



Universidad Nacional Autónoma de México

Maestría en Ciencias Matemáticas

Modelación estadística de datos espaciales

Proyecto 2

Profesor Dr. Carlos Díaz Ávalos

Elaborado por:

Alarcón González Edgar Gerardo

Índice

1. Resumen del Proyecto

2. Introducción

3. Análisis Exploratorio

3.1 Descripción de los datos

3.2 Mapas Descriptivos

4. Análisis sin Efecto Espacial

4.1 Forma del Modelo Lineal Generalizado

4.2 Implementación de algunos modelos no espaciales

4.3 Necesidad de un modelo espacial

5. Modelo SAR

5.1 Forma del Modelo SAR

5.2 Implementación de algunos modelos espaciales SAR

6. Modelo CAR

6.1 Forma del Modelo CAR

6.2 Implementación de algunos modelos espaciales CAR

7. Selección del Mejor Modelo

7.1 Comparativa de Modelos

7.2 Simulaciones

8. Conclusiones y respuestas concretas

8.1 Pregunta 1

8.2 Pregunta 2

8.3 Pregunta 3

9. Bibliografía

1. Resumen del Proyecto

El proyecto tiene como objetivo **estimar el riesgo relativo** de muerte de cuna para los años 1974-1978 y 1979-1984 para el estado de California del Norte por condado. Para esto, comenzaremos dando una introducción de qué es la muerte de cuna y algunos otros datos derivados de esta enfermedad. Posteriormente vamos a mostrar un análisis descriptivo de los datos con los que contamos para realizar un modelo estadístico que pueda estimar el riesgo relativo por condado. Luego, realizamos un modelo lineal generalizado que no contempla el efecto espacial con la finalidad de mostrar que es necesario acudir a esta rama de la estadística. Una vez expuesto el punto anterior, realizamos un par de modelos espaciales del tipo SAR y CAR, para finalmente buscar el mejor modelo por periodo y culminar respondiendo a tres preguntas importantes derivadas de este análisis.

2. Introducción

Los datos para este proyecto corresponden a un estudio sobre el riesgo relativo de muerte de cuna (Sudden Infant Death Syndrome) hecho en Carolina del Norte, EUA y publicado en 1989. Históricamente, este tipo de muertes podrían no tener una causa aparente y ocurre generalmente durante el sueño de un bebé, menor de un año de edad, que a primera instancia luce saludable. La denominación de este síndrome viene dada porque en la mayoría de los casos los infantes fallecen en sus cunas [1].

Aunque la causa de este padecimiento sea desconocida, parece que el también conocido como síndrome de muerte infantil podría asociarse con anomalías en la porción del cerebro del bebé que controla la respiración y la vuelta al estado de vigilia. Durante varios años, ha habido descubrimientos sobre algunos de los factores que podrían aumentar el riesgo de los menores a esta enfermedad.

Entre algunos de los factores que pueden aumentar el riesgo de este síndrome se sabe que pueden afectar: defectos cerebrales, bajo peso al nacer, infección respiratoria, dormir boca abajo o de costado, dormir sobre una superficie blanda, compartir una cama, calor excesivo, sexo, edad, raza, antecedentes familiares, fumador pasivo, ser prematuro, entre otros.

Para propósitos de este proyecto, vamos a buscar la existencia de alguna relación espacial con el riesgo de este síndrome, basándonos en la información antes mencionada.

3. Análisis Exploratorio

El objetivo de esta sección es conocer los datos con los que vamos a desarrollar la modelación del Riesgo Relativo. Así como conocer las variables auxiliares derivadas de nuestra información que nos ayudarán a esta tarea.

3.1 Descripción de los datos

Contamos con los siguientes datos.

- CNTY_ID, AREA, PERIMETER, CNTY_, NAME, FIPS, FIPSN0 y CRESS_ID son identificadores de los condados.
- BIR74, SID74, NWBIR74, BIR79, SID79 y NWBIR79 son los números de nacimientos totales, numero de muertes de cuna y nacimientos de niños “no blancos” para los años 1974-78 y 1979-1984 respectivamente.
- east, north, x, y, lon y lat son las coordenadas en diferentes proyecciones geográficas.
- L_id, M_id y geometry son características geométricas de los condados en el SIG (Sistema de Información Geográfica).

Como aclaración, basados en las variables anteriores, vamos a definir el **Riesgo Relativo** para el condado i como

$$RR_i := \frac{\frac{SID_i}{BIR_i}}{\sum_{BIR_i} \frac{SID_i}{BIR_i}} = \frac{SID_i}{BIR_i \sum_{BIR_i} \frac{SID_i}{BIR_i}} := \frac{SID_i}{E_i}$$

Donde E_i representa la cantidad de muertos “esperada” en el condado i . Otra manera de pensarlo, es que, el riesgo relativo es tal que

$$SID_i = RR_i E_i$$

3.2 Mapas Descriptivos

Comenzamos mostrando un mapa del estado de California del Norte (Figura 1). Este gráfico lo estaremos mostrando en repetidas ocasiones buscando explicar con estilo mapa de calor interacciones con diferentes resultados obtenidos y postulados en este proyecto. En las tablas al final del documento (Anexo A y B) podremos ver los nombres de los condados con base en el número expuesto.

Mapa de condados de California del Norte

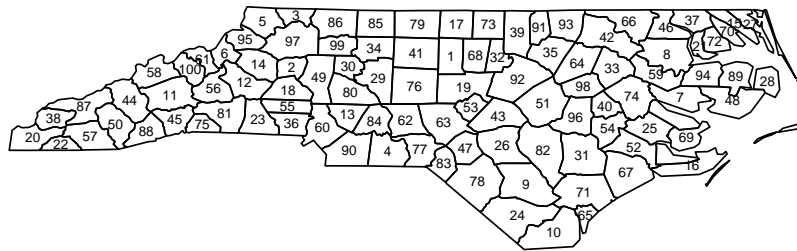


Figura 1: Estado de California del Norte con ID.

Nota: La enumeración del mapa viene dada por la variable CRESS_ID la cual consiste en un ordenamiento alfabético de los nombres de los condados en este estado. Esta enumeración fue propuesta en [2].

