

PROGRAMA INTERNACIONAL PARA LA CONSERVACIÓN DE LOS DELFINES
PANEL INTERNACIONAL DE REVISIÓN

47ª REUNIÓN

LA JOLLA, CALIFORNIA (EE.UU.)
5 DE JUNIO DE 2009

DOCUMENTO IRP-47-13

ANÁLISIS DE LOS REQUISITOS DE DATOS PARA LA COMPARACIÓN ANUAL DE LOS PROGRAMAS DE OBSERVADORES, Y PROPUESTA PARA LA ESTANDARIZACIÓN DE LOS DATOS DE LOS OBSERVADORES

1. ANTECEDENTES

La responsabilidad del muestreo por los observadores en el mar de las actividades de pesca de los buques de más de 363 t de capacidad de acarreo en el Océano Pacífico oriental (OPO) corresponde actualmente a ocho programas de observadores a bordo, el programa internacional de la CIAT y los programas nacionales de Colombia, Ecuador, México, Nicaragua, Panamá, la Unión Europea, y Venezuela.

Estos ocho programas son responsables de la capacitación de los observadores, el procesamiento de los datos, y el control de calidad de los datos. Como parte de los esfuerzos por mantener una comparabilidad de los datos entre los programas de observadores, desde 2000 se ha realizado, y presentado al Panel, una comparación anual de los resúmenes de los datos de cada uno de los programas nacionales con resúmenes de los datos del programa de la CIAT correspondientes a viajes de buques de ese mismo país. Las comparaciones incluyeron información sumaria tanto estadística como no estadística. Se presentaron resúmenes no estadísticos de la distribución espacial de los lances por tipo, la captura media por día, el porcentaje de lances por tipo, y el porcentaje de lances sobre delfines con mortalidad cero. (Los dos primeros no han sido incluidos desde 2004.) Se presentaron pruebas estadísticas de diferencias promedio del porcentaje de viajes sin lances sobre delfines, el número promedio de días por viaje, la mortalidad promedio por lance, el número promedio de infracciones de interferencia al observador por viaje, y el número promedio de infracciones de procedimiento por lance. Aunque no todos los programas de observadores han tenido actividad desde 2000, se descubrieron pocas diferencias estadísticamente significativas entre los datos del programa nacional y los datos de la CIAT correspondientes al mismo país. Se descubrieron diferencias estadísticamente significativas en una comparación de programas en 2000 en el caso del porcentaje de viajes sin lances sobre delfines y la mortalidad promedio por lance, y en una comparación de programas en 2004 en el caso del número promedio de días por viaje. Además, en una comparación de programas, se descubrieron diferencias estadísticamente significativas en las tasas de infracción en cada año. No obstante, las diferencias influyentes en tasas de infracción ocurrieron al principio del período, y por lo tanto afectaron la comparación subsiguiente, porque todos los datos disponibles desde 2000 fueron usados anualmente. Los resultados de la mayoría de las comparaciones estadísticas no fueron significativamente diferentes (en pruebas bilaterales con un nivel α de 5%). Por ende, medidos con estas comparaciones, los programas han sido generalmente comparables.

Históricamente, las pruebas estadísticas de las comparaciones entre programas han estado basadas en todos los datos disponibles desde 2000. Esto significa que en 2008, las comparaciones entre programas activos desde 2000 se basaron en ocho años de datos (es decir, datos de 2000-2007). No todos los programas de observadores han tenido actividad desde 2000, y por lo tanto, en el caso de algunos programas nacionales, es posible que se haya usado menos años de datos en el análisis estadístico. El número de años de datos que se debería incluir en las comparaciones no ha sido tratado explícitamente. La inclusión

de demasiado pocos años de datos puede llevar a resultados sin sentido, debido a variabilidad interanual inherente en los datos. La inclusión de todos los años de datos disponibles puede conducir a resultados que posiblemente reflejen indebidamente la primera parte del período sin tener relación significativa con la situación actual. En la 46ª reunión del PIR, se pidió a la Secretaría realizar un análisis del número de años que se podría usar razonablemente en las pruebas estadísticas de la comparación de los programas de observadores.

