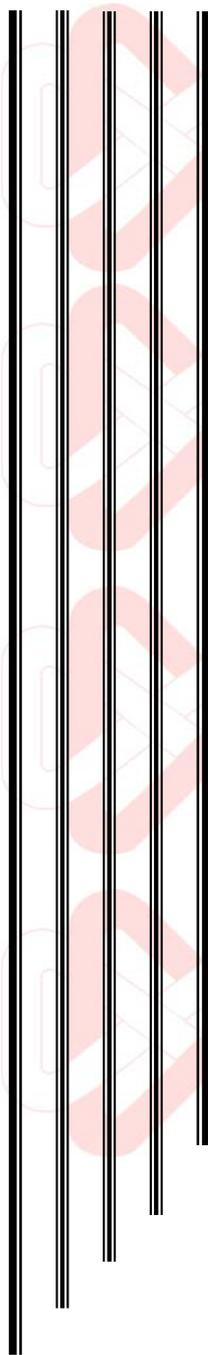




Centro de Investigación en Matemáticas, A.C.

CIMAT



**Determinación de Tiempos de
Paro de Estudios de
Degradación en Confiabilidad**

T E S I S

Que para obtener el grado de:

Maestría en Ciencias

con especialidad en:

Probabilidad y Estadística

P r e s e n t a:

Silvia Rodríguez Narciso

Guanajuato, Gto. México; 16 de Diciembre de 2004.

Para mi esposo Netzahualcóyotl,
mis hijos: Aarón, Miriam y Andrea
y mis padres: Margarita y Rubén.

AGRADECIMIENTOS

Deseo expresar mi profundo agradecimiento a mi asesor, Dr. Enrique R. Villa Diharce, por su paciencia y apoyo en todas las etapas de la tesis.

A la Dra. Ivonne Solís Trápala y al Dr. Rogelio Ramos Quiroga por sus sugerencias y por ser mis sinodales.

Al CONACyT por la beca recibida para la realización de la maestría.

Al Concyteg por la beca recibida para la realización del trabajo de tesis.

Al Cimat mi gratitud, en especial a los profesores del área de Probabilidad y Estadística.

Al la Dra. Belem Trejo Valdivia por las facilidades proporcionadas para la realización de la tesis.

Al Dr. Juan Zaragoza Almaraz, por el apoyo recibido durante la realización de la tesis.

A mis hermanos: José Luis, Gloria, Leticia, Alvaro, César, Armando y Salvador.

A mis amigos: María Guadalupe Russell Noriega, Armando Domínguez Molina, Blanca Abonce Tenorio y Miguel Angel Moreles Vázquez.

Contenido

1	Introducción	1
2	Degradación	6
2.1	Modelo general de degradación	8
2.2	Distribución del tiempo a la falla	10
2.3	Función de verosimilitud	11
2.4	Modelo lineal de efectos mixtos	14
2.4.1	Propiedades Básicas	15
2.4.2	Función de verosimilitud	16
2.5	Modelos	20
2.5.1	Modelo de intercepto aleatorio	20
2.5.2	Modelo de pendiente aleatoria	26
2.5.3	Modelo de intercepto y pendiente aleatorios	27
2.6	Estimación de parámetros	28

3	Planteamiento técnico	30
3.1	Primer Criterio	33
3.2	Segundo criterio	35
4	Verosimilitud	40
4.1	Propiedades de los estimadores	41
4.2	Función Score e información observada	45
4.3	Información de Fisher esperada	46
4.4	Información de Fisher esperada: casos	48
4.4.1	Modelo con intercepto aleatorio	49
4.4.2	Modelo con pendiente aleatoria	49
4.4.3	Modelo con intercepto y pendiente aleatorios	50
4.5	Cuantil t_p del tiempo a la falla	52
4.6	Cuantiles y estimadores de varianza de \hat{t}_p : casos	55
4.6.1	Modelo de intercepto aleatorio	56
4.6.2	Modelo de pendiente aleatoria	59
4.6.3	Modelo con intercepto y pendiente aleatoria	63
5	Ejemplo real: Modelo con intercepto y pendiente aleatorios	70
5.1	Análisis de datos	71
5.2	Planeación de una prueba	82
5.2.1	Estimaciones de los parámetros y su precisión	83

5.2.2	Primer criterio para tiempos de paro	89
5.2.3	Segundo criterio para tiempos de paro	92
6	Conclusiones	97
7	Apéndice	99
7.1	Función Score e información observada	99
7.2	Información de Fisher	103
7.2.1	Modelo con intercepto aleatorio	105
7.2.2	Modelo con pendiente aleatoria	106
7.2.3	Modelo con intercepto y pendiente aleatorios	107
7.3	Vector gradiente $\frac{\partial}{\partial \phi_r} t_p(\phi_r)$	110

Capítulo 1

Introducción

Cualquier estudio de confiabilidad se lleva a cabo bajo restricciones de costos y tiempo, por lo cual es importante llevar a cabo una planeación adecuada del diseño. En la planeación se deben considerar las siguientes características (o variables): tipo de prueba, tiempo de duración de la prueba, también llamado tiempo de censura, número de unidades en muestra, y la varianza o precisión en la inferencia o predicción.

En confiabilidad es de interés estimar algunos valores de la distribución del tiempo a la falla, como son, probabilidades de falla evaluadas en cierto tiempos, cuantiles de la distribución de vida, tasas de falla; además de otras. Estas inferencias y predicciones involucran extrapolación en el tiempo.

Para los fabricantes, el conocimiento de los cuantiles, en particular, es de gran utilidad para determinar acertadamente tiempos de garantía de sus productos y también para conocer el desempeño de estos a través del tiempo. Esta información es relevante también

en la etapa de desarrollo de los productos, en donde se requiere elevar la confiabilidad hasta alcanzar y superar niveles de confiabilidad de la competencia.

En la planeación del estudio de confiabilidad se debe escoger la prueba adecuada que resuelve el problema de interés. Algunas de las prueba de vida que se utilizan en confiabilidad son las siguientes: pruebas tradicionales de vida, pruebas de vida acelerada y pruebas de degradación; entre otras.

En pruebas tradicionales de vida se registran únicamente los tiempos de falla de las unidades. Cuando la duración de esta prueba es corta, o bien cuando las unidades con las cuales se trabaja son altamente confiables, el porcentaje de unidades que fallan es pequeño o cero. En estos casos, se suelen censurar los tiempos de falla de las unidades que no fallaron.

En las pruebas de vida acelerada las unidades se someten a niveles altos de esfuerzo. El esfuerzo se da en alguna variable ambiental como temperatura, voltaje, etc., que aceleran la presencia de fallas al aumentar su valor. Las pruebas de vida acelerada permiten obtener información de la confiabilidad en un tiempo reducido.

Las pruebas de degradación son útiles en sistemas altamente confiables y cuando existe alguna característica de las unidades cuya degradación sobre el tiempo o el uso esta relacionada con la confiabilidad. Estas pruebas pueden proporcionar información sobre el tiempo de vida de productos altamente confiables y los cuales no fallan bajo pruebas de confiabilidad tradicionales.

En estudios sobre confiabilidad, es común que se quiera conocer la vida media de

los productos o un percentil de la distribución del tiempo de vida, t_p con una cierta precisión especificada con anterioridad y al menor costo. En la planeación de cualquier prueba de vida existen factores como la duración del experimento, número de unidades en la muestra, etc. que afectan la precisión de las estimaciones, así como los costos de la prueba.

En las pruebas de vida acelerada, existen criterios para determinar planes de prueba óptimos. Los criterios consideran el tiempo de duración de la prueba, el número de unidades, los niveles de esfuerzo adecuados; entre otros. En particular para este tipo de pruebas se cuenta con herramientas para determinar el tiempo de duración de la prueba.

En pruebas de degradación, existen pocos estudios para determinar criterios óptimos de planes de vida. En la literatura revisada se encontraron los siguientes artículos que tratan sobre criterios para planes de vida. En Tseng and Yu (1997) se describe una regla de paro para experimentos de degradación. En Yu y Tseng (1998) se considera una extensión del artículo de Tseng y Yu (1997), para pruebas de degradación acelerada. En Yu y Tseng (1999) y Yu y Chiao (2002) se presentan diseños de degradación óptimos, tomando como base los factores de duración del experimento, frecuencia de observación y número de unidades. En el primer artículo, bajo la restricción de que el costo del experimento no exceda una cantidad fija, se minimiza la varianza de la estimación del cuantil t_p de la distribución del tiempo de vida. En el segundo, se minimiza el costo bajo la restricción de que la probabilidad de identificar correctamente los factores que afectan la confiabilidad de los productos sea mayor a un valor dado.

Debido al reducido número de criterios que existen, este trabajo tiene como objetivo presentar criterios para determinar tiempos de paro en experimentos de degradación.

En la tesis se proponen dos criterios para determinar el tiempo de paro de un experimento de degradación. El primer criterio consiste en detener el experimento cuando la varianza estimada del estimador \hat{t}_p se estabiliza alrededor de un valor o cuando alcanza un valor deseado. El segundo criterio se fundamenta en una función de costo y consiste en parar la prueba, cuando el costo global del experimento, como función del tiempo sea mínimo. El costo global esta dado como la suma del costo económico generado por la precisión en la inferencia más el costo de experimentación. Para un mismo estudio de degradación se comparan ambos criterios.

Para probar los criterios propuestos, se simularon diferentes escenarios para tres tipos de modelos lineales de efectos mixtos. Para cada tipo de modelo se programaron los dos criterios anteriores. Los programas se construyeron utilizando funciones de S-Plus; en particular la función LME, Linear Mixed-Effects.

El trabajo que se presenta esta organizado de la siguiente manera. En el Capítulo 2 se presentan conceptos y teoría general de modelos de degradación y de modelos lineales de efectos mixtos. También se muestran los modelos que se van a utilizar en la tesis. En el Capítulo 3 se presenta el planteamiento técnico de los criterios que se proponen. En el capítulo 4 se obtiene la herramienta técnica que le da soporte a los criterios de tiempos de paro. En este mismo capítulo se obtienen los cuantiles t_p y los estimadores de las varianzas de \hat{t}_p que se requieren para aplicar los criterios a los modelos: de intercepto

aleatorio y pendiente fija, intercepto fijo y pendiente aleatoria e intercepto y pendiente aleatorios.

En el Capítulo 5, para ilustrar el procedimiento propuesto en este trabajo se aplican los criterios propuestos a un conjunto de datos reales, los cuales provienen de un estudio de confiabilidad que se llevó a cabo en MABE. En el Capítulo 6 se muestran las conclusiones y problemas abiertos.

Finalmente, en el Apéndice se presentan el desarrollo algebraico utilizado para obtener los resultados que se manejan a lo largo de la tesis.

Capítulo 2

Degradación

Las pruebas de degradación se utilizan en estudios sobre productos de alta confiabilidad que posean alguna característica que mida su desgaste en el transcurso de tiempo o el uso. Esta característica debe estar relacionada con la confiabilidad del producto.

Al contar con mediciones repetidas, el análisis de degradación aporta más información del proceso de desgaste de un producto que observar sólo los tiempos a la falla.

Por otra parte, para reducir el tiempo de observación de la degradación, es usual aplicar pruebas de vida acelerada. Ya que, aún bajo técnicas de censura, el porcentaje de unidades que fallan puede ser pequeño, aun cuando el tiempo de duración de la prueba es razonable.

La degradación física puede ser medida en forma directa como función del tiempo o de manera indirecta como función del desempeño del producto. En general, los datos obtenidos en un estudio de degradación son difíciles de modelar, ya que en el desempeño

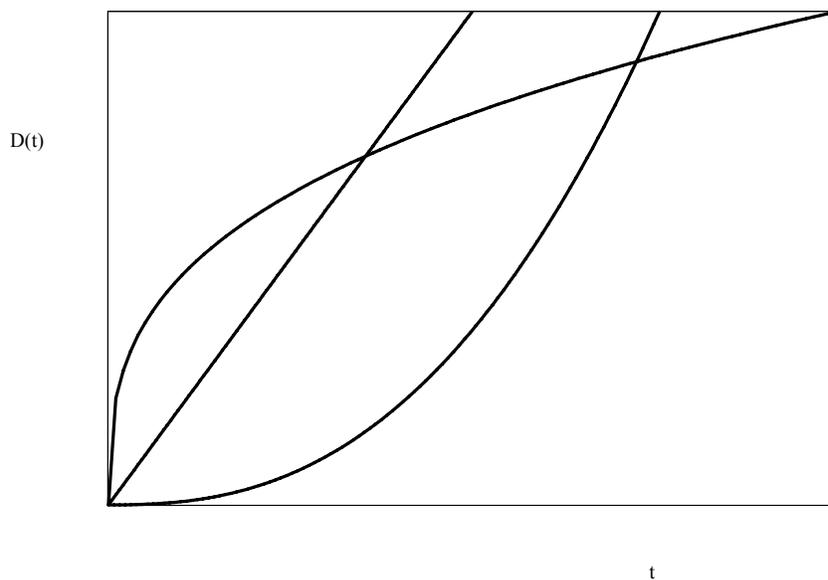


Figura 2-1: Trayectorias generales de degradación.

del producto puede existir más de una característica que este relacionada con el desgaste.

Las trayectorias de degradación de las unidades pueden tener diversas formas. Tres formas generales de curvas de degradación son: lineal, cóncava y convexa; las cuales se muestran en la Figura 2-1.

Los factores que influyen en la variabilidad de la degradación de las unidades son: condiciones iniciales, propiedades del material, condiciones de operación y ambientales y la variabilidad natural de las propias unidades.

2.1 Modelo general de degradación

Considere M unidades observadas en los tiempos $0 \leq t_{i1} < \dots < t_{in_i}$; para $i = 1, \dots, M$.

El modelo general que representa un proceso de degradación es de la forma

$$Y_{ij} = D(t_{ij}, \boldsymbol{\beta}_i) + \varepsilon_{ij}; \quad i = 1, \dots, M, \quad j = 1, \dots, n_i, \quad (2.1)$$

donde Y_{ij} es la observación j de la unidad i en el tiempo t_{ij} , ε_{ij} el término residual y $D(t, \boldsymbol{\beta})$ es una función conocida que depende del tiempo t y del vector de *efectos* $\boldsymbol{\beta}$.

La variable t puede ser el tiempo real, o tiempo de operación, o cualquier otra medida de uso adecuada. El vector de efectos $\boldsymbol{\beta}$ determina la forma de la función $D(t, \boldsymbol{\beta})$ y contiene efectos fijos (parámetros comunes a todas las unidades) o efectos aleatorios (asociados a cada unidad) o ambos. M es el número de unidades en la prueba y n_i es el total de mediciones de la i -ésima unidad. Además se asumirá que

$$\begin{aligned} \boldsymbol{\beta} &\sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \quad \text{y} \\ \varepsilon_{ij} &\sim N(0, \sigma^2), \end{aligned}$$

para cierto vector de medias $\boldsymbol{\mu}_\beta$ y matriz de covarianza $\boldsymbol{\Sigma}_\beta$, y que los errores ε_{ij} son **independientes e idénticamente distribuidos, e independientes de $\boldsymbol{\beta}$** . Para los efectos fijos de $\boldsymbol{\beta}$ se asume que tienen varianza cero. Los parámetros del modelo general (2.1) son $\boldsymbol{\mu}_\beta$, $\boldsymbol{\Sigma}_\beta$ y σ^2 . A σ^2 y a los parámetros funcionalmente independientes de $\boldsymbol{\Sigma}_\beta$, se

les conoce como componentes de varianza.

En la especificación de un modelo de degradación se requiere:

1. Determinar la forma de la función $D(t, \boldsymbol{\beta})$; la cual es sugerida por la teoría física o química.
2. Escoger la escala adecuada de la respuesta y y del tiempo t , para simplificar la forma de $D(t, \boldsymbol{\beta})$. Las trayectorias de $D(t, \boldsymbol{\beta})$ pueden ser lineales o no lineales. La transformación logaritmo para y o t permite simplificar el modelo, ya que en algunos casos reduce $D(t, \boldsymbol{\beta})$ a la forma lineal.
3. Especificar cuales de los elementos de $\boldsymbol{\beta}$ son efectos fijos y cuales son aleatorios. Para un primer análisis inferencial simple, se considera el supuesto de normalidad en los componentes aleatorios de $\boldsymbol{\beta}$. Existen otros tipos de modelos para los efectos aleatorios de $\boldsymbol{\beta}$ como la distribución lognormal, Weibull, valor extremo, entre otros.

Si el tiempo entre las mediciones es corto, es posible que los errores ε_{ij} de las degradaciones Y_{ij} , tomadas secuencialmente en la i -ésima unidad, presenten algún tipo de autocorrelación. La autocorrelación se puede reducir o hacer insignificante al alejar los tiempos de observación dentro de cada individuo, preferentemente con tiempos equidistantes. En algunos casos la autocorrelación es pequeña con respecto a la variabilidad de los efectos aleatorios. Cuando la autocorrelación entre los errores no es insignificante, se puede utilizar un modelo de series de tiempo para el término residual del modelo.

2.2 Distribución del tiempo a la falla

En las unidades de desgaste progresivo se asume que la falla ha ocurrido cuando la degradación alcanza un nivel crítico, D_f , definido previamente. Así, el tiempo de falla T se define como el tiempo cuando la trayectoria de degradación $D(t, \boldsymbol{\beta})$ cruza el nivel crítico D_f ; como se muestra en la Figura 2-2. Por lo anterior, un modelo $D(t, \boldsymbol{\beta})$ y un nivel crítico D_f definen la distribución del tiempo de falla como

$$F_T(t) = P(T \leq t) = P(D(t, \boldsymbol{\beta}) \geq D_f), \quad t > 0.$$

Esta función depende de la distribución de los efectos aleatorios $\boldsymbol{\beta}$ y de los parámetros del modelo, excepto σ^2 . En algunas ocasiones es posible obtener expresiones cerradas para $F_T(t)$, pero en general dichas expresiones no existen. Por ejemplo cuando $D(t, \boldsymbol{\beta})$ es no lineal y más de uno de los β_k son aleatorios, solo se puede evaluar a $F_T(t)$ por métodos numéricos como el método Monte Carlo. En la Sección (4.6) se obtienen expresiones analíticas de $F_T(t)$, cuando se tienen trayectorias de degradación lineales.

Por la definición de F_T , un percentil t_p de la distribución del tiempo de vida esta dado como

$$p = P(T \leq t_p) = F_T(t_p; \boldsymbol{\beta}) = P(D(t_p; \boldsymbol{\beta}) \geq D_f).$$

En secciones posteriores se mostrará la importancia que tiene un percentil en el área de confiabilidad.

Para evaluar $F_T(t)$ y hacer inferencia se requiere estimar los parámetros del mod-

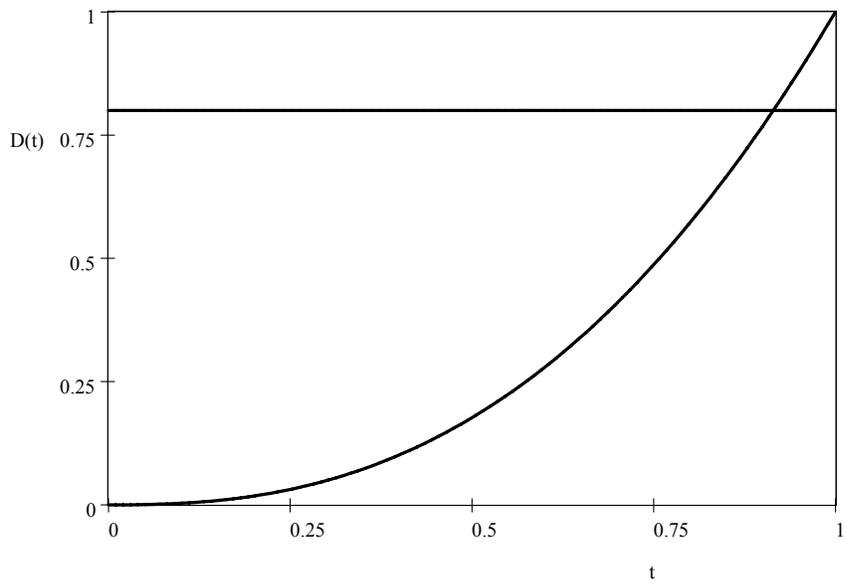


Figura 2-2: Nivel crítico de falla

elo de degradación. En este trabajo se utiliza la rutina LME de S-Plus para estimar los parámetros del modelo. Dicha rutina, utiliza el método de máxima verosimilitud o máxima verosimilitud restringida.

2.3 Función de verosimilitud

Los parámetros del modelo 2.1 son los efectos fijos β_i , el vector de medias y la matriz de covarianza de los efectos aleatorios μ_β y Σ_β y el parámetro del término residual σ . La estimación de los parámetros se puede realizar por el método de máxima verosimilitud (MV) o máxima verosimilitud restringida (MVR). Siguiendo el enfoque de Meeker y Escobar (1998), la función de verosimilitud se puede obtener de la siguiente manera.

Sea

$$\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})$$

el vector de observaciones para la unidad i . Del modelo general de degradación se tiene que

$$\mathbf{Y}_i | \boldsymbol{\beta}_i \sim N \left((D_{ij})_{j=1}^{n_i}, \sigma^2 \mathbf{I}_{n_i} \right),$$

donde

$$D_{ij} = D(t_{ij}, \boldsymbol{\beta}_i); \quad j = 1, \dots, n_i;$$

e \mathbf{I}_{n_i} es la matriz identidad de $n_i \times n_i$. Por lo que la densidad condicional de \mathbf{Y}_i dado $\boldsymbol{\beta}_i$ y la densidad marginal de \mathbf{Y}_i son

$$f_{\mathbf{Y}_i | \boldsymbol{\beta}_i}(\mathbf{y}_i | \boldsymbol{\beta}_i) = \prod_{j=1}^{n_i} \frac{1}{\sigma} \phi \left(\frac{y_{ij} - D_{ij}}{\sigma} \right)$$

y

$$\begin{aligned} f_{\mathbf{Y}_i}(\mathbf{y}_i) &= \int_{\mathbf{R}^q} f_{\mathbf{Y}_i, \boldsymbol{\beta}_i}(\mathbf{y}_i, \boldsymbol{\beta}_i) d\boldsymbol{\beta}_i = \int_{\mathbf{R}^q} f_{\mathbf{Y}_i | \boldsymbol{\beta}_i}(\mathbf{y}_i | \boldsymbol{\beta}_i) f_{\boldsymbol{\beta}_i}(\boldsymbol{\beta}_i) d\boldsymbol{\beta}_i \\ &= \frac{1}{\sigma^{n_i}} \int_{\mathbf{R}^q} \prod_{j=1}^{n_i} \phi \left(\frac{y_{ij} - D_{ij}}{\sigma} \right) f_{\boldsymbol{\beta}_i}(\boldsymbol{\beta}_i) d\boldsymbol{\beta}_i, \end{aligned}$$

respectivamente. Donde $\phi(y)$ es la densidad de la distribución Normal estándar y q es la dimensión del vector de efectos aleatorios en $\boldsymbol{\beta}_i$. Luego, la función de densidad conjunta

de las M unidades $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_M)$ es

$$f_{\mathbf{Y}}(\mathbf{y}) = \prod_{i=1}^M f_{\mathbf{Y}_i}(\mathbf{y}_i) = \frac{1}{\sigma^{n_1 + \dots + n_m}} \prod_{i=1}^M \int \prod_{j=1}^{n_i} \phi\left(\frac{y_{ij} - D_{ij}}{\sigma}\right) f_{\beta}(\beta) d\beta.$$

Entonces la función de verosimilitud esta dada como

$$L(\theta_{\beta}, \sigma; \text{datos}) = \frac{1}{\sigma^{n_1 + \dots + n_m}} \prod_{i=1}^M \int \prod_{j=1}^{n_i} \phi\left(\frac{y_{ij} - D_{ij}}{\sigma}\right) f_{\beta}(\beta; \theta_{\beta}) d\beta,$$

donde $\theta_{\beta} = (\boldsymbol{\mu}_{\beta}, \boldsymbol{\Sigma}_{\beta})$ y $f_{\beta}(\beta) = f_{\beta}(\beta; \theta_{\beta})$ es una densidad Normal multivariada. Como se puede observar, la función de verosimilitud tiene una forma compleja y computacionalmente es difícil de maximizar. Meeker y Escobar (1998), mencionan que cuando $D(t, \beta)$ es lineal, es más fácil obtener estimaciones por máxima verosimilitud.

Para obtener estimaciones por máxima verosimilitud, Pinheiro y Bates (2000) desarrollan una modificación del método de Lindstrom y Bates (1990), la cual se considera en las rutinas LME (Linear Mixed Effects) y NLME (Nonlinear Mixed Effects) que se encuentran implementadas en S-Plus.

El trabajo de tesis se enfocará en el estudio de los modelos de degradación con trayectorias lineales.

