

□

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

PANORÀMICA DELS MODELS MIXTOS

1. Model de regressió lineal
2. Model de regressió logística
3. Model lineal generalitzat (GLM)
4. Models mixtos

PANORÀMICA DELS MODELS MIXTOS

1. **Model de regressió lineal**
2. Model de regressió logística
3. Model lineal generalitzat (GLM)
4. Models mixtos

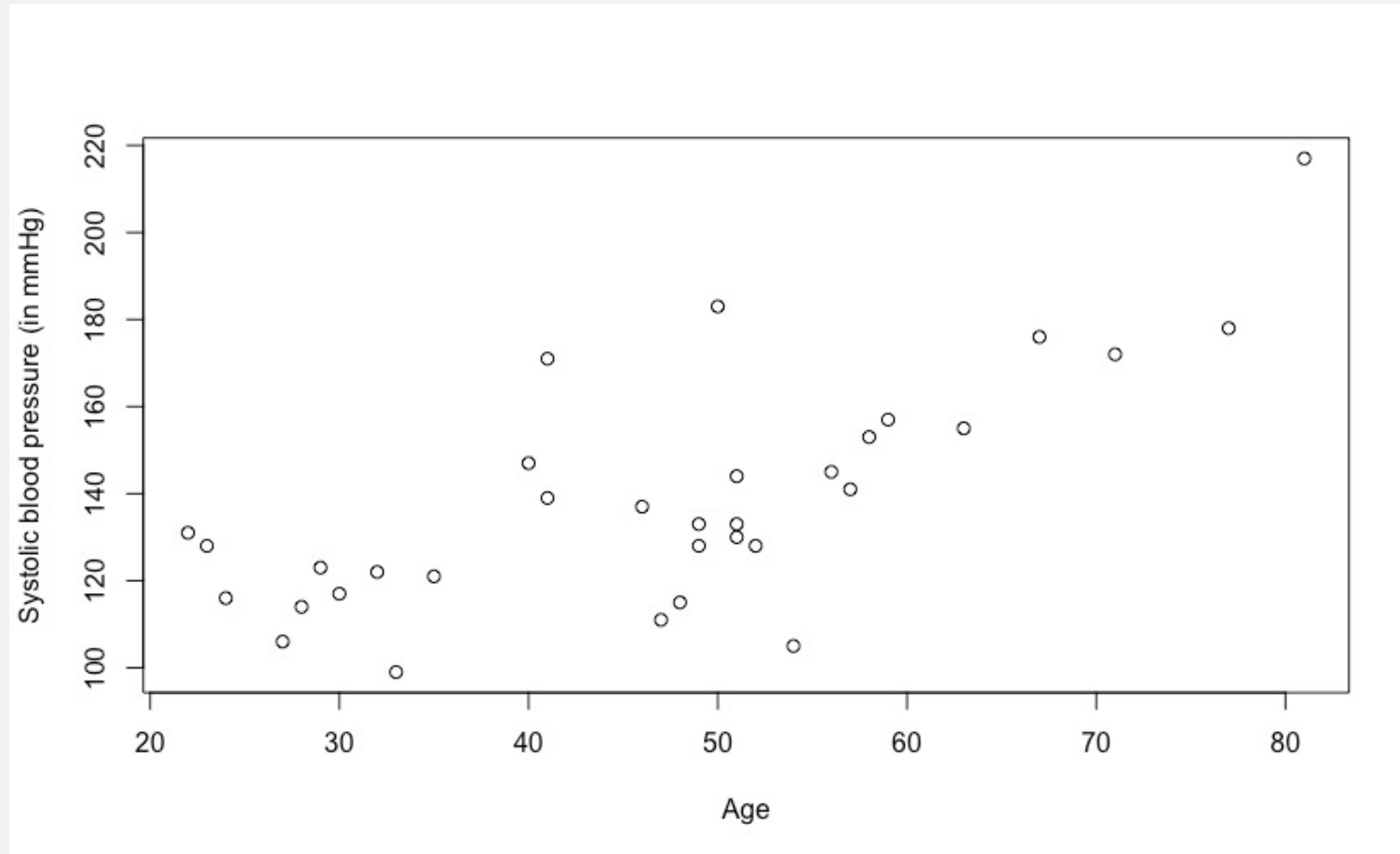
MODEL DE REGRESSIÓ LINEAL

Table 1 Age and systolic blood pressure (SBP) among 33 adult women

Age	SBP	Age	SBP	Age	SBP
22	131	41	139	52	128
23	128	41	171	54	105
24	116	46	137	56	145
27	106	47	111	57	141
28	114	48	115	58	153
29	123	49	133	59	157
30	117	49	128	63	155
32	122	50	183	67	176
33	99	51	130	71	172
35	121	51	133	77	178
40	147	51	144	81	217

