Universidad del País Vasco / Euskal Herriko Unibertsitatea

La importancia de una correcta estimación del LC50 en estudios toxicológicos

Primeras Jornadas Científicas de Estudiantes de la SEB

Nerea Gutiérrez

Con la colaboración de Irantzu Barrio, Jose María Lacave, Amaia Orbea e Inmaculada Arostegui

Valencia, 19 de enero del 2015

Índice

1. Introducción
2. Métodos
   - Máxima verosimilitud
   - Máxima verosimilitud penalizada
3. Simulación
   - Estudio realizado
   - Resultados
4. Conclusiones
5. Trabajo a futuro

Índice

1. Introducción
2. Métodos
   - Máxima verosimilitud
   - Máxima verosimilitud penalizada
3. Simulación
   - Estudio realizado
   - Resultados
4. Conclusiones
5. Trabajo a futuro

Índice

1. Introducción
2. Métodos
   - Máxima verosimilitud
   - Máxima verosimilitud penalizada
3. Simulación
   - Estudio realizado
   - Resultados
4. Conclusiones
5. Trabajo a futuro

Índice

1. Introducción
2. Métodos
   - Máxima verosimilitud
   - Máxima verosimilitud penalizada
3. Simulación
   - Estudio realizado
   - Resultados
4. Conclusiones
5. Trabajo a futuro

Índice

1. Introducción
2. Métodos
   - Máxima verosimilitud
   - Máxima verosimilitud penalizada
3. Simulación
   - Estudio realizado
   - Resultados
4. Conclusiones
5. Trabajo a futuro

Motivación I

Contexto

- Nuevas sustancias químicas en nuestro día a día.
- Para poder utilizarlas de forma segura, es necesario evaluar su toxicidad de una forma práctica.
- "La diferencia entre lo que cura y lo que mata es la dosis"(Paracelso).
- Se necesita establecer la dosis o concentración a la que las sustancias químicas se pueden utilizar de forma segura.
- Primera aproximación: El cálculo de la Concentración Letal 50 % (LC50).

Motivación I

Contexto

- Nuevas sustancias químicas en nuestro día a día.
- Para poder utilizarlas de forma segura, es necesario evaluar su toxicidad de una forma práctica.
- "La diferencia entre lo que cura y lo que mata es la dosis"(Paracelso).
- Se necesita establecer la dosis o concentración a la que las sustancias químicas se pueden utilizar de forma segura.
- Primera aproximación: El cálculo de la Concentración Letal 50 % (LC50).

Motivación I

Contexto

- Nuevas sustancias químicas en nuestro día a día.
- Para poder utilizarlas de forma segura, es necesario evaluar su toxicidad de una forma práctica.
- "La diferencia entre lo que cura y lo que mata es la dosis"(Paracelso).
- Se necesita establecer la dosis o concentración a la que las sustancias químicas se pueden utilizar de forma segura.
- Primera aproximación: El cálculo de la Concentración Letal 50 % (LC50).

Motivación I

Contexto

- Nuevas sustancias químicas en nuestro día a día.
- Para poder utilizarlas de forma segura, es necesario evaluar su toxicidad de una forma práctica.
- "La diferencia entre lo que cura y lo que mata es la dosis"(Paracelso).
- Se necesita establecer la dosis o concentración a la que las sustancias químicas se pueden utilizar de forma segura.
- Primera aproximación: El cálculo de la Concentración Letal 50 % (LC50).

Motivación I

Contexto

- Nuevas sustancias químicas en nuestro día a día.
- Para poder utilizarlas de forma segura, es necesario evaluar su toxicidad de una forma práctica.
- "La diferencia entre lo que cura y lo que mata es la dosis"(Paracelso).
- Se necesita establecer la dosis o concentración a la que las sustancias químicas se pueden utilizar de forma segura.
- Primera aproximación: El cálculo de la Concentración Letal 50 % (LC50).

Motivación II

¿Qué es el LC50?

- La concentración de una sustancia que provoca la muerte del 50 % de un grupo de organismos de prueba en un tiempo determinado.
- Permite medir la toxicidad a corto plazo de una sustancia, parámetro de referencia en Toxicología y la base de la Regulación del uso de estas sustancias.

Motivación II

¿Qué es el LC50?

