ESTUDO SOBRE A DISTRIBUIÇÃO DO n
NO DIMENSIONAMENTO DE AMOSTRAS

RUBENS LEITE DO CANTO BRAGA JUNIOR
Bacharel em Estatística

Orientador: Dr. Humberto de Campos

Dissertação apresentada à Escola
Superior de Agricultura "Luiz de
Queiroz", da Universidade de São
Paulo, para obtenção do título de
Mestre em Agronomia. Área de con-
centração: Estatística e Experi-
mentação Agronômica.

P I R A C I C A B A
Estado de São Paulo -- Brasil
Julho, 1986

*A minha família pelo bom
exemplo e pelo carinho*

D E D I C O

AGRADECIMENTOS

- Ao Dr. Humberto de Campos, Professor Titular do Departamento de Matemática e Estatística da ESALQ/USP, pela amizade e pela orientação segura deste trabalho.
- Aos Professores do Curso de Pós-Graduação em Estatística e Experimentação Agronômica da ESALQ/USP, na pessoa de seu Coordenador Prof. Décio Barbin, por sua acolhida e pelos ensinamentos recebidos.
- Ao Prof. Dr. Cássio Roberto de Melo Godoi, pelas óportunas sugestões.
- Ao Prof. Dr. Antonio Francisco Iemma, pelo apoio e incentivo recebidos durante a realização do curso.
- Aos Professores do Curso de Graduação no Bacharelado em Estatística na Universidade Federal de São Carlos, pelos ensinamentos básicos prestados.
- Aos Colegas João Gil de Luna, Geraldo Veríssimo de Souza Barbosa e Lusbi Antonio Herrera Sandoval pela amizade e troca de idéias durante o curso.
- Aos demais colegas do Curso de Pós-Graduação, pelo excelente convívio.

À FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo, pela bolsa de estudos concedida.

À EMPASC - Empresa Catarinense de Pesquisa Agropecuária S.A., pela bolsa de estudos concedida a partir do término da anterior.

Aos Funcionários do Departamento de Matemática e Estatística da ESALQ/USP, pela dedicação.

À Srt^a Rosa Maria Alves, Secretária do Departamento de Matemática e Estatística da ESALQ/USP, pela amizade e pelo trabalho de datilografia.

A Todos que, direta ou indiretamente, colaboraram para a realização deste trabalho.

Í N D I C E

	Pág.
RESUMO	vii
SUMMARY	x
1. INTRODUÇÃO	01
2. REVISÃO BIBLIOGRÁFICA	04
2.1. Amostragem	04
2.2. Alguns Conceitos Básicos de Estatística Matemática	08
3. MATERIAL	13
4. MÉTODOS	18
4.1. Obtenção da Equação de n	18
4.2. Amostragem Casual Simples	20
4.3. Amostragem Estratificada	21
4.4. Amostragem Sistemática	24
4.5. Amostragem por Conglomerados	25
5. RESULTADOS E DISCUSSÃO	28
5.1. Introdução	28
5.2. Amostragem Casual Simples	28
5.3. Amostragem Estratificada	42
5.4. Amostragem Sistemática	50
5.5. Amostragem por Conglomerados	50
6. ILUSTRAÇÃO DAS TÉCNICAS PROPOSTAS	59
6.1. Amostragem Casual Simples	59

Continuação.

	Pág.
6.2. Amostragem Estratificada	62
6.3. Amostragem Sistemática	66
6.4. Amostragem por Conglomerados	69
7. CONCLUSÕES	72
8. BIBLIOGRAFIA	75
APÊNDICE	78

ESTUDO SOBRE A DISTRIBUIÇÃO DO n
NO DIMENSIONAMENTO DE AMOSTRAS

Autor: RUBENS LEITE DO CANTO BRAGA JUNIOR

Orientador: Dr. HUMBERTO DE CAMPOS

R E S U M O

Os procedimentos utilizados na obtenção de dimensionamentos de amostras se baseiam, regra geral, em levantamentos anteriores ou nas informações contidas numa amostra inicial. Neste caso esta amostra gerará estimativas de parâmetros que serão usados na obtenção do melhor tamanho de amostra.

Desta forma estas estimativas são variáveis aleatórias, pois trabalhou-se com amostras iniciais tomadas aleatoriamente da população.

Os objetivos deste trabalho são os de estudar as distribuições de probabilidade da variável aleatória n (tamanho da amostra) para diversas técnicas de amostragem probabilística e com base nestas distribuições obter valores médios não aleatórios para n .

Supondo sempre populações normais, as principais conclusões foram as seguintes:

No caso da Amostragem Casual Simples a distribuição de probabilidade de n é dada por:

$$n \sim \frac{c}{(n_0-1)} \cdot \chi^2_{(n_0-1)} \quad , \quad \text{onde}$$

n_0 = tamanho da amostra inicial pré-fixada;

c = constante fixada a partir da precisão desejada para o levantamento.

Esta variável tem como valor médio e variância:

$$E(n) = c \quad ; \quad V(n) = \frac{2 \cdot c^2}{(n_0-1)} \quad .$$

Para a Amostragem Estratificada (Partilha Proporcional) tem-se:

$$n \sim \frac{c}{(n_0-L)} \cdot \chi^2_{(n_0-L)} \quad , \quad \text{onde}$$

L = número de estratos e

c tem o mesmo significado do caso anterior,

com média e variância iguais a

$$E(n) = c \quad ; \quad V(n) = \frac{2 \cdot c^2}{(n_0-L)} \quad .$$

Os resultados obtidos na Amostragem Sistemática são idênticos aos da Amostragem Casual Simples, quando é possível admitir-se que os elementos da população se sucedem de forma aleatória.

