UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
UNIDAD ACADÉMICA DEL CICLO PROFESIONAL Y POSGRADO
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

"EL CÁLCULO DE VARIANZA EN MUESTREOS COMPLEJOS. APLICACIÓN EN LA ENCUESTA DE NUTRICIÓN 1996."

T E S I S

QUE PARA OBTENER EL GRADO DE

MAESTRÍA EN ESTADÍSTICA E INVESTIGACIÓN DE OPERACIONES

P R E S E N T A:

CATALINA PALMER ARRACHE

DIRECTORA DE TESIS:
DRA. GUILLERMINA ESLAVA GÓMEZ

MÉXICO, D.F. 1999

A la memoria de mis padres,
Augusto Palmer y Carmen Dolores (Alole) Arrache.
No cabe duda que a veces, "la ausencia se llena de presencia",
por eso también dedico este trabajo a la innegable presencia de ellos
en mi diario vivir.

A mi esposo Juan y
a mis hijos Mariana, Mari Carmen y Juan,
con todo mi amor.

A mis hermanos Nia y Tito,
y a los ya también hermanos, Hugue e Ivonne.

A los niños cuyas medidas conformaron la base de datos utilizada en esta tesis,
y a la población a la cual expande esta muestra,
con la esperanza de que en futuras encuestas, en toda la República Mexicana,
la proporción de niños en la categoría NORMAL sea mayor.

ÍNDICE

INDICE DE TABLAS

AGRADECIMIENTOS

Mi principal agradecimiento se dirige a la gran institución mexicana que es la Universidad Pública, en particular la Universidad Nacional Autónoma de México, que me ha proporcionado una educación de primer nivel, a través de todos los profesores del Departamento de Probabilidad y Estadística del IIMAS, quienes poseen una gran calidad académica y humana. He de señalar que siendo extranjera, valoro de manera especial la oportunidad que se medió de estudiar un posgrado. Cuando estaba por comenzar la maestría, el Dr. Ignacio Méndez Ramírez, la Dra. Silvia Ruiz Velasco y el M. en C. Rafael Madrid hicieron lo que estaba a su alcance para gestionar que me fuera otorgada una beca, a la cual no tenía derecho por no tener la nacionalidad mexicana. Finalmente, no recibí otra beca más que la de la *santa institución del matrimonio*, sin embargo, siempre agradecí la confianza y el apoyo que demostraron.

Agradezco a la Dra. Guillermina Eslava Gómez todo el tiempo que dedicó a la dirección de este trabajo, siempre con una actitud de compromiso y sobre todo con mucha paciencia. Nuevamente agradezco al Dr. Ignacio Méndez por sus valiosas recomendaciones durante la elaboración de la tesis y en la lectura final de la misma. Igualmente, le estoy agradecida a la Dra. Silvia Ruiz Velasco, al Dr. Federico O'Reilly Togno y al M. en C. Jose Antonio Flores, por sus observaciones y comentarios a esta tesis.

Por otra parte, este trabajo no hubiera sido posible sin la colaboración del Dr. Abelardo Ávila Curiel y la Lic. Nut. Teresa Shamah Levy del Instituto Nacional de la Nutrición "Salvador Zubirán", a quienes agradezco el tiempo y la atención que prestaron, así como sus aportaciones. De la misma forma, agradezco a la Lic. Nut. Linda Barragán el tiempo que dedicó a explicarme y depurar aspectos de la base de datos.

En realidad todos los profesores que tuve son grandes maestros y a todos les estoy agradecida, pero quiero agradecer de manera especial al Dr. Federico O'Reilly Togno, no sólo por su labor como maestro, sino más bien, por haberme brindado apoyo emocional en los momentos difíciles.

También agradezco la amistad y los buenos momentos que pasé junto a mis compañeros de clase : Hortensia Moreno, Adriana López, Rosario Tufiño, Yolanda Hernández, Luis Barrientos, Saúl Salazar, Gabriel Núñez y Miguel Ángel Covarrubias.

Por supuesto, agradezco a mi esposo y a mis hijos la compañía, y el amor que me dan, lo que me motiva a seguir luchando. También agradezco el cariño y apoyo de toda la familia Climént, que es mi familia en México, en especial a: Don Juan, Doña Margarita, Margarita, Herminia y Miguel Climént, así como Carlos Soto, a quien le agradezco que me iluminara el despacho, lo que me permitió trabajar mejor.

Introducción

Hoy en día el uso de encuestas por muestreo está cobrando terreno en nuestra sociedad. Ejemplos claros de la aplicación de encuestas lo son las investigaciones sociales de órganos gubernativos o académicos y estudios de mercado de empresas productoras o publicitarias. En la mayor parte de estas encuestas, resulta inevitable el basarse en un diseño muestral complejo, es decir un diseño que se base en diferentes etapas de selección. Usualmente, la razón por la que son necesarios estos diseños es la búsqueda de eficiencia en términos de operatividad y de costos, otras veces se determina un diseño por considerar que se logrará mayor precisión para los estimadores considerados más importantes. Es natural que el esfuerzo de su levantamiento se dirija a conocer muchas interrogantes sobre la población de interés; de igual forma, se plantean diversos niveles de estudio, desde los meramente descriptivos hasta los inferenciales o analíticos.

En México son muchas las organizaciones gubernamentales que levantan encuestas, de ellas , la más relevante es el Instituto Nacional de Estadística, Geografía e Informática (INEGI). El tema que revisa esta tesis surgió precisamente del planteamiento hecho por un funcionario a la Dra. Eslava, en cuanto al problema que radicaba en el cálculo de varianzas en el dominio de estudio y en subclases. Cabe mencionar, que en estas instituciones, aún cuando parte del personal a cargo de encuestas no está conformado por estadísticos, propiamente hablando, es impresionante el dominio que adquieren en los aspectos operativos, administrativos, de informática y hasta de criterios de diseño.

Sin duda, estas instituciones hacen un levantamiento correcto, siguiendo un esquema probabilístico, con personal capacitado o la debida asesoría, pero usualmente se enfrentan a la etapa de análisis y presentación de resultados, con poco tiempo y elementos para resolver situaciones específicas. Es natural entonces, que los reportes se planteen en términos muy descriptivos, y aunque las bases de datos se dejan al alcance de investigadores u otras instituciones para su estudio, probablemente nunca son explotadas a su debida magnitud, sobre todo para efectos de toma de decisiones. Quizás, porque detrás del planteamiento específico del cálculo de varianza se deslindaba esa realidad, llamó la atención a la autora de esta tesis, al hacer la revisión bibliográfica, el hecho de que varias de las investigaciones en torno al tema, se desarrollaron en instituciones gubernamentales, como lo es la Oficina del Censo de E.U. No se pretende con este comentario establecer una comparación entre el desarrollo de la estadística en México y en países desarrollados, sino expresar el franco deseo de que la estadística cobre cada día mejores espacios, y sea aplicada de manera razonable y adecuada. El problema del cálculo de varianzas, y los métodos de remuestreo que aquí se abordan, en realidad representan un pequeño escalón en el estudio analítico de encuestas (lo que sería un tema mucho más abarcador y atractivo). Pero además representa un humilde intento de abordar problemas que se ven en la práctica y merecen una atención cuidadosa.

El trabajo que aquí se presenta como tesis consiste en una revisión de los métodos de cálculo de varianza en muestreos complejos, dilucidando sus ventajas y desventajas, a manera de guía para quien deseara aplicar alguno de ellos, y su aplicación en la solución de varios estimadores en la Encuesta Nacional de Alimentación y Nutrición en el Medio Rural 1996, que fuera llevada acabo por el Instituto Nacional de la Nutrición, "Salvador Zubirán". Como estos métodos responden a un problema de estimación, se comenzó por explorar qué resultados de inferencia existían en el contexto de muestreo de poblaciones finitas, que permitieran establecer algunas propiedades deseadas en los estimadores de varianza. De tal suerte, el Capítulo 1 resume los resultados más relevantes de inferencia que se relacionan con el tema de esta tesis. En realidad, el apoyo en los criterios clásicos de inferencia, para el caso de muestreo complejo, es muy poco; aunque mucho se ha investigado, pocos resultados son positivos. Sin embargo no deja de ser intrigante todo el desarrollo teórico que ha habido en torno a un área considerada por muchos de lo más práctico de la estadística. Los siguientes dos capítulos explican y comentan los métodos de cálculo de varianza basados en remuestreo. Se consideró conveniente discutir también el método de linealización porque ha sido el más difundido en programas de cómputo con cierta comercialización, y con el cual se han comparado los métodos de remuestreo. Se sigue la aplicación a la encuesta nacional señalada.

Para penetrar en el terreno de estudio, es fundamental especificar a qué situaciones o tipos de diseño se avoca esta tesis. Al hablar de un diseño complejo, existen varios aspectos que califican su complejidad, los cuales son señalados por Wolter(1985) y se citan a continuación:

1. El grado de complejidad del diseño muestral.

2. El grado de complejidad del estimador o estimadores.

3. La existencia de características o variables múltiples que son de interés.

4. El uso que se le de a la encuesta: descriptivo o analítico.

5. El tamaño o escala de la encuesta.

La primera interrogante que surge es el por qué de la importancia de obtener un estimador de varianza. Para que una encuesta sea útil, debe ser analizada e interpretada adecuadamente, sin importar el grado de complejidad que posea su diseño. Es primordial entonces, el contar con una medida de precisión para cada estimador reportado. De tal suerte, resulta imprescindible el obtener una buena estimación de la varianza de los estimadores considerados en una encuesta.

La varianza de una estadística que proviene de una encuesta es función de la de la forma del estimador en cuestión y del diseño muestral empleado. En casi todos los diseños de muestreo complejos, es posible conseguir un estimador insesgado de la varianza, cuando la estadística de interés es una combinacion lineal de las observaciones. Sin embargo, es muy común el hecho de que la función de los datos de la muestra con la que se construye el estimador sea una función no lineal, como son las razones, diferencia de razones, y coeficientes de regresión y correlación, entre otros. En estos casos, no se cuenta con una expresión para el estimador de la varianza, por lo que se hace necesario el apoyo de otras formas de estimarla. También se puede dar el caso de un diseño aparentemente muy simple, de donde interesa una estadística lineal, y que no se encuentre la expresión de la varianza del estimador. Un ejemplo concreto es el muestreo sistemático por conglomerados, con probabilidades proporcionales al tamaño.

En la medida en que la aplicación y teoría del muestreo han cobrado terreno, son cada vez más los estudios para establecer y comparar métodos para calcular la varianza de un estimador cuando no se tiene una expresión para su estimación. Algunos métodos se basan en el remuestreo, lo cual conlleva un gran trabajo computacional. Otro enfoque para resolver el problema es aproximar la estadística no lineal con una función lineal. Además, existen esfuerzos recientes para brindar soluciones que buscan la simplificación del cálculo; por ejemplo, el método de *función generalizada de varianza*, que busca relacionar la varianza a la esperanza del estimador. Este método no se discute en esta tesis, ya que éste trabajo se centró en métodos de remuestreo, a excepción del de *linealización*. Wolter (1985, cap. 5) discute los aspectos básicos sobre el anterior).

Para algunas de las técnicas de cálculo de varianza, existe cierta teoría que las apoya, sin embargo algunas de ellas se conocen y se utilizan pero no están avaladas por un desarrollo teórico. También cabe mencionar que varios de estos métodos no surgieron del problema específico de un muestreo complejo, por lo que ha tomado años el llevar su aplicación a este caso y revisar las propiedades de los estimadores que resultan. Solamente se han identificado dos métodos dirigidos desde su origen al caso de población finita y en particular, un muestreo diferente al aleatorio simple: *grupos aleatorios* y *repeticiones balanceadas*; siendo que el primero pretendía además de calcular la varianza, tener evidencia de otras fuentes de error; y el segundo surgió para dar solución a un diseño en particular, lo que se aprecia en el Capítulo 2 de esta tesis. Aún en estos casos, se ha dado gran atención por parte de estadísticos dedicados al área de muestreo, a estudiar sus propiedades, posibles mejoras y comparación con otros métodos. La historia de cómo llegaron a aplicarse otras técnicas, como el *jackknife* y el *bootstrap* a un muestreo complejo, es un poco más larga, por haber sido concebidas en el marco de una población infinita, y por no resultar tan evidente la mecánica más adecuada para ejecutarla en un muestreo complejo. Así pues, no es sino hasta finales de los '80 que se empiezan a publicar trabajos sobre este tema en particular, siendo el principal autor J.N.K. Rao de la Universidad de Carleton en Canadá (Ver referencias varias).

Quien se acerca a este tema, puede traer en su mente la idea preconcebida de que el cálculo de varianza en realidad no debe merecer tanta atención, que no representa mayor problema. Posteriormente, trás revisar algunos métodos, se puede percatar de que existen congruencias de unos a otros, y finalmente puede afianzar su posición inicial sosteniendo que "son demasiadas formas de hacer lo mismo". Probablemente esta persona no se ha percatado de que si bien es cierto que existe dificultad para estimar la varianza de funciones no lineales de la muestra, como se ha dicho antes, el mayor obstáculo se presenta no por **un** cierto estimador en particular, sino porque se tiene un **conjunto** de estimadores de diferente índole, para los que hay que considerar su estimación y su varianza; así también, aunque el diseño de la encuesta fuera el mismo para todos los estimadores, se encuentran diferentes niveles de no respuesta, como puede darse un interés o necesidad por efectuar una posestratificación. Todo ello sin duda ha sido el móvil de tantos estudios dirigidos a este tema: el poder conciliar la solución de diferentes estimadores y sus varianzas, junto con situaciones como la no respuesta y la pos-estratificación en un sólo método. La búsqueda de esta solución, a lo largo de la historia de la estadística en este siglo (que está por concluir), se dió por diversos caminos, trabajando distintas líneas. En efecto, existen congruencias entre unos métodos y otros, pero éstas se han detectado después de sus respectivos desarrollos. La solución final no se ha dado. Cada técnica tiene sus ventajas y desventajas. Lo que se muestra en este trabajo, es pues, un espectro de soluciones y una aplicación muy particular.

Capítulo 1

Resultados de inferencia en muestreo

Ya que el cálculo de varianza en un muestreo complejo, representa un problema en torno a una estimación, resulta conveniente señalar, aunque sea brevemente, ciertos resultados de inferencia estadística en el caso particular del muestreo en una población finita. Básicamente, la inferencia estadística es una sola, sin embargo, en el caso del muestreo se tienen dos particularidades que impiden la aplicación directa de sus leyes: En primer lugar, el análisis se refiere a una población finita, en la que sus elementos pueden ser debidamente identificados (o etiquetados) y en segundo lugar, las observaciones no son independientes e idénticamente distribuidas; una situación particular que explica el último punto es la correlación entre datos de un mismo conglomerado. Además de éstos aspectos, se añade el hecho de que en inferencia (general) no se habla sobre el concepto de *diseño*. Si se reflexiona un poco sobre este punto, quizás solamente se halle similitud con el manejo de condiciones experimentales en el área de Diseño de Experimentos. Sin embargo, en todo uso de estadística hay *diseño*, si se entiende por ello, la forma de obtener los elementos de estudio, los procedimientos de medición, etc. Lamentablemente, los textos de inferencia no lo mencionan, y todo se desarrolla en torno a variables aleatorias independientes e idénticamente distribuidas (iid), de donde se supone un muestreo aleatorio simple de poblaciones infinitas, o con reemplazo de poblaciones finitas. Pero no se dice que en realidad en poblaciones infinitas no se puede tomar una muestra aleatoria, sino que se toma de algún modo y se *supone que es aleatoria y las variables iid. Por otra parte, en poblaciones* finitas, se supone que la aleatoriedad la introduce el investigador al tomar muestras probabilísticas. En otras palabras, se dice que hay un proceso que produce la aleatoriedad.

Éste capítulo pretende señalar, dentro del marco de la inferencia, qué propiedades se pueden pedir de un estimador de varianza en un muestreo complejo y encontrar los elementos que determinen la utilidad de estimarla. A continuación se abordan varios puntos que son fundamentales en cualquier problema de estimación en inferencia.

1.1 Verosimilitud vs. diseño

En el ámbito del muestreo, existen dos corrientes conceptuales, las cuales se catalogan como: *inferencia por aleatorización* e *inferencia basada en modelo* (Smith, 1993) o bien como: *acercamiento de población fija* y *acercamiento de superpoblación* (Cassel, Sarndal, Wretman, 1977) respectivamente . Aplicando el enfoque de *inferencia por aleatorización*, las respuestas de interés Y_i no son aleatorias. Se consideran constantes o parámetros, por lo que se cuenta un vector de N parámetros,$\mathbf{y} = (\mathbf{y}_1, ..., \mathbf{y}_N)$, de los cuales, a través del muestreo, se conocerán n de ellos, bajo el supuesto de que no hay error en la medición.

Resulta conveniente hacer explícita cierta notación, la cual se apega a Cassel, Särdal y Wretman (1977, pág. 29):

$$
\begin{aligned}
\mathbf{y} = (y_1, y_2, ..., y_N) &= \text{Valores (fijos) de interés en la} \\
& \text{ población de tamaño N.} \\
k \in \mathrm{s} & \text{ Notación que indica que } k \text{ es un} \\
& \text{ componente de la secuencia s} \\
\mathrm{s} = (k_1, ..., k_{n(S)}) = (k : k \in \mathrm{s}) &= \text{Secuencia de unidades no} \\
& \text{ necesariamente distintas.} \\
n(S) &= \text{Tamaño de la muestra completa.} \\
s = (k : k \in s) &= \text{Muestra de unidades distintas,} \\
& \text{ sin orden.} \\
\mathbf{S} &= \text{Secuencia aleatoria que toma}
\end{aligned}
$$

$$
\begin{array}{rcl}
 & & \text{los valores } s \in \Omega * . \\
S & = & \text{Secuencia aleatoria que toma} \\
 & & \text{valores } s \in \Omega. \\
\Omega* = \{\ \mathbf{S}\ \} & = & \text{Conjunto de todas las} \\
 & & \text{muestras ordenadas, } \mathbf{s}\ . \\
\Omega = \{S\} & = & \text{Conjunto de todas las} \\
 & & \text{muestras no-ordenadas, s.} \\
\mathbf{D}\ = ((k, y_k)k \in \mathbf{S}) & = & \text{Var. aleatoria de datos ordenados.} \\
D = ((k, y_k); k \in S) & = & \text{Var. aleatoria de datos} \\
 & & \text{donde no se repiten unidades.} \\
\mathbf{d}\ = ((k, y_k)k \in \mathbf{s}) & = & \text{Realización de } \mathbf{D}. \\
d = ((k, y_k); k \in s) & = & \text{Realización de D.}
\end{array}
$$

Siendo y fija, ocurre que la variable aleatoria es entonces,

$$
\begin{array}{rcl}
I(\mathrm{s}) & = & [I(1,\mathrm{s}), ..., I(N,\mathrm{s})] \\
\text{donde, } I(i,\mathrm{s}) & = & \left\{ \begin{array}{ll} 1 & \text{si el elemento i} \in \text{ s,} \\ 0 & \text{en otro caso.} \end{array} \right\} \\
 & & \text{para, i= 1,...,N.}
\end{array}
$$

Es decir, lo aleatorio es la selección de los individuos en lugar de la característica que se pretende medir, de donde se deriva el hecho de que la distribución muestral de un estimador (bajo este enfoque) depende de la acción del estadístico, a través del diseño que elige. Ahora bien, recordando que se entiende por *verosimilitud* la probabilidad de observar una muestra, vista como función de los parámetros y denotando a la probabilidad de obtener una muestra en particular como P(s), se tiene que

$$
L(y) = \left\{ \begin{array}{ll} P(\mathbf{s}\ ) & \forall\ y_i \in \Omega \ \ (\text{i=1,...,N}) \\ 0 & \qquad \text{en otro caso.} \end{array} \right\}
$$

Sin embargo, bajo este esquema de pensamiento, resulta que la verosimilitud es constante en relación a los parámetros, y como consecuencia, la inferencia basada en la verosimilitud es imposible.

La inferencia *basada en modelo*, como lo cataloga Smith, sí se basa en la verosimilitud, pero ésta se refiere a un modelo que se supone para la población. De tal forma, basándose en los valores observados en la muestra, se pretende estimar los no muestreados. Las críticas a este enfoque radican principalmente en la suposición de un modelo en la población.

Por el contrario, la inferencia por aleatorización está desligada de la verosimilitud. Resulta interesante advertir que si se aplicara el principio de verosimilitud al caso de una encuesta, se tendría que para dos diseños muestrales, tales que las probabilidades de las observaciones $s(y_i,\ i \in S)$ en uno sean un múltiplo de las probabilidades en el segundo diseño, cualquier inferencia sobre la población, debería ser la misma en ambos casos.

Lo anterior significa que de aplicarse el principio de verosimilitud, las inferencias **no** deben depender del diseño de muestreo, ni siquiera a través de las probabilidades $P(\mathbf{s})$, de las muestras observadas. Quien conoce un poco los métodos generalizados en la práctica de muestreo se puede percatar que éstos se ubican en el enfoque de *aleatorización* o *población fija* y que son inconsistentes con el principio de verosimilitud. Los estimadores ampliamente utilizados dependen funcionalmente del diseño y se buscan criterios como insesgamiento y reducción de la varianza.

Existe otro planteamiento basado en la verosimilitud de los datos observados, y_s, mediante una distribución hipergeométrica multivariada (Royall, 1968), que permite, hasta cierto punto, encontrar estimadores máximo verosímiles del número de elementos de la población que poseen determinado valor de la variable de interés. Consecuentemente, se puede encontrar un estimador máximo verosímil para la media y otros estimadores. Sin embargo, esta solución difícilmente se da para otro diseño que no sea el aleatorio simple o el aleatorio simple estratificado.

Resumiendo lo que compete al interés de esta tesis, se tiene que los estimadores a los cuales se hará referencia no son *máximo verosímiles*, como no lo son los métodos de estimar su varianza, más importante aún, hay que ubicar que el enfoque del que parten los métodos discutidos en esta tesis es el de *población fija*. Pero, ¿Qué justificación existe para usar estimadores que contemplan el diseño? Como ya se mencionó, bajo este enfoque, lo aleatorio no son los valores de la variable observada, sino el proceso de selección. De tal forma, al introducir la dependencia hacia el diseño de muestreo en el estimador, se introduce información sobre la aleatorización. En consecuencia, no sólo se habla de determinado estimador, $\hat{\Theta}$, sino que se especifica lo que se llama una estrategia de muestreo $H(D, \hat{\Theta})$, donde, D representa el diseño de muestreo. Se puede ser más explícito aún indicando la estrategia como: $H(D(S, P), \hat{\Theta})$, pues el diseño está dado por las posibles muestras que se pueden obtener, S, y su probabilidad, P.

Por último, cabe señalar que el interés en una encuesta no está en conocer aquellos valores de la población que no se midieron en la muestra, lo que interesa es alguna función de todos los valores de la población. De ahí se deriva la lógica frecuentista en la que se basa el enfoque de *población fija*, de imaginar repetidas muestras del mismo tamaño y con el mismo diseño. Para hacer inferencia se define un estimador y se evalúa la distribución de este estimador en repetidas muestras. Más adelante, en 1.4, se describen las bases establecidas por Hansen, Madow y Tepping para hacer inferencia en muestreo, basados en este enfoque.

1.2 Suficiencia y varianza mínima

En un problema de estimación en Inferencia, los conceptos de *suficiencia* y *suficiencia minimal* son centrales. Es de interés saber si existen resultados sobre suficiencia en el caso de muestreo. La relación directa con el tema de esta tesis es saber si se pueden obtener estimadores de menor varianza a través de la aplicación del teorema de *Rao-Blackwell* y quizás un estimador uniformemente con varianza mínima (UMVUE). De existir esa posibilidad, tal debe ser el estimador a utilizarse, y se procedería con el cálculo de su varianza.

Siendo puntual, sí existen resultados sobre suficiencia en muestreo de poblaciones finitas. Existe un estimador suficiente minimal para el vector de parámetros de la población $y = (y_1, y_2, \ldots y_N)$, pero éste no es completo, por lo que es posible conseguir un estimador de menor varianza (aunque a veces no resulta un estimador práctico), pero no existe un UMVUE en la clase de todos los estimadores, dada por Λ. Más adelante se verá que al restringir la clase de estimadores sí existe UMVUE, siendo un caso particular, la media muestral.

El estimador suficiente minimal para $y = (y_1, y_2, \ldots y_N)$ está dado por la variable aleatoria que representa el conjunto de los datos etiquetados, en el que no se repiten unidades. Esta variable está dada por $D = ((k, y_k);\ k \in S)$ (incluida en la nomenclatura de la sección anterior). La definición de suficiencia en el caso de muestreo es equivalente al concepto en inferencia pero conviene explicitarla aquí:

Una estadística $Z = u(\mathbf{D})$ es **suficiente** para el parámetro $y = (y_1, y_2, \ldots y_N)$, si y sólo si, la distribución condicional de $\mathbf{D}$ dado $Z = z$, no depende de y, siendo que su distribución condicional está bien definida.

Una estadística $Z_1 = u_1(\mathbf{D})$ es **suficiente minimal** para $y = (y_1, y_2, \ldots y_N)$, si y sólo si, para cada estadística suficiente $Z = u(\mathbf{D})$, cada partición de P_u es un subconjunto de la partición inducida por Z_1, P_{u1}.

Finalmente, ocurre que $D = u_0(\mathbf{D})$ es suficiente y suficiente minimal para y, cuya demostración se aprecia en Cassel, Särdal y Wretman (1977, págs. 36-38). Este resultado no es del todo satisfactorio, pues simplemente los datos, sin considerar orden ni unidades repetidas, son la estadística suficiente. Como se dijo antes, es posible algunas aplicaciones del teorema de Rao-Blackwell, pero no ofrecen una solución práctica.

No existe UMVUE en la clase de *todos los estimadores insesgados*, ni en la clase de los *insesgados lineales*, y para casi todos los diseños tampoco lo existe para los *lineales insesgados homogéneos*. Casi todos estos resultados se deben a Godambe (1955) y Godambe y Joshi (1965). Sin embargo, Hege, Hanurav y Lanke [1] vieron que los resultados de Godambe se aplicaban al caso de un diseño que no fuera *de uni-conglomerado*; entendiéndose por un diseño *uni-conglomerado* aquél en el que cualquiera dos muestras s_1, s_2, donde $p(s_1) > 0$, $p(s_2) > 0$, ocurre que las muestras son disjuntas o son equivalentes en el sentido de que ambas contienen el mismo conjunto de unidades distintas. Así pues el mismo Godambe (1965) corroboró que una condición necesaria y suficiente para que exista un UMVUE en la clase de estimadores insesgados lineales homogéneos, es que el diseño sea *uni-conglomerado*. Bajo estas restricciones, se encuentra que el estimador Horvitz-Thompson es

[1] citados en Chaudhuri A, *Optimality of Samplinf Strategies*, Handbook of Statistics, Vol. 6, Elsevier Science Publishers B.V., 1988, pags. 47-96.

UMVUE en la clase de estimadores insesgados para la media. También lo es la media muestral.

En Sukhatme y colaboradores (1984, págs. 38-40) se aprecia la demostración de que restringiendo la clase de estimadores insesgados para la media a los lineales, la media muestral es el estimador con menor varianza, por lo tanto es el mejor estimador lineal insesgado de la media poblacional. También demostraron que para la clase de estimadores del tipo Horvitz-Thompson, dada por

$$T_{HT} = \sum_{i \in s} \beta_i Y_i \quad ,$$

donde, las β_i son el inverso de las probabilidades de inclusión, la media muestral es el único estimador insesgado y por lo tanto es el mejor estimador insesgado para la media poblacional[2], $\overline{Y}$. En Chaudhuri (1988), Chauduhuri y Vos(1988) y Cassel, Sardal y Wretman (1977) se discuten estos hallazgos, así como aspectos sobre admisibilidad de estimadores, pero para los propósitos de esta tesis no se considera necesario abundar más al respecto.

Con los hechos que se han planteado se entiende que el problema a abordar no tiene salida por los conceptos de inferencia clásica. Si se tiene un diseño complejo y estimadores que no son lineales, no se encuentra un estimador de varianza mínima, que sería el lógico a escoger, ni siquiera se sabe si un estimador insesgado será el más adecuado.

1.3 Propiedades de los estimadores

Otros criterios, quizás menos exigentes, para elegir un estimador en el caso de una población finita, son los de *insesgamiento*, *consistencia* y *error cuadrático medio*. A continuación se explican los criterios mencionados.

Estimador insesgado : Cuando el valor promedio del estimador, sobre todas las muestras posibles de tamaño n, es exactamente igual al valor poblacional, se dice que el estimador es insesgado. Es decir, sea $\hat{\Theta}$ un estimador insesgado de Θ, entonces:

[2] Sukhatme y colaboradores (1984, pág.81).

$$E(\hat{\Theta}) = \sum_{s \epsilon S} \hat{\Theta}_s P(s) = \Theta$$

donde $\hat{\Theta}_s$ es el valor de $\hat{\Theta}$ que corresponde a la muestra s.

Por otra parte, el sesgo de un estimador está dado por:

$$SESGO(\hat{\Theta}) = E(\hat{\Theta}) - \Theta$$

El sesgo resulta ser de especial interés pues los estimadores de las varianzas de muestreos complejos suelen ser sesgados. Ocurre además que el método *jackknife* se concibió como un estimador para reducir sesgo.

Estimador consistente: En el caso de poblaciones infinitas se dice que un estimador $\hat{\Theta}$ es consistente sí:

$$lim_{n \to \infty} \left(P[\mid \hat{\Theta} - \Theta \mid \; > \epsilon] \right) \longrightarrow 0$$

En otras palabras, el estimador $\hat{\Theta}$ es consistente si asume el valor Θ con una probabilidad que converge a 1, cuando el tamaño muestral aumenta indefinidamente. Cabe advertir que un estimador consistente no necesariamente será insesgado, pero si es sesgado, el sesgo tenderá a cero mientras n tiende a infinito. Ahora bien, para el caso de poblaciones finitas, Sukhatme et al(1984) define un estimador consistente como aquél que asume el valor de Θ cuando se toma la población total como muestra, es decir, cuando $n= N$.

Los estimadores usados con mayor frecuencia son consistentes pero no insesgados. Sin embargo, un estimador inconsistente puede ser útil, pues puede tener buena precisión cuando n es pequeña comparada con N.

Error cuadrático medio: Para un estimador $\hat{\Theta}$, éste se define como

$$ECM_{\Theta}\left[\hat{\Theta}\right] = E(\hat{\Theta} - \Theta)^2 = V(\hat{\Theta}) + SESGO^2$$

El error cuadrático medio es útil para comparar dos estimadores sesgados o un insesgado con un sesgado.

1.4 Bases de inferencia en muestreo

Los intervalos de confianza para datos que provienen de muestreos complejos, basados en la distribución *Normal(0,1)* o en la distribución t de *Student* son de uso generalizado. Sin embargo, el por qué se hacen de esta manera, a veces no es claro para los usuarios de estadística.

Hansen, Madow y Tepping (1983, págs. 776-793) describen los principios de inferencia desde el punto de vista de la aleatorización (Enfoque de *población fija*). Definen un diseño de muestreo probabilístico como aquél que consiste de:

a) Un plan de muestreo tal que cada miembro de la población tiene una probabilidad conocida, mayor a cero, de ser incluido en la muestra.

b) Procedimientos de inferencia tales que para muestras razonablemente grandes, la exactitud de las inferencias no dependen de la suposición de un modelo.

De esta conceptualización se deriva el hecho de que las inferencias se basan en distribuciones límites que son inducidas por la aleatorización; por lo tanto, un intervalo de confianza, en este contexto, es válido si se interpreta en términos de la probabilidad de aleatorización. Es decir, la probabilidad de que un intervalo de confianza contenga el valor estimado, es igual o mayor al valor nominal del nivel de confianza, **bajo la probabilidad inducida** por la forma en que se hizo la selección.

Así pues, las inferencias usualmente se derivan de la cantidad pivotal:

$$t = \frac{\hat{\Theta} - \Theta}{\sqrt{v(\hat{\Theta})}}$$

En la expresión de arriba, $v(\hat{\Theta})$ es un estimador de la varianza de $\hat{\Theta}$, y se supone que t se distribuye, al menos aproximadamente, como una N(0,1). Su aproximación a la distribución t de *Student* es más correcta, pero como las muestras en encuestas son grandes, se tiende a utilizar automáticamente la distribución Normal. Sin embargo, en el caso en que la varianza se calcula por métodos de remuestreo, los grados de libertad son menos y se debe observar el uso de la distribución t de Student. Más adelante en esta tesis se ve que estos intervalos son válidos para varios de los métodos discutidos. En particular, en la sección 1.6 se ven resultados para el *jackknife*, *repeticiones balanceadas* (*Half- sampling*) y *linealización*.

Si se tiene un intervalo de confianza, es lógico querer transformarlo para construir una prueba de hipótesis. Sin embargo, se ha visto que cuando las muestras son grandes, tales pruebas no son adecuadas. "Si la muestra tiene un tamaño suficientemente grande, la relación más débil y menos significativa aparecerá con significancia estadística"[3]. Este planteamiento ha sido por años uno de los puntos debatidos por los estadísticos bayesianos. Sin duda, lo razonable es expresar resultados en intervalos de confianza. Pero, con la reciente difusión de los métodos de estimación por remuestreo, se ha visto que éstos proveen una forma robusta de hacer pruebas de significancia [4]. Anteriormente, se señaló que al aplicar estas técnicas los grados de libertad son mucho menores, lo que hace a un lado el problema que representa un tamaño de muestra tan grande, es decir, el hecho de que en todas las pruebas de significancia con gran tamaño de muestra todo sea significativo aunque no sea importante. El cálculo de grados de libertad bajo diferentes métodos de remuestreo se muestra en el Apéndice B, citando a Rust y Rao (1996), pues es la única referencia que se encontró al respecto.

1.5 Aspectos de interés sobre la varianza

En el marco de inferencia, la varianza es uno de los parámetros que define la distribución sobre la cual se quiere inferir y es en este aspecto, en el que radica la importancia de estimarla. Asimismo, en el caso de muestreo de poblaciones finitas, se vió, en la sección anterior, que el estimador de la varianza o la desviación estándar juega un papel impor-

[3]Palabras de Yates citadas por Kish en: Kish, L.,Muestreo de Encuestas, Trillas, 1972, pág. 678.

[4]Rust K.F., Rao J.N.K., *Variance Estimation for Complex Surveys Using Replication Techniques*, Statistical Methods in Medical Research, 1996; 5, p.283-310.

tante para establecer intervalos de confianza y en algunos casos, hacer pruebas de hipótesis. Al respecto conviene citar a Rao (1997): "Ignorar el diseño de muestreo y analizar los datos de forma convencional puede llevar a una seria sub-estimación de la desviación estándar, niveles de pruebas inflados y diagnósticos erróneos".

Es claro entonces, que bajo el esquema de aleatorización es fundamental la consideración del diseño en los estimadores de varianza. A su vez, la estimación de la varianza en muestreos complejos tiene importancia como medida de la precisión del diseño; ésto permite poder establecer comparaciones entre distintos diseños. Obviamente, el interés de una encuesta no es estimar una varianza, pero sí lo es obtener varios estimadores de valores en la población, y éstos se quieren conocer con la mayor precisión posible. La comparación de diferentes diseños es útil, sobre todo, en situaciones en las que se requiere periódicamente levantar una encuesta similar, pues se puede optar por cambios en el diseño para un futuro proyecto. Por ejemplo se puede apreciar si cierta estratificación tuvo o no el efecto deseado en reducción de varianza.

Lo más común es la comparación con el muestreo aleatorio simple con reemplazo y sin reemplazo. Las medidas más difundidas en la práctica para este efecto son: el **DEFF**, cuando se compara contra el m.a.s. sin reemplazo (m.a.s.s.r.), y el $DEFT^2$, si se compara contra el m.a.s. con reemplazo (m.a.s.c.r.). Tales medidas son definidas por Kish (1989) de la siguiente manera:

$$DEFF(\hat{\Theta}) = \frac{Var(\hat{\Theta}_{\text{diseño}})}{Var(\hat{\Theta}_{\text{m.a.s.s.r.}})} \tag{1.1}$$

$$DEFT^2(\hat{\Theta}) = \frac{Var(\hat{\Theta}_{\text{diseño}})}{Var(\hat{\Theta}_{\text{m.a.s.c.r.}})} \tag{1.2}$$

Las verdaderas varianzas son desconocidas en la mayoría de los casos, por lo que normalmente, sólo se obtienen estimadores de dichas medidas, substituyendo en el numerador y denominador los estimadores de las varianzas correspondientes. En ocasiones esta medida sirve para simplificar la presentación de resultados y evitar un excesivo número de cálculos de varianza que consideren diseño . Cuando hay elementos para pensar que varias variables de la encuesta tienen un DEFF similar (usualmente se sabe por encuestas anteriores) sólo se exhiben estimadores basados en m.a.s. - los que calcula un paquete de cómputo fácilmente - y se indica el valor de DEFF para ese grupo de variables. De esta forma, quien está interesado en conocer una varianza o un intervalo, estima la varianza mediante el producto de la varianza de m.a.s. y el DEFF. Esta simplificación de resultados puede ser muy objetada porque depende mucho de la debida inspección de encuestas anteriores y la correcta agrupación de variables; sin embargo es una práctica muy difundida, y que se debe, en parte, a las dificultades que enfrenta un equipo de análisis de una encuesta. Kish (1989) discute ampliamente la utilidad de esta medida.

Se ha hablado de la necesidad de conocer la varianza, pero trás la breve exposición de resultados de inferencia en muestreo, resulta de interés establecer qué propiedades se pueden buscar en un estimador de varianza en un muestreo complejo, bajo un enfoque de *población fija*. Las propiedades que se revisan en los estimadores de varianza en general, son: insesgamiento y consistencia. En el caso de estimadores no lineales, en ocasiones se busca que el método de estimación, de ser aplicado a una estadística lineal (por ejemplo, la media), produzca el mismo estimador que se obtendría por la fórmula, según el diseño, y que la esperanza del estimador de varianza sea también la conocida para el caso. Obviamente, ésto no garantiza que al aplicarse el método a una estadística no lineal, se obtengan las mismas propiedades.

Por otra parte, es conveniente aclarar que muchas veces se habla de varianza o *error cuadrático medio* (ECM) casi indistintamente. Se sabe que

$$ECM(\hat{\Theta}) = Var(\hat{\Theta}) + Sesgo^2(\hat{\Theta}).$$

Como $Sesgo^2(\hat{\Theta})$ es de orden menor que $Var(\hat{\Theta})$, una aproximación para $ECM(\hat{\Theta})$ de primer orden, también lo sería para $Var(\hat{\Theta})$. Por tal razón, muchas veces lo que se obtiene es una aproximación del ECM y se toma como estimación o expresión para la varianza. Así pues, en ocasiones, se elige entre dos posibles estimadores de acuerdo al que tenga menor error cuadrático medio. En el caso de estimadores de varianza, cabe la posibilidad de comparar $ECM(\hat{Var}(\hat{\Theta}))$, pero llega a ocurrir que un estimador con menor ECM, no da los mejores intervalos, es decir, la probabilidad de inclusión en el intervalo es menor al valor nominal. Lo que se verá al respecto en esta tesis (en 4.3.) es que el comportamiento de los métodos de remuestreo, han sido inspeccionados a través de simulaciones por distintos autores, ya que no es posible hacerlo de forma analítica.

Es interesante mencionar que el mismo problema que representa el cálculo de varianza en un muestreo complejo, se da para calcular una matriz de varianza-covarianza. Estas matrices son necesarias especialmente, cuando se desea hacer un análisis multivariado. Nathan (1988) considera que el método de *linealización*, el de *repeticiones balanceadas* y el *jackknife* producen buenos estimadores para estas matrices.

Como se advirtió en la introducción, son pocos los resultados de inferencia clásica que servirán de apoyo en la discusión de los métodos de cálculo de varianza. No existen resultados en los que se pueda basar la elección de un estimador, pero sí existen otros que permiten hacer inferencia sobre los estimadores $\hat{\Theta}$, basándose en $\hat{Var}(\hat{\Theta})$, obtenida por distintos métodos.

1.6 Resultados específicos sobre los métodos de remuestreo

Krewski y Rao (1981) sentaron las bases de la teoría asintótica del estimador de linealización, el jackknife y el de repeticiones balanceadas (*Half-Sampling*). Esta teoría se da en el marco de una secuencia infinita de poblaciones finitas, $\{\prod_L\}_{L=1}{}^{\infty}$ con L estratos en $\{\prod_L\}$. Los resultados son válidos en el contexto de un diseño estratificado multietápico en el que las unidades primarias de muestreo se seleccionan *con reemplazo* y en las que se toman muestran independientes, cuando una unidad resulta seleccionada más de una vez. Además, el análisis de Krewski y Rao se avoca al caso en el que se tienen muchos estratos con pocas unidades primarias en cada estrato.

El interés radica en parámetros de la forma $\Theta = g(\overline{\boldsymbol{Y}})$, y el estimador correspondiente, $\hat{\Theta} = g(\overline{\boldsymbol{y}})$, siendo $g(.)$ una función real, $\overline{\boldsymbol{Y}}$ la media poblacional, ($\overline{\boldsymbol{Y}}= \sum_{h=1}^{L} W_h \overline{\boldsymbol{Y}}_h$), $\overline{\boldsymbol{y}}$ un estimador insesgado de $\overline{\boldsymbol{Y}}$ ($\overline{\boldsymbol{y}}= \sum_{h=1}^{L} W_h \overline{\boldsymbol{y}}_h$), y W_h el peso del estrato h, dado por la proporción de unidades muestreadas en el estrato y las unidades en la población ($W_h = \frac{N_h}{N}$). Se denomina por $\boldsymbol{Y_h}= (Y_{h1}, ..., Y_{hp})'$, los valores poblacionales de la unidad i-ésima del estrato h, por $\boldsymbol{y}_{hi}= (y_{hi1}, \ldots, y_{hip})'$ los valores muestrales respectivos; de manera similar, $\overline{\boldsymbol{Y}}_h= (\overline{Y}_{h1}, \ldots, \overline{Y}_{hp})'$ y $\overline{\boldsymbol{y}}_h= (\overline{y}_{h1}, \ldots, \overline{y}_{hp})'$, son los vectores de medias poblacionales y muestrales de los estratos. La matriz de varianza-covarianza de $\overline{\boldsymbol{y}}$ está dada por

$$\Sigma = \sum_{h=1}^{L} W_h^2 n_h^{-1} \Sigma_h$$

donde Σ_h es una matriz (p x p) cuyo elemento típico es de la siguiente forma:

$$\sigma_{hj,hj'} = N_h^{-1} \sum_{i=1}^{N_h} (Y_{hij} - \overline{Y}_{hj})(Y_{hij'} - \overline{Y}_{hj'}) \quad ,$$

Por otra parte, se supone que la secuencia de poblaciones satisface las siguientes condiciones de regularidad :

i) $\sum_{h=1}^{L} W_h E\{| y_{hij} - \overline{Y}_{hj}|^{2+\delta}\} = 0(1)$, para algún $\delta > 0$, $j = (1,\ldots,p)$.

ii) $max_{1 \leq h \leq L}\{n_h\} = 0(1)$

iii) $max_{1 \leq h \leq L}\{W_h\} = 0(L^{-1})$

iv) $n \sum_{h=1}^{L} W_h^2 n_h^{-1} \Sigma_h \rightarrow \Sigma_*$, donde Σ_* es p x p, definida positiva.

v) $\overline{Y}_k \rightarrow \mu_k$(finita) para k=1,...,p.

vi) Las primeras derivadas $g_j(.)$ de $g(.)$, son continuas en una vecindad de $\mu = (\mu_1,\ldots,\mu_p)$.

Sea $v_L(\hat{\Theta})$, $v_J(\hat{\Theta})$, $v_{RB}(\hat{\Theta})$ los estimadores de varianza obtenidos por linealización, jackknife y repeticiones balanceadas respectivamente; así también, t_L, t_J y t_{RB} son estadísticas que parten de los mismos métodos. Bajo las condiciones anteriores, Krewski y Rao (1981) demostraron los siguientes resultados:

a.)

$$\begin{aligned} n v_L(\hat{\Theta}) &\xrightarrow{p} \sigma_{\Theta}^2 \\ t_L = \frac{\hat{\Theta} - \Theta}{\sqrt{v_L(\hat{\Theta})}} &\xrightarrow{d} N(0,1) \end{aligned}$$

b.)

$$
\begin{aligned}
nv_J(\hat{\Theta}) &\xrightarrow{p} \sigma^2_{\Theta} \\
t_J = \frac{\hat{\Theta} - \Theta}{\sqrt{v_J(\hat{\Theta})}} &\xrightarrow{d} N(0,1)
\end{aligned}
$$

c.)

$$
\begin{aligned}
nv_{RB}(\hat{\Theta}) &\xrightarrow{p} \sigma^2_{\Theta} \\
t_{RB} = \frac{\hat{\Theta} - \Theta}{\sqrt{v_{RB}(\hat{\Theta})}} &\xrightarrow{d} N(0,1)
\end{aligned}
$$

Al examinar las condiciones iniciales del teorema, se aprecia que para considerar válidos estos resultados, se requiere del supuesto de que el número total de unidades $n = \sum n_h$, sea grande, y que ningún factor de expansión sea desproporcionadamente mayor a los demás. Los resultados expresan que los tres estimadores de varianza son asintóticamente consistentes para la verdadera varianza y que los intervalos de confianza asociados a la teoría normal también son válidos asintóticamente.

Capítulo 2

Métodos de remuestreo surgidos en la práctica

Por lo general, los problemas que se dan en la práctica necesitan de una inteligencia intuitiva que de soluciones rápidas. Posteriormente esas soluciones se pueden afinar, incorporando elementos de estudio más detenido. Este es el caso de los dos métodos de cálculo de varianza que se exponen en este capítulo. Ambos, el de *grupos aleatorios* y el de *repeticiones balanceadas* se iniciaron con lo que en un momento dado parecía una solución lógica y creativa. Posteriormente, se observaron ciertas deficiencias y los métodos fueron mejorados. A pesar de que datan de mediados de este siglo, los dos siguen siendo muy útiles; *repeticiones balanceadas* da solución a un diseño común en encuestas de gran cobertura; por otra parte, *Grupos Aleatorios* fue durante muchos años el único procedimiento factible de ser aplicado, y varios organismos internacionales reportan que no es hasta fechas recientes que han considerado la aplicación de otro método, si no es el caso de que continúan aplicándolo.

2.1 Grupos aleatorios

2.1.1 Caracterización del estimador

El método de Grupos Aleatorios, como ya se ha mencionado en la introducción, surgió en el trabajo de campo. Mahalanobis (1946), quien trabajaba en la India en problemas aplicados a la agricultura (en particular, la estimación de cosechas de diferentes productos) se vió en la necesidad de reducir las fuentes de error en sus encuestas. Es decir, estaba consciente de que el error total no consistía nada más de la variación del muestreo sino también de fallas humanas. Por tal razón, comenzó a diseñar muestras a las que llamó *interpenetrantes*, o *muestras a la mitad* ("half-sampling")[1] . Básicamente su técnica consistía en numerar consecutivamente las unidades de muestreo dentro de cada estrato o unidad primaria; posteriormente, separaba la muestra de acuerdo a dicho número en nones y pares, formando dos conjuntos de datos que daban dos estimaciones. De tal suerte, con las dos muestras estaba también en posibilidades de calcular la varianza muestral. Aunque Mahalanobis procedía en su análisis suponiendo la independencia entre esas dos muestras, es evidente que los estimadores de cada una de ellas estaban correlacionados porque se construían obteniendo grupos mutuamente exclusivos; por lo tanto, no existía tal independencia y en realidad, su estimación incurría en un sesgo.[2]

El término *interpenetrante* alude a la aplicación del método al evaluar márgenes de error diferentes a la variabilidad del fenómeno que se observa. Sin embargo, la idea de Mahalanobis fue considerada por autores como Hansen, Hurwitz y Madow (1953), llamándole la *técnica del conglomerado último* (*Ultimate cluster*) o del *grupo aleatorio*, dándole un carácter más general. Deming (1956) la utilizó con el simple propósito de obtener un estimador de varianza sin importar lo compli-

[1]Curiosamente Mahalanobis comenta que quiso llamarle *muestreo duplicado* (*Double sampling*), pero observó que este nombre producía en sus clientes la sensación de que la encuesta les costaba el doble, por lo que prefirió el término *Half-sampling*.

[2]El señalar un error de alguien que dejó una obra tan reconocida no es del agrado de la autora, pero resulta un hecho didáctico en el aspecto estadístico y de la vida en general.

cado que fuera el estimador o el diseño muestral y le llamó *Muestreo replicado.*

A pesar de que las muestras de Mahalanobis no eran precisamente independientes, el concepto general de la metodología tuvo aceptación porque ayudaba a resolver varios problemas. De ahí que se desarrollaran dos formas de llevar a cabo esta técnica, basadas en: *muestras independientes* o *muestras dependientes.* En el primer caso, se supone una población finita P, de la cual, bajo cierto diseño, se extraen k muestras, s_1,s_2, ..., s_k; siendo que cada una de ellas se repone en la población antes de elegir la siguiente. De tal suerte, se obtienen k *muestras independientes.* Si en lugar de esta serie de muestras independientes se cuenta con una muestra total, la que es dividida aleatoriamente en k grupos, entonces se estaría hablando de k *muestras dependientes.* El estimador de la varianza en ambos casos es el mismo; la única diferencia es que al ser las muestras dependientes, se introduce un sesgo. Más adelante se mostrarán las características del estimador en un caso u otro. La única referencia explícita sobre este método es Wolter (1985), a excepción de la versión llamada *grupos aleatorios repetidos*, que se discute en este capítulo.

De manera general, sean $\hat{\Theta}_1$,$\hat{\Theta}_2$,...,$\hat{\Theta}_k$ los estimadores de Θ evaluados en los k grupos aleatorios. Se considera también a $\bar{\hat{\Theta}}$ como estimador del parámetro de interés, Θ, que se consigue de la siguiente forma:

$$\bar{\hat{\Theta}} = \sum_{i=1}^{k} \frac{\hat{\Theta}_i}{k}$$

Bajo el esquema de grupos dependientes ocurre que la esperanza del estimador en el grupo i-ésimo no necesariamente corresponde al valor esperado de la población. Es decir, $E(\hat{\Theta}_i) = \mu_i$, donde μ_i no necesariamente es igual a $\Theta = \mu$. De ahí que la esperanza del estimador de grupos dependientes es la media de las esperanzas, o sea, $E(\bar{\hat{\Theta}}) = \sum_{i=1}^{k} \frac{\mu_i}{k}$. Por otra parte, si los grupos son independientes entonces el estimador es insesgado.

El estimador de la varianza de $\bar{\hat{\Theta}}$, tanto para el caso de grupos

Muestras independientes	Muestras dependientes
$\hat{\Theta}_1,\ldots,\hat{\Theta}_k$ independientes.	$\hat{\Theta}_1,\ldots,\hat{\Theta}_k$ correlacionadas.
$E(\hat{\Theta}_i) = \mu$ (i=1,...,k)	$E(\hat{\Theta}_i) = \mu_i$ (i=1,...,k)
$E(\overline{\hat{\Theta}}) = \mu = \Theta.$	$E(\overline{\hat{\Theta}}) = \overline{\mu}.$
$E(v(\overline{\hat{\Theta}})) = Var(\overline{\hat{\Theta}}).$	$E(v(\overline{\hat{\Theta}})) = Var(\overline{\hat{\Theta}})$
	$+\frac{1}{k(k-1)}\sum_{i=1}^{k}(\mu_i - \overline{\mu})^2$
	$-2\frac{\sum_{i=1}^{k}\sum_{j>i}^{k} Cov\{\hat{\Theta}_i,\hat{\Theta}_j\}}{k(k-1)}.$

Tabla 2.1: Propiedades de estimadores de grupos aleatorios.

dependientes o independientes, está dado por:

$$v(\overline{\hat{\Theta}}) = \frac{\sum_{i=1}^{k}(\hat{\Theta}_i - \overline{\hat{\Theta}})^2}{k(k-1)} \tag{2.1}$$

Este estimador es insesgado si los grupos son independientes pero incurre en un sesgo si existe dependencia entre los grupos. En la tabla 2.1 se esbozan ciertas características de estos estimadores, de acuerdo al tipo de las muestras.

Como se advierte en la tabla anterior, no existe, en ninguno de los casos, el supuesto de que las varianzas de los k estimadores, $\hat{\Theta}_1$, ..., $\hat{\Theta}_k$, sean iguales. Esto implica que los k grupos, aunque sean independientes, no tienen que extraerse bajo un mismo diseño muestral. La verdad es que el supuesto de igualdad de varianza no se requiere para que $v(\overline{\hat{\Theta}})$ sea insesgado, pero si se desea hacer inferencias sobre Θ, dicho supuesto es necesario. A continuación se exhibe la esperanza de $v(\overline{\hat{\Theta}})$, bajo el supuesto de que los grupos son independientes.

Facilita el proceso, el considerar que $v(\overline{\hat{\Theta}})$ (definido en (2.1)) se puede expresar como sigue:

$$v(\overline{\hat{\Theta}}) = \frac{1}{k(k-1)}\left[\sum_{i=1}^{k} \hat{\Theta}_i^{\,2} - k\overline{\hat{\Theta}}^2\right]$$

Así pues,

$$\begin{aligned}
E\left[v(\overline{\hat{\Theta}})\right] &= \frac{\sum_{i=1}^{k}\left(Var(\hat{\Theta}_i) + \mu^2\right) - k\left(Var(\overline{\hat{\Theta}}) + \mu^2\right)}{k(k-1)} \\
&= \frac{k^2 Var(\overline{\hat{\Theta}}) + k\mu^2 - kVar(\overline{\hat{\Theta}}) - k\mu^2}{k(k-1)} \\
&= \frac{(k^2 - k)Var(\overline{\hat{\Theta}})}{k(k-1)} \\
&= Var(\overline{\hat{\Theta}}).
\end{aligned}$$

Es posible que el lector haya advertido la posibilidad de utilizar otro estimador de Θ, diferente de $\overline{\hat{\Theta}}$. Éste sería el que se obtendría con la muestra total, compuesta de los k grupos. Es decir, una muestra de tamaño n, donde, $n = \sum_{i=1}^{k} n_i$, y n_i es obviamente, el tamaño de la muestra i-ésima. Si el estimador es lineal en las observaciones, este nuevo estimador, que se denomina $\hat{\Theta}$, será igual a $\overline{\hat{\Theta}}$. Pero generalmente,

$$\overline{\hat{\Theta}} \neq \hat{\Theta}$$

Si se utiliza $\hat{\Theta}$, entonces es lógico calcular su varianza como sigue.

$$v(\hat{\Theta}) = \frac{\sum_{i=1}^{k}(\hat{\Theta}_i - \hat{\Theta})^2}{k(k-1)}$$

Se ha visto que $v(\hat{\Theta}) \geq v(\overline{\hat{\Theta}})$. Pero al respecto, Wolter(1985, Capt. 2) señala el hecho de que en muestreos complejos y de tamaño de muestra grande, los dos estimadores de varianza mencionados, no difieren mucho.

Grupos aleatorios repetidos

Otra versión de la técnica de Grupos Aleatorios, según Kovar, Rao y Wu (1988), es la de *grupos aleatorios repetidos* (GAR), la cual se avoca al caso en que se muestrean igual número de unidades, U, en cada estrato. La muestra se divide en U grupos, tal que cada una de los conglomerados o unidades en cada estrato, cae en sólo una de esas submuestras; y se extraen los U estimadores, $\hat{\Theta}_u$, $(u = 1, ..., U)$ que les corresponden. Este proceso se repite un número arbitrario de veces, R, y la varianza se estima con respecto al estimador global $\hat{\Theta}$, o con respecto a la media del estimador en la repeticón u-ésima dada por: $\overline{\Theta}_r = \sum_{u=1}^{U} \hat{\Theta}_{ru}/U$, $(r = 1, ..., R)$, como se ve aquí:

$$v1_{GAR} = \frac{1}{R} \frac{\sum_{r=1}^{R} \sum_{u=1}^{U} (\hat{\Theta}_{ru} - \hat{\Theta})^2}{U(U-1)} \tag{2.2}$$

$$v2_{GAR} = \frac{1}{R} \frac{\sum_{r=1}^{R} \sum_{u=1}^{U} (\hat{\Theta}_{ru} - \bar{\Theta}_r)^2}{U(U-1)} \tag{2.3}$$

Esta forma de estimar la varianza consiste en obtener grupos dependientes, para una misma muestra total, un número repetido de veces. Entonces, dentro de cada réplica se tienen grupos dependientes, aunque dichas repeticiones son independientes entre sí. Con la información de la tabla 2.1 se da uno cuenta que la esperanza del estimador de varianza de GAR es la misma que la de grupos aleatorios Dependientes, ya que:

$$\begin{aligned} E\left[v1_{GAR}\right] &= \frac{1}{R}\sum_{r=1}^{R} E\left[v(\overline{\hat{\Theta}})\right] \\ &= E\left[v(\overline{\hat{\Theta}})\right] \end{aligned}$$

donde, $v(\overline{\hat{\Theta}})$ es el estimador de grupos aleatorios Dependientes.

Como un adelanto al siguiente método que se discute en este capítulo, se hace la notación de que cuando $U = 2$, el proceso resulta parecido al de *Muestreo por Mitades* o *repeticiones balanceadas* (RB) (Sección 2.2, y el inciso 2 de 4.4).

2.1.2 Consideraciones prácticas

En esta tesis se aborda el método de grupos aleatorios como una técnica para calcular la varianza, aunque se ha hecho mención de que en ocasiones se puede aplicar para comparar varianzas por encuestador u otro tipo de control. El método de muestras independientes, en particular, es apta para cumplir ambos propósitos.

Al construir grupos independientes, se recomienda que los encuestadores, así como otros trabajadores en la operación de la encuesta, se asignen a unidades que pertenezcan a un mismo grupo. De esta manera, se evita la existencia de correlación en los datos de diferentes grupos, debida a la fuente de error que puede introducir una persona al desempeñar su trabajo. Finalmente, la independencia entre las muestras sería lo más certero que se pued construir. Por otra parte, si un encuestador introduce cierto sesgo, éste se reflejará en los resultados que arroje su grupo, y se esperaría, que estas cifras contrastaran con las de los otros grupos. De ahí, la doble utilidad del método.

Sobre la construcción de muestras dependientes, es importante advertir que los grupos deben conservar el mismo diseño de muestreo que la muestra total. Igualmente, es recomendable, en el caso de muestras independientes pues así se garantizan las bases de inferencia, lo que se explica en la última sección de este capítulo. En el Apéndice A se exhiben las reglas dadas por Wolter (1985) para formar grupos conservando el diseño. A continuación se describe cómo se formarían grupos para un caso en particular. La tabla 3.2 muestra la estratificación que se estipuló para el estado de Jalisco en la ENAL96[3].

En esta encuesta las localidades dentro de los estratos representan conglomerados y las viviendas encuestadas en la localidad vienen a ser las unidades elementales de muestreo. (Mayores detalles sobre el diseño y otros aspectos de esta encuesta se aprecian en el capítulo 5). Bajo un diseño estratificado bietápico, como el que se utilizó en la encuesta, si se deseara formar 3 grupos independientes, se tendrían que

[3]Barragán L. R. Hernández, A. Ávila yM.A. Ávila, Cartografía, Encuesta Nacional de Alimentaciíon en el Medio Rural 1996, Instituto Nacional de Nutrición "Salvador Zubirán", México D.F., 1997.

seleccionar 3 muestras con el mismo diseño, donde los conglomerados son seleccionados con reemplazo, así como las viviendas encuestadas en cada conglomerado. No está por demás decir que el número de conglomerados por estrato debe ser igual en las 3 muestras. Además debe quedar claro que la decisión de aplicar este método para cálculo de varianzas se toma en la etapa de planeación y diseño, pues se trata de contar con varias muestras independientes. Como consecuencia resulta evidente que el costo del levantamiento de la encuesta se puede incrementar. Sin embargo, puede ocurrir que esas 3 muestras no se levanten a un mismo tiempo. Pudiera ser que, con objeto de evaluar cambios en la condición nutricional de un estado, éstas se levanten en diferentes períodos; en tal situación, las muestras son independientes y es posible el análisis comparativo entre ellas.

No.	Estrato	Localidades en estrato	Locs. encuestadas
1	HUEJUCAR	32	3
2	AMECA	35	3
3	TOMATLAN	34	3
4	HUERTALA	45	3
5	CHIQUILISTAN	36	3
6	TAMAZULA DE GORDIANO	41	3
7	TLAJOMULCO	34	3
8	IXTLAHUACAN DEL RIO	35	2
9	BARCALA	33	3
10	ATOTONILCO EL ALTO	38	3
11	TEPENTITLAN DE MORELOS	32	3
12	LAGOS DE MORENO	39	3

Tabla 2.2: Estratificación del Estado de Jalisco en la ENAL'96.