- La concentración de una sustancia que provoca la muerte del 50 % de un grupo de organismos de prueba en un tiempo determinado.
- Permite medir la toxicidad a corto plazo de una sustancia, parámetro de referencia en Toxicología y la base de la Regulación del uso de estas sustancias.

Motivación II

¿Qué es el LC50?

- La concentración de una sustancia que provoca la muerte del 50 % de un grupo de organismos de prueba en un tiempo determinado.
- Permite medir la toxicidad a corto plazo de una sustancia, parámetro de referencia en Toxicología y la base de la Regulación del uso de estas sustancias.

Estudio en Toxicología Ambiental

En un estudio sobre el efecto de las nanopartículas de plata en el desarrollo de los embriones de pez cebra del Grupo de Investigación de Biología Celular de la UPV/EHU, observamos que la estimación del valor LC50 podía no estar siempre correctamente calculada.

	Maltosa	Ionic	PVP-PEI
Software	**MV**	**MV**	**MV**
R	10,367	0,033	0,460
SPSS	10,367	0,033	0,460

Cuadro: Resultados obtenidos del LC50 según el método y software utilizados

WARNING: glm.fit: fitted probabilities numerically 0 or 1 occurred

Estudio en Toxicología Ambiental

En un estudio sobre el efecto de las nanopartículas de plata en el desarrollo de los embriones de pez cebra del Grupo de Investigación de Biología Celular de la UPV/EHU, observamos que la estimación del valor LC50 podía no estar siempre correctamente calculada.

Software	**Maltosa MV**	**Ionic MV**	**PVP-PEI MV**
R	10,367	0,033	0,460
SPSS	10,367	0,033	0,460

Cuadro: Resultados obtenidos del LC50 según el método y software utilizados

WARNING: glm.fit: fitted probabilities numerically 0 or 1 occurred

Estudio en Toxicología Ambiental

En un estudio sobre el efecto de las nanopartículas de plata en el desarrollo de los embriones de pez cebra del Grupo de Investigación de Biología Celular de la UPV/EHU, observamos que la estimación del valor LC50 podía no estar siempre correctamente calculada.

Software	**Maltosa MV**	**Ionic MV**	**PVP-PEI MV**
R	10,367	0,033	0,460
SPSS	10,367	0,033	0,460

Cuadro: Resultados obtenidos del LC50 según el método y software utilizados

WARNING: glm.fit: fitted probabilities numerically 0 or 1 occurred

Separación de datos

- División completa o casi completa de la variable respuesta.
- Si solamente hay una variable independiente continua X, entonces $\exists k \in \mathbb{R}$, tal que si $X < k$ todos son éxitos y si $X > k$ todos son fracasos o viceversa.
- Consecuencia: la no existencia de estimadores máximo verosimiles adecuados.
- Firth (1993) propuso una alternativa: máxima verosimilitud penalizada.

Separación de datos

- División completa o casi completa de la variable respuesta.
- Si solamente hay una variable independiente continua X, entonces $\exists k \in \mathbb{R}$, tal que si $X < k$ todos son éxitos y si $X > k$ todos son fracasos o viceversa.
- Consecuencia: la no existencia de estimadores máximo verosimiles adecuados.
- Firth (1993) propuso una alternativa: máxima verosimilitud penalizada.

Separación de datos

- División completa o casi completa de la variable respuesta.
- Si solamente hay una variable independiente continua X, entonces $\exists k \in \mathbb{R}$, tal que si $X < k$ todos son éxitos y si $X > k$ todos son fracasos o viceversa.
- Consecuencia: la no existencia de estimadores máximo verosimiles adecuados.
- Firth (1993) propuso una alternativa: máxima verosimilitud penalizada.

Separación de datos

- División completa o casi completa de la variable respuesta.
- Si solamente hay una variable independiente continua X, entonces $\exists k \in \mathbb{R}$, tal que si $X < k$ todos son éxitos y si $X > k$ todos son fracasos o viceversa.
- Consecuencia: la no existencia de estimadores máximo verosimiles adecuados.
- Firth (1993) propuso una alternativa: máxima verosimilitud penalizada.

