

**UNIVERSIDAD NACIONAL AUTONOMA DE
MEXICO**

**INSTITUTO DE INVESTIGACIONES EN
MATEMATICAS APLICADAS Y EN SISTEMAS
IIMAS**

**ANALISIS DESCRIPTIVO E INFERENCIAL DE LA
MORTALIDAD DE DELFINES, ASOCIADA A LA
ACTIVIDAD DE PESCA DEL ATUN EN UN VIAJE DE
PESCA EN EL PACIFICO ORIENTAL**

T E S I S I N A

**QUE PARA OBTENER EL DIPLOMA DE LA
ESPECIALIZACION EN ESTADÍSTICA APLICADA**

P R E S E N T A:

ARAVID TORRES GARDUÑO

ENERO 2009



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.

INDICE

INTRODUCCION.....	1
1. METODOS.	
1.1 Antecedentes	
1.1. Procedimiento de pesca del atún en asociación con delfines, utilizando red de cerco.....	3
1.1.2 Información sobre las variables y observaciones.....	6
1.1.3 Grupo de trabajo y del grupo de validación.....	7
1.2 Regresión logística	
1.2.1 El modelo logístico y la transformación logit.....	8
1.2.2 Cociente de momios (Odds Ratio).....	10
1.2.3 Interpretaciones del OR.....	13
1.2.4 Intervalo de confianza para el cociente de momios OR...14	
1.2.5 Método de Estimación de los coeficientes β_j en el modelo logístico	
1.2.5.1 Método de Máxima Verosimilitud.....	15
1.2.5.1.1 Estimación condicional y no condicional...16	
1.2.5.2 Método de ecuaciones de estimación generalizada (GEE).....	17
1.2.6 Prueba de Wald y Rao para la selección de variables....18	
1.2.6.1 Estadístico de Wald.....	18
1.2.6.2 Estadístico Rao	20
1.2.6.3 Pasos seguidos en el proceso de selección de variables utilizando el estadístico de Wald y Rao.....	21

1.2.7 Medidas de Ajuste del Modelo.....	22
1.2.7.1 Estadístico del cociente de verosimilitudes	24
1.2.7.2 Prueba de Hoemer-Lemeshow.....	25
1.2.7.3 Curva ROC.....	26
1.2.7.4 Sensibilidad y Especificidad.....	28
2. RESULTADOS Y DISCUSION	
2.1. Regresión logística	30
2.2 Ajuste del Modelo	
2.2.1 Calibración.....	33
2.2.2 Discriminación.....	34
2.2.3 Reproducibilidad.....	36
2.2.4 Transportabilidad.....	37
3. CONCLUSIONES.....	38
4. BIBLIOGRAFIA.....	40

I. INTRODUCCION

La mortalidad de los delfines debido a la pesca del atún es un fenómeno que ganó atención por el embargo al atún exportado de México hacia Estados Unidos. El motivo oficial de dicho embargo fue la matanza indiscriminada de delfines por parte de la flota atunera mexicana. A partir de este antecedente se crea el Acuerdo para la Conservación de Delfines ("el Acuerdo de La Jolla del 17 de junio de 1992"), donde se estableció el Programa Internacional para la Conservación de Delfines (PICD). El objetivo principal del Acuerdo, según la Comisión Interamericana del atún Tropical (CIAT), "...fue reducir la mortalidad de delfines en la pesquería cerquera sin perjudicar los recursos atuneros de la región y las pesquerías que dependen de los mismos."¹ Posteriormente se implementó el Acuerdo sobre el Programa Internacional para la Conservación de los Delfines (APICD), que de acuerdo con la CIAT "han logrado un éxito espectacular, demostrado por la reducción en la mortalidad de delfines incidental a la pesca". Pero, si bien los datos presentados por la CIAT en su página de internet (<http://www.iattc.org/DolphinSafeSPN.htm>) reflejan una drástica reducción en el número de delfines muertos, no demuestran que el programa sea el responsable de dicha reducción. Para validar las afirmaciones de la CIAT es necesario: Identificar los factores en la actividad pesquera que inciden en la mortalidad de los delfines y verificar que estos han sido modificados a lo largo del tiempo, de tal manera que han dado por resultado la disminución en la mortalidad de los delfines.

En este trabajo se determinan y cuantifican algunos factores de riesgo en la pesca del atún *Tunnus albacares* que inciden en la mortalidad del delfín

¹ www.iattc.org/PDFFiles/CIATq024SPN.pdf

Stenella attenuata y *Stenella longirostris* cuando delfines y atunes se encuentran asociados. Para lograr la identificación y cuantificación de riesgos se utilizó la regresión logística, que describe la probabilidad de muerte de los delfines como una función de factores de riesgo detectados dentro de la actividad de pesca del atún.

En el primer capítulo se presenta una descripción general de la realización de la faena de pesca de atún cuando se encuentra asociado con delfines, presentando los posibles factores de riesgo en el contexto de la actividad, a fin de entender la manera en que afectan a los delfines. También se describen los factores considerados como candidatos a factores de riesgo y la metodología estadística de la regresión logística incluyendo la justificación de su uso, el método para determinar si un factor es o no considerado de riesgo, la manera en que se cuantifica dicho riesgo y finalmente los criterios y procedimientos para medir la representatividad del modelo matemático obtenido.

En el capítulo 2 se presentan los factores de riesgo en la actividad pesquera que fueron encontrados mediante la regresión logística, el grado en que cada factor detectado incide en la probabilidad de muerte de un delfín, la validez del modelo matemático obtenido y la discusión sobre los resultados obtenidos.

Finalmente en la conclusión se plantean las cuestiones resueltas en este estudio y las propuestas para futuras investigaciones, así como sugerencias para su realización, basadas en la experiencia del presente trabajo.

1. METODOS

1.1 Antecedentes

1.1.1 Procedimiento de pesca del atún en asociación con delfines, utilizando red de cerco.

En este apartado se presenta una descripción general de cómo se realiza una faena de pesca de atún cuando está asociado con delfines, explicando el papel de los diferentes factores de riesgo analizados en este estudio.

El trabajo en la pesca del atún principia cada día, con horas de observación mediante lentes de alta potencia y frente a la pantalla del radar anticolidión, al aclarar el día, se buscan señales en el agua que permitan dar con un cardumen de atunes por ejemplo: agregaciones de pájaros o salpicaduras de delfines al nadar contra las olas entre otras. Los esfuerzos de búsqueda en alta mar se centran en 2 especies “*S. attenuata*” y “*S. longirostris*”. Una vez seleccionado el grupo de delfines asociado a un cardumen de atunes, el barco se aproxima hasta una distancia de 3 a 4 Km. y de ahí se bajan al mar lanchas rápidas que se encargan de arrear a los delfines para que naden en círculo y formen un grupo lo suficientemente compacto para cercarlos con la red. (Salvador Macotela, comunicación personal, 19 de octubre de 2004).

