ESTUDIO DEL COMPORTAMIENTO DE TEST TIPO SCORE Y WALD PARA DATOS BINARIOS CORRELACIONADOS*

Hachuel, L.; Boggio, G.; Wojdyla, D.; Cuesta, C.

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística de la Facultad de Ciencias Económicas y Estadística. Universidad Nacional de Rosario.

E mail: lhachuel@agatha.unr.edu.ar

OBJETIVO

Estudiar el comportamiento de estadísticas tipo score y Wald para datos binarios correlacionados a través de la evaluación del control del error de tipo I y II mediante un estudio de simulación.

METODOLOGIA

▪ ESTADÍSTICAS

X^2_{cs}: Cuasi-score
X^2_{SG}: Score generalizada
X^2_{WG}: Wald generalizada
X^2_{WGc}: Wald generalizada corregida

} se calculan bajo la estructura de correlación de independencia y AR(1)

▪ ALGORITMO DE GENERACIÓN DE DATOS

Algoritmo de generación de datos binarios correlacionados
(Servy, Hachuel y Wojdyla)

Este modelo fue diseñado originalmente para generar muestras de conglomerados cuyos elementos son pares de valores de dos variables categóricas, de forma tal que finalmente la muestra puede presentarse bajo la forma de una tabla de contingencia bivariada. Los conglomerados se generan por los k pasos de una cadena de Markov. Si se ignora una de las variables, dichos conglomerados se transforman en univariados y si esa variable es binaria, el algoritmo se puede utilizar para generar muestras de conglomerados univariados o de vectores de respuestas binarias correlacionadas. El modelo fija inicialmente el valor de la probabilidad de respuesta igual a 1, π, el tamaño k del conglomerado y especifica la matriz de transición M de la cadena de Markov. A partir de estos valores, se determina el vector de probabilidades iniciales resolviendo un sistema de ecuaciones. Es posible determinar, una vez especificados dichos parámetros las correlaciones entre las respuestas en pares de posiciones diferentes, ρ_{vj}.

▪ ESTUDIO DE SIMULACIÓN

El comportamiento de las estadísticas se estudia para el caso particular:

$$\ln\{\pi/(1-\pi)\} = \beta_1 + \beta_2 X \qquad X=0,1$$

Se ajusta el modelo a datos generados por el algoritmo de Servy et al.
Se eligen escenarios con esquemas de correlación semejante pero con probabilidad de respuesta igual a 1, π, baja, media y alta.
Se simulan muestras simples aleatorias de n=15, 30, 50, 70 y 100 conglomerados de tamaño fijo k=3.

Escenarios paramétricos

Escenario	π	$R=(\rho_{12};\rho_{13};\rho_{23})$
1	0.3	(0.76, 0.59, 0.77)
2	0.5	(0.80, 0.64, 0.80)
3	0.8	(0.67, 0.49, 0.73)

RESULTADOS

Nivel de significación real

Escenario 1

N	Estadísticas						
	X^2_{cs}	$X^2_{SG(ind)}$	$X^2_{WG(ind)}$	$X^2_{WGc(ind)}$	$X^2_{SG(AR1)}$	$X^2_{WG(AR1)}$	$X^2_{WGc(AR1)}$
15	7.70	3.00	4.60	2.90	2.70	4.40	2.90
30	6.50	5.40	5.10	4.40	4.70	4.70	4.00
50	6.60	6.00	5.70	5.00	6.20	6.00	5.20
70	6.40	5.90	5.80	5.00	5.40	5.50	4.80
100	6.50	6.00	6.00	5.60	6.10	6.30	6.00

Escenario 2

N	Estadísticas						
	X^2_{cs}	$X^2_{SG(ind)}$	$X^2_{WG(ind)}$	$X^2_{WGc(ind)}$	$X^2_{SG(AR1)}$	$X^2_{WG(AR1)}$	$X^2_{WGc(AR1)}$
15	9.70	5.70	6.40	4.20	5.70	6.40	4.30
30	6.30	5.00	5.00	3.70	5.40	5.10	4.50
50	5.80	5.00	4.50	3.00	4.70	4.50	3.60
70	4.90	4.50	4.20	4.00	4.20	4.30	3.90
100	5.10	5.10	4.80	4.70	4.60	4.60	4.60

Escenario 3

N	Estadísticas						
	X^2_{cs}	$X^2_{SG(ind)}$	$X^2_{WG(ind)}$	$X^2_{WGc(ind)}$	$X^2_{SG(AR1)}$	$X^2_{WG(AR1)}$	$X^2_{WGc(AR1)}$
15	7.10	1.60	4.40	2.40	1.10	3.50	1.90
30	6.30	4.70	5.20	4.10	4.20	5.80	4.20
50	6.50	6.20	5.90	5.40	5.50	5.80	5.50
70	5.30	5.10	4.90	4.60	480	4.60	4.30
100	6.20	6.00	5.80	5.60	6.10	5.80	5.70

- Estadística cuasi-score: comportamiento liberal para cualquier tamaño de la muestra.
- Estadística de score y Wald generalizadas: conservadoras para muestras muy pequeñas (escenarios 1 y 3)
- Estadística de Wald generalizada corregida: no se detecta la necesidad de la correción.

Potencias empíricas para diferentes alternativas

β_2=0.85 (Escenario 1 vs. 2)

N	Estadísticas						
	X^2_{cs}	$X^2_{SG(ind)}$	$X^2_{WG(ind)}$	$X^2_{WGc(ind)}$	$X^2_{SG(AR1)}$	$X^2_{WG(AR1)}$	$X^2_{WGc(AR1)}$
15	19.00	11.50	13.40	9.60	12.50	14.70	10.50
30	24.40	21.10	20.30	18.10	21.60	21.20	18.40
50	37.00	34.60	34.60	32.80	36.00	35.80	34.40
70	48.30	46.70	46.60	45.40	48.00	48.50	47.40
100	62.40	61.40	61.50	60.40	63.00	62.80	61.90

β_2=2.23 (Escenario 1 vs. 3)

N	Estadísticas						
	X^2_{cs}	$X^2_{SG(ind)}$	$X^2_{WG(ind)}$	$X^2_{WGc(ind)}$	$X^2_{SG(AR1)}$	$X^2_{WG(AR1)}$	$X^2_{WGc(AR1)}$
15	65.00	52.00	54.60	46.40	53.30	57.50	49.20
30	90.40	88.30	88.80	86.60	89.20	89.80	87.70
50	98.60	98.50	98.60	98.10	98.60	98.60	98.60
70	99.90	99.90	99.90	98.90	99.90	99.90	99.90
100	99.90	99.90	99.90	99.90	99.90	99.90	99.90

- Potencias similares para todos los tests.
- Alternativa β_2=0.85: potencias empíricas bajas (63% para n=100)
- Alternativa β_2=2.23: potencias empíricas que alcanzan el 85% para n=30

* Este trabajo se realiza en el marco del Proyecto de Investigación de la Secretaría de Ciencia y Tecnología de la UNR (PID 2001-2003): "Estudio del comportamiento de estadísticas para medidas repetidas en muestras pequeñas bajo escenarios múltiples".