por sexo, se observó que de los casos analizados 1,171 son niños y 1,123 son niñas. De este total 464 niños y niñas padecen algún grado de desnutrición que constituyen el 20.2%; el 11.3% o sea la mayor prevalencia 352 son niños y 8.8% son niñas.

Los varones representan un 51.%, del total de casos, de éstos el 4.2% registraron problemas de bajo peso y el 10.4% o sea 239 presentaban retardo en su crecimiento. Respecto a las niñas,

estás representan el 48.9% del total de casos, de las cuales el 8% (184) presentan desnutrición crónica, mientras que el 4.2% tienen bajo peso para su edad. Esto indica que existe una mayor prevalencia de niños que de niñas con algún grado de deterioro en su estado nutricional incidiendo con más fuerza el retardo en el crecimiento (desnutrición crónica).

CUADRO 4.3.
PREVALENCIA DE DÉFICIT NUTRICIONAL EN NIÑOS Y NIÑAS POR NIVEL DE POBREZA, SEGÚN MEDIDAS ANTROPOMÉTRICAS

MEDIDAS ANTROPOMÉTRICAS		NIVEL DE POBREZA				TOTAL	
		NO POBRE		POBRE			
		NÚMERO	PORCENTAJE	NÚMERO	PORCENTAJE	NÚMERO	PORCENTAJE
FALTA DE PESO (PESO/EDAD)		18	0.8	176	7.7	194	8.5
AGUDA (PESO/TALLA)		6	0.3	19	0.8	25	1.1
CRÓNICA (TALLA/EDAD)		42	1.8	381	16.6	423	18.4
TOTAL	Prevalencia	50	2.2	414	18.0	464	20.2
	Total	899	39.2	1395	60.8	2294	100.0

FUENTE: ENCUESTA DE NIVELES DE VIDA . MINISTERIO DE ECONÓMICA Y FINANZAS. 1997

La pobreza es uno de los flagelos que los pueblos de Latinoamérica sufren hoy en día, estos se deben a múltiples factores como el desempleo, la falta de recursos para alcanzar niveles educativos que permitan acceder a los mercados laborales actuales y la falta del estado en concretar políticas que ayuden a repartir más equitativamente las riquezas del país, entre otros. Esto hace que los problemas de desnutrición se agraven.

En Panamá esta situación es relativamente baja (18.4%), sin embargo, las cifras absolutas no reflejan lo mismo; ya que cerca de 50,000 niños están desnutrido y alrededor del 20.2% sufre de algún grado de desnutrición. Las disparidades debido a la condición de pobreza se expresan en prevalencias de 2.2% y 18.0% de niños (as) que sufren algún grado de desnutrición, procedentes de estratos no pobres y pobres, respectivamente, estas diferencia evidencia la desigualdad en las condiciones de vida de esto grupos de población.

GRAFICA 4.3.
NIÑOS Y NIÑAS QUE PRESENTAN DÉFICIT DE TALLA-EDAD. PESO-EDAD Y PESO-TALLA. SEGÚN NIVEL DE POBREZA

En esta Gráfica se observa que en los no pobres, los tres indicadores: talla para la edad (1.8%), peso para la edad (0.78%) y peso y talla (0.26%) están por debajo de la prevalencia total. En contra disposición con los pobres en donde los valores encontrados se aproximan al del total general; con relación a la desnutrición crónica esta alcanza valores iguales al 16.6%, mientras que el bajo peso para la edad es del orden del 7.7%. Estos resultados hacen fuerte referencia a que la pobreza esta íntimamente ligada al estado nutricional de los niños (as).

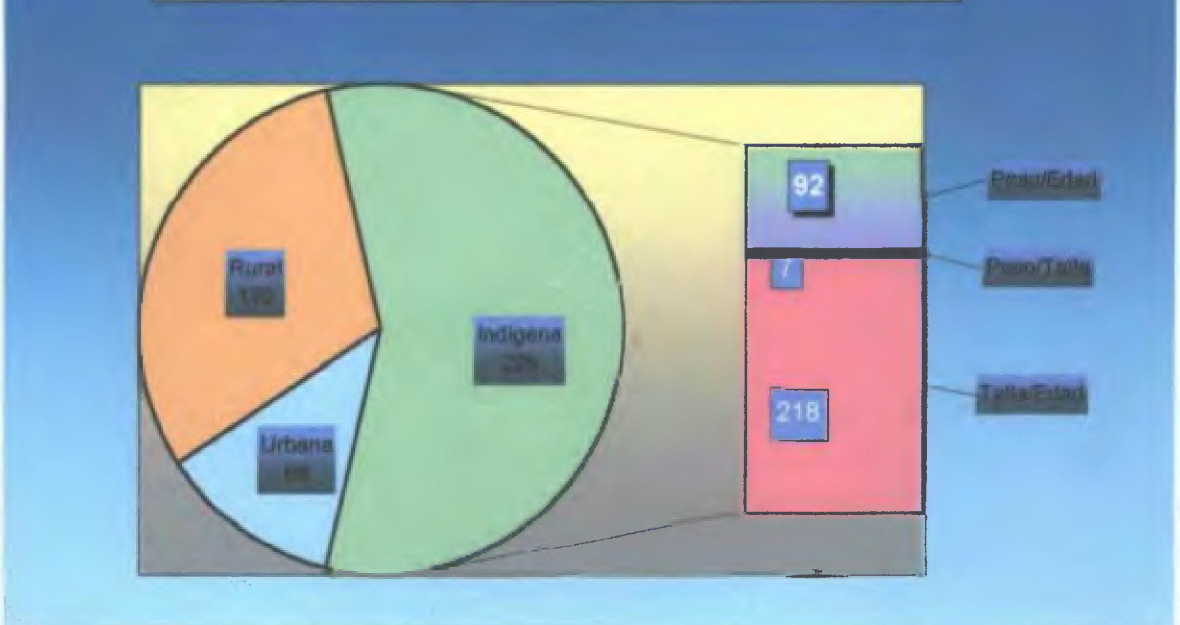
CUADRO 4.4.
PREVALENCIA DE DÉFICIT NUTRICIONAL EN NIÑOS Y NIÑAS POR AREA GEOGRAFICA, SEGÚN MEDIDAS ANTROPOMÉTRICAS

MEDIDAS ANTROPOMÉTRICAS		ÁREA GEOGRAFICA						TOTAL	
		URBANA		RURAL		INDÍGENA			
		NÚMERO	PORCEN-TAJE	NÚMERO	PORCEN-TAJE	NÚMERO	PORCEN-TAJE	NÚMERO	PORCENTAJE
FALTA DE PESO (PESO/EDAD)		27	1.2	75	3.3	92	4.0	194	8.5
AGUDA (PESO/TALLA)		8	0.3	10	0.4	7	0.3	25	1.1
CRÓNICA (TALLA/EDAD)		54	2.4	151	6.6	218	9.5	423	18.4
TOTAL	Prevalencia	68	3.0	170	7.4	226	9.9	464	20.2
	Total	876	38.2	980	42.8	438	19.1	2294	100.0

FUENTE: ENCUESTA DE NIVELES DE VIDA REALIZADA. MINISTERIO DE ECONOMÍA Y FINAZAS. 1997

Las disparidades existentes en el área geográfica dentro del territorio nacional son evidentes. El deterioro del estado nutricional de los niños y niñas del área indígena con un 9.5% y el rural con el 6.6% representan juntos el 16.1% de la totalidad de los casos desnutrición crónica que a nivel general registran el 18.4%. Esto es indicativo que la desnutrición está asociada a la situación geográfica y evidencia una menor disponibilidad de alimentos en las zonas rurales e indígenas respecto al área urbana; pues los recursos que existen para satisfacer las necesidades primarias de los niños (as) son limitados. Pues la relación de desnutrido urbano/ rural es de 1 a 7, esto quiere decir, que por cada niño (a) desnutrido en el área urbana hay 7 desnutridos en el área rural e indígena según este indicador.

GRAFICA 4.4
NIÑOS Y NIÑAS CON DÉFICIT NUTRICIONAL, SEGÚN AREA GEOGRAFICA



Los datos de la gráfica muestran que de los 464 niños (as) con algún grado de desnutrición que representan el 20.2%, casi la mitad 226 son del área indígena (9.9%). De éste total, 92 (4.0%) tienen bajo peso para la edad y 218 padecen desnutrición crónica.

4.2. MODELO DE REGRESIÓN LOGÍSTICA SIMPLE: VARIABLE INDEPENDIENTE CONTINUA

Con base a la información prevista por la ENV 97 sobre las características de la población menor de cinco años, procederemos a realizar un análisis de regresión aplicando el modelo logístico; con el propósito de determinar, si la edad del niño esta relacionada con el hecho de estar desnutrido. La Tabla 4.1 resume las características de la muestra respecto al estado nutricional, según la edad del niño o la niña.

TABLA 4.1.
DESCRIPCION DE LA RELACIÓN ESTADO NUTRICIONAL EDAD

Estado Nutricional	N	Valor Medio	Desviación Tipica	Rango	
				Mínimo	Máximo
Desnutrido (1)	423	2.4	1.29	0	4
No Desnutrido (0)	1871	1.83	1.44	0	4
Total	2294	1.94	1.43	0	4

La variable dependiente Estado Nutricional (ESTN) se define como: Y=1"Desnutrido", Y = 0 " No desnutrido" y el modelo de regresión logística construido para estudiar la relación entre la variable independiente continua " X_1 = Edad" y la variable dependiente "Y = Estado Nutricional" (ESTN).

4.2.1. Modelo de Regresión Logística

TABLA 4.2.
MODELO DE REGRESIÓN LOGÍSTICA SIMPLE: VARIABLE CONTINUA:
MODELO: LN (ESTN) = - 2.088 + 0.283 (EDAD)

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	. 95.0% para EXP(	
								Inferior	Superior
Paso 1[a]	EDAD	.283	.039	53.321	1	.000	1.328	1.230	1.432
	Constant	-2.088	.105	398.774	1	.000	.124		

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

El modelo de regresión logística resultante es

$$\ln(p/1-p) = -2\ 088 + 0\ 283\ (\text{EDAD})$$

Los coeficientes de regresión $\hat{\beta}_0$ y $\hat{\beta}_1$ puede interpretarse de la manera siguiente

a) La constante de regresión $\hat{\beta}_0$ es igual a -2 088 es negativa corresponde al valor del logaritmo natural del cociente de probabilidades es " P(Y = 1) = 1" cuando la edad es cero Al ser negativa, indica que la probabilidad estimada de estar desnutrido disminuye en los menores de 1 año

b) El signo positivo del coeficiente "$\hat{\beta}_1$", que estima el valor del parámetro poblacional "β_1", indica que la probabilidad estimada de estar desnutrido "P(Y=1)" aumenta conforme también aumenta la edad del niño También se puede interpretar que por cada aumento de 1 año en la edad, se estima que el logaritmo natural del coeficiente de probabilidades "P(Y=1)=1" se incrementará en 0 283

Estimación de Probabilidades

El modelo regresión logística también permite el cálculo de probabilidades aplicando la fórmulas descrita en el Capítulo III, podemos estimar la probabilidad de que un niño seleccionado al azar este desnutrido dado que tenga una edad específica Para un año estará dada por

$$P(Y = 1 / X_1 = 1\ año) = 1 / \left(1 + e^{-(\beta_0 + \beta_1 X_1)}\right)$$

$$= 1 / (1 + e^{-[-2\ 088 + 0\ 283)(1)]}$$

$$= 1 / (1 + e^{1\ 805})$$

$$= 0.1412$$

Si tiene 4 años de edad la probabilidad estimada de estar desnutrido sería de

$$P(Y = 1 / X_1 = 4 \text{ años}) = 1 / \left(1 + e^{-(\beta_0 + \beta_1 X_1)}\right)$$
$$= 1 / (1 + e^{-[-2\ 088 + 0\ 283)(4)]}$$
$$= 1 / (1 + e^{0\ 956})$$
$$= 0\ 2777$$

En contra disposición, la probabilidad estimada de que un niño no padezca desnutrición es de 0 8588, si tiene 1 años de edad y de 0 7223 si tiene 4 años de edad A continuación describiremos su cálculo puesto que P (Y = 1) = 0 1412 para X_1 = 1 año, entonces la probabilidad de no padecer desnutrición al primer año de vida se obtiene a través de su complemento, es decir

$$P(Y = 0 / X_1 = 1 \text{ año}) = 1 - P(Y = 1)$$
$$= 1 - 0\ 1412$$
$$= 0\ 8588$$

para el caso de X_1 = 4 años el resultado sería el siguiente

$$P(Y = 0 / X_1 = 4 \text{ años}) = 1 - P(Y = 1)$$
$$= 1 - 0\ 2777$$
$$= 0\ 7223$$

Las probabilidades obtenidas también se podrían estimar con la siguiente fórmula directa

$$P(Y = 1) = e^{(\beta_0 + \beta_1 X_1)} / \left(1 + e^{(\beta_0 + \beta_1 X_1)}\right)$$

La aplicación de la fórmula anterior para estimar la probabilidad de que un niño seleccionado al azar de primer año de edad padezca desnutrición es

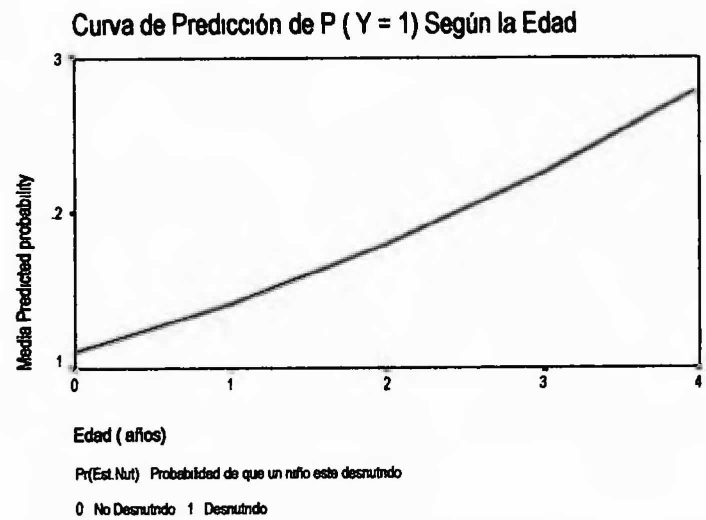
$$P(Y = 1 / X_1 = 1 \text{ año}) = e^{-[-2\ 088 + 0\ 283)(1)]} / (1 + e^{-[-2\ 088 + 0\ 283)(1)]})$$

$$= e^{-1\ 805} / (1 + e^{-1\ 805})$$

$$= 0\ 1412$$

La probabilidad estimada de que un niño (a) seleccionado al azar de la población este desnutrido en función de la edad se describe en el Gráfico 4 5.

GRAFICO 4.5.



La función expresada en la ecuación [16] no es otra cosa que una estimación de la curva de prevalencia de desnutrición en los menores de cinco años. Una forma de comprender como se construyen este tipo gráfico en regresión logística se muestra a continuación·

Primeramente se calcula la probabilidad P(Y=1) para los diversos valores de X, los que se encuentran en el rango de 0 a 4 años Estos valores serán las coordenadas que permitirán representar gráficamente los puntos (X, p) A continuación se presenta las probabilidades obtenidas a partir de la ecuación [16].

Por otra parte, supongamos que se quiere conocer a que edad la prevalencia de desnutrición infantil asciende a un 27% El valor de la probabilidad predicha es 0 27, "P(Y = 1) = 0 27" de que un niño este desnutrido se halla en los 3 86 años de edad, lo que se puede comprobar tanto en el gráfico anterior como con la aplicación de la siguiente fórmula

$$P(Y = 1) = 1 / \left(1 + e^{-(\hat{\beta}_0 + \hat{\beta}_1 X_1)} \right)$$

Reemplazando en P (Y = 1) el valor de la ecuación el valor igual a 0 27 y resolviendo algebraicamente la ecuación de primer grado tenemos

$$0\ 27 + 0\ 27 e^{-(\beta_0 + \beta_1 X_1)} = 1$$

$$0\ 27 \left((1 + e^{-(\hat{\beta}_0 + \hat{\beta}_1 X_1)} \right) = 1$$

$$0\ 27 e^{-(\hat{\beta}_0 + \hat{\beta}_1 X_1)} = 1 - 0\ 27$$

Aplicando logaritmo natural en ambos lado de la ecuación tendremos

$$In \quad e^{-(\beta_0 + \beta_1 X_1)} = In \quad 2.7037$$

$$-(\hat{\beta}_0 + \hat{\beta}_1 X_1) = 0\ 9946$$

$$X_1 = \frac{0\ 9946 + \hat{\beta}_0}{-\hat{\beta}_1}$$

Reemplazando los valores de $\hat{\beta}_0$ y $\hat{\beta}_1$ obtenemos

X_1 = - 1 0934 / -0 283=3 86 años

Este resultado nos indica que la probabilidad de que un niño (a) seleccionado al azar con 3 86 años de edad padezca desnutrición es del 27%

En el caso de la función logística simple, el valor X para el cual la función de probabilidad asciende a 0 5 es,

$$X_{MEDIANA} = {}^{-\hat{\beta}_0}\!\big/_{\hat{\beta}_1}$$

Este resultado se obtiene, siguiendo el esquema planteado anteriormente, si P(Y=1) = 0 5, entonces

$$0.5\left(1+e^{-(\hat{\beta}_0+\hat{\beta}_1 Xm)}\right)=1$$

$$e^{-(\hat{\beta}_0+\hat{\beta}_1 Xm)} = {}^{0.5}\!\big/_{0.5}$$

Aplicando logaritmo natural en ambos lado de la ecuación tendremos

$$-\hat{\beta}_0 - \hat{\beta}_1 X_m = 0$$

$$\therefore \quad X_{MEDIANA} = {}^{-\hat{\beta}_0}\!\big/_{\hat{\beta}_1}$$

4.2.2. Análisis del Modelo Estimado

Una vez estimado los parámetros del modelo procederemos a verificar la significancia y la bondad de ajuste del modelo.

__ **Evaluación de la Bondad del Ajuste**

- *Desviación (Desviance)*

Estadígrafo de Prueba·

$$D = 2\sum_{j=1}^{k}[y_j \; In\frac{y_j}{m_j\hat{p}_j} + (m_j - y_j)In\frac{m_j - y_j}{m_j - m_j(1-\hat{p}_j)}]$$

El estadístico de desviación sigue una distribución chi-cuadrado con j-(p+1) grados de libertad, donde j representa el número de observaciones en X y p el número de parámetros (g l = 5 - (2) = 3) k = 5: es el número de grupos de sujeto de estudio en relación a la predictora, o sea el número de patrones covariables: menores de 1 año, 1año, 2 años, 3 y 4 años.

y_j : son los números de sujetos con la característica de interés; es decir, los desnutridos por segmento de edad: 35, 87, 82, 110 y 109.

$\hat{p}_j$: es la probabilidad obtenida a través del Modelo ajustado: 0.113, 0.14124, 0.17817, 0.2246 y 0.278.

$m_j\hat{p}_j$: son los valores esperados correspondientes a los y_j y que predice el modelo ajustado

m_j : es el número de sujetos en cada grupo: 499, 478, 441, 418 y 458 respectivamente.

