



Universidad Austral de Chile

Facultad de Ciencias
Escuela de Ciencias

PROFESOR PATROCINANTE

Heriberto Figueroa Sánchez
Instituto de estadística
Facultad de Ciencias Económicas y
Administrativas

PROFESOR

CO-PATROCINANTE:

Carlos Jara Senn
Instituto de Zoología
Facultad de Ciencias

**BIOMETRÍA DEL CONFLICTO TAXONÓMICO ENTRE
AEGLA ABTAO SCHMITT Y *AEGLA RIOLIMAYANA* SCHMITT
(CRUSTACEA, DECAPODA, ANOMURA, AEGLIDAE)**

Tesis de grado presentada como parte
de los requisitos para optar al Grado
de Licenciado en Ciencias Biológicas

Oliver Rojas Bustamante

2005

Valdivia – Chile

ÍNDICE

1. RESUMEN	5
1.1 Summary	7
2. INTRODUCCIÓN	9
2.1 Hipótesis	12
2.2 Objetivos	13
3. MATERIAL Y MÉTODOS	14
3.1 Colección de los especímenes	14
3.2 Medición de especímenes y Tabulación de datos	17
3.2.1 Mediciones biométricas	17
3.2.2 Mediciones cualitativas	19
3.3 Metodología estadística	21
3.3.1 Análisis descriptivo	22
3.3.2 Evaluación de las Zonas de agrupación	23
Análisis de Componentes Principales.	23
Análisis de Varianza	25
3.3.3 Análisis Clasificadorio	26
3.3.4 Análisis de variables cualitativas	28
Comparación de las Zonas	28
Determinación de morfotipos.	30
4. RESULTADOS	32
4.1 Estadística descriptiva	32
4.1.1 Descripción gráfica de los datos	32
4.1.2 Estadígrafos Descriptivos	36
4.2 Evaluación de las Zonas de agrupación	38
4.2.1 Análisis de Componentes Principales	38
4.2.1 Análisis de Varianza	42
4.3 Análisis clasificadorio	47
4.3.1 Análisis discriminante de las Zonas de agrupación.	47
4.3.2 Análisis discriminante de los Bloques propuestos	51

	3
4.4 Análisis de datos cualitativos	54
4.4.1 Evaluación de las Zonas de agrupación.	54
4.4.2 Determinación de morfos	57
Determinación mediante puntos de corte (Cuantiles)	57
Determinación mediante probabilidades	59
5. DISCUSIÓN Y CONCLUSIONES	61
5.1 Descripción de las especies y síntesis de los resultados	61
5.2 Revisión histórica del problema.	62
5.3 Interpretación biológica de los resultados	63
5.4 Conclusiones y recomendaciones	64
6. REFERENCIAS	66

Dedicado:

A mi madre por su eterna paciencia y comprensión
a José Leiva e Isolina Quezada, por darme un hogar en Valdivia
a Cristian Salazar (Profesor informante) por su amistad, tiempo y confianza
y a Francisco Marín por su confianza y amistad.

1. RESUMEN

La descripción de las especies nominales *Aegla abtao* y *Aegla riolimayana* por Schmitt (1942^{a,b}) fue basada en un conjunto discreto de particularidades morfológicas y su relación con una distribución geográfica dividida por la Cordillera de los Andes. Posteriormente algunos investigadores (Ringuelet, 1960; Schmitt, 1942^b; Bahamonde & López, 1963; Jara, 1980) experimentaron dificultades en la identificación de ejemplares de *Aegla* provenientes de diversas localidades, destacando como problemática la continuidad morfológica observada. Por este motivo se propuso modificar el estatus taxonómico de tales entidades al rango de subespecies (Ringuelet, 1948). No obstante, el problema no se ha resuelto del todo y la incertidumbre taxonómica persiste.

En este estudio se evaluó, mediante procedimientos estadísticos, la relación entre estos dos morfotipos, para lo cual se analizaron 8 caracteres somatométricos y 12 caracteres cualitativos, los cuales fueron obtenidos de ejemplares de *Aegla* provenientes de cinco Zonas geográficas: Isla de Chiloé, Cuenca del Río Bueno, Río Calle-Calle y Río Cruces en Chile y de la Cuenca del Río Limay en Argentina. Los caracteres somatométricos fueron abordados mediante un Análisis de Componentes Principales como técnica de reducción de dimensionalidad y las distribuciones de sus valores en las Zonas evaluadas con un Análisis de Varianza. Posteriormente se intentó la construcción de una regla clasificatoria para los especímenes, aplicando un Análisis Discriminante. Para el caso de los caracteres morfológicos se construyó una variable de resumen para reflejar la situación de un espécimen en las 12 variables; a partir de esto se realizó un Análisis de Kruskal-Wallis y un Test de Dunn. Adicionalmente se planteó una metodología para la clasificación de organismos en las Zonas de agrupación. Los resultados muestran que los

especímenes de las Zonas más septentrionales (Argentina, Río Cruces y Calle-Calle) tienden a distribuirse de manera conjunta, tanto en términos de los caracteres somatométricos como de los caracteres cualitativos, encontrando en tales Zonas ejemplares que pueden ser identificados como morfotipo *riolimayana* y especímenes con características intermedias entre éste y el tipo *abtao*. Cabe destacar que no hay evidencia entre los morfos encontrados en estas Zonas y un patrón geográfico definido. La Cuenca de Río Bueno parece ser la Zona que más representa al morfotipo *Abtao* mostrando un buen ajuste en los análisis de discriminación. La Zona de Chiloé muestra un ajuste clasificatorio muy bajo para los parámetros del análisis, quedando relegada como grupo marginal y no representativo de la Zona muestral. Los caracteres morfológicos muestran claras tendencias de diferenciación morfológica y geográfica, sin embargo la cantidad de especímenes que presentan mezcla de caracteres posee una ponderación muy superior a los especímenes “puros”. A partir de estos resultados obtenidos, se puede concluir que los estudios anteriores carecen de las metodologías apropiadas para la resolución de los problemas planteados. De esta forma la explicación más plausible de la relación entre los morfotipos es que tengan un nivel de diferenciación fenotípica equivalente a subespecies, sin embargo, la existencia de una gran cantidad de especímenes con caracteres mezclados y la baja significancia de los patrones geográficos propuestos le restan poder a esta suposición. Al término de este estudio es posible concluir la imposibilidad de toma decisiones correctas, basándose en tan pocos aspectos de la biología de *aegla*, debido a la multidimensionalidad del problema; por este motivo solo resta recomendar extender estos resultados a las demás áreas de la biología de poblaciones.

1.1 Summary

The description by Schmitt (1942^{a,b}) of the nominal species *Aegla abtao* and *Aegla riolimayana* was based in a set of morphological particularities and its relationship with a disjoint geographical distribution across the Cordillera de los Andes. Later some scientists (Ringuelet, 1960; Schmitt, 1942^b; Bahamonde & López, 1963; Jara, 1980) had difficulties with the *aegla* identification from several places, observing and emphasizing in the apparent morphological continuity. Based on this problem Ringuelet, 1948 proposed to change the taxonomical status of this species to subspecies. Despite this, the problem has not been solved yet.

This research use statistical methods to evaluate the relationship between both morphotypes, thus it 8 somatometric characters and 12 qualitative characters from *aegla* specimens across five geographic zones were analyzed: Chiloé Island, Bueno river basin, Calle-Calle river basin, Cruces river basin and Limay river basin from Argentina. The somatometric characters were exanimated applying principal component analysis as a dimensional reduction technique, and their resulting factors distribution by means of an analysis of variance. Later the construction of a classification rule for the specimens was tried by means of a discriminant analysis. In the case of the qualitative characters, a summary variable was constructed to show the situation of the specimens in relation to the 12 variables, plus a Kruskal-Wallis test and a Dunn test. Additionally, a methodology for the classification of organisms in the grouping zones was considered.

The results show that the northern zones tend to the conjoint variation as much in term of the somatometrical and qualitative variables, finding units that may be identified as *riolimayana* type and intermediate specimens between that and the *abtao* type. However it's no evidence was found of a relation between the morphs found and their geographical distribution. The Bueno river basin is apparently the most representative zone for the *abtao* type and shows a good fit for statistical discrimination. However, the Chiloé Island shows a poor capacity of discrimination and is appears as a marginal zone, non representative from their geographical area. The qualitative characters analysis shows a morphological and geographic trend to the conjoint variation; however the amount of individuals with mixed characters is greater than the amount of “pure” specimens. Based on the obtained results it is possible to conclude that the previous investigations lacked the suitable methodology to solve the proposed problem. In this way, the most plausible explanation for the obtained results is that the morph types have a phenotypic level of differentiation equivalent to subspecies, situation that apparently is coherent. Nevertheless, the great amount of specimens with intermediate characteristics and the low significance of the proposed geographic patterns reduce the weight of this hypothesis. At the end, it is possible to conclude the impossibility to take correct decisions based in few aspects of the *aegla* biology, because the multidimensional nature of the problem. In addition, I must recommend to extend the investigation to other fields of population's biology, in order to argue more deeply the taxonomy of the Chilean and Argentinean *Aegla-abtao-riolimayana* populations.

2. INTRODUCCIÓN

En 1942 Waldo Schmitt anunció a la comunidad científica el descubrimiento de dos nuevas especies de cangrejos de agua dulce, *Aegla abtao* Schmitt, 1942^a y *Aegla riolimayana* Schmitt, 1942^b; Sin embargo señaló también la estrecha semejanza de la morfología externa existente entre los dos morfotipos descubiertos (Schmitt, 1942^b). Como parte del proceso de descripción, determinó como localidades tipo para estas especies a Zonas relativamente cercanas, separadas longitudinalmente por la Cordillera de los Andes, asignando para *Aegla abtao* en Chile, la localidad de Alerce (41°23'S, 72°55'W), en la Zona cercana al Lago Llanquihue y para *Aegla riolimayana* en Argentina, la localidad de Río Limay (41°05'S, 71°10'W), ubicada en la provincia de Río Negro. Según Schmitt (1942^b) las diferencias que se observan entre las dos especies, especialmente en la forma del caparazón, del rostrum, de las quelas y areola justifican el reconocimiento de estas dos especies diferentes.

Posteriormente Ringuelet (1948) señaló la dificultad experimentada en la determinación de ejemplares provenientes del área nororiente de la cuenca del Río Negro en el sur de Argentina como organismos asignables a *Aegla riolimayana* o *Aegla abtao*, tal dificultad se basó en la continuidad de los rasgos morfológicos exhibidos por los especímenes, de forma que algunos de los caracteres coincidieron con las características de *A. riolimayana* y otros con los de *A. abtao* generando una mezcla de caracteres en los organismos observados, basándose en esta observación señaló la posibilidad de reducir a *Aegla riolimayana* a una subespecie de *Aegla abtao* (Ringuelet, 1948), situación que se materializó (Ringuelet, 1960) posterior a la revisión de algunos ejemplares del Lago Lacar, en los cuales también se advertía una mezcla de los caracteres supuestamente propios de una u otra especie.

En investigaciones subsecuentes, Bahamonde & López (1963) usaron la propuesta de Ringuelet, adscribiendo ejemplares provenientes de la Isla de Chiloé, como *Aegla abtao riolimayana* situación que implica que la Zona en cuestión podría ser una extensión zoogeográfica del territorio argentino. Adicionalmente, Jara (1980) comparó ejemplares provenientes de la Cuenca del Río Valdivia con especímenes topotípicos de *A. abtao* y *A. riolimayana* determinándolos de manera vacilante como *A. riolimayana*, aumentando con esto la incertidumbre acerca de la distinción de los morfotipos propuestos por Schmitt, y de una asociación exclusiva existente entre *A. riolimayana* y el sistema lacustre argentino. La observación de la variabilidad en la morfología externa de ejemplares de *Aegla* ha sido también descrita en ejemplares adscritos a *Aegla abtao riolimayana* muestreados en la localidad Argentina del Lago Huechulafquén (Schuldt *et al*, 1988), haciendo hincapié en la amplia variabilidad exofenotípica de los ejemplares.

Las observaciones anteriormente descritas generan un escenario confuso, en el cual las especies propuestas por Schmitt se debilitan, en términos de la incertidumbre que han experimentado algunos investigadores en la toma de decisiones para adscribir especímenes de *Aegla* a los morfotipos *Abtao* o *Riolimayana* en algunas localidades en las que suelen encontrarse: Lago Lacar (Ringuelet, 1960), Neuquén (Schmitt, 1942^b; Ringuelet, 1960), Chiloé (Bahamonde & López, 1963) y Cuenca del Río Valdivia (Jara, 1980). El debilitamiento de las especies nominales *A. abtao* Schmitt y *A. riolimayana* Schmitt como morfotipos discriminables puede deberse a variadas causas, tales como que estas entidades biológicas sean en realidad subespecies o razas geográficas que traslapan sus márgenes de distribución integrando de manera secundaria sus características morfológicas externas o simplemente se trata de una sola especie que tiene una

amplia distribución geográfica y presenta una variabilidad exofenotípica en mosaico (Mayr, 1963; Mayr & Ashlock, 1991).

Si *A. abtao* y *A. riolimayana* fueran en realidad subespecies o razas geográficas, sería coherente que tras el escrutinio de una cantidad suficientemente grande de ejemplares que tengan la característica de abarcar el área de distribución de estos organismos se puedan discriminar o distinguir tres grupos de especímenes. Adicionalmente cabría esperar que cada uno de grupos de organismos tenga una localización geográfica determinada.

- A) Ejemplares que se ajusten a la diagnosis de *Aegla abtao*.
- B) Ejemplares que presenten una mezcla de caracteres.
- C) Ejemplares que se ajusten a la diagnosis de *Aegla riolimayana*.

De ser válida la posibilidad de que *A. abtao* y *A. riolimayana* fueran en realidad una sola especie con un amplio rango distributivo, podría esperarse que la variación de los caracteres morfológicos externos de los organismos no se relacionen con patrones geográficos y se observe una variación aleatoria de rasgos de uno y otro morfotipo en los especímenes de las distintas poblaciones que conforman su distribución geográfica.

El presente estudio pretende evaluar la hipótesis de que *Aegla abtao* Schmitt y *Aegla riolimayana* Schmitt son entidades taxonómicas bien definidas desde el punto de vista fenotípico y geográfico, para lo cual se propone la utilización de metodología estadística para la discriminación de los morfotipos a través de una serie de muestras provenientes de distintas poblaciones representativas de la distribución geográfica de ambas especies, intentando además que los estudios de Schmitt

(1942^{a,b}), Ringuelet (1948) y Jara (1980b) sean comparables con las inferencias a extraer de esta investigación, por esta razón se han considerado básicamente los mismos caracteres taxonómicos examinados por estos autores.

2.1 Hipótesis

Las hipótesis que serán puestas a prueba en este estudio derivan de las observaciones realizadas por algunos investigadores (Ringuelet, 1960; Schmitt, 1942^b; Bahamonde & López, 1963 y Jara, 1980), y apuntan al esclarecimiento de la incerteza taxonómica que evidencian las especies nominales *Aegla abtao* Schmitt y *Aegla riolimayana* Schmitt.

- *Aegla abtao* Schmitt y *Aegla riolimayana* Schmitt son entidades taxonómicas bien definidas desde el punto de vista fenotípico y geográfico, de manera que son entidades discriminables desde el punto de vista matemático estadístico.
- Existe un regla clasificatoria que permite a los especímenes muestreados ser adscritos a las especies nominales *Aegla abtao* Schmitt y *Aegla riolimayana* Schmitt, y que además evidencia un ajuste estadístico en términos de la probabilidad asociada al error del modelo matemático.

2.2 Objetivos

Objetivo general

Buscar y validar mediante modelos estadísticos las relaciones que subyacen a los patrones biométricos y morfológicos de las especies nominales *Aegla riolimayana* Schmitt y *Aegla abtao* Schmitt, a fin de encontrar antecedentes que ayuden definir la relación existente entre estas dos especies.

Objetivos específicos

- Evaluar y definir la relación matemático estadística existente entre las distintas cuencas hidrográficas y Zonas geográficas consideradas en el estudio.
- Ajustar un modelo probabilístico que permita evaluar las poblaciones componentes del rango distributivo de las especies en estudio, según su comportamiento numérico.
- Evaluar el ajuste y aplicabilidad de un sistema clasificatorio que agrupe a las distintas poblaciones que componen la muestra, además de evaluar la capacidad de las variables para discriminación de las especies en estudio.
- Escalar y evaluar descriptivamente la capacidad de los caracteres cualitativos disponibles que componen la muestra, en términos de la utilidad de las variables en la determinación de los morfotipos, distribución geográfica y determinación probabilística de los morfos que de estas variables se desprendan.
- Discusión del tratamiento taxonómico de las especies *Aegla riolimayana* y *Aegla abtao* a partir de las inferencias obtenidas en los análisis.

