

# Evaluación de la presencia de confusión en algunos miembros de la familia exponencial

## Evaluating the Presence of Confounding in Some Members of the Exponential Family

L. M. Acosta Avena<sup>a,\*</sup>

J. C. Salazar Uribe<sup>b</sup>

### Resumen

Una técnica recomendada en la literatura, para detectar si una variable es o no de confusión en un modelo de Cox, es aplicada en algunos modelos de la familia exponencial. El objetivo es determinar si esta técnica es válida en los modelos logístico, Poisson y lineal clásico. Una técnica alternativa, basada en los odds, es propuesta para el caso del modelo logístico. Los resultados muestran que las técnicas tienen un comportamiento similar cuando la confusión está presente en el modelo; sin embargo, la técnica propuesta resultó ser más efectiva en los escenarios en que hay ausencia de este fenómeno.

**Palabras clave:** Confusión, Familia exponencial, Modelo de Cox, Regresión.

### Abstract

A technique to detect confounding, which is recommended in the literature to detect confounders in the Cox models, is adapted and applied to some models of the exponential family. The aim is to assess whether this technique is useful in the logistic, Poisson and linear models. An alternative technique, based on the odds, is proposed for the case of the logistic model. The results show that the techniques have a similar behavior when confusion is present in the model; however, the proposed technique proved to be more effective in scenarios in which this phenomenon was absent.

**Key words:** Confounding, Exponential Family, Cox Models, Regression.

---

<sup>a</sup>Universidad de São Paulo.

\*Autor de correspondencia: linamav@ime.usp.br

<sup>b</sup>Universidad Nacional de Colombia, Sede Medellín.

## 1. Introducción

El término inglés “confounding” es usado por los epidemiólogos cuando una covariable está asociada tanto con el resultado de interés y un factor de riesgo principal [7]. La presencia de estas variables puede distorsionar el efecto observado de una exposición sobre un resultado; este efecto se refiere a una aparente relación, cuando la variable de confusión es la causante de la relación; o aparente falta de relación, cuando la variable de confusión atenúa la relación [15]. Por ejemplo, en tablas de contingencia, la asociación que se observa para varios grupos, puede invertir su dirección cuando los datos se combinan en un solo grupo. Esta inversión se conoce como la paradoja de Simpson [9].

Existen en la literatura estadística, varios métodos que permiten evaluar confusión, entre los más comunes se encuentran el riesgo relativo, las razones de odds y las estandarizaciones [1]; otros métodos como aleatorización, restricción y matching son implementados en el diseño de estudio y antes del proceso de recolección de los datos, con el objetivo de controlar este fenómeno [10, 11, 12]. Una discusión sobre el uso de estos métodos se puede ver Mickey [11] y Kamangar [10].

Otros procedimientos empleados para controlar los factores de confusión en los análisis son, la estratificación y los modelos multivariados [13]. Con la estratificación se trata de buscar grupos dentro de los cuales el factor de confusión no varíe y así evaluar la asociación entre la exposición y la enfermedad dentro de cada estrato del confusor, y luego emplear el estimador de Mantel - Haenszel para ajustar los resultados para cada estrato [13], con este método se determina que hay confusión cuando se encuentran diferencias entre los resultados crudos y los ajustados de acuerdo al estrato. Este método se vuelve poco manejable cuando se tiene un gran número de estratos o de covariables, o inclusive un número grande de posibles confusores. En estas situaciones es de gran utilidad implementar entonces modelos multivariados.

Wilson [14], Becher [2], Chao [4], Chen [5], Frank [6], Austin [1], entre otros; estudian los efectos que causa la presencia de estas variables en algunos modelos de regresión. Específicamente estudian algunos aspectos importantes en los modelos tales como la estimación de los coeficientes de regresión y sus ni-

veles de confianza, la probabilidad de error tipo I, los residuales, etc. Sin embargo, son pocos los estudios que se centran en los métodos para identificar variables de confusión en modelos de regresión.

Hosmer [8] proponen una técnica para detectar confusión en un modelo de Cox, la cual en adelante llamaremos *cambio porcentual*, la técnica consiste en comparar los tamaños de los efectos observados en la variable regresora cuando el modelo es ajustado con y sin la variable que es considerada como confusora. Se termina confusión en el modelo cuando existe un cambio superior a 0,20 en los tamaños en los efectos. Otro criterio para detectar confusión se encuentra en Bliss [3], estos autores encuentran que la regla del 10%<sup>1</sup> aplicada a las estimaciones de los odds, proporcionan una técnica simple y de fácil aplicación para determinar confusión en un modelo de regresión logística.

En este artículo es de particular interés estudiar -vía simulación- el comportamiento del criterio del cambio porcentual en los modelos logístico, Poisson y lineal clásico. El objetivo principal de este trabajo es determinar si el criterio es válido para estos modelos. Para esto se asumieron dos variables regresoras de una distribución normal multivariada y en las simulaciones se establecieron diferentes escenarios en los parámetros y tamaños muestrales. Para determinar la eficiencia del criterio en los modelos considerados, se usaron diferentes niveles de referencia.

El desarrollo de este artículo comienza con la sección 2, donde se presenta una descripción de algunos métodos para detectar confusión en modelos de regresión. En la sección 3, se describe el estudio de simulación que se implementó y se discuten los resultados. Finalmente, en la sección 4 se exponen las conclusiones.

## 2. Metodología

Antes de describir el estudio de simulación implementado en este trabajo, se abordan algunas técnicas utilizadas para identificar variables de confusión en modelos de regresión.

<sup>1</sup>Los tamaños de los efectos observados en los modelos de regresión difieren en una magnitud relativa de al menos el 10%

## 2.1. Algunos criterios para detectar confusión en modelos de regresión

Existen diferentes métodos para detectar la presencia de confusión en modelos de regresión. Entre ellos se encuentran, el criterio del cambio porcentual [8] y la regla del 10% [3], las cuales serán abordadas en esta sección, así como también el criterio del cambio relativo del odds, propuesto en este artículo.

### 2.1.1. Criterio de cambio porcentual

Hosmer [8] recomiendan que el porcentaje de cambio en la estimación ajustada se calcula como una medida de la cantidad de ajuste. El estimador del porcentaje de cambio, en general es definido como:

$$\Delta\hat{\beta}\% = \left| \frac{\hat{\theta} - \hat{\beta}}{\hat{\beta}} \right| \quad (1)$$

donde  $\hat{\theta}$  denota el estimador crudo del modelo que no contiene el potencial confusor y  $\hat{\beta}$  denota el estimador ajustado del modelo que incluye el potencial confusor.

A pesar de que no existen reglas en el porcentaje de cambio como medida de confusión, Hosmer [8] sugieren que ésta está presente en el modelo si la proporción  $\Delta\hat{\beta}\%$  es máyor a cualquiera entre [0,15,0,20]. Otros autores [11] recomiendan un nivel convencional de 0,20.