2. SIMULACIÓN

A petición de la 46ª reunión del Panel, se realizó una simulación para guiar la selección del número de años por usar en la comparación anual. Las simulaciones se basaron en la mortalidad promedio por lance porque, de los elementos incluidos de la comparación anual, la mortalidad promedio por lance es el que está más directamente relacionado con la comparabilidad de la mortalidad de delfines. Se simularon diferencias anuales medias entre los programas en mortalidad promedio por lance de 0,005, 0,01, 0,02, 0,05 y 0,10 animales por lance; éstas corresponden a diferencias medias en mortalidad entre programas (dentro de un país) de 5, 10, 20, 50 y 100 animales por mil lances. La simulación implicó la generación de 1.000 conjuntos de datos simulados. Cada conjunto de datos fue generado mediante una asignación aleatoria de los viajes reales de los observadores de un país a dos programas hipotéticos, añadiendo luego una cantidad constante a la mortalidad de cada lance sobre delfines de los viajes de uno de los dos programas. Para cada uno de los 1.000 conjuntos de datos simulados, se realizó una prueba de aleatorización de la diferencia entre programas en la mortalidad promedio por lance, usando 999 aleatorizaciones. Este es el mismo tipo de prueba usado en la comparación anual. (Ver el Anexo para los detalles de la prueba estadística usada en la comparación anual.) Se registró la proporción de las 1.000 pruebas (una prueba por conjunto de datos) que hubiera rechazado la hipótesis nula de ninguna diferencia en la mortalidad promedio por lance entre programas. Esta simulación fue repetida para diferentes números de años de datos (de dos a cinco años), y con datos de dos países diferentes. Las simulaciones basadas en los datos de los dos países son ilustrativos porque la variabilidad en los datos de tasa de mortalidad de un país (País B) es mayor que aquélla del otro país (País A). Los años de datos usados como base para la simulación fueron 2003-2007.

En la tabla siguiente se presentan los resultados de estas simulaciones. En la columna de ‘Ninguna diferencia’ se señala la proporción de las pruebas que rechazó la hipótesis nula cuando en realidad no hubo ninguna diferencia entre los programas. Se esperaría encontrar diferencias significativas en un 5% de las ocasiones (a partir de una prueba bilateral con un nivel α de 5%; valores entre 3,6% y 6,4% son aceptables). Se pueden usar los resultados de la simulación para guiar la selección del número de años de datos por usar en la comparación anual, bajo el supuesto que la diferencia real entre programas es constante a través de los años. Estas simulaciones podrían ser ampliadas para incluir, por ejemplo, diferencias no

| Años de datos usados en la prueba | Proporción de pruebas que rechazan la hipótesis de ninguna diferencia en mortalidad por lance (prueba bilateral con un nivel α de 5%) | | | | | |
|-----------------------------------|--|-------|-------|-------|-------|-------|
| | Diferencia simulada en la mortalidad media por lance (animales por lance) | | | | | |
| | Ninguna diferencia | 0.005 | 0.01 | 0.02 | 0.05 | 0.10 |
| País A | | | | | | |
| 2 | 0.058 | 0.079 | 0.131 | 0.397 | 0.985 | 1.0 |
| 3 | 0.045 | 0.078 | 0.214 | 0.571 | 1.0 | 1.0 |
| 4 | 0.048 | 0.097 | 0.243 | 0.709 | 1.0 | 1.0 |
| 5 | 0.061 | 0.102 | 0.309 | 0.837 | 1.0 | 1.0 |
| País B | | | | | | |
| 2 | 0.060 | 0.049 | 0.057 | 0.111 | 0.382 | 0.919 |
| 3 | 0.052 | 0.065 | 0.067 | 0.164 | 0.625 | 0.999 |
| 4 | 0.055 | 0.060 | 0.092 | 0.196 | 0.793 | 1.0 |
| 5 | 0.045 | 0.060 | 0.092 | 0.247 | 0.889 | 1.0 |

constantes en la mortalidad media por lance entre años. Esto requerirá dirección del PIR con respecto a los patrones específicos de diferencias que son de interés.