adapted from Colton T. Statistics in Medicine. Boston: Little Brown, 1974

MODEL DE REGRESSIÓ LINEAL

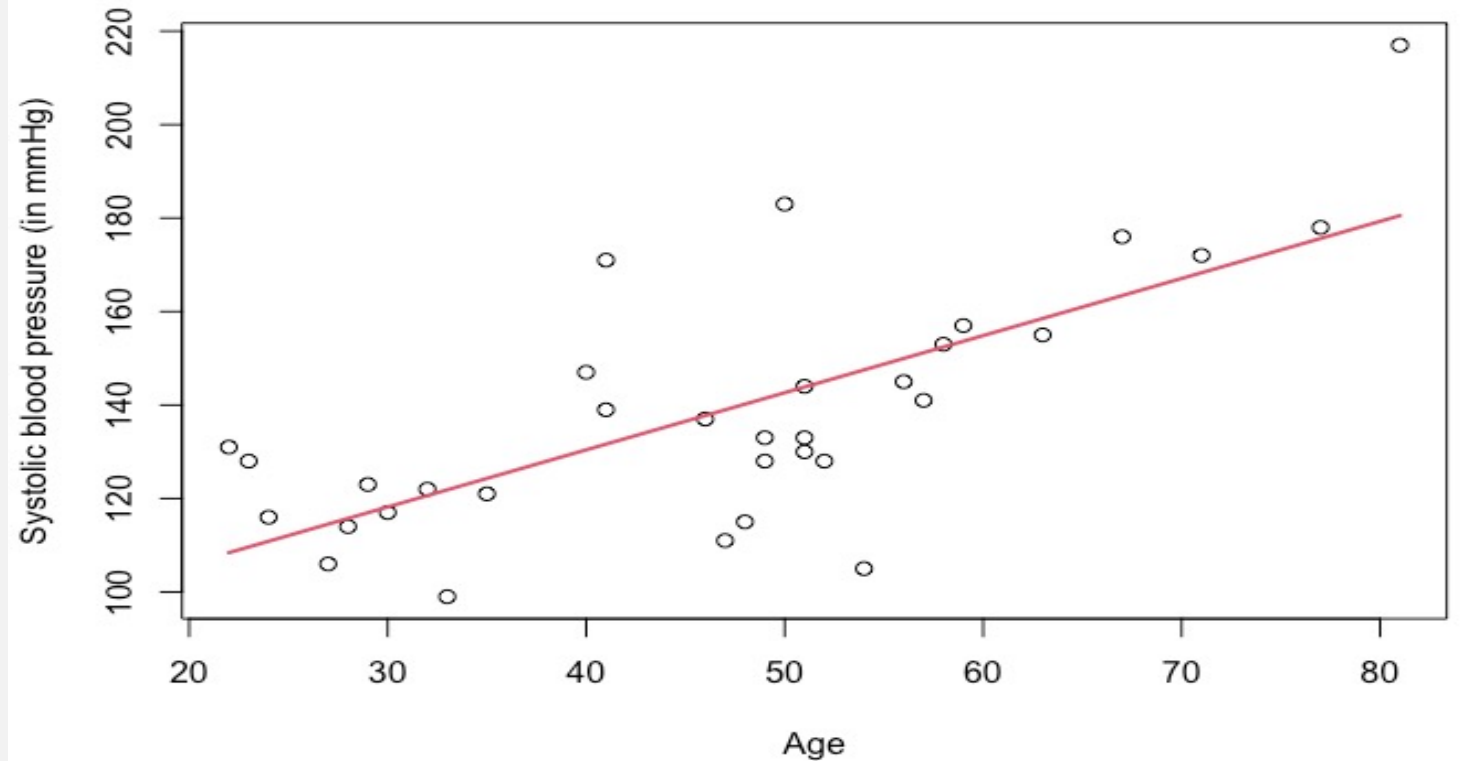


MODEL DE REGRESSIÓ LINEAL

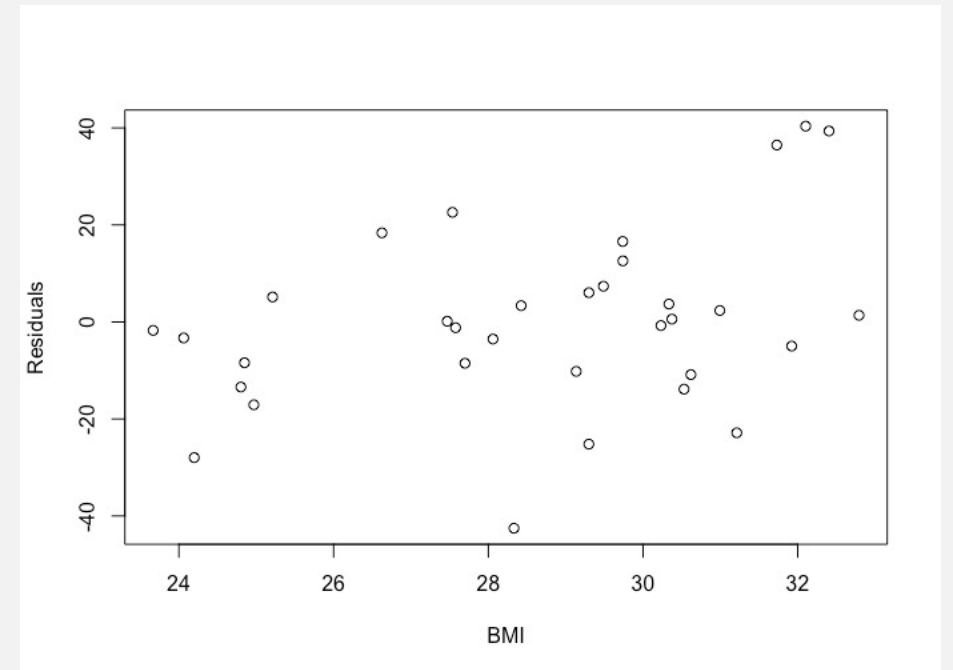
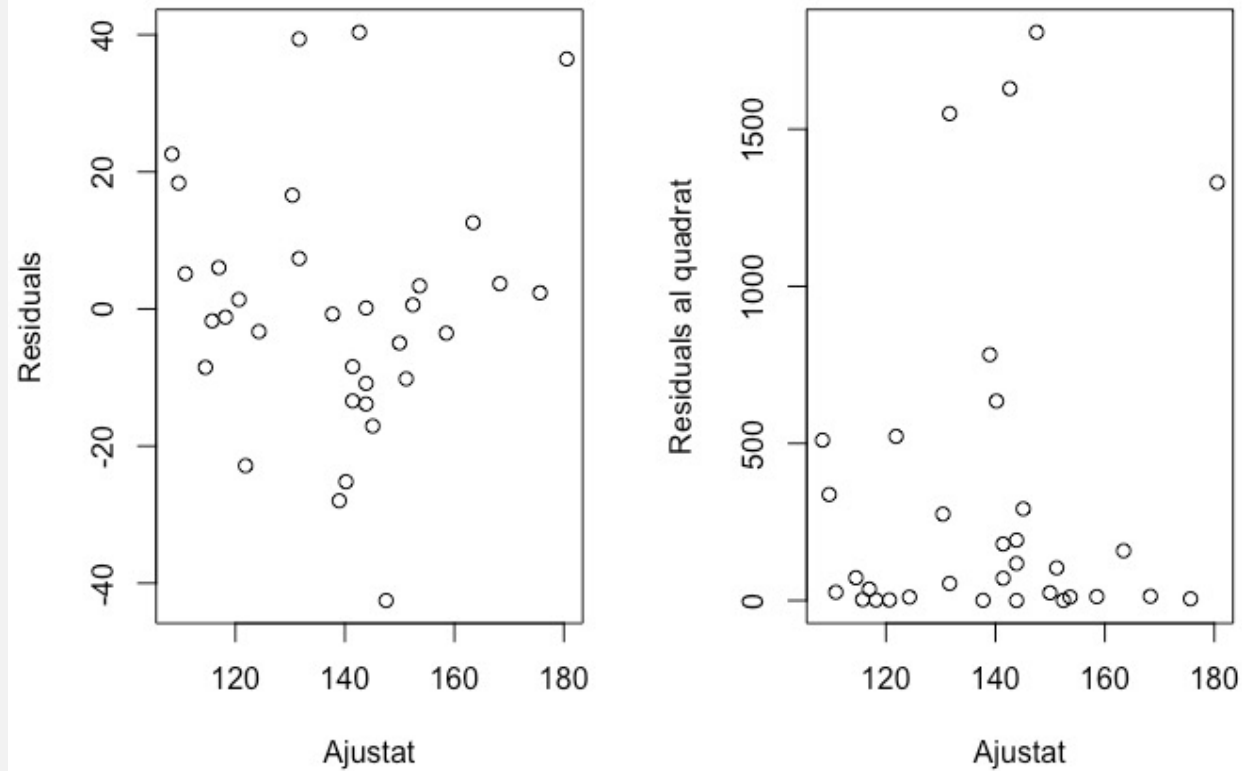
$$SBP_i = \beta_0 + \beta_1 age_i + u_i$$

$$\widehat{SBP}_i = \hat{\beta}_0 + \hat{\beta}_1 age_i$$

$$\widehat{SBP}_i = 81,5161 + 1,224 age_i$$

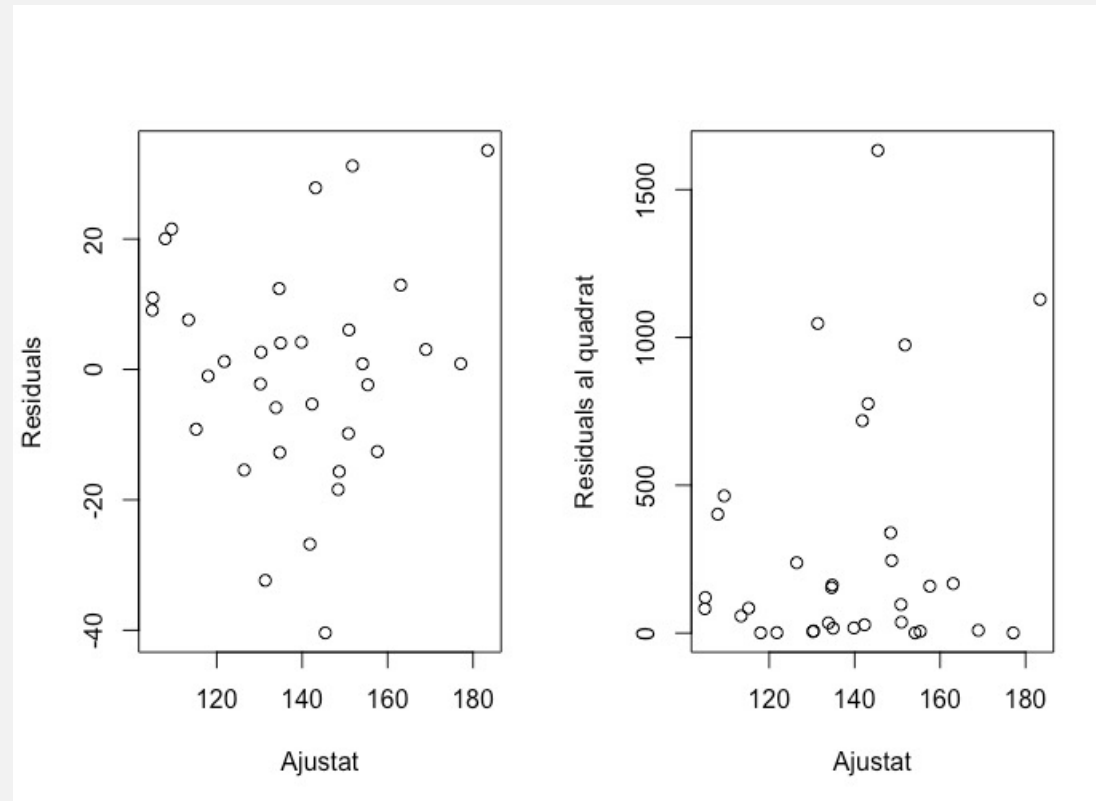


MODEL DE REGRESSIÓ LINEAL



MODEL DE REGRESSIÓ LINEAL

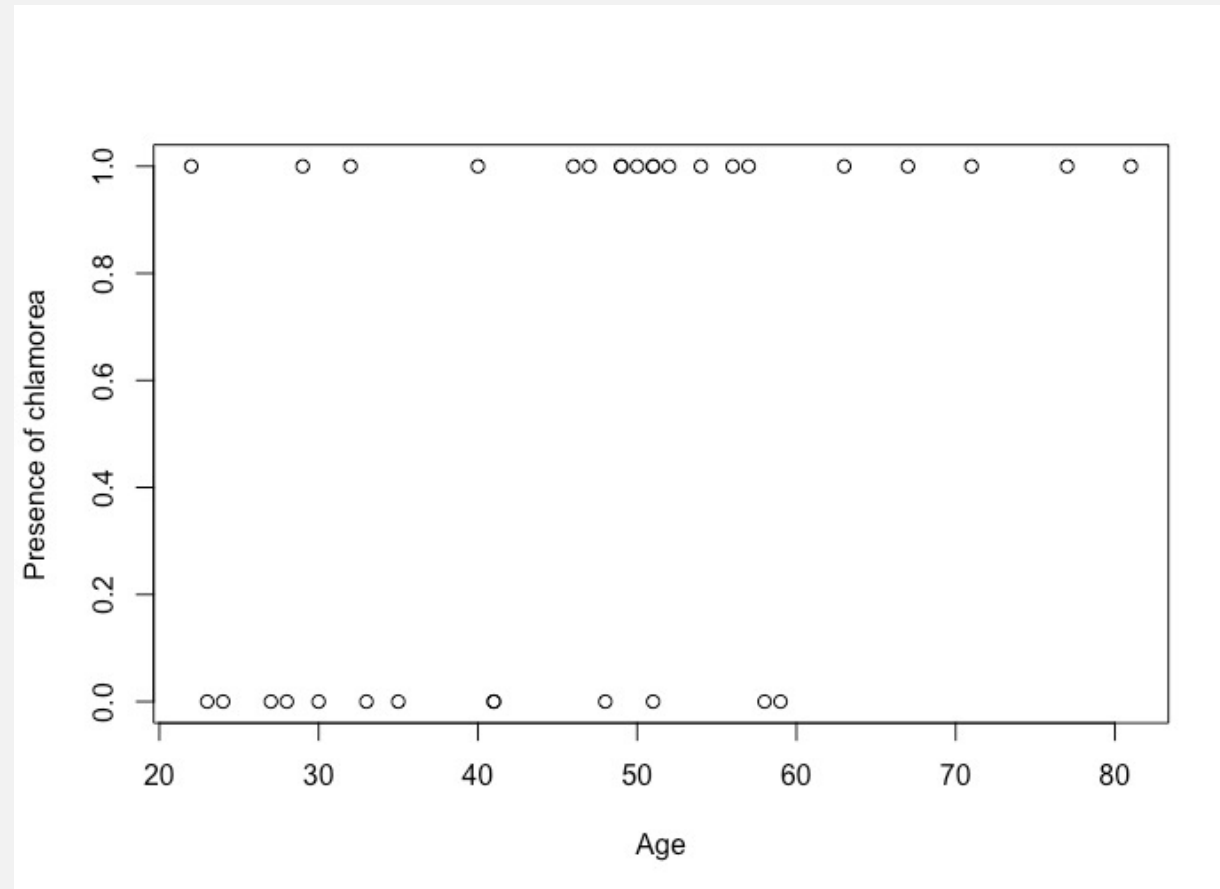
$$\widehat{SBP}_i = 8,9631 + 1,0538 \text{ age}_i + 2,8075 \text{ BMI}_i$$



