

MÉTODOS PARA OBTENÇÃO DE ESTRATIFICAÇÕES
APROXIMADAMENTE ÓTIMAS

ANNUZIA MARIA PONTES MOREIRA GOSSON

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ORIENTADOR:

PROF.DR. WILTON DE OLIVEIRA BUSSAB

- SÃO PAULO, DEZEMBRO DE 1982 -

Aos meus pais PERBOYRE e ANNUZIA
o agradecimento maior

À AGOSTINHO e JOÃO PAULO

AGRADECIMENTOS

Agradeço:

- ao meu orientador Prof. Dr. Wilton de Oliveira Bussab, da Universidade de São Paulo - USP -, cuja influência fêz-me optar pelo estudo de amostragem em geral e amostragem estratificada em particular, e que, pacientemente, superando dificuldades de distância e tempo, e de limitações de sua aluna, soube apontar o caminho seguro para a execução deste trabalho;
- ao Prof. Roberto Cláudio Frota Bezerra, da Universidade Federal do Ceará - UFC -, a quem como amigo - principalmente coube a tarefa de encorajar-me em todos os momentos deste trabalho, e de quem, como estudioso aplicado e reconhecido, pude receber valiosas contribuições;
- ao Prof. Dr. João Batista Ferreira Gomes Neto, da Universidade Federal do Ceará - UFC -, cuja amizade fê-lo, muitas vezes, sacrificar parte de seu tempo pessoal e profissional para resolver comigo as "passagens matemáticas";
- ao Prof. José Jackson Lima de Albuquerque, da Universidade Federal do Ceará - UFC -, a quem coube a tarefa de iniciar-me no estudo de amostragem, ainda ao tempo de graduação, e de quem, de forma amiga, ao longo dos anos, independentemente deste trabalho, venho recebendo orientação científica;
- a Comissão Estadual de Planejamento Agrícola - CEPA-CE - e a Universidade Federal do Ceará - UFC - pelo apoio recebido durante o período do Curso de Mestrado;
- a José Alves Ferreira - o Ferreirinha - cujo trabalho de datilógrafo possibilitou a publicação ordenada e clara deste trabalho.

INDICE

INTRODUÇÃO	1
CAP. 1 - MÉTODO DE ESTRATIFICAÇÃO DEVIDO A DALENIUS E HODGES.	4
1.1 - Introdução.	4
1.2 - Limite Ótimo dos Estratos	6
1.2.1 - Limites Ótimos de Estratificação para Alocação de Neyman.	7
1.2.2 - Limites Ótimos de Estratificação para Alocação Proporcional	9
1.3 - Soluções Aproximadas dos Pontos Ótimos de Estratificação.	10
1.3.1 - Primeira Aproximação.	10
1.3.2 - Segunda Aproximação	11
1.4 - Exemplo Numérico.	19
1.4.1 - Determinação de $Z(y)$ e $I_p(y)$ para $f(y) = e^{-y}$	19
1.4.2 - Cálculos Numéricos para L=2 Estratos.	20
1.4.3 - Cálculos Numéricos para L=3 Estratos.	23
1.4.4 - Cálculos Numéricos para L>3 Estratos.	25
1.4.5 - Cálculo de Variâncias	25
CAP. 2 - MÉTODO DE ESTRATIFICAÇÃO DEVIDO A EKMAN	28
2.1 - Introdução.	28
2.2 - Métodos de Aproximação de Limites Ótimos de Estratificação para uma Amplitude Finita.	30
2.3 - Métodos de Aproximação de Limites Ótimos de Estratificação para uma Amplitude Infinita.	37
2.4 - Primeira Aproximação para a Constante C_L	38
2.5 - Segunda Aproximação para a Constante C_L	40
2.6 - Exemplo Numérico.	40
2.6.1 - Determinação da Constante C_2 e do Limite do Estrato y_1 para L=2 Estratos	41
2.6.2 - Determinação da Constante C_3 e dos Limites dos Estratos $[y_h]$ para L=3 Estratos	42
2.6.2.1 - Primeira Aproximação para C_3 e $[y_h]$	42
2.6.2.2 - Segunda Aproximação para C_3 e $[y_h]$	44

2.6.3 - Determinação da Constante C_L e dos Limites dos Estratos $[y_h]$ para $L > 3$ Estratos	45
2.6.4 - Cálculo de Variâncias	45
CAP. 3 - MÉTODO DE ESTRATIFICAÇÃO USANDO O TOTAL	47
3.1 - Introdução	47
3.2 - Adaptação a Cálculos Numéricos	48
3.3 - Exemplo Numérico	49
3.3.1 - Cálculos Numéricos para $L=2$ Estratos	50
3.3.2 - Cálculos Numéricos para $L=3$ Estratos	51
3.3.3 - Cálculos Numéricos para $L > 3$ Estratos	53
3.4 - Cálculos de Variâncias	53
CAP. 4 - ESTRATIFICAÇÃO APROXIMADAMENTE ÓTIMA COM VARIÁVEL AUXILIAR	56
4.1 - Introdução	56
4.2 - Limite Ótimo dos Estratos	57
4.2.1 - Limites Ótimos de Estratificação para Alocação de Neyman	59
4.2.2 - Limites Ótimos de Estratificação para Alocação Proporcional	61
4.3 - Expressões Aproximadas para a Média e Variância Condicional	63
4.4 - Soluções Aproximadas dos Pontos Ótimos de Estratificação	69
4.5 - Uso da Regressão como Método de Estimação	78
4.6 - Exemplo Numérico	82
CAP. 5 - APLICAÇÃO DAS REGRAS DE ESTRATIFICAÇÃO ÓTIMA	86
5.1 - Introdução	86
5.2 - Aplicação das Regras	87
5.2.1 - O Estudo da População	87
5.2.2 - Método de Dalenius e Hodges	89
5.2.3 - Método de Ekman	89
5.2.4 - Igual Tamanho Agregado	91
5.2.5 - Comparação das Diferentes Regras	92
5.3 - Comparação de Métodos para Determinação dos Limites dos Estratos (Resumo)	94
CAP. 6 - CONCLUSÕES	99
RELAÇÃO BIBLIOGRÁFICA	102

INTRODUÇÃO

O uso de amostragem estratificada envolve quatro operações específicas:

- (i) a escolha de uma variável de estratificação;
- (ii) a escolha de um número L de estratos;
- (iii) a determinação do limite dos estratos;
- (iv) a escolha do tamanho n_h da amostra a ser tomada do h -ésimo estrato.

Um dos problemas sérios da estratificação é a divisão dos estratos.

Dalenius (1950) determinou equações cujas soluções nos fornecem limites ótimos de estratificação, y_h , para alocações de Neyman e Proporcional. Como a solução dessas equações apresenta dificuldades computacionais, vários métodos de aproximação para esses limites ótimos foram desenvolvidos.

Este trabalho se detém em aspectos envolvidos na determinação de limites, próximos do ótimo, para cada estrato da população em estudo. Nossa preocupação é determinar, para uma população contínua, os melhores limites que

vão compor os estratos. Os melhores limites são definidos como aqueles que dão uma variância mínima para a média populacional estimada de uma amostra estratificada de tamanho n , com escolhas ótimas de tamanho n_h nos estratos individuais. A estimativa da média populacional é $\bar{y}_{st} = \sum_h \omega_h \bar{y}_h$ onde \bar{y}_h é a média da amostra no h -ésimo estrato e ω_h é a proporção de unidades no mesmo estrato. Sua variância é igual a $V(\bar{y}_{st}) = \sum_h \frac{\omega_h^2 \sigma_h^2}{n_h}$, onde o esquema proposto é amostra aleatória simples com reposição em cada estrato.

Os três primeiros capítulos, apresentam métodos de estratificação aproximadamente ótimos quando a variável de estratificação é considerada como a própria variável de estimação. No primeiro capítulo, abordamos o método desenvolvido por Dalenius e Hodges (1959). De início mostramos as equações que determinam os limites reais ótimos de estratificação para as alocações de Neyman e Proporcional. A seguir encontramos soluções aproximadas para o conjunto que satisfaz as equações minimais e apresentamos um exemplo. No segundo capítulo desenvolvemos um outro método, devido a Ekman(1959), no qual são apresentados dois casos; um para densidades com amplitude finita, e outro para densidades com amplitude infinita. Finalizando o capítulo, apresentamos um exemplo. No terceiro, desenvolvemos uma regra sugerida por Hansen, Hurwitz e Madow (1953) que consiste em formar estratos de tal maneira que, na distribuição de y , o seu valor agregado é o mesmo para cada estrato. Também exemplificamos através de

uma aplicação do procedimento.

O quarto capítulo difere dos três primeiros na medida que a variável de estratificação x , não é a variável de estimação y . Na maioria das situações práticas a variável de estimação y é raramente conhecida. Então, a estratificação é feita com base em alguma variável x convenientemente escolhida (geralmente é o valor de y em um censo recente ou outra variável muito relacionada com y). A composição desse capítulo é iniciada com a determinação dos limites ótimos de estratificação. A seguir encontramos soluções aproximadas sintetizadas em quatro regras determinadas por Singh e Sukhatme (1969). Um item aborda em particular o uso da regressão como método de estimação do qual apresentamos um exemplo.

No quinto capítulo, fazemos uma aplicação das técnicas apresentadas nos três primeiros capítulos, a uma situação prática, com dados referentes às áreas de imóveis rurais de proprietários agrícolas do município de Independência-Ce.

Finalizando o trabalho, expomos, em um breve capítulo de conclusões, sugestões de estudo para alguns assuntos referentes à estratificação e que não foram abordados aqui.

CAPÍTULO 1

MÉTODO DE ESTRATIFICAÇÃO DEVIDO A DALENIUS E HODGES

1.1. Introdução

Considere uma população a ser estratificada em L estratos. Seja y_0, y_1, \dots, y_L os limites dos estratos, os estratos sendo numerados por $1, 2, \dots, L$, e seja μ_h, σ_h e ω_h , a média, o desvio padrão e a proporção de unidades no h -ésimo estrato.

ESTRATO	1	2	...	h	...	L
LIMITES	y_1	y_2	...	y_h	...	y_L
MÉDIA	μ_1	μ_2	...	μ_h	...	μ_L
DESVIO PADRÃO	σ_1	σ_2	...	σ_h	...	σ_L
PROPORÇÃO	ω_1	ω_2	...	ω_h	...	ω_L

Em amostragem estratificada, a estimativa da média populacional é $\bar{y}_{st} = \sum_h \omega_h \bar{y}_h$ onde \bar{y}_h é a média da amostra no h-ésimo estrato. Sua variância é igual a $V(\bar{y}_{st}) = \sum_h \frac{\omega_h^2 \sigma_h^2}{n_h}$ quando foi retirada uma amostra aleatória simples de n_h unidades com reposição da população com N_h elementos no estrato h , $h=1,2,\dots,L$.

Para um tamanho de amostra n fixado, e tomando alocação de Neyman, onde $n_h = n \frac{\omega_h \sigma_h}{\sum_h \omega_h \sigma_h}$, a variância da média é dada por

$$V(\bar{y}_{st})_N = \frac{1}{n} \left[\sum_h \omega_h \sigma_h \right]^2 \quad (1.1)$$

e para alocação proporcional, onde $n_h = n \frac{N_h}{N}$, a variância da média é dada por

$$V(\bar{y}_{st})_p = \frac{1}{n} \sum_h \omega_h \sigma_h^2 \quad (1.2)$$

O problema de escolher limites de estratificação y_h que minimizem $V(\bar{y}_{st})$ foi abordado primeiro por Dalenius (1950) que determinou equações mínimas fornecendo limites ótimos para alocações de Neyman e proporcional. No entanto, a solução dessas equações apresenta dificuldades de cálculos, sendo em alguns casos, até mais trabalhosa que minimizar $V(\bar{y}_{st})$ diretamente por tentativa. Para evitar essas dificuldades, Dalenius e Hodges, considerando o caso de alocação de Neyman,

propuseram tomar valores $[y_h]$ com base na frequência acumulada de $\sqrt{f(y)}$ onde $f(y)$ é uma função densidade.

No item 1.2 deste capítulo apresentaremos as equações que determinam os limites ótimos dos estratos. No item 1.3 serão abordados métodos de aproximação para o conjunto $[y_h]$ satisfazendo as equações minimais. Como ilustração, no item 1.5 apresentaremos um exemplo numérico para a função densidade $f(y) = e^{-y}$, $y \geq 0$, com a determinação de pontos limites de estratificação aproximadamente ótimos.

1.2. Limite Ótimo dos Estratos

Seja y a variável de estratificação com função densidade $f(y)$ e média

$$\mu = \int_{-\infty}^{\infty} t f(t) dt \quad (1.3)$$

a amplitude y_0, y_L da variável y é separada em L partes com pontos $y_1 < y_2 < \dots < y_{L-1}$, cada parte correspondendo a um estrato. Para o h -ésimo estrato:

$$\omega_h = \int_{y_{h-1}}^{y_h} f(t) dt \quad (1.4)$$

$$\mu_h = \frac{1}{\omega_h} \int_{y_{h-1}}^{y_h} t f(t) dt \quad (1.5)$$

$$\sigma_h^2 = \frac{1}{\omega_h} \int_{y_{h-1}}^{y_h} t^2 f(t) dt - \mu_h^2 \quad (1.6)$$

$$\mu = \sum_h \omega_h \mu_h \quad (1.7)$$

Seja $[y_h]$ o conjunto de pontos ótimos de estratificação para o qual $V(\bar{y}_{st})$ é mínima. Estes pontos $[y_h]$ são as soluções das equações minimais que são obtidas quando igualamos a zero as derivadas parciais de $V(\bar{y}_{st})$ com respeito a $[y_h]$. Nós obteremos essas equações para alocações de Neyman e proporcional.

1.2.1 Limites Ótimos de Estratificação para Alocação de Neyman

Nós desejamos minimizar $V(\bar{y}_{st})_N = \left[\sum_h \omega_h \sigma_h \right]^2 / n$ e que é equivalente a minimizar

$$\sum_h \omega_h \sigma_h \quad (1.8)$$

Para facilidade de cálculos iremos considerar $y_{h+1} = y_i$. Igualando a zero as derivadas parciais da equação (1.8) com respeito a y_h , nós chegamos a

$$\omega_h \frac{\partial \sigma_h}{\partial y_h} + \sigma_h \frac{\partial \omega_h}{\partial y_h} + \omega_i \frac{\partial \sigma_i}{\partial y_h} + \sigma_i \frac{\partial \omega_i}{\partial y_h} = 0 \quad (1.9)$$

mas,

$$\begin{aligned}
 \frac{\partial \sigma_h^2}{\partial y_h} &= \frac{1}{\omega_h} y_h^2 f(y_h) - \frac{f(y_h)}{\omega_h^2} \int_{y_{h-1}}^{y_h} t^2 f(t) dt - 2\mu_h \left[\frac{1}{\omega_h} y_h f(y_h) - \frac{f(y_h)}{\omega_h^2} \int_{y_{h-1}}^{y_h} t f(t) dt \right] \\
 &= \frac{1}{\omega_h} y_h^2 f(y_h) - \frac{f(y_h)}{\omega_h} \left[\sigma_h^2 + \mu_h^2 \right] - 2 \frac{\mu_h}{\omega_h} y_h f(y_h) + 2 \frac{\mu_h^2}{\omega_h} f(y_h) \\
 &= \frac{f(y_h)}{\omega_h} \left[y_h^2 - \sigma_h^2 - \mu_h^2 - 2\mu_h y_h + 2\mu_h^2 \right] \\
 &= \frac{f(y_h)}{\omega_h} \left[(y_h - \mu_h)^2 - \sigma_h^2 \right] \tag{1.10}
 \end{aligned}$$

De modo análogo

$$\frac{\partial \sigma_i^2}{\partial y_h} = - \frac{f(y_h)}{\omega_i} \left[(y_h - \mu_i)^2 - \sigma_i^2 \right] \tag{1.11}$$

$$\frac{\partial \omega_h}{\partial y_h} = f(y_h) = - \frac{\partial \omega_i}{\partial y_h} \tag{1.12}$$

$$\frac{\partial \sigma_h}{\partial y_h} = \frac{\partial}{\partial y_h} \sqrt{\sigma_h^2} = \frac{1}{2\sigma_h} \frac{\partial \sigma_h^2}{\partial y_h} \tag{1.13}$$

usando os resultados acima na equação (1.9), temos

$$\begin{aligned}
 \omega_h \frac{1}{2\sigma_h} \frac{f(y_h)}{\omega_h} \left[(y_h - \mu_h)^2 - \sigma_h^2 \right] + f(y_h) \sigma_h - \omega_i \frac{1}{2\sigma_i} \frac{f(y_h)}{\omega_i} \left[(y_h - \mu_i)^2 - \sigma_i^2 \right] \\
 - f(y_h) \sigma_i = 0
 \end{aligned}$$

$$\begin{aligned}
 \frac{f(y_h)}{2\sigma_h} \left[(y_h - \mu_h)^2 - \sigma_h^2 + 2\sigma_h^2 \right] &= \frac{f(y_h)}{2\sigma_i} \left[(y_h - \mu_i)^2 - \sigma_i^2 + 2\sigma_i^2 \right] \\
 \frac{(y_h - \mu_h)^2 + \sigma_h^2}{\sigma_h} &= \frac{(y_h - \mu_i)^2 + \sigma_i^2}{\sigma_i} \tag{1.14}
 \end{aligned}$$

onde $i = h+1, h = 1, 2, \dots, L-1$.

O conjunto $[y_h]$ satisfazendo equação (1,14) é o conjunto de pontos ótimos de estratificação que minimizam $V(\bar{y}_{st})_N$.

1.2.2. Limites Ótimos de Estratificação para Alocação Proporcional

Com a alocação proporcional, nós desejamos minimizar $V(\bar{y}_{st})_p = \left[\sum_h \omega_h \sigma_h^2 \right] / n$ o que é equivalente a minimizar $\sum_h \omega_h \sigma_h^2$. (1.15)

Iguälando a zero as derivadas parciais da equação (1.15) com respeito a y_h , temos

$$\omega_h \frac{\partial}{\partial y_h} \sigma_h^2 + \sigma_h^2 \frac{\partial \omega_h}{\partial y_h} + \omega_i \frac{\partial \sigma_i}{\partial y_h} + \sigma_i^2 \frac{\partial \omega_i}{\partial y_h} = 0 \quad (1.16)$$

usando as equações (1,10), (1,11) e (1,12) na equação acima, nós chegamos a

$$f(y_h) \left[(y_h - \mu_h)^2 - \sigma_h^2 \right] + f(y_h) \sigma_h^2 - f(y_h) \left[(y_h - \mu_i)^2 - \sigma_i^2 \right] - f(y_h) \sigma_i^2 = 0$$

$$f(y_h) (y_h - \mu_h)^2 = f(y_h) (y_h - \mu_i)^2$$

$$y_h^2 + \mu_h^2 - 2y_h \mu_h = y_h^2 + \mu_i^2 - 2y_h \mu_i$$

$$(\mu_h - \mu_i) (\mu_h + \mu_i - 2y_h) = 0$$

$$y_h = \frac{\mu_h + \mu_i}{2}, \quad i = h+1, \quad h = 1, 2, \dots, L-1. \quad (1.17)$$

Nós observamos que ambos os sistemas (1.14) e (1.17) são funções de parâmetros populacionais que por sua vez são funções das soluções destas equações. Devido a essas dificuldades de cálculo, Dalenius e Hodges desenvolveram um método de aproximação da solução exata $[y_h]$ satisfazendo eq. (1.14), bastante aplicável em situações práticas.

1.3. Soluções Aproximadas dos Pontos Ótimos de Estratificação

1.3.1. Primeira Aproximação

Seja $f(y)$ uma função densidade e y_0, y_1, \dots, y_L , pontos limites de estratificação que satisfazem equação (1.14). Nosso problema é encontrar um conjunto $[y'_h]$ que possa ser tomado como uma primeira aproximação de $[y_h]$.

Para determinação desse conjunto $[y'_h]$ considere

$$Z(y) = \int_{-\infty}^y \sqrt{f(t)} dt \quad (1.18)$$

Quando $y \rightarrow \infty$, $Z(y)$ aproxima-se de um limite superior H . As raízes $y'_1, y'_2, \dots, y'_{L-1}$ das equações

$$Z(y) = \frac{h}{t} H, \quad h = 1, 2, \dots, L-1 \quad (1.19)$$

podem ser tomadas para L grande, como primeira aproximação dos pontos y_1, y_2, \dots, y_{L-1} satisfazendo (1.14). Em outras palavras, o valor do $\frac{h}{t} \sqrt{f}$ é constante em todos os estratos.

Uma justificativa de Dalenius e Hodges (1959) para o exposto acima é dada como segue.

Quando o número L de estratos é grande, os estratos estarão próximos, e cada um deles terá distribuição aproximadamente constante dentro de um dado estrato. Assim,

$$\omega_h = \int_{y_{h-1}}^{y_h} f(t) dt \sim f_h (y_h - y_{h-1}) \quad (1.20)$$

onde f_h é o valor "constante" de $f(y)$ no estrato h .

$$\sigma_h \sim \frac{1}{\sqrt{12}} (y_h - y_{h-1}) \quad (1.21)$$

Substituindo essas aproximações em (1.1), nós encontramos

$$\begin{aligned} \sqrt{12} \sum_h \omega_h \sigma_h &\sim \sum_h f_h (y_h - y_{h-1})^2 \\ &= \sum_h \left[\sqrt{f_h} (y_h - y_{h-1}) \right]^2 \\ &= \sum_h (Z_h - Z_{h-1})^2 \end{aligned} \quad (1.22)$$

a última soma é mínima quando $(Z_h - Z_{h-1})$ é constante, ou seja,

$$(Z_h - Z_{h-1}) = \int_{y_{h-1}}^{y_h} \sqrt{f(t)} dt \sim \sqrt{f_h} (y_h - y_{h-1}) \quad (1.23)$$

1.3.2 Segunda Aproximação

Suponha que a função densidade $f(y)$ é estratificada em L estratos. Dois estratos consecutivos são especificados

por y_{h-1} , y_h e y_{h+1} . Para simplificar as fórmulas, considere $y_{h-1} = y_g$, $y_h = y_h$ e $y_{h+1} = y_i$. O intervalo (y_g, y_h) corresponde ao h-ésimo estrato, enquanto (y_h, y_i) ao i-ésimo estrato.

