



UNIVERSIDAD NACIONAL
AUTÓNOMA

Unidad Académica de los Ciclos Profesionales
y de Posgrado del CCH

Instituto de Investigaciones en Matemáticas Aplicadas
y en Sistemas

LA FUNCION DE INFLUENCIA EN EL ANALISIS DE

DATOS DE SUPERVIVENCIA

T E S I S

que para obtener el grado de
MAESTRA EN ESTADISTICA E
INVESTIGACION DE OPERACIONES
presenta la Actuaría
GLORIA MA. BELEM TREJO VALDIVIA

BIBLIOTECA
JUAN A. ESCALANTE H.
UNIDAD ACADEMICA DE
LOS CICLOS PROFESIONAL
Y DE POSGRADO / CCH
UNAM

MEXICO, D.F.

Enero 1985



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A mis papás, Coty y Rafael.
Por todas las cosas que me
han brindado, apoyo, com-
prensión y amor, entre otras.

A mis hermanos, Ana Laura y Rafael.
Por todo lo que hemos hecho juntos.

A mi sobrina Cynthia.
Por toda la alegría que
me da día a día.

A todos y cada uno de
aquellos con quienes he
podido compartir por lo
menos un momento de fe-
licidad.

A todos aquellos que
han estado conmigo cuando
los necesité, mis amigos.

A la memoria del pe-
queño Alfredito.

A G R A D E C I M I E N T O S

Quiero expresar mi mas sincero agradecimiento al Dr. Francisco J. Aranda por haberme orientado hacia el estudio de esta area de la Estadística que tanto me ha interesado, por su paciente y entusiasta dirección en el desarrollo de este trabajo así como por el gran interés y preocupación que ha demostrado por mi superación académica.

Agradezco muy especialmente al Dr. Ignacio Méndez por toda la ayuda que me ha brindado y porque sé -- que cuento con su amistad.

Agradezco al Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas (IIMAS) por todas las facilidades brindadas para el desarrollo de este -- trabajo.

TACTICA Y ESTRATEGIA

MI TACTICA ES MIRARTE
APRENDER COMO ERES
QUERERTE COMO ERES

MI TACTICA ES HABLARTE
Y ESCUCHARTE CONSTRUIR
CON PALABRAS
UN PUENTE INDESTRUCTIBLE

MI TACTICA ES QUEDARME
EN TU RECUERDO
NO SE COMO, NI SE
CON QUE PRETEXTO
PERO QUEDARME EN EL

MI TACTICA ES SER FRANCA
Y SABER QUE ERES FRANCO
Y QUE NO NOS VENDAMOS
SIMULACROS

PARA QUE ENTRE LOS DOS
NO HAY TELON NI ABISMOS

MI ESTRATEGIA ES EN CAMBIO
MAS PROFUNDA
Y MAS SIMPLE

MI ESTRATEGIA ES QUE
UN DIA CUALQUIERA
NO SE COMO NI SE
CON QUE PRETEXTO
POR FIN ME NECESITES.

Mario Benedetti

LA FUNCION DE INFLUENCIA EN EL ANALISIS DE

DATOS DE SUPERVIVENCIA

I N D I C E

1. INTRODUCCION	1
2. LA FUNCION DE INFLUENCIA	5
2.1 LA FUNCION O CURVA DE INFLUENCIA TEORICA	5
2.2 EJEMPLOS	8
2.2.1 LA MEDIA μ	8
2.2.2 LA VARIANZA σ^2	10
2.2.3 LA MEDIANA	11
2.2.4 EL CUANTIL q	16
2.2.5 LA MEDIA α -RECORTADA μ_α	18
2.2.6 ESTIMADORES M	23
2.2.7 ESTIMADORES L	25
2.2.8 ESTIMADORES R	27
2.2.9 COEFICIENTE DE CORRELACION	32
2.3 RELACION ENTRE LA FUNCION DE INFLUENCIA Y OTRAS FUNCIONES SIMILARES	35
2.3.1 CURVA DE SENSIBILIDAD	35
2.3.2 EL JACKKNIFE	38
2.4 LA FUNCION DE INFLUENCIA MUESTRAL	41
2.4.1 PRIMERA VERSION	42
2.4.1.1 EJEMPLOS	43
2.4.2 SEGUNDA VERSION	48
2.4.2.1 EJEMPLOS	49
2.4.3 TERCERA VERSION	59
3. DATOS DE SUPERVIVENCIA	61
3.1 INTRODUCCION	61
3.2 EJEMPLOS	62
3.3 DISTRIBUCION DEL TIEMPO DE FALLA O SUPERVIVENCIA	63

3.4 ESTIMACION DE LA FUNCION DE SUPERVIVENCIA	72
3.4.1 LA FUNCION DE SUPERVIVENCIA EMPIRICA	72
3.4.2 LA TABLA DE VIDA	73
3.4.3 LA TABLA DE VIDA MODIFICADA	75
3.4.4 EL ESTIMADOR KAPLAN-MEIER	77
3.5 MODELOS PARA TIEMPO DE FALLA	88
3.5.1 MODELO EXPONENCIAL	89
3.5.2 MODELO WEIBULL	90
3.5.3 MODELO LOG-NORMAL	91
3.5.4 MODELO GAMMA	93
3.6 MODELOS PARA POBLACIONES HETEROGENEAS	95
4. LA FUNCION DE INFLUENCIA EN DATOS DE SUPERVIVENCIA	100
4.1 INTRODUCCION	100
4.2 CURVA DE INFLUENCIA PROPUESTA PARA DATOS DE SUPERVIVENCIA	105
4.2.1 INTRODUCCION	105
4.2.2 VALOR PROMEDIO DEL ESTIMADOR KAPLAN-MEIER	106
4.2.3 LA CURVA DE INFLUENCIA	127
4.2.4 EJEMPLO	139
5. CONCLUSIONES	162
6. APENDICES	164
6.1 APENDICE 1	164
6.2 APENDICE 2	170
6.3 APENDICE 3	172
6.4 APENDICE 4	174
7. REFERENCIAS	179

1. INTRODUCCION

En cualquier análisis o estudio de tipo estadístico uno de los principales objetivos es el de extraer la mayor información posible acerca de una población cuando solo se observa una parte de ella. A menudo dicha información se obtiene en base a estimadores de los parámetros poblacionales. Dado este hecho, el valor y el comportamiento asintótico de estos estimadores dependerá principalmente de todas y cada una de las observaciones de la muestra obtenida, sin embargo, surge la pregunta ¿Efectivamente todas las observaciones tienen un peso igual en el resultado del análisis?.

Hay que tener cuidado al responderla pues aparentemente uno podría darle un igual peso a las observaciones y sin embargo estas afectan de manera diferente. Un ejemplo clásico de esta situación es en la estimación de la media poblacional para cualquier variable continua. Como ya sabemos, esta estimación se hace a partir de la media muestral o sea con el promedio aritmético de las observaciones obtenidas entonces, uno podría pensar que todas las observaciones tienen el mismo peso pues todos son afectados por el término $1/n$ por lo cual afectan al valor \bar{X} de igual manera.

Sin embargo esto no se da en la realidad, pues como también sabemos, uno de los problemas de la media muestral es que es muy sensible a valores extremos, por lo que si en la muestra se obtuviera un valor de este tipo, a pesar de estar ponderado por $1/n$, este "moverá" significativamente el valor de \bar{X} , lo cual no

ocurre con las observaciones localizadas en el rango esperado, es mas, entre mas cercanas estén dichas observaciones al valor de la media estas afectarán en menor grado el valor de \bar{X} .

El problema anterior se va a presentar de hecho en cualquier tipo de análisis estadístico, quizá no sea tan fácil detectarlo como en el ejemplo anterior pero hay que tomar en cuenta su existencia, de aquí que sea deseable tener una herramienta de tipo estadístico que nos permita principalmente

1. Apreciar la influencia o peso relativo de las observaciones individuales en el valor de una estadística o dentro de un procedimiento o análisis estadístico.
2. Cuando el procedimiento estadístico se refiere a la estimación de parámetros poblacionales, calcular las propiedades asintóticas de dichos estimadores.

Un medio relativamente reciente para lograr los objetivos anteriores es la llamada curva o función de influencia introducida por Hampel(1968).

El alcance de conocimiento sobre la función de influencia se ha centrado principalmente en la parte correspondiente a la estimación del parámetro poblacional, el cual no necesariamente es tratado como unidimensional, es decir, se tiene actualmente un buen conocimiento del comportamiento de estimadores utilizando la función de influencia de ellos para cualquier tipo de parámetro p-dimensional.

Dentro del análisis estadístico se encuentra el tema conocido como análisis de datos de supervivencia, en el cual, el parámetro

de interés no es de tipo p -dimensional o mejor dicho, es m -dimensional, es decir, el parámetro de estudio es toda una función, en donde dicha función refleja el comportamiento general del tiempo de falla de los individuos observados.

Una de las funciones más comúnmente utilizadas en este tipo de análisis es la función de supervivencia, la cual es esencialmente el complemento de la función de distribución del tiempo de falla. Para esta función se han propuesto varios estimadores pero el más usado es el estimador Kaplan-Meier.

Por lo anterior, era necesario, encontrar una expresión para la curva de influencia del estimador Kaplan-Meier de la función de supervivencia, la cual fue obtenida por Reid(1979). Esta función corresponde en realidad a dos funciones de influencia, una para las observaciones no censuradas y la otra para las que sí lo son. Hay que notar que en este tipo de estudios se tiene una característica especial en las observaciones muestrales que es la posibilidad de censura.

Sin embargo, resulta un poco complicado el manejo de la función de influencia teórica; también esta no permite medir la influencia de observaciones contaminadas, las cuales poseen una cierta importancia en el análisis de datos de supervivencia.

El presente trabajo tiene como finalidad presentar de manera general la teoría y aplicaciones desarrolladas acerca de la función de influencia, por lo cual se desarrolla una parte correspondiente a la función de influencia teórica. De igual

manera se presentan las versiones muestrales que se han propuesto para dicha función, pues como se verá posteriormente, su cálculo se basa en la suposición del conocimiento de la verdadera función de distribución de la población en estudio, lo cual no se da en la realidad. Además es conveniente tener una función de influencia muestral para evaluar el efecto de las observaciones realmente obtenidas en un caso dado. Cada una de las versiones muestrales se presentará con ejemplos de la aplicación que han tenido.

Dada la relativamente poca difusión que tiene el análisis de datos de supervivencia se presenta todo un capítulo acerca de este tema, en el cual se tratan las principales funciones que se estudian en este tipo de datos, así como las suposiciones, características generales, modelos, etc.

Se presenta brevemente el trabajo desarrollado por Reid así como las expresiones que obtuvo de las funciones de influencia utilizando para esto la descomposición del estimador Kaplan-Meier en funciones de sub-supervivencia propuesta por Peterson(1977).

Finalmente, se propone una expresión alternativa de la curva de influencia del estimador Kaplan-Meier de la función de supervivencia. El comportamiento de la expresión propuesta no es fácilmente obtenido de manera general por lo que se estudió dicho comportamiento para algunos casos particulares, estos casos involucran diferentes tamaños de muestra, diferentes niveles de censura y diferentes grados de contaminación.

2. LA FUNCION DE INFLUENCIA

2.1 LA FUNCION O CURVA DE INFLUENCIA TEORICA

Consideremos una población con función de distribución $F(x, \theta)$ $\equiv F$ donde θ es el parámetro poblacional desconocido (no necesariamente unidimensional). Supongamos que θ puede ser visto como (o aproximado por) una funcional lineal T , tal que:

$$T : M \rightarrow \Theta$$

donde M es el espacio de todas las medidas de probabilidad sobre (\mathcal{R}, B) con B el σ -álgebra de Borel y Θ es el espacio parametral en estudio, es decir,

$$\theta = T(F)$$

Dado $x \in \mathcal{R}$ consideremos la función de distribución \tilde{F} dada por

$$\tilde{F}(\cdot) = (1 - \epsilon)F(\cdot) + \epsilon \delta_x(\cdot)$$

con $\epsilon > 0$

donde

$$\delta_x(y) = \begin{cases} 0 & y < x \\ 1 & y \geq x \end{cases} .$$

Entonces la curva de influencia de T en el punto x , que no será otra cosa que un índice del peso que tiene dicho valor x sobre el

correspondiente valor θ a través de la funcional T , está dada por

$$CI_{T,F}(x; \theta) = \lim_{\epsilon \rightarrow 0^+} \frac{T(\tilde{F}) - T(F)}{\epsilon} = \lim_{\epsilon \rightarrow 0^+} \frac{T((1-\epsilon)F + \epsilon \delta_x) - T(F)}{\epsilon}.$$

Es decir, es esencialmente la primera derivada de la estadística vista como una funcional, en alguna distribución, que siguiendo con la idea de derivadas en el caso usual nos mide el cambio relativo de la función en el punto de interés, aunque en este caso el punto de interés es la función de distribución F .

Además, si T es suficientemente regular en el sentido que puede aproximarse satisfactoriamente con los primeros términos del desarrollo de Taylor en su versión para funcionales, y si G es otra función de distribución en el mismo espacio de tal manera que G está "cerca" (de acuerdo a una distancia entre funciones de distribuciones que se establezca) de F , se tendrá que

$$\begin{aligned} T(G) &= T(F) + \int CI_{T,F}(x; \theta) (dG - dF) + \dots \\ &= T(F) + \int CI_{T,F}(x; \theta) dG - \mu_0 + \dots, \end{aligned}$$

donde

$$\mu_0 = \int CI_{T,F}(x; \theta) dF.$$

Por otro lado sabemos que para una cierta función de distribución F , se puede tener una muy buena aproximación de esta

a través de la función de distribución empírica de una muestra extraída de la población en estudio de tamaño n suficientemente grande, que llamaremos F_n , por lo cual F_n es una función "cercana" a F y por lo tanto

$$T(F_n) = T(F) + \int CI_{T,F}(x; \theta) dF_n - \mu_0 + \dots$$

$$T(F_n) - T(F) = \frac{1}{n} \sum_{i=1}^n CI_{T,F}(x_i; \theta) - \mu_0 + \dots$$

y

$$\sqrt{n} (T(F_n) - T(F)) = \frac{1}{\sqrt{n}} \left[\sum_{i=1}^n CI_{T,F}(x_i; \theta) - n\mu_0 \right] + \dots$$

Si el tamaño de muestra n es grande, y considerando que la expresión anterior no es otra cosa que el promedio de una muestra aleatoria de la variable $CI_{T,F}(X)$, podemos aplicar el Teorema Central del Límite e ignorar los demás términos de la expansión. Por lo cual se obtendrá que el lado derecho de la expresión anterior tiene una distribución asintóticamente normal con media cero; entonces

$$\sqrt{n} (T(F_n) - T(F)) \underset{\text{asint.}}{\rightsquigarrow} N(0, J_{T,F})$$

con

$$J_{T,F} = \int (CI_{T,F}(x; \theta) - \mu_0)^2 dF$$

Como puede verse, los objetivos planteados en el capítulo anterior se alcanzan usando la curva de influencia, si consideramos un estimador de θ , $\hat{\theta}$, definido a través de una funcional $\hat{\theta} = T(\hat{F})$ en donde \hat{F} es algún estimador de F .

Finalmente, para tener una buena idea del comportamiento que tiene la curva de influencia por el hecho de estar definida como una derivada de una funcional es necesario establecer y considerar ciertos conceptos matemáticos los cuales se presentan brevemente en el Apéndice 1.

2.2 EJEMPLOS

A continuación se presentan una serie de ejemplos del cálculo de la función de influencia de varios estimadores de los parámetros más comúnmente utilizados en estadística.

2.2.1 LA MEDIA μ

Este parámetro está definido para toda medida de probabilidad con primer momento finito. La funcional que determina este parámetro está dada por

$$T(F) = \int y dF = \mu$$

Sean $\tilde{F} = (1-\epsilon)F + \epsilon\delta_x$ y $x \in \mathbb{R}$,

entonces

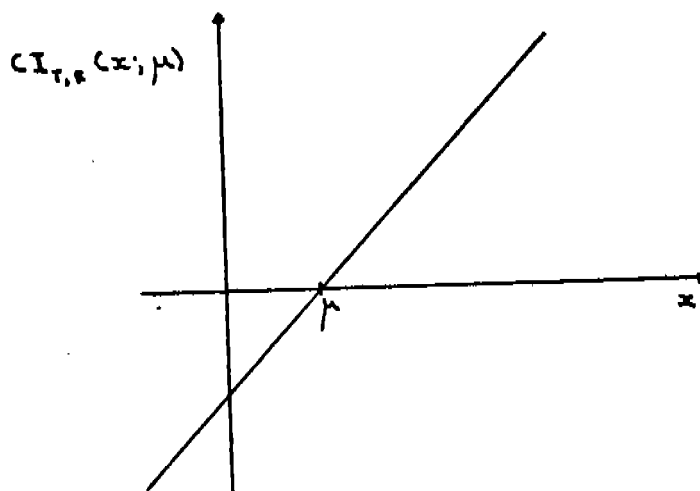
$$T(\tilde{F}) - T(F) = T((1-\epsilon)F + \epsilon\delta_x) - T(F)$$

$$\begin{aligned}
 &= \int y d((1-\epsilon)F + \epsilon\delta_x) - \int y dF \\
 &= (1-\epsilon) \int y dF + \epsilon \int y d\delta_x - \int y dF \\
 &= (1-\epsilon)\mu + \epsilon x - \mu = \epsilon(x - \mu)
 \end{aligned}$$

por lo que

$$CI_{T,F}(x; \mu) = \lim_{\epsilon \rightarrow 0^+} \frac{\epsilon(x - \mu)}{\epsilon} = x - \mu.$$

La siguiente gráfica muestra esta función de influencia.



Puede observarse que la CI no es una curva acotada, de donde se deduce que la media es un estimador muy sensible a valores extremos, además la influencia de cada observación es directamente proporcional al valor de dicha observación. Lo anterior son características ya conocidas, sin embargo esta expresión solo las está corroborando.

2.2.2 LA VARIANZA σ^2

Este parámetro está definido para toda medida de probabilidad con segundo momento finito y la funcional asociada al mismo está dada por

$$T(F) = \int (y - \mu)^2 dF = \sigma^2 \quad \text{donde} \quad \mu = \int y dF,$$

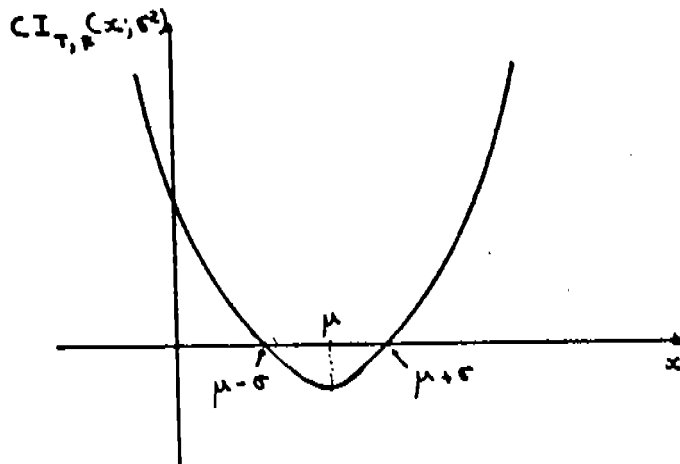
así que

$$\begin{aligned} T(\tilde{F}) - T(F) &= \int (y - \mu)^2 d((1-\epsilon)F + \epsilon\delta_x) - \int (y - \mu)^2 dF \\ &= (1-\epsilon) \int (y - \mu)^2 dF + \epsilon \int (y - \mu)^2 d\delta_x - \int (y - \mu)^2 dF \\ &= (1-\epsilon)\sigma^2 + \epsilon(x - \mu)^2 - \sigma^2 = \epsilon(x - \mu)^2 - \epsilon\sigma^2. \end{aligned}$$

Por lo tanto

$$CI_{T,F}(x; \sigma^2) = \lim_{\epsilon \rightarrow 0^+} \frac{\epsilon(x - \mu)^2 - \epsilon\sigma^2}{\epsilon} = (x - \mu)^2 - \sigma^2.$$

Esta función está representada gráficamente en la siguiente figura.



Puede observarse que este estimador también es sensible a valores extremos. Las observaciones alejadas de la media una desviación estándar no influyen en la estimación; además se tiene un valor mínimo de la curva de influencia cuando se tienen observaciones iguales a la media. Finalmente la influencia de las observaciones es la misma para valores simétricos alrededor de la media. Lo anterior también son características bastante familiares de la varianza que de nuevo se están confirmando con la expresión de la CI.

2.2.3 LA MEDIANA

Este parámetro está definido para toda función de distribución, y la funcional asociada a este es

$$T(F) = F^{-1}(1/2).$$

Consideremos ahora la identidad

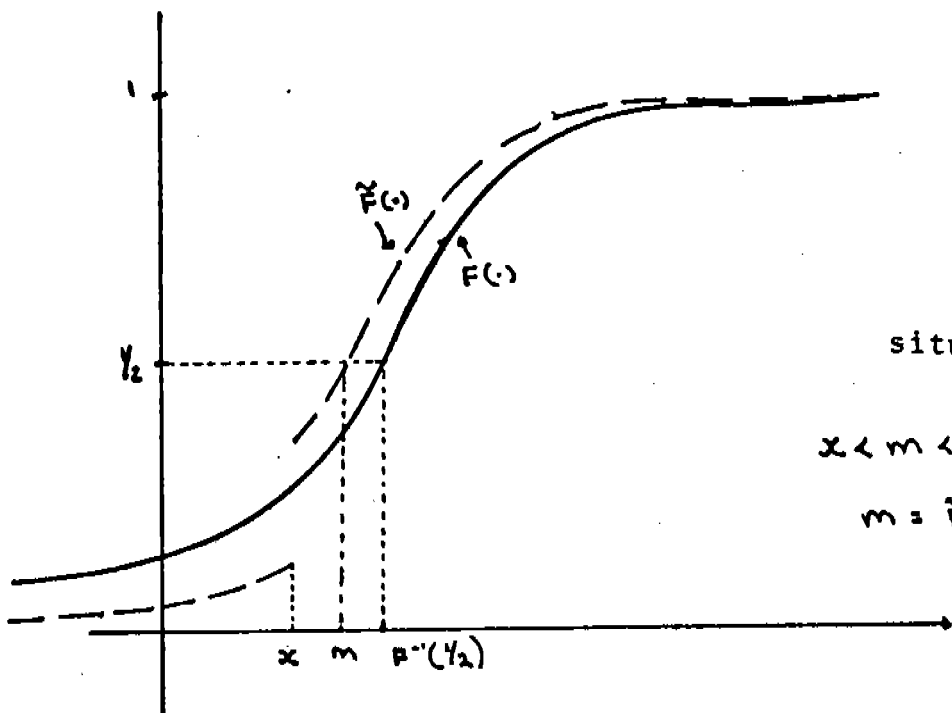
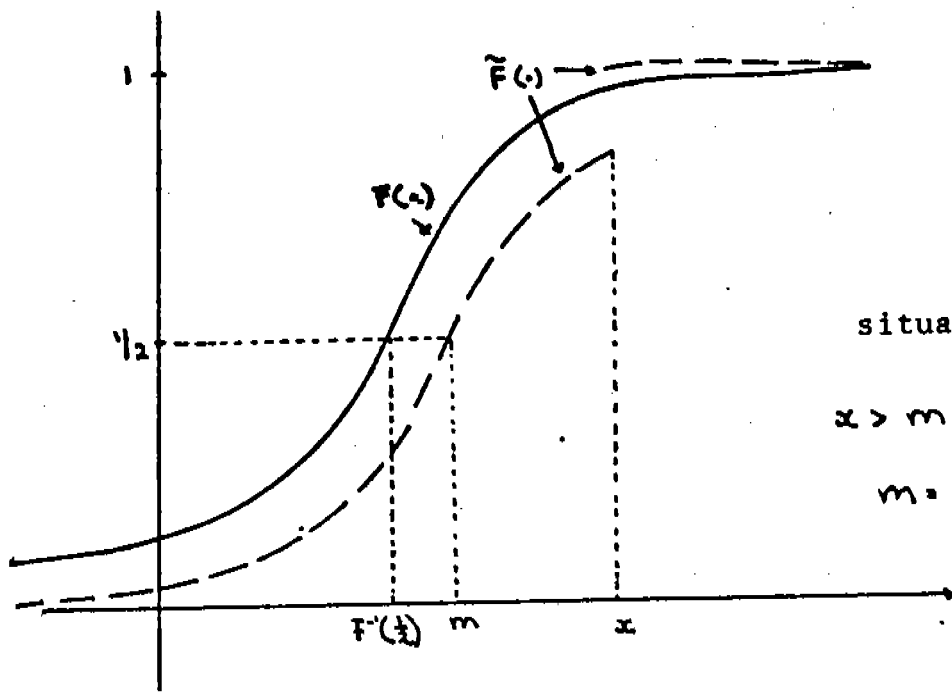
$$\tilde{F}(\tilde{F}^{-1}(1/2)) = 1/2 \quad \text{con} \quad \tilde{F} = (1-\epsilon)F + \epsilon\delta_x$$

$\epsilon > 0,$

por lo que

$$(1-\epsilon)F(\tilde{F}^{-1}(1/2)) + \epsilon\delta_x(\tilde{F}^{-1}(1/2)) = 1/2 -$$

Para este parámetro hay que notar que se tienen dos casos a estudiar; el primero es cuando el valor x de interés es menor que la mediana y el segundo cuando es mayor. Evidentemente si x es un valor igual a la mediana tendrá una influencia igual a cero en el estimador. Gráficamente las dos situaciones son



Para la situación 1, es decir, cuando $x > F^{-1}(1/2)$, $m = F^{-1}(1/2)$, tenemos

$$\begin{aligned}\tilde{F}(m) &= F(\tilde{F}^{-1}(1/2)) = 1/2 \\ &= (1-\epsilon)F(m) + \epsilon \Delta_x(m) \\ &= (1-\epsilon)F(m); \end{aligned}$$

por lo tanto

$$F(m) = \frac{1}{2(1-\epsilon)} \quad \text{y entonces} \quad m = F^{-1}\left(\frac{1}{2(1-\epsilon)}\right).$$

y ya que

$$(1-\epsilon)F(\tilde{F}^{-1}(1/2)) = (1-\epsilon)F(T(\tilde{F})) = 1/2,$$

derivando con respecto a ϵ tenemos

$$(1-\epsilon) f(T(\tilde{F})) \dot{T}(\tilde{F}) - F(T(\tilde{F})) = 0$$

y evaluando en $\epsilon=0$

$$f(T(F)) C I_{T,F}(x) - F(T(F)) = 0$$

$$f(F^{-1}(1/2)) C I_{T,F}(x) - F(F^{-1}(1/2)) = 0$$

$$f(F^{-1}(1/2)) C I_{T,F}(x) - \frac{1}{2} = 0.$$

Es decir

$$C I_{T,F}(x) = \frac{1}{2 f(F^{-1}(1/2))}.$$

Ahora bien, para la situación 2, es decir, cuando $x < F^{-1}(1/2)$, $m = F^{-1}(1/2)$, tenemos que

$$\begin{aligned}\tilde{F}(m) &= \frac{1}{2} = (1-\varepsilon)F(m) + \varepsilon\delta_x(m) \\ &= (1-\varepsilon)F(m) + \varepsilon;\end{aligned}$$

por lo tanto

$$F(m) = \frac{1/2 - \varepsilon}{1 - \varepsilon} = \frac{1 - 2\varepsilon}{2(1 - \varepsilon)},$$

es decir,

$$m = F^{-1}\left(\frac{1 - 2\varepsilon}{2(1 - \varepsilon)}\right).$$

y entonces

$$(1 - \varepsilon)F(\tilde{F}(1/2)) + \varepsilon = (1 - \varepsilon)F(T(\tilde{F})) + \varepsilon = \frac{1}{2};$$

derivando con respecto a ε tenemos

$$(1 - \varepsilon)f(T(\tilde{F}))\dot{T}(F) - F(T(\tilde{F})) + 1 = 0$$

y evaluando en $\varepsilon = 0$

$$f(T(F))CI_{T,F}(x) - F(T(F)) + 1$$

$$= f(F^{-1}(1/2))CI_{T,F}(x) - F(F^{-1}(1/2)) + 1$$

$$= f(F^{-1}(1/2))CI_{T,F}(x) + 1/2 = 0;$$

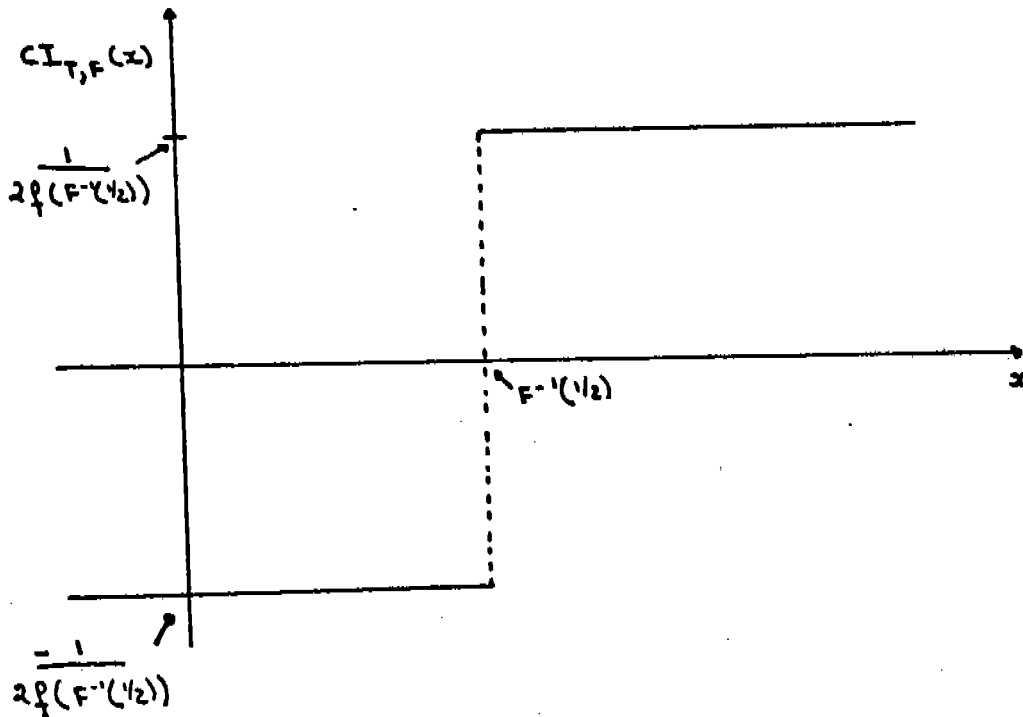
es decir,

$$CI_{T,F}(x) = \frac{-1}{2f(F^{-1}(1/2))}.$$

Resumiendo

$$CI_{T,F}(x) = \begin{cases} \frac{1}{2f(F^{-1}(1/2))} & \text{si } x > F^{-1}(1/2) \\ -\frac{1}{2f(F^{-1}(1/2))} & \text{si } x \leq F^{-1}(1/2) \end{cases}$$

Esta función se representa gráficamente en la siguiente figura.



2.2.4 EL CUANTIL q

Este parámetro está definido a través de la funcional

$$T(F) = F^{-1}(q) .$$

Para calcular la curva de influencia se considera la identidad

$$\tilde{F}(T(\tilde{F})) = q$$

con

$$\tilde{F} = (1 - \varepsilon)F + \varepsilon \delta_x .$$

Derivando con respecto a ε

$$-F(T(\tilde{F})) + \delta_x(T(\tilde{F})) + ((1-\varepsilon)f(T(\tilde{F})))\dot{T}(\tilde{F}) = 0,$$

y evaluando en $\varepsilon = 0$

$$-F(T(F)) + \delta_x(T(F)) + f(T(F))CI_{T,F}(x; q) = 0$$

$$-q + \delta_x(F^{-1}(q)) + f(F^{-1}(q))CI_{T,F}(x; q) = 0$$

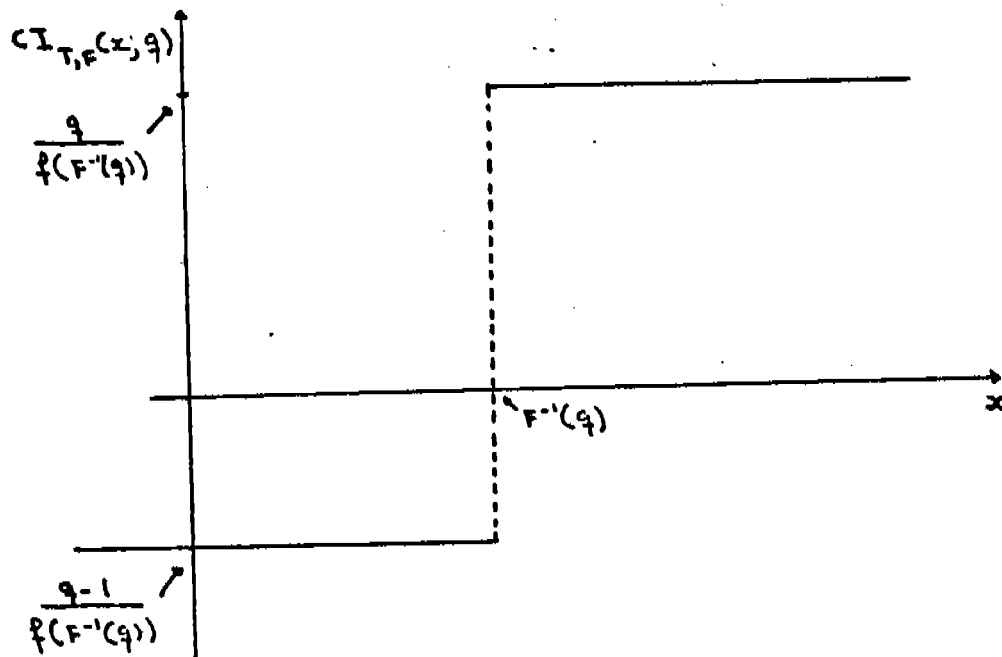
por lo tanto

$$CI_{T,F}(x; q) = \frac{q - \delta_x(F^{-1}(q))}{f(F^{-1}(q))} ;$$

es decir

$$CI_{T,F}(x; q) = \begin{cases} \frac{q}{f(F^{-1}(q))} & \text{si } x > F^{-1}(q) \\ \frac{q-1}{f(F^{-1}(q))} & \text{si } x \leq F^{-1}(q) \end{cases}$$

La situación se presenta gráficamente en la siguiente figura.



Debido a la forma de esta función, la influencia de cualquier observación solo depende de su localización con respecto al punto $F^{-1}(q)$. Esto es, si la nueva observación es menor que $F^{-1}(q)$, el nuevo cuantil estimado q será menor que el anterior al tener una influencia negativa y, por el contrario, si es mayor, el nuevo cuantil estimado será mayor pues la influencia es positiva.

