

Universidade Federal de Santa Catarina
Programa de Pós-Graduação em Engenharia de Produção

**ESTIMAÇÃO DO FATOR DE CONDIÇÃO PARA
PEIXES UTILIZANDO MODELOS LINEARES
GENERALIZADOS**

Dissertação de Mestrado

Clédina Regina Lonardan Acorsi

Florianópolis

2002

Clédina Regina Lonardan Acorsi

**ESTIMAÇÃO DO FATOR DE CONDIÇÃO PARA
PEIXES UTILIZANDO MODELOS LINEARES
GENERALIZADOS**

Dissertação apresentada ao Programa de Pós-Graduação em
Engenharia de Produção da Universidade Federal de Santa Catarina
como requisito parcial para a obtenção do grau de Mestre em
Engenharia de Produção

Orientadora: Silvia Modesto Nassar, Dra.

Florianópolis

2002

Clédina Regina Lonardan Acorsi

ESTIMAÇÃO DO FATOR DE CONDIÇÃO PARA PEIXES UTILIZANDO MODELOS LINEARES GENERALIZADOS

Esta dissertação foi julgada e aprovada para a obtenção do grau de **Mestre em Engenharia de Produção no Programa de Pós-Graduação em Engenharia de Produção** da Universidade Federal de Santa Catarina.

Florianópolis, 29 de novembro de 2002

Edison Pacheco Paladini, Dr.

Coordenador do Curso

BANCA EXAMINADORA

Prof^a. Silvia Modesto Nassar, Dra.

Orientadora

Prof. Pedro Alberto Barbetta, Dr.

Membro

Prof. Paulo José Ogliari, Dr.

Membro

Prof. Marcelo Menezes Reis, Dr.

Membro

Aos amores de minha vida:

Renam e Rayanne, por serem os melhores filhos que alguém pode desejar. Ao meu esposo Luiz Carlos por entender minhas ausências no nosso dia a dia. Ao meu irmão, Clever, que nos momentos mais difíceis tem sido a força intensa capaz de me impulsionar. À minha cunhada Márcia e meus sobrinhos Hudson e Flávia que nunca permitiram que eu me sentisse sozinha... À minha mãe, Maria Mariana, por seu carinho incondicional e muito especialmente ao meu pai, Romeu Lonardon, que do infinito onde se encontra, certamente ainda é capaz de sorrir feliz com esta minha realização...

AGRADECIMENTOS

À Universidade Federal de Santa Catarina.

Ao professor Doutor Gauss Moutinho Cordeiro, pelo exemplo de dedicação e generosidade ao definir-me a teoria a ser estudada.

Ao professor Doutor Luiz Carlos Gomes, por tão gentilmente confiar-me dados para pesquisa.

À professora Doutora Terezinha Aparecida Guedes pelas valiosas palavras em momentos cruciais.

Ao professor Josmar Mazucheli pelo apoio e atenção.

À professora Isolde Previdelli, pela viabilização do curso.

À Professora Rosângela Getirana Santana e Margareth C. T. Udo pela ajuda e disponibilidade perante minhas dúvidas.

A todos meus colegas de curso e em especial aos amigos e colegas de grupo Angela Maria Marcone Araujo, Sebastião

Gazola e Valentina de L. M. de P. Soares.

À minha orientadora Prof^a. Silvia Modesto Nassar, pelo acompanhamento pontual e competente.

Aos professores do Curso de Pós Graduação em Engenharia de Produção.

A todos os que direta ou indiretamente contribuíram para a realização desta pesquisa.

“O paradigma ecológico é alicerçado pela ciência moderna, mas se acha enraizado numa percepção da realidade que vai além do arcabouço científico, no rumo de uma consciência da unidade de toda a vida e da interdependência de suas múltiplas manifestações e de seus ciclos de mudança e transformação. Em última análise, essa profunda consciência ecológica é consciência espiritual. Quando o conceito de espírito humano é entendido como o modo de consciência em que o indivíduo se sente ligado ao cosmo como um todo, fica claro que a percepção ecológica é espiritual em sua essência mais profunda, e então não é surpreendente o fato de que a nova visão da realidade esteja em harmonia com as concepções das tradições espirituais”.

Fritjof Capra

RESUMO

ACORSI, Clédina Regina Lonardan Acorsi. **Estimação do fator de condição para peixes utilizando modelos lineares generalizados**. 2002. 115f. Dissertação (Mestrado em Engenharia de Produção) - Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis.

Este trabalho busca aliar através de procedimentos teóricos e práticos, conhecimentos de duas ciências: a Ictiologia e a Estatística. Nele descreve-se a construção de um indicador do “bem-estar do peixe”, denominado fator de condição do peixe representado por K . Este fator de condição fundamenta-se na relação peso-comprimento do animal, cujo modelo será construído segundo a teoria dos Modelos Lineares Generalizados. O estudo está baseado em um conjunto de 15.175 indivíduos pertencentes às espécies *Trachydoras paraguayensis*, *Liposarcus anisiti*, *Raphiodon vulpinus*, *Serrasalmus marginatus* e *Leporinus friderice*. A captura dos peixes foi realizada na planície de inundação do Alto Rio Paraná, sob coordenação do Núcleo de Pesquisas em Limnologia, Ictiologia e Aqüicultura, da Universidade Estadual de Maringá. O procedimento metodológico inicial desta pesquisa detalha o processo de construção de um modelo geral, capaz de definir o peso total de um peixe considerando, simultaneamente, as diversas espécies em estudo e sua classificação quanto ao sexo e comprimento. A segunda fase é a construção de modelos mais parcimoniosos, capazes de relatar o peso total do animal de maneira mais simplificada que a anterior. A partir destes modelos, é definido um fator de condição K , cuja determinação é o objeto central desta pesquisa. Durante o processo de construção dos modelos, a variável espécie mostra-se como uma variável significativa, enquanto o fator sexo é significativa somente para as espécies *Trachydoras paraguayensis* e *Raphiodon vulpinus*. Conseqüentemente, os sete fatores de condição do peixe apresentados, corresponderão às espécies *Liposarcus anisiti*, *Serrasalmus marginatus*, *Leporinus friderice*, às fêmeas das espécies *Trachydoras paraguayensis* e *Raphiodon vulpinus* e aos machos das espécies *Trachydoras paraguayensis* e *Raphiodon vulpinus*.

Palavras chaves: Modelos Lineares Generalizados; Peixe; Fator de condição.

ABSTRACT

ACORSI, Clédina Regina Lonardan Acorsi. **Estimação do fator de condição para peixes utilizando modelos lineares generalizados**. 2002. 115f. Dissertação (Mestrado em Engenharia de Produção) - Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis.

This work search to ally through theoretical and practical procedures, knowledge of two sciences: Ichthyology and Statistics. In him is described construction of an indicator of the "well-being of the fish", denominated factor of condition of the fish is described. This condition factor is based in the relationship weight-length of the fish, whose model will be built according to the theory of the Generalized Linear Models. The study is based on a group of 15.175 individuals belonging to the species *Trachydoras paraguayensis*, *Liposarcus anisiti*, *Raphiodon vulpinus*, *Serrasalmus marginatus* and *Leporinus friderice*. The capture of the fish was accomplished in the plain of flood of High Paraná River, under coordination of the Nucleus of Researches in Limnology, Ichthyology and Aquiculture (Nupélia), of the State University of Maringá. The procedure methodological initial of this research details the process of construction of a general model, capable to define the total weight of a fish considering, simultaneously, the several species in study and its classification with relationship to the sex and length. The second phase is the construction of more parsimonious models, capable to tell the total weight of the animal in a simplified way than the previous. Starting from these models, it is defined a condition factor K, whose determination is the central object of this research. During the process of construction of the models, the variable species is shown as a significant variable, while the factor sex is only significant for the species *Trachydoras paraguayensis* and *Raphiodon vulpines*. Consequently, the seven factors of condition of the presented fish, will correspond to the species *Liposarcus anisiti*, *Serrasalmus marginatus*, *Leporinus friderice*, to the females of the species *Trachydoras paraguayensis* and *Raphiodon vulpinus* and to the males of the species *Trachydoras paraguayensis* and *Raphiodon vulpinus*.

Key words: Generalized Lineal Models; Fish; Factor of Condition.

SUMÁRIO

1- INTRODUÇÃO.....	15
1.1 - PROBLEMA.....	17
1.2 - OBJETIVOS.....	18
1.2.1 - Objetivo Geral.....	18
1.2.2 - Objetivos Específicos.....	19
1.3 - Estrutura do Trabalho.....	19
2- DOMÍNIO DE APLICAÇÃO DA PESQUISA	21
2.1 - NUPÉLIA.....	21
2.2 - A PLANÍCIE DO RIO PARANÁ.....	22
2.3 - ESPÉCIES.....	24
2.4 - FATOR DE CONDIÇÃO.....	26
3- FUNDAMENTAÇÃO TEÓRICA.....	29
3.1 - DISTRIBUIÇÃO DA VARIÁVEL RESPOSTA.....	30
3.2 - MATRIZ DE PLANEJAMENTO (OU MATRIZ DO MODELO).....	34
3.3 - FUNÇÃO DE LIGAÇÃO.....	35
3.4 - O VETOR ESCORE.....	36
3.5 - MATRIZ INFORMAÇÃO.....	37
3.6 - ESCORE DE NEWTON-RAPHSON E FISHER.....	37
3.7 - ESTIMAÇÃO POR MÁXIMA VEROSSIMILHANÇA.....	38
3.8 - TESTES DE HIPÓTESES.....	41
3.9 - VERIFICAÇÃO DA QUALIDADE DO AJUSTE DO MODELO.....	42
3.10- TÉCNICAS DE DIAGNÓSTICO.....	44
4- PROCEDIMENTOS METODOLÓGICOS.....	49
4.1 - POPULAÇÃO.....	50
4.2 - VARIÁVEIS.....	50
4.3 - AMOSTRA E COLETA DE DADOS.....	50
4.4 - ETAPAS PARA A CONSTRUÇÃO DO MODELO.....	52
5- ANÁLISE DOS RESULTADOS.....	54
5.1 - CONSTRUÇÃO DE UM MODELO GERAL.....	54

5.1.1-Distribuição da variável resposta peso total (W_t).....	54
5.1.2-Identificação da função de ligação.....	55
5.1.3-A construção de um modelo.....	56
5.2 - CONSTRUÇÃO DE MODELOS POR ESPÉCIES.....	71
5.2.1-Distribuição da variável resposta peso total (W_t).....	72
5.2.2-Identificação da função de ligação.....	74
5.2.3-Construção do modelo.....	74
5.3 - COMPARAÇÃO ENTRE OS MODELOS.....	88
5.4 - FATOR DE CONDIÇÃO DO PEIXE.....	92
6- CONCLUSÃO.....	96
6.1 - RESULTADOS OBTIDOS.....	96
6.2 - LIMITAÇÕES DA PESQUISA.....	98
6.3 - SUGESTÕES PARA NOVAS PESQUISAS.....	99
REFERÊNCIAS.....	100
APÊNDICES.....	104
APÊNDICE A: ESTIMAÇÃO DE PARÂMETROS.....	104
APÊNDICE B: PARAMETRIZAÇÃO DO MODELO GERAL	108
APÊNDICE C: EXEMPLOS PARA O FATOR DE CONDIÇÃO DO PEIXE	111

LISTA DE TABELAS

Tabela 3.1- Distribuição das variáveis respostas.....	32
Tabela 3.2- Representação dos termos de um Modelo Linear Generalizado.....	35
Tabela 3.3- Ligações canônicas para as principais distribuições de probabilidade.....	36
Tabela 3.4- Funções de verossimilhança expressas em função da média e parâmetro de dispersão.....	40
Tabela 3.5- Deviance para algumas distribuições de probabilidade.....	43
Tabela 3.6- Estimativa pela máxima verossimilhança do parâmetro de escala ϕ	44
Tabela 5.1- Medidas descritivas da variável peso total (W_t).....	55
Tabela 5.2- Deviances para o modelo Gama com e sem pontos atípicos.....	57
Tabela 5.3- Parâmetros estimados para o modelo Gama – Geral.....	58
Tabela 5.4- Teste tipo 1 para os parâmetros do modelo Gama – Geral	58
Tabela 5.5- Deviance para o modelo Normal Inversa com e sem pontos atípicos.....	59
Tabela 5.6- Parâmetros estimados para o modelo Normal Inversa – Geral.....	60
Tabela 5.7- Teste tipo 1 para os parâmetros do modelo Normal Inversa – Geral.....	60
Tabela 5.8- Parâmetros estimados para a função Normal Inversa com ligação Potência (0,35).....	64
Tabela 5.9- Parâmetros estimados para a função Normal Inversa com ligação Potência (0,35), desconsiderando o fator sexo.....	64
Tabela 5.10- Teste tipo 1 para os parâmetros do modelo Normal Inversa–Geral desconsiderando o fator sexo.....	64
Tabela 5.11- Valores das deviances do modelo inicial e do modelo acrescido da covariável η^2	65
Tabela 5.12- Equações para o peso estimado pelo modelo Gama e fator de condição do peixe correspondente.....	68
Tabela 5.13- Medidas descritivas para o fator de condição determinado.....	69
Tabela 5.14- Medidas descritivas para o fator de condição do peixe determinado para cada espécie classificada por sexo.....	69

Tabela 5.15- Medidas descritivas para as espécies pesquisadas	73
Tabela 5.16- Correlações da variável W_t com as variáveis preditoras.....	73
Tabela 5.17- Deviances dos modelos propostos para cada espécie.....	75
Tabela 5.18- Parâmetros estimados para o modelo proposto para a espécie A, C, D, H e P.....	75
Tabela 5.19- Deviance dos modelos das espécies C, H e P sem os pontos atípicos.....	78
Tabela 5.20- Parâmetros estimados e seus respectivos intervalos de confiança segundo modelo proposto para cada espécie classificada segundo o sexo do peixe.....	78
Tabela 5.21- Teste tipo 1 para os parâmetros estimados através dos modelos propostos para as espécies C, H e P	79
Tabela 5.22- Valores das deviances do modelo inicial e acrescido da covariável $\hat{\eta}^2$	85
Tabela 5.23- Equações para o modelo desenvolvido para o peso das espécies dos peixes	87
Tabela 5.24 Parâmetros estimados segundo o modelo $W_t = aL_s^b$	89
Tabela 5.25 Parâmetros estimados segundo o modelo $W_t = aL_s^3$	90
Tabela 2.27 Parâmetros estimados segundo o modelo linear.....	90
Tabela 2.28 Percentuais das estimativas contidas nos intervalos dos erros especificados.....	91
Tabela 5.29- Fator de condição K do peixe à partir do modelo construído	92
Tabela 5.30- Medidas descritivas do fator de condição do peixe, K.....	94
Tabela C.1- Fator de condição K para o peixe considerando comprimentos do peixe com pesos distintos.....	114

LISTA DE FIGURAS

Figura 1.1	Fluxograma da estrutura do trabalho	19
Figura 2.1	<i>Trachydoras paraguayensis</i>	24
Figura 2.2	<i>Liposarcus anisiti</i>	24
Figura 2.3	<i>Raphiodon vulpinus</i>	25
Figura 2.4	<i>Serrasalmus marginatus</i>	25
Figura 2.5	<i>Leporinus friderice</i>	26
Figura 4.1	Mapa da área de amostragem	48
Figura 4.2	Fluxograma do processo de construção do Fator de Condição do Peixe, K	53
Figura 5.1-	Histograma da variável peso total	54
Figura 5.2-	Dispersão entre as variáveis comprimento padrão (L_s) e peso total (W_t)	55
Figura 5.3-	Diagrama de dispersão entre as variáveis comprimento e peso (a) com total de observações coletadas e (b) excluindo-se os pontos atípicos ao modelo estudado	58
Figura 5.4-	Dispersão entre as variáveis ajustadas e seus respectivos resíduos ordinários segundo o modelo (a) Gama e (b) Normal inversa	61
Figura 5.5-	Dispersão entre as variáveis ajustadas transformadas e seus respectivos resíduos padronizados das deviances para os modelos (a) Gama e (b) Normal Inversa	61
Figura 5.6-	Dispersão entre a ordem das observações e os resíduos padronizados.....	62
Figura 5.7-	Dispersão dos valores ajustados pelos valores observados (a) Modelo Gama e (b) Modelo Normal Inversa	62
Figura 5.8-	Gráficos dos valores ajustados pelos valores previstos para os dois modelos Normal Inversa em estudo	63
Figura 5.9-	Representação gráfica dos resíduos ordinários ordenados para (a) Modelo Gama e (b) Modelo Normal Inversa	65
Figura 5.10-	Sobreposição dos pontos ajustados pelos modelos aos valores observados, ordenados de forma crescente	66
Figura 5.11-	Plotagem dos valores observados contra a distribuição de probabilidade teórica	66
Figura 5.12-	Histograma para o fator de condição para os peixes da espécie A, C, H e P	68

Figura 5.13-Fator de condição para cada espécie identificadas de acordo com seu sexo	70
Figura 5.14-Plotagem dos valores observados contra os valores esperados pela probabilidade normal.....	71
Figura 5.15-Histogramas da variável peso total para as espécies em estudo	72
Figura 5.16-Dispersão entre a variável peso total e comprimento padrão para a espécie A, <i>Trachydoras paraguayensis</i> . Espécie C, <i>Liposarcus anisiti</i> . Espécie D, <i>Raphiodon vulpinus</i> . Espécie H, <i>Serrasalmus marginatus</i> . Espécie P, <i>Leporinus friderice</i>	74
Figura 5.17-Diagramas das variáveis comprimento padrão e peso total sem casos atípicos para as espécies C, <i>Liposarcus anisiti</i> ; espécie H, <i>Serrasalmus marginatus</i> ; espécie P, <i>Leporinus friderice</i> ; machos e fêmeas da espécie <i>Trachydoras paraguayensis</i> e machos e fêmeas da espécie <i>Raphiodon vulpinus</i>	77
Figura 5.18-Dispersão entre os valores ajustados e seus resíduos ordinários	80
Figura 5.19-Gráficos dos valores ajustados pelos resíduos padronizados da deviance	81
Figura 5.20-Dispersão dos valores ajustados transformados e resíduos padronizados da deviance	82
Figura 5.21-Dispersão entre a ordem das observações e os resíduos padronizados da deviance	83
Figura 5.22-Dispersão entre os valores ajustados e valores observados	84
Figura 5.23-Distribuição dos resíduos ordinários em ordem crescente	85
Figura 5.24-Sobreposição dos valores ajustados aos valores observados contra sua ordem crescente.....	88
Figura 5.25-Distribuição dos fatores de condição para os peixes coletados	93
Figura 5.26-Plotagem da probabilidade normal para a condição do peixe	94
Figura C.1- Sobreposição dos valores ajustados aos valores observados.....	112

1- INTRODUÇÃO

O século XX definiu-se como um marco na história da humanidade; período caracterizado por avanços fantásticos em quase todas as áreas do conhecimento humano. Entretanto, mesmo o homem tendo consolidado seu poder de atuação, os resultados obtidos nem sempre foram revertidos em benefício social ou mesmo individual: o impacto do crescimento econômico acelerado e o aumento populacional trouxeram perdas irreparáveis ao meio ambiente. Para atender às necessidades modernas, metade da superfície terrestre foi transformada. Cerca de 70% das áreas florestais da Europa e cerca de um quarto das espécies de aves da Ásia foi extinta. A cobertura vegetal total perdeu 10% de sua área neste período e as taxas de desmatamento ainda são alarmantes. Segundo projeções da ONU (Organização das Nações Unidas), para o ano de 2025, dois terços dos habitantes da terra terão algum tipo de dificuldade para conseguir água com regularidade.

Poluição, desmatamento e erosão têm colaborado diretamente para a diminuição da quantidade de água para o consumo humano. Além da questão da água potável, os ecossistemas formados por rios e lagos de água doce embora ocupem apenas 1% da superfície terrestre, concentram aproximadamente 40% das espécies de peixes e 12% dos demais animais e de acordo com o relatório Planeta Vivo do Fundo Mundial para a Natureza, 34% das espécies de peixes de água doce estão em risco de extinção. Todas estas evidências alarmantes fazem com que este ecossistema seja considerado o meio de vida natural mais ameaçado do nosso planeta e conseqüentemente, o principal problema ambiental deste milênio.

Para o World Resources Institute (WRI), as maiores ameaças à manutenção da vida nos rios e lagos são a construção de represas e a canalização de rios. De 1950 até o ano de 2000, o número de grandes barragens cresceu de 5.270 para mais de 36.500. Um exemplo do impacto destas obras sobre o meio ambiente está na construção da represa Pak Mum, na Tailândia, no início dos anos 90, que levou à extinção aproximadamente 150 espécies de peixes do rio Mum (Almanaque Abril, 2000).

Com a conscientização da gravidade do problema, muitos trabalhos estão sendo realizados buscando, não só alternativas e soluções, mas também a compreensão do processo que desencadeou e sustenta esta situação catastrófica. É com esta visão de pesquisa e análise como instrumento de preservação do meio ambiente que no ano de 1983, um grupo de pesquisadores da Universidade Estadual de Maringá, iniciou um projeto de estudo sobre a população de peixes do reservatório de Itaipu, financiado pela Itaipu Binacional com o objetivo de avaliar o impacto da construção da usina sobre a ictiofauna, bem como subsidiar medidas mitigadoras e de manejo pesqueiros. Este projeto inicial consolidou-se e expandiu sua abrangência dos estudos limnológicos, estendendo-se além dos reservatórios para pesquisas em planícies alagáveis. Em 1986, este grupo foi reconhecido como Núcleo multidisciplinar de pesquisa dentro da UEM, passando a denominar-se Nupélia, *Núcleo de Pesquisas em Limnologia, Ictiologia e Aqüicultura*.

Os dados trabalhados nesta pesquisa referem-se ao período de 1986 até 2001, coletados na planície de inundação do rio Paraná sob orientação e execução do *Núcleo de Pesquisas em Limnologia, Ictiologia e Aquicultura*, o *Nupélia*.

A contribuição deste trabalho é a construção de um modelo pela aplicação da teoria dos Modelos Lineares Generalizados (MLG), capaz de estimar o peso do peixe tendo como variáveis preditoras sua espécie, comprimento e sexo. Por meio deste modelo busca-se identificar uma relação peso-comprimento considerando apenas as variáveis que inferem significativamente na estimação dos parâmetros relevantes para a determinação de um *fator de condição* do peixe. Segundo Agostinho e Gomes (1997), o *fator de condição* possibilita inferências acerca da condição do indivíduo em relação ao seu ambiente, por ser capaz de refletir a ação de fatores biótipos e abiótipos. A relação peso-comprimento do peixe é aceita no meio científico com um indicador útil e prático no auxílio da identificação e avaliação geral do “bem-estar” da espécie.

Existem vários procedimentos disponíveis para a determinação do fator de condição do peixe. Bolger e Connolly, em artigo publicado pela *The Fisheries Society of the British Isles* (1989), faz uma avaliação de vários desses

indicadores de condição e afirma que a variedade desses índices bem como a multiplicidade de procedimentos estatísticos não se deve apenas à facilidade do uso dos softwares que possibilitam cálculos sofisticados em curto espaço de tempo, mas à diversidade de opiniões sobre os mais satisfatórios e às falhas nas análises dos pressupostos básicos inerentes à construção estatística desses índices.

1.1 PROBLEMA

Uma característica inerente ao ser humano é a sua capacidade de descrever, compreender e definir teorias para que, a partir de observações sistematizadas dos fatos, ser capaz de deduzir, inferir e até mesmo prever resultados através de procedimentos que minimizem erros. Com a evolução tecnológica, estes procedimentos tornam-se cada vez mais eficientes e sofisticados, rompendo fronteiras e desafiando o potencial criativo individual e coletivo de pesquisadores.

Este trabalho desenvolve conceitos teóricos e práticos sobre Modelos Lineares Generalizados (MLG), uma classe de modelos desenvolvida por Nelder e Wedderburn em 1972 que se caracteriza por ser mais abrangente que os modelos lineares clássicos. É esta teoria em parceria com observação dos dados levantados pelo Nupélia que incitou reflexões sobre a seguinte situação problema:

- É possível construir um modelo com princípios lineares para uma variável dependente, cuja distribuição é conhecida não-normal e variância não apresente homoscedasticidade?
- Pode-se, através deste ajuste, encontrar uma aplicação real de dados que auxiliem na obtenção de índices capazes de avaliar condições físicas e ambientais em estudos limnológicos?

Portanto, esta pesquisa justifica-se não apenas pela aplicabilidade de conceitos estatísticos em estudos ictiológicos, mas em especial por representar esforço conjunto de áreas distintas do conhecimento, cujo propósito é o de definir instrumentos e procedimentos que auxiliem na

preservação das espécies de peixes de água doce e, conseqüentemente, monitorar o ecossistema ao qual as espécies pertencem.

1.2 OBJETIVOS

Esta pesquisa está centrada em determinar um fator de condição do peixe considerando-se sua espécie, sexo e comprimento. Sabe-se que, em geral, os fatores de condição do peixe encontrados como referenciais na literatura disponível referem-se a resultados obtidos em pesquisas não compatíveis com o ambiente tropical e com o histórico de barramentos como é o caso desta pesquisa. Segundo Vazzoler *et al.* (1997), as alterações hidrológicas que as barragens hidrelétricas produzem, particularmente as relacionadas à atenuação e ao retardamento dos picos de cheias e aos pulsos de vazão determinados por exigências operacionais, têm considerável impacto sobre a ictiofauna de jusante.

Os modelos desenvolvidos para a relação peso-comprimento em geral são uma aplicação dos modelos lineares clássicos, que freqüentemente oferecem alguma restrição quanto aos seus pré-supostos básicos: homocedasticidade da variância e distribuição normal dos erros.

De acordo com Dobson (2002, p.43), uma das principais vantagens de se trabalhar com os Modelos Lineares Generalizados é que “muitas das melhores propriedades da distribuição normal são compartilhadas por uma classe mais ampla de distribuição chamada família de distribuição exponencial”. Portanto, os objetivos estarão fundados numa aplicação teórica dos Modelos Lineares Generalizados em uma base de dados reais que possibilite a determinação de um fator de condição do peixe para as espécies *Trachydora paraguayensis*, *Liposarcus anisiti*, *Raphiodon vulpinus*, *Serrasalmus marginatus* e *Leporinus friderice*, coletadas na planície de inundação do Alto Rio Paraná.

1.2.1 Objetivo Geral

Desenvolver um fator de condição do peixe utilizando a teoria dos

Modelos Lineares Generalizados.

1.2.2 Objetivos Específicos

- Verificar a significância dos fatores espécie e sexo para a construção do modelo.
- Validar os modelos construídos comparando-os a outros modelos aceitos no meio científico como adequados para definir a relação peso-comprimento do peixe.
- Estimar o fator de condição do peixe segundo o modelo especificado pela relação peso-comprimento.

1.3 ESTRUTURA DO TRABALHO

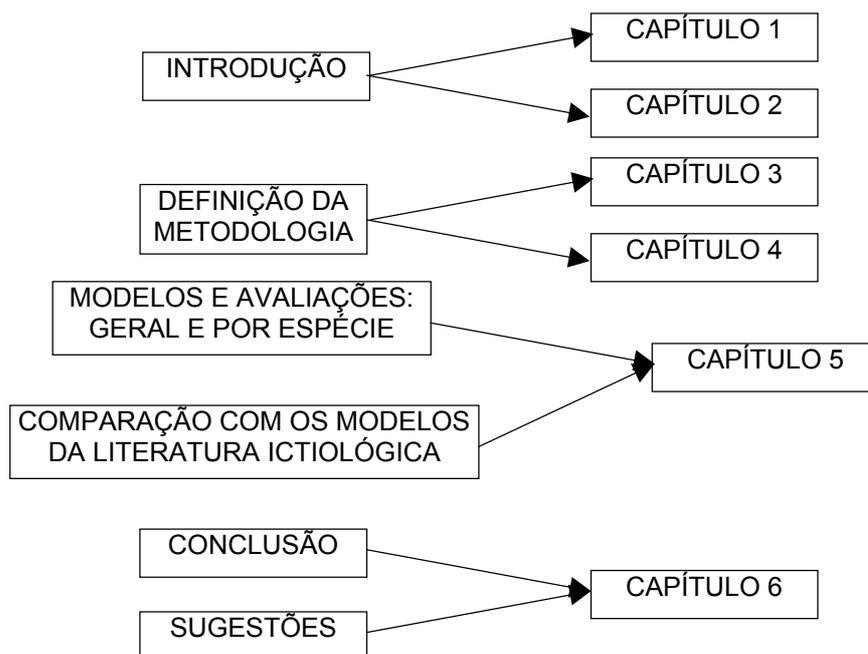


Figura 1.1 Fluxograma da estrutura do trabalho

A estrutura deste trabalho, na sua parte introdutória, é composta pelos capítulos 1 e 2. No Capítulo 1 define-se o problema a ser abordado e os objetivos a serem atingidos. O Capítulo 2 apresenta a estrutura básica do *Núcleo de Pesquisas em Limnologia, Ictiologia e Aqüicultura*, da Universidade Estadual de Maringá, informações sobre o Rio Paraná, as características

fundamentais das espécies estudadas e introduz-se a conceitualização de um *fator de condição K do peixe*.

A metodologia também compreende dois capítulos. No Capítulo 3 apresenta-se a fundamentação teórica da modelagem que será desenvolvida, enquanto que o Capítulo 4 expõe as principais características da população, o processo utilizado na amostragem e as etapas básicas desenvolvidas para construção de um Modelo Linear Generalizado.

No Capítulo 5 desenvolve-se a análise dos dados, num primeiro momento construindo-se um modelo geral para as espécies de peixes em estudo e, após verificar-se a significância das espécies, reconstrói-se os modelos avaliando sua qualidade, e então comparando-os a alguns outros modelos citados em trabalhos Ictiológicos. Tais procedimentos foram realizados com auxílio de softwares próprios.

Finalmente, no Capítulo 6 são expostas as conclusões obtidas através da pesquisa a partir dos quais sugere-se futuras pesquisas.

2- DOMÍNIO DE APLICAÇÃO DA PESQUISA

O Capítulo 2 oferece a caracterização ou o domínio de aplicação para a compreensão desta pesquisa. Nele é descrito o núcleo de pesquisa responsável pela coleta e identificação dos dados, o local onde a apreensão dos espécimes foi realizada, bem como a identificação e caracterização de cada uma das espécies trabalhadas. Por último, define-se o fator de condição do peixe.

2.1 NUPÉLIA

O Nupélia, *Núcleo de Pesquisas em Limnologia, Ictiologia e Aqüicultura*, da Universidade Estadual de Maringá, originou-se no ano de 1983, quando um grupo de docentes, técnicos e graduados da Universidade Estadual de Maringá - UEM, iniciou estudos sobre ecologia de populações de peixes do reservatório de Itaipu, projeto financiado pela Itaipu Binacional cujo objetivo era avaliar o impacto da construção da usina sobre a ictiofauna, bem como subsidiar medidas mitigadoras e de manejo pesqueiro. Gradativamente, a estes pesquisadores agregaram-se pesquisadores de outras áreas, organizando-se uma infra-estrutura de apoio e expandindo a abrangência dos estudos limnológicos, até que, em setembro de 1986 o Núcleo foi reconhecido formalmente como um Núcleo Multidisciplinar de Pesquisa dentro da UEM. Atualmente ele conta com 34 pesquisadores nas áreas de limnologia química e física, bacterioplâncton, perifíton, macrófitas aquática, fitoplâncton, zooplâncton, zoobentos, mata ciliar e peixes (autecologia, ecologia de populações e comunidades, genética e parasitologia), que atuam em reservatórios, rios, riachos, lagoas permanentes e temporárias da bacia do rio Paraná.

A estrutura física do Nupélia compõe-se de 3.100m² de edificações, uma base avançada de pesquisas nas margens do rio Paraná, 18 laboratórios nas várias áreas de sua atuação, biblioteca especializada, processamento de dados, artes gráficas, veículos e embarcações. Além das atividades de

pesquisa, também se dedica à formação de recursos humanos, mantendo uma visão sistêmica sobre questões ambientais, oferecendo estágios, cursos de especializações, mestrado e doutorado.

Este núcleo de pesquisa tem uma forte atuação na comunidade científica nacional e internacional, mantendo convênios e intercâmbios de informações, estagiários e pesquisadores com diversas instituições. Sua inserção regional também é relevante, atuando em conjunto com o setor produtivo, particularmente aqueles relacionados ao uso dos recursos hídricos, na solução de problemas ambientais e da pesca. Dedicase também em decodificar a linguagem científica de suas publicações e repassá-las através de jornais e programas de radiodifusão às populações ribeirinhas, especialmente para os pescadores.

Além de sua atuação junto às concessionárias hidrelétricas (Itaipu Binacional, Copel, Cesp, e Furnas), tem participado na elaboração de diretrizes de ação de manejo de recursos aquáticos, em especial dos pesqueiros e da pesca para a Eletrobrás, IBAMA, Secretarias Estaduais de Meio Ambiente e organizações não governamentais.

2.2 A PLANÍCIE DO RIO PARANÁ

A bacia do Prata, composta pelos rios Paraguai, Paraná e Uruguai é a segunda maior da América do Sul, responsável pela drenagem de uma área total de aproximadamente 3.100.000 km², o que corresponde a aproximadamente 10,5% do território brasileiro. Seu principal rio é o Paraná, devido à magnitude de sua descarga, à extensão de sua área tributária, à extensão de seu curso e pelos diversos aspectos dimensionais que caracterizam sua bacia fluvial, e cujos principais tributários são os rios Grande e o Paranaíba (formadores), Tietê, Paranapanema e Iguçu.

O rio Paraná ocupa a décima posição mundial em descarga e a quarta em drenagem. Estende-se por todo o centro-sul, desde as encostas dos Andes até a Serra do Mar, nas proximidades da costa atlântica, banhando uma região de aproximadamente 891.000 km² do sudeste brasileiro, área com a

maior densidade demográfica do país.

A ação do homem sobre a vegetação natural da região tem afetado seriamente o ambiente aquático desta bacia: aporte de elevadas cargas de biocidas e nutrientes decorrentes de esgotos domésticos e da atividade agrícola, desmatamento da vegetação riparia e, principalmente, construção de barragens, as quais têm alterado o regime natural da cheia e suprimido trechos lóticos e amplas áreas alagáveis do rio Paraná e de seus principais afluentes. Grandes extensões de seus principais afluentes do trecho superior do rio Paraná já são consideradas impróprias para o ser humano e para a vida aquática.

Neste contexto, merecem destaque os rios Grande, Tietê e Paranapanema que, além de receberem elevadas quantidades de compostos químicos das atividades agrícolas e urbanas, foram virtualmente transformados em cascatas de reservatórios a partir da década de 60. Porém, os afluentes que desembocam diretamente na área da planície que drenam terrenos dos Estados do Mato Grosso do Sul e Paraná, também apresentam elevado grau de degradação em virtude de aportes de esgotos, biocidas e fertilizantes agrícolas. Estes rios ainda não se encontram represados como aqueles dos estados de Minas Gerais e São Paulo.

Existem mais de 130 barragens na bacia (considerando-se apenas aquelas com alturas superiores a 10m); dos 809 km originais do rio, somente 230 ainda são de água corrente. Este último trecho não represado do rio Paraná apresenta amplo canal, ora com uma extensa planície fluvial com pequenas ilhas (mais de 300), ora com grandes ilhas e uma planície alagável mais restrita. A planície apresenta numerosos canais secundários e lagoas. As flutuações dos níveis da água, mesmo que prolongadas pelos represamentos, ainda mantêm a sazonalidade e uma amplitude média de cinco metros. Este remanescente de várzea tem importância fundamental na manutenção das espécies de peixes já eliminadas dos trechos superiores da bacia, especialmente espécies de grande porte que realizam extensas migrações reprodutivas.

