

# **Modelització espacial del COVID-19 a Catalunya per comarques**

**Una aproximació Bayesiana**

**Robert Torrell  
Nabil El Bachiri**

# Objectius del treball

- **Modelitzar les defuncions per COVID-19 (serie temporal) a Catalunya per comarques (anàlisi espacial).**
- **Variable a predir:** Defuncions per COVID-19 cada 1000 habitants.
- **Posibles covariables:** Edat, Educació, Densitat de població, renda garantida, atur, índex de vacunació(?), mesures de contenció(?), alguna altre més que podria ser interessant (?)

# Preguntes d'investigació

- Quines son les comarques on hi ha hagut més incidència de COVID-19? Com s'expandeix una epidemia en la nostra societat? Quins factors aturen o incrementen els seus efectes?
- Han sigut les comarques més pobres més o menys propenses a tenir morts per COVID-19? I si parléssim d'educació?
- Han sigut les comarques amb més atur més o menys propenses a tenir morts per COVID-19?
- Fins a quin punt ha condicionat l'edat la probabilitat de mort per COVID-19? I les enfermetats com l'obesitat, la diabetis, o el colesterol...? Per comarques

# Metodología

**Opció 1 - Model de regressió Bayesià:** Simple, pero no pot capturar la variabilitat entre comarques.

**Opció 2 - Model jeràrquic Bayesià:** Més complexe però captura la variabilitat de les covariables entre comarques. Millor inferència. Distribucions a posteriori per a tots els paràmetres.

# Fonts de dades

COVID-19 per comarques a Catalunya: <https://analisi.transparenciacatalunya.cat>

Educació, densitat, atur (...) per comarques: <https://www.idescat.cat/>

Renda garantida per comarques: <https://dretssocials.gencat.cat>

INE - Datos de movilidad en estado de alerta:  
[https://www.ine.es/covid/covid\\_movilidad.htm](https://www.ine.es/covid/covid_movilidad.htm)

# Exploració de les dades (I)

Tenim moltes dades:

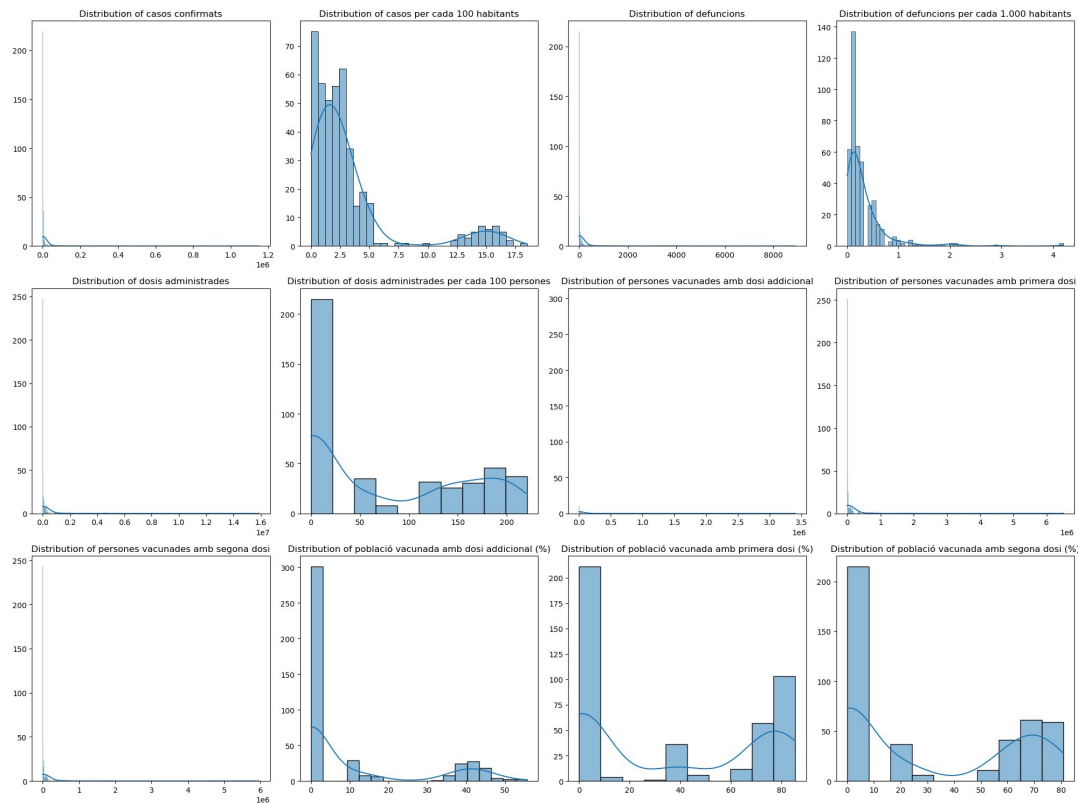
- Per temps (Semanal, mensual, trimestral, anual)
- Per ubicació (Comunitat autònoma, províncies, comarques, municipis)
- COVID-19: casos, propagació, danys
- Vacunació de la població
- Capacitat hospitalària
- Mobilitat de la població en estat d'alerta
- Índexs econòmics
- Índexs socioeconòmics