Cuando en un estudio de degradación se tienen observaciones sucesivas sobre cada una de las unidades en estudio, los datos son de tipo longitudinal. Esto implica que estos modelos de degradación caen dentro de la clase de modelos de efectos mixtos con un nivel de agrupación. Las trayectorias de degradación lineal que se estudian en este

trabajo se analizarán bajo el enfoque de modelos lineales de efectos mixtos, para un nivel de agrupación. En la siguiente sección se presenta una descripción de tales modelos.

2.4 Modelo lineal de efectos mixtos

Siguiendo el enfoque de Laird y Ware(1982) para un nivel de agrupación, el modelo lineal de efectos mixtos para el vector de respuesta \mathbf{Y}_i de la unidad i , esta dado como

$$\left\{ \begin{array}{l} \mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i \\ \mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Psi}), \\ \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2\boldsymbol{\Sigma}), \\ \mathbf{b}_1, \dots, \mathbf{b}_M, \boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_M \text{ son independientes,} \end{array} \right. \quad (2.2)$$

donde $1 \leq i \leq M$, $\boldsymbol{\beta}$ es el vector de efectos fijos de dimensión p y \mathbf{b}_i es el vector de efectos aleatorios de dimensión q . Además \mathbf{X}_i y \mathbf{Z}_i son matrices de covariables conocidas de dimensiones $n_i \times p$ y $n_i \times q$, respectivamente y $\boldsymbol{\varepsilon}_i$ es el vector de componentes del error del grupo i de dimensión n_i . Los efectos aleatorios \mathbf{b}_i no son observables. En algunos textos, a este modelo se le llama un modelo jerárquico. En el modelo se asume que la matriz de covarianzas $\boldsymbol{\Psi}$ es positiva definida. En algunas ocasiones se suele restringir la última suposición al considerar $\sigma^2\mathbf{I}_{n_i}$ en lugar de $\sigma^2\boldsymbol{\Sigma}$. En el presente trabajo se considera que los errores son independientes, lo cual permite considerar $\sigma^2\mathbf{I}_{n_i}$.

2.4.1 Propiedades Básicas

Del modelo (2.2) y dado que

$$E[\mathbf{b}_i] = \mathbf{0} \quad \text{y} \quad \text{var}[\mathbf{b}_i] = \mathbf{\Psi},$$

se sigue que

$$E[\mathbf{Y}_i | \mathbf{b}_i] = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i \quad \text{y} \quad \text{var}[\mathbf{Y}_i | \mathbf{b}_i] = \sigma^2\mathbf{I}_{n_i}.$$

Luego

$$E[\mathbf{Y}_i] = \mathbf{X}_i\boldsymbol{\beta} \quad \text{y} \quad \text{var}[\mathbf{Y}_i] = \mathbf{Z}_i\mathbf{\Psi}\mathbf{Z}_i^T + \sigma^2\mathbf{I}_{n_i}.$$

Por las propiedades de la distribución Normal se tiene que

$$\mathbf{Y}_i | \mathbf{b}_i \sim N(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \sigma^2\mathbf{I}_{n_i})$$

y

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{\Psi}\mathbf{Z}_i^T + \sigma^2\mathbf{I}_{n_i}).$$

A la distribución no condicional de \mathbf{Y}_i se le llama la representación marginal del modelo (2.2). Note que los resultados anteriores son análogos a los del modelo general de degradación.

En modelos lineales de efectos mixtos, los efectos fijos sirven para modelar la media de \mathbf{Y}_i ; mientras que los efectos aleatorios modelan la estructura de covarianza de \mathbf{Y}_i .

Estas características se pueden observar en el vector de medias y la matriz de covarianza del modelo marginal.

A partir de ahora, se denotará como $\boldsymbol{\theta}$ al vector que contiene los parámetros distintos de la matriz de covarianza $\boldsymbol{\Psi}$ y σ^2 y como $\boldsymbol{\phi} = (\boldsymbol{\beta}, \boldsymbol{\theta})$. Note que $\boldsymbol{\theta}$ contiene los componentes de varianza del modelo.

2.4.2 Función de verosimilitud

En la literatura existen diferentes formas de representar la función de verosimilitud correspondiente al modelo de efectos mixtos. La variedad de formas tienen como objetivo encontrar expresiones para la función de verosimilitud $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2; \text{datos})$ que computacionalmente sean más fáciles, o bien para mostrar propiedades de los Estimadores de Máxima Verosimilitud (EMV). En Lindstrom y Bates (1988) se presenta una forma sencilla e inmediata de representar la función de verosimilitud. Dicha forma es adecuada al tipo de modelos que se van a considerar en la tesis. La misma se describe a continuación.

Como

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^T + \sigma^2\mathbf{I}_{n_i}),$$

entonces, dada \mathbf{y}_i , la función de verosimilitud es

$$L(\boldsymbol{\phi}; \mathbf{y}_i) = \frac{1}{|\mathbf{V}_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right),$$

donde $\mathbf{V}_i = \mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^T + \sigma^2\mathbf{I}_{n_i}$.

De este modo dada la muestra aleatoria $\mathbf{y}_1, \dots, \mathbf{y}_M$ la función de verosimilitud es

$$L(\boldsymbol{\phi}; \text{datos}) = \prod_{i=1}^M \frac{1}{|\mathbf{V}_i|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})\right)$$

y la correspondiente logverosimilitud esta dada como

$$l(\boldsymbol{\phi}; \text{datos}) = -\frac{1}{2} \left(\sum_{i=1}^M \log |\mathbf{V}_i| + \sum_{i=1}^M (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right). \quad (2.3)$$

Como ya se mencionó, los criterios que se han planteado en este trabajo de tesis están en función de la información de Fisher esperada y la información observada de los parámetros del modelo. De acuerdo al material revisado para este fin, dicho cálculo resulta complicado, cuando se incrementa el número de parámetros. Bajo el modelo marginal, en el artículo de Lindstrom y Bates (1988) se presenta una forma de obtener la matriz Hessiana “perfil”, a partir de la logverosimilitud perfil de $\boldsymbol{\beta}$ y de los componentes distintos en $\boldsymbol{\Psi}$. Para obtener la información observada y la esperada, en este trabajo se sigue el enfoque de Lindstrom y Bates (1988), pero utilizando la función de logverosimilitud (2.3).

En Pinheiro y Bates(2000) se describe otra representación de $L(\boldsymbol{\phi}; \text{datos})$; la cual es compleja, pero computacionalmente es más fácil de manejar. Para obtener la función de verosimilitud de los parámetros, se “estiman” los efectos aleatorios \mathbf{b}_i , para $i = 1, \dots, M$. Para la estimación de los parámetros se utiliza el algoritmo LME; el cual considera esta función de verosimilitud.

En Pinheiro y Bates(2000) para hacer los cálculos más sencillos en la estimación de los parámetros, se considera una doble reparametrización de Ψ ; como se muestra a continuación

$$\Psi \rightarrow \left(\frac{\Psi^{-1}}{1/\sigma^2} = \Delta^T \Delta \right) \rightarrow \theta, \quad (2.4)$$

donde $\Delta^T \Delta$ representa una descomposición de Cholesky de Ψ . Además θ es un vector de los parámetros (sin restricción) que determinan a Δ . De este modo los parámetros del modelo son β y θ .

En Pinheiro y Bates (2000), para obtener la función de verosimilitud de los parámetros se considera que los efectos aleatorios no observados \mathbf{b}_i , para $i = 1, \dots, M$ son parte del modelo y que \mathbf{b}_i y ε_i son independientes. De este modo la verosimilitud $L(\phi; \text{datos})$ se puede expresar como

$$\begin{aligned} L(\phi; \text{datos}) &= \prod_{i=1}^M f_{\mathbf{Y}_i}(\mathbf{y}_i; \beta, \theta, \sigma^2) \\ &= \prod_{i=1}^M \int f_{\mathbf{Y}_i|\mathbf{b}_i}(\mathbf{y}_i; \mathbf{b}_i, \beta, \sigma^2) f_{\mathbf{b}_i}(\mathbf{b}_i; \theta) d\mathbf{b}_i. \end{aligned}$$

Dado \mathbf{b}_i , \mathbf{Y}_i tiene una distribución Normal multivariada $N(\mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i, \sigma^2\mathbf{I})$, es decir

$$f_{\mathbf{Y}_i|\mathbf{b}_i}(\mathbf{y}_i | \mathbf{b}_i, \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n_i/2}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i\|^2\right).$$

La densidad marginal \mathbf{b}_i es también una Normal multivariada:

$$\begin{aligned} f_{\mathbf{b}_i}(\mathbf{b}_i; \boldsymbol{\theta}) &= \frac{1}{(2\pi)^{q/2} |\boldsymbol{\Psi}|^{1/2}} \exp(-\mathbf{b}_i^T \boldsymbol{\Psi}^{-1} \mathbf{b}_i) \\ &= \frac{|\boldsymbol{\Delta}|}{(2\pi\sigma^2)^{q/2}} \exp\left(-\frac{1}{2\sigma^2} \|\boldsymbol{\Delta}\mathbf{b}_i\|^2\right). \end{aligned}$$

Sustituyendo la verosimilitud queda expresada como

$$\begin{aligned} &L(\boldsymbol{\phi}; \text{datos}) \\ &= \prod_{i=1}^M \frac{|\boldsymbol{\Delta}|}{(2\pi\sigma^2)^{n_i/2}} \int \frac{1}{(2\pi\sigma^2)^{q/2}} \exp\left[-\frac{1}{2\sigma^2} (\|\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i\|^2 + \|\boldsymbol{\Delta}\mathbf{b}_i\|^2)\right] d\mathbf{b}_i. \end{aligned}$$

Para obtener estimaciones de los parámetros, esta función de verosimilitud es utilizada en la función de S-Plus denominada LME (Linear Model Effects). En la función consideran dos métodos de estimación: máxima verosimilitud (MV) y máxima verosimilitud restringida (MVR). Con el método de MV los componentes de las varianzas tiende a subestimarse, esto se debe a que no se toman en cuenta los grados de libertad que se pierden al estimar los efectos fijos, $\boldsymbol{\beta}$. Para obtener la matriz de información observada se utiliza la rutina LME con el método de estimación de MV.

Para hacer inferencia sobre los parámetros, se consideran las propiedades asintóticas de los estimadores de máxima verosimilitud. En Pinheiro (1994), se muestra que bajo ciertas condiciones de regularidad, los estimadores de máxima verosimilitud del modelo lineal y de efectos mixtos, son consistentes y asintóticamente normales.

2.5 Modelos

En este trabajo los dos criterios para tiempos de paro que se proponen son analizados para modelos de degradación lineal. Específicamente los modelos de efectos mixtos que se consideran son los siguientes: modelo con intercepto aleatorio y pendiente fija, con intercepto fijo y pendiente aleatoria y el modelo con intercepto y pendiente aleatorios. Dichos modelos se presentan a continuación.

Cuando se tengan trayectorias no lineales, se buscará algún tipo de transformación para obtener trayectorias lineales. Por ejemplo, en muchos casos con la función logaritmo es posible obtener modelos lineales.

2.5.1 Modelo de intercepto aleatorio

El modelo más simple que se va a considerar es un modelo con intercepto aleatorio y pendiente fija. Este modelo se representa como

$$Y_{ij} = D(t_{ij}) + \sigma\varepsilon_{ij} = \beta_0 + \sigma_b b_i + \beta_1 t_{ij} + \sigma\varepsilon_{ij},$$

donde

β_0 es la media poblacional del intercepto,

β_1 es la pendiente,

$$b_i \sim N(0, 1), \quad \sigma_b > 0,$$

$$\varepsilon_{ij} \sim N(0, 1), \quad \sigma > 0.$$

En degradación el parámetro β_0 representa el desgaste inicial de las unidades y β_1 la tasa o velocidad de desgaste. Estos parámetros también son útiles para representar los tipos de materiales con los cuales están hechas las unidades. La colección de variables aleatorias b_i y ε_{ij} son independientes. En este caso las observaciones entre individuos son independientes, pero no dentro de cada individuo. Ya que para $j \neq k$

$$\begin{aligned} Cov(Y_{ij}, Y_{ik}) &= E(Y_{ij} - \beta_0 - \beta_1 t_{ij})(Y_{ik} - \beta_0 - \beta_1 t_{ik}) \\ &= E(\sigma_b b_i + \sigma \varepsilon_{ij})(\sigma_b b_i + \sigma \varepsilon_{ik}) \\ &= \sigma_b^2 E b_i^2 = \sigma_b^2, \\ \rho(Y_{ij}, Y_{ik}) &= \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2}. \end{aligned}$$

Observe que bajo las suposiciones del modelo, la correlación entre cualesquiera dos observaciones de la misma unidades es constante. Es decir, no depende de la distancia que existe entre ellas, ni de la unidad a la que pertenecen. Además, cuando la varianza

del error de medición es pequeña en relación a la varianza entre unidades, la correlación al interior de las unidades es cercana a uno.

Como

$$Y_{ij} | b_i \sim N(\beta_0 + \sigma_b b_i + \beta_1 t_{ij}, \sigma^2); \quad j = 1, \dots, n_i,$$

$$b_i \sim N(0, 1); \quad i = 1, \dots, m,$$

entonces

$$\mathbf{Y}_i \sim N(\beta_0 \mathbf{1}_i + \beta_1 \mathbf{t}_i, \boldsymbol{\Sigma}_i),$$

donde

$$\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}_i + \sigma_b^2 \mathbf{J}_i,$$

con $\mathbf{1}_i$ un vector de unos $n_i \times 1$ y \mathbf{J}_i es una matriz de unos de $n_i \times n_i$. Como

$$|\sigma^2 \mathbf{I}_i + \sigma_b^2 \mathbf{J}_i| = \sigma^{2(n_i-1)} (\sigma^2 + n_i \sigma_b^2) \quad \text{y}$$

$$(\sigma^2 \mathbf{I}_i + \sigma_b^2 \mathbf{J}_i)^{-1} = \frac{1}{\sigma^2} \left(\mathbf{I}_i - \frac{\sigma_b^2}{\sigma^2 + n_i \sigma_b^2} \mathbf{J}_i \right),$$

entonces la función de densidad de \mathbf{Y}_i es

$$f_{\mathbf{Y}_i}(\mathbf{y}) \propto \frac{1}{|\boldsymbol{\Sigma}_i|^{1/2}} e^{-\frac{1}{2}(\mathbf{y} - \beta_0 \mathbf{1}_i - \beta_1 \mathbf{t}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y} - \beta_0 \mathbf{1}_i - \beta_1 \mathbf{t}_i)} = \frac{1}{\sigma^{n_i-1} (\sigma^2 + n_i \sigma_b^2)^{1/2}} \times$$

$$\exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \beta_0 \mathbf{1}_i - \beta_1 \mathbf{t}_i)^T \left(\mathbf{I}_i - \frac{\sigma_b^2}{\sigma^2 + n_i \sigma_b^2} \mathbf{J}_i \right) (\mathbf{y} - \beta_0 \mathbf{1}_i - \beta_1 \mathbf{t}_i)\right).$$

$$\begin{aligned}
f_{\mathbf{Y}_i}(\mathbf{y}) &\propto \frac{1}{\sigma^{n_i-1}(\sigma^2+n_i\sigma_b^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-\beta_0\mathbf{1}_i-\beta_1\mathbf{t}_i)^T\left(\mathbf{I}_i-\frac{\sigma_b^2}{\sigma^2+n_i\sigma_b^2}\mathbf{J}_i\right)(\mathbf{y}-\beta_0\mathbf{1}_i-\beta_1\mathbf{t}_i)\right) \\
&= \frac{1}{\sigma^{n_i-1}(\sigma^2+n_i\sigma_b^2)^{1/2}} \exp\left(-\left(t_i+2y\right)\beta_0+n\beta_0^2-2\beta_1\sum_{j=1}^ny_jt_{ij}+\beta_1^2\sum_{j=1}^nt_{ij}^2\right. \\
&\quad \left.+\frac{\sigma_b^2}{\sigma^2+n_i\sigma_b^2}\left[y^2-2ny\cdot\beta_0+n^2\beta_0^2-2\beta_1(y-an)t\cdot\beta_1+t^2\beta_1^2\right]\right).
\end{aligned}$$

Note que, para $a, b, c \in \mathbf{R}$ y $\mathbf{t}, \mathbf{y} \in \mathbf{R}^n$,

$$\begin{aligned}
&(\mathbf{y}-a\mathbf{1}-b\mathbf{t})^T(\mathbf{I}-c\mathbf{J})(\mathbf{y}-a\mathbf{1}-b\mathbf{t}) \\
&= (\mathbf{y}-a\mathbf{1}-b\mathbf{t})^T(\mathbf{y}-a\mathbf{1}-b\mathbf{t})-c(\mathbf{y}-a\mathbf{1}-b\mathbf{t})^T\mathbf{J}(\mathbf{y}-a\mathbf{1}-b\mathbf{t}) \\
&= (\mathbf{y}-a\mathbf{1})^T(\mathbf{y}-a\mathbf{1})-2b(\mathbf{y}-a\mathbf{1})^T\mathbf{t}+b^2\mathbf{t}^T\mathbf{t} \\
&\quad -c(\mathbf{y}-a\mathbf{1}-b\mathbf{t})^T\mathbf{J}(\mathbf{y}-a\mathbf{1}-b\mathbf{t}) \\
&= \sum_{j=1}^n(y_j-a)^2-2b\sum_{j=1}^n(y_j-a)t_j+b^2\sum_{j=1}^nt_j^2 \\
&\quad -c\left[\left(\sum_{j=1}^ny_j-an\right)^2-2b\left(\sum_{j=1}^ny_j-an\right)\sum_{j=1}^nt_j+b^2\left(\sum_{j=1}^nt_j\right)^2\right] \\
&= \sum_{j=1}^ny_j^2-a(t+2y)+a^2n-2b\sum_{j=1}^ny_jt_j+b^2\sum_{j=1}^nt_j^2 \\
&\quad -c\left[y^2-2any+a^2n^2-2b(y-an)t+b^2t^2\right],
\end{aligned}$$

donde

$$y = \sum_{j=1}^n y_j \quad \text{y} \quad t = \sum_{j=1}^n t_j.$$

De este modo

$$\begin{aligned}
f_{\mathbf{Y}_i}(\mathbf{y}) &\propto \frac{1}{\sigma^{n_i-1}(\sigma^2+n_i\sigma_b^2)^{1/2}} e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\beta_0\mathbf{1}_i-\beta_1\mathbf{t}_i)^T\left(\mathbf{I}_i-\frac{\sigma_b^2}{\sigma^2+n_i\sigma_b^2}\mathbf{J}_i\right)(\mathbf{y}-\beta_0\mathbf{1}_i-\beta_1\mathbf{t}_i)} \\
&= \frac{1}{\sigma^{n_i-1}(\sigma^2+n_i\sigma_b^2)^{1/2}} \exp\left(- (t_i. + 2y.)\beta_0 + n\beta_0^2 + -2\beta_1\sum_{j=1}^n y_j t_{ij} + \beta_1^2\sum_{j=1}^n t_{ij}^2 \right. \\
&\quad \left. + \frac{\sigma_b^2}{\sigma^2 + n_i\sigma_b^2} [y.^2 - 2ny.\beta_0 + n^2\beta_0^2 - 2\beta_1(y. - an)t.\beta_1 + t.^2\beta_1^2] \right).
\end{aligned}$$

Considerando lo anterior, en McCulloch y Searle (2001) se obtienen expresiones analíticas para los estimadores de máxima verosimilitud para el caso balanceado. Dichas expresiones se presentan a continuación.

$$\begin{aligned}
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\bar{t}, \\
\hat{\beta}_1 &= \frac{\sum_{j=1}^n (t_j - \bar{t})^2}{\sum_{j=1}^n (t_j - \bar{t})(\bar{y}_{.j} - \bar{y})}, \\
\hat{\sigma}^2 &= \frac{1}{M(n-1)} \sum_{i=1}^M \sum_{j=1}^n \left[y_{ij} - \bar{y}_{.} - \hat{\beta}_1(x_j - \bar{x}_{.}) \right]^2 = \frac{SSR}{M(n-1)}, \\
\hat{\sigma}_b^2 &= \frac{1}{n} \left(\frac{1}{M} \sum_{i=1}^M \sum_{j=1}^n (y_{i.} - \bar{y}_{.})^2 - \hat{\sigma}^2 \right),
\end{aligned} \tag{2.5}$$

donde

$$\begin{aligned}\bar{y} &= \frac{1}{Mn} \sum_{i=1}^M \sum_{j=1}^n y_{ij} \quad \text{es el promedio global de la degradación o desgaste,} \\ \bar{t} &= \frac{1}{n} \sum_{j=1}^n t_{ij} \quad \text{es el tiempo promedio,} \\ \bar{y}_{\cdot j} &= \frac{1}{n} \sum_{i=1}^n y_{ij} \quad \text{es el promedio de desgaste para la unidad.}\end{aligned}$$

Nótese que los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ no dependen de los componentes varianza σ^2 y σ_b^2 . Esto implica que cuando se van a estimar los cuatro parámetros del modelo, primero se estiman β_0 y β_1 y después se maximiza la logverosimilitud perfil

$$l_p(\hat{\beta}_0, \hat{\beta}_1, \sigma^2, \sigma_b^2)$$

Las varianzas de los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ son

$$\begin{aligned}var\hat{\beta}_0 &= \frac{1}{m} \left(\sigma_b^2 + \frac{\sigma^2}{n_i} + \frac{\sigma^2 \bar{t}^2}{\sum_{j=1}^{n_i} (t_j - \bar{t})^2} \right), \\ var\hat{\beta}_1 &= \frac{\sigma^2}{m \sum_{j=1}^{n_i} (t_j - \bar{t})^2}.\end{aligned}$$

Para las varianzas de $\hat{\sigma}^2$ y $\hat{\sigma}_b^2$ se obtienen aproximaciones con la información de Fisher.

En las fórmulas para las varianzas se observa que σ_b^2 influye directamente en la varianza de $\hat{\beta}_0$. Además al separar e incrementar los tiempos de observación, se tiene una mayor precisión en las estimaciones de β_0 y β_1 .

Debido a que este modelo es uno de los más sencillos que se encuentran en el grupo de modelos lineales de efectos mixtos, fue sencillo obtener expresiones analíticas de los estimadores de máxima verosimilitud. Sin embargo, esto no sucede con los modelos que se presentan en seguida.

2.5.2 Modelo de pendiente aleatoria

El modelo de intercepto fijo y pendiente aleatoria es de la forma

$$Y_{ij} = D(t_{ij}) + \sigma\varepsilon_{ij} = \beta_0 + (\beta_1 + \sigma_b b_i) t_{ij} + \sigma\varepsilon_{ij},$$

donde

$$\beta_1 = \text{media poblacional de la pendiente,}$$

$$\sigma_b, \sigma > 0,$$

$$b_i \sim N(0, 1),$$

$$\varepsilon_{ij} \sim N(0, 1).$$

En forma matricial el modelo queda expresado como

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i,$$

$$\mathbf{b} \sim N(0, \sigma^2\boldsymbol{\Psi}),$$

$$\boldsymbol{\varepsilon}_i \sim N(0, \sigma^2\mathbf{I}_i),$$

donde

$$\mathbf{Z}_i = \begin{pmatrix} t_{i1} \\ t_{i2} \\ \vdots \\ t_{in_i} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}.$$

Este modelo es un poco más complejo que el anterior. En este trabajo se intentó obtener expresiones analíticas para los estimadores de MV. En la bibliografía revisada no se encontraron expresiones para los estimadores de MV.

2.5.3 Modelo de intercepto y pendiente aleatorios

El modelo que tiene intercepto y pendiente aleatorios se representa como

$$Y_{ij} = D(t_{ij}) + \sigma\varepsilon_{ij} = (\beta_0 + \sigma_0 b_{0i}) + (\beta_1 + \sigma_1 b_{1i}) t_{ij} + \sigma\varepsilon_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, m.$$

donde $\sigma, \sigma_0, \sigma_1 > 0$

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \\ \varepsilon_{ij} \sim N(0, 1),$$

con $-1 < \rho < 1$. Las variables aleatorias b_{0i}, b_{1i} son independientes de ε_{ij} . En este caso las observaciones entre individuos son independientes, más no dentro de cada individuo.

Bajo el supuesto de normalidad entre b_{0i} , y b_{1i} , la distribución de $D(t)$ es

$$D(t) \sim N(\beta_0 + \beta_1 t, \sigma_0^2 + \sigma_1^2 t^2 + 2t\rho\sigma_0\sigma_1).$$

A estos modelos también se les llama modelos de efectos aleatorios.

Este tipo de modelos representan trayectorias de degradación lineales, donde el desgaste inicial y la tasa de desgaste son diferentes para cada una de las unidades.

2.6 Estimación de parámetros

Como se mencionó anteriormente, en la estimación de parámetros se utiliza la función LME (Linear Mixed-Effects), la cual considera dos métodos de estimación: Máxima Verosimilitud (MV) y Máxima Verosimilitud Restringida (MVR). Además de estos métodos, para estimar los componentes de varianza del modelo se cuenta con el Análisis de Varianza (ANOVA). Cuando se tienen datos balanceados con el método ANOVA los estimadores de los componentes de varianza son insesgados. Si además de lo anterior se tiene normalidad de y , entonces los estimadores de los componentes son insesgados, de varianza mínima y sus varianzas muestrales están siempre disponibles. Sin embargo, cuando se tienen datos no balanceados los estimadores bajo este método no son únicos. En modelos lineales mixtos generalizados el ANOVA no proporciona resultados satisfactorios.