2.2.5 LA MEDIA α -RECORTADA μ_α

Esta estadística está definida para valores de α entre 0 y 1/2. La funcional que la define es

$$T(F) = \frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x dF$$

o equivalentemente, si $t=F(x)$ y $dt=dF$, entonces

$$T(F) = \frac{1}{1-2\alpha} \int_{\alpha}^{1-\alpha} F^{-1}(t) dt$$

Consideremos esta última expresión en $\tilde{F} = (1-e)F + e\delta_x$.
Derivando con respecto a e y evaluando en $e=0$ se tiene

$$\begin{aligned} CI_{T,F}(x; \mu_\alpha) &= \frac{1}{1-2\alpha} \int_{\alpha}^{1-\alpha} CI_{T^*,F}(x) dt \\ &= \frac{1}{1-2\alpha} \int_{\alpha}^{1-\alpha} \frac{t - \delta_x(F^{-1}(t))}{f(F^{-1}(t))} dt \end{aligned}$$

donde $T^* = F^{-1}(t)$

Si $x < F^{-1}(\alpha)$ tenemos $S_x(F^{-1}(t)) \equiv 1$ y entonces

$$CI_{T,F}(x; \mu_\alpha) = \frac{1}{1-2\alpha} \int_\alpha^{1-\alpha} \frac{t-1}{f(F^{-1}(t))} dt$$

Sea $t=F(s)$ y $dt=f(s)ds$; por lo tanto

$$CI_{T,F}(x; \mu_\alpha) = \frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \frac{F(s)-1}{f(s)} f(s) ds$$

$$= \frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} (F(s)-1) ds$$

$$= \frac{1}{1-2\alpha} \left[\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} F(s) ds - F^{-1}(1-\alpha) + F^{-1}(\alpha) \right].$$

Integrando por partes y posteriormente haciendo $s=F^{-1}(t)$, $ds = (1/f(s))ds$, obtenemos

$$\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} F(s) ds = (1-\alpha)F^{-1}(1-\alpha) - \alpha F^{-1}(\alpha) - \int_\alpha^{1-\alpha} F^{-1}(t) dt,$$

y por lo tanto

$$CI_{T,F}(x; \mu_\alpha) = \frac{1}{1-2\alpha} \left[(1-\alpha)F^{-1}(1-\alpha) - \alpha F^{-1}(\alpha) - \int_\alpha^{1-\alpha} F^{-1}(t) dt - F^{-1}(1-\alpha) + F^{-1}(\alpha) \right]$$

$$= \frac{1}{1-2\alpha} \left[F^{-1}(\alpha) - \alpha F^{-1}(1-\alpha) - \alpha F^{-1}(\alpha) - \int_{\alpha}^{1-\alpha} F^{-1}(t) dt \right].$$

Por otro lado si $x > F^{-1}(\alpha)$ entonces $\delta_x(F^{-1}(t)) \equiv 0$ y por lo tanto

$$CI_{T,F}(x; \mu_{\alpha}) = \frac{1}{1-2\alpha} \int_{\alpha}^{1-\alpha} \frac{t}{f(F^{-1}(t))} dt.$$

Ahora bien, haciendo $t=F(s)$ y $dt=f(s)ds$, obtenemos

$$\begin{aligned} CI_{T,F}(x; \mu_{\alpha}) &= \frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} F(s) ds \\ &= \frac{1}{1-2\alpha} \left[F^{-1}(1-\alpha) - \alpha F^{-1}(1-\alpha) - \alpha F^{-1}(\alpha) - \int_{\alpha}^{1-\alpha} F^{-1}(t) dt \right] \end{aligned}$$

En el caso restante con $F^{-1}(\alpha) \leq x \leq F^{-1}(1-\alpha)$ tenemos que

$$\begin{aligned} CI_{T,F}(x; \mu_{\alpha}) &= \frac{1}{1-2\alpha} \left[\int_{\alpha}^{F(x)} \frac{t}{f(F^{-1}(t))} dt \right. \\ &\quad \left. + \int_{F(x)}^{1-\alpha} \frac{t-1}{f(F^{-1}(t))} dt \right] \end{aligned}$$

Utilizando de nuevo $t=F(s)$ y $dt=f(s)ds$, finalmente tendremos que

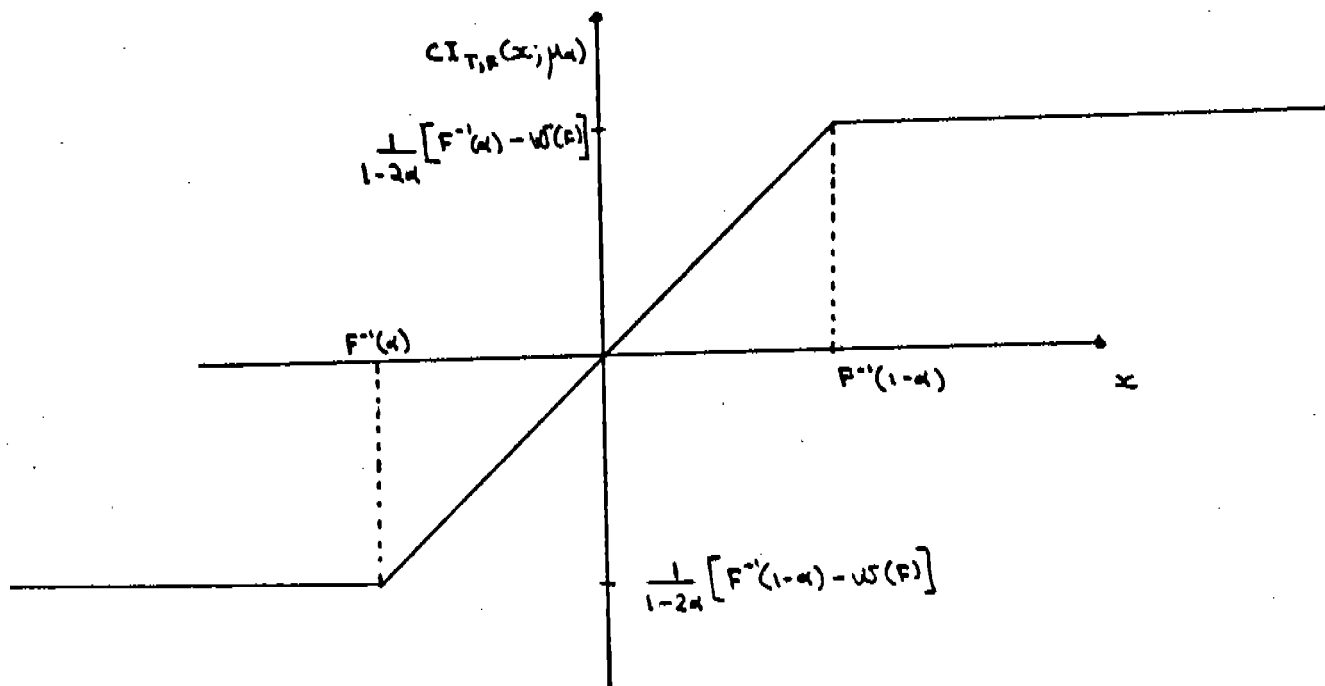
$$\begin{aligned}
 CI_{T,F}(x; \mu_\alpha) &= \frac{1}{1-2\alpha} \left[\int_{F^{-1}(\alpha)}^x F(s) ds + \int_x^{F^{-1}(1-\alpha)} (F(s)-1) ds \right] \\
 &= \frac{1}{1-2\alpha} \left[\int_{F^{-1}(\alpha)}^x F(s) ds + \int_x^{F^{-1}(1-\alpha)} F(s) ds - F^{-1}(1-\alpha) + x \right] \\
 &= \frac{1}{1-2\alpha} \left[x - F^{-1}(1-\alpha) + \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} F(s) ds \right] \\
 &= \frac{1}{1-2\alpha} \left[x - F^{-1}(1-\alpha) + F^{-1}(1-\alpha) - \alpha F^{-1}(1-\alpha) \right. \\
 &\quad \left. - \alpha F^{-1}(\alpha) - \int_\alpha^{1-\alpha} F^{-1}(t) dt \right] \\
 &= \frac{1}{1-2\alpha} \left[x - \alpha F^{-1}(1-\alpha) - \alpha F^{-1}(\alpha) \right. \\
 &\quad \left. - \int_\alpha^{1-\alpha} F^{-1}(t) dt \right].
 \end{aligned}$$

Resumiendo, se tiene que

$$CI_{T,F}(x; \mu_d) = \begin{cases} \frac{1}{1-2\alpha} [F^{-1}(\alpha) - W(F)] & \text{si } x < F^{-1}(\alpha) \\ \frac{1}{1-2\alpha} [x - W(F)] & \text{si } F^{-1}(\alpha) \leq x \leq F^{-1}(1-\alpha) \\ \frac{1}{1-2\alpha} [F^{-1}(1-\alpha) - W(F)] & \text{si } x > F^{-1}(1-\alpha) \end{cases}$$

donde $W(F) = \alpha F^{-1}(\alpha) + \alpha F^{-1}(1-\alpha) + \int_{\alpha}^{1-\alpha} F^{-1}(t) dt$.

La curva de influencia anterior para el caso de una distribución simétrica con media 0 se muestra en la siguiente figura.



Como es una función acotada y constante en las colas, la media α -recortada solo se ve influida por la posición de valores extremos y no por la magnitud de estos.

Hay que notar que si $\alpha \rightarrow 1/2$ se tiene la curva de influencia para la mediana.

2.2.6 ESTIMADORES M.

Un estimador M es cualquier estimador T definido a partir de una muestra aleatoria X_1, X_2, \dots, X_n , a través de un problema de minimización de la forma

$$\min \sum_i \rho(x_i; T_n),$$

o de una ecuación implícita

$$\sum_i \psi(x_i; T_n) = 0,$$

donde ρ es una función arbitraria y

$$\psi(x_i; T_n) = \frac{\partial}{\partial \theta} \rho(x; \theta).$$

Si $\rho(x, \theta) = -\ln(f(x, \theta))$ la solución del problema anterior produce el estimador máximo verosímil usual para el parámetro θ .

En este caso, la funcional $\theta = T(F)$ debe definirse de tal forma que

$$\int \Psi(s; T(F)) dF = 0$$

o que

$$\int \Psi(s; T(F)) f(s) ds = 0.$$

Evaluando de nuevo esta expresión en $\tilde{F} = (1 - \epsilon)F + \epsilon \delta_x$ y derivando con respecto a ϵ se obtiene

$$\int \left[\Psi(s; T(\tilde{F})) d(-F + \delta_x) + \left(\frac{\partial}{\partial \theta} \Psi(s; T(\tilde{F})) \right) \dot{T}(\tilde{F}) d\tilde{F} \right] = 0$$

$$- \int \Psi(s; T(\tilde{F})) dF + \int \Psi(s; T(F)) d\delta_x + \int \dot{T}(\tilde{F}) \frac{\partial}{\partial \theta} \Psi(s; T(\tilde{F})) d\tilde{F} = 0,$$

y evaluando en $\epsilon = 0$

$$- \int \Psi(s; T(F)) dF + \Psi(x; T(F)) + \int CI_{T,F}(x) \frac{\partial}{\partial \theta} \Psi(s; T(F)) dF = 0$$

lo que equivale a

$$\Psi(x; T(F)) + CI_{T,F}(x) \int \frac{\partial}{\partial \theta} \Psi(s, T(F)) dF = 0,$$

por lo que

$$CI_{T,F}(x) = \frac{-\Psi(x, T(F))}{\int \frac{\partial}{\partial \theta} \Psi(s, T(F)) dF}.$$

Hay que notar que la curva de influencia de un estimador M es proporcional a ψ .

2.2.7 ESTIMADORES L

Un estimador que es una combinación lineal de estadísticas de orden α , mas generalmente, de alguna función h de ellas, es llamado un estimador L , es decir, su forma será

$$T_n = \sum_{i=1}^n a_{ni} h(x_{(i)})$$

con

$$\sum_{i=1}^n a_{ni} = 1.$$

Por lo cual, la funcional $\theta = T(F)$ se define de la siguiente manera

$$T(F) = \int_0^1 h(F^{-1}(s)) dM(s),$$

donde $M(s)$ es una medida sobre $(0,1)$.

Sea

$$\tilde{F} = (1-\epsilon)F + \epsilon \delta_x,$$

$$T(\tilde{F}) = \int_0^1 h(\tilde{F}^{-1}(s)) dM(s)$$

y

$$\frac{\partial}{\partial \epsilon} T(\tilde{F}) = \int_0^1 h'(\tilde{F}^{-1}(s)) \frac{\partial}{\partial \epsilon} \tilde{F}^{-1}(s) dM(s).$$

Por lo tanto

$$CI_{T,F}(x) = \frac{\partial}{\partial \epsilon} T(\tilde{x}) \Big|_{\epsilon=0}$$

$$= \int_0^1 h'(F^{-1}(s)) CI_{T^*,F}(s) dM(s),$$

donde $T^*(F) = F^{-1}(s)$.

Entonces

$$CI_{T,F}(x) = \int_0^1 h'(F^{-1}(s)) \left[\frac{s - S_x(F^{-1}(s))}{f(F^{-1}(s))} \right] dM(s)$$

$$= \int_0^1 \frac{s h'(F^{-1}(s))}{f(F^{-1}(s))} dM(s)$$

$$- \int_0^1 \frac{h'(F^{-1}(s)) S_x(F^{-1}(s))}{f(F^{-1}(s))} dM(s)$$

y finalmente,

$$CI_{T,F}(x) = \int_0^1 \frac{s h'(F^{-1}(s))}{f(F^{-1}(s))} dM(s)$$

$$- \int_{F(x)}^1 \frac{h'(F^{-1}(s))}{f(F^{-1}(s))} dM(s).$$

2.2.8 ESTIMADORES R

La estimación R es un método no paramétrico que resulta de asignar rangos cuando la muestra fue obtenida de una distribución de tipo continuo.

La estimación R puede definirse como sigue:

Sean X_1, X_2, \dots, X_n y Y_1, Y_2, \dots, Y_n dos muestras aleatorias independientes de las distribuciones $F(x)$ y $G(x)$ respectivamente. Se unen las dos muestras en una sola de tamaño $2n$ y se definen los rangos R_i de las observaciones X_i 's en la muestra combinada y se considera además $a = a(i)$ alguna función que determina puntajes ("scores") de interés.

La estadística en estudio es

$$S_n = \frac{1}{n} \sum_{i=1}^n a(R_i)$$

o bien

$$S_n = \frac{1}{n} \sum_{i=1}^n a\left(\frac{R_i}{2n+1}\right).$$

Usualmente los puntajes son generados por alguna función J casi simétrica en el sentido que $J(1-t) = -J(t)$ para $0 \leq t \leq 1$.

Por lo tanto

$$S_n = \frac{1}{n} \sum_{i=1}^n J\left(\frac{R_i}{2n+1}\right),$$

que en términos de funcionales puede escribirse como

$$S(F, G) = \int J\left(\frac{1}{2}F(s) + \frac{1}{2}G(s)\right) dF.$$

Supongamos ahora que la muestra Y_1, \dots, Y_n se genera a partir de la muestra X_1, \dots, X_n mediante la relación $Y_i = 2\theta - X_i$, en donde θ es un valor desconocido, pero fijo, tal que cumple con $S(F, G) = 0$. La función de distribución G será entonces función de F , puesto que

$$\begin{aligned} G(y) &= P[Y \leq y] = P[2\theta - X \leq y] \\ &= P[X \geq 2\theta - y] = 1 - P[X \leq 2\theta - y] \\ &= 1 - F(2\theta - y), \end{aligned}$$

por lo cual

$$\begin{aligned} S(F, G) &= \int J\left(\frac{1}{2}F(s) + \frac{1}{2}(1 - F(2\theta - s))\right) dF \\ &= \int J\left(\frac{1}{2}(F(s) + 1 - F(2\theta - s))\right) dF = 0. \end{aligned}$$

Sea $K(s) = \frac{1}{2}(F(s) + 1 - F(2\theta - s))$, entonces

$$\int J(K(s)) dF = \int J(K(F^{-1}(t))) dt = 0,$$

donde $t = F(s)$ y $dt = dF$

Para calcular la curva de influencia consideremos $\theta = T(F)$ en la ecuación anterior y $\tilde{F} = (1 - \epsilon)F + \epsilon \delta_x$; es decir,

$$\int J(\tilde{K}(\tilde{F}^{-1}(t))) dt = 0.$$

Derivando con respecto a ϵ tenemos

$$\int J'(\tilde{K}(\tilde{F}^{-1}(t))) \frac{d}{d\epsilon} \tilde{K}(\tilde{F}^{-1}(t)) dt = 0$$

pero

$$\frac{d}{d\epsilon} \tilde{K}(\tilde{F}^{-1}(t)) = \frac{d}{d\epsilon} \left[\frac{1}{2} (t+1 - \tilde{F}(2T(\tilde{F}) - \tilde{F}^{-1}(t))) \right]$$

$$= -\frac{1}{2} \frac{d}{d\epsilon} \left[(1-\epsilon) F(2T(\tilde{F}) - \tilde{F}^{-1}(t)) \right. \\ \left. + \epsilon \delta_x(2T(\tilde{F}) - \tilde{F}^{-1}(t)) \right]$$

$$= -\frac{1}{2} \left[(1-\epsilon) f(2T(\tilde{F}) - \tilde{F}^{-1}(t)) \left(2\dot{T}(\tilde{F}) - \frac{d}{d\epsilon} \tilde{F}^{-1}(t) \right) \right. \\ \left. - F(2T(\tilde{F}) - \tilde{F}^{-1}(t)) + \delta_x(2T(\tilde{F}) - \tilde{F}^{-1}(t)) \right. \\ \left. + \epsilon \frac{d}{d\epsilon} \delta_x(2T(\tilde{F}) - \tilde{F}^{-1}(t)) \right].$$

Por lo tanto

$$\frac{d}{d\epsilon} \tilde{K}(\tilde{F}^{-1}(t)) \Big|_{\epsilon=0} =$$

$$= -\frac{1}{2} \left[f(2T(F) - F^{-1}(t)) (2CI_{T,F}(x) - \frac{t - \delta_x(F^{-1}(t))}{f(F^{-1}(t))}) \right]$$

$$- F(2T(F) - F^{-1}(t)) + \delta_x(2T(F) - F^{-1}(t)) \Big].$$

y sustituyendo en

$$\int J'(K(F^{-1}(t))) \frac{d}{d\varepsilon} \tilde{K}(\tilde{F}^{-1}(t)) \Big|_{\varepsilon=0} dt = 0,$$

tenemos que

$$\begin{aligned} & \int J'(K(F^{-1}(t))) \left[\delta_x(2T(F) - F^{-1}(t)) - F(2T(F) - F^{-1}(t)) \right. \\ & \quad + \frac{f(2T(F) - F^{-1}(t))}{f(F^{-1}(t))} (\delta_x(F^{-1}(t)) - t) \\ & \quad \left. + 2CI_{T,F}(x) f(2T(F) - F^{-1}(t)) \right] dt \\ & = \int J'(K(F^{-1}(t))) (\delta_x(2T(F) - F^{-1}(t)) - F(2T(F) - F^{-1}(t))) dt \\ & \quad + \int J'(K(F^{-1}(t))) \frac{f(2T(F) - F^{-1}(t))}{f(F^{-1}(t))} (\delta_x(F^{-1}(t)) - t) dt \\ & \quad + 2 \int CI_{T,F}(x) J'(K(F^{-1}(t))) f(2T(F) - F^{-1}(t)) dt = 0 \end{aligned}$$

Mediante el cambio de variable

$$y = 2T(F) - F^{-1}(t)$$

$$t = F(2T - y)$$

$$dt = -f(2T - y) dy$$

y utilizando el hecho de que

$$K(2T(F) - y) = \frac{1}{2} \left[F(2T(F) - y) + 1 - F(2T(F) - 2T(F) + y) \right]$$

$$\begin{aligned}
&= \frac{1}{2} [F(2T(F) - y) + 1 - F(y)] \\
&= 1 - \frac{1}{2} [F(y) + 1 - F(2T(F) - y)] \\
&= 1 - K(y)
\end{aligned}$$

podemos reescribir la primera de las tres integrales como

$$- \int J'(1 - K(y)) (\delta_x(y) - F(y)) f(2T - y) dy.$$

Por otro lado, la segunda integral puede reducirse utilizando

$$y = F^{-1}(t), \quad t = F(y) \quad y \quad dt = f(y) dy$$

para dar

$$\int J'(K(y)) (\delta_x(y) - F(y)) f(2T(F) - y) dy.$$

Por lo tanto tenemos la ecuación

$$\begin{aligned}
&- \int J'(1 - K(y)) (\delta_x(y) - F(y)) f(2T(F) - y) dy \\
&+ \int J'(K(y)) (\delta_x(y) - F(y)) f(2T(F) - y) dy \\
&+ 2CI_{T,F}(x) \int J'(K(y)) f(2T(F) - y) f(y) dy = 0
\end{aligned}$$

es decir,

$$CI_{T,F}(x) = \frac{\int [J'(1 - K(y)) - J'(K(y))] f(2T(F) - y) (\delta_x(y) - F(y)) dy}{2 \int J'(K(y)) f(2T(F) - y) f(y) dy}$$

2.2.9 COEFICIENTE DE CORRELACION

Consideremos una distribución bivariada con segundos momentos finitos, es decir,

Sea $\begin{pmatrix} X \\ Y \end{pmatrix}$ un vector aleatorio tal que

$$E \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \text{Var} \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

La funcional que determina el coeficiente de correlación ρ es

$$T(F) = \frac{\iint (s - \mu_1)(t - \mu_2) dF}{\sqrt{\iint (s - \mu_1)^2 dF} \sqrt{\iint (t - \mu_2)^2 dF}}$$

Sea $\tilde{F} = (1 - \epsilon)F + \epsilon\delta_{\tilde{x}}$ donde $\tilde{x} = \begin{pmatrix} X \\ Y \end{pmatrix}$, entonces

$$\frac{d}{d\epsilon} T(F) = \frac{A}{\left(\iint (s - \mu_1)^2 dF\right) \left(\iint (t - \mu_2)^2 dF\right)}$$

con

$$A = \sqrt{\iint (s - \mu_1)^2 d\tilde{F}} \sqrt{\iint (t - \mu_2)^2 d\tilde{F}} \frac{d}{d\epsilon} \iint (s - \mu_1)(t - \mu_2) d\tilde{F} \\ - \left(\iint (s - \mu_1)(t - \mu_2) dF\right) \frac{d}{d\epsilon} \sqrt{\iint (s - \mu_1)^2 d\tilde{F}} \sqrt{\iint (t - \mu_2)^2 d\tilde{F}}.$$

Nótese que

$$\frac{d}{d\epsilon} \iint (s - \mu_1)(t - \mu_2) d\tilde{F} = -\iint (s - \mu_1)(t - \mu_2) dF + \iint (s - \mu_1)(t - \mu_2) d\delta_{\tilde{x}} \\ = -\rho\sigma_1\sigma_2 + (x - \mu_1)(y - \mu_2).$$

También tenemos que

$$\begin{aligned} \frac{d}{d\epsilon} \sqrt{\iint (s-\mu_1)^2 d\tilde{F}} \sqrt{\iint (t-\mu_2)^2 d\tilde{F}} &= \\ &= \frac{\sqrt{\iint (s-\mu_1)^2 d\tilde{F}}}{2 \sqrt{\iint (t-\mu_2)^2 d\tilde{F}}} \left[-\iint (t-\mu_2)^2 dF + \iint (t-\mu_2)^2 dS_{\tilde{F}} \right] \\ &+ \frac{\sqrt{\iint (t-\mu_2)^2 d\tilde{F}}}{2 \sqrt{\iint (s-\mu_1)^2 d\tilde{F}}} \left[-\iint (s-\mu_1)^2 dF + \iint (s-\mu_1)^2 dS_{\tilde{F}} \right] \end{aligned}$$

lo cual evaluado en $\epsilon=0$ da

$$\begin{aligned} \frac{d}{d\epsilon} \sqrt{\iint (s-\mu_1)^2 d\tilde{F}} \sqrt{\iint (t-\mu_2)^2 d\tilde{F}} \Big|_{\epsilon=0} &= \\ &= \frac{\sqrt{\iint (s-\mu_1)^2 dF}}{2 \sqrt{\iint (t-\mu_2)^2 dF}} (-\sigma_2^2 + (y-\mu_2)^2) \\ &+ \frac{\sqrt{\iint (t-\mu_2)^2 dF}}{2 \sqrt{\iint (s-\mu_1)^2 dF}} (-\sigma_1^2 + (x-\mu_1)^2) \\ &= \frac{\sigma_1}{2\sigma_2} (-\sigma_2^2 + (y-\mu_2)^2) + \frac{\sigma_2}{2\sigma_1} (-\sigma_1^2 + (x-\mu_1)^2) \end{aligned}$$

Por lo tanto,

$$\begin{aligned} CI_{T,F}(x,y;\rho) &= \frac{d}{d\epsilon} T(\tilde{F}) \Big|_{\epsilon=0} \\ &= \frac{A'}{\left(\iint (s-\mu_1)^2 dF \right) \left(\iint (t-\mu_2)^2 dF \right)} \end{aligned}$$

donde ahora

$$A' = \sqrt{\iint (s - \mu_1)^2 dF} \sqrt{\iint (t - \mu_2)^2 dF} (-\rho \sigma_1 \sigma_2 + (x - \mu_1)(y - \mu_2)) \\ - \left(\iint (s - \mu_1)(t - \mu_2) dF \right) \left(\frac{\sigma_1}{2\sigma_2} (-\sigma_2^2 + (y - \mu_2)^2) + \frac{\sigma_2}{2\sigma_1} (-\sigma_1^2 + (x - \mu_1)^2) \right).$$

Entonces haciendo $x^* = \frac{x - \mu_1}{\sigma_1}$ y $y^* = \frac{y - \mu_2}{\sigma_2}$ tenemos

$$CI_{T,F}(x, y; \rho) = \frac{1}{\sigma_1^2 \sigma_2^2} \left[\sigma_1 \sigma_2 (-\rho \sigma_1 \sigma_2 + (x - \mu_1)(y - \mu_2)) \right. \\ \left. - \rho \sigma_1 \sigma_2 \left(-\frac{\sigma_1 \sigma_2}{2} + \frac{\sigma_1}{2\sigma_2} (y - \mu_2)^2 - \frac{\sigma_1 \sigma_2}{2} + \frac{\sigma_2}{2\sigma_1} (x - \mu_1)^2 \right) \right] \\ = \frac{1}{2} [2x^*y^* - \rho x^{*2} - \rho y^{*2}] \\ = \frac{1}{4} [(1 - \rho)(x^* + y^*)^2 - (1 + \rho)(x^* - y^*)^2] \\ = (1 - \rho^2) \left[\frac{1}{2} \left(\frac{x^* + y^*}{\sqrt{1 + \rho}} + \frac{x^* - y^*}{\sqrt{1 - \rho}} \right) \right] \left[\frac{1}{2} \left(\frac{x^* + y^*}{\sqrt{1 + \rho}} - \frac{x^* - y^*}{\sqrt{1 - \rho}} \right) \right].$$

Sean $z_1 = \frac{x^* + y^*}{\sqrt{1 + \rho}}$ y $z_2 = \frac{x^* - y^*}{\sqrt{1 - \rho}}$

es decir, la suma y diferencia estandarizadas entre x^* y y^* - respectivamente, entonces podemos reescribir

$$CI_{T,F}(x, y; \rho) = (1 - \rho^2) \left[\frac{1}{2} (z_1 + z_2) \right] \left[\frac{1}{2} (z_1 - z_2) \right]$$

O bien, haciendo

$$u_1 = \frac{1}{2} (z_1 + z_2) \quad y \quad u_2 = \frac{1}{2} (z_1 - z_2)$$

tenemos

$$CI_{T,F}(x, y; \rho) = (1 - \rho^2) u_1 u_2.$$

2.3 RELACION ENTRE LA FUNCION DE INFLUENCIA Y OTRAS FUNCIONES SIMILARES

Dada una muestra X_1, X_2, \dots, X_n de una cierta población y un estimador del parámetro poblacional $\theta = T(X_1, X_2, \dots, X_n) \equiv T_n(X)$ se han construido funciones de la misma muestra tales que nos permitan tener una idea del comportamiento de θ , es decir, funciones cuyo objetivo es esencialmente el mismo que tiene la función de influencia. Veamos ahora algunas de estas funciones

2.3.1 CURVA DE SENSIBILIDAD

Supongamos que a la muestra inicial se le ha añadido una observación con valor x con lo cual se tendrá una muestra de tamaño $n+1$. Sea $\theta_{n+1} = T_{n+1}(x)$ el estimador del parámetro calculado con la muestra conjunta. Hay que notar que θ_{n+1} es una función del valor x . La curva de sensibilidad del estimador se define como:

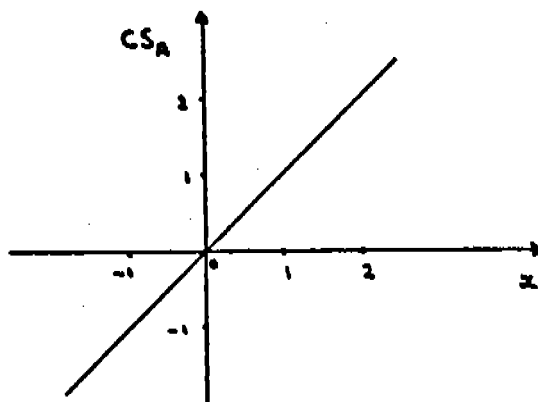
$$\begin{aligned} CS(x) &= T_{n+1}(x) - T_n(x) \\ &= \theta_{n+1} - \theta_n \end{aligned}$$

Esta función fue propuesta por Tukey(1970) para estudiar el comportamiento de los estimadores en muestras finitas. Una pequeña modificación es usada por Andrews, et.al.(1972) para estudiar el comportamiento de aproximadamente 68 estimadores de localización. Dicha modificación es

$$CS_A = n(T_{n+1}(x) - T_n(x)).$$

Algunos resultados que han sido obtenidos a partir de la curva de sensibilidad son:¹

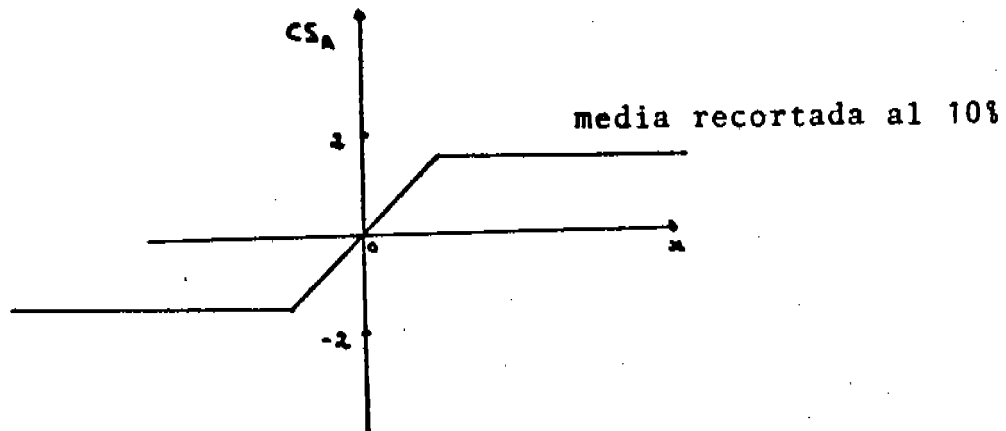
a) La media, que es un estimador no robusto, es muy sensible a observaciones extremas ya que la curva de sensibilidad no es acotada en las colas. Esto se muestra en la siguiente figura.



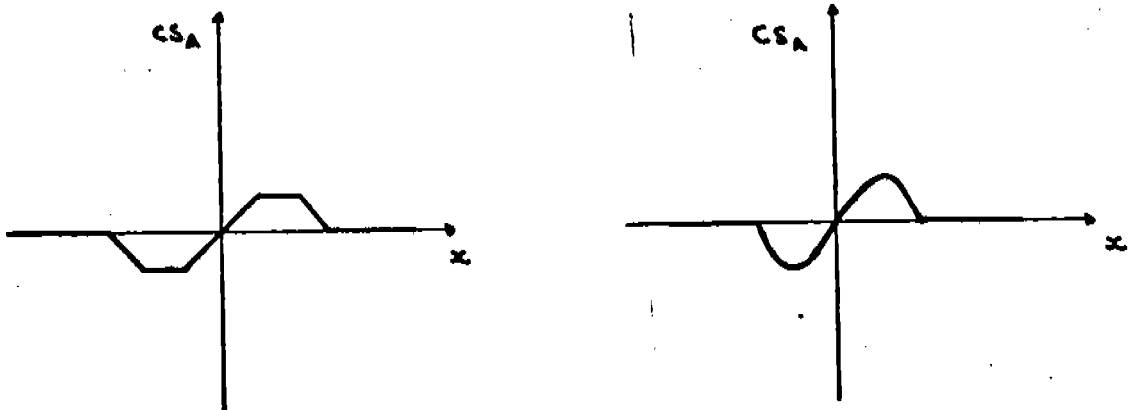
b) Otros estimadores (la media recortada al 10%, la mediana, etc.) tienen curvas de sensibilidad acotadas en los extremos. Esto se puede observar en la siguiente gráfica de la curva de sensibilidad para la media recortada al 10% .

¹

Las gráficas fueron construidas a partir de una muestra de tamaño 19 con media 0 de una población normal y añadiendo una observación con valor x



c) Estimadores de tipo Hampel son tales que $CS_A(x) = 0$ para valores grandes de $|x|$. Esto provoca que se ignora por completo la información de observaciones muy discrepantes. En la siguiente figura se muestran dos formas muy utilizadas.



Si además consideramos la curva de sensibilidad como una aproximación a la curva de influencia (de hecho Hogg afirma que la curva de influencia es una curva de sensibilidad asintótica dividida entre $1/n$), la varianza asintótica del estimador es aproximadamente

$$\int CS_A^2(x) dF(x) .$$

La aproximación CS discutida arriba puede justificarse por el hecho de que si consideramos la expresión que se utiliza en la construcción de la curva de influencia, es decir,

$$\frac{T(\tilde{F}) - T(F)}{\epsilon}$$

y sustituimos F por F_{n-1} , la función de distribución empírica para una muestra de tamaño $n-1$ (la cual a su vez es un buen estimador de F) y sustituimos ϵ por $1/n$, obtendremos

$$\frac{T\left(\left(1 - \frac{1}{n}\right)F_{n-1} + \frac{1}{n}\delta_x\right) - T(F_{n-1})}{(1/n)}$$

$$= n \left[T_n(x_1, x_2, \dots, x_{n-1}, x) - T_{n-1}(x_1, x_2, \dots, x_{n-1}) \right] .$$

2.3.2 EL JACKKNIFE

Sean X_1, X_2, \dots, X_n una muestra aleatoria y $\hat{\theta} = T(x_1, x_2, \dots, x_n)$ un estimador del parámetro θ basado en la muestra de tamaño n . Consideremos, además, una partición de las observaciones en g grupos de tamaño h ($n=gh$). Sea $\hat{\theta}_{-i}$ el correspondiente estimador basado en la muestra de tamaño $(g-1)h$ donde el i -ésimo grupo de tamaño h ha sido eliminado.

El i -ésimo pseudovalor del jackknife se define como:

$$\tilde{\theta}_i = g\hat{\theta} - (g-1)\hat{\theta}_{-i} \quad i=1,2,\dots,g.$$

El jackknife es entonces el promedio de los pseudovalores, es decir

$$\tilde{\theta} = \frac{1}{g} \sum_{i=1}^g \tilde{\theta}_i = g\hat{\theta} - \frac{g-1}{g} \sum_{i=1}^g \hat{\theta}_{-i}.$$

Sin embargo, la forma más usual del jackknife es considerar $h=1$ y por lo tanto $g=n$, con lo cual el i -ésimo pseudovalor será

$$\begin{aligned} \tilde{\theta}_i &= n\hat{\theta} - (n-1)\hat{\theta}_{-i} \\ &= nT_n(x_1, \dots, x_n) - (n-1)T_{n-1}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n), \end{aligned}$$

el cual es un indicador del peso que tiene la i -ésima observación en el valor del estimador.

Consideremos de nuevo la expresión

$$\frac{T(\tilde{F}) - T(F)}{e};$$

si sustituimos F por F , e por $-1/(n-1)$ y evaluamos la expresión en cada x_i , $i=1,2,\dots,n$, obtendremos

$$\begin{aligned} & \frac{T\left(\left(1 + \frac{1}{n-1}\right)F_n - \frac{1}{n-1} \delta_{x_i}\right) - T(F_n)}{\left(-\frac{1}{n-1}\right)} \\ &= (n-1) \left(T(F_n) - T\left(\frac{n}{n-1} F_n - \frac{1}{n-1} \delta_{x_i}\right) \right) \\ &= (n-1) (\hat{\theta} - \hat{\theta}_{-i}) = n\hat{\theta} - \hat{\theta} - (n-1)\hat{\theta}_{-i} \\ &= \tilde{\theta}_i - \hat{\theta}. \end{aligned}$$

Por lo tanto, las diferencias $\tilde{\theta}_i - \hat{\theta}$, $i=1,2,\dots,n$, forman una colección que es casi la versión finita de la curva de influencia; de aquí la varianza del estimador es aproximadamente

$$\sum_i \frac{(\tilde{\theta}_i - \hat{\theta})^2}{n(n-1)},$$

pues esta expresión es una aproximación de

$$\int (CI_{T,F}(x))^2 dF.$$

Consideremos ahora al estimador $\hat{\theta}$ como una funcional $T(X_1, X_2, \dots, X_n; w_1, w_2, \dots, w_n)$ que depende de los valores muestrales y de ponderaciones arbitrarias w_1, \dots, w_n de tales valores.

Así, para el jackknife ordinario se tiene que

$$\hat{\theta}_{-i} = T(X_1, \dots, X_n; 1/(n-1), \dots, 0, \dots, 1/(n-1))$$

i-ésimo lugar

Esta consideración es de interés para poder construir el jackknife infinitesimal. Sean $\epsilon < 1/n$ y

$$\hat{\theta}_{-i}(\epsilon) = T(X_1, \dots, X_n; \frac{1-\epsilon}{n}, \dots, \frac{1}{n} - \epsilon, \dots, \frac{1-\epsilon}{n}).$$

El jackknife infinitesimal es tal que las diferencias $\hat{\theta}_{-i} - \hat{\theta}$, dadas en el jackknife ordinario, son sustituidas por las derivadas con respecto a ϵ ; es decir,

$$d_i = \lim_{\epsilon \rightarrow 0^+} \frac{T(X_1, \dots, X_n; \frac{1-\epsilon}{n}, \dots, \frac{1}{n} - \epsilon, \dots, \frac{1-\epsilon}{n}) - T(X_1, \dots, X_n; \frac{1}{n}, \dots, \frac{1}{n})}{\epsilon}$$

$i = 1, 2, \dots, n.$

Con estos valores es posible aproximar la varianza del estimador por

$$\sum_i \frac{d_i^2}{n(n-1)},$$

pues se vuelve a tener con las d_i 's una versión finita de la curva de influencia.

2.4 LA FUNCION DE INFLUENCIA MUESTRAL

Como hemos visto, la curva de influencia teórica supone el conocimiento de la verdadera función de distribución de la

población en estudio. Sin embargo, aunque en general no es posible conocer dicha función de distribución, nos gustaría tener de alguna forma una estimación o aproximación a la curva de influencia.

Se han desarrollado algunas versiones para muestras finitas de la curva de influencia las cuales son de gran utilidad para analizar el comportamiento de estimadores. A continuación consideraremos tres de las versiones mas utilizadas.

