

□

ANÀlisi DE DADES ESPACIALS EN L'ÀMBIT DE L'EPIDEMIOLOGIA

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

8, 10, 14 i 16 de setembre de 2021

Grup de Recerca en Estadística, Econometria i Salut (GRECS), Universitat de Girona
CIBER d'Epidemiologia i Salut Pública(CIBERESP)

INTRODUCCIÓ AL CURS

1. Introducció al curs
2. Introducció a l'epidemiologia i l'estadística espacial
3. Panoràmica del models mixtos
4. Panoràmica del models mixtos - Pràctiques
5. Introducció a INLA i R INLA
6. R INLA - Pràctiques

Dimecres 8

Divendres 10

INTRODUCCIÓ AL CURS

- 7. Mapes de malalties. Estandardització de raons d'incidència i mortalitat
- 8. Mapes de malalties. Suavització de raons d'incidència i de mortalitat estandarditzades
- 9. Mapes de malalties – Pràctiques
- 10. Estudis d'associació geogràfica. Regressió ecològica espacial
- 11. Regressió ecològica espacial - Pràctiques

Dimarts 14

INTRODUCCIÓ AL CURS

- 12. Agrupació de casos
- 13. **Extensions: BYM2, processos puntuals, leaflet, pc priors**
- 14. Extensions – Pràctiques

} Dijous 16

EXTENSIONS

Veurem les següents extensions:

- Model BYM2
- Processos puntuals
- Leaflet (ho veurem en les pràctiques)
- pc priors (ho veurem en les pràctiques)

EXTENSIONS – EL MODEL BYM2

- En el model clàssic de Besag, York i Mollie, BYM (Besag *et al.*, 1991), la variació estructurada espacialment no és independent de la variació no estructurada (problema anomenat de la no identificabilitat).
- Com a conseqüència part de la dependència espacial (variació estructurada) podria ser realment heterogeneïtat (variació no estructurada) i viceversa.

EXTENSIONS – EL MODEL BYM2

- Hi ha formulacions alternatives al model de BYM, com els models de Leroux (Leroux *et al.*, 2000) i de Dean (Dean *et al.*, 2001), en què s'assegura que la variació espacial estructurada sigui independent de la no estructurada.
- No obstant això, cap dels dos models '*escala*' la variació espacial.
- Com a conseqüència, els hiperparàmetres depenen de l'estructura espacial del problema i no poden ser interpretats correctament.

EXTENSIONS – EL MODEL BYM2

- D'altra banda, en el context Bayesià l'elecció de les distribucions a priori dels hiperparàmetres (*priors*), pot tenir un impacte considerable en els resultats.
- En els models de Leroux i de Dean s'utilitzen priors estàndard que condueixen a un sobreajustament, *overfitting* (multicolinealitat).
- La principal conseqüència de l'*overfitting* és que els estimadors de les variàncies són més grans que els reals i, per tant, els intervals de credibilitat seran molt més amplis que l'esperat, el que implica que ens podria portar a no rebutjar la hipòtesi nul·la (que els coeficients siguin igual a zero) quan en realitat s'hauria de rebutjar.

EXTENSIONS – EL MODEL BYM2

- Simpson *et al.* (2017) van proposar una modificació del model BYM, anomenada BYM2, que soluciona aquests problemes, ja que escala la variació estructurada espacialment i utilitza priors que penalitzen la complexitat (anomenats PC priors).
- Aquests priors són robustos, en el sentit que no tenen impacte en els resultats i, a més, tenen una interpretació epidemiològica.

