

Aplicación del modelamiento log-lineal para el estudio de interacciones entre factores de

riesgo para cáncer de cuello uterino y la presencia de infección por VPH o anomalías en



la citología cervical



Ledyz Cuesta Herrera, Hugo Alexander Torres Mantilla, José Ángel Vásquez Coronel, Militza Quintero Vega, Rafael Borges Peña

Doctorado en modelamiento matemático aplicado
Universidad Católica del Maule, Talca, Chile

ledyz.cuesta@alu.ucm.cl, al.torres@mail.udes.edu.co,
jose.vasquez@alu.ucm.cl, milyq@hotmail.com, borgesr@gmail.com

1. Introducción

El cáncer de cuello uterino es una de las principales causas de morbilidad en mujeres a nivel mundial. La infección por el virus del papiloma humano (VPH) se ha identificado como uno de los factores etiológicos más importantes. La detección precoz de las neoplasias intraepiteliales permite realizar intervenciones para reducir la progresión a cáncer. La citología cervicouterina es la principal herramienta para la pesquisa de estas lesiones y del cáncer.

El presente trabajo de corte transversal estudió los factores asociados con la presencia de anomalías en la citología y con la infección por VPH confirmada por métodos moleculares en una muestra no probabilística de 1639 mujeres del municipio Santos Marquina del estado Mérida en Venezuela. Se les realizó citología, colposcopia, PCR para VPH y una encuesta indagando antecedentes reproductivos, sexuales y de exposición a carcinógenos, previo consentimiento informado. Para el modelamiento se tuvieron en cuenta 1114 casos válidos con datos completos.

2. El Modelo

Un modelo log lineal es un modelo lineal y aditivo que se obtiene al aplicar el logaritmo natural a las frecuencias esperadas de una tabla de contingencia, descomponiendo este logaritmo en partes atribuibles a los efectos filas y columnas como efectos principales, y las interacciones filas y columnas teniendo en cuenta que dichas interacciones son nulas si se cumple la hipótesis de independencia [1]. En este estudio tuvo en cuenta el modelo saturado, dado que en él no se impone ninguna restricción en las asociaciones entre las categorías para una tabla de contingencia de dos atributos, teniendo tantos grados de libertad como celdas tiene la tabla de contingencia. Por ejemplo, en el caso del modelo general para tres variables categóricas X, Y, Z, también conocido como modelo log lineal saturado se tiene:

$$\log(m_{ijk}) = \mu + \gamma_i^x + \gamma_j^y + \gamma_k^z + \gamma_{ij}^{xy} + \gamma_{ik}^{xz} + \gamma_{jk}^{yz} + \gamma_{ijk}^{xyz} \quad (1)$$

Se ajustaron dos modelos log-lineales por separado para las variables dicotomizadas de resultado Citología y VPH-PCR, usando así mismo covariables dicotómicas. Se realizó una selección por pasos hacia atrás con una probabilidad de eliminación del 5% partiendo desde un modelo saturado que incluye todas las interacciones. Para el modelado de la variable Citología, se consideraron como covariables el antecedente de más de un compañero sexual, el antecedente de gestación previa y el hábito de fumar (pese a no ser significativo).

Con respecto al modelo que incluye la VPH-PCR junto con las covariables significativamente asociadas en el bivariable, precocidad sexual, zona de residencia, antecedente de más de una pareja sexual y edad mayor a 35 años.

3. Resultados

El modelo para la Citología finalmente seleccionado no involucra términos mayores a interacciones entre dos variables. Solo las interacciones entre Gestación y Citología;

Promiscuidad y Citología; y Hábito de fumar y Promiscuidad, fueron retenidas en el modelo. El modelo predice adecuadamente la mayoría de los recuentos excepto en el caso de fumadoras, sin gestación previa y sin antecedente de más de una pareja sexual. Cabe mencionar que justo en la subcategoría anterior, con citología anormal, el recuento esperado es menor a 1, siendo 0.095, vulnerando parcialmente de esta manera los supuestos del modelo.