El método de MV es ampliamente conocido. El método de MVR considera en la función de verosimilitud funciones lineales de y , de la forma $k^T y$; donde k se especifica

de tal modo que $k^T y$ no contiene a ninguno de los efectos, los cuales también son parte del modelo de y . Es decir, en MV se reemplaza y con $k^T y$ y con ello se obtiene MVR.

Debido a la restricción de modelos bajo los cuales ANOVA proporciona resultados satisfactorios, se prefieren los métodos de MV y MVR. En Searle et. al (2001), Sección 6.8 se presentan una serie de razones por las cuales se prefieren los métodos de máxima verosimilitud sobre el método ANOVA.

Para decidir cuál de los métodos de máxima verosimilitud utilizar, a continuación se muestran las ventajas y desventajas de cada uno. Ambos métodos están fundamentados bajo el principio de máxima verosimilitud. El método de MV proporciona estimaciones de los efectos fijos, pero MVR por si mismo no. En cambio MVR presenta las siguientes ventajas sobre MV. Primero, con datos balanceados los estimadores de MVR coinciden con los estimadores de ANOVA; los cuales poseen propiedades deseables. Segundo, con MVR al estimar los componentes de varianza se consideran los grados de libertad que se pierden al estimar los efectos fijos; en cambio con MV no. Por último, dado que β no esta presente en MVR, los estimadores de los componentes de varianza son invariantes a β . Las propiedades anteriores permiten que se prefiera el método por MVR.

Capítulo 3

Planteamiento técnico

En un estudio de degradación, dos de los principales objetivos son reducir costos de experimentación y aumentar la precisión de estimación de los valores de interés; tales como el tiempo medio a la falla μ , o un cuantil de la distribución del tiempo a la falla, t_p . Estos objetivos se ven afectados por el tamaño de la muestra, la frecuencia de observación y el tiempo de terminación del experimento. Para conducir un estudio de degradación eficientemente, es necesario determinar adecuadamente dichos factores.

En la práctica, en la planeación de un experimento de degradación es común que se presente alguno de los siguientes escenarios:

1. Se fije con anterioridad el tamaño de la muestra y la frecuencia de observación.

En cuyo caso se debe determinar adecuadamente el tiempo de duración del experimento.

2. Se fije la frecuencia, entonces en el diseño de la prueba de degradación se pueden

determinar el número de unidades y el tiempo de terminación del experimento.

En Tseng y Yu (1997) se presenta una regla para determinar el tiempo de terminación de un experimento de degradación, considerando el comportamiento asintótico de un estimador de μ . En Yu y Tseng (1998) se propone una extensión del método descrito en Tseng y Yu (1997), para pruebas de degradación acelerada. Ambos artículos se basan en el comportamiento asintótico de las estimaciones de $\hat{\mu}$ al incrementar el tiempo de duración de la prueba y fijar el tamaño de muestra y la frecuencia de observación.

Por otro lado, en Yu y Tseng (1999) se plantea un diseño de degradación óptimo, bajo la restricción de que el costo total de experimentación no exceda un valor fijado con anterioridad, escogiendo los valores de los factores que minimicen la varianza de \hat{t}_p .

En una prueba de confiabilidad el interés se centra en una de la colas de la distribución del tiempo de vida, más que en hacer inferencia sobre los parámetros del modelo de degradación. En especial, se desea conocer, con una precisión alta, el tiempo en el cual falla un porcentaje p de unidades; es decir, se pretende estimar el percentil t_p con una varianza reducida de \hat{t}_p . Es por ello que los criterios siguientes se concentran en este hecho.

El primer criterio que se propone en este trabajo consiste en determinar un tiempo de paro adecuado y/o un tamaño de muestra, pero considerando el comportamiento de las estimaciones de la varianza de \hat{t}_p . El segundo criterio consiste en determinar un tiempo de paro y un tamaño de muestra adecuados que minimicen una función de costos global de la prueba.

En este punto es importante recordar que de acuerdo a la definición de la distribución del tiempo a la falla; ver Sección (2.2), un percentil t_p de la distribución no depende de σ^2 . A continuación se denotará como $\boldsymbol{\theta}_r$ al vector de componentes de varianza que no contenga a σ^2 y como $\boldsymbol{\phi}_r = (\boldsymbol{\beta}, \boldsymbol{\theta}_r)$.

Para la formulación de los criterios se asume lo siguiente:

1. Sea t_c un tiempo de paro del experimento. Se asume que es posible tener diferentes tiempos de paro del experimento. En cada tiempo de paro se estima el cuantil t_p del tiempo de vida. Un estimador intuitivo de t_p esta dado como función de los estimadores de máxima verosimilitud $\hat{\boldsymbol{\beta}}$ y $\hat{\boldsymbol{\theta}}_r$ esto es

$$t_p(\hat{\boldsymbol{\phi}}_r) = t_p(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}_r) = \hat{t}_p.$$

2. En cada t_c , se estima la varianza de \hat{t}_p , la cual denotamos como

$$\widehat{var} [\hat{t}_p]_c.$$

Como el estimador de t_p es una función de $\hat{\boldsymbol{\beta}}$ y $\hat{\boldsymbol{\theta}}_r$, para estimar $\widehat{var} [\hat{t}_p]$ se considera una aproximación, la cual se obtiene con el método delta. Para esto se supone que \hat{t}_p tiene segundas derivadas continuas con respecto a cada una de las componentes de $\hat{\boldsymbol{\beta}}$ y $\hat{\boldsymbol{\theta}}_r$. Un estimador aproximado de la varianza es de la forma

$$\widehat{var} [\hat{t}_p] \approx \frac{\partial t_p(\boldsymbol{\phi}_r)}{\partial(\boldsymbol{\phi}_r)} \widehat{var} [\hat{\boldsymbol{\phi}}_r] \frac{\partial t_p(\boldsymbol{\phi}_r)^T}{\partial(\boldsymbol{\phi}_r)},$$

donde $\frac{\partial t_p(\phi_r)}{\partial(\phi_r)}$ es el vector gradiente de $t_p(\phi_r)$, evaluado en $\hat{\phi}_r$.

En las Secciones siguientes se obtienen estimadores t_p y para $var[\hat{\phi}_r]$.

En las secciones posteriores se presenta desarrollo para obtener t_p y $var[\hat{t}_p]$ bajo los tres tipos de modelo de efectos que se encuentran en la Sección (2.5).

Es importante mencionar que la precisión de las estimaciones depende de la exactitud en las mediciones y de la dispersión (componentes de varianza) de las trayectorias de degradación.

En el segundo criterio, la función de costos que se va utilizar contiene costos de operación, costos de medición y costos de las unidades.

Como se puede ver los dos criterios para determinar el tiempo de duración adecuado en un experimento de degradación, influyen el tamaño de muestra M ; el tiempo de terminación t_c ; la varianza del estimador de t_p , $var[\hat{t}_p]$; y los costos por inferencia y observación, CTI y CTO , respectivamente. A continuación se describen detalladamente.

3.1 Primer Criterio

Intuitivamente, se espera que al incrementar el tiempo de la prueba de degradación y/o el tamaño de la muestra, las estimaciones de la varianza del cuantil decrecen; es decir, se obtienen estimaciones más precisas. Considerando esta tendencia, este criterio fija un nivel γ de $var[\hat{t}_p]$ y la prueba se detiene hasta alcanzar este nivel en función del tiempo de observación y del número de unidades. La regla de paro para este criterio es

Curvas de varianza para el cuantil 0.01

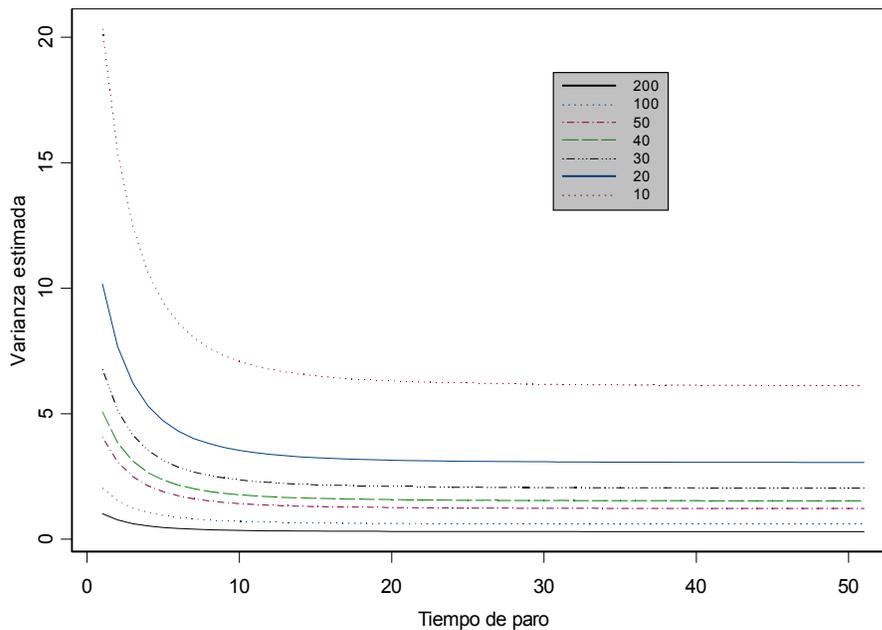


Figura 3-1: Estimaciones de la varianza del cuantil 10%, al crecer el tiempo de la prueba.

la siguiente.

$$t_{c^*} \text{ es un } t_c \text{ adecuado si } \rho_c = |\widehat{var} [\hat{t}_p]_c - \gamma| \leq \zeta; \text{ para toda } c \geq c^*,$$

donde $\zeta \equiv$ constante determinada.

Al observar el comportamiento de las estimaciones de $var [\hat{t}_p]_c$, se tiene otra variante de este criterio. El cual consiste en detener el proceso cuando el comportamiento de las estimaciones de $var [\hat{t}_p]_c$ tiende a estabilizarse; ver la Figura 3-1. Con ello se tiene una

variante de la regla de paro del para el primer criterio. La regla de paro es la siguiente

$$t_{c^*} \text{ es un } t_c \text{ adecuado si } \eta_c = \left| 1 - \frac{\text{var} [\hat{t}_p]_c}{\text{var} [\hat{t}_p]_{c-1}} \right| \leq \delta, \text{ para toda } c \geq c^*,$$

donde $\delta \equiv$ constante determinada.

3.2 Segundo criterio

El segundo criterio que se propone es útil cuando se puede diseñar un plan de prueba de vida. Se define una función de costo total CT , del estudio; la cual es función del tiempo de duración y esta dada como

$$CT(t) = CTI(t) + CTO(t),$$

donde

$$CTI(t) = K \times \text{var} [\hat{t}_p]$$

es la función que representa el costo por inferencia; y

$$CTO(t) = C_{obs} \times t_c \times M + C_{uni} \times M$$

representa el costo de observación de la prueba de vida más el costo de las unidades en muestra.

Donde K es una constante que representa el costo unitario por pérdida de precisión en la inferencia. En cualquier estudio, la precisión en la inferencia tiene un costo económico. Es claro que al incrementar la duración de la prueba se reducen las estimaciones de $var [\hat{t}_p]$ y en consecuencia la inferencia se realiza con mayor precisión. Esto implica que la toma de decisiones basada en la inferencia representa un menor costo económico. De este modo, el costo $CTI(t)$ es decreciente como función del tiempo de duración de la prueba.

Por otro lado, C_{op} es el costo unitario de observación. En este costo están incluidos los costos unitarios de operación, medición y mano de obra. Así mismo, C_{uni} representa el costo por unidad en prueba. Para un M fijo, la función $CTO(t)$ es creciente, ya que a mayor tiempo de duración, mayor es el costo total de ejecución del experimento. Este costo también es creciente como función de M .

Al sumar las funciones anteriores, se obtiene una función convexa, cuyo mínimo proporciona el costo global mínimo del experimento; ver Figura 3-2.

El segundo criterio que se propone combina los dos tipos de costo. De este modo el segundo criterio consiste en determinar el tiempo de paro que proporciona el menor costo total del estudio. La regla de paro es la siguiente

$$t_{c^*} \text{ es un } t_c \text{ adecuado si } CT(t_{c^*}) \leq CT(t_c) \text{ para toda } c \geq c^*,$$

Estos criterios surgen ante la necesidad de dar solución a los problemas que se plantean en la práctica.

Modelo 1. Funcion global de Costos.

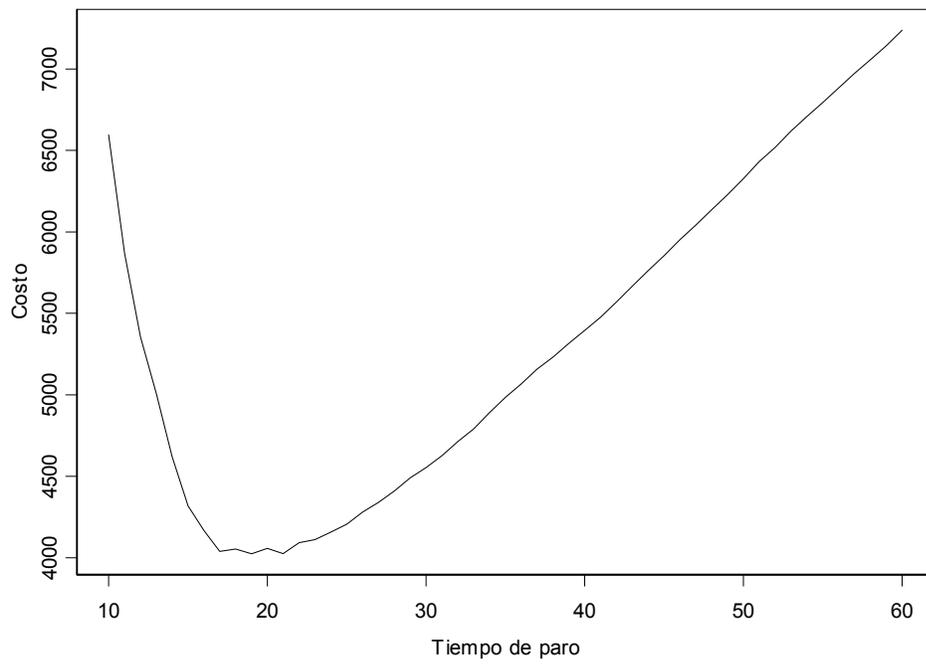


Figura 3-2: Gráfica de la función total de costo.

Como se puede observar, los criterios que se proponen en la tesis determinan tiempos de paro para diferentes tamaños de muestra, cuando la frecuencia de observación es fija y se tiene alguna estimación de los parámetros.

En la literatura revisada se encontraron algunos artículos donde se proponen procedimientos para diseñar pruebas de degradación. En Tseng y Yu (1997) proponen un método para determinar el tiempo de paro en un experimento de degradación, el cual se fundamenta en la propiedad límite del estimador de la media de vida $\hat{\mu}$, cuando el tiempo de duración de la prueba de degradación se va incrementando. El experimento se detiene cuando $\hat{\mu}$ se estabiliza.

En Yu y Tseng (1998) se presenta un procedimiento intuitivo para determinar el tiempo de paro en una prueba de degradación acelerada. El procedimiento es una combinación del enfoque que se muestra en Tseng y Yu (1997) con los modelos para pruebas de vida acelerada.

En Lu y Meeker (1993) se proponen modelos estadísticos y métodos para el análisis de datos de degradación, los cuales tienen como fin estimar una distribución del tiempo a la falla. Los modelos de degradación que se consideran son no lineales y de efectos mixtos. Los modelos trabajan bajo el supuesto de que el vector de efectos mixtos o una reparametrización de ellos siguen una distribución Normal multivariada.

En Yu y Tseng (1999) se propone un diseño óptimo para un estudio de degradación. El cual consiste en encontrar una combinación óptima de las variables: número de unidades, frecuencia de observación y tiempo de paro. Bajo la restricción de que el costo total (CT)

no exceda una cantidad específica (C), las variables óptimas de decisión (M , $freq$, y t_c) se obtienen al minimizar la varianza asintótica de un percentil estimado \hat{t}_p de la distribución del tiempo de vida.

Por último en Yu y Chiao (2002) se propone una regla de identificación bajo un diseño factorial fraccionado. Para esto se deriva un plan de prueba óptimo; el cual se obtiene al minimizar el costo total del experimento, bajo la restricción de que la probabilidad de identificar correctamente los efectos de los factores, sea mayor que una cantidad específica. La combinación óptima de las variables de decisión se determinan al minimizar el costo total, ya que la función de costo total se puede caracterizar por estas variables.

Capítulo 4

Verosimilitud

En este capítulo se presentan los elementos técnicos necesarios que se requieren en el planteamiento de los criterios de tiempos de paro ya mencionados. Como se sabe ambos criterios dependen fuertemente de las covarianzas de los estimadores de MV de los parámetros del modelo, esto es, dependen de la precisión de las estimaciones de máxima verosimilitud. En el presente capítulo, considerando las propiedades asintóticas de los estimadores de MV, se muestran dos medidas aproximadas para estimar $var [\hat{t}_p]$. Estas medidas son la información observada y la información de Fisher esperada evaluada en las estimaciones de máxima verosimilitud de los parámetros del modelo. De este modo, en la Sección (4.1) se presentan las propiedades asintóticas de los estimadores de MV, en la Sección (4.2) se obtienen la función Score y la matriz Hessiana de los parámetros del modelo lineal marginal de efectos mixtos. En la Sección (4.3) se obtiene la información esperada, las expresiones de los cuantiles de los modelos que se analizan en este

trabajo y finalmente, por el método delta, se obtienen expresiones para las varianzas de los cuantiles.

4.1 Propiedades de los estimadores

Bajo condiciones de regularidad, se tiene que el estimador de máxima verosimilitud $\hat{\phi}$ se distribuye asintóticamente como

$$\hat{\phi} \stackrel{\text{aprox}}{\sim} N(\phi, \mathbf{I}^{-1}(\phi)),$$

donde $\mathbf{I}(\phi)$ es la información esperada o información de Fisher. La información de Fisher esta dada como

$$\mathbf{I}(\phi) = -E \left[\frac{\partial^2 l(\phi)}{\partial \phi \partial \phi^T} \right],$$

donde $\frac{\partial^2 l(\phi)}{\partial \phi \partial \phi^T}$ son las segundas derivadas parciales de la función de logverosimilitud $l(\phi)$.

La información de Fisher esperada es una medida de la precisión promedio que puede ser obtenida sobre un número de repeticiones del experimento. Las segundas derivadas de la logverosimilitud indican la curvatura o concavidad de la función de verosimilitud alrededor del estimador de MV. Con tamaños de muestra grande la concavidad es mayor; lo cual se interpreta como una mayor precisión en las estimaciones.

Cuando no hay censura o no se tienen datos truncados, $\mathbf{I}(\phi)$ depende del modelo subyacente, de sus parámetros desconocidos ϕ , del tamaño de la muestra M y de las covariables.

La información de Fisher esperada es útil en la planeación de experimentos, ya que proporciona información antes de conducirlo, esto se puede ver en Fisher (1973). Por ejemplo, para la planeación de un prueba de degradación es posible que solo se conozcan la forma del modelo, pero con respecto a los parámetros se tenga información sobre un rango de posibles valores de ϕ . En esta situación la información de Fisher esperada puede ser un elemento importante para diseñar dicha prueba.

Bajo condiciones de regularidad, un estimador consistente de $\mathbf{I}(\phi)$ esta dado por la información observada,

$$\mathbf{J}(\phi) = -\frac{\partial^2 l(\phi)}{\partial \phi \partial \phi^T},$$

donde las derivadas son evaluadas en $\phi = \hat{\phi}$. También una estimación local de $\mathbf{I}(\phi)$ se obtiene al evaluar las derivadas en las estimaciones de máxima verosimilitud, como se muestra

$$\mathbf{I}(\hat{\phi}) = -E \left. \frac{\partial^2 l(\phi)}{\partial \phi \partial \phi^T} \right|_{\phi=\hat{\phi}}.$$

Este estimador es poco usual; ya que es difícil obtener los valores esperados. En Meeker y Escobar (1998) se menciona que no se tiene una ventaja clara al utilizar $\mathbf{I}(\hat{\phi})$, en lugar de $\mathbf{J}(\hat{\phi})$.

Efron y Hinkley (1978) estudian la información de Fisher esperada evaluada en el estimador de MV y la información observada como una medida de precisión de la estimación de MV de un parámetro. Desde un punto de vista frecuentista prefieren el inverso de la información observada sobre el inverso de la información esperada evaluada

en la estimación de MV como una aproximación de la varianza de $\hat{\phi}$, ya que proporciona mejores estimaciones.

Fisher (1973) enfatiza la importancia de usar la información observada después de conducir el experimento. En Díaz-Francés (1998) se menciona que en muestras finitas se debe tener precaución en usar $\mathbf{J}(\hat{\phi})$ e $\mathbf{I}(\hat{\phi})$ como aproximaciones de la varianza de $\hat{\phi}$. Además, se menciona que ninguna de estas matrices se puede usar de manera aislada como estimadores aproximados de las covarianzas de $\hat{\phi}$. Para considerar a $J(\theta)$ como una medida de la precisión, se debe considerar conjuntamente con la función de verosimilitud.

En este trabajo se consideran y analizan $\mathbf{J}(\hat{\phi})$ e $\mathbf{I}(\hat{\phi})$ como aproximaciones de la varianza de $\hat{\phi}$. En el análisis se observó que las estimaciones de la varianza de $\hat{\phi}$ con la información observada son menores que las que se obtienen con $\mathbf{I}(\hat{\phi})$. Para tamaños de muestra grandes son similares, esto se debe a que asintóticamente ambas convergen a $\mathbf{I}(\phi)$.

Tomando en cuenta que la función de verosimilitud contiene toda la información del parámetro en la muestra, se prefiere utilizar la información observada como indicador de la precisiones de la estimaciones de MV, puesto que $\mathbf{J}(\hat{\phi})$ contiene prácticamente toda la información sobre la curvatura de la función de máxima verosimilitud alrededor de $\hat{\phi}$. Esto es cuando $l(\phi; \text{datos})$ es simétrica. En otros casos, $\mathbf{J}(\hat{\phi})$ y derivadas de $l(\phi; \text{datos})$ de un orden mayor contienen la información de $l(\phi; \text{datos})$. De este modo, $\mathbf{J}(\hat{\phi})$ es una medida de precisión.

También es posible que se prefiera $\mathbf{J}(\hat{\phi})$, ya que con $\mathbf{I}(\hat{\phi})$ al integrar para obtener valores esperados puede suceder que se pierda información.

Por definición las segundas derivadas de la logverosimilitud indican la curvatura o concavidad de la función de verosimilitud alrededor del estimador de MV. Cuando el tamaño de la muestra es grande la concavidad es mayor; lo cual se interpreta como una mayor precisión en las estimaciones.

En confiabilidad se tiene interés en hacer inferencias sobre funciones de los parámetros, más que en los parámetros del modelo. Sea la función $g(\phi)$; la cual supongamos tiene segundas derivadas parciales continuas, con respecto a los elementos de ϕ . Para este tipo de funciones el estimador de MV de $g(\phi)$ es $\hat{g} = g(\hat{\phi})$, si la función es uno a uno. Por el método delta

$$var[\hat{g}] \approx \frac{\partial g(\phi)}{\partial \phi} var[\hat{\phi}] \frac{\partial g(\phi)^T}{\partial \phi},$$

donde $\frac{\partial g(\phi)}{\partial \phi}$ es el vector (o matriz) gradiente evaluado en $\hat{\phi}$. El método delta se describe en el apéndice.

En particular, el presente trabajo se concentra en la estimación de cuantiles t_p de la distribución del tiempo de vida y en las estimaciones de las varianzas de \hat{t}_p . Bajo un modelo de degradación $D(t; \beta, \theta)$, un cuantil del tiempo de vida es función de β y θ . Entonces el estimador de MV de $t_p(\phi)$ es \hat{t}_p . Para una gran variedad de modelos $D(t; \beta, \theta)$ no se tienen expresiones analíticas para $t_p(\beta, \theta)$, en cuyo caso estimaciones del cuantil se obtienen por métodos numéricos. Aquí se consideran tres tipos de modelos lineales y bajo los cuales se obtuvieron expresiones analíticas de t_p .

A continuación se obtienen la función Score y la matriz Hessiana, que se requieren para obtener estimaciones de la $var [\hat{t}_p]$.

4.2 Función Score e información observada

Como se recordará, en el modelo lineal de efectos mixtos los parámetros son $\phi = (\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)$, donde $\boldsymbol{\beta}$ es el vector de parámetros de efectos fijos, $\boldsymbol{\theta}$ al vector que contiene los parámetros distintos de la matriz de covarianza $\boldsymbol{\Psi}$.

