

Práctica 5

Maximiliano Vaca Montejano
364897
maximiliano.vaca@uabc.edu.mx

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

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

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 medición.
- **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:

1. Identifique qu'e factores trataria como aleatorios y cu'ales como fijos, en un análisis de modelo lineal mixto de estos datos.
2. Escriba un modelo que pueda constituir una base adecuada para comenzar a analizar los datos.
3. Realice un análisis completo de los datos, incluida la verificación del mod-elo. 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:

- A. Graficar residuales estandarizados vs. valores ajustados.
- B. Analizar la suposición de normalidad de los residuales a través de sus cuantiles.
- C. Analizar la suposición de normalidad de los efectos aleatorios a través de sus cuantiles.
- D. Ajustar un nuevo modelo mixto sobre los datos suponiendo que la variable Wafer representa un efecto aleatorio.
- E. Comparar su propuesta original con el modelo anterior usando la función anova.
- F. Determinar los intervalos de confianza de los parámetros al 99% usando la función intervals

(a) Identifique que factores trataria como aleatorios y cuales como fijos, en un análisis de modelo lineal mixto de estos datos.

In [2]:

```
library(nlme)
data(Oxide)

head(Oxide, 20) # 72 x 5

A nmGroupedData: 20 x 5
  Source Lot Wafer Site Thickness
<fct> <fct> <fct> <fct> <dbl>
1 1 1 1 1 2006
2 1 1 1 2 1999
3 1 1 1 3 2007
4 1 1 2 1 1980
5 1 1 2 2 1988
6 1 1 2 3 1982
7 1 1 3 1 2000
8 1 1 3 2 1998
9 1 1 3 3 2007
10 1 2 1 1 1991
11 1 2 1 2 1990
12 1 2 1 3 1988
13 1 2 2 1 1987
14 1 2 2 2 1989
15 1 2 2 3 1988
16 1 2 3 1 1985
17 1 2 3 2 1983
18 1 2 3 3 1989
19 1 3 1 1 2000
20 1 3 1 2 2004
```

- trata como fijos aquellos factores que queremos comparar directamente: *Source* y *Site*
- y como aleatorios los que pueden afectar las mediciones pero no son de interes principal, como lo son *Lot* y *Wafer*

(b) Escriba un modelo que pueda constituir una base adecuada para comenzar a analizar los datos.

In [3]:

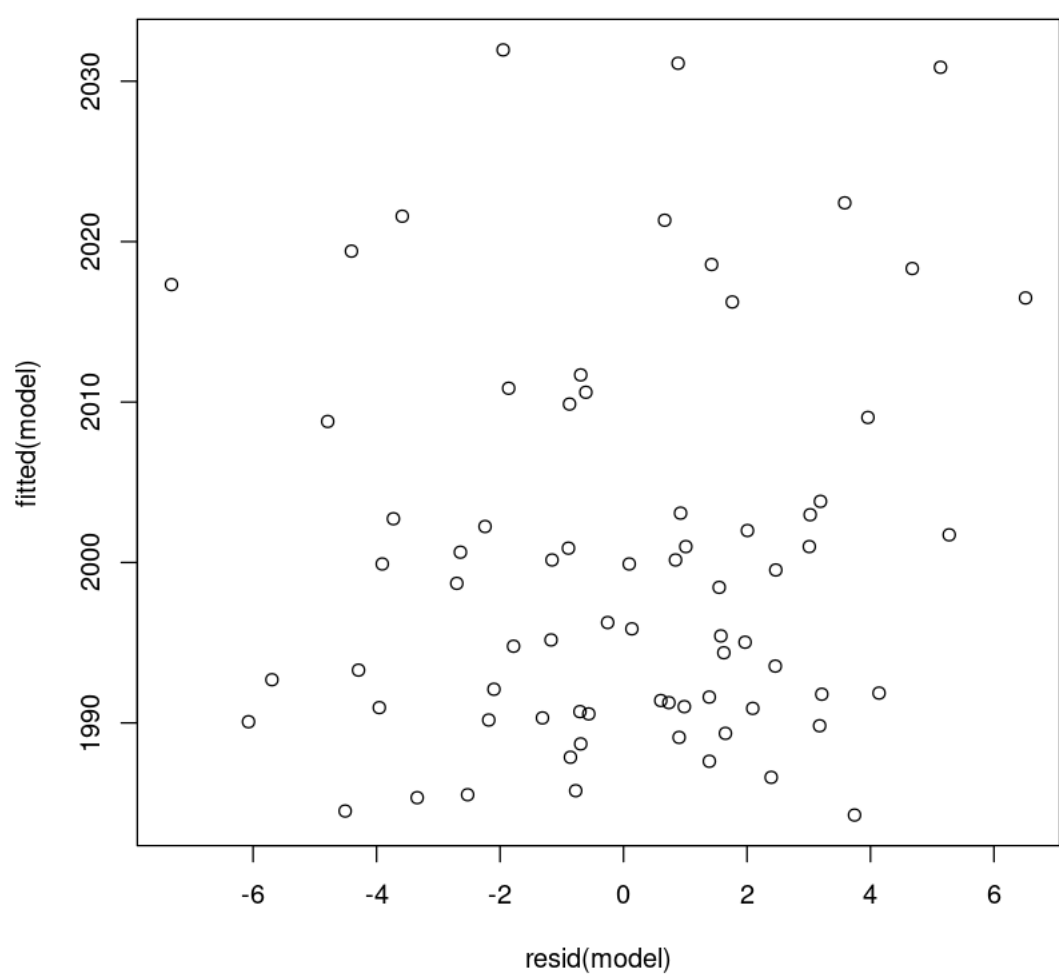
```
model <- lme(Thickness ~ Source + Site, random = ~ 1 | Lot/Wafer, data = Oxide)
```