No obstante estos resultados, la Secretaría sugiere que se cambie del ejercicio de la comparación anual realizada desde 2000 para enfocar el esfuerzo en mejorar la calidad de los datos obtenidos por todos los programas y en la estandarización de los datos obtenidos. Ya que se usan los datos de todos los programas colectivamente en los análisis científicos, es importante que no se introduzcan sesgos en la base de datos de observadores común como consecuencia de diferencias entre los programas en los procedimientos de toma y edición de datos. Por los motivos siguientes, el ejercicio actual de comparación anual no brinda un marco para asegurar ni la estandarización de los datos entre programas ni que los datos obtenidos sean necesariamente de la mejor calidad.

1. Dos programas pueden ser estadísticamente similares pero tener ambos datos de mala calidad. Se necesita contar con procedimientos detallados para la toma y edición de los datos para poder sentar las bases para mantener el nivel más alto posible de calidad de datos.
2. Lograr comparabilidad de datos entre programas es ante todo cuestión de estandarizar la capacitación de los observadores y los procedimientos de edición de datos. Por lo tanto, se deben sentar las bases para la comparabilidad de los datos mediante el uso de materiales comunes para la capacitación de los observadores, y directrices y procedimientos comunes para la revisión de los datos con los observadores y la edición de los mismos.
3. Una meta importante de cualquier procedimiento para la estandarización de datos entre los programas de observadores es que se identifiquen y marquen, de forma apropiada, los datos de calidad cuestionable, para poder excluirlos de análisis futuros en caso necesario. Las comparaciones anuales actuales no contemplan esta meta.

Para asegurar que los datos obtenidos por todos los programas sean comparables y de la mejor calidad posible, se recomienda que las directrices señaladas a continuación, que amplía el Anexo II del APICD, sean elaboradas y adoptadas por todos los programas. Algunas de las mismas fueron discutidas y acordadas en la segunda reunión de los programas de observadores nacionales y de la CIAT en octubre de 2007, y algunas, pero no todas, están siendo aplicadas por los diferentes programas de observadores.

3. DIRECTRICES PARA LA ESTANDARIZACIÓN Y CONTROL DE CALIDAD DE DATOS POR LOS PROGRAMAS DE OBSERVADORES

A continuación se presentan directrices sobre los siguientes aspectos de la estandarización y control de calidad de los datos.

1. Selección de observadores
2. Capacitación de observadores
3. Toma de datos
4. Edición de datos
5. Intercambio de datos
6. Colaboración adicional entre los programas

3.1. Selección de observadores

La selección de candidatos debería ser una actividad conjunta de todos los programas.

Los candidatos no deberían ser descalificados por su sexo, edad, o raza.

Todos los candidatos deben ser entrevistados.

Los candidatos deben satisfacer los siguientes requisitos mínimos:

- a. ser licenciado universitario con título en biología o especialidades relacionadas (oceanografía, ecología, biología pesquera, por ejemplo); se debe evitar reclutar licenciados en disciplinas técnicas

- b. Estar en buenas condiciones físicas y ser certificados por un médico para trabajo en el mar.
- c. Haber completado un curso de supervivencia en el mar, y presentar pruebas de haber aprobado una prueba de supervivencia en el mar realizada por una entidad gubernamental. Los observadores pagarán el costo del curso y de la prueba, pero serán reembolsados por su programa de observadores al ser contratados como observador

En todo caso posible, los candidatos deben tomar una prueba psicométrica para evaluar su aptitud psicológica con respecto a las exigencias del trabajo de observador pesquero.

3.2. Capacitación de observadores

Toda capacitación de observadores debe ser realizada en sesiones conjuntas de capacitación.

Los materiales usados para la capacitación de los observadores deben ser idénticos en todas las sesiones de capacitación y en todos los programas. Nuevos materiales deben ser circulados a todos los programas en borrador para sus comentarios.