Na Amostragem por Conglomerados em estágio único obteve-se a seguinte distribuição de probabilidade:

$$n \sim \frac{M \cdot c}{(n_0 - 1)} \cdot \chi^2_{(n_0 - 1)} \quad , \quad \text{onde}$$

n_0 = nº de unidades primárias da amostra inicial;

M = nº de unidades secundárias por unidade primária.

O valor médio e a variância de n são dadas por:

$$E(n) = M \cdot c \quad ; \quad V(n) = \frac{2 \cdot M^2 \cdot c^2}{(n_0 - 1)} .$$

Finalmente pode-se observar que em todas as técnicas apresentadas, o valor médio da variável tamanho de amostra é dado em função apenas da precisão adotada para o levantamento e portanto não é um valor aleatório.

STUDY ABOUT THE n DISTRIBUTION IN SAMPLE-SIZE

Author: RUBENS LEITE DO CANTO BRAGA JUNIOR

Adviser: Dr. HUMBERTO DE CAMPOS

S U M M A R Y

The procedures used in obtaining sample-size are based, usually, on previously taken observed values or on information contained in an initial sample. In this last case, this sample will furnish parameter estimates which will be used in getting the best sample-size.

Therefore, those estimates are actually random variables because the initial sample is taken randomly from the population.

The objectives of this dissertation are: the study of probability distributions of the random sample-size by several sample techniques and, based on this distributions, to obtain the expected values for the sample-size.

Considering only normal populations, the mean conclusions were:

In the case of Non-stratified Sampling the probability distribution of n is given by

$$n \sim \frac{c}{(n_0-1)} \cdot \chi^2_{(n_0-1)} \quad , \quad \text{where}$$

n_0 = initial sample-size;

c = fixed constant based on the desired precision.

The expected value and variance of this variable are:

$$E(n) = c \quad ; \quad V(n) = \frac{2 \cdot c^2}{(n_0-1)}$$

In the case of Stratified Sampling (proportional partition) we get

$$n \sim \frac{c}{(n_0-L)} \cdot \chi^2_{(n_0-L)} \quad , \quad \text{where}$$

L = number of partitions

c = same meaning as the previous case.

with expected mean and variance given by

$$E(n) = c \quad ; \quad V(n) = \frac{2 \cdot c^2}{(n_0-L)}$$

The results obtained in the Systematic Sampling are identical as the above cited, being possible to admit that the population elements are given in a random succession.

In the Cluster Sampling case (one stage), we get the following probability distribution

$$n \sim \frac{M \cdot c}{(n_0 - 1)} \cdot \chi^2_{(n_0 - 1)} \quad , \quad \text{where}$$

n_0 = number of primary units in the initial sample

M = number of secondary units by primary units.

The value and variance of n are given by

$$E(n) = M \cdot c \quad ; \quad V(n) = \frac{2 \cdot M^2 \cdot c^2}{(n_0 - 1)}$$

Finally, we can observe that in all techniques presented, the expected values of the sample-size is only depending on the precision selected and, therefore is not a random variable.

1. INTRODUÇÃO

O levantamento de uma população por amostragem é uma técnica amplamente utilizada nos dias atuais. Isto decorre das numerosas vantagens que este processo proporciona, dentre outras, o menor custo do levantamento, a maior rapidez de análise dos dados, etc.

Existem na literatura quatro técnicas fundamentais de amostragem, quais sejam, Amostragem Casual Simples, Amostragem Estratificada, Amostragem Sistemática e Amostragem por Conglomerados.

Em todas elas, um dos principais problemas a ser estudado pelos pesquisadores é o do dimensionamento da amostra. Uma amostra muito grande implica num desperdício de recursos e uma amostra muito pequena pode provocar resultados não confiáveis.

As principais questões a serem observadas quando se está pesquisando o melhor tamanho da amostra são:

- A especificação da precisão dos resultados que esta amostra deverá gerar;
- A obtenção de uma equação que relacione o valor de n (tamanho da amostra) com o grau de precisão adotado;
- A disponibilidade de recursos a serem empregados na realização do levantamento dos dados.

Como se vê, a grandeza da amostra será obtida principalmente em função da exatidão dos resultados e do custo a ser considerado.

Sabe-se que os procedimentos utilizados no dimensionamento de amostras se baseiam ou em informações anteriormente obtidas ou em uma pequena amostra inicial (amostra piloto) retirada da população a ser estudada. Neste caso esta amostra piloto gerará estimativas dos parâmetros populacionais (média e variância) as quais propiciarão a obtenção do melhor valor para n .

Deste modo, fica implícito que estas estimativas são variáveis aleatórias, pois a amostra piloto é retirada aleatoriamente da população. Assim, diferentes pesquisadores que estudarem uma mesma população, mesmo fixando o tamanho da amostra piloto obterão diferentes esti-

mativas de parâmetros populacionais, o que resultará numa diversidade de tamanhos de amostra, para um mesmo grau de precisão.

Os objetivos deste trabalho são os de estudar as distribuições de probabilidade da variável aleatória n (tamanho de amostra), quando se baseiam em levantamentos anteriores, para diversos métodos de amostragem probabilística.

Com base nestas distribuições de probabilidade obter os valores médios e intervalos de confiança para a variável aleatória n , fazendo com que o dimensionamento perca o caráter aleatório.

