

# **Análisis de Datos Longitudinales**

## **Maestría en Demografía y Estudios de Población**

Alejandra Marroig

Instituto de Estadística, Universidad de la República, Uruguay

August 2025

# Modalidad y contenidos del curso

## Modalidad:

- 15 horas de trabajo presencial  $\Rightarrow$  sesiones expositivas y taller
- 20 horas de trabajo domiciliario

## Contenido:

- 1 Contexto de investigaciones del cambio en el tiempo
- 2 Análisis exploratorio del cambio con datos longitudinales
- 3 Análisis de datos con el modelo multinivel
- 4 Aspectos avanzados de análisis de datos longitudinales

# Sistema de evaluación

Para tener derecho a examen se debe asistir al curso y hacer los trabajos

**Asistencia mínima:** 75% de las clases (3 clases y media)

## Evaluación:

- 1 Trabajos teórico prácticos en clase 30%
- 2 Trabajo domiciliario a entregar 70%

Fechas de examen:

- 1 29 de septiembre 2025
- 2 17 de diciembre 2025
- 3 2026 día a confirmar

# Bibliografía

Singer, J. D., & Willett, J. B. (2003). Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Parte I. Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780195152968.001.0001>

Ejemplos del libro: <https://stats.oarc.ucla.edu/other/examples/alda/>

## Opcional:

Hoffman, L. (2015). Longitudinal Analysis: Modeling Within-Person Fluctuation and Change (1.<sup>a</sup> ed.). Taylor & Francis.  
<https://www.routledge.com/Longitudinal-Analysis-Modeling-Within-Person-Fluctuation-and-Change/Hoffman/p/book/9780415876025>

# Presentación

**Presentación estudiantes:** nombre y 3 palabras

# Contenidos

## En esta clase veremos

- 1 **Contexto de investigaciones del cambio en el tiempo**
- 2 **Análisis exploratorio del cambio con datos longitudinales**
- 3 Análisis de datos con el modelo multinivel
- 4 Aspectos avanzados de análisis de datos longitudinales (no linealidades, flexibilización de la variable tiempo, otros)

# Contexto del análisis del cambio en el tiempo

Es muy usual que se investigue el cambio en el tiempo:

- Desarrollo infantil
- Envejecimiento
- Enfermedades
- etc.

Luego de una *intervención* también suele ser de interés investigar el cambio:

- Evolución resultados de exámenes luego de un entrenamiento
- Reducción del colesterol luego de cambio en medicación
- Evolución de células CD4 luego del diagnóstico de VIH

# ¿Cambios en el tiempo en qué disciplinas?

Los cambios pueden ser de diverso tipo:

- 1 Biológicos: desarrollo físico, menopausia, ...
- 2 Sociales/Económicos: índice de vulnerabilidad social, ingreso, años de educación, matrimonio, hijas/os, ...
- 3 Psicológicos: cognitivos, no cognitivos, *flourishing*, ...
- 4 etc.

El cambio se puede producir a nivel POBLACIONAL o INDIVIDUAL ⇒ NO SON LO MISMO!!!

**Trayectoria** o evento.

# Desarrollos metodológico para estudiar el cambio

**Laird & Waird (1982)**. Random-Effects Models for Longitudinal Data, *Biometrika*, <https://doi.org/10.2307/2529876>

Desarrollaron métodos estadísticos para este análisis, conocidos como:

- *individual growth models*
- *random coefficient models*
- *multilevel models*
- *mixed models*
- *hierarchical models*

# 40 años después. . .

Teniendo:

- 1 el marco metodológico de los modelos mixtos
- 2 datos longitudinales
- 3 software (R, STATA, MPlus, etc.)

se puede realizar investigación cuantitativa, específicamente en ciencia sociales, interesada en el estudio del cambio en el tiempo.

# Situaciones donde se podría estudiar el cambio

- Diseños de investigación **observacionales** o **experimentales**.
- Datos recolectados de forma **prospectiva** o **retrospectiva**.
- El tiempo se puede medir en diferentes **unidades** (días, meses, semestres, años, sesiones, etc.).
- Datos recogidos en un esquema **fijo** (todas las personas en la misma periodicidad) o **flexible** (cada persona tiene una periodicidad única)

# Dos tipos de preguntas sobre el cambio

- 1 ¿Cómo evoluciona el *outcome* en el tiempo?

*Patrón del cambio, within-individual change*

- 2 ¿Existen diferencias en estos cambios según características de los individuos?

*Estudio relacional, de asociaciones, interindividual differences in change, personas diferentes tendrán patrones distintos del cambio*

Un modelo que permite analizar ambas cosas es un modelo multinivel para analizar el cambio (***multilevel model for change***)

# Diseño de investigación para el cambio en el tiempo

Aspectos clave de un diseño de investigación para el cambio en el tiempo:

- 1 Múltiples ocasiones de medición, tres o más
- 2 Una medida sensible del tiempo
- 3 Un *outcome* cuyos valores cambian sistemáticamente en el tiempo

# Múltiples ocasiones de medición de los datos

- En términos estadísticos, los estudios *cross-sectional* confunden los efectos de la edad y de cohorte (y los efectos de la edad y de período) y son propensos al sesgo.
- Contar con dos ocasiones permite medir solamente el incremento (decremento) del *outcome*. **NO** dice sobre la forma del proceso de cambio y no permite distinguir el efecto de los errores de medida.
- ¿Cuántas olas se precisan para un estudio del cambio en el tiempo? Mínimo tres. Cuántas más mejor. Es “caro”. Es una pregunta difícil que se responde en cada situación concreta.

# Una medida sensible del tiempo

- El **tiempo** es el predictor fundamental en los análisis sobre el cambio.
- **¿Cuál es esa métrica del tiempo?** La respuesta puede no ser única. Adoptar métrica con más sentido para la pregunta de investigación y que sea más útil para el análisis del cambio en el *outcome*.
- **Restricción:** el cambio en esa métrica debe ser **monótono**.
- **Espacio entre olas:** puede ser cualquiera, depende del estudio específico.
- **Datos balanceados o no balanceados:** la cantidad de olas de cada individuo puede ser diferente (*time-unstructured*).

