

INSTITUTO DEL MAR DEL PERU

INFORME

Nº 79

FAO Proyecto PNUD/FAO-PER/76/022

PARTE I: "INVESTIGACION DE LA MERLUZA EN IMARPE"
Dr. David Armstrong

PARTE II: "RECOMENDACIONES PARA COMPUTACION DE DATOS DE
CAPTURAS DE LA PESQUERIA DE CONSUMO HUMANO"
Dr. William Hall

PARTE III: "PESQUERIA ARTESANAL/RECURSOS COSTEROS"
Dr. Martin Walsh

TRADUCCION Y EDICION DEL
EDITOR CIENTIFICO, A LANDA C.

CALLAO-PERU 1981

RECOMENDACIONES PARA COMPUTACION DE DATOS
DE CAPTURAS DE LA PESQUERIA DE CONSUMO
HUMANO

Informe del Consultor

William Hall

COMPUTER SPECIALIST

14 DE MARZO, 1980

1.1. CONSULTOR : W.B. Hall

FECHAS DE CONTRATO : 15 de Enero de 1980 - 14 de Marzo de 1980

LUGAR DE TRABAJO : Instituto del Mar, Perú

PROYECTO : PER/76/022

TERMINOS DE REFERENCIA

"Hacer un análisis del sistema existente en IMARPE para la colección, codificación y procesamiento de datos biológicos y estadísticos pesqueros procedentes de la pesquería de consumo humano. Desarrollar programas de computadora para analizar la información ya existente, teniendo en cuenta las instalaciones actuales y las que se pueden obtener a corto plazo".

"Preparar y conducir una serie de charlas destinadas a capacitar al personal en las técnicas de muestreo y el análisis de información biológico-pesquero".

"Ayudar en el planeamiento de trabajos futuros y plantear las posibles líneas de acción para mejorar la capacidad del Instituto para reco^llectar, manejar y analizar datos biológico-pesqueros".

"Preparar un informe final de las actividades y los resultados del trabajo realizado, con recomendaciones a la FAO e IMARPE".

VIAJE:	15 de Enero	Aberdeen - Roma
	16 - 18 de Enero	Roma
	19 de Enero - 11 de Marzo	IMARPE, Perú
	12 de Marzo - 14 de Marzo	Roma
	14 de Marzo	Roma - Aberdeen

GENERAL

La ocasión de esta consultoría no fue muy propicia. Se están dando con siderables cambios en cuanto a la forma en que se hará las computaciones en el laboratorio, esto tanto porque el equipo actual está siendo aumentado con partes adicionales como por, y principalmente, por la adquisición de una mi ni computadora. Aunque ninguno de estos nuevos aditamentos fue instalado du rante mi permanencia, todos ellos estarán en funcionamiento dentro de dos o tres meses. Además, el jefe del grupo de procesamiento de datos, mi contra parte, dejó el Instituto por otro trabajo a mediados de Febrero.

A pesar de todo esto, hubo varios requerimientos de consulta para ayu dar en el análisis de datos con el equipo existente, y gran parte de mi tiem po se usó en esos proyectos, incluyendo la preparación de algunos programas generales que pueden ser fácilmente transferidos a otra computadora. Sin em bargo, si esta consultoría hubiese coincidido con la instalación del nuevo equipo me hubiese ocupado más en establecer programas de entrenamiento para el personal, tal como se discute más adelante en este informe.

Quiero agradecer al Asesor Científico Principal del Proyecto, Mr. Homer Campbell por sus finezas durante mi estadía y por la forma en que sorteó mu chos de los obstáculos.

2.1. Situación actual de IMARPE

El Instituto tiene acceso a dos tipos de equipos de computación de datos.

(a) Una Hewlett Packard 9830 de mesa. Con una memoria de 4K, un traductor BASIC, lector de caseta, lector de tarjetas de 80 columnas y una impresora lenta. También hay un ploteador Hewlett Packard conectado a la calculadora pero no estuvo operativo durante mi tiempo de mi estadía.

(b) La computadora IBM de Pesca Perú. La mayor parte de los programas largos hechos por el personal del laboratorio están en FORTRAM para esta computadora. Si bien las facilidades que presta Pesca Perú son extensas, el centro de computación está situado a alguna distancia del Instituto. Naturalmente, los programas del Instituto tienen menor prioridad que los de Pesca Perú lo cual puede ocasionar demoras en el procesamiento de datos. Así también el personal gasta mucho tiempo a fin de conseguir transporte para llevar las tarjetas a tiempo para su computación.

El Instituto tiene un personal bastante numeroso ocupado con el procesamiento de datos. De este personal, dos pueden programar en BASIC para la H-P y tres en FORTRAM para la IBM. El resto del personal del grupo de computación se ocupa de la preparación de los datos, esto es, chequeando datos antes de su entrada, perforando tarjetas, codificando los datos en forma adecuada para su ingreso a las tarjetas. El jefe del grupo de computación, mi contraparte, quien también era el responsable de la mayor parte del trabajo de análisis de sistemas, dejó su empleo en el Instituto a mediados de mi consultoría.

La mayor parte de los trabajos hechos en la HP son análisis hasta una primera etapa solamente. Por ejemplo, hay varios programas para proyectar la frecuencia de longitudes de las muestras a nivel de toda la pesca de la embarcación, pero no hay programas para manejar los datos más allá de esta etapa. Esto se debe principalmente a la limitada capacidad del equipo, que no tiene medios de archivar la información excepto en cassettes, lo cual es muy lento, no muy confiable y de capacidad limitada pues el acceso a los datos es sólo secuencial. Por esta razón es que en setiembre de 1979 se pidió un sistema de disco para incrementar la capacidad de la computadora. Este equipo que podría mejorar notablemente la calidad y cantidad del procesamiento de datos que hace el Instituto no fue instalado en el tiempo de mi estadía.

Además del trabajo científico que se realiza en la calculadora, se gasta una buena cantidad de tiempo en programas administrativos para proveer listas de personal y varias clasificaciones.

- (a) programas para el procesamiento de datos.
- (b) programas estadísticos o matemáticos.
- (c) programas de dinámica de poblaciones.

2.1.1. Programas para el procesamiento de datos

Este tipo de programa se escribe normalmente para aplicarse específicamente dentro del laboratorio. Muchos de estos programas tienen elementos comunes como ser proyectar datos de frecuencias de longitudes a nivel de estrato; clasificar los datos de captura según el método, la estación, el área, etc. Los programas de este tipo deberían ser escritos en forma modular de modo que algunas de sus partes puedan ser extraídas y usadas como ladrillos para construir otros programas. FORTRAM, mediante su estructura de subrutinas, es ideal para este propósito, pero también programas en BASIC pueden ser escritos de tal modo que parte de ellos puedan ser extractados para ser usados en otros proyectos sin tener que comenzarse cada programa desde el primer paso.

A pesar de los intentos que se hagan para reducir el número de programas modulares, el trabajo en ellos representará la mayor parte del esfuerzo del grupo por los particulares requerimientos de cada caso.

2.1.2. Programas matemáticos y estadísticos

Solamente dos programas de estadística corrientes existen para uso de la HP9830, uno para la comparación de líneas de regresión y el otro para análisis de discriminantes. En un laboratorio científico existe generalmente una colección de este tipo de programas, así que es un tanto sorprendente que el grupo de procesamiento haya sido requerido para escribir solamente estos dos programas. Durante mi consultoría solamente dos programas más fueron escritos para ser incluidos en la colección, a saber, uno de regresión múltiple y otro para fijar una curva de regresión asintótica. (Ambos son descritos en detalle en la Sección 5).

El problema con esta clase de programas es que pueden no ser usados muy frecuentemente y que el usuario puede muy bien ha-

cer los cálculos necesarios con una calculadora corriente en menos tiempo que el necesario para escribir el programa. Además, la presión sobre un grupo programador con recursos limitados tiende usualmente a ser en el sentido de concentrarse en programaciones que tengan aplicación inmediata de modo que cálculos de este tipo son dejados para que los haga el mismo usuario.