Separación de datos

Conc.	Total	Fall. Maltosa	Fall. Ionic	Fall. PVP-PEI
0	36	1	1	0
0,001	30	1	0	1
0,01	30	1	2	0
0,1	30	1	30	30
1	30	1	30	30
5	30	5	30	30

Cuadro: Datos mortalidad según tratamiento

OBJETIVO

- Comparar diferentes estrategias de estimación del LC50.
- Estudiar su implementación en dos software estadísticos.

OBJETIVO

- Comparar diferentes estrategias de estimación del LC50.
- Estudiar su implementación en dos software estadísticos.

OBJETIVO

- Comparar diferentes estrategias de estimación del LC50.
- Estudiar su implementación en dos software estadísticos.

Metodología

- Sea Y una variable respuesta dicotómica, donde $Y = 1$ representa que el sujeto está muerto, mientras que $Y = 0$ representa que está vivo.
- Sea X una variable continua que representa el nivel de concentración al que han sido expuestos los sujetos.

MODELO

- Se explica Y en función de X, mediante un modelo lineal generalizado tal que $g(\mu) = \beta_0 + \beta_1 * X$, siendo g la función link y $\mu = E(Y|X)$.
- Consideramos como función link el probit, por ser la más común en el ámbito de la toxicología. Es decir, $g = \Phi^{-1}(\mu)$, donde Φ representa la función de distribución acumulada para la normal estandarizada.
- Reescribimos la ecuación anterior como $p_i = E(Y = 1|X) = \Phi(\beta_0 + \beta_1 * X)$

Metodología

- Sea Y una variable respuesta dicotómica, donde $Y = 1$ representa que el sujeto está muerto, mientras que $Y = 0$ representa que está vivo.
- Sea X una variable continua que representa el nivel de concentración al que han sido expuestos los sujetos.

MODELO

- Se explica Y en función de X, mediante un modelo lineal generalizado tal que $g(\mu) = \beta_0 + \beta_1 * X$, siendo g la función link y $\mu = E(Y|X)$.
- Consideramos como función link el probit, por ser la más común en el ámbito de la toxicología. Es decir, $g = \Phi^{-1}(\mu)$, donde Φ representa la función de distribución acumulada para la normal estandarizada.
- Reescribimos la ecuación anterior como $p_i = E(Y = 1|X) = \Phi(\beta_0 + \beta_1 * X)$

Metodología

- Sea Y una variable respuesta dicotómica, donde $Y = 1$ representa que el sujeto está muerto, mientras que $Y = 0$ representa que está vivo.
- Sea X una variable continua que representa el nivel de concentración al que han sido expuestos los sujetos.

MODELO

- Se explica Y en función de X, mediante un modelo lineal generalizado tal que $g(\mu) = \beta_0 + \beta_1 * X$, siendo g la función link y $\mu = E(Y|X)$.
- Consideramos como función link el probit, por ser la más común en el ámbito de la toxicología. Es decir, $g = \Phi^{-1}(\mu)$, donde Φ representa la función de distribución acumulada para la normal estandarizada.
- Reescribimos la ecuación anterior como $p_i = E(Y = 1|X) = \Phi(\beta_0 + \beta_1 * X)$

Metodología

- Sea Y una variable respuesta dicotómica, donde $Y = 1$ representa que el sujeto está muerto, mientras que $Y = 0$ representa que está vivo.
- Sea X una variable continua que representa el nivel de concentración al que han sido expuestos los sujetos.

MODELO

- Se explica Y en función de X, mediante un modelo lineal generalizado tal que $g(\mu) = \beta_0 + \beta_1 * X$, siendo g la función link y $\mu = E(Y|X)$.
- Consideramos como función link el probit, por ser la más común en el ámbito de la toxicología. Es decir, $g = \Phi^{-1}(\mu)$, donde Φ representa la función de distribución acumulada para la normal estandarizada.
- Reescribimos la ecuación anterior como $p_i = E(Y = 1|X) = \Phi(\beta_0 + \beta_1 * X)$

Metodología

- Sea Y una variable respuesta dicotómica, donde $Y = 1$ representa que el sujeto está muerto, mientras que $Y = 0$ representa que está vivo.
- Sea X una variable continua que representa el nivel de concentración al que han sido expuestos los sujetos.

