

Seminario Análisis Multinivel con RStudio

Malú Feo Serrato
(mafeo@ucm.es)



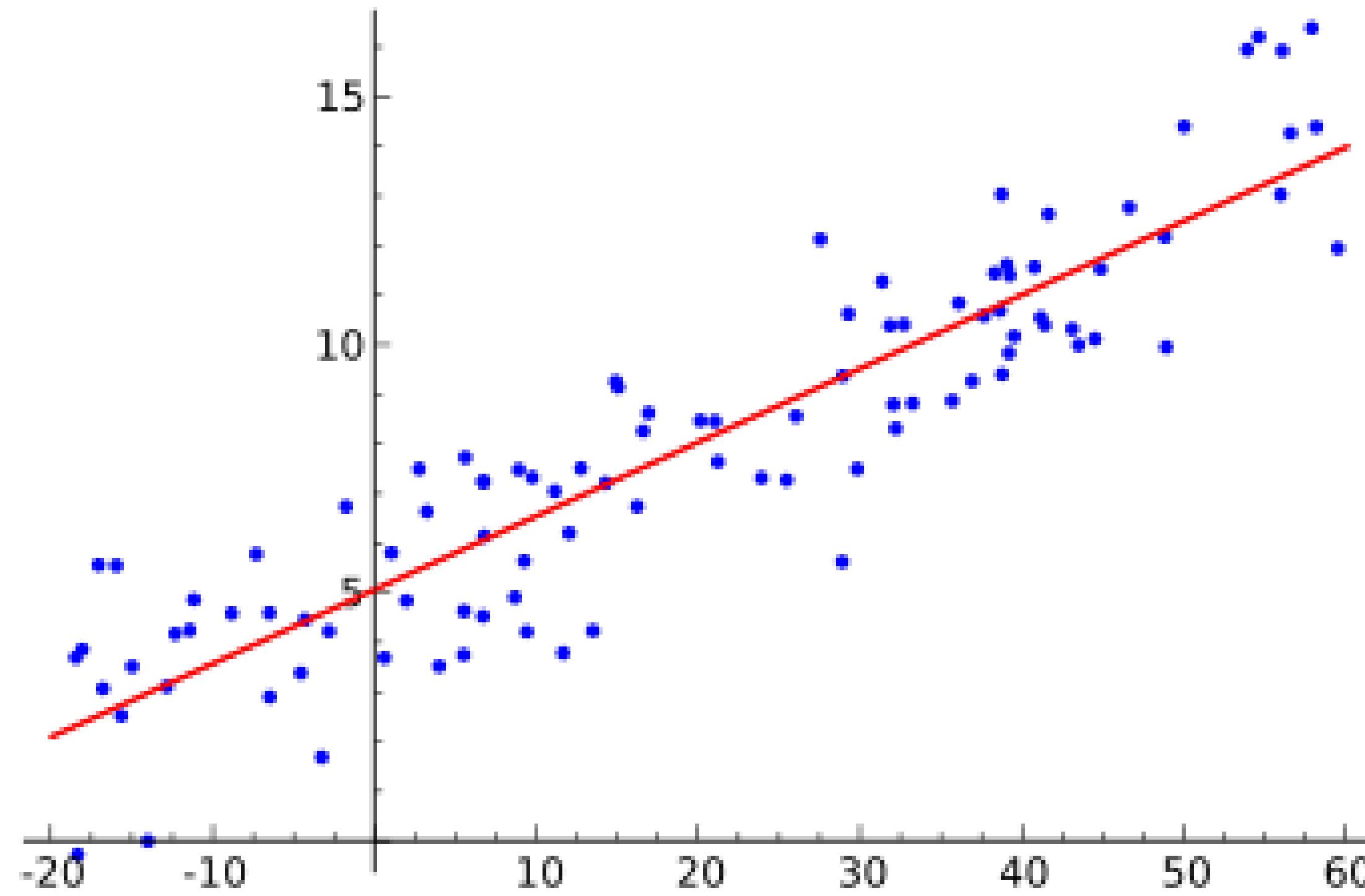
Introducción a la introducción del análisis multinivel

¿Qué es?

¿Por qué lo necesitamos?

¿Qué resuelve?





¿Qué hace la regresión lineal?