Muchos de estos programas existen ya sea en la literatura o como parte de la colección de programas de muchas instituciones que han usado computadoras por algún tiempo. El problema consiste en hacer que la información existente esté disponible para la gente que realmente requiere su uso, es decir para pequeños grupos, y en asegurarse que estén en una forma que pueda adaptarse rápidamente a sus equipos de computación. Abramson (1971) ha proporcionado varios programas en FORTRAM. Pero, con la creciente popularidad de los mini y de los micro computadores, puede que FORTRAM no sea el lenguaje apropiado para la diseminación de programas de computación relevantes a la investigación pesquera. BASIC es ciertamente más común para computadoras pequeñas y se enseña ampliamente como un lenguaje introductorio. La creciente popularidad de PASCAL vierte a este lenguaje en un poderoso competidor para el primer lugar. De todos modos, la preferencia de un lenguaje no es tan importante como

- (a) La decisión de cuales programas deben ser los incluidos en una lista.
- (b) La forma en que debe ser presentada la información.
- (c) Quien va a coordinar la redacción y la distribución de los programas.

La agencia para coordinar la recepción y la distribución de esta clase de información es obviamente FAO. La documentación de programas podría tomar la forma de hojas sueltas de modo similar al Suplemento de Algoritmos de la "Comunicaciones de la Asociación de Maquinaria de Computación", podría ser trimestral, según la disponibilidad de programas. Obviamente, se requiere de un formato estándar que contenga, por ejemplo:

- (a) Una descripción del método usado para el análisis con sus referencias apropiadas. Esta debería ser muy explícita, ya que las referencias más recientes sobre la materia

no están siempre disponibles de inmediato.

(b) Una lista de programas. El idioma podría ser cualquiera de los más comunes, FORTRAM, BASIC o PASCAL. Los autores debieran ser advertidos para que escriban en una forma standard en cuanto sea posible y que eviten el uso de formas que dependen del sistema usado (o por lo menos que expliquen en detalle estas características puesto que pueden no ser disponibles en otro tipo de computadora).

(c) Un juego de datos para la comprobación acompañado de detalles precisos para el input y la impresión de los resultados esperados. También deben ser mencionados el tipo de computadora en la que el programa fue probado así como cualquier detalle adicional sobre la precisión requerida.

Hay una amplia variedad de programas para ser incluidos en tal esquema. En un comienzo deben ser programadas las técnicas más comúnmente usadas y básicas y luego proceder las más avanzadas pero que no son de aplicación frecuente. Moller's "Manual of Methods in Aquatic Environment Research Part 5 Statistical Tests" es una excelente fuente de las técnicas estadísticas más frecuentemente aplicadas que pueden ser programadas, así también Bazigo's "Applied Fisheries Statistics".

2.1.3. Programas de dinámica de poblaciones

Los mismos problemas señalados arriba existen para los programas de computadoras para dinámica de poblaciones. En realidad son todavía más especializados y normalmente no existen fuera de los laboratorios dedicados a la pesquería o de los centros de enseñanza. Una lista de las técnicas más frecuentemente usadas podría prepararse, escribir los programas correspondientes en la forma indicada más arriba y dar a todo el conjunto una amplia difusión entre los grupos interesados.

3.1. Perspectivas del procesamiento de datos en IMARPE

Dos equipos nuevos llegarán al Instituto durante 1980 y afectarán considerablemente el procesamiento de datos.

Como parte de un programa de ayuda directa y en relación con el buque de investigaciones "Humboldt" que el gobierno alemán está dando al Perú, se pondrá a disposición del personal de IMARPE dos sistemas de computación. Uno de ellos será instalado en el Instituto. Durante el

primer año de la instalación un grupo de científicos alemanes, asociados con el Instituto, serán responsables del conjunto de accesorios de la computadora. Este grupo consistirá de un ecólogo, un dinámico de poblaciones, un especialista en huevos y larvas y un programador/analista de computadora.

El equipo a instalarse es un HP 1000 mini-computadora con capacidades para time sharing. Quiere decir que varias personas podrán realizar trabajos muy diferentes simultáneamente. En lo que al usuario concierne, tiene la impresión de ser él la única persona que está usando la máquina ya que la computadora por sí misma controla el tiempo de las diferentes tareas. Esto aumentará considerablemente el potencial del Instituto para hacer computaciones y será una oportunidad para el mejoramiento de la calidad así como para el aumento de la cantidad de las tareas de procesamiento de datos. Una lista del equipo propuesto puede encontrarse en el Apéndice 1 junto con una breve descripción del hardware y el software. Esta computadora será ensamblada aproximadamente al mismo tiempo que la HP9830 es reforzada con la adición de un sistema de memoria de disco. Esta situación es un problema para el personal de computación. Como quiera que estos dos tipos de equipo no son compatibles, habrá que decidir cual de ellos será el preferido para concentrar el trabajo. Es obvio que la HP1000 ofrece mayores oportunidades y que a ella deberá transferirse todos los trabajos corrientemente realizados por la HP9830, después de comprobar su permanente necesidad. Eventualmente, todos los trabajos de computación incluyendo los que corrientemente están procesados en FORTRAM para la computadora de Pesca Perú deberán ser adaptados para la HP1000.

La llegada de esta nueva computadora necesitará un rearrreglo de las tareas del grupo de computación.

(a) Personal de operación

Por lo menos dos de los miembros del personal deben ser entrenados como operadores. La operación de la computadora no será, casi seguramente, una labor de tiempo completo, la razón principal para que haya dos operadores es la de disponer de cobertura total. Dentro del cuarto de computación sólo debe ser permitida la presencia de gente entrenada y de su supervisor, ningún otro personal debe ser permitido para la operación de la computadora.

El entrenamiento debe abarcar desde el arranque hasta la última operación de funcionamiento en una forma ordenada; el montaje y

desmontaje de paquetes de discos y rollos de cinta, la descarga de la impresora, etc. El Instituto puede considerar la operación de la computadora sin atención ninguna fuera de las horas normales de trabajo puesto que muchas tareas de computación pueden ser realizadas sin la intervención del usuario ni del operador.

(b) Personal de programación

Esta computadora permite la programación tanto en BASIC como en FORTRAM. Los dos idiomas son de uso común y en el Instituto hay personal que los conoce. Las minicomputadoras están orientadas normalmente hacia el uso del BASIC puesto que este idioma se desarrolló para aplicarse a interacciones terminales mientras que el FORTRAM es más apto para aplicaciones de alto contenido calculatorio. Los programas en BASIC pueden ser escritos y desarrollados más rápidamente que los programas en FORTRAM, y para tareas pequeñas el BASIC es ciertamente el más apropiado. Según el interpretador de BASIC que contenga la computadora, los datos alfanuméricos pueden ser manejados muy eficientemente por programas en BASIC. Esto puede tener un efecto de largos alcances en la situación presente en que se encuentran muchos de los documentos de registro en los cuales se gasta buen tiempo en poner la información numérica en tarjetas perforadas antes de entrarla a la computadora.

Es necesario entrenar a los programadores profesionales en BASIC puesto que este lenguaje parece ser el más flexible para el nuevo equipo y también es probablemente el que da resultados más rápidos. Se requiere entrenamiento a varios niveles.

(i) para asegurarse que el programador conozca las principales características del lenguaje. Esto incluirá todos los comandos de operación del sistema, los conceptos fundamentales de programación, el manejo de variables "lineales" y un conocimiento del manejo de los sistemas de files de discos, cassettes y cintas magnetofónicas.

(ii) para seguir desarrollando las habilidades de los programadores profesionales seleccionados con el fin de que sean programadores de sistemas. Esto puede resultar necesario debido a las limitaciones en cuanto a software que proporciona el fabricante. Por ejemplo, la entrada directa de los datos desde el terminal al disco se hará inicialmente mediante un programa proporcionado por el fabricante denominado editor de sistema. Este programa es generalmente bastante largo y trata de manejar varias posibilidades. Por esta

razón pueden presentarse dificultades en su uso para ciertas aplicaciones y puede resultar necesario que especialistas del Instituto desarrollen programas especiales. Los problemas sobre entrada y verificación de datos pueden requerir un conocimiento detallado del sistema de software proporcionado.

Como parte de la HP1000 figuran terminales gráficas. Para un uso completo y efectivo de estos poderosos auxiliares para el análisis y la interpretación, será necesario desarrollar software especial. Por lo menos uno de los programadores tendrá que ser entrenado en técnicas de programación gráfica al más alto nivel posible. Si bien los métodos gráficos no son de uso tan común como la programación estadística o de dinámica de poblaciones, esta es un área que debe ser incluida en la lista de tópicos discutida en 2.1.2.