Essa situação, preocupante sob a ótica ambiental, faz com que a Planície do Alto Rio Paraná, assuma importância capital como um remanescente

ainda relativamente bem conservado de áreas alagáveis do rio Paraná. Deve-se enfatizar que esta planície representa o último trecho deste rio, em território brasileiro, onde ainda existe um ecossistema do tipo “rio-planície de inundação”(Junk *et al.*, 1989 e Neiff, 1990, citados por Vazzoler *et all.*, 1997).

Em face de sua importância como local representativo de um ecossistema rio planície de inundação, todo o remanescente ainda não represado da Planície de Inundação do Alto Rio Paraná, que se estende desde a foz do rio Paranapanema até o início do reservatório de Itaipu, onde se posicionavam as antigas e extintas Sete Quedas, foi recentemente transformado na Área de Proteção Ambiental das Ilhas e Várzeas do Rio Paraná, criada pelo Governo Federal. A metade inferior desta área pertence ao Parque Nacional de Ilha Grande e a foz do rio Ivinheima, também nela incluída, pertence ao Parque Estadual do Ivinheima.

2.3 ESPÉCIES

As espécies envolvidas neste estudo são identificadas a seguir.



Nome comum: Armadinho
Família: *Doradidae*
Ordem: *Siluriformes*
Classe: *Actinopterygii*

Figura 2.1: *Trachydoras paraguayensis*

Peixe inofensivo, de pequeno porte, encontrado basicamente no rio Paraná. Dessa espécie trabalhou-se inicialmente com 3117 indivíduos distribuídos entre 1665 fêmeas e 1452 machos. Para estes indivíduos, o peso mínimo e máximo foram respectivamente 3,16g e 74,73g, enquanto o comprimento variou de 4,70cm a 12,70cm (Figura 2.1).

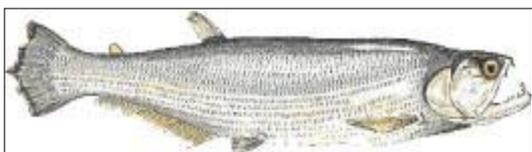


Nome comum: Cascudo
Família: *Loricariidae*
Ordem: *Siluriformes*
Classe: *Actinopterygii*

Figura 2.2: *Liposarcus anisiti*

Peixe de corpo cilíndrico e coberto por cutícula dura, de placas ósseas apresentando pequenos espinhos entre estas placas. Sua barriga é lisa e sem as placas ósseas (Figura 2.2). Alimenta-se de raspas de algas e detritos depositados sobre a vegetação alagada no período de enchentes. Reproduz-se no leito dos rios, em época das cheias e produzem ovos grandes que são depositados em locais de barrancos de rios. Sua carne é bastante apreciada pelos nativos ribeirinhos.

Foram trabalhados 367 indivíduos da espécie, sendo 200 fêmeas e 167 machos. O peso e o comprimento mínimo da amostra foi de 215,12g e 18,60cm, enquanto o peso e o comprimento máximo foi de 1.791,20g e 44,40cm, respectivamente.



Nome comum: Cachorra, peixe-cachorra, pirandira
Família: *Cynodontidae*
Ordem:
Classe:

Figura 2.3: *Raphiodon vulpinus*

Apresenta pequenas escamas, corpo alongado e comprido com boca grande e oblíqua (Figura 2.3). Possui dentes caninos com mandíbula apresentando um par de presas que se encaixa no maxilar superior. A nadadeira dorsal localizada na metade posterior do corpo, na mesma direção da anal e nadadeiras peitorais longas. Sua coloração é prata uniforme, mais escura na região dorsal. É um peixe de características piscívora, com reprodução entre os meses de novembro a março. Por possuir muita espinha e pouca carne, não possui grande valor comercial.

A amostra desta espécie compõe-se de 2161 indivíduos, dos quais 699 são fêmeas e 1462 machos. O peso e o comprimento mínimo da amostra foi de 15,90g e 14,70cm, enquanto o peso e o comprimento máximo foi de 3000,00g e 76,20cm, respectivamente.



Nome comum: – Piranha
Família: *Characidae*
Ordem: *Characiformes*
Classe: *Poissons*

Figura 2.4: *Serrasalmus marginatus*

Peixes carnívoro para o qual o odor de sangue, que captam à distância, exerce uma poderosa atração e desperta seus instintos. Após atacarem suas vítimas, em pouco tempo destroem e comem sua carne, restando apenas o esqueleto. Sua coloração é amarela dourada com escamas pequenas e brilhantes. Possuem grandes olhos e uma poderosa mandíbula, sobressaindo a inferior com dentes triangulares e afiados. É um peixe de águas quentes. É considerado um grande regulador da fauna ictica.

Amostrou-se 8340 casos desta espécie de peixe: 4179 fêmeas e 4161 machos, cujo peso mínimo foi de 1,10g e o máximo de 538,00g, enquanto que o comprimento mínimo foi de 4,30cm e o máximo de 26,20cm.



Nome comum: Piau
Família: *Anostomidae*
Ordem: *Characiformes*
Classe: *Actinopterygii*

Figura 2.5: *Leporinus friderice*

Possui escamas, corpo alongado e fusiforme com boca terminal um pouco inferior, com dentes incisivos e sem cúspides. Sua coloração é cinza (Figura 2.5). É um peixe onívoro, com tendência carnívora (principalmente insetos) ou frugívora (pequenas frutas e sementes). Vive principalmente nas margens dos rios, lagos e florestas inundadas, destacando-se como pesca de subsistência.

O tamanho amostrado para esta espécie foi de 1190 casos, distribuídos entre 802 fêmeas e 388 machos. A variação observada para o peso desta amostra é de 9,70g a 625,10g, enquanto que o comprimento é de 8,00cm a 30,60cm.

2.4 FATOR DE CONDIÇÃO

A relação peso-comprimento do peixe é aceita no meio científico com um indicador útil e prático no auxílio da identificação e avaliação geral do “bem-estar” da espécie. Tal consenso deve-se ao fato de que para um certo comprimento do peixe, quanto maior for seu peso, melhor deve ser sua

condição de sobrevivência. Segundo Agostinho (1997, p.144), “a estreita correlação entre o peso e comprimento descreve características estruturais dos indivíduos da população, fornecendo informações muito importantes acerca da população de determinada espécie”. O uso dos parâmetros da relação peso-comprimento na análise do bem-estar geral dos peixes, em pesquisas ictiológicas, é feita basicamente sob três modalidades: o fator de condição, a condição de Fulton e o peso relativo.

A variação do peso do indivíduo em relação ao seu comprimento é chamado de fator de condição e em geral é ajustado pela equação:

$$W = a \cdot L^b \Rightarrow K = \frac{W}{L^b} \quad (2.1)$$

onde K é o fator de condição do peixe, L é o comprimento do peixe e W o peso total do animal. Segundo Santos, 1978, Anderson e Gutreuter, 1992, citados por Agostinho e Gomes (1997, p.114) “a constante a indica o estado nutricional do peixe, cujo valor a pode variar em função do local e estágio gonadal”. O parâmetro b representa o tipo de incremento e pode assumir valores próprios de acordo com a espécie.

Experimentalmente, observa-se que o valor do parâmetro b em geral assume valores próximos a 3,0. Então, a relação (2,1) passa a ser definida como fator de condição de Fulton:

$$K = \frac{W}{L^3} \quad (2.2)$$

Segundo Gulland, Weatherley e Gill, citados por Agostinho e Gomes(1997, p.114) se o valor de b for bem próximo de 3,0 o crescimento é definido como isométrico, ou seja, as taxas de incremento das diferentes partes do corpo do peixe são iguais. Entretanto, segundo Agostinho e Gomes (1997, p.114), tal comportamento não é constante, pois muitas espécies mudam a forma com o crescimento corpóreo (incremento alométrico) e não há teoria que informe em quais casos os valores estimados de b podem ser esperados abaixo de 3,0 (alometria negativa) ou acima de 3,0 (alometria positiva). Contudo, admite-se que quando $b > 3$ o peixe está aumentando em peso e, supostamente, em volume, numa taxa maior do que a necessária para manter constantes as proporções corpóreas.

O peso relativo W_r determina o quanto o peso total (W_t) do peixe representa do valor do peso esperado (\hat{W}). Ou seja, pode-se definir o peso relativo pela equação:

$$W_r = \frac{W_t}{\hat{W}} \quad (2.3)$$

A utilização do fator de condição justifica-se por possibilitar inferências acerca da condição do indivíduo em relação ao seu ambiente, por refletir a ação conjunta de fatores biótipos e abiótipos. Agostinho (1997, p.131)

Pela aplicação da teoria dos Modelos Lineares Generalizados, busca-se um modelo que represente a relação peso-comprimento e, por meio deste modelo, define-se um fator de condição para o peixe, considerando as variáveis que são relevantes na construção do modelo.

3- MODELOS LINEARES GENERALIZADOS (MLG)

Ao procurar-se um bom modelo linear, capaz de relacionar satisfatoriamente a variável resposta Y com as variáveis preditoras X_1, X_2, \dots, X_p , é comum deparar-se com violações das hipóteses básicas para a construção do modelo, ou seja, os dados não apresentarem normalidade e tão pouco a variância constante. Neste caso, um procedimento comum é aplicar-se uma transformação, como por exemplo, a de Box-Cox, cujo objetivo é a resolução simultânea do problema: encontrar uma constante λ que produza a linearidade bem como a constância da variância. Nelder e Wedderburn em 1972 demonstraram a ineficiência de tal procedimento em algumas situações práticas e, no mesmo trabalho, desenvolvem uma classe de modelos os quais se distinguem dos modelos lineares clássicos basicamente pelo fato de que "os dados não precisam necessariamente apresentar uma distribuição normal e a média não é obrigatoriamente uma combinação linear dos parâmetros, mas sim uma função da média" (McCulloch e Searle, 2001, p.2.). São os modelos lineares generalizados.

De acordo com Cordeiro (1986, p.1), o Modelo Linear Generalizado, é definido por:

... uma distribuição de probabilidade, membro da família exponencial de distribuições, para a variável resposta, um conjunto de variáveis independentes descrevendo a estrutura linear do modelo e uma função de ligação entre a média da variável resposta e a estrutura linear.

Demétrio (2001, p.25), afirma que estes modelos "podem ser usados quando se tem uma única variável aleatória Y e associado a ela um conjunto de variáveis explicativas X_1, X_2, \dots, X_p " e que para "uma amostra de n observações (y_i, x_i) em que $x_i = (x_{1i}, x_{2i}, \dots, x_{pi})^T$ é o vetor coluna de variáveis explicativas". Estes modelos, ainda de acordo com Demétrio (2001, p.25), envolvem três componentes:

"i) **Componente aleatório**: representado por um conjunto de variáveis aleatórias independentes Y_1, Y_2, \dots, Y_k , cuja

distribuição faz parte da família exponencial na forma canônica com médias $\mu_1, \mu_2, \dots, \mu_k$. Ou seja:

$$E(Y_i) = \mu_i, \text{ onde } i=1, 2, \dots, n$$

ii) **Componente sistemático:** as variáveis explicativas entram na forma de uma soma de seus efeitos

$$\eta_i = \sum_{j=1}^p x_{ij} \beta_j = x_i^T \boldsymbol{\beta} \text{ ou } \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta},$$

sendo $\mathbf{X} = (x_1, x_2, \dots, x_n)^T$ a matriz do modelo, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ o vetor de parâmetros e $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^T$ o preditor linear.

iii) **Função de ligação:** uma função que liga o componente aleatório ao componente sistemático, ou seja, relaciona a média ao preditor linear, isto é,

$$\eta_i = g(\mu_i).$$

sendo $g(\cdot)$ uma função monótona e derivável.

É interessante ressaltar que a variável resposta pode variar até mesmo entre unidades idênticas das covariáveis (seção 3.2), justificando-se seu tratamento como uma variável aleatória. Conceitualmente, Y_i é trabalhada como um único resultado dos vários resultados possíveis, observados sob as mesmas circunstâncias, e estes resultados, em termos de uma distribuição de probabilidade pertencente à família exponencial (RODRIGUEZ, 2001, p.3).

Assume-se que a resposta esperada depende dos preditores $x_{i1}, x_{i2}, \dots, x_{ip}$. Então, no componente sistemático, a matriz do modelo \mathbf{X} , de ordem $n \times p$, com $(p < n)$, com posto p , “pode representar a presença ou ausência de um nível de um fator classificado em categorias, ou pode ser o valor de uma covariável quantitativa” (Cordeiro, 1986, p.5).

Nos próximos tópicos, os conceitos que necessários para a definição de um Modelo Linear Generalizado serão discutidas mais amplamente.

3.1 DISTRIBUIÇÃO DA VARIÁVEL RESPOSTA

Seja o conjunto de variáveis aleatórias independentes Y_1, Y_2, \dots, Y_n com médias $\mu_1, \mu_2, \dots, \mu_n$, respectivamente, tais que cada Y_i tem distribuição de probabilidade da família exponencial, obedecendo às seguintes propriedades:

- i) as distribuições de todos os Y_i apresentam a mesma forma;
- ii) cada Y_i tem a distribuição de probabilidade na forma canônica dependente de um único parâmetro θ_i . Então, pode-se escrever:

$$f(\mathbf{y}; \theta_i, \phi_i) = \exp \left[\frac{\{\mathbf{y}\theta_i - b(\theta_i)\}}{a(\phi_i)} + c(\mathbf{y}, \phi_i) \right] \quad (3.1)$$

onde $b(\cdot)$ e $c(\cdot)$ são funções conhecidas, θ_i é o parâmetro natural (ou canônico) da família de distribuições exponenciais e, em geral, $a(\phi) = \frac{\phi}{\rho}$, sendo ϕ chamado de parâmetro de dispersão e ρ um peso conhecido que varia de observação para observação.

A média e variância das distribuições de probabilidades da família exponencial são dadas por McCullagh e Nelder (1989, p.29) como:

$$E(Y_i) = \mu_i = b'(\theta_i) \quad (3.2)$$

$$\text{Var}(Y_i) = a(\phi)b''(\theta) = a(\phi)\mathbf{V}(\mu_i) = a(\phi)V_i \quad (3.3)$$

onde $V_i = \frac{d\mu_i}{d\theta_i}$ é freqüentemente chamada de função de variância e indica o quanto a variância de Y_i depende da média de Y_i . Como depende unicamente da média, o parâmetro natural θ_i pode ser expresso como: $\theta_i = \int V_i^{-1} d\mu_i = g(\mu_i)$

Dantas (1998, p.174) observa que "para a escolha de uma distribuição razoável, deve-se examinar os dados principalmente quanto aos seguintes aspectos: assimetria, natureza contínua ou descontínua e intervalo de variação".

As distribuições da família exponencial mais freqüentemente utilizadas para descrever as variáveis respostas de um Modelo Linear Generalizado são Normal, Poisson, Binomial, Gama e Normal Inversa.

Segundo Demétrio (2001, p.13), os identificadores da Família Exponencial para as distribuições citadas são os seguintes:

a) Normal, $N(\mu, \sigma^2)$: $a(\phi) = \sigma^2$, $\theta = \mu$, $b(\theta) = \frac{\theta^2}{2}$; $c(\mathbf{y}; \phi) = -\frac{1}{2} \left[\frac{\mathbf{y}^2}{\sigma^2} + \ln(2\pi\sigma^2) \right]$

$$e \quad \mu(\theta) = \theta.$$

b) Poisson, $P(\mu)$: $a(\phi) = 1$, $\theta = \ln(\mu)$, $b(\theta) = \exp(\theta)$, $c(\mathbf{y}; \phi) = -\ln y$ e $\mu(\theta) = \exp(\theta)$.

c) Binomial, $B(m, \pi)$, com m ensaios independentes e probabilidade de sucesso π : $a(\phi) = 1$, $\theta = \ln\left(\frac{\pi}{1-\pi}\right)$, $b(\theta) = m \cdot \ln(1 + e^\theta)$, e $c(\mathbf{y}; \phi) = \log\binom{m}{\mathbf{y}}$ e $\mu(\theta) = m \frac{\exp(\theta)}{1 + \exp(\theta)}$.

d) Gama, $G(\mu, \nu)$: $a(\phi) = \frac{1}{\nu}$, $\theta = -\frac{1}{\mu}$, $b(\theta) = -\ln(-\theta)$, e $c(\mathbf{y}; \phi) = \nu \cdot \ln(\mathbf{y}) - \ln \mathbf{y} - \ln \Gamma(\nu)$ e $\mu(\theta) = -\frac{1}{\theta}$.

e) Normal Inversa $IG(\mu, \sigma^2)$: $a(\phi) = \sigma^2$, $\theta = -\frac{1}{2\mu^2}$, $b(\theta) = -(-2\theta)^{1/2}$ e $c(\mathbf{y}; \phi) = -\frac{1}{2} \left[\ln(2\pi\sigma^2 \mathbf{y}^3) + \frac{1}{\sigma^2 \mathbf{y}} \right]$ e $\mu(\theta) = (-2\theta)^{-1/2}$.

A Tabela 3.1 apresenta as distribuições de probabilidades destas funções segundo os autores McCullagh e Nelder (1989) e Paula (2001).

Tabela 3.1 – Distribuições das variáveis respostas.

Distribuição e variação de \mathbf{y}	Parâmetro de dispersão ϕ	Função densidade de probabilidade	Valor esperado	Função de variância $V(\mu)$	Var(Y)
Normal - (μ, σ^2) $(-\infty, \infty)$	σ^2	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\mathbf{y}-\mu}{\sigma}\right)^2\right)$	μ	1	σ^2
Poisson - $P(\mu)$ 0, 1, 2, ...	1	$\frac{\mu^{\mathbf{y}} e^{-\mu}}{\mathbf{y}!}$	μ	μ	μ
Binomial - $B(m, \pi)$ 0, 1, ..., m	$1/m$	$\binom{m}{r} \pi^r (1-\pi)^{m-r}$	m	$\pi(1-\pi)$	$\frac{\pi(1-\pi)}{m}$
Gama - $G(\mu, \nu)$ $(0, \infty)$	$1/\nu$	$\frac{1}{\mathbf{y}\Gamma(\nu)} \left(\frac{\mathbf{y}}{\mu}\right)^\nu \exp\left(-\frac{\mathbf{y}}{\mu}\right)$	ν^2	μ^2	$\frac{\mu^2}{\nu}$
Normal Inversa- $IG(\mu, \sigma^2)$ $(0, \infty)$	σ^2	$\frac{1}{\sqrt{2\pi\sigma^3}} \exp\left(-\frac{1}{2\mathbf{y}}\left(\frac{\mathbf{y}-\mu}{\mu\sigma}\right)^2\right)$	μ	μ^3	$\sigma^2\mu^3$

Fonte: Paula, 2001. McCullagh e Nelder, 1989.

O modelo Normal é usado para o caso em que os dados são contínuos e apresentam um intervalo de variação de $-\infty$ a ∞ , “é aplicado na análise de variância com efeitos fixos, como modelo amostral e, mais comumente como

um modelo aproximado para distribuições desconhecidas. É o caso mais simples do modelo linear generalizado” (Cordeiro, 1997, p.13).

Segundo Cordeiro (1986, p.6), o modelo de Poisson é adotado para dados categorizados ou de contagem e, nestes casos, sua importância é equivalente à Normal para os dados contínuos. Suas principais características são: (a) proporciona, em geral, uma descrição satisfatória de dados experimentais cuja variância é proporcional à média; (b) pode ser deduzido teoricamente de princípios elementares com um mínimo de restrições; (c) se eventos ocorrerem independentemente e aleatoriamente no tempo, com taxa média de ocorrências constantes, o modelo determina o número de eventos, num tempo especificado...

“O modelo Binomial é usado especialmente no estudo de dados em forma de proporção, análise "probit" e logística (ou logit) bem como na análise de dados binários e regressão logística-linear” Cordeiro (1986, p.8).

O modelo Gama, segundo Cordeiro (1986, p.11), “é usado na análise de dados contínuos não-negativos que apresentam uma variância crescente com a média e, principalmente, quando o coeficiente da variação dos dados for aproximadamente constante”. Ainda, segundo este mesmo autor, este modelo também é usado na estimação de variâncias da análise de variância com efeitos aleatórios.

O modelo Normal Inversa é aplicado em análise de regressão com dados consideravelmente assimétricos e intervalos de variação positivos, testes de confiabilidade, análise seqüencial e análogo de análise de variância para classificações encaixadas (ou aninhadas). Tem distribuição assintótica normal.

Cordeiro (1997, p.15) sugere que quando a variância dos dados for maior que a média (super-dispersão), pode-se trabalhar com as distribuições Gama, Binomial Negativa ou Normal Inversa. Em caso da variância ser menor que a média (sub-dispersão), o modelo mais indicado é o Binomial.

3.2 MATRIZ DE PLANEJAMENTO (OU MATRIZ DO MODELO)

O componente sistemático do modelo, $\eta = \mathbf{X}\beta$, tem a matriz \mathbf{X} do modelo constituída por variáveis independentes denominadas de variáveis preditoras ou explicativas. Estas variáveis podem ser quantitativas, qualitativas ou uma combinação das mesmas.

As variáveis quantitativas contínuas são denominadas de covariáveis e, em geral, a uma covariável x corresponde um único parâmetro β , contribuindo com o termo $\mathbf{X}\beta$ para o modelo. Sua representação algébrica e no modelo são, respectivamente, $\lambda\mathbf{X}$ e \mathbf{X} . Segundo Wilkison e Rogers, citados por Amaral (1997, p.37), a notação das covariáveis é feita através das últimas letras maiúsculas do alfabeto: X, Y, Z e W.

As variáveis qualitativas são chamadas de fatores, os quais são compostos por categorias denominados de níveis. Um fator A de k níveis, pode ser representado pelo modelo $\eta = \alpha_i$, onde $i = 1, 2, \dots, k$. Sua representação no modelo é A. De maneira geral, a notação dos fatores é feita através das primeiras letras maiúsculas do alfabeto: A, B, C, ... enquanto os níveis são representados pelas letras minúsculas.

De acordo com Amaral (1997, p.37), para ajustar um modelo que possui fatores, é necessário utilizar variáveis indicadoras. Um fator com k níveis pode ser representado por k variáveis indicadoras

$$u_i = \begin{cases} 1, & \text{se ocorre o nível } i \\ 0, & \text{em caso contrário} \end{cases}$$

ou por

$$A = \alpha_1 u_1 + \alpha_2 u_2 + \dots + \alpha_k u_k, \quad (3.5)$$

onde α_i = valor do i-ésimo nível.

Pode ocorrer que um termo de interação entre os fatores pode ser incluído no modelo. Em experimentos fatoriais, onde existe apenas uma observação para cada combinação dos níveis dos fatores, se são colocadas todas as interações, obtém-se o modelo saturado. No caso de duas variáveis contínuas, a interação é obtida por incluir o termo $\beta_{12}x_1x_2$; se as variáveis são fatores, utiliza-se $(\alpha\beta)_{ij}$.

Finalmente, é possível ajustar uma componente que representará o efeito conjunto de um fator e de uma covariável contínua. Então, ao modelo com o fator A e a covariável \mathbf{X} , $\eta = \alpha_i + \beta\mathbf{X}$, ajusta-se o termo $\beta_i\mathbf{X}$ em lugar de $\beta\mathbf{X}$, $\eta = \alpha_i + \beta_i\mathbf{X}$, indicando que a declividade deve variar com o nível do fator, ou seja, a declividade β_i associada à covariável \mathbf{X} é suposta diferente para cada nível do fator A. O termo $\beta_i\mathbf{X}$ é chamado de misto.

Estas representações estão resumidas na tabela 3.3

Tabela 3.2 – Representação dos termos de um Modelo Linear Generalizado.

Tipo do termo	Fórmula algébrica	Fórmula do modelo
Covariável (contínuo)	$\lambda\mathbf{x}$	\mathbf{X}
Fator (quantitativo)	α_i	A
Misto	$\lambda_i\mathbf{x}$	A. \mathbf{X}
Composto	$(\alpha\beta)_{ij}$	A.B
Misto Composto	$\lambda_{ij}\mathbf{x}$	A.B. \mathbf{X}

Fonte: Amaral, 1997, p.38

3.3 FUNÇÃO DE LIGAÇÃO

O componente sistemático do Modelo Linear Generalizado admite a existência de uma função de ligação $g(\cdot)$ entre as médias das observações e a estrutura linear do modelo:

$$E(\mathbf{Y}) = g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} = \beta_0 + \beta_1x_1 + \dots + \beta_px_p, \quad \eta_i = g(\mu_i) \quad \text{e} \quad \mu_i = g^{-1}(\eta_i) \quad (3.4)$$

onde o preditor linear $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^T$, é uma função dos parâmetros desconhecidos $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ e $g(\cdot)$, suposta conhecida, é uma função monótona e diferenciável contínua. Em geral $g(\cdot)$ é não linear e \mathbf{X} é a matriz do modelo, totalmente conhecida e de posto completo (CORDEIRO, 1989).

Quando a função de ligação torna o preditor linear η_i igual ao parâmetro canônico θ , tem-se então uma ligação canônica. "Para cada tipo de distribuição existe uma ligação canônica que torna os efeitos sistemáticos aditivos, na escala dada pela ligação" (Dantas, 1998).

A função de ligação $g(\cdot)$ deve ser compatível com a distribuição proposta para o erro e ser escolhida de tal forma a facilitar a interpretação do modelo.

As ligações canônicas para as distribuições estudadas são apresentadas na tabela a seguir.

Tabela 3.3 – Ligações canônicas para as principais distribuições de probabilidade.

Distribuição	Ligação canônica
Normal	$\eta = \mu$
Poisson	$\eta = \ln(\mu)$
Binomial	$\eta = \ln\left(\frac{\mu}{1-\mu}\right)$
Gama	$\eta = \frac{1}{\mu}$
Normal Inversa	$\eta = \frac{1}{\mu^2}$

Fonte: McCullagh e Nelder, 1989, p.32

Segundo Cordeiro (1986,p.25):

“As ligações canônicas fornecem propriedades estatísticas de interesse ao modelo, tais como, suficiência, facilidade de cálculo, unicidade das estimativas de máxima verossimilhança e em alguns casos, simples interpretação. Entretanto não existe razão de se considerar sempre os efeitos sistemáticos como aditivos, na escala dada pela ligação canônica. A princípio, pode-se trabalhar com a ligação canônica quando não existir indicativos de outra preferível”.

Portanto, sempre que não houver ressalvas, é interessante usar-se as ligações canônicas.

3.4 O vetor escore

A primeira derivada da função de log-verossimilhança (Apêndice A) é chamada *escore de Fisher*, denotada por:

$$\mathbf{U}(\theta) = \frac{\partial \log L(\theta)}{\partial \theta} \quad (3.6)$$

Observe que a função escore é um vetor de primeiras derivadas, um para cada elemento de θ . Se a log-verossimilhança é côncava, pode-se encontrar o estimador de máxima verossimilhança igualando o escore a zero, ou seja, resolvendo o sistema de equações

$$\mathbf{U}(\hat{\theta}) = 0 \quad (3.7)$$

3.5 MATRIZ INFORMAÇÃO

O escore é um vetor aleatório com algumas propriedades estatísticas interessantes. Em particular, o escore avaliado no verdadeiro valor do parâmetro θ tem média zero.

$$E[\mathbf{U}(\theta)] = 0 \quad (3.8)$$

e a matriz de variâncias-covariâncias dada pela matriz de informação:

$$\text{Var}[\boldsymbol{\mu}(\theta)] = E[\boldsymbol{\mu}(\theta)\boldsymbol{\mu}'(\theta)] = \mathbf{I}(\theta) \quad (3.9)$$

Sob condições de regularidade moderada, a matriz informação também pode ser obtida como a esperança negativa do valor para a segunda derivada da log-verossimilhança,

$$\mathbf{I}(\theta) = -E\left[\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'}\right] \quad (3.10)$$

3.6 ESCORE DE NEWTON-RAPHSON E FISHER

Freqüentemente resoluções de equações de máxima-verossimilhança necessitam de procedimentos iterativos. Expandindo a função escore pela série de Taylor até os termos de primeira ordem, com $\hat{\theta}$ em torno do valor θ_0 , tem-se:

$$u(\hat{\theta}) \approx u(\theta_0) + \frac{\partial u(\theta)}{\partial \theta}(\hat{\theta} - \theta_0) \quad (3.11)$$

Denotando-se por \mathbf{H} a segunda derivada da matriz log-verossimilhança (conhecida como função Hessiana):

$$\mathbf{H}(\theta) = \frac{\partial^2 \log L}{\partial \theta \partial \theta'} = \frac{\partial u(\theta)}{\partial \theta} \quad (3.12)$$

e igualando a equação 3.11 a zero, a resolução para um dado $\hat{\theta}$, a aproximação de primeira ordem é:

$$\hat{\theta} = \theta_0 - \mathbf{H}^{-1}(\theta_0)u(\theta_0) \quad (3.13)$$

Este resultado alicerça o processo de aproximação iterativa para a resolução dos estimadores de máxima-verossimilhança, também conhecido como técnica de Newton-Raphson: partindo-se de um dado valor, usa-se a equação (3.13) para obter uma estimativa melhorada, repetindo-se o processo até que a diferença entre as estimativas sucessivas convirjam para zero.

Fisher sugere primeiramente repassarmos $(-H)$ pela matriz de informação resultando em

$$\hat{\theta} = \theta + I^{-1}(\theta_0)u(\theta_0) \quad (3.14)$$

conhecido como escore de Fisher.

3.7 ESTIMAÇÃO POR MÁXIMA VEROSSIMILHANÇA

Para estimar os parâmetros lineares β_1, \dots, β_p e conseqüentemente os preditores lineares η_1, \dots, η_n e os valores médios μ_1, \dots, μ_n , dentre os vários métodos existentes, neste trabalho aplica-se o da máxima verossimilhança.

De acordo com Cordeiro (1992, p.2), "a verossimilhança é interpretada como função do vetor de parâmetros para um conjunto de dados fixos, e serve para medir o quanto aqueles dados suportam uma hipótese sobre θ " e "a verossimilhança expressa a plausibilidade do vetor θ explicar os dados \mathbf{y} ".

Para introduzir a estimativa dos parâmetros β_1, \dots, β_p para os MLGs, convém antes definir a função de verossimilhança. Sejam y_1, y_2, \dots, y_n uma variável aleatória independente com função densidade $f(y_i; \theta)$ onde θ é um vetor de parâmetros desconhecidos. Então

$$f(\mathbf{y}; \theta) = \prod_{i=1}^n f_i(\mathbf{y}_i; \theta) = L(\theta; \mathbf{y}) \quad (3.15)$$

é chamada de função de verossimilhança.

O estimador de máxima verossimilhança de θ é definido como o valor $\hat{\theta}$ tal que:

$$L(\hat{\theta}; \mathbf{y}) \geq L(\theta; \mathbf{y}) \text{ para todo } \theta. \quad (3.16)$$

ou seja, a estimativa de máxima verossimilhança de θ é o valor $\hat{\theta}$ que maximiza a função de verossimilhança. Em geral, para desenvolver-se o processo de maximização, torna-se mais simples trabalhar-se com a função logaritmo de $L(\theta; \mathbf{y})$:

$$\ell(\theta; \mathbf{y}) = \log L(\theta; \mathbf{y}) = \ell(\theta) \quad (3.17)$$

chamada de função log-verossimilhança ou função suporte.

Como a função logarítmica é monótona, o processo de maximização de $L(\theta; \mathbf{y})$ e $\ell(\theta; \mathbf{y})$ são equivalentes e por serem as observações independentes, temos:

$$\ell(\theta) = \sum_{i=1}^n \log f(\mathbf{y}_i; \theta) \quad (3.18)$$

e então o estimador de máxima verossimilhança é:

$$\ell(\hat{\theta}) \geq \ell(\theta) \quad (3.19)$$

Sejam, então:

i) as variáveis independentes y_1, y_2, \dots, y_n variáveis respostas que satisfazem as propriedades de um MLG.

ii) $\hat{\beta}, \hat{\eta} = \mathbf{X}\hat{\beta}$ e $\hat{\mu} = g^{-1}(\hat{\eta})$ os estimadores de β, η e μ , respectivamente

iii) $L(\beta)$ o logaritmo da função de verossimilhança do MLG. Então:

$$L(\beta) = \sum_{i=1}^n \left[\frac{\mathbf{y}_i \theta_i - b(\theta_i)}{a(\theta_i)} + c(\mathbf{y}_i, \theta_i) \right] \quad (3.20)$$

que desenvolvida e reescrita em notação matricial, resulta nas seguintes equações:

$$\mathbf{X}^T \mathbf{W} \Delta \mathbf{y} = \mathbf{X}^T \mathbf{W} \Delta \mu \quad (3.21)$$

onde \mathbf{X} é a matriz do modelo e \mathbf{W} é uma matriz diagonal com

$$w_i = \frac{p_i}{\left[b''(\theta_i) \left(\frac{d\eta_i}{d\mu_i} \right)^2 \right]}$$

Como existem funções não lineares de β , (3.21) não pode ser resolvida

analiticamente, aplicando-se um método iterativo. Parte-se de um valor inicial $\hat{\beta}$, calcula-se o preditor linear $\hat{\eta}_i = \mathbf{x}_i^T \hat{\beta}$ e o aplicamos para obter os valores ajustados $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$. Com o desenvolvimento matemático destas equações, obtém-se uma estimativa melhorada de β :

$$\hat{\beta} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{z}_i \quad (3.22)$$

ou

$$\beta^{(m+1)} = (\mathbf{X}^T \mathbf{W}^m \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^m \mathbf{z}^m \quad (3.23)$$

onde \mathbf{Z} é um vetor resposta com $z_i^m = \mathbf{x}_i^T \beta^m + (y_i - \hat{\mu}_i) \frac{d\eta_i}{d\mu_i}$ e o expoente m indica o valor do termo na m -ésima iteração.

O procedimento é repetido até que os sucessivos estimadores escolhidos sejam menores que os valores especificados. McCullagh e Nelder (1989) provam que este algoritmo é equivalente ao Escore de Fisher e leva ao estimador de máxima verossimilhança

Conclui-se desta forma que o método escores em cada iteração corresponde a uma regressão entre \mathbf{z} e \mathbf{X} usando mínimos quadrados ponderados com matriz de pesos \mathbf{W} . Desta forma, quanto maior for a variância das observações, menor será seu peso no cálculo dos parâmetros.

Tabela 3.4 – Funções de verossimilhança expressas em função da média e parâmetro de dispersão.

Distribuição	Função de verossimilhança $l(\mu, \phi; \mathbf{y})$
Normal $-\infty < \mathbf{y} < \infty$	$-\frac{1}{2} \log(\phi) - \frac{1}{2\phi} (\mathbf{y} - \mu)^2$
Poisson $\mathbf{y} = 0, 1, 2, \dots$	$\mathbf{y} \log(\mu) - \mu$
Binomial $\mathbf{y} = r/m, r = 0, 1, 2, \dots, m$	$r \log(\mu) + (m - r) \log(1 - \mu)$
Gama $\mathbf{y} > 0$	$-\log\left(\mathbf{y} \Gamma\left(\frac{1}{\phi}\right)\right) + \frac{1}{\phi} \log\left(\frac{\mathbf{y}}{\mu\phi}\right) - \frac{\mathbf{y}}{\mu\phi}$
Normal Inversa $\mathbf{y} > 0$	$-\log(\mathbf{y}^3 \phi) - \frac{(\mathbf{y} - \mu)^2}{2\mathbf{y}\mu^2\phi}$

Fonte: SAS 8.00

3.8 TESTES DE HIPÓTESES

Os métodos de inferência nos modelos lineares generalizados fundamentam-se na teoria de máxima verossimilhança. Esta teoria admite três estatísticas para testar as hipóteses relativas aos parâmetros β 's:

- i) Wald
- ii) razão de verossimilhança e
- iii) escore

Considere o teste de hipótese nula simples $H_0: \beta = \beta_0$ onde β_0 é um vetor p -dimensional conhecido bem como suponha ϕ conhecido. Então se pode definir os testes

i. *Estatística de Wald* (ou de máxima verossimilhança): em geral é usada no caso de hipóteses relativas a um único coeficiente β_i . A estatística é

$$W = (\hat{\beta} - \beta_0)^T \hat{\text{Var}}^{-1}(\hat{\beta})^{-1} (\hat{\beta} - \beta_0) \quad (3.24)$$

onde $\hat{\text{Var}}(\hat{\beta})$ é a matriz de variância-covariância assintótica de $\hat{\beta}$ estimada em $\hat{\beta}$. Segundo Paula (2001, p.26), para os Modelos Lineares Generalizados, $\hat{\text{Var}}(\hat{\beta}) = \mathbf{I}^{-1}(\hat{\beta})$ o que permite reescrever a estatística de Wald da seguinte forma:

$$W = \phi(\hat{\beta} - \beta_0)^T (\mathbf{X}^T \mathbf{W} \mathbf{X}) (\hat{\beta} - \beta_0)$$

onde \mathbf{I} é a matriz informação calculada no ponto $\hat{\beta}$ e \mathbf{W} é a matriz diagonal de pesos com componentes $\omega_i = (d\mu/d\eta)_i^2 / V_i$ e a matriz $\mathbf{V} = \text{diagonal}\{V_1, \dots, V_n\}$.

ii) *Estatística da razão de máxima-verossimilhança*: aplicada geralmente a casos de hipóteses relativas a vários coeficientes β 's. Consideremos dois modelos encaixados w_1 e w_2 . Então $w_1 \subset w_2$, ou seja, w_1 é um subconjunto de w_2 . O objetivo é obter um modelo mais simples w_1 fixando alguns parâmetros de w_2 como zero, e então testar as hipóteses de que estes elementos são realmente zeros. A estatística para este teste é dada por:

$$-2\ln\lambda = 2[\ell(\hat{\beta}_{w_2}; \mathbf{y}) - \ell(\hat{\beta}_{w_1}; \mathbf{y})] \quad (3.25)$$

onde $\lambda = \frac{L(\hat{\beta}_{w_1}, \mathbf{y})}{L(\hat{\beta}_{w_2}, \mathbf{y})}$

iii. *Estatística do escore*: esta estatística é obtida a partir da função escore e é dada por:

$$\mathbf{E} = \mathbf{U}(\beta_0)' \mathbf{I}_0^{-1}(\beta_0) \mathbf{U}(\beta_0) \quad (3.26)$$

onde \mathbf{I}_0 é a matriz informação calculada no ponto β_0 .

Observa-se que o cálculo do teste da razão de máxima verossimilhança exige o ajuste de dois modelos (w_1 e w_2); no teste de Wald compara-se somente um modelo (w_2), enquanto que o teste do escore, às vezes, nenhum modelo.

Cordeiro (1986; p.43), afirma que “as três estatísticas são assintoticamente equivalentes e, segundo H_0 , convergem em distribuição para a variável χ_p^2 ”.

3.9 VERIFICAÇÃO DA QUALIDADE DO AJUSTE DO MODELO

Ajustar um conjunto de dados a um modelo linear generalizado significa substituir o vetor \mathbf{y} de observações pelo vetor $\hat{\boldsymbol{\mu}} = \mathbf{g}^{-1}(\mathbf{X}\hat{\boldsymbol{\beta}})$ de valores ajustados. Um modelo estatístico linear aditivo constitui-se de uma componente sistemática e uma componente aleatória, tendo por objetivo resumir razoavelmente as informações de n observações através de p parâmetros. Estas observações então podem ser ajustadas à modelos com até n parâmetros: quando $p = 1$ tem-se o modelo nulo (considerado simples demais) e quando $p = n$, o modelo é chamado de saturado ou completo (inadequado por não resumir os dados, apenas repeti-los).

Portanto, após uma combinação satisfatória da distribuição da variável resposta e da função de ligação, é necessário encontrar-se um valor p intermediário, $1 < p < n$, que resuma os dados e que torne irrelevante as informações perdidas no processo de ajuste, ou seja, encontrar um modelo

parcimonioso. Estatisticamente, deve-se comparar o modelo ajustado (ou em investigação) com o modelo saturado e verifica-se se a discrepância é significativa; uma grande discrepância não é tolerável, e então deve-se repetir todo o processo na busca de um outro ajuste mais adequado.

Segundo McCullagh e Nelder, Nelder e Wedderburn em 1972, propuseram como medida de discrepância a seguinte estatística:

$$\frac{D(\mathbf{y}; \hat{\boldsymbol{\mu}})}{a(\phi)} = -2[\ell(\hat{\boldsymbol{\mu}}; \mathbf{y}) - \ell(\mathbf{y}; \mathbf{y})] = D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) \quad (3.27)$$

onde $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ é chamada de deviance do modelo e $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ é o *scaled deviance*, expresso como um múltiplo do parâmetro de dispersão. Na relação 3.27, $\ell(\hat{\boldsymbol{\mu}}; \mathbf{y})$ é o máximo da log-verossimilhança para o modelo sob investigação com p parâmetros e $\ell(\mathbf{y}; \mathbf{y})$ é o máximo da verossimilhança para o modelo saturado com n parâmetros. O desvio é portanto uma medida de distância entre os valores ajustados $\hat{\boldsymbol{\mu}}$ em relação aos valores observados \mathbf{y} , ou do modelo sob investigação em relação ao modelo saturado. A Tabela 3.5 apresenta a deviance para os principais modelos.

Tabela 3.5 – Deviance para algumas distribuições de probabilidades.

Distribuições	Desvio
Normal	$\sum_{i=1}^n (\mathbf{y} - \hat{\boldsymbol{\mu}})^2$
Poisson	$2 \sum_{i=1}^n \left[\mathbf{y} \cdot \ell\left(\frac{\mathbf{y}}{\hat{\boldsymbol{\mu}}}\right) - (\mathbf{y} - \hat{\boldsymbol{\mu}}) \right]$
Binomial	$2 \sum_{i=1}^n \left\{ \left[\mathbf{y} \cdot \ell\left(\frac{\mathbf{y}}{\hat{\boldsymbol{\mu}}}\right) \right] + (m - \mathbf{y}) \cdot \ell\left[\frac{(m - \mathbf{y})}{(m - \hat{\boldsymbol{\mu}})}\right] \right\}$
Gama	$2 \sum_{i=1}^n \left[-\ell\left(\frac{\mathbf{y}}{\hat{\boldsymbol{\mu}}}\right) + \frac{\mathbf{y} - \hat{\boldsymbol{\mu}}}{\hat{\boldsymbol{\mu}}} \right]$
Normal inversa	$\sum_{i=1}^n \frac{(\mathbf{y} - \hat{\boldsymbol{\mu}})^2}{(\hat{\boldsymbol{\mu}}^2 \mathbf{y})}$

Fonte: McCullagh e Nelder, 1989, p.34.

De acordo com Cordeiro (1997, p.19), para o teste de um MLG¹, compara-se a deviance e seus graus de liberdade com alguma distribuição teórica de probabilidade. Para alguns modelos tais como Normal, Gama e Normal Inverso, o parâmetro ϕ , em geral é desconhecido, por esta razão, Dobson (2001, p.76), afirma que o deviance não pode ser usado diretamente como uma estatística da *bondade* do ajuste.

Mesmo assim, Cordeiro (1986, p.62) afirma que “na prática, se contenta em testar um MLG, sem muito rigor, comparando-se $D(\mathbf{y};\hat{\boldsymbol{\mu}})$ com o valor crítico $\chi_{n-p}^2(\alpha)$ da distribuição qui-quadrado a um nível de significância igual a α ”; se o deviance for maior que $\chi_{n-p}^2(\alpha)$, rejeita-se o modelo, caso contrário, o modelo é aceito. Particularmente, no “modelo Gama, o *deviance* estará bem aproximado por uma qui-quadrado com $(n-p)$ graus de liberdade a medida que o coeficiente de variação ficar próximo de zero” (Paula, 2001, p.17).

Assim, nos casos em que se pode fazer uma aproximação de uma qui-quadrado com $(n-p)$ graus de liberdade, um valor de $D^*(\mathbf{y};\hat{\boldsymbol{\mu}})$ próximo de $(n-p)$ pode ser a indicação de que o modelo ajustado aos dados é adequado.

Cordeiro apresenta as estimativas de máxima verossimilhança do parâmetro de escala ϕ , Tabela 3.6.

Tabela 3.6: Estimativa pela máxima verossimilhança do parâmetro de escala ϕ .

Modelo	Normal	Normal Inversa	Gama
ϕ	$\frac{n}{D^*}$	$\frac{n}{D^*}$	$\frac{n}{2D^*} \left[1 + \left(1 + \frac{2D^*}{3n} \right)^{1/2} \right]$

Fonte: Cordeiro, 1986, p. 68

No modelo Gama, a estimativa apresentada é aproximada.

3.10 TÉCNICAS DE DIAGNÓSTICO

"Por técnicas de diagnóstico entende-se a análise dos resíduos para detectar

¹ Modelos Lineares Generalizados

observações aberrantes e o estudo da influência de observações sobre o ajustamento global do modelo" (Cordeiro, 1997, p.31).

Define-se o "resíduo para a i -ésima observação como uma função que busca medir a discrepância entre o valor observado y_i e o valor ajustado $\hat{\mu}_i$ da i -ésima observação" (Paula, 1999). A expressão que representa o resíduo é:

$$R_i = h_i(y_i, \hat{\mu}_i)$$

onde h_i é uma função escolhida para estabilizar a variância ou induzir a simetria na distribuição amostral de R_i .

Segundo McCullagh e Nelder (1989, p.37), para "os modelos lineares generalizados amplia-se a definição dos resíduos dos modelos lineares, tornando-os aplicáveis a todas as distribuições que possam substituir à normal. É conveniente que estes resíduos possam ser usados com a mesma finalidade que os resíduos normais padronizados".

Dentre os vários tipos de resíduos, para os modelos lineares generalizados destacam-se:

i. *Resíduo de Pearson*: é o mais simples dos resíduos e sua desvantagem está em geralmente apresentar uma distribuição assimétrica para modelos não normais, cuja representação é

$$P_i = \frac{\phi_i^{1/2}(y_i - \hat{\mu}_i)}{\hat{V}_i^{1/2}} \quad (3.28)$$

onde V_i é dada pela tabela 3.1 e os valores de ϕ são, em geral, conhecidos a priori.

ii. *Resíduo de Anscombe*: este resíduo procura a normalização e a estabilização da variância através de uma transformação $N(y_i)$ da observação y_i . Tal resíduo é expresso por

$$A_i = \frac{\phi_i^{1/2}[N(y_i) - N(\hat{\mu})]}{N'(\hat{\mu})\hat{V}_i^{1/2}} \quad (3.29)$$

onde $N(\mu) = \int V^{-1/3}(\mu)d\mu$

iii) *Resíduo definido como componente do desvio*: representa a distância da observação y_i ao seu valor ajustado $\hat{\mu}_i$, medida na escala da log-verossimilhança. Tem-se que a distância $\pm D_i$ é bem modelada por uma normal padronizada se o modelo for adequado. Sua representação é

$$D_i^2 = 2\phi_i[V(y_i) - V(\hat{\mu}) + \theta(\hat{\mu}_i)(\hat{\mu}_i - y_i)] \quad (3.30)$$

Uma vantagem deste resíduo é que D_i não exige o conhecimento da função normalizadora.

Para o modelo normal não há diferença quanto a aplicabilidade de qualquer um dos resíduos, entretanto, quando o modelo não for suficientemente adequado aos dados, ou então deseja-se estudar as deficiências do modelo, deve-se empregar os resíduos de Anscombe e das componentes do desvio, pois neste caso, o de Pearson apresenta forte assimetria. Quando os modelos estão bem ajustados, as diferenças entre D_i e P_i devem ser pequenas, porém para modelos mal ajustados e/ou com discrepâncias, podem ocorrer diferenças consideráveis entre os resíduos.

Dobson (2002, p.33) afirma que os resíduos de maneira geral, deveriam ser independentes e ter uma distribuição aproximadamente normal com média zero e variância constante, também devem ser não-correlacionados com as variáveis explicativas. Conseqüentemente, os resíduos padronizados podem ser comparados à distribuição normal para identificar algum valor discrepante.

Então, após a determinação de R_i , "uma análise gráfica é útil para diagnosticar a eficácia do ajuste" (Dantas, p.179). A investigação gráfica pode ser feita através dos seguintes procedimentos:

Grafar os resíduos contra os valores esperados (ou alguma função dos valores ajustados). Se os pontos estiverem próximos de uma linha reta, isto demonstra a normalidade dos dados. Se os pontos divergirem desta reta, indica que as observações não pertencem à distribuição proposta para os dados.

Grafar os resíduos contra as variáveis preditoras incluídas no modelo. Se

o modelo for adequado, a distribuição dos pontos não deve apresentar qualquer padrão. Caso verifique-se um comportamento sistemático, termos adicionais ou alternativos podem ser incluídos no modelo.

Grafar os resíduos contra variáveis preditoras que não estejam no modelo. Se ocorrer algum padrão sistemático, sugere-se que variáveis adicionais devam ser incluídas.

Grafar resíduos contra os valores ajustados. Pode-se verificar a constância da variância (homoscedasticidade).

Grafar os resíduos contra a ordem seqüencial em que os valores y_i foram mensurados. Aplicado para verificar-se a independência entre as observações. Caso as variáveis realmente sejam independentes, os pontos devem flutuar aleatoriamente, sem qualquer padrão sistemático.

Outra opção importante é a deleção, na qual se compara o ajuste do modelo escolhido, considerando-se todos os pontos, com o ajuste do mesmo modelo sem os pontos atípicos.

Para auxiliar a identificação de pontos atípicos ao modelo, Dobson (2002, p.89), define como outlier a observação que não se ajusta bem ao modelo, enquanto que a observação influente é aquela com efeito relativamente grande nas inferências baseadas no modelo.

Demétrio (2001, p.65) afirma que quando uma observação está distante das outras em termos das variáveis explicativas ela pode ser, ou não, influente. Segundo McCullagh e Nelder, a medida para determinar a *leverage*, ou seja, o ponto de alavanca, é dada pelos elementos da diagonal da matriz “chapéu” H .

Pelo processo de ajuste,

$$\hat{\mu} = \mathbf{HY} = \sum_{j=1}^n h_{ij} Y_j = h_{i1} Y_1 + h_{i2} Y_2 + \dots + h_{ij} Y_j + \dots + h_{in} Y_n, \quad 1 \leq i \leq n$$

sendo que $h_i = h_{ii} = \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i$ depende da matriz do delineamento. Entretanto, como $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ é uma matriz de projeção, tem-se:

$$h_i = h_{ii} = \sum_{i'=1}^n h_{ii'}^2 = h_{ii}^2 + \sum_{i' \neq i} h_{ii'}^2,$$

$$\text{e } 0 \leq h_{ii} \leq 1, \sum_{i=1}^n h_{ii} = 1 \text{ e } r(\mathbf{H}) = \text{tr}(\mathbf{H}) = \sum_{i=1}^n h_{ii} = p$$

Portanto, segundo Demétrio (2001, 65), o valor ajustado $\hat{\mu}$ é a média ponderada dos valores observados e o peso da ponderação é o valor de h_{ii} .

Segundo Dobson (2002, p.89), um ponto pode ser considerado com alta *leverage*, para Demétrio ponto de alavanca, quando $h_{ii} > \frac{2p}{n}$, onde p é o número de parâmetros e n o número de observações trabalhadas.

As caracterizações e definições contidas neste capítulo permitem estabelecer procedimentos metodológicos que serão descritos no próximo capítulo. Tais procedimentos serão responsáveis por um tratamento adequado dos dados em estudo segundo a teoria dos Modelos Lineares Generalizados, conduzindo a resultados que deverão satisfazer os objetivos estabelecidos nesta pesquisa.

4- PROCEDIMENTOS METODOLÓGICOS

A estratégia metodológica proposta pelo presente trabalho para a determinação do fator de condição a partir da relação peso-comprimento, construída segundo aplicação da teoria dos Modelos Lineares Generalizados, toma como dados de aplicação as cinco espécies já citadas no Capítulo 2, supõe:

- representatividade e validade dos dados amostrados na análise estatística;
- utilização de um conjunto de métodos estatísticos concernentes aos Modelos Lineares Generalizados;
- utilização de software que forneça as análises propostas sobre os dados disponíveis.

A metodologia desenvolvida nesta pesquisa pautou-se em um processo de informações teóricas baseada na bibliografia apresentada. Os dados trabalhados na etapa de aplicação foram gentilmente cedidos pelo Nupélia e correspondem a uma amostragem criteriosa, coletada sistematicamente por um período de aproximadamente 10 anos.

Após determinar-se um modelo capaz de ajustar satisfatoriamente os dados, compara-se o modelo obtido a outros citados por Bolger e Connolly (1989, p.173-174). Esta etapa é um processo de validação do modelo, pois os modelos aos quais este novo modelo é comparado, são aceitos mundialmente em estudos biológicos.

A etapa final corresponde à determinação de uma equação que, a partir da relação peso-comprimento encontrada, seja capaz de encontrar um fator de condição K . Então, os dados são aplicados a esta equação e a seguir é feito um estudo sobre a distribuição dos valores resultantes.

Os cálculos desenvolvidos na execução de cada etapa foi realizado através da utilização dos software SAS, Statistica e Excel.

4.1 POPULAÇÃO:

As cinco espécies em estudo, *Trachydoras paraguayensis*, *Liposarcus anisiti*, *Raphiodon vulpinus*, *Serrasalmus marginatus* e *Leporinus friderice*, cujo habitat é a planície de inundação do rio Paraná.

4.2 VARIÁVEIS

Todo trabalho é desenvolvido considerando-se como variável resposta (ou variável dependente), a variável peso total do peixe, identificada por W_t . As variáveis preditoras constitui-se de uma covariável, o comprimento padrão (L_s), e pelos fatores espécie e sexo. O fator espécie compõe-se de cinco níveis: *Trachydoras paraguayensis* (A), *Liposarcus anisiti* (C), *Raphiodon vulpinus* (D), *Serrasalmus marginatus* (H) e finalmente, *Leporinus friderice* (P). Os níveis do fator sexo são identificados por fêmeas (F) e machos (M).

4.3 AMOSTRA E COLETA DE DADOS

O uso de amostragem em pesquisas Ictiológicas é um procedimento padrão que exige cuidadoso planejamento e execução. As estatísticas desenvolvidas nesta área de conhecimento, em geral, são comparativas, cujos resultados podem conduzir a importantes estimativas que auxiliam na avaliação e identificação de indicadores de crescimento, condições globais da saúde do peixe, depredações do meio, comportamento em comunidade e, até mesmo, a qualidade da água onde habitam.

Uma forte característica em Ictiologia, é que raramente pode-se examinar todos os indivíduos de uma população (como por exemplo, drenar uma lagoa). Então, segundo Ney (1993, p.138), a principal preocupação do cientista deve consistir na *suposição da coleta ser representativa da população*, pois se tal condição não ocorre, as estatísticas desenvolvidas são tendenciosas, ou seja, não estimam os verdadeiros parâmetros populacionais, podendo inclusive resultar em comprometimento de pesquisas futuras. Portanto, as principais características de um planejamento amostral são as de assegurar *consistência e representatividade* dos dados obtidos.

Nesta pesquisa, para assegurar os pressupostos básicos, as coletas foram realizadas através de procedimentos metódicos, repetitivos e diversificados, possibilitando a captura de espécies distintas e em várias fases de seu desenvolvimento. Os indivíduos são amostrados com o uso de redes de espera simples com malhagem de 3, 4, 5, 6, 7, 8, 10, 12, 14 e 16 centímetros entre nós opostos, e redes feiticeira com 7 e 8 centímetros entre os nós adjacentes. A escolha do aparelho de pesca em cada reservatório é feita com base em amostragem piloto para estimativa de eficiência de pesca. Após a captura, as amostras são fixadas em formol 4%, acondicionadas, com a identificação do local, data, horário e método de pesca, em tambores de polietileno, e conduzidas aos laboratórios de ictiologia do Nupélia onde são mensurados o comprimento total e padrão, o peso total, sexo, estágio de desenvolvimento gonadal, grau de repleção do estômago (escala de 0 a 3) e seu peso.

As coletas foram realizadas na planície de inundação do Alto Rio Paraná, a partir de 1986, sob a orientação do Nupélia.

Três regiões foram mais intensivamente estudadas dentro da Planície: i) baixo rio Ivinheima; ii) baixo rio Baía e iii) proximidades da calha principal do rio Paraná, nas imediações do município de Porto Rico. O levantamento dos dados foi realizado nos meses de novembro de 1986 a outubro de 1988 (mensais), março de 1993 a fevereiro de 1995 (mensais e bimestrais), março de 2000 a novembro de 2001 (trimestrais), em 36 estações de amostragem com abrangência suficiente para contemplar os distintos ambientes da região, tais como lagoas permanentes e temporárias com diferentes níveis de conectividade com a calha principal do rio, ressacos, canais e o leito dos principais rios do sistema (Ivinheima, Baía e Paraná). Para o caso específico das assembleias aquáticas, além das estações de amostragem alocadas na calha de cada um dos rios, foram estudadas quatro lagoas da planície, três conectadas ao rio Baía e uma ao rio Ivinheima. A Figura 3.1 mostra a área amostrada.

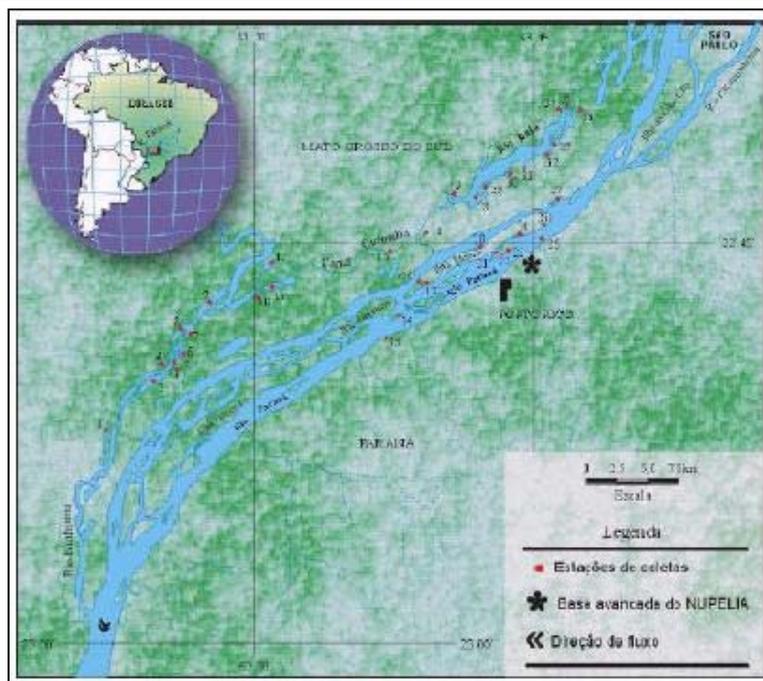


Figura 4.1: Mapa da área de amostragem

Nota: 1-Lagoa Peroba-LPER; 2-Lagoa Ventura-LVEN; 3-Lagoa do Zé do Paco-LZEP; 4-Canal do Ipoitã-CIPO; 5-Lagoa Boca do Ipoitã-LIPO; 6-Lagoa dos Patos-LPAT; 7-Lagoa Capivara-LCAP; 8-Rio Ivinheima-RIVI; 9-Lagoa do Finado Raimundo-LFRA; 10-Lagoa do Jacaré-LJAC; 11-Lagoa Sumida-LSUM; 12-Lagoa do Cervo-LCER; 13-Canal Cortado-CCOR; 14-Lagoa das Pombas-LPOM; 15-Canal Curutuba-CCUR; 16-S Ressaco do Manezinho-LMAN; 17-Lagoa do Osmar-LOSM; 18-Lagoa da Traíra-LTRA; 19-Lagoa do Guaraná-LGUA; 20-S Ressaco do Bilé-LBIL; 21-S Ressaco do Leopoldo-LLEO; 22-Lagoa Genipapo-LGEN; 23-Lagoa Clara-LCLA; 24-S Ressaco do Pau Véio-LPVE; 25-Rio Paraná-RPAR; 26-Lagoa Pousada-LPOU; 27-Lagoa das Garças-LGAR; 28-Rio Baía-RBAI; 29-Lagoa Fechada-LFEC; 30-Lagoa Pousada das Garças-PGA; 31-Lagoa dos Porcos-LPOR; 32-Lagoa do Aurélio-LAUR; 33-Baía Canal-CBAI; 34-Lagoa Maria Luiza-LMLU; 35-Lagoa do Gavião-LGAV; 36-Lagoa da Onça-LONÇ)

4.4 ETAPAS PARA A CONSTRUÇÃO DO FATOR DE CONDIÇÃO DO PEIXE

Com o propósito de facilitar a visualização das etapas desenvolvidas na construção do fator de condição do peixe, a Figura 3.2 apresenta um fluxograma, detalhando o processo de construção.

A primeira etapa da pesquisa é buscar garantir a confiabilidade e representatividade da amostra, codificando e identificando as variáveis de interesse para a pesquisa.

A estratégia desenvolvida para ser possível atingir os objetivos propostos apresenta quatro fases: a primeira caracteriza-se pela construção de um único modelo, capaz de relatar o peso do peixe tendo como variáveis preditoras sua espécie, sexo e comprimento. Após detectar a significância das espécies para a construção do modelo, passa-se para a segunda fase, reconstruindo um modelo individual para cada espécie, identificando para

quais delas o fator sexo é significativo. A terceira fase do processo preocupa-se em validar o modelo, comparando-o a outros modelos já aceitos e utilizados em trabalhos ictiológicos. Finalmente, a quarta e última etapa dedica-se à determinação do fator de condição K do peixe, condicionado a cada modelo especificado para a relação peso-comprimento da espécie.

Deve-se salientar que para a primeira e segunda etapa, o processo de construção do modelo compreende três etapas: i) determinação da distribuição da variável resposta; ii) identificação de uma função de ligação; iii) avaliação do modelo. Cada uma dessas etapas está minuciosamente relatada no Capítulo 5.

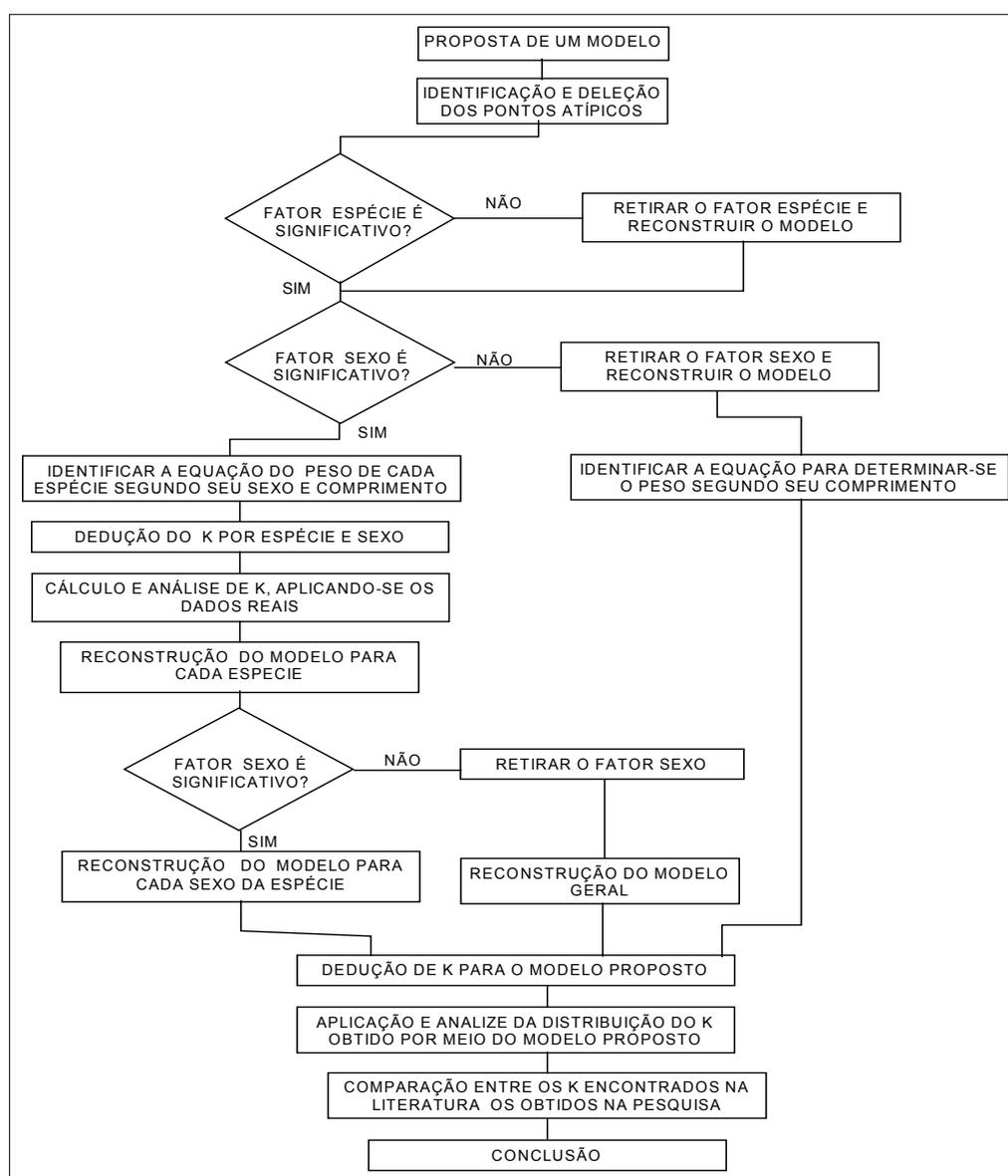


Figura 4.2: Fluxograma do processo de construção do Fator de Condição do Peixe, K.

5- ANÁLISE DOS RESULTADOS

O principal objetivo deste trabalho é a aplicação dos modelos lineares generalizados para a determinação de uma relação do peso-comprimento das espécies de peixes em estudo em função das variáveis definidas. Para se chegar a esta relação, faz-se necessário, primeiramente, a determinação de um modelo que ajuste satisfatoriamente os dados considerando a variável resposta peso total (W_t) em função das variáveis predictoras: covariável comprimento padrão (L_s) e os fatores espécie (A, C, D, H e P) e sexo (F, M). Nesta etapa, poder-se-á verificar uma das hipóteses básicas da pesquisa, a significância das espécies e seus respectivos sexos na determinação da variável peso.

5.1 CONSTRUÇÃO DO MODELO GERAL

A construção de um modelo geral justifica-se pelo fato de que, usualmente, encontra-se disponível na literatura, fatores de condição únicos, que não consideram as espécies ou o sexo dos animais. Uma preocupação neste capítulo é verificar a significância destes fatores.

5.1.1 Distribuição da variável resposta peso total (W_t)

Como foi definido no Capítulo 2, o primeiro passo para a construção do modelo é a identificação da distribuição da variável resposta W_t . A análise do comportamento descritivo da variável auxilia esta identificação. A figura 5.1 mostra o histograma enquanto que na Tabela 5.1 estão resumidas as principais medidas descritivas da variável peso.

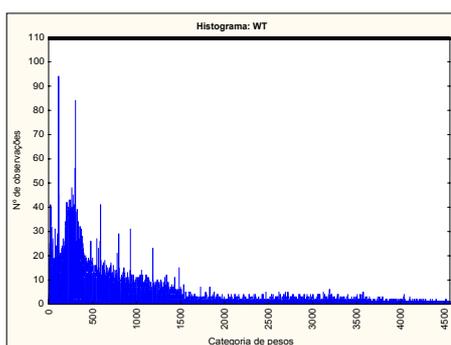


Figura 5.1: Histograma da variável peso total.

Os valores de W_t são contínuos e positivos (peso total sempre maior que zero), apresentam um alto grau de assimetria positiva e variabilidade. O peso médio total dos animais é de 111,68 gramas com desvio padrão de 177,07 gramas para um conjunto de dados que apresenta um valor mínimo de 1,10 gramas e máximo de 3000,00 gramas.

Tabela 5.1: Medidas descritivas da variável peso total (W_t).

Nº de casos	Média	Mediana	Moda	Desvio padrão	Assimetria	Curtose
15.175	111,6796	40,20	10,00	177,0706	3,7894	24,7899

Tais resultados sugerem como possíveis distribuições para a variável resposta, as distribuições Gama ou a Normal Inversa, pois, segundo Cordeiro (1997, p.2), “as distribuições Gama e Normal Inversa são associadas a dados contínuos assimétricos e positivos”.

5.1.2 Identificação da função de ligação:

O processo elementar para determinar as possíveis funções de ligações entre as covariáveis e a variável resposta é a construção do gráfico de dispersão dessas variáveis. É importante salientar que tal processo pode levar a funções de ligações equivocadas, especialmente em casos onde existam várias covariáveis; como o caso em estudo existe apenas uma covariável, este recurso é bastante útil.

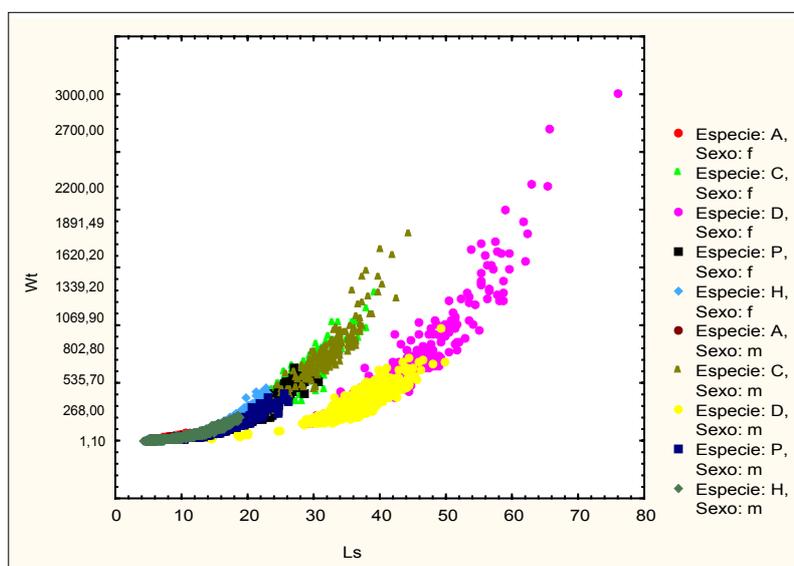


Figura 5.2: Dispersão entre as variáveis comprimento padrão (L_s) e peso total (W_t).

Observa-se pela Figura 5.2 a formação de dois grupos distintos: um grupo formado pelas espécies A, C, H e P e um outro formado apenas pela espécie D. Em ambos os casos, a função sugerida para representar a relação entre as variáveis é a função potência, ou seja, a função de ligação entre a média da variável peso total e o preditor linear.

5.1.3 A construção de um modelo

A construção do modelo considerará dois grupos: um grupo formado pelas espécies A, C, P e H e um outro grupo formado pela espécie D.

Para o primeiro grupo, dentre os vários modelos elaborados, foram dois os que mais se adequaram aos pré-supostos dos MLGs: o modelo cuja função de distribuição é a Gama com função de ligação potência de grau 0,35, ou potência $(0,35)^{*2}$ e o modelo cuja função de distribuição é a Normal Inversa com ligação potência (0,296).

Em ambos os casos, para melhorar a qualidade do modelo, retiram-se os pontos atípicos ao modelo, cujo procedimento de identificação seguiu a seguinte regra prática:

- i. Identificar pontos influentes, ou seja, pontos que produzem grandes efeitos nas inferências baseadas no modelo. De acordo com definição da seção 3.10 deste trabalho, são os pontos da diagonal da matriz chapéu $(X(X^T X)^{-1} X^T)$, cujo valor seja superior a duas vezes o número de parâmetros estimados dividido pelo tamanho da amostra, ou seja

$$h_{ii} \geq 2 \frac{9}{13014} = 0,0014$$

A determinação destes valores foi realizada pela aplicação dos softwares SAS e Statistica 6.0

- ii. Identificar os pontos considerados *outlier*, ou seja, os pontos ou

² Neste trabalho, o grau da função de ligação potência será indicado entre parênteses.

observações que não se ajustam bem ao modelo. Para tal, considerou-se *outlier* os pontos cujo valor absoluto do resíduo padronizado devido à deviance seja maior ou igual a 2,5.

Os pontos resultantes da intersecção desses dois itens, ou seja, *outlier* que sejam pontos influentes foram considerados atípicos ao modelo e por tal motivo, retirados do conjunto de dados em análise.

Os resultados obtidos pelos dois modelos propostos serão discutidos a seguir. O primeiro modelo proposto apresenta as seguintes características:

Variável resposta: peso total

Distribuição: Gama

Preditor linear: $\eta = \alpha + L_s + \text{Espécie} + \text{Sexo} + \text{Espécie} | \text{Sexo}$

Ligação: potência (0,35)

Tabela 5.2: Deviances para o modelo Gama com e sem pontos atípicos

Casos	Deviances	Graus de Liberdade (GL)	Estatística	Estatística/GL
Total	Deviance	13.005	299,28	0,023
	Deviance escalar	13.005	13.063,69	1,005
Sem atípicos	Deviance	12.791	211,11	0,017
	Deviance escalar	12.791	12.835,09	1,003

A Tabela 5.2 mostra que ao trabalhar as 13.014 observações, obtém-se uma deviance de 299,28. Retirando-se 214 pontos atípicos do modelo, cerca de 1,6 % do total dos casos, houve uma forte redução na deviance, passando esta a ser de 211,1095. Observa-se também que o valor da deviance escalar no segundo caso apresenta maior proximidade do tamanho da amostra (diferença de 49,69 e 35,09, respectivamente), mostrando o bom ajuste do modelo. Tais resultados, considerando-se também a grande quantidade de elementos trabalhados na amostra, permitem que se trabalhe com o conjunto de dados excluídos, ou seja, os pontos atípicos, obtendo-se assim, um modelo mais apurado, com maior representatividade para a determinação da relação peso-comprimento dos peixes. A Figura 5.3 permite comparar graficamente a relação peso-comprimento em ambas situações.

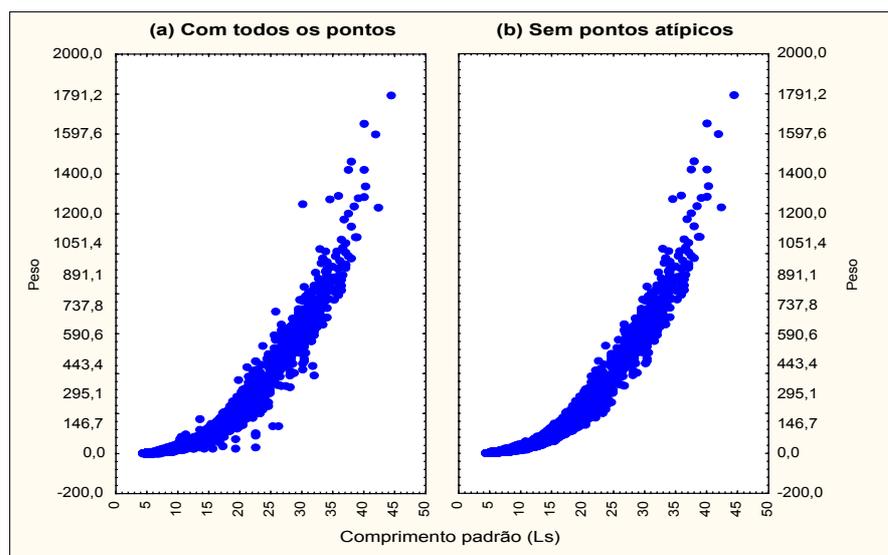


Figura 5.3: Diagrama de dispersão entre as variáveis comprimento e peso (a) com total de observações coletadas e (b) excluindo-se os pontos atípicos ao modelo estudado.

As estimativas dos parâmetros e as respectivas estatísticas de Wald, que testam a significância desses estimadores estão expostas na Tabela 5.3.

Tabela 5.3: Parâmetros estimados para o modelo Gama - Geral.

Efeitos	Valores Estimados	Estatística de Wald	p
Intercepto	-0,698	6.036,5	0,00001
Ls	0,362	671.117,2	0,00001
Espécie A	0,490	5.076,4	0,00001
Espécie C	-0,607	1.080,8	0,00001
Espécie P	-0,137	346,9	0,00001
Sexo f	0,045	55,9	0,00001
EspécieA*SexoF	-0,027	18,7	0,00001
EspécieC*SexoF	0,156	80,8	0,00001
EspécieP*SexoF	-0,087	139,1	0,00001

A Tabela 5.4 mostra os resultados do teste da razão da verossimilhança para os parâmetros, confirmando a significância dos mesmos através de elevados e significativos valores do χ^2 .

Tabela 5.4: Teste tipo 1 para os parâmetros do modelo Gama - Geral

Efeitos	Graus de liberdade	Log-verossimilhança	χ^2	p
Intercepto	1	-66.909,3		
Ls	1	-41.639,6	50.539,38	0,00001
Espécie	3	-37.173,3	8.932,47	0,00001
Sexo	1	-37.162,7	21,23	0,00001
Espécie*Sexo	3	-37.078,2	169,04	0,00001

Os parâmetros estimados apresentaram todos um alto grau de significância, ou seja, pode-se afirmar que a covariável comprimento e os fatores espécies e sexo, bem como interação sexo|espécie são significativos para a determinação do modelo do peso total do peixe, pois realmente os valores obtidos pelo teste da razão da verossimilhança apresentam valores absolutos bastante elevados.

O ajuste pelo segundo modelo proposto segue procedimentos semelhantes ao anterior, com as seguintes características:

Variável resposta: peso total

Distribuição: Normal Inversa

Preditor linear: $\eta = \alpha + L_s + \text{Espécie} + \text{Sexo} + \text{Espécie} | \text{Sexo}$

Ligação: potência (0,296)

Neste modelo, detectou-se 370 pontos atípicos, o que corresponde a aproximadamente 2,8% do total das observações trabalhadas. A Tabela 5.5 apresenta as deviances calculadas a partir dos dois conjuntos de dados: com as 13.014 observações e a seguir, sem os pontos atípicos, ou seja, com 12.644 observações.

Tabela 5.5: Deviance para o modelo Normal Inversa com e sem pontos atípicos

Casos	Deviances	Graus de Liberdade (GL)	Estatística	Estatística/GL
Total	Deviance	13.005	28,554	0,0022
	Deviance escalar	13.005	13.014,00	1,0007
Sem atípicos	Deviance	12.635	10,722	0,0008
	Deviance escalar	12.635	12.644,00	1,0007

Para o caso dos dados sem os pontos atípicos ao modelo, ocorre uma significativa redução na deviance; ela passa a ser aproximadamente 1,5 vezes menor do que a obtida através da amostra original enquanto que em ambos os casos, a deviance escalar é muito próxima do tamanho da amostra (aproximadamente, diferença nula em ambos os casos).

Como o tamanho da amostra é muito grande e dada a queda significativa no valor da deviance, o modelo será construído excluindo-se os pontos

atípicos. A Tabela 5.6 apresenta os parâmetros estimados por este novo modelo e suas respectivas estatísticas de Wald, mostrando que a espécie P e a interação entre a espécie A e sexo não são significativos para o modelo.

Tabela 5.6: Parâmetros estimados para o modelo Normal Inversa – Geral.

Efeitos	Valores Estimados	Estatística de Wald	p
Intercepto	-0,185	85,980	0,00001
Ls	0,260	438.391,000	0,00001
Espécie A	0,415	461,401	0,00001
Espécie C	-0,669	134,699	0,00001
Espécie P	0,034	2,903	0,08842
Sexo F	0,040	4,247	0,03931
EspécieA*SexoF	-0,030	2,408	0,12069
EspécieC*SexoF	0,142	6,117	0,01339
EspécieP*SexoF	-0,074	14,214	0,00016

O teste da razão da verossimilhança é apresentado na Tabela 5.7, e confirma a significância dos fatores espécies, sexo e interação entre os mesmos.

Tabela 5.7: Teste tipo 1 para os parâmetros do modelo Normal Inversa - Geral

Efeitos	Graus de liberdade	Log-verossimilhança	χ^2	p
Intercepto	1	-64.356,4		
Ls	1	-45.811,5	37089,82	0,00001
Espécie	3	-41.026,2	9570,56	0,00001
Sexo	1	-41.019,5	13,39	0,00025
Espécie*Sexo	3	-40.986,1	66,77	0,00001

Embora o modelo com função de distribuição Normal Inversa para a variável resposta apresente uma deviance significativamente inferior que a do modelo Gama, este fato não é suficiente para garantir que ele seja o melhor modelo para ajuste dos dados. A decisão quanto à escolha do modelo a ser aceito como adequado aos dados, deve considerar uma apurada análise dos resíduos.

Neste trabalho, os gráficos analisados seguem, basicamente, sugestões encontradas na obra de Cordeiro e Demétrio. Os gráficos trabalhados nesta análise são:

i) *Resíduos X Valores ajustados (ou alguma função dos valores ajustados):*

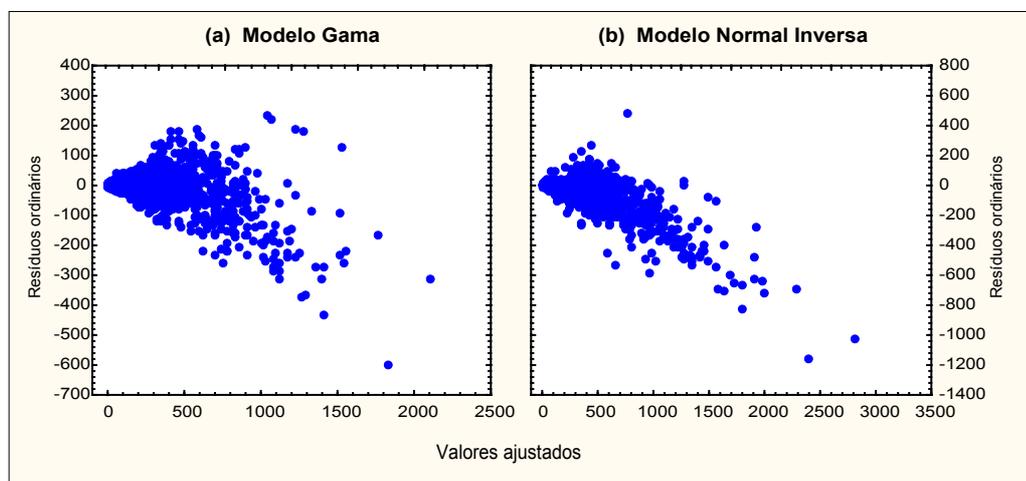


Figura 5.4: Dispersão entre as variáveis ajustadas e seus respectivos resíduos básicos segundo o modelo (a) Gama e (b) Normal Inversa.

Em ambos os gráficos da Figura 5.4, a variância não é constante; é recomendável o gráfico de algum tipo de resíduo padronizado versus os valores ajustados transformados de tal forma a ter-se uma variância constante (McCullagh e Nelder, 1989). Demétrio (2002, p.72) sugere a transformação $2\ln(\hat{\mu})$ para a distribuição Gama e $-2\hat{\mu}^{-1/2}$ para a distribuição Normal Inversa. Aplicando estas transformações, obtém-se os gráficos na Figura 5.5:

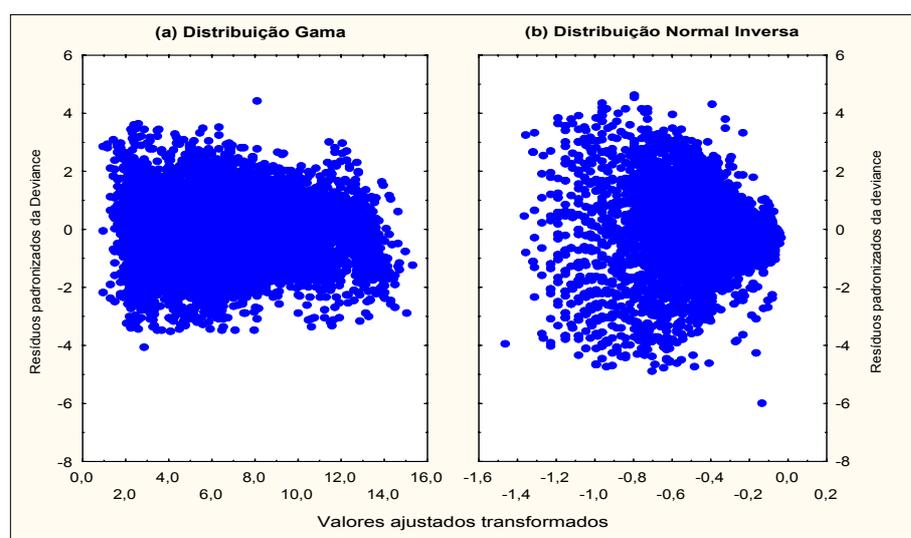


Figura 5.5: Dispersão entre as variáveis ajustadas transformadas e seus respectivos resíduos padronizados das deviances para os modelos (a) Gama e (b) Normal Inversa.

O padrão destes gráficos (Figura 5.5) é uma distribuição dos resíduos em

torno de zero com amplitude constante, porém no caso da distribuição Normal Inversa pode-se observar que esta condição não ocorre.

ii) *Resíduos x Ordem das observações*

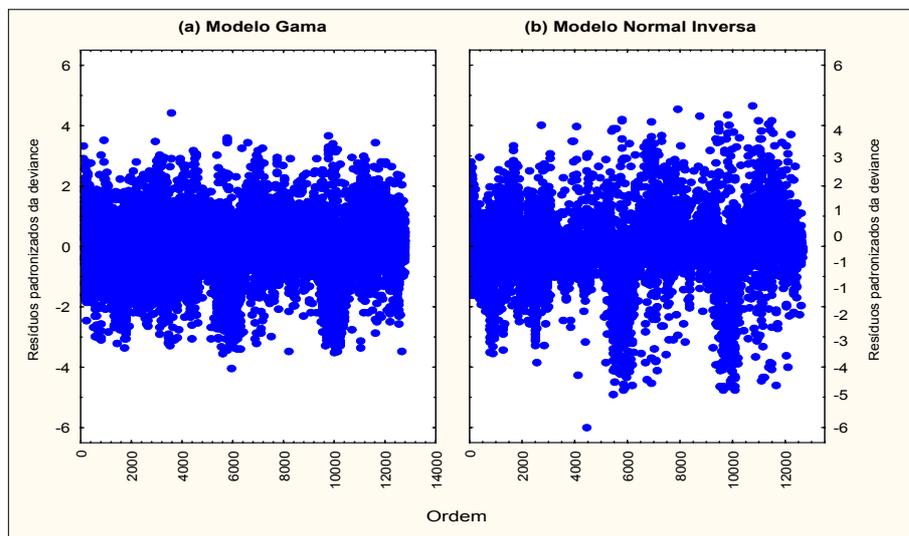


Figura 5.6: Dispersão entre a ordem das observações e os resíduos padronizados.

Este tipo de gráfico (Figura 5.6) verifica a independência entre as observações; para ambos os casos, observa-se que tal condição ocorre.

iii) *Valores observados x Valores ajustados*

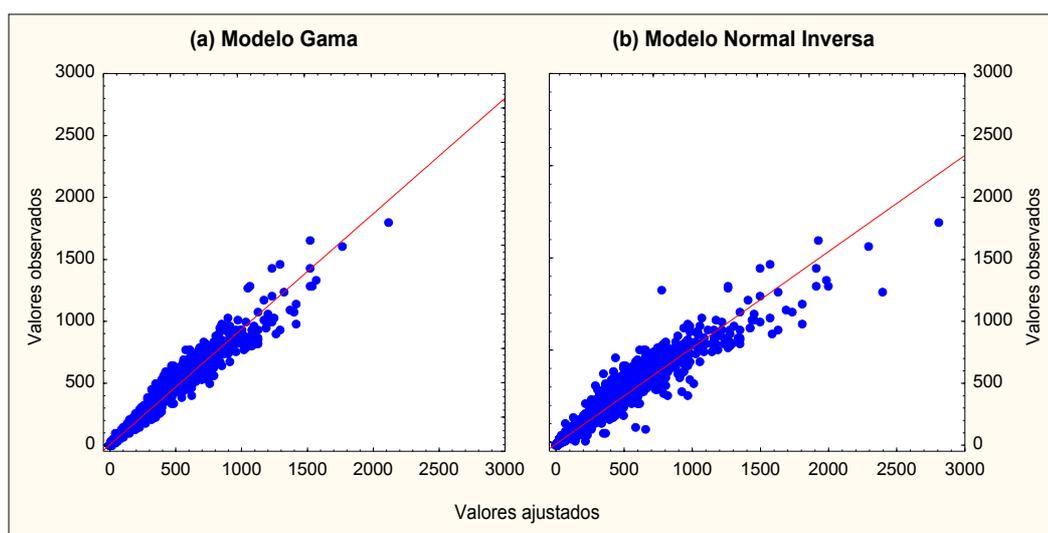


Figura 5.7: Dispersão dos valores ajustados pelos valores observados (a) Modelo Gama e (b) Modelo Normal Inversa.

Nos gráficos da Figura 5.7, a proximidade da disposição dos dados em

relação à 1ª bissetriz sugere a adequação dos dados ao modelo, particularmente, auxilia a aceitação da função de ligação proposta (AMARAL, 1997, p.97).

Observa-se que no caso da distribuição Gama, o comportamento do gráfico é mais aproximado de uma reta, com menor dispersão, enquanto que para a distribuição Normal Inversa há tendência de uma leve curvatura para baixo, o que no caso da ligação potência sugere a aplicação de um expoente maior (DEMÉTRIO, 2002, p.76). A partir desta sugestão, redefinindo o modelo, tomando a função de ligação potência (0,35), pode-se observar pela Figura 5.8, uma mudança de comportamento. Verifica-se que com este novo grau da função de ligação, a seqüência dos pontos resultantes da relação valores ajustados x valores observados aproxima-se mais da primeira bissetriz, enquanto a deviance do modelo passa de 10,808 para 10,804, apresentando uma redução de 0,004, ou seja, uma diferença não significativa.

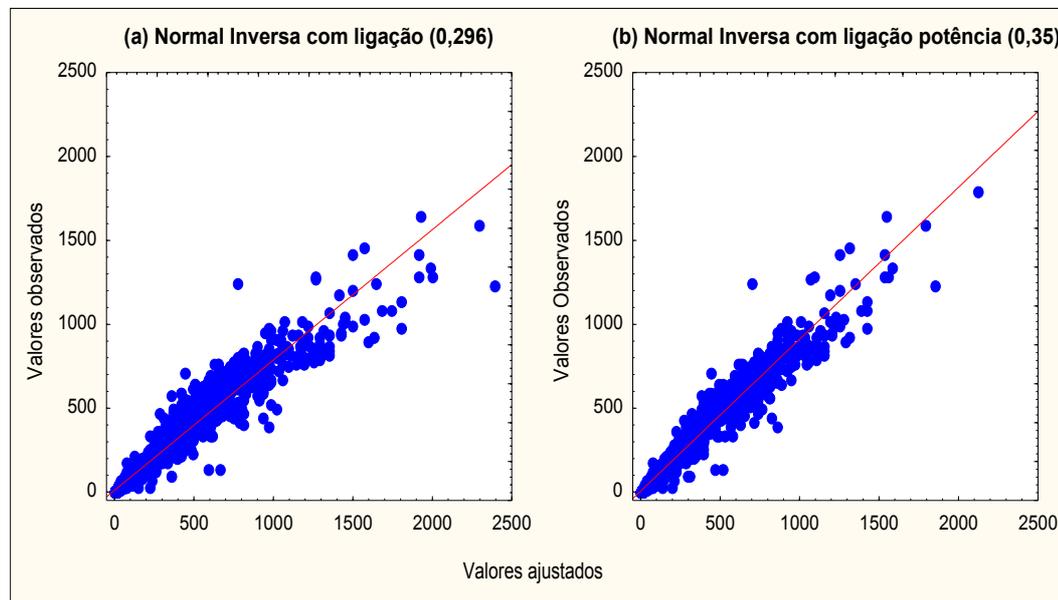


Figura 5.8: Gráficos dos valores ajustados pelos valores previstos para os dois modelos Normal Inversa em estudo.

Na Tabela 5.8 estão apresentados os parâmetros estimados segundo o novo modelo e o teste de Wald para verificar sua significância.

Tabela 5.8: Parâmetros estimados para a função Normal Inversa com ligação potência (0,35).

Efeitos	Valores Estimados	Estatística de Wald	p
Intercepto	-0,526	264,2	0,00001
Ls	0,354	435707,4	0,00001
Espécie A	0,386	148,2	0,00001
Espécie C	-0,418	19,5	0,00001
Espécie P	-0,107	11,1	0,00084
Sexo F	0,042	1,8	0,18166
EspécieA*Sexo F	-0,028	0,8	0,38313
EspécieC*Sexo F	0,148	2,5	0,11585
EspécieP*Sexo F	-0,080	6,1	0,01315

Nota: Espécie A, *Trachydoras paraguayensis*. Espécie C, *Liposarcus anisiti* e Espécie P, *Leporinus friderice*.

Para este modelo, o fator sexo não é significativo. Então, na busca de um modelo mais parcimonioso, refaz-se o modelo excluindo o fator sexo. Os novos parâmetros estão apresentados na Tabela 5.9:

Tabela 5.9: Parâmetros estimados para a função Normal Inversa com ligação potência (0,35), desconsiderando o fator sexo.

Efeitos	Valores Estimados	Estatística de Wald	p
Intercepto	-0,504	254,2	0,00001
Ls	0,354	431.515,7	0,00001
Espécie A	0,365	139,3	0,00001
Espécie C	-0,362	15,3	0,00009
Espécie P	-0,123	15,4	0,00009

Nota: Espécie A, *Trachydoras paraguayensis*. Espécie C, *Liposarcus anisiti* e Espécie P, *Leporinus friderice*.

A nova deviance é de 10,743. Pelo teste da razão da verossimilhança, todos os efeitos são bastante significativos;

Tabela 5.10: Teste tipo 1 para os parâmetros do modelo Normal Inversa-Geral, desconsiderando o fator sexo.

Efeitos	Graus de liberdade	Log-verossimilhança	χ^2	p
Intercepto	1	-64361.2		
Ls	1	-45276.6	38169.09	0,0001
Espécie	3	-41065.7	8421.91	0,0001

Para um teste mais formal, Demétrio (2002, p.73) sugere adicionar $\hat{\eta}^2$ como uma covariável extra e então examinar a mudança ocorrida na

deviance. Se houver uma redução drástica, é uma evidência de que a função de ligação não é satisfatória.

Tabela 5.11: Valores das deviances do modelo inicial e do modelo acrescido da covariável $\hat{\eta}^2$.

Distribuição de probabilidade	Deviance do modelo	Deviance acrescido de $\hat{\eta}^2$	Diferença
Gama	211,109	210,441	0,668
Normal Inversa (0,296)	10,722	10,582	0,140
Normal Inversa (0,35)	10,804	10,721	0,0831
Normal Inversa (0,35)-(Ls+Espécie)	10,743	10,678	0,065

Para todos os casos a diferença é extremamente pequena, indicando que a função de ligação é satisfatória (Tabela 5.11).

Como o conjunto de observações é extremamente grande, a comparação dos gráficos dos resíduos ordinários ordenados auxilia a compreensão do comportamento dos dois modelos. A análise deste gráfico mostra que o ajuste pela função de distribuição Gama tende subestimar os resultados, enquanto que pela distribuição Normal Inversa, a tendência é a superestimar o ajuste. O modelo Gama apresenta 99,98% dos resíduos ordinários dentro do intervalo de 231,54g até $-370,77g$ e destes, 73,63% encontram-se dentro de um intervalo de 5,00g a $-5,00g$. Pelo modelo Normal Inversa, 99,98% dos resíduos ordinários encontram-se entre no intervalo de 290,07g a $-437,63g$ dos quais 70,04% dos casos apresentam resíduos entre 5,00g e $-5,00g$. A Figura 5.9 resume tais informações.

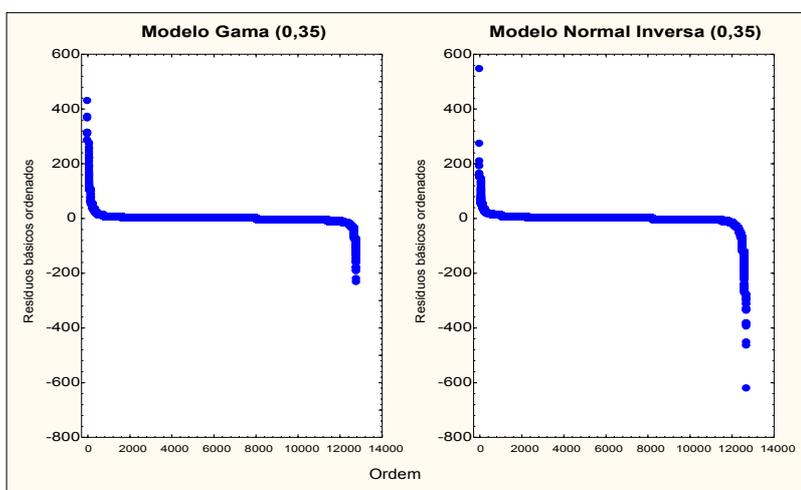


Figura 5.9: Representação gráfica dos resíduos ordinários ordenados para (a) Modelo Gama e (b) Modelo Normal Inversa.

É interessante salientar que a determinação da soma dos quadrados dos

resíduos assume um valor de 24.426.356,03g² para o modelo com a Normal Inversa, enquanto que o modelo Gama, embora construído com 155 pontos a mais que o outro modelo, apresenta um valor de 6.486.763,94g², ou seja, um total aproximadamente 3 vezes menor.

Finalmente, estabelecendo-se uma ordenação crescente para os dados observados e seus respectivos ajustes, o gráfico da sobreposição dos valores observados aos ajustados mostra uma relação com menor dispersão através do modelo Gama (Figura 5.10).

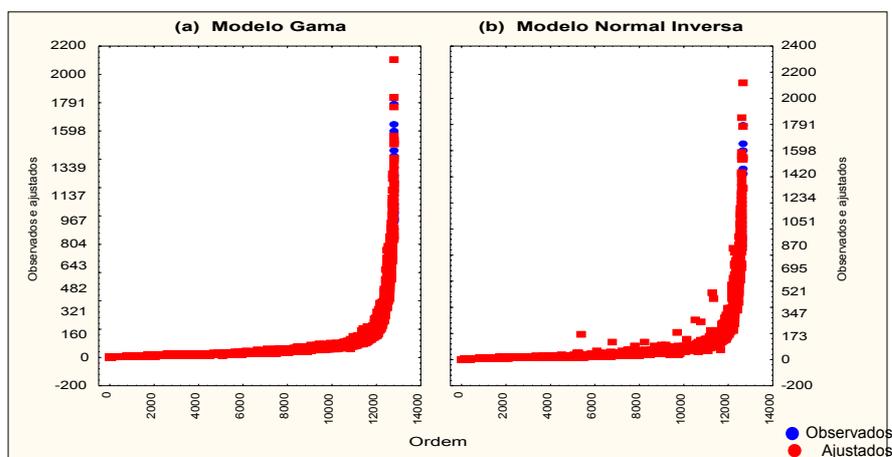


Figura 5.10: Sobreposição dos pontos ajustados pelos modelos aos valores observados, ordenados de forma crescente.

Também se pode avaliar a adequação da distribuição da variável resposta por meio da plotagem dos valores observados contra a distribuição teórica (Gama e Normal Inversa), conforme Figura 5.11.

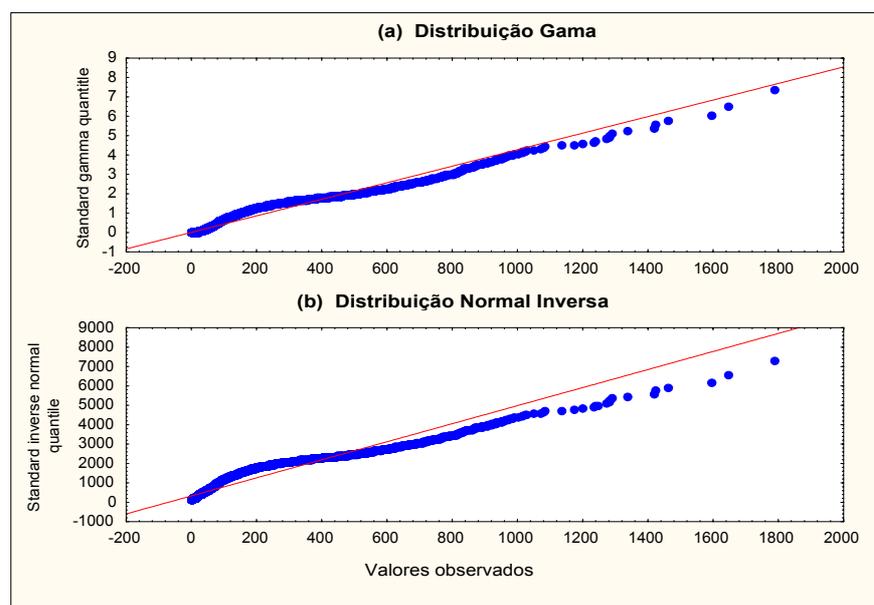


Figura 5.11: Plotagem dos valores observados contra a distribuição de probabilidade teórica.

Nota-se que, para a distribuição Gama, o gráfico tem uma disposição dos dados mais próxima de uma reta, indicando melhor ajuste dos dados observados com relação a este modelo.

Todas estas evidências indicam que o modelo mais adequado para definir a relação peso-comprimento das espécies de peixes em estudo é o modelo Gama, ou seja, os fatores de condição do peixe a serem determinados, terão como referencial o modelo de predição do peso do animal segundo o modelo Gama. Portanto, para que se obtenha a equação capaz de predizer o peso do animal conhecendo-se sua espécie, sexo e comprimento, é necessário considerar os parâmetros estimados apresentados na Tabela 5.3 e aplica-los à definição do preditor linear $\eta=g(\mu)$. A equação resultante para este preditor linear $\hat{\eta}$ é:

$$g(\hat{\mu}) = \hat{\eta} = [-0,698 + 0,362L_s + 0,490 A - 0,607 C - 0,137 P + 0,045 \text{SEXO}_f - 0,027(A \cdot \text{SEXO}_f) + 0,156(C \cdot \text{SEXO}_f) - 0,087(P \cdot \text{SEXO}_f)]^{35/100} \quad (5.1)$$

Conseqüentemente, para que se determine o vetor dos pesos médios esperados $\hat{\mu}$, é necessário que se encontre $g^{-1}(\mu)$. Dessa transformação resulta a seguinte equação de predição:

$$\hat{\mu} = [-0,698 + 0,362L_s + 0,490 A - 0,607 C - 0,137 P + 0,045 \text{SEXO}_f - 0,027(A \cdot \text{SEXO}_f) + 0,156(C \cdot \text{SEXO}_f) - 0,087(P \cdot \text{SEXO}_f)]^{100/35} \quad (5.2)$$

Isolando-se o intercepto da Equação 5.2, determina-se uma relação direta do peso e comprimento do peixe. Esta é a relação é definida como um *fator de condição K do peixe*.

A Tabela 5.12 mostra as equações de predição do peso do peixe, a equação que define o fator de condição do peixe e também a equação do fator de condição K do peixe. O processo de parametrização desenvolvido para a obtenção de tais equações está detalhado no Apêndice B.

Tabela 5.12: Equações para o peso estimado pelo modelo Gama e fator de condição do peixe correspondente.

Espécie	S e x o	Predição	Intercepto	Fator de condição
Trachydoras paraguayensis A	F	$\hat{\mu}_{AF} = (-0,190 + 0,362 \cdot Ls)^{100/35}$	$\hat{\mu}_{AF}^{0,35} - 0,362 \cdot Ls = -0,190$	$K_{AF} = \hat{\mu}_{AF}^{0,35} - 0,362 \cdot Ls$
	M	$\hat{\mu}_{AM} = (-0,226 + 0,362 \cdot Ls)^{100/35}$	$\hat{\mu}_{AM}^{0,35} - 0,362 \cdot Ls = -0,226$	$K_{AM} = \hat{\mu}_{AM}^{0,35} - 0,362 \cdot Ls$
Liposarcus anisitsi C	F	$\hat{\mu}_{CF} = (-1,104 + 0,362 \cdot Ls)^{100/35}$	$\hat{\mu}_{CF}^{0,35} - 0,362 \cdot Ls = -1,104$	$K_{CF} = \hat{\mu}_{CF}^{0,35} - 0,362 \cdot Ls$
	M	$\hat{\mu}_{CM} = (-1,505 + 0,362 \cdot Ls)^{100/35}$	$\hat{\mu}_{CM}^{0,35} - 0,362 \cdot Ls = -1,505$	$K_{CM} = \hat{\mu}_{CM}^{0,35} - 0,362 \cdot Ls$
Serrasalmus marginatus H	F	$\hat{\mu}_{HF} = (-0,441 + 0,362 \cdot Ls)^{100/35}$	$\hat{\mu}_{HF}^{0,35} - 0,362 \cdot Ls = -0,441$	$K_{HF} = \hat{\mu}_{HF}^{0,35} - 0,362 \cdot Ls$
	M	$\hat{\mu}_{HM} = (-0,447 + 0,362 \cdot Ls)^{100/35}$	$\hat{\mu}_{HM}^{0,35} - 0,362 \cdot Ls = -0,447$	$K_{HM} = \hat{\mu}_{HM}^{0,35} - 0,362 \cdot Ls$
Leporinus friderici P	F	$\hat{\mu}_{PF} = (-0,877 + 0,362 \cdot Ls)^{100/35}$	$\hat{\mu}_{PF}^{0,35} - 0,362 \cdot Ls = -0,877$	$K_{PF} = \hat{\mu}_{PF}^{0,35} - 0,362 \cdot Ls$
	M	$\hat{\mu}_{PM} = (-0,793 + 0,362 \cdot Ls)^{100/35}$	$\hat{\mu}_{PM}^{0,35} - 0,362 \cdot Ls = -0,793$	$K_{PM} = \hat{\mu}_{PM}^{0,35} - 0,362 \cdot Ls$

A aplicação do fator de condição aos dados observados, descrito pela relação peso-comprimento, fornece um conjunto de valores com a seguinte distribuição de frequência:

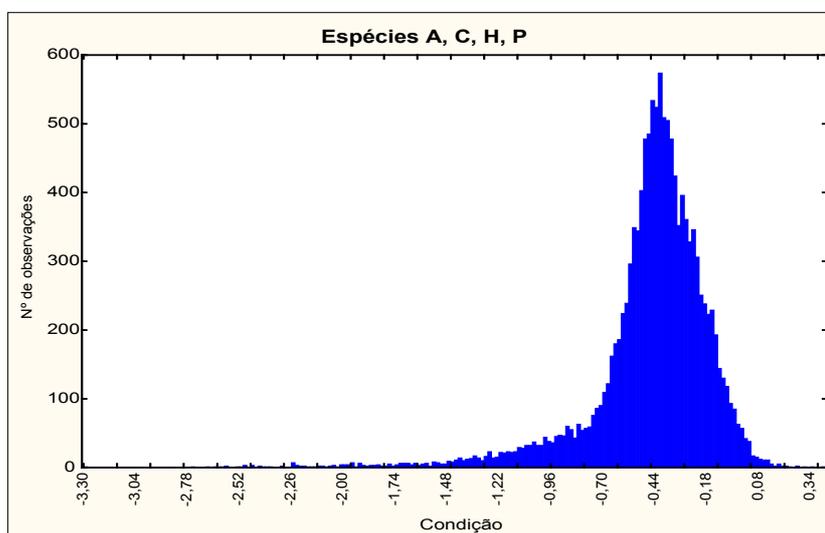


Figura 5.12: Histograma para o fator de condição para os peixes das espécies A, C, H e P.