### 2.1.2. Regla del 10% aplicada al odds

Basándose en la regla general del 10%, Bliss [3] evalúan el rendimiento de la regla del 10% aplicada en los odds ratios de un modelo logístico. De esta manera, los autores ajustan dos modelos de regresión logística: uno dado en la ecuación (2) que incluye sólo a la variable predictora ( $E$ ) y otro dado por la ecuación (3) que incluye  $E$  y  $C$ , donde  $C$  es el confusor.

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{\theta}_0 + \hat{\theta}_1 E \quad (2)$$

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{\gamma}_0 + \hat{\gamma}_1 E + \hat{\gamma}_2 C \quad (3)$$

Hay evidencia de confusión en este modelo, cuando la diferencia relativa entre las razones de odds, dada por la expresión (4) es superior al  $\pm 0,10$

$$R = \frac{\exp(\hat{\theta}_1) - \exp(\hat{\gamma}_1)}{\exp(\hat{\theta}_1)} = 1 - \exp(\hat{\gamma}_1 - \hat{\theta}_1) \quad (4)$$

### 2.1.3. Criterio del cambio relativo en el odds

Dada la importancia interpretativa del odds en un modelo de regresión logística, se considera más interesante mirar el efecto que tiene la presencia de una variable de confusión sobre el odds. Así que bajo el mismo enfoque del criterio del cambio porcentual dado por la ecuación (1), se utiliza como medida de confusión el cambio relativo en el odds. Por lo tanto el criterio propuesto está queda dado por la siguiente expresión

$$CP = \left| \frac{\exp\{\theta\} - \exp\{\beta\}}{\exp\{\beta\}} \right| \quad (5)$$

donde  $\exp\{\theta\}$  y  $\exp\{\beta\}$ , denotan los odds del modelo sin ajuste por confusión y con ajuste, respectivamente.

## 2.2. Estudio de simulación

Con el fin de estudiar el comportamiento del criterio del cambio porcentual en modelos lineales generalizados, se lleva a cabo un estudio de simulación en el cuál se consideran tres modelos de la familia exponencial: modelo logístico, modelo Poisson y modelo lineal clásico. En cada modelo, las covariables son generadas de una distribución normal bivariada, con vector de medias cero, y correlaciones  $\rho = -0,8, -0,6, -0,4, -0,2, 0, 0,2, 0,4, 0,6, 0,8$ ; tamaños muestrales  $n = 50, 100, 200, 500$  y 1000. Para generar la dependencia entre las covariables y la variable respuesta se utilizó la función de enlace de cada modelo

$$\text{Modelo logístico: } \frac{\exp\{\beta_0 + \beta_1 x_1 + \beta_2 x_2\}}{1 + \exp\{\beta_0 + \beta_1 x_1 + \beta_2 x_2\}}$$

$$\text{Modelo Poisson: } \ln(\eta) + x_1 \beta_1 + x_2 \beta_2$$

$$\text{Modelo lineal clásico: } \beta_0 + x_1 \beta_1 + x_2 \beta_2$$

por lo tanto se consideraron los siguientes escenarios en los parámetros:

1. En el modelo logístico, los valores conocidos de los parámetros  $\beta_0, \beta_1, \beta_2$  se tomaron de Austin [1], quienes consideran cinco escenarios de acuerdo a un estudio realizado sobre la relación entre el confusor, el factor de riesgo y la probabilidad de observar el resultado de interés. Debido a que en ese estudio, los escenarios 2 y

4 permitan un rango en la probabilidad de la respuesta similar a los escenarios 1 y 5 respectivamente, sólo se asumieron los siguientes

- Escenario 1:  $\beta_0 = 0, \beta_1 = 0, \beta_2 = 3$
- Escenario 2:  $\beta_0 = 0, \beta_1 = 0, \beta_2 = 0,5$
- Escenario 3:  $\beta_0 = -2,9, \beta_1 = 0, \beta_2 = 0,5$

En la figura 1, se observa que con el escenario 1 la probabilidad del resultado permite variar en un rango de muy bajo a muy alto; con el escenario 2 se permiten probabilidades moderadas y, con el escenarios 3 la probabilidad del resultado es baja.

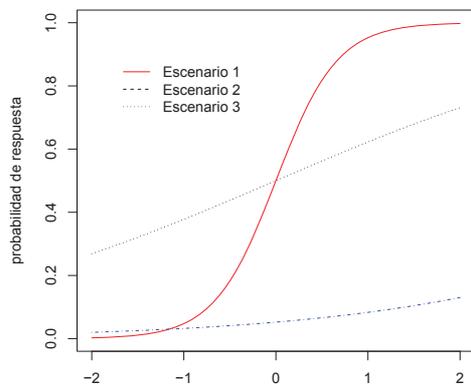


Figura 1. Escenarios para el modelo logístico.

2. En el modelo Poisson, la selección de los parámetros se hace en forma análoga al modelo logístico.

- Escenario 1:  $\beta_0 = 2,9, \beta_1 = -0,3, \beta_2 = -0,2$
- Escenario 2:  $\beta_0 = 1,9, \beta_1 = 0,1, \beta_2 = 0,4$
- Escenario 3:  $\beta_0 = -0,7, \beta_1 = 0,2, \beta_2 = 0,5$

Note en la figura 2 que con el escenario 1 se permite que  $\lambda$  tenga más variabilidad que en el 3 escenario.

3. Para el modelo lineal clásico, se trabajó sólo con un escenario en los parámetros, éstos fueron tomados arbitrariamente.

$$\beta_0 = -0,6, \beta_1 = 1,3, \beta_2 = 1,4$$

En las simulaciones el desempeño de (1) y de (5) se evaluó utilizando diferentes niveles de referencia  $\delta = 0,05, 0,10, 0,15, 0,20$ ; este desempeño se determina calculando la proporción de veces en que el

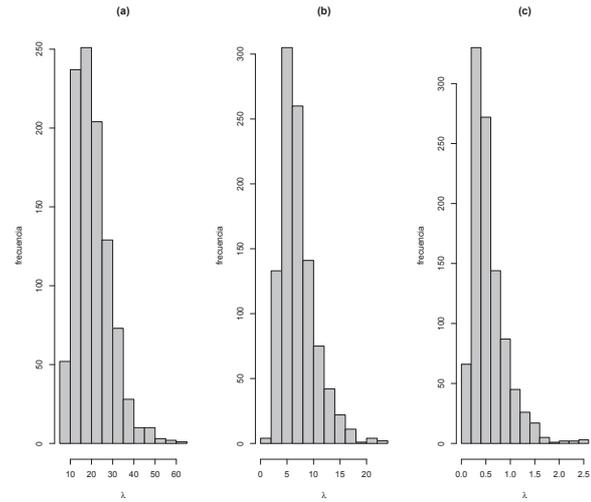


Figura 2. Escenarios para el modelo Poisson (a) Escenario 1, (b) Escenario 2, (c) Escenario 3.

criterio es inferior a estos niveles.