Las hipótesis nula y alternativa que se plantean para esta estadística son:

H_0 = El modelo es un modelo que se ajusta bien

H_1 = El modelo no es un modelo que se ajusta bien

TABLA 4.3.
PROBABILIDAD Y VALORES ESPERADOS PREDICHOS POR DESNUTRICIÓN
MODELO: LN(ESTN) = -2.088+0.283(EDAD)

Desnutridos (Y=1)	**Edad (en años)**										
	Menores de 1 año			1			2			3	4
Valores Observados	35			87			82			110	109
Valores Esperados	M_1	$\hat{P}_1$	$M_1\hat{P}_1$	M_2	$\hat{P}_2$	$M_2\hat{P}_2$	M_3	$\hat{P}_3$	$M_3\hat{P}_3$	$M_4\hat{P}_4$	$M_5\hat{P}_5$
	499	0.1103	55.04	478	0.14124	67.51	441	0.17917	79.01	93.88	127.32

El cálculo de la estadística de desviación a través de la Tabla 4.3. se realiza a continuación.

$$D = 2\{35\ln\left[\frac{35}{55.04}\right] + (499-35)\ln\left[\frac{499-35}{499-55.04}\right] + 87\ln\left(\frac{87}{67.51}\right) + (478-87)\ln\left[\frac{478-87}{478-67.50}\right] + \ldots + (458-109)\ln\left(\frac{458-109}{458-127.32}\right)$$

D = 2{-15 82 + 20 49 + 22 06 - 19 02+ 3 05 - 2 98 + 17 44 - 15 72 - 16 93 + 18 82}

=22 60

Regla de Decisión

Como D = 22 60 > $\chi_{0.95}^2$ (3) = 7 81, en consecuencia se rechaza la hipótesis nula, por lo que el modelo no es un modelo que se ajuste bien a los datos Luego entonces no es confiable hacer inferencias sobre la desnutrición con este modelo, sin embargo, por razones académicas continuaremos con dicho análisis luego de realizada la salvedad

_ ***Prueba Ji-cuadrado de Person***

Otra medida de bondad de ajuste de un modelo, es el estadístico χ^2, definido como

$$X = \sum \frac{(y_j - m_j \hat{p}_j)^2}{m_j \hat{p}_j} \sim \chi^2(j-(p+1))$$

En base a los datos de la Tabla 4 3 haremos el cálculo de esta estadística

$$\chi^2 = \frac{(35-55\,04)^2}{55\,04} + \frac{(87-67\,51)^2}{67\,1} + \frac{(82-79\,01)^2}{79\,01} + \frac{(110-93\,88)^2}{93\,88} + \frac{(109-107\,32)^2}{127\,32}$$

$= 7\,28 + 5\,63 + 0\,11 + 2\,77 + 2\,63$

$= 18\,42$

Regla de Decisión

Como 18 42 > $\chi^2_{0.95}$ (3) = 7 815, se rechaza la hipótesis nula, en consecuencia el modelo no se ajusta bien a los datos

_ ***Prueba de Hosmer-Lemeshow***

La prueba de Hosmer-Lemeshow para el modelo da como resultado·

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	22.028	3	.000

Como se observa en la tabla de la salidas del SPSS, el valor p=0.000, lo que evidencia que los datos no se ajustan adecuadamente al modelo. Como podemos observar se obtiene el mismo resultado con las tres pruebas estadística presentada.

__ **Evaluación de la Asunción de Linealidad de la Variable Continua**

La falta de bondad de ajuste puede provenir de una especificación incorrecta de la forma funcional de las predictoras. La utilización de métodos gráficos son una de las principales herramientas existentes para establecer una aproximación a la forma funcional correcta.

Un método de evaluar la adecuación de los datos al modelo es representar las logits estimadas para cada edad en el plano cartesiano, donde la logits se representan en el eje vertical y luego verificar si existe una tendencia lineal.

TABLA 4.4.
LOGITS POR DESNUTRICIÓN, SEGÚN EDAD
MODELO: LN(ESTN) =-2.088 + 0.283(EDAD)

Edad	Logits
0	-2.088
1	-1.805
2	-1.522
3	-1.239
4	-0.356

Una vez estimada las logits, se deben representar en un gráfico de dispersión y así podremos corroborar la existencia de una relación lineal. De no ser así, debemos someter nuestra variable predictora, en este caso EDAD, a algún tipo de transformación: logarítmica o la inclusión de términos elevados al cuadrado, al cubo, etc.

__ **Medidas de la Bondad del Ajuste**

Para la valoración de la bondad de ajuste del modelo de regresión logística a los datos utilizaremos medidas de bondad de ajuste análogas al coeficiente de determinación (R^2) de la regresión lineal

_ ***Pseudo - R^2 o R Cuadrado de Cox y Snell***

$$\text{Pseudo - } R^2 = 55\ 257 / (55\ 257 + 2{,}294)$$

$$= 0\ 024$$

En este estudio el Pseudo - R^2 = 0 024, lo que significa que el 2% de variación total de la desnutrición es explicada por la variable incluida en el modelo (EDAD)

- ***R - Cuadrado de Nagelkerke***

El R cuadrado de Nagelkerke según resultados de SPSS es igual a 0 041 el cual es someramente superior al Pseudo – R^2

Las dos pruebas anteriores producen valores entre 0 y 1 de manera que cuánto más se acerquen al valor 1 mayor será la bondad de ajuste

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	2137 768	024	039

__ **Comparación de Modelos**

Una pregunta que se intenta responder en este apartado es ¿ hay diferencias entre los niños (as) según su edad y la desnutrición infantil?, en otras palabras, ¿ está asociado el estado nutricional de los niños (as) con la edad? Estos cuestionamientos los podemos confirmar a través de los intervalos de confianza para la razón ods y la significancia de la variable

- ***Significancia de la Variable***

Con los resultados de la Tabla 4 2 obtenidas del modelo de regresión logística, se establecen las siguientes hipótesis nula y alternativa, con el propósito de determinar si la pendiente de la ecuación de regresión logística para la población es 0

$$H_0 \quad \beta_1 = 0$$

$$H_1 \quad \beta_1 \neq 0$$

La hipótesis nula (β_1) afirma que no existe relación entre la edad y el estado nutricional del niño (a) en la población

Para evaluar la significancia de la variable independiente EDAD en el modelo se compara el valor de la verosimilitud con la variable Y sin la variable en el modelo Esta estadística se expresa como

$$G = 2\left\{\sum_{i=1}^{n}[Y_i In(\hat{p}_i) + (1-y_i)In(1-\hat{p}_i)] - [n_1 In(n_1) + n_0 In(n_0) - nIn(n)]\right\}$$

Bajo la hipótesis nula de que β_1 es igual a cero, la estadística G se distribuye como uno chi-cuadrado con 1 grado de libertad

Los valores $n_1 = 423$, corresponden al número de sujetos con la característica de interés, $n_0 = 1,871$ es el total de individuos que no presentan la característica de interés y n es igual a 2,294 representa el número total de individuos en la muestra

Iniciaremos con el cálculo del segundo término de la ecuación anterior, el cual expresa como

$$n_1 \ln(n_1) + n_0 \ln(n_0) - n\ln(n) = 423 \ln(423) + 1{,}871 \ln(1{,}871) - 2{,}294 \ln(2{,}294)$$

$$= -1{,}096\ 5133 \qquad (76)$$

El valor obtenido -1,096 5133 corresponde al log verosimilitud del modelo con sólo el término constante El primer término de la ecuación que se distinguen por la sumatoria requiere del cálculo individual, es decir, caso por caso, o por el contrario a través de la agrupación de los datos. En el primero de los casos el cálculo individual requiere de siguiente información:

TABLA 4.5.
VALORES OBSERVADOS Y PROBABILIDAD PREDICHA, SEGÚN EDAD: MODELO LN (ESTN) = -2.088 + 0.283 (EDAD)

Ind	Edad	Y	$\hat{p}_i$	$Y \ln \hat{p}_i$	1-Y	$1-\hat{p}_i$	$(1-Y) \ln (1-\hat{p}_i)$
1	0	0	0.11	0	1	0.89	-0.11653
2	1	1	0.149	-1.959	0	0.859	0
3	2	0	0.179	0	1	0.821	-0.19723
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
2,294	3	0	0.225	0	1	0.775	-0.25489
		Σy_i=423		$\Sigma y_i \ln p_i$ = -69.3741	1,871		$\Sigma(1-y_j) \ln (1-p_j)$= -376.49134

Se nos haría imposible presentar la totalidad de los datos contenido en la muestra por lo que hicimos una agrupación de los mismos de manera de ofrecer una alternativa de comprensión de los resultados sin embargo podemos constatar que la complejidad de la sumatoria no es tal, ya que se resume en:

$$\ln(\hat{p}_j)$$

cuando el individuo presenta la característica de interés $y_j = 1$ y cuando no es así $y_j = 0$, entonces:

$$\ln(1-\hat{p}_j)$$

TABLA 4.6.
DISTRIBUCIÓN DEL ESTADO NUTRICIONAL, POR EDAD Y PROBABILIDAD PREDICHA: MODELO LN (ESTN) = -2.088 + 0.283(EDAD)

EDAD	DESNUTRIDOS Y =1	$\hat{p}_i$	$Y \ln \hat{p}_i$	NO DESNUTRIDO (1-Y)	$(1-\hat{p}_i)$	$(1-Y) \ln (1-\hat{p}_i)$
0	35	0.11	-77.254622	464	0.89	-54.07169
1	87	0.141	-170.4326	391	0.859	-59.42666
2	82	0.179	-141.0703	359	0.821	-70.806349
3	110	0.225	-164.08204	308	0.775	-78.506813
4	109	0.278	-139.53462	349	0.722	-113.67982
	Σy_i=423		$\Sigma y_i \ln p_i$ =692.3741	1,871		$\Sigma(1-y_j) \ln (1-\hat{p}_j)$= -376.49134

De la Tabla anterior se obtiene.

$$\sum_{i=1}^{n}[(Y_i \ln(\hat{p}_i) + (1-y_i)\ln(1-\hat{p}_i)] = \sum_{j=1}^{m} Y_i \ln \hat{p}_i - \sum_{j=1}^{n-m}(1-y_i)\ln(1-\hat{p}_i)$$

$$= -692\ 37418 - 376\ 49134$$

$$= -1{,}068\ 8655 \qquad (77)$$

el valor obtenido es el correspondiente a la log - verosimilitud del modelo con la variable EDAD

Utilizando los resultados [76] y [77] se obtiene la estadística G que esta dada por

$$G = 2\ \{-1{,}068\ 8655) - 2\ (\ -1{,}096\ 5123)\}$$

$$= -2{,}137\ 731\ +\ 2{,}193\ 0246$$

$$=\ 55\ 29$$

es fácil deducir que -2 log de la verosimilitud del modelo con la variable es igual a 2,137 731 y salvo algunas cifras decimales es similar al valor obtenido en la Tabla Resumen del Modelo (pág 86) de la salida del SPSS Por último el resultado de la estadística G = 55 29, corresponde a la denominada Prueba de Omnibus que presentamos a continuación de la salida del SPSS

Pruebas omnibus sobre los coeficientes del modelo

		Chi-cuadrado	gl	Sig
Paso 1	Paso	55 257	1	000
	Bloque	55 257	1	000
	Modelo	55 257	1	000

Como $G = 55\ 29 > \chi^2_{0\,95}\ (1) = 3\ 841$, podemos concluir que existen evidencia de que la variable EDAD es significativa prediciendo el estado nutricional del niño (a) También se deduce que el ajuste del modelo mejora significativamente con la adición de esta variable

El estadígrafo para el contraste de Wald también nos permite el análisis de significancia de las variables desde un punto de vista individual y no global como la Prueba de Omnibus

$$W = \frac{\hat{\beta}_1}{\hat{E}E(\hat{\beta}_1)} \sim N(0,1)$$
$$= \frac{0\ 283}{0\ 039} = 7\ 256$$

Regla de Decisión

Como el valor de la prueba W = 7 256 es mayor que el límite superior (Z = 1 96), se rechaza la hipótesis nula (Ho $\beta_1 = 0$)

__ **Estimación del Ods**

- Menores de 1 año de edad

El estimador del ods para la edad, es decir, el cociente de la probabilidad de estar desnutrido y la probabilidad de no estar desnutrido, si se tiene menos de 1 años de edad es

$$\text{Ods} = e^{\hat{\beta}_o}$$
$$= e^{-2\ 088}$$
$$= 0\ 1239$$

debido a que el ods es una cifra decimal podemos multiplicarlo por 10, lo que nos indicaría que por cada 10 niños menores de un año de edad no desnutrido existe aproximadamente 1 2 niños de la misma edad desnutrido Tenemos que señalar que los menores de un año tienen una edad de o años para efecto de nuestro estudio

- ***1 año de edad***

El ods para los niños con un 1 año de edad esta dado por

$$\text{Ods} = e^{\hat{\beta}_o + \hat{\beta}_1}$$
$$= e^{-2.088 + 0\ 283}$$
$$= 0\ 1645$$

¿ Cuál será el ods para 2 años de edad? Se calcula

$$e^{\hat{\beta}_0 + 2\hat{\beta}_1} = e^{-2\,088 + 2 \times 0\,283} = 0\,2183$$

que también se puede calcular como

$$0\,1239 \times (1\,328)^2 = 0\,2185$$

Para realizar el contraste de si es significativamente distinto de uno hay que estimar su varianza

Matriz de correlaciones

		Constante	EDAD
Paso 1	Constante	1 000	- 853
	EDAD	- 853	1 000

$$\begin{aligned} \mathrm{var}(\hat{\beta}_0 + 2\hat{\beta}_1) &= \hat{\mathrm{var}}(\hat{\beta}_0) + 4\,\hat{\mathrm{var}}(\hat{\beta}_1) + 4\,\mathrm{cov}(\hat{\beta}_0, \hat{\beta}_1) \\ &= \hat{E}E(\hat{\beta}_0)^2 + 4\hat{E}E(\hat{\beta}_1)^2 + 4cor(\hat{\beta}_0, \hat{\beta}_1) x\hat{E}E(\hat{\beta}_0) x\hat{E}E(\hat{\beta}_1) \\ &= (0\,105)^2 + 4\,(0\,039)^2 + 4\,(0\,853) \times 0\,105 \times 0\,039 \\ &= 3\,13686 \times 10^{-3} \end{aligned}$$

y, por tanto el estadígrafo de Wald para el contraste

$$W = \frac{(\hat{\beta}_i - \beta_i)^2}{\mathrm{var}(\hat{\beta}_i)} \approx \chi^2(1)$$

donde β_i es una constante

$$W = \frac{(\hat{\beta}_0 + 2\hat{\beta}_1)^2}{\mathrm{var}(\hat{\beta}_1)}$$

$$= \frac{(0\,2183)^2}{3\,13686 \times 10^{-3}} = 15\,19 \Rightarrow p = 0\,000$$

Como el valor p =0 000 < 0 05, se rechaza H_0 de que el coeficiente es igual a cero

Estimación de la Razón Ods

A continuación mediremos la magnitud de la asociación entre la edad y el estado nutricional asumiendo que se cumple la asunción o supuesto de linealidad entre la entre la logit de desnutrición y la edad

La razón ods estimada de estar desnutrido es creciente por cada incremento de un año de edad

$$R\hat{O} = e^{\hat{\beta}_1} = e^{(0\ 283)}$$

$$= 1\ 327$$

El estimador de la razón ods para el aumento en un año es 1 327, es decir el ods se multiplica por esa cantidad por cada aumento de un año y es significativamente distinto de uno (p = 0 000)

En el caso de variables continuas es recomendable estimar la razón ods a lo largo de un intervalo determinado de 10 años En nuestro caso utilizaremos 2 años por ser la edad máxima 4 años

$$R\ \hat{O}_{(2años)} = e^{[2 \times (+0\ 283)]} = 1\ 761$$

o su aproximación

$$(1.327)^2 = 1.761$$

Por cada aumento de un año de edad la probabilidad de estar desnutrido aumenta en un 32.7%

$$100\ (1\ 327\ - 1)\ =\ 32\ 7\%.$$

y por cada incremento de 2 años aumenta en.

$$100\ \ (1.761 - 1)\ \ = 76\ 1\%$$

Bajo el supuesto que la relación entre la longitud del estado nutricional y la edad es lineal se puede apreciar que el efecto del ods es multiplicativo y no aditivo, por lo que a aumento de dos años de edad, las ods anuales no se suman sino que se multiplican

$$R\hat{O}_{(2\ \text{años})} = R\hat{O}_{(1\ \text{año})} \times R\hat{O}_{(1\ \text{año})}$$

$$= 1\ 327 \times 1.327 = 1\ 761$$

__ **<u>Intervalos de Confianza para Razón Ods</u>**

Los resultados de la Tabla 4 2 también nos presentan la estimación de los intervalos de confianza para razón ods para un nivel de confianza de $\alpha = 0.05$

$$I\,C_{(R\hat{O})} = e^{[\ \hat{\beta}_1\ \pm (1\ 96 \times EE\,(\hat{\beta}_1\ ))]}$$

$$= e^{[\ 0\ 283\ \pm (1\ 96 \times 0\ 039)]}$$

$$= [1\ 230,\ 1\ 433]$$

$$1\ 230 \le R\hat{O} \le 1\ 433$$

El 95% IC para RÔ en un período de dos años se puede estimar de la siguiente forma·

$$I\,C_{(R\hat{O})} = e^{[\ (2x(0\ 283)\ \pm\ (1\ 96 \times 2 \times 0\ 039)\ ]}$$

$$= e^{(0\ 566\ \pm 0\ 15288)}$$

$$= [1.512,\ 2\ 052]$$

o su aproximación

$$(1\ 230)^2 = 1.512$$

$$(1\ 433)^2 = 2\ 053$$

4.3. MODELO DE REGRESIÓN LOGÍSTICA SIMPLE: VARIABLE INDEPENDIENTE DICOTÓMICA

Con el interés de predecir la proporción de desnutrición en función del nivel de pobreza se plantea el siguiente análisis de regresión logística. No sin antes destacar que esta variable no se incluirá en el modelo multivariado para evitar problemas de multicolinealidad que se produce cuando dos o más variables independientes se correlacionan entre sí de forma importante, lo que dificulta la estimación del efecto separado de cada unas de las variables en la predicción de los valores de P (Y = 1). Un ejemplo de correlación perfecta o multicolinealidad lo podría constituir un modelo de regresión logística en el que se incluyeran como variables independientes: nivel de educación del padre, la madre y el nivel de pobreza, como se puede observar ambas variables se saben están correlacionadas, ya que las personas pobres tienden a alcanzar un grado de escolaridad inferior a los no pobres.

TABLA 4.7.
DESCRIPCIÓN DE LA RELACIÓN ESTADO NUTRICIONAL Y EL NIVEL DE POBREZA

Estado Nutricional	Nivel de Pobreza		Total
	Pobre	No Pobre	
Desnutrido	381 $_a$	42 $_b$	423
No Desnutrido	1014 $_c$	857 $_d$	1871
Total	1,395	899	2,294

Los resultados de la Tabla 4.7 muestran que la frecuencia de niños (as) con desnutrición como era de esperarse es mayor en la población infantil pobre (381) en relación a la no pobre (42).