Recordemos que contamos con condados, de tal manera que más adelante realicemos un modelo espacial del tipo discreto definiendo una estructura de vecinos (Figura 2).

Hasta este punto, los gráficos que hemos mostrado dependen únicamente de aspectos espaciales. Ahora bien, realizaremos nuestro primer gráfico que realmente depende de los datos de interés por periodo, en [2] se propone una pequeña guía descriptiva de los datos. Una de las herramientas que se utilizan es una prueba de hipótesis derivada por el autor del mismo nombre **Choynowski**. Categorizando con cierto criterio a los condados como rangos “altos” y “bajos” al potencial de muerte de cuna (vistos de forma independiente), lo que busca esta prueba es esencialmente contrastar la siguiente hipótesis

$$H_0 : \text{El condado en cuestión NO pertenece al grupo asignado.}$$

Esta prueba la realizaremos para los periodos que estamos trabajando (1974-1978 y 1979-1984), con la finalidad de observar si existen zonas geográficas compuestas por conglomerados de condados que posean una tendencia visible. Este tipo de gráficos están relacionados con los llamados **Mapas de Probabilidad** (Figura 3). En este mapa, podemos observar que existen condados como lo son el 78 (*Robeson*) y el 24 (*Columbus*) que para ambos periodos pertenecen al mismo grupo gracias a la interpretación de su *p-value*, más aún, se puede ver a partir de esto que existen zonas hechas por conglomerados de condados que parecen pertenecer a un mismo grupo.

Ahora bien, vale la pena en este punto realizar los cálculos de los riesgos relativos para cada uno de los periodos y aplicar una prueba no paramétrica para contrastar la siguiente hipótesis nula de que los riesgos relativos **de forma**

Relación de vecinos por condado

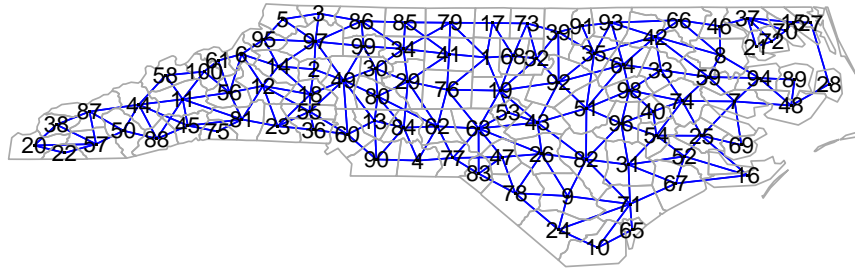


Figura 2: Vecinos de condados con ID.

Nota: Esta relación de vecinos se obtuvo con base en la función `readShapeSpatial()` en R. Aunque no coincide por la propuesta en [2].

Choynowski

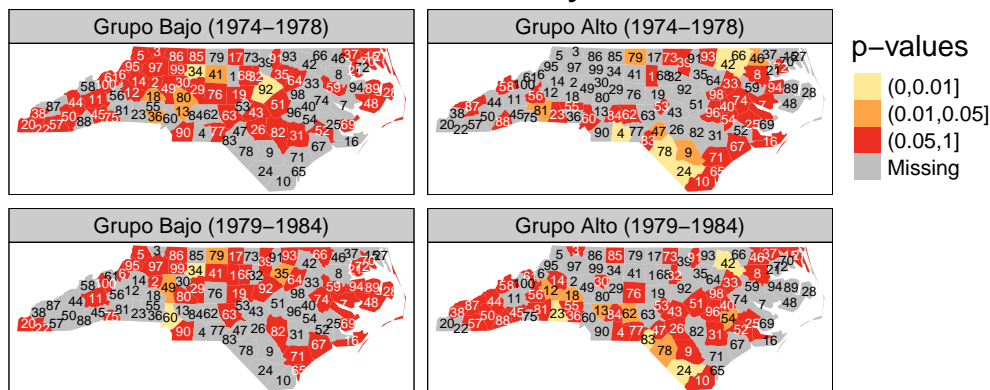


Figura 3: Prueba de Choynowski para los periodos de estudio.

Nota: Esto se obtuvo derivado de [3], donde se usa la función `choynowski()` en R y también se propone una implementación de código en el apartado descriptivo de [2].

general pensado que los condados son intercambiables y sin contemplar posición del mismo siguen una distribución estadísticamente equivalente.

$$H_0 : F_{RR(74)} = F_{RR(79)}$$

Esto se hará con la prueba de bondad de ajuste de Kolmogorov-Smirnov (que en código de R corresponde a `ks.test()`).

El resultado de esta prueba nos arroja un $p - value = 0.155 > 0.05$ y por lo tanto NO rechazamos H_0 . Algo que hasta cierto punto es sorprendente pues bajo las hipótesis mencionadas, el riesgo relativo tomando de forma indistinta cualquier condado conserva su distribución entre periodos. Esta prueba de bondad de ajuste se usará más adelante en la sección 7 donde usaremos simulaciones para observar el comportamiento de los modelos propuestos en las secciones de la 4 a la 6.

Una vez calculados los riesgos relativos en cada periodo, mostramos en el mapa un gráfico de intensidad por condado (Figura 4).