El proceso de lance de la red y recuperación, llamado *faena de pesca*, inicia cuando el lanchón es bajado del barco para sostener un extremo de la red, mientras tanto el barco se encarga de extenderla en forma de

círculo alrededor de los delfines y de los atunes, al finalizar la maniobra de encierro con el encuentro entre el barco y el lanchón unos minutos después, el extremo de la red que sostuvo el lanchón se pasa al barco. En este punto los delfines se han percatado de que se ha cerrado el cerco y generalmente tienden a nadar próximos al centro. Inmediatamente, en el barco, se inicia el cobrado del cable mediante el cual se cierra la red por el fondo, hasta que quede cerrada y asegurada. Ahora da inicio el acomodo de la red en el barco que consiste en comenzar a recoger la red del agua y subirla a bordo del barco. (Salvador Macotela, comunicación personal, 19 de octubre de 2004).

Cuando se han subido al barco aproximadamente $2/3$ partes de la longitud de la red, el barco inicia la *maniobra de liberación*. En dicha maniobra el barco es remolcado por el lanchón en reversa, entonces la red forma el canal de liberación y al hundirse las boyas en el extremo del canal, se permite la salida de la mayoría de los delfines por la superficie mientras estos son arreados desde una balsa hacia el extremo del canal. Una vez concluida dicha maniobra, se sigue subiendo la red hasta formarse un bolso, del que se extraen los atunes. Cuando termina la extracción de atunes continua la extracción de la red del agua y se acomoda sobre el barco, finalizando la faena cuando el lanchón esta a bordo del barco. (Salvador Macotela, comunicación personal, 19 de octubre de 2004).

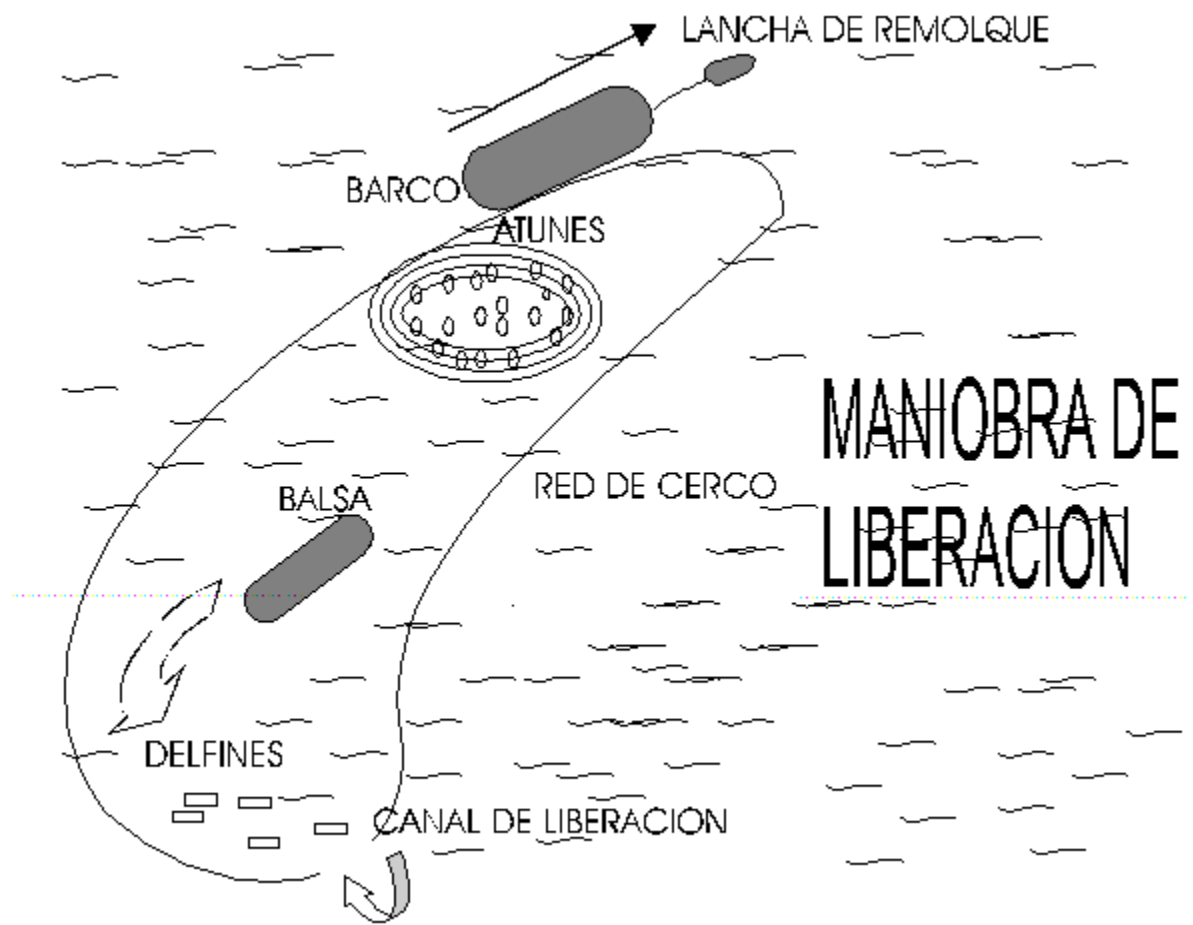


Figura 1. Diagrama de la maniobra de liberación

1.1.2 Información sobre las variables y observaciones.

Durante la actividad pesca de atún aleta amarilla, en 20 lances de pesca, a bordo de una embarcación atunera con bandera mexicana en el océano Pacífico y en un área de pesca cuya área de navegación comprendió entre los 123° W a los 142° W y de los 8° N a los 12° N se registró, en cada lance, no solo la mortalidad y sobrevivencia de los delfines involucrados, sino otros elementos parte de los cuales están descritos en la tabla 1.1 mismos que fueron considerados como posibles factores asociados a la mortalidad de los delfines. (Salvador Macotela, comunicación personal, 19 de octubre de 2004).