# *Outcome* con cambio sistemático en el tiempo

Las puntuaciones del *outcome* en el tiempo deben ser

- comparables
- válidas (medir el atributo que realmente pretende medir)
- preservar la precisión (minimizar errores de medición)

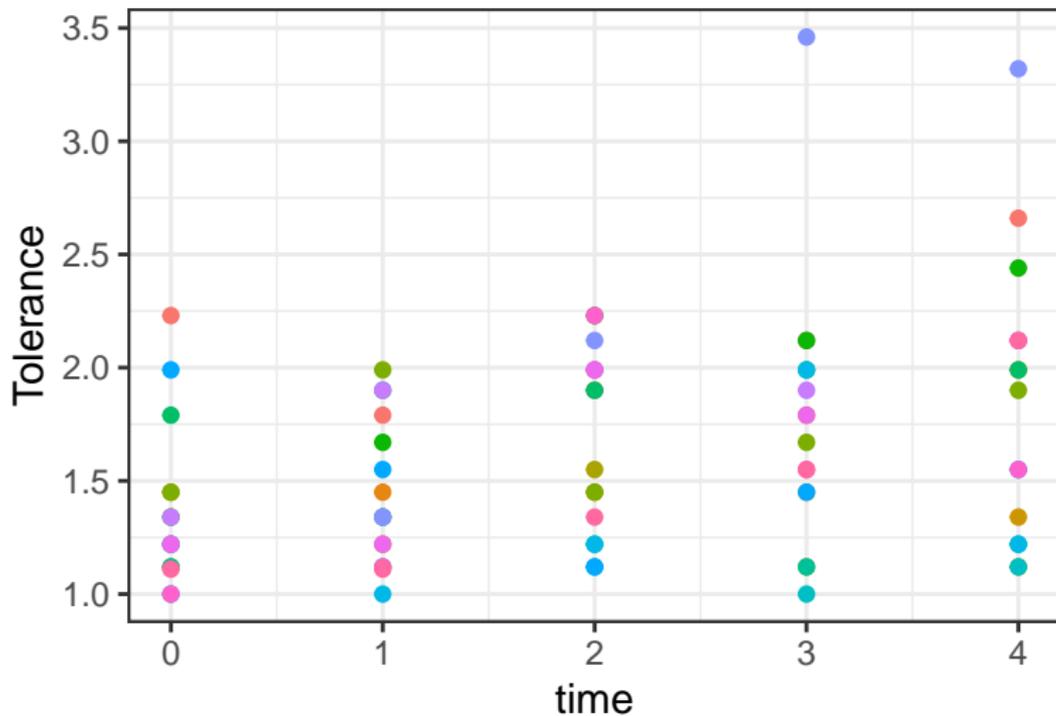
# ¿Y los datos?

# ¿Y los datos?

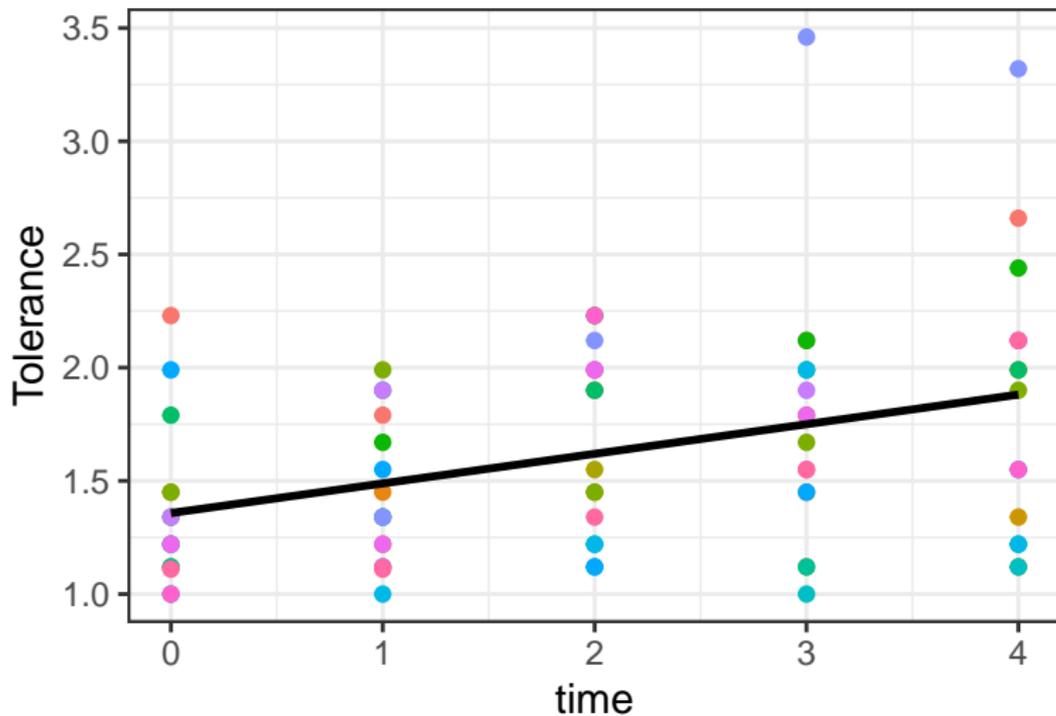
- 16 niñas/os de la Encuesta Nacional sobre la Juventud (NYS).
- Edades: 11, 12, 13, 14 y 15 años
- *Outcome*: tolerance

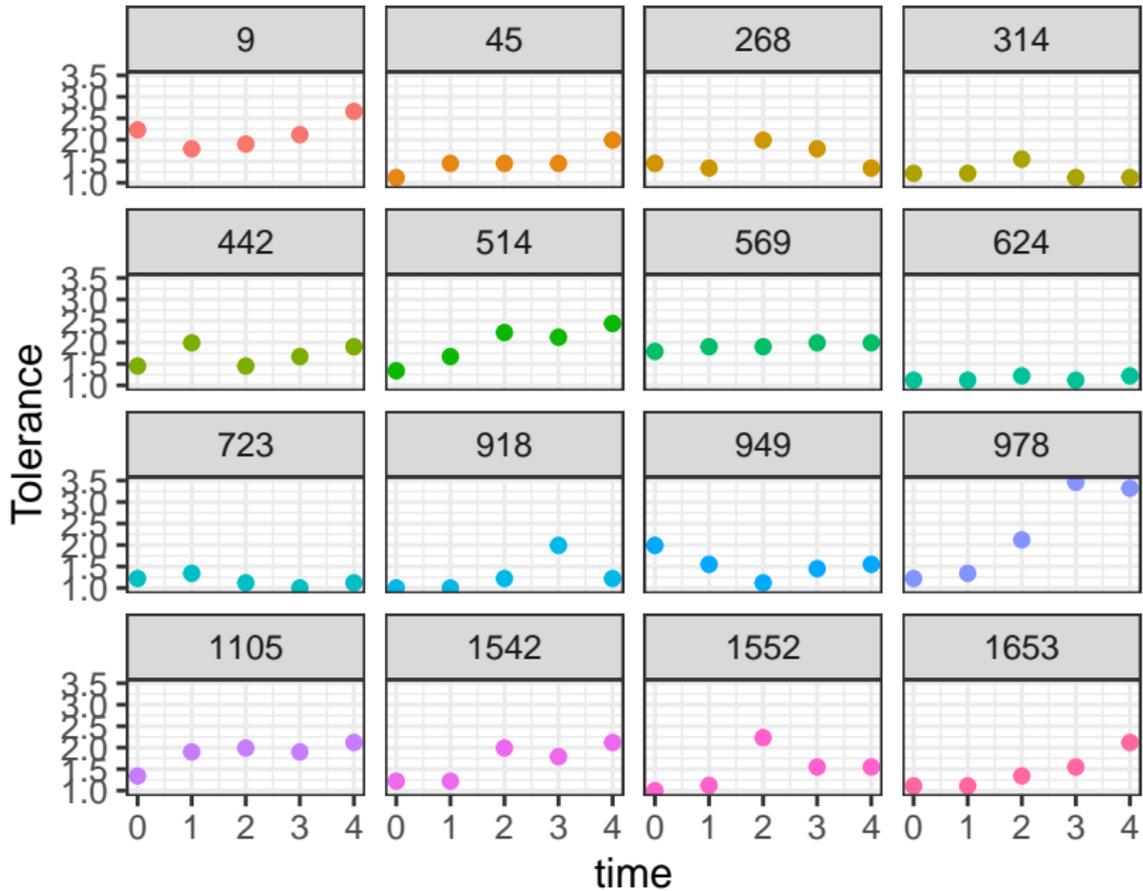
```
##      id age tolerance male exposure time idf
## 1    9  11      2.23    0      1.54    0    9
## 2    9  12      1.79    0      1.54    1    9
## 3    9  13      1.90    0      1.54    2    9
## 4    9  14      2.12    0      1.54    3    9
## 5    9  15      2.66    0      1.54    4    9
## 6   45  11      1.12    1      1.16    0   45
## 7   45  12      1.45    1      1.16    1   45
## 8   45  13      1.45    1      1.16    2   45
## 9   45  14      1.45    1      1.16    3   45
## 10  45  15      1.99    1      1.16    4   45
```