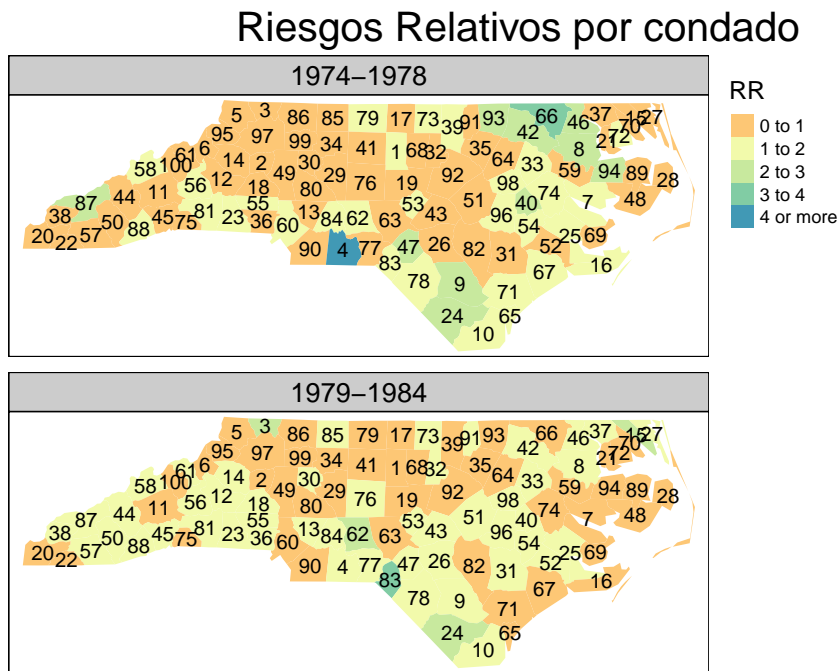


Figura 4: Riesgos Relativos en cada condado por periodo.

Gracias a la Figura 4, volvemos a ver la presencia de efecto espacial, debido al comportamiento que tienen la variable medida en la geografía del estado. En este mapa se puede observar que los vecinos parecen guardar una relación con respecto al RR. Por último en esta sección y aprovechando el cálculo de los riesgos relativos por periodo, vamos a mostrar un mapa donde observaremos el **cociente de riesgos relativos** en cada condado (Figura 5). Formalmente y con la finalidad de observar el impacto en magnitud de cada RR, el cálculo que proponemos para este cociente está dado para el estado i como

$$CRR_i = \frac{RR_i^{(74)} + 0.5}{RR_i^{(79)} + 0.5}$$

De tal manera que cuando $CRR_i > 1$ es porque el riesgo relativo disminuyó en el estado i **de un periodo a otro**, mientras que cuando $CRR_i = 1$ es porque se mantuvo y finalmente si $CRR_i < 1$ es porque el RR aumentó.

De la Figura 5 podemos concluir que hay condados como el 4 (*Anson*), 79 (*Rockingham*), 35 (*Franklin*) y 94 (*Washington*) que aumentan su RR disminuyó, por otro lado condados como el 3 (*Alleghany*) y 37 (*Gates*) aumentó, mientras que los que se ven en blanco, parece que se mantuvieron.

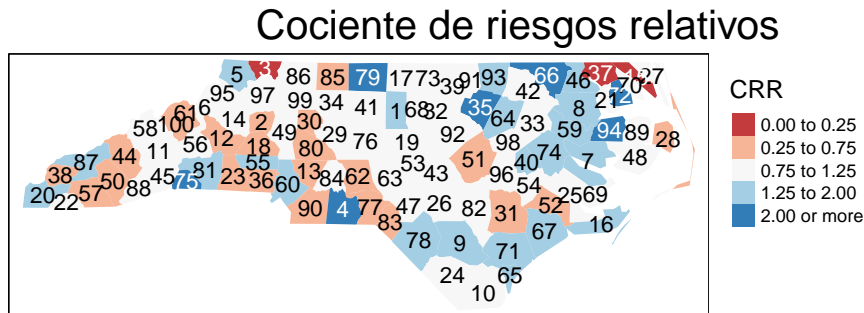


Figura 5: Cociente de Riesgos Relativos por condado

Una vez habiendo revisado una posible existencia de relación espacial de los datos, vamos a continuar justificando la incorporación de la metodología espacial primeramente con un argumento más clásico que nos oriente en esta dirección.

4. Análisis sin Efecto Espacial

En esta sección vamos a intentar omitir el uso de la estadística espacial, de tal manera que observaremos cómo este análisis nos presionará a la implementación de esta metodología.

4.1 Forma del Modelo Lineal Generalizado

Inspirados en [3], los modelos que se van a implementar aquí asumirán que las muertes de cuna dadas en el condado i son tales que $SID_i \sim Poisson(\mu_i)$, BIR_i son los nacimientos en el condado y X_i sus covariables. De esta manera nosotros modelamos

$$\log(\hat{\mu}_i) = \log(BIR_i) + \beta X_i \Leftrightarrow \log\left(\frac{\hat{\mu}_i}{BIR_i}\right) = \beta X_i$$

Esto significa que estaremos metiendo la variable BIR como *offset* del modelo. De tal manera que la sintaxis del mismo en código de R viene dada de forma genérica como

```
glm(SID ~ offset(log(BIR)) + X, family = "poisson")
```

A continuación vamos a estudiar el comportamiento de algunos candidatos de modelos lineales que puedan explicar el **Riesgo Relativo**.

4.2 Implementación de algunos modelos no espaciales

Con la finalidad de usar variables que no estén relacionadas directamente con la **posición espacial** vamos a implementar los modelos A y B como vienen a continuación.

A. Vamos a proponer un modelo que incluya las variables AREA, PERIMETER y NWBIR para cada uno de los periodos.

Tabla 1: Modelo NO espacial A para el periodo del 74

| | exp(coef) | CI(95 %) | p-value | Signif. |
|-------------|-----------|----------------|---------|---------|
| (Intercept) | 0.001 | (0.001,0.002) | 0 | *** |
| AREA | 6.34 | (0.649,64.302) | 0.115 | |
| PERIMETER | 1.052 | (0.791,1.367) | 0.716 | |
| NWBIR74 | 1 | (1,1) | 0.445 | |

Tabla 2: Modelo NO espacial A para el periodo del 79

| | exp(coef) | CI(95 %) | p-value | Signif. |
|-------------|-----------|----------------|---------|---------|
| (Intercept) | 0.002 | (0.002,0.003) | 0 | *** |
| AREA | 3.409 | (0.432,28.045) | 0.249 | |
| PERIMETER | 0.881 | (0.675,1.124) | 0.327 | |
| NWBIR79 | 1 | (1,1) | 0.009 | ** |

Tabla 3: Modelo NO espacial B para el periodo del 74

| | exp(coef) | CI(95 %) | p-value | Signif. |
|-------------|-----------|---------------|---------|---------|
| (Intercept) | 0.002 | (0.002,0.002) | 0 | *** |
| NWBIR74 | 1 | (1,1) | 0.937 | |