Al terminar el curso de capacitación, todos los candidatos deben aprobar un examen estandarizado para ser elegible para trabajar como observador. El examen estandarizado será desarrollado conjuntamente por los programas, y actualizado anualmente con material nuevo.

La capacitación de los observadores debe incluir una sesión dedicada al protocolo en el mar, y el tema debe ser cubierto en el manual de los observadores.

3.3. Toma de datos

Todos los programas deben usar los mismos formularios estandarizados para la toma de datos, y las mismas instrucciones y manuales.

Todos los programas deben usar las mismas guías de campo y claves de identificación de especies.

Todos los formularios de datos deben ser desarrollados conjuntamente entre los programas. Nuevos materiales deben ser circulados a todos los programas en borrador para sus comentarios.

3.4. Edición de datos

Los editores de datos deben, como mínimo, haber aprobado un curso de capacitación de observadores y, preferentemente, haber realizado al menos un viaje como observador.

Los procedimientos para las entrevistas de fin de viaje con los observadores deben ser estandarizados entre todos los programas.

Las entrevistas de fin de viaje deben seguir las directrices en un manual estandarizado (por desarrollar).

Las entrevistas de fin de viaje deben ser realizadas oportunamente, conforme a un calendario común compartido por todos los programas.

Todos los programas deben usar los mismos programas de computadora para la edición de los datos.

Todos los programas deben editar sus datos, tanto por ‘errores’ como por ‘avisos,’ de la misma manera. La edición de datos por todos los programas debe seguir las directrices en un manual estandarizado (por desarrollar).

Todos los programas deben seguir procedimientos estandarizados para identificar y marcar datos cuestionables, incluyendo identificaciones de especies y cantidades de captura y captura incidental. Estos procedimientos deben ser descritos en un manual estandarizado (por desarrollar).

El personal del programa debe repasar con cada observador los errores en sus datos antes de su próximo viaje, y entregarle una lista de los más recientes. El observador debe llevar esta lista consigo en su próximo viaje y revisarla.

El desempeño de los observadores debe ser evaluado usando criterios estandarizados compartidos por todos los programas.

El programa nacional de cada país y el programa de la CIAT en ese país deben intercambiar editores de datos anualmente para verificar que los procedimientos de entrevista y de edición usados sean idénticos.

3.5. Intercambio de datos

Los programas de la CIAT y nacionales deben:

- a. intercambiar semanalmente informes de las salidas y arribos de observadores en los buques del país respectivo.
- b. intercambiar datos preliminares de forma oportuna, conforme a un calendario acordado por todos los programas.
- c. intercambiar todos los datos editados finales de forma oportuna, conforme a un calendario acordado por todos los programas.

3.6. Colaboración adicional entre los programas

Se debe celebrar una reunión periódica (cada año o dos) del personal de los varios programas de observadores, incluyendo los individuos responsables de la capacitación de los observadores, de las entrevistas de fin de viaje y la edición de datos, y de las bases de datos. Las reuniones servirían de foro para la discusión de las actividades de los programas, actualizaciones/revisiones de los materiales de capacitación, procedimientos de edición, y programas de computadora, así como para la formación del personal.

ANEXO.

Descripción de la prueba de aleatorización usada en la comparación de los programas de observadores

Para evaluar estadísticamente diferencias entre programas, se usó una prueba de aleatorización para obtener una estimación de la probabilidad de que una diferencia anual media igual a aquella observada, o mayor, es ocasionada por la asignación aleatoria de viajes a programas. Se realizó la prueba asignando al azar viajes del conjunto de datos del PIR agrupados de un país en particular, por año, a dos programas, luego calculando la diferencia anual media simulada en la cantidad de interés (por ejemplo, mortalidad media por lance) entre los programas para la muestra aleatoria de viajes. Se simularon en total 4,999 muestras aleatorias de viajes. Se calculó el valor p para esta prueba como la proporción de las diferencias anuales medias simuladas con un valor absoluto igual a aquél observado, o mayor. Estos cálculos representan una aproximación de una prueba bilateral (de dos extremos) de la hipótesis nula de ninguna diferencia entre programas.