Debido a que se están asumiendo diferentes grados de correlación entre las variables explicativas, por lo tanto niveles bajos y altos de confusión en el modelo, se determina que el criterio es válido en los modelos mencionados anteriormente si la proporción de veces en la que  $\Delta\hat{\beta} \% < \delta$ :

- es cercana a cero y  $\rho \rightarrow \pm 1$ , es decir cuando existe un alto grado de confusión en el modelo.
- es cercana a uno y  $\rho \rightarrow 0$ , es decir cuando la confusión no esté presente en el modelo.

### 3. Resultados y discusión

La presentación de los resultados se desarrolla en dos partes, una para los modelos logístico y Poisson y otra para el modelo lineal clásico. Los resultados en los dos primeros modelos estudiados se encuentran a su vez divididos en los tres escenarios considerados, así mismo, éstos están divididos por cada nivel de referencia  $\delta$ .

**Tabla 1.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,05$ .

Modelo	$\rho$	n				
		50	100	200	500	1000
logístico	-0.8	0.0006(0.0002)	0(0)	0(0)	0(0)	0(0)
	-0.6	0.0044(0.002)	0.0006(0.0004)	0(0)	0(0)	0(0)
	-0.4	0.0186(0.0258)	0.0048(0.0076)	0.0006(0.0008)	0(0)	0(0)
	-0.2	0.0256(0.075)	0.024(0.0772)	0.0142(0.0494)	0.0016(0.0092)	0(0.0004)
	0.0	0.0268(0.1136)	0.025(0.166)	0.0264(0.261)	0.03(0.3806)	0.0264(0.5218)
	0.2	0.0274(0.0802)	0.0226(0.0758)	0.0144(0.0482)	0.0014(0.008)	0(0.0002)
	0.4	0.015(0.0208)	0.0054(0.0084)	0.0002(0.0004)	0(0)	0(0)
	0.6	0.0046(0.0042)	0.0002(0.0002)	0(0)	0(0)	0(0)
	0.8	0.0012(0.0008)	0(0)	0(0)	0(0)	0(0)
Poisson	-0.8	1	1	1	1	1
	-0.6	1	1	1	1	1
	-0.4	1	1	1	1	1
	-0.2	1	1	1	1	1
	0.0	1	1	1	1	1
	0.2	1	1	1	1	1
	0.4	1	1	1	1	1
	0.6	1	1	1	1	1
	0.8	1	1	1	1	1

**Tabla 2.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,10$ .

Modelo	$\rho$	n				
		50	100	200	500	1000
logístico	-0.8	0.003(0.0006)	0(0)	0(0)	0(0)	0(0)
	-0.6	0.0088(0.0064)	0.001(0.0006)	0(0)	0(0)	0(0)
	-0.4	0.0322(0.048)	0.0096(0.016)	0.001(0.002)	0(0)	0(0)
	-0.2	0.054(0.1584)	0.0456(0.1658)	0.0272(0.1186)	0.0036(0.042)	0.0002(0.0066)
	0.0	0.0584(0.2246)	0.056(0.3254)	0.0556(0.4838)	0.0556(0.6868)	0.0584(0.8522)
	0.2	0.0514(0.152)	0.0452(0.1438)	0.03(0.1062)	0.0028(0.0304)	0(0.0052)
	0.4	0.0306(0.0404)	0.0108(0.0164)	0.0012(0.002)	0(0)	0(0)
	0.6	0.0092(0.0072)	0.0006(0.0002)	0(0)	0(0)	0(0)
	0.8	0.0028(0.001)	0(0)	0(0)	0(0)	0(0)
Poisson	-0.8	1	1	1	1	1
	-0.6	1	1	1	1	1
	-0.4	1	1	1	1	1
	-0.2	1	1	1	1	1
	0.0	1	1	1	1	1
	0.2	1	1	1	1	1
	0.4	1	1	1	1	1
	0.6	1	1	1	1	1
	0.8	1	1	1	1	1

**Tabla 3.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,15$ .

Modelo	$\rho$	n				
		50	100	200	500	1000
logístico	-0.8	0.0042(0.0006)	0(0)	0(0)	0(0)	0(0)
	-0.6	0.0136(0.0116)	0.0014(0.0008)	0(0)	0(0)	0(0)
	-0.4	0.0488(0.0778)	0.0138(0.0304)	0.0012(0.0052)	0(0)	0(0)
	-0.2	0.087(0.2384)	0.066(0.2616)	0.0412(0.2154)	0.007(0.1166)	0.0004(0.0452)
	0.0	0.0908(0.3266)	0.0882(0.4692)	0.0868(0.6532)	0.0856(0.8622)	0.09(0.9662)
	0.2	0.0758(0.2214)	0.0696(0.2286)	0.0442(0.1774)	0.0056(0.0792)	0.0002(0.0222)
	0.4	0.0452(0.0642)	0.018(0.0272)	0.0016(0.004)	0(0)	0(0)
	0.6	0.0158(0.012)	0.0008(0.0008)	0(0)	0(0)	0(0)
	0.8	0.0044(0.0014)	0(0)	0(0)	0(0)	0(0)
Poisson	-0.8	1	1	1	1	1
	-0.6	1	1	1	1	1
	-0.4	1	1	1	1	1
	-0.2	1	1	1	1	1
	0.0	1	1	1	1	1
	0.2	1	1	1	1	1
	0.4	1	1	1	1	1
	0.6	1	1	1	1	1
	0.8	1	1	1	1	1

**Tabla 4.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,20$ .

Modelo	$\rho$	n				
		50	100	200	500	1000
logístico	-0.8	0.0066(0.0018)	0(0)	0(0)	0(0)	0(0)
	-0.6	0.0184(0.0178)	0.0018(0.002)	0(0)	0(0)	0(0)
	-0.4	0.0656(0.1078)	0.0206(0.0526)	0.0028(0.0106)	0(0)	0(0)
	-0.2	0.113(0.327)	0.0866(0.3636)	0.0566(0.346)	0.0104(0.2848)	0.0004(0.2006)
	0.0	0.12(0.427)	0.1172(0.59)	0.1194(0.7824)	0.1168(0.9494)	0.1242(0.9942)
	0.2	0.1038(0.2828)	0.0906(0.3084)	0.056(0.2624)	0.0082(0.1666)	0.0004(0.0846)
	0.4	0.0602(0.0902)	0.0246(0.039)	0.003(0.0068)	0(0)	0(0)
	0.6	0.0214(0.0152)	0.0018(0.001)	0(0)	0(0)	0(0)
	0.8	0.007(0.0022)	0(0)	0(0)	0(0)	0(0)
Poisson	-0.8	1	1	1	1	1
	-0.6	1	1	1	1	1
	-0.4	1	1	1	1	1
	-0.2	1	1	1	1	1
	0.0	1	1	1	1	1
	0.2	1	1	1	1	1
	0.4	1	1	1	1	1
	0.6	1	1	1	1	1
	0.8	1	1	1	1	1

En todas las tablas, los resultados representan la proporción de veces en las que el criterio del cambio porcentual fue inferior a los niveles de referencia asumidos (0.05,0.10,0.15,0.20) para cada modelo, y los resultados que se encuentran entre paréntesis, corresponden a el criterio propuesto para el modelo logístico.

