

XI CONGRESO LATINOAMERICANO DE BIOLOGÍA MATEMÁTICA

SoLaBIMa 2019

Universidad Católica del Maule

22-25 de Octubre de 2019, Talca, Chile



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*

Universidad Católica del Maule, Talca, Chile

Resumen

Se estudiaron las asociaciones entre los factores de riesgo para cáncer de cuello uterino y la presencia de infección por VPH de hallazgos anormales en la citología cervical en un municipio del estado Mérida, Venezuela. Se utilizó un análisis mediante un modelo log-lineal a 1114 casos válidos de mujeres participantes en un plan pesquisa de VPH y cáncer cervical. Se realizó una selección por pasos hacia atrás con una probabilidad de eliminación del 5% hasta llegar al mejor modelo jerárquico partiendo de las covariables que fueron significativas en un análisis bivariado previo. Para el modelado de la variable Citología, se consideraron como covariables número de compañeros sexuales, antecedente de gestación previa y hábito de fumar. Solo las interacciones entre Gestación y Citología, Promiscuidad y Citología, Hábito de fumar y Promiscuidad fueron retenidas en el modelo. Para el modelado de la VPH-PCR se utilizaron las covariables precocidad sexual, zona de residencia, Promiscuidad y edad mayor a 35 años. La cuádruple interacción entre Precocidad sexual, Promiscuidad, Zona de Residencia y la VPH-PCR fue retenida en el modelo. Dentro de las conclusiones se observa que el modelo log-lineal permitió detectar los términos que más influencia tienen en los conteos de celdas de manera conjunta así como las interacciones significativas.

Trabajo realizado en conjunto con:

Hugo Alexander Torres Mantilla¹, Universidad de Santander, Colombia.

José Angel Vásquez Coronel², Universidad Católica del Maule, Talca, Chile.

Militza Quintero Vega³, Universidad de los Andes, Venezuela.

Rafael Borges Peña⁴, Universidad de los Andes, Venezuela.



* e-mail: ledyz.cuesta@alu.ucm.cl

¹ e-mail: al.torres@mail.udes.edu.co

² e-mail: jose.vasquez@alu.ucm.cl

³ e-mail: milyq@hotmail.com

⁴ e-mail: borgesr@gmail.com

1. Introducción

Teniendo en cuenta la importancia que el conocimiento a tiempo de la infección por un tipo de Virus de Papiloma Humano (VPH) tiene en las mujeres para obtener un tratamiento adecuado y unido al hecho de que permanentemente estamos enfrentados a diversos mitos a través de los medios de comunicación, es importante no solo la lectura exacta de los resultados obtenidos después de una consulta, sino identificar las tendencias y variabilidad de los datos, así como detectar los posibles errores conscientes o inconscientes que pueden distorsionar la información presentada, para ello se propone un modelo estadístico que permita solucionar algunas dificultades presentes en las interpretaciones de los resultados después de una citología y las pruebas moleculares.

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 de las variables, en donde se puede establecer la independencia, ya sea conjunta, marginal e independencia condicional. (Narvaja Zarate & Chue Gallardo, 1997)

Con el tipo de modelo log-lineal se modeliza la asociación entre variables categóricas, analizando las asociaciones existentes entre dos o más atributos como las tablas de contingencias bidimensionales y multidimensionales. Para esto se debe tener en cuenta el modelo para tablas de contingencia RxC en el que se considera que los logaritmos de los valores estimados como un modelo lineal donde están presente 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.

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. La complejidad de dicho patrón dependerá no solo de los fenómenos que se consideren sino también de variables que existan. Por ejemplo, en el caso del modelo general para tres variables categóricas X, Y, Z , también conocido como modelo log-lineal saturado se tiene:

$$\log(m_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$$

Donde μ es el efecto medio, con superíndice de una letra representa un efecto principal, dos letras interacción doble y tres letras interacción triple.

Este modelo describe la serie $\{m_{ijk}\}$ teniendo valores positivos, donde cumple que $\sum_{i=1}^I \lambda_i^X = 0$, habiendo $I - 1$ parámetros $\{\lambda_i^X\}$ linealmente independientes, luego para los efectos de interacción se cumple que $\sum_{i=1}^I \lambda_{ij}^{XY} = \sum_{j=1}^J \lambda_{ij}^{XY} = 0$ donde $(I - 1)(J - 1)$ parámetros $\{\lambda_{ij}^{XY}\}$ son linealmente independientes de manera análoga para los otros efectos de interacción doble.