2. REVISÃO BIBLIOGRÁFICA

2.1. Amostragem

Numerosos autores apresentam métodos de obtenção de dimensionamentos de amostras associados às diferentes técnicas de amostragem. Cita-se a seguir alguns destes autores, uma vez que existe grande afinidade entre eles quanto à metodologia proposta.

HANSEN et alii (1953) definem as principais características da amostragem probabilística, assim como, obtém estimadores gerados pelas amostras e suas respectivas esperanças matemáticas.

Os estimadores da variância da média populacional são encontrados em KISH (1965), para todas as técnicas de amostragem probabilística. Estes estimadores são utilizados na obtenção dos dimensionamentos amostrais.

Também MACE (1964) trata de dimensionamentos de amostras, determinando os intervalos de confiança para os mais diversos tipos de contrastes entre parâmetros e a partir deles obtém o melhor tamanho de amostra para cada tipo.

YAMANE (1967) estrutura dimensionamento amostral através de dois princípios básicos, quais sejam, o princípio do máximo de exatidão dos resultados a serem obtidos relacionados com um mínimo de recursos dispensados na experimentação.

Um método para a obtenção das equações de alocação ótima do tamanho de amostra é apresentado por SUKHATAME (1956), que se utiliza da aproximação normal fornecida pelo Teorema do Limite Central e a partir da probabilidade $P\{|\bar{y} - \bar{Y}| \geq d\} = \alpha$ obtém o melhor valor para o tamanho da amostra.

Com base no método anterior COCHRAN (1965), dentre outros, obtém o dimensionamento amostral em populações consideradas como infinitas ($n \leq 0,05 \cdot N$), para as seguintes técnicas de amostragem:

AMOSTRAGEM CASUAL SIMPLES

$$n = \frac{t^2 \cdot s^2}{d^2}, \text{ onde}$$

n = tamanho da amostra em populações consideradas infinitas;

$t_{\alpha/2}$ = abscissa da curva de frequência t-Student, que define uma área $\alpha/2$;

s^2 = variância amostral;

d = semi-amplitude do intervalo de confiança.

AMOSTRAGEM ESTRATIFICADA (PARTILHA PROPORCIONAL)

$$n = \frac{t^2}{d^2} \left(\sum_{k=1}^K W_h s_h^2 \right)$$

W_h = peso do estrato h na população;

s_h^2 = variância amostral do estrato h ;

K = número de estratos.

Segundo o mesmo autor, estas expressões podem ser reestruturadas no caso de se conhecer o tamanho da população, com o uso da seguinte fórmula:

$$n' = \frac{n}{1 + \frac{n}{N}}, \text{ onde}$$

n' = tamanho da amostra em populações finitas,

N = tamanho da população.

Ainda sobre a expressão de dimensionamento de amostras estratificadas, SATTERTHWAITTE (1946) mostra

que a média ponderada de variâncias amostrais possui distribuição aproximada de uma χ^2 (qui-quadrado).

SUKHATAME e SUKHATAME (1970) também obtêm estas expressões de dimensionamentos de Amostras Estratificadas, minimizando o custo do levantamento, quando o valor da variância da média está fixado.

COCHRAN (1965) relaciona a Amostragem Sistemática com a Amostragem Casual Simples, quando se estuda populações consideradas como aleatórias.

Por sua vez, RAJ (1968) apresenta um estudo da Amostragem por Conglomerados em um e dois estágios, com suas respectivas estimativas de média e variância, as quais serão utilizadas na obtenção do melhor tamanho da amostra.

CAMPOS (1985) obtém o dimensionamento de amostras por Conglomerados em estágio único, com o número de unidades secundárias constante para todas as unidades primárias.

Supondo a população como infinita o tamanho da amostra será dado por:

$$n = \frac{t^2 s_1^2}{d^2} \quad , \quad \text{onde}$$

s_1^2 = variância amostral das unidades primárias.

2.2. Alguns Conceitos Básicos de Estatística Matemática

No estudo das distribuições de probabilidade das variáveis aleatórias tamanho de amostra relacionadas com as diversas técnicas de amostragem, serão utilizados importantes teoremas, axiomas e definições da estatística matemática.

No interesse de localizar o leitor será feita uma breve revisão bibliográfica dos conceitos de estatística matemática envolvidos neste trabalho.

As distribuições de probabilidade de variáveis amostrais são encontradas em MOOD et alii (1974), onde partindo de uma variável $X \sim N(\mu, \sigma^2)$ se verificam as demonstrações das relações:

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

$$U = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

Estes autores apresentam ainda os primeiros momentos de uma variável com distribuição de χ^2 (qui-quadrado).

Também COSTA NETO (1977) explana resumidamente sobre as distribuições amostrais da média $\bar{x}$ e da variância s^2 .

Estas mesmas definições podem ser encontradas em MEYER (1978), que por sua vez, apresenta ainda as definições de esperança e variância matemáticas de variáveis aleatórias, bem como as mais importantes propriedades destas definições, de valor preponderante no estudo de distribuições de probabilidade.

Alguns teoremas importantes nas deduções de variáveis são encontrados e provados em MURTEIRA (1979), dentre eles tem-se que para $X \sim N(\mu, \sigma^2)$ então:

$$v = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \text{ e ainda}$$

$$v^2 = \frac{(\bar{x} - \mu)^2}{\sigma^2/n} \sim \chi^2_{(1)}$$

A prova de que as estatísticas $\bar{x}$ (média) e s^2 (variância) amostrais são independentes sob normalidade, pode ser encontrada em MOOD et alii (1974). Esta prova é efetuada com o uso da função geradora de momentos.