2.4.1 PRIMERA VERSION

En este caso lo que se propone es aproximar la curva de influencia por medio de la diferencia entre $\hat{\theta}$, un estimador de θ basado en una muestra de tamaño n , y $\hat{\theta}^+$, un estimador de la misma forma que $\hat{\theta}$ que se obtiene de las n observaciones originales mas una observación adicional hipotética con valor x . Dada la construcción anterior, esta versión es llamada la curva de influencia predictiva.

Entonces la curva de influencia estimada $CI(x, \theta)$ está dada por

$$CI(x, \theta) = (n+1)(\hat{\theta}^+ - \hat{\theta}).$$

Como puede verse, esta expresión es prácticamente la misma que la de la curva de sensibilidad de Andrews, et. al. (1972) para la estimación del parámetro de localización.

2.4.1.1 EJEMPLOS

- LA MEDIA MUESTRAL.

Dada la muestra x_1, x_2, \dots, x_n el estimador de la media es

$$\hat{\theta} = \bar{x}_n = \frac{\sum_{i=1}^n x_i}{n},$$

y para la muestra x_1, x_2, \dots, x_n, x el estimador será

$$\hat{\theta}_+ = \bar{x}_{n+1} = \frac{\sum_{i=1}^n x_i + x}{n+1} = \frac{n\bar{x}_n + x}{n+1}.$$

Entonces

$$\begin{aligned} CI_+(x; \mu) &= (n+1) \left[\frac{n\bar{x}_n + x}{n+1} - \bar{x}_n \right] \\ &= (n+1) \left[\frac{n\bar{x}_n + x - n\bar{x}_n - \bar{x}_n}{n+1} \right] = x - \bar{x}_n. \end{aligned}$$

Por lo tanto, la influencia de una observación x sobre la media se incrementa linealmente con relación a la diferencia entre x y la media original.

Hay que notar la gran similitud entre esta expresión y la expresión de la verdadera curva de influencia encontrada anteriormente, es decir,

$$CI_{T,F}(x; \mu) = x - \mu.$$

- LA VARIANZA MUESTRAL.

En este caso

$$\hat{\theta} = S_n^2 = \sum_{i=1}^n \frac{(x_i - \bar{x}_n)^2}{n-1} = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}_n^2}{n-1}$$

y

$$\hat{\theta}_+ = S_{n+1}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x}_{n+1})^2 + (x - \bar{x}_{n+1})^2}{n}$$

y utilizando la igualdad

$$\begin{aligned} S_{n+1}^2 - S_n^2 &= \frac{n-1}{n} S_n^2 + \frac{1}{n+1} (x - \bar{x}_n)^2 - S_n^2 \\ &= -\frac{S_n^2}{n} + \frac{1}{n+1} (x - \bar{x}_n)^2 \end{aligned}$$

tenemos que

$$CI_+(x; \sigma^2) = (x - \bar{x}_n)^2 - \frac{n+1}{n} S_n^2.$$

Para n grande, si $|x - \bar{x}_n| < S_n$, la nueva estimación de la varianza, S_{n+1} , será menor que S_n , mientras que si $|x - \bar{x}_n| \rightarrow \infty$, la nueva estimación de la varianza se incrementará en forma cuadrática.

Otra vez se logra una forma muy similar a la curva de influencia teórica, por lo que las conclusiones obtenidas para esta última son válidas para la expresión muestral. La curva de influencia teórica en este caso es

$$CI_{T,F}(x; \sigma^2) = (x - \mu)^2 - \sigma^2.$$

- LA MEDIA α -RECORTADA

Para una muestra de tamaño n , la media α -recortada se obtiene quitando las h observaciones más pequeñas y las h observaciones más grandes y calculando la media aritmética de las restantes.

Como no necesariamente h es un entero, se suele utilizar el valor $h = \lceil \alpha n \rceil$ y entonces

$$\bar{x}_{\alpha, n} = \frac{1}{n-2h} \sum_{i=h+1}^{n-h} x_{(i)}$$

donde las $x_{(i)}$'s son los valores ordenados de las observaciones.

Si $h = \lceil \alpha n \rceil = \lceil \alpha(n+1) \rceil$ tenemos

$$\bar{x}_{\alpha, n+1} = \begin{cases} \frac{1}{n-2h+1} \sum_{i=h}^{n-h} x_{(i)} & \text{si } x \leq x_{(h)} \\ \frac{1}{n-2h+1} \sum_{i=h+1}^{n-h} \frac{x_{(i)} + x}{2} & \text{si } x_{(h)} < x < x_{(n-h+1)} \\ \frac{1}{n-2h+1} \sum_{i=h+1}^{n-h+1} x_{(i)} & \text{si } x \geq x_{(n-h+1)} \end{cases}$$

²

$\lceil v \rceil$ denota el mayor entero menor o igual que v

$$= \begin{cases} \frac{(n-2h)\bar{x}_{d,n} + x_{(h)}}{n-2h+1} & \text{si } x \leq x_{(h)} \\ \frac{(n-2h)\bar{x}_{d,n} + x}{n-2h+1} & \text{si } x_{(h)} < x < x_{(n-h+1)} \\ \frac{(n-2h)\bar{x}_{d,n} + x_{(n-h+1)}}{n-2h+1} & \text{si } x \geq x_{(n-h+1)} \end{cases}$$

por lo que

$$\bar{x}_{d,n+1} - \bar{x}_{d,n} = \begin{cases} \frac{x_{(h)} - \bar{x}_{d,n}}{n-2h+1} & \text{si } x \leq x_{(h)} \\ \frac{x - \bar{x}_{d,n}}{n-2h+1} & \text{si } x_{(h)} < x < x_{(n-h+1)} \\ \frac{x_{(n-h+1)} - \bar{x}_{d,n}}{n-2h+1} & \text{si } x \geq x_{(n-h+1)} \end{cases}$$

y entonces

$$C I_+(x; \mu_d) = \begin{cases} \frac{n+1}{n-2h+1} (x_{(h)} - \bar{x}_{d,n}) & \text{si } x \leq x_{(h)} \\ \frac{n+1}{n-2h+1} (x - \bar{x}_{d,n}) & \text{si } x_{(h)} < x < x_{(n-h+1)} \\ \frac{n+1}{n-2h+1} (x_{(n-h+1)} - \bar{x}_{d,n}) & \text{si } x \geq x_{(n-h+1)} \end{cases}$$

Si $[d_n] = [d_{(n+1)}] - 1$ tendremos

$$\bar{x}_{d,n+1} = \begin{cases} \sum_{h+1}^{n-h-1} \frac{x(i)}{n-2h-1} & \text{si } x \leq x_{(h+1)} \\ \sum_{h+2}^{n-h-1} \frac{x(i) + x}{n-2h-1} & \text{si } x_{(h+1)} < x < x_{(n-h-1)} \\ \sum_{h+2}^{n-h} \frac{x(i)}{n-2h-1} & \text{si } x \geq x_{(n-h-1)} \end{cases}$$

de donde

$$\bar{x}_{d,n+1} - \bar{x}_{d,n} = \begin{cases} \frac{\bar{x}_{d,n} - x_{(n-h-1)}}{n-2h-1} & \text{si } x \leq x_{(h+1)} \\ \frac{\bar{x}_{d,n} - x_{(h+1)} - x_{(n-h-1)} + x}{n-2h-1} & \text{si } x_{(h+1)} < x < x_{(n-h-1)} \\ \frac{\bar{x}_{d,n} - x_{(h+1)}}{n-2h-1} & \text{si } x \geq x_{(n-h-1)}, \end{cases}$$

y por lo tanto

$$C I_0(x; \mu_0) = \begin{cases} \frac{n+1}{n-2h-1} (\bar{x}_{d,n} - x_{(n-h-1)}) & \text{si } x \leq x_{(h+1)} \\ \frac{n+1}{n-2h-1} [\bar{x}_{d,n} - x_{(h+1)} - x_{(n-h-1)} + x] & \text{si } x_{(h+1)} < x < x_{(n-h-1)} \\ \frac{n+1}{n-2h-1} (\bar{x}_{d,n} - x_{(h+1)}) & \text{si } x \geq x_{(n-h-1)}. \end{cases}$$

En cualquiera de los 2 casos se muestra que la menor influencia se tiene cuando la observación adicional está en la parte central de la muestra. El cambio que sufre la media es constante pero no cero en observaciones en las colas de la muestra. En particular, la influencia de una observación discrepante en la media α -recortada no es cero como uno esperaría, pues las observaciones discrepantes se desechan, por lo que tienen la mayor influencia posible en cada uno de los lados.

2.4.2 SEGUNDA VERSION

Esta versión está enfocada a medir los efectos individuales de observaciones de la muestra, es decir, es particularmente útil para medir la influencia de las observaciones sobre el estimador $\hat{\theta}$. De aquí que esta versión sea llamada la curva de influencia muestral y está definida como

$$CI_{-i}(x_i; \theta) = (n-1)(\hat{\theta} - \hat{\theta}_{-i}) \quad (i=1, 2, \dots, n),$$

donde $\hat{\theta}_{-i}$ es un estimador de la misma forma que $\hat{\theta}$ pero calculado sin la i -ésima observación x_i de la muestra. Esta versión está muy relacionada con la técnica del jackknife de Tukey pues

$$CI(x_i; \theta) + \hat{\theta} = (n-1)(\hat{\theta} - \hat{\theta}_{-i}) + \hat{\theta} = n\hat{\theta} - (n-1)\hat{\theta}_{-i}$$

el cual es precisamente el i -ésimo pseudovalor (ver sección 2.3.2).

Dada la naturaleza de la curva de influencia muestral, esta tiende a ser mas utilizada que la curva CI⁺ pues no siempre es posible o es de interés obtener observaciones futuras.

2.4.2.1 EJEMPLOS

Los ejemplos que se presentan a continuación se refieren a aplicaciones de la curva de influencia muestral que son utilizadas en diferentes áreas estadísticas.

- DETECCION DE OBSERVACIONES INFLUYENTES Y DISCREPANTES EN ANALISIS DE REGRESION.

Consideremos el modelo

$$Y = X\beta + e,$$

donde

$Y_{n \times 1}$ es un vector de observaciones,

$X_{n \times p}$ es una matriz de rango completo de constantes conocidas,

$\beta_{p \times 1}$ es un vector de parámetros desconocidos,

y

$e_{n \times 1}$ es un vector de errores aleatorios tal que

$$E(e) = 0 \quad \text{y} \quad \text{Var}(e) = \sigma^2 I.$$

El estimador de β por mínimos cuadrados es $\hat{\beta} = (X'X)^{-1}X'Y$.

Sean $X_{(i)}$ la matriz de $(n-1) \times p$ obtenida a partir de X al

eliminar el i -ésimo renglón x_i , y $Y_{(i)}$ el vector de $(n-1) \times 1$ obtenido de Y al eliminar la i -ésima observación y_i .

Por lo tanto, el estimador de β al eliminar el i -ésimo caso (x_i y y_i) es

$$\hat{\beta}_{(i)} = (\bar{X}'_{(i)} \bar{X}_{(i)})^{-1} \bar{X}'_{(i)} Y_{(i)}$$

Sin perder generalidad supongamos que el n -ésimo caso es el que no se considerará, entonces

$$\bar{X} = \begin{bmatrix} \bar{X}_{(n)} \\ x'_n \end{bmatrix} \quad \text{y} \quad Y = \begin{bmatrix} Y_{(n)} \\ y_n \end{bmatrix},$$

y por lo tanto

$$\bar{X}'\bar{X} = \begin{bmatrix} \bar{X}'_{(n)} & x_n \end{bmatrix} \begin{bmatrix} \bar{X}_{(n)} \\ x'_n \end{bmatrix} = \bar{X}'_{(n)} \bar{X}_{(n)} + x_n x'_n.$$

De aquí que

$$\begin{aligned} (\bar{X}'_{(n)} \bar{X}_{(n)})^{-1} &= (\bar{X}'\bar{X} - x_n x'_n)^{-1} \\ &= (\bar{X}'\bar{X})^{-1} + \frac{(\bar{X}'\bar{X})^{-1} x_n x'_n (\bar{X}'\bar{X})^{-1}}{1 - x'_n (\bar{X}'\bar{X})^{-1} x_n}, \end{aligned}$$

y entonces

$$\begin{aligned} \hat{\beta}_{(i)} &= (\bar{X}'\bar{X})^{-1} \bar{X}' Y - \frac{(\bar{X}'\bar{X})^{-1} x_n x'_n (\bar{X}'\bar{X})^{-1} \bar{X}' Y}{1 - x'_n (\bar{X}'\bar{X})^{-1} x_n} \end{aligned}$$

$$\begin{aligned}
&= (\mathbf{X}'_n \mathbf{X}_n)^{-1} (\mathbf{X}'_n \mathbf{Y}_n + x_n y_n) - \frac{x_n \hat{\beta}}{1 - x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n} (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n \\
&= \hat{\beta}(n) + (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n y_n - \frac{x_n \hat{\beta}}{1 - x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n} (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n \\
&= \hat{\beta}(n) + (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n y_n + \frac{(\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n y_n}{1 - x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n} \\
&\quad - \frac{x_n \hat{\beta}}{1 - x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n} (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n \\
&= \hat{\beta}(n) + \frac{(\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n y_n - x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n y_n}{1 - x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n} \\
&\quad + \frac{x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n y_n}{1 - x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n} - \frac{x_n \hat{\beta}}{1 - x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n} (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n \\
&= \hat{\beta}(n) + \frac{(\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n y_n}{1 - x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n} - \frac{x_n \hat{\beta}}{1 - x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n} (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n \\
&= \hat{\beta}(n) + \left[\frac{y_n - x_n \hat{\beta}}{1 - x_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n} \right] (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_n'.
\end{aligned}$$

De igual manera se obtiene

$$\hat{\beta}_i = \hat{\beta}_i(n) + \left[\frac{y_i - x_i \hat{\beta}_i}{1 - x_i (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_i} \right] (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_i'$$

$i=1, 2, \dots, n$

y por lo tanto

$$\text{CI}_-(x_i, y_i; \beta) = (n-1) \left[\frac{y_i - x_i \hat{\beta}_i}{1 - x_i (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_i} \right] (\mathbf{X}'_n \mathbf{X}_n)^{-1} x_i'.$$

Por lo tanto se tiene un vector de p funciones de influencia el cual no siempre es fácil de analizar. Una alternativa fue propuesta por Cook (1977) quien define una cantidad equivalente a $CI(x_i, y_i; \beta)$ en el sentido que es una medida de la distancia entre $\hat{\beta}$ y $\hat{\beta}_{(i)}$.

La medida para cada i se denota por D_i , llamada la distancia de Cook, y está dada por

$$D_i = \frac{(CI(x_i, y_i; \beta))' X' X (CI(x_i, y_i; \beta))}{p \hat{\sigma}^2 (n-1)^2}$$

de donde valores grandes de D_i corresponderán a observaciones que al eliminarlas producen un gran cambio en la estimación del parámetro β , es decir, corresponderá a una observación influyente y/o discrepante.

Por otro lado, D_i también puede expresarse como

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})' X' X (\hat{\beta}_{(i)} - \hat{\beta})}{p \hat{\sigma}^2}$$

a partir de la cual, y teniendo la suposición de normalidad sobre los errores, se puede pensar que la significancia de la diferencia entre $\hat{\beta}$ y $\hat{\beta}_{(i)}$ puede ser analizada evaluada con

$$P_i = P [F_{(p, n-p)} \geq D_i] ,$$

aunque hay que notar que la función de distribución de la D_1 no es F , pero ésta se usa para proporcionar una escala de referencia conocida.

Así, si p_1 es grande no habrá gran diferencia entre $\hat{\beta}$ y $\hat{\beta}_1$,
 si p_1 es pequeño, se tiene diferencias grandes y por lo tanto una observación influyente y/o discrepante. (1)

Este análisis es equivalente a obtener la región de confianza para β calculada con todos los datos y verificar si $\hat{\beta}_1$ cae dentro o fuera de dicha región. (1)

El problema ahora es como diferenciar entre una observación influyente y una discrepante. Para esto consideremos:

$$D_i = \frac{\left[\left(\frac{y_i - x_i' \hat{\beta}}{1 - x_i' (X'X)^{-1} x_i} \right) x_i' (X'X)^{-1} \right] X'X \left[\left(\frac{y_i - x_i' \hat{\beta}}{1 - x_i' (X'X)^{-1} x_i} \right) (X'X)^{-1} x_i \right]}{p \hat{\sigma}^2}$$

$$= \left[\frac{y_i - x_i' \hat{\beta}}{1 - x_i' (X'X)^{-1} x_i} \right]^2 \frac{x_i' (X'X)^{-1} X'X (X'X)^{-1} x_i}{p \hat{\sigma}^2}$$

$$= \left[\frac{y_i - x_i' \hat{\beta}}{\hat{\sigma} \sqrt{1 - x_i' (X'X)^{-1} x_i}} \right]^2 \left[\frac{x_i' (X'X)^{-1} x_i}{1 - x_i' (X'X)^{-1} x_i} \right] \cdot \frac{1}{p}$$

Por lo tanto, D_i depende de tres cantidades relevantes, las tres referidas al conjunto de casos completo. Estas son:

- i) El número de parámetros p .
- ii) el i -ésimo residual estudentizado

$$t_i = \frac{y_i - x_i' \hat{\beta}}{\hat{\sigma} \sqrt{1 - x_i' (X'X)^{-1} x_i}},$$

y iii) el cociente entre la varianza del i -ésimo valor predicho y la varianza del i -ésimo residual r_i , es decir,

$$D_i = \frac{t_i^2 \text{Var}(\hat{y}_i)}{p \text{Var}(r_i)}$$

donde

$$\text{Var}(\hat{y}_i) = \sigma^2 x_i' (X'X)^{-1} x_i$$

y

$$\text{Var}(r_i) = \sigma^2 (1 - x_i' (X'X)^{-1} x_i)$$

respectivamente.

D_i puede ser grande si t_i o $x_i' (X'X)^{-1} x_i$ son grandes. Si t_i es grande se tendrá una observación aberrante, mientras que si $x_i' (X'X)^{-1} x_i$

x_1 es grande la observación será influyente. Nótese que una observación puede simultáneamente aberrante e influyente.

- DETECCION DE OBSERVACIONES DISCREPANTES EN EL COEFICIENTE DE CORRELACION.

Dada una muestra aleatoria de una distribución bivariada $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, el coeficiente de correlación muestral está dado por

$$r = \frac{\sum_j (x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_j (x_j - \bar{x})^2} \sqrt{\sum_j (y_j - \bar{y})^2}}$$

donde $\bar{x} = \sum_j \frac{x_j}{n}$ y $\bar{y} = \sum_j \frac{y_j}{n}$.

Si ahora no se considera la i -ésima pareja, se tendrá que el coeficiente de correlación es

$$r_{-i} = \frac{\sum_{j \neq i} (x_j - \bar{x}_{-i})(y_j - \bar{y}_{-i})}{\sqrt{\sum_{j \neq i} (x_j - \bar{x}_{-i})^2} \sqrt{\sum_{j \neq i} (y_j - \bar{y}_{-i})^2}}$$

donde $\bar{x}_{-i} = \sum_{j \neq i} \frac{x_j}{n-1}$ y $\bar{y}_{-i} = \sum_{j \neq i} \frac{y_j}{n-1}$.

Sin embargo, en este caso no es fácil obtener una expresión sencilla para la curva de influencia muestral

$$CI_{-}(x_i, y_i; \rho) = (n-1)(r - r_{ic}),$$

por lo que Devlin, et. al. (1975) propusieron utilizar la forma muestral análoga a la función de influencia teórica como una buena aproximación, es decir,

$$CI_{-}(x_i, y_i; \rho) \approx -\frac{1}{2} r (x_i^{*2} + y_i^{*2}) + x_i^{*} y_i^{*}$$

o equivalentemente,

$$CI_{-}(x_i, y_i; \rho) \approx (1-r^2) u_{i1} u_{i2},$$

donde u_{i1} y u_{i2} son las versiones muestrales de u_1 y u_2 , es decir,

$$u_{i1} = \frac{1}{2} \left[\frac{x_i^{*} + y_i^{*}}{\sqrt{1+r}} + \frac{x_i^{*} - y_i^{*}}{\sqrt{1-r}} \right]$$

$$u_{i2} = \frac{1}{2} \left[\frac{x_i^{*} + y_i^{*}}{\sqrt{1+r}} - \frac{x_i^{*} - y_i^{*}}{\sqrt{1-r}} \right]$$

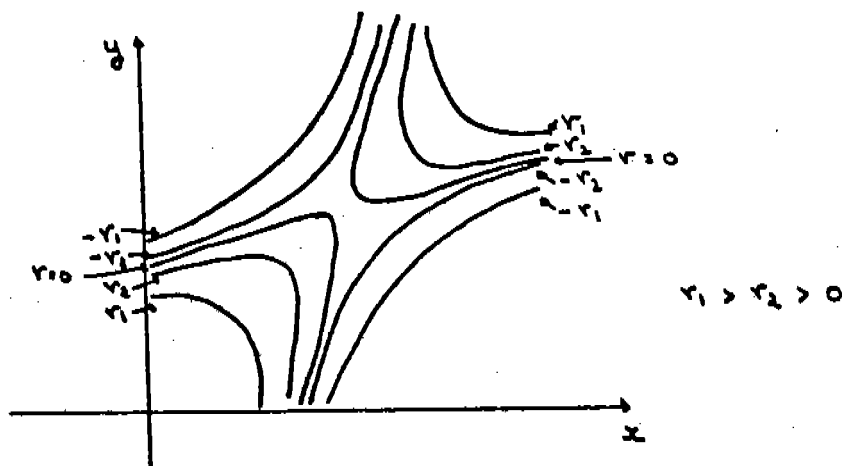
y donde

$$x_i^{*} = \frac{x_i - \bar{x}}{\sqrt{\sum_j \frac{(x_j - \bar{x})^2}{n-1}}}$$

y

$$y_i^{*} = \frac{y_i - \bar{y}}{\sqrt{\sum_j \frac{(y_j - \bar{y})^2}{n-1}}}$$

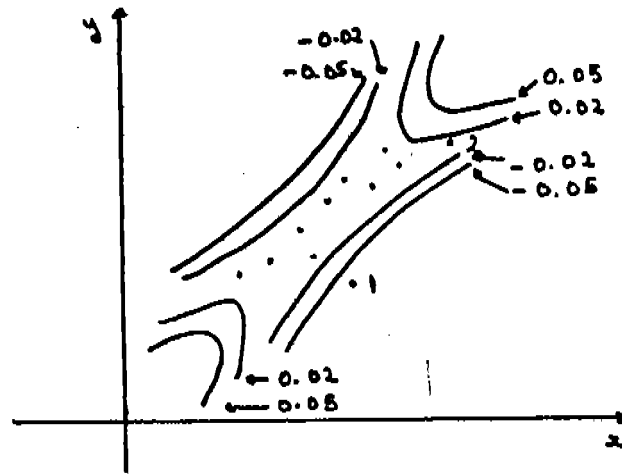
Para r fijo, la aproximación a $CI(x, y, \rho)$ es esencialmente una hipérbola como se muestra en la siguiente gráfica



Esta propiedad de la curva de influencia muestral es usada por Devlin, et. al. para desarrollar un método gráfico para detectar observaciones bivariadas que pueden distorsionar gravemente el valor de r .

Este método consiste en sobreponer la gráfica de dispersión simple de la muestra con la gráfica de las curvas de nivel (para algunos valores de r) y visualmente analizar la influencia de cada observación ya que las curvas de nivel indican aproximadamente que tanto se incrementa o decrementa r por cada una de las observaciones.

Por ejemplo, si la gráfica obtenida fuera la que se muestra en la siguiente figura, sugeriría que el coeficiente de correlación se incrementaría aproximadamente en 0.05 si el punto 1 fuera omitido de la muestra y decrecería aproximadamente un 0.02 si el punto 2 fuera omitido.



2.4.3 TERCERA VERSION

Esta última versión muestral fue propuesta por Mallows en 1973. le dió el nombre de función de influencia empírica ya que se obtiene al reemplazar F , la función de distribución de la población, en

$$CI_{T,F}(x) = \lim_{\epsilon \rightarrow 0^+} \frac{T((1-\epsilon)F + \epsilon \delta_x) - T(F)}{\epsilon}$$

por la función de distribución empírica de la muestra, es decir,

$$CIE_{T,F_n}(x) = \lim_{\epsilon \rightarrow 0^+} \frac{T((1-\epsilon)F_n + \epsilon \delta_x) - T(F_n)}{\epsilon},$$

donde F es una función escalonada con saltos de magnitud $1/n$ en cada una de las observaciones x_1, x_2, \dots, x_n .

De aquí que la función de influencia empírica sea una estimación no paramétrica de $CI_{T,F}(x)$, con la cual se puede analizar tanto la influencia de observaciones futuras como la influencia de las observaciones de la muestra.

Hay que notar que tanto CI^+ como CI^- pueden considerarse aproximaciones a la CIE al tomar en esta última $\epsilon = 1/(n+1)$ y $\epsilon = 1/(n-1)$ respectivamente.

La función de influencia empírica es un método equivalente al jackknife infinitesimal de Jaeckel, pues si recordamos, en este último se determina el i -ésimo pseudovalor por:

$$u_i = \lim_{\epsilon \rightarrow 0} \frac{\hat{\theta}((1-\epsilon)\underline{P} + \epsilon \underline{\delta}_i) - \hat{\theta}(\underline{P})}{\epsilon}$$

donde

$$\underline{P} = \frac{1}{n} (1, 1, \dots, 1) \quad \text{y} \quad \underline{\delta}_i = (0, 0, \dots, \underbrace{1}_{i\text{-ésimo lugar}}, \dots, 0)$$

i-ésimo lugar

por lo que u_i no es otra cosa que el valor de la función de influencia empírica evaluada en la i -ésima observación.

3. DATOS DE SUPERVIVENCIA

3.1 INTRODUCCION

En muchos estudios longitudinales, la variable de interés o variable respuesta es el tiempo que transcurre entre la ocurrencia de dos eventos especiales; el primero de ellos se refiere esencialmente a la entrada de un individuo al estudio y el segundo generalmente recibe el nombre de falla o muerte (que en algunos casos pueden estar relacionados con la aplicación de un tratamiento), por lo que la variable respuesta se conoce como tiempo de falla o tiempo de supervivencia.

Se supone que se tienen los tiempos de falla de n individuos y que estos tiempos son independientes. Una seria complicación en el análisis de tiempos de falla(o de supervivencia) es la posibilidad de censura, es decir, no poder observar a los individuos hasta la ocurrencia del segundo evento de interés.

La censura puede deberse a varias razones, como por ejemplo:

1. El estudio puede haber terminado mientras que algunos individuos aún están vivos (esto es, aún no ocurre el evento de interés para cada uno de estos individuos).
2. La falla puede presentarse en algunos individuos por causas que no son de interés para el estudio.
3. Algunos individuos en los que no se ha presentado la falla pueden retirarse del estudio sin que este haya concluido aún.

Por lo que la razón más importantes para desarrollar modelos especiales y procedimientos para datos de tiempo de falla(o tiempos de supervivencia) es la necesidad de obtener métodos de

análisis que consideren censura. Algunos de los procedimientos de análisis están basados en el supuesto que la población es homogénea, otros consideran que los individuos son expuestos a diferentes condiciones experimentales (que influyen en el tiempo de falla) y otros en donde se considera que el tiempo de falla se ve afectado por ciertas variables propias de los individuos que son tomadas como covariables.

También, en muchas situaciones es importante desarrollar procedimientos no paramétricos y robustos puesto que es muy poco el trabajo teórico que se tiene para una familia particular de distribuciones de tiempo de falla.

3.2 EJEMPLOS

Algunos ejemplos de datos de supervivencia son los siguientes, los cuales nos pueden dar idea que este tipo de estudios pueden desarrollarse en diferentes áreas:

1. Supongamos que en un estudio médico sobre Diabetes Mellitus se está interesado en analizar la edad máxima que puede llegar a tener una persona que padece dicha enfermedad. En este caso el segundo evento de interés es la muerte del individuo y la variable respuesta es el tiempo que transcurre desde que a un individuo se le diagnostica Diabetes Mellitus y es captado dentro del estudio hasta su muerte. En este caso los valores censurados pueden deberse a individuos que:
 - Mueren por otra causa que no sea esta enfermedad, por ejemplo envenenamiento, accidentes, etc.
 - Abandonan el estudio pues cambian de lugar de residencia y no es posible localizarlos posteriormente.
2. En un estudio demográfico sobre migración se decide captar a un cierto grupo de individuos y observarlos

durante 10 años para determinar el momento en el que cambian su lugar de residencia por primera vez. En este caso el segundo evento de interés es el cambio de residencia y la variable respuesta es el tiempo que transcurre desde que un individuo (con ciertas características propias definidas dentro del estudio mismo) es entrevistado por primera vez hasta el momento en que cambia de residencia. Las posibles causas de observaciones censuradas presentes en el estudio son por ejemplo:

- El individuo observado muera.
- El individuo no haya cambiado de residencia en los 10 años que dura el estudio.

3. En un análisis sobre el desarrollo de la Maestría en Estadística e Investigación de Operaciones se decide estudiar el tiempo que tarda un alumno de esta maestría en presentar el examen de grado. El segundo evento de interés en este caso es la presentación del examen de grado y la variable respuesta es el tiempo que transcurre desde que el alumno termina el 100% de los créditos hasta el momento en que presenta dicho examen.
4. En una inspección sobre control de calidad de cierto tipo de aparatos eléctricos se decide poner a trabajar un cierto número de estos y registrar el tiempo en el cual presentan la primera falla. Por lo tanto el evento de interés es la presencia de cualquier falla en estos aparatos y la variable respuesta es el tiempo que transcurre desde el momento en que es puesto a funcionar un aparato hasta que presenta la primera falla.

3.3 DISTRIBUCION DEL TIEMPO DE FALLA O SUPERVIVENCIA

Sea T una variable aleatoria no negativa que representa el tiempo de falla o tiempo de supervivencia de un individuo de una cierta población homogénea.

La distribución de T puede expresarse en términos de tres diferentes funciones relacionadas entre sí, cada una de ellas tiene una cierta utilidad y aplicación dentro del análisis de supervivencia. Dichas funciones son:

- i.- La función de supervivencia.
- ii.- La función de densidad de probabilidad de T.
- iii.- La función de riesgo.

i.- La función de supervivencia se define como la probabilidad de que T sea mayor o igual que t, un valor positivo dado. Es decir,

$$S(t) = P[T \geq t] \quad \text{con } 0 < t < \infty$$

independientemente de que T sea continua, discreta o mixta.

Esta función es monótona no-creciente y continua por la izquierda, además se tiene que $S(0) = 1$ y $\lim_{t \rightarrow \infty} S(t) = 0$.

Por lo tanto, la función de supervivencia es esencialmente el complemento con respecto a 1 de la función de distribución de T pues a lo mas difiere en $P[T=t]$ cuando esta probabilidad es positiva. Además para fines prácticos es mejor trabajar con $S(t)$ que con $F(t) = P [T \leq t]$.

ii.- Para definir la función de densidad de probabilidad de T se considera por separado el caso discreto y el caso continuo.

Supongamos en primer lugar que T es una variable aleatoria absolutamente continua; entonces la función de densidad de probabilidad de T se define como

$$f(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P[t \leq T < t + \Delta t]}{\Delta t},$$

pero ya que

$$\begin{aligned} P[t \leq T < t + \Delta t] &= P[T < t + \Delta t] - P[T \leq t] \\ &= 1 - P[T \geq t + \Delta t] - (1 - P[T \geq t]), \end{aligned}$$

entonces

$$\begin{aligned} P[t \leq T < t + \Delta t] &= P[T \geq t] - P[T \geq t + \Delta t] \\ &= S(t) - S(t + \Delta t), \end{aligned}$$

es decir,

$$f(t) = \lim_{\Delta t \rightarrow 0^+} \left[\frac{S(t) - S(t + \Delta t)}{\Delta t} \right]$$

por lo tanto

$$f(t) = - \frac{d}{dt} S(t).$$

De aquí fácilmente se obtiene que

$$S(t) = \int_t^{\infty} f(s) ds,$$

y además se cumple que

$$f(t) \geq 0 \quad \forall t \quad \text{y} \quad \int_0^{\infty} f(s) ds = 1$$

Supongamos ahora que T es discreta y que toma los valores $x_1 < x_2 < \dots$. Entonces la función de densidad de probabilidad de T está dada por

$$f(t) = \begin{cases} P[T = x_j] & \text{si } t = x_j \text{ para alguna } j \\ 0 & \text{en otro caso} \end{cases}$$

de donde se obtiene que la función de supervivencia es

$$S(t) = \sum_{j: x_j \geq t} f(x_j)$$

y se cumple que

$$f(t) \geq 0 \quad \forall t \quad \text{y} \quad \sum_j f(x_j) = 1$$

Supongamos ahora el caso más general, es decir, T tiene una distribución mixta, por lo que T tiene una componente continua y una discreta; entonces la función de densidad de probabilidad de T es la suma de las partes continua y discreta.

iii.- La función de riesgo también se define por separado para el caso de T continua, T discreta y T mixta.

Supongamos que T es absolutamente continua, entonces la función de riesgo mide la tasa instantánea de falla en $T=t$ condicionada a la supervivencia al tiempo t , es decir,

$$\lambda(t) = \lim_{\Delta t \rightarrow 0^+} \left[\frac{P[t \leq T < t + \Delta t \mid T \geq t]}{\Delta t} \right],$$

pero ya que

$$\begin{aligned} P[t \leq T < t + \Delta t \mid T \geq t] &= \frac{P[t \leq T < t + \Delta t, T \geq t]}{P[T \geq t]} \\ &= \frac{P[t \leq T < t + \Delta t]}{P[T \geq t]}, \end{aligned}$$

entonces

$$\lambda(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P[t \leq T < t + \Delta t]}{P[T \geq t] \Delta t} = \frac{1}{S(t)} f(t),$$

o, equivalentemente,

$$\lambda(t) = \frac{1}{S(t)} \left[-\frac{d}{dt} S(t) \right] = -\frac{d}{dt} \ln S(t),$$

por lo que

$$S(t) = e^{-\int_0^t \lambda(u) du}.$$

Por otro lado, se tiene que $f(t) = \lambda(t)S(t)$ por lo que

$$f(t) = \lambda(t) e^{-\int_0^t \lambda(u) du}$$

Supongamos ahora que T es discreta y que toma los valores $x_1 < x_2 < \dots$, entonces la función de riesgo sólo se define para x_j ($j=1,2,\dots$) y da la probabilidad condicional de falla en x_j , es decir,

$$\begin{aligned} \lambda(x_j) \equiv \lambda_j &= P[T=x_j | T \geq x_j] = \frac{P[T=x_j, T \geq x_j]}{P[T \geq x_j]} \\ &= \frac{P[T=x_j]}{P[T \geq x_j]} \\ &= \frac{f(x_j)}{S(x_j)} \quad j=1,2,3,\dots \end{aligned}$$

Ahora bien, utilizando el hecho de que

$$\begin{aligned} \frac{S(x_{j+1})}{S(x_j)} &= \frac{P[T \geq x_{j+1}]}{P[T \geq x_j]} = \frac{P[T \geq x_j] - P[T=x_j]}{P[T \geq x_j]} \\ &= \frac{S(x_j) - f(x_j)}{S(x_j)} = 1 - \frac{f(x_j)}{S(x_j)} = 1 - \lambda_j \\ & \quad j=1,2,\dots \end{aligned}$$

sea $t > 0$ tal que $x_{j-1} < t \leq x_j$ para algún j , entonces la función de supervivencia está dada por:

$$\begin{aligned}
 S(t) &= S(x_{j_0}) \cdot (1 - \lambda_{j_0-1}) S(x_{j_0-1}) \\
 &= (1 - \lambda_{j_0-1}) (1 - \lambda_{j_0-2}) S(x_{j_0-2}) \\
 &= \dots = (1 - \lambda_{j_0-1}) (1 - \lambda_{j_0-2}) \dots (1 - \lambda_1) S(x_1) \\
 &= \prod_{j|x_j < t} (1 - \lambda_j)
 \end{aligned}$$

ya que $S(x_1) = 1$. Entonces, la función de densidad de probabilidad es

$$\begin{aligned}
 f(x_j) &= \lambda_j S(x_j) \\
 &= \lambda_j \prod_{j|x_j < t} (1 - \lambda_j)
 \end{aligned}$$

Supongamos que T tiene una distribución mixta, por lo que tendrá una componente continua y una discreta, entonces la función de supervivencia considerará productos de términos para cada parte, es decir,

$$S(t) = e^{-\int_0^t \lambda_c(u) du} \prod_{j|x_j < t} (1 - \lambda_j),$$

en donde λ es la función de riesgo para la parte continua y $x_1 < x_2 < x_3 < \dots$ los puntos con probabilidad positiva.

En este caso la función de riesgo puede expresarse como

$$\lambda(t) dt = \lambda_c(t) dt + \sum_j \lambda_j \delta(t - x_j)$$

en donde $\delta(t - x_j)$ es una delta de Dirac.

