



UNIVERSIDADE DA CORUÑA



UNIVERSIDADE  
DE VIGO

Máster en Técnicas Estadísticas

# Método MOVER para el cálculo de intervalos de confianza de los índices RERI, AP y SI en caso de interacción aditiva: Implementación en R y aplicación a datos reales

*(Prácticas en la Unidad de Epidemiología Clínica del CHUS)*

Andrés Blanco Hortas

Julio 2012



## Resumen

En el marco de las actividades que se desarrollan en la Unidad de Epidemiología Clínica del Hospital Clínico Universitario de Santiago, encaminadas principalmente a tareas de asesoría y formación en metodología de la investigación clínico-epidemiológica, presentamos el siguiente trabajo que tiene por objetivo la implementación en R del método “MOVER” (method of variance estimates recovery) para el cálculo de intervalos de confianza en casos de interacción aditiva. Siguiendo el método propuesto por Zou (2008), se implementan los códigos para el cálculo del intervalo de confianza de los índices de exceso de riesgo (RERI), la proporción atribuible (AP) y el índice de sinergia (SI), debidos a la interacción; y se aplican a un caso real que trata acerca de la interacción entre un gen (CD14) y factores ambientales (consumo de alcohol).



# Índice general

Introducción .....	7
1. Formación .....	10
1.1. Seminarios asistidos .....	10
1.2. Seminarios impartidos .....	12
2. Asesoría metodológica .....	15
2.1. Código R para un análisis descriptivo .....	16
2.2. Análisis espacio-temporal de las quejas en el CHUS .....	16
3. MOVER .....	21
3.1. Interacción aditiva .....	22
3.2. Intervalos de confianza en interacción .....	24
3.3. Método MOVER .....	26
3.3.1. Estudio caso-control .....	31
3.3.2. Estudio cohortes .....	34
3.3.3. MOVER en regresión .....	37
3.4. Estudio de simulación .....	40
3.5. Aplicación a datos reales .....	44
3.6. Resumen .....	48
4. Valoración de las prácticas .....	50
Bibliografía .....	52
Anexo .....	55
Anexo A: Función MOVER caso-control .....	55
Anexo B: Función MOVER en cohortes .....	60
Anexo C: Función MOVER en regresión .....	66



# Introducción

Este documento es una memoria de unas prácticas realizadas en la Unidad de Epidemiología Clínica del Hospital Clínico Universitario de Santiago de Compostela.

Las tareas a realizar en una Unidad de Epidemiología son de diversa índole y van desde asesoramiento metodológico a investigadores, investigación propia o formación estadística.

Así, a lo largo de la estancia de prácticas en la Unidad, realicé distintas tareas relacionadas con los puntos anteriormente citados.

Una de las actividades más importantes que se realizan, consiste en unos seminarios o charlas internas entre los miembros de la Unidad, e impartidos por ellos mismos, con la intención de una formación continuada en estadística. Luego, tanto la persona que tiene que impartir el seminario como los asistentes, mantendrán un continuo aprendizaje gracias a la exposición de temas novedosos y posteriores discusiones. Así, en el capítulo 1 se realizará un resumen de algunos de los seminarios que impartí, que versan sobre la metodología bootstrap y la programación lineal entera, y otros en los que participé como asistente.

Otra de las tareas que se llevan a cabo en la Unidad es la de asesoramiento metodológico a investigadores. En el capítulo 2 se describen dos de los trabajos de asesoramiento y análisis estadístico que se realizaron en la Unidad. En uno de ellos se colaboró con otro miembro de la Unidad en la elaboración de un script de R para realizar un rápido análisis descriptivo de un conjunto de datos. El otro trabajo que se describe versa acerca de un estudio espacio-temporal acerca de las quejas registradas en el área sanitaria de Santiago de Compostela.

Finalmente, en el capítulo 3, se expone minuciosamente la principal tarea realizada en las prácticas. Se describe un nuevo método para el cálculo de intervalos de confianza de medidas de interés en interacción (Zou, 2008). El trabajo realizado

consistió en la implementación de este nuevo método en R, un estudio de simulación para evaluar el comportamiento del nuevo método y la aplicación del método a unos datos reales.

El código R de las funciones programadas para la utilización del método MOVER se pueden consultar en el Anexo. El Anexo A contiene las funciones para un estudio caso-control, el Anexo B para un estudio de cohortes y el Anexo C las funciones para el MOVER en regresión.

Estos trabajos descritos fueron las principales tareas que se realizaron durante la estancia en la Unidad y las que se exponen a lo largo del documento.



# Capítulo 1

## Formación

Ante la importancia que tiene un continuo aprendizaje y conocimiento de nuevas metodologías estadísticas, en la Unidad de Epidemiología Clínica del CHUS se organizan seminarios con una periodicidad mensual. Estos seminarios serán impartidos por los propios miembros de la Unidad.

En este capítulo se realizará un pequeño resumen tanto de los seminarios en los que se participó como asistente, como de los seminarios impartidos.

### 1.1. Seminarios asistidos

En esta sección se nombrarán los seminarios en los que se participó en calidad de asistente.

i) *Confeción de un artículo médico.*

Este seminario fue impartido por el Dr. Francisco Gude Sampedro. En él se exponían los pasos que se han de seguir para escribir un artículo en medicina, así como las distintas secciones que debe tener.

Una primera parte del seminario estuvo centrada en presentar las principales revistas epidemiológicas, describir el término *factor de impacto* de una revista y dar a conocer webs donde realizar consultas acerca de revistas médicas.

La segunda parte del seminario estuvo centrada en describir como se escribe un artículo en medicina y las distintas partes que debe tener. Estas partes son:

- 1°- Título.
- 2°- Resumen.
- 3°- Palabras clave.
- 4°- Introducción.
- 5°- Material y métodos.
- 6°- Resultados.
- 7°- Discusión.
- 8°- Agradecimientos.
- 9°- Bibliografía.

Se realizó una breve exposición de cada punto, explicando la manera correcta de redactarlo y la extensión adecuada que debe tener.

*ii) Introducción a las series de tiempo.*

El seminario fue impartido por Fiz Lagoa Labrador, alumno del máster en técnicas estadísticas realizando unas prácticas en la Unidad.

En este seminario se realizaba una introducción a las series de tiempo. Fue expuesta la metodología necesaria para realizar un análisis descriptivo de una serie temporal y una descomposición de la misma.

Se definió los procesos estocásticos y en particular los conceptos de ruido blanco, procesos autorregresivos y procesos de medias móviles. Se expuso la metodología para identificar cada tipo de proceso.

Por último se expuso el análisis de una serie temporal.

*iii) Confusión e interacción.*

Este seminario fue impartido por el Dr. Francisco Gude Sampedro. En él se trataban los importantes conceptos de interacción y confusión y la forma de identificarlos.

Después de definir cada concepto se expusieron varios ejemplos prácticos en los que identificar los dos procesos.

## 1.2. Seminarios impartidos

En esta sección se nombrarán los seminarios en los que se participó en calidad de ponente.

### *i) Introducción a la programación lineal y entera.*

En esta charla se expusieron las definiciones básicas de la investigación operativa y en concreto de la programación lineal y entera.

Se presentó el planteamiento de un problema de programación lineal, sus propiedades y el principal algoritmo para su resolución, el simplex. Se expusieron unos ejemplos sencillos de programación lineal.

Se trató la complejidad computacional y la importancia de la utilización de algoritmos de orden de velocidad polinomial.

En el último bloque de la presentación se presentaron los problemas de flujo a coste mínimo, que son una de las familias de problemas “sencillos” de resolver más grandes. Se expusieron los problemas clásicos de la investigación operativa, tales como: el problema del transporte, el problema del camino más corto o el problema de asignación.

Finalmente se resolvió, a modo de ejemplo de la utilización de la programación lineal y entera, un problema relacionado con el cariotipado de cromosomas.

### *ii) Métodos bootstrap.*

Con este seminario se pretendió exponer la idea del remuestreo bootstrap y su aplicación al cálculo de intervalos de confianza.

Se realizó una breve introducción histórica del método exponiendo el problema para el que Efron lo inventó, la estimación de la varianza de un estimador.

En una segunda parte de la exposición se presentaron las metodologías de distintos métodos bootstrap como el bootstrap percentil, bootstrap percentil estudentizado, bootstrap en regresión,...

Dentro de los bootstrap en regresión se trató el bootstrap homocedástico, su metodología y las limitaciones que tiene (que el modelo de regresión sea heterocedástico). Otro bootstrap en regresión del que se habló fue el Wild Bootstrap, con el cual se evita el problema de la heterocedasticidad de los datos.

Finalmente se presentaron varios casos concretos donde utilizar los distintos métodos bootstrap y mostrar como implementarlos en R. Al mismo tiempo, con estos ejemplos, se puso de manifiesto las limitaciones o problemas que podían tener distintos bootstrap y como otros podían subsanarlos.



## Capítulo 2

### Asesoría metodológica

A lo largo de la estancia en la Unidad de Epidemiología Clínica del Complejo Hospitalario Universitario de Santiago de Compostela, se realizaron tareas de asesoría estadística a investigadores, así como actividades internas de la Unidad.

Las temáticas que normalmente se abordan en la Unidad tratan temas donde el análisis estadístico versa sobre: Regresión logística, análisis de supervivencia, pruebas diagnósticas, análisis espacio-temporal, análisis de datos categóricos, modelos mixtos,...

Una parte importante del trabajo a realizar en la unidad consiste en el continuo desarrollo de software e implementación de códigos R, para la realización de análisis estadísticos.

En este capítulo se resumirá parte de estos trabajos realizados, exponiendo los resultados obtenidos.

En la sección 2.1 se expondrá la confección de un código R para realizar un rápido análisis descriptivo de un conjunto de datos, desarrollado por un compañero de la Unidad (Lagoa Labrador, Fiz)

En la sección 2.2 se resumirá los resultados estadísticos obtenidos en un estudio espacio-temporal.

## 2.1. Código R para un análisis descriptivo

Una de las tareas a las que hay que enfrentarse a la hora de realizar un análisis estadístico, consiste en un análisis descriptivo previo de los datos. Aunque es una tarea pesada no deja de ser necesaria e importante. Por este motivo, una de las tareas a realizar en las prácticas consistió en crear un código R para un rápido análisis descriptivo.

La creación de las funciones R para el análisis descriptivo fue llevado a cabo por Fiz Lagoa Labrador. Se realizó una discusión previa junto a él sobre la metodología a utilizar así como asesoramiento en la confección del código.

Las funciones programadas realizan un análisis descriptivo de la variable, sea continua o categórica. También se programaron funciones que realizan análisis estándares entre variables categóricas, continuas o entre factores de una variable continua sobre los factores de una categórica.

El código de las funciones así como una exhaustiva descripción de la metodología utilizada será descrita por Fiz Lagoa Labrador en su TFM.

## 2.2. Análisis espacio-temporal de las quejas en el CHUS

En esta sección se presentarán los resultados obtenidos en un análisis estadístico espacio-temporal acerca de quejas presentadas en el Área Sanitaria de Santiago de Compostela. Así, se estudiará la evolución temporal de las quejas y su posible asociación con el área geográfica de procedencia y con factores sociodemográficos.

El análisis temporal fue desarrollado por un compañero de la Unidad (Fiz Lagoa Labrador). Por tanto, en esta sección, nos centraremos en el análisis espacio-temporal.

La mayoría de las quejas son formuladas por gente que tiene su domicilio en el área sanitaria de Santiago de Compostela. Esta área está formado por 43 concellos pertenecientes a las provincias de Coruña y Pontevedra.

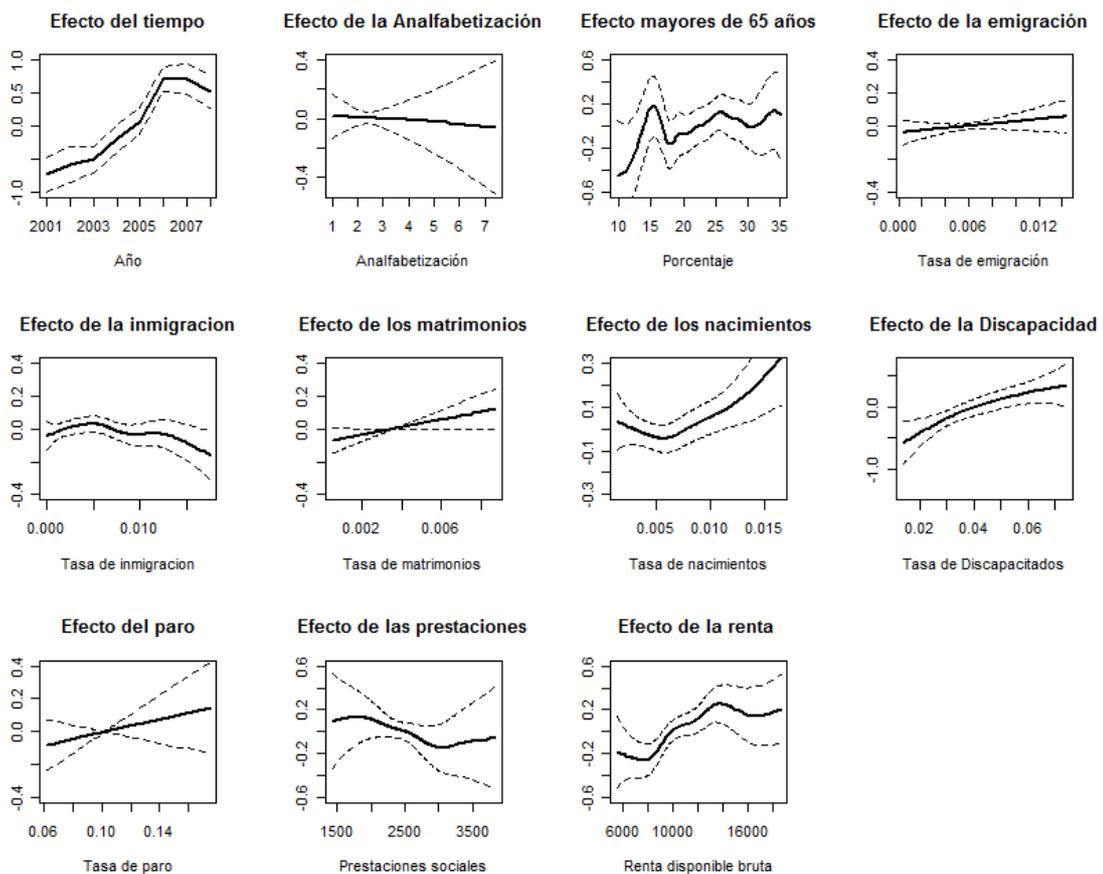
Se presentan los resultados de un modelo espacio-temporal ajustado, además de por el tiempo y el espacio, por otras covariables como: Índice de analfabetización, tasa de emigración e inmigración, porcentaje de personas mayores de 65 años, tasa de nuevos matrimonios y nacimientos, número de discapacitados e índices

socioeconómicos tales como tasa de paro, prestaciones sociales y la renta disponible bruta de cada concello.