Una situación muy distinta se daría si se cuenta con la muestra total, como aparece en la tabla 2.2, y se decide formar 3 grupos dependientes. En todos los estratos se encuestaron 3 conglomerados, excepto en Ixtlahuacan del Río (8), donde sólo se levantó la encuesta en dos localidades. Si simplemente se piensa en asignar una localidad de cada estrato por grupo, ocurre que uno de los grupos tendría un estrato sin datos, lo cual no es deseable. Se puede pensar en hacer dos grupos, y según la regla (iv) del Apéndice A, existe la opción de eliminar un conglomerado en cada uno de los 11 estratos restantes; o bien considerar que en cada grupo cerca de la mitad de los estratos tenga dos conglomerados y en el otro grupo, se cuente con una sola localidad. Posiblemente, ninguna de las alternativas sean satisfactorias, pues con cualquiera de ellas los diseños difieren bastante. Una opción para este caso (que no contempla Wolter) es el colapsar desde un inicio el estrato no. 8 con el más parecido entre estratos vecinos. De esta manera, es factible formar 3 grupos dependientes, basados en 11 estratos, de los cuales 10 de ellos tienen una localidad encuestada en cada grupo y el restante (que correspondería al no. 8 junto con el que se colapsó) incluiría en dos grupos dos localidades, y en el tercer grupo, una localidad. Esta alternativa parece dar los grupos más parecidos pero es obvio que no permite calcular la varianza dentro de estratos por el hecho de que de manera general, sólo hay una localidad por estrato en los grupos.

El ejemplo anterior sirvió para evidenciar que las reglas dadas por Wolter son de mucha ayuda, pero siempre acompañadas del criterio del estadístico, pues no resuelven todas las posibles situaciones que se pueden dar en la realidad.

2.1.3 Teoría entorno a grupos aleatorios

Las inferencias sobre Θ, cuando se calcula la varianza por el método de *muestras independientes*, se basan en el Teorema Central del Límite y en la distribución *t de Student*. Esto implica que se suponen $\Theta_1,\ldots,\Theta_k$, no sólo independientes, sino también que provienen de una distribución Normal (Θ, σ^2). Aunque la normalidad no se cumple exactamente, la teoría asintótica en muestreo, puede justificar el supuesto. Llama la atención igualmente, el que ahora se requiere suponer que las Θ_i se distribuyen idénticamente, con misma media y varianza, aun cuando la caracterización del estimador no contempla igualdad de varianza. Por lo tanto, si se opta por grupos aleatorios independientes, pero con distinto diseño, no sería factible el realizar inferencias sobre el parámetro Θ.

De manera general, cuando se estima la varianza del estimador de Θ mediante grupos aleatorios independientes, construidos bajo el mismo diseño, se considera el siguiente intervalo para Θ:

$$\left(\overline{\hat{\Theta}} - t_{k-1,\alpha/2}\sqrt{v(\overline{\hat{\Theta}})}\ ,\ \overline{\hat{\Theta}} + t_{k-1,\alpha/2}\sqrt{v(\overline{\hat{\Theta}})}\right),$$

donde $t_{k-1,\alpha/2}$ es el cuantil de la distribución t, con $k-1$ grados de libertad, y una probabilidad de obtener un valor mayor a él de $\alpha/2$. Con la misma base teórica se pueden realizar pruebas de hipótesis sobre Θ.

Claro está que cuando las muestras son dependientes, no se puede hacer inferencia sobre el parámetro Θ, ya que la misma construción de las muestras invalida el supuesto más elemental. Por otra parte, sí se cumple el supuesto de que los k estimadores son idénticamente distribuidos, pues se conserva el mismo diseño en todos los grupos. Cabría la pregunta de cuán robusto puede ser el intervalo dado por la distribución t, pero al respecto no se encontró nada en la literatura. Sin embargo, se tienen referencias por comunicaciones personales, de que éste método ha sido ampliamente utilizado por organismos internacionales durante décadas.

2.2 Repeticiones balanceadas

2.2.1 Caracterización del estimador

La técnica de Repeticiones Balanceadas ("Balanced Repeated Replication") también se conoce como *Muestreo por mitades* ("Half-sampling"), *Muestreo por mitades balanceadas* ("Balanced Half-samples), *Muestreo fraccional balanceado*, *Pseudorepetición* ("Pseudoreplication") o Método de semimuestras reiteradas (Mirás, 1985). Los inicios de esta metodología se dan en trabajos aplicados de muestreo, no en el ámbito teórico como ocurrió con el *jackknife* o el *bootstrap*. Aún así, está muy relacionado con estos métodos y con el de Grupos Aleatorios; como se mencionó en la introducción de la tesis, las congruencias entre estos métodos no se deben a una evolución de las técnicas sobre una misma línea, sino que se han encontrado tras análisis conjunto.

El lector que haya revisado el tema de grupos aleatorios en esta tesis, puede imaginar que la técnica de repeticiones balanceadas se debe también a Mahalanobis, debido a que el llamó a sus primeros diseños que contemplaban dos sub-muestras *"half-sampling"*. La coincidencia de nomenclatura parece deberse a la adopción del término por otros autores una vez que idearon una técnica a la que en realidad, le iba mejor dicho nombre. Sin embargo, aunque los diseños de Mahalanobis bien pudieron ser el preámbulo para la metodología que plantearon, no se menciona en la literatura revisada ninguna referencia de los iniciadores de el Muestreo por Mitades hacia Mahalanobis.

Autores como Wolter (1985) y Judkins (1987) señalan que el método tuvo sus orígenes en la década del '50, en el Departamento del Censo de Estados Unidos de América, en trabajos de W. N. Hurwitz, M. Gurney y colaboradores. Posteriormente la técnica fue formalizada por McCarthy en 1966. Luego, estudios basados en simulaciones (Kovar, 1985; Hansen y Tepping, 1985) reportaron que el estimador de la varianza jackknife se comportaba mejor que el de Repeticiones Balanceadas en cuanto a sesgo y estabilidad. De ahí que Robert Fay (Dippo, Fay y Morganstein, 1984), del Departamento del Censo de Estados Unidos, planteara una modificación a la técnica de repeticiones balanceadas que mejora su estabilidad.

Se abordará ahora la técnica repeticiones balanceadas en su forma simple y posteriormente se comentará la adecuación de Fay. El método se aplica al caso en que se tienen L estratos con dos unidades primarias de muestreo en cada uno. Es decir, el tamaño de muestra en cada estrato es 2, $n_h = 2$, y el número total de observaciones [4] es $n = \sum_{i=1}^{L} n_h = 2L$. Bajo este diseño de muestreo, si se fuera a implementar la metodología de grupos aleatorios **dependientes**, ocurriría que sólo se podrían formar dos grupos ($k = 2$), que conservaran la misma estructura de diseño, los cuales se podrían denominar por:

[4] Cabe la posibilidad de que estas observaciones se refieran a conglomerados, que de hecho es lo más usual. La sección 3.2.6 se refiere a esta situación

$(x_{11}, x_{21}, ..., x_{L1})$ y $(x_{12}, x_{22}, ..., x_{L2})$.

Interesa estimar la varianza de un estimador $\hat{\Theta}$, que es de la forma $\hat{\Theta} = g(\overline{X})$, por lo que al abordar el caso de $\overline{x}$, se sigue la generalización para otro $\hat{\Theta}$. Como referencia, se debe recordar que la media global bajo un diseño estratificado, considerando $n_h = 2$ (h=1,2,...,L), N_h es el tamaño del estrato h y $N = \sum_{h=1}^{L} N_h$, se obtiene de:

$$\begin{aligned} \overline{x}_{est} &= \sum_{h=1}^{L} W_h \overline{x}_h && (2.4) \\ \text{donde, } \overline{x}_h &= \frac{(x_{h1} + x_{h2})}{2} && (2.5) \end{aligned}$$

$$\text{y } W_h \text{ son los pesos de los estratos, } \quad W_h = \frac{N_h}{N} \qquad (2.6)$$

También conviene recordar que el estimador de la varianza de $\overline{x}_{est}$, en un muestreo estratificado simple, si la fracción de muestreo en cada estrato, n_h/N_h, es despreciable, se expresa como:

$$\begin{aligned} v(\overline{x}_{est}) &= \sum_{h=1}^{L} W_h^2 \frac{s_h^2}{2} \\ &= \sum_{h=1}^{L} \frac{W_h^2}{2} \left\{ \left[x_{h1} - \frac{(x_{h1} + x_{h2})}{2}\right]^2 + \left[x_{h2} - \frac{(x_{h1} + x_{h2})}{2}\right]^2 \right\} \\ &= \sum_{h=1}^{L} W_h^2 \frac{(x_{h1} - x_{h2})^2}{4} \\ &= \sum_{h=1}^{L} W_h^2 \frac{d_h^2}{4} && (2.7) \\ \text{donde, } d_h &= x_{h1} - x_{h2} \end{aligned}$$

Bajo el método de Grupos Aleatorios, se obtienen dos estimadores, $\widehat{\Theta}_1$ y $\widehat{\Theta}_2$, obviamente considerando los pesos W_h $(h = 1, 2, ...L)$ de los L estratos, de la siguiente manera:

$$\overline{x}_1 = \sum_{h=1}^{L} W_h x_{h1} \text{ ,de donde se obtiene, } \widehat{\Theta}_1 = g(\overline{x}_1).$$

De igual forma se calcula el estimador del segundo grupo, obteniendo $\overline{x}_2$ y $\widehat{\Theta}_2$. Aplicando (2.1) es inmediato que el estimador global sólo se basaría en un grado de libertad, lo que le daría poca estabilidad. La idea tras el método de Muestreo por Mitades, es considerar, *medias-muestras* compuestas por una unidad por estrato. Esto significa que se permite que las muestras se traslapen, pues se contemplan muestras con elementos comunes. De tal suerte, existen 2^L posibles *medias-muestras*, las cuales representan una forma de pseudo-repetición o pseudo-remuestreo.

De cada media-muestra se deriva un estimador de la media, $\overline{x}_i$, el cual se exhibe a continuación.

$$\overline{x}_i = \sum_{i=1}^{L} W_h(\delta_{h1i}x_{h1} + \delta_{h2i}x_{h2}) \tag{2.8}$$

$$\text{Donde } \delta_{h1i} = \left\{ \begin{array}{lll} 1 & \text{si la unidad (h,1)} \in & \text{la muestra i-ésima,} \\ 0 & \text{si la unidad (h,1)} \notin & \text{la muestra i-ésima.} \end{array} \right\}$$

$$\delta_{h2i} = 1 - \delta_{h1i}. \tag{2.9}$$

En el caso de un estimador lineal, si se consideran las 2^L estimaciones, se tiene que el promedio de ellas es igual al estimador de un diseño estratificado; lo cual se deriva del hecho de que cada unidad se encuentra en la mitad de las 2^L muestras. En ocasiones el número de estratos, L, puede ser muy grande. Por lo tanto, resulta lógico que se considere sólo una fracción de esas 2^L posibles medias-muestras; digamos k de ellas. Intuitivamente el lector ya puede darse cuenta de que es posible construir un estimador de la varianza de $\overline{x}_{est}$, mediante el promedio del cuadrado de las diferencias de los $\overline{x}_i$ obtenidos (i=1,2,...,k), con respecto a $\overline{x}_{est}$. Tal suposición es cierta, sin embargo, al inspeccionar un poco el estimador se aprecia que no se deben considerar cualesquiera k medias-muestras. Aquí es donde entra la colaboración de McCarthy (1966), puntualizando que se deben escoger k muestras *balanceadas*. A partir de este argumento, se entiende por el estimador de Muestreo por Mitades ("Half-Sampling"), el que se basa en 2^L muestras y sus respectivas estimaciones y por el estimador de repeticiones balanceadas ("Balanced Repeated Replication"), el que se basa en k medias-muestras balanceadas.

Antes de explicar qué son muestras balanceadas, es necesario examinar un poco el estimador en cuestión. De acuerdo a lo dicho en el párrafo anterior, se propone como estimador de la varianza de la media, basado en k muestras:

$$v_k(\overline{x}_{est}) = \sum_{i=1}^{k} \frac{(\overline{x}_i - \overline{x}_{est})^2}{k}. \tag{2.10}$$

Y análogamente, para cualquier $\widehat{\Theta}$ en general,

$$v_k(\hat{\Theta}) = \sum_{i=1}^{k} \frac{(\hat{\Theta}_i - \hat{\Theta})^2}{k}. \tag{2.11}$$

$\hat{\Theta}_i$ es el estimador obtenido de la muestra i-ésima, aplicando (2.8). Por $\widehat{\Theta}$ se entiende el estimador que se obtiene por las fórmulas usuales para diseño estratificado (Considerando $2L$ observaciones), o bien, $\widehat{\Theta} = g(\overline{x}_{est})$. Cabe señalar que en ocasiones se construye (2.11) con base en las desviaciones hacia

$$\widehat{\Theta}_{(.)} = \sum_{j=1}^{k} \frac{\widehat{\Theta}_j}{k}\ . \tag{2.12}$$

y no con respecto a $\widehat{\Theta}$.

Al examinar el caso de la varianza de una estadística lineal, $\hat{\Theta}_{lin}$, se encuentra el argumento de McCarthy para definir y requerir *muestras balanceadas*; por cuanto lo que sigue de esta discusión, se enfoca hacia este tipo de estimador. De manera general, una estadística lineal en la media muestral se expresa como [5]:

$$\begin{aligned}
\widehat{\Theta}_{lin} &= \sum_{h=1}^{L} \left\{ \mu_h + \frac{1}{n_h} \sum_{j=1}^{n_h} W_h x_{hj} \right\} \\
&= \sum_{i=1}^{L} \left\{ \mu_h + W_h \overline{x}_h \right\}
\end{aligned} \tag{2.13}$$

A partir de las variables indicadoras δ_{h1i} y δ_{h2i} definidas en 2.2.1, se puede construir otra variable indicadora, $\delta_h^{(i)}$, que corresponde a la

[5]Ver Efron (1982), pag.61.

muestra i-ésima, de la siguiente forma:

$$\delta_h^{(i)} = 2\delta_{h1i} - 1. \tag{2.14}$$

Así pues,

$$\delta_h^{(i)} = \left\{ \begin{array}{rl} 1 & \text{si la unidad (h,1) está en la muestra iésima} \\ -1 & \text{si la unidad (h,2) está en la muestra iésima} \end{array} \right\}$$

Ahora bien, cuando se considera la media-muestra i-ésima, la media del estrato h es igual al valor de la unidad incluida en ésta, es decir,

$$\overline{x}_h^i = \frac{x_{h1}(1 + \delta_h^{(i)}) + x_{h2}(1 - \delta_h^{(i)})}{2} \tag{2.15}$$

Si se hace la substitución de (2.15) en (2.13), y se desarrolla un poco el álgebra, se ve que el estimador lineal de cualquiera de las medias-muestras ($i = 1, 2, ..., 2^L$) es igual a (2.13) más otro término que depende de las $\delta_h^{(i)}$; lo que se desglosa aquí:

$$\begin{aligned} \widehat{\Theta}_{i_{lin}} &= \sum_{h=1}^{L} \left\{ \mu_h + \frac{W_h}{2}\Big[x_{h1}(1 + \delta_h^{(i)}) + x_{h2}(1 - \delta_h^{(i)})\Big] \right\} \\ &= \sum_{h=1}^{L} \left\{ \mu_h + \frac{\sum_{j=1}^{2} W_h x_{hj}}{2} + \delta_h^{(i)} \frac{x_{h1} - x_{h2}}{2} \right\} \\ &= \widehat{\Theta}_{lin} + \sum_{h=1}^{L} \delta_h^{(i)} W_h \frac{x_{h1} - x_{h2}}{2} \end{aligned} \tag{2.16}$$

Con estos elementos, (2.11) y (2.16), ya se puede desarrollar la varianza de $\widehat{\Theta}_{lin}$, a partir de k muestras.

$$\begin{aligned} v\{\widehat{\Theta}_{lin}\}_k &= \frac{1}{k}\sum_{j=1}^{k}(\hat{\Theta}_{i_{lin}} - \hat{\Theta}_{lin})^2 \\ &= \frac{1}{k}\sum_{j=1}^{k}\Big[\sum_{h=1}^{L} \delta_h^{(i)} W_h \frac{(x_{h1} - x_{h2})}{2}\Big]^2 \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{k}\sum_{j=1}^{k}\sum_{h=1}^{L} W_h^2 \frac{(x_{h1}-x_{h2})^2}{4} \\
&+ \frac{1}{k}\sum_{j=1}^{k}\sum_{h=1}^{L}\sum_{h'\neq h} \delta_h^{(j)}\delta_{h'}^{(j)} W_h W_{h'} \frac{(x_{h1}-x_{h2})(x_{h'1}-x_{h'2})}{4}
\end{aligned} \tag{2.17}
$$

El primer término en (2.17) se reduce a

$$
\sum_{h=1}^{L} W_h^2 \frac{(x_{h1}-x_{h2})^2}{4},
$$

lo que, si se compara con (2.7), se aprecia que es igual a la varianza que se obtendría mediante las fórmulas conocidas para muestreo estratificado. Por otra parte, el segundo término puede ser muy grande para ciertos conjuntos de medias-muestras, pero se hace cero si ocurre que

$$
\sum_{j=1}^{k} \delta_h^{(j)}\delta_{h'}^{(j)} = 0 \quad (\text{para } 1 \leq h < h' \leq L). \tag{2.18}
$$

El método de McCarthy busca la manera de obtener para una estadística lineal, la misma varianza que se obtendría a través de la fórmula (2.7), pero con k muestras, donde $k < 2^L$. Para tal efecto, el método de *repeticiones balanceadas* se basa en cualquiera de las k muestras de las posibles 2^L *medias-muestras*, que cumplan (2.18). Por supuesto, el conjunto de 2^L medias-muestras satisface tal requerimiento, ya que cada unidad se encuentra en 2^{L-1} medias-muestras [6], pero la ventaja del método consiste en necesitar un número pequeño de pseudo-repeticiones.

Antes se mencionó la posibilidad de calcular la varianza considerando las desviaciones con repecto a $\hat{\Theta}_{(.)}$, dado en (2.12). Si así se hiciera y la estadística en cuestión fuera lineal, se observaría que

$$
\begin{aligned}
\hat{\Theta}_{lin_{(.)}} &= \sum_{j=1}^{k} \frac{\hat{\Theta}_{j_{lin}}}{k} \qquad (2.19) \\
&= \hat{\Theta}_{lin} \qquad (2.20)
\end{aligned}
$$

[6] Es decir, $\sum_{j=1}^{2^L} \delta_h^{(j)}\delta_{h'}^{(j)} = \underbrace{\sum}_{j\ni\delta_h^{(j)}=1} \delta_{h'}^{(j)} - \underbrace{\sum}_{j\ni\delta_h^{(j)}=-1} \delta_{h'}^{(j)} = 0$

si se satisface el siguiente requisito por las medias-muestras elegidas:

$$\sum_{j=1}^{k} \delta_h^{(j)} = 0. \text{ (h=1,2,...,L)}. \tag{2.21}$$

Lo cual se deduce claramente al expandir $\hat{\Theta}_{lin_{(.)}}$;

$$\hat{\Theta}_{lin_{(.)}} = \hat{\Theta}_{lin} + \frac{1}{k}\sum_{h=1}^{L} W_h \frac{x_{h1} - x_{h2}}{2} \sum_{j=1}^{k} \delta_h^{(j)}.$$

Se dice que las medias-muestras poseen *balance ortogonal completo* si además de satisfacer (2.18), cumplen también con (2.21). No hay que olvidar que todas estas definiciones y características se atribuyen al caso de una estadística lineal en la media. Es concebible el que se busquen estimadores para estadísticas no lineales que no se alteren, al aplicarlos a casos lineales. Pero lo cierto es que aún utilizando medias-muestras balanceadas, para el caso no lineal no se garantiza ninguna propiedad de las mencionadas; en particular, el promedio de las estimaciones $\hat{\Theta}^{(i)}$ $(i = 1, 2, ..., k)$, no será igual al estimador global obtenido por reglas de muestreo estratificado, basado en $2L$ observaciones; sobre la varianza, resulta obvio el equivalente del argumento anterior, ya que si se pudiera calcular el estimador de la varianza mediante las reglas generales de muestreo, en el caso no lineal, no tendría sentido ni éste ni ningún método. Sin embargo, sí existe una buena razón para considerar medias-muestras balanceadas, pues Krewski y Rao (1981) demostraron resultados asintóticos que permiten hacer inferencia basada en la varianza obtenida de ellas.

2.2.2 Modificación Fay

El estimador de la varianza de repeticiones balanceadas, probó ser muy inestable al aplicarse a estadísticas no lineales. En particular, se detectó que el estimador es problemático al calcular la varianza de una razón, ya que el denominador puede ser cero o cercano a cero para algunas medias-muestras. Si se inspecciona (2.15), se aprecia que tácitamente, se atribuye a las unidades dentro de cada estrato un peso de 0 ó 2. Robert Fay, (Dippo, Fay y Morganstein, 1984) sugirió cambiar tales pesos para evitar los problemas mencionados. Su idea original era usar ponderaciones de 0.5 y 1.5, pero hizo el planteamiento general. En vez de los pesos $(1+\delta_h^{(i)})$ y $(1-\delta_h^{(i)})$, como se observan en 2.15, las unidades de los estratos se ponderan por $(1+\delta_h^{(i)}(1-\lambda))$ y $(1-\delta_h^{(i)}(1-\lambda))$; de tal suerte, se considera el siguiente estimador de la media de un estrato h $(h=1,2,...,L)$:

$$\overline{x}_{hFay}^{(i)} = \frac{x_{h1}[1+\delta_h^{(i)}(1-\lambda)] + x_{h2}[1-\delta_h^{(i)}(1-\lambda)]}{2}. \tag{2.22}$$

El estimador de la varianza consiste en las desviaciones del estimador *Fay* con respecto al estimador de muestreo estratificado. En otras palabras, a partir de (2.22) se consigue $\overline{x}_{Fay} = \sum_{h=1}^{L} W_h \overline{x_h}_{Fay}$; luego, se busca $\hat{\Theta}_{Fay}^{(j)} = g(\overline{x}_{Fay})$ $(j=1,2,...,k)$, para finalmente calcular

$$v_{Fay}(\hat{\Theta}) = \sum_{j=1}^{k} \frac{(\hat{\Theta}_{Fay}^{(i)} - \hat{\Theta})^2}{k(1-\lambda)^2}. \tag{2.23}$$

Nótese que si $\lambda = 0$, entonces la estimación de Fay coincide con la de *repetiones balanceadas*, presentada anteriormente, tanto en la estimación de la media, como en la varianza. Igualmente debe llamar la atención la incorporación del término $\frac{1}{(1-\lambda)^2}$ en el estimador de la varianza. Al respecto, Judkins (1987, pág. 492) explica que el error cuadrático medio disminuye en la medida en que λ se acerca a la unidad, pero se transforma en un estimador razonable de varianza al multiplicarse por dicho factor. En el mismo artículo, Judkins hace ver que se puede pensar que $100(1-\lambda)$ es un factor de *perturbación*. O sea, en una aplicación de repeticiones balanceadas sin la adecuación de Fay, la perturbación es del 100%, pero si se ejecuta la modificación con $\lambda = 0.9$, la perturbación es del 10%. La determinación óptima de λ no ha sido dada por ningún autor. Los resultados de simulaciones de repeticiones balanceadas, adecuación de Fay y jackknife para razones de variables normales, de Judkins (1987, pág. 493), mostraron que cuando los errores estándares de los estratos son pequeños, todos los métodos son esencialmente equivalentes, pero si las variabilidad dentro de los estratos es grande y λ se acerca a 0, el estimador de la varianza es muy malo. Además ocurre que, sin importar la varianza de los estratos, la modificación de Fay y el jackknife convergen en la medida en que λ se acerca a la unidad. Por otra parte, dichas simulaciones de Judkins (1987) revelan que la modificación Fay conserva las características del caso lineal y reduce el sesgo y la varianza del estimador de la varianza de una razón.

2.2.3 Otros estimadores para estadísticas no lineales

Una peculiaridad de esta técnica es que existen varios estimadores para el caso no lineal. Para empezar, es necesario señalar que por cada conjunto de medias-muestras balanceadas, se tiene directamente otro que también es balanceado; éste es aquél cuyos indicadores $\delta_h^{(j)}$ ($j = 1, 2, ..., k$ y $h = 1, 2, ..., L$) son el negativo del primero, razón por la cual se dice que es su *complemento.* Si el estimador es lineal, cualquier conjunto de medias-muestras balanceadas va a dar el mismo estimador, por lo que no tiene caso analizar dos conjuntos balanceados; pero cuando el estimador no es lineal, se obtienen resultados diferentes.

Sea $\hat{\Theta}_{(j)}^c$ el complemento de la media-muestra i-ésima. En seguida se deducen otras formas de calcular la varianza, tal como se aprecia aquí:

$$v_k^c(\hat{\Theta}) = \sum_{j=1}^{k} \frac{(\hat{\Theta}_{(j)}^c - \hat{\Theta})^2}{k} \tag{2.24}$$

$$\overline{v}_k(\hat{\Theta}) = \frac{v_k^c(\hat{\Theta}) + v_k(\hat{\Theta})}{2} \tag{2.25}$$

Wolter (1985) es quien presenta éstos y otros estimadores para el caso no lineal. Su libro no abarca la modificación de Fay que se discutió en la sección anterior. Sin embargo, es claro que si se utiliza el estimador Fay, se pueden calcular $v_k{}^c(\hat{\Theta})$ y $\overline{v}_k(\hat{\Theta})$, de la siguiente forma:

$$v_{k_{Fay}}^c(\hat{\Theta}) = \sum_{j=1}^{k} \frac{(\hat{\Theta}_{(j)_{Fay}}^c - \hat{\Theta})}{k(1-\lambda)} \tag{2.26}$$

$$\overline{v}_{k_{Fay}}(\hat{\Theta}) = \frac{v_{k_{Fay}}^c(\hat{\Theta}) + v_{k_{Fay}}(\hat{\Theta})}{2} \tag{2.27}$$

2.2.4 Consideraciones prácticas

Elección de las muestras balanceadas

Dada la descripción del estimador de repeticiones balanceadas, el primer problema práctico al que se enfrenta el estadístico que quiere aplicar esta técnica, es la selección de un conjunto de medias-muestras balanceadas. En primer lugar se pregunta qué valor de k es factible considerar, y luego se enfrenta a los requerimientos (2.18) y (2.21).

Esos cuestionamientos encontraron una solución en el trabajo de Plackett y Burman de 1946 (citado por Wolter, 1985), que aunque se trata de diseños factoriales 2^n, fue igualmente útil para este problema de muestreo. Ellos construyeron matrices ortogonales, de dimensión m x m, donde m es múltiplo de 4, y sus entradas son 1 ó -1. (Éstas son conocidas en matemáticas como matrices *Hadamard*).

Las columnas de estas matrices satisfacen (2.18), por lo que se identifican los estratos con las columnas y los renglones con las repeticiones (medias-muestras). A continuación se ofrecen advertencias en cuanto a la determinación de k, dado un diseño basado en L estratos:

a.) Se tiene un **balance ortogonal completo** si

$$k > L \ , \ k \text{ múltiplo de } 4$$

b.) Se tiene **balance**, pero no balance ortogonal completo si

$$k = L, \ \text{k y L múltiplos de } 4$$

c.) No existe balance alguno si

$$k < L$$

La recomendación más evidente es pues, escoger k como el valor más pequeño que satisface el inciso (a.). De esta manera se consigue además minimizar los cálculos. Por otra parte, es necesario basar la selección de las medias-muestras en la matriz Hadamard de dimensión k x k, una vez que se haya elegido el número de repeticiones.

Muestreo estratificado bietápico

Un aspecto importante a considerar es la generalización de esta técnica al caso en que las dos unidades muestreadas en cada estrato son conglomerados, de lo que se sigue otra etapa de muestreo. Se puede revisar en Cochran (1977) o Sukhatme (1984), el estimador insesgado para la media bajo este diseño, el cual está dado por:

$$\overline{x}_{\mathrm{BI}} = \sum_{h=1}^{L} W_h \frac{1}{\overline{\overline{M}}_h} \frac{1}{2} \sum_{i=1}^{2} M_{hi} \overline{x}_{hi} \tag{2.28}$$

Ahora el estimador lleva el subíndice "BI" para recordar que se aplica al muestreo bietápico. Por otra parte, a continuación se aclara la nomenclatura de (2.28).

- $W_h = \frac{N_h}{N}$, donde N_h es el número de conglomerados del estrato h $(h = 1, 2, \ldots, L)$, N es el total de conglomerados, es decir, $N = \sum_{h=1}^{L} N_h$.

- M_{hi} es el tamaño del conglomerado i-ésimo $(i = 1, 2)$ en el estrato h.

- $\overline{M}_h$ es el tamaño promedio de los conglomerados del estrato h. Sea $M_{h0} = \sum_{i=1}^{N_h} M_{hi}$, entonces,

$$\overline{M}_h = \frac{M_{h0}}{N_h}.$$

- $\overline{x}_{hi}$ es la media muestral del conglomerado i-ésimo $(i = 1, 2)$ en el estrato h. Es decir, si en el conglomerado hi se muestrean m_{hi} elementos,

$$\overline{x}_{hi} = \frac{1}{m_{hi}} \sum_{j=1}^{m_{hi}} x_{hij}.$$

Para aplicar la forma general para calcular la varianza de repeticiones balanceadas, (2.10) y (2.11), se debe comenzar por revisar el estimador que se basa en k medias-muestras balanceadas. Se supone que δ_h^j se define como antes y que se cumple (2.18). La media estimada por la media-muestra número j se expresa como:

$$\begin{aligned}\overline{x}_{\mathrm{BI}}^{(j)} &= \sum_{h=1}^{L} W_h \frac{1}{\overline{M}_h} \frac{M_{h1}\overline{x}_{h1}(1+\delta_h^{(j)}) + M_{h2}\overline{x}_{h2}(1-\delta_h^{(j)})}{2} \\ &= \overline{x}_{\mathrm{BI}} + \sum_{h=1}^{L} \frac{W_h}{\overline{M}_h} \delta_h^{(j)} \frac{M_{h1}\overline{x}_{h1} - M_{h2}\overline{x}_{h2}}{2} \end{aligned} \tag{2.29}$$

Con el estimador de la varianza ocurre algo similar a lo visto en el caso de muestreo estratificado simple.

$$v_{\mathrm{BI}}^{k} = \sum_{j=1}^{k} \frac{(\overline{x}_{\mathrm{BI}}^{(j)} - \overline{x}_{\mathrm{BI}})^2}{k} = \frac{1}{k} \sum_{j=1}^{k} \sum_{h=1}^{L} \frac{W_h^2}{\overline{M}_h^2} \frac{M_{h1}\overline{x}_{h1} - M_{h2}\overline{x}_{h2}}{4} \tag{2.30}$$

Más adelante, en 2.2.5. se explica cómo se puede llevar a cabo la aplicación de repeticiones balanceadas para estimar la varianza, de una forma directa, sin importar lo complicado del diseño o el estimador. También se hacen algunos comentarios sobre la posibilidad de aplicar esta técnica en diseños que no se ajustan exactamente al planteado, el cual se basa en dos conglomerados por estrato.

2.2.5 Generalizaciones

Interesa saber, cuán complicada es la aplicación de Repeticiones balanceadas para obtener estimadores no lineales en muestreo complejo. Rust y Rao (1996) dan un procedimiento con el que se resuelve cualquier estimador bajo cualquier diseño. Se consideran los factores de expansión para cada unidad elemental muestreada; por ejemplo, en un muestreo bietápico estratificado por conglomerados, se puede hablar de w_{hij}, el factor de la unidad j-ésima, del conglomerado i-ésimo del estrato h. Cada uno de estos pesos se altera de la siguiente forma [7] :

$$w_{hij}^{(k)} = \left\{ \begin{array}{rl} (2-\lambda)w_{hij} & \text{Si el conglomerado i-ésimo del estrato } h \\ & \text{está en la media-muestra } k \\ (\lambda)w_{hij} & \text{en otro caso.} \end{array} \right\}$$

Posteriormente se obtiene el estimador $\hat{\Theta}_i$, para cada una de las medias-muestras (i=1,...,k), con los pesos ajustados. Por ejemplo, si la

[7] Se sigue la notación de la sección 3.2.2

estadística de interés es un vector de coeficientes de regresión, β, en el modelo aplicado a los datos de una muestra:

$$y_{\mathbf{S}} = X_{\mathbf{S}}\beta + \epsilon_{\mathbf{S}} \ ,$$

donde $y_{\mathbf{S}}$ es el vector de respuestas en la muestra, $X_{\mathbf{S}}$ es la matriz de variables independientes o auxiliares y $W_{\mathbf{S}}$ es una matriz diagonal que contiene los factores de expansión de cada unidad elemental muestreada; entonces el estimador de β es [8]:

$$\hat{\beta} = (X'_{\mathbf{S}} W_{\mathbf{S}} X_{\mathbf{S}})^{-1} X'_{\mathbf{S}} W_{\mathbf{S}} y_{\mathbf{S}}$$

Quiere decir, que en cada media-muestra se obtendría un $\beta^{(k)}$ ajustando la matriz de pesos,$W_{\mathbf{S}}$ a $W_{\mathbf{S}}^{(k)}$, siguiendo la misma regla. Luego se obtiene la varianza de $\hat{\beta}$ mediante la expresión de la varianza que considera la modificación Fay, (2.23). Lo mismo se haría para cualquier otro estimador. Se debe apreciar que resulta mucho más fácil el calcular un mismo estimador muchas veces, que encontrar y proporcionar las fórmulas adecuadas al personal que se encarga del cómputo en el análisis de una encuesta.

Se puede decir que una debilidad de esta técnica es el hecho de que se enfoque a casos donde el tamaño de muestra en el estrato es $n_h = 2$, pues esta situación limita sus posibilidades de aplicación. Sin embargo, la revisión bibliográfica revela que este procedimiento es muy estudiado y utilizado en el Departamento del Censo de Estados Unidos. L.R.Ernst y T.R. Williams (1987) presentan un estudio muy singular, donde se colapsan estratos, de manera que pueden aplicar la técnica de repeticiones balanceadas.

El colapso de estratos consiste en unir o *colapsar* dos o más estratos en uno, de manera que se crea una nueva partición del conjunto de datos. Usualmente esta práctica se lleva a cabo cuando se tienen estratos con una sola unidad primaria de muestreo, pues en tal situación no se suele conseguir un estimador insesgado ni consistente de la varianza, ni siquiera para estadísticas lineales.

En esta tesis no se quiere entrar en detalles sobre el estimador de estratos colapsados, sino solamente mostrar que existen alternativas

[8]Ver a Kott (1994)

de acción ante un problema práctico, como el que presentan Ernst y Williams (1987). Para mayor información sobre este estimador se puede consultar a dichos autores y Wolter (1985, pags. 47- 54), o Cochran (1977).

Capítulo 3

Métodos de estimación de la varianza surgidos en el marco de poblacion infinita

En este capítulo se presentan los métodos de cálculo de varianza que surgieron en el marco de población infinita, los cuales son: *Linearización, jackknife y bootstrap.* La linealización por series de Taylor no se basa en el remuestreo pero ha sido una técnica muy difundida desde mediados de los setentas, ya que se puede aplicar a través de varios paquetes de cómputo. Por otra parte, las investigaciones sobre la aplicación del jackknife y del bootstrap en muestreos complejos se comienzan a dar a mediados de los ochentas; y no es hasta los noventas que algunas instituciones adoptan alguno de estos dos métodos formalmente. La referencia cronológica explica el por qué las investigaciones dirigidas a la comparación de métodos de cálculo de varianza (que se encontraron en la bibliografía), exploran las nuevas técnicas contrastándolas con la Linearización. Por tal motivo, se consideró necesario presentar su desarrollo, junto con el jackknife y el bootstrap, los cuales son métodos de remuestreo que están cobrando auge en la solución de problemas de muestreos complejos, gracias al gran desarrollo de la computación.

3.1 Linealización

El método de linealización por series de Taylor es quizás el más utilizado en aplicaciones con encuestas de tamaño considerable, donde se hace necesario el uso de paquetes computarizados de difusión comercial. Por ejemplo, el Instituto Nacional de Estadística Geografía e Informática (INEGI), reporta en el ENAID que obtuvo las varianzas mediante el programa *Clusters*, el cual usa este método.

El método consiste en expresar un parámetro no-lineal, Θ, como una función de medias de otras variables. La expansión por series de Taylor provee una aproximación lineal de la estadística de interés. De tal forma, se busca la varianza de esa aproximación, lo que representa un estimador sesgado de la estadística no-lineal.

Entre las primeras referencias encontradas respecto al uso de series de Taylor para estimar la varianza de una estadística, se hallan dos trabajos: Woodruff (1971) y Woodruff y Causey (1976). En el primero se advierte la utilidad del método en el caso de muestras complejas, y en el segundo, se describe un programa en el que se ha implementado la aproximación por series de Taylor para calcular errores de muestreo.

3.1.1 Caracterización del estimador

Cuando el parámetro de interés es de la forma $\Theta = f(\mathbf{Y})$, donde $\boldsymbol{Y}$ es un vector p-variado [1] de parámetros de la población, $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_p)$, resulta natural elegir el estimador $\hat{\Theta} = f(\hat{\mathbf{Y}})$. Sin embargo la varianza de tal estimador no se obtiene directamente, en la mayoría de las ocasiones.

Bajo el supuesto de que $f(\hat{\mathbf{Y}})$ posee derivadas de segundo orden en una hiperesfera que contiene a $\hat{\mathbf{Y}}$ y a $\mathbf{Y}$, entonces, al aplicar la expansión de Taylor se tiene:

$$f(\hat{\mathbf{Y}}) = f(\mathbf{Y}) + \sum_{j=1}^{p} \frac{\partial f(\mathbf{Y})}{\partial y_j}(\hat{Y}_j - Y_j) + R_{n_S}(\hat{\mathbf{Y}},\mathbf{Y}) \qquad (3.1)$$

[1] Nótese que $\boldsymbol{Y}$ es una función p-variada de $y = (y_1, ..., y_N)$, lo que de acuerdo a la notación del capítulo 1 representa un vector de valores en la población de tamaño N. Pero ahora se considera $y_i = [y_{i1}, y_{i2}, ..., y_{ip}]'$, es decir, la función de interés es función de varias variables en la población.

$R_{n_S}(\widehat{\mathbf{Y}},\mathbf{Y})$ es el residuo[2], al desechar los términos de segundo orden y más, el cual depende del tamaño de muestra y está dado en términos de los parámetros poblacionales y los estimadores.

El error cuadrático medio del estimador $f(\widehat{\mathbf{Y}})$ está dado por:

$$ECM(f(\hat{\mathbf{Y}})) = E\{(f(\widehat{\mathbf{Y}}) - f(\mathbf{Y}))^2\}$$

Ya se comentó en el capítulo 1 que en ocasiones se considera una aproximación del ECM como la aproximación de la varianza, debido a que el $Sesgo^2(\hat{\Theta})$ es de orden menor a $Var(\hat{\Theta})$. Por lo tanto,

$$\begin{aligned} ECM(f(\hat{\mathbf{Y}})) &\approx Var\{f(\widehat{\mathbf{Y}})\} \\ &= Var\{\sum_{j=1}^{p} \frac{\partial f(\mathbf{Y})}{\partial y_j}(\hat{Y}_j - Y_j)\} \\ &= \sum_{j=1}^{p}\sum_{i=1}^{p} \frac{\partial f(\mathbf{Y})}{\partial y_j}\frac{\partial f(\mathbf{Y})}{\partial y_i} Cov(\hat{Y}_j,\hat{Y}_i) \ . \end{aligned}$$

Así pues, al considerar la matriz de varianza-covarianza de $\hat{\mathbf{Y}}$, $\xi_{n_{(\mathbf{S})}}$, y el vector *1* x p de derivadas, d, cuyos elementos son de la forma $d_j = \frac{\partial f(\mathbf{Y})}{\partial y_j}$, se llega a:

$$ECM(f(\hat{\mathbf{Y}})) \approx d\xi_{n_{(\mathbf{S})}}d' \tag{3.2}$$

p. Por otra parte, el vector de derivadas se consigue evaluando en los valores muestrales de la siguiente forma:

$$\hat{d}_j = \frac{\partial f(\widehat{\mathbf{Y}})}{\partial y_j}$$

Sin embargo Woodruff (1971) advirtió que si las estadísticas $\hat{Y}_j$ están dadas por una combinación lineal de observaciones,

$$\hat{Y}_j = \sum_{i=1}^{n_S} w_i y_{ij} \text{ donde j=1,...,p.}$$

---

[2] n_S es el tamaño de la muestra completa, como se definió en el capítulo 1.

entonces el cálculo se simplifica a un problema univariado. El razonamiento de Woodruff se muestra a continuación:

$$
\begin{aligned}
ECM(\hat{\Theta}) &= Var[\sum_{j=1}^{p} \frac{\partial f(\mathbf{Y})}{\partial y_j} \hat{Y}_j] \\
&= Var[\frac{\partial f(\mathbf{Y})}{\partial y_j} \sum_{i=1}^{n_\mathbf{S}} w_i y_{ij}] \\
&= Var(\sum_{i=1}^{n_\mathbf{S}} w_i \sum_{j=1}^{p} \frac{\partial f(\mathbf{Y})}{\partial y_j} y_{ij}) \\
&= Var(\sum_{i=1}^{n_\mathbf{S}} w_i v_i) \text{ donde, } v_i = \frac{\partial f(\mathbf{Y})}{\partial y_j} y_{ij} \; .
\end{aligned}
$$

La estimación de esas v_i se realiza como se mostró antes en el caso del vector d. Es decir,

$$
\hat{v}_i = \frac{\partial f(\widehat{\mathbf{Y}})}{\partial y_j} y_{ij}
$$

3.1.2 Consideraciones prácticas

Se vió que se puede llegar a la expresión de la varianza de una estadística que es combinación lineal de las observaciones, ponderadas por la derivada de la función original, evaluada en valores muestrales obtenidos de la misma muestra. Lo cual implica, que se debe requerir un estimador insesgado de $Var(y_{ij})$.

Muchas veces, se puede aplicar este método para calcular varianzas casi a ciegas, si se tiene acceso a programas que implementen esta técnica. Sin embargo, aún para estos programas pueden haber ciertas dificultades al buscar las derivadas en cuestión. Ejemplos de situaciones difíciles lo son los coeficientes de correlacion múltiple y parcial. En tales casos pueden ser útiles los métodos descritos en Woodruff y Casey (1976).

Por otra parte, hay que tener presente que los términos en $\xi_{n(\mathbf{S})}$, o bien, la expresión para $Var(\sum_{i=1}^{n_\mathbf{S}} w_i v_i)$, dependerán del diseño de muestreo. Wolter (1985) advierte que es posible que para calcularla, sea necesario aplicar algún otro método para la estimación de la varianza.

Es útil mostrar un ejemplo de la aplicación de esta técnica. Sea $\hat{R}$ el estimador de razón entre dos variables, tal que $\hat{R} = \frac{\overline{y}}{\overline{x}}$. De acuerdo a (3.2), se debe determinar el vector de dos dimensiones d, y la matriz $\xi_{n(\mathbf{s})}$, los cuales son:

$$\hat{\xi}_{n(\mathbf{s})} = \begin{bmatrix} \widehat{Var}(\overline{x}) & \widehat{Cov}(\overline{x},\overline{y}) \\ \widehat{Cov}(\overline{x},\overline{y}) & \widehat{Var}(\overline{y}) \end{bmatrix}$$

$$d = \begin{bmatrix} -\frac{\overline{y}}{\overline{x}^2} \\ \frac{1}{\overline{x}} \end{bmatrix}$$

Por lo tanto, el estimador de la varianza de una razón, de acuerdo al método de linealización, está dado por:

$$\widehat{Var}(\hat{R}) = \left(\frac{\widehat{Var}(\overline{x})}{\overline{x}^2}\hat{R}^2 + \frac{\widehat{Var}(\overline{y})}{\overline{x}^2} - 2\frac{\widehat{Cov}(\overline{x},\overline{y})}{\overline{x}^2}\hat{R} \right) \qquad (3.3).$$

Se puede verificar en Cochran (1977) y otros autores que esta aproximación de la varianza obtendia por linealización, es el estimador que usualmente se proporciona por fórmula. Cuando se conoce $\overline{X}$, el valor poblacional, se usa $\overline{X}^2$ en lugar de $\overline{x}^2$ en (3.3). Es evidente que cada uno de los elementos de $\hat{\xi}_{n(\mathbf{s})}$ se tiene que estimar de acuerdo al diseño muestral, por lo que la evaluación de esta expresión podría hacerse bastante pesada. Cabe señalar que si se conoce el valor poblacional, $\overline{X}$, se usa $\overline{X}^2$ en lugar de $\overline{x}^2$ en (3.3).

3.2 Jackknife

Hasta hace unos años, el único método para cálculo de varianza en muestreos complejos que aplicaban muchas instituciones era el de grupos aleatorios Dependientes, o bien la linealización. El creciente desarrollo de los equipos de cómputo ha facilitado la exploración de métodos de intenso trabajo computacional. Tal es el caso del *jackknife*, el cual, cada vez está siendo adoptado por más analistas de encuestas en el mundo, en parte porque es más simple de aplicar que el *bootstrap*, además de tener cierto apoyo teórico y de que los estudios empíricos revelan, en general, un buen comportamiento.

3.2.1 Caracterización del estimador

La técnica a la que se ha denominado por *jackknife* partió de un estimador del coeficiente de correlación serial, con correción del sesgo, el cual fue creado por Quenouille (1949). Años más tarde, Tukey (1958), retomó el trabajo de Quenouille y propuso un estimador de la varianza y el sesgo de una estadística. De manera general, el jackknife es un método no-paramétrico basado en remuestreo. Sus inicios se dieron en el contexto de poblaciones infinitas aunque más tarde fue analizado en caso de poblaciones finitas, por lo que finalmente es aplicable al caso de encuestas.

Sin alejarnos del punto central de la tesis, es posible revisar la lógica tras la idea de Quenouille para estimar el sesgo, y posteriormente ver cómo fue aprovechada por Tukey para conseguir un estimador de la varianza. En primer lugar, Quenouille(1949) advierte que para una estadística t_n, basada en una muestra de tamaño n, que se pueda expandir a través de una serie de Taylor, que sea consistente, con cumulantes finitos, se cumple que:

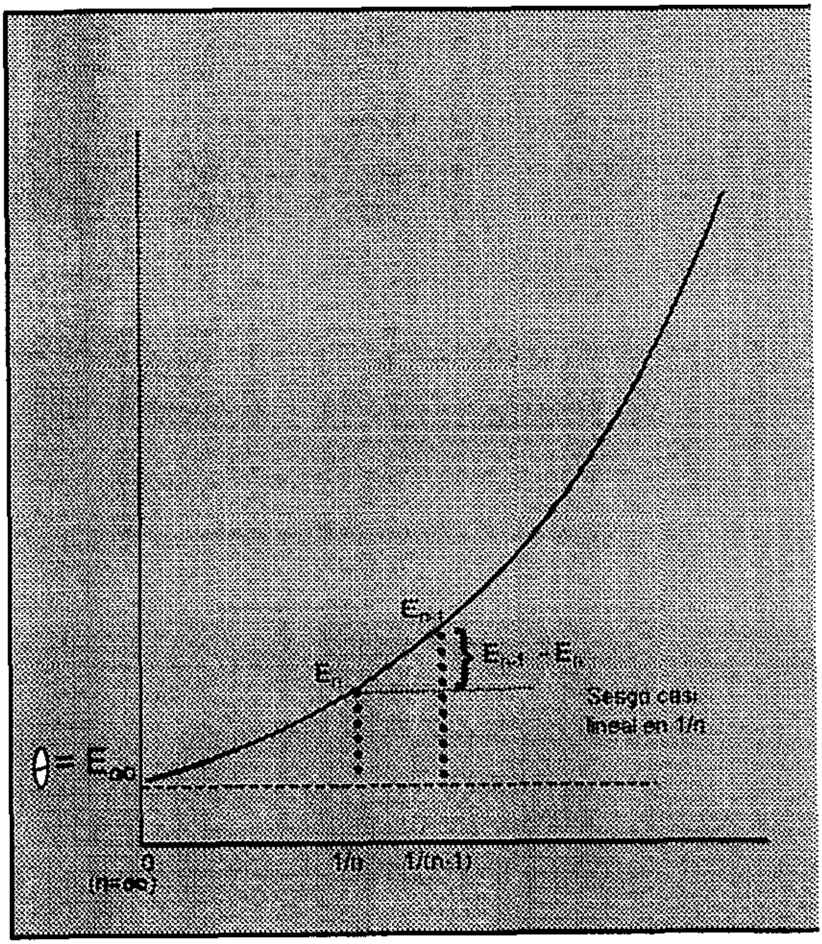
$$\mathrm{E}(t_n - \Theta) = \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3} + \ldots$$

Igualmente, si se calculara la misma estadística, pero con $n-1$ elementos, se tendría que :

$$\mathrm{E}(t_{n-1} - \Theta) = \frac{a_1}{n-1} + \frac{a_2}{(n-1)^2} + \frac{a_3}{(n-1)^3} + \ldots$$

Es claro que bajo los supuestos indicados, el sesgo disminuye, a medida que n crece. Efron (1982) da una explicación gráfica al planteamiento de Quenouille. (Para entenderla mejor, conviene ver la figura 3.1.) Sea E_n la esperanza del estimador t_n, basado en un tamaño de muestra n, entonces la gráfica de E_n vs $\frac{1}{n}$ estaría dada por una curva

Figura 3.1: Gráfica tomada de Efron (1982), que explica la lógica tras el Jackknife original.



Nota: E_n representa la esperanza del estimador basado en n datos.

creciente y cóncava, que parte del punto $(0, \Theta)$, el cual correspondería al caso $n = \infty$. Aproximando tal curva con una línea, se tendría que:

$$\frac{E_n - E_\infty}{E_{n-1} - E_n} = \frac{1/n}{1/(n-1) - 1/n}$$

Se entiende por E_∞, el verdadero valor del parámetro, o Θ. De ahí que el sesgo del estimador basado en n elementos, está dado por $E_n - E_\infty$, lo cual, resolviendo la relación anterior, resulta en

$$Sesgo = E_n - E_\infty = (n-1)(E_{(n-1)} - E_n).$$

De la misma relación se obtiene

$$\Theta = E_\infty = nE_n - (n-1)E_{(n-1)}.$$

La propuesta de Quenouille es obtener, a partir de la muestra de tamaño n, un estimador basado en los n elementos, $\hat{\Theta}$, junto con n estimadores basados en $(n-1)$ elementos cada uno, denominados por, $\hat{\Theta}_{(i)}$. Así pues, se calcula el promedio de los últimos

$$\hat{\Theta}_{(.)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\Theta}_{(i)}. \tag{3.4}$$

El estimador del sesgo está dado por

$$\widehat{Sesgo} \equiv (n-1)(\hat{\Theta}_{(.)} - \hat{\Theta}).$$

También se deduce un nuevo estimador con corrección del sesgo,

$$\tilde{\Theta} = \hat{\Theta} - \widehat{Sesgo} = n\hat{\Theta} - (n-1)\hat{\Theta}_{(.)}.$$

Se puede verificar que si bien $\hat{\Theta}$ tiene un sesgo del orden de $1/n$, el sesgo del nuevo estimador es del orden $1/n^2$.

Basado en el planteamiento de Quenouille, Tukey (1958) sugirió calcular n *pseudovalores*, denominados $\tilde{\Theta}_i$, los cuales se construyen a partir del estimador que se consigue al dejar el dato i fuera de la muestra. Como se ve a continuación el pseudovalor de Tukey corresponde directamente al estimador con correción del sesgo de Quenouille.

$$\tilde{\Theta}_i = \hat{\Theta} + (n-1)(\hat{\Theta} - \hat{\Theta}_{(i)}) = n\hat{\Theta} - (n-1)\hat{\Theta}_{(i)} \tag{3.5}$$

El estimador *jackknife* de Θ , estaría dado por:

$$\bar{\tilde{\Theta}}_J = \frac{1}{n}\sum_{i=1}^{n}\tilde{\Theta}_i. \tag{3.6}$$

La estimación de la varianza se basa en las desviaciones de los pseudovalores hacia su media, de la siguiente forma

$$\widehat{VAR}_{J1} = \frac{1}{n(n-1)}\sum_{\alpha=1}^{n}(\tilde{\Theta}_i - \bar{\tilde{\Theta}}_J)^2 \tag{3.7}$$

Es más común asociar el jackknife directamente a los estimadores $\hat{\Theta}_{(i)}$ (i=1,...,n), quizás porque es más conocida la aplicación del método a problemas de regresión, donde no se involucran pseudovalores como los que propone Tukey para un problema de muestreo. Así pues, se puede ver que (3.7) es equivalente a:

$$\widehat{VAR}_{J1} = \frac{(n-1)}{n}\sum_{i=1}^{n}(\hat{\Theta}_{(i)} - \hat{\Theta}_{(.)})^2. \tag{3.8}$$

Por lo tanto, es factible obtener el estimador de la varianza jackknife, sin necesidad de calcular los pseudovalores de Tukey. Por otra parte, en la práctica, $\widehat{VAR}_{J1}$ no sólo se utiliza para estimar la varianza de $\bar{\tilde{\Theta}}$ sino también la de $\hat{\Theta}$. De hecho, en ocasiones, también se usa el estimador:

$$\widehat{VAR}_{J2} = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\tilde{\Theta}_{(i)} - \hat{\Theta})^2 \tag{3.9}$$

3.2.3 Aplicación del jackknife en muestreo

Se ha explicado la naturaleza del estimador jackknife, pero lo que nos ocupa es el saber cómo se aplicaría este método en el caso de una encuesta. Hasta ahora, lo planteado se asocia a la situación de un muestreo aleatorio simple, donde se calculan n estimadores al eliminarse cada dato . La lista de pasos que se presenta a continuación se refiere a eliminación de un *grupo*, en lugar de un dato. De esta forma, se entiende que la mecánica que se presenta, tomada de Wolter (1985), es el primer intento de plantear la aplicación del jackknife en una manera generalizada para un muestreo complejo. Lo primero que salta a la vista es que se supone que se cuenta con k grupos de m elementos cada uno; lo cual acarrería el problema de formar esos grupos, o de determinar el tamaño de k. En realidad, la mecánica que se expone a continuación no es la mejor para un muestreo complejo, pero se exhibe porque representa una versión que en ocasiones es de utilidad, incluso en problemas de regresión.

De manera general, la estimación por jackknife consiste en los siguientes pasos:

1. Se calcula un estimador $\hat{\Theta}$ de Θ, basado en toda la muestra y respetando el diseño.

2. La muestra es particionada en k grupos con m observaciones cada uno. Es decir, se supone n múltiplo de k ($n = mk$).

3. Se calculan k estimadores, con la misma forma de $\hat{\Theta}$ pero omitiendo en cada ocasión uno de los grupos. Es decir, se obtienen k estimaciones con $m(k-1)$ unidades a las que se denominan mediante $\hat{\Theta}_{(i)}$ (donde $i = 1, ..., k$).

4. Se calcula, para cada grupo, su pseudovalor denominado por: $\tilde{\Theta}_i$, de la siguiente forma.
$$\tilde{\Theta}_i = k\hat{\Theta} - (k-1)\hat{\Theta}_{(i)}$$

5. Se obtiene el estimador jackknife promediando los pseudovalores, como en (3.6).
$$\bar{\tilde{\Theta}}_J = \frac{\sum_{i=1}^{k} \tilde{\Theta}_i}{k}$$

Así pues,

$$\begin{aligned}
\bar{\tilde{\Theta}}_J &= k\hat{\Theta} - (k-1)\frac{\sum_{(i)=1}^{k} \hat{\Theta}_{(i)}}{k} \\
&= \sum_{(i)=1}^{k} \hat{\Theta}_{(i)}(\frac{1}{k} - 1) + \sum^{k} \hat{\Theta}
\end{aligned}$$

$$= \frac{1}{k}\sum_{(i)=1}^{k}\hat{\Theta}_i - \sum_{i=1}^{k}(\hat{\Theta}_{(i)} - \hat{\Theta})$$

Vemos que $\overline{\tilde{\Theta}}_J$ no es simplemente el promedio de las estimaciones de *k grupos*, sino que se contempla un término que aporta información sobre la desviación simple de las estimaciones de cada grupo con respecto al estimador general.

Wolter (1985) indica que cuando se tiene un diseño por conglomerados, se deben considerar como los grupos a ir eliminando, los conglomerados primarios. Es decir, aquellos conglomerados que están constituídos por unidades de segunda, tercera etapa, etc., hasta las unidades elementales.

Resulta necesario mostrar el mecanismo a seguir para aplicar el jackknife bajo un diseño más complicado. Nuevamente, Wolter (1985) es la primera referencia explícita hacia un diseño más complicado. Cabe señalar que, según su planteamiento, los pseudovalores son ajustados para que el estimador jackknife sea igual al que resulta en el caso de una estadística lineal. Además, es importante añadir las siguientes advertencias, que son válidas para cualquier diseño estratificado, sea simple, o de varias etapas:

- Si se desea estimar la varianza dentro de cada estrato, entonces se debe conseguir el estimador jackknife del estrato, de acuerdo al diseño dentro de éste.

- Si se desea estimar la varianza total entre todos los estratos, entonces se deben crear grupos aleatorios bajo la misma estratificación, que contengan elementos de cada uno de los estratos .

Ahora se brinda el esquema general para un diseño estratificado por conglomerados, en el que se incluyen L estratos con n_h (h=1,2,...,L) conglomerados en cada uno. La aplicación del jackknife consiste en ir eliminando conglomerados, como se aprecia en los siguientes pasos:

1. Se calcula $\hat{\Theta}_{(hi)}$ de la misma forma que $\hat{\Theta}$ pero no se incluye el conglomerado i del estrato h.

2. Se obtiene un estimador del estrato promediando sobre todos los $\hat{\Theta}_{(hi)}$. De forma explícita :

$$\hat{\Theta}_{(h.)} = \sum_{i=1}^{n_h} \frac{\hat{\Theta}_{(hi)}}{n_h} \tag{3.10}$$

3. Dado $n = \sum_{h=1}^{L} n_h$, se obtiene un estimador global dado por:

$$\hat{\Theta}_{(..)} = \sum_{h=1}^{L} \sum_{i=1}^{n_h} \frac{\hat{\Theta}_{(hi)}}{n}$$

4. Se obtiene el promedio de las estimaciones de los estratos, al que se denomina por $\overline{\Theta}_{(..)}$.

$$\overline{\Theta}_{(..)} = \sum_{h=1}^{L} \frac{\hat{\Theta}_{(h.)}}{L} \tag{3.11}$$

5. Se procede a calcular los pseudovalores, que para este caso son como sigue:

$$\hat{\Theta}_{hi} = (LW_h + 1)\hat{\Theta} - LW_h\hat{\Theta}_{(hi)} \tag{3.12}$$

$$\begin{array}{rl} \text{donde,} & i = 1, ..., n_h; \;\; h = 1, ..., L. \\ W_h = & (n_h - 1)(1 - \frac{n_h}{N_h}) \text{ si es muestreo sin reemplazo} \\ & (n_h - 1) \text{ para muestreo con reemplazo.} \end{array}$$

6. Finalmente, el estimador jackknife de Θ se define como:

$$\begin{aligned} \hat{\Theta}' &= \sum_{h=1}^{L} \sum_{i=1}^{n_h} \frac{\hat{\Theta}_{hi}}{Ln_h} \\ &= (1 + \sum_{h=1}^{L} W_h)\hat{\Theta} - \sum_{h=1}^{L} W_h\hat{\Theta}_{(h.)} \end{aligned}$$

7. El estimador de la varianza para el muestreo estratificado es:

$$v(\hat{\Theta}) = \sum_{h=1}^{L} \frac{W_h}{n_h} \sum_{i=1}^{n_h} \left(\hat{\Theta}_{(hi)} - \hat{\Theta}_{(h.)}\right)^2 \tag{3.13}$$

W_h es como se definió en (3.12), del inciso 5.

Es evidente, que se requerirán tantas iteraciones *jackknife* como sea el total de conglomerados. Del desglose anterior, se desprende también que de esta forma se obtiene un estimador de la varianza global, mas no se obtienen los componentes de varianza. Por otra parte, se observa que se consideran factores de correción por población finita para conglomerados en cada estrato.