Pela análise gráfica, observa-se uma forte assimetria à esquerda ao se trabalhar com o modelo geral, desconsiderando-se a individualidade das espécies. A Tabela 5.13 expõe as medidas descritivas para o fator de condição geral do peixe.

Tabela 5.13: Medidas descritivas para o fator de condição determinado.

Média	Mediana	Moda	Mínimo	Máximo	Desvio Padrão	Assimetria	Curtose
-0,4552	-0,4122	-0,4029	-3,3028	0,4193	0.3039	-2.1987	8.8785

Para identificar de maneira mais precisa a existência de casos em que a assimetria seja mais forte, pode-se analisar o comportamento do fator de condição do peixe para cada espécie de acordo com o sexo. A Tabela 5.14 expõe as características descritivas sob tais condições.

Tabela 5.14: Medidas descritivas para o fator de condição do peixe determinado para cada espécie classificada por sexo.

Espécie	N	Média	Mediana	Mínimo	Máximo	Desvio Padrão	Assimetria	Curtose
Fêmea A	1.650	-0,197	-0,195	-0,678	0,288	0,138	-0,032	0,251
Macho A	1.445	-0,239	-0,233	-0,698	0,191	0,127	-0,231	0,238
Fêmea C	190	-1,200	-1,182	-2,627	-0,061	0,513	-0,199	-0,157
Macho C	165	-1,610	-1,607	-3,303	-0,407	0,543	-0,240	-0,426
Fêmea H	4.089	-0,428	-0,429	-1,356	0,419	0,149	0,155	1,954
Macho H	4.090	-0,442	-0,439	-1,410	0,157	0,140	-0,148	1,528
Fêmea P	789	-0,964	-0,923	-2,253	-0,060	0,261	-0,889	2,329
Macho P	382	-0,896	-0,869	-1,776	-0,361	0,229	-0,545	0,631

Nota: Espécie A, *Trachydoras paraguayensis*. Espécie C, *Liposarcus anisiti*. Espécie D, *Raphiodon vulpinus*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*.

Apenas as fêmeas da espécie *Serrasalmus marginatus* apresentam assimetria positiva. As demais espécies apresentam assimetria negativa. Quanto à curtose, somente a espécie *Liposarcus anisiti* pode ser classificada como platicúrtica. As outras espécies aproximam-se mais de uma distribuição leptocúrtica.

A Figura 5.13 mostra os histogramas das distribuições de frequência para cada uma das espécies de peixe classificadas quanto ao seu sexo. A

observação de tais gráficos confirma o comportamento levemente assimétrico em cada situação, conforme já avaliado na Tabela 5.14

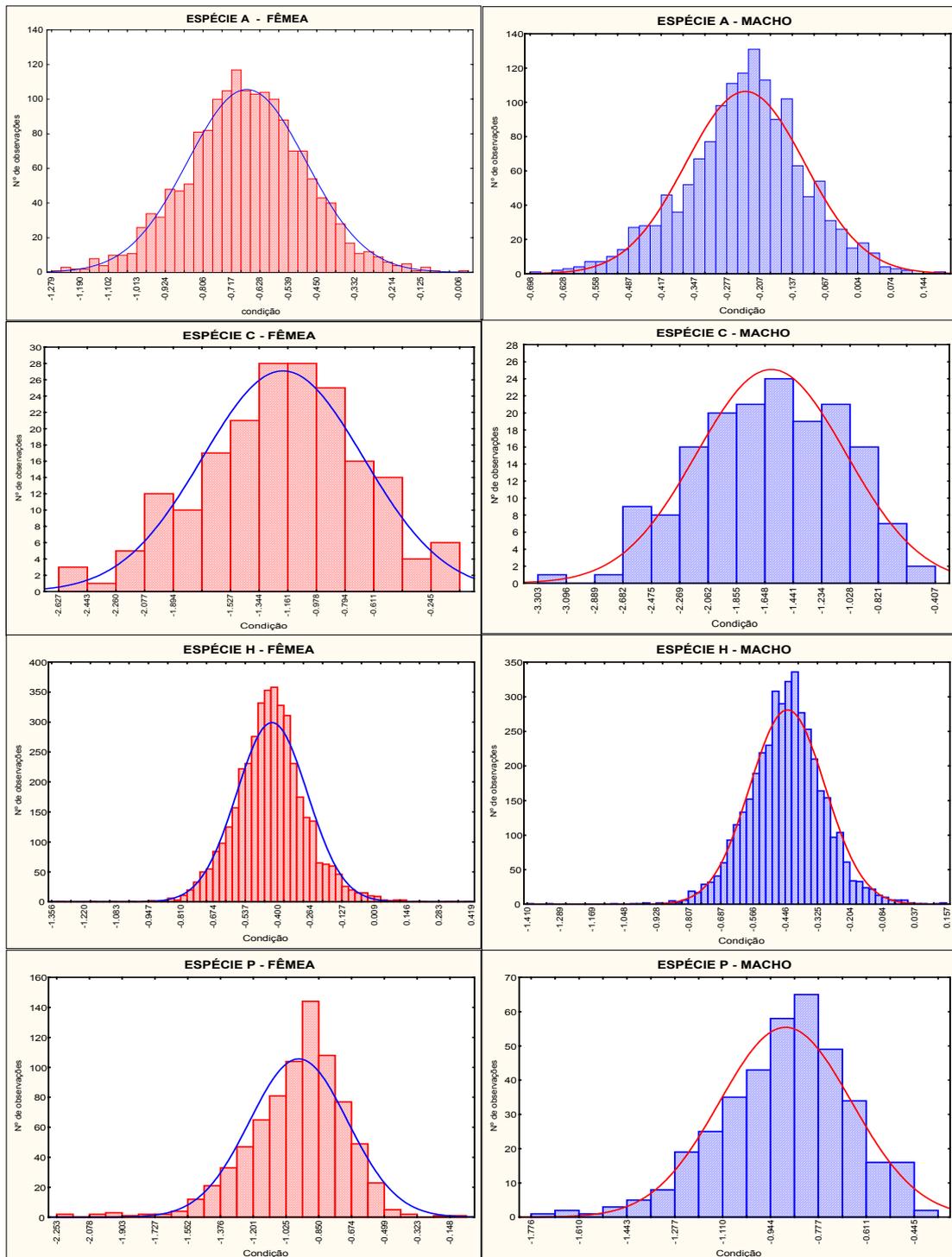


Figura 5.13: Fator de condição para cada espécie identificadas de acordo com seu sexo.

A análise destes gráficos apresenta comportamento bastante próximo de uma distribuição normal. Tal distribuição também pode ser avaliada pela plotagem dos valores observados contra as probabilidades normais dos

valores esperados, conforme Figura 5.14.

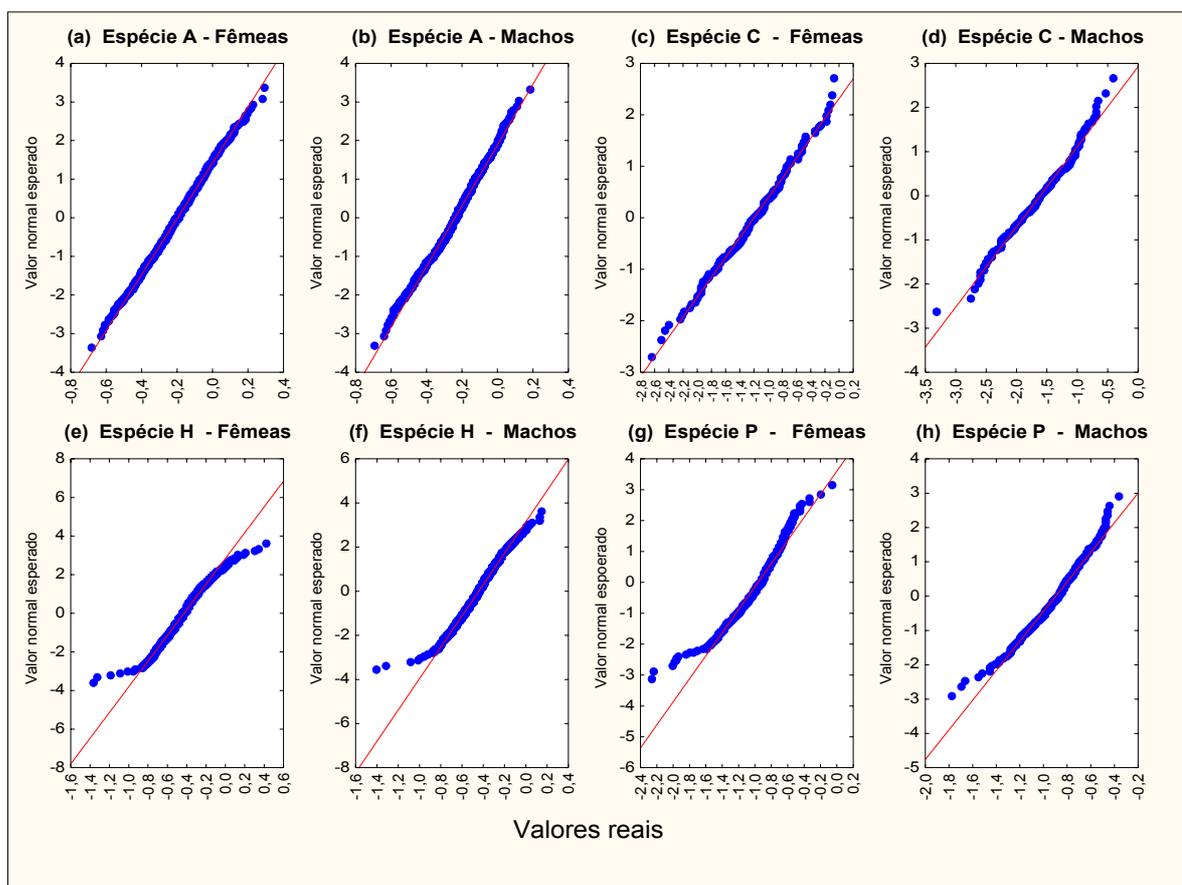


Figura 5.14: Plotagem dos valores observados contra os valores esperados pela probabilidade normal.

Nota: Espécie A, *Trachydoras paraguayensis*. Espécie C, *Liposarcus anisiti*. Espécie D, *Raphiodon vulpinus*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*.

Observa-se que para as espécies H e P, os gráficos não apresentam comportamento padrão, ou seja, plotagem aproximada de uma reta, o que indica uma inadequação à distribuição normal, em especial quanto ao comportamento dos pontos extremos.

No tópico 5.2, serão investigados novos modelos na expectativa de uma melhor distribuição para o fator de condição.

5.2 CONSTRUÇÃO DE MODELOS POR ESPÉCIES

Ao investigar um modelo geral para descrever o peso dos animais, além da variável comprimento, verificou-se que a espécie e o sexo também são relevantes na construção do modelo. Buscando uma modelagem mais apurada e cujo resultado simplifique a aplicação dos dados, neste tópico será

construído um modelo para cada espécie, tendo como variáveis preditoras a covariável comprimento padrão e o fator sexo. Caso se confirme a significância do fator sexo, outro modelo será construído, tomando separadamente a espécie por sexo. Depois de determinado um modelo com bom ajuste dos dados, então é estabelecido o modelo para o fator de condição e sua distribuição.

5.2.1 Distribuição da variável resposta peso total (W_t)

A variável peso total caracteriza-se por valores contínuos não negativos, e como demonstrado pela Figura 5.15, com comportamento assimétrico positivo bastante acentuado para todas as espécies.

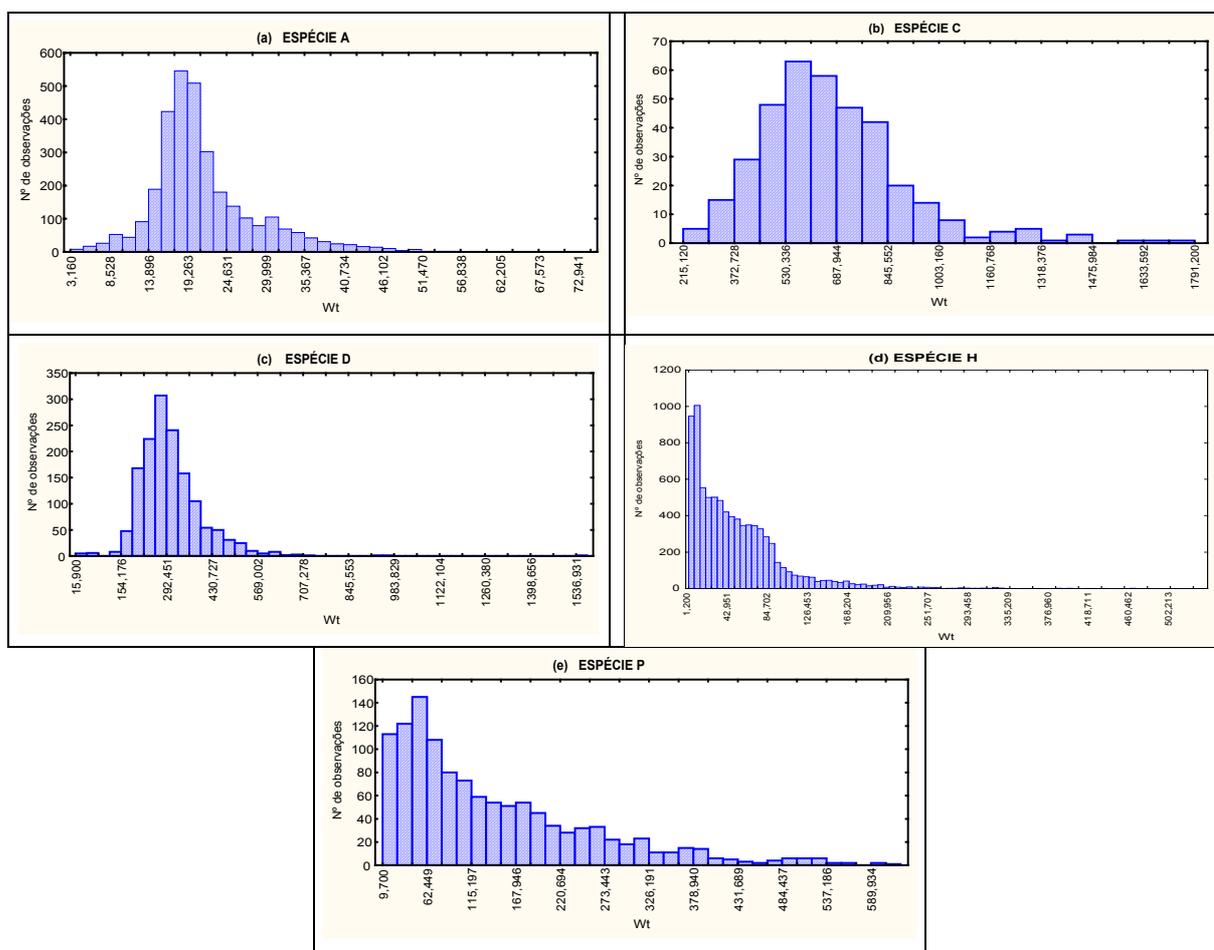


Figura 5.15: Histogramas da variável peso total para as espécies em estudo.

Nota: Espécie A, *Trachydoras paraguayensis*. Espécie C, *Liposarcus anisiti*. Espécie D, *Raphiodon vulpinus*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*.

As medidas descritivas das espécies, também identificadas por sexo, expostas na Tabela 5.15, confirmam tal assimetria bem como identificam nas espécies D, H e P um maior grau de dispersão. Para estas espécies, o coeficiente de variação é de 59,77%, 97,44% e 82,52%, respectivamente.

Tabela 5.15: Medidas descritivas para as espécies pesquisadas.

Espécie	Sexo	Nº de casos	Média	Mediana	Moda	Mínimo	Máximo	Desvio padrão	Assimetria	Curtose
A	Geral	3.117	21,49	19,80	20,00	3,16	74,73	7,42	1,35	3,16
	Fêmea	1.665	24,17	3,20	74,73	3,20	74,73	8,54	0,06	1,42
	Macho	1.452	18,43	3,16	44,30	3,16	44,30	4,12	0,06	3,33
C	Geral	367	672,36	635,80	Múltipla	215,12	1.791,2	230,83	1,31	3,19
	Fêmea	200	598,73	566,15	Múltipla	215,12	1.280,60	182,54	1,06	1,93
	Macho	167	760,54	732,90	737,80	232,37	1.791,20	251,36	1,25	2,87
D	Geral	2.161	368,15	312,70	Múltipla	15,90	3.000,00	220,06	4,38	29,80
	Fêmea	699	501,89	418,90	Múltipla	59,70	3.000,00	321,10	3,02	12,98
	Macho	1.462	304,21	289,83	309,90	15,90	1.571,50	98,43	2,55	21,94
H	Geral	8.340	50,07	37,90	10,00	1,10	538,00	48,79	2,32	9,44
	Fêmea	4.179	55,60	37,20	10,00	1,10	538,00	58,61	2,25	7,52
	Macho	4.161	44,52	38,60	10,00	1,20	214,60	35,50	1,06	1,26
P	Geral	1.190	141,00	104,22	45,40	9,70	625,10	116,36	1,34	1,58
	Fêmea	802	165,00	135,50	Múltipla	9,70	625,10	125,31	1,09	0,76
	Macho	388	91,39	68,05	79,80	10,8	405,50	73,710	1,55	2,37

Nota: Espécie A, *Trachydoras paraguayensis*. Espécie C, *Liposarcus anisiti*. Espécie D, *Raphiodon vulpinus*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*.

Observa-se que existe uma forte correlação positiva entre a variável resposta, peso total, e a covariável comprimento padrão, conforme Tabela 5.16.

Tabela 5.16: Correlações da variável W_t com a covariável L_s .

VARIÁVEL	Espécie A	Espécie C	Espécie D	Espécie H	Espécie P
L_s	0,89	0,89	0,90	0,91	0,94

Nota: Espécie A, *Trachydoras paraguayensis*. Espécie C, *Liposarcus anisiti*. Espécie D, *Raphiodon vulpinus*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*.

Como já foi citado no tópico 5.1.1, de acordo com Cordeiro (1997, p.2), “as distribuições Gama e Normal Inversa são associadas a dados contínuos assimétricos e positivos”. As características aqui apresentadas conduzem a aceitação da distribuição Gama como a distribuição da variável W_t para todas as espécies em estudo. Posteriormente, alguns testes serão realizados para detectar a adequação deste procedimento.

5.2.2 Identificação da função de ligação

Como no item 5.1.2, o procedimento mais simples sugerido para a identificação da possível função de ligação do MLG é a construção do diagrama entre a variável resposta e sua covariável, L_s (covariável única).

A Figura 5.16 apresenta os respectivos diagramas identificando a relação por espécie e sexo do animal. Desta construção gráfica, tem-se a função potência como indicação da mais provável função de ligação para os modelos.

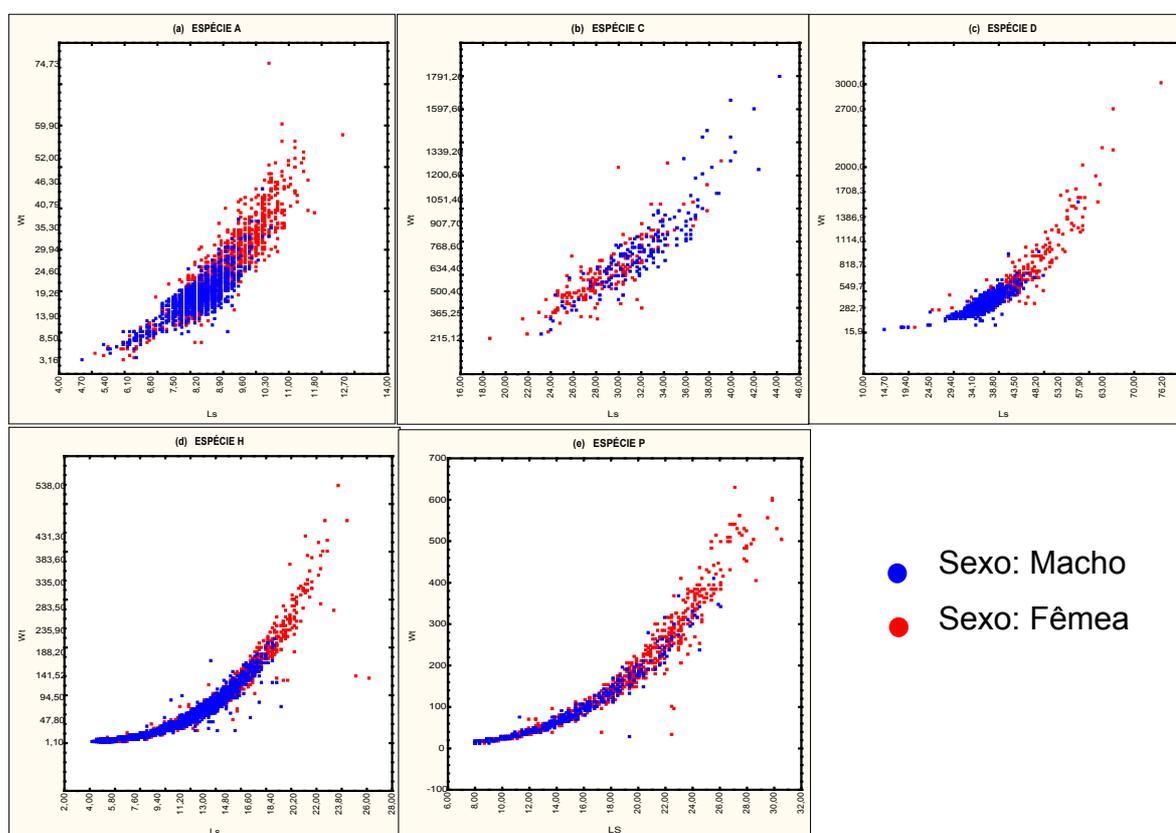


Figura 5.16: Dispersão entre a variável peso total e comprimento padrão para a espécie A, *Trachydoras paraguayensis*. Espécie C, *Liposarcus anisiti*. Espécie D, *Raphiodon vulpinus*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*.

5.2.3 Construção dos modelos

Vários modelos foram testados, porém só serão expostos os que melhor satisfizeram os pressupostos básicos dos MLGs. Para a escolha destes modelos, partiu-se da comparação das deviances resultantes, bem como a análise dos respectivos resíduos, os modelos propostos para cada espécie foram:

Distribuição: Gama

Ligação: Espécie A: Potência (0,30); Espécie C: Potência (0,32); Espécie D: Potência (0,23); Espécie H: Potência (0,31); Espécie P: Potência (0,31)

Preditor Linear: $\eta = \alpha + Ls + \text{Sexo}$

A Tabela 5.17 apresenta as deviances e suas respectivas estatísticas para os modelos propostos para cada espécie. Observa-se que para todas as espécies, a deviance escalar tem valores próximos ao tamanho de cada amostra, o que indica boa qualidade dos modelos sugeridos.

Tabela 5.17: Deviances dos modelos propostos para cada espécie.

Espécie	Estatística	Graus de Liberdade (gl)	Estatística	Estatística/gl
A	Deviance	3.114	59,375	0,019
	Deviance escalar	3.114	3.126,864	1,004
C	Deviance	364	7,721	0,021
	Deviance escalar	364	368,282	1,012
D	Deviance	2.158	48,516	0,023
	Deviance escalar	2.158	2.169,056	1,005
H	Deviance	8.337	187,732	0,023
	Deviance escalar	8.337	8.371,171	1,004
P	Deviance	1.187	23,710	0,020
	Deviance escalar	1.187	1.193,943	1,006

Nota: Espécie A, *Trachydoras paraguayensis*. Espécie C, *Liposarcus anisiti*. Espécie D, *Raphiodon vulpinus*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*.

Pela Tabela 5.18, verifica-se que para as espécies A e D, o comprimento padrão, intercepto e sexo são significativos, porém para as espécies C, H e P, o fator sexo não é significativo.

Tabela 5.18: Parâmetros estimados para o modelo proposto para a espécie A, C, D, H e P

Espécie	Efeitos	Valores Estimados	Estatística de Wald	p	Limite inferior	Limite superior
A	Intercepto	0,330	332,41	0,00	0,295	0,366
	Ls	0,251	13.738,14	0,00	0,247	0,255
	Sexo f	0,018	87,34	0,00	0,014	0,021
C	Intercepto	1,891	127,91	0,00	1,563	2,219
	Ls	0,199	1.291,74	0,00	0,188	0,210
	Sexo f	0,034	2,62	0,11	-0,007	0,075

Continua

Tabela 5.18: Parâmetros estimados para o modelo proposto para a espécie A, C, D, H e P

D	Intercepto	1,100	1.814,95	0,00	1,050	1,151
	Ls	0,073	11.398,53	0,00	0,072	0,074
	Sexo f	0,020	35,73	0,00	0,014	0,027
H	Intercepto	-0,089	531,56	0,00	-0,097	-0,082
	Ls	0,285	462.906,67	0,00	0,285	0,286
	Sexo f	-0,000	0,02	0,88	-0,097	-0,082
P	Intercepto	0,056	8,41	0,00	0,018	0,093
	Ls	0,262	41.757,50	0,00	0,260	0,265
	Sexo f	0,000	0,00	0,97	0,018	0,093

Nota: Espécie A, *Trachydoras paraguayensis*. Espécie C, *Liposarcus anisiti*. Espécie D, *Raphiodon vulpinus*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*.

A partir destes resultados, novos modelos foram propostos: para as espécies C, H e P, onde o sexo não foi considerado significativo, trabalhou-se o conjunto de dados desconsiderando-se o fator sexo. Já para as espécies onde este fator foi relevante, espécies A e D, na busca de ajustes mais parcimoniosos, optou-se por considerar modelos individuais para cada espécie, agrupando-os separadamente por fêmeas e machos. O grau da função de ligação proposto para as fêmeas da espécie A (*Trachydoras paraguayensis*) é de 0,4 enquanto que para os machos dessa mesma espécie é de 0,73. Para a espécie D (*Raphiodon vulpinus*) o grau proposto para as fêmeas foi de 0,13 enquanto que para os machos o grau foi de 0,30. Conseqüentemente, o preditor linear de todos os modelos trabalhados passa a ser $\eta = \alpha + Ls$.

A próxima etapa do trabalho foi a identificação dos pontos atípicos ao modelo, cujo processo encontra-se definido neste trabalho no tópico 5.1.3. A Figura 5.17 mostra graficamente a relação peso-comprimento de acordo com os modelos escolhidos desconsiderando os pontos atípicos.

Para a espécie C, *Liposarcus anisitsi*, detectou-se 6 pontos atípicos, aproximadamente 1,6% do total da amostra; quanto à espécie H, *Serrasalmus marginatus*, de um total de 8.340 casos, retirou-se 145 observações atípicos, cerca de 1,7% da amostra; para a espécie P, *Leporinus friderici*, retirou-se 12 pontos atípicos, o que corresponde a aproximadamente 1,0% do total da amostra.

Para as fêmeas da espécie A, *Trachydoras paraguayensis*, o total de

pontos atípicos ao modelo em construção corresponde a 16 observações, enquanto que para os machos, 11. Tais retiradas correspondem respectivamente a 0,96% e 0,76% do total inicial do tamanho da amostra. Quanto às fêmeas da espécie D, *Raphiodon vulpinus*, foram detectados 15 pontos atípicos ao modelo e para os machos, 13 pontos, correspondendo respectivamente a 2,1% e 0,9% das observações para cada sexo.

Mesmo com a retirada dos pontos atípicos, ainda se tem amostras de tamanhos significativamente grandes com relação a cada espécie. Portanto, é justificável a continuidade do ajuste dos dados desconsiderando-se os pontos atípicos na busca de modelos mais representativos.

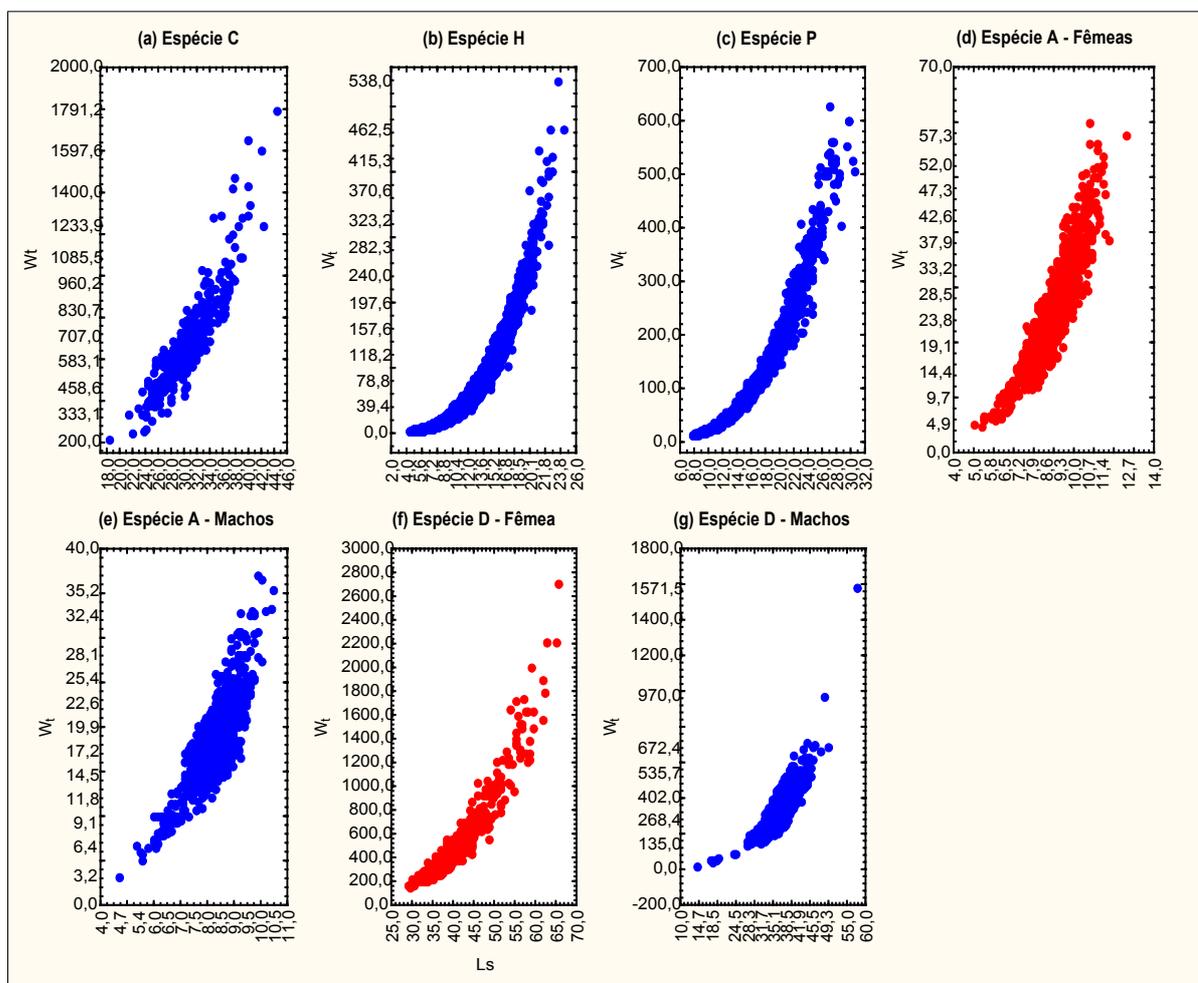


Figura 5.17: Diagramas das variáveis comprimento padrão e peso total sem casos atípicos para as espécies C, *Liposarcus anisiti*; espécie H, *Serrasalmus marginatus*; espécie P, *Leporinus friderice*; machos e fêmeas da espécie *Trachydoras paraguayensis* e machos e fêmeas da espécie *Raphiodon vulpinus*.

Na Tabela 5.19 estão apresentadas as deviances para estes modelos segundo o novo conjunto de dados.

Tabela 5.19: Deviance dos modelos das espécies C, H e P sem os pontos atípicos.

Espécie	Deviances	Graus de Liberdade (GL)	Estatística	Estatística/GL	Diferença
C	Deviance	359	6,0511	0,0169	1,67
	Deviance escalar	359	362,0057	1,0084	
H	Deviance	8.193	127,6511	0,0156	60,08
	Deviance escalar	8.193	8.216,2197	1,0028	
P	Deviance	1.176	13,8687	0,0118	9,87
	Deviance escalar	1.176	1.180,3069	1,0037	
A _F	Deviance	1.647	28,0500	0,0170	5,69
	Deviance escalar	1.647	1.653,6617	1,0040	
A _M	Deviance	1.439	19,9293	0,0138	3,66
	Deviance escalar	1.439	1.444,3139	1,0037	
D _F	Deviance	682	11,3069	0,0166	5,81
	Deviance escalar	682	685,8793	1,0057	
D _M	Deviance	1.447	23,2663	0,0161	7,65
	Deviance escalar	1.447	1.452,8673	1,0041	

Nota: Espécie C, *Liposarcus anisiti*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*. Espécie A_F, *Trachydoras paraguayensis* fêmea e A_M, *Trachydoras paraguayensis* macho. Espécie D_F, *Raphiodon vulpinus* fêmea e D_M, *Raphiodon vulpinus* macho.

Após a retirada dos pontos atípicos, houve queda na deviance para todos os casos em estudo, enquanto que a deviance escalonada apresenta valores ainda mais próximos ao tamanho das amostras. Estes indícios apontam para uma boa adequação do ajuste dos dados ao modelo proposto.

O cálculo das estimativas dos parâmetros e as respectivas estatísticas do teste de Wald confirmam o alto grau de significância das mesmas, pois para elevados valores dessas estatísticas, observa-se todos os valores de p estão muito próximos de zero. Quanto aos intervalos de confiança das estimativas, foram construídos tomando-se um nível de confiança de 95%. A Tabela 5.20 apresenta estes resultados.