- Relacionar/predicir una variable dependiente (Y) continua
- Con una o más variables independientes (X)
- Suponiendo que:
 - Los errores son normales
 - La varianza es constante
 - Las observaciones son **independientes**

El problema de la independencia

- En ciencias sociales, raramente se cumple (objeto de estudio en ciencias sociales: personas)
- Las personas están agrupadas en contextos:
 - Aulas, colegios, barrios, países, organizaciones...
- Esto viola un supuesto fundamental del modelo lineal clásico

¿Por qué importa esto?

Si ignoramos la estructura jerárquica podemos caer en... :

- Subestimación errores estándar
- Sobreestimación significación estadística
- Menor precisión explicativa
- Conclusiones “engañosas”

Ejemplo 1: sistema educativo

- Nivel 1 → Estudiantes (nota final, motivación...)
- Nivel 2 → Clase (clima del aula, docente, turno mañana/tarde...)

Estudiantes del mismo grupo se parecen entre sí (comparten condiciones que pueden influir en la VD)

Ejemplo 1: sistema educativo

Jerarquías posibles:

- Estudiantes ↔ Clases ↔ Escuelas ↔ Sistemas educativos
- Personas ↔ Barrios ↔ Ciudades ↔ Regiones ↔ Países

Cuanto más complejo el fenómeno, más niveles pueden ser relevantes

Ejemplo 1: sistema educativo



¿Qué nos aporta? (Método)

- Respeta la estructura de los datos
- Separa varianza individual/"contextual"
- Modela intersecciones y pendientes aleatorias
- Permite variables a distintos niveles
- Mejora la precisión de los coeficientes y reduce errores estándar sesgados

¿Qué nos aporta? (Ontología/Epistemología)

- Supera el individualismo metodológico
- Reconoce la agencia situada. Acepta que las estructuras sociales influyen (sin determinismo)
- Enfoca los fenómenos ecológicamente
- Ignorar la estructura de los datos distorsiona tanto las inferencias estadísticas como las interpretaciones teóricas

Conclusión del bloque

El análisis multínivel no es solo una técnica, es una forma de pensar: relacional-contextual -ecológica

Nos permite comprender procesos sociales complejos con mayor rigor.