Finalmente para la interacción triple se cumple que:

$$\sum_{i=1}^I \lambda_{ijk}^{XYZ} = \sum_{j=1}^J \lambda_{ijk}^{XYZ} = \sum_{k=1}^K \lambda_{ijk}^{XYZ} = 0$$

El total de número de parámetros linealmente independientes (incluyendo m) es:

$$1 + (I - 1) + (J - 1) + (K - 1) + (I - 1)(K - 1) + (J - 1)(K - 1) + (I - 1)(J - 1)(K - 1) = IJK$$

Los datos de una tabla de contingencia bajo el análisis log-lineal también pueden ser modelados como casos particulares de regresión logística, que surge como una aplicación del modelo lineal generalizado. Sin embargo, cabe destacar que los modelos log-lineales para tablas de contingencia modelizan la asociación entre variables categóricas, mientras que los modelos de regresión logística describen como una variable categórica de tipo respuesta depende de un grupo de variables explicativas.

2. Resultados

Se ajustó un modelo log-lineal para las variables dicotomizadas de resultado Citología y VPH-PCR; el análisis log-lineal se realizó por separado para las variables Citología y VPH-PCR de forma aislada e involucrando las covariables con asociaciones significativas. Se debe tener en cuenta que el análisis log-lineal no considera a las variables de resultado como variables dependientes en el modelo, dado que se pretende modelar los conteos de celda y todas las variables incluidas son tratadas de igual manera. Se seleccionaron solo las covariables significativamente asociadas con la variable de resultado, o cercanas al nivel de significancia, como entradas para un modelo inicial saturado. Se realizó una selección por pasos hacia atrás con una probabilidad de eliminación del 5%. Se debe tener presente que el procedimiento de pasos hacia atrás solo trabaja con modelos jerárquicos, de este modo primero prueba la interacción de mayor orden y solo en caso de rechazarla, prueba las interacciones de menor orden que involucren a sus componentes.

Modelo log-lineal para la Citología

Para el modelado de la variable Citología, se consideraron como covariables el número de compañeros sexuales, el antecedente de gestación previa y el hábito de fumar (pese a no ser significativo). Se cuenta con un total de 1114 casos válidos.

El modelo seleccionado no involucra términos mayores a interacciones entre dos variables. Solo las interacciones entre Gestación y Citología; Promiscuidad y Citología; 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 en la subcategoría anterior, con citología anormal, el recuento esperado es menor a 1, siendo 0.095, vulnerando parcialmente 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.

Cuadro 1: Frecuencias y residuos de casillas

Hábito de Fumar	Gestaciones Dicotomizadas	Compañeros sexuales Dicotomizados	Resultado de Citología Dicotomizado	Observado		Esperado		Residuos	Residuos Típicos
				Recuento	%	Recuento	%		
No	0	1	Normal	65,000	5,8	63,855	5,7	1,145	0,143
			Anormal	1,000	0,1	0,888	0,1	0,112	0,118
		>1	Normal	60,000	5,4	59,579	5,3	0,421	0,054
			Anormal	1,000	0,1	1,639	0,1	-0,639	-0,499
	≥ 1	1	Normal	404,000	36,3	403,082	36,2	0,918	0,046
			Anormal	15,000	1,3	17,175	1,5	-2,175	-0,525
>1		Normal	377,000	33,8	376,095	33,8	0,905	0,047	
		Anormal	31,000	2,8	31,687	2,8	-0,687	-0,122	
Si	0	1	Normal	3,000	0,3	6,846	0,6	-3,846	-1,470
			Anormal	1,000	0,1	0,095	0,0	0,905	2,932
		>1	Normal	16,000	1,4	13,720	1,2	2,280	0,616
			Anormal	0,000	0,0	0,377	0,0	-0,377	-0,614
	≥ 1	1	Normal	45,000	4,0	43,217	3,9	1,783	0,271
			Anormal	3,000	0,3	1,841	0,2	1,159	0,854
		>1	Normal	83,000	7,5	86,606	7,8	-3,606	-0,387
			Anormal	9,000	0,8	7,297	0,7	1,703	0,631

Modelo log-lineal para la presencia de VPH por PCR

El modelo que incluye la VPH-PCR junto con las covariables significativamente asociadas en análisis bivariado, precocidad sexual, zona de residencia, antecedente de más de una pareja sexual y edad mayor a 35 años. De igual manera que con el algoritmo de selección del modelo en el caso anterior se depuró siguiendo un proceso de eliminación hacia atrás hasta llegar al mejor modelo jerárquico.