Tabela 5.20: Parâmetros estimados e seus respectivos intervalos de confiança segundo modelo proposto para cada espécie classificada segundo o sexo do peixe.

Espécie	Efeitos	Valores Estimados	Estatística de Wald	p	Limite inferior	Limite superior
C	Intercepto	2,0003	214,11	0,00	1,7324	2,2682
	Ls	0,1955	1.843,26	0,00	0,1866	0,2044
H	Intercepto	-0,0893	724,56	0,00	-0,0958	-0,0828
	Ls	0,2857	642.661,33	0,00	0,2850	0,2864

Continua

Tabela 5.20: Parâmetros estimados e seus respectivos intervalos de confiança segundo modelo proposto para cada espécie classificada segundo o sexo do peixe.

P	Intercepto	0,1491	122,78	0,00	0,1227	0,1755
	Ls	0,2443	76.621,61	0,00	0,2426	0,2460
A_F	Intercepto	-0,6649	286,84	0,00	-0,7418	-0,5879
	Ls	0,4738	10.930,67	0,00	0,4649	0,4827
A_M	Intercepto	-5,9606	1.189,46	0,00	-6,2993	-5,6219
	Ls	1,7297	6.507,99	0,00	1,6877	1,7717
D_F	Intercepto	1,28478	15.508,85	0,00	1,2646	1,3050
	Ls	0,02275	7.980,56	0,00	0,0222	0,0232
D_M	Intercepto	0,3476	40,61	0,00	0,2407	0,4545
	Ls	0,1437	8.765,13	0,00	0,1407	0,1467

Nota: Espécie C, *Liposarcus anisiti*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*. Espécie A_F, *Trachydoras paraguayensis* fêmeas e A_M *Trachydoras paraguayensis* machos. Espécie D_F, *Raphiodon vulpinus* fêmeas e D_M, *Raphiodon vulpinus* macho.

O teste tipo 1 (calculado pelo software Statistica), consiste de um ajuste seqüencial do modelo, iniciando do modelo nulo, ou seja, considerando apenas o intercepto, até chegar ao modelo especificado. Nele, calcula-se a razão de verossimilhança para cada par de sucessivos modelos. Embora já tenhamos o cálculo do teste Wald, pelo teste da razão da verossimilhança reafirma-se a adequação do modelo proposto. Os altos valores do Chi-quadrado confirmam a significância das estimativas dos parâmetros.

Tabela 5.21: Teste tipo 1 para os parâmetros estimados através dos modelos propostos para as espécies C, H e P.

Espécie	Efeitos	Graus de liberdade	Log-verossimilhança	χ^2	p
C	Ls	1	-2.440,4	666,28	0,00
	Sexo	1	2.107,3		
H	Ls	1	-40.289,7	35.083,04	0,00
	Sexo	1	22.748,2		
P	Ls	1	-6.951,54	4.946,42	0,00
	Sexo	1	-4.478,33		
A _F	Ls	1	-5.753,03	3.224,85	0,00
	Sexo	1	-4.140,61		
A _M	Ls	1	-4.078,32	1.906,54	0,00
	Sexo	1	-3.125,05		
D _F	Ls	1	-4.678,60	1.892,97	0,00
	Sexo	1	-3.732,12		
D _M	Ls	1	-8.547,32	2.525,35	0,00
	Sexo	1	-7.284,64		

Legenda: Espécie C, *Liposarcus anisiti*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*. Espécie A_F, *Trachydoras paraguayensis* fêmeas e A_M *Trachydoras paraguayensis* machos. Espécie D_F, *Raphiodon vulpinus* fêmeas e D_M, *Raphiodon vulpinus* machos.

Uma das condições para se trabalhar com os modelos lineares generalizados é que a aceitação do modelo proposto não pode basear-se unicamente no cálculo e interpretação das deviances ou pelo nível de significância dos estimadores (técnicas formais). É necessária a análise gráfica dos resíduos (técnicas informais) para que haja a corroboração dos resultados e conclusões e, em especial, para a verificação de possíveis afastamentos das suposições feitas para o modelo, especialmente para a parte aleatória e para a parte sistemática do modelo.

Portanto, a próxima etapa do trabalho será a realização de tais análises, trabalhando-se com os gráficos dos mesmos.

i) Resíduos x Valores ajustados (ou alguma função dos valores ajustados)

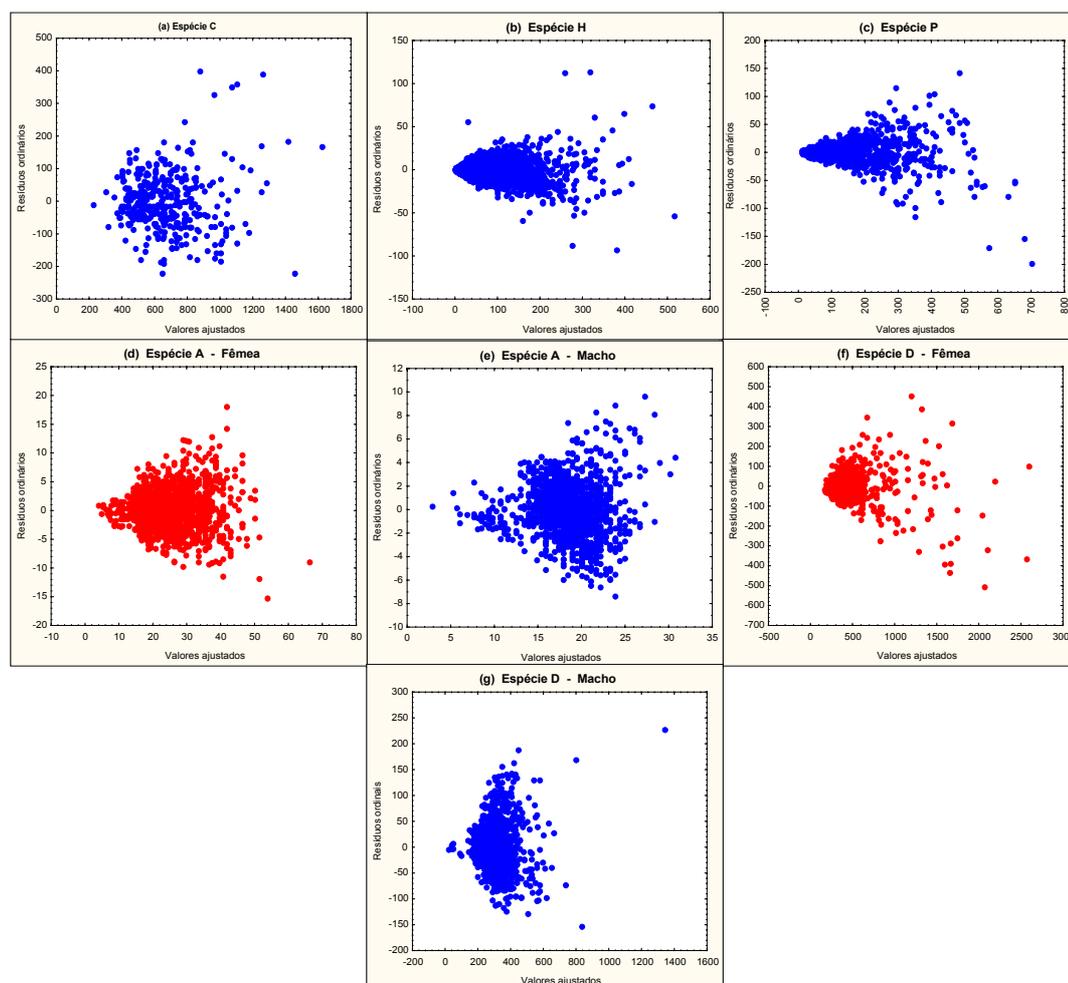


Figura 5.18: Dispersão entre os valores ajustados e seus resíduos ordinários.

Este tipo de gráfico é utilizado para verificar o comportamento da variável avaliando sua respectiva variância. Primeiramente serão grafados os

resíduos ordinários pelos valores ajustados. Através deste procedimento pretende-se comparar, posteriormente, o mesmo tipo de gráfico, substituindo os resíduos ordinários por resíduos padronizados.

Em todos os gráficos da Figura 5.18, identificam-se evidências do aumento dos resíduos conforme crescem os valores ajustados. Dobson (2002, p.33) propõe a utilização de um tipo de resíduo padronizado. De acordo com McCullagh e Nelder, é aconselhável a estabilização da variância através de alguma transformação dos valores ajustados. A transformação na variável resposta sugerida por Demétrio (2001, p.72) para tornar constante a variância da distribuição Gama é $2\ln(\hat{\mu})$. A Figura 5.19 apresenta os valores ajustados pelos resíduos padronizados da deviance, enquanto a Figura 5.20 mostra o comportamento da variável ajustada transformada pelo mesmo tipo de resíduos do caso anterior.

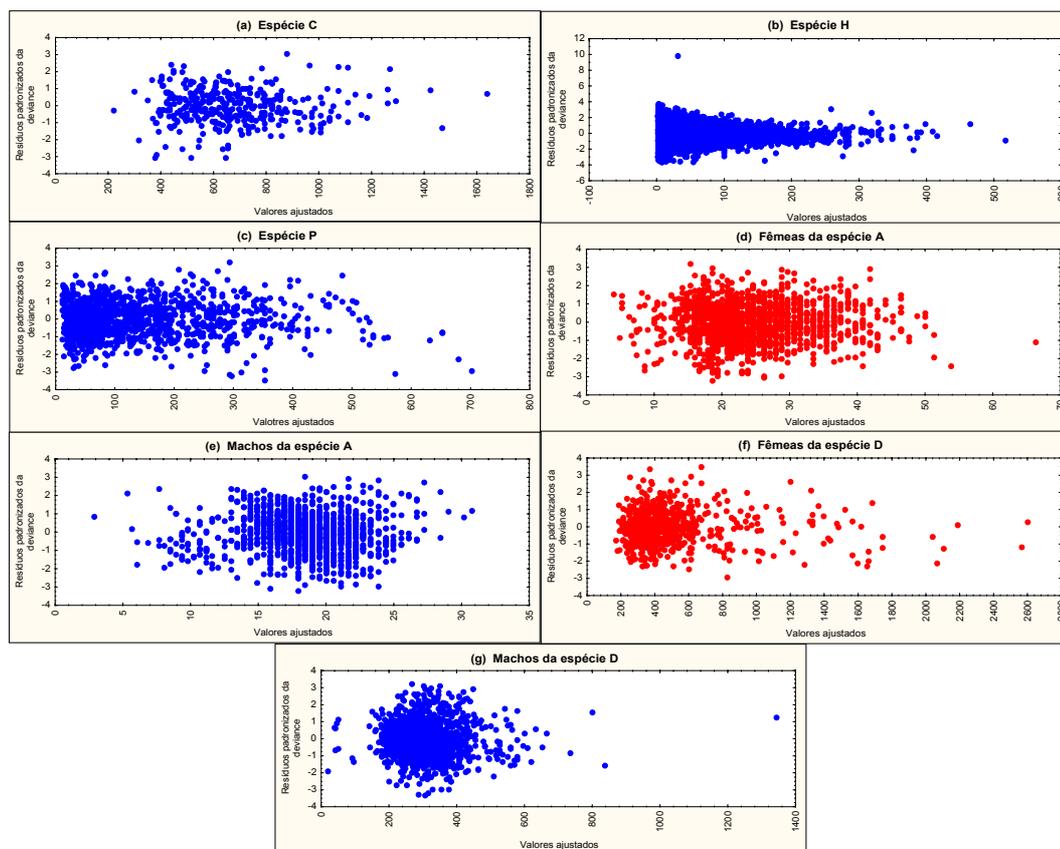


Figura 5.19: Gráficos dos valores ajustados pelos resíduos padronizados da deviance.

Embora tenha havido mudança no comportamento gráfico com a alteração para o resíduo padronizado da deviance, conforme sugerido, é de interesse

verificar este mesmo resíduo plotado contra a variável ajustada transformada, conforme Figura 5.20.

A análise dos gráficos que se encontram na Figura 5.20, mostra que, mesmo após a transformação da variável ajustada no caso da espécie H, a dispersão dos resíduos indica que a variância dos mesmos diminui com o aumento da média da variável resposta, ou seja, a média do peso total do peixe e, conseqüentemente, de acordo com Amaral (1997, p.103), “a função de variância $V(\mu) = \mu^2$ pode não ser adequada”.

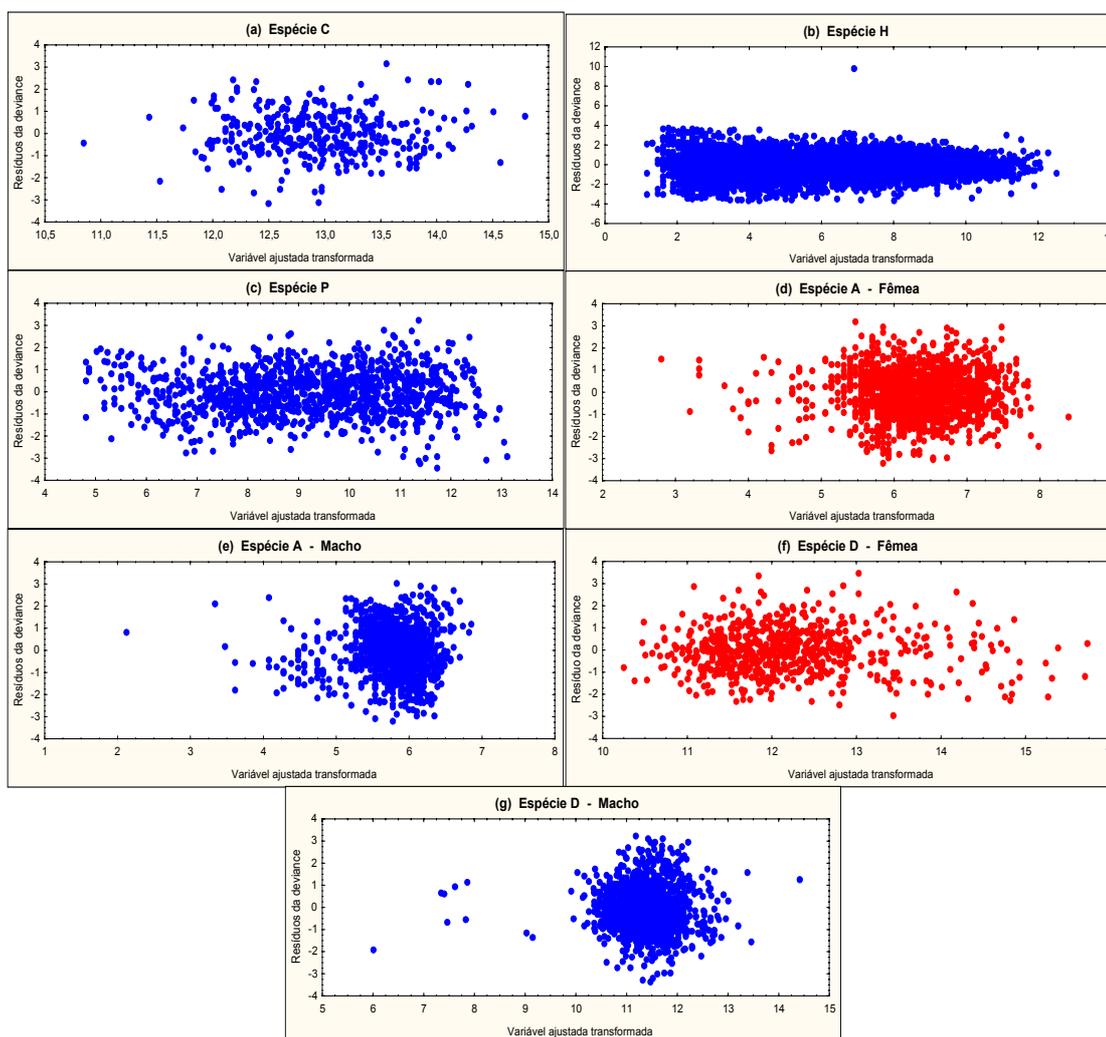


Figura 5.20: Dispersão dos valores ajustados transformados e resíduos padronizados da deviance

ii) *Resíduos x Ordem das observações*

O principal objetivo deste gráfico é verificar a independência entre as observações; mesmo que o tempo não seja uma variável incluída no modelo,

é aconselhável a aplicação deste gráfico, pois ele pode levar a identificação de padrões suspeitos, devido ou não à variável tempo. Em situações onde tal fato não ocorre, seu padrão é uma distribuição dos resíduos em torno do zero, com amplitude constante.

Na Figura 5.21 tem-se os gráficos que representam tais variáveis para cada modelo estudado.

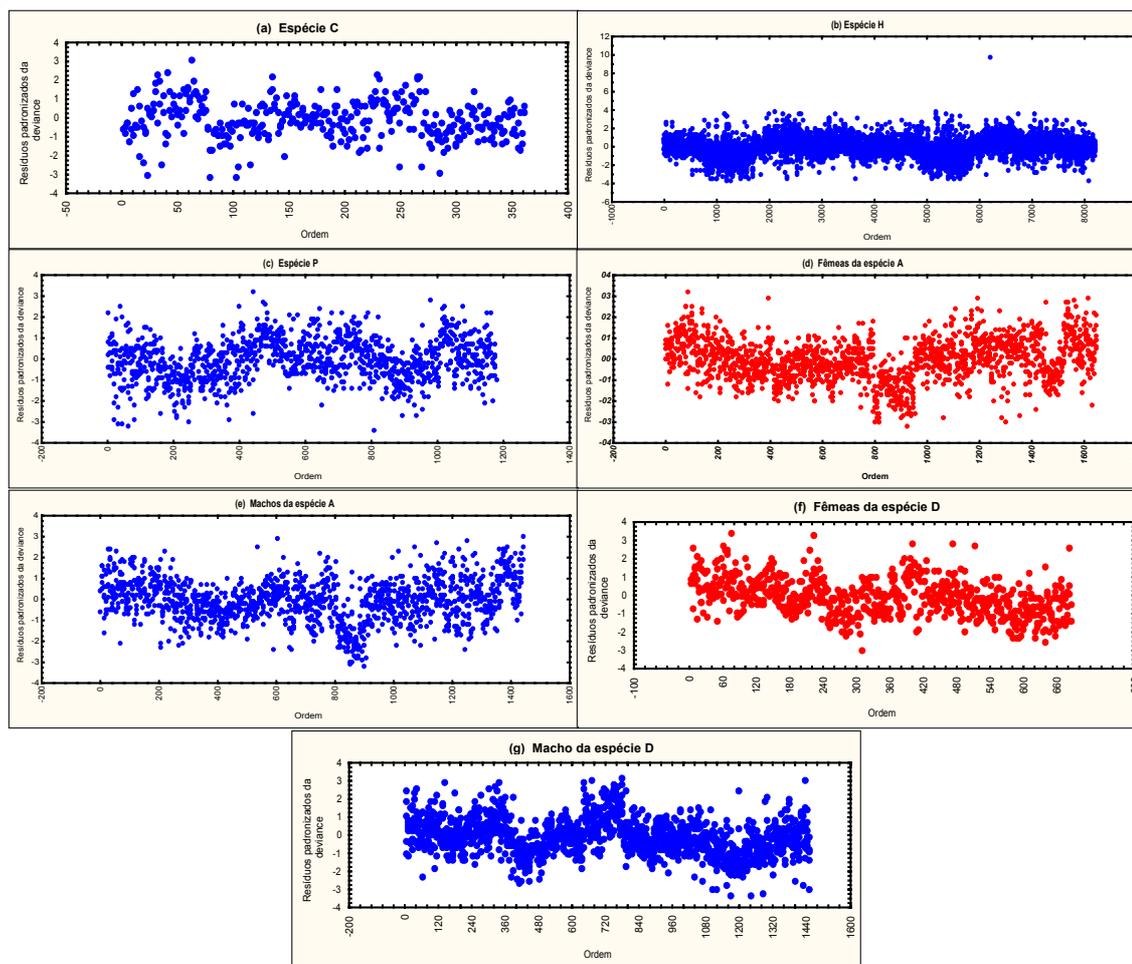


Figura 5.21: Dispersão entre a ordem das observações e os resíduos padronizados da deviance.

Os gráficos correspondentes ao modelo de cada espécie não apresentam padrões sistemáticos. Considera-se, portanto, que existe independência entre as observações.

iii) Valores observados x Valores ajustados

Neste gráfico, quando a disposição dos pontos se aproxima da 1ª bissetriz, ou seja, tem o comportamento da reta $y=x$, de acordo com Amaral

(1997, p.97), é um indicativo de que os dados estão bem ajustados ao modelo, particularmente, auxilia a avaliação da função de ligação.

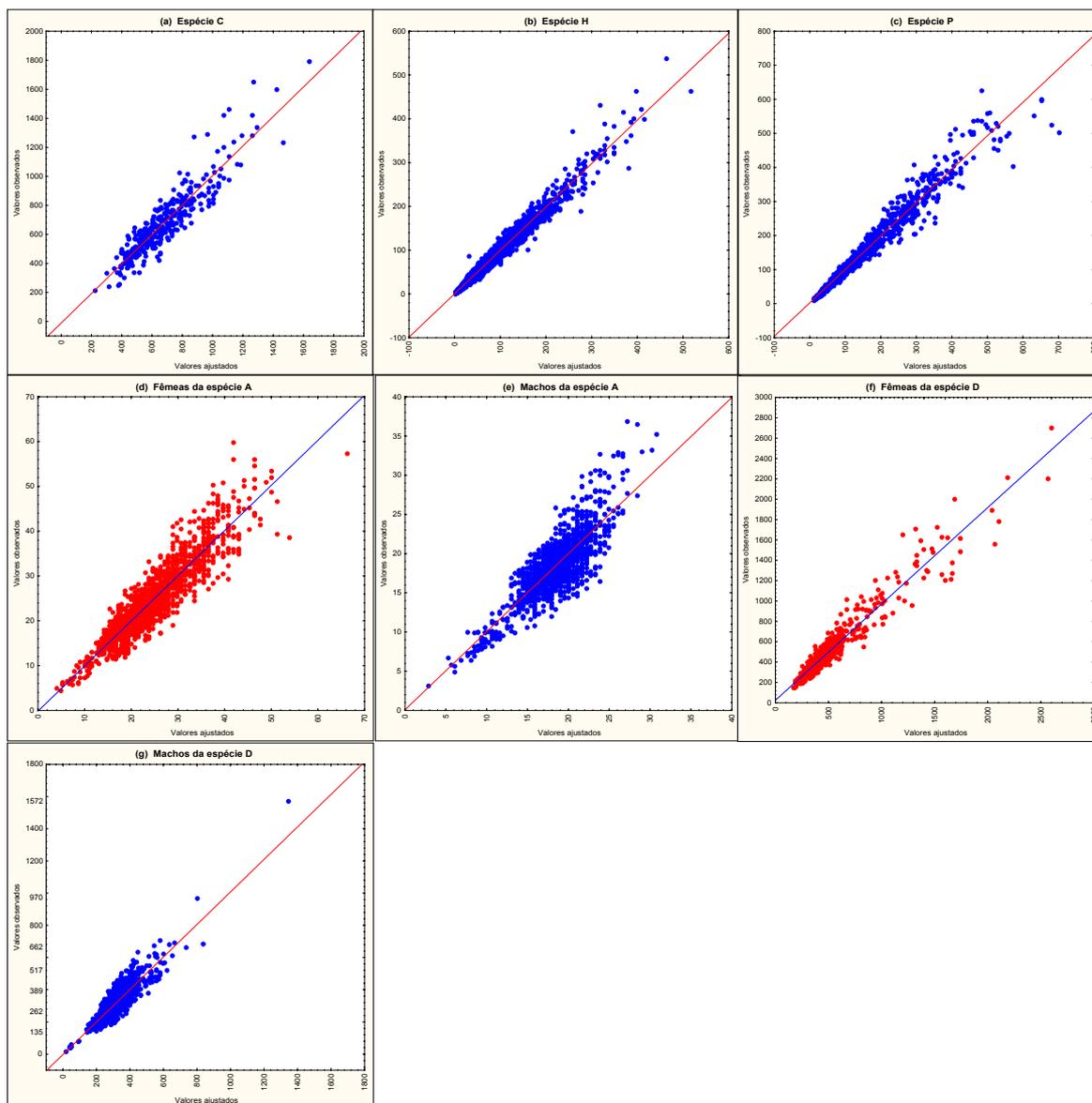


Figura 5.22: Dispersão entre os valores ajustados e valores observados.

Observando-se os gráficos contidos na Figura 5.22, nota-se uma boa proximidade dos dados em relação à 1ª bissetriz em todos os casos analisados.

Além da análise gráfica, também se aplicou o teste formal, sugerido no capítulo anterior, para verificar a adequação da função de ligação, ou seja, tomou-se o quadrado do preditor linear (η^2) como covariável adicional e comparou-se a nova deviance obtida para o modelo. A Tabela 5.22 apresenta os resultados e comparações para este procedimento.

Tabela 5.22: Valores das deviances do modelo inicial e acrescido da covariável $\hat{\eta}^2$.

Espécie	Deviance do modelo	Deviance acrescido de $\hat{\eta}^2$	Diferença
C	6,051	6,049	0,0023
H	127,651	127,642	0,0095
P	13,869	13,868	0,0011
A _F	28,050	28,015	0,0350
A _M	19,929	19,924	0,0057
D _F	11,307	11,124	0,1827
D _M	23,266	23,263	0,0032

Nota: Espécie C, *Liposarcus anisiti*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*. Espécie A_F, *Trachydoras paraguayensis* fêmeas e A_M *Trachydoras paraguayensis* machos. Espécie D_F, *Raphiodon vulpinus* fêmeas e D_M, *Raphiodon vulpinus* macho.

A maior redução da deviance provocada pela adição de $\hat{\eta}^2$ como uma covariável do modelo foi de 0,1827, que é um decréscimo insignificante. Portanto, pode-se considerar a função de ligação como sendo bastante adequada.

A Figura 5.23 apresenta os gráficos dos resíduos ordinários dispostos em ordem crescente.

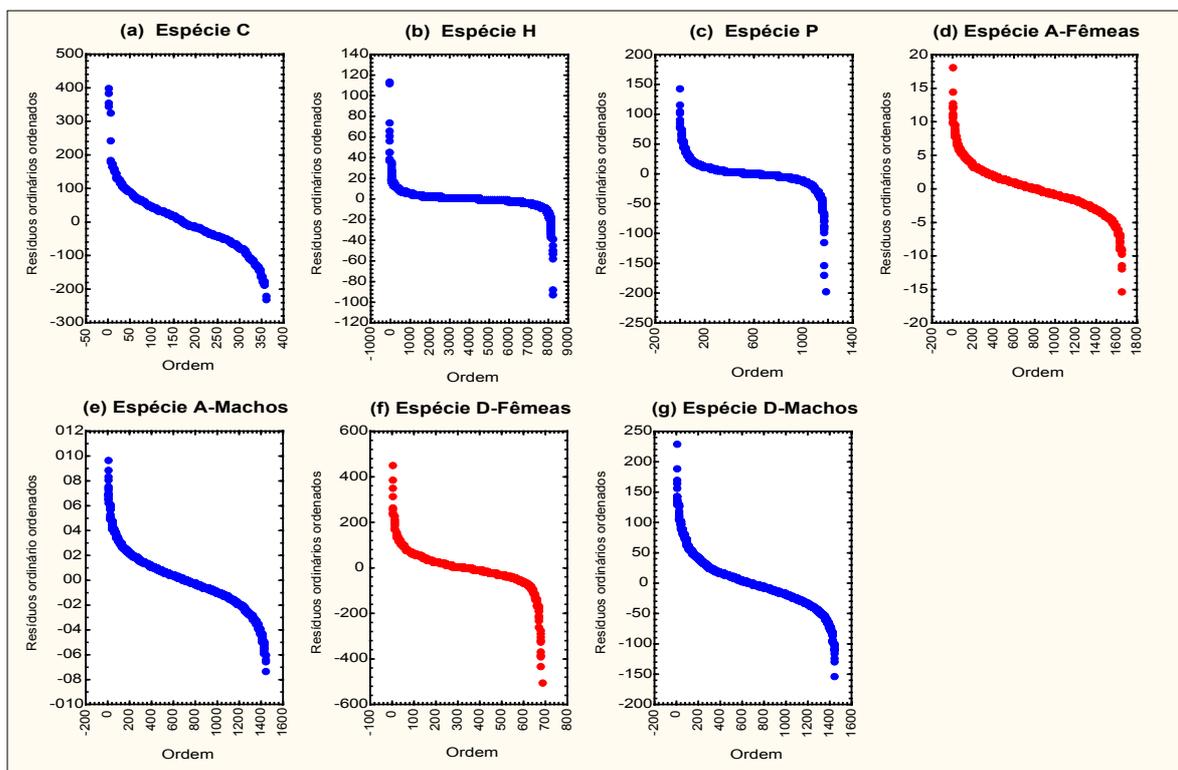


Figura 5.23: Distribuição dos resíduos ordinários em ordem decrescente.

Além da análise visual dos resíduos ordinários, uma análise descritiva dos resíduos ordinários pode ser feita por meio de algumas medidas. Tal procedimento auxilia a comparação dos resultados obtidos pelo modelo geral (tópico 5.1) e pelo modelo individual das espécies por sexo. Os dados apresentados nos parágrafos seguintes referem-se aos resíduos ordinários resultantes do ajuste ao modelo. Nestas informações, resíduos positivos correspondem a casos onde o valor estimado é menor que o observado, enquanto que os resíduos negativos referem-se a casos em que a estimativa apresenta valores superiores ao observado.

Espécie C: 98,62% dos resíduos ordinários encontram-se no intervalo de -220,77 a 182,42 gramas, variação residual que corresponde a menos de um desvio padrão do peso amostral, enquanto o intervalo interquartil situa-se entre -53,2 e 50,2 gramas. A soma do quadrado dos resíduos segundo este modelo, com 361 observações, é de 2.941.794,06 g² enquanto que pelo uso do modelo geral, com 354 observações, é de 5.033.508,85 g².

Espécie H: 99,84% dos indivíduos desta espécie tem o peso ajustado dentro de um intervalo correspondente a um desvio padrão do peso amostral, ou seja, resíduos que variam de -44,76 a 45,56 gramas. O intervalo interquartil é de -2,1 a 2,0 gramas. Este modelo produz uma soma de quadrados dos resíduos de 375.682,8 g² para um total de 8.195 observações. A soma do quadrado dos resíduos é de 473.835,1g² para os 7.379 casos analisados pelo modelo geral das espécies.

Espécie P: Nesta espécie, o intervalo dos pesos ajustados dentro de um desvio padrão do peso amostral corresponde a resíduos que variam de 114,52 a -115,57 gramas, abrangendo 99,66% dos casos, enquanto o intervalo interquartil é de -6,3 a 6,6 gramas. A soma dos quadrados dos resíduos de 600.162,39g² para um total de 1.178 observações. Pelo modelo geral, com 1.171 casos analisados para esta espécie, tem-se para a soma dos quadrados dos resíduos um valor de 952.491,38g².

Fêmeas da espécie A: 97,88% dos indivíduos desta espécie tem o peso ajustado dentro de um intervalo correspondente a um desvio padrão do peso amostral, ou seja, resíduos que variam de -8,42 a 8,44 gramas, enquanto o intervalo interquartil é de -2,00 a 1,8 gramas. A soma dos quadrados dos

resíduos é de 17.666,33g² para um total de 1.649 observações, enquanto que pelo modelo geral, para 1.646 casos, a soma dos quadrados dos resíduos é de 11.256,09 g².

Machos da espécie A: 93,06% dos casos desta espécie e sexo estão contidos num intervalo cujos resíduos variam de -4,09 a 4,11 gramas e não ultrapassam a um desvio padrão do peso amostral. A medida interquartil varia de -1,31 à 1,28gramas e a soma do quadrado dos resíduos para as 1.441 observações é de 7.141,07g² enquanto para o modelo geral ela é de 9.017,26g², para um total de 1445 observações.

Fêmeas da espécie D: 98,54% dos casos estudados tem o peso ajustado com resíduos ordinários inferiores a um desvio padrão amostral, situados entre -302,61 a 315,31 gramas. A medida interquartil varia de -34,6 a 32,2 gramas e a soma do quadrado dos resíduos para as 684 observações é de 4.793.463,28g².

Machos da espécie D: Para esta espécie, o intervalo dos pesos ajustados dentro de um desvio padrão do peso amostral corresponde a resíduos que variam de -98,40 a 98,03 gramas, abrangendo 96,34% dos casos, enquanto que a medida interquartil varia de -25,5 a 19,7 gramas e a soma do quadrado dos resíduos para as 1.449 observações é de 2.601.146,01g².

Os modelos construídos para o relato do peso total do peixe considerando sua espécie, sexo e comprimento são representados pelas equações contidas na Tabela 5.23. Isolando-se os interceptos estimados destas equações, tem-se a relação básica para a determinação do fator de condição do peixe, objetivo desta pesquisa. Tais resultados servirão como valores médios comparativos quando forem aplicados dados reais.

Tabela 5.23: Equações para o modelo desenvolvido para o peso das espécies dos peixes.

Espécie	Sexo	Equação do modelo	Intercepto
Trachydoras paraguayensis A	F	$\hat{\mu}_{AF} = (-0,6649 + 0,4738 \cdot Ls)^{100/40}$	$\hat{\mu}_{AF}^{0,40} - 0,4738 \cdot Ls = -0,6649$
	M	$\hat{\mu}_{AM} = (-5,9606 + 1,7297 \cdot Ls)^{100/73}$	$\hat{\mu}_{AM}^{0,73} - 1,7297 \cdot Ls = -5,9606$
Raphiodon vulpinus D	F	$\hat{\mu}_{DF} = (1,2848 + 0,0228 \cdot Ls)^{100/13}$	$\hat{\mu}_{DF}^{0,13} - 0,0228 \cdot Ls = 1,2848$
	M	$\hat{\mu}_{DM} = (0,3476 + 0,1437 \cdot Ls)^{100/30}$	$\hat{\mu}_{DM}^{0,30} - 0,1437 \cdot Ls = 0,3476$

Continua

Tabela 5.23: Equações para o modelo desenvolvido para o peso das espécies dos peixes.

Liposarcus anisitsi - C	$\hat{\mu}_C = (2,0003 + 0,1955 \cdot Ls)^{100/32}$	$\hat{\mu}_C^{0,32} - 0,1955 \cdot Ls = 2,0003$
Serrasalmus marginatus - H	$\hat{\mu}_H = (-0,0893 + 0,2857 \cdot Ls)^{100/31}$	$\hat{\mu}_H^{0,31} - 0,2857 \cdot Ls = -0,0893$
Leporinus friderici - P	$\hat{\mu}_P = (0,1491 + 0,2443 \cdot Ls)^{100/31}$	$\hat{\mu}_P^{0,31} - 0,2443 \cdot Ls = 0,1491$

Após a definição do modelo, um outro procedimento com o propósito de ainda observar a qualidade do ajuste dos dados, é a sobreposição aos valores observados, dispostos em ordem crescente, aos seus respectivos valores estimados como na Figura 5.24.