Los contextos sociales no son ruido; son parte del modelo teórico

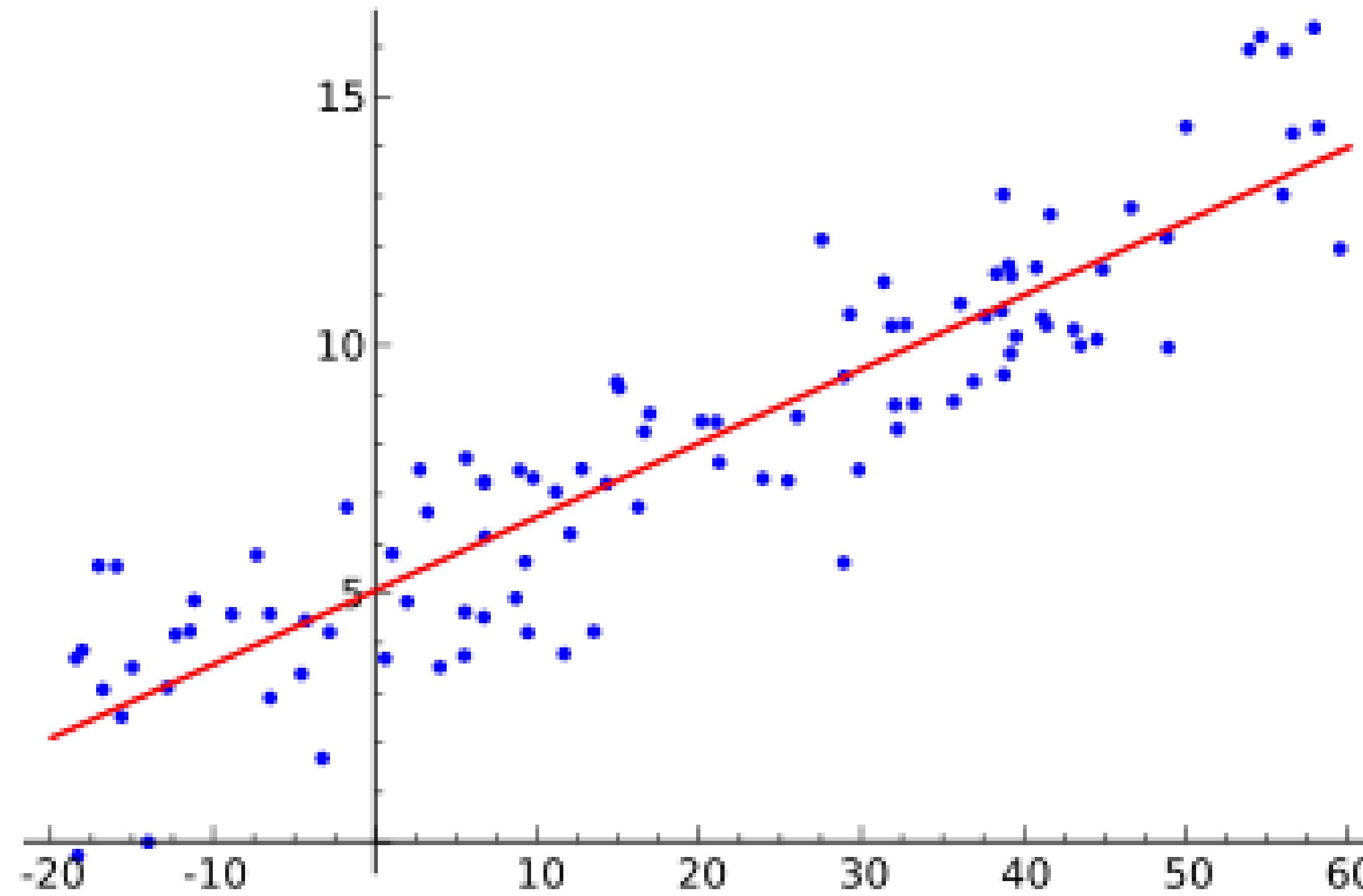
Introducción al modelado multinivel

¿En qué 'partes' se divide el modelo? → fileteo del modelo

Conceptos clave

La lógica del análisis multinivel





Anatomía del modelo de regresión lineal

i se refiere a un sujeto,
 $\{1, \dots, i, \dots, N\}$

$$Y_i = \beta_0 + \beta_1 + e_i$$

Supone que todos los casos:

- Tienen la misma intersección (β_0)
- Tienen la misma pendiente (β_1)
- Los errores son independientes (e_i)

Anatomía del modelo de regresión lineal

β_0 (intersección):

- Es el valor esperado de Y cuando X=0. Es donde la recta corta el eje Y

β_1 (pendiente o slope):

- Cambio esperado en Y por cada unidad de cambio en X. Se asume que esta relación es constante para todas las personas.

Pongamos que nuestra VD (Y_i) es el rendimiento académico de Sara en el examen del Seminario . El rendimiento académico viene definido por una VI (motivación)

Entonces, ¿cómo son los efectos de la
regresión lineal clásica?

Supone que todos los estudiantes, en todas las clases, siguen la misma relación motivación → rendimiento

Supongamos datos jerárquicos...

Estudiantes en clases:

- Cada estudiante tiene una nota final (Y)
- Están agrupados en clases (nivel 2)

¿Es razonable pensar que todas las clases tienen el mismo promedio?

Efectos fijos y aleatorios

Efectos fijos:

- Son coeficientes constantes en toda la muestra.
- Se interpreta como el promedio general del efecto de una variable.
- Representan una estimación única que se aplica por igual a todos los grupos.

Efectos fijos y aleatorios

Incluye:

- Intersección fija: nivel medio de la variable dependiente cuando todas las predictoras están en cero (o en su media si se centran).
- Pendiente fija: cambio promedio en la variable dependiente por cada unidad de cambio en el predictor.

