

□

ANÁLISIS DE DATOS ESPACIALES EN EL ÁMBITO DE LA EPIDEMIOLOGÍA

Prof. Dr. Maria A Barceló y Prof. Dr. Marc Saez

8, 10, 14 y 16 de septiembre de 2021

Grupo de Investigación en Estadística, Econometría y Salud (GRECS), Universidad de Girona
CIBER de Epidemiología y Salud Pública (CIBERESP)

INTRODUCCIÓN AL CURSO

1. Introducción al curso
2. Introducción a la epidemiología y la estadística espacial
3. Panorámica de los modelos mixtos
4. Panorámica de los modelos mixtos - Prácticas
5. Introducción a INLA y R INLA
6. R INLA - Prácticas

Miércoles 8

Viernes 10

INTRODUCCIÓN AL CURSO

- 7. Mapas de enfermedades. Estandarización de razones de incidencia y mortalidad
 - 8. Mapas de enfermedades. Suavización de razones de incidencia y de mortalidad estandarizadas
 - 9. Mapas de enfermedades – Prácticas
 - 10. Estudios de asociación geográfica. Regresión ecológica espacial
 - 11. Regresión ecológica espacial - Prácticas
- Martes 14

INTRODUCCIÓN AL CURSO

12. Agrupación de casos

13. **Extensiones: BYM2, procesos puntuales, leaflet, pc priors**

14. Extensiones – Prácticas

} Jueves 16

EXTENSIONES

Veremos las siguientes extensiones:

- Modelo BYM2
- Procesos puntuales
- Leaflet (lo veremos en las prácticas)
- pc priors (lo veremos en las prácticas)

EXTENSIONES – EL MODELO BYM2

- En el modelo clásico de Besag, York y Mollie, BYM (Besag *et al.*, 1991), la variación estructurada espacialmente no es independiente de la variación no estructurada (problema llamado de la no identificabilidad).
- Como consecuencia, parte de la dependencia espacial (variación estructurada) podría ser realmente heterogeneidad (variación no estructurada) y viceversa.

EXTENSIONES — EL MODELO BYM2

- Existen formulaciones alternativas al modelo de BYM, como los modelos de Leroux (Leroux *et al.*, 2000) y de Dean (Dean *et al.*, 2001), en los que se asegura que la variación espacial estructurada sea independiente de la no estructurada.
- Sin embargo, ninguno de los dos modelos '*escala*' la variación espacial.
- Como consecuencia de ello, los hiperparámetros dependen de la estructura espacial del problema y no pueden ser interpretados correctamente.

EXTENSIONES – EL MODELO BYM2

- Por otro lado, en el contexto Bayesiano, la elección de las distribuciones a priori de los hiperparámetros (*priors*), puede tener un impacto considerable en los resultados.
- En los modelos de Leroux y de Dean se utilizan priors estándar que conducen a un sobreajustamiento, *overfitting* (multicolinealidad).
- La principal consecuencia del overfitting es que los estimadores de las varianzas son más grandes que los reales y, por lo tanto, los intervalos de credibilidad serán mucho más amplios que lo esperado, lo que implica que podríamos llegar a no rechazar la hipótesis nula (de que los coeficientes sean igual a cero) cuando en realidad deberíamos haberla rechazado.

EXTENSIONES – EL MODELO BYM2

- Simpson *et al.* (2017) propusieron una modificación del modelo BYM, denominada BYM2, que soluciona estos problemas, ya que escala la variación estructurada espacialmente y utiliza priors que penalizan la complejidad (llamados PC priors).
- Estos priors son robustos, en el sentido de que no tienen impacto en los resultados y, además, tienen una interpretación epidemiológica.