(iii) para introducir una amplia clase de usuarios a la programación y análisis de computadora. Puesto que el sistema puede ser usado por varias personas al mismo tiempo, otros grupos además del personal de procesamiento deben ser motivados para usar las facilidades de computación ofrecidas. Entrenamiento en cuanto a como manejar un terminal, como engranarlo al sistema y como correr un programa será también necesario. Además, el entrenamiento en programación en BASIC debe ser impartido a miembros seleccionados de cada departamento, especialmente a aquellos que muestren interés y aptitudes para este tipo de trabajo.

El entrenamiento tendrá que ser muy específico. Los manuales que proporciona el fabricante normalmente no son adecuados para el entrenamiento de personas sin experiencia y por tanto debe prepararse material de entrenamiento especial para estos cursos. Las ilustraciones que se den en las clases deben ser engranadas a problemas específicos dentro de la actividad del laboratorio, v.g., además de enseñar al personal de otros departamentos los principios básicos del uso de computadoras y de la escritura de programas en BASIC, el curso debe incluir ilustraciones tomadas de la dinámica de poblaciones o de problemas hidrográficos.

El objetivo de todo este entrenamiento es mejorar la calidad del procesamiento de datos en el Instituto y reducir la cantidad de cálculos que se hacen a mano en cada departamento. El uso apropiado de la computadora liberará a los biólogos de cálculos repetitivos y les permitirá avanzar en proyectos más apropiados tales como

el establecimiento de programas para la determinación de edad de especies seleccionadas.

Se está preparando un cuarto especial con ambiente controlado para el uso de la computadora. Este equipo está mucho más expuesto a los daños por fluctuaciones del voltaje y las partes sensitivas tales como los discos y cintas magnéticas pueden dañarse cuando ocurre una falla violenta de energía. Es esencial que esta clase de equipo sea mantenido por un calificado ingeniero electrónico en funciones permanentes. La mayor parte de las instalaciones tienen un contrato de servicio con los representantes de los fabricantes y es muy recomendable que IMARPE haga un arreglo con Hewlett Packard para mantener su equipo una vez que el proyecto sea entregado por los alemanes. Normalmente un contrato de mantenimiento cuesta alrededor del 10% del costo anual del equipo.

4.1. Análisis de datos de muestreo biológico por computadora

El objetivo de un esquema de muestreo biológico es proporcionar estimados actuales de las principales características de la población en estudio. Un análisis por computadora debe juntar los diversos elementos de las estadísticas de captura y esfuerzo, frecuencias y distribución de longitudes y, cuando sea aplicable, proporcionar estimados de las cantidades pescadas en cada grupo de edad. Además, un sistema completo debe también proporcionar información sobre control y administración para permitir que el personal supervisor monitoree los avances.

La intención original de este trabajo fue ver si un sistema unificado de análisis de datos podría iniciarse a fin de prevenir la duplicación de programas que performan esencialmente los mismos cálculos. Sin embargo, muy pronto resultó aparente que hay tantas diferencias importantes en la forma de coleccionar los datos de merluza que se decidió que era necesaria la existencia de dos sistemas distintos de computación. Muchas de las partes componentes de estos programas son similares y pueden ser escritos en tal forma que eviten la duplicación de esfuerzo. El método de análisis que aquí se propone es principalmente para el procesamiento de datos de merluza, pero las líneas generales son similares para otras especies.

Primero es necesario definir los estratos de muestreo o clasificaciones según las cuales se colecta los datos. Para un esquema de computadora la lista debe ser completa e incluir todas las clasificaciones probables de darse durante la vida del programa. Tiempo, espacio, tipo

de arte, son generalmente las clasificaciones principales, considerando se cada unidad homogénea desde el punto de vista del muestreo. Por el momento las unidades de muestreo son el mes, cuadrado Marsden, flotas de alta mar y costeras y, como hay importantes diferencias entre los dos sexos, el sexo también es una clasificación esencial.

Para el análisis de muestreo de mercado hay tres elementos de los datos que tienen que juntarse en el sistema de computación.

4.1.1. Estadísticas de captura y esfuerzo

Este es el elemento más vital. Sin un método seguro y eficiente de coleccionar estadísticas de captura el resto del programa de muestreo biológico resultará menos efectivo puesto que el objetivo de muestreo es proyectar las estadísticas a toda la captura.

La clasificación para la captura es:

- | | |
|---------------------------------|------|
| (i) flota de alta mar y costera | (2) |
| (ii) mes | (12) |
| (iii) cuadrado Marsden | (8) |

La Zona 1 tiene índices del 1 al 5 y Chicama, Chimbote, Samanco y Casma son los puertos para esta zona. Para la Zona 2 los puertos son Huarmey, Supe, Vegueta, Huachi y la posición de los índices es del 6 al 9. Para la Zona 3, Chancay, Callao y Pucusana con índices de posición 10 - 11. Zona 4 con puertos Tambo de Mora y Pisco e índices 12 - 15; y para la Zona 5 los puertos de Atico, La Planchada, Mollendo e Ilo y los índices de posición del 16 al 23.

Ejemplo 1. Puerto: Chimbote Area: Chao Profundidad: 49
Por el código del puerto CHIM se decide buscar la posición del file desde el índice 1 al 5 a fin de encontrar la posición de CHAO. Esta es encontrada en el índice 4. Con la profundidad 49 se busca el file de posición 4. Esta no es mayor que la profundidad 50 en el índice 3 dentro de la posición del índice 4. Ahora se accede directamente al file de zona isoparalitoral al índice (I, J) 4,3 para obtener 3083.

Esto es simplemente un esbozo de un método que podría ser usado. La razón principal para presentarlo en detalle es tratar de evitar la transcripción de los datos de documentos primarios a documentos secundarios antes de seguir con el procesamiento. Por lo general esta es la razón por la cual se dan las demoras en la

etapa de procesamiento y por lo que a menudo el personal experimentado puede hacer una entrada alfanumérica directamente desde el documento primario de registro.

Hay varios puntos que deben notarse acerca de este método.

(a) Debe tomarse medidas para la inclusión de nuevas posiciones. Por ahora la mayor parte de las posiciones más comunes de pesca deben estar definidas y las otras posiciones pueden fácilmente ser añadidas al final de la lista pero dentro del índice correspondiente.

(b) Entre corrida y corrida del programa, todos los files estarán contenidos en disco. Sin embargo deben ser trasladados directamente a la memoria de uso al comienzo de cada corrida para evitar una búsqueda excesiva del contenido de los discos.

(c) La profundidad asociada con algunas de las posiciones todavía tiene que ser definida completamente.

(d) El file de zona no es absolutamente necesario ya que el área isoparalitoral puede ser calculada del índice con numeración I, J mediante la expresión.

$$1000 * J + 5 * I + 65 \quad I \text{ impar}$$

$$1000 * J + 5 * I + 68 \quad I \text{ par}$$

Los tipos de artes están diferenciados pero parecería más realístico agregar inicialmente el arrastre para la flota costera. Se podría preveer la captura de merluza por otras flotas que usan artes pelágicas. Si fuera posible obtener información segura de las capturas de estas artes, no habría problema en incluirlas en este esquema.

Información bien detallada sobre la captura y el esfuerzo de la flota de alta mar se obtiene corrientemente a partir de información por lance de embarcaciones escogidas. El muestreo hecho por observadores a bordo debe dar buenos estimados de las capturas de merluza dentro de la captura total y podría escribirse un pequeño programa para entrada de datos en su apropiada clasificación.

Los detalles de captura y esfuerzo de la flota costera están disponibles en tarjetas perforadas. Varios análisis preliminares se requieren antes de poder aceptarse la validez del agregado de

los datos de todas las embarcaciones, pero si se asume que un mes por área es una clasificación apropiada para la flota costera se podría escribir un programa para agregar los datos en el file de captura y esfuerzo.

Se asume que este proyecto será programado para la HP1000 y que el lenguaje será el BASIC. Los interpretadores para BASIC permiten normalmente secuencias de dos vías mientras que las secuencias discutidas aquí, y que serán guardadas en discos, son por lo menos de tres vías. Un procedimiento simple de índices puede resolver este problema. Por ejemplo, en estos datos se dará espacio para tres flotas, doce meses y ocho áreas. Se almacenan tres medidas, captura, horas de pesca y horas de pesca X el tonelaje bruto de registro. Se asume también que la HP1000 tiene un sistema de almacenamiento virtual en su disco de modo que las secuencias de los discos pueden ser manejadas por el programa como si estuviesen en la memoria y que cada elemento puede ser directamente llamado.

