

Documento base

Pablo Álvarez Arnedo

2026-02-11

Table of contents

Capítulos

1 + 1

[1] 2

1 Introducción

```
1 + 1
```

```
[1] 2
```

2 Capítulo 2: Datos longitudinales

2.1 ¿Qué son los datos longitudinales?

Los **datos longitudinales** son aquellos que se recopilan observando repetidamente las mismas unidades (individuos, regiones, células, etc.) a lo largo del tiempo. Este tipo de datos es común en estudios donde se busca evaluar cómo evolucionan ciertas características o mediciones bajo distintas condiciones o tratamientos.

2.1.1 Características principales

1. **Medidas repetidas:** Cada unidad tiene varias observaciones en diferentes momentos temporales.
2. **Estructura jerárquica:** Las observaciones están agrupadas por unidades (e.g., pacientes, regiones).
3. **Dependencia entre observaciones:** Las mediciones dentro de la misma unidad tienden a estar correlacionadas.

2.1.2 Ejemplos de datos longitudinales

1. **Ámbito biosanitario:** Medidas repetidas de presión arterial en un grupo de pacientes durante un tratamiento.
2. **Educación:** Evaluación de los puntajes de un estudiante a lo largo de varios exámenes anuales.
3. **Ciencias sociales:** Encuestas de opinión realizadas periódicamente a las mismas personas.

2.2 ¿Por qué no se puede usar la estadística clásica?

La **estadística clásica** (e.g., regresión lineal simple) supone que todas las observaciones son independientes entre sí. Sin embargo, en datos longitudinales, esta suposición no se cumple debido a la correlación entre observaciones tomadas de la misma unidad.

2.2.1 Problemas al aplicar técnicas clásicas

1. **Subestimación de la incertidumbre:** Ignorar la dependencia lleva a errores estándar más pequeños, inflando la significancia estadística.
2. **Modelos mal ajustados:** No considerar la estructura jerárquica puede llevar a conclusiones erróneas.
3. **Violación de las hipótesis básicas:** En regresión lineal simple, se viola la suposición de independencia de los errores.

2.2.2 Ejemplo conceptual

Consideremos un conjunto de datos en el que medimos un marcador biomédico en varios pacientes durante 3 años consecutivos. Si intentáramos aplicar una regresión lineal simple para modelar la evolución del marcador en función del tiempo, ignoraríamos que las mediciones de un mismo paciente están correlacionadas.

```
# Ejemplo conceptual: Datos longitudinales simulados
set.seed(123)
data_long <- data.frame(
  id = rep(1:5, each = 3),
  year = rep(0:2, times = 5),
  biomarker = c(50, 52, 54, 48, 49, 51, 55, 57, 59, 47, 49, 50, 53, 54, 56)
)
# CON REGRESIÓN LINEAL
model_incorrect <- lm(biomarker ~ year, data = data_long)
summary(model_incorrect)
```

Call:

```
lm(formula = biomarker ~ year, data = data_long)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.9667	-3.1167	-0.2667	2.2333	5.0333

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	50.567	1.368	36.956	1.49e-14 ***
year	1.700	1.060	1.604	0.133

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.352 on 13 degrees of freedom
Multiple R-squared: 0.1652, Adjusted R-squared: 0.101
F-statistic: 2.573 on 1 and 13 DF, p-value: 0.1327

En el modelo de regresión lineal, las mediciones de un paciente afectan a las estimaciones de los demás, lo que genera resultados sesgados

2.3 Modelos mixtos

Para analizar datos longitudinales de manera adecuada, se deben emplear modelos mixtos, que permiten: - Capturar la variabilidad entre individuos mediante efectos aleatorios. - Modelar la correlación entre observaciones dentro de una misma unidad. - Incluir covariables tanto a nivel individual como grupal.

2.3.1 Ventajas de los modelos mixtos

- Flexibilidad para incluir efectos específicos por individuo o grupo.
- Estimación precisa de la incertidumbre, respetando la dependencia entre observaciones.
- Generalización a estructuras de datos complejas.