(c) Realice un análisis completo de los datos, incluida la verificación del mod-elo. 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

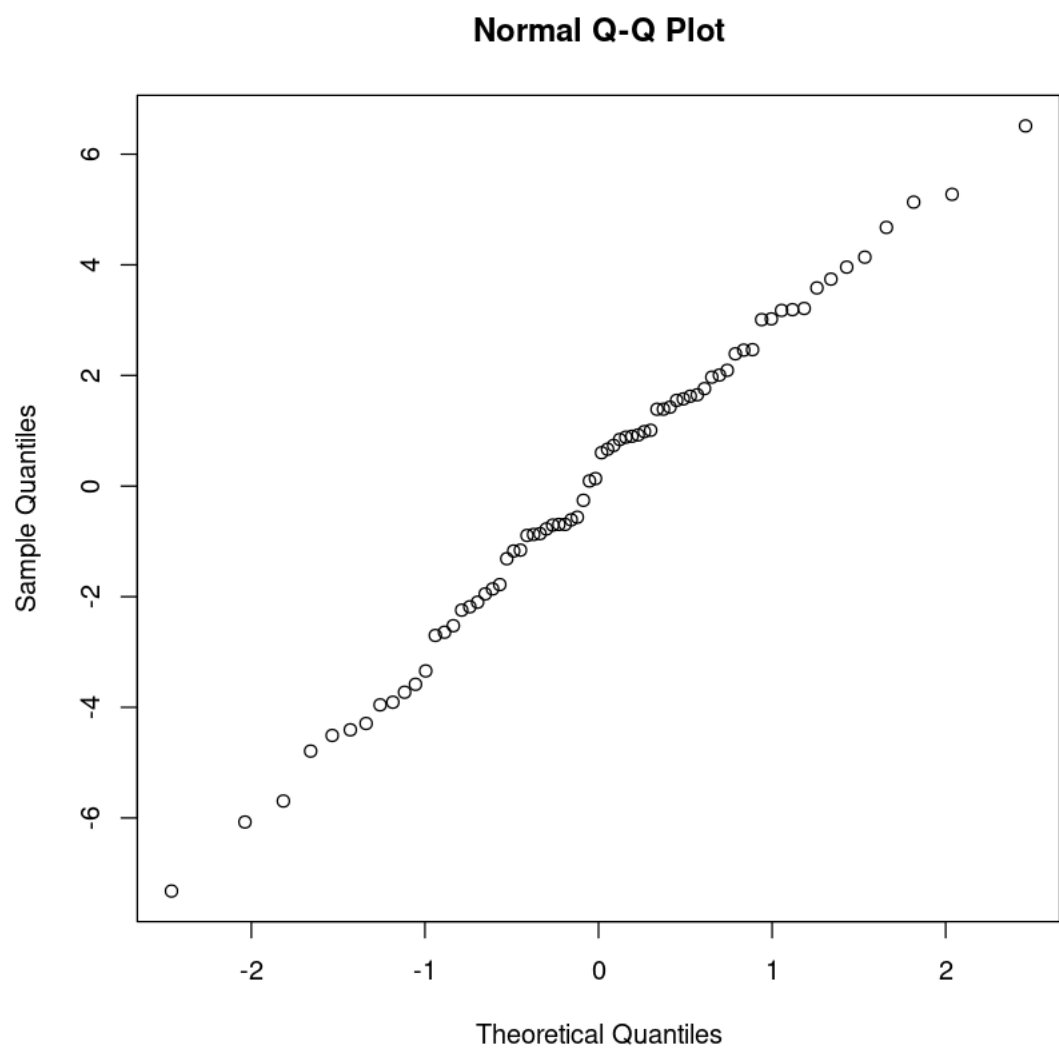
In [6]:

```
#residuales estandarizados vs. valores ajustados
plot(resid(model), fitted(model))
```