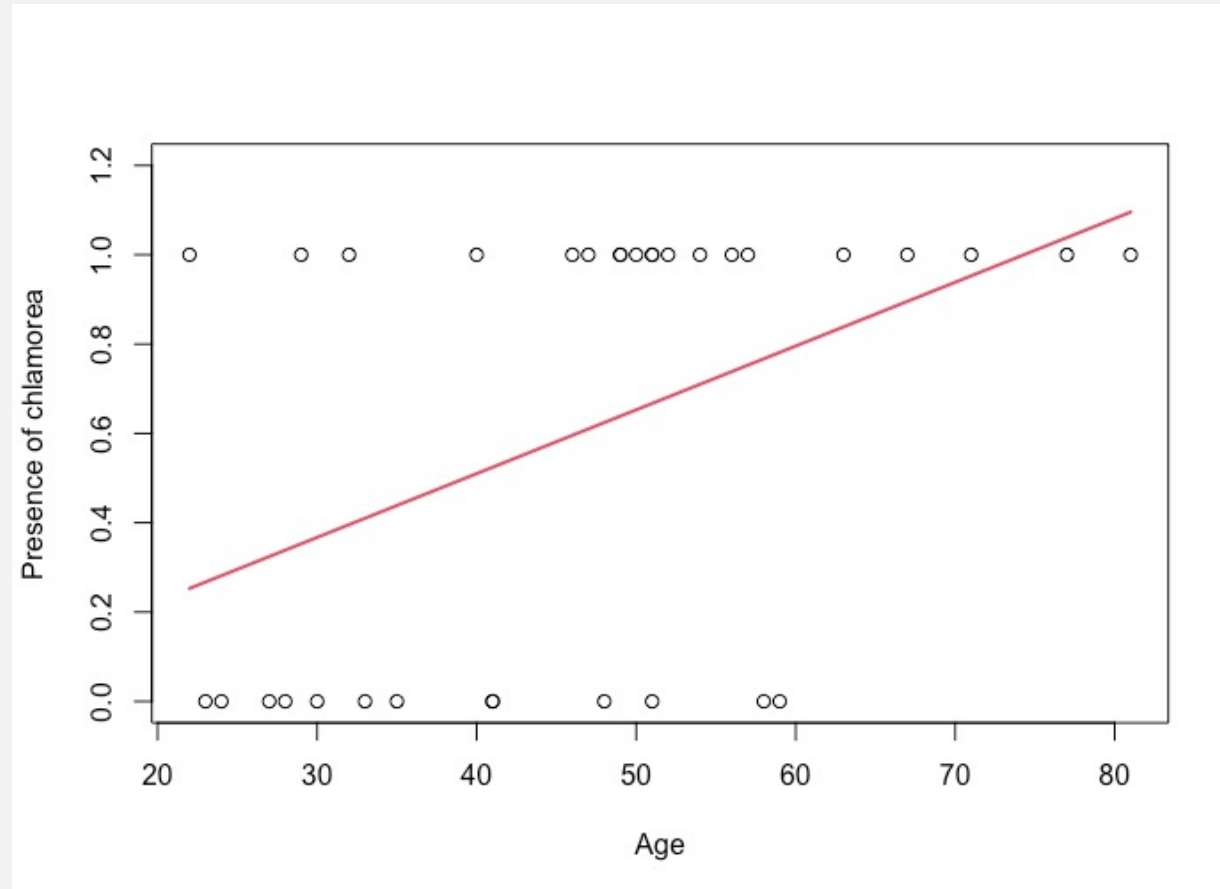
PANORÀMICA DELS MODELS MIXTOS

1. Model de regressió lineal
2. **Model de regressió logística**
3. Model lineal generalitzat (GLM)
4. Models mixtos

MODEL DE REGRESSIÓ LOGÍSTICA



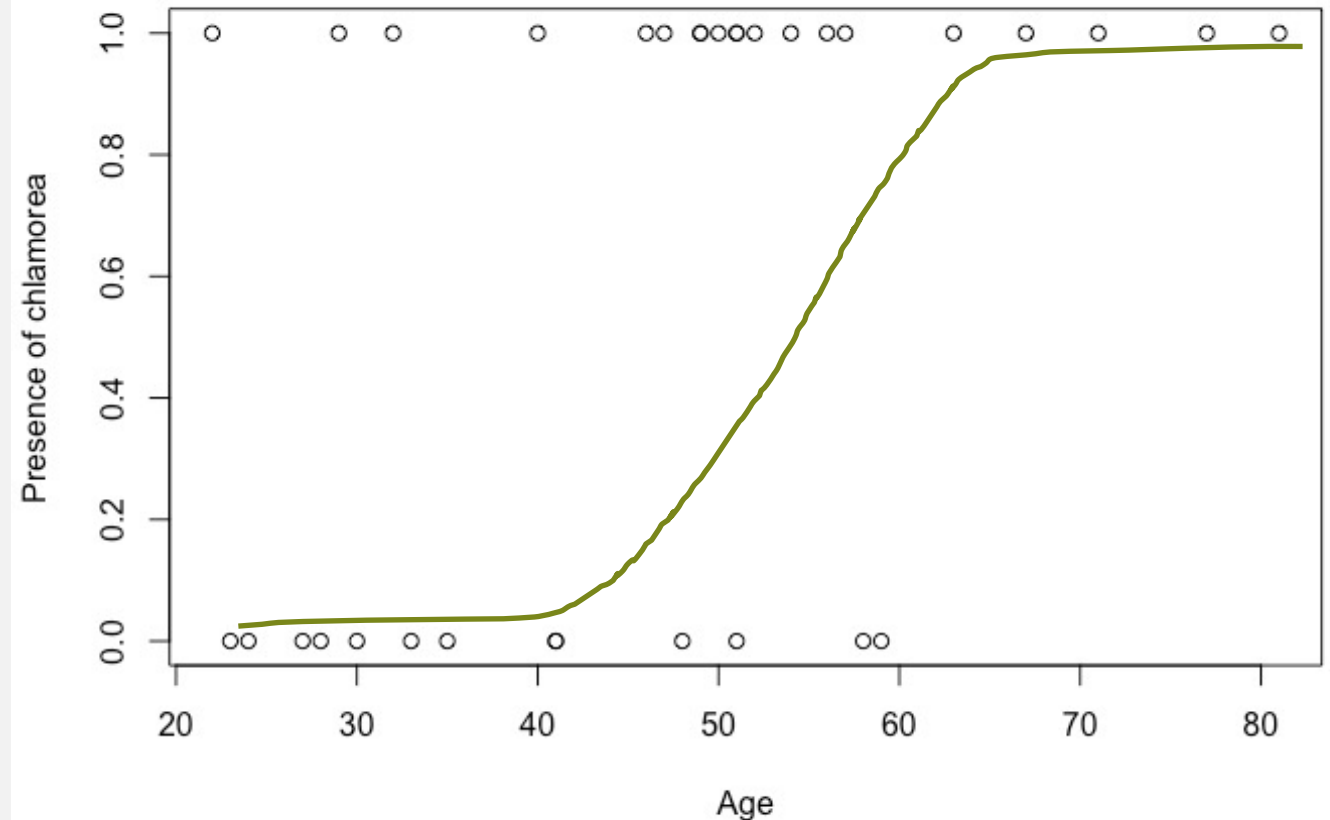
MODEL DE REGRESSIÓ LOGÍSTICA



MODEL DE REGRESSIÓ LOGÍSTICA

$$Prob(chlamorea = 1) = \frac{\exp^{\beta_0 + \beta_1 age_i}}{1 + \exp^{\beta_0 + \beta_1 age_i}}$$

$$\ln\left(\frac{Prob(chlamorea_i = 1)}{1 - Prob(chlamorea_i = 1)}\right) = \beta_0 + \beta_1 age_i$$



PANORÀMICA DELS MODELS MIXTOS

1. Model de regressió lineal
2. Model de regressió logística
- 3. Model lineal generalitzat (GLM)**
4. Models mixtos

MODEL LINEAL GENERALITZAT (GLM)

- Tal i com el seu nom indica, el **model lineal generalitzat, GLM** (McCullagh i Nelder, 1989), generalitza el model de regressió lineal per variables dependents no normals, com per exemple, variables comptadores (distribució de Poisson) o variables dicotòmiques (distribució binomial), permetent així utilitzar el mateix tipus de modelització, especificació, estimació i diagnòstic.
- El model lineal generalitzat és una extensió del model lineal quan les observacions són independents, però les suposicions sobre la normalitat de les pertorbacions no es compleixen.