La función de logverosimilitud del modelo lineal de efectos mixtos esta dada como

$$l(\phi; \text{datos}) = -\frac{1}{2} \left(\sum_{i=1}^M \log |\mathbf{V}_i| + \sum_{i=1}^M \left[(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] \right),$$

donde $\mathbf{V}_i(\boldsymbol{\theta}, \sigma^2) = \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \sigma^2 \mathbf{I}$.

La función Score $Sc(\phi)$ y la matriz Hessiana $\mathbf{J}(\phi)$, como se prueba en el apéndice están dadas por

$$Sc(\phi) = -\frac{1}{2} \left[\begin{array}{c} -2 \sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\beta} \mathbf{X}_i) \\ \sum_{i=1}^M \left(\text{tr} \left(\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \right) - (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{A}_{ik} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right) \end{array} \right]$$

y

$$\mathbf{J}(\boldsymbol{\phi}) = \begin{pmatrix} \sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i & \sum_{i=1}^M \mathbf{X}_i^T \mathbf{A}_{ik} \mathbf{r}_i \\ \sum_{i=1}^M \mathbf{X}_i^T \mathbf{A}_{ik} \mathbf{r}_i & \frac{1}{2} \sum_{i=1}^M \left(\text{tr} \left[\mathbf{V}_i^{-1} \frac{\partial^2 \mathbf{V}_i}{\partial \theta_h \partial \theta_k} - \mathbf{A}_{ih} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \right] - \mathbf{r}_i^T \frac{\partial \mathbf{A}_{ik}}{\partial \theta_h} \mathbf{r}_i \right) \end{pmatrix},$$

donde

$$\mathbf{r} = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}, \quad \mathbf{A}_{ik} = \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \mathbf{V}_i^{-1}$$

y

$$\frac{\partial \mathbf{A}_{ik}}{\partial \theta_h} = \mathbf{V}_i^{-1} \left(-\frac{\partial \mathbf{V}_i}{\partial \theta_h} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} + \frac{\partial^2 \mathbf{V}_i}{\partial \phi_h \partial \phi_k} - \frac{\partial \mathbf{V}_i}{\partial \theta_k} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_h} \right) \mathbf{V}_i^{-1}.$$

4.3 Información de Fisher esperada

La información de Fisher esperada, por definición es menos el valor esperado de los elementos de la matriz información observada.

Como se prueba en el apéndice, la información de Fisher esperada correspondiente al modelo lineal de efectos mixtos, queda expresada como

$$\begin{aligned} \mathbf{I}(\boldsymbol{\phi}) &= \begin{pmatrix} \sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^M \text{tr} \left[\mathbf{V}_i^{-1} \frac{\partial^2 \mathbf{V}_i}{\partial \theta_h \partial \theta_k} - \mathbf{A}_{ih} \frac{\partial \mathbf{V}_i}{\partial \theta_k} - \frac{\partial \mathbf{A}_{ik}}{\partial \theta_h} \mathbf{V}_i \right] \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^M \text{tr} \left[\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_h} \right] \end{pmatrix}, \end{aligned} \quad (4.1)$$

la cual es diagonal por bloques. En la Sección (7.2) del Apéndice se presenta el desarrollo algebraico para obtener la información de Fisher esperada.

Dado que

$$E \left[\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\theta}^T} \right] = 0 \quad \text{y} \quad E \left[\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \sigma^2} \right] = 0,$$

entonces $\boldsymbol{\beta}$ es ortogonal con $\boldsymbol{\theta}$ y σ .

Al ser $\mathbf{I}(\boldsymbol{\phi})$ diagonal por bloques implica que la matriz inversa $[\mathbf{I}(\boldsymbol{\phi})]^{-1}$ también es diagonal por bloques. Por lo tanto, el estimador de máxima verosimilitud $\hat{\boldsymbol{\beta}}$ es asintóticamente no correlacionado con $\hat{\boldsymbol{\theta}}$ y $\hat{\sigma}$.

De lo anterior también se puede concluir que

$$\hat{\boldsymbol{\beta}} \stackrel{\text{aprox}}{\sim} N \left(\boldsymbol{\beta}, \sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right)$$

y

$$\begin{pmatrix} \hat{\boldsymbol{\theta}} \\ \hat{\sigma} \end{pmatrix} \stackrel{\text{aprox}}{\sim} N \left(\begin{pmatrix} \boldsymbol{\theta} \\ \sigma \end{pmatrix}, \mathbf{I}_{sub}^{-1}(\boldsymbol{\theta}, \sigma) \right),$$

donde

$$\mathbf{I}_{sub}^{-1}(\boldsymbol{\theta}, \sigma) = \frac{1}{2} \sum_{i=1}^M \text{tr} \left[\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_h} \right],$$

es una submatriz de $\mathbf{I}(\boldsymbol{\phi})$.

Es importante mencionar que las propiedades asintóticas anteriores son válidas para los estimadores obtenidos por máxima verosimilitud, así como para los estimadores por máxima verosimilitud restringida.

Cuando se calculan intervalos de confianza, la función LME de S-Plus reparametriza las desviaciones estándar como su logaritmo y las correlaciones ρ , como logit generalizados; es decir

$$\log \sigma^2 \quad \text{y} \quad \log \frac{1 + \rho}{1 - \rho},$$

con $-1 < \rho < 1$.

Esta reparametrización se le llama parametrización natural y tiene la ventaja que los parámetros no están restringidos, ya que toman cualquier valor real. Además permite que se cumplan mejor las propiedades de normalidad en los estimadores de MV.

Dado que en este trabajo se consideran los parámetros originales del modelo y las estimaciones se obtienen con la función LME de S-Plus, entonces se realiza la transformación correspondiente al hacer inferencia sobre los parámetros originales del modelo.

4.4 Información de Fisher esperada: casos

En la tesis se analizan tres tipos de trayectorias lineales: modelo con intercepto aleatorio y pendiente fija, con intercepto fijo y pendiente aleatoria y el modelo con intercepto y pendiente aleatorios. En cada caso las unidades tienen el mismo número de mediciones, es decir, los datos están balanceados. Tomemos, $n_i = n$, para $i = 1, 2, \dots, M$. La información de Fisher esperada correspondiente a cada uno de los modelos de interés se obtiene a continuación.

4.4.1 Modelo con intercepto aleatorio

En el modelo con intercepto aleatorio la matriz de covarianza de \mathbf{Y}_i es

$$\mathbf{V}_i = \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \sigma^2 \mathbf{I} = \sigma_b^2 \mathbf{Z}_i \mathbf{Z}_i^T + \sigma^2 \mathbf{I}$$

con $\mathbf{Z}_i^T = (1, 1, \dots, 1)_{1 \times n_i}$. En este caso $\boldsymbol{\theta} = (\theta_h, \theta_k) = (\sigma_b, \sigma)$.

Sustituyendo la derivadas correspondientes en la información de Fisher esperada que se presentó en la Sección anterior, la información de Fisher esperada para los parámetros de este modelo es de la forma

$$\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \begin{pmatrix} \sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i & 0 & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^M \text{tr} [4\sigma_b^2 \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T] & \frac{1}{2} \sum_{i=1}^M \text{tr} [4\sigma_b \sigma \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1}] \\ 0 & \frac{1}{2} \sum_{i=1}^M \text{tr} [4\sigma_b \sigma \mathbf{V}_i^{-1} \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T] & \frac{1}{2} \sum_{i=1}^M \text{tr} [4\sigma^2 \mathbf{V}_i^{-1} \mathbf{V}_i^{-1}] \end{pmatrix};$$

donde $\text{tr} [4\sigma_b \sigma \mathbf{V}_i^{-1} \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T] = \text{tr} [4\sigma_b \sigma \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1}]$. El desarrollo algebraico para obtener esta matriz se presenta en el Apéndice.

4.4.2 Modelo con pendiente aleatoria

En el modelo con pendiente aleatoria la matriz de covarianza de \mathbf{Y}_i

$$\mathbf{V}_i = \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \sigma^2 \mathbf{I} = \sigma_b^2 \mathbf{Z}_i \mathbf{Z}_i^T + \sigma^2 \mathbf{I}$$

con $\mathbf{Z}_i^T = (t_1, t_2, \dots, t_{n_i})_{1 \times n_i}$. En este caso $\boldsymbol{\theta} = (\theta_h, \theta_k) = (\sigma_b, \sigma)$.

Entonces la información de Fisher esta dada como

$$\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\phi}) = \begin{pmatrix} \sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i & 0 & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^M \text{tr} [4\sigma_b^2 \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T] & \frac{1}{2} \sum_{i=1}^M \text{tr} [4\sigma_b \sigma \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1}] \\ 0 & \frac{1}{2} \sum_{i=1}^M \text{tr} [4\sigma_b \sigma \mathbf{V}_i^{-1} \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T] & \frac{1}{2} \sum_{i=1}^M \text{tr} [4\sigma^2 \mathbf{V}_i^{-1} \mathbf{V}_i^{-1}] \end{pmatrix}.$$

El desarrollo para obtener la información esperada se presenta en la Sección (7.2.2) del Apéndice.

4.4.3 Modelo con intercepto y pendiente aleatorios

En el modelo con intercepto y pendiente aleatoria, la matriz de covarianza de variable de respuesta \mathbf{Y}_i es

$$\mathbf{V}_i = \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \sigma^2 \mathbf{I},$$

donde $\mathbf{Z}_i^T = \begin{pmatrix} 1 & \dots & 1 \\ t_{i1} & \dots & t_{in_i} \end{pmatrix}$. En este caso $\boldsymbol{\Psi} = \begin{pmatrix} \sigma_0^2 & \sigma_0 \sigma_1 \rho \\ \sigma_0 \sigma_1 \rho & \sigma_1^2 \end{pmatrix}$, con $-1 < \rho < 1$ y $\boldsymbol{\theta} = (\sigma_0, \sigma_1, \rho, \sigma)$.

En forma similar a los modelos anteriores, la información esperada se obtiene al

sustituir las derivadas correspondientes en 4.1. Así

$$\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \begin{pmatrix} I_{11} & 0 & 0 & 0 & 0 \\ 0 & I_{22} & I_{23} & I_{24} & I_{25} \\ 0 & I_{32} & I_{33} & I_{34} & I_{35} \\ 0 & I_{42} & I_{43} & I_{44} & I_{45} \\ 0 & I_{52} & I_{53} & I_{54} & I_{55} \end{pmatrix},$$

donde los elementos de la matriz $\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\theta})$ están dados como

$$I_{11} = \sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i,$$

$$I_{22} = \frac{1}{2} \sum_{i=1}^M \text{tr} \left[\mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 2\sigma_0 & \sigma_1 \rho \\ \sigma_1 \rho & 0 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 2\sigma_0 & \sigma_1 \rho \\ \sigma_1 \rho & 0 \end{pmatrix} \mathbf{Z}_i^T \right],$$

$$I_{33} = \frac{1}{2} \sum_{i=1}^M \text{tr} \left[\mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \rho \\ \sigma_0 \rho & 2\sigma_1 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \rho \\ \sigma_0 \rho & 2\sigma_1 \end{pmatrix} \mathbf{Z}_i^T \right],$$

$$I_{44} = \frac{1}{2} \sum_{i=1}^M \text{tr} \left[\mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \sigma_1 \\ \sigma_0 \sigma_1 & 0 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \sigma_1 \\ \sigma_0 \sigma_1 & 0 \end{pmatrix} \mathbf{Z}_i^T \right],$$

$$I_{55} = \frac{1}{2} \sum_{i=1}^M \text{tr} [4\sigma^2 \mathbf{V}_i^{-1} \mathbf{V}_i^{-1}],$$

$$\begin{aligned}
I_{23} = I_{32} &= \frac{1}{2} \sum_{i=1}^M \text{tr} \left[\mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 2\sigma_0 & \sigma_1\rho \\ \sigma_1\rho & 0 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0\rho \\ \sigma_0\rho & 2\sigma_1 \end{pmatrix} \mathbf{Z}_i^T \right], \\
I_{24} = I_{42} &= \frac{1}{2} \sum_{i=1}^M \text{tr} \left[\mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 2\sigma_0 & \sigma_1\rho \\ \sigma_1\rho & 0 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0\sigma_1 \\ \sigma_0\sigma_1 & 0 \end{pmatrix} \mathbf{Z}_i^T \right], \\
I_{25} = I_{52} &= \frac{1}{2} \sum_{i=1}^M \text{tr} \left[2\sigma \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 2\sigma_0 & \sigma_1\rho \\ \sigma_1\rho & 0 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \right], \\
I_{34} = I_{43} &= \frac{1}{2} \sum_{i=1}^M \text{tr} \left[\mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0\sigma_1 \\ \sigma_0\sigma_1 & 0 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0\rho \\ \sigma_0\rho & 2\sigma_1 \end{pmatrix} \mathbf{Z}_i^T \right], \\
I_{35} = I_{53} &= \frac{1}{2} \sum_{i=1}^M \text{tr} \left[2\sigma \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0\rho \\ \sigma_0\rho & 2\sigma_1 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \right], \\
I_{45} = I_{54} &= \frac{1}{2} \sum_{i=1}^M \text{tr} \left[2\sigma \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0\sigma_1 \\ \sigma_0\sigma_1 & 0 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \right].
\end{aligned}$$

Las matrices de derivadas parciales de \mathbf{V}_i se encuentran en la Sección (7.2.3) del Apéndice.

4.5 Cuantil t_p del tiempo a la falla

Un cuantil t_p de la distribución del tiempo a la falla se define como,

$$p = P(T \leq t_p) = F_T(t_p; \phi_r) = P(D(t_p; \phi_r) \geq D_f).$$

El interés en este trabajo se centra en la estimación de cuantiles inferiores, $0 < p < 0.5$. Algunas veces t_p no tiene forma cerrada, afortunadamente para los modelos que aquí se analizan se obtuvieron formas cerradas para t_p .

Note que en la definición de cuantil no interviene el parámetro σ . Recordemos que ϕ_r contiene los parámetros del modelo, excepto a σ .

Por la propiedad de invarianza de los estimadores de MV, el estimador de MV de t_p es

$$\hat{t}_p = t_p(\hat{\phi}_r).$$

Al ser un estimador de MV posee las propiedades asintóticas de este tipo de estimadores. Así, para muestras grandes \hat{t}_p se distribuye aproximadamente como Normal con media t_p y matriz de covarianza $var[\hat{t}_p]$.

Como la varianza de \hat{t}_p no se puede obtener en forma directa, se obtiene una aproximación con el método delta. Este método se puede aplicar siempre que t_p como función de los parámetros tenga segundas derivadas continuas con respecto a los elementos de ϕ_r . Suponiendo lo anterior, se tiene que

$$var[\hat{t}_p] \approx \left[\frac{\partial t_p(\phi_r)}{\partial \phi_r} \right] var[\hat{\phi}_r] \left[\frac{\partial t_p(\phi_r)}{\partial \phi_r} \right]^T, \quad (4.2)$$

donde $\frac{\partial t_p(\phi_r)}{\partial \phi_r}$ es el vector gradiente de $t_p(\phi_r)$, evaluado en $\hat{\phi}_r$.

Observe que una estimación de $var[\hat{t}_p]$ queda determinada al estimar $var[\hat{\phi}_r]$. De los resultados obtenidos en la Sección (4.3) para los modelos lineales de efectos mixtos,

dos estimadores aproximados de $\text{var} [\hat{\phi}_r]$ son:

$$\widehat{\text{var}}_1 (\hat{\phi}_r) = \mathbf{J} (\hat{\phi}_r)^{-1},$$

y

$$\widehat{\text{var}}_2 (\hat{\phi}_r) = \mathbf{I} (\phi_r)^{-1} |_{\phi_r = \hat{\phi}_r},$$

donde $\mathbf{J} (\phi_r) = \left[-\frac{\partial^2 l(\phi_r)}{\partial \phi_r \partial \phi_r^T} \right]$ es la submatriz de $\mathbf{J} (\phi)$ que no contiene las derivadas con respecto a σ^2 e $\mathbf{I} (\phi_r)$ es la submatriz de $\mathbf{I} (\phi)$ que no contiene los valores esperados de las derivadas con respecto a σ^2 . Al sustituir estos estimadores en 4.2 se generan las dos estimadores locales de $\text{var} [\hat{t}_p]$,

$$\text{var}_1 [\hat{t}_p] = \frac{\partial t_p (\phi_r)}{\partial \phi_r} \mathbf{J} (\hat{\phi}_r)^{-1} \frac{\partial t_p (\phi_r)^T}{\partial \phi_r}$$

y

$$\text{var}_2 [\hat{t}_p] = \frac{\partial t_p (\phi_r)}{\partial \phi_r} \mathbf{I} (\phi_r)^{-1} |_{\phi_r = \hat{\phi}_r}^T \frac{\partial t_p (\phi_r)}{\partial \phi_r},$$

con

$$\mathbf{J} (\hat{\phi}_r) = \left(\begin{array}{cc} \sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i & \sum_{i=1}^M \mathbf{X}_i^T \mathbf{A}_{ik} \mathbf{r}_i \\ \sum_{i=1}^M \mathbf{X}_i^T \mathbf{A}_{ik} \mathbf{r}_i & \frac{1}{2} \sum_{i=1}^M \left(\text{tr} \left[\mathbf{V}_i^{-1} \frac{\partial^2 \mathbf{V}_i}{\partial \phi_h \partial \phi_k} - \mathbf{A}_{ih} \frac{\partial \mathbf{V}_i}{\partial \phi_k} \right] - \mathbf{r}_i^T \frac{\partial \mathbf{A}_{ik}}{\partial \phi_h} \mathbf{r}_i \right) \end{array} \right) \Bigg|_{\phi_r = \hat{\phi}_r}$$

y

$$\mathbf{I}(\boldsymbol{\phi}_r) = \begin{pmatrix} \sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^M \text{tr} \left[\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \phi_k} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \phi_h} \right] \end{pmatrix},$$

donde las derivadas parciales son evaluadas en $\hat{\boldsymbol{\phi}}_r$ y donde

$$\mathbf{r} = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}, \quad \mathbf{A}_{ik} = \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \mathbf{V}_i^{-1}$$

y

$$\frac{\partial \mathbf{A}_{ik}}{\partial \theta_h} = \mathbf{V}_i^{-1} \left(-\frac{\partial \mathbf{V}_i}{\partial \theta_h} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} + \frac{\partial^2 \mathbf{V}_i}{\partial \phi_h \partial \phi_k} - \frac{\partial \mathbf{V}_i}{\partial \theta_k} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_h} \right) \mathbf{V}_i^{-1}.$$

Las estimaciones de $\text{var}_1[\hat{t}_p]$ y $\text{var}_2[\hat{t}_p]$ se obtienen al evaluar las expresiones anteriores en las estimaciones de MV del vector de parámetros $\boldsymbol{\phi}_r$. Mediante estudios de simulación se estimarán y se compararán ambas varianzas.

En la siguiente sección se obtienen expresiones para los cuantiles del tiempo a la falla; así como las varianzas de los \hat{t}_p , para cada uno de los modelos de degradación en estudio.

4.6 Cuantiles y estimadores de varianza de \hat{t}_p : casos

Dado que cada uno de los modelos que aquí se estudian tiene sus propias características, fue necesario obtener las expresiones del cuantil t_p y la varianza de su estimador $\text{var}_{t_p}(\hat{\boldsymbol{\phi}}_r)$ para cada modelo. Para las simulaciones de los criterios de tiempos de duración se creó un programa para cada caso. A continuación se obtiene las expresiones de

los cuantiles y las varianzas aproximadas.

4.6.1 Modelo de intercepto aleatorio

Dado el modelo con intercepto aleatorio y pendiente fija,

$$Y_{ij} = \beta_0 + \sigma_0 b_i + \beta_1 t_{ij} + \sigma \varepsilon_{ij},$$

sea T el tiempo de falla para un nivel de falla $D_f > \beta_0$. Luego

$$\begin{aligned} F_T(t) &= P(T \leq t) = P(\beta_0 + \sigma_0 b + \beta_1 t \geq D_f) \\ &= P\left(b \geq \frac{D_f - \beta_0 - \beta_1 t}{\sigma_0}\right) = 1 - \Phi\left(\frac{D_f - \beta_0 - \beta_1 t}{\sigma_0}\right) \\ &= \Phi\left(\frac{\beta_0 + \beta_1 t - D_f}{\sigma_0}\right) = \Phi\left(\frac{\beta_1}{\sigma_0} \left(t - \frac{D_f - \beta_0}{\beta_1}\right)\right). \end{aligned}$$

Entonces la densidad de T es

$$f_T(t) = \frac{1}{\sqrt{2\pi}} \frac{\sigma_0}{\beta_1} e^{-\frac{1}{2} \frac{\sigma_0^2}{\beta_1^2} \left(t - \frac{D_f - \beta_0}{\beta_1}\right)^2},$$

la cual resulta ser una Normal:

$$T \sim N\left(\frac{D_f - \beta_0}{\beta_1}, \frac{\sigma_0^2}{\beta_1^2}\right).$$

La distribución de T depende de tres parámetros: β_0 , β_1 y σ_0 . Note que, bajo este modelo, $P(T \leq 0) = \Phi\left(-\frac{D_f - \beta_0}{\sigma_0}\right) > 0$. Por lo que se espera que $D_f \gg \beta_0 + \sigma_0$. El modelo es bueno si la componente aleatoria del intercepto es pequeña con respecto a la componente fija y que $D_f > \beta_0$. Pues si esto se cumple, $P(T \leq 0)$ se puede ser insignificante.

Sea t_p un cuantil de interés, con $0 < p < 1$. Luego,

$$t_p = \frac{D_f - \beta_0 + \sigma_0 z_p}{\beta_1}.$$

Entonces por la propiedad de invarianza el estimador de MV de t_p es

$$t_p(\hat{\phi}_r) = \frac{D_f - \hat{\beta}_0 + \hat{\sigma}_0 z_p}{\hat{\beta}_1}.$$

Por el método delta una aproximación de la $\text{var}_{t_p}(\hat{\phi}_r)$ se obtiene de la siguiente forma.

Sea

$$g(\beta_0, \beta_1, \sigma_0) = t_p(\phi_r),$$

luego,

$$\frac{\partial t_p}{\partial(\beta_0, \beta_1, \sigma_0)} = \left(-\frac{1}{\beta_1}, -\frac{D_f - \beta_0 + \sigma_0 z_p}{\beta_1^2}, \frac{z_p}{\beta_1} \right).$$

Entonces por el método delta,

$$\text{var} \hat{t}_p = \begin{pmatrix} -\frac{1}{\beta_1} \\ -\frac{D_f - \beta_0 + \sigma_0 z_p}{\beta_1^2} \\ \frac{z_p}{\beta_1} \end{pmatrix}^T \mathbf{I}^{-1}(\beta_0, \beta_1, \sigma_0) \begin{pmatrix} -\frac{1}{\beta_1} \\ -\frac{D_f - \beta_0 + \sigma_0 z_p}{\beta_1^2} \\ \frac{z_p}{\beta_1} \end{pmatrix},$$

donde las derivadas son evaluadas en $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_0)$ e

$$\mathbf{I}(\beta_0, \beta_1, \sigma_0) = \begin{pmatrix} \sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^M \text{tr} [4\sigma_0^2 \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T] \end{pmatrix}.$$

El elemento $\sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i$ es una matriz de 2×2 y $\mathbf{Z}_i^T = (1, 1, \dots, 1)_{1 \times n}$.

Los estimadores de $\text{var} [\hat{t}_p]$ que se presentaron en la Sección (4.4) para el modelo con intercepto aleatorio quedan expresados como

$$\widehat{\text{var}}_1 \hat{t}_p = \begin{pmatrix} -\frac{1}{\hat{\beta}_0} \\ -\frac{D_f - \hat{\beta}_0 + \hat{\sigma}_0 z_p}{\hat{\beta}_1^2} \\ \frac{z_p}{\hat{\beta}_1} \end{pmatrix}^T \mathbf{J}^{-1}(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_0) \begin{pmatrix} -\frac{1}{\hat{\beta}_0} \\ -\frac{D_f - \hat{\beta}_0 + \hat{\sigma}_0 z_p}{\hat{\beta}_1^2} \\ \frac{z_p}{\hat{\beta}_1} \end{pmatrix}$$

y

$$\widehat{var}_2 \hat{t}_p = \begin{pmatrix} -\frac{1}{\hat{\beta}_0} \\ -\frac{D_f - \hat{\beta}_0 + \hat{\sigma}_0 z_p}{\hat{\beta}_1^2} \\ \frac{z_p}{\hat{\beta}_1} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^m \mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^m \text{tr} \left[4\hat{\sigma}_0^2 \hat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \right] \end{pmatrix}^{-1} \\ \times \begin{pmatrix} -\frac{1}{\hat{\beta}_0} \\ -\frac{D_f - \hat{\beta}_0 + \hat{\sigma}_0 z_p}{\hat{\beta}_1^2} \\ \frac{z_p}{\hat{\beta}_1} \end{pmatrix},$$

donde $\hat{\mathbf{V}}_i(\hat{\boldsymbol{\theta}}) = \hat{\sigma}_0^2 \mathbf{Z}_i \mathbf{Z}_i^T + \hat{\sigma}^2 \mathbf{I}$.