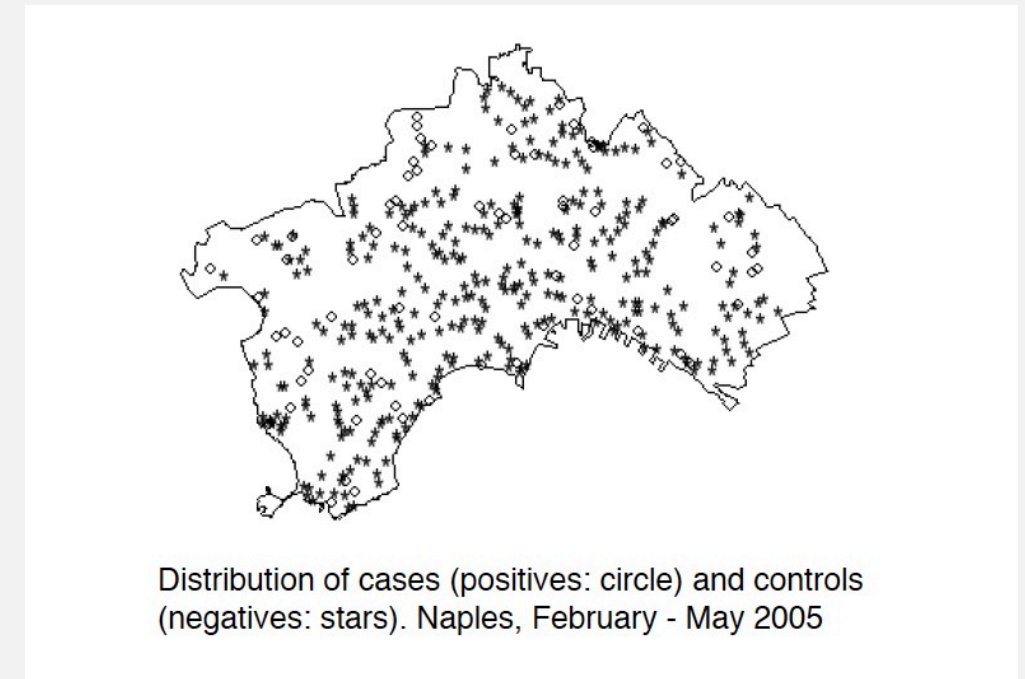
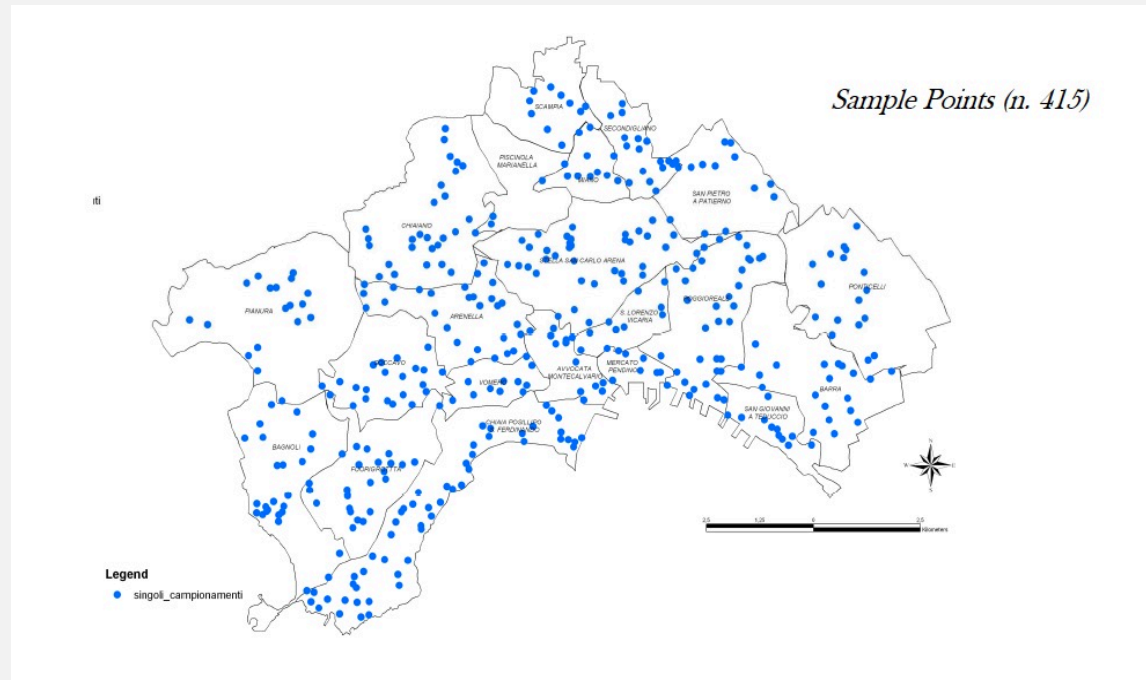
EXTENSIONS – PROCESSOS PUNTUALS

- Els **processos puntuals** es corresponen amb variables aleatòries Bernouilli.
- En els processos puntuals ***es coneix la localització exacta del cas i aquesta és aleatòria.***

EXTENSIONS – PROCESSOS PUNTUALS

- En aquest cas, ens pot interessar:
 - Valorar si els esdeveniments segueixen un patró espacial determinat (agregació, forma regular, etc.)
 - Estudiar si un patró observat s'associa amb alguna variable (exposició a alguna variable ambiental, com la contaminació atmosfèrica; proximitat a focus contaminants; context socioeconòmic, etc.).

EXTENSIONS – PROCESSOS PUNTUALS



EXTENSIONS – PROCESSOS PUNTUALS – LOG-GAUSSIAN COX PROCESS (LGCP)

- El model LGCP és l'anàleg del model lineal gaussià utilitzat per a dades geoestadístics, quan es modelitzen dades en forma de processos puntuals.
- No obstant això, Diggle *et al.* (2013) proposen utilitzar aquest model per aproximar l'ajustament de dades espacials de qualsevol tipus (és a dir, dades d'àrea, geoestadístics i processos puntuals).

EXTENSIONS – PROCESSOS PUNTUALS – LOG-GAUSSIAN COX PROCESS (LGCP)

Però, els diferents tipus de disseny espacial, què son en realitat? processos? models? mètodes?

