

□

# 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

# PANORÁMICA DE LOS MODELOS MIXTOS

1. Modelo de regresión lineal
2. Modelo de regresión logística
3. Modelo lineal generalizado (GLM)
4. Modelos mixtos

# PANORÁMICA DE LOS MODELOS MIXTOS

1. **Modelo de regresión lineal**
2. Modelo de regresión logística
3. Modelo lineal generalizado (GLM)
4. Modelos mixtos

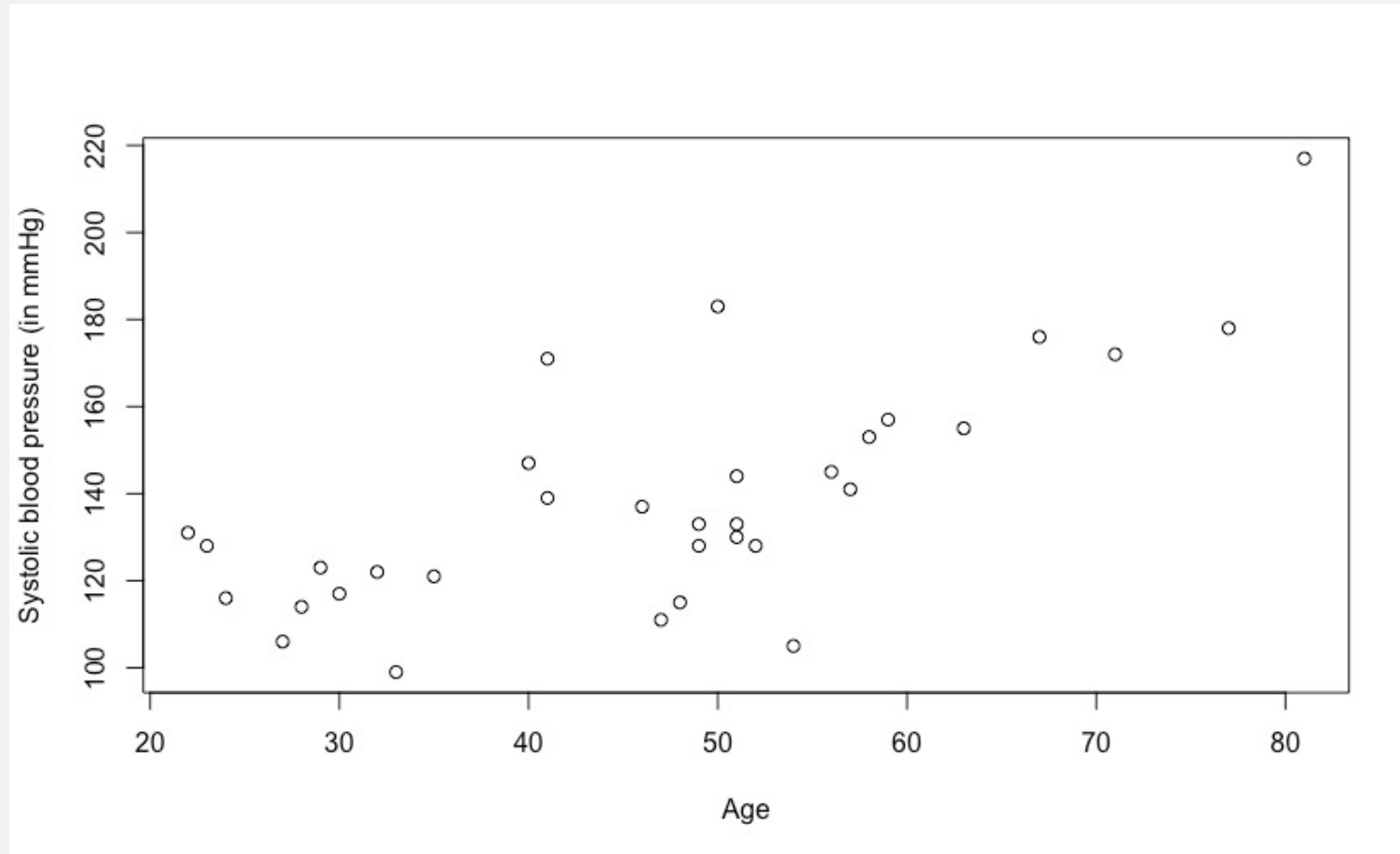
# MODELO DE REGRESIÓN LINEAL

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

# MODELO DE REGRESIÓN LINEAL



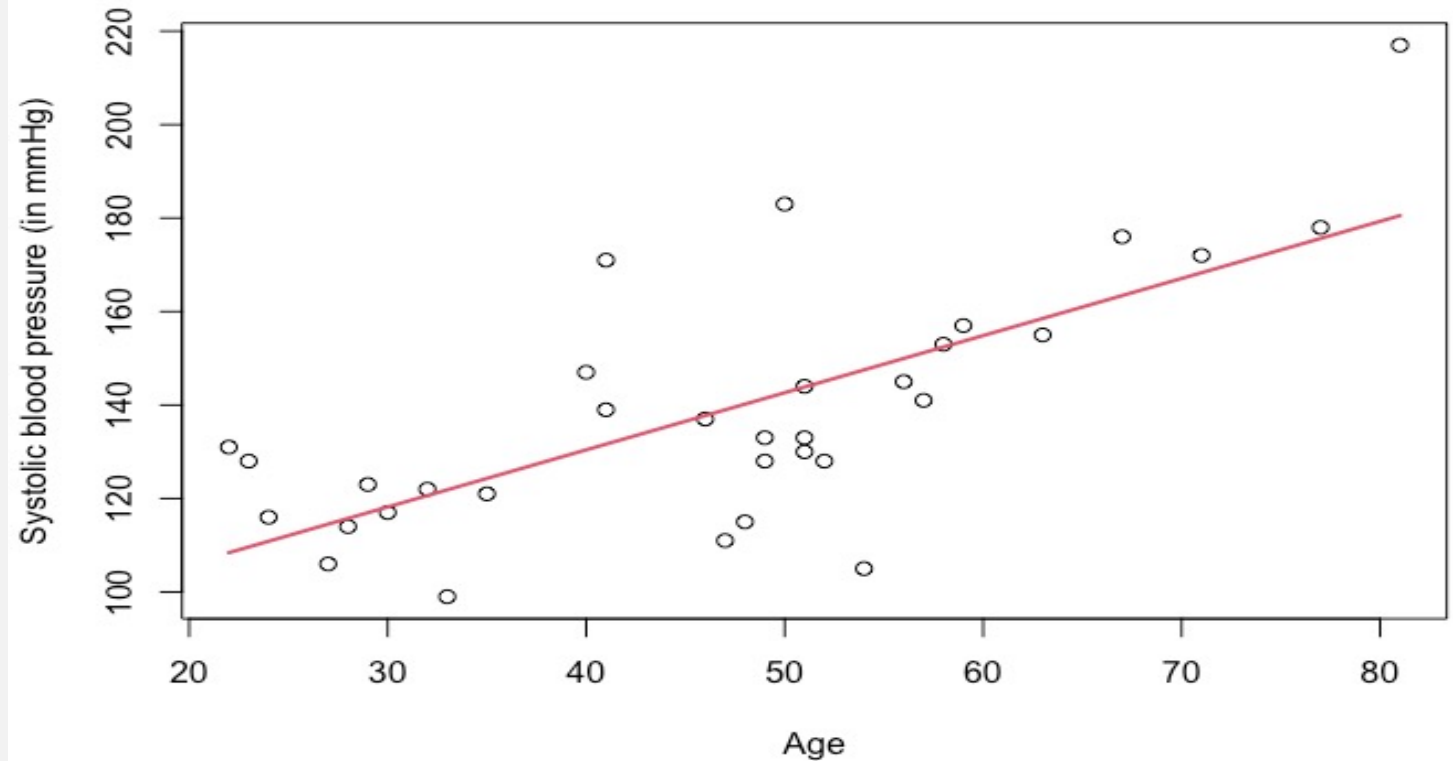