4.3.1. Estimación de Odds y Razón Odds (Tabla 2 x 2)

Utilizando los datos de la Tabla 4.7. para estimar el valor de ods. En este sentido el ods estimada de que un niño pobre este desnutrido es

$$\text{Ods} = a/c$$

$$= 381/1014 = 0\ 3757$$

mientras que el ods estimada de que un niño (a) no pobre este desnutrido es

$$\text{Ods} = b/d$$

$$= 42/857 = 0\ 049$$

El cociente entre ambos ods es igual a la razón ods

$$R\hat{O} = 0\ 3757/0\ 049$$

$$= 7\ 667$$

También se puede estimar la RÔ mediante la fórmula

$$R\hat{O} = (a \times d) / (b \times c)$$

$$= 381 \times 587/42 \times 1014$$

$$= 7\ 667$$

Los resultados obtenidos nos permiten concluir que el ods estimado de que un niño(a) pobre este desnutrido es 7 7 veces el ods de un niño (a) no pobre Es decir, los niños pobres están desnutrido 7 7 veces más que los no pobres

La razón ods complementaría de no estar desnutrido es

$$R\hat{O}_{\text{No Desnutrido}} = (1{,}014/381)/ (857/42)$$

$$= 0\ 130$$

__ **Intervalos de Confianza para Estimación de la Razón Ods**

Cuando calculamos la RÔ debemos determinar si este valor es diferente de 1, para ello construimos un intervalo de confianza

El cálculo de dicho intervalo para RÔ para $\alpha = 0\ 05$, utilizando los datos de la Tabla 4 8 se realiza de la siguiente forma

a) Logaritmo natural de 7 667 = 2 037

$$b)\ Error\ \ estandar = \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$$

$$= \sqrt{\frac{1}{381}+\frac{1}{42}+\frac{1}{1014}+\frac{1}{857}}$$

$$= \sqrt{0.0285872} = 0\ 16908$$

c) Para $\alpha = 0\ 05$, el intervalo de confianza de logaritmo de RO es.

$$\text{RÔ} = 2\ 037 \pm 1\ 96 \times 0.16908$$

$$= [1.7056, 2\ 3684]$$

d) El antilogaritmo de estos límites es:

$$e^{1\ 7056} = 5.05$$

$$e^{2\ 3684} = 10.680$$

$$5\ 505 \leq \text{RÔ} \leq 10\ 680$$

Como intervalo no incluye al valor 1, podemos concluir que la RÔ es estadísticamente significativa al nivel $\alpha = 0\ 05$

4.3.2. Análisis de Regresión Logística

Fueron incluidos en el análisis 2,294 datos donde la variable dependiente Y define el estado nutricional (ESTN) del niño (a) como Y = 1 " Desnutrido" o Y = 0 " No Desnutrido". La variable independiente " X" define el nivel de pobreza (LINEA 2PC) del niño (a) como· X = 1 " Pobre" o X = 0 " No Pobre "

__ Estimación de los Parámetros del Modelo

_ Método de Máxima Verosimilitud

En regresión logística como en otros modelos de regresión el método de estimación de los parámetros es el de máxima verosimilitud Este criterio de estimación tiene una base intuitiva, se trata de estimar los coeficientes del modelo bajo el supuesto de que lo que ha ocurrido, la experiencia observada, es lo más probable de lo que podía haber ocurrido

Por ejemplo, si se quiere estimar la probabilidad de que un niño menor de cinco años sea pobre, para la cual se toma una muestra aleatoria de 2,294 niños, de los cuales 1,395 son pobres, luego entonces la probabilidad de ser pobre es el cociente 1,945/2,294=0 608

A continuación veremos como estimar esta probabilidad según el método de máxima verosimilitud Ya que la condición de pobreza en los niños es independiente, se puede demostrar que la probabilidad de que en una muestra de <u>n</u> niños haya <u>r</u> de ellos pobres se puede calcular mediante la expresión

$$P (Y = r) = (r^{n}) P^{r} (1-P)^{(n-r)}$$

Reemplazando los valores de n, r, y p tenemos

$$P(Y = r) = \binom{2,294}{1,395} P^{1,395}(1 - P)^{899} \qquad (78)$$

donde p representa la probabilidad de ser pobre, esta expresión la denotaremos por L(p) estimada como función de parámetro p al estimarse se denomina función de verosimilitud

La estimación de máxima verosimilitud de p que es el valor que hace que la función de verosimilitud sea lo más grande posible, lo representamos con el símbolo

$\hat{p}$

Debido a que el máximo de una función coincide con el máximo de su logaritmo, le aplicaremos logaritmo a la expresión [78]

$$\text{Log}(L) = 1{,}395 \log(p) + 899 \log(1-p)$$

Luego la primera derivada respecto a p obtenemos

$$\frac{\partial Log(L)}{\partial p} = 1{,}395\frac{1}{p} + 899\frac{-1}{1-p}$$

Igualando a cero y despejando en función p tenemos

$$1{,}395\frac{1}{p} + 899\frac{-1}{1-p} = 0$$

$$1{,}395p + 899 = 1{,}395$$

$$\hat{p} = 1{,}395/2{,}294$$
$$= 0\ 608$$

es la estimación de máxima verosimilitud de la probabilidad de que un niño menor de cinco años sea pobre, en base a la muestra de niños elegida

Estimación de los Parámetros

Para estimar los parámetros del modelo utilizaremos como ya se mencionó anteriormente el método de máxima verosimilitud En este análisis consideraremos la variable predictora dicotómica nivel de pobreza (LINEA 2PC), categorizada como pobre y no pobre por lo cual tendremos dos muestras de tamaño n_1 y n_2 respectivamente Sea Y la variable respuesta que define el estado nutricional del niño (a) como desnutrido y no desnutrido Si denotamos Y_1 la variable número de individuos que presentan las características de interés (desnutrición) en la muestra n_1 de la población, e Y_2 el análogo en la muestra n_2. Sean p_1 y p_2 las probabilidades de presentar la característica en esas dos poblaciones y representamos y_1 y y_2 los valores Y_1 e Y_2 en los dos cohortes elegidos. Por otra parte la variable predictora X dicotómica con valores X = 1 " pobre" y X = 0 " no pobre".

TABLA 4.8.
ESTADO NUTRICIONAL Y NIVEL DE POBREZA EN NIÑOS /AS MENORES DE CINCO AÑOS

Estado Nutricional	Nivel de Pobreza		TOTAL
	Pobres (x = 1)	No Pobres (x = 0)	
Desnutrido (Y = 1)	$Y_1 = 381$	$Y_2 = 42$	423
No desnutridos (Y = 0)	$n_1 - Y_1 = 1{,}014$	$n_2 - Y_2 = 857$	1,871
TOTAL	$n_1 = 1{,}395$	$n_2 = 899$	$n = 2{,}294$

En estas condiciones la probabilidad en la primera cohorte se hayan encontrado Y_1 con la característica de interés viene dada por la expresión:

$$P(Y_1 = y_1) = \binom{n_1}{Y_1} P_1^{Y_1} (1 - P_1)^{n_1 - y_1}$$

y, de igual forma la probabilidad de que la segunda cohorte Y_2 individuos presenten la característica de interés será:

$$P(Y_2 = y_2) = \binom{n_2}{y_2} P_2^{y_2} (1 - P_1)^{n_2 - y_2}$$

Dado que existe independencia entre las dos variables la función de probabilidad conjunta estará dada por el producto de ambas funciones:

$$P(Y_1 = y_1).P(Y_2 = y_2) = \prod_{i=1}^{2} P(Y_i = y_i) = \prod_{i=1}^{2} \binom{n_i}{y_i} P_i^{y_i} (1 - p_i)^{n_i - y_i}$$

donde el símbolo $\prod_{i=1}^{2}$ representa el producto de las dos probabilidades.

Debido a que p_i depende de X a través de los parámetros β_1, por lo que el producto anterior es una función de β_0 y β_1.

Luego función de verosimilitud queda:

$$L(\beta_0,\beta_1) = \log\{L(\beta_0,\beta_1)\}$$

$$L(\beta_0,\beta_1) = \log\prod_{i=1}^{2} P(Y_i = y_i) = \sum_{i=1}^{2}\log\{P(Y_i = y_i)\}$$

$$= \sum_{i=1}^{2}\{y_i \log(p_i) + (n_i - y_i)\log(1 - p_i)\}$$

Sustituyendo p_i por su valor

$$p_i = \frac{e^{\beta_0+\beta_1 X_1}}{1+e^{\beta_0+\beta_1 X_1}}$$

Por propiedad de logaritmo

$$L(\beta_0,\beta_1) = \sum_{i=1}^{2}\left[Y_i \log\left(\frac{e^{\beta_0+\beta_1 X_1}}{1+e^{\beta_0+\beta_1 X_1}}\right) + (n_i - y_i)\log\left(1 - \frac{e^{\beta_0+\beta_1 X_1}}{1+e^{\beta_0+\beta_1 X_1}}\right)\right]$$

$$= \sum_{i=1}^{2} y_i(\beta_0 + \beta_1 X_1) - \sum_{i=1}^{2} y_i \log\left(1+e^{\beta_0+\beta_1}\right) - (n_i - y_i)\log(1+e^{\beta_0+\beta_1 X_1})$$

Restando los factores comunes

$$L(\beta_o,\beta_1) = \sum_{i=1}^{2} y_i(\beta_0 + \beta_1 X_1) - \sum_{i=1}^{2}\log n_i\left(1+e^{\beta_0+\beta_1 X_1}\right)$$

El máximo de la función se consigue derivando parcialmente respecto a β_0 y β_1 e igualando a cero ambas derivadas. La derivada del logaritmo de la función de verosimilitud respecto al parámetro β_0.

$$\frac{\partial L(\beta_0,\beta_1)}{\partial \beta_0} = \sum_{i=1}^{2} Y_i - \sum_{i=1}^{2} n_i \frac{e^{\beta_0+\beta_i X_i}}{1+e^{\beta_0+\beta_i X_i}}$$

Tenemos que:

$$Y_1 + Y_2 - n_1 \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} - \frac{n_2 e^{\beta_0}}{1+e^{\beta_0}} = 0$$

y la derivada parcial respecto a β_1 es:

$$\frac{\partial L(\beta_0,\beta_1)}{\partial \beta_1} = \sum_{i=1}^{2} y_i x_i - \sum_{i=1}^{2} n_i x_i \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}$$

Reemplazando el valor de Y_1 en la ecuación anterior tenemos:

$$Y_1 - \frac{n_1 e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} = 0 \qquad (79)$$

$$Y_2 - \frac{n_2 e^{\beta_0}}{1+e^{\beta_0}} = 0 \qquad (80)$$

Despejando β_o de la ecuación [80]

$$n_2 e^{\beta_0} = Y_2(1+e^{\beta_0})$$

Aplicado logaritmo natural

$$\beta_0 \log n_2 - \beta_0 \log Y_2 = \log Y_2$$

$$\beta_0 = \frac{\log Y_2}{\log n_2 - \log Y_2} = \log \frac{Y_2}{n_2 Y_2} \tag{81}$$

Despejando β_1 de la ecuación [79]

$$n_1 e^{\beta_0 + \beta_1} = Y_1 (1 + e^{\beta_0 + \beta_1})$$

Aplicando logaritmo natural y ordenando los factores

$$(\beta_0 + \beta_1) \log n_1 = \log Y_1 + (\beta_0 + \beta_1) \log Y_1$$

Sustituyendo β_0 por su valor en ecuación [81] tenemos:

$$\beta_1 = \log \frac{Y_1}{n_1 - Y_1} - \log \frac{Y_2}{n_2 - Y_2} \qquad \Rightarrow \qquad \beta_1 = -\beta_0 + \frac{\log Y_1}{\log n_1 - \log Y_1}$$

Por propiedad de logaritmo,

$$\beta_1 = \log \left[\frac{\frac{Y_1}{n_1 - Y_1}}{\frac{Y_2}{n_2 - Y_2}} \right]$$

Como se aprecia en este último resultado, la estimación de coeficiente $\hat{\beta}_1$ es el logaritmo de la razón ods. Reemplazando los valores de la Tabla 4.8, obtendremos los siguientes resultados: $n_1 = 1{,}395$, $n_2 = 899$, donde $Y_1 = 381$ y $Y_2 = 42$ son los individuos con la característica de interés (desnutridos) por lo que:

$$\hat{\beta}_0 = \log\frac{42}{857} = -3.016 \quad \text{y} \quad \hat{\beta}_1 = \log\left[\frac{\frac{381}{1{,}014}}{\frac{42}{857}}\right] = 2.037$$

El resultado obtenido del SPSS del modelo de regresión logistica es el siguiente:

TABLA 4.9.
MODELO DE REGRESIÓN LOGÍSTICA SIMPLE
MODELO LN (ESTN) = -3.016 + 2.037 (LINEA2PC)

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	C. 95.0% para EXP(B)	
								Inferior	Superior
Paso 1[a]	LINEA2PC	2.037	.169	145.135	1	.000	7.667	5.504	10.679
	Constante	-3.016	.158	364.139	1	.000	.049		

a. Variable(s) introducida(s) en el paso 1: LINEA2PC.

La interpretación de los coeficientes del modelo:

- La constante del modelo es $\hat{\beta}_0 = -3.016$
- El coeficiente de regresión "$\hat{\beta}_1 = 2.037$" de la variable nivel de pobreza (LINEA 2PC) es positivo, lo que indica que el logaritmo del ods de que un niño seleccionado al azar este desnutrido es 2.037 si es pobre en comparación a los no pobres.

<u>Estimación de la Probabilidad</u>

Luego del análisis de la significancia de los coeficientes el modelo de regresión logística procederemos a estimar probabilidades.

Podemos afirmar que el logit de estar desnutrido para un niño pobre, LINEA2PC= 1, es:

$$\ln (p /1\text{-}p) = -3.016 + 2.037\ (1) = -0.979$$

Por lo que la probabılıdad de estar desnutrıdo entre los nıños pobres será

$$P(Y = 1 / X = 1) = \frac{e^{-0\ 979}}{1 + e^{-0\ 979}}$$

$$= 0\ 3757 / (1 + 0.3757)$$

$$= 0\ 273$$

es decır, el modelo predıce que el 27.3% de los niños pobres están desnutrıdo. Por lo que se espera que de los 1,395 nıños pobres de la Tabla 4.8 haya un total de·

1,395 x 0 273 = 381 nıños desnutrıdos

Si aplıcamos los mısmos cálculos a la probabılıdad predıcha de que un nıño no pobre este desnutrido " P (Y = 1) si X = 0 " se obtıene el sıguıente valor

$$\ln [1/(1-p)] = -3\ 016 + 2\ 037\ (0) = -\ 3\ 016$$

el antılogarıtmo de ln [p/(1-p)] es·

$$p / (1-p) = e^{-3\ 016} = 0\ 049$$

Luego,

$$P (Y = 0 / X = 0) = 0\ 049 / (1 + 0\ 049)$$

$$= 0\ 0467$$

por lo que entre los 899 nıños no pobre es esperable que haya 42 nıños desnutrıdos

899 x 0 0467 = 42

Este resultado nos ındıca que la probabilıdad de que un nıño no pobre este desnutrıdo es mínıma e ıgual a 0.047. La probabılıdades complementarıa de que un niño no este desnutrıdo sı es pobre " X = 1" es de

$$P (Y = 0/ X = 1) = 1 - P(Y=1/ X = 1)$$

$$= 1 - 0\ 27 = 0.73$$

y la probabilidad complementaria de no estar desnutrido si el niño es no pobre " X = 0" será

$$P(Y=0/X=0)=1-P(Y=1/X=0)$$

$$=1-0\ 0467$$

$$=0.9533$$

4.3.3. Análisis del Modelo Estimado

Una vez estimado los parámetros del modelo procederemos a evaluar si el modelo se ajusta bien a los datos En este contexto, se plantean una serie de pruebas tales como a) Desviación, b) La Prueba G o Chi-cuadrado y c) Medida de bondad de ajuste

__ **Desviación**

Para determinar si el modelo se ajusta correctamente a los datos, se plantea el siguiente esquema de hipótesis·

Ho . El modelo es un modelo que se ajusta bien a los datos

H_1 El modelo es un modelo que no se ajusta bien a los datos

Estadígrafo de Prueba.

$$D=2\sum_{j=1}^{k}[y_j\ In\frac{y_j}{m_j\hat{p}_j}+(m_j-y_j)In\frac{m_j-y_j}{m_j-m_j(1-\hat{p}_j)}]$$

$$D=2\left\{381\ln\left(\frac{381}{380\,84}\right)+(1{,}395-381)\ln\left(\frac{6{,}395-381}{1{,}395-380\,84}\right)+42\ln\left(\frac{42}{41.98}\right)+(899-42)\ln\left(\frac{899-42}{899-41\,98}\right)\right\}$$

$$=2\{0\ 16-0\ 16+0.02-0\ 02\}$$

$$=0$$

TABLA 4.10.
PROBABILIDAD Y VALORES ESPERADOS
PREDICHO POR EL MODELO : ln (p/ 1-p) = -3.016 + 2.037 LINEA 2PC

Desnutridos	Pobres			No Pobres		
Observados	381			42		
Esperados o Predicho	m_1	$\hat{P}_1$	$m_1\hat{P}_1$	m_2	$\hat{P}_2$	$m_2\hat{P}_2$
	1,395	0.273	380.84	899	0.0467	41.98

Regla de Decisión:

Cuando los datos están agrupados la estadística se distribuyen como una chi-cuadrado con k-v grado de libertad. En nuestro caso k =2, por que tenemos patrones distintos de la variable predictora (pobre y no pobre) y v =2, por que se estiman dos parámetros. Por lo tanto, el modelo ajustado tiene o grado de libertad. Sin embargo, siempre que la estadística G sea menor que el chi-cuadrado correspondiente no hay evidencia de mal ajuste la ajuste. Podemos en consecuencia aceptar la hipótesis nula de que el modelo es un modelo que se ajusta bien a los datos.

__ **Medidas de Bondad de Ajuste**

Las medidas de bondad ajuste que analizaremos son el Pseudo - R^2 o R cuadrado de Cox y Snell y el R cuadrado de Nagelkerke.

$$\text{Pseudo - } R^2 = \text{chi-cuadrado} / (\text{chi-cuadrado} + n)$$

$$= 217.798 / (217.798 + 2294) = 0.09$$

Lo que nos muestra que el 9% de la varianza total de la desnutrición infantil es explicada por la variable incluida en el modelo (nivel de pobreza). Es normal que su valor sea bajo cuando los datos son individuales.

TABLA 4.11.