### 3.1. Modelo logístico y Modelo Poisson

#### 3.1.1. Escenario 1

En las tablas 1 a 4 se presentan los resultados obtenidos para este escenario. En estas tablas se puede observar que para el modelo Poisson, la proporción de veces en las que  $\Delta\hat{\beta}\% < \delta$  fue exactamente 1, esto sin importar el grado de correlación entre las variables, el tamaño de la muestra y el nivel de referencia  $\delta$ , indicando con esto que la presencia del fenómeno de la confusión no estaba presente en el modelo cuando realmente lo estaba, por lo tanto, para este modelo y bajo este escenario el criterio no funciona. Sin embargo, cuando  $\rho = 0$  o sea cuando no hay confusión, el criterio acertó en todas los modelos que se ajustaron.

En el modelo logístico, los criterios  $\Delta\hat{\beta}\%$  y  $CP$  se comportan de manera similar y tienen buen desempeño, principalmente cuando existe algún grado de correlación entre las covariables y los tamaños en las muestras aumentan, pues en estos casos las proporciones son cercanas a cero y por tanto logran detectar la presencia de confusión en el modelo. Sin embargo cuando  $\rho = \pm 0,2$  y  $\delta \geq 0,10$  el criterio  $\Delta\hat{\beta}\%$  es más efectivo que el propuesto, puesto que las proporciones son más bajas y por lo tanto percibe con mayor exactitud la presencia del fenómeno en el modelo.

La figura 3 ilustra el comportamiento de los criterios  $\Delta\hat{\beta}\%$  y  $CP$  cuando en el modelo logístico se simularon covariables con  $\rho = 0$ . Se puede observar en la figura 3(a) que la proporción de veces en las que  $\Delta\hat{\beta}\% < \delta$  es menor a 0,15, señalando con esto que el fenómeno de la confusión se encuentra presente en aproximadamente el 80% de los modelos, cuando realmente no lo está. En la figura 3(b) las proporciones en las que  $CP < \delta$  son acercanas a 1 a medida que  $n$  y  $\delta$  aumentan, esto indica que este criterio es más efectivo que el anterior.

#### 3.1.2. Escenario 2

En la tablas de la 5 a 8 se observa que en el modelo Poisson, la proporción de veces en las que  $\Delta\hat{\beta}\% < \delta$  disminuye a medida de aumenta  $n$  y que el grado de correlación entre las variables es más fuerte; esto tal vez se deba a que el modelo se simula con una varianza menor a la que se consideró en el escenario anterior.

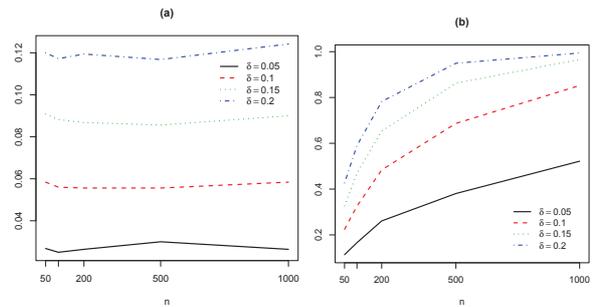


Figura 3. Comportamiento de los criterios (a)  $\Delta\hat{\beta}\%$  y (b)  $CP$ , en el modelo logístico cuando  $\rho = 0$ .

La figura 4 muestra el comportamiento del criterio del cambio porcentual propuesto por Hosmer & Lemeshow [8] en ambos modelos cuando  $\delta = 0,20$  y correlaciones positivas entre las covariables. Note que en estos casos y para  $n \geq 100$  el criterio es bastante eficiente en los dos modelos, sin embargo, en el modelo logístico se encuentra que cuando  $\rho = 0$  el criterio no tiene buen desempeño, ya que las proporciones son inferiores a 0,5.

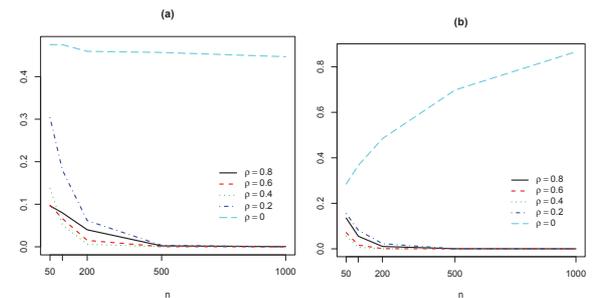


Figura 4. Comportamiento del criterio  $\Delta\hat{\beta}\%$  cuando  $\delta = 0,20$  y  $\rho = 0,8, 0,6, 0,4, 0,2, 0$  en (a) Modelo logístico, (b) Modelo Poisson.

**Tabla 5.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,05$ .

Modelo	$\rho$	n				
		50	100	200	500	1000
logístico	-0.8	0.0254(0.0612)	0.021(0.0504)	0.0074(0.0248)	0.0002(0.0014)	0(0)
	-0.6	0.0238(0.066)	0.0126(0.0464)	0.0018(0.012)	0(0)	0(0)
	-0.4	0.033(0.1248)	0.0108(0.0598)	0.0006(0.0102)	0(0)	0(0)
	-0.2	0.0936(0.3216)	0.0438(0.261)	0.007(0.1606)	0(0.0466)	0(0.0072)
	0.0	0.1852(0.551)	0.1606(0.6896)	0.1422(0.8354)	0.1328(0.9728)	0.1326(0.9978)
	0.2	0.1036(0.328)	0.0428(0.249)	0.0046(0.1426)	0(0.0304)	0(0.002)
	0.4	0.0342(0.1126)	0.0088(0.0522)	0.0008(0.0068)	0(0.0002)	0(0)
	0.6	0.0266(0.069)	0.0102(0.0362)	0.0024(0.0108)	0(0.0002)	0(0)
	0.8	0.026(0.058)	0.0178(0.049)	0.0082(0.026)	0.0004(0.0012)	0(0.0004)
Poisson	-0.8	0.1384	0.0594	0.0128	0	0
	-0.6	0.0718	0.0182	0.0018	0	0
	-0.4	0.0478	0.01	0.0006	0	0
	-0.2	0.0592	0.0192	0.0036	0	0
	0.0	0.0994	0.1012	0.1314	0.2078	0.2894
	0.2	0.0624	0.024	0.0048	0	0
	0.4	0.046	0.0066	0.0002	0	0
	0.6	0.0712	0.0166	0.0008	0	0
	0.8	0.1346	0.0552	0.0104	0.0004	0