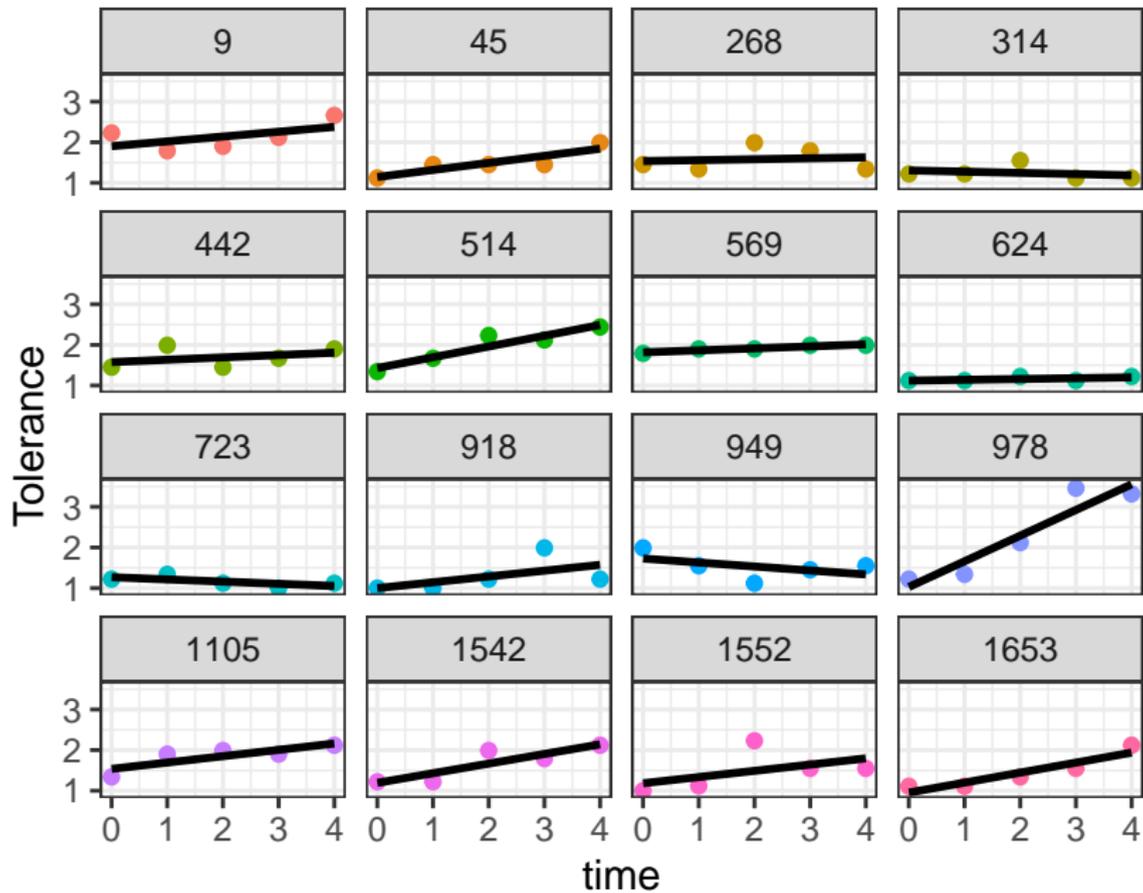
## Trayectoria de tolerancia en jóvenes



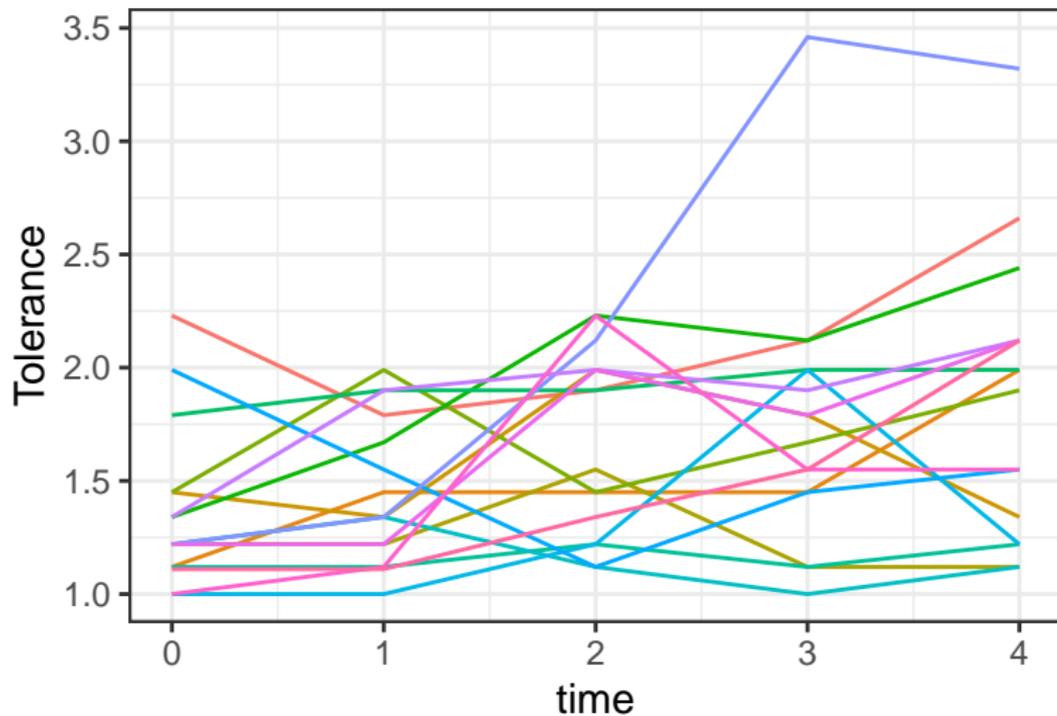
## Trayectoria de tolerancia en jóvenes



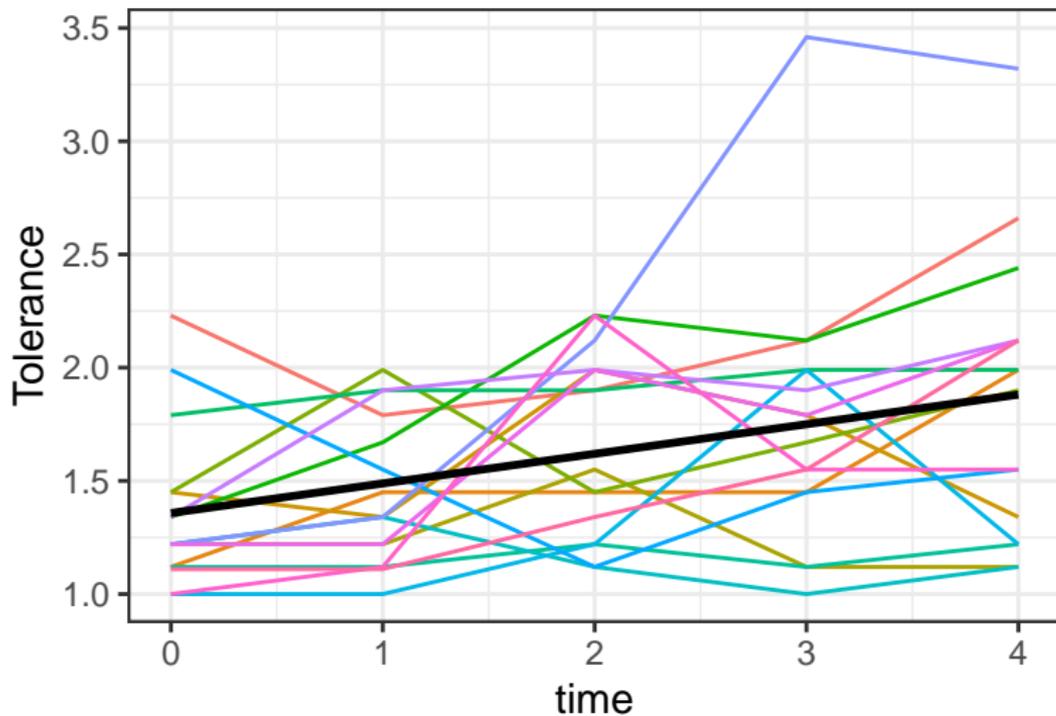




## Spaghetti plot



## Spaghetti plot



# Datos reales de la ENDIS

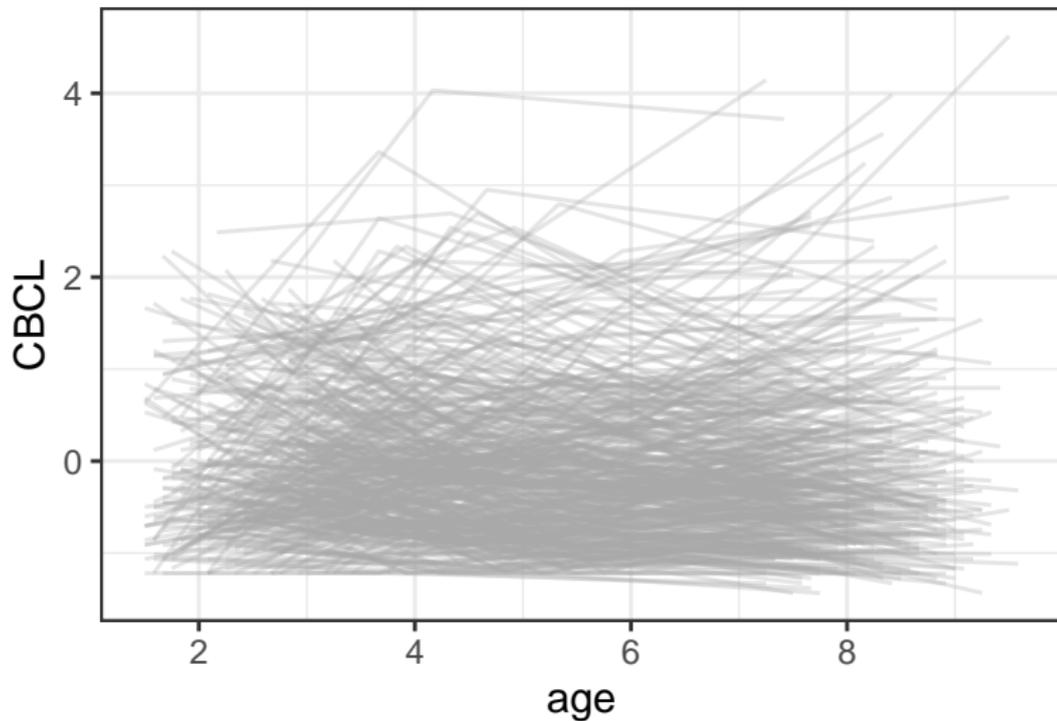
```
head(endis)
```

```
## # A tibble: 6 x 8
##   idnum   ola   age  sexo  tering  dtercil  cuidapa  zptcbcl
##   <dbl> <dbl> <dbl> <dbl> <dbl>   <dbl>   <dbl>   <dbl>
## 1     1     1   2.5     1     3       0         1  -0.139
## 2     1     2   4.67    1     3       0         1  -1.12
## 3     1     3   8.33    1     3       0         1  -1.12
## 4     2     1   2.67    0     2       0         0   0.170
## 5     2     2   5.17    0     2       0         0   0.119
## 6     2     3   8.25    0     2       0         0   2.34
```

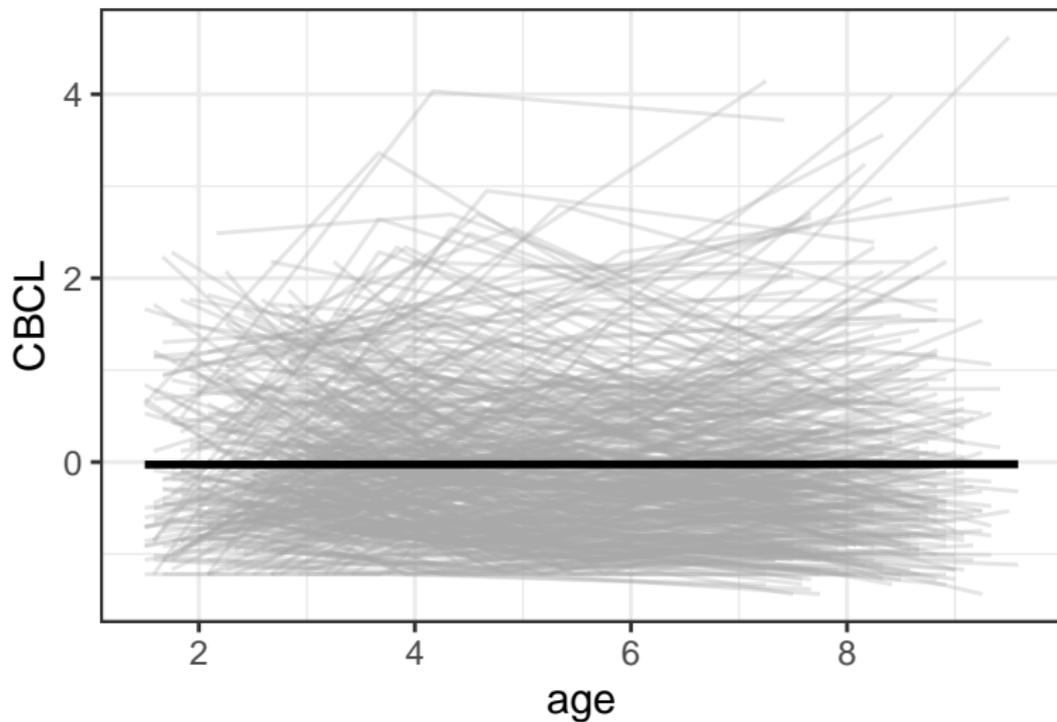
```
length(unique(endis$idnum))
```

```
## [1] 408
```

## Spaghetti plot



## Spaghetti plot



# Datos reales Satisfacción con la vida - SHARE

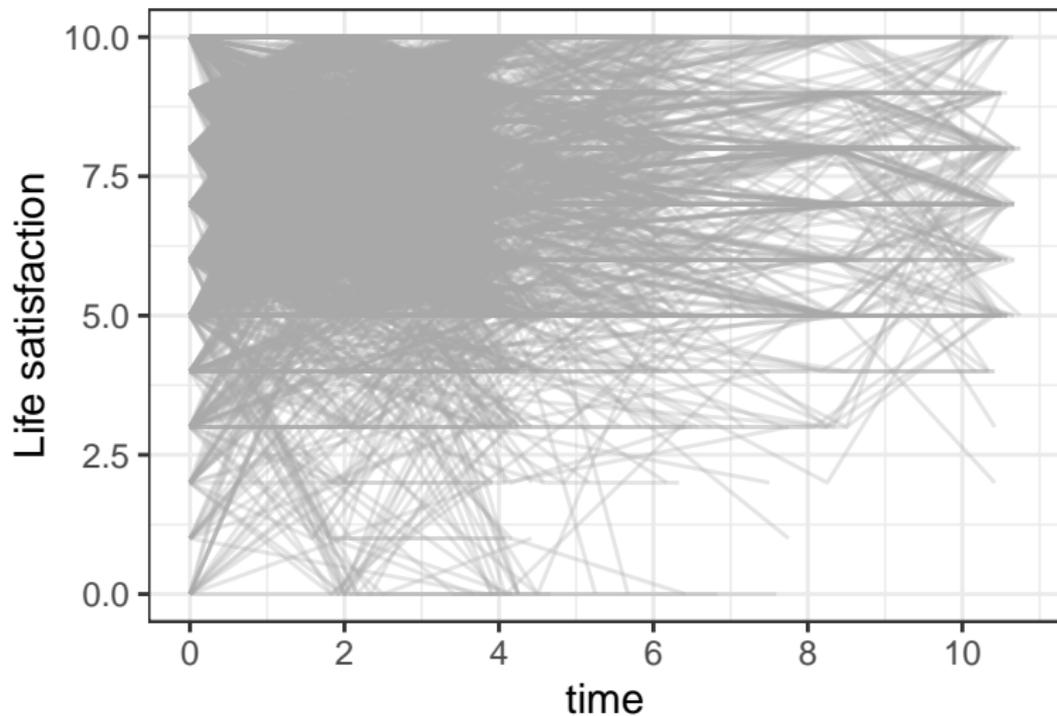
```
head(lifesat[,c("idnum", "lifesatis", "time", "caregiver", "bage")])
```

```
## # A tibble: 6 x 6
##   idnum lifesatis  time caregiver  bage female
##   <dbl>    <dbl> <dbl>    <dbl> <dbl> <dbl>
## 1     46         5  0         1  64.8     1
## 2     46        10  1.67       0  64.8     1
## 3     46         3  5.83       0  64.8     1
## 4     51         4  0         1  62.1     1
## 5     51         7  1.92       0  62.1     1
## 6     51         6  3.75       0  62.1     1
```

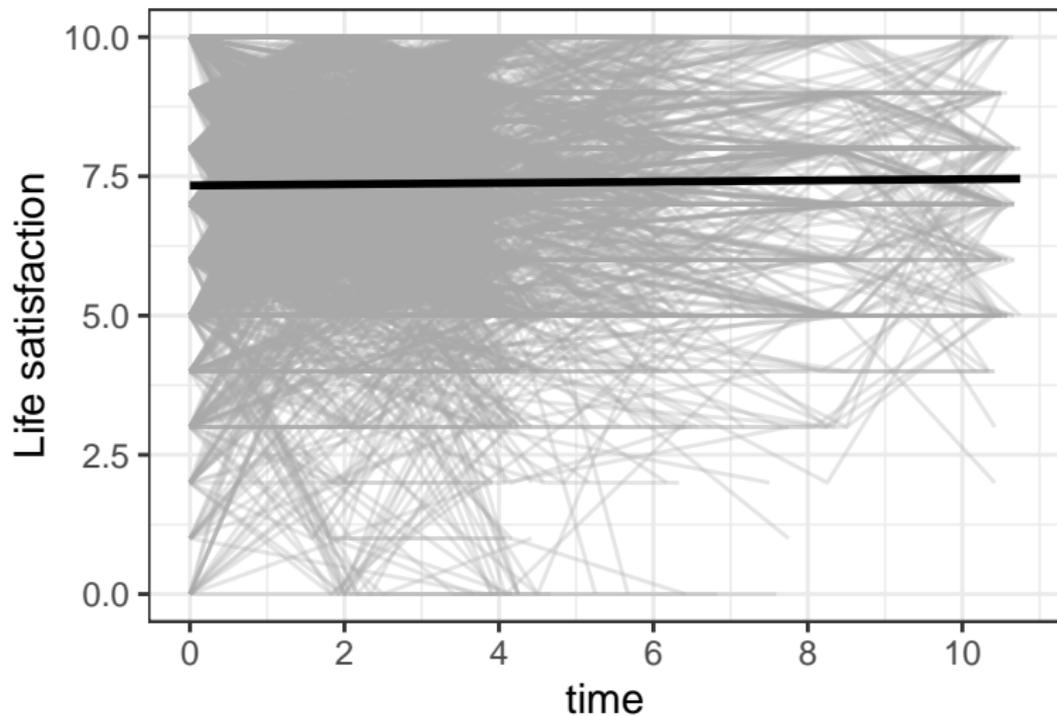
```
length(unique(lifesat$idnum))
```

```
## [1] 2395
```