Rao (1996) considera una forma mucho más expedita de llevar a cabo el *jackknife* en un muestreo estratificado bietápico o cualquier muestreo complejo. Su método se aplica a estimadores que se consiguen como función de los factores de expansión. Por ejemplo, en un muestreo estratificado bietápico por conglomerados, con L estratos, n_h conglomerados muestreados en el estrato h y m_{hi} unidades encuestadas en el conglomerado i-ésimo del estrato h y a su vez, w_{hij} representa el factor de expansión de la unidad j-ésima, del conglomerado i-ésimo del estrato h, se puede considerar el siguiente estimador de la media de X:

$$\hat{\bar{X}} = \frac{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} w_{hij}x_{hij}}{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} w_{hij}} \tag{3.14}$$

De igual forma, la correlación lineal entre dos variables x y y, bajo el diseño mencionado, se estima como sigue:

$$\hat{\rho_{xy}} = \frac{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} w_{hij}(x_{hij}-\hat{\bar{x}})(y_{hij}-\hat{\bar{y}})}{\sqrt{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} w_{hij}(x_{hij}-\hat{\bar{x}})^2}\sqrt{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} w_{hij}(y_{hij}-\hat{\bar{y}})^2}} \tag{3.15}$$

Rao al igual que Wolter, estipula que se consigue cada réplica del *jackknife* eliminando un conglomerado de un estrato, pero su metodología se sintetiza en un ajuste de los factores de expansión. En cada iteración, suponga que se elimina el conglomerado l del estrato k y sean w_{hij}^{kl} los factores de expansión ajustados como se indica a continuación:

$$w_{hij}^{kl} = \left\{ \begin{array}{rll} w_{hij} & (h \neq k) & \text{Para las unidades que no están en el} \\ & & \text{estrato donde se está eliminando} \\ & & \text{un conglomerado.} \\ w_{hij}\frac{n_h}{n_h-1} & (\mathrm{i} \neq l) & \text{Para las unidades en el estrato } k \\ & & \text{pero que no son del conglomerado } l. \\ 0 & (\mathrm{i} = l) & \text{Para las unidades del conglomerado} \\ & & \text{que se elimina.} \end{array} \right\}$$

Se obtiene el estimador $\hat{\Theta}_{(kl)}$ de la misma forma que $\hat{\Theta}$ (quitando el conglomerado l-ésimo del estrato k), pero con los nuevos factores de expansión. Luego la varianza de $\hat{\Theta}$ se estima de forma similar a (3.9),

$$v_J(\hat{\Theta}) = \sum_{h=1}^{L} \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} (\hat{\Theta}_{(hi)} - \hat{\Theta})^2 \quad , \tag{3.16}$$

o bien, de acuerdo a (3.8), si se obtienen las diferencias con respecto a $\hat{\Theta}_{(..)} = \sum^{L} \sum^{n_h} \hat{\Theta}_{(hi)}/n$ $(n = \sum_{h=1}^{L} n_h)$:

$$v_J(\hat{\Theta}) = \sum_{h=1}^{L} \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} (\hat{\Theta}_{(hi)} - \hat{\Theta}_{(..)})^2 \quad . \tag{3.17}$$

Esta mecánica, resulta mucho más directa y general, además de que sus estimadores son de forma similar a los que arroja el planteamiento original del jackknife. El estimador de la varianza dado por Wolter(1985) para un muestreo estratificado bietápico, y éste dado por Rao, no son iguales. Se observa que el estimador de Rao no incluye factores de corrección por población finita (fcpf). Al respecto, Canty y Davison (1998) toman la misma expresión de Rao, incorporando los fcpf de la siguiente forma:

$$v_J(\hat{\Theta}) = \sum_{h=1}^{L} \frac{n_h - 1}{n_h} \left(1 - \frac{n_h}{N_h}\right) \sum_{i=1}^{n_h} (\hat{\Theta}_{(hi)} - \hat{\Theta})^2 \quad . \tag{3.18}$$

En el capítulo 6 se verá que esta última versión no fue calculada en la aplicación. Por otra parte, se vió que el estimador de Rao y el de Wolter son numéricamente muy cercanos, lo que implica también, probablemente, que las fracciones de muestreo eran pequeñas.

3.2.4 Inferencia basada en el estimador jackknife

Inicialmente, se estudiaron las propiedades del estimador jackknife en el contexto de poblaciones infinitas. En 1958 Tukey sugirió que los estimadores individuales de las submuestras se podían considerar independientes e idénticamente distribuidos. De tal suerte, propuso el estimador de la varianza que se presentó en la sección 3.2.1. como $\widehat{VAR}_{J1}$ (3.8) . Pero no se conformó con hacer una simple estimación de la varianza. La idea de independencia lo llevó a considerar los pseudovalores, $\tilde{\Theta}_i$, como las x_i en el caso $\hat{\Theta} = \overline{x}$. De tal suerte, propuso basar pruebas de hipótesis e intervalos de confianza para Θ en la estadística $\overline{\tilde{\Theta}}_J$, dada en (3.6). Es decir, un intervalo para Θ con una confianza de $1-\alpha$, estaría dado por:

$$\overline{\tilde{\Theta}} \pm t_{\alpha/2,n-1}\sqrt{\widehat{VAR}_{J1}} \tag{3.19}$$

donde $t_{\alpha/2,n-1}$ es el cuantil de la distribución t con $n-1$ grados de libertad y probabilidad de obtener una t mayor de $\alpha/2$. Finalmente, se comprobó que las sospechas de Tukey eran acertadas en el caso asintótico, aún en el caso de población finita. Ésto se debe al teorema de Krewski y Rao (1981), que como se discutió en el capítulo 1, permite la creación de intervalos de confianza basados en la Normal(0,1) y en el estimador de varianza jackknife, cuando la estadística es suave.

Debido a este resultado o a los mecanismos que siempre se utilizan en el área de muestreo, es usual ver que los intervalos de confianza que se construyen a partir de estadísticas que provienen de muestras grandes (como suele ser el caso de muestreos complejos), se basan en la Normal estándar, debido a que se considera que el tamaño de muestra así lo permite. Sin embargo, Rao (1996) hace ver que los grados de libertad del estimador de varianza obtenido por un método de remuestreo dependen del número de iteraciones o repeticiones que se hacen, por lo que, en la mayoría de los casos, los intervalos deben basarse en la distribución *t de Student* y no en la Normal. En el anexo C se ofrece información sobre el cálculo de grados de libertad.

Recientemente Yung y Rao (1996) establecieron la consistencia de la varianza jackknife bajo ajustes de post-estratificación y no-respuesta. Estos resultados son importantes porque avalan su aplicación en situaciones donde con frecuencia la obtención de estimadores es laboriosa.

3.2.5 El problema de la mediana en el jackknife

Como se ha dicho anteriormente, los resultados asintóticos que se tienen para el jackknife son para estadísticas suaves. Se ha visto que el estimador de varianza jackknife para los cuantiles, o la mediana, en particular, no es consistente. Lo que implica, que aún teniendo muestras grandes el estimador de la varianza se puede alejar mucho del verdadero valor. Esta inconsistencia se ha demostrado (Efron 1985, Capt.3) en el caso de un muestreo aleatorio simple, en el marco de población infinita. Aunque no existen estudios teóricos sobre el comportamiento del jackknife cuando se trata de un muestreo complejo o una población finita, se conserva la alarma y se recomienda no aplicar este estimador de remuestreo. No obstante, Rao, (1996, 1997) plantea la posibilidad de que en un muestreo complejo, donde el jackknife se realiza eliminando conglomerados, no exista tal inconsistencia. Por lo pronto, se consideró de interés exponer en qué consiste el problema de la estimación de la varianza de la mediana por jackknife.

La inconsistencia resulta más evidente cuando n es par. Sean $(x_{(1)}, \ldots, x_{(m)}, x_{(m+1)}, \ldots, x_{(n)})$ las estadísticas de orden, de una muestra de tamaño n, donde $n = 2m$. Entonces la mediana es el punto medio entre $x_{(m)}$ y $x_{(m+1)}$. Claro está que al ejecutar el jackknife, se realizan $2m$ iteraciones. Cuando se elimina algún dato, $x_{(i)}$, tal que $(i) \leq (m)$, entonces, el estimador de la mediana es $x_{(m+1)}$; mientras que si el valor que se elimina es tal que $(i) \geq (m+1)$, entonces, la mediana es estimada por $x_{(m)}$. Así pues, de las $2m$ estimaciones, m de ellas están dadas por $x_{(m)}$ y las otras m, por $x_{(m+1)}$. El promedio de las n iteraciones jackknife es igual al parámetro pues:

$$\frac{m(x_{(m)} + x_{(m+1)})}{n} = \frac{x_{(m)} + x_{(m+1)}}{2}.$$

Pero la varianza jackknife, dada en (3.9), resulta ser:

$$\begin{aligned}
var(mediana)_J &= \frac{n-1}{n}\{\mathrm{m}(\mathrm{x}_{(m+1)} - \frac{\mathrm{x_m} + \mathrm{x_{(m-1)}}}{2})^2 \\
&\quad + m(x_{(m)} - \frac{x_m + x_{(m+1)}}{2})^2\} \\
&= \frac{n-1}{2}\left[\left(\frac{x_{(m+1)} - x_{(m)}}{2}\right)^2 + \left(\frac{x_m - x_{(m+1)}}{2}\right)^2\right] \\
&= \frac{n-1}{4}\left(x_{(m+1)} - x_{(m)}\right)^2 \quad .
\end{aligned}$$

De manera general, existe un teorema que dice [3] que si $X_1, \ldots, X_n$ son

[3]ver Mood (1974), pág. 255

variables aleatorias independientes e idénticamente distribuidas, con densidad $f(.)$ y función de distribución $F(.)$, estrictamente monótona, y con solución única para el cuantil p; entonces, la estadística que estima dicho percentil p, $\hat{X}_{np}$, se distribuye como una Normal con media μ_p y varianza $\frac{p(1-p)}{n\ f(\mu_p)^2}$. En particular, en el caso de la mediana $p = 0.5$ y lo que ocurre con la verdadera varianza del estimador de este cuantil es[4]:

$$Var(mediana) \rightarrow \frac{1}{4n\ f(\mu_p)^2},$$

lo cual significa que si la $f(.)$ es $N(\mu, \sigma^2)$, entonces la varianza del estimador de la mediana es igual a $\pi\sigma^2/2n$. Lo grave del jackknife es que su estimador de varianza, bajo un m.a.s., no se acerca en el límite a la verdadera varianza, y como se puede apreciar, tiende a incrementarse según n crece. Ya que se mencionó el poco conocimiento que se tiene de lo que ocurre en muestreos complejos sobre este respecto, resultará interesante ver en el capítulo 6, los resultados que se obtuvieron de la varianza jackknife de la mediana en la ENAL'96, y su comparación con el estimador de repeticiones balanceadas, el cual no posee éste problema para los cuantiles. (Claro está, no existe referencia sobre la verdadera varianza.)

[4]Ver Efron (1985), Cap.3, pág. 16.

3.3 Bootstrap

3.3.1 Caracterización del estimador

Efron (1979) dió a conocer una técnica para estimar la distribución acumulada de probabilidad, F, de una variable aleatoria. Al presentar el *Bootstrap*, como llamó al método, resaltó su relación con el ya conocido *Jackknife*, además de argüir mayor generalidad. Demostró que el Jackknife es una aproximación lineal del bootstrap, lo cual, sin la referencia cronológica, podría pensarse que ocurriera al contrario (que primero se desarrollara el bootstrap y luego se aproximara linealmente). Cabe resaltar que aunque el bootstrap se concibió en el marco de una población infinita, Efron (1979, 1982) consideró brevemente el caso de población finita. Posteriormente, autores como Rao y Wu (1985, 1988) y Kovar, Rao y Wu (1988) incursionaron en la aplicación del bootstrap en problemas de muestreo, para obtener estimadores de la varianza e intervalos de confianza.

El problema más elemental se basa en la observación de una muestra independiente de tamaño n de una variable aleatoria X, cuya distribución de probabilidad F es desconocida. Se tiene,

$$X_i \sim F \qquad \text{donde, } X_i \text{ i.i.d., } \quad \text{i=1,2,...,n.}$$

Por otra parte, se observa $x = (x_1,x_2,...,x_n)$, la realización de $X = (X_1,X_2,...,X_n)$. Ahora bien si se tiene la variable aleatoria $R(X,F)$, que depende de X y F desconocida, se desearía estimar la distribución muestral de R con base en la observación x. Para tal efecto Efron (1979) propone la consecución de los siguientes pasos, que describen el *bootstrap* en su forma más simple:

1. Se construye la distribución empírica, $\hat{F}$, atribuyendo una masa de $1/n$ a cada observación $x_1,x_2,...,x_n$.

2. Se considera

$$X_i^* \sim \hat{F} \qquad \text{donde, } X_i^* \text{ i.i.d., } \quad \text{i=1,2,...,n.}$$

A partir de $\hat{F}$ se obtiene la muestra "bootstrap" de tamaño n,$x^* = (x_1^*,x_2^*,...,x_n^*)$; lo que equivaldría a obtener dicha muestra, de

$(x_1,x_2,\ldots,x_n)$ con reposición. Con base en x^* se obtiene $R^*(X^*,\hat{F})$, el valor "bootstrap" de R.

3. Se lleva a cabo el inciso 2 repetidas veces dejando $\hat{F}$ fija, para así aproximar la distribución muestral de $R(X,F)$ mediante la distribución de las $R^*(X^*,\hat{F})$.

Es claro que si lo que se deseara fuera aproximar la distribución de X_i, el estimador máximo verosímil, no paramétrico, de F es $\hat{F}$, o sea, la distribución empírica; ante lo cual, no tendría sentido intentar afinar $\hat{F}$. Es importante hacer énfasis en el objetivo del bootstrap, porque si no es bien comprendido, pudiera parecer que es una solución redundante a un problema que ya sabemos resolver.

Cuando se obtiene la muestra $(x_1,x_2,\ldots,x_n)$, se tienen n observaciones de X, y como ya mencionamos, podemos estimar $F_X(X) = F$. Sin embargo, si nos interesa $R(X,F) = t(X)$, sólo contamos con una observación, es decir, se tiene una muestra de tamaño $n = 1$, de $t(X)$. Obviamente, nos vemos imposibilitados para estimar $F_{T(X)}(T(X))$, pues además, $F_X(X)$ es desconocida.

La solución de Efron valora a $\hat{F}$ como máximo verosímil, pues parte de la similitud que espera de esta distribución con F, para simular una muestra de $R(X,F) = t(X)$, representada por los valores $R^*(x,\hat{F}) = t^*(x)$; de lo que se sigue, estimar la distribución de R, con la de R^*, a la cual se denomina *distribución bootstrap.* Queda claro entonces que el paso clave radica en cómo se calcule la distribución bootstrap. Para resolver este problema fundamental Efron(1979) propone 3 alternativas, que se describen a continuación:

1. Cálculo teórico directo. (Obviamente, este método no siempre es posible aplicarlo).

2. Aproximación Monte Carlo. Se generan repetidas observaciones de $X*$, conservando $\hat{F}$ fija, se calculan los valores correspondientes de $R^*(X_1^*,\hat{F})$,$R^*(X_2^*,\hat{F})$,...,$R^*(X_n^*,\hat{F})$, y el histograma de estos últimos, lo que se tomará como la aproximación de la distribución bootstrap.

3. Utilizar algún método de linealización por series de Taylor para obtener la media y varianza de la distribución bootstrap de R^*.

La segunda alternativa de las arriba mencionadas representa la forma en que se ha generalizado la solución de la distribución bootstrap. Con base en ella, se puede explicar la mecánica y la lógica en cuestión: A partir de los valores "bootstrap", $R^*(X*,\hat{F})$, se construye a su vez lo que sería la distribución empírica de una estadística, $t^*(x)$, basada en la empírica de X; la cual viene a ser la distribución bootstrap deseada.

Es obvio que en la medida en que $\hat{F}$ se parezca a F, la estimación de la distribución de R, será mejor; ésto es válido en toda la estadística ya que si la muestra es representativa, la inferencia es mejor. Esta dependencia sobre la muestra se presenta al utilizar cualquier estimador no-paramétrico. Por otra parte, Efron (1979) puntualiza que el que la distribución bootstrap de R^* sea buen estimador de la distribución de $R = t(X)$, también dependerá de la forma de $R(X, F)$; aspecto que se relaciona con las cantidades pivotales. Un ejemplo claro es que se espera que la distribución de $R(X, F) = \frac{t(X)-E_F(t)}{\sqrt{Var_F(t)}}$ se aproxime mejor que la de $R(X, F) = t(X)$. Más adelante en esta tesis, se discutirá cómo este principio se ha aplicado para mejorar los intervalos de confianza.

Se ha revisado ya el fundamento general del *bootstrap*, por lo que se proseguirá con la aplicación de este método al problema que nos ocupa: el cálculo de la varianza de un estimador en un muestreo complejo. Del planteamiento de Efron (1979), se deduce que utilizando el mismo mecanismo con el que se estima la distribución de la estadística de interés, se obtendrá igualmente, una estimación de los parámetros de dicha distribución. Rao y Wu (1985) explican la versión más simple de la aplicación del bootstrap al caso de un muestreo de una población finita, a la cual denominan "el bootstrap ingenuo" (*naive bootstrap*).

Supongamos que se tiene una población finita de la que interesa conocer el parámetro Θ, que es una función del vector de medias de la población, $\overline{\boldsymbol{X}} = (\overline{X}_1,\overline{X}_2,\ldots,\overline{X}_p)^T$. Se considera el estimador $\hat{\Theta} = g(\overline{\boldsymbol{X}})$[1], que aun siendo una función no-lineal de las observaciones, puede ser calculado, pero no así su varianza. En el caso de un muestreo aleatorio simple, de una población finita, el "bootstrap ingenuo" se lleva a cabo mediante los siguientes pasos:

---

[1]Krewski y Rao (1981) demostraron que el estimador $\hat{\Theta} = g(\overline{\boldsymbol{X}})$se puede expresar como $\hat{\Theta} = g(\overline{x})$

1. Partiendo de la muestra observada, se selecciona una muestra aleatoria simple con reemplazo, cuyos valores se denominan $\{x_i\}_1^n$. Se calculan $\overline{x}^* = \frac{1}{n}\sum_1^n x_i^*$, y $\hat{\Theta}^* = g(\overline{x}^*)$.

2. Se repite el paso anterior una gran cantidad de veces, B, y cada estimación de Θ se denomina $\hat{\Theta}^{*1}$,$\hat{\Theta}^{*2}$,...,$\hat{\Theta}^{*B}$. Se obtiene entonces la media de estas estimaciones,

$$\hat{\Theta}_a^* = \sum_{b=1}^{B} \hat{\Theta}^{*b}/B.$$

3. El estimador de la varianza de $\hat{\Theta} = g(\overline{x})$ está dado por:[2]

$$v_b(a) = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\Theta}^{*b} - \hat{\Theta}_a^*)^2 \tag{3.1}$$

Se puede apreciar que el estimador de la varianza es la aproximación Monte Carlo a

$$Var(\hat{\Theta}^*) = E(\hat{\Theta}^* - E_*\hat{\Theta}^*)^2,$$

donde $E_*\hat{\Theta}^*$ es la esperanza bajo la distribución bootstrap.

Es de esperar que si Rao y Wu (1988) llamaron a la metodología anterior "ingenua", ellos hallan presentado un cambio en la técnica, con una justificación adecuada . A continuación se presentará la aplicación del bootstrap a muestreos más elaborados. En la discusión del muestreo estratificado se verá el método propuesto por Rao y Wu, que contempla el cálculo de pseudovalores.

3.3.2 Aplicación del bootstrap en muestreo

Muestreo estratificado

De acuerdo a Rao y Wu (1988) la extensión al muestreo estratificado requiere de "bootstraps" independientes en cada estrato, por cada

[2]**Nota: Aquí la numeración de las ecuaciones vuelve a comenzar en la 3.1, y las referencias que se hacen son sólo hacia las expresiones de esta sección que trata sobre el bootstrap.**

iteración que se realiza (b=1,...,B). Es decir, se obtiene una muestra aleatoria con reemplazo en cada estrato,de tamaño n_h, de entre las unidades que están contenidas en la muestra total, las cuales se pueden denotar por $\{x_{hi}\}$. Se calcula, para cada estrato,

$$\overline{x}_h^* = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}^*$$

Nótese que la nomenclatura que incluye un "*", se refiere a los valores obtenidos de la simulación bootstrap. Por ejemplo, x_{hi}^* significa el estimador del total de una unidad que ha sido seleccionada (en el bootstrap) y $\overline{x}_h^*$ es el promedio del estrato h obtenido en una iteración bootstrap. Así también se obtiene $\overline{x}^* = \sum W_h \overline{x}_h^*$ y $\hat{\Theta}^* = g(\overline{x}^*)$, donde W_h son los pesos de los estratos. Al igual que antes, se repite el muestreo (bootstrap) por estrato y la obtención de estimadores, que también se llaman $\hat{\Theta}^{*1}$,...,$\hat{\Theta}^{*B}$, seguida del promedio de éstos, $\hat{\Theta}_a^* = \frac{\sum \hat{\Theta}^{*b}}{B}$. El estimador de la varianza está dado por (3.1) de la sección 3.

Al examinar el caso lineal, con $p = 1$, $\hat{\Theta}^* = \sum W_h \overline{x}_h^* = \overline{x}^*$, Rao y Wu (1988) compararon la varianza que se obtendría mediante el bootstrap, $var_b(\overline{x}^*)$ con $var(\overline{x})$ y advirtieron que $\frac{var_b(\overline{x}^*)}{var(\overline{x})}$ *no converge a 1 en probabilidad* a menos que el número de estratos sea fijo y $n_h \to \infty$ en cada estrato. Como consecuencia, $var_b(\overline{x}^*)$ no es un estimador consistente de la varianza de $\overline{x}$, ni tampoco lo es la varianza bootstrap de cualquier estadística no lineal.

Para solucionar este problema, Rao y Wu (1988) propusieron otro método de realizar el bootstrap, que se diferencía del anterior en la obtención de pseudovalores antes de calcular el estimador bootstrap. La propuesta es como sigue:

1. Se obtiene una muestra aleatoria simple en cada estrato, de tamaño m_h, $\{x_{hi}^*\}_{i=1}^{m_h}$, a partir de la muestra observada $\{x_{hi}\}_{i=1}^{n_h}$. Teniendo en mente que $\overline{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$ es la media del estrato h que resulta de la muestra original (total), se calculan,

$$\tilde{x}_{hi} = \overline{x}_h + \frac{\sqrt{m_h}}{\sqrt{n_h - 1}}(x_{hi}^* - \overline{x}_h)$$

$$\begin{aligned}
\tilde{x}_h &= \frac{1}{m_h}\sum_{i=1}^{m_h}\tilde{x}_{hi} \\
\tilde{x} &= \sum W_h\tilde{x}_h \\
\tilde{\Theta} &= g(\tilde{x})
\end{aligned}$$

Cabe advertir que si el muestreo es sin reemplazo, en lugar de la expresión anterior para $\tilde{x}_{hi}$, se deberá considerar (3.3) de la sección 3.

2. Se repite lo anterior un gran número de veces[3], B, y se obtienen, $\tilde{\Theta}^1$,...,$\tilde{\Theta}^B$, así como el promedio de éstos, $\tilde{\Theta}_a^* = \frac{\sum_{b=1}^{B}\tilde{\Theta}^b}{B}$.

3. El estimador de la varianza está dado por :

$$\tilde{v}_b(a) = \frac{1}{B-1}\sum_{b=1}^{B}(\tilde{\Theta}^b - \tilde{\Theta}_a^*)^2 \tag{3.2}$$

Obviamente el estimador representa la estimación Monte Carlo de $E(\tilde{\Theta} - E_*\tilde{\Theta})^2$.

Con la adecuación que se acaba de señalar, ocurre que en el caso lineal, (3.2) se reduce a $var(\overline{x})$, para cualquier tamaño elegido, m_h. En el caso no lineal, $\tilde{v}_b(a)$ es consistente para $Var(\hat{\Theta})$. Por otra parte, cabe advertir que la normalidad asintótica de $\frac{(\tilde{\Theta}-\hat{\Theta})}{\sqrt{v_b}}$ se cumple para m_h comparable a n_h, en cada estrato. Si se elige $m_h = n_h - 1$ el método propuesto por Rao y Wu se reduce al bootstrap ingenuo, en el que se toman muestras de tamaño $n_h - 1$ en cada estrato. En tal situación, los pseudovalores $\tilde{x}_{hi}$ son iguales a las observaciones simuladas en el bootstrap ingenuo, es decir, $\tilde{x}_{hi} = x_{hi}^*$. Si ocurre que $n_h = 2$, Rao y Wu recomiendan tomar $m_h = 3$ o 4, ya que para $m_h = 1$ la varianza obtenida por el método de "Muestreo por Mitades" (*Half-sampling*) o repeticiones balanceadas sería más estable.

Es importante advertir que si el muestreo es estratificado simple **sin reemplazo** entonces, los pseudovalores contemplan un factor por

[3]Se entiende que el paso 1 se repite B veces, por lo que cada uno de las estadísticas que allí se mencionan llevarían un índice b ($b = 1, ..., B$).

correción finita de la siguiente forma:

$$\tilde{x}_{hi} = \overline{x}_h + \frac{\sqrt{m_h}}{\sqrt{n_h - 1}}\sqrt{1 - f_h}(x^*_{hi} - \overline{x}_h) \tag{3.3}$$

donde, $f_h = n_h/N_h$. Además, aunque se elija $m_h = n_h - 1$, si el muestreo es sin reemplazo, ocurre que $\tilde{x}_{hi} \neq x^*_{hi}$. El objetivo de estos ajustes por pseudovalores es lograr que la varianza bootstrapde una estadística lineal sea igual a la conocida. Este proceso de reescalamiento, según Sitter (1992) tiene el inconveniente de que en muestreos complejos se necesitan varias estadísticas de resumen, además de que los ajustes varían para distintos diseños y puede llegar a ocurrir que algún $\hat{\Theta}^b$ sea negativo, aún cuando por definición, $\hat{\Theta} \geq 0$.

Muestreo por conglomerados en dos etapas

El caso particular que ahora se discutirá es la aplicación del bootstrapcuando se tiene una población con N conglomerados, cada uno con M_i unidades. Se seleccionan n conglomerados **sin reemplazo** y m_i unidades en el conglomerado i-ésimo de los seleccionados, a través de un muestreo aleatorio simple sin reemplazo, nuevamente. Usualmente se denota por M_0 el tamaño de la población, es decir, $M_0 = \sum_1^N M_i$; lo que muchas veces es desconocido.

Para aplicar un bootstrapahora hay que llevar a cabo dos etapas de aleatorización. Primero se elige una muestra aleatoria simple de n conglomerados *con reemplazo*. Luego se selecciona una muestra aleatoria con reemplazo de m_i elementos, entre las m_i unidades de la muestra original, en el conglomerado correspondiente. Si un mismo conglomerado se selecciona varias veces, se realizan tantos muestreos independientes como veces haya sido elegido. La secuencia anterior produciría una iteración del Bootstrap, por lo que habría que repetirla B veces, para llegar a obtener el estimador bootstrapde la varianza.

Con el objeto de ser más explícita, hay que entrar en detalles de notación. Se recordará que el estimador en cuestión, $\hat{\Theta}$, se puede expresar como $\hat{\Theta} = g(\hat{\overline{X}})$. Ahora bien, la expresión para el estimador insesgado de $\overline{X}$ es:

$$\hat{\overline{X}} = \frac{\hat{X}}{M_0} = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{X}_i}{\overline{M}_0} \tag{3.4}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}\frac{M_i X_{ij}}{\overline{M}_0}$$

$$\text{Puesto que, } \hat{X} = \frac{N}{n}\sum_{i=1}^{n} M_i \overline{x}_i,$$

$$\overline{x}_i = \frac{1}{m_i}\sum_{i=1}^{m_i} x_{ij}$$

$$\text{y, } \overline{M}_0 = \frac{M_0}{N}.$$

Por otra parte, sea

$x_{ij}^{**} =$ el valor de x, del elemento j del conglomerado i-ésimo simulado por el Bootstrap.

$M_i^* =$ el tamaño del conglomerado iésimo del Bootstrap.

$m_i^* =$ el tamaño de muestra en el Bootstrap, en el conglomerado i-ésimo.

$\hat{X}_i^* =$ el estimador del total en el conglomerado i-ésimo del Bootstrap.

$$\begin{aligned}
\hat{X}_i^* &= M_i^* \overline{x}_i^{**} \\
\overline{x}_i^{**} &= \sum_j x_{ij}^{**}/m_i^* \\
\lambda_1^2 &= \frac{n}{n-1}\left(1-\frac{n}{N}\right) \\
\lambda_{2i}^{*2} &= \frac{n}{N}\left(1-\frac{m_i}{M_i}\right)\frac{m_i^*}{m_i^*-1}
\end{aligned}$$

Rao y Wu indican que se deben calcular los siguientes valores:

$$\tilde{x}_{ij} = \hat{\overline{X}} + \lambda_1\left[\frac{\hat{X}_i^*}{\overline{M}_0} - \hat{\overline{X}}\right] + \lambda_{2i}^*\left[\frac{M_i^* x_{ij}^{**}}{\overline{M}_0} - \frac{\hat{X}_i^*}{\overline{M}_0}\right] \tag{3.5}$$

y

$$\tilde{\overline{X}} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i^*}\sum_{j=1}^{m_i^*}\tilde{x}_{ij}$$

$$= \hat{\overline{X}} + \frac{\lambda_1}{n}\sum_{i=1}^{n}\left[\frac{\hat{X}_i^*}{\overline{\overline{M}}_0} - \hat{\overline{X}}\right] + \frac{1}{n}\sum_{i=1}^{n}\lambda_{2i}^*\left[\frac{\hat{X}_i^{**}}{\overline{\overline{M}}_0} - \frac{\hat{X}_i^*}{\overline{\overline{M}}_0}\right] \tag{3.6}$$

Por supuesto, para la ejecución del Bootstrap, se deben repetir las aleatorizaciones y los cálculos arriba mencionados, una gran cantidad de veces, B. Al igual que en los otros casos vistos, se obtienen B estimaciones de $\hat{\Theta}$, y el promedio de ellas. Finalmente el estimador de la varianza de $\hat{\Theta}$ se consigue de manera equivalente a (3.2) de la sección 3.

Estratificado bietápico por conglomerados

En el caso del Jackknife y de repeticiones balanceadas se mostró un procedimiento de llevarlos a cabo que simplifica muchos cálculos, ajustando los factores de expansión, y que se hace extensivo a cualquier diseño y estimador. Con la misma lógica, Rust y Rao (1996), describen lo que probablemente es la forma más corta de hacer un bootstrap, en un muestreo estratificado bietápico por conglomerados. Aquí la generalización a cualquier diseño no es directa porque el ajuste de los pesos no es tan evidente. Para cada réplica del *bootstrap*, en cada estrato h, se eligen por m.a.s. con reemplazo z_h ($z_h > 0$) conglomerados de los n_h que existen en la muestra del estrato. (Como se aprecia se simula la elección de conglomerados y no de unidades dentro de conglomerados.) En cada iteración, se obtiene, para cada conglomerado de todos los estratos, el número de veces que fue seleccionado, al que se denomina $r_{hi}^{(b)}$. Es obvio que $0 < r_{hi}^{(b)} \leq z_h$. El ajuste de los factores de expansión, en la iteración b, se obtiene de la siguiente forma:

$$w_{hij}^b = w_{hij}\left(\left[1 - \sqrt{\frac{z_h}{(n_h-1)}}\,\right] + \sqrt{\frac{z_h}{(n_h-1)}}\frac{n_h}{z_h}r_{hi}^{(b)}\right). \tag{3.7}$$

Anteriormente se vio que para un diseño estratificado, se recomendaba la obtención de unos pseudovalores y varias estadísticas por estrato, de tal suerte que el estimador resultaba el conocido si se trataba de una estadística lineal. Este ajuste de pesos simplifica el proceso que inicialmente recomendaba Rao, pero se aplica solamente al estratificado por conglomerados. Obviamente, esta metodología no proporciona elementos para encontrar los componentes de varianza, por lo que, de tener interés en ello, habría que replantear la solución, considerando la aleatorización dentro de conglomerados.

Aunque ya se han visto ejemplos de cómo se estiman estadísticas que se expresan en términos de los factores de expansión, se presentará el caso de la mediana estimada por bootstrap, bajo el diseño que aquí tocamos.

En el caso en que se quiere estimar un cuantil y su varianza [4], como puede ser la mediana, se requiere estimar parte de la función de distribución muestral de la variable de interés, $F_{n(\mathbf{s})}$. Sea $y = (y_1, \ldots, y_N)$ el vector de valores de dicha variable en la población. El cuantil "p" en la población se estima de:

$$\xi_p = F_{n(\mathbf{s})}^{-1}(p) = inf\left[y_i : \ F_{n(\mathbf{s})}(y) \ \geq \ p\right] \tag{3.8}$$

La estimación de $F_{n(\text{bf s})}$ se hace creando variables indicadoras, I_z, tales que

$$\begin{aligned} I(z) = I(y_{hij} \ \leq \ z) \ &= \ 1 \ \text{Si} \ \ y_{hij} \leq z \\ &= \ 0 \ \text{si no se cumple lo anterior.} \end{aligned}$$

En la expresión anterior y_{hij} denotaría que el diseño es estratificado bietápico por conglomerados y que este valor representa la respuesta de la unidad j-ésima del conglomerado i-ésimo del estrato h. Kovar, Rao y Wu (1988) recomiendan basar la estimación de la función de distribución en el pseudovalor $\tilde{y}_{hi}$, que se consigue mediante (3.2) o (3.3) [5].

De tal suerte la función de distribución (considerando el mismo diseño), evaluada en z, $F_{n(\mathbf{s})}(z)$, se obtiene de [6]:

$$F_{n(\mathbf{s})}(z) = \frac{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} w_{hij}^{b} I(z)_{hij}}{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} w_{hij}^{b}}$$

[4] Ver Kovar, Rao, Wu (1988) y Kish (1972)

[5] Así pues (3.9) se obtiene de$I(z) = I(\tilde{y}_{hij} \ \leq \ z)$.

[6] De manera general, considerando que la muestra total es de $n(\mathbf{s})$ unidades, se estimaría mediante:

$$F_{n(\mathbf{s})}(z) = \frac{\sum_{i=1}^{n(\mathbf{s})} w_i^{(b)} I(z)_i}{\sum_{i=1}^{n(\mathbf{s})} w_i^{(b)}}$$

Esta función se debe evaluar en varios puntos (z), hasta encontrar el valor del cuantil deseado, como indica (3.8). La varianza del cuantil se consigue repitiendo el proceso de aleatorización y cálculos B veces y utilizando el estimador dado en (3.2), donde es claro que $\tilde{\Theta}^b$ es el cuantil que se estima en la iteración número b.

3.3.3 Intervalos de confianza basados en el estimador *bootstrap*

Se vió que Krewski y Rao (1981) dieron elementos para basar los intervalos de confianza en la distribución normal estándar cuando el estimador de la varianza es del tipo Jackknife, de linealización o de repeticiones balanceadas. En el caso del Bootstrap, los intervalos sólo se basan en la propia distribución que éste genera. Chaudhuri y Stenger (1992) abordan brevemente el *Bootstrap* y la forma de hacer intervalos, así también Kovar, Rao y Wu (1988).

La manera más sencilla empieza por la elaboración de un histograma de los estimadores que se obtuvieron en las B iteraciones (En realidad, podría tenerse solamente los datos ordenados y no la gráfica en sí). Se consigue el valor de $\hat{\Theta}$ que tiene un área a su izquierda de 100 $\alpha/2$ (el que se denomina $\hat{\Theta}_{\alpha/2,l}$) e igualmente, aquél valor del estimador que tiene la misma área a su derecha ($\hat{\Theta}_{\alpha/2,u}$). Siguiendo el *método del percentil*, como comúnmente se conoce lo que se acaba de describir, finalmente, se considera que

$$\left(\hat{\Theta}_{\alpha/2,l},\ \hat{\Theta}_{\alpha/2,u} \right)$$

es un intervalo de confianza para Θ con un nivel de 100 $(1-\alpha)$

Existe otra alternativa para elaborar intervalos basados en un *Bootstrap*. Éste consiste en obtener estadísticas de la forma de una *t de Student*, a partir de los estimadores de cada iteración y la varianza bootstrap, como sigue:

$$t^b = \frac{\left(\tilde{\Theta}^b - \tilde{\Theta}_a\right)}{\sqrt{\tilde{v}_J^b}} \quad ,$$

donde $\tilde{v}_J^b$ es el estimador de varianza Jackknife aplicado a los datos simulados en la iteración b del Bootstrap. Posteriormente, se encuentran los valores de $t_{\alpha/2,l}$ y $t_{\alpha/2,u}$, de la misma forma que $\hat{\Theta}_{\alpha/2,l}$ y $\hat{\Theta}_{\alpha/2,u}$. O sea, observando el área que queda a la izquierda y derecha, respectivamente, de estos estimadores. El intervalo para Θ, con un nivel de confianza de $100(1-\alpha)$es la varianza Jackknife calculada para toda la muestra, se da de la siguiente forma:

$$\left(\hat{\Theta} \;-\; t_{\alpha/2,u}\sqrt{\tilde{v}_J}\ ,\ \hat{\Theta} \;-\; t_{\alpha/2,l}\sqrt{\tilde{v}_J}\ \right) \tag{3.9}$$

Si se recuerda, al discutir los inicios del *bootstrap*, se comentó que era de esperar que se lograra una mejor o más rápida aproximación a la distribución de $R(X,F) = \frac{t(X)-E_F(t)}{\sqrt{Var_F(t)}}$ que a la de $R(X,F) = t(X)$; tal es la justificación del intervalo dado en (3.9). Kovar, Rao y Wu (1988) en un estudio empírico encontraron que para los casos de estimador de razón y coeficiente de correlación, esta lógica para hallar un intervalo sólo probó ser mejor en el caso de intervalos de un sólo lado. La gran dificultad operativa de los intervalos bootstrapes que además de requerir de un gran número de aleatorizaciones se suma el hecho de que necesitan del cálculo de la varianza jackknife $B+1$ veces (Una vez para toda la muestra y B veces para las iteraciones).

Capítulo 4

Discusión sobre los métodos

Resulta conveniente comentar los distintos aspectos que caracterizan los estimadores de varianza que se han presentado en esta tesis y las posibilidades de aplicarlos. Este capítulo comienza por dar a conocer la paquetería de cómputo que resuelve el cálculo de varianza por alguno de estos métodos. Se puntualizan las características de cada método y posteriormente se comparan los métodos considerando los hallazgos de varios autores.

4.1 Paquetería existente

Aunque no le fue posible a la autora de esta tesis inspeccionar la paquetería que se menciona en varias referencias bibliográficas, considera que es importante comentar sobre el apoyo existente en programas de cómputo. Los programas que comúnmente se usan para el análisis de encuestas, como el SPSS o BMDP no proveen rutinas que apliquen alguno de las técnicas de remuestreo ni linealización. El SAS, según se encontró en la literatura, tiene acceso a varias rutinas, pero sólo en equipo de cómputo mayor.

De entre todos los métodos que se revisan en esta tesis, el que ha gozado de mayor difusión es el de linealización (es por eso que se incluyó su descripción). Rust y Rao (1996) comentan que la aplicación de técnicas de remuestreo se ha hecho, mayormente, a base de programación, y ésto ha causado que aquellos equipos de analistas que no son conformados por personas con bastante conocimiento de estadística se sientan más confortables con la linealización cuando es factible a través de un programa. Los paquetes que aplican la linealización son : **SESUDAAN/ SUDAAN, PC CARP/ SUPERCARP**. SUDAAN y PC CARP son versiones para computadora personal de SESUDAAN y SUPERCARP, respectivamente. Lepidus Carlson, Johnson y Cohen (1993) revisan estos últimos paquetes y sus versiones originales (para equipo mayor) en cuanto a la dificultad de programación, preparación de la base de datos y otros aspectos. Nathan (1988) comenta que el SUPERCARP y el **MINICARP** proveen la estimación (por linealización) de matrices de varianza-covarianza para análisis multivariados. Al respecto, también comenta Nathan que el paquete **OSIRIS IV-PSALMS** se basa en repeticiones balanceadas.

Algunas instituciones han preferido desarrollar sus propios programas basándose en métodos de remuestreo. Un caso que comentan Rust y Rao (1996) es el *National Assesment of Educational Progress*, en E.U., que realiza sus estimaciones con el método jackknife desde 1984. Igualmente, Rao(1997) reporta que para el análisis de la *Encuesta de Salud del Corazón de Canadá* se desarrolló un paquete que utiliza el jackknife, llamado, **JACKVAR**.

Rust y Rao(1988) mencionan el programa **CPLX**, que hace análisis de modelos log-lineales utilizando el jackknife. Otros programas o rutinas, de acuerdo con ellos son: **WesVarPC**; y las rutinas de SAS, WESVAR, WESREG y WESLOG. Nathan(1988) también menciona una rutina que se accesa por SAS, llamada **SURREGR**, que sirve para la estimación de coeficientes de regresión. (No se especifica en qué técnica se basa.)

4.2 Observaciones sobre los métodos

Después de la exposición hecha en los capítulos anteriores, es conveniente resumir ciertas propiedades o aspectos de cada uno de los métodos. A continuación se desglosan los aspectos más relevantes de los distintos estimadores presentados.

1. linealización por Series de Taylor

- En poblaciones infinitas se asegura que el residuo R_n tiende a cero, pero en el caso de muestreo de una población finita, sólo se supone que es de orden menor al término lineal de la serie de Taylor, por lo que pudiera darse el caso de que en realidad no converja.
- Los intervalos dados con base en la NORMAL(0,1) o la *t de Student* y la varianza estimada mediante linealización, son válidos en el marco estipulado por Krewski y Rao (1981)[1].
- El estimador de la varianza de linealización es asintóticamente consistente para la verdadera varianza de estadísticas suaves (que se pueden expresar como función de la media o el total), de acuerdo a las condiciones expuestas en el capítulo 1, sección 1.6.
- En ocasiones resulta necesario estimar varianzas y covarianzas de estadísticas lineales, bien por fórmula o por algún método de remuestreo, considerando el diseño, para finalmente conseguir el estimador dado por linealización.
- Para cada estimador, hay que calcular las derivadas que vengan al caso y las varianzas que se mencionan en el punto anterior.
- Aunque es posible manejar la no-respuesta y la pos-estratificación para ciertos estimadores, no es sencillo en general.

2. Grupos Aleatorios Independientes

   - Este método sirve para medir otras fuentes de error, como puede ser el sesgo introducido por un equipo de levantamiento de encuesta.
   - Puede ser incosteable si se desea un buen tamaño de muestra en cada grupo.
   - Tiene pocos grados de libertad porque normalmente se obtienen pocas muestras (2 o 3).
   - Tanto el estimador del parámetro Θ (lineal o no lineal) y de su varianza, son insesgados.

[1]Ver capítulo 1 .

- Si $\hat{\Theta}$ es lineal entonces el estimador que se obtendría con la aglomeración de todos los grupos y el promedio de los grupos, son iguales.
- Los intervalos de confianza basados en la Normal(0,1), son válidos con base en el Teorema Central del Límite.

3. Grupos dependientes

   - Tanto el estimador del parámetro como el de su varianza son sesgados.
   - Existe inseguridad en cuanto al número óptimo de grupos que se deben formar y además el establecer éste, depende de las posibilidades que ofrece el diseño.
   - Los grupos se deben formar respetando el diseño original.
   - Las posibilidades de hacer inferencia son nulas pues no existen bases teóricas para ello.
   - A pesar de lo anterior, durante décadas ha sido ampliamente utilizado por organismos gubernamentales internacionales, por ser el único método factible de ser implementado.

4. Repeticiones balanceadas

   - Está dirigido a un diseño en particular, un estratificado donde se tienen dos unidades por estrato. (Tales unidades suelen ser conglomerados).
   - Cuando se aplica el método a un estimador lineal,$\hat{\Theta}$, tanto éste como el estimador de la varianza, son insesgados e iguales al estimador obtenido por fórmula. Pero, es claro que si $\hat{\Theta}$ no es lineal no se garantiza ninguna de estas propiedades.
   - Su programación necesita que se incluyan la matrices Hadamar necesarias, o bien una rutina que las genere.
   - El estimador de la varianza de repeticiones balanceadas es asintóticamente consistente para la verdadera varianza de estadísticas suaves (que se pueden expresar como función de la media o el total), de acuerdo a las condiciones expuestas en el capítulo 1, sección 1.6.

- Los intervalos dados con base en la NORMAL(0,1) o la *t de Student* y la varianza estimada mediante repeticiones balanceadas, son válidos en el marco estipulado por Krewski y Rao (1981)[2]

- Es recomendable considerar adecuadamente los grados de libertad, de acuerdo al número de estratos y hacer los intervalos basados en la *t de Student*[3].

- La no-respuesta y pos-estratificación es fácil de manejar, mediante ajustes de factores de expansión.

5. Jackknife

- Para **m.a.s. con reemplazo**, y en el caso de $\hat{\Theta}$ lineal, el estimador jackknife es igual al obtenido por fórmula; y si se resuelve el *jackknife* eliminando un dato a la vez, la misma reproducción se obtiene con la varianza. Con m.a.s. sin reemplazo, se mantienen estas propiedades, considerando la corrección de los pseudo-valores de Wolter (1985).

- Se aplica fácilmente a cualquier diseño.

- Los intervalos dados con base en la NORMAL(0,1) o la *t de Student* y la varianza estimada mediante jackknife, son válidos en el marco estipulado por Krewski y Rao (1981), como se mencionó para la linealización y Repeticiones Balanceadas.

- Se debe cuidar la elección adecuada de los grados de libertad, como se indica en el anexo C.

- El estimador de la varianza jackknife es asintóticamente consistente para la verdadera varianza de estadísticas suaves (que se pueden expresar como función de la media o el total), de acuerdo a las condiciones expuestas en el capítulo 1, sección 1.6.

- El manejo de no-respuesta y pos-estratificación es sencillo y además existe un resultado sobre la consistencia de la varian-

[2]Ver capítulo 1 .

[3]Ver el anexo C

za jackknife bajo pos-estratificación y ajustes de los pesos por no-respuesta (Rust y Rao (1996), pág. 294)[4].

- El *jackknife* de eliminar un dato a la vez, el marco de $y_1, \ldots, y_n$, variables aleatorias i.i.d., es asintóticamante inconsistente para estadísticas no-suaves, como la mediana. Pero se cree que en un muestreo complejo, donde se elimina un conglomerado en cada iteración, no ocurre ésto. (Aspecto bajo estudio, *comentado por* Rao(1997)).

6. Bootstrap

- No se reproduce el estimador lineal ni su varianza si se aplica un *bootstrap ingenuo.* Para recuperar el estimador lineal y tener consistencia, es necesario aplicar el bootstrap considerando los ajustes expuestos en la sección 3.4.
- Se aplica a cualquier diseño, aunque resulta mucho más laboriosa su programación cuanto más complicado es éste.
- Se consigue una aproximación de la distribución de la estadística de interés, lo que no se tiene con otros métodos.
- Los intervalos e inferencias no se basan en la distribución NORMAL ni en la t, sino que es una técnica no-paramétrica, y las inferencias se basan en la propia distribución que se aproxima. Se ha encontrado que es mejor generar estadísticas t en cada iteración y crear los intervalos con base en ellas. Pero ésto requiere del cómputo de $(B+1)$ estimaciones de varianza jackknife.
- Es el método que requiere de mayor tiempo y esfuerzo en computación, porque se deben incluir rutinas que simulen la selección de unidades de muestreo y como se dijo en el punto anterior, hay que calcular varianzas jackknife para obtener mejores intervalos.

[4]Los autores citan una tesis de doctorado de W. Yung, en la Universidad de Carletón.

4.3 Comparaciones empíricas de los métodos para el caso de muestreo

Hasta la fecha, no ha habido un estudio analítico que compare los diferentes estimadores de varianza en el marco de una población finita y un muestreo complejo. El comportamiento de estos estimadores se revisan empíricamente, a través de estudios de simulación y aplicaciones. Uno de los estudios más completos, bajo un diseño estatificado simple (basado en 32 estratos) es el de Kovar, Rao y Wu (1988), donde se comparan : jackknife, bootstrap, repeticiones balanceadas y linealización por series de Taylor. En este estudio se obtuvieron estimadores de varios parámetros (razón, coeficiente de regresión, coeficiente de correlación) y sus varianzas una gran cantidad de veces (1000) para ser considerados como los parámetros poblacionales. Por esta razón se obtuvieron los ECM de los estimadores de los distintos métodos, los sesgos, una medida de estabilidad, y la cobertura de los intervalos. El hallazgo más importante fue establecer la similitud de la linealización con el jackknife, en el sentido que los sesgos y estimadores de varianza de ambos son equiparables; de hecho, resultó que estos estimadores son de menor sesgo y ECM que los estimadores bootstrap o de repeticiones balanceadas. Por otra parte, el bootstrap y repeticiones balanceadas producen estimadores similares. En general las diferencias entre los 4 métodos son pocas. Sin embargo, en este estudio por simulación se vió que las varianzas estimadas por jackknife y linealización son más precisas, pero los intervalos que dan tienden a estar un poco por debajo del nivel nominal; mientras que el bootstrap y repeticiones balanceadas tienden a sobreestimar la varianza, por lo que lo últimos dan intervalos más conservadores (especialmente el bootstrap).

Contrapuesto a lo que toca de este resultado al jackknife, se encontró el trabajo teórico de Efron y Stein (1981) donde se demuestra, en el marco de población infinita y m.a.s., que para una función simétrica, $S(X_1, ..., X_N)$, el estimador de la varianza jackknife es conservador, es decir, su esperanza es mayor o igual que la verdadera varianza; siendo que igualmente demuestra que bajo el mismo contexto, el sesgo del estimador de varianza bootstrap es menor al del jackknife. Aunque el estudio de Kovar, Rao y Wu (1988) se concretó a un diseño estratificado simple llama la atención el que dilucide que los resultados en el marco teórico y de m.a.s., no necesariamente son válidos en un muestreo complejo.

Por otra parte, Sitter (1992) publica un estudio de simulación similar al de los autores citados arriba, donde compara los resultados del jackknife, linealización y varias versiones de bootstrap. Considera 8 poblaciones bajo diseños estratificados simples sin reemplazo, con distintos números de estratos y unidades. Se vió que par el estimador de razón, el jackknife y la linealización producen resultados similares en cuanto a sesgo y estabilidad, y se comportan mejor que los estimadores bootstrap. Para el estimador de regresión, el bootstrap y el jackknife son equiparables, pero el bootstrap resultó un poco mejor para el estimador de correlación. En cuanto a la mediana, sólo incluyó la estimación de la varianza jackknife de la mediana para una población con 32 estratos. Se encontró que el error nominal en las colas de los intervalos era grandes (18.6 y 8.0 %). Los errores nominales en los intervalos jackknife para la mediana en el estudio de Kovar, Rao y Wu (1988) son un poco menores pero se aprecian medidas de sesgo y estabilidad poco deseables.

4.4 Otros aspectos interesantes encontrados

Existen detalles sobre la congruencia de los métodos o la lógica tras ciertos planteamientos que se pretenden ahora comentar. Hay aspectos que pueden parecer irrelevantes, como el preguntarse por qué se divide por k en unos estimadores y por $k(k-1)$, en otros. Pero, se prefiere dejar claras todo tipo de sutilezas, por simples que parezcan.

1. Sin duda, la pregunta de arriba es la que viene a la mente de muchas personas al revisar estas técnicas. En el caso de Grupos Aleatorios y jackknife, la sumatoria del cuadrado de las diferencias se divide[5] por $k(k-1)$. Como estos estimadores se elaboraron de manera que en el caso de Θ lineal, en particular la media, coincidan con los estimadores conocidos, se tiene que al dividir por

[5]En el caso del jackknife conviene revisar la sección 3.2.1 y ver la expresión (3.7).

$k-1$, se estima la varianza de las observaciones, pero para estimar la varianza de la media de las observaciones se requiere dividir también por k; lo que hace la fórmula de estos métodos es emular la varianza de la media para conseguir la de un estimador cualquiera. Por otro lado, ocurre que en el bootstrap, las estadísticas $\hat{\Theta}^b$ (b=1,.., B), obtenidas en cada iteración, son las observaciones ("las x_i"), que se suponen provienen de una misma distribución y de las que se quiere estimar su varianza; por lo que la suma del cuadrado de diferencias se divide entre $B-1$ (algunos autores dividen por B). Queda la duda de por qué entonces, bajo repeticiones balanceadas se divide sólo por k. Si se revisa el desarrollo para una estadística lineal en 2.2.1, se comprende que el estimador expresado de esta forma coincide con el estimador lineal. Es decir, el razonamiento que llevó a su fórmula es el mismo del jackknife y Grupos Aleatorios.

2. Se mencionó, al hablar de grupos aleatorios repetidos (GAR), que cuando $n_h = 2$ su proceso era similar al de Repeticiones Balanceadas. La similitud consiste en el hecho de que las réplicas de repeticiones balanceadas o *Muestreo por Mitades* vienen a ser como si de cada repetición de GAR se escogiera una de las muestras. Al formar grupos aleatorios repetidos, se hacen grupos dependientes pues la pertenencia de una unidad a un grupo la excluye de otro, lo que conlleva que a pesar de realizar este proceso R veces, el sesgo del estimador no se elimina. Al aplicar repeticiones balanceadas, el sesgo por dependencia entre grupos no aparece ya que no se considera simultáneamente una media-muestra y su complemento. Cada réplica representa una aleatorización independiente por estrato, de ahí que el método no tenga las inconveniencias de GAR para hacer inferencias.

3. Para encontrar una coincidencia entre Grupos Aleatorios y el jackknife, basta considerar $\hat{\Theta} = \sum_{\alpha=1}^{K} \hat{\Theta}_\alpha / K$, es decir, se considera un jackknife en el que se eliminan grupos. El estimador de varianza de Grupos Aleatorios coincidiría con el jackknife si se tuvieran K grupos, donde cada uno es la muestra total menos uno de los subgrupos.

4. Cuando $n_h = 2$ el estimador jackknife requiere de $2L$ iteraciones (L =no. de estratos), lo que es mayor que las que se necesitan para ejecutar repeticiones balanceadas (Un múltiplo de 4 mayor que L). Sin embargo, Rust y Rao (1996) consideran que se puede realizar un jackknifè basado en L réplicas, con el siguiente estimador de varianza que sólo requiere calcular $\hat{\Theta}_{(h1)}$, el estimador de Θ quitando la unidad 1 de cada estrato, omitiéndose el cálculo de $\hat{\Theta}_{(h2)}$:

$$v_J = \sum_{h=1}^{L} (\hat{\Theta}_{(h1)} - \hat{\Theta})^2$$

5. El planteamiento bootstrap original, consiste en la simulación de selección de muestras a partir de la distribución empírica,$\hat{F}_n$; lo que en m.a.s. está dada por el vector $(\frac{1}{n}, \ldots, \frac{1}{n})$. Si se pensara en el jackknife, bajo el mismo contexto, se tendría que lo que se hace es obtener n estimaciones que corresponden a distintas funciones empíricas dadas por las permutaciones de

$$(0, \frac{1}{n-1}, \ldots, \frac{1}{n-1}).$$

Cada permutación del vector anterior es una estimación de $F_{(n-1)}$, que se utiliza para estimar algún parámetro de F_n. El bootstrap, sin embargo, se basa en $\hat{F}_n$ para simular muestras y estimar sus parámetros.

6. Relacionado al comentario anterior, en un muestreo complejo,la aplicación del jackknife lo que se hace es considerar todas las posibles formas en que ocurre que elementos del vector que se exhibe a continuación sean 0, de acuerdo a la definición en la sección 3.2.2. :

$$(\frac{1}{w^*_{111}}, \frac{1}{w^*_{112}}, \ldots, \frac{1}{w^*_{hij}}, \ldots, \frac{1}{w^*_{Ln_L m_{Ln_L}}})$$

7. Después de ver lo complicado de tantas aleatorizaciones que requiere el bootstrap, surgió la interrogante de cuán robusto y adecuado pudiera ser una versión bootstrap para muestreo complejo,

que hiciera un sólo proceso de aleatorización, basándose en el vector de la fracción de los factores de expansión de cada unidad, con respecto a la suma total de factores de expansión:

$$(\frac{w_{111}}{\sum w}, \frac{w_{112}}{\sum w}, \ldots, \frac{w_{hij}}{\sum w}, \ldots, \frac{w_{Ln_L m_{Ln_L}}}{\sum w}) , \tag{4.1}$$

(donde, $\sum W$ = la suma de todos los factores de expansión) en la muestra.

A este vector la información del diseño le es inherente. Es decir, esto representaría retomar el problema de muestreo complejo con el concepto original, pero considerando la probabilidad de selección dado el diseño. Claro está, existiría el inconveniente de que las muestras simuladas no tendrían el mismo número de unidades por estrato o conglomerado, pero quizás el estimador de varianza y el intervalo fuera robusto a este percance. Los estimadores a considerar deberían ser tales que se pudieran expresar en términos de los factores de expansión.

8. La extensión al caso multivariado se consigue, con cualquiera de los métodos mencionados, substituyendo la matriz

$$(\hat{\Theta}_i - \hat{\Theta})(\hat{\Theta}_i - \hat{\Theta})^T$$

por la suma de cuadrados de las diferencias entre el valor estimado en una iteración y el estimador global. Obviamente, en tal caso, $\hat{\Theta}$ es p-variado, e igualmente $\hat{\Theta}_i$, el estimador de una réplica o iteración.

4.5 Guía de decisión

Después de tantos comentarios, queda la interrogante de cómo elegir uno entre estos métodos. Es cierto que no existen reglas al respecto, incluso todos los autores sobre el tema dicen que no se puede determinar un mejor método, sino que la elección depende del diseño, el estimador y las condiciones de trabajo. No obstante, a continuación se trata de esbozar lineamientos para la decisión de elegir una de estas técnicas.

A. Si el diseño es estratificado con dos unidades de muestreo por estrato, es lógico utilizar el estimador de repeticiones balanceadas, con la adecuación de Fay. Este estimador será similar al bootstrap y requiere menos iteraciones que un jackknife o bootstrap. Además tiene la ventaja de que se comporta igual para estadísticas suaves o no suaves.

B. Si son necesarios muchos ajustes por no-respuesta o pos-estratificación es recomendable elegir entre el jackknife, bootstrap o repeticiones balanceadas.

C. Sólo si se cuenta con los recursos económicos y la necesidad de detectar otras fuentes de error, se recomienda elegir Grupos Independientes.

D. El método de linealización puede ser el más adecuado a aplicarse cuando se cuenta con un paquete de cómputo que lo resuelva, no sean muchos o muy diferentes los estimadores de interés, además de que entre estos estimadores no estén estadísticas de orden (o funciones de ellas) y no haya problemas serios de no-respuesta o pos-estratificación.

E. Si existe un verdadero interés por conocer una aproximación a la función de distribución del estimador, se recomienda indudablemente el bootstrap, aunque teniedo presente que ese estimador de la función está condicionado al diseño.

F. Por el momento, aplicar el jackknife para estimar estadísticas de orden (como los percentiles) y su varianza, sólo tendría propósitos de investigación, puesto que, como se mencionó, en el caso de poblaciones infinitas el jackknife ha probado se inconsistente para estas estadísticas, pero aún no se aprecia su comportamiento en un muestreo complejo, donde se van eliminando conglomerados.

G. Los últimos estudios [6] parecen indicar que de manera general el jackknife es el método con cualidades de tipo operativo que lo

[6] Ver Rao (1997).

hacen fácil de implementar para distintos diseños y situaciones -como la no-respuesta y la pos-estratificación- exceptuando el caso de las estadísticas de orden, que está bajo estudio.

H. Si entre los estimadores de interés se encuentra alguna estadística de orden, y no se tiene el diseño que rige en Repeticiones Balanceadas, ni se tienen grupos independientes, entonces se debería proceder con el bootstrap, o con grupos aleatorios repetidos el cual, en tal situación, es recomendado por Kovar, Rao y Wu (1988). También sería factible obtener un intervalo para dicha estadística con el método de Woodruff [7] el cual es específico para este tipo de estadísticas.

I. El método de Grupos Dependientes es el menos recomendado. La principal razón es el que no existe ningún fundamento para hacer inferencias.

[7] No se incluyó en la discusión de esta tesis pero se encontró citado en Kovar, Rao y Wu (1988).

Capítulo 5

ENAL'96

La Encuesta Nacional de Alimentación y Nutrición en el Medio Rural 1996, se llevó a cabo con el gran esfuerzo del Instituto Nacional de la Nutrición "Salvador Zubirán" (INNSZ) y el apoyo institucional en el trabajo de campo del DIF, IMSS-Solidaridad, Instituto Nacional Indigenista, Secretaría de Salud y los Gobiernos de los Estados y las Delegaciones de SEDESOL. Como se aprecia en la descripción de la encuesta que se ofrece en este capítulo, se plantearon cinco objetivos generales y para su consecución, la recaudación de información por vivienda fue muy extensa, abarcando 10 cédulas de interrogantes. Los resultados fueron expuestos por los coordinadores de la encuesta en un documento [1] que incluye tablas de las estadísticas descriptivas de todas las variables de interés, por estado y a nivel nacional, y cierta discusión sobre las implicaciones de tales cifras. Intervalos de confianza sólo fueron expuestos para los indicadores de desnutrición en los niños que son : peso para la edad, peso para la talla y talla para la edad.