La secuencia del disco puede ser lineal, siendo su dimensión el producto de las clasificaciones $(3 \times 12 \times 8 \times 3) = 864$. La disposición del almacenado es

istema clasificatorio para los datos de captura y esfuerzo

	ENE.			FEB.			MAR.			C	DIC.		
	C	HF	TH	C	HF	TH	C	HF	TH		C	HF	TH
Flota 1 (A) Area 1	C	HF	TH	C	HF	TH	C	HF	TH	C	C	HF	TH
Indice	1	2	3	4	5	6	7	8	9	34	35	36
(B) Area 2	C	HF	TH	C	HF	TH	C	HF	TH			
Indice	37	38	39	40	41	42	43	44	45		70	71	72
.													
(H) Area 8	C	HF	TH										
	254	254	255								286	287	288
Flota 2 Area 1	C	HF	TH								C	HF	TH
	289	290	291								322	323	324
Area 8											574	575	576
Flota 3 Area 1	577	578	579								610	611	612
.													
Area 8	829										862	863	864

C= captura HF= Horas de pesca TH= CRT X horas de pesca
 Para acceder cualquier lugar de almacenamiento el método de índice es

$$(8 \times 12 \times 3) \times (\text{Código flota} - 1) + (12 \times 3) \times (\text{Código área} - 1) + 3 \times (\text{Meses} - 1) + D$$

- D = 1 Captura
- = 2 Horas de pesca
- = 3 Ton. X Horas

Por ejemplo si se desea conocer la captura para diciembre (12) para la Flota 2 y el área E (5) el índice es

$$(8 \times 12 \times 3) \times 1 + (12 \times 3) \times 4 + 3 \times 11 + 1 = 466$$

y para la flota 1 Mes 2 Área B el índice para tons X hora es

$$(8 \times 12 \times 3) \times 0 + (12 \times 3) \times 1 + (3) \times 1 + 3 = 42$$

Este esquema de índices es muy flexible y los datos pueden ser agregados rápidamente a partir de un file arreglado de esta manera. Se requiere un conocimiento más detallado de la computadora para ver si el esquema que se ofrece aquí es el óptimo puesto que esto depende de las varias maneras en que pueden agregarse

los datos y en la cantidad de información transferida al elemento intermediario entre el disco y la memoria de trabajo. El límite para el índice más alto probablemente será 32,767 si la HP1000 está limitada a los 16 bit palabra de los elementos de llamada.

4.1.2. Datos de frecuencia de longitud

Normalmente se toman muestras replicadas de la longitud de merluza dentro de cada estrato. Se toman dos clases de datos una para una mezcla de peces machos y hembras y la otra con identificación de sexo. Es obvio que la gran diferencia entre machos y hembras hace necesario que los datos de cada sexo sean considerados separadamente.

Aunque L_{50} para machos es considerada como 68 cm y para las hembras mayor de 100 cm la pesquería generalmente se limita entre 30 y 80 cm. El almacenaje de los datos podría entonces tener estos últimos límites poniéndose todo lo que es menor de 30 en la categoría 30 y los que es mayor de 80 en la categoría 80. No es probable que este agrupamiento abierto en los extremos pueda sesgar cualquier medida de población. El arreglo para el almacenaje de los datos de longitud en discos podría tomar la forma: 3 flotas X 12 meses X 8 áreas X 2 sexos X 51 clases = 29,376 posiciones. La estructura del file podría ser como la que se describe más adelante y como el BASIC comienza en el lugar con índice 0 la DIMENSION sería (29375)

Estructura del file para los datos de frecuencia de longitud

		ENERO				FEBRERO				DICIEMBRE			
ta	Area Sex Long.	30	31	32...80	30	31	32...80		30	31	32...80		
1	M Indic.	0	1	2...50	102	103	104 152	1122	1123	1124..1172		
	F	51	52	53...101	153	154	155 203	1173	1174	1175..1230		
2	M	1124	1225	1226 1274	1326		1376						
	F	1275		1325	1377		1427		2447		
			
			
			
8	M	28152	28153...	28205	28254	28304		29274	29324		
	F	28203	28254	28305	28355		29325	29375		

El índice para este esquema de file es

$$(12 \times 8 \times 2 \times 51) \times (\text{Código Flota} - 1) + (51 \times 12 \times 2) \times (\text{Código Area} - 1) \\ + 51 \times 2 \times (\text{mes} - 1) + 51 \times (\text{Código sexo} - 1) + 1$$

Código Sexo = 1 para machos

2 para hembras

I = 00 a 50 para longitudes 30 a 80 cm

Ejemplo: La frecuencia de un pez macho de 30 cm en Feb. de la flota 3 en el área 8 es

$$(12 \times 8 \times 2 \times 51) \times 2 + (51 \times 12 \times 2) \times 7 + (51 \times 2) \times 1 + (51 \times 0) + 0 = \\ \text{índice } 28254$$

Es probable que este sea un arreglo razonablemente óptimo de almacenaje ya que las distribuciones de las longitudes de machos y hembras dentro de un estrato ocupan áreas contiguas y esta es la agregación más probable para posteriores programas de análisis. La dimensión de esta matriz lineal es bastante próxima al máximo permitido de modo que si se considera más de 80 áreas este arreglo tendrá que ser dividido en dos files diferentes para cada sexo.

Para entrar estos datos deberá usarse directamente la entrada terminal mientras que la entrada por programa debe ser reducida a un mínimo. Debe también considerarse la verificación leyendo los datos originales de los discos y chequeándolos bajo control de programa. Una manera posible de entrar los datos de frecuencia

de longitud una vez que se ha entrado los encabezamientos (área, flota, mes, sexo, peso desembarcado, peso muestreado) es identificar la primera longitud como negativa y de aquí en adelante entrar cada observación acompañándola de RETURN, la última frecuencia también debe estar precedida por un signo negativo. Por ejemplo

LONGITUD	FRECUENCIA	PUEDE ENTRARSE COMO
30	1	- 30
31	2	1
32	5	2
33	0	5
34	4	0
35	2	4
36	1	2
37	0	1
38	1	0
39		- 38
40		1

y el programa podría localizar esta información en la posición apropiada a una longitud.

Hay muchas ocasiones en las que las muestras de longitud se toman sin saberse el peso de la muestra. Los datos de este tipo podrían muy bien ser proyectados al nivel de bote calculando el peso de la muestra estimado según alguna de las relaciones dadas en la sección 5. Estas relaciones podrían ser incorporadas a los programas. Se asume que por lo menos una muestra de longitud por sexo se colecta de cada estrato y que el último cálculo (después de que cada muestra ha sido proyectada al nivel de embarcación y sumada a través del estrato) es dividir la frecuencia entre los sexos. Los datos son entonces proyectados a nivel de flota recurriendo a los datos de captura y la estimación final de población se almacena en file de disco de la distribución de longitudes tal como se ha descrito más arriba.

4.1.3. Datos de edad

Se asume que se dispondrá de datos de edad para este análisis de computadora. Aún si los datos de edad no están disponibles para cada estrato, es posible utilizar los datos de longitud/edad de otro estrato, probablemente contiguo. Para que tal cosa suceda, es fácil diseñar un dispositivo apropiado en el programa, el dispositivo debe permitir al biólogo operador designar

una alternativa cuando falta una edad longitud e una muestra.

Los datos de longitud edad pueden ser alimentados a la computadora de dos modos, o como una serie de longitudes, edades y sexos individuales o tomando los mismos datos pero clasificados en forma de una distribución de frecuencias. El primer método es preferible en cuanto reduce la preparación manual de los datos, pero es mucho más probable que la clave longitud edad incorporada a la computadora no coincida exactamente con la distribución de longitud proveniente de la muestra de flota para cada intervalo de longitud. Si la clave longitud edad no tiene los datos correspondientes a un intervalo de longitud que aparece en la muestra, es necesario que exista algún modo automático para proporcionar la información que falta; de otro modo el programa se detendrá y los datos tendrán que ser devueltos al biólogo para que proporcione la observación adecuada. La existencia de un sistema automático es lo preferible.

