

Modelos Lineales con R

Práctica 5: Método de Newton y Algoritmo EM en Modelos Lineales Mixtos

Profesor: Andrés García Medina
andres.garcia.medina@uabc.edu.mx

Fecha de entrega: jueves 25 de abril, 2024 (12pm).

Instrucciones: Subir un documento pdf a classroom con las respuestas de cada uno de los ejercicios solicitados. Adjunta el código fuente en formato `.r` o `.ipynb`. Justificar detalladamente cada una de sus respuestas.

Ejercicio 1 (50 pts)

Los datos `Oxide` de la biblioteca `nlme` contiene datos de control de calidad en la industria de semiconductores. El objetivo del ejercicio es investigar las fuentes de variabilidad en el espesor de las capas de óxido en obleas de silicio. El marco de datos contiene las siguientes columnas:

- **Thickness:** es el espesor de la capa de óxido (en nanómetros, hasta donde yo sé).
- **Source:** es un factor de dos niveles que indica de cuál de los dos posibles proveedores provino la muestra.
- **Site:** es un factor de 3 niveles, que indica en cuál de los tres sitios de la oblea de silicio se midió el espesor.
- **Lot:** es una variable de factor cuyos niveles indican de qué lote particular de obleas de silicio proviene la medicion.
- **Wafer:** es una variable factor con niveles que etiquetan las obleas individuales examinadas.

Los investigadores están interesados en descubrir si existen diferencias sistemáticas entre las dos fuentes y esperan que el espesor pueda variar sistemáticamente en los tres sitios; sólo están interesados en los lotes y obleas en la medida en que sean representativos de una población más amplia de lotes y obleas.

Se espera su colaboración en el proyecto realizando las siguientes actividades:

- (a) Identifique qué factores trataría como aleatorios y cuáles como fijos, en un análisis de modelo lineal mixto de estos datos.
- (b) Escriba un modelo que pueda constituir una base adecuada para comenzar a analizar los datos.
- (c) Realice un análisis completo de los datos, incluida la verificación del modelo. Su objetivo debe ser identificar las fuentes de variabilidad del espesor en los datos y cualquier efecto fijo que cause la variabilidad del espesor. Se sugiere analizar específicamente lo siguiente:
 1. Graficar residuales estandarizados vs. valores ajustados.
 2. Analizar la suposición de normalidad de los residuales a través de sus cuantiles.
 3. Analizar la suposición de normalidad de los efectos aleatorios a través de sus cuantiles.
 4. Ajustar un nuevo modelo mixto sobre los datos suponiendo que la variable `Wafer` representa un efecto aleatorio.
 5. Comparar su propuesta original con el modelo anterior usando la función `anova`.
 6. Determinar los intervalos de confianza de los parámetros al 99% usando la función `intervals`
- (d) Estime el modelo inicial del inciso (b) optimizando la función de máxima verosimilitud en lugar de la función objetivo por defecto de la librería `lme`. Discuta las posibles diferencias.
- (e) Estime el modelo inicial del inciso (b) considerando 10 iteraciones del algoritmo EM, seguido de 1000 iteraciones del algoritmo de Newton. Discuta las posibles diferencias.
- (f) Estime el modelo inicial del inciso (b) considerando el método de optimización denominado algoritmo de recocido simulado en lugar del método de Newton. Discuta las posibles diferencias.

Nota: Se recomienda leer material de apoyo titulada: *Crecimiento de árboles: un ejemplo usando lme*

Ejercicio 2 (30 pts)

Acceder al siguiente recurso electrónico: <https://stephens999.github.io/fiveMinuteStats/index.html>

- (a) Leer las secciones:
 - Introduction to Mixture Models

- Introduction to the EM algorithm
 - A variational view of the EM algorithm
 - EM for Gaussian Mixture Models
 - Archive: Alternative Introduction to Mixture Models
- (b) Realizar un ensayo de 500 palabras (aproximadamente 1 página) discutiendo como implementaría las ideas de la lectura en relación al algoritmo EM visto en clase en el contexto de modelos lineales mixtos. Se puede apoyar de ecuaciones, gráficos, diagramas, o recursos que considere necesarios.

Ejercicio 3 (20 pts)

Descargar el libro: Pinheiro, J., & Bates, D. (2001). Mixed-effects models in S and S-PLUS. Springer, Statistics and Computing.

- (a) Revisar Unidad 1
- (b) Resolver ejercicio 1 de final de capítulo.

Nota: EL libro se puede descargar de los recursos electronicos de UABC.