Otra función que también puede usarse en el análisis de la distribución de T , es la función de riesgo acumulada que se define como

$$A(t) = \int_0^t \lambda(u) du = \int_0^t \lambda_c(u) du + \sum_{j|x_j < t} \lambda_j$$

Por otro lado, la función de supervivencia $S(t)$, puede ser expresada en términos de la función $\Lambda(t)$ independientemente de la naturaleza de T , de la siguiente manera:

$$s(t) = \prod_0^t (1 - d\Lambda(u)),$$

donde \prod_0^t representa el producto integral, el cual está definido como

$$\prod_a^b (1 - d\Lambda(u)) = \lim_{r \rightarrow \infty} \prod_{k=1}^r [1 - (\Lambda(u_k) - \Lambda(u_{k-1}))]$$

con $a = u_0 < u_1 < \dots < u_r = b$ y el límite se toma cuando $r \rightarrow \infty$ y $(u_k - u_{k-1}) \rightarrow 0$.

Esta expresión se cumple puesto que para la parte continua se tiene que

$$\prod_0^t (1 - \lambda_c(u) du) = e^{-\int_0^t \lambda_c(u) du}$$

y para la parte discreta

$$\prod_0^t (1 - d \sum \lambda_j(u - x_j)) = \prod_{j: x_j < t} (1 - \lambda_j).$$

Además, de esta manera se puede ver que la probabilidad de

supervivencia al tiempo t es el producto de las probabilidades condicionales de supervivencia para intervalos infinitesimales hasta el tiempo t .

Dadas las relaciones entre la función de distribución de T , la función de supervivencia, la función de densidad de probabilidad y la función de riesgo, bastará estimar, ajustar o modelar una de estas funciones para tener idea del comportamiento de las otras, con lo que se tendrá un conocimiento amplio acerca de T .

3.4 ESTIMACION DE LA FUNCION DE SUPERVIVENCIA

Para la función de supervivencia $S(t) = P\{T \geq t\}$ se han propuesto diferentes estimadores. Entre los más comunes se pueden citar:

1. La función de supervivencia empírica.
2. La tabla de vida.
3. La tabla de vida modificada.
4. El estimador Kaplan-Meier.

3.4.1 LA FUNCION DE SUPERVIVENCIA EMPIRICA

Dado que $S(t) = P\{T \geq t\} = 1 - P\{T < t\}$, se puede pensar que una primera forma de estimar $S(t)$ se obtendrá si estimamos la función de distribución de T .

Se sabe que dada una muestra aleatoria de una población homogénea³, el mejor estimador de la función de distribución es la función empírica $F_n(t)$, donde

3

En el sentido de que cada sujeto de la población tiene la misma distribución del tiempo de falla

$$F(t) = \frac{\text{Número de valores muestrales } < t}{n}$$

con n = tamaño de la muestra.

Por lo tanto si tenemos una muestra aleatoria no censurada de tamaño n de tiempos de falla, los cuales fueron observados de una población homogénea, la función de supervivencia muestral

$$S(t) = \frac{\text{Número de valores muestrales } > t}{n}$$

será una buena estimación de $S(t)$. Esta función es escalonada decreciente con saltos de tamaño $1/n$ después de cada tiempo de falla observado.

Sin embargo, como se dijo anteriormente, los datos de supervivencia muy a menudo presentan censura, por lo cual esta estimación no es la más conveniente.

3.4.2 LA TABLA DE VIDA

Este estimador describe la distribución de tiempos de supervivencia experimentados por un grupo homogéneo de individuos cuando estos tiempos están sujetos a censura. Es el más antiguo de los estimadores de la función de supervivencia.

La tabla de vida es un resumen de los datos de supervivencia agrupados en intervalos, es decir, presenta el número de fallas y tiempos de supervivencia censurados en cada intervalo.

Supongamos que los datos están agrupados en intervalos I_1, I_2, \dots, I_k tales que:

$$I_j = [b_{j-1}, b_j] \quad j = 1, \dots, k$$

$$j = 1, \dots, k$$

donde $b_0 = 0$ y $b_k = \infty$

y que en I_j caen c_j tiempos censurados y t_j tiempos de falla.

Sean n_j el número de individuos expuestos al riesgo al inicio del j -ésimo intervalo, es decir,

$$n_j = \sum_{l \geq j} (t_l + c_l)$$

Entonces, el estimador de la probabilidad condicional de falla en el intervalo I_j dada la supervivencia al inicio de I_j es:

$$\hat{P}[T \in I_j \mid T \geq b_0 + \dots + b_{j-1}] = \hat{q}_j = \begin{cases} 1 & \text{si } n_j = 0 \\ \frac{t_j}{n_j - c_j/2} & \text{si } n_j > 0 \end{cases}$$

El término $c_j/2$ en el denominador se utiliza como ajuste por el hecho de que no todos los n_j individuos están expuestos al riesgo durante todo el intervalo I_j . Esto es una aproximación bajo el supuesto que los tiempos de los c_j individuos se distribuyen en

forma uniforme o por lo menos simétrica en I_j , lo cual es válido para intervalos pequeños.

Por lo tanto, la estimación de la función de supervivencia evaluada al final del intervalo I_j es

$$\tilde{S}(b_0 + \dots + b_j) = \tilde{P} [T \geq b_0 + \dots + b_j]$$

pero

$$\begin{aligned} \tilde{P} [T \geq b_0 + \dots + b_j] &= \tilde{P} [T \geq b_0 + b_1 | T \geq b_0] \tilde{P} [T \geq b_0 + \dots + b_2 | T \geq b_0 + b_1] \\ &\dots \tilde{P} [T \geq b_0 + b_1 + \dots + b_j | T \geq b_0 + \dots + b_{j-1}] \end{aligned}$$

y

$$\tilde{P} [T \geq b_0 + \dots + b_l | T \geq b_0 + \dots + b_{l-1}] = 1 - \tilde{q}_l \quad l = 1, 2, \dots, j$$

por lo tanto

$$\tilde{S}(b_0 + \dots + b_j) = \prod_{i=1}^j (1 - \tilde{q}_i) \quad j = 1, 2, \dots, k.$$

Hay que notar que para utilizar esta estimación no es necesario conocer los tiempos de falla y los de censura, basta saber cuantos de ellos se tienen en cada uno de los intervalos considerados. Por lo anterior esta forma de estimación es muy pobre pues se pierde la información sobre los tiempos observados; además, depende de la elección de los intervalos I_j .

3.4.3 LA TABLA DE VIDA MODIFICADA

Esta forma de estimar la función de supervivencia se utiliza cuando se tienen los datos en una tabla de vida y además se conocen los tiempos de falla y los tiempos de censura de todos los individuos observados.

Por lo tanto, en el intervalo I_j se tienen registrados los tiempos $t_{j1}, t_{j2}, \dots, t_{jr_j}$ de los cuales c_j fueron tiempos de censura y t_{jr_j} tiempos de falla ($r_j = c_j + t_j$). Si suponemos además que la función de riesgo $\lambda_j(t)$ es una función escalonada con valores constantes λ_j en cada intervalo I_j entonces un estimador de λ_j es

$$\hat{\lambda}_j = \frac{t_j}{TS_j}$$

$$j = 1, 2, \dots, k,$$

donde TS_j es el tiempo de supervivencia observado en el intervalo I_j , es decir,

$$TS_j = \sum_{k=1}^{r_j} (t_{jk} - \sum_{i=0}^{k-1} b_i) + n_{j+1} b_j.$$

Finalmente, el estimador de la función de supervivencia es

$$\hat{S}(t) = e^{-[\hat{\lambda}_j(t - \sum_{i=0}^{j-1} b_i) + \sum_{i=0}^{j-1} \hat{\lambda}_i b_i]}$$

si $t \in I_j$, $j = 1, 2, \dots, k$.

el cual es continuo en t a diferencia de los anteriores

estimadores, pero sin embargo, depende también de la elección de los intervalos.

3.4.4 EL ESTIMADOR KAPLAN-MEIER

En 1958 Kaplan y Meier muestran como puede ser obtenido un estimador simple de la función de supervivencia sin hacer ningún agrupamiento de los tiempos de falla y de censura. De hecho los supuestos que hacen son de que el grupo de individuos observados es homogéneo y que los tiempos de supervivencia han sido registrados exactamente, es decir, sin redondeos o truncaciones. Este estimador también se le conoce como el estimador del producto-límite, y su construcción es como sigue:

Sean $t_1 < t_2 < \dots < t_k$ los tiempos de falla observados en una muestra de tamaño n de una población homogénea con función de supervivencia $S(t)$.

Supongamos que al tiempo t_j ($j=1, \dots, k$) se registran d_j fallas y c_j censuras se presentan en el intervalo $[t_j, t_{j+1})$ a los tiempos $t_{j1}, t_{j2}, \dots, t_{jc}$ con $j=0, 1, \dots, k$ en donde $t_0 = 0$ y $t_{k+1} = \infty$.

Sean n_j el número de individuos expuestos al riesgo inmediatamente antes de t_j , es decir,

$$n_j = \sum_{l=j}^k (d_l + c_l).$$

Por lo tanto, la probabilidad de falla al tiempo t_j será

$$S(t_j) - S(t_j + 0) \quad j=1,2,\dots,k$$

en donde $S(t_j + 0) = \lim_{x \rightarrow 0^+} S(t_j + x)$.

Supongamos que el tiempo de censura observado t_{jl} ($j=1,\dots,k$ y $l=1,2,\dots,c$) sólo nos indica que el tiempo de falla no observado es mayor que t_{jl} y que la contribución de esta censura a la función de verosimilitud que se propone es

$$P[T > t_{jl}] = S(t_{jl} + 0)$$

Entonces, la función de verosimilitud será

$$\mathcal{L} = \prod_{j=0}^k [(S(t_j) - S(t_j + 0))^{d_j} \prod_{l=1}^{c_j} S(t_{jl} + 0)]$$

la cual puede considerarse, para una muestra fija, como una función sobre el espacio de funciones de supervivencia. El estimador máximo-verosímil de $S(t)$ será la función de supervivencia $\hat{S}^0(t)$ que maximice \mathcal{L} .

$\hat{S}^0(t)$ tendrá que ser discontinua en los tiempos de falla observados, puesto que en caso contrario $S(t_j + 0) = S(t_j)$ y por lo tanto $\mathcal{L} = 0$ con lo cual no se tendrá un máximo.

Además si $t_{jl} \geq t_j$, $S(t_j + 0)$ será máximo si se cumple que $S(t_{jl} + 0) = S(t_j + 0)$ para $j=1,2,\dots,k$, $l=1,2,\dots,c$ y $S(t_{ol}) = 1$ para $l=1,2,\dots,c$.

Por lo tanto, $\hat{S}^0(t)$ tendrá que ser una función de supervivencia discreta con componentes de riesgo $\lambda_1, \lambda_2, \dots, \lambda_k$ en t_1, t_2, \dots, t_k respectivamente. De aquí que

$$\hat{S}^0(t_j) = \prod_{l=1}^{j-1} (1 - \hat{\lambda}_l)$$

y

$$\hat{S}^0(t_j + 0) = \prod_{l=1}^j (1 - \hat{\lambda}_l),$$

donde ahora el problema es encontrar $\lambda_l, l=1, 2, \dots, k$ que maximicen \mathcal{L} .

Entonces,

$$\begin{aligned} \mathcal{L} &= \prod_{j=0}^k [(s(t_j) - s(t_j + 0))^{d_j} \prod_{l=1}^{c_j} s(t_{j+l} + 0)] \\ &= [s(t_0) - s(t_0 + 0)]^{d_0} \prod_{l=1}^{c_0} s(t_{0+l} + 0) \prod_{j=1}^k [(s(t_j) - s(t_j + 0))^{d_j} \prod_{l=1}^{c_j} s(t_{j+l} + 0)] \\ &= \prod_{j=1}^k (s(t_j) - s(t_j + 0))^{d_j} \prod_{l=1}^{c_j} s(t_{j+l} + 0) \\ &= \prod_{j=1}^k (s(t_j) - s(t_j + 0))^{d_j} \prod_{l=1}^{c_j} s(t_j + 0) \\ &= \prod_{j=1}^k (s(t_j) - s(t_j + 0))^{d_j} (s(t_j + 0))^{c_j} \\ &= \prod_{j=1}^k \left[\prod_{l=1}^{j-1} (1 - \lambda_l) - \prod_{l=1}^j (1 - \lambda_l) \right]^{d_j} \left[\prod_{l=1}^j (1 - \lambda_l) \right]^{c_j} \end{aligned}$$

BIBLIOTECA
JUAN A. ESCALANTE II.
UNIDAD ACADÉMICA DE
LOS CICLOS PROFESIONAL
Y DE POSGRADO / CCH
UNAM

$$\begin{aligned}
&= \prod_{j=1}^k \left[\lambda_j \prod_{\lambda=1}^{j-1} (1-\lambda_\lambda) \right]^{d_j} \left[\prod_{\lambda=1}^j (1-\lambda_\lambda) \right]^{c_j} \\
&= \prod_{j=1}^k \lambda_j^{d_j} \left[\prod_{\lambda=1}^{j-1} (1-\lambda_\lambda)^{d_j} \right] \left[\prod_{\lambda=1}^j (1-\lambda_\lambda)^{c_j} \right] \\
&= \prod_{j=1}^k \lambda_j^{d_j} (1-\lambda_j)^{c_j} \left[\prod_{\lambda=1}^{j-1} (1-\lambda_\lambda)^{c_j+d_j} \right] \\
&= \lambda_1^{d_1} (1-\lambda_1)^{c_1} \lambda_2^{d_2} (1-\lambda_2)^{c_2} (1-\lambda_1)^{c_2+d_2} \lambda_3^{d_3} (1-\lambda_3)^{c_3} (1-\lambda_1)^{c_3+d_3} \\
&\quad (1-\lambda_2)^{c_3+d_3} \dots \lambda_{k-1}^{d_{k-1}} (1-\lambda_{k-1})^{c_{k-1}} (1-\lambda_1)^{c_{k-1}+d_{k-1}} (1-\lambda_2)^{c_{k-1}+d_{k-1}} \\
&\quad \dots (1-\lambda_{k-2})^{c_{k-1}+d_{k-1}} \lambda_k^{d_k} (1-\lambda_k)^{c_k} (1-\lambda_1)^{c_k+d_k} \\
&\quad \dots (1-\lambda_{k-1})^{c_k+d_k} \\
&= \lambda_1^{d_1} \lambda_2^{d_2} \dots \lambda_k^{d_k} (1-\lambda_1)^{c_1+c_2+d_2+\dots+c_k+d_k} \\
&\quad (1-\lambda_2)^{c_2+c_3+d_3+\dots+c_k+d_k} \dots (1-\lambda_{k-1})^{c_{k-1}+c_k+d_k} (1-\lambda_k)^{c_k} \\
&= \lambda_1^{d_1} \lambda_2^{d_2} \dots \lambda_k^{d_k} (1-\lambda_1)^{n_1-d_1} (1-\lambda_2)^{n_2-d_2} \dots (1-\lambda_k)^{n_k-d_k}
\end{aligned}$$

y por lo tanto

$$\mathcal{L} = \prod_{j=1}^k \lambda_j^{d_j} (1-\lambda_j)^{n_j-d_j}.$$

El problema se reduce a maximizar esta última expresión o equivalentemente a maximizar

$$\begin{aligned} \ln \mathcal{L} &= \ln \prod_{j=1}^k \lambda_j^{d_j} (1-\lambda_j)^{n_j-d_j} \\ &= \sum_{j=1}^k (d_j \ln \lambda_j + (n_j - d_j) \ln(1-\lambda_j)) . \end{aligned}$$

Esto equivale a resolver la expresión

$$\frac{\partial \ln \mathcal{L}}{\partial \lambda_i} = \frac{d_i}{\lambda_i} + \frac{(n_i - d_i)(-1)}{1 - \lambda_i} = 0$$

$$i = 1, 2, \dots, k .$$

Por lo tanto, el estimador de máxima-verosimilitud para λ_i será

$$\hat{\lambda}_i = \frac{d_i}{n_i} \quad i = 1, 2, \dots, k .$$

Además

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda_i^2} = - \frac{d_i}{\lambda_i^2} - \frac{(n_i - d_i)}{1 - \lambda_i^2} < 0$$

y

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda_i \partial \lambda_j} = 0$$

por lo que la matriz de segundas derivadas es diagonal con todos los elementos de la misma negativos, es decir, es una matriz negativa definida lo que asegura que se tiene un máximo.

Entonces el estimador Kaplan-Meier es

$$\hat{S}^{\circ}(t) = \prod_{j|t_j < t} (1 - \hat{\lambda}_j) = \prod_{j|t_j < t} \left(\frac{n_j - d_j}{n_j} \right).$$

Sin embargo, para los desarrollos posteriores de este trabajo es más conveniente utilizar la expresión propuesta por Peterson (1977) del estimador Kaplan-Meier, es decir, como una función de dos funciones empíricas de subsupervivencia.

Sean $X_1^{\circ}, X_2^{\circ}, \dots, X_n^{\circ}$ los verdaderos tiempos de supervivencia de n individuos extraídos de una población homogénea con función de distribución $F^{\circ}(x)$. Para cada X_i° se tiene asociada una variable de censura independiente Y_i° en donde supondremos que $Y_1^{\circ}, Y_2^{\circ}, \dots, Y_n^{\circ}$ son variables aleatorias independientes idénticamente distribuidas con función de distribución $H(y)$.

Para cada individuo observamos dos valores

$$X_i = \min(X_i^{\circ}, Y_i^{\circ})$$

$$\delta_i = I_{\{X_i \leq Y_i^{\circ}\}}$$

en donde δ_i indica si X_i es un valor censurado ($\delta_i = 0$) o no ($\delta_i = 1$).

Sea $F(x)$ la función de distribución de las X_i 's.

$$\begin{aligned} F(t) &= P[X_i \leq t] = 1 - P[X_i > t] = 1 - P[\min(X_i, Y_i) > t] \\ &= 1 - P[X_i > t, Y_i > t] = 1 - P[X_i > t] P[Y_i > t] \\ &= 1 - (1 - F^o(t))(1 - H(t)). \end{aligned}$$

Definimos la función de subdistribución para una observación no censurada como

$$F^u(t) = P[X_i \leq t, \delta_i = 1],$$

y la función de subdistribución para una observación censurada como

$$F^c(t) = P[X_i \leq t, \delta_i = 0].$$

Entonces se tienen las siguientes relaciones:

$$i) F^u(\infty) = P[X_i \leq \infty, \delta_i = 1] = P[\delta_i = 1]$$

$$ii) F^c(\infty) = P[X_i \leq \infty, \delta_i = 0] = P[\delta_i = 0]$$

$$iii) F^u(t) + F^c(t)$$

$$= P[X_i \leq t, \delta_i = 1] + P[X_i \leq t, \delta_i = 0]$$

$$= P[X_i \leq t, \delta_i = 0 \text{ o } \delta_i = 1]$$

$$= P[X_i \leq t] = F(t)$$

de aquí el nombre de subdistribuciones a F^u y a F^c .

$$iv) dF^u(t) = (1 - H(t)) dF^c(t)$$

la demostración de iv) se presenta en el Apéndice 2.

Cada una de las funciones de distribución puede estimarse a partir de la correspondiente función empírica de la muestra, es decir,

$$\tilde{F}(t) = F_n(t) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq t\}}$$

$$\tilde{F}^u(t) = F_n^u(t) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq t, \delta_i = 1\}}$$

$$\tilde{F}^c(t) = F_n^c(t) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq t, \delta_i = 0\}}.$$

Además se tendrán las funciones de supervivencia $S^0(t) = 1 - F^0(t)$ y $S(t) = 1 - F(t)$ junto con las funciones de

subsupervivencia $S^u(t) = P[X_1 \geq t, \delta_1 = 1]$ y $S^o(t) = P[X_1 \geq t, \delta_1 = 0]$ y sus versiones muestrales.

Entonces, el estimador de Kaplan-Meier para la función de supervivencia $S^o(t)$ es

$$\hat{S}^o(t) = \begin{cases} \prod_{i: X_{(i)} \leq t} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}} & \text{para } t \leq X_{(n)} \\ 0 & \text{si } \delta_{(n)} = 1 \text{ para } t > X_{(n)} \\ \text{no definido} & \text{si } \delta_{(n)} = 0 \text{ para } t > X_{(n)} \end{cases}$$

en donde $X_{(1)} < X_{(2)} < \dots < X_{(n)}$

$$y \quad \delta_{(i)} = I_{\{X_{(i)} \text{ no censurada}\}}$$

El estimador Kaplan-Meier $\hat{S}^o(t)$ es una función escalonada con saltos sólo en las observaciones no censuradas. El tamaño del salto en cada observación no censurada es una función del número de observaciones y el patrón de pérdida que se ha presentado antes de la falla. Una observación adicional cambiará todos los tamaños de los saltos del estimador. Si la nueva observación es no censurada, un salto extra será introducido en $\hat{S}^o(t)$, sin embargo, si la nueva observación es censurada, no habrá ningún salto adicional. Si una observación es eliminada también cambiarán todos los tamaños de los saltos del estimador, pero si la observación es no censurada disminuirá en 1 el número de saltos

mientras que si es censurada no cambiará el número de saltos del estimador. Esto constituye la diferencia esencial en el efecto de una observación (ya sea nueva o de la muestra misma) que hace natural la expresión propuesta por Peterson.

Como puede verse, un inconveniente del estimador Kaplan-Meier es que si el último valor observado corresponde a un valor censurado quedará indefinido el valor de este estimador. Una forma de eliminar este problema es definir $\hat{S}^0(t) = 0$ para $t > X_{(n)}$ independientemente de la naturaleza de la última observación siempre y cuando el tamaño de muestra n sea grande. Sin embargo, aunque en este tipo de estudios el tamaño de muestra suele no ser lo suficientemente grande como para considerar el caso anterior, más adelante se propone una forma de resolver dicho problema de manera general.

Si $\delta_i = 1$ para $i=1,2,\dots,n$, es decir, si no hay observaciones censuradas, el estimador Kaplan-Meier se reduce a la función de supervivencia empírica usual.

Las propiedades de este estimador han sido estudiadas por varios autores, los cuales han demostrado que:

- Es el estimador máximo-verosímil.
- Es un estimador fuertemente consistente.
- Es asintóticamente normal.
- Visto como un proceso estocástico en t , converge débilmente a un proceso gaussiano.

La importancia de la censura (o la no censura) en la

construcción del estimador $\hat{S}^o(t)$ es más clara si se representa este estimador en términos de las funciones de subsupervivencia empíricas $S^u_n(t)$ y $S^c_n(t)$. Esto es posible puesto que Peterson mostró que para la verdadera función de supervivencia $S^o(t)$ se cumple que

$$S^o(t) = e^{\int_0^t \frac{dS^u(x)}{S^u(x) + S^c(x)}} e^{\sum_{x \leq t} \ln \left[\frac{S^u(x^+) + S^c(x^+)}{S^u(x^-) + S^c(x^-)} \right]}$$

donde la integración se realiza sobre la unión de intervalos abiertos de puntos menores que t para los cuales $S^u(\cdot)$ es continua y la suma es sobre los puntos x 's los cuales son puntos de discontinuidad de $S^u(\cdot)$.

Por lo tanto, si S^u es continua para todo t , la suma es cero y

$$S^o(t) = e^{\int_0^t \frac{dS^u(x)}{S^u(x) + S^c(x)}} ,$$

mientras que si S^u es completamente discreta, la integral vale cero y

$$S^o(t) = e^{\sum_{x \leq t} \ln \left[\frac{S^u(x^+) + S^c(x^+)}{S^u(x^-) + S^c(x^-)} \right]}$$

La expresión de $S^o(t)$ no es otra cosa más que la generalización de la relación

$$S^o(t) = 1 - F^o(t) = e^{\ln(1 - F^o(t))} = e^{-\int_0^t \frac{dF^o(x)}{1 - F^o(x)}} .$$

Entonces el estimador $\hat{S}^0(t)$ de Kaplan-Meier puede expresarse como

$$\hat{S}^0(t) = e^{\int_0^t \frac{dS_n^u(x)}{S_n^u(x) + S_n^c(x)}} = e^{\sum_{x \leq t} \ln \left[\frac{S_n^u(x^+) + S_n^c(x^+)}{S_n^u(x^-) + S_n^c(x^-)} \right]}$$

y ya que la integral en el lado derecho es igual a cero pues $S_n^u(x)$ no es continua entonces,

$$\hat{S}^0(t) = e^{\sum_{x \leq t} \ln \left[\frac{S_n^u(x^+) + S_n^c(x^+)}{S_n^u(x^-) + S_n^c(x^-)} \right]}$$

3.5 MODELOS PARA TIEMPO DE FALLA

Como se dijo anteriormente, el análisis de datos de supervivencia no solo se puede realizar a través de la estimación de la función de supervivencia sino que también se puede hacer si se modela el comportamiento de la función de riesgo o de la función de densidad de probabilidad.

Cuando se tiene una población homogénea y el tiempo de falla es continuo se han propuesto los siguientes modelos paramétricos sobre la distribución de T.

1. Modelo Exponencial.
2. Modelo Weibull.
3. Modelo Log-normal.
4. Modelo Gamma.

Los más comúnmente utilizados son el modelo exponencial (que juega un papel similar en este tipo de análisis al papel que tiene la distribución normal dentro de la inferencia estadística) y el modelo Weibull puesto que estos admiten formas cerradas de las probabilidades de las colas de la distribución y por lo tanto expresiones simples para las funciones de supervivencia y riesgo.

Los otros dos modelos son menos utilizados dados sus inconvenientes computacionales que por lo general se presentan.

Otra vez, sea $T > 0$ una variable aleatoria que representa el tiempo de falla de un individuo, y sea $Y = \text{Log}(T)$.

3.5.1 MODELO EXPONENCIAL

Este modelo se utiliza para el caso en el que la función de riesgo es constante sobre el rango de T , es decir,

$$\lambda(t) = \lambda > 0 \quad t > 0$$

en otras palabras, la tasa instantánea de falla es independiente de t por lo que el cambio condicional de falla en un intervalo de tiempo de longitud específica es el mismo, esto es, la propiedad de pérdida de memoria de la distribución exponencial.

La función de supervivencia está dada por

$$s(t) = e^{-\int_0^t \lambda(s) ds} = e^{-\int_0^t \lambda ds} = e^{-\lambda t}$$

y la función de densidad de probabilidad es

$$f(t) = - \frac{d}{dt} S(t) = \lambda e^{-\lambda t}$$

Por lo tanto, una forma empírica de checar que un modelo exponencial es apropiado para un conjunto de datos de supervivencia es graficar el logaritmo de un estimador de la función de supervivencia contra t ; esta gráfica debe ser aproximadamente una línea recta a través del origen.

3.5.2 MODELO WEIBULL

Este modelo es una importante generalización del modelo exponencial, la cual se obtiene al suponer una fuerte dependencia del riesgo con relación a t . Esta dependencia se refleja a través de 2 parámetros positivos, λ y p , de la siguiente forma:

$$\lambda(t) = \lambda p (\lambda t)^{p-1} \quad t > 0$$

Esta función de riesgo es monótona decreciente para $p < 1$ y monótona creciente para $p > 1$. Se reduce al modelo exponencial cuando $p=1$.

La función de supervivencia es

$$S(t) = e^{-\int_0^t \lambda(s) ds}$$

pero

$$\int_0^t \lambda p (\lambda s)^{p-1} ds = (\lambda t)^p$$

por lo que

$$S(t) = e^{-(\lambda t)^p}$$

y la función de densidad de probabilidad es

$$f(t) = -\frac{d}{dt} S(t) = \lambda p (\lambda t)^{p-1} e^{-(\lambda t)^p}$$

Entonces la forma empírica de comprobar lo apropiado de este modelo, es graficar $\log(-\log S(t))$ contra $\log(t)$; si es correcto este modelo se obtendrá aproximadamente una línea recta cuya pendiente es un estimador de p y la ordenada al origen es un estimador de $-\log(p)$.

3.5.3 MODELO LOG-NORMAL

En este modelo se supone que el comportamiento de T se ajusta a

$$T = e^{\alpha + \beta W} \quad \text{en donde } W \sim N(0,1)$$

y notando que

$$W = \frac{\ln T - \alpha}{\beta} \quad \text{y} \quad \frac{dW}{dT} = \frac{1}{\beta T}$$

entonces la función de densidad de probabilidad de T es

$$f(t) = \frac{1}{\beta t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\ln t - \alpha}{\beta} \right)^2}$$

o bien, denotando

$$\beta = \frac{1}{p} \quad \text{y} \quad \alpha = -\ln \lambda \quad \text{tenemos}$$

$$f(t) = (2\pi)^{-1/2} p t^{-1} e^{-\frac{1}{2} p^2 (\ln(t\lambda))^2}$$

La función de supervivencia es

$$\begin{aligned} S(t) &= \int_t^{\infty} f(s) ds = 1 - \int_0^t f(s) ds \\ &= 1 - \int_0^t \frac{1}{\sqrt{2\pi}} p s^{-1} e^{-\frac{1}{2} p^2 (\ln(s\lambda))^2} ds \end{aligned}$$

Utilizando $r = p \ln(\lambda s)$ se tiene que

$$S(t) = 1 - \int_{-\infty}^{p \ln(\lambda t)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} r^2} dr = 1 - \Phi(p \ln(\lambda t)).$$

La función de riesgo es

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{(2\pi)^{-1/2} p t^{-1} e^{-\frac{1}{2} p^2 (\ln(t\lambda))^2}}{1 - \Phi(p \ln(\lambda t))},$$

la cual vale cero en $t=0$, se incrementa hasta un máximo y después decrece asintóticamente a cero cuando $t \rightarrow \infty$.

Este modelo es útil cuando no hay censura en las observaciones puesto que es simple trabajarlo, pero al existir censura se complica muchísimo la parte computacional.

3.5.4 MODELO GAMMA

Este modelo es otra forma de generalizar el modelo exponencial, en donde T se modela a partir de la relación

$$T = e^{\alpha + W}$$

donde W tiene la función de densidad de probabilidad

$$g(w, k) = \frac{e^{-kw} - e^{-w}}{\Gamma(k)}$$

$$\alpha, k > 0$$

por lo que W tiene una distribución asimétrica negativamente. Esta asimetría decrece al crecer k y cuando k=1 se tiene el modelo exponencial. De las expresiones anteriores es fácil mostrar que

$$f(t) = \frac{1}{t} \frac{1}{\Gamma(k)} e^{-k(\ln t - \alpha)} - e^{-(\ln t - \alpha)}$$

o bien, haciendo

$$d = -\ln \lambda$$

$$f(t) = \frac{1}{t\Gamma(k)} (\lambda t)^k e^{-\lambda t} = \frac{\lambda (\lambda t)^{k-1} e^{-\lambda t}}{\Gamma(k)}$$

La función de supervivencia es

$$\begin{aligned} S(t) &= 1 - \int_0^t f(s) ds \\ &= 1 - \int_0^t \frac{\lambda (\lambda s)^{k-1} e^{-\lambda s}}{\Gamma(k)} ds \end{aligned}$$

y haciendo el cambio de variable $r = \lambda s$

$$\begin{aligned} S(t) &= 1 - \int_0^{\lambda t} \frac{r^{k-1} e^{-r}}{\Gamma(k)} dr \\ &= 1 - I_k(\lambda t) \end{aligned}$$

donde

$$I_k(\lambda t) = \int_0^{\lambda t} \frac{r^{k-1} e^{-r}}{\Gamma(k)} dr$$

La función de riesgo es

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\lambda (\lambda t)^{k-1} e^{-\lambda t}}{\Gamma(k)(1 - I_k(\lambda t))} \quad t > 0,$$

la cual es una función monótona creciente desde 0 para $k > 1$, y monótona decreciente desde ∞ para $k < 1$. En cualquier caso se aproxima a λ cuando $t \rightarrow \infty$.

3.6 MODELOS PARA POBLACIONES HETEROGENEAS

Cuando no se tiene una población homogénea no es posible utilizar los modelos anteriores para hacer un buen análisis de los datos de supervivencia.

Si suponemos que la heterogeneidad de la población se debe a que existen variables explicativas las cuales tienen una dependencia sobre el tiempo de falla, podemos generalizar los modelos anteriores de tal manera que añadimos la información de las variables de cada individuo de la muestra.

Por lo tanto supondremos que para cada individuo se observarán el tiempo de falla T y un vector $z' = (z_1, z_2, \dots, z_p)$ de variables explicativas o covariables. El principal problema será modelar y determinar la relación entre T y z' .

En este caso la función de riesgo al tiempo t para un individuo con covariables z puede modelarse ya sea considerando un modelo aditivo o un modelo multiplicativo.

Si el modelo es aditivo, entonces

$$\lambda(t, z) = \lambda_1(t) + g(z)$$

donde $\lambda_1(t)$ representa el riesgo común a todos los individuos al tiempo t .

Hay que notar que si $g(z)$ es idénticamente 0 entonces se utilizarán los modelos vistos para poblaciones homogéneas.

Si el modelo es multiplicativo, entonces

$$\lambda(t, z) = \lambda_1(t)g(z)$$

aquí también $\lambda_1(t)$ es el riesgo común a todos los individuos al tiempo t .

Como puede verse, estos modelos son demasiado generales pues tanto $\lambda_1(t)$ como $g(z)$ son funciones arbitrarias lo cual hace que un análisis general sea complicado. Para facilitar el análisis se añade el supuesto de que $g(z)$ es una función de z , pero la contribución de este vector es en forma lineal, es decir,

$$g(z) = g(z' \beta)$$

en donde $\beta' = (\beta_1, \beta_2, \dots, \beta_p)$ es un vector de parámetros.

Por lo tanto

$$\lambda(t, z) = \lambda_1(t)g(z' \beta)$$

y la elección de $g(\cdot)$ puede depender de los datos que están siendo considerados. Pero la elección más comúnmente usada es $g(x) = e^x$, por lo que el modelo de regresión para el riesgo será

$$\lambda(t, z) = \lambda_1(t)e^{z' \beta}$$

en donde, como se dijo antes, β es un vector $p \times 1$ de parámetros desconocidos y $\lambda_1(t)$ es una función arbitraria desconocida que da la función de riesgo para un individuo con condiciones estándar $z' = Q'$.

Dentro de los modelos de regresión para la función de riesgo el más simple es aquel en el que se supone $\lambda_1(t)$ constante; con esto se obtiene el llamado modelo de regresión exponencial.

En este caso la función de riesgo será

$$\lambda(t, z) = \lambda e^{z' \beta}$$

con lo cual se especifica que el logaritmo de la tasa de falla al tiempo t es una función lineal de las covariables z .

La función de densidad condicional de T dado z es

$$\begin{aligned} f(t, z) &= \lambda(t, z) e^{-\int_0^t \lambda(s, z) ds} \\ &= \lambda e^{z' \beta} e^{-\int_0^t \lambda e^{z' \beta} ds} \\ &= \lambda e^{z' \beta} e^{-\lambda t e^{z' \beta}} \end{aligned}$$

y la función de supervivencia condicional de T dado z es

$$\begin{aligned}
 S(t; z) &= e^{-\int_0^t \lambda(s, z) ds} \\
 &= e^{-\int_0^t \lambda e^{z/\beta} ds} \\
 &= e^{-\lambda t e^{z/\beta}}
 \end{aligned}$$

El siguiente modelo más simple es considerar una familia doble-paramétrica de las funciones de riesgo, con lo que se obtiene el llamado modelo de regresión Weibull.

Este modelo expresa la función de riesgo condicional al tiempo t dado z como

$$\lambda(t, z) = p(\lambda t)^{p-1} z^\beta e^{-z/\beta}$$

es decir, el efecto que se considera de las covariables es multiplicativo sobre la función de riesgo Weibull.

La función de supervivencia condicional de T dado z es

$$\begin{aligned}
 S(t; z) &= e^{-\int_0^t \lambda(s, z) ds} \\
 &= e^{-\int_0^t p(\lambda s)^{p-1} e^{z/\beta} ds} \\
 &= e^{-t(\lambda t)^{p-1} e^{z/\beta}}
 \end{aligned}$$

y la función de densidad condicional de T dado z es

$$f(t, \underline{z}) = \lambda(t, \underline{z}) S(t, \underline{z})$$

$$= p(\lambda t)^{p-1} e^{\underline{z}'\underline{A} - t} - t(\lambda t)^{p-1} e^{\underline{z}'\underline{A}}$$

4. LA FUNCION DE INFLUENCIA EN DATOS DE SUPERVIVENCIA

4.1 INTRODUCCION

Como ya se ha visto, la curva de influencia está perfectamente definida solo en el caso en el que el parámetro de estudio sea un vector p -dimensional ($p \geq 1$). Por otro lado, cuando se analizan datos de supervivencia, el parámetro de estudio por lo general es una función, la cual puede ser la función de supervivencia, la función de densidad de probabilidad o la función de riesgo. Por lo tanto se presenta un problema al tratar de encontrar la función de influencia al estudiar datos de supervivencia, en particular cuando se usa como estimación de esta el estimador Kaplan-Meier (K-M).

Una solución al problema anterior fue propuesta por Reid (1981), quien basada en la representación del estimador K-M dada por Peterson, encontró la curva de influencia teórica de la función de supervivencia así como la curva de influencia del estimador K-M.

Para calcular la curva de influencia de $S(t)$, se calculó en primer lugar la curva de influencia de la función de riesgo acumulada $\Lambda(t)$, la cual se obtiene a continuación.