Para los grupos de edad de merluza de 1 a 10 podría considerarse que el requerimiento de espacio es de 10 clases anuales X 3 flotas X 12 meses X 8 áreas X 2 sexos = 5760 posiciones de almacenaje. Para un análisis completo se requiere tres files de esas dimensiones:

- (1) números a cada edad dentro de cada estrato
- (2) longitudes promedios a cada edad dentro de cada estrato
- (3) pesos promedios a cada edad dentro de cada estrato

Los números a cada edad dentro de cada estrato se calculan a partir de la tabla longitud edad que ha sido proyectada a los desembarques en número de cada flota, y el mismo procedimiento es válido para las longitudes promedios y los pesos promedios para cada edad. Los pesos promedios deben ser calculados usando una ecuación para cada mes que debe estar guardada en la computadora y convirtiéndolo la tabla longitud edad en una tabla peso edad.

La clasificación detallada de los índices para estos tres files va de acuerdo con el mismo procedimiento indicado más arriba para la clasificación de índices de los files de distribución de longitudes

Cuando todos los datos han sido juntados en estos files, ya no es necesario retener los datos originales en la computadora y las observaciones entonces pueden archivararse en cinta magnética. Teniendo los datos en esta forma cualquier combinación de ellos es

posible, los programas para obtener, por ejemplo, la proyección de la distribución de frecuencias a nivel anual, o los números para cada edad a nivel de estación, etc., resultan razonablemente directos y fáciles de escribir. Si se puede introducir a la computadora los datos de años anteriores en la forma indicada, estos datos constituirían una base bastante flexible para la subsecuente obtención de otros estimados que se requieren.

Como se puede ver en la tabla de coeficientes de regresión mensuales, estos varían considerablemente. Para el período de dos años considerado el coeficiente "b" para machos varía de 1.70 a 2.39 y aquel para hembras de 2.22 a 3.20. Los coeficientes para junio son especialmente bajos en ambos sexos y años. El nivel generalmente bajo para los machos resulta probablemente de un limitado rango de longitudes de las muestras. El error estándar para la mayoría de las pendientes individuales es del orden de + 0.1.

Los datos para machos fueron separados de los de hembras puesto que a simple vista sus pendientes difieren notablemente. Para el caso de los machos, un análisis de variancia demostró que no hay diferencias de mes a mes.

A pesar de las diferencias significativas entre las pendientes de cada año, es necesario obtener un resumen de los datos a fin de resaltar los rasgos principales, en consecuencia se decidió combinar las pendientes de los dos años. Esto se hizo ponderando cada pendiente promedio de 2.1214 y de 2.9405 para las hembras. Los datos del 78 y el 79 se combinaron para dar las siguientes relaciones mensuales.

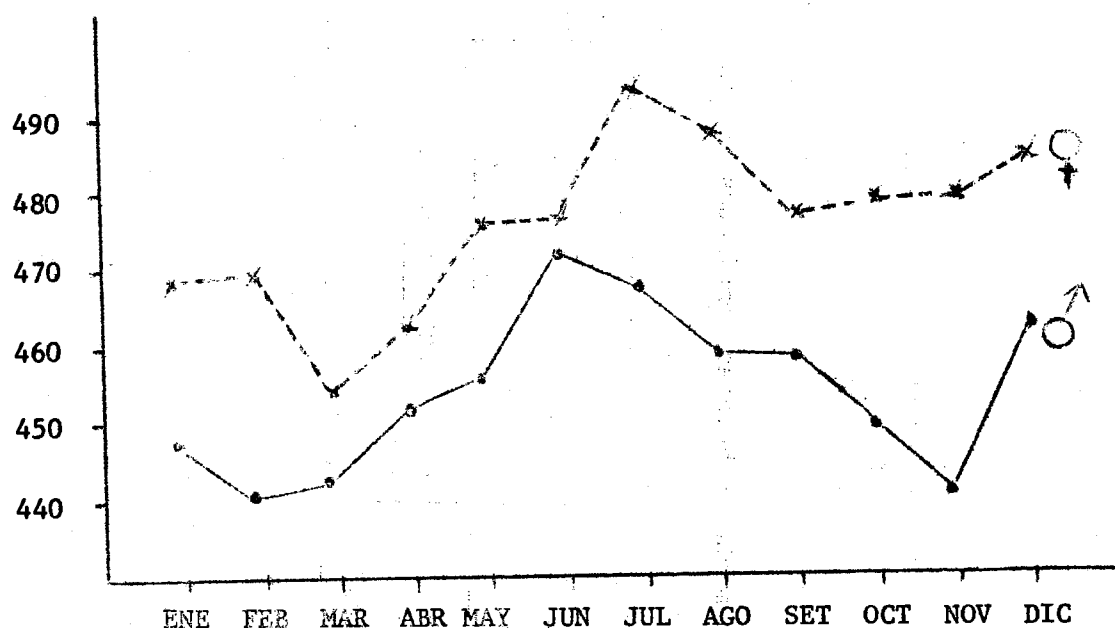
Coeficientes deregresión "a" para los datos mensuales combinados de 1978 y 1979

	ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO	
Machos	0.1790	0.1760	0.1770	0.1810	0.1825	0.1884	
Hembras	0.009114	0.009168	0.008859	0.009027	0.009271	0.009270	
	JULIO	AGOSTO	SEPTIEMBRE	OCTUBRE	NOVIEMBRE	DICIEMBRE	b
Machos	0.1867	0.1837	0.1839	0.1799	0.1766	0.1846 *	2.1214
Hembras	0.009638	0.009520	0.009279	0.009311	0.009296	0.009410 *	2.9405

* sólo datos 1978.

Estas ecuaciones podrían incluirse en un programa de computadora para poder estimar el peso en un mes particular cuando sólo se dispona de la longitud. Como se puede ver en la figura 1, hay un definido patrón estacional de los cambios en el peso para una longitud dada; para los machos este cambio es de un 7% entre mínimo y máximo y de un 9% para las hembras. Estas magnitudes porcentuales de variación deben ser tomadas en cuenta para un análisis más detallado de los datos de merluza llevado a cabo por computadora.

Variación estacional del peso calculado para peces de 40 cm.



Además de las relaciones separadas para machos y hembras, se necesita también ecuaciones para estimar las muestras combinadas. Asumiendo que la razón entre sexos a través de todo el año es la misma que la observada en estas muestras, se obtuvieron ecuaciones para machos y para hembras combinando los estimados mensuales de los dos años tal como sigue.

Relaciones longitud/peso para machos y hembras combinadas, datos de 1978 y 1979.

b
2.7579

ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO
0.01771	0.01764	0.01769	0.01769	0.01814	0.01830
JULIO	AGOSTO	SEPTIEMBRE	OCTUBRE	NOVIEMBRE	DICIEMBRE
0.01876	0.01857	0.01809	0.01813	0.01781	0.01846

Es obvio que estas relaciones pueden todavía combinarse más ya que entre algunas de ellas no hay diferencias significativas. Sin embargo, no hay mucha ventaja en proseguir la combinación puesto que lo que se desea es usar estas ecuaciones en un programa de computación para analizar los datos de frecuencia de longitud.

No ha sido muy afortunado que los datos para diciembre de 1979 no estuvieran disponibles al momento de hacer este análisis, porque los valores para ese mes se salen por lo general del patrón general de las otras observaciones. El todo se puede ajustar cuando los datos de diciembre están disponibles.

4.1.4. Control de files

Como los files de datos van recibiendo datos para varios programas en varias oportunidades a lo largo del año, es conveniente establecer también files de progreso a fin de monitorizar la entrada de datos. Además, puede que haya la necesidad de proporcionar a la administración información sobre, por ejemplo, la cantidad de peces medidos en cada área en cada mes, o el número de otolitos leídos en cada estrato. Información de este tipo requiere la instalación de files adicionales.

(i) Files para monitorizar la entrada de datos

Hay que entrar tres conjuntos distintos de datos antes de hacer un cálculo completo para un estrato, a saber, debe entrar se la captura, igualmente tiene que entrarse y verificarse las muestras de frecuencia de longitud de todas las embarcaciones para el estrato, así también los datos de longitud edad. Cada una de estas entradas pueden hacerse en tiempos diferentes, pero la verificación de la entrada de los datos de longitud edad es la que determina que los cálculos se inserten en los files de datos biológicos. Para asegurarse que en el estrato están presentes todos los datos se puede preparar un file de control en el que se coloca un 1 cuando los datos de captura son perforados al cual se le añade otro 1 cuando los datos de captura han sido verificados. Cuando los datos de longitud son perforados se añade 10 a la localización apropiada y cuando estos datos son verificados debe añadirse 100 al file. Para los datos de longitud edad los números respectivos son 1000 para la entrada y 10,000 para la verificación.