Nombre de la variable	Tipo de Variable	Descripción
Mortalidad	Dicotómica 0 Sobrevivió 1 Murió	Indica el estado para cada delfín al término de la faena de pesca
Presencia de colapsamiento del canal de liberación	Dicotómica 0 Ausencia 1 Presencia	Si hubo o no colapso del canal de liberación que impida la salida de los delfines.
Turno	Dicotómica 0 Matutino 1 Vespertino	El turno en que se realizó la maniobra de pesca
Distancia de la observación	Cuantitativa continua,	La lejanía del barco de un grupo de delfines asociado a un cardumen de

	expresada en millas náuticas ²	atunes al momento de ser vistos.
Duración de la maniobra de liberación	Cuantitativa continua, expresada en minutos	Tiempo que se invirtió para formar el canal de liberación.
Numero de Lance	Cualitativa	Indica el lance al que pertenece la observación correspondiente.

Tabla 1.1 Descripción de las variables registradas durante el viaje de pesca usadas en el presente trabajo.

Durante los 20 lances, se registró la mortalidad o supervivencia de un total de 54300 delfines, de los cuales, 53,245 delfines sobrevivieron y 1,055 delfines murieron.

1.1.3 Grupo de trabajo y del grupo de validación

El grupo de trabajo se refiere a las observaciones que serán usadas para generar el modelo y encontrar las relaciones entre el fenómeno de mortalidad y el resto de las variables estudiadas. El grupo de validación comprende a las observaciones a través de las cuales el modelo obtenido será evaluado en su capacidad para predecir la mortalidad o supervivencia de un delfín. Para formar el grupo de trabajo y el de validación basta dividir aleatoriamente la población en dos grupos³

² Una milla náutica equivale a 1.852 kilómetros.

³ http://www.hrc.es/bioest/Reglog_10.html

Dado que la proporción de delfines muertos es muy baja respecto al total de delfines, tomar una muestra aleatoria simple para conformar el grupo de trabajo a partir del registro total de delfines vivos y muertos no permitiría tener una muestra suficientemente grande de delfines muertos a fin de encontrar la relación entre el fenómeno de la mortalidad y los posibles factores asociados, de modo que, se consideró un tamaño de muestra a priori, tanto para delfines muertos como delfines vivos, superior a 100 individuos. Es decir, el grupo de trabajo consta de 116 registros de delfines vivos y 104 de delfines muertos y el grupo de validación de 53,129 registros de delfines vivos y 951 delfines muertos. En total la muestra consiste en 220 registros de delfines para el grupo de trabajo y los 54080 registros de delfines restantes formaron parte del grupo de validación del modelo.

1.2 Regresión logística

1.2.1 El modelo logístico y la transformación logit

La regresión logística permite modelar la probabilidad de ocurrencia de un evento que solo toma dos posibles valores, como función de una serie de variables explicativas. En este caso el evento es la muerte o supervivencia de cada delfín y las variables explicativas son los factores descritos en la tabla 1.1. El modelo logístico permite no solo determinar los factores de riesgo asociados a la mortalidad de los delfines sino cuantificar el grado de influencia de estos sobre la variable de respuesta

La relación entre la mortalidad y sus posibles factores queda expresada por la siguiente ecuación:

$$P(R = 1 | X_1, X_2, \dots, X_n) = \frac{1}{1 + e^{-(\alpha + \sum \beta_j X_j)}} \quad (1.1)$$

donde:

R es la variable de mortalidad con dos posibles resultados 0 (supervivencia) y 1 (muerte).

β_j es el coeficiente correspondiente al factor de riesgo número j que tiene posiblemente alguna relación con la mortalidad.

x_j es el factor de riesgo número j posiblemente asociado a la mortalidad.

Solo con fines de simplificar la notación en la expresión 1.1 se realiza el siguiente cambio:

$$P(R = 1 | X_1, X_2, \dots, X_n) = P(X) \quad (1.2)$$

Entonces el modelo logístico de 1.1 también puede expresarse como:

$$P(X) = \frac{1}{1 + e^{-(\alpha + \sum \beta_j X_j)}} \quad (1.3)$$

La transformación logit denotada como $\text{logit } P(X)$, esta dada por

$$\log it P(X) = \ln \left[\frac{P(X)}{1 - P(X)} \right] \quad (1.4)$$

Mediante procedimientos algebraicos y usando a la expresión 1.3 y 1.4 se puede obtener que:

$$\log it P(X) = \ln \left[\frac{P(X)}{1 - P(X)} \right] = \alpha + \sum \beta_j X_j \quad (1.5)$$

de manera que:

$$\frac{P(X)}{1 - P(X)} = e^{(\alpha + \sum \beta_j X_j)} \quad (1.6)$$

La expresión 1.6 se le conoce como Odds o momio y es el resultado de dividir la probabilidad de ocurrencia de un evento entre la probabilidad de que este no ocurra.

1.2.2 Cociente de momios. (Odds Ratio)

El Odds Ratio o cociente de momios es la razón de dos momios denotados como $odds_1$ y $odds_0$ donde el subíndice indica dos individuos de dos grupos que son comparados y permite establecer una cuantificación del cambio en la probabilidad de ocurrencia de la variable de respuesta, en

nuestro caso la mortalidad de un delfín, a partir del cambio o presencia de uno o más factores en el grupo 1 con relación al grupo 0. Es la única medida de asociación directamente estimada del modelo logístico sin requerir supuestos especiales sobre los datos.

$$OR = \frac{odds_1}{odds_0} \quad (1.7)$$

Con base en la expresión 1.6 podemos escribir el momio para el grupo 1 como:

$$\frac{P(X_1)}{1 - P(X_1)} = e^{(\alpha + \sum \beta_j X_{1j})} \quad (1.8)$$

Y de la misma forma para el grupo 0:

$$\frac{P(X_0)}{1 - P(X_0)} = e^{(\alpha + \sum \beta_j X_{0j})} \quad (1.9)$$

Al considerar las expresiones 1.7, 1.8 y 1.9 el cociente de momios queda como:

$$OR = \frac{odds_1}{odds_0} = \frac{\frac{P(X_1)}{1 - P(X_1)}}{\frac{P(X_0)}{1 - P(X_0)}} = \frac{e^{(\alpha + \sum \beta_j X_{1j})}}{e^{(\alpha + \sum \beta_j X_{0j})}} = e^{\sum_{j=1}^k \beta_j (X_{1j} - X_{0j})}$$

Por lo que el cociente de momios queda expresado así:

$$OR = e^{\sum_{j=1}^k \beta_j (X_{1j} - X_{0j})} \quad (1.10)$$

En particular para el caso en que hay un cambio de una unidad en un solo factor, digamos el factor número j , para el grupo 1 y mientras el resto de los factores permanece igual en ambos grupos, la expresión 1.10 se simplifica de la siguiente manera:

$$e^{\sum_{j=1}^k \beta_j (X_{1j} - X_{0j})} = e^{\beta_1(0) + \dots + \beta_j(1) + \dots + \beta_k(0)} = e^{0 + \dots + \beta_j + \dots + 0} = e^{\beta_j}$$

Entonces para cuantificar la contribución al cociente de momios al incrementar una unidad de un factor en un grupo y mientras el resto de los factores permanece sin cambio en ambos grupos, se usa la siguiente igualdad:

$$OR_j = e^{\beta_j} \quad (1.11)$$

La ecuación 1.11 también funciona en el caso de un factor dicotómico, pero en este caso cuantifica la contribución al cociente de momios de un factor j cuando esta presente en un grupo y ausente en otro, mientras el resto de los factores permanece sin cambio en ambos grupos.