La función de riesgo acumulada $\Lambda(t)$, vista como una funcional de S^u y S^c puede expresarse como

$$\Lambda(t) = - \int_0^t \frac{dS^u(y)}{S^u(y) + S^c(y)} - \sum_{x \leq t} \ln \left[\frac{S^u(x^+) + S^c(x^+)}{S^u(x^-) + S^c(x^-)} \right]$$

la función de influencia evaluada en un punto x se calcula de manera puntual para t , es decir, para cada t fija se obtienen 2 curvas de influencia para cada observación con valor x , pues como ya se dijo anteriormente el efecto que tiene la observación x en la función de supervivencia (y en este caso en la función de riesgo acumulada) depende de la naturaleza de la misma, es decir, si es o no una observación censurada.

Si x es una observación no censurada, la curva de influencia está dada por

$$CI_u(\Lambda(t), S^u, S^c, x) = \int_0^{\min(x,t)} \frac{dS^u(y)}{(S^u(y) + S^c(y))^2} + \frac{I\{x \leq t\}}{S^u(x) + S^c(x)},$$

y si x es una observación censurada

$$CI_c(\Lambda(t), S^u, S^c, x) = \int_0^{\min(x,t)} \frac{dS^u(y)}{(S^u(y) + S^c(y))^2}.$$

Para encontrar las curvas de influencia para $S(t)$ se considera a $S(t)$ como una funcional de S^u y S^c a través de la función:

$$s(t) = e^{-\Lambda(t)}$$

por lo que

$$s(t) = e^{\int_0^t \frac{dS^u(y)}{S^u(y) + S^c(y)}} e^{\sum_{x \leq t} \ln \left[\frac{S^u(x^+) + S^c(x^+)}{S^u(x^-) + S^c(x^-)} \right]},$$

de donde se obtiene

$$CI_{\mu}(S(t), S^u, S^c, x) = S(t) \left[\int_0^{\min(x,t)} \frac{dS^u(y)}{(S^u(y) + S^c(y))^2} + \frac{I\{x \leq t\}}{S^u(x) + S^c(x)} \right]$$

y

$$CI_c(S(t), S^u, S^c, x) = S(t) \int_0^{\min(x,t)} \frac{dS^u(y)}{(S^u(y) + S^c(y))^2}.$$

Hay que notar que estas expresiones se tendrán para cada $t > 0$ considerado, por lo que en realidad se tienen una infinidad de funciones de influencia evaluadas en cada observación x de interés, es decir, la influencia que tiene la observación x sobre el valor de $S(t)$ estará en el conjunto

$$\{CI_{\mu}(S(t), S^u, S^c, x), CI_c(S(t), S^u, S^c, x) \text{ con } t > 0\}$$

Finalmente, para obtener las funciones de influencia correspondientes al estimador K-M se utilizan las expresiones anteriores pero considerando las versiones muestrales de cada una de la funciones que aparecen en las mismas; es decir, se sustituye $S(t)$ por $\hat{S}^o(t)$, S^u por $S_n^u(t)$ y $S^c(t)$ por $S_n^c(t)$ donde n es el tamaño de la muestra.

Por lo tanto

$$CI_{\mu}(\hat{S}^o(t), S_n^u, S_n^c, x) = \hat{S}^o(t) \left[\int_0^{\min(x,t)} \frac{dS_n^u(y)}{(S_n^u(y) + S_n^c(y))^2} + \frac{I\{x \leq t\}}{S_n^u(x) + S_n^c(x)} \right]$$

y

$$CI_c(\hat{S}^o(t), S_n^u, S_n^c, x) = \hat{S}^o(t) \int_0^{\min(x,t)} \frac{dS_n^u(y)}{(S_n^u(y) + S_n^c(y))^2} .$$

Con las expresiones anteriores se pueden observar ciertas características del estimador K-M: El término

$$\int_0^{\min(x,t)} \frac{dS_n^u(y)}{(S_n^u(y) + S_n^c(y))^2}$$

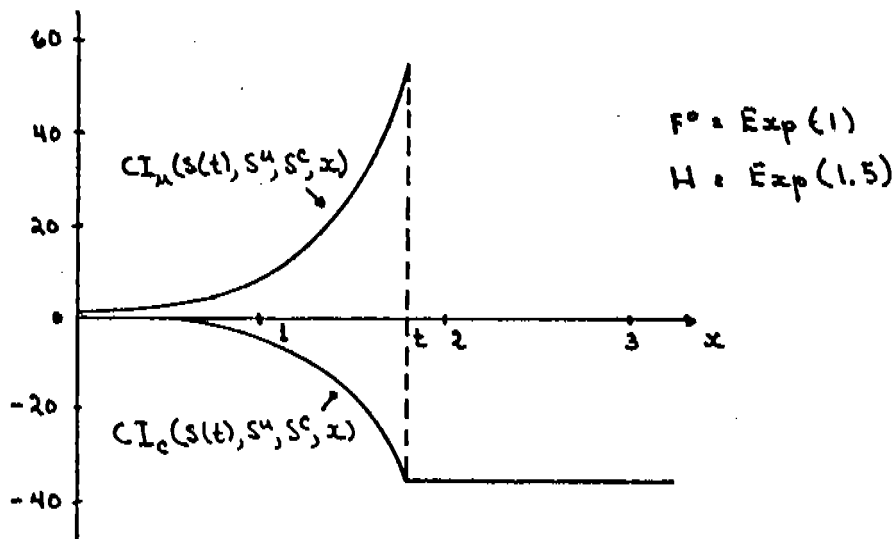
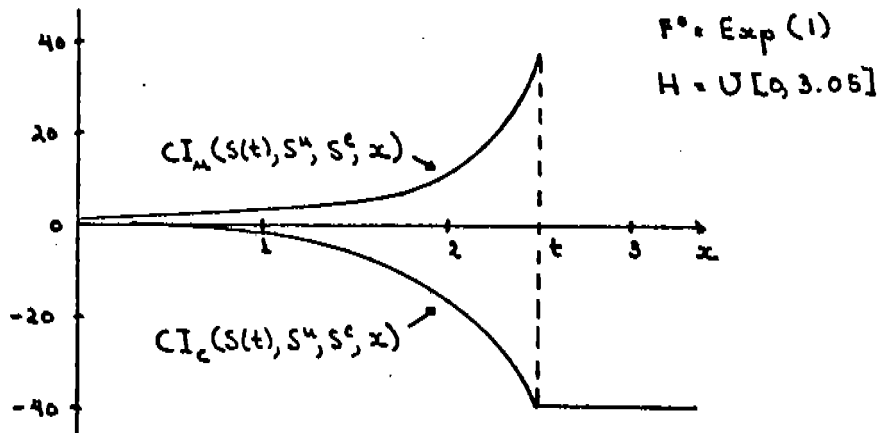
representa esencialmente (pues $\hat{S}^o(t)$ no depende del valor de x) el cambio en el tamaño de cada salto en $\hat{S}^o(t)$ cuando una nueva observación x es añadida a la muestra, y el término

$$\frac{I_{\{x \leq t\}}}{S_n^u(x) + S_n^c(x)}$$

representa el salto adicional que se introduce cuando la nueva observación es no censurada.

Las dos curvas de influencia para $\hat{S}^o(t)$ toman el mismo valor y permanecen ya constantes para $x > t$ puesto que una observación en $x > t$ afecta a $\hat{S}^o(t)$ solo a través de un cambio en el tamaño del salto.

Finalmente, Reid(1981) presenta un ejemplo gráfico de las funciones de influencia teóricas para el caso en el que la función de distribución F^o es exponencial y la distribución de la censura H es uniforme o exponencial. Las gráficas obtenidas se presentan en la siguiente figura.



Una cosa que hay que notar es que esta solución no nos permite analizar la influencia de posibles observaciones contaminantes ya que la construcción de la curva de influencia para el estimador

K-M propuesta por Reid se basa en el supuesto que la muestra aleatoria proviene de una población idénticamente distribuida.

4.2 CURVA DE INFLUENCIA PROPUESTA PARA DATOS DE SUPERVIVENCIA

4.2.1 INTRODUCCION

En la sección anterior se mostró brevemente la solución propuesta por Reid(1979) para el problema de calcular la función de influencia para el estimador K-M. Como puede observarse, el manejo de dicha solución resulta un poco complicado pues la curva de influencia para el estimador K-M se construye de manera puntual, esto es, se tiene un indicador de la influencia de una observación con valor x , ya sea censurada o no, sobre el estimador K-M evaluado en t para cada t fijo.

Por lo anterior surgió la idea de tratar de construir otra solución para el problema en cuestión, de tal manera que esta nueva solución tuviera una forma simple de cálculo, que nos llevara a interpretaciones sencillas y con la cual se obtuviera un indicador de la influencia de una observación (censurada o no) sobre el estimador K-M de forma global y no de manera puntual.

Por otro lado, al tratar de medir una influencia global parece natural tratar de encontrar, en primer lugar, alguna medida (función de la muestra) que sea sensible a la forma del estimador K-M y, posteriormente, contruir la función de influencia en base a dicha medida.

La medida buscada debe ser tal que refleje el comportamiento global del estimador K-M, esto es, que con este valor se pueda

determinar la existencia o ausencia de saltos muy grandes en el estimador K-M o, equivalentemente, la de intervalos en los que exista mucha censura; también, con este valor se debe poder determinar la presencia de observaciones contaminantes, es decir, observaciones muy grandes y/o muy pequeñas con respecto a las que esperamos.

Una medida que cumple con las características deseadas es el llamado valor promedio (VP) del estimador K-M; dicho concepto se describirá en la siguiente sección. Así, la solución estará basada en VP y permitirá medir la influencia que tienen sobre el estimador K-M las observaciones muestrales tanto censuradas como no censuradas, las cuales a su vez pueden provenir de una población contaminada.

4.2.2 VALOR PROMEDIO DEL ESTIMADOR KAPLAN-MEIER

Supongamos que se tiene una muestra aleatoria Y_1, Y_2, \dots, Y_n la cual proviene de una población contaminada, es decir,

$$Y_i \sim (1-p_o)F + p_o F'$$

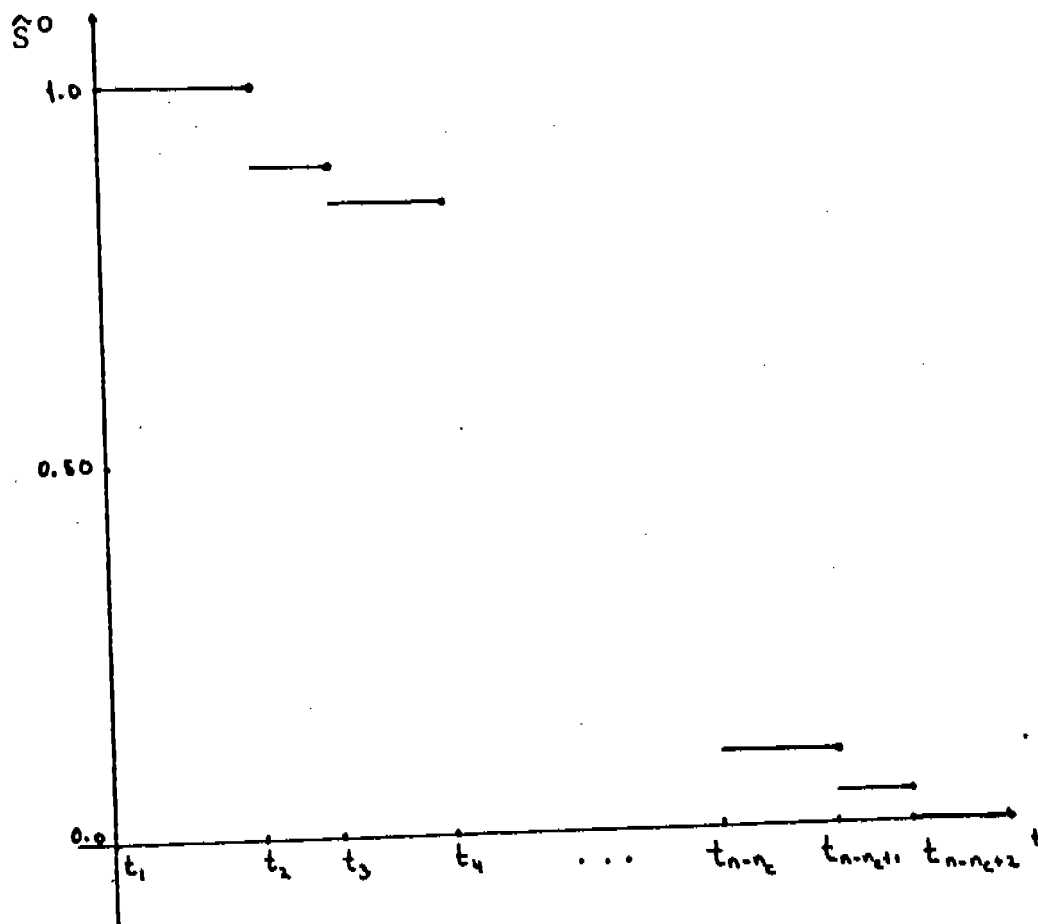
donde p_o es la proporción de contaminación, y F y F' son funciones de distribución de tiempos de falla.

Supongamos que la censura está distribuida uniformemente y que el porcentaje observado de valores censurados en la muestra es $(p_c \times 100)\%$. Denotemos además por $n_c = n \cdot p_c$ el número de observaciones censuradas.

Por comodidad tomemos el origen temporal como $t_1 = 0$ y denotemos

por $t_2, t_3, \dots, t_{n-n+1}$ los tiempos de falla observados. Con el objeto de eliminar la posible indeterminación del estimador K-M para x mayores que la última observación si esta fuera censurada, introduciremos una observación ficticia en $t_{n-n+2} = Y_{(n)} + h$, donde h tiene un valor positivo, pequeño y arbitrario escogido por el investigador. Ya que una nueva falla mayor que $Y_{(n)}$ probablemente habría ocurrido dentro de una desviación estándar posterior al valor de $Y_{(n)}$, es razonable hacer h igual a una desviación estándar de la distribución estudiada (es decir, de F) por lo que en este trabajo así se tomó. En la práctica F no es conocida y h se tiene que fijar de alguna otra forma, posiblemente utilizando información adicional acerca del fenómeno. Entonces, independientemente de $Y_{(n)}$ sea censurada o no, el estimador valdrá cero a partir del valor $Y_{(n)} + h$, aunque si $Y_{(n)}$ es no censurada el valor de cero empezará precisamente en $Y_{(n)} = t_{n-n+1}$.

Con los tiempos de falla anteriores se calcula el estimador K-M de la función de supervivencia, el cual se muestra gráficamente en la siguiente figura



Sean $D_i = \hat{S}^0(t_i) - \hat{S}^0(t_{i+1})$ $i=1,2,\dots,n-n_c+1$ las magnitudes de los saltos del estimador. Por lo tanto, el area bajo la curva es

$$\begin{aligned} \text{Area} &= (t_2 - t_1) \hat{S}^0(t_1) + (t_3 - t_2) \hat{S}^0(t_2) + \dots \\ &+ (t_{n-n_c+1} - t_{n-n_c}) \hat{S}^0(t_{n-n_c}) \\ &+ (t_{n-n_c+2} - t_{n-n_c+1}) \hat{S}^0(t_{n-n_c+1}) \end{aligned}$$

$$\begin{aligned}
&= t_2 (\hat{S}^{\circ}(t_1) - \hat{S}^{\circ}(t_2)) + t_3 (\hat{S}^{\circ}(t_2) - \hat{S}^{\circ}(t_3)) \\
&+ \dots \\
&+ t_{n-n_c+1} (\hat{S}^{\circ}(t_{n-n_c}) - \hat{S}^{\circ}(t_{n-n_c+1})) \\
&+ t_{n-n_c+2} \hat{S}^{\circ}(t_{n-n_c+1}) \\
&= t_2 D_1 + t_3 D_2 + \dots + t_{n-n_c+1} D_{n-n_c} + t_{n-n_c+2} D_{n-n_c+1} \\
&= \sum_{i=1}^{n-n_c+1} t_{i+1} D_i
\end{aligned}$$

por lo que definimos el valor promedio (VP) del estimador K-M para el caso en el que hay censura

$$VP = \begin{cases} \frac{\text{Area}}{t_{n-n_c+2}} = \sum_{i=1}^{n-n_c+1} \frac{t_{i+1} D_i}{t_{n-n_c+2}} & \text{si } Y(n) \text{ censurada} \\ \frac{\text{Area}}{t_{n-n_c+1}} = \sum_{i=1}^{n-n_c+1} \frac{t_{i+1} D_i}{t_{n-n_c+1}} & \text{si } Y(n) \text{ no censurada} \end{cases}$$

mientras que para el caso en el que no hay censura ($n_c = 0$), el

estimador K-M se transforma en el "complemento" a 1 de la función de distribución empírica de la muestra, y se tendrá entonces (ver la siguiente gráfica)

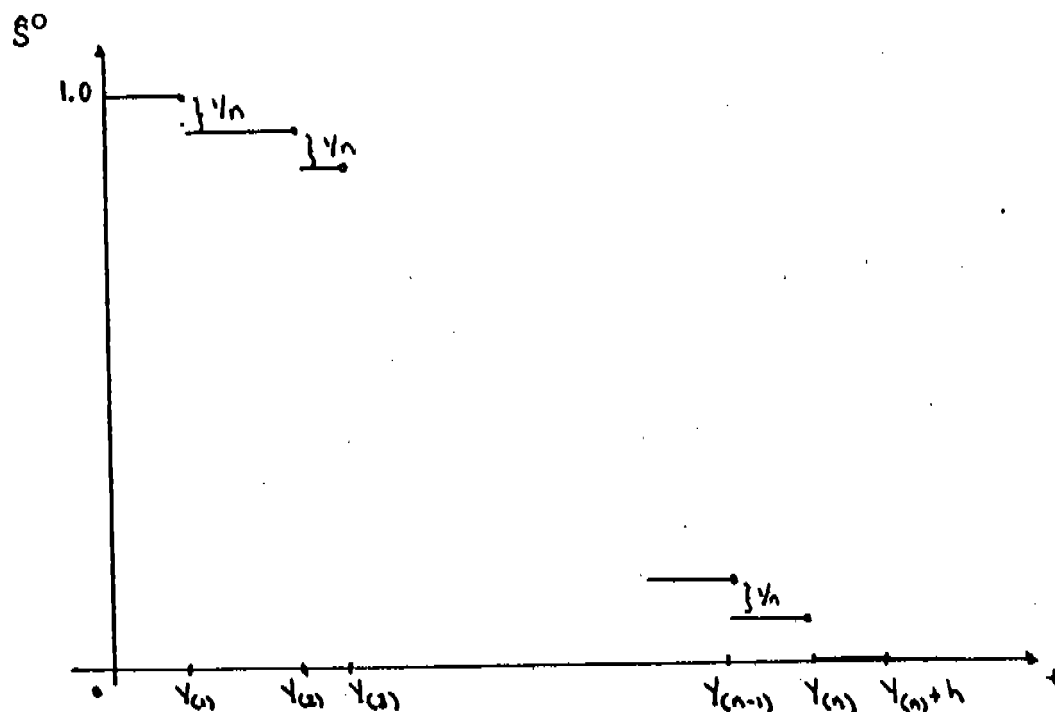
$$t = 0, \quad t = Y_i \quad i=2,3,\dots,n+1,$$

$$t_{n+2} = Y_{(n)} + h,$$

$$D_i = 1/n \quad i=1,\dots,n$$

y

$$D_{n+1} = 0,$$



por lo que el area bajo la curva es

$$\begin{aligned}
 \text{Area} &= t_2 D_1 + t_3 D_2 + \dots + t_{n+1} D_n + t_{n+2} D_{n+1} \\
 &= Y_{(1)} \left(\frac{1}{n}\right) + Y_{(2)} \left(\frac{1}{n}\right) + \dots + Y_{(n)} \left(\frac{1}{n}\right) + (Y_{(n)} + h)(0) \\
 &= \sum_{i=1}^n Y_{(i)} \left(\frac{1}{n}\right) = \sum_{i=1}^n \frac{Y_{(i)}}{n} = \bar{Y}
 \end{aligned}$$

y el valor promedio es

$$VP = \frac{\bar{Y}}{Y_{(n)}}$$

Entonces, independientemente de las características de la muestra, se podrá calcular el valor promedio (VP) del estimador K-M.

Ahora bien, una forma de analizar el comportamiento global del estimador K-M en una muestra particular, es comparar el VP obtenido con respecto a su esperanza bajo las condiciones de la muestra en cuestión. Por lo anterior es necesario encontrar la esperanza de VP en presencia o ausencia de contaminación. Discutiremos a continuación el caso mas sencillo, en el que no existe censura en la muestra, y en el que las funciones F y F' corresponden a distribuciones exponenciales. Sean

$$Y_i \sim (1-p_0) \text{Exp}(\lambda) + p_0 \text{Exp}(\theta) \quad i=1, \dots, n$$

en donde λ es el parámetro de interés y θ es el parámetro de contaminación.

La esperanza de VP está dada por

$$E(VP) = E\left(\frac{\bar{Y}}{Y_{(n)}}\right) \approx \frac{E(\bar{Y})}{E(Y_{(n)})}$$

Ahora bien,

$$f_{Y_i}(x) \equiv f(x) = \begin{cases} (1-p_0)\lambda e^{-\lambda x} + p_0\theta e^{-\theta x} & x, \lambda, \theta > 0 \\ 0 & 0 \leq p_0 \leq 1 \\ & \text{en otro caso} \end{cases}$$

y por lo tanto

$$\begin{aligned} E(\bar{Y}) = E(Y_i) &= (1-p_0) \int_0^{\infty} \lambda x e^{-\lambda x} dx + p_0 \int_0^{\infty} \theta x e^{-\theta x} dx \\ &= \frac{1-p_0}{\lambda} + \frac{p_0}{\theta} \end{aligned}$$

Para obtener $E(Y_{(n)})$ se hará por separado para el caso que no haya contaminación y para el caso que si la haya.

Caso 1.- $Y_i \sim \text{Exp}(\lambda)$

En este caso

$$f(y) = \begin{cases} \lambda e^{-\lambda y} & y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$F(y) = \begin{cases} 0 & y \leq 0 \\ 1 - e^{-\lambda y} & y > 0 \end{cases},$$

por lo que

$$\begin{aligned} f_{Y(n)}(y) &\equiv f_n(y) = n (F(y))^{n-1} f(y) \\ &= n \lambda (1 - e^{-\lambda y})^{n-1} e^{-\lambda y}. \end{aligned}$$

Entonces

$$\begin{aligned} E(Y(n)) &= \int_0^{\infty} y n \lambda (1 - e^{-\lambda y})^{n-1} e^{-\lambda y} dy \\ &= n \lambda \int_0^{\infty} y (1 - e^{-\lambda y})^{n-1} e^{-\lambda y} dy, \end{aligned}$$

pero

$$(1 - x)^{n-1} = \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i x^i,$$

por lo que

$$\begin{aligned} E(Y(n)) &= n \lambda \int_0^{\infty} y \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i (e^{-\lambda y})^i e^{-\lambda y} dy \\ &= n \lambda \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \int_0^{\infty} y e^{-\lambda(i+1)y} dy \\ &= n \lambda \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \frac{1}{\lambda^2 (i+1)^2} \\ &= \frac{n}{\lambda} \sum_{i=0}^{n-1} \binom{n-1}{i} \frac{(-1)^i}{(i+1)^2} \end{aligned}$$

$$\begin{aligned}
&= \frac{n}{\lambda} \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-1-k)!} \frac{(-1)^k}{(k+1)^2} \\
&= \frac{1}{\lambda} \sum_{k=0}^{n-1} \frac{n!}{(k+1)!(n-(k+1))!} \frac{(-1)^k}{(k+1)} \\
&= \frac{1}{\lambda} \sum_{k=0}^{n-1} \binom{n}{k+1} \frac{(-1)^k}{k+1} = \frac{1}{\lambda} \sum_{k=1}^n \binom{n}{k} \frac{(-1)^{k+1}}{k} \\
&= \frac{1}{\lambda} \sum_{k=1}^n \frac{1}{k} ,
\end{aligned}$$

la última igualdad es demostrada en el apéndice 3.

Caso 2.- $Y_i \sim (1 - p_0)\text{Exp}(\lambda) + p_0\text{Exp}(\theta)$
 En este caso

$$f(y) = \begin{cases} (1-p_0)\lambda e^{-\lambda y} + p_0\theta e^{-\theta y} & y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

y, denotando $q_0 = 1 - p_0$, tenemos que

$$\begin{aligned}
F(y) &= \int_0^y (q_0\lambda e^{-\lambda x} + p_0\theta e^{-\theta x}) dx \\
&= q_0(-e^{-\lambda x}|_0^y) + p_0(-e^{-\theta x}|_0^y) \\
&= q_0(1 - e^{-\lambda y}) + p_0(1 - e^{-\theta y}) \\
&= q_0 - q_0 e^{-\lambda y} + p_0 - p_0 e^{-\theta y} \\
&= 1 - (q_0 e^{-\lambda y} + p_0 e^{-\theta y}).
\end{aligned}$$

Por lo tanto,

$$\begin{aligned}
 E(Y_{(n)}) &= \int_0^{\infty} n y [1 - (q_0 e^{-\lambda y} + p_0 e^{-\theta y})]^{n-1} [\lambda q_0 e^{-\lambda y} + \theta p_0 e^{-\theta y}] dy \\
 &= n \lambda q_0 \int_0^{\infty} y [1 - (q_0 e^{-\lambda y} + p_0 e^{-\theta y})]^{n-1} e^{-\lambda y} dy \\
 &\quad + n \theta p_0 \int_0^{\infty} y [1 - (q_0 e^{-\lambda y} + p_0 e^{-\theta y})]^{n-1} e^{-\theta y} dy.
 \end{aligned}$$

Ahora bien,

$$\begin{aligned}
 [1 - (q_0 e^{-\lambda y} + p_0 e^{-\theta y})]^{n-1} &= \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i (q_0 e^{-\lambda y} + p_0 e^{-\theta y})^i \\
 &= \binom{n-1}{0} (-1)^0 (q_0 e^{-\lambda y} + p_0 e^{-\theta y})^0 \\
 &\quad + \binom{n-1}{1} (-1)^1 (q_0 e^{-\lambda y} + p_0 e^{-\theta y})^1 \\
 &\quad + \binom{n-1}{2} (-1)^2 (q_0 e^{-\lambda y} + p_0 e^{-\theta y})^2 \\
 &\quad + \binom{n-1}{3} (-1)^3 (q_0 e^{-\lambda y} + p_0 e^{-\theta y})^3 \\
 &\quad + \dots \\
 &\quad + \binom{n-1}{n-2} (-1)^{n-2} (q_0 e^{-\lambda y} + p_0 e^{-\theta y})^{n-2} \\
 &\quad + \binom{n-1}{n-1} (-1)^{n-1} (q_0 e^{-\lambda y} + p_0 e^{-\theta y})^{n-1}
 \end{aligned}$$

$$\begin{aligned}
&= 1 - \binom{n-1}{1} (q_0 e^{-\lambda y} + p_0 e^{-\theta y}) \\
&+ \binom{n-1}{2} (q_0^2 e^{-2\lambda y} + 2 p_0 q_0 e^{-(\lambda+\theta)y} + p_0^2 e^{-2\theta y}) \\
&- \binom{n-1}{3} (q_0^3 e^{-3\lambda y} + 3 q_0^2 p_0 e^{-(2\lambda+\theta)y} + 3 q_0 p_0^2 e^{-(\lambda+2\theta)y} \\
&\quad + p_0^3 e^{-3\theta y})
\end{aligned}$$

+ ...

$$+ (-1)^{n-1} \binom{n-1}{n-1} (q_0^{n-1} e^{-(n-1)\lambda y} + (n-1) q_0^{n-2} p_0 e^{-(n-2)\lambda+\theta} y)$$

$$+ \binom{n-1}{2} q_0^{n-3} p_0^2 e^{-(n-3)\lambda+2\theta} y + \dots$$

$$= \left[1 - \binom{n-1}{1} q_0 e^{-\lambda y} + \binom{n-1}{2} q_0^2 e^{-2\lambda y} - \binom{n-1}{3} q_0^3 e^{-3\lambda y} \right. \\
\left. + \dots + (-1)^{n-1} \binom{n-1}{n-1} q_0^{n-1} e^{-(n-1)\lambda y} \right]$$

$$+ \left[- \binom{n-1}{1} p_0 e^{-\theta y} + \binom{n-1}{2} 2 p_0 q_0 e^{-(\lambda+\theta)y} \right. \\
\left. + \dots + (-1)^{n-1} \binom{n-1}{n-1} (n-1) q_0 p_0^{n-2} e^{-(n-2)\lambda+\theta} y \right]$$

$$+ \left[\binom{n-1}{2} p_0^2 e^{-2\theta y} - \binom{n-1}{3} \binom{3}{2} p_0^2 q_0 e^{-(\lambda+2\theta)y} \right. \\
\left. + \dots + (-1)^{n-1} \binom{n-1}{n-1} \binom{n-1}{2} p_0^2 q_0^{n-3} e^{-(n-3)\lambda+2\theta} y \right]$$

+ ...

$$= \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i q_0^i e^{-i\lambda y} \quad (A)$$

+

$$+ \sum_{i=1}^{n-1} \binom{n-1}{i} (-1)^i i q_0^{i-1} p_0 e^{-(i-1)\lambda + \theta} y \quad (b)$$

$$+ \sum_{i=2}^{n-1} \binom{n-1}{i} (-1)^i \binom{i}{2} q_0^{i-2} p_0^2 e^{-(i-2)\lambda + 2\theta} y \quad (c)$$

+ ...

Entonces, si denotamos por I_1 el primer término de la expresión para $E(Y_{(n)})$,

$$I_1 = n \lambda q_0 \int_0^{\infty} y [1 - (q_0 e^{-\lambda y} + p_0 e^{-\theta y})]^{n-1} e^{-\lambda y} dy$$

$$\approx n \lambda q_0 \int_0^{\infty} y [A + B + C] e^{-\lambda y} dy.$$

Para evaluar esta expresión tenemos

$$\int_0^{\infty} y A e^{-\lambda y} dy = \int_0^{\infty} y \left[\sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i q_0^i e^{-i\lambda y} \right] e^{-\lambda y} dy$$

$$= \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i q_0^i \int_0^{\infty} y e^{-\lambda(i+1)y} dy$$

$$= \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i q_0^i \left(\frac{1}{\lambda(i+1)} \right)^2,$$

pero

$$\binom{n-1}{i} \left(\frac{1}{i+1} \right) = \frac{(n-1)!}{i!(n-1-i)!} \cdot \frac{1}{i+1} = \frac{n!}{n(i+1)!(n-i-1)!} = \frac{1}{n} \binom{n}{i+1},$$

por lo que

$$\int_0^{\infty} y A e^{-\lambda y} dy = \frac{1}{n \lambda^2} \sum_{i=0}^{n-1} \binom{n}{i+1} \frac{(-1)^i q_0^i}{i+1} = \frac{1}{n \lambda^2} \sum_{i=1}^n \binom{n}{i} \frac{(-1)^{i-1} q_0^{i-1}}{i}$$

$$= \frac{1}{n \lambda^2 q_0} \sum_{i=1}^n \binom{n}{i} \frac{(-1)^{i-1} q_0^i}{i}.$$

Por otro lado,

$$\begin{aligned} \int_0^{\infty} y B e^{-\lambda y} dy &= \int_0^{\infty} y \left[\sum_{i=1}^{n-1} \binom{n-1}{i} (-1)^i i q_0^{i-1} p_0 e^{-(i-1)\lambda + \theta} y \right] dy \\ &= \sum_{i=1}^{n-1} \binom{n-1}{i} (-1)^i i q_0^{i-1} p_0 \int_0^{\infty} y e^{-(i\lambda + \theta) y} dy \\ &= \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i i q_0^{i-1} p_0}{(i\lambda + \theta)^2} \\ &= \frac{p_0}{q_0} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i i q_0^i}{(i\lambda + \theta)^2}, \end{aligned}$$

$$\begin{aligned} \int_0^{\infty} y C e^{-\lambda y} dy &= \int_0^{\infty} y \left[\sum_{i=2}^{n-1} \binom{n-1}{i} (-1)^i \binom{i}{2} q_0^{i-2} p_0^2 e^{-(i-2)\lambda + 2\theta} y \right] e^{-\lambda y} dy \\ &= \sum_{i=2}^{n-1} \binom{n-1}{i} (-1)^i \binom{i}{2} q_0^{i-2} p_0^2 \int_0^{\infty} y e^{-(i-1)\lambda + 2\theta} y} dy \\ &= \sum_{i=2}^{n-1} \binom{n-1}{i} (-1)^i \binom{i}{2} q_0^{i-2} p_0^2 \left(\frac{1}{(i-1)\lambda + 2\theta} \right)^2 \\ &= \frac{1}{2} \left(\frac{p_0}{q_0} \right)^2 \sum_{i=2}^{n-1} \binom{n-1}{i} \frac{(-1)^i i(i-1) q_0^i}{((i-1)\lambda + 2\theta)^2} \end{aligned}$$

Entonces

$$\begin{aligned} I_1 &\approx n\lambda q_0 \left[\frac{1}{n\lambda^2 q_0} \sum_{i=1}^n \binom{n}{i} \frac{(-1)^{i-1} q_0^i}{i} \right. \\ &\quad + \frac{p_0}{q_0} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i i q_0^i}{(i\lambda + \theta)^2} \\ &\quad \left. + \frac{1}{2} \left(\frac{p_0}{q_0} \right)^2 \sum_{i=2}^{n-1} \binom{n-1}{i} \frac{(-1)^i i(i-1) q_0^i}{((i-1)\lambda + 2\theta)^2} \right] \\ &= \frac{1}{\lambda} \sum_{i=1}^n \binom{n}{i} \frac{(-1)^{i-1} q_0^i}{i} + n\lambda p_0 \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i i q_0^i}{(i\lambda + \theta)^2} \\ &\quad + \frac{n\lambda p_0^2}{2q_0} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i i(i-1) q_0^i}{((i-1)\lambda + 2\theta)^2} \end{aligned}$$

Sea I_2 el segundo término de la expresión para $E(Y_{(n)})$.

Entonces

$$I_2 = n\theta p_0 \int_0^{\infty} y [1 - (q_0 e^{-\lambda y} + p_0 e^{-\theta y})]^{n-1} e^{-\theta y} dy$$

$$\approx n\theta p_0 \int_0^{\infty} y [A + B] e^{-\theta y} dy,$$

En este caso, tenemos que

$$\int_0^{\infty} y A e^{-\theta y} dy = \int_0^{\infty} y \left[\sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i q_0^i e^{-i\lambda y} \right] e^{-\theta y} dy$$

$$= \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i q_0^i \int_0^{\infty} y e^{-(i\lambda + \theta)y} dy$$

$$= \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i q_0^i \left(\frac{1}{i\lambda + \theta} \right)^2 = \sum_{i=0}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(i\lambda + \theta)^2}$$

y

$$\int_0^{\infty} y B e^{-\theta y} dy = \int_0^{\infty} y \left[\sum_{i=1}^{n-1} \binom{n-1}{i} (-1)^i i q_0^{i-1} p_0 e^{-(i-1)\lambda y} \right] e^{-\theta y} dy$$

$$= \sum_{i=1}^{n-1} \binom{n-1}{i} (-1)^i i q_0^{i-1} p_0 \int_0^{\infty} y e^{-((i-1)\lambda + \theta)y} dy$$

$$= \sum_{i=1}^{n-1} \binom{n-1}{i} (-1)^i i q_0^{i-1} p_0 \left(\frac{1}{(i-1)\lambda + \theta} \right)^2$$

$$= \frac{p_0}{q_0} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i i q_0^i}{((i-1)\lambda + \theta)^2}$$

por lo que

$$I_2 = n\theta p_0 \left[\sum_{i=0}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(i\lambda + \theta)^2} + \frac{p_0}{q_0} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i i q_0^i}{((i-1)\lambda + \theta)^2} \right]$$

$$= n\theta p_0 \sum_{i=0}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(i\lambda + \theta)^2} + \frac{n p_0}{\theta} + \frac{n\theta p_0^2}{q_0} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i i q_0^i}{((i-1)\lambda + \theta)^2}$$

Por lo tanto,

$$E(Y_{(n)}) = I_1 + I_2$$

$$= \frac{1}{\lambda} \sum_{i=1}^n \binom{n}{i} \frac{(-1)^{i-1} q_0^i}{i} + \frac{n p_0}{\theta}$$

$$+ n p_0 \left[\sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i (i\lambda) q_0^i}{(i\lambda + \theta)^2} + \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i \theta q_0^i}{(i\lambda + \theta)^2} \right]$$

$$+ \frac{n p_0^2}{2 q_0} \left[\sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i i (i-1)\lambda q_0^i}{((i-1)\lambda + 2\theta)^2} + \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i i (2\theta) q_0^i}{((i-1)\lambda + 2\theta)^2} \right]$$

$$+ \dots$$

$$= \frac{1}{\lambda} \sum_{i=1}^n \binom{n}{i} \frac{(-1)^{i-1} q_0^i}{i} + \frac{n p_0}{\theta} + n p_0 \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{i\lambda + \theta}$$

$$+ \frac{n p_0^2}{2 q_0} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i i q_0^i}{((i-1)\lambda + 2\theta)} + \dots$$

$$= \frac{1}{\lambda} \sum_{i=1}^n \left(\frac{1-p_0^i}{i} \right) + \frac{n p_0}{\theta} - \frac{n p_0}{\lambda} \sum_{i=1}^{n-1} \left(\frac{1-p_0^i}{i} \right)$$

$$+ \frac{n p_0 \theta}{\lambda^2} \sum_{i=1}^{n-1} \frac{1}{i} \sum_{r=1}^i \left(\frac{1-p_0^r}{r} \right) + \frac{n p_0 \theta^2}{\lambda^3} \sum_{i=1}^{n-1} \frac{1}{i} \sum_{r=1}^i \frac{1}{r} \sum_{s=1}^r \left(\frac{1-p_0^s}{s} \right)$$

$$+ \frac{n p_0 \theta^3}{\lambda^3} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(i\lambda + \theta) i^3}$$

$$+ \frac{n p_0^2}{2 q_0} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i i q_0^i}{((i-1)\lambda + 2\theta)}$$

Esta última igualdad se obtiene simplificando cada uno de los tres términos considerados; dichas simplificaciones se presentan en el apéndice 4.