In [8]:

```
#analizar la suposición de normalidad de los residuales a través de sus cuantiles
qqnorm(resid(model))
```

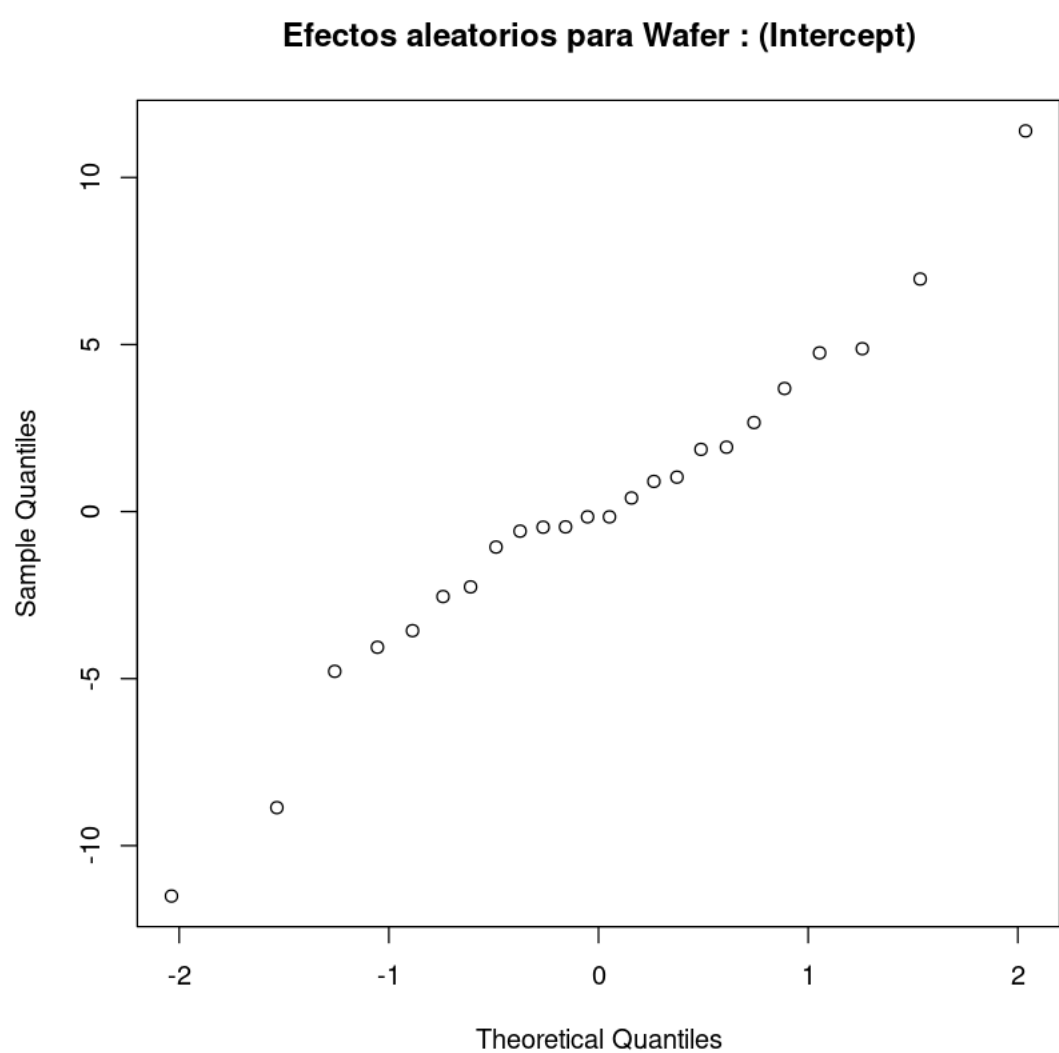
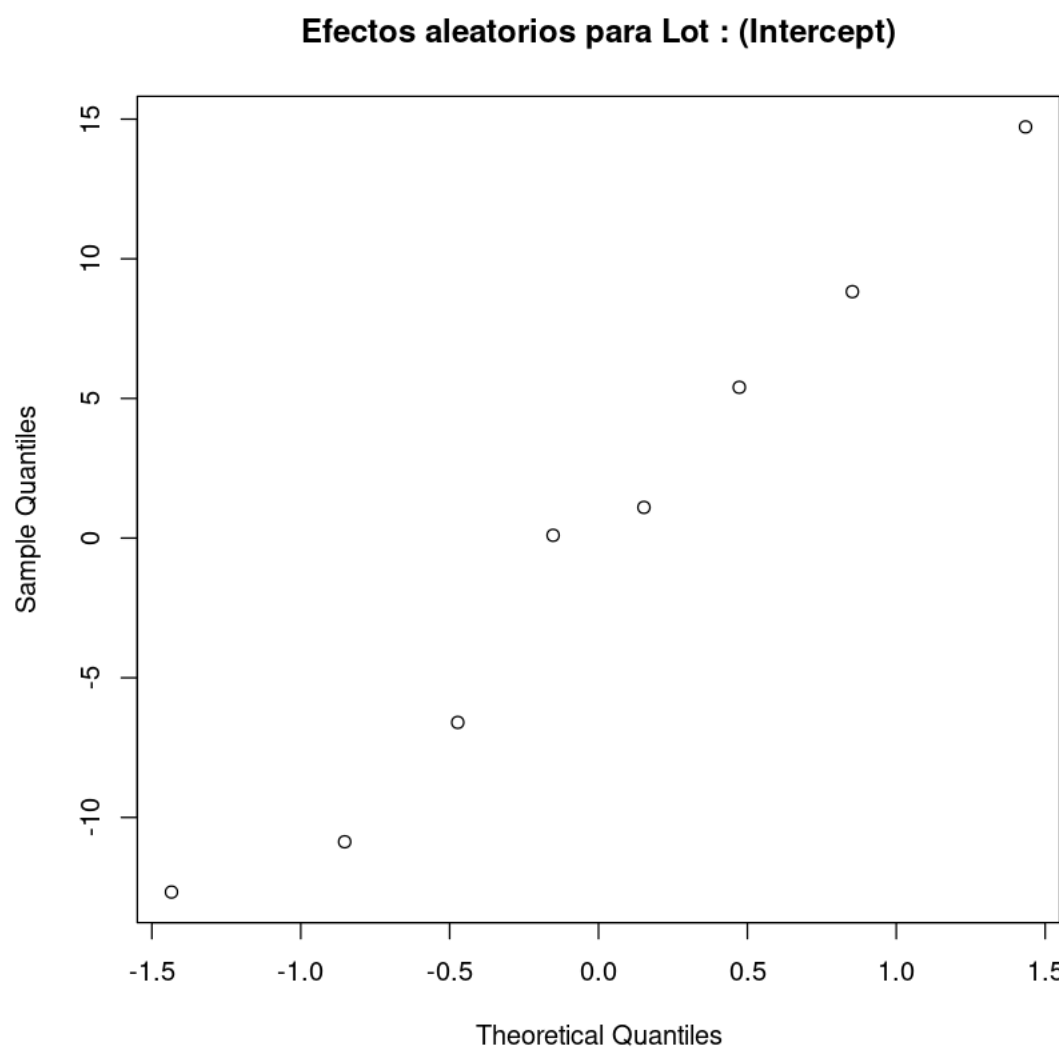