1.2.3 Interpretaciones del OR

Si en la expresión 1.11 se sustituye el valor de β_j por el valor del coeficiente asociado a un factor de riesgo de interés, el valor obtenido puede ser utilizado para determinar si dicho factor tiene influencia sobre la mortalidad de los delfines y es aplicable no solo cuando se trata de un factor con solo dos posibles valores como el caso de la presencia o ausencia de colapsamiento del canal de liberación sino también para una variable continúa como el tiempo de duración de la maniobra de liberación.

Si al sustituir en 1.11 el valor del OR obtenido es igual a uno se considera que el factor de interés no contribuye ni a favor ni en contra al riesgo de mortalidad de los delfines, porque el momio de muerte para el grupo donde hay un incremento de una unidad del factor de riesgo o presencia de éste es el mismo que para el grupo donde dicho incremento o factor esta ausente.

Si el valor del OR resulta mayor de uno entonces se trata de un factor que incrementa el riesgo de muerte de los delfines pues el momio de muerte para un delfín es mayor entre los expuestos al factor con relación a los no expuestos. Y un OR menor que uno indica que el factor localizado disminuye el riesgo de muerte de los delfines ya que el momio de muerte de un delfín es menor entre los expuestos al factor que entre los no expuestos.

.

1.2.4 Intervalo de confianza para el cociente de momios OR

Para poder obtener un valor del OR es necesario estimar los coeficientes β_j que se expresan como $\hat{\beta}_j$ y la estimación puntual del OR debe ir siempre acompañada de su intervalo de confianza, generalmente al 95% de confianza.

Un intervalo de confianza del 95% que incluye el valor 1 indica ausencia de asociación entre el factor de estudio y la variable de respuesta con una probabilidad de acierto del 95%. Pero si el intervalo excluye el valor 1 entonces, con una confianza del 95%, si existe asociación estadísticamente significativa entre el factor de estudio y la variable de respuesta. El intervalo de confianza se calcula como sigue:

$$e^{\left(\hat{\beta}_j \pm Z_{\alpha/2} * \sigma_{\hat{\beta}_j} \right)} \quad (1.12)$$

donde:

$\hat{\beta}_j$ es la estimación del coeficiente asociado al factor j obtenido por el método de máxima verosimilitud

$Z_{\alpha/2}$ es el valor de la distribución normal asociado a una confiabilidad que generalmente es del 95%

σ_{β_j} es el valor de la desviación estándar de la estimación del coeficiente asociado al factor j obtenido por el método de máxima verosimilitud.

Existen diversos métodos mediante los cuales es posible obtener un estimador $\hat{\beta}_j$ para cada coeficiente β_j , los métodos utilizados en este estudio se describe en la siguiente sección.

1.2.5 Método de Estimación de los coeficientes β_j en el modelo logístico

1.2.5.1 Método de Máxima Verosimilitud

El Método de Máxima Verosimilitud es una de las varias alternativas desarrolladas para estimar los parámetros en un modelo de regresión logística.

"A diferencia de otros el Método de Máxima Verosimilitud no requiere restricciones de ninguna clase en las características de las variables independientes .Es decir, usando la estimación de máxima verosimilitud, las variables independientes pueden ser nominales, ordinales o de intervalo. " (Kleinbaum 2002, p.105)

"La función de verosimilitud representa la probabilidad conjunta de observar los datos que han sido colectados. El término probabilidad conjunta se refiere a una probabilidad que combina las contribuciones de todas las observaciones en el estudio." (Kleinbaum 2002, p.110)

El método de máxima verosimilitud elige los coeficientes β_i en el modelo de regresión logística que maximizan la función de verosimilitud. Es decir mediante este método se eligen aquellos valores de β_i que maximizan la probabilidad de observar los datos registrados. Dichos coeficientes se obtienen mediante algoritmos computacionales como el método de Newton-Raphson.

1.2.5.1.1 Estimación condicional y no condicional

"Existen dos alternativas para realizar estimaciones a través del método de máxima verosimilitud, el método no condicional y el método condicional. " (Kleinbaum 2002, p.105). "El primero es utilizado si el número de parámetros es relativamente pequeño con relación al número de observaciones, mientras que el condicional es preferido si el número de parámetros es relativamente grande al número de observaciones" (Kleinbaum 2002, p.106); de manera que, el método correspondiente para este estudio es el método no condicional.

1.2.5.2 Método de ecuaciones de estimación generalizada (GEE)

La estimación de los coeficientes del modelo logístico para datos correlacionados es diferente en comparación al hacerlo para datos no correlacionados. En el presente estudio existe la posibilidad de que los datos presenten correlación entre si ya que algunas de las observaciones usadas dentro del modelo de trabajo pertenecen a un mismo lance de pesca, es decir, puede presentarse correlación entre observaciones dentro de un mismo lance. En este caso se utiliza el método de ecuaciones de estimación generalizada, abreviado como GEE, no solo para obtener una estimación de los coeficientes sino para obtener un correcto ajuste de las desviaciones estándar de los coeficientes estimados y en consecuencia un cálculo correcto de la precisión de estos. (Hosmer 2000, p.312,313)

El método GEE requiere hacer una suposición acerca de la naturaleza del tipo de correlación. La correlación predeterminada que utiliza la mayoría de los paquetes estadísticos y que se utiliza en este estudio se llama correlación intercambiable y supone que la correlación entre pares de respuestas es constante, en otras palabras, la correlación que existe para los delfines de un mismo lance es la misma, independientemente del lance al que pertenecen. En una amplia variedad de datos se obtiene un buen resultado utilizando esta suposición (Hosmer 2000, p.313). Al utilizar dicho método no se espera un cambio en el valor de los coeficientes estimados pero si una diferencia en la significancia de estos.