## Exploració de les dades (II)

Només combinant casos COVID-19 + defuncions + vacunació per trimestre...

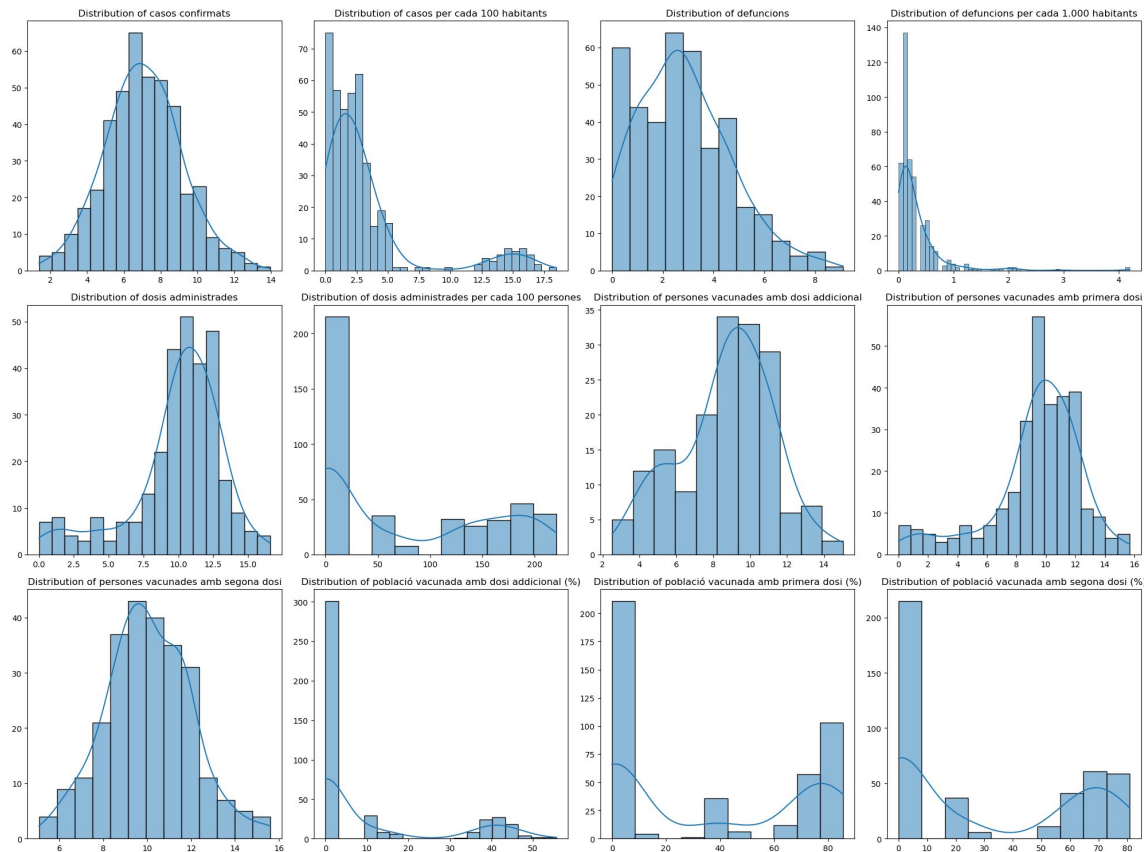
comarques i Aran	trimestre	casos confirmats	casos per cada 100 habitants	defuncions	defuncions per cada 1.000 habitants	dosis administrades	dosis administrades per cada 100 persones	persones vacunades amb dosi addicional	persones vacunades amb primera dosi	persones vacunades amb segona dosi	població vacunada amb dosi addicional (%)	població vacunada amb primera dosi (%)	població vacunada amb segona dosi (%)
Alt Camp	2020/T1	75.0	0.2	10.0	0.2	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
Alt Camp	2020/T2	93.0	0.2	21.0	0.5	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
Alt Camp	2020/T3	277.0	0.6	2.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
Alt Camp	2020/T4	932.0	2.1	14.0	0.3	41.0	0.10	0.0	41.0	0.0	0.0	0.1	0.0
Alt Camp	2021/T1	838.0	1.9	22.0	0.5	4136.0	9.23	0.0	2921.0	1215.0	0.0	6.5	2.7