# MODELO DE REGRESIÓN LINEAL

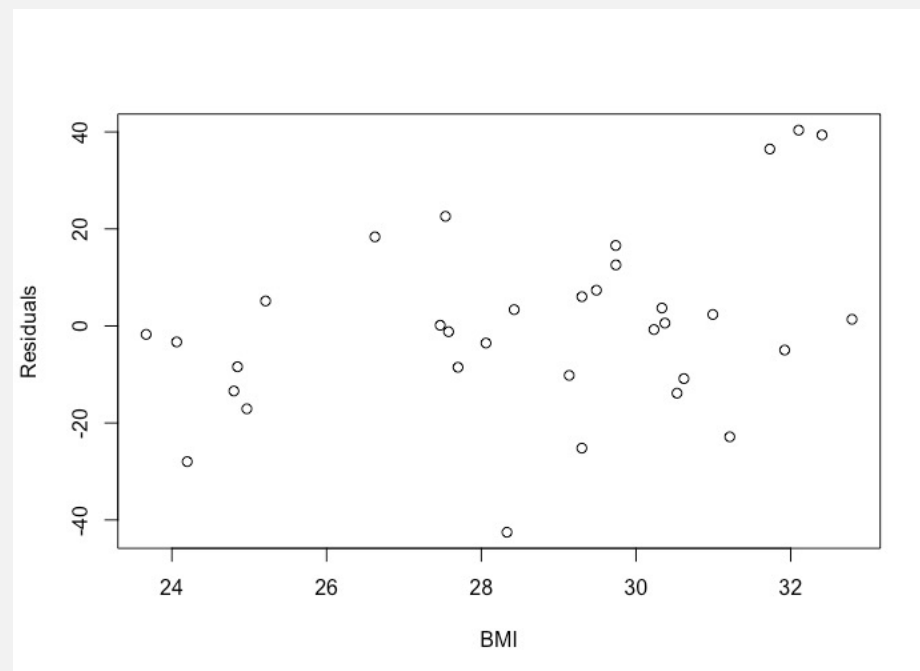
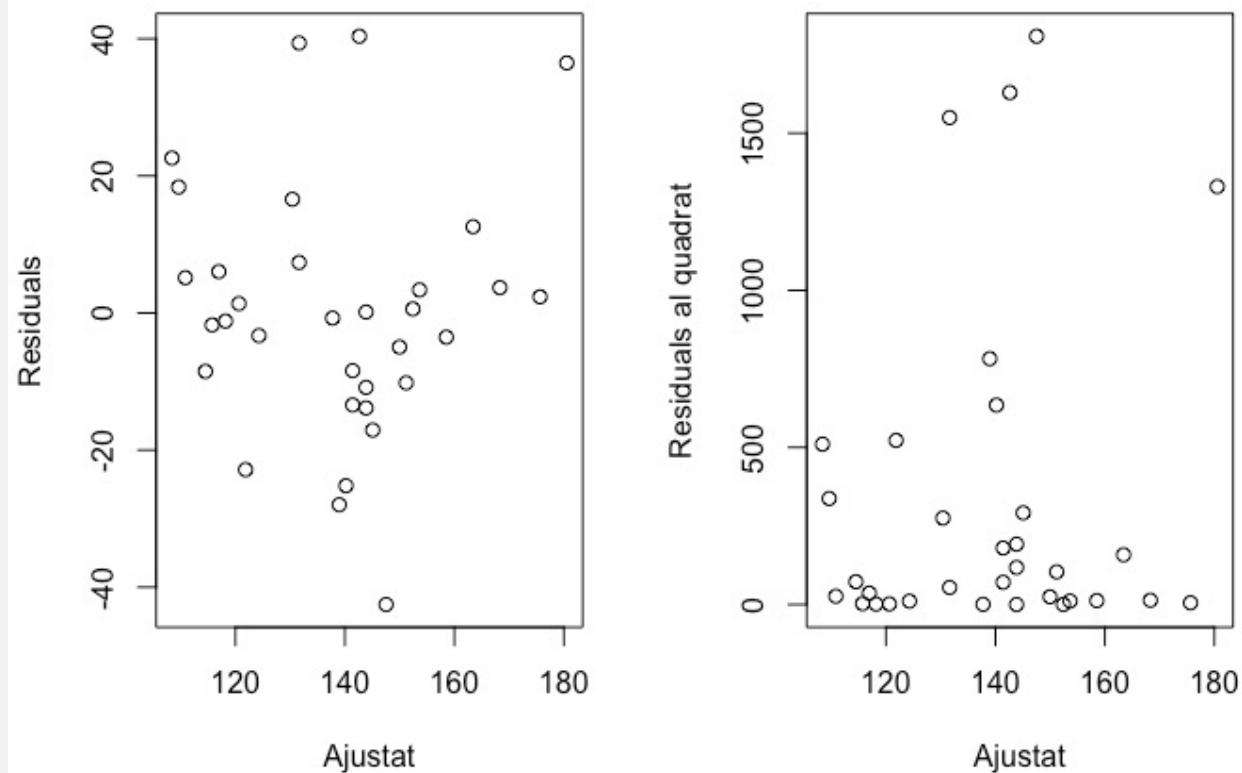
$$SBP_i = \beta_0 + \beta_1 age_i + u_i$$

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

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



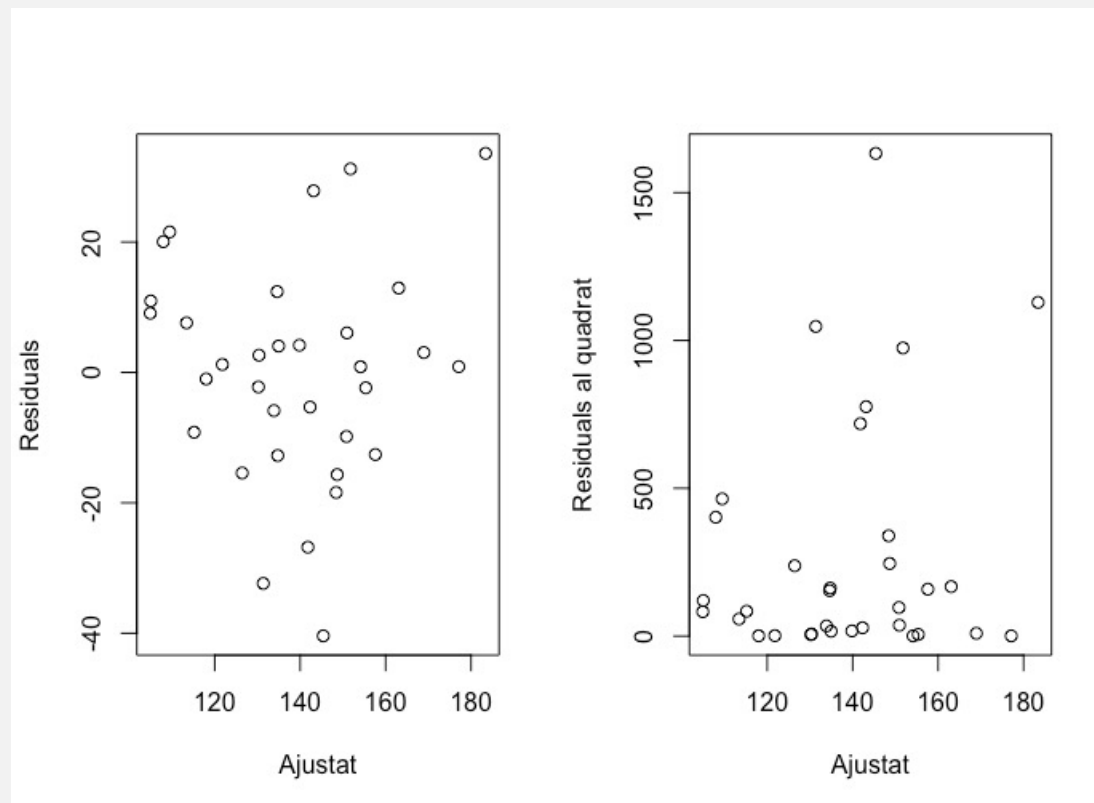
# MODELO DE REGRESIÓN LINEAL



## 3. Panorámica de los modelos mixtos

# MODELO DE REGRESIÓN LINEAL

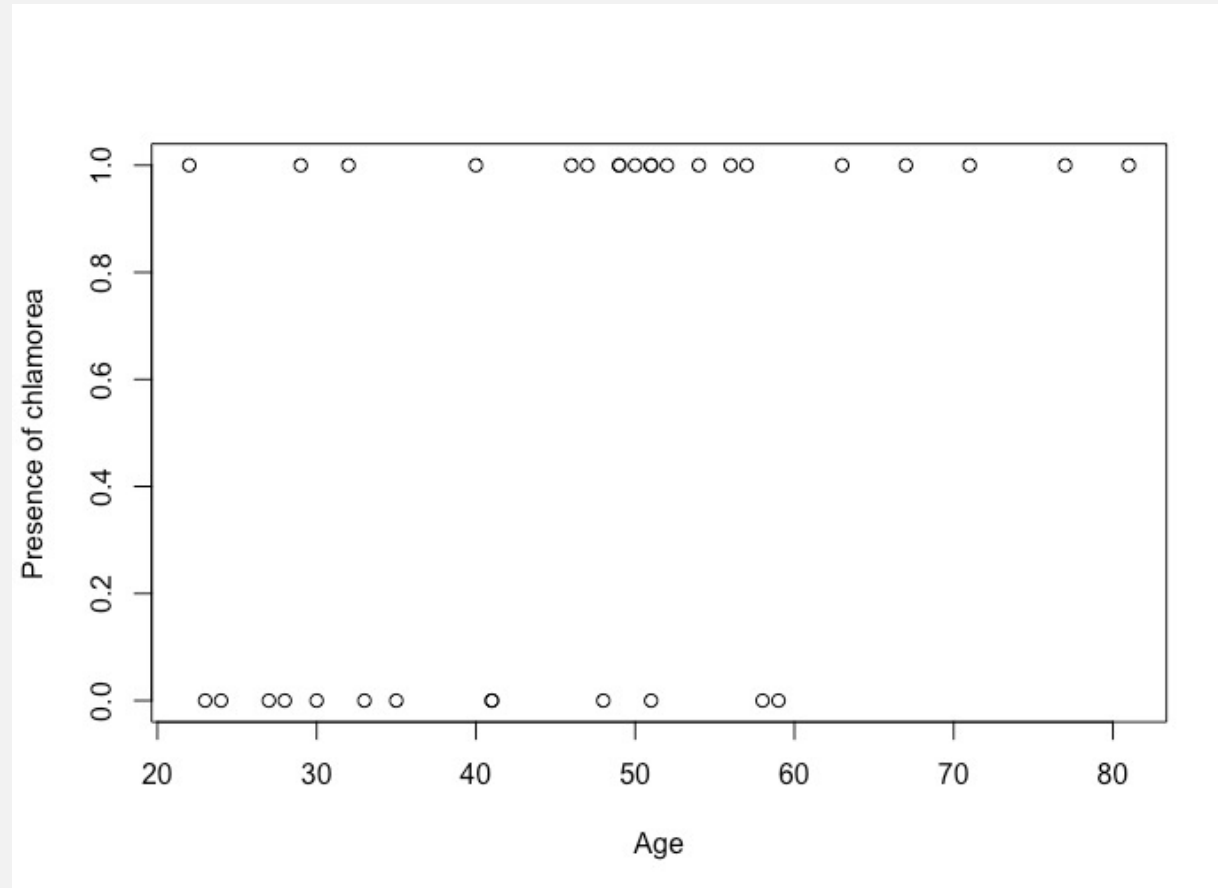
$$\widehat{SBP}_i = 8,9631 + 1,0538 \text{ age}_i + 2,8075 \text{ BMI}_i$$