Cuando se facilitó a la autora de esta tesis la base de datos, ya los resultados mencionados en el párrafo anterior se habían presentado; sin embargo, los responsables de esta encuesta manifestaron su preocupación por no haber podido extraer más información que apoyara cabalmente los objetivos. Se sostuvo una franca comunicación con ellos, para llegar a una propuesta de análisis que a la par, sirviera como

[1] Ávila A., T. Shamah, A. Chávez, Encuesta Nacional de Alimentación y Nutrición en el Medio Rural 1996, Instituto Nacional de Nutrición "Salvador Zubirán", documento interno.

la parte aplicativa de esta tesis y proporcionara elementos de apoyo al Instituto Nacional de Nutrición, en cuanto a la identificación de problemas nutricionales. Entre los aspectos discutidos, se le atribuyó mayor relieve -por considerarse primordial- el problema que existía para interpretar los resultados de los tres indicadores de desnutrición, situación ante la cual, resultó de mucha utilidad un documento de la Organización Mundial de la Salud[2], que los mismos coordinadores proporcionaron. Se planteó la idea de obtener estimadores de razón que correspondieran a categorías nutricionales construidas a partir de los tres indicadores internacionales, y ésta fue acogida con agrado, por lo que se prosiguió la fase de aplicación en este trabajo.

5.1 Objetivos generales de la encuesta

La ENAL'96 es la cuarta Encuesta Nacional de Alimentación en el Medio Rural Mexicano. Ésta y sus predecesoras se concibieron como un instrumento de diagnóstico para la planeación y evaluación de las acciones de bienestar social y combate a la pobreza. En particular, este tipo de encuesta, pretende llegar a definir la magnitud y características de los problemas nutricionales en la población de niños de cinco años y menores[3], en las zonas rurales de todos los Estados y áreas geográficas de México. Los objetivos generales de esta encuesta planteados por Ávila (1997) y colaboradores son los que siguen:

- **Conocer la situación actual alimentaria y nutricional en el medio rural mexicano.**

- **Comparar las condiciones actuales con las reportadas en las Encuestas de 1974, 1979, y 1989.**

[2]Organización Mundial de la Salud, *Measuring Change in Nutritional Status*, Ginebra, Suiza, 1983

[3]Todos los documentos sobre la ENAL'96, se refieren a **niños menores de cinco años**. Sin embargo, al examinar la base de datos y las estadísticas descriptivas que publicó el Instituto Nacional de Nutrición, se observa que en realidad se consideraron **niños menores de seis años**. Por otra parte, las edades de los niños se calcularon a partir de la fecha de nacimiento.

- **Identificar zonas en condiciones críticas de alimentación y nutrición.**

- **Identificar factores de riesgo asociado con desnutrición en las comunidades y familias estudiadas.**

- **Coadyuvar al establecimiento y perfeccionamiento de sistemas locales de vigilancia epidemiológica de la nutrición.**

5.2 Diseño muestral y descripción

Persiguiendo los objetivos ya planteados, la encuesta no sólo recabó los indicadores antropométricos de los niños, sino también aspectos que describen las viviendas y sus condiciones sanitarias, recursos para la alimentación de la familia, composición familiar y actividades de sus miembros, migración, información sobre las actividades agrícolas, datos del desarrollo reproductivo de las mujeres[4] y consumo familiar de alimentos. En total, la encuesta de una sola vivienda se compuso de diez cédulas u hojas de interrogantes; las cuales se vaciaron en una base de datos que consiste de once módulos (la última cédula, que es el recordatorio de lo que comió la madre de familia el día anterior, se dividió en dos módulos), a manera de archivos, que son tan anchos como el número de preguntas y tan largos como el número de respuestas, siendo su largo máximo 38,232. No todos los módulos tienen el mismo largo porque cada variable tiene un índice de no-respuesta diferente; además, el tamaño del módulo 7, el cual contiene la información de los niños, depende de los niños revisados y no de las viviendas encuestadas[5] Con repecto a este punto, también ocurre que no en todas las casa hay niños de la edad requerida en el estudio. En el apéndice D se muestra una copia de la cédula no. 7, en la que se captaron los datos de los

[4] En particular: edad actual, edad de la menarquía, número de embarazos, número de abortos, hijos nacidos vivos e hijos que aún viven.

[5] Tras revisar datos faltantes, incoherencias, o errores de captura, se tuvo un total de 26,700 registros de niños menores de 6 años, por lo que los estimadores que aquí se presentan se basan en un gran total de 26,700 datos.

niños menores de seis años. En esta tesis, sólo se muestra esa cédula porque el análisis que se presenta se basa en datos contenidos en dicha hoja de levantamiento de información. En ella se aprecian las siguientes preguntas:

1. ¿Cuántos niños menores de 5 años hay en esta familia?

2. ¿Quién es la madre del niño? (Se anota una clave que corresponde a la identificación de la madre en el módulo 4.

   De cada niño revisado, y para un máximo de tres niños se capta lo siguiente:

3. Nombre del niño.

4. Sexo

5. Fecha de Nacimiento.

6. Peso en kgs.(actual).

7. Talla en cm. (actual).

8. ¿Fue alimentado al seno materno?

9. ¿Durante cuantos meses?

10. ¿Fue alimentado regularmente con leche en biberón los primeros 12 meses?

11. Si la respuesta fue sí: ¿A qué edad inició?

12. ¿A qué edad recibió por primera vez otros alimentos distintos a la leche o líquidos?

13. ¿Ha estado enfermo los últimos 15 días?

14. Si la respuesta fue sí: ¿De qué se enfermó?
(1)Diarrea; (2)Infección respiratoria ; (3)Otra.

Como bien indica su nombre, la encuesta se dirigió a comunidades rurales. De acuerdo a Ávila y colaboradores (1997) se definió la población de interés como **las familias residentes en las localidades rurales del país cuya población era de 500 a 2500 habitantes, con población económicamente activa dedicada principalmente al sector primario, de acuerdo con la información del X Censo Nacional de Población y Vivienda, 1990 (INEGI)**. Aunque de manera general, esta definición es válida, se advierte que cuando las variables de interés se refieren a los niños (que es el caso de los estimadores obtenidos en esta tesis), la población a la que se extrapolan los resultados, está dada por **los niños menores de seis años, residentes en la localidades descritas anteriormente**.

El diseño muestral, que consistió en un estratificado bietápico por conglomerados en cada estado, fue realizado con la asesoría del Dr. Ignacio Méndez Ramírez. Hay que señalar que no se consideró a cada estado como un dominio de estudio. Sin embargo, sí se contemplaron 5 grandes regiones como dominios de estudio para los tres indicadores de estado nutricional: peso para la edad, peso para la talla y talla para la edad. Las cinco regiones o zonas son:

1. **Noroeste**: Baja California, baja California Sur, Sonora, Chihuahua, Sinaloa y Durango.

2. **Noreste**: Coahuila, Nuevo León, Tamaulipas, Zacatecas, San Luis Potosí y Aguascalientes.

3. **Centro Occidente**: Nayarit, Jalisco, Colima, Guanajuato, Querétaro, y Michoacán.

4. **Centro**: México, Hidalgo, Tlaxcala, Puebla, Morelos y Veracruz.

5. **Sur**: Guerrero, Oaxaca, Chiapas, Tabasco, Campeche, Yucatán y Quintana Roo.

El concepto básico para la construcción del diseño, fue establecer el mayor número posible de estratos, cada uno con pocas localidades encuestadas; en efecto originalmente se plantearon dos localidades por estrato pero las autoridades de algunos estados prefirieron que este número se aumentara en algunos estratos. Como el diseño contempló las zonas mencionadas como dominios, se hizo la advertencia de que sólo si las varianzas por estados eran pequeñas, entonces se arrojarían resultados por estado.

Sobre la determinación de los estratos y selección de localidades vale la pena citar a Barragán y colaboradores (1997), en la Cartografía de la encuesta (Se hace la observación de que en este documento muchas veces utilizan la palabra *región* en lugar de *estrato*):

Las regiones al interior de cada estado se construyeron respetando siempre los contornos de la división municipal, así como, en la medida de lo posible, las regionalizaciones político-administrativas locales. Habida cuenta de que la unidad mínima de desagregación fue el municipio, siempre que las cirscunstancias lo permitieron, se trató de identificar una región con un municipio. Esto fue posible en los estados de Tabasco, Baja California, Baja California Sur, Campeche y, prácticamente, Sinaloa. En el caso de Oaxaca, estado que contiene la cuarta parte de los municipios del país, se determinó conformar las regiones a partir de los treinta distritos en los que se divide tradicionalmente al estado para fines administrativos. En el resto del País, las regiones correspondientes se establecieron de acuerdo con la capacidad de levantamiento de encuestas de las instituciones participantes, procurando que cada región conformada tuviera alrededor de 30 mil habitantes en localidades rurales del rango requerido. Para la conglomeración de los municipios, de acuerdo con la estrategia del muestreo estratificado, se buscó que las regiones definidas fuesen los más homogéneas posible, para lo cual se utilizaron criterios socioeconómicos, geográficos, etnográficos, fisiográficos, de lugar central, continuidad territorial, y vías de comunicación.

De acuerdo con el *XI* Censo General de Población y Vivienda 1990, existen en el país 13,187 localidades rurales con población entre 500 y 2,500 habitantes, donde viven casi 13 millones de individuos. Desde hace casi 20 años la población rural del país ha permanecido prácticamente estancada por lo que no se consideró indispensable hacer proyecciones de crecimiento demográfico. Una vez definidas las regiones [estratos] y asignado el universo de las localidades a los 372 estratos, se procedió a seleccionar aleatoriamente al menos dos localidades por estrato.

Se observa que el diseño en cada estado es diferente en cuanto a número de estratos y muchas veces también, en cuanto a número de localidades muestreadas por estrato. Como se da a entender, la unidad primaria de muestreo estuvo conformada por las localidades, y las unidades secundarias por las viviendas dentro de cada localidad; de lo que se deriva que la unidad de aplicación de la encuesta fue la familia, siendo ésta definida como **el conjunto de personas unidas o no por parentesco, que viven bajo el mismo techo, comparten el gasto y la alimentación doméstica**. En cada estrato, se seleccionó, mediante muestreo aleatorio simple (m.a.s.) sin reemplazo, un mínimo de dos localidades, y en cada localidad se seleccionaron de la misma forma, 50 familias. Las familias que no quisieron participar, o que por alguna otra razón no se consiguió su respuesta, fueron excluidas de la muestra.

En el apéndice D se aprecia que en la base de datos existe una variable por familia que indica el número de niños menores de 5 años que viven en al casa[6], pero solamente hay espacio para tomar medidas de 3 niños. De aquí se desprende el hecho de que, al hablar de medidas o variables pertinentes a los niños, en realidad se dieron 3 etapas de muestreo puesto que hubo viviendas en las que no se recabó la información de todos los niños, bien porque alguno no estaba al momento de la

[6]Se señaló que los resultados de la encuesta se dieron para niños **menores de 6 años**, aunque todas los documentos hablan de los **menores de 5 años**. Se supone que se anotó como respuesta a esta pregunta el número de niños de 5 años o menores. Posteriormente se verá que esta variable es necesaria para estimar el tamaño de la población de niños que interesa en este estudio.

entrevista, o bien porque había más de tres. Cabe mencionar que no en todas las viviendas hay niños en el rango de edades que interesan, por lo que a la hora de obtener estimadores, los factores de expansión se construyeron con base al total de viviendas encuestadas, no al total de viviendas encuestadas donde hubo niños de la edad de interés; ya que de otra manera, la estimación de niños por vivienda estaría inflada.

A nivel nacional, la muestra se levantó en 31 estados[7], y se conformó de 372 estratos, 855 localidades, y un total de 38,232 encuestas a nivel familiar. En la tabla 5.1 se desglosa el número de estratos en cada estado, las localidades y el total de familias encuestadas. Se incluye también en la misma tabla, la población rural en los estados como referencia de la magnitud de la población bajo estudio. El dato de población rural se extrajo de la base de datos, sumando la población rural por estrato. En la Cartografía de la encuesta (ver Barragán y colaboradores (1997))se da un dato de población rural por estado que difiere levemente del que provino de la base de datos. De cualquier forma, se reporta que dicha cifra se basa en el XI Censo General de Población y Vivienda 1990.

Una observación importante es que en el estado de Michoacán hubo un estrato (el estrato 18) donde sólo se contó con información de una localidad (para efectos del análisis que aquí se presenta). Se decidió colapsar este estrato con el más próximo de los estratos que lo circundaban, por lo que fue el estrato 20 el que acogió la localidad encuestada en el 18. De tal suerte, este trabajo se basa en 371 estratos y 854 localidades, lo que quizás logre apreciar el lector en el apéndice TA, donde se muestran algunos programas en Fortran.

5.3 Indicadores internacionales sobre estado nutricional

Para entender los estimadores que se escogieron en este trabajo, es preciso entrar en una descripción de los indicadores internacionales de estado nutricional. A continuación se ofrece una explicación somera de

[7]La definición que se dió de la población de interés explica por sí sola, la razón por la cual no se consideró en la muestra al Distrito Federal

ESTADO	Pob. Rural	% Pob.	Estra-tos	Locs. enc.	Fams. enc.
Aguascalientes	98,150	0.75	5	10	466
Baja California	86,147	0.66	2	6	224
B.C. Sur	22,754	0.18	4	11	425
Campeche	109,826	0.85	8	24	1,173
Coahuila	139,885	1.08	5	15	678
Colima	40,628	0.31	4	12	514
Chiapas	951,969	7.33	25	57	2566
Chihuahua	183,117	1.41	4	9	439
Durango	289,059	2.23	8	21	947
Guanajuato	797,424	6.14	19	38	1,322
Guerrero	726,506	5.59	20	54	2,631
Hidalgo	601,492	4.63	15	32	1,430
Jalisco	460,792	3.55	12	35	1,718
México	1,010,980	7.78	27	54	2,322
Michoacán	818,561	6.30	20	40	1,876
Morelos	134,771	1.04	11	25	1,095
Nayarit	199,846	1.54	5	10	381
Nuevo León	66,424	0.51	3	9	447
Oaxaca	1,090,849	8.40	30	60	2,377
Puebla	1,017,802	7.84	25	59	2,738
Querétaro	237,067	1.82	9	18	759
Quintana Roo	78,418	0.60	3	9	429
San Luis Potosí	426,411	3.28	11	32	1,470
Sinaloa	439,349	3.38	14	33	1,307
Sonora	182,858	1.41	6	12	518
Tabasco	530,008	4.08	17	34	1,505
Tamaulipas	156,922	1.21	5	12	582
Tlaxcala	135,852	1.05	4	9	426
Veracruz	1,375,557	10.59	36	77	3687
Yucatán	196,386	1.51	6	17	811
Zacatecas	384,670	2.96	9	20	969
Nacional	**12,990,480**	**100.0**	**372**	**855**	**38,232**

Tabla 5.1: Esquema del diseño de la ENAL'96.

estas estadísticas.

5.3.1 Definición de los indicadores

La OrganizaciónMundial de la Salud (OMS), (OMS, 1983, págs. 11, 24) ha establecido tres medidas básicas para evaluar un estado nutricional. Éstas son: **peso para la edad**, **talla para la edad** y **peso para la talla** . El cálculo de dichos indicadores se desprende de las medidas de peso y talla y la ubicación del individuo en la población de referencia que le corresponde de acuerdo a su edad y sexo, o sexo y talla, según sea el caso. Se construyen estadísticas de manera similar a como se estandariza una variable normal, por lo que éstas se conocen como los "score z" de peso para la edad (PEDZ), talla para la edad (TEDZ) y peso para la talla (PETZ). Específicamente,

$$\text{PEDZ} = \frac{\text{Peso del individuo} - \text{Mediana del peso en la población de referencia } \mathbf{A}}{\text{Desviación Estándar (del peso) en la población de referencia } \mathbf{A}}$$

$$\text{TEDZ} = \frac{\text{Talla del individuo} - \text{Mediana de la talla en la población de referencia}}{\text{Desviación Estándar (de talla)en la población de referencia } \mathbf{A}}$$

$$\text{PETZ} = \frac{\text{Peso del individuo} - \text{Mediana del peso en la población de referencia } \mathbf{B}}{\text{Desviación Estándar(del peso) en la población de referencia } \mathbf{B}}$$

La población de referencia **A** es aquella que se compone de individuos con la misma edad y sexo del sujeto a quien se está evaluando, mientras que los que forman parte de la población de referencia **B** tienen su misma talla y sexo. Las distribuciónes de estas poblaciones son facilitadas por la OMS (1983); obviamente sus tablas corresponden a las estimaciones que resultaron de levantamientos de información, a nivel internacional.

Las poblaciones de referencia se han establecido para edades con precisión de meses y estaturas en decímetros (centímetros y un lugar decimal). Si se supone una distribución normal para cada una de las variables PED TED y PET [8], entonces las estadísticas PEDZ, TEDZ y PETZ pretenden ubicar en qué lugar de la población de referencia se halla el individuo. La Organización Mundial de la Salud (1983) indica que se deben establecer criterios para decidir cuáles individuos están con niveles bajos de nutrición. De manera general, esta organización considera que los valores de PEDZ, TEDZ y PETZ por debajo de 2 desviaciones estándar de la mediana (de la población de referencia) corresponden a los sujetos de bajos niveles nutricionales, pero de ninguna manera lo da por el criterio universal. También recomienda la inspección de individuos que están a entre una y dos desviaciones estándares por debajo de la mediana.

El INNSZ se encargó del cálculo de estos tres indicadores y los proporcionó en la base de datos. Por otra parte, categorizó las variables PEDZ, TEDZ y PETZ como se indica a continuación:

NORMAL "Score z" igual o mayor a -1.

LEVE "Score z" menor que -1 y mayor que -2.

MODERADA "Score z" menor o igual a -2 y mayor a -3.

SEVERA "Score z" menor o igual a -3.

Más adelante se aborda el significado que se puede atribuir a los distintos niveles de PEDZ, TEDZ y PETZ.

5.3.2 Interpretación de los indicadores

En la tabla 5.2 se reproducen los estimadores de la proporción de niños en la categoría NORMAL, que presentó el INNSZ en los resultados de la encuesta. Quien conoce el problema de nutrición en México se da cuenta que el indicador *talla para la edad* (TEDZ) produce resultados que se apegan a lo esperado en el sentido de que los estados de Chiapas,

[8] Haciendo a un lado el planteamiento de población fija en muestreo, es factible ver el problema con otro lente y pensar en estas variables como aleatorias.

Guerrero, Yucatán y Oaxaca están entre aquellos donde la proporción de niños "normales" es menor; mientras que el indicador *peso para la talla* (PETZ) parece arrojar resultados engañosos pues estados del Sur tienen niveles equiparables o mejores que los estados del Norte.

En el documento de la Organización Mundial de la Salud (1983) se ofrece una interpretación conjunta de las medidas, para determinar el estado nutricional de un sujeto en particular. Por ejemplo, se dice que un niño con PETZ normal, pero con índices de PEDZ y TEDZ bajos, está siendo actualmente alimentado de manera normal pero tiene historial de mala nutrición en etapas anteriores de su vida, o incluso en su gestación. Ésta y otras interpretaciones explícitas en este documento son:

PETZ	**PEDZ**	**TEDZ**	**Interpretación**
Normal	Bajo	Bajo	Nutrición actual normal pero con historial de desnutrición
Normal	Normal	Normal	Nutrición normal
Bajo	Bajo	Alto	Desnutrición actual seria
Bajo	Bajo	Normal	Desnutrición actual (moderada)
Bajo	Normal	Alto	desnutrición actual
Alto	Alto	Bajo	Problema de obesidad
Alto	Normal	Bajo	Actualmente sobrenutrido pero con historial de desnutrición

Aunque el desglose anterior no muestra todas las posibles combinaciones de los tres indicadores, resulta evidente que con la misma lógica es posible llegar a construir categorías que reflejen la situación nutricional en distintas zonas del País. Como criterio global, la OMS opina que todos los niños con *peso para la talla* (PETZ) bajo están sufriendo actualmente desnutrición y que el grado de ésta depende de los valores de los otros dos indicadores. Por otra parte, siempre queda al criterio de los expertos el determinar los puntos críticos para decir si un indicador es bajo o normal.

ESTADO	Talla para la edad % Normal	Peso para la edad % Normal	Peso para la talla % Normal
Aguascalientes	62.9	67.6	83.6
Baja California	77.7	83.5	87.5
B.C. Sur	62.5	77.6	80.8
Campeche	29.6	50.1	87.9
Coahuila	64.5	73.5	85.1
Colima	55.1	64.4	63.7
Chiapas	28.4	50.4	84.4
Chihuahua	62.1	67.2	77.6
Durango	61.3	76.4	81.8
Guanajuato	49.6	55.4	75.7
Guerrero	28.7	36.9	75.1
Hidalgo	36.7	54.0	84.3
Jalisco	69.2	75.3	78.6
México	43.9	56.8	82.0
Michoacán	48.5	69.0	86.3
Morelos	40.2	57.2	86.6
Nayarit	59.0	68.8	74.7
Nuevo León	58.5	64.7	73.4
Oaxaca	29.1	45.5	83.6
Puebla	33.5	46.7	77.3
Querétaro	43.1	56.3	82.7
Quintana Roo	35.1	51.8	83.9
San Luis Potosí	43.4	59.1	83.5
Sinaloa	64.8	73.6	82.1
Sonora	75.8	87.0	82.2
Tabasco	45.3	54.6	79.0
Tamaulipas	57.9	70.1	78.7
Tlaxcala	40.9	49.5	79.0
Veracruz	42.7	56.2	81.9
Yucatán	26.9	38.4	76.4
Zacatecas	57.6	66.2	82.4
Nacional	44.1	57.2	81.1

Tabla 5.2: Estimadores puntuales por Estado, y nacionales, de la proporción de niños en la categoría NORMAL según los tres indicadores de estado nutricional. (Estimaciones obtenidas por el INNSZ.)

5.4 Categorización

Siguiendo el razonamiento planteado en la sección anterior, se crearon categorías mediante la combinación de los niveles NORMAL, LEVE, MODERADA y SEVERA (de PEDZ, TEDZ y PETZ) construidos previamente por los expertos en nutrición; al respecto cabe señalar que no se hizo una primera y única categorización, sino que se pasó por varios intentos. Así es como en el primer acercamiento se crearon 6 categorías y al examinar las estadísticas descriptivas por grupo, se decidió desglosar más algunas de ellas. El segundo planteamiento consistió en ocho grupos de los que se obtuvieron resultados preliminares para someterlos a la consideración de los responsables de la encuesta; resultó entonces, el deseo de ellos de ampliar el intervalo de valores del nivel NORMAL. La justificación para el cambio de este criterio fue el evitar confusiones con el efecto de las etapas de crecimiento en los niños o posibles enfermedades, donde usualmente se observa una pequeña disminución en los indicadores; también se dió el hecho de que prefirieron ser más liberales admitiendo cierta desventaja, en cuanto a medidas de talla, que pudieran tener los grupos indígenas ante las poblaciones de referencia internacionales. De tal suerte, se consideraron los siguientes niveles de los indicadores:

NORMAL	"Score z" igual o mayor a -1.6.
LEVE	"Score z" menor que -1.6 y mayor o igual que -2.
MODERADA	"Score z" menor que -2 y mayor o igual a -3.
SEVERA	"Score z" menor que -3.

En el tercer acercamiento se construyeron ocho grupos nutricionales de la siguiente manera:

Grupo	Nivel de: PETZ	PEDZ y TEDZ
0	NORMAL	Ambos NORMAL
1	NORMAL	Uno de ellos es NORMAL y el otro no.
2	NORMAL	Ambos LEVE
3	No NORMAL	Al menos uno es MODERADA o SEVERA (Ambos no normales).
4	No NORMAL	Ambos NORMAL
5	No NORMAL	Uno de ellos es NORMAL y el otro no.
6	LEVE	Ambos LEVE .
7	No NORMAL (Ver siguiente col.)	Al menos uno de PEDZ, TEDZ y PETZ es MODERADA o SEVERA.

Para facilitar la interpretación de estos grupos se proporciona en la tabla 5.3 las estadísticas descriptivas básicas, obtenidas por m.a.s. con todos los datos de los niños encuestados (Un total de 26,700 niños menores de seis años, encuestados en 31 estados). Con ésta y otras tablas[9], suponiendo m.a.s., se identificaron características particulares de los grupos. Se vió que en el grupo 1 había un buen número de niños chaparritos y gorditos y se consideró adecuado el segregarlos. Es decir, se consideró la creación de un aglomerado formado por los sujetos con PETZ y PEDZ normal pero TEDZ no normal. Al hacer ésto, obviamente, quedó un grupo de niños con PETZ y TEDZ normal pero con PEDZ bajo; dada la amplitud del nivel NORMAL para los tres indicadores, fue perceptible el hecho de que en estos casos los valores de PETZ y TEDZ se encontraban en la frontera inferior de lo normal, por lo que, en conjunto, se podían considerar como casos de niños en situación nutricional cercana a lo normal. Por otra parte, se vió lo mismo sobre los menores del grupo 2. De ahí que posteriormente se decidiera reagrupar los casos del grupo 0, el 2 y los del grupo 1 que no eran "chaparritos gorditos", en una categoría que incluiría a los

[9]Se hicieron tablas de acuerdo a los niveles NORMAL, LEVE, MODERADA y SEVERA de los tres indicadores pero éstas no se presentan aquí. Al respecto de estas tablas es importante hacer énfasis en que se elaboraron con el paquete SPSS, como parte exploratoria y que estos estimadores obtenidos por m.a.s. no se consideran los definitivos o los más adecuados

Grupo N %	**Estadística**	**PETZ**	**PEDZ**	**TEDZ**
0	*Media*	0.216	-0.035	-0.243
	Mínimo	-1.6	-1.6	-1.6
N=13,150	*Máximo*	5.0	4.8	5.0
% = 49.25	*Desv.Est. de Media*	0.009	0.009	0.009
1	*Media*	1.011	-0.867	-2.484
	Mínimo	-1.6	-3.5	-5.0
N=6847	*Máximo*	5.0	2.7	-0.2
%= 25.64	*Desv.Est. de Media*	0.016	0.009	0.010
2	*Media*	-0.834	-1.824	-1.860
	Mínimo	-1.5	-2.0	-2.0
N=351	*Máximo*	-0.1	-1.7	-1.7
%= 1.31	*Desv.Est. de Media*	0.014	0.005	0.006
3	*Media*	-0.325	-2.279	-3.217
	Mínimo	-1.6	-4.9	-5.0
N=3840	*Máximo*	3.7	-1.7	-1.7
%=14.38	*Desv.Est. de Media*	0.013	0.008	0.013
4	*Media*	-2.308	-0.802	1.880
	Mínimo	-5.0	-1.6	-0.8
N=818	*Máximo*	-1.7	2.0	5.0
%= 3.06	*Desv.Est. de Media*	0.021	0.025	0.048
5	*Media*	-2.591	-2.355	-0.295
	Mínimo	-4.9	-4.8	-1.6
N=1022	*Máximo*	-1.7	-1.7	4.7
% =3.83	*Desv.Est. de Media*	0.024	0.016	0.034
6 N=0				
7	*Media*	-2.348	-3.261	-2.694
	Mínimo	-5.0	-5.0	-5.0
N =672	*Máximo*	-1.7	-1.7	-2.2
%= 2.52	*Desv.Est. de Media*	0.025	0.023	0.030

Tabla 5.3: Estadísticas descriptivas, por grupos, a nivel Nacional, de los indicadores PETZ, PEDZ y TEDZ, calculadas sin considerar diseño muestral.

"normales y casi normales".

Los individuos del grupo 3 son muy cortos de estatura —se vió que el 67% de los casos con TEDZ en nivel *severo* están en ese grupo— los índices de PETZ son normales pero el PEDZ es bastante bajo. Se consideró este grupo importante porque refleja problemas de desnutrición en etapas anteriores, y por las medidas de TEDZ y PEDZ, se considera que sufrieron seria desnutrición. El grupo 4 se distingue por incluir los niños que son muy altos pero flacos para su talla. Estos representan niños que están sufriendo desnutrición al momento de la encuesta. Muy similar a éste último es el grupo 5, donde se aglomeran niños que están bien en talla para la edad (TEDZ) pero son flacos para su edad y talla (en general). Ya que estos dos grupos tienen interpretaciones similares, se decidió posteriormente juntarlos en una sola clase. En la muestra no hubo ningún caso con todos los niveles leves, por lo que se dió un cero muestral en la categoría 6. Finalmente, el grupo 7 incluye los sujetos en estado de mayor gravedad, pues sus elementos tienen todos los indicadores muy bajos, advirtiendo la existencia de desnutrición en el momento de la encuesta y en etapas anteriores en la vida del niño.

Tal como se ha explicado en los párrafos anteriores, se construyeron otras categorías, a partir de las ya expuestas. Además de los criterios señalados se vió que era conveniente contar con menos clases y evitar estimadores con valores muy pequeños, ya que suelen ser inestables. Se tuvo especial cuidado para aglomerar los grupos 0, 2 y los "chaparritos gorditos" del grupo 1, obteniendo las estadísticas de orden de los estados, de acuerdo al grupo 0 (normales) y a la nueva categoría, y se detectó que no se daban cambios en la relación comparativa de los 31 estados (Los pocos cambios que se dieron fueron en estados con orden consecutivo) .

Quizás ha sido pesada la explicación de cómo se llegó a establecer la categorización, pero resultaba indispensable hacerlo para aportar una descripción del problema y entender su contexto; lo cual representa una fase del trabajo del estadístico para poder llegar a soluciones que tengan sentido. Es por ello que estas últimas secciones parecen haberse apartado del tema central de la tesis, a pesar de que el proceso en sí y la comunicación con los expertos en nutrición, engendraron experiencias didácticas para la autora.

Finalmente, lo que sigue de este trabajo, se basa en el análisis de los siguientes grupos:

Normales y Casi Normales: Abarca los grupos 0, 2 y los que restaron del grupo 1 que no forman parte de la siguiente categoría que se menciona aquí. Como consecuencia, incluye individuos con todos sus índices normales, o bien que tienen PETZ *normal*, PEDZ y TEDZ *leve*, o PETZ y TEDZ *normal*. Se considera que los niños que pertenecen a este grupo no sufren problema de desnutrición. Es importante hacer notar que en este grupo se pueden encontrar sujetos muy altos, corpulentos o con problemas de obesidad, puesto que el nivel *normal* en los tres indicadores en realidad incluye lo normal y más arriba de lo normal.

Bajitos Gorditos: Se incluyen los niños con PETZ y PEDZ *normal* pero TEDZ no normal. Una interpretación para este grupo es que se compone de lo sujetos que tuvieron desnutrición en etapas anteriores pero actualmente están sobrenutridos. Probablemente los niños de este grupo tienen una dieta no balanceada. Es posible que en este grupo también estén incluidos menores con problemas fisiológicos que causan su gordura, pero no se cuenta con elementos para distinguirlos.

Mal para su edad: Incluye los casos del grupo 3 (anteriormente expuesto). Los menores que componen este grupo tienen el PETZ *normal* pero los niveles de PEDZ y TEDZ son bajos, al menos uno de ellos *moderado* o *severo*. esta categoría representa los niños que han sufrido desnutriciónen etapas anteriores de su vida o en la gestación pero actualmente reciben una nutrición adecuada.

Mal para la talla: Esta clase abarca los grupos 4 y 5 descritos antes. Se compone de niños que están bien de etatura o son más altos de lo normal, pero su peso no es adcuado para su talla y a veces tampoco para su edad. En palabras comunes, son altos o de estatura normal pero flacos. Se entiende que los sujetos incluidos en este grupo están sufriendo desnutrición al momento de la encuesta.

Mal para la edad y talla: Esta última categoría acoge a los casos más graves, aquellos en que todos los niveles son bajos, y al menos uno de los tres indicadores es *Moderado* o *Severo*. La interpretación de estado nutricional para este grupo es que los menores aquí clasificados están sufriendo desnutrición actualmente y la sufrieron en etapas anteriores de su vida.

La conveniencia de este tipo de categorías estriba en que su interpretación facilita la detección de problemas de desnutrición. Es importante aclarar que si se quisiera hacer la misma clasificación en alguna encuesta futura, se recomienda considerar los grupos 0-8 descritos antes, con la salvedad de segregar del grupo 1 los niños bajos de estatura y gorditos (por lo que quedarían 9 clases). Posteriormente se puede ver si conviene resumir los resultados en cinco clases como se hizo en este trabajo.

Capítulo 6

Aplicación y resultados

En este capítulo se presenta la aplicación del jackknife y Repeticiones Balanceadas, para calcular la varianza de varios estimadores y los intervalos de confianza asociados a ellos. Las estadísticas de interés, como se mencionó en el capítulo 5, se refieren a la Encuesta Nacional de Alimentación en el Medio Rural, 1996 (ENAL'96). Se comienza por la exposición de los estimadores utilizados y posteriormente se muestran e interpretan los resultados.

6.1 Estimadores

La fase aplicativa de esta tesis consiste en la obtención de la varianza de estimadores de razón y de la mediana de los indicadores PEDZ, PETZ y TEDZ. Los estimadores de razón en cuestión, se refieren a la proporción de niños en cada uno de los cinco grupos señalados en el Capítulo 5, por estado, región y a nivel nacional. Las estimaciones de varianzas (de los estimadores de razón) por estado y nacional, se realizaron a través del jackknife y por medio de las fórmulas conocidas que aproximan la estimación del error cuadrático medio (ECM), que se derivan de la linearización. Así también, en los estados donde los estratos consisten de únicamente dos localidades o unidades primarias de muestreo, se aplicó el método de Repeticiones Balanceadas (*Balanced Half-Sampling*) con la adecuación de Fay y $\lambda = 0.5$; también se calculó con otros valores de λ (0.3 y 0.7) y sin tal adecuación, pero todos los resultados fueron tan similares que se decidió mostrar solamente un caso ($\lambda = 0.5$). Los estimadores por regiones se obtuvieron sólo por jackknife.

Es importante aclarar que la estimación de la varianza de la mediana tiene un carácter exploratorio, ya que como se indicó en 3.2.4., el jackknife es inconsistente para la varianza de las estadísticas de orden (y en general las no suaves), bajo un m.a.s. y población infinita, pero prácticamente no se conoce su comportamiento para estas estadísticas, en un muestreo complejo. Es por ello que solamente se exhiben las estimaciones de la mediana y su varianza en los estados en los que resultó factible aplicar Repeticiones Balanceadas, para lograr, de esta manera, cierto punto de comparación. (No se aplicó Repeticiones Balanceadas por región pues no hubo ninguna donde todos los estados tuvieran dos localidades muestreadas en cada estrato). Vale la pena añadir que hay que tener presente que las verdaderas varianzas son desconocidas, por lo que simplemente es posible apreciar cómo se asemejan o se alejan las varianzas de medianas jackknife y Repeticiones Balanceadas, contando también con la referencia de sus similitudes para los estimadores de razón y la media; siendo que ésta última se incluyó con el objetivo de ampliar el marco de referencia.

6.1.1 Estimadores de razón

La proporción de niños en determinada categoría se expresa como el total estimado de niños clasificados en ella, entre el total estimado de niños en la edad de interés que viven en comunidades rurales; o bien, como el promedio (estimado) por vivienda rural de niños menores de seis años en la categoría de interés, entre el promedio (estimado) por vivienda rural de niños de dicha edad. Como no se conoce el tamaño de dicha población, se requiere estimar igualmente el denominador, lo que conlleva que el estimador de interés sea de razón. A continuación se da cierta nomenclatura[1] para posteriormente, dar las expresiones de los estimadores. Como se realizan estimaciones por estado, debe quedar claro que los subíndices de la nomenclatura que sigue, en realidad corren

[1]La razón por la que algunos conceptos tienen doble nomenclatura es porque se quiso conservar la relación con los nombres de variables en la base de datos y también la relación con la nomenclatura propia de estadística.

de esta manera: $h = 1, ..., L^E$, donde L^E es el número de estratos en el estado E; $i = 1, ..., N_h^E$, donde N_h^E es el total de localidades en el estrato h del estado E; $j = 1, ..., M_{hi}^E$, donde M_{hi}^E es el total de viviendas en la localidad i-ésima del estrato h, del estado E. Sin embargo para simplificar un poco la notación se consideró que se podía obviar el superíndice E en las expresiones de los estimadores que siguen en esta sección y 6.1.2.

y_{hij}^c	El número de niños menores de seis años en la familia j-ésima de la localidad i-ésima del estrato h, que son clasificados en la categoría c
x_{hij}	El número de niños menores de seis años en la familia j-ésima de la localidad i-ésima del estrato h.
xe_{hij}	El número de niños menores de seis años **encuestados** en la familia j-ésima de la localidad i-ésima del estrato h.

$VIVENC_{hi} = m_{hi}$	Viviendas encuestadas en la localidad i-ésima del estrato h.
$LOCENC_h = n_h$	Localidades encuestadas en el estrato h.
$VIVIEN_{hi} = M_{hi}$	Total de viviendas en la localidad i-ésima del estrato h.
$LOCXESTR_h = N_h$	Total de localidades rurales que reunen los requisitos de inclusión en el estrato h. ($VIVIEN_{hi}$ y $LOCXESTR_h$ son conocidas.)
$VIVESTR_h$, $VIVEDO$	Total de viviendas en el estrato y en el estado respectivamente. (Valores conocidos.)

W_h Peso de un estrato de un estado (Se usó W_h = VIVESTR$_h$/VIVEDO$_E$)

$Fcasa_{hij} = \frac{x_{hij}}{xe_{hij}}$ Factor de expansión por la tercera etapa de muestreo implícita en el diseño

$FVIV_{hi} = \frac{VIVIEN_{hi}}{VIVENC_{hi}}$ Factor de expansión por vivienda (Segunda etapa de muestreo).

$FLOC_h = \frac{LOCXESTR_h}{LOCENC_h}$ Factor de expansión por localidad (Primera etapa de muestreo).

$POBRURAL_E$ Población rural en el Estado.

POB Población rural en la República Mexicana. ($POBRURAL_E$ y POB conocidos.)

$\Pi_E = \frac{POBRURAL_E}{POB}$ Peso del Estado E con respecto a población rural en el Estado y Nacional.

Se decidió utilizar un estimador de razón combinado a nivel *estado*, y un estimador separado a nivel *Nacional* cuya formulación se basa en dos etapas de estratificación, pues para este efecto, los Estados son considerados estratos. Cochran (1977, pág. 167) explica que cuando la razón en cuestión no es constante de estrato a estrato, el estimador separado es más presiso pues su sesgo se desvanece. Por otra parte, también recomienda el uso del estimador *combinado* cuando la muestra es chica y del *separado* cuando la muestra en cada estrato es grande. Con base en estas recomendaciones, y bajo la sospecha de que los estimadores tuvieran mayores diferencias entre estados que dentro de los estados, se consideró adecuado el uniformar el método de estimación por estado y a nivel nacional como se indicó al inicio de este párrafo. Por otra parte, a nivel región el estimador que se aplicó fue el combinado, puesto que era de esperar que los estimadores de los estados vecinos se comportaran de forma similar y además la muestra de cada estrato seguía siendo pequeña.

El estimador de razón combinado para la proporción de niños menores de seis años que viven en comunidades rurales y son clasificados en la categoría c, en un Estado E (o región) [2], de acuerdo al diseño muestral de la encuesta está dado por:

$$\begin{aligned}
\hat{R}_E^c &= \frac{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} FLOC_h\ FVIV_{hi}\ Fcasa_{hij}\ y_{hij}^c}{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} FLOC_h\ FVIV_{hi}\ x_{hij}} \qquad (6.1)\\
&= \frac{\frac{1}{VIVEDO_E}\left\{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} FLOC_h\ FVIV_{hi}\ Fcasa_{hij}\ y_{hij}^c\right\}}{\frac{1}{VIVEDO_E}\left\{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} FLOC_h\ FVIV_{hi}\ x_{hij}\right\}}\\
&= \frac{\overline{y}}{\overline{x}}
\end{aligned}$$

Hansen, Hurwitz y Madow (1953) dan la expresión anterior pero en Sukhatme y cols. (1984) se encuentra una forma del estimador en la que son explícitos los pesos de los estratos, W_h, de la siguiente forma[3]:

$$\hat{R}_E^c = \frac{\sum_{h=1}^{L} W_h \overline{\overline{y}}_{est}}{\sum_{h=1}^{L} W_h \overline{\overline{x}}_{est}} \qquad (6.2)$$

Siguiendo la misma nomenclatura y suponiendo que el peso de los estratos de un estado está dado en función de las viviendas; es decir, $W_h = \frac{\text{Viviendas rurales en el estrato}}{\text{Viviendas rurales en el Estado}}$, se deriva que [4] :

$$\sum_{h=1}^{L} W_h \overline{\overline{y}}_{est} = \sum_{h=1}^{L} W_h \left[\frac{1}{\mathrm{LOCENC}_h}\sum_{i=1}^{n_h}\frac{\mathrm{VIVIEN}_{hi}}{\mathrm{VIVESTR}_h}\ \mathrm{LOCXESTR}_h\ \overline{y}_{hi}\right] \qquad (6.3)$$

$$= \frac{1}{\mathrm{VIVEDO}_E}\left[\sum_{h=1}^{L}\frac{\mathrm{LOCXESTR}_h}{\mathrm{LOCENC}_h}\sum_{i=1}^{n_h}\frac{\mathrm{VIVIEN}_{hi}}{\mathrm{VIVESTR}_h}\ \mathrm{VIVESTR}_h \overline{y}_{hi}\right]$$

[2] El estimador para una región tiene la misma expresión, pero en tal caso se consideran el total de estratos en todos los estados que conforman la región, $L = \sum L_E$.

[3] La definición de $\overline{\overline{y}}_{est}$ se encuentra implícita en el desarrollo de (6.3).

[4] Nótese que en el desarrollo de (6.3) se encuentran implícitas las definiciones de $\overline{\overline{y}}_{est}$ y $\overline{y}_{hi}$.

$$
\begin{aligned}
&= \frac{1}{\mathrm{VIVEDO}_E}\left[\sum^{L}\mathrm{FLOC}_h \sum^{n_h} \frac{\mathrm{VIVIEN}_{hi}}{\mathrm{VIVENC}_{hi}} \sum_{j=1}^{m_{hi}} \mathrm{Fcasa}_{hij} y^c_{hij}\right] \\
&= \frac{1}{\mathrm{VIVEDO}_E}\left[\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} \mathrm{FLOC}_h\ \mathrm{FVIV}_{hi}\ \mathrm{Fcasa}_{hij}\ y^c_{hij}\right]
\end{aligned}
$$

El denominador se puede desarrollar de igual forma, recordando que no se considera en éste la tercera etapa de selección (Por lo que no se incluye el factor $Fcasa_{hij}$) y se llega a que ambas expresiones son iguales si el peso de los estratos se da en términos de los elementos para la segunda etapa de muestreo, que en este caso fueron las viviendas. Se debe observar que (6.1) representa la razón de las medias de y y x, por vivienda. Por otro lado, (6.2) se podría obtener con pesos W_h que fuesen función de otra variable, como por ejemplo, población rural. Sin embargo, en tal caso los dos estimadores no coincidirían. Cabe resaltar que los estimadores que se expresan en términos de los factores de expansión son de construcción más lógica y operativamente son más sencillos de aplicar.

El estimador de razón separado para la proporción a nivel Nacional de niños menores de seis años que viven en comunidades rurales y son clasificados en la categoría c, está dado por:

$$R^c_{MEX} = \sum_{E=1}^{31} \Pi_E\ R^c_E \tag{6.4}$$

Se observa que de acuerdo a la nomenclatura antes expuesta, los pesos de los estados se basan en la población rural. Igualmente, se pudo haber tomado otra variable para este efecto. Por ejemplo, era factible el dar los pesos por viviendas rurales, pero de existir diferencias en número de habitantes por vivienda en los estados, entonces resultaba más adecuado el considerar la población rural directamente, por lo cual se decidió usar este dato como ponderador. Cabría la duda de obtener este estimador separado que pondera los estados o considerar uno similar que se basara en regiones, en lugar de estados. Si se examinan los pesos en (6.4), se debe notar que a cada estrato que pertenece a un estado en particular, se le atribuye un peso a nivel nacional de Π_E/L_E, lo cual, en la mayoría de los casos, debe ser una buena aproximación a su peso verdadero. Si se considerara un estimador separado por regiones, como el que se mencionó arriba, en realidad se tendrían cinco grandes estratos (en la expresión del estimador) y se estaría igualando el peso de todos los estratos de una región. De tal suerte, se consideró más adecuado y menos riesgoso el incluir la mayor cantidad posible de estratos a nivel nacional (en la expresión del estimador), sin llegar a los 372 que en verdad se dieron (lo que evitó una programación especial del jackknife a nivel Nacional, además del programa que hace estimaciones por estado). En resumen, en la elección de este estimador, se involucró la filosofía de simplificar los métodos guardando la mayor coherencia posible.

En el Capítulo 3 se mostró que el método de linealizaciónproduce el estimador de la varianza para un estimador de razón que usualmente se ve en libros de texto. Así pues el estimador de la varianza de (6.1) de acuerdo a esta aproximación, y según lo plantean Hansen, Hurwitz y Madow (1953)[5], se encuentra de la siguiente forma:

$$\widehat{Var}(\hat{R}_E^c) = \frac{1}{\hat{X}^2}\sum_{h=1}^{L}\frac{LOCXESTR_h^2}{LOCENC_h}\, s_{ch}^2$$

$$\text{donde, } s_{ch}^2 = s_{chY}^2 + (\hat{R}_E^c)^2\, s_{chX}^2 - 2\hat{R}_E^c\, S_{chYX}$$

$$s_{chY}^2 = \frac{\sum_{i=1}^{n_h}\left(y'_{hi} - \frac{\sum_i^{n_h} y'_{hi}}{n_h}\right)^2}{n_h - 1}$$

$$y'_{hi} = FVIV_{hi}\left(\sum_{j=1}^{m_{hij}} Fcasa_{hij}\, y_{hij}^c\right)$$

$$x'_{hi} = FVIV_{hi}\left(\sum_{j=1}^{m_{hij}} xe_{hij}\right)$$

[5]Para quien revise esta referencia se advierte que el desarrollo presentado por los autores citados se avoca al caso de la "varianza relativa" que se define como la varianza del coeficiente de variación al cuadrado. Pero de este estimador se deriva con facilidad el estimador de la varianza del estimador de razón. El uso de la varianza relativa radica en que es una estadística transportable ("portable"), según Kish (1977).

$$\begin{aligned}\hat{X} &= \sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} FLOC_h\ FVIV_{hi}\ x_{hij}\\ s_{chYX} &= \frac{\sum_{i=1}^{n_h}\left(y'_{hi}-\frac{\sum_i^{n_h} y'_{hi}}{n_h}\right)\left(x'_{hi}-\frac{\sum_i^{n_h} x'_{hi}}{n_h}\right)}{n_h-1}\end{aligned}$$

Por otra parte, el estimador de la varianza de (6.4) se da en función de las varianzas en cada estado, de la siguiente manera:

$$\widehat{Var}(R^c_{MEX}) = \sum_{E=1}^{31}\Pi_E^2\ \widehat{Var}(\hat{R}^c_E) \tag{6.5}$$

Se aprecia en la expresión anterior que una vez que se consiguen los estimadores de varianza por estado, la estimación a nivel nacional es fácil de obtener, pues los pesos son conocidos. En realidad no se aplicó un jackknife a nivel nacional para obtener las varianzas, sino que se obtuvo la ponderación de las varianzas de los estados, $\widehat{V}_{E_J}(\hat{\Theta}_E)$, de esta manera:

$$\begin{aligned}\widehat{V}_{N_J}(\hat{\Theta}) &= \sum_{E=1}^{31}\Pi_E^2\ \widehat{V}_{E_J}(\hat{\Theta}_E)\\ &= \sum_{E=1}^{31}\Pi_E^2\ \left\{\sum_{h=1}^{L_E}\frac{n_h-1}{n_h}\sum_{i=1}^{n_h}(\hat{\Theta}_{E(hi)}-\overline{\hat{\Theta}}_E)^2\right\}\\ &= \sum_{E=1}^{31}\left\{\sum_{h=1}^{L_E}\frac{n_h-1}{n_h}\sum_{i=1}^{n_h}(\Pi\hat{\Theta}_{E(hi)}-\Pi\overline{\hat{\Theta}}_E)^2\right\}\end{aligned} \tag{6.6}$$

donde, $\hat{\Theta}_E$ es el estimador del estado; $\overline{\hat{\Theta}}_E$ es el promedio de las iteraciones jackknife en el estado E lo cual representa otro estimador a nivel estado) y $\hat{\Theta}_{E(hi)}$ es el estimador del estado E, al quitar la localidad i-ésima del estrato h. Ahora, interesa saber, cómo compara esta expresión a la de un estimador jackknife basado en tantas iteraciones como localidades en la Nación. Para establecer tal comparación es conveniente revisar las expresiones de varios estimadores. En primer lugar, de acuerdo a la definición del estimador separado en (6.4), el estimador nacional sería:

$$\hat{\Theta}_N = \sum_{E=1}^{31} \Pi \hat{\Theta}_E. \tag{6.7}$$

Alguien podría sugerir otro estimador nacional que se basara en la ponderación de los estimadores jackknife, de la siguiente forma:

$$\hat{\overline{\Theta}}_N = \sum_{E=1}^{31} \Pi \hat{\overline{\Theta}}_E. \tag{6.8}$$

O bien, siendo muy puntuales, se podría pedir el estimador jackknife dado por el promedio de todas las iteraciones a nivel nacional, en cuyo caso, el estimador nacional sería :

$$\hat{\overline{\Theta}}_{N_J} = \frac{\sum_{E=1}^{3} 1 \sum_{h=1}^{L_E} \sum_{i=1}^{n_h} \left(\sum_{e \neq E} \Pi_e \hat{\Theta}_e + \Pi_E \hat{\Theta}_{E(hi)} \right)}{(\text{ Localidades encuestadas})} \tag{6.9}$$

Ahora bien, si se realizara un jackknife para todo el País, la expresión de la varianza sería,

$$\sum_{E=1}^{31} \sum_{h=1}^{L_E} \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} \left(\hat{\Theta}_{E(hi)} - A \right)^2 \tag{6.10}$$

donde A pudiera ser alguna de las expresiones en (6.9), (6.8) o (6.7).

Por otra parte, ocurre que el estimador nacional, al quitar la localidad i-ésima es :

$$\hat{\Theta}_{N_{E(hi)}} = \sum_{e \neq E} \Pi_e \hat{\Theta}_e + \Pi_e \hat{\Theta}_{E(hi)}. \tag{6.11}$$

Con estos elementos, se puede volver a revisar la expresión en (6.6). Sumando y restando el primer término de (6.11) dentro del paréntesis de (6.6), se tiene,

$$\widehat{V}_{N_J} = \sum_{E=1}^{31} \sum_{h=1}^{L_E} \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} \Bigg(\left(\Pi_E \hat{\Theta}_{E(hi)} + \sum_{e \neq E} \Pi_e \hat{\Theta}_e \right)$$

$$
\begin{aligned}
& - \Big(\sum_{e\neq E} \Pi_e \hat{\Theta}_e + \Pi_E \hat{\overline{\Theta}}_E\Big)\Bigg)^2 \\
= & \sum_{E=1}^{31} \sum_{h=1}^{L_E} \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} \Big(\hat{\Theta}_{E(hi)} - B\Big)^2
\end{aligned}
\qquad (6.12)
$$

Se observa que B no es igual por definición a (6.7), (6.8) ni a (6.9). Aunque parezca muy reiterativo, se recuerda que $\hat{\overline{\Theta}}_E$, es el estimador del estado E, correspondiente al promedio de las iteraciones jackknife, en las que se quita una localidad a la vez. Como este estimador puntual es muy parecido al global, es decir, a aquél en que no se elimina ninguna localidad (se constató en los resultados de la aplicación), se tiene que $B \approx A$, y por lo tanto, se considera (6.6) una buena aproximación de (6.10). Por otra parte, se debe notar que si en (6.6) se consideran las diferencias con respecto a $\hat{\Theta}_E$ (el estimador global, que no es el promedio de iteraciones jackknife), entonces, $B = A$, y las dos formas de calcular la varianza son idénticas. Así pues, se tiene que el estimador que se aplicó a nivel nacional es muy similar al que hubiera requerido 854 iteraciones (pues hay 854 localidades) y en términos numéricos se pueden considerar iguales.

Sesgo relativo del estimador de razón

Otro aspecto del estimador de razón que merece discusión es su sesgo y estabilidad. Dada la definición del estimador de razón, resulta evidente que si el denominador se acerca a cero, los estimadores de varianza serán inestables. Kish (1987, pág. 133) recomienda verificar que el coeficiente de variación del denominador (que en esta aplicación es el promedio de niños menores de 6 años por vivienda en el estado, o el total de niños de esta edad en el estado) sea menor a 0.2 o 0.1. Dicha recomendación se basa en el hecho de que el *sesgo relativo* del estimador de razón, definido por:

$$\frac{|Sesgo(\hat{R})|}{\sigma(\hat{R})},$$

disminuye en tanto menor sea el coeficiente de variación del denominador. Des Raj (1980, pág.101) explica el origen de dicha relación, lo que se expone a continuación.

Sea $R = \frac{Y}{X}$, el valor poblacional y $\hat{R} = \frac{\overline{y}}{\overline{x}}$ su estimador; se considera,

$$Cov(\hat{R}, \overline{x}) = E(\hat{R}\overline{x}) - E(\hat{R})E(\overline{x}) = \overline{Y} - E(\hat{R})\overline{X}.$$

Con base en la expresión anterior para la covarianza, se llega a una expresión del sesgo, de esta forma:

$$\begin{aligned} \frac{Cov(\hat{R}, \overline{x})}{\overline{X}} &= R - E(\hat{R}) \quad , \\ Sesgo(\hat{R}) &= E(\hat{R}) - R = \frac{-Cov(\hat{R}, \overline{x})}{\overline{X}} \quad . \end{aligned} \tag{6.13}$$

Luego, ocurre que la covarianza de interés también se define como, $Cov(\hat{R}, \overline{x}) = -\rho(\hat{R}, \overline{x})\sigma(\hat{R})\sigma(\overline{x})$, donde ρ es el coeficiente de correlación lineal entre $\hat{R}$ y $\overline{x}$. Por lo que se sigue que,

$$\begin{aligned} Sesgo(\hat{R}) &= -\rho(\hat{R}, \overline{x})\sigma(\hat{R})\frac{\sigma(\overline{x})}{\overline{X}} \quad \text{y, finalmente,} \\ \frac{|Sesgo(\hat{R})|}{\sigma(\hat{R})} &\leq \frac{\sigma(\overline{x})}{\overline{X}} = C_{\overline{x}} \end{aligned} \tag{6.14}$$

De (6.14), y recordadndo la definición de Error Cuadratico Medio (ECM), se deriva que :

$$ECM \leq \sigma^2(\hat{R})\left(1 + C_{\overline{x}}^2\right) \quad . \tag{6.15}$$

Es decir, en la medida que aumenta $C_{\overline{x}}$, el coeficiente de variacion del denominador del estimador de razon combinado, el error cuadrático medio del estimador crece.

Con el desarrollo anterior debe resultar evidente la importancia de la verificación de los coeficientes de variación del denominador de (6.1), a nivel estatal y regional, pues de hecho, se encontró que en algunos estados no es adecuado hacer inferencias sobre estos estimadores. Al respecto, cabe otra aclaración, ya que como las razones que se estiman en este trabajo son en sí proporciones, puede suscitarse en el lector inquietud porque en algunas categorías (a nivel estado o región estos valores son muy pequeños (menores a 0.2); situación ante la cual, de haberse obtenido los estimadores de proporción, se hubiera señalado la falta de estabilidad de su varianza. No interesa aquí entrar en detalles del estimador de proporción, pero sí puntualizar que son dos estimadores diferentes y que el trabajo que aquí se presenta, como es lógico, ha atendido las advertencias que le corresponden al estimador de razón.

6.1.2 Estimador de la mediana

La estimación de la mediana de los indicadores de estado nutricional provee información para comparar una población determinada con la población de referencia. Cuando se dan muestreos periódicos, un cambio en la mediana puede significar mejoría o deterioro de una población, según sea el caso. En un corto período de tiempo, tras la implementación de un programa de ayuda alimentaria, por ejemplo, es usual apreciar un cambio en la mediana del Peso para la Talla, aunque los otros indicadores no cambien (OMS(1983, págs. 20, 32)). Si se quisiera llevar a cabo en futuras encuestas este tipo de inspección, sería indispensable un estimador de la varianza de la mediana. Sin embargo, en un muestreo complejo, la estimación de la varianza de la mediana no es tan evidente como la de la media, ya que la mediana es un cuantil, o estadística no suave.

Aunque se comentó al inicio del capítulo, se reitera que la estimación de la mediana se hizo en un plan exploratorio, pues se conoce poco sobre el comportamiento del jackknife al estimar su varianza en un muestreo complejo. A continuación se presenta el estimador de la mediana que se obtuvo en cada iteración jackknife o réplica de Repeticiones Balanceadas, basado en Kish (1972, pág. 572) y Kovar, Rao y Wu (1988, pág. 27).

De manera general, el estimador de la mediana está dado por $\tilde{x}$, que se define de la siguiente forma:

$$\tilde{x} = F_n^{-1}(0.5) = min\{z : F_n(z) \geq 0.5\},$$

donde, F_n es la distribución acumulada muestral.

La estimación de la mediana, como de cualquier otro cuantil, requiere del cálculo de la distribución muestral contemplando el diseño, en las cercanías del valor estimado. Para evaluar la distribución muestral en un punto z, se requiere de una variable indicadora, $I_{hij}(z)$ [6], tal que :

$$I_{hij}(z) = \begin{array}{ll} 1 & \text{si el elemento } hij \text{ es menor o igual a } z, \\ 0 & \text{en otro caso.} \end{array}$$

Luego,

$$F_n(z) = \frac{\sum_{h=1}^{L} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} FLOC_h \; FVIV_{hi} \; Fcasa_{hij} \; I_{hij}(z)}{\sum_{h=1}^{L} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} FLOC_h \; FVIV_{hi} \; Fcasa_{hij}} \qquad (6.16)$$

El primer valor para el cual se cumple que $F_n(z)$ es mayor o igual a 0.5, es el estimador de la mediana. Pero, en esta aplicación en particular, la precisión de los indicadores PEDZ, PETZ y TEDZ está dada en décimas, por lo que se interpoló, de manera similar a cuando se busca la mediana de datos agrupados (Ott, Mendenhall y Larson (1978), pag. 97), para evitar que todas las iteraciones resultaran en el mismo valor.

Así es como, si C es un valor tal que $F_n(C) < 0.5$, pero $F_n(C + 0.1) > 0.5$, entonces,

$$\tilde{x} = C + \frac{0.1}{F_n(C+0.1) - F_n(C)} \Big\{ 0.5 - F_n(C) \Big\}. \qquad (6.17)$$

Quiere decir que el estimador de la mediana tiene una precisión mayor a los datos de los indicadores, de acuerdo a la probabilidad acumulada hasta el valor C.

No existe una expresión definida para el estimador de varianza de la mediana en un muestreo complejo. La estimación de ésta se realizó mediante el jackknife y Repeticiones Balanceadas, siguiendo las expresiones correspondientes que aparecen en los capítulos 2 y 3. Más adelante, se comentará cómo comparan las estimaciones de ambos métodos, lo cual es de interés por la poca experiencia que se tiene con este estimador en un muestreo complejo.

---

[6] **La nomenclatura que aquí se usa se refiere a un diseño estratificado bietápico, lo que obviamente cambiaría para otro diseño.**

6.1.3 Estimador de la media

Se estimó la media de los indicadores PEDZ, PETZ y TEDZ en los estados donde habían dos localidades (conglomerados) encuestadas por estrato, para comparar las estimaciones de jackknife y Repeticiones Balanceadas, en cuanto a su similitud en el caso de este estimador y en el caso de la mediana. La media se estima como el cociente de la suma de los datos ponderados por los factores de expansión entre la sumatoria de los factores de expansión:

$$\hat{\bar{x}} = \frac{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} FLOC_h \; FVIV_{hi} \; Fcasa_{hij} \sum_{k=1}^{\eta_{hij}} y_{hijk}}{\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}}\sum_{k=1}^{\eta_{hij}} FLOC_h \; FVIV_{hi} \; Fcasa_{hij}} \qquad (6.18)$$

donde, η_{hij} es el número de niños encuestados en la vivienda j-ésima, de la localidad i-ésima en el estrato h; y_{hijk} es el valor de PEDZ, PETZ o TEDZ, según sea al caso, para el niño k-ésimo de la vivienda j, en la localidad i, estrato h; el resto de la notación corresponde a lo expuesto anteriormente en este capítulo. Como ya se dijo, se consideraron los estimadores de varianza Jakknife y Repeticiones Balanceadas.

6.2 Selección de método

En el capítulo 4 se resumieron los hallazgos que se han reportado sobre los métodos de remuestreo. En general, al jackknife se le han encontrado buenos atributos en el caso de estadísticas suaves, como es la fundamentación teórica para establecer intervalos de confianza y su consistencia. También resulta más sencilla su programación que la del bootstrap, además de que su estimador se ajusta fácilmente a situaciones de pos-estratificación, que aunque no se da en la ENAL'96, es interesante tener la noción práctica de su operación. Por otra parte, el estimador por fórmulas, que provienen de una linearización, es el más difundido en libros de muestreo, y el que proveen algunos paquetes. Por este motivo, se decidió aplicar el Jakknife para el caso de los estimadores de razón de la proporción de niños por grupo nutricional, además del estimador de fórmula, y así verificar la similitud entre ellos.