# Exploració de les dades (III)

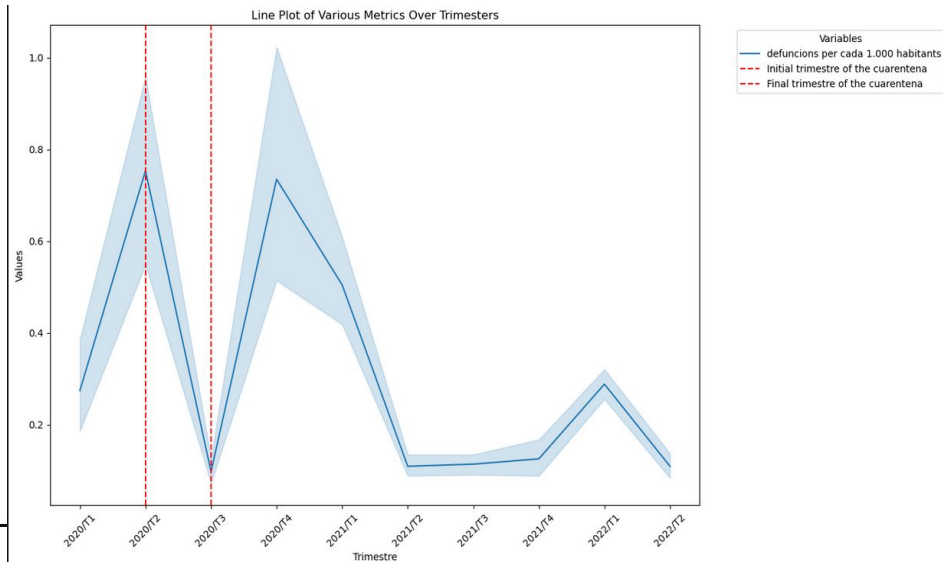
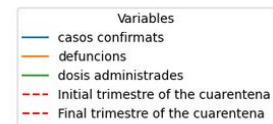
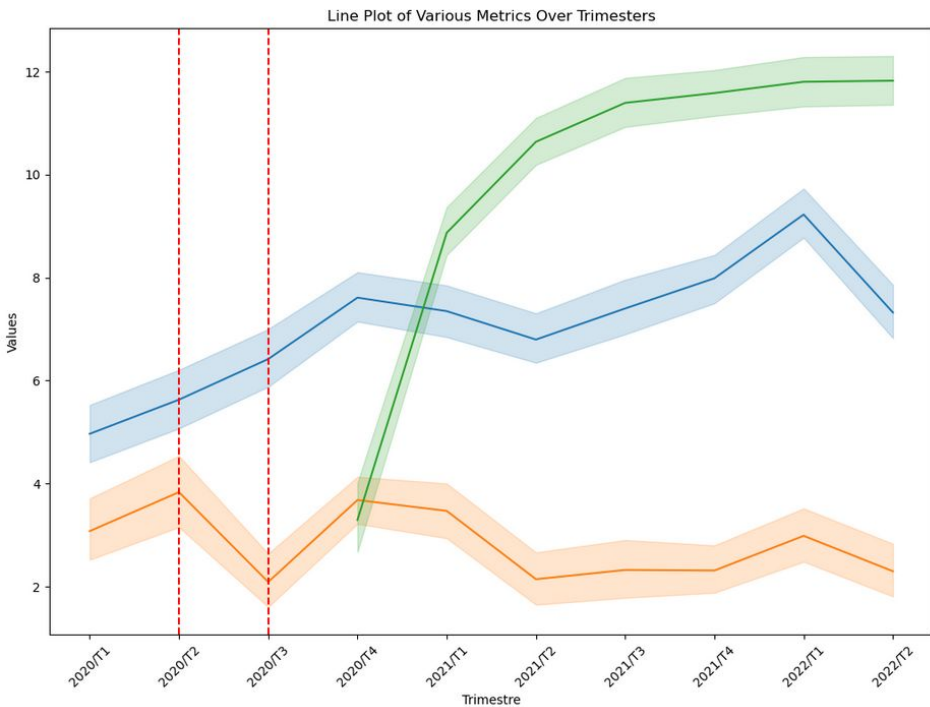