3. MATERIAL Y MÉTODOS

3.1 Colección de los especímenes

Para la realización del presente estudio se emplearon las mediciones corporales y caracteres morfológicos de ejemplares identificados indistintamente como *Aegla abtao* o *Aegla riolimayana* recolectadas entre 1977 y 1992, almacenadas en la colección IZUA (Instituto de Zoología de la Universidad Austral de Chile). El material fue colectado de diversas localidades, las cuales comprenden una amplia Zona geográfica, limitada latitudinalmente por la cuenca del Río Cruces (39°22'S aprox.) en el norte y la Isla Grande de Chiloé (42°40'S aprox.) por el sur; Los límites longitudinales del espectro muestral abarcan por el este a algunas localidades de Argentina, hasta las cercanías del mar Chileno por el oeste (rango aproximado entre 73°54'W y 71°05'W). Las distintas muestras fueron agrupadas como afluentes de una cuenca hidrográfica o como pertenecientes a una Zona geográfica delimitada de acuerdo al detalle de la Tabla III-1.

Tabla III-1: Zonas de agrupación de las muestras.

Zona	ID	N
Isla grande de Chiloé	1	26
Cuenca del Río Bueno	2	194
Cuenca del Río Calle - Calle	3	252
Cuenca del Río Cruces	4	165
Argentina	5	96

Cada Zona geográfica o cuenca hidrográfica está compuesta por muestras provenientes de distintas localidades, las cuales fueron nominadas según el río o lago del cual provienen y en ocasiones señalando alguna localidad cercana que las identifique (Tabla III-2), para luego ser asignadas a algunas de las Zonas de clasificación (Tabla III-1).

Tabla III-2: Muestras componentes de las Zonas de agrupación.

Localidad	Procedencia	Nº de colección	N	ID
Lago Notuco	42°40'S 73°54'W	ZUA C-77 C-85	22	1
Río Puntra	42°10'S 73°43'W	ZUA C-369	4	1
Lago Puyehue	40°40'S 72°36'W	ZUA C-411	25	2
Lago Rupanco	40°46'S 72°40'W	ZUA C-460	30	2
Río Lollehue	40°10'S 72°53'W	ZUA C-314	30	2
Río Maule (Tegualda)	40°00'S 73°25'W	ZUA C-305	29	2
Río Alerce	41°23'S 72°55'W	ZUA C-190 C-191	26	2
Lago Pirehueico	39°52'S 71°53'W	ZUA C-93 C-104	13	3
Río Fuy	39°51'S 71°55'W	ZUA C-33	25	3
Río Liquiñe	39°44'S 71°51'W	ZUA C-465	20	3
Río Neltume	39°48'S 72°59'W	ZUA C-466	16	3
Lago Calafquén	39°34'S 72°15'W	ZUA C-91	30	3
Río Coñaripe	39°35'S 71°59'W	ZUA C-29	24	3
Lago Riñihue	39°46'S 72°31'W	ZUA C-18 C-19	57	3
Río Quinchilca	39°51'S 72°45'W	ZUA C-347	21	3
Río Pichico	39°35'S 72°40'W	ZUA C-461	30	3
Río Calle-Calle	39°47'S 72°58'W	ZUA C-136	16	3
Río Chesque	39°22'S 72°37'W	IZUA C-159	6	4
Río Cruces (Ciruelos)	39°32'S 72°53'W	IZUA C-115	9	4
Río Cruces (Loncoche)	39°22'S 72°40'W	IZUA C-161 C-463	30	4
Río Cruces (Lanco)	39°27'S 72°50'W	IZUA C-459	30	4
Río Melefquén	39°35'S 72°29'W	IZUA C-462	30	4
Río Leufucade (Aylín)	39°31'S 72°39'W	IZUA C-371	30	4
Río Leufucade (Purulón)	39°29'S 72°44'W	IZUA C-458	30	4
Río Chimehuín	39°55'S 71°05'W	IZUA C-228	35	5
Río Curruhué	39°59'S 71°06'W	IZUA C-232	22	5
Lago Lacar	40°10'S 71°27'W	IZUA C-199	4	5
Río Limay	41°05'S 71°10'W	IZUA C-73	22	5

La captura de los especímenes fue realizada mediante dos procedimientos: Hasta medio metro de profundidad, la recolección consistió en la búsqueda manual de organismos bajo las rocas del litoral de los lagos y los lechos de los ríos. Entre medio y un metro de profundidad la recolección en los ríos fue asistida con una red de captura, arrastrada a contra corriente y en los lagos con ayuda de nasas cebadas con trozos de pescado dispuestas en el sublitoral entre 8 y 12 horas. Posteriormente los especímenes fueron sacrificados por inmersión en una solución 9:1 de Alcohol 70° y Formaldehído 7% y conservados a posterior en una solución de alcohol 70°.

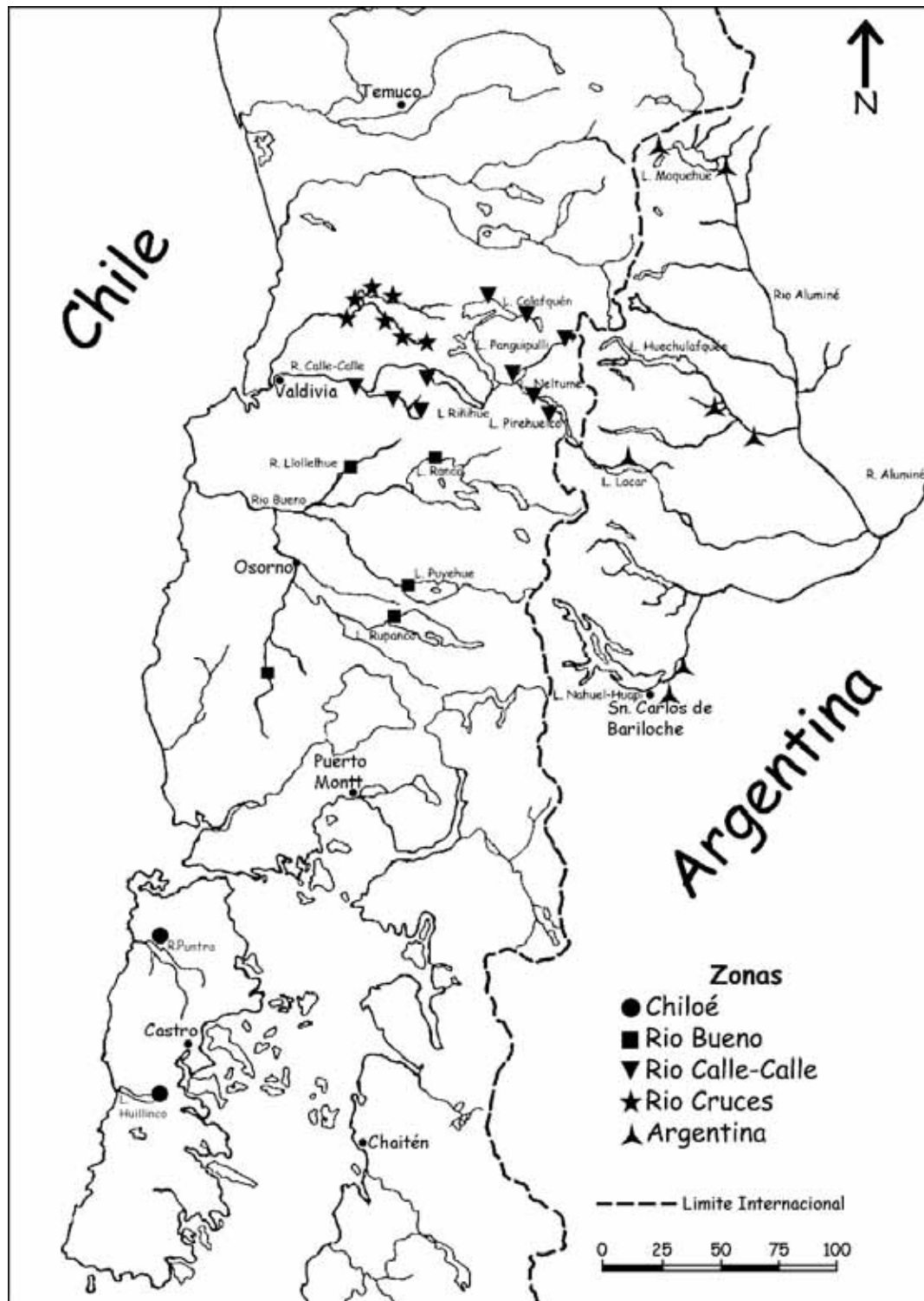


Figura III-1: Mapa de las localidades de muestreo y su asignación a las Zonas de agrupación.

3.2 Medición de especímenes y Tabulación de datos

Como punto de partida para las mediciones, se consideró que los especímenes colectados y seleccionados para el estudio procurasen la adultez, situación condicionada por la presencia del tubo genital completamente desarrollado en las coxas del 8° par de teracópodos y complementariamente por la presencia de heteroquelia.

3.2.1 Mediciones biométricas

Para la construcción de la base de datos, se consideraron seis medidas corporales (Figura III-2; Tabla III-3) previamente definidas por Ringuelet (1948 y 1949) y Jara (1980), las cuales se advierten en las descripciones de Schmitt (1942^b), y que fueron obtenidas mediante el uso de un calibrador digital, con una imprecisión asociada de 100 μm .

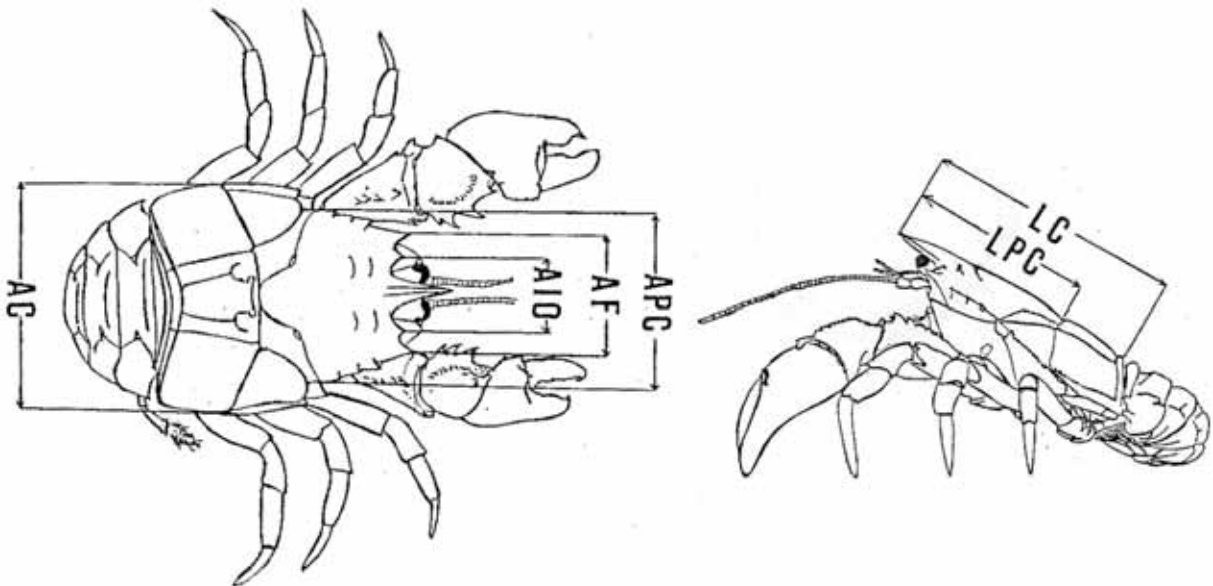


Figura III-2: Mediciones externas de aegla usadas en los análisis biométricos.

Adicionalmente a las somatometrías ya mencionadas se consideraron dos variables de tipo merístico o discreto (Figura III-3; Tabla III-3): El número de escamas cuticulares sobre la cara dorsoexterna del lóbulo carpal de las quelas y el número de dientes gástricos, los cuales fueron contabilizados disectando los estómagos longitudinalmente a través del eje ventromedial, presionando la Zona hasta expulsar desde el interior los osículos zigocárdicos, deshidratando los estómagos en Etanol absoluto, secándolos sobre papel absorbente y transparentándolos en Xilol durante 15 minutos, para posteriormente fijarlos con Rhenohistol (MerckTM) sobre un portaobjetos, lo que permitió su observación y contabilización mediante el uso de una lupa estereoscópica. El número de escamas carpales y de dientes gástricos fueron incluidas en los análisis, la primera porque al comparar especímenes de *A. abtao* y *A. riolimayana* se advirtió que en *A. abtao* las escamas parecen ser más numerosas y la segunda debido a la existencia de diferencias en el número de ellas en la comparación de especímenes de las especies en estudio (Jara 1981). Las estructuras del estómago cardinal fueron referidas a Paterson (1968).

Tabla III-3: Descripción de las mediciones biométricas registradas.

ID	Descripción
LC	Largo máximo del cefalotórax, desde el ápice del rostrum hasta el borde posterior del caparazón.
LPC	Largo precervical, entre el ápice del rostrum y el surco cervical.
APC	Ancho precervical, sobre los lóbulos hepáticos terceros.
AC	Ancho máximo del cefalotórax
AF	Ancho de la frente, entre los ápices de las espinas anterolaterales.
AIO	Ancho interorbitario, entre los ápices de las espinas orbitarias.
DIEN	Número total de dientes cigocárdicos del estomago
ESC	Número de escamas cuticulares en la cara dorso externa del lóbulo carpal de las quelas

3.2.2 Mediciones cualitativas

Adicionalmente se consideraron 12 caracteres cualitativos (Tabla III-4) presumiblemente discriminantes de los morfotipos *Abtao* y *Riolimayana* (Jara, 1980), los cuales fueron convertidos en una matriz binaria de datos en la que se utilizó la tipificación de valor uno (1) para los caracteres positivos del Tipo *Abtao* y de valor cero (0) para el Tipo *Riolimayana* para cada espécimen en las 12 variables. Cabe señalar que las observaciones correspondientes a este tipo de variables no se corresponden a los organismos de la matriz de datos con variables biométricas, es por esto que entre las dos bases de datos existen diferencias en la cantidad de observaciones y de localidades muestreadas (Tabla III-5).

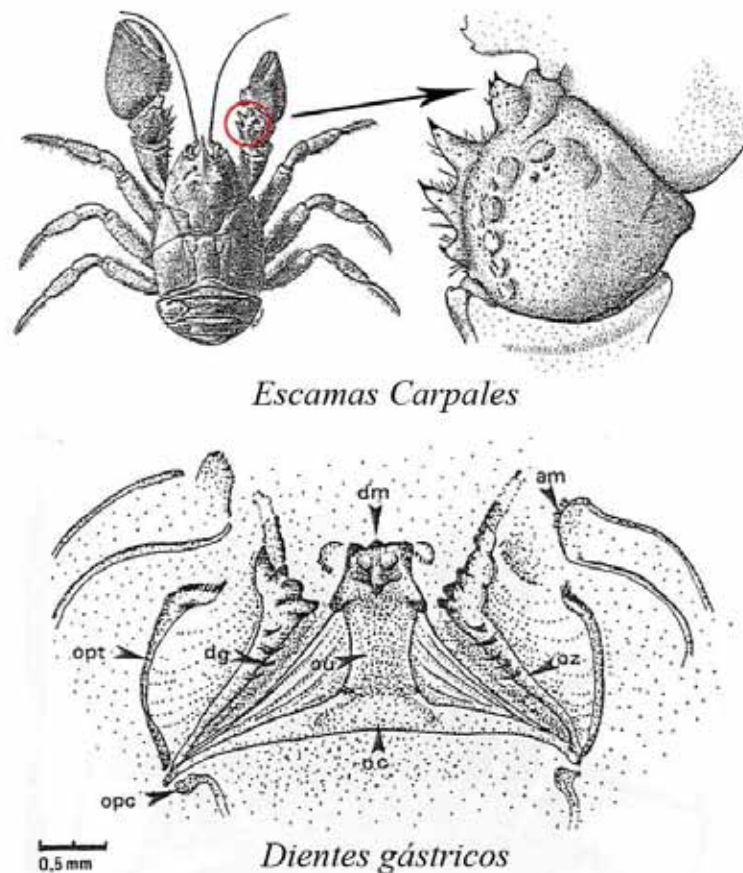


Figura III-3: Arriba: Escamas cuticulares sobre las caras dorso-externa del lóbulo carpal de la quela derecha. Abajo: vista interna del estomago cardial dg, dientes gástricos.