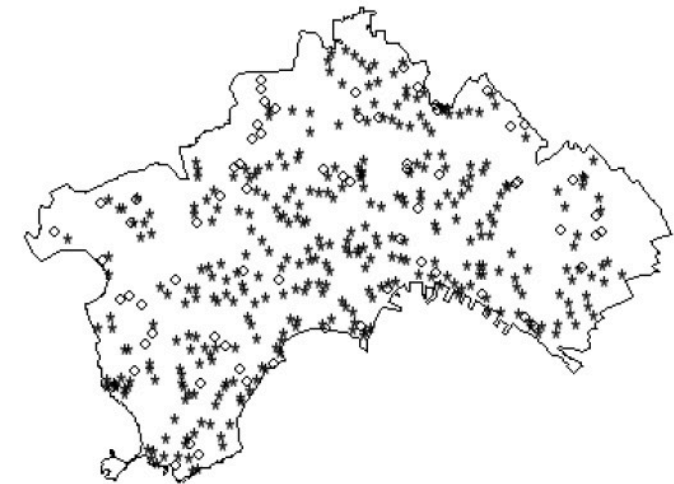
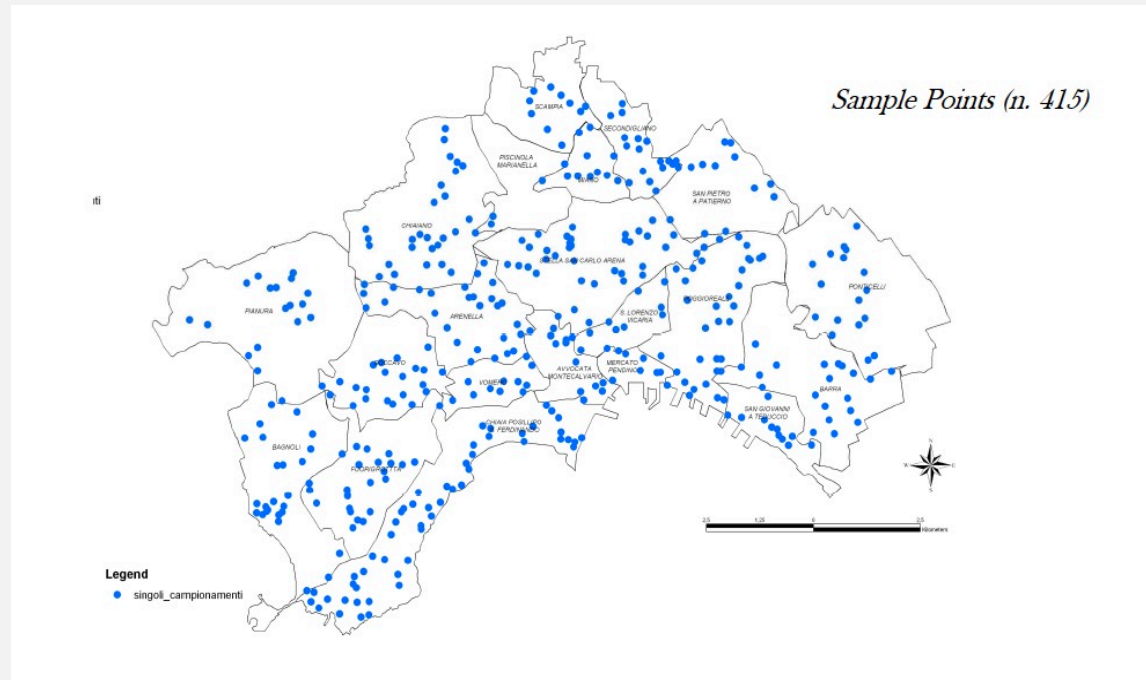
EXTENSIONES — PROCESOS PUNTUALES

- Los **procesos puntuales** se corresponden con variables aleatorias Bernouilli.
- En los procesos puntuales ***se conoce la localización exacta del caso y esta es aleatoria.***

EXTENSIONES – PROCESOS PUNTUALES

- En este caso, nos puede interesar:
 - Valorar si los sucesos siguen un patrón espacial determinado (agregación, forma regular, etc.)
 - Estudiar si un patrón observado se asocia con alguna variable (exposición a alguna variable ambiental, como la contaminación atmosférica; proximidad a focos contaminantes; contexto socioeconómico, etc.).

EXTENSIONES – PROCESOS PUNTUALES



Distribution of cases (positives: circle) and controls (negatives: stars). Naples, February - May 2005

EXTENSIONES — PROCESOS PUNTUALES — LOG-GAUSSIAN COX PROCESS (LGCP)

- El model LGCP es el análogo al modelo lineal gaussiano utilizado para datos geoestadísticos, cuando se modelizan datos en forma de procesos puntuales.
- Sin embargo, Diggle *et al.* (2013) proponen utilizar este modelo para aproximar el ajuste de datos espaciales de cualquier tipo (es decir, datos de área, geoestadísticos y procesos puntuales).

EXTENSIONES – PROCESOS PUNTUALES – LOG-GAUSSIAN COX PROCESS (LGCP)

Pero, los diferentes tipos de diseño espacial, ¿qué son en realidad? ¿procesos? ¿modelos? ¿métodos?

- El 2012, Diggle propone un cambio de paradigma y redefine la estadística espacial como ***‘un conjunto de modelos y métodos estadísticos que pretenden ayudar a los científicos a entender fenómenos espaciales, que no pueden ser observados directamente, sino de forma indirecta, con información incompleta’***, en forma de datos en rejilla, procesos puntuales y datos geoestadísticos.
- El model estadístico que propuso, como modelo único base, es el **modelo log Cox**.

EXTENSIONES — PROCESOS PUNTUALES — LOG-GAUSSIAN COX PROCESS (LGCP)

- Es decir, Diggle unificó la estadística espacial, tal y como McCullagh y Nelder (1989) unificaron los modelos lineales generalizados.