MODEL LINEAL GENERALITZAT (GLM)

- La distribució de les pertorbacions es pot escollir dins del conjunt de distribucions que pertanyen a la família de **distribucions exponencial** d'un paràmetre que inclou les distribucions Binomial, Poisson, Binomial Negativa, Gamma, Beta, Inversa Gaussiana i dins del qual la distribució Normal és un cas especial.

MODEL LINEAL GENERALITZAT (GLM)

- L'objectiu dels GLM és el de descriure la dependència de la resposta (o variable dependent) y , respecte de les variables explicatives, $E(y | x) = \mu$.
- Els GLM es componen de dues funcions: la funció 'vinçle' (link), $g()$, i la funció de 'variància', $v()$. De vegades també hi ha un paràmetre d'escala, ϕ .

MODEL LINEAL GENERALITZAT (GLM)

- Així, per exemple, quan la **variable dependent és contínua**, el GLM s'especificaria com (equivalent a una regressió lineal):

$$\mu = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}$$

$$Var(y | x) = \phi$$

El vincle és lineal i la variància constant.

MODEL LINEAL GENERALITZAT (GLM)

- Quan la **variable dependent és dicotòmica**, el GLM (equivalent a una regressió logística), s'especificaria:

$$\ln\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}$$

$$Var(y|x) = \phi\mu(1 - \mu)$$

En aquest cas, el vincle és un lògit. El paràmetre ϕ es denomina de sobredispersió (si $\phi > 1$) o de subdispersió (si $\phi < 1$).

MODEL LINEAL GENERALITZAT (GLM)

- Quan la **variable dependent és discreta (o comptadora)**, el GLM (equivalent a una regressió de Poisson), s'especificaria:

$$\ln(\mu) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}$$

$$\text{Var}(y|x) = \phi\mu$$

El vincle és logarítmic. El paràmetre ϕ es denomina de sobredispersió (si $\phi > 1$) o de subdispersió (si $\phi < 1$).

PANORÀMICA DELS MODELS MIXTOS

1. Model de regressió lineal
2. Model de regressió logística
3. Model lineal generalitzat (GLM)
4. **Models mixtos**

MODELS MIXTOS

Els **dissenys mixtos** es caracteritzen per considerar simultàniament dues o més dimensions d'anàlisi.

Els **dissenys mixtos** comprenen:

- els **dissenys multinivell** o de nivells múltiples (també anomenats jeràrquics)
- els **dissenys longitudinals** o de mesures repetides.

MODELS MIXTOS – DISSENYS MULTINIVELL

- Tenen una estructura jeràrquica, amb agrupacions de dades en grups o clústers.
- En la literatura, aquests nivells jeràrquics es denominen nivell 1, nivell 2, etc.; estadi 1, estadi 2, etc.; nivell individual i nivell poblacional; nivell individual i nivell clúster.
- Per exemple, quan tenim països i regions, tenim dos nivells jeràrquics.

MODELS MIXTOS – DISSENYS MULTINIVELL

Exemple: Pacients dins de centres de salut

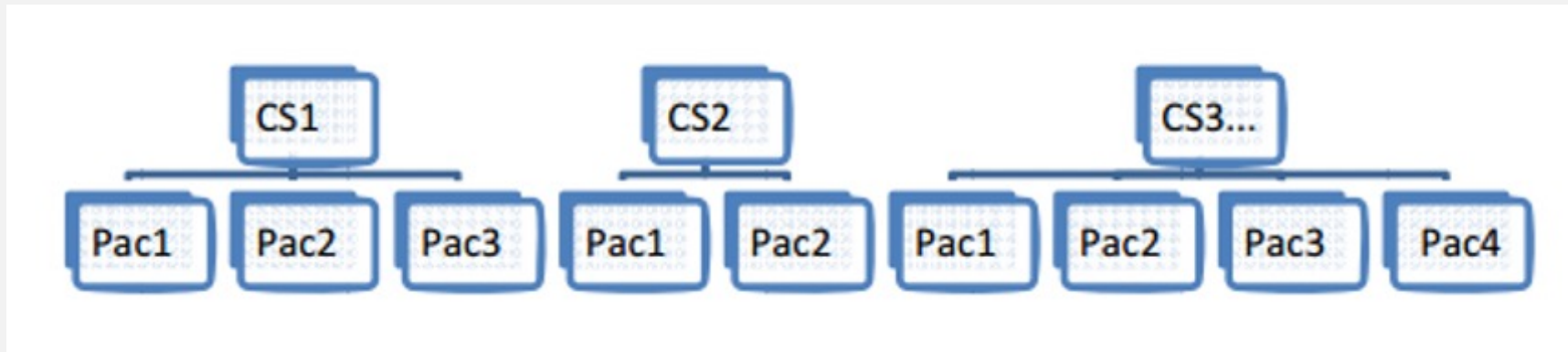


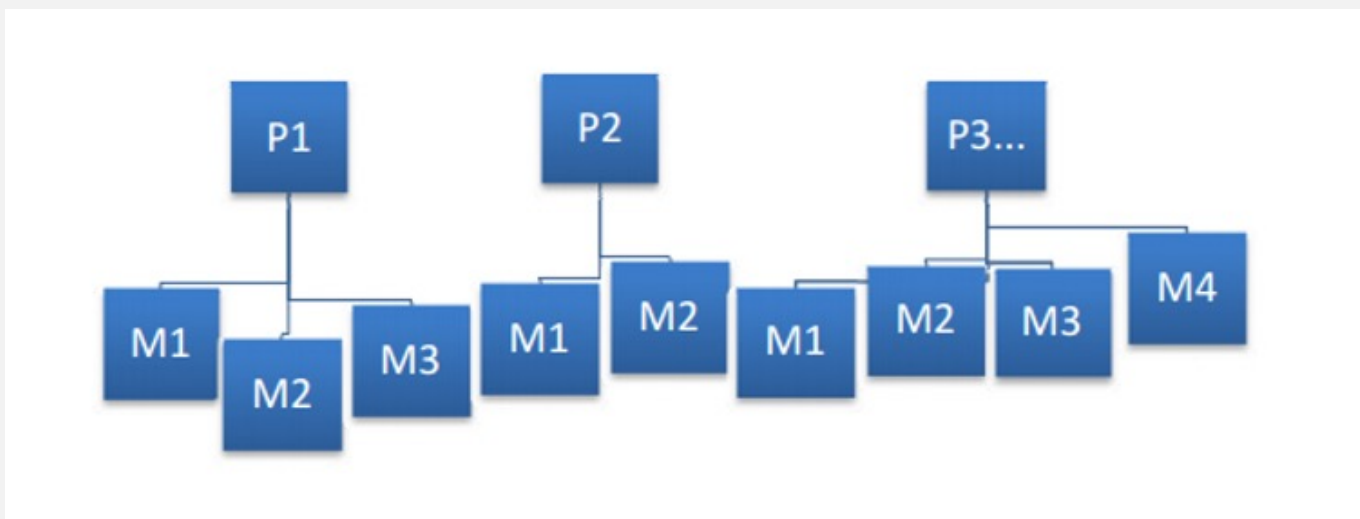
Diagrama de 2 nivells ennierrats

MODELS MIXTOS – DISSENYS LONGITUDINALS

- La característica més peculiar dels **estudis longitudinals** es la mesura repetida al llarg del temps de cada individu o objecte d'estudi.
- En Econometria, els dissenys longitudinals més freqüentment utilitzats són els **models de dades de panell** o també **panells de dades**.
- Un **model econòmic de dades de panell** és aquell que inclou una mostra d'agents econòmics o d'interès per un període determinat de temps, combinant ambdós tipus de dades (dimensió temporal i estructural).

MODELS MIXTOS – DISSENYS LONGITUDINALS

Exemple: Mesures repetides, dades de panell



Classificació i diagrama d'unitats d'un disseny de mesures repetides, mesures enriades en pacients

MODELS MIXTOS – DISSENYS LONGITUDINALS

Tipus de panell

- Els panells amb un nombre gran d'unitats individuals (observacions transversals), s'anomenen **panells micro** (exemple, moltes empreses).
- Mentre que si tenen un gran nombre d'observacions temporals s'anomenen **panells macro** (exemple, evolució d'una variable en el temps).
- Quan es donen les dues situacions, tant una dimensió transversal gran com una dimensió temporal gran, s'anomena **camp aleatori** (random field).

MODELS MIXTOS – DISSENYS LONGITUDINALS

A diferència dels dissenys longitudinals que s'utilitzen en altres disciplines, en Economia, les dades solen observar-se en intervals regulars.