HOEL (1980) trata da teoria de amostragem e demonstra através de deduções e também pela função geradora de momentos, que a soma de variáveis aleatórias independentes, identicamente distribuídas e com distribuição normal possui distribuição normal com parâmetros: soma das médias e soma das variâncias, ou seja, se $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \dots, n$ então

$$Y = \sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$$

A mediana de uma variável aleatória do tipo discreta ou contínua é definida por HOOG e CRAIG (1978) da seguinte forma:

Mediana de uma variável X é o valor x tal que $P(X < x) \leq 0,5$ e $P(X \leq x) \geq 0,5$ e para o caso contínuo $P(X < x) = P(X \leq x) = 0,5$.

O Teorema do Limite Central é aplicado por HOEL et alii (1978) na obtenção de aproximações normais que facilitam a construção de intervalos de confiança.

As funções densidades de probabilidade de variáveis aleatórias com distribuições Normal, t -Student, χ^2 (qui-quadrado) e F -Snedecor todas centrais, são definidas por HOOG e GRAIG (1978). Sendo que em particular a f. d.p. de uma variável $F \sim F(r_1, r_2)$ é dada por:

$$g(f) = \begin{cases} \frac{\Gamma\left(\frac{r_1+r_2}{2}\right) \cdot \left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} \cdot (f)^{\frac{r_1}{2}-1}}{\Gamma\left(\frac{r_1}{2}\right) \cdot \Gamma\left(\frac{r_2}{2}\right) \cdot \left(1 + \frac{r_1 f}{r_2}\right)^{\left(\frac{r_1+r_2}{2}\right)}}, & f > 0 \\ 0, & f \leq 0 \end{cases}$$

DIXON e MASSEY (1969) fornecem tabelas mais completas da distribuição F-Snedecor, inclusive para nível de significância $\alpha = 0,50$, usado na obtenção da mediana da variável n (tamanho da amostra).

Através de formas quadráticas, GRAYBILL (1976) obtém as funções densidades de probabilidade de variáveis com distribuições χ^2 (qui-quadrado), F-Snedecor e t-Student não centrais. No caso da F não central a f.d.p. é dada por:

$$F'(w, n_1, n_2, \lambda) = \begin{cases} \sum_{j=0}^{\infty} \frac{\lambda^j \cdot e^{-\lambda} \cdot \Gamma\left(\frac{2j+n_1+n_2}{2}\right) \cdot \left(\frac{n_1}{n_2}\right)^{\frac{n_1+n_2}{2}} \cdot w^{\frac{n_1+2j-2}{2}}}{j! \cdot \Gamma\left(\frac{n_2}{2}\right) \cdot \Gamma\left(\frac{2j+n_1}{2}\right) \cdot \left(1 + \frac{n_1 \cdot w}{n_2}\right)^{\frac{n_1+n_2+2j}{2}}}, & w > 0 \\ 0, & w \leq 0 \end{cases}$$

onde $\lambda = \frac{\mu_1 \mu}{2\sigma^2}$ é o parâmetro de não centralidade.

GRAYBILL (1961) também define as distribuições χ^2 e F não centrais, e em particular a distribuição da estatística:

$$u = \frac{\bar{y}^2 \cdot n \cdot (n-1)}{\sum_{i=1}^n (y_i - \bar{y})^2} \sim F'(u, 1, n, \frac{n \cdot \mu^2}{2 \cdot \sigma^2})$$

TIKU (1966) apresenta várias técnicas de aproximações para a obtenção de probabilidades de variáveis com distribuição F não central, entre elas a aproximação normal obtida por SEVERO e ZELEN (1960).

3. MATERIAL

Anesar de haver-se considerado este estudo mais pelo aspecto teórico, sentiu-se a necessidade de ilustrar, com exemplos, às diversas técnicas de dimensionamentos amostrais abordadas.

No caso de Amostragem Casual Simples foi utilizada uma amostra de tamanho $n_0 = 36$, com dados de áreas (em ha) cultivadas com mandioca, em propriedades de até 25 ha do estado da Bahia. Estes dados constam da Tabela 1.

TABELA 1 - Áreas (em ha) cultivadas com mandioca, no estado da Bahia.

4,7	3,6	2,7	8,8	6,7	6,2	9,0	7,8	7,1
4,3	4,2	9,1	6,4	5,2	2,7	12,6	10,1	5,0
6,6	7,9	12,2	11,2	6,5	7,8	9,5	8,3	13,2
8,1	8,6	5,9	10,9	7,2	8,5	6,5	8,7	6,8

Ilustrando a técnica de Amostragem Estratificada utilizou-se uma população com 300 dados de fornecimentos de cana-de-açúcar (em t.) nas usinas da região de Piracicaba. Esta população está ordenada de forma decrescente em toneladas de cana-de-açúcar e pode ser encontrada na Tabela I do apêndice.

A população de fornecimentos foi subdividida "a priori" em 5 estratos de onde foram retiradas amostras iniciais conforme se observa na Tabela 2.

TABELA 2 - Dados de uma amostra estratificada de 40 fornecimentos (em t.)
de cana-de-açúcar.

ESTRATOS	CLASSES	Nh	nh	FORNECIMENTOS (em t.)
1	< 43,9	120	15	27, 4, 14, 19, 30, 16, 21, 39, 1, 12, 24, 7, 9, 34, 6
2	43,9 - 126,4	80	8	87, 66, 126, 75, 58, 106, 98, 50
3	126,4 - 366,4	50	6	201, 306, 146, 173, 358, 244
4	366,4 - 948,3	30	6	523, 706, 456, 878, 383, 577
5	≥ 948,3	20	5	1122, 2299, 3943, 1593, 948

Ordenando a população vista no exemplo anterior alfabeticamente pelos nomes dos fornecedores, como se vê na Tabela II do apêndice, foi possível exemplificar a técnica de Amostragem Sistemática, pois deste modo os elementos da população (fornecimento da cana em t.) se sucedem de forma aleatória.

Assim a população foi subdividida em 30 grupos de 10 observações, seguindo a ordem alfabética, sendo sorteado um número i de 1 a 10 e com isto obteve-se uma amostra de 30 elementos, tomando-se sempre o i-ésimo elemento de cada grupo. Os dados desta amostra sistemática são encontrados na Tabela 3.

TABELA 3 - Dados de uma amostra sistemática de fornecimentos (em t.) de cana-de-açúcar na região de Piracicaba.

15	70	878	3	15	236	316	50	99	16	7
177	628	948	48	41	17	150	39	181	27	8
8	46	1456	38	79	13	30	3			

Para exemplificar a Amostragem por Conglomerados, foram tomados dados de produtividade de cana-de-açúcar (T.C.H.) encontrados em CAMPOS (1984). Considerou-se

um experimento com 5 variedades em 4 blocos casualizados, com 2 espaçamentos a nível de subparcelas, instalado em 3 locais distintos.

Descaracterizando-se os espaçamentos e blocos e tomando as 15 interações entre locais e variedades como unidades primárias, tem-se 8 unidades secundárias por unidade primária, cujos dados são apresentados na Tabela 4.

TABELA 4 - Dados de produtividade de cana-de-açúcar (TCH) em 15 conglomerados.

U.P.	UNIDADES SECUNDÁRIAS								TOTAIS
1	100,4	97,3	84,2	103,6	81,7	92,6	78,8	91,5	730,1
2	83,4	60,7	71,3	62,6	90,7	58,4	65,2	58,9	551,2
3	62,6	68,4	71,4	57,8	60,5	61,7	67,2	56,3	505,9
4	76,4	63,8	57,2	68,3	78,6	60,7	60,3	63,6	528,9
5	72,6	54,2	60,5	57,4	65,7	56,3	51,3	52,6	470,6
6	49,2	53,2	60,6	54,2	40,5	50,6	57,2	48,6	414,1
7	105,8	89,7	84,9	97,1	102,0	86,3	85,7	100,2	751,7
8	101,8	96,3	110,0	90,5	86,3	90,7	92,4	85,8	753,8
9	78,4	83,8	93,2	85,7	73,8	80,9	84,5	83,8	664,1
10	97,1	93,2	86,8	92,7	90,4	95,6	90,4	89,5	735,7
11	105,0	101,4	97,2	89,6	94,3	91,7	93,5	81,8	754,5
12	80,7	92,0	97,3	93,7	72,6	85,4	94,8	87,4	703,9
13	65,6	62,7	70,3	69,5	69,7	60,4	75,2	77,5	550,9
14	57,3	50,2	61,3	65,2	51,6	53,4	51,9	48,7	439,6
15	50,3	61,8	65,2	83,6	45,7	63,6	53,6	68,5	492,3

T.C.H. = tonelada de cana por hectare.

4. MÉTODOS

4.1. Obtenção da Equação de n

Como foi visto em 1., uma das principais questões de interesse, quando se estuda o tamanho ótimo da amostra, é a obtenção de uma equação que relacione o valor n (tamanho de amostra) com o grau de precisão adotado.

O procedimento usual para se dimensionar uma amostra é o seguinte.

Fixa-se a precisão d (semi-amplitude do intervalo de confiança) a ser considerada no experimento e toma-se a probabilidade $P\{|\bar{y} - \bar{Y}| \geq d\} = \alpha$, onde $1-\alpha$ é o coeficiente de confiança do intervalo.

Deste modo ao dividir-se membro a membro, a probabilidade acima, pelo desvio padrão da média populacional obtém-se:

$$P \left\{ \frac{|\bar{y} - \bar{Y}|}{s(\bar{y})} \geq \frac{d}{s(\bar{y})} \right\} = \alpha$$

É sabido que pela aproximação normal obtida do Teorema do Limite Central,

$$\frac{|\bar{y} - \bar{Y}|}{s(\bar{y})} \sim Z_{\alpha/2},$$

onde $Z_{\alpha/2}$ é um limite superior da distribuição $N(0, 1)$

Assim tem-se:

$$Z_{\alpha/2} \geq \frac{d}{s(\bar{y})}$$

o que equivale à $d \leq Z_{\alpha/2} s(\bar{y})$

Como geralmente não se conhece o desvio padrão da média populacional ($S(\bar{y})$), o procedimento usual é estimá-lo através de uma pequena amostra inicial de tamanho n_0 . Deste modo,

$$P \left\{ \frac{|\bar{y} - \bar{Y}|}{s(\bar{y})} \geq \frac{d}{s(\bar{y})} \right\} = \alpha, \text{ e neste caso,}$$

$$\frac{|\bar{y} - \bar{Y}|}{s(\bar{y})} \sim t_{(n_0-1); \alpha/2}$$

é uma aproximação da distribuição t-Student com (n_0-1) gl, e nível de probabilidade $\alpha/2$. Daí tem-se:

$$t_{n_0-1; \alpha/2} \geq \frac{d}{s(\bar{y})}$$

o que equivale à $d \leq t_{n_0-1; \alpha/2} \cdot s(\bar{y})$

A partir desta expressão, obtém-se as equações que relacionam a variável n com a precisão adotada, para as diversas técnicas de amostragem.