EXTENSIONES – PROCESOS PUNTUALES – LOG-GAUSSIAN COX PROCESS (LGCP)

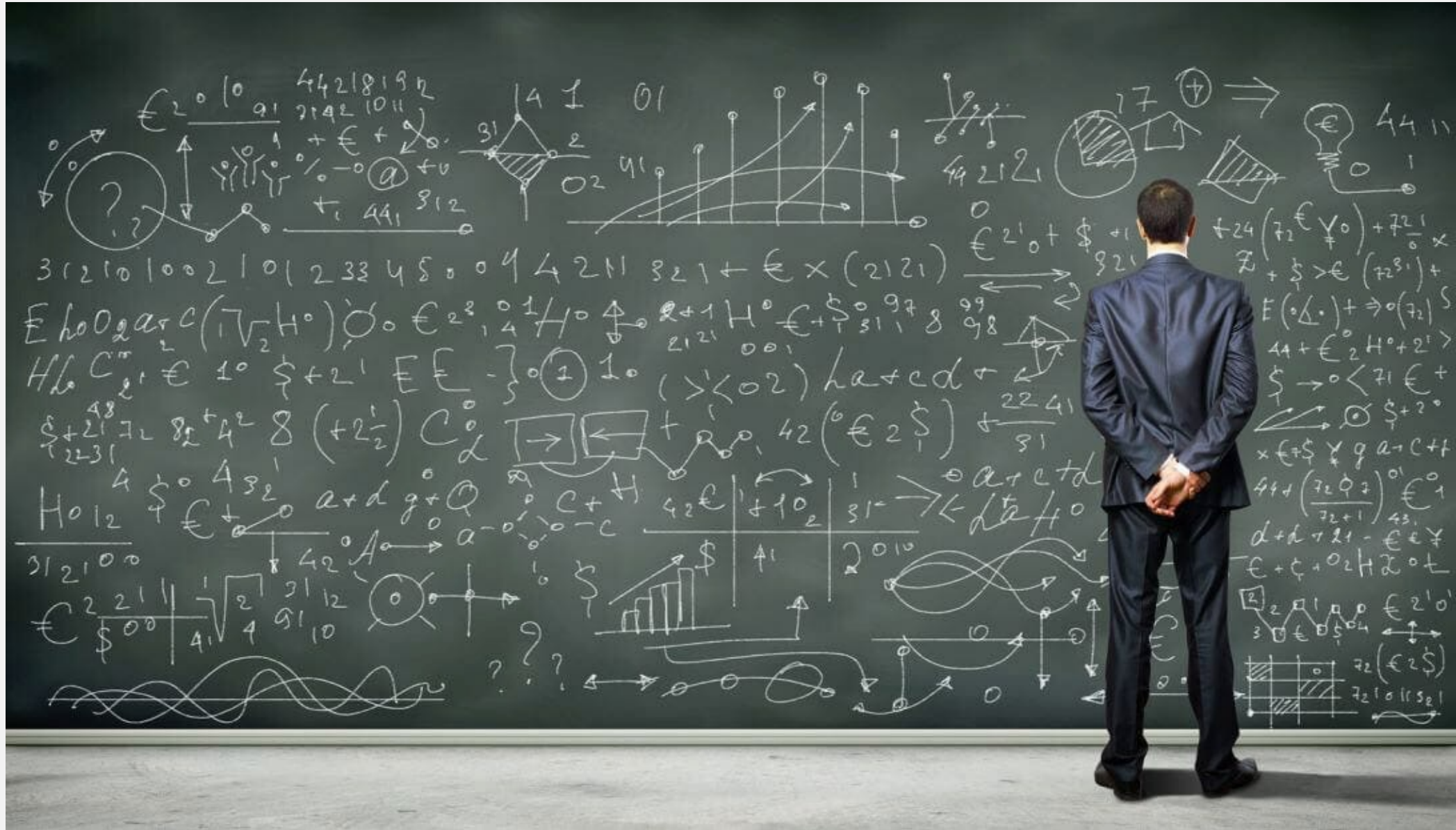
- Condicional al verdadero riesgo en la localización x_i , la probabilidad de que ocurra un caso, $P(x_i)$, $i = 1, \dots, n$, en esta localización se distribuye como una binomial.

$$Y_i | P(x_i) \sim \text{Binomial}(E_i, P(x_i))$$

- El modelo es (la función vínculo puede ser otra ...)

$$\log \mu = \beta_0 + \eta_{ji} + S(x_{ji}) + \tau_{ji} + \text{offset}(E_i)$$

$$\text{Cov}(S(x_i), S(x_{i'})) = \frac{\sigma^2}{2^{\nu-1} \Gamma(\nu)} (\kappa \|x_i - x_{i'}\|)^{\nu} K_{\nu}(\kappa \|x_i - x_{i'}\|)$$



I CAN... BUT I WON'T