# Exploració de les dades (IV)



# Exploració de les dades (V)



# Model lineal (I)

$$y_n \sim \text{normal}(\alpha + \beta X_n, \sigma).$$

$y_i$  es el nombre de defuncions per 1000 habitants.

Els coeficients son:

- Casos per cada 100 habitants (%)
- Vacunes administrades per cada 100 persones (%)
  - De nomes la primera dosi
  - De la segona dosi
  - De més dosis
- El trimestre de l'observació

Avantatge:

- Es senzill d'implementar e interpretar

Inconvenients:

- Assumeix linealitat (Temps!)
- Coeficients per a tot el període
- No té en compte l'espai
- Només descriu defuncions per tot Catalunya
- Mal model per dades zero inflades (vacunes)

# Model lineal (II)

$$y \sim \text{Normal}(x \cdot \beta + \alpha, \sigma)$$

$$\alpha \sim \text{Gamma}(40, 12)$$

$$\beta \sim \text{Normal}(0, 10)$$

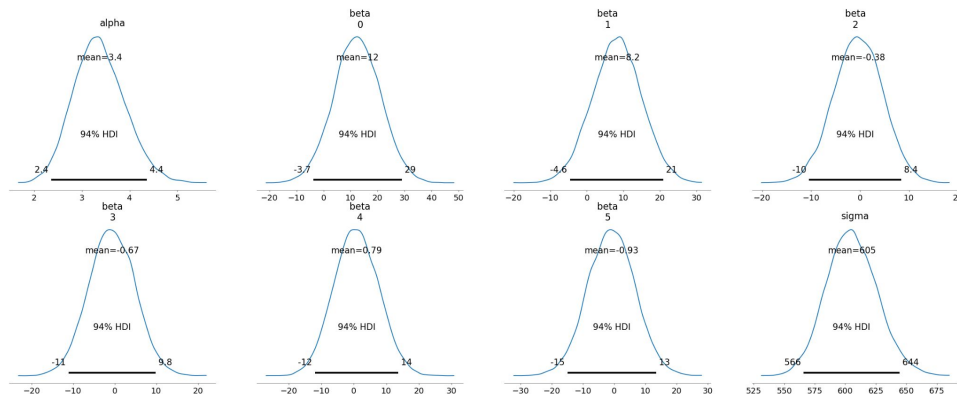
$$\sigma \sim \text{Normal}(0, 1000)$$

MCMC:

3 cadenes, 5000 samples

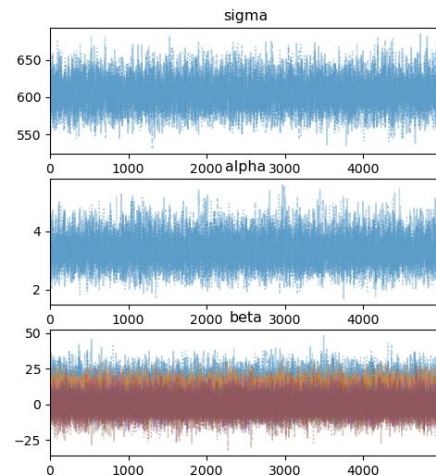
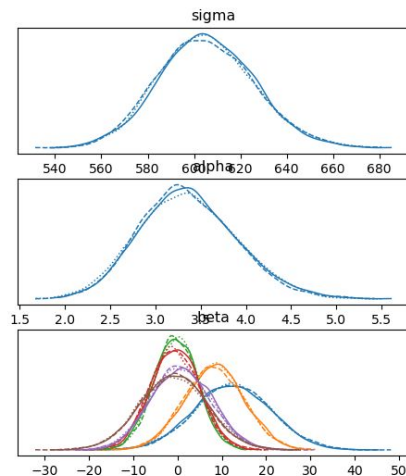
	count	mean	std	min	25%	50%	75%	max
parameters								
lp_	15000.0	-2958.775273	2.020194	-2979.372841	-2959.897806	-2958.434974	-2957.298510	-2954.970575
accept_stat_	15000.0	0.940974	0.082816	0.314294	0.919323	0.976133	0.996385	1.000000
stepsize_	15000.0	0.039460	0.002699	0.036964	0.036964	0.038207	0.043208	0.043208
treedepth_	15000.0	6.370533	0.611062	2.000000	6.000000	6.000000	7.000000	7.000000
n_leapfrog_	15000.0	102.541600	32.493757	3.000000	63.000000	127.000000	127.000000	255.000000
divergent_	15000.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
energy_	15000.0	2962.770070	2.837859	2955.905986	2960.722509	2962.449701	2964.450850	2982.531385
alpha	15000.0	3.362623	0.537385	1.669531	2.988201	3.338651	3.711301	5.606036
beta.1	15000.0	12.460398	8.666171	-21.291341	6.654698	12.422940	18.360196	48.286120
beta.2	15000.0	8.212883	6.768184	-20.022732	3.676272	8.270088	12.780005	31.385372
beta.3	15000.0	-0.383622	5.003980	-20.170311	-3.761994	-0.390719	3.016713	18.230594
beta.4	15000.0	-0.672809	5.567965	-23.191714	-4.415623	-0.710464	3.145025	22.034634
beta.5	15000.0	0.794877	6.840283	-26.985717	-3.876458	0.730201	5.478496	30.788306
beta.6	15000.0	-0.925771	7.589027	-32.239238	-6.090909	-0.921879	4.187625	28.063942
sigma	15000.0	605.423727	20.841281	531.270765	591.083271	604.788351	619.074979	685.215622

# Model lineal (III)



Massa poc informatiu...