**Tabla 6.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,10$ .

Modelo	$\rho$	n				
		50	100	200	500	1000
logístico	-0.8	0.054(0.1258)	0.0428(0.1052)	0.0176(0.0566)	0.0006(0.007)	0(0)
	-0.6	0.046(0.1462)	0.0292(0.114)	0.0068(0.039)	0(0.0028)	0(0)
	-0.4	0.0648(0.2762)	0.0246(0.1918)	0.0022(0.0898)	0(0.019)	0(0.002)
	-0.2	0.1752(0.5678)	0.082(0.6024)	0.0164(0.6142)	0(0.664)	0(0.7726)
	0.0	0.3048(0.7902)	0.291(0.9144)	0.2716(0.9828)	0.2576(0.9998)	0.2546(1)
	0.2	0.181(0.5454)	0.089(0.5458)	0.0146(0.5414)	0(0.5226)	0(0.5224)
	0.4	0.068(0.2352)	0.022(0.1562)	0.0016(0.0666)	0(0.0066)	0(0.0002)
	0.6	0.0502(0.139)	0.023(0.084)	0.0048(0.0346)	0(0.0022)	0(0.0002)
	0.8	0.0492(0.1108)	0.0386(0.0994)	0.0196(0.0546)	0.0006(0.006)	0.0002(0.0006)
Poisson	-0.8	0.1384	0.0594	0.0128	0	0
	-0.6	0.0718	0.0182	0.0018	0	0
	-0.4	0.0488	0.0102	0.0006	0	0
	-0.2	0.0832	0.0356	0.0084	0	0
	0.0	0.168	0.191	0.2478	0.402	0.5494
	0.2	0.093	0.043	0.0096	0	0
	0.4	0.0486	0.0066	0.0002	0	0
	0.6	0.0712	0.0166	0.0008	0	0
	0.8	0.1348	0.0552	0.0104	0.0004	0

**Tabla 7.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,15$ .

Modelo	$\rho$	n				
		50	100	200	500	1000
logístico	-0.8	0.0796(0.1794)	0.0632(0.1698)	0.027(0.1002)	0.0012(0.0264)	0(0.0036)
	-0.6	0.0688(0.2462)	0.0436(0.2074)	0.0112(0.1248)	0(0.0294)	0(0.0038)
	-0.4	0.1038(0.4388)	0.042(0.4132)	0.0044(0.3476)	0(0.2546)	0(0.1742)
	-0.2	0.2436(0.7468)	0.1248(0.833)	0.032(0.9092)	0.0024(0.982)	0(0.9984)
	0.0	0.4028(0.8938)	0.3964(0.9756)	0.3768(0.998)	0.3632(1)	0.3584(1)
	0.2	0.2452(0.6952)	0.13(0.7642)	0.0358(0.8274)	0.0008(0.927)	0(0.9804)
	0.4	0.1034(0.3678)	0.0348(0.2996)	0.0034(0.2254)	0(0.1064)	0(0.039)
	0.6	0.0752(0.2156)	0.0346(0.1544)	0.0106(0.0836)	0(0.0138)	0(0.0006)
	0.8	0.0746(0.1676)	0.0612(0.1462)	0.0286(0.0856)	0.0014(0.0172)	0.0002(0.0028)
Poisson	-0.8	0.1384	0.0594	0.0128	0	0
	-0.6	0.0718	0.0182	0.0018	0	0
	-0.4	0.0502	0.0102	0.0006	0	0
	-0.2	0.1094	0.0526	0.015	0	0
	0.0	0.225	0.2814	0.3704	0.5718	0.7366
	0.2	0.123	0.0602	0.0144	0.0002	0
	0.4	0.051	0.0066	0.0002	0	0
	0.6	0.0712	0.0166	0.0008	0	0
	0.8	0.135	0.0552	0.0104	0.0004	0

**Tabla 8.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,20$ .

Modelo	$\rho$	n				
		50	100	200	500	1000
logístico	-0.8	0.1028(0.2462)	0.0792(0.2424)	0.036(0.1788)	0.002(0.074)	0(0.0238)
	-0.6	0.0932(0.3552)	0.0578(0.3408)	0.0132(0.2724)	0(0.1694)	0(0.081)
	-0.4	0.5926(0.5926)	0.0562(0.6502)	0.0078(0.679)	0(0.7736)	0(0.854)
	-0.2	0.2996(0.8646)	0.174(0.9448)	0.0574(0.9894)	0(1)	0(1)
	0.0	0.4752(0.994)	0.475(0.9932)	0.4594(1)	0.4568(1)	0.4468(1)
	0.2	0.3036(0.7948)	0.1804(0.8776)	0.0616(0.9528)	0.0034(0.9952)	0(1)
	0.4	0.1372(0.4936)	0.0518(0.4786)	0.006(0.456)	0(0.431)	0(0.3988)
	0.6	0.0976(0.2934)	0.0664(0.238)	0.015(0.159)	0.0004(0.063)	0(0.0126)
	0.8	0.0964(0.2234)	0.0798(0.1974)	0.04(0.1228)	0.0022(0.0418)	0.0004(0.0084)
Poisson	-0.8	0.1384	0.0594	0.0128	0	0
	-0.6	0.0718	0.0182	0.0018	0	0
	-0.4	0.053	0.0102	0.0006	0	0
	-0.2	0.14	0.0706	0.0234	0.0008	0
	0.0	0.2848	0.3666	0.4844	0.698	0.8664
	0.2	0.1548	0.0818	0.0226	0.0002	0
	0.4	0.0548	0.0066	0.0002	0	0
	0.6	0.0712	0.0166	0.0008	0	0
	0.8	0.1352	0.0552	0.0104	0.0004	0

En relación al criterio propuesto se observa que es sensible cuando existe algún grado de correlación entre las covariables y cuando  $\delta$  aumenta, puesto que la proporción de veces en las que fue menor a  $\delta$  es bastante alta, indicando que el fenómeno se encuentra en el modelo cuando realmente no está. A manera de ilustración, observe que cuando  $\rho = \pm 0,2$ , sin importar el tamaño de la muestra el criterio no detecta la confusión en más de la mitad de los modelos que fueron realmente simulados con ese fenómeno.

### 3.1.3. Escenario 3

En las tablas 9 y 10 para el modelo logístico, se observa que a medida que el grado de correlación entre las covariables es más fuerte, la proporción de veces en que  $\Delta\hat{\beta}\% < \delta$  disminuye, y que cuando se aumenta el tamaño de la muestra tienden a ser nula. Este comportamiento también se observa en el modelo Poisson cuando  $n \geq 500$ .

En cuanto a la ausencia de correlación entre las covariables ( $\rho = 0$ ), se encuentran proporciones más altas en el modelo Poisson, pero al igual que en el modelo logístico, éstas son inferiores a 0,5 cuando  $\delta = 0,05$  o  $\delta = 0,10$  figura 5(e) y figura 6(e), respectivamente), excepto en el modelo Poisson cuando  $n \geq 500$  y  $\delta = 0,10$ ; es decir en estos casos  $\Delta\hat{\beta}\%$  indica que la confusión está presente en dicho modelo, en más de la mitad de los modelos simulados sin este fenómeno.