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	1975 227	091	147

Como se puede observar el Pseudo-R^2 y el R-Cuadrado de Nagelkerke mejoraron respecto al modelo con la variable edad Ambos expresan la relación moderada entre las variables incluidas en el modelo

__ **Significancia de los Coeficientes del Modelo**

- Estadística G

Determinaremos si existe una relación significativa entre la variable X y Y al probar si β_1 (la pendiente verdadera) es igual a cero La hipótesis nula y alternativa se pueden establecer de la siguiente forma.

$$H_0 \cdot \beta_1 = 0$$

$$H_1 \quad \beta_1 \neq 0$$

Estadígrafo de Prueba

$$G = 2\left\{\sum_{i=1}^{n}\left[y_j \ln(\hat{p}_j) + (1 - y_j)\ln(1 - \hat{p}_j)\right] - \left[n_1 \ln(n_1) + n_0 \ln n_0 - n \ln \quad n\right]\right\}$$

La primera sumatoria es la log verosimilitud del modelo con la variable nivel de pobreza y el segundo término es la log verosimilitud del modelo con sólo el término constante

El cálculo de la estadística G no resulta difícil por ser la variable predictora dicotómica, permitiendo agrupar los datos en dos patrones en relación a la predictora. pobre y no pobre (j = 2)

Los resultados se muestran a continuación :

TABLA 4.12.
VALORES ESPERADOS Y PROBABILIDAD PREDICHA

Nivel de Pobreza	Y	$\hat{P}$	$Y \ln \hat{p}$	(1-y)	$(1 - \hat{p})$	$(1-Y) \ln (1-\hat{p})$
Pobre	381	0.273	-494.65	1,014	.727	-323.29
No Pobre	42	0.0467	-128.69	857	0.9533	-40.99
	$\Sigma Y_j = 423$		$\Sigma Y_i \ln \hat{p}_i = -623.34$			$\Sigma(1-Y_i) \ln (1 - \hat{p}_i) = -364.28$

Luego la verosimilitud del modelo con la variable :

$$\sum_{j=1}^{n}[y_j \ln(\hat{p}_j)+(1-y_j)\ln(1-\hat{p}_j)]=\sum_{j=1}^{m}Y_j \ln \hat{p}_j+\sum_{j=1}^{n-m}(1-y_i)\ln(1-\hat{p})$$

$$= -623.34 + (-364.28)$$

$$= -987.62$$

De este resultado se puede deducir el valor de la -2log de la verosimilitud del modelo igual a:

$$2\,(-987.62) = 1,975.24$$

La log verosimilitud del modelo con solo el término constante es:

$$n_1 \ln (n_1) + n_0 \ln (n_0) - n \ln (n) = 423 \ln (423) + 1,871 \ln (1,871) - 2,294 \ln (2,294)$$

$$= -1096.5123$$

Luego la estadística G es igual a la diferencia la verosimilitud del modelo y la log verosimilitud del modelo con solo el término constante:

$$G = 2 \{ -987.62 - (-1,096.5123) \}$$

$$= 217.78$$

Regla de Desición

Como 217 78 > χ^2 (1) = 3 841, podemos deducir que el ajuste de modelo mejora con la inclusión de la variable nivel de pobreza. En otras palabras esta variable es significante prediciendo el estado nutricional del niño o niña

Los resultados obtenidos a través del programa SPSS se muestra a continuación, los cuales coinciden con los cálculos manuales.

Pruebas omnibus sobre los coeficientes del modelo

		Chi-cuadrado	gl	Sig
Paso 1	Paso	217 798	1	000
	Bloque	217 798	1	000
	Modelo	217 798	1	000

En síntesis podemos inferir que el modelo de regresión logística objeto de análisis es estadísticamente significativo con un valor chi-cuadrado de 217.798 y un valor p=0.000; lo que indica que la probabilidad de obtener un valor chi-cuadrado; con un mayor grado de libertad, mayor que 217 798 es 0 Por lo tanto se rechaza la hipótesis nula de que β_1 es cero en la población

Un segundo método equivalente para probar la existencia de una relación entre las variables X y Y consiste en establecer una estimación de intervalo de confianza β_1 y determinar si el valor supuesto ($\beta_1 = 0$) esta incluido en el intervalo.

El modelo de regresión logística permite el cálculo del error estándar y el contraste de hipótesis y la estimación de los intervalos de confianza en torno al coeficiente $\hat{\beta}_1$ y el valor de la RÔ

La fórmula para el cálculo del error estándar de $\hat{\beta}_1$ es

$$\hat{\mathrm{E}}\mathrm{E}(\hat{\beta}_1) = \sqrt{(1/n_1 p_1 q_1 + 1/n_2 p_2 q_2)}$$

donde "n_1" es el número de niños pobres es; " n_2" es el número de niños no pobres, " p_1" la proporción de niños desnutridos pobres y "p_2" la proporción de niños desnutridos no pobres, "q_1" es la proporción de niños pobres no desnutridos y "q_2" es la proporción de niños no pobres no desnutridos

Los resultados obtenidos de la muestra de la Tabla 4 8 son los siguientes.

$n_1 = 1{,}395$ $\qquad$ $n_2 = 899$

$p_1 = 381/1395 = 0\ 2731$ $\qquad$ $p_2 = 42/899 = 0\ 04672$

$q_1 = 1014/1395 = 0\ 7269$ $\qquad$ $q_2 = 857/899 = 0\ 9533$

Remplazando los valores en la fórmula anterior tenemos

$$\hat{E}E(\hat{\beta}_1) = \sqrt{\left[\frac{1}{1395(0\ 2731)(0\ 7269)}\right] + \left[\frac{1}{899(0.04672)(0\ 9533)}\right]}$$

$$= 0\ 169$$

El coeficiente " β_1" también se puede estimar a partir de la fórmula:

$$\hat{\beta}_1 = \ln(p_1/q_1) - \ln(p_2/q_2)$$

$$= \ln\left(0\ 2731/0\ 7269\right) - \ln\left(0\ 04672/0\ 9533\right)$$

$$= 2\ 037$$

La estimación de intervalos de confianza de $\hat{\beta}_1$ se obtendrán mediante la siguiente fórmula

$$\hat{\beta}_1 \pm Z_{\alpha/2}\ \hat{E}E(\hat{\beta}_1) \quad \begin{aligned} &= 2\ 037 \pm\ 1\ 96\ (0.169) \\ &= 2\ 037 \pm\ 0\ 33124 \\ &1.7058 \le \beta_1 \le 2\ 3682 \end{aligned}$$

Hay que tomar en consideración sin embargo que los estimadores habituales de la asociación no son los coeficientes $\hat{\beta}_i$ sino los de la razón ods, por lo tanto éstos son los intervalos que se calculan frecuentemente

__ **Estimación Ods**

- Nivel de Pobreza: Pobre

El logaritmo natural del ods estimado de que un niño pobre este desnutrido, LINEA 2PC = 1, será

$$\text{logit} = \ln (\text{ods})_p = -3\ 016 + 2\ 037\ (1) \tag{82}$$

Aplicando antilogaritmo obtenemos el valor del ods

$$(\text{ods})_p = e^{-3\ 016 + 2\ 037\ (1)} = 0\ 3757 \tag{83}$$

- Nivel de Pobreza: No Pobre

El logaritmo natural del ods estimada de que un niño no pobre este desnutrido, LINEA2PC = 0, será

$$\text{logit} = \ln (\text{ods})_{np} = -3\ 016 + 2\ 037\ (0)$$

Luego aplicando antilogaritmo, obtenemos el ods

$$(\text{ods})_{np} = e^{-3\ 016 + 2\ 037\ (0)} \tag{84}$$

$$= e^{-3\ 016}$$

$$= 0\ 049 \tag{85}$$

__ Razón Ods

La estimación de la razón ods se basa en la diferencia de los logits; o en su lugar en el cociente del ods para los niños pobres respecto a los no pobres

El logits, es decir, el logaritmo natural del ods de un niño no pobre este desnutrido, LINEA2PC = 0, será

$$\ln\left(\frac{p}{1-p}\right)_{Np} = \hat{\beta}_0 + \hat{\beta}_1(0) = \hat{\beta}_0 \qquad (86)$$

El logit por desnutrición de que un niño pobre, LINEA2PC = 1, será

$$\ln\left(\frac{p}{1-p}\right)_{p} = \hat{\beta}_0 + \hat{\beta}_1(1) = \hat{\beta}_0 + \hat{\beta}_1 \qquad (87)$$

De la ecuación [87] se despeja $\hat{\beta}_1$ y se reemplaza $\hat{\beta}_o$ por su valor, según la ecuación [86] tendríamos que.

$$\hat{\beta}_1 = \ln\left(\frac{p}{1-p}\right)_{p} - \hat{\beta}_0$$

$$\hat{\beta}_1 = \ln\left(\frac{p}{1-p}\right)_{p} - \ln\left(\frac{p}{1-p}\right)_{Np} \qquad (88)$$

Este último resultado es la diferencia de la logits o ln (p/1-p) por desnutrición de un niño pobre respecto a uno no pobre

Siguiendo el procedimiento matemático para la obtención de la RÔ, tenemos que por propiedad de logaritmo la ecuación [88] se convierte en

$$\hat{\beta}_1 = \ln\left[\frac{\left(\frac{p}{1-p}\right)_{p}}{\left(\frac{p}{1-p}\right)_{Np}}\right]$$

Luego para obtener RÔ debemos aplica el antilogaritmo

$$R\hat{O} = e^{\hat{\beta}_1}$$

Utilizando el resultado de la ecuación [88] sobre la diferencia del logaritmo natural del ods de desnutrición en niños pobres y no pobres respectivamente, tendríamos de manera directa que el ln RÔ es igual a la diferencia entre las ecuaciones [87] y [86].

$$\ln R\hat{O} = \hat{\beta}_o - \hat{\beta}_o + \hat{\beta}_1$$

$$R\hat{O} = e^{\hat{\beta}_1}$$

Realizando el cálculo matemático de la RÔ a partir de la ecuación [88] y utilizando los resultados obtenidos en [82] y [84] tendríamos que la RÔ por desnutrición de un niño pobre respecto a un no pobre es de.

$$\ln(R\hat{O}) = -3\ 016 + 2\ 037\,(1) - [-3\ 016 + 2.037(0)]$$

$$= 2\ 037$$

Luego la razón ods es igual a .

$$R\hat{O} = e^{2\ 037}$$
$$= 7\ 668$$

La razón ods también se puede estimar a partir del cociente de las dos ods, obtenidas en [83] y [85] Así tenemos que el ods de que un niño este desnutrido si es pobre es

$$e^{[(-3\ 016+2\ 037(1)]} = 0.3757$$

mientras que el ods de estar desnutrido un niño no pobre sería

$$e^{[-3\ 016+2\ 037(0)]} = 0.049$$

Luego la razón ods también se podría estimar mediante el cociente de ambas ods

0 3757/0 049 = 7 667

Este valor ındica que al seleccıonar un nıño al azar, sı es desnutrıdo es 7.7 veces más frecuente que sea pobre a que sea no pobre Como podemos observar los resultados para la razón ods obtenidas con el modelo de regresıón logístıca coıncıden con los cálculos ınıcıales con la Tabla 4 7.

__ **Intervalos de confianza para RÔ**

Antes de hacer inferencıa sobre la RÔ debemos conocer si su valor es dıferente de 1. Para tal efecto se construyen los ıntervalos de confianza El intervalo de confianza al 95% para RÔ esta dado por.

$$e^{\hat{\beta}_1 \pm Z_{\alpha/2}\hat{E}E(\hat{\beta}_1)}$$

que para nuestro estudıo correspondera a

$$e^{2\ 037 \pm 1.96(0\ 169)}$$

$$(e^{1\ 70576}, e^{2\ 36824}) = (5\,504, 10.679)$$

Como el ıntervalo no incluye al 1 podemos afirmar que el rıesgo de desnutrıción asocıado al nıvel de pobreza es estadístıcamente sıgnıficatıvo, al margen de error prefijado

Este ıntervalo tambıén nos ındıca que con una confianza del 95%, se espera que la verdadera RÔ el parámetro a estımar se encuentre como mínımo en 5 504 y como máxımo 10 68, dicho de otra manera, con una confianza del 95% podemos decır que un niño pobre esta como poco 5 5 veces y como mucho a 10 68 veces más rıesgo de padecer desnutrición que un nıño no pobre.

Una vez verificada de sıgnıficancıa de la razón ods podemos concluır lo sıguıente la ods estımada por desnutrıcıón sı el niño es pobre es 7 688, el ods de un nıño no pobre. En otras palabras, la probabılıdad de que un nıño pobre este desnutrıdo es un 666 76% veces mayor en comparacıón a los niños no pobre

$$100\,(e^{2\,037} - 1) = 100\,(7\,668 - 1)$$

$$= 666\,76\%$$

Tambıén puede ınterpretarse que el rıesgo de desnutrıcıón es aproxımadamente 7 7 veces más alto en los nıños pobres que entre los no pobres.

4.4. MODELO DE REGRESIÓN LOGÍSTICA: CON UNA VARIABLE CATEGÓRICA

El modelo que se plantea a contınuacıón tıene como propósıto determınar, sı el área geográfica donde resıde el nıño, esta relacıonada con el hecho de estar desnutrıdo. La varıable dependiente Estado Nutrıcıonal se define como Y = 1 "Desnutrıdo", Y= 0 " No desnutrıdo", mientras que la varıable independıente es Área Geográfica " X_1 = Área" y puede asumır tres categorías urbana, rural e ındígena Para poder ıncluir esta varıable cualıtatıva en el modelo de regresıón logística, debemos prımeramente crear una varıable ındıcadora o dummy Así por ejemplo, la varıable Área Rural adoptaría un valor “1” para los niños y niñas que resıden en esa área y “0” para todos los ındivıduos que no están representado por esa categoría de área

Para crear una variable ındıcadora prımeramente debemos elegır una categoría de referencıa En nuestro estudıo el área referencıal X_0 será el área urbana por ser una categoría numerosa y por razones de fiabılıdad estadístıca, ya que sı tomamos una categoría con pocas observacıones, todas las comparacıones con las categorías restantes van a ser pocos fiables

El área urbana presenta la mejor sıtuación de la nıñez; así esperaremos que los coeficientes del modelo regresión logística sean positivos, denotando de esta forma el incremento de la desnutrición infantil que se dan en las áreas rurales e indígenas en relación al área urbana.

Codificación Tipo Indicador con Referencia a la Primera Categoría X_0

La variable independiente, AREA Geográfica definida como: Area (0) = "Urbana", Area (1) = "Rural" y Area (2) = "Indígena". La tres variables indicadoras o dummy que representarán las tres categorías de área se construirán a través de un diseño de variables indicadores y que se presentan a continuación:

TABLA 4.13.
VARIABLES INDICADORAS DE AREA GEOGRÁFICAS GENERAL

Área Geográfica	$X_{1\ urbana}$	$X_{2\ rural}$	$X_{3\ indígena}$
Área (0)	1	0	0
Área (1)	0	1	0
Área (2)	0	0	1

En la Tabla anterior al incluir las dos variables Área (1) y Área (2) en el modelo de regresión logística, la categoría de referencia AREA "urbana" estará representada por los valores "0" de esas dos variables.

De esta forma nuestro modelo genera valores β_i y $\hat{RO}$ que revelarán el incremento del fenómeno de la desnutrición en el área rural e indígena respecto al área urbana (categoría de referencia) según nuestra hipótesis.

Los datos de la Tabla 4.14. que se muestra más adelante nos permite estimar los diferentes valores de Razón Ods, adoptando la categoría de área urbana como referencia de la siguiente forma:

$$R\hat{O}_{(\text{área rural})} = (822 \times 151) \ / \ (829 \times 54) = 2.773$$

$$R\hat{O}_{(\text{área indígena})} = (822 \times 281) \ / \ (220 \times 54) = 15.084$$

TABLA 4.14.
DESCRIPCION DE LA RELACION ESTADO NUTRICIONAL Y ÁREA GEOGRÁFICA

Estado Nutricional	Area			Total
	Rural	Indígena	Urbana	
No Denutrido	829	220	822	1871
Desnutrido	151	218	54	423
Total	**980**	**438**	**876**	**2,294**

4.4.1 Análisis de Regresión Logística

Con el propósito de determinar si el área geográfica está relacionada con el hecho de estar desnutrido, se incluye la variable ÁREA en el modelo de regresión logística, según el esquema descrito en la Tabla 4.12. La ecuación de regresión logística resultante será:

$$\ln [p / (1-p)] = -2.722 + 1.019 \text{ AREA } (1) + 2.713 \text{ AREA } (2)$$

__ **Estimación de Probabilidades**

El modelo de regresión logística también permite estimar probabilidades usando la fórmula:

$$P(Y = 1) = \left[1 \Big/ \left(1 + e^{-(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2)}\right) \right]$$

Así, para el Área Indígena donde ÁREA (1) = 0 y ÁREA (2) = 1, la probabilidad estimada que un niño seleccionada al azar este desnutrido será:

$$P(Y=1/X= \text{ÁREA } (2) = 1) = [1 / 1+ e^{-(-2.722+1.019(0) + 2.713(1))}]$$

$$= 1 / (1 + e^{-.009}) = 0.4978$$

Para el Área (1) = "Rural" donde ÁREA (1) = 1 y ÁREA (2) = 0, la probabilidad estimada sería de.

$$P(Y=1/ \text{Área } (1)) = [1 / e^{-(-2\ 722+1\ 019\ (1) + 2\ 713(0))}]$$

$$= 1 / (1 + e^{-(-1\ 703)}) = 0\ 1541$$

En el caso del Área (0) = "Urbana" donde ÁREA (1) = 0 y ÁREA (2) = 0, la probabilidad estimada estará dada por

$$P(Y=1/ \text{Área } (0)) = [1 / e^{-(-2\ 722 + 0\ 381\ (0) + 1\ 911\ (0))}]$$

$$= 1 / (1 + e^{2\ 722}) = 0\ 0616$$

Cabe señalar que los resultados obtenidos son equivalentes a los que se obtendrían a través de la muestra Por ejemplo, la probabilidad estimada de que un niño seleccionado al azar en el área indígena este desnutrido es·

$$P\ (Y=1 \backslash \text{AREA } (2)) = 218/438$$

$$= 0\ 4978$$

4.4.2. Análisis del Modelo Estimado

__ <u>**Evaluación de la Bondad de Ajuste**</u>

Para medir la bondad de ajuste del modelo, se analizan los siguientes indicadores. a) Desviación; b) Prueba de Hosmer - Lemeshow, y c) Medidas de bondad de ajuste

_ ***Desviación***

Para esta estadística se plantea el siguiente esquema de hipótesis nula y alternativa.

Ho : El modelo es un modelo que ajusta bien

H_1 · El modelo no es un modelo que ajusta bien

<u>Estadígrafo de Prueba.</u>

$$D = 2\sum_{j=1}^{k}[y_j \ \ In\frac{y_j}{m_j \hat{p}_j} + (m_j - y_j)In\frac{m_j - y_j}{m_j - m_j(1-\hat{p}_j)}]$$

$$D = 2\ \{151 \ln (151 / 151.02) + (980 - 151) \ln [(980 - 151)/ (980-151.02)] + 218 \ln (218/218.04) + (438-218) \ln [(438-218) / (438-218.04)] + 54 \ln (54/50.04) + (822-54) \ln [(822-54) / (822-50.64)]\}$$

$=0.22$

Donde $j = 3$ por los 3 patrones distintos de la predictora (rural, indígena y urbana) y $v = 2$ por que se estiman 2 parámetros. Por lo tanto, el modelo ajustado tiene 1 grado de libertad.