La principal causa de los bajos recuentos esperados son las desproporciones de casos en las distintas variables dicotómicas. Existe una baja frecuencia de fumadoras, además una baja frecuencia de mujeres sin antecedente de gestación y finalmente una baja frecuencia de citologías anormales. Esto permite la presencia de valores esperados demasiado bajos. Aun así el análisis log lineal permite detectar los términos que más influencia tienen en los conteos de celdas de manera conjunta.

En cambio del modelo para la infección por VPH es notable que la cuádruple interacción vista entre precocidad, antecedente de más de una pareja sexual, zona de Residencia y la VPH-PCR fuera retenida en el modelo debido a que su eliminación perjudicaba significativamente el ajuste del modelo resultante. Se puede observar que el efecto de la interacción entre la Zona de Residencia y la VPH-PCR muestra que en la zona centro la frecuencia relativa de VPH-PCR positivos aumenta respecto a la zona rural.

Este aumento se presenta siempre excepto en el subgrupo de paciente con antecedente de más de una pareja sexual y con inicio de vida sexual tardía. Este fenómeno en particular causa que dicha cuádruple interacción sea significativa. Otros términos retenidos por el modelo fueron la interacción entre el antecedente de más de una pareja sexual y la precocidad sexual, y la interacción entre el grupo etario y la precocidad sexual, siendo ambas relaciones claras debido a la naturaleza de las mismas variables.