## Spaghetti plot



## Spaghetti plot



# Datos reales de índice de fragilidad - SHARE

```
head(frailty[,c("idnum", "country", "age", "migrant", "female", "educ")])
```

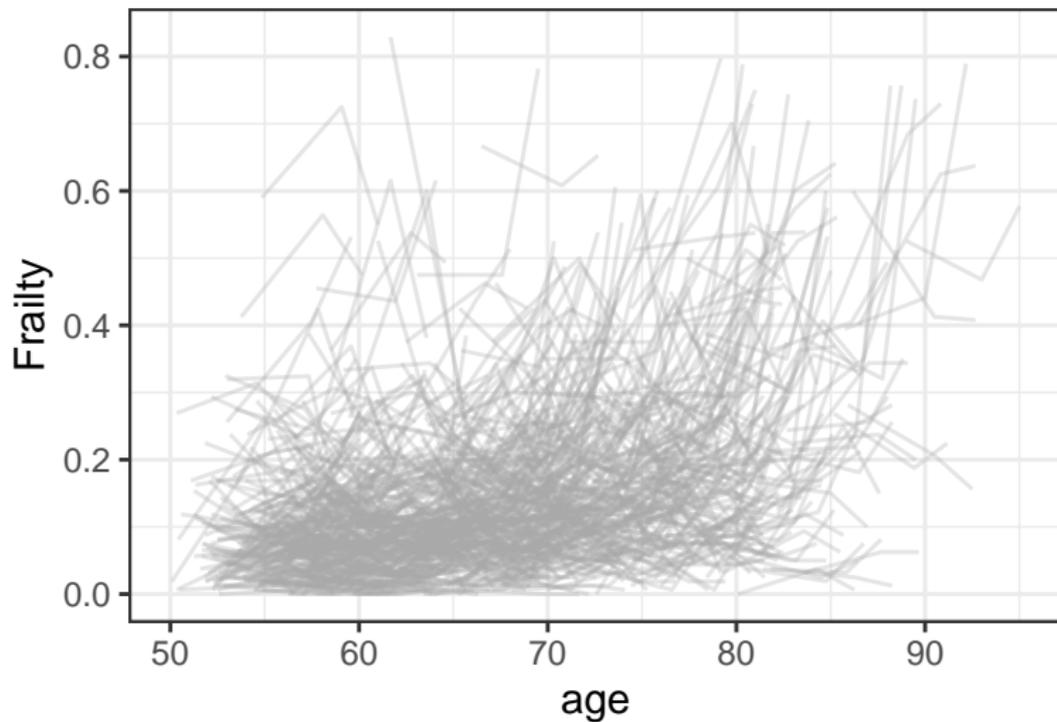
```
## # A tibble: 6 x 7
```

```
##   idnum country   age migrant female educ frailty
##   <dbl> <chr>    <dbl>  <dbl>  <dbl> <dbl>  <dbl>
## 1  90648 Croatia  56.8      0      0     12 0.0125
## 2  90648 Croatia  61.2      0      0     12 0.00625
## 3  90648 Croatia  63.2      0      0     12 0.00625
## 4  90649 Croatia  57.4      0      1     11 0.0500
## 5  90649 Croatia  61.8      0      1     11 0.1
## 6  90649 Croatia  63.8      0      1     11 0.0500
```

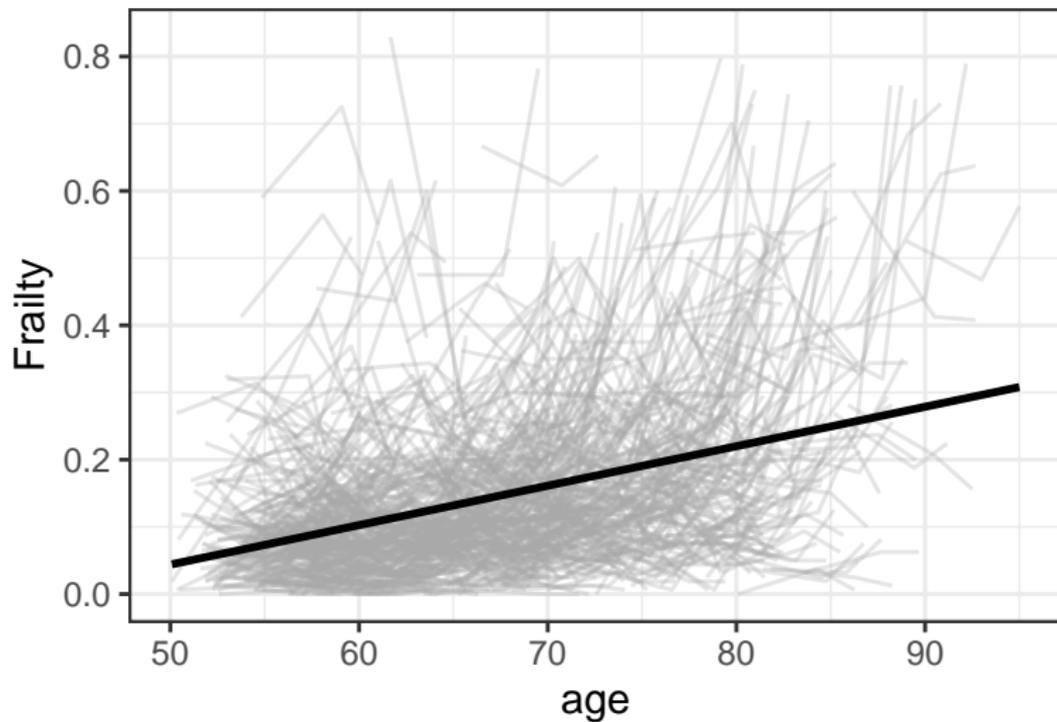
```
length(unique(frailty$idnum))
```

```
## [1] 911
```