1.2.6 Prueba de Wald y Rao para la selección de variables

Existen varias pruebas para determinar si el valor de un coeficiente dentro del modelo de regresión logística es estadísticamente distinto de cero que equivale a que su OR respectivo sea estadísticamente distinto de uno, entre ellas están la prueba del cociente de verosimilitud, la prueba de Wald y la prueba de Rao. Dado que para muestras grandes el cociente de verosimilitudes y la prueba de Wald dan resultados similares, se decidió utilizar el estadístico de Wald y el estadístico de Rao en el proceso de eliminación y selección de variables.

1.2.6.1 Estadístico de Wald

Para cualquier variable independiente x_j seleccionada, si β_j es el coeficiente asociado a x_j en la ecuación de regresión logística, el estadístico de Wald permite constatar la hipótesis nula:

$$H_0 : \beta_j = 0$$

Que se interpreta diciendo que la información que se pierde al eliminar la variable asociada al coeficiente β_j no es significativa. De este modo en cada etapa del proceso de selección de variables, la variable candidata de ser eliminada será aquella que presente la menor confiabilidad asociada al estadístico de Wald. Y será eliminada si dicho valor es mayor que un valor crítico prefijado

"La prueba estadística de Wald es una generalización de una estadística t o Z . Es una función de la diferencia entre el estimador de máxima verosimilitud (EMV) y su valor bajo la hipótesis nula, normalizado sobre una estimación de la desviación estándar del EMV." (Harrell 2001, p.184)

Es decir:

$$W = [\hat{\beta}_j - \beta_0]^2 / [s(\hat{\beta}_j)]^2 \quad (1.13)$$

La expresión 1.13 se distribuye aproximadamente como una Ji cuadrada con un grado de libertad bajo H_0 . Y para el caso particular en que

$H_0 : \beta_j = 0$ el estadístico es:

$$Z = \frac{\hat{\beta}_j}{s(\hat{\beta}_j)} \quad \text{es aproximadamente Normal (0,1)}$$

"El estadístico para la prueba de Wald se calcula como la división de la estimación del coeficiente de interés entre su error estándar. Este cociente tiene aproximadamente una distribución normal (0,1) para muestras grandes. El cuadrado de esta estadística Z es aproximadamente una estadística Ji cuadrada con un grado de libertad. ." (Kleinbaum 2002, p.135).

1.2.6.2 Estadístico Rao (Puntuación eficiente de Rao)

Para cualquier variable independiente x_j no seleccionada, si β_j es el parámetro asociado a x_j en la ecuación de regresión logística, el estadístico de Rao permite constatar la hipótesis nula:

$$H_0 : \beta_j = 0$$

Que se interpreta diciendo que si la variable asociada al coeficiente β_j fuera seleccionada en el siguiente paso, la información que aportaría no sería significativa. En este caso en cada etapa de selección de variables, la candidata a ser seleccionada será aquella que presente la mayor confiabilidad asociada al estadístico Rao. Y será seleccionada si el valor de significancia es menor a un valor predeterminado

El estadístico es:

$$S = U(\beta_0)^2 / I(\beta_0) \quad (1.14)$$

formado por:

$$U(\beta_j) = \partial \text{Log } L / \partial \beta_j$$

$$I(\beta_j) = -\partial^2 \text{Log } L / \partial \beta_j^2$$

donde L es la función de verosimilitud. La expresión 1.14 se distribuye aproximadamente como una Ji cuadrada con un grado de libertad bajo H_0 .

La expresión 1.13 y 1.14 pueden ser generalizadas para el caso en el que se prueba la hipótesis sobre más de un parámetro de manera simultánea, pueden consultarse dichas formulas en Hosmer.(2000) y Collett (2003)

1.2.6.3 Pasos seguidos en el proceso de selección de variables utilizando el estadístico de Wald y Rao

En el primer paso se introduce la variable que presenta el valor máximo asociado al estadístico Rao y que al mismo tiempo debe de ser menor a .01, en otro caso el proceso termina y no es posible realizar la construcción del modelo que describe la mortalidad de los delfines.

En el siguiente paso se introduce la variable que presenta el valor máximo asociado al estadístico Rao y que al mismo tiempo debe de ser menor a .01, en otro caso el proceso termina y el modelo se construye con la variable seleccionada en el paso anterior.

En el tercer paso se introduce la variable que presenta el valor máximo asociado al estadístico Rao y que al mismo tiempo debe de ser menor a .01. Si, al introducir una variable, la significancia máxima asociada al estadístico Wald para las variables previamente seleccionadas es mayor a

0.05, antes de proceder a la selección de una nueva variable, se eliminara la variable correspondiente.

Cuando ninguna variable tenga un valor asociado al estadístico de Wald mayor a .05 se vuelve al tercer paso. Dicho punto del proceso se repite hasta que ninguna variable no seleccionada tenga una significancia asociada al estadístico de Rao menor a 0.01 y ninguna de las seleccionadas tenga un valor asociado al estadístico de Wald mayor a 0.05

1.2.7 Medidas de Ajuste del Modelo

Un modelo de regresión logística puede utilizarse con dos propósitos:

Estimativo: El interés se centra en estimar la relación de una o más variables independientes con la variable dependiente. En este caso es conveniente usar los intervalos de confianza para poder medir el grado de precisión de los coeficientes e^{β_j} asociados a los factores de riesgo.

Predictivo: el objetivo del modelo es clasificar correctamente a un nuevo individuo que no haya sido utilizado para la elaboración del modelo.

En este caso es necesario determinar el ajuste entre las observaciones y las predicciones del modelo, que se mide evaluando la calibración y la discriminación del modelo. Además para los modelos destinados a la predicción se busca determinar el grado en que el modelo puede generalizarse a diferentes poblaciones de estudio.

La calibración compara el número predicho de eventos con el número observado en grupos de individuos, mientras que la discriminación evalúa el grado en que el modelo distingue entre individuos en los que ocurre el evento y los que no.

La capacidad en que el modelo puede generalizarse a diferentes poblaciones se refiere a que el modelo puede realizar predicciones válidas en individuos diferentes de aquellos en los que se ha generado y tiene también dos componentes: reproducibilidad (capacidad del modelo de realizar predicciones válidas en individuos no incluidos en la muestra con la que se ha generado, pero procedentes de la misma población) y transportabilidad (capacidad de realizar predicciones válidas en individuos procedentes de una población distinta pero relacionada).