Tabla III-4: Variables cualitativas, ID y tipificación.

ID	Morfo ABTAO	Morfo RIOLIMAYANA
A	Rostro triangular, ancho, dorsalmente deprimido en el tercio distal. Moderadamente excavado a ambos lados de la carena.	Rostro triangular de base angosta, acuminado o estiliforme, notoriamente excavado a ambos lados de la carena.
B	Escamas de la carena rostral de igual tamaño.	Escamas de la carena rostral desiguales. Mayores las apicales.
C	Eminencias epi y protogástricas nodulares, prominentes.	Prominencias epi y protogástricas mal definidas.
D	Areola ancha, limitada por surcos laterales subparalelos.	Areola angosta delimitada por surcos laterales convergentes hacia atrás.
E	Lóbulo carpal como espina grande casi siempre acompañada de una menor, inserta sobre su borde posterior. Escamas espiniformes adicionales sobre el cuerpo del lóbulo.	Lóbulo carpal como una sola espina grande, sin espínulas ni escamas adicionales.
F	Borde dorsal del isquio de los quelípedos con tubérculos multiespinulados. Nódulo dorso medial distal notorio y escamoso.	Borde dorsal del isquio de los quelípedos con tubérculos de ápice único. nódulo dorso medial distal inconspicuo.
G	Margen ventral interno del isquio de los quelípedos con un cono agudo en el extremo distal y uno menor en el proximal. Entre ambos el borde es tuberculado y escamoso.	Margen ventral interno del isquio de los quelípedos con un tubérculo agudo distal y uno menor proximal. Entre ambos el borde es liso o levemente sinuoso.
H	Márgenes rostrales escamosos en toda su extensión.	Márgenes rostrales sin escamas cerca del ápice.
I	Frente angosta (apariciencia).	Frente ancha (apariciencia).
J	Area gástrica anterior ancha y protuberante que declina bruscamente hacia la frente.	Area gástrica anterior angosta y poco prominente que declina suavemente hacia la frente.
K	Dorso del caparazón fuertemente convexo. Areas demarcadas por surcos profundos.	Dorso del caparazón moderadamente convexo, aplanado. Surcos suturales poco profundos.
L	Superficie del caparazón muy punteada, pilosa y/o escamosa.	Superficie del caparazón lisa brillante o levemente punteada, glabra, escamas escasas o ausentes.
Tab.	1	0

La tabulación de los datos se realizó a partir de las planillas originales usadas en el registro de las mediciones corporales de los especímenes, utilizando para su registro digital la planilla electrónica Excel 2003.

Tabla III-5: Detalle de las localidades muestrales para las variables cualitativas, Zonas asignadas e ID correspondiente.

Muestra	N	Zona	ID
Lago Notuco	24	Chiloé	1
Lago Puyehue	15	Río Bueno	2
Lago Ranco	9		
Lago Rupanco	15		
Río Alerce	4		
Río Llollehue	16		
Río Maule	14		
Río Vald. (B.Sn. Javier)	11	Calle-Calle	3
Río Coñaripe	8		
Lago Calafquen	13		
Lago Pirehieico	5		
Lago Riñihue	12		
Río Liquiñe	9		
Río Neltume	17		
Río Pichico	15		
Río Chesque	6	Río Cruces	4
Río Cruces (Lanco)	14		
Río Cruces (Loncoche)	13		
Río Leufucade (Aylin)	11		
Río Leufucade (Purulon)	13		
Río Melefquen	7		
Río Chimehuin	13	Argentina	5
Río Curruhue	14		
Lago Lacar	5		
Río Limay	14		

3.3 Metodología estadística

La selección de una metodología estadística para el análisis y descripción de un conjunto de datos, debieran siempre contar con el respaldo y fundamentos de un diseño experimental, sin embargo, en ausencia de este, la elección de una metodología apropiada implicó para esta investigación, la realización de una gran variedad de pruebas estadísticas y análisis descriptivos a partir de muchas de las combinaciones posibles de las variables a disposición, a fin de cumplir con los objetivos propuestos y el respeto de las suposiciones y restricciones matemáticas involucradas, puesto que las inferencias realizadas deben suplir de manera satisfactoria la omisión de un diseño experimental. Este sistema de selección metodológica de ensayo y error es

denominado *Data Mining*, y para el presente estudio ha requerido de la utilización conjunta de los programas computacionales detallados a continuación.

- Office Excel 2003 Copyright © Microsoft Corporation. 2003
- SYSTAT 11.0 Copyright © Systat Software, Inc. 2004
- SigmaStat 3.11 Copyright © Systat Software, Inc. 2004

3.3.1 Análisis descriptivo

El análisis descriptivo de los datos fue realizado solamente en las variables numéricas, mediante la construcción gráfica realizada con el paquete estadístico SYSTAT 11, tal análisis cuenta con:

Diagramas de probabilidad. Este gráfico contrasta la distribución de los datos de prueba y la distribución normal esperada, esperando para ello un ajuste lineal perfecto. Las desviaciones de este patrón pueden ser: Un arco bajo la curva de existir sesgo derecho o sobre la curva para la asimetría izquierda, en cuando al aplanamiento una forma en Z acostada mostraría una platicurtosis y una S acostada una leptocurtosis.

Diagramas de caja. Representaciones que evalúan la distribución de algunos cuantiles de las variables continuas usadas. La caja posee el tamaño del rango intercuartílico, medida útil (aunque poco precisa) para comparar gráficamente la posición de los grupos. La comparación de los bigotes de las cajas es útil para evaluar la dispersión y los puntos más allá de 1,5 cajas (o más) son útiles en la detección de observaciones extremas que pudieran distorsionar resultados.

Gráficos de barras. Los gráficos de barras representan la distribución de las variables merísticas, intentando con ello evaluar la forma, dispersión y centro de tales variables, en relación a las Zonas de agrupación.

Estadígrafos descriptivos. (Spiegel, 2001^a) Aún cuando se espera que las conclusiones que puedan ser extraídas de las construcciones gráficas sean un reflejo de la realidad matemática subyacente, se dispondrá de los estadígrafos básicos de descripción: Media aritmética y mediana como estadígrafos de posición central; Rango, Desviación estándar y Varianza a como exponentes de la dispersión; Coeficiente de variación (C.V.) como medida comparativa de la dispersión de las Zonas; Y los coeficientes de Kurtosis y Simetría que miden el aplanamiento y sesgo de la distribución de datos, respectivamente.

3.3.2 Evaluación de las Zonas de agrupación

Dado que las muestras tomadas en las distintas localidades fueron identificadas como afluentes de una cuenca hidrográfica o como pertenecientes a un sector geográfico, generando de esta forma Zonas de agrupación (Tabla III-1) que conforman una variable de agregación constituida por cinco celdas, resulta coherente evaluar si los organismos que las componen son diferentes entre sí formando o no poblaciones independientes.

Análisis de Componentes Principales.

El primer paso en el contraste de las Zonas de agrupación creadas, se plantea como una reducción de dimensionalidad mediante un Análisis de Componentes Principales con el objeto de construir variables de resumen que permitan resolver las diferencias entre las Zonas de una manera más

sencilla, desde el punto de vista teórico y estadístico. Para la realización de este análisis se utilizó el paquete estadístico SYSTAT 11.

El Análisis de Componentes Principales (Dallas, 2000^a) es una técnica matemática que transforma un conjunto de variables correlacionadas en un conjunto menor de variables ortogonales denominadas Componentes Principales o Factores. El análisis explora una matriz de correlación o covarianza para determinar las raíces latentes o Eigenvalores, los cuales son una medida de ponderación de la varianza de cada uno de los Factores a extraer, de esta forma el primer Factor acumulará la mayor cantidad de la varianza de la muestra, el siguiente Factor la segunda magnitud de varianza, etc. Con estas medidas de carga se construyen los vectores ponderados de las variables o Eigenvectores que determinan las cargas correspondientes a cada variable dentro de un Factor, con el objeto de construir ecuaciones lineales de las variables originales y poder obtener las puntuaciones factoriales. Como cada Factor acumula una porción de la varianza muestral, determinada para cada componente en base a los Eigenvalores es de esperar que los primeros Factores a extraer contengan una gran cantidad de información, lo cual permitiría que con uno o dos Factores se pueda resumir la información de las seis variables utilizadas (Dallas, 2002^b).

Se valoró el grado de correlación de las variables mediante la matriz de dispersión (Zuwaylif, 1977^a), consistente en gráficos de dispersión para las combinaciones de variables, de manera que de observar una manifiesta colinealidad, estas se podrían considerar como buenas candidatas al Análisis de Componentes Principales. Para esta prueba sólo se consideraron las variables somatométricas externas, por considerarlas homogéneas en relación a su escala de medición, características distributivas, dispersión, etc. La extracción de los Componentes Principales se

realizó a partir de la matriz de Covarianzas, intentando con esto que la dispersión de los Factores a extraer mantengan las características de las variables utilizadas y faciliten la incorporación de nuevas observaciones al modelo (Stenson & Wilkinson, 2004), adicionalmente la extracción no consideró la rotación de los Factores. Posteriormente se almacenaron las puntuaciones de los dos primeros Factores para ser utilizados en un Análisis de Varianza.

Análisis de Varianza

Con las puntuaciones obtenidas del primer Factor del Análisis de Componentes Principales, se plantea la realización de un Análisis de Varianza, el cual, al considerar el uso de una variable independiente que contiene la información de seis variables, genera la necesidad de considerar un análisis de residuales, para dar mayor soporte a las inferencias que puedan obtenerse.

El Análisis de Varianza de un Factor (Weimer, 1999) es un procedimiento matemático ideal para la comparación de grupos en una variable; Extrae la variabilidad atribuible al azar como las desviaciones de los valores de cada observación, a la media del grupo correspondiente, obteniendo así la Varianza del error; Adicionalmente extrae la variabilidad atribuible a la existencia de grupos como las desviaciones que presentan las medias de cada grupo a la media de toda la muestra, generando así la Varianza de los tratamientos (o grupos). Con estas dos medidas de la variabilidad muestral es posible determinar que la diferencia que pueda existir entre ellas es, en términos de probabilísticos, causa del azar o de un real efecto de la existencia de grupos, lo que determinaría que estos pudieran ser (o no) poblaciones independientes. La determinación de cuales grupos son distintos de otros se determinó mediante comparaciones pareadas usando el Test de Scheffé (Scheffé, 1953), el cual permite, además de saldar el problema de contar con una

muestra no balanceada, elevar el nivel de significación para las combinaciones de pares por considerar el Test como conservador en comparación a otras pruebas disponibles (cita).

Para la evaluación de los supuestos del análisis se utilizaron de forma combinada, diagramas de probabilidad normal y diagramas de cajas, considerando que esta metodología tiene la suficiencia requerida dado el tamaño muestral (Hair, 1999). A partir de esto se realizó un Test de Varianza de una vía utilizando como variable independiente el primer Factor extraído del Análisis de Componentes Principales, las comparaciones de pareadas se realizaron mediante un Test de Scheffé. La normalidad de los residuales se evaluó con un diagrama de probabilidad y la homocedasticidad de estos mediante un gráfico de dispersión de medidas repetidas en las medias estimadas (Systat, 2004). La influencia de las observaciones en el cálculo de los coeficientes (Besley, 1980) fue evaluada mediante el diagrama de dispersión de las distancias de Cook y la influencia de las Zonas de agrupación en la varianza de los tratamientos mediante un gráfico de barras del estadístico Leverage calculado para cada celda.

3.3.3 Análisis Clasificadorio

El complemento del Test de Varianza es la construcción de una regla clasificatoria para los organismos asignados a las Zonas de agrupación o combinaciones de estas, mediante un Análisis Discriminante, el cual fue probado con la ayuda del paquete estadístico SYSTAT 11.

El Análisis Discriminante (Morrison, 1976) es una técnica que explora cuales de las variables introducidas o combinaciones de estas son las que mejor separan (o discriminan) a los grupos de prueba, a los cuales se han asignado de antemano las observaciones. Extrayendo las distancias de

Mahalanobis desde la matriz combinada de covarianzas, que cuantifican la cercanía (o lejanía) de una observación cualquiera al centroide de su grupo, a partir de esto se establecen los límites de tales nubes de puntos formadas por los grupos a evaluar, mediante una combinación lineal de variables independientes que permitan la asignación de las observaciones a tales grupos. De esta forma se logran tres objetivos: La determinación de la o las variables que mejor discriminan entre los grupos, la bondad del ajuste de las funciones clasificatorias y la posibilidad de incorporar nuevas observaciones al modelo (Engelman, 2004^b).

Para la construcción de las funciones de clasificación se usaron las ocho variables biométricas disponibles, es decir, las mediciones corporales del caparazón de los especímenes y las variables merísticas, por considerar que éstas, al margen de su distribución, pudieran aportar en la discriminación de los grupos. Teniendo en cuenta el tamaño de la muestra y la robustez del modelo, es de esperar que de las violaciones a los supuestos, como la desigualdad de las matrices de covarianzas entre las celdas y la distribución no normal de algunas variables no afecte considerablemente la clasificación (Hair, 1999). Para la realización del Análisis Discriminante, se usaron probabilidades previas iguales para cada grupo, la determinación de la importancia de las variables en la separación de los grupos con el estadístico F de remoción (Engelman, 2004^b), resultante de un Análisis de varianza de una vía para cada variable usando las restantes como covariables, y la validación cruzada de las observaciones usando las funciones discriminantes calculadas.

3.3.4 Análisis de variables cualitativas

El análisis de los caracteres cualitativos obedece a un intento por extraer la capacidad de las variables binarias para diferenciar entre las Zonas de agrupación y a la eventual determinación de la combinación de variables que mejor representan a los organismos pertenecientes a los morfos en estudio o a las agrupaciones que puedan inferirse.

Comparación de las Zonas

La evaluación de las Zonas de agrupación no obedece a una metodología estándar y se basa más bien en una secuencia deductiva que tiene como primer componente la determinación de cuales de las variables binarias disponibles muestran alguna utilidad para diferenciar a las Zonas en estudio, mediante la realización de un Test de X^2 (Zuwaylif, 1977^b), para el cual se calculó la frecuencia del morfo tipificado como uno (1) en la matriz binaria de acuerdo a:

$$f_{(x)} = \frac{\sum_{i=1}^n x_i}{n} \cdot 100 \quad \begin{array}{l} x = 1 \text{ (En matriz binaria)} \\ n = \text{Tamaño Zona} \end{array}$$

De manera que de observar una frecuencia menor a 50% implicaría un predominio del Tipo *Riolimayana* (tipificado como cero) y una frecuencia sobre el 50% una mayor cantidad de individuos del Tipo *Abtao*. Calculado este estadístico para cada variable y cada Zona, se probó cuales variables difieren (en alguno de sus cinco componentes) significativamente de la frecuencia promedio (valor esperado). Seleccionando así un subconjunto de las variables originales, que permitirán determinar la existencia de diferencias entre las Zonas. Este análisis fue realizado con el software Excel 2003.

El siguiente paso en la secuencia deductiva generada para la comparación de las Zonas fue que: Si en un espécimen cualquiera, se evalúa la presencia de varios caracteres, asignando un 1 (uno) a la manifestación de caracteres Tipo *Abtao* y un 0 (cero) a la presencia de caracteres de tipo *Riolimayana*, obteniendo con esto una matriz binaria que representa esta información, es posible concatenar estos caracteres y representar a un espécimen cualquiera mediante la combinación binaria (de la forma 011000101) que le corresponde, de acuerdo a la cantidad de variables disponibles. De esta forma, las combinaciones que presenten muchos 0 (cero) tendrían una mayor cercanía con el Tipo *Riolimayana*, u *Abtao* en el caso de presentar muchos 1 (uno), permitiendo que cada individuo pueda ser resumido como la suma de la cantidad de unos o ceros que presenta. En teoría esto permitiría evaluar la situación de ambos morfotipos, en relación a algún modelo de agrupación. Por ejemplo, de existir dos morfos predominantes (simpátricos o alopátricos) los conteos de estos tendrían forma de U al graficarlos, con muchos individuos del Tipo *Abtao*, pocos intermedios y muchos del Tipo *Riolimayana*; Lo contrario sería la disposición gráfica de los conteos en forma de campana, lo cual privilegiaría las combinaciones intermedias.

De acuerdo a lo anterior, si en cada espécimen se cuenta la cantidad de caracteres tipificados como 1 (uno) en las variables seleccionadas en el Test de X^2 se estará formando una nueva variable que resume a los registros de la matriz binaria y que permitirá la realización de un Test de Kruskal-Wallis (Spiegel, 2001) para comparar las Zonas y un Test de Dunn (Hollander & Wolfe, 1973) para la comparación pareada, usando para esto el paquete estadístico SigmaStat 3.11. La nueva variable de resumen podrá generar en un espécimen cualquiera, un valor muy bajo (o simplemente cero) que indicaría la prevalencia del Tipo *Riolimayana*, un valor igual o menor

al número de variables usadas indicando la prevalencia del Tipo *Abtao* o un valor intermedio entre cero y el número de variables indicando una mezcla de caracteres.

Determinación de morfotipos.

Una vez que se determinó la relación existente entre las Zonas de agrupación, es posible usar la variable de resumen de las observaciones binarias construida para el Análisis de rangos, como medida de clasificación de los organismos, usando un gráfico de cuantiles de las distribuciones porcentuales de la variable en los grupos de prueba (Weimer, 1999). De esta forma es posible la determinación gráfica del solapamiento de los grupos a evaluar, delimitando el 95% de las distribuciones de cada agrupación, para determinar con esto los valores que puede tomar la variable de resumen de observaciones binarias en los sectores de independencia o solapamiento. De existir sectores dentro de las distribuciones de los grupos que no muestren superposición con otros, los valores que la variable adquiriera en tales sectores determinarían la identidad de cada grupo.

Una alternativa para la clasificación de organismos en las agrupaciones que se observen en el Análisis de rangos, resulta de la determinación de las variables que puedan tener la capacidad de diferenciar a tales grupos, y determinar con esto los morfos representativos para cada grupo como combinaciones binarias de las variables seleccionadas. En base a esto se realizó un Test de X^2 (Zuwaylif, 1977^b) para seleccionar las variables que presenten desviaciones significativas del promedio esperado de frecuencias para el Tipo *Abtao* (o *Riolimayana*), para minimizar con esto la cantidad de variables y posibles combinaciones binarias. Posteriormente se evaluaron las variables seleccionadas concatenándolas para producir las combinaciones para cada observación

y se tabularon según las probabilidades de pertenencia a alguno de los grupos propuestos, esperando encontrar combinaciones exclusivas para cada uno de éstos.

4. RESULTADOS

4.1 Estadística descriptiva

4.1.1 Descripción gráfica de los datos

La construcción gráfica permite apreciar visualmente algunas de las características de los datos tales como su distribución, dispersión, colinealidad, normalidad, etc. Esto permite obtener una visión aproximada de la realidad matemática contenida en la muestra y formarse una idea de las posibles metodologías a usar. Es por esto que se han graficado algunos aspectos de las variables en sus Zonas de agrupación, derivadas de la posición geográfica de las localidades muestrales.

La alta variabilidad de los datos, unido a la elevada sensibilidad de las pruebas estadísticas para probar la normalidad en grandes muestras, lleva a considerar la presunción de normalidad de las variables a través de las celdas de agrupación, bajo una perspectiva descriptiva y no numérica, mediante diagramas de probabilidad (Gráfico IV-1), en los cuales puede advertirse que las variaciones de la distribución normal que se observan, tienden a manifestar un patrón a través de las Zonas de agrupación y no se aprecian grandes distorsiones. Se observa que la Zona de Chiloé, presenta un tamaño muestral muy inferior al de las demás Zonas (Tabla III-1), situación que genera la necesidad de observar con detención el comportamiento de este grupo en posteriores análisis, dada la posibilidad de que las observaciones que lo componen sean no representativas de esta Zona, en relación a su distribución se advierte un ligero sesgo izquierdo que se mantiene a través de las variables. Para la Zona de Río Bueno la desviación de la normal más evidente es la platicurtosis que presentan sus variables, exceptuando a AIO (Ancho Interocular) que muestra un leve sesgo derecho; En Zona de Río Calle-Calle la situación es muy similar al de la Zona

anterior, presentando platicurtosis en todas sus variables. La Zona de Río Cruces se ajusta mucho a la distribución normal esperada y bajo los términos de esta inspección gráfica podría considerarse con normalidad en todas sus variables. Por último la Zona Argentina presenta una mezcla muy leve de platicurtosis y sesgo izquierdo. En lo sucesivo se considerará que la desviación de la normalidad que presentan las variables no es suficiente para considerarlas atípicas y que el uso de transformaciones no se justifica.

La evaluación gráfica del comportamiento de las Zonas muestrales dentro de cada variable, descritas en términos de las diferencias aparentes entre los rangos intercuartílicos dispuestos en los diagramas de cajas (Gráfico IV-2) ayudarán en una primera aproximación (descriptiva) de las relaciones entre las celdas de agrupación. Puede observarse que las distintas Zonas se comportan de manera muy similar en las distintas variables, lo que hace pensar en la posibilidad de que estas presenten una variación conjunta. El patrón que describen las Zonas en las variables muestran a Chiloé con especímenes de un tamaño inferior en relación a las otras Zonas, la Cuenca del Río Bueno muestra un comportamiento opuesto al presentar a los organismos con la mayor talla de la población en estudio, aunque también con una gran dispersión, sin embargo, lo más destacable es la disposición en bloque de las Zonas más septentrionales (Río Calle-Calle, Río Cruces y Argentina) con medianas y dispersiones similares. Otro aspecto de las cajas es la presencia de algunas observaciones atípicas, sin embargo, al no ser extremas no serán, consideradas como outliers.

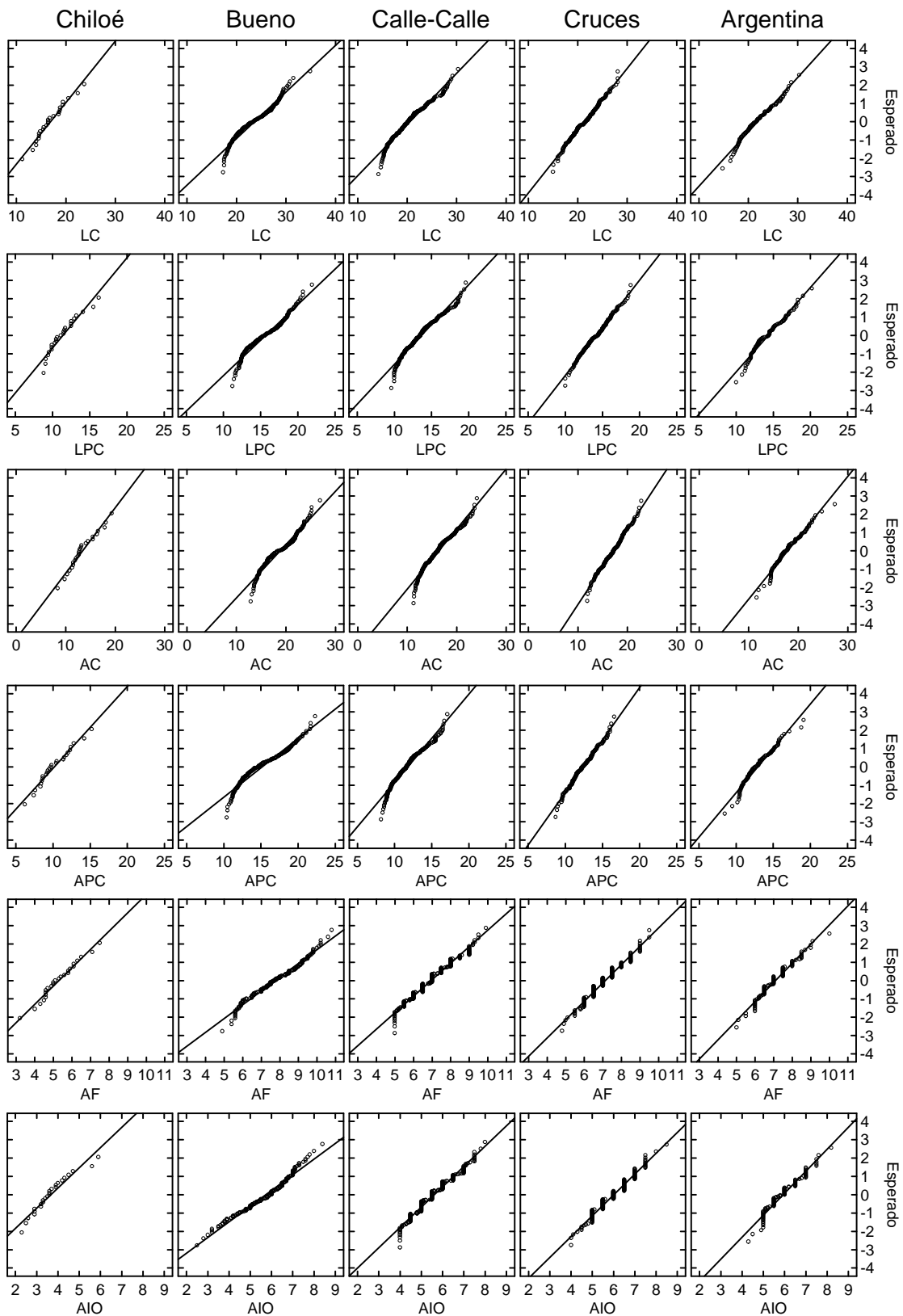


Gráfico IV-1: Diagramas de probabilidad de las variables y las Zonas de agrupación.

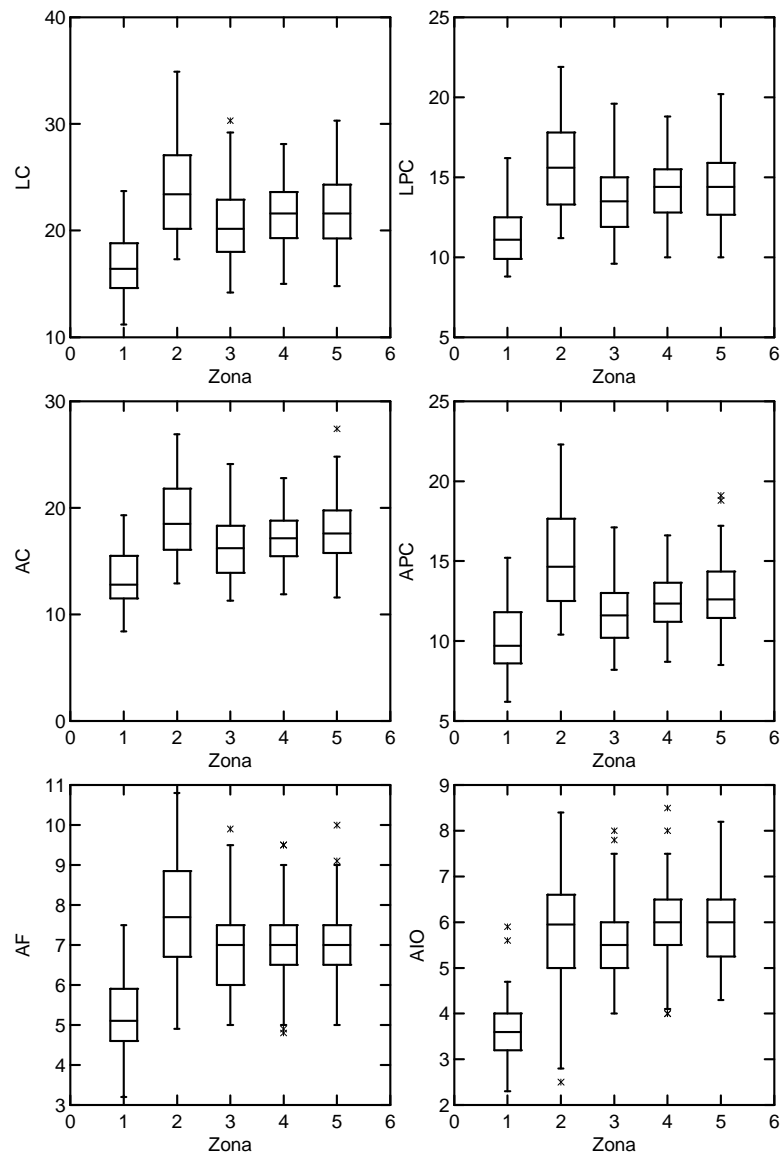


Gráfico IV-2: Diagramas de caja de las variables, divididas por Zona. Chiloé:1 Río Bueno:2 Calle-Calle:3 Río Cruces:4 Argentina:5.

Las mediciones de los especímenes consideran variables de tipo discreto o métrico (Gráfico IV-3). Tal condición no las hace candidatas en primera instancia a las pruebas paramétricas, sin embargo, la posible capacidad para diferenciar entre las distintas Zonas crea la necesidad de considerarlas para tales fines.

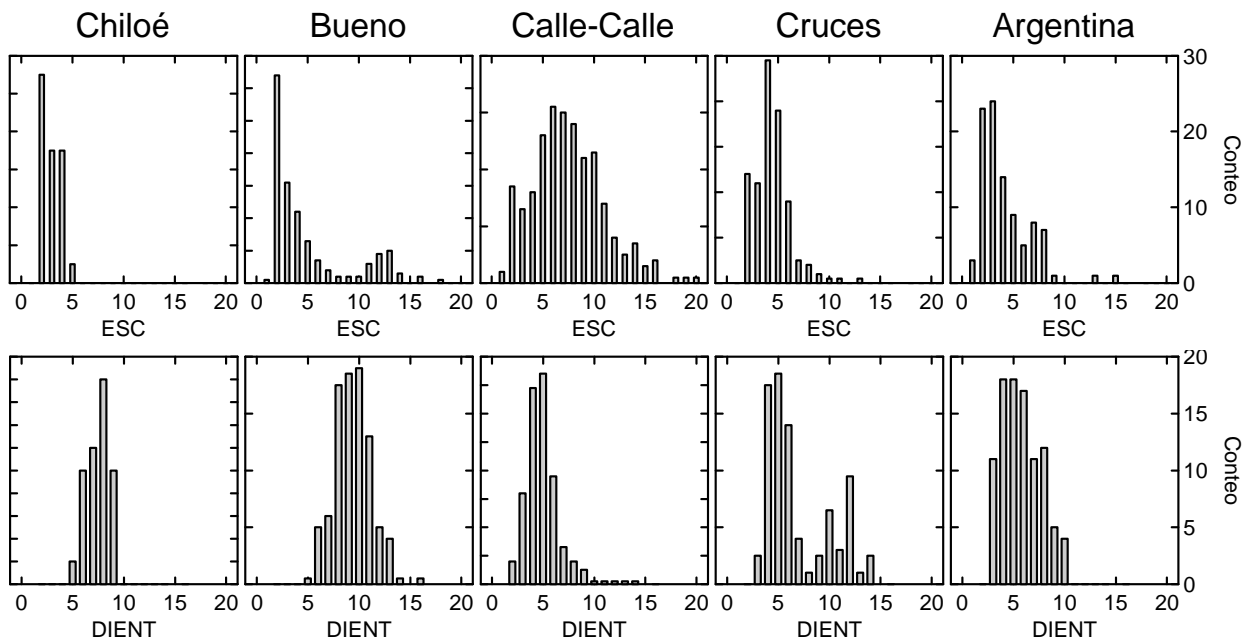


Gráfico IV-3: Gráfico de barras de las variables discretas, divididas por Zonas. ESC: Número de escamas carpales y DIENT: Número de dientes gástricos.

Desde el punto de vista descriptivo llama la atención la heterogeneidad que estas variables presentan, la aparente presencia de grupos dentro de las Zonas de análisis y la escasa interpretación que de ellas se desprende.

4.1.2 Estadígrafos Descriptivos

De acuerdo con el análisis Gráfico de los datos, la Tabla IV-1 muestra de forma numérica las observaciones realizadas anteriormente, las cuales concuerdan y reafirman la idea que los datos, a pesar de presentar variaciones no deseables en algunos aspectos, presentan también grandes oportunidades para el análisis.

Tabla IV-1: Estadísticos descriptivos de las variables y las Zonas en análisis.