TABLA 4.15.
PROBABILIDAD Y VALORES ESPERADOS POR DESNUTRICIÓN, SEGÚN, PREDICE EL MODELO: $\ln(p/1-p) = -2.722 + 1.019$ AREA (1) + 2.713 AREA (2)

Desnutrido (Y=1)	AREA								
	Rural			Indígena			Urbana		
Observados	151			218			54		
Esperados o predichos	m_1	$\hat{P}_1$	$m_1 \hat{P}_1$	m_2	$\hat{P}_2$	$m_2 \hat{P}_2$	m_3	$\hat{P}_3$	$m_{13} \hat{P}_3$
	980	0.1541	151.02	438	0.4978	218.04	822	0.0616	50.64

<u>Regla de Decisión:</u>

Como $D = 0.22 < \chi_{0.95}^2 (1) = 3.841$, en consecuencia se acepta la hipótesis nula, es decir, el modelo es un modelo que se ajusta bien a los datos, al nivel prefijado del 5%.

- Prueba de Hosmer y Lemeshow

El ajuste del modelo es muy bueno según la prueba de Hosmer y Lemeshow, ya que un valor alto de $p = 0.997$ está asociado con un buen ajuste de los datos con el modelo.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	.000	1	.997

_ Medidas de Bondad de Ajuste

El valor del Pseudo – R^2 ó R cuadrado de Cox y Snell es igual a 0.13. Lo que no muestra que el 13% de la varianza total de la desnutrición infantil es explicada por la variable incluida en el modelo AREA. Es normal que su valor sea bajo cuando los datos son individuales, sin embargo indica la presencia de una asociación moderada Mientras que el R - Cuadrado de Nagelkerke = 0 223, revela una asociación alta entre ambas variables

TABLA 4.16.

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	1854 977	137	223

__ *Significancia de la Variable*

El modelo de regresión logística descrito es estadísticamente significativo con un valor chi-cuadrado de 338 047 y un valor p = 0 00 Los resultados de la prueba de Omnibus permiten rechazar la hipótesis nula (H_o) de que los coeficientes β_i es cero, es decir, (H_o $\beta_{área(1)} = \beta_{área(2)}=0$)

Pruebas omnibus sobre los coeficientes del modelo

		Chi-cuadrado	gl	Sig
Paso 1	Paso	338 047	2	000
	Bloque	338 047	2	000
	Modelo	338 047	2	000

Estadígrafo de Prueba

Cabe destacar que el cálculo de la estadística chi-cuadrado se obtiene a través de la fórmula

$$G = 2\left\{\sum_{i=1}^{n}\left[y_i In\ \hat{p}_i + (1-y_i) In(1-\hat{p}_i)\right] - \left[n_1 In(n_1) + n_0\ In(n_0) - n\ In(n)\right]\right\}$$

Así, desarrollando el primer término, o sea la log verosimilitud del modelo con la variable AREA a partir de los datos de la Tabla 4 17 que se presenta más adelante

$$\sum_{i=1}^{n}[(Y_i \ln(\hat{p}_i) + (1-y_i)\ln(1-\hat{p}_i)] = \sum_{j=1}^{m} Y_i \ln \hat{p}_i - \sum_{j=1}^{n-m}(1-y_i)\ln(1-\hat{p}_i)$$

$$= -584.96 - 342.53$$

$$= -927.49$$

Este valor se puede obtener a partir de la Tabla 4.16. de la salida del SPSS equivalente a –2log de la verosimilitud igual a –2(-927.49) = 1,854.977.

El segundo término de la ecuación referente a la log verosimilitud del modelo con sólo la constante es igual a:

$$n_1 \ln(n_1) + n_0 \ln(n_0) - n\ln(n) = 423 \ln(423) + 1{,}871 \ln(1{,}871) - 2{,}294 \ln(2{,}294)$$

$$= -1{,}096.5133$$

La estadística G es igual a:

$$G = 2\{-927.499 - (-1{,}096.51)\}$$

$$= 338.04$$

<u>Regla de Decisión:</u>

Por lo tanto, como $G = 338.04 > \chi^2_{0.95}(2) = 5.991$, el ajuste del modelo mejora con la adición de la variable AREA. También existe evidencia que la variable AREA es significativa en la predicción del estado nutricional.

TABLA 4.17.
DISTRIBUCIÓN DEL ESTADO NUTRICIONAL, POR EDAD Y PROBABILIDAD PREDICHA: MODELO LN (ESTN) = -2.722 + 1.019 AREA (1) + 2.713 AREA (2)

Edad	Desnuridos Y =1	$\hat{p}_i$	$Y \ln \hat{p}_i$	No Desnutrido (1-Y)	$(1-\hat{p}_i)$	$(1-Y)\ln(1-\hat{p}_i)$
Rural	151	0.1541	-282.39	829	0.8459	-138.74
Indígena	218	0.4978	-152.07	220	0.5022	-151.53
Urbana	54	0.0616	-150.50	822	0.9384	-52.26
	$\Sigma y_i = 423$		$\Sigma y_i \ln p_i = -584.96$			$\Sigma(1-y_j)\ln(1-\hat{p}_j) = -342.53$

Los coeficientes estimados del modelo ajustado·

$$\ln [\ ESTN] = 2\ 722 + 1\ 019\ AREA\ (1) + 2\ 713\ AREA\ (2)$$

TABLA 4.18.
MODELO: ln (ESTN) = -2.722 + 1.019 AREA (1) + 2.713 AREA (2)

Variables en la ecuación

		B .	E T	Wald	gl	Sig	Exp(B)	I C 95 0% para EXP(B)	
								Inferior	Superior
Paso 1[a]	AREA			304 647	2	000			
	AREA(1)	1 019	166	37 709	1	000	2 771	2 002	3 837
	AREA(2)	2 713	170	255 071	1	000	15 077	10 807	21 034
	Constante	-2 722	140	375 670	1	000	066		

a Variable(s) introducida(s) en el paso 1 AREA

Los coeficientes de regresión $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ puede interpretarse de la manera siguiente·

- La constante de regresión $\hat{\beta}_0$ es igual a -2.722
- El coeficiente $\hat{\beta}_1 = 1\ 019$ de la variable AREA (1) corresponde al logaritmo de la razón ods de padecer desnutrición en el área rural respecto al área urbana, es decir, los niños y niñas en el área rural están a $e^{1\ 019} = 2\ 771$ veces más riesgo de sufrir desnutrición que los del área urbana.
- El coeficiente $\hat{\beta}_2 = 2\ 713$ de la variable AREA (2) corresponde al logaritmo de la razón de padecer desnutrición en el área indígena respecto al área urbana, es decir, los niños y niñas en esta área están a $e^{2\ 713} = 15.077$ veces más riesgo de sufrir desnutrición que los del área urbana

Cabe destacar que la RÔ estimadas son estadísticamente significativas al no incluir el valor 1

4.4.3. Estimación del Logit y el Ods

Una vez valorada la idoneidad del modelo, se puede medir la magnitud de la asociación de la variable dependiente con las categorías de la variables indicadoras del área.

Área Urbana

Para el ÁREA URBANA, tenemos que AREA(1)=0 y AREA(2) = 0, por lo tanto tendremos que el logit esta dado por

$$\ln\,[p/\ 1\text{-}p]_{AREA(0)} = \hat{\beta}_0 + \hat{\beta}_1\ (0) + \hat{\beta}_2\ (0) = \hat{\beta}_0 \tag{89}$$

$$\ln\,[p/\ 1\text{-}p]_{AREA(0)} = -2\ 722 + 1\ 019\ (0) + 2\ 713\ (0) = -2\ 722$$

Luego el logit o el logaritmo del ods estimada de desnutrición en el AREA URBANA es igual a

$$\ln\,(\text{ods}) = -2\ 722$$

De esta forma el ods para esta área es igual a

$$e^{\hat{\beta}_0} = e^{-2\ 722} = 0\ 0657$$

Área Rural

Para el AREA RURAL, tenemos que AREA (1) = 1 y AREA (2) = 0 luego la estimación del modelo será.

$$\ln\,[p/\ 1\text{-}p]_{AREA(1)} = \hat{\beta}_0 + \hat{\beta}_1\ \text{AREA}\ (1) + \hat{\beta}_2\ \text{AREA}\ (0)$$

$$\ln\,[p/1\text{-}p]_{AREA\ (1)} = \hat{\beta}_0 + \hat{\beta}_1\ \text{AREA}\ (1) \tag{90}$$

Reemplazando por los coeficientes estimados

$$\ln\,[p/\ 1\text{-}p]_{AREA(1)} = -2\ 722 + 1\ 019\ (1) + 2\ 713\ (0)$$

$$\ln\,[p/\ 1\text{-}p]_{AREA(1)} = -2\ 722 + 1\ 019\ (1)$$

El ods para el ÁREA RURAL es

$$e^{\hat{\beta}_0 + \hat{\beta}_1} = e^{-2\ 722 + 1\ 019} = e^{-1\ 703} = 0\ 1821$$

Área Indígena

Para el AREA INDÍGENA, tenemos que AREA (1) = 0 y AREA (2) =1 el logaritmo del ods o la estimación de la logit del modelo, esta dado por

$$\ln [p/ 1\text{-}p]_{AREA(2)} = \hat{\beta}_0 + \hat{\beta}_1 \text{ AREA } (0) + \hat{\beta}_2 \text{ AREA } (1)$$

o en su lugar

$$\ln [p / 1\text{-}p]_{AREA (2)} = \hat{\beta}_0 + \hat{\beta}_2 \qquad (91)$$

Reemplazando por los coeficientes estimados

$$\ln [p/ 1\text{-}p]_{AREA(2)} = -2\ 722 + 1.019\ (0) + 2\ 713\ (1)$$

$$\ln [p/ 1\text{-}p]_{AREA(2)} = -2.722 + 2\ 713$$

Luego el ods para el AREA INDÍGENA es

$$e^{\hat{\beta}_0 + \hat{\beta}_2} = e^{-2\ 722 + 2\ 713} = e^{-0\ 009} = 0\ 991$$

4.4.4. Estimación de la Razón Ods

— **Área Rural**

La razón ods resulta de la diferencia entre el logit del AREA RURAL y el AREA URBANA. Es decir el riesgo de un niño de padecer desnutrición en el ÁREA RURAL respecto a la URBANA se obtiene a continuación

Despejando $\hat{\beta}_1$ de la expresión (90) y reemplazando $\hat{\beta}_o$ por $\ln [p / 1\text{-}p]_{AREA (0)}$, según resultado obtenido en (89) tenemos:

$$\hat{\beta}_1 = \ln [p / 1\text{-}p]_{AREA (1)} - \hat{\beta}_0$$

$$\hat{\beta}_1 = \ln [p / 1\text{-}p]_{AREA (1)} - \ln [p / 1\text{-}p]_{AREA (0)} \qquad (92)$$

por propiedad de logaritmo,

$$\hat{\beta}_1 = \ln\left[{(p/1-p)_{AREA(1)}} \Big/ {(p/1-p)_{AREA\,(0)}} \right]$$

$$\hat{\beta}_1 = \ln \hat{RO}$$

así se cumple,

$$\hat{\beta}_1 = 1\ 019$$

Este resultado es el logaritmo natural de la razón ods de desnutrición del AREA RURAL respecto al AREA URBANA Por tanto, la RÔ será igual a su antilogaritmo

$$e^{\hat{\beta}_1} = e^{1\ 019}$$

$$= 2\ 770$$

Otra forma de llegar a este resultado, sería mediante la división del ods del área rural, respecto a del área urbana

$$R\hat{O} = \frac{0\ 1821}{0.0657} = 2\ 771$$

__ **Área Indígena**

La razón ods del área indígena respecto a la urbana, se obtiene por la diferencia entre el logit del área indígena y urbana

$$\ln [\ p\ /\ 1\text{-}p]_{AREA\ (2)} = \hat{\beta}_0 + \hat{\beta}_2$$

$$\hat{\beta}_2 = \ln [\ p\ /\ 1\text{-}p]_{AREA\ (2)} - \hat{\beta}_0$$

$$\hat{\beta}_2 = \ln [\ p\ /\ 1\text{-}p]_{AREA\ (2)} - \ln [\ p\ /\ 1\text{-}p]_{AREA\ (0)} \qquad (92)$$

por propiedad del logaritmo

$$\hat{\beta}_2 = \ln\left[{(p/1-p)_{AREA(2)}}\Big/{(p/1-p)_{AREA(0)}}\right]$$

$$\hat{\beta}_2 = \ln \hat{RO}$$

Realizando la diferencia, según resultado [92] se tiene

$$\hat{\beta}_2 = [\ -2\ 722 + 1\ 019\ (0) + 2\ 713\ (1)] - [\ -2\ 722 + 1\ 019\ (0) + 2\ 713\ (0)]$$

$$= 2\ 713$$

Este resultado corresponde al logaritmo natural de RÔ de padecer desnutrición en el área rural respecto a la urbana Por tanto, aplicando antilogaritmo

$$R\hat{O} = e^{-2\ 713} = 15\ 09$$

Los niños y niñas indígenas estarán 15 09 veces más en riesgo de estar desnutrido que los niños urbanos En otras palabras, la probabilidad de desnutrición de un niño de área indígena es un 1,407 4 veces mayor en comparación con los del área urbana

$$100(e^{2\ 713} - 1) = 1{,}407\ 4$$

Si queremos comparar el área indígena con el área rural no hay más que calcular sus respectivas logits y realizar sus diferencias, haciendo el ejercicio matemático previo tendremos   Logit del AREA RURAL es igual a

$$\ln\ [p/\ 1\text{-}p]_{AREA(1)} = -2\ 722 + 1.019 = \hat{\beta}_0 + \hat{\beta}_1 \qquad (93)$$

y la logit del ÁREA INDÍGENA

$$\ln\ [p/\ 1\text{-}p]_{AREA(2)} = -2\ 722 + 2\ 713 = \hat{\beta}_0 + \hat{\beta}_2 \qquad (94)$$

Realizando la diferencia entre [93] y [94] tenemos

$$\hat{\beta}_2 - \hat{\beta}_1 = \ln\ [p/\ 1\text{-}p]_{AREA(2)} - \ln\ [p/\ 1\text{-}p]_{AREA(1)}$$

$$\hat{\beta}_2 - \hat{\beta}_1 = \ln\left[\left(\frac{p}{1-p}\right)_{AREA(2)} \Big/ \left(\frac{p}{1-p}\right)_{AREA(1)}\right]$$

$$\hat{\beta}_2 - \hat{\beta}_1 = \ln R\hat{O}$$

por tanto,

$$R\hat{O} = e^{\hat{\beta}_2 - \hat{\beta}_1}$$

$$= e^{2\ 713\ -\ 1\ 019}$$

$$= 5\ 441$$

Podemos entonces destacar, que para estimar la RÔ entre dos categorías diferentes sólo bastaría hacer la diferencia entre sus respectivas logit, donde la categoría de referencia asume el signo negativo y posteriormente se le calcula el antilogaritmo. En el caso nuestro, la diferencia directa sería:

$$-2.722 + 2.713 - (-2.722 + 1.019) = 1.694$$

y su antilogaritmo

$$e^{1.694} = 5.441 \cong 5.4$$

Otra forma sería a través de la división entre los ods de ambas AREA:

$$R\hat{O} = \frac{0.9910}{0.1821} = 5.441$$

__ **Codificación Desviación con Referencia a la Primera Categoría de X_0**

La codificación utilizada anteriormente es la más utilizada en las ciencias de la salud; sin embargo no es la única. Otra forma de construir variables indicadoras para variables AREA podría ser la que se muestra en la Tabla 4.19. basada en la codificación de Desviación.

TABLA 4.19.
CODIFICACIÓN DESVIACIÓN DE LA VARIABLE AREA CON REFERENCIA A LA PRIMERA CATEGORÍA: AREA URBANA

Área Geográfica	AREA(1)	AREA(2)
Área (0): Urbana	-1	-1
Área (1): Rural	1	0
Área (2): Indígena	0	1

En este caso la ecuación del modelo ajustado viene dada por:

$$\ln [p / 1\text{-}p] = -1.478 + 0.225 \text{ AREA } (1) - 1.469 \text{ AREA } (2)$$

La salida del SPSS:

TABLA 4.20.
RESULTADO DEL MODELO DE REGRESIÓN LOGÍSTICA
MODELO : In [p / 1-p] = -1.478 - 0.225 AREA (1) + 1.469 AREA (2)

Variables en la ecuación

		B	E T	Wald	gl	Sig	Exp(B)
Paso 1[a]	AREA			304 647	2	000	
	AREA(1)	- 225	082	7 559	1	006	799
	AREA(2)	1 469	084	303 051	1	000	4 345
	Constante	-1 478	064	535 961	1	000	228

a Vanable(s) introducida(s) en el paso 1 AREA

El logit para un niño residente en el área urbana, AREA (1) = AREA (2) = -1, será

$$\text{In }[p/\ 1\text{-}p]_{AREA(0)} = -1\ 478 - 0\ 225\ (-1) + 1\ 469(-1)$$

y para el área rural, AREA (1) = 1 y AREA (2) = 0 su logit será

$$\text{In }[p/\ 1\text{-}p]_{AREA(1)} = -1\ 478 - 0\ 225 + 1\ 469\ (0)$$

Luego restando las logits se obtiene

$$\hat{\beta}_0 + \hat{\beta}_1 - (\hat{\beta}_0 - \hat{\beta}_1 - \hat{\beta}_2) = 2\hat{\beta}_1 + \hat{\beta}_2$$

Entonces el riesgo de desnutrición de un niño o niña residente en el área rural respecto a la urbana será

$$R\hat{O} = e^{2\hat{\beta}_1+\hat{\beta}_2}$$
$$= e^{2(-0\ 225)+1\ 469}$$
$$= 2\ 770$$

Como pudimos constatar los resultados obtenidos son similares a los del primer esquema de codificación de variables indicadoras

Con respecto al último esquema de codificación los coeficientes de la categoría AREA (1) igual a -0 225, es el logit de padecer desnutrición en el área rural, respecto a la categoría promedio de las tres áreas

4.5. MODELO DE REGRESIÓN LOGÍSTICA MULTIVARIADO CON INTERACCION

4.5.1. Modelo 1: Un Ajuste Multivariado con Interacción

A continuación analizaremos si para los datos objetos de estudio existen una interacción entre las variables independientes X_1 = ALFABET que define el nivel de alfabetismo de la madre como. X_1= "ALFABETA" ó X_1= 1 "ANALFABETA" y la variable X_2 = "CANTPERS" que define el número de personas que habitan el hogar como X_2 = 1 "1 a 4 personas" ó X_2 = 1 "5 y más personas" y la variable independiente "Y" (ESTN) que define el estado nutricional del niño como Y = 1 "desnutrido" ó Y = 0 "no desnutrido" El análisis se realizará con datos procedentes de las áreas rural e indígena, se excluyó el área urbana por no ser significativa la variable de interés CANTPERS, lo que podría deberse a que el tamaño de la muestra disminuyó en 1,573 casos, debido a pérdida de datos al incluir en el análisis la variable ALFABET como resultado de la no respuesta, la misma fue dicotomizada.

