



Ciencia UANL

Universidad Autónoma de Nuevo León

ciencia@mail.uanl.mx

ISSN (Versión impresa): 1405-9177

MÉXICO

2006

Peter B. Mandeville

TEMA 10: TAMAÑO DE LA MUESTRA V

*Ciencia UANL*, enero-marzo, año/vol. IX, número 001

Universidad Autónoma de Nuevo León

Monterrey, México

pp. 100-103

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal

Universidad Autónoma del Estado de México





PETER B. MANDEVILLE

*To call in the statistician after the experiment is done may be no more than asking him to perform a postmortem examination: he may be able to say what the experiment died of.*

Sir Ronald A. Fisher

Recientemente un alumno me preguntó cuántas repeticiones se necesitan para una regresión logística múltiple en un estudio de la permeabilidad del foramen oval en niños. La probabilidad de que cierre durante el primer año de vida es igual a 0.75, por lo cual la probabilidad de que no cierre es igual a 0.25.\*

Hay por lo menos dos enfoques para el cálculo del tamaño de la muestra para regresión logística múltiple.

El primero se basa en el análisis de potencia.<sup>1</sup> Hsieh<sup>2</sup> presenta tablas que extienden el uso de la fórmula de Whittemore,<sup>3</sup> la cual se basa en la distribución muestral de la estadística de Wald.<sup>4</sup> Hsieh *et al.*<sup>5</sup> sugieren que se utilice la fórmula para comparar proporciones ajustando por el factor de inflación de la varianza (VIF) para estimar el número de repeticiones para regresión logística múltiple. Hosmer y Lemeshow<sup>4</sup> utili-

zan modificaciones de Whittemore<sup>3</sup> y Hsieh<sup>2</sup> para estimar el tamaño de la muestra para regresión logística múltiple. Agresti<sup>6</sup> recomienda el uso de Hsieh *et al.*,<sup>5</sup> y Whittemore<sup>3</sup> como una guía con muchas limitaciones para estimar el tamaño de la muestra para regresión logística múltiple. PASS 6.0<sup>7</sup> recurre a las fórmulas de Hsieh<sup>2</sup> y Whittemore<sup>3</sup> con el mismo fin.

El uso de estas fórmulas requiere datos de estudios previos o un estudio piloto. No existen estudios previos y un estudio piloto no es factible, dado que se requeriría por lo menos un año para llevarse a cabo.

El segundo enfoque está basado en la simulación Monte Carlo.<sup>8</sup> La metodología utilizada está descrita en Concato *et al.*<sup>9</sup> y Concato *et al.*<sup>10</sup> La motivación se describe en Harrell *et al.*<sup>11</sup> y Concato *et al.*<sup>12</sup> Este enfoque está basado en la confiabilidad de los coeficientes de regresión, y no en el análisis de potencia. Como tal, se utiliza el sesgo, precisión y significancia de los coeficientes de regre-

sión y presencia de asociaciones falsas. Tres diferentes tipos de errores fueron discutidos.

Error de tipo 1. Sobreajuste, *overfitting*, sucede cuando demasiadas variables se incluyen en el modelo final.

Error de tipo 2. Bajo ajuste, *underfitting*, sucede cuando variables importantes están omitidas del modelo final.

Error de tipo 3. Ajuste paradójico, *paradoxical fitting*, sucede cuando una variable tiene una dirección de asociación opuesta a la dirección verdadera.

Peduzzi *et al.*<sup>8</sup> mencionan que cuando el número de eventos por variable (EPV) son de 10 o más, no hubo grandes problemas con respecto a los criterios mencionados anteriormente.

Con base en estos estudios, varias autoridades<sup>4, 13, 35</sup> recomiendan el uso de un EPV mínimo de 10 para regresión logística múltiple. Van Belle<sup>20</sup> dice que todas las recomendaciones de un EPV de 10 están ba-

\* Agradezco al Dr. Mauricio Pierdant Pérez, quien propuso la pregunta.

sadas en el estudio de Peduzzi *et al.*<sup>8</sup> Vittinghoff *et al.*<sup>14</sup> notan que esta regla puede ser muy problemática. Katz<sup>13</sup> reconoce que se pueden necesitar un EPV de más de 10. Hosmer y Lemeshow<sup>4</sup> dice que Peduzzi *et al.*<sup>8</sup> demuestran que se necesita un EPV mínimo de 10 para evitar problemas con los límites de confianza de Wald y las pruebas de significancia de Wald.