Efectos fijos y aleatorios

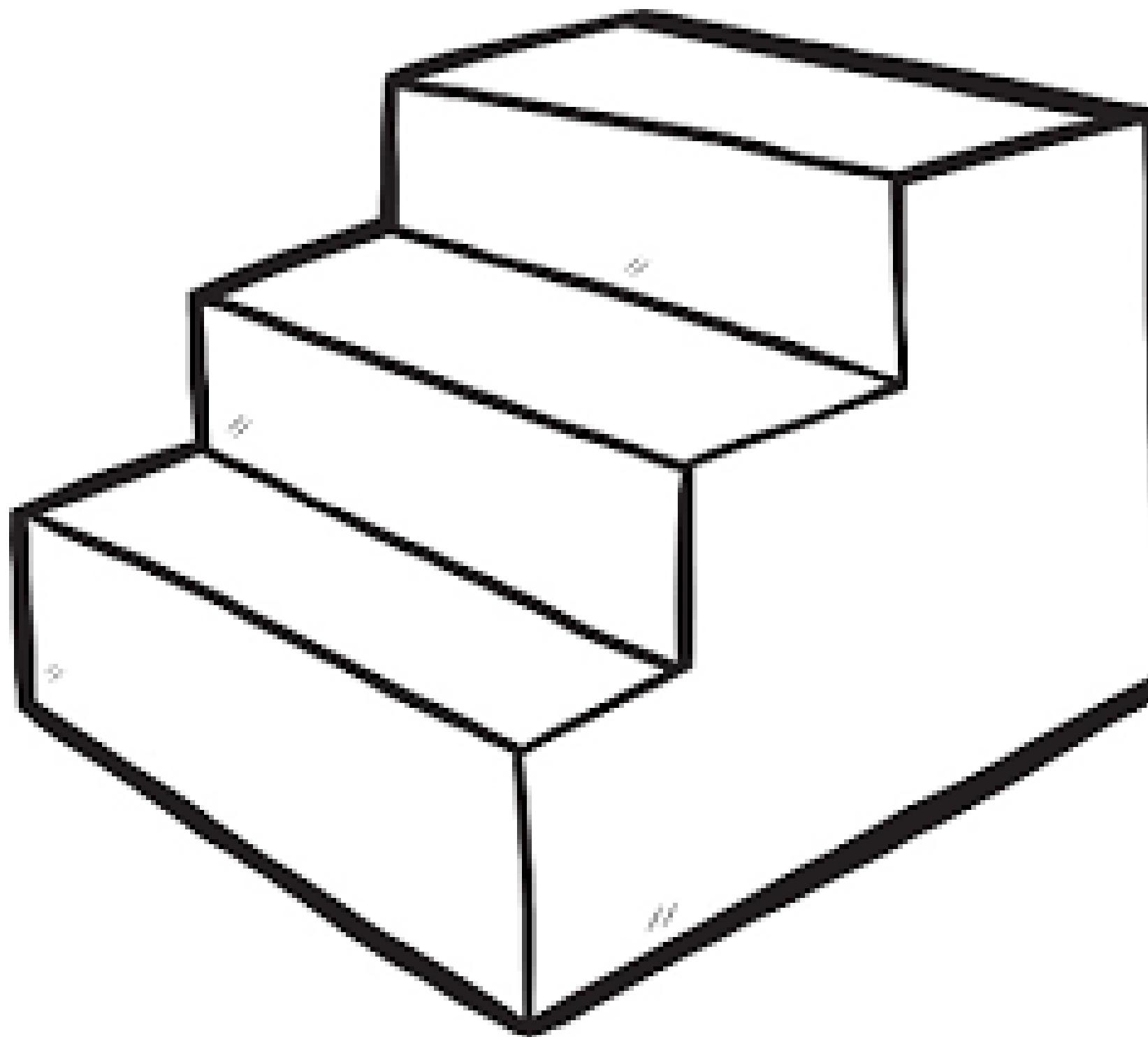
Efectos aleatorios

- Son coeficientes que varían entre unidades superiores (colegios, clases, países...).
- Permiten modelar la heterogeneidad entre contextos.
- Pueden aplicarse tanto a la intersección como a las pendientes.

Efectos fijos y aleatorios

Efectos aleatorios

- Intersección aleatoria: permite que cada grupo (ej. cada clase) tenga un punto de partida distinto.
- Pendiente aleatoria: permite que el efecto de una variable (ej. motivación) no sea igual en todos los grupos.