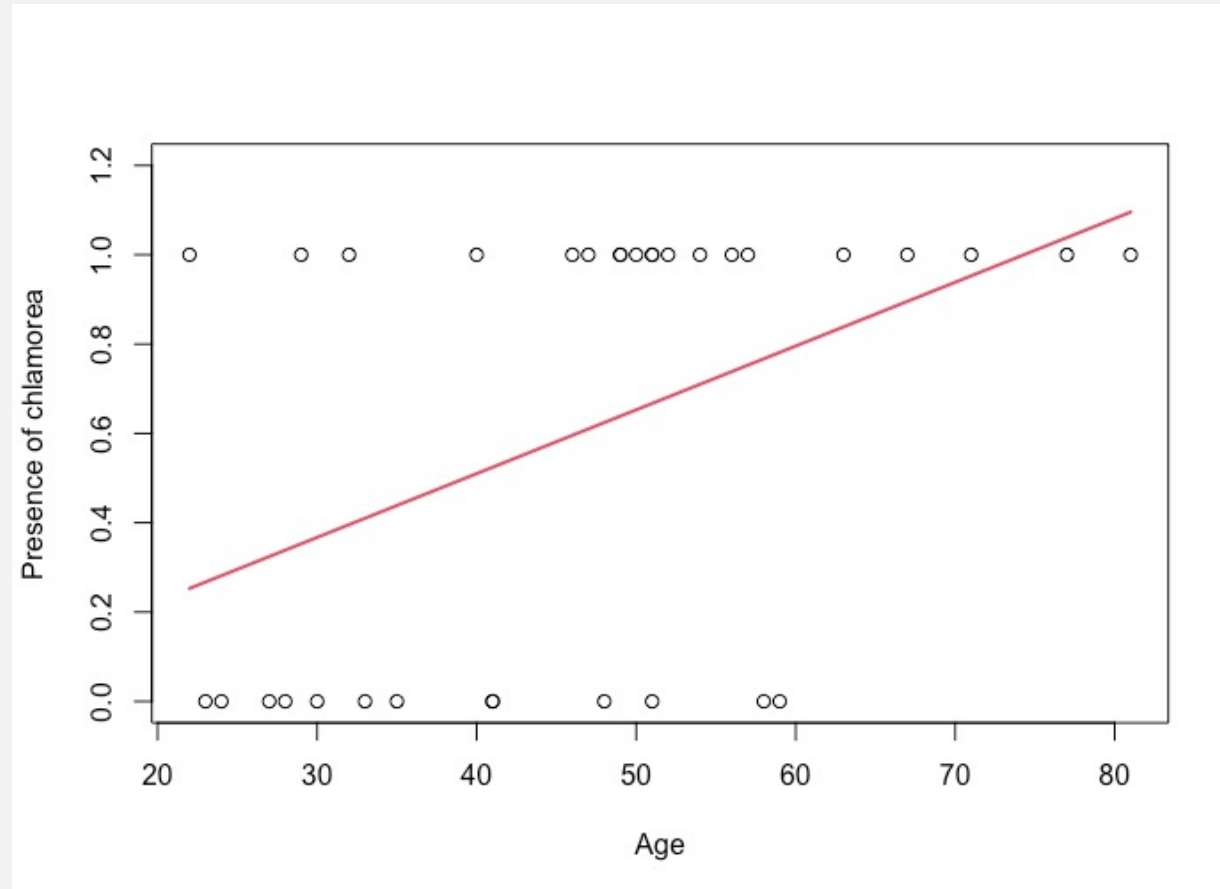
# PANORÁMICA DE LOS MODELOS MIXTOS

1. Modelo de regresión lineal
2. **Modelo de regresión logística**
3. Modelo lineal generalizado (GLM)
4. Modelos mixtos

# MODELO DE REGRESIÓN LOGÍSTICA



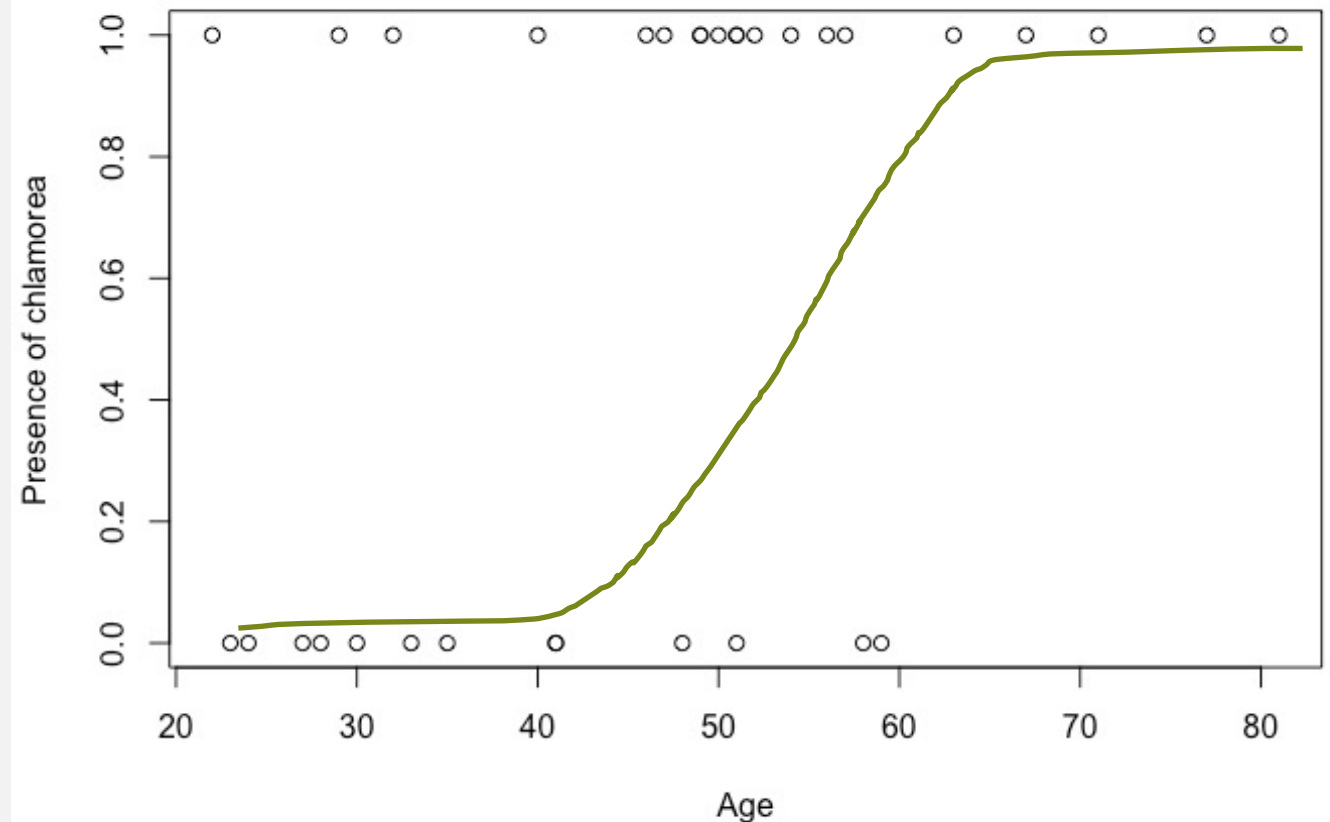
# MODELO DE REGRESIÓN LOGÍSTICA



# MODELO DE REGRESIÓN LOGÍSTICA

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

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



# PANORÁMICA DE LOS MODELOS MIXTOS

1. Modelo de regresión lineal
2. Modelo de regresión logística
- 3. Modelo lineal generalizado (GLM)**
4. Modelos mixtos



## MODELO LINEAL GENERALIZADO (GLM)

- Tal y como su nombre indica, el **modelo lineal generalizado, GLM** (McCullagh y Nelder, 1989), generaliza el modelo de regresión lineal para variables dependientes no normales, como por ejemplo, variables de recuento (distribución de Poisson) o variables dicotómicas (distribución binomial), permitiendo así utilizar el mismo tipo de modelización, especificación, estimación y diagnóstico.
- El modelo lineal generalizado es una extensión del modelo lineal cuando las observaciones son independientes, pero los supuestos sobre la normalidad de las perturbaciones no se cumplen.