Es notable que la cuádruple interacción vista en el cuadro 2 entre Precocidad, Promiscuidad, Zona de Residencia y la VPH-PCR fuera retenida en el modelo debido a que su eliminación perjudicaba significativamente el ajuste del modelo resultante.

Cuadro 2: Cuádruple interacción entre Precocidad, Promiscuidad, Zona de Residencia y VPH-PCR

				Compañeros Sexuales Dicotomizados							
				1				>1			
				Edad Inicio de Vida Sexual				Edad Inicio de Vida Sexual			
				≤ 18 años		> 18 años		≤ 18 años		> 18 años	
Recuento		% del N° de la columna		Recuento		% del N° de la columna		Recuento		% del N° de la columna	
Zona Dicotomizada	Zona Rural	Diagnóstico de VPH por PCR Dicotomizado	Negativo	174	63,0 %	129	69,4 %	213	60,0 %	64	62,7 %
		Positivo	102	37,0 %	57	30,6 %	142	40,0 %	38	37,3 %	
	Zona Centro	Diagnóstico de VPH por PCR Dicotomizado	Negativo	78	80,4 %	53	84,1 %	119	72,6 %	28	66,7 %
		Positivo	102	37,0 %	57	30,6 %	142	40,0 %	38	37,3 %	

3. Conclusiones

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. Al retenerse ese término de orden 4, todos los términos de órdenes inferiores que involucren esas cuatro variables también se retienen al no ser evaluados explícitamente por el método

de eliminación hacia atrás. 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.

Referencias

- [1] ÁLVAREZ HERNÁNDEZ, G., & DELGADO DELAMORA, J. (2015). *Diseño de Estudios Epidemiológicos. I. El estudio Transversal: Tomando una Fotografía de la Salud y la Enfermedad*. Bol Clin Infant Edo Son, 32(1), 26-34.
- [2] BONNEZ, W. (2002). *Papillomavirus*. En R. Richman, R. Whitley, & F. Hayden, *Clinical Virology (Segunda ed., págs. 557-596)*. American Society for Microbiology.
- [3] FIARC WORKING GROUP ON EVALUATION OF CERVICAL CANCER SCREENING PROGRAMMES. (1986). *Screening for squamous cervical cancer: duration of low risk after negative results of cervical cytology*.
- [4] MANTEL, N., & HAENSZEL, W. (1959). *Statistical aspects of the analysis of data from retrospective studies of disease*. J Natl Cancer Inst(22), 719-748.
- [5] NARVAJA ZARATE, J., & CHUE GALLARDO, J. (1997). *Análisis de datos categóricos usando modelos loglineales*. Anales Científicos UNALM, 1-14.
- [6] PAGANO, M., & GAUVREAU, K. (2001). *Fundamentos de Bioestadística (Segunda ed.)*. México, D.F., México: Thomson Learning.
- [7] QUINTERO VEGA, M., CRUZ GÓMEZ, J. F., BASTIDAS, M., MÁRQUEZ, L., & PUIG PONS, J. (MARZO DE 2008). *Detección y tipificación de virus del papiloma humano (VPH) mediante PCR- FFLP*. Revista.
- [8] SOCIEDAD AMERICANA CONTRA EL CÁNCER. (2014). *Datos y Estadísticas sobre el Cáncer entre los Hispano/Latinos 2012-2014*. American Cancer Society Inc, Georgia. Atlanta: American Cancer Society Inc.
- [9] TAMAYO ACEVEDO, L., VALENCIA ARREDONDO, M., ESCOBAR BOTERO, S., & VILLA GIRALDO, M. (2001). *Tendencia de la infección por el Virus del Papiloma humano (VPH) en usuarias del Servicio de Citología del Laboratorio Docente Asistencial de la Escuela de Bacteriología y Laboratorio Clínico de la Universidad de Antioquia. 1993-1998*. Revista de la Facultad de Farmacia.
- [10] ZAMORA PALMA, A., & TERRÉS SPEZIALE, A. (ENERO-MARZO DE 1998). *Infección por virus del papiloma humano en mujeres y hombres mexicanos*. Revista Mexicana de Patología Clínica, 45(1), 1-7.