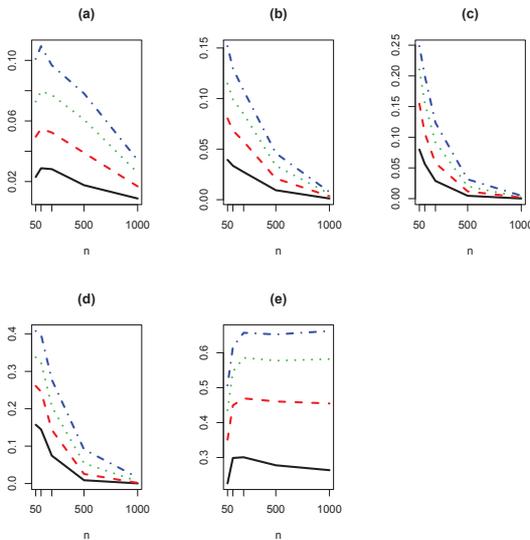


Figura 5. Comportamiento del criterio  $\Delta\hat{\beta}\%$  en el modelo logístico cuando (a)  $\rho = 0,8$ , (b)  $\rho = 0,6$ , (c)  $\rho = 0,4$ , (d)  $\rho = 0,2$ , (e)  $\rho = 0$ .

En la figura 5, se nota que el criterio  $\Delta\hat{\beta}\%$  mantiene un buen desempeño en el modelo logístico y que además mejora cuando  $\rho = 0$ , pues a medida de que se aumenta el tamaño de la muestra y el nivel de referencia  $\delta$ , las proporciones son cada vez más grandes. Cuando  $\rho = 0,8, 0,6, 0,4$  en el modelo Poisson, el criterio tiene el mismo comportamiento figura 6(a), (b), (c), mostrando que para  $n \geq 500$  este criterio tiene un desempeño en este modelo.

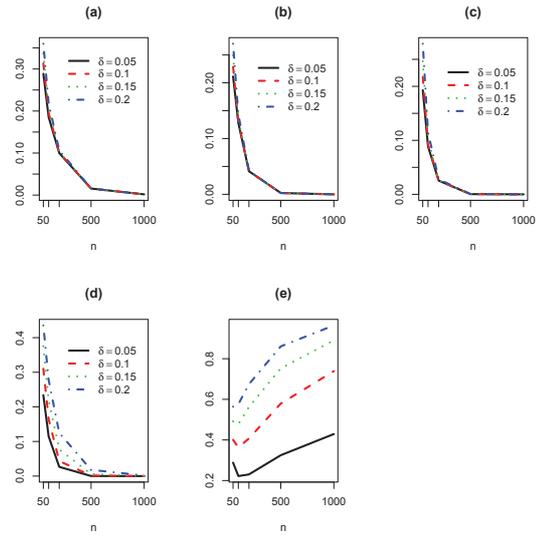


Figura 6. Comportamiento del criterio  $\Delta\hat{\beta}\%$  en el modelo Poisson cuando (a)  $\rho = 0,8$ , (b)  $\rho = 0,6$ , (c)  $\rho = 0,4$ , (d)  $\rho = 0,2$ , (e)  $\rho = 0$ .

Los criterios en estudio tienen comportamientos muy similares cuando en el modelo logístico las covariables se simulan con algún grado de correlación, puesto que en la mayoría de los casos la diferencia entre las proporciones tiende a ser nula, salvo en  $\rho = -0,2$  donde las proporciones de CP llegan a ser superiores a 0,5, mientras que con el criterio  $\Delta\hat{\beta}\%$  éstas tienden a cero. Para  $\rho = 0$  las proporciones aumentan cuando  $n$  y  $\delta$  aumentan pero con el criterio propuesto se obtienen valores más cercanos a 1.

### 3.2. Modelo Lineal clásico

En las tablas de la tablas 13 a la 16, se muestra el comportamiento del criterio  $\Delta\hat{\beta}\%$  cuando se trabaja con un modelo lineal clásico.

Cuando  $\rho$  toma valores  $\pm 0,8$  y  $\pm 0,6$ , la proporción de veces en la que  $\Delta\hat{\beta}\% < \delta$  es nula, esto significa que para los  $N = 5000$  modelos simulados bajo un alto grado de confusión, este criterio detectó que en todas el fenómeno realmente estaba presente.

**Tabla 9.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,05$ .

Modelo	$\rho$	n				
		50	100	200	500	1000
logístico	-0.8	0.0256(0.09)	0.0262(0.0512)	0.0216(0.0562)	0.0148(0.0434)	0.0086(0.0254)
	-0.6	0.0408(0.1134)	0.0378(0.0728)	0.0284(0.076)	0.0126(0.0404)	0.0014(0.0062)
	-0.4	0.0724(0.171)	0.0584(0.1332)	0.0348(0.1072)	0.0092(0.0398)	0(0.0058)
	-0.2	0.1584(0.3194)	0.1398(0.3294)	0.0068(0.273)	0.0122(0.145)	0.002(0.0572)
	0.0	0.2258(0.437)	0.2986( 0.6008)	0.3006(0.757)	0.2776(0.9268)	0.2636( 0.986)
	0.2	0.1572(0.3166)	0.1448(0.3224)	0.0746(0.2654)	0.0088(0.1282)	0.0004(0.0468)
	0.4	0.08(0.1762)	0.0568(0.1274)	0.0288(0.1016)	0.0046(0.0316)	0.0002(0.0066)
	0.6	0.0394(0.113)	0.0336(0.0724)	0.0274(0.075)	0.0094(0.034)	0.0012(0.0074)
	0.8	0.023(0.0862)	0.0288(0.0562)	0.0282(0.059)	0.0176(0.0526)	0.0088(0.0226)
Poisson	-0.8	0.27	0.1976	0.1086	0.029	0.0042
	-0.6	0.2172	0.122	0.0522	0.0052	0.0002
	-0.4	0.1868	0.0962	0.0266	0.0018	0
	-0.2	0.221	0.1096	0.029	0.0008	0
	0.0	0.2882	0.222	0.2302	0.3248	0.429
	0.2	0.2338	0.1148	0.0266	0.0002	0
	0.4	0.193	0.0886	0.0254	0.0004	0
	0.6	0.2114	0.128	0.0414	0.0024	0
	0.8	0.2886	0.1854	0.1	0.0156	0.0018