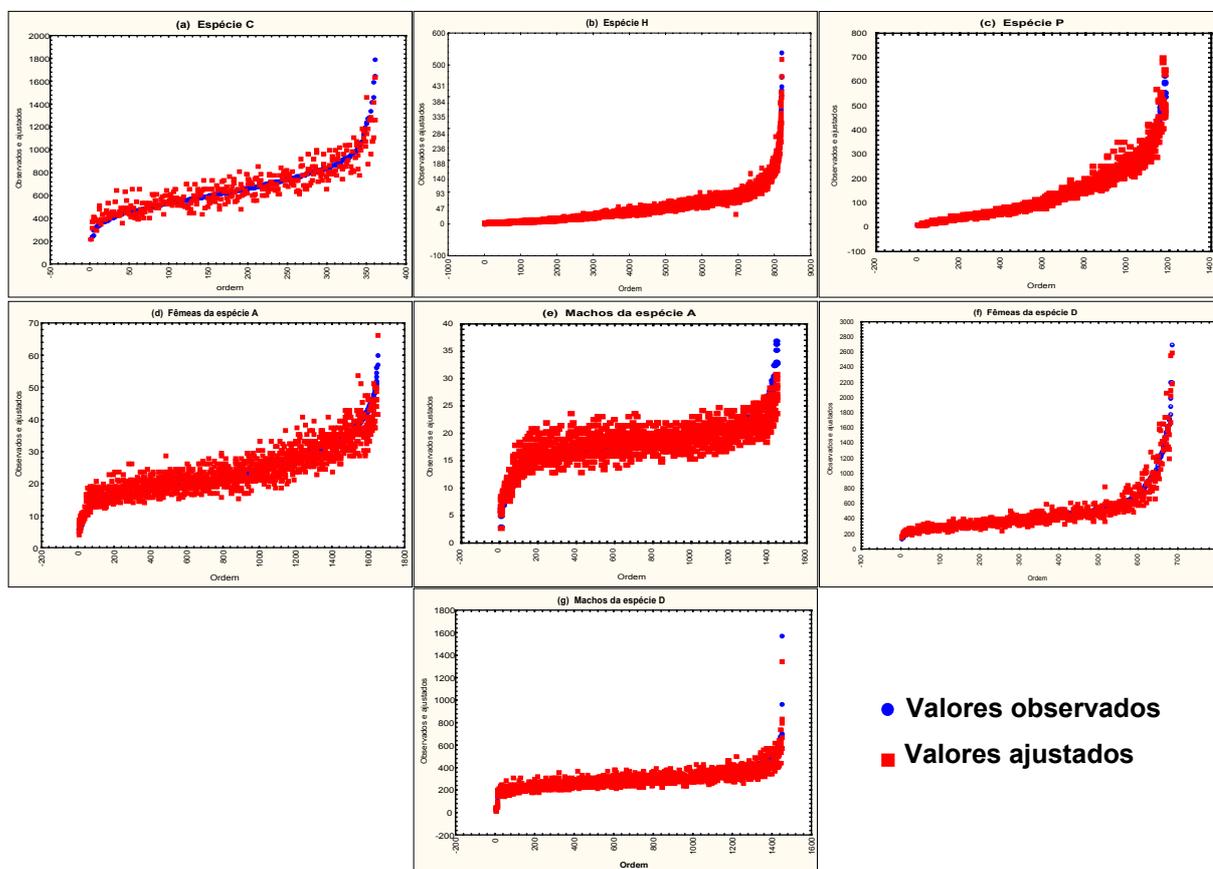


Figura 5.24: Sobreposição dos valores ajustados aos valores observados contra sua ordem crescente.

5.3 COMPARAÇÃO ENTRE OS MODELOS

Finalmente, é relevante estabelecer comparações entre os resultados obtidos por meio deste modelo e por outros modelos, aqueles que mais freqüentemente são utilizados em pesquisas ictiológicas. De acordo com Agostinho e Gomes (1997, p.114), em geral a relação peso e comprimento

padrão é ajustada pelo modelo (i) $W_t = aL_s^b$, onde a , como já citado no Capítulo 2, indica o estado nutricional do peixe e b em geral assume valores próximos a 3. Neste caso, quando b é assumido como a constante 3, tem-se o modelo que pelo qual gera-se o fator de condição de Fulton, (ii) $W_t = aL_s^3$ (CONNOLLY,1989,P.175). Um outro ajuste usualmente encontrado na literatura é a utilização do modelo linear com transformação logarítmica, descrito através da equação (iii) $\log(W_t) = \log(a) + b \log(L_s)$ (CONNOLLY,1989,p.175).

Os parâmetros a e b do modelo $W_t = aL_s^b$ apresentados na Tabela 5.25 e o parâmetro a do modelo $W_t = aL_s^3$, exposto na Tabela 5.26 foram estimados pelo método de Gauss-Newton. Os intervalos de confiança apresentados em ambos os modelos correspondem a um nível de confiança de 95%.

Tabela 5.24: Parâmetros estimados segundo o modelo $W_t = aL_s^b$

ESPÉCIE	a	INTERVALO		B	INTERVALO	
		Mínimo	Máximo		Mínimo	Máximo
C	0,16	0,103	0,222	2,43	2,324	2,535
H	0,01	0,014	0,015	3,27	3,256	3,277
P	0,02	0,019	0,024	3,03	2,989	3,068
A_F	0,04	0,032	0,042	2,96	2,900	3,021
A_M	0,12	0,100	0,146	2,37	2,775	2,455
D_F	0,0016	0,001	0,002	3,39	3,327	3,446
D_M	0,0031	0,002	0,004	3,20	3,131	3,271

Nota: Espécie C, *Liposarcus anisiti*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*. Espécie A_F, *Trachydoras paraguayensis* fêmeas e A_M, *Trachydoras paraguayensis* machos. Espécie D_F, *Raphiodon vulpinus* fêmeas e D_M, *Raphiodon vulpinus* macho.

Observa-se que somente as espécies P e A_F apresentam, segundo a teoria, um crescimento isométrico (intervalo da estimação de b contém 3). As espécies C e A_M tem crescimento alométrico negativo enquanto que para as espécies H, D_F e D_M, verifica-se alometria positiva.

O modelo $W_t = aL_s^3$ assume que as espécies, em geral, apresentam um crescimento isométrico. O modelo trabalhado no parágrafo anterior demonstra que tal afirmativa é inadequada, entretanto, por ser este um modelo amplamente utilizado, em especial por sua praticidade quanto ao cálculo do fator de condição do peixe, sua eficiência também será avaliada.

Tabela 5.25: Parâmetros estimados segundo o modelo $W_t = aL_s^3$.

ESPÉCIE	a	INTERVALO	
		Mínimo	Máximo
C	0,022	0,0217	0,0224
H	0,030	0,0304	0,0305
P	0,023	0,0231	0,0234
A _F	0,034	0,0334	0,0338
A _M	0,032	0,0314	0,0318
D _F	0,007	0,0070	0,0072
D _M	0,006	0,0064	0,0065

Nota: Espécie C, *Liposarcus anisiti*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*. Espécie A_F, *Trachydoras paraguayensis* fêmeas e A_M *Trachydoras paraguayensis* machos. Espécie D_F, *Raphiodon vulpinus* fêmeas e D_M, *Raphiodon vulpinus* macho.

O modelo linear proposto pela literatura citada foi construído segundo a teoria dos Modelos Lineares Generalizados, tomando-se como função de distribuição da variável resposta transformada $\ln(W_t)$ a distribuição de probabilidade Normal e como função de ligação, a função identidade. A Tabela 2.27 expõe os parâmetros estimados juntamente com seus intervalos com 95% de confiança.

Tabela 5.26: Parâmetros estimados segundo o modelo linear.

ESPÉCIE	Ln(a)	INTERVALO		B	INTERVALO	
		Mínimo	Máximo		Mínimo	Máximo
C	-1,52	-1,8853	-1,1569	2,34	2,2343	2,4478
H	-4,45	-4,4668	-4,4269	3,35	3,3424	3,3592
P	-4,08	-4,1469	-4,0229	3,10	3,0806	3,1247
A _F	-3,41	-3,5381	-3,2830	3,00	2,9458	3,0629
A _M	-2,36	-2,5232	-2,2064	2,49	2,4137	2,5637
D _F	-6,15	-6,4102	-5,8910	3,31	3,2385	3,3788
D _M	-5,37	-5,6224	-5,1138	3,08	3,0140	3,1530

Nota: Espécie C, *Liposarcus anisiti*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*. Espécie A_F, *Trachydoras paraguayensis* fêmeas e A_M *Trachydoras paraguayensis* machos. Espécie D_F, *Raphiodon vulpinus* fêmeas e D_M, *Raphiodon vulpinus* macho.

O método utilizado para comparar as estimativas segundo cada modelo, é similar ao desenvolvido por Gazola (2002, p.80). Após a estimação do peso total para cada observação de cada espécie, calcula-se os percentuais com relação aos pesos reais e através destes valores estabelece-se os limites para o intervalo do erro nas estimativas. Estabelecidos tais faixas de erros, verifica-se o percentual de casos estimados que estão contidos dentro destes intervalos através dos resíduos ordinários determinados. Os percentuais obtidos encontram-se na Tabela 5.28.

Tabela 5.27: Percentuais das estimativas contidas nos intervalos dos erros especificados.

ESPÉCIE	ERROS	$W_T = a \cdot L_s^b$	$W_T = a' \cdot L_s^3$	$\ln(W_T) = \ln(a'') + b'' \cdot \ln(L_s)$	$\hat{\mu} = (\alpha + \beta \cdot L_s)^h$
<i>Liposarcus anisiti</i> C	5%	28,81	22,99	29,92	30,19
	10%	58,17	45,15	59,28	59,28
	15%	76,45	66,65	76,73	77,56
	20%	87,53	80,89	88,92	88,64
	30%	96,68	94,74	96,68	96,34
	40%	98,34	99,72	98,06	98,06
	50%	99,45	100,00	99,45	99,17
	55%	100,00		100,00	100,00
<i>Serrasalmus marginatus</i> H	5%	70,59	65,65	70,40	70,69
	10%	82,73	73,94	82,57	83,06
	15%	89,84	80,30	90,37	90,32
	20%	94,06	85,23	94,59	94,55
	30%	97,22	91,02	98,12	97,93
	40%	98,71	94,18	99,33	99,28
	50%	99,40	96,03	99,78	99,77
	55%	97,57	96,67	99,90	99,90
	60%	99,82	97,36	99,97	99,94
	70%	99,98	98,17	100,00	100,00
	80%	100,00	98,82		
<i>Leporinus friderice</i> P	5%	34,13	32,34	34,80	34,97
	10%	61,97	60,53	64,86	64,35
	15%	81,58	79,46	83,87	83,70
	20%	90,92	89,39	93,97	93,63
	30%	97,79	97,28	98,64	98,64
	40%	99,49	99,49	99,74	99,57
	50%	100,00	100,00	100,00	100,00
<i>Trachydoras paraguayensis</i> A _F (Fêmeas)	5%	30,75	30,69	30,62	30,50
	10%	57,45	57,07	57,19	56,76
	15%	76,23	76,41	76,59	76,23
	20%	87,45	87,02	87,75	87,39
	30%	96,12	96,30	96,60	96,36
	40%	98,61	98,67	98,73	98,79
	50%	99,70	99,70	99,88	99,70
	55%	99,94	99,94	100,00	99,94
60%	100,00	100,00		100,00	
<i>Trachydoras paraguayensis</i> A _M (Machos)	5%	35,11	33,80	34,17	36,57
	10%	62,80	58,29	62,66	63,43
	15%	77,93	76,61	78,28	78,83
	20%	88,90	87,79	90,15	89,73
	30%	97,02	97,22	97,36	97,29
	40%	99,31	99,44	99,51	99,44
	50%	99,93	99,93	100,00	100,00
55%	100,00	100,00			
<i>Raphiodon vulpinus</i> D _F (Fêmeas)	5%	30,26	22,81	31,87	32,16
	10%	54,39	47,51	58,19	55,85
	15%	73,68	64,04	75,73	75,44
	20%	87,13	77,92	88,45	88,01
	30%	97,22	94,59	97,95	97,51
	40%	99,56	98,25	99,71	99,85
	50%	99,85	100,00	100,00	100,00
55%	100,00				

Continua

Tabela 5.27: Percentuais das estimativas contidas nos intervalos dos erros especificados.

ESPÉCIE	ERROS	$W_T = a \cdot L_S^b$	$W_T = a' \cdot L_S^3$	$\ln(W_T) = \ln(a'') + b'' \cdot \ln(L_S)$	$\hat{\mu} = (\alpha + \beta \cdot L_S)^h$
<i>Raphiodon vulpinus</i> Df (Machos)	5%	32,22	32,23	32,44	31,88
	10%	58,73	57,83	59,08	58,59
	15%	78,88	77,23	79,16	78,33
	20%	87,34	87,23	88,75	88,13
	30%	97,65	97,31	97,65	97,65
	40%	99,38	99,24	99,38	99,38
	50%	99,79	99,72	99,79	99,79
	55%	99,86	99,86	99,93	99,86
	60%	100,00	100,00	100,00	100,00

Por tais resultados, verifica-se que as estimativas apresentam percentuais de erros com pequenas diferenças, porém, somente para as espécies *Trachydoras paraguayensis* fêmeas e *Raphiodon vulpinus* macho verificou-se um menor número de casos nas menores faixas de erros, uma diferença para o índice mais elevado de 0,25% e 0,56%, respectivamente. Para o modelo apresentado nesta pesquisa, a faixa que admite um erro máximo de 30% contém no mínimo 96,34% das estimativas (*Liposarcus anisiti*) e no máximo 98,64% dos casos (*Leporinus friderice*). Dentre os erros verificados na Tabela 5.28, os intervalos que mais se aproximam são os que correspondem aos modelos: linear e linear generalizado. Entretanto, o primeiro modelo exige a transformação das variáveis, pois a regressão necessitou da transformação logarítmica dos dados, enquanto que, para o modelo linear generalizado, tal procedimento não foi necessário, facilitando sua aplicabilidade. Todas estas considerações levam a concluir que os dados ajustados apresentam uma boa aderência ao modelo.

5.4 FATOR DE CONDIÇÃO DO PEIXE

Finalmente, a partir das equações apresentadas na Tabela 5.23, pode-se definir uma fórmula para o fator de condição do peixe:

Tabela 5.28: Fator de condição K do peixe a partir do modelo construído.

Espécie	Fator de condição	
	Fêmea	Macho
<i>Trachydoras paraguayensis</i>	$K_{AF} = \hat{\mu}_{AF}^{0,40} - 0,4738 \cdot L_S$	$K_{AM} = \hat{\mu}_{AM}^{0,73} - 1,7297 \cdot L_S$
<i>Raphiodon vulpinus</i>	$K_{DF} = \hat{\mu}_{DF}^{0,13} - 0,0228 \cdot L_S$	$K_{DM} = \hat{\mu}_{DM}^{0,30} - 0,1437 \cdot L_S$
<i>Liposarcus anisiti</i>	$K_C = \hat{\mu}_C^{0,32} - 0,1955 \cdot L_S$	
<i>Serrasalmus marginatus</i>	$K_H = \hat{\mu}_H^{0,31} - 0,2857 \cdot L_S$	
<i>Leporinus friderice</i>	$K_P = \hat{\mu}_P^{0,31} - 0,2443 \cdot L_S$	

Aplicando-se os dados observados às relações contidas na Tabela 2.29, obtêm-se os valores do fator de condição K. Os resultados obtidos, quando apresentados sob a forma de histogramas, apresentam uma distribuição muito próxima da Normal, conforme Figura 5.25.

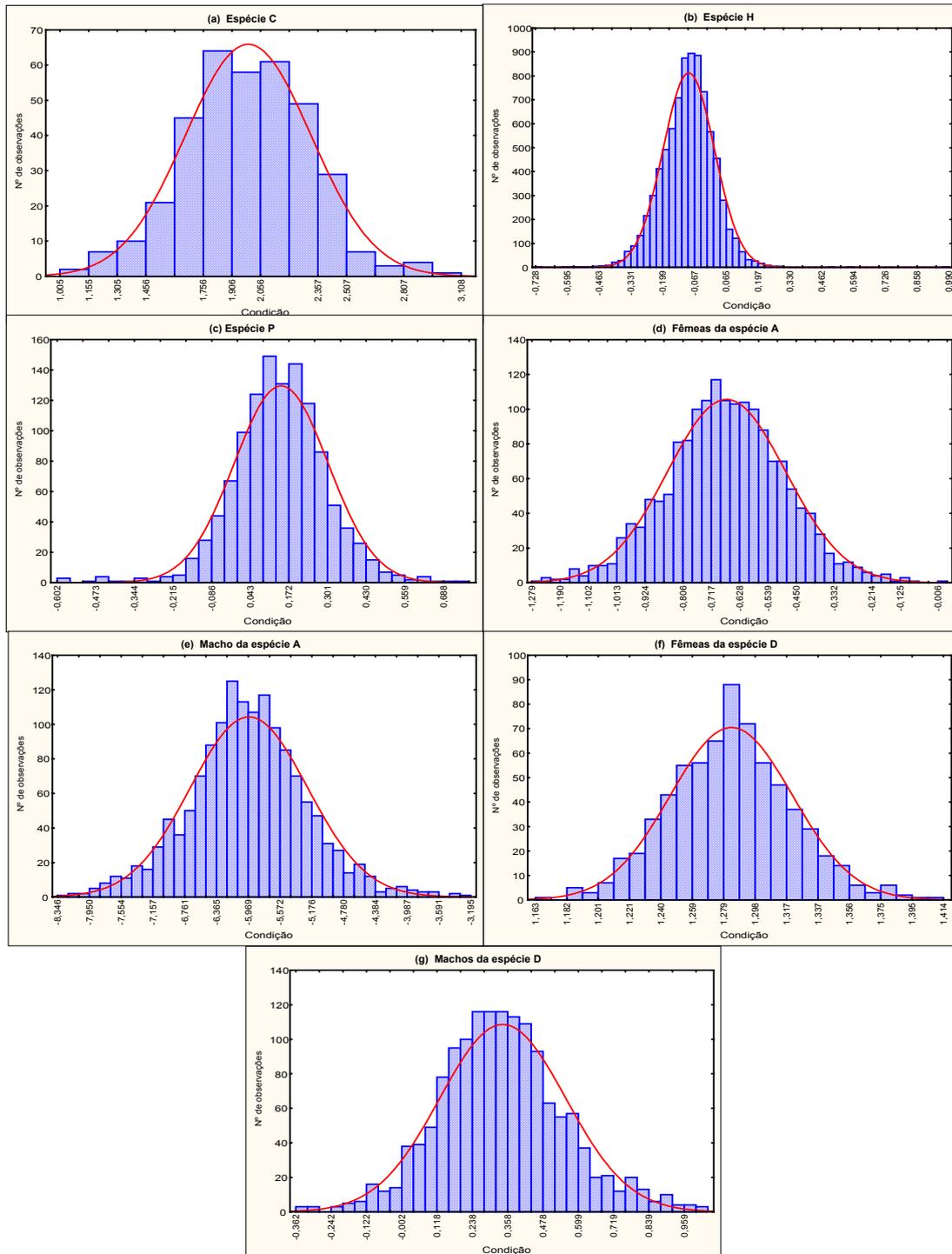


Figura 5.25: Distribuição dos fatores de condição para os peixes coletados.

A normalidade dos fatores de condição do peixe também pode ser verificada por meio do gráfico dos valores observados contra os valores esperados segundo a distribuição normal, conforme mostrado na Figura 5.26.

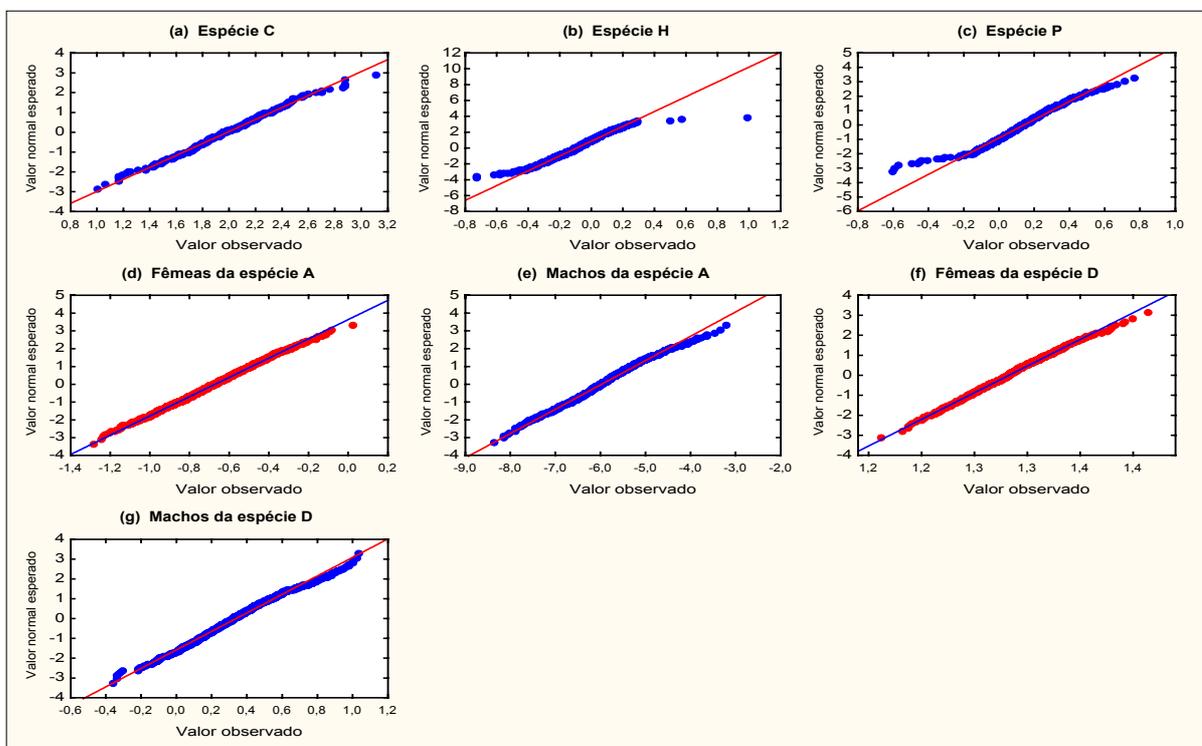


Figura 5.26: Plotagem da probabilidade normal para a condição do peixe.

As medidas descritivas da variável K confirmam os valores esperados para as médias. Para todas as espécies, o fator de condição médio encontram-se dentro do intervalo estimado para o intercepto (fator de condição) apresentado na Tabela 5.30.

Tabela 5.29: Medidas descritivas do fator de condição do peixe, K.

Espécie	n	Média	Desvio Padrão	Mínimo	Máximo
C	361	1,986	0,328	1,005	3,108
H	8.195	-0,094	0,106	-0,728	0,990
P	1.178	0,142	0,156	-0,602	0,773
A _F	1.649	-0,672	0,184	-1,279	0,023
A _M	1.441	-5,973	0,728	-8,346	-3,195
D _F	684	1,283	0,037	1,163	1,414
D _M	1.449	0,338	0,213	-0,362	1,039

Nota: Espécie C, *Liposarcus anisiti*. Espécie H, *Serrasalmus marginatus*. Espécie P, *Leporinus friderice*. Espécie A_F, *Trachydoras paraguayensis* fêmeas e A_M *Trachydoras paraguayensis* machos. Espécie D_F, *Raphiodon vulpinus* fêmeas e D_M, *Raphiodon vulpinus* macho.

Comparando-se as médias dos fatores aplicados aos dados observados com os resultados das estimações do intersepto da Tabela 5.23, pode-se observar que estão todas dentro do intervalo estimado.

6- CONCLUSÃO

Quando se estuda um fator de condição do peixe, K, em geral, observa-se que os mesmos não consideram as diferenças por espécies ou sexo dos animais. Por esta razão, o processo de investigação e adequação dos modelos propostos nesta pesquisa partiu de um modelo geral, etapa na qual comprovou-se que o fator espécie era significativa para cada uma das espécies em estudo. Conseqüentemente, propôs-se um novo modelo para cada uma das cinco espécies em estudo e então se verificou a significância do fator sexo.

Nas espécies *Trachydoras paraguayensis* e *Raphiodon vulpinus* o fator sexo mostrou-se significativa; então, se construiu um modelo para machos e fêmeas de cada uma dessas espécies. Já para as espécies *Liposarcus anisitsi*, *Serrasalmus marginatus* e *Leporinus friderice*, este fator não se apresentou significativo e por tal motivo foi retirado e um modelo final foi proposto.

Após apresentar um modelo capaz de descrever o peso do peixe em função do seu comprimento, classificado segundo sua espécie e sexo, como o fator de condição do peixe proposto nesta pesquisa não foi encontrado na literatura consultada, houve um cuidado especial em compara-lo a alguns outros modelos que, tradicionalmente, são aceitos em pesquisas da área. Os resultados obtidos mostram-se bastante satisfatórios, conforme tópico 5.2.4, o que demonstra a validade do modelo proposto.

Outro aspecto relevante deste trabalho, é que a população alvo pertence a uma região de clima tropical, cujo habitat é caracterizado por alagamentos e barragens, conforme descrito no Capítulo 2.

6.1 RESULTADOS OBTIDOS

Os modelos construídos apresentam características lineares na função da média para a variável peso total do peixe; uma variável que de acordo com o

desenvolvimento da pesquisa, comporta-se segundo uma distribuição de probabilidade Gama. A função de ligação entre as médias das observações e a estrutura linear do modelo, em todos os casos, foi uma função potência, cujo grau variou de espécie para espécie.

Um dos principais objetivos deste trabalho é estabelecer relações mais simplificadas do peso em função do comprimento do animal. Os modelos individuais obtidos foram os seguintes:

Trachydoras paraguayensis, A:

$$\text{Fêmeas - } A_F: \hat{\mu}_{AF} = (-0,665 + 0,474 \cdot Ls)^{100/40}$$

$$\text{Machos - } A_M: \hat{\mu}_{AM} = (-5,961 + 1,730 \cdot Ls)^{100/73}$$

Raphiodon vulpinus, D:

$$\text{Fêmeas - } D_F: \hat{\mu}_{DF} = (1,285 + 0,023 \cdot Ls)^{100/13}$$

$$\text{Machos - } D_M: \hat{\mu}_{DM} = (0,348 + 0,144 \cdot Ls)^{100/30}$$

$$\text{Liposarcus anisitsi, C: } \hat{\mu}_C = (2,000 + 0,196 \cdot Ls)^{100/32}$$

$$\text{Serrasalmus marginatus, H: } \hat{\mu}_H = (-0,089 + 0,286 \cdot Ls)^{100/31}$$

$$\text{Leporinus friderici, P: } \hat{\mu}_P = (0,149 + 0,244 \cdot Ls)^{100/31}$$

Por meio do ajuste dos dados a estes modelos, pode-se encontrar uma equação fundamentada na relação peso-comprimento, que poderá ser utilizada como um indicador do bem estar do animal, bem como a distribuição desses fatores.

Trachydoras paraguayensis, A:

$$\text{Fêmeas - } A_F: K_{AF} = \hat{\mu}_{AF}^{0,40} - 0,4738 \cdot Ls$$

$$\text{Machos - } A_M: K_{AM} = \hat{\mu}_{AM}^{0,73} - 1,7297 \cdot Ls$$

Raphiodon vulpinus, D:

$$\text{Fêmeas - } D_F: K_{DF} = \hat{\mu}_{DF}^{0,13} - 0,0228 \cdot L_s$$

$$\text{Machos - } D_M: K_{DM} = \hat{\mu}_{DM}^{0,30} - 0,1437 \cdot L_s$$

$$\text{Liposarcus anisitsi, C: } K_C = \hat{\mu}_C^{0,32} - 0,1955 \cdot L_s$$

$$\text{Serrasalmus marginatus, H: } K_H = \hat{\mu}_H^{0,31} - 0,2857 \cdot L_s$$

$$\text{Leporinus friderici, P: } K_P = \hat{\mu}_P^{0,31} - 0,2443 \cdot L_s$$

Os intervalos obtidos para cada um dos fatores de condição do peixe, K, representados pelas equações anteriores, são os seguintes:

Tabela 6.1: Intervalos encontrados para os fatores de condição do peixe, K, segundo a espécie e sexo do animal.

Espécies	Mínimo	Máximo
<i>Trachydoras paraguayensis</i> - Fêmeas	-1,279	0,023
<i>Trachydoras paraguayensis</i> - Machos	-8,346	-3,195
<i>Raphiodon vulpinus</i> - Fêmeas	1,163	1,414
<i>Raphiodon vulpinus</i> - Machos	-0,362	1,039
<i>Liposarcus anisitsi</i>	1,005	3,108
<i>Serrasalmus marginatus</i>	-0,728	0,990
<i>Leporinus friderici</i>	-0,602	0,773

Uma importante característica destes fatores de condição, K, é a maior sensibilidade às diferenças do comprimento do animal que os demais modelos (Apêndice C), o que auxilia a avaliação das condições gerais físicas e ambientais na qual o animal vive.

6.2 LIMITAÇÕES DA PESQUISA

Os resultados obtidos limitam-se ao estudo das cinco espécies consideradas, levando-se em consideração o ambiente ao qual pertencem. Os modelos podem ser aplicados como previsão dos pesos dos espécimes cujas características, peso e sexo, estejam dentro das dimensões encontrados nas amostras trabalhadas.

6.3 SUGESTÕES PARA NOVAS PESQUISAS:

Investigar os ajustes obtidos pelos modelos mostrados no tópico 5.3, pois nesta pesquisa não houve a preocupação de verificar seus pressupostos, já que os mesmo são citados em outros trabalhos da área. Também as espécies H, *Serrasalmus marginatus*, e os machos da espécie *Raphiodon vulpinus*, não apresentam um comportamento adequado segundo a construção do modelo, sugerindo investigações mais aprofundadas. Tais procedimentos podem auxiliar na melhor compreensão e aplicabilidade do fator de condição do peixe, K, trabalhados nesta pesquisa, bem como definir limites mais precisos para tais fatores.

REFERÊNCIAS

AGOSTINHO, Angelo A.; GOMES, Luiz C. **Reservatório de Segredo. Bases ecológicas para o manejo.** Maringá, EDUEM, 1997.

Almaque Abril 2000. São Paulo: Editora Abril, 2000

AMARAL, Getúlio José Amorim do. **Modelo de Alocação de Recursos entre as instituições Federais de Ensino Superior:** Uma Aplicação dos Modelos Lineares Generalizados. 1997. Dissertação, UFPE, Recife.

Bacia do Prata. Disponível em <http://www.ibama.gov.br/pescaamadora/locais/b_doprata.htm>. Acesso em 30.05.02

Bacia do Paraná. Disponível em:<<http://www.geologiabrasil.hpg.ig.com.br/LINKS/BACIAS/baciaparana.htm>>. Acesso em 31.05.02

BOLGER, T; CONNOLLY, P. L. **The selection of suitable índices for the measurement and analysis of fish condition.** J. Fish Biol. (1989) 34, 171-182

BUSSAB, Wilton O.; MORETTIN, Pedro A. **Estatística Básica – Métodos Quantitativos.** Ed.4, São Paulo: Atual Editora Ltda, 1987.

BUSTOS, Oscar. **Estimação robusta no modelo de posição.** 13º Colóquio Brasileiro De Matemática, IMPA, Rio de Janeiro, 1981.

CAPRA, Fritjof. **O tao da Física.** Ed. 16, São Paulo: Editora Cultrix Ltda, 1995.

CORDEIRO, Gauss Moutinho. **Introdução à teoria de verossimilhança.** 10º SINAPE, Rio de Janeiro, 1992.

CORDEIRO, Gauss Moutinho. **Modelos Lineares Generalizados.** 1ª Jornada Regional de Estatística de Maringá, 1997.

CORDEIRO, Gauss Moutinho. **Modelos Lineares Generalizados.** VII SINAPE, Campinas, São Paulo, 1986.

COSTA NETO, Pedro Luiz de Oliveira. **Estatística**. Ed.2, São Paulo: Editora Edgard Blücher Ltda, 2002.

DANTAS, Rubens Alves. **Engenharia de Avaliações – introdução à Metodologia Científica**. São Paulo: Pini, 1998.

DEMÉTRIO, Clarice Garcia Borges. **Modelos Lineares Generalizados na Experimentação Agrônômica**. 5º Simpósio de Estatística aplicada à experimentação agrônômica, 38ª Reunião anual da sociedade internacional de biometria. Porto Alegre, 1993.

DEMÉTRIO, Clarice Garcia Borges. **Modelos Lineares Generalizados na Experimentação Agrônômica**. ESALQ/USP, Piracicaba, SP, Maio/2001.

DOBSON, Annette J. **A Introduction to Generalized Linear Models**. Ed.2, New York: Chapman & Hall/CRC, 2002.

DRUMOND, Fátima Brant; WERKEMA, Maria Cristina Catarino; AGUIAR, Sílvio. **Análise de Variância: Comparações de várias situações**. Série ferramentas da qualidade, v.6, Belo Horizonte: Editora Littera Maciel Ltda, 1996.

EVANS, Merran; HASTINGS, Nicholas e PEACOCK. **Statistical Distributions**. Ed.2, New York: John Wiley & Sons, Inc, 1993.

GAZOLA, Sebastião. **Construção de um modelo de regressão para avaliação de imóveis**. 2002. Dissertação – Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis.

GRETEL, Villamonte. **Proposta metodológica para construção de indicadores para a avaliação da rede escolar brasileira**. 2001. Dissertação – Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis.

HORTON, Nicholas J.; LAIRD, Nan M. **Maximum Likelihood Analysis of Generalized Linear Models with Missing Covariates**. Statistical Methods in Medical Research, 8: 37-50, 1998.

McCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. Ed.2, London: Chapman & Hall, 1989.

McCULLOCH, Charles E.; SEARLE, Shayle R. **Generalized, Linear, and Mixed Models**. New York: John Wiley & Sons, Inc, 2001.

MOOD, Alexander M.; GRAYBILL, Franklin A. e BOES, Duane C. **introduction to the Theory of Statistics**. Tokyo: MacGraw Hill Kogakusha Ltda, 1974.

NETER, J. WASSERMAN, W. KUTNER, M. H. e NACHTSHELM, C. J. **Applied Linear Regression Models**. Ed.3, Boston: Times Mirror Hiher Group, Inc., 1996.

NEY, John J. **Practical use of Biological Statistics**. In KOHLER, Christopher C.; HUBERT, Wayne A. **Inland Fisheries Management in north America**.. Maryland, USA: Editora American Fisheries Society, 1993, 6: 137-158.

PAULA, Gilberto A. **Modelos de Regressão com apoio computacional**. Instituto de Matemática e Estatística, Universidade de São Paulo. São Paulo, 1999.

PAULA, Gilberto A. **Modelos de Regressão com apoio computacional**. Instituto de Matemática e Estatística, Universidade de São Paulo. São Paulo, 2001.

Pesca Amadora. Disponível em <<http://www.ibama.gov.br/pescaamadora/peixes.htm>>. Acesso em: 30.05.02

PORTELLA, Maria Célia; VERANI, José Roberto; CARNEIRO, Dalton José; CESTAROLLI, Marcos Antonio. **GROWTH PERFORMANCE OF *Prochilodus lineatus* (=Prochilodus scrofa) LARVAE AND FRY- METHOD OF ANALYSIS BASED ON THE RELATIVE CONDITION FACTOR**. Boletim do Instituto de Pesca, São Paulo, 26(2): 129-135, 2000.

RATKOWSKY, David A. **Nonlinear Regression Modeling – A Unified Pratical Approach**. New York: MARCEL DEKKER, INC.,1983.