El método de Repeticiones Balanceadas (*Balanced Half- Sampling*) tiene ventajas similares al jackknife pero se adscribe al diseño de dos conglomerados por estrato. La generalización a otros diseños es posible pero requiere de la colapsación de conglomerados. En el caso de estadísticas no-suaves no tiene el problema de inconsistencia que el jackknife. Se sabe además que su estimador es el más parecido al bootstrap, así como el estimador de linealizaciónse parece al jackknife. De tal suerte, se decidió aplicar el estimador de Repeticiones Balanceadas en los ocho estados donde el diseño era adecuado para ello: Aguascalientes, Guanajuato, México, Nayarit, Oaxaca, Querétaro, Sonora y Tabasco.

La estimación de la mediana y su varianza, se realizó como estudio exploratorio. Se llevó a cabo en los estados arriba mencionados para contar así con el estimador de Repeticiones Balanceadas junto con el jackknife.

6.3 Resultados

6.3.1 Determinación de estados en los que no son válidas las inferencias

Se consideró conveniente el empezar a hablar de resultados distinguiendo aquellos estados en los que se pueden hacer inferencias sobre los estimadores de razón combinados, de manera que el lector tenga conciencia de ésto al ver las tablas. De hecho, ellas incluyen los estimadores de todos los estados porque es necesario dar a conocer estos resultados. Se recordará, que en el capítulo anterior se dijo que el plan de muestreo se realizó considerando cinco grandes regiones como dominios de estudio para los tres indicadores de estado nutricional. De ahí que aunque los estados no fueron diseñados como dominios, teniendo las precauciones que aquí se señalan, se puede determinar si es válido hacer inferencias en algunos de ellos. Como se vió en (6.14), cuando el coeficiente de variación del denominador del estimador de razón combinado es grande su sesgo relativo se incrementa y por ende su error cuadrático medio. Kish (1987) recomienda verificar que este coeficiente sea menor a 0.1 o 0.2. En las tablas E.1 o E.2 se incluye la información sobre este coeficiente en los estados. Se aprecia que en 5 estados el valor de este coeficiente excede 0.2, los cuales son: **Baja California Sur, Chihuahua, Nuevo León Quintana Roo y Tlaxcala**. Por lo tanto, se debe de tener presente que no es adecuado hacer inferencias en estos estados. Por otra parte, se puede decir que hay 11 estados de la República en los cuales no existe ningún problema para hacer inferencias sobre la proporción de niños en las categorías de nutrición porque este coeficiente tiene un valor menor a 0.1. Éstos son: **Aguascalientes, Chiapas, Guerrero, Jalisco, México, Oaxaca, Puebla, San Luis Potosí Sinaloa, Tabasco y Veracruz**. En los 15 estados restantes el coeficiente tiene un valor menor a 0.2 y mayor a 0.1, por lo que según Kish(1987), podría considerarse un criterio más liberal que el de 0.1 y considerar también las inferencias en tales estados. (Éstos últimos son: **Baja California, Campeche, Coahuila, Colima, Durango, Guanajuato, Hidalgo, Michoacán, Morelos, Nayarit, Querétaro, Sonora, Tamaulipas, Yucatán y Zacatecas**).

Sin embargo, en una encuesta se dan otros problemas y los criterios, muchas veces, no se pueden circunscribir a un aspecto. El INNSZ reportó que detectó problemas de captura o confiabilidad de datos en los estados de Baja California Sur, Baja California, Colima y algo en Quintana Roo. Los problemas más serios fueron en Baja California Sur, pues de hecho, al hacer este trabajo se identificaron más problemas en los datos. Por tal motivo, en los estados mencionados se deben guardar reservas en cuanto a sus resultados; no sólo para los estimadores de razón sino para todos los que se fuesen a obtener.

6.3.2 Resultados sobre los estimadores de razón

La varianza de los estimadores de razón obtenida por fórmula se consiguió elaborando un programa en SPSS (versión 6.0), el cual se incluye en el Apéndice F. La autora se percató de que son muchos los detalles a cuidar para calcular esta varianza por fórmula, y si otra persona hubiera sido quien lo programara, hubiera sido más difícil la corrección de errores. Por otro lado, la varianza jackknife se obtuvo desarrollando un programa en Fortran32 (ver Apéndice G). Se realizaron los cálculos de acuerdo al ajuste de factores de expansión que propone Rao (1996), y de acuerdo al método de Wolter (1985) que considera explícitamente los factores de correción por población finita y la obtención de Pseudo-valores. En la tabla E. 2 se aprecia que ambos estimadores jackknife proveen resultados muy cercanos. El único estado que presenta una diferencia importante entre los dos estimadores es Hidalgo. Sin embargo, el estimador de linealización, que sí considera factores de corrección por población finita fue muy parecido al de Rao, aún en Hidalgo.

La tabla E. 1 muestra los estimadores jackknife, los obtenidos por linealización (o fórmula) y los estimadores de razón que asumen un muestreo aleatorio simple; de donde se deriva el efecto de diseño, o DEFF, que también se exhibe. En todos los estados y estimadores es evidente la similitud de la varianza jackknife y de la varianza obtenida por fórmula. Se aprecia que el DEFF varía entre estados y entre distintas variables para un mismo estado. Al respecto se ve que en los estados cuyo coeficiente de variación del denominador (del estimador de razón combinado)es grande, tiende ha haber algún DEFF menor a 0.5. Otros estados que presentan DEFF muy pequeños son Baja

California, Colima, Durango y Sonora; siendo que los coeficientes de variación de Colima y Sonora son cercanos a 0.20 pero los de Baja California y Durango son de 0.1447 y 0.1239. (Se debe recordar, que como se vió en (6.14), el sesgo relativo del estimador de razón combinado disminuye en la medida en que el coeficiente de variación del denominador es menor).

En los estados donde se muestrearon dos localidades en cada estrato, se obtuvo también la varianza de acuerdo a Repeticiones Balanceadas (*Balanced Half-Sampling*), ajustando un balance ortogonal completo y aplicando la adecuación de Fay con factor de 0.5. En el Apéndice G se muestra el programa en Fortran32 que se desarrolló para este propósito. Cabe mencionar que se consideró también la varianza dada por el complemento de la matriz Hadamard utilizada (Se invirtieron los 1's por -1, y viceversa), y al promedio de ambas varianzas. Las matrices Hadamard en las que se basó la estimación fueron las dadas por Wolter(1985), teniendo cuidado de no usar las columnas de unos, para poder lograr el balance ortogonal.

Más adelante, en este capítulo, se encuentra la tabla 6.1, en la que se pueden apreciar las estimaciones de razón jackknife y de Repeticiones Balanceadas (R.B.). En la tabla 6.4 se aprecia el cociente de ambos estimadores de varianza (Var(jackknife) / Var(R.B.)). En los estados donde el número de estratos es mayor, y por tanto así lo es el número de iteraciones o réplicas, ambos estimadores son muy cercanos. No ocurre ésto en Nayarit, Querétaro y Sonora, estados con menos de 10 estratos. Si el teorema de Krewski y Rao que se expuso en 1.6 tuviera efecto en esta aplicación, se encontraría que este resultado es coherente , pues uno de los supuestos en que se basa la demostración de la consistencia de estos estimadores de varianza para estadísticas suaves, es que se tienen un número de estratos grande y pocas unidades primarias de muestreo en ellos. Sin embargo, en esta aplicación se tiene un m.a.s. sin reemplazo de unidades primarias, lo cual no cumple con otro de los requerimientos de este teorema (m.a.s. con reemplazo de las unidades en los estratos).

Otro aspecto que se observa es que en Aguascalientes, a pesar de tener 5 estratos, los estimadores no resultaron tan diferentes (el jackknife es aproximadamente 1.5 veces el de R.B.), lo cual se deba a que su coeficiente de variación es menor a 0.1; mientras en los otros estados mencionados con pocos estratos, la varianza jackknife es en muchos casos, más de 10 veces la de R.B., siendo que sus coeficientes de variación son cercanos a 0.2. Es decir, pareciera ser que aún cuando el número de estratos fuera chico, si el coeficiente de variación del denominador del estimador de razón es pequeño, las varianzas obtenidas por ambas técnicas tenderán a parecerse. En relación a ésto, se encontró un comentario de J.N.K. Rao en un artículo de Kish (1977), diciendo que cuando más grande es dicho coeficiente de variación, tiende a haber diferencias significativas entre Repeticiones Balanceadas y linealización.

Por otra parte, en el Estado de México y Oaxaca los estimadores son prácticamente iguales y se encuentran dos factores que pueden influir en ello: ambos tienen muchos estratos y coeficientes de variación del denominador menor a 0.1. En los demás estados se observa la tendencia a que la varianza jackknife sea mayor, lo que repercute en intervalos más amplios. Este resultado es contrario a lo que encontraron Kovar, Rao y Wu (1988) y lo cual se platica en 4.3., ya que en su estudio, ellos vieron que las varianzas de Repeticiones Balanceadas eran más conservadoras. No obstante, en este caso se tienen situaciones diferentes, pues en el estudio citado se tenían 32 estratos y un muestreo estratificado simple y aquí ocurre que los estados donde se observa que la varianza jackknife es mayor, es en los que tienen pocos estratos.

Es importante traer a colasión otro resultado del estudio de Kovar, Rao y Wu (1988, pág. 35-37). Para el caso del etimador de razón, (en poblaciones simuladas de 32 estratos, con $n_h = 2$), se vió que a medida que el coeficiente de variación del denominador crece (> 0.1), R.B. y bootstrap tienden a sobreestimar la varianza. Llama la atención que en la aplicación que aquí se ha realizado, cuando el número de estratos es chico, parece que ocurre lo inverso, pues el estimador jackknife es mayor al de R.B. Se deriva de esta observación que en estudios futuros se pudieran considerar distintas combinaciones de coeficientes de variación y número de estratos para comparar los estimadores de métodos de remuestreo, considerando su sesgo y estabilidad.

En la tabla 6.13 se encuentran los resultados de estos estimadores por regiones. Como se mencionó en le capítulo 5, estas regiones fueron consideradas como dominios de estudio, para los tres indicadores, al hacer el plan de muestreo. Como las categorías que nos interesan se basan en estos indicadores, es lógica la consecuencia de que los estimadores de razón de proporciones en tales grupos sean más precisos a nivel región. Lo primero que salta a la vista es que los coeficientes de variación son menores a 0.1 en las cinco regiones. Aunado a ésto, los intervalos de confianza son más precisos que los obtenidos por estado (de manera general). Otro factor que beneficia una buena estimación a través del jackknife es que cada región se compone de una gran cantidad de estratos. No se pudo contar con la estimación de Repeticiones Balanceadas por región porque no se dieron las condiciones adecuadas de diseño. Tampoco se consideró de interés obtener la estimación de linealización (fórmulas) por regiones porque ya a nivel estado se había observado una gran similitud; en realidad, a nivel región no se estableció comparación alguna, sino simplemente, se quisieron obtener las estimaciones puntuales e intervalos, para propósitos inferenciales en los dominios de estudio que se establecieron en el diseño.

6.3.3 Resultados del estimador de la mediana

Las estimaciones de la mediana y su varianza se obtuvieron a través de programas en Fortran32, los que se pueden apreciar en el Apéndice G. Sólo se incluyen resultados de los estados en que habían dos localidades encuestadas por estrato, para poder comparar la varianza jackknife con la de R.B. En la tabla 6.2 se exhiben estos resultados. Conviene señalar que debido a que la precisión de los datos es de décimas, y que se tuvo que caer en una interpolación para estimar la mediana, los resultados se presentan con presición de milésimas y no con la misma de los estimadores de razón.

Como no se tiene referencia de los valores poblacionales, no es posible determinar si la varianza jackknife de la mediana en un muestreo complejo como el de esta aplicación, es adecuado o no. Sin embargo, como la varianza de la mediana de R.B. no tiene los problemas que la jackknife presenta en el marco de m.a.s. en población infinita, se pretendió comparar el desempeño de ambas varianzas con otros estimadores, como son los estimadores de razón ya mencionados y la media de los indicadores nutricionales (PEDZ, PETZ y TEDZ). La tabla 6.3 contiene los resultados de la estimación de la media a través de las dos técnicas. Si se observa nuevamente la tabla 6.4, se aprecia que las estimaciones de varianza de las medias comparan de forma similar al comportameinto de éstas en los estimadores de razón. Además se nota que la estimación jackknife de la varianza suele ser mayor y que en los estados con más estratos la similitud entre ambas estimaciones es muy grande.

Lo que ocurre en el caso de la mediana es que esa disyuntiva entre los estados con pocos o muchos estratos se hace más marcada. Donde el número de muestra es mayor la mediana parece ser tan estable (en cuanto a que se asemejen los métodos) como los otros estimadores suaves. En la tabla 6.4 se observa que en los estados con mayor número de estratos, Oaxaca y México, el estimador de varianza jackknife tiende a ser un poco menor al de Repeticiones balanceadas, lo que hace pensar que la relación encontrada por Kovar, Rao y Wu (1988) en su estudio[1], se de básicamente, en las situaciones en que hay muchos estratos. Estos autores encontraron que el estimador jackknife era más preciso y el de Repeticiones Balanceadas, más conservador. Cabe señalar que el estimador jackknife de la mediana de PEDZ en Oaxaca es cerca de la mitad del estimador R.B., los intervalos de confianza son muy similares. Pero sobre los intervalos se habla a continuación.

6.3.4 Intervalos de confianza

En la tabla E.3 se exhiben los intervalos de confianza de dos colas, con $\alpha = 0.05$, de los estimadores de razón de la proporción de niños por grupo nutricional, para todos los estados de la República Mexicana. Los intervalos se basan en $(n - L)$ grados de libertad, es decir el número total de localidades encuestadas en el estado menos el número de estratos. Cuando el diseño es de dos conglomerados por estrato, los grados de libertad están dados por el número de estratos. Se aprecia que la tabla E.3 también incluye los intervalos que se obtendrían si se hiciera caso omiso a los grados de libertad que corresponden a un método de remuestreo y se basaran en la $N(0,1)$, lo que en ocasiones puede repercutir en ciertas inferencias, ya que los intervalos basados en la $N(0,1)$ son más cortos.

La tabla 6.5 de este capítulo presenta los intervalos de confianza de los estimadores de razón, de acuerdo a la técnica jackknife y la de R.B.

[1] Ver sección 4.3 de esta tesis

Se observa que en los estados con pocos estratos, y por ende, pocos grados de libertad, se incurre en el problema de límites inferiores menores a cero, en particular en categorías con proporción muy baja. Casi invariablemente, cuando se da esta situación, son los dos métodos los que incurren en ella; a excepción del intervalo para los niños *Mal para la edad y talla* en Sonora, en cuyo caso, el dado por el jackknife tiene límite inferior mayor a cero, lo que no ocurre con el de R.B. Es importante ver que en la medida en que el coeficiente de variación del denominador del estimador de razón es menor (menor a 0.1), los intervalos dados por ambos métodos son más parecidos. De tal suerte, los intervalos para Sonora, Querétaro y Nayarit dados por el jackknife son más amplios que los que se obtuvieron por R.B. y los de Aguascalientes, México, Oaxaca, y Tabasco (estados donde el coeficiente fue menor a 0.1) son prácticamente iguales. El coeficiente de Guanajuato fue de 0.1064 y los intervalos de ambas técnicas también son bastante parecidos. Llama la atención el caso de Aguascalientes, pues aunque los grados de libertad son pocos, la coincidencia de los intervalos obtenidos por los dos métodos es también muy buena. Esto da a pensar que cuando el coeficiente de variación (del denominador del estimador de razón) es chico, la congruencia entre jackknife y R.B. se da aún con tamaño de muestra pequeño.

Se formularon los intervalos de confianza de dos colas, con un α de 0.05, para la mediana según la estimación jackknife y R.B., como se ve en la tabla 6.6. Conviene recordar que precisión de los indicadores en la base de datos es de décimas, por lo que ahora resulta más evidente que los estados con mayor número de grados de libertad: Tabasco, Guanajuato, México y Oaxaca, son los que presentan clara congruencia entre ambos métodos.

6.3.5 Interpretación y discusión general

Para facilitar la interpretación se elaboraron las tablas 6.7 - 6.11, en las que se ven los estados ordenados de acuerdo al límite superior de los intervalos de confianza jackknife. En la tabla 6.7 se observa que los estados con coeficiente de variación del denominador mayor a 0.2 tienen una longitud de intervalo mayor a la de los estados con coeficiente menor a 0.1.(los cuales, en general, son más cortos de 0.1).

Igualmente, en la tabla 6.8 se nota que el largo de intervalo de los estados con coeficiente grande es aproximadamente de 0.12 a 0.2, pero en aquellos donde el coeficiente es menor a 0.1 el largo de intervalo es menor (o alrededor) de 0.08. De manera similar se pueden inspeccionar las tablas 6.9-6.11 y percatarse de que en efecto, cuando el coeficiente de variación del denominador del estimador de razón es grande, los intervalos son más amplios, y muchas veces llegan a ser ridículos; tal como es el caso del intervalo para Baja California Sur en la tabla 6.9, que se da de 0 a 0.30 aproximadamente (Se debe recordar además que el límite inferior se acotó en cero cuando resultaba ser menor.) Si se consulta la tabla 6.12 se advierte que los estados con algunos intervalos demasiado amplios son los que contaron con menor tamaño de muestra. Son evidentes los problemas en Baja California Sur (donde, de hecho, se reportaron problemas de confiabilidad o extravío de datos), al estimar la proporción de niños *Mal para la edad, Mal para la talla y Mal para edad y talla.* Como estas categorías son las que representan los niños en peor estado nutricional, es claro que no se deben arrojar conclusiones sobre este estado, debido a la magnitud de sus intervalos. Se recomienda aumentar el tamaño de muestra en este estado en futuras encuestas (de querer hacer inferencias para el estado) y supervisar de manera más cerca la captura de sus datos.

Otros estados que presentan intervalos demasiado amplios para la categoría *Mal para la talla*, son Chihuahua, Nayarit, Tamaulipas Nuevo León, Colima y Baja California. Todos ellos tuvieron un tamaño de muestra chico (ver tabla 6.12); como se dijo antes, el INNSZ reportó también problemas de confiabilidad de datos y de captura en Colima y Baja California (aunque aquí fue menor el problema que en Baja California Sur) . En Chihuahua se presenta además la situación del cero muestral en la categoría *Mal para la edad y talla.*

Ya que la categoría *mal para la talla* se refiere a niños en mala situación en la actualidad, se recomienda hacer esfuerzos en los estados arriba mencionados por incrementar su tamaño de muestra en encuestas futuras y por cuidar la calidad de la información. El estado de Quintana Roo es un caso similar a los anteriores pues presenta algunos intervalos demasiado amplios, como ocurre con el de niños *Bajitos garditos* y los que están *Mal para la edad.*

En general , si se observa la tabla E.3, se ve que los intervalos que se basan en la $N(0,1)$ *son más estrechos* y *el tipo* de problema que se está planteando no es tan evidente como cuando se usan los intervalos calculados a partir de una distribución t *con grados* de *libertad* que varían según el número de estratos y localidades.

La tabla 6.13 resume los resultados más importantes para propósitos inferenciales, pues en ella se exhiben los resultados obtenidos mediante el jackknife para las cinco regiones del país que se planificaron como dominios de estudio. Examinando los intervalos de confianza al 95%, resulta notorio el hecho de que en los estados del Noroeste y Noreste la proporción de niños normales es mucho mayor que en el Centro y Sur, mientras que la proporción de Centro Occidente es similar a la que se da en el Noreste. Se observa la tendencia, cuanto más al Sur, a una mayor proporción de niños bajitos gorditos. Esto merece una adecuada interpretación de expertos en nutrición o quizás en antropología, a fin de determinar, qué tanto se puede atribuir a diferencias étnicas o de otra índole, o si hay algún problema de mala nutrición escondido tras estas cifras (no falta de nutrición, pero sí malos hábitos alimenticios). El mismo patrón de comportamiento se observa en la categoría *Mal para la edad*, es decir, los estados del Sur y Centro tienen mayor proporción que los del Norte y Centro Occidente. Por otra parte, se vuelve a notar similitud entre el Centro Occidente y Noreste. En este grupo se clasificaron niños que reflejan que tuvieron desnutrición en etapas anteriores de su vida. Por ello, hubiera resultado interesante analizar este grupo cuando menos, por edades; además de analizar las variables captadas sobre la alimentación del niño en sus primeras etapas (lactancia y ablactación). También puede revelar la necesidad de incluir en futuras encuestas datos sobre la alimentación de mujeres embarazadas.

Los resultados de la categoría *Mal para la talla* son preocupantes pues reflejan una situación similar en todo el País. Al respecto se debe recordar que se consiera que esta categoría incluye a los niños que están sufriendo desnutrición al momento de la encuesta. Por otra parte, la proporción de niños en peor situación, agrupados bajo el nombre *Mal para la edad y talla*, vuelve a ser mayor en los estados del Centro y Sur.

Al final de la tabla E.2, se aprecian los resultados nacionales, por categorías, basados en el estimador de razón separado. Aunque es importante contar con estas cifras globales, se considera que los resultados más importantes a nivel nacional, en términos interpretativos, se desprenden de mejor manera de las estadísticas por regiones o zonas.

Sobre la utilización de los métodos de remuestreo en esta encuesta, en particular el jackknife, cabe mencionar que fue mucho más rápida y confiable que la estimación por fórmulas. Al decir confiable me refiero a que los errores en su programación fueron de mucho menor trascendencia que los que se dieron en la obtención de varianzas por fórmulas. Esto se traduce en que es más factible entrenar a un equipo de técnicos o programadores que colaboren en el análisis de encuestas, para evaluar varianzas por jackknife. Existe la ventaja de trabajar de la misma forma estimadores más complejos, pos-estratificación y no respuesta, sin importar lo complejo del diseño.

La aplicación de Repeticiones Balanceadas tiene la desventaja sobre el jackknife de necesitar las matrices Hadamard, las cuales hay que proporcionarlas, o bien establecer la rutina que las cree. Por otra parte, no tiene el problema teórico con las estadísticas no suaves de inconsistencia de la varianza. Pero, el ejemplo que aquí se pudo desarrollar parece apoyar la idea de que en un diseño complejo, donde se eliminan conglomerados y no unidades aisladas, la varianza jackknife es muy similar a la R.B. cuando el número de estratos es grande.

Al respecto, un punto importante de notar es que el que la varianza de R.B. sea consistente para la mediana significa que según el tamaño de muestra crece, ésta se acerca a la verdadera varianza. Por lo tanto, en los casos donde se tienen pocos estratos y pocas localidades encuestadas, importaría más conocer el sesgo para el estimador en cuestión, que la consistencia. De aquí se desprende que en el caso del estimador de razón, cuando el coeficiente de variación del denominador es chico, el sesgo relativo del estimador es menor, lo que hace esperar que cualquiera que sea el sesgo del método utilizado (jackknife o R.B.), éste sea también pequeño y los estimadores calculados por diferentes técnicas se parezcan. Tal parece ser lo que ocurrió en el caso particular de Aguascalientes.(Buena aproximación entre los métodos a pesar de que el tamaño de muestra chico se reflejó en algunos intervalos amplios, como se ve en las tablas 6.9 y 6.10).

6.4 Conclusiones

Esta tesis consistió en la revisión de técnicas de remuestreo y la aplicación de algunas de ellas en la ENAL'96, para evaluar la varianza e intervalos de estimadores de razón y la mediana. Resultó gratificante el desarrollar un tema sobre técnicas poco explotadas en México y descubrir en otras personas interés en ellas (en la medida en que se desarrolló el tema). A continuación se exponen las conclusiones que se derivan de este trabajo, contemplando, en mayor medida, el aspecto estadístico. En segundo término, se resumen los resultados pertinentes a la nutrición, y de los cuales se hace la advertencia, de que requieren mayor interpretación de expertos en esa área.

1. Entre todos los métodos revisados, no se puede determinar que alguno de ellos sea *el mejor*. Sin embargo se pueden observar las recomendaciones expuestas en 4.5 para decidir en un momento dado qué método es adecuado y factible de aplicar para una situación particular.

2. Los métodos alternativos de cálculo de varianza en muestreos complejos merecen mayor difusión en México para así mejorar la calidad de los reportes de encuestas y proveer datos más cercanos a la realidad, evitando sobre todo, el cálculo de varianzas como si se tratara de muestreo aleatorio simple.

3. Para el caso de los estimadores de razón, en un muestreo complejo, el jackknife representa una forma sencilla y eficaz de obtener estimadores de varianza.

4. Se corrobora la gran similitud del estimador de varianza jackknife y el de linealización por series de Taylor. En la aplicación aquí realizada, el estimador por fórmula (derivado de la linealización) incluyó factores de corrección por población finita y el jackknife basado en Rao (ver varias referencias) no. Sin embargo, aún así

las dos estimaciones fueron prácticamente iguales en todos los estados, incluyendo Hidalgo, donde el estimador jackknife basado en Wolter (1985) fue el único discordante.

5. Se observó que para el caso del estimador de razón, cuando el coeficiente del denominador es menor a 0.1, las estimaciones de varianza jackknife y de Repeticiones Balanceadas (*Balanced Half-Sampling*) son más cercanas entre sí.

6. Respecto al punto anterior, se pudo ver que cuando el número de estratos es grande y el coeficiente de variación del denominador chico, los dos métodos señalados coinciden en sus estimaciones. (No hubo un estado donde se tuvieran muchos estratos y coeficiente grande, pero esto se debe a que en la medida de que el tamaño de muestra crece, el coeficiente de variación decrece). Por otra parte, en Aguascalientes se dió la situación de pocos estratos (5) y coeficiente menor a 0.1 y se observó una buena congruencia entre los dos métodos. En los estados con pocos estratos y coeficiente de variación grande (mayor a 0.1), se observó que la estimación de varianza por jackknife era mucho mayor a la de R.B. Sin embargo no se puede determinar cuál de los dos está más cercano al verdadero valor, pues es desconocido.

7. Los intervalos de confianza de estimadores de razón en los estados con coeficiente de variación del denominador grande son muy amplios y confirman la inadecuación de hacer inferencias en tales estados.

8. A nivel región se obtuvieron coeficientes de variación del denominador muy pequeños y por ende, intervalos precisos, lo que permite las inferencias sobre aspectos nutricionales a este nivel. Éste resultado era de esperar pues la muestra se diseñó considerando las regiones como dominios de estudio para los indicadores (PEDZ, PETZ y TEDZ) y las categorías para las que se

calculan las razones se basan en ellos.

9. El hecho de que en este ejemplo las varianzas e intervalos de la mediana obtenidos por Repeticiones Balanceadas (R.B.) y jackknife en los estados donde había alrededor de 30 estratos, fueran tan parecidos sustenta la necesidad de indagar en futuras investigaciones la posibilidad de que bajo ciertas condiciones, en un muestreo complejo, el estimador de la varianza jackknife (de la mediana) no sea inconsistente. (Esto se desprende de que el estimador de R.B. no es inconsistente y que ambas estimaciones resultaron similares).

10. Se encontró que la combinación de niveles de los tres indicadores de estado nutricional da lugar a la formación de categorías que permiten describir mejor la situación de los niños en determinada zona. En particular, en esta aplicación, se lograron identificar 5 grupos con adecuada representatividad. En un estudio futuro, se aconseja formar primero 9 grupos, como se menciona al final del capítulo 5. Es probable que los expertos en nutrición se interesen en explorar los mismos grupos, fijando límites distintos a los aquí utilizados. Pero si desean establecer comparaciones en el tiempo, se deben construir los grupos de la misma manera. También sería de interés, en un análisis estadístico posterior, hacer un análisis de conglomerados por niño, basado en los tres indicadores.

11. Se detecta que la proporción de niños menores de seis años, con estado nutricional *Normal o casi Normal* es mayor en los estados del Noroeste y Noreste, con intervalos de confianza al 95% de (71.12-74.77) y (59.32-64.44), en comparación con los estados del Centro y Sur, cuyos intervalos resultaron de (45.89 - 50.03) y (39.88 - 43.54), respectivamente. Pero la proporción en los estados de Centro Occidente es similar a la del Noreste, con un intervalo dado por (57.39-62.75).

En los estados del Centro y Sur, la proporción de niños *Bajitos Gorditos* es mayor que en las otras regiones, interpretación de lo cual debe ser del cuidado de un experto (En los estados del Centro y Sur las estimaciones y sus intervalos son menores pero cercanos a 30%, y en las otras zonas, es menor del 20%).

El grupo de los niños *Mal para la edad*, exhibe una proporción que crece cuanto más al Sur esté la región, ocurriendo la menor proporción en el Noroeste (4.17-6.37), y la mayor en el Sur (18.13-21.53); Se vuelve a observar similitud en las zonas Noreste y Centro Occidente, con intervalos de confianza al 95% de (8.60-12.53) y (7.89-11.01), repectivamente.

La clasificación *Mal para la talla* exhibió una situación similar en todo el País, ya que en el Noroeste el intervalo obtenido fue de (5.53- 10.58) y en el Sur de (5.34-6.88).

Se detectó una mayor proporción de niños *Mal para la edad y talla* en los estados del Centro y Sur, con intervalos dados por (2.40-3.41) y (2.67-3.84), respectivamente. Se encontró que dicha proporción en el Centro Occidente es un poco menor (1.32-2.40); y nuevamente, en los estados del norte, se encuentra un mejor panorama con intervalos para estas proporciones de (0.23-1.18) en el Noroeste y (0.83-1.91) en el Noreste.

12. Finalmente, aunque este trabajo fue aplicativo y de revisión, surgen tres propuestas para posibles investigaciones o trabajos en un futuro, que se plantean de forma muy general:

    a. Considerar el inciso7 de la sección 4.4, que trata de otra posible conceptualización del bootstrap en un muestreo complejo y comparar sus estimaciones con las de otras versiones del bootstrap, lo cual debe ser bajo un estudio de simulación. Al respecto se señala que Sitter (1992) compara tres versiones del bootstrap.

    b. Considerar una muestra tan grande como la que representa la ENAL'96 (u otra cualquiera) como si fuera una población;

Luego, establecer diferentes diseños con distintos números de estratos y unidades primarias de muestreo. (El tomar la gran muestra como población serviría para poder calcular valores poblacionales). Comparar el sesgo y estabilidad de los estimadores de la varianza de la mediana, obtenidos por varios métodos de remuestreo, incluyendo el jackknife.

c. Analizar una muestra proporcionada por INEGI, para la que haya interés en la implementación de métodos alternativos para el cálculo de varianza de estimaciones tanto a nivel nacional como a nivel de subclases. (Se expone aquí porque ya el INEGI ha manifestado esta necesidad).

TABLA 6.1 : ESTIMADORES DE RAZON Y SUS VARIANZAS,
OBTENIDOS POR REPETICIONES BALANCEADAS, JACKKNIFE Y FORMULA,
EN LOS ESTADOS CON DOS LOCALIDADES MUESTREADAS POR ESTRATO

ESTADO		REPETICIONES BALANCEADAS* (Half-Sampling)				JACKKNIFE		FORMULA o Linearizacion	
L= no. de estratos n= niños encuestados	GRUPO NUTRICIONAL	ESTIMADOR DE RAZON**	VARIANZA	VAR. COMPLEMENTO	MEDIA DE VARIANZAS	ESTIMADOR DE RAZON**	VARIANZA	ESTIMADOR DE RAZON	VARIANZA
1 AGUASCALIENTES									
L = 5	NORMALES y casi N,	0,671295	0,000436	0,000436	**0,000436**	0,671309	**0,000632**	0,671347	**0,000626**
n=385	BAJITOS GORDITOS	0,150036	0,000296	0,000296	**0,000296**	0,150033	**0,000327**	0,150159	**0,000297**
cv = 0.068466	MAL PARA LA EDAD	0,091842	0,000357	0,000357	**0,000357**	0,091889	**0,000564**	0,091869	**0,000515**
	MAL PARA LA TALLA	0,071799	0,000228	0,000228	**0,000228**	0,071746	**0,000341**	0,071514	**0,000239**
	MAL EDAD Y TALLA	0,015027	0,000044	0,000044	**0,000044**	0,015022	**0,000063**	0,015111	**0,000057**
11 GUANAJUATO									
L= 19	NORMALES y casi N,	0,585936	0,000421	0,000415	**0,000418**	0,586051	**0,000539**	0,586011	**0,000538**
n=845	BAJITOS GORDITOS	0,180433	0,000419	0,000413	**0,000416**	0,180556	**0,000516**	0,180556	**0,000486**
cv= 0.106484	MAL PARA LA EDAD	0,105298	0,000192	0,000196	**0,000194**	0,105145	**0,000239**	0,105159	**0,000229**
	MAL PARA LA TALLA	0,093071	0,000057	0,000061	**0,000059**	0,092979	**0,000092**	0,092989	**0,000083**
	MAL EDAD Y TALLA	0,035263	0,000021	0,000022	**0,000021**	0,035269	**0,000059**	0,035284	**0,000056**
15 MEXICO									
L=27	NORMALES y casi N,	0,533425	0,000554	0,000564	**0,000559**	0,533731	**0,000553**	0,533757	**0,000664**
n=1209	BAJITOS GORDITOS	0,268736	0,000412	0,000415	**0,000413**	0,268386	**0,000414**	0,268416	**0,00039**
cv= 0.097556	MAL PARA LA EDAD	0,136376	0,000122	0,000122	**0,000122**	0,13611	**0,000117**	0,136056	**0,000109**
	MAL PARA LA TALLA	0,054562	0,000143	0,000147	**0,000145**	0,054748	**0,000148**	0,054723	**0,000089**
	MAL EDAD Y TALLA	0,006902	0,000007	0,000007	**0,000007**	0,007024	**0,000007**	0,007048	**0,000007**
18 NAYARIT									
L= 5	NORMALES y casi N,	0,638343	0,000514	0,000514	**0,000514**	0,638913	**0,0039**	0,637705	**0,003404**
n=207	BAJITOS GORDITOS	0,180257	0,000138	0,000138	**0,000138**	0,17982	**0,001537**	0,180209	**0,001439**
cv= 0.148956	MAL PARA LA EDAD	0,074245	0,000256	0,000256	**0,000256**	0,074105	**0,000732**	0,074775	**0,000687**
	MAL PARA LA TALLA	0,095639	0,000040	0,000040	**0,000040**	0,095601	**0,000926**	0,095646	**0,000572**
	MAL EDAD Y TALLA	0,011516	0,000064	0,000064	**0,000064**	0,011561	**0,000066**	0,011665	**0,000061**
20 OAXACA									
L= 30	NORMALES y casi N,	0,404475	0,000392	0,000429	**0,000410**	0,404345	**0,000408**	0,404333	**0,000342**
1728	BAJITOS GORDITOS	0,301328	0,000399	0,000407	**0,000403**	0,301366	**0,000402**	0,30132	**0,000384**
cv= 0.090278	MAL PARA LA EDAD	0,211758	0,000449	0,000451	**0,000450**	0,211899	**0,00045**	0,211909	**0,000426**
	MAL PARA LA TALLA	0,052913	0,000080	0,000085	**0,000083**	0,052892	**0,000082**	0,052922	**0,000066**
	MAL EDAD Y TALLA	0,029526	0,000025	0,000025	**0,000025**	0,029498	**0,000025**	0,029514	**0,000024**

* *Repeticiones balanceadas con balace ortogonal completo y la adecuacion de Fay, factor =0.5.*
Var. Complemento es la varianza de las replicas dadas por el complemento de la matriz Hadamard.
** Se exhibe la media de las iteraciones.

...TABLA 6.1 : ESTIMADORES DE RAZON Y SUS VARIANZAS, OBTENIDOS POR REPETICIONES BALANCEADAS, JACKKNIFE Y FORMULA, EN LOS ESTADOS CON DOS LOCALIDADES MUESTREADAS POR ESTRATO

ESTADO L= no. de estratos n= niños encuestados	GRUPO NUTRICIONAL	REPETICIONES BALANCEADAS* (Half-Sampling) ESTIMADOR DE RAZON**	VARIANZA	VAR. COMPLEMENTO	MEDIA DE VARIANZAS	JACKKNIFE ESTIMADOR DE RAZON**	VARIANZA	FORMULA o Linearizacion ESTIMADOR DE RAZON	VARIANZA
L= 30	*NORMALES y casi N.*	0,404475	0,000392	0,000429	**0,000410**	0,404345	**0,000408**	0,404333	**0,000342**
1728	*BAJITOS GORDITOS*	0,301328	0,000399	0,000407	**0,000403**	0,301366	**0,000402**	0,30132	**0,000384**
cv= 0.090278	*MAL PARA LA EDAD*	0,211758	0,000449	0,000451	**0,000450**	0,211899	**0,00045**	0,211909	**0,000426**
	MAL PARA LA TALLA	0,052913	0,000080	0,000085	**0,000083**	0,052892	**0,000082**	0,052922	**0,000066**
	MAL EDAD Y TALLA	0,029526	0,000025	0,000025	**0,000025**	0,029498	**0,000025**	0,029514	**0,000024**
22 QUERETARO									
L=9	*NORMALES y casi N.*	0,553715	0,000045	0,000044	**0,000045**	0,554111	**0,000722**	0,553556	**0,000477**
n=651	*BAJITOS GORDITOS*	0,228116	0,000061	0,000062	**0,000061**	0,227926	**0,000308**	0,228162	**0,00028**
cv= 0.191388	*MAL PARA LA EDAD*	0,155317	0,000225	0,000217	**0,000221**	0,154936	**0,000689**	0,155423	**0,000627**
	MAL PARA LA TALLA	0,048646	0,000080	0,000080	**0,000080**	0,04878	**0,000137**	0,048619	**0,000121**
	MAL EDAD Y TALLA	0,014206	0,000003	0,000003	**0,000003**	0,014248	**0,000031**	0,01424	**0,000028**
26 SONORA									
L=6	*NORMALES y casi N.*	0,768257	0,000092	0,000101	**0,000097**	0,768715	**0,000965**	0,768588	**0,00097**
n=243	*BAJITOS GORDITOS*	0,133309	0,000193	0,000195	**0,000194**	0,133197	**0,000376**	0,13336	**0,000359**
cv= 0.187546	*MAL PARA LA EDAD*	0,025261	0,000119	0,000119	**0,000119**	0,024944	**0,000309**	0,024864	**0,000281**
	MAL PARA LA TALLA	0,068768	0,000032	0,000025	**0,000029**	0,068712	**0,000202**	0,068776	**0,000382**
	MAL EDAD Y TALLA	0,004405	0,000001	0,000001	**0,000001**	0,004431	**0,000012**	0,004412	**0,000011**
27 TABASCO									
L= 17	*NORMALES y casi N.*	0,58079	0,000254	0,000250	**0,000252**	0,58076	**0,000287**	0,581191	**0,000454**
n=1367	*BAJITOS GORDITOS*	0,213225	0,000300	0,000295	**0,000297**	0,21323	**0,000307**	0,213235	**0,000292**
cv= 0.61831	*MAL PARA LA EDAD*	0,117903	0,000018	0,000018	**0,000018**	0,117906	**0,000047**	0,11777	**0,000043**
	MAL PARA LA TALLA	0,07248	0,000044	0,000044	**0,000044**	0,072502	**0,00007**	0,072259	**0,000067**
	MAL EDAD Y TALLA	0,015603	0,000037	0,000037	**0,000037**	0,015603	**0,000042**	0,015544	**0,000039**

* *Repeticiones balanceadas* con balace ortogonal completo y la adecuacion de Fay, factor =0.5.
Var. Complemento es la varianza de las replicas dadas por el complemento de la matriz Hadamard.

** Se exhibe la media de las iteraciones.

TABLA 6.2 : ESTIMADORES DE LA **MEDIANA** DE LOS TRES INDICADORES NUTRICIONALES, Y SUS VARIANZAS, OBTENIDOS MEDIANTE JACKKNIFE Y REPETICIONES BALANCEADAS EN LOS ESTADOS CON DOS LOCALIDADES ENCUESTADAS POR ESTRATO

NO.	ESTADO L = no. estratos n= niños enc.	INDICADOR NUTRICIONAL	Jackknife Mediana**	Jackknife Varianza	REPETICIONES BALANCEADAS* (Balanced Half-Sampling) Mediana**	Varianza	V(Complemento)	Media de Vars.
1	AGUASCALIENTES							
	L=5	*Peso para la Edad*	-0,502	**0,006601**	-0,501	0,002145	0,002145	**0,002145**
	n=385	*Peso para la Talla*	-0,021	**0,001757**	-0,021	0,000291	0,000291	**0,000291**
		Talla para la Edad	-0,664	**0,015055**	-0,666	0,005046	0,005046	**0,005046**
11	GUANAJUATO							
	L= 19	*Peso para la Edad*	-0,909	**0,010048**	-0,909	0,007264	0,008092	**0,007678**
	n=845	*Peso para la Talla*	-0,194	**0,016098**	-0,187	0,013345	0,013911	**0,013628**
		Talla para la Edad	-1,131	**0,005424**	-1,132	0,002545	0,002579	**0,002562**
15	MEXICO							
	L=27	*Peso para la Edad*	-0,846	**0,005124**	-0,849	0,004773	0,005237	**0,005005**
	n=1209	*Peso para la Talla*	0,123	**0,003039**	0,122	0,002783	0,00278	**0,0027815**
		Talla para la Edad	-1,366	**0,011277**	-1,368	0,009914	0,01149	**0,010702**
18	NAYARIT							
	L=5	*Peso para la Edad*	-0,553	**0,019390**	-0,549	0,00043	0,00043	**0,00043**
	n=207	*Peso para la Talla*	0,008	**0,026707**	-0,006	0,010468	0,010468	**0,010468**
		Talla para la Edad	-0,619	**0,024550**	-0,615	0,001293	0,001293	**0,001293**
20	OAXACA							
	L=30	*Peso para la Edad*	-1,212	**0,001356**	-1,207	0,003113	0,001903	**0,002508**
	n=1728	*Peso para la Talla*	0,047	**0,004945**	0,047	0,00622	0,005919	**0,0060695**
		Talla para la Edad	-1,799	**0,004957**	-1,800	0,004846	0,005159	**0,0050025**
22	QUERETARO							
	L=9	*Peso para la Edad*	-0,857	**0,006873**	-0,857	0,001797	0,001492	**0,0016445**
	n=651	*Peso para la Talla*	0,074	**0,006114**	0,073	0,000958	0,00089	**0,000924**
		Talla para la Edad	-1,287	**0,004380**	-1,285	0,000349	0,000351	**0,00035**
26	SONORA							
	L=6	*Peso para la Edad*	0,044	**0,037332**	0,036	0,021089	0,018609	**0,019849**
	n=243	*Peso para la Talla*	0,320	**0,048977**	0,304	0,043526	0,043649	**0,0435875**
		Talla para la Edad	-0,043	**0,097062**	-0,04	0,0546	0,048985	**0,0517925**
27	TABASCO							
	L= 17	*Peso para la Edad*	-0,934	**0,005094**	-0,934	0,002532	0,002469	**0,0025005**
	n=1367	*Peso para la Talla*	-0,110	**0,006276**	-0,109	0,004084	0,004528	**0,004306**
		Talla para la Edad	-1,230	**0,001133**	-1,229	0,001013	0,00094	**0,0009765**

* Repeticiones Balanceadas con balance ortogonal completo y la adecuacion de Fay (factor =0.5).
V(complemento) = varianza de replicas dadas por el complemento de la matriz Hadamard .
**Se exhibe la media de las iteraciones o replicas.

TABLA 6.3 : ESTIMADORES DE LA MEDIA DE LOS TRES INDICADORES NUTRICIONALES, Y SUS VARIANZAS, OBTENIDOS MEDIANTE JACKKNIFE Y REPETICIONES BALANCEADAS EN LOS ESTADOS CON DOS LOCALIDADES ENCUESTADAS POR ESTRATO

NO.	ESTADO (L = no. estratos)	INDICADOR NUTRICIONAL	Jackknife Media**	Jackknife Varianza	REPETICIONES BALANCEADAS* (Balanced Half- Sampling) Media**	Varianza	V(Complemento)	Media de Vars.
1	**AGUASCALIENTES**							
	L=5	*Peso para la Edad*	-0,422	**0,006004**	-0,433	0,003401	0,003401	**0,003401**
		Peso para la Talla	0,027	**0,004985**	0,035	0,004175	0,004175	**0,004175**
		Talla para la Edad	-0,653	**0,026368**	-0,664	0,013155	0,013155	**0,013155**
11	**GUANAJUATO**							
	L= 19	*Peso para la Edad*	-0,795	**0,008635**	-0,800	0,006797	0,00694	**0,006869**
		Peso para la Talla	-0,112	**0,015020**	-0,113	0,014398	0,01433	**0,014364**
		Talla para la Edad	-1,058	**0,006278**	-1,062	0,004260	0,004138	**0,004199**
15	**MEXICO**							
	L=27	*Peso para la Edad*	-0,732	**0,002518**	-0,739	0,002523	0,002521	**0,002522**
		Peso para la Talla	0,173	**0,002614**	0,175	0,002738	0,002731	**0,002735**
		Talla para la Edad	-1,283	**0,009194**	-1,292	0,009040	0,009195	**0,009118**
18	**NAYARIT**							
	L=5	*Peso para la Edad*	-0,469	**0,013076**	-0,475	0,002095	0,002095	**0,002095**
		Peso para la Talla	0,005	**0,021640**	0,002	0,005935	0,005935	**0,005935**
		Talla para la Edad	-0,635	**0,077107**	-0,640	0,007608	0,007608	**0,007608**
20	**OAXACA**							
	L=30	*Peso para la Edad*	-1,059	**0,004829**	-1,064	0,004857	0,004815	**0,004836**
		Peso para la Talla	0,099	**0,007921**	0,102	0,007760	0,008227	**0,007994**
		Talla para la Edad	-1,758	**0,006317**	-1,763	0,006058	0,006596	**0,006327**
22	**QUERETARO**							
	L=9	*Peso para la Edad*	-0,793	**0,004830**	-0,801	0,001004	0,000966	**0,000985**
		Peso para la Talla	0,095	**0,005243**	0,103	0,000617	0,000607	**0,000612**
		Talla para la Edad	-1,280	**0,008829**	-1,290	0,001517	0,001508	**0,001513**
26	**SONORA**							
	L=6	*Peso para la Edad*	0,218	**0,038753**	0,220	0,029445	0,029400	**0,029423**
		Peso para la Talla	0,381	**0,068982**	0,383	0,058103	0,057863	**0,057983**
		Talla para la Edad	0,059	**0,029769**	0,059	0,024003	0,023266	**0,023635**
27	**TABASCO**							
	L= 17	*Peso para la Edad*	-0,793	**0,003197**	-0,801	0,001964	0,002021	**0,001993**
		Peso para la Talla	-0,051	**0,002719**	-0,059	0,001656	0,001593	**0,001625**
		Talla para la Edad	-1,118	**0,003921**	-1,126	0,003216	0,003283	**0,003250**

* Repeticiones Balanceadas con balance ortogonal completo y la adecuacion de Fay (factor =0.5).

V(complemento) = varianza de replicas dadas por el complemento de la matriz Hadamard .

**Se exhibe la media de las iteraciones o replicas.

TABLA 6.4 : COMPARACION DE LOS ESTIMADORES DE VARIANZA JACKKNIFE (1) Y REPETICIONES BALANCEADAS (2), PARA LA MEDIA, MEDIANA Y LOS ESTIMADORES DE RAZON DE LA PROPORCION POR GRUPO NUTRICIONAL, MEDIANTE EL COCIENTE [(1) / (2)]

ESTADO L= no. de estratos n= niños encuestados		ESTIMADORES DE RAZON POR GRUPO NUTRICIONAL	(1) / (2)	PESO PARA LA EDAD MEDIA CV*	PESO PARA LA EDAD MEDIANA	PESO PARA LA TALLA MEDIA	PESO PARA LA TALLA MEDIANA	TALLA PARA LA EDAD MEDIA	TALLA PARA LA EDAD MEDIANA
1	AGUASCALIENTES	NORMALES y casi N,	1,4484	1,7654	3,0774	1,1940	6,0378	2,0044	2,9836
	L = 5	BAJITOS GORDITOS	1,1047						
	n=385	MAL PARA LA EDAD	1,5783	0.068466					
		MAL PARA LA TALLA	1,4984						
		MAL EDAD Y TALLA	1,4255						
11	GUANAJUATO	NORMALES y casi N,	1,2881	1,2572	1,3087	1,0457	1,1812	1,4951	2,1171
	L= 19	BAJITOS GORDITOS	1,2406						
	n=845	MAL PARA LA EDAD	1,2311	0.106484					
		MAL PARA LA TALLA	1,5467						
		MAL EDAD Y TALLA	2,7593						
15	MEXICO	NORMALES y casi N,	0,9889	0,9984	1,0238	0,9559	1,0926	1,0084	1,0537
	L=27	BAJITOS GORDITOS	1,0016						
	n=1209	MAL PARA LA EDAD	0,9581	0,097556					
		MAL PARA LA TALLA	1,0202						
		MAL EDAD Y TALLA	0,9942						
18	NAYARIT	NORMALES y casi N,	7,5882	6,2415	45,0930	3,6462	2,5513	10,1350	18,9869
	L= 5	BAJITOS GORDITOS	11,1167						
	n=207	MAL PARA LA EDAD	2,8538	0.148956					
		MAL PARA LA TALLA	22,8983						
		MAL EDAD Y TALLA	1,0360						
20	OAXACA	NORMALES y casi N,	0,9946	0,9986	0,5407	0,9909	0,8147	0,9984	0,9909
	L= 30	BAJITOS GORDITOS	0,9974						
	1728	MAL PARA LA EDAD	0,9991	0.090278					
		MAL PARA LA TALLA	0,9931						
		MAL EDAD Y TALLA	0,9989						

Header above the value columns: ESTIMADOR DE VARIANZA JACKKNIFE / ESTIMADOR DE VARIANZA REPETICIONES BALANCEADAS

* CV= coeficiente de variacion del denominador del est. de razon combinado.

...TABLA 6.4 : COMPARACION DE LOS ESTIMADORES DE VARIANZA JACKKNIFE (1) Y REPETICIONES BALANCEADAS (2), PARA LA MEDIA, MEDIANA Y LOS ESTIMADORES DE RAZON DE LA PROPORCION POR GRUPO NUTRICIONAL, MEDIANTE EL COCIENTE [(1) / (2)]

ESTIMADOR DE VARIANZA JACKKNIFE / ESTIMADOR DE VARIANZA REPETICIONES BALANCEADAS								
ESTADO L= no. de estratos n= niños encuestados	*ESTIMADORES DE RAZON* *POR GRUPO NUTRICIONAL*	(1) / (2)	*PESO PARA LA EDAD* MEDIA CV*	*PESO PARA LA EDAD* MEDIANA	*PESO PARA LA TALLA* MEDIA	*PESO PARA LA TALLA* MEDIANA	*TALLA PARA LA EDAD* MEDIA	*TALLA PARA LA EDAD* MEDIANA
	MAL PARA LA TALLA	0,9931						
	MAL EDAD Y TALLA	0,9989						
22 QUERETARO	*NORMALES y casi N,*	16,1150	4,9036	4,1794	8,5670	6,6169	5,8374	12,5143
L=9	*BAJITOS GORDITOS*	5,0267						
n=651	*MAL PARA LA EDAD*	3,1186	0.191388					
	MAL PARA LA TALLA	1,7204						
	MAL EDAD Y TALLA	10,9736						
26 SONORA	*NORMALES y casi N,*	9,9703	1,3171	1,8808	1,1897	1,1236	1,2596	1,8741
L=6	*BAJITOS GORDITOS*	1,9365						
n=243	*MAL PARA LA EDAD*	2,6033	0.187546					
	MAL PARA LA TALLA	7,0529						
	MAL EDAD Y TALLA	23,7062						
27 TABASCO	*NORMALES y casi N,*	1,1379	1,6045	2,0372	1,6737	1,4575	1,2066	1,1603
L= 17	*BAJITOS GORDITOS*	1,0328						
n=1367	*MAL PARA LA EDAD*	2,6281	0.061831					
	MAL PARA LA TALLA	1,5903						
	MAL EDAD Y TALLA	1,1342						

* CV= coeficiente de variacion del denominador del est. de razon combinado.

TABLA 6.5 : INTERVALOS DE CONFIANZA AL 95% PARA LOS ESTIMADORES DE RAZON, SEGUN LAS VARIANZAS JACKKNIFE Y REPETICIONES BALANCEADAS EN LOS ESTADOS CON DOS LOCALIDADES ENCUESTADAS POR ESTRATO

ESTADO NO.	ESTADO NOMBRE	GRUPO NUTRICIONAL	G. L. (I - L)	REPETICIONES BALANCEADAS LIM. INF.	REPETICIONES BALANCEADAS LIM. SUP.	JACKKNIFE LIM. INF.	JACKKNIFE LIM. SUP.
1	AGUASCALIENTES		5				
	cv* = 0.68466	*NORMALES y casi N.*		0,617599	0,724991	0,606686	0,735932
		BAJITOS GORDITOS		0,105809	0,194263	0,103549	0,196517
		MAL PARA LA EDAD		0,043248	0,140436	0,030841	0,152937
		MAL PARA LA TALLA		0,033020	0,110578	0,024277	0,119215
		MAL EDAD Y TALLA		0,000000	0,032116	0,000000	0,035425
11	GUANAJUATO		19				
	cv= 0.106484	*NORMALES y casi N.*		0,543122	0,628750	0,537459	0,634643
		BAJITOS GORDITOS		0,137747	0,223119	0,133012	0,228100
		MAL PARA LA EDAD		0,076135	0,134461	0,072788	0,137502
		MAL PARA LA TALLA		0,076929	0,109213	0,072903	0,113055
		MAL EDAD Y TALLA		0,025585	0,044941	0,019192	0,051346
15	MEXICO		27				
	cv= 0.097556	*NORMALES y casi N.*		0,484905	0,581945	0,485480	0,581982
		BAJITOS GORDITOS		0,227020	0,310452	0,226637	0,310135
		MAL PARA LA EDAD		0,113702	0,159050	0,113916	0,158304
		MAL PARA LA TALLA		0,029849	0,079275	0,029786	0,079710
		MAL EDAD Y TALLA		0,001457	0,012347	0,001595	0,012453
18	NAYARIT		5				
	cv= 0.148956	*NORMALES y casi N.*		0,580067	0,696619	0,478380	0,799446
		BAJITOS GORDITOS		0,150031	0,210483	0,079042	0,280598
		MAL PARA LA EDAD		0,033076	0,115414	0,004557	0,143653
		MAL PARA LA TALLA		0,079292	0,111986	0,017378	0,173824
		MAL EDAD Y TALLA		0,000000	0,032034	0,000000	0,032444
20	OAXACA		30				
	cv= 0.090278	*NORMALES y casi N.*		0,363110	0,445840	0,363093	0,445597
		BAJITOS GORDITOS		0,260327	0,342329	0,260419	0,342313
		MAL PARA LA EDAD		0,168416	0,255100	0,168576	0,255222
		MAL PARA LA TALLA		0,034355	0,071471	0,034398	0,071386
		MAL EDAD Y TALLA		0,019309	0,039743	0,019287	0,039709
22	QUERETARO		9				
	cv= 0.191388	*NORMALES y casi N.*		0,538573	0,568857	0,493327	0,614895
		BAJITOS GORDITOS		0,210409	0,245823	0,188225	0,267627
		MAL PARA LA EDAD		0,121693	0,188941	0,095557	0,214315
		MAL PARA LA TALLA		0,028459	0,068833	0,022302	0,075258
		MAL EDAD Y TALLA		0,010404	0,018008	0,001653	0,026843
26	SONORA		6				
	cv= 0.187546	*NORMALES y casi N.*		0,744184	0,792330	0,692703	0,844727
		BAJITOS GORDITOS		0,099213	0,167405	0,085750	0,180644
		MAL PARA LA EDAD		0,000000	0,051919	0,000000	0,067957
		MAL PARA LA TALLA		0,055673	0,081863	0,033935	0,103489
		MAL EDAD Y TALLA		0,002664	0,006146	0,000000	0,012907
27	TABASCO		17				
	cv= 0.61831	*NORMALES y casi N.*		0,547284	0,614296	0,545017	0,616503
		BAJITOS GORDITOS		0,176850	0,249600	0,176263	0,250197
		MAL PARA LA EDAD		0,108981	0,126825	0,103442	0,132370
		MAL PARA LA TALLA		0,058483	0,086477	0,054850	0,090154
		MAL EDAD Y TALLA		0,002764	0,028442	0,001930	0,029276

* I = localidades encuestadas en el estado. L = estratos. cv = coef. de var. del denominador del est. combinado.
Se encuentran sombreados los intervalos cuyo límite inferior resultó ser menor a cero, por lo que aparece "0" como dicho límite.