Anatomía del modelo nulo (sin predictores)

$$Y_{ij} = \beta_{0j} + e_{ij}$$

Error aleatorio

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

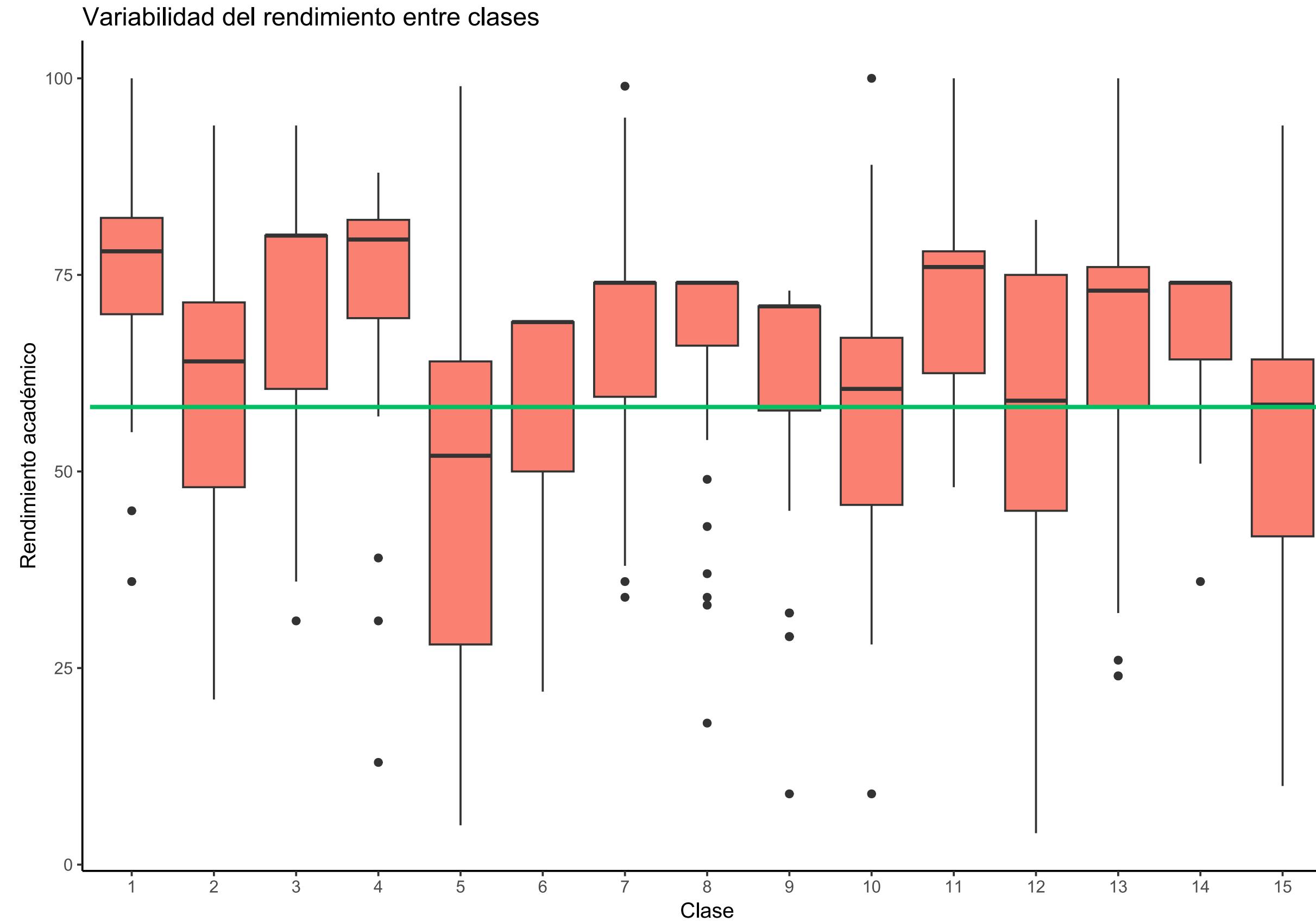
Intersección específica del grupo j

Intersección global promedio

Desviación aleatoria del grupo respecto al promedio

```
graph TD; Yij[Y_{ij}] --> Beta0j[β_{0j}]; Beta0j --> Gamma00[γ_{00}]; Gamma00 --> Eij[e_{ij}]; Eij --> Error[Error aleatorio]; Beta0j --> U0j[u_{0j}]; U0j --> Desv[Desviación aleatoria del grupo respecto al promedio];
```

El modelo estima un nivel promedio de Y por grupo, sin explicar por qué varía. Aquí la pendiente no aparece todavía, solo interesa cuánto varía el promedio entre grupos.