En la figura 1 se muestran los efectos suavizados de cada covariable. Puede observarse como el tiempo tiene un efecto creciente y no lineal. Los efectos de la tasa de analfabetización, porcentaje de mayores de 65 años, la tasa de emigración y las prestaciones sociales no son significativos.

El efecto de la tasa de inmigración sí es significativo y decreciente. En cambio, los efectos de la tasa de nuevos matrimonios, tasa de nacimientos, tasa de discapacitados, el número de parados y la renta bruta disponible tienen un efecto significativo y creciente. Los efectos de la tasa de nuevos matrimonios, tasa de discapacidad y número de parados, siendo significativos, son lineales.

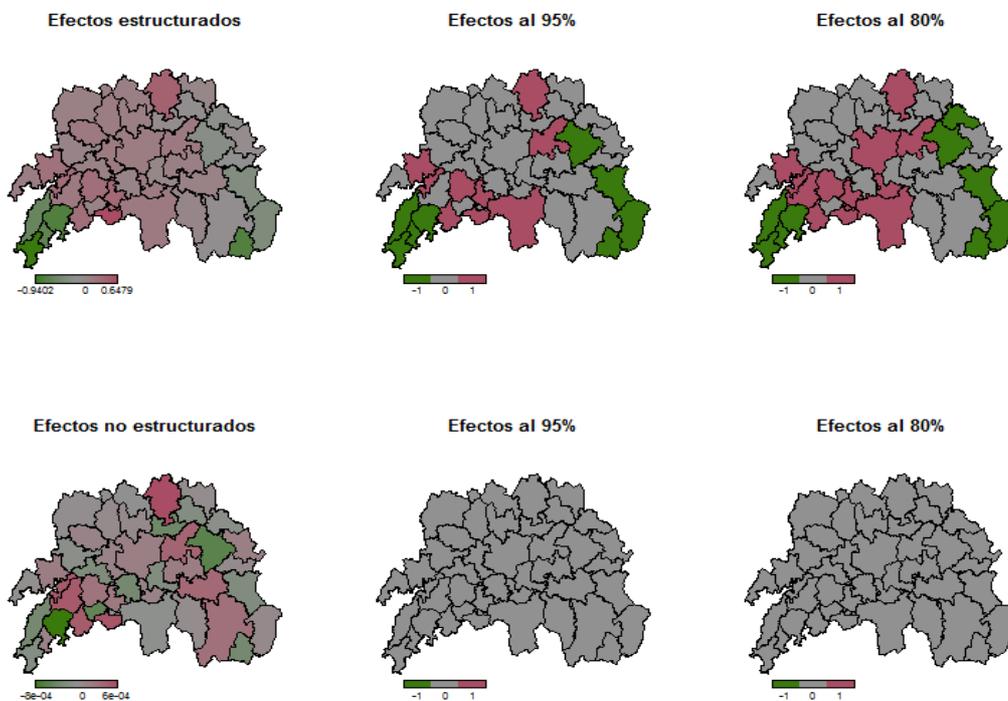
**Figura 1: Efecto de las covariables en el modelo espacio-temporal.**



En la figura 2 se presentan los efectos espaciales del modelo. Se puede observar que se producen un menor número de quejas en las zonas del Suroeste y del Sudeste del área sanitaria de Santiago de Compostela. Es decir, en la península del Barbanza, concretamente en los concellos de Porto do Son, Ribeira, Pobra do Caramiñal y Boiro. La otra zona en la que se producen menos quejas es en la comarca de Deza, en los concellos de Lalín, Dozón, Rodeiro, Agolada, Vila de Cruces y Silleda. Estos efectos además son significativos al 95% de confianza.

La zona central del área sanitaria de Santiago de Compostela es en la que más quejas se producen, siendo también significativos estos efectos.

**Figura 2: Efectos espaciales en el modelo espacio-temporal**



La distribución espacial de las quejas tiene un comportamiento muy marcado, con dos zonas del área sanitaria de Santiago de Compostela con significativamente menor número de quejas. Estas dos zonas son la península del Barbanza y la comarca de Deza. El menor número de quejas puede ser achacable al hecho de una menor utilización del Complejo Hospitalario, ya sea por la existencia de otros centros de salud en la zona o por la lejanía a Santiago de Compostela.

En cuanto a la influencia de factores socioeconómicos, a la vista de los resultados, se puede concluir que al aumentar la emigración, nuevos matrimonios, nacimientos y renta son factores que hacen aumentar el número de quejas. En cambio, el aumento de la inmigración y de las prestaciones sociales produce una disminución en el número de quejas.



## Capítulo 3

# MOVER

En este capítulo se tratará la interacción aditiva y se presentará el método MOVER para el cálculo de los intervalos de confianza (IC) de las cantidades de interés en la interacción.

Se realizará una pequeña exposición del significado de interacción y la diferencia entre interacción biológica y matemática. Se remarcará la importancia que tiene la interacción en los estudios epidemiológicos actuales donde se está a tratar en muchos casos la interacción gen-medio ambiente.

Existen unas ciertas cantidades de interés en la interacción que la definen. El cálculo de de sus intervalos de confianza es de gran importancia. En este capítulo se expondrá los IC propuestos por Hosmer-Lemeshow y una pequeña crítica a ellos. Se presentará otro método para el cálculo de los IC, el método MOVER y su implementación en R.

Para comprobar el buen funcionamiento de este método, se realizará un estudio de simulación, con distintos escenarios, y se compararán los resultados obtenidos con el método MOVER y con distintos métodos bootstrap.

Por último, se utilizará el método MOVER para calcular los IC de los índices de interés en interacción con unos datos reales. Los datos, que versan acerca de la interacción gen-medio ambiente, son datos de un estudio previo en el que se calculó el IC de un índice de la interacción mediante el método de Hosmer-Lemeshow. En este capítulo se calcularán por el método MOVER y se compararán ambos IC.

### 3.1. Interacción aditiva

En las últimas décadas muchos autores han descrito la importancia de la interacción en los estudios epidemiológicos. Una de las vías de investigación abierta en los últimos años, y de una gran importancia (Tiret, 2002), es la interacción de factores ambientales-gen. En esta sección trataremos de definir el término interacción, que como veremos no es una tarea trivial, y su utilidad.

Una primera dificultad que surge a la hora de definir la interacción es que es un concepto que define dos cosas diferentes, a las que llamaremos *interacción estadística* e *interacción biológica*. Son varios los autores que han tratado la cuestión de la confusión que crea el concepto de interacción y sus dos significados (Rothman, 1980; Ahlbom, 2005).

*Interacción estadística* es el contexto donde el término interacción tiene un significado establecido y específico en estadística, donde se usa para caracterizar una situación en la que los efectos no son aditivos. Es decir, la *interacción estadística* se refiere a la necesidad de incluir una interacción, o producto, como término de un modelo estadístico para el buen ajuste de unos datos.

*Interacción biológica*, en cambio, se puede definir como la interdependencia entre dos o más causas para producir un efecto. Es decir, muchas causas de enfermedad dependen de la presencia de otro factor para acrecentar o disminuir su efecto, y esto es lo que se entiende como *interacción biológica*.

A lo largo de todo el capítulo lo que se tratará será la *interacción biológica*. Como identificarla, representarla y medirla serán tareas que se tratarán en el capítulo. De aquí en adelante se denotará la *interacción biológica* por interacción a secas.

Por tanto, el término interacción se usa en epidemiología para describir una situación en donde dos o más factores de riesgo modifican el efecto cada uno del otro en relación a un cierto efecto.

Así, interacción, significa que el efecto de una variable difiere dependiendo de la presencia o ausencia de otra variable (variable modificadora). Si la presencia de la variable modificadora acentúa el efecto de la otra variable se dirá que es una situación de *sinergismo* (interacción positiva). En cambio, si la presencia de la variable modificadora hace disminuir o desaparecer el efecto de la otra variable se dirá que es una situación de *antagonismo* (interacción negativa).

Un mínimo de tres factores son necesarios para que exista el fenómeno de interacción. Designaremos por A al principal factor de riesgo, la variable dependiente como Y, y el tercer factor (el posible efecto modificador) como Z. Se puede definir interacción de dos formas diferentes:

- Interacción ocurre cuando el efecto del factor de riesgo A sobre la variable Y no es homogéneo en los estratos formados por la tercera variable Z.
- Interacción ocurre cuando el efecto conjunto de A y Z observado difiere de lo que cabría esperar por la suma (interacción aditiva) o multiplicación (interacción multiplicativa) de los efectos de A y Z por separado.

Esta segunda acepción tiene una interpretación más intuitiva y conlleva una forma más sencilla de medir la interacción. En este sentido, puede hablarse de dos tipos de interacción, *interacción aditiva* e *interacción multiplicativa*.

La presencia de *interacción aditiva* se evalúa comparando los efectos observados con los esperados de los factores A y Z. Denotemos por  $AR$  el riesgo atribuible a un factor de riesgo. Se denota por A y Z la presencia de los factores de riesgo y por  $\bar{A}$  y  $\bar{Z}$  su ausencia. Así, la ausencia de interacción vendrá definida por la igualdad entre el riesgo atribuible observado y el riesgo atribuible esperado en presencia de los dos factores. El riesgo atribuible esperado se define como:

$$Esp AR_{AZ} = Obs AR_{A\bar{Z}} + Obs AR_{\bar{A}Z}$$

Es decir, el riesgo atribuible esperado debido a los factores A y Z es igual a la suma de los riesgos atribuibles observados de los factores A y Z por separado.

En caso de no darse la igualdad  $Esp AR_{AZ} = Obs AR_{AZ}$ , se diría que existe una situación de interacción.

La *interacción multiplicativa* se evalúa observando si el riesgo relativo ( $RR$ ) observado en presencia de los dos factores A y Z es igual al esperado, definido como:

$$Esp RR_{AZ} = Obs RR_{A\bar{Z}} \times Obs RR_{\bar{A}Z}$$

La igualdad entre el riesgo relativo esperado y el observado en presencia de los dos factores será indicio de no interacción, y si hay diferencias entonces se hablará de sí interacción.

A lo largo del capítulo trataremos la interacción aditiva, que es la que tiene una interpretación más directa y fácil de usar en el entorno de la epidemiología. Definiremos tres índices de interés en la interacción aditiva y como estos la definen.

Rothman discute en su libro (Rothman, 1986) acerca de la estimación de la interacción aditiva, usando modelos de regresión logística, definida como la aditividad de efectos de dos factores. Define tres índices de interés, el exceso de riesgo relativo (RERI), la proporción atribuible (AP) y el índice de sinergia (SI) debidos a la interacción.

Se denota por A y B la presencia de los factores de riesgo y por  $\bar{A}$  y  $\bar{B}$  su ausencia. Sea  $RR$  el riesgo relativo, se definen los tres índices como sigue:

$$RERI = RR(AB) - RR(A\bar{B}) - RR(\bar{A}B) + 1$$

$$AP = \frac{RERI}{RR(AB)} = \frac{RR(AB) - RR(A\bar{B}) - RR(\bar{A}B) + 1}{RR(AB)}$$

$$SI = \frac{RR(AB) - 1}{RR(A\bar{B}) + RR(\bar{A}B) - 2}$$

La ausencia de interacción viene definida por  $RERI = AP = 0$  y  $SI=1$ .

Por tanto, es de gran importancia poder calcular intervalos de confianza (IC) para estos índices. El propio Rothman propuso un método para calcular el intervalo de confianza, aunque los más extendidos en la literatura son los presentados por Hosmer-Lemeshow.

## 3.2. Intervalos de confianza en interacción

En esta sección presentaremos los métodos propuestos por Rothman (Rothman, 1976) y Hosmer-Lemeshow (Hosmer-Lemeshow, 1992) para el cálculo de los IC de los índices en interacción definidos en la sección anterior (RERI, AP y SI).

Como se ha comentado en la sección anterior, Rothman definió los índices de interés en la interacción. En un artículo (Rothman, 1976) presentaba una metodología, basada en la distribución log-normal, para el cálculo del IC para el índice de sinergia. Si  $S$  denota este índice, los límites del IC serán:

$$S_L = \exp\left(\ln(\hat{S}) - Z_{1-\alpha/2} SE(\ln(\hat{S}))\right)$$

$$S_U = \exp\left(\ln(\hat{S}) + Z_{1-\alpha/2} SE(\ln(\hat{S}))\right)$$

donde  $SE(\ln(\hat{S}))$  es el error estándar del logaritmo natural de  $\hat{S}$ , y  $Z_{1-\alpha/2}$  es el cuantil  $\alpha/2$  de la distribución normal estándar.

Algunos años más tarde, Hosmer y Lemeshow presentaron otro método para el cálculo de los intervalos de confianza para estos índices (Hosmer and Lemeshow, 1992). Este método, basado en el método delta, ha sido el más extendido en la literatura.

Los autores proponen una aproximación asintótica simple (SA) para el cálculo de los intervalos de confianza. Si  $S$  denota la cantidad de interés, su IC será:

$$\hat{S} \mp 1.96\sqrt{\hat{v}(\hat{S})}$$

Para calcular la varianza estimada se usa el método delta. Si  $h(\hat{\theta})$ ,  $\hat{\theta} = (\theta_1, \dots, \theta_n)$ , denota una reparametrización de  $S$  la varianza estimada será:

$$v[h(\hat{\theta})] = \sum_{i=1}^n h_i^2 \hat{v}(\hat{\theta}_i) + 2 \sum_{i \neq j} h_i h_j \text{cov}(\hat{\theta}_i, \hat{\theta}_j); \quad h_i = \frac{\partial h}{\partial \theta_i}$$

Proponen obtener las OR's a partir de un modelo de regresión logístico. Así, puede realizarse una reparametrización de los índices,  $h(\hat{\theta})$ , donde  $\hat{\theta} = (\theta_1, \dots, \theta_n)$  serán los coeficientes de la regresión. En esta situación, el cálculo de las varianzas y las covarianzas es sencillo.