	Zona	LC	LPC	AC	APC	AF	AIO	ESC	DIENT
N de casos	Chiloé	26	26	26	26	26	26	26	26
	Río Bueno	179	179	179	179	179	179	179	179
	Calle-Calle	252	252	252	252	252	252	252	252
	Río Cruces	165	165	165	165	165	165	165	165
	Argentina	96	96	96	96	96	96	96	96
Media	Chiloé	17,246	11,602	13,723	10,450	5,415	3,769	2,923	7,462
	Río Bueno	23,480	15,583	18,699	15,055	7,722	5,697	5,000	9,397
	Calle-Calle	20,404	13,532	16,270	11,705	6,903	5,569	7,524	4,944
	Río Cruces	21,507	14,245	17,178	12,473	7,136	6,019	4,406	6,939
	Argentina	21,999	14,418	18,033	13,023	7,145	5,977	4,177	5,781
Mediana	Chiloé	16,40	11,30	12,85	9,75	5,20	3,60	3	8
	Río Bueno	23,30	15,50	18,20	14,60	7,70	5,90	3	9
	Calle-Calle	20,10	13,45	16,15	11,60	7,00	5,50	7	5
	Río Cruces	21,60	14,40	17,20	12,40	7,00	6,00	4	6
	Argentina	21,65	14,40	17,70	12,70	7,00	6,00	3	6
Rango	Chiloé	15,70	8,60	13,30	10,80	5,10	4,20	3	4
	Río Bueno	19,70	11,30	15,30	12,70	6,10	5,90	17	11
	Calle-Calle	16,70	10,60	13,70	8,90	5,00	4,10	19	12
	Río Cruces	15,20	10,20	12,70	8,90	4,70	4,50	11	11
	Argentina	21,50	13,60	20,40	14,20	7,00	5,70	14	7
Desviación Estándar	Chiloé	3,467	2,252	3,127	2,530	1,129	1,022	0,935	1,140
	Río Bueno	3,942	2,584	3,440	3,081	1,351	1,170	3,975	1,828
	Calle-Calle	3,606	2,329	3,051	2,074	1,119	0,912	3,534	1,746
	Río Cruces	2,987	1,962	2,477	1,775	0,985	0,870	1,791	3,121
	Argentina	3,618	2,326	3,253	2,244	1,061	0,909	2,471	1,926
Varianza	Chiloé	12,021	5,074	9,779	6,403	1,274	1,044	0,874	1,298
	Río Bueno	15,538	6,678	11,831	9,491	1,826	1,369	15,798	3,342
	Calle-Calle	13,004	5,424	9,308	4,300	1,252	0,832	12,489	3,049
	Río Cruces	8,923	3,850	6,133	3,151	0,971	0,756	3,206	9,740
	Argentina	13,087	5,411	10,581	5,035	1,126	0,826	6,105	3,710
C.V.	Chiloé	0,201	0,194	0,228	0,242	0,208	0,271	0,320	0,153
	Río Bueno	0,168	0,166	0,184	0,205	0,175	0,205	0,795	0,195
	Calle-Calle	0,177	0,172	0,188	0,177	0,162	0,164	0,470	0,353
	Río Cruces	0,139	0,138	0,144	0,142	0,138	0,144	0,406	0,450
	Argentina	0,164	0,161	0,180	0,172	0,149	0,152	0,592	0,333
Coeficiente de Kurtosis	Chiloé	1,290	0,707	0,417	0,549	0,715	1,262	-0,998	-0,725
	Río Bueno	-0,898	-1,047	-1,064	-0,906	-0,832	-0,352	0,577	0,367
	Calle-Calle	-0,339	-0,301	-0,442	-0,355	-0,410	-0,478	0,502	5,241
	Río Cruces	-0,338	-0,219	-0,230	-0,325	-0,386	-0,431	3,788	-0,735
	Argentina	1,364	1,415	2,814	2,917	3,740	2,828	3,898	-0,681
Coeficiente de sesgo	Chiloé	0,971	1,063	0,792	0,806	0,662	1,145	0,480	-0,337
	Río Bueno	0,169	0,126	0,091	0,409	-0,010	-0,442	1,353	0,290
	Calle-Calle	0,445	0,454	0,452	0,528	0,257	0,295	0,627	1,652
	Río Cruces	0,165	0,203	0,106	0,251	0,169	0,054	1,339	0,812
	Argentina	0,824	0,885	1,132	1,213	1,204	1,145	1,639	0,408

4.2 Evaluación de las Zonas de agrupación

Como se señaló en el Capítulo 3 (Material y métodos) se cuenta con ocho variables biométricas (Tabla IV-2) de las cuales, dos son de tipo merístico, y serán abordadas con la precaución de no afectar la potencia de los análisis, es por esto que al proponer una reducción de dimensionalidad como punto de partida en la realización de un análisis paramétrico para la comparación de las Zonas, estas no serán consideradas por estar medidas en otra escala y porque se espera que las variables a usar sean similares en sus características y componentes.

Tabla IV-2: Variables muestreadas.

ID	Descripción
LC	Largo máximo del cefalotórax, desde el ápice del rostrum hasta el borde posterior del caparazón.
LPC	Largo precervical, entre el ápice del rostrum y el surco cervical.
APC	Ancho precervical, sobre los lóbulos hepáticos terceros.
AC	Ancho máximo del cefalotórax
AF	Ancho de la frente, entre los ápices de las espinas anterolaterales.
AIO	Ancho interorbitario, entre los ápices de las espinas orbitarias.
DIEN	Número total de dientes cigocárdicos del estomago
ESC	Número de escamas cuticulares en la cara dorso externa del lóbulo carpal de las quelas

4.2.1 Análisis de Componentes Principales

El primer paso en la reducción de dimensionalidad es la evaluación del grado de asociación lineal de las variables, ya que el análisis seleccionado para tal fin produce combinaciones lineales de las variables usadas y cuyos coeficientes dependen del grado de asociación que estas presenten, esta condición se evaluó mediante la matriz de dispersión (Gráfico IV-4), el cual muestra una gran tendencia a la variación conjunta de las variables, escenario conocido como colinealidad; Cabe destacar que la colinealidad disminuye a medida que las variables cuantifican estructuras más pequeñas, lo cual implica que estas poseen una menor relación lineal con las grandes medidas corporales y las hace más independientes, a pesar de esto, las características observadas en la

matriz de dispersión convierte a las variables seleccionadas en candidatas ideales para la realización de un Análisis de Componentes Principales.

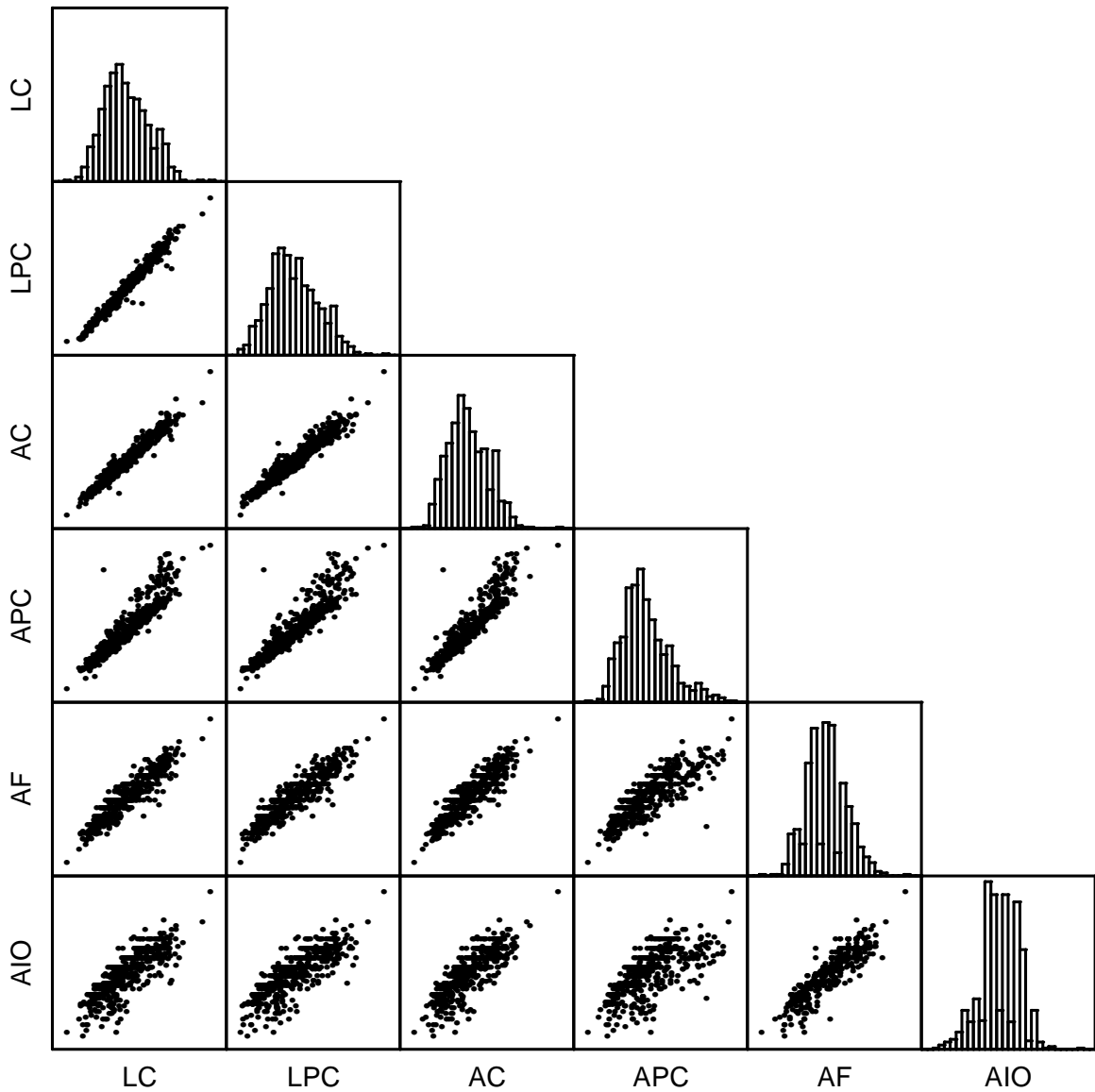


Gráfico IV-4: Matriz de dispersión de las variables continuas.

Como las variables seleccionadas están medidas en la misma escala, se considera que lo más apropiado para la estimación de los Factores es la utilización de la matriz de Covarianzas, considerando los resultados sin rotar. Como se observa, la varianza de los Factores extraídos

(Tabla IV-3) cumple toda expectativa y nos permite la utilización del primer componente como medida de resumen de las seis variables seleccionadas ya que acumula más del 95% de la variabilidad de la muestra, situación que puede observarse descriptivamente en el Gráfico de sedimentación (Gráfico IV-5), donde se representan los eigenvalores (análogos a la varianza) de cada uno de los Factores.

Tabla IV-3: Eigenvalores y porcentajes de la varianza de los dos primeros componentes.

Factor	Eigenvalor	Varianza
1	39,872	95,66%
2	0,937	2,25%

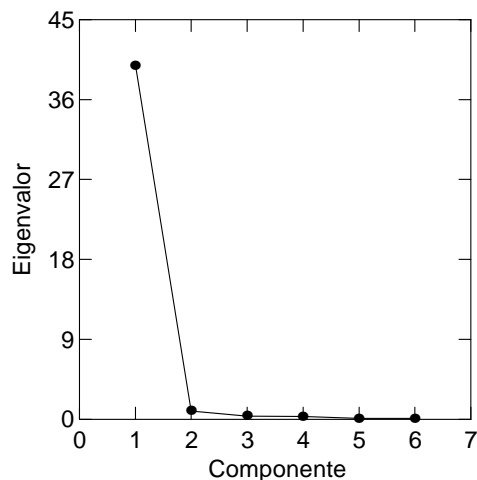


Gráfico IV-5: Gráfico de sedimentación de los componentes principales.

Otro de los aspectos relevantes del análisis tiene que ver con la ponderación de cada variable en la construcción de los Factores (Gráfico IV-6). Al observar las ponderaciones de las variables resulta evidente la relación que existe con la magnitud de las estructuras que cuantifican, siendo el largo del caparazón la variable que más carga acumula, el ancho del caparazón la segunda, etc. (Tabla IV-4). Puede observarse también la relación de ponderación de las variables entre los dos Factores extraídos.

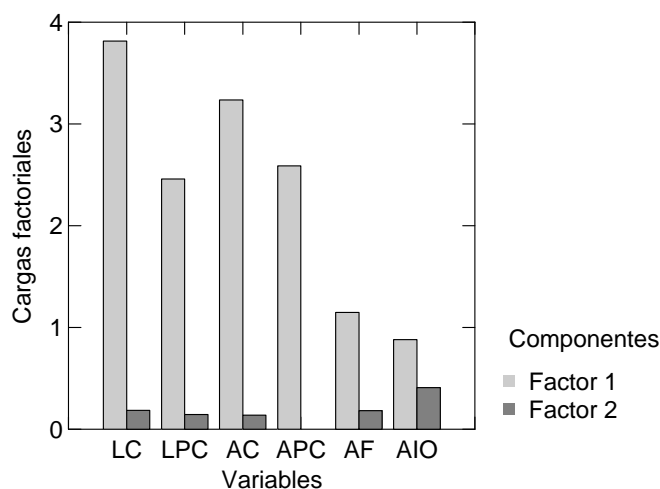


Gráfico IV-6: Gráfico de las cargas factoriales de los dos primeros componentes extraídos.

Una de las características más útiles del Análisis de Componentes Principales es la posibilidad de incorporar nuevas observaciones al modelo, a través de los coeficientes de la combinación lineal construida para cada Factor (Tabla IV-4), es decir, sumando el resultado de multiplicar el valor que toma cada variable para un espécimen dado (nueva observación) por el coeficiente correspondiente, obteniendo así el valor que le corresponde a esta nueva observación en alguno de los Factores extraídos; De esta forma la utilidad de resumir seis biometrías en una resulta evidente, tanto en los análisis que a continuación se disponen, como en futuros experimentos.

Tabla IV-4: Coeficientes, cargas factoriales (ponderaciones) y eigenvectores de los primeros dos componentes principales.

Variable	Coeficientes Factor 1	Cargas Factor 1	Eigenvector Factor 1	Coeficientes Factor 2	Cargas Factor 2	Eigenvector Factor 2
LC	0,096	3,814	0,604	0,200	0,187	0,194
AC	0,081	3,237	0,513	0,150	0,141	0,145
APC	0,065	2,588	0,410	-0,867	-0,812	-0,839
LPC	0,062	2,460	0,390	0,156	0,146	0,151
AF	0,029	1,148	0,182	0,195	0,183	0,189
AIO	0,022	0,883	0,140	0,437	0,409	0,423

4.2.1 Análisis de Varianza

Los resultados del análisis de componentes principales nos permiten la evaluación de las Zonas de agrupación de una manera sencilla y elegante, mediante un Análisis de Varianza de un Factor, calculado con el primer Componente (Factor) extraído del análisis anterior. El primer paso en la realización del análisis es evaluar gráficamente el comportamiento que tiene este Factor a través de las Zonas de agrupación (Tabla IV-5).

Tabla IV-5: Tipificación de las Zonas de agrupación según ID.

Zona	ID	N
Isla grande de Chiloé	1	26
Cuenca del Río Bueno	2	194
Cuenca del Río Calle - Calle	3	252
Cuenca del Río Cruces	4	165
Argentina	5	96

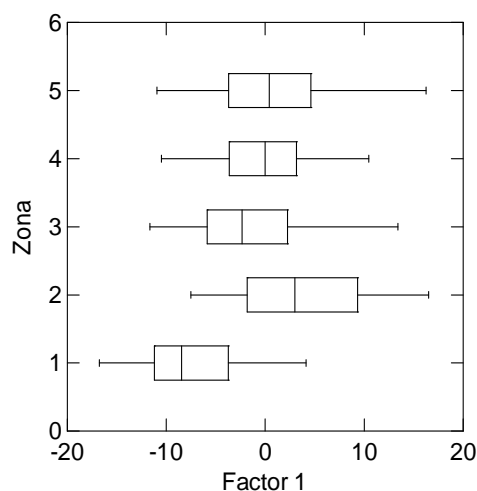


Gráfico IV-7: Diagrama de cajas para las Zonas de agrupación. Chiloé:1 Río Bueno:2 Calle-Calle:3 Río Cruces:4 Argentina:5.

La evaluación de las distribuciones muestra un gran solapamiento de las Zonas en estudio (Gráfico IV-7), dispersiones similares y la ausencia de observaciones extremas que bordeen a las cajas. El gráfico de probabilidad indica que la normalidad de las celdas (Gráfico IV-8) muestra en

general un buen ajuste de las observaciones a la curva normal esperada, y a pesar de la platicurtosis que evidencia la Cuenca de Río Bueno no se considera necesaria la aplicación de transformación alguna.

