Tema 22: Muestreo multietápico

A high intraclass correlation means that the corrected sample size is considerably lower, and this means less precise standard errors! At the end of the day, this makes sense; if observations on a beach are highly correlated, we cannot treat them as independent observations. Why then bother taking many observations per beach? Perhaps we should sample more beaches with fewer observations per beach?
Alain F. Zuur[1]

Peter B. Mandeville

Un marco muestral es una relación completa de las unidades en el universo del estudio. El muestreo estratificado requiere que los elementos en este universo sean divididos en grupos, llamados estratos, antes de empezar el proceso de muestreo. Cada unidad es asignada a un estrato basado en el conocimiento previo de las características de la unidad. Entonces, se seleccionan muestras aleatorias independientes de cada estrato.[2]

Un muestreo polietápico utiliza más que una etapa de selección para formar la muestra. La primera etapa utiliza las unidades de muestreo de mayor tamaño que se llaman unidades de muestreo primarios, PSUs *primary sampling units*, mientras que la en etapa final se utilizan las unidades de muestreo más pequeño denominadas unidades de muestreo secundarias, SSUs *secondary sampling units*. Si el objeto es investigar alumnos, entonces los PSUs pueden ser las escuelas y los SSUs los alumnos dentro de las escuelas seleccionadas. Si el objeto es investigar pacientes, entonces los PSUs pueden ser los hospitales, y los SSUs los pacientes dentro de los hospitales seleccionados.

La pregunta es ¿cómo tomar una muestra estratificada cuando no existe un marco muestral de los SSUs? Una estrategia es efectuar un muestreo multietápico, tomando primero una muestra aleatoria de los PSUs y entonces levantar marcos muestrales de cada PSU seleccionado, los cuales se utilizarán para efectuar muestras aleatorias de los SSUs en cada PSU.[2-4]

Cada etapa del muestreo tiene covariables asociadas que permiten la estratificación de la muestra después de que ésta es tomada con respecto a cualquier combinación de las covariables. En algunos casos se pueden estratificar los agrupamientos antes de la primera etapa, por ejemplo, si las escuelas son particulares o públicas.

El tamaño de la muestra está directamente ligado al tipo de análisis, y se puede pensar estos modelos como extensiones de modelos lineales, pero sería una equivocación analizar los datos como si los elementos fueran muestras aleatorias simples. En educación, por ejemplo, los alumnos toman clases en salones que están en escuelas, y el comportamiento de los alumnos en un mismo salón estará correlacionado, como también lo estará el comportamiento de los alumnos en la misma escuela. Se necesitan incluir estas correlaciones en el análisis para efectuar inferencias correctas de la investigación. El análisis debe ser efectuado con modelos mixtos.[5]

Otra característica de datos multietápicos es que los elementos en el mismo agrupamiento son más homogéneos que los de distintos agrupamientos. Generalmente se mide el grado de homogeneidad de los agrupamientos con un coeficiente de correlación intraclase, ICC. Si la correlación es alta, los agrupamientos son homogéneos dentro de cada grupo o muy diferentes de un agrupamiento a otro.[6] En general si el ICC es pequeño, los agrupamientos solamente son un poco diferentes uno de otro. Si la ICC es muy pequeña o igual a cero, no existen diferencias de los agrupamientos para las variables de interés. Un ICC de cero quiere decir que los agrupamientos no tienen consecuencias para la relación entre las variables de interés, por lo que pueden ser ignorados en el análisis. Por suponer una correlación intraclase y modelar esta correlación, la estructura anidada de los datos es tomada en cuenta. El ignorar una correlación intraclase tiene consecuencias para la confiabilidad de los resultados solamente si la correlación es significativa y sustancial.[6] La ICC aplicable en estos casos es la ICC(1) de Bartko[7] y la ICC(1,1) de Shrout and Fleiss.[8,9]

En el modelo de medias no condicionadas, *unconditional means model*, el componente fijo es el intercepto y el componente aleatorio son los interceptos aleatorios.[9] Se utilizó R versión 2.9.2[10] con la función *lme()* del paquete *nlme*[11] para efectuar los cálculos. Los datos *bh1996* están en el paquete *multilevel*.[12]

```
> library(multilevel)
> library(nlme)
> data(bh1996)
>Null.Model<-lme(WBEING~1,random=
~1|GRP,data=bh1996,
  control=list(opt=»optim»))
```

Se utilizó la función *VarCorr()* del paquete *nlme* para efectuar las estimaciones de las varianzas para un objeto *lme*.

```
> VarCorr(Null.Model)
GRP = pdSymm(1)
Variance StdDev
(Intercept) 0.03580079 0.1892110
Residual 0.78949727 0.8885366
```

La estimación de $\tau 00$ (varianza entre-agrupamientos o intercepto) es 0.035, y el estimado de σ^2 (varianza dentro-agrupamiento o residuo) es 0.789.

```
> 0.03580079/(0.03580079+0.78949727)
[1] 0.04337922
```

La estimación de la ICC es 0.04, $ICC = \frac{\tau 00}{\tau 00 + \sigma^2}$, la proporción de la varianza, se debe al nivel de agrupamiento.[13]

Un concepto útil es el efecto del diseño, deff *design effect*, que indica cómo el diseño seleccionado, en este caso el diseño multietápico, afecta al error estándar de los parámetros.[14]