4.6.2 Modelo de pendiente aleatoria

Dado el modelo con intercepto fijo y pendiente aleatoria

$$Y_{ij} = \beta_0 + (\beta_1 + \sigma_1 b_i) t_{ij} + \sigma \varepsilon_{ij},$$

donde

$$\beta_0, \beta_1, \sigma_1, \sigma > 0, \quad b_i \sim N(0, 1), \quad \varepsilon_{ij} \sim N(0, 1)$$

Sea T la variable aleatoria que representa el tiempo a la falla para un nivel crítico de falla D_f ; con $D_f > \beta_0$. Una unidad falla cuando la trayectoria de degradación $D(t, \theta) \geq D_f$.

Luego la distribución del tiempo a la falla se puede obtener como

$$\begin{aligned}
 P(T \leq t) &= P(\beta_0 + (\beta_1 + \sigma_1 b)t \geq D_f) \\
 &= P\left(b \geq \frac{D_f - \beta_0 - \beta_1 t}{\sigma_1 t}\right) \\
 &= 1 - \Phi\left(\frac{D_f - \beta_0 - \beta_1 t}{\sigma_1 t}\right) \\
 &= \Phi\left(\frac{\beta_1 t - (D_f - \beta_0)}{\sigma_1 t}\right) = \Phi\left(\frac{\beta_1}{\sigma_1} - \frac{D_f - \beta_0}{\sigma_1} \frac{1}{t}\right).
 \end{aligned}$$

Note que bajo este modelo, $P(T < \infty) = \Phi\left(\frac{\beta_1}{\sigma_1}\right)$. Por lo que se espera $\beta_1 \gg \sigma_1$ para que la distribución anterior sea una buena aproximación. Por ejemplo, $\beta_1 \geq 5\sigma_1$. Esto quiere decir que el modelo de pendiente aleatoria es bueno siempre que la componente aleatoria sea significativamente menor que la componente fija.

La densidad de T es

$$f_T(t) = \frac{D_f - \beta_0}{\sigma_1} \phi\left(\frac{\beta_1}{\sigma_1} - \frac{D_f - \beta_0}{\sigma_1} \frac{1}{t}\right)$$

A este tipo de distribuciones se les conoce como distribuciones de Berstein.

A partir del cuantil z_p de la Normal estándar, se puede obtener el cuantil t_p del tiempo a la falla, de la manera siguiente

$$p = \Phi(z_p) = \Phi\left(\frac{\beta_1}{\sigma_1} - \frac{D_f - \beta_0}{\sigma_1} \frac{1}{t_p}\right).$$

Esto implica que

$$z_p = \frac{\beta_1}{\sigma_1} - \frac{D_f - \beta_0}{\sigma_1} \frac{1}{t_p}.$$

Despejando, t_p queda expresado como

$$t_p = \frac{D_f - \beta_0}{\beta_1 - z_p \sigma_1},$$

bajo la restricción de $D_f \geq \beta_0$. Por la propiedad de invarianza el estimador de MV de t_p

es

$$t_p(\hat{\phi}_r) = \frac{D_f - \hat{\beta}_0}{\hat{\beta}_1 - z_p \hat{\sigma}_1}.$$

Por el método delta una aproximación de la $\text{var} t_p(\hat{\phi}_r)$ se obtiene de la siguiente forma.

Sea

$$g(\beta_0, \beta_1, \sigma_1) = t_p(\phi_r),$$

luego el gradiente es

$$\frac{\partial t_p}{\partial (\beta_0, \beta_1, \sigma_1)} = \begin{pmatrix} -\frac{1}{\beta_1 - z_p \sigma_1} \\ -\frac{D_f - \beta_0}{(\beta_1 - z_p \sigma_1)^2} \\ \frac{(D_f - \beta_0) z_p}{(\beta_1 - z_p \sigma_1)^2} \end{pmatrix}^T.$$

Entonces

$$\text{var} [\hat{t}_p] = \begin{pmatrix} -\frac{1}{\beta_1 - z_p \sigma_1} \\ -\frac{D_f - \beta_0}{(\beta_1 - z_p \sigma_1)^2} \\ \frac{(D_f - \beta_0) z_p}{(\beta_1 - z_p \sigma_1)^2} \end{pmatrix}^T \mathbf{I}^{-1}(\beta_0, \beta_1, \sigma_1) \begin{pmatrix} -\frac{1}{\beta_1 - z_p \sigma_1} \\ -\frac{D_f - \beta_0}{(\beta_1 - z_p \sigma_1)^2} \\ \frac{(D_f - \beta_0) z_p}{(\beta_1 - z_p \sigma_1)^2} \end{pmatrix},$$

donde las derivadas son evaluadas en $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_1)$ y $\mathbf{Z}_i^T = (t_1, t_2, \dots, t_{n_i})$. Con

$$\mathbf{I}(\beta_0, \beta_1, \sigma_1) = \begin{pmatrix} \sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^M \text{tr} [4\sigma_1^2 \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T] \end{pmatrix}.$$

Los estimadores de $\text{var} [\hat{t}_p]$ que se presentaron en la Sección (4.5) para el modelo con pendiente aleatoria quedan expresados como

$$\widehat{\text{var}}_1 \hat{t}_p = \begin{pmatrix} -\frac{1}{\hat{\beta}_0} \\ -\frac{D_f - \hat{\beta}_0 + \hat{\sigma}_1 z_p}{\hat{\beta}_1^2} \\ \frac{z_p}{\hat{\beta}_1} \end{pmatrix}^T \mathbf{J}^{-1}(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_1) \begin{pmatrix} -\frac{1}{\hat{\beta}_0} \\ -\frac{D_f - \hat{\beta}_0 + \hat{\sigma}_1 z_p}{\hat{\beta}_1^2} \\ \frac{z_p}{\hat{\beta}_1} \end{pmatrix}$$

y

$$\begin{aligned} \widehat{\text{var}}_2 \hat{t}_p &= \begin{pmatrix} -\frac{1}{\hat{\beta}_0} \\ -\frac{D_f - \hat{\beta}_0 + \hat{\sigma}_1 z_p}{\hat{\beta}_1^2} \\ \frac{z_p}{\hat{\beta}_1} \end{pmatrix}^T \left(\begin{pmatrix} \sum_{i=1}^M \mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^M \text{tr} [4\hat{\sigma}_1^2 \hat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T] \end{pmatrix}^{-1} \right) \\ &\times \begin{pmatrix} -\frac{1}{\hat{\beta}_0} \\ -\frac{D_f - \hat{\beta}_0 + \hat{\sigma}_1 z_p}{\hat{\beta}_1^2} \\ \frac{z_p}{\hat{\beta}_1} \end{pmatrix}, \end{aligned}$$

donde $\hat{\mathbf{V}}_i = \hat{\sigma}_1^2 \mathbf{Z}_i \mathbf{Z}_i^T + \hat{\sigma}^2 \mathbf{I}$.

4.6.3 Modelo con intercepto y pendiente aleatoria

Dado el modelo para trayectorias lineales con intercepto y pendiente aleatorias

$$Y_{ij} = D(t_{ij}) + \sigma\varepsilon_{ij} = (\beta_0 + \sigma_0 b_{0i}) + (\beta_1 + \sigma_1 b_{1i})t_{ij} + \sigma\varepsilon_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, M.$$

donde

$$\sigma, \sigma_0, \sigma_1 > 0, \quad \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right); \quad \varepsilon_{ij} \sim N(0, 1).$$

Aquí las variables aleatorias b_{0i} y b_{1i} son independientes de ε_{ij} . En este caso las observaciones entre individuos son independientes, más no dentro de cada individuo.

Bajo el supuesto de normalidad entre b_{0i} , y b_{1i} , la distribución de $D(t)$ es

$$D(t) \sim N(\beta_0 + \beta_1 t, \sigma_0^2 + \sigma_1^2 t^2 + 2t\rho\sigma_0\sigma_1).$$

Dado el nivel crítico de falla D_f , la distribución del tiempo a la falla se puede obtener

como

$$\begin{aligned}
F_T(t) &= P(\beta_0 + \beta_1 t + \sigma_0 b^0 + \sigma_1 b^1 t \geq D_f) \\
&= P(\sigma_0 b^0 + \sigma_1 b^1 t \geq D_f - \beta_0 - \beta_1 t) \\
&= P\left(\frac{\sigma_0 b_0 + \sigma_1 b_1 t}{\sqrt{\sigma_0^2 + \sigma_1^2 t^2 + 2t\rho\sigma_0\sigma_1}} > \frac{D_f - \beta_0 - \beta_1 t}{\sqrt{\sigma_0^2 + \sigma_1^2 t^2 + 2t\rho\sigma_0\sigma_1}}\right) \\
&= P\left(\frac{\sigma_0 b_0 + \sigma_1 b_1 t}{\sqrt{\sigma_0^2 + \sigma_1^2 t^2 + 2t\rho\sigma_0\sigma_1}} < \frac{\beta_0 + \beta_1 t - D_f}{\sqrt{\sigma_0^2 + \sigma_1^2 t^2 + 2t\rho\sigma_0\sigma_1}}\right) \\
&\approx \Phi\left(\frac{\beta_0 + \beta_1 t - D_f}{\sqrt{\sigma_0^2 + \sigma_1^2 t^2 + 2t\rho\sigma_0\sigma_1}}\right). \tag{4.3}
\end{aligned}$$

De hecho la aproximación

$$F_T(t) \approx \Phi\left(\frac{\beta_0 + \beta_1 t - D_f}{\sqrt{\sigma_0^2 + \sigma_1^2 t^2 + 2t\rho\sigma_0\sigma_1}}\right)$$

es buena si $(D_f - \beta_0)/\sigma_0$ y β_1/σ_1 son grandes. Por ejemplo

$$\frac{D_f - \beta_0}{\sigma_0}, \frac{\beta_1}{\sigma_1} > 3. \tag{4.4}$$

Lo anterior se debe a que en $t = 0$ y para $t \rightarrow \infty$, se debe cumplir

$$F_T(0) = 0 \approx \Phi\left(-\frac{D_f - \beta_0}{\sigma_0}\right)$$

y

$$\begin{aligned}
F_T(\infty) &= \lim_{t \rightarrow \infty} F_T(t) = 1 \\
&\approx \lim_{t \rightarrow \infty} \Phi \left(\frac{\beta_0 + \beta_1 t - D_f}{\sqrt{\sigma_0^2 + \sigma_1^2 t^2 + 2t\sigma_0\sigma_1\rho}} \right) \\
&= \lim_{t \rightarrow \infty} \Phi \left(\frac{\beta_0 - D_f}{\sqrt{\sigma_0^2 + \sigma_1^2 t^2 + 2t\sigma_0\sigma_1\rho}} + \frac{\beta_1 t}{\sqrt{\sigma_0^2 + \sigma_1^2 t^2 + 2t\sigma_0\sigma_1\rho}} \right) \\
&= \lim_{t \rightarrow \infty} \Phi \left(\frac{\beta_1}{\sqrt{\frac{\sigma_0^2}{t^2} + \frac{2\sigma_0\sigma_1\rho}{t} + \sigma_1^2}} \right) \\
&= \Phi \left(\frac{\beta_1}{\sigma_1} \right).
\end{aligned}$$

Esto significa que en la práctica la distribución del tiempo a la falla es la expresión (4.3), siempre que la variabilidad explicada de la degradación por los términos comunes de las unidades es mucho mayor que la variabilidad explicada por los términos dentro de las unidades.

Como la varianza de $D(t_{ij})$ depende del tiempo, entonces la distribución de T no es una distribución Normal. A este tipo de distribuciones se les conoce como distribuciones de *Bernstein*.

En el presente trabajo se requiere estimar un cuantil t_p inferior de T . El cuantil se puede obtener de la distribución aproximada para T , como se muestra en seguida. Dado que

$$p = F_T(t_p) = \Phi \left(\frac{\beta_0 + \beta_1 t_p - D_f}{\sqrt{\sigma_0^2 + \sigma_1^2 t_p^2 + 2t_p\sigma_0\sigma_1\rho}} \right) = \Phi(z_p),$$

entonces

$$z_p = \frac{\beta_0 + \beta_1 t_p - D_f}{\sqrt{\sigma_0^2 + \sigma_1^2 t_p^2 + 2t_p \sigma_0 \sigma_1 \rho}} \leq 0.$$

De lo anterior se obtiene la ecuación

$$(\beta_1^2 - z_p^2 \sigma_1^2) t_p^2 - 2 [(D_f - \beta_0) \beta_1 + z_p^2 \sigma_0 \sigma_1 \rho] t_p + (D_f - \beta_0)^2 - z_p^2 \sigma_0^2 = 0.$$

Resolviendo la ecuación para t_p se tiene

$$\begin{aligned} t_p &= \frac{(D_f - \beta_0) \beta_1 + z_p^2 \sigma_0 \sigma_1 \rho}{\beta_1^2 - z_p^2 \sigma_1^2} \\ &= \pm \frac{\sqrt{[(D_f - \beta_0) \beta_1 + z_p^2 \sigma_0 \sigma_1 \rho]^2 - (\beta_1^2 - z_p^2 \sigma_1^2) [(D_f - \beta_0)^2 - z_p^2 \sigma_0^2]}}{\beta_1^2 - z_p^2 \sigma_1^2} \end{aligned} \quad (4.5)$$

Aquí hay dos restricciones naturales

$$0 < t_p < \frac{D_f - \beta_0}{\beta_1}$$

y

$$[(D_f - \beta_0) \beta_1 + z_p^2 \sigma_0 \sigma_1 \rho]^2 > (\beta_1^2 - z_p^2 \sigma_1^2) [(D_f - \beta_0)^2 - z_p^2 \sigma_0^2].$$

Esta última se puede cumplir fácilmente para $0.01 \leq p \leq 0.5$ si se satisfacen las desigualdades en la expresión (4.4).

Sea

$$g(\beta_0, \beta_1, \sigma_b) = t_p(\hat{\phi}_r).$$

Defina

$$\begin{aligned}
 A &= \beta_1^2 - z_p^2 \sigma_1^2, \\
 B &= (D_f - \beta_0) \beta_1 + z_p^2 \sigma_0 \sigma_1 \rho, \\
 C &= (D_f - \beta_0)^2 - z_p^2 \sigma_0^2, \\
 D &= \{B^2 - AC\}^{1/2}.
 \end{aligned}$$

Entonces t_p se puede expresar como

$$t_p = \frac{B - D}{A}.$$

El estimador de MV de t_p es

$$\hat{t}_p = \frac{B(\hat{\phi}_r) - D(\hat{\phi}_r)}{A(\hat{\phi}_r)}.$$

El vector gradiente $\frac{\partial}{\partial \phi_r} t_p(\phi_r)$ se obtiene en la Sección (7.3) del apéndice. Los elementos

de $\frac{\partial}{\partial \phi_r} t_p$ son:

$$\begin{aligned}
 \frac{\partial}{\partial \beta_0} t_p &= -\frac{1}{AD} [(\beta_1 - B) D + A(D_f - \beta_0)], \\
 \frac{\partial}{\partial \beta_1} g(\beta_0, \beta_1, \sigma_b) &= \frac{1}{A^2} \left\{ [(D_f - \beta_0)(D - B) + \beta_1 C] \frac{A^2}{D} - 2\beta_1(B - D) \right\}, \\
 \frac{\partial}{\partial \sigma_0} g(\beta_0, \beta_1, \sigma_0, \sigma_1) &= \frac{z_p^2}{AD} (D\sigma_1\rho - B\sigma_1\rho - A\sigma_0),
 \end{aligned}$$

$$\frac{\partial}{\partial \sigma_1} g(\beta_0, \beta_1, \sigma_0, \sigma_1) = \frac{z_p^2}{A^2} \left\{ [D\sigma_0\rho - B\sigma_0\rho - \sigma_1 C] \frac{A}{D} + 2\sigma_1(B - D) \right\},$$

$$\frac{\partial}{\partial \rho} g(\beta_0, \beta_1, \sigma_0, \sigma_1, \rho) = \frac{z_p^2 \sigma_0 \sigma_1}{A} (D - B).$$

Entonces

$$\text{var} [\hat{t}_p] = \left(\frac{\partial}{\partial \phi_r} t_p(\phi_r) \right)^T \mathbf{I}^{-1}(\beta_0, \beta_1, \sigma_0, \sigma_1, \rho) \left(\frac{\partial}{\partial \phi_r} t_p(\phi_r) \right),$$

donde las derivadas son evaluadas en $\hat{\phi}_r = (\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_0, \hat{\sigma}_1, \hat{\rho})$ y $\mathbf{Z}_i^T = \begin{pmatrix} 1 & \cdots & 1 \\ t_{i1} & \cdots & t_{ini} \end{pmatrix}$, con $\mathbf{I}(\beta_0, \beta_1, \sigma_0, \sigma_1, \rho)$; como se encuentra expresada en la Subsección (5.4.3).

Los estimadores de $\text{var} [\hat{t}_p]$ que se presentaron en la Sección (4.5) para el modelo con intercepto y pendiente aleatoria quedan expresados como

$$\widehat{\text{var}}_1 [\hat{t}_p] = \left(\frac{\partial}{\partial \phi_r} t_p(\phi_r) \right)^T \mathbf{J}^{-1}(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_0, \hat{\sigma}_1, \hat{\rho}) \left(\frac{\partial}{\partial \phi_r} t_p(\phi_r) \right),$$

donde las derivadas se evalúan en $\hat{\phi}_r = (\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_0, \hat{\sigma}_1, \hat{\rho})$; y

$$\widehat{\text{var}}_2 [\hat{t}_p] = \left(\frac{\partial}{\partial \phi_r} t_p(\phi_r) \right)^T \mathbf{I}^{-1}(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_0, \hat{\sigma}_1, \hat{\rho}) \left(\frac{\partial}{\partial \phi_r} t_p(\phi_r) \right),$$

donde $\mathbf{I}(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_0, \hat{\sigma}_1, \hat{\rho})$ es como en la Subsección (5.4.3) pero evaluada en $\hat{\phi}_r = (\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_0, \hat{\sigma}_1, \hat{\rho})$. Además en este caso,

$$\hat{\mathbf{V}}_i = \mathbf{Z}_i \hat{\Psi} \mathbf{Z}_i^T + \hat{\sigma}^2 \mathbf{I}_n,$$

$$\text{con } \mathbf{Z}_i^T = \begin{pmatrix} 1 & \cdots & 1 \\ t_{i1} & \cdots & t_{in_i} \end{pmatrix} \text{ y } \widehat{\Psi} = \begin{pmatrix} \hat{\sigma}_0^2 & \hat{\sigma}_0 \hat{\sigma}_1 \hat{\rho} \\ \hat{\sigma}_0 \hat{\sigma}_1 \hat{\rho} & \hat{\sigma}_1^2 \end{pmatrix}, \text{ con } -1 < \hat{\rho} < 1.$$

Debido a la complejidad de los estimadores $\widehat{var}_1 [\hat{t}_p]$ y $\widehat{var}_2 [\hat{t}_p]$ en este caso sólo se representan en forma genérica. Pero se pueden obtener explícitamente a partir de los resultados de la Sección (7.2.3).

Capítulo 5

Ejemplo real: Modelo con intercepto y pendiente aleatorios

En este capítulo se presenta la aplicación de los criterios planteados para tiempos de paro en estudios de degradación. Inicialmente se muestra el análisis de un conjunto de datos reales, provenientes de un estudio de confiabilidad; el cual se llevó a cabo en el Centro de Tecnología y Productos de MABE en la ciudad de Querétaro. Posteriormente con estos datos se muestra la aplicación de los dos criterios de paro para un modelo de degradación lineal con intercepto y pendiente aleatorios. Bajo este escenario, se planea el tiempo de paro antes de iniciar el estudio. Esto significa que se planea el tiempo de duración del estudio con la opción de determinar el número de unidades en la prueba. También se considera el caso cuando el estudio de degradación está en proceso y con la información que se esta recopilando se decide el momento de terminar dicho estudio. Finalmente,

en la aplicación del segundo criterio se planea la prueba, pero minimizando los costos económicos de la misma.

5.1 Análisis de datos

Los datos reales utilizados para ilustrar el procedimiento propuesto en este trabajo para la terminación de una prueba de degradación en confiabilidad, fueron obtenidos en un estudio realizado para la comparación de la calidad del producto ofrecido por dos proveedores diferentes, Peña (2003).

Los datos son mediciones del cambio de color a lo largo del tiempo de una laca que se aplica a una lámina de acero. El estudio de degradación del color de la laca es de gran interés, ya que este tipo de laca se aplica a las láminas de acero que se utilizan en los frentes de algunos modelos de estufas que fabrican en MABE. Este interés se debe a la necesidad de mejorar continuamente la calidad de los productos. En el proceso de degradación de la laca, esta tiende a amarillarse. Entonces en el estudio se tiene interés en determinar el tiempo en que una proporción específica de unidades alcanza un nivel de amarillamiento, tal que se declara como un defecto de apariencia. Es decir, se desea determinar un cuantil t_p de la distribución de los tiempos a la falla por amarillamiento.

Los datos corresponden a cambio de color observados a través del tiempo. El estudio se realizó a una temperatura $167^{\circ}C$, con 15 unidades y se tomaron 10 mediciones en cada unidad.

En la Figura 5-1 se puede apreciar que las trayectorias de degradación el desgaste del

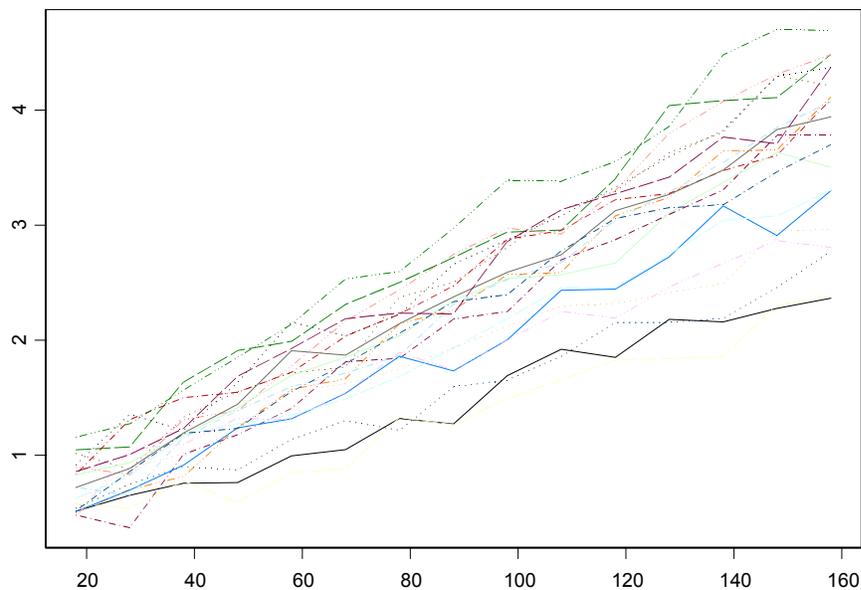


Figura 5-1: Trayectorias del desgaste por amarillamiento.

color en las lacas son de tipo lineal y que un modelo de degradación con una covariable podría ajustarse adecuadamente a los datos. En esta Figura también se observa que los interceptos y las pendientes parecen ser aleatorios, ya que son diferentes para cada unidad.

Como un estudio preliminar, se realizó un análisis de regresión lineal simple para cada unidad. Esto con el fin de identificar efectos aleatorios en modelo; así identificar la estructura de covarianza entre tales efectos.