Se puede escribir un pequeño programa para exhibir el contenido de este file, el cual tendrá las mismas dimensiones que

el del número de estratos. Una entrada de 11,112 dentro de cualquier célula significa que todos los datos para ese estrato están disponibles pero 00112, por ejemplo, indicaría que los datos de captura han sido entrados y verificados, que los datos de longitud también han sido entrados y verificados pero que los datos de longitud edad están todavía ausentes.

(ii) File para la colección de datos

El propósito de este file es proporcionar información al día acerca de los datos que han sido recolectados. Por ejemplo, el número de embarcaciones muestreadas dentro de cada estrato puede calcularse como parte de los datos de frecuencia de longitud entrados, así también puede considerarse el número de peces medidos. Esta información puede insertarse en un file clasificado en la forma descrita arriba y para exhibirla es posible escribir los apropiados programas.

(iii) File para las zonas insoparalitorales

Para cada rango de profundidad hay una disposición de zonas paralitorales. La lista de las clasificaciones de estas zonas, en el mismo orden de los rangos de profundidad, puede incluirse en el file de zonas de disco.

File para zonas isoparalitorales

	Indice	1	2	3	4	5
Posición	1	1070	2070	3070	4070	
	2	1073	2073	3073	4073	
	3	1080	2080	3080	4080	
	4	1083	2083	3083	4083	5083
	5	1090	2090	3090	4090	
	6	1093	2093	3093	4093	
	.					
	.					
	.					
	.					
	.					
	.					
	.					
	.					
	.					
	.					
	.					
	N	1180	2180	3180	4180	

Este conjunto de files proporciona un marco básico que después puede desarrollarse. Se puede sugerir varias formas de salidas, pero en esta etapa lo que se intenta es sólo proveer de una base de datos fácilmente entendible y de la cual los datos puedan ser extractados y combinados mediante programas interactuantes. Los biólogos pueden ellos mismos iniciar estos programas de acuerdo a la forma de output que sea requerida. Se asume que todos los programas para esta parte serán escritos en BASIC que ofrece un tiempo de desarrollo considerablemente más corto que el que toma un programa en FORTRAM para la HP 1000. La tabla que sigue es un breve sumario de los principales rasgos de este sistema y muestra la asociación sugerida entre varios files.

Los principios generales de este modo de atacar el problema pueden aplicarse a otras especies como sardina, jurel, etc. aunque en esta sección la única especie considerada ha sido la merluza.

Resumen de los Programas de Entrada de Datos que se Necesitan
para crear una Base de Datos

DATOS	OPERACION	FILES
(1) Entrada de captura	Entrar la captura en disco. Datos de mantenimiento en forma sin estructura hasta que se verifican	El file monitor añade 1
(2) Verificación de captura	Verificar captura en línea	El file monitor añade 1 El file de datos de captura es puesto al día para los estratos entrados.
(3) Entrada de datos de longitud	Entrar datos de frecuencia de longitud en cada embarcación. Los datos se mantienen en forma sin estructura hasta que se verifiquen.	Un 10 es añadido al file monitor de entrada
(4) Verificación de datos de longitud	Verifica los datos de frecuencia de longitudes entrados en (3). Chequea que los datos de captura requeridos estén disponibles, proyecta la captura al nivel del peso desembarcado por la embarcación y suma las muestras replicadas dentro del estrato. Finalmente proyecta la distribución de frecuencias al nivel de la captura de la flota dentro del estrato, lo cual se almacena luego en el file apropiado.	Añade 100 a los datos del file monitor de entrada. Entra el número de embarcaciones muestreadas y el número de peces medidos en el file de selección de datos Chequea los datos de captura disponibles. Entra la distribución de frecuencias para la flota en el file de frecuencias de longitud.
(5) Entrada de datos de longitud edad	Entrar los datos de longitud edad y sexo por estrato. Los datos se mantienen en forma sin estructura hasta ser verificados.	Se añade 1000 a los datos del file monitor de entrada.
(6) Verificación de datos longitud edad	Verifica los datos de longitud edad. Proyecta la clave longitud de la flota y calcula los números a edad, las longitudes y pesos promedio a edad. Produce un informe sumario de los detalles del estrato. Los resultados se almacenan en los files apropiados.	Se añade 10000 a los datos del file monitor. El número de otolitos leídos disponibles en el file de colección chequea que los datos para proyectar la distribución de frecuencias de este estrato estén disponibles.

Da acceso al file de numerosa edad en la clasificación apropiada. Almacena las longitudes promedio a edades en los files respectivos.

Un ejemplo de los cálculos se da en el informe de Armstrong Dic. 1979.

5.1. Análisis de longitud peso para jurel

El propósito de este análisis fue el proporcionar ecuaciones usables en la computadora para evaluar el peso a edad además de la longitud a edad. Con frecuencia existen diferencias estacionales, así como también diferencias por área, en las relaciones longitud peso. Los datos para estos análisis fueron proporcionados por la sección de alimento, donde se colectó los datos y se hicieron análisis hasta alcanzar ecuaciones de regresión para los datos de áreas y trimestres individuales. Los análisis de este tipo general deben hacerse usando todos los datos existentes sobre longitud/peso a fin de establecer los patrones estacionales y poder incluir las posibles diferencias por área en los análisis de rutina del programa de muestreo de mercado.

Los datos para este análisis consistieron en los de peso y longitud colectados en el verano, otoño, invierno y primavera en cinco puertos, Paíta, Chimbote, Callao, Pisco e Ilo que abarcan toda la extensión de la costa peruana. La Tabla 5 da los coeficientes de pendiente (b) de las regresiones individuales entre log longitud y log peso. Esta tabla también incluye el número de observaciones en cada muestra.

Pendientes de la Regresión para cada Area y Estación y Numero de Observaciones en cada Muestra

	Pendientes de Regresión					N° de Observaciones				
	PAITA	CHIMBOTE	CALLAO	PISCO	ILO	PAITA	CHIMBOTE	CALLAO	PISCO	ILO
VERANO	2.79	3.05	2.95	2.94	2.80	275	52	290	179	231
OTOÑO	2.84	3.16	2.81	3.25	2.88	550	253	414	217	210
INVIERNO	2.80	3.11	3.23.26	3.22	3.13	190	37	344	75	212
IMAVE.	2.69	2.68	2.78	2.81	2.84	450	224	286	156	73

Aunque normalmente se comparan las ecuaciones de regresión usando los promedios de cuadrados dentro de cada muestra como una estimación del

error, se ha considerado que es apropiado para estos datos comparar las pendientes mediante un análisis de variancia de dos entradas sin replicas entre las clases. La tabla de análisis de variancia es:

AV. Fuente de Variación g.l.	S de C	C.M.	F
Puertos 4	0.2031	0.0508	3.79 (P< 0.05)
Estaciones 3	0.3134	0.1055	7.87 (P< 0.01)
Resto 12	0.1603	0.0134	
TOTAL 19	0.6768		

La fuente mayor de variación entre puertos está en las estimaciones de "b" para Paita que son consistentemente más bajas que las demás. Al remover esta área y rehacer el análisis ya no aparece la diferencia entre puertos, pero si diferencias significativas entre estimaciones. Los promedios estacionales dependiente

Verano	Otoño	Invierno	Primavera
2.93	3.02	3.18	2.78

El error estándar de la diferencia entre dos promedios cualesquiera es ± 0.09 .

Cada ecuación individual fue recalculada con la pendiente apropiada para la estación. La comparación entre las entersecciones ajustadas no demostró diferencia entre áreas pero si entre estaciones.

En resumen, se necesita cuatro ecuaciones separadas para representar la relación entre la longitud y peso totales del jurel. Las ecuaciones para las áreas Central y Sur son:

Verano	$W = 0.009737 L^{2.9337}$
Otoño	$W = 0.006677 L^{3.0249}$
Invierno	$W = 0.003607 L^{3.1802}$
Primavera	$W = 0.01767 L^{2.7773}$

Para Paita también es apropiado usar una relación diferente para cada estación

Verano	$W = 0.01781 L^{2.7906}$
Otoño	$W = 0.01465 L^{2.8357}$
Invierno	$W = 0.01646 L^{2.8029}$
Primavera	$W = 0.02494 L^{2.6948}$

Estas ecuaciones pueden ser incluidas en un programa de computación para evaluar el peso cuando se da la longitud para un área y una estación determinada.