**Tabla 10.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,10$ .

Modelo	$\rho$	n				
		50	100	200	500	1000
logístico	-0.8	0.0468(0.1212)	0.0484(0.1008)	0.0446(0.1096)	0.035(0.0914)	0.019(0.0536)
	-0.6	0.0812(0.1746)	0.0718(0.1546)	0.0554(0.156)	0.0236(0.0938)	0.0026(0.0254)
	-0.4	0.1408(0.2814)	0.1078(0.266)	0.0672(0.2336)	0.0186(0.141)	0.0012(0.0636)
	-0.2	0.2552(0.4738)	0.2416(0.5504)	0.1476(0.5864)	0.032(0.5928)	0.0018(0.6082)
	0.0	0.351(0.6054)	0.4494(0.7942)	0.4694(0.933)	0.4606(0.9956)	0.4548(1)
	0.2	0.261(0.4698)	0.2458(0.5274)	0.1436(0.531)	0.0256(0.4924)	0.0018(0.484)
	0.4	0.1546(0.29)	0.1068(0.2454)	0.058(0.2132)	0.0118(0.1164)	0.0018(0.0448)
	0.6	0.0804(0.1726)	0.0684(0.1528)	0.0576(0.1512)	0.021(0.0818)	0.0034(0.0248)
	0.8	0.0496(0.121)	0.0548(0.1052)	0.0524(0.11)	0.039(0.1008)	0.0168(0.0456)
Poisson	-0.8	0.2706	0.1976	0.1086	0.029	0.0042
	-0.6	0.2196	0.1222	0.0522	0.0052	0.0002
	-0.4	0.1974	0.097	0.0266	0.0018	0
	-0.2	0.2748	0.1464	0.0406	0.0012	0
	0.0	0.4008	0.363	0.4074	0.5792	0.7388
	0.2	0.3104	0.1646	0.0434	0.0014	0
	0.4	0.2174	0.0924	0.0256	0.0004	0
	0.6	0.2286	0.1306	0.0416	0.0024	0
	0.8	0.3134	0.1952	0.1006	0.0156	0.0018

**Tabla 11.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,15$ .

Modelo	$\rho$	n				
		50	100	200	500	1000
logístico	-0.8	0.0688(0.157)	0.0748(0.1518)	0.0688(0.172)	0.0502(0.1532)	0.0284(0.0982)
	-0.6	0.1186(0.2404)	0.106(0.2354)	0.0818(0.244)	0.0372(0.1798)	0.005(0.0886)
	-0.4	0.102(0.384)	0.159(0.4002)	0.1006(0.4016)	0.0264(0.3456)	0.0026(0.2824)
	-0.2	0.333(0.5962)	0.3214(0.707)	0.212(0.8028)	0.058(0.9008)	0.0056(0.9656)
	0.0	0.4354(0.7004)	0.546(0.8792)	0.5858(0.9782)	0.5774(1)	0.582(1)
	0.2	0.3376(0.5774)	0.3234(0.6704)	0.2058(0.731)	0.0542(0.8144)	0.006(0.8908)
	0.4	0.2104(0.3844)	0.1562(0.3594)	0.0918(0.3426)	0.02(0.2494)	0.0024(0.1626)
	0.6	0.1148(0.2282)	0.0992(0.223)	0.0848(0.22)	0.0336(0.1442)	0.0058(0.0598)
	0.8	0.0728(0.1582)	0.0796(0.1548)	0.0772(0.1644)	0.061(0.1482)	0.0258(0.0786)
Poisson	-0.8	0.2722	0.1978	0.1086	0.029	0.0042
	-0.6	0.2216	0.1226	0.0522	0.0052	0.0002
	-0.4	0.2096	0.0986	0.0266	0.0018	0
	-0.2	0.3258	0.1918	0.0644	0.0052	0
	0.0	0.4914	0.4772	0.5624	0.7524	0.8904
	0.2	0.3752	0.2186	0.0772	0.0058	0.0002
	0.4	0.2464	0.1024	0.026	0.0004	0
	0.6	0.249	0.1358	0.0418	0.0024	0
	0.8	0.337	0.2066	0.103	0.0156	0.0018

**Tabla 12.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,20$ .

Modelo	$\rho$	n				
		50	100	200	500	1000
logístico	-0.8	0.0906(0.194)	0.1002(0.203)	0.0942(0.2474)	0.0676(0.229)	0.0358(0.1672)
	-0.6	0.1548(0.303)	0.1408(0.3244)	0.1074(0.3474)	0.0438(0.3086)	0.0072(0.234)
	-0.4	0.2558(0.4794)	0.203(0.541)	0.126(0.576)	0.038(0.6266)	0.0044(0.6714)
	-0.2	0.3958(0.6814)	0.3954(0.8186)	0.2718(0.9162)	0.0918(0.9868)	0.0142(0.999)
	0.0	0.5068(0.77)	0.6168(0.9286)	0.6574(0.9904)	0.6526(1)	0.6624(1)
	0.2	0.408(0.6592)	0.3982(0.7722)	0.2778(0.8554)	0.0912(0.944)	0.0162(0.986)
	0.4	0.2484(0.4662)	0.2006(0.4592)	0.1236(0.475)	0.0314(0.4302)	0.0048(0.398)
	0.6	0.1518(0.2842)	0.1298(0.2926)	0.108(0.2856)	0.046(0.2124)	0.0076(0.1302)
	0.8	0.101(0.1954)	0.1094(0.2038)	0.0968(0.2216)	0.0782(0.195)	0.0342(0.113)
Poisson	-0.8	0.273	0.1978	0.1086	0.029	0.0042
	-0.6	0.2254	0.1236	0.0522	0.0052	0.0002
	-0.4	0.2258	0.1048	0.0266	0.0018	0
	-0.2	0.3786	0.2464	0.1026	0.0172	0.0006
	0.0	0.5638	0.576	0.6746	0.8608	0.964
	0.2	0.4354	0.2778	0.1248	0.018	0.0016
	0.4	0.279	0.1148	0.028	0.0004	0
	0.6	0.2702	0.1436	0.0422	0.0024	0
	0.8	0.3602	0.218	0.1066	0.0156	0.0018

**Tabla 13.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,05$ .

$\rho$	n				
	50	100	200	500	1000
-0.8	0	0	0	0	0
-0.6	0	0	0	0	0
-0.4	0.0042	0	0	0	0
-0.2	0.0968	0.0484	0.0146	0.0002	0
0	0.2552	0.352	0.485	0.7058	0.8452
0.2	0.1028	0.0546	0.0134	0	0
0.4	0.0046	0	0	0	0
0.6	0	0	0	0	0
0.8	0	0	0	0	0

**Tabla 14.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,10$ .

$\rho$	n				
	50	100	200	500	1000
-0.8	0	0	0	0	0
-0.6	0	0	0	0	0
-0.4	0.0108	0.0004	0	0	0
-0.2	0.2076	0.1306	0.0594	0.0068	0.0004
0	0.4894	0.646	0.8062	0.9636	0.9972
0.2	0.2158	0.1398	0.0558	0.0062	0.0002
0.4	0.0128	0.0004	0	0	0
0.6	0	0	0	0	0
0.8	0	0	0	0	0

**Tabla 15.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,15$ .

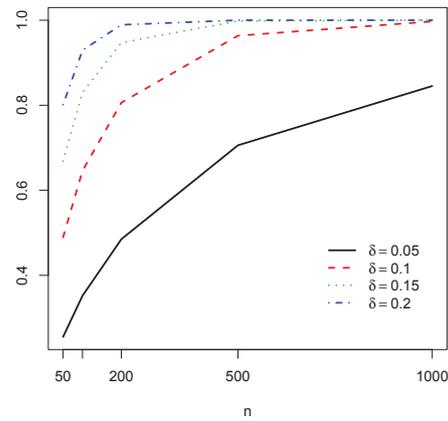
$\rho$	n				
	50	100	200	500	1000
-0.8	0	0	0	0	0
-0.6	0	0	0	0	0
-0.4	0.0244	0.0024	0	0	0
-0.2	0.3368	0.2622	0.1892	0.0814	0.0256
0	0.6686	0.83	0.9474	0.998	1
0.2	0.3386	0.2758	0.1884	0.0844	0.0236
0.4	0.029	0.0036	0	0	0
0.6	0.0002	0	0	0	0
0.8	0	0	0	0	0

En la figura 7, se observa que a medida que los tamaños muestrales ( $n$ ) y el nivel de referencia ( $\delta$ ) aumentan y que no existe confusión en el modelo ( $\rho = 0$ ), el desempeño del criterio mejora. Note por ejemplo en la tabla 15 donde  $\delta = 0,20$ , el criterio detecta efectivamente la ausencia de confusión en más del 80% de los modelos simulados, esto sin importar el tamaño en las muestras.