Entonces, para el caso de no contaminación tendremos que

$$E(VP) \approx \frac{1/\lambda}{1/\lambda \sum_{i=1}^n \frac{1}{x_i}} = \frac{1}{\sum_{i=1}^n \frac{1}{x_i}},$$

y para el caso de contaminación

$$E(VP) \approx \frac{\frac{1-p_0}{\lambda} + \frac{p_0}{\theta}}{\left[\frac{1}{\lambda} \sum_{i=1}^n \left(\frac{1-p_0^i}{x_i} \right) + \frac{n p_0}{\theta} - \frac{n p_0}{\lambda} \sum_{i=1}^{n-1} \left(\frac{1-p_0^i}{x_i} \right) \right.}$$

$$\left. + \frac{n p_0 \theta}{\lambda^2} \sum_{i=1}^{n-1} \frac{1}{x_i} \sum_{r=1}^i \left(\frac{1-p_0^r}{r} \right) \right.$$

$$\left. + \frac{n p_0 \theta^2}{\lambda^3} \sum_{i=1}^{n-1} \frac{1}{x_i} \sum_{r=1}^i \frac{1}{r} \sum_{s=1}^r \left(\frac{1-p_0^s}{s} \right) + \dots \right]$$

La expresión anterior se puede aproximar en la práctica tomando únicamente los términos indicados en el denominador. Para una p_0 fija esta aproximación será mejor para λ / θ grande.

Por otro lado se analizó la bondad de esta aproximación con relación a $E(VP)$ (es decir $E(\bar{Y}/Y^{(n)})$), para lo cual se realizaron simulaciones de las variables Y_1, \dots, Y_n para diversos valores de n, λ, θ y p .

Con estas simulaciones se estudió además el efecto de la contaminación en el estimador K-M y se observó que cuando la contaminación es por la izquierda (es decir, cuando $\theta > \lambda$, lo cual implica que las observaciones contaminantes son muy pequeñas con relación a las esperadas), esta no afecta esencialmente a la forma del estimador K-M, mientras que si la contaminación es por la derecha ($\theta < \lambda$), el estimador se verá afectado fuertemente ya que cuando la diferencia $\lambda - \theta$ se incrementa se tenderá a observar una cola larga en el estimador con lo cual aparentemente se tendría un promedio de vida mayor. Por lo anterior, los desarrollos y situaciones estudiadas solo consideran contaminación por la derecha.

Con estas simulaciones se produjeron 1000 muestras para cada una de las 60 combinaciones de los siguientes (para cada λ solo se utilizaron los valores de θ que se encuentran en el mismo renglón, y para $p = 0$ el parámetro θ no interviene):

n	λ	θ	$(p \times 100)\%$
20,40,60	0.10	0.050,0.020	0.5,10
	0.05	0.025,0.010	
	10.00	5.000,2.000	
	20.00	10.000,4.000	

Los resultados que se obtuvieron fueron aproximando el denominador de la última expresión por los 5 primeros términos.

n=20 $\lambda = 0.10$

θ	p o	E(Y)	E(Y) (n)		E(VP)	
			Simul.	Aprox.	Simul.	Aprox.
-----	0	10	36.020	35.977	0.2776	0.2779
0.05	0.05	10.5	39.841	39.556	0.2636	0.2655
0.05	0.10	11	43.262	43.332	0.2543	0.2539
0.02	0.05	12	62.884	61.043	0.1908	0.1966
0.02	0.10	14	81.696	86.639	0.1714	0.1616

n=20 $\lambda = 0.05$

θ	p o	E(Y)	E(Y) (n)		E(VP)	
			Simul.	Aprox.	Simul.	Aprox.
-----	0	20	72.040	71.955	0.2776	0.2779
0.025	0.05	21	79.682	79.113	0.2636	0.2655
0.025	0.10	22	86.523	86.663	0.2543	0.2539
0.010	0.05	24	125.769	122.087	0.1908	0.1966
0.010	0.10	28	163.391	173.277	0.1714	0.1616

n=20 $\lambda = 10$

θ	p o	E(Y)	E(Y) (n)		E(VP)	
			Simul.	Aprox.	Simul.	Aprox.
-----	0	0.1	0.3602	0.3598	0.2776	0.2779
5	0.05	0.105	0.3984	0.3956	0.2636	0.2655
5	0.10	0.11	0.4326	0.4333	0.2543	0.2539
2	0.05	0.12	0.6288	0.6104	0.1908	0.1966
2	0.10	0.14	0.8170	0.8664	0.1714	0.1616

n=20

 $\lambda = 20$

θ	p o	E(Y)	E(Y (n))		E(VP)	
			Simul.	Aprox.	Simul.	Aprox.
----	0	0.05	0.1801	0.1799	0.2776	0.2779
10	0.05	0.0525	0.1992	0.1978	0.2636	0.2655
10	0.10	0.055	0.2163	0.2167	0.2543	0.2539
5	0.05	0.0575	0.3013	0.2925	0.1908	0.1966
5	0.10	0.065	0.3793	0.4022	0.1714	0.1616

Como puede verse, la esperanza de VP no depende finalmente de los valores particulares de λ y θ sino de la relación que existe entre ellos, más explícitamente de la razón λ/θ . Lo anterior se justifica por el hecho que el valor esperado de una función no depende de la escala de medición de la variable sino de los valores particulares de dicha función, y en este caso la escala de medición está referida a λ y θ . Además, regresando a las expresiones de E(VP) se tiene:

Para el caso de no contaminación

$$E(VP) = \frac{1}{\sum_{i=1}^n \frac{1}{i}},$$

que evidentemente no depende de λ .

Para el caso de contaminación

$$E(VP) \approx \frac{\frac{1-p_0}{\lambda} + \frac{p_0}{\theta}}{\left[\frac{1}{\lambda} \sum_{i=1}^n \left(\frac{1-p_0^i}{\lambda} \right) + \frac{np_0}{\theta} - \frac{np_0}{\lambda} \sum_{i=1}^{n-1} \left(\frac{1-p_0^i}{\lambda} \right) \right.}$$

$$+ \frac{np_0\theta}{\lambda^2} \sum_{i=1}^{n-1} \frac{1}{i} \sum_{r=1}^i \left(\frac{1-p_0^r}{r} \right)$$

$$\left. + \frac{np_0\theta^2}{\lambda^2} \sum_{i=1}^{n-1} \frac{1}{i} \sum_{r=1}^i \sum_{s=1}^r \left(\frac{1-p_0^s}{s} \right) \right]$$

multiplicando numerador y denominador por θ

$$E(VP) \approx \frac{(1-p_0) \left(\frac{\theta}{\lambda} \right) + p_0}{\left[\left(\frac{\theta}{\lambda} \right) \sum_{i=1}^n \left(\frac{1-p_0^i}{\lambda} \right) + np_0 - np_0 \left(\frac{\theta}{\lambda} \right) \sum_{i=1}^{n-1} \left(\frac{1-p_0^i}{\lambda} \right) \right.}$$

$$+ np_0 \left(\frac{\theta}{\lambda} \right)^2 \sum_{i=1}^{n-1} \frac{1}{i} \sum_{r=1}^i \left(\frac{1-p_0^r}{r} \right)$$

$$\left. + np_0 \left(\frac{\theta}{\lambda} \right)^3 \sum_{i=1}^{n-1} \frac{1}{i} \sum_{r=1}^i \sum_{s=1}^r \left(\frac{1-p_0^s}{s} \right) \right]$$

por lo que solo depende de λ/θ . Por lo anterior, las tablas correspondientes a tamaños de muestra 40 y 60 solo se presentarán en términos de la relación entre λ y θ (en todos los casos, la relación λ/θ es mayor que 1). Obteniéndose

n=40

λ/θ	p_0	E(VP)	
		Simul.	Aprox.
-----	0	0.2335	0.2337
2	0.05	0.2180	0.2205
2	0.10	0.2095	0.2054
5	0.05	0.1461	0.1387
5	0.10	0.1295	0.1064

n=60

λ/θ	p_0	E(VP)	
		Simul.	Aprox.
-----	0	0.2127	0.2137
2	0.05	0.1974	0.1946
2	0.10	0.1874	0.1821
5	0.05	0.1250	0.1113
5	0.10	0.1107	0.0933

Como puede apreciarse en las tablas obtenidas, al utilizar las expresiones propuestas para E(VP) se tienen muy buenas aproximaciones al valor obtenido por simulación, el cual a su vez se espera que sea semejante al verdadero valor. Por lo tanto, se pueden utilizar estas expresiones, por lo menos, en muestras con un tamaño entre 20 y 60, aunque personalmente no creo que se deban excluir los demás casos.

Al analizar globalmente las tablas anteriores, se puede afirmar

que la $E(VP)$ puede ser utilizada satisfactoriamente para estudiar la influencia que tiene la contaminación sobre el estimador K-M. Como puede verse, a medida que aumentan el grado de contaminación y la razón λ/θ , la $E(VP)$ va disminuyendo, lo cual es lógico pues si se tienen observaciones en promedio muy grandes, el estimador K-M tendrá una cola larga lo que para efectos de area bajo la curva no contribuye significativamente pero en cuanto al valor $Y_{(n)}$ este aumentará por lo que se espera que VP sea menor.

4.2.3 LA CURVA DE INFLUENCIA

Ahora bien, ya que en el planteamiento anterior se supuso la no existencia de censura, se intentó obtener expresiones análogas para $E(VP)$ en el caso de que existieran observaciones censuradas. Sin embargo, no fue posible encontrar la expresión correspondiente para $E(VP)$ pero, a pesar de esto, se atacó el problema de manera diferente basándose también en el VP del estimador K-M.

Es claro que una forma de medir globalmente la influencia que tiene el patrón de censura de una muestra particular en el estimador K-M es comparar, de alguna manera, el valor promedio de este estimador con el que se obtendría de una muestra con características iguales (tamaño, porcentaje de contaminación y razón λ/θ) pero sin censura.

Sin embargo, es generalmente de mas interés determinar una medida de la influencia que tiene cada una de las observaciones por separado en el estimador K-M. Para este caso es necesario analizar la influencia de cada observación bajo las siguientes situaciones:

1. La influencia que tiene la observación debida a su presencia en la muestra analizada, la cual puede medirse mediante la eliminación de la observación y comparando los resultados obtenidos con y sin la observación. Esto se hizo utilizando el concepto de función de influencia muestral discutida en el capítulo 2.
2. La influencia que tiene la observación debida a su naturaleza (censurada o no censurada), la cual puede medirse mediante un cambio de estado, es decir, pasar de censurada a no censurada o viceversa, y entonces comparar los valores de VP entre si.

Para cualquiera de las situaciones anteriores se propone medir la influencia de las observaciones de la siguiente manera. Sean VP el valor promedio del estimador K-M de la muestra obtenida y VP_i el valor promedio del mismo estimador de la muestra pero con la modificación requerida de la i-ésima observación. Entonces si CI_i denota la influencia de la i-ésima observación sobre el estimador K-M, CI_i estará dada por

$$CI_i = \frac{F(VP) - F(c_i VP_i)}{F(VP)}$$

en donde, F es la función de distribución de la variable aleatoria VP (valor promedio del estimador K-M) con las condiciones de contaminación, censura y tamaño de la muestra original y c_i es 1 si la modificación de la i-ésima observación fue un cambio de estado y es $n/(n-1)$ si la i-ésima observación fue eliminada (debido, esto último a que al eliminar una observación el tamaño de muestra cambia, por lo que era necesario una corrección al valor VP_i para que puedan ser comparables los valores F).

De acuerdo a la forma propuesta de CI, la influencia de cada observación es, esencialmente, el cambio relativo que tiene la función de distribución evaluada en el valor promedio con una modificación en la i -ésima observación con respecto al valor de la función de distribución evaluada en el valor promedio del estimador K-M de la muestra original. Es decir, se está comparando simplemente el valor de una estadística bajo la muestra original y bajo una situación particular (dada por una modificación en la i -ésima observación) de interés. Esto no es otra cosa que la idea de la función de influencia muestral, vista en el capítulo 2 por lo que se espera que efectivamente sea un índice que mida la influencia.

Por lo tanto, solo resta encontrar la forma de la función de distribución F para cualesquiera condiciones de la muestra. Sin embargo no fue posible encontrar una expresión para F de manera general para los diferentes valores de tamaño de muestra (n), contaminación (p y λ/θ) y censura (p_c), pero se aproximó mediante la función de distribución empírica del VP para algunos valores de n , p , λ/θ y p_c (proporción de censura en la muestra) utilizando para esto simulación de situaciones dadas por la combinación de los siguientes valores y a partir de 1000 réplicas para cada una de las situaciones consideradas.

Los valores estudiados fueron:

$$n = 20, 40 \text{ y } 60$$

$$p = 0, 5 \text{ y } 10 \%$$

$$\lambda/\theta = 2 \text{ y } 5$$

$$p_c = 0, 10, 20 \text{ y } 30 \%$$

Por lo tanto, para cada tamaño de muestra se obtuvieron 20 funciones de distribución empíricas, las cuales están dadas por:

Función empírica número	%censura	% contaminación	λ/θ
1	0	0	-
2	10	0	-
3	20	0	-
4	30	0	-
5	0	10	2
6	10	10	2
7	20	10	2
8	30	10	2
9	0	10	5
10	10	10	5
11	20	10	5
12	30	10	5
13	0	5	2
14	10	5	2
15	20	5	2
16	30	5	2
17	0	5	5
18	10	5	5
19	20	5	5
20	30	5	5

Además, se obtuvieron las gráficas de estas funciones. Estas gráficas permiten, además de calcular la CI, analizar y comparar el comportamiento global de dichas funciones al cambiar los valores de p_c , p_o y/o λ/θ . Algunas de estas gráficas se presentan al final de esta sección.

En la figura 1, se presentan las gráficas de las funciones de distribución empíricas ya suavizadas, correspondientes a un tamaño de muestra 40, un valor arbitrario de λ , sin contaminación y para 4 diferentes grados de censura.

A partir de esta gráfica se puede analizar el efecto de la censura, a diferentes niveles, sobre el comportamiento de VP. Hay que notar que las funciones de distribución reflejan un cierto paralelismo en la parte central lo que hace pensar que podríamos tener esencialmente la misma variable pero corrida a la derecha, por lo que a mayor grado de censura se tenderá a obtener valores de la VP más grandes pero con menor variabilidad puesto que el valor máximo no cambia radicalmente con el grado de censura.

Además hay un cruce con la función correspondiente a no censura, lo cual puede deberse a muestras en las que la presencia de censura en las observaciones no tiene un peso significativo en el VP.

Este patrón general se observa también en la figura 2 la cual involucra las mismas situaciones anteriores pero para un tamaño de muestra 60, aunque en esta última los valores observados de VP tienden a ser un poco más pequeños pues es mucho más probable que se obtengan valores máximos "grandes" en una muestra de tamaño 60 que en una muestra de tamaño 40.

En las figuras 3 y 4 se presentan las gráficas de las funciones de distribución para tamaños de muestra 40 y 60 respectivamente, un valor arbitrario de λ en presencia de un 10% de contaminación con $\lambda/\theta = 5$ y para 4 diferentes grados de censura.

En las gráficas hay que notar que a pesar de presentarse contaminación en la muestra, el patrón general de comportamiento antes descrito sigue existiendo con la única diferencia en el

rango de valores del VP considerado, es decir, se tiende a tener valores más pequeños (correspondientes a valores máximos grandes en la muestra) no solo por tener un tamaño mayor de muestra sino porque se está contaminando por la derecha, o sea con valores en promedio mayores que los esperados.

Además, el "corrimiento" que sufren las funciones de distribución es menor en presencia de contaminación, lo cual sugiere analizar las gráficas con diferentes grados de contaminación.

En la figura 4 se nota que el cruce de la función de distribución sin censura solo se presenta para la correspondiente función de distribución a un 10% de censura puesto que, para el tamaño de muestra considerado en este caso (60) un 20% o más de censura corresponde a bastantes observaciones (12 o más) por lo que no es lógico pensar en valores de la VP con igual valor de la función de distribución con (20% o más) y sin censura.

En la gráfica 5. se muestran las funciones de distribución correspondientes a un tamaño de muestra 40, un valor arbitrario de λ con un 5% de contaminación y 10% de censura para 3 diferentes valores del parámetro de contaminación.

Al analizar el patrón de comportamiento que tiene el VP cuando el parámetro de contaminación (por la derecha) cambia, se observa que a mayor valor del cociente λ/θ serán más probables valores pequeños del VP puesto que en la muestra se tendrán valores en promedio cada vez mayores que el resto.

Para este tipo de gráficas no hay cruce entre las funciones de distribución. Sin embargo, se nota un gran cambio al pasar de $\lambda/\theta = 2$ a $\lambda/\theta = 5$ lo cual refleja la importancia que tiene el parámetro de contaminación (o el cociente λ/θ) a pesar de tener solamente un 5% de contaminación.

Hay que añadir que este comportamiento también se presenta para otros niveles de censura, de contaminación y tamaño de muestra (aunque no se muestran aquí todas las gráficas). Como ejemplo de esto se presenta la figura 6 que corresponde a las funciones de distribución para un tamaño de muestra 60, un valor de λ arbitrario con un 10% de contaminación, 20% de censura y con los mismos valores del parámetro θ de contaminación, es decir, λ/θ igual a 1,2 y 5.

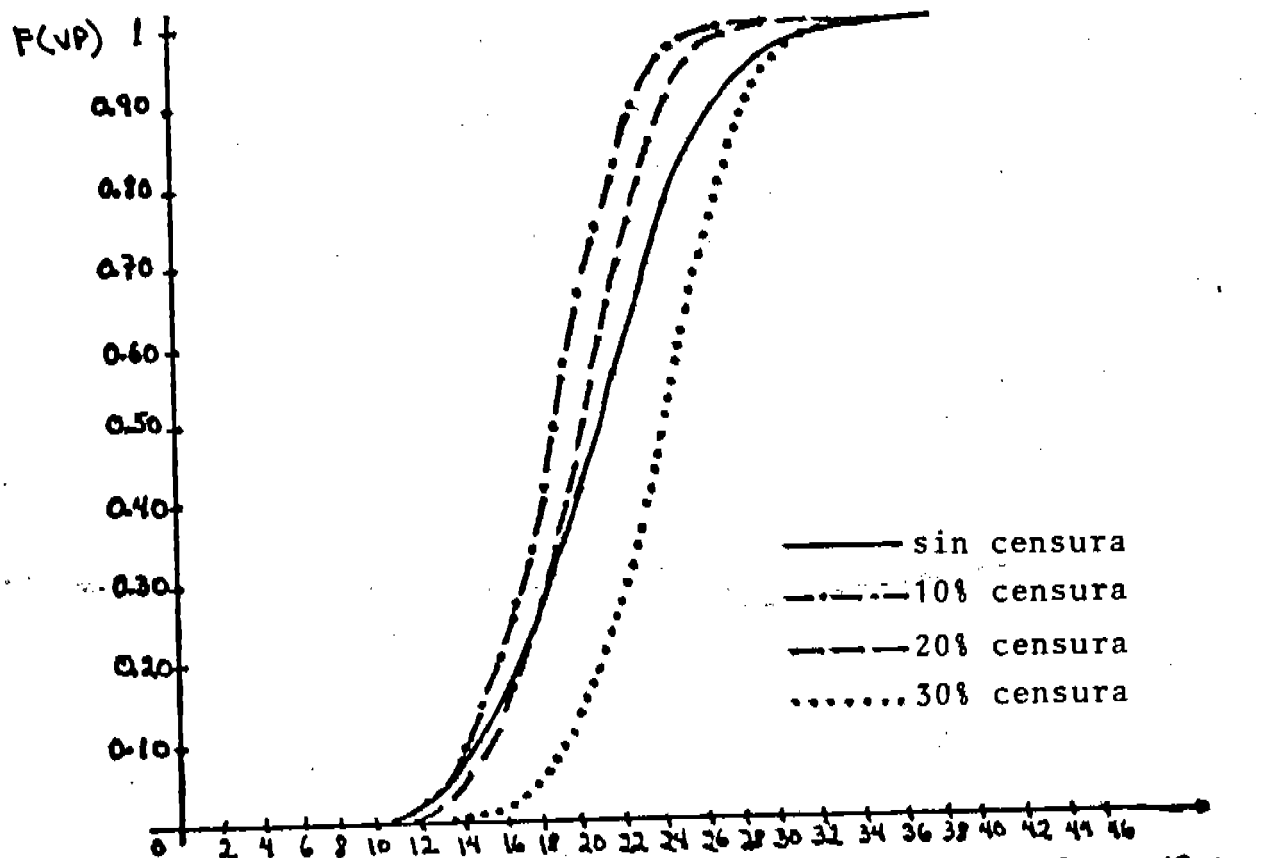
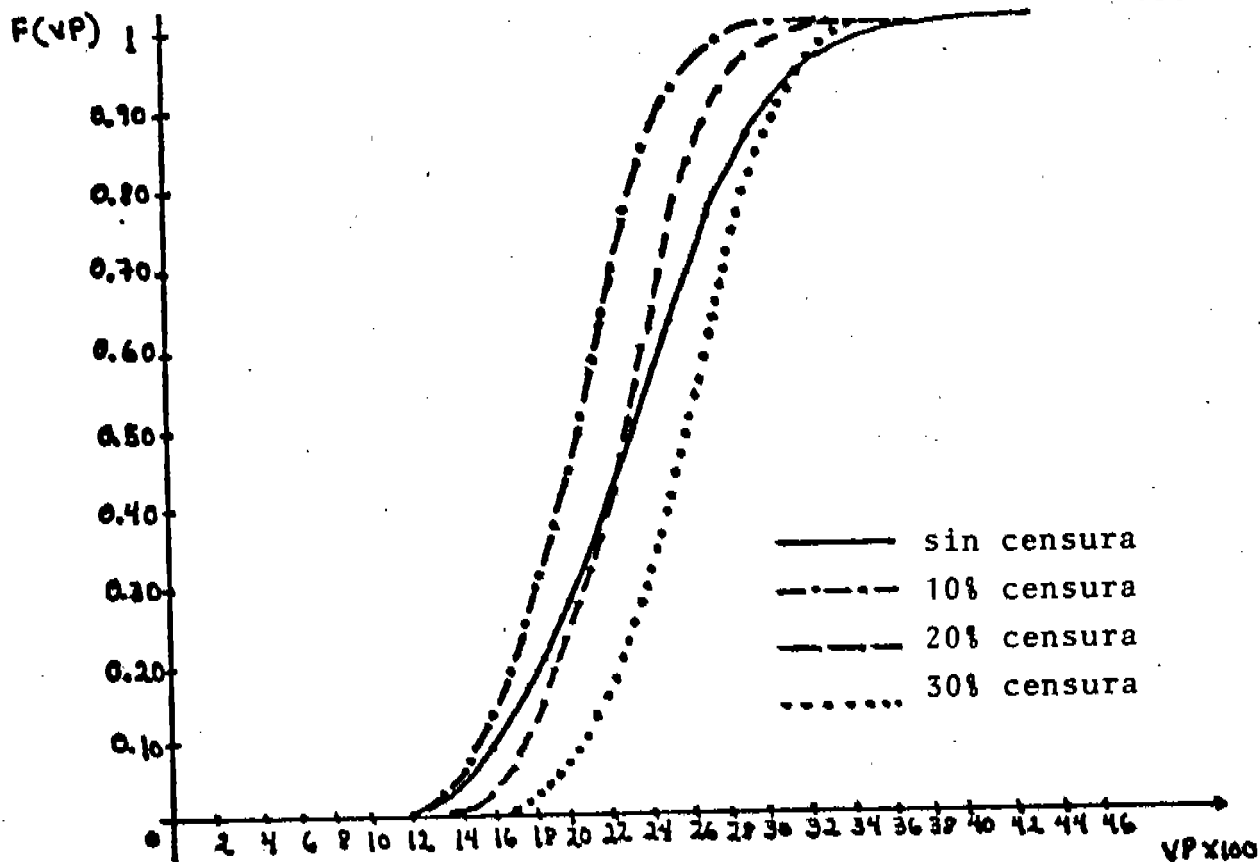
Ahora bien, al analizar entre sí las figuras 5 y 6 hay que notar que a pesar de las diferentes características consideradas, las 2 gráficas en sí son semejantes, lo que podría estar reflejando que el patrón de censura y el de contaminación se compensan de tal manera que no producen cambios radicales en las funciones de distribución.

Finalmente, para analizar el efecto que tiene el grado de contaminación, se presentan las figuras 7 y 8. La gráfica 7 corresponde a distribuciones con un tamaño de muestra 40, un valor arbitrario de λ , sin censura y con $\lambda/\theta = 2$, mientras que la figura 8 corresponde a muestras de tamaño 60 con λ arbitrario, 30% de censura y $\lambda/\theta = 5$. Como puede apreciarse en las 2 figuras, al aumentar el grado de contaminación se sufre

un corrimiento a la izquierda lo cual es debido, como se esperaba, a que los valores máximos grandes aparecen con mayor probabilidad lo que ocasiona que el valor VP disminuya. Al tener curvas casi paralelas (figura 7) y/o que no tienen cruce alguno (figura 8) se muestra que para estas situaciones nunca va a ser igual tener un grado mayor o menor de contaminación, es decir, la contaminación va a tener un efecto significativo sobre el VP y dicho efecto incrementará a medida que el porcentaje de contaminación aumente.

La diferencia mas marcada entre las curvas se presenta en la figura 8, pues al pasar de un grado a otro de contaminación se ven involucradas 3 observaciones contaminantes adicionales, que en conjunto tienen un efecto significativo.

En la gráfica correspondiente a estas mismas condiciones pero con un tamaño de muestra 20 (la cual no se presenta), el cambio entre las funciones es casi insignificante pues aquí solo se involucra 1 observación contaminante adicional al cambiar de un nivel a otro de contaminación.



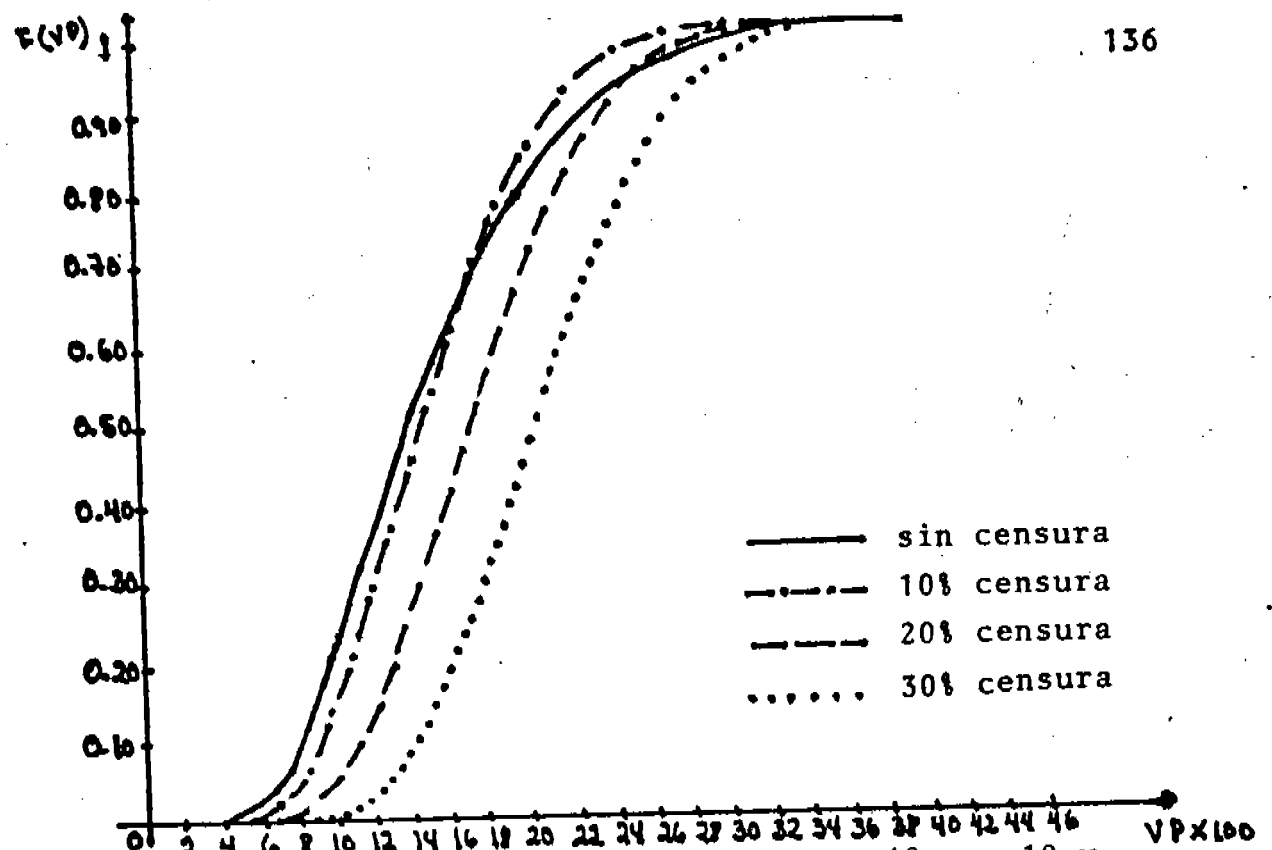


Figura 3. Función de distribución de VP para $n=40$, $p_0=10$ y $\lambda/\theta = 5$.

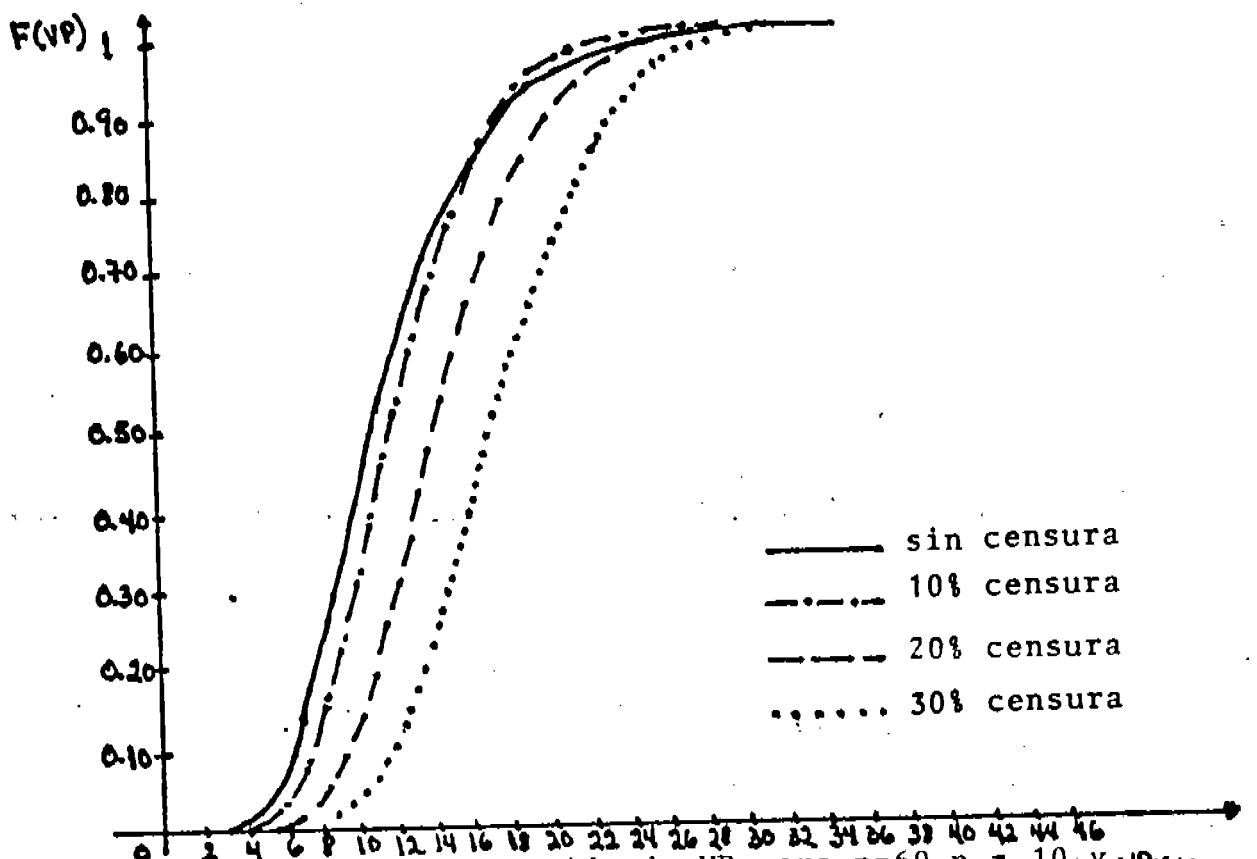


Figura 4. Función de distribución de VP para $n=60$, $p_0=10$ y $VP \times 100$.

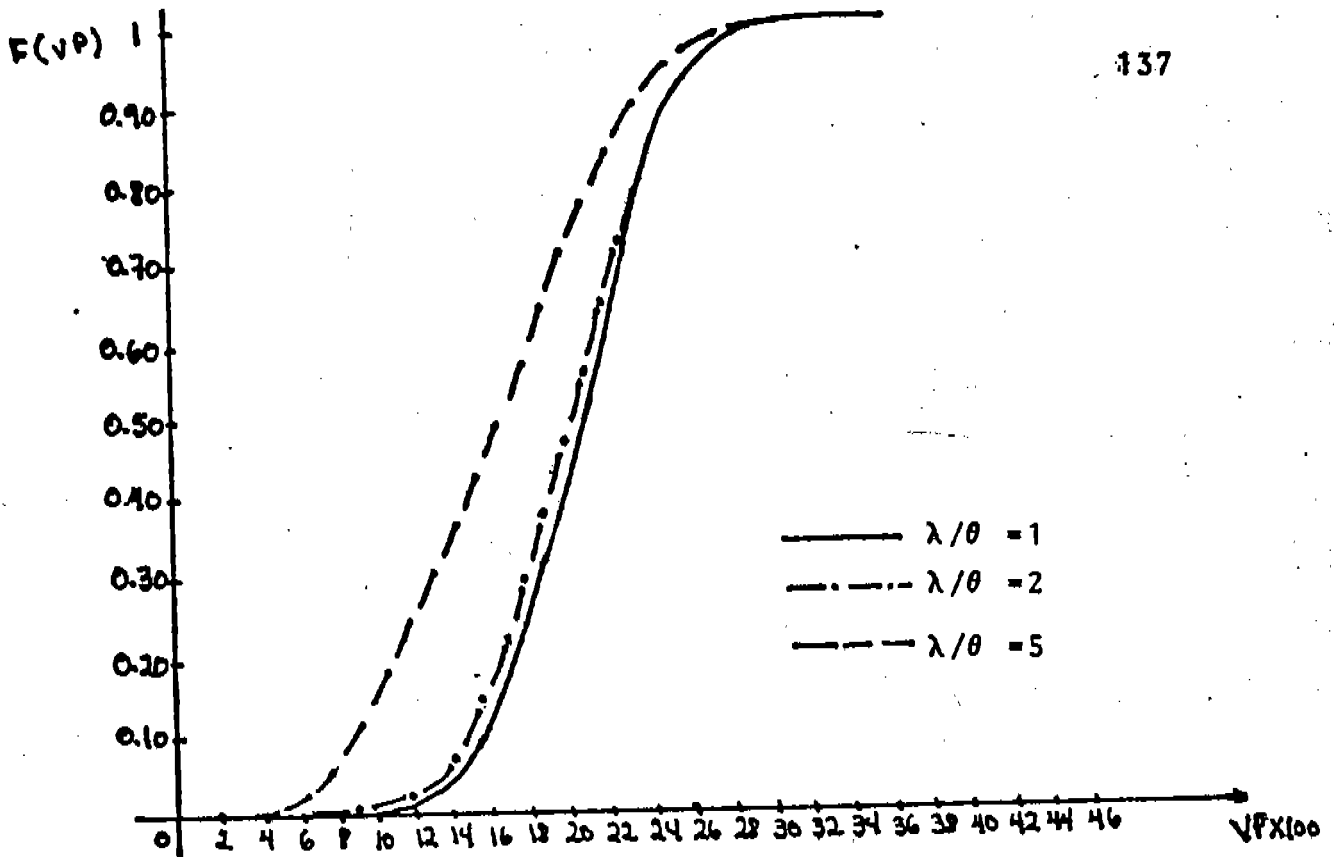


Figura 5. Función de distribución de VP para $n=40$, $p_0=5$ y $p_c=10$.