Para estudiar la interacción se debe crear variables productos X_1 X_2 (Nivel de Analfabetismo y Miembros en el hogar) donde el modelo completo está dado por

$$\ln (p/q) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_1 X_2$$

Si en caso más simple que se refleja en la ecuación anterior, se verifica que el coeficiente de X_1 no es ahora constante, sino que depende de X_2 En este contexto

$$\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_1 X_2 = \hat{\beta}_0 + (\hat{\beta}_1 + \hat{\beta}_3 X_2) X_1 + \hat{\beta}_2 X_2 \qquad (95)$$

El grado en que influye el aumento de X_1 en una unidad (es decir, la razón ods asociada a X_1) es igual a

$$e^{(\hat{\beta}_1 + \hat{\beta}_3 X_2)} = e^{\hat{\beta}_1} e^{(\hat{\beta}_2 X_2)}$$

Para estudiar el efecto del analfabetismo de la madre (ALFABET) sobre el hecho de que el niño(a) este desnutrido, efectuaremos el siguiente análisis de regresión logística Sospechamos que la cantidad de personas en el hogar podría ser un potenciador del efecto negativo atribuido al analfabetismo de la madre que al carecer de conocimientos, desconoce como preparar una dieta balanceada sobre todo con los escasos recursos con que cuenta Situación que tal vez se hace más difícil de acuerdo a la cantidad de personas en el hogar Para comprobar este hallazgo definimos el modelo de regresión logística

$$\ln (p/q) = \hat{\beta}_0 + \hat{\beta}_1 \text{ ALFABET} + \hat{\beta}_2 \text{ CANTPERS} + \hat{\beta}_3 \text{ ALFABET x CANTPERS}$$

donde los denominados efectos principales son los términos correspondiente a CANTPERS y ALFABET como el término de interacción correspondiente al producto (ALFABET) x (CANTPER)

Para este esquema del modelo el coeficiente $\hat{\beta}_0$ no significa nada, $\hat{\beta}_1$ es el aumento del logaritmo del ods de madre analfabeta ($X_1 = 1$) en un hogar con al menos cuatro miembros ($X_2 = 0$), $\hat{\beta}_2$ es el aumento del logaritmo de la ods por provenir de un hogar con 5 o más miembros ($X_2 = 1$) con respecto a un hogar con al menos 4 miembros, sin ser la madre analfabeta ($X_1 = 0$) y $\hat{\beta}_3$ modeliza la posible interacción o sobre aumento por ambas característica

El modelo de interacción cuando los niños (as) proceden de un hogar donde el número de miembros es de cinco o más (CANTPERS = 1) toma la forma siguiente

$$\ln (p/1\text{-}p) = (\hat{\beta}_0 + \hat{\beta}_2) + (\hat{\beta}_1 + \hat{\beta}_3) \text{ ALFABET}$$

y para los hogares con menos de 5 miembros (CANTPERS = 0) el modelo es

$$\ln (p/1\text{-}p) = \hat{\beta}_0 + \hat{\beta}_1 \text{ ALFABET}$$

El modelo descrito contempla la posibilidad de una asociación diferente entre el nivel alfabetismo y desnutrición en cada una de las categorías de la variable CANTPERS pues:

$$R\hat{O}_{CANTPERS\ <5} = e^{\hat{\beta}_1} \quad (96) \qquad y \qquad R\hat{O}_{CANTPERS\ \geq 5} = e^{\hat{\beta}_1+\hat{\beta}_3} \quad (97)$$

donde $\hat{\beta}_1 + \hat{\beta}_3$ es el coeficiente del nivel de alfabetismo del modelo donde el hogar tiene cinco miembros o más. Correspondiéndole $\hat{\beta}_1$ al modelo donde el hogar tiene menos de cinco persona. A continuación se presenta el esquema del modelo logístico en interacción.

TABLA 4.21.

MODELO DE REGRESIÓN LOGÍSTICA MULTIVARIADO CON INTERACCIÓN

Nivel de alfabetismo	Cantpers	Modelo reemplazando los valores X_1 y X_2	Términos a interpretar	Interpretación de los coeficientes
ALFABET ($X_1 = 0$)	Variable independiente dicotomica 0: menos de 5 pers. 1: 5 y más pers.	$In(p/1-p) = \hat{\beta}_0 + \hat{\beta}_2\ X_2$	$\hat{\beta}_2\ X_2$	Es el aumento del logaritmo de la ods, que resulta de comparar $X_1 = 1$ (CANTPERS ≥ 5) respecto a $X_2 = 0$ (CANTPERS < 5) cuando la madre es ALFABETA ($X_1 = 0$)
ANALFABET ($X_1 = 1$)	$X_2 = 0$ (CANTPERS<5)	$In(p/1-p) = \hat{\beta}_0 + \hat{\beta}_1$	$\hat{\beta}_1$	Es el aumento del logaritmo del ods de una madre ANALFABETA ($X_1 = 1$) que vive en un hogar con al menos 4 personas ($X_2 = 0$)
	$X_1 = 1$ (CANTPERS ≥ 5)	$In(p/1-p) = \hat{\beta}_0 + \hat{\beta}_1\ X_3$	$\hat{\beta}_1 + \hat{\beta}_3$	Es el aumento del logaritmo de la ods de una madre analfabeta ($X_1 = 1$) que vive en un hogar con cinco o más miembros ($X_2 = 1$)

La salida del SPSS sobre el resumen del procedimiento se muestra a continuación:

Pruebas omnibus sobre los coeficientes del modelo

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	35.370	3	.000
	Bloque	35.370	3	.000
	Modelo	35.370	3	.000

Con la prueba de logaritmo del coeficiente de verosimilitud, el modelo completo es significativo ($p = 0.000$).

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	868 374	043	064

La bondad de ajuste del modelo final, el R cuadrado de Cox y Snell toma el valor de 0 043, normal que su valor sea pequeño, pues no cabe esperar que sea alto cuando los datos son individuales

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig
1	000	2	1 000

Con la prueba de Hosmer y Lemeshow si el ajuste es bueno, un valor alto de p predicha se asociará (con una frecuencia parecida a la p con el resultado 1 de la variable binomial)

TABLA 4.22.
RESULTADOS DEL MODELO DE REGRESIÓN LOGÍSTICA MULTIVARIDO CON INTERACCION:
MODELO: LN (ESTN) = -1.618 + 1.344 (ALFABET) + 0.717 CANTPERS -0.839 ALFABET*CANTPERS

Variables en la ecuación

		B	E T	Wald	gl	Sig	Exp(B)	C 95 0% para EXP(B	
								Inferior	Superior
Paso 1 [a]	ALFABET	1 344	333	16 302	1	000	3 834	1 997	7 361
	CANTPERS	717	189	14 394	1	000	2 049	1 414	2 967
	ALFABET by CANTPE	- 839	420	3 992	1	046	432	190	984
	Constante	-1 618	135	144 275	1	000	198		

a Variable(s) introducida(s) en el paso 1 ALFABET, CANTPERS ALFABET * CANTPERS

Con la prueba de Wald (W^2= 144 275 > $\chi_{0.95(1)}$ = 3 84) para el término de interacción se rechaza la hipótesis nula de no existencia de interacción, aunque la misma (p = 0 046) es muy cercana a 0 05

Los coeficientes de regresión pueden interpretarse de la siguiente manera

- El coeficiente de la variable nivel de alfabetismo (ALFABET) es positivo, lo que indica que el logaritmo del ods de que un niño (a) seleccionado al azar este desnutrido aumenta en 1 344, si este proviene de un hogar con madre analfabeta y con al menos 4 miembros

- El coeficiente $\hat{\beta}_2$ de la variable cantidad de personas en el hogar (CANTPERS) es positivo, lo que indica que el logaritmo de la ods0 de que un niño (a) seleccionado al azar este desnutrido aumenta en 0.717 si este proviene de un hogar con cinco o más miembros respecto a un hogar con al menos 4 miembros, sin ser la madre analfabeta
- $\hat{\beta}_3$ modeliza la posible interacción o sobre aumento por ambas característica Si $\hat{\beta}_3$ es mayor que cero el riesgo del analfabetismo en el deterioro nutricional será mayor en los hogares con 5 miembros o más respecto aquellos con menos de 5 miembros Como $\hat{\beta}_3$ es negativo, entonces el riesgo del analfabetismo de la madre, disminuye en 0 839 en los hogares con cinco o más miembros respecto a aquellas con al menos 4 miembros es decir el tamaño del hogar potencia el efecto del analfabetismo

En cuanto a los coeficientes del modelo de regresión logística éstos nos permiten calcular la razón ods para una madre analfabeta ($X_1 = 1$) y un grupo familiar de al menos 4 miembros ($X_2 = 0$) usando la expresión [96] es de

$$\begin{aligned} R\hat{O}_{CANTPERS} &= e^{\hat{\beta}_1} \\ &= e^{1\,344} \\ &= 3\,834 \end{aligned}$$

y su intervalo de confianza al 95% es

$$\begin{aligned} e^{\hat{\beta}_1 \pm 1\,96 \hat{E}E(\hat{\beta}_1)} &= e^{1\,344 \pm 1\,96(0\,333)} \\ &= [1\,996{,}7\,365] \end{aligned}$$

Hay que agregar que la salida del SPSS y los paquetes estadísticos en general calculan los intervalos de confianza asumiendo que no hay términos de interacción y por tanto, son sólo parcialmente valida cuando existe interacción

La razón ods de deterioro nutricional en la niñez menor de cinco años procedente de un hogar donde la madre es analfabeta y un grupo familiar de 5 o más personas ($X_2 = 1$) se obtiene de la ecuación [97] y es igual a

$$
\begin{aligned}
\hat{R}O_{CANTPERS \geq 5} &= e^{\hat{\beta}_1 + \hat{\beta}_3} \\
&= e^{1\ 344 - 0\ 839} \\
&= e^{0\ 505} \\
&= 1\ 657
\end{aligned}
$$

Para calcular su intervalo de confianza se necesita estimar la varianza de $\hat{\beta}_1 + \hat{\beta}_3$, pero como podemos observar la expresión [95] denota que un valor constante de la variable X_2 en nuestro caso CANTPERS depende de la suma de dos estimadores, es decir

$$(\hat{\beta}_1 + \hat{\beta}_3 \text{ CANTPERS})$$

Para estimar esta varianza utilizaremos la matriz de correlación que es la que produce la salida de SPSS

Matriz de correlaciones

		Constant	ALFABET	CANTPERS	ALFABET by CANTPERS
Paso 1	Constant	1 000	- 405	- 713	321
	ALFABET	- 405	1 000	289	- 793
	CANTPERS	- 713	289	1 000	- 450
	ALFABET by CANTPERS	321	- 793	- 450	1 000

$$
\begin{aligned}
\text{var}(\hat{\beta}_1 + \hat{\beta}_3) &= \text{var}(\hat{\beta}_1) + \text{var}(\hat{\beta}_3) + 2\ \text{cov}(\hat{\beta}_1, \hat{\beta}_3) \\
&= [\hat{E}E(\hat{\beta}_1)]^2 + [\hat{E}E(\hat{\beta}_3)]^2 + 2\ \hat{c}ov(\hat{\beta}_1, \hat{\beta}_3)\, x\, \hat{E}E(\hat{\beta}_1)\, x\, \hat{E}E(\hat{\beta}_3) \\
&= 0\ 333^2 + 0\ 420^2 - 2 \text{ x } 0\ 793 \text{ x } 0\ 333 \text{ x } 0\ 420 \\
&= 0\ 065471
\end{aligned}
$$

y por tanto, su intervalo de confianza al 95%

$$
\begin{aligned}
e^{(\hat{\beta}_1 + \hat{\beta}_3) \pm 1\ 96 \hat{E}E(\hat{\beta}_1 + \hat{\beta}_3)} &= e^{0\ 50 \pm 1\ 96 \sqrt{0\ 065471}} \\
&= [1.0034, 2.736]
\end{aligned}
$$

Obsérvese que los resultados muestran evidencia de interacción, ya que los resultados son distintos para el grupo familiar de más de 5 personas (RÔ = 1.657) respecto al grupo de al menos 4 miembros en la familia (RÔ = 3.834) que duplica el riesgo asociado al nivel de alfabetismo.

TABLA 4.23.
RAZÓN ODDS E INTERVALOS DE CONFIANZA

CANTPERS	$\hat{\beta}_1 + (\text{CANTPERS})\,\hat{\beta}_3$	$\hat{RO} = e^{\hat{\beta}_1 + (\text{CANTPERS})\hat{\beta}_3}$	LC 95.0% para Exp.	
			Inferior	Superior
0	1.344	3.834	1.997	7.361
1	0.505	1.657	1.0034	2.736

Si consideramos la relación desnutrición y alfabetismo sin controlar el hecho de que el tamaño del hogar sea de al menos 4 miembros o mayor de cinco miembros, entonces tendremos un modelo simple, tal estimación viene dada por:

TABLA 4.24.
Modelo: LN (ESTN) = -1.289 + 0.934 ALFABET

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	C. 95.0% para EXP(B	
								Inferior	Superior
Paso 1[a]	ALFABET	.934	.201	21.698	1	.000	2.546	1.718	3.772
	Constante	-1.289	.093	190.268	1	.000	.275		

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

Esta Tabla permite estimar la razón ods que según este modelo su valor viene dado por:

$$\hat{RO} = e^{0.934} = 2.546$$

Los resultados de las diferentes razón ods se pueden derivarse de las siguientes tablas de datos.

TABLA 4.25.
ESTADO NUTRICIONAL Y NIVEL DE ALFABETISMO

Estado Nutricional	Nivel de Alfabetismo	
	Analfabeta	Alfabeta
Desnutrido	54	146
No desnutrido	77	530
Total	**131**	**676**

$$R\hat{O} = {(axd)}\big/{(bxc)}$$
$$= {(54\ X\ 530)}\big/{(146\ X\ 77)}$$
$$= 2.546$$

Los estimadores correspondiente al riesgos asociados al nivel de alfabetismo de la madre en los hogares con CANTPERS < 5 y CANTPERS ≥ 5 lo obtendremos a partir de la siguiente tablas:

TABLA 4.26.
ESTADO NUTRICIONAL Y NIVEL DE ALFABETISMO EN HOGARES CON CANTPERS < 5

Estado nutricional	Nivel de alfabetismo	
	Analfabeta	Alfabeta
Desnutrido	19	66
No desnutrido	25	333
Total	44	339

$$R\hat{O} = {(19\ X\ 333)}\big/{(66\ X\ 25)}$$
$$= 3.834$$

TABLA 4.27.
ESTADO NUTRICIONAL Y NIVEL DE ALFABETISMO EN HOGARES CON CANTPERS ≥5

Estado nutricional	Nivel de alfabetismo	
	Analfabeta	Alfabeta
Desnutrido	35	80
No desnutrido	52	197
Total	87	277

$$R\hat{O} = {(35\ X\ 197)}\big/{(80\ X\ 52)}$$
$$= 1.657$$

La razón ods para un hogar con cinco o más miembros (CANTPERS=1) con madre analfabeta (ALFABET=1) respecto a un hogar con al menos cuatro miembros

(CANTPERS=0) sin ser la madre analfabeta (ALFABET=0) se puede obtener a partir de la Tabla siguiente:

TABLA 4.28.
ESTADO NUTRICIONAL, SEGÚN CATEGORIAS: ANALFABETISMO Y TAMAÑO DEL HOGAR

Estado nutricional	Nivel de alfabetismo	
	Analfabeta (CANTPERS<5)	Alfabeta (CANTPERS≥5)
Desnutrido	35	66
No desnutrido	52	333
Total	**87**	**339**

$$R\hat{O} = \frac{(35 \; X \; 333)}{(66 \; X \; 52)}$$
$$= 3.396$$

4.5.2. Modelo 2: Dos Ajuste Simple Independientes

El modelo 2 comprende un análisis de regresión logística simple. Para el cual separamos la muestra en dos grupos: los niños que viven en hogares con al menos cuatro miembros y los niños que viven en hogares con cinco o más personas y se aplicó el modelo de regresión logística en cada muestra por separado.

Las estimaciones de los parámetros para el modelo de regresión simple se muestran en las siguientes Tablas:

TABLA 4.29.
RESULTADO DEL MODELO DE REGRESIÓN LOGÍSTICA SIMPLE:
MODELO: LN (ESTN) = -1.618 + 1.344 ALFABET
CANTPERS < 5

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	C. 95.0% para EXP(B	
								Inferior	Superior
Paso 1[a]	ALFABET	1.344	.333	16.302	1	.000	3.834	1.997	7.361
	Constante	-1.618	.135	144.275	1	.000	.198		

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

TABLA 4.30.
RESULTADO DEL MODELO DE REGRESIÓN LOGÍSTICA SIMPLE:
MODELO: LN (ESTN.) = -0.901 + 0.505 ALFABET
CANTPERS ≥ 5

Variables en la ecuación

		B	E T	Wald	gl	Sig	Exp(B)	95 0% para EXP(	
								Inferior	Superior
Paso 1[a]	ALFABE	505	256	3 905	1	048	1 657	1 004	2 736
	Constant	- 901	133	46 206	1	000	406		

a Variable(s) introducida(s) en el paso 1 ALFABET

4.5.3. Modelo 3: Un solo Ajuste multivariado

El ajuste a un modelo multivariado es un enfoque dependiente y más apropiado si los tamaños de las submuestras son pequeños. Este un modelo en la cual se incluyen dos variables independientes, es decir, además de la variable ALFABET, también se incluyen la variable CANTPERS, que define a las subpoblaciones referentes al tamaño del hogar

TABLA 4.31.
RESULTADOS DEL MODELO DE REGRESIÓN LOGÍSTICA MULTIVARIADO
MODELO: LN (ESTN) = - 1.536 + 0.808 ALFABET + 0.551 CANTPERS

Variables en la ecuación

		B	E T	Wald	gl	Sig	Exp(B)	C 95 0% para EXP(B	
								Inferior	Superior
Paso 1[a]	ALFABET	808	205	15 528	1	000	2 244	1 501	3 354
	CANTPERS	551	169	10 603	1	001	1 735	1 245	2 418
	Constante	-1 536	125	151 232	1	000	215		

a Variable(s) introducida(s) en el paso 1 ALFABET, CANTPERS

Para este caso, la ecuación de regresión esta dada por

$$P(Y=1)=\frac{1}{1+e^{-(\beta_0+\beta_1(ALFABET)+\beta_2(CANTPERS))}} \qquad (98)$$

donde CANTPERS = 0, si el niño o niña procede de un hogar con al menos 4 miembros y 1 si el tamaño del hogar es mayor o igual a 5

$$P(Y=1)_{CANTPERS\geq 5}=\frac{1}{1+e^{-(\beta^{*}_0+\beta_1(ALFABET))}} \qquad (99)$$

Sustituyendo CANTPERS = 1 en la ecuación [98] se obtendría·

donde $\hat{\beta}^{*}_0 = \hat{\beta}_0 + \hat{\beta}_1$

Para CANTPERS = 0, se tendrá·

$$P(Y=1)_{CANTPERS\ <5} = \frac{1}{1+e^{-(\beta_0+\beta_1(ALFABET))}} \tag{100}$$

Las ecuaciones [99] y [100] reflejan los patrones de la desnutrición para un hogar con CANTPERS ≥ 5 y CANTPERS <5 respectivamente

El valor de la pendiente es constante para dos modelos descrito . Utilizando los datos de la Tabla 4.31. que resume el modelo [98] obtendremos las siguientes estimaciones de los parámetros.

$$\hat{\beta}_0 = -1\ 536;\ \hat{\beta}_1 = 0\ 808,\ \ \hat{\beta}_2 = 0\ 551$$

Reemplazando estos valores en las fórmulas [99] y [100] asumen las formas siguientes, tomando en consideración que $\beta^{*}{}_0 = -0\ 985$

$$P(Y=1)_{CANTPERS\geq 5} = \frac{1}{1+e^{-(-0\ 985+0\ 808(ALFABET))}}$$

$$P(Y=1)_{CANTPERS<5} = \frac{1}{1+e^{-(-1\ 536+0\ 808(ALFABET))}}$$

4.5.4. Consideraciones Finales

En el modelo 3 se asume que no hay interacción, por lo cual el modelo 1 plantea un enfoque más adecuado. Más adelante podremos demostrar que este último modelo con 616 individuos objetos de análisis arrojó el mismo resultado que el modelo 2, cuando se utilizó submuestras separadas de manera independiente con 277 y 339 individuos de área indígenas y rurales respectivamente. Situación que no ocurre siempre, sin embargo en este caso la muestra es grande y probablemente sean equivalentes en lo esencial el modelo 1 y el modelo 2 A continuación veremos la similitud entre el modelo 1 y 2.