Tabla 4: Modelo NO espacial B para el periodo del 79

| | exp(coef) | CI(95 %) | p-value | Signif. |
|-------------|-----------|---------------|---------|---------|
| (Intercept) | 0.002 | (0.002,0.002) | 0 | *** |
| NWBIR79 | 1 | (1,1) | 0.012 | * |

B. Vamos a proponer un modelo que incluya únicamente la variable NWBIR para cada uno de los periodos.

En general, podemos ver que realmente las covariables no parecen ser significativas, observando que en particular el coeficiente de NWBIR79 es estimado como muy cercano a cero. De tal manera que parece que, al menos con los datos que se nos presentan, no podemos explicar la media de las muertes por el síndrome de cuna. De tal manera que nuestro posible mejor candidato es el modelo tal que únicamente incluye al intercepto (β_0). En otras palabras, nuestro “**modelo 1**” será tal que

$$\log(\hat{\mu}_i) = \log(BIR_i) + \beta_0 \Leftrightarrow \frac{\hat{\mu}_i}{BIR_i} = e^{\beta_0} \Leftrightarrow \hat{\mu}_i = BIR_i e^{\beta_0}$$

Algo que, aunque el coeficiente β_0 es de hecho significativo, quizás no sea del todo conveniente ya que en el fondo estamos diciendo que para cada estado i

$$\widehat{RR}_i = \frac{\frac{\hat{\mu}_i}{BIR_i}}{\frac{\sum \hat{\mu}_i}{\sum BIR_i}} = \frac{e^{\beta_0}}{\frac{\sum BIR_i e^{\beta_0}}{\sum BIR_i}} = 1$$

De tal manera que el mapa de los residuales para el riesgo relativo coincide prácticamente con el que obtenemos en la Figura 4. Algo que como ya se había comentado, parece tener presente un efecto espacial sobre los datos.

4.3 Necesidad de un modelo espacial

Con todo esto veamos una prueba estadística que nos permitirá contrastar la hipótesis sobre el comportamiento espacial de los datos, pero primero, construiremos herramientas para poder aplicar avanzar en este sentido. Comenzamos mostrando la matriz de pesos (Figura 6) de los vecinos destacando que con base en nuestros datos, tenemos una estructura donde todos los condados cuentan con un vecino.

Una vez que ya tenemos esta matriz, lo que haremos es contrastar la hipótesis

$$H_0 : \text{Los datos NO presentan una correlación espacial.}$$



Figura 6: Matriz de pesos de los vecinos.

Nota: Esta matriz se obtuvo a partir de las funciones `poly2nb` y `nb2mat()` en R.

Esta prueba se le conoce por su autor del mismo nombre **Moran** y se le aplicará a los **residuales del Modelo 1**. Los resultados para cada periodo los vemos en la Tabla 5. *Nota:* La función en R para aplicar esta prueba es `Moran.I()`.

Tabla 5: Resultados de la prueba de Moran por periodo.

| Periodo | p-value | Signif. |
|-----------|---------|---------|
| 1974-1978 | 0.000 | *** |
| 1979-1984 | 0.018 | * |

De donde se concluye que ambos periodos en efecto tienen correlación espacial y por lo tanto debemos aplicar algún modelo correspondiente. Antes de comenzar una sección, vamos a introducir una nueva variable que llamaremos *both* construida como una combinación de las variables `L_id` y `M_id` lo cual nos da como resultado lo que vemos en la Figura 7.

5. Modelo SAR

En esta sección haremos la implementación de modelos **SAR** para la modelación del riesgo relativo.

5.1 Forma del Modelo SAR

En un estilo de *pseudo-código* la manera en que estaremos haciendo la implementación del modelo será la siguiente incluyendo la estructura de vecinos que construimos anteriormente.

```
lagsarlm(RR ~ BIR+X, listw = vecinos)
```

Donde `X` son las covariables que introduciremos los modelos.

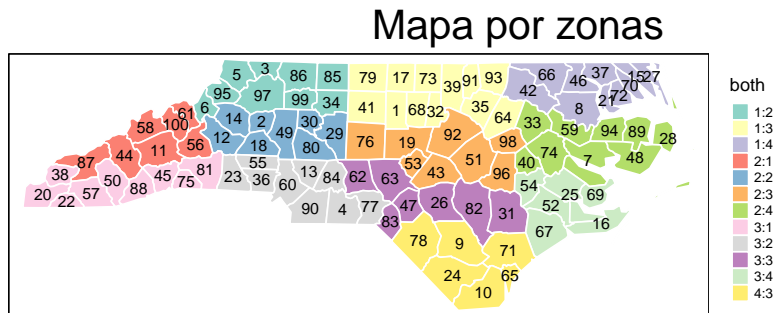


Figura 7: Mapa por zonas both.

Nota: La variable both se creó con base en la función `paste()` en R de las variables `L_id` y `M_id`.

5.2 Implementación de algunos modelos espaciales SAR

- Como un primer acercamiento, meteremos al modelo solo la variable BIR para ambos periodos (Tablas 6 y 7).

Tabla 6: Modelo (2) espacial para el periodo del 74

| | coef | p-value | Signif. |
|-------------|-------|---------|---------|
| (Intercept) | 1.043 | 0 | *** |
| BIR74 | 0 | 0.884 | |

Tabla 7: Modelo (2) espacial para el periodo del 79

| | coef | p-value | Signif. |
|-------------|-------|---------|---------|
| (Intercept) | 0.864 | 0 | *** |
| BIR79 | 0 | 0.555 | |

Con la finalidad de dar potencia a la variable BIR y a la covariable both que construimos anteriormente, procedemos quitando el intercepto del modelo.

- Con respecto al modelo anterior, quitamos el intercepto y la variable both (Tabla 8).

