

Universidade Estadual de Campinas - UNICAMP
Instituto de Matemática, Estatística e Ciência da Computação - IMECC

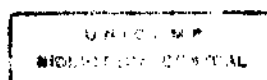
**Comparação das Estimações do Tamanho de uma
População por Captura-Recaptura utilizando
Amostragem Simples, Múltipla e Sequencial**

Luz Jeanet Bullón Camarena

Prof. Dra. Cíclia Yuko Wada
Orientadora

Campinas - SP


1995



**COMPARAÇÃO DAS ESTIMAÇÕES DO TAMANHO DE UMA
POPULAÇÃO POR CAPTURA-RECAPTURA UTILIZANDO
AMOSTRAGEM SIMPLES, MÚLTIPLA E SEQUENCIAL**

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por **Luz Jeanet Bullón Camarena** e aprovada pela Comissão Julgadora.

Campinas, 14 de fevereiro de 1995.


Profa. Dra. Círcia Yuko Wada.

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de MESTRE em Estatística.

Aos meus pais,
Olga e Justino

MEU AGRADECIMENTO
aos que contribuíram na realização
do presente trabalho

À Profa. Cicília Yuko Wada pela orientação do presente trabalho;

À Profa. Clarice Azevedo de Luna Freire pela sua colaboração e entusiasmo permanentes;

Ao IMECC pela oportunidade concedida e a CAPES e FAEP pelo suporte financeiro;

Aos professores, colegas e amigos vinculados de alguma forma à minha formação;

Ao povo brasileiro que me fez sentir parte sua, por fazer da minha estadia uma experiência inesquecível;

Ao povo peruano, meu povo, que embora longe dos olhos manteve cativos meus pensamentos e sentimentos;

Aos meus irmãos pelo incentivo e carinho.

RESUMO

Neste trabalho, são apresentados os modelos e esquemas de amostragem no método de Captura-Recaptura, C-R, para estimação do tamanho de uma população fechada N . A comparação dos estimadores destes modelos é realizada através dos intervalos de confiança baseados na distribuição assintótica e intervalos de confiança obtidos utilizando o método bootstrap.

Primeiramente, são considerados os modelos de C-R de duas amostras (ou ocasiões) com os estimadores de Petersen (1896) e Chapman (1952), também é estudada a distribuição assintótica do estimador de Petersen, (Sen, 1987). Em segundo lugar, são considerados os modelos de múltiplas amostras, com os estimadores de Leite et al. (1988) e o de Petersen (ponderado) multiamostrai (Sen, 1987). Em terceiro lugar, é realizado o estudo de métodos de C-R de amostragem sequencial (Sen, 1987) tanto no caso do tamanho de amostra fixo como estocástico.

Para os casos de amostragem simples e múltipla, e para amostragem sequencial de tamanho de amostra fixo, o método bootstrap apresenta-se como uma alternativa para a avaliação das estimativas de N e a estimação da sua variabilidade.

Simulações de experimentos de C-R são feitas apresentando-se as estimativas e os intervalos de confiança em cada esquema de amostragem. Os programas de computação correspondentes foram desenvolvidos usando o software estatístico *SAS* e o software matemático *Mathematica*.

ÍNDICE

Capítulo 1: Método Bootstrap na Captura-Recaptura	1
1.1 Introdução	1
1.2 Estimativa Bootstrap do Erro Padrão	5
1.3 Intervalos de Confiança Baseados nos Percentis Bootstrap	6
Capítulo 2: Estimação do Tamanho da População por Amostragem Simples.....	9
2.1 Introdução	9
2.2 Estimador de Petersen na Amostragem Direta Simples.....	10
2.3 Estimador de Petersen na Amostragem Inversa Simples	12
2.4 Distribuição do Estimador de Petersen na Amostragem Simples	14
2.5 Metodologia Bootstrap nos Experimentos de C-R para Duas Amostras	16
Capítulo 3: Estimação do Tamanho da População por Amostragem Múltipla	31
3.1 Introdução	31
3.2 Estimador Multiamostrai para Amostragem Direta	31
3.3 Estimador Multiamostrai para Amostragem Inversa.....	34
3.4 Estimador de Petersen (Ponderado) Multiamostrai	36
3.5 Distribuição Assintótica do Estimador de Petersen no Caso Multiamostrai.....	40
3.6 Metodologia Bootstrap nos Experimentos de C-R para Três Amostras	43
Capítulo 4: Estimação do Tamanho da População por Amostragem Sequencial	52
4.1 Introdução	52
4.2 Amostragem Sequencial com o Tamanho de Amostra Fixo	53

4.2.1 Distribuição Assintótica do Estimador Sequencial para	
Amostra de Tamanho Fixo	57
4.2.2 Simulação de um Experimento de C-R de Amostragem Sequencial e	
Cálculo da Estimativa do Tamanho de População e Intervalos de	
Confiança Assintóticos	62
4.2.3 Intervalos de Confiança Bootstrap	67
4.3 Amostragem Sequencial com o Tamanho de Amostra Estocástico	68
4.3.1 Simulação de um Experimento de C-R de Amostragem Sequencial	
com Tamanho de Amostra Estocástico e Cálculo da Estimativa do	
Tamanho da População	70
Capítulo 5: Considerações Finais	73
Apêndice	75
Referências Bibliográficas	104

CAPÍTULO 1

MÉTODO BOOTSTRAP NA CAPTURA-RECAPTURA

1.1 Introdução

Um dos problemas que têm preocupado os biólogos e ecologistas é a estimação da abundância animal, ou dito de uma outra forma, a estimação do tamanho de uma população animal.

Os estatísticos têm realizado estudos sobre os métodos apropriados para tal problema, entre eles os denominados de captura e recaptura (C-R), assim chamados porque se baseiam na captura de uma amostra de uma população de tamanho desconhecido, marcação conveniente da amostra, libertação da mesma e posterior recaptura. Se a amostragem de recaptura for feita uma única ocasião, trata-se do método de amostragem simples; se houver duas ou mais ocasiões, o método é de amostragem múltipla.

Os modelos de C-R, aplicáveis não somente a populações animais, mas também a populações humanas, são classificados segundo o tipo de população a ser estimada. Os tipos de população para os quais dirigem-se os modelos de C-R são:

1. População Fechada - é uma população cujo tamanho permanece constante durante o período de investigação, isto é, não existe recrutamento (nascimentos ou imigração) nem perdas (mortes ou emigração). Estas são suposições fortes e nunca completamente verdadeiras em uma população biológica, porém, se o estudo for apropriadamente planejado pode-se ter, pelo menos aproximadamente, uma população fechada.
2. População Aberta - é uma população na qual existe um ou mais tipos de recrutamento e perdas durante o percurso do experimento.

Em ambos os casos, somente o componente demográfico do conceito fechado varia; o componente geográfico, isto é, a delimitação por fronteiras permanece constante já que os parâmetros da população referem-se a uma área sem a qual eles não teriam significado.

Os métodos de C-R para estimação do tamanho de população em uma população fechada podem ter os esquemas de amostragem direta, de amostragem inversa ou de amostragem seqüencial.

Petersen (1896), considerou o problema de C-R em uma situação simples consistindo do seguinte. Uma amostra de n_1 unidades é capturada sem reposição de uma população de N (desconhecido) unidades; os indivíduos da amostra são marcados convenientemente e logo libertados de forma que fiquem misturados completamente com o restante da população. Depois de passado um tempo, uma segunda amostra de n_2 unidades é obtida com ou sem reposição da mesma população, obtendo-se m_2 indivíduos marcados. Assumindo que os indivíduos marcados e não marcados estão igualmente representados na segunda amostra, pode-se escrever

$$\frac{m_2}{n_1} = \frac{n_2 - m_2}{N - n_1} \left(= \frac{n_2}{N} \right)$$

obtendo-se o estimador clássico de Petersen:

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

Um exemplo típico, ilustrativo do método de C-R para estimar o tamanho de uma população fechada é o seguinte (Otis et al. (1978)): Numa área a ser estudada, delimitada fisicamente, quer-se estimar o tamanho de uma população de uma determinada espécie animal. Para isso, colocam-se armadilhas igualmente espaçadas, (por exemplo 144 armadilhas dispostas em 12 filas por 12 colunas e separadas 7 m. entre si). No começo do estudo, uma amostra de tamanho n_1 é tomada da população, os animais são marcados para sua identificação futura e são retornados à população no mesmo ponto de sua captura. Depois de passado um tempo determinado (geralmente um dia) para que os animais marcados e não marcados se misturem, uma segunda amostra de tamanho n_2 é tomada. Esta segunda amostra contém tanto indivíduos marcados (m_2), como não marcados ($n_2 - m_2$), e todos os animais capturados são retornados à população. Portanto, pode-se estimar o tamanho N da população através do estimador de Petersen dado acima.

O método de C-R de amostragem múltipla pode ser considerado uma extensão do método de Petersen para uma série de s (≥ 2) amostras de tamanhos fixados ou não n_1, \dots, n_s , respectivamente. Em cada amostra a partir da segunda, selecionada sem

reposição registra-se o número de indivíduos marcados obtidos (m_i , $i=2,\dots,s$, $m_1=0$). Todos os indivíduos da amostra recebem uma nova marca antes de retornar à população. As marcas podem ser distinguíveis entre si de uma ocasião para outra ou não. Seja M_i o número de indivíduos marcados na população exatamente antes da i -ésima amostra ser capturada. Então

$$M_i = \sum_{j < i} (n_j - m_j), \quad i=1,\dots,s+1.$$

Note-se que $M_1=m_1=0$, $M_2=n_1$ e que M_{s+1} é o número total de indivíduos distintos observados durante o experimento.

Se no exemplo acima o procedimento de amostragem continuar por s ocasiões, tem-se um exemplo de recapturas múltiplas. Considere-se por exemplo $s=3$, então n_3 é o tamanho da terceira amostra, m_3 o número de animais marcados capturados na terceira amostra, marcas que podem provir da primeira ou da segunda amostragem. O quadro abaixo ilustra a notação relacionada a este experimento:

	1ª ocasião	2ª ocasião	3ª ocasião
Tamanho da amostra	n_1	n_2	n_3
Nº de indivíduos marcados			
capturados na amostra	$m_1=0$	m_2	m_3
Nº de indivíduos distintos			
marcados antes da ocasião s	$M_1=0$	$M_2=n_1$	$M_3=n_1 + (n_2 - m_2)$

O objetivo desta dissertação é estudar alguns dos métodos de C-R para estimação do tamanho N de população, considerando o tipo fechada, esquemas de amostragem direta e inversa, tanto simples como múltipla, e esquema seqüencial, sob um enfoque que propicie:

1. O entendimento desses métodos por profissionais das áreas que os utilizam na prática, e
2. A escolha talvez mais adequada de um método.

Alguns autores propõem estimadores para o cálculo do tamanho da população e derivam a precisão desses estimadores (Chapman (1951) e Bailey (1951)). Entretanto, outros estudos e modelos propõem estimadores cujas variâncias não podem ser determinadas analiticamente no caso da amostragem múltipla (Chapman (1952), Schnabel (1958), Darroch (1958), Cormack (1969), Otis et al. (1978) e Seber (1965)).

Para a determinação da eficiência é necessário a comparação das variâncias desses estimadores. O método estatístico bootstrap pode ser utilizado para calcular a variância e fazer a comparação no caso do método de C-R quando a amostragem é simples, e no caso da amostragem múltipla para solucionar o problema da falta da expressão analítica desta variância.

O método estatístico bootstrap foi desenvolvido por Efron(1979), com o objetivo de solucionar o cálculo da precisão dos estimadores cuja solução teórica é difícil. Basea-se na analogia na qual os dados observados assumem o papel de uma população básica. Como resultado, a variância bootstrap, a distribuição, e intervalos de confiança são obtidos pela retirada de sucessivas amostras com reposição da amostra original. O bootstrap é uma abordagem para a avaliação de estimativas e estatísticas em geral. As principais vantagens do método estão no fato de possibilitarem análises estatísticas livres de suposições de modelos teóricos para a distribuição dos dados dando margem à exploração de propriedades amostrais de interesse, independentemente de suas formas analíticas.

O método bootstrap mostra-se útil nos modelos de C-R na estimação de variâncias e intervalos de confiança para as estimativas do tamanho de população. Como estabelecido em Buckland e Garthwaite (1991), no experimento de C-R ambos métodos bootstrap paramétrico e não paramétrico podem ser aplicados. Para ser aplicado o método paramétrico dois tipos de análises podem ser realizadas: o primeiro, condicional aos tamanhos observados das amostras da C-R e o segundo, não condicional a estes tamanhos. As distribuições amostrais assumidas para a geração das réplicas bootstrap dependem desta escolha e podem ser a hipergeométrica e a multinomial respectivamente. O bootstrap não paramétrico assume implicitamente a distribuição multinomial.

O objetivo deste capítulo é fazer a revisão do método bootstrap quando aplicado ao método de C-R e apresentado basicamente nos estudos de Buckland e Garthwaite (1991) e Buckland (1983). Na seção 2 é apresentada a metodologia bootstrap para a obtenção da variância bootstrap do estimador quando o método de C-R é de amostragem direta simples e múltipla. O método bootstrap descrito é o paramétrico e as análises são a condicional e a não condicional aos tamanhos observados das amostras. Na seção 3 é tratada a estimação bootstrap por intervalos de confiança, intervalos que são baseados nos percentis bootstrap e intervalos de confiança quando é realizada a correção devida à falha da condição de simetria da distribuição das estimativas bootstrap do parâmetro.

1.2 Estimativa Bootstrap do Erro Padrão

Suponha que exista uma população de tamanho N e é utilizado o esquema de amostragem direta múltipla. Seja s (≥ 2) o número de ocasiões de amostragem e seja M_{s+1} o número total de indivíduos capturados durante o experimento.

Baseado na informação acima, o objetivo do problema de C-R é a estimação do tamanho da população e a avaliação desta estimativa por meio do erro padrão, bem como fornecer intervalos de confiança para o total populacional.

Nesta seção serão descritos o método bootstrap (Buckland e Garthwaite (1991)) para avaliar a estimativa do erro padrão para amostragem simples e múltipla.

Suponha que o experimento consiste de duas amostras ou de recaptura simples, onde n_i é o número de indivíduos na amostra i , ($i=1,2$) e m_i é o número de indivíduos com marca capturados na i -ésima amostra, ($m_1=0$). Portanto o estimador de Petersen N_p ou alguma modificação dele (serão vistos no capítulo 2) podem ser facilmente calculados com as informações acima. Com a finalidade de obter o erro padrão dessa estimativa através do método bootstrap considera-se a análise não condicional aos tamanhos das amostras assumindo que as histórias de captura seguem uma distribuição multinomial $(N, p_{11}, p_{10}, p_{01}, p_{00})$, onde os índices das probabilidades associadas à multinomial indicam a história de captura dos indivíduos, isto é, a seqüência 11 significa a captura do indivíduo nas duas ocasiões, 10 significa que o indivíduo foi capturado na primeira ocasião e não foi capturado na segunda, 01 significa que o indivíduo não foi capturado na primeira amostra e sim na segunda e finalmente 00 representa a história dos indivíduos nunca capturados. Considerando $\hat{p}_i = n_i/\hat{N}$ a probabilidade estimada de captura na amostra i , as probabilidades p_{ij} correspondentes à multinomial são estimadas pelas \hat{p}_{ij} e obtidas como segue,

$$\begin{aligned}\hat{p}_{11} &= \hat{p}_1 \hat{p}_2 = m_2 / \hat{N} \\ \hat{p}_{10} &= \hat{p}_1 (1 - \hat{p}_2) = (n_1 - m_2) / \hat{N} \\ \hat{p}_{01} &= (1 - \hat{p}_1) \hat{p}_2 = (n_2 - m_2) / \hat{N} \\ \hat{p}_{00} &= (1 - \hat{p}_1) (1 - \hat{p}_2) = (\hat{N} - (n_1 + n_2 - m_2)) / \hat{N}.\end{aligned}$$

Similares análises são feitas quando há mais do que duas amostras. Os $2^s + 1$ parâmetros da distribuição multinomial assumida são estimados. O modelo fornece a estimativa \hat{N} de N e as probabilidades de cada história de captura são estimadas pela multiplicação das probabilidades de captura estimadas na amostra i , \hat{p}_i ($i=1, \dots, s$). Por exemplo, no experimento de três amostras, a probabilidade estimada da história 101 é

$\hat{p}_1(1-\hat{p}_2)\hat{p}_3$. Para s amostras, a distribuição associada às histórias de captura é Multinomial $(N, p_{11...11}, p_{11...10}, \dots, p_{00...00})$. A análise utilizando esta distribuição é denominada nesta dissertação de método 1.

Na análise não condicional aos tamanhos das amostras também pode ser assumida a distribuição multinomial $(M_{s+1}, p_{11}, p_{10}, p_{01})$, onde os parâmetros são o número total de indivíduos diferentes observados durante o experimento, e as probabilidades correspondentes às histórias de captura dos indivíduos alguma vez observados. Considerando-se a distribuição multinomial com estes parâmetros, a análise será denominada de método 2.

As réplicas bootstrap são obtidas pela geração de observações da distribuição multinomial ajustada. Para a obtenção de observações que seguem distribuição multinomial com probabilidades estimadas \hat{p}_{ij} ($i, j=0,1$), geram-se N (ou M_{s+1}) observações de uma distribuição uniforme $(0,1)$ atribuindo valor 1 à categoria (ou história de captura) caso a observação pertença ao intervalo definido entre 0 e a probabilidade de captura estimada correspondente. Assim, obtém-se o número de observações n_{ij} das categorias. Uma réplica bootstrap fornece valores de $n_1=n_{10}+n_{11}$, $n_2=n_{01}+n_{11}$ e $m_2=n_{11}$, e portanto em cada réplica pode-se calcular a estimativa bootstrap de N . O procedimento é repetido um número grande B de vezes obtendo-se B estimativas \hat{N}_j^* ($j=1, \dots, B$). A variância bootstrap de \hat{N} é simplesmente a mesma dos \hat{N}_j^* ,

$$\hat{V}(\hat{N}) = [\sum \hat{N}_j^{*2} - (\sum \hat{N}_j^*)^2 / B] / (B - 1). \quad (1.1)$$

Uma outra análise que pode ser feita é condicional aos tamanhos observados das amostras. Neste caso a distribuição assumida é a hipergeométrica. A análise utilizando esta distribuição será denominada de método 3. As réplicas bootstrap são obtidas pela geração de observações desta distribuição com parâmetros N, n_1, n_2 onde N é substituído por sua estimativa (estimador de Petersen ou alguma modificação dele). Condicionada a n_1, n_2 a distribuição gera valores para m_2 , sendo possível portanto obter valores simulados \hat{N}_j^* ($j=1, \dots, B$).

1.3 Intervalos de Confiança baseados no Percentil Bootstrap

Nesta seção descreve-se a técnica para a construção de intervalos de confiança para o parâmetro N baseados nos percentis da distribuição bootstrap das estimativas do

tamanho da população, isto é, descreve-se como é que os percentis do histograma bootstrap que representa a distribuição das replicações \hat{N}^* , podem ser usados para definir limites de confiança. Esta técnica devida a Efron (1982) é descrita por Buckland (1983) para os modelos de C-R.

Suponha que as observações tenham sido geradas das distribuições ajustadas como explicadas na seção anterior e denominadas de métodos 1, 2 e 3, e portanto calculadas as replicações bootstrap \hat{N}^* . Seja \hat{G} a função acumulada da distribuição bootstrap de \hat{N}^* . Para um valor α dado entre 0 e 0,5 definem-se

$$\begin{aligned}\hat{N}_{inf}(\alpha) &= \hat{G}^{-1}(\alpha) \\ \hat{N}_{sup}(\alpha) &= \hat{G}^{-1}(1-\alpha)\end{aligned}$$

Dados que pela definição $\hat{G}^{-1}(\alpha)$ e $\hat{G}^{-1}(1-\alpha)$ são os percentis 100α -ésimo e $100(1-\alpha)$ -ésimo da distribuição bootstrap, respectivamente, o método consiste em tomar o intervalo $[\hat{N}_{inf}(\alpha), \hat{N}_{sup}(\alpha)]$ como um intervalo central de $100(1-2\alpha)\%$ de confiança para N .

A situação descrita refere-se à situação ideal na qual o número de replicações bootstrap é infinita. Na prática e quando é realizado um estudo de Monte Carlo, o número de replicações B é um número finito, grande, e \hat{G} é aproximado por

$$\hat{G}(n) = \# \{ \hat{N}_j^* \leq n \} / B$$

onde $\# \{ \hat{N}_j^* \leq n \}$ é o número de vezes nas quais verifica-se a desigualdade e j varia de 1 até B . Logo, geram-se B replicações bootstrap de \hat{N}^* , independentes, ordenam-se de forma crescente os valores \hat{N}_j^* denotando-os por $\hat{N}_{(k)}^*$. Então os limites do intervalo estarão dados por $\hat{N}_{inf} = \hat{N}_{(k1)}^*$ e $\hat{N}_{sup} = \hat{N}_{(k2)}^*$ onde $k1 = (B+1)\alpha$ e $k2 = (B+1)(1-\alpha)$. Em geral, $k1$ e $k2$ serão não inteiros e ambos deverão ser arredondados ao valor inteiro mais próximo.

Para obter intervalos de confiança de 95%, $B=1000$ seria uma escolha satisfatória enquanto valores menores para B resultaria inadequado. (Buckland(1984)).

Os intervalos de confiança obtidos pela técnica descrita acima seriam exatos se o componente devido à simulação não existisse. Buckland (1984) examinou a condição sob a qual a técnica seria exata e portanto a variação Monte Carlo ignorada, e descreveu-a como uma condição de simetria da distribuição das estimativas obtidas. Efron(1982) mostra como o método do percentil pode ser corrigido quando esta condição falha.

Essencialmente a correção baseia-se na transformação da distribuição em uma outra satisfazendo a condição de simetria. O intervalo de confiança achado utilizando esta correção é chamado de intervalo de confiança de vício corrigido.

Suponha \hat{N} , um estimador consistente do tamanho da população N , cuja função de distribuição acumulada é $G(\hat{N}|N)$. Suponha que um conjunto de dados forneça a estimativa \hat{N}_0 para N , então a função de distribuição acumulada estimada está dada por $\hat{G}(\hat{N}|N=\hat{N}_0)$. Supondo que B valores são simulados independentemente e ordenados de forma crescente de $\hat{N}_{(1)}^*$ até $\hat{N}_{(B)}^*$, um intervalo para N , de $(1-2\alpha)\%$ de confiança estará dado por $[\hat{N}_{(k1)}^*, \hat{N}_{(k2)}^*]$, onde $k1=(B+1)\alpha$ e $k2=(B+1)(1-\alpha)$.

Agora, suponha $\hat{N}_m^* < \hat{N}_0 < \hat{N}_{m+1}^*$, isto é, m das B replicações bootstrap produzem estimativas para N menores que \hat{N}_0 . Define-se $Z_0 = \Phi^{-1}(m/B)$, onde Φ denota a função de distribuição acumulada normal padrão. Logo, o procedimento corrigido utiliza as estimativas simuladas para estimar α_1 inferior tal que $\alpha_1 = \Phi(2Z_0 - Z_\alpha)$ e α_2 superior tal que $\alpha_2 = 1 - \Phi(2Z_0 + Z_\alpha)$, pontos de $\hat{G}(\hat{N}|N=\hat{N}_0)$ os quais fornecem limites de confiança de aproximadamente $100(1-2\alpha)\%$. Aqui, Z_α é o ponto α superior para a normal padrão de forma que $\Phi(Z_\alpha) = 1 - \alpha$. Os valores $k1$ e $k2$ podem agora ser aproximados por $k'1 = (B+1)\alpha_1$ e $k'2 = (B+1)(1-\alpha_2)$ e o intervalo de $100(1-2\alpha)\%$ de confiança é dado por $[\hat{N}_{(k'1)}^*, \hat{N}_{(k'2)}^*]$.

CAPÍTULO 2

ESTIMAÇÃO DO TAMANHO DA POPULAÇÃO POR AMOSTRAGEM SIMPLES

2.1 Introdução

O método de C-R de amostragem simples ou duas amostras, utilizado para a estimação do tamanho de uma população, pode ser descrito da seguinte forma: uma amostra de n_1 unidades é tomada sem reposição de uma população de N (desconhecido) unidades, e marcada convenientemente e logo libertada de forma que fique misturada completamente com o restante da população. Toma-se uma segunda amostra de n_2 unidades da mesma população, obtendo-se m_2 indivíduos marcados. Assumindo que os indivíduos marcados e não marcados estão igualmente representados na segunda amostra, pode-se escrever

$$\frac{m_2}{n_2} = \frac{n_1 - m_1}{N - n_1} \left(= \frac{n_1}{N} \right)$$

obtendo-se o estimador clássico de Petersen

$$\hat{N}_P = \left[\frac{n_1 n_2}{m_2} \right] \quad (2.1)$$

onde $[x]$ significa o maior inteiro menor ou igual que x .

Este princípio de igualdade das proporções constitui a base de muitos dos mais complexos modelos de C-R. Se o esquema de amostragem considera n_2 fixo, a amostragem é chamada direta e se n_2 for aleatório (m_2 fixo) a amostragem é chamada inversa. O estimador de Petersen na amostragem direta pode ser justificado pela utilização de dois modelos probabilísticos dependendo da forma da seleção da segunda amostra. O modelo é hipergeométrico se a segunda amostra for obtida sem reposição ou é binomial se esta for obtida com reposição. O estimador de Petersen para amostragem inversa é justificado pela distribuição hipergeométrica negativa.

Nas seções 2.2 e 2.3 apresenta-se o estimador de Petersen na amostragem direta

e na inversa respectivamente. Na seção 2.4 é descrita a derivação da distribuição do estimador na amostragem simples. Finalmente na seção 2.5 são apresentados alguns resultados da aplicação da metodologia bootstrap aos métodos de C-R de amostragem simples.

2.2 Estimador de Petersen na Amostragem Direta Simples

Para \hat{N} estimar adequadamente o tamanho da população N devemos estabelecer um conjunto de suposições as quais não são mutuamente exclusivas:

1. A população é fechada de forma que N é constante.
2. Todos os indivíduos têm a mesma probabilidade de serem capturados na primeira amostra.
3. A marca não afeta a captura do indivíduo.
4. A segunda amostra é aleatória simples.
5. Os indivíduos não perdem suas marcas durante o tempo entre as duas amostras.
6. Todas as marcas são registradas na recaptura dos indivíduos na segunda amostra.

Dependendo da forma de seleção da segunda amostra, os modelos podem ser hipergeométrico ou binomial.

a) Modelo hipergeométrico.

Se as suposições 1, 4, 5 e 6 são satisfeitas (4 dependendo da validade de 2 e 3), a função de distribuição de probabilidade de m_2 dados N , n_1 e n_2 é

$$P(m_2 | N, n_1, n_2) = \frac{\binom{n_1}{m_2} \binom{N - n_1}{n_2 - m_2}}{\binom{N}{n_2}}, \quad \max \{0, n_1 + n_2 - N\} \leq m_2 \leq \min \{n_1, n_2\}. \quad (2.2)$$

A esperança e a variância de m_2 são, respectivamente,

$$E(m_2) = \frac{n_1 n_2}{N} \quad \text{e} \quad V(m_2) = \frac{n_1 n_2}{N} \left(1 - \frac{n_1}{N} \right) \left(\frac{N - n_1}{N} \right).$$

Para achar o estimador de máxima verossimilhança (EMV) de N , consideram-se os quocientes,

os quocientes,

$$\frac{P(m_2|N, n_1, n_2)}{P(m_2|N-1, n_1, n_2)} \geq 1 \text{ e } \frac{P(m_2|N, n_1, n_2)}{P(m_2|N+1, n_1, n_2)} \geq 1.$$

Utilizando o primeiro,

$$\frac{\binom{n_1}{m_2} \binom{N-n_1}{n_2-m_2}}{\binom{N}{n_2}} \geq \frac{\binom{n_1}{m_2} \binom{N-1-n_1}{n_2-m_2}}{\binom{N-1}{n_2}}$$

obtém-se,

$$N \leq \frac{n_1 n_2}{m_2}.$$

Utilizando o segundo quociente,

$$\frac{\binom{n_1}{m_2} \binom{N-n_1}{n_2-m_2}}{\binom{N}{n_2}} \geq \frac{\binom{n_1}{m_2} \binom{N+1-n_1}{n_2-m_2}}{\binom{N+1}{n_2}}$$

obtém-se,

$$N \geq \frac{n_1 n_2}{m_2} - 1.$$

Assim, o EMV de N , \hat{N} satisfaz

$$\frac{n_1 n_2}{m_2} - 1 \leq \hat{N} \leq \frac{n_1 n_2}{m_2}$$

e dado que N é inteiro, então \hat{N} é o estimador de Petersen \hat{N}_p

$$\hat{N} = \hat{N}_p = \left\lceil \frac{n_1 n_2}{m_2} \right\rceil \quad (2.3)$$

Note-se que para todo N , n_1 , n_2 finitos, existe uma probabilidade positiva de m_2 ser igual a zero, isto é,

$$\frac{\binom{N-n_1}{n_2}}{\binom{N}{n_2}} > 0$$

(a não ser que $n_2 > N - n_1$) daí, $\hat{N} = +\infty$ com probabilidade positiva e por tanto, \hat{N} não tem momentos (de qualquer ordem) finitos. Para eliminar este problema, uma modificação do estimador deve-se a Chapman (1951) e é a seguinte:

$$\hat{N}_{ch} = \left[\frac{(n_1+1)(n_2+1)}{(m_2+1)} - 1 \right], \quad (2.4)$$

onde m_2 tem distribuição hipergeométrica com parâmetros N, n_1, n_2 . \hat{N}_{ch} possui momentos finitos e é não viciado, isto é, $E(\hat{N}_{ch})=N$, se $n_1+n_2 \geq N$. A estimação da variância de \hat{N}_{ch} para amostras grandes é aproximadamente,

$$\hat{V}(\hat{N}_{ch}) = \frac{(n_1+1)(n_2+2)(n_1-m_2)(n_2-m_2)}{(m_2+1)^2(m_2+2)}. \quad (2.5)$$

b) Modelo binomial.

Este modelo é adequado na situação menos comum da amostragem feita com reposição, por exemplo, se os indivíduos são somente observados no lugar de recapturados. Na maioria das vezes é usado como aproximação à distribuição hipergeométrica. Nesta situação a distribuição de probabilidade de m_2 é

$$P(m_2|N, n_1, n_2) = \binom{n_2}{m_2} \left(\frac{n_1}{N} \right)^{m_2} \left(1 - \frac{n_1}{N} \right)^{n_2 - m_2}, \quad m_2=0, 1, \dots, n_2 \quad (2.6)$$

A esperança e variância de m_2 são respectivamente,

$$E(m_2) = \frac{n_1 n_2}{N} \quad \text{e} \quad V(m_2) = n_2 \frac{n_1}{N} \left(1 - \frac{n_1}{N} \right)$$

Também neste caso, o EMV é o estimador de Petersen, tendo esperança infinita pois m_2 tem probabilidade positiva de assumir o valor zero e sua variância não existe. Para resolver este problema, Bailey (1951) propôs o seguinte estimador modificado

$$\hat{N}_b = \left[\frac{n_1(n_2+1)}{(m_2+1)} \right], \quad (2.7)$$

onde m_2 tem distribuição binomial com parâmetros $n_2, n_1/N$. A estimação da variância de \hat{N}_b para amostras grandes é

$$\hat{V}(\hat{N}_b) = \frac{n_1^2(n_2+1)(n_1-m_2)}{(m_2+1)^2(m_2+2)}. \quad (2.8)$$

Se n_2/N for suficientemente pequeno pode-se esquecer a não reposição e considerar \hat{N}_b no lugar de \hat{N}_{ch} pois na prática apresentam diferenças muito pequenas.

2.3 Estimador de Petersen na Amostragem Inversa Simples

Uma outra sugestão para eliminar o problema da probabilidade positiva de m_2 ser igual a zero é utilizar o esquema de amostragem inversa. Neste esquema, na segunda ocasião as unidades podem ser seleccionadas uma a uma sem reposição até que um número pre-fixado m_2 (≥ 1) de unidades marcadas sejam obtidas. Neste procedimento, o número de unidades necessárias na segunda amostra para obter as m_2 marcadas, n_2 , é uma variável aleatória tendo valores inteiros positivos e cuja distribuição é dada por

$$P(n_2|N, n_1, m_2) = \frac{\binom{n_1}{m_2-1} \binom{N-n_1}{n_2-m_2}}{\binom{N}{n_2-1}} \frac{n_1 - (m_2-1)}{N - (n_2-1)}, \quad m_2 \leq n_2 \leq (N - n_1) + m_2 \quad (2.9)$$

A distribuição acima é conhecida como hipergeométrica negativa, hipergeométrica inversa ou hipergeométrica de tempo de espera (Johnson & Kotz (1989)). A esperança e variância de n_2 são respectivamente,

$$E(n_2) = \frac{m_2(N+1)}{(n_1+1)} \quad \text{e} \quad V(n_2) = \frac{m_2(N+1)(N-n_1)(n_1-m_2+1)}{(n_1+1)^2(n_1+2)}. \quad (2.10)$$

Para achar o EMV de N consideram-se os quocientes

$$\frac{P(n_2|N, n_1, m_2)}{P(n_2|N-1, n_1, m_2)} \geq 1 \quad \text{e} \quad \frac{P(n_2|N, n_1, m_2)}{P(n_2|N+1, n_1, m_2)} \geq 1.$$

Da primeira desigualdade obtém-se

$$\frac{\binom{n_1}{m_2-1} \binom{N-n_1}{n_2-m_2}}{\binom{N}{n_2-1}} \frac{n_1 - (m_2-1)}{N - (n_2-1)} \geq \frac{\binom{n_1}{m_2-1} \binom{N-1-n_1}{n_2-m_2}}{\binom{N-1}{n_2-1}} \frac{n_1 - (m_2-1)}{N-1 - (n_2-1)}$$

e portanto,

$$N \leq \frac{n_1 n_2}{m_2}.$$

Da segunda desigualdade obtém-se

$$\frac{\binom{n_1}{m_2-1} \binom{N-n_1}{n_2-m_2}}{\binom{N}{n_2-1}} \frac{n_1 - (m_2-1)}{N - (n_2-1)} \geq \frac{\binom{n_1}{m_2-1} \binom{N+1-n_1}{n_2-m_2}}{\binom{N+1}{n_2-1}} \frac{n_1 - (m_2-1)}{N+1 - (n_2-1)}$$

e portanto,

$$N \geq \frac{n_1 n_2}{m_2} - 1.$$

Assim o EMV de N , satisfaz

$$\frac{n_1 n_2}{m_2} - 1 \leq \hat{N} \leq \frac{n_1 n_2}{m_2},$$

obtendo-se o estimador

$$\hat{N}_a = \left\lceil \frac{n_1 n_2}{m_2} \right\rceil,$$

que embora apresente-se igual ao de Petersen na fórmula, difere nos fatos de $m_2 (\geq 1)$ ser prefixado e $n_2 (\geq m_2)$ ser uma variável aleatória. Portanto pode-se mostrar que $P(\hat{N}_a < \infty) = 1$ ou $P(\hat{N}_a = \infty) = 0$ e \hat{N}_a tem momentos finitos de qualquer ordem positivo. Agora, por (2.10), a esperança e a variância de \hat{N}_a são respectivamente,

$$E(\hat{N}_a) = \frac{n_1}{(n_1+1)} (N+1) \quad \text{e} \quad V(\hat{N}_a) = \frac{n_1^2 (N+1) (N-n_1) (n_1-m_2+1)}{m_2 (n_1+1)^2 (n_1+2)}$$

Um estimador modificado, \hat{N}_c , foi proposto por Chapman (1952) o qual é não viciado para N ,

$$\hat{N}_c = \frac{n_2 (n_1+1)}{m_2} - 1 \tag{2.11}$$

e cuja variancia é a seguinte

$$V(\hat{N}_c) = \frac{(N+1) (N-n_1) (n_1-m_2+1)}{m_2 (n_1+2)}.$$

2.4 Distribuição do Estimador de Petersen para Amostragem Simples

A distribuição assintótica abaixo é relacionada com o estimador de Petersen tanto para o esquema de amostragem direta como para o esquema de amostragem inversa.

Assumindo que N é grande, e que para $\alpha, \beta > 0$, tem-se

$$n_1/N \rightarrow \alpha, \quad n_2/N \rightarrow \beta, \text{ com } 0 < \alpha, \beta \leq 1. \quad (2.12)$$

Se

$$\frac{m_2 - n_2\alpha}{(N\alpha(1-\alpha)\beta(1-\beta))^{1/2}} \rightarrow y,$$

então pela aproximação normal à hipergeométrica,

$$P(m_2|N, n_1, n_2) \rightarrow \frac{1}{(N\alpha\beta(1-\alpha)(1-\beta))^{1/2}} \frac{1}{(2\pi)^{1/2}} \exp(-y^2/2)$$

isto é,

$$m_2 \sim N(n_2\alpha, N\alpha\beta(1-\alpha)(1-\beta)).$$

Portanto,

$$(m_2 - n_2\alpha) \sim N(0, N\alpha\beta(1-\alpha)(1-\beta))$$

e também

$$N^{-1/2} (m_2 - n_2\alpha) \sim N(0, \alpha\beta(1-\alpha)(1-\beta)).$$

Como $\hat{N}_p = \frac{n_1 n_2}{m_2} = g(m_2)$, pelo teorema de Slutsky:

$$N^{-1/2} (g(m_2) - g(n_2\alpha)) \sim N(0, \alpha\beta(1-\alpha)(1-\beta)[g'(n_2\alpha)]^2)$$

onde por (2.12) $g'(m_2)|_{m_2 = n_2\alpha} = -\frac{1}{\alpha\beta}$

e em consequencia,

$$N^{-1/2} \left(\frac{n_1 n_2}{m_2} - \frac{n_1 n_2}{n_2 \alpha} \right) \sim N(0, \alpha\beta(1-\alpha)(1-\beta)(-1/\alpha\beta)^2)$$

ou

$$N^{-1/2} (\hat{N}_p - N) \sim N(0, (1-\alpha)(1-\beta)/\alpha\beta). \quad (2.13)$$

Pode-se observar que $(1-\alpha)(1-\beta)/\alpha\beta$ é mínimo (para um γ dado tal que $\alpha+\beta=\gamma>0$) quando $\alpha=\beta=\gamma/2$, isto é, no modelo de duas amostras, idealmente, os tamanhos de amostra n_1, n_2 devem estar muito próximos quando não iguais.

A mesma distribuição assintótica definida em (2.13) pertence ao estimador \hat{N}_p

quando $m_2 \rightarrow N\alpha\beta$ (isto é, $m/N\alpha\beta \rightarrow 1$) de forma que o valor esperado de n_2 estará próximo de $N\beta$ e comparável a (2.12).

2.6 Metodologia Bootstrap nos experimentos de C-R para Duas Amostras

Nesta seção são apresentados os resultados da aplicação da metodologia bootstrap aos experimentos de C-R de amostragem simples, como visto no capítulo 1, com o objetivo principal de comparar as estimativas do erro padrão da estimação do tamanho da população, N , assim como os intervalos de confiança para o mesmo, baseados no estimador de Petersen modificado, e os baseados em bootstrap.

A aplicação consiste primeiro, na simulação de experimentos de C-R para duas amostras, quando a amostragem é direta e sem reposição, para uma população hipotética de tamanho N , e amostras dos tamanhos n_1 e n_2 respectivamente.

A construção do experimento de C-R de amostragem simples consiste nos seguinte passos:

1. Atribuir valor 0 a todos os N indivíduos da população.
2. Selecionar uma amostra aleatória sem reposição de tamanho n_1 e atribuir valor 1 aos indivíduos selecionados.
3. Selecionar n_2 indivíduos da mesma população em 1, por amostragem aleatória sem reposição, atribuindo valor 1 aos selecionados.
4. Contagem dos indivíduos selecionados na primeira amostra e na segunda, isto é, número de indivíduos com valor 1 nos passos 2 e 3.
5. O número total de indivíduos marcados na população após as duas seleções (ou antes da terceira amostra) é calculado através de $M_3 = n_1 + n_2 - m_2$.

Foram considerados dois tamanhos de população: 100 e 1000 indivíduos. Para ambos tamanhos, sete experimentos foram construídos. Com $N=100$, os tamanhos amostrais considerados variaram de 15 a 75 na primeira amostra e de 15 a 50 na segunda (Tabela 1). Esta variação dos tamanhos das amostras foi feita a partir do questionamento que surge na prática: as amostras devem ter tamanhos similares, ou então, a captura deve ser maior que a recaptura ou vice versa, a fim de produzir “melhores” estimativas de N . Para o experimento 1, os tamanhos são pequenos e muito próximos entre si: 15 e 16; no experimento 2 estes tamanhos são maiores que o anterior e também próximos entre si: 25 e 24; para o experimento 3, mantém-se o tamanho da primeira amostra do experimento 1, fazendo com que a segunda seja da ordem dos tamanhos do experimento 2, resultando estes tamanhos: 15 e 25; no experimento 4 os

tamanhos do experimento 3 invertem-se: 25 e 15. Nos experimentos 5, 6 e 7, a idéia foi dar a uma das amostras um tamanho consideravelmente grande, permanecendo a outra com o tamanho da ordem do experimento 1, assim, o experimento 5 tem os tamanhos: 75 e 16, o experimento 6: 50 e 16, finalmente o experimento 7: 15 e 50 respectivamente.

O conjunto de resultados da simulação dos experimentos, apresentados na Tabela 1 são fornecidos pelo programa PROG1.SAS realizado utilizando o software SAS (Apêndice).

TABELA 1. Simulação de experimentos de C-R com duas amostras para o tamanho de população 100.

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
n_1	15	25	15	25	75	50	15
n_2	16	24	25	15	16	16	50
m_2	3	5	2	5	12	5	7
M_3	28	44	38	35	79	61	58

As estatísticas acima são utilizadas para estimar o tamanho da população para cada experimento através do estimador de Petersen modificado (ou de Chapman), dado em (2.4)

$$\hat{N}_{ch} = \left[\frac{(n_1+1)(n_2+1)}{(m_2+1)} - 1 \right].$$

Foi visto na seção 2.2 que $\hat{N}_{ch} \underset{\alpha}{\sim} N(N, V(\hat{N}_{ch}))$ com $V(\hat{N}_{ch})$ estimada por

$$v(\hat{N}_{ch}) = \frac{(n_1+1)(n_2+1)(n_1-m_2)(n_2-m_2)}{(m_2+1)^2(m_2+2)}.$$

Portanto, o intervalo de confiança assintótico é calculado por

$$\text{I.C.} \left[\hat{N}_{ch} \pm Z_{\alpha} [v(\hat{N}_{ch})]^{1/2} \right] = 1 - \alpha .$$

Também, por (2.13) a distribuição assintótica de \hat{N}_p padronizado é a que segue,

$$N^{-1/2} (\hat{N}_p - N) \sim N(0, (1-\alpha)(1-\beta)/\alpha\beta)$$

logo, é encontrado o valor para $v(\hat{N}_p)$ substituindo N pela estimativa obtida \hat{N}_{ch} , assim como um intervalo de confiança assintótico calculado por

$$\text{I.C.} \left[\hat{N}_{ch} \pm Z_\alpha [v(\hat{N}_p)]^{1/2} \right] = 1 - \alpha.$$

Na Tabela 2 são apresentados para cada um dos experimentos, os resultados da estimação do tamanho da população, \hat{N}_{ch} , as estimativas das variâncias assintóticas, $v(\hat{N}_{ch})$ e $v(\hat{N}_p)$ assim como os intervalos de confiança assintóticos utilizando estas variâncias.

TABELA 2. Estimativa de Petersen modificado, erro padrão e Intervalo de confiança assintótico para $N=100$.

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
\hat{N}_{ch}	67	107	137	68	98	143	101
$[v(\hat{N}_{ch})]^{1/2}$	22,86	31,11	58,43	17,99	11,65	41,05	21,97
$[v(\hat{N}_p)]^{1/2}$	27,20	34,80	70,65	20,33	12,41	45,95	24,30

Intervalo de confiança de aproximação normal

L.inf. L.sup. 22 112 46 168 22 252 33 103 75 121 63 223 58 144

Intervalo de confiança assintótico (baseado em $v(\hat{N}_p)$)

L.inf. L.sup. 14 120 39 175 0 275 28 108 74 122 53 233 53 149

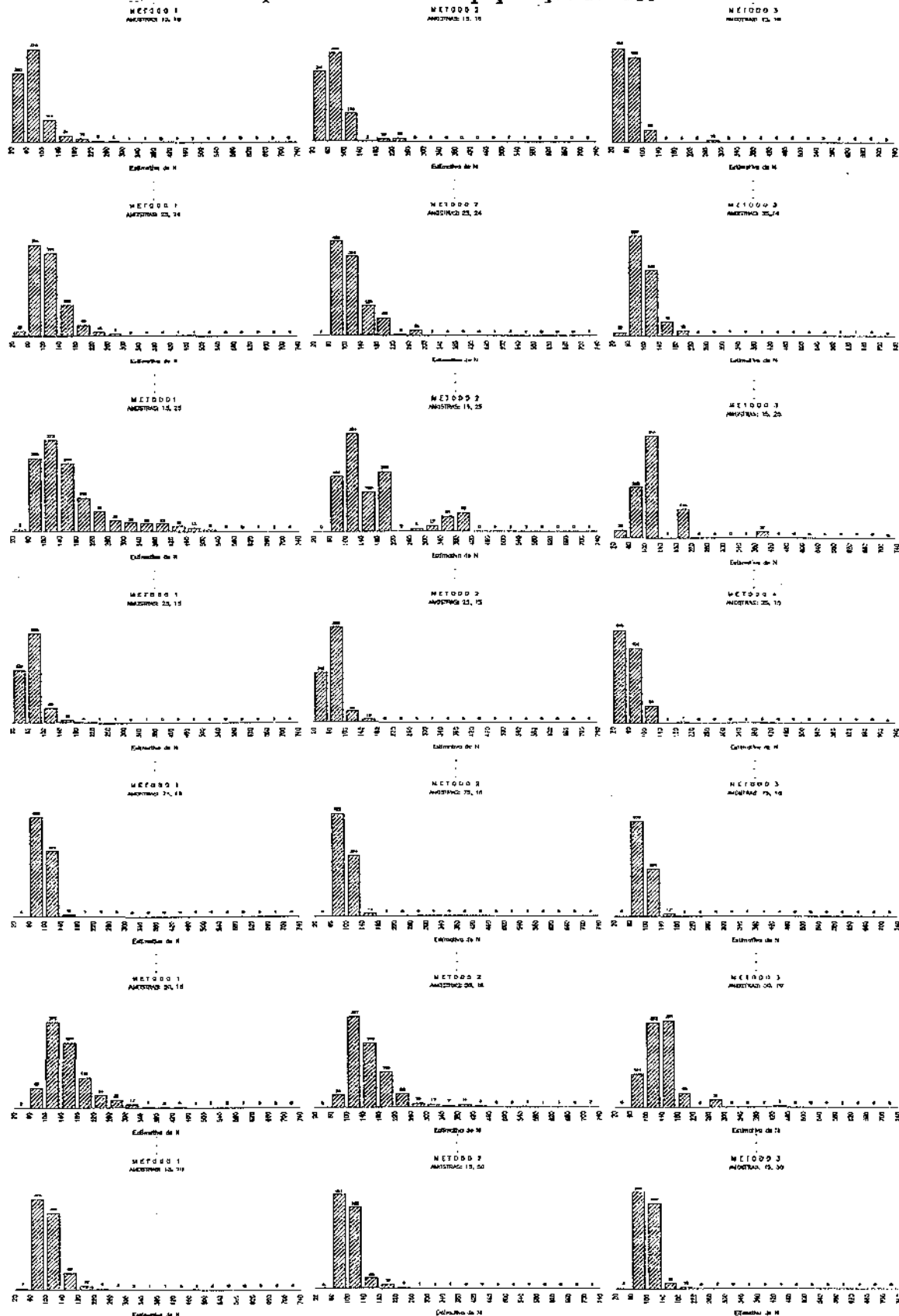
Uma alternativa para a avaliação das estimativas do tamanho da população obtidas pelos métodos de C-R é a metodologia bootstrap. Os métodos bootstrap são utilizados na estimação da variância das estimativas assim como na construção de intervalos de confiança para o parâmetro (Seção 3 do Capítulo 1). Como apresentados nessa seção, três métodos bootstrap podem ser aplicados aos experimentos de C-R de amostragem simples e direta. Estes métodos são denominados de método 1 e método 2 quando as análises são não condicionais aos tamanhos de amostra observados e de método 3 quando a análise é condicional a estes tamanhos. Considerando o vetor

$\eta=(n_{10},n_{01},n_{11},n_{00})$ indicando as frequências das histórias de captura na primeira e segunda amostragem no total da população $N=n_{10}+n_{01}+n_{11}+n_{00}$, a distribuição de probabilidade associada a ele é multinomial com parâmetros $(\hat{N}_{ch}, p_{10}, p_{01}, p_{11}, p_{00})$ ou com parâmetros $(M_3, p_{10}, p_{01}, p_{11})$, que correspondem respectivamente aos métodos 1 e 2. No caso do método 3 a distribuição associada a m_2 é hipergeométrica (\hat{N}_{ch}, n_1, n_2) .

O número de replicações bootstrap, B , considerado nos experimentos é de 1000, desde que 100 replicações resultou em estimativas bastante grandes dos erros e 2000 não indicou um ganho considerável nas estimativas dos erros padrão quando comparado com as obtidas com 1000. Este valor de $B=1000$ concorda com a sugestão dada em Buckland(1984).

A figura seguinte mostra os histogramas das 1000 replicações bootstrap das estimativas de N para cada experimento e para cada método analisado. Na figura, os três primeiros gráficos correspondem aos três métodos bootstrap aplicados ao primeiro experimento, os gráficos restantes correspondem aos experimentos 2 a 7 respectivamente. Os histogramas mostram a assimetria da distribuição das estimativas nos três métodos, sendo que comparativamente o método 1 apresenta maior proporção de valores extremos à direita.

Histogramas das replicações bootstrap por método aplicado para o tamanho de população $N=100$



Na Tabela 3, são apresentadas as estimativas bootstrap do erro padrão de \hat{N}_{ch} , calculadas a partir de (1.1),

$$\hat{V}(\hat{N}) = [\sum \hat{N}_j^{*2} - (\sum \hat{N}_j^*)^2 / B] / (B - 1).$$

Também são apresentados os intervalos de confiança para N , baseados nos percentis bootstrap e os intervalos do percentil de vício corrigido para os três métodos bootstrap (como visto na seção 1.3).

Observa-se que as estimativas do erro padrão resultam próximas e comparáveis entre si nos três métodos bootstrap, mas estas em conjunto, discrepam do valor da estimativa obtida pela fórmula assintótica. Só no experimento 5, de tamanhos amostrais 75 e 16 respectivamente, estas estimativas resultam próximas devido ao fato da primeira amostra ser irrealisticamente grande.

TABELA 3. Estimativa bootstrap do erro padrão e Intervalos de confiança.

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
Estimativa bootstrap do erro padrão de \hat{N}_{ch}							
Método 1	39,17	49,15	93,26	29,01	13,64	64,51	31,20
Método 2	37,25	49,32	82,64	27,44	13,24	70,28	29,42
Método 3	31,77	39,51	67,76	22,29	13,12	53,50	24,45
Intervalo de Confiança baseado no percentil bootstrap							
Método 1	36 195	64 249	66 441	42 149	79 132	86 303	69 189
Método 2	41 220	70 191	78 383	48 129	82 133	92 349	72 187
Método 3	37 135	64 215	58 415	40 137	79 128	85 288	67 162
Intervalo de Confiança do percentil de vício corrigido							
Método 1	34 175	63 230	63 399	39 139	80 135	82 299	69 190
Método 2	43 223	69 191	71 363	47 121	82 135	92 349	72 186
Método 3	37 135	64 215	68 415	40 137	75 116	85 288	67 162

A estimação de N , pontual e por intervalo, é realizada utilizando o programa PROG2.SAS (Apêndice). Este programa é o que deve ser usado na prática quando dados reais são analisados fornecendo os valores amostrais n_1 , n_2 , m_2 e M_3 , em experimentos de C-R de amostragem simples e direta.

Na seguinte tabela são mostrados os comprimentos dos intervalos obtidos pelos três métodos e para cada experimento.

TABELA 4. Comprimento dos Intervalos de confiança.

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
Intervalos assintóticos baseados em							
$v(\hat{N}_{ch})$	90	122	230	70	46	160	86
$v(\hat{N}_p)$	106	136	275	80	48	180	96
Intervalos baseados no percentil bootstrap							
Método 1	159	185	375	107	53	217	120
Método 2	179	121	305	81	51	257	115
Método 3	98	151	357	97	49	203	95
Intervalos do percentil de vício corrigido							
Método 1	141	167	337	100	55	217	121
Método 2	180	122	292	74	53	257	114
Método 3	98	151	347	97	41	203	95

A representação gráfica dos intervalos permite comparar os intervalos baseados na distribuição assintótica do estimador de N , com os obtidos pelo método bootstrap. Como mostrado para o experimento 1 a maneira de exemplo, tem-se que o primeiro intervalo (a) representa o intervalo de aproximação normal baseado na variância de \hat{N}_{ch} , é o de menor comprimento embora assuma uma simetria da distribuição das estimativas a qual não é verificada pela análise do histograma acima. O intervalo (b) representa o intervalo assintótico baseado na variância de \hat{N}_p . Os três intervalos seguintes (c, d, e) são os baseados nos percentis bootstrap obtidos pelos métodos 1, 2 e 3 respectivamente, e os restantes (f, g, h) correspondem aos intervalos do percentil de vício corrigido dos

métodos 1, 2 e 3 respectivamente. O intervalo obtido assumindo a distribuição hipergeométrica (e) é o de menor comprimento e este não é alterado ao ser corrigido pelo vício.

a.	22	—————	112
b.	14	—————	120
c.	36	—————	195
d.	41	—————	220
e.	37	—————	135
f.	34	—————	175
g.	43	—————	228
h.	37	—————	135

Análises similares nos experimentos 2 a 7, revelam que em todos os métodos e todos os experimentos, os intervalos do percentil de vício corrigido resultam pelo menos de igual comprimento, quando não menores, comparados aos obtidos sem a correção. Os intervalos obtidos pelo método 3 são os de menor mudança quando é feita a correção pelo vício. Em geral, o método 2 fornece os intervalos de menor comprimento para os experimentos 2, 3 e 4, isto é, a distribuição multinomial ($M_{3,P_{10},P_{01},P_{11}}$), fornece os intervalos de menor comprimento para os experimentos com os tamanhos das amostras 25 e 24, 15 e 25, 25 e 15, enquanto para os experimentos 5, 6 e 7 o método fornecendo os intervalos menores é o 3, isto é, para os experimentos com tamanhos das amostras 75 e 16, 50 e 16, 15 e 50, a distribuição que gera as réplicações bootstrap fornecendo os intervalos de menor comprimento é a hipergeométrica (\hat{N}_{ch,n_1,n_2}).

Um estudo de simulação pode ser realizado para estimar o nível de confiança real dos intervalos baseados nos percentis bootstrap. Este estudo consiste na obtenção de 100 intervalos de 95% de confiança nominal, baseados em 1000 réplicações bootstrap cada um deles. Em cada intervalo verifica-se se o valor do parâmetro $N=100$, está ou não contido no intervalo. Finalmente, o novo nível de confiança é obtido da proporção do número de intervalos, dentre os 100, que contêm o valor do parâmetro.

Na Tabela 5 são apresentados os níveis de confiança estimados por este procedimento, para cada um dos sete experimentos e para os três métodos aplicados. Estes resultados são fornecidos pelo programa PROG3.SAS.

Os resultados da tabela mostram que nenhum dos métodos estima uniformemente os níveis de confiança em todos os experimentos, mas, os três métodos fornecem uma boa estimativa do nível de confiança no caso do experimento 2 o qual trata-se de tamanhos das amostras muito próximos entre si e relativamente grandes (0,25N). Também, tem-se que estes níveis estimados são menores do que 95% em todos os métodos para o experimento 6 (tamanhos das amostras de 50 e 16 respectivamente).

TABELA 5. Estimação do nível de confiança real dos intervalos do percentil de 95% de confiança nominal.

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
Método 1	91	96	96	90	96	89	94
Método 2	92	98	91	89	92	89	91
Método 3	81	98	95	90	89	88	94

A seguir é apresentada uma segunda aplicação da metodologia bootstrap aos experimentos de C-R de amostragem simples e direta, para um outro tamanho de população. O valor de $N=1000$ tem sido escolhido com a finalidade de fazer uma comparação com a aplicação anterior onde $N=100$ e pelo fato deste tamanho ser mais possível de se achar na prática. Foram seguidos os procedimentos já descritos nesta seção, tanto para a simulação dos experimentos quanto para a aplicação da metodologia bootstrap a estes experimentos. Simulou-se sete experimentos de C-R de amostragem simples e direta com os tamanhos amostrais variando de 150 a 750 a primeira amostra e de 150 a 500 a segunda.

Na seguinte tabela são apresentadas as estatísticas relevantes destes experimentos, tais como m_2 , o número de indivíduos marcados capturados na segunda amostra, e M_3 , o número total de indivíduos diferentes marcados durante o experimento (ou antes da captura da terceira amostra).

TABELA 6. Simulação de experimentos de C-R com duas amostras para o tamanho de população 1000.

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
n_1	150	250	150	250	750	500	150
n_2	160	240	250	150	160	160	500
m_2	26	65	34	47	113	88	82
M_3	284	425	366	353	797	572	568

As estatísticas resultantes da simulação dos experimentos apresentadas na tabela anterior, foram utilizadas no cálculo da estimativa do tamanho da população mediante o estimador modificado de Chapman mostrado em (2.4). Também foram calculadas as estimativas das variâncias assintóticas de \hat{N}_{ch} e de \hat{N}_p mostradas em (2.5) e (2.13), logo elas são utilizadas na obtenção de intervalos assintóticos de 95% de confiança para N.

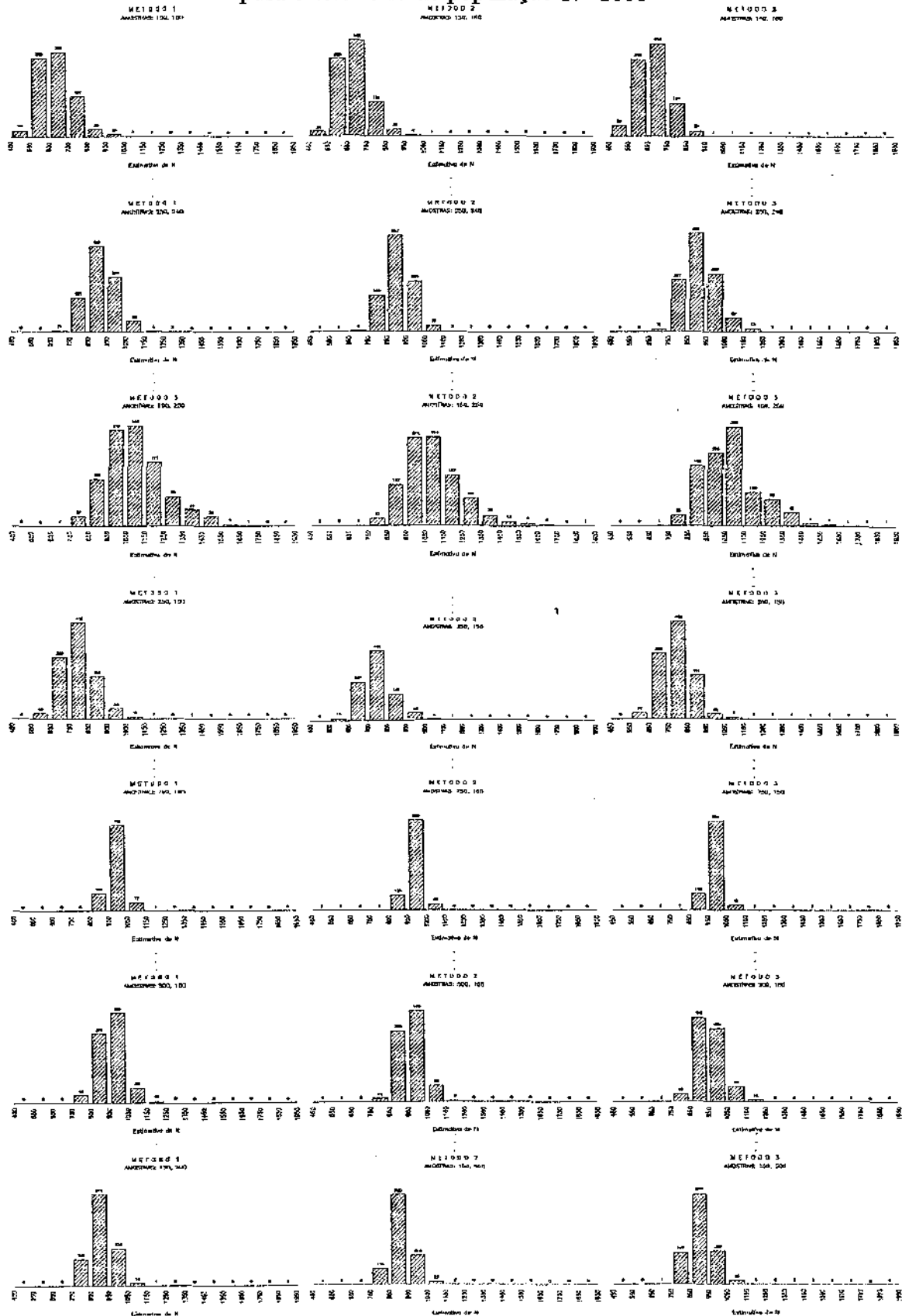
TABELA 7. Estimativa de Petersen modificado, erro padrão e Intervalos de confiança assintóticos.

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
\hat{N}_{ch}	899	915	1081	788	1059	905	910
$[v(\hat{N}_{ch})]^{1/2}$	140,57	81,85	146,62	83,70	49,18	57,89	60,91
$[v(\hat{N}_p)]^{1/2}$	143,99	82,74	149,34	84,93	49,51	58,42	61,49
Intervalo de Confiança de aproximação normal							
Lim. inf.	623	755	794	624	963	792	791
Lim. sup.	1175	1075	1368	952	1155	1018	1029
Intervalo de Confiança assintótico (baseado em $v(\hat{N}_p)$)							
Lim. inf.	617	753	788	622	962	791	789
Lim. sup.	1181	1077	1374	954	1156	1019	1031

parâmetro assim como na obtenção de intervalos de confiança é aplicada a estes experimentos simulados. O número de replicações bootstrap, de forma similar à aplicação anterior foi de $B=1000$.

A figura seguinte mostra os histogramas das 1000 replicações bootstrap das estimativas de N para cada experimento e para cada método aplicado. Os três primeiros gráficos correspondem ao experimento 1 e os três métodos bootstrap aplicados. Similarmente, os gráficos seguintes correspondem aos experimentos 2 a 7. A assimetria da distribuição das estimativas de N é menos evidente do que no caso $N=100$, valores extremos à direita também estão presentes, embora as frequências sejam muito pequenas.

Histogramas das replicações bootstrap por método aplicado para o tamanho de população $N=1000$



Na Tabela 8 são apresentados os resultados bootstrap para os experimentos simulados, tais são as estimativas bootstrap do erro padrão de \hat{N}_{ch} , calculadas de (1.1) e os intervalos de confiança para N , baseados nos percentis bootstrap e os intervalos do percentil de vício corrigido como descritos no capítulo 1.

TABELA 8. Estimativa bootstrap do erro padrão e intervalos de confiança.

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
Estimativa bootstrap do erro padrão de \hat{N}_{ch}							
Método 1	157,09	85,39	156,76	86,93	49,97	59,94	61,57
Método 2	152,87	80,94	151,36	81,92	47,40	52,30	58,58
Método 3	141,37	85,57	150,74	85,12	49,17	60,25	62,53
Intervalo de Confiança baseado no percentil bootstrap							
Mét. 1 L.inf.	680	780	834	646	969	797	807
L.sup.	1303	1118	1461	986	1171	1036	1043
Mét. 2 L.inf.	698	790	872	664	984	816	809
L.sup.	1274	1097	1479	980	1166	1018	1049
Mét. 3 L.inf.	674	774	841	641	966	805	803
L.sup.	1214	1098	1402	970	1161	1046	1049
Intervalo de Confiança do percentil de vício corrigido							
Mét. 1 L.inf.	679	778	855	649	969	790	801
L.sup.	1298	1116	1493	1002	1171	1031	1030
Mét. 2 L.inf.	696	790	870	657	984	812	818
L.sup.	1273	1105	1472	965	1166	1010	1059
Mét. 3 L.inf.	674	764	860	641	958	797	803
L.sup.	1214	1098	1456	946	1150	1033	1049

Observa-se que para todos os experimentos e todos os métodos, os valores obtidos para as estimativas bootstrap do erro padrão quando comparadas com o erro padrão da estimativa de Chapman, melhoram significativamente daquelas obtidas tamanho da população 100, isto é, a análise de um tamanho de população maior faz com que a

metodologia bootstrap forneça estimativas do erro padrão mais similares à de Chapman.

Na tabela seguinte são apresentados os comprimentos dos intervalos de confiança obtidos pelos três métodos bootstrap e para todos os experimentos.

TABELA 9. Comprimento dos Intervalos de confiança.

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
Intervalos assintóticos baseados em							
$[v(\hat{N}_{ch})]^{1/2}$	552	320	574	328	192	226	238
$[v(\hat{N}_p)]^{1/2}$	564	324	586	332	194	230	242
Intervalos baseados no percentil bootstrap							
Método 1	623	338	627	340	202	239	236
Método 2	576	307	607	316	182	202	240
Método 3	540	324	561	329	195	241	246
Intervalos do percentil de vício corrigido							
Método 1	619	338	638	353	202	241	229
Método 2	577	315	602	308	182	198	241
Método 3	540	334	596	305	192	236	246

Uma representação gráfica dos intervalos de confiança obtidos pode ser feita a fim de comparação. Assim, para o experimento 2, os dois primeiros gráficos, (a, b) representam os intervalos assintóticos baseados nas variâncias de \hat{N}_{ch} e de \hat{N}_p , respectivamente, os três gráficos seguintes (c, d, e), representam os intervalos baseados no percentil bootstrap e os restantes, (f, g, h) os intervalos do percentil de vício corrigido.

a.	755	1075
b.	753	1077
c.	780	1118
d.	790	1097
e.	774	1098
f.	778	1116
g.	790	1105
h.	764	1098

Na Tabela 10 são apresentadas as estimações do nível de confiança real dos intervalos baseados no percentil bootstrap. Estas estimações foram obtidas simulando 100 intervalos, e calculando a proporção deles contendo o valor de N . O nível de confiança nominal foi 95%.

TABELA 10. Estimação do nível de confiança real dos intervalos do percentil de 95% de confiança nominal.

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
Método 1	95	93	96	92	95	95	93
Método 2	94	93	96	91	93	92	91
Método 3	95	93	96	91	96	94	94

Em geral as estimativas do nível de confiança real foram mais próximos do valor nominal do que as obtidas para o tamanho de população $N=100$. Os três métodos estimaram homogeneamente o nível confiança no caso do experimento 3 com os tamanhos das amostras 150 e 250.

CAPÍTULO 3

ESTIMAÇÃO DO TAMANHO DA POPULAÇÃO POR AMOSTRAGEM MÚLTIPLA

3.1 Introdução

A estimação do tamanho de população N por C-R pode-se basear em amostragem múltipla, isto é, em s (≥ 2) amostras ou ocasiões de amostragem. Em cada ocasião registra-se o número de indivíduos marcados, capturados na amostra, e todos, marcados e não marcados ganham uma nova marca antes de serem retornados à população. As marcas podem ser distinguíveis de uma ocasião para outra ou não. Também em amostragem múltipla a amostragem pode ser direta (em cada ocasião o tamanho da amostra é fixo), ou inversa (em cada ocasião o número de indivíduos marcados é fixado e o tamanho da amostra é aleatório).

Nas seguintes seções estudam-se os métodos de estimação do tamanho da população baseados em amostragem múltipla: na seção 3.2 quando a amostragem é direta e na seção 3.3 quando a amostragem é inversa. Na seção 3.4 estuda-se o estimador de Petersen (ponderado) multiamostrai. Na seção 3.5 estuda-se a distribuição assintótica do estimador multiamostrai. Por último na seção 3.6 apresentam-se resultados da metodologia bootstrap quando aplicados aos métodos de C-R de amostragem múltipla.

3.2 Estimador Multiamostrai para Amostragem Direta

O procedimento de amostragem direta múltipla, fixa os tamanhos das s amostras n_1, \dots, n_s , respectivamente. As amostras são selecionadas sem reposição e a partir da segunda, registra-se o número de indivíduos marcados capturados na amostra, (m_i , $i=2, \dots, s$, $m_1=0$) e os não marcados ($u_i=n_i-m_i$). Todos os indivíduos da amostra recebem uma nova marca antes de retornar à população.

Seja M_i o número de indivíduos marcados na população, exatamente antes da i -ésima amostra ser capturada, então

$$M_i = \sum_{j < i} u_j, \quad i=1, \dots, s+1.$$

Note-se que $M_1=m_1=0$, $M_2=u_1=n_1$ e para $k \geq 2$,

$$M_{k+1} = M_k + n_k - m_k. \quad (3.1)$$

Note-se também que $M_{s+1} = \sum_{j=1}^s u_j$ é o número de indivíduos distintos observados durante o experimento.

Então,

$$P(m_2, \dots, m_s | n_1, \dots, n_s, N) = \prod_{i=2}^s \frac{\binom{M_i}{m_i} \binom{N-M_i}{n_i-m_i}}{\binom{N}{n_i}}. \quad (3.2)$$

Para obter o EMV de N estabelecem-se

$$P(m_2, \dots, m_s | n_1, \dots, n_s, N) \geq P(m_2, \dots, m_s | n_1, \dots, n_s, N-1),$$

$$P(m_2, \dots, m_s | n_1, \dots, n_s, N) \geq P(m_2, \dots, m_s | n_1, \dots, n_s, N+1).$$

Da primeira desigualdade, obtem-se

$$\prod_{i=2}^s \frac{\binom{M_i}{m_i} \binom{N-M_i}{n_i-m_i}}{\binom{N}{n_i}} \geq \prod_{i=2}^s \frac{\binom{M_i}{m_i} \binom{N-M_i-1}{n_i-m_i}}{\binom{N-1}{n_i}}.$$

Simplificando-a obtem-se

$$\prod_{i=2}^s \frac{(N-M_i)/(N-M_i-n_i+m_i)}{N/(N-n_i)} = \prod_{i=2}^s \frac{(N-M_i)(N-n_i)}{N(N-M_i-n_i+m_i)} \geq 1$$

Utilizando (3.1) na expressão acima, tem-se

$$\prod_{i=2}^s \frac{(N-M_i)(N-n_i)}{N(N-M_{i+1})} \geq 1$$

ou

$$\frac{1}{N^{s-1}} \frac{(N-M_2)(N-n_2)}{N-M_3} \frac{(N-M_3)(N-n_3)}{N-M_4} \dots \frac{(N-M_s)(N-n_s)}{N-M_{s+1}} \geq 1$$

Como $M_2=n_1$, então

$$\frac{1}{N^{s-1}} \frac{1}{N-M_{s+1}} \prod_{i=1}^s (N-n_i) \geq 1$$

Portanto, o EMV multiamostrai de N, \hat{N}_m satisfaz

$$\prod_{i=1}^s \left(1 - \frac{n_i}{N}\right) \geq 1 - \frac{M_{s+1}}{N}$$

ou

$$1 - \frac{M_{s+1}}{\hat{N}_m} \leq \left(1 - \frac{n_1}{\hat{N}_m}\right) \dots \left(1 - \frac{n_s}{\hat{N}_m}\right). \quad (3.3)$$

Agora, da segunda desigualdade tem-se

$$\prod_{i=2}^s \frac{\binom{M_i}{m_i} \binom{N-M_i}{n_i-m_i}}{\binom{N}{n_i}} \geq \prod_{i=2}^s \frac{\binom{M_i}{m_i} \binom{N-M_i+1}{n_i-m_i}}{\binom{N+1}{n_i}}$$

simplificando-a, obtém-se

$$1 \geq \prod_{i=2}^s \frac{(N+1-M_i)/(N+1-M_i-n_i+m_i)}{(N+1)/(N+1-n_i)} = \prod_{i=2}^s \frac{(N+1-M_i)(N+1-n_i)}{(N+1)(N+1-M_i-n_i+m_i)}$$

Utilizando (3.1) na expressão acima, tem-se

$$1 \geq \prod_{i=2}^s \frac{(N+1-M_i)(N+1-n_i)}{(N+1)(N+1-M_{i+1})}$$

ou

$$1 \geq \frac{1}{(N+1)^{s+1}} \frac{1}{N+1-M_{s+1}} \prod_{i=2}^s (N+1-M_i)(N+1-n_i)$$

Como $M_2 = n_1$, então,

$$(N+1)^{s+1}(N+1-M_{s+1}) \frac{1}{(N+1)^s} \geq \prod_{i=1}^s \frac{(N+1-n_i)}{N+1}.$$

Portanto, o EMV multiamostrai de N, \hat{N}_m também satisfaz

$$1 - \frac{M_{s+1}}{\hat{N}_m+1} \geq \left(1 - \frac{n_1}{\hat{N}_m+1}\right) \dots \left(1 - \frac{n_s}{\hat{N}_m+1}\right). \quad (3.4)$$

O estimador \hat{N}_m para $s > 2$, isto é, quando a amostragem é múltipla, é encontrado através da solução das equações não lineares (3.3) e (3.4) simultaneamente. Contudo, Darroch (1958) mostrou que o EMV de N é a maior raiz do polinômio dado por

$$\left(1 - \frac{M_{s+1}}{\hat{N}_m}\right) = \prod_{i=1}^s \left(1 - \frac{n_i}{\hat{N}_m}\right)$$

Leite et al. (1988) apresentam uma expressão mais explícita para \hat{N}_m dada por

$$\hat{N}_m = \begin{cases} M_{s+1} + N_0 - 1, & \text{se } \max\{n_1, \dots, n_s\} \leq M_{s+1} < \sum_{i=1}^s n_i \\ \infty & \text{se } M_{s+1} = \sum_{i=1}^s n_i \end{cases} \quad (3.5)$$

onde

$$N_0 = \min \left\{ n \in \mathbb{N}^* : \prod_{i=1}^s (M_{s+1} + n - n_i) < n(M_{s+1} + n)^{s-1} \right\}.$$

A unicidade deste estimador é garantida exceto quando

$$\prod_{i=1}^s (M_{s+1} + N_0 - 1 - n_i) = (N_0 - 1)(M_{s+1} + N_0 - 1)^{s-1},$$

em cujo caso os dois estimadores são $M_{s+1} + N_0 - 1$ e $M_{s+1} + N_0 - 2$.

3.3 Estimador Multiamostrai para Amostragem Inversa

Como no caso da amostragem direta, a amostragem inversa múltipla é uma generalização do método de amostragem simples, para mais de duas ocasiões. Neste procedimento, captura-se aleatoriamente e sem reposição uma amostra de n_1 indivíduos, marca-se a amostra convenientemente e retorna-se à população. Logo, para cada uma das s ocasiões seguintes ($s \geq 2$) fixa-se o número de indivíduos marcados m_j ($j=2, \dots, s$) a serem obtidos. Seleciona-se então sem reposição amostras aleatórias até obter em cada uma delas esse número de indivíduos marcados. O tamanho de cada amostra, n_j , é aleatório. Todos os indivíduos da amostra recebem uma nova marca antes de serem retornados à população.

Seja M_i o número de indivíduos marcados na população exatamente antes da i -ésima amostra ser capturada. Então

$$M_i = \sum_{j < i} (n_j - m_j), \quad i=1, \dots, s+1.$$

Note-se que $M_1 = m_1 = 0$, $M_2 = n_1$. Note-se também que $M_{s+1} = \sum_{j=1}^s (n_j - m_j)$ é o número de indivíduos distintos observados durante o experimento. Logo, a função de distribuição

está dada por

$$P(n_2, \dots, n_s | n_1, m_2, \dots, m_s, N) = \prod_{i=2}^s \frac{\binom{M_i}{m_i-1} \binom{N-M_i}{n_i-m_i}}{\binom{N}{n_i-1}} \frac{M_i - (m_i-1)}{N - (n_i-1)}. \quad (3.6)$$

Para obter o EMV de N estabelecem-se as condições

$$P(m_2, \dots, m_s | n_1, \dots, n_s, N) \geq P(m_2, \dots, m_s | n_1, \dots, n_s, N-1),$$

$$P(m_2, \dots, m_s | n_1, \dots, n_s, N) \geq P(m_2, \dots, m_s | n_1, \dots, n_s, N+1).$$

Das duas condições acima obtêm-se as seguintes desigualdades

$$\prod_{i=2}^s \frac{(N-M_i)(N-n_i)}{N(N-M_{i+1})} \geq 1 \quad \text{e} \quad \prod_{i=2}^s \frac{(N+1-M_i)(N+1-n_i)}{(N+1)(N+1-M_{i+1})} \leq 1$$

e dado que elas são exatamente iguais às obtidas no procedimento da amostragem direta, segue que o EMV de N , \hat{N}_{mi} , deve satisfazer às desigualdades seguintes

$$1 - \frac{M_{s+1}}{\hat{N}_{mi}} \leq \left(1 - \frac{n_1}{\hat{N}_{mi}}\right) \dots \left(1 - \frac{n_s}{\hat{N}_{mi}}\right),$$

$$1 - \frac{M_{s+1}}{\hat{N}_{mi}+1} \geq \left(1 - \frac{n_1}{\hat{N}_{mi}+1}\right) \dots \left(1 - \frac{n_s}{\hat{N}_{mi}+1}\right),$$

cujas solução simultânea fornece o estimador multiamostrai (como visto em 3.2).

Similarmente, a expressão mais explícita (3.5) obtida por Leite et al (1988) para \hat{N}_{mi} também é útil na amostragem inversa,

$$\hat{N}_{mi} = \begin{cases} M_{s+1} + N_0 - 1 & \text{se } \sum_{i=1}^s n_i > M_{s+1} \\ \infty & \text{se } \sum_{i=1}^s n_i = M_{s+1} \end{cases}$$

onde

$$N_0 = \min\{n \in \mathbb{N}^*: \prod_{i=1}^s (M_{s+1} + n - n_i) < n(M_{s+1} + n)^s\}.$$

Mas, como neste caso $m_i \geq 1$, $i=2, \dots, s$, então

$$M_i = \sum_{j < i} (n_j - m_j) = \sum n_j - \sum m_j < \sum n_j$$

Logo, o EMV para N é $\hat{N}_{mi} = M_{s+1} + N_0 - 1$, não existindo a possibilidade de ser ∞ . Além disso, a unicidade deste estimador é garantida exceto quando

$$\prod_{i=1}^s (M_{s+1} + N_0 - 1 - n_i) = (N_0 - 1)(M_{s+1} + N_0 - 1)^{s-1}$$

em cujo caso os dois estimadores são $M_{s+1} + N_0 - 1$ e $M_{s+1} + N_0 - 2$.

Chapman (1952) propôs o seguinte estimador para N para o esquema de amostragem inversa

$$\hat{N}_{mc} = \frac{1}{s} \sum_{i=2}^s \left(\frac{n_i(M_i+1)}{m_i} - 1 \right). \quad (3.7)$$

Pode-se notar que \hat{N}_{mc} é uma média aritmética de estimadores de Petersen modificados dado que se em cada ocasião i ($=2, \dots, s$) iguala-se a proporção de marcados na população à proporção de marcados na i -ésima amostra, ou seja $\frac{M_i}{N_i} = \frac{m_i}{n_i}$, obtendo-se para cada ocasião, o estimador de Petersen para N , $\hat{N}_i = \frac{M_i n_i}{m_i}$ no momento da i -ésima amostragem.

3.4 Estimador de Petersen (ponderado) Multiamostrai

Sen & Sen (1981) sugerem um estimador alternativo no caso de s amostras como sendo a média ponderada dos $s-1$ estimadores de Petersen para duas amostras $n_i M_i / m_i$. Para isso, consideraram a distribuição condicional de m_i dados M_i e n_i como dado em (3.2),

$$L_i(m_i | n_i, M_i) = \frac{\binom{M_i}{m_i} \binom{N-M_i}{n_i-m_i}}{\binom{N}{n_i}}, \quad i=2, \dots, s. \quad (3.8)$$

Com a função de verossimilhança acima, o EMV de N , \hat{N}_i , é o estimador de Petersen usual

$$\hat{N}_i = \frac{n_i M_i}{m_i}, \quad i=2, \dots, s \quad (3.9)$$

Observe-se que L_i não leva em conta as marcas múltiplas dos indivíduos, caso contrário, o numerador teria i fatores no lugar de dois fatores. Em consequência, a função de verossimilhança parcial (Cox(1975)) é utilizada,

$$L^{(s)}(m_2, \dots, m_s | n_1, \dots, n_s) = \prod_{i=2}^s L_i(m_i | n_i, M_i) = \prod_{i=2}^s \frac{\binom{M_i}{m_i} \binom{N-M_i}{n_i-m_i}}{\binom{N}{n_i}}. \quad (3.10)$$

Logo,

$$\log L^{(s)} = \sum_{i=2}^s \left\{ \log(N-M_i)! + \log(N-n_i)! - \log N! - \log(N-M_i-n_i+m_i)! \right. \\ \left. + \text{termos não envolvendo } N \right\}.$$

O EMV de N seria \hat{N}_m , solução de $(\partial/\partial N) \log L^{(s)} = 0$, mas este poderia não resultar um número inteiro e a solução não seria exata. Assim, usando a derivada juntamente com a aproximação de Stirling para fatoriais, obtem-se

$$\log L^{(s)} = \sum_{i=2}^s \left\{ \log \sqrt{2\pi} (N-M_i)^{N-M_i+1/2} \exp(-(N-M_i)) \right. \\ + \log \sqrt{2\pi} (N-n_i)^{N-n_i+1/2} \exp(-(N-n_i)) - \log \sqrt{2\pi} N^{N+1/2} \exp(-N) \\ - \log \sqrt{2\pi} (N-M_i-n_i+m_i)^{N-M_i-n_i+m_i+1/2} \exp(-(N-M_i-n_i+m_i)) \\ \left. + \text{termos não envolvendo } N \right\}.$$

Derivando-se com relação a N , obtem-se

$$\frac{\partial}{\partial N} \log L^{(s)} = \sum_{i=2}^s \left\{ \frac{N-M_i+1/2}{N-M_i} + \log(N-M_i) - 1 + \frac{N-n_i+1/2}{N-n_i} + \log(N-n_i) - 1 \right. \\ \left. - \frac{N+1/2}{N} - \log N + 1 - \frac{N-M_i-n_i+m_i+1/2}{N-M_i-n_i+m_i} - \log(N-M_i-n_i+m_i) + 1 \right\} \\ = \sum_{i=2}^s \left\{ -\log \frac{N(N-M_i-n_i+m_i)}{(N-M_i)(N-n_i)} \right. \\ \left. + \frac{1}{2} \left(\frac{1}{N-M_i} + \frac{1}{N-n_i} - \frac{1}{N} - \frac{1}{N-M_i-n_i+m_i} \right) + O(N^{-2}) \right\}. \quad (3.11)$$

Mas, como

$$N(N-M_i-n_i+m_i) = (N-M_i)(N-n_i) + (Nm_i - M_i n_i),$$

tem-se que

$$\frac{\partial}{\partial N} \log L^{(s)} = \sum_{i=2}^s \left\{ -\log \left(1 + \frac{Nm_i - M_i n_i}{(N-M_i)(N-n_i)} \right) \right.$$

$$+ \frac{1}{2} \left(\frac{1}{N-M_i} + \frac{1}{N-n_i} - \frac{1}{N} - \frac{1}{N-M_i-n_i+m_i} \right) + O(N^{-2}) \}. \quad (3.12)$$

Pode-se notar que, se na expressão anterior são omitidos os termos $\frac{1}{2} \left(\frac{1}{N-M_i} + \frac{1}{N-n_i} - \frac{1}{N} - \frac{1}{N-M_i-n_i+m_i} \right) + O(N^{-2})$, a equação se reduz a

$$1 - \frac{M_{s+1}}{N} = \left(1 - \frac{n_1}{N} \right) \dots \left(1 - \frac{n_s}{N} \right)$$

mas neste caso o estimador obtido não poderia ser expressado em termos dos estimadores de Petersen individuais.

Seja

$$Z_i = \frac{m_i - n_i M_i / N}{\sqrt{N}}, \quad i=2, \dots, s \quad (3.13)$$

logo,

$$E Z_i = 0, \quad E [Z_i Z_j] = 0, \quad i \neq j=2, \dots, s \quad (3.14)$$

desde que $E m_i = n_i M_i / N$ e Z_i, Z_j são independentes.

Agora

$$E Z_i^2 = E [E (Z_i^2 | n_i M_i)]$$

$$\begin{aligned} E(Z_i^2 | n_i M_i) &= \frac{1}{N} E m_i^2 - 2 \frac{n_i M_i}{N} E m_i + \frac{n_i^2 M_i^2}{N^3} \\ &= \frac{1}{N} \left\{ \frac{n_i(n_i-1)M_i(M_i-1)}{N(N-1)} + \frac{n_i M_i}{N} \right\} - \frac{2n_i M_i}{N^2} \frac{n_i M_i}{N} + \frac{n_i^2 M_i^2}{N^2} \\ &= \frac{n_i(n_i-1)M_i(M_i-1)}{N^2(N-1)} + \frac{n_i M_i}{N^2} - \frac{2n_i^2 M_i^2}{N^3} + \frac{n_i^2 M_i^2}{N^3} \\ &= \frac{n_i M_i}{N^2} \left(\frac{(n_i-1)(M_i-1)}{(N-1)} + 1 - \frac{n_i M_i}{N} \right) \\ &= \frac{n_i (N-n_i)M_i(N-M_i)}{N^3(N-1)}. \end{aligned}$$

Portanto,

$$E Z_i^2 = \frac{n_i(N-n_i)}{N^3(N-1)} E [M_i (N-M_i)] < \frac{1}{16}, \quad i=2, \dots, s \quad (3.15)$$

Logo, de (3.14) e (3.15), e pelo lema de Chevyshev,

$$|Z_i| = O_p(1), \quad i=2,\dots,s.$$

Pelo anterior, (3.12) pode-se escrever

$$\frac{\partial}{\partial N} \log L^{(s)} = - \sum_{i=2}^s \frac{Z_i N^{3/2}}{(N - M_i)(N - n_i)} + O_p(N^{-1}). \quad (3.16)$$

Agora, usando só os termos relevantes da expressão acima, o EMV \hat{N}_{pp} de N vai estar dado pela solução de

$$\sum_{i=2}^s \frac{Z_i^* (\hat{N}_{pp})^{3/2}}{(\hat{N}_{pp} - M_i)(\hat{N}_{pp} - n_i)} = 0$$

onde Z_i^* é definido em (3.13) com N substituído por \hat{N}_{pp} , isto é,

$$\sum_{i=2}^s \frac{\hat{N}_{pp} m_i - M_i n_i}{(\hat{N}_{pp} - M_i)(\hat{N}_{pp} - n_i)} = 0$$

por tanto,

$$\sum_{i=2}^s \frac{\hat{N}_{pp} m_i}{(\hat{N}_{pp} - M_i)(\hat{N}_{pp} - n_i)} = \sum_{i=2}^s \frac{M_i n_i}{(\hat{N}_{pp} - M_i)(\hat{N}_{pp} - n_i)}$$

e

$$\hat{N}_{pp} = \frac{\sum_{i=2}^s \frac{M_i n_i}{(\hat{N}_{pp} - M_i)(\hat{N}_{pp} - n_i)}}{\sum_{i=2}^s \frac{m_i}{(\hat{N}_{pp} - M_i)(\hat{N}_{pp} - n_i)}}$$

agora, por (3.9) obtem-se

$$\hat{N}_{pp} = \frac{\sum_{i=2}^s \frac{\hat{N}_i m_i}{(\hat{N}_{pp} - M_i)(\hat{N}_{pp} - n_i)}}{\sum_{i=2}^s \frac{m_i}{(\hat{N}_{pp} - M_i)(\hat{N}_{pp} - n_i)}}, \quad (3.17)$$

que é o estimador de Petersen (ponderado) para amostragem direta múltipla.

Para \hat{N}_{pp} podem ser formuladas soluções iterativas como a que segue. Seja

$$w_i^{(1)} = \frac{m_i}{(\hat{N}_i - M_i)(\hat{N}_i - n_i)} \quad i=2,\dots,s$$

e seja

$$\hat{N}_{pp}^{(1)} = \frac{\sum_{i=2}^s w_i^{(1)} \hat{N}_i}{\sum_{i=2}^s w_i^{(1)}}$$

então para $r \geq 1$,

$$w_i^{(r)} = \frac{m_i}{(\hat{N}_{pp}^{(r)} - M_i)(\hat{N}_{pp}^{(r)} - n_i)} \quad i=2, \dots, s$$

logo,

$$\hat{N}_{pp}^{(r+1)} = \frac{\sum_{i=2}^s w_i^{(r)} \hat{N}_i}{\sum_{i=2}^s w_i^{(r)}}$$

O processo iterativo deverá parar logo que a diferença entre $\hat{N}^{(r)}$ e $\hat{N}^{(r+1)}$ seja o suficientemente pequena.

Uma expressão um pouco diferente pode ser encontrada para (3.17) substituindo \hat{N}_i pelo estimador sugerido por Chapman e dado em (2.4), sendo este último não viciado se $n_i + M_i \geq n$, $i=1, \dots, s$.

3.5 Distribuição Assintótica do Estimador de Petersen no caso Multiamostrai

Sen e Sen (1982) estudaram as propriedades assintóticas do estimador de Petersen, dado em (3.17), para múltiplas amostras.

Tem-se de (3.1), para $k \geq 2$,

$$M_{k+1}^o = E M_{k+1} = E M_k + n_k - E m_k = M_k^o + n_k - \frac{n_k M_k^o}{N}$$

também, $M_2^o = M_2 = n_1$

Sejam

$$\frac{n_i}{N} = \alpha_i, \quad i=1, \dots, s \quad (0 < \alpha_i < 1) \quad (3.18)$$

$$\gamma_k = \frac{M_k^o}{N} \quad (3.19)$$

então de (3.18) e (3.19)

$$\gamma_k = \alpha_{k-1} + \gamma_{k-1} (1 - \alpha_{k-1})$$

e conseqüentemente,

$$\begin{aligned} \gamma_k &= \alpha_{k-1} + \alpha_{k-2}(1 - \alpha_{k-1}) + \alpha_{k-3}(1 - \alpha_{k-2})(1 - \alpha_{k-1}) + \alpha_1(1 - \alpha_2) \dots (1 - \alpha_{k-1}) \\ &= 1 - (1 - \alpha_{k-1}) [1 - \alpha_{k-2} - \alpha_{k-3} (1 - \alpha_{k-2}) - \dots - \alpha_1(1 - \alpha_2) \dots (1 - \alpha_{k-2})] \\ &= 1 - (1 - \alpha_{k-1}) (1 - \alpha_{k-2}) [1 - \alpha_{k-3} - \dots - \alpha_1 (1 - \alpha_2) \dots (1 - \alpha_{k-3})] \\ &= 1 - (1 - \alpha_{k-1}) (1 - \alpha_{k-2}) \dots (1 - \alpha_1) \\ &= 1 - \beta_{k-1} \quad , \quad k=2, \dots, s \end{aligned}$$

Agora, a esperança da função escore em (3.8) é zero, isto é,

$$E \left(\frac{\partial}{\partial N} \log L_i (m_i | n_i, M_i) \mid n_i, M_i \right) = 0 \quad , \quad i=2, \dots, s$$

Então por (3.10)

$$E \left(\frac{\partial}{\partial N} \log L^{(s)} \right) = 0$$

e dados n_i e M_i , m_i tem distribuição hipergeométrica e Z_i é assintoticamente normal com média zero e variância (condicional) $n_i(N - n_i)M_i(N - M_i)/N^3(N - 1)$, $i=2, \dots, s$. Logo, usando uma cadeia de argumentos condicionais, tem-se que para N grande,

$$N^{-1/2} \frac{\partial}{\partial N} \log L^{(s)} = N^{-1/2} \left\{ \sum_{i=2}^s \frac{Z_i N^{3/2}}{(N - M_i)(N - n_i)} + O_p(N^{-1}) \right\}$$

é também assintoticamente normal.

Por (3.9) e (3.11) tem-se

$$-\frac{\partial^2}{\partial N^2} \log L^{(s)} = \frac{1}{N} \sum_{i=2}^s \frac{M_i n_i}{(N - M_i)(N - n_i)} - \sum_{i=2}^s \frac{Z_i N^{5/2}}{\{(N - M_i)(N - n_i)\}^2} + O_p(N^{-2})$$

de forma que pela expansão de M_i ao redor de M_i^0 , tem-se

$$\begin{aligned} E \left(-N \frac{\partial^2}{\partial N^2} \log L^{(s)} \right) &= \sum_{i=2}^s \frac{M_i^0 n_i}{(N - M_i^0)(N - n_i)} + O(N^{-1}) \\ &= \sum_{i=2}^s \frac{\alpha_i (1 - \beta_{i-1})}{\beta_i} + O(N^{-1}) \end{aligned} \quad (3.20)$$

Também de (3.16)

$$N^{-1/2} \frac{\partial}{\partial N} \log L^{(s)} = \sum_{i=2}^s \beta_i^{-1} Z_i + O(N^{-1/2}) \quad (3.21)$$

Logo, de (3.20) e (3.21) e pelo teorema de Slutsky, conclui-se que para N grande,

$$N^{-1/2} (\hat{N}_{pp} - N) \xrightarrow{D} N(0, \sigma^2) \quad (3.22)$$

onde

$$\sigma^2 = \left\{ \sum_{i=2}^s \frac{\alpha_i}{\beta_i} (1 - \beta_{i-1}) \right\}^{-1}. \quad (3.23)$$

Note-se que $\sigma_2^2 = \left\{ \frac{\alpha_2}{\beta_2} (1 - \beta_1) \right\}^{-1} = (1 - \alpha_1)(1 - \alpha_2)/\alpha_1\alpha_2$. Esta expressão concorda com a variância assintótica do estimador de Petersen como dada em (2.12) fazendo $\beta = \alpha_2$.

Pode-se mostrar que a variância (assintótica) do estimador multiamostrado \hat{N}_{pp} é menor que a variância do estimador clássico de Petersen (amostragem simples ou duas amostras) quando considerase o total das amostras múltiplas n_2, \dots, n_s próximo ao tamanho n_2 da segunda amostra. Efetivamente, considere-se o caso $s=3$, fazendo $\alpha_1 = \alpha$, $\alpha_2 + \alpha_3 = \beta$, $n_2 + n_3 = n_2'$. Então, comparando a variância em (3.22) para σ_2^2 com sua expressão em (2.12), tem-se

$$\begin{aligned} \frac{\alpha_2}{\beta_2} (1 - \beta_1) + \frac{\alpha_3}{\beta_3} (1 - \beta_2) &\geq \frac{\alpha_1(\alpha_2 + \alpha_3)}{(1 - \alpha_1)(1 - \alpha_2 - \alpha_3)} \\ \frac{\alpha_2\alpha_1}{(1 - \alpha_1)(1 - \alpha_2)} + \frac{\alpha_3(\alpha_1 + \alpha_2 - \alpha_1\alpha_2)}{(1 - \alpha_1)(1 - \alpha_2)(1 - \alpha_3)} &\geq \frac{\alpha_1\alpha_2}{(1 - \alpha_1)(1 - \alpha_2 - \alpha_3)} + \frac{\alpha_1\alpha_3}{(1 - \alpha_1)(1 - \alpha_2 - \alpha_3)} \end{aligned}$$

para todo $0 < \alpha_1, \alpha_2, \alpha_3 < 1$, $(\alpha_2 + \alpha_3 \leq 1)$ de forma que $\sigma_2^2 \geq \sigma_3^2$. Similarmente, a desigualdade anterior verifica-se fazendo $\alpha = \alpha_1 + \alpha_2$ e $\beta = \alpha_3$.

3.7 Metodologia Bootstrap nos Experimentos de C-R para Três Amostras

Nesta seção são apresentados os resultados da aplicação da metodologia bootstrap aos experimentos de C-R para três amostras e com amostragem direta, como tratado no capítulo 1. O objetivo principal é fornecer estimativas bootstrap do erro padrão da estimativa do tamanho da população N , assim como intervalos de confiança para N baseados no percentil bootstrap.

Na aplicação, primeiro simulam-se experimentos de C-R para três amostras, quando a amostragem é direta e sem reposição.

Suponha-se o tamanho de população N , e os tamanhos das amostras, n_1 , n_2 e n_3 respectivamente. A construção do experimento consiste no seguinte:

1. Atribuir valor 0 a todos os indivíduos da população.
2. Selecionar uma amostra aleatória sem reposição de tamanho n_1 e atribuir valor 1 aos indivíduos selecionados.
3. Selecionar n_2 indivíduos da mesma população em 1, por amostragem aleatória sem reposição, atribuindo valor 1 aos selecionados.
4. Selecionar n_3 indivíduos da mesma população em 1, por amostragem aleatória sem reposição, atribuindo valor 1 aos selecionados.
5. Contagem dos indivíduos selecionados na primeira amostra e na segunda, isto é número de indivíduos com valor 1 obtidos em 2 e 3 (representado por m_2).
6. Contagem dos indivíduos selecionados na primeira ou na segunda amostra e selecionado na terceira amostra, isto é, número de indivíduos com valor 1 obtidos em 2 ou 3 e 4 (representado por m_3).
7. O número total de indivíduos marcados na população após a i -ésima ocasião (ou antes da $i + 1$ -ésima ocasião) é calculado através de $M_i = \sum_{j < i} (n_j - m_j)$.

O procedimento de simulação foi realizado para o tamanho de população de 1000 indivíduos. Para esta população foram simulados sete experimentos com tamanhos amostrais comparáveis aos considerados no caso do método de C-R de amostragem simples. Em cada experimento, a primeira amostra é do tamanho da primeira amostra do respectivo experimento para duas amostras, já as amostras 2 e 3, foram ambas iguais à metade da segunda amostra do experimento para duas amostras. Por exemplo, no experimento 2 para duas amostras os tamanhos amostrais foram $n_1=250$ e $n_2=240$, para o caso de três amostras $n_1=250$, e $n_2=n_3=120$ e assim sucessivamente para o resto dos experimentos.

Na Tabela 11 são apresentadas as estatísticas resultantes da simulação, tais como m_2 , o número de indivíduos marcados capturados na segunda amostra, m_3 , o número de indivíduos marcados na primeira ou na segunda amostra e capturados na terceira, M_3 , o número total de indivíduos diferentes marcados antes da captura da terceira amostra, e M_4 o número total de indivíduos diferentes marcados durante o experimento. Elas são fornecidas pelo programa PROG4.SAS (Apêndice), realizado utilizando o software SAS.

TABELA 11. Simulação dos experimentos de C-R para três amostras e tamanho de população 1000.

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
n_1	150	250	150	250	750	500	150
n_2	80	120	125	75	80	80	250
n_3	80	120	125	75	80	80	250
m_2	14	36	20	22	64	40	45
m_3	16	40	32	29	58	49	83
M_2	150	250	150	250	750	500	150
M_3	216	334	255	303	766	540	355
M_4	280	414	348	349	788	571	522

Com os dados da tabela acima, é feito o cálculo das estimativas do tamanho da população para cada experimento, utilizando dois estimadores, o de Leite et al. (3.5)

$$\hat{N}_m = \begin{cases} M_{s+1} + N_0 - 1 & \text{se } \max \{n_1, \dots, n_s\} \leq M_{s+1} < \sum_{i=1}^s n_i \\ \infty & \text{se } M_{s+1} = \sum_{i=1}^s n_i \end{cases}$$

onde

$$N_0 = \min \left\{ n \in N^* : \prod_{i=1}^s (M_{s+1} + n - n_i) < n(M_{s+1} + n)^{s-1} \right\}.$$

e o estimador multiamostrai de Petersen (ponderado) (3.17)

$$\hat{N}_{pp} = \frac{\sum_{i=2}^s \frac{\hat{N}_i m_i}{(\hat{N}_{pp} - M_i)(\hat{N}_{pp} - n_i)}}{\sum_{i=2}^s \frac{m_i}{(\hat{N}_{pp} - M_i)(\hat{N}_{pp} - n_i)}}$$

Para obter \hat{N}_m , o valor de N_0 é encontrado da condição acima e verifica-se sua unicidade da condição estabelecida em (3.5). Para \hat{N}_{pp} uma solução iterativa é formulada, fazendo com que a diferença entre duas estimativas consecutivas seja no máximo de 0,001 vezes o valor imediatamente anterior. Também, por (3.22),

$$N^{-1/2} (\hat{N}_{pp} - N) \sim \mathcal{N} \left(0, \left\{ \sum_{i=2}^s \frac{\alpha_i}{\beta_i} (1 - \beta_{i-1}) \right\}^{-1} \right)$$

logo, é encontrado o valor para $v(\hat{N}_{pp})$ substituindo N pela estimativa obtida \hat{N}_{pp} , assim como um intervalo de confiança assintótico é calculado por

$$\text{I.C.} \left[\hat{N}_{pp} \pm Z_{\alpha} [v(\hat{N}_{pp})]^{1/2} \right] = 1 - \alpha.$$

Na Tabela 12 são apresentadas as estimativas do tamanho da população assim calculadas, para cada experimento. As estimativas \hat{N}_m resultaram maiores do que \hat{N}_{pp} em todos os experimentos. Também é apresentada a estimativa da variância assintótica, $v(\hat{N}_{pp})$, assim como o intervalo de confiança assintótico utilizando esta variância.

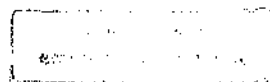
TABELA 12. Estimativas de Leite e de Petersen multiamostrais.

	Experimento						
	1	2	3	4	5	6	7
\hat{N}_m	980	927	975	811	996	932	1001
\hat{N}_{pp}	935	913	951	793	992	923	992
$[v(\hat{N}_{pp})]^{1/2}$	142,28	79,17	106,49	84,10	41,53	61,50	62,52
Intervalo de Confiança assintótico (baseado em $v(\hat{N}_{pp})$)							
L.inf.	656	758	742	628	911	802	869
L.sup.	1214	1068	1160	958	1073	1044	1115

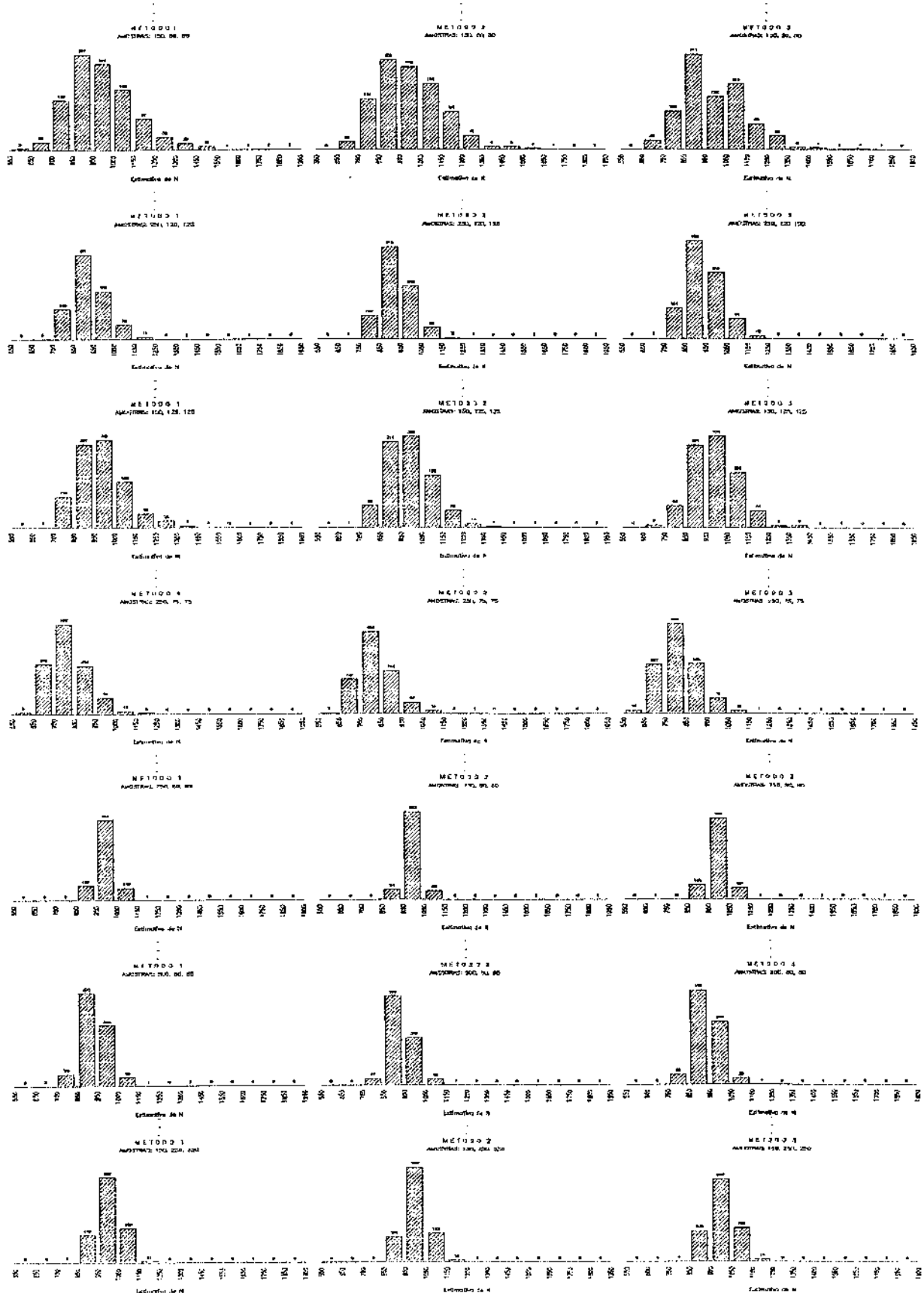
Como foi visto no capítulo 1, o método bootstrap na técnica de C-R de amostragem múltipla, mostra-se como uma solução para a estimação da variância das estimativas do tamanho da população assim como para a construção de intervalos de confiança para o parâmetro N . Os três métodos bootstrap são aplicados a estes experimentos, sendo denominados de método 1, 2 e 3 respectivamente. No dois primeiros, a análise é não condicional aos tamanhos observados das amostras, logo, considera-se o vetor $\eta = (n_{111}, n_{110}, \dots, n_{000})$, indicando as frequências das histórias de captura nas três amostras como tendo distribuição multinomial $(\hat{N}_{pp}, p_{111}, \dots, p_{000})$ no caso do método 1, e tendo distribuição multinomial $(M_4, p_{111}, \dots, p_{001})$ no caso do método 2. No terceiro método, a análise é condicional aos tamanhos amostrais e assume a distribuição hipergeométrica generalizada, onde m_2 tem distribuição hipergeométrica (\hat{N}_{pp}, M_2, n_2) e m_3 hipergeométrica (\hat{N}_{pp}, M_3, n_3) .

O número de replicações bootstrap é de $B=1000$ que como no caso da amostragem simples, provou ser mais adequado assim como é o valor sugerido em Buckland (1984).

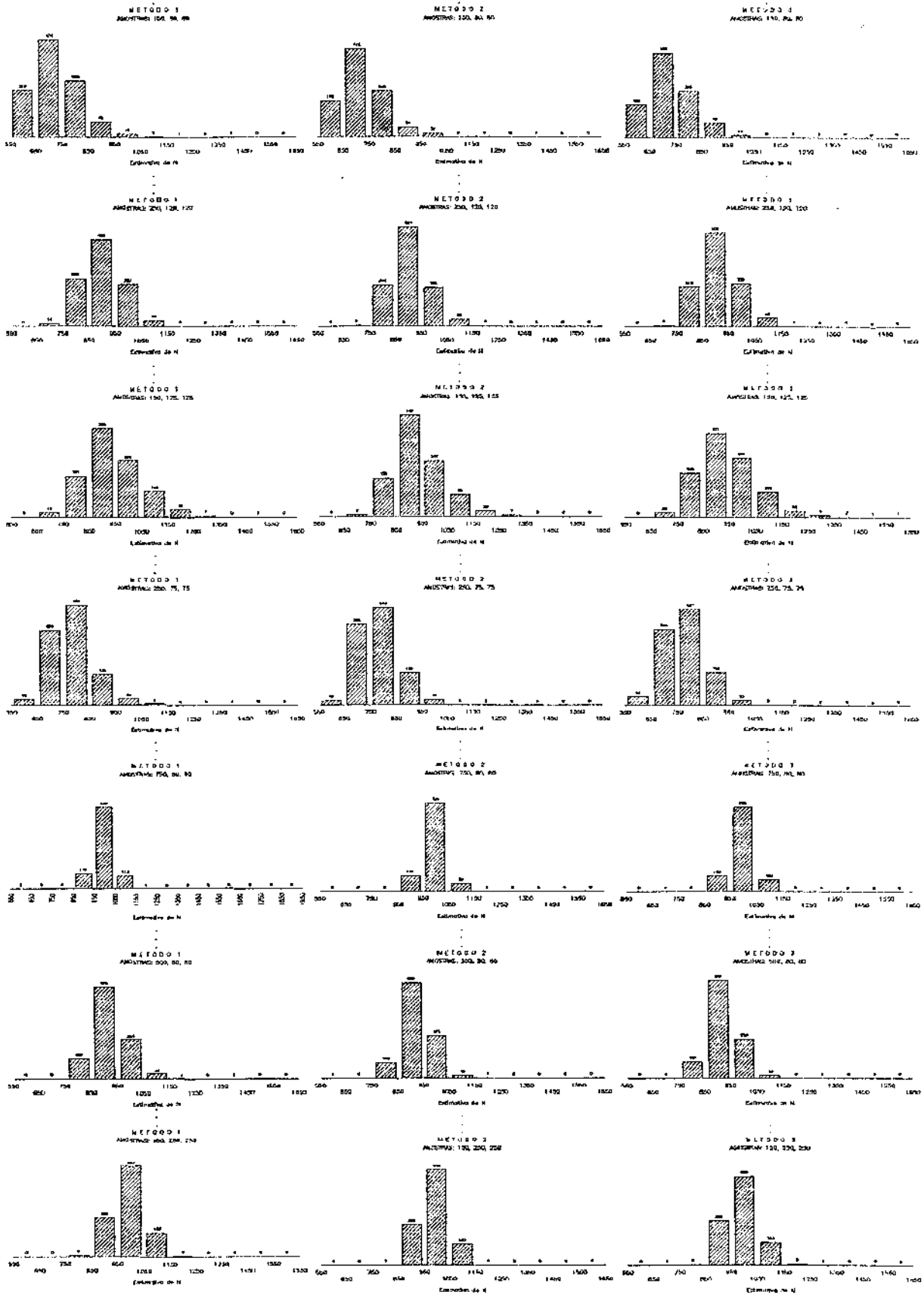
Nas figuras seguintes são apresentados os histogramas das estimativas \hat{N}_m e \hat{N}_{pp} respectivamente, para cada experimento e para cada método aplicado. Em cada figura, os três primeiros gráficos correspondem aos três métodos bootstrap para o primeiro experimento e assim sucessivamente para os experimentos 2 a 7. A assimetria das distribuições das estimativas não é tão evidente quanto o foi nos experimentos de C-R de amostragem simples para o tamanho de população $N=1000$. Nota-se a existência de valores extremos à direita com frequências muito baixas.



Histogramas das replicações bootstrap para o estimador de Leite et al. para o tamanho de população $N=1000$



Histogramas das replicações bootstrap para o estimador de Petersen para o tamanho de população $N=1000$



Nas tabelas 13 e 14 são apresentadas, para cada um dos experimentos simulados, as estimativas bootstrap do erro padrão de \hat{N}_m e \hat{N}_{pp} respectivamente calculadas utilizando (1.1),

$$\hat{V}(\hat{N}) = \{ \sum \hat{N}_j^{*2} - (\sum \hat{N}_j^*)^2 / B \} / (B - 1),$$

assim como os intervalos de confiança para N , baseados nos percentis bootstrap e os intervalos do percentil de vício corrigido (como visto na seção 1.3). Estes resultados são fornecidos pelos programas PROG5.SAS e PROG6.SAS respectivamente. (Apêndice)

TABELA 13. Estimativa bootstrap do erro padrão e intervalos de confiança para a estimativa multiamostrai (Petersen ponderado).

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
Estimativa bootstrap do erro padrão de \hat{N}_{pp}							
Método 1	134,11	78,86	108,00	82,75	41,79	63,06	62,58
Método 2	133,34	73,27	96,97	75,35	38,57	58,87	52,39
Método 3	141,15	75,60	111,05	79,53	42,43	58,47	60,59
Intervalo de Confiança baseado no percentil bootstrap							
Met. 1 L.inf.	690	766	760	643	910	809	873
L.sup.	1223	1069	1177	978	1078	1058	1121
Met. 2 L.inf.	717	783	781	652	921	818	896
L.sup.	1220	1072	1159	955	1074	1042	1098
Met. 3 L.inf.	686	774	756	639	921	811	875
L.sup.	1253	1070	1205	960	1083	1042	1112
Intervalo de Confiança do percentil de vício corrigido							
Met. 1 L.inf.	750	788	799	671	919	822	880
L.sup.	1345	1110	1235	1052	1089	1077	1138
Met. 2 L.inf.	758	800	820	688	924	831	910
L.sup.	1383	1111	1276	1026	1081	1076	1117
Met. 3 L.inf.	741	784	794	672	921	827	882
L.sup.	1435	1073	1299	1010	1083	1070	1132

TABELA 14. Estimativa bootstrap do erro padrão e intervalos de confiança para a estimativa multiamostrai de Leite et al.

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
Estimativa bootstrap do erro padrão de \hat{N}_m							
Método 1	166,58	83,03	120,15	93,87	41,23	64,44	64,32
Método 2	157,29	75,26	110,91	88,32	37,61	59,51	56,29
Método 3	157,36	83,75	113,21	91,25	43,22	62,96	64,81
Intervalo de Confiança baseado no percentil bootstrap							
Met. 1 L.inf.	746	789	794	668	925	817	885
L.sup.	1394	1111	1279	1033	1084	1074	1136
Met. 2 L.inf.	753	798	813	677	929	835	908
L.sup.	1346	1092	1234	1033	1080	1069	1128
Met. 3 L.inf.	746	784	796	672	925	825	880
L.sup.	1349	1130	1219	1018	1089	1069	1144
Intervalo de Confiança do percentil de vício corrigido							
Met. 1 L.inf.	755	800	799	672	925	813	877
L.sup.	1421	1137	1286	1061	1084	1069	1130
Met. 2 L.inf.	749	804	807	676	931	842	906
L.sup.	1330	1098	1229	1029	1080	1080	1122
Met. 3 L.inf.	726	775	770	661	917	816	874
L.sup.	1349	1093	1189	993	1079	1055	1123

Pode-se observar que as magnitudes obtidas para o erro padrão das estimativas são menores do que as obtidas com amostragem simples para o mesmo tamanho de população. Também, em geral os intervalos resultaram de menor comprimento quando feita esta comparação.

Nas tabelas 15 e 16 são apresentados os comprimentos dos intervalos de confiança obtidos pelos três métodos bootstrap e para cada um dos experimentos além do intervalo assintótico baseado na estimação da variância de \hat{N}_{pp} .

TABELA 15. Comprimento dos Intervalos de confiança obtidas das estimativas de Petersen Multiamostrat (Ponderado).

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
Intervalo assintótico baseado em							
$v(\hat{N}_{pp})$	558	310	418	330	162	242	246
Intervalos baseados no percentil bootstrap							
Método 1	533	303	417	335	168	249	248
Método 2	503	289	378	303	153	224	202
Método 3	567	296	449	321	162	231	237
Intervalos do percentil de vício corrigido							
Método 1	595	322	436	381	170	255	258
Método 2	625	311	447	338	157	245	207
Método 3	694	289	505	338	162	243	250

TABELA 16. Comprimento dos Intervalos de confiança obtidas das estimativas de Leite et al.

	E x p e r i m e n t o						
	1	2	3	4	5	6	7
Intervalos baseados no percentil bootstrap							
Método 1	648	322	485	365	159	257	251
Método 2	593	294	421	456	151	234	220
Método 3	603	356	413	346	164	244	264
Intervalos do percentil de vício corrigido							
Método 1	666	337	487	389	159	256	253
Método 2	586	294	422	353	149	248	216
Método 3	623	216	419	332	162	239	249

CAPÍTULO 4

ESTIMAÇÃO DO TAMANHO DA POPULAÇÃO BASEADO EM AMOSTRAGEM SEQUENCIAL

4.1 Introdução

Os estimadores do tamanho de população N obtidos pelos métodos de C-R e dos esquemas de amostragem direta e inversa, descritos nos capítulos anteriores, apresentam erros amostrais que decrescem conforme o tamanho de amostra n cresce. Como N na maioria dos casos é um número razoavelmente grande, poderia-se sugerir um tamanho de amostra grande para reduzir tais erros na estimação. Contudo, a técnica de C-R envolve um custo operacional (tanto para a captura dos indivíduos da amostra como para sua marcação e libertação) de forma que é preciso ter em conta este custo na escolha de um esquema de amostragem adequado. Este custo pode não depender (sensivelmente) do tamanho da população, mas sim do tamanho da amostra e da técnica de amostragem. Como N é desconhecido não há uma solução ótima para n a priori e portanto pode-se recorrer aos métodos sequenciais a fim de alcançar esta otimização, pelo menos no contexto assintótico.

A amostragem sequencial consiste em capturar os indivíduos aleatoriamente um a um, marcá-los convenientemente (ou remarca-los, caso tenham sido marcados previamente), e libertá-los antes da captura do seguinte indivíduo. Assim sendo, pode-se optar pelo procedimento da escolha do tamanho de amostra ser estocástico ou não estocástico. As variáveis que interessam neste tipo de amostragem são M_k , o número de indivíduos marcados (não fazendo distinção entre indivíduos com uma ou múltiplas marcas) exatamente antes da k -ésima escolha e X_k representando se o indivíduo capturado na k -ésima escolha é ou não um indivíduo marcado. Assim, $M_1=0$, $M_2=1$, $M_3=2$ se o segundo indivíduo capturado foi não marcado, ou $M_3=1$ se o segundo indivíduo capturado já havia sido marcado, e assim por diante. Também, $M_{k+1} \geq M_k$ para todo k , e

$$M_{k+1} = M_k + (1 - X_k), \quad k \geq 1$$

onde

$$X_k = \begin{cases} 1, & \text{se a } k\text{-ésima escolha corresponde a um indivíduo marcado} \\ 0, & \text{caso contrário} \end{cases}$$

O quadro seguinte ilustra um experimento de C-R de amostragem sequencial, com a notação estabelecida e para um tamanho de amostra de 12, observa-se que as escolhas 5, 7 e 11 corresponderam a indivíduos marcados. O número de indivíduos marcados durante o experimento ou equivalentemente o número de marcados na população antes da escolha, eventual, do indivíduo 13 é $M_{13}=9$.

k	1	2	3	4	5	6	7	8	9	10	11	12
X_k	0	0	0	0	1	0	1	0	0	0	1	0
$1 - X_k$	1	1	1	1	0	1	0	1	1	1	0	1
M_k	0	1	2	3	4	4	5	5	6	7	8	8

Neste capítulo o interesse é estudar estes procedimentos sequenciais para a estimação de N . Na seção 4.2, é realizada a revisão da amostragem sequencial com tamanho de amostra fixo, também é mostrada a distribuição assintótica do estimador sequencial. Na seção 4.3, o estuda-se a amostragem sequencial no caso deste tamanho ser estocástico, e a distribuição assintótica. Finalmente são apresentados resultados obtidos por simulação para ambos os casos.

4.2 Amostragem Sequencial com o Tamanho de Amostra Fixo

O objetivo desta seção é descrever o método de estimação sequencial com o tamanho de amostra fixo, suas propriedades assintóticas, e também algumas considerações sobre sua eficiência assintótica com respeito ao estimador de Petersen.

Seja M_k o número de indivíduos marcados existentes na população exatamente antes da captura do k -ésimo indivíduo, $k \geq 1$. Logo, $M_0=M_1=0$, $M_2=1$, $M_{k+1} \geq M_k$, para todo $k \geq 1$, e

$$M_{k+1} = M_k + (1 - X_k), \quad k \geq 1 \quad (4.1)$$

onde

$$X_k = \begin{cases} 1, & \text{se a } k\text{-ésima escolha corresponde a um indivíduo marcado} \\ 0, & \text{caso contrário} \end{cases} \quad (4.2)$$

A função de probabilidade condicional de X_k , dados X_1, \dots, X_{k-1} , segue a lei Bernoulli,

$$f_k(X_k | X_1, \dots, X_{k-1}) = \left(\frac{M_k}{N}\right)^{X_k} \left(1 - \frac{M_k}{N}\right)^{1-X_k} \quad k \geq 2 \quad (4.3)$$

Note-se que esta probabilidade condicional depende de X_1, \dots, X_{k-1} somente através de M_k . Então na amostragem sequencial de indivíduos marcados à n -ésima escolha ($n \geq 2$), a função de verossimilhança está dada por

$$L_n(N) = \prod_{k=2}^n \frac{1}{N^{n-1}} \left\{ M_k^{X_k} (N - M_k)^{1-X_k} \right\}. \quad (4.4)$$

$L_n(N)$ também pode ser interpretado como uma função de verossimilhança parcial pelo fato das definições de X_k e M_k não fazerem distinção entre marcas simples e marcas múltiplas dos indivíduos.

Para achar o EMV de N estabelecem-se as condições

$$\begin{aligned} \frac{L_n(N)}{L_n(N-1)} &= \frac{\prod_{k=2}^n \left\{ M_k^{X_k} (N - M_k)^{1-X_k} \right\} / N}{\prod_{k=2}^n \left\{ M_k^{X_k} (N-1 - M_k)^{1-X_k} \right\} / (N-1)} \\ &= \left(1 - \frac{1}{N}\right)^{n-1} \prod_{k=2}^n \left(1 - \frac{1}{N - M_k}\right)^{-(1-X_k)} \geq 1 \end{aligned}$$

e

$$\begin{aligned} \frac{L_n(N)}{L_n(N+1)} &= \frac{\prod_{k=2}^n \left\{ M_k^{X_k} (N - M_k)^{1-X_k} \right\} / N}{\prod_{k=2}^n \left\{ M_k^{X_k} (N+1 - M_k)^{1-X_k} \right\} / (N+1)} \\ &= \left(1 - \frac{1}{N+1}\right)^{-(n-1)} \prod_{k=2}^n \left(1 - \frac{1}{N+1 - M_k}\right)^{1-X_k} \geq 1 \end{aligned}$$

portanto, o EMV de N satisfaz as condições:

$$\left(1 - \frac{1}{\hat{N}_n - M_2}\right)^{1-X_2} \cdots \left(1 - \frac{1}{\hat{N}_n - M_n}\right)^{1-X_n} \leq \left(1 - \frac{1}{\hat{N}_n}\right)^{n-1} \quad (4.5)$$

$$\left(1 - \frac{1}{\hat{N}_n + 1 - M_2}\right)^{1-X_2} \cdots \left(1 - \frac{1}{\hat{N}_n + 1 - M_n}\right)^{1-X_n} \geq \left(1 - \frac{1}{\hat{N}_n + 1}\right)^{n-1} \quad (4.6)$$

Equivalentemente, pode-se maximizar $\log L_n(N)$ e obter o EMV \hat{N}_n de N como solução de

$$\begin{aligned} \frac{\partial}{\partial N} \log L_n(N) &= \frac{\partial}{\partial N} \left\{ -(n-1) \log N + \sum_{k=2}^n \{X_k \log M_k + (1-X_k) \log (N-M_k)\} \right\} \\ &= -\frac{n-1}{N} + \sum_{k=2}^n \frac{1-X_k}{N-M_k} = 0 \end{aligned}$$

portanto,

$$\frac{1-X_2}{\hat{N}_n - M_2} + \cdots + \frac{1-X_n}{\hat{N}_n - M_n} = \frac{n-1}{\hat{N}_n} \quad (4.7)$$

Para obter o EMV de N , as desigualdades (4.5) e (4.6) devem ser satisfeitas conjuntamente ou de maneira similar (4.7) deve ser solucionada.

A equação (4.7) pode ser expressa de forma mais reduzida desde que $M_{k+1} \geq M_k$ para todo k e $(1-X_k)=0$ ou 1, dependendo se a k -ésima escolha corresponde a um indivíduo marcado ou não, portanto, $M_{k+1} - M_k = 0$ ou 1. Logo, se R é definido como sendo o número de vezes, na amostra de tamanho n , nas quais escolheu-se um indivíduo marcado (recapturado), tem-se os seguintes resultados possíveis para valores de n de 2, 3, 4, e 5, com \hat{N}_n representado por y .

	$R=0$	$R=1$	$R=2$	$R=3$	$R=4$
$n=2$	$M_3=2$ $\frac{1}{y-1}=\frac{1}{y}$	$M_3=1$ $0=\frac{1}{y}$			
$n=3$	$M_4=3$ $\frac{1}{y-1}+\frac{1}{y-2}=\frac{2}{y}$	$M_4=2$ $\frac{1}{y-1}=\frac{2}{y}$	$M_4=1$ $0=\frac{2}{y}$		
$n=4$	$M_5=4$ $\frac{1}{y-1}+\frac{1}{y-2}+\frac{1}{y-3}=\frac{3}{y}$	$M_5=3$ $\frac{1}{y-1}+\frac{1}{y-2}=\frac{3}{y}$	$M_5=2$ $\frac{1}{y-1}=\frac{3}{y}$	$M_5=1$ $0=\frac{3}{y}$	
$n=5$	$M_6=5$ $\frac{1}{y-1}+\frac{1}{y-2}+\frac{1}{y-3}+\frac{1}{y-4}=\frac{4}{y}$	$M_6=4$ $\frac{1}{y-1}+\frac{1}{y-2}+\frac{1}{y-3}=\frac{4}{y}$	$M_6=3$ $\frac{1}{y-1}+\frac{1}{y-2}=\frac{4}{y}$	$M_6=2$ $\frac{1}{y-1}=\frac{4}{y}$	$M_6=1$ $0=\frac{4}{y}$

Note-se que dado $R=r$, para um tamanho de amostra n , as r recapturas podem ocorrer em quaisquer das $\binom{n-1}{r}$ posições produzindo a mesma equação e portanto a mesma estimativa de N .

Com base nos resultados da tabela acima, (4.7) pode-se escrever como

$$\sum_{j=1}^M \frac{1}{\hat{N}_n - j} = \frac{n-1}{\hat{N}_n} \quad (4.8)$$

onde M representa a $M_{k+1} - 1$.

Da maneira similar, pela definição de M_{k+1} , as inequações (4.5) e (4.6) podem ser expressas de forma mais reduzida como

$$\begin{aligned} (\hat{N}_n)^{n-1} (\hat{N}_n - M_{n+1}) &\leq (\hat{N}_n - 1)^n \\ (\hat{N}_n + 1)^{n-1} (\hat{N}_n - M_{n+1} - 1) &\geq (\hat{N}_n)^n \end{aligned}$$

ou equivalentemente,

$$\left(\frac{\hat{N}_n - M_{n+1}}{\hat{N}_n - 1} \right) - \left(\frac{\hat{N}_n - 1}{\hat{N}_n} \right)^{n-1} \leq 0 \quad (4.9)$$

$$\left(\frac{\hat{N}_n}{\hat{N}_n + 1} \right)^{n-1} - \left(\frac{\hat{N}_n - M_{n+1} - 1}{\hat{N}_n} \right) \leq 0 \quad (4.10)$$

A equação (4.8), quando solucionada utilizando o “software” matemático *Mathematica*, fornece as raízes da equação $\frac{\partial}{\partial N} \log L_n(N)=0$, e portanto caberia verificar se efetivamente são pontos de máximo. As expressões (4.9) e (4.10) juntamente, utilizando o mesmo software, determinam graficamente a região da solução para \hat{N}_n (Gráfico 4.1). Ambas formas de solução tem limitações quanto a valores grandes de n .

Uma maneira alternativa de estimar N é a partir de $L_n(n)$ em (4.4), o que elimina as limitações citadas acima. Para a expressão do $\log L_n(N)$, o mesmo software encontra graficamente o valor de \hat{N}_n que a maximiza, isto é o valor que faz zero a primeira derivada e que produz valor negativo para a segunda derivada (Gráfico 4.2). As expressões são as seguintes,

$$\begin{aligned}\log L_n(N) &= -(n-1) \log N + \sum_{k=2}^n X_k \log M_k + \sum_{k=2}^n (1-X_k) \log (N-M_k) \\ \frac{\partial}{\partial N} \log L_n(N) &= -\frac{n-1}{N} + \sum_{k=2}^n \frac{1-X_k}{N-M_k} \\ \frac{\partial^2}{\partial N^2} \log L_n(N) &= \frac{n-1}{N^2} - \sum_{k=2}^n \frac{1-X_k}{(N-M_k)^2}.\end{aligned}$$

4.2.1 Distribuição Assintótica do Estimador Sequencial para amostra de Tamanho Fixo

Sen (1982) tem estudado as propriedades assintóticas do estimador sequencial \hat{N}_n , as quais detalham-se a seguir:

Seja $\beta = \beta(X_i, i \leq k)$ a σ -álgebra gerada por $\{X_i, i \leq k\}$ para $k \geq 1$. Então por (4.1) e (4.3), para $k \geq 1$,

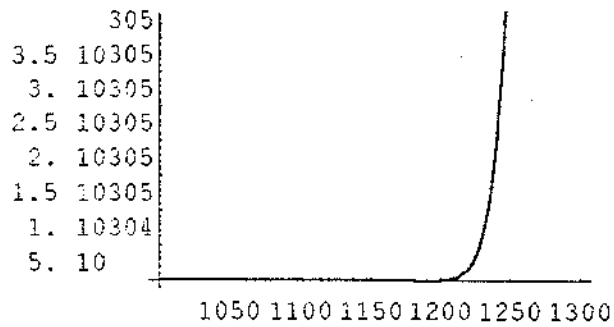
$$M_{k+1}^o = E M_{k+1} = E \{ E (M_{k+1} | \beta) \}.$$

Dado que condicionado ao valor de M_k , X_k pode assumir os valores 1 e 0 com probabilidades M_k/N e $1 - M_k/N$ respectivamente, então

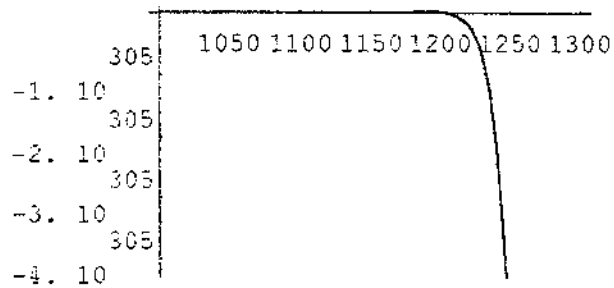
$$\begin{aligned}M_{k+1}^o &= E \left\{ M_k + \frac{N-M_k}{N} \right\} = 1 + \frac{N-1}{N} E M_k \\ &= \sum_{i=0}^k \left(\frac{N-1}{N} \right)^i = N \left\{ 1 - \left(\frac{N-1}{N} \right)^k \right\} \\ &= N \left\{ 1 - \left(1 - \frac{1}{N} \right)^k \right\}\end{aligned}\tag{4.11}$$

Gráfico 4.1 Região de solução para \hat{N}_n

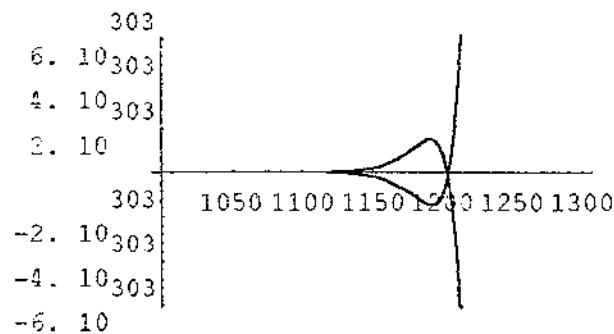
```
a=Plot[(x^99)(x-96)-(x-1)^100,{x,1000,1300}]
b=Plot[x^100-((x+1)^99)(x-95),{x,1000,1300}]
Show[a,b]
```



-Graphics-



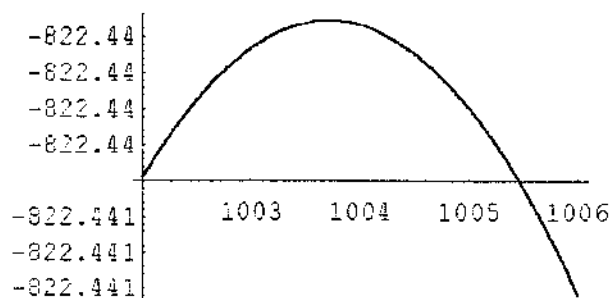
-Graphics-



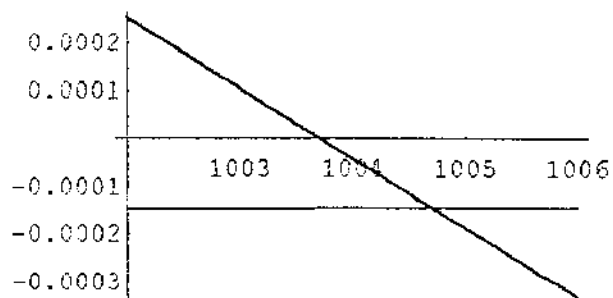
-Graphics-

Gráfico 4.2 Solução gráfica maximizando a verossimilhança

```
f=-499Log[y]+Sum[Log[y-i]},{i,1,393}];
g=Dt[f,y];
h=Dt[f,{y,2}];
Plot[f,{y,1002,1006}]
Plot[{g,h},{y,1002,1006}]
```



-Graphics-



-Graphics-

Similarmente,

$$\begin{aligned}
\delta_{k+1}^2 &= \text{Var}\{M_{k+1}\} = E \{ [M_{k+1} - E(M_{k+1} | \beta)] + [E(M_{k+1} | \beta) - M_{k+1}^o] \}^2 \\
&= E \{ M_{k+1}^2 - [E(M_{k+1} | \beta)]^2 - 2M_{k+1}M_{k+1}^o + 2E(M_{k+1} | \beta)M_{k+1}^o \\
&\quad + [E(M_{k+1} | \beta) - M_{k+1}^o]^2 \} \\
&= E M_{k+1}^2 - M_{k+1}^{o2} - E [E(M_{k+1} | \beta)]^2 + M_{k+1}^o E [E(M_{k+1} | \beta)] - M_{k+1}^{o2} \\
&\quad - M_{k+1}^2 + M_{k+1}^o E [E(M_{k+1} | \beta)] + E [E(M_{k+1} | \beta) - M_{k+1}^o]^2 \\
&= E \left\{ \frac{M_k(N - M_k)}{N^2} \right\} + \left(\frac{N-1}{N} \right)^2 \delta_k^2 \\
&= \sum_{i=0}^{k-2} \left(\frac{N-1}{N} \right)^{2i} E \left\{ \frac{M_{k-i}(N - M_{k-i})}{N^2} \right\} \\
&\leq \frac{N}{4} \left(1 - \left(1 - \frac{1}{N} \right)^{2(k-1)} \right) < \frac{N}{4}
\end{aligned} \tag{4.12}$$

adicionalmente,

$$n = N\alpha, \quad 0 < \alpha < K^*, \quad K^* < \infty, \tag{4.13}$$

onde α não precisa ser menor que 1 pois a amostragem sequencial pode continuar enquanto $M_n < N$, e ainda para $n > N$, a probabilidade de $M_n < N$ é positiva. Então, por (4.11) e (4.13),

$$\begin{aligned}
E\left(\frac{M_n}{N}\right) &= 1 - \left(1 - \frac{1}{N}\right)^{n-1} \\
&= 1 - \left(1 - \frac{\alpha}{N}\right)^{n-1}
\end{aligned}$$

por tanto, $\frac{M_n}{N}$ aproxima-se de $1 - e^{-\alpha}$. Também, por (4.12)

$$V\left(\frac{M_n}{N}\right) \leq \frac{1}{4N}, \tag{4.14}$$

de forma que pela desigualdade de Chevyshev, $\frac{M_n}{N} \rightarrow 1 - e^{-\alpha}$ em probabilidade, conforme $n \rightarrow \infty$.

Continuando o procedimento demonstrado em Sen (1982), tem-se que conforme n cresce,

$$N^{-1/2} (\hat{N}_n - N) \sim \mathcal{N} \left(0, \frac{1}{e^\alpha - \alpha - 1} \right) \quad (4.15)$$

Considera-se agora a eficiência assintótica relativa (ARE) do estimador de \hat{N}_n com respeito a \hat{N}_p definido em (2.3), ambos para amostras de tamanho n , satisfazendo (4.13), sendo que α não é necessariamente menor que 1 (mas sim ≤ 2), isto é, $K^*=2$, tem-se

$$\begin{aligned} e(\hat{N}_p, N_n) &= \frac{\text{Variância assintótica de } (\hat{N}_p - N)/N^{1/2}}{\text{Variância assintótica de } (\hat{N}_n - N)/N^{1/2}} \\ &= \sigma_2^2 (e^\alpha - \alpha - 1) \\ &= \frac{(1 - \alpha_1)(1 - \alpha_2)}{\alpha_1 \alpha_2} (e^{-\alpha_1 - \alpha_2} - \alpha_1 - \alpha_2 - 1) \end{aligned} \quad (4.16)$$

e para um valor $\alpha (= \alpha_1 + \alpha_2)$ dado, e é um mínimo para $\alpha_1 = \alpha_2 = \alpha/2$ e seu valor está dado por

$$\left(\frac{2 - \alpha}{\alpha} \right)^2 (e^\alpha - \alpha - 1). \quad (4.17)$$

Note-se que conforme $\alpha \downarrow 0$, e converge a 2, o que indica que para α pequeno (ou seja, $n/N \downarrow 0$) o estimador de Petersen nunca terá mais do que 50% da eficiência do estimador sequencial \hat{N}_n . Por outro lado, a figura é diferente quando α não é pequeno. Enquanto $\alpha \leq 0,7656$, e em (4.15) é maior que 1, mas, se α assume valores maiores de 0,7656 (até o limite superior 2), e permanecerá menor de 1 e tenderá a 0 conforme $\alpha \rightarrow 2$. Então a maior eficiência do estimador de C-R sequencial é claramente mostrada respeito ao estimador clássico de Petersen para duas amostras quando α é menor que 0,7656 (isto é, cada amostra de tamanho $n_1 = n_2 = 0,383N$). Na prática, geralmente n_1/N e n_2/N são ambos pequenos, portanto a amostragem sequencial vai ser mais conveniente.

4.2.2 Simulação de um experimento de C-R de Amostragem Sequencial e cálculo das Estimativas do Tamanho da População e Intervalos de Confiança Assintóticos.

Na simulação de um experimento de C-R de amostragem sequencial para um tamanho de população N e tamanho de amostra n , foram seguidos os passos abaixo.

1. Atribuir valor 0 aos N indivíduos da população e definir $M=0$.
2. Selecionar aleatoriamente um indivíduo dos N e atribuir valor 1 àquele selecionado.
3. Fazer $M=M+1$ se o indivíduo selecionado tiver valor 0, ou $M=M$ se tiver valor 1.
4. Repetir 2 e 3 até o número de vezes n .
4. Encontrar o número R de vezes que foi selecionado um indivíduo com valor 1.

Pode ser notado que M conta o número de vezes que é selecionado um indivíduo novo ou não marcado. Também pode-se notar que $n=R+M$, é dizer, o tamanho de amostra é igual ao número de indivíduos recapturados somado ao número total de indivíduos marcados durante o experimento ou antes da $(n+1)$ -ésima captura. M é equivalente a M_{n+1} definido em (4.1).

Para o tamanho de população de 1000 indivíduos foram considerados os tamanhos de amostra de 100, 200, 300, 400 e 500. Para cada tamanho amostral foram simulados 1000 experimentos obtendo-se as frequências de R , (ou equivalentemente de M_{n+1}), logo é encontrado o valor da estimativa de \hat{N}_n mediante o software *Mathematica*, o valor de $v(\hat{N}_n)$ é encontrado da expressão (4.15) e finalmente é calculado um intervalo de 95% de confiança de aproximação normal utilizando o valor 1,96 da tabela da distribuição normal padrão. Nas tabelas seguintes são apresentados os resultados das simulações por tamanho amostral.

Observa-se que para os valores de R iguais a 0 e n , não é possível estimar N porque não existe um valor para \hat{N}_n (maior que M_{n+1}) que maximize a verossimilhança em (4.7). Também, pode-se observar que conforme R cresce, os valores para \hat{N}_n e $v(\hat{N}_n)$ são decrescentes .

TABELA 17. Simulação de 1000 experimentos de amostragem
sequencial para o tamanho de população 1000
e tamanho de amostra 100

R	Freq(%)	M_{n+1}	\hat{N}_n	$\sqrt{v(\hat{N}_n)}$	Int. de Confiança	
0	1,8	100	--	--	--	--
1	3,2	99	4917	4859,7	0	14442
2	8,4	98	2442	1694,9	0	5764
3	13,7	97	1617	910,1	0	3400
4	20,6	96	1204	582,0	62	2346
5	16,7	95	957	411,4	150	1763
6	15,2	94	792	308,6	187	1397
7	10,0	93	674	241,3	200	1147
8	5,6	92	585	194,4	204	966
9	2,8	91	516	160,4	202	830
10	0,9	90	461	134,9	197	725
11	0,8	89	416	115,2	190	642
12	0,3	88	379	99,8	183	575

TABELA 18. Simulação de 1000 experimentos de amostragem
sequencial para o tamanho de população 1000
e tamanho de amostra 200

R	Freq(%)	M_{n+1}	\hat{N}_n	$\sqrt{v(\hat{N}_n)}$	Int. de Confiança	
8	0,4	192	2421	830,7	792	4049
9	0,1	191	2144	691,1	789	3492
10	1,0	190	1923	586,0	775	3071
11	1,5	189	1742	504,3	754	2730
12	1,8	188	1591	439,3	730	2452
13	4,5	187	1464	387,1	705	2223
14	4,9	186	1354	343,6	680	2028
15	8,2	185	1260	307,9	657	1863
16	10,0	184	1176	277,1	633	1719
17	11,8	183	1103	251,2	611	1595
18	9,3	182	1038	228,9	589	1487
19	10,3	181	980	209,6	569	1390
20	8,6	180	927	192,4	550	1304
21	8,5	179	880	177,6	532	1228
22	5,3	178	837	164,4	515	1159
23	4,2	177	797	152,4	498	1096
24	3,5	176	761	141,9	483	1039
25	2,9	175	728	132,5	468	988
26	1,9	174	697	123,9	454	940
27	1,1	173	669	116,3	441	897
31	0,1	169	574	91,6	394	754
32	0,1	168	553	86,4	384	722

TABELA 19. Simulação de 1000 experimentos de amostragem
sequencial para o tamanho de população 1000
e tamanho de amostra 300

R	Freq(%)	M_{n+1}	\hat{N}_n	$\sqrt{v(\hat{N}_n)}$	Int. de Confiança	
24	0,2	276	1728	328,8	1084	2372
26	0,2	274	1624	299,0	1038	2210
27	0,4	273	1560	281,1	1009	2111
28	0,2	272	1500	264,7	981	2019
29	0,9	271	1445	250,0	955	1935
30	1,2	270	1394	236,6	930	1858
31	1,8	269	1345	223,9	906	1784
32	2,6	268	1300	212,5	884	1716
33	3,7	267	1257	201,7	862	1652
34	2,9	266	1217	191,9	841	1593
35	4,1	265	1180	183,0	821	1539
36	5,8	264	1144	174,4	802	1486
37	7,5	263	1110	166,5	784	1436
38	7,1	262	1078	159,1	766	1390
39	6,4	261	1048	152,3	749	1347
40	6,7	260	1019	155,8	733	1305
41	8,0	259	992	139,9	718	1266
42	5,9	258	966	134,2	703	1229
43	6,5	257	941	128,8	688	1194
44	6,7	256	917	123,8	674	1160
45	4,5	255	894	119,0	661	1127
46	4,0	254	872	114,4	648	1096
47	3,5	253	852	110,4	636	1068
48	2,9	252	832	106,3	624	1040
49	0,6	251	813	102,6	612	1014
50	2,6	250	794	98,8	600	988
51	1,5	249	776	95,3	589	963
52	0,6	248	760	92,3	579	941
53	0,3	247	743	89,1	568	918
54	0,4	246	727	86,1	558	896
56	0,2	244	698	80,7	540	856
58	0,1	242	670	75,7	522	818

TABELA 20. Simulação de 1000 experimentos de amostragem
sequencial para o tamanho de população 1000
e tamanho de amostra 400

R	Freq(%)	M_{n+1}	\hat{N}_n	$\sqrt{v(\hat{N}_n)}$	Int. de Confiança	
52	0,2	348	1398	176,0	1053	1743
55	0,4	345	1314	159,9	1001	1627
56	0,6	344	1289	155,2	985	1593
57	0,2	343	1264	150,5	969	1559
58	0,6	342	1240	146,1	954	1526
59	1,2	341	1216	141,7	938	1494
60	2,9	340	1193	137,5	923	1463
61	2,5	339	1171	133,6	909	1433
62	2,9	338	1150	129,9	895	1405
63	2,0	337	1130	126,4	882	1378
64	4,7	336	1110	122,9	869	1351
65	4,6	335	1091	119,6	857	1325
66	5,4	334	1072	117,4	842	1302
67	5,4	333	1054	114,3	830	1278
68	6,6	332	1035	110,1	819	1251
69	6,8	331	1019	107,5	808	1230
70	5,8	330	1003	104,8	797	1209
71	5,4	329	986	102,1	786	1186
72	6,4	328	971	99,6	776	1166
73	5,5	327	956	97,2	765	1147
74	5,3	326	941	95,6	754	1128
75	3,4	325	926	92,5	745	1107
76	4,5	324	912	90,3	735	1089
77	4,2	323	898	88,1	725	1071
78	2,9	322	885	86,1	716	1054
79	3,3	321	872	84,1	707	1037
80	1,7	320	859	82,1	698	1020
81	1,3	319	847	80,3	690	1004
82	1,3	318	835	78,5	681	989
83	0,5	317	823	76,7	673	973
84	0,6	316	812	75,1	665	959
85	0,3	315	800	73,3	656	944
86	0,3	314	789	71,7	648	930
88	0,2	312	768	68,7	633	903
90	0,1	310	748	65,9	619	877

TABELA 21. Simulação de 1000 experimentos de amostragem
sequencial para o tamanho de população 1000
e tamanho de amostra 500

R	Freq(%)	M_{n+1}	\hat{N}_n	$\sqrt{v(\hat{N}_n)}$	Int. de Confiança	
87	0,1	413	1262	118,40	1030	1494
88	0,1	412	1246	116,1	1018	1474
89	0,1	411	1230	113,1	1007	1453
91	0,7	409	1199	109,3	985	1413
92	1,0	408	1184	107,1	974	1394
93	0,5	407	1169	105,0	963	1375
94	1,2	406	1155	103,0	953	1357
95	0,9	405	1141	101,1	943	1339
96	2,0	404	1127	99,1	933	1321
97	1,8	403	1114	97,3	923	1305
98	3,5	402	1100	95,4	913	1287
99	3,0	401	1087	93,6	904	1270
100	3,6	400	1075	92,0	895	1255
101	3,6	399	1062	90,2	885	1239
102	3,6	398	1050	88,6	875	1224
103	4,6	397	1038	87,0	865	1209
104	5,6	396	1026	85,4	855	1193
105	5,6	395	1015	84,0	855	1180
106	6,6	394	1004	82,5	845	1166
107	5,1	393	993	81,1	834	1152
108	5,0	392	982	79,7	826	1138
109	5,9	391	971	78,2	818	1124
110	5,1	390	961	77,0	810	1112
111	3,4	389	950	75,6	802	1098
112	4,0	388	940	74,3	794	1086
113	4,6	387	930	73,0	787	1073
114	3,4	386	921	71,9	780	1062
115	2,2	385	911	70,7	772	1050
116	3,9	384	902	69,6	766	1038
117	0,8	383	892	68,3	758	1026
118	2,3	382	883	67,2	751	1015
119	1,4	381	874	66,1	744	1004
120	0,9	380	865	65,0	738	992
121	1,2	379	857	64,1	731	983
122	1,0	378	848	63,0	725	971
123	1,2	377	840	62,3	718	962
125	0,3	375	823	60,0	705	941
126	0,6	374	815	59,1	699	931
129	0,2	371	792	56,4	681	903

4.2.3 Intervalos de Confiança Bootstrap

A seguir é descrita a metodologia bootstrap aplicada a um experimento de C-R de amostragem seqüencial a fim de obter um intervalo de confiança para N baseado no percentil bootstrap.

Na idéia do bootstrap de reutilizar a amostra de forma que esta faça as vezes de uma “nova população”, (seção 1.2 Cap. 1), serão utilizadas as estatísticas R e M_{n+1} , resultantes da simulação de um experimento. Esta população fica conformada por n indivíduos dos quais R são “0” e $(n - R)$ “1”. A partir dela são selecionadas B amostras com reposição, logo para cada uma delas encontram-se os R_j^* e $(M_{n+1}^*)_j$, ($j=1, \dots, B$) as quais são utilizadas no cálculo das B estimativas \hat{N}_n^* . A distribuição das B estimativas é usada para definir limites de confiança para N , como estabelecido na seção 1.3 do capítulo 1.

Na Tabela 22 são apresentados os intervalos de confiança baseados no percentil bootstrap para os tamanhos de amostra de 100, 200, 300, 400 e 500 e alguns resultados por cada tamanho.

TABELA 22. Intervalos bootstrap de 95%
de confiança.

n	R	Lim. Inf.	Lim. Sup.	Comprimento
100	4	516	4917	4401
	5	516	2442	1926
	6	461	2442	1981
200	16	797	2144	1347
	17	697	1742	1045
	19	669	1464	795
300	37	852	1624	772
	38	832	1500	668
	41	794	1394	600
400	68	823	1289	466
	69	800	1289	489
	72	800	1314	514
500	105	815	1230	415
	106	857	1155	298
	109	807	1100	293

A modo de ilustração, a Tabela 22A apresenta a distribuição de frequências de $B=1000$ replicações bootstrap para o tamanho de amostra 100 com $R=4$, $M_{101}=96$, referentes ao intervalo de confiança do percentil da primeira linha da Tabela 22.

TABELA 22A. Replicações Bootstrap para o
resultado $R=4$ e $M_{101}=96$.

R	Freq	Perc(%)	Freq.Acum.	Perc.Acum.
0	19	1,9	19	1,9
1	93	9,3	112	11,2
2	128	12,8	240	24,0
3	191	19,1	431	43,1
4	179	17,9	610	61,0
5	159	15,9	769	76,9
6	112	11,2	881	88,1
7	60	6,0	941	94,1
8	26	2,6	967	96,7
9	19	1,9	986	98,6
10	7	0,7	993	99,3
11	4	0,4	997	99,7
12	3	0,3	1000	100,0

Comparando os intervalos de confiança assintóticos com os intervalos bootstrap na amostragem sequencial, pode-se notar que para tamanhos de amostra menores (100, 200 e 300) os comprimentos dos intervalos obtidos por bootstrap são maiores do que os comprimentos dos intervalos, já para o tamanho de amostra de 500, os comprimentos dos intervalos bootstrap são menores.

4.3 Amostragem Sequencial com o Tamanho de Amostra Estocástico.

Na seção anterior foi tratado o procedimento de amostragem sequencial considerando o caso do tamanho da amostra n fixo e $n/N \rightarrow \alpha$, para algum $\alpha > 0$. Nesta seção é tratado o procedimento quando o tamanho da amostra não é fixo. Neste contexto vai ser definida a regra de parada a qual produz uma amostra de tamanho n ,

assumindo valores inteiros não negativos.

Tem-se que a função de probabilidade definida em (4.3) pode também ser descrita convenientemente em termos de um modelo de urna simples. Suponha-se que em uma urna existam N bolas, todas brancas, onde N é desconhecido. Repetidamente, escolhe-se estas bolas aleatoriamente uma a uma, observa-se sua cor e substitui-se por uma bola preta, de forma que antes de qualquer escolha, existam N bolas na urna.

Seja W_n o número de bolas brancas observadas nas n primeiras escolhas, $n \geq 1$. Note-se que $W_n \leq n$ para cada $n \geq 1$, assim como $W_0=0$ e $W_1=1$. Também, W_n é não decrescente em n (≥ 1) e pode-se definir M_k como em (4.1) e (4.2) e notar que

$$W_k = M_{k+1} \text{ para cada } k \geq 1 \quad (4.18)$$

Também, note-se que para cada $k \geq 1$, $k - W_k$ é o número de escolhas (nas k primeiras) nas quais apareceu uma bola preta. Então, o esquema de captura sequencial pode ser descrito equivalentemente em termos da sequência $\{W_k, k \geq 1\}$.

Samuel (1968), no contexto da estimação sequencial de N , tem considerado a variável de parada t_c assim como tem feito a conjectura de que sua forma padronizada tem distribuição assintoticamente normal conforme $N \rightarrow \infty$, baseado no fato de W_k ser não decrescente em n , isto é, $W_n \geq W_{n-1}$, para cada $n \geq 1$. A variável de parada é definida a seguir.

Para cada $c > 0$, fixo,

$$\begin{aligned} t_c &= \inf \{n : n \geq (c+1)W_n\} \\ &= \inf \{n : (n - W_n)/W_n \geq c\} \end{aligned} \quad (4.19)$$

onde t_c pode tomar somente os valores $[k(c+1)]$, $k=1,2,\dots$, e $W_{t_c}=m$ sempre que $t_c=[m(c+1)]$, $m \geq 1$. Aqui $[x]$ denota o maior inteiro menor ou igual que x . Desde que $W_n \leq N$, a regra de parada é limitada, portanto pára com probabilidade 1.

O resultado a estabelecer provem do procedimento seguido em Sen (1982 a), nele consideram-se princípios de invariância para a sequência parcial $\{W_k, k \leq n\}$ incorporando-os na formulação de importantes teoremas de convergência para as variáveis de parada $\{t_c; c > 0\}$. Neste contexto, ele considera uma representação de W_n em termos de uma combinação linear de diferenças de martingales. Sen (1982 b) com base no anterior, estabelece a normalidade assintótica do EMV de N para o tamanho de amostra com parada aleatória t_c .

A distribuição limite achada para o EMV de N é a mesma de (4.15) só que no caso de n estocástico tem-se uma sequência $\{v_n\}$ de variáveis aleatórias tendo valores inteiros positivos, tais que

$$\frac{v_n}{n} \rightarrow 1, \text{ conforme } n \rightarrow \infty \quad (4.20)$$

então pode-se concluir que sob (4.13),

$$| N^{-1/2} (\hat{N}_{v_n} - N) - N^{-1/2} (\hat{N}_n - N) | \xrightarrow{p} 0 \quad (4.21)$$

e daí,

$$N^{-1/2} (\hat{N}_{v_n} - N) \sim N \left(0, \frac{1}{e^\alpha - \alpha - 1} \right) \quad (4.22)$$

Agora, se $n (= n_N)$ é tal que $N^{-1} n_N^2 \rightarrow \infty$ conforme $N \rightarrow \infty$, sendo que $N^{-1} n_N$ pode não convergir a um número positivo, então no limite deste caso degenerado, $N^{-1} n_N \rightarrow 0$,

$$N^{-1/2} n (N^{-1} \hat{N}_{v_n} - 1) \sim N(0, 2) \quad (4.23)$$

Note-se de passagem, que para o estimador de Petersen de duas amostras, sempre que $n_1 = n_2 = n/2$ e se $n (= n_N)$ crescente, mas $N^{-1} n_N \rightarrow 0$ conforme $N \rightarrow \infty$, tem-se

$$N^{-1/2} n (N^{-1} \hat{N}_p - 1) \sim N(0, 4) \quad (4.24)$$

A modificação anterior também aplica-se à normalidade assintótica no caso da amostragem de recapturas múltiplas, sempre que $N^{-1} n_N \rightarrow 0$, quando $N \rightarrow \infty$.

4.3.1 Simulação de um experimento de C-R de amostragem Sequencial com tamanho de amostra estocástico.

A seguir é descrito o procedimento de simulação de um experimento de C-R de amostragem sequencial quando o tamanho de amostra é estocástico. Assumindo uma população de N indivíduos e um tamanho de amostra n , os passos seguidos neste procedimento foram,

1. Atribuir valor 0 (cor branca às bolas da urna) aos N indivíduos da população.
2. Definir a variável W_n indicando o número de bolas brancas observadas.

3. Estabelecer os valores de c , constante da variável de parada.
4. Selecionar os indivíduos um a um até a variável de parada chegar ao valor cW_n .
5. Recuperar o valor do tamanho de amostra.

Na seguinte tabela são apresentados os resultados da simulação de um experimento de C-R de amostragem sequencial com tamanho amostral estocástico. Para o tamanho de população 1000, são estabelecidos valores de W_n , o número de bolas brancas observadas nas primeiras n escolhas. Também foram estabelecidos diversos valores de c , sugerido na literatura (Sen, 1982a), como um valor adequadamente grande.

TABELA 23. Simulação de Experimentos de amostragem sequencial e tamanho de amostra estocástico.

c	W_n	n	c	W_n	n
10	10	121	60	10	140
	20	191		20	223
	30	256		30	235
	40	288		40	303
	50	310		50	325
20	10	139	70	10	175
	20	190		20	195
	30	242		30	247
	40	292		40	261
	50	321		50	317
30	10	121	80	10	175
	20	213		20	216
	30	230		30	237
	40	290		40	296
	50	318		50	311
40	10	138	90	10	177
	20	190		20	224
	30	248		30	234
	40	347		40	311
	50	353		50	341
50	10	97	100	10	117
	20	222		20	188
	30	292		30	242
	40	293		40	301
	50	333		50	322

Pode-se observar que o tamanho amostral, n , e em consequência o EMV de N ,

\hat{N}_{v_n} , depende de W_n , o número de bolas brancas (ou número de indivíduos diferentes capturados no experimento) observadas na parada, e não da regra de parada t_c , embora sua distribuição sim dependa da regra. Logo, para qualquer valor de c , n apresenta comportamento similar, sendo função de W_n . Na tabela seguinte são apresentadas as estimativas de \hat{N}_{v_n} assim como os intervalos de confiança assintóticos para o valor de $c=10$.

TABELA 24. Estimação de N pontual e por intervalo na amostragem sequencial de tamanho de amostra estocástico.

c	W_n	n	\hat{N}_{v_n}	$\sqrt{v(\hat{N}_{v_n})}$	Int. de Confiança	
10	10	121	766	241,3	293	1238
	20	191	890	189,6	518	1262
	30	256	1039	177,4	691	1387
	40	288	962	139,2	689	1235
	50	310	872	110,5	655	1089

Pode-se notar que os comprimentos dos intervalos de confiança decrescem conforme o tamanho de amostra se incrementa. Também, estes intervalos quando comparados com aqueles obtidos na amostragem sequencial de tamanho de amostra fixo, resultaram de menor comprimento e mais acurados, para tamanhos de amostra comparáveis.

CAPÍTULO 5

CONSIDERAÇÕES FINAIS

Neste trabalho, procurou-se estabelecer os aspectos da aplicação da metodologia bootstrap aos modelos de C-R para estimação do tamanho de uma população animal, N a fim de avaliar os estimadores obtidos. Desde que os modelos de C-R podem estar dirigidos a populações fechadas ou abertas, modelos para a primeira população em menção foram analisados quando a amostragem é direta e inversa, tanto para duas amostras (ou ocasiões) como para múltiplas amostras. Também é analisado o caso de amostragem sequencial para os tamanhos de amostra fixo e estocástico.

No estudo tanto do caso de duas amostras quanto de múltiplas amostras bem como amostragem sequencial é apresentada a distribuição assintótica do estimador de N , completando-se a demonstração esquematizada em Sen (1987). Em todos os casos comparam-se as medidas de variabilidade de \hat{N} .

A metodologia bootstrap mostrou-se de utilidade na estimação da variabilidade das estimativas, assim como no fornecimento de intervalos de confiança para N . Para a aplicação do bootstrap paramétrico em modelos de amostragem direta, foram consideradas duas análises: condicional aos tamanhos observados das amostras, em cujo caso a distribuição que gera as replicações bootstrap é assumida hipergeométrica, e não condicional a estes tamanhos, assumindo neste caso a distribuição multinomial das histórias de captura (Buckland e Garthwaite (1991)). O melhor desempenho de uma análise com respeito a outra não é evidente, as vantagens poderiam ser notadas em termos dos recursos computacionais utilizados.

Para a aplicação do bootstrap feita em termos de simulações, considerou-se um conjunto de experimentos de C-R de amostragem direta, de duas e três ocasiões e com tamanhos de amostra comparáveis. O desempenho do método bootstrap melhora para tamanhos maiores de população. A distribuição do estimador de N nas replicações

bootstrap para uma população maior é mais simétrica, isto último significa a não necessidade de fazer a correção pelo vício do intervalo de percentil obtido.

A estimação do verdadeiro nível de confiança dos intervalos percentis, feita por simulação para os experimentos de duas amostras, revela uma proximidade do valor teórico.

Para um mesmo tamanho de população, e tamanhos de amostra comparáveis, o modelo de três amostras proporciona melhores estimativas, tanto pontuais quanto por intervalo, comparado com o modelo de duas amostras. Nos modelos de três amostras as variâncias assintóticas (Sen, 1987) resultaram de menor valor para todos os experimentos.

Em todos os experimentos de três amostras, as estimativas de Petersen multiamostrais \hat{N}_{pp} , resultaram menores do que as de Leite \hat{N}_m , assim como as variâncias bootstrap resultaram menores e os intervalos de menor comprimento.

Em relação à amostragem seqüencial é de importância ter chegado por simulação, às distribuições das estatísticas dos experimentos de forma de abranger os possíveis resultados e as estimativas do tamanho da população. Comprovou-se que com tamanhos menores de amostra pode-se obter estimativas mais acuradas. Portanto, sempre que seja praticável recomendaria-se a amostragem seqüencial.

Também em relação à amostragem seqüencial ainda devem ser estudados os enfoques presentes em Sen (1987), com o objetivo de encontrar um intervalo de confiança de amplitude de percentagem fixo assim como um intervalo de custo e risco mínimo.

APÊNDICE

Nesta seção são apresentados os programas computacionais citados através dos capítulos desta dissertação. Os programas, desenvolvidos utilizando o “software” estatístico *SAS* (Statistical Analysis System) fazem uso de variáveis e procedimentos “macro”, identificados por %.

PROG1.SAS

```
/* .....  
SIMULACAO DE UM EXPERIMENTO DE C-R DE DUAS AMOSTRAS CALCULO DA  
ESTIMATIVA DE CHAPMAN, VARIANCIA, E INTERVALO DE CONFIANCA ASSINTOTICO  
..... */  
  
filename amostra'a:\amostra.dat';  
  
%let npop=1000; /*TAMANHO DA POPULACAO */  
%let s=2; /* NUMERO DE AMOSTRAS A SEREM SELECIONADAS */  
%let na1=150; /* TAMANHO DA PRIMEIRA AMOSTRA */  
%let na2=160; /* TAMANHO DA SEGUNDA AMOSTRA */  
%let sem1=1111; /* SEMENTE PARA GERACAO DA AMOSTRA ALEATORIA */  
  
%macro prog1; /*INICIO DO MACRO “PROG1” DISPENSAVEL SE DADOS REAIS FOSSEM A  
*/  
SER ANALISADOS */  
file amostra;  
  
array a{&s,&npop};  
do j=1 to &s;  
  do i=1 to &npop;  
    a{j,i}=0; /* POPULACAO INICIALMENTE COM VALORES “0” */  
  end;  
end;  
array ni{&s};  
ni1=&na1; ni2=&na2;  
do j=1 to &s;  
  do k=1 to ni{j};  
    verifica='f'; /* VERIFICACAO DA SELECAO DE UMA AMOSTRA SEM REPOSICAO */  
    do until (verifica='v');  
      u=int(ranuni(&sem1)*&npop)+1;  
      if a{j,u}=0 then do;  
        a{j,u}=1;  

```

```

        verifica='v';
    end;
end;
end;
end;
array mi{&s};
do j=1 to &s;
    num=0;
    do i=1 to &npop;
        m=0;
        if a{j,i} = 1 then do;
            do k=1 to j-1;
                m=m+a{k,i};
            end;
            if m >= 1 then num=num+1; /* CONTAGEM DE INDIVIDUOS SELECIONADOS NA
                                     SEGUNDA E PRIMEIRA AMOSTRA*/
        end;
    end;
    mi{j}=num;
end;
array mmi{&s};
do j=1 to &s;
    num=0;
    do k=1 to j;
        num=num+(ni{k}-mi{k});
    end;
    mmi{j}=num; /* NUMERO TOTAL DE INDIVIDUOS DIFERENTES CAPTURADOS NO
                  EXPERIMENTO*/
end;

* ESTIMACAO DO TAMANHO DA POPULACAO;

nest=int(((mmi{1}+1)*(ni{2}+1)/(mi{2}+1)-1);
vnest=nest*((ni{1}-mi{2})*(ni{2}-mi{2}))/((mi{2}+1)*(mi{2}+2));
snest=vnest**.5;
linf=round(nest-1.96*snest);
if linf<0 then linf=0;
lsup=round(nest+1.96*snest);
tamanho=lsup-linf;
put nil-ni&s mil-mi&s mmil-mmi&s nest;

%mend prog1; /* FIM DO MACRO "PROG1"*/

data amostra; %prog1; /*CHAMADA A EXECUCAO DO MACRO PROG1*/

proc print;
    var nil-ni&s mi&s mmil-mmi&s nest snest;
    title 'ESTATISTICAS DO EXPERIMENTO DE RECAPTURA SIMPLES';
proc print;

    var linf lsup tamanho;
    title 'INTERVALO DE CONFIANCA NORMAL';
run;

```

PROG2.SAS

```

/* .....
      CALCULO DA VARIANCIA BOOTSTRAP, INTERVALO DO PERCENTIL
      E INTERVALO DO PERCENTIL DE VICIO CORRIGIDO
..... */

```

```

filename inter1'a:\inter1.dat';
filename inter2'a:\inter2.dat';
filename inter3'a:\inter3.dat';

```

```

%let s=2;
%let na1=150;
%let na2=160;
%let max=2500;
%let sem2=5678;
%let zalfa=probit(.975);
%let alfa=.025;
%let b=1000;

```

```

*REPLICACOES BOOTSTRAP MULTINOMIAL (N, p01, p10, p11);

```

```

data multil;
infile 'a: ' ; /* INVOCA EXECUCAO DO MACRO PROG1 OU ENTRADA DE DADOS REAIS*/

```

```

input nil-ni&s mil-mi&s mmil-mmi&s nest;
file inter1;
N=nest;
NEb=0; n10=0; n01=0;
s=0; me=0; menores=0;
do j=1 to &b;
  soma1=0;
  soma2=0;
  n11=0;
  do i=1 to n;
    r=ranuni(0);
    if r <= (nil-mi2)/n then soma1=soma1+1;
    if r > (nil-mi2)/n & r <= nil/n then n11=n11+1;
    if r > nil/n & r <= (nil+ni2-mi2)/n then soma2=soma2+1;
  end;
  n10 = soma1+n11;
  n01 = soma2+n11;
  NEb = int((n10+1)*(n01+1)/(n11+1)-1);
  if neb<n then menores=menores+1;
  put neb;

```

```

* Variancia de NEb;
  aux=neb-me;
  s=s+(j-1)*(aux)**2/j;
  me=me+aux/j;
end;
s=(s/(&b-1))**.5;
keep s menores;
proc print;
  var s menores;
title'VARIANCIA BOOTSTRAP MULTINOMIAL (N,p10,p11,p01)';

```

```

data multi3;
set multi1;
prop=menores/&b;
zzero=probit(prop);
alfa1=probnorm(2*zzero-&zalfa);
alfa2=1-probnorm(2*zzero+&zalfa);
do k=1 to &b;
    output;
end;
keep alfa1 alfa2;

*Intervalo do percentil para neb;
data multi1;
infile inter1;
input neb;
proc sort data=multi1;
    by neb;

data multi5;
set multi1;
n=_n_;
if n=round((&b+1)*&alfa) then output;
if n=round((&b+1)*(1-&alfa)) then output;
proc print;
    var n neb;
title'INTERVALO DE CONFIANCA DO PERCENTIL';

data multi7;
merge multi1 multi3;
data multi7;
set multi7;
n=_n_;
if n=round((&b+1)*&alfa1) then output;
if n=round((&b+1)*(1-&alfa2)) then output;
proc print;
    var n neb;
title 'INTERVALO DE CONFIANCA DE VICIO CORRIGIDO';

*REPLICACOES BOOTSTRAP MULTINOMIAL (M, p01, p10, p11);
data multi2;
infile amostra;
input n1-ni&s mil-mi&s mmil-mmi&s nest;
file inter2;
M=mmi2; N=nest;
NEb=0; n10=0; n01=0;
s=0; me=0; menores=0;
do j=1 to &b;
    soma1=0;
    soma2=0;
    n11=0;
    do i=1 to m;
        r=(m/n)*ranuni(0);
        if r <= (n1-mi2)/n then soma1=soma1+1;
        if r > (n1-mi2)/n & r <= n1/n then n11=n11+1;

```

```

        if r > n11/n & r <= (m/n) then soma2=soma2+1;
    end;
    n10 = soma1+n11;
    n01 = soma2+n11;
    neb = int((n10+1)*(n01+1)/(n11+1)-1);
    if neb<n then menores=menores+1;
    put neb;
* Variancia de NEb;
    aux=neb-me;
    s=s+(j-1)*(aux)**2/j;
    me=me+aux/j;
end;
s=(s/(&b-1))**.5;
keep s menores;
proc print;
    var s menores;
title 'VARIANCIA BOOTSTRAP MULTINOMIAL (M,p01,p11,p10)';

data multi4;
set multi2;
prop=menores/&b;
zzero=probit(prop);
alfa1=probnorm(2*zzero-&zalfa);
alfa2=1-probnorm(2*zzero+&zalfa);
do k=1 to &b;
    output;
end;
keep alfa1 alfa2;

*Intervalo do percentil para Neb;
data multi2;
infile inter2;
input neb;
proc sort data=multi2;
    by neb;

data multi6;
set multi2;
n=_n_;
if n=round((&b+1)*&alfa) then output;
if n=round((&b+1)*(1-&alfa)) then output;
proc print;
    var n neb;
    title 'INTERVALO DE CONFIANCA DO PERCENTIL';

data multi8;
merge multi2 multi4;
data multi8;
set multi8;
n=_n_;
if n=round((&b+1)*alfa1) then output;
if n=round((&b+1)*(1-alfa2)) then output;
proc print;
    var n neb;
title 'INTERVALO DE CONFIANCA DE VICIO CORRIGIDO';

```

*REPLICACOES BOOTSTRAP HIPERGEOMETRICO;

data hiper1;

infile amostra;

input ni1-ni&s mi1-mi&s mmi1-mmi&s nest;

file inter3;

N=nest;

array a{&max};

do i=1 to n;

 a{i}=0;

end;

*Selecao da primeira amostra;

do j=1 to &na1;

 verifica='f';

 do until (verifica='v');

 u=int(ranuni(&sem2)*n)+1;

 if a{u}=0 then do;

 a{u}=u;

 verifica='v';

 end;

 end;

end;

*Ordenacao da primeira amostra;

array a1m{&na1};

j=1;

k=1;

do while(k<=&na1);

 if a{j}=0 then j=j+1;

 else do;

 a1m{k}=a{j};

 a{j}=0;

 k=k+1;

 j=j+1;

 end;

end;

*Selecao das b segundas amostras;

Neb=0; s=0; me=0; menores=0;

do kk=1 to &b;

 do j=1 to &na2;

 verifica='f';

 do until (verifica='v');

 u=int(ranuni(0)*n)+1;

 if a{u}=0 then do;

 a{u}=u;

 verifica='v';

 end;

 end;

 end;

*Ordenacao da segunda amostra;

array a2m{&na2};

j=1;

k=1;

do while(k<=&na2);

 if a{j}=0 then j=j+1;

 else do;

 a2m{k}=a{j}; a{j}=0;

```

        k=k+1;
        j=j+1;
    end;
end;
*Calculo dos mi para cada segunda amostra;
m=0;
i=1;
j=1;
do while (i<=&na1 & j<=&na2);
    if a1m{i}=a2m{j} then do;
        m=m+1;
        j=j+1;
        i=i+1;
    end;
    else do;
        if a1m{i}>a2m{j} then j=j+1;
        else i=i+1;
    end;
end;
neb=int((&na1+1)*(&na2+1)/(m+1)-1);
if neb<n then menores=menores+1;
put neb;
*Variancia de Neb;
aux=neb-me;
s=s+(kk-1)*(aux)**2/kk;
me=me+aux/kk;
end;
s=(s/(&b-1))**.5;
keep s menores;
proc print;
    var s menores;
title'VARIANCIA BOOTSTRAP HIPERGEOMETRICO (N, n1, n2)';

data hiper2;
set hiper1;
prop=menores/&b;
zzero=probit(prop);
alfa1=probnorm(2*zzero-&zalfa);
alfa2=1-probnorm(2*zzero+&zalfa);
do k=1 to &b;
    output;
end;
keep alfa1 alfa2;

*Intervalo do Percentil;
data hiper1;
infile inter3;
input neb;
proc sort data=hiper1;
    by neb;

data hiper3;
set hiper1;
n=_n_;
if n=round((&b+1)*&zalfa) then output;
if n=round((&b+1)*(1-&zalfa)) then output;

```



```

proc print;
  var n neb;
title 'INTERVALO DE CONFIANCA DO PERCENTIL';

data hiper4;
merge hiper1 hiper2;
data hiper4;
set hiper4;
  n=_n_;
if n=round((&b+1)*alfa1) then output;
if n=round((&b+1)*(1-alfa2)) then output;
proc print;
  var n neb;
title 'INTERVALO DE CONFIANCA DE VICIO CORRIGIDO';
run;

```

PROG3.SAS

```

/* .....
      ESTIMACAO DO NIVEL DE CONFIANCA REAL
      DOS INTERVALOS DO PERCENTIL
..... */

```

```
filename amostral'a:\amostral.dat';
```

```

%let npop=1000;
%let s=2;
%let na1=150;
%let na2=160;
%let max=2500;
%let alfa=.025;
%let b=1000;

```

```
%macro teste;
```

```
file amostral;
```

```
data amostra; %progl;
```

```
*REPLICACOES BOOTSTRAP MULTINOMIAL (N, p01, p10, p11);
```

```

data mult1;
infile amostral;
input ni1-ni&s mi1-mi&s mmi1-mmi&s nest;
N=nest;
NEb=0; n10=0; n01=0;
do j=1 to &b;
  soma1=0;
  soma2=0;
  n11=0;
  do i=1 to n;
    r=ranuni(0);
    if r <= (ni1-mi2)/n then soma1=soma1+1;
    if r > (ni1-mi2)/n & r <= ni1/n then n11=n11+1;
    if r > ni1/n & r <= (ni1+ni2-mi2)/n then soma2=soma2+1;

```

```

end;
n10 = soma1+n11;
n01 = soma2+n11;
NEb = int((n10+1)*(n01+1)/(n11+1)-1);
output;
end;
proc sort data=multin1;
    by neb;
data multin1;
set multin1;
n=_n_;
if n=round((&b+1)*&alfa) then do;
    if neb < &npop then limites=1;
    output;
end;
if n=round((&b+1)*(1-&alfa)) then do;
    if neb > &npop then limites=1;
    output;
end;
proc means data=multin1 noprint;
    var limites;
    output out=temp3 sum=lim;
data temp3;
set temp3;
if lim=2 then limites='contido';
    else limites='nao-cont';
output;
keep limites;
proc append data=temp3 base=multin2;

*REPLICACOES BOOTSTRAP MULTINOMIAL (M, p01, p10, p11);
data multin3;
infile amostra1;
input n11-ni&s mi1-mi&s mmi1-mmi&s nest;
M=mmi2; N=nest;
NEb=0; n10=0; n01=0;
do j=1 to &b;
    soma1=0;
    soma2=0;
    n11=0;
    do i=1 to m;
        r=(m/n)*ranuni(0);
        if r <= (ni1-mi2)/n then soma1=soma1+1;
        if r > (ni1-mi2)/n & r <= ni1/n then n11=n11+1;
        if r > ni1/n & r <= (m/n) then soma2=soma2+1;
    end;
    n10 = soma1+n11;
    n01 = soma2+n11;
    neb = int((n10+1)*(n01+1)/(n11+1)-1);
    output;
end;
proc sort data=multin3;
    by neb;
data multin3;
set multin3;

```

```

n=_n_;
if n=round((b+1)*alfa) then do;
  if neb < npop then limites=1;
  output;
end;
if n=round((b+1)*(1-alfa)) then do;
  if neb > npop then limites=1;
  output;
end;
proc means data=multin3 noprint;
  var limites;
  output out=temp4 sum=lim;
data temp4;
set temp4;
if lim=2 then limites='contido';
  else limites='nao-cont';
output;
keep limites;
proc append data=temp4 base=multin4;

```

*REPLICACOES BOOTSTRAP HIPERGEOMETRICO;

```

data hiper1;
infile amostra1;
input ni1-ni&s mi1-mi&s mmil-mmi&s nest;
N=nest;
array a{&max};
do i=1 to n;
  a{i}=0;
end;
*Selecao da primeira amostra;
do j=1 to &na1;
  verifica='f';
  do until (verifica='v');
    u=int(ranuni(0)*n)+1;
    if a{u}=0 then do;
      a{u}=u;
      verifica='v';
    end;
  end;
end;
*Ordenacao da primeira amostra;
array alm{&na1};
j=1;
k=1;
do while(k<=&na1);
  if a{j}=0 then j=j+1;
  else do;
    alm{k}=a{j};
    a{j}=0;
    k=k+1;
    j=j+1;
  end;
end;
*Selecao das b segundas amostras;
Neb=0;

```

```

do kk=1 to &b;
  do j=1 to &na2;
    verifica='f';
    do until (verifica='v');
      u=int(ranuni(0)*n)+1;
      if a{u}=0 then do;
        a{u}=u;
        verifica='v';
      end;
    end;
  end;
end;
*Ordenacao da segunda amostra;
array a2m{&na2};
j=1;
k=1;
do while(k<=&na2);
  if a{j}=0 then j=j+1;
  else do;
    a2m{k}=a{j}; a{j}=0;
    k=k+1;
    j=j+1;
  end;
end;
*Calculo dos mi para cada segunda amostra;
m=0;
i=1;
j=1;
do while (i<=&na1 & j<=&na2);
  if a1m{i}=a2m{j} then do;
    m=m+1;
    j=j+1;
    i=i+1;
  end;
  else do;
    if a1m{i}>a2m{j} then j=j+1;
    else i=i+1;
  end;
end;
neb=int((&na1+1)*(&na2+1)/(m+1)-1);
output;
end;
proc sort data=hiper1;
  by neb;
data hiper1;
set hiper1;
n=_n_;
if n=round((&b+1)*&alfa) then do;
  if neb<&npop then limites=1;
  output;
end;
if n=round((&b+1)*(1-&alfa)) then do;
  if neb>&npop then limites=1;
  output;
end;
proc means data=hiper1 noprint;
  var limites;

```

```

        output out=temp5 sum=lim;
data temp5;
set temp5;
if lim=2 then limites='contido';
    else limites='nao-cont';
output;
keep limites;
proc append data=temp5 base=hiper2;

%mend teste; /* FIM DO MACRO */

data interval; %teste; /* REPETIR AS VEZES QUE O MACRO VAI SER EXECUTADO */

*OBTER AS FREQUENCIAS 'CONTIDO/NAO-CONTIDO';
proc freq data=multin2;
    tables limites;
    title 'BOOTSTRAP MULTINOMIAL (N, p10, p01, p11)';
proc freq data=multin4;
    tables limites;
    title 'BOOTSTRAP MULTINOMIAL (M, p10, p01, p11)';
proc freq data=hiper2;
    tables limites;
    title 'BOOTSTRAP HIPERGEOMETRICO (N, n1,n2)';
run;

```

PROG4.SAS

```

/* .....
    EXPERIMENTO DE C-R PARA TRES AMOSTRAS
    CALCULO DAS ESTIMATIVA DE PETERSEN E LEITE
    ..... */

filename amostras'b:\amostras.dat';

%let npop=1000;
%let nam=3;
%let na1=150;
%let na2=250;
%let na3=250;
%let sem=7247;

%macro prog4; /*MACRO PARA SIMULACAO, PRESCINDIVEL SE DADOS REAIS EXISTEM
    PARA TRES AMOSTRAS*/
file amostras;
array a{&nam,&npop};

```

```

do j=1 to &nam;
  do i=1 to &npop;
    a{j,i}=0;
  end;
end;
array ni{&nam};
ni1=&na1; ni2=&na2; ni3=&na3;
do j=1 to &nam;
  do k=1 to ni{j};
    verifica='f';
    do until (verifica='v');
      u=int(ranuni(&sem)*&npop)+1;
      if a{j,u}=0 then do;
        a{j,u}=1;
        verifica='v';
      end;
    end;
  end;
end;
array mi{&nam};
do j=1 to &nam;
  num=0;
  do i=1 to &npop;
    m=0;
    if a{j,i} = 1 then do;
      do k=1 to j-1;
        m=m+a{k,i};
      end;
      if m >= 1 then num=num+1;
    end;
  end;
  mi{j}=num;
end;

array mmi{&nam};
do j=1 to &nam;
  num=0;
  do k=1 to j;
    num=num+(ni{k}-mi{k});
  end;
  mmi{j}=num;
end;

array Nest{&nam};
do k= 2 to &nam;
  Nest{k}= int((ni{k}+1)*(mmi{k-1}+1)/(mi{k}+1))-1;
end;

iteracao=0;
Nast= 1;
Nant= 1;
erro= 20;
do until (abs(erro) <= 0.01);
  count1= 0;
  count2= 0;
  do i= 2 to &nam;

```

```

        count1= count1+ Nest{i}*mi{i}/((Nast-mmi{i-1})*(Nast-ni{i}));
        count2= count2+ mi{i}/((Nast-mmi{i-1})*(Nast-ni{i}));
    end;
    Nant= Nast;
    Nast= count1/(count2);
    erro= (Nast-Nant)/(Nant);
    iteracao=iteracao+1;
end;
npp=round(nast);

aa=0-mi2-mi3;
bb=2*(mmi3*aa)+(ni1*ni2+ni1*ni3+ni2*ni3);
cc=(mmi3-ni1)*(mmi3-ni2)*(mmi3-ni3);
term=sqrt((bb**2)-4*aa*cc);
raiz1=(0-bb+term)/(2*aa);
raiz2=(0-bb-term)/(2*aa);
nzero=0;
if raiz1<0 & raiz2<0 then nzero=0;
if raiz1<0 & raiz2>0 then nzero=raiz2;
if raiz1>0 & raiz2>0 then nzero=raiz1;
nzero=int(nzero+1);
nle=mmi3+nzero-1;
put ni1-ni&nam mi1-mi&nam mmi1-mmi&nam npp nle;

%mend prog4; /* FIM DO MACRO PROG4*/

data amostras; %prog4;

proc print;
    var ni1-ni&nam mi1-mi&nam mmi1-mmi&nam npp nle;
run;

```

P R O G 5 . S A S

```

/* .....
    CALCULO DA VARIANCIA BOOTSTRAP, INTERVALOS DO PERCENTIL
    PARA A ESTIMATIVA DE PETERSEN (TRES AMOSTRAS)
    ..... */

filename boot11'a:\boot11.dat';
filename boot12'a:\boot12.dat';
filename boot13'a:\boot13.dat';

%let na1=150;
%let na2=80;
%let na3=80;
%let alfa=.025;
%let zalfa=probit(.975);
%let b=1000;
%let sem=1111;
%let nmax=2500;

*REPLICACOES BOOTSTRAP MULTINOMIAL (N, p1,p2,...);

```

```

data multin1;
infile 'a:\ amostras.dat;
input ni1-ni&nam mi1-mi&nam mm1-mmi&nam npp nle;

file boot11;
n=npp;
p100=ni1*(n-ni2)*(n-ni3)/n**3;
p010=(n-ni1)*ni2*(n-ni3)/n**3;
p001=(n-ni1)*(n-ni2)*ni3/n**3;
p110=ni1*ni2*(n-ni3)/n**3;
p101=ni1*(n-ni2)*ni3/n**3;
p011=(n-ni1)*ni2*ni3/n**3;
p111=ni1*ni2*ni3/n**3;

neb=0;
s=0; me=0; menores=0;
do kk=1 to &b;
  c100=0; c010=0; c001=0; c110=0; c101=0; c011=0; c111=0;
  do j=1 to n;
    r=ranuni(0);
    if r <= p100 then c100=c100+1;
    if r > p100 & r <= p100+p010 then c010=c010+1;
    if r > p100+p010 & r <= p100+p010+p001 then c001=c001+1;
    if r > p100+p010+p001 & r <= p100+p010+p001+p110 then c110=c110+1;
    if r > p100+p010+p001+p110 & r <= p100+p010+p001+p110+p101
      then c101=c101+1;
    if r > p100+p010+p001+p110+p101 & r <= p100+p010+p001+p110+p101+p011
      then c011=c011+1;
    if r > p100+p010+p001+p110+p101+p011 &
      r <= p100+p010+p001+p110+p101+p011+p111 then c111=c111+1;
  end;
  array nib{&nam};
  nib1=c100+c110+c101+c111;
  nib2=c010+c110+c011+c111;
  nib3=c001+c101+c011+c111;
  array mib{&nam};
  mib1=0; mib2=c110+c111; mib3=c101+c011+c111;
  array mmib{&nam};
  mmib1=nib1; mmib2=nib1+nib2-mib2; mmib3=nib1+nib2+nib3-mib2-mib3;
  array nestb{&nam};
  do k= 2 to &nam;
    Nestb{k}= int((nib{k}+1)*(mmib{k-1}+1)/(mib{k}+1))-1;
  end;

  Nastb= 1;
  Nantb= 1;
  erro= 20;
  do until (abs(erro) <= 0.001);
    count1= 0;
    count2= 0;
    do i= 2 to &nam;
      count1= count1+ Nestb{i}*mib{i}/((Nastb-mmib{i-1})*(Nastb-nib{i}));
      count2= count2+ mib{i}/((Nastb-mmib{i-1})*(Nastb-nib{i}));
    end;
    Nantb= Nastb;
    Nastb= count1/(count2);

```



```

    neb=round(nastb);
    erro=(Nastb-Nantb)/(Nantb);
end;
if neb < n then menores=menores+1;
put neb;

*Variancia de neb;
aux=neb-me;
s=s+(kk-1)*(aux)**2/kk;
me=me+aux/kk;
end;
s=(s/(&b-1))**.5;
keep s menores;
proc print;
    var s menores;
title'VARIANCIA BOOTSTRAP MULTINOMIAL (N, p...)';

data mult2;
set mult1;
prop=menores/&b;
zzero=probit(prop);
alfa1=probnorm(2*zzero-&zalfa);
alfa2=1-probnorm(2*zzero+&zalfa);
do k=1 to &b;
    output;
end;
keep alfa1 alfa2;

* INTERVALO DO PERCENTIL;
data mult1;
infile boot1;
input neb;
proc sort data=mult1;
    by neb;
data mult3;
set mult1;
n=_n_;
if n=round((&b+1)*&alfa) then output;
if n=round((&b+1)*(1-&alfa)) then output;
proc print;
    var n neb;
title'INTERVALO DE CONFIANCA DO PERCENTIL';

data mult4;
merge mult1 mult2;
data mult4;
set mult4;
n=_n_;
if n=round((&b+1)*alfa1) then output;
if n=round((&b+1)*(1-alfa2)) then output;
proc print;
    var n neb;
title'INTERVALO DO PERCENTIL DE VICIO CORRIGIDO';

*REPLICACOES BOOTSTRAP MULTINOMIAL (M, p1,p2,...);

```

```

data multin5;
infile 'a:\amostras.dat';
input ni1-ni&nam mi1-mi&nam mmi1-mmi&nam npp nle;

file boot12;

m=mmi3; n=npp;
p100=ni1*(n-ni2)*(n-ni3)/n**3;
p010=(n-ni1)*ni2*(n-ni3)/n**3;
p001=(n-ni1)*(n-ni2)*ni3/n**3;
p110=ni1*ni2*(n-ni3)/n**3;
p101=ni1*(n-ni2)*ni3/n**3;
p011=(n-ni1)*ni2*ni3/n**3;
p111=ni1*ni2*ni3/n**3;

neb=0;
s=0; me=0; menores=0;
do kk=1 to &b;
  c100=0; c010=0; c001=0; c110=0; c101=0; c011=0; c111=0;
  do j=1 to m;
    r=(m/n)*ranuni(0);
    if r <= p100 then c100=c100+1;
    if r > p100 & r <= p100+p010 then c010=c010+1;
    if r > p100+p010 & r <= p100+p010+p001 then c001=c001+1;
    if r > p100+p010+p001 & r <= p100+p010+p001+p110 then c110=c110+1;
    if r > p100+p010+p001+p110 & r <= p100+p010+p001+p110+p101
      then c101=c101+1;
    if r > p100+p010+p001+p110+p101 & r <= p100+p010+p001+p110+p101+p011
      then c011=c011+1;
    if r > p100+p010+p001+p110+p101+p011 &
      r <= p100+p010+p001+p110+p101+p011+p111 then c111=c111+1;
  end;
  array nib{&nam};
  nib1=c100+c110+c101+c111;
  nib2=c010+c110+c011+c111;
  nib3=c001+c101+c011+c111;
  array mib{&nam};
  mib1=0; mib2=c110+c111; mib3=c101+c011+c111;
  array mmib{&nam};
  mmib1=nib1; mmib2=nib1+nib2-mib2; mmib3=nib1+nib2+nib3-mib2-mib3;
  array nestb{&nam};
  do k= 2 to &nam;
    Nestb{k}= int((nib{k}+1)*(mmib{k-1}+1)/(mib{k}+1))-1;
  end;

Nastb= 1;
Nantb= 1;
erro= 20;
do until (abs(erro) <= 0.001);
  count1= 0;
  count2= 0;
  do i= 2 to &nam;
    count1= count1+ Nestb{i}*mib{i}/((Nastb-mmib{i-1})*(Nastb-nib{i}));
    count2= count2+ mib{i}/((Nastb-mmib{i-1})*(Nastb-nib{i}));
  end;
  erro= (count1-count2)/count1;
end;

```

```

Nantb= Nastb;
Nastb= count1/(count2);
neb=round(nastb);
erro=(Nastb-Nantb)/(Nantb);
end;
if neb < n then menores=menores+1;
put neb;

*Variancia de neb;
aux=neb-me;
s=s+(kk-1)*(aux)**2/kk;
me=me+aux/kk;
end;
s=(s/(&b-1))*5;
keep s menores;
proc print;
    var s menores;
title'VARIANCIA BOOTSTRAP MULTINOMIAL (M, p...);

data multin6;
set multin5;
prop=menores/&b;
zzero=probit(prop);
alfa1=probnorm(2*zzero-&zalfa);
alfa2=1-probnorm(2*zzero+&zalfa);
do k=1 to &b;
    output;
end;
keep alfa1 alfa2;

* INTERVALO DO PERCENTIL;
data multin5;
infile boot12;
input neb;
proc sort data=multin5;
    by neb;
data multin7;
set multin5;
n=_n_;
if n=round((&b+1)*&alfa) then output;
if n=round((&b+1)*(1-&alfa)) then output;
proc print;
    var n neb;
title'INTERVALO DE CONFIANCA DO PERCENTIL';

data multin8;
merge multin5 multin6;
data multin8;
set multin8;
n=_n_;
if n=round((&b+1)*alfa1) then output;
if n=round((&b+1)*(1-alfa2)) then output;
proc print;
    var n neb;
title'INTERVALO DO PERCENTIL DE VICIO CORRIGIDO';

```

```
*REPLICACOES BOOTSTRAP HIPERGEOMETRICO;
```

```
data hiper1;
infile 'a:\amostras.dat';
input ni1-ni&nam mil-mi&nam mmil-mmi&nam npp nle;
```

```
file boot13;
```

```
n=npp;
neb=0; s=0; me=0; menores=0;
```

```
do kk=1 to &b;
array a{&nam,&nmax};
do j=1 to &nam;
  do i=1 to n;
    a{j,i}=0;
  end;
end;
```

```
array ni{&nam};
do j=1 to &nam;
  do k=1 to ni{j};
    verifica='f';
    do until (verifica='v');
      u=int(ranuni(j)*n)+1;
      if a{j,u}=0 then do;
        a{j,u}=1;
        verifica='v';
      end;
    end;
  end;
end;
array mib{&nam};
do j=1 to &nam;
  num=0;
  do i=1 to n;
    m=0;
    if a{j,i} = 1 then do;
      do k=1 to j-1;
        m=m+a{k,i};
      end;
      if m >= 1 then num=num+1;
    end;
  end;
  mib{j}=num;
end;
```

```
array mmib{&nam};
do j=1 to &nam;
  num=0;
  do k=1 to j;
    num=num+(ni{k}-mib{k});
  end;
  mmib{j}=num;
end;
```

```

array Nestb{&nam};
do k= 2 to &nam;
  Nestb{k}= int((ni{k}+1)*(mmib{k-1}+1)/(mib{k}+1))-1;
end;

Nastb= 1;
Nantb= 1;
erro= 20;
do until (abs(erro)<=0.001);
  count1= 0;
  count2= 0;
  do i= 2 to &nam;
    count1= count1+ Nestb{i}*mib{i}/((Nastb-mmib{i-1})*(Nastb-ni{i}));
    count2= count2+ mib{i}/((Nastb-mmib{i-1})*(Nastb-ni{i}));
  end;
  Nastb= Nastb;
  Nantb= count1/(count2);
  erro= (Nastb-Nantb)/(Nantb);
  neb =round(nastb);
end;
if neb<n then menores=menores+1;
put neb;

*Variância de Neb;
aux=neb-me;
s=s+(kk-1)*(aux)**2/kk;
me=me+aux/kk;
end;
s=(s/(&b-1))*5;
keep s menores;
proc print;
  var s menores;
title'VARIANCIA BOOTSTRAP HIPERGEOMETRICO (N, M4)';

data hiper2;
set hiper1;
prop=menores/&b;
zzero=probit(prop);
alfa1=probnorm(2*zzero-&zalfa);
alfa2=1-probnorm(2*zzero+&zalfa);
do k=1 to &b;
  output;
end;
keep alfa1 alfa2;

*Intervalo do Percentil;
data hiper1;
infile boot13;
input neb;
proc sort data=hiper1;
  by neb;

data hiper3;
set hiper1;

```

```

n=_n_;
if n=round((b+1)*alfa) then output;
if n=round((b+1)*(1-alfa)) then output;
proc print;
var n neb;
title 'INTERVALO DE CONFIANCA DO PERCENTIL';

data hiper4;
merge hiper1 hiper2;
data hiper4;
set hiper4;
n=_n_;
if n=round((b+1)*alfa1) then output;
if n=round((b+1)*(1-alfa2)) then output;
proc print;
var n neb;
title 'INTERVALO DE CONFIANCA DE VICIO CORRIGIDO';
run;

```

P R O G 6 . S A S

```

/* .....
      CALCULO DA VARIANCA BOOTSTRAP, INTERVALOS DO PERCENTIL
      PARA A ESTIMATIVA MULTIAMOSTRAL DE LEITE
..... */

```

```

filename boot11'a:\boot11.dat';
filename boot12'a:\boot12.dat';
filename boot13'a:\boot13.dat';

```

```

%let nam=3;
%let na1=150;
%let na2=80;
%let na3=80;
%let alfa=.025;
%let zalfa=probit(.975);
%let b=1000;
%let sem=1111;
%let nmax=2500;

```

*REPLICACOES BOOTSTRAP MULTINOMIAL (N, p1,p2,...);

```

data multin1;
infile 'a:\amostras.dat';
input ni1-ni&nam mi1-mi&nam mmil-mmi&nam nle;

```

```

file boot11;
n=nle;
p100=ni1*(n-ni2)*(n-ni3)/n**3;
p010=(n-ni1)*ni2*(n-ni3)/n**3;
p001=(n-ni1)*(n-ni2)*ni3/n**3;
p110=ni1*ni2*(n-ni3)/n**3;
p101=ni1*(n-ni2)*ni3/n**3;
p011=(n-ni1)*ni2*ni3/n**3;

```

```

p111=ni1*ni2*ni3/n**3;

neb=0;
s=0; me=0; menores=0;

do kk=1 to &b;
c100=0; c010=0; c001=0; c110=0; c101=0; c011=0; c111=0;
do j=1 to n;
r=ranuni(0);
if r <= p100 then c100=c100+1;
if r > p100 & r <= p100+p010 then c010=c010+1;
if r > p100+p010 & r <= p100+p010+p001 then c001=c001+1;
if r > p100+p010+p001 & r <= p100+p010+p001+p110 then c110=c110+1;
if r > p100+p010+p001+p110 & r <= p100+p010+p001+p110+p101
then c101=c101+1;
if r > p100+p010+p001+p110+p101 & r <= p100+p010+p001+p110+p101+p011
then c011=c011+1;
if r > p100+p010+p001+p110+p101+p011 &
r <= p100+p010+p001+p110+p101+p011+p111 then c111=c111+1;
end;
array nib{&nam};
nib1=c100+c110+c101+c111;
nib2=c010+c110+c011+c111;
nib3=c001+c101+c011+c111;
array mib{&nam};
mib1=0; mib2=c110+c111; mib3=c101+c011+c111;
array mmib{&nam};
mmib1=nib1; mmib2=nib1+nib2-mib2; mmib3=nib1+nib2+nib3-mib2-mib3;

aa=0-mib2-mib3;
bb=2*(mmib3*aa)+(nib1*nib2+nib1*nib3+nib2*nib3);
cc=(mmib3-nib1)*(mmib3-nib2)*(mmib3-nib3);
term=sqrt((bb**2)-4*aa*cc);
raiz1=(0-bb+term)/(2*aa);
raiz2=(0-bb-term)/(2*aa);
nzero=0;
if raiz1<0 & raiz2<0 then nzero=0;
if raiz1<0 & raiz2>0 then nzero=raiz2;
if raiz1>0 & raiz2>0 then nzero=raiz1;
nzero=int(nzero+1);
neb=mmib3+nzero-1;

if neb < n then menores=menores+1;
put neb;

*Variancia de neb;
aux=neb-me;
s=s+(kk-1)*(aux)**2/kk;
me=me+aux/kk;
end;
s=(s/(&b-1))**.5;
keep s menores;
proc print;
var s menores;
title'VARIANCIA BOOTSTRAP MULTINOMIAL (N, p...);

```

```

data multin2;
set multin1;
prop=menores/&b;
zzero=probit(prop);
alfa1=probnorm(2*zzero-&zalfa);
alfa2=1-probnorm(2*zzero+&zalfa);
do k=1 to &b;
    output;
end;
keep alfa1 alfa2;

* INTERVALO DO PERCENTIL;
data multin1;
infile boot11;
input neb;
proc sort data=multin1;
    by neb;
data multin3;
set multin1;
n=_n_;
if n=round((&b+1)*&alfa) then output;
if n=round((&b+1)*(1-&alfa)) then output;
proc print;
    var n neb;
title'INTERVALO DE CONFIANCA DO PERCENTIL';

data multin4;
merge multin1 multin2;
data multin4;
set multin4;
n=_n_;
if n=round((&b+1)*alfa1) then output;
if n=round((&b+1)*(1-alfa2)) then output;
proc print;
    var n neb;
title'INTERVALO DO PERCENTIL DE VICIO CORRIGIDO';
run;

*REPLICACOES BOOTSTRAP MULTINOMIAL (M, p1,p2,...);

data multin5;
infile 'a:\amostras.dat';
input ni1-ni&nam mil-mi&nam mmi1-mmi&nam nle;

file boot12;

m=mmi3; n=nle;
p100=ni1*(n-ni2)*(n-ni3)/n**3;
p010=(n-ni1)*ni2*(n-ni3)/n**3;
p001=(n-ni1)*(n-ni2)*ni3/n**3;
p110=ni1*ni2*(n-ni3)/n**3;
p101=ni1*(n-ni2)*ni3/n**3;
p011=(n-ni1)*ni2*ni3/n**3;
p111=ni1*ni2*ni3/n**3;

neb=0;

```



```

s=0; me=0; menores=0;
do kk=1 to &b;
  c100=0; c010=0; c001=0; c110=0; c101=0; c011=0; c111=0;
  do j=1 to m;
    r=(m/n)*ranuni(0);
    if r <= p100 then c100=c100+1;
    if r > p100 & r <= p100+p010 then c010=c010+1;
    if r > p100+p010 & r <= p100+p010+p001 then c001=c001+1;
    if r > p100+p010+p001 & r <= p100+p010+p001+p110 then c110=c110+1;
    if r > p100+p010+p001+p110 & r <= p100+p010+p001+p110+p101
      then c101=c101+1;
    if r > p100+p010+p001+p110+p101 & r <= p100+p010+p001+p110+p101+p011
      then c011=c011+1;
    if r > p100+p010+p001+p110+p101+p011 &
      r <= p100+p010+p001+p110+p101+p011+p111 then c111=c111+1;
  end;
  array nib{&nam};
  nib1=c100+c110+c101+c111;
  nib2=c010+c110+c011+c111;
  nib3=c001+c101+c011+c111;
  array mib{&nam};
  mib1=0; mib2=c110+c111; mib3=c101+c011+c111;
  array mmib{&nam};
  mmib1=nib1; mmib2=nib1+nib2-mib2; mmib3=nib1+nib2+nib3-mib2-mib3;

  aa=0-mib2-mib3;
  bb=2*(mmib3*aa)+(nib1*nib2+nib1*nib3+nib2*nib3);
  cc=(mmib3-nib1)*(mmib3-nib2)*(mmib3-nib3);
  term=sqrt((bb**2)-4*aa*cc);
  raiz1=(0-bb+term)/(2*aa);
  raiz2=(0-bb-term)/(2*aa);
  nzero=0;
  if raiz1<0 & raiz2<0 then nzero=0;
  if raiz1<0 & raiz2>0 then nzero=raiz2;
  if raiz1>0 & raiz2>0 then nzero=raiz1;
  nzero=int(nzero+1);
  neb=mmib3+nzero-1;

  if neb < n then menores=menores+1;
  put neb;

  *Variância de neb;
  aux=neb-me;
  s=s+(kk-1)*(aux)**2/kk;
  me=me+aux/kk;
end;
s=(s/(&b-1))**.5;
keep s menores;
proc print;
  var s menores;
title 'VARIANCIA BOOTSTRAP MULTINOMIAL (M, p...)';

data multin6;
set multin5;
prop=menores/&b;
zzero=probit(prop);

```

```

alfa1=probnorm(2*zzero-&zalfa);
alfa2=1-probnorm(2*zzero+&zalfa);
do k=1 to &b;
    output;
end;
keep alfa1 alfa2;

* INTERVALO DO PERCENTIL;
data multin5;
infile boot12;
input neb;
proc sort data=multin5;
    by neb;
data multin7;
set multin5;
n=_n_;
if n=round((&b+1)*&alfa) then output;
if n=round((&b+1)*(1-&alfa)) then output;
proc print;
    var n neb;
title'INTERVALO DE CONFIANCA DO PERCENTIL';

data multin8;
merge multin5 multin6;
data multin8;
set multin8;
n=_n_;
if n=round((&b+1)*alfa1) then output;
if n=round((&b+1)*(1-alfa2)) then output;
proc print;
    var n neb;
title'INTERVALO DO PERCENTIL DE VICIO CORRIGIDO';

*REPLICACOES BOOTSTRAP HIPERGEOMETRICO;
data hiper1;
infile 'a:\amostras.dat';
input ni1-ni&nam mil-mi&nam mmil-mmi&nam nle;

file boot13;

n=nle;
neb=0; s=0; me=0; menores=0;

do kk=1 to &b;
array a{&nam,&nmax};
do j=1 to &nam;
    do i=1 to n;
        a{j,i}=0;
    end;
end;

array ni{&nam};
do j=1 to &nam;
    do k=1 to ni{j};
        verifica='f';
        do until (verifica='v');

```

```

        u=int(ranuni(j)*n)+1;
        if a{j,u}=0 then do;
            a{j,u}=1;
            verifica='v';
        end;
    end;
end;
array mib{&nam};
do j=1 to &nam;
    num=0;
    do i=1 to n;
        m=0;
        if a{j,i} = 1 then do;
            do k=1 to j-1;
                m=m+a{k,i};
            end;
            if m >= 1 then num=num+1;
        end;
    end;
    mib{j}=num;
end;

array mmib{&nam};
do j=1 to &nam;
    num=0;
    do k=1 to j;
        num=num+(ni{k}-mib{k});
    end;
    mmib{j}=num;
end;

aa=0-mib2-mib3;
bb=2*(mmib3*aa)+(ni1*ni2+ni1*ni3+ni2*ni3);
cc=(mmib3-ni1)*(mmib3-ni2)*(mmib3-ni3);
term=sqrt((bb**2)-4*aa*cc);
raiz1=(0-bb+term)/(2*aa);
raiz2=(0-bb-term)/(2*aa);
nzero=0;
if raiz1<0 & raiz2<0 then nzero=0;
if raiz1<0 & raiz2>0 then nzero=raiz2;
if raiz1>0 & raiz2>0 then nzero=raiz1;
nzero=int(nzero+1);
neb=mmib3+nzero-1;

if neb<n then menores=menores+1;
put neb;

*Variância de Neb;
aux=neb-me;
s=s+(kk-1)*(aux)**2/kk;
me=me+aux/kk;
end;
s=(s/(&b-1))**.5;
keep s menores;
proc print;

```

```

var s menores;
title 'VARIANCIA BOOTSTRAP HIPERGEOMETRICO (N, M4)';

data hiper2;
set hiper1;
prop=menores/&b;
zzero=probit(prop);
alfa1=probnorm(2*zzero-&zalfa);
alfa2=1-probnorm(2*zzero+&zalfa);
do k=1 to &b;
    output;
end;
keep alfa1 alfa2;

*Intervalo do Percentil;
data hiper1;
infile boot13;
input neb;
proc sort data=hiper1;
    by neb;

data hiper3;
set hiper1;
n=_n_;
if n=round((&b+1)*&alfa) then output;
if n=round((&b+1)*(1-&alfa)) then output;
proc print;
    var n neb;
title 'INTERVALO DE CONFIANCA DO PERCENTIL';

data hiper4;
merge hiper1 hiper2;
data hiper4;
set hiper4;
n=_n_;
if n=round((&b+1)*alfa1) then output;
if n=round((&b+1)*(1-alfa2)) then output;
proc print;
    var n neb;
title 'INTERVALO DE CONFIANCA DE VICIO CORRIGIDO';
run;

```

P R O G 7 . S A S

```

/* .....
SIMULACAO DE UM EXPERIMENTO DE C-R
COM AMOSTRAGEM SEQUENCIAL
..... */

%let npop=1000;
%let na=301; /* tamanho amostral + 1 */

data a1;
sem=98765;

```

```

array a{&npop};
do i=1 to &npop;
  a{i}=0;
end;

array x{&na};
array m{&na};

do i=1 to (&na-1);
  xk=1;
  u=int(ranuni(sem)*&npop)+1;
  if a{u}=0 then do;
    a{u}=1;
    xk=0;
  end;
  x{i}=xk;
  m{1}=0;
  m{i+1}=m{i}+(1-x{i});
end;

rec=&na-1-m&na;
keep sem x1-x&na rec m1-m&na;

proc print;
  var x1-x&na sem rec;
proc print;
  var m1-m&na;
run;

```

PROG8.SAS

```

/* .....
  BOOTSTRAP PARA O RESULTADO DE UM EXPERIMENTO
    DE C-R DE AMOSTRAGEM SEQUENCIAL
  ..... */

%let na=200;

%macro amostras;

array x{&na};
rec=0;
do i=1 to &na;
  x{i}=0;
  u=int(ranuni(0)*&na)+1;
  if u<5 then x{i}=1;
  rec=rec+x{i};
end;
keep rec;

proc append data=a1 base=a2;

%mend amostras;

```

```
data a1; %amostras; /* REPETIR O NUMERO DE VEZES DAS REPLICACOES BOOTSTRAP
*/
```

```
proc freq data=a2;
  tables rec;
title'frequencia de recapturas na amostragem sequencial';
run;
```

PROG9.SAS

```
/* .....
SIMULACAO DE UM EXPERIMENTO DE C-R SEQUENCIAL
E TAMANHO DE AMOSTRA ESTOCASTICO
..... */
```

```
options ls=72;
options nodate;
```

```
%let npop=1000;
```

```
data a1;
```

```
do c=10 to 100 by 10;
```

```
  array a{&npop};
  do i=1 to &npop;
    a{i}=0;
  end;
```

```
  do j=10 to 50 by 10;
```

```
    np=&npop;
    na=0;
    wn=0;
    do until(wn=j);
      u=int(ranuni(94752+j))*&npop)+1;
      if a{u}=1 then wn=wn+1;
      if a{u}=0 then a{u}=1;
      na=na+1;
    end;
```

```
    tc=c*wn;
    output;
```

```
  end;
end;
```

```
drop a1-a&npop i u;
```

```
proc print;
```

```
title1 'SIMULACAO DE UM EXPERIMENTO DE C-R DE AMOSTRAGEM';
```

```
title2 'SEQUENCIAL E TAMANHO DE AMOSTRA ESTOCASTICO';
```

```
run;
```

REFERÊNCIA BIBLIOGRÁFICA

- Bailey, N. T. J. (1951). "On Estimating the Size of Mobile Population from Recapture Data". *Biometrika* **38**, 293-306.
- Buckland, S. T. (1984). "Monte Carlo Confidence Intervals". *Biometrics* **40**, 811-817.
- Buckland, S. T. and Garthwaite, P. H. (1991). "Quantifying Precision of Mark-Recapture Estimates Using the Bootstrap and Related Methods". *Biometrics* **47**, 255-268.
- Casajús, M. I. (1990). *Estimadores de Máxima Verossimilhança para os Parâmetros de Alguns Modelos de Captura-Recaptura em População Animal*. Dissertação de Mestrado: Instituto de Matemática e Estatística-USP, São Paulo.
- Chapman, D. (1952). "Inverse, Multiple and Sequential Sample Censuses". *Biometrics* **8**, 286-306.
- Cormack, R. M. (1969). "The Statistics of Capture-Recapture Methods". *Oceanogr. Mar. Biol. Ann. Rev.* **6**, 455-506.
- Cox, D. R. (1975). "Partial Likelihood". *Biometrika* **62**, 269-276.
- Cramér, H. (1974). *Mathematical Methods of Statistics*. Princeton University Press.
- Darroch, J. N. (1958). "The Multiple-Recapture Census I: Estimation of a Closed Population". *Biometrika* **45**, 343-359.
- Efron, B. (1982). "The Jackknife, the Bootstrap and the Other Resampling Plans". *Soc. Ind. Appl. Math. CBMS- Natl. Sci. Found. Monogr.* **38**
- Johnson, N. L. and Kotz, S. (1969). *Discrete Distributions*. Boston: Houghton Mifflin.
- Leite, J. G., Oishi, J. and Pereira, C. A. B. (1988). "A Note on the Exact Maximum Likelihood Estimation of the Size of a Finite and Closed Population". *Biometrika* **75**, 178-80.
- Otis, D. L., Burnham, K. P., White, G. C., Anderson, D. R. (1978). "Statistical Inference from Capture Data on Closed Animal Population". *Wildlife Monographs* **62**, 1-135.
- Petersen, G. G. J. (1896). "The Yearly Immigration of Young Plaice into the Limfjord from the German Sea". *Rep. Danish Biol. Stat.* **6**, 1-68.
- Pollock, K. H. (1991). "Modelling Capture, Recapture and Removal Statistics for Estimation of Demographic Parameters for Fish and Wildlife Population: Past,

- Present and Future". *JASA* **86**, 225-238.
- Pollock, K. H., Nichols, J. D., Hines, J. E. and Brownie, C. (1990). "Statistical Inference for Capture-Recapture Experiments". *Wildlife Monographs* **107**, 1-97.
- Samuel, E. (1968). "Sequential Maximum Likelihood Estimation of the Size of a Population". *Ann. Math. Statist.* **39**, 1057-1068.
- Schnabel, Z. E. (1938). "The Estimation of the Total Fish Population of a Lake". *American Mathematical Monthly* **45**, 348-352.
- Seber, G. A. F. (1965). "A Note on the Multiple Recapture Census". *Biometrika* **52**, 249-259.
- Seber, G. A. F. (1983). *The Estimation of Animal Abundance and Related Parameters*. (2nd ed.), London: Charles W. Griffin.
- Seber, G. A. F. (1986). "A Review of Estimating Animal Abundance". *Biometrics* **42**, 267-292.
- Seber, G. A. F. (1992). "A Review of Estimating Animal Abundance II". *International Statistical Review* **60**, 129-166.
- Sen, Pranab K. (1987). "Sequential Estimation of the Size of a Finite Population". *ReBraPe* **1**, 113-137.
- Sen, Pranab K. (1982 a). "A Renewal Theorem for an Urn Model". *The Annals of Probability*, **10**, 838-843.
- Sen, Pranab K. (1982 b). "On Asymptotic Normality in Sequential Sampling Tagging". *Sankhyā, Ser. A*, **44**, 352-363.
- Sen, A. R. and Sen, P. K. (1981). "Schnabel-type Estimators for Closed Populations with Multiple marking". *Sankhyā, Ser. B*, **43**, 68-80.
- White, G. C., Anderson, D. R., Burnham, K. P. and Otis, D. L. (1982). *Capture-Recapture and Removal Methods for Sampling Closed Populations*. Los Alamos, NM: Los Alamos National Laboratory.