4.2. Amostragem Casual Simples (A.C.S.)

Como se sabe, quando a amostra é obtida pela técnica da A.C.S. o valor estimado do desvio padrão da média, quando se trabalha com populações consideradas infinitas ($n \leq 0,05.N$), é dado por:

$$s(\bar{y}) = s/\sqrt{n} \quad , \quad \text{onde}$$

n = tamanho da amostra;

N = tamanho da população;

s = estimativa do desvio padrão, fornecida pela amostra inicial.

Pela técnica da Amostragem Casual Simples a semi-amplitude do intervalo de confiança é dada por:

$$d = t_{(n_0-1); \alpha/2} \cdot s/\sqrt{n}$$

Daí tem-se:

$$d^2 = t^2 \cdot \frac{s^2}{n} ,$$

donde,

$$n = \frac{t^2 \cdot s^2}{d^2} \quad (4.2)$$

Uma vez obtida a equação que relacione o valor de n (tamanho de amostra) com o grau de precisão adotado, pode-se fixar o valor da precisão d de duas maneiras:

- 1 - Diretamente, fixando $d = K$, onde K é uma constante, ou
- 2 - Tomando $d = K_1 \cdot \bar{y}$, onde K_1 é constante e $\bar{y}$ é a média da amostra inicial, ou seja, d é uma fração da média amostral.

4.3. Amostragem Estratificada

Em levantamentos amostrais é comum deparar-se com populações extremamente heterogêneas, que se estudadas pela técnica da Amostragem Casual Simples acarretariam graves problemas de imprecisão.

Nestes casos, divide-se a população em subpopulações mais homogêneas, que serão chamadas de estratos.

Admite-se então que para cada estrato h tem-se:

N_h : nº de unidades do estrato h na população;

n_h : nº de unidades do estrato h na amostra;

y_{h_i} : observação i do estrato h ;

$W_h = \frac{N_h}{N}$: peso do estrato h na população;

$\bar{Y}_h = \frac{\sum_{i=1}^{N_h} y_{h_i}}{N_h}$: valor médio do estrato h ;

$\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{h_i}}{n_h}$: estimativa do valor médio do estrato h ;

$S_h^2 = \frac{\sum_{i=1}^{N_h} (y_{h_i} - \bar{Y}_h)^2}{N_h - 1}$: variância do estrato h ;

$s_h^2 = \frac{\sum_{i=1}^{n_h} (y_{h_i} - \bar{y}_h)^2}{n_h - 1}$: estimativa da variância do estrato h .

Este trabalho fixar-se-á, para facilitar a obtenção da distribuição de probabilidade da variável n , na partilha proporcional onde $n_h/n = N_h/N$.

A estimativa do valor médio da população é dada por:

$$\bar{y}_{est} = \sum_{i=1}^L W_h \bar{y}_h ,$$

onde L é o número de estratos que subdividem a população.

Quando se considera a população como infinita, a variância do estimador da média, em cada estrato é dada por:

$$V(\bar{y}_h) = \frac{Sh^2}{nh} ,$$

cujo estimador é

$$\hat{V}(\bar{y}_h) = \frac{sh^2}{nh} .$$

Consequentemente a estimativa da variância da média populacional é obtida por:

$$\hat{V}(\bar{y}_{est}) = \sum_{h=1}^L W_h^2 \hat{V}(\bar{y}_h) = \sum_{h=1}^L W_h^2 \frac{sh^2}{nh} .$$

Adotando a partilha proporcional ($nh=n.Nh/N$) e ainda fixando a variância da média em $V= d^2/t^2$ tem-se:

$$V = \sum_{h=1}^L W_h^2 \frac{sh^2}{nh} = \sum_{h=1}^L \frac{Nh^2}{N^2} \cdot \frac{sh^2}{\frac{n.Nh}{N}} = \sum_{h=1}^L W_h \frac{sh^2}{n} .$$

Portanto,

$$n = \frac{1}{V} \cdot \sum_{h=1}^L W_h sh^2 = \frac{t^2}{d^2} \cdot \sum_{h=1}^L W_h sh^2 \quad (4.3) .$$

Assim obteve-se uma equação que relaciona o tamanho da amostra com a precisão adotada na Amostragem Estratificada (partilha proporcional).

4.4. Amostragem Sistemática

Supondo que as N unidades da população em estudo possuam uma ordenação qualquer, pode-se tomar uma amostra de tamanho n , subdividindo a população em n grupos iguais de tamanho K (onde $K = N/n$).

Retirando-se uma unidade x ao acaso, do primeiro grupo e tomando as seguintes unidades: $K+x$, $2K+x$, $3K+x$, ..., $(n-1)K+x$ dos grupos posteriores, montaria-se a amostra.

Este processo de seleção de uma amostra é chamado de Amostragem Sistemática.

Algumas vezes a amostragem sistemática é aplicada em populações onde a ordenação das unidades é efetivamente aleatória. Isto ocorre quando a ordenação dos elementos da população não se associa com a variável que tem-se interesse de estudar.

Neste caso a ordem da população é chamada de aleatória e nesta situação é de se esperar que a Amos-tragem Sistemática possa ser interpretada como uma Amos-tra-gem Casual Simples.

Portanto, nestas circunstâncias, pode-se estender as técnicas de dimensionamento de amostra obtidas na A.C.S. para dados provenientes de uma amostragem siste-mática, sem incorrer em grandes erros.