La reproducibilidad puede ser evaluada dividiendo aleatoriamente los datos en dos grupos: en uno de ellos (grupo de trabajo) se ajusta el modelo y en el otro (grupo de validación) se observa su comportamiento, en cuanto a la capacidad de clasificar adecuadamente a los individuos. Pero la transportabilidad requiere una muestra de otra población.

Las pruebas estadísticas que pueden ser usadas para evaluar la calibración es el cociente de verosimilitudes y la prueba de Hosmer-Lemeshow, mientras que como medida para determinar el grado de discriminación se utiliza el área bajo la curva ROC.

1.2.7.1 Estadístico del cociente de verosimilitudes.

Una medida del grado de bondad o falta de ajuste de un modelo es comparar la verosimilitud de un modelo que ajusta perfectamente a los datos con el modelo elegido. El modelo que ajusta perfectamente a los datos no aporta ningún tipo de información resumida sobre estos por tener un número de coeficientes igual al número de observaciones y por tanto no se usa en la práctica, su máxima verosimilitud se denota como \hat{L}_f , mientras que en el modelo calculado, la máxima verosimilitud se denota por \hat{L}_c .

Para comparar \hat{L}_c con \hat{L}_f , es conveniente utilizar menos dos veces el logaritmo del cociente de ambos máximos de verosimilitudes, denotados por D:

$$D = -2\log(\hat{L}_c/\hat{L}_f) = -2\{\log \hat{L}_c - \log \hat{L}_f\} \quad (1.15)$$

En el caso en que se trabaja con observaciones de respuesta y_i binarias el estadístico D solo depende de las probabilidades ajustadas por el modelo calculado, de modo que dicho estadístico no puede decir nada acerca de la correspondencia entre los valores observados y sus correspondientes probabilidades ajustadas. Consecuentemente, el estadístico D para el caso en que se ajusta un modelo de respuesta binaria

no puede ser usado como una medida representativa de la bondad de ajuste de un modelo. (Collett 2003, p.69).

1.2.7.2 Prueba de Hosmer-Lemeshow

Es una prueba para evaluar la bondad del ajuste de un modelo de regresión logística. Se divide la población en g grupos iguales de acuerdo con los percentiles de las probabilidades predichas por el modelo. (Hosmer 2000, p.148)

$$G_{HL}^2 = \sum_{j=1}^g \frac{(O_j - n_j \pi_j)^2}{n_j \pi_j (1 - \pi_j)}$$

donde:

n_j es el número de observaciones en el grupo j.

O_j número observado de casos en el grupo j.

π_j es la probabilidad estimada promedio para el grupo j.

La distribución del estadístico es aproximadamente Ji cuadrada con g - 2 grados de libertad.

1.2.7.3 Curva ROC

El área bajo la curva ROC es un buen indicador de la discriminación de un modelo logístico. Representa, para todos los pares posibles de individuos formados por un individuo en el que ocurrió el evento y otro en el que no, la proporción de los que el modelo predice una mayor probabilidad para el que tuvo el evento. Gráficamente esta proporción esta representada con el área bajo la curva ROC y mientras mayor es el área comprendida por debajo de la curva, mejor es la discriminación del modelo logístico. Para obtener la curva ROC, se calcula la sensibilidad y especificidad para cada uno de los diferentes valores observados en los datos (gráfica 1.1), con la sensibilidad en el eje de las Y, y [1-Especificidad] en el eje de las X. El cálculo de la sensibilidad y la especificidad está descrito en la siguiente sección.

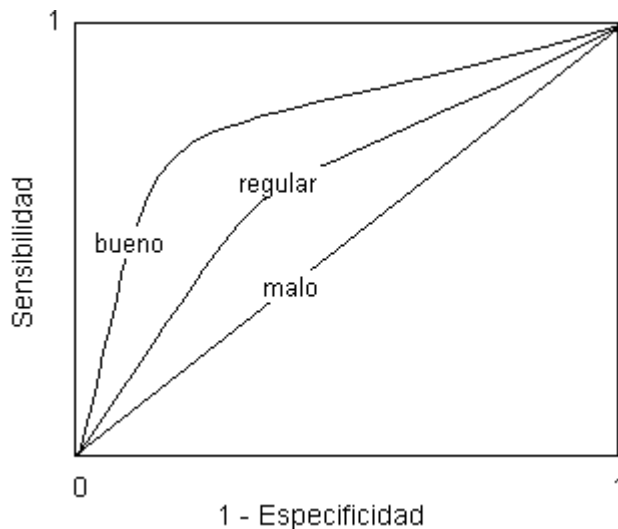


Gráfico 1.1 Curvas ROC y el grado de discriminación que representan

El rango de valores del área bajo la curva va desde 0,5 (el modelo no discrimina) hasta 1 (discriminación perfecta).

Generalmente se utiliza el estimador por intervalo para el área bajo la curva ROC de modo que rechazar la hipótesis de que el área es igual a 0.5 (con una significancia menor a 0.05 o un intervalo de confianza que no contiene al 0,5), proporciona evidencia de que el modelo logístico tiene capacidad de discriminación. A partir de un área de 0.7 la discriminación del modelo se considera aceptable.

Una interpretación equivalente del valor del área bajo la curva ROC es la siguiente: un área de 0,8 significa que un individuo seleccionado aleatoriamente del grupo de individuos quienes presentaron el evento de interés tiene un valor asignado por el modelo logístico mayor que uno seleccionado aleatoriamente del grupo de individuos que no presentaron el evento de interés en el 80 % de las veces.

1.2.7.4 Sensibilidad y Especificidad

A partir de un conjunto de datos, es posible definir los conceptos de sensibilidad y especificidad necesarios para poder construir la gráfica ROC

		Observado		
		Presencia	Ausencia	Total
Predicción del modelo	Presencia	A	B	a+b
	Ausencia	C	D	c+d
Totales		a+c	b+d	

Tabla 1.2 Comportamiento observado y predicho de los individuos con respecto a la ausencia o presencia de característica de interés.

Sensibilidad. Es la proporción de individuos que presentaron el evento de interés y que fueron identificados como tales por el modelo, llamados también, verdaderos positivos.

$$\text{Sensibilidad} = S = \frac{a}{a + c}$$

En contraste, existe la presencia de falsos positivos o 1-especificidad, es decir, individuos señalados como poseedores de la característica de interés cuando en realidad no la poseen.