Els dissenys de **dades de panell** poden ser:

- **Equilibrats (balanced)**, quan les observacions es repeteixen el mateix nombre de vegades en totes les unitats individuals.
- **No equilibrats (unbalanced)**, quan existeix almenys una unitat individual en la que les observacions no es repeteixen el mateix nombre de vegades que en la resta.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Avantatges generals dels estudis longitudinals

Els dissenys de **dades de panell** comparats amb altres dissenys unidimensionals:

- **Tenen un major nombre d'observacions** i, per tant, de graus de llibertat. Això implicarà una major eficiència (o precisió) en les estimacions.
- **Permeten capturar més variabilitat en les variables.** Això implicarà una menor probabilitat de cometre error tipus II, és a dir, tenen una major potència estadística.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Avantatges particulars dels estudis longitudinals

Els dissenys de **dades de panell** comparats amb els dissenys de sèries temporals:

- **Aprofiten la variabilitat transversal.** Si les variables incloses no presenten una excessiva variabilitat temporal, però sí transversal, el disseny de dades de panell permetrà estimacions dels paràmetres amb major precisió (eficiència).
- Respecte als dissenys transversals, aquests dissenys **aprofiten la variabilitat temporal.** Algunes variables poden presentar variabilitat temporal, però no transversal. Per tant, el seu efecte només es podria captar amb una dimensió temporal.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Avantatges particulars dels estudis longitudinals

Per últim, i segurament més important, els dissenys de **dades de panell** permeten incloure **factors no observables** pels quals no es disposa d'informació. Aquests factors, anomenats **heterogeneïtat**, poden ser de dos tipus:

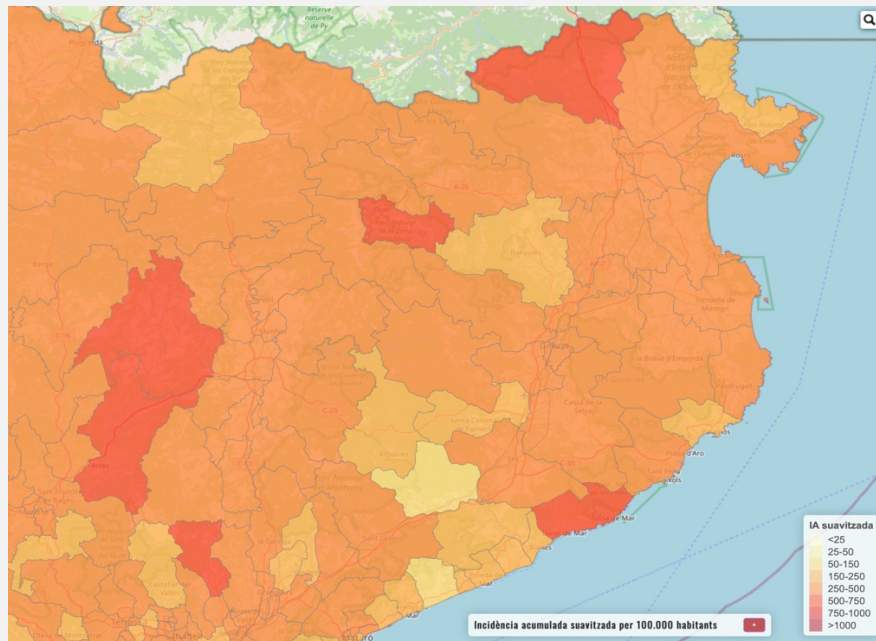
- **Heterogeneïtat individual.** Específica a cada individu o unitat d'estudi i constant en el temps. Correspon a la dimensió transversal.
- **Heterogeneïtat temporal.** Comú a tots els individus o unitats d'estudi i variant en el temps, però constant per les observacions transversals. Correspon a la dimensió temporal.

EXEMPLE PRÀCTIC

Per a tal d'entendre millor els dos tipus d'heterogeneïtat utilitzarem, com a exemple, informació sobre la pandèmia del COVID-19.

Mostrem, en primer lloc, la representació de la incidència acumulada per 100.000 habitants, en els últims 14 dies, en un mapa per municipis, centrat en la Regió Sanitària de Girona (la qual coincideix pràcticament amb la província de Girona, amb l'excepció de La Cerdanya, que pertany a la Regió Sanitària d'Alt Pirineu i Aran), en el període corresponent a la tercera onada (desembre 2020-febrer 2021).

EXEMPLE PRÀCTIC



Font: COVIDCAT (<https://ubidi.shinyapps.io/covidcat/>)

Observem que durant la tercera onada la majoria de municipis de la Regió tenien **una incidència entre 250 i 500 casos per 100.000 habitants** en els darrers 14 dies, però també observem **alguns entre 150 i 250** (Cadaqués, Llançà, Palamós, Santa Coloma de Farners, Arbúcies, Sant Hilari de Sacalm, Banyoles, Ribes de Freser i Campdevàrol) i **uns pocs entre 500 i 750** (Lloret de Mar, Tossa de Mar, Olot i una agrupació de municipis entorn a Agullana en l'Alt Empordà).

EXEMPLE PRÀCTIC

Com s'explica, per exemple, que Banyoles, Besalú i Olot, sent municipis veïns, tinguin incidències tant diferents? Banyoles en un rang 150-250, Besalú en un rang 250-500 i Olot en un rang 500-750.

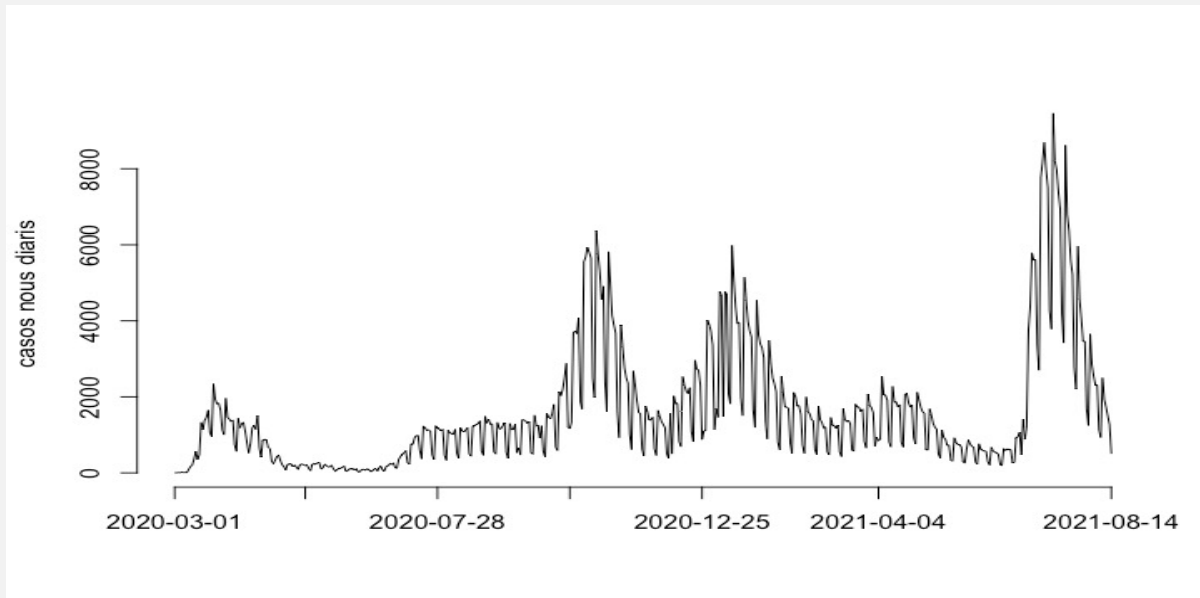
EXEMPLE PRÀCTIC

Com s'explica, per exemple, que Banyoles, Besalú i Olot, sent municipis veïns, tinguin incidències tant diferents? Banyoles en un rang 150-250, Besalú en un rang 250-500 i Olot en un rang 500-750.

Aquestes diferències i, en general, la distribució no homogènia de la incidència en el territori, podrien ser degudes a l'existència de variables no observades, específiques a cada unitat transversal (municipi en aquest cas) i que podrien no variar en el temps. Això és el que s'anomena **heterogeneïtat individual**.

EXEMPLE PRÀCTIC

Mostrem ara, l'evolució temporal del casos nous diaris en Catalunya, des de l'1 de març de 2020 fins el 14 d'agost de 2021.



Font de les dades: Registre de casos de COVID-19 realitzats a Catalunya. Segregació per sexe i àrea bàsica de salut (ABS)
(<https://analisi.transparenciacatalunya.cat/Salut/Registre-de-casos-de-COVID-19-realitzats-a-Catalun/xuwf-dxjd>)

EXEMPLE PRÀCTIC

S'observen amb claredat les cinc onades. Com es veu, tant la magnitud (alçada dels pics) com la durada varia entre les diferents onades. Per què?

EXEMPLE PRÀCTIC

S'observen amb claredat les cinc onades. Com es veu, tant la magnitud (alçada dels pics) com la durada varia entre les diferents onades. Per què?

Aquestes diferències podrien ser degudes a variables no observades, específiques a cada unitat temporal (les onades, en aquest cas) i, presumiblement, comunes a totes les unitats transversals (ABS en aquest cas). Aquestes diferències reben el nom d'**heterogeneïtat temporal**.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Especificació del model de dades de panell (o model mixt longitudinal)

Començant amb una resposta lineal (o normal), un model de dades de panell es pot especificar de la següent manera:

$$y_{it} = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3it} + \cdots + \beta_k x_{kit} + u_{it}$$

On el subíndex $i = 1, \dots, N$ denota les unitats individuals (N unitats individuals) i el subíndex $t = 1, \dots, t_i$ ($T = \max(t_i)$, número de períodes) denota les diferents observacions repetides per cada unitat individual.

Per exemple, i =cada pacient i t =tres observacions dels pacients en el temps.