Los propios autores realizaron un estudio de simulación (Assman, Hosmer, Lemeshow, 1996) para comprobar el buen funcionamiento del método. En el estudio comparan los resultados con los obtenidos por tres métodos bootstrap. Los resultados obtenidos por los IC asintóticos son buenos, sin ser espectaculares, siendo claramente peores que los obtenidos por uno de los métodos bootstrap (bootstrap percentil).

Una de las grandes limitaciones de los IC asintóticos es que se fuerzan a ser simétricos. En el estudio de simulación que realizan Hosmer-Lemeshow esta circunstancia queda de manifiesto. En la gran mayoría de los escenarios que presentan, el error de cobertura está mucho más balanceado hacia un lado del intervalo que hacia el otro. En cambio, el bootstrap percentil, obtiene un error de cobertura bastante simétrico, lo que es deseable.

A lo largo del capítulo se expondrá una alternativa para la construcción de los intervalos de confianza para las medidas de interés en la interacción aditiva. Esta alternativa, propuesta por Zou (Zou, 2008), es el resultado de incorporar límites de confianza asimétricos para los riesgos relativos. Nos referiremos a este nuevo método como MOVER (method of variance estimates recovery), por el hecho de que utiliza, o recupera, las varianzas estimadas para el cálculo de los límites de los intervalos de confianza para los riesgos relativos, y las utiliza para construir el IC para cada índice en la interacción.

### 3.3. Método MOVER

En esta sección se presenta la metodología para el cálculo de los intervalos de confianza para las medidas de interés en la interacción mediante el método *MOVER*.

Sean A y B dos factores de riesgo, la presencia o ausencia de estos factores se denota por 1 o 0 respectivamente. Por tanto pueden darse cuatro combinaciones, 11, 10, 01 y 00. Sea  $RR_{ij}$  el riesgo relativo correspondiente a la combinación  $ij$  ( $i,j=0,1$ ). Se definen el RERI, AP y SI como sigue:

$$RERI = RR_{11} - RR_{10} - RR_{01} + 1$$

$$AP = \frac{RERI}{RR_{11}} = \frac{1}{RR_{11}} - \frac{RR_{10}}{RR_{11}} - \frac{RR_{01}}{RR_{11}} + 1$$

$$SI = \frac{RR_{11} - 1}{RR_{10} + RR_{01} - 2}$$

Hay que tener en cuenta que en un estudio de caso-control se debe substituir los riesgos relativos (RR) por odds ratios (OR). En los estudios de cohortes se trabajará con RR, al ser posible utilizarlos, ya que las odds ratios tienden a exagerar los riesgos relativos (Knol, 2007).

Los índices RERI y AP pueden reparametrizarse ambos como

$$\theta_1 - \theta_2 - \theta_3 + 1$$

Donde  $\theta_1 = RR_{11}$ ,  $\theta_2 = RR_{10}$ ,  $\theta_3 = RR_{01}$  para el RERI, y  $\theta_1 = \frac{1}{RR_{11}}$ ,  $\theta_2 = \frac{RR_{10}}{RR_{11}}$ ,  $\theta_3 = \frac{RR_{01}}{RR_{11}}$  para la AP.

Si aplicamos una transformación logaritmo a SI:

$$\ln SI = \ln(RR_{11} - 1) - \ln(RR_{10} + RR_{01} - 2)$$

Por tanto, el logaritmo del índice SI puede reparametrizarse como

$$\theta_1 - \theta_2$$

La idea básica del método MOVER para el cálculo de los intervalos de confianza consiste en no perder la información que se obtiene al ser posible el cálculo

de intervalos de confianza asimétricos para los RR. La estrategia es usar estos intervalos de confianza, sin destruir su asimetría, para construir unos intervalos asimétricos para una combinación lineal de RR.

A continuación se mostrará la construcción del intervalo de confianza bilateral al  $(1-\alpha)\%$  para  $\theta_1 + \theta_2$ . Los IC para cualquier combinación lineal se obtienen por inducción a partir de estos.

Por tanto, consideremos la combinación de RR  $\theta_1 + \theta_2$ . Se consideran primeramente las dos estimaciones  $\hat{\theta}_1$  y  $\hat{\theta}_2$  independientes. Entonces los límites inferior y superior serán

$$L = (\hat{\theta}_1 + \hat{\theta}_2) - z_{\alpha/2} \sqrt{v(\hat{\theta}_1) + v(\hat{\theta}_2)}$$

$$U = (\hat{\theta}_1 + \hat{\theta}_2) + z_{\alpha/2} \sqrt{v(\hat{\theta}_1) + v(\hat{\theta}_2)}$$

Donde  $z_{\alpha/2}$  denota el cuantil  $\alpha/2$  de la distribución normal estándar.

Los términos  $v(\hat{\theta}_i)$  ( $i=1,2$ ) pueden ser calculados por dos aproximaciones: Una asumiendo que  $v(\hat{\theta}_i)$  es independiente de  $\theta_i$ , y otra que no hace esta suposición. Los intervalos de confianza para  $\theta_i$  de la primera aproximación son simétricos mientras que los de la segunda son asimétricos y normalmente mejores (Efron y Tibshirani, 1993).

En el método MOVER se utilizarán estos intervalos asimétricos y a partir de ellos se construirán los límites del IC para una combinación lineal de RR.

De la relación existente entre intervalos de confianza y contraste de hipótesis, el intervalo debe contener todos los valores de los parámetros que no son rechazados por el test. En otras palabras, L (el límite inferior) es el mínimo valor de  $\theta_1 + \theta_2$  satisfaciendo

$$\frac{[(\hat{\theta}_1 + \hat{\theta}_2) - (\theta_1 + \theta_2)]^2}{v(\hat{\theta}_1 + \hat{\theta}_2)} = z_{\alpha/2}^2$$

Para reflejar el hecho de que la varianza de  $\hat{\theta}_1 + \hat{\theta}_2$  en general depende del verdadero valor de  $\theta_1 + \theta_2$ , la varianza estimada para obtener L debe ser estimada en una vecindad de L o  $\min(\theta_1) + \min(\theta_2)$ . Entre los límites de los dos intervalos de confianza para  $\theta_i$  ( $l_1, u_1, l_2, u_2$ ), serían  $\theta_1 + \theta_2 = l_1 + l_2$ . Lo cual implica que la varianza puede ser estimada con  $\theta_1 = l_1$  y  $\theta_2 = l_2$ .

De nuevo, por la dualidad de IC y contraste de hipótesis

$$l_i = \hat{\theta}_i - z_{\alpha/2} \sqrt{\hat{v}(\hat{\theta}_i)}, \quad i = 1, 2$$

que recupera la varianza estimada como

$$\hat{v}(\hat{\theta}_i) = \frac{(\hat{\theta}_i - l_i)^2}{z_{\alpha/2}^2}$$

Introduciendo estas estimaciones de las varianzas en los límites de  $\theta_1 + \theta_2$  se tiene que

$$L = (\hat{\theta}_1 + \hat{\theta}_2) - \sqrt{(\hat{\theta}_1 - l_1)^2 + (\hat{\theta}_2 - l_2)^2}$$

$$U = (\hat{\theta}_1 + \hat{\theta}_2) + \sqrt{(u_1 - \hat{\theta}_1)^2 + (u_2 - \hat{\theta}_2)^2}$$

Estos límites están contruidos en caso de independecia entre  $\hat{\theta}_1$  y  $\hat{\theta}_2$ . En caso de existir dependecia entre los parámetros se pueden extender como sigue. Sea  $r$  la correlación entre  $\hat{\theta}_1$  y  $\hat{\theta}_2$ , luego los límites del intervalo de confianza son

$$L = (\hat{\theta}_1 + \hat{\theta}_2) - \sqrt{(\hat{\theta}_1 - l_1)^2 + (\hat{\theta}_2 - l_2)^2 + 2r(\hat{\theta}_1 - l_1)(\hat{\theta}_2 - l_2)}$$

$$U = (\hat{\theta}_1 + \hat{\theta}_2) + \sqrt{(u_1 - \hat{\theta}_1)^2 + (u_2 - \hat{\theta}_2)^2 + 2r(u_1 - \hat{\theta}_1)(u_2 - \hat{\theta}_2)}$$

Estos serían los límites del IC para la combinación lineal  $\theta_1 + \theta_2$  en caso de existir dependecia entre los parámetros.

Los límites del IC para cualquier combinación lineal de RR se obtienen por inducción. Así, el intervalo de confianza  $(L, U)$  para una combinación lineal de RR  $\sum_i c_i \theta_i$  sería:

$$L = \sum_i c_i \hat{\theta}_i - \sqrt{\sum_i [c_i \hat{\theta}_i - \min(c_i l_i, c_i u_i)]^2 + 2 \text{cov}_L}$$

$$U = \sum_i c_i \hat{\theta}_i + \sqrt{\sum_i [c_i \hat{\theta}_i - \max(c_i l_i, c_i u_i)]^2 + 2 \text{cov}_U}$$

Donde

$$\text{cov}_L = \sum_{i < j} \text{sgn}(c_i c_j) r_{ij} [c_i \hat{\theta}_i - \min(c_i l_i, c_i u_i)] \times [c_j \hat{\theta}_j - \min(c_j l_j, c_j u_j)]$$

$$\text{cov}_L = \sum_{i < j} \text{sgn}(c_i c_j) r_{ij} [c_i \hat{\theta}_i - \max(c_i l_i, c_i u_i)] \times [c_j \hat{\theta}_j - \max(c_j l_j, c_j u_j)]$$

y

$$\text{sgn}(c_i c_j) = \begin{cases} +1, & c_i c_j > 0 \\ -1, & c_i c_j < 0 \end{cases}$$

El RERI y AP se reparametrizar como  $\theta_1 - \theta_2 - \theta_3 + 1$ , por lo expuesto anteriormente, los límites de sus intervalos de confianza serán:

$$\begin{aligned} L = & 1 + \hat{\theta}_1 - \hat{\theta}_2 - \hat{\theta}_3 - [(\hat{\theta}_1 - l_1)^2 + (u_2 - \hat{\theta}_2)^2 + (u_3 - \hat{\theta}_3)^2 \\ & - 2r_{12}(\hat{\theta}_1 - l_1)(u_2 - \hat{\theta}_2) - 2r_{13}(\hat{\theta}_1 - l_1)(u_3 - \hat{\theta}_3) \\ & + 2r_{23}(u_2 - \hat{\theta}_2)(u_3 - \hat{\theta}_3)]^{1/2} \end{aligned} \quad (1)$$

$$\begin{aligned} U = & 1 + \hat{\theta}_1 - \hat{\theta}_2 - \hat{\theta}_3 + [(u_1 - \hat{\theta}_1)^2 + (\hat{\theta}_2 - l_2)^2 + (\hat{\theta}_3 - u_3)^2 \\ & - 2r_{12}(l_1 - \hat{\theta}_1)(\hat{\theta}_2 - l_2) - 2r_{13}(u_1 - \hat{\theta}_1)(\hat{\theta}_3 - l_3) \\ & + 2r_{23}(\hat{\theta}_2 - l_2)(\hat{\theta}_3 - l_3)]^{1/2} \end{aligned}$$

El logaritmo del índice de sinergia se reparametriza como  $\theta_1 - \theta_2$ , los límites de su intervalo de confianza serán

$$L = \hat{\theta}_1 - \hat{\theta}_2 - [(\hat{\theta}_1 - l_1)^2 + (\hat{\theta}_2 - l_2)^2 + 2r(\hat{\theta}_1 - l_1)(\hat{\theta}_2 - l_2)]^{1/2} \quad (2)$$

$$U = \hat{\theta}_1 - \hat{\theta}_2 + [(u_1 - \hat{\theta}_1)^2 + (u_2 + \hat{\theta}_2)^2 + 2r(u_1 - \hat{\theta}_1)(u_2 + \hat{\theta}_2)]^{1/2}$$

Para obtener los límites del IC para el SI, bastará con aplicar la función exponencial a estos límites.

Puede verse (Zou, 2008) que estos límites del IC para el SI calculados mediante el método MOVER son idénticos a los propuestos por Rothman (Rothman, 1976), basados en la distribución log-normal, los cuales tienen un cálculo mucho más sencillo. La metodología sería la siguiente: Si la cantidad de interés es  $S$ , los límites de su IC serían:

$$S_L = \exp(\ln(\hat{S}) - Z_{1-\alpha/2} \sqrt{v(\ln \hat{S})}) = \hat{S} \times \exp(-Z_{1-\alpha/2} \sqrt{v(\ln \hat{S})})$$

$$S_L = \exp(\ln(\hat{S}) + Z_{1-\alpha/2} \sqrt{v(\ln \hat{S})}) = \hat{S} \times \exp(+Z_{1-\alpha/2} \sqrt{v(\ln \hat{S})})$$

Por tanto, para obtener los IC por el MOVER para RERI, AP y SI se necesitará realizar el cálculo de los IC para los parámetros (obtener  $l_i$  y  $u_i$ ) así como las correlaciones entre ellos.

Para los parámetros no se asumirán intervalos de confianza simétricos, sino que se utilizarán los basados en la distribución log-normal, más sensibles y que pueden ser asimétricos (Rothman, 1976).

$$(l_i, u_i) = \theta_i \exp\left(\mp z_{\alpha/2} \sqrt{v(\ln \theta_i)}\right)$$

Las correlaciones se pueden obtener como  $r_{ij} = \frac{\text{cov}(\theta_i, \theta_j)}{\sqrt{v(\theta_i)v(\theta_j)}}$

Para el cálculo de las correlaciones, por tanto, es necesario obtener las varianzas de los parámetros y las covarianzas entre ellos. Mediante la aplicación el método delta, se tiene que si  $h(\theta) = h(\theta_i / i = 1, \dots, n)$ , la varianza de  $h$  será:

$$v[h(\theta)] = \sum_{i=1}^n h_i^2 v(\theta_i) + 2 \sum_{i \neq j} h_i h_j \text{cov}(\theta_i, \theta_j)$$

donde  $h_i = \frac{\partial h}{\partial \theta_i}$ .