TABLA 6.6 : INTERVALOS DE CONFIANZA AL 95% PARA LA MEDIANA DE LOS TRES INDICADORES DE ESTADO NUTRICIONAL, DE ACUERDO A LAS ESTIMACIONES JACKKNIFE Y REPETICIONES BALANCEADAS EN LOS ESTADOS DONDE HABIAN DOS LOCALIDADES ENCUESTADAS POR ESTRATO

NO.	ESTADO	G.L. (n-L)	INDICADOR NUTRICIONAL	INTERVALOS DE CONFIANZA 95% PARA LA MEDIANA			
				REPETICIONES BALANCEADAS		JACKKNIFE	
				Lim. Inf.	Lim. Sup.	Lim. Inf.	Lim. Sup.
1	AGUASCALIENTES	5					
			Peso para la Edad	-0,711	-0,293	-0,620	-0,382
			Peso para la Talla	-0,129	0,087	-0,065	0,023
			Talla para la Edad	-0,979	-0,349	-0,849	-0,483
11	GUANAJUATO	19					
			Peso para la Edad	-1,119	-0,699	-1,092	-0,726
			Peso para la Talla	-0,460	0,072	-0,431	0,057
			Talla para la Edad	-1,285	-0,977	-1,238	-1,026
15	MEXICO	27					
			Peso para la Edad	-0,993	-0,699	-0,994	-0,704
			Peso para la Talla	0,010	0,236	0,014	0,230
			Talla para la Edad	-1,584	-1,148	-1,580	-1,156
18	NAYARIT	5					
			Peso para la Edad	-0,911	-0,195	-0,602	-0,496
			Peso para la Talla	-0,412	0,428	-0,269	0,257
			Talla para la Edad	-1,022	-0,216	-0,707	-0,523
20	OAXACA	30					
			Peso para la Edad	-1,287	-1,137	-1,309	-1,105
			Peso para la Talla	-0,097	0,191	-0,112	0,206
			Talla para la Edad	-1,943	-1,655	-1,944	-1,656
22	QUERETARO	9					
			Peso para la Edad	-1,045	-0,669	-0,949	-0,765
			Peso para la Talla	-0,103	0,251	0,004	0,142
			Talla para la Edad	-1,437	-1,137	-1,327	-1,243
26	SONORA	6					
			Peso para la Edad	-0,429	0,517	-0,309	0,381
			Peso para la Talla	-0,222	0,862	-0,207	0,815
			Talla para la Edad	-0,805	0,719	-0,597	0,517
27	TABASCO	17					
			Peso para la Edad	-1,085	-0,783	-1,040	-0,828
			Peso para la Talla	-0,277	0,057	-0,247	0,029
			Talla para la Edad	-1,301	-1,159	-1,295	-1,163

TABLA 6.7 : ORDENAMIENTO DE LOS ESTADOS DE ACUERDO A LOS INTERVALOS DE CONFIANZA AL 95%, BASADOS EN EL JACKKNIFE, PARA LA PROPORCION DE NIÑOS NORMALES Y CASI NORMALES

ESTADOS (En orden de lím. Sup.)	I. C. 95% JACKKNIFE Normales y casi normales	
	Lím. Inferior	Lím. Superior
31 YUCATAN	0,266767	0,391585
7 CHIAPAS	0,330698	0,403002
21 PUEBLA	0,362904	0,429442
12 GUERRERO	0,348648	0,434964
20 OAXACA	0,363093	0,445597
4 CAMPECHE	0,368336	0,506764
13 HIDALGO	0,400708	0,509930
23 QUINTANA ROO	0,345034	0,525578
30 VERACRUZ	0,492586	0,572658
15 MEXICO	0,485480	0,581982
6 COLIMA	0,408372	0,610302
22 QUERETARO	0,493327	0,614895
27 TABASCO	0,545017	0,616503
24 SAN LUIS POTOSI	0,500934	0,627500
11 GUANAJUATO	0,537459	0,634643
19 NUEVO LEON	0,557268	0,635110
17 MORELOS	0,474763	0,639563
29 TLAXCALA	0,416739	0,644405
16 MICHOACAN	0,525397	0,644645
32 ZACATECAS	0,591801	0,683083
28 TAMAULIPAS	0,588480	0,698480
3 B.C. SUR	0,561000	0,715284
10 DURANGO	0,621204	0,727040
1 AGUASCALIENTES	0,606686	0,735932
14 JALISCO	0,684341	0,750537
25 SINALOA	0,689042	0,754564
8 CHIHUAHUA	0,676075	0,772531
5 COAHUILA	0,588989	0,783847
18 NAYARIT	0,478380	0,799446
26 SONORA	0,692703	0,844727
2 BAJA CALIFORNIA	0,741883	0,906515

Nota: No se deben hacer inferencias en los estados sombreados, pues cv >0.2

TABLA 6.8 : ORDENAMIENTO DE LOS ESTADOS DE ACUERDO A LOS INTERVALOS DE CONFIANZA AL 95%, BASADOS EN EL JACKKNIFE, PARA LA PROPORCION DE NIÑOS **BAJITOS GORDITOS**

ESTADOS (En orden de lim. Sup.)	I. C. 95% JACKKNIFE *Bajitos gorditos*	
	Lim. Inferior	Lim. Superior
14 JALISCO	0,114164	0,162412
25 SINALOA	0,094869	0,164789
2 BAJA CALIFORNIA	0,000000	0,166034
5 COAHUILA	0,095327	0,176401
26 SONORA	0,085750	0,180644
1 AGUASCALIENTES	0,103549	0,196517
3 B. C. SUR	0,072454	0,203430
10 DURANGO	0,151000	0,214354
32 ZACATECAS	0,124984	0,215090
8 CHIHUAHUA	0,026930	0,216386
11 GUANAJUATO	0,133012	0,228100
27 TABASCO	0,176263	0,250197
12 GUERRERO	0,204928	0,255856
19 NUEVO LEON	0,077484	0,262934
30 VERACRUZ	0,210663	0,264853
24 SAN LUIS POTOSI	0,172292	0,265388
22 QUERETARO	0,188225	0,267627
28 TAMAULIPAS	0,102342	0,268808
16 MICHOACAN	0,210609	0,273275
18 NAYARIT	0,079042	0,280598
17 MORELOS	0,209153	0,291217
29 TLAXCALA	0,170702	0,292472
15 MEXICO	0,226637	0,310135
21 PUEBLA	0,246954	0,312868
6 COLIMA	0,118381	0,321205
13 HIDALGO	0,237650	0,338746
31 YUCATAN	0,278393	0,340177
20 OAXACA	0,260419	0,342313
4 CAMPECHE	0,273219	0,351855
7 CHIAPAS	0,327491	0,394679
23 QUINTANA ROO	0,265878	0,423776

Nota: No se deben hacer inferencias en los estados sombreados, pues cv >0.2

TABLA 6.9 : ORDENAMIENTO DE LOS ESTADOS DE ACUERDO A LOS INTERVALOS DE CONFIANZA AL 95%, BASADOS EN EL JACKKNIFE, PARA LA PROPORCION DE NIÑOS **MAL PARA LA EDAD**

ESTADOS (En orden de lím. Sup.)	I. C. 95% JACKKNIFE Mal para la Edad	
	Lim. Inferior	Lim. Superior
2 BAJA CALIFORNIA	0,035222	0,050928
14 JALISCO	0,027441	0,052607
25 SINALOA	0,035038	0,067192
26 SONORA	0,000000	0,067957
10 DURANGO	0,027108	0,082944
28 TAMAULIPAS	0,006411	0,086109
6 COLIMA	0,020204	0,095282
16 MICHOACAN	0,060171	0,121581
8 CHIHUAHUA	0,025327	0,122331
27 TABASCO	0,103442	0,132370
19 NUEVO LEON	0,020966	0,132990
11 GUANAJUATO	0,072788	0,137502
18 NAYARIT	0,004557	0,143653
1 AGUASCALIENTES	0,030841	0,152937
32 ZACATECAS	0,046305	0,154401
15 MEXICO	0,113916	0,158304
24 SAN LUIS POTOSI	0,104231	0,166481
30 VERACRUZ	0,122421	0,168119
5 COAHUILA	0,021548	0,181416
13 HIDALGO	0,120942	0,192302
29 TLAXCALA	0,079589	0,192461
17 MORELOS	0,081343	0,207941
22 QUERETARO	0,095557	0,214315
21 PUEBLA	0,169743	0,215367
7 CHIAPAS	0,153943	0,216261
23 QUINTANA ROO	0,071819	0,227071
4 CAMPECHE	0,155180	0,242484
20 OAXACA	0,168576	0,255222
31 YUCATAN	0,151446	0,294018
12 GUERRERO	0,214627	0,300077
3 B. C. SUR	0,000000	[illegible]

Nota: No se deben hacer inferencias en los estados sombreados, pues cv >0.2

TABLA 6.10 : LISTADO DE LOS ESTADOS ORDENADOS DE ACUERDO A LOS INTERVALOS DE CONFIANZA AL 95%, BASADOS EN EL JACKKNIFE, PARA LA PROPORCION DE NIÑOS **MAL PARA LA TALLA**

RANG	ESTADOS (En orden de lim. Sup.)		I. C. 95% JACKKNIFE Mal Para la Talla	
	NO.	NOMBRE	Lim. Inferior	im. Superio
1	4	CAMPECHE	0,020643	0,040073
2	17	MORELOS	0,018885	0,059125
3	12	GUERRERO	0,046896	0,070244
4	20	OAXACA	0,034398	0,071386
5	30	VERACRUZ	0,048964	0,075140
6	22	QUERETARO	0,022302	0,075258
7	15	MEXICO	0,029786	0,079710
8	7	CHIAPAS	0,047300	0,080396
9	24	SAN LUIS POTOSI	0,035182	0,084570
10	27	TABASCO	0,054850	0,090154
11	13	HIDALGO	0,032471	0,097841
12	5	COAHUILA	0,024828	0,098458
13	16	MICHOACAN	0,035112	0,103456
14	26	SONORA	0,033935	0,103489
15	25	SINALOA	0,056293	0,107391
16	21	PUEBLA	0,058572	0,110940
17	32	ZACATECAS	0,058535	0,111359
18	14	JALISCO	0,081039	0,111441
19	11	GUANAJUATO	0,072903	0,113055
20	3	B. C. SUR	0,000000	0,113505
21	23	QUINTANA ROO	0,000000	0,117440
22	29	TLAXCALA	0,025118	0,118794
23	1	AGUASCALIENTES	0,024277	0,119215
24	31	YUCATAN	0,054160	0,122356
25	10	DURANGO	0,037328	0,127754
26	8	CHIHUAHUA	0,007095	0,153325
27	18	NAYARIT	0,017378	0,173824
28	28	TAMAULIPAS	0,026843	0,210797
29	19	NUEVO LEON	0,018423	0,269205
30	2	BAJA CALIFORNIA	0,000000	0,295545
31	6	COLIMA	0,073983	0,301847

Nota: No se deben hacer inferencias en los estados sombreados, pues cv >0.2

TABLA 6.11 : LISTADO DE LOS ESTADOS ORDENADOS DE ACUERDO A LOS INTERVALOS DE CONFIANZA AL 95%, BASADOS EN EL JACKKNIFE, PARA LA PROPORCION DE NIÑOS **MAL PARA LA EDAD Y TALLA**

RANG	ESTADOS (En orden de lím. Sup.)		I. C. 95% JACKKNIFE Mal para Edad y Talla	
	NO.	NOMBRE	Lim. Inferior	Lim. Superior
1	8	CHIHUAHUA	———	———
2	2	**BAJA CALIFORNIA**	0,000000	0,005944
3	10	**DURANGO**	0,001312	0,009954
4	15	**MEXICO**	0,001595	0,012453
5	26	**SONORA**	0,000000	0,012907
6	14	**JALISCO**	0,002158	0,013860
7	28	**TAMAULIPAS**	0,000000	0,017206
8	32	**ZACATECAS**	0,000000	0,017776
9	17	**MORELOS**	0,000000	0,019511
10	16	**MICHOACAN**	0,005096	0,020658
11	22	**QUERETARO**	0,001653	0,026843
12	27	**TABASCO**	0,001930	0,029276
13	30	**VERACRUZ**	0,015016	0,029580
14	25	**SINALOA**	0,000465	0,030359
15	4	**CAMPECHE**	0,009112	0,032334
16	18	**NAYARIT**	0,000000	0,032444
17	19	**NUEVO LEON**	0,000000	0,033428
18	24	**SAN LUIS POTOSI**	0,009233	0,034189
19	23	**QUINTANA ROO**	0,000544	0,034450
20	1	**AGUASCALIENTES**	0,000000	0,035425
21	5	**COAHUILA**	0,000000	0,036539
22	7	**CHIAPAS**	0,009452	0,036780
23	20	**OAXACA**	0,019287	0,039709
24	6	**COLIMA**	0,010446	0,039978
25	29	**TLAXCALA**	0,011323	0,048397
26	11	**GUANAJUATO**	0,019192	0,051346
27	13	**HIDALGO**	0,012983	0,056427
28	21	**PUEBLA**	0,034755	0,058455
29	31	**YUCATAN**	0,028539	0,072559
30	12	**GUERRERO**	0,045497	0,078265
31	3	**B. C. SUR**	0,000000	0,146946

Notas: No se deben hacer inferencias en los estados sombreados pues cv > 0.2.
En el estado de Chihuahua se dio un cero muestral en esta categoria.

TABLA 6.12 : TAMAÑOS DE MUESTRA EN PRIMERA (LOCALIDADES) Y SEGUNDA ETAPA (NIÑOS) POR ESTADO

ESTADO		ESTRATOS	TAMAÑO DE MUESTRA	
			LOCALIDADES	NIÑOS
1	AGUASCALIENTES	5	10	385
2	BAJA CALIFORNIA	2	6	164
3	B. C. SUR	4	11	90
4	CAMPECHE	8	24	1024
5	COAHUILA	5	15	409
6	COLIMA	4	12	247
7	CHIAPAS	25	57	1897
8	CHIHUAHUA	4	9	250
10	DURANGO	8	21	581
11	GUANAJUATO	19	38	845
12	GUERRERO	20	54	1887
13	HIDALGO	15	32	997
14	JALISCO	12	35	886
15	MEXICO	27	54	1209
16	MICHOACAN	20	40	1435
17	MORELOS	11	25	669
18	NAYARIT	5	10	207
19	NUEVO LEON	3	9	206
20	OAXACA	30	60	1728
21	PUEBLA	25	59	2413
22	QUERETARO	9	18	651
23	QUINTANA ROO	3	9	209
24	SAN LUIS POTOSI	11	32	1192
25	SINALOA	14	33	742
26	SONORA	6	12	243
27	TABASCO	17	34	1367
28	TAMAULIPAS	5	12	332
29	TLAXCALA	4	9	380
30	VERACRUZ	36	77	2635
31	YUCATAN	6	17	674
32	ZACATECAS	9	20	746

TABLA 6.13 : RESULTADOS DE LOS ESTIMADORES DE RAZON
E INTERVALOS DE CONFIANZA AL 95%, OBTENIDOS MEDIANTE EL JACKKNIFE,
POR GRANDES REGIONES DE LA REPUBLICA MEXICANA

REGION *Estados en la region* CV= Coeficiente de variacion del denominador del est. de razon	**TAMAÑOS EN LA MUESTRA DE:**			**GRUPO NUTRICIONAL**	**ESTIMADOR JACKKNIFE**		**INTERVALO DE CONFIANZA AL 95% ****	
	ESTRA-TOS	*LOCS.*	*FAMI-LIAS**		*PROP.*	*VARIANZA*	*LIM. INFERIOR*	*LIM SUPERIOR*
1 NOROESTE	38	92	1564					
Baja California, B.C.Sur,				*NORMALES y casi N*	0,729457	0,00008247	0,711250	0,747664
Sonora,Chihuahua,				*BAJITOS GORDITOS*	0,130186	0,0001308	0,107257	0,153115
Sinaloa, Durango.				*MAL PARA LA EDAD*	0,052695	0,00003029	0,041661	0,063729
CV= 0.08472				*MAL PARA LA TALLA*	0,08056	0,00015899	0,055280	0,105840
				MAL EDAD Y TALLA	0,007102	0,00000557	0,002370	0,011834
2 NORESTE	38	98	2316					
Coahuila, Nuevo Leon,				*NORMALES y casi N*	0,618843	0,00016347	0,593268	0,644418
Tamaulipas, Zacatecas, San				*BAJITOS GORDITOS*	0,183761	0,00011587	0,162229	0,205293
Luis Potosi, Aguascalientes				*MAL PARA LA EDAD*	0,105689	0,0000966	0,086029	0,125349
CV= 0.05896				*MAL PARA LA TALLA*	0,077972	0,00004357	0,064769	0,091175
				MAL EDAD Y TALLA	0,013735	0,00000718	0,008375	0,019095
3 CENTRO OCCIDENTE	68	153	3042					
Nayarit, Jalisco,				*NORMALES y casi N*	0,600743	0,00018128	0,573973	0,627513
Colima, Guanajuato,				*BAJITOS GORDITOS*	0,206682	0,00009286	0,187522	0,225842
Queretaro , Michoacan				*MAL PARA LA EDAD*	0,094516	0,0000614	0,078936	0,110096
CV= 0.06172				*MAL PARA LA TALLA*	0,079439	0,0000515	0,065170	0,093708
				MAL EDAD Y TALLA	0,018621	0,00000737	0,013223	0,024019
4 CENTRO	118	256	5881					
Mexico, Hidalgo,				*NORMALES y casi N.*	0,479611	0,00010965	0,458906	0,500316
Tlaxcala, Puebla,				*BAJITOS GORDITOS*	0,262861	0,00006835	0,246514	0,279208
Morelos, Veracruz.				*MAL PARA LA EDAD*	0,160165	0,00003611	0,148283	0,172047
CV= 0.03819				*MAL PARA LA TALLA*	0,068318	0,00002931	0,057613	0,079023
				MAL EDAD Y TALLA	0,029044	0,0000065	0,024003	0,034085
5 SUR	109	255	6185					
Guerrero, Oaxaca,Chiapas,				*NORMALES y casi N*	0,417104	0,00008565	0,398813	0,435395
Tabasco, Campeche,				*BAJITOS GORDITOS*	0,290987	0,00007209	0,274207	0,307767
Yucatan, Quintana Roo.				*MAL PARA LA EDAD*	0,198278	0,00007389	0,181289	0,215267
CV= 0.03853				*MAL PARA LA TALLA*	0,061132	0,0000152	0,053427	0,068837
				MAL EDAD Y TALLA	0,0325	0,00000884	0,026624	0,038376

* Se indica el numero de familias encuestadas, en las que hubo niños menores de 6 años.

** Los intervalos se basan en una "t" con grados de libertad dados por: #locs. encuestadas -estratos.

Bibliografía

ANDERSON C. Y LENNART NORDBERG (1994), "A Method for variance Estimation of non-linear functions of Totals in surveys - Theory and Software Imlementation", *Journal of Official Statistics*, vol 10, no 4, pags. 395-405.

ÁVILA A., T. SHAMAH Y A. CHÁVEZ (1997) , "Encuesta Nacional de Alimentación y Nutrición en el Medio Rural 1996", Documento interno del Instituto Nacional de la Nutrición "Salvador Zubirán".

BARRAGÁN L.M., R. AMBROCIO, A. ÁVILA, M.A. ÁVILA (1997) "Cartografía. Encuesta Nacinal de Alimentación y Nutrición en el medio rural 1996", Documento interno del Instituto Nacional de la Nutrición "Salvador Zubirán".

BINDER D.A. Y Z. PATAK (1994), "Use of Estimating Functions for Estimation from Complex Surveys", *Journal of the American Statistical Association*, vol. 89, no. 427, "Theory and Methods".

BRICK J.M. Y G. KALTON (1996), "Handling Missing Data in Survey Research", *Statistical Methods in Medical Research*, 5, págs. 215-238.

CANTY A.J., Y A.C. DAVISON (1998), "Variance estimation for Two Complex Surveys in Switzerland", Documento interno de *Swiss Federal Instirute of Technology*, bajo el contrato de *Swiss Federal Statistical Office.*

CASSEL C.M., C.E. SÄRNDAL Y J.H. WRETMAN (1977), "Foundations of Inference in Survey Sampling", John Wiley & Sons, E.U.A.

COCHRAN W.G. (1977), "Sampling Techniques", John Wiley & Sons, E.U.A.

CHAUDHURI A. Y H. STENGER (1992), "Survey Sampling", Marcel Dekker Inc., E.U.A.

CHAUDHURI A. Y J.W.E. VOS (1988), "Unified Theory and Strategies of Survey Sampling", *North-Holland Series in Statistics and Probability*, Elsevier Science Publishers B.V., Países Bajos.

DEMING W.E. (1956), "On simplification of Sampling Design through Replication with Equal Probabilities and Without Stages", *Journal of the American Statistical Association*, 51, págs. 24-53.

DES RAJ (1980), Teoría del Muestreo, Fondo de Cultura Económica, México.

DEVILLE J.C., C.E. SÄRNDAL (1994), "Variance Estimation for the Regression Imputed Horvitz-Thompson Estimator", *Journal of Official Statistics*, vol 10, no. 4, pags. 381- 394.

DIPPO, C.S., R.E. FAY Y D.H. MORGANSTEIN (1984), "Computing variances from Complex Samples with Replicate Weights", *ASA Survey Research (1984)*, pags. 489-494.

EFRON B. (1979), "THE 1977 RIETZ LECTURE ; Bootstrap Methods: Another look at the Jackknife", *The Annals of Statistics*, vol. 7, no. 1, pags. 1-26.

EFRON B. (1985), "The Jackknife, the Bootstrap and Other Resampling Plans", Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 2da impresión, 1985.

EFRON B. Y C. STEIN (1981), "The Jackknife Estimate of Variance", *The Annals of Statistics*, vol. 9, no.3, pags. 586-596.

ERNST L. Y T. R. WILLIAMS (1987), "Some Aspects of Estimating variances by Half-sample Replication in CPS", *ASA*, *Proceedings of the Section on Survey Research Methods*, pags. 480-485.

GALFOND G.J., M. E. GARVIN Y J. C. THOMPSON (1987), "A Simple Approximate Variance Estimate for a Complex Unequal Probability Sample Survey", *ASA*, *Proceedings of the Section on Survey Research Methods*, pags. 755- 760.

GODAMBE V.P.(1955), "A Unified Theory of Sampling from Finite Populations", *Journal of the Royal Statistical Society*, B **17**, pags. 269-278.

GODAMBE V.P. (1966), " A New Approach to Sampling from Finite Populations. Sufficiency and Linear Estimation", *J R Statist. Soc.* , B, 28, págs. 310-328.

GODAMBE V.P. Y V.M. JOSHI (1965), "Admissibility and Bayes Estimation in Sampling Finite Populations", *Ann. Math. Statis.*, 36, págs. 1707-1722.

GODAMBE V. P., THOMPSON M.E. (1986), " Parameters of Superpopulation and Survey population: Their Relationship and Estimation", *International Statistical Review*, **54**, 2, pags. 127- 138.

GRAUBARD B. I. Y KORN E.L. (1996), "Modelling the Sampling Design in the Analysis of Health Surveys", *Statistical Methods in Medical Research*, **5**, pags. 263-281.

HANSEN M.H., W.N. HURWITZ Y W.G. MADOW (1953), "Sample Survey Methods and Theory", vol I. y II., John Wiley & Sons, New York.

HANSEN M.H., W.G. MADOW Y B.J. TEPPING (1983), "An Evaluation of Model-Dependent and Probability-Sampling Inferences in Sample Surveys", *Journal of the American Statistical Association*, vol. 78, no 384, "Invited Papers Section".

JUDKINS D.R. (1987), "Modified Balaced Repeated Replication",*ASA, Proceedings of the Section on Survey Research Methods.*

KISH L. (1992), "Weighting for Unequal P_i", *Journal of Offical Statistics*, vol. 8, no. 2 pags. 183-200.

KISH L. (1989), "Deffs: Why, When, and How? A Review", *ASA Proceedings of Survey Research Methods Section 1989* , pags. 209-211.

KISH L.(1972), Muestreo de Encuestas, Primera Edición en Español, Editorial Trillas, S.A.

KISH L. (1977), "Robustness in Survey sampling", *Bulletin of International Statistical Institute.* vol. 47, pág. 515- 533.

KISH L. (1987), "Statistical Design for Research", John Wiley & Sons, E.U.A.

KOTT P.S. (1994), "Reweighting and Variance Estimation for the Characteristics of Business Owners Survey", *Journal of Official Statistics*, vol. 10, no 4, pags. 407- 418.

KOVAR J.G., J.N.K. RAO, Y C.F.J. WU (1988), "Bootstrap and other methods to measure errors in survey estimates", *Canadian Journal of Statistics 1988*, 16, pags. 25- 46.

KREWSKI D. Y J.N.K. RAO (1981), "Inference from stratified samples: proporties of the linearization, Jackknife and Balanced Repeated Replication Methods", *The Annals of Statistics*, vol9, no. 5, pags. 1010-1019.

LAPLANTE M. P. , (1991), "Estimating variances for Health and Disability Domains in the National Health Interview Survey", *Proceedings of the Section on Survey Research Methods, ASA*, pags. 684-689.

LEPIDUS CARLSON B., A.E. JOHNSON, S.B. COHEN (1993), "An Evaluation of the Use of Personal Computers for Variance Estimation with Complex Survey Data", *Journal of Offical Statistics*, vol. 9, no. 4, pags. 795-814.

LEVY P.S.(1981), "Sampling of popualtions: Methods and Applications", 2da edición, 1981, Capítulo 12.

MAHALANOBIS P.C. (1946), "Recent Experiments in Statistical Sampling in the Indian Statistical Institute", *Journal of the Royal Statistical Society*, vol. **CIX**, Parte **IV**, pags. 326-378.

MCCARTHY P.J. (1966), "Replication: An approach to the Analysis of Data from Complex Surveys, *Vital and Health Statistics*, Serie 2, no.14, *National Center for Health Statistics, Public health Service*, Washington, D.D.;

citado por Wolter (1985).

MIRÁS J. (1985), Elementos de Muestreo para Poblaciones Finitas, Instituto Nacional de Estadísstica, Madrid, España.

NATHAN G. (1988), "Inference Based on Data from Complex Sample Designs", P.R. Krishnaiah y C.R. Rao: editores,*Handbook of Statistics*, vol. 6, Elsevier Science Publishers B.V., pags. 247-266.

ORGANIZACIÓN MUNDIAL DE LA SALUD (OMS) (1983), "Measuring Change in Nutritional Status", Ginebra, Suiza,1983.

OTT L., W. MENDENHALL Y R.F. LARSON (1978), "Statistics: A Tool for the Social Sciences", Duxbury Press, North Scituate, Massachusetts.

PFEFFERMANN D. (1996), "The use of sampling weights for survey data analysis", *Statistical Methods in Medical Research*, 5, pags. 239-261.

QUENOUILLE M.H. (1940), "Aproximate Tests of Correlation in Time Series", *Journal of the Royal Statistical Society*, B, 11, págs. 68-84.

QUENOUILLE M.H. (1956), "Notes on Bias in Estimation", *Biometrika*, **43**, pags. 353-360.

RAO, J.N.K. (1988), "Variance Estimation in Sample Surveys", *Handbook of Statistics*, Vol. 6, Elsevier Science Publishers B.V., pags. 427-447.

RAO J.N.K. Y C.F.J. WU (1985), "Inference from Stratified Samples; Second-order Analysis of the Three Methods for Non-linear Statistics", *Journal of the American Statistical Associatio*, 80, págs. 620-630.

RAO J.N.K. Y C.F.J. WU (1988), "Resampling Inference with complex Survey Data", *Journal of the American Statistical Association*,vol. 83, no. 401, Theory and Methods, pags. 231-241.

RAO, J.N.K. (1997), "Developments in sample survey theory: an appraisal", *The Canadian Journal of Statistics*, vol. 25, no.1, pags.1-21.

RUST K.F., J.N.K. RAO (1996), "Variance estimation for complex surveys using replication techniques", *Statistical Methods in Medical Research*, 5, pags. 283-310.

SKINNER C.J., D.HOLT Y T.M.F. SMITH (editores) (1989), "Analysis of Complex Surveys", John Wiley & Sons, Gran Bretaña.

SITTER R.R. (1992), "Comparing Three Bootstrap Methods for Survey Data", *The canadian Journal of Statistics*, vol. 20, no. 2, 1992, págs. 135-154.

SMITH T.M.F. ( ), "First Morris Hansen Lecture: Sample Surveys 1975-1990; An Age of Reconciliation? ", *International Statistical Review*.

SUKHATME P.V., B.V. SUKHATME, S. SUKHATME Y C. ASOK (1984), "Sampling Theory of Surveys with Applications", *Iowa State University Press* (Ames, Iowa, E.U.A.) e *Indian Society of Agricultural Statistics* (Nueva Delhi, India), 3a. edición.

TUKEY J.W. (1958), "Bias and Confidence in not Quite Large Samples", *Annals of Mathematical Statistics*, 29, pág. 614.

WOLTER K.M. (1985), "Introduction to Variance Estimation", *Springer-Verlag*, New York, N.Y. U.S.A.

WOODRUFF R. S. (1971), "A Simple Method for Approximating the Variance of a Complicated Estimate", *Journal of the American Statistical Association*, 66, págs. 411-414.

WOODRUFF R. S. Y B. D. CAUSEY (1976), " Computerized Method for Approximating the Variance of a Complicated Estimate", *Journal of the American Statistical Association*, vol. 71, no. 354, pags. 315- 321.

YUNG W. Y J.N.K. RAO (1996), "Jackknife Variance Estimators under Stratified Multistage Sampling", *Survey Methodology*, 22, págs. 23-31.

Apéndice A

Reglas para Formar Grupos Aleatorios

A continuación se citan las reglas dadas por Wolter (1985, pág. 31) para formar grupos aleatorios dependientes:

1. Si se tiene un amuestra de tamaño n, de una sola etapa, seleccionada por muestreo aleatorio simple sin reemplazo o muestreo proporcional al tamaño sin reemplazo, entonces los grupos aleatorios se forman dividiendo la muestra original aleatoriamente. Esto significa que el primer grupo aleatorio se obtiene seleccionando una muestra aleatoria simple sin reemplazo de tamaño $m = \frac{n}{k}$, de la muestra original; el segundo grupo, seleccionado una m.a.s. sin reemplazo de tamaño m, de las $n - m$ unidades restantes; y así sucesivamente. Si $\frac{n}{k}$ no es entero, es decir, $n = km + q$ con $0 < q < k$, entonces lse pueden eliminar las q unidades que sobran de los k grupos. Una alternativa es asignar una de las unidades a cada uno de los q primeros grupos.

2. Si se seleccionó una muestra sistemática de tamaño n, con un solo punto de inicio, y con probabilidades iguales o diferentes, entonces los grupos aleatorios se deben formar dividiendo la muestra original sistemáticamente. Esto se puede llevar a cabo, generando un entero entre 1 y k, digamos α^*. La primera unidad en la muestra total se puede asignar al grupo α^*, la segunda al grupo α^*+1, y

así sucesivamente, bajo un esquema de módulo k.

3. En muestreos multietápicos, los grupos aleatorios se deben formar dividiendo los conglomerados de última etapa, es decir, el agregado de todas las unidades elementales seleccionadas de la misma Unidad Primaria de Muestreo, en k grupos. La Unidad Primaria de Muestreo se debe considerar como una sola unidad al formar los Gruos Aleatorios.La verdadera división de los conglomerados de última etapa en grupos aleatorios se hace de acuerdo a las reglas (1) o (2), dependiendo de la naturaleza del diseño muestral de primera etapa. Si el diseño es m.a.s. sin reemplazo o proporcional al tamaño sin reemplazo, entonces se debe usar la regla (1), mientras que la regla (2) se usa para muestreo sistemático. Se requiere de especial cuidado cuando se aplica esta regla de *conglomerado último* a las Unidades Primarias de Muestreo que son auto-representativas (elegidas con probabilidad 1) . Desde el punto de vista de estimación de la varianza, una Unidad auto-representativa se debe considerar un estrto aparte, y las unidades utilizadas en la primera etapa del submuestreo son la base para la formación de grupos aleatorios.

4. En muestreo estratificado, existen dos opciones. En primer lugar, si deseamos estimar la varianza dentro de un estrato, entonces se aplican las reglas (1), (2) o (3), dependiendo de la naturaleza del diseño muestral dentro del estrato. Por ejemplo, la regla (3) se usa si hay un dieño multietápico dentro del estrato. En segundo lugar, si deseamos estimar la varianza total entre todos los estratos, entonces cada grupo aleatorio debe ser una muestra estratificada conformada por unidades de todos los estratos. En este caso, se obtiene el primer grupo seleccionando una m.a.s. sin reemplazo de tamaño $m_h = \frac{n_h}{k}$ de las n_h unidades originales en el estrato h, para $h = 1, 2, ..., L$ El segundo grupo aleatorio se obtiene de la misma maneraseleccionando de las $n_h - m_h$ unidades en el estrato h. Los grupos aleatorios restantes se construyen de manera similar. Si hay observaciones excedentes en algún es-

trato, es decir, $n_h = km_h + q_h$, pueden quedar eliminadas de los k grupos, o añadidos de uno en uno, a los primeros q_h grupos. Si la muestra original se selecciona sistemáticamente dentro de los estratos, entonces los grupos aleatorios también se deben construir sistemáticamente. En otras palabras, cada grupo aleatorio se debe componer de una submuestra sistemática de la muestra total, en cada estrato.

5. Si se va a construir un estimador para un esquema de muestreo doble, tal como, muestreo doble para estratificación, o muestreo doble para el estimador de razón, (ver Cochran(1977, Capítulo 12)), entonces las n' unidades de muestreo seleccionadas en la muestra inicial deben ser divididas en k grupos aleatorios. La división debe ser aleatoria para diseños m.a.s. sin reemplazo y muestreo proporcional al tamaño sin reemplazo y sistemáticamante para muestreos sistemáticos. Cuando n' no es un entero múltiplo de k, cualquiera de los procedimientos planteados en la regla (1) para lidear con las unidades excedentes, puede aplicarse. La muestra de segunda fase, digamos de tamaño n, se divide en grupos aleatorios de acuerdo a la división de la muestra inicial. este procedimiento se usa cuando ambas, la muestra inicial y la de segunda fase, se selecionan antes de la formación de grupos aleatorios. Alternativamente, para algunas aplicaciones, es posible formar grupos aleatorios después de la selección de la muestra inicial pero antes de la selección de la muestra de segunda fase.En este caso, la muestra n' se divide en k grupos aleatorios y la muestra de segunda fase se obtiene independientemente seleccionado $m = \frac{n}{k}$ unidades de cada grupo aleatorio.

Apéndice B

Ajustes para la Pos-estratificación

En ocasiones la variable por la que se quisiera estratificar no está disponible, por lo que el diseño se basa en un estratificación regional o geográfica, y luego se recurre a la pos-estratificación. Cuando las características utilizadas para definir los pos-estratos están correlacionadas con la variable de interés, este ajuste lleva a una reducción de la varianza; por otra parte, el estimador posestratificado es aproximadamente insesgado cuando la no-respuesta es aleatoria (Brick y Kalton, 1996), es decir que ésta no está localizada en una clase de una posible partición de la población .

Hansen, Hurwitz y Madow (1953) dan una aproximación de la varianza del estimador de la media, en el caso de una pos-estratificación. Sin embargo, en la práctica siempre resulta problemática la supervisión a los programadores y la aplicación de este método cuando son varios los estimadores o las variables de interés. Obviamente, cuando el estimador no es lineal, el cálculo de la varianza no se basa en una expresión directa. Muchas veces es posible utilizar la linearización pero esto se dificulta cuando interesan estimadores de diversa índole.

El Jackknife y el Método de Repeticiones Balanceadas ("Balanced Half-Sampling") resuelven el cálculo de varianza bajo pos-estratificación con mucha facilidad. A continuación se explica brevemente el ajuste propuesto por Rust y Rao (1996) para resolver este problema mediante los métodos de remuestreo que se acaban de mencionar.

Sea a, $a = 1, ..., A$, el número de posestrato. Sea T_a el total fijo para el posestrato a. Sea w_j el peso inicial para la unidad j-ésima, antes de posestratificar. Se considera el siguiente factor de ajuste para el posestrato a:

$$\hat{p}_a = \frac{T_a}{\sum_{j \epsilon a} w_j}$$

Los pesos de posestratificación son :

$$w_j^* = \hat{p}_a w_j.$$

Así pues, ocurre que:

$$\sum_{j \epsilon a} w_j^* = T_a \text{ para todo } a \text{ y cualquier muestra}$$

Ahora bien, de acuerdo a la descripción de los métodos ofrecida en esta tesis, al aplicar el Jackknife o Repeticiones Balanceadas, se ajustan los pesos w_j, obteniendo $w_j^{(k)}$. Por lo que la solución de una posestratificación sólo requeriría del volver a ajustar los pesos, es decir:

$$w_j^{(k)*} = \hat{p}_a^{(k)} w_j^{(k)}.$$

Finalmente, estos pesos son los que se utilizan en la expresión del estimador (cualquiera que sea) y se aplican las expresiones de estiamdor de la varianza que venga al caso, según el método de remuestreo que se implementó.

Apéndice C

Grados de Libertad

Cuando se utiliza un estimador de varianza para un muestreo complejo, basado en remuestreo, los grados de libertad son un número mucho menor que el tamaño de la muestra. Esto es un aspecto importante de advertir pues en ocasiones sí puede representar una diferencia en la precisión de los intervalos y las inferencias. Rust y Rao (1996) dan elementos para obtener los grados de libertad en el caso de un muestreo complejo. Siendo que dicha referencia ha sido la única fuente (encontrada) que aborde este tema, se decidió presentar en este apéndice los puntos principales que exponen estos autores, como un complemento importante al tema de la tesis.

El enfoque utilizado requiere de varios supuestos, por lo que no siempre esta solución provee de un cálculo adecuado de grados de libertad. Se supone un m.a.s. y una distribución normal para una variable de interés, x, i.e., $x \sim N(\overline{x}, \sigma^2)$. En tal caso, se sabe que $n(n-1)v(\hat{\overline{x}})/\sigma^2$ se distribuye Ji-cuadrada con (n-1) grados de libertad $(\chi^2_{(n-1)})$. De ahí que,

$$V(v(\hat{\overline{x}})) = 2\sigma^4/n^2(n-1).$$

Como, por otra parte, $E(v(\hat{\overline{x}})) = \sigma^2/n$, ocurre que

$$V(v(\hat{\overline{x}})) = 2(E(v))^2/(n-1),$$

de donde se despejan los grados de libertad (n-1), como

$$(n-1) = \frac{2(E(v))^2}{V(v(\hat{\overline{x}}))}.$$

La expresión anterior se generaliza para encontrar los grados de libertad, d:

$$d = \frac{2(E(v))^2}{V(v(\hat{\bar{x}}))}. \tag{C.1}$$

Cuando se tiene un diseño estratificado multietápico con dos Unidades Primarias de Muestreo por estrato. Se considera el estimador del total, dado por:

$$\hat{X} = \sum_{h=1}^{L}\sum_{i=1}^{2} x_{hi} \ ,$$

donde, x_{hi} es el estimador del total de la unidad primaria o conglomerado i-ésimo del estrato h. La varianza de x_{hi} se denota por V_h. Entonces, la varianza del estimador del total, se expresa como:

$$V(\hat{X}) = 2\sum_{h=1}^{L} V_h.$$

Por otra parte, ocurre que:

$$V(v(\hat{X})) = V(v_{jackknife}(\hat{X})) = V(v_{Rept.Balnc.}(\hat{X})) = \sum_{h=1}^{L} 2(\beta_h + 1)V_h^2.$$

Cuando las x_{hi} se ditribuyen como una Normal, entonces $\beta_h = 3$. Bajo este supuesto, y utilizando (??), se llega a:

$$d = \frac{\left(\sum_{h=1}^{L} V_h\right)^2}{\sum_{h=1}^{L} V_h^2} \tag{C.2}$$

Se ve que el máximo valor que puede tener esta expresión es L, el número de estratos. Pero no hay que olvidar que esta solución es para un diseño en particular. De manera general, sea n_h, el número de Unidades Primarias de muestreo en el estrato h, entonces,

$$d = \frac{\left(\sum_{h=1}^{L} V_h/n_h\right)^2}{\sum_{h=1}^{L} \frac{V_h^2}{n_h^2(n_h-1)}} \tag{C.3}$$

Si todos los n_h son iguales y lo mismo ocurre con todos los V_h/n_h, los grados de libertad son $(n - L)$, donde $n = \sum_{h=1}^{L} n_h$, es el total de Unidades Primarias de Muestreo en la muestra.

Rao también explica que para estimadores complejos, se puede tratar de usar el mismo enfoque basado en (??),

$$d = \frac{2(E(v(\hat{\Theta})))^2}{V(v(\hat{\Theta}))},$$

pero que en tales situaciones en difícil encontrar los componentes de los estratos de $E(v(\hat{\Theta}))$ y de $V(v(\hat{\Theta}))$. Sin embargo, en casi todas las cirscunstancias, se ha encontrado que los grados de libertad no exceden $(n - L)$, y suelen ser bastantes menores.

7.- PREESCOLARES

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

(SI NO HAY MARCAR CON 0 Y PASAR DIRECTAMENTE A LA SECCION 8
SI HAY MAS DE 3 NIÑOS ANOTE LA INFORMACION SIGUIENTE EN
ESTE MODULO, SOLO DE LOS 3 MAS PEQUEÑOS)

¿QUIEN ES LA MADRE DEL NIÑO?

ANOTAR LA CLAVE DE LA MADRE ASIGNADA EN
EL MODULO 4 (COMPOSICION FAMILIAR), SI LA
MADRE DEL NIÑO NO VIVE EN EL HOGAR ANOTAR "99"

NIÑO 1 NIÑO 2 NIÑO 3

NIÑO 1

NOMBRE DEL NIÑO

APELLIDO PATERNO APELLIDO MATERNO NOMBRE

SEXO

M = MASCULINO
F = FEMENINO

FECHA DE NACIMIENTO

DIA MES 199 AÑO

PESO

K G

TALLA

CMS

LACTANCIA MATERNA

¿FUE ALIMENTADO AL SENO MATERNO?

SI NO

¿DURANTE CUANTOS MESES?

SI ACTUALMENTE AUN TOMA, ANOTAR 93

USO DE BIBERON

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

SI NO

SI LA RESPUESTA FUE SI ¿A QUE EDAD INICIO?

MESES

ABLACTACION

¿A QUE EDAD RECIBIO POR PRIMERA VEZ OTROS ALIMENTOS DISTINTOS A LA LECHE O LIQUIDOS?

MESES

SI NO HA SIDO AUN ABLACTADO, ANOTAR 99

ENFERMEDADES

¿HA ESTADO ENFERMO LOS ULTIMOS 15 DIAS?

SI NO

SI LA RESPUESTA FUE SI ¿DE QUE SE ENFERMO?

1. DIARREA
2. INFECCION RESPIRATORIA
3. OTRA

NIÑO 2

NOMBRE DEL NIÑO

APELLIDO PATERNO APELLIDO MATERNO NOMBRE

SEXO

M = MASCULINO
F = FEMENINO

FECHA DE NACIMIENTO

DIA MES 199 AÑO

PESO

K G

TALLA

CMS

LACTANCIA MATERNA

¿FUE ALIMENTADO AL SENO MATERNO?

SI NO

¿DURANTE CUANTOS MESES?

SI ACTUALMENTE AUN TOMA, ANOTAR 93

USO DE BIBERON

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

SI NO

SI LA RESPUESTA FUE SI ¿A QUE EDAD INICIO?

MESES

ABLACTACION

¿A QUE EDAD RECIBIO POR PRIMERA VEZ OTROS ALIMENTOS DISTINTOS A LA LECHE O LIQUIDOS?

MESES

SI NO HA SIDO AUN ABLACTADO, ANOTAR 99

ENFERMEDADES

¿HA ESTADO ENFERMO LOS ULTIMOS 15 DIAS?

SI NO

SI LA RESPUESTA FUE SI ¿DE QUE SE ENFERMO?

1. DIARREA
2. INFECCION RESPIRATORIA
3. OTRA

NIÑO 3

NOMBRE DEL NIÑO

APELLIDO PATERNO APELLIDO MATERNO NOMBRE

SEXO

M = MASCULINO
F = FEMENINO

FECHA DE NACIMIENTO

DIA MES 199 AÑO

PESO

K G

TALLA

CMS

LACTANCIA MATERNA

¿FUE ALIMENTADO AL SENO MATERNO?

SI NO

¿DURANTE CUANTOS MESES?

SI ACTUALMENTE AUN TOMA, ANOTAR 93

USO DE BIBERON

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

SI NO

SI LA RESPUESTA FUE SI ¿A QUE EDAD INICIO?

MESES

ABLACTACION

¿A QUE EDAD RECIBIO POR PRIMERA VEZ OTROS ALIMENTOS DISTINTOS A LA LECHE O LIQUIDOS?

MESES

SI NO HA SIDO AUN ABLACTADO, ANOTAR 99

ENFERMEDADES

¿HA ESTADO ENFERMO LOS ULTIMOS 15 DIAS?

SI NO

SI LA RESPUESTA FUE SI ¿DE QUE SE ENFERMO?

1. DIARREA
2. INFECCION RESPIRATORIA
3. OTRA

TABLA E.1 : ESTIMADOR Y VARIANZA JACKKNIFE, LA ESTIMACION POR FORMULA DE LOS ESTIMADORES DE RAZON, LA BASADA EN UN MUESTREO ALEATORIO SIMPLE, Y EL DEFT CORRESPONDIENTE

Nota: No se deben hacer inferencias cuando cv > 0.2.

ESTADO NO, NOMBRE		GRUPO NUTRICIONAL	ESTIMADOR JACKKNIFE*	EST, RAZON COMBINADO**	EST, RAZON M,A,S,***	VAR JACKKNIFE*	VAR (Rcomb,)**	VAR (Rmas)***	DEFF
1 AGUASCALIENTES									
L=5	**#**	*NORMALES y casi N.*	0,671309	0,671347	0,657143	0,000632	0,000626	0,000795	0,7874
n= 385	**#**	*BAJITOS GORDITOS*	0,150033	0,150159	0,158442	0,000327	0,000297	0,00036	0,8250
cv=0.068466	**#**	*MAL PARA LA EDAD*	0,091889	0,091869	0,096104	0,000564	0,000515	0,000252	2,0437
		MAL PARA LA TALLA	0,071746	0,071514	0,072727	0,000341	0,000239	0,000178	1,3427
		MAL EDAD Y TALLA	0,015022	0,015111	0,015584	0,000063	0,000057	0,000039	1,4615
2 BAJA CALIFORNIA									
L=2		*NORMALES y casi N.*	0,824199	0,822218	0,768293	0,000879	0,000789	0,001176	0,6709
n=164		*BAJITOS GORDITOS*	0,045203	0,043814	0,103659	0,001894	0,001736	0,000524	3,3130
cv=0.144709		*MAL PARA LA EDAD*	0,043075	0,042996	0,054878	0,000008	0,000007	0,0003	0,0233
		MAL PARA LA TALLA	0,085505	0,088651	0,04878	0,005723	0,003294	0,00029	11,3586
		MAL EDAD Y TALLA	0,002018	0,002321	0,02439	0,000002	0,000001	0,000144	0,0069
3 B. C. SUR									
L= 4		*NORMALES y casi N.*	0,638132	0,647499	0,722222	0,001064	0,000838	0,002653	0,3159
n=90		*BAJITOS GORDITOS*	0,137942	0,139227	0,122222	0,000767	0,000529	0,00152	0,3480
cv=0.292971		*MAL PARA LA EDAD*	0,123643	0,117197	0,055556	0,005742	0,003066	0,000598	5,1271
		MAL PARA LA TALLA	0,05641	0,057479	0,077778	0,000583	0,000259	0,001053	0,2460
		MAL EDAD Y TALLA	0,043872	0,038599	0,022222	0,001878	0,000997	0,000244	4,0861
4 CAMPECHE									
L=8		*NORMALES y casi N.*	0,43755	0,43774	0,445313	0,001066	0,001008	0,000282	3,5745
n=1024		*BAJITOS GORDITOS*	0,312537	0,312249	0,302734	0,000344	0,000294	0,000213	1,3803
cv= 0.104915		*MAL PARA LA EDAD*	0,198832	0,198745	0,188477	0,000424	0,000429	0,000155	2,7677
		MAL PARA LA TALLA	0,030358	0,03042	0,038086	0,000021	0,000021	0,000037	0,5676
		MAL EDAD Y TALLA	0,020723	0,020846	0,025391	0,00003	0,00003	0,000023	1,3043

* El estimador jackknife es el que propone Rao (1996), ajustando los factores de expansion.

** Estimador por las formulas que provienen de la linearizacion.

***Estimador que asume un muestreo aleatorio simple.

TABLA E.1 : ESTIMADOR Y VARIANZA JACKKNIFE, LA ESTIMACION POR FORMULA DE LOS ESTIMADORES DE RAZON, LA BASADA EN UN MUESTREO ALEATORIO SIMPLE, Y EL DEFT CORRESPONDIENTE

Nota: No se deben hacer inferencias cuando cv > 0.2.

ESTADO NO. NOMBRE	GRUPO NUTRICIONAL	ESTIMADOR JACKKNIFE*	EST, RAZON COMBINADO**	EST, RAZON M,A,S,***	VAR JACKKNIFE*	VAR (Rcomb,)**	VAR (Rmas)***	DEFF
5 COAHUILA								
L=5	*NORMALES y casi N.*	0,686418	0,685303	0,684597	0,001912	0,001334	0,000624	2,1378
n=409	*BAJITOS GORDITOS*	0,135864	0,136843	0,158924	0,000331	0,000375	0,00033	1,1364
cv=0.183529	*MAL PARA LA EDAD*	0,101482	0,101732	0,085575	0,001287	0,001026	0,000206	4,9806
	MAL PARA LA TALLA	0,061643	0,061512	0,05868	0,000273	0,00022	0,000134	1,6418
	MAL EDAD Y TALLA	0,014594	0,01461	0,012225	0,000097	0,000086	0,000029	2,9655
6 COLIMA								
L=4	*NORMALES y casi N.*	0,509337	0,507111	0,530364	0,001917	0,001316	0,001284	1,0249
n=247	*BAJITOS GORDITOS*	0,219793	0,220826	0,190283	0,001934	0,001269	0,000671	1,8912
cv= 0.197409	*MAL PARA LA EDAD*	0,057743	0,058108	0,060729	0,000265	0,000228	0,000242	0,9421
	MAL PARA LA TALLA	0,187915	0,188345	0,186235	0,002441	0,001279	0,000682	1,8754
	MAL EDAD Y TALLA	0,025212	0,025609	0,032389	0,000041	0,000019	0,000122	0,1557
7 CHIAPAS								
L= 25	*NORMALES y casi N.*	0,36685	0,367183	0,396943	0,000315	0,000309	0,000149	2,0738
n=1897	*BAJITOS GORDITOS*	0,361085	0,360928	0,337902	0,000272	0,000267	0,000128	2,0859
cv= 0.083302	*MAL PARA LA EDAD*	0,185102	0,184939	0,176595	0,000234	0,000232	0,000083	2,7952
	MAL PARA LA TALLA	0,063848	0,063863	0,066421	0,000066	0,000053	0,000034	1,5588
	MAL EDAD Y TALLA	0,023116	0,023087	0,02214	0,000045	0,000044	0,000012	3,6667
8 CHIHUAHUA								
L=4	*NORMALES y casi N.*	0,724303	0,724842	0,716	0,000352	0,000407	0,001188	0,3426
n=250	*BAJITOS GORDITOS*	0,121658	0,123901	0,124	0,001358	0,001005	0,000481	2,0894
cv=0.333081	*MAL PARA LA EDAD*	0,073829	0,072895	0,068	0,000356	0,000309	0,000333	0,9279
	MAL PARA LA TALLA	0,08021	0,078362	0,092	0,000809	0,000597	0,000344	1,7355
	MAL EDAD Y TALLA	0	0	0	0	0	0	——

* El estimador jackknife es el que propone Rao (1996), ajustando los factores de expansion.

** Estimador por las formulas que provienen de la linearizacion.

***Estimador que asume un muestreo aleatorio simple.

TABLA E.1 : ESTIMADOR Y VARIANZA JACKKNIFE, LA ESTIMACION POR FORMULA DE LOS ESTIMADORES DE RAZON, LA BASADA EN UN MUESTREO ALEATORIO SIMPLE, Y EL DEFT CORRESPONDIENTE

Nota: No se deben hacer inferencias cuando cv > 0.2.

ESTADO NO, NOMBRE	GRUPO NUTRICIONAL	ESTIMADOR JACKKNIFE*	EST, RAZON COMBINADO**	EST, RAZON M,A,S,***	VAR JACKKNIFE*	VAR (Rcomb,)**	VAR (Rmas)***	DEFF
10 DURANGO								
L= 8	*NORMALES y casi N.*	0,674122	0,673611	0,671256	0,0006	0,000827	0,000505	1,6376
n=581	*BAJITOS GORDITOS*	0,182677	0,182426	0,184165	0,000215	0,000236	0,000278	0,8489
cv=0.123927	*MAL PARA LA EDAD*	0,055026	0,055084	0,051635	0,000167	0,000178	0,000091	1,9560
	MAL PARA LA TALLA	0,082541	0,08314	0,084337	0,000438	0,000635	0,000146	4,3493
	MAL EDAD Y TALLA	0,005633	0,005739	0,008606	0,000004	0,000006	0,000015	0,4000
11 GUANAJUATO								
L=19	*NORMALES y casi N.*	0,586051	0,586011	0,582249	0,000539	0,000538	0,000389	1,3830
n=845	*BAJITOS GORDITOS*	0,180556	0,180556	0,173964	0,000516	0,000486	0,000187	2,5989
cv=0.106484	*MAL PARA LA EDAD*	0,105145	0,105159	0,114793	0,000239	0,000229	0,000126	1,8175
	MAL PARA LA TALLA	0,092979	0,092989	0,091124	0,000092	0,000083	0,000099	0,8384
	MAL EDAD Y TALLA	0,035269	0,035284	0,03787	0,000059	0,000056	0,000058	0,9655
12 GUERRERO								
L=20	*NORMALES y casi N.*	0,391806	0,392367	0,420244	0,000451	0,000365	0,000164	2,2256
n=1887	*BAJITOS GORDITOS*	0,230392	0,230391	0,219396	0,000157	0,000129	0,000098	1,3163
cv=0.066246	*MAL PARA LA EDAD*	0,257352	0,256959	0,244833	0,000442	0,000351	0,000114	3,0789
	MAL PARA LA TALLA	0,05857	0,05865	0,061473	0,000033	0,000023	0,000031	0,7419
	MAL EDAD Y TALLA	0,061881	0,061634	0,054054	0,000065	0,000053	0,000027	1,9630
13 HIDALGO								
L=15	*NORMALES y casi N.*	0,455319	0,460365	0,469408	0,00067	0,000611	0,000305	2,0033
n=997	*BAJITOS GORDITOS*	0,288198	0,286675	0,28987	0,000574	0,000404	0,000218	1,8532
cv= 0.105027	*MAL PARA LA EDAD*	0,156622	0,15718	0,148445	0,000286	0,000186	0,000144	1,2917
	MAL PARA LA TALLA	0,065156	0,063185	0,060181	0,00024	0,000152	0,000062	2,4516
	MAL EDAD Y TALLA	0,034705	0,032596	0,032096	0,000106	0,000071	0,000038	1,8684

* El estimador jackknife es el que propone Rao (1996), ajustando los factores de expansion.

** Estimador por las formulas que provienen de la linearizacion.

***Estimador que asume un muestreo aleatorio simple.

TABLA E.1 : ESTIMADOR Y VARIANZA JACKKNIFE, LA ESTIMACION POR FORMULA DE LOS ESTIMADORES DE RAZON, LA BASADA EN UN MUESTREO ALEATORIO SIMPLE, Y EL DEFT CORRESPONDIENTE

Nota: No se deben hacer inferencias cuando cv > 0.2.

ESTADO NO, NOMBRE	**GRUPO** NUTRICIONAL	ESTIMADOR JACKKNIFE*	EST, RAZON COMBINADO**	EST, RAZON M,A,S,***	VAR JACKKNIFE*	VAR (Rcomb,)**	VAR (Rmas)***	DEFF
14 JALISCO								
L=12	*NORMALES y casi N.*	0,717439	0,717678	0,714447	0,000256	0,000247	0,000313	0,7891
n=886	*BAJITOS GORDITOS*	0,138288	0,138066	0,137698	0,000136	0,000151	0,000152	0,9934
cv=0.090785	*MAL PARA LA EDAD*	0,040024	0,040128	0,042889	0,000037	0,000037	0,000056	0,6607
	MAL PARA LA TALLA	0,09624	0,096133	0,095937	0,000054	0,000063	0,000106	0,5943
	MAL EDAD Y TALLA	0,008009	0,007996	0,009029	0,000008	0,000008	0,00001	0,8000
15 MEXICO								
L=27	*NORMALES y casi N.*	0,533731	0,533757	0,56493	0,000553	0,000664	0,000267	2,4869
n=1209	*BAJITOS GORDITOS*	0,268386	0,268416	0,25062	0,000414	0,00039	0,000169	2,3077
cv=0.097556	*MAL PARA LA EDAD*	0,13611	0,136056	0,123242	0,000117	0,000109	0,000104	1,0481
	MAL PARA LA TALLA	0,054748	0,054723	0,052109	0,000148	0,000089	0,000048	1,8542
	MAL EDAD Y TALLA	0,007024	0,007048	0,009098	0,000007	0,000007	0,000007	1,0000
16 MICHOACAN								
L=19	*NORMALES y casi N.*	0,585021	0,585243	0,579094	0,000822	0,000621	0,000207	3,0000
n=1435	*BAJITOS GORDITOS*	0,241942	0,241802	0,233449	0,000227	0,000163	0,000135	1,2074
cv=0.131001	*MAL PARA LA EDAD*	0,090876	0,090833	0,09338	0,000218	0,000153	0,000071	2,1549
	MAL PARA LA TALLA	0,069284	0,069267	0,080836	0,00027	0,000144	0,000059	2,4407
	MAL EDAD Y TALLA	0,012877	0,012855	0,01324	0,000014	0,00001	0,00001	1,0000
17 MORELOS								
L=11	*NORMALES y casi N.*	0,557163	0,557681	0,578475	0,001476	0,000841	0,000462	1,8203
n=669	*BAJITOS GORDITOS*	0,250185	0,249776	0,231689	0,000366	0,000187	0,00028	0,6679
cv= 0.131155	*MAL PARA LA EDAD*	0,144642	0,144647	0,139013	0,000871	0,000521	0,000195	2,6718
	MAL PARA LA TALLA	0,039005	0,038903	0,040359	0,000088	0,00004	0,000058	0,6897
	MAL EDAD Y TALLA	0,009004	0,008993	0,010463	0,000024	0,000014	0,000015	0,9333

* El estimador jackknife es el que propone Rao (1996), ajustando los factores de expansion.

** Estimador por las formulas que provienen de la linearizacion.

***Estimador que asume un muestreo aleatorio simple.

TABLA E.1 : ESTIMADOR Y VARIANZA JACKKNIFE, LA ESTIMACION POR FORMULA DE LOS ESTIMADORES DE RAZON, LA BASADA EN UN MUESTREO ALEATORIO SIMPLE, Y EL DEFT CORRESPONDIENTE

Nota: No se deben hacer inferencias cuando cv > 0.2.

ESTADO NO, NOMBRE	GRUPO NUTRICIONAL	ESTIMADOR JACKKNIFE*	EST, RAZON COMBINADO**	EST, RAZON M,A,S,***	VAR JACKKNIFE*	VAR (Rcomb,)**	VAR (Rmas)***	DEFF
18 NAYARIT								
L=5	*NORMALES y casi N.*	0,638913	0,637705	0,652174	0,0039	0,003404	0,001434	2,3738
n=207	*BAJITOS GORDITOS*	0,17982	0,180209	0,183575	0,001537	0,001439	0,000797	1,8055
cv=0.148956	*MAL PARA LA EDAD*	0,074105	0,074775	0,067633	0,000732	0,000687	0,000325	2,1138
	MAL PARA LA TALLA	0,095601	0,095646	0,086957	0,000926	0,000572	0,000397	1,4408
	MAL EDAD Y TALLA	0,011561	0,011665	0,009662	0,000066	0,000061	0,000046	1,3261
19 NUEVO LEON								
L= 3	*NORMALES y casi N.*	0,596189	0,595749	0,597087	0,000253	0,000281	0,001538	0,1827
n=206	*BAJITOS GORDITOS*	0,170209	0,170615	0,150485	0,001436	0,000837	0,00066	1,2682
cv= 0.232378	*MAL PARA LA EDAD*	0,076978	0,077018	0,067961	0,000524	0,000401	0,000304	1,3191
	MAL PARA LA TALLA	0,143814	0,143929	0,169903	0,002626	0,00123	0,000828	1,4855
	MAL EDAD Y TALLA	0,01281	0,012689	0,014563	0,000071	0,000071	0,000068	1,0441
20 OAXACA								
L=30	*NORMALES y casi N.*	0,404345	0,404333	0,407407	0,000408	0,000342	0,000172	1,9884
n=1728	*BAJITOS GORDITOS*	0,301366	0,30132	0,284144	0,000402	0,000384	0,000124	3,0968
cv= 0.090278	*MAL PARA LA EDAD*	0,211899	0,211909	0,215856	0,00045	0,000426	0,000111	3,8378
	MAL PARA LA TALLA	0,052892	0,052922	0,056713	0,000082	0,000066	0,000033	2,0000
	MAL EDAD Y TALLA	0,029498	0,029514	0,03588	0,000025	0,000024	0,00002	1,2000
21 PUEBLA								
L=25	*NORMALES y casi N.*	0,396173	0,396262	0,407791	0,000268	0,000373	0,000123	3,0325
n=2413	*BAJITOS GORDITOS*	0,279911	0,279822	0,276005	0,000263	0,00026	0,00009	2,8889
cv= 0.070055	*MAL PARA LA EDAD*	0,192555	0,192632	0,195193	0,000126	0,000139	0,000072	1,9306
	MAL PARA LA TALLA	0,084756	0,084643	0,074182	0,000166	0,000108	0,000032	3,3750
	MAL EDAD Y TALLA	0,046605	0,046641	0,04683	0,000034	0,000042	0,000019	2,2105

* El estimador *jackknife es el que* propone Rao (1996), ajustando los factores de expansion.

** Estimador por las formulas que provienen de la linearizacion.

***Estimador que asume un muestreo aleatorio simple.

TABLA E.1 : ESTIMADOR Y VARIANZA JACKKNIFE, LA ESTIMACION POR FORMULA DE LOS ESTIMADORES DE RAZON, LA BASADA EN UN MUESTREO ALEATORIO SIMPLE, Y EL DEFT CORRESPONDIENTE

Nota: No se deben hacer inferencias cuando cv > 0.2.

ESTADO NO. NOMBRE	GRUPO NUTRICIONAL	ESTIMADOR JACKKNIFE*	EST. RAZON COMBINADO**	EST. RAZON M.A.S.***	VAR JACKKNIFE*	VAR (Rcomb.)**	VAR (Rmas)***	DEFF
22 QUERETARO								
L=9	*NORMALES y casi N.*	0,554111	0,553556	0,565284	0,000722	0,000477	0,000475	1,0042
n=651	*BAJITOS GORDITOS*	0,227926	0,228162	0,215054	0,000308	0,00028	0,000282	0,9929
cv= 0.191388	*MAL PARA LA EDAD*	0,154936	0,155423	0,15361	0,000689	0,000627	0,000236	2,6568
	MAL PARA LA TALLA	0,04878	0,048619	0,047619	0,000137	0,000121	0,000079	1,5316
	MAL EDAD Y TALLA	0,014248	0,01424	0,018433	0,000031	0,000028	0,000027	1,0370
23 QUINTANA ROO								
L=3	*NORMALES y casi N.*	0,435305	0,433718	0,38756	0,001361	0,001792	0,001446	1,2393
n=209	*BAJITOS GORDITOS*	0,344827	0,34628	0,344498	0,001041	0,001321	0,001119	1,1805
cv= 0.252427	*MAL PARA LA EDAD*	0,149345	0,149127	0,181818	0,001009	0,000999	0,0008	1,2488
	MAL PARA LA TALLA	0,053025	0,053502	0,066986	0,000693	0,000455	0,000351	1,2963
	MAL EDAD Y TALLA	0,017497	0,017373	0,019139	0,000048	0,000042	0,000089	0,4719
24 SAN LUIS POTOSI								
L=11	*NORMALES y casi N.*	0,564217	0,564447	0,576342	0,000926	0,001042	0,000286	3,6434
n=1192	*BAJITOS GORDITOS*	0,21884	0,21888	0,209732	0,000501	0,000519	0,000151	3,4371
cv= 0.081463	*MAL PARA LA EDAD*	0,135356	0,135317	0,139262	0,000224	0,000231	0,000114	2,0263
	MAL PARA LA TALLA	0,059876	0,059751	0,05453	0,000141	0,000116	0,000046	2,5217
	MAL EDAD Y TALLA	0,021711	0,021606	0,020134	0,000036	0,000036	0,000019	1,8947
25 SINALOA								
L= 14	*NORMALES y casi N.*	0,721803	0,721998	0,725067	0,000245	0,000308	0,000325	0,9477
n=742	*BAJITOS GORDITOS*	0,129829	0,129389	0,132075	0,000279	0,000263	0,000162	1,6235
cv= 0.089434	*MAL PARA LA EDAD*	0,051115	0,051274	0,045822	0,000059	0,000074	0,000064	1,1563
	MAL PARA LA TALLA	0,081842	0,081834	0,086253	0,000149	0,000181	0,000114	1,5877
	MAL EDAD Y TALLA	0,015412	0,015506	0,010782	0,000051	0,000084	0,000014	6,0000

* El estimador jackknife es el que propone Rao (1996), ajustando los factores de expansion.