MODELS MIXTOS – DISSENYS LONGITUDINALS

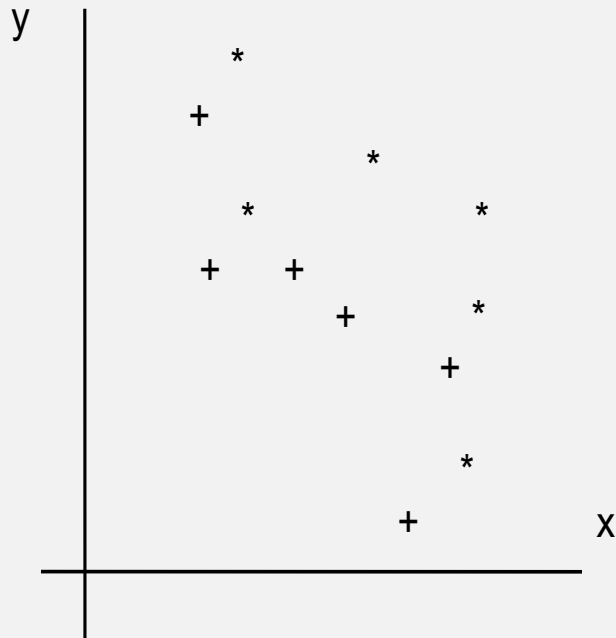
Especificació del model de dades de panell (o model mixt longitudinal)

Exemple: De forma simplificada, suposem que volem estimar l'efecte de l'edat (x) en la variació de la comprensió lectora (y) de sis nens menors de 6 anys, $i = 1, 2, 3, 4, 5, 6$ ($N = 6$). Els nens s'observen en dos moments del temps, dos períodes $t = 1 (+), 2(*)$ ($T = 2$) (model equilibrat). Diem que el model està equilibrat perquè es disposen de mesures de la capacitat lectora per a tots els nens i per a tots els períodes. En aquest cas, la variable explicativa és temps dependent, ja que l'edat dels nens és diferent en els dos moments del temps en el que són observats.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Exemple: En aquest cas, l'especificació del nostre model serà:

$$y_{it} = \beta_0 + \beta_1 x_{it} + u_{it}$$



MODELS MIXTOS – DISSENYS LONGITUDINALS

Especificació del model de dades de panell (o model mixt longitudinal)

Segons l'objectiu proposat, les dades de panell es poden aproximar marginal o condicionalment.

MODELS MIXTOS – DISSENYS LONGITUDINALS

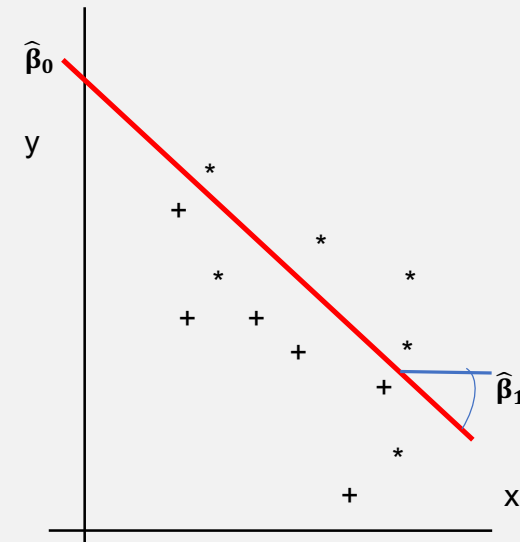
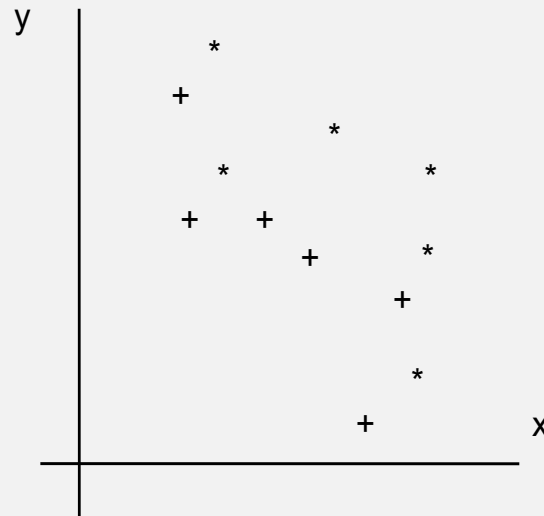
Especificació del model de dades de panell (o model mixt longitudinal). Enfocament marginal

Utilitzem **l'enfocament marginal** quan volem fer **inferències “poblacionals”**, és a dir, si el que volem és explicar la relació entre la variable dependent i les variables explicatives amb independència de la variabilitat intraindividual.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Especificació del model de dades de panell (o model mixt longitudinal). Enfocament marginal

En el nostre exemple:



MODELS MIXTOS – DISSENYS LONGITUDINALS

Especificació del model de dades de panell (o model mixt longitudinal). Enfocament marginal

L'ordenada en l'origen $\hat{\beta}_0$ i el coeficient associat a la variable explicativa $\hat{\beta}_1$ són comuns a tots els individus. No existeix heterogeneïtat individual. Dit d'altra manera, tots els efectes (de les variables explicatives, inclosa l'ordenada en l'origen) són fixes.

Observem que, en promig, a més edat del nen, menys variació en la comprensió lectora.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Especificació del model de dades de panell (o model mixt longitudinal). Enfocament marginal

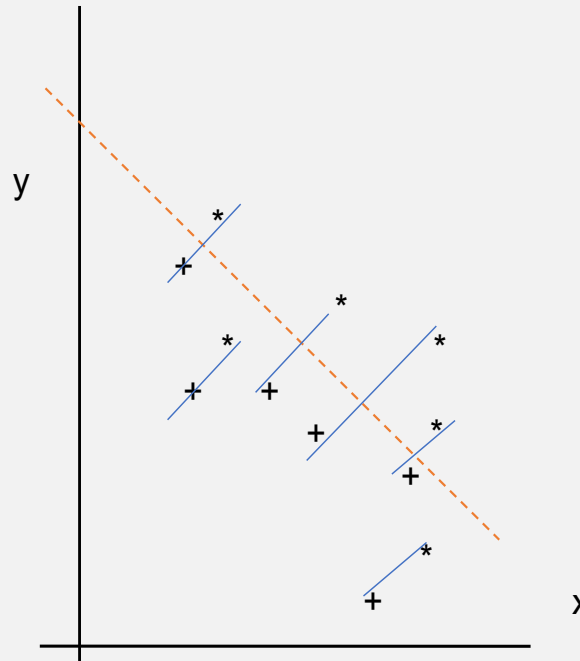
En **l'enfocament marginal**, es tracta d'estimar els paràmetres corresponents a la mitjana, β . Molt rarament són d'interès els paràmetres de la covariància. De fet, aquests solen ser tractats com una nosa en l'enfocament marginal, controlats, però no estimats.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Especificació del model de dades de panell (o model mixt longitudinal). Enfocament condicional

En l'enfocament condicional, volem fer inferències “individuals”.

En el nostre exemple:



MODELS MIXTOS – DISSENYS LONGITUDINALS

Especificació del model de dades de panell (o model mixt longitudinal). Enfocament condicional

En aquest cas, el coeficient associat a la variable explicativa $\hat{\beta}_1$ és aproximadament igual per a tots els individus. Però, l'ordenada en l'origen $\hat{\beta}_0$ és diferent per a cadascun d'ells. Existeix heterogeneïtat individual.

Amb independència del comportament promig, per cada nen en particular, a més edat, més variació de la comprensió lectora.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Especificació del model de dades de panell (o model mixt longitudinal). Enfocament condicional

En **l'aproximació condicional**, es modelitza simultàniament la mitjana de la variable dependent (la variabilitat inter-individual) i l'estructura de covariàncies o correlacions (la variabilitat intra-individual). En aquest enfocament, els paràmetres que defineixen la correlació tenen el mateix interès que els corresponents a la mitjana, a vegades inclús més.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Especificació del model de dades de panell (o model mixt longitudinal)

Fixeu-vos que en l'enfocament marginal, interessa el promig (o la inferència poblacional).

En l'enfocament condicional, interessa cada nen (o la inferència individual).

MODELS MIXTOS – DISSENYS LONGITUDINALS

Suposem que especifiquem el següent model:

$$y_{it} = \beta x_{it} + a_i + \tau_t + u_{it}$$

On a_i captura **l'heterogeneïtat individual no observable** (factors específics a cada unitat individual i que no varien en el temps); i τ_t captura **l'heterogeneïtat temporal**, comuna a totes les unitats individuals.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Model pooled

L'especificació més senzilla del model, anomenada **model pooled**, consisteix en ignorar l'heterogeneïtat ($a_i = a$ i $\tau_t = 0$):

$$y_{it} = a + \beta x_{it} + u_{it}$$

i estimar el model per Mínims Quadrats Ordinaris (MQO). Aquest mètode d'estimació es coneix com **Pooled Least Squares** (PLS).

MODELS MIXTOS – DISSENYS LONGITUDINALS

Model pooled. Problemes d'ignorar l'heterogeneïtat en un disseny mixt

És a dir, tot i que tenim un disseny mixt, ometem variables rellevants (l'heterogeneïtat).

Això significa que es comet un error d'especificació. Per tant, **l'estimació PLS dels paràmetres estarà esbiaixada i amb variàncies mal calculades.**

Possible solució:

Estimar el model mitjançant mètodes d'estimació alternatius ja sigui mitjançant l'aproximació marginal o bé mitjançant l'aproximació condicional. Nosaltres utilitzarem l'enfocament condicional.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

Els **efectes aleatoris** és l'enfocament més conegut d'entre els condicionals. Suposa que la dependència que pot existir entre diferents observacions de la variable dependent (és a dir, entre les respostes repetides d'un mateix individu), es deguda a que els coeficients de la regressió (de la mitjana) no són els mateixos per a tots els individus.