Especificidad. Es la proporción de individuos que no presentaron la característica de interés y que fueron identificados como tales por el modelo logístico, también llamados verdaderos negativos.

$$\text{Especificidad} = E = \frac{d}{b + d}$$

La contraparte de los verdaderos negativos consiste de aquellos individuos que fueron identificados como no poseedores de la característica de interés cuando en realidad si la poseen.

2. RESULTADOS Y DISCUSION

2.1 Regresión logística

Utilizando el método de máxima verosimilitud y el método GEE se estimaron los valores del Odds Ratio correspondientes a cada uno de los factores de riesgo y seleccionado aquellos que inciden en la mortalidad de los defines y el grado en que lo hacen. Los resultados están presentados en la tabla 2.1

No existe una variación significativa en el valor de los coeficientes ni su nivel de significancia asociado obtenido a través de ambos métodos,. El mayor cambio fue el de la significancia del coeficiente del turno en que se realiza el lance, con un cambio de 0.009 sin considerar correlación dentro de cada lance a 0.055 usando el método GEE, dicho valor fue considerado como significativo dada su cercanía con el valor 0.05.

Variable	β_j	Sig.	e^{β_j}	Intervalo de confianza e^{β_j} (95%)	
				Límite Inferior	Límite Superior
Turno	-0.9783	0.055	0.3759	0.1384	1.0211
Colapsamiento del canal de liberación	2.2947	0.000	9.9217	3.3951	28.9948
Distancia de la Observación	0.2478	0.004	1.2812	1.0834	1.5151

Duración de la maniobra de liberación	0.1030	0.007	1.1085	1.0292	1.1940
---------------------------------------	--------	-------	--------	--------	--------

Tabla 2.1 Modelo seleccionado mediante el método Forward

El colapsamiento del canal de liberación implica que el medio principal mediante la cual los delfines son liberados fracasa y por tanto se espera un incremento considerable en el riesgo de muerte de un delfín. De acuerdo a modelo descrito en tabla 2.1 el momio de muerte para un delfín es aproximadamente 9 veces superior al momio de no morir si se presenta el colapsamiento del canal de liberación que si este no se presenta. El colapsamiento del canal de liberación es entonces el factor de mayor riesgo al que están expuestos los delfines durante la faena de pesca. Por lo que si se desea reducir la mortalidad de los delfines asociada a la pesca del atún, es necesario determinar, en un futuro estudio, las fallas en la formación del canal de liberación de delfines que producen su colapsamiento.

En relación al tiempo en que toma en realizarse la maniobra de liberación puede suponerse que el incremento en el tiempo de dicha maniobra incrementa el estrés del encierro al que son sometidos los delfines provocando una mayor mortalidad de éstos y como consecuencia el incremento de un minuto en la realización de la maniobra de liberación, incrementa el momio de que un delfín muera en más del 10%.

A diferencia de colapsamiento del canal de liberación o la duración de la maniobra para realizar este, la distancia en la que se realiza el avistamiento del grupo de delfines asociado con el cardumen de atunes o el turno en que se realiza el lance no parecen tener un efecto directo sobre la mortalidad de los delfines. Un incremento de una milla náutica en la distancia a la que es observado el cardumen de atunes, incrementa el riesgo de que un delfín muera en un 24 % aproximadamente y el riesgo de morir para un delfín capturado en el turno vespertino es aproximadamente un 62% menor en comparación a aquellos delfines que fueron capturados en el turno matutino. Es probable que ambos factores esté relacionados a otros que si afectan de forma directa el riesgo de morir de un delfín, los cuales pueden ser determinados en un nuevo estudio, lo que a su vez, permitiría determinar las medidas necesarias para minimizar o eliminar la mortalidad provocada por tales factores ya que aunque la comisión interamericana del atún tropical asegura que la mortalidad anual es cercana a cero, la mortalidad total para los años 2004, 2005 y 2006 registran 1469, 1151 y 886 delfines muertos respectivamente que, si bien en promedio es una mortalidad por lance de menos de un delfín, se desconoce el total de delfines capturados al año sobre los cuales se registraron las mortalidades reportadas.

Existe la posibilidad de que la reducción en la mortalidad reportada sea provocada por una disminución importante en el número de delfines que componen las manadas capturadas más que por las medidas implementadas en protección a los delfines.

2.2 Ajuste del Modelo

2.2.1 Calibración

Mediante la prueba de Hosmer-Lemshow (tabla 2.2) se comparan el número de delfines muertos predicho por el modelo de regresión logística con el número observado para cada uno de los 8 grupos formados. Y se efectúa la prueba Ji Cuadrada (Tabla 2.3) donde se observa que no existen diferencias significativas entre el número de delfines muertos observados con el número de delfines muertos predichos por el modelo logístico.

Grupo	Delfines muertos		Delfines vivos		Total
	Observados	Esperados	Observados	Esperados	
1	26	26.663	2	1	28
2	21	20.625	3	3	24
3	17	17.425	9	8	26
4	14	13.362	8	8.638	22
5	13	14.304	18	16.696	31
6	10	8.403	14	15.597	24
7	10	7.939	14	16.061	24
8	5	7.28	36	33.72	41

Tabla 2.2 Comparación entre el número predicho por el modelo con el número observado de delfines muertos para cada grupo.

Valor Cuadrada	Ji	Grados de libertad	Significancia
2.8578		6	.8265

Tabla 2.3 Resultados de la prueba de Ji Cuadrada para comparar diferencias entre el número de delfines muertos observados con el número de delfines muertos predichos por el modelo logístico en cada grupo.