Anatomía del modelo intersección aleatoria y pendiente fija

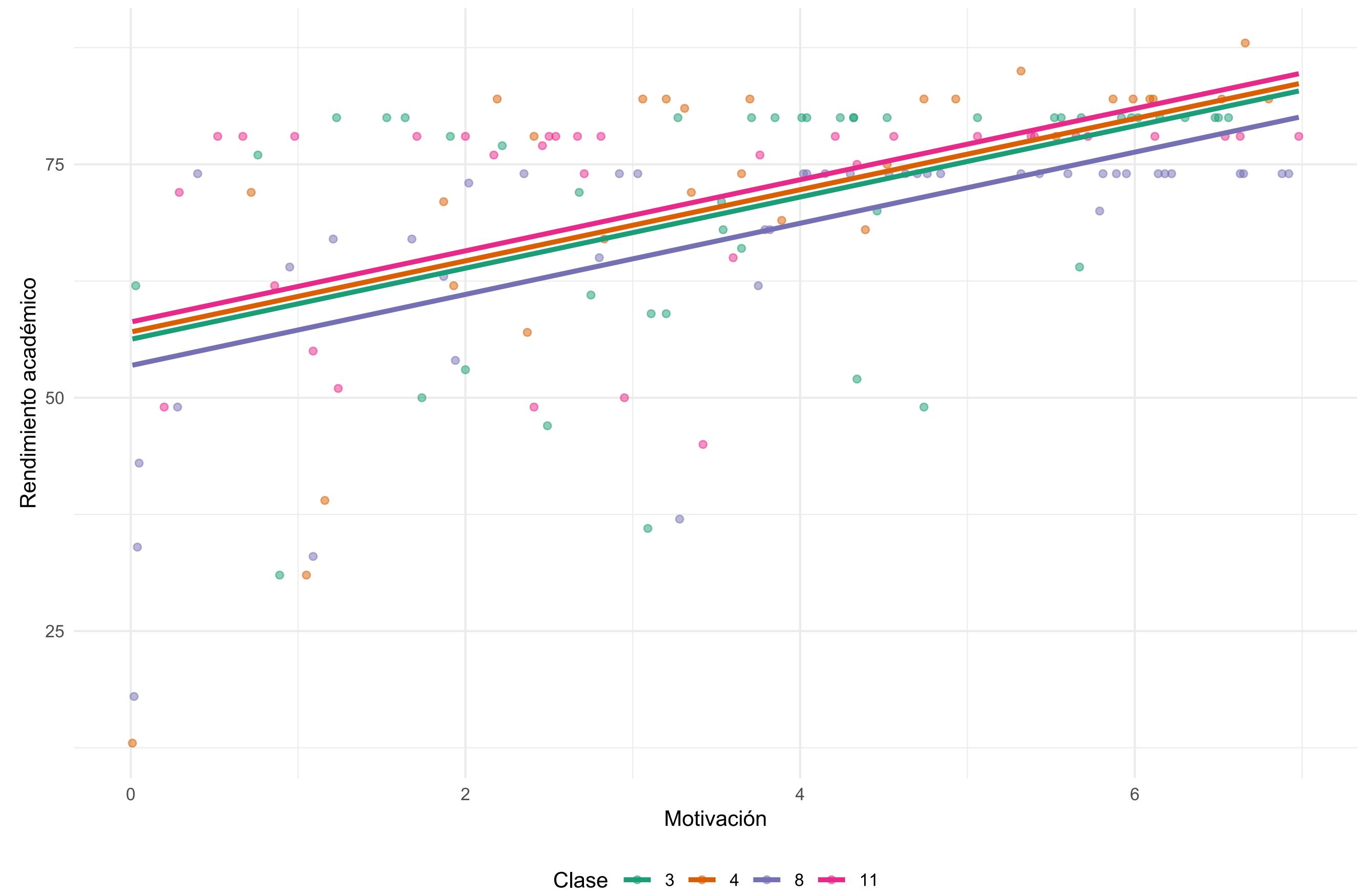
$$Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$

↓
Intersección varía entre grupos (intersección aleatoria) → permite que cada grupo tenga su propio “punto de partida”.