En la Figura 5-2 se aprecian diferencias entre los interceptos; así como entre las pendientes. Esto parece indicar que el intercepto al igual que la pendiente son efectos

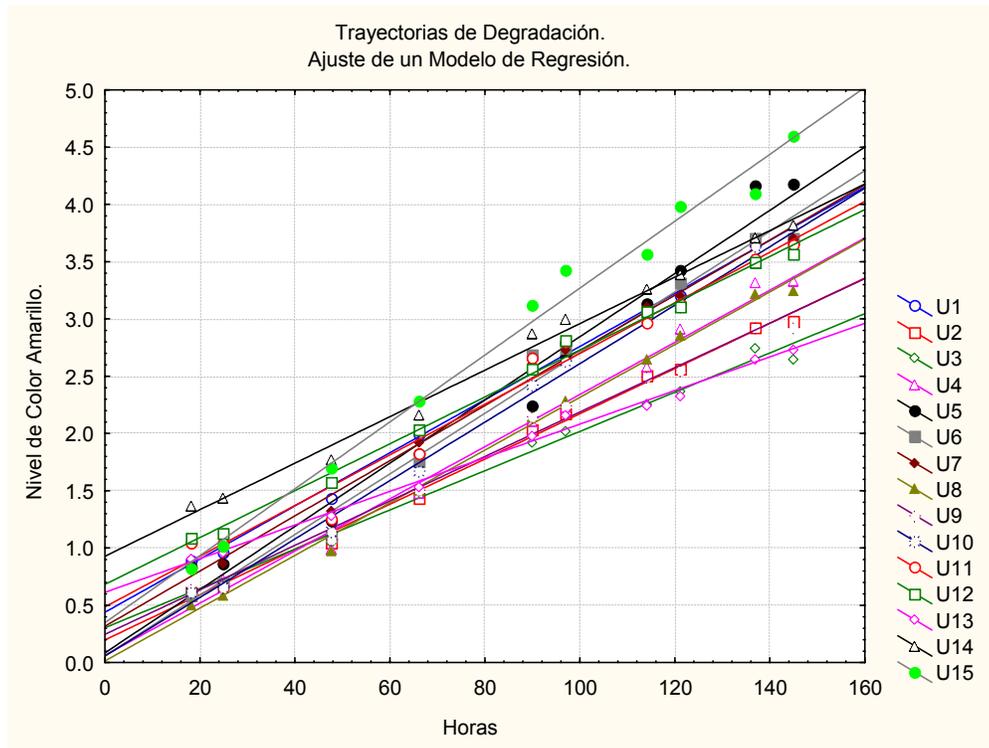


Figura 5-2: Modelo de regresión ajustado a cada unidad.

aleatorios. Al graficar los intervalos de confianza del intercepto y pendiente de cada uno de los modelos lineales, se observa que la mayoría de ellos no se intersecta; ver Figura 5-3. Este comportamiento corrobora que los dos efectos aleatorios, ya mencionados, son necesarios para representar la variabilidad entre las unidades.

De este modo, se propone el siguiente modelo lineal de efectos aleatorios,

$$Y_{ij} = D(t_{ij}) + \sigma\varepsilon_{ij} = (\beta_0 + \sigma_b b_i^0) + (\beta_1 + \sigma_b b_i^1) t_{ij} + \sigma\varepsilon_{ij},$$

donde $b_i^0 \sim N(0, 1)$, $b_i^1 \sim N(0, 1)$ y $\varepsilon_{ij} \sim N(0, 1)$. Esta forma de representar los modelos

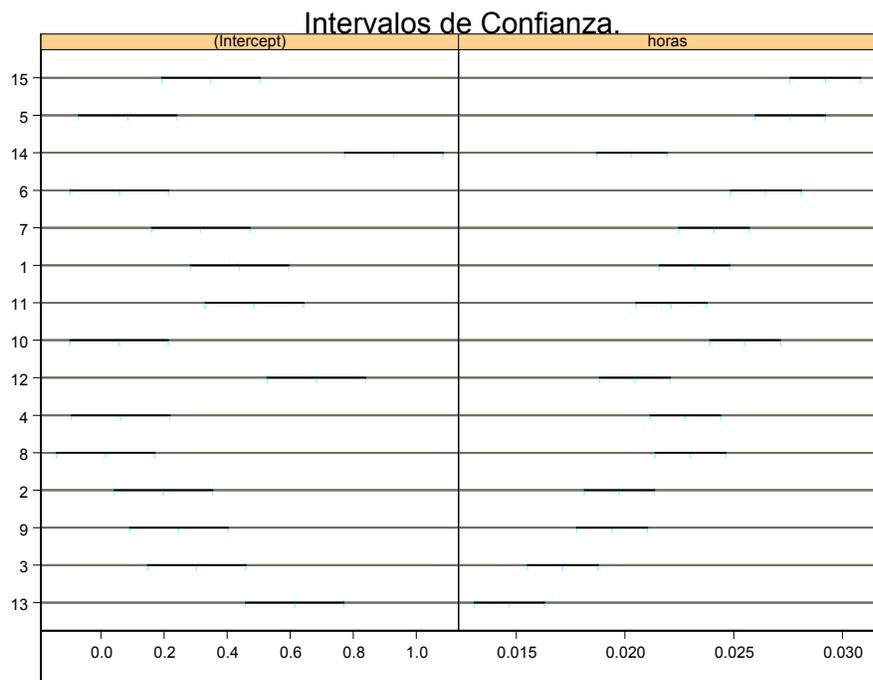


Figura 5-3: Intervalos de Confianza de los parámetros de la regresión en cada unidad.

de efectos mixtos sigue el enfoque de Laird y Ware (1982).

Es importante mencionar que también se ajustó un modelo lineal con intercepto aleatorio y pendiente fija, con los mismos supuestos distribucionales de normalidad. Al comparar los dos modelos mediante un análisis de varianza se observó un p -valor muy pequeño para la prueba de razón de verosimilitud; la cual es una evidencia de que el modelo con pendiente e intercepto aleatorios proporciona una mejor descripción de los datos que el modelo con intercepto aleatorio.

Modelo ajustado

Para estimar los parámetros del modelo se utilizó la función LME de S-Plus, con el método de máxima verosimilitud restringida. Las estimaciones obtenidas son las siguientes

$$\hat{\beta}_0 = 0.323, \quad \hat{\beta}_1 = 0.0223, \quad \hat{\sigma}_0 = 0.2557, \quad \hat{\sigma}_1 = 0.00383, \quad \rho = -0.434 \quad \text{y} \quad \hat{\sigma} = 0.111.$$

De esta tabla se observa que la variabilidad de efecto aleatorio del intercepto es grande, en relación a la varianza del efecto de la pendiente. Existe también una correlación negativa entre los efectos aleatorios.

	Estimación	Error Estándar	g. l.	t -valor	p -valor
β_0	.32308	.06719	134	4.8082	.0001
β_1	.02238	.00098	134	22.7305	.0001

En la Tabla anterior sobre la significancia de los coeficientes del modelo, se tienen p -valores pequeños; lo cual permite concluir que los coeficientes fijos en el modelo son

altamente significativos. En la Tabla de intervalos de confianza para los parámetros del modelo la mayoría son estrechos, con excepción del intervalo para la correlación entre los

	Límite inferior	Estimación	Límite superior
β_0	0.191	0.323	0.4551
β	0.0204	0.0223	0.0243
σ_0	0.1408	0.2462	0.4303
σ_1	0.0021	0.0036	0.0062
ρ	-0.840	-0.431	0.2887
σ	.0929	0.111	0.133

efectos aleatorios.

En Figura ?? se muestra también que el modelo estimado se ajusta adecuadamente a las trayectorias de degradación.

Validación del modelo

Para validar el modelo, en seguida se realiza un diagnóstico sobre los supuestos distribucionales del modelo. En los modelos que se han trabajado en la tesis se consideran dos suposiciones distribucionales: una dentro de las unidades y la otra entre unidades.

- Análisis de residuales. En la Figura 5-5 se observa que los residuales se distribuyen simétricamente alrededor del cero y que aproximadamente se satisface el supuestos de homocedasticidad entre los errores. Además se tiene que los residuales no siguen un comportamiento definido, lo cual muestra que existe independencia entre los errores. La gráfica de papel normal en la Figura 5-6, muestra que los errores aproximadamente se distribuyen como una Normal.

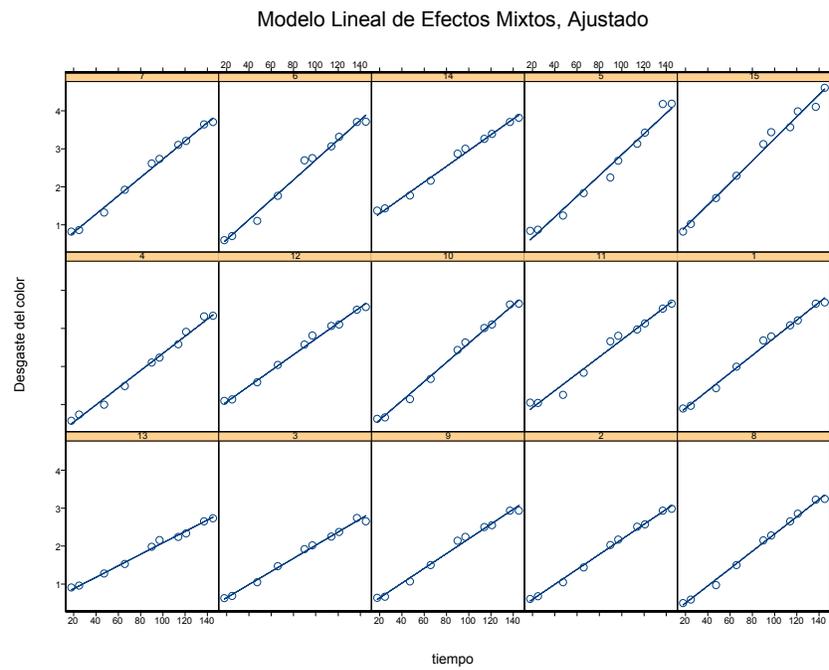


Figura 5-4: Modelo de efectos mixtos ajustado a los datos sobre el amarillamiento.

Valores ajustados contra Residuales.

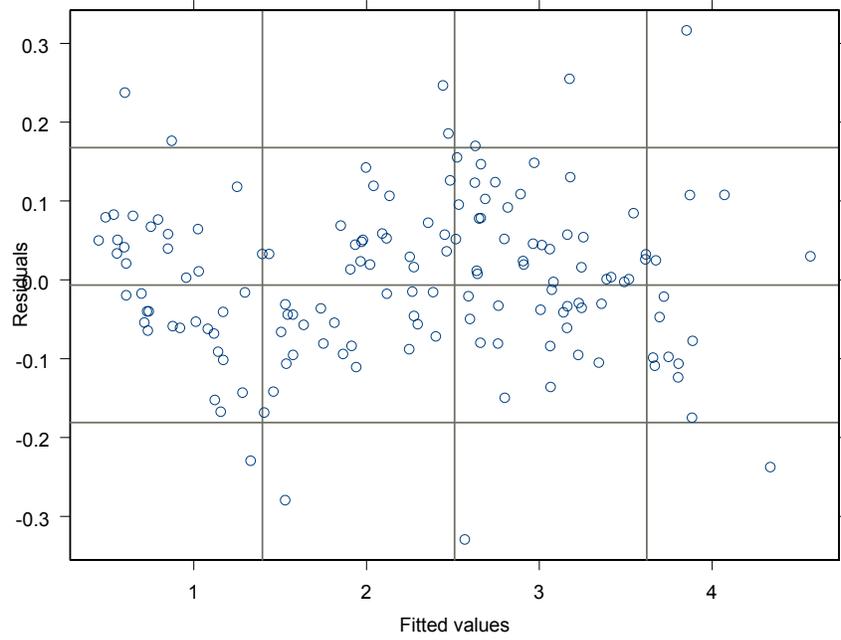


Figura 5-5: Gráfica de valores ajustados vs residuales.

Grafica en papel normal para los residuales

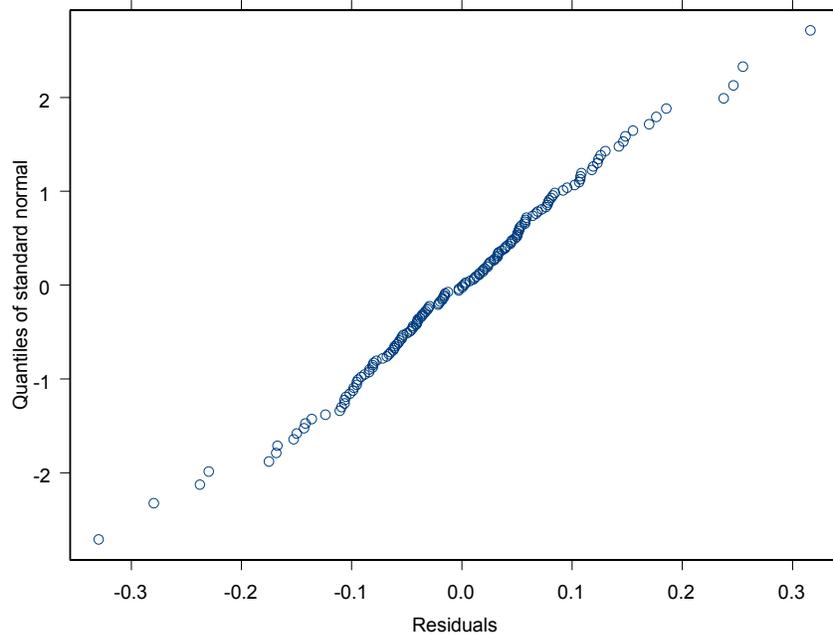


Figura 5-6: Gráfica de papel normal para los residuales del modelo ajustado.

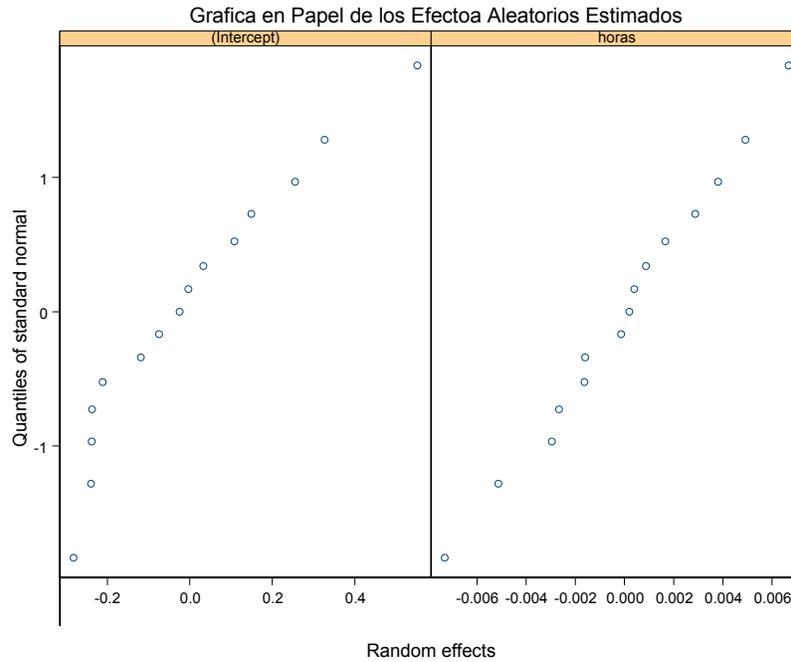


Figura 5-7: Gráficas de papel normal de los efectos aleatorios.

- Validación de las suposiciones sobre los efectos aleatorios. La suposición de normalidad (marginal) de los efectos aleatorios se cumple de manera aproximada; a pesar de que en el intercepto se observa cierta asimetría y algunos outliers; ver Figura 5-7.

En la Figura 5-8 se observa que existe mayor variabilidad en el efecto del intercepto que en la pendiente. Además se tiene una ligera correlación negativa entre los efectos aleatorios del modelo.

En resumen se puede concluir que el modelo con intercepto y pendiente aleatorios explica adecuadamente los datos. Además, en la validación del modelo se tiene que los

Grafica de Dispersion de los Efectos Aleatorios.

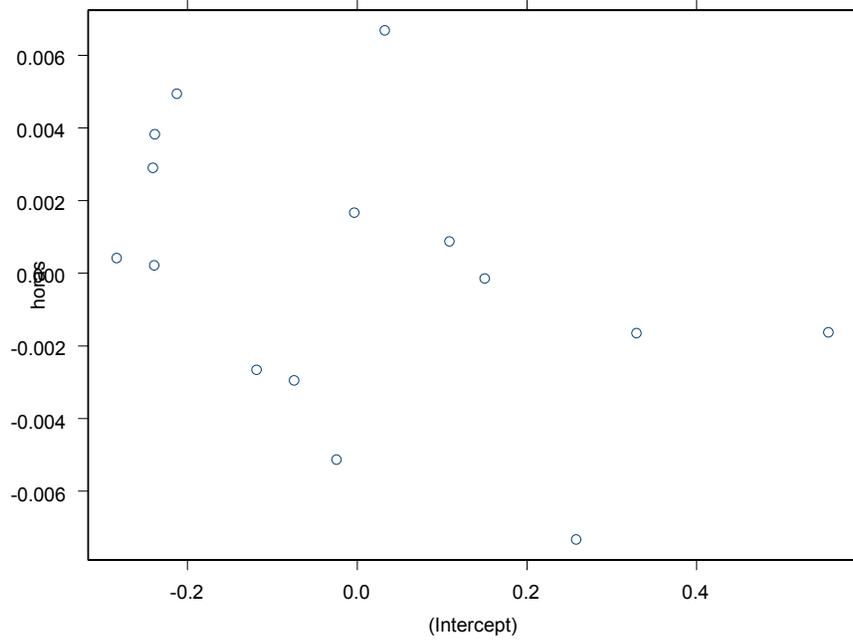


Figura 5-8: Diagrama de dispersión de efectos estimados del intercepto vs tiempo.

supuestos de los errores y efectos aleatorios se cumplen de manera aproximada.

5.2 Planeación de una prueba

Se desea planear una prueba sobre el desgaste del color que sufren las lacas. Este estudio tiene como objetivo determinar, con una precisión alta, el tiempo en el cual el 1% y el 5% de las unidades alcanzan el nivel crítico de degradación $D_f = 5$. Es decir, se desea estimar los cuantiles $t_{.01}$ y $t_{.05}$ de la distribución de vida con un alto nivel de precisión.

Para planear la prueba se cuenta con la información sobre el tipo de modelo de degradación y con las estimaciones de los parámetros del modelo que se encuentra en la sección anterior.

Antes de planear la prueba vamos a considerar tres escenarios:

- Se determina el tiempo de paro antes de iniciar el estudio. Esto significa que se planea el tiempo de duración del estudio con la opción de determinar el número de unidades en la prueba.
- También se considera el caso cuando el estudio de degradación está en proceso y con la información que se esta recopilando se decide el momento de terminar dicho estudio.
- Se planea la prueba, pero minimizando los costos económicos de la misma.

En el primero y tercer caso, en la planeación de la prueba nos apoyaremos en la simulación.

Como se mencionó en el Capítulo 1, en la planeación de la prueba es importante determinar el tiempo de duración del experimento, el número de unidades que se van a someter a prueba y la frecuencia de inspección. En la práctica es común que la frecuencia de observación es fija y no forma parte de las variables que se utilizan en la planeación del estudio de degradación. Por esta razón en los criterios de planeación que aquí se proponen, la frecuencia es fija.

Para una frecuencia de observación fija de 3 horas se obtuvieron estimaciones de MV para diferentes tiempos de terminación de experimento, t_c , y para los tamaños de muestra $M = (10, 20, 30, \dots, 150)$. La primera medición se consideró a las 15 horas, el primer tiempo de paro fue a las 45 horas y el último a las 264 horas. Cada unidad en prueba tiene el mismo número de mediciones y en los mismos tiempos de observación.

Para obtener las estimaciones se construyó un programa con las funciones de S-Plus, para lo cual se simularon diferentes escenarios.

5.2.1 Estimaciones de los parámetros y su precisión

En las Figuras 5-9 y 5-11 se presentan las estimaciones de β_0 , β_1 , σ_0 , σ_1 , σ y ρ . En estas figuras se observa que para un tamaño de muestra fijo, las gráficas que representan las estimaciones del parámetro se comportan en forma errática para tiempos cortos en la duración del experimento y tienden a estabilizarse alrededor de un valor cuando t_c se incrementa. Así mismo, al incrementar el tamaño de muestra las gráficas de estimaciones tienden a estabilizarse más rápidamente y el valor alrededor del cual se estabilizan esta

más próximo del verdadero valor del parámetro. Esta tendencia es similar en cualquiera de la serie de estimaciones de los diferentes parámetros del modelo. En la gráfica sobre las estimaciones de β_0 y σ_0 se observa que requiere de un mayor t_c para estabilizarse. Esto se debe a que la varianza del efecto aleatorio del intercepto es mayor que los otros componentes de varianza. Las gráficas con las estimaciones de σ para diferentes tamaños de muestra tienden a estabilizarse rápidamente al valor poblacional del parámetro; ver la Figura 5-11.

Este comportamiento quiere decir que después de un cierto tiempo de duración del experimento, la información que proporcionan las trayectorias ya no mejora las estimaciones de los parámetros. Es decir, que después de un tiempo de observación la precisión de las estimaciones ya no mejora. Un comportamiento similar se observa en las estimaciones de todos los parámetros.

Nótese también que cada una de las series de estimaciones de un parámetro específico se aproximan a diferentes valores, para tamaños de muestra diferentes. Esto se debe a que el modelo con los dos efectos aleatorios causa una alta variabilidad. Desde el punto de vista teórico, para tamaños de muestra grandes (en este caso mayores de 300 unidades) la serie de estimaciones se aproxima mejor al verdadero valor del parámetro. Sin embargo, en este caso los tamaños de muestra considerados son pequeños para ver algún comportamiento asintótico al verdadero valor del parámetro. En el estudio no se consideran tamaños de muestra mayores, ya que en este trabajo se pretende simular escenarios reales.

Estimaciones de los parámetros

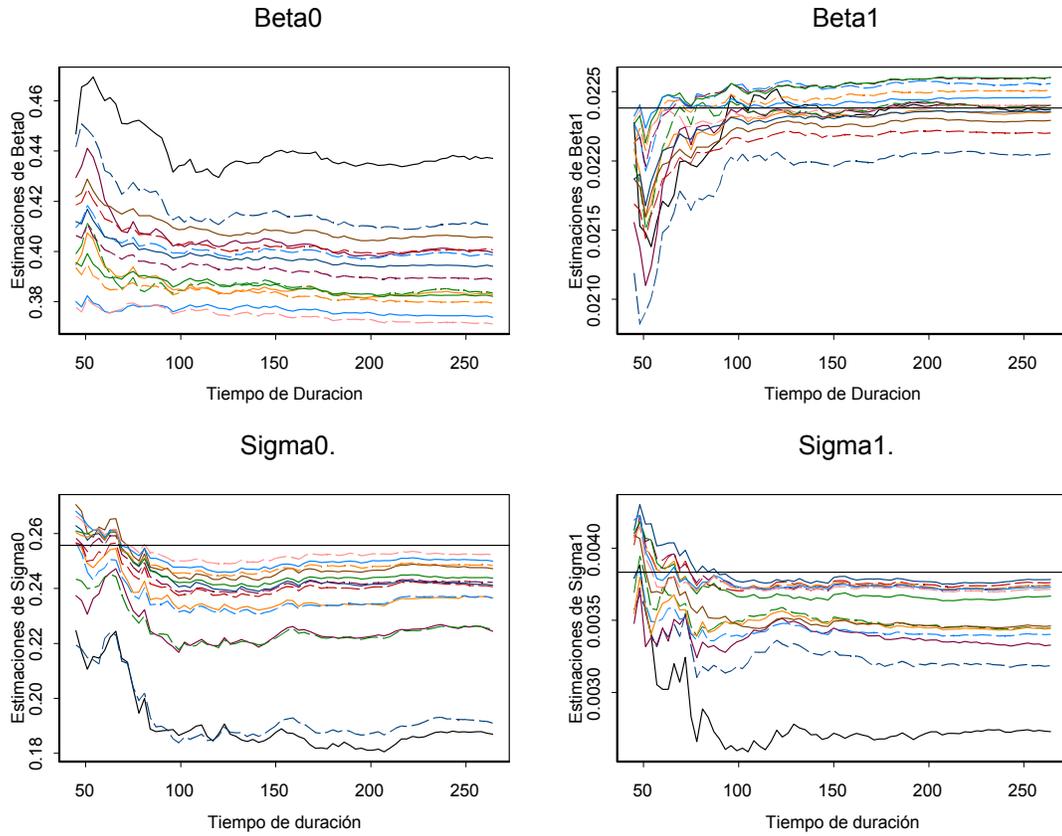


Figura 5-9: Estimaciones de β_0 , β_1 , σ_0 y σ_1 , para diferentes tiempos de paro.

Por las propiedades asintóticas de los estimadores de máxima verosimilitud, se consideró el inverso de la matriz de información como una medida aproximada de la precisión de las estimaciones. Las gráficas de las precisiones correspondientes a las estimaciones de los parámetros se presentan en las Figuras 5-9 y 5-11. En estas se puede apreciar que dado un tamaño de muestra fijo, las curvas de varianzas estimadas se comportan erráticamente para valores de t_c cortos, pero al incrementar la duración del experimento tienden a decrecer y a estabilizarse. Esto significa que después de un cierto tiempo de observación del experimento, la ganancia en la precisión de las estimaciones es muy baja o casi cero, a pesar de tener mayor información al seguir corriendo la prueba de degradación. Note que al incrementar el número de unidades en muestra, las estimaciones de las varianzas son menores y están más cercanas al cero. Esto es que al incrementar la muestra, se tienen estimaciones con una precisión mayor. Las estimaciones de β_0 y σ_0 son menos precisas; esto se debe a que la varianza del efecto aleatorio del intercepto con el cual se generaron las trayectorias de degradación es alta. Por otro lado, las estimaciones de σ tienen una alta precisión.

Por lo anterior, resulta conveniente detener el experimento de degradación cuando las estimaciones y las varianzas de sus estimadores tiendan a estabilizarse; ya que la ganancia en precisión es baja, pero la pérdida económica es alta.