Para  $\rho \pm 0,4$  y  $n = 50$ , se observa que  $\Delta\hat{\beta}\%$  tiende a equivocarse aproximadamente entre el 1% y el 6% de las simulaciones, aunque podría pensarse que este “pequeño error” sucede porque el tamaño de muestra

**Tabla 16.** Proporción de veces en las que el cambio porcentual  $\Delta\hat{\beta}\% < \delta = 0,20$ .

$\rho$	n				
	50	100	200	500	1000
-0.8	0	0	0	0	0
-0.6	0.0006	0	0	0	0
-0.4	0.0556	0.0104	0.0002	0	0
-0.2	0.4518	0.4434	0.4088	0.3694	0.3184
0	0.8016	0.9294	0.9892	1	1
0.2	0.4596	0.4564	0.416	0.3756	0.3218
0.4	0.0596	0.0122	0.0014	0	0
0.6	0.0004	0	0	0	0
0.8	0	0	0	0	0



**Figura 7.** Comportamiento del criterio  $\Delta\hat{\beta}\%$  en el modelo lineal clásico cuando  $\rho = 0$ .

no es lo suficientemente grande, debido a que cuando  $n \geq 100$ , el criterio funciona perfectamente. Análogamente sucede cuando  $\rho \pm 0,2$ , donde el desempeño del criterio es cada vez más bajo cuando se aumenta el nivel de referencia.

#### 4. Conclusiones

En este artículo se evaluó el desempeño del criterio de cambio porcentual propuesto por Hosmer [8] como una posible medida de confusión en modelos de la familia exponencial.

Se encontró que independientemente de la probabilidad del resultado, el nivel de referencia, el tamaño de la muestra y el grado de confusión, el criterio  $\Delta\hat{\beta}\%$  mostró tener buen desempeño en el modelo logístico, sin embargo éste tiende a fallar cuando no existe correlación entre las covariables ( $\rho = 0$ ), principalmente en el escenario 1, donde la probabilidad del resultado se permitía variar en un rango de muy bajo a muy alto. Al igual que el criterio del cambio porcentual, el criterio propuesto tiene un desempe-

ño excelente a lo largo de los escenarios simulados, pero es más confiable para detectar la ausencia de confusión; cuando se emplee este criterio como medida de confusión, se recomienda utilizar un nivel de referencia  $\delta = 0,10$ .

El principal hallazgo es que cuando se simula la mayor variabilidad en la variable respuesta del modelo Poisson, es decir bajo el primer escenario, el criterio falló completamente en todos los modelos simulados con confusión, incluso con un tamaño de muestra grande. En cambio, el criterio fue efectivo en todos los modelos simulados cuyas covariables tenían  $\rho = 0$ .

En forma general, el criterio estudiado mostró a lo largo de los escenarios de simulación, tener mejor desempeño en el modelo lineal clásico, donde un nivel de referencia del 5% es suficiente para que dicho criterio fuera capaz de detectar en más del 90% de las situaciones simuladas con el fenómeno de la confusión; pero para garantizar la veracidad del criterio en situaciones donde no haya confusión, se recomienda utilizar  $\delta = 0,15$ .

## Referencias

- [1] P. Austin y L. Brunner, "Inflation of the Type I Error Rate When a Continuous Confounding Variable is Categorized in Logistic Regression Analysis", *Statistics in Medicine*, vol. 23, pp. 1159-1178, 2004.
- [2] H. Becher, "The concept of Residual Confounding in Regression Models and some applications", *Statistics in Medicine*, 11, pp. 1747-1758, 1992.
- [3] R. Bliss, J. Weinberg, T. Webster y V. Vieira, "Determining the Probability Distribution and Evaluating Sensitivity and False Positive Rate of a Confounder Detection Method Applied To Logistic Regression", *J Biom Bioestat*, vol. 3, no. 4, pp. 1-14, 2013.
- [4] W. Chao, M. Palta y T. Young, "Effect of omitted Confounders on the Analysis of Correlated Binary Data.", *Biometrics*, vol. 53, no. 2, pp. 678-689, 1997.
- [5] C. Chen, D.P. Chock y S.L. Winkler, "A Simulation Study of Confounding in Generalized Linear Models for Air Pollution Epidemiology", *Environmental Health Perspectives*, vol. 107, no. 3, pp. 217-222, 1999.
- [6] K. A. Frank, "Impact of a Confounding Variable on the Inference of a Regression Coefficient", *Sociological Methods and Research*, vol. 29, no. 2, pp. 147-194, 2002.
- [7] D. W. Hosmer y S. Lemeshow, *Applied logistic Regression*, John Wiley & Sons, Inc., New York, 2000.
- [8] D. W. Hosmer y S. Lemeshow, *Applied Survival Analysis*, John Wiley & Sons, Inc., New York, 1999.
- [9] K. Schneiter y J. Symanzik, "An Applet for the Investigation of Simpson's Paradox", *Journal of Statistics Education*, vol. 21, no. 1, pp. 1-20, 2013.
- [10] F. Kamangar, "Confounding Variables in Epidemiologic Studies: Basics and Beyond", *Arch Iran Med*, vol. 15, no. 6, pp. 508-516, 2012.
- [11] J. Mickey, y S. Greenland, "A study of the impact of confounder selection criteria of effect estimation.", *American Journal of epidemiology*, vol. 129, no. 1, pp. 125-137, 1989.
- [12] S. C. Newman, *Biostatistical Methods in Epidemiology*, John Wiley & Sons, Inc., New York, 2001.
- [13] M. A. Pourhoseingholi, A. R. Baghestani y M. Vahedi, "How to Control Confounding Effects by Statistical Analysis", *Gastroenterol Hepatol Bed Bench*, vol. 5, no. 2, pp. 79-83, 2012.
- [14] S. Wilson, y I. Gordon, "Calculating Sample Sizes in the Presence of Confounding Variables", *Journal of the Royal Statistical Society Series C (Applied Statistics)*, vol. 35, no. 2, pp.207-213, 1986.
- [15] M. Woodward, *Epidemiology. Study Design and Data Analysis*, 2 edn, Boca Ratón, Chapman & Hall/CRC, 1999.