Algunos individuos han interpretado EPV como el número de variables explicativas utilizadas en la regresión logística. Están equivocados porque EPV se refiere al número de términos que serán incluidas en la regresión. Peduzzi *et al.*<sup>8</sup> solamente utilizaron factores dicotómicos y covariables que tienen 1 grado de libertad cada uno y contribuye un término en la regresión. Si un factor tiene más que dos niveles, entonces los grados de libertad (gl) son iguales al número de niveles del factor menos uno y la regresión tendrá gl términos.

Desgraciadamente esto ha sido interpretado por algunos investigadores médicos como una recomendación para un EPV de 10, lo que ha resultado en estudios con poca potencia e investigadores tristes.

De acuerdo con Peduzzi *et al.*,<sup>8</sup> se estima el número de repeticiones para regresión logística múltiple con:

$$n = \frac{Sgl \times EPV}{pm}$$

Donde  $n$  es igual al número mínimo de repeticiones,  $Sgl$  es igual a

la suma de los grados de libertad de las variables explicativas para estar incluidas en el modelo final y  $pm$  es la probabilidad menor. Siempre se utiliza el valor menor entre la probabilidad del evento y del complemento.<sup>8, 13</sup>

En una comunicación personal del 5 noviembre 2001, el Dr. Feinstein dijo que:

*As a general rule, my colleagues and I currently believe that EPV of  $\geq 20$  is desirable for logistic, Cox, and linear multiple regression, and that EPV = 10 should be a minimum.*

Harrell<sup>15</sup> dice que un EPV de 15 resultará en un modelo confiable para estimaciones futuras. En una comunicación personal, el 20 septiembre de 2005, el Dr. Harrell dijo que todavía recomienda un EPV de 15 y que el valor fue seleccionado como un compromiso entre los que proponen un EPV de 10 y los que proponen un EPV de 20.

Katz<sup>13</sup> y Agresti<sup>6</sup> dicen que si el cálculo del número de repeticiones para un análisis bivariado demuestra que no tiene suficiente repeticiones para demostrar el efecto, entonces definitivamente no tiene suficientes repeticiones para un análisis multivariable. Ellos sugieren que se empiece el proceso por determinar el número de repeticiones con el cálculo en el caso bivariado.

En este caso, en el 25% de los nacimientos, el foramen oval no se cierra dentro de un año y en el 75% de los nacimientos el foramen oval

se cierra. La probabilidad menor es 0.25. De acuerdo con el Dr. Feinstein, se utilizan un EPV de 20. Hay un gl en el modelo, por lo cual el número mínimo de repeticiones es igual a

$$n = \frac{1 \times 20}{0.25} = 80$$

Dado que esto es lo equivalente a probar dos proporciones con el diseño de dos muestras independientes al azar, se puede utilizar la función *power.prop.test* en R versión 2.1.1<sup>16</sup> para estimar la precisión,  $\delta$ , que tendrá con una potencia de 0.8.

```
>
power.prop.test(n=40,p1=0.25,power=0.8)
```

Two-sample comparison of proportions power calculation

```
n = 40
p1 = 0.25
p2 = 0.5525606
sig.level = 0.05
power = 0.8
alternative = two.sided
```

Note: n is number in \*each\* group

El tamaño del efecto  $\delta$  será de  $0.5526 - 0.25 = 0.3026$ , que quiere decir que tendrá la capacidad de detectar diferencias en las proporciones de 0.3026 o más como significativas, mientras diferencias menores que 0.3026 no serán significativas.

Se repite el proceso con 10 repeticiones por gl por EPV.

>

```
power.prop.test(n=20,p1=0.25,power=0.8)
```

Two-sample comparison of proportions power calculation

```
n = 20
p1 = 0.25
p2 = 0.6789486
sig.level = 0.05
power = 0.8
alternative = two.sided
```

Note: n is number in \*each\* group

El tamaño del efecto  $\delta$  será de  $0.6789-0.25=0.4289$ , lo cual quiere decir que tendrá la capacidad de detectar diferencias en las proporciones de 0.4289 o más como significativas, mientras diferencias menores que 0.4289 no serán significativas.

Para proporciones, la capacidad de detectar una diferencia de 20% está clasificada como la capacidad de detectar una pequeña diferencia, la capacidad de detectar una diferencia de 50% está clasificada como la capacidad de detectar una diferencia mediana y la capacidad de detectar una diferencia de 80% está clasificada como la capacidad de detectar una diferencia grande.<sup>1, 17,18</sup>

Machin y Campbell<sup>19</sup> sugieren que estos valores son aplicables a estudios médicos.