MODELO

- Se explica Y en función de X, mediante un modelo lineal generalizado tal que $g(\mu) = \beta_0 + \beta_1 * X$, siendo g la función link y $\mu = E(Y|X)$.
- Consideramos como función link el probit, por ser la más común en el ámbito de la toxicología. Es decir, $g = \Phi^{-1}(\mu)$, donde Φ representa la función de distribución acumulada para la normal estandarizada.
- Reescribimos la ecuación anterior como $p_i = E(Y = 1|X) = \Phi(\beta_0 + \beta_1 * X)$

Metodología

- El LC50 será el valor de la concentración para el cual la probabilidad de mortalidad sea $p_i = 0,5$, de forma que $X = \frac{\Phi^{-1}(0{,}5)-\beta_0}{\beta_1} = \frac{-\beta_0}{\beta_1}$
- Estimaremos β_0 y β_1 con el método más conveniente.

Metodología

- El LC50 será el valor de la concentración para el cual la probabilidad de mortalidad sea $p_i = 0,5$, de forma que $X = \frac{\Phi^{-1}(0{,}5)-\beta_0}{\beta_1} = \frac{-\beta_0}{\beta_1}$
- Estimaremos β_0 y β_1 con el método más conveniente.

Metodología: Máxima verosimilitud

- Objetivo: Maximizar la verosimilitud del modelo.
- La verosimilitud se define como $L = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}$, y la log verosimilitud como

$$lnL = l = \sum_{i=1}^{n}[y_1 lnp_i + (1-y_i)ln(1-p_i)]$$

- Para estimar β_0 y β_1, derivamos la ecuación anterior y la igualamos a 0. Como las ecuaciones obtenidas son no lineales, debemos utilizar métodos iterativos como el Fisher Scoring.
- Problemas: Excesiva desviación del sesgo en tamaños muestrales pequeños y el estimador no existe o diverge a infinito cuando se observa una "separación de datos".

Metodología: Máxima verosimilitud

- Objetivo: Maximizar la verosimilitud del modelo.
- La verosimilitud se define como $L = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}$, y la log verosimilitud como

$$lnL = l = \sum_{i=1}^{n}[y_1 lnp_i + (1-y_i)ln(1-p_i)]$$

- Para estimar β_0 y β_1, derivamos la ecuación anterior y la igualamos a 0. Como las ecuaciones obtenidas son no lineales, debemos utilizar métodos iterativos como el Fisher Scoring.
- Problemas: Excesiva desviación del sesgo en tamaños muestrales pequeños y el estimador no existe o diverge a infinito cuando se observa una "separación de datos".

Metodología: Máxima verosimilitud

- Objetivo: Maximizar la verosimilitud del modelo.
- La verosimilitud se define como $L = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}$, y la log verosimilitud como

$$lnL = l = \sum_{i=1}^{n}[y_1 lnp_i + (1-y_i)ln(1-p_i)]$$

- Para estimar β_0 y β_1, derivamos la ecuación anterior y la igualamos a 0. Como las ecuaciones obtenidas son no lineales, debemos utilizar métodos iterativos como el Fisher Scoring.
- Problemas: Excesiva desviación del sesgo en tamaños muestrales pequeños y el estimador no existe o diverge a infinito cuando se observa una "separación de datos".

Metodología: Máxima verosimilitud

- Objetivo: Maximizar la verosimilitud del modelo.
- La verosimilitud se define como $L = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}$, y la log verosimilitud como

$$lnL = l = \sum_{i=1}^{n}[y_1 lnp_i + (1-y_i)ln(1-p_i)]$$

- Para estimar β_0 y β_1, derivamos la ecuación anterior y la igualamos a 0. Como las ecuaciones obtenidas son no lineales, debemos utilizar métodos iterativos como el Fisher Scoring.
- Problemas: Excesiva desviación del sesgo en tamaños muestrales pequeños y el estimador no existe o diverge a infinito cuando se observa una "separación de datos".

Metodología: Máxima verosimilutd penalizada

- Firth (1993) propuso un método para eliminar el sesgo de los estimadores obtenidos con máxima verosimilitud.
- La función de logverosimilitud es:

$$LnL^F = l^F = l + \frac{1}{2} ln|I(\beta)|,$$

donde $I(\beta)$ es la matriz de información.

- Este método ha sido considerado una solución ideal al problema de la separación (Heinze, 2002).