→ Fija → misma para todos los grupos → se estima una sola vez.

Ejemplo educativo: todos los colegios comparten el mismo efecto de motivación sobre rendimiento, pero tienen promedios diferentes.

Modelo de intersección aleatoria y pendiente fija



Anatomía del modelo intersección y pendiente aleatoria

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

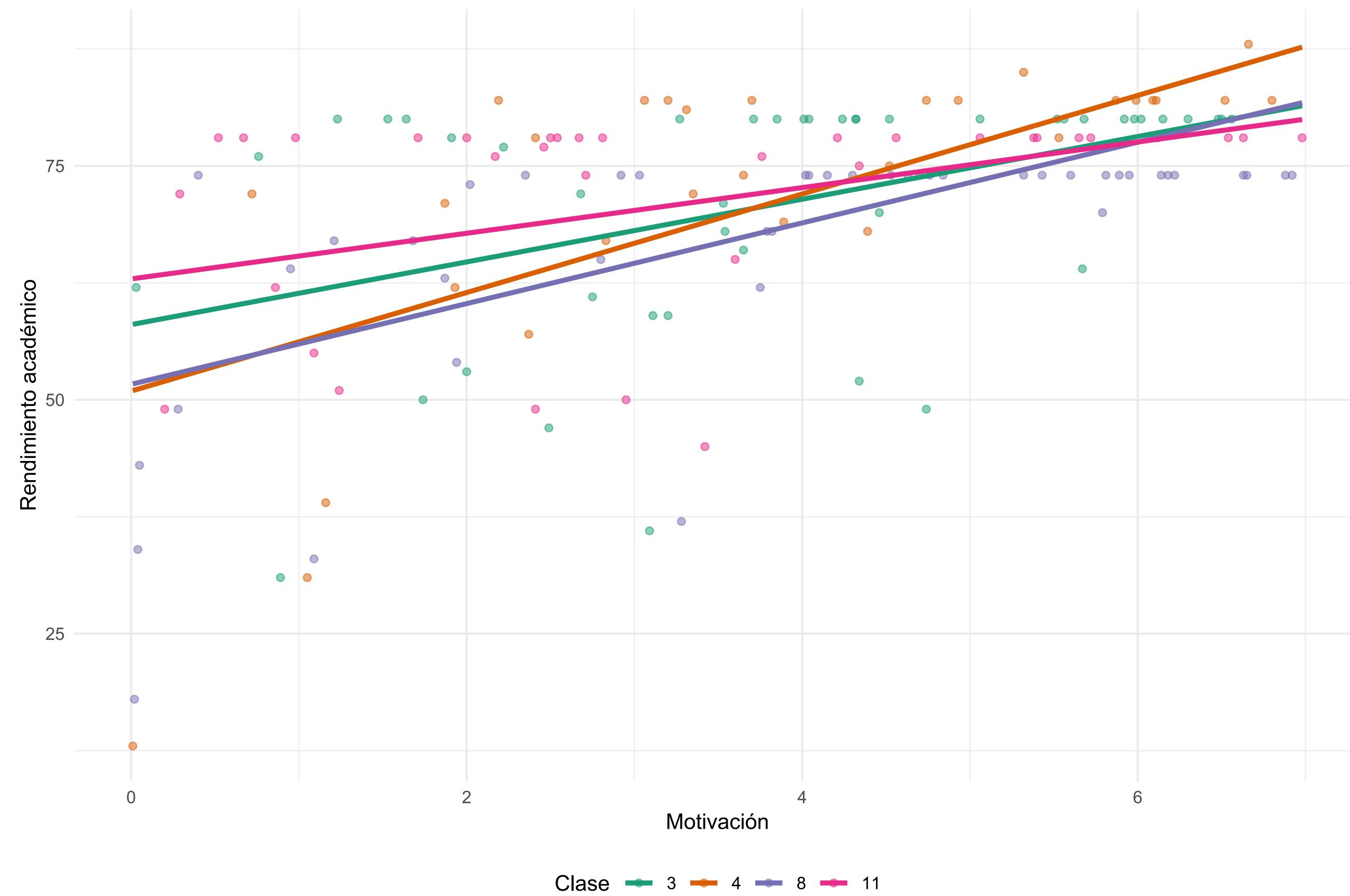
pendiente promedio

desviación aleatoria del grupo
respecto a la pendiente media

pendiente específica del grupo j

Cada grupo tiene su propia línea de regresión, con su propio nivel base y su propia 'sensibilidad' a X.