Las estimaciones de las varianzas y las covarianzas de los parámetros, se realizarán mediante el método delta expuesto, teniendo en cuenta que los parámetros son RR o factores entre RR. A partir de ellas se calcularán las correlaciones entre ellos como sigue (Zou, 2008):

$$\text{cov}(RR_{10}, RR_{01}) = RR_{10} \times RR_{01} \times \text{cov}(\ln RR_{10}, \ln RR_{01})$$

$$v(RR_{10}) = RR_{10}^2 \times v(\ln RR_{10})$$

$$\text{corr}(RR_{10}, RR_{01}) = \frac{\text{cov}(RR_{10}, RR_{01})}{\sqrt{v(RR_{10})v(RR_{01})}} = \frac{\text{cov}(\ln RR_{10}, \ln RR_{01})}{\sqrt{v(\ln RR_{10})v(\ln RR_{01})}}$$

Las demás varianzas, covarianzas y correlaciones se calculan de forma análoga.

A continuación se mostrarán los cálculos para estudios caso-control y de cohortes. También para el caso de introducir la interacción mediante un modelo de regresión logística, con el cual se pueden adjuntar más covariables al modelo, y extraer las OR a partir él.

### 3.3.1 Estudio caso-control

En esta sección presentamos los intervalos de confianza para las medidas de interés en interacción en un estudio de caso-control. El RERI, AP y SI están definidas en términos de RR pero, ante la imposibilidad del cálculo de RR en un estudio de caso-control, se aproximarán estos por las odds-ratios (OR).

$$RERI = OR_{11} - OR_{10} - OR_{01} + 1 = \theta_1 - \theta_2 - \theta_2 + 1$$

$$AP = \frac{RERI}{OR_{11}} = \frac{1}{OR_{11}} - \frac{OR_{10}}{OR_{11}} - \frac{OR_{01}}{OR_{11}} + 1 = \theta_1 - \theta_2 - \theta_2 + 1$$

$$\ln SI = \ln(OR_{11} - 1) - \ln(OR_{10} + OR_{01} - 2) = \theta_1 - \theta_2$$

En la tabla.1 se muestra la representación de la interacción de dos factores de riesgo mediante una tabla 4x2, idóneas para representar estas situaciones (Botto, 2001), así como unas cantidades de interés, las OR.

**Tabla 1. Tabla 4x2 caso-control**

Factores		Variable Dependiente		
G	E	SI	NO	OR
1	1	a	b	$OR_{11} = \frac{ah}{bg}$
1	0	c	d	$OR_{10} = \frac{ch}{dg}$
0	1	e	f	$OR_{01} = \frac{eh}{fg}$
0	0	g	h	1

En la sección 3.3 se presentaron los intervalos de confianza para las reparametrizaciones del RERI, AP y SI. Es necesario el cálculo de los intervalos de confianza de los parámetros, así como la correlación entre ellos.

Por tanto es preciso obtener las varianzas de los logaritmos de los parámetros. Estos cálculos los realiza Rothman (Rothman, 1976). A continuación presentamos los resultados:

$$\begin{aligned}
 v(\ln OR_{11}) &= \frac{1}{a} + \frac{1}{b} + \frac{1}{g} + \frac{1}{h} & v\left(\ln \frac{1}{OR_{11}}\right) &= v(\ln OR_{11}) \\
 v(\ln OR_{10}) &= \frac{1}{c} + \frac{1}{d} + \frac{1}{g} + \frac{1}{h} & v\left(\ln \frac{OR_{10}}{OR_{11}}\right) &= \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \\
 v(\ln OR_{01}) &= \frac{1}{e} + \frac{1}{f} + \frac{1}{g} + \frac{1}{h} & v\left(\ln \frac{OR_{01}}{OR_{11}}\right) &= \frac{1}{a} + \frac{1}{b} + \frac{1}{e} + \frac{1}{f}
 \end{aligned}$$

Los IC de los parámetros así como las correlaciones entre ellos de los distintos índices en la interacción serían:

$$- \quad i / RERI = OR_{11} - OR_{10} - OR_{01} + 1 = \theta_1 - \theta_2 - \theta_2 + 1$$

$$r_{12} = \frac{\frac{1}{g} + \frac{1}{h}}{\sqrt{v(\ln OR_{11})v(\ln OR_{10})}} \quad (l_1, u_1) = OR_{11} \exp\left(\mp 1.96\sqrt{v(\ln OR_{11})}\right)$$

$$r_{13} = \frac{\frac{1}{g} + \frac{1}{h}}{\sqrt{v(\ln OR_{11})v(\ln OR_{01})}} \quad (l_2, u_2) = OR_{10} \exp\left(\mp 1.96\sqrt{v(\ln OR_{10})}\right)$$

$$r_{23} = \frac{\frac{1}{g} + \frac{1}{h}}{\sqrt{v(\ln OR_{10})v(\ln OR_{01})}} \quad (l_3, u_3) = OR_{01} \exp\left(\mp 1.96\sqrt{v(\ln OR_{01})}\right)$$

$$- \quad ii / AP = \frac{1}{OR_{11}} - \frac{OR_{10}}{OR_{11}} - \frac{OR_{01}}{OR_{11}} + 1 = \theta_1 - \theta_2 - \theta_2 + 1$$

$$r_{12} = \frac{\frac{1}{a} + \frac{1}{b}}{\sqrt{v\left(\ln \frac{1}{OR_{11}}\right)v\left(\ln \frac{OR_{10}}{OR_{11}}\right)}} \quad (l_1, u_1) = \frac{1}{OR_{11}} \exp\left(\mp 1.96\sqrt{v\left(\ln \frac{1}{OR_{11}}\right)}\right)$$

$$r_{13} = \frac{\frac{1}{a} + \frac{1}{b}}{\sqrt{v\left(\ln \frac{1}{OR_{11}}\right)v\left(\ln \frac{OR_{01}}{OR_{11}}\right)}} \quad (l_2, u_2) = \frac{OR_{10}}{OR_{11}} \exp\left(\mp 1.96\sqrt{v\left(\ln \frac{OR_{01}}{OR_{11}}\right)}\right)$$

$$r_{23} = \frac{\frac{1}{a} + \frac{1}{b}}{\sqrt{v\left(\ln \frac{OR_{10}}{OR_{11}}\right)v\left(\ln \frac{OR_{01}}{OR_{11}}\right)}} \quad (l_3, u_3) = \frac{OR_{01}}{OR_{11}} \exp\left(\mp 1.96\sqrt{v\left(\ln \frac{OR_{01}}{OR_{11}}\right)}\right)$$

$$- \quad iii / \ln SI = \ln(OR_{11} - 1) - \ln(OR_{10} + OR_{01} - 2) = \theta_1 - \theta_2$$

$$v(OR_{11} - 1) = OR_{11}^2 v(\ln OR_{11})$$

$$v(\ln(OR_{11} - 1)) = \frac{OR_{11}^2 v(\ln OR_{11})}{(OR_{11} - 1)^2}$$

$$v(OR_{10} + OR_{01} - 2) = OR_{10}^2 v(\ln OR_{10}) + OR_{01}^2 v(\ln OR_{01}) + 2OR_{10}OR_{01} \left( \frac{1}{g} + \frac{1}{h} \right)$$

$$v(\ln(OR_{10} + OR_{01} - 2)) = \frac{v(OR_{10} + OR_{01} - 2)}{(OR_{10} + OR_{01} - 2)^2}$$

Entonces;

$$r = \frac{(OR_{10} + OR_{01}) \left( \frac{1}{g} + \frac{1}{h} \right)}{\sqrt{v(\ln OR_{11}) v(OR_{10} + OR_{01} - 2)}}$$

$$(l_1, u_1) = \ln(OR_{11} - 1) \exp \left[ \mp 1.96 \sqrt{v(\ln(\ln(OR_{11} - 1)))} \right]$$

$$(l_2, u_2) = \ln(OR_{10} + OR_{01} - 2) \exp \left[ \mp 1.96 \sqrt{v(\ln(\ln(OR_{10} + OR_{01} - 2)))} \right]$$

Estos son los valores de los límites de los IC de los parámetros y sus correlaciones. Para el cálculo de los IC para los índices de interés bastaría con substituir estos valores en las ecuaciones (1) y (2).

Como se ha comentado anteriormente, el método MOVER para el SI tiene una alta complejidad en sus cálculos y ofrecen unos resultados idénticos a los obtenidos por la aproximación log-normal. Por tanto, es preferible la utilización de este último método.

En el Anexo A se puede consultar el código de la implementación en R del método MOVER para el cálculo de los IC de RERI, AP y SI en un estudio de caso-control.

### 3.3.2 Estudio de cohortes

En esta sección presentamos los intervalos de confianza para las medidas de interés en interacción en un estudio de cohortes. Las definiciones de los índices RERI, AP y SI en términos de RR son:

$$RERI = RR_{11} - RR_{10} - RR_{01} + 1 = \theta_1 - \theta_2 - \theta_2 + 1$$

$$AP = \frac{RERI}{RR_{11}} = \frac{1}{RR_{11}} - \frac{RR_{10}}{RR_{11}} - \frac{RR_{01}}{RR_{11}} + 1 = \theta_1 - \theta_2 - \theta_2 + 1$$

$$\ln SI = \ln(RR_{11} - 1) - \ln(RR_{10} + RR_{01} - 2) = \theta_1 - \theta_2$$

En la tabla.2 se muestra la representación de la interacción de dos factores de riesgo mediante una tabla 4x2. También se muestran los RR.

**Tabla 2. Tabla 4x2 cohortes**

Factores		Variable Dependiente				
G	E	SI	NO	Total	RR	
1	1	a	b	$n_{11}$	$RR_{11} = \frac{an_{00}}{gn_{11}}$	
1	0	c	d	$n_{10}$	$RR_{10} = \frac{cn_{00}}{gn_{10}}$	
0	1	e	f	$n_{01}$	$RR_{01} = \frac{en_{00}}{gn_{01}}$	
0	0	g	h	$n_{00}$	1	

Como en el caso de un estudio de caso-control, es necesario realizar el cálculo de las varianzas de los logaritmos de los parámetros. Presentamos los resultados de estos cálculos (Rothman, 1976):

$$\begin{aligned}
 v(\ln RR_{11}) &= \frac{1}{a} - \frac{1}{n_{11}} + \frac{1}{g} - \frac{1}{n_{00}} & v\left(\ln \frac{1}{RR_{11}}\right) &= v(\ln RR_{11}) \\
 v(\ln RR_{10}) &= \frac{1}{c} - \frac{1}{n_{10}} + \frac{1}{g} - \frac{1}{n_{00}} & v\left(\ln \frac{RR_{10}}{RR_{11}}\right) &= \frac{1}{a} - \frac{1}{n_{11}} + \frac{1}{c} - \frac{1}{n_{10}} \\
 v(\ln RR_{01}) &= \frac{1}{e} - \frac{1}{n_{01}} + \frac{1}{g} - \frac{1}{n_{00}} & v\left(\ln \frac{RR_{01}}{RR_{11}}\right) &= \frac{1}{a} - \frac{1}{n_{11}} + \frac{1}{e} - \frac{1}{n_{01}}
 \end{aligned}$$

Los IC de los parámetros y las correlaciones entre ellos de los distintos índices RERI, AP y SI serían:

$$- \quad i / RERI = RR_{11} - RR_{10} - RR_{01} + 1 = \theta_1 - \theta_2 - \theta_2 + 1$$

$$r_{12} = \frac{\frac{1}{g} - \frac{1}{n_{00}}}{\sqrt{v(\ln RR_{11})v(\ln RR_{10})}} \quad (l_1, u_1) = RR_{11} \exp\left(\mp 1.96\sqrt{v(\ln RR_{11})}\right)$$

$$r_{13} = \frac{\frac{1}{g} - \frac{1}{n_{00}}}{\sqrt{v(\ln RR_{11})v(\ln RR_{01})}} \quad (l_2, u_2) = RR_{10} \exp\left(\mp 1.96\sqrt{v(\ln RR_{10})}\right)$$

$$r_{23} = \frac{\frac{1}{g} - \frac{1}{n_{00}}}{\sqrt{v(\ln RR_{10})v(\ln RR_{01})}} \quad (l_3, u_3) = RR_{01} \exp\left(\mp 1.96\sqrt{v(\ln RR_{01})}\right)$$

$$- \quad ii / AP = \frac{1}{RR_{11}} - \frac{RR_{10}}{RR_{11}} - \frac{RR_{01}}{RR_{11}} + 1 = \theta_1 - \theta_2 - \theta_2 + 1$$

$$r_{12} = \frac{\frac{1}{a} + \frac{1}{b}}{\sqrt{v\left(\ln \frac{1}{RR_{11}}\right)v\left(\ln \frac{RR_{10}}{RR_{11}}\right)}} \quad (l_1, u_1) = \frac{1}{RR_{11}} \exp\left(\mp 1.96\sqrt{v\left(\ln \frac{1}{RR_{11}}\right)}\right)$$

$$r_{13} = \frac{\frac{1}{a} + \frac{1}{b}}{\sqrt{v\left(\ln \frac{1}{RR_{11}}\right)v\left(\ln \frac{RR_{01}}{RR_{11}}\right)}} \quad (l_2, u_2) = \frac{RR_{10}}{RR_{11}} \exp\left(\mp 1.96\sqrt{v\left(\ln \frac{RR_{01}}{RR_{11}}\right)}\right)$$

$$r_{23} = \frac{\frac{1}{a} + \frac{1}{b}}{\sqrt{v\left(\ln \frac{RR_{10}}{RR_{11}}\right)v\left(\ln \frac{RR_{01}}{RR_{11}}\right)}} \quad (l_3, u_3) = \frac{RR_{01}}{RR_{11}} \exp\left(\mp 1.96\sqrt{v\left(\ln \frac{RR_{01}}{RR_{11}}\right)}\right)$$

$$- \text{iii} / \ln SI = \ln(RR_{11} - 1) - \ln(RR_{10} + RR_{01} - 2) = \theta_1 - \theta_2$$

$$v(RR_{11} - 1) = RR_{11}^2 v(\ln RR_{11})$$

$$v(\ln(RR_{11} - 1)) = \frac{RR_{11}^2 v(\ln RR_{11})}{(RR_{11} - 1)^2}$$

$$v(RR_{10} + RR_{01} - 2) = RR_{10}^2 v(\ln RR_{10}) + RR_{01}^2 v(\ln RR_{01}) + 2RR_{10}RR_{01} \left( \frac{1}{g} - \frac{1}{n_{00}} \right)$$

$$v(\ln(RR_{10} + RR_{01} - 2)) = \frac{v(RR_{10} + RR_{01} - 2)}{(RR_{10} + RR_{01} - 2)^2}$$

Entonces;

$$r = \frac{(RR_{10} + RR_{01}) \left( \frac{1}{g} - \frac{1}{n_{00}} \right)}{\sqrt{v(\ln RR_{11}) v(RR_{10} + RR_{01} - 2)}}$$

$$(l_1, u_1) = \ln(RR_{11} - 1) \exp \left[ \mp 1.96 \sqrt{v(\ln(\ln(RR_{11} - 1)))} \right]$$

$$(l_2, u_2) = \ln(RR_{10} + RR_{01} - 2) \exp \left[ \mp 1.96 \sqrt{v(\ln(\ln(RR_{10} + RR_{01} - 2)))} \right]$$

Para el cálculo de los IC de los índices bastaría con substituir los límites de los IC de los parámetros y sus correlaciones en las fórmulas (1) y (2).