Metodología: Máxima verosimilutd penalizada

- Firth (1993) propuso un método para eliminar el sesgo de los estimadores obtenidos con máxima verosimilitud.
- La función de logverosimilitud es:

$$LnL^F = l^F = l + \frac{1}{2} ln|I(\beta)|,$$

donde $I(\beta)$ es la matriz de información.

- Este método ha sido considerado una solución ideal al problema de la separación (Heinze, 2002).

Metodología: Máxima verosimilutd penalizada

- Firth (1993) propuso un método para eliminar el sesgo de los estimadores obtenidos con máxima verosimilitud.
- La función de logverosimilitud es:

$$LnL^F = l^F = l + \frac{1}{2} ln|I(\beta)|,$$

  donde $I(\beta)$ es la matriz de información.
- Este método ha sido considerado una solución ideal al problema de la separación (Heinze, 2002).

Aplicación en diferentes software

SPSS

- Máxima verosimilitud

- Máxima verosimilitud
- Máxima verosimilitud penalizada
  - Paquete `logistf`, agosto 2013
  - Paquete `brglm`, noviembre 2013

Estudio de simulación

Comparativa software

- Sea X una variable que toma valores 0, 0,1, 0,5, 1, 5 y 10.
- Sea $Y \sim Be(p)$, donde p es la probabilidad de mortalidad asociada a cada nivel de concentración X.

	Tamaño muestral	Probabilidad de mortalidad		
Conc.	N=186	Esc. 1	Esc. 2	Esc. 3
0	36	0,001	0	0
0,1	30	0,01	0	0
0,5	30	0,02	0,001	0,05
1	30	0,05	0,1	0,15
5	30	0,1	1	1
10	30	0,3	1	1

Estudio de simulación

Comparativa software

- Sea X una variable que toma valores 0, 0,1, 0,5, 1, 5 y 10.
- Sea $Y \sim Be(p)$, donde p es la probabilidad de mortalidad asociada a cada nivel de concentración X.

	Tamaño muestral	Probabilidad de mortalidad		
Conc.	N=186	Esc. 1	Esc. 2	Esc. 3
0	36	0,001	0	0
0,1	30	0,01	0	0
0,5	30	0,02	0,001	0,05
1	30	0,05	0,1	0,15
5	30	0,1	1	1
10	30	0,3	1	1

Estudio de simulación

Comparativa software

- Sea X una variable que toma valores 0, 0,1, 0,5, 1, 5 y 10.
- Sea $Y \sim Be(p)$, donde p es la probabilidad de mortalidad asociada a cada nivel de concentración X.

	Tamaño muestral	**Probabilidad de mortalidad**		
Conc.	**N=186**	**Esc. 1**	**Esc. 2**	**Esc. 3**
0	36	0,001	0	0
0,1	30	0,01	0	0
0,5	30	0,02	0,001	0,05
1	30	0,05	0,1	0,15
5	30	0,1	1	1
10	30	0,3	1	1

Simulación: Resultados

Comparativa software

	Software	Método MV	Método MVP
Escenario 1	R	13,658	13,690
	SPSS	13,658	-
Escenario 2	R	1,166	2,682
	SPSS	-	-
Escenario 3	R	1,414	2,152
	SPSS	1,414	-

Cuadro: Resultados obtenidos del LC50 según el método y software utilizados

Estudio de simulación

Comparativa métodos en R

- Bajo las mismas condiciones que en la comparativa de software.
- Se generaron 400 réplicas para un tamaño muestral $N = 186$.