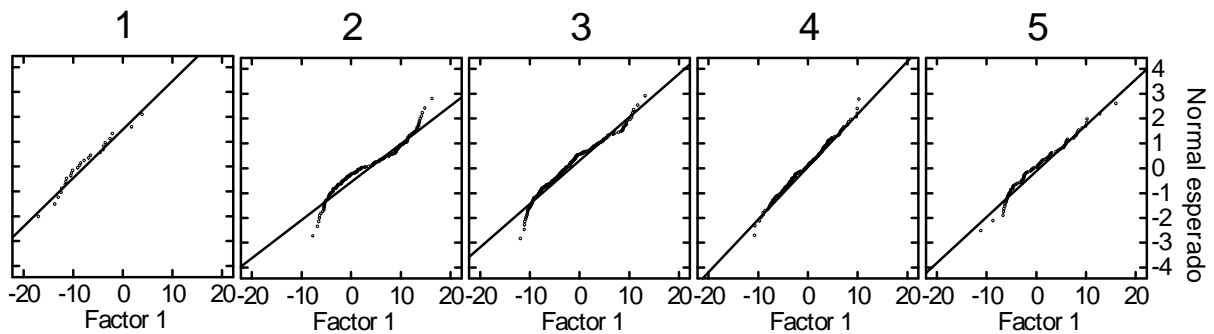


Gráfico IV-8: Gráfico de probabilidad para las celdas del factor. Chiloé:1 Río Bueno:2 Calle-Calle:3 Río Cruces:4 Argentina:5.

La Tabla ANOVA (Tabla IV-6) muestra que al menos una de las Zonas es significativamente distinta de las demás. Se observa que la varianza de los tratamientos es muy grande en relación al error, razón por la cual el estadístico F posee un valor tan grande, el cual puede ser consecuencia de una sobre estimación de la varianza de los tratamientos producida por el menor tamaño que la Zona de Chiloé evidencia en todos los gráficos.

Tabla IV-6: Tabla resumen Análisis de Varianza.

Variable dependiente: Factor 1			N: 706	R: 0,417	R2: 0,174
Fuente	Suma cuadrados	Gl.	Varianza	F	P-Valor
Zona	4616,633	4	1154,158	22,059	0,000
Error	36676,320	701	52,320		

Según los contrastes pareados evaluados mediante el Test de Scheffé (Tabla IV-7) se observa un bloque homogéneo en términos de sus medias que agrupa a las Zonas de Argentina, Calle-Calle y Río Cruces, las cuales puede ser consideradas como una sola población; La situación opuesta enfrenta a las Zonas de Río Bueno y Chiloé, esta última presentando la media más pequeña y

puede considerarse como una población marginal, muy separada de los demás grupos. No obstante, esta Zona muestral debe ser analizada con cautela, debido a la reducida cantidad de observaciones y ante la posibilidad de que la única localidad que le da forma sea no representativa de la Isla de Chiloé.

Tabla IV-7: Matriz de comparaciones múltiples para el Test de Scheffé.

Varianza del error: 52.320 con 701 grados de libertad

Matriz de las diferencias entre pares de medias					
	Chiloé	Río Bueno	Calle-Calle	Río Cruces	Argentina
Chiloé	0				
Río Bueno	11,290	0			
Calle-Calle	5,872	-5,419	0		
Río Cruces	7,377	-3,914	1,505	0	
Argentina	8,225	-3,066	2,353	0,848	0

Matriz de probabilidad de las comparaciones pareadas					
	Chiloé	Río Bueno	Calle-Calle	Río Cruces	Argentina
Chiloé	1				
Río Bueno	0,000	1			
Calle-Calle	0,004	0,000	1		
Río Cruces	0,000	0,000	0,352	1	
Argentina	0,000	0,022	0,111	0,930	1

A fin de ser responsable con el uso de una sola variable o factor que contiene la información de seis variables, se realizará un análisis de los residuales obtenidos del ANOVA para contrastar algunos supuestos que guardan relación con el ajuste de los datos al modelo lineal aplicado. La evaluación de la normalidad de los residuales (Gráfico IV-9), presenta bajo la construcción gráfica, un buen ajuste a la curva normal esperada. Sin embargo, puede observarse la característica forma de Z horizontal, que advierte de un pequeño aplanamiento de la función de densidad o platicurtosis. No obstante, lo que más interesa para el caso de los residuales se relaciona con la simetría de la distribución, elemento de correcto comportamiento en este caso.

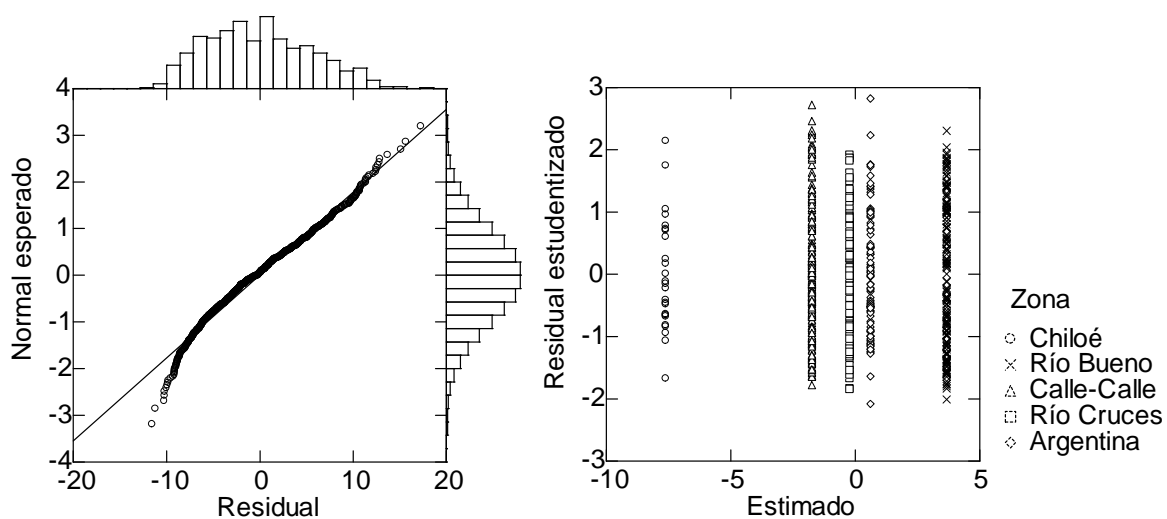


Gráfico IV-9: Diagrama de probabilidad normal para los residuales y residuales estudentizados versus las medias estimadas.

Otro de los aspectos ya evaluados y que tiene su contraparte en los supuestos de los residuales es la homocedasticidad de los grupos (Gráfico IV-9), situación a evaluar mediante los residuales estudentizados, que representan la normalización de estos bajo la distribución t de Student (con $N - p - 1$ grados de libertad, siendo N : tamaño muestral, p : Número de predictores) y que además son útiles en la detección de outliers u observaciones extremas, las cuales de existir presentarían valores superiores a tres. Como puede apreciarse, no se observan casos con estas características y se advierte la similitud de los rangos de cada Zona, las cuales están ordenadas según las medias estimadas por el modelo. Resulta fácil distinguir la situación señalada en el Test de Scheffé, donde la Zona de Chiloé es la más pequeña en la talla de los organismos, seguida por el bloque de las Zonas más septentrionales y por último la Zona de Río Bueno que presenta la mayor talla.

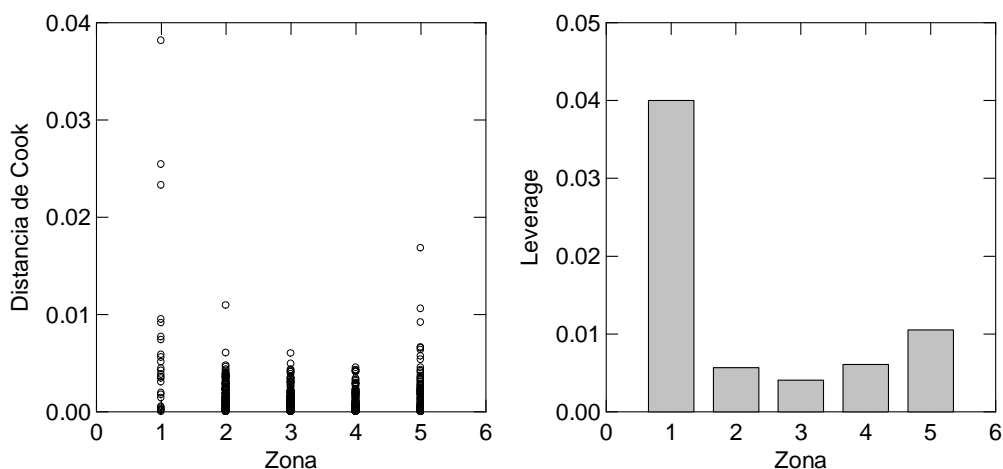


Gráfico IV-10: Medidas de ponderación del ANOVA; Distancia de Cook y Leverage. Chiloé:1 Río Bueno:2 Calle-Calle:3 Río Cruces:4 Argentina:5.

Las medidas de influencia (Gráfico IV-10) muestran en la distancia de Cook las observaciones que más afectan en la estimación de los parámetros, puede verse que en la Zona de Chiloé existen casos que pueden presentar gran influencia en el cálculo de la varianza del error, sin embargo, no se consideran críticos (aunque en el límite), dado el gran tamaño muestral. El estudio de la situación opuesta, correspondiente al estadístico Leverage que detecta la influencia y presencia de outliers en el campo de la variable independiente, explica en parte el cálculo de un valor F tan grande en el ANOVA (Tabla IV-6) y la elevada proporción que tiene la varianza de los tratamientos frente al error. El valor que alcanza este estadístico en relación a la Zona de Chiloé (0,04), a pesar de presentarse gráficamente muy grande, se encuentra dentro de valores de riesgo (0,04 – 0,05) y no se considera necesario recalculer el estadístico F.

4.3 Análisis clasificatorio

Aún cuando el Análisis de Varianza entregue una visión clara de la distribución de las poblaciones existentes en la muestra, esta inferencia no permite en la práctica la asignación de observaciones a algunas de las agrupaciones observadas, de manera que resulta coherente intentar la construcción de una regla clasificatoria que permita tomar decisiones en la adscripción de nuevos ejemplares a alguno de los grupos disponibles o propuestos.

4.3.1 Análisis discriminante de las Zonas de agrupación.

La construcción de una regla que permita clasificar los especímenes, se evaluó en primera instancia en las cinco Zonas de Agrupación iniciales, lo que permitirá evaluar el comportamiento de éstas en términos de la clasificación, usando para ello todas las variables disponibles (Tabla IV-2).

Las Funciones Discriminantes (Tabla IV-8), necesarias para clasificar tanto la muestra usada (validación cruzada) como a nuevas observaciones, son combinaciones lineales de las variables usadas, es decir, para clasificar una observación cualquiera se debe sumar, incluyendo la constante, el resultado de multiplicar el valor que toma cada variable en el nuevo espécimen por el coeficiente que le corresponde. De esta forma, al comparar el valor que toma el cálculo de un espécimen en cada una de las cinco funciones, el mayor valor observado determinará la Zona a la cual este pertenece.

Tabla IV-8: Funciones discriminantes lineales.

	Chiloé	Bueno	Calle-Calle	Cruces	Argentina
Constante	-26,349	-47,107	-27,321	-31,972	-28,168
LC	0,128	0,196	0,897	1,325	1,129
LPC	5,025	5,325	3,224	3,133	2,076
AC	-3,125	-5,607	-3,244	-3,280	-1,862
APC	0,663	3,308	0,063	-0,358	-0,427
AF	6,183	8,212	4,373	2,925	2,288
AIO	-6,346	-5,878	0,163	2,332	2,031
ESC	0,091	0,414	0,397	0,017	-0,016
DIENT	2,003	2,547	1,507	1,928	1,611

Una de las finalidades del análisis discriminante es la determinación de la contribución de las variables en la discriminación de los grupos (Tabla IV-9), esto se consigue mediante la evaluación del estadístico F de remoción, el cual evalúa el valor discriminante de una variable ajustando el efecto que tienen las demás en la clasificación (Engelman, 2004^b). Puede observarse que LC es la variable que más contribuye a la discriminación, sin embargo, también posee el menor valor de tolerancia lo que significa que su capacidad discriminatoria, a pesar de ser la de mayor cuantía es dependiente de las demás variables, lo cual como se vió anteriormente se relaciona con la elevada correlación que esta variable tiene con las demás. En la situación contraria, la variable AIO es la que menos contribuye a la separación de los grupos y su valor de tolerancia habla también de que su dependencia de otras variables es menor que para LC, situación ya evaluada descriptivamente (Gráfico IV-4). Otra consideración es que las variables merísticas introducidas en este análisis presentan una contribución moderada y una alta independencia (relativa) en relación a las demás variables, lo que las posiciona como un aporte en la discriminación, por lo que la decisión de incluirlas en el análisis se ve reafirmada.

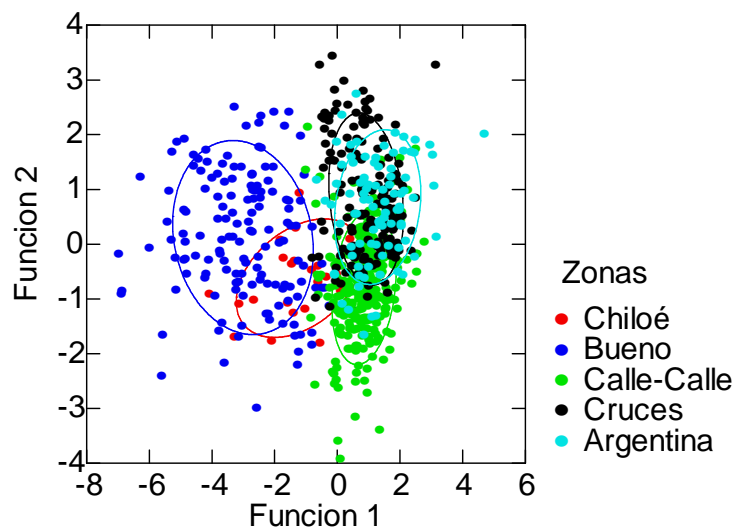
Tabla IV-9: Estadístico F de remoción y Tolerancia.

Variables	F-remoción	Tolerancia
LC	2,320	0,01778
LPC	9,760	0,02892
AC	32,920	0,03719
APC	64,350	0,09060
AF	30,260	0,08847
AIO	104,860	0,12407
ESC	39,060	0,82984
DIENT	52,940	0,96549

Quizás el resultado más importante de un análisis discriminante es la matriz de clasificaciones (Tabla IV-10), que representa el resumen del modelo. Puede apreciarse que las clasificaciones de las observaciones en las Zonas de agrupación, es heterogénea y en promedio, aunque la clasificación no es mala (72%), su aplicabilidad es reducida. La mejor clasificación la evidencia la Zona de Chiloé (87%), y la menor corresponde a la Cuenca del Río Cruces (56%). La idea de la existencia de tres grupos clasificatorios (en lugar de cinco) propuesta con anterioridad en el Análisis de Varianza, tiende a parecer coherente al observar que el bloque compuesto por las tres Zonas más septentrionales presentan una dispersión clasificatoria que funciona en forma de bloque. Cabe destacar el leve solapamiento que presenta la Cuenca del Río Bueno clasificando 13 observaciones en Calle-Calle. Como medida descriptiva el Gráfico de dispersión de las dos primeras funciones discriminantes canónicas (Gráfico IV-11), las cuales acumulan un 91,7% de la variabilidad de la muestra, exponen de forma gráfica la situación descrita en la matriz de clasificaciones.

Tabla IV-10: Matriz de clasificaciones. Filas clasificadas en las columnas.

	Chiloé	Bueno	Calle-Calle	Cruces	Argentina	Total
Chiloé	21 87,5%	3 12,5%	0	0	0	24 100,0%
Bueno	19 12,8%	115 77,7%	13 8,8%	0	1 0,7%	148 100,0%
Calle-Calle	0	0	191 75,8%	34 13,5%	27 10,7%	252 100,0%
Cruces	1	0	35 21,6%	91 56,2%	35 21,6%	162 100,0%
Argentina	0	0	13 14,3%	18 19,8%	60 65,9%	91 100,0%
Bien clasificado	87,5%	77,7%	75,8%	56,2%	65,9%	72,6%
Mal clasificado	12,5%	22,3%	24,2%	43,8%	34,1%	27,4%
Clasificación zona	41	118	252	143	123	677

**Gráfico IV-11:** Diagrama de dispersión de las dos primeras funciones discriminantes canónicas agrupadas por Zona.

La validación cruzada (Tabla IV-12) es el procedimiento mediante el cual se estiman los valores de las funciones discriminantes de las Zonas para cada observación y se asignan éstas a los grupos, es decir, se emplea la regla de discriminación creada para clasificar a las observaciones que se usaron para construirla. Aunque no ideal, es un buen método para evaluar la variabilidad en el ajuste de la clasificación. Puede observarse que en general los resultados de la validación cruzada son muy similares a la matriz de clasificación original, quizás con un poco menos de precisión pero la suficiente como para considerar que la muestra tiene un buen ajuste al modelo.