De nuevo, el cálculo del IC para el SI conlleva unos desarrollos complejos, por tanto es preferible su cálculo mediante la aproximación log-normal.

En el Anexo B se puede consultar el código de la implementación en R del método MOVER para el cálculo de los IC de RERI, AP y SI en un estudio de cohortes.

### 3.3.3 MOVER en regresión

En esta sección se realizará el cálculo de los IC para el RERI, AP y SI en caso de interacción, a partir de la información extraída de un modelo de regresión. Estos índices se definen, como se ha expuesto en secciones anteriores, en términos de RR. Los RR pueden ser aproximados por las OR calculadas a partir del modelo de regresión ajustado. De esta forma, aparte de la interacción pura de dos factores, se puede introducir en el modelo el efecto de otras covariables.

Se considera por tanto el modelo de regresión

$$Z \approx X + Y + X \times Y + \sum_{i=1}^n U_i$$

donde  $Z$  es la variable de interés,  $X$  e  $Y$  los factores de interacción y  $\{U_1, \dots, U_n\}$  el resto de covariables que entran en el modelo. En la tabla 3 se muestran las características del modelo de regresión.

**Tabla 3. Mover en regresión**

	Coeficientes	Varianzas y Covarianzas		
G	$b_1$	$v_1$	$v_{12}$	$v_{13}$
E	$b_2$		$v_2$	$v_{23}$
G*E	$b_3$			$v_3$

Por tanto, los índices de interés se definen como:

$$RERI = e^{b_1+b_2+b_3} - e^{b_1} - e^{b_2} + 1 = \theta_1 - \theta_2 - \theta_2 + 1$$

$$AP = \frac{RERI}{RR_{11}} = \frac{1}{e^{b_1+b_2+b_3}} - \frac{1}{e^{b_2+b_3}} - \frac{1}{e^{b_1+b_3}} + 1 = \theta_1 - \theta_2 - \theta_2 + 1$$

$$\ln SI = \ln(e^{b_1+b_2+b_3} - 1) - \ln(e^{b_1} + e^{b_2} - 2) = \theta_1 - \theta_2$$

Como se ha expuesto anteriormente, es necesario el cálculo de las varianzas de los parámetros  $\theta_i$  así como las covarianzas y correlaciones entre ellos. Así, por ejemplo, si  $\theta_1 = e^{b_1+b_2+b_3}$  y  $\theta_2 = e^{b_1}$ , entonces, aplicando el método delta se tiene que:

$$\begin{aligned}
 r_{12} &= \frac{\text{cov}(e^{b_1+b_2+b_3}, e^{b_1})}{\sqrt{\text{var}(e^{b_1+b_2+b_3}) \times \text{var}(e^{b_1})}} \\
 &= \frac{e^{b_1+b_2+b_3} e^{b_1} \text{cov}(b_1+b_2+b_3, b_1)}{\sqrt{(e^{b_1+b_2+b_3})^2 \text{var}(b_1+b_2+b_3) \times (e^{b_1})^2 \text{var}(b_1)}} \\
 &= \frac{\text{cov}(b_1+b_2+b_3, b_1)}{\sqrt{\text{var}(b_1+b_2+b_3) \times \text{var}(b_1)}}
 \end{aligned}$$

El resto de correlaciones se calcularían de modo análogo.

Es preciso obtener, por tanto, el valor de las varianzas y covarianzas de varias combinaciones lineales de los coeficientes de la regresión, los  $b_i$ . Los valores de estas, de acuerdo a la tabla 3 y aplicando las propiedades de covarianzas y varianzas, serían:

$$\begin{aligned}
 v(b_1+b_2+b_3) &= v_1+v_2+v_3+2(v_{12}+v_{13}+v_{23}) \\
 v(b_1+b_3) &= v_1+v_3+2v_{13} \\
 v(b_2+b_3) &= v_2+v_3+2v_{23} \\
 \text{cov}(b_1+b_2+b_3, b_1) &= v_1+v_{12}+v_{13} \\
 \text{cov}(b_1+b_2+b_3, b_2) &= v_2+v_{12}+v_{23} \\
 \text{cov}(b_1+b_2+b_3, b_2+b_3) &= v_2+v_3+v_{12}+v_{13}+2v_{23} \\
 \text{cov}(b_1+b_2+b_3, b_1+b_3) &= v_1+v_3+v_{12}+v_{23}+2v_{13} \\
 \text{cov}(b_2+b_3, b_1+b_3) &= v_3+v_{12}+v_{13}+v_{23}
 \end{aligned}$$

En base a estos resultados, los IC para el RERI, AP y ln(SI) serían:

$$- \quad i / RERI = e^{b_1+b_2+b_3} - e^{b_1} - e^{b_2} = \theta_1 - \theta_2 - \theta_2 + 1$$

$$r_{12} = \frac{\text{cov}(b_1+b_2+b_3, b_1)}{\sqrt{v(b_1+b_2+b_3)v(b_1)}} \quad (l_1, u_1) = e^{b_1+b_2+b_3} \exp\left(\mp 1.96\sqrt{v(b_1+b_2+b_3)}\right)$$

$$r_{13} = \frac{\text{cov}(b_1+b_2+b_3, b_2)}{\sqrt{v(b_1+b_2+b_3)v(b_2)}} \quad (l_2, u_2) = e^{b_1} \exp\left(\mp 1.96\sqrt{v(b_1)}\right)$$

$$r_{23} = \frac{\text{cov}(b_1, b_2)}{\sqrt{v(b_1)v(b_2)}} \quad (l_3, u_3) = e^{b_2} \exp\left(\mp 1.96\sqrt{v(b_2)}\right)$$

$$- \quad ii / AP = \frac{1}{e^{b_1+b_2+b_3}} - \frac{1}{e^{b_2+b_3}} - \frac{1}{e^{b_1+b_3}} + 1 = \theta_1 - \theta_2 - \theta_3 + 1$$

$$r_{12} = \frac{\text{cov}(b_1 + b_2 + b_3, b_2 + b_3)}{\sqrt{v(b_1 + b_2 + b_3)v(b_2 + b_3)}} \quad (l_1, u_1) = \frac{1}{e^{b_1+b_2+b_3}} \exp\left(\mp 1.96\sqrt{v(b_1 + b_2 + b_3)}\right)$$

$$r_{13} = \frac{\text{cov}(b_1 + b_2 + b_3, b_1 + b_3)}{\sqrt{v(b_1 + b_2 + b_3)v(b_1 + b_3)}} \quad (l_2, u_2) = \frac{1}{e^{b_2+b_3}} \exp\left(\mp 1.96\sqrt{v(b_2 + b_3)}\right)$$

$$r_{23} = \frac{\text{cov}(b_2 + b_3, b_1 + b_3)}{\sqrt{v(b_2 + b_3)v(b_1 + b_3)}} \quad (l_3, u_3) = \frac{1}{e^{b_1+b_3}} \exp\left(\mp 1.96\sqrt{v(b_1 + b_3)}\right)$$

$$- \quad iii / \ln SI = \ln(e^{b_1+b_2+b_3} - 1) - \ln(e^{b_1} + e^{b_2} - 2) = \theta_1 - \theta_2$$

$$v(\ln(e^{b_1+b_2+b_3} - 1)) = \left[ \frac{e^{b_1+b_2+b_3}}{e^{b_1+b_2+b_3} - 1} \right]^2 v(b_1 + b_2 + b_3)$$

$$v(e^{b_1} + e^{b_2} - 2) = e^{2b_1} v_1 + e^{2b_2} v_2 + 2e^{b_1+b_2} v_{12}$$

$$v(\ln(e^{b_1} + e^{b_2} - 2)) = \frac{v(e^{b_1} + e^{b_2} - 2)}{(e^{b_1} + e^{b_2} - 2)^2}$$

Entonces;

$$r = \frac{\left[ e^{b_1} (v_1 + v_{12} + v_{13}) + e^{b_2} (v_2 + v_{12} + v_{23}) \right]}{\sqrt{v(b_1 + b_2 + b_3)v(e^{b_1} + e^{b_2} - 2)}}$$

$$(l_1, u_1) = \ln(e^{b_1+b_2+b_3} - 1) \exp\left[\mp 1.96\sqrt{v(\ln(\ln(e^{b_1+b_2+b_3} - 1)))}\right]$$

$$(l_2, u_2) = \ln(e^{b_1} + e^{b_2} - 2) \exp\left[\mp 1.96\sqrt{v(\ln(\ln(e^{b_1} + e^{b_2} - 2)))}\right]$$

Para el cálculo de los IC de los índices bastaría con substituir los límites de los IC de los parámetros y sus correlaciones en las fórmulas (1) y (2).

Igual que en los casos anteriores, el IC proporcionado por el MOVER para el SI es idéntico al obtenido por la aproximación log-normal, que tiene un cálculo mucho más sencillo.

En el Anexo C se puede consultar el código de la implementación en R del método MOVER para el cálculo de los IC de RERI, AP y SI mediante un modelo de regresión.

### 3.4. Estudio de simulación

Para la evaluación del comportamiento del método MOVER implementado, se realizó un estudio de simulación. Estudios anteriores (Zou, 2008) ponen de manifiesto el mejor comportamiento del método MOVER frente a los IC asintóticos presentados por Hosmer-Lemeshow. En esta sección se comparan los IC obtenidos a partir del método MOVER con los IC obtenidos con 3 métodos bootstrap diferentes: percentil, pivotal y bootstrap acelerado y corrector de sesgo.

#### *Metodología*

Se exponen 2 escenarios de simulación (tabla 1). El primer escenario que se considera es una situación donde existe interacción ( $OR_{11} = 20, OR_{10} = 5, OR_{01} = 4$ ). El segundo escenario considerado es una situación de no interacción ( $OR_{11} = 8, OR_{10} = 5, OR_{01} = 4$ ). Cada escenario se replica para dos tamaños muestrales, uno con un tamaño moderado ( $n=500$ ), y otro con un tamaño grande ( $n=2000$ ).

En cada repetición de la simulación se generan independientemente dos variables de bernoulli  $X_1, X_2$  de tamaño  $n$  ( $n=500, 2000$ ). A partir de ellas, se crean las variables *dummy* de interacción  $X_{11}, X_{10}, X_{01}, X_{00}$  y, mediante la función *logit*, con las OR's adecuadas en cada caso, se genera la variable dependiente  $Y$ , es decir:

$$Y \approx Ber(p(X_0, X_1))$$

con  $p = \frac{(\beta_0 + \log(OR_{01})X_{01} + \log(OR_{10})X_{10} + \log(OR_{11})X_{11})}{1 + (\beta_0 + \log(OR_{01})X_{01} + \log(OR_{10})X_{10} + \log(OR_{11})X_{11})}$ , donde  $\beta_0 = \log(0.5)$ , para asegurarnos que la prevalencia poblacional es igual a 0.5.

Para comparar los IC obtenidos mediante el método MOVER se programaron 3 métodos bootstrap. En el bootstrap percentil (*BI*) se realizan las remuestras de los datos originales y a partir de la distribución bootstrap se estiman los percentiles de la distribución muestral para las cantidades de interés (RERI, AP y SI). Si denotamos por  $X_{0.025}^*$  y  $X_{0.975}^*$  los percentiles 2.5° y 97.5° estimados de la distribución muestral, el intervalo de confianza al 95% por el método bootstrap percentil será:

$$(X_{0.025}^*, X_{0.975}^*)$$

En el bootstrap pivotal (*B2*), se resta la cantidad de interés original a la calculada en cada remuestra. Sobre esta distribución muestral se calculan los percentiles 2.5° y 97.5°, los denotamos por  $Y_{0.025}^*$  y  $Y_{0.975}^*$ . El IC al 95% para  $Y$ , donde  $Y$  es el índice de interés, calculado por el bootstrap pivotal será:

$$(\hat{Y} + Y_{0.025}^*, \hat{Y} + Y_{0.975}^*)$$

El tercer bootstrap (*BCa*) es el bootstrap acelerado y corrector de sesgo (Efron y Tibshirani, 1993). Este método ofrece unos intervalos similares al bootstrap percentil pero en vez de ser  $X_{0.025}^* < X < X_{0.975}^*$  los límites del IC serán  $X_{\alpha_L}^* < X < X_{\alpha_U}^*$  donde  $L$  y  $U$  son elegidos para hacer las medianas de los intervalos insesgadas y ajustadas por asimetría. Los valores  $\alpha_L$  y  $\alpha_U$  se eligen para que tener la misma probabilidad acumulada que  $z_L$  y  $z_U$ , definidos como:

$$z_L = \frac{z_0 - z_{1-\alpha/2}}{1 - a(z_0 - z_{1-\alpha/2})} + z_0 ; \quad z_U = \frac{z_0 + z_{1-\alpha/2}}{1 - a(z_0 + z_{1-\alpha/2})} + z_0$$

El valor  $z_0$  mide que las medianas sean insesgadas y se define por:  $P(Z \leq z_0) = p_0$  donde  $p_0$  es la proporción de veces que la cantidad de interés en las remuestras es menor o igual a la original ( $\hat{\theta}_b \leq \hat{\theta}$ ). El valor  $a$  mide asimetría en los datos y viene dado por:

$$a = \frac{\sum (\hat{\theta}_{( )} - \hat{\theta}_{-i})^3}{6 \left[ \sum (\hat{\theta}_{( )} - \hat{\theta}_{-i})^2 \right]^{3/2}}$$

donde  $\hat{\theta}_{-i}$  es la estimación de  $\theta$  sin la observación  $i$ -ésima de los datos y  $\hat{\theta}_{( )}$  es la media de los valores  $\hat{\theta}_{-i}$ . En el caso de que  $a=0$  y  $z_0=0$ , el bootstrap BCa es el mismo que el bootstrap percentil.