## Spaghetti plot



## Spaghetti plot



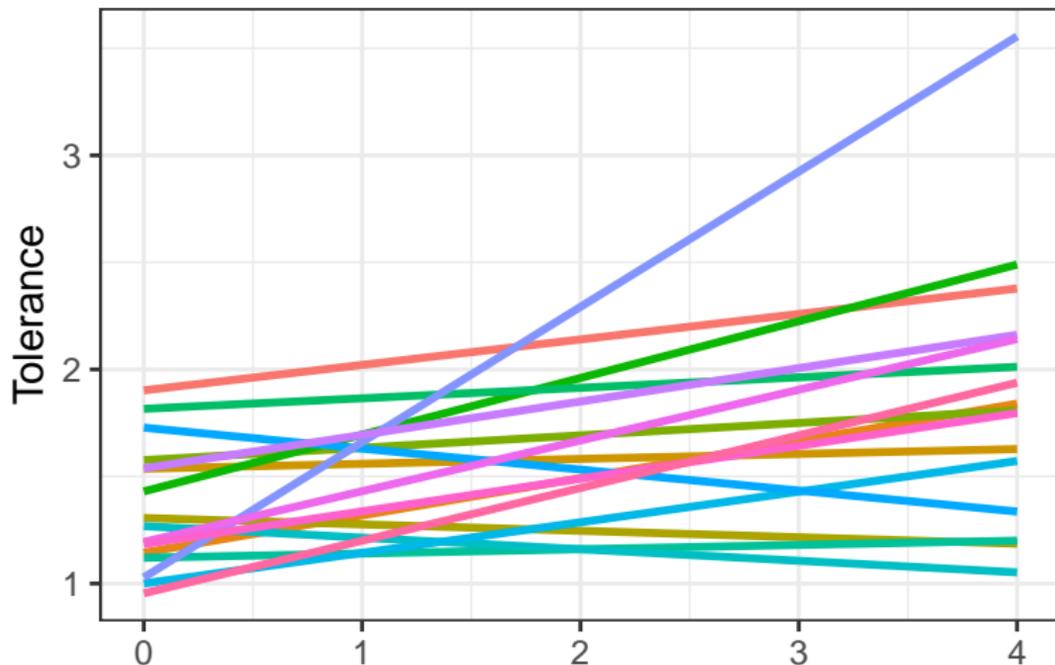
# ¿Todos los individuos cambian de la misma manera?

Volvamos al ejemplo de la tolerancia en los jóvenes

# ¿Todos los individuos cambian de la misma manera?

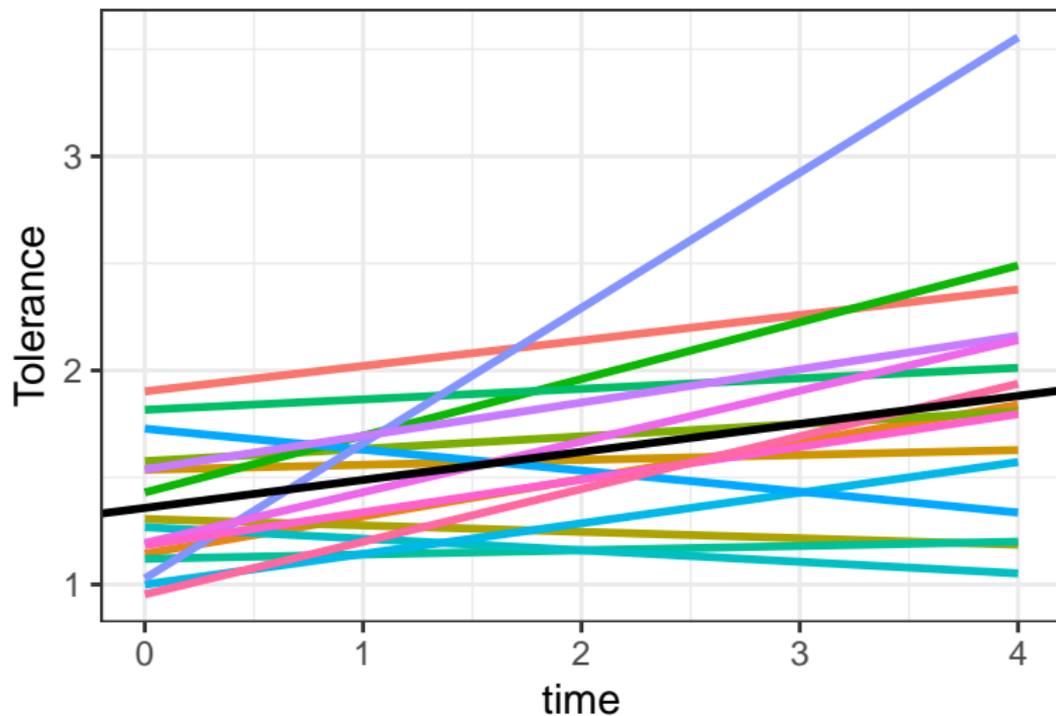
Volvamos al ejemplo de la tolerancia en los jóvenes

## Trayectorias MCO por participante



# ¿Todos los individuos cambian de la misma manera?

Trayectorias MCO por participante



# Diferencias entre individuos en trayectorias

- El cambio en la tolerancia en jóvenes entre 11 y 15 años es **positivo y bajo magnitud**. La pendiente promedio fue 0.13, aumento promedio por año (en escala 1 a 4).
- **Aunque la trayectoria de cambio promedio es un buen resumen, hay que tomarlo con cautela**: la forma de la trayectoria de cambio promedio puede ser muy distinta de la forma de las trayectorias individuales de las que se deriva.

# Diferencias entre individuos en trayectorias

- El cambio en la tolerancia en jóvenes entre 11 y 15 años es **positivo y bajo magnitud**. La pendiente promedio fue 0.13, aumento promedio por año (en escala 1 a 4).
- **Aunque la trayectoria de cambio promedio es un buen resumen, hay que tomarlo con cautela:** la forma de la trayectoria de cambio promedio puede ser muy distinta de la forma de las trayectorias individuales de las que se deriva.

## Esto significa que

**nunca** se debe inferir la forma de las trayectorias de cambio individuales a partir de la forma de su promedio.

# ¿Cómo aproximar las diferencias entre individuos en el cambio?

Estadísticas descriptivas de los interceptos y pendientes individuales estimados:

- 1 Medias muestrales
- 2 Varianzas (desvíos) muestrales
- 3 Correlaciones muestrales  $\Rightarrow$  ¿Están relacionados el estado inicial observado y la tasa de cambio?