Tabla IV-12: Validación cruzada. Matriz de clasificaciones de las observaciones muestreadas.

	Chiloé	Bueno	Calle-Calle	Cruces	Argentina	Total
Chiloé	21 87,5%	3 12,5%	0	0	0	24 100,0%
Bueno	22 14,9%	110 74,3%	15 10,1%	0	1 0,7%	148 100,0%
Calle-Calle	0	0	191 75,8%	34 13,5%	27 10,7%	252 100,0%
Cruces	1 0,6%	0	35 21,9%	89 55,6%	35 21,9%	160 100,0%
Argentina	0	0	13 14,3%	20 22,0%	58 63,7%	91 100,0%
Bien clasificado	87,5%	74,3%	75,8%	55,6%	63,7%	71,4%
Mal clasificado	12,5%	25,7%	24,2%	44,4%	36,3%	28,6%
Clasificación zona	44	113	254	143	121	675

4.3.2 Análisis discriminante de los Bloques propuestos

Como ya se ha mencionado, la observación realizada en el Análisis de Varianza, del apartado anterior, para las comparaciones pareadas (Tabla IV-7), se ve reafirmado en el proceso de construcción de un sistema clasificatorio de las observaciones y parece apoyar la existencia de un sistema de agrupación de las Zonas geográficas, que concentra a las tres Zonas septentrionales y las dos australes formando tres bloques (Tabla IV-13).

Tabla IV-13: Asignación de las Zonas a los Bloques de agrupación propuestos.

Bloque	Zona	N muestra	N bloque
Chiloé	Isla grande de Chiloé	26	26
Río Bueno	Cuenca del Río Bueno	194	194
Norte	Cuenca del Río Calle - Calle	252	513
Norte	Cuenca del Río Cruces	165	
Norte	Argentina	96	

Resulta evidente la necesidad de construir una regla de clasificación para los grupos propuestos mediante un Análisis Discriminante, para esto se ha recodificado la variable de agrupación (Tabla IV-13) y se han construido las funciones discriminantes correspondientes (Tabla IV-14).

Tabla IV-14: Funciones discriminantes de los grupos propuestos.

	Chiloé	Río Bueno	Norte
Constante	-24,996	-45,185	-26,911
LC	-0,031	0,001	0,877
LPC	4,846	5,043	2,860
AC	-2,831	-5,175	-2,769
APC	0,768	3,430	-0,004
AF	6,473	8,499	4,016
AIO	-7,032	-6,676	0,285
ESC	0,190	0,525	0,347
DIENT	1,825	2,323	1,446

El valor relativo de ponderación de las variables en la discriminación de Bloques (Tabla IV-15) se ve modificada en relación a la discriminación de cinco Zonas. De esta forma puede observarse que la variable que más contribuye a la separación de los grupos corresponde a LC, seguida de ESC (Número de escamas carpales) y la que menos contribuye es AIO seguida de APC, etc. Los valores de tolerancia de las variables se mantienen debido a que la correlación de las variables no cambia y sólo muestra variaciones el poder discriminante de las variables dada la nueva variable de agrupación.

Tabla IV-15: Estadístico F de remoción y Tolerancia.

Variabes	F-remoción	Tolerancia
LC	3,220	0,01741
LPC	13,350	0,02881
AC	40,440	0,03836
APC	123,380	0,08858
AF	42,700	0,09187
AIO	175,200	0,12722
ESC	10,380	0,83169
DIENT	68,160	0,94698

La matriz de clasificaciones para el contraste de los tres Bloques propuestos (Tabla IV-16) muestra que a diferencia de la utilización de las cinco Zonas, la clasificación es mucho más homogénea, alcanzando una eficacia cercana al 90% situación que sí merece significación práctica en la asignación de especímenes a alguno de los tres grupos propuestos. No se ha

dispuesto de la validación cruzada por presentar ésta una clasificación igual a la estimación inicial, lo que habla del buen ajuste que tienen los Bloques propuestos al modelo.

Tabla IV-16: Matriz de clasificaciones, filas clasificadas en las columnas

	Chiloé	Río Bueno	Norte	Total
Chiloé	21 87,5%	3 12,5%	0	24 100,0%
Río Bueno	19 12,8%	119 80,4%	10 6,8%	148 100,0%
Bloque Norte	5 1,0%	0	500 99,0%	505 100,0%
Bien clasificado	87,5%	80,4%	99,0%	89,0%
Mal clasificado	12,5%	19,6%	1,0%	11,0%
Clasificación zona	45	122	510	677

El gráfico de dispersión de las puntuaciones de las observaciones para las funciones discriminantes canónicas (Gráfico IV-12) muestra descriptivamente los resultados de la clasificación para los bloques propuestos. Bajo este prisma resulta fácil observar la gran dispersión de algunas observaciones que se escapan a la elipse de concentración muestral.

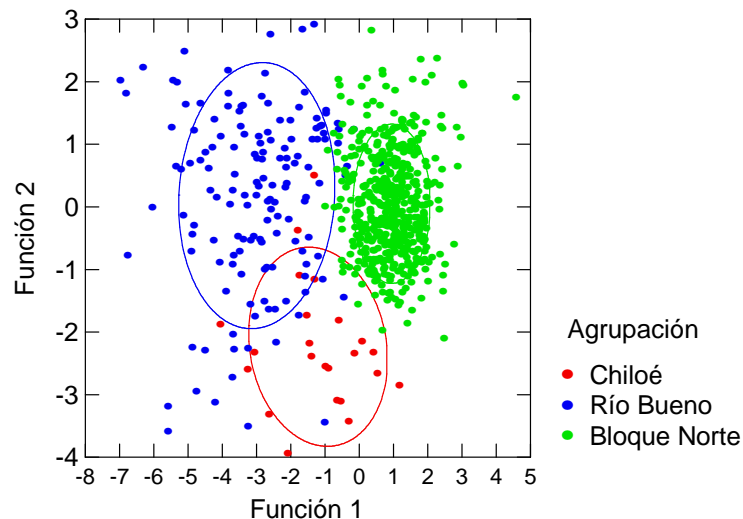


Gráfico IV-12: Diagrama de dispersión de las dos primeras funciones discriminantes canónicas agrupadas por Bloque.

4.4 Análisis de datos cualitativos

La estrategia seleccionada para el análisis de los caracteres cualitativos, que fueron tabulados en una matriz binaria de presencia ausencia, obedece a un intento por determinar cuales de las variables presumiblemente diagnósticas (Tabla IV-18) de los morfos *Abtao* y *Riolimayana* (Jara, 1980) presentan diferencias para los Zonas de agrupación, evaluar la existencia de grupos al interior de la matriz, e intentar definir el o los morfos que definen a las Zonas o grupos predeterminados.

4.4.1 Evaluación de las Zonas de agrupación.

El primer paso del análisis de la matriz binaria es evaluar la capacidad de las variables disponibles para diferenciar entre las Zonas de agrupación (Tabla III-1), mediante una prueba de bondad de ajuste a la frecuencia promedio, a través de las Zonas.

Tabla IV-17: Test X^2 . Frecuencia esperada correspondiente al promedio de las frecuencias de las Zonas; P-Valor con 4 grados de libertad.

Zona	A	B	C	D	E	F	G	H	I	J	K	L
Chiloé	83,3	87,5	45,8	58,3	100,0	8,3	8,3	83,3	29,2	33,3	12,5	66,7
Río Bueno	82,2	78,1	63,0	78,1	93,2	87,7	82,2	84,9	39,7	58,9	26,0	87,7
Calle-Calle	71,1	66,7	46,7	68,9	77,8	66,7	24,4	80,0	58,9	37,8	16,7	64,4
Río Cruces	64,1	28,1	43,8	81,3	90,6	70,3	40,6	92,2	46,9	35,9	10,9	78,1
Argentina	67,4	56,5	34,8	47,8	76,1	45,7	4,3	95,7	56,5	82,6	54,3	19,6
Freq. Esperada	73,6	63,4	46,8	66,9	87,5	55,7	32,0	87,2	46,2	49,7	24,1	63,3
Chi-Square	4,13	33,11	8,92	11,55	4,83	66,40	124,28	1,93	12,98	35,55	53,19	43,28
P-Valor	0,388	0,000	0,063	0,021	0,305	0,000	0,000	0,749	0,011	0,000	0,000	0,000

El Test de X^2 (Tabla IV-17) muestra las variables cuyos componentes de agrupación se desvían de manera significativa del valor esperado (frecuencia promedio) para cada Zona. Como era de esperar, no todas las variables presentan significación, de manera que las variables B, D, F, G, I, J, K y L (Tabla III-4) quedan seleccionadas como el conjunto de variables que mejor separa a las

Zonas y pueden ser consideradas como un aporte en la discriminación de los grupos existentes en la matriz binaria.

Con las ocho variables que poseen la mayor posibilidad de distinguir entre las Zonas, se procedió a la construcción de una medida que resume estas variables binarias. Con ese fin se calculó, en las observaciones disponibles la frecuencia de atributos de Tipo *Abtao* (Tipificado como 1) que cada espécimen presenta. Como puede observarse, (Gráfico IV-13) la medida de resumen muestra una preferencia de combinaciones binarias de tipo intermedio sin privilegio de los morfotipos extremos. Con esta estimación de cuán *Riolimayana* o *Abtao* es cada individuo muestreado se compararon las Zonas de agrupación mediante una Prueba de rangos de Kruskal-Wallis y un Test de Dunn para el caso de las comparaciones pareadas.

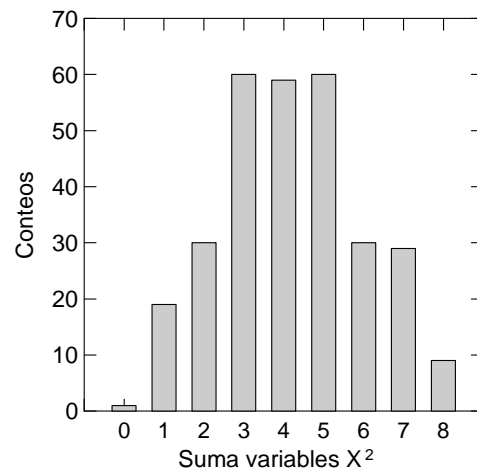


Gráfico IV-13: Gráfico de Barras de la medida de resumen de las ocho variables binarias seleccionadas en el Test de X^2 .

Tabla IV-18: Test de Kruskal-Wallis, para la medida de resumen de las variables seleccionadas en el Test de X^2 .

H = 50,116 con 4 grados de libertad. (P = <0,001)

Zona	N	Mediana	25%	75%
Chiloé	24	3	2,0	4,0
Bueno	73	5	4,0	7,0
Calle-Calle	90	4	3,0	5,0
Cruces	64	4	3,0	5,0
Argentina	46	3	2,0	5,0

El Test de Kruskal-Wallis (Tabla IV-18) muestra la existencia de diferencias significativas en al menos una de las Zonas de agrupación. La Zona de Río Bueno presenta la mediana de mayor cuantía y se convierte en el grupo más representativo del Tipo *Abtao* con la mayor cantidad de atributos que determinan este morfo. En relación a este análisis es importante señalar que los rangos intercuartílicos indican que a excepción de la Zona de Río Bueno existe un gran solapamiento de las distribuciones de las Zonas restantes.

Tabla IV-19: Test de Dunn. Comparaciones pareadas de las Zona de agrupación. Las comparaciones significativas están señaladas con “Si”.

Comparación	Dif. Ranking	Q	P <0,01
2 y 1	115,435	5,71	Si
2 y 5	82,935	5,13	Si
2 y 4	67,505	4,59	Si
2 y 3	62,585	4,63	Si
3 y 1	52,850	2,68	No
4 y 1	47,930	2,33	No
5 y 1	32,500	1,50	No
3 y 5	20,350	1,31	No
4 y 5	15,430	0,93	No
3 y 4	4,920	0,35	No

Las comparaciones pareadas realizadas con el Test de Dunn (Tabla IV-19) muestran de manera probabilística lo ya señalado en relación a la diferencia entre los especímenes de la Cuenca de Río Bueno y las demás Zonas, las cuales dejan en evidencia su solapamiento al mostrar la inexistencia de diferencias significativas entre ellas, asignando a la Zona de Río Bueno una

distribución marginal dentro de la muestra. En lo sucesivo sólo se considerarán como Zonas de agrupación a Río Bueno y al bloque formado por los grupos restantes.

4.4.2 Determinación de morfos

Como la Zona de Río Bueno y las restantes Zonas generan dos bloques estadísticamente distintos se hace necesario generar una regla para asignar futuras observaciones a uno de estos dos grupos. Se han determinado dos estrategias tendientes a resolver esta problemática.

Determinación mediante puntos de corte (Cuantiles)

Una manera de asignar nuevas observaciones a los grupos descubiertos con el Test de Kruskal-Wallis es la determinación de los puntos de corte resultante de la superposición del 95% acumulado de las distribuciones de estos bloques de agrupación.

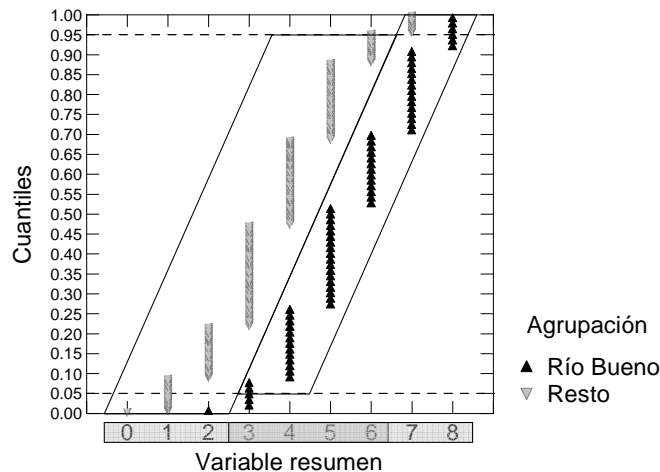


Gráfico IV-14: Diagrama de cuantiles.

El diagrama de cuantiles (Gráfico IV-14) construido para determinar el punto de corte de las distribuciones de Río Bueno y el resto de las Zonas muestra que el 95% de las observaciones del

bloque correspondiente a la Cuenca del Río Bueno se encuentra entre 3 y 8, mientras que el bloque de las Zonas restantes tiene rangos entre 0 y 6, creando con esto una Zona de solapamiento entre 3 y 6 lo cual significa que para un espécimen cualquiera la presencia de entre 7 y 8 caracteres de tipo *Abtao* implica su adscripción al bloque de Río Bueno y la presencia de entre 0 y 2 caracteres de tipo *Abtao* se asocia al bloque de las Zonas restantes. La presencia o ausencia de estos caracteres está determinada por el uso de las ocho variables seleccionadas en el Test de X^2 y empleadas en la construcción de la variable de resumen de las combinaciones binarias y probadas en el análisis de rangos. Los resultados para esta alternativa de clasificación (Gráfico IV-15) muestran que la distribución de los especímenes clasificados como representantes del morfotipo *riolimayana* tienden a aumentar de acuerdo al patrón de conexión de las Zonas entre Río Bueno y Argentina, la situación opuesta se muestra para los especímenes de morfo *abtao*, los cuales se presentan en mayor cuantía en la cuenca del Río Bueno, decreciendo a medida que se avanza hacia el norte a través de las Zonas. Por último, se destaca la gran cantidad de especímenes que presentan una mezcla de caracteres entre los dos morfos estudiados, sin mostrar algún patrón de variación geográfica.

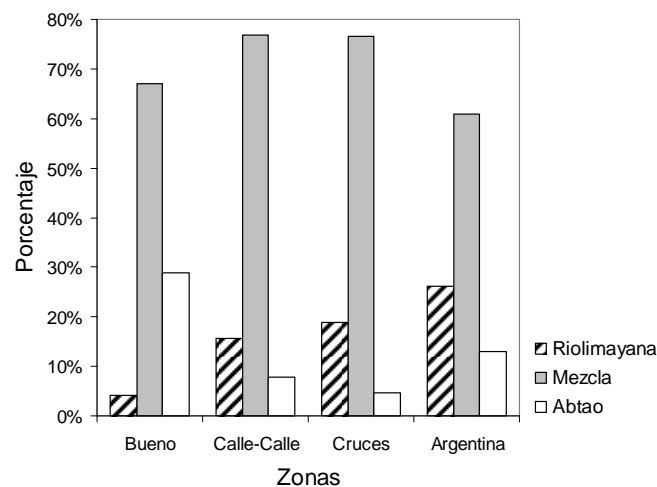


Gráfico IV-15: Resultados de la clasificación de especímenes usando el criterio de cuantiles.