**Tabla 1: Escenarios elegidos para la simulación**

Escenario	OR <sub>11</sub>	OR <sub>10</sub>	OR <sub>01</sub>	RERI	AP	SI
Interacción	20	5	4	12	0.6	2.7
No interacción	8	5	4	0	0	1

*Resultados de la simulación*

El primer escenario de simulación se plantea una situación donde existe interacción, con unas OR's de  $OR_{11}=20$ ,  $OR_{10}=5$ ,  $OR_{01}=4$ . En este escenario se tendrían unos valores de los índices de interés de  $RERI=12$ ,  $AP=0.6$  y  $SI=2.7$ . En la tabla 2 se muestran los resultados del estudio. El método MOVER tiene una excelente cobertura, cercana al valor nominal (95%), y en algunos casos incluso mejor que las de los métodos bootstrap.

También se ha computado la longitud media de los intervalos obtenidos en la simulación. Es interesante comprobar cómo, con un tamaño muestral pequeño ( $n=500$ ), el MOVER consigue una cobertura similar a los métodos bootstrap con una amplitud media menor. En el caso del RERI, la amplitud media es de 35, mientras que en los bootstrap percentil y pivotal supera 45, y el BCa es de 40. En el caso del SI también es sensiblemente menor el MOVER y en el caso de la AP los resultados son muy similares. Con un tamaño muestral grande ( $n=2000$ ), la longitud media de los IC de los tres índice son prácticamente idénticas en los 4 métodos.

**Tabla 2: Estudio de simulación 1**

<i>n</i>			MOVER	B1	B2	BCa
500	Cobertura	RERI	95.5	94.3	94.9	96.2
		AP	94.4	94.8	95	95.8
		SI	95.5	95	95.1	95.8
	Amplitud	RERI	35.6	46.1	45.4	40.07
		AP	0.624	0.602	0.602	0.618
		SI	5.326	6.777	6.730	6.040
2000	Cobertura	RERI	94.9	94.3	94.6	96
		AP	94.8	94.6	95.1	95.7
		SI	95	94.8	95.4	95.7
	Amplitud	RERI	14.04	14.51	14.56	14.11
		AP	0.28	0.27	0.27	0.27
		SI	2.11	2.17	2.18	2.16

Cobertura y amplitud media de los IC para el RERI, AP, SI en el estudio de simulación 1. Tamaños muestrales de 500 y 2000, y unas OR de  $OR_{11}=20$ ,  $OR_{10}=5$ ,  $OR_{01}=4$ . Sí hay interacción. B1: Bootstrap cuantil. B2: bootstrap pivotal. BCa: bootstrap acelerado y con corrección de sesgo. 500 remuestras en cada bootstrap.

El segundo escenario de simulación está planteado con una situación de no interacción. Las OR's consideradas son  $OR_{11}=8$ ,  $OR_{10}=5$ ,  $OR_{01}=4$ . Con esta situación, los valores de los índices serían  $RERI=0$ ,  $AP=0$  y  $SI=1$ . En la tabla 3 se muestran los resultados del estudio. El método MOVER nuevamente tiene una cobertura excelente y competitiva con los métodos bootstrap.

La amplitud media de los IC es muy similar en todos los casos, siendo mucho menores que las del primer escenario de simulación, donde sí había interacción.

**Tabla 3: Estudio de simulación 2**

<i>n</i>			MOVER	B1	B2	BCa
500	Cobertura	RERI	94.6	94.2	94.2	95.7
		AP	94	94.2	94.1	95
		SI	95	94.2	94.1	95.2
	Amplitud	RERI	11.40	11.69	11.73	11.63
		AP	1.25	1.20	1.20	1.19
		SI	1.42	1.51	1.51	1.50
2000	Cobertura	RERI	95.1	95.5	95.8	96
		AP	95.3	95.5	95.8	95.8
		SI	95.2	95.5	95.8	95.8
	Amplitud	RERI	4.58	4.56	4.57	4.57
		AP	0.55	0.54	0.54	0.53
		SI	0.63	0.63	0.63	0.65

Cobertura y amplitud media de los IC para el RERI, AP, SI en el estudio de simulación 2. Tamaños muestrales de 500 y 2000, y unas OR de  $OR_{11}=8$ ,  $OR_{10}=5$ ,  $OR_{01}=4$ . No interacción. B1: Bootstrap cuantil. B2: bootstrap pivotal. BCa: bootstrap acelerado y con corrección de sesgo. 500 remuestras en cada bootstrap.

A la vista de los resultados obtenidos en la simulación, el método MOVER tiene unos resultados excelentes. En tamaños muestrales grandes obtiene unos resultados incluso mejores que los métodos bootstrap, manteniéndose competitivo frente a estos con tamaños muestrales pequeños.

### 3.5. Aplicación a datos reales

En esta sección se aplicará el método MOVER para calcular los IC del RERI, AP y SI en unos datos reales. Se utilizarán los datos de un estudio que trata la interacción gen por medio ambiente. En este estudio fue necesario el cálculo del IC para el RERI y se realizó mediante la aproximación asintótica de Hosmer-Lemeshow. Se volverán a calcular por el método MOVER y se compararán los resultados.

#### *Introducción*

Diversos estudios han puesto de manifiesto que los niveles de inmunoglobulina E (IgE) aumentan en personas que consumen alcohol. Sin embargo este aumento no es uniforme, existiendo aumentos importantes en bebedores excesivos mientras que otros presentan niveles totalmente normales. Por otra parte, también existen evidencias de que la interacción de factores ambientales y genes relacionados con la respuesta inmune (como el *CD14*) puede jugar un papel importante en los niveles de IgE.

Así, en 2006, Campos et al (Campos et al., 2006) publicaron un estudio cuyo objetivo era investigar una posible interacción entre el gen *CD14* (polimorfismo *-159/T*) y el consumo de alcohol como causante del aumento de los niveles de IgE.

#### *Análisis descriptivo*

En el estudio participaron 415 individuos, de los cuales 140 eran abstemios, 112 consumidores moderados de alcohol (1-280 g/semana) y 163 consumían alcohol en cantidades altas (>280 g/semana). Se consideró una clasificación dicotómica, un grupo constituido por los abstemios y consumidores moderados de alcohol, y otro grupo formado por los consumidores excesivos de alcohol.

La IgE se categorizó en dos grupos: i) IgE alta, con valores superiores o iguales a 170 IU/mL; ii) IgE normal si presentaban valores menores a 170 IU/mL. El 31% (n=127) de los individuos presentaban valores de IgE altos.

Existen 3 posibles genotipos del polimorfismo *-159/T* del gen *CD14*: TT, CC y CT. Estos se clasificaron en dos grupos, uno donde está presente el alelo C (CC y CT) y otro grupo donde no está presente este alelo (TT). El 76% (n=316) de los individuos están en el grupo de genotipo con presencia de alelo C, y los restantes en el grupo de genotipo de alelos TT.

En el estudio publicado, los autores realizaron dos tipos de ajustes. Uno en el que presentan los resultados de la interacción cruda de estas variables, y otro ajustando el modelo con otras potenciales variables de confusión tales como edad, sexo, hábito tabáquico, antecedentes de cirrosis y ser atópico.

En la tabla 1 se describen un conjunto de variables que se encuentran asociadas a los niveles de IgE. Todas estas variables han sido descritas en distintos estudios con

posible influencia sobre los valores de IgE, por lo que parece recomendable introducirlas en el modelo que investigue la interacción entre consumo de alcohol y el gen CD/14.

En la tabla 2 se describe la relación entre las variables que vamos a introducir en el modelo y la presencia o ausencia del alelo C en el genotipo del polimorfismo -159/T del gen *CD14* (grupos CC/CT y TT) (tabla 2). Se observa como las distribuciones de las variables en los dos grupos son similares.

**Tabla 1. Características generales de acuerdo con los niveles de IgE**

Variable	>170 IU/mL	0-170 IU/mL	p-valor
Edad (años)	49 (18-82)	50 (19-92)	0.69
Sexo (masculino)	87 (68.5)	127 (44.1)	<0.01
Fumador (activo)	59 (46.4)	68 (23.6)	<0.01
Alcohol (>280 g/sem)	88 (69.3)	75 (26.1)	<0.01
Cirrosis (ocurrencia)	25 (19.5)	10 ( 3.5)	<0.01
Test alérgico (positivo)	49 (38.6)	46 (15.9)	<0.01

Los datos están presentados en medianas y rangos (continuas) o número absoluto y proporción (categóricas). p-valores obtenidos por test de Mann-Withney (continuas) o por el test chi-cuadrado (categóricas).

**Tabla 2. Características generales de acuerdo con la clasificación del polimorfismo CD14/-159**

Variable	CC/CT	TT	p-valor
Edad (años)	49 (18-92)	50 (19-86)	0.91
Sexo (masculino)	161 (50.9)	53 (53.5)	0.73
Fumador (activo)	100 (31.6)	27 (27.3)	0.48
Alcohol (>280 g/sem)	125 (39.5)	38 (38.4)	0.92
Cirrosis (ocurrencia)	23 (7.3)	12 (12.1)	0.19
Test alérgico (positivo)	72 (22.8)	23 (23.2)	0.96

Los datos están presentados en medianas y rangos (continuas) o número absoluto y proporción (categóricas). p-valores obtenidos por test de Mann-Withney (continuas) o por el test chi-cuadrado (categóricas).

*Interacción*

Se presentan los resultados de la interacción entre consumo de alcohol y el gen *CD14*. En el artículo original se calculó el RERI en la interacción cruda y en la ajustada por otras variables de interés, ofreciendo su IC al 95% calculado mediante la aproximación asintótica (tabla 3).

Existen evidencias de interacción positiva, con un valor de RERI de 6.87 en interacción cruda y de 7.46 en la interacción ajustada por edad, sexo, ser fumador, haber padecido cirrosis y tener resultado positivo en el test alérgico (ser atópico). En ambos casos el valor es significativo.

**Tabla 3: Tabla 4x2 con las características de la interacción Alcohol-Genotipo CD14/-159 para niveles de IgE**

Alcohol	Genotipo		Casos	Controles	OR cruda	OR ajustada
	CD14/-159					
No	TT		8	53	1 (referencia)	1 (referencia)
No	CC/CT		31	160	1.28	1.27
Si	TT		12	26	3.05	1.54
Si	CC/CT		76	49	10.2	9.27
RERI crudo (IC)			RERI ajustado (IC)			
10.2-3.05-1.28+1=6.87 (0.80 – 13.07)			9.27-1.54-1.27+1=7.46 (0.57 – 14.34)			

Casos: Individuos con niveles de IgE > 170 IU/mL. Controles: Individuos con niveles de IgE < 170 IU/mL.  
 Modelo ajustado por edad, sexo, ser fumador, padecer cirrosis y test alérgico positivo.  
 OR: odds ratios, IC: intervalos de confianza al 95%.

*Aplicación método MOVER*

En este apartado nuestro objetivo es calcular los IC de los índices de interés en interacción mediante el método MOVER, y comparar los resultados obtenidos con los ofrecidos por el método asintótico.

Se calculan, por tanto, el RERI, AP y SI y sus respectivos IC por aproximación asintótica, por el método MOVER y por los métodos bootstrap acelerado y corrector de sesgo (BCa), y bootstrap cuantil (B1) (tabla 4). Se realizan los cálculos para el modelo de interacción ajustado por las variables edad, sexo, ser fumador, haber padecido cirrosis y tener resultado positivo en el test alérgico.

Por construcción, los IC asintóticos de Hosmer-Lemeshow están forzados a ser simétricos, lo que implica una limitación. Por ejemplo, para el RERI se observa que los límites del IC asintótico toman valores más bajos, 0.57 y 14.34, que el método MOVER, 3.5 y 20.7, los cuales son más similares a los obtenidos por los métodos bootstrap. En cambio, en el caso de la AP el límite inferior es sensiblemente más alto (0.61 por 0.51), siendo, nuevamente, el método MOVER más próximo a los métodos bootstrap. En el caso del SI, los dos métodos (asintótico y MOVER) calculan el mismo IC por construcción (Zou, 2008).

**Tabla 4: Intervalos de Confianza de los índices de interés por los distintos métodos.**

Índice	Valor	H-L	MOVER	B1	BCa
RERI	7.46	0.574 – 14.34	3.572 – 20.73	3.58 – 20.74	3.26 – 18.78
AP	0.804	0.61 – 0.99	0.51 – 0.98	0.56 – 0.974	0.49 – 0.952
SI	10.12	0.796 – 129.6	0.796 – 129.6	2.34 – 69.05	-5.28 – 837.8

Para los métodos bootstrap B1 (cuantil) y BCa (acelerado y corrector de sesgo) se realizan 500 remuestras.  
H-L: método asintótico de Hosmer-Lemeshow.

Puede observarse que el método MOVER ofrece unos resultados más próximos a los métodos bootstrap que el método asintótico. Zou realizó un estudio de simulación (Zou, 2008) donde comparaba ambos métodos y pone de manifiesto un mejor comportamiento del MOVER frente a los asintóticos.

Este estudio puede tomarse como ejemplo de un posible error del método asintótico. En el caso del RERI, el IC está muy cerca de contener al cero, lo que nos llevaría a considerar el índice no significativo. En cambio, por el método MOVER es claramente significativo. El peor comportamiento del método asintótico podría llevarnos a consideraciones erróneas. En este sentido, el método MOVER es más fiable.

## 3.6. Resumen

A lo largo del capítulo se ha tratado el concepto de interacción. Más concretamente la *interacción biológica* que, como se ha expuesto, es un concepto distinto a *interacción matemática*. También hemos visto que la forma de medir la interacción puede realizarse con dos filosofías diferentes, la aditiva y la multiplicativa. Encuadrada dentro de la epidemiología, es más intuitiva y sencilla de medir la interacción aditiva y, por tanto, es la que se trató a lo largo del capítulo.

La interacción aditiva queda caracterizada por tres índices (RERI, AP y SI) y de ahí que tenga un especial interés el cálculo de sus intervalos de confianza. En este capítulo se presentó la metodología descrita por Zou (método MOVER) (Zou, 2008) para el cálculo los intervalos de confianza para estas tres cantidades como alternativa a los de Rothman y Hosmer-Lemeshow.

Una de las tareas en las prácticas llevadas a cabo en la Unidad de Epidemiología consistió en la implementación del método MOVER en R.

En un estudio de simulación (Zou, 2008) se ponía de manifiesto un mejor comportamiento de los IC calculados por el MOVER frente a los asintóticos de Hosmer-Lemeshow. En la sección 3.4 se realizó otro estudio para evaluar el comportamiento de las funciones programadas. Los resultados fueron satisfactorios, estando en todos los escenarios de la simulación, la cobertura de IC muy cercana al valor nominal. Además, en ese estudio, se compararon los resultados con tres métodos bootstrap, siendo los resultados del MOVER muy parejos a ellos.

Finalmente, en la sección 3.5, se utilizó el método MOVER implementado para el cálculo de los IC en un caso real. Se obtuvieron con el MOVER los IC que se habían calculado en un artículo previo mediante el método asintótico. Como era de esperar, los dos métodos ofrecieron unos resultados diferentes. Para poder discriminar cual de los dos métodos ofrecían unos IC más fiables, se calcularon también por dos métodos bootstrap, los cuales tenían unos resultados próximos al método MOVER. Así, en principio, esto nos hace suponer que el método MOVER ofrece unos resultados más exactos que el método asintótico en este caso concreto.



## Valoración de las prácticas

El máster en técnicas estadísticas posee una gran carga teórica, por lo que creo que un alumno que lo cursa acaba con unos grandes conocimientos de distintas metodologías estadísticas.

Una posible dificultad que puede surgir a un estudiante del máster es la de poner en práctica todos los conocimientos adquiridos. Una cosa es tener muchos fundamentos teóricos asumidos y otra bien diferente es saber qué hacer con ellos.

Cerca de finalizar mi formación académica en el máster, esta era la sensación que tenía. Dudaba de la capacidad que tendría de llevar a la práctica los conocimientos adquiridos a lo largo del máster. Así que la realización de unas prácticas me parecía que podrían ser el complemento ideal a la formación académica.

Así, finalmente, surgió la posibilidad de realizar unas prácticas en la Unidad de Epidemiología Clínica del CHUS. A lo largo de la estancia en la Unidad continué con la formación académica, aprendiendo nuevas metodologías estadísticas, y, tanto o más importante, poniéndolas en práctica.

Durante el periodo de prácticas aprendí que las posibles tareas a realizar van mucho más allá de realizar un análisis estadístico. La recogida de los datos, la depuración de los mismos y análisis previos son tanto o más importantes y necesarios que el propio análisis estadístico.

En las prácticas obtuve una visión más general a la hora de realizar un estudio. Saber planificar las tareas que se realizarán, tener claro el objetivo final de dicho estudio y obrar en consecuencia. Es importante saber resolver los problemas, realizar el análisis estadístico, como saber qué problemas resolver, la planificación del estudio. Estos son pasos a seguir en los que no tenía experiencia o no consideraba competencia de un estadístico y que con la estancia en la Unidad comprendí la importancia que tienen y cogí experiencia en desarrollarlos.

Por lo tanto, la estancia de prácticas en la Unidad de Epidemiología, ha sido el complemento ideal a mi formación académica, tanto por continuar con el aprendizaje metodológico estadístico como por la experiencia ganada en los otros factores de importancia antes descritos que un analista de datos debe poseer.



## Bibliografía

Ahlbom A, Alfredsson L. Interaction: A word with two meanings creates confusion. *European Journal of Epidemiology* 2005;20:563–564.

Assmann SF, Hosmer DW, Lemeshow S, et al. Confidence intervals for measures of interaction. *Epidemiology* 1996;7:286–90.

Botto LD, Khoury MJ. Commentary: facing the challenge of gene-environment interaction: the two-by-four table and beyond. *Am J Epidemiol* 2001;153:1016–20.

Campos J, Gude F, Quintero C, Vidal C, Gonzalez-Quintela A. Gene by Environment Interaction: The  $\_159C/T$  Polymorphism in the Promoter Region of the CD14 Gene Modifies the Effect of Alcohol Consumption on Serum IgE Levels. *Alcohol Clin Exp Res* 2006;30:7-14.

Efron B, Tibshirani RJ. *An introduction to the bootstrap*. New York, NY: Chapman & Hall/CRC, 1993.

Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology* 1992;3:452–6.

Kalilani L, Atashili J. Measuring additive interaction using odds ratios. *Epidemiologic Perspectives & Innovations* 2006,3:5.

Knol MJ, van der Tweel I, Grobbee DE, et al. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. *Int J Epidemiol* 2007;36:1111–18.

Rothman KJ. The estimation of synergy or antagonism. *Am J Epidemiol* 1976;103:506-11.

Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *Am J Epidemiol* 1980;112:467–70.

Rothman KJ. *Modern epidemiology*. Boston, MA: Little, Brown & Co, 1986.

Szklo M, Nieto FJ. Epidemiology: beyond the basics. 2nd ed. Sudbury, MA: Jones and Bartlett Publishers, Inc, 2006.

Tiret L, Gene-environment interaction: a central concept in multifactorial diseases. Proceedings of the Nutrition Society 2002;61:457-63.

Zou GY, On the estimation of additive interaction by use the four-by-two table and beyond. Am J Epidemiol 2008;168:212-24.



## Anexo A

### Función MOVER en caso-control

- **Función *mover.control*.**

Argumentos de entrada:      “var”: variable dependiente.  
                                   “inter”: data frame con las variables de interacción.

Argumentos de salida:       Devuelve un data frame con los valores de los tres índices  
                                   y sus intervalos de confianza

```
mover.control<-function(var,inter){

  t<-tabla1(var,inter)
  res1<-parametros.reri1(t[1],t[2],t[3],t[4],t[5],t[6],t[7],t[8])
  res2<-parametros.ap1(t[1],t[2],t[3],t[4],t[5],t[6],t[7],t[8])
  res3<-parametros.si1(t[1],t[2],t[3],t[4],t[5],t[6],t[7],t[8])
  lim.reri<-limites1(res1[2],res1[3],res1[4],res1[8],res1[5],res1[6],res1[7])
  lim.ap<-limites1(res2[2],res2[3],res2[4],res2[8],res2[5],res2[6],res2[7])
  lim.si<-limites11(res3[1],res3[3])

  valores<-c(res1[1],res2[1],res3[2],res3[1])
  L<-c(lim.reri$IC.L,lim.ap$IC.L,lim.si$IC.L,round(exp(lim.si$IC.L),digits=3))
  U<-c(lim.reri$IC.U,lim.ap$IC.U,lim.si$IC.U,round(exp(lim.si$IC.U),digits=3))

  out<-data.frame("value"=valores,"IC.L"=L,"IC.U"=U,
                 row.names=c("RERI", "AP", "lnSI", "SI"))

}
```

- **Función *tabla1*.**

Argumentos de entrada:      “var”: variable dependiente.  
                                   “inter”: data frame con las variables de interacción.

Argumentos de salida:       Devuelve un vector con los valores de la tabla 4x2 de las variables que interactúan y la variable dependiente.

```
tabla1<-function(var,inter){
  A<-inter[,1]
  B<-inter[,2]
  AB<-numeric()
  AB[A==1&B==1]<-3
  AB[A==1&B==0]<-2
  AB[A==0&B==1]<-1
  AB[A==0&B==0]<-0

  t<-table(AB,var)
  res<-c(t[8],t[4],t[7],t[3],t[6],t[2],t[5],t[1])
}
```

- **Función *parámetros.reri1*.**

Argumentos de entrada:      Los 8 valores de la tabla 4x2 de interacción (el resultado de *tabla1*).

Argumentos de salida:       Devuelve un vector con el valor de RERI y los parámetros necesarios para el cálculo del IC mediante MOVER.

```
parametros.reri1<-function(a,b,c,d,e,f,g,h){
  OR11<-(a*h)/(b*g)
  OR10<-(c*h)/(d*g)
  OR01<-(e*h)/(f*g)

  var11<-1/a+1/b+1/h+1/g
  var10<-1/c+1/d+1/h+1/g
  var01<-1/e+1/f+1/h+1/g
  cov.reri<-1/g+1/h

  RERI<-round(OR11-OR10-OR01+1,digits=3)

  res<-c(RERI,OR11,OR10,OR01,var11,var10,var01,cov.reri)
}
```

- **Función *parámetros.ap1*.**

Argumentos de entrada: Los 8 valores de la tabla 4x2 de interacción (el resultado de *tabla1*).

Argumentos de salida: Devuelve un vector con el valor del AP y los parámetros necesarios para el cálculo del IC mediante MOVER.

```
parametros.ap1<-function(a,b,c,d,e,f,g,h){
  OR11<-(a*h)/(b*g)
  OR10<-(c*h)/(d*g)
  OR01<-(e*h)/(f*g)

  var1.11<-1/a+1/b+1/h+1/g
  var10.11<-1/a+1/b+1/c+1/d
  var01.11<-1/a+1/b+1/e+1/f
  cov.ap<-1/a+1/b

  RERI<-round(OR11-OR10-OR01+1,digits=3)
  AP<-round(RERI/OR11,digits=3)

  res<-c(AP,1/OR11,OR10/OR11,OR01/OR11,var1.11,var10.11,var01.11,cov.ap)
}
```

- **Función *parámetros.si1*.**

Argumentos de entrada: Los 8 valores de la tabla 4x2 de interacción (el resultado de *tabla1*).

Argumentos de salida: Devuelve un vector con el valor del SI, logaritmo del SI y la varianza del logaritmo del SI necesarios para el cálculo del IC.

```
parametros.si1<-function(a,b,c,d,e,f,g,h){
  OR11<-(a*h)/(b*g)
  OR10<-(c*h)/(d*g)
  OR01<-(e*h)/(f*g)

  var11<-1/a+1/b+1/h+1/g
  var10<-1/c+1/d+1/h+1/g
  var01<-1/e+1/f+1/h+1/g
  tita1<-OR11-1
  tita2<-OR10+OR01-2
```

```

Sest<-tita1/tita2

var1<-((OR11^2)*var11)/(OR11-1)^2
var2<-((OR10^2)*var10+(OR01^2)*var01+
        2*OR10*OR01*(1/g+1/h))/(OR10+OR01-2)^2
cov<-((OR10+OR01)*(1/g+1/h))

varlnS<-var1+var2-cov
lnSI<-round(log(Sest),digits=3)

res<-c(Sest,lnSI,varlnS)
}

```

- **Función *limites1*.**

Argumentos de entrada:      “tita1”: Primer parámetro de la reparametrización.  
                                   “tita2”: Segundo parámetro de la reparametrización.  
                                   “tita3”: Tercer parámetro de la reparametrización.  
                                   “cov”: Covarianza entre los parámetros.  
                                   “var1”: Logaritmo de la varianza de tita1.  
                                   “var2”: Logaritmo de la varianza de tita2.  
                                   “var3”: Logaritmo de la varianza de tita3.

Argumentos de salida:        Devuelve una lista con los límites del IC MOVER para los datos del argumento de entrada.

```

limites1<-function(tita1,tita2,tita3,cov,var1,var2,var3){

  r12<-cov/sqrt(var1*var2)
  r13<-cov/sqrt(var1*var3)
  r23<-cov/sqrt(var2*var3)

  l1<-tita1*exp(-1.96*sqrt(var1))
  u1<-tita1*exp(1.96*sqrt(var1))
  l2<-tita2*exp(-1.96*sqrt(var2))
  u2<-tita2*exp(1.96*sqrt(var2))
  l3<-tita3*exp(-1.96*sqrt(var3))
  u3<-tita3*exp(1.96*sqrt(var3))

  L<-round(1+tita1-tita2-tita3-sqrt((tita1-l1)^2+(u2-tita2)^2+(u3-tita3)^2-
    2*r12*(tita1-l1)*(u2-tita2)-2*r13*(tita1-l1)*(u3-tita3)+
    2*r23*(u2-tita2)*(u3-tita3)),digits=3)

  U<-round(1+tita1-tita2-tita3+sqrt((u1-tita1)^2+(tita2-l2)^2+(tita3-l3)^2-

```

```
2*r12*(u1-tita1)*(tita2-l2)-2*r13*(u1-tita1)*(tita3-l3)+
2*r23*(tita2-l2)*(tita3-l3)),digits=3)
```

```
list(IC.L=L,IC.U=U)
}
```

- **Función *limites11*.**

Argumentos de entrada:    “Sest”: Valor estimado del SI  
                               “varlnS”: Valor de la varianza del logaritmo del SI.

Argumentos de salida:     Devuelve una lista con los límites del IC para los datos del argumento de entrada.

```
limites11<-function(Sest,varlnS){
  L<-round(log(Sest)-1.96*sqrt(varlnS),digits=3)
  U<-round(log(Sest)+1.96*sqrt(varlnS),digits=3)
  list(IC.L=L,IC.U=U)
}
```

## Anexo B

### Función MOVER en cohorte

- **Función *mover.cohort*.**

Argumentos de entrada:     “var”: variable dependiente.  
                                   “inter”: data frame con las variables de interacción.

Argumentos de salida:       Devuelve un data frame con los valores de los tres índices  
                                   y sus intervalos de confianza

```
mover.cohort<-function(var,inter){

  t<-tabla2(var,inter)
  res1<-parametros.eri2(t[1],t[2],t[3],t[4],t[5],t[6],t[7],t[8],t[9],t[10],t[11],t[12])
  res2<-parametros.ap2(t[1],t[2],t[3],t[4],t[5],t[6],t[7],t[8],t[9],t[10],t[11],t[12])
  res3<-parametros.si2(t[1],t[2],t[3],t[4],t[5],t[6],t[7],t[8],t[9],t[10],t[11],t[12])
  lim.eri<-limites2(res1[2],res1[3],res1[4],res1[8],res1[5],res1[6],res1[7])
  lim.ap<-limites2(res2[2],res2[3],res2[4],res2[8],res2[5],res2[6],res2[7])
  lim.si<-limites22(res3[1],res3[3])

  valores<-round(c(res1[1],res2[1],res3[2],res3[1]),digits=3)
  L<-c(lim.eri$IC.L,lim.ap$IC.L,lim.si$IC.L,round(exp(lim.si$IC.L),digits=3))
  U<-c(lim.eri$IC.U,lim.ap$IC.U,lim.si$IC.U,round(exp(lim.si$IC.U),digits=3))

  data.frame("value"=valores,"IC.L"=L,"IC.U"=U,
            row.names=c("RERI","AP","lnSI","SI"))
}
```

- **Función *tabla2*.**

Argumentos de entrada:      “var”: variable dependiente.  
                                   “inter”: data frame con las variables de interacción.