2.2.2 Discriminación

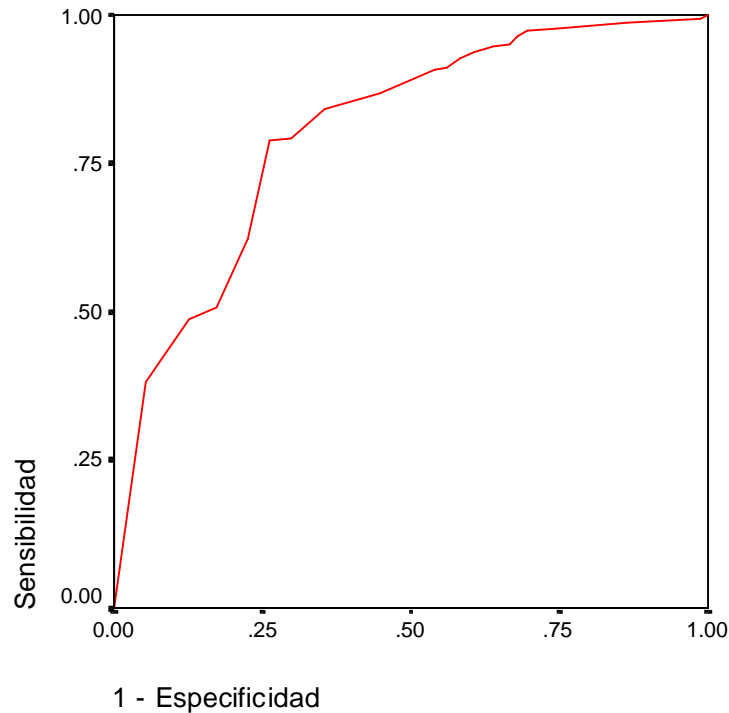
El área bajo la curva ROC (tabla 2.4) se interpreta diciendo que un delfín seleccionado al azar del grupo de delfines muertos tiene una probabilidad de muerte asignada por el modelo logístico mayor que uno seleccionado aleatoriamente del grupo de delfines vivos en el 80.3 % de las veces, es decir, la probabilidad con la que el modelo distingue entre delfines muertos y los delfines vivos es adecuada.

Area	Error Estándar	Significancia	Intervalo de Confianza	
			Límite Inferior	Límite Superior
0.803	0.006	.000	0.790	0.815

Tabla 2.4 Resultados del área bajo la curva ROC

Aproximadamente para una sensibilidad del 80% existe una presencia de falsos positivos de menos del 30%. En todos los casos donde la sensibilidad es de 50% y menos la presencia de falsos positivos es del 25% o menos. Para una sensibilidad mayor al 80% la presencia de falsos

positivos es menor al 75% aproximadamente.(Gráfica 2.1). Es decir, aunque el modelo pierde especificidad a medida que aumenta su sensibilidad esta pérdida no alcanza un grado que haga dudar de la confiabilidad del modelo para diferenciar mediante el modelo logístico las condiciones de mortalidad o sobrevivencia para los delfines.



Gráfica 2.1 Curva ROC para las predicciones realizadas a través del modelo logístico

2.2.3 Reproducibilidad

Como se había descrito en la sección 1.1.3 para el análisis de la mortalidad de delfines el grupo de trabajo consistió en una muestra de 220 delfines, compuesta de 116 delfines que sobrevivieron a las maniobras de pesca y 104 delfines que no sobrevivieron, tomada del total de 54300 observaciones del viaje de pesca, de manera que el grupo de validación consta de 54080 observaciones.

En el grupo formado por las nuevas observaciones, el modelo clasificó correctamente 74% de estas, en particular clasificó correctamente 73.95% delfines vivos como tales y 78.65% delfines muertos como tales.(Tabla 2.5) Entonces los resultados obtenidos a partir de las 220 observaciones son aplicables para las 54080 observaciones restantes.

Observados	Predichos		Porcentaje Correcto
	Delfines Vivos	Delfines Muertos	
Delfines Vivos	39289	13840	73.95%
Delfines Muertos	203	748	78.65%
Porcentaje total de observaciones bien clasificadas			74.03%

Tabla 2.5 Datos correctamente clasificados para el grupo de validación con su porcentaje correspondiente

2.2.4 Transportabilidad

Por no contarse con datos de otros viajes de pesca, es decir con datos correspondientes a otras poblaciones, no es posible medir la transportabilidad del modelo.

3. CONCLUSIONES

La regresión logística resultó una técnica adecuada para modelar el riesgo de muerte de los delfines asociado a la actividad de pesca del atún obteniéndose, entre otros resultados, que el colapsamiento del canal de liberación es el factor de riesgo más importante para incrementar la mortalidad. Sin embargo, los datos obtenidos corresponden solamente a un viaje de pesca, por lo que el trabajo deja abierta la posibilidad a futuras investigaciones destinadas a resolver las siguientes cuestiones:

- Mostrar si los factores de riesgo detectados en el viaje de pesca analizado lo son también en otros viajes de pesca.
- Determinar con exactitud las causas que provocan que el colapsamiento del canal de liberación y la forma en que el tiempo en que toma su realización incide en la mortalidad de los delfines.
- Registrar nuevos posibles factores de riesgo a fin de encontrar aquellos que inciden directamente en la mortalidad de los delfines y que están asociados a otros que no la afectan de manera evidente como el caso turno en que se realiza el lance de pesca y la distancia en la que se realiza el avistamiento del grupo de delfines asociado con el cardumen de atunes.
- Determinar si la temporada en que se realiza la captura del atún o aspectos relacionados a la embarcación que la realiza pueden ser factores de riesgo en la mortalidad de los delfines.

- Determinar si existe o no una reducción en las poblaciones de delfines, en particular las que están asociadas con los cardúmenes de atún y al mismo tiempo estudiar la evolución de estas en el tiempo.

Los futuros trabajos de investigación deben estar basados en un muestreo de datos que incluya observaciones no solo sobre varios lances de pesca en cada embarcación sino también sobre varias embarcaciones a lo largo del tiempo.

Y en un sentido más amplio, es necesario desarrollar e implementar nuevas formas de obtener recursos para alimentar a la población humana. Dichas formas tienen que considerar a la naturaleza como una prioridad, no solo para garantizar una fuente de alimento segura para el futuro, sino para permitir la recuperación y preservación de las diversas especies y ecosistemas que han sido afectados por la inconsciente actividad humana.

4. BIBLIOGRAFIA

- Canavos George C. 1991. *Probabilidad y Estadística*. Mc Graw Hill. Mexico. Primera Edición. 651 p.
- Collett, David. 2003. *Modeling binary data*. Chapman&Hall/CRC. . Florida, E.U. Segunda Edición. 387 p
- Cook Thomas. 2007. *Introduction to Statistical Methods for Clinical Trials*. Chapman & Hall. E.U. Primera Edición. 464 p
- Ferrán A. Magdalena. 1996. *SPSS para Windows*. McGRAW-HILL. México. Primera edición. 580 p.
- Grimm L. 1993. *Statistical Applications for the Behavioral Sciences*. John Wiley & Sons Inc. New York, E.U. Primera Edición. 608 p
- Harrell, Frank E Jr. 2001. *Regression modelling strategies*. Springer-Verlag. New York, E.U. Primera Edición. 568 p
- Hosmer, W. David. 2000. *Applied logistic regression*. J. Wiley. New York. E.U. Segunda Edición. 307 p.
- Kleinbaum G. David. 2002. *Logistic Regression: A Self-Learning Text*. Springer-Verlag. New York, E.U. Segunda Edición. 513 p