Determinación mediante probabilidades

La metodología recién expuesta se basa en la determinación de especímenes mediante la cantidad de atributos observados para una variable que resume las observaciones binarias y no hace referencia a cuales atributos son los que representan a los morfotipos. Es por esto que se plantea una metodología alternativa para salvar este problema. Este método alterno implica minimizar la cantidad de variables a considerar, para reducir así la cantidad posible de combinaciones que estas generan. Con tal objetivo se realizó un Test de X^2 (Tabla IV-20) dirigido a excluir las variables que no presentan diferencias significativas en la frecuencia de atributos de uno u otro morfotipo frente a la frecuencia esperada para los nuevos grupos propuestos.

Tabla IV-20: Test de X^2 para el ajuste de variables a los grupos propuestos.

Zona	A	B	C	D	E	F	G	H	I	J	K	L
Resto	69,6	55,8	43,3	67,0	83,5	57,1	23,2	87,1	51,8	46,0	22,3	59,4
Río Bueno	82,2	78,1	63,0	78,1	93,2	87,7	82,2	84,9	39,7	58,9	26,0	87,7
Freq. Esperada	75,9	66,9	53,2	72,5	88,3	72,4	52,7	86,0	45,8	52,4	24,2	73,5
Chi-Square	1,037	3,707	3,654	0,852	0,529	6,436	32,999	0,026	1,589	1,592	0,284	5,445
P-Valor	0,31	0,05	0,06	0,36	0,47	0,01	0,00	0,87	0,21	0,21	0,59	0,02

El Test de X^2 (Tabla IV-20) realizado sobre las variables binarias originales muestra que en relación a B, F, G y L existen preferencias significativas por uno u otro morfo. Estas variables se concatenaron para generar las combinaciones binarias, a partir de estas se calcularon las probabilidades de pertenencia a los bloques de clasificación, para cada uno de los morfos presentes, además de los intervalos de confianza respectivos.

Tabla IV-21: Probabilidades de pertenencia a un bloque de agrupación para las combinaciones binarias, límites confianza 95% y porcentajes del morfo sobre el total.

Morfo B F G L	Porcentaje por morfos						Total Morfo	% del Total
	Limite Inf.	Río Bueno	Limite Sup.	Limite Inf.	Resto	Limite Sup.		
0000				82,1%	100,0%	100,0%	100%	7,8%
0001	0,4%	11,8%	38,7%	56,7%	88,2%	98,7%	100%	6,0%
0100				69,4%	100,0%	100,0%	100%	4,3%
0101	0,0%	4,0%	22,2%	74,1%	96,0%	99,9%	100%	8,9%
0110				31,9%	100,0%	100,0%	100%	1,4%
0111	16,6%	35,5%	56,4%	41,6%	64,5%	81,8%	100%	11,0%
1000	0,2%	6,7%	23,8%	73,3%	93,3%	99,3%	100%	10,7%
1001	1,0%	9,4%	26,7%	70,7%	90,6%	98,2%	100%	11,4%
1010	0,0%	50,0%	99,5%	0,0%	50,0%	99,5%	100%	0,7%
1100	0,0%	5,9%	31,0%	63,9%	94,1%	99,9%	100%	6,0%
1101	0,2%	7,4%	26,1%	70,7%	92,6%	99,2%	100%	9,6%
1110	0,3%	30,0%	67,4%	27,2%	70,0%	93,9%	100%	3,6%
1111	71,3%	86,5%	94,8%	4,5%	13,5%	27,2%	100%	18,5%
Total		25,3%			74,7%		100%	100%

La tabla de probabilidad (Tabla IV-21) muestra para una combinación binaria cualquiera la probabilidad de pertenecer a uno de los dos bloques de agrupación, con sus respectivos límites de confianza al 95%. La gran variabilidad que se observa para los límites de confianza en algunos de los morfos se debe principalmente a la escasa cantidad de observaciones que acumula esa combinación, por ejemplo la combinación 1110 representa sólo el 3,6% de las observaciones. De todas las combinaciones presentes en la muestra para las variables que se cree son las que más discriminan entre los bloques propuestos, 1111 es la única que prevalece para la Zona de Río Bueno, lo cual implica que este es el morfo más representativo de este bloque. Sin embargo, de seleccionar un espécimen que presente esta combinación, la probabilidad de que pertenezca a la Cuenca del Río Bueno es tan sólo de 86,5%, en el mejor de los casos un 94,8%. Dicho de otra forma, de obtener cualquier combinación binaria para las variables B, F, L y G lo más probable es, que de no ser 1111 corresponda a un individuo perteneciente a alguna de las cuatro Zonas que excluye la presencia de Río Bueno.

5. DISCUSIÓN

5.1 Descripción de las especies y síntesis de los resultados

Cuando Schmitt (1942^{a,b}) describió a *Aegla abtao* y *Aegla riolimayana* como nuevas especies para la ciencia, basó su propuesta en una combinación discreta de caracteres morfológicos relativos la forma del rostrum y del caparazón, los cuales al combinarse con los patrones de distribución geográfica, configuraron un escenario, que en su momento juzgó como argumentos suficientes para el reconocimiento de las especies.

En el presente estudio se ha intentado evaluar las decisiones y consideraciones hechas por Schmitt (1942^{a,b}), Ringuelet (1960) y Jara (1996) sobre la relación existente entre *Aegla abtao* y *Aegla riolimayana*. Para tal fin se analizaron algunas de las biometrías y caracteres morfológicos que los autores ya mencionados utilizaron con anterioridad, cuidando de cumplir a cabalidad todos los aspectos teóricos implicados. El resultado de los análisis muestra que existe, para el rango geográfico contemplado (Ver Pág. 11), una serie de poblaciones biológicas asociadas a las cuencas de los principales ríos del Sur de Chile y de la provincia de Río Negro en Argentina, las cuales tienden a agruparse, definiendo un patrón de dispersión que sitúa a las Cuencas de los Ríos Calle-Calle, Cruces y Limay (Argentina) como un conjunto homogéneo de poblaciones, en las cuales los organismos que las componen no se diferencian significativamente entre sí, tanto para las somatometrías como para los caracteres morfológicos contemplados. De esta forma se puede concluir que para este conjunto de poblaciones, existen individuos que pueden ser considerados como representantes del morfotipo *riolimayana* y organismos en los cuales se pueden observar una mixtura de características de los tipos *abtao* y *riolimayana*, sin embargo la distinción de

especímenes para estas tres cuencas no muestra un patrón geográfico asociado. La Cuenca del Río Bueno se presenta como una población muy homogénea y diferente de las demás, cuyos componentes pueden ser considerados con propiedad como exponentes del morfotipo *abtao*. Por otra parte, los confusos resultados que presenta la Zona de Chiloé llevan a pensar que el reducido número de especímenes con los que se cuenta no son, en términos estadísticos, representativos de esta amplia Zona muestral.

5.2 Revisión histórica del problema.

Considerando los resultados de los análisis estadísticos realizados, resulta evidente que los fundamentos de Schmitt (1942^{a,b}) para la descripción de las especies se ven debilitados, probablemente porque en tal decisión el elemento más influyente lo constituye la relación geográfica evaluada en ese instante para los morfotipos en cuestión. Sin embargo, como se ha mostrado, tanto en relación a la dificultad para la determinación de las especies por parte de algunos investigadores, como en las estimaciones probabilísticas construidas, no existe evidencia para sostener que el morfotipo *riolimayana* sea exclusivo de los ríos trasandinos, al contrario, presenta una distribución aún mayor, abarcando dos importantes cuencas hidrográficas del sur de Chile. Por otra parte, la tentativa de Ringuelet (1948) por reducir a *Aegla riolimayana* a subespecie de *Aegla abtao* carece de rigurosidad, debido que el fundamento de tal propuesta considera una gran cantidad de especímenes trasandinos y solo algunos especímenes chilenos, presumiblemente representativos del morfotipo *abtao*, por esta situación considero que tal propuesta no merece mayor crédito dada la ausencia de un marco muestral conveniente que abarque una distribución geográfica representativa de los organismos y un tamaño muestral adecuado. La observación de que *Aegla riolimayana* es un sinónimo menor de *Aegla abtao*,

propuesta por Jara (1996) es un intento metodológico coherente encaminado a saldar el problema a través de metodologías estadísticas. Sin embargo, a pesar de contar con un marco muestral robusto, en términos de la extensión geográfica de la muestra y el tamaño de esta, el estudio posee algunas falencias estadísticas en relación a la potencia y el resguardo de los supuesto matemáticos de los análisis, los cuales fueron abordados de manera excesivamente conservadora. Tal situación provoca que la elección de las metodologías estadísticas evaluadas generen resultados confusos, lo cual deriva en que las decisiones que se desprenden de estos sean también incorrectos. Cabe destacar que el presente estudio hace uso de la muestra empleada por Jara (1996), y puede ser considerada como una corrección metodológica y estadística de tal estudio.

5.3 Interpretación biológica de los resultados

A partir de los resultados obtenidos surge una interpretación de tipo biológico, para explicar la dinámica poblacional observada en los dos morfotipos atribuidos por Schmitt (1942^{a,b}) a especies diferentes. Tal interpretación implicaría que los taxa involucrados tengan un nivel de diferenciación fenotípica equivalente a subespecies (Mayr, 1969), lo cual parece coherente a la luz de los resultados que se desprenden de la metodología para la determinación de morfos derivada de los caracteres morfológicos (Gráfico IV-15). Bajo el prisma de esta explicación la presencia de especímenes con caracteres mezclados podría deberse a hibridación (Mayr, 1969), sin embargo, la presencia de un cinturón híbrido presupone también la existencia de poblaciones geográficamente separadas (alopátricas) y morfológicamente distintas (subespecies), ubicadas en los límites del territorio donde se observan los organismos con caracteres mezclados. En este caso la mezcla de caracteres adquiere una ponderación muy alta en relación a los especímenes “puros”, los cuales a pesar de mostrar tendencias de diferenciación morfológica y geográfica se

ven disminuidos por la baja presencia de estos en las Zonas. De esta forma cualquier tentativa de validar la proposición de subespecies carece de significación debido a la gran cantidad de organismos con características intermedias entre los dos morfotipos. No resultaría imprudente señalar que los morfotipos en cuestión puedan interpretarse como razas de una sola especie (Mayr, 1969), sin embargo esta consideración resulta poco práctica dada la escasa objetividad que existe en la aplicación de este concepto.

Se debe señalar que aunque Ringuelet (1960^a), Bahamonde y López (1963) y Schuldt et al. (1988) optaron por emplear la nominación subespecífica, para darle una connotación taxonómica a la variabilidad fenotípica observada en especímenes del Lago Lacar, la Isla de Chiloé y el Lago Huechulafquen, respectivamente, su decisión no parece haber estado basada en la aplicación del concepto de subespecie propuesto por Mayr (1969).

5.4 Conclusiones y recomendaciones

Concluyendo la evaluación de la problemática que ha abarcado este trabajo se debe señalar que la hipótesis de que *Aegla abtao* Schmitt y *Aegla riolimayana* Schmitt son entidades bien definidas desde el punto de vista fenotípico y geográfico (Ver Pág. 12) ha sido refutada, por considerar que no existen elementos suficientes para su validación. Sin embargo, los resultados han mostrado una coherencia matemática y teórica suficiente para suponer alternativas lógicas a esta hipótesis. En relación a los objetivos (ver Pág. 13) propuestos debe destacarse que se han cumplido a cabalidad, evidenciando el buen ajuste de la muestra a los modelos matemáticos empleados.

Por otra parte y con la responsabilidad que merece la investigación de las poblaciones biológicas, se debe señalar que resulta difícil, sino imposible, tomar una decisión acerca de la real situación taxonómica, geográfica y morfológica de estos especímenes sin contar con los antecedentes que definen la vida de éstos en su medio natural, tales como su ecología, etología, genética, etc. Por lo cual, el presente estudio representa sólo una pequeña fracción de la realidad multidimensional de los organismos estudiados, enfatizando la necesidad de ampliar estos resultados a los demás ámbitos de la biología poblacional. Debieran también corregirse algunas de las debilidades metodológicas de esta investigación, como contar con una muestra representativa de las Isla de Chiloé, una ampliación del número de afluentes de la Cuenca del Río Bueno, la evaluación de la influencia de los procesos geológicos asociados y la realización de análisis mediante sistemas de información geográfica.

6. REFERENCIAS

Bahamonde, N. & M.T. López. 1963. Decápodos de Aguas Continentales en Chile. Investigaciones Zoológicas Chilenas, 10:123-149.

Besley, D., Kuh, E. & Welsh, R. 1980. Regression diagnostics: Identifying influential data and sources of collinearity. Ed. John Wiley y Sons. Página 153 – 180.

Dallas E.J. 2000^a. Métodos multivariados aplicados al análisis de datos. Páginas: (93 – 142)^a, (100)^b

Engelman, L. 2001^{a,b}. Discriminant Analysis. En: Systat 11 Statistics guide, Statistics I. SYSTAT Software, Inc. Páginas (I 301)^a, (I 328 – I 329)^b.

Hair Joseph F. et al (1999) Análisis Multivariante. 5º Edición Pearson Educación S.A. Madrid España pp. 2 – 27.

Hollander, M. & D.A. Wolfe. 1973. Nonparametric Statistical Methods. Ed. John Wiley & Sons. pp. 503.

Jara, C.G. 1980. Taxonomía y Distribución del Género *Aegla* Leach (Crustacea, Decapoda, Anomura) en el Sistema Hidrográfico del Río Valdivia (Chile). Tesis, Facultad de Ciencias, Universidad Austral de Chile, Valdivia. 132 pp.

Jara, C.G. 1981. Exploración del valor taxonómico de la armadura gástrica de *Aegla* (Crust.: Anom.). Archivos de Biología y Medicina Experimentales, 14(3):271.

Jara, C.G. 1996. Taxonomía, sistemática y zoogeografía de las especies chilenas del género *Aegla* leach (Crustacea, Decapoda, Anomura, Aeglidae). Tesis Doctoral, Facultad de Ciencias Biológicas y Recursos Naturales, Universidad de Concepción, Concepción. 119 pp.

- Mayr, E. & P.H. Ashlock. 1991. Principles of Systematic Zoology. Ed. McGraw Hill. New York. 475 pp.
- Mayr, E. 1963. Especies animales y evolución. Harvard University Press, Cambridge, Massachusetts.
- Morrison, D.F. 1976. Multivariate Statistical Methods. Ed. Mc Graw-Hill. pp. 415.
- Paterson, N.F. 1968. The anatomy of the Cape Rock Lobster, *Jasus lalandii* (H. Milne-Edwards). Annals of the South African Museum, 51:1-232.
- Ringuelet, R.A. 1948. Los cangrejos argentinos del género *Aegla* de Cuyo y la Patagonia. Revista del Museo de La Plata, 5:297-347.
- Ringuelet, R.A. 1949. La Prioridad en el Estudio Biométrico de los Decápodos Anomuros del Género *Aegla* Leach. Notas del Museo de La Plata, 14:119-121.
- Ringuelet, R.A. 1960. Notas sobre *Aegla* de Argentina y Paraguay (Crust. Decap. Anomura). Physis, B. Aires, 21(61):231-239.
- Scheffé, H. 1953. A method for judging all contrasts in the analysis of variance. Biometrika, 40:87-104.
- Schmitt, W.L. 1942^a. Two New Species of *Aegla* from Chile. Revista Chilena de Historia Natural, 44(1940):25-31.
- Schmitt, W.L. 1942^b. The Species of *Aegla*, Endemic South American Freshwater Crustaceans. Proceedings of the United States National Museum, 91(3132):431-520.

Schuldt, M., P. Núñez, W. Mersing, A. del Valle, y M. Manceñido. 1988. *Aegla* (Crustacea Anomura) en el Lago Huechulafquén (Neuquén, Argentina) y algunas implicancias filogenéticas para Aeglidae del Centro-Oeste de Argentina. *Anales de la Sociedad Científica Argentina*, 217(50):27-37.

Spiegel, M. 2001^{a,b}. Probabilidad y estadística. Ed. McGraw-Hill. Páginas: 82 – 91^a, 374^b.

Stenson, H. & Wilkinson L. 2004. Factor Analysis. En: *Systat 11 Statistics guide, Statistics I*. SYSTAT Software, Inc. Páginas I 359 – I 398

Systat, 2004. *Systat 11 Language Reference user guide*. SYSTAT Software, Inc. Página: 87. En: www.systat.com

Weimer, R. 1999^{a,b}. Estadística. Ed. Continental. Páginas 581 – 605^a, 81^b.

Zuwaylif, F. 1977^{a,b}. Estadística general aplicada. Fondo Educativo Interamericano. Páginas: 308^a, 254^b.