4.5. Amostragem por Conglomerados

Em levantamentos amostrais, muitas vezes a unidade amostral, também chamada de unidade primária e formada por um grupo de unidades secundárias, constituindo assim o que denota-se por conglomerados.

Conforme a tomada dos elementos dentro de cada conglomerado pode-se ter diversos tipos de amostragem. Este trabalho fixa-se na Amostragem por Conglomerados em estágio único com número de elementos dentro de cada conglomerado constante, ou seja, são levantadas todas as unidades secundárias dos conglomerados selecionados.

Admita-se então que:

N = nº de unidades primárias na população;

n = nº de unidades primárias na amostra;

M = nº de unidades secundárias em cada U.P.;

y_{ij} = elemento j do conglomerado i , com $j=1,2,\dots,M$;

$y_{i.} = \sum_{j=1}^M y_{ij}$ = total das U.S. do conglomerado i , $i=1,2,\dots,N$;

$\bar{Y}$ = valor médio das unidades primárias;

$\bar{\bar{Y}}$ = valor médio das unidades secundárias;

$$\text{onde } \bar{Y} = \frac{\sum_{i=1}^N y_{i.}}{N} \quad \text{e} \quad \bar{\bar{Y}} = \frac{\sum_{i=1}^N y_{i.}}{N \cdot M} = \frac{\bar{Y}}{M} .$$

Suas estimativas, a nível de amostra, são respectivamente:

$$\bar{y} = \frac{\sum_{i=1}^n y_{i.}}{n} \quad \text{e} \quad \bar{\bar{y}} = \frac{\sum_{i=1}^n y_{i.}}{n.M} = \frac{\bar{y}}{M}$$

A fim de estruturar as variâncias dos valores médios, admitiu-se:

S_1^2 = variância das unidades primárias e

S^2 = variância das unidades secundárias, onde

$$S_1^2 = \frac{\sum_{i=1}^N y_{i.}^2 - \frac{(\sum y_{i.})^2}{N}}{N - 1}$$

$$S^2 = \frac{\sum_{i,j} y_{ij}^2 - \frac{(\sum_{i,j} y_{ij})^2}{N.M}}{N.M - 1}$$

Suas estimativas são obtidas pelas mesmas fórmulas, mas considerando apenas as n U.P. de amostra e consequentemente substituindo N por n .

Admitindo uma Amostragem Casual Simples para a tomada das U.P. e fixando a semi-amplitude do intervalo de confiança para média $\bar{y}$, ou seja,

$$d = t \cdot s(\bar{y}) ,$$

para uma população tida como infinita, tem-se:

$$s(\bar{y}) = \frac{s_1}{\sqrt{n}} \quad .$$

Consequentemente,

$$d = \frac{t \cdot s_1}{\sqrt{n}} \implies d^2 = \frac{t^2 \cdot s_1^2}{n}$$

e portanto a equação de n (tamanho de amostra) é dada por:

$$n = \frac{t^2 \cdot s_1^2}{d^2} \quad (4.5)$$

5. RESULTADOS E DISCUSSÃO

5.1. Introdução

Com a obtenção das equações da variável aleatória n (tamanho de amostra) para as diversas técnicas de amostragem, pode-se agora obter as distribuições de probabilidade destas variáveis.

5.2. Amostragem Casual Simples

Partindo-se de um tamanho de amostra inicial n_0 pré-fixado e de um nível α de probabilidade convenientemente selecionado, resulta um valor de t constante. Nestas condições ao fixar-se o valor $d = K$ tem-se que,

$$a = \frac{t^2}{K^2} \quad \text{é uma constante.}$$

Assim a expressão (4.2) do valor de n fica: $n = a.s^2$, onde s^2 é a única variável aleatória (dependente da amostra).

Supondo uma população normalmente distribuída, com média zero e variância igual a um ($Y \sim N(0, 1)$), então o valor da esperança matemática da variável aleatória n será dado por:

$$E(n) = E(a.s^2) = a.E(s^2) = a.1 = a$$

pois s^2 é um estimador não viciado da variância populacional.

Sabe-se ainda que se $Y \sim N(0, 1)$, então:

$$(n_0-1).s^2 \sim \chi^2_{(n_0-1)} \quad (\text{qui-quadrado com } n_0-1 \text{ gl.})$$

então pode-se afirmar:

$$Q = (n_0-1).s^2 = \frac{(n_0-1)}{a} . n \sim \chi^2_{(n_0-1)}$$

portanto:

$$n = \frac{a}{(n_0-1)} . Q \sim \frac{a}{(n_0-1)} . \chi^2_{(n_0-1)}$$

Logo, a variável aleatória n tem distribuição de qui-quadrado com n_0-1 g.l., multiplicada por $\frac{a}{n_0-1}$.

ma constante que só depende do tamanho da amostra inicial n_0 e do grau de precisão d desejado.

Com isto pode-se verificar que o valor médio da variável aleatória n é realmente igual à a pois:

$$E(n) = \left[\frac{a \cdot Q}{(n_0 - 1)} \right] = \frac{a}{(n_0 - 1)} E(Q) = \frac{a(n_0 - 1)}{(n_0 - 1)} = a.$$

Pode-se também obter o valor da variância desta variável, da seguinte maneira:

$$\begin{aligned} V(n) &= V \left[\frac{a \cdot Q}{(n_0 - 1)} \right] = \frac{a^2}{(n_0 - 1)^2} \cdot V(Q) = \frac{a^2}{(n_0 - 1)^2} \cdot 2(n_0 - 1) = \\ &= \frac{2 \cdot a^2}{(n_0 - 1)}. \end{aligned}$$

Por outro lado, tomando d pela segunda alternativa ($d = K_1 \cdot \bar{y}$) e novamente fixando o tamanho n_0 da amostra inicial e o nível α de probabilidade tem-se:

$$n = \frac{t^2 \cdot s^2}{(K_1 \cdot \bar{y})^2} = \frac{t^2 \cdot s^2}{K_1^2 \cdot \bar{y}^2},$$

onde $\frac{t^2}{K_1^2} = b$ é uma constante.

