



UNIVERSIDADE FEDERAL DE LAVRAS

**ESTIMATIVA DE COMPONENTES DE ROTATIVIDADE E
TAMANHO DA FROTA DE VEÍCULOS ESTACIONADOS NO
CENTRO DE LAVRAS – MG VIA ENSAIOS DE CAPTURA-
RECAPTURA**

JOSÉ OTAVIANO DE OLIVEIRA

2003

JOSÉ OTAVIANO DE OLIVEIRA

**ESTIMATIVA DE COMPONENTES DE ROTATIVIDADE E
TAMANHO DA FROTA DE VEÍCULOS ESTACIONADOS NO
CENTRO DE LAVRAS – MG VIA ENSAIOS DE CAPTURA-
RECAPTURA**

Dissertação apresentada á Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

Orientador
Prof. Júlio Sílvio de Sousa Bueno Filho

LAVRAS
MINAS GERAIS – BRASIL
2003

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Oliveira, José Otaviano de

Estimativa de componentes de rotatividade e tamanho da frota de veículos estacionados no centro de Lavras-MG via ensaios de captura-recaptura / Otaviano José de Oliveira. -- Lavras : UFLA, 2003.

94 p. : il.

Orientador: Júlio Silvio de Sousa Bueno Filho.

Dissertação (Mestrado) – UFLA.

Bibliografia.

1. Modelo Linear Generalizado 2. Captura-recaptura. 3. Modelo de Poisson. I. Universidade Federal de Lavras. II. Título.

CDD-519.5

JOSÉ OTAVIANO DE OLIVEIRA

**ESTIMATIVA DE COMPONENTES DE ROTATIVIDADE E
TAMANHO DA FROTA DE VEÍCULOS ESTACIONADOS NO
CENTRO DE LAVRAS – MG VIA ENSAIOS DE CAPTURA-
RECAPTURA**

Dissertação apresentada á Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

Aprovado em Sexta-feira, 28 de fevereiro de 2003.

Prof. Dr. Júlio Neil Cassa Louzada

UFLA

Prof. Dr. Lucas Monteiro Chaves

UFLA



Prof. Júlio Sílvio de Sousa Bueno Filho
UFLA
(Orientador)

LAVRAS
MINAS GERAIS – BRASIL
2003

Ao meu pai, Gondes , minha mãe
Maria das Dores ("in memoriam") e
aos meus irmãos

OFEREÇO

À minha esposa Suely,
ao meu filho Carlos Otávio
e sobrinha Patrícia

DEDICO

AGRADECIMENTOS

A **DEUS**, por tudo...

À Escola Agrotécnica Federal Antônio José Teixeira pela oportunidade, em especial ao diretor geral Ariomar Rodrigues dos Santos.

À Universidade Federal de Lavras (UFLA) e ao Departamento de Ciências Exatas (DEX) pela oportunidade.

À Coordenadoria de Aperfeiçoamento de Pessoal do Ensino Superior (CAPES), pela concessão da bolsa de estudos.

Aos professores e funcionários do DEX pelo convívio e transmissão de ensinamentos, de forma especial ao professor Mário.

Ao professor orientador Júlio pela orientação, sugestões e paciência na elaboração deste trabalho.

Aos amigos Daniel, Antônio Carlos, Clifford, Rafael, Luiz, Arquiteclino, Paulo José, Júlio, José Waldemar, Janser e Osvaldo.

À turma de 2001, Marcelo Cirillo, José Airton, Alessandra, José Ermelindo, João Marcos, Washington, Luís Alberto, Lúcia, Ana Lúcia, Élcio, Waldemar, pela amizade.

Aos professores membros da banca examinadora, Júlio Louzada e Lucas pelas sugestões e críticas apresentadas.

Ao Sinésio pelo atenção e trabalho de digitação desta monografia.
Aos demais colegas pelo o incentivo, companheirismo e amizade.

À Universidade Federal de Ouro Preto especialmente ao professor Américo e ao professor Murilo.

À Faculdade de Filosofia Ciências e Letras de Caratinga e aos colegas da turma de 95.

SUMÁRIO

Resumo geral.....	i
Abstract.....	ii
CAPÍTULO 1.....	1
1 Introdução geral.....	2
1.1 Objetivos.....	3
1.2 Referências bibliográficas.....	5
CAPÍTULO 2.....	6
2. Referencial teórico.....	7
2.1 Modelos de população aberta.....	7
2.1.1 Modelo de Jolly-Seber.....	7
2.1.2 Combinação de modelos de população fechada e aberta	9
2.2 Modelos de população fechada.....	10
2.2.1 Modelo de Lincoln-Petersen.....	10
2.2.2 Modelo de Schnabel.....	11
2.2.3 Modelos gerais para população fechada.....	13
2.3 Abordagem log-linear em população aberta (Cormack.1989).....	17
2.3.1 O modelo biológico-demográfico.....	17
2.3.2 Parâmetros do modelo de regressão e seus correspondentes biológicos	19
2.3.3 Correção para modelos de captura.....	22

2.3.4 Tabelas de contingência	23
2.3.5 Caracterização dos modelos log-lineares Poisson.....	24
2.3.6 Forma canônica.....	25
2.4 Abordagem bayesiana em população fechada (Smith – 1988).....	26
2.4.1 Teorema de bayes.....	27
2.4.2 Informação a priori e a função de verossimilhança.....	29
2.4.3 Teoria da decisão.....	30
2.5 Referências bibliográficas.....	33
CAPÍTULO 3.....	35
Resumo	36
Abstract.....	37
3.1 Introdução.....	38
3.2 Material e métodos.....	39
3.2.1 Delimitação da área de estudo.....	39
3.2.2 Matriz do modelo.....	40
3.2.3 Solução das equações.....	46
3.2.4 Algoritmo de ajuste do modelo linear generalizado a dados com distribuição Poisson.....	48
3.2.5 Critério para análise e seleção dos modelos.....	50
3.3 Resultados e discussão.....	53
3.3.1 Freqüência das histórias de captura.....	53
3.3.2 Estimativas dos parâmetros dos modelos.....	54
3.3.3 Intervalo de confiança para o tamanho populacional...	61
3.4 Referências bibliográficas.....	64

CAPÍTULO 4	65
Resumo	66
Abstract	67
4.1 Introdução	68
4.2 Material e métodos	70
4.2.1 Descrição do experimento	70
4.2.2 Estimação do tamanho populacional	71
4.3 Resultados e discussão	76
4.3.1 Posteriori conjunta	76
4.3.2 Resultados para a priori uniforme	77
4.3.3 Funções de risco e estimadores pontuais	78
4.3.4 Estimadores de máxima verossimilhança	81
4.3.5 Resultados para a priori $\Gamma 1, M_{i+1}$	81
4.3.6 Funções de risco e estimadores pontuais	81
4.3.7 Intervalo de confiança simétrico	84
4.4 Referências bibliográficas	87
CAPÍTULO 5 (considerações gerais)	88
5.1 Populações abertas (modelo generalizado de regressão)	88
5.2 Modelo de Schnabel (população fechada)	89
5.3 Conclusão	91
6 Referências bibliográficas	92

RESUMO GERAL

OLIVEIRA, José Otaviano de. Estimativa de componentes de rotatividade e tamanho da frota de veículos estacionada no centro de Lavras - MG via ensaios de captura-recaptura. 2003. 94 p. Dissertação (Mestrado em Agronomia/ Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras-MG*

Os modelos de captura-recaptura podem ser aplicados a estudos de tamanho e de dinâmica de populações, notadamente aquelas de difícil censo completo, como o é caso de populações selvagens (peixes, répteis, anfíbios, insetos). Tais modelos têm sido utilizados para estimar tamanhos de populações em diversas aplicações ecológicas e ou sócio-econômicas. O objetivo deste trabalho foi estimar o tamanho populacional da frota de veículos estacionada no centro da cidade de Lavras – MG e detectar seus possíveis componentes de rotatividade em um determinado período. Foram aplicadas duas abordagens ao problema. Inicialmente partiu-se dos modelos de captura para populações abertas para detectar as possíveis entradas e saídas de veículos da frota estacionada ou outras fontes de variação que pudessem causar algum viés nas estimativas. Neste caso foram utilizados os modelos de regressão Poisson (log-lineares). Verificou-se, por meio dos modelos log-lineares, que a população apresenta comportamento fechado. Na segunda abordagem utilizou-se o modelo de Schnabel (1938), sob o enfoque bayesiano, para estimar o tamanho da frota com menor erro, supondo que a população é fechada. As estimativas pontuais obtidas para o tamanho populacional nas abordagens foram muito semelhantes. A estimativa do tamanho da frota ficou em 6817 veículos, com intervalo de credibilidade de 95% entre 4360 e 9785 veículos.

* Prof. orientador: Júlio Sílvio de Sousa Bueno Filho – UFLA.

CAPÍTULO 1

INTRODUÇÃO GERAL

1 INTRODUÇÃO GERAL

Os modelos de captura-recaptura têm sido aplicados para estimar parâmetros demográficos (tamanho populacional, sobrevivência recrutamento, emigração e imigração) em populações selvagens. Recentemente também vem sendo aplicados a um grande número de problemas fora desta área. Outra aplicação está no estudo da reincidência em criminalidade, aplicada a populações de presidiários, contagem de sem tetos, usuários de drogas entre outros. Este tipo de população não permite um censo completo de fácil realização. Estes modelos são particularmente importantes nesta situação.

Historicamente, a aplicação dos modelos de captura-recaptura iniciou com Laplace, em 1786, mas sua aplicação crescente se deu apenas no início do século XX com o modelo de Lincoln-Petersen, aplicado ao estudo do tamanho de populações utilizando apenas duas ocasiões de amostragem sendo uma captura, em que todos os indivíduos são marcados e uma recaptura. Extensão deste modelo para múltiplas ocasiões de amostragem, foi apresentada por Schnabel em 1938 (Pereira, 1995) ; (Schwarz & Seber 1999) e Krebs (1989).

Jolly (1965) e Seber (1965) citado por Jolly (1982), apresentaram o modelo para populações abertas onde consideraram adições e deleções na população em estudo. Uma abordagem complementar do modelo tradicional de Jolly-Seber é dada por Cormack (1989) o qual substitui a aplicação tradicional da distribuição multinomial por uma distribuição Poisson que é mais flexível em termos de facilitar a obtenção de sua verossimilhança.

O problema de estimação do tamanho de uma população pela técnica de captura-recaptura, em sua forma mais simplificada, consiste em capturar uma amostra desta população, marcar seus elementos de algum modo e devolvê-la à população original; em seguida, recaptura-se uma nova amostra da população. A

partir dos resultados, obtém-se uma estimativa da população através da igualdade das proporções na população e na amostra recapturada.

Smith (1988) apresenta, em seu trabalho, uma abordagem bayesiana para encontrar o tamanho de população finita fechada (modelo de Schnabel), sob o enfoque da Teoria da Decisão mostrando diversas priores e diversas funções de perda. Neste modelo é usado apenas um tipo de marca, isto é, a população se divide em marcados, capturados alguma vez durante a série de amostras, e os não marcados, nunca capturado antes.

No presente trabalho, os modelos de captura-recaptura foram aplicados ao estudo do comportamento da frota de veículos estacionada na área central da cidade de Lavras – MG a fim de detectar a rotação de veículos (entrada e saída) na área de estacionamento. Inicialmente assumiu-se que esta população era aberta e que um modelo apropriado para o estudo é o chamado Jolly-Seber (J-S), com aplicação dos modelos log-lineares, que utiliza a história de captura para detectar parâmetros de entrada e de saída de veículos. Por meio destes é possível estimar a rotação dos grupos de veículos (o que será feito no Capítulo III).

Além disso, foi adotado um modelo de população fechada para buscar melhorar as estimativas de tamanho populacional, utilizando uma abordagem bayesiana (o que está apresentado no capítulo IV). No capítulo V são apresentadas as considerações finais sobre a aplicação de cada uma das abordagens consideradas e a conclusão geral.

1.1 Objetivos

Os objetivos principais deste trabalho foram:

- Estimar os componentes de rotatividade da frota de veículos estacionada na centro da cidade de Lavras – MG;

- Discutir aspectos dos modelos lineares generalizados para o caso da distribuição Poisson, aplicados ao modelo de captura;
- Estimar o tamanho da frota de veículos estacionados no centro de Lavras – MG;
- Discutir alguns aspectos gerais do processo de estimação do tamanho populacional, considerando a abordagem bayesiana.

1.2 REFERÊNCIAS BIBLIOGRÁFICAS

CORMACK, R. M. Log-linear models for capture-recapture. **Biometrics**, Alexandria, v.45, n. 3, p. 395-413, June 1989.

JOLLY, G. M. Mark-recapture models with parameters constant in time. **Biometrics**, Alexandria, v. 38, n. 3, p.301-321, June 1982.

KREBS, W. J. C. **Ecological Methodology**. New York: Editora & Row, Publishers, 1989.

PEREIRA, J. E. **Análise de experimentos de captura-recaptura através de modelos log-lineares**. 1995. 106 p. Dissertação (Mestrado) – Escola Superior de Agricultura “Luiz de queiroz”, Piracicaba, SP.

SCHWARZ, C. J; SEBER G. A. F. Estimating animal abundance: review III. **Statistics Science**, n. 14, p. 427-456. 1999.

CAPÍTULO 2

REFERENCIAL TEÓRICO

2 REFERENCIAL TEÓRICO

2.1 Modelos de população aberta

No estudo de componentes demográficos por meio de ensaios de captura-recaptura, notadamente aqueles de longa duração, não é possível assumir que a população é fechada. Neste caso deve-se considerar o modelo para população aberta. Uma população é genericamente considerada aberta quando apresentar migrações, nascimentos ou mortes ou alguma combinação destes fatores. Isto é o que normalmente ocorre com populações biológicas, as quais estão sofrendo variação demográfica por deleções ou adições de forma permanente ou temporária.

2.1.1 Modelo de Jolly-Seber

O modelo básico que aborda esta situação foi introduzido independentemente por Jolly e Seber em 1965, ambos citados por Jolly (1982) e Krebs (1989) desenvolveram o modelo de produto de multinomiais. São quatro modelos que visam incorporar diferentes fontes de variação da probabilidade de captura e detectar nascimentos, mortes, migração e a razão de sobrevivência dos indivíduos. O modelo A é o mais completo e, segundo Brown Jr. (2001), possui mais parâmetros a estimar que o número de variáveis observáveis. A função de verossimilhança se torna muito complexa e algumas restrições devem ser adotadas. O modelo B tem razão de sobrevivência assumido constante com a probabilidade de captura variando no tempo. Já no modelo C, é assumida a constância para a probabilidade de captura e a razão de sobrevivência é variável a cada ocasião amostral. O modelo D, o menos parametrizado de todos, apresenta razão de sobrevivência e probabilidade de captura constantes.

Algumas restrições prévias para este modelo são apresentadas e descritas por Krebs (1989) e extraídas de Pereira (1995):

- Cada animal presente na população no momento da i -ésima amostra ($i=1, 2, \dots, s$) tem a mesma probabilidade de captura (p_i).

Esta é uma situação bastante restritiva para a utilização do modelo, e pode ser detectada pelo modelo de Cormack (1989).

- Cada animal marcado, presente na população imediatamente após a i -ésima amostra, tem a mesma probabilidade de sobrevivência (ϕ_i) até a $(i+1)$ -ésima ocasião de amostragem ($i=1, 2, \dots, s-1$).

Assume-se a probabilidade de sobrevivência constante. O modelo de Cormack (1989) é mais flexível, porém as estimativas mais precisas de sobrevivência necessitam de modelos específicos.

- As marcas não são perdidas.

A perda das marcas pode ser a causa de viés nas estimativas por falta de identificação e, conseqüentemente, da composição das frequências dos grupos. Outro problema é causado pelo constrangimento do indivíduo durante o processo de captura causada pela armadilha no caso de estudo de populações selvagens ou de armadilhas de amostragem tipo “trap happy”. Neste tipo de armadilha, o indivíduo é favorecido e sua probabilidade de captura será maior que a dos demais que ainda não foram capturados.

- Todas as amostragens são instantâneas e cada soltura é feita imediatamente após a captura.

Embora este modelo seja considerado o mais importante, tanto no aspecto prático quanto teórico; apresenta alguns inconvenientes como as estimativas fora do espaço paramétrico, como o caso de estimativas de probabilidade de taxas de sobrevivência maiores que um ou taxas negativas para nascimentos e mortes. A sugestão dada por Cormack (1989) neste caso, é simplesmente desconsiderar o parâmetro e proceder a um novo ajuste do

modelo. O modelo apresenta um conjunto de restrições que limita o seu uso em algumas situações. São os casos, por exemplo, da migração temporária e da diferença de probabilidade de captura entre indivíduos em uma mesma ocasião de amostragem.

Para solucionar o problema da dependência entre as amostras, podem-se aplicar os modelos log-lineares. Estes modelos permitem modelar a dependência e, ao mesmo tempo, transformar estes efeitos em parâmetros demográficos estimáveis. O efeito do comportamento também pode ser detectado por interações de segunda ordem entre as amostras vizinhas. Em algumas situações não é possível estimar isoladamente os efeitos, mas algumas combinações de fatores. Para modelar migrações temporárias é necessário modelos específicos.

Para estudar populações que apresentem alguma das restrições anteriores, pode-se utilizar o modelo de delineamento robusto proposto por Pollock (1982) e apresentado Nichols et al.,1994.

2.1.2 Combinação de modelos de população fechada e aberta

O delineamento amostral robusto à heterogeneidade de captura ou resposta à armadilha, ou ambos, foi introduzido por Pollock (1982) citado por Nichols et al.(1994), com o objetivo de reduzir o viés das estimativas do tamanho populacional e de outros parâmetros. Este modelo permite uma análise que usa metodologias de modelos para população fechada e aberta conjuntamente. O delineamento consiste na seqüência de capturas espaçadas no tempo. Considere a representação de um experimento de amostragem de captura-recaptura dada pela Figura 2.1.1 em que se têm S períodos de amostragem primários (por exemplo, anos) e, dentro de cada um desses, períodos de amostragem secundários (por exemplo, meses, semanas).

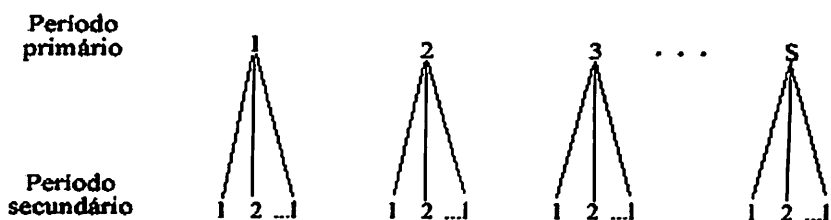


FIGURA 2.1.1 – Esquema para amostragem pelo delineamento robusto.

Nesta situação, é possível estimar componentes demográficos considerando a população fechada nos períodos primários. O modelo de Jolly-Seber é aplicado no estudo entre as amostras dentro de cada período primário. Este delineamento também pode ser usado para estimar emigração temporária. É possível aplicar este mesmo delineamento para estimar, separadamente, recrutamento e imigração. Embora este modelo seja importante em diversas situações práticas, ele não será abordado neste trabalho.

2.2 Modelos de população fechada

2.2.1 Modelo de Lincoln-Petersen

Uma população é considerada fechada quando apresentar, mesmo que aproximadamente, uma clausura geográfica (área restrita) e demográfica (não há nascimento, migrações ou mortes). Efetivamente, a população deverá permanecer imutável no sentido demográfico durante o período que ocorrer o estudo. Obviamente, em se tratando de população biológica, esta condição é muito restritiva se o tempo de estudo for longo em relação ao período de

sobrevivência e de reprodução dos indivíduos. O Procedimento básico para a estimação do tamanho populacional, pelo o modelo de Lincoln – Petersen, é o mais simples entre todos os modelos de Captura-recaptura aplicado à população fechada, consiste em marcar um número de indivíduos em um curto período e soltá-los novamente. Em seguida, recapturar e checar os marcados dividindo-os em duas etapas distintas: marca e recaptura. A segunda amostra deverá ser aleatória para o método ser válido; isto é, marcados e não marcados devem ter a mesma chance de serem capturados na segunda amostra. Os dados obtidos são:

n_1 : número de indivíduos marcados na 1ª amostra;

n_2 : total de indivíduos capturados na 2ª amostra;

m_2 : número de indivíduos na 2ª amostra que são marcados;

N : tamanho populacional

A expressão do estimador de máxima verossimilhança é dada pela expressão:

$$\hat{N} = \frac{n_1 n_2}{m_2}$$

Este estimador só pode ser considerado nos caso de amostras independentes. Segundo Chao (2001), se houver dependência entre as amostras, o estimador acima é viesado. Se as duas amostras são positivamente correlacionadas por uma amostragem tipo “trap-happy”, o tamanho populacional será subestimado. Se amostragem for correlacionada negativamente, haverá uma superestimação do tamanho populacional.

2.2.2 Modelo de Schnabel

A condição de clausura geográfica e demográfica é desejável para utilização dos modelos de populações fechadas em situações de recaptura única ou de múltiplas ocasiões de recaptura. Krebs (1989) descreve o modelo de

Schnabel como um modelo idêntico ao modelo Linclon-Petersen (L-P) para uma série de amostras. É usado apenas um tipo de marca, isto é, a população se divide em marcados, capturados alguma vez durante a série de amostras, e os não marcados, nunca capturado antes. As pressuposições do modelo de Schnabel são idênticas às do modelo de Petersen, para o qual não se considera retorno, resposta à armadilha ou variação temporal das condições da população que afetam seu retorno.

- Considere uma população fechada de tamanho N da qual retiram-se amostras em k ocasiões, sendo $k \geq 1$, tamanho n_i ($i = 0, 1, \dots, s$). Na primeira amostra tem-se n_0 indivíduos.

Na i 'ésima ocasião amostral têm-se:

- n_i indivíduos capturados, dos quais m_i são marcados;
- Os $(n_i - m_i)$ indivíduos são não marcados e devolvidos à população original.
- Seja

$$M_i = \sum_{j=0}^{i-1} (n_j - m_j)$$

que denota o número marcados na população exatamente antes da i -ésima ocasião amostral e $m_0 = 0$, em que n_j é o número de animais capturados na j -ésima amostra e m_j é o número de animais marcados capturados na j -ésima amostra. Esse estimador equivale a uma extensão do estimador de Petersen para j ocasiões de amostragem. Desta forma, considerando a distribuição binomial em cada ocasião de amostragem, um estimador para N é dado por:

$$\hat{N} = \frac{\sum n_i M_i}{\sum m_i}$$

Uma distribuição apropriada ao estudo de captura-recaptura é a hipergeométrica:

$$P(m_i | N, n_i) = \frac{\binom{M_i}{m_i} \binom{N_i - M_i}{n_i - m_i}}{\binom{N}{n_i}}$$

que para o caso de duas amostras possui a seguinte função de verossimilhança:

$$L(n_1, n_2, m_2 | N, p_1, p_2) = \binom{N}{n_1} p_1^{n_1} (1-p_1)^{N-n_1} \binom{N-n_1}{n_2-m_2} \binom{n_1}{m_2} p_2^{m_2} (1-p_2)^{n_1-m_2}$$

Esta expressão pode ser aproximada pela distribuição binomial para n ocasiões de amostragem:

$$L(N, p_1, \dots, p_l) \propto \frac{N!}{(N-r)!} \prod_{i=1}^l p_i^{n_i} (1-p_i)^{N-p_i}$$

Para a situação em que M_j/N é pequena, e m_j pode ser tomado como uma variável Poisson, então \hat{N} é uma estimativa exata de máxima verossimilhança para N .

2.2.3 Modelos gerais para população fechada

Para o estudo dos modelos gerais, foi utilizada a revisão apresentada por Schwarz & Seber (1999), que apresenta as referências dos trabalhos recentes em captura-recaptura.

O modelo de Schnabel (Krebs, 1989) é tido como o modelo que dá origem ao estimador clássico do tamanho populacional em experimento de captura-recaptura envolvendo múltiplas ocasiões de amostragem. O modelo M_t (modelo de Schnabel) é um caso particular do conjunto de oito modelos introduzidos por Otis et al. (1978) e descrito por (Pereira, 1995), para o qual se assume que cada animal na população tem a mesma probabilidade de captura em cada ocasião de amostragem, ou seja, igual ao retorno em cada amostra.

Modelos que levam em consideração a probabilidade de captura variando com a heterogeneidade (h), resposta comportamental à armadilha (b),

variação temporal (t) e todas as possíveis combinações entre estes fatores estão disponíveis para análise de dados de experimento de captura-recaptura, com múltiplas ocasiões de amostragem em populações fechadas. O problema mais crítico é heterogeneidade de captura, que não pode ser modelada pelas oito combinações anteriormente descritas. Neste caso, a proposição é que se tome subpopulações que possam ser mais homogêneas. Modelos de estratificação da população foram apresentados como solução parcial para este problema.

A figura 1 apresenta a relação entre os oito modelos com todas as possíveis relações entre eles.

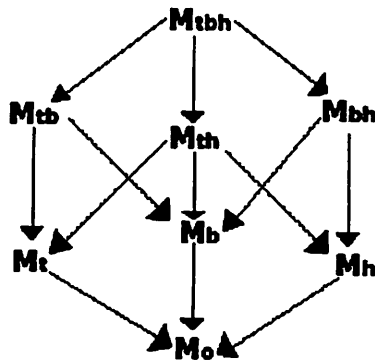


FIGURA 2.2.1 Relação entre os oito modelos (a seta indica que o modelo é um caso especial imediato do outro) extraído de Pereira (1995).

Estes modelos diferem na forma e na quantidade de parâmetros introduzidos para modelar algum aspecto da probabilidade de captura.

Os oito modelos que são obtidos através das considerações citadas são: (M_o , M_{tb} , M_b , M_{bh} , M_t , M_{th} , M_{tbh}), que serão descritos brevemente a seguir:

Modelo M_0

Este é o modelo mais simples, pois assume que as probabilidades de captura são constantes. Desta forma, todos os elementos, marcados e não marcados, tem a mesma probabilidade de serem apanhados em quaisquer das ocasiões amostrais semelhantemente a um experimento com bolas em uma caixa, com reposição dos elementos. A abordagem por meio dos modelos lineares generalizados pode ser aplicada a este caso mais simplificado de captura, considerando somente a constante do modelo de regressão.

Modelo M_1

A probabilidade de captura varia somente com o tempo, ou seja, ao longo das ocasiões de amostragem. O estimador de Schnabel é frequentemente utilizado com este modelo, mas é apenas uma aproximação do estimador de máxima verossimilhança introduzido por Darroch (1958) citado por Pereira (1995).

Modelo M_b :

Este é o modelo mais simples que permite resposta comportamental à armadilha. Uma característica importante deste modelo é que os animais não contribuem com qualquer informação para estimação do tamanho da população após a primeira captura.

Os indivíduos que já foram capturados possuem probabilidade diferente de serem recapturados em relação aos que ainda não foram. O comportamento “trap shy” é observado pela inibição do indivíduo já capturado e o “trap happy” o indivíduo se sente atraído pela armadilha diferentemente dos indivíduos ainda não capturados.

Modelo M_h :

Este modelo permite heterogeneidade de probabilidade de captura entre os animais, mas não permite modelar a variação nas probabilidades de captura no tempo. A modelagem por meio dos modelos log-lineares é semelhante ao modelo M_l . A interpretação de dependência entre as amostras é diretamente justificada por algum comportamento em relação à armadilha (Evans et al., 1994)

Modelo M_{th}

Este modelo permite variação nas probabilidades de captura devido a ambos, tempo e comportamento (dependência de armadilha). Aqui a probabilidade de captura é modelada separadamente da dependência da variação comportamental em relação à armadilha ou ao processo amostral. Cormack (1989) apresenta uma parametrização, através da qual é possível identificar e converter parâmetros de regressão do modelo log-linear para este modelo.

Modelo M_{th} :

Este modelo assume heterogeneidade com variação nas probabilidades de captura no tempo e relação ao comportamento à armadilha. Este comportamento pode ser “trap shy” quando o indivíduo é afugentado pela armadilha, levando à obtenção de superestimativas dos valores populacionais, ou “trap happy”, quando o indivíduo se sente atraído pela armadilha, o que causa uma subestimativa do tamanho populacional. Chao et al.(1989) desenvolveram um estimador para o tamanho da população em que utilizam técnicas de reamostragem.

Modelo M_{bh}

Este modelo assume heterogeneidade e a dependência de armadilha, para descrever o comportamento de uma população. A estimativa para o tamanho populacional pode ser encontrada utilizando os modelos log-lineares.

Modelo M_{tbb}

Este modelo é importante sob o aspecto conceitual, entretanto apresenta problemas de estimatabilidades e identificação de parâmetros. Neste caso, o tamanho da população não é estimável, a não ser que algumas restrições sejam efetuadas, mas apenas algumas combinações dos parâmetros de interesse.

2.3 Abordagem log-linear em população aberta (Cormack – 1989)

O modelo descrito por Cormack (1989) considera a frequência das histórias de captura, assumindo que estas frequências são variáveis com uma distribuição Poisson, e aplica a teoria dos modelos lineares generalizados para encontrar as estimativas de alguns componentes do modelo enquanto outros só podem ser estimados fazendo alguma restrição.

Evans et al (1994) apresentaram uma teoria geral para modelar os dados de captura-recaptura em população fechada. Os autores desenvolveram os modelos M_0 , M_t , M_b e M_{tb} utilizando a abordagem de modelos log-lineares.

2.3.1 O modelo biológico-demográfico

Os modelos log-lineares são muito usados para modelar dependência entre amostra e heterogeneidade de captura. Eles foram introduzidos no estudo de captura-recaptura por Fienberg (1972), e posteriormente desenvolvido por

Cormack (1989), que relata as conversões entre os parâmetros do modelo e os parâmetros biológicos de interesse. Evans et al. (1994) desenvolveram um método geral para analisar tais modelos utilizou as funções dos parâmetros biológicos.

As estimativas para o tamanho populacional, probabilidade de captura, sobrevivência e migração para os períodos amostrais podem ser obtidas a partir do modelo completo e são descritas a seguir:

p_i : é a probabilidade de captura de um indivíduo não marcado na i -ésima amostra;

ϕ_i : é a probabilidade de um indivíduo sobreviver da i -ésima para $(i + 1)$ -ésima amostra;

χ_i : é a probabilidade de que um indivíduo não seja visto após a i -ésima amostra;

$1/\psi_i$ é a probabilidade de que um indivíduo vivo e não marcado na população no momento da $(i + 1)$ -ésima amostra estava na população, no momento da i -ésima amostra. Este parâmetro refere-se à entrada de novos indivíduos, de forma permanente, dentro da área pesquisada.

$1/\lambda_i$, é a probabilidade de que um indivíduo vivo na população no momento da i -ésima amostra, mas não observado em ocasiões subseqüentes, ainda esteja vivo na população no momento da $(i + 1)$ -ésima amostra.

As demais estimativas podem ser obtidas diretamente dos valores estimados para as frequências de captura:

Probabilidades de sobrevivência ϕ_i

Os parâmetros ϕ_i representam a fração de indivíduos que sobreviveram de uma ocasião de amostragem para a seguinte. Estes parâmetros não podem ser

estimados na presença de migrações, mas apenas algumas combinações destas estimativas (a menos que se use alguma restrição nestas). Há, no entanto, delineamentos (e os respectivos modelos) específicos para quando o interesse for estimar parâmetros de sobrevivência.

$$\text{Nascimentos: } B_i = N_i(1 - p_i) \phi \Psi_i;$$

Por nascimento entende-se a quantidade de entrada de novos indivíduos em um período amostral específico. Este valor é diretamente proporcional ao parâmetro de entrada obtido diretamente do modelo de regressão. Nota-se aqui também, que esta quantidade somente é estimável admitindo algumas restrições. É importante observar que entradas e saídas temporárias não podem ser estimadas por este modelo com precisão.

$$\text{Parâmetros de entrada: } \lambda_i = \chi / [\phi(1 - p_{i+1}) \chi_{i+1}];$$

As estimativas λ_i representam as entradas de indivíduos de um período para o outro. Estes parâmetros somente são estimáveis com algumas restrições, já que existem mais parâmetros do que estatísticas observáveis.

As estimativas para o tamanho populacional, probabilidade de captura, sobrevivência e número de entrada de veículos para os períodos amostrais podem ser obtidas a partir do modelo completo. As demais estimativas podem ser obtidas diretamente dos valores estimados para as frequências de captura. Nas seções seguintes será apresentada a conversão dos parâmetros biológicos para os parâmetros de rotatividade da frota estacionada.

2.3.2 Parâmetros do modelo de regressão e seus correspondentes biológicos

Os coeficientes de regressão apresentam alguma combinação de fatores de interesse para estimar o tamanho da frota em estudo. As tabelas a seguir,

mostram os parâmetros correspondentes à dinâmica da frota e seus respectivos modelos a serem estudados em cada caso específico. Estes modelos são apresentados na seção 2.3.5, separadamente, considerando as mais diversas situações amostrais. Na tabela 2.3.1 estão todas as combinações dos fatores que possuam algum interesse para três ocasiões amostrais.

TABELA 2.3.1 Correspondência entre parâmetros de ajuste e parâmetros biológicos para o modelo M_1 de Schnabel por Cormack (1989) com $S = 3$.

Parâmetros	Parâmetros de ajuste	Parâmetros biológicos
Media Geral	u	$\ln(Np_1p_2p_3)$
	$u_1,$	$\ln(1-p_1)/p_1,$
Efeitos Principias	$u_2,$	$\ln(1-p_2)/p_2,$
	u_3	$\ln(1-p_3)/p_3$
Interações	$u_{12},$	0
	u_{13}, u_{23}	0

u : representa a constante do modelo;

u_1, u_2 e u_3 : efeitos principais representando cada uma das ocasiões amostrais;

u_{12} : representa a interação entre os fatores.

Esta interação é muito importante para se obter alguma estimativa que seja interessante para a obtenção de estimativas de parâmetros de rotatividade ou algum problema para detectar o comportamento amostral. Parâmetros de comportamento são detectados pela presença de interação significativa de segunda ordem entre os períodos amostrais próximos

$$g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i}$$

Para este exemplo a matriz modelo de delineamento amostral para os efeitos estudados fica:

$$\begin{bmatrix} \hat{m}_{000} \\ \hat{m}_{100} \\ \hat{m}_{010} \\ \hat{m}_{110} \\ \hat{m}_{001} \\ \hat{m}_{101} \\ \hat{m}_{011} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{bmatrix}$$

A conversão dos parâmetros de regressão e parâmetros do tamanho populacional e da probabilidade de captura é imediata para o caso de populações fechada com a probabilidade de captura variando no tempo (ocasião de amostragem).

TABELA 2.3.2 Três períodos amostrais: Modelo M_{tb} de para população fechada.(extraído de Pereira, 1995)

Parâmetros	Parâmetros do modelo	do Parâmetros de componentes de Rotação
Media peral	U	$\ln(Np_1c_2c_3)$
Efeitos principias	$u_1,$	$\ln[(1-p_1)p_{11}] / p_1c_2,$
	$u_2,$	$\ln[(1-c_2)] / c_2,$
	u_3	$\ln[(1-c_3) / c_3]$
Interações	$u_{12},$	$\ln\{(1-p_2) c_2 p_3 / [p(1-c_2) c_3]\}$
	u_{23}	0
	u_{13}	0

A interação de segunda ordem na tabela 3.2 representa o componente do parâmetro de dependência comportamental em relação ao processo amostral.

Genericamente, os coeficientes de regressão têm correspondência direta com os parâmetros biológicos em estudo geral de populações abertas de acordo com a tabela a seguir:

TABELA 2.3.3 Correspondência entre parâmetros do modelo de regressão e parâmetros de componentes de rotação para um experimento com s amostras considerando a população aberta.

	Parâmetros do modelo	Parâmetros do modelo de componente de rotação
Media Geral	u	$\ln[Np_1\phi_1 p_2\phi_2 \dots p_{s-1}(1 - \chi_{s-1})]$
Efeitos Principais	$u_{1i},$ $u_{ij},$ u_s	$\ln[(1-p_1)\psi_1/p_1],$ $\ln[(1-p_i)/p_i],$ $\ln[\chi_2/(1-\chi_2\phi_2)]$
Interações	$u_{12}, u_{123}, \dots u_{1234} \dots$ $u_{(s-1)s}, u_{(s-2)(s-1)s} \dots$ Demais interações	$\ln(\psi_2), \ln(\psi_3), \ln(\psi_4), \dots$ $\ln(\lambda_{s-2}), \ln(\lambda_{s-3}), \ln(\lambda_{s-4})$ Assumido como zero

Fonte: Cormack (1989)

2.3.3 Correção para modelos de captura

Quando as frequências de algumas histórias de captura são muito baixas, isto pode provocar um viés nas estimativas dos parâmetros de regressão e, conseqüentemente, dos parâmetros demográficos de interesse. Evans & Bonett (1994) apresentaram uma sugestão de correção com base em trabalhos de simulação, para populações fechadas, a qual consiste em adicionar uma constante c a cada célula da tabela de frequências. Se o valor esperado da i -ésima frequência é denotado pela $E(F_a) = \mu_i$, então $\log(\mu_i)$ é estimado por \log de $(F_a + c)$ com a finalidade minimizar o viés. O estudo realizado por Plackett (1981) citado por Evans & Bonett (1994), mostrou que a constante $c = 0,5$, o

que implica que a adição desta constante a cada casela da tabela de contingência de múltiplas entradas reduz o viés do estimador obtido por meio dos modelos log-lineares. Este procedimento é significativamente importante quando a população em estudo é pequena.

2.3.4 Tabelas de contingência

No estudo específico dos modelos de captura-recaptura, Fienberg (1972) apresentou um trabalho de aplicação de tabelas de contingência de dupla entrada para o estudo de uma população fechada. No artigo, ele considera as ocasiões amostrais em classes mutuamente exclusivas com frequência f_a , em que a indica a ordem das classes. As c classes podem ser representadas em tabela de múltiplas entradas semelhantes a um experimento fatorial 2^c e seus efeitos principais e interações poderiam ser estudadas.

Considerando os dados obtidos a partir das histórias de captura em tabela de contingência incompleta das classes mutuamente exclusiva, pode-se estudá-los através de uma estrutura multiplicativa para as médias do modelo Poisson apropriada para estes casos.

As histórias de captura com frequências F_a são consideradas variáveis aleatórias independentes com distribuição de Poisson com médias m_a , sendo esta distribuição apropriada para aplicação da teoria dos modelos lineares generalizados. Tais modelos permitem associar a distribuição da variável resposta de modo que a esperança desta, apresente uma relação funcional com um preditor linear.

2.3.5 Caraterização dos modelos log-lineares Poisson

No estudo específico dos modelos de captura-recaptura, Fienberg (1972) apresentou o primeiro trabalho de aplicação de tabelas de contingência de dupla entrada para o estudo de uma população fechada.

Os modelos lineares generalizados incluem os modelos lineares com erros com distribuição normal, os modelos logísticos, para resposta com distribuição binomial, os modelos log-lineares com distribuição multinomial e os modelos de regressão para dados para respostas com distribuição Poisson caracterizados e apresentados com mais detalhes por McCullagg & Nelder (1989). Muitos outros modelos estatísticos, cuja distribuição pertença à família de distribuições exponencial e que possuam uma função de ligação apropriada, podem ser agrupados nos modelos lineares generalizados. Tais modelos são constituídos pelo seguintes componentes:

- A variável resposta Y_i segue alguma distribuição da família exponencial;
- A função de ligação , $g(\cdot)$, é uma função monótona e diferenciável, definida por:

▪

$$g(\mu)=\eta$$

e que

$$\eta = X_i' \beta$$

é o preditor linear

$$\beta = (\beta_0, \beta_1, \beta_2 \dots \beta_q)$$

$p < n$, é um vetor de parâmetros desconhecidos a serem estimados e

$X^T = (x_{11} \dots x_{1q})^T$ representa os valores das q variáveis explicativas.

Esta metodologia foi aplicada no tratamento do problema de captura-recaptura por Cormack (1989) para descrever componentes da dinâmica demográfica da população em estudo, ligando-os às estimativas dos componentes do modelo de regressão. Será considerada, neste trabalho, a caracterização dada por ele aos modelos biológicos adaptados ao estudo de rotação de veículos em áreas de estacionamento.

2.3.6 Forma canônica

A função distribuição Poisson, assumindo a que F_a tenha esta distribuição é:

$$P(f_a; m_a) = \frac{m_a^{f_a} e^{-m_a}}{f_a!} I_A^{f_a}$$

com $m_a > 0$ e $A = \{0, 1, \dots\}$

possui a forma canônica

$$P(f_a; m_a) = \exp \left[f_a \ln m_a - m_a - \ln f_a! \right]$$

A forma canônica geral é:

$$P(f_a; \theta_a; \phi) = \exp \left\{ \frac{1}{a(\phi)} \left[f_a \theta_a - b(\theta_a) + c(f_a; \phi) \right] \right\}$$

Comparando as duas equações têm-se:

$$\begin{aligned} a(\phi) &= 1 & b(\theta_a) &= m_a = e^{\theta_a} \\ \theta_a &= \ln m_a & c(f_a; \phi) &= -\ln f_a! \end{aligned}$$

Para a especificação do modelo, o vetor de parâmetros θ_a não é de interesse direto, e sim um conjunto menor de parâmetros $\beta_0, \beta_1, \beta_2, \dots, \beta_\varphi$ tais que sejam uma combinação linear destes parâmetros.

2.4 Abordagem bayesiana em população fechada (Smith, 1988)

Smith (1988, 1991) apresentou um procedimento bayesiano para a estimação do tamanho populacional, aplicando o modelo M_t sob o enfoque da teoria da decisão. O autor considerou um grupo de funções de perda, para as quais obteve as estimas de menor risco para o inverso do tamanho populacional. Neste trabalho seguiremos detalhadamente os passos de sua implementação, sugerindo apenas distribuições de probabilidades “a priori” conjugadas, porém informativas. Tal abordagem é também encontrada nos trabalhos apresentados a seguir, os quais são baseados em construção de distribuição a priori conjugados da família Gama ou Beta.

Smith (1988, 1991) descreve a análise bayesiana para modelos de captura em múltiplas ocasiões amostrais (modelo de Schnabel), em que as probabilidades de captura variam entre as ocasiões amostrais. O autor utiliza prioris Beta para a probabilidade de captura e $1/N$ para o tamanho populacional.

Em outro contexto que não o da técnica de captura-recaptura, Salow & Palka (1996) utilizaram uma abordagem bayesiana para estimação do tamanho de grupo de animais, considerando a função de verossimilhança binomial com prioris Beta, e encontraram uma distribuição a posteriori da família Beta. A função de perda quadrática também foi a escolhida na obtenção de estimativas.

King & Brooks (2001) consideram o problema da estimação do tamanho populacional usando a técnica de captura-recaptura por meio de modelos log-lineares em população fechada correspondentes a uma série de listas incompletas, utilizando os dados dispostos em tabelas de contingência.

Fienberg (1999), citado por King & Brooks (2001), adota a família de prioris N^c em que c é uma constante positiva que é semelhante à construção Gama – Poisson – Gama. A estimação a posteriori é abordada pela Teoria da decisão. Os autores utilizaram a função de perda da forma geral

$$P\left[\left(\hat{\theta}_x, \theta\right)\right] = \frac{\left(\hat{\theta}_x - \theta\right)^2}{\theta^x}$$

com $x = (0, 1, 2)$

Brown Jr. (2001) apresenta referências da abordagem bayesiana aplicada ao estudo de populações fechadas, destacando o trabalho de Smith (1988), referência central desta monografia em que é dada ênfase ao modelo M_1 de Schnabel, que inclui somente a variação de probabilidade de captura no tempo. Na modelagem foi aplicado prioris da família da distribuição probabilidades Beta (a, b), prioris para a probabilidade de captura e $1/N$ para o tamanho populacional, refletindo as informações do período amostral anterior.

Smith (1988), utilizando a técnica de captura-recaptura para estimação do tamanho populacional considerando o modelo de Schnabel, aplicou o método de Bayes para a estimação do tamanho populacional. O autor considerou uma aproximação Poisson para a distribuição de probabilidade Hipergeométrica. As prioris com distribuição Gama (a, b) foram construídas de forma a refletir o tamanho populacional. Esta função de distribuição de probabilidade tem a particular propriedade de ser conjugada com a função de verossimilhança (Poisson), chegando-se a uma função densidade “a posteriori” conjunta da família Gama. Foram consideradas as funções de perda chi-quadrado, quadrática, módulo e zero-um (0 para a região de acerto e 1 para a região de erro).

2.4.1 Teorema de Bayes

Uma vez aceito que há “graus de confiança” sobre os parâmetros antes mesmo que se realizem os experimentos, tais eventos podem ser descritos por distribuições de probabilidades. Neste caso, é desejável encontrar métodos algébricos eficientes e flexíveis para incorporar estas informações à expressão

matemática em problemas de estimação. A função do teorema de Bayes é incorporar as noções subjetivas a respeito do parâmetro que se quer estimar, por meio de uma distribuição de probabilidade, à função de verossimilhança dos dados.

Gelman et al. (1995) apresentam ao teorema de Bayes em termos da distribuição de probabilidade da seguinte forma:

Seja a função de probabilidade conjunta para o parâmetro, para a variável aleatória y . Esta função de distribuição de probabilidade conjunta pode ser escrita como o produto da função distribuição ou da função densidade de probabilidade correspondente à informação a priori $p(\theta)$ e a distribuição amostral $p(y|\theta)$

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta) p(y|\theta)}{p(y)}$$

em que $p(y) = \sum p(\theta) p(y|\theta)$ para o caso discreto ou $p(y) = \int p(\theta) p(y|\theta) d\theta$ para o caso contínuo. Como $p(y)$ não depende do valor paramétrico θ para um valor fixo de y , então pode-se considerá-la como uma constante de uniformização da “posteriori” e escrever:

$$p(\theta|y) \propto p(\theta) p(y|\theta).$$

Neste sentido, o Teorema de Bayes é usado para tornar operacional e quantitativo o processo de modificar probabilidade (representando “grau de confiança” em proposições), tendo em vista a incorporação de novas informações (geralmente como resultados de experimentos, considerações teóricas ou mesmo observação casual)

2.4.2 Informação a priori e a função de verossimilhança

Em algumas situações experimentais, a informação a priori pode ser útil na obtenção de estimativas mais confiáveis. Muitas vezes esta informação é resultado de observações casuais, consideração teórica e não provenientes de um conjunto particular de dados. Em qualquer situação, ou mesmo se ambos os tipos de informação são disponíveis, é necessário avaliar a sua importância no sentido de aumentar a precisão da estimativa dos parâmetros. A informação a priori pode ser obtida por meio de dados passados, gerados por experimentos anteriores e disponíveis para uso em análises futuras. Se for possível agrupar estas informações de modo que se possa representá-las por meio de uma distribuição de probabilidade, é possível, então, incorporá-la à função de verossimilhança.

DeGroot (1970) apresenta a construção e aplicação da distribuição a priori, que caracteriza perfeitamente informação anteriormente disponível e que pode ser substancialmente útil em termos de aumentar a qualidade da inferência do parâmetro ou do vetor paramétrico sobre a distribuição a posteriori. O método utilizado por ele é baseado em estatísticas suficientes para os parâmetros estudados. No caso presente, serão usadas prioris conjugadas da família de distribuições Gama.

Borges (2001) descreve o método de obtenção de distribuições a posteriori conjugadas a partir de prioris não informativas para as distribuições da família exponencial. Aplicou a estimação bayesiana a um conjunto de dados com distribuição multinomial, considerando priori uniforme.

Gelman, et al. (1995) apresentam descrição completa da metodologia bayesiana detalhando as vantagens do método ao incorporar informações subjetivas por meio de uma distribuição de probabilidade ponderando a probabilidade de realização da variável aleatória. Os autores apresentam o

teorema de Bayes, que torna operacional o processo para incorporar a informação a priori às evidências da amostra. As conclusões estatísticas sobre um parâmetro pela abordagem são mais realistas. A principal diferença deste método é a possibilidade de quantificar as informações subjetivas que se tem a respeito do parâmetro de interesse.

2.4.3 Teoria da Decisão

As afirmações probabilísticas a respeito do parâmetro (singular ou vetor de parâmetros) que representa o verdadeiro estado da natureza são feitas sob condições de incerteza. É necessário então, tomar uma decisão que, de alguma forma, torne a estimativa obtida o mais próximo possível do verdadeiro valor paramétrico. Os elementos necessários para a escolha de um bom estimador sob o enfoque da Teoria da decisão utilizando a abordagem bayesiana são:

A distribuição de probabilidade dos parâmetros condicional aos dados a chamada distribuição “a posteriori” $p(\theta | y)$;

A função de Perda: $L[\hat{\theta}_x, \theta]$;

A função de Risco: $R(\theta) = E\{L[\hat{\theta}, \theta]\}$

O problema de estimação utilizando a teoria da decisão tem como base a escolha de uma função denominada função de perda. Esta função é escolhida de forma a atender algum objetivo específico. O espaço paramétrico cujos elementos representam o verdadeiro estado da natureza é dado por Θ . Define-se espaço de decisão $D \subset \Theta$ para cada um dos elementos que representam uma decisão que pode ser tomada e a função de perda P , definida sobre $D \times L(d, \theta)$, representando a perda resultante da decisão d quando θ é o verdadeiro do estado da natureza. A função de decisão δ é definida sobre a variável aleatória

dentro de D , e o problema é escolher função δ que minimize a perda de alguma maneira. Para cada função de perda considerada existe uma função de risco a ela associada. O que se busca é um valor $\hat{\theta}$ que torne o risco mínimo.

Segundo Silvey (1995) a Teoria da Decisão foi introduzida por Wald em 1946. Do ponto de vista matemático, um procedimento de decisão amostral está inserido dentro do espaço de decisão das duas alternativas elementares d_0 e d_1 . Estas decisões devem ser confrontadas e a decisão de aceitar uma delas como verdadeira implica em rejeitar a outra. Desta maneira, o problema se inicia com a escolha de uma função de decisão. Em particular no problema de estimação pontual. Para ver isto considere o caso em que se tem um único valor paramétrico real θ .

Bickel & Doksum (1977) apresentam um capítulo dedicado ao estudo da Teoria da Decisão. Em geral, o processo de tomada de decisão, em face de algumas alternativas propostas, é cercado de incertezas. Estas incertezas podem ser reduzidas admitindo uma função de perda que deve ser minimizada. A função de perda leva a uma alternativa de menor risco, sendo a função de risco então definida como a esperança da função de perda considerada.

A função de Perda é útil para analisar, segundo Silvey (1995), as consequências de uma decisão sobre o verdadeiro valor paramétrico. Na prática, a função de perda $L\left[\left(\hat{\theta}_x, \theta\right)\right]$, a qual expressa o afastamento da estimativa quando $\hat{\theta}_x$ é o verdadeiro “estado natural”, está ligada as observações particulares da variável aleatória e à escolha de uma função distribuição a priori.

Para uma determinada classe de problemas é possível definir um conjunto de ações através das quais se pode assumir que todos os elementos deste conjunto podem ser ordenados em preferência pela chamada Função Perda. Uma função de perda típica é a função de perda quadrática:

$$L\left[\left(\hat{\theta}_x, \theta\right)\right] = \left(\hat{\theta}_x - \theta\right)^2.$$

Existem outras formas mais genéricas para a função de perda

$$L\left[\left(\hat{\theta}_x, \theta\right)\right] = \frac{\left(\hat{\theta}_x - \theta\right)^2}{\theta^x}$$

em que x pode assumir valores inteiros (0, 1, 2).

A função de risco está associada à função de perda escolhida, já que esta é a esperança da perda a posteriori. Estes dois conceitos, segundo Mood et al (1974), são usados para assegurar a qualidade de um bom estimador. As informações adicionais introduzidas por meio da distribuição a priori conduzem a uma solução de menor risco. Este critério pode ser utilizado para obter um estimador ótimo para o parâmetro de interesse. Para uma seqüência de estimadores, será preferido o menor risco.

Outras funções de perda podem ser admitidas. Um grupo de uso geral é apresentado a seguir, do qual um exemplo muito usado é a função de perda é a quadrática (que se obtém fazendo $x=0$):

$$L\left[\left(\hat{\theta}_x, \theta\right)\right] = \frac{\left(\hat{\theta}_x - \theta\right)^2}{\theta^x}.$$

A esperança matemática da função de perda a posteriori fica então:

$$E\{L(d, \theta)\} = \int_{\Theta} L(d, \theta) \pi(\theta|x) d\theta$$

Desta forma, encontra-se a função de menor risco a posteriori da forma:

$$R_d = E\{L(d, \theta)\} = \int_{\Theta} L(d, \theta) \pi(\theta) p(\theta|x) d\theta$$

Ainda que as funções estejam determinadas, na prática a perda depende das observações particulares x_i , obtidas na amostra aleatória.

2. 5 REFERÊNCIAS BIBLIOGRÁFICAS

BICKEL, P. J.; DOKSUM, K. A. **Mathematical statistics, basic ideas and selected topics.** Oakland, California, 1977.

BROWN Jr., G. G. **Comparing bayesian, maximum likelihood and classical estimates for the jolly-seber Model.** Carolina: North Carolina State University, 2001. 150 p.

BORGES, C. C. **Um método para obtenção de famílias conjugadas utilizando o teorema de Bayes e Prioris não informativas.** 2001. 55 p. Dissertação (Mestrado em Ciências Exatas) – Universidade Federal de Lavras, Lavras, MG.

CHAO A. Estimating population size for sparse data in capture-recapture experiments. **Biometrics**, Alexandria, v. 45, n. 2, p. 427-438, June 1989.

CHAO, A. An overview of closed capture-recapture models. **Journal of Agricultural, Biological, and Environmental Statistics**, Washington, v. 6, n. 2, p. 158-175, June 2001.

CORMACK, R. M. Log-linear models for capture-recapture. **Biometrics**, Alexandria, v.45, n. 2, p. 395-413, June 1989.

DeGROOT, M. H. **Optimal statistical decisions.** New York: McGraw-Hill, 1970. 489 p.

EVANS M. A.; BONETT D. G.; McDONALD L. L.; A general theory for modeling capture-recapture data from a closed population. **Biometrics**, Alexandria, v. 50, n. 3, p. 396-405, June 1994.

FIENBERG, S. E. The multiple recapture census for closed populations and incomplete 2^k contingency tables. **Biometrika**, London, v. 59, n. 3, p.591, Dec. 1972.

FREEMAN, P. R. Sequential recapture. **Biometrika**, London, v. 60, n. 1, p. 141-153, Apr. 1973.

FREEMAN, P. R. A numerical comparison between sequential tagging and sequential recapture. **Biometrika**, London, v. 60, n. 3, p. 499-508, Dec. 1973.

- GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian data analysis**. New York: Chapman & Hall, 1995.\
- JOLLY, G. M. Mark-recapture models with parameters constant in time. **Biometrics**, Alexandria, v. 38, n. 2, p.301-321, June 1982.
- KING, R.; BROOKS, S. P. On the Bayesian analysis of population size. **Biometrika**, London, v. 88, n. 2, p. 317-336, June 2001.
- KREBS, W. J. C. **Ecological Methodology**. New York: Editora & Row, Publishers, 1989.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3. ed. Tokyo: McGraw-Hill, 1974. 564 p.
- McCULLAGH, P.; NELDER, J. A. **Generalized linear models**. 2. ed. London: Chapman & Hall, 1989. 511 p.
- NICHOLS, J. D. ; HINES J. E.; POLLOCK, K. H.; HINZ, R. L.; LINK, W. A. Estimating breeding proportions and testing hypotheses about costs of reproduction with capture-recapture data. **Ecology**, Washington, v. 75, n.7, p. 2052-2065, Oct. 1995.
- PEREIRA, J. E. **Análise de experimentos de captura-recaptura através de modelos log-lineares**. 1995. 106 p. Dissertação (Mestrado) – Escola Superior de Agricultura “Luiz de queiroz”, Piracicaba, SP.
- SILVEY, S. D. **Statical Inference**. New York: Chapman & Hall, 1995.
- SMITH, P. J. Bayesian Methods for Multiple Capture-Recapture Surveys, **Biometrics**, Alexandria, v. 44, n. 4, p.1177-1189, Dec. 1988.
- SMITH, P. J. Bayesian analyses for a multiple capture-cecapture model, **Biometrika**, London, v. 78, n. 2, p.399-407, June 1991.
- SOLOW R.; A, PALKA, D. On Bayesian Estimation of Group Size, **Biometrics**, Alexandria, v. 52, n. 1, p. 335-340, Mar. 1996.

CAPÍTULO 3

DINÂMICA DA FROTA DE VEÍCULOS QUE ESTACIONA NO CENTRO DE LAVRAS, MG.

RESUMO

OLIVEIRA, José Otaviano de. Dinâmica da frota de veículos que estaciona no centro de Lavras – MG. 2003. 94 p. Dissertação (Mestrado em Agronomia/Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras-MG.*

A técnica de captura-recaptura foi utilizada para estimar diversos aspectos da dinâmica populacional da frota de veículos estacionada no centro de Lavras – MG no período das férias semestrais que inclui a última semana de agosto até a segunda semana de setembro de 2002. Inicialmente foi aplicado o modelo de população aberta de Jolly-Seber. Foram utilizados modelos generalizados Poisson (log-lineares), segundo apresentado por Cormack (1989). Tais modelos requerem registros de cada indivíduo e sua história de captura anterior, compondo grupos mutuamente exclusivos com a mesma história de captura. As placas dos veículos foram utilizadas em sua marcação. Os efeitos principais (representado pelas as ocasiões de captura) e as interações entre os fatores foram transformados em estimativas dos parâmetros de interesse de cada caso. Não foi detectada qualquer fonte de interação significativa entre as recapturas, o que indica uma população fechada geográfica e demograficamente. Estimativas pontuais apresentaram valor próximo de 6.000 veículos, no entanto os intervalos de confiança foram muito amplos. Sugere-se que o emprego de modelos mais parcimoniosos, que desprezem a dinâmica da população, poderiam levar a estimativas mais precisas do tamanho da frota.

* Comitê Orientador: Júlio Sílvio de Sousa Bueno Filho - UFLA (Orientador).

ABSTRACT

OLIVEIRA, José Otaviano de. Dynamics of the convoy of vehicles parking at Lavras – MG city center. 2003. 94 p. Dissertation (Master Program in Agronomy/ Major in Statistics and Agricultural Experimentation) Universidade Federal de Lavras.*

Capture-recapture techniques were used to estimate aspects of the populational dynamics of the convoy of vehicles parking at the Lavras-MG city center during the inter-term holiday period from August last week until September 24, 2002. We started from the open population model of Jolly-Seber. Generalized linear models were used (log-linear), according to Cormack (1989). Such models demand individual records and also its previous capture history. The vehicles' plates were used as marks. The main effects (represented by capture occasions) and the interactions among factors were transformed in estimates of the interest parameters of each case. No source of significative interaction was detected, and this suggests a geographically and demographically closed population. Point estimates values were near 6.000 vehicles, but the confidence regions were very broad. It was suggested that parsimonious models that do not allow population dynamic estimates should be considered to get more accurate estimates of the number of the convoy.

* Guidance Committee: Júlio Sílvio de Sousa Bueno Filho-UFLA (Major Professor).

3.1 INTRODUÇÃO

Os modelos de captura-recaptura têm sido aplicados para estimar parâmetros de dinâmica populacional (sobrevivência, recrutamento, emigração e imigração) em populações selvagens. Estimativas podem ser obtidas a partir de informações detalhadas que vêm do uso de identificações individuais e da produção de histórias de capturas para cada indivíduo, em uma seqüência de amostras ou de períodos amostrais.

Jolly e Seber em 1965 apresentaram o modelo para populações abertas em que consideraram adições e deleções na população em estudo. Através do modelo J-S, é possível determinar outros parâmetros de interesse biológico através dos quais pode-se estudar todo o comportamento da dinâmica populacional.

Neste capítulo aplicou-se a abordagem dada por Cormack (1989), que faz uso dos modelos lineares generalizados para estimar os componentes demográficos de interesse. Por esta metodologia, cada indivíduo é marcado com marcas únicas o que permite identificá-lo e, assim, obter as classes mutuamente exclusivas. Estas classes são estudadas a partir de uma tabela de contingência, apurando a frequência de cada uma delas. Cada uma dessas frequências é considerada uma variável com distribuição assumida Poisson. Esta distribuição é da família de distribuições exponencial, o que permite a aplicação da teoria dos modelos lineares generalizados.

O objetivo do presente estudo é detectar parâmetros de rotatividade na frota de veículos estacionada no centro de Lavras, MG, num período atípico em que se suspeitava que fosse possível detectar tais fenômenos (entrada e saída de veículos).

3.2 MATERIAL E MÉTODOS

3.2.1 – Delimitação da área de estudo

Neste experimento foi considerada uma área de referência da cidade de Lavras – MG, que se definiu como “Área Central”, delimitada pela praça Doutor Augusto Silva, a rua Santana até a praça Dona Josephina e a área de estacionamento do Hospital Vaz Monteiro. Foram tomadas amostras de veículos estacionados em que a marca considerada era o número da placa do veículo.

Essencialmente, o procedimento de amostragem baseado em história de captura consiste em identificar cada elemento da amostra através de uma marca única, neste caso a placa do veículo. A população considerada neste experimento consistiu de todos automóveis de categoria particular estacionados na “Área Central”. Uma observação de interesse para pesquisa é que em nenhuma amostra toda a área foi percorrida.

A escolha desta área deve-se ao maior fluxo de veículos e por ser a maior área de estacionamento da cidade. Foi desenvolvida uma rotina no software R-1.6.1 (Iraka & Gentleman, 1996) para a contagem dos grupos mutuamente exclusivos para compor as histórias de captura. Os objetivos foram identificar componentes da dinâmica populacional da frota de veículos entre dois períodos distintos: as férias semestrais de 2002 e o início das aulas, e verificar a influência da frota de veículos dos estudantes no crescimento ou decréscimo da população estacionada.

As amostras foram coletadas de acordo com o seguinte esquema:

1) Primeiro período amostral:

26/08/2002 – 14:00 h às 16:30 h – segunda-feira.

28/08/2002 – 14:30 h às 16:05 h – quarta-feira.

30/08/2002 – 8:30 h às 9:30 h – sexta-feira.

Este período amostral coincide com a última semana das aulas da UFLA e início do período de férias. O objetivo era inicialmente detectar o movimento “migratório” dos veículos no período.

2) segundo período amostral

02/09/2002 – 10:30 h às 11:50 h – segunda-feira.

04/09/2002 – 14:30 h às 16:05 h – quarta-feira.

06/09/2002 – 16:00 h às 17:00 h – sexta-feira.

O segundo período foi coletado durante as férias escolares e primeira semana de retorno;

3) Terceiro período amostral:

16/09/2002 – 13:30 h às 14:45 h – segunda-feira.

18/09/2002 – 11:50 h às 16:05 h – quarta-feira.

20/09/2002 – 16:00 h às 17:00 h – sexta-feira.

O terceiro período coincide com a semana de estabilização de retorno dos alunos. O objetivo novamente era verificar a entrada ou recrutamento significativo de veículo na área.

4) Quarto período amostral:

24/09/2002 – 8:30 h às 10:45 h – terça-feira.

O quarto período amostral é composto de uma amostragem única, tomada separadamente por ser um dia chuvoso, um fator atípico, o que poderia provocar alguma variação as estimativas.

3.2.2 Matriz do modelo

Seja X a matriz de variáveis indicadoras que representa a presença (0) e (1) ausência do indivíduo na amostra i :

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Por exemplo, no presente estudo, têm-se quatro ocasiões de amostragem. O primeiro vetor coluna de X representa a constante do modelo de regressão proposto. Os demais vetores colunas representam as ocasiões de captura. As linhas de X representam as histórias de captura do indivíduo durante as ocasiões amostrais. Os demais parâmetros de interesse biológicos são obtidos a partir de operações nas colunas da matriz. Tem-se $a = (2^5 - 1)$ observações; qualquer modelo retirado daí pode conter no máximo $(2^5 - 1)$ parâmetros. É interessante escolher aqueles que tenham algum interesse do ponto de vista da dinâmica da população. Cormack (1989, 1993) apresentou um conjunto de parâmetros de interesse possíveis de serem obtidos a partir do modelo de regressão por conversão (back transformation) ou interpretação direta dos parâmetros de regressão.

Para estudar os componentes de rotação dos carros será apresentado, a seguir, o modelo mais completo, o qual possui todos os possíveis parâmetros de interesse. A primeira coluna da matriz representa o comportamento da constante do modelo. As quatro seguintes representam os efeitos principais referentes aos períodos de amostragem. As duas seguintes se referem às interações de segunda e terceira ordem, em ordem crescente, definidas como o produto direto da

primeira pela segunda (ψ_2) e da primeira pelas segunda e terceiras colunas da matriz X do delineamento amostral (ψ_3).

O produto do último pela penúltima (λ_2) e da última pela penúltima e antepenúltima (λ_1) representa as interações de saída. A última coluna, representa a interação que é relativa puramente à detecção de algum comportamento em relação ao processo amostral ou armadilha. Isto é facilmente observado a partir da matriz do modelo:

$$\begin{bmatrix}
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\
 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0
 \end{bmatrix}
 \times
 \begin{bmatrix}
 \hat{\beta}_0 \\
 \hat{\beta}_1 \\
 \hat{\beta}_2 \\
 \hat{\beta}_3 \\
 \hat{\beta}_4 \\
 \hat{\beta}_5 \\
 \hat{\beta}_6 \\
 \hat{\beta}_7 \\
 \hat{\beta}_8 \\
 \hat{\beta}_9
 \end{bmatrix}
 =
 \begin{bmatrix}
 \ln \hat{m}_{0000} \\
 \ln \hat{m}_{1000} \\
 \ln \hat{m}_{0100} \\
 \ln \hat{m}_{1100} \\
 \ln \hat{m}_{0010} \\
 \ln \hat{m}_{1010} \\
 \ln \hat{m}_{0110} \\
 \ln \hat{m}_{1110} \\
 \ln \hat{m}_{0001} \\
 \ln \hat{m}_{1001} \\
 \ln \hat{m}_{0101} \\
 \ln \hat{m}_{1101} \\
 \ln \hat{m}_{0011} \\
 \ln \hat{m}_{1011} \\
 \ln \hat{m}_{0111}
 \end{bmatrix}$$

Os vetores de parâmetros $\hat{\beta}_{is}$ são obtidos iterativamente por algum procedimento matemático de aproximação numérica.

Alguns componentes de rotação podem ser retirados diretamente dos parâmetros do modelo de regressão. Os demais podem ser obtidos por combinações lineares dos parâmetros do modelo e convertidos para valores estimados dos parâmetros de interesse. Alguns deles não podem ser estimados, a não ser que alguma restrição seja efetuada.

As frequências das classes dos grupos mutuamente exclusivos estimadas contem tais combinações em escala logarítmica. Os parâmetros estimados (β_5 e β_6) representam as interações de entrada e os (β_7 e β_8), as interações de saída. Assim, a solução para a situação mais geral possível pode ser retirada diretamente da equação de regressão ou dos valores ajustados.

Os modelos estudados neste exemplo foram:

a) População aberta com modelo composto de entrada e saída e dependência – M1:

$$g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \\ + \beta_5 x_{1i} x_{2i} + \beta_6 x_{1i} x_{2i} x_{3i} + \beta_7 x_{4i} x_{3i} + \beta_8 x_{2i} x_{3i}$$

Este é o modelo mais completo do ponto de vista da parametrização proposta por Cormack (1989). Todas combinações dos parâmetros de interesse são contempladas.

b) População aberta (somente entrada) – M2:

$$g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{1i} x_{2i} + \beta_6 x_{1i} x_{2i} x_{3i}$$

Neste modelo discute-se o efeito de entrada de novos veículos na Área central de estacionamento. Tais efeitos são estimados pela interação em ordem crescente entre os períodos amostrais. Estes efeitos só têm significado se apresentarem estimativas positivas, caso contrário serão descartados do modelo e em seguida obtém-se um novo ajuste.

c) População aberta (somente saída) – M3:

$$g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{4i} x_{3i} + \beta_6 x_{4i} x_{3i} x_{2i}$$

Da mesma forma como nos modelos anteriores, atua-se trocando apenas a ordem das interações, que são tomadas da última para a primeira, como apresentado no modelo acima. São sete parâmetros para descrever a dinâmica da rotação, para quinze observações de classes mutuamente exclusivas. O número de parâmetro do modelo é relativamente alto se comparado com um modelo de regressão tradicional, mas neste caso o objetivo é encontrar possíveis interações de ordem específica que levem a estimação de parâmetros interessantes a estudar a rotação dos veículos dentro da área de estacionamento.

d) População aberta com modelo composto de entrada e saída – M4:

$$g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{1i} x_{2i} + \beta_6 x_{1i} x_{2i} x_{3i} + \beta_7 x_{4i} x_{3i} + \beta_8 x_{4i} x_{3i} x_{2i}$$

Com este modelo é possível estimar os parâmetros de entrada e saída. Estes parâmetros, quando significativos, apontam para uma população aberta em termos de rotação dos seus indivíduos de forma permanente. Se nenhum deles for significativo em nenhuma ordem, pode-se concluir que a população é fechada.

Se for encontrada alguma estimativa fora do espaço paramétrico, todos os demais parâmetros devem ser retirados do modelo e, em seguida, efetuado um novo ajuste (Cormack, 1989).

e) População fechada – M6:

$$g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i}$$

Este modelo é composto somente pelo os efeitos principais, como no exemplo descrito na seção anterior. O coeficiente da média geral representa a ligação entre tamanho populacional e as probabilidades de captura em cada período amostral.

f) População fechada com esforço constante nas amostras centrais – M7:

$$g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 (x_{2i} + x_{3i}) + \beta_3 x_{4i}$$

Por meio deste modelo, pode-se determinar se os períodos amostrais apresentam um esforço uniforme de amostragem ou não. No caso abordado, considerou-se a possibilidade de que as duas amostras centrais possam ser estudadas como se fossem uma amostra única. Isto implica em ganho de graus de liberdade pelo modelo. Sobretudo em experimento de longa duração, este pode ser um método para o agrupamento de amostras dentro de grupos ou períodos segundo algum critério pré-estabelecido.

g) População fechada com esforço constante nas três últimas amostras – M8:

$$g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 (x_{2i} + x_{3i} + x_{4i})$$

Todos a modelos a seguir são idênticos ao modelo anterior. O que varia é a forma de agrupar os períodos amostrais, a fim de detectar a possibilidade de melhorar o ajuste.

h) População fechada com esforço constante nas três primeiras amostras – M9:

$$g(\mu_i) = \beta_0 + \beta_1 (x_{1i} + x_{2i} + x_{3i}) + \beta_2 x_{4i}$$

i) População fechada com esforço totalmente constante – M10:

$$g(\mu_i) = \beta_0 + \beta_1 (x_{1i} + x_{2i} + x_{3i} + x_{4i})$$

No caso do modelo M10, a equação de regressão é composta apenas pelo intercepto e o coeficiente de regressão, sendo este modelo considerado de esforço inteiramente constante. O modelo considerando apenas o intercepto pode ser utilizado para representar populacional. Esta representação equivale ao modelo descrito como M_0 na literatura como o modelo mais simples dentre todos.

3.2.3 Solução das equações

Soluções para alguns parâmetros para o modelo de rotação somente são obtidas utilizando restrições para alguns deles. Para estimar o parâmetro tamanho populacional considerando o modelo completo com quatro ocasiões de amostragem tem-se o grupo de equações apresentadas no modelo genérico.

• Se for assumida a restrição de que saídas e entradas são constantes $\lambda_i = \chi_i / [\phi_i (1 - p_{i+1}) \chi_{i+1}]$; com $\chi_i = \chi_{i+1}$ tem-se então a estimativa da sobrevivência, que para o caso, significa $\lambda_i = 1 / [\phi_i (1 - p_{i+1})]$ ou $\phi_i = 1 / [\lambda_i (1 - p_{i+1})]$ todos os elementos do segundo membro são diretamente estimáveis pela equação de regressão. As estimativas dos λ_i são obtidas diretamente dos parâmetros do modelo de regressão e as estimativas de dos p_i são obtidas a partir dos valores ajustados

$$\frac{m_{1000}}{m_{0000}} = \frac{(1 - p_1) \Psi_1}{p_1};$$

$$\frac{m_{0100}}{m_{0000}} = \frac{(1 - p_2)}{p_2};$$

$$\frac{m_{0010}}{m_{0000}} = \frac{(1 - p_3)}{p_3};$$

$$\frac{m_{0001}}{m_{0000}} = \frac{\chi_3}{1 - \chi_3};$$

Na primeira equação acima, verifica-se que Ψ_1 não é isoladamente estimável. Pode-se então assumir que não houve nenhuma entrada naquele período e que a probabilidade, definida na seção anterior: $1/\Psi_1 = 1$, torna p_1 estimável.

TABELA 3.2.1 - Parâmetros dos modelos com quatro ocasiões de captura e sua correspondência com os fatores correspondentes à dinâmica de rotação.

Parâmetros do Modelo de rotação	Combinações lineares dos parâmetros de regressão
$Np_1 \phi_1 p_2 \phi_2 p_3 (1-\chi_3)$	β_0
$N(1-p_1) \phi_1 \Psi_1 p_2 \phi_2 p_3 (1-\chi_3)$	$\beta_0 + \beta_1$
$Np_1 \phi_1 (1-p_2) \phi_2 p_3 (1-\chi_3)$	$\beta_0 + \beta_2$
$N(1-p_1) \phi_1 (1-p_2) \phi_2 p_3 \Psi_1 \Psi_2 (1-\chi_3)$	$\beta_0 + \beta_1 + \beta_2$
$Np_1 \phi_1 p_2 \phi_2 (1-p_3) (1-\chi_3)$	$\beta_0 + \beta_3$
$N(1-p_1) \phi_1 \Psi_1 p_2 \phi_2 (1-p_3) (1-\chi_3)$	$\beta_0 + \beta_1 + \beta_3$
$Np_1 \phi_1 (1-p_2) \phi_2 (1-p_3) (1-\chi_3)$	$\beta_0 + \beta_2 + \beta_3 + \beta_9$
$N(1-p_1) \phi_1 (1-p_2) \phi_2 (1-p_3) \Psi_1 \Psi_2 \Psi_3 (1-\chi_3)$	$\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_6 + \beta_9$
$Np_1 \phi_1 p_2 \phi_2 p_3 \chi_3$	$\beta_0 + \beta_4$
$N(1-p_1) \phi_1 \Psi_1 p_2 \phi_2 p_3 \chi_3$	$\beta_0 + \beta_1 + \beta_4$
$Np_1 \phi_1 p_2 \phi_2 p_3 \chi_3$	$\beta_0 + \beta_2 + \beta_4$
$N(1-p_1) \phi_1 \Psi_1 \Psi_2 (1-p_2) \phi_2 p_3 \chi_3$	$\beta_0 + \beta_1 + \beta_2 + \beta_4 + \beta_5$
$Np_1 \phi_1 \lambda_1 p_2 p_3 \chi_3$	$\beta_0 + \beta_3 + \beta_4 + \beta_7$
$N(1-p_1) \phi_1 \Psi_1 \lambda_1 p_2 \phi_2 p_3 \chi_3$	$\beta_0 + \beta_1 + \beta_3 + \beta_4 + \beta_7$
$Np_1 \phi_1 \lambda_1 \lambda_2 \phi_2 \lambda_1 \lambda_2 (1-p_2) (1-p_3) \chi_3$	$\beta_0 + \beta_2 + \beta_3 + \beta_4 + \beta_7 + \beta_8$

Para encontrar as estimativas dos demais parâmetros, pode-se obtê-los diretamente das operações sobre os parâmetros do modelo de regressão. Os parâmetros de entrada são representados por β_6 e β_7 . Então basta tomá-los diretamente do modelo de regressão. O mesmo deverá acontecer com as interações de entrada.

3.2.4 Algoritmo de ajuste do modelo linear generalizado a dados com distribuição Poisson

Suponha que n observações com distribuição Poisson independentes f_1, f_2, \dots, f_a são tais que a i -ésima observação, $i = 1, 2, \dots, a$, tem distribuição Poisson com parâmetros m_i . Também suponha que o valor transformado da probabilidade de resposta para a i -ésima observação está relacionado a uma combinação linear de q variáveis explanatórias, isto é,

$$g(m_a) = \beta_1 x_{1i} + \dots + \beta_q x_{qi},$$

em que $g(m_a)$ pode ser a transformação logarítmica de m_i . O componente linear deste modelo será denotado por η_a , também que $g(m_i) = \eta_i$. A função g é a função de ligação de um modelo linear generalizado (Nelder & Wedderburn, 1972).

Para ajustar este modelo usando o método iterativo da equação usando a definição dos modelos lineares generalizados são necessárias expressões para $u(\beta)$ e $I(\beta)$. Uma forma alternativa a este procedimento é resolver estas equações a partir das expressões $\hat{\beta}_{r+1} = \hat{\beta}_r - H^{-1}(\hat{\beta}_r)u(\hat{\beta}_r)$.

O logaritmo da função de verossimilhança para n observações com distribuição Poisson é dado por

$$\ln L(\beta) = \sum_{a=1}^n -m_a + f_a \ln m_a - \ln f_a!,$$

com um valor da estimativa original para os valores da usando a quadrados mínimos ponderados da forma a seguir: Seja $\hat{\eta}_0$ correspondendo à estimativa derivada da função ligadora $\eta = g(m_a)$, define-se o ajuste da variável resposta como:

$$z_{i0} = \hat{\eta}_{i0} + (y_i - \mu_{i0}) \frac{d\eta}{d\mu} \Big|_{\mu_i = \mu_{i0}}$$

define-se a ponderação w_{i0} por:

$$\frac{1}{w_{i0}} = \left(\frac{d\eta}{d\mu} \Big|_{\mu_i = \mu_{i0}} \right)^2 Vf(\mu_{i0})$$

Para obter o ajuste do modelo de regressão por meio dos quadrados mínimos ponderados, faz-se

$$z_{i0} = \sum_{j=1}^p \beta_{,j1} x_{ji}$$

para obter os parâmetros β_1 . Usa-se este novo parâmetro para obter uma nova estimativa para o preditor linear $\hat{\eta}_1$ e assim, sucessivamente, até convergir.

Também, como

$$\eta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_q x_{qi}$$

tem-se, então

$$\frac{\partial \eta_i}{\partial \beta_j} = x_{ji}$$

Se X é a matriz de valores ($n \times q$) das q variáveis explanatórias, W é a matriz diagonal ($n \times n$) cuja i -ésimo elemento diagonal é w_i , em que:

$$w_{i0} = \frac{1}{\left(\frac{d\eta}{d\mu} \Big|_{\mu_i = \mu_{i0}} \right)^2 Vf(\mu_{i0})}$$

z^* é o vetor ($n \times 1$) cujo i -ésimo componente é z^*_i .

O vetor de escores eficientes, $u(\beta)$, pode ser escrito como $u(\beta) = X'W y^*$.

Desta forma, a matriz de informação de Fisher é dada por $X'WX$ cuja inversa é a matriz de covariâncias assintóticas.

Substituindo para $u(\hat{\beta})$ e $I(\hat{\beta})$, encontra-se que a estimativa de β para a $(r + 1)$ -ésima iteração é

$$\hat{\beta}_{r+1} = \hat{\beta}_r + (X'W_r X)^{-1} X'W_r y_r^*$$

em que W_r é a matriz diagonal dos valores de w_i para o r -ésimo ciclo e y_r^* , é o valor de valores de y_i^* para o r -ésimo ciclo.

Então,

$$\hat{\beta}_{r+1} = (X'W_r X)^{-1} [X'W_r (X\hat{\beta}_r + y_r^*)]$$

$$\hat{\beta}_{r+1} = (X'W_r X)^{-1} X'W_r z_r$$

Para um determinado valor inicial β_r atribuído, estima-se β_{r+1} no passo seguinte. O processo termina quando o módulo da diferença entre β_{r+1} e β_r convergir para zero.

O procedimento de ajuste foi efetuado usando o software livre R, (Iraka & Gentleman, 1996) através do procedimento GLM, que além do ajuste, fornece a composição da análise de deviance e dos resíduos.

3.2.5 Critério para análise e seleção de modelos

O critério utilizado para verificação de ajuste e seleção dos modelos é baseado na comparação das deviances do modelo em estudo com relação ao modelo maximal ou saturado em que cada observação é representada por um

parâmetro. Os modelos hierarquizados podem ser comparados entre si a fim de verificar se eles são equivalentes ou não.

Para testar a adequação do modelo linear generalizado aplica-se na prática, sem muito rigor para pequenas amostras, a diferença das “Scaled Deviance”, que para a distribuição Poisson equivale a “Deviance” dos dados, e confronta-se este resultado com o valor crítico de uma distribuição de qui-quadrado com $(n-p)$ graus de liberdade a um nível de significância desejado. O modelo será rejeitado se o valor da deviance for superior ao valor crítico da distribuição de qui-quadrado; caso contrário, o modelo é aceito. Para a distribuição Poisson, usada neste trabalho $S_p = D_p$ já seu parâmetro de dispersão é igual 1 (Demetrio, 1993)

Análise de Deviance

Se existir uma seqüência de modelos os quais são originados de um modelo maior (modelos hierárquicos), Cordeiro & Paula (1986) afirmam que estes podem ser testados contra o modelo com maior número de parâmetros. Desta forma, a diferença entre as deviances dos dois representam um valor que tem distribuição aproximado de uma distribuição χ^2 .

Sejam dois modelos M_1 e M_2 , em que M_1 possui p parâmetros e M_2 , q parâmetros e $p > q$ então $D_p - D_q \cap \chi^2$ (assintótica). $\chi^2_{cal} = D_p - D_q$ e χ^2 tabelados $\chi^2(p - q)$ graus de liberdade a um nível de significância estabelecido. Pode-se testar a hipótese de igualdade dos modelos e verificar qual deles é mais bem ajustado. Rejeita-se a hipótese de igualdade quando o valor calculado pela deviance é maior que valor crítico da distribuição χ^2 .

Análise de Resíduo

Geralmente, após um estudo em que se aplica o modelo de regressão, torna-se necessário utilizar a análise dos resíduos. Os resíduos indicados são:

a) Resíduos ordinários

$$r = y - \hat{u},$$

b) Resíduos de Poisson

$$r_p = \frac{y - \hat{u}}{\sqrt{V(\hat{u})}},$$

em que y representa as observações

\hat{u} : valor ajustado;

$V(\hat{u})$: variância;

c) os resíduos de Pearson estandardizados internamente

$$r_p = \frac{y - \hat{u}}{\sqrt{\hat{\phi} V(\hat{u})(1 - h_{ii})}},$$

em que h_{ii} representa um elemento da diagonal da matriz projetor ortogonal.

3.3 RESULTADOS E DISCUSSÃO

3.3.1 Frequência das histórias de captura

Observou-se que dos quinze padrões de história de captura, os que apresentaram as maiores frequências foram as f_8 , f_{12} , f_{14} e f_{15} com 142, 153, 107 e 170 veículos, respectivamente. Estas histórias de capturas coincidem exatamente com os períodos de captura, o que seria esperado para o caso de os modelos melhor ajustados não envolverem interações. Existem também algumas caselas com zeros, o que sugere a correção de Evans & Bonnet (1994), que propõe somar $(0,5)^{(s-1)}$ a cada casela. A Tabela 3.3.1 registra os padrões das quinze classes mutuamente exclusivas, obtidas das quatro ocasiões amostrais.

TABELA 3.3.1 Histórias de captura com os valores observados para a população de veículos na ordem padrão do fatorial 2^s

Ocasião 1	Ocasião 2	Ocasião 3	Ocasião 4	Frequência
0	0	0	0	0
1	0	0	0	1
0	1	0	0	0
1	1	0	0	4
0	0	1	0	0
1	0	1	0	6
0	1	1	0	3
1	1	1	0	142
0	0	0	1	0
1	0	0	1	1
0	1	0	1	5
1	1	0	1	153
0	0	1	1	3
1	0	1	1	107
0	1	1	1	170

Verifica-se pela tabela que as frequências maiores são concentradas nas ocasiões amostrais. São raras as recapturas para medir as interações de terceira ordem. Isto pode conduzir a problemas na estimação dos componentes de rotatividade de interesse. Na tabela 3.3.2 são apresentadas as correções sugeridas por Evans et al. (1994).

TABELA 3.3.2 Histórias de captura com os valores corrigidos para a população de veículos.

Hist. captura		Valores corrigidos								
f_1	a f_8	0,125	1,13	0,13	4,13	0,13	6,13	3,13	142,13	
f_9	a f_{16}	0,125	1,13	5,13	153,13	3,13	107,13	170,13		*

* Classe não observável.

Não se verificou nenhuma alteração nas estimativas com a correção proposta. Optou-se então, pelo o uso dos dados na forma original.

3.3.2 Estimativas dos parâmetros dos modelos

Modelo completo: entrada, saída e comportamento em relação à armadilha

Nas Tabelas 3.3.3 a 3.3.7 são agrupados os resultados do ajuste dos modelos considerados, a partir do modelo mais completo, descrito abaixo:

$$E(m_a) = u + o_1 + o_2 + o_3 + o_4 + e_1 + e_2 + s_1 + s_2 + c_1$$

Neste modelo, os parâmetros de regressão são representados por:

o_i : ocasiões amostrais;

e_j : interações de entrada de novos carros na frota estacionada (“nascimentos ou imigrações”);

s_k : correspondentes à saída dos veículos (“mortes ou emigrações”);

c_i : correspondente ao comportamento amostral dos elementos com a mesma história de captura (efeito de armadilha).

São dez os parâmetros para descrever o comportamento de uma amostra de quinze dados representando as classes mutuamente exclusivas.

Na Tabela 3.3.3 são apresentadas as estimativas para os parâmetros de regressão. Do ponto de vista da captura-recaptura, isto é importante para detectar possíveis interações de interesse para a estimação de componentes da renovação (rotação) da frota estacionada.

TABELA 3.3.3 Estimativas de máxima verossimilhança dos parâmetros de regressão para um modelo de população aberta com entrada e saída de veículos, além de dependência da “armadilha”.

<i>Coefficiente</i>	<i>Estimativa</i> (β)	$X\beta$	<i>Erro padrão</i>	z	$Pr(> z)$
(Intercepto)	-6,4694	0,0015	1,0089	-6,4120	0,0000**
o_1	3,6463	38,3326	0,5848	6,2350	0,0000**
o_2	4,5884	98,3370	1,0119	4,5340	0,0000**
o_3	4,5890	98,3960	0,9272	4,9500	0,0000**
o_4	3,4595	31,8011	0,4542	7,6170	0,0000**
e_1	-0,1995	0,8191	0,7405	-0,2690	0,7880
e_2	0,4104	1,5074	0,7394	0,5550	0,5790
e_1	-0,5507	0,5765	0,6181	-0,8910	0,3730
e_2	1,1285	3,0910	0,7176	1,5730	0,1160
c_1	-1,6094	0,2000	1,0950	-1,4700	0,1420

** Significativo ao nível de 1% de probabilidade.

Os parâmetros de entrada e saída de veículos não foram significativos no período, o que sugere que se estudem modelos mais simples (o que se fará nas próximas seções). Outros valores de estimativas são pouco importantes. Como

exemplo, da terceira para quarta ocasião, o valor da taxa de entrada é de igual a 1,51 veículos, o que é insignificante diante do tamanho populacional estimado.

Para as interações de saída, verificou-se que apenas s_2 está no espaço paramétrico, porém seu efeito não é significativo. Nenhum efeito comportamental em relação à armadilha foi verificado. Observam-se estimativas negativas para os parâmetros de entrada e saída que não possuem significado plausível, indicando má especificação do modelo.

O modelo completo é o mais geral do ponto de vista da parametrização “biológica” que considera parâmetros de entrada (e_i) ou saída permanentes dos indivíduos presentes na população no momento em que se inicia o estudo. Apresenta, ainda, parâmetros que incluem o comportamento do indivíduo em relação à armadilha ou a forma como ele é capturado. Novamente nestes modelos, as estimativas de interação não apresentaram significância pela estatística z . A presença de muitos parâmetros faz, no entanto, que seja pouco preciso para o estudo do tamanho populacional total, para o que devemos procurar modelos mais simples.

Os valores preditos e intervalos de confiança para cada história de captura, no modelo de população aberta com todos os parâmetros (entrada, saída e dependência da armadilha), são apresentados na Tabela 3.3.4. Estes valores apenas confirmam as considerações anteriores, sendo que a última linha da tabela reflete a previsão do tamanho da população não observada (8047, [2539;25497]).

Modelos de interesse podem ser retirados do modelo completo por redução do modelo mais parametrizado. Observa-se que a frota foi estimada em $8047 + 595 = 8642$ veículos (última casela mais número total de carros observados), com um intervalo de confiança bastante amplo: [3134; 26092].

TABELA 3.3.4 - Valores preditos e intervalos de confiança para cada história de captura, no modelo de população aberta, com todos os parâmetros (entrada, saída e dependência da armadilha).

f_a	Valores preditos	$X\beta$	Erro padrão	L I	L S
0,000	0,0016	-6,4694	1,0093	0,0002	0,0112
1,000	0,0594	-2,8230	0,8331	0,0116	0,3042
0,000	0,1524	-1,8810	0,6276	0,0446	0,5216
4,000	4,7866	1,5658	0,4475	1,9911	11,5068
0,000	0,1525	-1,8803	0,7011	0,0386	0,6028
6,000	5,8475	1,7660	0,4085	2,6256	13,0226
3,000	3,0000	1,0986	0,5774	0,9676	9,3017
142,000	142,0000	4,9558	0,0839	120,4639	167,3862
0,000	0,0493	-3,0099	0,9083	0,0083	0,2924
1,000	1,8897	0,6364	0,7074	0,4723	7,5603
5,000	4,8476	1,5785	0,4474	2,0168	11,6513
153,000	152,2134	5,0253	0,0810	129,8688	178,4026
3,000	2,7966	1,0284	0,5779	0,9010	8,6802
107,000	107,2034	4,6747	0,0965	88,7296	129,5234
170,000	170,0000	5,1358	0,0767	146,2732	197,5755
-	8047,0000	8,9930	0,5884	2539,0000	25497,0000

Utilizando as equações e as restrições propostas no item 3.7 da seção anterior têm-se as estimativas dos demais componentes da dinâmica rotacional dos veículos. Tais valores estão relacionados na Tabela 3.3.5 abaixo.

Verifica-se que a primeira estimativa de sobrevivência apresenta-se fora do espaço paramétrico. Ambas não têm interpretação realista porque as estimativas que geraram não diferem estatisticamente de zero. Desta forma ambas podem tomadas com o valor unitário o que conduz novamente a um modelo mais simples para o tamanho populacional.

TABELA 3.3.5 - Estimativas das probabilidades de captura nas quatro ocasiões amostrais e respectivas probabilidades de sobrevivência

Probabilidade de captura	
p_1	0,025424
p_2	0,010067
p_3	0,010061
χ_2	0,030487

Análise de Deviance

Os modelos encaixados de população fechada não são equivalentes pelo teste de qui-quadrado. Portanto o M6 (M_1 de Schnabel) foi considerado o mais representativo para o conjunto de dados. Assim, não é possível aplicar modelos com quaisquer combinações de esforço constante de amostragem neste caso.

TABELA 3.3.6 - Análise de deviance para os modelos de população fechada

Modelo	GL	Deviance	dif. Deviance	dif. GL	$P(\chi^2_v > Dev)$
M1	5	5,1477	-	-	-
M2	8	8,2075	3,0638	3	0,3818 ^{ns}
M3	8	7,9636	2,8159	3	0,4209 ^{ns}
M3	6	7,4975	2,3496	1	0,1253 ^{ns}
M5	9	7,9947	2,847	4	0,5837 ^{ns}
M6	10	8,8937	3,7465	5	0,5865 ^{ns}

ns: não significativo

TABELA 3.3.7 - Análise de deviance para os modelos de população aberta

Modelo	GL	Deviance	dif. Dev.	dif. GL	$P(\chi^2_v > Dev)$
M6	10	8,8937	-	-	-
M7	11	16,6166	7,7429	1	0,0054 **
M8	12	17,7646	8,7646	2	0,0125 *
M9	12	23,5391	14,6454	2	0,0007 **
M10	13	23,5452	14,6515	3	0,0021 **

** Significativo ao nível de 1% de probabilidade.

* Significativo ao nível de 5% de probabilidade.

Verificou-se que os modelos de população aberta não diferem, estatisticamente entre si pelo teste adotado, porém o modelo M6 de população fechada é mais parcimonioso entre todos os apresentados, sendo aceito para representar o tamanho da frota.

Modelo reduzido: população fechada com esforço variável.

$$E(m_a) = u + o_1 + o_2 + o_3 + o_4$$

TABELA 3.3.8 Estimativas de máxima verossimilhança para os parâmetros para população fechada com esforço variável de captura.

Parâmetros	Estimativa	Erro padrão	z value	Pr(> z)
(Intercepto)	-5,8985	0,6071	-9,7150	0,0000**
o_1	3,4425	0,2151	16,0080	0,0000**
o_2	3,8814	0,2198	17,6600	0,0000**
o_3	3,5441	0,2158	16,4220	0,0000**
o_4	3,5955	0,2163	16,6250	0,0000**

** Significativo ao nível de 1% de probabilidade.

O tamanho populacional, neste estudo, é obtido somando-se as classes observadas com o valor estimado da classe não observada. Isto pode ser

visualizado facilmente no modelo de regressão considerando a população fechada, como é o caso presente, assim como nos demais modelos.

TABELA 3.3.9 Valores preditos e intervalo de confiança para as frequências de histórias de captura considerando o modelo de população fechada.

Fa	Preditos	X beta	Erro padrão	Limite Inferior	Limite Superior
0	0,0027	-5,8985	0,6073	0,0008	0,0090
1	0,0858	-2,4560	0,4101	0,0384	0,1916
0	0,1330	-2,0171	0,4086	0,0597	0,2963
4	4,1593	1,4254	0,2157	2,7253	6,3480
0	0,0950	-2,3544	0,4099	0,0425	0,2121
6	2,9686	1,0881	0,2198	1,9296	4,5673
3	4,6043	1,5270	0,2149	3,0216	7,0159
142	143,9512	4,9695	0,0817	122,6479	168,9548
0	0,1000	-2,3030	0,4098	0,0448	0,2232
1	3,1253	1,1395	0,2193	2,0333	4,8038
5	4,8472	1,5784	0,2144	3,1841	7,3790
153	151,5467	5,0209	0,0797	129,6248	177,1759
3	3,4596	1,2412	0,2185	2,2543	5,3093
107	108,1631	4,6836	0,0937	90,0096	129,9777
170	167,7582	5,1225	0,0759	144,5645	194,6731
,	5245,0000	8,5650	0,2169	3429,0000	8023,0000

O intervalo de confiança ao nível de 95% de probabilidade para o tamanho populacional é de [4024; 8618], um intervalo mais plausível que o intervalo par população aberta. A estimativa pontual foi de 5245 veículos (aproximadamente um quarto da frota total da cidade).

Quando se tomam os modelos de esforço constante de amostragem, há um acréscimo no valor da deviance. Este acréscimo é significativo sendo portanto, estes modelos estatisticamente diferentes do modelo para população fechada com esforço variável. O comportamento global da frota circulante é fechado, não havendo fuga ou entrada permanente de veículos na área estudada,

o que dá indícios de um comportamento aproximadamente fixo da frota estacionada. Mesmo em um área muito pequena em relação à área total da zona urbana, o número de veículos estacionados naquela região pode ser considerado grande. Não foi verificada alteração significativa em relação ao período de férias e o início das aulas (final de agosto até a terceira semana de setembro).

A correção de Evans & Bonett. (1994) não afetou de forma marcante as estimativas dos parâmetros, não havendo alteração das componentes demográficas. As estimativas corrigidas são mais confiáveis. O esforço de amostragem não pode ser considerado constante, devido à variabilidade na intensidade de amostragem.

3.3.3 Intervalo de confiança para o tamanho populacional

A Tabela 3.3.10 mostra os valores médios e os intervalos de confiança para as estimativas do tamanho populacional, na variável transformada, para os dez modelos considerando. Verificou-se que os modelos considerando população fechada com esforço constante ou sem esforço constante (M_j) foram os que apresentaram intervalos de confiança mais concentrados.

TABELA 3.3.10 Estimativas do tamanho da frota de veículos estacionados no centro de Lavras – MG.

Modelos	Grau de liberdade	Deviance	Valor médio	Intervalo de confiança	
				Limite inferior	Limite superior
M1	5	5,1477	8642	3134	26096
M2	8	8,2075	5048	3146	8367
M3	8	7,9636	6993	4046	12457
M4	6	7,4954	5959	2896	13102
M5	9	7,9947	6993	4046	12457
M6	10	8,8937	5840	4024	8618
M7	11	16,6166	5861	4038	8649
M8	12	17,7646	5864	4040	8654
M9	12	23,5391	5883	4052	8682
M10	13	23,5452	5282	4051	8683

A obtenção das estimativas dos parâmetros revela que é necessário o reajuste do modelo sem o parâmetro s_1 , cuja estimativa é negativa. Quando feito o novo ajuste sem s_1 , a diferença de deviance também não é significativa indicando que o ajuste não foi fortemente afetado pela remoção de qualquer parâmetro.

As estimativas de entrada, saída e permanência de veículos na frota através dos modelos apresentados nas tabelas anteriores, não se alteram significativamente pela troca dos modelos, mas a variância do tamanho populacional cresce acentuadamente.

Houve uma forte concentração de observações nos períodos amostrais e poucos elementos recapturados da população marcada $m_{1101}=152$; $m_{1011}= 107$; $m_{0111} = 170$; $m_{1110} = 142$. Este fato confunde-se com o efeito da dependência de armadilha (Pereira 1995).

Pelo padrão dos resíduos, não houve evidências de heterogeneidade na população. É razoável assim, considerar que a probabilidade de captura não varia de veículo para veículo.

O efeito dessas manipulações nas estimativas do tamanho da população mostrou para os modelos de população fechada, com e sem esforço constante, apresentam intervalos de confiança semelhantes. Para o modelo de população aberta foi de com intervalo de confiança de [3134; 26092] ao nível de 95% de probabilidade. A inclusão de parâmetros com estimativas negativas e ou não significativas aumenta as estimativas de variância e as amplitudes dos IC (intervalos de confiança).

Para populações abertas é possível obter uma exata equivalência entre os modelos J-S e os modelos log-lineares ou qualquer sub-modelo deste por exemplo, omitindo-se entradas ou saídas, ou especificando uma probabilidade de captura constante ou quaisquer outros subconjuntos de amostras.

Outra vantagem da análise de população aberta é que estimativas negativas para nascimentos e mortes podem ser eliminadas. A dependência de armadilha também pode ser incluída através da medida da interação dupla entre os períodos amostrais.

3.4 REFERÊNCIAS BIBLIOGRÁFICAS

CORDEIRO, G. M.; GILBERTO, A. P. Modelos de regressão para análise de dados univariados. In: COLÓQUIO BRASILEIRO DE MATEMÁTICA. 17., 1989, Rio de Janeiro. Rio de Janeiro: IMPA-CNPq, 1989. 353 p.

CORMACK, R. M. Log-linear models for capture-recapture. **Biometrics**, Alexandria, v.45, n. 2, p. 395-413, June 1989.

CORMACK, R. M. Variances of mark-recapture estimates. **Biometrics**, Alexandria, v. 49, n. 4, p. 1188-1193, Dec. 1993.

DEMÉTRIO, C. G. B. Modelos lineares generalizados na experimentação agrônômica. In: SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 5.; REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 38., 1993, Porto Alegre. 113 p.

EVANS M. A.; BONETT D. G.; Bias reduction for multiple capture-recapture estimator closed population. **Biometrics**, Alexandria, v. 50, n. 2, p. 388-395, June 1994.

IRAKA, R.; GENTLEMAN R. R: A language for data analyzes and graphics. **Journal of Computation and Graphics Statistics**. Alexandria, v. 5, n. 2, p. 299-314, Sept. 1996.

NELDER, J.A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society A**, London, v.135, n. 3, p. 370-384, 1972.

PEREIRA, J. E. **Análise de experimentos de captura-recaptura através de modelos log-lineares**. 1995. 106 p. Dissertação (Mestrado) – Escola Superior de Agricultura “Luiz de queiroz”, Piracicaba, SP.

CAPÍTULO IV

**NÚMERO DE VEÍCULOS QUE ESTACIONAM NO CENTRO DE
LAVRAS - MG**



RESUMO

OLIVEIRA, José Otaviano de. Número de veículos que estacionam no centro de Lavras – MG. 2003. 94 p. Dissertação (Mestrado em Agronomia/ Estatística e Experimentação Agropecuária) Universidade Federal de Lavras, Lavras MG.*

A estimação do tamanho populacional pela técnica de captura-recaptura é um processo largamente usado por cientistas que trabalham com populações selvagens. Schnabel (1938) foi quem estendeu o modelo inicial de Lincoln-Petersen para estudos envolvendo mais de uma ocasião amostral. O modelo M_h de Schnabel descreve uma população específica que é fechada geográfica e demograficamente e a única fonte de dependência introduzida é o tempo de captura. O objetivo deste trabalho foi implementar este modelo na estimação do tamanho da frota de veículos que estacionava no centro da cidade de Lavras – MG, no ano de 2002. A função de verossimilhança dos dados foi aproximada por uma densidade Poisson, o que permitiu a construção de priori-posteriori conjugadas da família Gama, o que facilitou o trabalho de inferência. A inferência foi realizada aplicando a Teoria da Decisão com a utilização de dois tipos de funções de perda para obter estimativas de risco mínimo. A melhor estimativa do tamanho da frota que utiliza o centro como estacionamento foi de 6273 veículos, com intervalo de credibilidade confiança de 95% entre 4260 e 9785 veículos.

* Comitê Orientador: Júlio Sílvio de Sousa Bueno Filho - UFLA (Orientador), .

ABSTRACT

OLIVEIRA, José Otaviano de. **Number of vehicles parking at Lavras-MG city center.** 2003. 94 p. Dissertation. (Master Program in Agronomy/ Major in Statistics and Agricultural Experimentation) – Universidade Federal de Lavras.*

Estimates of population sizes by means of capture-recapture techniques is a process broadly used by scientists working with wild populations. Schnabel (1938) extended Lincoln-Petersen's pioneering model to studies involving more than one sampling occasion. Schnabel's M_t model describes a specific population that is closed both geographical and demographically and the only source of introduced dependence is time of capture. The goal of this work was to implement this model to estimate the number of the vehicles parking at Lavras - MG city center, during 2002. The likelihood function was approximated by Poisson density, that allows conjugated prior-posteriors from the Gamma family, to get an easy inference framework. The inference was carried out using Decision Theory with two different kinds of loss functions to get minimum risk estimates. The best estimate was 6273 parking vehicles with a credibility interval ranging from 4260 to 9785 vehicles.

* Guidance Committee: Júlio Sílvio de Sousa Filho - UFLA (Major Professor), .

4.1 INTRODUÇÃO

A técnica de captura-recaptura é em geral aplicada a levantamentos em que um censo completo da população é de difícil realização. O modelo M_t de Schnabel é a generalização mais comum dos processo de estimação proposto originalmente por Lincoln e Petersen, quando se tem mais de uma ocasião amostral (Krebs, 1989).

O fato de se combinar diversas ocasiões amostrais de recaptura para estimar apenas o tamanho populacional permite uma inferência mais adequada, que será o assunto considerado neste capítulo.

Quanto mais se estuda um determinado fenômeno é possível, maior é a informação prévia a uma nova observação. Muitas vezes é possível fazer conjecturas a respeito de um conjunto de informações sobre o objeto de estudo. Se de alguma maneira o conhecimento que se tem puder ser expresso na forma de uma distribuição de probabilidades sobre o parâmetro estudado, é possível incorporar tal conhecimento ao processo de estimação a partir dos dados amostrais.

A abordagem Bayesiana permite incorporar, na análise, ambos os tipos de informações: anterior ao experimento por meio da expressão matemática da função de distribuição a priori e da evidência da experiência expressa pela função de verossimilhança da amostra. A distribuição a posteriori é então obtida como a expressão da distribuição conjunta sobre a qual é efetuada a inferência.

Neste trabalho buscou-se utilizar prioris conjugadas para uma facilidade conceitual, e, principalmente, algébrica na modelagem das diversas recapturas. Desta forma, foi empregado um modelo Gama-Poisson-Gama (priori-verossimilhança-posteriori), que torna mais simples a inferência.

O objetivo principal deste trabalho foi discutir alguns aspectos gerais do processo de estimação sob o enfoque bayesiano, na análise de experimento de captura-recaptura em populações fechada. Aplicou-se esta abordagem em um exemplo particular, que foi o de estimar o tamanho da frota de veículos que estacionam no centro de Lavras, MG, durante o um período do ano de 2002. Utilizou-se, para tanto o modelo de Schnabel, que assume ser a população fechada, conforme proposta metodológica detalhada em Smith (1988, 1991).

4.2 MATERIAL E MÉTODOS

4.2.1 Descrição do experimento

A dinâmica da frota de veículos aqui considerada foi composta de nove períodos amostrais, sendo uma captura e oito recapturas, compreendendo os subperíodos amostrais descritos na primeira parte deste trabalho. A primeira captura e oito recapturas foram obtidas em três semanas diferentes, a partir da última semana de agosto de 2002 até a segunda semana de setembro a partir dos quatro períodos amostrais primários. As amostras foram coletadas em três dias da semana: Segunda, quarta e sexta-feira.

O objetivo foi estimar o tamanho da frota estacionada e verificar a influência da redução de veículos, devido às férias semestrais, sobre este número. Foi considerado o modelo M_t de Schnabel utilizando uma abordagem bayesiana do problema. A frota foi considerada fechada em termos de rotatividade.

Notação usada nas demonstrações

k – número de ocasiões amostrais;

i – ocasião amostral $i = 1, 2, \dots, k$;

n_i – número de indivíduos marcados na amostra i ;

m_i – número de indivíduos marcados na amostra i ;

$$M_i = \sum_{j=0}^{i-1} (n_j - m_j) - \text{soma dos indivíduos marcados até a ocasião } i-1.$$

Estes são todos os indivíduos capturados menos os marcados presentes em cada amostra. A estatística M_i é importante na obtenção dos estimadores do tamanho populacional.

p_i = probabilidade de captura de um indivíduo na ocasião amostra;

$\sum_{i=1}^n m_i = s$: total de indivíduos marcados na população até a amostra i . ocasião i ;

$$\sum_{j=0}^I n_j M_j = T_i$$

A soma do produto do número de indivíduos do número de indivíduos capturados na amostra i pelo o numero de marcados naquela mesma amostra é dada por:

$\lambda_i = n_i m_i / N$: é a constante da distribuição de Poisson contendo as estatísticas observáveis e o inverso do tamanho populacional;

$w = 1/N$ é o inverso do tamanho populacional $0 \leq w \leq 1$;

No estudo do tamanho populacional foi considerado o inverso do tamanho populacional. Esta mudança algébrica permite encontrar uma distribuição a priori conjugada (Smith 1989).

4.2.2 Estimação do tamanho populacional

Seber (1982), citado por Smith (1988), mostra que este problema pode ser modelado por uma função densidade de probabilidade hipergeométrica em que dos N indivíduos da população, numa determinada ocasião de captura (i),

$$P(m_i | N, n_i) = \frac{\binom{M_i}{m_i} \binom{N_i - n_i}{n_i - m_i}}{\binom{N}{n_i}}$$

subdividem-se em dois grupos: marcados e não marcados. Verifica-se claramente que a característica desta distribuição é apropriada aos problemas desta natureza.

É possível provar que para n_i grande e M_i/N pequeno, a função de verossimilhança pode ser aproximada para função densidade de probabilidade de Poisson.

$$f(m_i|\lambda) = \frac{\lambda^{m_i} e^{-\lambda}}{m_i!}$$

A Poisson é um caso limite da distribuição hipergeométrica. Esta aproximação é justificada porque se pode obter, através do teorema de Bayes, uma função de distribuição conjugada a posteriori de fácil manejo algébrico. Frequentemente existe uma informação a respeito do parâmetro que se quer estimar. Este será o caso em estudo, para o qual é repetido muitas vezes o processo amostral (recapturas) para estudar a dinâmica do tamanho da frota. Em outras ocasiões podem haver noções subjetivas em que os valores do parâmetro são mais plausíveis que outros. Também neste caso, a informação a priori a respeito de $w=1/N$ pode ser modelada usando a distribuição Gama (a, b).

Como informação a priori utilizada foi a de que número de veículos estacionados tomados da população será maior ou igual ao total de marcados até determinada ocasião amostral. Esta informação foi incorporada ao processo de estimação, melhorando sua precisão.

A Informação inicial foi modelada por uma distribuição Gama (a, b):

$$f(w) = \frac{b^a}{\Gamma(a)} \cdot w^{a-1} e^{-bw},$$

que contém exclusivamente o parâmetro w . O parâmetro w é igual ao inverso do tamanho da frota, considerado desta forma por motivações algébricas de se conseguir uma posteriori da mesma família da distribuição a posteriori. As constantes a e b são escolhidas de forma a representar a confiança que se tem nas informações sobre o parâmetro N .

em que $w = 1/N$ e N é o tamanho populacional; N guarda a seguinte relação com $\lambda_i = n_i m_i / N$.

A distribuição de Poisson é frequentemente utilizada em estudo de dados em forma de contagem. Se um conjunto de dados m_i segue a distribuição de Poisson com parâmetro λ_i , então a função de distribuição de probabilidades:

$$m_i = 0, 1, 2, \dots,$$

e para o vetor $m_i = (m_1, \dots, m_n)$ de observações i. i. d. em que com $t(m_i) = \sum_{i=1}^n m_i$

é uma estatística suficiente. Nós podemos reescrever a verossimilhança:

$$L[f(m_i | \lambda)] = \prod_{i=1}^n \frac{\lambda_i^{m_i} e^{-\lambda_i}}{m_i!}$$

que é condicional à observação do parâmetro λ_i . Decompondo o termo λ_i e retirando seus termos observáveis tem-se:

$$\lambda_i = \frac{n_i m_i}{N} = n_i m_i w$$

A distribuição de Probabilidade a Posteriori (condicional) é:

$$f(w | S_i; T_i) \propto L[f(m_i | \lambda_i)] f(w)$$

Aplicando as expressões da verossimilhança e da distribuição a priori, segundo o teorema de Bayes, tem-se a expressão:

$$f(\lambda|m_i) = \prod_{i=1}^n \frac{\lambda_i^{m_i} e^{-\lambda_i}}{m_i!} \cdot \frac{b^a}{\Gamma(a)} w^{a-1} e^{-bw}$$

Desprezando os termos constantes, agrupando os termos de mesma base e substituindo as duas estatísticas :

$$\sum_{j=0}^l n_j M_j = T_i$$

$$\sum_{i=1}^n m_i = S_i$$

chega-se à expressão proporcional a posteriori:

$$f(w|S_i T_i) \propto w^{S_i+a-1} e^{-w(b+T_i)}$$

Observa-se que a função de verossimilhança e a distribuição “a posteriori” são conjugadas naturais tais que o núcleo das duas distribuições são idênticos.

A constante de proporcionalidade resultante do teorema de Bayes para o caso da distribuição de probabilidade apresentada é dada por:

$$A = 1 / \int_0^{1/M_{i+1}} w^{S_i+a-1} e^{-(T_i+b)w} dw$$

em que M_{i+1} corresponde ao número de indivíduos marcados na população na ocasião i somados aos elementos capturados na amostra $(i + 1)$.

As funções de perda considerada por Smith (1988) foram:

$$P(w - \hat{w}) = \frac{(w - \hat{w})^2}{w^x}$$

em que x pode assumir os valores (0, 1, 2). Para $x = 0$, tem-se a função de perda quadrática e a estimativa de menor risco corresponde à média da posteriori. Esta é a função de perda mais comum utilizada em trabalhos de captura-recaptura.

Outra função de perda é a chamada função de perda qui-quadrado é obtida para $x=1$ e corresponde a uma estimativa de risco mínimo a posteriori que se afasta da moda a posteriori para distribuições mais assimétrica. O estimador pontual, neste caso, fica ponderado pelo valor paramétrico. Considerando uma função de perda qualquer, associa-se a ela uma função de risco.

No presente estudo, a função estudada é a distribuição a posteriori sobre a qual deseja-se fazer as inferências sobre o parâmetro que representa o tamanho da frota de veículos. A função de risco é definida como a esperança da função de perda:

$$R = E[P(w - \hat{w})].$$

A justificativa para o uso de funções de perda (Teoria da Decisão) pode ser encontrada em livros textos de vários autores, entre eles Silvey (1995), Mood et al (1974) e Bickel & Doksum (1977)

A terceira função de perda considerada é a perda de qui-quadrado proporcional, considerada aqui para mostrar as diversas estimativas de risco mínimo. Neste caso, o estimador fica ponderado pelo quadrado do valor paramétrico.

As estimativas de menor risco foram confrontadas com as estimativas de máxima verossimilhança, que para o caso de se considerar a priori uniforme, eqüivale ao estimador de Schnabel. O intervalo de confiança obtido para estimativas seguiu o padrão clássico de intervalo simétrico.

4.3 RESULTADOS E DISCUSSÃO

4.3.1 Posteriori conjunta

A distribuição conjugada usada (Gama - Poisson - Gama), permitiu encontrar uma função distribuição de probabilidade a posteriori de fácil manejo algébrico. Desta forma, partindo de uma distribuição a priori Gama e assumindo que os dados tivessem uma distribuição de Poisson, chegou-se a uma distribuição a posteriori também Gama. Na tabela 4.3.1 são apresentados todos os dados relativos à captura e às recapturas dos veículos e as estatísticas observáveis necessárias à obtenção do tamanho da frota. O experimento consistiu de uma amostra inicial de $n_i = 49$ elementos capturados e mais oito ocasiões de recaptura subseqüentes, correspondentes aos subperíodos amostrais. O número de itens marcados está na segunda coluna e nas colunas seguintes estão as demais estatísticas obtidas a partir das duas primeiras. Para cada uma das funções de distribuição a priori usadas foi obtida uma posteriori conjunta da família Gama.

TABELA 4.3.1 Dados relativos à contagem da população da frota de veículos estacionados na área central em diversas ocasiões amostrais coletadas em Lavras -MG. Agosto a setembro de 2002.

Amostra	n_i	m_i	S_i	M_i	$\Sigma(n_i M_i)$
0	49	0	0	0	0
1	45	0	0	49	2205
2	80	1	1	94	9725
3	33	2	3	173	15434
4	87	3	6	204	33182
5	116	6	12	288	66590
6	41	2	14	398	82908
7	99	7	21	437	126171
8	45	2	23	529	149976

Inicialmente considerou-se a função de distribuição de probabilidade a priori ($a = 0$; $b = 0$) e verificou-se que as estimativas obtidas são as mesmas das de máxima verossimilhança dos dados. Para cada ocasião amostral, a partir da terceira, (figuras de 4.3.2 a 4.3.5), é possível construir uma distribuição adequada aos dados que corresponde exatamente à função de verossimilhança.

4.3.2 Resultados para a priori uniforme

A vantagem do método Bayesiano está na incorporação direta de uma significativa quantidade de informação a priori. A função de distribuição a priori usada foi, inicialmente, uma priori uniforme Gama ($a = 0$; $b = 0$). Esta função permite obter a posteriori que é somente a própria verossimilhança dos dados. Para este caso, a função não possui um máximo e, então, sua estimativa de moda não existe.

Sob o enfoque da Teoria da Decisão, é possível utilizar funções de perda e funções de risco que tornem a perda mínima, possibilitando, assim, encontrar estimativas ótimas (de menor risco a posteriori). Aqui foram consideradas as funções perda quadrática correspondentes à esperança e a moda da “posteriori” respectivamente. Este enfoque permite encontrar estimativas ótimas considerando a melhor solução de Bayes, aquela que minimiza os risco a posteriori.

Na tabela 4.3.2, as estimativas pontuais para o tamanho populacional para o estimador de Schnabel é média da distribuição “a posteriori”. A suposição de independência é subtendida aqui para as inferências “a posteriori” sejam válidas. A média a posteriori corresponde exatamente ao estimador de risco mínimo considerando uma perda quadrática.

TABELA 4.3.2. Comparação do estimador de Schnabel com o estimador bayesiano considerando as funções de perda quadrática e a moda para os dados da tabela 1 utilizando priori $a = 0$ e $b = o$.

Ocasião amostral	Estimador Schnabel	Moda a posteriori	Média a Posteriori
3	9725	-	9725
4	5145	7712	5145
5	5530	6638	5530
6	5549	6054	5549
7	5922	6378	5922
8	6008	6308	6008
9	6521	6817	6521

Na tabela acima, introduzimos o estimador clássico de Schnabel (modelo M_1) para as $(s - 2)$ ocasiões amostrais e os estimadores bayesianos considerando a função de perda quadrática. Note que o estimador clássico de Schnabel (estimador de máxima verossimilhança) é igual à esperança da posteriori considerando a função de perda quadrática e uma priori uniforme. As estimativas pontuais representam, respectivamente, as estimativas de risco mínimo para a função de perda considerada.

4.3.3 Funções de risco e estimadores pontuais

A função de risco torna-se particularmente importante em processos amostrais continuados em que se quer determinar um ponto ótimo para a finalização da amostragem. Percebe-se que os riscos associados às funções de

perda são ligados à amostragem progressiva. Recomenda-se plotar o risco versus a ocasião amostral como um indicativo para a finalização do processo amostral. No início da amostragem o risco é alto, mas decresce com o aumento dos períodos, pois toda informação sobre os parâmetros provém apenas das amostras atualizadas.

Na figura 4.3.1 tem-se a distribuição “a posteriori” para w para a terceira ocasião amostral considerando a priori uniforme. Verifica-se que a distribuição é assimétrica e que possui uma dispersão relativamente alta. Nas figuras subseqüentes, pode-se verificar que os gráficos da posteriori conjunta são mais concentrados.

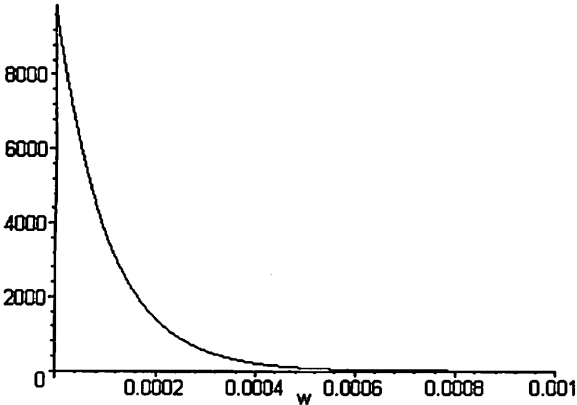


FIGURA 4.3.1 Função de distribuição de probabilidade a posteriori, considerando a priori uniforme ($a = 0$; $b = 0$); para a terceira ocasião amostral.

Esta função não apresenta um máximo definido, não sendo possível determinar a estimativa modal. As estimativas pontuais apresentam riscos elevados se comparada com uma distribuição dos períodos amostrais seguintes.

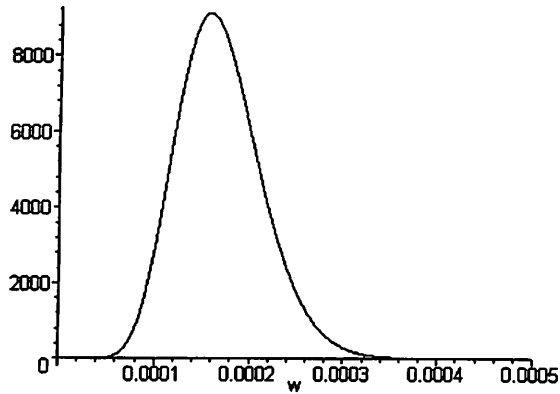


FIGURA 4.3.2 Função de distribuição de probabilidade a posteriori, considerando a priori uniforme ($a = 0$; $b = 0$); para a sexta ocasião amostral.

É facilmente verificado que na sexta ocasião há uma grande concentração da distribuição a posteriori devida a continuidade do processo de amostragem. Pode-se afirmar que estas estimativas são seguramente de risco menor que as estimativas anteriores.

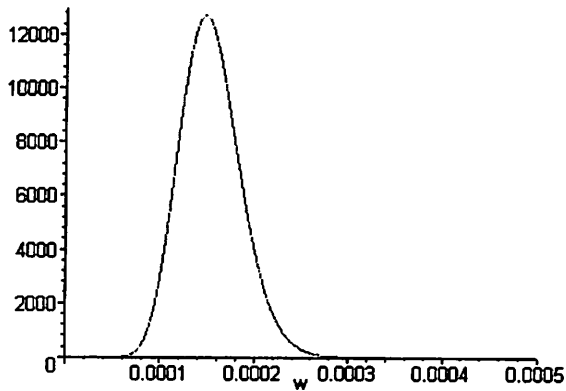


FIGURA 4.3.3 Função de distribuição de probabilidade a posteriori, considerando a priori uniforme ($a = 0$; $b = 0$); para a nona ocasião amostral

Na última ocasião amostral, verificou-se que a posteriori conjunta é ainda mais concentrada, sendo, portanto, preferível na estimação pontual porque permite obter as estimativas de menor risco entre todas as distribuições apresentadas anteriormente.

4.3.4 Estimadores de máxima verossimilhança

Quando se utiliza apenas a informação das amostras, tem-se a estimativa de máxima verossimilhança. A estimativa pontual de menor risco considerando uma função de perda quadrática e priori uniforme é exatamente a estimativa de máxima verossimilhança. Na tabela 4.3.2, este fato pode ser verificado. Estes valores são idênticos aos valores obtidos através estimador clássico de Schnabel.

4.3.5 Resultados para a priori $\Gamma(1, M_{i+1})$

Esta priori foi escolhida com base na informação de que o tamanho da frota e não é fixo, mas atualizado para a cada uma das ocasiões amostrais. A inferência bayesiana efetuada sobre a posteriori, mostra os gráficos de menor dispersão, que contém amostras em maior número de ocasiões e com a priori $a=1$ e $b= M_{i+1}$; $\Gamma(1, M_{i+1})$

4.3.6 Funções de risco e estimadores pontuais

Serão apresentadas aqui as estimativas pontuais correspondentes às funções de perda quadrática, qui-quadrado e qui-quadrado proporcional, cuja

estimativa de risco mínimo é algum ponto, que tende para a estimativa de máxima verossimilhança a posteriori para amostras grandes.

Para o caso de se considerar a segunda priori, as estimativas tornam-se muito mais concentradas em torno do verdadeiro valor paramétrico, devido a quantidade de informação contida.

TABELA 4.3.3 Comparação do Estimador de Schnabel com estimadores bayesianos considerando as funções de perda módulo e quadrática para os dados da tabela 1 utilizando priori $a = 1$ e $b = M_{i+1}$.

Ocasião amostal	Estimador Schnabel	Moda a posteriori	Média a posteriori
3	9275	-	4949
4	5145	7717	3909
5	5530	6636	4781
6	5549	6054	5153
7	5922	6377	5556
8	6008	6309	5759
9	6520	6817	6273

Na figura 4.3.4 e 4.3.5 são apresentadas as priori usadas $a= 1$ e $b= M_{i+1}$, para ocasião 3 e para a ocasião 8. Nota-se que as curvas para priori $a= 1$ e $b= M_{i+1}$ refletindo o tamanho populacional atual são mais informativas, apresentando a forma mais concentrada em torno do verdadeiro valor paramétrico. Isto indica que as estimativas considerando a priori informativa são de riscos menores que as estimativas da priori uniforme.

TABELA 4.3.4 Estimadores bayesianos considerando as funções de perda qui-quadrado, Qui-quadrado proporcional a mediana para os dados da tabela 1 utilizando priori $a = 1$ e $b = M_{i+1}$.

Ocasião amostral	Qui-quadrado	Qui-quadrado proporcional	Mediana
4	5213	7819	4250
5	5578	6694	5000
6	5582	6090	5290
7	5953	6411	5680
8	6008	6335	5845
9	6546	6843	6358

Nota-se que as estimativas, considerando uma função de perda tipo modular, são exatamente a mediana da função de distribuição de probabilidade a posteriori.

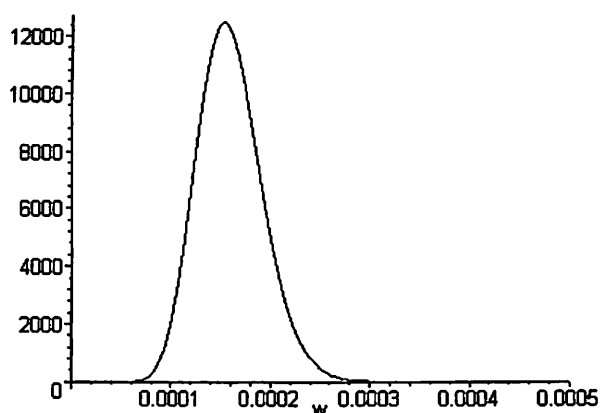


FIGURA 4.3.4 - Função de distribuição de probabilidade a posteriori, considerando a priori uniforme ($a = 1$; $b = M_{i+1}$) para a sexta ocasião amostral.

Verifica-se uma grande concentração da probabilidade em torno do verdadeiro valor paramétrico. Esta concentração indica uma menor variabilidade das informações da amostra, que neste caso ficam mais restritas um determinado patamar. É possível verificar, ainda, que as estimativas considerando as mais diversas funções de perda tendem a valores muito próximos.

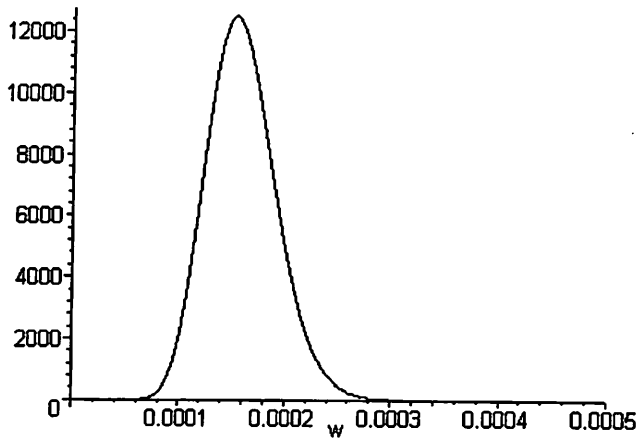


FIGURA 4.3.5 Função de distribuição de probabilidade a posteriori, considerando a priori uniforme ($a = 1$; $b = M_{i+1}$); para a nona ocasião amostral

4.3.7 Intervalo de confiança simétrico

As estimativas dos intervalos de confiança para as ocasiões amostrais estão na tabela a seguir e correspondem ao intervalo simétrico clássico. Observa-se que os intervalos são mais concentrados para últimas ocasiões e bem mais concentrados para a priori informativa ($1, M_{i+1}$). É possível afirmar que o uso da priori informativa pode reduzir a quantidade de períodos amostrais para obter estimativas de riscos iguais.

Quando se considera a “priori” informativa, os intervalos de confiança são mais concentrados (menor dispersão) e atingem uma estabilização dos riscos mais rapidamente. Estimativas obtidas a partir da sexta ocasião amostral são tão confiáveis em termos de preciso quanto as estimativas obtidas na nona ocasião considerando a priori uniforme.

TABELA 4.3.5 Intervalo com 95% de probabilidade de confiança (e metade do nível de significância para cada extremo) para o tamanho populacional considerando da terceira ocasião amostral até a nona ocasião, considerando prioris não informativas.

Ocasião amostral	Moda a posteriori	Média a Posteriori	Intervalo de Confiança	
			Limite Inferior	Limite Superior
3	-	9725	2700	400,000
4	7712	5141	2250	25,000
5	6638	5530	2890	15,050
6	6054	5549	3380	10,720
7	6378	5922	3765	10,800
8	6308	6008	4050	9,700
9	6817	6521	4500	10,280

Os intervalos de confiança são bastante amplos para os períodos iniciais e são mais concentrados para os períodos finais de amostragem.

TABELA 4.3.6 - Intervalo com 95% de probabilidade de confiança (e metade do nível de significância para cada extremo) para o tamanho populacional considerando da terceira ocasião amostral até a nona ocasião, considerando a priori $(1 - M_i + 1)$

Ocasião amostral	Moda a posteriori	Média a Posteriori	Intervalo de Confiança	
			Limite Inferior	Limite Superior
3	-	4949	1700	39.000
4	7717	3909	2790	16.200
5	6636	4781	2560	11.890
6	6054	5153	3195	9.675
7	6377	5556	3545	9.925
8	6309	5759	3945	9185
9	6817	6273	4360	9785

4.4 REFERÊNCIAS BIBLIOGRÁFICAS

BICKEL, P. J.; DOKSUM, K. A. **Mathematical statistics, basic ideas and selected topics.** Oakland, California, 1977.

KREBS, W. J. C. **Ecological Methodology.** New York: Editora & Row, Publishers, 1989.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics.** 3. ed. Tokyo: McGraw-Hill, 1974. 564 p.

SILVEY, S. D. **Statistical Inference.** New York: Chapman & Hall, 1995.
p

SMITH, P. J. Bayesian Methods for Multiple Capture-Recapture Surveys, **Biometrics**, Alexandria, v. 44, n. 4, p.1177-1189, Dec. 1988.

SMITH, P. J. Bayesian analyses for a multiple capture-recapture model, **Biometrika**, London, v. 78, n. 2, p.399-407, June 1991.

CAPÍTULO 5

CONSIDERAÇÕES GERAIS

5.1 Populações Abertas (modelo generalizado de regressão)

- A frota de veículos que estacionam no centro de Lavras possui um comportamento demográfico fechado, não havendo, portanto, entrada ou saída significativa de veículos. Isto implica diretamente em afirmar que a área de estacionamento não apresenta rotação no sentido de receber um contingente significativo de novos carros no período estudado.
- A técnica de captura-recaptura pode ser aplicada para estimar, além do tamanho populacional, aspectos diversos como a entrada e saída de elementos de um determinado grupo, desde que seja possível identificá-los por meio de marcas únicas.
- Ao comparar as estimativas obtidas por meio dos modelos lineares generalizados, os modelos de população fechada apresentaram o melhor desempenho na obtenção de estimativas de tamanho populacional.
- Os modelos log-lineares são mais flexíveis quanto à identificação de possíveis fontes de heterogeneidade de captura quando o processo se estende por muitas ocasiões de amostragem, como no presente caso. O fato de não se encontrarem efeitos de entrada e saída, no entanto, revela que o período foi insuficiente, ou que a população é fechada para o estacionamento.

- O modelo M_1 não deve ser diretamente aplicado quando se suspeitar de alguma fonte de heterogeneidade de captura que não somente o tempo (ocasião de captura).
- Sempre que for possível uma identificação única do indivíduo e uma composição de grupos mutuamente exclusivos de acordo com as histórias de captura o modelo log-linear oferece mais informações a respeito da dinâmica populacional. Isto porque, neste caso é possível modelar a heterogeneidade de captura e identificar a entrada e saída de indivíduos pertencentes a uma determinada população.
- A análise dos modelos propostos mostrou que o melhor dentre eles, foi o modelo para população fechada. Entretanto, nenhum modelo pode ser rejeitado pela sua Deviance.
- As estimativas para o tamanho da população, considerando os diversos modelos não foram significativamente diferentes em termos de estimativa pontual. Houve, no entanto, estimativas fora do espaço paramétrico, em especial para parâmetros da dinâmica populacional.
- Modelos mais simples levaram a menores erros de predição das estimativas (intervalos de confiança mais estreitos).
- Os tamanhos populacionais para as duas abordagens foram semelhantes com a população estimada próxima de 6.000 veículos estacionados com intervalos de [4024; 8618] para o modelo de regressão de Poisson.

5.2 Modelo de Schnabel (População Fechada)

- As estimativas obtidas a partir da abordagem bayesiana apresentam diferenças com o uso da priori uniforme e da priori informativa. Os intervalos de confiança

para as estimativas (de menor risco) utilizando a priori informativa são de menor amplitude.

- O uso da priori informativa permite que a distribuição a posteriori seja mais concentrada e atinja mais rapidamente a estabilização do intervalo de previsão.
- A priori informativa utilizada é intuitiva e melhorou consideravelmente a análise.
- O tamanho populacional da frota, para as diferentes funções de perda considerada, é próximo ao valor obtido pelo estimador de Schnabel. O intervalo de confiança (ao nível de 95%, simétrico) para a última ocasião amostral foi de [4500; 10280] veículos para a priori uniforme e [4360; 9785] veículos para a priori informativa ($1, M_{i+1}$).
- O modelo de Schnabel (M_i) não deve ser diretamente aplicado quando se suspeitar de alguma fonte de heterogeneidade de captura que não somente o tempo (ocasião de captura). Este tipo de efeito não foi detectado nos dados.

5.3 CONCLUSÃO

Recomenda-se o uso da abordagem bayesiana baseada em prioris conjugadas e próprias não apenas para a estimação no modelo de Schnabel (tamanho de populações fechadas), mas também para o estudo de parâmetros da dinâmica populacional.

A frota estacionada no centro de Lavras, MG, constitui uma população fechada e foi estimada em 6273 veículos [4360;9785], o que significa menos de um terço da população de veículos da cidade.

6 REFERÊNCIAS BIBLIOGRÁFICAS

BICKEL, P. J.; DOKSUM, K. A. **Mathematical statistics, basic ideas and selected topics.** Oakland, California, 1977.

BROWN Jr., G. G. **Comparing bayesian, maximum likelihood and classical estimates for the jolly-seber Model.** Carolina: North Carolina State University, 2001. 150 p.

BORGES, C. C. **Um método para obtenção de famílias conjugadas utilizando o teorema de Bayes e Prioris não informativas.** 2001. 55 p. Dissertação (Mestrado em Ciências Exatas) – Universidade Federal de Lavras, Lavras, MG.

CHAO A. Estimating population size for sparse data in capture-recapture experiments. **Biometrics**, Alexandria, v. 45, n. 2, p. 427-438, June 1989.

CHAO, A. An overview of closed capture-recapture models. **Journal of Agricultural, Biological, and Environmental Statistics**, Washington, v. 6, n. 2, p. 158-175, June 2001.

CORDEIRO, G. M.; GILBERTO, A. P. Modelos de regressão para análise de dados univariados. In: COLÓQUIO BRASILEIRO DE MATEMÁTICA. 17., 1989, Rio de Janeiro. Rio de Janeiro: IMPA-CNPq, 1989. 353 p.

CORMACK, R. M. Log-linear models for capture-recapture. **Biometrics**, Alexandria, v.45, n. 2, p. 395-413, June 1989.

CORMACK, R. M. Variances of mark-recapture estimates. **Biometrics**, Alexandria, v. 49, n. 4, p. 1188-1193, Dec. 1993.

DEMÉTRIO, C. G. B. Modelos lineares generalizados na experimentação agrônômica. In: SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 5.; REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 38., 1993, Porto Alegre. 113 p.

DeGROOT, M. H. **Optimal statistical decisions.** New York: McGraw-Hill, 1970. 489 p.

EVANS M. A.; BONETT D. G.; Bias reduction for multiple capture-recapture estimator closed population. **Biometrics**, Alexandria, v. 50, n. 2, p. 388-395, June 1994.

EVANS M. A.; BONETT D. G.; McDONALD L. L.; A general theory for modeling capture-recapture data from a closed population. **Biometrics**, Alexandria, v. 50, n. 2, p. 396-405, June 1994.

FIENBERG, S. E. The multiple recapture census for closed populations and incomplete 2^k contingency tables. **Biometrika**, London, v. 59, n. 3, p.591, Dec. 1972.

FREEMAN, P. R. Sequential recapture. **Biometrika**, London, v. 60, n. 1, p. 141-153, Apr. 1973.

FREEMAN, P. R. A numerical comparison between sequential tagging and sequential recapture. **Biometrika**, London, v. 60, n. 3, p. 499-508, Dec. 1973.

GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian data analysis**. New York: Chapman & Hall, 1995.

IRAKA, R.; GENTLEMAN R. R: A language for data analyzes and graphics. **Journal of Computation and Graphics Statistics**. Alexandria, v. 5, n. 2, p. 299-314, Sept. 1996.

JOLLY, G. M. Mark-recapture models with parameters constant in time. **Biometrics**, Alexandria, v. 38, n. 2, p.301-321, June 1982.

KING, R.; BROOKS, S. P. On the Bayesian analysis of population size. **Biometrika**, London, v. 88, n. 2, p. 317-336, June 2001.

KREBS, W. J. C. **Ecological Methodology**. New York: Editora & Row, Publishers, 1989.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3. ed. Tokyo: McGraw-Hill, 1974. 564 p.

McCULLAGH, P.; NELDER, J. A. **Generalized linear models**. 2. ed. London: Chapman & Hall, 1989. 511 p.

NELDER, J.A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society A**, London, v.135, n. 3, p. 370-384, 1972.

NICHOLS, J. D. ; HINES J. E.; POLLOCK, K. H.; HINZ, R. L.; LINK, W. A. Estimating breeding proportions and testing hypotheses about costs of reproduction with capture-recapture data. **Ecology**, Washington, v. 75, n.7, p. 2052-2065, Oct. 1995.

PEREIRA, J. E. **Análise de experimentos de captura-recaptura através de modelos log-lineares**. 1995. 106 p. Dissertação (Mestrado) – Escola Superior de Agricultura “Luiz de queiroz”, Piracicaba, SP.

SILVEY, S. D. **Statistical Inference**. New York: Chapman & Hall, 1995.

P

SCHWARZ, C. J; SEBER G. A. F. Estimating animal abundance: review III. **Statistics Science**, n. 14, p. 427-456. 1999.

SMITH, P. J. Bayesian Methods for Multiple Capture-Recapture Surveys, **Biometrics**, Alexandria, v. 44, n. 4, p.1177-1189, Dec. 1988.

SMITH, P. J. Bayesian analyses for a multiple capture-recapture model, **Biometrika**, London, v. 78, n. 2, p.399-407, June 1991.

SOLOW R.; A, PALKA, D. On Bayesian Estimation of Group Size, **Biometrics**, Alexandria, v. 52, n. 1, p. 335-340, Mar. 1996.