En una muestra aleatoria simple, el error estándar de la media aritmética está relacionado con el tamaño de la muestra, por la fórmula

$$\text{error estándar} = \frac{\text{desviación estándar}}{\sqrt{\text{tamaño de la muestra}}}$$

Se puede utilizar esta fórmula para indicar el tamaño de la muestra requerida en una muestra aleatoria simple, si se desea un error estándar dado.[13]

Cuando se utilizan muestras de dos etapas, es necesario tomar en cuenta los agrupamientos de los datos en la determinación del tamaño de la muestra. El deff indica por cuánto se debe ajustar el tamaño de la muestra en el denominador de la ecuación anterior, por razón del diseño utilizado.[13] Está definido como:

$$\text{deff} = \frac{\text{Error estándar con este diseño}}{\text{Error estándar con el diseño estándar}}$$

con el diseño estándar definido como el diseño utilizando una muestra simple aleatoria con el mismo tamaño de la muestra al nivel uno.[14] El deff de una muestra de dos etapas con agrupamientos del mismo tamaño está dado por $deff = 1 + (n-1)\rho_I$. Esta fórmula indica que una muestra de dos etapas es menos atractiva cuando ρ_I (**ICC**) y el tamaño de los agrupamientos aumentan.[13]

En el ejemplo de Snijders y Bosker[13] se supone que se está estudiando el grado de la satisfacción de los pacientes con los tratamientos de sus médicos. Además, se supone que algunos médicos tienen más pacientes satisfechos que otros, resultando en una ICC de 0.30. Los investigadores utilizarán un muestreo de dos etapas dado que esto es más barato que seleccionar los pacientes al azar. Primero se seleccionaron aleatoriamente 100 médicos y de cada médico se seleccionaron cinco pacientes al azar, los cuales fueron entrevistados. En este caso el deff es $1 + (5-1) \times 0.30 = 2.20$. Cuando se estima el error estándar de la media aritmética, no se puede suponer que las observaciones son independientes y el tamaño efectivo de la muestra que se debe utilizar para estimar el error estándar es igual a $N_{effective} = \frac{Nn}{deff}$, donde N es el número de agrupamientos seleccionados. Para el ejemplo se encuentra que $N_{effective} = \frac{100 \times 5}{2.20} = 227$, por lo cual la muestra de dos etapas con un total de 500 pacientes es equivalente a una muestra aleatoria simple de 227 pacientes.

También se puede derivar el tamaño de la muestra total con un diseño de dos etapas para un nivel de precisión deseado, suponiendo que la ICC es conocida y *n* fijo por consideraciones de presupuesto o tiempo. La regla general es: el tamaño de la muestra requerido aumenta cuando la ICC aumenta y también cuando aumenta el número de elementos que se desea seleccionar de cada agrupamiento. Numéricamente $N_{ts} = N_{srs} + N_{srs}(n-1)\rho_I$, donde N_{ts} refiere al tamaño total de la muestra cuando se utiliza una muestra de dos etapas, mientras que N_{ts} refiere al tamaño de la muestra si fue utilizada una muestra aleatoria simple.[13]

En la práctica, la ICC es desconocida, pero muchas veces es posible estimarla con base en investigaciones anteriores,[13] de no ser así es preciso un estudio piloto.

Referencias

1. Alain F. Zuur, Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev, and Graham M. Smith. 2009. Mixed Effects Models and Extensions in Ecology with R. Statistics for Biology and Health. Springer Science+Business Media, LLC, New York, NY, USA.
2. Gary T. Henry. 1990. Practical Sampling. Applied Social Research Methods Series Volume 21. Sage Publications, Inc., Newbury Park, CA, USA.
3. Floyd J. Fowler, Jr. 2002. Survey Research Methods. Third Edition. Applied Social Research Methods Series Volume 1. Sage Publications, Inc., Newbury Park, CA, USA.
4. Paul S. Levy and Stanley Lemeshow. 1999. Sampling of Populations: Methods and Applications. Third Edition. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, NY, USA.
5. Brady T. West, Kathleen B. Welch, and Andrzej T. Galecki. 2007. Linear Mixed Models: A Practical Guide Using Statistical Software. Chapman and Hall/CRC, Boca Raton, FL, USA.
6. Ita Kreft and Jan de Leeuw. 1998. Introducing Multilevel Modeling. ISM Introducing Statistical Methods. Sage Publications Ltd., London, UK.
7. J.J. Bartko. 1976. On various intraclass correlation reliability coefficients. *Psychological Bulletin, 83*, 762-765.
8. Shrout, P.E., and Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.
9. Paul Bliese. 2009. Multilevel Modeling in R (2.3): A Brief Introduction to R, the multilevel package and the nlme package. URL http://www.R-project.org.
10. R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0,

URL http://www.R-project.org.

11. Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar and the R Core team (2009). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-96. URL http://www.R-project.org.
12. Paul Bliese. 2008. multilevel: Multilevel Functions. R package version 2.3. URL http://www.R-project.org.
13. Tom A.B. Snijders and Roel J. Bosker. 1999. Multilevel Analysis: An introduction to basic and advanced multilevel modeling. Sage Publications Ltd., London, UK.
14. Tom A.B. Snijders. 2005. Power and Sample Size in Multilevel Modelling. Pages 1570-1573 in B. S. Everitt and D.C. Howell (ed). 2005. Encyclopedia of Statistics in Behavioral Science. Volume 3. John Wiley & Sons, Ltd., Chichester, UK.