5.2. Análisis de regresión múltiple para experimentos de metabolismo de ancho- veta

Un científico de IMARPE solicitó ayuda en relación a un problema de análisis de regresión múltiple. El propósito era la estimación de logaritmo natural de la tasa de metabolismo como una función de log.nat. del peso en gm, de la temperatura del agua en grados centígrados, de la velocidad en cm por segundo y del factor de condición del pez. La evaluación del pez. La evaluación de la ecuación para diferentes conjuntos de datos había sido hecha a mano, pero anteriores conjuntos de datos habían sido evaluados con un programa incluido en el paquete de estadísticas para las ciencias sociales (SPSS) de una Universidad Norteamericana. En el Instituto no existía un programa archivado para este problema.

Se escribió un programa de análisis de regresión múltiple para la HP9830 en BASIC. Fue escrito en términos generales para que pueda ser incluido en el archivo general de IMARPE. En su forma presente es bastante limitado aunque suficiente para el problema particular para el que fue escrito.

5.2.1. Programa: MURLEG

Idioma : HP9830 BASIC

Descripción: Fijar una ecuación de regresión múltiple a un conjunto de hasta 10 variables

Este programa calcula:

- (a) La matriz de correlación de las variables independientes.
- (b) El análisis de variancia de todas las variables incluidas en el análisis. En este programa no se incluyen procedimientos paso por paso, lo cual es una obvia limitación. Un pequeño programa general de este tipo requeriría un análisis de variancia de los efectos de incluir cada variable después de haberse ajustado las anteriores. A lo largo de un proceso paso a paso de esta naturaleza también deberían calcularse y mostrarse los coeficientes de regresión asociados así como los errores estándar de dichos coeficientes.
- (c) Los coeficientes de regresión para todas las variables independientes incluidas en el análisis y los valores T que se obtienen dividiendo los coeficientes de regresión por sus respectivos errores estándar.

Este programa fue probado usando los datos que se dan en Snee decor and Cochran "Statistical Methods" pp.405-408. Se acompaña

las entradas y salidas de prueba así como un listado del programa. El programa también se probó con la corrida de datos del programa de multiple regresión del paquete SPSS en su versión 7.0.

Este programa BASIC utiliza como nombres de las variables una letra y un número. En caso de extenderse este programa se aconseja que se cambie el nombre de las variables por algo de mayor significado si el "interprete" de la computadora lo permite.

El método para calcular la matriz de información

$$// \sum (X_i - \bar{X}_i) (X_i - \bar{X}_i) //$$

está diseñado para reducir la acumulación de errores por redondeamiento a un mínimo. El algoritmo usado para la suma de cuadrados

$$SOS = SOS + (I - I) * (X - XBAR) * (X - XBAR)/I$$

y para el promedio

$$MEAN = MEAN + (I - I) * MEAN/I + X/I$$

La rutina para la inversión de matrices es la que proporciona la math matrix ROM del calculador. Para reducir los errores de cálculo en la inversión, la matriz de información se transforma a la matriz de correlación antes de la inversión, y luego se reconvierte después de la inversión. Esto resulta en una razonable correspondencia con los valores usados para la prueba, pero no se ha tratado de ver como funcionaría el programa en caso de variables independientes altamente correlacionadas.

5.2.2. Entrada de datos

- 1 V - El número total de variables en el conjunto de datos que va a analizarse incluyendo las dependientes e independientes. Los datos se entran en el orden $X_1, X_2, X_3, \dots, X_{V-1}, Y$
- 2 N - El número total de observaciones en el set. Una observación contiene V - variables.

3 Variables 1 - V Observación 1

4 Variables 1 - V Observación 2

5 " " " " " " " " " " " " 3

6 " " " " " " " " " " " " 4

N+2 " " " " " " " " " " " " N

Con el listado del programa se acompaña un conjunto de datos como ejemplo.

5.3. Método propuesto para el análisis de los Partes de Viaje de anchoveta

Las hojas diarias de bitácora que cada pescador de anchoveta llena para dar detalles del esfuerzo y captura de cada lance están en uso actualmente por más de cinco años. El método actualmente en uso para analizar los resultados requiere una considerable cantidad de codificación y perforación de tarjetas antes de poder entrar los datos a la computadora. Esto consume mucho tiempo y la transcripción de datos de una parte a otra del documento puede resultar en errores y en una considerable atrazo de los análisis de captura y esfuerzo.

El ítem que ocupa la mayor parte del tiempo es la codificación de la identificación de la zona paralitoral, que no es dada directamente por el pescador, pues él generalmente proporciona una localidad y una profundidad. IMARPE ha preparado una lista de localidades asociadas con profundidad y de ellas es de donde se saca la zona.

El actual sistema de computador no permite cambios significativos en el método de análisis, pero la Hewlett Packard 1000 proporciona nuevas oportunidades. Este equipo es capaz de manejar datos alfa-numéricos (string variables) y no está limitado por las convenciones rígidas con que se opera los formatos de tarjetas perforadas. Puede ser conveniente considerar la entrada de más información alfabética y dejar a la computadora que busque sus files para la información asociada.

(a) Los puertos de salida y de llegada podrían ser identificados digamos por un código de cuatro letras CHIM - Chimbote (12), CHIC - Chicama (11), SAMA (14) etc.

(b) Un file de embarcaciones puede guardarse en los discos del computador junto con las características de la embarcación, y la búsqueda en el file podría ser por nombre en vez de código si fuera necesario.

(c) Se puede preparar un sistema para codificar las zonas isoparalitorales automáticamente cuando se da una referencia de lugar y profundidad. Esto requerirá la construcción de tres files separados, a saber.

(i) File de posición

Esto consistirá en una matriz de tres o cuatro letras de código para cada nombre de lugar, por ejemplo, Antártica podría ser ANT o ANTA; El Barranco podría ser EBAR.

La lista de nombres debe ser ordenada de tal modo que las posiciones más frecuentemente pescadas esten al comienzo de la lista en cada índice. La lista también debe ser chequeada por los sinónimos, es decir, por ejemplo, la lista corriente contine ISLA LA VIUDA e ISLA VIUDA y las dos han sido codificadas como IVIU. Uno de estos nombres tendrá que ser cambiado.

(ii) File de profundidad

Asociada con cada conjunto de nombres de localidades en una serie de rangos de profundidad 0 - 25, 26 - 38, 39 - 60, 45 - 100. Los contornos de profundidad identificados son las zonas paralitorales más probables. El file de profundidad clasificado en el mismo modo que el file de puertos contendrá el límite superior del rango de profundidad en cada celda. Normalmente se requiere para describir la zona paralitoral intervalos de cuatro clases pero ocasionalmente también de cinco.

Posición Índice	1	2	3	4	5
1	25	38	60	100	
2	30	63	90	105	
3	20	46	76	90	
4	30	40	50	70	70
FILE DE	55	60	80	90	
PROFUNDIDAD	60	88			
"					
"					
"					
N	155	500			

(e) En este esquema el código de zona comienza en 10. Probablemente debería comenzar en la zona más al norte que haya sido pescada en todos los tiempos a fin de hacer el sistema lo más flexible posible.

5.4. Curva de crecimiento asimptótico

Este programa fue escrito durante mi consultoría principalmente para familiarizarme con la Hp pero también por un pedido de uno de los biólogos del Instituto. En el archivo de programas no hay ningún programa para ajustar este tipo de relación.

El método de ajustar la relación asintótica

$$y = a + b P^t$$

se describe en Stevens 1951 y proporciona un estimado de máximo probabilidad de los parámetros a , b y ρ . El modelo matemático expresado por esta forma puede ser relacionado a la curva de crecimiento de Von Bertalanffy

$$y = L_{\infty} (1 - e^{-k(t-t_0)})$$

y permite obtener los parámetros L_{∞} , k y t_0 . Este programa puede usarse con observaciones ponderadas o sin ponderar. Como los cálculos son bastante extensos y consisten en elevar a potencias fraccionales, es mejor codificar la variable t si, por ejemplo, es expresada en días por un período largo. (Esto es sólo para evitar la acumulación de errores propios del cálculo).