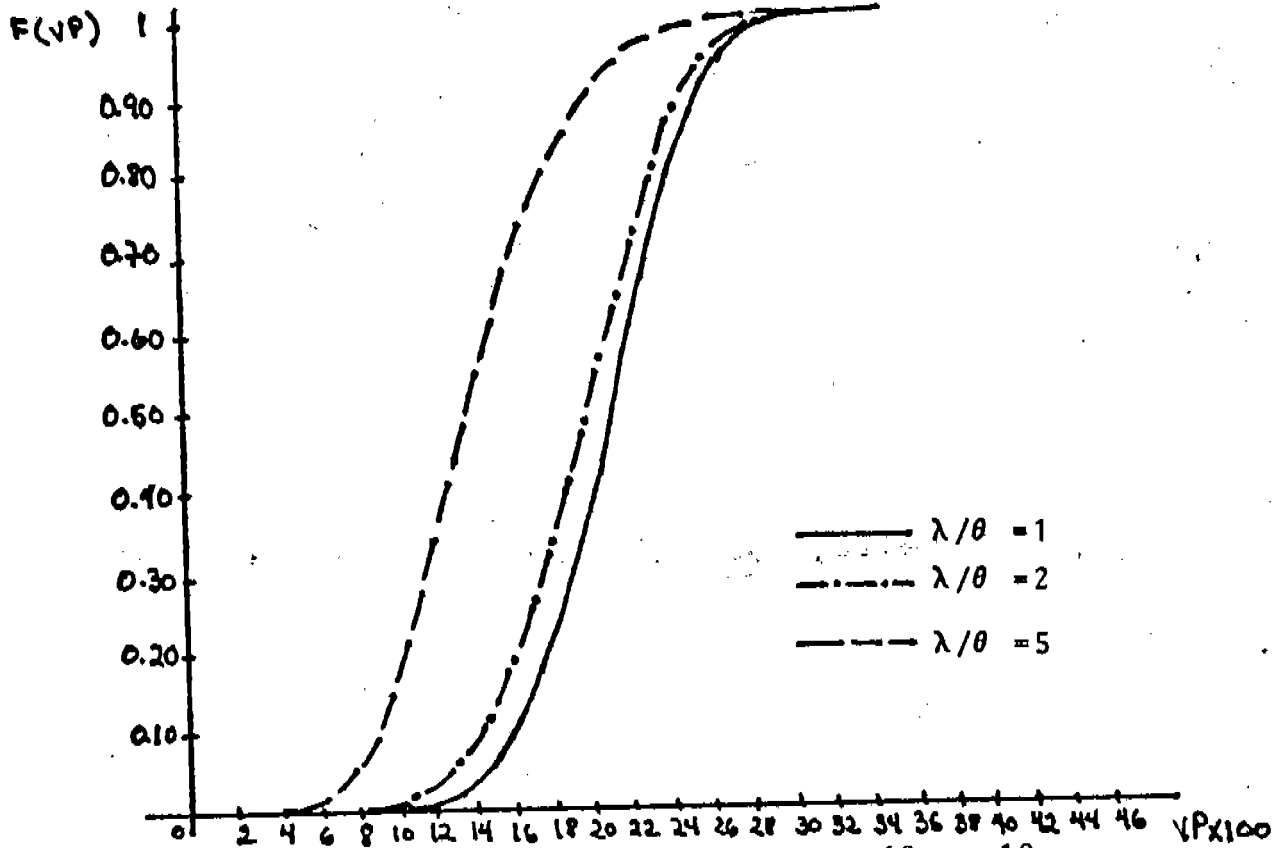


Figura 6. Función de distribución de VP para $n=60$, $p_0=10$ y $p_c=10$.

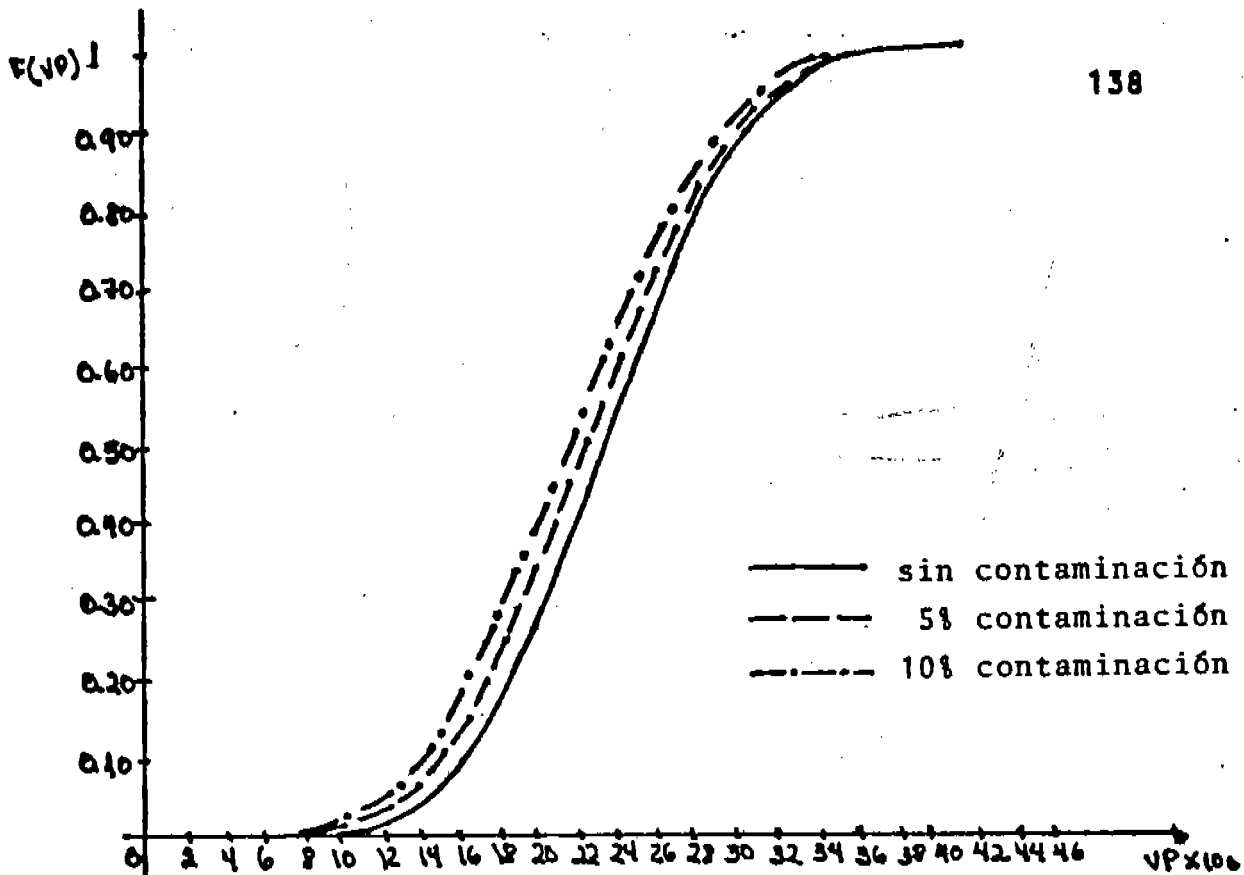


Figura 7. Función de distribución de VP para $n=40$, sin censura y $\lambda/\theta=2$.

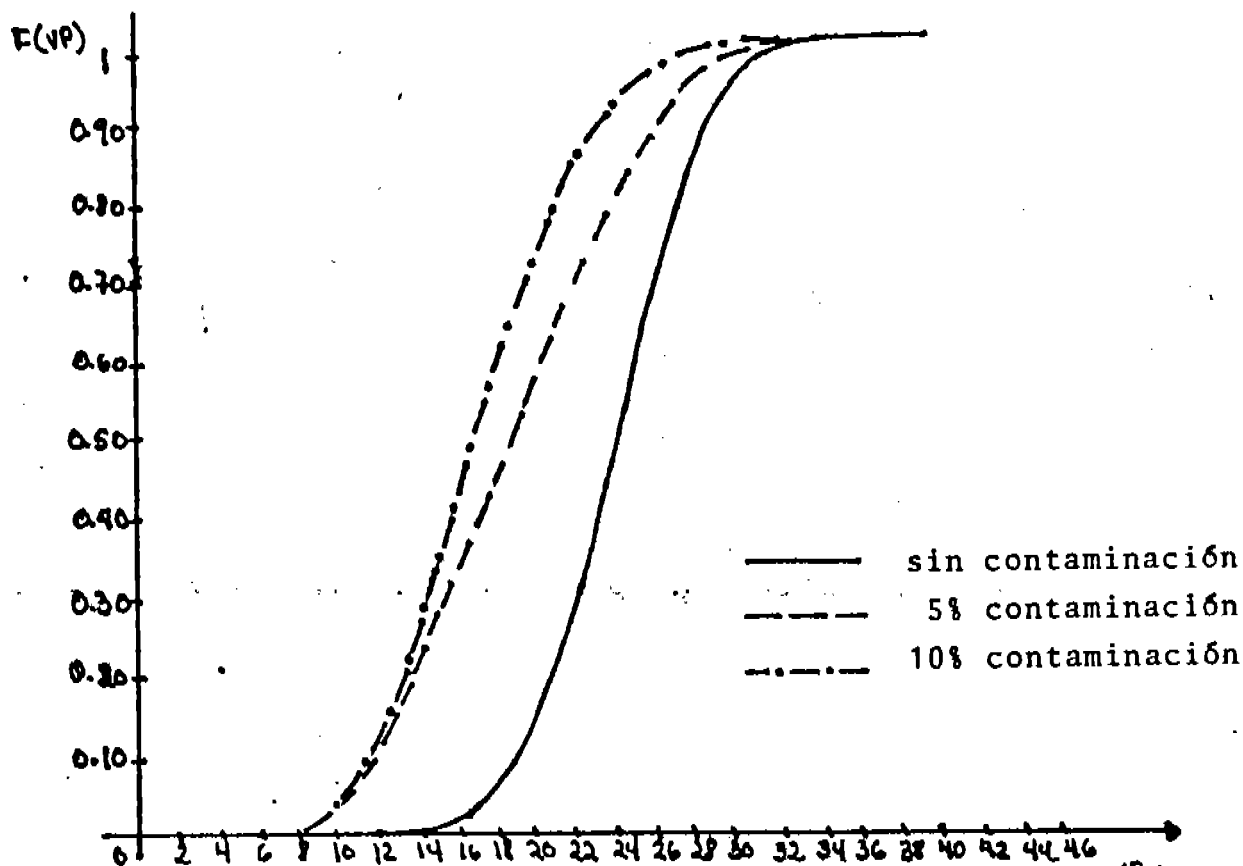


Figura 8. Función de distribución de VP para $n=60$, $p_c=30$ y $\lambda/\theta=5$.

4.2.4 EJEMPLO

Para mostrar el uso e interpretación de la curva de influencia propuesta se presenta el siguiente ejemplo. Para ello se generó una muestra de tamaño 20 con un valor de $\lambda = 0.1$, $\theta = 0.05$ con un grado de contaminación del 10% y un 30% de censura.

Las observaciones de la muestra ya ordenada fueron

0.473	2.848	5.885	11.542
0.803	3.185*	7.603*	12.404
1.035*	3.584	8.721	14.945*
1.282	4.443*	9.672	31.211*
2.661	4.509	11.087	37.238

en donde * denota una observación censurada.

El estimador K-M que se obtiene a partir de estos datos es

$$\hat{S}^0(t) = \left\{ \begin{array}{ll} 0.000 & t < 0 \\ 1.000 & 0 \leq t < 0.473 \\ 0.950 & 0.473 \leq t < 0.803 \\ 0.900 & 0.803 \leq t < 1.282 \\ 0.847 & 1.282 \leq t < 2.661 \\ 0.794 & 2.661 \leq t < 2.848 \\ 0.741 & 2.848 \leq t < 3.584 \\ 0.684 & 3.584 \leq t < 4.509 \\ 0.622 & 4.504 \leq t < 5.885 \\ 0.560 & 5.885 \leq t < 8.721 \\ 0.490 & 8.721 \leq t < 9.672 \\ 0.420 & 9.672 \leq t < 11.087 \\ 0.350 & 11.087 \leq t < 11.542 \\ 0.280 & 11.542 \leq t < 12.404 \\ 0.210 & 12.404 \leq t < 37.238 \\ 0.000 & t \geq 37.238 \end{array} \right.$$

Las magnitudes de los saltos son

D	=	1-0.950	=	0.050	D	=	0.062
1					8		
D	=	0.050			D	=	0.070
2					9		
D	=	0.053			D	=	0.070
3					10		
D	=	0.053			D	=	0.070
4					11		
D	=	0.053			D	=	0.070
5					12		
D	=	0.057			D	=	0.070
6					13		
D	=	0.062			D	=	0.210
7					14		

Entonces el area bajo la curva es

$$\text{Area} = 12.829109$$

el VP del estimador K-M es

$$\text{VP} = 0.3445$$

y el valor de la función de distribución es

$$F(\text{VP}) = 0.84125$$

Analizaremos ahora el efecto que tienen cada una de las 6 observaciones censuradas sobre el estimador K-M. Como se ha mencionado ya, esto se hará considerando un cambio de estado en estas observaciones y entonces se calculará el nuevo valor promedio del estimador K-M tomando en cuenta que cada una de ellas pasa a ser una observación no censurada.

Recordando que la función de influencia está dada por

$$CI_1 = \frac{F(VP) - F(c VP)}{F(VP)}$$

se tendrá para este ejemplo que

$$CI_1 = \frac{0.84125 - F(VP)}{0.84125} \quad i=3,7,9,12,18,19$$

- Para la observación número 3 (1.035) tenemos que el nuevo valor del estimador K-M es

$$\hat{S}_3^0(t) = \left\{ \begin{array}{ll} 0.000 & t < 0 \\ 1.000 & 0 \leq t < 0.473 \\ 0.950 & 0.473 \leq t < 0.803 \\ 0.900 & 0.803 \leq t < 1.035 \\ 0.850 & 1.035 \leq t < 1.282 \\ 0.800 & 1.282 \leq t < 2.661 \\ 0.750 & 2.661 \leq t < 2.848 \\ 0.700 & 2.848 \leq t < 3.584 \\ 0.646 & 3.584 \leq t < 4.509 \\ 0.587 & 4.509 \leq t < 5.885 \\ 0.529 & 5.885 \leq t < 8.721 \\ 0.463 & 8.721 \leq t < 9.672 \\ 0.397 & 9.672 \leq t < 11.087 \\ 0.330 & 11.087 \leq t < 11.542 \\ 0.264 & 11.542 \leq t < 12.404 \\ 0.198 & 12.404 \leq t < 37.238 \\ 0.000 & t \geq 37.238 \end{array} \right.$$

y entonces tendremos

$$Area_3 = 12.1717$$

$$VP_3 = 0.3267$$

$$F(VP)_3 = 0.74292$$

por lo que

$$CI_3 = \frac{0.84125 - 0.74292}{0.84125} = 0.1169$$

- Para la observación número 7 (3.185) tendremos

$$S_7^0(t) = \left\{ \begin{array}{ll} 0.000 & t < 0 \\ 1.000 & 0 \leq t < 0.473 \\ 0.950 & 0.473 \leq t < 0.803 \\ 0.900 & 0.803 \leq t < 1.282 \\ 0.847 & 1.282 \leq t < 2.661 \\ 0.794 & 2.661 \leq t < 2.848 \\ 0.741 & 2.848 \leq t < 3.185 \\ 0.688 & 3.185 \leq t < 3.584 \\ 0.635 & 3.574 \leq t < 4.509 \\ 0.578 & 4.509 \leq t < 5.885 \\ 0.520 & 5.885 \leq t < 8.721 \\ 0.455 & 8.721 \leq t < 9.672 \\ 0.390 & 9.672 \leq t < 11.087 \\ 0.325 & 11.087 \leq t < 11.542 \\ 0.260 & 11.542 \leq t < 12.404 \\ 0.195 & 12.404 \leq t < 37.238 \\ 0.000 & t \geq 37.238 \end{array} \right.$$

y entonces

$$Area_7 = 12.1117$$

$$VP_7 = 0.3253$$

$$F(VP_7) = 0.73228$$

y

$$CI_7 = 0.1295$$

- Para la observación número 9 (4.443) tendremos

$$\hat{S}_9^0(t) = \left\{ \begin{array}{ll} 0.000 & t < 0 \\ 1.000 & 0 \leq t < 0.473 \\ 0.950 & 0.473 \leq t < 0.803 \\ 0.900 & 0.803 \leq t < 1.282 \\ 0.847 & 1.282 \leq t < 2.661 \\ 0.794 & 2.661 \leq t < 2.848 \\ 0.741 & 2.848 \leq t < 3.584 \\ 0.684 & 3.584 \leq t < 4.443 \\ 0.627 & 4.443 \leq t < 4.509 \\ 0.570 & 4.509 \leq t < 5.885 \\ 0.513 & 5.885 \leq t < 8.721 \\ 0.449 & 8.721 \leq t < 9.672 \\ 0.385 & 9.672 \leq t < 11.087 \\ 0.321 & 11.087 \leq t < 11.542 \\ 0.257 & 11.542 \leq t < 12.404 \\ 0.192 & 12.404 \leq t < 37.238 \\ 0.000 & t \geq 37.238 \end{array} \right.$$

y los nuevos valores son

$$\text{Area}_9 = 12.0657$$

$$\text{VP}_9 = 0.3240$$

$$\text{F}(\text{VP}_9) = 0.7224$$

y

$$\text{CI}_9 = 0.1413$$

- Para la observación número 12 (7.603) tendremos

$$\hat{S}_{12}^0(t) = \left\{ \begin{array}{ll} 0.000 & t < 0 \\ 1.000 & 0 \leq t < 0.473 \\ 0.950 & 0.473 \leq t < 0.803 \\ 0.900 & 0.803 \leq t < 1.282 \\ 0.847 & 1.282 \leq t < 2.661 \\ 0.794 & 2.661 \leq t < 2.848 \\ 0.741 & 2.848 \leq t < 3.584 \\ 0.684 & 3.584 \leq t < 4.509 \\ 0.622 & 4.509 \leq t < 5.885 \\ 0.560 & 5.885 \leq t < 7.603 \\ 0.498 & 7.603 \leq t < 8.721 \\ 0.435 & 8.721 \leq t < 9.672 \\ 0.373 & 9.672 \leq t < 11.087 \\ 0.311 & 11.087 \leq t < 11.542 \\ 0.249 & 11.542 \leq t < 12.404 \\ 0.187 & 12.404 \leq t < 37.238 \\ 0.000 & t \geq 37.238 \end{array} \right.$$

y los valores correspondientes son

$$\text{Area}_{12} = 12.0182$$

$$\text{VP}_{12} = 0.3227$$

$$F(\text{VP}_{12}) = 0.71252$$

y

$$\text{CI}_{12} = 0.1530$$

- Para la observación número 18 (14.945) el nuevo valor del estimador K-M es

$$\hat{S}_{18}^0(t) = \left\{ \begin{array}{ll} 0.000 & t < 0 \\ 1.000 & 0 \leq t < 0.473 \\ 0.950 & 0.473 \leq t < 0.803 \\ 0.900 & 0.803 \leq t < 1.282 \\ 0.847 & 1.282 \leq t < 2.661 \\ 0.794 & 2.661 \leq t < 2.848 \\ 0.741 & 2.848 \leq t < 3.584 \\ 0.684 & 3.584 \leq t < 4.509 \\ 0.622 & 4.509 \leq t < 5.885 \\ 0.560 & 5.885 \leq t < 8.721 \\ 0.490 & 8.721 \leq t < 9.672 \\ 0.420 & 9.672 \leq t < 11.087 \\ 0.350 & 11.087 \leq t < 11.542 \\ 0.280 & 11.542 \leq t < 12.404 \\ 0.210 & 12.404 \leq t < 14.945 \\ 0.140 & 14.945 \leq t < 37.238 \\ 0.000 & t \geq 37.238 \end{array} \right.$$

y los valores correspondientes son

$$\text{Area}_{18} = 11.2693$$

$$\text{VP}_{18} = 0.3026$$

$$F(\text{VP}_{18}) = 0.56336$$

y

$$\text{CI}_{18} = 0.3303$$

- Por último, para la observación número 19 (31.211) el estimador K-M que se obtiene en este caso es igual al obtenido para 14.945 con la pequeña diferencia que en lugar de aparecer 14.945 en los 2 penúltimos intervalos aparece el valor 31.211. Entonces, en este caso los valores son

$$\text{Area}_{19} = 12.4075$$

$$VP_{19} = 0.3332$$

$$F(VP_{19}) = 0.78496$$

$$y \quad CI_{19} = 0.0669$$

Resumiendo los resultados de estos 6 cambios se tiene la siguiente tabla.

Observación que deja de ser cen- surada.	VP _i	F(VP _i)	CI _i
1.035	0.3267	0.74292	0.1169
3.185	0.3253	0.73228	0.1295
4.443	0.3240	0.72240	0.1413
7.603	0.3227	0.71252	0.1530
14.945	0.3026	0.56336	0.3303
31.211	0.3332	0.78496	0.0669

Por lo tanto, si consideramos que entre mas cercana a cero sea la CI esto indicará una menor influencia (pues el valor del VP será muy parecido al de la muestra original) entonces en este ejemplo se puede afirmar que la observación censurada número 19, 31.211, tiene la menor influencia posible pues la función de distribución del VP solo se ve afectado en un 6% de su valor, el cual no se espera que sea significativo.

Por otro lado, la observación censurada número 18, 14.945, tiene una influencia muy fuerte, pues en principio es la mayor de todas y al mismo tiempo está reflejando que el hecho que dicha observación sea censurada va a modificar el valor de la función

de distribución en un 33% lo cual es significativo pues es casi una tercera parte de su valor.

Las otras 4 observaciones tienen un efecto medianamente significativo pues variarán el valor de la función de distribución entre un 11 y un 15% aproximadamente, el cual va aumentando a medida que la observación censurada sea mayor.

Hay que añadir que la forma de afectar es en sentido opuesto pues se tiene una disminución en el valor de la función de distribución si la curva de influencia es positiva mientras que se tendrá un aumento si esta curva es negativa.

Ahora analizaremos el efecto que tienen algunas de las observaciones no censuradas, debido a su naturaleza, en el VP. Para esto se escogieron 4 de las 14 observaciones de interés y se les consideró como observaciones censuradas una por una. La forma de calcular la curva de influencia en este caso es igual que en el análisis anterior.

- Si consideramos la observación número 1 (0.473) como si fuera censurada, el estimador K-M será

$$\hat{S}_1^o(t) =$$

0.000		t < 0
1.000	0	t < 0.803
0.947	0.803	t < 1.282
0.892	1.282	t < 2.661
0.836	2.661	t < 2.848
0.780	2.848	t < 3.584
0.720	3.584	t < 4.509
0.655	4.509	t < 5.885
0.589	5.885	t < 8.721
0.516	8.721	t < 9.672
0.442	9.672	t < 11.087
0.368	11.087	t < 11.542
0.295	11.542	t < 12.404
0.221	12.404	t < 37.238
0.000		t > 37.238

y tendremos

$$\text{Area}_1 = 13.4795$$

$$\text{VP}_1 = 0.3620$$

$$F(\text{VP}_1) = 0.9072$$

y

$$\text{CI}_1 = -0.0784$$

- Si la observación número 8 (3.584) fuera censurada, el estimador K-M será

$$\hat{S}_8^o(t) = \left\{ \begin{array}{ll} 0.000 & t < 0 \\ 1.000 & 0 \leq t < 0.473 \\ 0.950 & 0.473 \leq t < 0.803 \\ 0.900 & 0.803 \leq t < 1.282 \\ 0.847 & 1.282 \leq t < 2.661 \\ 0.794 & 2.661 \leq t < 2.848 \\ 0.741 & 2.848 \leq t < 4.509 \\ 0.674 & 4.509 \leq t < 5.885 \\ 0.606 & 5.885 \leq t < 8.721 \\ 0.531 & 8.721 \leq t < 9.672 \\ 0.455 & 9.672 \leq t < 11.087 \\ 0.379 & 11.087 \leq t < 11.542 \\ 0.303 & 11.542 \leq t < 12.404 \\ 0.227 & 12.404 \leq t < 37.238 \\ 0.000 & t \geq 37.238 \end{array} \right.$$

$$\text{Area}_8 = 13.6416$$

$$\text{VP}_8 = 0.3663$$

$$F(\text{VP}_8) = 0.91838$$

y

$$\text{CI}_8 = -0.09165$$

- Si la observación número 15 (11.087) fuera censurada, el estimador K-M será

$$\hat{S}_{15}^o(t) = \left\{ \begin{array}{ll} 0.000 & t < 0 \\ 1.000 & 0 \leq t < 0.473 \\ 0.950 & 0.473 \leq t < 0.803 \\ 0.900 & 0.803 \leq t < 1.282 \\ 0.847 & 1.282 \leq t < 2.661 \\ 0.794 & 2.661 \leq t < 2.848 \\ 0.741 & 2.848 \leq t < 3.584 \\ 0.684 & 3.584 \leq t < 4.509 \\ 0.622 & 4.509 \leq t < 5.885 \\ 0.560 & 5.885 \leq t < 8.721 \\ 0.490 & 8.721 \leq t < 9.672 \\ 0.420 & 9.672 \leq t < 11.542 \\ 0.336 & 11.542 \leq t < 12.404 \\ 0.252 & 12.404 \leq t < 37.238 \\ 0.000 & t \geq 37.238 \end{array} \right.$$

y los valores correspondientes son

$$\text{Area}_{15} = 13.9517$$

$$\text{VP}_{15} = 0.3747$$

$$F(\text{VP}_{15}) = 0.94398$$

y

$$\text{CI}_{15} = -0.1221$$

- Si la observación número 20 (37.238) fuera censurada, el estimador K-M será

$$\hat{S}_{20}^0(t) = \left\{ \begin{array}{lll} 0.000 & & t < 0 \\ 1.000 & 0 & \leftarrow t < 0.473 \\ 0.950 & 0.473 & \leftarrow t < 0.803 \\ 0.900 & 0.803 & \leftarrow t < 1.282 \\ 0.847 & 1.282 & \leftarrow t < 2.661 \\ 0.794 & 2.661 & \leftarrow t < 2.848 \\ 0.741 & 2.848 & \leftarrow t < 3.584 \\ 0.684 & 3.584 & \leftarrow t < 4.509 \\ 0.622 & 4.509 & \leftarrow t < 5.885 \\ 0.560 & 5.885 & \leftarrow t < 8.721 \\ 0.490 & 8.721 & \leftarrow t < 9.672 \\ 0.420 & 9.672 & \leftarrow t < 11.087 \\ 0.350 & 11.087 & \leftarrow t < 11.542 \\ 0.280 & 11.542 & \leftarrow t < 12.404 \\ 0.210 & 12.404 & \leftarrow t < 47.238 \\ 0.000 & & t > 47.238 \end{array} \right.$$

En este caso, para la construcción del estimador K-M se dió a h el valor de una desviación estándar de la población estudiada, es decir, $h = 1/\lambda = 10$, por lo que

$$t_{20-6+2} = t_{16} = Y_{(20)} + 10 = 37.238 + 10 = 47.238$$

y los valores en este caso son

$$\text{Area}_{20} = 14.9282$$

$$\text{VP}_{20} = 0.3160$$

$$F(\text{VP}_{20}) = 0.666$$

y

$$\text{CI}_{20} = 0.2083$$

Resumiendo los resultados de las 4 últimas situaciones, se llega a la siguiente tabla:

Observación que pasa a ser censu- rada.	VP i	F(VP) i	CI i
0.473	0.3620	0.90720	-0.07840
3.584	0.3663	0.91838	-0.09165
11.087	0.3747	0.94398	-0.12210
37.238	0.3160	0.66600	0.20830

Al analizar los valores de la curva de influencia de estas 4 observaciones llegamos a que la observación número 20, 37.238, tiene la mayor influencia entre las observaciones no censuradas pues tiene un efecto tal que modifica aproximadamente un 20% de su valor a la función de distribución; además, esta modificación es en sentido inverso, es decir, disminuye el correspondiente valor del VP.

Por otro lado, las otras 3 observaciones tienen un efecto tal que aumenta el valor de VP al tener un valor negativo de la curva de influencia. Lo anterior es lógico si pensamos que, al ser estas observaciones censuradas (con valores pequeños), se tendrá que los verdaderos tiempos de falla son mayores por lo que el tiempo de vida esperado es mas grande; esto se refleja en un VP mayor.

Además, la observación mas influyente entre las analizadas es la mas cercana al verdadero tiempo promedio de vida. Sin embargo, no hay que olvidar que en este caso se está estudiando el efecto por no censura y no por la presencia de la observación en la muestra.

Finalmente se analizará el efecto que tienen las observaciones, tanto censuradas como no censuradas, por el hecho de estar presentes en la muestra obtenida.

Hay que recordar que en este caso el cambio que se produce en cada paso nos lleva a calcular el VP a partir de una muestra de tamaño 19 con lo cual la constante de corrección es

$$c_1 = 20/19$$

por lo que la curva de influencia está dada por

$$CI_{-1} = \frac{0.84125 - F((20/19)VP_{-1})}{0.84125}$$

en donde el subíndice (-1) indica que la i-ésima observación es eliminada.

En este caso se escogieron 4 observaciones de las 20 para dicho análisis.

- Si se elimina la observación número 2, 0.803, la cual es no censurada, el estimador K-M será

$$\hat{S}_{-2}^0(t) = \left\{ \begin{array}{lll} 0.000 & & t < 0 \\ 1.000 & 0 & \leftarrow t < 0.473 \\ 0.947 & 0.473 & \leftarrow t < 1.282 \\ 0.892 & 1.282 & \leftarrow t < 2.661 \\ 0.836 & 2.661 & \leftarrow t < 2.848 \\ 0.780 & 2.848 & \leftarrow t < 3.584 \\ 0.720 & 3.584 & \leftarrow t < 4.509 \\ 0.655 & 4.509 & \leftarrow t < 5.885 \\ 0.589 & 5.885 & \leftarrow t < 8.721 \\ 0.516 & 8.721 & \leftarrow t < 9.672 \\ 0.442 & 9.672 & \leftarrow t < 11.087 \\ 0.368 & 11.087 & \leftarrow t < 11.542 \\ 0.295 & 11.542 & \leftarrow t < 12.404 \\ 0.221 & 12.404 & \leftarrow t < 37.238 \\ 0.000 & & t \rightarrow 37.238 \end{array} \right.$$

por lo que los nuevos valores son

$$\text{Area}_{-2} = 13.4621$$

$$\text{VP}_{-2} = 0.3615$$

$$F((20/19)\text{VP}_{-2}) = F(0.3805) = 0.9627$$

y

$$\text{CI}_{-2} = -0.1444$$

- Si se elimina la observación número 7, 3.185, la cual es censurada, el estimador K-M será

$$\hat{S}_{-7}^0(t) = \left\{ \begin{array}{lll} 0.000 & & t < 0 \\ 1.000 & 0 & \leq t < 0.473 \\ 0.947 & 0.473 & \leq t < 0.803 \\ 0.895 & 0.803 & \leq t < 1.282 \\ 0.839 & 1.282 & \leq t < 2.661 \\ 0.783 & 2.661 & \leq t < 2.848 \\ 0.727 & 2.848 & \leq t < 3.584 \\ 0.671 & 3.584 & \leq t < 4.509 \\ 0.610 & 4.509 & \leq t < 5.885 \\ 0.549 & 5.885 & \leq t < 8.721 \\ 0.480 & 8.721 & \leq t < 9.672 \\ 0.412 & 9.672 & \leq t < 11.087 \\ 0.343 & 11.087 & \leq t < 11.542 \\ 0.275 & 11.542 & \leq t < 12.404 \\ 0.206 & 12.404 & \leq t < 37.238 \\ 0.000 & & t \geq 37.238 \end{array} \right.$$

y los valores son

$$\text{Area}_{-7} = 12.6150$$

$$\text{VP}_{-7} = 0.3388$$

$$F((20/19)\text{VP}_{-7}) = F(0.3566) = 0.88976$$

y

$$\text{CI}_{-7} = -0.0576$$

- Si se elimina la observación número 14, 9.672, la cual es no censurada, el estimador K-M será

$$\hat{S}_{-14}^0(t) = \left\{ \begin{array}{ll} 0.000 & t < 0 \\ 1.000 & 0 \leq t < 0.473 \\ 0.947 & 0.473 \leq t < 0.803 \\ 0.895 & 0.803 \leq t < 1.282 \\ 0.839 & 1.282 \leq t < 2.661 \\ 0.783 & 2.661 \leq t < 2.848 \\ 0.727 & 2.848 \leq t < 3.584 \\ 0.666 & 3.584 \leq t < 4.509 \\ 0.600 & 4.509 \leq t < 5.885 \\ 0.533 & 5.885 \leq t < 8.721 \\ 0.457 & 8.721 \leq t < 11.087 \\ 0.381 & 11.087 \leq t < 11.542 \\ 0.305 & 11.542 \leq t < 12.404 \\ 0.228 & 12.404 \leq t < 37.238 \\ 0.000 & t \geq 37.238 \end{array} \right.$$

y entonces

$$\text{Area}_{-14} = 13.1970$$

$$\text{VP}_{-14} = 0.3544$$

$$F_{-14}((20/19)\text{VP}_{-14}) = F(0.3731) = 0.9386$$

y

$$\text{CI}_{-14} = -0.1157$$

- Por último, si se elimina la observación número 18, 14.945, la cual es censurada pero con la característica adicional que su censura es la que mas influye en el VP del estimador K-M, el estimador K-M será

$S_{-18}^0(t) =$	{	0.000		$t < 0$
		1.000	0	$t < 0.473$
		0.947	0.473	$t < 0.803$
		0.895	0.803	$t < 1.282$
		0.839	1.282	$t < 2.661$
		0.783	2.661	$t < 2.848$
		0.727	2.848	$t < 3.584$
		0.666	3.584	$t < 4.509$
		0.600	4.509	$t < 5.885$
		0.533	5.885	$t < 8.721$
		0.457	8.721	$t < 9.672$
		0.381	9.672	$t < 11.087$
		0.305	11.087	$t < 11.542$
		0.228	11.542	$t < 12.404$
		0.152	12.404	$t < 37.238$
		0.000		$t > 37.238$

y entonces

$$\text{Area}_{-18} = 11.0976$$

$$\text{VP}_{-18} = 0.2980$$

$$F((20/19)\text{VP}_{-18}) = F(0.3137) = 0.65105$$

y

$$\text{CI}_{-18} = 0.2261$$

La siguiente tabla resume estas últimas 4 situaciones

Observación eli- minada de la ---- muestra	VP -1	F((20/19)VP) -1	CI -1
0.803	0.3615	0.9627	-0.1444
3.185	0.3388	0.8898	-0.0576
9.672	0.3544	0.9386	-0.1157
14.945	0.2980	0.6511	0.2261

De nuevo se encuentra que la observación número 18, 14.945, es la observación cuya presencia en la muestra es mas altamente influyente, ya que, como se dijo anteriormente, el hecho de ser censurada implica que el verdadero tiempo de falla sea mayor.

Por otro lado, la presencia de la observación censurada número 7, 3.185, no causa mucho efecto pues, como se puede apreciar, solo afecta en un 5% al valor de la función de distribución.

Ahora bien, las otras 2 observaciones son medianamente influyentes en el mismo sentido, es decir, las 2 están contribuyendo a aumentar el valor de la VP. Este efecto es mas marcado en la primera de ellas, pues al eliminar observaciones pequeñas se está produciendo un aumento en el tiempo de vida medio.

Para concluir, se presentan las siguientes 2 figuras en las que se representa al estimador K-M de la muestra original y los

estimadores K-M para las 2 modificaciones sobre la observación número 18, las cuales nos van a permitir apreciar el cambio que sufre este estimador pues, como se mostró arriba dicha observación es la que tiene mayor influencia.

Figura 9. Estimador Kaplan-Meier para una muestra de tamaño 20 considerando todas las observaciones ($\hat{S}^0(t)$) y con un cambio de estado para la observación número 18 ($\hat{S}^0_{18}(t)$).