Definindo

$$I_p(y) = \int_{-\infty}^y t^p f(t) dt \quad (1.24)$$

as médias condicionais μ_h e μ_i , e as variâncias σ_h^2 e σ_i^2 dos estratos podem ser expressos em termos de $I_p(y)$ como segue:

$$\begin{aligned} \mu_h &= \frac{\int_{y_g}^{y_h} t f(t) dt}{\int_{y_g}^{y_h} f(t) dt} = \frac{\int_{-\infty}^{y_h} t f(t) dt - \int_{-\infty}^{y_g} t f(t) dt}{\int_{-\infty}^{y_h} f(t) dt - \int_{-\infty}^{y_g} f(t) dt} \\ &= \frac{I_1(y_h) - I_1(y_g)}{I_0(y_h) - I_0(y_g)} \end{aligned} \quad (1.25)$$

$$\begin{aligned} \sigma_h^2 &= \frac{\int_{y_g}^{y_h} t^2 f(t) dt}{\int_{y_g}^{y_h} f(t) dt} - \mu_h^2 = \frac{\int_{-\infty}^{y_h} t^2 f(t) dt - \int_{-\infty}^{y_g} t^2 f(t) dt}{\int_{-\infty}^{y_h} f(t) dt - \int_{-\infty}^{y_g} f(t) dt} - \mu_h^2 \\ &= \frac{I_2(y_h) - I_2(y_g)}{I_0(y_h) - I_0(y_g)} - \mu_h^2 \end{aligned} \quad (1.26)$$

Definindo

$$J_{ph} = I_p(y_h) - I_p(y_g) \quad (1.27)$$

temos

$$\mu_h = \frac{J_{1h}}{J_{oh}} \quad (1.28)$$

$$\begin{aligned} \sigma_h^2 &= \frac{J_{2h}}{J_{oh}} - \frac{J_{1h}^2}{J_{oh}^2} \\ &= \frac{J_{oh} J_{2h} - J_{1h}^2}{J_{oh}^2} \end{aligned} \quad (1.29)$$

De maneira análoga

$$\mu_i = \frac{J_{1i}}{J_{oi}} \quad (1.30)$$

$$\sigma_i^2 = \frac{J_{oi} J_{2i} - J_{1i}^2}{J_{oi}^2} \quad (1.31)$$

substituindo os valores da equação (1.14) em termos de J_{ph} e J_{pi} temos

$$\frac{\frac{J_{oh} J_{2h} - J_{1h}^2}{J_{oh}^2} + \left(y_h - \frac{J_{1h}}{J_{oh}} \right)^2}{\sqrt{\frac{J_{oh} J_{2h} - J_{1h}^2}{J_{oh}^2}}} = \frac{\frac{J_{oi} J_{2i} - J_{1i}^2}{J_{oi}^2} + \left(y_h - \frac{J_{1i}}{J_{oi}} \right)^2}{\sqrt{\frac{J_{oi} J_{2i} - J_{1i}^2}{J_{oi}^2}}} \quad (1.32)$$

Desenvolvendo o lado esquerdo de (1.32) temos

$$\frac{\frac{J_{oh} J_{2h} - J_{1h}^2 + J_{oh}^2 y_h^2 + J_{1h}^2 - 2y_h J_{oh} J_{1h}}{J_{oh}^2}}{\frac{\sqrt{J_{oh} J_{2h} - J_{1h}^2}}{J_{oh}}} = \frac{J_{2h} - 2y_h J_{1h} + y_h^2 J_{oh}}{\sqrt{J_{oh} J_{2h} - J_{1h}^2}} \quad (1.33)$$

Procedendo de modo análogo para o lado direito da equação (1.32), podemos escrever

$$\frac{J_{2h} - 2y_h J_{1h} + y_h^2 J_{oh}}{\sqrt{J_{oh} J_{2h} - J_{1h}^2}} - \frac{J_{2i} - 2y_h J_{1i} + y_h^2 J_{oi}}{\sqrt{J_{oi} J_{2i} - J_{1i}^2}} = 0 \quad (1.34)$$

Para simplificar, esta equação pode ser escrita como

$$A_h - B_h = \Delta_h = 0 \quad (1.35)$$

O conjunto $[y_h]$ satisfazendo equação (1.35) corresponde à estratificação de mínima variância. Se nós substituirmos algum outro valor, digamos y'_h , nós iremos denotar os lados esquerdo e direito da equação (1.34) por A'_h e B'_h respectivamente e sua diferença por Δ'_h .

Em geral não podemos supor que o conjunto $[y'_h]$ derivado da equação (1.19) satisfaça a equação (1.34). Descreveremos a seguir, um método de ajustamento do conjunto inicial $[y'_h]$ em um conjunto $[y''_h]$ satisfazendo equação (1.34).

Considere uma distribuição retangular $f(y) = 1$, $0 \leq y \leq 1$. Para um número L de estratos, vamos considerar três estratos consecutivos com índices g, h, i . Neste caso, temos

$$\left. \begin{aligned} J_{0h} &= I_0(y'_h) - I_0(y'_g) = \int_0^{y'_h} dt - \int_0^{y'_g} dt = (y'_h - y'_g) \\ J_{1h} &= I_1(y'_h) - I_1(y'_g) = \int_0^{y'_h} t dt - \int_0^{y'_g} t dt = \frac{1}{2}(y'^2_h - y'^2_g) \\ J_{2h} &= I_2(y'_h) - I_2(y'_g) = \int_0^{y'_h} t^2 dt - \int_0^{y'_g} t^2 dt = \frac{1}{3}(y'^3_h - y'^3_g) \end{aligned} \right\} (1.36)$$

De modo análogo

$$\left. \begin{aligned} J_{0i} &= (y'_i - y'_h) \\ J_{1i} &= \frac{1}{2}(y'^2_i - y'^2_h) \\ J_{2i} &= \frac{1}{3}(y'^3_i - y'^3_h) \end{aligned} \right\} (1.37)$$

Substituindo os valores (1.36) e (1.37) na equação (1.33) temos

$$\begin{aligned} A'_h &= \frac{J_{2h} - 2y'_h J_{1h} + y'^2_h J_{0h}}{\sqrt{J_{0h} J_{2h} - J_{1h}^2}} \\ &= \frac{1/3(y'^3_h - y'^3_g) - y'_h(y'^2_h - y'^2_g) + y'^2_h(y'_h - y'_g)}{\sqrt{1/3(y'_h - y'_g)(y'^3_h - y'^3_g) - 1/4(y'^2_h - y'^2_g)^2}} \\ &= \frac{1/3 \left[(y'^3_h - y'^3_g) - 3y'_h(y'^2_h - y'^2_g) + 3y'^2_h(y'_h - y'_g) \right]}{1/2\sqrt{3} \left[\sqrt{4(y'_h - y'_g)(y'^3_h - y'^3_g) - 3(y'^2_h - y'^2_g)^2} \right]} \end{aligned}$$

$$\begin{aligned}
 &= \frac{2}{\sqrt{3}} \frac{(y_h'^3 - y_g'^3 + 3y_g'^2 y_h' - 3y_g' y_h'^2)}{\sqrt{y_g'^4 - 4y_g' y_h'^3 - 4y_g'^3 y_h' + 6y_g'^2 y_h'^2 + y_h'^4}} \\
 &= \frac{2}{\sqrt{3}} \frac{(y_h' - y_g')(y_h' - y_g')^2}{\sqrt{(y_h' - y_g')^4}} \\
 &= \frac{2}{\sqrt{3}} (y_h' - y_g') \tag{1.38}
 \end{aligned}$$

De modo análogo

$$\begin{aligned}
 B_h' &= \frac{J_{2i} - 2y_h' J_{li} + y_h'^2 J_{oi}}{\sqrt{J_{oi} J_{2i} - J_{li}^2}} \\
 &= \frac{2}{\sqrt{3}} (y_i' - y_h') \tag{1.39}
 \end{aligned}$$

Desta forma

$$\left. \begin{aligned}
 \frac{\partial A_h'}{\partial y_g'} &= \frac{\partial B_h'}{\partial y_h'} = \frac{-2}{\sqrt{3}} \\
 \frac{\partial A_h'}{\partial y_h'} &= \frac{\partial B_h'}{\partial y_i'} = \frac{2}{\sqrt{3}}
 \end{aligned} \right\} \tag{1.40}$$

as expressões do tipo $\frac{\partial A_h'}{\partial y_i'}$, $\frac{\partial B_h'}{\partial y_g'}$ etc, são iguais a zero .

De $\Delta_h' = A_h' - B_h'$ nós derivamos as seguintes expressões

$$\left. \begin{aligned} \frac{\partial \Delta'_h}{\partial y'_g} &= \frac{\partial \Delta'_h}{\partial y'_i} = \frac{-2}{\sqrt{3}} \\ \frac{\partial \Delta'_h}{\partial y'_h} &= \frac{4}{\sqrt{3}} \end{aligned} \right\} \quad (1.41)$$

Nós temos um conjunto $[y'_h]$ com os correspondentes Δ'_h e desejamos determinar um conjunto $[y''_h]$ com $|\Delta''_h| < |\Delta'_h|$. Pelo teorema do valor médio, nós temos

$$\Delta''_h = \Delta'_h + (y''_g - y'_g) \frac{\partial \Delta'_h}{\partial y'_g} + (y''_h - y'_h) \frac{\partial \Delta'_h}{\partial y'_h} + (y''_i - y'_i) \frac{\partial \Delta'_h}{\partial y'_i}, \quad (1.42)$$

$h = 1, 2, \dots, g, h, i, \dots, L-1$, com $y''_0 = y'_0$ e $y''_L = y'_L$. Resolvendo este sistema de equações para y''_h , nós encontramos o conjunto desejado.

A matriz M de $\frac{\partial \Delta'_h}{\partial y'_j}$, $j=1, 2, \dots, g, h, i, \dots, L-1$, admitindo distribuição aproximadamente retangular, é a seguinte

$$M = \frac{2}{\sqrt{3}} \begin{pmatrix} 2 & -1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & -1 & 2 \end{pmatrix} \quad (1.43)$$

Se necessário, o processo pode ser repetido para determinar um terceiro conjunto $[y_h''']$ etc.

Como passo final nesse processo de ajustamento, espera-se que essas aproximações desenvolvidas para o caso retangular, sejam mantidas com bons resultados para outros casos.

1.4. Exemplo Numérico

Seja $f(y) = e^{-y}$, $y \geq 0$. Desejamos determinar pontos de estratificação y_1, y_2, \dots, y_{L-1} para $L = 2, 3, 4$ e 5 estratos tais que $V(\bar{y}_{st})_N = \left[\sum_h \omega_h \sigma_h \right]^2 / n$ seja mínima. Aqui, $V(\bar{y}_{st})_N$ é a variância da estimativa da média populacional, quando a amostra foi retirada sob alocação de Neyman.

Primeiramente vamos encontrar um conjunto $[y_h']$ como visto na secção 1.3.1 e que é a primeira aproximação para o conjunto $[y_h]$ satisfazendo equação (1.14). A seguir, faremos a segunda aproximação para o conjunto $[y_h]$ usando as equações (1.45) da secção 1.3.2. Finalmente, calcularemos $V(\bar{y}_{st})_N$ e apresentaremos um resumo dos resultados na tabela 1.4.

1.4.1. Determinação de $Z(y)$ e $I_p(y)$ para $f(y) = e^{-y}$

Da definição de $Z(y)$ e de $I_p(y)$, nós temos

$$Z(y) = \int_0^y \sqrt{e^{-t}} dt = 2(1 - e^{-y/2})$$

com $Z(0) = 0$ e $Z(\infty) = 2$.

$$I_p(y) = \int_0^y t^p e^{-t} dt, \text{ tal que}$$

$$I_0(y) = \int_0^y e^{-t} dt = 1 - e^{-y}.$$

$$I_1(y) = \int_0^y t e^{-t} dt = 1 - e^{-y} - ye^{-y}$$

$$I_2(y) = \int_0^y t^2 e^{-t} dt = 2 - e^{-y} (y^2 + 2y + 2)$$

1.4.2. Cálculos Numéricos para L=2 Estratos

Da equação (1.19), a primeira aproximação y_1' é dada pela raiz da seguinte equação

$$\int_0^y \sqrt{e^{-t}} dt = \frac{1}{2} \int_0^\infty \sqrt{e^{-t}} dt$$

$$2(1 - e^{-y/2}) = \frac{1}{2} \cdot 2 = 1$$

Neste caso, $y_1' = -2 \ln 1/2 = 1,39$.

Usando as funções $I_p(y)$ nós temos

$$I_0(0) = 0, \quad I_0(y_1') = 0,7509, \quad I_0(\infty) = 1$$

$$I_1(0) = 0, \quad I_1(y_1') = 0,4047, \quad I_1(\infty) = 1$$

$$I_2(0) = 0, \quad I_2(y_1') = 0,3280, \quad I_2(\infty) = 2$$

e de $J_{ph} = I_p(y_h) - I_p(y_g)$ temos

$$J_{o1}' = 0,7509, \quad J_{o2}' = 0,2491$$

$$J_{11}' = 0,4047, \quad J_{12}' = 0,5953$$

$$J_{21}' = 0,3280, \quad J_{22}' = 1,6720$$

TABELA 1.1

RESULTADOS NUMÉRICOS DA PRIMEIRA APROXIMAÇÃO PARA PONTOS
ÓTIMOS DE ESTRATIFICAÇÃO DE $f(y) = e^{-y}$ PARA $L=2$

	$y_0' = 0$	$y_1' = 1.39$	$y_2' = \infty$
$I_0(y_h')$	0,0000	0,7509	1.0000
$I_1(y_h')$	0,0000	0,4047	1.0000
$I_2(y_h')$	0,0000	0,3280	2.0000
J_{o1}'	0,7509		
J_{o2}'		0,2491	
J_{11}'	0,4047		
J_{12}'		0,5953	
J_{21}'	0,3280		
J_{22}'		1,6720	

nós calculamos

$$A_1' = \frac{J_{21}' - 2y_1' J_{11}' + y_1'^2 J_{o1}'}{\sqrt{J_{o1}' J_{21}' - J_{11}'^2}}$$

$$= \frac{0,3280 - 2 \times 1,39 \times 0,4047 + 1,39^2 \times 0,7509}{\sqrt{0,7509 \times 0,3280 - 0,4047^2}} = 2,2761$$

$$B_1' = \frac{J_{22}' - 2y_1' J_{12}' + y_1'^2 J_{02}'}{\sqrt{J_{02}' J_{22}' - J_{12}'^2}}$$

$$= \frac{1,6720 - 2 \times 1,39 \times 0,5953 + 1,39^2 \times 0,2491}{\sqrt{0,2491 \times 1,6720 - 0,5953^2}} = 2,0000$$

$$\Delta_1' = A_1' - B_1' = 2,2761 - 2,0000 = 0,2761$$

como Δ_1' é consideravelmente maior que zero, nós desejamos ajustar y_1' .

Da equação (1.45), temos

$$y_1'' = y_1' - \frac{\sqrt{3}}{4} \Delta_1' = 1,39 - \frac{\sqrt{3}}{4} \times 0,2761 = 1,27$$

Usando as equações $I_p(y)$ e J_{ph} , nós calculamos os resultados apresentados na tabela 1.2

TABELA 1.2
RESULTADOS NUMÉRICOS DA SEGUNDA APROXIMAÇÃO PARA PONTOS
ÓTIMOS DE ESTRATIFICAÇÃO DE $f(y) = e^{-y}$ PARA $L=2$

	$y_0'' = 0$	$y_1'' = 1,27$	$y_2'' = \infty$
$I_0(y_h'')$	0,0000	0,7192	1,0000
$I_1(y_h'')$	0,0000	0,3625	1,0000
$I_2(y_h'')$	0,0000	0,2721	2,0000
J_{01}''		0,7192	
J_{02}''			0,2808
J_{11}''		0,3625	
J_{12}''			0,6375
J_{21}''		0,2721	
J_{22}''			1,7279

$$A_1'' = \frac{J_{21}'' - 2y_1'' J_{11}'' + y_1''^2 J_{01}''}{\sqrt{J_{01}'' J_{21}'' - J_{11}''^2}} = 2,0162$$

$$B_1'' = \frac{J_{22}'' - 2y_1'' J_{12}'' + y_1''^2 J_{02}''}{\sqrt{J_{02}'' J_{22}'' - J_{12}''^2}} = 2,0007$$

E $\Delta_1'' = A_1'' - B_1'' = 2,0162 - 2,0007 = 0,0115$ é muito mais próximo de zero que Δ_1' . O ponto y_1'' é então, muito mais próximo do ponto y_1' que o ponto y_1' .

Na parte 1.4.5 calcularemos as variâncias correspondentes a y_1' e y_1'' .

1.4.3. Cálculos Numéricos para L=3 Estratos

A primeira aproximação y_1' e y_2' são dadas pelas raízes das seguintes equações

$$\left. \begin{aligned} 2(1 - e^{-y/2}) &= \frac{1}{3} \cdot 2 \\ 2(1 - e^{-y/2}) &= \frac{2}{3} \cdot 2 \end{aligned} \right\}$$

resultando em $y_1' = -2\ln(2/3) = 0,81$ e

$$y_2' = -2\ln(1/3) = 2,20$$

Usando as funções $I_p(y)$ e J_{ph} nós calculamos os resultados apresentados na tabela 1.3.

TABELA 1.3

RESULTADOS NUMÉRICOS DA PRIMEIRA APROXIMAÇÃO PARA PONTOS
 ÓTIMOS DE ESTRATIFICAÇÃO DE $f(y) = e^{-y}$ PARA $L=3$

	$y'_0 = 0$	$y'_1 = 0,81$	$y'_2 = 2,20$	$y'_3 = \infty$
$I_0(y'_h)$	0,0000	0,5551	0,8892	1,0000
$I_1(y'_h)$	0,0000	1,1948	0,6454	1,0000
$I_2(y'_h)$	0,0000	0,0977	0,7546	2,0000
J'_{01}	0,5551			
J'_{02}		0,3341		
J'_{03}			0,1108	
J'_{11}	1,1948			
J'_{12}		0,4506		
J'_{13}			0,3546	
J'_{21}	0,0977			
J'_{22}		0,6569		
J'_{23}			1,2454	

$$A'_1 = \frac{J'_{21} - 2y'_1 J'_{11} + y_1'^2 J'_{01}}{\sqrt{J'_{01} J'_{21} - J_1'^2}} = 1,1466$$

$$B'_1 = \frac{J'_{22} - 2y'_1 J'_{12} + y_1'^2 J'_{02}}{\sqrt{J'_{02} J'_{22} - J_1'^2}} = 1,1401$$

$$A'_2 = \frac{J'_{22} - 2y'_2 J'_{12} + y_2'^2 J'_{02}}{\sqrt{J'_{02} J'_{22} - J_1'^2}} = 2,2726$$

$$B_2' = \frac{J_{23}' - 2y_2' J_{13}' + y_2'^2 J_{03}'}{\sqrt{J_{03}' J_{23}' - J_{13}'^2}} = 2,0007$$

Então

$$\Delta_1' = A_1' - B_1' = 1,1466 - 1,1401 = 0,0065$$

$$\Delta_2' = A_2' - B_2' = 2,2726 - 2,0007 = 0,2719$$

a segunda aproximação y_1'' e y_2'' é dada por

$$y_1'' = y_1' - \frac{\sqrt{3}}{2 \times 3} \left[(3-1)\Delta_1' + \Delta_2' \right] = 0,73$$

$$y_2'' = y_1' - \frac{\sqrt{3}}{2 \times 3} \left[\Delta_1' + (3-1)\Delta_2' \right] = 2,04$$

1.4.4. Cálculos Numéricos para $L > 3$ Estratos

Cálculos semelhantes foram feitos para $L=4$ e $L=5$ estratos. Como são inteiramente análogos ao apresentado acima, não serão apresentados detalhadamente aqui. Os resultados estão resumidos na tabela 1.4.

1.4.5. Cálculo de Variâncias

Para o cálculo de $nV(\bar{y}_{st}) = \left[\sum_h \omega_h \sigma_h \right]^2$ precisamos calcular o valor de ω_h e σ_h , proporção e desvio padrão para $h = 1, 2, \dots, L$.

Para $L=2$, temos $y_0' = 0$, $y_1' = 1,39$ e $y_2' = \infty$. Usando as definições $I_p(y)$ e J_{ph} temos

$$\omega_1' = \int_0^{1,39} e^{-t} dt = I_0(1,39) = J_{o1}' = 0,7509$$

$$\omega_2' = \int_{1,39}^{\infty} e^{-t} dt = 1 - 0,7509 = J_{o2}' = 0,2491$$

Das equações (1.28) e (1.29) temos

$$\mu_h = \frac{J_{1h}}{J_{oh}} \quad e \quad \sigma_h^2 = \frac{J_{oh} J_{2h} - J_{1h}^2}{J_{oh}^2}$$

Então

$$\mu_1' = \frac{J_{11}'}{J_{o1}'} = \frac{0,4047}{0,7509} = 0,5390$$

$$\mu_2' = \frac{J_{12}'}{J_{o2}'} = \frac{0,5953}{0,2491} = 2,3898$$

$$\sigma_1' = \sqrt{\frac{J_{o1}' J_{21}' - J_{11}'^2}{J_{o1}'^2}} = \sqrt{\frac{0,7509 \times 0,3280 - (0,4047)^2}{(0,7509)^2}} = 0,3825$$

$$\sigma_2' = \sqrt{\frac{J_{o2}' J_{22}' - J_{12}'^2}{J_{o2}'^2}} = \sqrt{\frac{0,2491 \times 1,6720 - (0,5953)^2}{(0,2491)^2}} = 1,0005$$

$$\begin{aligned} nV(\bar{y}_{st}) &= \left[\sum_h \omega_h \sigma_h \right]^2 \\ &= (0,7509 \times 0,3825 + 0,2491 \times 1,0005)^2 = 0,2855 \end{aligned}$$

Com a segunda aproximação, temos $y_o''=0$, $y_1''=1,27$ e $y_2'' = \infty$

$$\omega_1'' = 0,7192, \quad \sigma_1'' = 0,3525$$

$$\omega_2'' = 0,2808, \quad \sigma_2'' = 0,9996$$

$$nV(\bar{y}_{st}) = (0,7192 \times 0,3525 + 0,2808 \times 0,9996)^2 = 0,2855 .$$

Para $L > 2$ estratos o procedimento é análogo.