** Estimador por las formulas que provienen de la linearizacion.

***Estimador que asume un muestreo aleatorio simple.

TABLA E.1 : ESTIMADOR Y VARIANZA JACKKNIFE, LA ESTIMACION POR FORMULA DE LOS ESTIMADORES DE RAZON, LA BASADA EN UN MUESTREO ALEATORIO SIMPLE, Y EL DEFT CORRESPONDIENTE

Nota: No se deben hacer inferencias cuando cv > 0.2.

ESTADO NO, NOMBRE	GRUPO NUTRICIONAL	ESTIMADOR JACKKNIFE*	EST, RAZON COMBINADO**	EST, RAZON M,A,S,***	VAR JACKKNIFE*	VAR (Rcomb,)**	VAR (Rmas)***	DEFF
26 SONORA								
L=6	*NORMALES y casi N.*	0,768715	0,768588	0,769547	0,000965	0,00097	0,000825	1,1758
n=243	*BAJITOS GORDITOS*	0,133197	0,13336	0,123457	0,000376	0,000359	0,000464	0,7737
cv= 0.187546	*MAL PARA LA EDAD*	0,024944	0,024864	0,020576	0,000309	0,000281	0,000082	3,4268
	MAL PARA LA TALLA	0,068712	0,068776	0,078189	0,000202	0,000382	0,000329	1,1611
	MAL EDAD Y TALLA	0,004431	0,004412	0,00823	0,000012	0,000011	0,000033	0,3333
27 TABASCO								
L=17	*NORMALES y casi N.*	0,58076	0,581191	0,607169	0,000287	0,000454	0,000236	1,9237
n=1367	*BAJITOS GORDITOS*	0,21323	0,213235	0,219459	0,000307	0,000292	0,000134	2,1791
cv=0.061831	*MAL PARA LA EDAD*	0,117906	0,11777	0,108998	0,000047	0,000043	0,000076	0,5658
	MAL PARA LA TALLA	0,072502	0,072259	0,051939	0,00007	0,000067	0,000036	1,8611
	MAL EDAD Y TALLA	0,015603	0,015544	0,012436	0,000042	0,000039	0,00001	3,9000
28 TAMAULIPAS								
L=5	*NORMALES y casi N.*	0,64348	0,644342	0,674699	0,000541	0,000585	0,000897	0,6522
n=332	*BAJITOS GORDITOS*	0,185575	0,185757	0,171687	0,001239	0,001083	0,000453	2,3907
cv= 0.116067	*MAL PARA LA EDAD*	0,04626	0,046298	0,048193	0,000284	0,00022	0,000163	1,3497
	MAL PARA LA TALLA	0,11882	0,117829	0,099398	0,001513	0,000613	0,000281	2,1815
	MAL EDAD Y TALLA	0,005866	0,005773	0,006024	0,000023	0,000017	0,000018	0,9444
29 TLAXCALA								
L= 4	*NORMALES y casi N.*	0,530572	0,529848	0,505263	0,001961	0,002201	0,000822	2,6776
n=380	*BAJITOS GORDITOS*	0,231587	0,231415	0,247368	0,000561	0,000492	0,000503	0,9781
cv= 0.218980	*MAL PARA LA EDAD*	0,136025	0,136753	0,15	0,000482	0,000592	0,000381	1,5538
	MAL PARA LA TALLA	0,071956	0,072348	0,065789	0,000332	0,000208	0,000161	1,2919
	MAL EDAD Y TALLA	0,02986	0,029635	0,031579	0,000052	0,000053	0,000118	0,4492

* *El estimador jackknife es el que propone Rao (1996), ajustando los factores de expansion.*

** Estimador por las formulas que provienen de la linearizacion.

***Estimador que asume un muestreo aleatorio simple.

TABLA E.1 : ESTIMADOR Y VARIANZA JACKKNIFE, LA ESTIMACION POR FORMULA DE LOS ESTIMADORES DE RAZON, LA BASADA EN UN MUESTREO ALEATORIO SIMPLE, Y EL DEFT CORRESPONDIENTE

Nota: No se deben hacer inferencias cuando cv > 0.2.

ESTADO NO. NOMBRE		GRUPO NUTRICIONAL	ESTIMADOR JACKKNIFE*	EST. RAZON COMBINADO**	EST. RAZON M.A.S.***	VAR JACKKNIFE*	VAR (Rcomb.)**	VAR (Rmas)***	D E F F
30	VERACRUZ								
	L=36	*NORMALES y casi N.*	0,532622	0,532529	0,513472	0,000393	0,000406	0,000118	3,4407
	n=2635	*BAJITOS GORDITOS*	0,237758	0,237826	0,249336	0,00018	0,000177	0,000078	2,2692
	cv= 0.066841	*MAL PARA LA EDAD*	0,14527	0,145241	0,149146	0,000128	0,000123	0,000055	2,2364
		MAL PARA LA TALLA	0,062052	0,062106	0,065275	0,000042	0,000055	0,000026	2,1154
		MAL EDAD Y TALLA	0,022298	0,022299	0,02277	0,000013	0,000012	0,000009	1,3333
31	YUCATAN								
	L=6	*NORMALES y casi N.*	0,329176	0,328716	0,345697	0,000804	0,001185	0,000408	2,9044
	n=674	*BAJITOS GORDITOS*	0,309285	0,308406	0,290801	0,000197	0,000226	0,000331	0,6828
	cv=0.116179	*MAL PARA LA EDAD*	0,222732	0,223876	0,225519	0,001049	0,001672	0,000298	5,6107
		MAL PARA LA TALLA	0,088258	0,088606	0,08457	0,00024	0,00027	0,000137	1,9708
		MAL EDAD Y TALLA	0,050549	0,050397	0,053412	0,0001	0,000107	0,000072	1,4861
32	ZACATECAS								
	L=9	*NORMALES y casi N.*	0,637442	0,637893	0,646113	0,00043	0,000494	0,000413	1,1961
	n=746	*BAJITOS GORDITOS*	0,170037	0,169942	0,176944	0,000419	0,000396	0,000221	1,7919
	cv= 0.147150	*MAL PARA LA EDAD*	0,100353	0,100258	0,081769	0,000603	0,000567	0,000109	5,2018
		MAL PARA LA TALLA	0,084947	0,084719	0,087131	0,000144	0,000125	0,000124	1,0081
		MAL EDAD Y TALLA	0,00722	0,007187	0,008043	0,000023	0,000022	0,000011	2,0000

 # L = no. de estratos; n= niños encuestados;
cv = coeficiente de variación del denominador del estimador de razón.

* El estimador jackknife es el que propone Rao (1996), ajustando los factores de expansion.

** Estimador por las formulas que provienen de la linearizacion.

***Estimador que asume un muestreo aleatorio simple.

TABLA E.2 : ESTIMADOR Y VARIANZA JACKKNIFE (SEGUN RAO Y WOLTER) JUNTO CON LA ESTIMACION POR FORMULA DEL ESTIMADOR DE RAZON . POR ESTADO Y A NIVEL NACIONAL

Nota: No se deben hacer inferencias en los estados donde cv > 0.2

ESTADO No. NOMBRE	GRUPO NUTRICIONAL	JACKKNIFE ESTIMADOR RAO*	JACKKNIFE ESTIMADOR Wolter*	Formula/lin. EST. RAZON COMBINADO	JACKKNIFE VAR RAO*	JACKKNIFE VAR Wolter*	Formula/lin. VAR (Rcomb.)
5 COAHUILA							
L=5	*NORMALES y casi N.*	0,686418	0,683681	0,685303	0,001912	0,00174	0,001334
n=409	*BAJITOS GORDITOS*	0,135864	0,138665	0,136843	0,000331	0,000441	0,000375
cv=0.183529	*MAL PARA LA EDAD*	0,101482	0,102368	0,101732	0,001287	0,001327	0,001026
	MAL PARA LA TALLA	0,061643	0,060495	0,061512	0,000273	0,000225	0,00022
	MAL EDAD Y TALLA	0,014594	0,014791	0,01461	0,000097	0,00011	0,000086
6 COLIMA							
L=4	*NORMALES y casi N.*	0,509337	0,507064	0,507111	0,001917	0,001767	0,001316
n=247	*BAJITOS GORDITOS*	0,219793	0,213136	0,220826	0,001934	0,002738	0,001269
cv= 0.197409	*MAL PARA LA EDAD*	0,057743	0,057877	0,058108	0,000265	0,000344	0,000228
	MAL PARA LA TALLA	0,187915	0,193597	0,188345	0,002441	0,002619	0,001279
	MAL EDAD Y TALLA	0,025212	0,028321	0,025609	0,000041	0,000029	0,000019
7 CHIAPAS							
L= 25	*NORMALES y casi N.*	0,36685	0,371494	0,367183	0,000315	0,00036	0,000309
n=1897	*BAJITOS GORDITOS*	0,361085	0,360065	0,360928	0,000272	0,000281	0,000267
cv= 0,083302	*MAL PARA LA EDAD*	0,185102	0,182333	0,184939	0,000234	0,000244	0,000232
	MAL PARA LA TALLA	0,063848	0,063518	0,063863	0,000066	0,000074	0,000053
	MAL EDAD Y TALLA	0,023116	0,022591	0,023087	0,000045	0,000046	0,000044
8 CHIHUAHUA							
L=4	*NORMALES y casi N.*	0,724303	0,725553	0,724842	0,000352	0,000367	0,000407
n=250	*BAJITOS GORDITOS*	0,121658	0,135089	0,123901	0,001358	0,001262	0,001005
cv=0.333081	*MAL PARA LA EDAD*	0,073829	0,069153	0,072895	0,000356	0,000354	0,000309
	MAL PARA LA TALLA	0,08021	0,070205	0,078362	0,000809	0,000777	0,000597
	MAL EDAD Y TALLA	0	0	0	0	0	0

* El estimador de Rao se obtiene ajustando los factores de expansion; El estimador Wolter contempla la obtencion de pseudovalores y fcpf .

**L= estratos; n= niños enc.; cv = coef var. del denomin. del est. de razon.

TABLA E.2 : ESTIMADOR Y VARIANZA JACKKNIFE (SEGUN RAO Y WOLTER)
JUNTO CON LA ESTIMACION POR FORMULA DEL ESTIMADOR DE RAZON .
POR ESTADO Y A NIVEL NACIONAL

Nota: No se deben hacer inferencias en los estados donde cv > 0.2

ESTADO No. NOMBRE	GRUPO NUTRICIONAL	JACKKNIFE ESTIMADOR RAO*	JACKKNIFE ESTIMADOR Wolter*	Formula/lin. EST. RAZON COMBINADO	JACKKNIFE VAR RAO*	JACKKNIFE VAR Wolter*	Formula/lin. VAR (Rcomb.)
1 AGUASCALIENTES							
L=5 **	*NORMALES y casi N.*	0,671309	0,671493	0,671347	0,000632	0,000631	0,000626
n= 385 **	*BAJITOS GORDITOS*	0,150033	0,150743	0,150159	0,000327	0,000326	0,000297
cv=0.068466 **	*MAL PARA LA EDAD*	0,091889	0,091811	0,091869	0,000564	0,000564	0,000515
	MAL PARA LA TALLA	0,071746	0,07045	0,071514	0,000341	0,000341	0,000239
	MAL EDAD Y TALLA	0,015022	0,015502	0,015111	0,000063	0,000063	0,000057
2 BAJA CALIFORNIA							
L=2	*NORMALES y casi N.*	0,824199	0,817621	0,822218	0,000879	0,000892	0,000789
n=164	*BAJITOS GORDITOS*	0,045203	0,038352	0,043814	0,001894	0,001868	0,001736
cv=0.144709	*MAL PARA LA EDAD*	0,043075	0,042784	0,042996	0,000008	0,000008	0,000007
	MAL PARA LA TALLA	0,085505	0,098806	0,088651	0,005723	0,005642	0,003294
	MAL EDAD Y TALLA	0,002018	0,002437	0,002321	0,000002	0,000003	0,000001
3 B. C. SUR							
L= 4	*NORMALES y casi N.*	0,638132	0,672737	0,647499	0,001064	0,001778	0,000838
n=90	*BAJITOS GORDITOS*	0,137942	0,138164	0,139227	0,000767	0,000969	0,000529
cv=0.292971	*MAL PARA LA EDAD*	0,123643	0,106096	0,117197	0,005742	0,005363	0,003066
	MAL PARA LA TALLA	0,05641	0,060431	0,057479	0,000583	0,000777	0,000259
	MAL EDAD Y TALLA	0,043872	0,022576	0,038599	0,001878	0,001801	0,000997
4 CAMPECHE							
L=8	*NORMALES y casi N.*	0,43755	0,449197	0,43774	0,001066	0,001206	0,001008
n=1024	*BAJITOS GORDITOS*	0,312537	0,306177	0,312249	0,000344	0,000369	0,000294
cv= 0.104915	*MAL PARA LA EDAD*	0,198832	0,191883	0,198745	0,000424	0,000531	0,000429
	MAL PARA LA TALLA	0,030358	0,031612	0,03042	0,000021	0,000028	0,000021
	MAL EDAD Y TALLA	0,020723	0,021131	0,020846	0,00003	0,000042	0,00003

* *El estimador de Rao se obtiene ajustando* los factores de expansion; El estimador Wolter contempla la obtencion de pseudovalores y fcpf .

**L= estratos; n= niños enc.; cv = coef var. del denomin. del est. de razon.

TABLA E.2 : ESTIMADOR Y VARIANZA JACKKNIFE (SEGUN RAO Y WOLTER)
JUNTO CON LA ESTIMACION POR FORMULA DEL ESTIMADOR DE RAZON .
POR ESTADO Y A NIVEL NACIONAL

Nota: No se deben hacer inferencias en los estados donde cv > 0.2

ESTADO No. NOMBRE	GRUPO NUTRICIONAL	JACKKNIFE ESTIMADOR RAO*	JACKKNIFE ESTIMADOR Wolter*	Formula/lin. EST. RAZON COMBINADO	JACKKNIFE VAR RAO*	JACKKNIFE VAR Wolter*	Formula/lin. VAR (Rcomb.)
10 DURANGO							
L= 8	*NORMALES y casi N.*	0,674122	0,673789	0,673611	0,0006	0,000777	0,000827
n=581	*BAJITOS GORDITOS*	0,182677	0,181775	0,182426	0,000215	0,000254	0,000236
cv=0.123927	*MAL PARA LA EDAD*	0,055026	0,054928	0,055084	0,000167	0,000191	0,000178
	MAL PARA LA TALLA	0,082541	0,083521	0,08314	0,000438	0,000702	0,000635
	MAL EDAD Y TALLA	0,005633	0,005986	0,005739	0,000004	0,000007	0,000006
11 GUANAJUATO							
L=19	*NORMALES y casi N.*	0,586051	0,585296	0,586011	0,000539	0,000538	0,000538
n=845	*BAJITOS GORDITOS*	0,180556	0,180553	0,180556	0,000516	0,000515	0,000486
cv=0.106484	*MAL PARA LA EDAD*	0,105145	0,105397	0,105159	0,000239	0,000239	0,000229
	MAL PARA LA TALLA	0,092979	0,093177	0,092989	0,000092	0,000092	0,000083
	MAL EDAD Y TALLA	0,035269	0,035558	0,035284	0,000059	0,000059	0,000056
12 GUERRERO							
L=20	*NORMALES y casi N.*	0,391806	0,402671	0,392367	0,000451	0,000392	0,000365
n=1887	*BAJITOS GORDITOS*	0,230392	0,229098	0,230391	0,000157	0,000139	0,000129
cv=0.066246	*MAL PARA LA EDAD*	0,257352	0,250552	0,256959	0,000442	0,000377	0,000351
	MAL PARA LA TALLA	0,05857	0,059149	0,05865	0,000033	0,000033	0,000023
	MAL EDAD Y TALLA	0,061881	0,058562	0,061634	0,000065	0,000057	0,000053
13 HIDALGO							
L=15	*NORMALES y casi N.*	0,455319	0,540787	0,460365	0,00067	0,000519	0,000611
n=997	*BAJITOS GORDITOS*	0,288198	0,261704	0,286675	0,000574	0,000443	0,000404
cv= 0.105027	*MAL PARA LA EDAD*	0,156622	0,16601	0,15718	0,000286	0,000216	0,000186
	MAL PARA LA TALLA	0,065156	0,032297	0,063185	0,00024	0,000179	0,000152
	MAL EDAD Y TALLA	0,034705	0,00078	0,032596	0,000106	0,000079	0,000071

* El estimador de Rao se obtiene ajustando los factores de expansion; El estimador Wolter contempla la obtencion de pseudovalores y fcpf .
**L= estratos; n= niños enc.; cv = coef var. del denomin. del est. de razon.

TABLA E.2 : ESTIMADOR Y VARIANZA JACKKNIFE (SEGUN RAO Y WOLTER)
JUNTO CON LA ESTIMACION POR FORMULA DEL ESTIMADOR DE RAZON .
POR ESTADO Y A NIVEL NACIONAL

Nota: No se deben hacer inferencias en los estados donde cv > 0.2

ESTADO No. NOMBRE	GRUPO NUTRICIONAL	JACKKNIFE ESTIMADOR RAO*	JACKKNIFE ESTIMADOR Wolter*	Formula/lin. EST. RAZON COMBINADO	JACKKNIFE VAR RAO*	JACKKNIFE VAR Wolter*	Formula/lin. VAR (Rcomb.)
14 JALISCO							
L=12	*NORMALES y casi N.*	0,717439	0,712726	0,717678	0,000256	0,000266	0,000247
n=886	*BAJITOS GORDITOS*	0,138288	0,137366	0,138066	0,000136	0,000168	0,000151
cv=0.090785	*MAL PARA LA EDAD*	0,040024	0,04241	0,040128	0,000037	0,000041	0,000037
	MAL PARA LA TALLA	0,09624	0,099027	0,096133	0,000054	0,000056	0,000063
	MAL EDAD Y TALLA	0,008009	0,008493	0,007996	0,000008	0,000009	0,000008
15 MEXICO							
L=27	*NORMALES y casi N.*	0,533731	0,534401	0,533757	0,000553	0,000553	0,000664
n=1209	*BAJITOS GORDITOS*	0,268386	0,269174	0,268416	0,000414	0,000413	0,00039
cv=0.097556	*MAL PARA LA EDAD*	0,13611	0,13469	0,136056	0,000117	0,000116	0,000109
	MAL PARA LA TALLA	0,054748	0,054091	0,054723	0,000148	0,000148	0,000089
	MAL EDAD Y TALLA	0,007024	0,007644	0,007048	0,000007	0,000007	0,000007
16 MICHOACAN							
L=19	*NORMALES y casi N.*	0,585021	0,58897	0,585243	0,000822	0,000612	0,000621
n=1435	*BAJITOS GORDITOS*	0,241942	0,239889	0,241802	0,000227	0,00017	0,000163
cv=0.131001	*MAL PARA LA EDAD*	0,090876	0,09027	0,090833	0,000218	0,000164	0,000153
	MAL PARA LA TALLA	0,069284	0,068372	0,069267	0,00027	0,000202	0,000144
	MAL EDAD Y TALLA	0,012877	0,0125	0,012855	0,000014	0,000011	0,00001
17 MORELOS							
L=11	*NORMALES y casi N.*	0,557163	0,558447	0,557681	0,001476	0,001071	0,000841
n=669	*BAJITOS GORDITOS*	0,250185	0,24913	0,249776	0,000366	0,000279	0,000187
cv= 0.131156	*MAL PARA LA EDAD*	0,144642	0,144827	0,144647	0,000871	0,000658	0,000521
	MAL PARA LA TALLA	0,039005	0,03837	0,038903	0,000088	0,000059	0,00004
	MAL EDAD Y TALLA	0,009004	0,009227	0,008993	0,000024	0,000017	0,000014

* El estimador de Rao se *obtiene ajustando* los factores de expansion; El estimador Wolter contempla la obtencion de pseudovalores y fcpf .
**L= estratos; n= niños enc.; cv = coef var. del denomin. del est. de razon.

TABLA E.2 : ESTIMADOR Y VARIANZA JACKKNIFE (SEGUN RAO Y WOLTER)
JUNTO CON LA ESTIMACION POR FORMULA DEL ESTIMADOR DE RAZON .
POR ESTADO Y A NIVEL NACIONAL

Nota: No se deben hacer inferencias en los estados donde cv > 0.2

ESTADO No. NOMBRE		GRUPO NUTRICIONAL	JACKKNIFE ESTIMADOR RAO*	JACKKNIFE ESTIMADOR Wolter*	Formula/lin. EST. RAZON COMBINADO	JACKKNIFE VAR RAO*	JACKKNIFE VAR Wolter*	Formula/lin. VAR (Rcomb.)
18	NAYARIT							
	L=5	NORMALES y casi N.	0,638913	0,631985	0,637705	0,0039	0,003893	0,003404
	n=207	BAJITOS GORDITOS	0,17982	0,182018	0,180209	0,001537	0,001531	0,001439
	cv=0.148956	MAL PARA LA EDAD	0,074105	0,077963	0,074775	0,000732	0,00073	0,000687
		MAL PARA LA TALLA	0,095601	0,095864	0,095646	0,000926	0,000925	0,000572
		MAL EDAD Y TALLA	0,011561	0,01217	0,011665	0,000066	0,000066	0,000061
19	NUEVO LEON							
	L= 3	NORMALES y casi N.	0,596189	0,598227	0,595749	0,000253	0,000275	0,000281
	n=206	BAJITOS GORDITOS	0,170209	0,169784	0,170615	0,001436	0,001134	0,000837
	cv= 0.232378	MAL PARA LA EDAD	0,076978	0,078747	0,077018	0,000524	0,000516	0,000401
		MAL PARA LA TALLA	0,143814	0,142368	0,143929	0,002626	0,002362	0,00123
		MAL EDAD Y TALLA	0,01281	0,010873	0,012689	0,000071	0,000096	0,000071
20	OAXACA							
	L=30	NORMALES y casi N.	0,404345	0,403969	0,404333	0,000408	0,000408	0,000342
	n=1728	BAJITOS GORDITOS	0,301366	0,300124	0,30132	0,000402	0,000402	0,000384
	cv= 0.090278	MAL PARA LA EDAD	0,211899	0,212213	0,211909	0,00045	0,000449	0,000426
		MAL PARA LA TALLA	0,052892	0,0537	0,052922	0,000082	0,000082	0,000066
		MAL EDAD Y TALLA	0,029498	0,029938	0,029514	0,000025	0,000025	0,000024
21	PUEBLA							
	L=25	NORMALES y casi N.	0,396173	0,39796	0,396262	0,000268	0,000321	0,000373
	n=2413	BAJITOS GORDITOS	0,279911	0,279464	0,279822	0,000263	0,000277	0,00026
	cv= 0.070055	MAL PARA LA EDAD	0,192555	0,194342	0,192632	0,000126	0,000148	0,000139
		MAL PARA LA TALLA	0,084756	0,081842	0,084643	0,000166	0,000184	0,000108
		MAL EDAD Y TALLA	0,046605	0,046392	0,046641	0,000034	0,000046	0,000042

* El estimador de Rao se obtiene ajustando los factores de expansion; El estimador Wolter contempla la obtencion de pseudovalores y fcpf .
**L= estratos; n= niños enc.; cv = coef var. del denomin. del est. de razon.

TABLA E.2 : ESTIMADOR Y VARIANZA JACKKNIFE (SEGUN RAO Y WOLTER)
JUNTO CON LA ESTIMACION POR FORMULA DEL ESTIMADOR DE RAZON .
POR ESTADO Y A NIVEL NACIONAL

Nota: No se deben hacer inferencias en los estados donde cv > 0.2

ESTADO No. NOMBRE	GRUPO NUTRICIONAL	JACKKNIFE ESTIMADOR RAO*	JACKKNIFE ESTIMADOR Wolter*	Formula/lin. EST. RAZON COMBINADO	JACKKNIFE VAR RAO*	JACKKNIFE VAR Wolter*	Formula/lin. VAR (Rcomb.)
22 QUERETARO							
L=9	*NORMALES y casi N.*	0,554111	0,548901	0,553556	0,000722	0,000715	0,000477
n=651	*BAJITOS GORDITOS*	0,227926	0,230142	0,228162	0,000308	0,000306	0,00028
cv= 0.191388	*MAL PARA LA EDAD*	0,154936	0,15951	0,155423	0,000689	0,000684	0,000627
	MAL PARA LA TALLA	0,04878	0,047271	0,048619	0,000137	0,000136	0,000121
	MAL EDAD Y TALLA	0,014248	0,014176	0,01424	0,000031	0,00003	0,000028
23 QUINTANA ROO							
L=3	*NORMALES y casi N.*	0,435305	0,424524	0,433718	0,001361	0,001799	0,001792
n=209	*BAJITOS GORDITOS*	0,344827	0,348771	0,34628	0,001041	0,001334	0,001321
cv= 0.252427	*MAL PARA LA EDAD*	0,149345	0,158029	0,149127	0,001009	0,00128	0,000999
	MAL PARA LA TALLA	0,053025	0,050872	0,053502	0,000693	0,000828	0,000455
	MAL EDAD Y TALLA	0,017497	0,017804	0,017373	0,000048	0,000049	0,000042
24 SAN LUIS POTOSI							
L=11	*NORMALES y casi N.*	0,564217	0,566288	0,564447	0,000926	0,001012	0,001042
n=1192	*BAJITOS GORDITOS*	0,21884	0,221496	0,21888	0,000501	0,000568	0,000519
cv= 0.081463	*MAL PARA LA EDAD*	0,135356	0,133669	0,135317	0,000224	0,000252	0,000231
	MAL PARA LA TALLA	0,059876	0,058183	0,059751	0,000141	0,000166	0,000116
	MAL EDAD Y TALLA	0,021711	0,020385	0,021606	0,000036	0,000039	0,000036
25 SINALOA							
L= 14	*NORMALES y casi N.*	0,721803	0,720171	0,721998	0,000245	0,000294	0,000308
n=742	*BAJITOS GORDITOS*	0,129829	0,128406	0,129389	0,000279	0,00028	0,000263
cv= 0.089434	*MAL PARA LA EDAD*	0,051115	0,051443	0,051274	0,000059	0,000081	0,000074
	MAL PARA LA TALLA	0,081842	0,085146	0,081834	0,000149	0,000183	0,000181
	MAL EDAD Y TALLA	0,015412	0,014851	0,015506	0,000051	0,000093	0,000084

* *El estimador de Rao* se obtiene ajustando los factores de expansion; El estimador Wolter contempla la obtencion de pseudovalores y fcpf .
**L= estratos; n= niños enc.; cv = coef var. del denomin. del est. de razon.

TABLA E.2 : ESTIMADOR Y VARIANZA JACKKNIFE (SEGUN RAO Y WOLTER)
JUNTO CON LA ESTIMACION POR FORMULA DEL ESTIMADOR DE RAZON .
POR ESTADO Y A NIVEL NACIONAL

Nota: No se deben hacer inferencias en los estados donde cv > 0.2

ESTADO No. NOMBRE	GRUPO NUTRICIONAL	JACKKNIFE ESTIMADOR RAO*	JACKKNIFE ESTIMADOR Wolter*	Formula/lin. EST. RAZON COMBINADO	JACKKNIFE VAR RAO*	JACKKNIFE VAR Wolter*	Formula/lin. VAR (Rcomb.)
26 SONORA							
L=6	NORMALES y casi N.	0,768715	0,767871	0,768588	0,000965	0,000961	0,00097
n=243	BAJITOS GORDITOS	0,133197	0,134301	0,13336	0,000376	0,000375	0,000359
cv= 0.187546	MAL PARA LA EDAD	0,024944	0,02439	0,024864	0,000309	0,000304	0,000281
	MAL PARA LA TALLA	0,068712	0,06914	0,068776	0,000202	0,000202	0,000382
	MAL EDAD Y TALLA	0,004431	0,004299	0,004412	0,000012	0,000012	0,000011
27 TABASCO							
L=17	NORMALES y casi N.	0,58076	0,587434	0,581191	0,000287	0,000287	0,000454
n=1367	BAJITOS GORDITOS	0,21323	0,213311	0,213235	0,000307	0,000307	0,000292
cv=0.061831	MAL PARA LA EDAD	0,117906	0,115811	0,11777	0,000047	0,000047	0,000043
	MAL PARA LA TALLA	0,072502	0,068741	0,072259	0,00007	0,00007	0,000067
	MAL EDAD Y TALLA	0,015603	0,014688	0,015544	0,000042	0,000042	0,000039
28 TAMAULIPAS							
L=6	NORMALES y casi N.	0,64348	0,645035	0,644342	0,000541	0,000705	0,000585
n=332	BAJITOS GORDITOS	0,185575	0,188252	0,185757	0,001239	0,001209	0,001083
cv= 0.116067	MAL PARA LA EDAD	0,04626	0,045959	0,046298	0,000284	0,000238	0,00022
	MAL PARA LA TALLA	0,11882	0,11504	0,117829	0,001513	0,001161	0,000613
	MAL EDAD Y TALLA	0,005866	0,005706	0,005773	0,000023	0,000019	0,000017
29 TLAXCALA							
L= 4	NORMALES y casi N.	0,530572	0,530606	0,529848	0,001961	0,002193	0,002201
n=380	BAJITOS GORDITOS	0,231587	0,229962	0,231415	0,000561	0,000549	0,000492
cv= 0.218980	MAL PARA LA EDAD	0,136025	0,135329	0,136753	0,000482	0,000643	0,000592
	MAL PARA LA TALLA	0,071956	0,074119	0,072348	0,000332	0,000326	0,000208
	MAL EDAD Y TALLA	0,02986	0,029977	0,029635	0,000052	0,000057	0,000053

* El estimador de Rao se obtiene ajustando los factores de expansion; El estimador Wolter contempla la obtencion de pseudovalores y fcpf .
**L= estratos; n= niños enc.; cv = coef var. del denomin. del est. de razon.

TABLA E.2 : ESTIMADOR Y VARIANZA JACKKNIFE (SEGUN RAO Y WOLTER)
JUNTO CON LA ESTIMACION POR FORMULA DEL ESTIMADOR DE RAZON .
POR ESTADO Y A NIVEL NACIONAL

Nota: No se deben hacer inferencias en los estados donde cv > 0.2

ESTADO		GRUPO	JACKKNIFE		Formula/lin.	JACKKNIFE		Formula/lin.
No.	NOMBRE	NUTRICIONAL	ESTIMADOR RAO*	ESTIMADOR Wolter*	EST. RAZON COMBINADO	VAR RAO*	VAR Wolter*	VAR (Rcomb.)
30	VERACRUZ							
	L=36	*NORMALES y casi N.*	0,532622	0,531428	0,532529	0,000393	0,000399	0,000406
	n=2635	*BAJITOS GORDITOS*	0,237758	0,23897	0,237826	0,00018	0,000187	0,000177
	cv= 0.066841	*MAL PARA LA EDAD*	0,14527	0,143488	0,145241	0,000128	0,000131	0,000123
		MAL PARA LA TALLA	0,062052	0,06361	0,062106	0,000042	0,000043	0,000055
		MAL EDAD Y TALLA	0,022298	0,022543	0,022299	0,000013	0,000013	0,000012
31	YUCATAN							
	L=6	*NORMALES y casi N.*	0,329176	0,33087	0,328716	0,000804	0,001472	0,001185
	n=674	*BAJITOS GORDITOS*	0,309285	0,305806	0,308406	0,000197	0,000272	0,000226
	cv=0.116179	*MAL PARA LA EDAD*	0,222732	0,224314	0,223876	0,001049	0,001988	0,001672
		MAL PARA LA TALLA	0,088258	0,088089	0,088606	0,00024	0,000429	0,00027
		MAL EDAD Y TALLA	0,050549	0,050932	0,050397	0,0001	0,000126	0,000107
32	ZACATECAS							
	L=9	*NORMALES y casi N.*	0,637442	0,639019	0,637893	0,00043	0,000446	0,000494
	n=746	*BAJITOS GORDITOS*	0,170037	0,168052	0,169942	0,000419	0,000422	0,000396
	cv= 0.147150	*MAL PARA LA EDAD*	0,100353	0,101772	0,100258	0,000603	0,000603	0,000567
		MAL PARA LA TALLA	0,084947	0,083953	0,084719	0,000144	0,000145	0,000125
		MAL EDAD Y TALLA	0,00722	0,007192	0,007187	0,000023	0,000023	0,000022

		Estimador de razon SEPARADO			*Est. De Varianzadel est. SEPARADO*		
		JACKKNIFE		Formula /	JACKKNIFE		Formula /
NACIONAL		RAO	Wolter	Lineariz.	RAO	Wolter	Lineariz.
	NORMALES y casi N.	0,526664874	0,53185181	0,52695888	2,7375E-05	2,7091E-05	2,7574E-05
	BAJITOS GORDITOS	0,238469191	0,23711539	0,2383906	1,84168E-05	1,8197E-05	1,7063E-05
	MAL PARA LA EDAD	0,140498287	0,14025239	0,14050901	1,3399E-05	1,3333E-05	1,2408E-05
	MAL PARA LA TALLA	0,070572584	0,06885175	0,07046621	7,41629E-06	7,353E-06	5,5937E-06
	MAL EDAD Y TALLA	0,023795139	0,02200297	0,0236753	1,81219E-06	1,858E-06	1,7234E-06

* El estimador de Rao se obtiene ajustando los factores de expansion; El estimador Wolter contempla la obtencion de pseudovalores y fcpf .

**L= estratos; n= niños enc.; cv = coef var. del denomin. del est. de razon.

TABLA E.3: INTERVALOS DE CONFIANZA BASADOS EN LA VARIANZA JACKKNIFE
PARA LOS ESTIMADORES DE RAZON, CONSIDERANDO LA DISTRIBUCION T CON (n - L) GRADOS DE LIBERTAD Y LA NORMAL

ESTADO		GRUPO	G. L.	ESTIMADOR	VAR	I.C. Basado en (n-L) g.l.		I. C. Basado en N(0,1)	
NO.	NOMBRE	NUTRICIONAL	(n - L)	JACKKNIFE	JACKKNIFE	Lim. Inf.	Lim. Sup.	Lim. Inf.	Lim. Sup.
1	**AGUASCALIENTES**								
	L=5 *	*NORMALES y casi N.*	5	0,671309	0,000632	0,606686	0,735932	0,622035	0,720583
	n= 385 *	*BAJITOS GORDITOS*	5	0,150033	0,000327	0,103549	0,196517	0,114590	0,185476
	cv=0.068466 *	*MAL PARA LA EDAD*	5	0,091889	0,000564	0,030841	0,152937	0,045342	0,138436
		MAL PARA LA TALLA	5	0,071746	0,000341	0,024277	0,119215	0,035552	0,107940
		MAL EDAD Y TALLA	5	0,015022	0,000063	0,000000	0,035425	0,000000	0,030579
2	**BAJA CALIFORNIA**								
	L=2	*NORMALES y casi N,*	4	0,824199	0,000879	0,741883	0,906515	0,766089	0,882309
	n=164	*BAJITOS GORDITOS*	4	0,045203	0,001894	0,000000	0,166034	0,000000	0,130502
	cv=0.144709	*MAL PARA LA EDAD*	4	0,043075	0,000008	0,035222	0,050928	0,037531	0,048619
		MAL PARA LA TALLA	4	0,085505	0,005723	0,000000	0,295545	0,000000	0,233780
		MAL EDAD Y TALLA	4	0,002018	0,000002	0,000000	0,005944	0,000000	0,004790
3	**B. C. SUR**								
	L= 4	*NORMALES y casi N,*	7	0,638132	0,001064	0,561000	0,715264	0,574199	0,702065
	n=90	*BAJITOS GORDITOS*	7	0,137942	0,000767	0,072454	0,203430	0,083660	0,192224
	cv=0.292971	*MAL PARA LA EDAD*	7	0,123643	0,005742	0,000000	0,302825	0,000000	0,272164
		MAL PARA LA TALLA	7	0,056410	0,000583	0,000000	0,113505	0,009085	0,103735
		MAL EDAD Y TALLA	7	0,043872	0,001878	0,000000	0,146345	0,000000	0,128810
4	**CAMPECHE**								
	L=8	*NORMALES y casi N,*	16	0,437550	0,001066	0,368336	0,506764	0,373557	0,501543
	n=1024	*BAJITOS GORDITOS*	16	0,312537	0,000344	0,273219	0,351855	0,276184	0,348890
	cv= 0.104915	*MAL PARA LA EDAD*	16	0,198832	0,000424	0,155180	0,242484	0,158473	0,239191
		MAL PARA LA TALLA	16	0,030358	0,000021	0,020643	0,040073	0,021376	0,039340
		MAL EDAD Y TALLA	16	0,020723	0,000030	0,009112	0,032334	0,009988	0,031458
5	**COAHUILA**								
	L=5	*NORMALES y casi N.*	10	0,686418	0,001912	0,588989	0,783847	0,600714	0,772122
	n=409	*BAJITOS GORDITOS*	10	0,135864	0,000331	0,095327	0,176401	0,100205	0,171523
	cv=0.183529	*MAL PARA LA EDAD*	10	0,101482	0,001287	0,021548	0,181416	0,031167	0,171797
		MAL PARA LA TALLA	10	0,061643	0,000273	0,024828	0,098458	0,029258	0,094028
		MAL EDAD Y TALLA	10	0,014594	0,000097	0,000000	0,036539	0,000000	0,033898

* L=no. estratos; n= total de niños encuestados en el estado; cv= coef. de variacion del denominador del est. de razon.

Nota: **No se deben hacer inferencias en los estados donde cv >0.2.

TABLA E.3: INTERVALOS DE CONFIANZA BASADOS EN LA VARIANZA JACKKNIFE
PARA LOS ESTIMADORES DE RAZON, CONSIDERANDO LA DISTRIBUCION T CON (n - L) GRADOS DE LIBERTAD Y LA NORMAL

ESTADO NO. NOMBRE	GRUPO NUTRICIONAL	G. L. (n - L)	ESTIMADOR JACKKNIFE	VAR JACKKNIFE	I.C. Basado en (n-L) g.l. Lim. Inf.	I.C. Basado en (n-L) g.l. Lim. Sup.	I. C. Basado en N(0,1) Lim. Inf.	I. C. Basado en N(0,1) Lim. Sup.
6 COLIMA								
L=4	*NORMALES y casi N,*	*8*	0,509337	0,001917	0,408372	0,610302	0,423521	0,595153
n=247	*BAJITOS GORDITOS*	*8*	0,219793	0,001934	0,118381	0,321205	0,133598	0,305988
cv= *0.197409*	*MAL PARA LA EDAD*	*8*	0,057743	0,000265	0,020204	0,095282	0,025837	0,089649
	MAL PARA LA TALLA	*8*	0,187915	0,002441	0,073983	0,301847	0,091078	0,284752
	MAL EDAD Y TALLA	*8*	0,025212	0,000041	0,010446	0,039978	0,012662	0,037762
7 CHIAPAS								
L= 25	*NORMALES y casi N,*	*32*	0,366850	0,000315	0,330698	0,403002	0,332063	0,401637
n=1897	*BAJITOS GORDITOS*	*32*	0,361085	0,000272	0,327491	0,394679	0,328760	0,393410
cv= 0.083302	*MAL PARA LA EDAD*	*32*	0,185102	0,000234	0,153943	0,216261	0,155120	0,215084
	MAL PARA LA TALLA	*32*	0,063848	0,000066	0,047300	0,080396	0,047925	0,079771
	MAL EDAD Y TALLA	*32*	0,023116	0,000045	0,009452	0,036780	0,009968	0,036264
8 CHIHUAHUA								
L=4	*NORMALES y casi N,*	*5*	0,724303	0,000352	0,676075	0,772531	0,687530	0,761076
n=250	*BAJITOS GORDITOS*	*5*	0,121658	0,001358	0,026930	0,216386	0,049430	0,193886
cv=0.333081	*MAL PARA LA EDAD*	*5*	0,073829	0,000356	0,025327	0,122331	0,036848	0,110810
	MAL PARA LA TALLA	*5*	0,080210	0,000809	0,007095	0,153325	0,024462	0,135958
	MAL EDAD Y TALLA	*5*	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000
10 DURANGO								
L=8	*NORMALES y casi N,*	*13*	0,674122	0,000600	0,621204	0,727040	0,626112	0,722132
n=581	*BAJITOS GORDITOS*	*13*	0,182677	0,000215	0,151000	0,214354	0,153938	0,211416
cv=0.123927	*MAL PARA LA EDAD*	*13*	0,055026	0,000167	0,027108	0,082944	0,029697	0,080355
	MAL PARA LA TALLA	*13*	0,082541	0,000438	0,037328	0,127754	0,041521	0,123561
	MAL EDAD Y TALLA	*13*	0,005633	0,000004	0,001312	0,009954	0,001713	0,009553
11 GUANAJUATO								
L=19	*NORMALES y casi N,*	*19*	0,586051	0,000539	0,537459	0,634643	0,540547	0,631555
n=845	*BAJITOS GORDITOS*	*19*	0,180556	0,000516	0,133012	0,228100	0,136033	0,225079
cv=0.106484	*MAL PARA LA EDAD*	*19*	0,105145	0,000239	0,072788	0,137502	0,074844	0,135446
	MAL PARA LA TALLA	*19*	0,092979	0,000092	0,072903	0,113055	0,074179	0,111779
	MAL EDAD Y TALLA	*19*	0,035269	0,000059	0,019192	0,051346	0,020214	0,050324

* L=no. estratos; n= total de niños encuestados en el estado; cv= coef. de variacion del denominador del est. de razon.

Nota: **No se deben hacer inferencias en los estados donde cv >0.2.

TABLA E.3: INTERVALOS DE CONFIANZA BASADOS EN LA VARIANZA JACKKNIFE
PARA LOS ESTIMADORES DE RAZON, CONSIDERANDO LA DISTRIBUCION T CON (n - L) GRADOS DE LIBERTAD Y LA NORMAL

ESTADO NO. NOMBRE		GRUPO NUTRICIONAL	G. L. (n - L)	ESTIMADOR JACKKNIFE	VAR JACKKNIFE	I.C. Basado en (n-L) g.l. Lim. Inf.	Lim. Sup.	I. C. Basado en N(0,1) Lim. Inf.	Lim. Sup.
12	**GUERRERO**								
	L=20	*NORMALES y casi N.*	34	0,391806	0,000451	0,348648	0,434964	0,350182	0,433430
	n=1887	*BAJITOS GORDITOS*	34	0,230392	0,000157	0,204928	0,255856	0,205833	0,254951
	cv=0.066246	*MAL PARA LA EDAD*	34	0,257352	0,000442	0,214627	0,300077	0,216145	0,298559
		MAL PARA LA TALLA	34	0,058570	0,000033	0,046896	0,070244	0,047311	0,069829
		MAL EDAD Y TALLA	34	0,061881	0,000065	0,045497	0,078265	0,046079	0,077683
13	**HIDALGO**								
	L=15	*NORMALES y casi N.*	17	0,455319	0,000670	0,400708	0,509930	0,404586	0,506052
	n=997	*BAJITOS GORDITOS*	17	0,288198	0,000574	0,237650	0,338746	0,241240	0,335156
	cv= 0.105027	*MAL PARA LA EDAD*	17	0,156622	0,000286	0,120942	0,192302	0,123475	0,189769
		MAL PARA LA TALLA	17	0,065156	0,000240	0,032471	0,097841	0,034792	0,095520
		MAL EDAD Y TALLA	17	0,034705	0,000106	0,012983	0,056427	0,014526	0,054884
14	**JALISCO**								
	L=12	*NORMALES y casi N.*	23	0,717439	0,000256	0,684341	0,750537	0,686079	0,748799
	n=886	*BAJITOS GORDITOS*	23	0,138288	0,000136	0,114164	0,162412	0,115431	0,161145
	cv=0.090785	*MAL PARA LA EDAD*	23	0,040024	0,000037	0,027441	0,052607	0,028102	0,051946
		MAL PARA LA TALLA	23	0,096240	0,000054	0,081039	0,111441	0,081837	0,110643
		MAL EDAD Y TALLA	23	0,008009	0,000008	0,002158	0,013860	0,002465	0,013553
15	**MEXICO**								
	L=27	*NORMALES y casi N.*	27	0,533731	0,000553	0,485480	0,581982	0,487640	0,579822
	n=1209	*BAJITOS GORDITOS*	27	0,268386	0,000414	0,226637	0,310135	0,228506	0,308266
	cv=0.097556	*MAL PARA LA EDAD*	27	0,136110	0,000117	0,113916	0,158304	0,114909	0,157311
		MAL PARA LA TALLA	27	0,054748	0,000148	0,029786	0,079710	0,030904	0,078592
		MAL EDAD Y TALLA	27	0,007024	0,000007	0,001595	0,012453	0,001838	0,012210
16	**MICHOACAN**								
	L=19	*NORMALES y casi N.*	21	0,585021	0,000822	0,525397	0,644645	0,528827	0,641215
	n=1435	*BAJITOS GORDITOS*	21	0,241942	0,000227	0,210609	0,273275	0,212412	0,271472
	cv=0.131001	*MAL PARA LA EDAD*	21	0,090876	0,000218	0,060171	0,121581	0,061937	0,119815
		MAL PARA LA TALLA	21	0,069284	0,000270	0,035112	0,103456	0,037078	0,101490
		MAL EDAD Y TALLA	21	0,012877	0,000014	0,005096	0,020658	0,005543	0,020211

* L=no. estratos; n= total de niños encuestados en el estado; cv= coef. de variacion del denominador del est. de razon.

Nota: **No se deben hacer inferencias en los estados donde cv >0.2.

TABLA E.3: INTERVALOS DE CONFIANZA BASADOS EN LA VARIANZA JACKKNIFE
PARA LOS ESTIMADORES DE RAZON, CONSIDERANDO LA DISTRIBUCION T CON (n - L) GRADOS DE LIBERTAD Y LA NORMAL

ESTADO NO. NOMBRE	GRUPO NUTRICIONAL	G. L. (n - L)	ESTIMADOR JACKKNIFE	VAR JACKKNIFE	I.C. Basado en (n-L) g.l. Lim. Inf.	I.C. Basado en (n-L) g.l. Lim. Sup.	I. C. Basado en N(0,1) Lim. Inf.	I. C. Basado en N(0,1) Lim. Sup.
17 MORELOS								
L=11	*NORMALES y casi N.*	14	0,557163	0,001476	0,474763	0,639563	0,481862	0,632464
n=669	*BAJITOS GORDITOS*	14	0,250185	0,000366	0,209153	0,291217	0,212688	0,287682
cv= 0.131155	*MAL PARA LA EDAD*	14	0,144642	0,000871	0,081343	0,207941	0,086797	0,202487
	MAL PARA LA TALLA	14	0,039005	0,000088	0,018885	0,059125	0,020619	0,057391
	MAL EDAD Y TALLA	14	0,009004	0,000024	0,000000	0,019511	0,000000	0,018606
18 NAYARIT								
L=5	*NORMALES y casi N.*	5	0,638913	0,003900	0,478380	0,799446	0,516511	0,761315
n=207	*BAJITOS GORDITOS*	5	0,179820	0,001537	0,079042	0,280598	0,102979	0,256661
cv=0.148956	*MAL PARA LA EDAD*	5	0,074105	0,000732	0,004557	0,143653	0,021076	0,127134
	MAL PARA LA TALLA	5	0,095601	0,000926	0,017378	0,173824	0,035958	0,155244
	MAL EDAD Y TALLA	5	0,011561	0,000066	0,000000	0,032444	0,000000	0,027484
19 NUEVO LEON								
L= 3	*NORMALES y casi N.*	6	0,596189	0,000253	0,557268	0,635110	0,565013	0,627365
n=206	*BAJITOS GORDITOS*	6	0,170209	0,001436	0,077484	0,262934	0,095936	0,244482
cv= 0.232378	*MAL PARA LA EDAD*	6	0,076978	0,000524	0,020966	0,132990	0,032112	0,121844
	MAL PARA LA TALLA	6	0,143814	0,002626	0,018423	0,269205	0,043375	0,244253
	MAL EDAD Y TALLA	6	0,012810	0,000071	0,000000	0,033428	0,000000	0,029325
20 OAXACA								
L=30	*NORMALES y casi N.*	30	0,404345	0,000408	0,363093	0,445597	0,364755	0,443935
n=1728	*BAJITOS GORDITOS*	30	0,301366	0,000402	0,260419	0,342313	0,262068	0,340664
cv= 0.090278	*MAL PÀRA LA EDAD*	30	0,211899	0,000450	0,168576	0,255222	0,170321	0,253477
	MAL PARA LA TALLA	30	0,052892	0,000082	0,034398	0,071386	0,035143	0,070641
	MAL EDAD Y TALLA	30	0,029498	0,000025	0,019287	0,039709	0,019698	0,039298
21 PUEBLA								
L=25	*NORMALES y casi N.*	34	0,396173	0,000268	0,362904	0,429442	0,364086	0,428260
n=2413	*BAJITOS GORDITOS*	34	0,279911	0,000263	0,246954	0,312868	0,248125	0,311697
cv= 0.070055	*MAL PARA LA EDAD*	34	0,192555	0,000126	0,169743	0,215367	0,170554	0,214556
	MAL PARA LA TALLA	34	0,084756	0,000166	0,058572	0,110940	0,059503	0,110009
	MAL EDAD Y TALLA	34	0,046605	0,000034	0,034755	0,058455	0,035176	0,058034

* L=no. estratos; n= total de niños encuestados en el estado; cv= coef. de variacion del denominador del est. de razon.
**Nota: No se deben hacer inferencias en los estados donde cv >0.2.

TABLA E.3: INTERVALOS DE CONFIANZA BASADOS EN LA VARIANZA JACKKNIFE
PARA LOS ESTIMADORES DE RAZON, CONSIDERANDO LA DISTRIBUCION T CON (n - L) GRADOS DE LIBERTAD Y LA NORMAL

ESTADO NO. NOMBRE	GRUPO NUTRICIONAL	G. L. (n - L)	ESTIMADOR JACKKNIFE	VAR JACKKNIFE	I.C. Basado en (n-L) g.l. Lim. Inf.	Lim. Sup.	I. C. Basado en N(0,1) Lim. Inf.	Lim. Sup.
22 QUERETARO								
L=9	*NORMALES y casi N,*	9	0,554111	0,000722	0.493327	0,614895	0,501446	0,606776
n=651	*BAJITOS GORDITOS*	9	0,227926	0,000308	0,188225	0,267627	0,193528	0,262324
cv= 0.191388	*MAL PARA LA EDAD*	9	0,154936	0,000689	0,095557	0,214315	0,103488	0,206384
	MAL PARA LA TALLA	9	0,048780	0,000137	0,022302	0,075258	0,025839	0,071721
	MAL EDAD Y TALLA	9	0,014248	0,000031	0,001653	0,026843	0,003335	0,025161
23 QUINTANA ROO								
L=3	*NORMALES y casi N,*	6	0,435305	0,001361	0,345034	0,525576	0,362997	0,507613
n=209	*BAJITOS GORDITOS*	6	0,344827	0,001041	0,265878	0,423776	0,281589	0,408065
cv= 0.252427	*MAL PARA LA EDAD*	6	0,149345	0,001009	0,071619	0,227071	0,087086	0,211604
	MAL PARA LA TALLA	6	0,053025	0,000693	0,000000	0,117440	0,001428	0,104622
	MAL EDAD Y TALLA	6	0,017497	0,000048	0,000544	0,034450	0,003918	0,031076
24 SAN LUIS POTOSI								
L=11	*NORMALES y casi N,*	21	0,564217	0,000926	0,500934	0,627500	0,504574	0,623860
n=1192	*BAJITOS GORDITOS*	21	0,218840	0,000501	0,172292	0,265388	0,174969	0,262711
cv= 0.081463	*MAL PARA LA EDAD*	21	0,135356	0,000224	0,104231	0,166481	0,106021	0,164691
	MAL PARA LA TALLA	21	0,059876	0,000141	0,035182	0,084570	0,036602	0,083150
	MAL EDAD Y TALLA	21	0,021711	0,000036	0,009233	0,034189	0,009951	0,033471
25 SINALOA								
L= 14	*NORMALES y casi N,*	19	0,721803	0,000245	0,689042	0,754564	0,691124	0,752482
n=742	*BAJITOS GORDITOS*	19	0,129829	0,000279	0,094869	0,164789	0,097091	0,162567
cv= 0.089434	*MAL PARA LA EDAD*	19	0,051115	0,000059	0,035038	0,067192	0,036060	0,066170
	MAL PARA LA TALLA	19	0,081842	0,000149	0,056293	0,107391	0,057917	0,105767
	MAL EDAD Y TALLA	19	0,015412	0,000051	0,000465	0,030359	0,001415	0,029409
26 SONORA								
L=6	*NORMALES y casi N,*	6	0,768715	0,000965	0,692703	0,844727	0,707829	0,829601
n=243	*BAJITOS GORDITOS*	6	0,133197	0,000376	0,085750	0,180644	0,095191	0,171203
cv= 0.187546	*MAL PARA LA EDAD*	6	0,024944	0,000309	0,000000	0,067957	0,000000	0,059398
	MAL PARA LA TALLA	6	0,068712	0,000202	0,033935	0,103489	0,040855	0,096569
	MAL EDAD Y TALLA	6	0,004431	0,000012	0,000000	0,012907	0,000000	0,011221

* L=no. estratos; n= total de niños encuestados en el estado; cv= coef. de variacion del denominador del est. de razon.
Nota: **No se deben hacer inferencias en los estados donde cv >0.2.

TABLA E.3: INTERVALOS DE CONFIANZA BASADOS EN LA VARIANZA JACKKNIFE
PARA LOS ESTIMADORES DE RAZON, CONSIDERANDO LA DISTRIBUCION T CON (n - L) GRADOS DE LIBERTAD Y LA NORMAL

ESTADO NO. NOMBRE	GRUPO NUTRICIONAL	G. L. (n - L)	ESTIMADOR JACKKNIFE	VAR JACKKNIFE	I.C. Basado en (n-L) g.l. Lim. Inf.	Lim. Sup.	I. C. Basado en N(0,1) Lim. Inf.	Lim. Sup.
27 TABASCO								
L=17	*NORMALES y casi N.*	17	0,580760	0,000287	0,545017	0,616503	0,547555	0,613965
n=1367	*BAJITOS GORDITOS*	17	0,213230	0,000307	0,176263	0,250197	0,178888	0,247572
cv=0.061831	*MAL PARA LA EDAD*	17	0,117906	0,000047	0,103442	0,132370	0,104469	0,131343
	MAL PARA LA TALLA	17	0,072502	0,000070	0,054850	0,090154	0,056103	0,088901
	MAL EDAD Y TALLA	17	0,015603	0,000042	0,001930	0,029276	0,002901	0,028305
28 TAMAULIPAS								
L=5	*NORMALES y casi N.*	7	0,643480	0,000541	0,588480	0,698480	0,597892	0,689068
n=332	*BAJITOS GORDITOS*	7	0,185575	0,001239	0,102342	0,268808	0,116584	0,254566
cv= 0.116067	*MAL PARA LA EDAD*	7	0,046260	0,000284	0,006411	0,086109	0,013229	0,079291
	MAL PARA LA TALLA	7	0,118820	0,001513	0,026843	0,210797	0,042581	0,195059
	MAL EDAD Y TALLA	7	0,005866	0,000023	0,000000	0,017206	0,000000	0,015266
29 TLAXCALA								
L= 4	*NORMALES y casi N.*	5	0,530572	0,001961	0,416739	0,644405	0,443777	0,617367
n=380	*BAJITOS GORDITOS*	5	0,231587	0,000561	0,170702	0,292472	0,185164	0,278010
cv= 0.218980	*MAL PARA LA EDAD*	5	0,136025	0,000482	0,079589	0,192461	0,092994	0,179056
	MAL PARA LA TALLA	5	0,071956	0,000332	0,025118	0,118794	0,036243	0,107669
	MAL EDAD Y TALLA	5	0,029860	0,000052	0,011323	0,048397	0,015726	0,043994
30 VERACRUZ								
L=36	*NORMALES y casi N.*	41	0,532622	0,000393	0,492586	0,572658	0,493767	0,571477
n=2635	*BAJITOS GORDITOS*	41	0,237758	0,000180	0,210663	0,264853	0,211462	0,264054
cv= 0.066841	*MAL PARA LA EDAD*	41	0,145270	0,000128	0,122421	0,168119	0,123095	0,167445
	MAL PARA LA TALLA	41	0,062052	0,000042	0,048964	0,075140	0,049350	0,074754
	MAL EDAD Y TALLA	41	0,022298	0,000013	0,015016	0,029580	0,015231	0,029365
31 YUCATAN								
L=6	*NORMALES y casi N.*	11	0,329176	0,000804	0,266767	0,391585	0,273600	0,384752
n=674	*BAJITOS GORDITOS*	11	0,309285	0,000197	0,278393	0,340177	0,281775	0,336795
cv=0.116179	*MAL PARA LA EDAD*	11	0,222732	0,001049	0,151446	0,294018	0,159251	0,286213
	MAL PARA LA TALLA	11	0,088258	0,000240	0,054160	0,122356	0,057894	0,118622
	MAL EDAD Y TALLA	11	0,050549	0,000100	0,028539	0,072559	0,030949	0,070149

* L=no. estratos; n= total de niños encuestados en el estado; cv= coef. de variacion del denominador del est. de razon.
Nota: **No se deben hacer inferencias en los estados donde cv >0.2.

TABLA E.3: INTERVALOS DE CONFIANZA BASADOS EN LA VARIANZA JACKKNIFE
PARA LOS ESTIMADORES DE RAZON, CONSIDERANDO LA DISTRIBUCION T CON (n - L) GRADOS DE LIBERTAD Y LA NORMAL

ESTADO		GRUPO	G. L.	ESTIMADOR	VAR	I.C. Basado en (n-L) g.l.		I. C. Basado en N(0,1)	
NO.	NOMBRE	NUTRICIONAL	(n - L)	JACKKNIFE	JACKKNIFE	Lim. Inf.	Lim. Sup.	Lim. Inf.	Lim. Sup.
32	**ZACATECAS**								
	L=9	*NORMALES y casi N.*	*11*	0,637442	0,000430	0,591801	0,683083	0,596799	0,678085
	n=746	*BAJITOS GORDITOS*	*11*	0,170037	0,000419	0,124984	0,215090	0,129917	0,210157
	cv= 0.147150	*MAL PARA LA EDAD*	*11*	0,100353	0,000603	0,046305	0,154401	0,052223	0,148483
		MAL PARA LA TALLA	*11*	0,084947	0,000144	0,058535	0,111359	0,061427	0,108467
		MAL EDAD Y TALLA	*11*	0,007220	0,000023	0,000000	0,017776	0,000000	0,016620

* L=no. estratos; n= total de niños encuestados en el estado; cv= coef. de variacion del denominador del est. de razon.
Nota: **No se deben hacer inferencias en los estados donde cv >0.2.

```
/* PROGRAMA QUE CALCULA LAS VARIANZAS POR FORMULA */

/* este programa se llama: varnewco.sps  */
get  file='c:\enal96\modulo7\porcasac.sav'.

/* do if (chiquis =0 ).   */
/* compute chiquis = nmenor. */
/* else.  */
/* end if.  */
/*  save outfile='c:\enal96\varjul\porcasac.sav'.   */

SAVE OUTFILE='C:\ENAL96\varjul\compvar.Sav'
   /COMPRESSED.

COMPUTE FCASA=CHIQUIS/NMENOR.
COMPUTE DES0=FCASA* DES0_1.
COMPUTE DES1=FCASA * DES1_1.
COMPUTE DES2=FCASA * DES2_1.
COMPUTE DES3=FCASA * DES3_1.
COMPUTE DES4=FCASA * DES4_1.
COMPUTE DES5=FCASA * DES5_1.
COMPUTE DES6=FCASA * DES6_1.
COMPUTE DES7= FCASA* DES7_1.
save outfile='c:\enal96\varjul\compvar.sav'.
AGGREGATE
  /OUTFILE='C:\ENAL96\varjul\MENORXC.SAV'
  /BREAK=claveedo estrato clavemun claveloc
  /des0_1_1 = SUM(des0_1) /des1_1_1 = SUM(des1_1) /des2_1_1 = SUM(des2_1)
 /des3_1_1 = SUM(des3_1) /des4_1_1 - SUM(des4_1) /des5_1_1 = SUM(des5_1)
 /des6_1_1 = SUM(des6_1) /des7_1_1 = SUM(des7_1)
    /D0 D1 D2 D3 D4 D5 D6 D7 =
    SUM(DES0 DES1 DES2 DES3 DES4 DES5 DES6 DES7 )
  /vivloc= first(vivloc)
  /menor = first(menor)
  /UNOS= SUM(NMENOR)
    /NUMENCU = FIRST(NUMENCU)
  /locestr LOCXESTR =first(locestr  LOCXESTR)
/chicoloc -sum(chiquis)   /VIVIEN =FIRST(VIVIEN).

/*  AHORA SE VAN A CALCULAR LOS ESTIMADORES DE RAZON    */

GET FILE ='C:\ENAL96\varjul\MENORXC.SAV'.
COMPUTE FVIV = VIVIEN/NUMENCU.
COMPUTE FLOC =LOCXESTR/LOCESTR.
COMPUTE X0LOC=  DES0_1_1 * FVIV.
COMPUTE X1LOC = DES1_1_1 * FVIV.
COMPUTE X2LOC = DES2_1_1 * FVIV.
COMPUTE X3LOC = DES3_1_1 * FVIV.
COMPUTE X4LOC=  DES4_1_1 * FVIV.
COMPUTE X5LOC = DES5_1_1 * FVIV.
COMPUTE X6LOC = DES6_1_1 * FVIV.
COMPUTE X7LOC = DES7_1_1 * FVIV.
COMPUTE YLOC = CHICOLOC * FVIV.
COMPUTE YUNOS =UNOS * FVIV.
COMPUTE  M0 = D0 * FVIV.
COMPUTE M1 = D1 * FVIV.
COMPUTE M2 = D2 * FVIV.
COMPUTE M3 = D3 * FVIV.
COMPUTE M4 = D4 * FVIV.
COMPUTE M5 = D5 * FVIV.
COMPUTE M6 = D6*FVIV.
COMPUTE M7 = D7 * FVIV.
AGGREGATE
    OUTFILE='C:\ENAL96\varjul\PASO1.SAV'
    /BREAK CLAVEEDO ESTRATO
    /X0 X1 X2 X3 X4 X5 X6 X7 Y YU=
    SUM(X0LOC X1LOC X2LOC X3LOC X4LOC X5LOC
    X6LOC X7LOC  YLOC YUNOS)
    /M0 M1 M2 M3 M4 M5 M6 M7 =
    SUM(M0 M1 M2 M3 M4 M5 M6 M7)
    /  FLOC = FIRST(FLOC).
GET FILE='C:\ENAL96\varjul\PASO1.SAV'.
COMPUTE HUNOS =FLOC * YU.
```

```
COMPUTE HY = FLOC* Y.
COMPUTE H0=FLOC* X0.
COMPUTE H1 =FLOC*X1.
COMPUTE H2=FLOC*X2.
COMPUTE H3=FLOC* X3.
COMPUTE H4 =FLOC*X4.
COMPUTE H5=FLOC*X5.
COMPUTE H6 =FLOC*X6.
COMPUTE H7=FLOC*X7.
COMPUTE H20=FLOC*M0.
COMPUTE H21=FLOC*M1.
COMPUTE H22=FLOC*M2.
COMPUTE H23 =FLOC*M3.
COMPUTE H24 = FLOC*M4.
COMPUTE H25=FLOC* M5.
COMPUTE H26 = FLOC*M6.
COMPUTE H27=FLOC* M7.
AGGREGATE
    /OUTFILE='C:\ENAL96\varjul\PASO2.SAV'
    /BREAK = CLAVEEDO
    /TIPO0 TIPO1 TIPO2 TIPO3 TIPO4
TIPO5 TIPO6 TIPO7 MENORES5 TUNOS = SUM (H0 H1 H2
H3 H4 H5 H6 H7 HY HUNOS)
    /M2T0 M2T1 M2T2 M2T3 M2T4 M2T5 M2T6 M2T7 =
    SUM(H20 H21 H22 H23 H24 H25 H26 H27).
GET FILE='C:\ENAL96\varjul\PASO2.SAV'.
COMPUTE RAZON_0 = TIPO0/TUNOS.
COMPUTE RAZON_1 = TIPO1/TUNOS.
COMPUTE RAZON_2 = TIPO2/TUNOS.
COMPUTE RAZON_3 = TIPO3/TUNOS.
COMPUTE RAZON_4 = TIPO4/TUNOS.
COMPUTE RAZON_5 = TIPO5/TUNOS.
COMPUTE RAZON_6 = TIPO6/TUNOS.
COMPUTE RAZON_7 = TIPO7/TUNOS.
COMPUTE RA20 = M2T0/MENORES5.
COMPUTE RA21 = M2T1/MENORES5.
COMPUTE RA22 = M2T2/MENORES5.
COMPUTE RA23 = M2T3/MENORES5.
COMPUTE RA24 = M2T4/MENORES5.
COMPUTE RA25 = M2T5/MENORES5.
COMPUTE RA26 = M2T6/MENORES5.
COMPUTE RA27 = M2T7/MENORES5.
COMPUTE SUMA2= RA20 +RA21 +RA22 +RA23 +RA24
+ RA25 + RA26 + RA27 .
COMPUTE SUMA1= RAZON_0 + RAZON_1 + RAZON_2 +RAZON_3 +
RAZON_4 + RAZON_5 + RAZON_6 + RAZON_7.


SAVE OUTFILE='C:\ENAL96\varjul\PASO2.SAV'.




get file='c:\enal96\varjul\menorxc.sav'.
compute  x0hi= des0_1_1/ numencu.
compute  x1hi= des1_1_1/ numencu.
compute  x2hi= des2_1_1/ numencu.
compute  x3hi= des3_1_1/ numencu.
compute  x4hi= des4_1_1/ numencu.
compute  x5hi= des5_1_1/ numencu.
compute  x6hi= des6_1_1/ numencu.
compute  x7hi= des7_1_1/ numencu.
compute  xmenor =  unos / numencu.
compute  xmenor2 =chicoloc /numencu.
compute ymenor =xmenor2 * vivien.
COMPUTE  THI0= d0/numencu *VIVIEN.
COMPUTE  THI1 = d1/numencu* VIVIEN.
COMPUTE  THI2 = d2/numencu* VIVIEN.
COMPUTE THI3 = d3/numencu * VIVIEN.
COMPUTE THI4 = d4/numencu* VIVIEN.
COMPUTE  THI5 = d5/numencu* VIVIEN.
COMPUTE THI6 = d6/numencu * VIVIEN.
COMPUTE THI7 = d7/numencu * VIVIEN.
save outfile='c:\enal96\varjul\menorxc.sav'.
```

```
get file='C:\ENAL96\varjul\compvar.Sav'.
match files  table= 'c:\enal96\varjul\menorxc.sav'
/file=*
/by claveedo estrato clavemun claveloc.
save outfile='c:\enal96\varjul\compvar.sav'.
compute dif10 =(des0_1 - X0HI).
compute dif11 =(des1_1 - X1HI).
compute dif12 =(des2_1 -X2HI).
compute dif13 =(des3_1 - X3HI).
compute dif14 =(des4_1 -X4HI).
compute dif15 =(des5_1 - X5HI).
compute dif16 =(des6_1 - X6HI).
compute dif17 =(des7_1 - X7HI).
compute diflmeno= (nmenor - xmenor).
save outfile='c:\enal96\varjul\compvar.sav'.

compute d02= dif10**2.
compute d12=dif11**2.
compute d22=dif12**2.
compute d32=dif13**2.
compute d42= dif14**2.
compute d52= dif15**2.
compute d62= dif16**2.
compute d72=dif17**2.
compute dmenor2= diflmeno**2.
compute d0dm=dif10 * diflmeno.
compute d1dm=dif12 * diflmeno.
compute d2dm=dif13 * diflmeno.
compute d3dm=dif14 * diflmeno.
compute d4dm=dif15 * diflmeno.
compute d5dm=dif16 * diflmeno.
compute d6dm=dif17 * diflmeno.
compute d7dm=dif17 * diflmeno.
save outfile='c:\enal96\varjul\compvar.sav'.
/*  fin bloque 1  */


AGGREGATE
  /OUTFILE='C:\ENAL96\varjul\S2HIAGG.SAV'
  /BREAK=claveedo estrato clavemun claveloc
  /d1dm_1 = SUM(d1dm) /d2dm_1 = SUM(d2dm) /d3dm_1 = SUM(d3dm) /d4dm_1 =
  SUM(d4dm) /d5dm_1 = SUM(d5dm) /d6dm_1 = SUM(d6dm) /d7dm_1 = SUM(d7dm)
 /d0dm_1 = SUM(d0dm) /d02_1 = SUM(d02) /d12_1 = SUM(d12) /d22_1 = SUM(d22)
 /d32_1 = SUM(d32) /d42_1 = SUM(d42) /d52_1 = SUM(d52) /d62_1 = SUM(d62)
 /d72_1 = SUM(d72) /dmenor_1 = SUM(dmenor2) /vivien_1 = FIRST(vivien)
 /numencu = FIRST(numencu) /chicoloc = FIRST(chicoloc) /famxestr =
  FIRST(famxestr) /locestr = FIRST(locestr) /locxestr = FIRST(locxestr)
/vivloc=first(vivloc)
/estado =first(estado) /VIVIEN =FIRST(VIVIEN)
/xmenor=first(xmenor)
/x0hi x1hi x2hi x3hi x4hi x5hi x6hi x7hi =
 first(x0hi x1hi x2hi x3hi x4hi x5hi x6hi x7hi).

/* bloque2   */

get file='c:\enal96\varjul\s2hiagg.sav'.
DO IF (CLAVEEDO NE 3).
/* EN B.C. SUR HAY DOS LOCALIDADES CON UNA SOLA CASA ENCUESTADA */
/* POR ESO ESTOY EXCLUYENDO A ESTE ESTADO DEL CALCULO DEL  */
/* SEGUNDO COMPONENTE DE VARIANZA   */

compute s20me=d0dm_1/(numencu -1)+ (numencu-vivloc) *xmenor*x0hi.
compute s21me=d1dm_1/(numencu -1)+ (numencu-vivloc) *xmenor*x1hi.
compute s22me=d2dm_1/(numencu -1)+ (numencu-vivloc) *xmenor*x2hi.
compute s23me=d3dm_1/(numencu -1)+ (numencu-vivloc) *xmenor*x3hi.
compute s24me=d4dm_1/(numencu -1)+ (numencu-vivloc) *xmenor*x4hi.
compute s25me=d5dm_1/(numencu -1)+ (numencu-vivloc) *xmenor*x5hi.
compute s26me=d6dm_1/(numencu -1)+ (numencu-vivloc) *xmenor*x6hi.
compute s27me=d7dm_1/(numencu -1)+ (numencu-vivloc) *xmenor*x7hi.
compute s20hi= d02_1 / (numencu -1)+(numencu -vivloc) *x0hi**2.
compute s21hi= d12_1 / (numencu -1)+(numencu -vivloc) *x1hi**2.
```

```
compute s22hi= d22_1 / (numencu -1)+(numencu -vivloc) *x2hi**2.
compute s23hi= d32_1 / (numencu -1)+(numencu -vivloc) *x3hi**2.
compute s24hi= d42_1 / (numencu -1)+(numencu -vivloc) *x4hi**2.
compute s25hi= d52_1 / (numencu -1)+(numencu -vivloc) *x5hi**2.
compute s26hi= d62_1 / (numencu -1)+(numencu -vivloc) *x6hi**2.
compute s27hi= d72_1 / (numencu -1)+(numencu -vivloc) *x7hi**2.
COMPUTE S2MEHI = DMENOR_1 / (numencu -1)+(numencu-vivloc)*
                xmenor**2.
ELSE.
END IF.
save outfile='c:\enal96\varjul\S2AGGS3.SAV'.

/* EN EL SIGUIENTE PASO SE LEEN LOS VALORES DE LAS RAZONES */
/* Y TOTALES  ESTIMADOS   */

MATCH FILES   /FILE=*
 /TABLE='C:\ENAL96\varjul\PASO2.SAV'
 /BY claveedo
 /DROP= TIPO0 TIPO1 TIPO2 TIPO3 TIPO4 TIPO5
TIPO6 TIPO7 RAZON_0 RAZON_1 RAZON_2
RAZON_3 RAZON_4 RAZON_5 RAZON_6 RAZON_7.
EXECUTE.
SAVE OUTFILE='C:\ENAL96\varjul\S2HIAGG.SAV'.

/*  bloque 3  */

GET FILE='C:\ENAL96\varjul\S2HIAGG.SAV'.
DO IF (CLAVEEDO NE 3) .

COMPUTE S20= (S20HI + (RA20)**2 *S2MEHI -2*RA20*S20ME )
* (VIVIEN**2/numencu *(VIVIEN-numencu)/VIVIEN)/(LOCESTR) .
compute s2hi0= s20 *(LOCXESTR**2/LOCESTR).
COMPUTE S21 = (S21HI +RA21**2 *S2MEHI - 2* RA21 *S21ME )
* (VIVIEN**2/numencu *(VIVIEN-numencu)/VIVIEN)/(LOCESTR).
compute s2hi1 = s21 *(LOCXESTR**2/LOCESTR).

COMPUTE S22 =(S22HI +RA22 **2 *S2MEHI - 2*RA22 * S22ME )
* (VIVIEN**2/numencu *(VIVIEN-numencu)/VIVIEN)/(LOCESTR).
compute s2hi2 = s22 *(LOCXESTR**2/LOCESTR).

COMPUTE S23 =(S23HI +RA23 **2 *S2MEHI - 2*RA23*S23ME )
* (VIVIEN**2/numencu *(VIVIEN-numencu)/VIVIEN)/(LOCESTR) .
compute s2hi3 = s23 *(LOCXESTR**2/LOCESTR).

COMPUTE S24= (S24HI + RA24**2 *S2MEHI -2*RA24*S24ME )
* (VIVIEN**2/numencu *(VIVIEN-numencu)/VIVIEN)/(LOCESTR) .
compute s2hi4 = s24 *(LOCXESTR**2/LOCESTR).


COMPUTE S25 = (S25HI +RA25**2*S2MEHI - 2*RA25*S25ME )
* (VIVIEN**2/numencu *(VIVIEN-numencu)/VIVIEN)/(LOCESTR).
compute s2hi5 = s25 *(LOCXESTR**2/LOCESTR).

COMPUTE S26 = (S26HI + RA26**2*S2MEHI- 2*RA26*S26ME )
* (VIVIEN**2/numencu *(VIVIEN-numencu)/VIVIEN)/(LOCESTR).
compute s2hi6 = s26 *(LOCXESTR**2/LOCESTR).

COMPUTE S27=( S27HI + RA27**2 *S2MEHI - 2*RA27*S27ME )
* (VIVIEN**2/numencu *(VIVIEN-numencu)/VIVIEN)/(LOCESTR).
compute s2hi7 = s27 *(LOCXESTR**2/LOCESTR).
ELSE IF (CLAVEEDO =3).
COMPUTE S2HI0 =9999.
COMPUTE S2HI1 = 9999.
COMPUTE S2HI2 =9999.
COMPUTE S2HI3 =9999.
COMPUTE S2HI4=9999.
COMPUTE S2HI5=9999.
COMPUTE S2HI6 = 9999.
COMPUTE S2HI7 = 9999.
COMPUTE S20 =9999.
COMPUTE S21=9999.
COMPUTE S22=9999.
COMPUTE S23 =9999.
COMPUTE S24 =9999.
```

```
COMPUTE S25 =9999.
COMPUTE S26 = 9999.
COMPUTE S27 = 9999.
END IF.

SAVE OUTFILE='C:\ENAL96\varjul\S2HIAGG.SAV'.

/* bloque 4 */
get file ='C:\ENAL96\varjul\S2HIAGG.SAV'.
AGGREGATE
/OUTFILE='C:\ENAL96\varjul\S2HILOC.SAV'
/BREAK = CLAVEEDO ESTRATO
/S2LOC0 S2LOC1 S2LOC2 S2LOC3 S2LOC4 S2LOC5 S2LOC6 S2LOC7=
  SUM(S2HI0 S2HI1 S2HI2 S2HI3 S2HI4 S2HI5 S2HI6 S2HI7)
/NORMALES TIPO1 TIPO2 TIPO3 TIPO4 TIPO5 TIPO6 TIPO7 =
  FIRST(M2T0 M2T1 M2T2 M2T3 M2T4 M2T5 M2T6 M2T7)
/estado =first(estado)   /locestr = first(locestr)
/LOCXESTR  menores5 = FIRST(LOCXESTR  menores5)
/sh0 sh1 sh2 sh3 sh4 sh5 sh6 sh7=
  sum(s20  s21 s22 s23 s24 s25 s26 s27 ).
/* bloque 5 */

GET FILE='C:\ENAL96\varjul\S2HILOC.SAV'.

AGGREGATE
    /OUTFILE='C:\ENAL96\varjul\S2HIEDO.SAV'
    /BREAK = CLAVEEDO
    /S2EST0 S2EST1 S2EST2 S2EST3 S2EST4 S2EST5 S2EST6 S2EST7 =
    SUM(S2LOC0 S2LOC1 S2LOC2 S2LOC3 S2LOC4 S2LOC5 S2LOC6 S2LOC7)
    /menores5 NORMALES TIPO1 TIPO2 TIPO3 TIPO4 TIPO5 TIPO6 TIPO7 =
     FIRST( menores5 NORMALES TIPO1 TIPO2 TIPO3 TIPO4
     TIPO5 TIPO6 TIPO7 )
/estado =first(estado).

GET FILE='C:\ENAL96\varjul\S2HIEDO.SAV'.
DO IF (CLAVEEDO NE 3).
COMPUTE COMP20= S2EST0/MENORES5**2.
COMPUTE COMP21 = S2EST1/MENORES5**2.
COMPUTE COMP22= S2EST2 /MENORES5**2.
COMPUTE COMP23 = S2EST3/MENORES5**2.
COMPUTE COMP24 = S2EST4 /MENORES5**2.
COMPUTE COMP25 = S2EST5 / MENORES5**2.
COMPUTE COMP26 =S2EST6 /MENORES5**2.
COMPUTE COMP27= S2EST7 /MENORES5**2.

END IF.

SAVE OUTFILE= 'C:\ENAL96\varjul\S2HIEDO.SAV'.
/*  bloque 6   */

/* LO QUE SIGUE ES PARA EL CALCULO DEL PRIMER  */
/*  COMPONENTE DE VARIANZA   */

GET FILE='c:\enal96\varjul\menorxc.sav'.
AGGREGATE
    /OUTFILE='C:\ENAL96\varjul\MENORXC2.SAV'
    /BREAK CLAVEEDO ESTRATO
    /TH0 TH1 TH2 TH3 TH4 TH5 TH6 TH7 THMENOR =
      SUM(THI0 THI1 THI2 THI3 THI4 THI5 THI6 THI7 ymenor).
MATCH FILES /FILE=*
 /TABLE ='C:\ENAL96\varjul\MENORXC2.SAV'
 /BY CLAVEEDO ESTRATO.
SAVE OUTFILE='C:\ENAL96\varjul\MENORXC.SAV'.
COMPUTE MEDH0= TH0 / LOCESTR.
COMPUTE MEDH1 = TH1/LOCESTR.
COMPUTE MEDH2 = TH2/LOCESTR.
COMPUTE MEDH3 = TH3/LOCESTR.
COMPUTE MEDH4 = TH4 /LOCESTR.
COMPUTE MEDH5 = TH5 /LOCESTR.
COMPUTE MEDH6 = TH6 /LOCESTR.
COMPUTE MEDH7 = TH7 /LOCESTR.
COMPUTE MEDMENH = THMENOR /LOCESTR.
COMPUTE DIFY = (ymenor - MEDMENH).
COMPUTE DIF0Y = (THI0 - MEDH0) * DIFY.
```

```
COMPUTE DIF1Y = (THI1 - MEDH1) * DIFY.
COMPUTE DIF2Y = (THI2 - MEDH2) * DIFY.
COMPUTE DIF3Y = (THI3 - MEDH3) * DIFY.
COMPUTE DIF4Y = (THI4 - MEDH4 ) * DIFY.
COMPUTE DIF5Y = (THI5 - MEDH5) * DIFY.
COMPUTE DIF6Y = (THI6 - MEDH6) * DIFY.
COMPUTE DIF7Y = (THI7 - MEDH7) * DIFY.

COMPUTE DIFY2 = DIFY * DIFY.
COMPUTE DIF02 = (THI0 - MEDH0) **2.
COMPUTE DIF12 = (THI1 - MEDH1) **2.
COMPUTE DIF22 = (THI2 - MEDH2) **2.
COMPUTE DIF32 = (THI3 - MEDH3) **2.
COMPUTE DIF42 = (THI4 - MEDH4 ) **2.
COMPUTE DIF52 = (THI5 - MEDH5) **2.
COMPUTE DIF62 = (THI6 - MEDH6) **2.
COMPUTE DIF72 = (THI7 - MEDH7) **2.
SAVE OUTFILE='C:\ENAL96\varjul\MENORXC.SAV'.

/*  bloque 7 */

GET FILE='C:\ENAL96\varjul\MENORXC.SAV'.
AGGREGATE
/OUTFILE='C:\ENAL96\varjul\COMPS1.SAV'
/BREAK CLAVEEDO ESTRATO
/P0Y P1Y P2Y P3Y P4Y P5Y P6Y P7Y =
SUM( DIF0Y DIF1Y DIF2Y DIF3Y DIF4Y DIF5Y DIF6Y DIF7Y)
/SQ0 SQ1 SQ2 SQ3 SQ4 SQ5 SQ6 SQ7 SQY =
SUM(Dif02 Dif12 Dif22 Dif32 Dif42 Dif52 Dif62 Dif72 Dify2)
/LOCESTR LOCXESTR = FIRST(LOCESTR LOCXESTR).

/*  bloque  bloque  */

GET FILE='C:\ENAL96\varjul\COMPS1.SAV'.
/* encontre que en el estado 16 se dan menos de dos locs. */
/* en estrato 18   */

/*  FILTER OFF.  */
/* USE ALL.  */
/* SELECT IF((claveedo ne 16)). */
/* EXECUTE .  */

MATCH FILES   /file= *
/TABLE = 'C:\ENAL96\varjul\S2HILOC.SAV'
/BY CLAVEEDO ESTRATO
/DROP = S2LOC0 S2LOC1 S2LOC2 S2LOC3 S2LOC4 S2LOC5
        S2LOC6 S2LOC7 NORMALES TIPO1 TIPO2
        TIPO3 TIPO4 TIPO5 TIPO6 TIPO7.
SAVE OUTFILE='C:\ENAL96\varjul\COMPS1.SAV'.

get file='C:\ENAL96\varjul\COMPS1.SAV'.
MATCH FILES /FILE=*
 /TABLE='C:\ENAL96\varjul\PASO2.SAV'
 /BY claveedo
 /DROP= TIPO0 TIPO1 TIPO2 TIPO3 TIPO4 TIPO5
TIPO6 TIPO7  RAZON_0 RAZON_1 RAZON_2
RAZON_3 RAZON_4 RAZON_5 RAZON_6 RAZON_7.
EXECUTE.

SAVE OUTFILE='C:\ENAL96\varjul\COMPS1.SAV'.

COMPUTE SCH0Y= P0Y /(LOCESTR -1).
COMPUTE SCH1Y= P1Y /(LOCESTR -1).
COMPUTE SCH2Y= P2Y /(LOCESTR -1).
COMPUTE SCH3Y= P3Y /(LOCESTR -1).
COMPUTE SCH4Y= P4Y /(LOCESTR -1).
COMPUTE SCH5Y= P5Y /(LOCESTR -1).
COMPUTE SCH6Y= P6Y /(LOCESTR -1).
COMPUTE SCH7Y= P7Y /(LOCESTR -1).
COMPUTE SCHY = SQY /  (LOCESTR -1).
COMPUTE SCH0= SQ0 /(LOCESTR -1).
COMPUTE SCH1= SQ1 /(LOCESTR -1).
COMPUTE SCH2= SQ2 /(LOCESTR -1).
COMPUTE SCH3= SQ3 /(LOCESTR -1).
```

```
COMPUTE SCH4= SQ4 /(LOCESTR -1).
COMPUTE SCH5= SQ5 /(LOCESTR -1).
COMPUTE SCH6= SQ6 /(LOCESTR -1).
COMPUTE SCH7= SQ7 /(LOCESTR -1).
SAVE OUTFILE='C:\ENAL96\varjul\COMPS1.SAV'.

/*   BLOQUE   */

COMPUTE S2CH0 = SCH0 + RA20**2 * SCHY -2*RA20 * SCH0Y.
COMPUTE S2CH1 = SCH1 + RA21**2 * SCHY - 2*RA21 * SCH1Y.
COMPUTE S2CH2 = SCH2 + RA22**2 * SCHY - 2*RA22 * SCH2Y.
COMPUTE S2CH3 = SCH3 + RA23**2 * SCHY - 2*RA23 * SCH3Y.
COMPUTE S2CH4 = SCH4 + RA24**2 * SCHY - 2*RA24 * SCH4Y.
COMPUTE S2CH5 = SCH5 + RA25**2 * SCHY - 2*RA25 * SCH5Y.
COMPUTE S2CH6 = SCH6 + RA26**2 * SCHY - 2*RA26 * SCH6Y.
COMPUTE S2CH7 = SCH7 + RA27**2 * SCHY - 2*RA27 * SCH7Y.

SAVE OUTFILE='C:\ENAL96\varjul\COMPS1.SAV'.


COMPUTE S21H0 = S2CH0 - SH0.
COMPUTE S21H1 = S2CH1 - SH1.
COMPUTE S21H2 = S2CH2 - SH2.
COMPUTE S21H3 = S2CH3 - SH3.
COMPUTE S21H4 = S2CH4 - SH4.
COMPUTE S21H5 = S2CH5 - SH5.
COMPUTE S21H6 = S2CH6 - SH6.
COMPUTE S21H7 = S2CH7 - SH7.

COMPUTE GLOBAL0= S2CH0 *(LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.
COMPUTE GLOBAL1= S2CH1 *(LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.
COMPUTE GLOBAL2= S2CH2 *(LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.
COMPUTE GLOBAL3= S2CH3 *(LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.
COMPUTE GLOBAL4= S2CH4 *(LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.
COMPUTE GLOBAL5= S2CH5 *(LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.
COMPUTE GLOBAL6= S2CH6 *(LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.
COMPUTE GLOBAL7= S2CH7 *(LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.


COMPUTE VARH0 = S21H0 * (LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.


COMPUTE VARH1 = S21H1 * (LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.

COMPUTE VARH2 = S21H2 * (LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.

COMPUTE VARH3 = S21H3 * (LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.

COMPUTE VARH4 = S21H4 * (LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.

COMPUTE VARH5 = S21H5 * (LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.

COMPUTE VARH6 = S21H6 * (LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.

COMPUTE VARH7 = S21H7 * (LOCXESTR**2/LOCESTR)
* (LOCXESTR -LOCESTR) / LOCXESTR.

SAVE OUTFILE='C:\ENAL96\varjul\COMPS1.SAV'.
```

```
/*   BLOQUE· */
GET FILE='C:\ENAL96\varjul\COMPS1.SAV'.
AGGREGATE
    /OUTFILE='C:\ENAL96\varjul\comps2.sav'
    /BREAK = CLAVEEDO
    /VAR0 VAR1 VAR2 VAR3 VAR4 VAR5 VAR6 VAR7 =
    SUM( VARH0 VARH1 VARH2 VARH3 VARH4 VARH5 VARH6 VARH7)
    /NORMALES TIPO1 TIPO2 TIPO3 TIPO4 TIPO5 TIPO6 TIPO7 =
    FIRST(M2T0 M2T1 M2T2 M2T3 M2T4 M2T5 M2T6 M2T7)
    /GLOBAL0 GLOBAL1 GLOBAL2 GLOBAL3 GLOBAL4 GLOBAL5
    GLOBAL6 GLOBAL7  = SUM(GLOBAL0 GLOBAL1 GLOBAL2 GLOBAL3
GLOBAL4 GLOBAL5  GLOBAL6 GLOBAL7 )
    /MENORES5=FIRST(MENORES5)
    /ESTADO = FIRST(ESTADO).
GET FILE='C:\ENAL96\varjul\COMPS2.SAV'.
COMPUTE GLOBA0=GLOBAL0/MENORES5**2.
COMPUTE GLOBA1 = GLOBAL1/MENORES5**2.
COMPUTE GLOBA2 = GLOBAL2 / MENORES5**2.
COMPUTE GLOBA3 = GLOBAL3 / MENORES5**2.
COMPUTE GLOBA4=  GLOBAL4/MENORES5**2.
COMPUTE GLOBA5 = GLOBAL5/MENORES5**2.
COMPUTE GLOBA6 = GLOBAL6 / MENORES5**2.
COMPUTE GLOBA7 = GLOBAL7 / MENORES5**2.


COMPUTE COMP10=VAR0/menores5**2.
COMPUTE COMP11=VAR1/menores5**2.
COMPUTE COMP12 = VAR2/menores5**2.
COMPUTE COMP13= VAR3/ menores5**2.
COMPUTE COMP14 = VAR4 / menores5**2.
COMPUTE COMP15 = VAR5 / menores5**2.
COMPUTE COMP16 = VAR6/ MENORES5**2.
COMPUTE  COMP17 = VAR7 / MENORES5**2.
SAVE OUTFILE='C:\ENAL96\varjul\COMPS2.SAV'.
MATCH FILES
/FILE=*
/TABLE = 'C:\ENAL96\varjul\S2HIEDO.SAV'
/BY CLAVEEDO.
SAVE OUTFILE = 'C:\ENAL96\varjul\COMPnew0.SAV'.
```

```
      PROGRAM nacionjk
c Este programa calcula las varianzas jackknife para todos los estados.
C      PARA CORRER EL PROGRAMA PARA UN ESTADO DETERMINADO
C      SE DEBE CONSULTAR EL ARCHIVO NEDO.DAT Y CHECAR LA
C      CUARTA COL., PUES DA EL NUMERO DE ENCUESTAS EN EL NOMBRES
C      CON ESE DATO SE CORRIGE LA DIMENSION DEL ARREGLO EDOX( )
C      TAMBIEN HAY QUE PONER EN EL "DO" DE LEER LOS DATOS DEL NOMBRES,
C      LA UBICACION CORRECTA EN EL ARCHIVO NEDO(  ).
C      ES DECIR, SI EDO< 9 , LA UBICAION DE FILA ES EL # DE NOMBRES;
C      SI EDO > 9 ES EL (#EDO -1), YA QUE NO HAY DATOS DEL D.F.

      DIMENSION NCASA(854, 5) , NEDO(31, 4)  ,NEST(371, 3)
      DIMENSION DATOS(18988,19)
      integer ncasa , nedo , nest
      INTEGER VA ,VACASA, VALOC, VALOEDO ,E , I2, H,  I3, J,I ,K , F, S
      REAL*8   X1EST, X2EST, X3EST, X4EST, X5EST, YEST
      REAL*8   R1, R2, R3, R4, R5

      DIMENSION X(77, 10) , x_mean(5) , V(5)
      CHARACTER NOMBREs(31)*15 ,   CLASE(5)*18

      REAL*8 X , X_MEAN    , V
      OPEN(4,FILE='C:\FORT32\ITERAR.DAT', STATUS='OLD')
      OPEN(6,FILE= 'C:\FORT32\samples\ESTIMADO.DAT', STATUS='OLD')
      OPEN (5,FILE='C:\FORT32\SAMPLES\NCASA.DAT', STATUS='UNKNOWN')
      OPEN (3,FILE='C:\FORT32\NEDO.DAT', STATUS='UNKNOWN')
      OPEN (2, FILE='C:\FORT32\NEST.DAT', STATUS='UNKNOWN')
      OPEN (1,FILE='C:\FORT32\DATOS.DAT', STATUS='UNKNOWN')

      NOMBRES(1)='AGUASCALIENTES '
      NOMBRES(2)='BAJA CALIFORNIA'
      NOMBRES(3)='B. C. SUR      '
      NOMBRES(4)='CAMPECHE       '
      NOMBRES(5)='COAHUILA       '
      NOMBRES(6)='COLIMA         '
      NOMBRES(7)='CHIAPAS        '
      NOMBRES(8)='CHIHUAHUA      '
      NOMBRES(9)='DURANGO        '
      NOMBRES(10)='GUANAJUATO     '
      NOMBRES(11)='GUERRERO       '
      NOMBRES(12)='HIDALGO        '
      NOMBRES(13)='JALISCO        '
      NOMBRES(14)='MEXICO         '
      NOMBRES(15)='MICHOACAN      '
      NOMBRES(16)='MORELOS        '
      NOMBRES(17)='NAYARIT        '
      NOMBRES(18)='NUEVO LEON     '
      NOMBRES(19)='OAXACA         '
      NOMBRES(20)='PUEBLA         '
      NOMBRES(21)='QUERETARO      '
      NOMBRES(22)='QUINTANA ROO   '
```

```
      NOMBRES(23)='SAN LUIS POTOSI'
      NOMBRES(24)='SINALOA        '
      NOMBRES(25)='SONORA         '
      NOMBRES(26)='TABASCO        '
      NOMBRES(27)='TAMAULIPAS     '
      NOMBRES(28)='TLAXCALA       '
      NOMBRES(29)='VERACRUZ       '
      NOMBRES(30)='YUCATAN        '
      NOMBRES(31)='ZACATECAS      '

         CLASE(1)  ='NORMALES y casi N.'
         CLASE(2)  ='BAJITOS GORDITOS  '
         CLASE(3)  ='MAL PARA LA EDAD  '
         CLASE(4)  ='MAL PARA LA TALLA '
         CLASE(5)  ='MAL EDAD Y TALLA  '




c     SE LEEN ARCHIVOS QUE CONTIENEN INFORMACION SOBRE CUANTOS
C     ESTRATOS , LOCALIDADES Y ENCUESTAS HAY POR ESTADO, POR ESTRATO
C      Y POR LOCALIDAD. ( SEGUN CORRESPONDA)
      DO 10 J=1 , 854

      READ (5,*)  (NCASA(J,I), I=1,5)
C     NCASA INDICA: NO. DE ENCUESTAS POR
C                  ESTADO, ESTRATO MUNICIPIO Y LOCALIDAD
   10 CONTINUE
      CLOSE(5)
      DO 20 J=1 , 31
      READ (3, *) (NEDO(J,I), I=1,4)
   20 CONTINUE
      CLOSE(3)
C     NEDO INDICA POR ESTADO, NO. DE ESTRATOS, NO. TOTAL DE LOCALIDADES
C      EN EL ESTADO Y NO. TOTAL DE ENCUESTAS EN EL ESTADO.
      DO 30 J=1, 371
      READ(2, *) (NEST(J,I) ,I= 1, 3)
   30 CONTINUE
      CLOSE(2)
C      NEST INDICA EN CADA ESTRATO (LA PRIMERA COL. ES ESTADO
C      Y LA SEGUNDA EL ESTRATO), EL NO. DE LOCALIDADES ENCUESTADAS
C      EN EL ESTRATO.
      DO 40 J=1, 18988
      READ(1, *) (DATOS(J,I) , I= 1, 13)

   40  CONTINUE
      CLOSE(1)
C     DATOS CONTIENE POR ORDEN DE COL.: ESTADO, ESTRATO, MUNICIPIO, LOCALID
AD
C     ENCUESTA ; LAS COLUMNAS 6-10 CONTIENE EL NUMERO DE MENORES EN CATEGOR
IA
C     1-5 (SEGUN CORRESPONDA EL ORDEN) QUE HAY EN LA CASA ENCUESTADAS; LA C
```

```
OL. 11
C     CONTIENE EL NUMERO DE NINOS ENCUESTADOS EN LA CASA.; LA COL. 12 TIENE
 EL FACTOR DE
C     EXPANSION Y LA 13 TAMBIEN (EN EL PROGRAMA SE AJUSTA ESA COL, DE ACUER
DO
C      AL METODO JACKKNIFE.
C NOTESE QUE LA COL. 11 TIENE AL NO. DE NINOS ENCUESTADOS.  ESTO ES PORQUE
C EL FACTOR EN COL. 12 CONTEMPLA EL FACTOR POR CASA (MENORES EN LA CASA/MEN
ORES ENCUESTADOS);
C AL USAR NINOS ENCUESTADOS, ESTE VALOR SE CANCELA Y QUEDA MENORES EN LA CA
SA, QUE ES LO QUE
C ME INTERESA.(PARA ESTIMAR "MENORES" NO VIENE AL CASA LA TERCERA ETAPA DE
MUESTREO,
C POR LO QUE NO SE USA EL FACTOR POR CASA.

      VACASA=0
      I2=0
      VA=0
      VALOC=0

       DO 900 S=1, 32
       IF (S.EQ.9) THEN
       CYCLE
       ELSE
       END IF
        WRITE(*,'(A)') ' EL ESTADO : '
         WRITE(*,'(I9)') S

       E= S

      IF (E.EQ.1) THEN
      VA=0
      VALOC=0
      F=E
      I1=1
      I2 = NEDO(F,4) + I2
      ELSE
          IF (E.GE.9) THEN
          F= E-1
          ELSE
          F= E
          END IF

       I1 =I2+1
       I2 = NEDO(F, 4) +I2

      END IF
```

```
      VALOEDO =0
      DO 42 K=1,5
      V(K) =0
      X_MEAN(K) =0

   42 CONTINUE
C

C      INICIA EL CICLO POR ESTRATO, DENTRO DEL NOMBRES "E"
          DO 800  H=1 , NEDO(F,2)


          VA= VA+1
          I3=VACASA + 1
       DO 50 I=I1  , I2
          DATOS(I,13) = DATOS(I,12)
             DO 45 J=14, 19
             DATOS(I,J) = DATOS(I,13)* DATOS(I,(J-8))

   45        CONTINUE
   50     CONTINUE
C
C       SE CAMBIA EL VALOR DEL FACTOR A LOS DATOS DEL ESTRATO DE DONDE
C       ESTAREMOS ELIMINANDO LOCALIDADES
C
        DO 54 WHILE  ((DATOS(I3,1).EQ.E). AND .(DATOS(I3,2).EQ.H))
        DATOS(I3,13) = DATOS(I3,12)*(NEST(VA,3)/(NEST(VA,3) - 1))

                DO 52 J=14,19
                DATOS(I3,J) = DATOS(I3,13)*DATOS(I3,(J-8))

   52           CONTINUE
             I3=I3+1

   54        CONTINUE

C
C      AHORA SIGUE EL CICLO POR LOCALIDAD, DENTRO DEL ESTRATO "H", NOMBRES
"E"
C
                DO 700 J=1 , NEST(VA,3)
                VALOC=VALOC+1
                VALOEDO= VALOEDO+1
C       SE LE DA UN FACTOR DE EXP. DE "0" A LAS ENCUESTAS DE LA LOCALIDAD
C       QUE SE ESTA ELIMINANDO

                 CASA1L= VACASA
                   DO 60 I=1, NCASA(VALOC,5)
                   VACASA = VACASA +1
                   DATOS(VACASA,13)= 0
                   DATOS(VACASA,14)=0
```

```
                         DATOS(VACASA,15)=  0
                         DATOS(VACASA,16)=0
                         DATOS(VACASA,17)=  0
                         DATOS(VACASA,18)=0
                         DATOS(VACASA,19)=0

   60                    CONTINUE

C     SE INICIALIZAN EN "0" LAS VARIABLES QUE CONTENDRAN LA SUMATORIA
C     QUE DARA , A FIN DE CUENTAS, EL ESTIMADOR
                 X1EST=0
                 X2EST=0
                 X3EST=0
                 X4EST=0
                 X5EST=0
                 YEST=0
C     EL SIGUIENTE CICLO HACE LA SUMATORIA PARA OBTENER EL ESTIMADOR
C     DE MENORES EN CADA GRUPO

                         DO 70 I=I1,  I2
                         X1EST = X1EST+DATOS(I,14)
                         X2EST= X2EST + DATOS(I,15)
                         X3EST= X3EST+DATOS(I,16)
                         X4EST = X4EST+DATOS(I,17)
                         X5EST = X5EST+DATOS(I,18)
                         YEST = YEST +DATOS(I,19)
   70                      CONTINUE

C     SE OBTIENEN LOS ESTIMADORES DE RAZON, DIVIDIENDO EL ESTIMADOR
C     DE MENORES EN UN GRUPO, POR EL ESTIMADOR DEL TOTAL DE MENORES
C

                 IF (YEST.EQ.0) THEN
                 R1=9999
                 R2=9999
                 R3=9999
                 R5=9999
                 WRITE(*,'(A)') '  ERROR, UNA MUESTRA DA 0 MENORES '
                 WRITE(*,'(I9)') H
                 GO TO 999
                 ELSE
                 R1= X1EST/YEST
                 R2=X2EST/YEST
                 R3=X3EST/YEST
                 R4=X4EST/YEST
                 R5=X5EST/YEST
                 END IF


                 X(VALOEDO,1) = R1
                 X(VALOEDO,2) = R2
                 X(VALOEDO, 3) = R3
```

```
                   X(VALOEDO,4)  = R4
                   X(VALOEDO,5)  = R5



C       PARA LA PROXIMA VUELTA EN EL ESTRATO, QUIERO QUITAR LOS CEROS Y PONE
R
C       EL FACTOR DE EXPANSION CORREGIDO POR ELIMINAR OTRA LOCALIDAD.
      DO 200 I=1, NCASA(VALOC,5)
      CASA1L=CASA1L+ 1
      DATOS(CASA1L,13)=DATOS(CASA1L,12)*NEST(VA,3)/ (NEST(VA,3) -1)
      DO 190 K=14,19
      DATOS(CASA1L,K) = DATOS(CASA1L,13)*DATOS(CASA1L,(K-8))
  190 CONTINUE
  200 CONTINUE
C            SE TERMINA EL CICLO POR LOCALIDAD
  700                  CONTINUE

c       EL SIGUIENTE CONTINUE ES PARA EL DO POR ESTRATO
  800 CONTINUE

          DO 80 K=1,5
              do 75 I=1,VALOEDO
              x_mean(k) = x_mean(k) + x(I,k)
c              X_1 = X_1 + X(I,1)
c              X_2 = X_2+ X(I,2)
c              X_3 = X_3 +X(I,3)
c              X_4= X_4 + X(I,4)
c              X_5 = X_5 + X(I,5)
  75         continue
            x_mean(k) = x_mean(k) / valoedo
  80          CONTINUE
C       SE OBTIENEN LOS PROMEDIOS DE TODOS LOS JACKNNIFE DEL NOMBRES


          DO 95 k=1,5
            do 90 J = 1, VALOEDO
          x(j,(k+5)) = (x(j,k) - x_mean(k))**2
          V(K) = V(K) + X(J,(K+5))*(NEST(VA,3) -1)/ NEST(VA,3)
  90               CONTINUE
  95            CONTINUE
         CLOSE(1, STATUS='DELETE')
C


C        SE ESCRIBE UN ARCHIVO CON LAS ITERACIONES
         WRITE(4,'(A)') '
         WRITE(4,'(I4\)') S
         WRITE(4,'(A20\)' )  NOMBREs(F)
         WRITE(4,500 ) (X(1,J), J=1,5)
         DO 510 I=2,VALOEDO
```

```
      WRITE(4,505) (X(I,J) , J=1,5)
500   FORMAT( F12.6,F12.6, F12.6 , F12.6, F12.6 )
505   FORMAT(24X, F12.6,F12.6, F12.6 , F12.6, F12.6 )
510   CONTINUE
      WRITE(4,'(A)') '
C     SE ESCRIBE UN ARCHIVO CON LOS ESTIMADORES Y VARIANZAS

      WRITE(6,'(I4\)') S
      WRITE(6,'(A20)' )  NOMBREs(F)
     do 520 i=1,5
     WRITE(6,515) clase(i)
515     FORMAT(24x, a20\)
     write(6,516) x_mean(i)
516     format(f12.6\)
     write(6,517) v(i)
517    format( f12.6)
520 continue


C         EL SIGUIENTE CONTINUE ES PARA EL CICLO  POR ESTADO
900 CONTINUE
    CLOSE  (4, STATUS='KEEP')
    CLOSE  (6, STATUS='KEEP')

999   END
```

```
      PROGRAM halfsCOM
C PROG. QUE CALCULA UN PORCENTIL POR HALF-SAMPLING , DE UNA MUESTRA
C Y SU COMPLEMENTO

      DIMENSION NCASAN(854, 4) , NEDON(31, 4)  ,NEST(371, 3)
      DIMENSION NINOS(26700,10),PERCENTI(32,4),HADAM(32,32)
      DIMENSION HADAM8(8,8),HADAM12(12,12),HADAM20(20,20),HADAM28(28,28)
      DIMENSION  HADAM32(32,32), MEDFORMU(93,2), COMPLEM(32,4)
      integer NCASAN , NEDON , NEST, INUT, MEPASE ,LOC1
      INTEGER r1,r,E,I1,I2,I3,H,J,I,F,S
      REAL*4  P, C , ESTIMO, ESTANT, NUMERADO, DENOMINA
      REAL*4   PERCENTI, NINOS , MEDFORMU, COMPLEM, SPER, SPERC

      INTEGER HADAM8, HADAM12,HADAM20, HADAM28
      INTEGER HADAM32

      CHARACTER NOMBREs(31)*15      , CUAL*20

      REAL*4 PERMED, PERMEDC, V1,V2, V1C, V2C, V1M, V2M

      OPEN(4,FILE='C:\FORT32\PERC\ITERAR.DAT', STATUS='OLD')
      OPEN(6,FILE= 'C:\FORT32\PERC\ESTIMADO.DAT', STATUS='OLD')
      OPEN (5,FILE='C:\FORT32\PERC\NCASAN.DAT', STATUS='UNKNOWN')
      OPEN (3,FILE='C:\FORT32\PERC\NEDON.DAT', STATUS='UNKNOWN')
      OPEN (2, FILE='C:\FORT32\PERC\NEST.DAT', STATUS='UNKNOWN')
      OPEN (1,FILE='C:\FORT32\PERC\NINOS.DAT', STATUS='OLD')
      OPEN (7,FILE='C:\FORT32\PERC\HADAM8.DAT',STATUS='OLD')
      OPEN (8,FILE='C:\FORT32\PERC\HADAM12.DAT',STATUS='OLD')
      OPEN (9,FILE='C:\FORT32\PERC\HADAM20.DAT',STATUS='OLD')
      OPEN (10,FILE-'C:\FORT32\PERC\HADAM28.DAT',STATUS='OLD')
      OPEN (11,FILE='C:\FORT32\PERC\HADAM32.DAT',STATUS='OLD')
      OPEN (12,FILE='C:\FORT32\PERC\MEDFORMU.DAT',STATUS='OLD')
      NOMBRES(1)='AGUASCALIENTES '
      NOMBRES(2)='BAJA CALIFORNIA'
      NOMBRES(3)='B. C. SUR      '
      NOMBRES(4)='CAMPECHE       '
      NOMBRES(5)='COAHUILA       '
      NOMBRES(6)='COLIMA         '
      NOMBRES(7)='CHIAPAS        '
      NOMBRES(8)='CHIHUAHUA      '
      NOMBRES(9)='DURANGO        '
      NOMBRES(10)='GUANAJUATO     '
      NOMBRES(11)='GUERRERO       '
      NOMBRES(12)='HIDALGO        '
      NOMBRES(13)='JALISCO        '
      NOMBRES(14)='MEXICO         '
      NOMBRES(15)='MICHOACAN      '
      NOMBRES(16)='MORELOS        '
      NOMBRES(17)='NAYARIT        '
      NOMBRES(18)='NUEVO LEON     '
      NOMBRES(19)='OAXACA
```

```
      NOMBRES(20)='PUEBLA          '
      NOMBRES(21)='QUERETARO       '
      NOMBRES(22)='QUINTANA ROO    '
      NOMBRES(23)='SAN LUIS POTOSI'
      NOMBRES(24)='SINALOA         '
      NOMBRES(25)='SONORA          '
      NOMBRES(26)='TABASCO         '
      NOMBRES(27)='TAMAULIPAS      '
      NOMBRES(28)='TLAXCALA        '
      NOMBRES(29)='VERACRUZ        '
      NOMBRES(30)='YUCATAN         '
      NOMBRES(31)='ZACATECAS       '




c SE LEEN ARCHIVOS QUE CONTIENEN INFORMACION SOBRE CUANTOS
C ESTRATOS , LOCALIDADES Y ENCUESTAS HAY POR ESTADO, POR ESTRATO
C Y POR LOCALIDAD. ( SEGUN CORRESPONDA)
      DO 10 J=1 , 854

      READ (5,*)   (NCASAN(J,I), I=1,4)
C NCASAN INDICA: NO. DE ENCUESTAS POR
C ESTADO, ESTRATO, Y LOCALIDAD
   10 CONTINUE
      CLOSE(5)
      DO 20 J=1 , 31
      READ (3, *) (NEDON(J,I), I=1,4)
   20 CONTINUE
      CLOSE(3)
C NEDON INDICA POR ESTADO, NO. DE ESTRATOS, NO. TOTAL DE LOCALIDADES
C EN EL ESTADO Y NO. TOTAL DE ENCUESTAS A NINOS  EN EL ESTADO.
      DO 30 J=1, 371
      READ(2, *) (NEST(J,I) ,I= 1, 3)
   30 CONTINUE
      CLOSE(2)
C NEST INDICA EN CADA ESTRATO (LA PRIMERA COL. ES ESTADO
C Y LA SEGUNDA EL ESTRATO), EL NO. DE LOCALIDADES ENCUESTADAS
C EN EL ESTRATO.
      DO 40 J=1, 26700
      READ(1, *) (NINOS(J,I), I= 1, 8)

   40  CONTINUE
      CLOSE(1)
C NINOS CONTIENE POR ORDEN: ESTADO , ESTRATO, CLAVE, ENCUESTA,
c  PEDZ, TEDZ, PETZ, FACT. EXP.
c
       DO 300 J=1, 8
       READ(7,*) (HADAM8(J,I), I=1,8)
  300  CONTINUE
```

```
      CLOSE(7)
      DO 310 J=1, 12
      READ(8,*) (HADAM12(J,I), I=1,12)
 310  CONTINUE
      CLOSE(8)
      DO 320 J=1, 20
      READ(9,*) (HADAM20(J,I), I=1,20)
 320  CONTINUE
      CLOSE(9)
      DO 330 J=1, 28
      READ(10,*) (HADAM28(J,I), I=1,28)
 330  CONTINUE
      CLOSE(10)
      DO 340 J=1, 32
      READ(11,*) (HADAM32(J,I), I=1,32)
 340  CONTINUE
      CLOSE(11)
      DO 345 J=1,93
      READ(12,*) (MEDFORMU(J,I), I=1,2)
 345     CONTINUE
      CLOSE(12)


c SI SE QUIERE UN PERCENTIL DE PEDZ; ENTONCES, INUT=5
C SI SE QUIERE PERCENTIL DE   PETZ             INUT =6
C SI SE QUIERE PERCENTIL DE   TEDZ             INUT =7
c LA VARIABLE P  CONTIENE EL PERCENTIL QUE SE DESEA ESTIMAR
     INUT = 7
      P = 0.50

      IF (INUT.EQ.5) THEN
      CUAL='PESO PARA LA EDAD    '
      ELSE IF (INUT.EQ.6) THEN
      CUAL='PESO PARA LA TALLA   '
      ELSE
      CUAL='TALLA PARA LA EDAD   '
      END IF
      WRITE(4,'(A\)')  CUAL
        WRITE(4,400)  ' PERCENTIL ', P
 400    FORMAT(A,4X,F6.3)
      WRITE(6,'(A)') '                  '
        WRITE(6,'(A)')  CUAL
        WRITE(6,440)  ' PERCENTIL ', P
 440    FORMAT(A,4X,F6.3)



     I2=0


      DO 900 S=1, 32
       IF (S.EQ.9) THEN
```

```
      CYCLE
      END IF
      E= S
       IF (E.EQ.1) THEN
     VA=0
     VALOC=0
     F=E
     I1=1
     I2 = NEDON(F,4)
     ELSE
         IF (E.GE.9) THEN
         F= E-1
         ELSE
         F= E
         END IF

      I1 =I2+1
      I2 = NEDON(F, 4) +I2

     END IF
     IF (E.GT.27) THEN
     EXIT
     END IF

     IF ((E.EQ.1).OR.(E.EQ.11).OR.(E.EQ.15)) THEN
     GO TO 200
     ELSE IF ((E.EQ.18).OR.(E.EQ.20).OR.(E.EQ.22)) THEN
     GO TO 200
     ELSE IF ((E.EQ.26).OR.(E.EQ.27) ) THEN
     GO TO 200
     ELSE
     CYCLE
     END IF


200   IF ((E.EQ.1).OR.(E.EQ.18).OR.(E.EQ.26) ) THEN
     R=8
     DO 100 I=1, R
     DO 103 J=1, R
     HADAM(I,J) = HADAM8(I,J)
103     CONTINUE
100     CONTINUE
     ELSE  IF (E.EQ.22) THEN
         R=12
         DO 107 I=1, R
         DO 105 J=1, R
         HADAM(I,J) = HADAM12(I,J)
105     CONTINUE
107     CONTINUE
         ELSE IF ((E.EQ.11) .OR. (E.EQ.27)) THEN
             R=20
             DO 111 I=1, R
```

```
                DO 109 J=1, R
                HADAM(I,J)  =  HADAM20(I,J)
  109           CONTINUE
  111          CONTINUE
               ELSE IF (E.EQ.15) THEN
                   R=28
                   DO 115 I=1, R
                    DO 113 J=1, R
                    HADAM(I,J)  =  HADAM28(I,J)
  113              CONTINUE
  115              CONTINUE
                   ELSE IF (E.EQ.20) THEN
                       R=32
                       DO 119 I=1, R
                       DO 117 J=1, R
                       HADAM(I,J)  =  HADAM32(I,J)
  117                  CONTINUE
  119                  CONTINUE
                       END IF


      I3 = I1
      DO  45  H=1, NEDON(F,2)
      LOC1=NINOS(I3,3)
          DO 52 WHILE(NINOS(I3,2).EQ.H)
          IF (NINOS(I3,3).EQ.LOC1) THEN
          NINOS(I3,3) =1
          ELSE
          NINOS(I3,3)  =  2
          END IF
          I3 =I3 +1
   52       CONTINUE
   45 CONTINUE

c Comienza el ciclo que busca los estimadores de cada replica .

       DO 800 R1=1, R
C
       DO 50 I=I1   , I2
       H = NINOS(I,2)
      NINOS(I,9)=NINOS(I,8)*(1+  (-1)**(NINOS(I,3)  +1)*HADAM(R1,H+1)*0.5)

   50      CONTINUE
C

C LO QUE SIGUE ES LA PARTE QUE BUSCA EL PERCENTIL DESEADO, ESPECIFICADO
C POR LA VARIABLE P QUE SE DA AL INICIOO DEL PROGRAMA.
C
       ESTANT = P
       C=-1
       MEPASE =0
   80  NUMERADO =0
```

```
      DENOMINA =0
      I4 = I1
      DO 90 WHILE (NINOS(I4,1). EQ.E)
      IF (NINOS(I4,INUT).LE.C) THEN
      NINOS(I4,10) = 1
      ELSE
      NINOS(I4,10) = 0
      END IF
      NINOS(I4,10) = NINOS(I4,9) * NINOS(I4,10)
      NUMERADO = NINOS(I4,10) +NUMERADO
      DENOMINA = NINOS(I4,9) + DENOMINA

      I4=I4 +1

   90 CONTINUE


      ESTIMO = NUMERADO / DENOMINA

      IF (ESTIMO.EQ.P)  THEN
        PERCENTI(R1,1) =C
        PERCENTI(R1,3) =ESTIMO
        GO TO 800
      ELSE  IF (ESTIMO.LT.P) THEN
               IF (MEPASE.EQ.1) THEN
               PERCENTI(R1,1)= C+0.1
               PERCENTI(R1,3) = ESTANT
               GO TO 120
                               ELSE
               C=C+0.1
               GO TO 80
                             END IF
            ELSE
                MEPASE=1
                ESTANT = ESTIMO
                C=C - 0.1
                GO TO 80
            END IF





  120 PERCENTI(R1,1)=C+ 0.1/(ESTANT-ESTIMO)*(P-ESTIMO)


  800    CONTINUE
c Ahora es cuando se busca el estimador del complemento.

        DO 1800 R1=1, R
C
```

```
      DO 1050 I=I1   , I2
      H = NINOS(I,2)
     NINOS(I,9)=NINOS(I,8)*(1+(-1)**(NINOS(I,3)+1)*(-HADAM(R1,H+1))*.5)

 1050     CONTINUE
C

C LO QUE SIGUE ES LA PARTE QUE BUSCA EL PERCENTIL DESEADO, ESPECIFICADO
C POR LA VARIABLE P QUE SE DA AL INICIOO DEL PROGRAMA.
C
      ESTANT = P
      C=-1
      MEPASE =0
 1080 NUMERADO =0
      DENOMINA =0
      I4 = I1
      DO 1090 WHILE (NINOS(I4,1). EQ.E)
      IF (NINOS(I4,INUT).LE.C) THEN
      NINOS(I4,10) = 1
      ELSE
      NINOS(I4,10) = 0
      END IF
      NINOS(I4,10) = NINOS(I4,9) * NINOS(I4,10)
      NUMERADO = NINOS(I4,10) +NUMERADO
      DENOMINA = NINOS(I4,9) + DENOMINA

      I4=I4 +1

 1090 CONTINUE


      ESTIMO = NUMERADO / DENOMINA

      IF (ESTIMO.EQ.P)  THEN
        COMPLEM(R1,1) =C
        COMPLEM(R1,3) =ESTIMO
        GO TO 1800
      ELSE  IF (ESTIMO.LT.P) THEN
               IF (MEPASE.EQ.1) THEN
               COMPLEM(R1,1)= C+0.1
               COMPLEM(R1,3) = ESTANT
               GO TO 1120
                                ELSE
               C=C+0.1
               GO TO 1080
                              END IF
           ELSE
              MEPASE=1
              ESTANT = ESTIMO
              C=C - 0.1
              GO TO 1080
           END IF
```

```
1120 COMPLEM(R1,1)=C+ 0.1/(ESTANT-ESTIMO)*(P-ESTIMO)


1800    CONTINUE


        SPERC = 0
        SPER= 0
        V1 = 0
        V2=0
        V1C =0
        V2C =0
        DO 150  I=1, R
        SPER = PERCENTI(I,1) + SPER
        SPERC = COMPLEM(I,1) + SPERC
150    CONTINUE
        PERMED = SPER/R
        PERMEDC = SPERC/R

     DO 160 I=1,R
     J= F+(31*(INUT - 5))
     PERCENTI(I,2)=(PERCENTI(I,1)-MEDFORMU(J,2))**2
     COMPLEM(I,2) = (COMPLEM(I,1) - MEDFORMU(J,2))**2
     PERCENTI(I,4)=(PERCENTI(I,1) -PERMED)**2
     COMPLEM(I,4) = (COMPLEM(I,1) - PERMED)**2
         V1=V1 + PERCENTI(I,2)
         V2=V2+PERCENTI(I,4)
         V1C=V1C +COMPLEM(I,2)
         V2C = V2C+ COMPLEM(I,4)
 160     CONTINUE
         V1 = V1/ (R* (0.5)**2)
         V2=V2/  (R* (0.5)**2)
         V1C = V1C/ (R* (0.5)**2)
         V2C=V2C/  (R* (0.5)**2)
         V1M =(V1 + V1C)/2
         V2M = (V2 +V2C)/2
         WRITE(*,*) V1M

C


C        SE ESCRIBE UN ARCHIVO CON LAS REPLICAS


         WRITE(4,'(A)') '                    '
```

```
      WRITE(4,'(I4\)') S
      WRITE(4,'(A20\)' )  NOMBREs(F)
      WRITE(4,500 ) PERCENTI(1,1), PERCENTI(1,3)
      DO 510 I=2, R
      WRITE(4,505) PERCENTI(I,1), PERCENTI(I,3)
500   FORMAT( F9.4,4X,E12.6 )
505   FORMAT(24X, F9.4, 4X, E12.6)
510   CONTINUE
      WRITE(4,'(A)') '                    '
C    SE ESCRIBE UN ARCHIVO CON EL  ESTIMADOR Y SU VARIANZA


      WRITE(6,'(I4\)') S
      WRITE(6,'(A20\)' )  NOMBREs(F)
      WRITE(6,515) PERMED, V1,V2,V1C, V2C, V1M, V2M
515   FORMAT(6X, F9.3\,4X, 6(2X,F9.6))

     DO 170 I=1,  R
     PERCENTI(I,1) =0
     PERCENTI(I,2) = 0
     PERCENTI(I,3) =0
    COMPLEM(I,1) =0
     COMPLEM(I,2) = 0
     COMPLEM(I,3) =0
170     CONTINUE



C EL SIGUIENTE CONTINUE ES PARA EL CICLO  POR ESTADO
900 CONTINUE
    CLOSE  (4, STATUS='KEEP')
    CLOSE  (6, STATUS='KEEP')

    WRITE(*,'(A\)') 'I3 ES:'
    WRITE(*,*)  I3
    WRITE(*,'(A\)') 'I1 E I2 SON:    '
    WRITE(*,*) I1, I2


999    END
```