

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 fue posible inspeccionar la paquetería que se menciona en varias referencias bibliográficas, se 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 este trabajo, 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, principalmente, a base de programación, y esto 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 construidos en base a 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.

¹Ver Capítulo 1.

- 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 ó 3)
- Tanto el estimador del parámetro Θ (lineal o no lineal) y de su varianza, son insesgados.
- 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 Hadamard 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.

²Ver Capítulo 1.

- 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 varianza *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, \dots, y_n , variables aleatorias i.i.d., es asintóticamente 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 esto. (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.

³Ver el anexo C.

⁴Los autores citan una tesis de doctorado de W. Yung, en la Universidad de Carleton.

- 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 esto 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.

4.3 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 estratificado 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 cuatro métodos son pocas. Sin embargo, en este estudio por simulación se vio 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 los 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, $S(X_1, \dots, X_N)$, 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 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 anteriormente, donde compara los resultados del *jackknife*, linealización y varias versiones de *bootstrap*. Considera ocho poblaciones bajo diseños estratificados simples sin reemplazo, con distintos números de estratos y unidades. Se vió que 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 grande (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 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 divide⁵ 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 $k-1$, 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 $\hat{\Theta}^b$ ($b=1, \dots, B$), 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

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

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 *jackknife* 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 $(\frac{1}{n}, \dots, \frac{1}{n})$. 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

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

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, en 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.:

$$\left(\frac{1}{w_{111}^*}, \frac{1}{w_{112}^*}, \dots, \frac{1}{w_{hij}^*}, \dots, \frac{1}{w_{L_n L_n L_n}^*} \right)$$

7. Después de ver lo complicado de tantas aleatorizaciones que requiere el *bootstrap*, surgió la interrogante de cuán robusta y adecuada 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:

$$\left(\frac{w_{111}}{\sum w}, \frac{w_{112}}{\sum w}, \dots, \frac{w_{hij}}{\sum w}, \dots, \frac{w_{L_n L_m L_n L}}{\sum w} \right), \quad (4.1)$$

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

A este vector le es inherente la información del diseño. 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 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 para 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 teniendo 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 ser 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 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.

⁶Ver Rao (1997).

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