Relatório Técnico 2000. Área de Pesquisas Ecológicas de Longa Duração PELD – Nupélia -. Disponível em <<http://www.peld.uem.br/Relat2000/index.htm>>. Acesso em: 10.04.2002

RODRIGUEZ, Gérman. **Generalized Linear Models**. Disponível em: <http://data.princeton.edu/wws509/notes/default.htm>. Acesso em 25 de janeiro de 2002.

ROSEGRANT, Mark W.; CAI, Ximing; CLINE, Sarah A. **Global Water Outlook to 2025 – Averting na Impending Crisis**. International Food Policy Research Institute, 09.2002. Disponível em < www.ifpri.org>. Acesso em: 20/10/2002

SÁ, Maria de Fátima P. de; BARBIERI, Geraldo; VERANI, José Roberto. **ANÁLISE DO COMPORTAMENTO DE *Cyprinus carpio*, *Prochilodus cearensis* E *Colossoma***

macropomum EM EXPERIMENTO DE POLICULTIVO, EMBASADO NOS FATORES DE CONDIÇÃO. Boletim do Instituto de Pesca, São Paulo, 26(2): 181-187, 2000.

SEARLE, Shayle R. **Linear Models**. New York: John Wiley & Sons, Inc, 1971.

SPIEGEL, Murray R. **CÁLCULO AVANÇADO**. São Paulo: Editora McGraw-Hill do Brasil LTDA, 1974.

The SAS System for Windows. Version 8.02 TS nível 02M0. Windows Version 4.10.2222. Cary, NC, USA [1999 – 2001].

The Statistica Version 6,0. StatSoft Séries 0602 for Windows. **STATISTICA Eletronic Manual**. Tulsa, USA [1982–2002]

VAZZOLER, Anna Emília A. De M.; AGOSTINHO, Angelo A.; HAHN, Norma Segatti. **A Planície de Inundação do Alto do Rio Paraná: aspectos físicos, biológicos e socioeconômicos**. Maringá, EDUEM, 1997.

WERKEMA, Maria Cristina Catarino; AGUIAR, Sílvio. **Planejamento e análise de experimentos: como identificar as principais variáveis influentes em um processo**. Série ferramentas da qualidade, v.8, Belo Horizonte: Editora Littera Maciel Ltda, 1996.

WONNACOTT, Thomas H.; WONNACOTT, Ronald J. **ESTATÍSTICA APLICADA À ECONOMIA E À ADMINISTRAÇÃO**. Rio de Janeiro: Livros Técnicos e Científicos Editora S.A., 1981.

APÊNDICE A: ESTIMAÇÃO DE PARÂMETROS

A.1 ESTIMADOR E ESTIMATIVA

Ao trabalhar-se com situações que envolvam amostras as quais, embora retiradas de uma população cujos valores se apresentem segundo um modelo de distribuição probabilística conhecida, tenha seus parâmetros populacionais desconhecidos, faz-se necessário estimar esses parâmetros baseados nos resultados amostrais. Segundo Costa Neto (p.58, 2002), “um *estimador* do parâmetro populacional é a quantidade calculada em função dos elementos da amostra”, enquanto que *estimativa* corresponde a “cada valor particular assumido por um estimador”.

A.2 PROPRIEDADES DOS ESTIMADORES

Ao se estimar parâmetros populacionais desconhecidos, para que o estimador $\hat{\theta}$ seja considerado um “bom” estimador e possa caracterizar devidamente uma distribuição probabilística, é necessário que satisfaça as seguintes propriedades:

(i) *Justeza ou não-tendenciosidade*: Segundo Costa Neto (p.59,2002), “um estimador é não-tendencioso (justo, não-viciado ou não-viesado) quando sua esperança for o próprio parâmetro que se pretende estimar”, ou seja,

$$E(\hat{\theta}) = \theta$$

(ii) *Consistência*: Um estimador $\hat{\theta}$ é dito consistente para θ se o limite da probabilidade tal que $|\hat{\theta}_n - \theta| \geq \xi$ é igual a zero quando n tende para o infinito, ou seja;

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| \geq \xi) = 0 \quad \text{onde } \xi > 0$$

Em outras palavras, “um estimador $\hat{\theta}$ é dito consistente se sua tendenciosidade e sua variância tenderem ambas para zero, quando $n \rightarrow \infty$ ” (Wonnacott, p.168,1981).

(iii) *Eficiência*: Um estimador $\hat{\theta}$ é dito eficiente quando as seguintes condições

forem satisfeitas:

(a) a variância σ^2 é menor que a variância de qualquer outro estimador θ_n^* que satisfaça a condição (b).

(b) $\sqrt{n}(\hat{\theta}_n - \theta)$ tem distribuição normal assintótica com média 0 e variância σ^2 , onde n é o tamanho da amostra.

(iv) *Variância mínima*: Um estimador $\hat{\theta}$ tem variância mínima se:

$$E(\hat{\theta} - E(\hat{\theta}))^2 \leq E(\theta^* - E(\theta^*))^2$$

onde θ^* é qualquer outro estimador para θ . Se um estimador é eficiente, então é consistente e não-tendencioso no limite, mas não necessariamente não-tendencioso para amostras de tamanho finito.

v) *Suficiência*: “Seja x_1, x_2, \dots, x_n uma amostra aleatória de função densidade $f(x_1, \dots, x_n; \theta_1, \dots, \theta_k)$. Uma estatística $S = s(x_1, x_2, \dots, x_n)$ é definida como estatística suficiente se, e somente se, a distribuição condicional de x_1, x_2, \dots, x_n dado $S = s$ não depende de θ para qualquer valor s de S .” (MOOD, p.301, 1974)

(vi) *Completo*: Um estimador $\hat{\theta}$ de θ é um estimador completo se não existir um estimador não-tendencioso de zero na função de frequência, $f(\hat{\theta}, \theta)$ exceto o próprio zero.

(vii) *Invariância*: Um estimador $\hat{\theta}$ de θ é um estimador invariante para uma certa classe de transformações g , se o estimador é $g(\hat{\theta})$ quando a transformação muda o parâmetro para $g(\theta)$.

A.3 MÉTODO DE ESTIMAÇÃO PELA MÁXIMA VEROSSIMILHANÇA

Existem diversos métodos para se estimar um parâmetro. Neste trabalho é utilizado o da Máxima Verossimilhança, o qual “consiste basicamente em tomar para parâmetro, o valor que maximize a função de verossimilhança correspondente ao resultado obtido através de uma amostra aleatória” (LONARDAN, p.70, 1985).

De acordo com German (Apêndice A, p.1, 2001), a função de verossimilhança para uma amostra aleatória de n observações independentes $\mathbf{y} = (y_1, \dots, y_n)^T$ é

$$f(\mathbf{y}; \boldsymbol{\theta}) = \prod_{i=1}^n f_i(y_i; \boldsymbol{\theta}) = L(\mathbf{y}; \boldsymbol{\theta}) = L(\boldsymbol{\theta})$$

onde θ é um vetor de parâmetros desconhecidos $\theta=(\theta_1,\dots,\theta_p)^\top$.

Observa-se que neste método é necessário conhecer a função de frequência dos dados trabalhados.

Para Gauss (p.5, 1992), “para estimarmos o vetor de parâmetros verdadeiros θ_0 nós escolhemos o vetor de parâmetros que maximiza a função de verossimilhança no domínio de θ . Então a Estimativa de Máxima Verossimilhança de θ é o vetor $\hat{\theta}$ que maximiza $L(\theta)$ em Ω^3 , isto é, $L(\hat{\theta}) \geq L(\theta)$ para todo $\theta \in \Omega$ ”.

Equivalentemente, segundo Dobson (p.11,2002), “ $\hat{\theta}$ é o valor que maximiza a função log-verossimilhança $\ell(\theta)=\log L(\theta)$, pois a função logarítmica é monótona. Então $\ell(\hat{\theta}) \geq \ell(\theta)$ para todo $\theta \in \Omega$.”

Em geral é mais simples trabalhar com a função log-verossimilhança que com a própria função de verossimilhança.

Por este processo fundamentar-se na determinação de um valor que maximize uma função, os teoremas que seguem são de grande valia.

Teorema B.1: Se $Y=f(x_1,\dots, x_n)$ é uma função de n variáveis, e se todas derivadas parciais $\frac{\partial Y}{\partial x_i}$ são contínuas, então Y apresenta máximo e mínimo somente nos

pontos onde $\frac{\partial Y}{\partial x_1} = \frac{\partial Y}{\partial x_2} = \dots = \frac{\partial Y}{\partial x_n} = 0$.

Teorema B.2: Se $f(x_1,\dots, x_n)$ é tal que toda primeira e segunda derivadas parciais sejam contínuas, então para o ponto onde $\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \dots = \frac{\partial f}{\partial x_n} = 0$, a função tem:

- (i) um mínimo, se a matriz \mathbf{K} , onde o ij -ésimo elemento de \mathbf{K} é $\frac{\partial^2 f}{\partial x_i \partial x_j}$, for definida positiva;
- (ii) Um máximo, se a matriz $(-\mathbf{K})$ é definida positiva.

Então, o estimador $\hat{\theta}$ é obtido pela diferenciação da função de log-verossimilhança em relação a cada elemento θ_j de θ e resolvendo-se as equações

³ Ω é denominado espaço paramétrico, o conjunto dos valores possíveis para θ .

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_j} = 0 \quad \text{para } j=1, \dots, p$$

este sistema de equações em geral é resolvido por técnicas iterativas. É necessário, então, verificar se as soluções estão em Ω , e se elas correspondem a máximos locais de $\ell(\boldsymbol{\theta})$. Portanto, pelo Teorema B.2, é que a matriz de segundas derivadas

$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$ avaliada no ponto $\hat{\boldsymbol{\theta}}$ seja negativa definida.

As estimativas determinadas pelo método da máxima verossimilhança, segundo Dobson (p.11,2002) apresentam importantes propriedades, tais como: invariância, consistência, suficiência, eficiência assintótica e normalidade assintótica.

APÊNDICE B - PARAMETRIZAÇÃO DO MODELO GERAL

Neste apêndice desenvolve-se o procedimento geral para a parametrização dos modelos expressos pelas equações contidas na Tabela 5.12.

As estimativas dos parâmetros dos modelos desenvolvidos neste trabalho foram obtidas pelo uso do Software Statistica 6,0, no qual o método de parametrização aplicado às variáveis preditoras categóricas, ou fatores, é o do sigma-restrito. Neste método, atribui-se os valores 1, 0 e -1 aos diferentes níveis de cada fator, de tal forma que “os coeficientes da regressão reflitam as mudanças de unidades na variável dependente para cada mudança de unidade na variável preditora” (Statistica Electronic Manual, sigma-restricted model).

O item 5.1 desta pesquisa apresenta um modelo geral para a estimativa do peso total do peixe considerando a covariável comprimento padrão (L_s) e dois fatores: espécie e sexo. O fator espécie apresenta quatro níveis: A (*Trachydoras paraguayensis*), C (*Liposarcus anisiti*), P (*Raphiodon vulpinus*) e H (*Serrasalmus marginatus*) e o fator sexo (S) com dois níveis, fêmea e macho. Tais variáveis preditoras associadas ao método do sigma-restrito definem a matriz de planejamento X composta por um intercepto, uma covariável, dois fatores e a interação entre os fatores, ou seja,

$$X = \begin{matrix} & \text{Int} & L_s & A & C & P & S & A*S & C*S & P*S \\ \begin{matrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{matrix} & \begin{matrix} L_{s_1} \\ L_{s_2} \\ L_{s_3} \\ L_{s_4} \\ L_{s_5} \\ L_{s_6} \\ L_{s_7} \\ L_{s_8} \end{matrix} & \begin{matrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ -1 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ -1 \\ -1 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ -1 \\ -1 \end{matrix} & \begin{matrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \end{matrix} & \begin{matrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 1 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ -1 \\ 1 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ -1 \\ 1 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ -1 \\ 1 \end{matrix} \end{matrix}$$

Pela expressão 3.4, tem-se que $\hat{\eta} = g(\hat{\mu}) = X\hat{\beta}$ e $\mu = g^{-1}(\eta_i)$, onde X é a matriz de planejamento (ou matriz do modelo), $\hat{\beta}$ é o vetor dos parâmetros estimados da regressão, cujos valores estão apresentados na tabela 5.3 e $\hat{\mu}$ é o vetor dos pesos

médios esperados para cada peixe de acordo com sua espécie, comprimento e sexo. Logo,

$$\begin{bmatrix} \hat{\mu}_{AF}^{0,35} \\ \hat{\mu}_{AM}^{0,35} \\ \hat{\mu}_{CF}^{0,35} \\ \hat{\mu}_{CM}^{0,35} \\ \hat{\mu}_{PF}^{0,35} \\ \hat{\mu}_{PM}^{0,35} \\ \hat{\mu}_{HF}^{0,35} \\ \hat{\mu}_{HM}^{0,35} \end{bmatrix} = \begin{bmatrix} 1 & Ls_1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & Ls_2 & 1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 1 & Ls_3 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & Ls_4 & 0 & 1 & 0 & -1 & 0 & -1 & 0 \\ 1 & Ls_5 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & Ls_6 & 0 & 0 & 1 & -1 & 0 & 0 & -1 \\ 1 & Ls_7 & -1 & -1 & -1 & 1 & -1 & -1 & -1 \\ 1 & Ls_8 & -1 & -1 & -1 & -1 & 1 & 1 & 1 \end{bmatrix} * \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{bmatrix}$$

cuja solução resulta em equações individuais, como segue.

i. Peso médio estimado para as fêmeas da espécie A (*Trachydoras paraguayensis*):

$$\hat{\mu}_{AF}^{0,35} = (1)b_0 + b_1Ls_1 + (1)b_2 + (1)b_5 + (1)b_6 = (b_0 + b_2 + b_5 + b_6) + b_1Ls_1$$

$$\hat{\mu}_{AF}^{0,35} = (-0,698 + 0,490 + 0,045 - 0,027) + 0,362Ls_1$$

$$\hat{\mu}_{AF} = (-0,190 + 0,362 Ls_1)^{100/35} \quad (A.1)$$

ii. Peso médio estimado para os machos da espécie A (*Trachydoras paraguayensis*):

$$\hat{\mu}_{AM}^{0,35} = (1)b_0 + b_1Ls_1 + (1)b_2 + (-1)b_5 + (-1)b_6 = (b_0 + b_2 - b_5 - b_6) + b_1Ls_1$$

$$\hat{\mu}_{AM}^{0,35} = (-0,698 + 0,490 - 0,045 + 0,027) + 0,362Ls_1$$

$$\hat{\mu}_{AM} = (-0,226 + 0,362 Ls_1)^{100/35} \quad (A.2)$$

iii. Peso médio estimado para as fêmeas da espécie C (*Liposarcus anisitsi*):

$$\hat{\mu}_{CF}^{0,35} = (1)b_0 + b_1Ls_1 + (1)b_3 + (1)b_5 + (1)b_7 = (b_0 + b_3 + b_5 + b_7) + b_1Ls_1$$

$$\hat{\mu}_{CF}^{0,35} = (-0,698 - 0,607 + 0,045 + 0,156) + 0,362Ls_1$$

$$\hat{\mu}_{CF} = (-1,104 + 0,362 Ls_1)^{100/35} \quad A.3$$

iv. Peso médio estimado para os machos espécie C (*Liposarcus anisitsi*):

$$\hat{\mu}_{CM}^{0,35} = (1)b_0 + b_1Ls_1 + (1)b_3 + (-1)b_5 + (-1)b_7 = (b_0 + b_3 - b_5 - b_7) + b_1Ls_1$$

$$\hat{\mu}_{CM}^{0,35} = (-0,698 - 0,607 - 0,045 - 0,156) + 0,362Ls_1$$

$$\hat{\mu}_{CM} = (-1,506 + 0,362Ls_1)^{100/35} \quad A.4$$

v. Peso médio estimado para as fêmeas da espécie P (*Leporinus friderici*):

$$\hat{\mu}_{PF}^{0,35} = (1)b_0 + b_1Ls_1 + (1)b_4 + (1)b_5 + (1)b_8 = (b_0 + b_4 + b_5 + b_8) + b_1Ls_1$$

$$\hat{\mu}_{PF}^{0,35} = (-0,698 - 0,137 + 0,045 - 0,087) + 0,362Ls_1$$

$$\hat{\mu}_{PF} = (-0,877 + 0,362Ls_1)^{100/35} \quad A.5$$

vi. Peso médio estimado para os machos da espécie P (*Leporinus friderici*):

$$\hat{\mu}_{PM}^{0,35} = (1)b_0 + b_1Ls_1 + (1)b_4 + (-1)b_5 + (-1)b_8 = (b_0 + b_4 - b_5 - b_8) + b_1Ls_1$$

$$\hat{\mu}_{PM}^{0,35} = (-0,698 - 0,137 - 0,045 + 0,087) + 0,362Ls_1$$

$$\hat{\mu}_{PM} = (-0,793 + 0,362Ls_1)^{100/35} \quad A.6$$

vii. Peso médio estimado para as fêmeas da espécie H (*Serrasalmus marginatus*):

$$\hat{\mu}_{HF}^{0,35} = (1)b_0 + b_1Ls_1 + (-1)b_2 + (-1)b_3 + (-1)b_4 + (1)b_5 + (-1)b_6 + (-1)b_7 + (-1)b_8$$

$$\hat{\mu}_{HF}^{0,35} = (b_0 - b_2 - b_3 - b_4 + b_5 - b_6 - b_7 - b_8) + b_1Ls_1$$

$$\hat{\mu}_{HF}^{0,35} = (-0,698 - 0,490 + 0,607 + 0,137 + 0,045 + 0,027 - 0,156 + 0,087) + 0,362Ls_1$$

$$\hat{\mu}_{HF} = (-0,441 + 0,362Ls_1)^{100/35} \quad A.7$$

viii. Peso médio estimado para os machos da espécie H (*Serrasalmus marginatus*):

$$\hat{\mu}_{HM}^{0,35} = (1)b_0 + b_1Ls_1 + (-1)b_2 + (-1)b_3 + (-1)b_4 - (1)b_5 + (1)b_6 + (1)b_7 + (1)b_8$$

$$\hat{\mu}_{HM}^{0,35} = (b_0 - b_2 - b_3 - b_4 - b_5 + b_6 + b_7 + b_8) + b_1Ls_1$$

$$\hat{\mu}_{HM}^{0,35} = (-0,698 - 0,490 + 0,607 + 0,137 - 0,045 - 0,027 + 0,156 - 0,087) + 0,362Ls_1$$

$$\hat{\mu}_{HM} = (-0,447 + 0,362Ls_1)^{100/35}$$

APÊNDICE C – EXEMPLOS PARA O FATOR DE CONDIÇÃO DO PEIXE

Neste apêndice busca-se verificar a aplicabilidade do Fator de Condição K do peixe, construído segundo a teoria dos Modelos Lineares Generalizados, e compará-lo a outros modelos citados em obras de ictiologia, conforme citado neste trabalho (seção 5.3). Os fatores de condição segundo cada modelo são, respectivamente:

$$(i) \mathbf{K} = \frac{\mathbf{W}}{\mathbf{L}^b} \quad (ii) \mathbf{K} = \frac{\mathbf{W}}{\mathbf{L}^3} \quad (iii) \mathbf{K} = \ln(\mathbf{W}) - b\ln(\mathbf{L}) \quad (iv) \mathbf{K} = \mathbf{W}^\lambda - b_1\mathbf{L}$$

A determinação do valor de K para cada espécie de peixe é feita tendo como referencial um certo comprimento L, contido no intervalo de observações encontradas na pesquisa. A este comprimento são atribuídos pesos W, cujos valores podem ser considerados: superiores, inferiores ou compatíveis ao modelo especificado. A seguir, os resultados obtidos são confrontados com os intervalos identificados para cada modelo. Caso o resultado obtido para este valor particular de K esteja contido dentro do intervalo definido pelo modelo, considera-se o desenvolvimento do peixe adequado. Se o valor calculado para K não pertencer a este intervalo, há indícios que o peixe apresenta um desenvolvimento corporal inadequado ao padrão, sugerindo, então, um estudo mais aprofundado das condições gerais em que o animal vive.

Para auxiliar a interpretação de tais resultados, a Figura C-1 mostra os peso estimados sobrepostos aos pesos observados. Os gráficos apresentados nesta figura foram construídos segundo cada espécie e modelo, e têm o propósito de fornecer condições para uma comparação global entre o ajuste produzido pela aplicação dos Modelos Lineares Generalizados e os demais modelos.

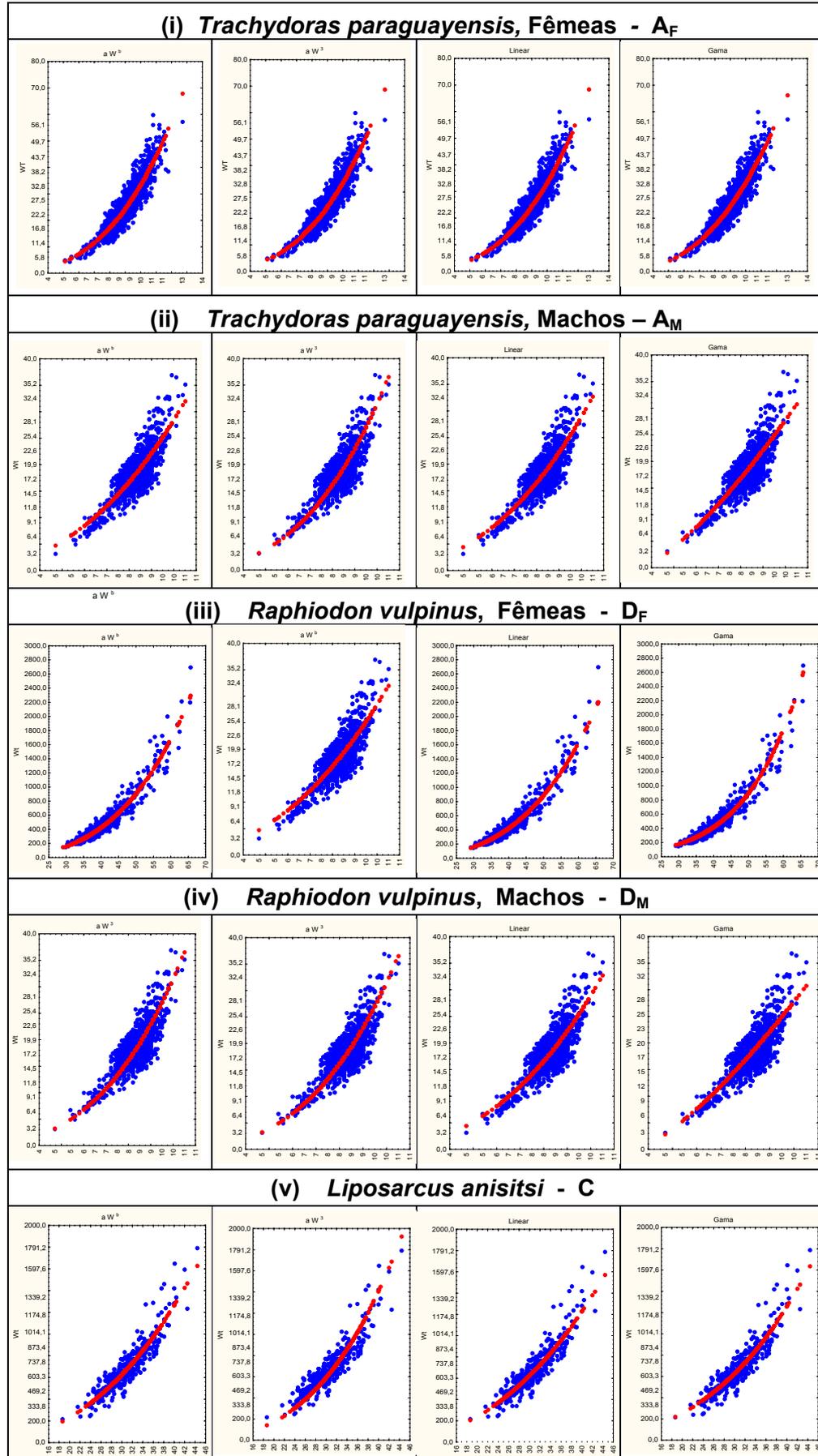


Figura C-1: Sobreposição dos valores ajustados aos valores observados. (Continua)

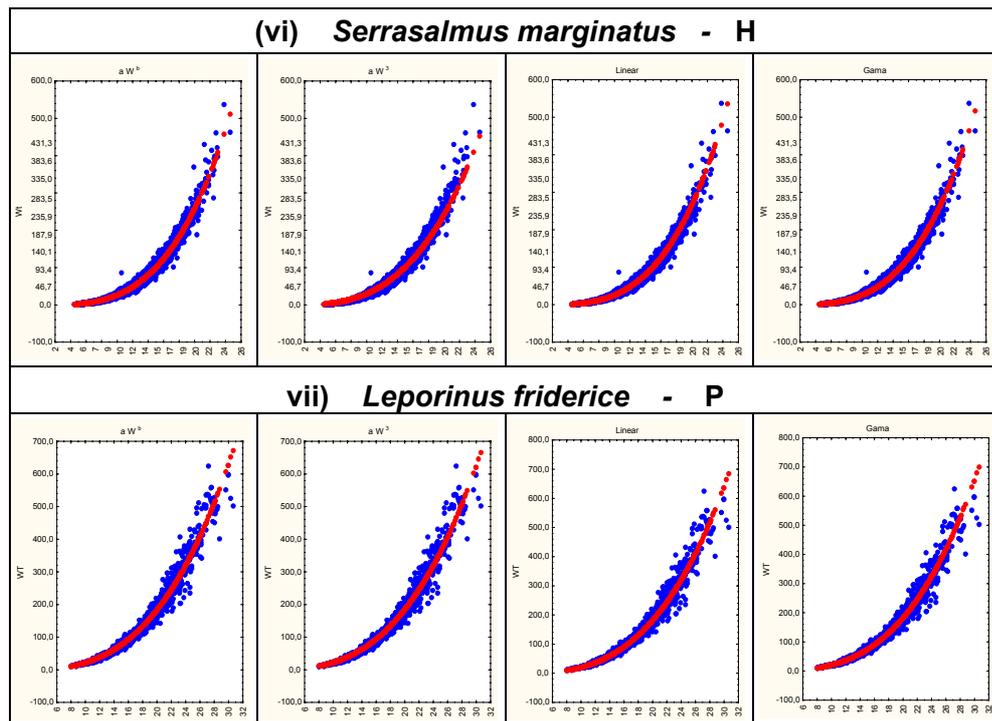


Figura C-1: Sobreposição dos valores ajustados aos valores observados.

Nota-se que para as fêmeas da espécie *Trachydoras paraguayensis* (A_F), a espécie *Serrasalmus marginatus* (H) e a espécie *Leporinus friderice* (P), não há uma grande discrepância entre o ajuste dos dados obtidos por qualquer um dos quatro modelos. O modelo Gama para os machos da espécie *Trachydoras paraguayensis* (A_M) e os machos da espécie *Raphiodon vulpinus* (D_M) mostra uma tendência a subestimar o peso total do animal, apresentando um desvio maior para comprimentos mais elevados. Já as fêmeas da espécie *Raphiodon vulpinus* (D_F) apresentam o ajuste pelo modelo Gama bem mais próximos aos valores observados, enquanto os demais modelos apresentam uma tendência a subestimar os pesos. A espécie *Liposarcus anisitsi* © apresenta os menores resíduos segundo o modelo Gama.

Os resultados expostos na Tabela C-1 foram obtidos conforme especificado no segundo parágrafo deste apêndice. Nesta mesma tabela estão contidos os intervalos de cada Fator de Condição K, segundo cada espécie e modelo. Os valores em vermelho correspondem aos indivíduos que, segundo o modelo, não tem desenvolvimento corpóreo de acordo com o padrão proposto.

Tabela C-1: Fator de condição K para o peixe considerando comprimentos do peixe com pesos distintos.

Espécie	W	L	MODELOS			
			$K = \frac{W}{L^b}$	$K = \frac{W}{L^3}$	$K = \ln(W) - b_1 \cdot \ln(L)$	$K = W^\lambda - b_1 \cdot L_s$
<i>Trachydoras paraguayensis</i> Fêmeas	Intervalos		(0,024 ; 0,054)	(0,022 ; 0,050)	(0,022 ; 0,050)	(-1,279 ; 0,023)
	13,0	9	0,020	0,018	0,018	-1,474
	15,5	9	0,023	0,021	0,021	-1,271
	18,0	9	0,027	0,025	0,025	-1,087
	24,7	9	0,037	0,034	0,034	-0,658
	37,8	9	0,057	0,052	0,052	0,011
	38,0	9	0,057	0,052	0,052	0,020
	38,2	9	0,057	0,052	0,052	0,029
	20,0	10	0,022	0,020	0,020	-1,424
	34,5	10	0,038	0,035	0,035	-0,616
	49,3	10	0,054	0,049	0,049	0,017
	49,6	10	0,054	0,050	0,050	0,028
	49,7	10	0,054	0,050	0,050	0,032
<i>Trachydoras paraguayensis</i> Machos	Intervalo		(0,081 ; 0,171)	(0,021 ; 0,046)	(0,064 ; 0,133)	(-8,346 ; -3,195)
	4,2	6,5	0,050	0,015	0,040	1,125
	10,0	6,5	0,118	0,036	0,095	1,273
	17,6	6,5	0,208	0,064	0,166	1,366
	17,7	6,5	0,210	0,064	0,167	1,416
	14,0	9,0	0,077	0,019	0,059	1,162
	30,0	9,0	0,164	0,041	0,126	1,270
	40,0	9,0	0,219	0,055	0,168	1,314
	14,0	9,0	0,077	0,019	0,059	1,418
	15,9	9,8	0,071	0,017	0,054	-9,417
	23,5	9,8	0,105	0,025	0,080	-6,931
	35,0	9,8	0,157	0,037	0,119	-3,549
	41,5	9,8	0,186	0,044	0,141	-1,774
45,0	9,8	0,201	0,048	0,153	-0,850	
<i>Raphiodon vulpinus</i> Fêmeas	Intervalo		(0,001 ; 0,002)	(0,005 ; 0,011)	(0,001 ; 0,003)	(1,163 ; 1,414)
	87,0	29,1	0,001	0,0045	0,001	1,125
	160,5	29,1	0,002	0,007	0,002	1,273
	230,0	29,1	0,003	0,009	0,003	1,366
	278,0	29,1	0,003	0,011	0,004	1,416
	216,0	39,5	0,001	0,004	0,001	1,113
	270,0	39,5	0,001	0,004	0,001	1,172
	277,0	39,5	0,001	0,004	0,001	1,179
	420,0	39,5	0,002	0,007	0,002	1,294
	700,0	54,0	0,001	0,004	0,001	1,115
	790,0	54,0	0,001	0,005	0,001	1,152
	820,0	54,0	0,001	0,005	0,002	1,164
	1300,0	54,0	0,002	0,008	0,002	1,311
<i>Raphiodon vulpinus</i> Machos	Intervalo		(0,002 ; 0,005)	(0,004 ; 0,009)	(0,003 ; 0,007)	(-0,362 ; 1,039)
	10,0	14,7	0,002	0,003	0,003	-0,117
	16,0	14,7	0,003	0,005	0,004	0,185
	20,0	14,7	0,004	0,006	0,005	0,344
	47,0	14,7	0,009	0,015	0,012	1,062
	300,0	41,2	0,002	0,004	0,003	-0,385
	400,0	41,2	0,003	0,006	0,004	0,114
	650,5	41,2	0,004	0,009	0,007	1,061
	850,0	41,2	0,006	0,012	0,009	1,645
	438,0	50,0	0,002	0,004	0,003	-0,984
	550,0	50,0	0,002	0,004	0,003	-0,546
	690,0	50,0	0,003	0,006	0,004	-0,079
	1210,0	50,0	0,004	0,010	0,007	1,226
1300,0	50,0	0,005	0,010	0,008	1,409	

Tabela C-1: Fator de condição K para o peixe considerando comprimentos do peixe com pesos distintos.

Espécie	W	L	MODELOS			
			$K = \frac{W}{L^b}$	$K = \frac{W}{L^3}$	$K = \ln(W) - b_1 \cdot \ln(L)$	$K = W^\lambda - b_1 \cdot L_s$
<i>Liposarcus anisitsi</i>	Intervalo		(0,106 ; 0,233)	(0,015 ; 0,035)	(0,145 ; 0,321)	(1,005 ; 3,108)
	100,0	18,6	0,082	0,016	0,107	0,729
	220,0	18,6	0,181	0,034	0,235	1,982
	300,0	18,6	0,247	0,047	0,321	2,568
	350,0	18,6	0,288	0,054	0,374	2,882
	390,0	18,6	0,321	0,061	0,417	3,111
	500,0	36,5	0,080	0,010	0,110	0,170
	650,0	36,5	0,104	0,013	0,144	0,810
	817,8	36,5	0,131	0,017	0,181	1,416
	1500,0	36,5	0,240	0,031	0,331	3,248
	1050,0	44,4	0,104	0,012	0,147	0,583
	1300,0	44,4	0,129	0,015	0,182	1,239
	1790,0	44,4	0,178	0,020	0,250	2,308
2250,0	44,4	0,223	0,026	0,314	3,142	
2320,0	44,4	0,230	0,027	0,324	3,258	
<i>Serrasalmus marginatus</i>	Intervalo		(0,014 ; 0,030)	(0,016 ; 0,033)	(0,012 ; 0,024)	(-0,728 ; 0,990)
	1,2	4,5	0,009	0,013	0,008	-0,228
	8,0	4,5	0,058	0,088	0,052	0,620
	15,0	4,5	0,110	0,165	0,097	1,030
	78,0	16,4	0,008	0,018	0,007	-0,826
	100,0	16,4	0,011	0,023	0,009	-0,517
	150,0	16,4	0,016	0,034	0,013	0,042
	278,0	16,4	0,030	0,063	0,024	1,038
	295,0	24,6	0,008	0,020	0,006	-1,199
	360,0	24,6	0,010	0,024	0,008	-0,827
470,0	24,6	0,013	0,032	0,010	-0,293	
850,0	24,6	0,024	0,057	0,019	1,065	
<i>Leporinus friderice</i>	Intervalo		(0,014 ; 0,030)	(0,016 ; 0,033)	(0,012 ; 0,024)	(-0,602 ; 0,773)
	11,0	8,0	0,020	0,021	0,017	0,149
	25,4	8,0	0,047	0,050	0,040	0,771
	25,6	8,0	0,047	0,050	0,041	0,778
	25,7	8,0	0,047	0,050	0,041	0,781
	15,0	12,4	0,007	0,008	0,006	-0,714
	43,0	12,4	0,021	0,023	0,018	0,180
	70,0	12,4	0,034	0,037	0,029	0,703
	75,0	12,4	0,036	0,039	0,031	0,784
	100,0	20,0	0,011	0,013	0,009	-0,717
	115,0	20,0	0,013	0,014	0,011	-0,533
145,0	20,0	0,017	0,018	0,013	-0,208	
200,0	20,0	0,023	0,025	0,019	0,282	
258,0	20,0	0,029	0,032	0,024	0,706	

É interessante observar que, além das considerações já exposta neste trabalho, o modelo Gama proposto oferece maior sensibilidade quanto aos valores de K. Um exemplo está nas fêmeas da espécie *Raphiodon vulpinus*, pois para pesos entre 216 a 277 gramas, uma variação de 61 gramas no peso, os demais modelos não detectaram mudanças no fator, enquanto pelo modelo Gama, a variação do Fator de Condição K variou de 1,113 a 1,179 unidades, havendo uma mudança na classificação do indivíduo.