Como se puede ver en las Figuras 5-9, 5-10 y 5-11 para un mismo tamaño de muestra las series de estimaciones de los parámetros y las varianzas de sus estimadores se estabilizan en diferentes tiempos de observación, t_c . Resulta difícil determinar un tiempo

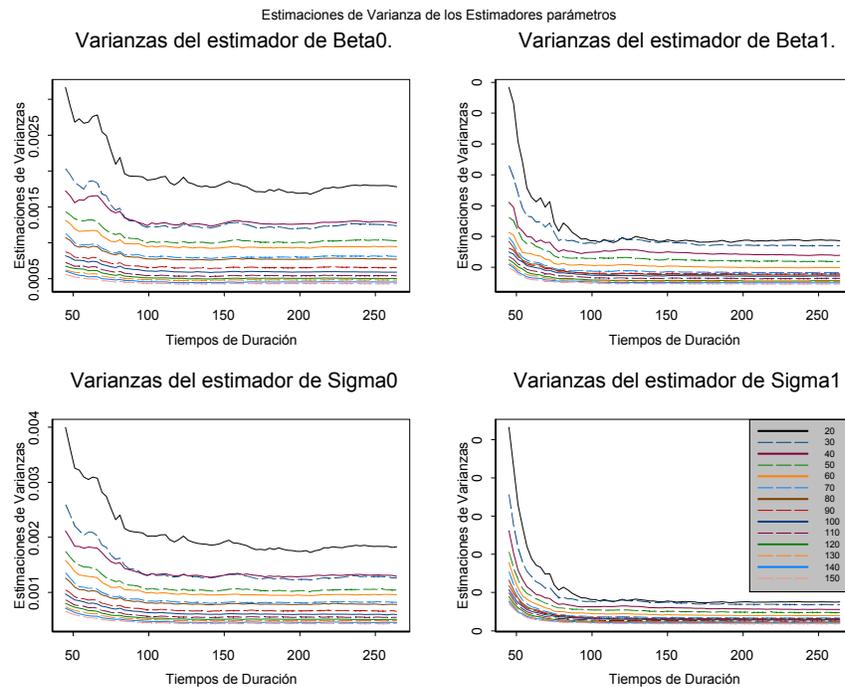


Figura 5-10: Estimaciones de varianzas de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\sigma}_0$ y $\hat{\sigma}_1$, para diferentes tiempos de paro.

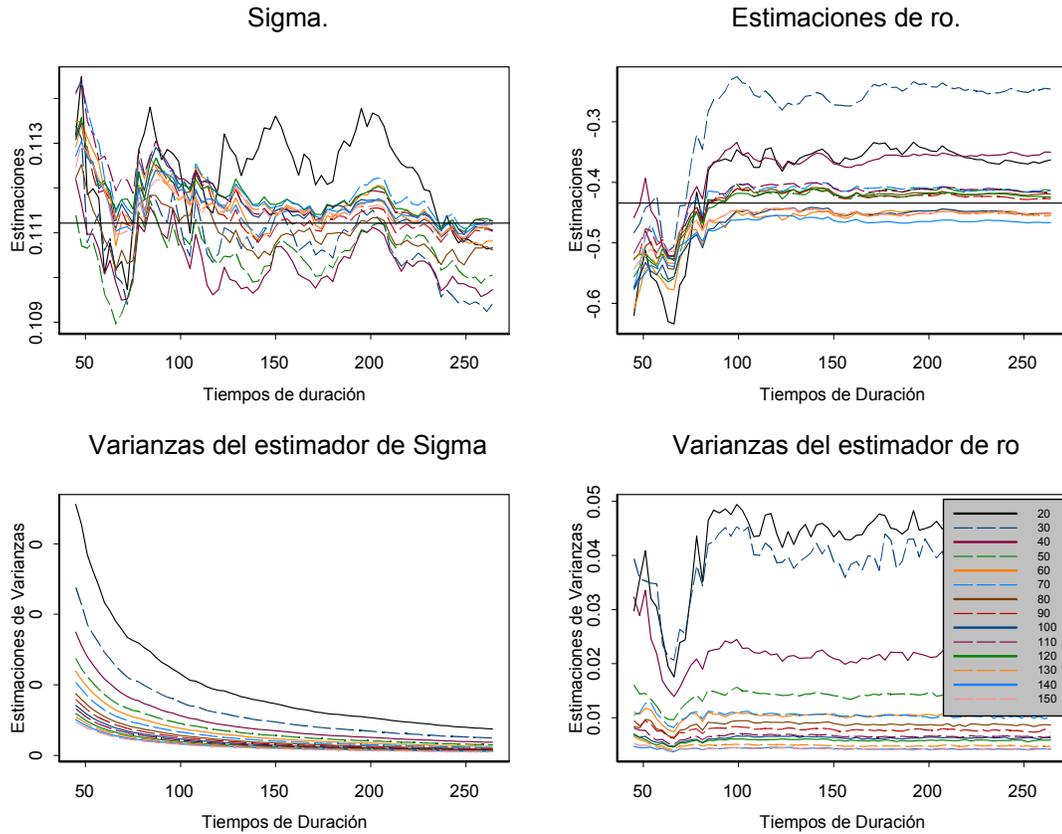


Figura 5-11: Estimaciones de σ^2 , ρ , $\text{var}(\hat{\sigma}^2)$ y $\text{var}(\hat{\rho})$, para diferentes tiempos de paro.

de paro adecuado, tomando en cuenta de manera simultánea las tendencias anteriores; excepto cuando se tiene interés en un parámetro específico.

El criterio para detener la prueba de degradación se basa en los objetivos del experimento.

Como se mencionó al inicio de esta Sección, el objetivo es estimar los cuantiles $t_{.01}$ y $t_{.05}$ del tiempo de vida; entonces para determinar el tiempo de paro, se consideran los criterios propuestos en la Secciones (3.1) y (3.2).

5.2.2 Primer criterio para tiempos de paro

Para la misma frecuencia, tamaños de muestra y tiempos de terminación que se mencionaron en la Subsección anterior, se estimaron los cuantiles $t_{.01}$ y $t_{.05}$, y las varianzas de sus estimadores, para diferentes tiempos de duración, t_c . En la Sección (4.5.3) se obtuvo la expresión analítica del estimador de MV del cuantil t_p y por el método delta se muestran dos expresiones que permiten obtener estimaciones de la varianza de \hat{t}_p . Las cuales como se recordará se denotan como $var_1 [\hat{t}_p]$ y $var_2 [\hat{t}_p]$. Además también se obtiene $var [\hat{t}_p]$. Antes de obtener las estimaciones se verificó que se cumplieran las restricciones 4.4 para poder utilizar \hat{t}_p .

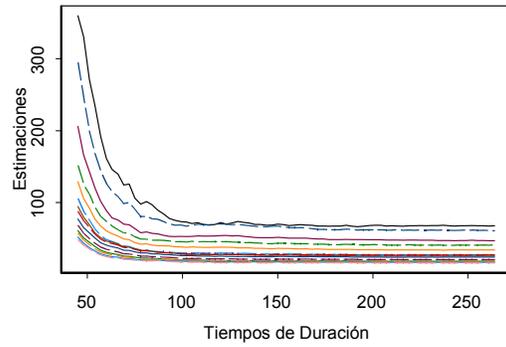
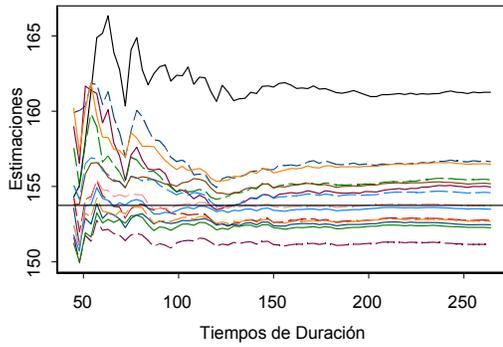
En la Figura 5-12 se presentan las gráficas de las estimaciones de $t_{.01}$, $var_1 [\hat{t}_p]$ y $var_2 [\hat{t}_p]$ para los tiempos de terminación t_c y los tamaños de muestra que se presentan en la Sección (5.2). En la gráfica de estimaciones del cuantil $t_{.01}$ se observa claramente que, para un tamaño de muestra fijo, la serie de estimaciones del cuantil 0.01 tiene una tendencia errática para t_c pequeños, pero tiende a estabilizarse cuando el tiempo de duración del experimento crece. Además, cuando el tamaño de muestra crece, las gráficas de estimaciones son más suaves, se estabilizan más rápido y además las estimaciones están más cerca del valor poblacional del cuantil. Note que este comportamiento es similar al que muestran las estimaciones de los parámetros del modelo.

Con respecto a las estimaciones de la varianza del cuantil, también se aprecia que bajo un tamaño de muestra M fijo, las curvas de las varianzas decrecen con el incremento en la terminación de la prueba, pero después de un cierto tiempo tienden a estabilizarse.

Estimaciones del Cuantil 0.01 y de las Varianzas de su Estimador

Estimaciones del Cuantil 0.01

Varianzas con Información Observada



Varianzas con Información Esperada Evaluada

Varianzas con Información Espera Teórica

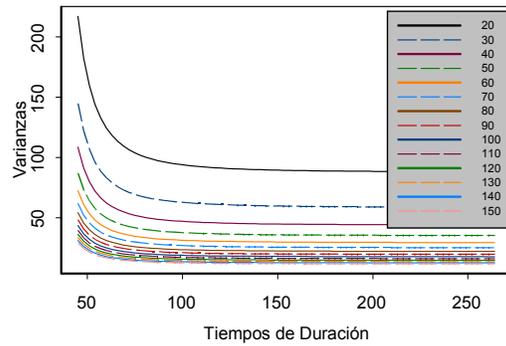
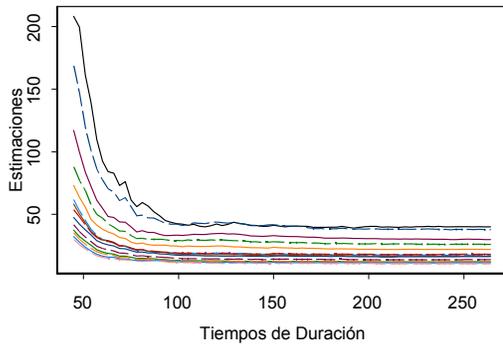


Figura 5-12: Estimaciones de $t_{.01}$, $\text{var}_1(\hat{t}_{.01})$ y $\text{var}_2(\hat{t}_{.01})$ y los valores simulados de $\text{var}(\hat{t}_{.01})$.

Esta misma tendencia se observa en las estimaciones de $var_1 [\hat{t}_p]$ y $var_2 [\hat{t}_p]$. Es claro, que cuando se incrementa el número de unidades en la prueba, las estimaciones de las varianzas de $\hat{t}_{0.01}$ son menores y empiezan a estabilizarse para tiempos de terminación reducidos. Esto quiere decir que para un número de unidades pequeño en la prueba, el tiempo de terminación debe ser mayor hasta observar que se estabilice la precisión que cuando se tiene un mayor número de unidades. pero después de este tiempo la ganancia en precisión es mínima.

En resumen cuando se tienen tamaños de muestra pequeños las estimaciones del cuantil son más erráticas y están más alejadas del valor poblacional; las precisiones de las estimaciones son más pobres y el tiempo de terminación de la prueba donde se estabiliza es mayor que cuando el número de unidades en prueba es mayor. Sin embargo, en ambos casos la ganancia en la precisión de las estimaciones ya no se incrementa, a pesar de que la prueba continua.

Al aplicar el primer criterio propuesto en la tesis con $M = 20$ y $M = 60$ se observa en la gráfica de las estimaciones de $var_1 [\hat{t}_p]$ que en 20 unidades la gráfica se estabiliza alrededor de las 100 horas y que con 60 unidades se estabiliza cerca de las 75 horas de duración del experimento. En la gráfica de las estimaciones de $var_2 \hat{t}_p$ es claro que las curvas de precisión para estos tamaños de muestra se estabilizan alrededor de los mismos tiempos.

Con el primer criterio propuesto, se tienen los siguientes planes de prueba de degradación para el estudio del desgaste del color en las lacas.

1. Para una frecuencia de observación de 3 horas y un tamaño de muestra de $M = 20$, el tiempo adecuado de terminación la prueba es de $t_c = 100$ horas. Con la desventaja de que las precisiones de las estimaciones del cuantil $t_{0.01}$ pueden ser pequeñas.
2. Para una frecuencia de observación de 3 horas y un tamaño de muestra de $M = 60$, el tiempo adecuado de terminación la prueba es de $t_c = 75$ horas. Donde las precisiones de las estimaciones que se obtengan del cuantil $t_{0.01}$ son mejores que las del plan de prueba anterior.

Resultados similares se obtienen al planear la prueba para estimar el cuantil $t_{0.05}$. Aunque para este cuantil se obtienen mejores estimaciones que para el cuantil $t_{0.01}$, esto es porque el primer cuantil esta más cerca de la media de tiempo a la falla.

5.2.3 Segundo criterio para tiempos de paro

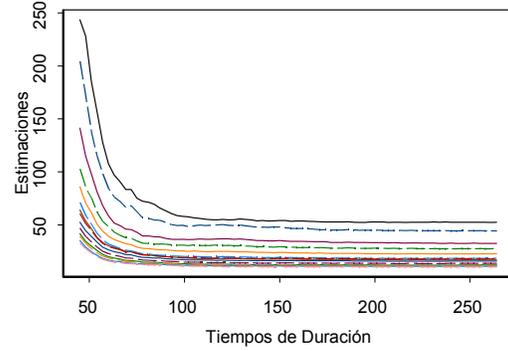
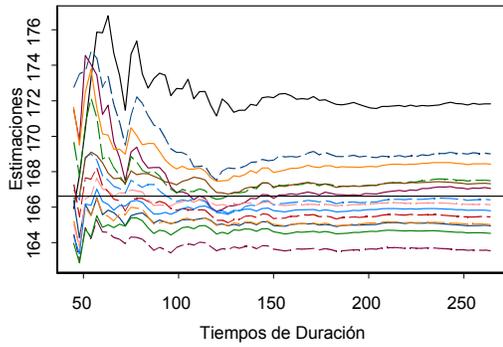
Como se mencionó en la Sección (3.2) el segundo criterio que se propone combina los costos por precisión en la inferencia, costo de observación y el costo de la unidades en la prueba. Este criterio consiste en determinar el tiempo de paro que proporciona el menor costo total de la prueba de degradación. Este criterio utiliza las estimaciones de la varianza del cuantil que se obtienen en el programa para el primer criterio. En este caso la varianza del cuantil $t_{0.01}$.

Para la planeación del estudio de degradación del desgaste de las lacas, se simularon

Estimaciones del Cuantil 0.01 y de las Varianzas de su Estimador

Estimaciones del Cuantil 0.05

Varianzas con Información Observada



Varianzas con Información Esperada Evaluada

Varianzas con Información Espera Teórica

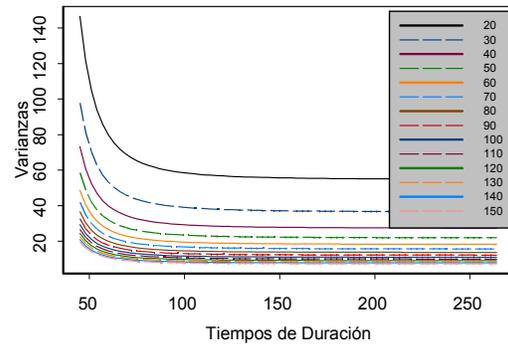
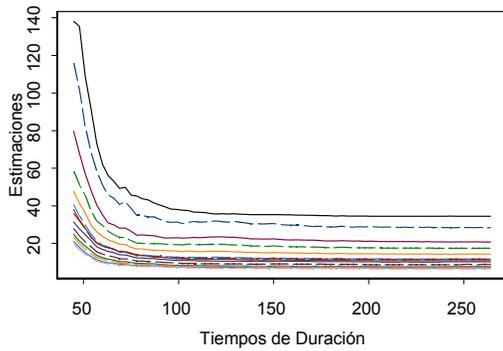


Figura 5-13: Estimaciones de $t_{.05}$, $\text{var}_1(\hat{t}_{.05})$ y $\text{var}_2(\hat{t}_{.05})$ y los valores simulados de $\text{var}(\hat{t}_{.05})$.

diferentes escenarios con tamaños de muestra de $M = (10, 20, 30, \dots, 150)$. La función de costo que se considera en este estudio es

$$CT(t) = 2500 \times var(\hat{t}_{0.01})_c + 20 \times M \times t_c + 30 \times M,$$

donde $2500 \times var(\hat{t}_{0.01})_c$ corresponde al costo por la precisión de la estimación del cuantil $t_{0.01}$, $20 \times M \times t_c$ es el costo de observación y $30 \times M$ el costo de las unidades en la prueba, al tiempo de paro t_c .

Los tiempos de terminación son los mismos que se consideran en el primer criterio.

En la Figura 5-14 se aprecia que las curvas de costo son convexas y por lo tanto tienen un valor mínimo. Para una prueba con un número fijo de unidades la curva de costo tiende a decrecer para tiempos de paro cortos. Esto es porque cuando el tiempo de terminación empieza a crecer la precisión de la estimación del cuantil se mejora y en consecuencia el costo por inferencia disminuye, pero el costo de observación no continua siendo bajo. En cambio, cuando el tiempo de observación aún mayor, la ganancia en la precisión en las estimaciones no es significativa, pero el costo de observación crece linealmente. Con este criterio, el tiempo de paro adecuado es en donde la función de costo tiene su mínimo.

A partir del segundo. Para el estudio sobre el desgaste del color, con este criterio se proponen los siguientes planes de pruebas de degradación.

1. Para una muestra de $M = 20$ unidades, el costo mínimo de la prueba de degradación

Funciones de Costo para Diferente Numero de Unidades

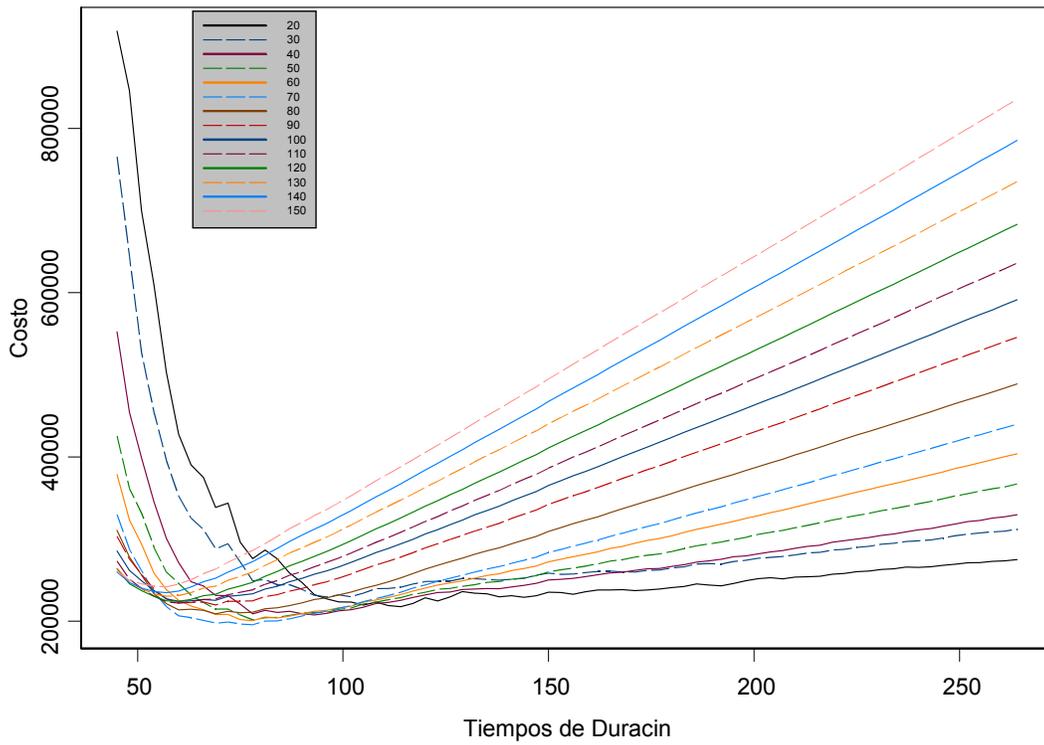


Figura 5-14: Funciones de costo total para diferentes tiempos de paro y tamaños de muestra.

se obtiene en un tiempo aproximado de terminación de 100 horas. Entonces el tiempo de paro adecuado sería de 100 horas y se estimaría $t_{0.01}$ con una aproximada de $\widehat{var} [\hat{t}_{0.01}] \approx 90$.

2. Para una muestra de $M = 60$ unidades, el costo mínimo de la prueba de degradación se obtiene alrededor de un tiempo de terminación de 80 horas. Entonces el tiempo de paro adecuado bajo este criterio sería de 100 horas y se estimaría $t_{0.01}$ con una aproximada de $\widehat{var} [\hat{t}_{0.01}] \approx 35$.

Nótese que los tiempos de paro que se determinan con ambos criterios, son similares.

En conclusión, los dos criterios propuestos proporcionan herramientas adecuadas para diseñar planes de pruebas de degradación en el estudio del desgaste del color en las lacas. Los criterios son adecuados para diferentes escenarios que se pueden presentar en la práctica.

Capítulo 6

Conclusiones

Los dos criterios propuestos en la tesis son útiles para determinar tiempos de paro adecuados en pruebas de degradación, bajo dos enfoques diferentes. El primer criterio consiste en detener el experimento cuando las estimaciones de la varianza de un cuantil alcanzan un valor específico o cuando se estabilizan. El segundo criterio combina la precisión y el costo económico de las estimaciones y consiste en parar la prueba, cuando el costo total del experimento, como función del tiempo sea mínimo.

Estos dos criterios proporcionan herramientas adecuadas para diseñar planes de pruebas de degradación en alguno de los siguientes escenarios.

1. Cuando la frecuencia de medición y el número de unidades se fijan anticipadamente.

En tal caso se determina el tiempo de duración del experimento.

2. Cuando la frecuencia de observación es fija, entonces en el diseño de la prueba de degradación se pueden determinar el número de unidades y el tiempo de termi-

nación.

Estos criterios también son útiles, cuando el estudio de degradación está en proceso y con la información que se esta recopilando se decide el momento de terminar la prueba.

En el presente trabajo se analizaron trayectorias lineales de degradación para datos balanceados y bajo los supuestos distribucionales de normalidad. Extensiones de este trabajo se pueden llevar a cabo al utilizar datos no balanceados, supuestos distribucionales diferentes al de normalidad y modelos de degradación no lineales.

Capítulo 7

Apéndice

En esta parte se presenta el desarrollo de algunos de los resultados que se utilizan en el trabajo de tesis.

7.1 Función Score e información observada

La función de log verosimilitud para el modelo lineal de efectos mixtos, bajo el supuesto de normalidad es

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}; \text{datos}) = -\frac{1}{2} \left(\sum_{i=1}^M \log |\mathbf{V}_i| + \sum_{i=1}^M (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right),$$

donde $\mathbf{V}_i(\boldsymbol{\theta}) = \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \sigma^2 \mathbf{I}_{n_i}$ y $\boldsymbol{\theta}$ contiene los elementos diferentes de la matriz de covarianza $\boldsymbol{\Psi}$ y la varianza del error, σ .

Para hacer los cálculos sencillos al obtener la función Score y la matriz Hessiana se

derivan por separado los dos términos de la función $l(\boldsymbol{\beta}, \boldsymbol{\theta}; \text{datos})$.