Tabla 8: Modelo (3) espacial para los periodos del 74 y del 79

| | coef | p-value | Signif. | | coef | p-value | Signif. |
|---------|-------|---------|---------|---------|-------|---------|---------|
| BIR74 | 0 | 0.564 | | BIR79 | 0 | 0.18 | |
| both1:2 | 0.421 | 0.135 | | both1:2 | 0.568 | 0.012 | * |
| both1:3 | 1.191 | 0 | *** | both1:3 | 0.614 | 0.003 | ** |
| both1:4 | 1.408 | 0 | *** | both1:4 | 1.11 | 0 | *** |
| both2:1 | 0.87 | 0.003 | ** | both2:1 | 0.893 | 0 | *** |
| both2:2 | 0.545 | 0.067 | . | both2:2 | 0.752 | 0.002 | ** |
| both2:3 | 0.993 | 0.001 | ** | both2:3 | 0.902 | 0 | *** |
| both2:4 | 1.084 | 0 | *** | both2:4 | 0.335 | 0.122 | |
| both3:1 | 0.802 | 0.004 | ** | both3:1 | 0.699 | 0.001 | ** |
| both3:2 | 1.388 | 0 | *** | both3:2 | 1.066 | 0 | *** |
| both3:3 | 1.296 | 0 | *** | both3:3 | 1.438 | 0 | *** |
| both3:4 | 1.155 | 0 | *** | both3:4 | 0.911 | 0 | *** |
| both4:3 | 1.761 | 0 | *** | both4:3 | 0.996 | 0 | *** |

Más adelante haremos una comparación entre todos los modelos para seleccionar la mejor propuesta.

6. Modelo CAR

En esta sección haremos la implementación de modelos **CAR** para la modelación del riesgo relativo.

6.1 Forma del Modelo CAR

En un estilo de *pseudo-código* la manera en que estaremos haciendo la implementación del modelo será la siguiente incluyendo la estructura de vecinos que construimos anteriormente.

```
spautolm(RR ~ BIR+X, listw = vecinos, weights = BIR, family = "CAR")
```

Donde X son las covariables que introduciremos los modelos.

6.2 Implementación de algunos modelos espaciales CAR

- Como un primer acercamiento, meteremos al modelo solo la variable BIR para ambos periodos (Tablas 9 y 10)

Tabla 9: Modelo (4) espacial para el periodo del 74

| | coef | p-value | Signif. |
|-------------|-------|---------|---------|
| (Intercept) | 1.121 | 0 | *** |
| BIR74 | 0 | 0.074 | . |

Tabla 10: Modelo (4) espacial para el periodo del 79

| | coef | p-value | Signif. |
|-------------|-------|---------|---------|
| (Intercept) | 1.145 | 0 | *** |
| BIR79 | 0 | 0.008 | ** |

Con la finalidad de dar potencia a la variable BIR y a la covariable both que construimos anteriormente, procedemos quitando el intercepto del modelo.

- Con respecto al modelo anterior, quitamos el intercepto y la variable both (Tabla 11).

Más adelante haremos una comparación entre todos los modelos para seleccionar la mejor propuesta .

7. Selección del Mejor Modelo

7.1 Comparativa de Modelos

Para seleccionar el modelo, nos vamos a fijar en las estadísticas de la devianza, el AIC y el BIC. Esto lo podemos ver en la Tabla 12 para ambos periodos. De donde concluiremos, basados en el comportamiento de los datos que los modelos seleccionados serán el 4 y el 5 para los periodos del 74 y 79 respectivamente. Procedemos a ver un mapa de los residuales de estos modelos (Figura 8) y también de los valores ajustados (Figura 9).

De las Figuras 8 y 9 vemos que el modelo para el periodo del 74 NO parece explicar el efecto espacial por lo que quizás necesitaríamos más variables para hacerlo, mientras que para el periodo del 79 logramos hacer que el modelo absorba el efecto espacial y esto se ve con los residuales tomando un comportamiento que parece “más” aleatorio. Esto se puede verificar con la prueba de Moran realizada en la Tabla 13.

Tabla 11: Modelo (5) espacial para los periodos del 74 y del 79

| | coef | p-value | Signif. | | coef | p-value | Signif. |
|---------|-------|---------|---------|---------|-------|---------|---------|
| BIR74 | 0 | 0.072 | . | BIR79 | 0 | 0 | *** |
| both1:2 | 0.56 | 0.002 | ** | both1:2 | 0.84 | 0 | *** |
| both1:3 | 1.126 | 0 | *** | both1:3 | 1.059 | 0 | *** |
| both1:4 | 1.96 | 0 | *** | both1:4 | 1.387 | 0 | *** |
| both2:1 | 0.795 | 0 | *** | both2:1 | 1.158 | 0 | *** |
| both2:2 | 0.597 | 0 | *** | both2:2 | 1.081 | 0 | *** |
| both2:3 | 0.968 | 0 | *** | both2:3 | 1.265 | 0 | *** |
| both2:4 | 1.353 | 0 | *** | both2:4 | 0.825 | 0 | *** |
| both3:1 | 1.169 | 0 | *** | both3:1 | 1.099 | 0 | *** |
| both3:2 | 1.157 | 0 | *** | both3:2 | 1.323 | 0 | *** |
| both3:3 | 1.214 | 0 | *** | both3:3 | 1.569 | 0 | *** |
| both3:4 | 1.318 | 0 | *** | both3:4 | 1.2 | 0 | *** |
| both4:3 | 1.765 | 0 | *** | both4:3 | 1.385 | 0 | *** |

Tabla 12: Comparativa de los modelos para los periodos del 74 y del 79

| | Devianzas | AIC | BIC | | Devianzas | AIC | BIC |
|----------|-----------|----------|----------|----------|-----------|----------|----------|
| Modelo 1 | 203.34364 | 510.7536 | 513.3588 | Modelo 1 | 167.85426 | 502.1247 | 504.7299 |
| Modelo 2 | 59.94556 | 240.6239 | 251.0445 | Modelo 2 | 36.02393 | 190.5166 | 200.9372 |
| Modelo 3 | 47.59432 | 239.5479 | 278.6254 | Modelo 3 | 28.28504 | 188.9840 | 228.0615 |
| Modelo 4 | NA | 228.2305 | 238.6511 | Modelo 4 | NA | 174.2651 | 184.6858 |
| Modelo 5 | NA | 207.5008 | 246.5783 | Modelo 5 | NA | 173.4027 | 212.4803 |

Tabla 13: Resultados de la prueba de Moran sobre los residuales de los modelos seleccionados por periodo.

| Periodo | p-value | Signif. |
|-----------|---------|---------|
| 1974-1978 | 0.000 | *** |
| 1979-1984 | 0.266 | |

7.2 Simulaciones

Con base en el modelo obtenido para ambos periodos y con la finalidad de dar respuesta a una de las preguntas de la siguiente sección, procedemos a hacer simulaciones de los riesgos relativos con base en los modelos obtenidos. La idea consiste en utilizar estos modelos para asignar a los parámetros de los modelos (los coeficientes β) una distribución Normal con media el valor estimado y varianza el error estándar que tiene el modelo con la finalidad de obtener simulaciones de los Riesgos Relativos (variable respuesta) generando así una muestra \underline{RR}_i para cada uno de los condados y cada uno de los periodos.

Este procedimiento se realizó internamente en el código y posteriormente se realizó una prueba de bondad de ajuste similar a la realizada en la sección 3.2 para contrastar la hipótesis nula de tener distribuciones estadísticamente equivalentes pero esta vez con la oportunidad de hacerlo por condado y por construcción del modelo, incluyendo también el efecto aleatorio.

$$H_0 : F_{\underline{RR}_i^{(74)}} = F_{\underline{RR}_i^{(79)}}$$

En este caso concluiremos basándonos en los *p-values* obtenidos de este procedimiento para interpretar “si la hipótesis nula no se rechaza entonces no hubo diferencias significativas de entre los riesgos relativos de cada periodo”. Los resultados obtenidos se encuentran de forma resumida en la tabla 14

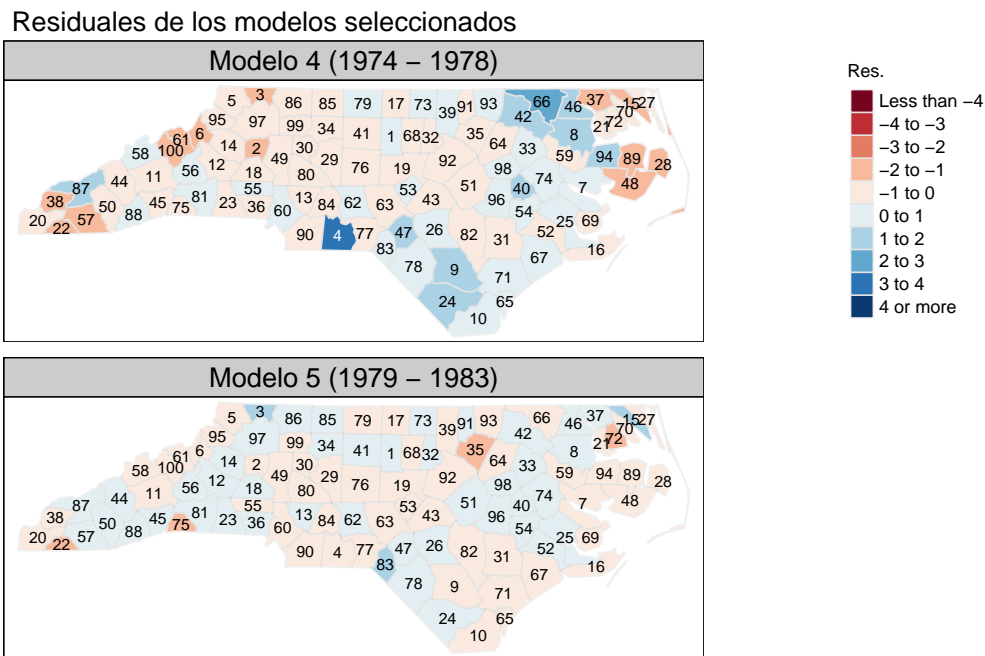


Figura 8: Residuales de los modelos seleccionados con base en la Tabla 12.

Tabla 14: Cantidad de condados donde hubo un cambio de un periodo a otro basados en los modelos.

| Hubo Cambio | No hubo cambio |
|-------------|----------------|
| 98 | 2 |

Y de hecho los estados que no cambiaron fueron el 60 (*Mecklenburg*) y el 92 (*Wake*). Debemos tener en mente que este resultado se piensa de tal manera que ya contemplamos a todos los estados. Es decir, si por ejemplo un condado tiene muchas muertes por síndrome de cuna en el primer periodo y en el siguiente le tiene aún más muertes, en comparativa con el resto de los condados éste mantuvo su posición.

8. Conclusiones y respuestas concretas

En esta sección daremos solución en concreto con base en los resultados que hemos establecido a lo largo del proyecto de las incógnitas que fueron planteadas para la elaboración del mismo.

8.1 Pregunta 1

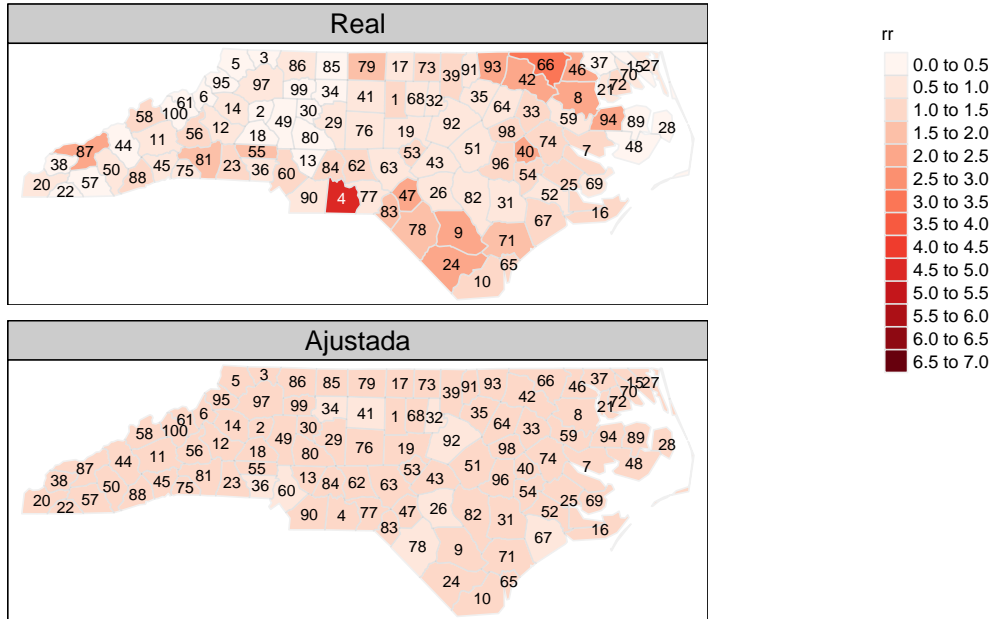
¿Existen diferencias significativas de entre los riesgos relativos de cada periodo?

Teniendo una tasa de cambio del 98 % de acuerdo a lo realizado en la sección 7.2, podemos decir que sí hubo cambio entre los periodos. Muy a pesar de que en la sección 3.2 parecía que no habría cambios, sin embargo, vimos que el efecto espacial afectó esta decisión.

8.2 Pregunta 2

¿Hay alguna zona del estado donde los riesgos y las diferencias entre periodos sean mas altas o bajas?

Ajuste del Modelo (4) CAR periodo 74



Ajuste del Modelo (5) CAR periodo 79

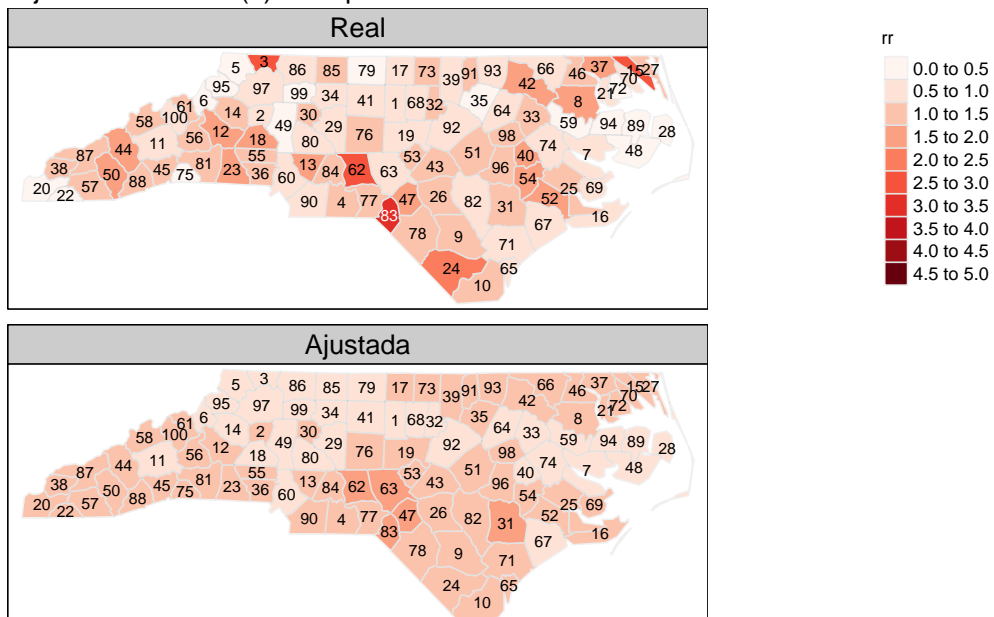


Figura 9: Valores ajustados de los modelos seleccionados con base en la Tabla 12.

Con base en la sección 3.2 - Figura 5 vemos que en general sí hay cambio entre los periodos, nuevamente, las zonas que están diferente de color blanco, consideramos que han sufrido un cambio representativo de un periodo a otro.

8.3 Pregunta 3

¿Cuál fue el mejor modelo ajustado y cual fue el criterio de comparación entre modelos para llegar a esa conclusión?

Esto se aclara en la sección 7.1, pero en concreto, son modelos derivados de la metodología estadística-espacial CAR.

9. Bibliografía

Referencias

- [1] [Mayoclinic.org, Sudden Infant Death Syndrome.](#)
- [2] [Noel A. C. Cressie, Statistics for Spatial Data. John Wiley & Sons, INC.](#)
- [3] [Roger Bivand, Introduction to the North Carolina SIDS data set \(re-revised\).](#)
- [4] [Autor desconocido, Repositorio de CRAN.](#)

Nota: Toda la bibliografía viene dada con un hipervínculo que los llevará a la publicación original por internet (Actualizado al 14/Enero/2021).