Argumentos de salida:       Devuelve un vector con los valores de la tabla 4x2 de las variables que interactúan y la variable dependiente y los individuos totales.

```
tabla2<-function(var,inter){

  A<-inter[,1]
  B<-inter[,2]
  AB<-numeric()
  AB[A==1&B==1]<-3
  AB[A==1&B==0]<-2
  AB[A==0&B==1]<-1
  AB[A==0&B==0]<-0

  t<-table(AB,var)
  n11<-t[8]+t[4]
  n10<-t[7]+t[3]
  n01<-t[6]+t[2]
  n00<-t[5]+t[1]

  c(t[8],t[4],t[7],t[3],t[6],t[2],t[5],t[1],n11,n10,n01,n00,t)

}
```

- **Función *parámetros.reri2*.**

Argumentos de entrada:      Los 8 valores de la tabla 4x2 de interacción y los 4 valores de los individuos totales (el resultado de *tabla2*).

Argumentos de salida:       Devuelve un vector con el valor de RERI y los parámetros necesarios para el cálculo del IC mediante MOVER.

```
parametros.reri2<-function(a,b,c,d,e,f,g,h,n11,n10,n01,n00){

  RR11<-(a*n00)/(n11*g)
  RR10<-(c*n00)/(n10*g)
  RR01<-(e*n00)/(n01*g)
```

```

var11<-1/a-1/n11+1/g-1/n00
var10<-1/c-1/n10+1/g-1/n00
var01<-1/e-1/n01+1/g-1/n00
cov.reri<-1/g-1/n00

RERI<-RR11-RR10-RR01+1

c(RERI,RR11,RR10,RR01,var11,var10,var01,cov.reri)
}

```

- **Función *parámetros.ap2*.**

Argumentos de entrada: Los 8 valores de la tabla 4x2 de interacción y los 4 valores de los individuos totales (el resultado de *tabla2*).

Argumentos de salida: Devuelve un vector con el valor del AP y los parámetros necesarios para el cálculo del IC mediante MOVER.

```

parametros.ap2<-function(a,b,c,d,e,f,g,h,n11,n10,n01,n00){

  RR11<-(a*n00)/(n11*g)
  RR10<-(c*n00)/(n10*g)
  RR01<-(e*n00)/(n01*g)

  var1.11<-1/a-1/n11+1/g-1/n00
  var10.11<-1/a-1/n11+1/c-1/n10
  var01.11<-1/a-1/n11+1/e-1/n01
  cov.ap<-1/a-1/n11

  RERI<-RR11-RR10-RR01+1
  AP<-RERI/RR11

  c(AP,1/RR11,RR10/RR11,RR01/RR11,var1.11,var10.11,var01.11,cov.ap)

}

```

- **Función *parámetros.si2*.**

Argumentos de entrada: Los 8 valores de la tabla 4x2 de interacción y los 4 valores de los individuos totales (el resultado de *tabla2*).

Argumentos de salida: Devuelve un vector con el valor del SI, logaritmo del SI y la varianza del logaritmo del SI necesarios para el cálculo del IC.

```
parametros.si2<-function(a,b,c,d,e,f,g,h,n11,n10,n01,n00){

  RR11<-(a*n00)/(n11*g)
  RR10<-(c*n00)/(n10*g)
  RR01<-(e*n00)/(n01*g)

  tita1<-RR11-1
  tita2<-RR10+RR01-2
  Sest<-tita1/tita2

  var1<-1/a-1/n11+1/g-1/n00
  var2<-RR10^2*(1/c-1/n10+1/g-1/n00)+
    RR01^2*(1/e-1/n01+1/g-1/n00)+2*RR10*RR01*(1/g+1/n00)
  vartita1<-RR11^2*var1/(RR11-1)^2
  vartita2<-1/(RR10+RR01-2)^2*var2
  cov<-(RR10+RR01)*(1/g-1/n00)

  varlnS<-vartita1+vartita2-cov
  lnSI<-round(log(Sest),digits=3)

  res<-c(Sest,lnSI,varlnS)

}
```

- **Función *limites2*.**

Argumentos de entrada:      “tita1”: Primer parámetro de la reparametrización.  
                                   “tita2”: Segundo parámetro de la reparametrización.  
                                   “tita3”: Tercer parámetro de la reparametrización.  
                                   “cov”: Covarianza entre los parámetros.  
                                   “var1”: Logaritmo de la varianza de tita1.  
                                   “var2”: Logaritmo de la varianza de tita2.  
                                   “var3”: Logaritmo de la varianza de tita3.

Argumentos de salida:       Devuelve una lista con los límites del IC MOVER para los  
                                   datos del argumento de entrada.

```
limites2<-function(tita1,tita2,tita3,cov,var1,var2,var3){

  r12<-cov/sqrt(var1*var2)
  r13<-cov/sqrt(var1*var3)
  r23<-cov/sqrt(var2*var3)

  l1<-tita1*exp(-1.96*sqrt(var1))
  u1<-tita1*exp(1.96*sqrt(var1))
  l2<-tita2*exp(-1.96*sqrt(var2))
  u2<-tita2*exp(1.96*sqrt(var2))
  l3<-tita3*exp(-1.96*sqrt(var3))
  u3<-tita3*exp(1.96*sqrt(var3))

  L<-round(1+tita1-tita2-tita3-sqrt((tita1-l1)^2+(u2-tita2)^2+(u3-tita3)^2-
    2*r12*(tita1-l1)*(u2-tita2)-2*r13*(tita1-l1)*(u3-tita3)+
    2*r23*(u2-tita2)*(u3-tita3)),digits=3)

  U<-round(1+tita1-tita2-tita3+sqrt((u1-tita1)^2+(tita2-l2)^2+(tita3-l3)^2-
    2*r12*(u1-tita1)*(tita2-l2)-2*r13*(u1-tita1)*(tita3-l3)+
    2*r23*(tita2-l2)*(tita3-l3)),digits=3)

  list(IC.L=L,IC.U=U)
}
```

- **Función *limites22*.**

Argumentos de entrada: “Sest”: Valor estimado del SI  
“varlnS”: Valor de la varianza del logaritmo del SI.

Argumentos de salida: Devuelve una lista con los límites del IC para los datos del argumento de entrada.

```
limites22<-function(Sest,varlnS){  
  
  L<-round(log(Sest)-1.96*sqrt(varlnS),digits=3)  
  
  U<-round(log(Sest)+1.96*sqrt(varlnS),digits=3)  
  
  list(IC.L=L,IC.U=U)  
  
}
```

## Anexo C

### Función MOVER en regresión

- **Función *mover.reg*.**

Argumentos de entrada:      “data”: Data frame con únicamente las variables que entran en el modelo.  
                                   “dep”: Posición de la variable dependiente en data.  
                                   “inter”: Vector con las posiciones de las variables de interacción en data.

Argumentos de salida:        Devuelve un data frame con los valores de los tres índices y sus intervalos de confianza

```
mover.reg<-function(data,dep,inter){

  mod<-modelo(data,dep,inter)
  res1<-param.reri(mod[[2]])
  res2<-param.ap(mod[[2]])
  res3<-param.si(mod[[2]])
  lim.reri<-limites.reg(res1[[2]])
  lim.ap<-limites.reg(res2[[2]])
  lim.lnSI<-limites.si(res3[[3]])

  valores<-c(res1[[1]],res2[[1]],res3[[2]],res3[[1]])
  L<-c(lim.reri$IC.L,lim.ap$IC.L,lim.lnSI$IC.L,
       round(exp(lim.lnSI$IC.L),digits=3))
  U<-c(lim.reri$IC.U,lim.ap$IC.U,lim.lnSI$IC.U,
       round(exp(lim.lnSI$IC.U),digits=3))

  data.frame("value"=valores,"IC.L"=L,"IC.U"=U,
            row.names=c("RERI","AP","lnSI","SI"))
}
```

- **Función *modelo*.**

Argumentos de entrada:      “data”: Data frame con únicamente las variables que entran en el modelo.  
                                   “dep”: Posición de la variable dependiente en data.  
                                   “inter”: Vector con las posiciones de las variables de interacción.

Argumentos de salida:       Devuelve una lista con el modelo de regresión con interacción y el summary del modelo

```

modelo<-function(data,dep,inter){

  a<-inter[1]
  b<-inter[2]
  c<-dep
  A<-data[,a]
  B<-data[,b]
  AB<-A*B
  var<-data[,c]
  cov<-data[,-c(a,b,c)]

  datos<-data.frame(AB,A,B,var,cov)
  modelo<-glm(var~.,data=datos,family=binomial,na.action=na.omit)

  list(modelo,summary(modelo))
}

```

- **Función *param.reri*.**

Argumentos de entrada:      “res”: El summary de un modelo de regresión con interacción (el resultado de la función *modelo*)

Argumentos de salida:       Devuelve una lista con el valor del RERI y un vector con los parámetros necesarios para el cálculo del IC MOVER

```

param.reri<-function(res){

  cov<-res$cov.scaled[-1,-1]
  est<-res$coefficients[2:4,1]

  v3<-cov[1,1]

```

```

v23<-cov[1,2]
v13<-cov[1,3]
v2<-cov[2,2]
v12<-cov[2,3]
v1<-cov[3,3]
var1<-v1+v2+v3+2*(v12+v13+v23)
var2<-v2
var3<-v1
cov1<-v1+v12+v13
cov2<-v12+v2+v23
cov3<-v12

tita1<-exp(est[1]+est[2]+est[3])
tita2<-exp(est[2])
tita3<-exp(est[3])
RERI<-round(tita1-tita2-tita3+1,digits=3)
param<-c(tita1,tita2,tita3,cov1,cov2,cov3,var1,var2,var3)

list(RERI,param)
}

```

- **Función *param.ap*.**

Argumentos de entrada: “res”: El summary de un modelo de regresión con interacción (el resultado de la función *modelo*)

Argumentos de salida: Devuelve una lista con el valor del AP y un vector con los parámetros necesarios para el cálculo del IC MOVER

```

param.ap<-function(res){

  cov<-res$cov.scaled[-1,-1]
  est<-res$coefficients[2:4,1]

  v3<-cov[1,1]
  v23<-cov[1,2]
  v13<-cov[1,3]
  v2<-cov[2,2]
  v12<-cov[2,3]
  v1<-cov[3,3]
  var1<-v1+v2+v3+2*(v12+v13+v23)
  var2<-v1+v3+2*v13
  var3<-v2+v3+2*v23
}

```

```

cov1<-v12+v13+v2+2*v23+v3
cov2<-v1+v12+2*v13+v23+v3
cov3<-v12+v23+v13+v3

tita1<-1/exp(est[1]+est[2]+est[3])
tita2<-1/exp(est[1]+est[3])
tita3<-1/exp(est[1]+est[2])
AP<-round(tita1-tita2-tita3+1,digits=3)
param<-c(tita1,tita2,tita3,cov1,cov2,cov3,var1,var2,var3)

list(AP,param)
}

```

- **Función *param.si*.**

Argumentos de entrada: “res”: El summary de un modelo de regresión con interacción (el resultado de la función *modelo*)

Argumentos de salida: Devuelve una lista con el valor del SI, del logaritmo del SI y la varianza del logaritmo del SI.

```

param.si<-function(res){

  cov<-res$cov.scaled[-1,-1]
  est<-res$coefficients[2:4,1]

  v1<-cov[1,1]
  v12<-cov[1,2]
  v13<-cov[1,3]
  v2<-cov[2,2]
  v23<-cov[2,3]
  v3<-cov[3,3]

  tita1<-exp(est[1]+est[2]+est[3])-1
  tita2<-exp(est[2])+exp(est[3])-2
  Sest<-tita1/tita2

  s3<-exp(est[1]+est[2]+est[3])/(exp(est[1]+est[2]+est[3])-1)
  s1<-s3-exp(est[2])/(exp(est[2])+exp(est[3])-2)
  s2<-s3-exp(est[3])/(exp(est[2])+exp(est[3])-2)

  varlnS<-v2*s1^2+v3*s2^2+v1*s3^2+2*s1*s2*v23+2*s1*s3*v12+2*s2*s3*v13
  lnSI<-round(log(tita1)-log(tita2),digits=3)
}

```

```

SI<-round(exp(lnSI),digits=3)
param<-c(Sest,varlnS)

list(SI,lnSI,param)
}

```

- **Función *limites.reg.***

Argumentos de entrada: “param”: Vector con los valores de los parámetros, las covarianzas entre ellos y los logaritmos de las varianzas de los parámetros (resultado de las funciones *param.reri* o *param.ap*)

Argumentos de salida: Devuelve una lista con los límites del IC MOVER para los datos del argumento de entrada.

```

limites.reg<-function(param){

  tita1<-param[1];tita2<-param[2];tita3<-param[3];
  cov1<-param[4];cov2<-param[5]
  cov3<-param[6];var1<-param[7];var2<-param[8];var3<-param[9]

  r12<-cov1/sqrt(var1*var2)
  r13<-cov2/sqrt(var1*var3)
  r23<-cov3/sqrt(var2*var3)

  l1<-tita1*exp(-1.96*sqrt(var1))
  u1<-tita1*exp(1.96*sqrt(var1))
  l2<-tita2*exp(-1.96*sqrt(var2))
  u2<-tita2*exp(1.96*sqrt(var2))
  l3<-tita3*exp(-1.96*sqrt(var3))
  u3<-tita3*exp(1.96*sqrt(var3))

  L<-round(1+tita1-tita2-tita3-sqrt((tita1-l1)^2+(u2-tita2)^2+(u3-tita3)^2-
    2*r12*(tita1-l1)*(u2-tita2)-2*r13*(tita1-l1)*(u3-tita3)+
    2*r23*(u2-tita2)*(u3-tita3)),digits=3)

  U<-round(1+tita1-tita2-tita3+sqrt((u1-tita1)^2+(tita2-l2)^2+(tita3-l3)^2-
    2*r12*(u1-tita1)*(tita2-l2)-2*r13*(u1-tita1)*(tita3-l3)+
    2*r23*(tita2-l2)*(tita3-l3)),digits=3)

  list(IC.L=L,IC.U=U)
}

```

- **Función *limites.si*.**

Argumentos de entrada: “param”: vector con los valores del SI y la varianza del logaritmo del SI.

Argumentos de salida: Devuelve una lista con los límites del IC para los datos del argumento de entrada.

```
limites.si<-function(param){
  Sest<-param[1];varlnS<-param[2]
  L<-round(log(Sest)-1.96*sqrt(varlnS),digits=3)
  U<-round(log(Sest)+1.96*sqrt(varlnS),digits=3)
  list(IC.L=L,IC.U=U)
}
```