## MODELO LINEAL GENERALIZADO (GLM)

- La distribución de las perturbaciones se puede escoger dentro del conjunto de distribuciones que pertenecen a la familia de **distribuciones exponenciales** de un parámetro que incluye las distribuciones Binomial, Poisson, Binomial Negativa, Gamma, Beta, Inversa Gaussiana y dentro del cual la distribución Normal es un caso especial.

## MODELO LINEAL GENERALIZADO (GLM)

- El objetivo de los GLM es el de describir la dependencia de la respuesta (o variable dependiente)  $y$ , respecto de las variables explicativas,  $E(y | x) = \mu$ .
- Los GLM se componen de dos funciones: la función 'vínculo' (link),  $g()$ , y la función de 'variancia',  $v()$ . A veces también hay un parámetro de escala,  $\phi$ .

## MODELO LINEAL GENERALIZADO (GLM)

- Así, por ejemplo, cuando la **variable dependiente es continua**, el GLM se especificaría como (equivalente a una regresión lineal):

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

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

El vínculo es lineal y la variancia constante.

## MODELO LINEAL GENERALIZADO (GLM)

- Cuando la **variable dependiente es dicotómica**, el GLM (equivalente a una regresión logística), se especificaría:

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

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

En este caso, el vínculo es un logit. El parámetro  $\phi$  se denomina de sobredispersión (si  $\phi > 1$ ) o de subdispersión (si  $\phi < 1$ ).

## MODELO LINEAL GENERALIZADO (GLM)

- Cuando la **variable dependiente es discreta (o de recuento)**, el GLM (equivalente a una regresión de Poisson), se especificaría:

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

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

El vínculo es logarítmico. El parámetro  $\phi$  se denomina de sobredispersión (si  $\phi > 1$ ) o de subdispersión (si  $\phi < 1$ ).

# PANORÁMICA DE LOS MODELOS MIXTOS

1. Modelo de regresión lineal
2. Modelo de regresión logística
3. Modelo lineal generalizado (GLM)
4. **Modelos mixtos**

# MODELOS MIXTOS

Los **diseños mixtos** se caracterizan por considerar simultáneamente dos o más dimensiones de análisis.

Los **diseños mixtos** comprenden:

- los **diseños multinivel** o de niveles múltiples (también denominados jerárquicos)
- los **diseños longitudinales** o de medidas repetidas.



## MODELOS MIXTOS – DISEÑOS MULTINIVEL

- Tienen una estructura jerárquica, con agrupaciones de datos en grupos o clústeres.
- En la literatura, estos niveles jerárquicos se denominan nivel 1, nivel 2, etc.; estadio 1, estadio 2, etc.; nivel individual y nivel poblacional; nivel individual y nivel clúster.
- Por ejemplo, cuando tenemos países y regiones, tenemos dos niveles jerárquicos.

# MODELOS MIXTOS – DISEÑOS MULTINIVEL

Ejemplo: Pacientes dentro de centros de salud

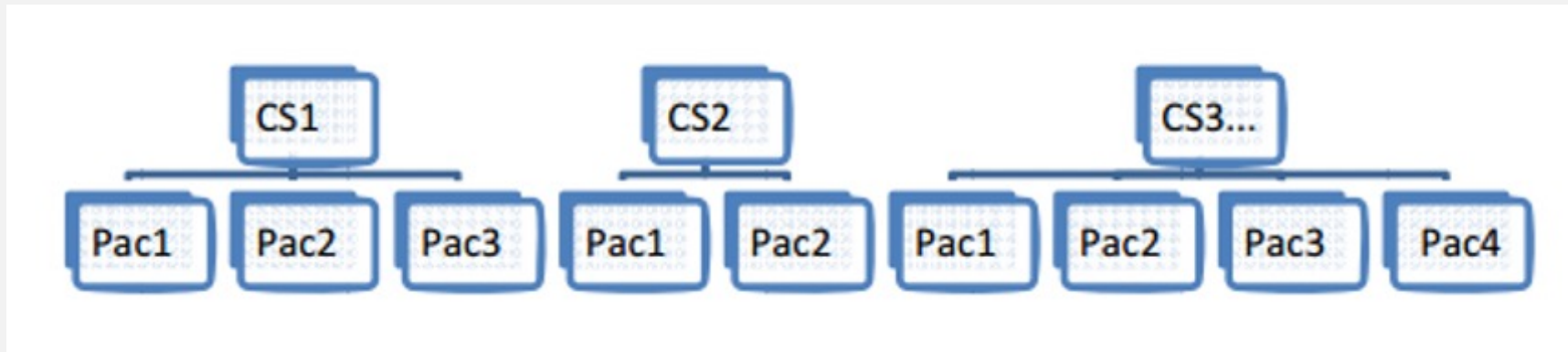


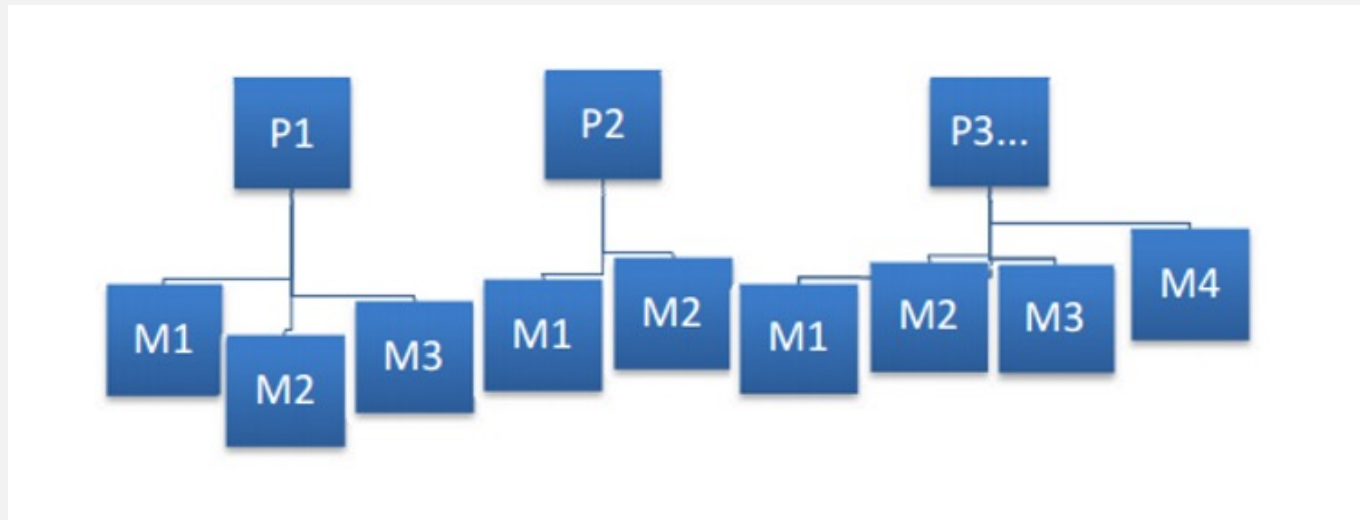
Diagrama de 2 niveles anidados

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

- La característica más peculiar de los **estudios longitudinales** es la medición repetida a lo largo del tiempo de cada individuo u objeto de estudio.
- En Econometría, los diseños longitudinales más frecuentemente utilizados son los **modelos de datos de panel**, también conocidos con el nombre de **paneles de datos**.
- Un **modelo económico de datos de panel** es aquel que incluye una muestra de agentes económicos o de interés para un período determinado de tiempo, combinando ambos tipos de datos (dimensión temporal y estructural).

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

**Ejemplo:** Medidas repetidas, datos de panel



Clasificación y diagrama de unidades de un diseño de medidas repetidas, medidas anidadas en pacientes

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Tipos de panel

- Los paneles con un número grande de unidades individuales (observaciones transversales), se denominan **paneles micro** (por ejemplo, muchas empresas).
- Por el contrario, si tienen un gran número de observaciones temporales se llaman **paneles macro** (por ejemplo, la evolución de una variable en el tiempo).
- Cuando se dan las dos situaciones, tanto una dimensión transversal grande como una dimensión temporal grande, se denomina **campo aleatorio** (random field).