Na tabela abaixo, apresentamos um resumo dos resultados obtidos para $L=2,3,4$ e 5 estratos para a função $f(y)=e^{-y}$.

TABELA 1.4

PRIMEIRA E SEGUNDA APROXIMAÇÃO PARA PONTOS ÓTIMOS DE ESTRATIFICAÇÃO DE $f(y)=e^{-y}$ PARA $L=2,3,4$ e 5 ESTRATOS E SUAS VARIÂNCIAS

NÚMERO DE ESTRATOS	PONTOS DE ESTRATIFICAÇÃO					$nV(\bar{y}_{et})$ $= \left[\sum_h \omega_h \sigma_h \right]^2$
	1 ^a APROX.	y_1'	y_2'	y_3'	y_4'	
2	2 ^a APROX.	y_1''	y_2''	y_3''	y_4''	
	1 ^a	1,39				0,2877
3	2 ^a	1,27				0,2855
	1 ^a	0,81	2,20			0,1345
4	2 ^a	0,73	2,04			0,1339
	1 ^a	0,58	1,39	2,77		0,0777
5	2 ^a	0,52	1,27	2,61		0,0774
	1 ^a	0,45	1,02	1,83	3,22	0,0503
	2 ^a	0,39	0,92	1,68	3,02	0,0501

Observamos na tabela acima para a função $f(y) = e^{-y}$, que não há uma melhora substancial nas variâncias quando do uso da segunda aproximação com as diferenças entre elas tornando-se ainda menores à medida que o número de estratos cresce. Isto nos sugere que a primeira aproximação já fornece bons limites de estratificação.

CAPÍTULO 2

MÉTODO DE ESTRATIFICAÇÃO DEVIDO A EKMAN

2.1. Introdução

Suponha que desejamos determinar pontos $y_1 < y_2 < \dots < y_{L-1}$ tais que $V(\bar{y}_{st})_N = n^{-1} \left[\sum_h \omega_h \sigma_h \right]^2$ seja mínima. Aqui, $V(\bar{y}_{st})_N$ é a variância da média sob alocação de Neyman, ω_h e σ_h são a proporção e o desvio padrão no estrato h .

Delenius mostrou que os pontos $[y_h]$ que minimizam $\sum_h \omega_h \sigma_h$ são aqueles que satisfazem as equações (1.14), igual a.

$$\frac{\sigma_h^2 + (y_h - \mu_h)^2}{\sigma_h} = \frac{\sigma_{h+1}^2 + (y_h - \mu_{h+1})^2}{\sigma_{h+1}}, \quad h=1, 2, \dots, L-1.$$

Como já foi comentado no capítulo anterior, a solução destas equações envolvem o conhecimento de μ_h e σ_h , média e variância no h -ésimo estrato, que por sua vez não podem ser calculados até que $[y_h]$ seja conhecido, levando a complicados cálculos iterativos. Desta forma, alguns estu-

diosos - e em particular Ekman-(1959) procuraram soluções aproximadas para $[y_h]$ satisfazendo as equações minimais.

Ekman mostrou que para uma densidade $f(y)$ com amplitude finita, os pontos $[y_h]$ satisfazendo $(y_h - y_{h-1}) \omega_h = C_L$, $h=1,2,\dots,L$, onde C_L é uma constante que depende do número L de estratos, são uma aproximação para a solução exata.

Para uma densidade $f(y)$ com amplitude infinita, a equação acima não pode ser aplicada, mas com alguma modificação, chegamos às equações

$$(y_h - y_{h-1}) \omega_h = C_L, \quad h=2,3,\dots,(L-1)$$

$$\frac{3f(y_1)}{4} \left[\frac{\sigma_1^2 + (y_1 - \mu_1)^2}{\sigma_1} \right] = \frac{3f(y_{L-1})}{4} \left[\frac{\sigma_L^2 + (y_{L-1} - \mu_L)^2}{\sigma_L} \right] = C_L$$

que podem ser usadas no lugar das equações obtidas para uma densidade com amplitude finita.

Os resultados obtidos neste capítulo, serão derivados sob a suposição de L grande, correspondendo a pequenos intervalos $(y_h - y_{h-1})$. Quando L se aproxima do infinito, pode-se provar que a solução exata e esta aproximação, são equivalentes. Na prática, no entanto, L é frequentemente pequeno, dificilmente superior a 5, o que certamente não satisfaz a suposição de L grande. Verifica-se entretanto, que mesmo para poucos estratos, $L=2,3$ etc, a aproximação acima produz bons resultados. Ekman comparou os resultados obtidos de três funções densidade, $f(y) = 2(1-y)$, $f(y) = e^{-y}$ e $f(y) = ye^{-y}$,

calculando as variâncias da média para estratos com pontos satisfazendo a solução mínima e com pontos aproximados, para $L=2,3,4$ e 5 estratos. As diferenças encontradas foram muito pequenas, o que justifica plenamente o uso das soluções aproximadas.

Neste capítulo, a título de ilustração, nós trabalharemos apenas com a função densidade $f(y) = e^{-y}$.

É conveniente observar que para funções densidade com uma longa "cauda", como por exemplo distribuições χ^2 com um grande número de graus de liberdade, são esperados resultados mais "pobres".

Um outro ponto a ser abordado neste capítulo, é a determinação de aproximações para a constante C_L , nos itens 2.4 e 2.5.

2.2. Método de Aproximação de Limites Ótimos de Estratificação para uma Amplitude Finita

Considere uma função densidade $f(y)$ com amplitude finita (y_0, y_L) e tal que as derivadas de primeira e segunda ordem, $f'(y)$ e $f''(y)$ existem e são contínuas sobre a amplitude total. Para maior facilidade de cálculo nós introduzimos uma função $H(y)$ da seguinte maneira:

$$H'''(y) = f(y)$$

$$H''(y) = \int_{-\infty}^y H'''(t) dt$$

$$H'(y) = \int_{-\infty}^y H''(t) dt \quad (2.1)$$

$$H(y) = \int_{-\infty}^y H'(t) dt$$

onde $H'(y)$, $H''(y)$, $H'''(y)$ existem e são contínuas.

Usando integração parcial, nós obtemos

$$\int_a^b f(y) dy = \int_{-\infty}^b H'''(y) dy - \int_{-\infty}^a H'''(y) dy = H''(b) - H''(a) \quad (2.2)$$

$$\begin{aligned} \int_a^b y f(y) dy &= \int_{-\infty}^b y H'''(y) dy - \int_{-\infty}^a y H'''(y) dy \\ &= bH''(b) - \int_{-\infty}^b H''(y) dy - aH''(a) + \int_{-\infty}^a H''(y) dy \\ &= bH''(b) - H'(b) - aH''(a) + H'(a) \\ &= [bH''(b) - aH''(a)] - [H'(b) - H'(a)] \quad (2.3) \end{aligned}$$

$$\begin{aligned} \int_a^b y^2 f(y) dy &= \int_{-\infty}^b y^2 H'''(y) dy - \int_{-\infty}^a y^2 H'''(y) dy \\ &= b^2 H''(b) - 2 \int_{-\infty}^b y H''(y) dy - a^2 H''(a) + 2 \int_{-\infty}^a y H''(y) dy \\ &= b^2 H''(b) - 2bH'(b) + 2 \int_{-\infty}^b H'(y) dy - a^2 H''(a) + 2aH'(a) \\ &\quad - 2 \int_{-\infty}^a H'(y) dy \end{aligned}$$

$$= [b^2 H''(b) - a^2 H''(a)] - 2 [bH'(b) - aH'(a)] + 2 [H(b) - H(a)] \quad (2.4)$$

Desenvolvendo $\int_{y_{h-1}}^{y_h} (y_h - y)^k f(y) dy$ para $k=0,1,2$ temos

$$\left. \begin{aligned} \int_{y_{h-1}}^{y_h} (y_h - y)^0 f(y) dy &= \int_{y_{h-1}}^{y_h} f(y) dy \\ \int_{y_{h-1}}^{y_h} (y_h - y) f(y) dy &= y_h \int_{y_{h-1}}^{y_h} f(y) dy - \int_{y_{h-1}}^{y_h} y f(y) dy \\ \int_{y_{h-1}}^{y_h} (y_h - y)^2 f(y) dy &= y_h^2 \int_{y_{h-1}}^{y_h} f(y) dy - 2y_h \int_{y_{h-1}}^{y_h} y f(y) dy + \int_{y_{h-1}}^{y_h} y^2 f(y) dy \end{aligned} \right\} (2.5)$$

por outro lado, sabemos que

$$\left. \begin{aligned} \omega_h \mu_h &= \int_{y_{h-1}}^{y_h} y f(y) dy \\ \omega_h (\sigma_h^2 + \mu_h^2) &= \int_{y_{h-1}}^{y_h} y^2 f(y) dy \end{aligned} \right\} (2.6)$$

Usando as equações de (2.2) a (2.6), obtemos o seguinte resultado

$$\begin{aligned} \omega_h [\sigma_h^2 + (y_h - \mu_h)^2] &= \omega_h [\sigma_h^2 + \mu_h^2] - 2y_h \omega_h \mu_h + \omega_h y_h^2 \\ &= \int_{y_{h-1}}^{y_h} y^2 f(y) dy - 2y_h \int_{y_{h-1}}^{y_h} y f(y) dy + y_h^2 \int_{y_{h-1}}^{y_h} f(y) dy \end{aligned}$$

$$\begin{aligned}
 &= 2 \left[H(y_h) - H(y_{h-1}) \right] - 2 \left[y_h H'(y_h) - y_{h-1} H'(y_{h-1}) + y_h^2 H''(y_h) \right. \\
 &\quad \left. - y_{h-1}^2 H''(y_{h-1}) - 2y_h \left\{ - \left[H'(y_h) - H'(y_{h-1}) \right] + \left[y_h H''(y_h) - y_{h-1} H''(y_{h-1}) \right] \right\} \right. \\
 &\quad \left. + y_h^2 \left[H''(y_h) - H''(y_{h-1}) \right] \right] \\
 &= 2 \left\{ H(y_h) - H(y_{h-1}) + y_{h-1} H'(y_{h-1}) - y_h H'(y_{h-1}) + y_h y_{h-1} H''(y_{h-1}) \right. \\
 &\quad \left. - \frac{y_{h-1}^2 H''(y_{h-1})}{2} - \frac{y_h^2 H''(y_{h-1})}{2} \right\} \\
 &= 2 \left\{ H(y_h) - \left[H(y_{h-1}) + (y_h - y_{h-1}) H'(y_{h-1}) + \frac{(y_h - y_{h-1})^2}{2!} H''(y_{h-1}) \right] \right\} \quad (2.7)
 \end{aligned}$$

$$\omega_h(y_h - \mu_h) = \omega_h y_h - \omega_h \mu_h$$

$$\begin{aligned}
 &= y_h \int_{y_{h-1}}^{y_h} f(y) dy - \int_{y_{h-1}}^{y_h} y f(y) dy \\
 &= y_h \left[H''(y_h) - H''(y_{h-1}) \right] + H'(y_h) - H'(y_{h-1}) - y_h H''(y_h) - y_{h-1} H''(y_{h-1}) \\
 &= \left\{ H'(y_h) - \left[H'(y_{h-1}) + (y_h - y_{h-1}) H''(y_{h-1}) \right] \right\} \quad (2.8)
 \end{aligned}$$

$$\begin{aligned}
 \omega_h &= \int_{y_{h-1}}^{y_h} f(y) dy \\
 &= H''(y_h) - H''(y_{h-1}) \quad (2.9)
 \end{aligned}$$

Nota-se que o lado direito das equações (2.7), (2.8) e (2.9) consistem de $H(y_h)$, $H'(y_h)$ e $H''(y_h)$, menos os primeiros termos em suas respectivas expansões de Taylor no ponto

$y=y_{h-1}$. Continuando estas expansões até a quinta, quarta e terceira ordem respectivamente, os seguintes resultados são obtidos

$$\omega_h \left[\sigma_h^2 + (y_h - \mu_h)^2 \right] = 2 \left[\frac{(y_h - y_{h-1})^3}{3!} H'''(y_{h-1}) + \frac{(y_h - y_{h-1})^4}{4!} H^{(4)}(y_{h-1}) + \frac{(y_h - y_{h-1})^5}{5!} H^{(5)}(\xi_1) \right]$$

$$\omega_h (y_h - \mu_h) = \left[\frac{(y_h - y_{h-1})^2}{2!} H'''(y_{h-1}) + \frac{(y_h - y_{h-1})^3}{3!} H^{(4)}(y_{h-1}) + \frac{(y_h - y_{h-1})^4}{4!} H^{(5)}(\xi_2) \right]$$

$$\omega_h = \left[(y_h - y_{h-1}) H'''(y_{h-1}) + \frac{(y_h - y_{h-1})^2}{2!} H^{(4)}(y_{h-1}) + \frac{(y_h - y_{h-1})^3}{3!} H^{(5)}(\xi_3) \right]$$

onde ξ_i são os pontos no intervalo (y_{h-1}, y_h) , $H^{(4)}(y_{h-1})$ e $H^{(5)}(y_{h-1})$ são as derivadas de 4^a e 5^a ordem da função H no ponto $y=y_{h-1}$, e que equivalem a $f'(y_{h-1})$ e $f''(y_{h-1})$ respectivamente. Para simplificação de cálculos notaremos $K=(y_h - y_{h-1})$ e H''' , $H^{(4)}$ e $H^{(5)}$ para as respectivas derivadas de H no ponto $y = y_{h-1}$.

Então, temos

$$\omega_h \left[\sigma_h^2 + (y_h - \mu_h)^2 \right] = \left[\frac{K^3}{3} H''' + \frac{K^4}{12} H^{(4)} + o(K^5) \right] \quad (2.10)$$

$$\omega_h (y_h - \mu_h) = \left[\frac{K^2}{2} H''' + \frac{K^3}{6} H^{(4)} + o(K^4) \right] \quad (2.11)$$

$$\omega_h = \left[KH''' + \frac{K^2}{2} H^{(4)} + o(K^3) \right] \quad (2.12)$$

Multiplicando (2.10) por (2.12), obtemos

$$\omega_h^2 \sigma_h^2 + (y_h - \mu_h)^2 = \frac{K^4}{3} H'''^2 + \frac{K^5}{4} H''' H^{(4)} + o(K^6) \quad (2.13)$$

e elevando (2.11) ao quadrado, temos

$$\omega_h^2 (y_h - \mu_h)^2 = \frac{K^4}{4} H''''^2 + \frac{K^5}{6} H'''' H^{(4)} + O(K^6) \quad (2.14)$$

Subtraindo (2.14) de (2.13), a seguinte identidade é obtida

$$(\omega_h \sigma_h)^2 = \frac{K^4}{12} \left[H'''' (H'''' + KH^{(4)}) + O(K^2) \right] \quad (2.15)$$

Elevando (2.10) ao quadrado, determinamos

$$\omega_h^2 \left[\sigma_h^2 + (y_h - \mu_h)^2 \right]^2 = \frac{K^6}{9} \left[H'''' (H'''' + \frac{K}{2!} H^{(4)}) + O(K^2) \right] \quad (2.16)$$

Para L grande, os termos de maior ordem em $K = (y_h - y_{h-1})$ podem ser negligenciados.

Dividindo (2.16) por (2.15) e substituindo H'''' e $H^{(4)}$ por f e f' , temos

$$\left[\frac{\sigma_h^2 + (y_h - \mu_h)^2}{\sigma_h} \right] \sim \frac{4}{3} K \times \frac{Kf + K^2 f' / 2!}{f + Kf'}$$

Ou seja:

$$\left[\frac{\sigma_h^2 + (y_h - \mu_h)^2}{\sigma_h} \right]^2 \sim \frac{4}{3} (y_h - y_{h-1}) \cdot \frac{(y_h - y_{h-1})f(y_{h-1}) + (y_h - y_{h-1})^2 f'(y_{h-1}) / 2!}{f(y_{h-1}) + (y_h - y_{h-1}) f'(y_{h-1})} \quad (2.17)$$

Observe que

$$\begin{aligned} \omega_h &= \int_{y_{h-1}}^{y_h} f(y) dy \\ &= (y_h - y_{h-1})f(y_{h-1}) + \frac{(y_h - y_{h-1})^2}{2!} f'(y_{h-1}) + \frac{(y_h - y_{h-1})^3}{3!} f''(y_{h-1}) + O(K^4) \end{aligned} \quad (2.18)$$

e

$$\begin{aligned} \frac{\partial \omega_h}{\partial y_h} &= f(y_h) \\ &= f(y_{h-1}) + (y_h - y_{h-1}) f'(y_{h-1}) + \frac{(y_h - y_{h-1})^2}{2!} f''(y_{h-1}) + O(K^3) \end{aligned} \quad (2.19)$$

No numerador do segundo fator do lado direito da equação (2.17) temos então os dois primeiros termos da expansão de Taylor de ω_h no ponto $y=y_{h-1}$, enquanto que o denominador é a derivada parcial do numerador com respeito a y_h , isto é, os dois primeiros termos da expansão de $f(y_h)$ no ponto $y=y_{h-1}$. Aproximando mais uma vez e negligenciando os termos de maior ordem na expansão de ω_h e $f(y_h)$, obtemos

$$\left[\frac{\sigma_h^2 + (y_h - \mu_h)^2}{\sigma_h} \right]^2 \sim \frac{4(y_h - y_{h-1}) \omega_h}{3f(y_h)} \quad (2.20)$$

Procedendo de modo análogo

$$\left[\frac{\sigma_{h+1}^2 + (y_h - \mu_{h+1})^2}{\sigma_{h+1}} \right]^2 \sim \frac{4(y_{h+1} - y_h) \omega_{h+1}}{3f(y_h)} \quad (2.21)$$

Aplicando este resultado para a equação da solução exata, temos

$$(y_h - y_{h-1}) \omega_h \sim (y_{h+1} - y_h) \omega_{h+1} \quad (2.22)$$

Finalmente, aplicando (2.22) para $h=1, 2, \dots, L-1$, nós determinamos a seguinte aproximação para a solução exata

$$(y_h - y_{h-1}) \omega_h = C_L \quad (2.23)$$

onde C_L é uma constante.

2.3. Método de Aproximação de Limites Ótimos de Estratificação para uma Amplitude Infinita

Se $y_0 = -\infty$ e $y_L = \infty$, a aproximação (2.23) pode ainda ser aplicada para $L=2,3,\dots (L-1)$. Desenvolvendo as identidades (2.20) e (2.21) para $h=1$ e $(L-1)$ respectivamente, podemos escrever

$$\left. \begin{aligned} \left[\frac{\sigma_1^2 + (y_1 - \mu_1)^2}{\sigma_1} \right]^2 &\sim \frac{4 C_L}{3f(y_1)} \\ \left[\frac{\sigma_L^2 + (y_{L-1} - \mu_L)^2}{\sigma_L} \right]^2 &\sim \frac{4 C_L}{3f(y_{L-1})} \end{aligned} \right\} \quad (2.24)$$

Ou seja:

$$\left. \begin{aligned} (y_h - y_{h-1}) \omega_h &= C_L, \quad h=2,3,\dots, L-1 \\ \frac{3}{4} f(y_1) \left[\frac{\sigma_1^2 + (y_1 - \mu_1)^2}{\sigma_1} \right]^2 &= \frac{3}{4} f(y_{L-1}) \left[\frac{\sigma_L^2 + (y_{L-1} - \mu_L)^2}{\sigma_L} \right]^2 = C_L \end{aligned} \right\} \quad (2.25)$$

Com as equações (2.25) uma aproximação para uma amplitude infinita pode ser obtida. Para muitas distribuições, os cálculos no desenvolvimento da segunda identidade da equação (2.25) são simples. Por exemplo, para $f(y) = e^{-y}$

$$\frac{3f(y_{L-1})}{4} \left[\frac{\sigma_L^2 + (y_{L-1} - \mu_L)^2}{\sigma_L} \right]^2 = 3f(y_{L-1})$$

pois

$$\left[\frac{\sigma_L^2 + (y_{L-1} - \mu_L)^2}{\sigma_L} \right]^2 = \left\{ \frac{1 + [y_{L-1} - (1-y_{L-1})]^2}{1} \right\}^2 = 4$$

2.4. Primeira Aproximação para a Constante C_L

Com o conhecimento da constante C_L , o conjunto $[y_h]$ satisfazendo (2.23) ou (2.25) pode ser facilmente determinado. Portanto, se faz necessário determinar pelo menos uma aproximação para C_L . O desenvolvimento de uma aproximação para C_L , é feito a partir da suposição de que a constante C_{L-r} já foi obtida.

Suponha que para uma densidade com amplitude finita $(y_L - y_0)$, o conjunto de pontos y_1, y_2, \dots, y_{L-1} foi encontrado tal que as relações (2.23) são satisfeitas (existe sempre uma única solução). De $(y_h - y_{h-1})\omega_h = C_L$ derivamos a identidade

$$\sum_h \left(\frac{y_h - y_{h-1}}{y_L - y_0} \right) \omega_h = \frac{L C_L}{(y_L - y_0)} \quad (2.26)$$

que pode ser considerada como uma média ponderada de $(y_h - y_{h-1}) / (y_L - y_0)$ por ω_h ou vice-versa. Para L grande

$$\frac{(y_h - y_{h-1})}{(y_L - y_0)} \sim \frac{1}{L} \quad (2.27)$$

pois $(y_h - y_{h-1}) = C = \text{constante}$, é assintoticamente correto quando os termos, a partir do primeiro, em (2.15) e (2.16) são negligenciados. Assim

$$(\omega_h \sigma_h)^2 \sim \frac{(y_h - y_{h-1})^4}{12} H'''^2(y_{h-1}) \quad (2.28)$$

$$\omega_h^2 \left[\sigma_h^2 + (y_h - \mu_h)^2 \right]^2 \sim \frac{(y_h - y_{h-1})^6}{9} H'''^2(y_{h-1}) \quad (2.29)$$

Então

$$\frac{\sigma_h^2 + (y_h - \mu_h)^2}{\sigma_h} \sim \frac{2}{\sqrt{3}} (y_h - y_{h-1}) \quad (2.30)$$

De maneira análoga

$$\frac{\sigma_{h+1}^2 + (y_h - \mu_{h+1})^2}{\sigma_{h+1}} \sim \frac{2}{\sqrt{3}} (y_{h+1} - y_h) \quad (2.31)$$

aplicando (2.30) e (2.31) em (1.14) temos $(y_h - y_{h-1}) \sim (y_{h+1} - y_h)$, e para $h=1, 2, \dots, L-1$, uma aproximação para a equação minimal é $(y_h - y_{h-1}) = C = \text{constante}$.

Desta forma, (2.26) torna-se igual a

$$\frac{1}{L} \sum_h \omega_h = \frac{1}{L} = \frac{L C_L}{(y_L - y_0)} \quad (2.32)$$

assim

$$C_L = \frac{(y_L - y_0)}{L^2} \quad (2.33)$$

E, para $C_{L-r} = \frac{(y_L - y_0)}{(L-r)^2}$, chegamos a

$$C'_L = \frac{(L-r)^2 C_{L-r}}{L^2} \quad (2.34)$$

como uma primeira aproximação da constante C_L . Adiante segue um exemplo numérico que torna mais claro a diferença entre C_L e C'_L .

2.5. Segunda Aproximação para a Constante C_L

Supondo que um conjunto $[y_h]$ satisfazendo (2.23) foi encontrado para algum número de estratos igual a $L-r$, uma primeira aproximação pode ser obtida de (2.34). Um conjunto $[y'_h]$ pode ser encontrado tal que todos os valores $(y'_h - y'_{h-1})\omega_h$ são em geral iguais a C'_L , com exceção de um valor $(y'_j - y'_{j-1})\omega_j$ (por exemplo $j=L$) que é diferente de C'_L . Então uma segunda aproximação C''_L para C_L pode ser obtida de

$$C''_L = \frac{(L-1)C'_L + (y'_j - y'_{j-1})\omega_j}{L} \quad (2.35)$$

Procedendo de maneira análoga, uma nova aproximação C'''_L pode ser encontrada para C_L , e assim por diante. A constante $C_L^{(i)}$ converge para C_L e o conjunto $[y_h^{(i)}]$ corresponde ao conjunto $[y_h]$ satisfazendo (2.23).

2.6. Exemplo Numérico

Seja $f(y) = e^{-y}$, $0 \leq y < \infty$. Para esta função densidade determinaremos o conjunto $[y_h]$ satisfazendo (2.25) usando o método de aproximação de Ekman para $L=2,3,4$ e 5 estratos.

No final deste item apresentaremos a tabela 2.1 com um resumo dos resultados com os pontos de estratificação e as variâncias associadas para estratificação ótima e estratificação baseada no princípio de aproximação de Ekman. Comparando essas variâncias, podemos observar que os pontos de estratificação calculados por esse método de aproximação pro-

duzem excelentes resultados para esta função densidade, mesmo para um número de estratos tão pequenos.

2.6.1. Determinação da Constante C_2 e do Limite do Estrato y_1 para L=2 Estratos

De (2.25) temos

$$\left. \begin{aligned} y_1(1 - e^{-y_1}) &= C_2 \\ 3e^{-y_1} &= C_2 \end{aligned} \right\}$$

Usando o método de aproximação de Newton (1965) considere

$$g(y_1) = 3e^{-y_1} - y_1(1 - e^{-y_1}) = (3 + y_1)e^{-y_1} - y_1 = 0$$

$$\frac{\partial g}{\partial y_1} = - \left[1 + (2 + y_1)e^{-y_1} \right]$$

$$y_1'' = y_1' + \frac{(3 + y_1') - y_1'}{1 + (2 + y_1')e^{-y_1'}}$$

Como as raízes estão entre 1 e 1,5 (pois $g(1) = 0,4715$ e $g(1,5) = -0,4959$), podemos escolher para aproximação inicial $y_1' = 1$, que determina

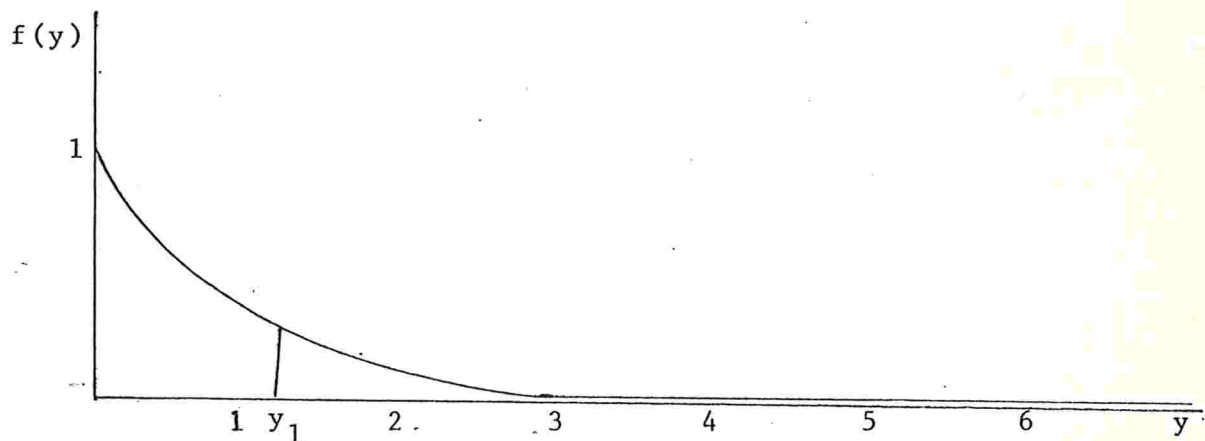
$$y_1'' = 1 + \frac{4e^{-1} - 1}{1 + 3e^{-1}} = 1,224$$

O valor de y_1'' será colocado no lugar de y_1' até que dois valores consecutivos sejam encontrados iguais, ou diferentes por quantidades negligenciáveis. Então

$$y_1''' = 1,224 + \frac{4,224 e^{-1,224} - 1,224}{1 + 3,224 e^{-1,224}} = 1,223$$

$$y_1^{(4)} = 1,223 + \frac{4,223 e^{-1,223} - 1,223}{1 + 3,223 e^{-1,223}} = 1,223$$

Logo $y_1 = 1,223$ e $C_2 = 1,223(1 - e^{-1,223}) = 0,8737$



2.6.2. Determinação da Constante C_3 e dos Limites dos Estratos $[y_h]$ para $L=3$ Estratos

2.6.2.1. Primeira Aproximação para C_3 e $[y_h]$

Para calcular a primeira aproximação para a constante C_3 vamos utilizar a constante C_2 na equação (2.34). Neste caso temos $r=1$ e

$$C_3' = \frac{(3-1)^2 \times 0,8737}{3^2} = 0,3883$$

De (2.25) temos

$$3 e^{-y_2'} = 0,3883 \therefore y_2' = 2,045$$

Por outro lado

$$\omega_1' = \int_0^{y_1'} e^{-t} dt = (1 - e^{-y_1'})$$

Então

$$y_1'(1 - e^{-y_1'}) = 0,3883$$

Para determinar y_1' , considere

$$g(y_1) = y_1(1 - e^{-y_1}) - 0,3883 = 0$$

$$\frac{\partial g}{\partial y_1} = y_1 e^{-y_1} + 1 - e^{-y_1}$$

$$y_1'' = y_1' + \frac{y_1'(1 - e^{-y_1'}) - 0,3883}{y_1' e^{-y_1'} + 1 - e^{-y_1'}}$$

Como as raízes estão entre 0,5 e 1 (pois $g(0,5) = -0,1916$ e $g(1) = 0,2438$), podemos escolher para aproximação inicial $y_1' = 0,5$, que determina

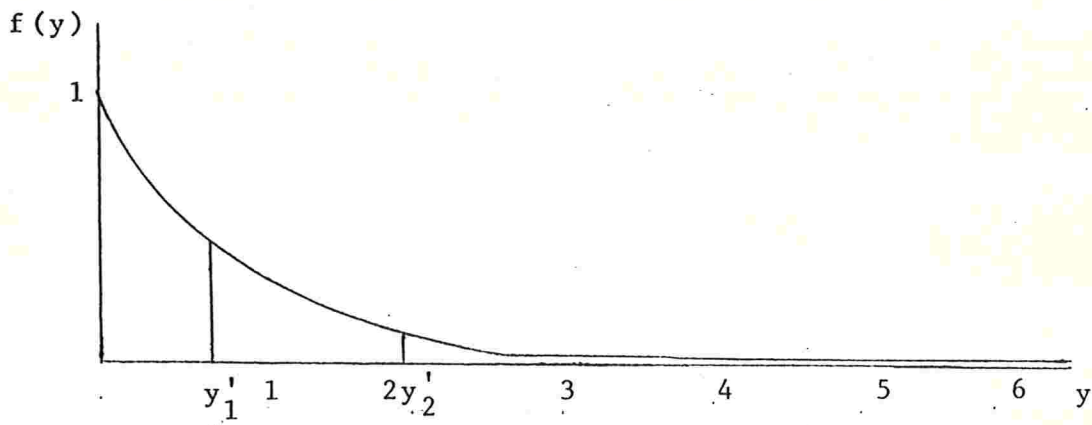
$$y_1'' = 0,5 - \frac{0,5(1 - e^{-0,5}) - 0,3883}{0,5 e^{-0,5} - e^{-0,5} + 1} = 0,775$$

O valor de y_1'' será colocado no lugar de y_1' até que dois valores consecutivos sejam encontrados iguais, ou diferentes por quantidades negligenciáveis. Então

$$y_1''' = 0,775 - \frac{0,775(1 - e^{-0,775}) - 0,3883}{0,775 e^{-0,775} - e^{-0,775} + 1} = 0,742$$

$$y_1^{(4)} = 0,742 - \frac{0,742(1 - e^{-0,742}) - 0,3883}{0,742 e^{-0,742} - e^{-0,742} + 1} = 0,742$$

Logo $y_1'(1 - e^{-y_1'}) = 0,3883$ implica em $y_1' = 0,742$



2.6.2.2. Segunda Aproximação para C_3 e $[y_h]$

Da equação (2.35), temos

$$C_3'' = \frac{(3-1)C_3' + (y_2' - y_1')\omega_2'}{3}$$

Mas

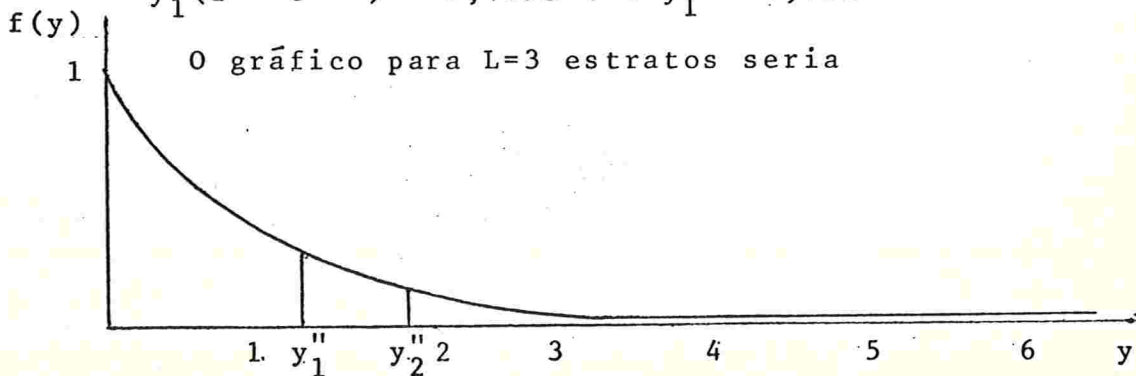
$$\omega_2' = \int_{0,742}^{2,045} e^{-t} dt = 0,3468$$

Então

$$C_3'' = \frac{2 \times 0,3883 + (2,045 - 0,742) 0,3468}{3} = 0,4095$$

Então $3 e^{-y_2''} = 0,4095 \therefore y_2'' = 1,991$

$y_1''(1 - e^{-y_1''}) = 0,4095 \therefore y_1'' = 0,766$



2.6.3. Determinação da Constante C_L e dos Limites dos Estratos $[y_h]$ para $L > 3$ Estratos

Para um número de estratos maior que três, o procedimento é análogo ao desenvolvido para $L=3$ estratos.

Um resumo dos resultados para $L=2,3,4$ e 5 estratos pode ser visto na tabela 2.1.

2.6.4. Cálculo de Variâncias

Desejamos calcular $n V(\bar{y}_{st})_N = \left[\sum_h \omega_h \sigma_h \right]^2$.

Para $L=2$ estratos, temos $y_0=0$, $y_1=1,223$ e $y_2=\infty$ e

$$\omega_1 = \int_0^{1,223} e^{-t} dt = 0,7057$$

$$\omega_2 = \int_{1,223}^{\infty} e^{-t} dt = 0,2943$$

$$\mu_1 = \frac{1}{\omega_1} \int_0^{1,223} t e^{-t} dt = 0,4898$$

$$\mu_2 = \frac{1}{\omega_2} \int_{1,223}^{\infty} t e^{-t} dt = 2,2233$$

$$\sigma_1^2 = \frac{1}{\omega_1} \left[\int_0^{1,223} t^2 e^{-t} dt \right] - \mu_1^2 = 0,1159 \quad \therefore \sigma_1 = 0,3404$$

$$\sigma_2^2 = \frac{1}{\omega_2} \left[\int_{1,223}^{\infty} t^2 e^{-t} dt \right] - \mu_2^2 = 0,9995 \quad \therefore \sigma_2 = 0,9997$$

$$\begin{aligned}
 n v(\bar{y}_{st})_N &= \left[\sum_h \omega_h \sigma_h \right]^2 \\
 &= [0,7057 \times 0,3404 + 0,2943 \times 0,9997]^2 \\
 &= 0,2856
 \end{aligned}$$

Para $L > 2$ estratos, o procedimento é análogo.

TABELA 2.1

PONTOS ÓTIMOS E APROXIMADAMENTE ÓTIMOS DE ESTRATIFICAÇÃO DE $f(y) = e^{-y}$ PARA $L=2,3,4$ e 5 ESTRATOS E SUAS VARIÂNCIAS

NÚMERO DE ESTRATOS	PONTOS DE ESTRATIFICAÇÃO					CONST.	$n v(\bar{y}_{st})_N = \left[\sum_h \omega_h \sigma_h \right]^2$
	1 ^a APROX.	y ₁ '	y ₂ '	y ₃ '	y ₄ '	C _L '	
	2 ^a APROX. MIN. (1)	y ₁ "	y ₂ "	y ₃ "	y ₄ "	C _L "	
2	1 ^a	1,223	-	-	-		0,2856
	min.	1,262	-	-	-	0,8737	0,2855
3	1 ^a	0,742	2,045	-	-	0,3883	0,1334
	2 ^a	0,766	1,991	-	-	0,4095	0,1333
	min.	0,764	2,026	-	-	0,4071	0,1332
4	1 ^a	0,545	1,295	2,572	-	0,2292	0,0769
	2 ^a	0,553	1,317	2,547	-	0,2349	0,0769
	min.	0,551	1,315	2,577	-	0,2345	0,0768
5	1 ^a	0,430	0,977	1,730	2,995	0,1501	0,0500
	2 ^a	0,433	0,985	1,748	2,982	0,1522	0,0500
	min.	0,431	0,982	1,746	3,008	0,1521	0,0500

(1) Os valores mínimos foram retirados de Ekman (1959)

Observando a tabela acima notamos que a 1^a e a 2^a aproximações obtidas pelo método de Ekman apresentam variâncias praticamente iguais a aquelas obtidas pela solução de variância mínima.

CAPÍTULO 3

MÉTODO DE ESTRATIFICAÇÃO USANDO O TOTAL

3.1. Introdução

Esta regra, sugerida por Mahalanobis (1952) e por Hansen, Hurwitz e Madow (1953), é comumente usada, e consiste em formar estratos, usando a distribuição de y de tal maneira, que o valor agregado total da variável y é o mesmo para cada estrato, ou seja

$$\omega_h \mu_h = C = \text{constante} \quad (3.1)$$

Dalenius e Hodges (1957), observaram que esta regra não pode ser de validade geral porque os limites que ela fornece, variam se a escala de origem varia, considerando que a solução exata para o problema, é invariante sob mudança de origem. Assim, a relação desta regra com a equação (1.14) não está ainda muito clara.

Uma justificativa dada por Cochran (1961) é a de

que esta regra pode ser derivada da equação (1.22) na qual $(Z_h - Z_{h-1}) = \text{Const.}$ é o mesmo que fazer $\omega_h \sigma_h = \text{Const.}$ Neste caso, quando os limites próximos do ótimo, levam a coeficientes de variação μ_h / σ_h aproximadamente iguais em todos os estratos, a relação $\omega_h \sigma_h$ igual a uma constante faz com que $\omega_h \mu_h$ também seja igual a uma constante. Este resultado sugere que o procedimento aqui discutido deveria dar melhores soluções naquelas situações em que temos os coeficientes de variação aproximadamente constantes em todos os estratos.

A inclusão da regra de aproximação neste trabalho justifica-se por seu vasto uso em situações práticas.

3.2. Adaptação a Cálculos Numéricos

Da definição de $I_p(y) = \int_{-\infty}^y t^p f(t) dt$ nós temos

$$I_1(y_h) - I_1(y_{h-1}) = \int_{y_{h-1}}^{y_h} t f(t) dt \quad (3.2)$$

A condição $\omega_h \mu_h = C$ para $h=1, 2, \dots, L$ na determinação do conjunto $[y'_h]$ é equivalente a

$$\left. \begin{aligned} I_1(y'_L) - I_1(y'_{L-1}) &= C \\ I_1(y'_{L-1}) - I_1(y'_{L-2}) &= C \\ \dots & \\ I_1(y'_1) - I_1(y'_0) &= C \end{aligned} \right\} \quad (3.3)$$

$$\begin{aligned} \text{pois } \omega_h \mu_h &= \int_{y_{h-1}}^{y_h} t f(t) dt \\ &= I_1(y_h) - I_1(y_{h-1}) . \end{aligned}$$

Somando as L equações em (3.3), chegamos a

$$I_1(y'_L) - I_1(y'_0) = L C . \tag{3.4}$$

Como $I_1(y'_0) = 0$, nós temos

$$\frac{I_1(y'_L)}{L} = C \tag{3.5}$$

Considerando $I_1(y'_L) = H$, nós temos

$$C = \frac{H}{L} \tag{3.6}$$

Substituindo o valor de C acima, na equação (3.3) obtemos

$$\left. \begin{aligned} I_1(y'_{L-1}) &= \frac{(L-1)}{L} H \\ I_1(y'_{L-2}) &= \frac{(L-2)}{L} H \\ \dots &\dots \dots \dots \dots \dots \dots \dots \\ I_1(y'_1) &= \frac{1}{L} H \end{aligned} \right\} \tag{3.7}$$

do qual calculamos o conjunto $[y'_h]$.

3.3. Exemplo Numérico

Como ilustração, usaremos este método de aproximação dos limites ótimos de estratificação, na função densidade $f(y) = e^{-y}$, $y \geq 0$.

Primeiramente iremos determinar o valor de $I_1(y'_L)=H$, e a partir deste valor, o conjunto $[y'_h]$ usando equações(3.7).

Por definição

$$I_1(y'_L) = \int_0^{\infty} t e^{-t} dt = 1 = H .$$

3.3.1. Cálculos Numéricos para L=2 Estratos

Da equação (3.7) temos

$$I_1(y'_1) = \frac{1}{2}$$

Por outro lado

$$I_1(y'_1) = \int_0^{y'_1} t e^{-t} dt$$

Logo

$$1 - e^{-y'_1} - y'_1 e^{-y'_1} = 0,5$$

Usando aproximação de Newton, considere

$$g(y_1) = e^{-y_1} + y_1 e^{-y_1} - 0,5 = 0$$

$$\frac{\partial g}{\partial y_1} = -y_1 e^{-y_1}$$

$$y_1'' = y_1' + \frac{e^{-y_1'} + y_1' e^{-y_1'} - 0,5}{y_1' e^{-y_1'}}$$

Como as raízes estão entre 1,5 e 2 (pois $g(1,5) = 0,0578$ e $g(2) = -0,0940$), podemos escolher para aproximação inicial $y_1' = 1,5$, que determina

$$y_1'' = 1,5 + \frac{e^{-1,5} + 1,5 e^{-1,5} - 0,5}{1,5 e^{-1,5}} = 1,673$$

O valor de y_1'' será colocado no lugar de y_1' até que dois valores consecutivos sejam encontrados iguais, ou diferentes por quantidades negligenciáveis.

Então

$$y_1''' = 1,673 + \frac{e^{-1,673} + 1,673 e^{-1,673} - 0,5}{1,673 e^{-1,673}} = 1,678$$

$$y_1^{(4)} = 1,678 + \frac{e^{-1,678} + 1,678 e^{-1,678} - 0,5}{1,678 e^{-1,678}} = 1,678$$

Logo $(1 - e^{-y_1'} - y_1' e^{-y_1'}) = 0,5$ implica em $y_1' = 1,678$. Aproximando para duas casas decimais, podemos considerar $y_1' = 1,68$.

3.3.2. Cálculos Numéricos para L=3 Estratos

Da equação (3.7) temos

$$\left. \begin{aligned} I_1(y_1') &= \int_0^{y_1'} t e^{-t} dt = \frac{1}{3} \\ I_2(y_2') &= \int_0^{y_2'} t e^{-t} dt = \frac{2}{3} \end{aligned} \right\}$$

Desta forma

$$\left. \begin{aligned} 1 - e^{-y_1'} - y_1' e^{-y_1'} &= 0,3333 \\ 1 - e^{-y_2'} - y_2' e^{-y_2'} &= 0,6667 \end{aligned} \right\}$$

Para encontrar o valor de y_1' , considere

$$g(y_1) = e^{-y_1} + y_1 e^{-y_1} - 0,6667 = 0$$

$$\frac{\partial g}{\partial y_1} = -y_1 e^{-y_1}$$

$$y_1'' = y_1' + \frac{e^{-y_1'} + y_1' e^{-y_1'} - 0,6667}{y_1' e^{-y_1'}}$$

Como as raízes estão entre 1 e 1,5 (pois $g(1) = 0,0691$ e $g(1,5) = -0,1089$), podemos escolher para aproximação inicial $y_1' = 1$, que determina

$$y_1'' = 1 + \frac{e^{-1} + e^{-1} - 0,6667}{e^{-1}} = 1,1877$$

O valor de y_1'' será colocado no lugar de y_1' até que dois valores consecutivos sejam encontrados iguais, ou diferentes por quantidades negligenciáveis. Então

$$y_1''' = 1,1877 + \frac{e^{-1,1877} + 1,1877 e^{-1,1877} - 0,6667}{1,1877 e^{-1,1877}} = 1,1887$$

$$y_1^{(4)} = 1,1887 + \frac{e^{-1,1887} + 1,1887 e^{-1,1887} - 0,6667}{1,1887 e^{-1,1887}} = 1,1887$$

Logo $1 - e^{-y_1'} - y_1' e^{-y_1'} = 0,3333$ implica em $y_1' = 1,1887$ que aproximado para duas casas decimais, torna-se igual a 1,19.

Procedendo de modo análogo para $g(y_2) = e^{-y_2} + y_2 e^{-y_2} - 0,3333 = 0$ nós encontramos $y_2' = 2,30$.

3.3.3. Cálculos Numéricos para L>3 Estratos

Procedimentos análogos ao item anterior, são usados para determinar o conjunto $[y'_h]$ para um número L>3 estratos. A tabela 3.1 apresenta os resultados para L=2,3,4 e 5 estratos, e suas variâncias.

3.4. Cálculo de Variâncias

Para o cálculo de $n V(\bar{y}_{st}) = \left[\sum_h \omega_h \sigma_h \right]^2$ precisamos calcular o valor de ω_h e σ_h , proporção e desvio padrão para $h=1,2,\dots,L$.

Para L=2 estratos, temos $y'_0 = 0$, $y'_1 = 1,68$ e $y'_2 = \infty$.

Usando as definições $I_p(y)$ e J_{ph} , equações (1.24) e (1.27) dadas no Capítulo 1, nós podemos escrever

$$\omega'_1 = \int_0^{1,68} e^{-t} dt = I_0(1,68) = J'_{01} = 0,8136$$

$$\omega'_2 = \int_{1,68}^{\infty} e^{-t} dt = 1 - I_0(1,68) = J'_{02} = 0,1864$$

Das equações (1.28) e (1.29) temos

$$\mu_h = \frac{J_{1h}}{J_{0h}} \quad e \quad \sigma_h^2 = \frac{J_{0h} J_{2h} - J_{1h}^2}{J_{0h}^2}$$

Então

$$\mu'_1 = \frac{J'_{11}}{J'_{01}} = \frac{0,5005}{0,8136} = 0,6152$$

pois

$$J'_{11} = I_1(y'_1) - I_1(y'_0) = I_1(y'_1) = \int_0^{1,68} t e^{-t} dt = 0,5005$$

$$e \mu'_2 = \frac{J'_{12}}{J'_{02}} = \frac{0,4995}{0,1864} = 2,6797$$

$$\text{pois } J'_{12} = I_1(y'_2) - I_1(y'_1) = 1 - 0,5005 = 0,4995$$

Por outro lado

$$\begin{aligned} \sigma'_1 &= \sqrt{\frac{J'_{01} J'_{21} - J'^2_{11}}{J'^2_{01}}} \\ &= \sqrt{\frac{0,8136 \times 0,4750 - (0,5005)^2}{(0,8136)^2}} = 0,4532 \end{aligned}$$

onde

$$J'_{21} = I_2(y'_1) - I_2(y'_0) = I_2(y'_1) = \int_0^{1,68} t^2 e^{-t} dt = 0,4750$$

E

$$\begin{aligned} \sigma'_2 &= \sqrt{\frac{J'_{02} J'_{22} - J'^2_{12}}{J'^2_{02}}} \\ &= \sqrt{\frac{0,1864 \times 1,5250 - (0,4995)^2}{(0,1864)^2}} = 1,0002 \end{aligned}$$

onde

$$\begin{aligned} J'_{22} &= I_2(y'_2) - I_2(y'_1) \\ &= 2 - 0,4750 = 1,5250 \end{aligned}$$

Então, para L=2 estratos

$$n V(\bar{y}_{st}) = \left[\sum_{h=1}^2 \omega_h \sigma_h \right]^2$$

$$= (0,8136 \times 0,4532 + 0,1864 \times 1,0002)^2 = 0,3082$$

Para um número de estratos maior que dois, o procedimento é análogo.

Na tabela abaixo, apresentamos um resumo dos resultados obtidos para L=2,3,4 e 5 estratos para a função $f(y)=e^{-y}$.

TABELA 3.1

PONTOS DE ESTRATIFICAÇÃO DE $f(y) = e^{-y}$ PARA L = 2, 3, 4 e 5
ESTRATOS E SUAS VARIÂNCIAS

NÚMERO DE ESTRATOS	PONTOS DE ESTRATIFICAÇÃO				n V(\bar{y}_{st}) = $[\sum \omega_h \sigma_h]^2$
	y' ₁	y' ₂	y' ₃	y' ₄	
2	1,68	-	-	-	0,3082
3	1,19	2,30	-	-	0,1556
4	0,96	1,68	2,69	-	0,0950
5	0,82	1,38	2,02	2,99	0,0639

Comparando as variâncias obtidas na tabela acima com as variâncias mínimas (tabela 2.1) iguais a 0,2855, 0,1332, 0,0768 e 0,0500 para L = 2, 3, 4 e 5 estratos respectivamente, observamos que os limites de estratificação obtidos deste método não estão muito próximos do ótimo, o que nos leva a concluir que para a função $f(y) = e^{-y}$, dos três métodos apresentados, este é o que apresenta resultados mais "pobres".

CAPÍTULO 4

ESTRATIFICAÇÃO APROXIMADAMENTE ÓTIMA COM VARIÁVEL AUXILIAR

4.1. Introdução

Nos capítulos anteriores nós vimos métodos aproximados de estratificação ótima quando a variável de estratificação é considerada como a própria variável de estimação. Do ponto de vista matemático esse procedimento é perfeitamente correto. No entanto, do ponto de vista prático, não há condição de que se possa considerar a variável de estimação e estratificação como as mesmas. O que acontece geralmente é que a variável de estimação y , é raramente conhecida, e é desejável estratificar com base em alguma variável x convenientemente escolhida (por exemplo, o valor da variável de estimação em um censo recente).

Neste caso, a população em estudo é dividida em L estratos e uma amostra aleatória simples estratificada de tamanho n é retirada, o tamanho da amostra no h -ésimo estrato

sendo n_h e tal que $\sum_h n_h = n$. Se y é a variável em estudo, um estimador não viesado da média populacional é dado por $\bar{y}_{st} = \sum_h \omega_h \bar{y}_h$ onde ω_h é a proporção de unidades no h -ésimo estrato e \bar{y}_h é a média da amostra no estrato h .

Ignorando as frações de correção para populações finitas, a variância da estimativa \bar{y}_{st} , sob alocação de Neyman é dada por $V(\bar{y}_{st})_N = \left[\sum_h \omega_h \sigma_{hy} \right]^2 / n$, enquanto que sob alocação proporcional, é dada por $V(\bar{y}_{st})_p = \sum_h \omega_h \sigma_{hy}^2 / n$.

Nós iremos considerar o problema da estratificação ótima na variável auxiliar x . Supondo conhecidas a função de dependência de y em x , e a função $V(y/x)$, determinaremos equações mínimas fornecendo limites ótimos de estratificação para alocações de Neyman e proporcional.

Como estas equações não podem ser resolvidas facilmente, apresentaremos no item 4.4, vários métodos desenvolvidos por Singh e Sukhatme (1969) para encontrar soluções aproximadas para a solução exata.

No item 4.5 estudaremos o caso particular da regressão linear na estratificação e no final do capítulo apresentaremos um exemplo numérico para a função densidade $f(x) = e^{-x+1}$, $x \geq 1$ e admitindo regressão linear de y em x .

4.2. Limite Ótimo dos Estratos

Suponha y como variável de estimação e seja

$$y = c(x) + e \quad (4.1)$$

onde $c(x)$ é uma função de x e tal que $E(e|x) = 0$ e $V(e|x) = \psi(x) > 0$ para todo x no intervalo (a,b) com $(b-a) < \infty$. Seja $f(x,y)$ a densidade conjunta de (x,y) e $f(x)$ a densidade marginal de x . Então nós temos

$$\left. \begin{aligned} \omega_h &= \int_{x_{h-1}}^{x_h} f(x) dx \\ \mu_{hy} = \mu_{hc} &= \frac{1}{\omega_h} \int_{x_{h-1}}^{x_h} c(x)f(x) dx \\ \sigma_{hy}^2 &= \sigma_{hc}^2 + \mu_h \psi \end{aligned} \right\} \quad (4.2)$$

onde (x_{h-1}, x_h) são os limites do h -ésimo estrato, $\mu_h \psi$ é o valor esperado de $\psi(x)$ e σ_{hc}^2 é a variância de $c(x)$ no h -ésimo estrato.

Usando as relações acima, a expressão da variância para alocação de Neyman se reduz a

$$\begin{aligned} V(\bar{y}_{st})_N &= \left[\sum_h \omega_h \sigma_{hy} \right]^2 / n \\ &= \left[\sum_h \omega_h \sqrt{\sigma_{hc}^2 + \mu_h \psi} \right]^2 / n \end{aligned} \quad (4.3)$$

e para alocação proporcional

$$\begin{aligned} V(\bar{y}_{st})_p &= \left[\sum_h \omega_h \sigma_{hy}^2 \right] / n \\ &= \left[\sum_h \omega_h (\sigma_{hc}^2 + \mu_h \psi) \right] / n \end{aligned} \quad (4.4)$$

Seja $[x_h]$ o conjunto de pontos ótimos de estratificação na amplitude (a,b) de x , para o qual a $V(\bar{y}_{st})$ é mínima. Estes pontos $[x_h]$ são as soluções das equações minimais que são obtidas quando igualamos a zero as derivadas parciais de $V(\bar{y}_{st})$ com respeito a $[x_h]$. Nós obteremos estas equações para alocações de Neyman e proporcional.

4.2.1. Limites Ótimos de Estratificação para Alocação de Neyman

Nós desejamos minimizar $V(\bar{y}_{st})_N = \left[\sum_h \omega_h \sqrt{\sigma_{hc}^2 + \mu_h \varphi} \right]^2 / n$, o que é equivalente a minimizar

$$\sum_h \omega_h \sqrt{\sigma_{hc}^2 + \mu_h \varphi} \quad (4.5)$$

Igualando a zero as derivadas parciais da equação (4.5) com relação a x_h , nós chegamos a

$$\omega_h \frac{\partial}{\partial x_h} \sqrt{(h)} + \sqrt{(h)} \frac{\partial \omega_h}{\partial x_h} + \omega_i \frac{\partial}{\partial x_h} \sqrt{(i)} + \sqrt{(i)} \frac{\partial \omega_i}{\partial x_h} = 0 \quad (4.6)$$

onde $i=(h+1)$, $(h) = \sigma_{hc}^2 + \mu_h \varphi$ e $(i) = \sigma_{ic}^2 + \mu_i \varphi$

$$\frac{\partial \omega_h}{\partial x_h} = f(x_h) = - \frac{\partial \omega_i}{\partial x_h} \quad (4.7)$$

$$\begin{aligned} \frac{\partial (h)}{\partial x_h} &= \frac{1}{\omega_h} c^2(x_h) f(x_h) - \frac{f(x_h)}{\omega_h^2} \int_{x_{h-1}}^{x_h} c^2(x) f(x) dx - 2\mu_{hc} \left\{ \frac{1}{\omega_h} [c(x_h) f(x_h)] \right. \\ &\quad \left. - \frac{f(x_h)}{\omega_h^2} \int_{x_{h-1}}^{x_h} c(x) f(x) dx \right\} + \frac{1}{\omega_h} \varphi(x_h) f(x_h) - \frac{f(x_h)}{\omega_h^2} \int_{x_{h-1}}^{x_h} \varphi(x) f(x) dx \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{\omega_h} c^2(x_h) f(x_h) - \frac{f(x_h)}{\omega_h} (\sigma_{hc}^2 + \mu_{hc}^2) - \frac{2\mu_{hc}}{\omega_h} c(x_h) f(x_h) \\
 &+ \frac{2\mu_{hc}^2}{\omega_h} f(x_h) + \frac{1}{\omega_h} \varphi(x_h) f(x_h) - \frac{f(x_h)}{\omega_h} \mu_h \varphi \\
 &= \frac{f(x_h)}{\omega_h} \left[c^2(x_h) - \sigma_{hc}^2 - \mu_{hc}^2 - 2\mu_{hc} c(x_h) + 2\mu_{hc}^2 + \varphi(x_h) - \mu_h \varphi \right] \\
 &= \frac{f(x_h)}{\omega_h} \left\{ [c(x_h) - \mu_{hc}]^2 - \sigma_{hc}^2 + \varphi(x_h) - \mu_h \varphi \right\} \quad (4.8)
 \end{aligned}$$

Procedendo de maneira análoga, temos

$$\frac{\partial(i)}{\partial x_h} = - \frac{f(x_h)}{\omega_i} \left\{ [c(x_h) - \mu_{ic}]^2 - \sigma_{ic}^2 + \varphi(x_h) - \mu_i \varphi \right\} \quad (4.9)$$

Usando os resultados acima na equação (4.6) obtemos

$$\begin{aligned}
 &\frac{\omega_h}{2 \sqrt{\sigma_{hc}^2 + \mu_h \varphi}} \cdot \frac{f(x_h)}{\omega_h} \left\{ [c(x_h) - \mu_{hc}]^2 - \sigma_{hc}^2 + \varphi(x_h) - \mu_h \varphi \right\} + f(x_h) \sqrt{\sigma_{hc}^2 + \mu_h \varphi} \\
 &- \frac{\omega_i}{2 \sqrt{\sigma_{ic}^2 + \mu_i \varphi}} \cdot \frac{f(x_h)}{\omega_i} \left\{ [c(x_h) - \mu_{ic}]^2 - \sigma_{ic}^2 + \varphi(x_h) - \mu_i \varphi \right\} \\
 &- f(x_h) \sqrt{\sigma_{ic}^2 + \mu_i \varphi} = 0 \\
 &\frac{[c(x_h) - \mu_{hc}]^2 + \sigma_{hc}^2 + \varphi(x_h) + \mu_h \varphi}{\sqrt{\sigma_{hc}^2 + \mu_h \varphi}} = \frac{[c(x_h) - \mu_{ic}]^2 + \sigma_{ic}^2 + \varphi(x_h) + \mu_i \varphi}{\sqrt{\sigma_{ic}^2 + \mu_i \varphi}} \quad (4.10)
 \end{aligned}$$

$$i = h+1, \quad h = 1, 2, \dots, L-1.$$

O conjunto $[x_h]$ satisfazendo equação (4.10) é o conjunto de pontos ótimos de estratificação que minimizam $V(\bar{y}_{st})_N$.

4.2.2. Limites Ótimos de Estratificação para Alocação Proporcional

Com alocação proporcional, nós desejamos minimizar

$$\begin{aligned} V(\bar{y}_{st})_p &= \left[\sum_h \omega_h (\sigma_{hc}^2 + \mu_{h\varphi}) \right] / n \\ &= \left[\sum_h \omega_h \sigma_{hc}^2 + \sum_h \omega_h \mu_{h\varphi} \right] / n \\ &= \left[\sum_h \omega_h \sigma_{hc}^2 + \mu \varphi \right] / n \end{aligned} \quad (4.11)$$

onde $\mu_{\varphi} = \sum_h \omega_h \mu_{h\varphi}$ é um parâmetro populacional e portanto constante. Assim, minimizar esta variância é equivalente a minimizar $\sum_h \omega_h \sigma_{hc}^2$.

Igualando a zero as derivadas parciais de $\sum_h \omega_h \sigma_{hc}^2$ com respeito a x_h , temos

$$\omega_h \frac{\partial}{\partial x_h} \sigma_{hc}^2 + \sigma_{hc}^2 \frac{\partial \omega_h}{\partial x_h} + \omega_i \frac{\partial}{\partial x_h} \sigma_{ic}^2 + \sigma_{ic}^2 \frac{\partial \omega_i}{\partial x_h} = 0 \quad (4.12)$$

$$\begin{aligned} \frac{\partial \sigma_{hc}^2}{\partial x_h} &= \frac{1}{\omega_h} c^2(x_h) f(x_h) - \frac{f(x_h)}{\omega_h} (\sigma_{hc}^2 + \mu_{hc}^2) - \frac{2\mu_{hc}}{\omega_h} c(x_h) f(x_h) \\ &+ \frac{2\mu_{hc}^2}{\omega_h} f(x_h) = \frac{f(x_h)}{\omega_h} \left\{ [c(x_h) - \mu_{hc}]^2 - \sigma_{hc}^2 \right\} \end{aligned} \quad (4.13)$$

De maneira análoga

$$\frac{\partial}{\partial x_h} \sigma_{ic}^2 = - \frac{f(x_h)}{\omega_i} \left\{ \left[c(x_h) - \mu_{ic} \right]^2 - \sigma_{ic}^2 \right\} \quad (4.14)$$

Substituindo as equações acima em (4.12) temos

$$\omega_h \frac{f(x_h)}{\omega_h} \left\{ \left[c(x_h) - \mu_{hc} \right]^2 - \sigma_{hc}^2 \right\} + f(x_h) \sigma_{hc}^2 - \omega_i \frac{f(x_h)}{\omega_i} \left\{ \left[c(x_h) - \mu_{ic} \right]^2 - \sigma_{ic}^2 \right\} + f(x_h) \sigma_{ic}^2 = 0$$

ou seja

$$c(x_h) - 2c(x_h)\mu_{hc} + \mu_{hc}^2 = c(x_h) - 2c(x_h)\mu_{ic} + \mu_{ic}^2$$

Então

$$-2c(x_h) \left[\mu_{hc} - \mu_{ic} \right] + (\mu_{hc} - \mu_{ic})(\mu_{hc} + \mu_{ic}) = 0$$

$$(\mu_{hc} - \mu_{ic}) \left[\mu_{hc} + \mu_{ic} - 2c(x_h) \right] = 0$$

$$c(x_h) = \frac{\mu_{hc} + \mu_{ic}}{2} \quad (4.15)$$

Nós observamos que ambos os sistemas (4.10) e (4.15) são funções de parâmetros populacionais que por sua vez são funções de soluções destas equações. Devido a essa dificuldade não é possível determinar soluções exatas. Nós encontraremos soluções aproximadas. Para facilidade de cálculo iremos encontrar a seguir, expressões aproximadas para a média e variância condicional em intervalos pequenos.

4.3. Expressões Aproximadas para a Média e Variância Condicional

Suponha que as funções $f(x)$, $\varphi(x)$ e $c(x)$ são limitadas e possuem as duas primeiras derivadas contínuas para todo x em (a,b) . Considere as seguintes identidades devidas a Ekman (1959).

$$\begin{aligned} I_i(y,x) &= \int_y^x (t-y)^i f(t) dt \\ &= \sum_{j=0}^3 \frac{K^{i+j+1}}{j!(i+j+1)} f^{(j)}(y) + o(K^{i+5}) \end{aligned} \quad (4.16)$$

onde $f^{(j)}(y)$ é a j -ésima derivada de $f(t)$ em $t=y$ e $K=x-y$.

$$\begin{aligned} I_i(y,x) &= \int_x^y (t-x)^i f(t) dt \\ &= \sum_{j=0}^3 \frac{(-K)^{i+j+1}}{j!(i+j+1)} f^{(j)}(y) + o(K^{i+5}) \end{aligned} \quad (4.17)$$

Seja $\mu_{\varphi}(y,x)$ a esperança condicional de $\varphi(t)$ no intervalo (y,x) tal que

$$\mu_{\varphi}(y,x) = \frac{\int_y^x \varphi(t) f(t) dt}{\int_y^x f(t) dt} \quad (4.18)$$

Do teorema de Taylor nós temos

$$\begin{aligned} \varphi(t) &= \varphi(y) + (t-y)\varphi'(y) + \frac{(t-y)^2}{2!}\varphi''(y) + \frac{(t-y)^3}{3!}\varphi'''(y) + o[(t-y)^4] \\ &= \sum_{j=0}^3 \frac{(t-y)^j}{j!}\varphi^{(j)}(y) + o[(t-y)^4] \end{aligned} \quad (4.19)$$

onde $\varphi^{(j)}(y)$ é a j -ésima derivada de $\varphi(t)$ em $t=y$. Então

$$\begin{aligned} \mu_{\varphi}(y,x) \int_y^x f(t)dt &= \mu_{\varphi}(y,x) I_0(y,x) \\ &= \int_y^x \varphi(t) f(t) dt \\ &= \int_y^x \left[\sum_{j=0}^3 \frac{(t-y)^j}{j!} \varphi^{(j)}(y) + o(t-y)^4 \right] f(t) dt \end{aligned} \quad (4.20)$$

mas

$$\begin{aligned} &\int_y^x \left[\sum_{j=0}^3 \frac{(t-y)^j}{j!} \varphi^{(j)}(y) + o(t-y)^4 \right] f(t) dt = \\ &= \int_y^x \varphi(y) f(t) dt + \int_y^x (t-y) \varphi'(y) f(t) dt + \int_y^x \frac{(t-y)^2}{2!} \varphi''(y) f(t) dt \\ &\quad + \int_y^x \frac{(t-y)^3}{3!} \varphi'''(y) f(t) dt + o(K^3) \\ &= \varphi(y) I_0(y,x) + \varphi'(y) I_1(y,x) + \frac{\varphi''(y)}{2!} I_2(y,x) + \frac{\varphi'''(y)}{3!} I_3(y,x) + o(K^4) \\ &= \sum_{j=0}^3 \frac{\varphi^{(j)}(y)}{j!} I_j(y,x) + o(K^4) \end{aligned} \quad (4.21)$$

Então

$$\mu_{\varphi}(y,x) = \sum_{j=0}^3 \frac{\varphi^{(j)}(y)}{j!} \frac{I_j(y,x)}{I_0(y,x)} + o(K^4) \quad (4.22)$$

onde $\varphi^{(j)}(y)$ é a j -ésima derivada de $\varphi(t)$ em $t=y$. Usando as

equações (4.16) para determinar $I_j(y,x)$ temos

$$\left. \begin{aligned} I_0(y,x) &= f(y)K + \frac{f'(y)}{2!} K^2 + \frac{f''(y)}{3!} K^3 + \frac{f'''(y)}{4!} K^4 + o(K^5) \\ I_1(y,x) &= \frac{f(y)}{2} K^2 + \frac{f'(y)}{3} K^3 + \frac{f''(y)}{8} K^4 + o(K^5) \\ I_2(y,x) &= \frac{f(y)}{3} K^3 + \frac{f'(y)}{4} K^4 + o(K^5) \\ I_3(y,x) &= \frac{f(y)}{4} K^4 + o(K^5) \end{aligned} \right\} \quad (4.23)$$

Considere

$$H_j(y,x) = \frac{I_j(y,x)}{I_0(y,x)}, \quad j = 1, 2, 3. \quad (4.24)$$

Para determinar o valor de $H_1(y,x)$, considere o produto

$$\begin{aligned} I_0 [A + BK + CK^2 + DK^3 + o(K^4)] &= \\ = A f K + \left(\frac{Af'}{2} + Bf \right) K^2 + \left(\frac{Af''}{6} + \frac{Bf'}{2} + Cf \right) K^3 \\ + \left(\frac{Af'''}{24} + \frac{Bf''}{6} + \frac{Cf'}{2} + Df \right) K^4 + o(K^5) \end{aligned} \quad (4.25)$$

Igualando o lado direito da equação (4.25) a $I_1(y,x)$ temos

$$A = 0, \quad B = \frac{1}{2}, \quad C = \frac{f'}{12f} \quad \text{e} \quad D = \frac{f''f - f'^2}{24f^2}$$

Logo

$$H_1(y,x) = \frac{1}{2} K + \frac{f'}{12f} K^2 + \frac{f''f - f'^2}{24f^2} K^3 + o(K^4) \quad (4.26)$$

De maneira análoga

$$H_2(y,x) = \frac{1}{3} K^2 + \frac{f'}{12f} K^3 + o(K^4) \quad (4.27)$$

$$H_3(y,x) = \frac{1}{4} K^3 + o(K^4) \quad (4.28)$$

Substituindo os resultados acima na equação (4.22) obtemos

$$\mu_{\varphi}(y,x) = \varphi \left\{ 1 + \frac{\varphi'}{2\varphi} K + \left[\frac{\varphi'f' + 2f\varphi''}{12f\varphi} \right] K^2 + \left[\frac{ff''\varphi' + ff'\varphi'' + f^2\varphi''' - f'^2\varphi'}{24f^2\varphi} \right] K^3 + o(K^4) \right\} \quad (4.29)$$

onde as funções φ , f e suas derivadas são calculadas no ponto $t = y$.

Usando a expressão (4.17) para os valores de $I_j(y,x)$, $j = 0, 1, 2, 3$ e procedendo de maneira análoga, temos

$$\mu_{\varphi}(y,x) = \varphi \left\{ 1 - \frac{\varphi'}{2\varphi} K + \left[\frac{\varphi'f' + 2f\varphi''}{12f\varphi} \right] K^2 - \left[\frac{ff''\varphi' + ff'\varphi'' + f^2\varphi''' - f'^2\varphi'}{24f^2\varphi} \right] K^3 + o(K^4) \right\} \quad (4.30)$$

onde as funções φ , f e suas derivadas são calculadas em $t=x$.

Seja $\sigma_{\varphi}^2(y,x)$ a variância condicional de $\varphi(t)$ no intervalo (y,x) . Então

$$\sigma_{\varphi}^2(y,x) = \mu_{\varphi^2}(y,x) - [\mu_{\varphi}(y,x)]^2 \quad (4.31)$$

mas

$$\mu_{\varphi^2}(y,x) = \varphi^2 \left\{ 1 + \frac{\varphi'}{\varphi} K + \left[\frac{2\varphi\varphi'f' + 4\varphi'^2f + 4\varphi\varphi''f}{12f\varphi^2} \right] K^2 + \left[\frac{\varphi\varphi'ff'' + \varphi'^2ff' + \varphi\varphi''ff' + 3\varphi'\varphi''f^2 + \varphi\varphi''''f^2 - \varphi\varphi'f'^2}{12f^2\varphi^2} \right] K^3 + o(K^4) \right\}$$

$$\begin{aligned} [\mu_{\varphi}(y,x)]^2 = & \varphi^2 \left\{ 1 + \frac{\varphi'}{\varphi} K + \left[\frac{\varphi'^2}{4\varphi^2} + \frac{2\varphi'f' + 4f\varphi''}{12f\varphi} \right] K^2 \right. \\ & \left. + \left[\frac{\varphi'^2 f' + 2\varphi'\varphi'''f}{12f\varphi^2} + \frac{\varphi'ff'' + \varphi''ff' + \varphi'''f^2 - \varphi'f'^2}{12f^2\varphi} \right] K^3 + o(K^4) \right\} \end{aligned}$$

Então

$$\begin{aligned} \sigma_{\varphi}^2(y,x) &= \frac{\varphi'^2}{12} K^2 + \frac{\varphi'\varphi''}{12} K^3 + o(K^4) \\ &= \frac{K^2\varphi'^2}{12} \left[1 + \frac{\varphi''}{\varphi'} K + o(K^2) \right] \end{aligned} \quad (4.32)$$

onde a função φ e suas derivadas são calculadas no ponto $t=y$.

Usando os resultados acima, outras aproximações podem ser obtidas. Multiplicando as expansões das séries para $\mu_{\varphi}(y,x)$ nos pontos $t=y$ e $t=x$ e tomando a raiz quadrada nós obtemos

$$\mu_{\varphi}(y,x) = \sqrt{\varphi(y)\varphi(x)} [1 + o(K^2)] \quad (4.33)$$

Outro resultado de interêsse pode ser obtido de

$$I_o(y,x) \mu_{\varphi}(y,x) = \int_y^x \varphi(t)f(t) dt$$

quando tomamos $\varphi(t) = t^2$ e usando a equação (4.33)

$$I_o(y,x) xy [1 + o(K^2)] = \int_y^x t^2 f(t) dt$$

$$I_o(y,x) = \frac{1}{xy} \int_y^x t^2 f(t) dt [1 + o(K^2)] \quad (4.34)$$

Temos ainda o seguinte resultado, quando expandimos $\sqrt[\lambda]{f(t)}$ no ponto $t=y$

$$\begin{aligned} \left[\int_y^x \sqrt[\lambda]{f(t)} dt \right]^\lambda &= \left\{ \int_y^x \left[f(y)^{1/\lambda} + \frac{(t-y)}{\lambda} \frac{f'(y)}{f(y)^{1-1/\lambda}} + o(t-y)^2 \right] dt \right\}^\lambda \\ &= \left[K f(y)^{1/\lambda} + \frac{K^2}{2\lambda} \frac{f'(y)}{f(y)^{1-1/\lambda}} + o(K^3) \right]^\lambda \\ &= K^\lambda f(y) \left[1 + \frac{K}{2\lambda} \frac{f'(y)}{f(y)} + o(K^2) \right]^\lambda \\ &= K^\lambda f(y) \left\{ \sum_{t=0}^{\infty} \binom{\lambda}{t} \left[1 + \frac{K}{2\lambda} \frac{f'(y)}{f(y)} \right]^{\lambda-t} [o(K^2)]^t \right\} \\ &= K^\lambda f(y) \left[\left(1 + \frac{K}{2\lambda} \frac{f'(y)}{f(y)} \right)^\lambda + o(K^2) \right] \\ &= K^\lambda f(y) \left[\sum_{t=0}^{\infty} \binom{\lambda}{t} 1^{\lambda-t} \left(\frac{K}{2\lambda} \frac{f'(y)}{f(y)} \right)^t + o(K^2) \right] \\ &= K^\lambda f(y) \left[1 + \frac{\lambda K}{2\lambda} \frac{f'(y)}{f(y)} + o(K^2) \right] \\ &= K^{\lambda-1} \left[f(y)K + \frac{K^2}{2} f'(y) + o(K^2) \right] \\ &= K^{\lambda-1} \int_y^x f(t) dt [1 + o(K^2)] \quad (4.35) \end{aligned}$$

4.4. Soluções Aproximadas dos Pontos Ótimos de Estratificação

Para encontrar soluções aproximadas para a equação (4.10) nós desenvolvemos a expansão da série deste sistema de equações no ponto x_h por ser o limite comum dos estratos h e $h+1$. Resultados obtidos na seção anterior serão usados na expansão dos dois lados da equação minimal, onde (y, x) será substituído por (x_h, x_{h+1}) para a expansão do lado direito e por (x_{h-1}, x_h) para o lado esquerdo.

Primeiramente iremos desenvolver o lado direito da equação (4.10).

Seja $K_i = x_{h+1} - x_h$. Usando a equação (4.29) nós temos

$$\mu_{ic} = c \left\{ 1 + \frac{c'}{2c} K_i + \left[\frac{c'f' + 2fc''}{12fc} \right] K_i^2 + \left[\frac{ff''c + ff'c'' + f^2c''' - c'f'^2}{24f^2c} \right] K_i^3 + o(K_i^4) \right\}$$

$$\mu_{ic}^2 = c^2 \left\{ 1 + \frac{c'}{c} K_i + \left[\frac{3fc'^2 + 2cc'f' + 4fcc''}{12fc^2} \right] K_i^2 \right.$$

$$\left. + \left[\frac{c'^2ff' + 2f^2c'c'' + ff''cc' + ff'cc'' + f^2cc''' - cc'f'^2}{12f^2c^2} \right] K_i^3 + o(K_i^4) \right\}$$

$$[c(x_h) - \mu_{ic}]^2 = c^2 + \mu_{ic}^2 - 2c\mu_{ic}$$

$$= \left[\frac{3fc'^2 + 2cc'f' + 4fcc''}{12f} - \frac{2cc'f' + 4fcc''}{12f} \right] K_i^2$$

$$\begin{aligned}
 & + \left[\frac{c'^2 ff' + 2f^2 c' c'' + ff'' cc' + ff' cc'' + f^2 cc''' - cc' f'^2}{12f^2} \right. \\
 & \quad \left. - \frac{ff'' cc' + ff' cc'' + f^2 cc''' - cc' f'^2}{12f^2} \right] K_i^3 + o(K_i^4) \\
 & = \frac{c'^2}{4} K_i^2 + \left[\frac{c'^2 f' + 2fc' c''}{12f} \right] K_i^3 + o(K_i^4) \\
 & = \frac{K_i^2}{4} \left\{ c'^2 + \left[\frac{c'^2 f' + 2fc' c''}{3f} \right] K_i + o(K_i^2) \right\} \quad (4.36)
 \end{aligned}$$

$$\begin{aligned}
 \varphi(x_h) + \mu_i \varphi & = \varphi \left\{ 2 + \frac{\varphi'}{2\varphi} K_i + \left[\frac{\varphi' f' + 2f\varphi''}{12f\varphi} \right] K_i^2 \right. \\
 & \quad \left. + \left[\frac{ff'\varphi' + ff'\varphi'' + f^2\varphi''' - f'^2\varphi'}{24f^2\varphi} \right] K_i^3 + o(K_i^4) \right\} \quad (4.37)
 \end{aligned}$$

$$\begin{aligned}
 \sigma_{ic}^2 & = \frac{K_i^2 c'^2}{12} \left[1 + \frac{c''}{c'} K_i + o(K_i^2) \right] \\
 & = \frac{K_i^2}{12} \left[c'^2 + c' c'' K_i + o(K_i^2) \right] \quad (4.38)
 \end{aligned}$$

Então $\left\{ [c(x_h) - \mu_{ic}]^2 + \sigma_{ic}^2 + \varphi(x_h) + \mu_i \varphi \right\}$ pode ser escrito como

$$\begin{aligned}
 2\varphi + \frac{\varphi\varphi'}{2\varphi} K_i + \left[\frac{\varphi\varphi' f' + 2f\varphi\varphi'' + 4c'^2 f\varphi}{12f\varphi} \right] K_i^2 \\
 + \left[\frac{ff''\varphi\varphi' + ff'\varphi\varphi'' + f^2\varphi\varphi''' - f'^2\varphi\varphi'}{24f^2\varphi} \right] K_i^3 + o(K_i^4)
 \end{aligned}$$

$$\begin{aligned}
 & + \left. \left[\frac{c'^2 \varphi f' + 2\varphi f c' c''}{12f\varphi} + \frac{c' c'' \varphi}{12\varphi} \right] K_i^3 + o(K_i^4) \right\} \\
 = & 2\varphi \left\{ 1 + \frac{\varphi'}{4\varphi} K_i + \left[\frac{\varphi' f' + 2f\varphi'' + 4c'^2 f}{24f\varphi} \right] K_i^2 \right. \\
 & \left. + \left[\frac{2c' f f' + 6f^2 c' c'' + f f' \varphi'' + f^2 \varphi''' - f'^2 \varphi'}{48f^2 \varphi} \right] K_i^3 + o(K_i^4) \right\} \quad (4.39)
 \end{aligned}$$

Por outro lado,

$$\begin{aligned}
 \mu_i \varphi + \sigma_{ic}^2 = & \varphi \left\{ 1 + \frac{\varphi'}{2\varphi} K_i + \left[\frac{\varphi' f' + 2f\varphi'' + f c'^2}{12f\varphi} \right] K_i^2 \right. \\
 & \left. + \left[\frac{f f'' \varphi' + f f' \varphi'' + f^2 \varphi''' - \varphi' f'^2 + 2f^2 c' c''}{24f^2 \varphi} \right] K_i^3 + o(K_i^4) \right\} \quad (4.40)
 \end{aligned}$$

onde as funções φ , f , c e suas derivadas são calculadas no ponto $t=x_h$. Usando as relações (4.39) e (4.40), nós obtemos em simplificação

$$\frac{[c(x_h) - \mu_{ic}]^2 + \sigma_{ic}^2 + \varphi(x_h) + \mu_i \varphi}{\sqrt{\sigma_{ic}^2 + \mu_i \varphi}} = 2 \sqrt{\varphi} \left[1 + B_2 K_i^2 + B_3 K_i^3 + o(K_i^4) \right] \quad (4.41)$$

onde

$$\begin{aligned}
 B_2 = & \frac{4\varphi c'^2 + \varphi'^2}{32\varphi^2} \\
 B_3 = & \frac{1}{96f\sqrt{\varphi'}} \frac{d}{dx_h} \left(\frac{4f\varphi c'^2 + f\varphi'^2}{\varphi^{3/2}} \right) \\
 = & \frac{8f'\varphi^2 c'^2 + 16f\varphi^2 c' c'' + 2\varphi f' c'^2 + 4f\varphi\varphi'\varphi'' - 4f\varphi\varphi' c'^2 - 3f\varphi'^3}{192f\varphi^3} \quad (4.42)
 \end{aligned}$$

Procedendo de modo análogo para o lado esquerdo da equação (4.10) chegamos a

$$\frac{[c(x_h) - \mu_{hc}]^2 + \sigma_{hc}^2 + \varphi(x_h) + \mu_h \varphi}{\sqrt{\sigma_{hc}^2 + \mu_h \varphi}} = 2\sqrt{\varphi} [1 + B_2 K_h^2 - B_3 K_h^3 + O(K_h^4)] \quad (4.43)$$

onde $K_h = x_h - x_{h-1}$. O sistema de equação (4.10) pode então ser reescrito da forma

$$K_h^2 [B_2 - B_3 K_h + O(K_h^2)] = K_i^2 [B_2 + B_3 K_i + O(K_i^2)] \quad (4.44)$$

ou seja

$$K_h^2 \left[\frac{4\varphi c'^2 + \varphi'^2}{32\varphi^2} - \frac{1}{96f\sqrt{\varphi}} \frac{d}{dx_h} \left(\frac{4f\varphi c'^2 + f\varphi'^2}{\varphi^{3/2}} \right) K_h + O(K_h^2) \right] \\ = K_i^2 \left[\frac{4\varphi c'^2 + \varphi'^2}{32\varphi^2} + \frac{1}{96f\sqrt{\varphi}} \frac{d}{dx_h} \left(\frac{4f\varphi c'^2 + f\varphi'^2}{\varphi^{3/2}} \right) K_i + O(K_i^2) \right] \quad (4.45)$$

que por sua vez pode ser reescrito da forma

$$K_h^2 \int_{x_{h-1}}^{x_h} g(t)f(t)dt [1 + O(K_h^2)] = K_i^2 \int_{x_h}^{x_{h+1}} g(t)f(t)dt [1 + O(K_i^2)] \quad (4.46)$$

onde

$$g(t) = \frac{\varphi'^2(t) + 4\varphi(t)c'^2(t)}{[\varphi(t)]^{3/2}} \quad (4.47)$$

Sob a suposição de um grande número de estratos, as amplitudes $K_h = (x_h - x_{h-1})$ são pequenas e suas maiores potências podem ser negligenciadas. Então o sistema de equações

minimal (4.10) ou equivalentemente o sistema de equações(4.46) pode ser aproximado por

$$K_h^2 \int_{x_{h-1}}^{x_h} g(t)f(t) dt = c_1 = \text{constante}, h = 1,2,\dots,L \quad (4.48)$$

onde os termos de ordem $O(m^4)$, $m = \sup_{(a,b)} K_h$ foram negligenciados em ambos os lados da equação (4.46).

Desta forma $\int_{x_{h-1}}^{x_h} g(t)f(t)dt = O(m)$ em vista do fato de $g(t)f(t)$

ser limitada para todo x em (a,b) . Além disso, Singh e Sukhatme (1969) observaram que para uma função $Q(x_{h-1}, x_h)$ tal que

$$K_h^2 \int_{x_{h-1}}^{x_h} g(t) f(t) dt = Q(x_{h-1}, x_h) \left[1 + O(K_h^2) \right] \quad (4.49)$$

o sistema de equações minimais (4.10) pode para o mesmo grau de aproximação em (4.48) ser aproximado por

$$Q(x_{h-1}, x_h) = \text{constante}, h = 1,2,\dots, L . \quad (4.50)$$

O conjunto de soluções aproximadas dadas pela equação (4.48) são muito boas e bem próximas da solução exata (4.10). No entanto, estas soluções aproximadas podem envolver dificuldades de cálculo. Descreveremos a seguir, algumas regras no sentido de evitar algumas destas dificuldades.

Regra 1 - Se na equação (4.44) nós negligenciamos os termos de ordem $O(m^3)$, $m = \sup_{(a,b)} K_h$, $K_h = x_h - x_{h-1}$, então

$$K_h^2 B_2 O(K_h^3) = K_i^2 B_2 + O(K_i^3) \text{ implica em}$$

$$K_h = \text{constante} = \frac{b-a}{L}, \quad h = 1, 2, \dots, L \quad (4.51)$$

e os pontos de estratificação aproximadamente ótimos são dados por

$$\left. \begin{aligned} x_h &= a + h \frac{(b-a)}{L}, \quad h = 1, 2, \dots, L-1 \\ x_0 &= a, \quad x_L = b \end{aligned} \right\} \quad (4.52)$$

Não podemos esperar resultados muito bons deste conjunto de soluções, mas ele deve ser usado nos casos onde não se conhece muito a respeito de $g(x)$ e $f(x)$. No entanto, usando eq. (4.48) e as várias identidades provadas na secção 4.3, outras regras de aproximações podem ser obtidas.

Regra 2 - Considere as equações (4.33) e (4.48) escritas abaixo:

$$\mu_{\varphi}(x_{h-1}, x_h) = \sqrt{\varphi(x_{h-1}) \varphi(x_h)} \quad [1 + O(K_h^2)]$$

$$e \quad K_h^2 \int_{x_{h-1}}^{x_h} g(t)f(t) dt = \text{constante}$$

Por outro lado

$$\begin{aligned} \int_{x_{h-1}}^{x_h} g(t)f(t) dt &= \omega_h \mu_g(x_{h-1}, x_h) \\ &= \omega_h \sqrt{g(x_{h-1})g(x_h)} \quad [1 + O(K_h^2)] \end{aligned}$$

Então uma aproximação para a solução exata é tal que

$$K_h^2 \int_{x_{h-1}}^{x_h} g(t)f(t) dt = K_h^2 \omega_h \sqrt{g(x_{h-1})g(x_h)} = c_2 \quad (4.53)$$

quando negligenciamos os termos de maior ordem.

Regra 3 - Usando (4.34) e (4.48) escritas abaixo

$$I_o(x_{h-1}, x_h) = \frac{1}{x_{h-1}x_h} \int_{x_{h-1}}^{x_h} t^2 f(t) dt [1 + o(K_h^2)]$$

$$K_h^2 \int_{x_{h-1}}^{x_h} g(t)f(t) dt = \text{constante}$$

Então, desprezando os termos de maior ordem, temos

$$\int_{x_{h-1}}^{x_h} g(t)f(t) dt = \frac{1}{x_{h-1}x_h} \int_{x_{h-1}}^{x_h} t^2 g(t)f(t) dt$$

E uma outra aproximação para o sistema de equação minimal é obtida de

$$\frac{K_h^2}{x_{h-1}x_h} \int_{x_{h-1}}^{x_h} t^2 g(t)f(t) dt = c_3 \quad (4.54)$$

Regra 4 - Usando a identidade (4.35) igual a

$$\left[\int_{x_{h-1}}^{x_h} \sqrt[\lambda]{f(t)} dt \right]^\lambda = K_h^{\lambda-1} \int_{x_{h-1}}^{x_h} f(t) dt [1 + o(K_h^2)]$$

nós podemos escrever

$$\int_{x_{h-1}}^{x_h} g(t)f(t) = \frac{1}{K_h^{1/\lambda-1}} \left[\int_{x_{h-1}}^{x_h} [g(t)f(t)]^\lambda dt \right]^{1/\lambda}$$

Então

$$\begin{aligned} K_h^2 \int_{x_{h-1}}^{x_h} g(t)f(t) dt &= \frac{K_h^2}{K_h^{1/\lambda-1}} \left[\int_{x_{h-1}}^{x_h} [g(t)f(t)]^\lambda dt \right]^{1/\lambda} \\ &= K_h^{\frac{3\lambda-1}{\lambda}} \left[\int_{x_{h-1}}^{x_h} [g(t)f(t)]^\lambda dt \right]^{1/\lambda} \\ &= \left\{ K_h^{3\lambda-1} \int_{x_{h-1}}^{x_h} [g(t)f(t)]^\lambda dt \right\}^{1/\lambda} = \text{constante} \end{aligned} \quad (4.55)$$

no caso particular de $\lambda = 1/2$ temos

$$\left[K_h \int_{x_{h-1}}^{x_h} \sqrt{g(t)f(t)} dt \right]^2 = c_4 \quad (4.56)$$

E para $\lambda = 1/3$

$$\left[\int_{x_{h-1}}^{x_h} \sqrt[3]{g(t)f(t)} dt \right]^3 = \text{constante}$$

ou seja

$$\int_{x_{h-1}}^{x_h} \sqrt[3]{g(t)f(t)} dt = c_5 \quad (4.57)$$

Em todas as regras apresentadas o valor da constante c_i deve ser determinado. É difícil encontrar o valor exato

dessas constantes, com exceção de c_5 , cujo valor é

$$\frac{1}{L} \int_a^b \sqrt[3]{g(t)f(t)} dt .$$

A justificativa deste valor de c_5 é semelhante a dada por Dalenius e Hodges no caso do \sqrt{f} .

Todas estas aproximações são consideradas para o caso de alocação de Neyman. No entanto, nós podemos verificar que o sistema de equações mínimas para alocação proporcional pode ser considerado como um caso particular do sistema de equações dando pontos ótimos de estratificação sob alocação de Neyman. Assim, o sistema de equações (4.10) se reduz ao (4.15), se consideramos $\varphi(x)$ constante e $\sigma_{hc}^2 = \sigma_{ic}^2$ para $i = h+1$ e $h = 1, 2, \dots, L-1$. Desta forma

$$\frac{[c(x_h) - \mu_{hc}]^2 + \sigma_{hc}^2 + \varphi(x_h) + \mu_h \varphi}{\sqrt{\sigma_{hc}^2 + \mu_h \varphi}} = \frac{[c(x_h) - \mu_{ic}]^2 + \sigma_{ic}^2 + \varphi(x_h) + \mu_i \varphi}{\sqrt{\sigma_{ic}^2 + \mu_i \varphi}}$$

se reduz a

$$[c(x_h) - \mu_{hc}]^2 = [c(x_h) - \mu_{ic}]^2$$

$$c^2(x_h) + \mu_{hc}^2 - 2c(x_h)\mu_{hc} = c^2(x_h) + \mu_{ic}^2 - 2c(x_h)\mu_{ic}$$

$$[\mu_{hc} - \mu_{ic}][\mu_{hc} + \mu_{ic} - 2c(x_h)] = 0$$

$$c(x_h) = \frac{\mu_{hc} + \mu_{ic}}{2}$$

Sob estas suposições, $g(t) \doteq [c'(t)]^2$ e assim as várias aproximações do sistema de equações dando pontos ótimos

de estratificação sob alocação proporcional podem ser obtidas dos correspondentes resultados para alocação de Neyman para $g(t) = [c'(t)]^2$.

4.5. Uso da Regressão como Método de Estimação

Neste item vamos considerar o caso particular da estimativa regressão combinada na estratificação ótima.

A estimativa regressão combinada é apropriada quando se supõe que os verdadeiros coeficientes de regressão, β_h , têm o mesmo valor em todos os estratos. Para calcular o valor médio \bar{y}_{lc} , primeiro achamos $\bar{y}_{st} = \sum_h \omega_h \bar{y}_h$ e $\bar{x}_{st} = \sum_h \omega_h \bar{x}_h$ e então

$$\bar{y}_{lc} = \bar{y}_{st} + b(\bar{X} - \bar{x}_{st}) \quad (4.58)$$

A estimativa \bar{y}_{lc} é não tendenciosa para μ desde que o valor de b seja predeterminado e tal que

$$b = \left[\sum_h \omega_h^2 S_{hxy} / n_h \right] \div \left[\sum_h \omega_h^2 S_{hx}^2 / n_h \right]$$

e no h -ésimo estrato

S_{hxy} = covariância amostral entre x e y

S_{hx}^2 = variância amostral da variável x .

A variância da estimativa \bar{y}_{lc} é dada por

$$v(\bar{y}_{lc}) = \sum_h \omega_h^2 (\sigma_{hy}^2 - 2\beta\sigma_{hxy} + \beta^2\sigma_{hx}^2) / n_h \quad (4.59)$$

$$\text{onde } \beta = \left[\sum_h \omega_h^2 \sigma_{hxy} / n_h \right] \div \left[\sum_h \omega_h^2 \sigma_{hx}^2 / n_h \right]$$

sob alocação de Neyman temos

$$V(\bar{y}_{\ell_c})_N = \left[\sum_h \omega_h \sqrt{\sigma_{hy}^2 - 2\beta\sigma_{hxy} + \beta^2\sigma_{hx}^2} \right]^2 / n \quad (4.60)$$

A expressão da variância acima, é uma função dos limites dos estratos da variável x.

Determinaremos a seguir, os limites ótimos dos estratos que correspondem $V(\bar{y}_{\ell_c})_N$ mínima com respeito a esses limites.

Suponha que a função densidade conjunta de (x,y) na população, é contínua e que a relação entre y e x é da forma

$$y = c(x) + e \quad (4.61)$$

onde $c(x)$ é uma função de x e e é tal que $E(e/x) = 0$ e $V(e/x) = \varphi(x) > 0$ para todo x no intervalo (a,b) de x com $(b-a) < \infty$. Sob esse modelo

$$\left. \begin{aligned} \sigma_{hy}^2 &= \sigma_{hc}^2 + \mu_{h\varphi} \\ \sigma_{hxy} &= \sigma_{hxc} \end{aligned} \right\} \quad (4.62)$$

onde no h-ésimo estrato

σ_{hc}^2 = variância de $c(x)$

σ_{hxc} = covariância entre x e $c(x)$

$\mu_{h\varphi}$ = valor esperado de $\varphi(x)$

De (4.60) e (4.62), nós temos a variância da estimativa regressão combinada como

$$\begin{aligned} V(\bar{y}_{\ell c}) &= \left[\sum_h \omega_h \sqrt{\sigma_{hy}^2 - 2\beta\sigma_{hxy} + \beta^2\sigma_{hx}^2} \right]^2 / n \\ &= \left[\sum_h \omega_h \sqrt{\sigma_{hc}^2 - 2\beta\sigma_{hxc} + \beta^2\sigma_{hx}^2 + \mu_{h\varphi}} \right]^2 / n \end{aligned}$$

seja $\lambda(x) = c(x) - \beta x$. Então, $\sigma_{h\lambda}^2 = \sigma_{hc}^2 + \beta^2\sigma_{hx}^2 - 2\beta\sigma_{hxc}$ e então

$$V(\bar{y}_{\ell c}) = \left[\sum_h \omega_h \sqrt{\sigma_{h\lambda}^2 + \mu_{h\varphi}} \right]^2 / n \quad (4.63)$$

Esta variância é a mesma que foi obtida na relação (4.3) do item 4.2 deste capítulo, com $\sigma_{h\lambda}^2$ no lugar de σ_{hc}^2 . As equações minimais que fornecem os limites ótimos dos estratos e os métodos de encontrar soluções aproximadas neste caso, são portanto as mesmas que foram determinadas nos itens 4.2 e 4.4 respectivamente.

A estimativa regressão é geralmente empregada quando a regressão de y em x é linear, ou seja,

$$y = \alpha + \beta x \quad (4.64)$$

neste caso, $c(x) = \alpha + \beta x$ e $\lambda(x) = \alpha$, tal que $\sigma_{\lambda}^2 = 0$, e assim

$$V(\bar{y}_{\ell c})_N = \left[\sum_h \omega_h \sqrt{\mu_{h\varphi}} \right]^2 / n \quad (4.65)$$

Minimizar $V(\bar{y}_{\ell c})_N$ é equivalente a minimizar

$$\sum_h \omega_h \sqrt{\mu_{h\varphi}} \quad (4.66)$$

Igualando a zero as derivadas da equação (4.66) com respeito a x_h , temos

$$\omega_h \frac{\partial}{\partial x_h} \sqrt{\mu_h} \varphi + \sqrt{\mu_h} \varphi \frac{\partial \omega_h}{\partial x_h} + \omega_i \frac{\partial}{\partial x_h} \sqrt{\mu_i} \varphi + \sqrt{\mu_i} \varphi \frac{\partial \omega_i}{\partial x_h} = 0 \quad (4.67)$$

onde $i = h+1$.

$$\frac{\partial \omega_h}{\partial x_h} = f(x_h) = - \frac{\partial \omega_i}{\partial x_h} \quad (4.68)$$

$$\mu_h \varphi = \int_{x_{h-1}}^{x_h} \varphi(t) f(t) dt / \omega_h \quad (4.69)$$

$$\begin{aligned} \frac{\partial \mu_h \varphi}{\partial x_h} &= \frac{\left[\omega_h \varphi(x_h) f(x_h) - f(x_h) \int_{x_{h-1}}^{x_h} \varphi(t) f(t) dt \right]}{\omega_h^2} \\ &= \frac{\omega_h \varphi(x_h) f(x_h) - \omega_h \mu_h \varphi f(x_h)}{\omega_h^2} \\ &= \frac{f(x_h)}{\omega_h} \left[\varphi(x_h) - \mu_h \varphi \right] \end{aligned} \quad (4.70)$$

De modo análogo

$$\frac{\partial \mu_i \varphi}{\partial x_h} = - \frac{f(x_h)}{\omega_i} \left[\varphi(x_h) - \mu_i \varphi \right] \quad (4.71)$$

Usando estes resultados acima em (4.67), temos

$$\omega_h \frac{f(x_h) [\varphi(x_h) - \mu_h \varphi]}{2\omega_h \sqrt{\mu_h \varphi}} + f(x_h) \sqrt{\mu_h \varphi} - \omega_i \frac{f(x_h) [\varphi(x_h) - \mu_i \varphi]}{2\omega_i \sqrt{\mu_i \varphi}} - f(x_h) \sqrt{\mu_i \varphi} = 0 \quad (4.72)$$

Então

$$\frac{\varphi(x_h) - \mu_h \varphi}{2 \sqrt{\mu_h \varphi}} + \sqrt{\mu_h \varphi} = \frac{\varphi(x_h) - \mu_i \varphi}{2 \sqrt{\mu_i \varphi}} + \sqrt{\mu_i \varphi}$$

$$\frac{\varphi(x_h) + \mu_h \varphi}{\sqrt{\mu_h \varphi}} = \frac{\varphi(x_h) + \mu_i \varphi}{\sqrt{\mu_i \varphi}} \quad (4.73)$$

onde $i = h+1, \quad h = 1, 2, \dots, L-1$

Neste caso, o valor de g dado pela expressão (4.47) torna-se igual a

$$g(t) = \frac{\varphi'^2(t)}{[\varphi(t)]^{3/2}} \quad (4.74)$$

Então, se a função $p(x) = g(x) f(x)$ onde $f(x)$ é a função densidade de x , é limitada e possui as duas primeiras derivadas contínuas para todo x em (a, b) , os vários métodos de soluções aproximadas para a equação (4.73) podem ser obtidos dos métodos determinados no item 4.4.

4.6. Exemplo Numérico

Como ilustração, considere a função densidade de x sendo $f(x) = e^{-x+1}, \quad 1 \leq x \leq \infty$. Para função de variância

condicional $\varphi(x)$ nós tomaremos $\varphi(x) = ax$ onde $a > 0$ é uma constante. Suporemos ainda regressão linear de y em x na população por ser exatamente a situação onde o uso de estimativa da regressão linear é recomendada. Desta forma, $c(x) = \alpha + \beta x$.

Para determinar a constante a iremos considerar que 75% da variação total é explicada pela regressão. Desenvolvemos a $V(y)$ e ρ^2 em função de $E(\varphi(x))$ e $V(c(x))$. Usando algumas substituições encontramos a em função de $E(x)$, $\text{Var}(x)$ e ρ^2 . Então, para calcular o valor de a é suficiente usar $\rho^2 = 0,75$ e calcular o valor de $E(x)$ e $V(x)$ para a função densidade $f(x) = e^{-x+1}$.

Para ter uma amplitude finita, a distribuição exponencial foi truncada em $x = 6$ e a amplitude foi dividida em 20 classes de igual comprimento. Para determinar os pontos de estratificação aproximadamente ótimo, foi usada a equação (4.57) da regra 4. A função

$$g(x) = \frac{\varphi'(x)}{[\varphi(x)]^{3/2}} = \frac{\sqrt{a}}{x\sqrt{x}}$$

foi calculada nos pontos médios dos intervalos de classe e então multiplicada por ω_h e denominada por $p(x)$. A raiz cúbica de $p(x)$ foi encontrada para as 20 classes e seus valores acumulados. Os pontos de estratificação foram então obtidos tomando iguais intervalos no acumulado total.

TABELA 4.1

CÁLCULO DO LIMITE DOS ESTRATOS

Classes	ω_h	$g(x)$	$p(x)$	$\text{cum} \sqrt[3]{p(x)}$
1,00 - 1,25	0,22120	0,31472	0,06962	0,4114
1,25 - 1,50	0,17227	0,23292	0,04012	0,7537
1,50 - 1,75	0,13416	0,18129	0,02432	1,0435
1,75 - 2,00	0,10449	0,14627	0,01528	1,2916
2,00 - 2,25	0,08137	0,12123	0,00986	1,5061
2,25 - 2,50	0,06337	0,10260	0,00650	1,6927
2,50 - 2,75	0,04936	0,08830	0,00436	1,8561
2,75 - 3,00	0,03844	0,07704	0,00296	1,9997
3,00 - 3,25	0,02994	0,06798	0,00203	2,1264
3,25 - 3,50	0,02331	0,06057	0,00141	2,2386
3,50 - 3,75	0,01816	0,05441	0,00099	2,3382
3,75 - 4,00	0,01414	0,04923	0,00070	2,4268
4,00 - 4,25	0,01101	0,04482	0,00049	2,5058
4,25 - 4,50	0,00858	0,04104	0,00035	2,5765
4,50 - 4,75	0,00668	0,03776	0,00025	2,6396
4,75 - 5,00	0,00520	0,03489	0,00018	2,6962
5,00 - 5,25	0,00405	0,03237	0,00013	2,7470
5,25 - 5,50	0,00316	0,03014	0,00010	2,7927
5,50 - 5,75	0,00246	0,02815	0,00007	2,8337
5,75 - 6,00	0,00191	0,02637	0,00005	2,8706

TABELA 4.2

LIMITES DOS ESTRATOS APROXIMADAMENTE ÓTIMOS PARA L=2,3,4 e 5 NO CASO DA DISTRIBUIÇÃO EXPONENCIAL.

L	PONTOS DE ESTRATIFICAÇÃO			
	x_1	x_2	x_3	x_4
2	2,167			
3	1,675	2,850		
4	1,474	2,167	3,308	
5	1,369	1,855	2,545	3,645

TABELA 4.3

VARIÂNCIAS E EFICIÊNCIA RELATIVA DA DISTRIBUIÇÃO

L	$n V(\bar{y}_{lc})$	EFICIÊNCIA RELATIVA (%)
1	0,27596	100,00
2	0,26227	105,22
3	0,26040	105,98
4	0,25967	106,27
5	0,25937	106,40

Na tabela acima temos representadas as eficiências relativas dos diversos números de estratos quando comparadas com a variância da amostra aleatória simples (L=1). Da tabela concluímos que o melhoramento na diminuição da variância quando aumentamos o número de estratos tende a uma estabilidade.

CAPÍTULO 5

APLICAÇÃO DAS REGRAS DE ESTRATIFICAÇÃO ÓTIMA

5.1. Introdução

O estudo de unidades elementares de uma característica quantitativa y apresenta interessantes problemas amostrais quando o universo é caracterizado por uma distribuição de frequência altamente assimétrica, como, por exemplo, pacientes de hospitais, matrículas em colégios, empréstimos bancários, área de estabelecimentos agrícolas, etc.

Segundo Hess e Sethi (1966), "Estratificação para melhorar estimativas para a população total, torna-se mais importante na presença de assimetria, e quando a variabilidade e a disparidade em tamanhos nos grupos aumentam".

Interessantes trabalhos têm sido feitos para o problema de dividir uma função de frequência $f(y)$ em L estratos quando se deseja estimar a média populacional de uma amostra aleatória estratificada.

Para ilustrar os métodos descritos nos três primeiros capítulos foi tomada uma distribuição de área dos imóveis rurais de proprietários agrícolas do município de Independência-Ce. Em casos onde um mesmo produtor é dono de mais de um imóvel, a soma das áreas de suas propriedades foi considerada como uma unidade de seleção. Os dados foram retirados da "Relação de Cobrança" obtida no Departamento de Cadastros e Tributação do INCRA e referem-se ao recadastramento de 1980. Uma melhor descrição desta população é dada no item 5.2.1 deste capítulo.

Os critérios de "Estratificação Aproximadamente Ótima", dados por Dalenius e Hodges (Cap. 1), Ekman (Cap. 2) e o de Estratos de Igual Tamanho Agregado (Cap. 3), foram aplicados para L=2,3,4 e 5 estratos. Os métodos descritos nos referidos capítulos foram desenvolvidos para distribuições de frequências teóricas. Cochran (1961) comparou oito distribuições assimétricas, representativas daquelas encontradas em amostragem de instituições do tipo no qual existem muitas instituições pequenas e poucas muito grandes. Um resumo do trabalho de Cochran é apresentado no item 5.3 deste capítulo. Neste trabalho, as regras são aplicadas a uma distribuição de frequência para dados agrupados, com os agrupamentos feitos a partir da amplitude de área das propriedades.

5.2. Aplicação das Regras

5.2.1. O Estudo da População

A população estudada é constituída dos proprietários de imóveis rurais no município de Independência-Ce classifi-

cados de acordo com a área total de suas propriedades. Dessa forma, um grande número de pequenos e médios proprietários e poucos muito grandes compreende o universo em estudo.

Quando os proprietários de imóveis rurais são ordenados pelo tamanho das propriedades (em hectares), observamos 5% dos maiores proprietários detendo 42,7% da área total, enquanto os 80% menores estão de posse de apenas 27,3%. A tabela 5.1 mostra essa distribuição de área onde pode ser observado esse alto nível de concentração.

TABELA 5.1

DISTRIBUIÇÃO DE ÁREA DAS PROPRIEDADES
ORDENADAS PELO TAMANHO

% DE PROPRIETÁRIO (ordenados p/tam. de área)	PARTICIPAÇÃO PERCENTUAL NA ÁREA TOTAL	PERCENTUAL ACUMULADO
0 - 5	0,15	0,15
5 - 10	0,29	0,44
10 - 15	0,41	0,85
15 - 20	0,52	1,37
20 - 25	0,64	2,01
25 - 30	0,81	2,82
30 - 35	0,93	3,75
35 - 40	1,13	4,88
40 - 45	1,36	6,24
45 - 50	1,66	7,90
50 - 55	1,97	9,87
55 - 60	2,28	12,15
60 - 65	2,70	14,85
65 - 70	3,32	18,17
70 - 75	4,04	22,21
75 - 80	5,05	27,26
80 - 85	6,77	34,03
85 - 90	9,33	43,36
90 - 95	13,93	57,29
95 - 100	42,71	100,00

Para a aplicação das regras descritas nos três primeiros capítulos, consideramos uma distribuição de frequências dos proprietários de acordo com o tamanho do imóvel que é a variável de estratificação deste trabalho. As classes de área foram agrupadas a partir da amplitude total de 2215 propriedades. A descrição de cada regra pode ser vista nos subitens seguintes

5.2.2. Método de Dalenius e Hodges

Neste método, os valores de $\sqrt{f_h}$ são acumulados como na coluna (4) da tabela 5.2. A regra consiste em tomar iguais intervalos no cum $\sqrt{f_h}$, ou seja, cada estrato deve conter cum $\sqrt{f_h}$ dividido por L na escala cum $\sqrt{f_h}$. Assim, para L=3 estratos, por exemplo, temos $195,24/3 = 53,08$. O melhor que pode ser feito é dividir os estratos em 0-8%, 8-28% e 28-100% onde o valor do cum $\sqrt{f_h}$ é 53,08, 50,54 e 55,62 respectivamente. O quadro abaixo apresenta os limites dos estratos para L=2, 3, 4 e 5.