El modelo multivariado con interacción que se ajustado que se presentó en el enfoque 1 es el siguiente:

$$P(Y=1)=\frac{1}{1+e^{-(\hat{\beta}_0+\hat{\beta}_1(ALFABET)+\hat{\beta}_2(CANTPERS)-\hat{\beta}_3(ALFABET)(CANTPERS))}} \qquad (101)$$

Si CANTPERS = 0, entonces el modelo descrito en [101] se reduce a

$$P(Y=1)_{CANTPERS<5}=\frac{1}{1+e^{-(\hat{\beta}_0+\hat{\beta}_1(ALFABET))}}$$

El cual es similar al obtenido en el modelo 2, Tabla 4 30 donde

$$P(Y=1)_{CANTPERS<5}=\frac{1}{1+e^{-(1\ 618+1\ 344(ALFABET))}}$$

Si CANTPERS = 1, entonces el argumento de la función se expresa como

$$\hat{\beta}_0+\hat{\beta}_1(ALFABET)+\hat{\beta}_2+\hat{\beta}_3(ALFABET)=(\hat{\beta}_0+\hat{\beta}_2)+(\hat{\beta}_1+\hat{\beta}_3)(ALFABET)$$

con $\hat{\beta}_0^{*}=\hat{\beta}_0+\hat{\beta}_2$ *y* $\hat{\beta}^{*}=\hat{\beta}_1+\hat{\beta}_3$

$$P(Y=1)_{CANTEPERS\ \geq 5}=\frac{1}{1+e^{-(\hat{\beta}_0^{*}+\hat{\beta}_1^{*}(ALFABET))}}$$

Cuando no hay interacción $(\hat{\beta}_3=0)$ la pendiente coinciden pero no así cuando existe dicha interacción El ajuste del modelo de regresión logística multivariado con interacción dio lugar a las estimaciones siguientes

$$\hat{\beta}o=-1\ 618,\quad \hat{\beta}_1=1\ 344$$

$$\hat{\beta}_2=0\ 717,\quad \hat{\beta}_3=-0\ 839 \qquad (102)$$

de donde se deduce que

$$\hat{\beta}^{*}=0\ 505 \qquad y \qquad \hat{\beta}_0^{*}=-0901 \qquad (103)$$

Como se puede observar los resultados [102] y [103] son iguales a los obtenidos mediante el modelo 2, Tabla 4 29 y Tabla 4.30 respectivamente

4.6. MODELO DE REGRESIÓN LOGÍSTICA MULTIVARIADO

Para determinar el efecto de algunos factores socio-económicos y demográficos en la desnutrición infantil se plantea un análisis de regresión logística multivariado que pueda explicar correctamente este fenómeno.

Se trabajo con información procedente de la Encuesta de Niveles de Vida de 1997, desarrollada por el entonces Ministerio de Planificación y Política Económica en conjunto con el Banco Interamericano de Desarrollo De esta fueron seleccionados los datos referente a la población infantil menor de cinco años que ascendieron a un total de 2,294 casos

En el primer capítulo, hacíamos referencia de las hipótesis alternativas o de investigación, en esta sección procederemos a comprobar las mismas a través del análisis de regresión logística multivariado Para ello, las variables incluidas en las hipótesis también formaron parte del análisis de regresión, con excepción de la variable AFILIACIÓN la cual no fue significativa para explicar el fenómeno de la desnutrición por lo cual fue excluida del estudio

4.6.1. Variable de Investigación

Se seleccionaron aquellas variables que según investigaciones realizadas están asociadas a la desnutrición infantil. La variable Estado Nutricional denota la variable dependiente, para la cual se tuvo que recodificar sus valores anteriores donde 0 = No y 1 = Si, los que nos indican la ausencia o presencia de la característica en estudio; es decir, desnutrición infantil, según el indicador talla/edad, está por debajo de 2 D.E de la mediana de talla de niños (as) de la misma edad y sexo en la población de referencia (NCHS)

TABLA 4.32.
VARIABLES DE INVESTIGACIÓN

Variable Dependiente	Variables Predictoras
Y = Estado Nutricional	Sexo del Niño (a): SEXO
Se definen como:	Mujer : 1 Hombres: 0
Y = 0" No Desnutrido" Y = 1"Desnutrido"	Edad del Niño (a): EDNIÑO
	Menor de 1 año: EDNIÑO (0) 1 a 2 años: EDNIÑO (1) 3 a 4 años: EDNIÑO (2)
	Categoría de Referencia: Menor de 1 año
	Área Geográfica: AREA Urbana : Area (0) Rural : Area (1) Indígena : Area (2)
	Categoría de Referencia: Área Urbana
	Afiliación:
	Sin seguridad social: 1 Con seguridad social: 0
	Gastos Médicos anuales en B/.: GASTMED
	Ninguno : Gastmed (1) Menos de B/.20 : Gastmed (2) Más de B/.20 : Gastmed (0)
	Categoría de Referencia: Gastmed (0)
	Tamaño del Hogar : MIEMBROS
	Variable continua, con un valor mínimo de 2 y máximo de 14 personas en el hogar.

4.6.2 Selección del Modelo y Significancia de las Variables

El método que utilizaremos es el de incorporación progresiva de variables (forward). Este método consiste en ir añadiendo los posibles predictores de uno a uno, manteniendo en el modelo las que son estadísticamente significativos a un determinado nivel de significación "α" y descartando los que no lo son. El criterio de significación estadística que utilizaremos es: $\alpha = 0.05$.

La variable AFILIACIÓN no se incluyó en el proceso, pues se detuvo debido a que no aporta una mejora estadísticamente significativa a la bondad de ajuste del modelo.

La Tabla nos muestra los resultados de aplicar una estrategia de modalización estadística hacia adelante en los datos.

PASO 1: Inclusión de la Variable Indicadora AREA:

MODELO: $\ln(\text{ESTN}) = \hat{\beta}_0 + \hat{\beta}_1\,\text{AREA}(1) + \hat{\beta}_2\,\text{AREA}(2)$

Número de observaciones	2,294	
Chi-cuadrado (2)	338.047	Prob>|chi-2|=0.000
Log Verosimilitud de la constante	-1,096.512	Pseudo -R2 = 0.139
Log Verosimilitud el del modelo con la variable	-927.489	R cuadrado de Nagelkerke = 0.223

El modelo después paso 1 es:

$$\ln(\text{ESTN}) = -2.722 + 1.019\,\text{AREA}(1) + 2.713\,\text{AREA}(2)$$

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	. 95.0% para EXP	
								Inferior	Superior
Paso 1[a]	AREA			304.647	2	.000			
	AREA(1)	1.019	.166	37.709	1	.000	2.771	2.002	3.837
	AREA(2)	2.713	.170	255.071	1	.000	15.077	10.807	21.034
	Constante	-2.722	.140	375.670	1	.000	.066		

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

Este modelo explica el 13 0% de la varianza de la variable ESTN, según el indicador Pseudo -R^2 y el 22 3% según el indicador R- Cuadrado de Nagelkerke

Si el análisis se realiza con la prueba de razón de verosimilitud la hipótesis nula y alternativa para determinar si el coeficiente de regresión significativamente diferente de 0 son

$$Ho \quad \beta_1 = \beta_2 = 0$$

$$H_1 \quad \text{al menos una } \beta_j \neq 0$$

Estadístico de Prueba

$$G = -2\ln \left[\frac{(\text{Verosimilitud del modelo con la constante } (m_0))}{(\text{Verosimilitud del modelo seleccionado } (m))} \right]$$

donde el valor G se distribuye χ^2 con g l = número de parámetros p del modelo En este caso

$$G = -2\ln(Lm_0/Lm) = -2[\ln lm_0 - \ln lm]$$
$$= -2[-1096\ 512 - (-927\ 489)]$$
$$= 2{,}193\ 024 - 1{,}854\ 978$$
$$= 338\ 047$$

donde Lm_0 es el log verosimilitud del modelo cuando solo incluye la constante y lm es la log verosimilitud del modelo con las dos variables indicadoras que definen el AREA (Ver cálculo en pág 121)

Regla de Desición

Los resultados indican que las variables indicadoras del "Área" (1) y Área (2)" son estadísticamente significativos al nivel seleccionado (α = 0 05) La prueba de la razón de

PASO 2. Inclusión de la Variable Indicadora EDNIÑO:

MODELO: $\ln(ESTN) = \hat{\beta}_0 + \hat{\beta}_1\, AREA(1) + \hat{\beta}_2\, AREA(2) + \hat{\beta}_3\, EDNIÑO(1) + \hat{\beta}_4\, EDNIÑO(2)$

Número de observaciones	2, 294	Pseudo - R^2 = 0.171
Chi-cuadrado (4)	431.188	Prob. > |chi - 2| = 0.000
Log Verosimilitud	-880.918	R cuadrado de Nagelkerke = 0.278

El modelo después del paso 2 es:

$$\ln(ESTN) = -4.074 + 1.019\, AREA(1) + 2.848\, AREA(2) + 1.321\, EDNIÑO(1) + 1.777\, EDNIÑO(2)$$

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95.0% para EXP(B) Inferior	Superior
Paso 2[b]	AREA			311.637	2	.000			
	AREA(1)	1.009	.168	36.236	1	.000	2.742	1.975	3.809
	AREA(2)	2.848	.175	264.019	1	.000	17.248	12.234	24.317
	EDNIÑO			74.169	2	.000			
	EDNIÑO(1)	1.321	.209	39.933	1	.000	3.747	2.487	5.644
	EDNIÑO(2)	1.777	.208	73.160	1	.000	5.910	3.933	8.879
	Constante	-4.074	.233	305.226	1	.000	.017		

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

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

Este modelo explica el 17.1% o el 27.8% de la varianza de la variable ESTN, según las medidas Pseudo - R^2 y R- cuadrado de NAGELKERKE respectivamente.

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	1854.977	.137	.223
2	1761.836	.171	.278

La inclusión definitiva de las variables indicadoras que definen la EDAD del niño en el modelo viene determinada por los resultados de la prueba de la razón de verosimilitud (G) que mide la mejora en la bondad de ajuste del modelo con la introducción de la variable EDAD. La hipótesis nula y alternativa para determinar si los coeficientes de regresión de la variable indicadora EDAD son significativamente distintos de 0 son

$$Ho \quad \beta_3 = \beta_4 = 0$$

$$H_1 \quad \text{al menos una } \beta_i \neq 0$$

La estadística G se estima mediante la fórmula

G = D [Modelo sin la variable Adicional (m-h)] - D[modelo con la variable (m)]

Equivalentemente,

$$G = -2 \ln [L_{m-2} - lm] = 2 (\ln L_{m-2} - \ln lm)$$

donde $\ln L_{m-2}$ es la log verosimilitud del modelo descrito en el paso 1 y ln Lm es la log verosimilitud del modelo que incluye la variable independiente AREA y la variable EDAD que se resume en paso 2

$$G = - 2 (\ln L_{m-2} - \ln L_m)$$

$$= - 2 \{ (-927\ 489) - (-880\ 918)\}$$

$$= -1854\ 978 \ - \ 1761\ 836 \ = \ 93\ 141$$

<u>Regla de Decisión</u>

Respecto a la prueba de razón de verosimilitud podemos rechazar la hipótesis nula ($Ho \cdot \beta_3 = \beta_4 = 0$) dado que $G = 93\ 141 > \chi^2_{0\,95}(2) = 5\ 991$, por lo que las variables indicadoras EDADNIÑO deben permanecer en el modelo

Pruebas omnibus sobre los coeficientes del modelo

		Chi-cuadrado	gl	Sig
Paso 2	Paso	93 141	2	000
	Bloque	431 188	4	000
	Modelo	431 188	4	000

El valor chi-cuadrado (4 g l) es igual a

$$\text{Chi-cuadrado } (4) = 338\ 047 + 93\ 141$$

$$= 431\ 188$$

Regla Decisión:

Como el valor $\chi^2_{0.95}(4) = 9.488 < 431.188$ se rechaza Ho, por consiguiente al menos uno de los coeficientes estimados es distinto de 0 en la población. Por lo que en la población factores tales como AREA y EDAD están asociados con el hecho de estar desnutrido.

Por otra parte, el estadístico de Wald permite comprobar individualmente que todos los coeficientes son diferentes de 0.

Paso 3. Inclusión de la Variable Predictora MIEMBROS:

Siguiendo con el procedimiento de inclusión de variables en el modelo se tiene que la variable predictora a incluir es MIEMBROS que define el tamaño del hogar.

MODELO: $\ln(ESTN) = \hat{\beta}_0 + \hat{\beta}_1 \text{ AREA(1)} + \hat{\beta}_2 \text{ AREA (2)} + \hat{\beta}_3 \text{ EDNIÑO (1)} + \hat{\beta}_4 \text{ EDNIÑO (2)} + \hat{\beta}_5 \text{MIEMBROS}$

Número de observaciones	2,294	
Chi-cuadrado (5)	468.397	Prob > |chi - 2| =0.000
Log Verosimilitud	-862.314	Pseudo - R^2 = 0.185
		R Cuadrado de Nagelkerke = 0.300

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95.0% para EXP(B)	
								Inferior	Superior
Paso 3[c]	AREA			219.442	2	.000			
	AREA(1)	.940	.169	30.963	1	.000	2.560	1.838	3.565
	AREA(2)	2.530	.182	193.583	1	.000	12.551	8.788	17.925
	MIEMBROS	.112	.018	37.702	1	.000	1.119	1.079	1.159
	EDNIÑO			77.070	2	.000			
	EDNIÑO(1)	1.352	.211	40.858	1	.000	3.863	2.553	5.847
	EDNIÑO(2)	1.832	.210	75.896	1	.000	6.248	4.137	9.435
	Constante	-4.809	.270	316.180	1	.000	.008		

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

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

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

Este modelo explica el 18.5% o el 30% de la varianza de la variable ESTN, según las medidas Pseudo – R^2 y R-Cuadrado NAGELKERKE respectivamente.

El modelo después del Paso 3 es

$$\ln(ESTN) = -4\ 809 + 0\ 940\ AREA(1) + 2\ 530\ AREA(2) + 1\ 352\ EDNIÑO(1) + 1\ 832\ EDNIÑO(2) + 0\ 112\ MIEMBROS$$

La hipótesis nula y la alternativa para determinar si el coeficiente de regresión del tamaño del hogar (MIEMBROS) es significativamente diferente de 0 son

$$H_o \quad \beta_5 = 0$$

$$H_1 \quad \beta_5 \neq 0$$

Estadístico de Prueba

El valor "p" de la razón "W" indica que la inclusión de la variable cuantitativa MIEMBROS mejora la bondad de ajuste del modelo La prueba de razón de verosimilitud corrobora este resultado

$$G = -2\ [(\ln L_{m-4} - Lm)$$

$$= -2[-880\ 918 - (-862.314)]$$

$$= 1{,}761\ 836 - 1{,}724\ 628 = 37\ 208$$

Regla de Decisión

Puesto que $G = 37\ 208 > \chi^2_{0\ 95}(1) = 3\ 81$, se rechaza Ho $\beta_5 = 0$ por lo que la variable predictora MIEMBROS mejora la bondad de ajuste del modelo y por lo tanto debe permanecer en el modelo

Paso 4: Inclusión de la Variable Predictora GASTMED:

La próxima variable a incluir en el modelo es GASTMED (Gasto médico anuales) variable indicadora cuya categoría de referencia es "De 20 y más balboas anuales en gasto de salud"

MODELO: $\ln(\text{ESTN}) = \hat{\beta}_0 + \hat{\beta}_1 \text{ AREA(1)} + \hat{\beta}_2 \text{ AREA (2)} + \hat{\beta}_3 \text{ EDNIÑO (1)} + \hat{\beta}_4 \text{ EDNIÑO (2)} + \hat{\beta}_5 \text{ MIEMBROS} + \hat{\beta}_6 \text{ GASTMED (1)} + \hat{\beta}_7 \text{ GASTMED (2)}$

Número de observaciones	2,294	
Chi-cuadrado (7)	484.176	Prob > |chi - 2| =0.000
Log Verosimilitud	-854.424	Pseudo - R^2 = 0.190 R Cuadrado de Nagelkerke = 0.312

Este modelo explica el 19% o el 31.2% de la varianza de la variable ESTN, según las medidas Pseudo - R^2 y R cuadrado de NAGELKERKE respectivamente.

El modelo después del paso 4 es:

$$\ln(\text{ESTN}) = -5.921 + 0.884 \text{ AREA (1)} + 2.432 \text{ AREA (2)} + 1.354 \text{ EDNIÑO (1)} + 1.832 \text{ EDNIÑO (2)} + 0.112 \text{ MIEMBROS} + 1.259 \text{ GASTMED (1)} + 1.334 \text{ GASTMED (2)}$$

La hipótesis nula y alternativa para determinar si el coeficiente de regresión de la variable predictora GASTMED es significativamente diferente de cero son:

$$H_o: \beta_6 = \beta_7 = 0$$

$$H_1: \beta_6 = \beta_7 \neq 0$$

<u>Estadístico de Prueba</u>:

$$G = -2\,[(\ln l_{m-5} - \ln lm)$$

$$= -2\,[-862.314 - (-854.424)]$$

$$= 1{,}724.628 - 1{,}708.848 = 15.78$$

<u>Regla de Decisión:</u>

Como $G = 15.78 > \chi^2_{0.95}(2) = 5.991$, se rechaza la hipótesis nula, lo que comprueba la significancia estadística del aporte de la variable indicadora GASTMED en el modelo. Con la estadística de WALD, también se puede corroborar de manera individual el resultado de la estadística G.

Pruebas omnibus sobre los coeficientes del modelo

		Chi-cuadrado	gl	Sig.
Paso 4	Paso	15.780	2	.000
	Bloque	484.176	7	.000
	Modelo	484.176	7	.000

La salida del SPSS del modelo se resume en:

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95.0% para EXP(B)	
								Inferior	Superior
Paso 4[d]	AREA			196.097	2	.000			
	AREA(1)	.884	.170	27.098	1	.000	2.421	1.735	3.377
	AREA(2)	2.432	.184	173.755	1	.000	11.376	7.925	16.331
	MIEMBROS	.104	.018	32.186	1	.000	1.110	1.070	1.150
	EDNIÑO			76.953	2	.000			
	EDNIÑO(1)	1.354	.211	41.126	1	.000	3.873	2.561	5.859
	EDNIÑO(2)	1.829	.210	75.908	1	.000	6.231	4.128	9.403
	GASTMED			11.122	2	.004			
	GASTMED(1)	1.259	.402	9.825	1	.002	3.522	1.603	7.740
	GASTMED(2)	1.334	.400	11.112	1	.001	3.798	1.733	8.322
	Constante	-5.921	.455	169.291	1	.000	.003		

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

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

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

d. Variable(s) introducida(s) en el paso 4: GASTMED.