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

A diferencia de los diseños longitudinales que se utilizan en otras disciplinas, en Economía, los datos suelen observarse en intervalos regulares.

Los diseños de **datos de panel** pueden ser:

- **Equilibrados (balanced)**, cuando las observaciones se repiten el mismo número de veces en todas las unidades individuales.
- **No equilibrados (unbalanced)**, cuando existe al menos una unidad individual en la que las observaciones no se repiten el mismo número de veces que en el resto.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Ventajas generales de los estudios longitudinales

Los diseños de **datos de panel** comparados con otros diseños unidimensionales:

- **Tienen un mayor número de observaciones** y, por lo tanto, de grados de libertad. Esto implicará una mayor eficiencia (o precisión) en las estimaciones.
- **Permiten capturar más variabilidad en las variables.** Esto implicará una menor probabilidad de cometer error tipo II, es decir, tienen una mayor potencia estadística.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Ventajas particulares de los estudios longitudinales

Los diseños de **datos de panel** comparados con los diseños de series temporales:

- **Aprovechan la variabilidad transversal.** Si las variables incluídas no presentan una excesiva variabilidad temporal, pero sí transversal, el diseño de datos de panel permitirá estimaciones de los parámetros con mayor precisión (eficiencia).
- Respecto a los diseños transversales, estos diseños **aprovechan la variabilidad temporal.** Algunas variables pueden presentar variabilidad temporal, pero no transversal. Por lo tanto, su efecto solo se podría capturar mediante una dimensión temporal.



# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Ventajas particulares de los estudios longitudinales

Por último, y seguramente más importante, los diseños de **datos de panel** permiten incluir **factores no observables** para los cuáles no se dispone de información. Estos factores, llamados **heterogeneidad**, pueden ser de dos tipos:

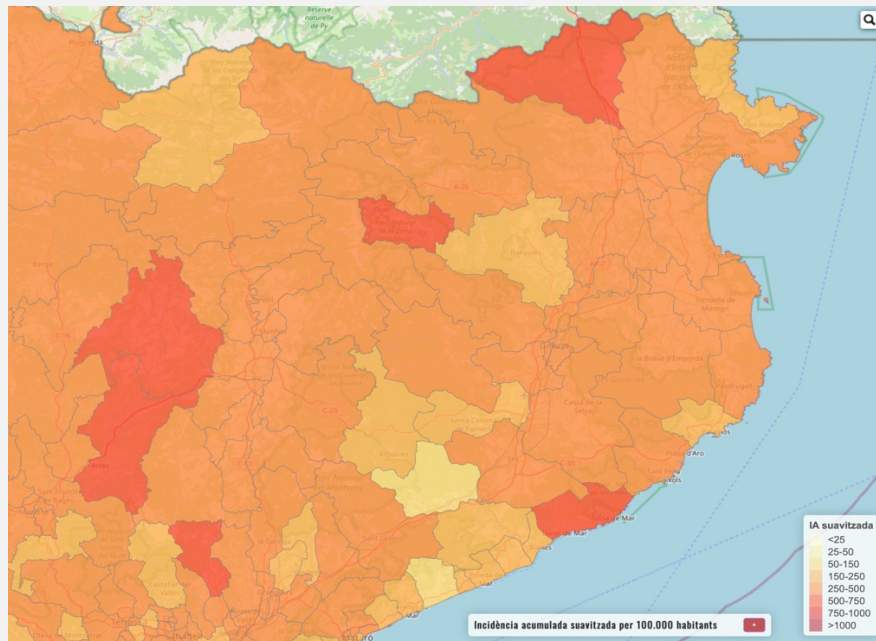
- **Heterogeneidad individual.** Específica a cada individuo o unidad de estudio y constante en el tiempo. Corresponde a la dimensión transversal.
- **Heterogeneidad temporal.** Común a todos los individuos o unidades de estudio y variante en el tiempo, pero constante para las observaciones transversales. Corresponde a la dimensión temporal.

## EJEMPLO PRÁCTICO

Con el propósito de entender mejor los dos tipos de heterogeneidad utilizaremos, como ejemplo, información sobre la pandemia de la COVID-19.

Mostramos, en primer lugar, la representación de la incidencia acumulada por 100.000 habitantes, en los últimos 14 días, en un mapa por municipios, centrado en la Región Sanitaria de Girona (la que coincide prácticamente con la provincia de Girona, a excepción de La Cerdanya, que pertenece a la Región Sanitaria del Alt Pirineu i Aran), en el período correspondiente a la tercera ola (diciembre 2020-febrero 2021).

## EJEMPLO PRÁCTICO



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

Observamos que durante la tercera ola la mayoría de municipios de la Región tenían **una incidencia entre 250 y 500 casos por 100.000 habitantes** en los últimos 14 días, pero también observamos **algunos entre 150 y 250** (Cadaqués, Llançà, Palamós, Santa Coloma de Farners, Arbúcies, Sant Hilari de Sacalm, Banyoles, Ribes de Freser y Campdevàrol) y **unos pocos entre 500 y 750** (Lloret de Mar, Tossa de Mar, Olot y una agrupación de municipios entorno a Agullana en el Alt Empordà).

## EJEMPLO PRÁCTICO

¿Como se explica, por ejemplo, que Banyoles, Besalú y Olot, siendo municipios vecinos, tengan incidencias tan diferentes? Banyoles en un rango 150-250, Besalú en un rango 250-500 y Olot en un rango 500-750.

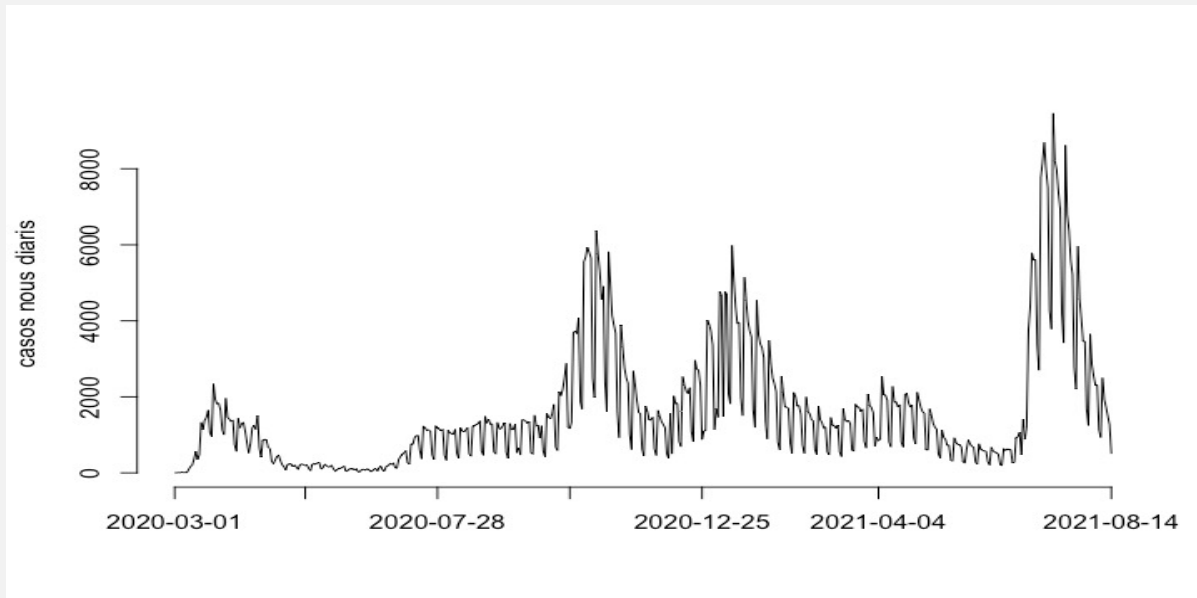
## EJEMPLO PRÁCTICO

¿Como se explica, por ejemplo, que Banyoles, Besalú y Olot, siendo municipios vecinos, tengan incidencias tan diferentes? Banyoles en un rango 150-250, Besalú en un rango 250-500 y Olot en un rango 500-750.

Estas diferencias y, en general, la distribución no homogenea de la incidencia en el territorio, podrían ser debidas a la existencia de variables no observadas, específicas a cada unidad transversal (municipio en este caso) y que podrían no variar en el tiempo. Esto es lo que se llama **heterogeneidad individual**.

## EJEMPLO PRÁCTICO

Enseñamos ahora la evolución temporal de los casos nuevos diarios en Cataluña, desde el 1 de marzo de 2020 hasta el 14 de agosto de 2021.



Fuente de datos: Registro de casos de COVID-19 realizados en Cataluña. Segregación por sexo y área básica de salud (ABS)  
(<https://analisi.transparenciacatalunya.cat/Salut/Registre-de-casos-de-COVID-19-realitzats-a-Catalun/xuwf-dxjd>)

## EJEMPLO PRÁCTICO

Se observan claramente las cinco olas. Como se ve, tanto la magnitud (altura de los picos) como la duración varían entre las diferentes olas. ¿Por qué?