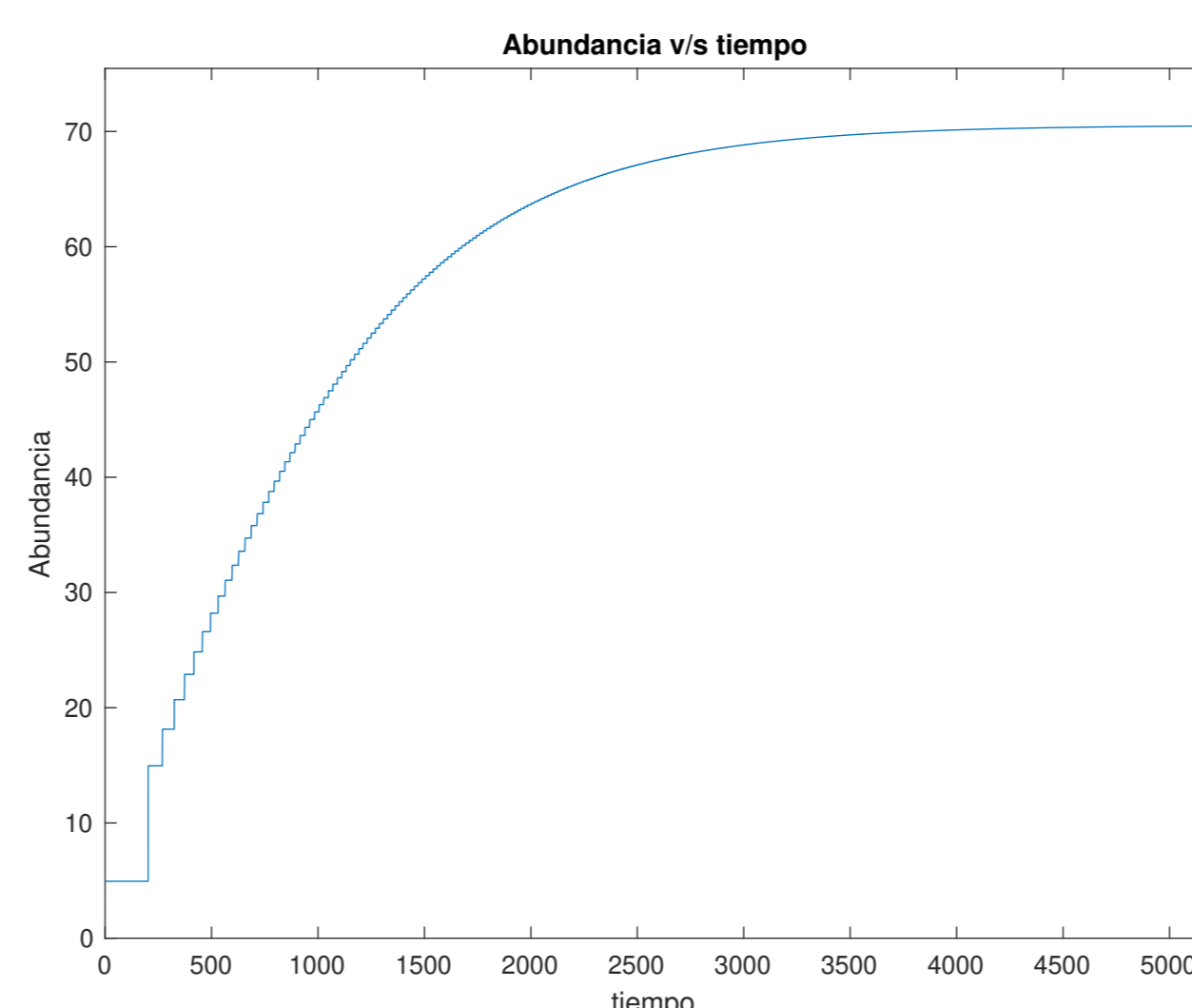


Figura 1: Descripción Figura

4. Conclusiones

Nuestro enfoque establece ... Nuestros resultados indican ...

5. Agradecimientos

Este trabajo es realizado con el soporte ...

$$K_{ce}(G, C) = \sum_{i < j, (i, j) \in E} x_{ij} + \sum_{i < j, (i, j) \notin E} (1 - x_{ij}), \quad (2)$$

$$x_{ij} = \begin{cases} 0, & \text{if } \mathcal{L}_C(i) \cap \mathcal{L}_C(j) \neq \emptyset, \\ 1, & \text{if } \mathcal{L}_C(i) \cap \mathcal{L}_C(j) = \emptyset. \end{cases}$$

Hábito de fumar	Gestación previa	Pareja sexual no exclusiva	Citología anormal	Observado	Esperado	Residuos	Residuos tipificados
No	No	No	No	65 (5.8%)	63.9 (5.7%)	1.145	0.143
			Si	1 (0.1%)	0.9 (0.1%)	0.122	0.118
		Si	60 (5.4%)	59.6 (5.3%)	0.421	0.054	
	Si	No	No	1 (0.1%)	1.6 (0.1%)	-0.639	-0.499
			Si	404 (36.3%)	403.1 (36.2%)	0.918	0.046
		Si	15 (1.3%)	17.2 (1.5%)	-2.175	-0.525	
Si	No	No	No	377 (33.8%)	376.1 (33.8%)	0.905	0.047
			Si	31 (2.8%)	31.7 (2.8%)	-0.687	-0.122
		Si	3 (0.3%)	6.8 (0.6%)	-3.846	-1.470	
	Si	No	No	1 (0.1%)	1 (0.1%)	0.905	2.832
			Si	16 (1.4%)	13.7 (1.2%)	2.280	0.616
		Si	0 (0.0%)	0.4 (0.0%)	-0.377	-0.614	
Si	No	No	45 (4%)	43.2 (3.9%)	1.783	0.271	
		Si	3 (0.3%)	1.8 (0.2%)	1.159	0.854	
	Si	83 (7.5%)	86.6 (7.8%)	-3.606	-0.387		
			Si	9 (0.8%)	7.3 (0.7%)	1.703	0.631