Paso 5. Inclusión de la Variable Predictora SEXO

MODELO: $\ln(\text{ESTN}) = \hat{\beta}_0 + \hat{\beta}_1\,\text{AREA(1)} + \hat{\beta}_2\,\text{AREA (2)} + \hat{\beta}_3\,\text{EDNIÑO (1)} + \hat{\beta}_4\,\text{EDNIÑO (2)} + \hat{\beta}_5\,\text{MIEMBROS} + \hat{\beta}_6\,\text{GASTMED (1)} + \hat{\beta}_7\,\text{GASTMED (2)} + \hat{\beta}_8\,\text{SEXO}$

Número de observaciones	2,294	
Chi-cuadrado (7)	488.684	Prob > $\lvert$chi - 2$\rvert$=0.000
Log Verosimilitud	-851.9865	Pseudor R^2 = 0.192 R cuadrado de NAGALKERKE = 0.312

El modelo después del paso 5 es:

In (ESTN) = -5.786 + 0.868 AREA (1) + 2.432 AREA (2) + 1.346 EDNIÑO (1) +

1.831 EDNIÑO (2) + 0.104 MIEMBROS + GASTMED (1) +

GASTMED (2) - 0.262 SEXO

TABLA 4.33.

RESULTADOS DEL MODELO DE REGRESIÓN LOGÍSTICA MULTIVARIADO

Variables en la ecuación

		B	E T	Wald	gl	Sig	Exp(B)	I C 95 0% para EXP(B)	
								Inferior	Superior
Paso 5[e]	AREA			195 042	2	000			
	AREA(1)	868	170	26 039	1	000	2 382	1 707	3 325
	AREA(2)	2 423	185	172 057	1	000	11 276	7 851	16 194
	SEXO	- 262	124	4 486	1	034	769	603	981
	MIEMBROS	104	018	32 243	1	000	1 110	1 071	1 151
	EDNIÑO			77 125	2	000			
	EDNIÑO(1)	1 346	211	40 619	1	000	3 843	2 540	5 815
	EDNIÑO(2)	1 831	210	75 960	1	000	6 238	4 133	9 415
	GASTMED			11 208	2	004			
	GASTMED(1)	1 251	402	9 692	1	002	3 495	1 590	7 685
	GASTMED(2)	1 339	400	11 173	1	001	3 814	1 740	8 360
	Constante	-5 786	458	159 389	1	000	003		

a Variable(s) introducida(s) en el paso 1 AREA

b Variable(s) introducida(s) en el paso 2 EDNIÑO

c Variable(s) introducida(s) en el paso 3 MIEMBROS

d Variable(s) introducida(s) en el paso 4 GASTMED

e Variable(s) introducida(s) en el paso 5 SEXO

El 19 2% o el 31 2% de la variabilidad total de la desnutrición es explicada por este modelo por las variables incluidas en el mismo, según la medida Pseudo-R^2 y R -Cuadrado de NAGELKERKE respectivamente

La hipótesis nula y alternativa que se plantea para determinar si el coeficiente de regresión de la variable SEXO es significativamente diferente de 0 son

$$Ho \quad \beta_8 = 0 \quad vr \quad H_1 \quad \beta_8 \neq 0$$

<u>Estadístico de Prueba</u>

$$G = -2\,[(\ln L_{m-7} - \ln Lm)$$

$$= -2\,[\,-854\ 424 - (-\,852\ 1705)]$$

$$= 1{,}708\ 848 - 1{,}704\ 341 = 4\ 507$$

<u>Regla de Decisión</u>

Como G = 4 507 > $\chi^2_{0\,95}$ (1) = 3 84, se rechaza la hipótesis nula de que el coeficiente de la variable sexo es cero, además se verifica la significancia estadística de esta variable por ende debe permanecer en el modelo

También se puede analizar la Prueba de Ómnibus donde el valor chi-cuadrado con 8 grados de libertad que es igual a

Chi-cuadrado (8) = 338 047 + 93 141 + 37 208 + 15 780 + 4 507

= 488 684

Las hipótesis que se plantean para contrastar son

Ho $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$

H_1 al menos una $\beta_j \neq 0$

<u>Regla de Decisión</u>

Como el valor $\chi^2_{0\,95}$ (8) = 15 507 < 488 684 se rechaza Ho, por consiguiente al menos uno de los coeficientes estimados es distinto de 0 en la población

La estadística de WALD que a diferencia de la chi-cuadrado verifica individualmente la significancia de los coeficientes, también permite rechazar Ho Por lo tanto en la población factores tales como el AREA, SEXO, EDAD, TAMAÑO DEL HOGAR y GASTMED están asociadas con el hecho de estar desnutrido

Pruebas omnibus sobre los coeficientes del modelo

		Chi-cuadrado	gl	Sig
Paso 5	Paso	4 507	1	034
	Bloque	488 684	8	000
	Modelo	488 684	8	000

PASO 6. Inclusión de la Variable Predictora AFILIACIÓN:

El computador no ejecuta este paso por no ser significante la variable predictora

4.6.4. Bondad de Ajuste del Modelo

La Prueba de Hosmer y Lemeshow ordena y dıvıde los datos en decıles y luego compara el número de desnutrıdo observado en cada decıl con el esperado Un ȷı-cuadrado no sıgnıficatıva para esta prueba ındıca que el modelo se aproxıma al saturado Así la prueba resulta no sıgnıficativa lo que ındica que el modelo se aproxıma al saturado, lo que sugıere que el ajuste es bueno.

Prueba de Hosmer y Lemeshow

Paso	Chı-cuadrado	gl	Sıg
5	4 881	8	770

Las medıdas de bondad de ajuste Pseudo – R cuadrado y Cuadrado de Nagelkerke ındican que el modelo explıca el 19 2% y el 31 2% de la varıabılıdad total de la desnutrıcıón por las varıables ıncluıdas en el mısmo Esta relacıón revela la existencıa de una relacıón moderada a alta

Resumen de los modelos

Paso	-2 log de la verosımılıtud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
5	1704 341	192	312

4.6.5. Medida de Eficiencia Predictiva: Curva ROC

La eficacıa predıctıva del modelo y el cálculo del punto o probabılıdad de corte más adecuado se puede realızar a partır de las denomınada Curva ROC

El área debajo de la curva es una medida conjunta de la eficacia predictiva del modelo El valor de está área en el modelo seleccionado es de 0 817, lo que indica una capacidad o eficacia predictiva moderada a alta. Un modelo que carece de poder predictivo presenta un área baja la curva de 0 50.

TABLA 4.33.

Área bajo la curva

Variables resultado de contraste Predicted probability

			Intervalo de confianza asintótico al 95%	
Área	Error típ [a]	Sig asintótica [b]	Límite inferior	Límite superior
817	012	000	794	.839

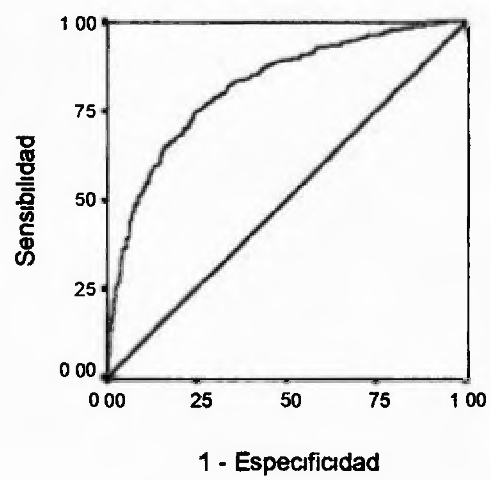
La variable (o variables) de resultado de contraste Predicted probability tiene al menos un empate entre el grupo de estado real positivo y el grupo de estado real negativo Los estadísticos pueden estar sesgados

a Bajo el supuesto no paramétrico

b Hipótesis nula área verdadera = 0,5

GRAFICA 4.6

Curva ROC



Los segmentos diagonales son producidos por los empates

4.6.6. Interpretación de los Resultados del Modelo de Regresión Logística

Luego de verificar la idoneidad del modelo se pueden extraer algunas conclusiones sobre los resultados del mismo, que nos permitirán comprobar las hipótesis planteadas inicialmente en nuestro estudio:

__ **Factores de Riesgo**

Del modelo logístico generado podemos observar que el signo de los coeficientes de algunas variables es positivo, eso significa que la variable en cuestión aumenta la probabilidad del suceso en estudio lo que es lo mismo aumenta la probabilidad de desnutrición infantil. En efecto se tiene; el resumen siguiente:

Variable	β	Exp (β) = RÔ
AREA (1)	+ 0.868	2.382
AREA (2)	+2.423	11.276
EDNIÑO (1)	+1.346	3.843
EDNIÑO (2)	+1.831	6.238
GASTMED (1)	+1.251	3.495
GASTMED (2)	+1.339	3.814
MIEMBROS	+0.104	1.110

__ **Interpretación de los Coeficientes**

- En primer lugar, el ÁREA geográfica influye sobre la probabilidad de sufrir desnutrición, de manera que los niños que residen en el área rural tienen una RÔ=2.382 significativamente distinta de cero. La influencia del área rural sobre el hecho de estar desnutrido puede ser mayor pues se observa que el intervalo de confianza para la razón ods en el extremo superior es de 3 325 Después de este análisis multivariante ya se puede sacar una primera conclusión señalando que los niños y niñas residentes en el área rural están a 2 382 veces más riesgo de sufrir desnutrición que aquellos que residen el área urbana

La influencia del área indígena sobre el hecho de estar desnutrido es mayor pues se observa una RÔ =11 276 que es significativamente distinta de cero La influencia del área indígena puede ser mayor si se observa que el intervalo superior es de 16 194 Podemos entonces concluir que los niños residentes en el área indígena están a 11 276 veces más riesgo de sufrir desnutrición que aquellos residentes en el área urbana

- Con respecto a la variable SEXO, esta influye también en que un niño tenga una mayor probabilidad de estar desnutrido Su efecto es negativo, lo cual es indicativo de que el riesgo de desnutrición disminuye en el sexo femenino (RÔ = 0 769) respecto al masculino (RÔ =1 00) Es decir, los niños tienen un 23 1% veces más riesgo de sufrir desnutrición que las niñas

- El tamaño del hogar (MIEMBROS) donde vive el niño consiste en que la desnutrición aumenta en un 11% por cada un miembro que vive en el hogar

- La edad influye sobre la probabilidad de estar desnutrido, de manera que los niños y niñas entre 1 a 2 años tienen un riesgo de sufrir desnutrición (RÔ = 3 843) casi cuatro veces mayor que un niño menor de un año En los niños entre 3 a 4 años el riesgo (RÔ = 6 238) es seis veces mayor respecto a los menores de un año

- Con respecto a los gastos anuales (GASTMED) invertidos en salud, el grupo que no invierte en salud y los que invierten al menos 20 balboas anuales, están en claros riesgo

(RÔ = 3 495 y RÔ = 3 814) con tres y cuatro veces más posibilidades de sufrir desnutrición respectivamente, en relación con aquellos hogares que invierten más de 20 00 balboas anuales

— **<u>Predicción del Riesgo</u>**

El valor predictivo de cada variable independiente o bien del modelo en su conjunto Siguiendo los ejemplos de las secciones anteriores, lo abordaremos ahora para obtener el valor predictivo del riesgo asociado a padecer desnutrición infantil Para ello partiremos de la ecuación siguiente

In (ESTN) = -5 786 + 0 868 AREA (1) + 2 423 AREA (2) + 1 346 EDNIÑO (1) +

1 831 EDNIÑO (2) + 0 104 MIEMBROS + 1 251 GASTMED (1) +

1 339 GASTMED (2) - 0 262 SEXO

Para un individuo procedente del ÁREA indígena cuya edad oscile entre 3 y 4 años, del sexo femenino (S_1 = 1) un tamaño de hogar de 5 miembros y un gasto anual de salud de cero La ecuación predice un riesgo alto del 49% de padecer desnutrición Así tendremos que

In (ESTN) = -5 786 + 0 868 (0) + 2 423 (1) + 1 346 (0) + 1 831 (1) + 0 104 (5)

+ 1 251 (1) + 1 339 (0) - 0 262 (1) = -0 023

Luego ln (ESTN) = ln (ods)

= -0 023

Por lo tanto

$$\text{Ods} = e^{(-0\ 023)} = 0\ 9773$$

La probabilidad estará dada por

$$p = \frac{ods}{1+ods} = \frac{0.9773}{1+0.9773} = 0.4943$$

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIÓN

Este estudio tiene como finalidad determinar el efecto de factores socio -económicos y demográficos en el estado nutricional de los niños menores de 5 años de Panamá El estudio se baso en los resultados obtenidos de la Encuesta de Niveles de Vida realizada en 1997 donde se trataron un total de 2,294 casos El método estadístico central utilizado en este estudio consistió en un análisis multivariado, mediante la técnica de regresión logística

Las variables utilizadas para medir tales efectos fueron EDAD del niño, SEXO, número de MIEMBROS en el hogar, AREA geográfica de residencia y GASTOS médicos anuales El estado nutricional del niño esta medido en base al indicador antropométrico Talla/Edad, que define el deterioro del estado nutricional como crónico y se basa en el retardo en su crecimiento.

A continuación se resumen los hallazgos más importantes de este estudio

1 Las medidas antropométricas que evalúan el estado nutricional del niño o la niña, definidas por peso/edad, peso/talla y talla/edad, muestran que aproximadamente el 20.2% de éstos presentaban algún forma de desnutrición

2 El indicador peso para la talla que estima la desnutrición aguda o actual, revela que el 1 1% de los niños y niñas analizados padecían este tipo de desnutrición. Su mayor incidencia se registraba en los niños con 1 a 3 años de edad La proporción era ligeramente superior en el sexo masculino (0 7%) respecto al femenino con 0 4% En cuanto al área geográfica no existen mayores diferencia relativas entre éstas, siendo su prevalencia del 0.4% en el área rural y del 0 3% en el área indígena y urbana respectivamente

3 El indicador antropométrico peso para la edad refleja que cerca del 8 5% de los niños y niñas menores de cinco años analizados, presentaban problema nutricional sin poder discriminar si esta desnutrición era crónica o aguda Al diferenciar por sexo, se observa igual prevalencia entre hombres y mujeres con un 4 2% respectivamente Con relación a la edad, este indicador presenta la misma tendencia que indicador anterior con una mayor frecuencia de casos en las edades de 1 y 3 años, con un 2 6% y 1 8% respectivamente

Al diferenciar por área de residencia la prevalencia es mayor en el área indígena con el 4 0% y rural con el 3 3% contra un 1 2% del área urbana

4 La evaluación nutricional de los niños y niñas menores de cinco años, basado en su retardo en su crecimiento utilizando como indicador antropométrico talla/edad, muestra una prevalencia de deterioro nutricional pasada o actual del 18 4% de los casos analizados Al diferenciar por sexo el 10 4% de los hombres y el 8 0% de las mujeres estaban desnutridos. Con relación a la edad, se observa una mayor porcentaje de desnutrición crónica en los mayores de 1 año de edad La prevalencia de desnutrición en los niños y niñas con 3 y 4 años de edad es de 4 8% y 4 7% respectivamente En los niños (as) con 1 año de edad la proporción es de el 3 8% y de un 3 6% en aquellos con 2 años de edad, frente a una proporción menor 1 51% observada en los menores de un año

Al diferenciar por área de residencia la prevalencia es de 9 5% en el área indígena, 6 6% en el área rural y 2 4% en el área urbana Por nivel de pobreza, el 16 6% de los niños y niñas desnutridos son pobres y un porcentaje (1 8%) menor son no pobres.

Los resultados obtenido del modelo de regresión logística multivariado nos permite hacer la siguiente conclusiones

5 Los niños residentes en el área indígena ofrecen un riesgo 11 276 veces mayor de sufrir desnutrición, los del área rural 2 382 veces mayor que aquellos que residen el área urbana

6 El riesgo de desnutrición disminuye en el sexo femenino (RÔ = 0 769) respecto al masculino (RÔ =1 00) Es decir, los niños tienen un 23 1% veces más riesgo de sufrir desnutrición que las niñas

7 Los niños y niñas entre 1 a 2 años tienen un riesgo de sufrir desnutrición (RÔ = 3 843) casi cuatro veces mayor y los de 3 a 4 años de edad seis veces mayor (RÔ = 6 238) en relación a los menores de un año

8 Las familias que no invierte en salud tienen un riesgo tres veces mayor (RÔ = 3 495) y los que invierten al menos 20 balboas anuales (RÔ = 3 814), están en claros riesgo de sufrir desnutrición en relación con aquellas familias que invierten más de 20 00 balboas anuales

9 En resumen se tiene que los grupos que tienen un riesgo especial de tener un deficiente estado nutricional son aquellos niños y niñas con las siguientes características mayores de un año, de sexo masculino, residentes en las áreas rurales e indígenas, con una inversión anual en salud menor de 20 balboas y un grupo familiar numeroso

RECOMENDACIÓN

1 Es importante que el estado panameño mejore la calidad y eficiencia de la educación, así como ampliar su cobertura, ya que esta constituye la mejor inversión a favor de los grupos más desfavorecidos, mejora la productividad y aumenta los ingresos y por ende contribuye a disminuir la prevalencia de desnutrición

2. Una prioridad de las políticas estatales es incrementar la producción de los alimentos básicos, como factor para reducir la pobreza y conseguir la seguridad alimentaría en el país

3 Es importante que los organismos rectores de la salud establezcan programas de educación sexual dirigidas a mujeres jóvenes en edades fértiles de áreas preferiblemente rurales e indígenas para tratar de reducir el número de embarazos y disminuir los riesgos de desnutrición asociados a la fecundidad temprana y el tamaño (numeroso) de los miembros en el hogar

4 Es necesario que los estudios que se realicen acerca de estado nutricional del niño incluyan variables asociadas a factores maternos, tales como espaciamiento de los nacimientos, hijos no deseados, edad de la madre al nacimiento del niño (a) y paridez, información fundamental para realizar un análisis profundo sobre la incidencia de la desnutrición infantil

BIBLIOGRAFÍA

Berenson, Mark Estadística Básica en Administración Conceptos y Aplicaciones, Prentice Hall México 1996

Hanke, John Estadística para los Negocios, McGraw-Hill/Irwin España 1994

Kohler, Heinz Estadística para Negocios y Economía CECSA México 1998

Sánchez, Emilio Regresión Logística en Salud Pública Escuela Andaluza de Salud Pública España 2000

Silva, Luis Carlos Excursión a la Regresión Logística en Ciencias de la Salud Díaz Santos España 1995

Perfil y Características de los Pobres Ministerio de Economía y Finanzas 1997

Manual del SPSS 2000

Conferencias

Moreno, Oscar Taller sobre Orientación Alimentaria y Conservación de Alimentos Instituto Nacional de la Nutrición México 1990