## EJEMPLO PRÁCTICO

Se observan claramente las cinco olas. Tal y como se ve, tanto la magnitud (altura de los picos) como la duración varían entre las diferentes olas. ¿Por qué?

Estas diferencias podrían ser debidas a variables no observadas, específicas a cada unidad temporal (las olas, en este caso) y, presumiblemente, comunes a todas las unidades transversales (ABS en este caso). Estas diferencias reciben el nombre de **heterogeneidad temporal**.



# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Especificación del modelo de datos de panel (o modelo mixto longitudinal)

Empezando con una respuesta lineal (o normal), un modelo de datos de panel puede especificarse de la siguiente manera:

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

donde el subíndice  $i = 1, \dots, N$  denota las unidades individuales ( $N$  unidades individuales) y el subíndice  $t = 1, \dots, t_i$  ( $T = \max(t_i)$ , número de períodos) denota las diferentes observaciones repetidas para cada unidad individual.

Por ejemplo,  $i$ =cada paciente y  $t$ =tres observaciones de los pacientes en el tiempo.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

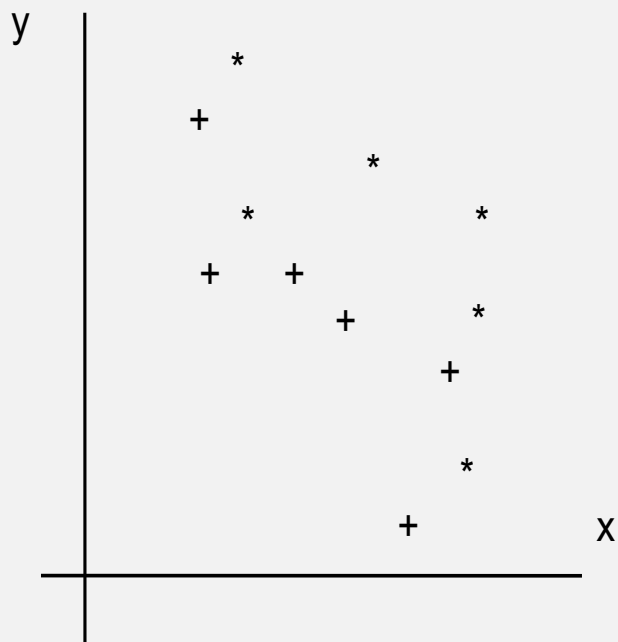
## Especificación del modelo de datos de panel (o modelo mixto longitudinal)

**Ejemplo:** De forma simplificada, supongamos que queremos estimar el efecto de la edad ( $x$ ) en la variación de la comprensión lectora ( $y$ ) de seis niños menores de 6 años,  $i = 1, 2, 3, 4, 5, 6$  ( $N = 6$ ). Los niños se observan en dos momentos del tiempo, dos períodos  $t = 1 (+), 2(*)$  ( $T = 2$ ) (modelo equilibrado). Decimos que el modelo está equilibrado porque se dispone de medidas de la capacidad lectora para todos los niños y para todos los períodos. En este caso, la variable explicativa es tiempo dependiente, ya que la edad de los niños es diferente en los dos momentos del tiempo en el que son observados.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

**Ejemplo:** En este caso, la especificación de nuestro modelo será:

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



# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Especificación del modelo de datos de panel (o modelo mixto longitudinal)

Según el objetivo propuesto, los datos de panel pueden aproximarse marginal o condicionalmente.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

Especificación del modelo de datos de panel (o modelo mixto longitudinal).

## Enfoque marginal

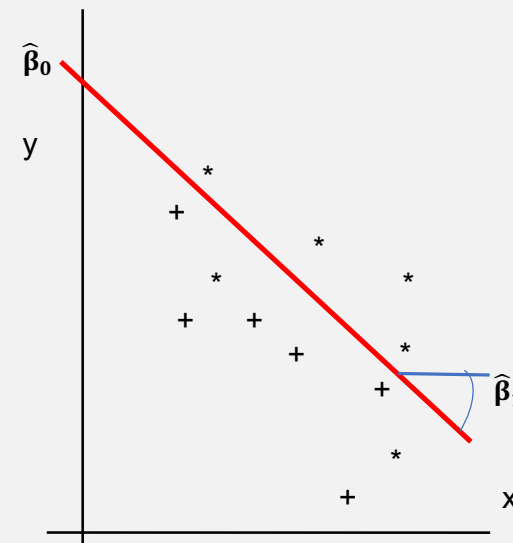
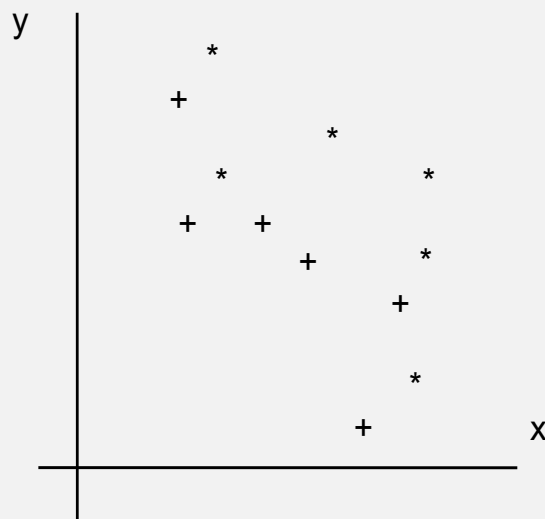
Utilizamos **el enfoque marginal** cuando queremos hacer **inferencias “poblacionales”**, es decir, si lo que queremos es explicar la relación entre la variable dependiente y las variables explicativas con independencia de la variabilidad intraindividual.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

Especificación del modelo de datos de panel (o modelo mixto longitudinal).

Enfoque marginal

En nuestro ejemplo:



# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

**Especificación del modelo de datos de panel (o modelo mixto longitudinal).**

**Enfoque marginal**

La ordenada en el origen  $\hat{\beta}_0$  y el coeficiente asociado a la variable explicativa  $\hat{\beta}_1$  son comunes a todos los individuos. No existe heterogeneidad individual. Dicho de otra manera, todos los efectos (de las variables explicativas, incluida la ordenada en el origen) son fijos.

**Observamos que, en promedio, a más edad del niño, menos variación en la comprensión lectora.**

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

Especificación del modelo de datos de panel (o modelo mixto longitudinal).

## Enfoque marginal

En el **enfoque marginal**, se trata de estimar los parámetros correspondientes a la media,  $\beta$ . Muy raramente son de interés los parámetros de la covarianza. De hecho, estos suelen ser tractados como una molestia en el enfoque marginal, controlados, pero no estimados.



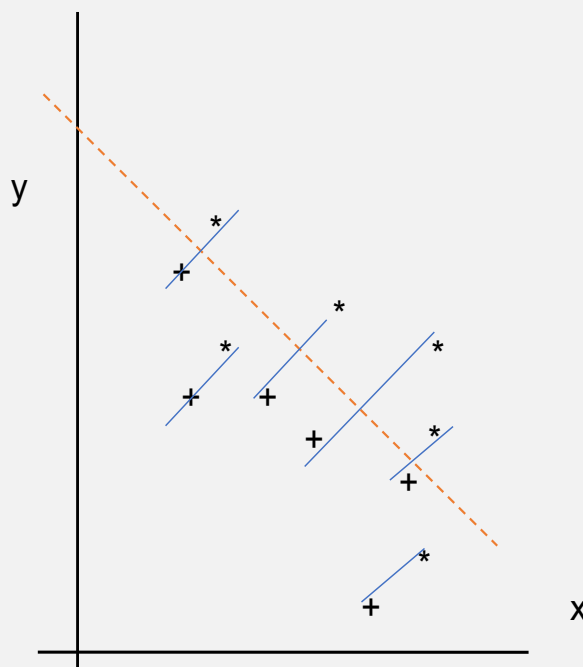
# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

Especificación del modelo de datos de panel (o modelo mixto longitudinal).

## Enfoque condicional

En el enfoque condicional, queremos hacer inferencias “individuales”.

En nuestro ejemplo:



# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

**Especificación del modelo de datos de panel (o modelo mixto longitudinal).**

**Enfoque condicional**

En este caso, el coeficiente asociado a la variable explicativa  $\hat{\beta}_1$  es aproximadamente igual para todos los individuos. Pero, la ordenada en el origen  $\hat{\beta}_0$  es diferente para cada uno de ellos. Existe heterogeneidad individual.

**Con independencia del comportamiento promedio, para cada niño en particular, a más edad, más variación de la comprensión lectora.**

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

Especificación del modelo de datos de panel (o modelo mixto longitudinal).

## Enfoque condicional

En la **aproximación condicional**, se modeliza simultáneamente la media de la variable dependiente (la variabilidad inter-individual) y la estructura de covarianzas o correlaciones (la variabilidad intra-individual). En este enfoque, los parámetros que definen la correlación tienen el mismo interés que los correspondiente a la media, a veces incluso más.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Especificación del modelo de datos de panel (o modelo mixto longitudinal)

Fijaros que en el enfoque marginal, interesa el promedio (o la inferencia poblacional). En el enfoque condicional, interesa cada niño (o la inferencia individual).