Tabla 15: Anexo A - Datos de los Estados

| CRESS_ID | Nombre | BIR74 | SID74 | BIR79 | SID79 | CRESS_ID | Nombre | BIR74 | SID74 | BIR79 | SID79 |
|----------|-----------|-------|-------|-------|-------|----------|------------|-------|-------|-------|-------|
| 1 | Alamance | 4672 | 13 | 5767 | 11 | 26 | Cumberland | 20366 | 38 | 26370 | 57 |
| 2 | Alexander | 1333 | 0 | 1683 | 2 | 27 | Currituck | 508 | 1 | 830 | 2 |
| 3 | Alleghany | 487 | 0 | 542 | 3 | 28 | Dare | 521 | 0 | 1059 | 1 |
| 4 | Anson | 1570 | 15 | 1875 | 4 | 29 | Davidson | 5509 | 8 | 7143 | 8 |
| 5 | Ashe | 1091 | 1 | 1364 | 0 | 30 | Davie | 1207 | 1 | 1438 | 3 |
| 6 | Avery | 781 | 0 | 977 | 0 | 31 | Duplin | 2483 | 4 | 2777 | 7 |
| 7 | Beaufort | 2692 | 7 | 2909 | 4 | 32 | Durham | 7970 | 16 | 10432 | 22 |
| 8 | Bertie | 1324 | 6 | 1616 | 5 | 33 | Edgecombe | 3657 | 10 | 4359 | 9 |
| 9 | Bladen | 1782 | 8 | 2052 | 5 | 34 | Forsyth | 11858 | 10 | 15704 | 18 |
| 10 | Brunswick | 2181 | 5 | 2655 | 6 | 35 | Franklin | 1399 | 2 | 1863 | 0 |
| 11 | Buncombe | 7515 | 9 | 9956 | 18 | 36 | Gaston | 9014 | 11 | 11455 | 26 |
| 12 | Burke | 3573 | 5 | 4314 | 15 | 37 | Gates | 420 | 0 | 594 | 2 |
| 13 | Cabarrus | 4099 | 3 | 5669 | 20 | 38 | Graham | 415 | 0 | 488 | 1 |
| 14 | Caldwell | 3609 | 6 | 4249 | 9 | 39 | Granville | 1671 | 4 | 2074 | 4 |
| 15 | Camden | 286 | 0 | 350 | 2 | 40 | Greene | 870 | 4 | 1178 | 4 |
| 16 | Carteret | 2414 | 5 | 3339 | 4 | 41 | Guilford | 16184 | 23 | 20543 | 38 |
| 17 | Caswell | 1035 | 2 | 1253 | 2 | 42 | Halifax | 3608 | 18 | 4463 | 17 |
| 18 | Catawba | 5754 | 5 | 6883 | 21 | 43 | Harnett | 3776 | 6 | 4789 | 10 |
| 19 | Chatham | 1646 | 2 | 2398 | 3 | 44 | Haywood | 2110 | 2 | 2463 | 8 |
| 20 | Cherokee | 1027 | 2 | 1173 | 1 | 45 | Henderson | 2574 | 5 | 3679 | 8 |
| 21 | Chowan | 751 | 1 | 899 | 1 | 46 | Hertford | 1452 | 7 | 1838 | 5 |
| 22 | Clay | 284 | 0 | 419 | 0 | 47 | Hoke | 1494 | 7 | 1706 | 6 |
| 23 | Cleveland | 4866 | 10 | 5526 | 21 | 48 | Hyde | 338 | 0 | 427 | 0 |
| 24 | Columbus | 3350 | 15 | 4144 | 17 | 49 | Iredell | 4139 | 4 | 5400 | 5 |
| 25 | Craven | 5868 | 13 | 7595 | 18 | 50 | Jackson | 1143 | 2 | 1504 | 5 |

Tabla 16: Anexo B - Datos de los Estados

| CRESS_ID | Nombre | BIR74 | SID74 | BIR79 | SID79 | CRESS_ID | Nombre | BIR74 | SID74 | BIR79 | SID79 |
|----------|-------------|-------|-------|-------|-------|----------|--------------|-------|-------|-------|-------|
| 51 | Johnston | 3999 | 6 | 4780 | 13 | 76 | Randolph | 4456 | 7 | 5711 | 12 |
| 52 | Jones | 578 | 1 | 650 | 2 | 77 | Richmond | 2756 | 4 | 3108 | 7 |
| 53 | Lee | 2252 | 5 | 2949 | 6 | 78 | Robeson | 7889 | 31 | 9087 | 26 |
| 54 | Lenoir | 3589 | 10 | 4225 | 14 | 79 | Rockingham | 4449 | 16 | 5386 | 5 |
| 55 | Lincoln | 2216 | 8 | 2817 | 7 | 80 | Rowan | 4606 | 3 | 6427 | 8 |
| 56 | McDowell | 1946 | 5 | 2215 | 5 | 81 | Rutherford | 2992 | 12 | 3543 | 8 |
| 57 | Macon | 797 | 0 | 1157 | 3 | 82 | Sampson | 3025 | 4 | 3447 | 4 |
| 58 | Madison | 765 | 2 | 926 | 2 | 83 | Scotland | 2255 | 8 | 2617 | 16 |
| 59 | Martin | 1549 | 2 | 1849 | 1 | 84 | Stanly | 2356 | 5 | 3039 | 7 |
| 60 | Mecklenburg | 21588 | 44 | 30757 | 35 | 85 | Stokes | 1612 | 1 | 2038 | 5 |
| 61 | Mitchell | 671 | 0 | 919 | 2 | 86 | Surry | 3188 | 5 | 3616 | 6 |
| 62 | Montgomery | 1258 | 3 | 1598 | 8 | 87 | Swain | 675 | 3 | 883 | 2 |
| 63 | Moore | 2648 | 5 | 3534 | 5 | 88 | Transylvania | 1173 | 3 | 1401 | 4 |
| 64 | Nash | 4021 | 8 | 5189 | 7 | 89 | Tyrrell | 248 | 0 | 319 | 0 |
| 65 | New Hanover | 5526 | 12 | 6917 | 9 | 90 | Union | 3915 | 4 | 5273 | 9 |
| 66 | Northampton | 1421 | 9 | 1606 | 3 | 91 | Vance | 2180 | 4 | 2753 | 6 |
| 67 | Onslow | 11158 | 29 | 14655 | 23 | 92 | Wake | 14484 | 16 | 20857 | 31 |
| 68 | Orange | 3164 | 4 | 4478 | 6 | 93 | Warren | 968 | 4 | 1190 | 2 |
| 69 | Pamlico | 542 | 1 | 631 | 1 | 94 | Washington | 990 | 5 | 1141 | 0 |
| 70 | Pasquotank | 1638 | 3 | 2275 | 4 | 95 | Watauga | 1323 | 1 | 1775 | 1 |
| 71 | Pender | 1228 | 4 | 1602 | 3 | 96 | Wayne | 6638 | 18 | 8227 | 23 |
| 72 | Perquimans | 484 | 1 | 676 | 0 | 97 | Wilkes | 3146 | 4 | 3725 | 7 |
| 73 | Person | 1556 | 4 | 1790 | 4 | 98 | Wilson | 3702 | 11 | 4706 | 13 |
| 74 | Pitt | 5094 | 14 | 6635 | 11 | 99 | Yadkin | 1269 | 1 | 1568 | 1 |
| 75 | Polk | 533 | 1 | 673 | 0 | 100 | Yancey | 770 | 0 | 869 | 1 |