De fet, la interpretació més senzilla, és la que suposa que l'heterogeneïtat individual es deu a factors no observables (o variables omeses) fixes en el temps, però variables entre individus. Aquests factors són els que provoquen la dependència entre les diferents observacions de la variable dependent.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

Per tant, aquest model **suposa que els coeficients de la regressió no són els mateixos per tots els individus:**

$$y_{it} = \beta_{it}x_{it} + a_i + \tau_t + u_{it}$$

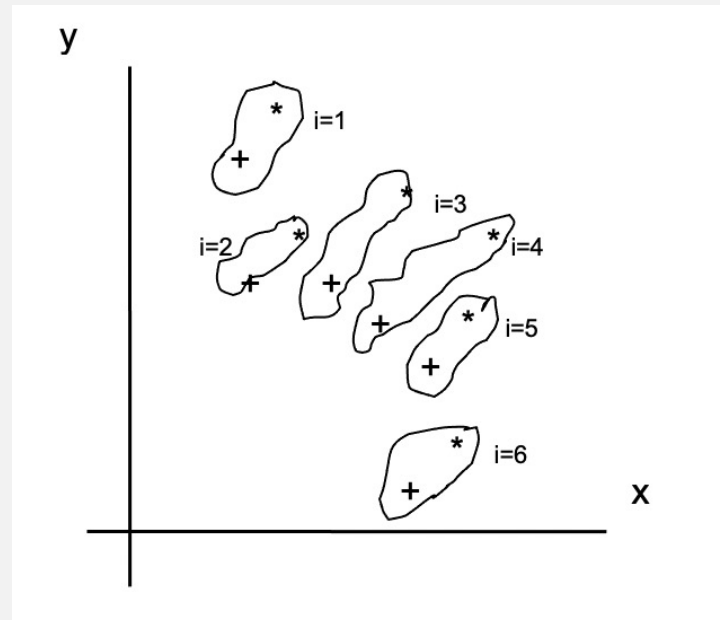


Poden variar entre individus

MODELS MIXTOS – DISSENYS LONGITUDINALS

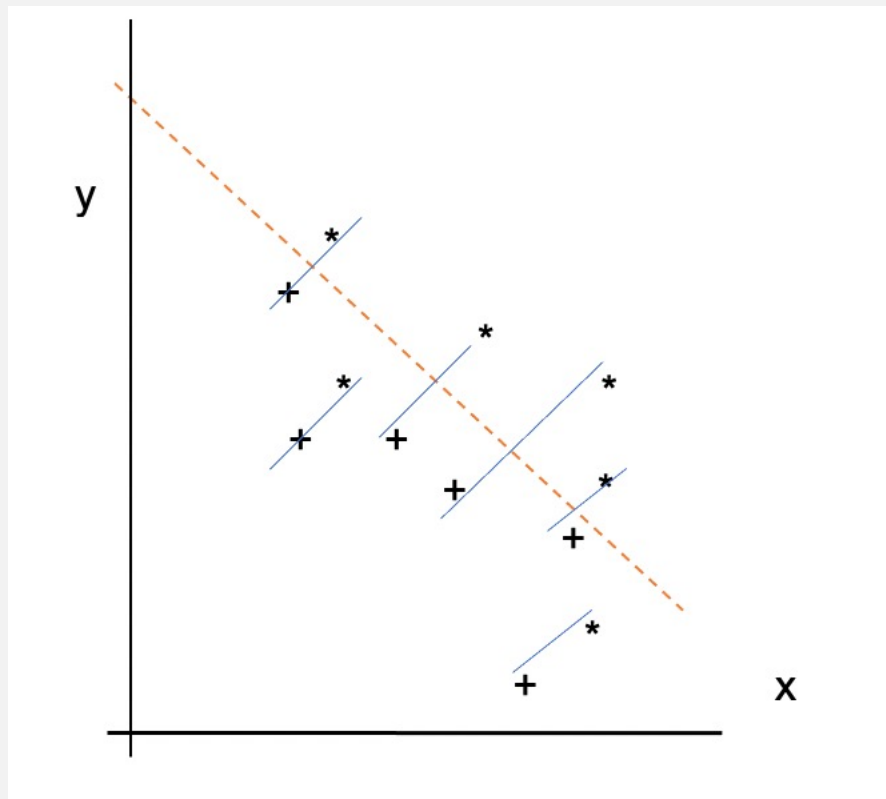
Enfocament condicional. Model d'efectes aleatoris

Suposem l'exemple que hem vist abans, en el que vam considerar 6 nens i dos períodes (representem el primer període mitjançant el símbol + i el segon mitjançant el símbol *):



MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris



En aquest cas, el coeficient associat a la variable explicativa és aproximadament el mateix per a tots els nens. És a dir, la comprensió lectora varia més o menys de la mateixa manera per tots els nens.

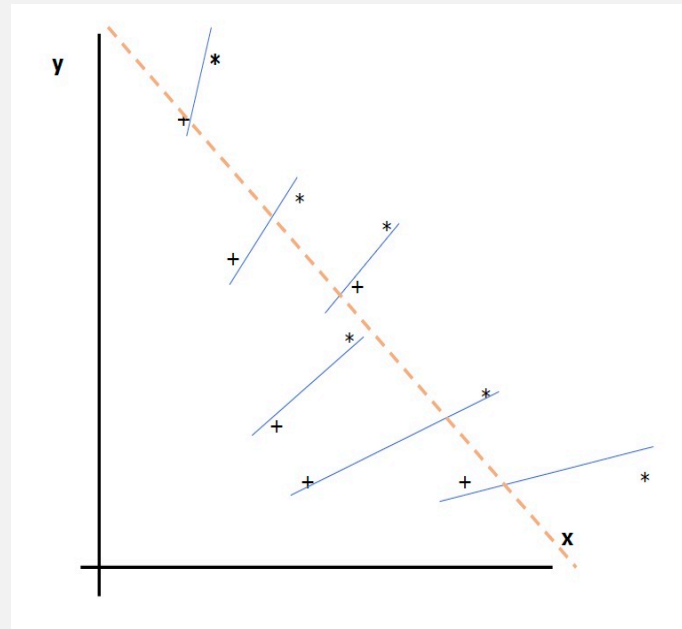
Pel contrari, l'ordenada en l'origen és diferent per cada un d'ells. La comprensió lectora basal és diferent.

Existeix heterogeneïtat individual!!!

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

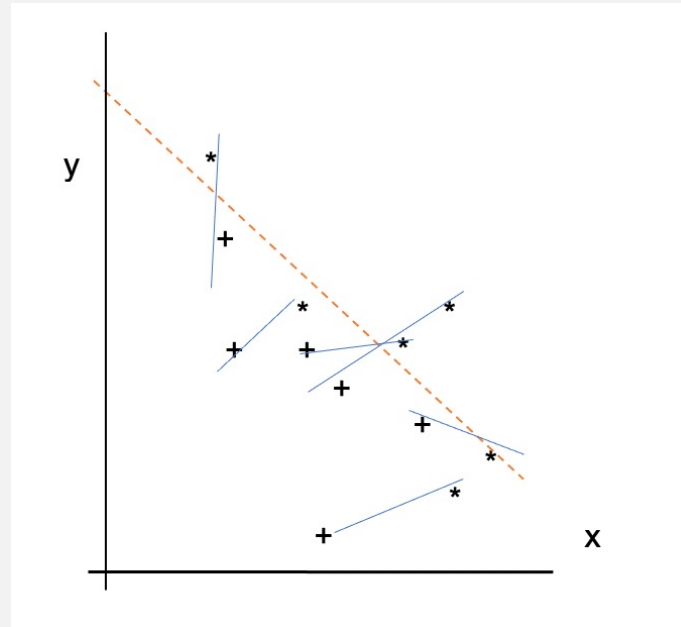
En el següent exemple, el paràmetre associat a la variable explicativa i no l'associat a l'ordenada en l'origen, és el que és l'efecte aleatori.



MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

També es podrien donar les dues situacions. És a dir, que tant l'ordenada en l'origen com la pendent siguin efectes aleatoris.



MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

- En Econometria, quan es parla d'un **model d'efectes aleatoris**, ens referim precisament al primer model. És a dir, només existeix un efecte aleatori que es correspon amb l'ordenada en l'origen (és a dir, heterogeneïtat individual no observada).

$$y_{it} = \beta_{0i} + \beta_1(x_{it} - x_{it-1}) + \beta_p x_{it} + u_{it}$$



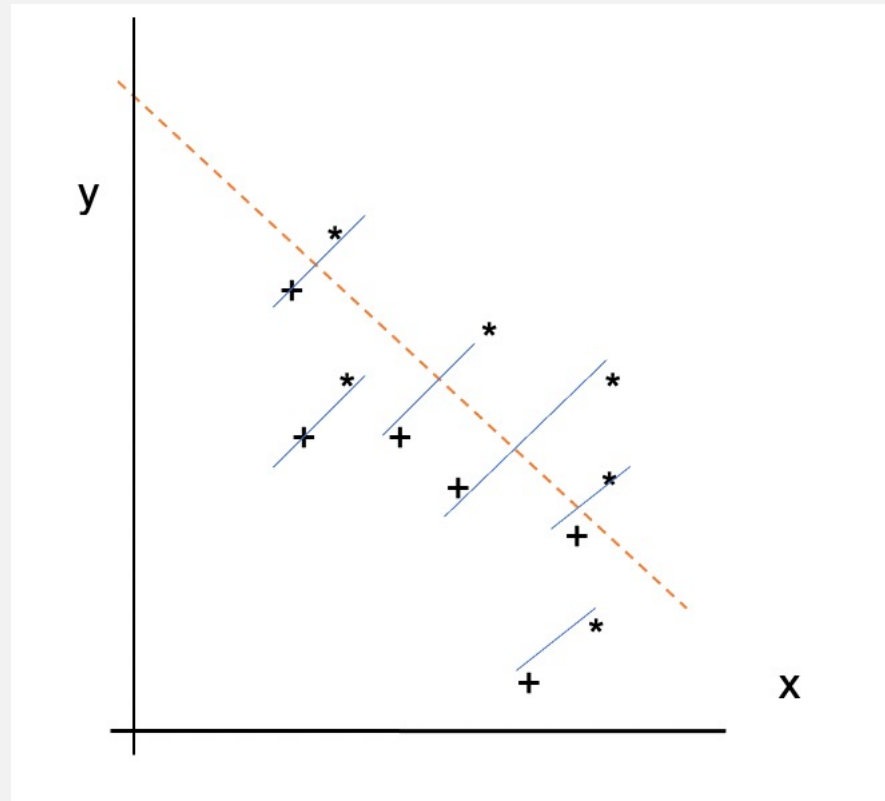
- En canvi, en el segon model, estaríem parlant d'un **model de coeficients aleatoris**. És a dir, en aquest cas l'efecte aleatori estaria associat a una variable explicativa.

$$y_{it} = \beta_0 + \beta_{1i}(x_{it} - x_{it-1}) + \beta_p x_{it} + u_{it}$$



MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris



MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

En resum i seguint amb el nostre exemple, en el que l'efecte aleatori és en l'ordenada en l'origen però no en la pendent:

$$\begin{array}{ll} \text{Per a } i = 1 & y_{1t} = \beta_{01} + \beta_1 x_{1t} + u_{1t} \\ \text{Per a } i = 2 & y_{2t} = \beta_{02} + \beta_1 x_{2t} + u_{2t} \\ \text{Per a } i = 3 & y_{3t} = \beta_{03} + \beta_1 x_{3t} + u_{3t} \\ \text{Per a } i = 4 & y_{4t} = \beta_{04} + \beta_1 x_{4t} + u_{4t} \\ \text{Per a } i = 5 & y_{5t} = \beta_{05} + \beta_1 x_{5t} + u_{5t} \\ \text{Per a } i = 6 & y_{6t} = \beta_{06} + \beta_1 x_{6t} + u_{6t} \end{array}$$

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

El problema és que per a cada individu sol haver-hi molt poques observacions per estimar els β a partir únicament de (y_{it}, x_{it}) . Per exemple, en el nostre cas tenim 7 paràmetres desconeguts ($6 \beta_{oi}, i = 1, \dots, 6$ i $1 \beta_1$) amb 12 observacions (6 nens x 2 períodes). **ESTIMACIONS INEFICIENTS!!!**

Haurem d'estimar el model mitjançant un **model lineal generalitzat mixt (GLMM)**.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris. Model lineal generalitzat mixt

Haurem de suposar que els β no són més que realitzacions independents d'alguna distribució de probabilitat (la normal en models lineals) amb mitjana β :

$$\begin{aligned}y_{it} &= \beta_{0i} + \beta_1 x_{it} + u_{it} \\ \beta_{0i} &= \beta_0 + \eta_i\end{aligned}$$

Resumint, hem especificat un model mixt amb un efecte fixe (β_1) i un efecte aleatori (β_{0i}), la qual cosa significa que β_{0i} pot variar entre els individus ($i = 1, 2, \dots, 6$).

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

Observeu que, en el model d'efectes aleatoris, dins de l'enfocament condicional, també estimarem els paràmetres poblacionals, β_0 i β_1 , corresponents a l'enfocament marginal.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

Ara, a més a més de suposar:

$$\begin{aligned}E(u_{it}) &= 0 \\E(u_{it}^2) &= \sigma_u^2 \text{ constant} \\E(u_{it} \times u_{js}) &= 0 \quad \forall i \neq j \quad \forall t \neq s\end{aligned}$$

Haurem de suposar també:

$$\begin{aligned}E(\eta_i) &= 0 \\E(\eta_i^2) &= \sigma_\eta^2 \text{ constant} \\E(\eta_i, \eta_j) &= 0 \quad \forall i \neq j \\E(u_{it}, \eta_j) &= 0 \quad \forall i, t, j\end{aligned}$$

Les dues últimes hipòtesis són particularment importants.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

- En primer lloc, $E(\eta_i, \eta_j) = 0 \forall i \neq j$, significa que els nivells superiors han de ser independents entre sí. L'incompliment d'aquesta hipòtesis (anomenada efectes creuats o crossover) exigeix la utilització de mètodes d'estimació molt més complexos que els explicats aquí (per exemple, mètodes Bayesians).
- En segon lloc, $E(u_{it}, \eta_j) = 0 \forall i, t, j$, exigeix que els efectes aleatoris han de ser independents dels efectes fixos. Altrament, els efectes han de ser efectivament completament aleatoris.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

Anem a aplicar les dues últimes hipòtesis en el nostre exemple.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

- En primer lloc, $E(\eta_i, \eta_j) = 0 \forall i \neq j$, significa que els nivells superiors han de ser independents entre sí.

Seguint amb el nostre exemple, els nens són el nivell superior, mentre que el temps constitueix el nivell inferior, ja que cada nen “conté” (de fet, és observat) dos períodes de temps. Així doncs, la variació en la comprensió lectora d'un nen no depèn de la variació en la comprensió lectora de qualsevol altre. És a dir, els nens són independents entre ells, pel que es compleix aquest hipòtesi.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

Ara bé, podria passar que els nens fossin família (per exemple, suposem que els nens 2 i 5 són germans). En aquest cas, la variació de la comprensió lectora del nen 2 i del nen 5 no serien independents, incomplint-se la hipòtesis.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

Però, perquè això es compleixi, n'hi ha prou en re-especificar el model afegint un nivell addicional:

$$\begin{aligned}y_{kit} &= \beta_{0k} + \beta_1 x_{kit} + u_{kit} \\ \beta_{0k} &= \beta_0 + \eta_k\end{aligned}$$

On k ($k = 1$ nens 2 i 5; $k = 2$ nen 3; $k = 3$ nen 4; $k = 4$ nen 6) denota la família, i ($i = 1, 2, \dots, 6$) denota el nen i t ($t = +, *$) denota el període. Ara, les unitats del nivell superior (famílies) sí que són independents, complint-se la hipòtesis.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

- En segon lloc, $E(u_{it}, \eta_j) = 0 \forall i, t, j$, exigeix que els efectes aleatoris han de ser independents dels efectes fixos.

En el nostre exemple, suposem que en el model que estem utilitzant, les diferències en les ordenades en l'origen entre els nens s'expliquen per alguna variable, per exemple, el sexe del nen. És a dir:

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

És a dir:

$$\begin{aligned}y_{it} &= \beta_{oi} + \beta_1 x_{it} + u_{it} \\ \beta_{oi} &= \beta_0 + \gamma \text{sexe}_i + \eta_i\end{aligned}$$

Així, aïllant:

$$\begin{aligned}y_{it} &= \beta_0 + \gamma \text{sexe}_i + \eta_i + \beta_1 x_{it} + u_{it} \\ y_{it} &= \beta_0 + \beta_1 x_{it} + u_{it} + \gamma \text{sexe}_i + \eta_i\end{aligned}$$

I, finalment, arreglant termes:

$$y_{it} = \beta_0 + \beta_1 x_{it} + v_{it}$$

on $v_{it} = u_{it} + \gamma \text{sexe}_i + \eta_i$.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris

Evidentment, v_{it} estarà correlacionat amb u_{it} , incomplint-se la hipòtesis que **els efectes aleatoris siguin independents dels efectes fixos** i els estimadors seran inconsistents.

Perquè es torni a complir, s'han d'incloure totes les variables explicatives en l'equació principal (la de la mitjana) i deixar l'efecte completament aleatori. És a dir:

$$y_{it} = \beta_{oi} + \beta_1 x_{it} + \gamma \text{sexe}_i + u_{it}$$
$$\beta_{oi} = \beta_0 + \eta_i$$

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris. Estimació del model

Si es compleixen les dues hipòtesis, els models **s'estimen per mètodes basats en la màxima versemblança**. El problema és que només existeix una funció de versemblança completa quan no existeix correlació, existeixin o no efectes aleatoris. Però, es pot definir la quasi-versemblança, que té exactament la mateixa forma que la derivada parcial de la versemblança completa.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Enfocament condicional. Model d'efectes aleatoris. Estimació del model

La maximització de la quasi-versemblança és molt complexa. Per altra banda, els mètodes Bayesians, als quals podem recórrer, impliquen l'avaluació total de la quasi-versemblança, per lo que solen ser molt intensius computacionalment.

Alternativament, s'utilitza la pseudo-versemblança o pseudo-versemblança penalitzada, PQL, la versemblança restringida (REML) o el mètode de Laplace.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Resumint:

- L'estimador d'efectes fixes (aproximació marginal) permet estimar el model sota supòsits menys restrictius que l'estimador d'efectes aleatoris.
- Tot i això, si es compleixen les condicions de regularitat, l'estimador d'efectes aleatoris és més eficient que el d'efectes fixes. Però pot ser inconsistent si existeix correlació entre les variables explicatives i la heterogeneïtat individual.
- Per mètodes Bayesians, els estimadors d'efectes aleatoris sempre són consistents.

MODELS MIXTOS – DISSENYS LONGITUDINALS

Recordeu que les condicions de regularitat a les que ens estem referint en el segon punt són:

$$\begin{aligned}E(u_{it}) &= 0 \\Var(u_{it}) &= E(u_{it}^2) = \sigma_u^2 \text{ constant} \\Cov(u_{it}, u_{js}) &= E(u_{it} \times u_{js}) = 0 \quad \forall i \neq j \quad \forall t \neq s\end{aligned}$$