Zona	Grupo etario	Pareja sexual no exclusiva	Precocidad sexual	VPH positivo	Observado	Esperado	Residuos	Residuos tipificados
Rural	≤ 35	No	No	No	91 (7.1%)	100.3 (7.8%)	-9.308	-0.929
				Si	65 (5.1%)	60.0 (4.7%)	4.951	0.639
			Si	23 (1.8%)	28.0 (2.2%)	-4.950	-0.936	
		Si	No	No	112 (8.7%)	112.6 (8.8%)	-0.630	-0.059
				Si	88 (6.9%)	85.7 (5.7%)	2.267	0.245
			Si	22 (1.7%)	21.4 (1.7%)	0.628	0.136	
	> 35	No	No	No	14 (2.8%)	16.3 (1.3%)	-2.268	-0.562
				Si	83 (6.5%)	79.2 (6.2%)	3.842	0.432
			Si	37 (2.7%)	36.0 (2.8%)	0.973	0.162	
		Si	No	No	73 (5.7%)	76.8 (6.0%)	-3.844	-0.439
				Si	34 (2.7%)	35.0 (2.7%)	-0.974	-0.165
			Si	101 (7.9%)	102.5 (8.0%)	-1.453	-0.144	
Urbana	≤ 35	No	No	52 (4.1%)	54.4 (4.2%)	-2.450	-0.332	
			Si	42 (3.3%)	40.5 (3.2%)	1.455	0.228	
		Si	24 (1.9%)	21.5 (1.7%)	2.452	0.528		
	> 35	No	No	47 (3.7%)	47.8 (3.7%)	-0.766	-0.111	
			Si	15 (1.2%)	15.7 (1.2%)	-0.695	-0.175	
		Si	23 (1.8%)	22.2 (1.7%)	0.767	0.163		
Urbana	≤ 35	No	No	8 (0.6%)	7.3 (0.6%)	0.695	0.257	
			Si	73 (5.7%)	68.1 (5.3%)	4.918	0.596	
		Si	27 (2.1%)	27.7 (2.2%)	-0.737	-0.140		
	> 35	No	No	8 (0.6%)	12.9 (1.0%)	-4.919	-1.368	
			Si	6 (0.5%)	5.3 (0.4%)	0.737	0.321	
		Si	31 (2.4%)	31.0 (2.4%)	0.047	0.008		
> 35	No	No	4 (0.3%)	3 (0.2%)	0.955	0.548		
		Si	30 (2.3%)	30.0 (2.3%)	-0.048	-0.009		
	Si	2 (0.2%)	3 (0.2%)	-0.956	-0.556			
> 35	No	No	46 (3.6%)	47.3 (3.7%)	-1.286	-0.187		
		Si	18 (1.4%)	18.6 (1.5%)	-0.628	-0.145		
	Si	20 (1.6%)	18.7 (1.5%)	1.287	0.267			
			Si	8 (0.6%)	7.4 (0.6%)	0.628	0.231	

	K	gl	Razón de verosimilitudes		Pearson		Iteraciones
			χ^2	p-value	χ^2	p-value	
Efectos de orden K superiores	1	15	2414.1	0.000	3534.2	0.000	0
	2	11	38.9	0.000	36.8	0.000	2
	3	5	7.0	0.219	9.9	0.078	3
	4	1	2.1	0.150	1.8	0.185	2
Efectos de orden K	1	4	2375.5	0.000	3497.4	0.000	0
	2	6	31.6	0.000	26.9	0.000	0
	3	4	5.0	0.292	8.2	0.086	0
	4	1	2.1	0.150	1.8	0.185	0

	K	gl	Razón de verosimilitudes		Pearson		Iteraciones
			χ^2	p-value	χ^2	p-value	
Efectos de orden K superiores	1	31	738.6	0.000	739.7	0.000	0
	2	26	142.0	0.000	136.6	0.000	2
	3	16	15.2	0.509	14.1	0.593	4
	4	6	7.3	0.296	6.9	0.329	4
	5	1	0.1	0.809	0.1	0.810	3
	5	1	0.1	0.809	0.1	0.810	3
Efectos de orden K	1	5	596.6	0.000	603.1	0.000	0
	2	10	126.8	0.000	122.5	0.000	0
	3	10	7.9	0.635	7.2	0.710	0
	4	5	7.2	0.205	6.9	0.232	0
	5	1	0.1	0.809	0.1	0.810	0
	5	1	0.1	0.809	0.1	0.810	0

Referencias

- [1] Narvaja Zarate, J., Chue Gallardo, J. (1997). Análisis de datos categóricos usando modelos loglineales. Anales Científicos UNALM, 1-14
- [2] Boggs, C. (1992). Resource Allocation: Exploring Connections between Foraging and Life History. Functional Ecology 6:508-518.