Modelo de intersección y pendiente aleatorias



¿Cuándo considerar análisis multinivel?

El ICC (Intra-Class Correlation coefficient) qué proporción de la varianza total de la variable dependiente se debe a diferencias entre grupos (colegios, clases...)

- Se calcula a partir del modelo nulo (sin predictores)
- Si hay agrupación en los datos, el ICC NO será 0.
- Valores próximos a 1 → mayor variabilidad debida a los grupos
- Se suele informar del porcentaje → multiplicar por 100.

¿Cuándo considerar análisis multinivel?

$$\text{ICC} = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}$$

$$\text{ICC} = \frac{\sigma_{\text{grupo}}^2}{\sigma_{\text{grupo}}^2 + \sigma_{\text{residual}}^2}$$

Incluso con ICC bajos, puede justificarse el multinivel si hay predictores de nivel 2 que tengan sentido a nivel teórico

¿Y con qué modelo nos quedamos?

El criterio más importante es el test de razón de verosimilitudes (Likelihood Ratio Test, LRT).

- Sirve para comparar dos modelos anidados (uno más simple y otro más complejo)
- Un modelo está anidado en otro si puede obtenerse eliminando uno o más parámetros del modelo más complejo.
- Modelo nulo < modelo intersección fija y pendiente aleatoria < modelo intersección fija y pendiente aleatoria

¿Y con qué modelo nos quedamos?

$$\text{LRT} = -2 \cdot (\log L_{\text{modelo simple}} - \log L_{\text{modelo complejo}})$$

Se compara con una distribución chi-cuadrado con k grados de libertad, donde k = diferencia en el número de parámetros

- Si el test es significativo, el modelo complejo mejora sustancialmente el ajuste
- Si no lo es, preferimos el modelo más simple (parsimonia)

¿Y con qué modelo nos quedamos?

Criterio	Qué mide	Mejor modelo
AIC (Akaike)	Compromiso entre ajuste y complejidad	Menor AIC
BIC (Bayesian)	Penaliza más la complejidad con muestras grandes	Menor BIC
LogLik	Log-verosimilitud	Más alto = mejor ajuste

Consideraciones previas al análisis multínivel

Base de datos: long vs. wide

Centrado de variables: dar significado al 0



Formato de base de datos: ¿long o wide?

Formato	Características	Cuándo se usa
Wide	Una fila por sujeto o unidad	Diseños multivel clásicos (ej. sujetos en grupos)
Long	Varias filas por sujeto	Diseños longitudinales (medidas repetidas por persona)

Centrado: dar sentido al 0

Tipo de centrado	¿Qué significa el 0?	Cuándo usarlo
Grand-mean	Media total	Efecto comparado al promedio total
Group-mean	Media de su grupo	Aísla efecto dentro de grupo

Análisis multinivel en RStudio



malufese



lmer

```
model<- lmer(rendimiento ~ motivacion + (motivacion | clase), data = data, REML = F)
```

```
summary(modelo_pendiente_aleatoria)
```

Bibliografía

- Hox, J., Moerbeek, M., & Van de Schoot, R. (2017). Multilevel analysis: Techniques and applications. Routledge.
- Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning variation in multilevel models. *Understanding statistics: statistical issues in psychology, education, and the social sciences*, 1(4), 223-231.
- Snijders, T. A., & Bosker, R. (2011). Multilevel analysis: An introduction to basic and advanced multilevel modeling.
- Hox, J. (1998). Multilevel modeling: When and why. In Classification, data analysis, and data highways: proceedings of the 21st Annual Conference of the Gesellschaft für Klassifikation eV, University of Potsdam, March 12–14, 1997 (pp. 147-154). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Sommet, N., & Morselli, D. (2021). Keep calm and learn multilevel linear modeling: A three-step procedure using SPSS, Stata, R, and Mplus. *International Review of Social Psychology*, 34(1).