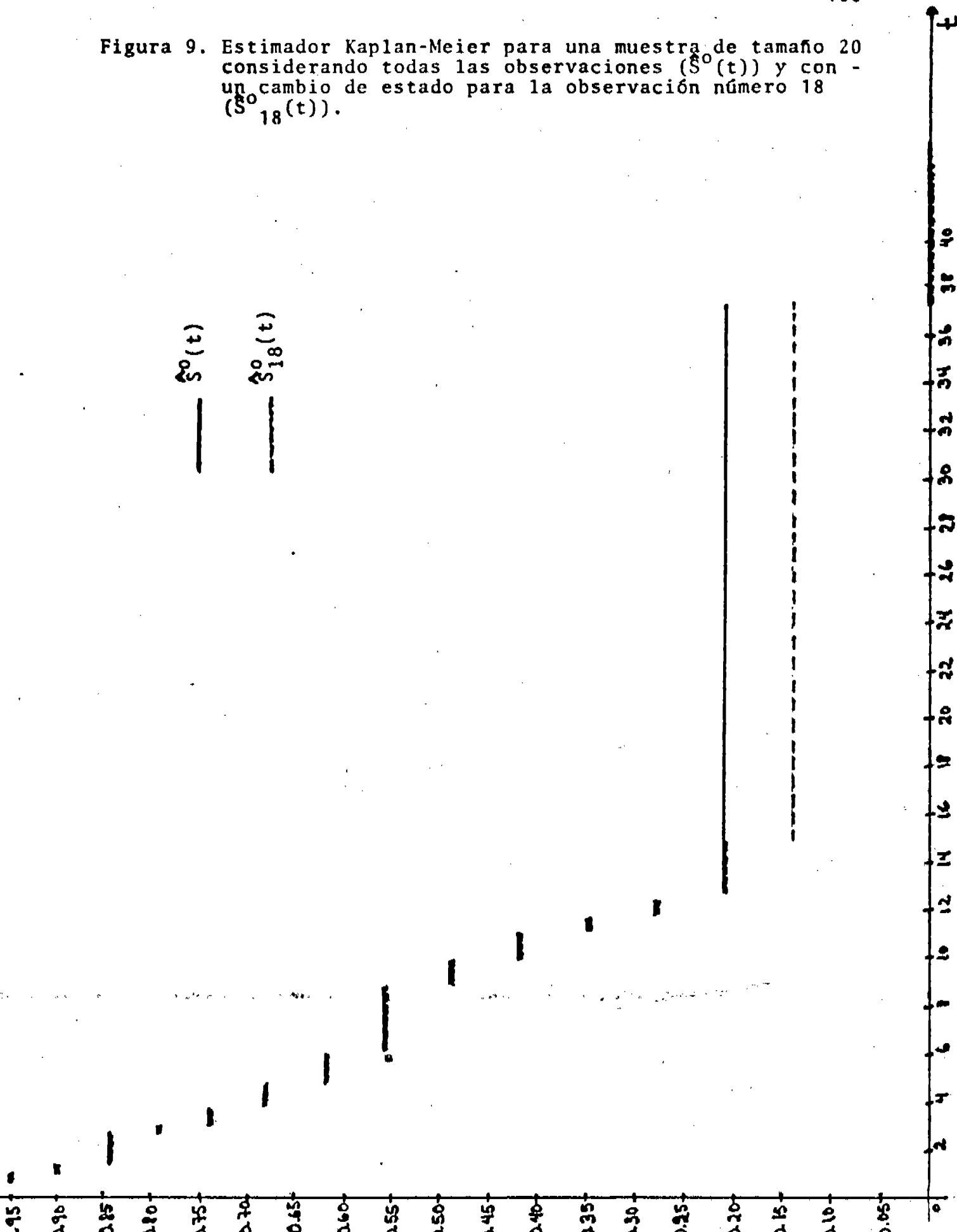
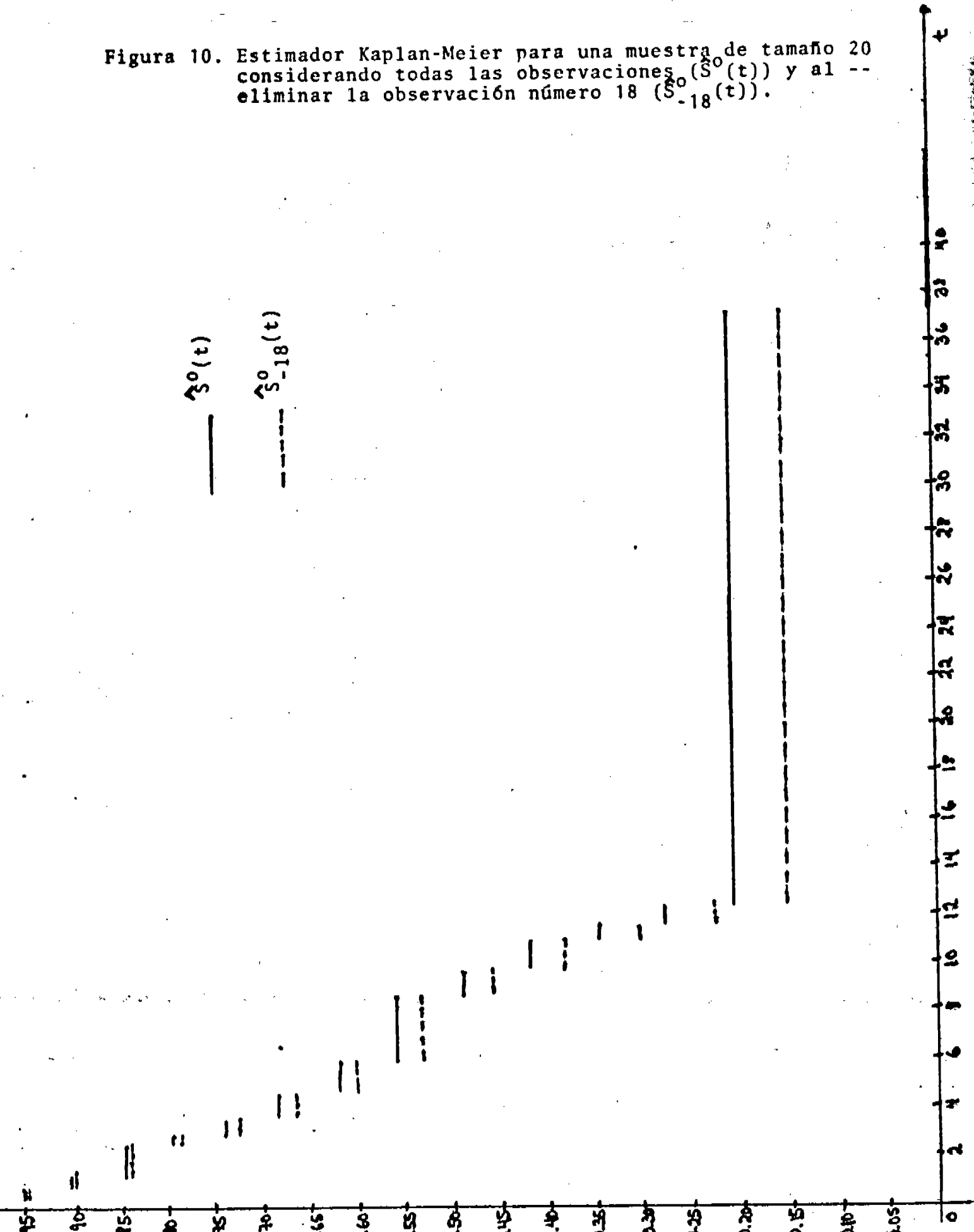


Figura 10. Estimador Kaplan-Meier para una muestra de tamaño 20 considerando todas las observaciones ($\hat{S}_0^0(t)$) y al -- eliminar la observación número 18 ($\hat{S}_{-18}^0(t)$).



5. CONCLUSIONES

La expresión que ha sido propuesta en esta tesis como medida de influencia tiene una forma simple y permite estudiar fácilmente la influencia de las observaciones muestrales tanto censuradas como no censuradas, así como la influencia de posibles observaciones contaminantes.

A pesar de no tener una expresión de la función de distribución del valor promedio del estimador Kaplan-Meier (VP), con la curva de influencia propuesta se obtuvieron, en todas las situaciones estudiadas, los resultados que se deseaban con respecto a la detección del efecto de las observaciones en el estimador Kaplan-Meier. Aunque estos resultados solo fueron estudiados para el caso exponencial, se espera que la forma propuesta se comporte de razonablemente para otros modelos de tiempo de falla.

Una ventaja adicional de la expresión analizada es que no depende de λ , el inverso del tiempo de falla medio esperado, en el caso exponencial, sino de la razón λ/θ , en donde θ es el parámetro de contaminación. Lo anterior nos evita el problema de tener una función de distribución para cada valor de λ posible y además nos da una posibilidad de generalización para cada valor constante de λ/θ .

Cabe mencionar que la función de influencia considerada se obtuvo después de haber propuesto y estudiado otras dos alternativas. La primera de ellas se refería a analizar la influencia utilizando como estadística la máxima magnitud de los saltos del estimador Kaplan-Meier obtenido, pero tal propuesta no

fue satisfactoria, entre otras cosas, por no tomar en cuenta la posición de dicho salto, la cual es importante sobre todo para las observaciones contaminadas. La segunda propuesta fue considerar como estadística el area bajo la curva del estimador Kaplan-Meier. Aunque esta alternativa no tenía el mismo defecto de la primera dependía del parámetro desconocido λ , por lo cual sería necesario tener otra muestra de la misma población, pero sin contaminación y censura para obtener una buena estimación de este parámetro.

Finalmente, un pequeño inconveniente del método propuesto en este trabajo es que para su cálculo se debe contar con la función de distribución del VP para la situación dada por la muestra particular obtenida. Sin embargo, para evitar este inconveniente se está actualmente estudiando el comportamiento de una curva de influencia diferente que consiste, esencialmente, en comparar directamente el cambio relativo que sufre el VP al realizar un cambio en cada una de las observaciones. Es decir, la nueva expresión es

$$CI_1 = \frac{VP_1 - c VP_1}{VP_1}$$

Los primeros análisis que se han realizado con esta expresión muestran resultados equivalentes con los obtenidos a partir de la expresión utilizada en este trabajo por lo que, al parecer, también produce resultados satisfactorios.

6. APENDICES

6.1 APENDICE 1

La existencia de la curva de influencia de un estimador depende de las características y propiedades matemáticas de la funcional que determine el parámetro considerado, dentro de estas propiedades se encuentra el que la funcional sea lineal y derivable en algún sentido. En este apéndice se dará una breve idea de tales requerimientos.

Supongamos que se tiene una métrica d_* o por lo menos una pseudo-métrica, en el espacio M de las medidas de probabilidad, tal que:

1. Es compatible con la topología (débil) de M , en el sentido que:

$$\{ F \mid d_*(G, F) < \epsilon \}$$

es abierto para toda $\epsilon > 0$.

2. Es compatible con la estructura afín de M , es decir, Si

$$F_\gamma = (1 - \gamma)F_0 + \gamma F_1$$

entonces

$$d_*(F_t, F_s) = o(|t-s|) \quad \forall t, s$$

Ejemplos de métricas que cumplen estas 2 características son:

- Métrica de Lévy

$$d_L(F, G) = \inf \{ \epsilon > 0 \mid \forall x \quad F(x-\epsilon) - \epsilon \leq G(x) \leq F(x+\epsilon) + \epsilon \}$$

- Métrica de Prohorov

$$d_p(F, G) = \inf \{ \epsilon > 0 \mid F(A) \leq G(A^\delta) + \epsilon \quad \forall A \in \mathcal{B} \}$$

donde

$$A^\delta = \left\{ x \in \mathbb{R}^n \mid \inf_{y \in A} d(x, y) \leq \delta \right\}$$

es una δ -vecindad cerrada de A .

- Métrica Lipschitz acotada. Con la suposición que la distancia d en \mathbb{R}^n está acotada por 1, en caso contrario se puede reemplazar por la métrica

$$\frac{d(x, y)}{1 + d(x, y)}$$

la métrica entonces está dada por

$$d_{BL}(F, G) = \sup \left| \int \psi dF - \int \psi dG \right|$$

donde el supremo se toma sobre todas las funciones ψ que satisfacen la condición Lipschitz

$$|\psi(x) - \psi(y)| \leq d(x, y).$$

- Métrica de variación total.

$$d_{VT}(F, G) = \sup_{A \in \mathcal{A}} |F(A) - G(A)|$$

Hay que hacer notar que aquí F es usada tanto para la medida como para la función de distribución.

Definición 1.-

Se dice que una funcional estadística T es Fréchet diferenciable en F si y solo si existe una funcional lineal L de M en Θ que depende de F tal que para toda $G \in M$

$$|T(G) - T(F) - L(G-F)| = o(d_*(F, G))$$

donde $|\cdot|$ es la métrica en Θ espacio parametral.

Definición 2.-

Si T es una funcional estadística Fréchet diferenciable en toda $F \in M$ entonces se dice que T es Fréchet diferenciable en M .

Supongamos que para F fija existen L_1 y L_2 funcionales lineales que satisfacen la definición 1, entonces

$$|L_1(G-F) - L_2(G-F)|$$

$$= |T(G) - T(F) - L_2(G-F) - (T(G) - T(F) - L_1(G-F))|$$

$$\begin{aligned} & \leq |T(G) - T(F) - h_2(G-F)| \\ & \quad + |T(G) - T(F) - h_1(G-F)| \\ & = o(d_x(F, G)) \end{aligned}$$

por lo cual L es esencialmente única para cada $F \in M$.

Esta forma de derivación, es una generalización propuesta en 1925 por Fréchet, de la derivada clásica de funciones reales de variable real, es decir, de la siguiente proposición

Una función $f: \mathbb{R} \rightarrow \mathbb{R}$ tiene derivada m en el punto $a \in \mathbb{R}$ si y solo si

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \text{s.} \quad \left| \frac{f(x) - f(a)}{x - a} - m \right| \leq \varepsilon \quad \text{si} \quad |x - a| < \delta$$

o equivalentemente

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \text{s.} \quad |f(x) - f(a) - m(x-a)| \leq \varepsilon |x-a| \quad \text{si} \quad |x-a| < \delta$$

Un primer paso par la generalización, es considerar 2 espacios lineales normados E_1 y E_2 y una función $f: E_1 \rightarrow E_2$. La derivada de f en un punto $a \in E_1$ será definida por una transformación lineal $L (:E_1 \rightarrow E_2)$ tal que se satisface

$$\|f(x) - f(a) - L(x-a)\| \leq \epsilon \|x-a\| \quad \text{si} \quad \|x-a\| < \delta$$

Hay que notar que si $E_1 = E_2 = \mathbb{R}$ y $L: \mathbb{R} \rightarrow \mathbb{R}$ tal que $x \rightarrow mx$, se tiene la derivada usual.

Ahora si consideramos $E_1 = M$, $E_2 = \Theta$, $f=T$, tendremos entonces esencialmente la diferencial de Fréchet.

Teorema 1.-

Sea T una funcional Fréchet diferenciable en F , continua en una vecindad de F , entonces $L=T'(F)$ la derivada de Fréchet en F es una funcional lineal continua y acotada, y además se puede representar como

$$L(G-F) = \int \psi_F dG$$

con ψ_F continua y acotada que cumple con

$$\int \psi_F dF = 0$$

Desafortunadamente, el concepto de diferenciabilidad de Fréchet es muy fuerte y en muchos casos la derivada de Fréchet no existe o si existe es muy difícil de demostrarlo. Por lo anterior surgió un concepto más débil de diferenciabilidad de funcionales, el cual es más sencillo de trabajar y que para la existencia de la función de influencia es suficiente, este concepto es la

diferenciabilidad de Gateaux la cual también es llamada a veces la derivada de Volterra.

Definición 3.-

Se dice que una funcional estadística T es Gateaux diferenciable en F si y solo si existe una funcional lineal L (de M en Θ) que depende de F tal que para toda $G \in M$

$$\lim_{t \rightarrow 0} \frac{T(F_t) - T(F)}{t} = L(G - F)$$

con

$$F_t = (1-t)F + tG$$

Como puede verse, la derivada de Gateaux es un límite que se calcula sobre dominios convexos de M , es decir, considerando funciones de distribución cuya forma es

$$F + t(G - F) = (1-t)F + tG, \quad 0 \leq t \leq 1$$

Por lo tanto, dado el parámetro de la población de interés como una funcional de la función de distribución asociada, la curva de influencia teórica existirá siempre y cuando dicha funcional sea diferenciable en el sentido Gateaux.

6.2 APENDICE 2

Demostrar que $dF^u(t) = (1 - H(t)) dF^o(t)$

$$\begin{aligned} F^u(t) &= P[X_i \leq t, S_i = 1] \\ &= P[S_i = 1] - P[X_i \geq t, S_i = 1], \end{aligned}$$

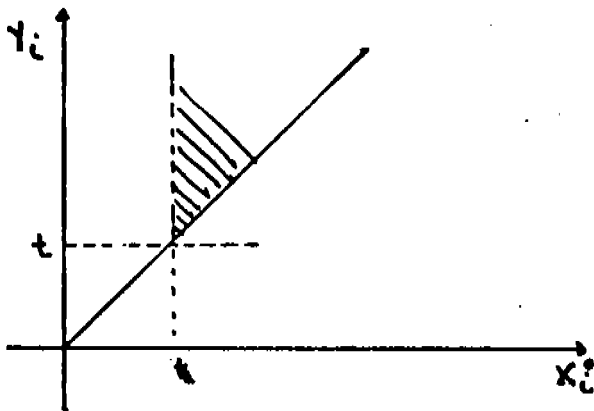
pero

$$\begin{aligned} P[S_i = 1] &= P[X_i^o \leq Y_i] = \int_0^{\infty} \int_x^{\infty} f^o(x) h(y) dy dx \\ &= \int_0^{\infty} f^o(x) (1 - H(x)) dx = \int_0^{\infty} (1 - H(x)) dF^o \end{aligned}$$

y

$$\begin{aligned} P[X_i \geq t, S_i = 1] &= P[X_i \geq t, X_i^o \leq Y_i] \\ &= P[\min(X_i^o, Y_i) \geq t, X_i^o \leq Y_i] \\ &= P[X_i^o \geq t, Y_i \geq t, X_i^o \leq Y_i] \end{aligned}$$

El evento que nos permite evaluar la última expresión está representado por el área rayada en la figura siguiente.



Por lo que

$$\begin{aligned}
 P[X_i > t, \delta_i = 1] &= \int_t^\infty \int_t^r f^0(s) h(r) ds dr \\
 &= \int_t^\infty h(r) \left[\int_t^r f^0(s) ds \right] dr \\
 &= \int_t^\infty h(r) [F^0(r) - F^0(t)] dr = \int_t^\infty h(r) F^0(r) dr \\
 &\quad - \int_t^\infty h(r) F^0(t) dt = \int_t^\infty F^0(r) dH - F^0(t) [1 - H(t)].
 \end{aligned}$$

Integrando por partes, tenemos que

$$\begin{aligned}
 P[X_i > t, \delta_i = 1] &= F^0(r)H(r) \Big|_t^\infty - \int_t^\infty H(r) dF^0 - F^0(t) \\
 &\quad + F^0(t)H(t) \\
 &= 1 - F^0(t) - \int_t^\infty H(r) dF^0 \\
 &= \int_t^\infty dF^0 - \int_t^\infty H(r) dF^0 = \int_t^\infty (1 - H(r)) dF^0.
 \end{aligned}$$

Por lo tanto

$$\begin{aligned}
 P[X_i \leq t, \delta_i = 1] &= \int_0^\infty (1 - H(x)) dF^0 - \int_t^\infty (1 - H(x)) dF^0 \\
 &= \int_0^t (1 - H(x)) dF^0
 \end{aligned}$$

y entonces

$$\begin{aligned}
 F^u(t) &= \int_0^t (1 - H(x)) dF^0 \\
 dF^u(t) &= (1 - H(t)) dF^0(t).
 \end{aligned}$$

6.3 APENDICE 3

Demostrar que
$$\sum_{i=1}^n \binom{n}{i} \frac{(-1)^{i-1}}{i} = \sum_{i=1}^n \frac{1}{i}$$

La demostración se hará por inducción.

Para $n=1$, tenemos que

$$\sum_{i=1}^1 \binom{1}{i} \frac{(-1)^{i-1}}{i} = \binom{1}{1} \frac{(-1)^0}{1} = 1$$

y

$$\sum_{i=1}^1 \frac{1}{i} = 1$$

Suponiendo que se cumple para $k-1$, es decir,

$$\sum_{i=1}^{k-1} \binom{k-1}{i} \frac{(-1)^{i-1}}{i} = \sum_{i=1}^{k-1} \frac{1}{i},$$

se demostrará para k .

$$\begin{aligned} \sum_{i=1}^k \binom{k}{i} \frac{(-1)^{i-1}}{i} &= \sum_{i=1}^{k-1} \binom{k}{i} \frac{(-1)^{i-1}}{i} + \binom{k}{k} \frac{(-1)^{k-1}}{k} \\ &= \sum_{i=1}^{k-1} \binom{k}{i} \frac{(-1)^{i-1}}{i} + \frac{(-1)^{k-1}}{k}. \end{aligned}$$

Ahora bien,

$$\begin{aligned} \binom{k}{i} - \binom{k-1}{i} &= \frac{k!}{i!(k-i)!} - \frac{(k-1)!}{i!(k-1-i)!} \\ &= \frac{k(k-1)!}{i!(k-i)!} - \frac{(k-1)!(k-i)}{i!(k-i)!} \\ &= \frac{(k-1)!}{i!(k-i)!} [k - (k-i)] = \frac{(k-1)!}{i!(k-i)!} i \end{aligned}$$

$$= \frac{k!}{i!(k-i)!} \left(\frac{i}{k}\right) = \binom{k}{i} \frac{i}{k}$$

por lo que

$$\binom{k}{i} = \binom{k-1}{i} + \frac{i}{k} \binom{k}{i}.$$

Entonces

$$\begin{aligned} \sum_{i=1}^{k-1} \binom{k}{i} \frac{(-1)^{i-1}}{i} &= \sum_{i=1}^{k-1} \left[\binom{k-1}{i} + \binom{k}{i} \frac{i}{k} \right] \frac{(-1)^{i-1}}{i} \\ &= \sum_{i=1}^{k-1} \binom{k-1}{i} \frac{(-1)^{i-1}}{i} + \frac{1}{k} \sum_{i=1}^{k-1} \binom{k}{i} (-1)^{i-1} \\ &= \sum_{i=1}^{k-1} \frac{1}{i} - \frac{1}{k} \sum_{i=1}^{k-1} \binom{k}{i} (-1)^i \\ &= \sum_{i=1}^{k-1} \frac{1}{i} - \frac{1}{k} \left[\sum_{i=0}^k \binom{k}{i} (-1)^i - \binom{k}{0} (-1)^0 - \binom{k}{k} (-1)^k \right] \\ &= \sum_{i=1}^{k-1} \frac{1}{i} - \frac{1}{k} [0 - 1 - (-1)^k] \\ &= \sum_{i=1}^{k-1} \frac{1}{i} + \frac{1}{k} - \frac{(-1)^{k-1}}{k} = \sum_{i=1}^k \frac{1}{i} = \frac{(-1)^{k-1}}{k} \end{aligned}$$

Finalmente

$$\sum_{i=1}^k \binom{k}{i} \frac{(-1)^{i-1}}{i} = \sum_{i=1}^k \frac{1}{i} - \frac{(-1)^{k-1}}{k} + \frac{(-1)^{k-1}}{k} = \sum_{i=1}^k \frac{1}{i}.$$

6.4 APENDICE 4

Simplificación de $\sum_{i=1}^n \binom{n}{i} \frac{(-1)^{i-1} a_0^i}{x}$.

puesto que $\frac{a_0^i}{x} = \int_0^{a_0} x^{i-1} dx$, se tiene que

$$\begin{aligned} \sum_{i=1}^n \binom{n}{i} \frac{(-1)^{i-1} a_0^i}{x} &= \sum_{i=1}^n \binom{n}{i} (-1)^{i-1} \int_0^{a_0} x^{i-1} dx \\ &= \int_0^{a_0} \sum_{i=1}^n \binom{n}{i} (-1)^{i-1} x^{i-1} dx \end{aligned}$$

pero

$$\begin{aligned} \sum_{i=1}^n \binom{n}{i} (-1)^{i-1} x^{i-1} &= -\frac{1}{x} \sum_{i=1}^n \binom{n}{i} (-1)^i x^i = -\frac{1}{x} \left[\sum_{i=0}^n \binom{n}{i} (-1)^i x^i - 1 \right] \\ &= -\frac{1}{x} [(1-x)^n - 1] = \frac{1 - (1-x)^n}{x}. \end{aligned}$$

Por lo que

$$\begin{aligned} \sum_{i=1}^n \binom{n}{i} \frac{(-1)^{i-1} a_0^i}{x} &= \int_0^{a_0} \frac{1 - (1-x)^n}{x} dx \\ &= \int_1^{1-a_0} \frac{1-y^n}{1-y} (-dy) = \int_{1-a_0}^1 \frac{1-y^n}{1-y} dy \\ &= \int_{p_0}^1 (1+y+y^2+\dots+y^{n-1}) dy \\ &= \left(y + \frac{y^2}{2} + \frac{y^3}{3} + \dots + \frac{y^n}{n} \Big|_{p_0}^1 \right) \\ &= 1 - p_0 + \frac{1}{2} - \frac{p_0^2}{2} + \frac{1}{3} - \frac{p_0^3}{3} + \dots + \frac{1}{n} - \frac{p_0^n}{n}. \end{aligned}$$

$$= \sum_{i=1}^n \left(\frac{1 - \rho_0^i}{i} \right).$$

Simplificación de $\sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(i\lambda + \theta)}$.

$$\begin{aligned} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(i\lambda + \theta)} &= \frac{1}{\lambda} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i (i\lambda) q_0^i}{(i\lambda + \theta) i} \\ &= \frac{1}{\lambda} \sum_{i=1}^{n-1} \binom{n-1}{i} (-1)^i \left(\frac{i\lambda}{i\lambda + \theta} \right) \frac{q_0^i}{i} \\ &= \frac{1}{\lambda} \sum_{i=1}^{n-1} \binom{n-1}{i} (-1)^i \left(1 - \frac{\theta}{i\lambda + \theta} \right) \frac{q_0^i}{i} \\ &= \frac{1}{\lambda} \left[\sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{i} - \theta \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(i\lambda + \theta) i} \right] \\ &= \frac{1}{\lambda} \left[- \sum_{i=1}^{n-1} \left(\frac{1 - \rho_0^i}{i} \right) - \theta \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(i\lambda + \theta) i} \right]. \end{aligned}$$

ahora bien,

$$\begin{aligned} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(i\lambda + \theta) i} &= \frac{1}{\lambda} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i (i\lambda) q_0^i}{(i\lambda + \theta) i^2} \\ &= \frac{1}{\lambda} \sum_{i=1}^{n-1} \binom{n-1}{i} (-1)^i \left(\frac{i\lambda}{i\lambda + \theta} \right) \frac{q_0^i}{i^2} \\ &= \frac{1}{\lambda} \sum_{i=1}^{n-1} \binom{n-1}{i} (-1)^i \left(1 - \frac{\theta}{i\lambda + \theta} \right) \frac{q_0^i}{i^2} \\ &= \frac{1}{\lambda} \left[\sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{i^2} - \theta \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(i\lambda + \theta) i^2} \right] \end{aligned}$$

pero

$$\begin{aligned}
 \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda q_0^\lambda}{\lambda^2} &= \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda}{\lambda} \left(\frac{q_0^\lambda}{\lambda} \right) \\
 &= \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda}{\lambda} \int_0^{q_0} x^{\lambda-1} dx \\
 &= \int_0^{q_0} \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda x^{\lambda-1}}{\lambda} dx \\
 &= \int_0^{q_0} -\frac{1}{x} \left(\sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^{\lambda-1} x^\lambda}{\lambda} \right) dx \\
 &= \int_0^{q_0} -\frac{1}{x} \left(\sum_{\lambda=1}^{n-1} \frac{1 - (1-x)^\lambda}{\lambda} \right) dx \\
 &= -\sum_{\lambda=1}^{n-1} \frac{1}{\lambda} \left(\int_0^{q_0} \frac{1 - (1-x)^\lambda}{x} dx \right) \\
 &= -\sum_{\lambda=1}^{n-1} \frac{1}{\lambda} \sum_{r=1}^{\lambda} \left(\frac{1 - p_0^r}{r} \right),
 \end{aligned}$$

y

$$\begin{aligned}
 \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda q_0^\lambda}{(\lambda\lambda + \theta)\lambda^2} &= \frac{1}{\lambda} \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda (\lambda\lambda)}{(\lambda\lambda + \theta)\lambda^3} q_0^\lambda \\
 &= \frac{1}{\lambda} \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} (-1)^\lambda \left(\frac{\lambda\lambda}{\lambda\lambda + \theta} \right) \frac{q_0^\lambda}{\lambda^3} \\
 &= \frac{1}{\lambda} \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} (-1)^\lambda \left(1 - \frac{\theta}{\lambda\lambda + \theta} \right) \frac{q_0^\lambda}{\lambda^3} \\
 &= \frac{1}{\lambda} \left[\sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda q_0^\lambda}{\lambda^3} - \theta \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda q_0^\lambda}{(\lambda\lambda + \theta)\lambda^3} \right]
 \end{aligned}$$

además

$$\begin{aligned}
 \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda q_0^\lambda}{\lambda^3} &= \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda}{\lambda^2} \left(\frac{q_0^\lambda}{\lambda} \right) \\
 &= \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda}{\lambda^2} \int_0^{q_0} x^{\lambda-1} dx \\
 &= \int_0^{q_0} \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda}{\lambda^2} x^{\lambda-1} dx \\
 &= \int_0^{q_0} \frac{1}{x} \left(\sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda x^\lambda}{\lambda} \right) dx \\
 &= \int_0^{q_0} \frac{1}{x} \left(- \sum_{\lambda=1}^{n-1} \frac{1}{\lambda} \sum_{r=1}^{\lambda} \frac{1 - (1-x)^r}{r} \right) dx \\
 &= - \sum_{\lambda=1}^{n-1} \frac{1}{\lambda} \sum_{r=1}^{\lambda} \frac{1}{r} \int_0^{q_0} \frac{1 - (1-x)^r}{x} dx \\
 &= - \sum_{\lambda=1}^{n-1} \frac{1}{\lambda} \sum_{r=1}^{\lambda} \frac{1}{r} \sum_{s=1}^r \left(\frac{1 - p_0^s}{s} \right).
 \end{aligned}$$

Por lo que finalmente se tiene

$$\begin{aligned}
 \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda q_0^\lambda}{\lambda \lambda + \theta} &= \frac{1}{\lambda} \left[- \sum_{\lambda=1}^{n-1} \left(\frac{1 - p_0^\lambda}{\lambda} \right) - \theta \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda q_0^\lambda}{(\lambda \lambda + \theta) \lambda} \right] \\
 &= - \frac{1}{\lambda} \sum_{\lambda=1}^{n-1} \left(\frac{1 - p_0^\lambda}{\lambda} \right) - \frac{\theta}{\lambda} \sum_{\lambda=1}^{n-1} \binom{n-1}{\lambda} \frac{(-1)^\lambda q_0^\lambda}{(\lambda \lambda + \theta) \lambda}
 \end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{\lambda} \sum_{i=1}^{n-1} \left(\frac{1-p_0^i}{\lambda} \right) - \frac{\theta}{\lambda} \left[\frac{1}{\lambda} \left(\sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{\lambda^2} \right. \right. \\
&\quad \left. \left. - \theta \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(\lambda\lambda + \theta) i^2} \right) \right] \\
&= -\frac{1}{\lambda} \sum_{i=1}^{n-1} \left(\frac{1-p_0^i}{\lambda} \right) - \frac{\theta}{\lambda^2} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{\lambda^2} \\
&\quad - \frac{\theta^2}{\lambda^2} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(\lambda\lambda + \theta) i^2} \\
&= -\frac{1}{\lambda} \sum_{i=1}^{n-1} \left(\frac{1-p_0^i}{\lambda} \right) - \frac{\theta}{\lambda^2} \left(- \sum_{i=1}^{n-1} \frac{1}{i} \sum_{r=1}^i \left(\frac{1-p_0^r}{r} \right) \right) \\
&\quad - \frac{\theta^2}{\lambda^2} \left(\left(\frac{1}{\lambda} \right) \left(\sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{\lambda^3} - \theta \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(\lambda\lambda + \theta) i^3} \right) \right) \\
&= -\frac{1}{\lambda} \sum_{i=1}^{n-1} \left(\frac{1-p_0^i}{\lambda} \right) + \frac{\theta}{\lambda^2} \sum_{i=1}^{n-1} \frac{1}{i} \sum_{r=1}^i \left(\frac{1-p_0^r}{r} \right) \\
&\quad - \frac{\theta^2}{\lambda^3} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{i^3} + \frac{\theta^3}{\lambda^3} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(\lambda\lambda + \theta) i^3} \\
&= -\frac{1}{\lambda} \sum_{i=1}^{n-1} \left(\frac{1-p_0^i}{\lambda} \right) + \frac{\theta}{\lambda^2} \sum_{i=1}^{n-1} \frac{1}{i} \sum_{r=1}^i \left(\frac{1-p_0^r}{r} \right) \\
&\quad + \frac{\theta^2}{\lambda^3} \sum_{i=1}^{n-1} \frac{1}{i} \sum_{r=1}^i \frac{1}{r} \sum_{s=1}^r \left(\frac{1-p_0^s}{s} \right) \\
&\quad + \frac{\theta^3}{\lambda^3} \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{(-1)^i q_0^i}{(\lambda\lambda + \theta) i^3} .
\end{aligned}$$

7. REFERENCIAS

Andrews, D.F., Bichel, P.J., et.al. (1972)

Robust Estimates of Location: Survey and Advances.
Princeton, N.J.. Princeton University Press.

Breslow, N.E. (1975)

Analysis of survival data under the proportional hazards
model.

Int. Stat. Rev., Vol. 43, 45-58.

Breslow, N.E., y Crowley, J. (1974)

A large sample study of the table and product limit
estimates under censorship.

Ann. Statist., Vol. 2, 437-453.

Brown, A.L., y Page, A. (1970)

Elements of functional analysis.

London. Van Nostrand Reinhold Company.

Cook, R.D. (1977)

Detection of influential observations in linear regression.

Technometrics, Vol. 19, 15-18.

Cook, R. D. (1979)

Influential observations in linear regression.

J. Amer. Statist. Assoc., Vol. 74, 169-174.

Cook, R. D., y Weisberg, S. (1980)

Characterizations of an empirical influence function for detecting influential cases in regression.

Technometrics, Vol. 22, 495-508.

Cox, D. R. (1972)

Regression models with life tables.

J. Roy. Statist. Soc., Ser. B, Vol. 34, 187-200.

Devlin, S. J., Gnanadesikan, R., y Kettenring, J. R. (1975)

Robust estimation and outlier detection with correlation coefficients.

Biometrika, Vol. 62, 531-545.

Draper, N. R., y John, J. A. (1981)

Influential observations and outliers in regression.

Technometrics, Vol. 23, 21-26.

Efron, B. (1981)

Nonparametric estimates of standar error: The jackknife, the bootstrap and other methods.
Biometrika, Vol. 68, 589-599.

Efron, B. y Gong, G. (1983)

A leisurely look at the bootstrap, the jackknife and cross-validation.
The american statistician, Vol. 37, 36-48.

Everitt, B. S. y Hand, D. J. (1981)

Finite mixture distributions-monographs on applied probability and statistics.
London, Chapman and Hall.

Filippova, A. A. (1962)

Mises' Theorem on the asymptotic behavior of functionals of empirical distribution functions and its statistical applications.
Theory of Probatility and its Applications, Vol. 7, 24-57.

Johnson, N. L., y Kotz, S. (1970)

Distributions in Statistics. Continuous univariate distributions. Vol. 2.
Boston. Houghton Mifflin.

Hall, G. J., Rogers W. y Pregibon, D. (1982)

Outliers matter in survival analysis.

The rand paper series, No. P-6761.

Hampel, F. R. (1971)

A general qualitative definition of robustness.

Ann. Math. Statist., Vol. 42, 1887-1896.

Hampel, F. R. (1974)

The influence curve and its role in robust estimation.

J. Amer. Statist. Assoc., Vol. 69, 383-393.

Hampel, F. R. (1978)

Modern trends in the theory of robustness.

Math. Oper. Statis., Ser. Statistics, Vol. 9, 425-442.

Hastings, N. A. J. y Peacock, J. A. (1975)

Statistical Distributions.

London. Butterworths.

Hinkley, D.V. (1977)

Jackknifing in unbalanced situations.

Technometrics, Vol. 19, 285-292.

Hogg, R.V. (1979)

An introduction to robust estimation en Robustness in
Statistics. Edit. Launer y Wilkinson. 1-17.

New York. Academic Press.

Huber, P.J. (1964)

Robust estimation of a location parameter.

Ann. Math. Statist., Vol. 35, 73-101.

Huber, P.J. (1977)

Robust Statistics.

Philadelphia. SIAM.

Kalbfleisch, J.D. y Prentice, R.L. (1980)

The Statistical Analysis of Failure Time Data.

New York. John Wiley and Sons.

Kaplan, E.L., y Meier, P. (1958)

Nonparametric estimation from incomplete observations.

J. Amer. Statist. Assoc., Vol. 53, 457-481.

Miller, R.G. (1974)

The jackknife—a review.

Biometrika, Vol. 61, 1-15.

Peterson, A.V. (1977)

Expressing the Kaplan-Meier estimator as a function of empirical sub-survival functions.

J. Amer. Statist. Assoc., Vol. 32, 854-858.

Reid, N. (1979)

Influence functions for censored data.

Ph.D. thesis, Department of Statistics, Stanford University.

Reid, N. (1981)

Influence functions for censored data.

Ann. Statist., Vol. 9, 78-92.

Sarhan, A. y Greenberg, B. (1962)

Contributions to order statistics. Vol. II

New York. John Wiley and Sons.

Von Mises, R. (1947)

On the asymptotic distribution of differentiable
statistical functions.

Ann. Math. Statist., Vol. 18, 309-348.

Von Mises, R. y Geiringer, H. (1964)

Mathematical Theory of Probability and Statistics.

New York. Academic Press.

Weisberg, S. (1980)

Applied Linear Regression.

New York. John Wiley and Sons.