Las primeras derivadas

Las primeras derivadas del primer término de l son

$$\frac{\partial}{\partial \boldsymbol{\beta}} \log |\mathbf{V}_i| = 0,$$

$$\frac{\partial}{\partial \theta_k} \log |\mathbf{V}_i| = \text{tr} \left[\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \right].$$

Desarrollando el segundo término se tiene que

$$(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) = \mathbf{y}_i^T \mathbf{V}_i^{-1} \mathbf{y}_i - \mathbf{y}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{y}_i + \boldsymbol{\beta}^T \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \boldsymbol{\beta}.$$

Para este término las primeras derivadas son

$$\frac{\partial}{\partial \boldsymbol{\beta}} \left[(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] \tag{7.1}$$

$$= -[\mathbf{y}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i]^T - [\mathbf{y}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i]^T + 2\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \boldsymbol{\beta} \tag{7.2}$$

$$= -2(\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{y}_i - \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \boldsymbol{\beta})$$

$$= -2\mathbf{X}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \tag{7.3}$$

y

$$\frac{\partial}{\partial \theta_k} \left[(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] \quad (7.4)$$

$$= - (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}), \quad (7.5)$$

Las segundas derivadas

Las segundas derivadas del primer término están dadas como,

$$\frac{\partial}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} \log |\mathbf{V}_i| = 0,$$

$$\frac{\partial}{\partial \theta_h \partial \theta_k} \log |\mathbf{V}_i| = \text{tr} \left[\mathbf{V}_i^{-1} \frac{\partial^2 \mathbf{V}_i}{\partial \theta_h \partial \theta_k} - \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_h} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \right],$$

$$\frac{\partial}{\partial \boldsymbol{\beta} \partial \theta_k} \log |\mathbf{V}_i| = 0.$$

Las segundas derivadas del segundo término son

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} \left[(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] &= -2 \frac{\partial}{\partial \boldsymbol{\beta}^T} (\mathbf{X}_i^T \mathbf{V}_i^{-1} [\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}]) \\ &= 2 \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i, \end{aligned} \quad (7.6)$$

$$\begin{aligned}
& \frac{\partial^2}{\partial\theta_h\partial\theta_k} \left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \right] \\
&= -\frac{\partial}{\partial\theta_h} \left((\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} \frac{\partial\mathbf{V}_i}{\partial\theta_k} \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \right) \\
&= -(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \left[\frac{\partial\mathbf{V}_i^{-1}}{\partial\theta_h} \frac{\partial\mathbf{V}_i}{\partial\theta_k} \mathbf{V}_i^{-1} + \mathbf{V}_i^{-1} \frac{\partial}{\partial\theta_h} \left(\frac{\partial\mathbf{V}_i}{\partial\theta_k} \mathbf{V}_i^{-1} \right) \right] (\mathbf{y}_i - \boldsymbol{\beta}\mathbf{X}_i) \\
&= -(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \times \\
&\quad \left[-\mathbf{V}_i^{-1} \frac{\partial\mathbf{V}_i}{\partial\theta_h} \mathbf{V}_i^{-1} \frac{\partial\mathbf{V}_i}{\partial\theta_k} \mathbf{V}_i^{-1} + \mathbf{V}_i^{-1} \left(\frac{\partial^2\mathbf{V}_i}{\partial\theta_h\partial\theta_k} \mathbf{V}_i^{-1} - \frac{\partial\mathbf{V}_i}{\partial\theta_k} \mathbf{V}_i^{-1} \frac{\partial\mathbf{V}_i}{\partial\theta_h} \mathbf{V}_i^{-1} \right) \right] (\mathbf{y}_i - \boldsymbol{\beta}\mathbf{X}_i) \\
&= -(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} \left[-\frac{\partial\mathbf{V}_i}{\partial\theta_h} \mathbf{V}_i^{-1} \frac{\partial\mathbf{V}_i}{\partial\theta_k} + \frac{\partial^2\mathbf{V}_i}{\partial\theta_h\partial\theta_k} - \frac{\partial\mathbf{V}_i}{\partial\theta_k} \mathbf{V}_i^{-1} \frac{\partial\mathbf{V}_i}{\partial\theta_h} \right] \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\beta}\mathbf{X}_i),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial\theta_k\partial\boldsymbol{\beta}} \left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \right] &= -2\frac{\partial}{\partial\theta_k} \left[\mathbf{X}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\beta}\mathbf{X}_i) \right] \\
&= 2\mathbf{X}_i^T \mathbf{V}_i^{-1} \frac{\partial\mathbf{V}_i}{\partial\theta_k} \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\beta}\mathbf{X}_i).
\end{aligned}$$

Sumando las primeras derivadas de los dos términos de la log verosimilitud, la función Score que se obtiene es

$$S_c(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2} \left[\begin{array}{c} -2 \sum_{i=1}^m \mathbf{X}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\beta}\mathbf{X}_i) \\ \sum_{i=1}^m \left(\text{tr} \left(\mathbf{V}_i^{-1} \frac{\partial\mathbf{V}_i}{\partial\theta_k} \right) - (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{A}_{ik} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \right) \end{array} \right].$$

Luego, menos la suma de la segundas derivadas de los dos términos se obtiene la infor-

mación observada, la cual queda expresada como

$$\mathbf{J}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \begin{pmatrix} \sum_{i=1}^m \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i & \sum_{i=1}^m \mathbf{X}_i^T \mathbf{A}_{ik} \mathbf{r}_i \\ \sum_{i=1}^m \mathbf{X}_i^T \mathbf{A}_{ik} \mathbf{r}_i & \frac{1}{2} \sum_{i=1}^m \left(\text{tr} \left[\mathbf{V}_i^{-1} \frac{\partial^2 \mathbf{V}_i}{\partial \theta_h \partial \theta_k} - \mathbf{A}_{ih} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \right] - \mathbf{r}_i^T \frac{\partial \mathbf{A}_{ik}}{\partial \theta_h} \mathbf{r}_i \right) \end{pmatrix},$$

donde

$$\mathbf{r} = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}, \quad \mathbf{A}_{ik} = \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \mathbf{V}_i^{-1} \quad (7.7)$$

y

$$\frac{\partial \mathbf{A}_{ik}}{\partial \theta_h} = \mathbf{V}_i^{-1} \left(-\frac{\partial \mathbf{V}_i}{\partial \theta_h} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} + \frac{\partial^2 \mathbf{V}_i}{\partial \theta_h \partial \theta_k} - \frac{\partial \mathbf{V}_i}{\partial \theta_k} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_h} \right) \mathbf{V}_i^{-1}. \quad (7.8)$$

7.2 Información de Fisher

Por definición, la información de Fisher esta dada como

$$\begin{aligned} \mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\phi}) &= E \mathbf{J}(\boldsymbol{\beta}, \boldsymbol{\phi}) \\ &= \begin{pmatrix} \sum_{i=1}^m E [\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i] & \sum_{i=1}^m E [\mathbf{X}_i^T \mathbf{A}_{ik} \mathbf{r}_i] \\ \sum_{i=1}^m E [\mathbf{X}_i^T \mathbf{A}_{ik} \mathbf{r}_i] & \frac{1}{2} \sum_{i=1}^m E \left(\text{tr} \left[\mathbf{V}_i^{-1} \frac{\partial^2 \mathbf{V}_i}{\partial \theta_h \partial \theta_k} - \mathbf{A}_{ih} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \right] - \mathbf{r}_i^T \frac{\partial \mathbf{A}_{ik}}{\partial \theta_h} \mathbf{r}_i \right) \end{pmatrix}; \end{aligned}$$

es decir,

$$\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\phi}) = \begin{pmatrix} \sum_{i=1}^m E [\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i] & \sum_{i=1}^m E [\mathbf{X}_i^T \mathbf{A}_{ik} \mathbf{r}_i] \\ \sum_{i=1}^m E [\mathbf{X}_i^T \mathbf{A}_{ik} \mathbf{r}_i] & \frac{1}{2} \sum_{i=1}^m E \left(\text{tr} \left[\mathbf{V}_i^{-1} \frac{\partial^2 \mathbf{V}_i}{\partial \theta_h \partial \theta_k} - \mathbf{A}_{ih} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \right] - \mathbf{r}_i^T \frac{\partial \mathbf{A}_{ik}}{\partial \theta_h} \mathbf{r}_i \right) \end{pmatrix}.$$

Dado que

$$\begin{aligned} E \mathbf{r}_i &= E [\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}] = 0 \quad y \\ E \text{tr} \left(\mathbf{r}_i^T \frac{\partial \mathbf{A}_{ik}}{\partial \theta_h} \mathbf{r}_i \right) &= \text{tr} \left\{ \frac{\partial \mathbf{A}_{ik}}{\partial \theta_h} E [\mathbf{r}_i \mathbf{r}_i^T] \right\} = \text{tr} \left\{ \frac{\partial \mathbf{A}_{ik}}{\partial \theta_h} \mathbf{V}_i \right\}, \end{aligned}$$

entonces la información de Fisher queda expresada como

$$\begin{aligned} \mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\phi}) &= \begin{pmatrix} \sum_{i=1}^m \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^m \text{tr} \left[\mathbf{V}_i^{-1} \frac{\partial^2 \mathbf{V}_i}{\partial \theta_h \partial \theta_k} - \mathbf{A}_{ih} \frac{\partial \mathbf{V}_i}{\partial \theta_k} - \frac{\partial \mathbf{A}_{ik}}{\partial \theta_h} \mathbf{V}_i \right] \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^m \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^m \text{tr} \left[\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_h} \right] \end{pmatrix}; \end{aligned}$$

La última igualdad se sigue al sustituir las expresiones de \mathbf{A}_{ih} y $\frac{\partial \mathbf{A}_{ik}}{\partial \theta_h}$; ver (7.7) y (7.8).

El desarrollo se muestra en seguida.

$$\begin{aligned}
& \mathbf{V}_i^{-1} \frac{\partial^2 \mathbf{V}_i}{\partial \theta_h \partial \theta_k} - \mathbf{A}_{ih} \frac{\partial \mathbf{V}_i}{\partial \theta_k} - \frac{\partial \mathbf{A}_{ik}}{\partial \theta_h} \mathbf{V}_i \\
= & \mathbf{V}_i^{-1} \frac{\partial^2 \mathbf{V}_i}{\partial \theta_h \partial \theta_k} - \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_h} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} + \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_h} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \mathbf{V}_i^{-1} \mathbf{V}_i - \mathbf{V}_i^{-1} \frac{\partial^2 \mathbf{V}_i}{\partial \theta_h \partial \theta_k} \mathbf{V}_i^{-1} \mathbf{V}_i \\
& + \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_h} \mathbf{V}_i^{-1} \mathbf{V}_i \\
= & \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_h} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k}.
\end{aligned}$$

Dado que en la tesis cada uno de los modelos considerados se trabaja por separado, a continuación se obtiene la información esperada en los modelos ya mencionados.

7.2.1 Modelo con intercepto aleatorio

En el modelo con intercepto aleatorio la matriz de covarianza de \mathbf{Y}_i es

$$\mathbf{V}_i(\boldsymbol{\theta}) = \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \sigma^2 \mathbf{I} = \sigma_0^2 \mathbf{Z}_i \mathbf{Z}_i^T + \sigma^2 \mathbf{I},$$

con $\mathbf{Z}_i^T = (1, 1, \dots, 1)_{1 \times n_i}$. En este caso $\boldsymbol{\theta} = (\theta_h, \theta_k) = (\sigma_0, \sigma)$. Entonces

$$\begin{aligned}
\frac{\partial \mathbf{V}_i}{\partial \sigma_0} &= 2\sigma_0 \mathbf{Z}_i \mathbf{Z}_i^T, & \frac{\partial^2 \mathbf{V}_i}{\partial^2 \sigma_0} &= 2\mathbf{Z}_i \mathbf{Z}_i^T, \\
\frac{\partial \mathbf{V}_i}{\partial \sigma} &= 2\sigma \mathbf{I}, & \frac{\partial^2 \mathbf{V}_i}{\partial^2 \sigma} &= 2\mathbf{I}, & \frac{\partial \mathbf{V}_i}{\partial \sigma \partial \sigma_0} &= 0.
\end{aligned}$$

De este modo la información esperada es

$$\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \begin{pmatrix} \sum_{i=1}^m \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i & 0 & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^m \text{tr} [4\sigma_0^2 \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T] & \frac{1}{2} \sum_{i=1}^m \text{tr} [4\sigma_0 \sigma \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1}] \\ 0 & \frac{1}{2} \sum_{i=1}^m \text{tr} [4\sigma_0 \sigma \mathbf{V}_i^{-1} \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T] & \frac{1}{2} \sum_{i=1}^m \text{tr} [4\sigma^2 \mathbf{V}_i^{-1} \mathbf{V}_i^{-1}] \end{pmatrix};$$

donde $\text{tr} [4\sigma_0 \sigma \mathbf{V}_i^{-1} \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T] = \text{tr} [4\sigma_0 \sigma \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1}]$.

7.2.2 Modelo con pendiente aleatoria

En el modelo con pendiente aleatoria la matriz de covarianza de \mathbf{Y}_i

$$\mathbf{V}_i(\boldsymbol{\theta}) = \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \sigma^2 \mathbf{I} = \sigma_1^2 \mathbf{Z}_i \mathbf{Z}_i^T + \sigma^2 \mathbf{I}$$

con $\mathbf{Z}_i^T = (t_1, t_2, \dots, t_{n_i})_{1 \times n_i}$. En este caso $\boldsymbol{\theta} = (\theta_h, \theta_k) = (\sigma_1, \sigma)$. Entonces las derivadas son

$$\begin{aligned} \frac{\partial \mathbf{V}_i}{\partial \theta_h} &= 2\sigma_1 \mathbf{Z}_i \mathbf{Z}_i^T, & \frac{\partial^2 \mathbf{V}_i}{\partial^2 \theta_h} &= 2\mathbf{Z}_i \mathbf{Z}_i^T, \\ \frac{\partial \mathbf{V}_i}{\partial \theta_k} &= 2\sigma \mathbf{I}, & \frac{\partial^2 \mathbf{V}_i}{\partial^2 \theta_k} &= 2\mathbf{I}, & \frac{\partial \mathbf{V}_i}{\partial \theta_h \partial \theta_k} &= 0. \end{aligned}$$

De este modo la información esperada es

$$\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \begin{pmatrix} \sum_{i=1}^m \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i & 0 & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^m \text{tr} [4\sigma_1^2 \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T] & \frac{1}{2} \sum_{i=1}^m \text{tr} [4\sigma_1 \sigma \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1}] \\ 0 & \frac{1}{2} \sum_{i=1}^m \text{tr} [4\sigma_1 \sigma \mathbf{V}_i^{-1} \mathbf{V}_i^{-1} \mathbf{Z}_i \mathbf{Z}_i^T] & \frac{1}{2} \sum_{i=1}^m \text{tr} [4\sigma^2 \mathbf{V}_i^{-1} \mathbf{V}_i^{-1}] \end{pmatrix}.$$

7.2.3 Modelo con intercepto y pendiente aleatorios

La matriz de covarianza de \mathbf{Y}_i

$$\mathbf{V}_i(\boldsymbol{\theta}) = \mathbf{Z}_i \boldsymbol{\Psi} \mathbf{Z}_i^T + \sigma^2 \mathbf{I}$$

donde $\mathbf{Z}_i^T = \begin{pmatrix} 1 & \cdots & 1 \\ t_{i1} & \cdots & t_{in_i} \end{pmatrix}$. En este caso $\boldsymbol{\Psi} = \begin{pmatrix} \sigma_0^2 & \sigma_0 \sigma_1 \rho \\ \sigma_0 \sigma_1 \rho & \sigma_1^2 \end{pmatrix}$ con $-1 < \rho < 1$.

Aquí $\boldsymbol{\phi} = (\theta_h, \theta_k, \theta_l) = (\sigma_0, \sigma_1, \rho, \sigma)$.

Seguendo el esquema de los modelos anteriores, las derivadas son

$$\begin{aligned}
\frac{\partial \mathbf{V}_i}{\partial \sigma_0} &= \mathbf{Z}_i \begin{pmatrix} 2\sigma_0 & \sigma_1\rho \\ \sigma_1\rho & 0 \end{pmatrix} \mathbf{Z}_i^T; & \frac{\partial^2 \mathbf{V}_i}{\partial^2 \sigma_0} &= \mathbf{Z}_i \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \mathbf{Z}_i^T, \\
\frac{\partial \mathbf{V}_i}{\partial \sigma_1} &= \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0\rho \\ \sigma_0\rho & 2\sigma_1 \end{pmatrix} \mathbf{Z}_i^T, & \frac{\partial^2 \mathbf{V}_i}{\partial^2 \sigma_1} &= \mathbf{Z}_i \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} \mathbf{Z}_i^T, \\
\frac{\partial \mathbf{V}_i}{\partial \rho} &= \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0\sigma_1 \\ \sigma_0\sigma_1 & 0 \end{pmatrix} \mathbf{Z}_i^T, & \frac{\partial^2 \mathbf{V}_i}{\partial \sigma_0 \partial \sigma_1} &= \mathbf{Z}_i \begin{pmatrix} 0 & \rho \\ \rho & 0 \end{pmatrix} \mathbf{Z}_i^T, \\
\frac{\partial \mathbf{V}_i}{\partial \sigma} &= 2\sigma \mathbf{I}, & \frac{\partial^2 \mathbf{V}_i}{\partial^2 \sigma} &= 2\mathbf{I}, & \frac{\partial^2 \mathbf{V}_i}{\partial \sigma_0 \partial \rho} &= \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_1 \\ \sigma_1 & 0 \end{pmatrix} \mathbf{Z}_i^T, \\
\frac{\partial^2 \mathbf{V}_i}{\partial \sigma_1 \partial \rho} &= \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix} \mathbf{Z}_i^T.
\end{aligned}$$

Entonces la información esperada es de la forma

$$\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \begin{pmatrix} I_{11} & 0 & 0 & 0 & 0 \\ 0 & I_{22} & I_{23} & I_{24} & I_{25} \\ 0 & I_{32} & I_{33} & I_{34} & I_{35} \\ 0 & I_{42} & I_{43} & I_{44} & I_{45} \\ 0 & I_{52} & I_{53} & I_{54} & I_{55} \end{pmatrix},$$

donde los elementos de la matriz $I(\boldsymbol{\beta}, \boldsymbol{\theta})$ están dados como

$$I_{11} = \sum_{i=1}^m \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i,$$

$$I_{22} = \frac{1}{2} \sum_{i=1}^m \text{tr} \left[\mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 2\sigma_0 & \sigma_1 \rho \\ \sigma_1 \rho & 0 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 2\sigma_0 & \sigma_1 \rho \\ \sigma_1 \rho & 0 \end{pmatrix} \mathbf{Z}_i^T \right],$$

$$I_{33} = \frac{1}{2} \sum_{i=1}^m \text{tr} \left[\mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \rho \\ \sigma_0 \rho & 2\sigma_1 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \rho \\ \sigma_0 \rho & 2\sigma_1 \end{pmatrix} \mathbf{Z}_i^T \right],$$

$$I_{44} = \frac{1}{2} \sum_{i=1}^m \text{tr} \left[\mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \sigma_1 \\ \sigma_0 \sigma_1 & 0 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \sigma_1 \\ \sigma_0 \sigma_1 & 0 \end{pmatrix} \mathbf{Z}_i^T \right],$$

$$I_{55} = \frac{1}{2} \sum_{i=1}^m \text{tr} [4\sigma^2 \mathbf{V}_i^{-1} \mathbf{V}_i^{-1}],$$

$$I_{23} = I_{32} = \frac{1}{2} \sum_{i=1}^m \text{tr} \left[\mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 2\sigma_0 & \sigma_1 \rho \\ \sigma_1 \rho & 0 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \rho \\ \sigma_0 \rho & 2\sigma_1 \end{pmatrix} \mathbf{Z}_i^T \right],$$

$$I_{24} = I_{42} = \frac{1}{2} \sum_{i=1}^m \text{tr} \left[\mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 2\sigma_0 & \sigma_1 \rho \\ \sigma_1 \rho & 0 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \sigma_1 \\ \sigma_0 \sigma_1 & 0 \end{pmatrix} \mathbf{Z}_i^T \right],$$

$$I_{25} = I_{52} = \frac{1}{2} \sum_{i=1}^m \text{tr} \left[2\sigma \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 2\sigma_0 & \sigma_1 \rho \\ \sigma_1 \rho & 0 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \right],$$

$$I_{34} = I_{43} = \frac{1}{2} \sum_{i=1}^m \text{tr} \left[\mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \sigma_1 \\ \sigma_0 \sigma_1 & 0 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \rho \\ \sigma_0 \rho & 2\sigma_1 \end{pmatrix} \mathbf{Z}_i^T \right],$$

$$I_{35} = I_{53} = \frac{1}{2} \sum_{i=1}^m \text{tr} \left[2\sigma \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \rho \\ \sigma_0 \rho & 2\sigma_1 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \right],$$

$$I_{45} = I_{54} = \frac{1}{2} \sum_{i=1}^m \text{tr} \left[2\sigma \mathbf{V}_i^{-1} \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_0 \sigma_1 \\ \sigma_0 \sigma_1 & 0 \end{pmatrix} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \right].$$

7.3 Vector gradiente $\frac{\partial}{\partial \phi_r} t_p(\phi_r)$

Para obtener una aproximación de la varianza de $t_p(\hat{\phi}_r)$ en el modelo con pendiente e intercepto aleatorio, se requiere calcular el vector de gradiente $\frac{\partial}{\partial \phi_r} t_p(\phi_r)$. Sus elementos se obtienen a continuación:

$$\begin{aligned} \frac{\partial}{\partial \beta_0} t_p(\phi_r) &= \frac{1}{\beta_1^2 - z_p^2 \sigma_1^2} \left(-\beta_1 - |z_p| \frac{2(D_f - \beta_0) \sigma_1^2}{2\sqrt{(D_f - \beta_0)^2 \sigma_1^2 + (\beta_1^2 - z_p^2 \sigma_1^2) \sigma_0^2}} \right) \\ &= -\frac{1}{\beta_1^2 - z_p^2 \sigma_1^2} \left(\beta_1 + \frac{(D_f - \beta_0) |z_p| \sigma_1^2}{\sqrt{(D_f - \beta_0)^2 \sigma_1^2 + (\beta_1^2 - z_p^2 \sigma_1^2) \sigma_0^2}} \right) \\ &= -\frac{1}{A} \left(\beta_1 + \frac{(D_f - \beta_0) |z_p| \sigma_1^2}{D} \right), \end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \beta_1} t_p &= \frac{1}{A^2} \left(A \left[(D_f - \beta_0) - \frac{\beta_1 |z_p| \sigma_0^2}{D} \right] - 2\beta_1 (B - |z_p| D) \right) \\
&= \frac{1}{A^2} \left(\frac{A}{D} [(D_f - \beta_0) D - \beta_1 |z_p| \sigma_0^2] - 2\beta_1 (B - |z_p| D) \right) \\
&\quad \frac{1}{AD} ((D_f - \beta_0) D - |z_p| \beta_1 \sigma_0^2) - 2 \frac{\beta_1 (B - |z_p| D)}{A^2},
\end{aligned}$$

$$\frac{\partial}{\partial \sigma_0} t_p = -\frac{|z_p| \sigma_0}{D},$$

$$\begin{aligned}
\frac{\partial}{\partial \sigma_1} t_p &= \frac{1}{A^2} \left(A \left[-\frac{|z_p|}{D} [(D_f - \beta_0)^2 - z_p^2 \sigma_0^2] \sigma_1 \right] - 2(B - |z_p| D) z_p^2 \sigma_1 \right) \\
&= -\frac{\sigma_1}{A^2} \left(\frac{A}{D} |z_p| [(D_f - \beta_0)^2 - z_p^2 \sigma_0^2] + 2(B - |z_p| D) z_p^2 \right), \tag{7.9}
\end{aligned}$$

donde

$$A = \beta_1^2 - z_p^2 \sigma_1^2,$$

$$B = (D_f - \beta_0) \beta_1 + z_p^2 \sigma_0 \sigma_1 \rho,$$

$$C = (D_f - \beta_0)^2 - z_p^2 \sigma_0^2,$$

$$D = \{B^2 - AC\}^{1/2}.$$

Bibliografía

- [1] Díaz-Frances, E. (1998). *Scientific Application of Maximum Likelihood in Multiparametric Problems*. Ph. D. Thesis, Centro de Investigación en Matemáticas (CIMAT), Guanajuato, México.
- [2] Efron, B. and D. Hinkley (1978). Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information. *Biometrika*, **65**, No. 3, 457-487.
- [3] Laird, N. M. and J. H. Ware (1982). Random-effects models for longitudinal data, *Biometrics*, **38**, 963-974.
- [4] Lindstrom, M.J. and D.M. Bates (1988). Newton-Raphson and EM Algorithms for Linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, **83**, 1014-1022.
- [5] Lindstrom, M.J. and D.M. Bates (1990). Nonlinear mixed effects models for repeated measure data, *Biometrics* **46** , 673-687.

- [6] Lu, C. J. and W. Q. Meeker (1993). Using Degradation Measures to Estimate a Time-to-Failure Distribution, *Technometrics*, **34**, 161-174.
- [7] McCulloch, Ch. E. and S. R. Searle (2001). *Generalized, Linear, and Mixed Models*. John Wiley & Sons.
- [8] Meeker, W. Q. and L. A. Escobar (1998). *Statistical Methods for Reliability Data*. John Wiley & Sons, Inc.
- [9] Peña, E. (1993). *Desarrollo de una laca para recubrir acero inoxidable*. Tesina de Especialidad en Ingeniería de la Calidad, CIMAT.
- [10] Pinheiro, J. C. *Topics in Mixed-Effects Models* (1994). Ph. D. thesis, University of Wisconsin, Madison, WI.
- [11] Pinheiro, J. C. and D. M. Bates (2000). *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag .
- [12] Searle, S. R., G. Casella, and C. E. McCulloch (1992). *Variance Components*. John Wiley & Sons, New York, .
- [13] Tseng, S. T. and H. F. Yu (1997). A termination rule for degradation experiments, *IEEE Transactions on Reliability*, **46**, 130-133.
- [14] Yu, H. F. and S. T. Tseng (1998). On-line procedure for terminating an accelerated degradation test, *Statistical Sinica*, **8**, 207-220.

- [15] Yu, H. F. and Ch. Chiao (2002). An Optimal Designed Degradation Experiment for Reliability Improvement, *IEEE Transactions on Reliability*, **51**, No. 4, 427-433.
- [16] Yu, H. F. and S. T. Tseng (1999). Designing a degradation experiment. *Naval Research Logistics*, 689-706.