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

Supongamos que especificamos el siguiente modelo:

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

donde  $a_i$  captura **la heterogeneidad individual no observable** (factores específicos a cada unidad individual y que no varían en el tiempo); y  $\tau_t$  captura **la heterogeneidad temporal**, común a todas las unidades individuales.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Modelo pooled

La especificación más sencilla del modelo, llamada **modelo pooled**, consiste en ignorar la heterogeneidad ( $a_i = a$  y  $\tau_t = 0$ ):

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

y estimar el modelo por Mínimos Cuadrados Ordinarios (MCO). Este método de estimación se conoce como **Pooled Least Squares** (PLS).

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Modelo pooled. Problemas de ignorar la heterogeneidad en un diseño mixto

Es decir, aun cuando tenemos un diseño mixto, omitimos variables relevantes (la heterogeneidad).

Esto significa que se comete un error de especificación. Por lo tanto, **la estimación PLS de los parámetros estará sesgada y sus varianzas estarán mal calculadas.**

### Posible solución:

Estimar el modelo mediante métodos de estimación alternativos ya sea mediante la aproximación marginal o bien mediante la aproximación condicional. Nosotros utilizaremos el enfoque condicional.

[3. Panorámica de los modelos mixtos](#)

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

El modelo de **efectos aleatorios** es el más conocido de entre los condicionales. Supone que la dependencia que puede existir entre las observaciones de la variable dependiente (és decir, entre las respuestas repetidas de un mismo individuo), se debe a que los coeficientes de la regresión (de la media) no son los mismos para todos los individuos.

De hecho, la interpretación más sencilla, es la que supone que la heterogeneidad individual se debe a factores no observables (o variables omitidas) fijas en el tiempo, pero variables entre individuos. Estos factores son los que provocan la dependencia entre las diferentes observaciones de la variable dependiente.

3. Panorámica de los modelos mixtos



# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

Por lo tanto, este modelo **supone que los coeficientes de la regresión no son los mismos para todos los individuos:**

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

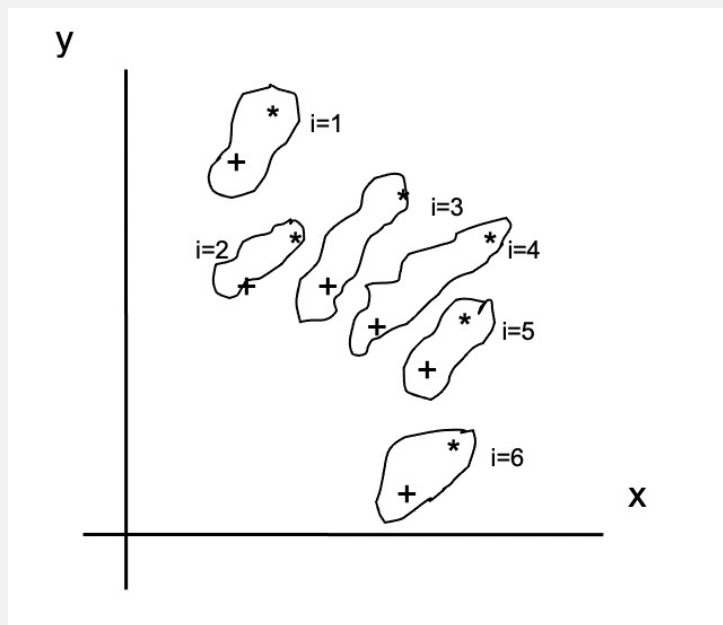


Pueden variar entre individuos

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

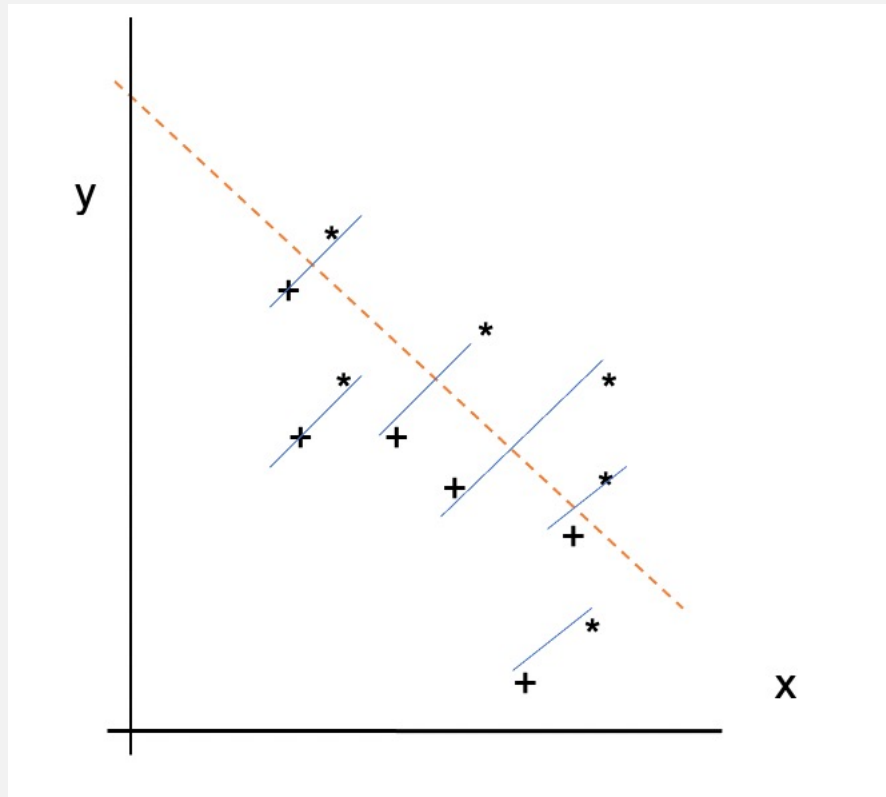
## Enfoque condicional. Modelo de efectos aleatorios

Supongamos el ejemplo que hemos visto antes, en el que consideramos 6 niños y dos períodos (representamos el primer período mediante el símbolo + y el segundo mediante el símbolo \*):



# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios



En este caso, el coeficiente asociado a la variable explicativa es aproximadamente el mismo para todos los niños. Es decir, la comprensión lectora varía más o menos de la misma manera para todos los niños.

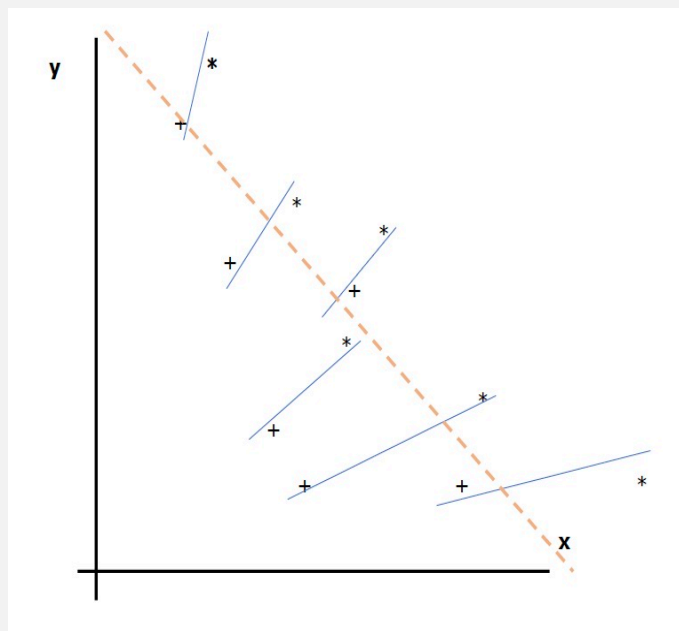
Por el contrario, la ordenada en el origen es diferente para cada uno de ellos. La comprensión lectora basal es diferente.

**Existe heterogeneidad individual!!!**

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

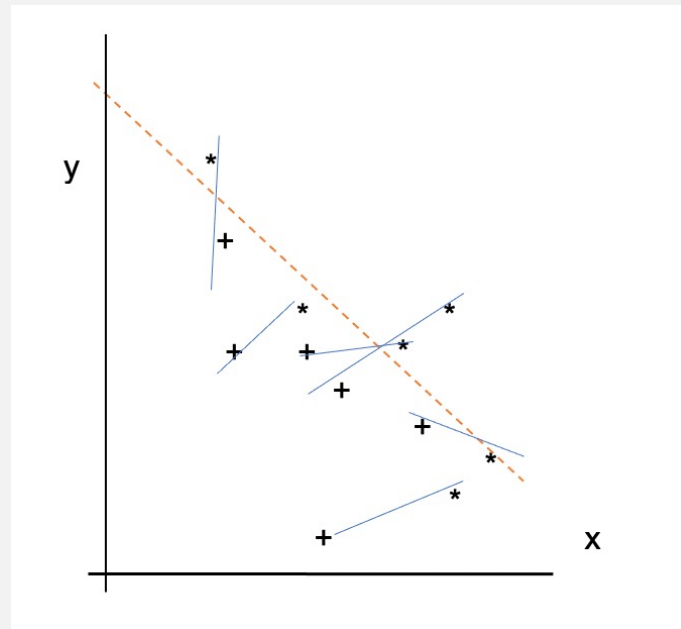
En el siguiente ejemplo, el parámetro asociado a la variable explicativa y no el asociado a la ordenada en el origen, es el que es el efecto aleatorio.



# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

También podrían darse las dos situaciones. Es decir, que tanto la ordenada en el origen como la pendiente sean efectos aleatorios.



# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

- En Econometría, hablamos de **modelo de efectos aleatorios**, nos referimos precisamente al primer modelo. Es decir, sólo existe un efecto aleatorio que se corresponde con la ordenada en el origen (es decir, heterogeneidad individual no observada).

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



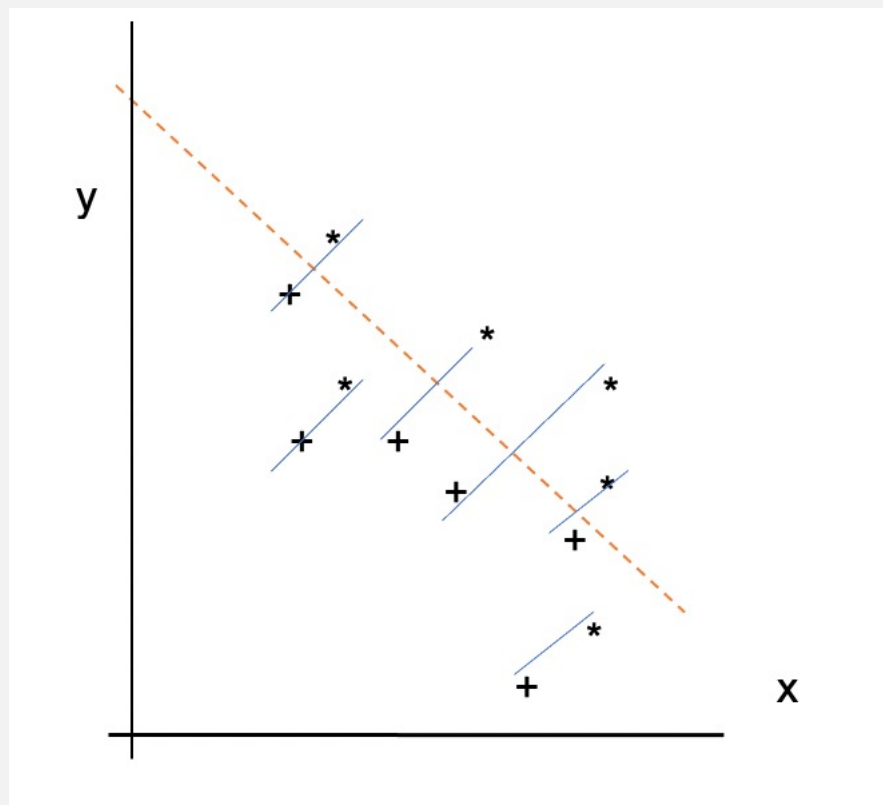
- En cambio, en el segundo modelo, estaríamos hablando de un **modelo de coeficientes aleatorios**. En este caso el efecto aleatorio estaría asociado a una variable explicativa.

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



# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios



# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

En resumen y siguiendo con nuestro ejemplo, en el que el efecto aleatorio está en la ordenada en el origen pero no en la pendiente:

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



# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

El problema consiste en que cada individuo suele tener muy pocas observaciones para estimar los  $\beta$  a partir únicamente de  $(y_{it}, x_{it})$ . Por ejemplo, en nuestro caso tenemos 7 parámetros desconocidos ( $6 \beta_{oi}, i = 1, \dots, 6$  i  $1 \beta_1$ ) con 12 observaciones (6 niños x 2 períodos). **ESTIMACIONES INEFICIENTES!!!**

Deberemos estimar el modelo mediante un **modelo lineal generalizado mixto (GLMM)**.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

Deberemos suponer que los  $\beta$  no son más que realizaciones independientes de alguna distribución de probabilidad (la normal en modelos lineales) con media  $\beta$ :

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

Resumiendo, hemos especificado un modelo mixto con un efecto fijo ( $\beta_1$ ) y un efecto aleatorio ( $\beta_{0i}$ ), lo que significa que  $\beta_{0i}$  puede variar entre individuos ( $i = 1, 2, \dots, 6$ ).

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

Observad que, en el modelo de efectos aleatorios, dentro del enfoque condicional, también estimaremos los parámetros poblacionales,  $\beta_0$  i  $\beta_1$ , correspondientes al enfoque marginal.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

Ahora, además de suponer:

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

Deberemos suponer también:

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

Las dos últimas hipótesis son particularmente importantes.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

- En primer lugar,  $E(\eta_i, \eta_j) = 0 \forall i \neq j$ , significa que los **niveles superiores deben ser independientes entre sí**. El incumplimiento de esta hipótesis (llamada efectos cruzados o crossover) exige la utilización de métodos de estimación mucho más complejos que los explicados aquí (por ejemplo, métodos Bayesianos).
- En segundo lugar,  $E(u_{it}, \eta_j) = 0 \forall i, t, j$ , exige que **los efectos aleatorios sean independientes de los efectos fijos**. De lo contrario, los efectos deben ser efectivamente completamente aleatorios.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

Ilustraremos ahora las dos últimas hipótesis en nuestro ejemplo.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

- En primer lugar,  $E(\eta_i, \eta_j) = 0 \forall i \neq j$ , significa que los niveles superiores deben ser independientes entre sí.

Siguiendo con nuestro ejemplo, los niños son el nivel superior, mientras que el tiempo constituye el nivel inferior, ya que cada niño “contiene” (de hecho, se observa) dos períodos de tiempo. Por lo tanto, la variación en la comprensión lectora de un niño no depende de la variación en la comprensión lectora de cualquier otro. Es decir, los niños son independientes entre ellos, por lo que se cumple esta hipótesis.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

Ahora bien, podría ocurrir que los niños pertenecieran a la misma familia (por ejemplo, supongamos que los niños 2 y 5 son hermanos). En este caso, la variación de la comprensión lectora del niño 2 y del niño 5 no serían independientes, incumpléndose la hipótesis.



# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

Pero, para que esto se cumpla, basta con re-especificar el modelo añadiendo un nivel adicional:

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

donde  $k$  ( $k = 1$  niños 2 y 5;  $k = 2$  niño 3;  $k = 3$  niño 4;  $k = 4$  niño 6) denota la familia,  $i$  ( $i = 1, 2, \dots, 6$ ) denota el niño y  $t$  ( $t = +, *$ ) denota el período. Ahora, las unidades del nivel superior (familias) sí que son independientes, cumpliéndose la hipótesis.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

- En segundo lugar,  $E(u_{it}, \eta_j) = 0 \forall i, t, j$ , exige que los efectos aleatorios sean independientes de los efectos fijos.

En nuestro ejemplo, supongamos que en el modelo que estamos utilizando, las diferencias en las ordenadas en el origen entre los niños se expliquen por alguna variable, por ejemplo, el sexo del niño.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

Es decir:

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

Así, aislando:

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

Y, finalmente, arreglando términos:

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

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

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios

Evidentemente,  $v_{it}$  estará correlacionado con  $u_{it}$ , incumpléndose la hipótesis que **los efectos aleatorios sean independientes de los efectos fijos** y los estimadores serán inconsistentes.

Para que vuelva a cumplirse, deben incluirse todas las variables explicativas en la ecuación principal (la de la media) y dejar el efecto completamente aleatorio. Es decir:

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

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios. Estimación del modelo

Si se cumplen las dos hipótesis, los modelos **se estiman por métodos basados en la máxima verosimilitud**. El problema es que sólo existe una función de verosimilitud completa cuando no existe correlación, existan o no efectos aleatorios. Pero, se puede definir la cuasi-verosimilitud, que tiene exactamente la misma forma que la derivada parcial de la verosimilitud completa.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Enfoque condicional. Modelo de efectos aleatorios. Estimación del modelo

La maximización de la cuasi-verosimilitud es muy compleja. Por otro lado, los métodos Bayesianos, a los que podemos recurrir, implican la evaluación total de la cuasi-verosimilitud, por lo que suelen ser muy intensivos computacionalmente.

Alternativamente, se utiliza la pseudo-verosimilitud o pseudo-verosimilitud penalizada, PQL, la verosimilitud restringida (REML) o el método de Laplace.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

## Resumiendo:

- El estimador de efectos fijos (aproximación marginal) permite estimar el modelo bajo supuestos menos restrictivos que el estimador de efectos aleatorios.
- Aún así, si se cumplen las condiciones de regularidad, el estimador de efectos aleatorios es más eficiente que el de efectos fijos. Pero puede ser inconsistente si existe correlación entre las variables explicativas y la heterogeneidad individual.
- Por métodos Bayesianos, los estimadores de efectos aleatorios siempre son consistentes.

# MODELOS MIXTOS – DISEÑOS LONGITUDINALES

Recordad que las condiciones de regularidad a las que nos referimos en el segundo punto son:

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