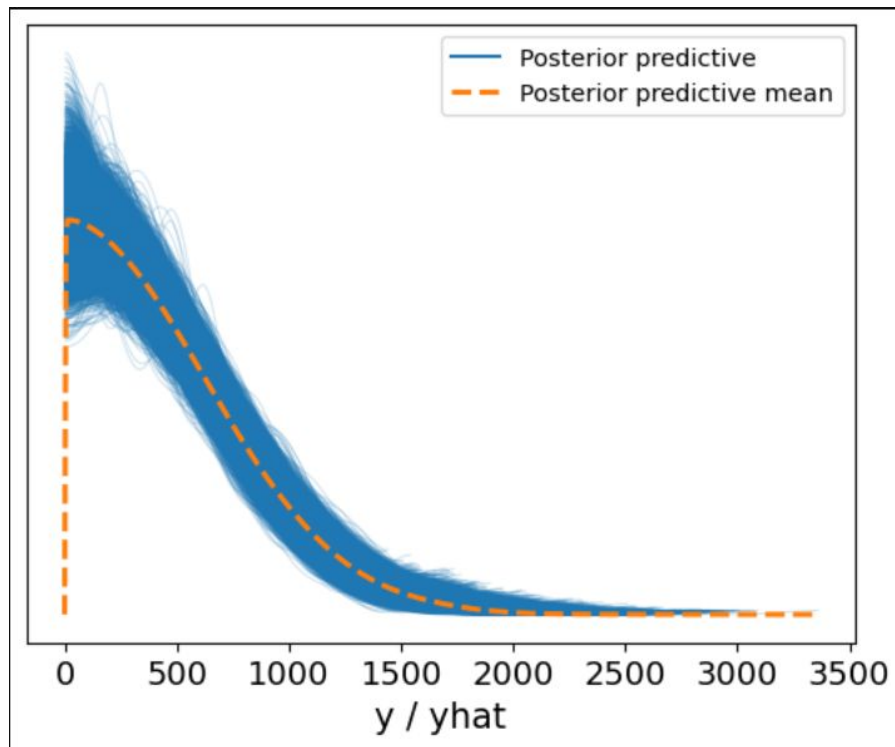
Pero el model convergeix molt be!



## Model lineal (IV)

Posteriori predictiva no informativa:

- Centrat en 0...
- ... Pero massa variança!
- Considera bastant possibles uns valors massa alts.



## Model jeràrquic (I)

$$Y_i \sim \text{Poisson}(\lambda_i)$$

$\lambda_i$  és la taxa de defuncions a la comarca  $i$ .

### **Logaritme de la Mitjana:**

El logaritme de  $\lambda_i$  es modela com una combinació lineal de covariables i un efecte aleatori espacial:

$$\log(\lambda_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i$$

## Model jeràrquic (II)

Els efectes aleatoris  $u_i$  es modelitzen utilitzant un model condicionalment autoregressiu (CAR) o un model d'efectes aleatoris espacials suaus:

$$\mathbf{u} \sim \text{MVN}(\mathbf{0}, \tau_u^{-1} \mathbf{R})$$

$\mathbf{u}$  és el vector d'efectes aleatoris espacials per a totes les comarques.

$\mathbf{R}$  és una matriu que defineix l'estructura de dependència espacial entre les comarques, basada en la contigüïtat o la distància entre elles.

$\tau_u$  és el paràmetre de precisió dels efectes aleatoris, que controla la variabilitat espacial.



## Model jeràrquic (III)

**Distribucions a priori per als paràmetres del model:**

$$\beta_j \sim \text{Normal}(0, \sigma_\beta^2)$$

**Distribució a Priori per a la Precisió dels Efectes Aleatoris:**

$$\tau_u \sim \text{Gamma}(a, b)$$

# Preguntes

- Variable a predir: Contagis? Defuncions?
  - Problema de la variable contagis: Esbiaixos, covariates diferents?
- Covariates a inclur: Hi ha alguna que sigui interessant de posar?
- Escala: Seria millor fer-ho en un altre escala (eg. per barris a Barcelona). Si es així, on trobar les dades?
  
- Com modelitzar adientment l'evolució del COVID en l'espai? Quins models espacials alternatius proposaríeu?
- Com es pot capturar tant informació espacial com temporal de les dades?  
Com combinar i normalitzar tantes dades tan heterogènees?