In [13]:

```
#efectos aleatorios del modelo
random_effects <- ranef(model)

for (term in names(random_effects)) {
  #efectos aleatorios
  effects <- random_effects[[term]]

  #gráfico QQ si es vector
  if (is.vector(effects)) {
    qqnorm(effects, main = paste("Efectos aleatorios para", term))
  } else {
    #gráfico QQ para cada columna
    for (col in colnames(effects)) {
      qqnorm(effects[, col], main = paste("Efectos aleatorios para", term, ":", col))
    }
  }
}
```



In [10]:

```
#suponiendo Wafer como efecto aleatorio
model2 <- lme(Thickness ~ Source + Site, random = ~ 1 | Wafer, data = Oxide)
```

In [11]:

```
#Comparar
anova(model, model2)
```

A anova.lme: 2 x 9									
	call	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
	<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<fct>	<dbl>	<dbl>
model	lme.formula(fixed = Thickness ~ Source + Site, data = Oxide, random = ~1 Lot/Wafer)	1	7	455.7619	471.2985	-220.8810		NA	NA
model2	lme.formula(fixed = Thickness ~ Source + Site, data = Oxide, random = ~1 Wafer)	2	6	554.7459	568.0629	-271.3729	1 vs 2	100.9839	9.273318e-24

In [12]:

```
#intervalos de confianza de los parámetros al 99%
intervals(model, level = 0.99)
```

Approximate 99% confidence intervals

Fixed effects:			
	lower	est.	upper
(Intercept)	1979.325947	1994.916667	2010.507386
Source2	-20.177583	10.083333	40.344250
Site2	-3.022988	-0.250000	2.522988
Site3	-1.939655	0.833333	3.606321

Random Effects:			
Level: Lot			
sd((Intercept))	4.784201	10.94954	25.06008
Level: Wafer			
sd((Intercept))	3.590726	5.982932	9.968868

Within-group standard error:			
	lower	est.	upper
2.732989	3.574940	4.676271	

In []: