

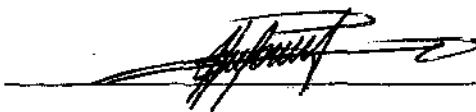
APLICAÇÃO DO MÉTODO BOOTSTRAP
NA ESTIMAÇÃO DA VARIÂNCIA
DO ESTIMADOR DE RAZÃO

JAIR BISCOLA

APLICAÇÃO DO MÉTODO BOOTSTRAP NA ESTIMAÇÃO
DA VARIÂNCIA DO ESTIMADOR DE RAZÃO

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. Jair Biscola e aprovada pela Comissão Julgadora.

Campinas, 20 de março de 1985.



Prof.Dr. Sebastião de Amorim
Orientador

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

MARÇO/1985.

Dedico o presente trabalho
a minha esposa Eliane e a
meus filhos Paulo Henrique
e Iara Cristina.

AGRADECIMENTOS

Inicialmente, deixo aqui a minha imensa gratidão ao Prof. Sebastião de Amorim, pela segura orientação que me foi dispensada durante a elaboração desta dissertação.

Quero também registrar os meus agradecimentos aos incentivos recebidos de minha esposa, meus pais e irmãos, meus colegas e professores, que muito contribuíram para o desempenho das tarefas desenvolvidas.

Finalmente, reservo meus agradecimentos à UNICAMP, CAPES e ao Departamento de Matemática da UFMS, este principalmente na figura do Prof. Eronides de Jesus Biscola, às contribuições recebidas, sem as quais não seria possível a realização deste trabalho.

INDICE

INTRODUÇÃO	i
CAPÍTULO I - ESTIMADORES DE RAZÃO: COMO ESTIMAR SUA VARIÂNCIA	1
1.1. Introdução	1
1.2. As Dificuldades de se Estimar a Variância	7
1.3. Estimadores Propostos	9
1.4. Propriedades Assintóticas de Alguns Estimadores	11
1.5. Os Critérios de Wu	12
CAPÍTULO II - O MÉTODO "BOOTSTRAP" PARA ESTIMAÇÃO DE DISTRIBUIÇÕES AMOSTRAIS	14
2.1. O "Jackknife"	14
2.2. O "Bootstrap" como uma Variante do "Jackknife"	16
2.3. Regressão e o "Bootstrap"	18
2.4. Possíveis Usos do "Bootstrap" no Problema da Estimacão do Estimador de Razão	24
CAPÍTULO III - O ESTIMADOR "BOOTSTRAP" DA VARIÂNCIA DO ESTIMADOR DE RAZÃO, EM DIVERSAS SITUAÇÕES	28
3.1. Um Problema Típico	28

3.2. O Efeito de t no Desempenho Relativo dos 6 Estimadores de $V(\hat{\bar{Y}}_R)$	36
3.3. Situações em que o "Bootstrap" Oferece Melhores Soluções	45
3.4. Um Estudo Monte Carlo do Desempenho Relativo dos 5 Estimadores de $V(\hat{\bar{Y}}_R)$ para $t = 2$ e X Proveniente de uma Distribuição Gama	46
APÊNDICE	50
1. Programas Computacionais	50
2. Material Processado	57
REFERÊNCIA BIBLIOGRÁFICA	64

INTRODUÇÃO

Os estimadores de razão têm amplas aplicações práticas no problema de estimação de \bar{Y} em planos de amostragem (Cochran [1977]).

A estimação da variância de $\hat{\bar{Y}}_R$, o estimador de razão de \bar{Y} é, contudo, um problema que tem desafiado as abordagens teóricas. A menos de situações muito particulares, os estimadores da variância de $\hat{\bar{Y}}_R$ são apenas soluções aproximadas que supõem algumas condições restritivas na população (Rao [1969], Rao e Kuzik [1974], Cochran [1977]).

O objetivo do presente trabalho é explorar o método "bootstrap" (Efron [1979]) na estimação da variância do estimador de razão, desenvolvendo alguns estimadores, e comparando-os com estimadores usuais deste parâmetro.

Desenvolvemos o assunto em três capítulos e dois apêndices. No capítulo I damos uma breve introdução ao estimador de razão, destacando as dificuldades existentes para se estimar a variância do mesmo. Apresentamos os estimadores da variância propostos na literatura, bem como as propriedades assintóticas de alguns deles.

Finalmente apresentamos os critérios de comparação entre estimadores usuais estabelecidos por Wu [1982].

No capítulo II discutimos o método "bootstrap" (Efron [1979]),

para estimação de distribuições amostrais, relacionando-o ao "jackknife" (Quenouille [1956], Miller [1974], Efron [1979]), e apresentamos uma possibilidade do uso do "bootstrap" no problema da regressão (Freedman [1981]), por sua associação natural ao problema do estimador de razão. Finalmente explorando estas e outras idéias desenvolvemos alternativas de aplicação do "bootstrap" ao problema da estimação da variância de $\hat{\bar{Y}}_R$.

No capítulo III, apresentamos o estimador "bootstrap" da variância de $\hat{\bar{Y}}_R$ em diversas situações, e no contexto de um problema real típico. Desenvolvemos uma análise comparativa dos estimadores v_0, v_1, v_2 e v_3 (Wu [1982]) entre si e com dois estimadores "bootstrap". Considerando o modelo $Y = r \cdot x + \epsilon$, com $\text{Var}(\epsilon) = k \cdot x^t$, estudamos o efeito de t na eficiência relativa daqueles estimadores. Verifica-se aqui que o método "bootstrap" oferece uma alternativa vantajosa para o problema da estimação da variância de $\hat{\bar{Y}}_R$ para $t = 2$.

Na última parte desse trabalho são apresentados os apêndices I e II com listagens completas dos programas Fortran utilizados, bem como algumas tabelas mais detalhadas do material apresentado no capítulo III.

CAPÍTULO I

ESTIMADORES DE RAZÃO: COMO ESTIMAR SUA VARIÂNCIA

1.1. INTRODUÇÃO

Considere uma população finita com N unidades, às quais está associada uma característica numérica y , com valores y_1, y_2, \dots, y_N . Um dos problemas básicos da teoria da amostragem consiste em estimar $\bar{Y} = \frac{1}{N} \cdot \sum_{i=1}^N y_i$, a média populacional da característica numérica y .

Baseado numa amostragem aleatória simples, sem reposição, a média amostral \bar{y} é um estimador de \bar{Y} , com propriedades estatísticas que podem ser facilmente estabelecidas. É simples provar que

$$E(\bar{y}) = \bar{Y}$$

$$\text{Var}(\bar{y}) = \frac{1}{n} \cdot S^2 \cdot \left(1 - \frac{n}{N}\right) \quad (1.1.1)$$

onde $S^2 = \frac{1}{N-1} \cdot \sum_{i=1}^N (y_i - \bar{Y})^2$, sendo $s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$ um estimador

não viciado de s^2 (Cochran [1977], pág. 26).

Contudo são frequentes situações reais em que existe associado aos indivíduos de uma população, um outro característico numérico x , do qual se conhece a média populacional \bar{X} . Em situações como esta, dependendo da forma de associação entre x e y , alguns estimadores de \bar{Y} podem ser desenvolvidos que oferecem vantagens substanciais em relação a \bar{y} .

Em particular, quando a população puder ser considerada como proveniente de uma superpopulação na qual

$$y = r \cdot x + \varepsilon,$$

onde r é uma constante positiva e ε é uma variável aleatória tal que

$$E(\varepsilon) = 0$$

$$V(\varepsilon) = k \cdot x^t$$

para constantes positivas k e t , verifica-se que o estimador de razão de \bar{Y} , definido por $\hat{\bar{Y}}_R = \hat{R} \cdot \bar{X}$, com $\hat{R} = \frac{\bar{y}}{\bar{x}}$ apresenta, em geral, propriedades estatísticas que o tornam uma alternativa vantajosa à média amostral \bar{y} . Aqui denominamos $R = \bar{Y} \cdot \bar{X}^{-1}$, a razão populacional, relacionada a r , está referente a superpopulação.

Ilustraremos com um exemplo:

Considere as 50 maiores cidades do interior de São Paulo em 1970 com as respectivas populações (em milhares de habitantes) em 1970 (x_i) e 1980 (y_i).

TABELA 1.1.1: As cinquenta maiores cidades do interior paulista, por população em 1970, com respectivas populações em milhares de habitantes nos censos de 1970 e 1980.

CIDADES	Pop. 70	Pop. 80	CIDADES	Pop. 70	Pop. 80
Americana	66	121	Limeira	91	145
Andradina	52	48	Lins	46	60
Araçatuba	108	129	Lorena	46	57
Araraquara	100	127	Mauá	101	204
Araras	53	65	Mogi das Cruzes	139	197
Assis	57	67	Osasco	283	469
Barretos	65	72	Ourinhos	49	60
Bauru	132	196	Pereira Barreto	52	46
Botucatu	52	64	Pindamonhangaba	48	69
Bragança Paulista	63	83	Piracicaba	152	213
Campinas	368	658	Pres. Prudente	106	136
Carapicuíba	55	182	Ribeirão Preto	213	316
Catanduva	58	73	Rio Claro	78	110
Cubatão	58	78	Santo André	419	550
Diadema	79	224	Santos	346	415
Franca	83	147	S.B. dos Campos	201	420
Guaratinguetá	69	84	S. C. do Sul	150	163
Guarujá	94	150	São Carlos	85	119
Guarulhos	237	526	S.J. Rio Preto	121	187
Itapetininga	63	84	S.J. dos Campos	148	284
Itapeva	49	65	São Vicente	116	190
Itu	49	73	Sorocaba	176	266
Jacareí	61	114	Suzano	55	100
Jaú	56	74	Taubaté	110	168
Jundiaí	89	257	Tupã	52	57

Seja o problema de se estimar a população média nestas cidades, no ano de 1980, com uma amostra de, digamos, 10 cidades.

Daremos a seguir, uma comparação Monte Carlo entre o desempenho do estimador de razão $\hat{\bar{Y}}_R$ com o estimador de média simples \bar{y} .

Com base em 500 repetições independentes do processo, obtemos $\hat{\bar{Y}}_R = 175,07$ e $\bar{y} = 172,56$.

O valor real da população média, em 1980, é $\bar{Y} = 175,12$.

Sabe-se que \bar{y} é um estimador não viciado de \bar{Y} , o que não ocorre com $\hat{\bar{Y}}_R$. Os histogramas a seguir, contudo, revelam uma maior precisão de $\hat{\bar{Y}}_R$.

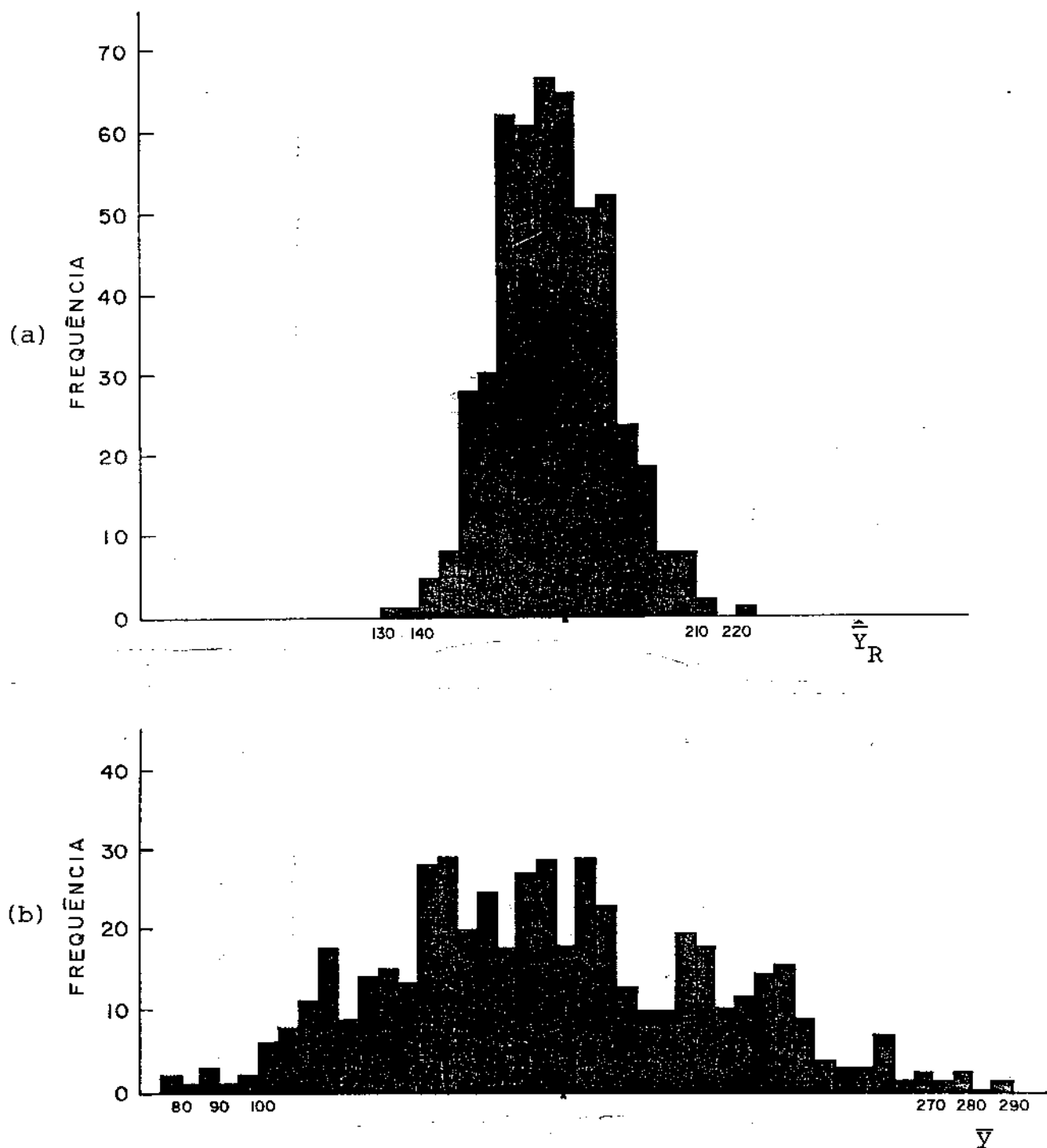


Figura 1.1 - Histograma baseado em 500 estimativas de \bar{Y} .

(a) com estimadores de razão

(b) com estimadores de média simples.

Com base nas 500 repetições obtivemos $\text{Var}(\bar{y}) = 1603,58$ e $\text{Var}(\hat{\bar{Y}}_R) = 228,10$. A distância entre o 5º e o 95º percentil foi de 129,7 para o estimador de média simples, e de 41,9 para o estimador de razão. Evidentemente a utilização do estimador de razão introduziu, neste caso, melhoras substanciais.

O estimador de razão como definido anteriormente, e diversas variantes do mesmo (Tin [1965], Rao [1969], Rao e Rao [1971], Rao e Kuzik [1974]), têm tido amplas aplicações e despertado grande interesse teórico.

Como estimador de \bar{Y} , $\hat{\bar{Y}}_R$ é em geral viciado exceto em algumas situações particulares (Hartley e Ross [1954]). Sua consistência no caso de população finita é óbvia; em situações mais gerais, é facilmente estabelecível (Scott e Wu [1981]).

Ao longo deste trabalho denotaremos, para simplicidade, $V(\hat{\bar{Y}}_R)$ a variância do estimador de razão $\hat{\bar{Y}}_R$. Sua importância decorre de $\hat{\bar{Y}}_R$ ser frequentemente a opção adotada para se estimar \bar{Y} em situações reais. Entretanto apenas expressões aproximadas para este parâmetro são conhecidas. Em (Cochran [1977]) temos uma aproximação geralmente aceita para $V(\hat{\bar{Y}}_R)$, da qual trataremos mais adiante.

A estimação deste parâmetro se constitui ainda num problema sem solução definitiva.

Estimadores alternativos de $V(\hat{\bar{Y}}_R)$ foram desenvolvidos em Cochran [1977], Royall e Eberhard [1975], Royall e Cumberland [1978], Rao [1981] e Wu [1982], entre outros.

O estabelecimento de propriedades estatísticas desses estimadores, bem como o desenvolvimento de estimadores mais adequados a diversas situações se constitui num problema atual de grande interesse na pesquisa estatística.

As dificuldades inerentes à abordagem teórica deste problema têm sido em parte contornadas por estudos de simulação Monte Carlo, visando estabelecer as eficiências relativas de diversos estimadores de $V(\hat{\bar{Y}}_R)$ em uma variedade de situações.

1.2. AS DIFICULDADES DE SE ESTIMAR A VARIÂNCIA.

A procura de uma expressão exata para $V(\hat{\bar{Y}}_R)$ bem como de um estimador adequado para este parâmetro, tem tido muita repercussão dentro dos meios estatísticos, haja visto o número de artigos

publicados sobre o assunto desde os anos 60 até os dias de hoje.

Uma solução teórica aproximada deste problema, encontra-se em (Cochran [1977], pág. 153).

$$V(\hat{\bar{Y}}_R) = \left[\frac{1-f}{n} \right] \cdot \frac{1}{N-1} \cdot \sum_{i=1}^N (y_i - R\bar{x}_i)^2 \quad (1.2.1)$$

onde $f = \frac{n}{N}$ é a fração de amostragem. Na obtenção da expressão acima, \bar{x} é assumido igual a \bar{X} , permitindo um tratamento algébrico simples.

Hoje existe um número considerável de estimadores de $V(\hat{\bar{Y}}_R)$ propostos, com desempenhos relativos dependentes da situação particular em que são aplicados.

Recentemente, uma variedade de artigos têm revelado esforço no sentido de esclarecer qual é o melhor estimador a ser usado em diversas situações reais.

1.3. ESTIMADORES PROPOSTOS.

A estimação de $V(\hat{\bar{Y}}_R)$, em geral tem sido feita segundo os estimadores convencionais (Cochran [1977])

$$v_o = \left[\frac{1-f}{n} \right] \cdot \frac{1}{n-1} \cdot \sum_{i=1}^n (y_i - \hat{R}x_i)^2 \quad (1.3.1)$$

e

$$\begin{aligned} v_2 &= \left[\frac{1-f}{n} \right] \cdot \left[\frac{\bar{X}}{\bar{x}} \right]^2 \cdot \frac{1}{n-1} \cdot \sum_{i=1}^n (y_i - \bar{R}x_i)^2 = \\ &= \left[\frac{\bar{X}}{\bar{x}} \right]^2 \cdot v_o. \end{aligned} \quad (1.3.2)$$

Em (Royall e Eberhard [1975]) apresenta-se um estimador de $V(\hat{\bar{Y}}_R)$

$$v_H = v_o \cdot \frac{\bar{x}_c \cdot \bar{X}}{\bar{x}^2} \cdot \left[1 - \frac{c_x^2}{n} \right]^{-1} \quad (1.3.3)$$

onde:

\bar{x}_c : média das $(N-n)$ unidades que não são da amostra.

c_x^2 : coeficiente de variação de x ,

e que é assintoticamente equivalente a v_2 .

Ainda referente ao estimador v_H , encontra-se um estudo empírico em (Royall e Cumberland [1978]), no qual se conclui que a teoria convencional usada em trabalhos de amostragem não proporciona uma comparação sobre a eficiência relativa de v_0 e v_2 .

Posteriormente, Rao [1981] compara três alternativas de estimadores do parâmetro $V(\hat{\bar{Y}}_R)$. Um dos três estimadores considerados é

$$\hat{v}_c = \left[\frac{1-f}{n} \right] \cdot S_d^2, \quad (1.3.4)$$

onde $d_i = (y_i - \hat{R}x_i)$ e $(n-1) \cdot S_d^2 = \sum_{i=1}^n d_i^2$, que coincide com v_0 .

O outro estimador baseado nos mínimos quadrados é

$$\hat{v}_w = \left[\frac{1-f}{n} \right] \cdot \frac{\bar{X} \cdot \bar{x}_c}{\bar{X}} \cdot \hat{\delta} \quad (1.3.5)$$

onde $(n-1) \cdot \hat{\delta} = \sum_{i=1}^n (d_i^2/x_i)$ e finalmente o estimador "jackknife",

$$\begin{aligned} \hat{v}_R &= \frac{\bar{X}^2}{\bar{X}} \cdot \hat{v}_c \\ &= v_1 \cdot \bar{X}, \end{aligned} \quad (1.3.6)$$

baseado no procedimento sugerido por Quenouille [1956] e Tukey [1958] que abordaremos no capítulo II.

Observa-se que, alguma atenção tem sido devotada ao problema da estimação de $V(\hat{\bar{Y}}_R)$. Contudo, um estudo mais aprofundado encontra-se em (Wu [1982]), no qual é proposta a família de estimadores de $V(\hat{\bar{Y}}_R)$

$$v_g = \left[\frac{\bar{X}}{\bar{x}} \right]^g \cdot v_0, \quad (1.3.7)$$

à qual pertencem v_0 e v_2 , e se estabelece uma comparação entre v_0 , v_1 e v_2 que apresentaremos no parágrafo 1.5.

1.4. PROPRIEDADES ASSINTÓTICAS DE ALGUNS ESTIMADORES.

No sentido de complementar o estudo referente aos estimadores propostos no parágrafo anterior, (Scott e Wu [1981]) apresenta alguns teoremas como resultados preliminares. Com base nesses resultados os autores estabeleceram a normalidade assintótica de $\hat{\bar{Y}}_R$ e, provando que a razão $\frac{\bar{X}}{\bar{x}}$ converge em probabilidade para 1, finalmente concluem que os estimadores v_0 e v_2 são assintoticamente consistentes.

Considerando-se que v_0 é assintoticamente consistente e que a razão $\frac{\bar{X}}{\bar{x}}$ converge em probabilidade para 1, conclui-se pelo

teorema de Slutsky (Bickel e Doksum [1977]) que todos os estimadores da família v_g são assintoticamente consistentes.

1.5. OS CRITÉRIOS DE WU.

Considerando-se as dificuldades que porventura possam surgir na escolha do melhor estimador para $V(\hat{\bar{Y}}_R)$, para situações particulares, encontra-se em (Wu [1982]) uma contribuição importante. Supondo que a população em estudo provém de uma superpopulação que satisfaz o modelo de regressão $y = rx + \varepsilon$, com $E(\varepsilon) = 0$ e $\text{Var}(\varepsilon) = \sigma^2 \cdot x^t$, desenvolve uma análise comparativa entre v_0 , v_1 e v_2 com respeito à variância e ao vício destes estimadores.

Analisando primeiramente as variâncias dos estimadores em estudo constata-se que:

- i) Para $t = 0$, v_0 é o estimador ótimo de $V(\hat{\bar{Y}}_R)$.
- ii) Se $t = 1$, o estimador ótimo é v_1 .
- iii) Para $t \geq 1$ temos que v_1 e v_2 são ambos melhores do que v_0 para estimar $V(\hat{\bar{Y}}_R)$.

Quanto à análise dos vícios destes estimadores, constata-se que:

- i) Para $t=0$, $v_{-\frac{1}{2}}$ é o estimador menos viciado de $V(\hat{\bar{Y}}_R)$.

ii) Para $t=1$, v_2 e v_{-1} são os estimadores menos viciados de $V(\bar{Y}_R)$.

Cabe destacar que a comparação dos estimadores para $t \neq 0$ ou $t \neq 1$ torna-se muito difícil, e faz-se necessário a suposição de uma distribuição para x .

CAPÍTULO II

O MÉTODO "BOOTSTRAP" PARA ESTIMAÇÃO DE DISTRIBUIÇÕES AMOSTRAIS

2.1. O "JACKKNIFE"

O "jackknife" é uma técnica estatística introduzida por Quenouille em 1949 e posteriormente desenvolvida por Tukey [1950], e desde então amplamente investigada por, entre outros, Durbin [1959], Schusany [1971], Miller [1974]. O nome "jackknife" foi atribuído por Tukey, indicando que o método é um instrumento prático, com variadas possibilidades de aplicação na Estatística.

Considere Y_1, Y_2, \dots, Y_n iid $\sim F$, e sejam $\hat{\theta}_n(Y_1, Y_2, \dots, Y_n)$ e $\theta(F)$, respectivamente uma estatística e um funcional em F .

Se $E(\hat{\theta}_n - \theta) = \frac{a_1}{n} + O(\frac{1}{n^2})$, o método "jackknife" permite construir uma outra estatística $\tilde{\theta}_n$ associada a $\hat{\theta}_n$, tal que

$$E(\tilde{\theta}_n - \theta) = O(\frac{1}{n^2}) ,$$

ou seja, o "jackknife" permite eliminar a componente em $\frac{1}{n}$ do vício de $\hat{\theta}_n$ como estimador de θ (F).

Para se construir $\tilde{\theta}_n$, divide-se a amostra em g grupos de tamanho h (com $n = g \cdot h$), e seja $\hat{\theta}_{-i}$, $i=1,2,\dots,g$ a estatística $\hat{\theta}_{n-h}$, aplicada à amostra de $n-h$ observações após a eliminação do i -ésimo grupo. A estes valores Tukey em um trabalho não publicado denominou "pseudo-valores" (Miller [1974]). O estimador "jackknife" $\tilde{\theta}_n$ é então dado por

$$\tilde{\theta}_n = g \cdot \hat{\theta}_n - \frac{(g-1)}{g} \cdot \sum_{i=1}^g \hat{\theta}_{-i} . \quad (2.1.1)$$

O estimador "jackknife", tido como de difícil aplicação, devido a quantidade de cálculo envolvido, tem despertado mais interesse agora, com a ampliação e popularização do uso de computadores rápidos e baratos.

O método "jackknife" foi originalmente concebido com a finalidade de reduzir o vício de estimadores pela eliminação do termo em $\frac{1}{n}$ do mesmo. Tukey [1958], todavia chama a atenção para a possibilidade do método na estimação da distribuição amostral de $\hat{\theta}_n$. Encontra-se em (Miller [1974]) uma aplicação do "jackknife" em estimação de intervalos de confiança para θ , e em outros artigos como (Rao [1981] e Rao e Kuzik [1974]) aplica-se o método na estimação de $V(\hat{Y}_R)$.

2.2. O "BOOTSTRAP" COMO UMA VARIANTE DO "JACKKNIFE"

O "bootstrap" é uma técnica estatística relacionada ao "jackknife", já que como esta, se baseia em reamostragens dentro de uma amostra. Introduzido recentemente (Efron [1979]), tem se mostrado uma ferramenta eficiente e versátil no tratamento de uma variedade de problemas estatísticos.

A partir de Efron [1979] verifica-se um interesse crescente pelo assunto, haja visto a quantidade de artigos recentemente publicados. A idéia básica do "bootstrap" é bastante simples.

Seja X_1, X_2, \dots, X_n , iid $\sim F$. Sejam $\hat{T}_n = \hat{T}_n(X_1, X_2, \dots, X_n)$ estatísticas reais e $T_n(F)$ funcionais reais tais que $\sqrt{n} \cdot (\hat{T}_n - T_n(F))$ tenha distribuição $H_n(x, F)$. Um problema básico em Estatística consiste em estimar $H_n(x, F)$. De fato,

$$E_{H_n(x, F)}(X) \quad \text{e} \quad V_{H_n(x, F)}(X)$$

são respectivamente o vício e a variância de \hat{T}_n (Beran [1981]).

O método "bootstrap" oferece uma opção simples, embora computacionalmente intensiva, de se estimar $H_n(x, F)$. Seja \hat{F} o estimador não paramétrico de máxima verossimilhança de F (isto é, a distribuição que coloca massa igual a n^{-1} em cada ponto amostrado x_i), temos que o estimador por reamostragem (bootstrap) de

$H_n(x, F)$ é dada por $H_n(x, \hat{F})$.

Na prática, por amostragem aleatória simples com reposição, constrõem-se amostras "bootstrap" $X_1^*, X_2^*, \dots, X_n^*$ iid $\sim \hat{F}$, e calcula-se $\hat{T}_n^* = \hat{T}_n(X_1^*, X_2^*, \dots, X_n^*)$. Tal processo é repetido um grande número de vezes, produzindo-se

$$\hat{T}_{n_1}^*, \hat{T}_{n_2}^*, \dots, \hat{T}_{n_m}^* \text{ iid } \sim H_n(x, \hat{F}),$$

tem-se então a distribuição empírica $\hat{H}(x, \hat{F})$ que, sob condições bastante fracas, pode ser feita tão próxima quanto se deseje $H_n(x, \hat{F})$, bastando escolher \underline{m} suficientemente grande.

É intuitivo que se $H_n(x, F)$ depende suavemente de F , o estimador de reamostragem (bootstrap), $H_n(x, \hat{F})$ será fracamente consistente, no sentido de convergir para a mesma distribuição que $H_n(x, F)$.

De fato, Efron [1979] provou a consistência fraca sob diversas suposições, e mais tarde, Bickel e Freedman [1981], provaram que a consistência é, tipicamente, forte.

Encontra-se em (Efron [1979]) e (Diaconis e Efron [1983]), a título de ilustração, um exemplo interessante sobre a estimação de intervalos de confiança para o coeficiente de correlação entre GPA e LSAT médios de calouros nas 82 faculdades de direito nos EUA.

O "bootstrap" pode ser visto como uma generalização do "jackknife" (Rubin [1981]), contudo tem-se mostrado mais versátil e de aplicação mais generalizada. Seu uso tem-se expandido muito mais rapidamente, devido também ao fato de ter sido introduzido numa época em que os computadores alcançam enorme popularidade, ao contrário do "jackknife", introduzido ainda na década de 40 e divulgado por Tukey nos anos 50, quando os computadores eletrônicos eram ainda máquinas de uso restrito e extremamente caras.

Esta popularização do uso do computador explica o crescente interesse nos anos recentes, pelos métodos de reamostragem em geral, e pelo "bootstrap" em particular.

2.3. REGRESSÃO E O "BOOTSTRAP".

Valendo-se da versatilidade do método "bootstrap" abordaremos a aplicação do método na análise de regressão, por sua associação natural ao problema do estimador de razão.

Um estudo detalhado sobre o assunto, encontra-se em (Freedman [1981]).

Seja o modelo

$$\underline{y} = \underline{X} \cdot \underline{\beta} + \underline{\varepsilon} , \quad (2.3.1)$$

com $E(\varepsilon) = 0$ e $V(\varepsilon) = \Sigma$, tipicamente $\varepsilon \sim N_n [0, I_n \cdot \sigma^2]$ e sejam

$$\hat{\underline{y}} = \underline{X} \cdot \hat{\underline{\beta}} \quad \text{e} \quad \underline{e} = \underline{y} - \hat{\underline{y}} \quad (2.3.2)$$

onde $\hat{\underline{\beta}}$ é um estimador de $\underline{\beta}$.

Uma idéia para a aplicação do "bootstrap" à análise de regressão, conforme apresentada por Freedman [1981], consiste em reamostrar sobre as componentes de \underline{e} , criando-se versões "bootstrap" \underline{e}^* . Em seguida reconstitue-se valores de \underline{y} , construindo-se versões \underline{y}^* com

$$\underline{y}^* = \hat{\underline{y}} + \underline{e}^* \quad , \quad (2.3.3)$$

ajusta-se então o modelo a \underline{y}^* , obtendo-se $\hat{\underline{\beta}}^*$.

Contudo algum reajuste nos \underline{e}_i 's é necessário para compensar a redução da variância pela perda de graus de liberdade.

A título de ilustração, trataremos da aplicação da idéia básica acima ao caso clássico de regressão linear simples.

Seja

$$y = \alpha + \beta \cdot x + \varepsilon \quad (2.3.4)$$

com $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2$ e $\tilde{y} = \hat{\alpha} + \hat{\beta} \cdot x$, onde $\hat{\alpha}$ e $\hat{\beta}$ são estimadores de mínimos quadrados de α e β . Então,

$$e_i = y_i - \hat{\beta} \cdot x_i, \quad i=1,2,3,\dots,n.$$

Podemos reamostrar sobre os n valores e_i e construir

$$Y_i^* = \hat{y} + e_i^* . \quad (2.3.5)$$

Porém a reconstituição não seria adequada pois $E(\sum_{i=1}^n e_i^2) = (n-2) \cdot \sigma^2$. Multiplicando cada e_i , por $\sqrt{\frac{n}{n-2}}$, estaríamos restaurando a variabilidade original de ε , compensando a redução da mesma pelo ajustamento e consequente perda de graus de liberdade.

Assim trabalhamos com $\sqrt{\frac{n}{n-2}} \cdot e_i$ e reamostramos para compor e_i^* .

Aplicaremos esta idéia para os dados apresentados abaixo, retirados de Draper e Smith [1981], página 9.

TABELA 2.3.1. Dados relativos a 25 observações das variáveis Y (consumo mensal de vapor, em libras) e X (temperatura atmosférica, em graus Fahrenheit)

i	Valor da Variável	
	Y_i	X_i
1	10,98	35,3
2	11,13	29,7
3	12,51	30,8
4	8,40	58,8
5	9,27	61,4
6	8,73	71,3
7	6,36	74,4
8	8,50	76,7
9	7,82	70,7
10	9,14	57,5
11	8,24	46,4
12	12,19	28,9
13	11,88	28,1
14	9,57	39,1
15	10,94	46,8
16	9,58	48,5
17	10,09	59,3
18	8,11	70,0
19	6,83	70,0
20	8,88	74,5
21	7,68	72,1
22	8,47	58,1
23	8,86	44,6
24	10,36	33,4
25	11,08	28,6

Para os dados da tabela acima, determinamos o ajuste $Y = 13,62 - 0,0798X$ pelo método dos mínimos quadrados. Efetuamos 1000 ajustes "bootstrap" do tipo $Y = \hat{a} + \hat{b} \cdot X$ e obtivemos $\hat{a}(\text{médio}) = 13,63$, $\text{Var}(\hat{a}) = 0,298$, $\hat{b}(\text{médio}) = -0,0799$ e $\text{Var}(\hat{b}) = 9,7 \times 10^{-6}$.

Com base em 10 ajustes "bootstrap" selecionado ao acaso, elaboramos a figura a seguir que tem por finalidade ilustrar a variabilidade associada ao ajuste efetuado pelo método dos mínimos quadrados.

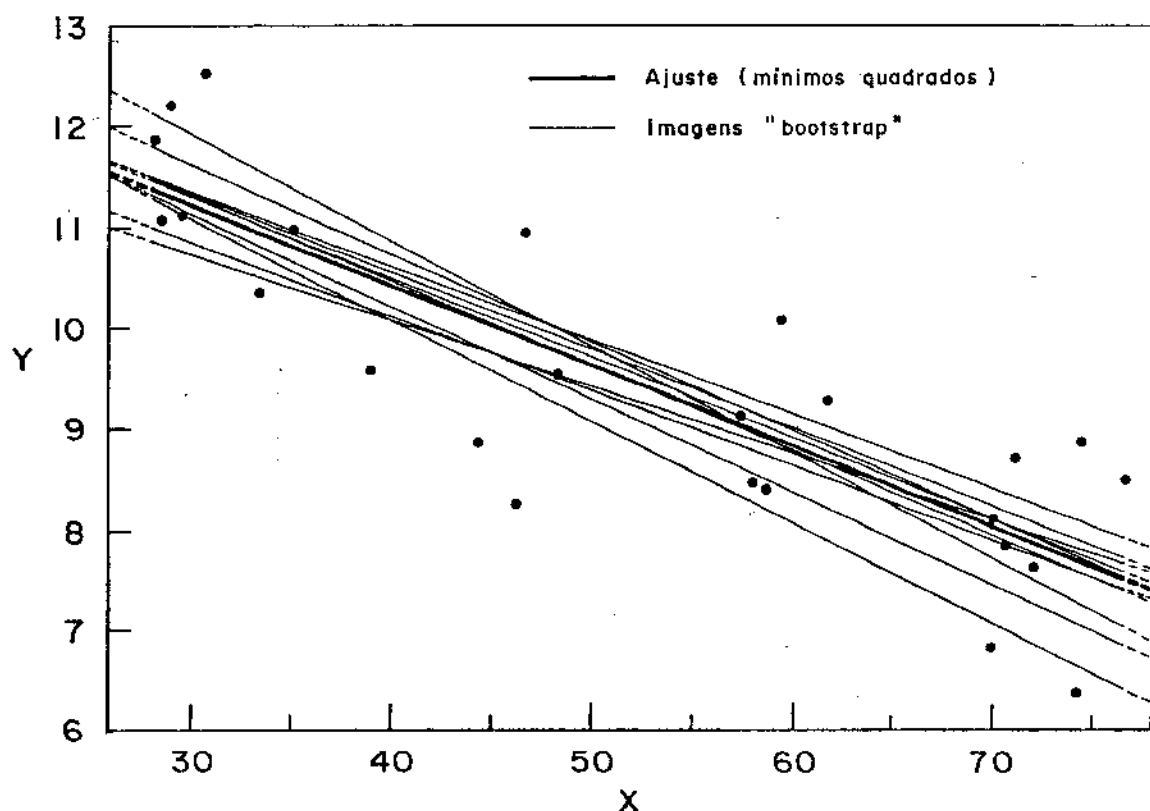


Fig.2.3.1 - Comparação entre o ajuste pelo método dos mínimos quadrados com dez imagens "bootstrap".

A seguir apresentamos uma figura, no sentido de ilustrar a comparação entre a distribuição "bootstrap" dos \hat{b} e a distribuição \underline{t} da teoria normal.

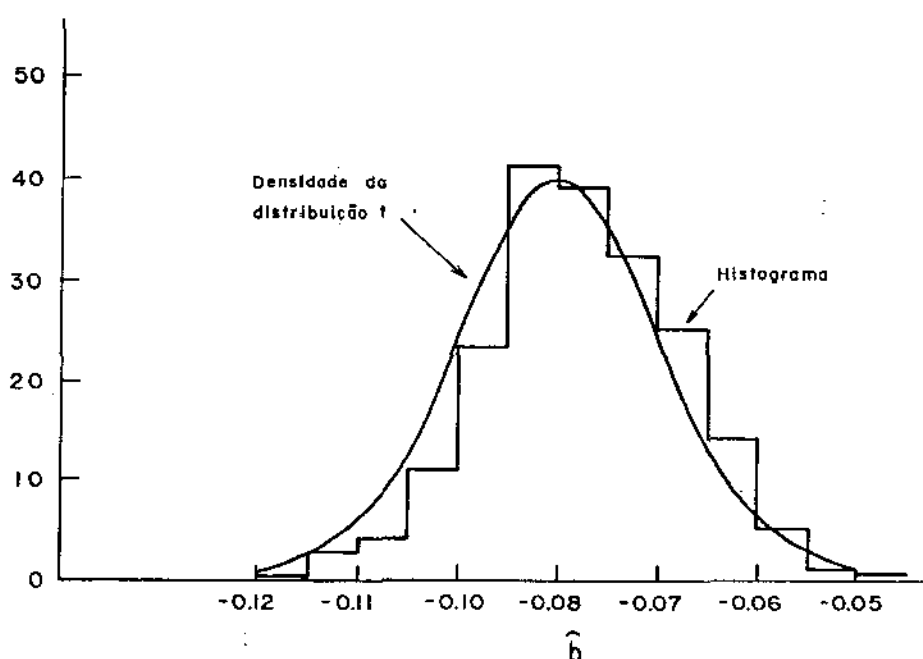


Fig.2.3.2 - Comparação entre a distribuição "bootstrap" dos \hat{b} com a distribuição $\underline{t}(23)$ da teoria normal.

Ressaltamos aqui que a utilização de imagens "bootstrap" oferece uma idéia visual bastante expressiva da variabilidade associada a um ajuste. Pode ser aplicada a situações mais complexas, com dois ou mais preditores.

2.4. POSSÍVEIS USOS DO "BOOTSTRAP" NO PROBLEMA DA ESTIMAÇÃO DA VARIÂNCIA DO ESTIMADOR DE RAZÃO.

Aplicando o "bootstrap" no problema da estimação de $V(\hat{\bar{Y}}_R)$ adotaremos duas abordagens alternativas.

Primeiramente, por meio do processo de reamostragem sobre os resíduos estimados, e_i , e posterior reconstituição dos y_i , discutido na seção anterior, desenvolvemos um estimador de $V(\hat{\bar{Y}}_R)$ do seguinte modo:

Considere uma amostra aleatória simples $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$, $\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$, ..., $\begin{pmatrix} x_n \\ y_n \end{pmatrix}$ extraída de uma população $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$, $\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$, ..., $\begin{pmatrix} x_N \\ y_N \end{pmatrix}$,

suposta proveniente de uma superpopulação, onde

$$y = r \cdot x + \varepsilon, \quad (2.4.1)$$

com $E(\varepsilon) = 0$ e $V(\varepsilon) = k \cdot x^t$.

Temos $\hat{\bar{Y}}_R = \frac{\bar{Y}}{\bar{X}} \cdot \bar{X}$, assim

$$e_i = y_i - \hat{R} \cdot x_i.$$

Reamostrando sobre os e_i 's e reconstituindo y_i 's, temos

$$y_i^* = \hat{y} + e_i^*, \text{ consequentemente}$$

$$\hat{\bar{Y}}_R^* = \frac{\sum_{i=1}^n y_i^*}{\sum_{i=1}^n x_i} \cdot \bar{X} \quad (\text{estimador "bootstrap" de razão}). \quad (2.4.2)$$

Temos então o seguinte resultado.

TEOREMA. Sob o procedimento descrito anteriormente,

$$\text{Var}(\hat{\bar{Y}}_R^* \mid \text{amostra}) = \left[\frac{\bar{X}}{n \cdot \bar{x}} \right]^2 \cdot \sum_{i=1}^n e_i^2.$$

PROVA:

De (2.4.2.), temos que

$$\begin{aligned} \hat{\bar{Y}}_R^* &= \frac{\bar{X}}{n \cdot \bar{x}} \cdot \sum_{i=1}^n y_i^* \\ &= \frac{\bar{X}}{n \cdot \bar{x}} \cdot \sum_{i=1}^n (\hat{R} \cdot x_i + e_i^*) \end{aligned}$$

$$= \frac{\bar{X}}{n \cdot \bar{X}} \cdot \left(\hat{R} \cdot n \cdot \bar{X} + \sum_{i=1}^n e_i^* \right)$$

$$= \hat{R} \cdot \bar{X} + \frac{\bar{X}}{n \cdot \bar{X}} \cdot \sum_{i=1}^n e_i^*$$

Assim,

$$\text{Var}(\hat{Y}_R^* | \text{amostra}) = \text{Var}\left(\hat{R} \cdot \bar{X} + \frac{\bar{X}}{n \cdot \bar{X}} \cdot \sum_{i=1}^n e_i^*\right)$$

$$= \left[\frac{\bar{X}}{n \cdot \bar{X}} \right]^2 \cdot \text{Var}\left(\sum_{i=1}^n e_i^*\right)$$

$$= \left[\frac{\bar{X}}{n \cdot \bar{X}} \right]^2 \cdot \sum_{i=1}^n \text{Var}(e_i^*)$$

$$= \left[\frac{\bar{X}}{n \cdot \bar{X}} \right]^2 \cdot n \cdot \text{Var}(e_i^*)$$

$$= \left[\frac{\bar{X}}{n \cdot \bar{X}} \right]^2 \cdot \sum_{i=1}^n e_i^2$$

Neste trabalho denotaremos

$$v_1^*(\hat{\bar{Y}}_R) = \left[\frac{\bar{X}}{n \cdot \bar{X}} \right]^2 \cdot \sum_{i=1}^n e_i^2 .$$

Uma abordagem alternativa à aplicação do método "bootstrap" na estimação de $V(\hat{\bar{Y}}_R)$ consiste basicamente em reamostrar em cima da própria amostra $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$, $\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$, ..., $\begin{pmatrix} x_n \\ y_n \end{pmatrix}$ e não sobre os resíduos. Aplicando o "bootstrap" à amostra original obtida, teremos amostras "bootstrap" $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}^*$, $\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}^*$, ..., $\begin{pmatrix} x_n \\ y_n \end{pmatrix}^*$ e estimadores "bootstrap" de razão $\hat{\bar{Y}}_R^* = \frac{\bar{Y}^*}{\bar{X}^*} \cdot \bar{X}$. Assim após um número m , suficientemente grande, de repetição deste processo, estimamos $V(\hat{\bar{Y}}_R)$ que denotaremos por

$$v_2^*(\hat{\bar{Y}}_R) = \sum_{i=1}^m (\hat{\bar{Y}}_{R_i}^* - \bar{\hat{\bar{Y}}_R^*})^2 / (m-1) .$$

CAPÍTULO III

O ESTIMADOR "BOOTSTRAP" DA VARIÂNCIA DO ESTIMADOR DE RAZÃO EM DIVERSAS SITUAÇÕES

3.1. UM PROBLEMA TÍPICO

Conforme o exemplo estudado na seção 1.1, no qual considera-se os dados da tabela 1.1.1, vimos que o estimador de razão, $\hat{\bar{Y}}_R$, introduziu melhoras substanciais na estimação de \bar{Y} .

Considerando-se que a população original, provém de uma superpopulação, à qual está associada o modelo de regressão

$$y = r \cdot x + \varepsilon ,$$

onde r é uma constante positiva e os ε_i 's são independentes e identicamente distribuídos, com

$$E(\varepsilon) = 0 \quad \text{e} \quad V(\varepsilon) = \sigma^2 \cdot x^t .$$

Um problema que surge naturalmente, é o de se estimar $V(\hat{\bar{Y}}_R) = E(\hat{\bar{Y}}_R - \bar{Y})^2$. Entretanto, não existe uma expressão exata para

$V(\hat{\bar{Y}}_R)$. Frequentemente emprega-se a fórmula aproximada

$$V(\hat{\bar{Y}}_R) \doteq \left[\frac{1-f}{n} \right] \cdot \left[\frac{\sum_{i=1}^N (y_i - R x_i)^2}{N-1} \right] \quad (3.1.1)$$

introduzida na sessão 1.1, que pode ser escrita em formas alternativas como

$$V(\hat{\bar{Y}}_R) \doteq \left[\frac{1-f}{n} \right] \cdot [S_y^2 + R^2 \cdot S_x^2 - 2 \cdot R \rho S_y S_x] \quad (3.1.2)$$

onde

S_x^2 : variância populacional de x ;

S_y^2 : variância populacional de y ;

ρ : coeficiente de correlação entre x e y .

ou ainda

$$V(\hat{\bar{Y}}_R) \doteq \left[\frac{1-f}{n} \right] \cdot [C_{yy}^2 + C_{xx}^2 - 2 \cdot C_{yx}] \quad (3.1.3)$$

onde

C_{xx} : coeficiente de variação de x

C_{yy} : coeficiente de variação de y

C_{yx} : covariância relativa entre x e y .

A aproximação para $V(\hat{\bar{Y}}_R)$ apresentada acima, em geral oferece bons resultados para amostras não muito pequenas.

A título de ilustração apresentamos uma comparação de $V(\hat{\bar{Y}}_R)$, referente aos dados da Tabela 1.1.1, para diversos tamanhos de amostras, com os valores aproximados dados pela expressão (3.1.1). Os valores verdadeiros de $V(\hat{\bar{Y}}_R)$ foram aproximados por amostragem Monte-Carlo. O programa Fortran que foi utilizado é apresentado no apêndice 1.

TABELA 3.1.1. Valor Monte Carlo e Valor Aproximado de $V(\hat{\bar{Y}}_R)$, para amostras de tamanhos diferentes, com NMC repetições Monte Carlo.

N	NMC	Valor Monte Carlo	Valor Aproximado
5	5000	493,2	461,7
10	2500	215,4	205,2
15	1700	126,6	136,8
20	1250	85,5	76,9
25	1000	54,0	51,3
30	840	35,7	34,2
35	700	22,9	22,0
40	625	14,1	12,9
45	550	6,0	5,7

A figura a seguir ilustra melhor a relação entre $V(\hat{\bar{Y}}_R)$ e a aproximação proposta.

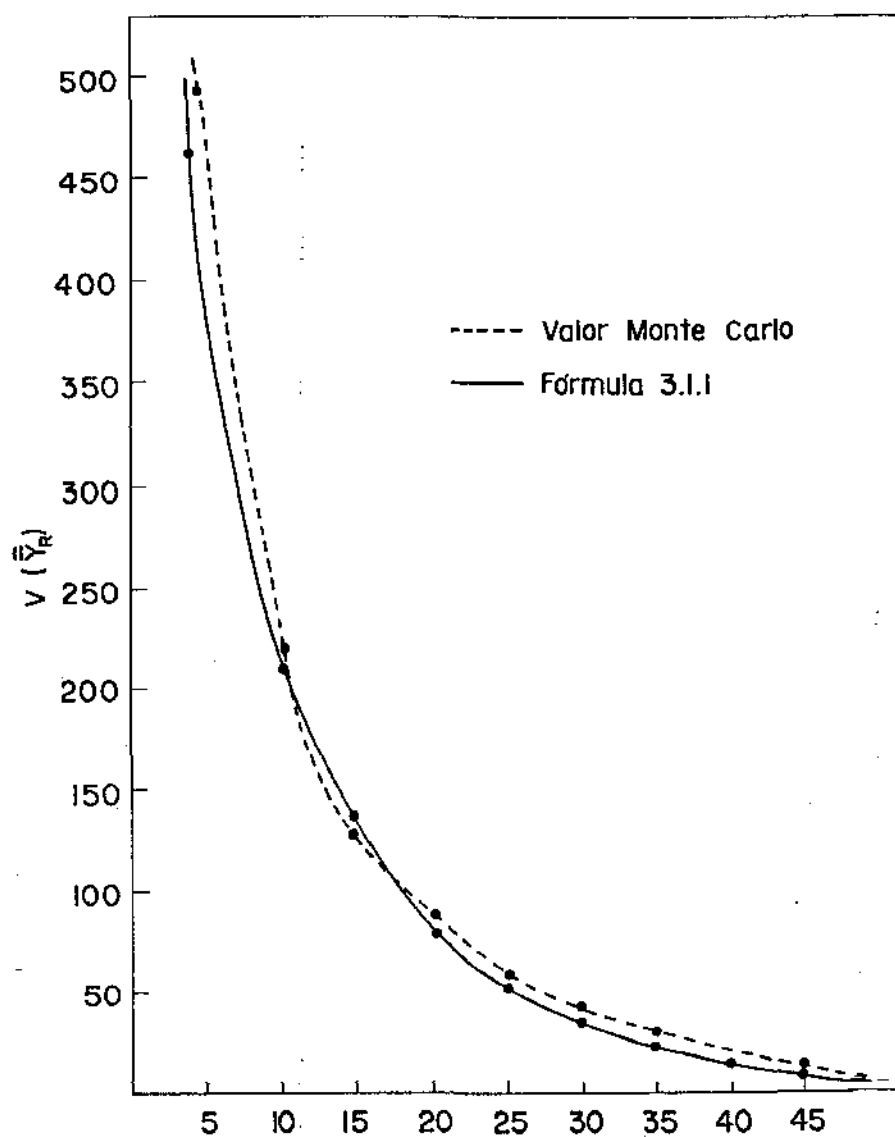


Fig.3.1.1 - Relação entre $V(\hat{\bar{Y}}_R)$, obtida por Monte Carlo, a a aproximação proposta.

No sentido de oferecer uma contribuição ao problema apresentado, desenvolvemos um estudo comparativo entre estimadores da família v_g , para $g = 1, 2$ e 3 , e as duas variantes "bootstrap" que introduzimos no capítulo anterior.

Como a população em estudo é finita, tem-se a necessidade de introduzir o fator de correção, $(1 - \frac{n}{N})$, nas expressões dos estimadores "bootstrap". A comparação entre os quatro estimadores da família v_g e os dois estimadores "bootstrap" que propomos, é feita, via Monte Carlo, para amostras de tamanho 10, 25 e 40.

Os resultados obtidos são apresentados na tabela 3.1.2., que contém informações sobre o vício e a variância, com o objetivo de acompanhar o trabalho desenvolvido em (Wu [1982]). O erro médio quadrático, E.M.Q., será utilizado para determinar o melhor estimador em cada caso.

TABELA 3.1.2. Estimativas obtidas com base em 3 tamanhos de amostras extraídas da população original: N = 10, 25 e 40.

NBS = 1.000 (número de repetições "bootstrap")

NMC = 1.000 (número de repetições Monte Carlo)

		v_0	v_1	v_2	v_3	$v_1^*(\hat{\bar{Y}}_R)$	$v_2^*(\hat{\bar{Y}}_R)$
N = 10	Média	173,8	186,4	213,3	261,3	229,9	188,3
	Vício	-41,56	-29,00	-2,100	45,96	14,56	-27,05
	Variância	13456	11664	18225	48777	30615	11592
	E.M.Q.	15183	12505	18229	50859	30827	12326
N = 25	Média	39,31	49,25	63,24	83,34	79,36	47,86
	Vício	-14,62	-4,000	9,310	29,41	25,43	-6,070
	Variância	216,1	252,8	484,0	1398,8	998,6	349,7
	E.M.Q.	429,8	268,8	570,7	2263,7	1645,2	386,5
N = 40	Média	8,751	12,54	18,24	26,98	26,96	12,44
	Vício	-7,880	-1,54	4,160	12,90	12,88	-1,64
	Variância	6,200	9,000	22,18	83,72	73,61	8,526
	E.M.Q.	68,29	11,37	39,49	250,1	239,5	11,22

A tabela acima fornece os elementos essenciais, como o vício, a variância e o E.M.Q., à análise do desempenho relativo dos diversos estimadores propostos.

Quanto ao vício dos estimadores, observa-se que:

(1) Para $N = 10$, os estimadores v_0 , v_1 , v_2 e $v_2^*(\hat{\bar{Y}}_R)$ subestimam $V(\hat{\bar{Y}}_R)$, enquanto que v_3 e $v_1^*(\hat{\bar{Y}}_R)$ superestimam, sendo que o estimador v_2 apresenta menor vício.

(2) Para $N = 25$, os estimadores v_0 , v_1 e $v_2^*(\hat{\bar{Y}}_R)$ subestimam $V(\hat{\bar{Y}}_R)$, enquanto que v_2 , v_3 e $v_1^*(\hat{\bar{Y}}_R)$ superestimam, sendo que os estimadores menos viciados são v_1 , v_2 e $v_2^*(\hat{\bar{Y}}_R)$.

(3) Para $N = 40$, os estimadores v_0 , v_1 e $v_2^*(\hat{\bar{Y}}_R)$ subestimam $V(\hat{\bar{Y}}_R)$, enquanto que v_2 , v_3 e $v_1^*(\hat{\bar{Y}}_R)$ superestimam e os estimadores menos viciados são v_1 , v_2 e $v_2^*(\hat{\bar{Y}}_R)$.

Quanto à variância, tem-se:

(1) Para $N = 10$, o estimador de menor variância é $v_2^*(\hat{\bar{Y}}_R)$ que apresenta um vício menor que o seu concorrente mais próximo v_1 .

(2) Para $N = 25$ e $N = 40$, o estimador de menor variância é v_0 , sendo que o seu vício é relativamente alto quando comparado com o vício dos seus seguidores mais próximos v_1 e $v_2^*(\hat{\bar{Y}}_R)$, cujos erros médios quadráticos são inferiores aos de v_0 .

Observa-se que, para $N = 10, 25$ e 40 , v_1 e $v_2^*(\hat{\bar{Y}}_R)$ possuem E.M.Q. substancialmente menor que os demais estimadores.

Com esta análise fica evidenciado que os estimadores v_1 e $v_2^*(\hat{\bar{Y}}_R)$ são substancialmente melhores que os demais para estimar $V(\hat{\bar{Y}}_R)$, no exemplo considerado. Procuraremos, a seguir, desenvolver alguns resultados mais gerais.

3.2. O EFEITO DE t NO DESEMPENHO RELATIVO DOS 6 ESTIMADORES DE $V(\hat{\bar{Y}}_R)$.

Com a finalidade de explorar o desempenho dos estimadores "bootstrap", a partir dos dados originais considerados na sessão 1.1, criaremos novos valores de y mantendo fixos os valores de 1.970 e variando os de 1.980, no sentido de ajustar os dados a modelos pré-definidos. Assim, ajustando os dados a modelos do tipo $y = r \cdot x + \varepsilon$, com $\text{Var}(\varepsilon) = \sigma^2 \cdot x^t$. Processaremos as variações para diferentes valores de t , iguais a 0, 1 e 2. O programa Fortran que foi utilizado é apresentado no Apêndice 1.

Tendo em vista a necessidade de garantir que as características dos modelos efetivamente se manifestam nos dados e, assim, nos resultados, criamos 5 conjuntos de dados para cada valor de t . O desempenho médio de cada estimador entre os 5 conjuntos é considerado.

De posse dos 3 grupos populacionais $t = 0, 1$ e 2 , levantamos os valores de $V(\hat{\bar{Y}}_R)$ em função de N , por amostragem Monte Carlo, sendo que os dados obtidos encontram-se na tabela a seguir.

TABELA 3.2.1. Valores médios de $V(\hat{\bar{Y}}_R)$, obtidos dos 5 conjuntos de dados, para diferentes valores de t e N .

NMC : número de repetições Monte Carlo.

	N	NMC	$V(\hat{\bar{Y}}_R)$ Médio
$t = 0$	10	2500	239,6
	25	1000	54,67
	40	625	13,27
$t = 1$	10	2500	216,3
	25	1000	52,34
	40	625	13,58
$t = 2$	10	2500	207,7
	25	1000	54,13
	40	625	12,72

A tabela acima fornece os valores médios de $V(\hat{\bar{Y}}_R)$ para o cálculo do vício dos estimadores $v_0, v_1, v_2, v_3, v_1^*(\hat{\bar{Y}}_R)$ e $v_2^*(\hat{\bar{Y}}_R)$, que serão úteis na análise do desempenho dos mesmos.

Ainda, com base nos grupos populacionais criados, efetuamos a retirada de 3 amostras de tamanho $N = 10, 25$ e 40 de cada conjunto de dados, sendo que para cada uma das amostras obtivemos estimativas de $V(\hat{\bar{Y}}_R)$ gerados pelos estimadores em estudo. Considerando o grande número de populações e amostras envolvidas, apresentaremos a seguir três tabelas com os valores médios obtidos em cada grupo populacional, bem como uma análise do desempenho do conjunto de estimadores para cada tabela.

TABELA 3.2.2. Valores médios obtidos com 5 amostras de 5 populações (para $t = 0$).

NBS = 1.000 (número de repetições "bootstrap")

NMC = 1.000 (número de repetições Monte Carlo)

$t = 0$		v_0	v_1	v_2	v_3	$v_1^*(\hat{\bar{Y}}_R)$	$v_2^*(\hat{\bar{Y}}_R)$
N = 10	Média	196,4	226,1	277,3	368,1	296,3	256,8
	Vício	-43,18	-13,43	37,72	128,6	56,61	13,27
	Variância	7806	12289	37532	97021	47497	23140
	E.M.Q.	9713	12479	39034	114181	50769	23472
N = 25	Média	41,85	53,87	65,08	96,09	87,60	54,12
	Vício	-12,82	-0,806	10,41	41,42	32,92	-0,550
	Variância	123,1	244,2	725,5	2594,2	1282,4	338,2
	E.M.Q.	291,3	250,0	879,7	4326,8	2370,3	340,7
N = 40	Média	8,764	12,61	18,90	28,26	26,44	13,50
	Vício	-4,502	-0,654	5,634	14,99	13,18	0,546
	Variância	3,272	7,447	31,11	130,8	76,64	12,01
	E.M.Q.	24,12	8,363	63,60	356,7	259,6	12,82

Através desta tabela, efetuamos uma análise do desempenho dos estimadores $v_0, v_1, v_2, v_3, v_1^*(\hat{\bar{Y}}_R)$ e $v_2^*(\hat{\bar{Y}}_R)$ em relação ao vício e variância dos mesmos.

Em relação ao vício, observa-se que:

(1) Para $N = 10$, os estimadores v_0 e v_1 subestimam $V(\hat{\bar{Y}}_R)$, enquanto que $v_2, v_3, v_1^*(\hat{\bar{Y}}_R)$ e $v_2^*(\hat{\bar{Y}}_R)$ superestimam, sendo que o estimador $v_2^*(\hat{\bar{Y}}_R)$ é o menos viciado entre eles.

(2) Para $N = 25$, os estimadores v_0, v_1 e $v_2^*(\hat{\bar{Y}}_R)$ subestimam $V(\hat{\bar{Y}}_R)$, enquanto que v_2, v_3 e $v_1^*(\hat{\bar{Y}}_R)$ superestimam e o estimador menos viciado é $v_2^*(\hat{\bar{Y}}_R)$.

(3) Para $N = 40$, os estimadores v_0 e v_1 subestimam $V(\hat{\bar{Y}}_R)$, enquanto que os demais superestimam e o estimador menos viciado é $v_2^*(\hat{\bar{Y}}_R)$.

Em relação à variância, pode-se observar que:

Para $N = 10, 25$ e 40 , o estimador de menor variância é v_0 , satisfazendo os critérios de Wu, entretanto, apresenta um vício relativamente alto quando comparado com o vício de v_1 e $v_2^*(\hat{\bar{Y}}_R)$. O estimador v_1 apresenta o menor E.M.Q. para $N = 25$ e 40 . Assim, para $t = 0$ os estimadores "bootstrap" propostos não devem ser utilizados pois têm desempenho inferior a v_1 .

TABELA 3.2.3. Valores médios obtidos com 5 amostras de 5 populações (para $t = 1$).

NBS = 1.000 (número de repetições "bootstrap")

NMC = 1.000 (número de repetições Monte Carlo)

$t = 1$		v_0	v_1	v_2	v_3	$v_1^*(\hat{\bar{Y}}_R)$	$v_2(\hat{\bar{Y}}_R)$
N = 10	Média	175,8	189,5	218,0	268,1	235,3	195,5
	Vício	-41,10	-27,40	1,060	51,16	18,42	-21,42
	Variância	16706	15267	23072	54140	42318	14784
	E.M.Q.	18499	16089	23145	56881	42768	15292
N = 25	Média	39,96	51,29	65,09	86,03	83,58	51,58
	Vício	-12,37	-4,052	12,75	36,69	31,24	-0,758
	Variância	447,1	325,4	834,4	2127,2	1443,9	347,2
	E.M.Q.	501,4	381,7	1002,0	3273,4	2425,5	349,9
N = 40	Média	8,815	12,65	18,42	27,65	23,78	12,61
	Vício	-4,767	-0,934	4,840	14,07	10,19	-0,712
	Variância	15,49	8,513	39,46	120,8	98,95	13,48
	E.M.Q.	36,57	10,70	63,22	319,7	205,4	14,24

Com base nesta tabela, analisamos o desempenho dos estimadores $v_0, v_1, v_2, v_3, v_1^*(\hat{\bar{Y}}_R)$ e $v_2^*(\hat{\bar{Y}}_R)$, considerando a variância e o vício associado a eles.

Em relação ao vício, constata-se que:

Para $N = 10, 25$ e 40 os estimadores v_0, v_1 e $v_2^*(\hat{\bar{Y}}_R)$ subestimam $V(\hat{\bar{Y}}_R)$, enquanto que os estimadores v_2, v_3 e $v_1^*(\hat{\bar{Y}}_R)$ superestimam, sendo que os estimadores menos viciados são v_1 e $v_2^*(\hat{\bar{Y}}_R)$, exceto o estimador v_2 que é o menos viciado para $N = 10$, satisfazendo os critérios estabelecidos por Wu. Para $N = 20$, todavia, o vício de v_2 é maior, aparentemente contradizendo aquele critério.

Quanto à variância, observa-se que:

Para todos os casos analisados, os estimadores v_0, v_1 e $v_2^*(\hat{\bar{Y}}_R)$ apresentam variâncias menores, sendo que os estimadores v_1 e $v_2^*(\hat{\bar{Y}}_R)$ são os melhores para estimar $V(\hat{\bar{Y}}_R)$. Destacamos aqui a coerência dos resultados obtidos com os critérios estabelecidos por Wu.

Tomando como critério o E.M.Q., a conclusão se mantém, já que o vício tem sido pequeno em relação ao desvio padrão, $\sqrt{V(\hat{\bar{Y}}_R)}$, portanto com contribuição pouco relevante ao E.M.Q.

TABELA 3.2.4. Valores médios obtidos com 5 amostras de 5 populações (para $t = 2$).

NBS = 1.000 (número de repetições "bootstrap")

NMC = 1.000 (número de repetições Monte Carlo)

$t = 2$		V_0	V_1	V_2	V_3	$v_1^*(\hat{\bar{Y}}_R)$	$v_2^*(\hat{\bar{Y}}_R)$
N = 10	Média	165,6	168,1	184,8	216,9	202,8	160,7
	Vício	-45,08	-39,58	-22,86	9,220	-4,920	-47,04
	Variância	16065	12349	113476	27121	27471	7182
	E.M.Q.	18120	14004	14303	28132	27794	9646
N = 25	Média	38,29	46,92	59,98	77,66	76,81	45,83
	Vício	-15,84	-7,210	5,850	23,53	22,68	-8,296
	Variância	526,2	302,2	330,8	1285,7	1250,4	222,0
	E.M.Q.	696,4	351,5	373,2	1872,1	1786,7	302,6
N = 40	Média	8,758	12,46	18,00	26,40	26,81	12,17
	Vício	-3,982	-0,280	5,264	13,66	14,07	-0,572
	Variância	25,88	9,759	12,12	81,81	87,17	8,201
	E.M.Q.	28,91	10,39	43,70	268,7	285,4	8,649

Por meio desta tabela, efetuamos uma análise do desempenho dos estimadores $v_0, v_1, v_2, v_3, v_1^*(\hat{\bar{Y}}_R)$ e $v_2^*(\hat{\bar{Y}}_R)$, através do vício e variância dos mesmos.

Em relação ao vício, observa-se que:

(1) Para $N = 10$, todos os estimadores subestimam $V(\hat{\bar{Y}}_R)$, exceto v_3 que superestima e o menos viciado é $v_1^*(\hat{\bar{Y}}_R)$.

(2) Para $N = 25$, os estimadores v_0, v_1 e $v_2^*(\hat{\bar{Y}}_R)$ subestimam $V(\hat{\bar{Y}}_R)$, enquanto que v_2, v_3 e $v_1^*(\hat{\bar{Y}}_R)$ superestimam, sendo que os menos viciados são v_1, v_2 e $v_2^*(\hat{\bar{Y}}_R)$, satisfazendo os critérios de Wu.

(3) Para $N = 40$, os estimadores v_0, v_1 e $v_2^*(\hat{\bar{Y}}_R)$ subestimam $V(\hat{\bar{Y}}_R)$ enquanto que v_2, v_3 e $v_1^*(\hat{\bar{Y}}_R)$ superestimam, sendo que os menos viciados são v_1 e $v_2^*(\hat{\bar{Y}}_R)$, satisfazendo os critérios de Wu.

Quanto à variância, para todos os casos, observa-se que $v_2^*(\hat{\bar{Y}}_R)$ apresenta um desempenho substancialmente melhor que os demais. Com relação a v_1 , o estimador que apresentou o segundo melhor desempenho, $v_2^*(\hat{\bar{Y}}_R)$ apresentou E.M.Q. 32, 14 e 17 por cento menor, para $N = 10, 25$ e 40 respectivamente. O estimador $v_2^*(\hat{\bar{Y}}_R)$ apresenta o menor E.M.Q. para todos os casos. Assim, para $t = 2$, $v_2^*(\hat{\bar{Y}}_R)$ é o melhor estimador de $V(\hat{\bar{Y}}_R)$.

3.3. SITUAÇÕES EM QUE O "BOOTSTRAP" OFERECE MELHORES SOLUÇÕES.

Considerando a análise efetuada na sessão anterior, envolvendo diferentes valores de t , observa-se que para $t=2$, ou seja, $\text{Var}(Y|X) \propto X^2$, o estimador "bootstrap" oferece solução vantajosa comparado com outros estimadores de uso frequente.

Abordaremos as situações em que o estimador "bootstrap" oferece melhores soluções, considerando-se a variância, o vício e o E.M.Q. associado a ele:

(i) Para $t = 0$, os estimadores v_0 e v_1 possuem menor variância. Contudo $v_2^*(\hat{\bar{Y}}_R)$ parece apresentar menor vício, sendo que o estimador v_1 apresenta o menor E.M.Q. para $N = 25$ e 40 .

(ii) Para $t = 1$, os estimadores v_1 e $v_2^*(\hat{\bar{Y}}_R)$ apresentam menor variância e menor E.M.Q., sendo que o menos viciado é $v_2^*(\hat{\bar{Y}}_R)$.

(iii) Para $t = 2$, o estimador $v_2^*(\hat{\bar{Y}}_R)$ é o estimador ótimo para estimar $V(\hat{\bar{Y}}_R)$, entre os 6 propostos.

No sentido de investigar ainda mais a eficiência do estimador "bootstrap", exploraremos na sessão seguinte uma situação envolvendo 400 pares (X,Y) de tal modo que X tem distribuição gama e Y é tal que $E(Y|X) = X$ e $V(Y|X) \propto X^2$.

3.4. UM ESTUDO MONTE CARLO DO DESEMPENHO RELATIVO DOS 5 ESTIMADORES DE $V(\hat{\bar{Y}}_R)$ PARA $t = 2$ e X PROVENIENTE DE UMA DISTRIBUIÇÃO GAMA.

Com o objetivo de investigar o desempenho do conjunto de estimadores v_0, v_1, v_2, v_3 e $v_2^*(\hat{\bar{Y}}_R)$ em uma situação, onde X e Y apresentam certas características especiais. Criamos uma situação, envolvendo 400 pares (X, Y) tais que:

$$X \sim \Gamma(5, 20) \quad \text{com}$$

$$E(X) = 100, \quad V(X) = 2000$$

e Y foi construído segundo o modelo

$$Y = R \cdot X + c_1 \cdot X^{c_2} \cdot Z$$

com $R = 1,0$, $c_1 = 0,1$, $c_2 = 1,0$ e $Z \sim N(0,1)$. Assim

$$E(Y|X) = X \quad \text{e} \quad V(Y|X) = 0,01 X^2.$$

Para a população ($N = 400$), temos que

$$\bar{X} = 102,180$$

$$\bar{Y} = 102,589$$

$$R = 1,004$$

$$V(X) = 2180,000$$

$$V(Y) = 2148,140$$

$$\text{corr}(x,y) = 0,9665$$

Com base em 10.000 amostras de tamanho 10, sem reposição, obtivemos $\bar{\hat{Y}}_R = 102,670$ e $\bar{\bar{Y}} = 102,827$, sendo que $\text{Var}(\bar{\hat{Y}}_R) = 13,79$ e $\text{Var}(\bar{\bar{Y}}) = 210,15$. Neste caso, observa-se que o estimador de razão $\bar{\hat{Y}}_R$ oferece uma grande vantagem na estimação de \bar{Y} quando comparado com o estimador amostral \bar{y} . Entretanto, o nosso objetivo é estimar $V(\bar{\hat{Y}}_R)$.

No sentido de dar uma solução para o problema criado, desenvolvemos um estudo comparativo entre os estimadores v_0, v_1, v_2, v_3 e o estimador $v_2^*(\bar{\hat{Y}}_R)$ que apresentou melhor desempenho nos estudos realizados anteriormente.

Através de simulação amostral Monte Carlo, calculamos $V(\bar{\hat{Y}}_R)$ para amostras de diversos tamanhos, sendo que os resultados obtidos encontram-se na tabela a seguir.

TABELA 3.4.1. Valor Monte Carlo de $V(\bar{\hat{Y}}_R)$, para amostras de tamanhos diferentes, com NMC repetições Monte Carlo.

N	NMC	$V(\bar{\hat{Y}}_R)$ Monte Carlo
10	2500	13,05
25	1000	5,60
40	550	2,94

Retirando amostras de tamanhos $N = 10, 25$ e 40 da população constituída pelos 400 pares (X,Y) e aplicando os estimadores $v_o, v_1, v_2, v_3, v_2^*(\hat{\bar{Y}}_R)$, obtivemos as estimativas computadas na tabela a seguir.

TABELA 3.4.2. Estimativas obtidas com base em 3 tamanhos de amostra extraídas dos 400 pares (X,Y) .

NBS = 400 (número de repetições "bootstrap")

NMC = 80 (número de repetições Monte Carlo)

		v_o	v_1	v_2	v_3	$v_2^*(\hat{\bar{Y}}_R)$
$N = 10$	Média	10,81	10,81	11,02	11,45	10,11
	Vício	-2,24	-2,24	-2,03	-1,60	-2,94
	Variância	109,9	77,54	60,83	50,05	43,51
	E.M.Q.	114,9	82,55	64,95	52,61	52,15
$N = 25$	Média	4,698	4,711	4,763	4,855	4,908
	Vício	-0,902	-0,889	-0,837	-0,745	-0,692
	Variância	5,997	4,683	3,388	3,610	3,453
	E.M.Q.	6,715	5,473	4,088	4,165	3,932
$N = 40$	Média	3,152	0,187	0,237	0,183	0,129
	Vício	0,212	0,087	0,237	0,183	0,129
	Variância	2,548	2,231	2,053	2,009	1,946
	E.M.Q.	2,593	2,266	2,109	2,042	1,963

Analisando a tabela acima, observa-se que o estimador $v_2^*(\hat{\bar{Y}}_R)$ é o menos viciado, exceto para $N = 10$ que apresenta um vício maior que os demais, entretanto, é o estimador que possui menor erro médio quadrático.

Constata-se ainda, que o estimador $v_2^*(\hat{\bar{Y}}_R)$ é aquele que apresenta menor variância entre os estimadores utilizados. Assim, conclui-se que $v_2^*(\hat{\bar{Y}}_R)$ é o estimador ótimo para estimar $V(\hat{\bar{Y}}_R)$, entre os 5 considerados, embora com performance bem próxima à de v_3 .

O "bootstrap" é uma ferramenta estatística de enorme versatilidade, e suas amplas possibilidades na estatística apenas começam a ser exploradas. Em nosso caso, verificamos que o "bootstrap" oferece vantagens substanciais na estimação de $V(\hat{\bar{Y}}_R)$, um problema importante atualmente, numa situação de frequente ocorrência na prática, ou seja, quando $E(Y|X) = R \cdot x$ e $V(Y|X) \propto X^2$, isto é, o desvio relativo de Y é constante.

APÊNDICES

1. PROGRAMAS COMPUTACIONAIS.

Apresentamos neste apêndice as listagens completas dos programas FORTRAN utilizados no PDP-10 da UNICAMP e uma específica — ção do gerador de números aleatórios.

O programa (MCarlo-FOR) possibilita a obtenção dos valores de $V(\hat{\bar{Y}}_R)$ por amostragem Monte Carlo para diferentes tamanhos de amostras retiradas da população em estudo.

O programa (PROG5.FOR) tem por finalidade, obter as estimativas de $V(\hat{\bar{Y}}_R)$ através dos estimadores $v_0, v_1, v_2, v_3, VBS1 = v_1^*(\hat{\bar{Y}}_R)$ e $VBS2 = v_2^*(\hat{\bar{Y}}_R)$.

Os programas estão dimensionados para populações de tamanho no máximo 50. Nos casos em que se deseje um estudo com populações de tamanho superior a 50, basta alterar as instruções DIMENSION.

Nos programas citados, utilizamos o gerador de números aleatórios da sub-rotina RAN do computador PDP-10 da UNICAMP.

```

C
C *****
C *****  P R O G R A M A   M C A R L O . F O R   *****
C *****
C
C DIMENSION X(50),Y(50)
C INTEGER FLAG(50)
C
C *****
C *****  LEITURA DOS DADOS *****
C *****
C
C READ(3,10) (X(I),Y(I), I=1,50)
10  FORMAT(4X,2F5.0)
C
C *****
C *****  CALCULANDO XBARRA E IMPRIMINDO OS DADOS *****
C *****
C
C SOMAX=0.0
C DO 20 I=1,50
C     SOMAX=SOMAX+X(I)
C     TYPE 15, X(I),Y(I)
15  FORMAT(10X,2F7.0)
20  CONTINUE
C     XBP=SOMAX/50.0
C
C *****
C *****  ENTRANDO N, NMC E A SEMENTE *****
C *****
C
C 25  CONTINUE
C     TYPE 30
C 30  FORMAT('  ENTRE N , NMC E SEMENTE ')
C     ACCEPT 40, N, NMC,M
C 40  FORMAT(3I)
C
C *****
C *****  ZERANDO OS ACUMULADORES GLOBAIS *****
C *****
C
C SYR=0.0
C SYR2=0.00
C CALL SETRAN(M)
C
C *****
C *****  INICIANDO O PROCESSO MONTE CARLO *****
C *****

```

```

DD 100 IMC=1,NMC

*****
***** ZERANDO OS ACUMULADORES DA AMOSTRA *****
*****

DO 50 I=1,50
  FLAG(I)=0
  CONTINUE
50  SXA=0.0
  SYA=0.0

*****
***** RETIRANDO UMA AMOSTRA ALEATORIA SIMPLES SEM REPOSICAO *****
*****

DO 60 IN=1,N
  K=INT(50*PAN(M))+1
  IF (FLAG(K).EQ.1) GO TO 55
  FLAG(K)=1
  SXA=SXA+X(K)
  SYA=SYA+Y(K)
  YB=SYA/10.0
60  CONTINUE

*****
***** CALCULANDO YR E ACUMULANDO ESTES VALORES *****
*****

  YR=(SYA/SXA)*XBP
  SYR=SYR+YR
  SYR2=SYR2+YR**2
100 CONTINUE

*****
***** CALCULANDO A VARIANCIA DE YR *****
*****

FNMC=FLDAT(NMC)
YRB=SYR/FNMC
VYR=(SYR2-FNMC*YRB**2)/(FNMC-1.0)

*****
***** IMPRIMINDO OS RESULTADOS *****
*****

TYPE 110,N,NMC,YRB,VYR
110 FORMAT(10X,'N=',I3,'/',10X,'NMC=',I6,'/',10X,'YRB=',F10.4,
  /,10X,'VYR=',F10.4,/)
GO TO 25
STOP
END

```

```

C *****
C ***** PROGRAMA PROCE.FOR *****
C *****
C
C DIMENSION Y(50),X(50),XA(50),YA(50),E(50)
C INTEGER FLAG(100)
C
C *****
C ***** LEITURA DOS DADOS *****
C *****
C
10 READ(3,10) (X(I),Y(I),I=1,50)
   FORMAT(4X,F10.2,10X,F10.2)
   NP=50
   FNP=FLDAT(NP)
   SOMAX=0.0
   SOMAY=0.0
   SOMAX2=0.0
   SOMAY2=0.0
   SOMAXY=0.0
C
C *****
C ***** CALCULANDO OS PARAMETROS POPULACIONAIS *****
C *****
C
20 DO 20 I=1,NP
   SOMAX=SOMAX+X(I)
   SOMAY=SOMAY+Y(I)
   SOMAX2=SOMAX2+X(I)**2
   SOMAY2=SOMAY2+Y(I)**2
   SOMAXY=SOMAXY+X(I)*Y(I)
   CONTINUE
   XBP=SOMAX/FNP
   YBP=SOMAY/FNP
   S2XP=(SOMAX2-FNP*XBP**2)/(FNP-1.0)
   S2YP=(SOMAY2-FNP*YBP**2)/(FNP-1.0)
   COVXY=(SOMAXY-FNP*XBP*YBP)/(FNP-1.0)
   CORRXY=COVXY/SORT(S2XP*S2YP)
   RP=SOMAY/SOMAX
30 TYPE 30,XBP,YBP,S2XP,S2YP,CORRXY,RP
   FORMAT(10X,'X MEDIO NA POP=',F10.4,/,
*       10X,'Y MEDIO NA POP=',F10.4,/,
*       10X,'S2X NA POP=',F10.4,/,
*       10X,'S2Y NA POP=',F10.4,/,
*       10X,'CORRXY NA POP=',F10.4,/,
*       10X,'R NA POP=',F10.4,/)
C
C *****
C ***** ENTRANDO OS ELEMENTOS DA SIMULACAO *****
C *****

```

```

      TYPE 40
40    FORMAT(//, ' QUAL O TAMANHO DA AMOSTRA? ')
      ACCEPT 50, NA
50    FORMAT(I)
      FNA=FLOAT(NA)
      TYPE 60
60    FORMAT(//, ' QUANTAS REPETICOES BODISTRAP? ')
      ACCEPT 62, NBS
62    FORMAT(I)

      FNBS=FLOAT(NBS)
      TYPE 64
64    FORMAT(//, ' QUANTAS REPETICOES MONTE-CARLO? ')
      ACCEPT 70, NMC
70    FORMAT(I)
      TYPE 80
80    FORMAT( ' ENTRE A SEMENTE ' )
      ACCEPT 90, SEED
90    FORMAT(G)
      CALL SETRAN(SEED)

C      *****
C      ***** INICIO DO PROCESSO MONTE CARLO *****
C      *****
C

105   DO 200 IMC=1, NMC
      DO 110 K=1, NP
          FLAG(K)=0
110   CONTINUE

      SXA=0
      SYA=0
      SY2A=0
      SX2A=0
      SXYA=0
      DO 120 I=1, NP
          U=RAN(SEED)
          K=INT(FNP*U)+1
          IF (FLAG(K).EQ.1) GOTO 120
          XA(I)=X(K)
          YA(I)=Y(K)
          FLAG(K)=1
          SXA=SXA+XA(I)
          SYA=SYA+YA(I)
          SY2A=SY2A+YA(I)**2
          SX2A=SX2A+XA(I)**2
          SXYA=SXYA+XA(I)*YA(I)
120   CONTINUE

      RA=SYA/SXA
      XBA=SXA/FNA
      YR=RA*XBP

```

```

RX=XBP/XBA
SRAXA=0.0
DO 130 I=1, NA
  SRAXA=SRAXA+RA*XA(I)
130  CONTINUE
V0=((1.0-FNA/FNP)/FNA)*(SY2A-2*RA*SYVA+SX2A*RA**2)/(FNA-1.0)
V1=V0*RX
V2=V1*RX
V3=V2*RX
SE2=0.0
DO 150 I=1, NA
  E(I)=YA(I)-RA*XA(I)
  SE2=SE2+E(I)**2
150  CONTINUE
VBS1=(RX/FNA)**2*SE2

*****
***** RETIRANDO AS NBS AMOSTRAS BOOTSTRAP *****
*****

S2YBS2=0.0
S2YBS=0.0
SYRBS=0.0
SYRBS2=0.0
***** INICIO DO BOOTSTRAP *****
DO 160 IBS=1,NBS
  SXBS=0.0
  SYBS=0.0
  SEBS=0.0
  DO 160 I=1,NA
    U=RAN(SEED)
    K=INT(FNA*U)+1
    SEBS=SEBS+E(K)
    SXBS=SXBS+XA(K)
    SYBS=SYBS+YA(K)
160  CONTINUE
  YBS1=((SRAXA+SEBS)/SXA)*XRP
  YBS2=(SYBS/SXBS)*XBP
  SYRBS=SYRBS+YBS1
  SYRBS2=SYRBS2+YBS1**2
  S2YBS=S2YBS+YBS2
  S2YBS2=S2YBS2+YBS2**2
180 CONTINUE
YRBSB=SYRBS/FNBS
Y2RBSB=S2YBS/FNBS
VBS2=(S2YBS2-FNBS*Y2RBSB**2)/(FNBS-1.0)
WRITE(20,190)YR,YRBSB,Y2RBSB
190 FORMAT(3F8.2)

```

0
 0
 0
 0

 ***** IMPRIMINDO OS RESULTADOS *****

```

19_  WRITE(4,192)V0,V1,V2
      FORMAT(24X,3F8.2)
194  WRITE(5,194)V3,V002,V0S1
      FORMAT(18X,3F8.2)
200  CONTINUE
      TYPE 210
210  FORMAT('***** ENTRE IND=1 P/ CONTINUAR OU IND=0 P/PARAR***')
      ACCEPT 220, IND
220  FORMAT(I)
      IF (IND.EQ.1) GOTO 105
      STOP
      END
  
```


2. MATERIAL PROCESSADO.

Os resultados das impressões do material processado pelo computador foram organizados em forma de tabelas, para diversos tamanhos de amostras, dependendo da necessidade de cada estudo.

Apresentamos primeiramente as tabelas que fornecem os valores de $V(\hat{\bar{Y}}_R)$ obtidos por amostragem Monte Carlo, mediante o programa (MCarlo.FOR) introduzido no apêndice 1, para os três grupos populacionais com amostras de diversos tamanhos.

Finalmente, apresentamos um conjunto de tabelas que fornecem as estimativas de $V(\hat{\bar{Y}}_R)$ obtidos pelos estimadores $v_0, v_1, v_2, v_3, v_1^*(\hat{\bar{Y}}_R)$ e $v_2^*(\hat{\bar{Y}}_R)$, mediante o programa (PROG5.FOR) introduzido no apêndice 1, para amostras de tamanhos $N = 10, 25$ e 40 extraídas de cada conjunto de dados dos grupos populacionais.

TABELA 2.1. Valores de $V(\bar{Y}_R)$ para as 5 populações geradas com $t=0$, para amostras de diversos tamanhos.

NMC: número de repetições Monte Carlo.

N	NMC	Pop.1	Pop.2	Pop.3	Pop.4	Pop.5
5	5000	605,7	567,9	671,3	618,9	650,8
10	2500	228,8	228,2	261,6	237,8	241,4
15	1700	128,4	129,5	138,9	127,1	134,6
20	1250	81,25	85,28	87,59	82,53	83,15
25	1000	58,06	54,76	54,28	54,24	52,02
30	840	34,47	35,90	37,23	31,34	37,17
35	700	22,00	24,63	19,81	21,25	24,90
40	625	11,77	14,40	13,14	13,95	13,07
45	550	6,739	6,811	5,811	5,776	5,506

TABELA 2.2. Valores de $V(\hat{\bar{Y}}_R)$ para as 5 populações geradas com $t=1$, para amostras de diversos tamanhos.

NMC: número de repetições Monte Carlo.

N	NMC	Pop.6	Pop.7	Pop.8	Pop.9	Pop.10
5	5000	476,1	489,1	499,1	502,2	540,3
10	2500	207,5	205,7	231,0	211,7	228,6
15	1700	120,6	127,6	134,0	123,2	132,8
20	1250	77,38	80,66	79,42	86,10	78,40
25	1000	50,29	53,20	54,33	52,15	51,72
30	840	32,13	37,44	35,76	36,62	33,18
35	700	23,83	24,39	23,07	22,63	20,98
40	625	12,60	13,91	14,16	13,72	13,52
45	550	6,690	5,452	5,682	5,444	5,888

TABELA 2.3. Valores de $V(\hat{\bar{Y}}_R)$ para as 5 populações geradas com $t=2$, para amostras de diversos tamanhos.

NMC: número de repetições Monte Carlo.

N	NMC	Pop.11	Pop.12	Pop.13	Pop.14	Pop.15
5	5000	447,6	341,9	506,4	455,4	427,3
10	2500	215,6	190,0	215,2	215,7	202,0
15	1700	132,7	112,2	130,4	124,7	123,8
20	1250	80,77	76,46	83,34	85,51	84,27
25	1000	53,10	52,88	53,67	58,03	52,96
30	840	35,05	38,22	35,77	38,61	34,46
35	700	22,70	21,91	23,65	21,55	21,81
40	625	12,94	11,90	13,73	12,80	12,23
45	550	6,631	5,581	6,616	6,333	4,959

TABELA 2.4. Variância e vício dos estimadores v_0 , v_1 , v_2 , v_3 , $v_1^*(\hat{Y}_R)$ e $v_2^*(\hat{Y}_R)$ para 15 amostras de diferentes tamanhos, retiradas dos 5 conjuntos de dados gerados para $t = 0$.

NBS = 1000 (número de repetições "bootstrap")

NMC = 1000 (número de repetições Monte Carlo)

	v_0		v_1		v_2		v_3		$v_1^*(\hat{Y}_R)$		$v_2^*(\hat{Y}_R)$	
	Var.	Vício	Var.	Vício	Var.	Vício	Var.	Vício	Var.	Vício	Var.	Vício
N=10	9960	-38,00	14884	-9,700	34969	37,90	103684	115,3	43597	48,70	21667	14,40
	9623	-37,00	11449	-16,80	25281	21,00	78400	85,60	40000	53,30	15576	13,20
	6512	-54,72	14161	-17,44	45796	44,62	103216	146,57	51619	50,05	28764	14,67
	6648	-42,90	8630	-12,20	41616	39,70	101044	154,4	47703	71,20	25856	11,20
	6288	-43,30	12321	-11,00	40000	45,40	98764	140,9	54568	59,80	23839	12,90
N=25	153,8	-16,11	252,8	-4,850	734,4	4,200	2570,5	34,58	1376,4	29,26	372,5	-2,510
	116,6	-13,62	263,7	-1,750	683,3	10,38	2526,1	41,51	1279,5	32,86	329,1	-1,320
	104,0	-11,84	252,8	0,530	870,3	9,200	2624,5	39,92	1268,8	33,38	327,8	-1,090
	144,0	-12,49	264,1	0,940	819,1	10,51	2809,8	44,94	1269,1	33,93	333,8	0,360
	97,22	-10,04	187,7	1,100	655,4	17,76	2440,4	46,16	1218,0	35,19	327,6	1,810
N=40	3,920	-3,616	7,344	-0,590	29,59	6,100	125,4	14,39	71,57	15,71	10,63	0,800
	3,803	-5,535	7,182	-1,570	26,42	4,460	118,8	13,79	73,62	12,82	10,24	0,970
	2,756	-4,054	8,123	0,130	36,84	6,540	126,4	16,54	88,36	14,94	15,05	0,450
	3,386	-5,279	8,545	-1,320	34,93	4,730	151,3	14,16	76,04	7,390	12,82	-0,060
	2,496	-4,026	6,003	0,080	27,77	6,340	132,3	16,07	73,62	15,02	11,29	0,570

TABELA 2.5. Variância e vício dos estimadores $v_0, v_1, v_2, v_3, v_1^*(\hat{Y}_R)$ e $v_2^*(\hat{Y}_R)$ para 15 amostras de diferentes tamanhos, retiradas dos 5 conjuntos de dados gerados para $t = 2$.

NBS = 1000 (número de repetições "bootstrap")

NMC = 1000 (número de repetições Monte Carlo)

v_0			v_1		v_2		v_3		$v_1^*(\bar{Y}_R)$		$v_2^*(\bar{Y}_R)$			
Var.		Vício	Var.		Vício		Var.		Vício		Var.		Vício	
N=10	20164	-48,30	16129	-44,00	19600	-27,60	37249	4,500	32978	-9,700	10983	-54,10		
	20736	-52,20	14461	-56,10	11449	-54,20	11236	-45,50	24586	-36,50	6658	-74,30		
	11449	-44,80	9880	-31,40	13456	-3,700	39204	45,50	23594	10,30	6196	-28,50		
	15876	-41,60	13456	-33,70	15376	-13,10	32041	24,50	27159	2,100	7744	-39,60		
	12100	-38,50	7637	-32,70	7499	-15,60	15876	17,10	29036	9,200	4329	-38,70		
N=25	707,6	-15,51	364,8	-6,920	445,2	5,020	1536,6	21,99	1421,3	20,49	309,8	-8,910		
	729,0	-16,75	510,8	-9,130	547,6	1,360	1260,3	16,05	1709,8	21,15	292,9	-11,81		
	404,0	-14,19	182,3	-3,740	196,0	11,10	1361,6	32,67	1082,4	29,55	183,6	-3,240		
	466,6	-18,89	225,0	-11,76	240,3	3,300	1474,6	20,13	1132,3	16,11	204,0	-11,95		
	324,0	-13,86	228,0	-4,500	225,0	8,470	795,2	26,80	906,0	26,09	119,9	-5,570		
N=40	38,81	-4,199	13,73	-0,500	16,89	5,020	116,6	13,37	105,7	13,50	11,70	-0,860		
	31,02	-3,212	12,82	0,290	18,49	5,440	68,39	13,12	110,3	14,83	10,24	-0,360		
	19,18	-4,833	6,973	-1,000	7,080	4,760	78,42	13,55	61,40	13,52	6,450	-1,100		
	25,81	-4,057	9,020	-0,2500	10,56	5,500	93,70	14,33	80,65	14,24	7,952	-0,370		
	14,60	-3,608	6,250	-0,060	7,560	5,600	51,98	13,95	77,79	14,26	4,666	-0,170		

TABELA 2.6. Variância e vício dos estimadores v_0 , v_1 , v_2 , v_3 , $v_1^*(\hat{Y}_R)$ e $v_2^*(\hat{Y}_R)$ para 15 amostras de diferentes tamanhos, retiradas dos 5 conjuntos de dados gerados para $t = 1$.

NBS = 1000 (número de repetições "bootstrap")

NMC = 1000 (número de repetições Monte Carlo)

	v_0		v_1		v_2		v_3		$v_1^*(\hat{Y}_R)$		$v_2^*(\hat{Y}_R)$	
	Var.	Vício	Var.	Vício	Var.	Vício	Var.	Vício	Var.	Vício	Var.	Vício
N=10	12996	-31,60	10000	-21,50	15625	3,100	42025	48,60	29036	16,80	10000	-19,80
	14884	-37,60	13225	-24,50	20449	2,800	54756	51,60	45283	19,80	13087	-17,95
	16129	-55,00	16129	-40,00	24025	-11,50	50176	36,70	34745	5,6	16384	-30,09
	18496	-31,40	15376	-17,00	22500	14,40	56664	71,00	42600	37,20	14400	-11,60
	21025	-49,90	21609	-34,00	32761	-3,5	67081	47,90	59927	12,70	20050	-27,70
N=25	234,1	-11,81	213,2	-3,830	510,8	15,14	1823,3	40,83	1085,7	35,08	226,5	0,190
	412,1	-14,11	295,8	-3,390	817,9	10,16	2227,8	35,61	1291,9	28,09	333,3	-2,290
	445,2	-14,25	313,3	-3,950	870,3	10,54	2199,6	36,27	1751,4	29,02	384,2	-2,200
	529,0	-12,20	380,3	-4,360	985,9	12,35	2401,0	31,22	1701,5	31,06	396,0	-1,070
	615,0	-9,510	424,4	-4,150	987,1	15,56	1984,0	39,52	1389,0	32,90	396,1	1,580
N=40	9,303	-3,706	6,554	0,120	21,63	5,870	79,93	14,65	96,47	9,900	8,796	-0,620
	15,68	-5,308	8,352	-1,550	40,58	4,150	84,09	12,88	92,90	9,770	9,479	-1,170
	15,13	-5,285	8,123	-1,390	39,31	4,480	129,9	13,49	99,10	10,47	15,52	-1,24
	18,84	-4,829	10,11	-0,990	47,20	4,790	151,3	15,60	98,10	7,920	16,30	-0,750
	18,49	-4,705	9,425	-0,860	48,58	4,910	158,8	13,74	108,2	12,91	17,30	-0,220

REFERENCIA BIBLIOGRÁFICA

- BERAN, R. (1983) - Estimated Sampling Distributions: The Bootstrap and competitors - Annals of statistics, 10, 212-225.
- BICKEL, P.J. and DOKSUM, K.A. (1977) - Mathematical Statistics - Holden - Day, S. Francisco.
- BICKEL, P.J., and FREEDMAN, D.A. (1981) - Some Asymptotic Theory for the Bootstrap - Annals of Statistics, 9, 1196-1217.
- COCHRAN, W.G. (1977) - Sampling Techniques - 3rd-ed - John Wiley & Sons, New York.
- DIACONIS, P. and EFRON, B. (1983) - Computer - Intensive Methods in Statistics - Scientific American, 96-107.
- DRAPPER, N.R. and SMITH, H. (1981) - Applied Regression Analysis, second Edition.
- DURBIN, J. (1959) - A note on the application of Quenouille's method of bias reduction to the estimation of ratios - Biometrika, 46, 477-480.

- EFRON, B. (1979) - Bootstrap methods: another look at the jackknife - Annals of Statistics, 7, 1-26.
- EFRON, B. (1979) - Computers and the Theory of Statistics: Thinking the Unthinkable - SIAM, 21, 462-477.
- FREEDMAN, D.A. (1981) - Bootstrapping Regression Models - Annals of Statistics, 9, 1218-1228.
- HARTLEY, H.O., and ROSS, A. (1954) - Unbiased ratio estimators - Nature, 174, London, 270-271.
- MILLER, R.G. (1974) - The jackknife - a review - Biometrika, 61, 1-15.
- QUENOUILLE, M.H. (1956) - Note on bias in estimation - Biometrika, 43, 353-360.
- RAO, P.S.R.S. and RAO, J.N.K. (1971) - Small sample results for ratio estimators - Biometrika, 58, 625-630.
- RAO, J.N.K. and KUZIK, R.A. (1974) - Sampling errors in ratio estimation - Sankhyā, C 36, 43-58.

- RAO, P.S.R.S. (1981) - Estimation of the mean Square Error of the Ratio estimator - Academic Press.
- ROYALL, R.M. and EBERHARDT, K.R. (1975) - Variance estimates for the ratio estimator. - Sankhyā, C 37, 43-52.
- ROYALL, R.M. and CUBERLAND, W.G. (1978) - Variance estimation in finite population sampling - Journal of the American Statistical Association, 73, 351-358.
- RUBIN, D.B. (1981) - The Bayesian Bootstrap - Annals of Statistics, 9, 130-134.
- SCHUCANY, W.R., GRAY, H.L. & OWEN, O.B. (1971) - On bias reduction in estimation - J. Am. Statist. ass., 66, to appear.
- SCOTT, A.J. & WU, C.F. (1982) - On the asymptotic distribution of ratio and regression estimators. J. Am. Statist. Assoc., 76, 98-102.
- TIN, M. (1965) - Comparison of some ratio estimators - J. Am. Statist. Ass., 60, 287-307.

- TUKEY, J.W. (1958) - Bias and confidence in not-quite - large sample - Annals of Mathematical Statistics (Abstract) 29, 614.
- WU, C.F. (1982) - Estimation of variance of the ratio estimator - Biometrika, 69, 183-189.