- El 2012, Diggle proposa un canvi de paradigma i re-defineix la estadística espacial com ***'un conjunt de models i mètodes estadístics que pretenen ajudar els científics a entendre fenòmens espacials, que no poden ser observats directament, sinó de forma indirecta, amb informació incompleta'***, en forma de dades en reixeta, processos puntuals i dades geoestadístics.
- El model estadístic que va proposar, com a model únic base, és el **model log Cox**.

EXTENSIONS — PROCESSOS PUNTUALS — LOG-GAUSSIAN COX PROCESS (LGCP)

- És a dir, Diggle va unificar l'estadística espacial, tal com McCullagh i Nelder (1989) van unificar els models lineals generalitzats.

EXTENSIONS – PROCESSOS PUNTUALS – LOG-GAUSSIAN COX PROCESS (LGCP)

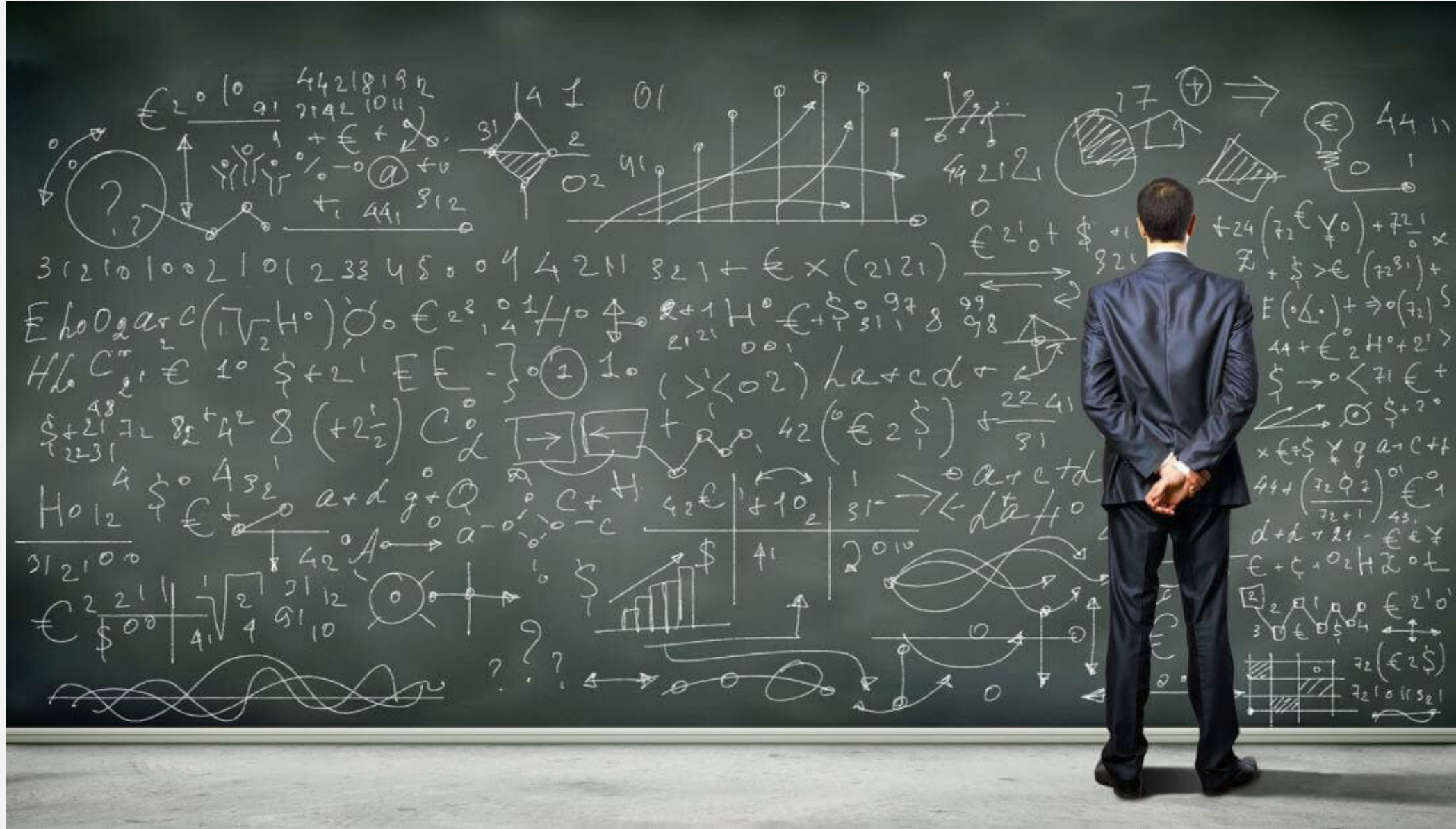
- Condicional al vertader risc en la localització x_i , la probabilitat que ocorri un cas, $P(x_i)$, $i = 1, \dots, n$, en aquesta localització es distribueix com una binomial.

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

- El model és (el vincle pot ser un altra...)

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

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



I CAN... BUT I WON'T