Estudio de simulación

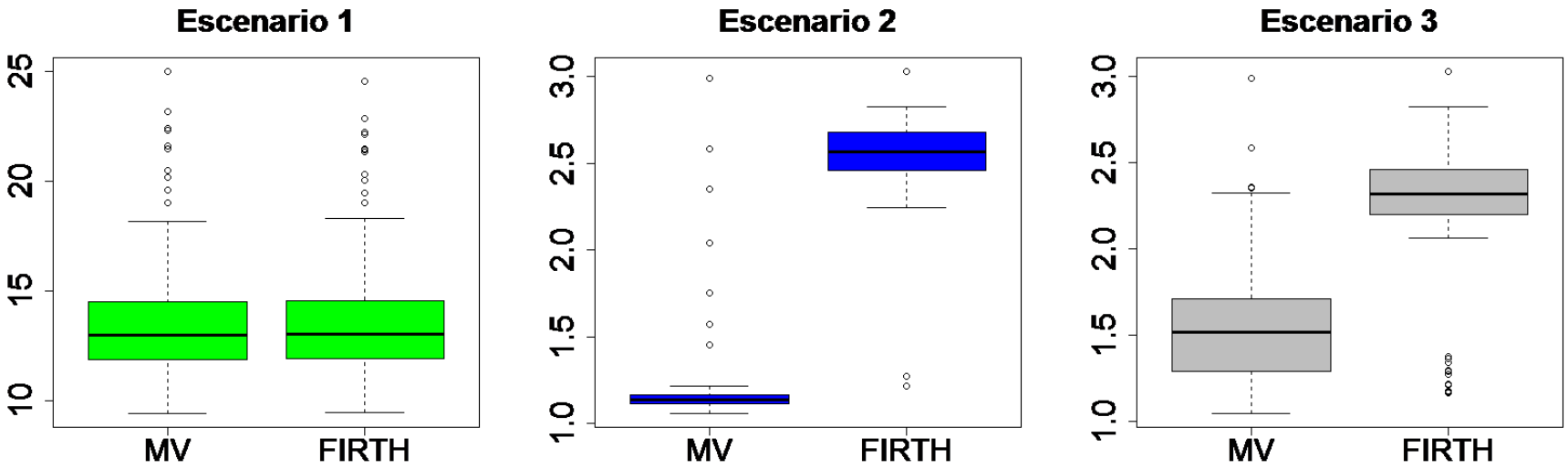
Comparativa métodos en R

- Bajo las mismas condiciones que en la comparativa de software.
- Se generaron 400 réplicas para un tamaño muestral $N = 186$.

Simulación: Resultados

Comparativa métodos en R

Figura: Resultados gráficos para el LC50 con 400 réplicas, $N = 186$



Conclusiones

- La estimación del LC50 depende del método de estimación empleado.
- Reportar el método de estimación y software utilizado.
- Especial cuidado en tamaños muestrales pequeños (menor que 25, 30 por grupo experimental) y grandes variaciones en los efectos de la concentración (escenarios tipo 2 o 3).

Conclusiones

- La estimación del LC50 depende del método de estimación empleado.
- Reportar el método de estimación y software utilizado.
- Especial cuidado en tamaños muestrales pequeños (menor que 25, 30 por grupo experimental) y grandes variaciones en los efectos de la concentración (escenarios tipo 2 o 3).

Conclusiones

- La estimación del LC50 depende del método de estimación empleado.
- Reportar el método de estimación y software utilizado.
- Especial cuidado en tamaños muestrales pequeños (menor que 25, 30 por grupo experimental) y grandes variaciones en los efectos de la concentración (escenarios tipo 2 o 3).

Trabajo a futuro

1. Considerar otros tamaños muestrales.
2. Comparar otros métodos además del propuesto por Firth para la estimación de los parámetros en situaciones de separación de datos.
3. Comparar las diferencias entre usar el link logit y el probit.

Trabajo a futuro

1. Considerar otros tamaños muestrales.
2. Comparar otros métodos además del propuesto por Firth para la estimación de los parámetros en situaciones de separación de datos.
3. Comparar las diferencias entre usar el link logit y el probit.

Trabajo a futuro

1. Considerar otros tamaños muestrales.
2. Comparar otros métodos además del propuesto por Firth para la estimación de los parámetros en situaciones de separación de datos.
3. Comparar las diferencias entre usar el link logit y el probit.

Referencias I

Heinze G. and Schemper M. (2002)
A solution to the problem of separation in logistic regression.
Statistics in Medicine, **21(1)**,2409 – 2419.

Heinze G. (2006)
A comparative investigation of methods for logistic regression with separated or nearly separated data.
Statistics in medicine,, **25(1)**,4216 – 4226.

Firth D. (1993)
Bias Reduction of maximun likelihood estimates.
Biometrika, **80**,27-38.

Godínez-Jaimes F., Ramírez-Valverde G., Reyes-Carreto R., Ariza-Hernandez F.J., Barrera-Rodriguez E.
La colinealidad y la separación en los datos en el modelo de regresión logística
Red de Revistas Científicas de América Latina, el Caribe, España y Portugal, **vol. 46, núm. 4** pp. 411-425

¡¡Muchas Gracias!!