# Interceptos y pendientes de trayectorias lineales individuales

Modelos de regresión MCO independientes para  $c/$  individuo de la Tolerancia como función lineal del tiempo ( $n = 16$ ).

```
indiv <- unique(tolerance.pp$id)
beta0 <- NULL
beta1 <- NULL

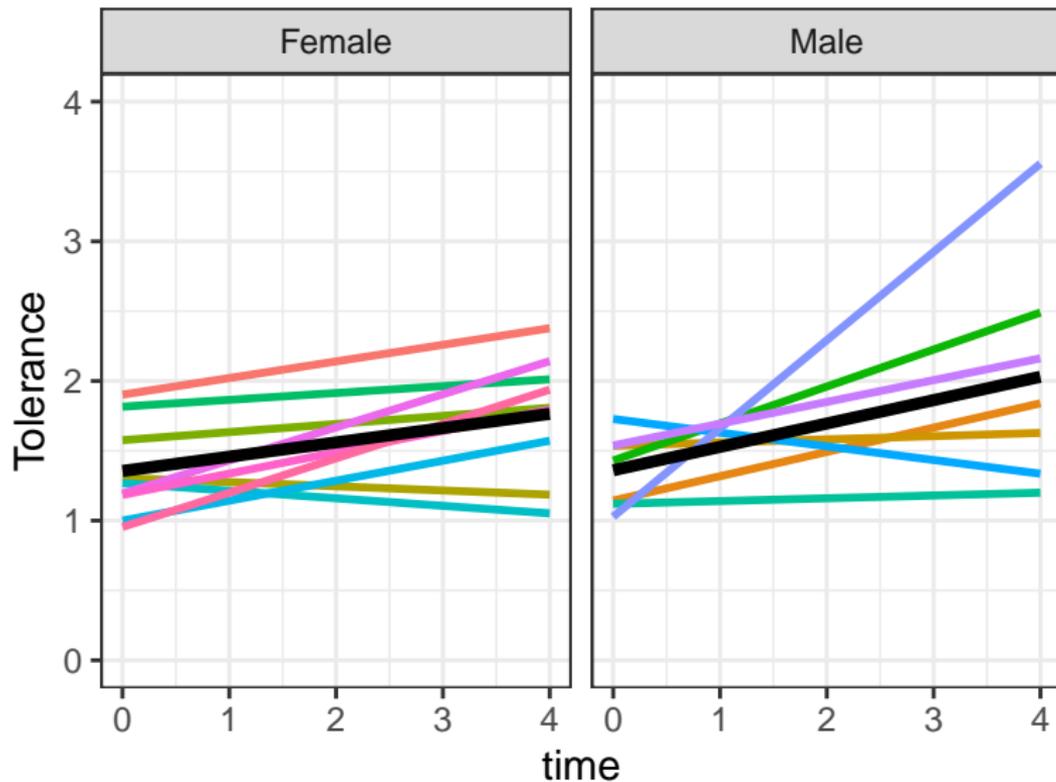
for (j in 1:16) {
  beta0[j] <- coef(lm(tolerance ~ time,
                     data = tolerance.pp %>%
                       filter(id==indiv[j]))) [1]
  beta1[j] <- coef(lm(tolerance ~ time,
                     data = tolerance.pp %>%
                       filter(id==indiv[j]))) [2]
}
```

# Estadísticas descriptivas de parámetros de trayectorias individuales

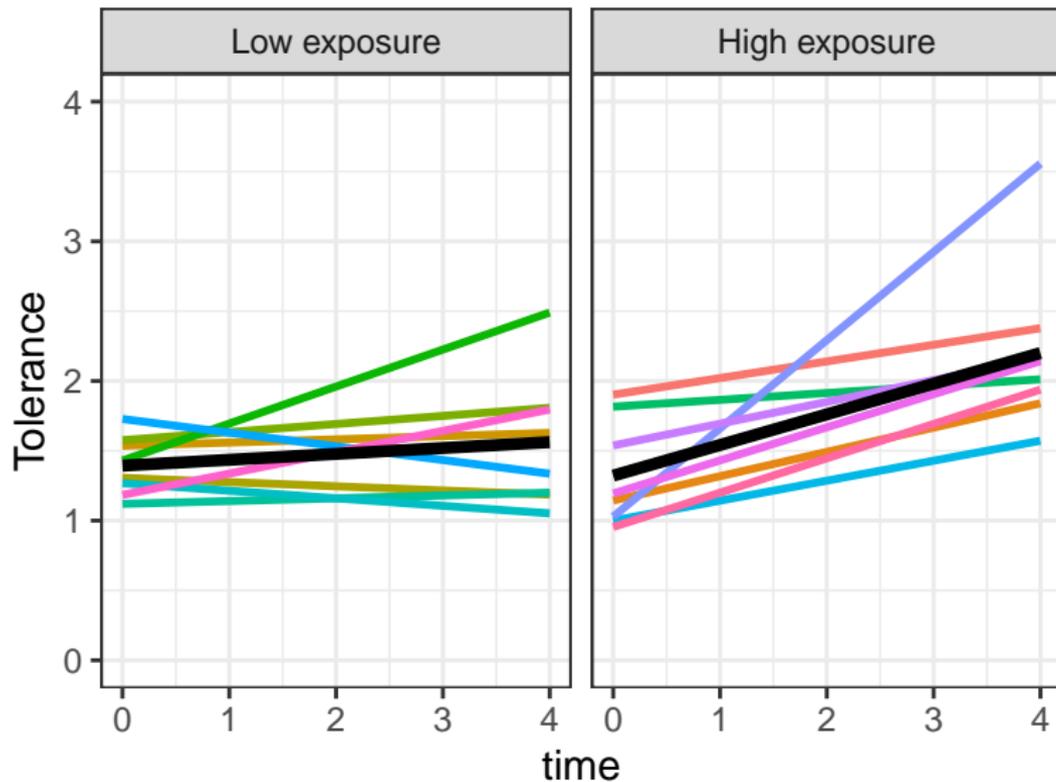
```
data.frame(beta0=beta0,beta1=beta1) %>%  
  summarise(  
    int_m = mean(beta0),  
    int_d = sd(beta0),  
    pen_m = mean(beta1),  
    pen_d = sd(beta1),  
    cor   = cor(data.frame(beta0=beta0,beta1=beta1))[1,2]  
  ) %>%  
  mutate_if(is.numeric, ~ round(., 2))
```

```
##   int_m int_d pen_m pen_d cor  
## 1  1.36  0.3  0.13  0.17 -0.45
```

## Relación con variables invariantes - *Female/Male*



## Relación con variables invariantes - *Low/High exp*



# Un comentario sobre MCO

- El método MCO asume independencia y homocedasticidad de los residuos.
- Supuesto muy poco probable en datos longitudinales: residuos suelen ser autocorrelacionados y heterocedásticos a lo largo del tiempo *within-individual*.
- Pero las estimaciones MCO pueden ser útiles con fines exploratorios. Aunque sean menos eficientes, siguen proporcionando estimaciones insesgadas del intercepto y la pendiente del cambio individual.

**¡¡Gracias!!**

**alejandra.marroig@fcea.edu.uy**