Parece que el número de repeticiones para el cumplimiento de la confiabilidad satisfactoria de los coeficientes de regresión para pruebas de Wald no son congruentes con el número de repeticiones de un análisis de potencia. La recomendación de un EPV de 10 permite la detección de diferencias de entre las proporciones de 43%, mientras que la recomendación de un EPV de 20 permite detectar diferencias entre las proporciones de 30%, cuando la probabilidad del evento es de 0.25.

Debe hacerse notar que habrá diferentes valores de  $\delta$  para diferentes probabilidades del evento.

## Referencias

1. Jacob Cohen (1988). *Statistical Power Analysis for the Behavioral Sciences*. Second edition. Lawrence Erlbaum Associates, Publishers, Hillsdale, NY, USA.
2. F. Y. Hsieh (1989). *Sample Size Tables for Logistic Regression*. *Statistics in Medicine*, Vol. 8, 795-802.
3. A. Whittemore (1981). *Sample size for logistic regression with small response probability*. *Journal of the American Statistical Association*, 76, 27-32.
4. David W. Hosmer y Stanley Lemeshow (2000). *Applied*

*Logistic Regression*. Second Edition. *Wiley Series in Probability and Statistics*. John Wiley & Sons, Inc., New York, NY, USA.

5. F. Y. Hsieh, Daniel A. Bloch y Michael D. Larsen (1998). *A Simple Method of Sample Size Calculation for Linear and Logistic Regression*. *Statistics in Medicine* 17, 1623-1634.
6. Alan Agresti (2002). *Categorical Data Analysis*. Second edition. *Wiley Series in Probability and Statistics*. John Wiley & Sons, Inc., New York, NY, USA.
7. Jerry L. Hintze (1996). *PASS 6.0 User's Guide*. Kaysville, Utah. USA.
8. Peter Peduzzi, John Concato, Elizabeth Kemper, Theodore R. Holford y Alvan R. Feinstein. (1996). *A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis*. *J Clin Epidemiol* Vol. 49, No. 12, pp. 1373-1379.
9. John Concato, Peter Peduzzi, Theodore R. Holford y Alvan R. Feinstein (1995). *Importance of Events per Independent Variable in Proportional Hazards Analysis I. Background, Goals, and General Strategy*. *J Clin Epidemiol* Vol. 48, No. 12, pp. 1495-1501.
10. John Concato, Peter Peduzzi, Alvan R. Feinstein y Theodore R. Holford (1996). *Importance of Events per Independent Variable in Proportional Hazards Regression Analysis II. Accuracy*

- and Precision of Regression Estimates. *J Clin Epidemiol* Vol. 48, No. 12, pp. 1503-1510.
11. Frank E. Harrell, Jr, Kerry L. Lee, David B. Matchar y Thomas A. Reichert. (1985). *Regression Models for Prognostic Prediction: Advantages, Problems, and Suggested Solutions*. *Cancer Treat Rep* 69:1071-1077, 1985.
  12. John Concato, Alvan R. Feinstein y Theodore R. Holford (1993). The Risk of Determining Risk with Multivariable Models. *Annals of Internal Medicine* 1993; 119:201-210.
  13. Mitchell H. Katz (1999). *Multivariable Analysis: A Practical Guide for Clinicians*. Cambridge University Press, Cambridge, UK.
  14. Eric Vittinghoff, David V. Glidden, Stephen C. Shiboski y Charles E. McCulloch. (2005). *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. Statistics for Biology and Health. Springer Science+Business Media, Inc., New York, NY, USA.
  15. Frank E. Harrell, Jr. (2001b). Re: [S] model fitting with logistic regression. S-news@lists.biostat.wustl.edu 6 Apr 2001.
  16. R Development Core Team (2000). *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>.
  17. Kevin R. Murphy y Brett Myers (1998). *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*. Erlbaum Associates, Publishers, Mahwah, NJ, USA.
  18. Kevin R. Murphy y Brett Myers (2004). *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*. Second edition. Lawrence Erlbaum Associates, Publishers, Mahway, NJ.
  19. David Machin y Michael J. Campbell (2005). *Design of Studies for Medical Research*. John Wiley & Sons, Ltd., Chichester, UK.
  20. Gerald van Belle (2002). *Statistical Rules of Thumb*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, NY, USA.