LIMITE DO ESTRATO (%)					
L	1	2	3	4	5
2	0-16	16-100	-	-	-
3	0- 8	8- 28	28-100	-	-
4	0- 4	4- 16	16- 40	40-100	-
5	0- 4	4- 12	12- 24	24- 52	52-100

5.2.3. Método de Ekman

Nesta regra, nós igualamos o produto da frequência acumulada dentro do estrato pela sua extensão. O primeiro

passo é acumular os valores de f, coluna (6) da tabela 5.2. O método é um pouco trabalhoso de aplicar porque o valor da $\sum_h \omega_h (y_h - y_{h-1})$ não é constante, dependendo tanto de L como da posição dos limites. Assim, para um dado L não fica muito claro a forma de igualar $\omega_h (y_h - y_{h-1})$ em estratos individuais. Uma orientação sugerida por Ekman é calcular o produto para L=1, dado neste caso por $Q = (\text{cum } f)(y_L - y_0) = 2215 \times 100 = 221500$.

TABELA 5.2
CÁLCULO DOS LIMITES DOS ESTRATOS PARA OS
TRÊS MÉTODOS PARA L = 2, 3, 4 e 5.

(1)	(2)	(3)	(4)	(5)	(6)
%	μ_h	f_h	$\text{cum} \sqrt{f_h}$	$\text{cum } f_h \mu_h$	$\text{cum } f_h$
0 - 4	2	1017	31,89	2034	1017
4 - 8	6	449	53,08	4728	1466
8 - 12	10	228	68,18	7008	1694
12 - 16	14	117	79,00	8646	1811
16 - 20	18	77	87,77	10032	1888
20 - 24	22	71	96,20	11594	1959
24 - 28	26	55	103,62	13024	2014
28 - 32	30	28	108,91	13864	2042
32 - 36	34	31	114,48	14918	2073
36 - 40	38	35	120,40	16248	2108
40 - 44	42	14	124,13	16836	2122
44 - 48	46	9	127,13	17250	2131
48 - 52	50	7	129,78	17600	2138
52 - 56	54	8	132,61	18032	2146
56 - 60	58	10	135,77	18612	2156
60 - 64	62	8	138,60	19108	2164
64 - 68	66	7	141,24	19570	2171
68 - 72	70	10	144,40	20270	2181
72 - 76	74	11	147,72	21084	2192
76 - 80	78	7	150,37	21630	2199
80 - 84	82	4	152,37	21958	2203
84 - 88	86	4	154,37	22302	2207
88 - 92	90	3	156,10	22572	2210
92 - 96	94	2	157,51	22760	2212
96 - 100	98	3	159,24	23054	2215

Para L estratos, o valor constante por estrato é aproximadamente igual a Q/L^2 . Esta relação é exata para uma distribuição retangular, mas tende a dar um resultado muito alto para distribuições assimétricas. Como exemplo, tomemos $L=3$ estratos. Temos $Q/L^2 = 221500/9 = 24611$. A melhor partição é então 0-8%, 8-32% e 32-100% com valores $\omega_h(y_h - y_{h-1})$ iguais a 11728, 13824 e 11764 respectivamente. Para $L=2, 3, 4$ e 5 estratos os limites estão no quadro abaixo.

L	LIMITE DOS ESTRATOS (%)				
	1	2	3	4	5
2	0-16	16-100	-	-	-
3	0- 8	8- 32	32-100	-	-
4	0- 4	4- 16	16- 40	40 100	-
5	0- 4	4- 12	12- 24	24- 48	48-100

5.2.4. Igual Tamanho Agregado

Em geral, os pontos médios da distribuição de frequências podem ser usados como estimativas dos valores de μ_h . Nesta distribuição os pontos médios são 2, 6, 10 etc como na coluna (2) da tabela 5.2.

Na aplicação desta regra, os valores de $f_h \mu_h$ são acumulados como na coluna (5) da tabela 5.2. O tamanho agregado em cada estrato será dado por $\sum f_h \mu_h / L$. Assim, para $L=3$, por exemplo, temos $23054/3 = 7685$. Para o primeiro es-

trato o valor mais próximo é 7008 para o intervalo de 0-12%. Dessa forma $2 \times 7685 = 15370$, leva a uma escolha para o segundo estrato, o intervalo de 12-36% onde com $f_h \mu_h$ é 14918. Os resultados para L=2,3,4 e 5 estratos estão a seguir.

L	LIMITE DOS ESTRATOS (%)				
	1	2	3	4	5
2	0-24	24-100	-	-	-
3	0-12	12- 36	36-100	-	-
4	0- 8	8- 24	24- 48	48-100	-
5	0- 8	8- 16	16- 32	32- 60	60-100

5.2.5. Comparação das Diferentes Regras

Os métodos de Dalenius e Hodges e o de Ekman tiveram desempenhos aproximadamente iguais em todos os tamanhos de estratos e exatamente iguais para L=2 e 4, notando-se, no geral, um pequeno ganho com o método de Ekman (ver tabela 5.3). A regra para "igual tamanho agregado" apresentou variâncias superiores aos dois métodos anteriores em todos os casos. O relativo insucesso para "igual tamanho agregado" é a pressuposição da razão σ_h / μ_h ser aproximadamente constante, o que não acontece neste exemplo.

A seguir são apresentadas três tabelas comparativas das variâncias.

TABELA 5.3

VARIÂNCIAS ENCONTRADAS PARA AS TRÊS DIFERENTES REGRAS

L	cum $\sqrt{f_h}$	$\omega_h (y_h - y_{h-1})$	$\omega_h \mu_h$
2	7 577,25	7 577,25	8 497,31
3	3 384,83	3 375,31	3 561,22
4	1 817,08	1 817,08	1 925,08
5	1 132,17	1 120,80	1 360,16

TABELA 5.4

COMPARAÇÃO DA VARIÂNCIA NO ESTRATO L = 2, 3, 4 e 5 COM A VARIÂNCIA EM AAS (L = 1)

L	cum $\sqrt{f_h}$	$\omega_h (y_h - y_{h-1})$	$\omega_h \mu_h$
2	0,2000	0,2000	0,2243
3	0,0894	0,0891	0,0940
4	0,0480	0,0480	0,0508
5	0,0299	0,0296	0,0359

TABELA 5.5

COMPARAÇÃO DA VARIÂNCIA NO ESTRATO L=2,3,4 e 5 COM O ESTRATO ANTERIOR

L	cum $\sqrt{f_h}$	$\omega_h (y_h - y_{h-1})$	$\omega_h \mu_h$	MÉDIA	$(L-1)^2 / L^2$
2	0,2000	0,2000	0,2243	0,2081	0,2500
3	0,4467	0,4455	0,4191	0,4371	0,4444
4	0,5368	0,5383	0,5406	0,5386	0,5625
5	0,6231	0,6168	0,7065	0,6488	0,6944

Comparando as variâncias em cada regra com a variância no caso de amostra aleatória simples (AAS), ou seja, L=1,

como também $\text{Var } L/\text{Var}(L-1)$, constatou-se ganho pelo aumento do número de estratos em todos os três casos.

Como aproximação válida para a distribuição retangular, Dalenius (1957) sugeriu que $\text{Var } L/\text{Var}(L-1) = (L-1)^2/L^2$. Esta fórmula dá 0,250, 0,444, 0,562 e 0,694 para $L=2,3,4$ e 5 respectivamente. As médias para as três regras na tabela 5.5 ficam próximos destes valores, mas, superficialmente, abaixo deles.

Em linhas gerais, os resultados encontrados para o exemplo aqui utilizado são semelhantes ao de Cochran (1961) quando ele verificou o desempenho de quatro regras para delimitação dos limites dos estratos em oito distribuições de frequências. Devido a importância deste trabalho, a seguir apresentaremos um resumo, com suas principais conclusões.

5.3. Comparação de Métodos para Determinação dos Limites dos Estratos (Resumo)

As oito distribuições escolhidas para análise, são assimétricas com uma longa extremidade positiva e são representativas daquelas encontradas em amostragens de instituições do tipo no qual existem muitas unidades pequenas e poucas grandes.

As oito variáveis são como segue:

- Renda bruta ajustada por taxa de retorno nos Estados Unidos em 1951;

- Número de estudantes em quatro anos de colégios dos Estados Unidos em 1952-53;
- População das cidades dos Estados Unidos em 1940;
- Recursos de grandes bancos comerciais dos Estados Unidos;
- Número de fazendas por unidade de área amostral em Seneca Country, N.Y.;
- Proporção de empréstimos bancários destinados a (i) comércio e indústria; (ii) bens imóveis; (iii) agrícolas.

O artigo compara quatro regras que foram propostas para aproximação dos melhores limites dos estratos, onde uma função de frequência $f(y)$ é subdividida em $L=2,3$ e 4 estratos.

Os quatro métodos de aproximação são:

- Método proposto por Dalenius e Hodges (ver Cap. 1);
- Método proposto por Ekman (ver Cap. 2)
- Estratos de Igual Tamanho Agregado (ver Cap. 3);
- Método proposto por Durbin (1948) a partir de uma revisão na tese de Dalenius.

Em resumo, a regra de Durbin consiste em tomar áreas iguais sob uma distribuição de frequências com densidade entre a distribuição original e uma distribuição retangular, Desta forma, considerando $F(y)$ o acumulado de $f(y)$, uma distribuição retangular $r(y) = \frac{F(y)}{y_L - y_0}$ pode ser formada sobre a mesma amplitude. Os limites dos estratos são obtidos tomando intervalos iguais no acumulado de $1/2[r(y) + f(y)]$.

Com relação a precisão relativa das regras, foram

determinados para cada estrato e cada regra, os valores de $V(\bar{y}_{st}) = [\sum_h \omega_h \sigma_h]^2$. Em seguida, os limites que minimizam esta variância foram encontrados por tentativa. As razões V/V_{\min} estão na tabela 5.6 onde os oito conjuntos de dados são arranjados de cima para baixo pelo afastamento crescente de uma distribuição retangular. Nesta tabela pode-se observar que o método de Dalenius e Hodges e o de Ekman são os de melhor desempenho. A regra de Dalenius e Hodges (cum $\sqrt{f_h}$) coincidiu com o mínimo em 12 dos 24 casos enquanto o de Ekman [$\omega_h (y_h - y_{h-1})$] coincide em 6 dos 24 casos e é apenas ligeiramente superior na maioria dos outros casos. Em ambas as regras, a média das razões de variância para os 24 casos foi aproximadamente 1,03. As regras de igual tamanho agregado ($\omega_h \mu_h$) e de Durbin [cum(r+f)] apresentaram resultados mais pobres. A média obtida nas 24 razões foi de 1,29 e 1,20 respectivamente.

TABELA 5.6

RAZÕES DE $V(\bar{y}_{st})$ p/ $V_{\min}(\bar{y}_{st})$ PARA AS DIFERENTES REGRAS

L	Empréstimos Agrícolas				Empréstimos de Bens Imóveis			
	cum \sqrt{f}	$\omega_h \mu_h$	$\mu_h (y_h - y_{h-1})$	cum(r+f)	cum \sqrt{f}	$\omega_h \mu_h$	$\omega_h (y_h - y_{h-1})$	cum(r+f)
2	1,03	1,36	1,03	1,03	1 *	1,19	1,06	1,06
3	1,03	1,89	1,05	1 *	1 *	1,28	1 *	1,07
4	1 *	1,96	1 *	1 *	1 *	1,39	1,05	1,10
	Empréstimos Industriais				Recursos de Bancos			
2	1 *	1,07	1,07	1,07	1,03	1,01	1,03	1,03
3	1*	1,35	1,07	1,14	1 *	1,07	1,04	1,12
4	1,01	1,42	1,01	1,21	1,09	1,05	1,09	1,22
	Estudantes de Colégio				População da cidade			
2	1,02	1,06	1,02	1,02	1 *	1,03	1 *	1 *
3	1,08	1,18	1,00	1,08	1 *	1 *	1,01	1,16
4	1,08	1,16	1,02	1,10	1,03	1,02	1,00	1 *
	Nº de Faz. por Unid. Amostrais				Renda Bruta			
2	1 *	1 *	1,16	1,16	1,25	1,70	1 *	1 *
3	1 *	1,22	1 *	1,45	1,02	1,81	1,08	2,76
4	1,02	1 *	1 *	1,39	1 *	1,69	1,05	1,52

* Indica que os limites dados pelas regras foram aqueles que fizeram a variância mínima.

Como $V(\bar{y}_{st})$ decresce quando o número de estratos é crescente, as sucessivas razões V_2/V_1 , V_3/V_2 e V_4/V_3 foram calculadas para cada distribuição arranjada na tabela 5.7 em ordem crescente de assimetria da esquerda para a direita.

Comparando $V_L/V_{(L-1)}$ com $(L-1)^2/L^2$ foram obtidos resultados bem próximos da média para as 8 distribuições.

TABELA 5.7

RAZÃO DE V_{\min} PARA L ESTRATOS PARA V_{\min} PARA (L-1)ESTRATOS

L	Distribuição de Frequência								MÉDIA	$\frac{(L-1)^2}{L^2}$
	1	2	3	4	5	6	7	8		
2	0,225	0,266	0,288	0,205	0,178	0,200	0,281	0,214	0,232	0,250
3	0,402	0,459	0,431	0,410	0,426	0,434	0,377	0,457	0,424	0,444
4	0,539	0,555	0,537	0,530	0,623	0,550	0,444	0,572	0,544	0,562

Com duas das distribuições (estudantes de colégio, 1952-53 e população da cidade, 1940) foi estudada a eficiência da estratificação sobre este conjunto de dados na precisão de pesquisas posteriores. Para os colégios, distribuições bivariadas foram trabalhadas para matrículas de 1952-53 e 1955-56 e para matrícula 1952-53 e 1958-59, simulando pesquisas com estratos construídos dos dados de 3 e 6 anos. Semelhantemente, a população da cidade de 1940 foi testada com os dados da população da cidade em 1950.

Para cada regra a tabela 5.8 mostra o valor médio de V/V_{\min} nos períodos anterior e posterior.

TABELA 5-8

EFICIÊNCIA RELATIVA DAS REGRAS EM PERÍODOS ANTERIORES E POSTERIORES - VALORES DE V/V_{\min} -

Distribuição	Regra			
	$\text{cum } \sqrt{f}$	$\omega_h \mu_h$	$\omega_h (y_h - y_{h-1})$	$\text{cum}(r+f)$
Anterior	1,03	1,09	1,01	1,06
Posterior	1,05	1,10	1,04	1,06

Para as duas distribuições (anterior e posterior) todas as regras foram "boas" quando aplicadas a dados correntes e quase igualmente "boas" quando usadas alguns anos depois.

CAPÍTULO 6

CONCLUSÕES

As técnicas de estratificação aproximadamente ótima apresentadas neste trabalho são importantes na utilização de Planos de Amostragem Estratificada.

Dos métodos aqui apresentados vale ressaltar o desempenho da regra de Dalenius e Hodges (Cap. 1) e da regra de Ekman (Cap. 2). A regra de Dalenius e Hodges, em particular, parece ser mais indicada que a de Ekman, pela simplicidade de cálculos e ótimos resultados em situações de natureza prática.

No exemplo apresentado no capítulo 5, obtivemos conclusões semelhantes as de Cochran em seu trabalho de "Comparação de Métodos para Determinação dos Limites dos Estratos", escrito em 1961 e onde ele compara oito distribuições de frequências.

Vale assinalar que a utilização das regras poderia ter gerado melhor partição, se maior número de intervalos de classe fôsse utilizado.

Outra observação a ser feita é que, às vezes, a distribuição de frequências original contém comprimentos desiguais com intervalos mais largos no final da distribuição de y. Neste caso, nós temos $f_h = \frac{\omega_h}{d_h}$ (densidade no intervalo h) onde ω_h é a proporção no intervalo h e d_h é a amplitude. Assim, se um intervalo de comprimento d passa para um outro de comprimento ud, o valor de \sqrt{f} na regra de Dalenius e Hodges deve ser multiplicado por \sqrt{u} na formação do cum \sqrt{f} .

Outro aspecto a ser ressaltado é a não utilização da estratificação com uso de uma variável auxiliar (Cap. 4) na estratificação que fizemos para o tamanho dos imóveis rurais no município de Independência-Ce. Nossa justificativa para este fato é que não conhecíamos uma função de regressão de alguma variável correlacionada com a área total do imóvel, tais como renda, área do imóvel em período anterior, etc.

Alguns tópicos relacionados com Estratificação Ótima, que não foram abordados aqui, podem dar continuidade a este trabalho. Entre eles sugerimos:

- Análise por agrupamento para Estratificação usado em procedimentos amostrais de múltiplos estágios onde a população é frequentemente pequena e tal que o número de fatores de estratificação que podem ser empregados é limitado. Golder e Yeomans (1973), Herler e Day (1975) e Yeomans e Golder (1975) são alguns dos autores com artigos escritos sobre este assunto;

- O uso do modelo Normal Bivariado em Estratificação Ótima e Aproximadamente Ótima, abordado por Anderson, Kish e Cornell (1976) onde eles comparam limites determinados pela regra do \sqrt{f} devido a Dalenius e Hodges (1959) e a regra do $\sqrt[3]{f}$ desenvolvida por Thomsen (1976) e concluem desempenho ligeiramente superior para a regra do \sqrt{f} . Ainda sobre este assunto, Brown (1976) desenvolve um trabalho sobre a "Eficiência da Amostragem Estratificada na Estimação de Parâmetros para uma População Normal Bivariada".
- Estratificação Ótima com duas variáveis, estudado por Ghosh (1963), numa tentativa de estender a teoria de Dalenius (1950) de Estratificação univariada, para duas variáveis, e por Thomsen (1977) que analisa os efeitos da estratificação quando duas variáveis são usadas;
- O caso mais geral da estratificação de múltiplas variáveis e múltiplos objetivos abordado por Kish e Anderson (1978).

As sugestões acima mencionadas nos levam a crer que muitos estudos ainda podem ser feitos no sentido de continuar e aprofundar esta monografia.

RELAÇÃO BIBLIOGRÁFICA

01. Anderson, D.W., Kish, L. and Cornell, R.G. (1976): "Quantifying Gains from Stratification for Optimum and Approximately Optimum Strata Using a Bivariate Normal Model". *Journal of the American Statistical Association*, 71, 887-892.
02. Brown, G.H. (1976): "The Efficiency of Stratified Sampling in the Estimation of Parameters for a Bivariate Normal Population". *Applied Statistics*, 25, 1-7.
03. Cochran, W.G., (1961): "Comparisons of Methods for Determining Stratum Boundaries", *Bulletim of the International Statistical Institute*, 38(2), 345-358.
04. _____ (1963): *Sampling Techniques*. John Wiley and Sons, New York.
05. Courant, R. (1965): *Cálculo Diferencial e Integral*, Vol. I Ed. Globo, Porto Alegre.
06. Dalenius, T. and Hodges, J.L. (1959): "Minimum Variance Stratification". *Journal of the American Statistical Association*, 54, 88-101.
07. Ekman, G. (1959): "An Approximation Useful in Univariate Stratification". *The Annals of Mathematical Statistics*, 30, 219-229.

08. Ekman, G. (1959): "Approximate Expressions of the Conditional Mean and Variance Over Small Intervals of a Continuous Distribution". *The Annals of Mathematical Statistics*, 30, 1131-1134.
09. Ghosh, S.P. (1963): "Optimum Stratification with two Characters". *The Annals of Mathematical Statistics*, 34, 866-872.
10. Golder, P.A. and Yeomans, K.A. (1973): "The Use of Cluster Analysis for Stratification". *Applied Statistics*, 22, 213-219.
11. Herler, R.M. and Day, G.S. (1975): "A Supplementary Note on the Use of Cluster Analysis for Stratification" *Applied Statistics*, 24, 342-344.
12. Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953): *Sample Survey Methods and Theory*, Vol. I, John Wiley and Sons, New York.
13. Hess, J., Sethi, V.K. and Balakrishnam T.R. (1966): "Stratification a Pratical Investigation", *Journal of the American Statistical Association*, 61, 74-90.
14. Kish, L. (1965): *Survey Sampling*, John Wiley and Sons, New York.
15. Kish, L. and Anderson, D.W. (1978): "Multivariate and Multipurpose Stratification", *Journal of the American Statistical Association*, 73, 24-34.
16. Mahalandbis, P.C. (1952): "Some Aspects of the Design of Sample Survey", *Sankhyā*, 12, 1-7,

17. Raj, D. (1964): "On Forming Strata of Equal Aggregate Size". *Journal of the American Statistical Association*, 59, 481-486.
18. Serfling, R.J. (1968): "Approximately Optimum Stratification". *Journal of the American Association*, 63, 1298 - 1309.
19. Sethi, V.K. (1963): "A Note on Optimal Stratifications of Populations for Estimating the Population Mean", *Australian Journal of Statistics*, 5, 20-33,
20. Sing, R. (1971): "Approximately Optimum Stratification on the Auxiliary Variable", *Journal of the American Association*, 66, 829-833.
21. Singh, R. and Sukhatme B.V. (1969): "Optimum Stratification". *Annals of the Institute of Statistical Mathematics*, 21, 515-528.
22. _____ (1971): "Optimum Stratification in Sampling with Varying Probabilistics". *Annals of the Institute of Statistical Mathematics*, 24, 485-494,
23. _____ (1973): "Optimum Stratification with Ratio and Regression Methods of Estimation". *Annals of the Institute of Statistical Mathematics*, 25. 627-633.
24. Thomsen, J. (1976): "A Comparison of Approximately Optimal Stratification Given Proportional Allocation with Other Methods of Stratification and Allocation", *Metrika*, 23, 15-25

25. Thomsen, J. (1977): "On the Effect of Stratification when two Stratifying Variables are Used". *Journal of the American Statistical Association*, 72, 149-153.
26. Yeomans, K.A. and Golder, P.A. (1975): "Further Observations on the Stratification of Birmingham Wards by Clustering: a Riposte". *Applied Statistics*, 24, 345-346.