El método requiere una estimación inicial de ρ que cae entre 0 y 1. Si el primer estimado es muy alejado del verdadero valor de ρ puede necesitarse mucho tiempo para que la serie converja si es que eventualmente converge. Un estimado inicial puede proveerse automáticamente si a los datos se ajusta un polinomio de segundo orden no ponderado; el estimado se obtiene de la relación

$$\frac{Y_{t_{max}} - Y_{t_{mid}}}{Y_{t_{mid}} - Y_{t_{min}}} \quad 1/h$$

$$\text{donde } h = X_{t_{max}} - X_{t_{mid}}$$

Se entra entonces en un ciclo iterativo para refinar cada vez más el estimado inicial de ρ con un incremento $\Delta\rho$. El ciclo se detiene cuando $\Delta\rho < 0.001$ cuando 10 ciclos se han completado. El número de ciclos necesario para obtener el estimado final de ρ es un output como también lo son L_{∞} , k y t_0 más sus errores estándar. Normalmente el procedimiento convergirá con suficiente exactitud dentro de dos o tres ciclos.

En su forma presente este programa requiere que las entradas se hagan directamente del teclado. Normalmente los datos para este tipo de análisis no son muy numerosos y su entrada directa es apropiada. Todo lo que se necesita entrar es el número de observaciones seguidas de la variable independiente (normalmente el tiempo), la variable dependiente y un factor de ponderación. Este último puede ser 1, cuando el análisis ponderado no es necesario. La entrada es entonces.

N- N° de observaciones

1ra. observación; tiempo (t), longitud (y) factor de ponderación (w)

2da. observación, tiempo (t), longitud (y) factor de ponderación (w).
 nésima " " " "

Este programa puede ser fácilmente adaptado para usarse como cualquier computador que tenga un interpretador de BASIC.

5.5. Análisis de longitud/peso para merluza

La estimación del peso a partir de la longitud necesita ser particularmente exacta en el caso de la merluza pues este es el método que se usa para proyectar la distribución de longitudes de la muestra hasta el nivel de la embarcación muestreada. Las muestras no se pesan directamente y sus pesos son necesario estimarlos a partir de la relación longitud peso. Ya que el factor para elevar el valor de la muestra al valor de la embarcación puede ser bastante grande, es esencial que las relaciones longitud peso sean tan exactas como sea posible y que además reflejen los patrones estacionales si es que existen.

Grandes volúmenes de datos se han recolectado durante los muestreos biológicos los cuales contienen longitud total en cm y pesos en gm de modo que ha sido posible sumarizar los datos de dos años como un paso preparatorio para incluir las ecuaciones apropiadas en un programa de computadora. Se examinó las muestras para 1978 y 1979. Sin embargo, los datos para diciembre del 79 no estaban disponibles en el momento del análisis. Se examinaron separadamente los machos y las hembras y se escribió un programa para analizar estos datos.

El método de análisis es bastante similar al descrito en la sección 5.1. Las muestras analizadas fueron 46 con un total de 4,290 observaciones. El número de observaciones y los estimados originales de las pendientes fueron como sigue

Estimado de los coeficientes de regresión ajustada a los datos del 78 y 79

	AÑO	ENE.	FEB.	MAR.	ABR.	MAY.	JUN.	JUL.	AGO.	SET.	OCT.	NOV.	DIC.
Machos	1978	2.11	1.70	2.39	2.22	2.03	1.78	2.28	2.13	2.14	2.39	1.74	2.08
	1979	2.01	2.08	1.96	2.13	2.02	1.80	2.19	2.39	2.14	2.36	2.11	Sin mues tra disp.
Hembras	1978	2.03	2.06	2.51	2.22	2.48	2.24	2.28	2.94	2.70	3.17	3.02	3.20
	1979	3.05	2.85	2.87	3.20	3.08	2.26	3.03	3.10	3.03	2.95	3.01	Sin mues tra disp.

Número de observaciones de cada muestra

	AÑO	ENE.	FEB.	MAR.	ABR.	MAY.	JUN.	JUL.	AGO.	SET.	OCT.	NOV.	DIC.
Machos	1978	183	122	120	142	68	96	96	66	77	75	92	104
	1979	136	92	112	107	85	67	105	95	70	104	113	Sin mues tra disp.
Hembras	1978	115	105	84	85	82	79	54	109	74	125	108	73
	1979	64	108	113	69	65	34	121	55	81	98	62	Sin mues tra disp.

Una buena cantidad de peces fue medida para estos estudios. La colección de datos no fue hecha solamente para evaluar la relación longitud/peso sino para una amplia gama de medidas. Pero hay que hacer notar que cantidades tan grandes de datos requieren algún tiempo para ser procesados y que la misma precisión (o tal vez una mayor) puede obtenerse mediante un esquema de muestreo selectivo.

Por ejemplo, es sabido que para la mayor parte de especies la relación entre peso y longitud es de la forma $W = aL^b$, o sea, que es lineal si se convierte a logaritmos. La precisión de esta relación puede ser definida en términos de la precisión de los estimados de a y b . La variancia de b es $s^2 / \sum (l - \bar{l})^2$ donde s^2 es el cuadrado medio residual después de haberse ajustado la regresión. Esta variancia puede ser minimizada haciendo un máximo del denominador $\sum (l - \bar{l})^2$. La forma más fácil de maximizar $\sum (l - \bar{l})^2$ es escoger un conjunto de longitudes mínimas y máximas únicamente.

A modo de ilustración, consideremos una especie que normalmente se encuentra entre 20 y 40 cm. Si se tomasen 10 medidas por cada centímetro a través de todo el rango, se tendrían 210 medidas con $\sum (l - \bar{l})^2 = 1.7131$. Sin embargo si sólo se tomasen las medidas seleccionadas a 20 y 40 cm el mismo valor $\sum (l - \bar{l})^2$ podría obtenerse con sólo 38 peces a cada una de estas longitudes, es decir, un total de solo 76. Obviamente, esto tomará un tiempo más largo para hacer el muestreo inicialmente pero se reducirá el tiempo del análisis. El punto principal es que cuando una relación se la conoce a priori como lineal, el esquema de muestreo más eficiente es concentrarse en la medición de individuos pequeños y grandes en la población.

Se escribió un pequeño programa para sumarizar los datos de longitud peso en forma que estén disponibles de inmediato para referencia. Este es un programa bastante general que puede usarse para cualquier especie y a intervalos de longitud de 1 ó 0.5 cm. Se añade un ejemplo de output.

REFERENCIAS

- Abramson N.J. (1971). FAO Fish. Tech. Pap., (101)**
Computer programs for fish stock assessment.
- Armstrong D.W. FAO Project PER/76/022. Consultancy Report December 1979 (Unpublished).**
- Bazigos G.P. (1975). FAO Fish. Tech. Pap., (135)**
Applied fisheries statistics.
- Moller F. (1979). FAO Fish. Tech. Pap., (182)**
Manual of methods in aquatic environment research. Part 5.
Statistical tests.
- Snedecor G.W. and Cochran W.G. Statistical Methods. The Iowa State University Press. 1973, Iowa.**
- Stevens W.L. (1951). Asymptotic Regression Biometrics 7. 247-267.**

APENDICE 1. LISTA DEL EQUIPO DE COMPUTACIÓN PARA EL IMARPE
DEL PROYECTO ALEMAN

Hewlett Packard 1000 Mini-Computer System

(a) Hardware Configuration

- Memory size : 128 K bytes. This can be extended to 2 megabytes.
- Disk : 1 Fixed Head. Capacity 2.4 megabytes, 2 Replaceable Head. Capacity 2.4 megabytes.
(An additional 4.9 megabyte capacity disk system is to be built into the computer).
- System console : HP 2645 visual display unit with built-in mini-cassette reader. Each cassette has 110 K byte capacity.
- Terminal devices : 8 - data entry keyboards
2 - small HP2621A terminals
1 - HP9284A graphics visual display unit
- Punch card devices : 1 - 80 column punch card reader
1 - card punch
- Printer : 1 - HP2631G 180 character per second
136 character per line matrix printer
- Magnetic tape : 1 - 9 track, 800 b.p.i IBM/ANSI compatible magnetic tape system.

(b) Software

The system software includes a memory management system to handle several users simultaneously. It also has a file management system, an interactive editor, a multi-terminal monitor, a debugger and loader, a system library and a time base generator.

Included is an alphanumeric store which is extendable to 12K.

A FORTRAN IV compiler is provided along with a multi-terminal BASIC interpreter with extended language capability.

Graphics software for the 9284A graphics terminal is provided and there is a V24 plotter interface for a CALCOMP plotter with associated software.