Assim o valor de n depende agora de duas variáveis aleatórias s^2 e $\bar{y}$.

Supondo-se uma população normal padronizada, tem-se:

$$(n_o - 1) \cdot s^2 \sim \chi^2_{(n_o - 1)}$$

e pode-se ainda afirmar que:

$$\bar{y} \sim N(0, 1/n_o) \quad .$$

Assim,

$$\frac{\bar{y}}{\sqrt{1/n_o}} \sim N(0, 1)$$

e por sua vez,

$$\left[\frac{\bar{y}}{\sqrt{1/n_o}} \right]^2 = n_o \cdot \bar{y}^2 \sim \chi^2_{(1)} \quad .$$

Como pode-se provar $\bar{y}$ e s^2 são variáveis aleatórias independentes sob normalidade e deste modo tem-se:

$$F = \frac{\frac{(n_o - 1) \cdot s^2}{(n_o - 1)}}{\frac{n_o \cdot \bar{y}^2}{1}} \sim F_{(n_o - 1, 1)} \quad , \quad \text{logo}$$

$$F = \frac{s^2}{n_0 \cdot \bar{y}^2} = \frac{1}{n_0 \cdot b} \left(b \cdot \frac{s^2}{\bar{y}^2} \right) = \frac{n}{n_0 \cdot b}, \text{ portanto}$$

$$n = b \cdot n_0 \cdot F \sim b \cdot n_0 \cdot F(n_0 - 1, 1)$$

Deste modo, a variável aleatória n tem distribuição de F-Snedecor com $n_0 - 1$ e 1 q.l., multiplicada por constantes que dependem apenas do tamanho da amostra inicial n_0 e do valor da semi-amplitude do intervalo de confiança desejado.

Neste caso, torna-se impossível obter o valor da esperança matemática da variável aleatória n , pois como se sabe, a esperança de uma variável com distribuição $F(n_1, n_2)$ só existe para valores de $n_2 > 2$. Assim faz-se necessário encontrar uma outra medida de tendência central para esta variável.

Uma saída é a de obter-se o valor da me-diana desta distribuição, da seguinte forma:

Seja Y uma variável aleatória com distribuição $F(n_0 - 1, 1)$, então a função densidade de probabilidade desta variável é dada por:

$$f(y) = \frac{(n_0-1) \left(\frac{n_0-1}{2}\right) \cdot y^{\frac{n_0-3}{2}}}{B\left(\frac{n_0-1}{2}; \frac{1}{2}\right) (1 + (n_0-1) \cdot y)^{n_0/2}} I_{R_+}(y).$$

onde:

$$I_{R_+}(y) = \begin{cases} 1, & \text{se } y \geq 0 \\ 0, & \text{caso contrário} \end{cases}$$

$B\left(\frac{n_0-1}{2}; \frac{1}{2}\right)$ é a função beta com parâmetros $\frac{n_0-1}{2}$ e $\frac{1}{2}$.

Pode-se então obter o valor da mediana de Y por intermédio da integral

$$\int_0^{md} f(y) dy = \frac{1}{2},$$

pois a mediana de uma v.a. contínua Y é um valor md , tal que, $P(Y \leq md) = 1/2$.

Nestas circunstâncias o valor da mediana da variável aleatória n é dado por:

$$\text{mediana}(n) = n_0 \cdot b \cdot md.$$

As maiores dificuldades são encontradas quando se tenta estender os resultados anteriores para amostras provenientes de populações normalmente distribuídas, com média e variância desconhecidas.

A falta de conhecimento destes parâmetros irá gerar problemas na obtenção da distribuição de probabilidade da variável aleatória n , os quais serão abordados a seguir.

Quando se fixa o valor de $d = K$, onde K é uma constante tem-se, como já se viu, $n = a.s^2$. Considerando $Y \sim N(\mu, \sigma^2)$, então

$$Q_1 = \frac{(n_0 - 1) \cdot s^2}{\sigma^2} \sim \chi^2_{(n_0 - 1)}$$

e por analogia ao caso anterior,

$$n \sim \sigma^2 \cdot \frac{a}{(n_0 - 1)} \cdot \chi^2_{(n_0 - 1)}.$$

Como não se conhece o valor da variância populacional σ^2 , não se pode obter a distribuição de probabilidade da variável aleatória n , com o uso da técnica anteriormente apresentada.

Para resolver o problema da falta de conhecimento deste parâmetro propõe-se a seguinte técnica:

Adota-se para valor da precisão $\underline{d}$ uma fração do desvio padrão populacional, ou seja, $d = K_2 \sigma$ onde K_2 é uma constante, e deste modo tem-se:

$$n = \frac{t^2 \cdot s^2}{d^2} = \frac{t^2 \cdot s^2}{K_2^2 \cdot \sigma^2}$$

Como n_0 e conseqüentemente t estão fixados "a priori", resulta que:

$$c = \frac{t^2}{K_2^2} \text{ é uma constante e portanto,}$$

$$n = c \cdot \frac{s^2}{\sigma^2} .$$

Por analogia ao visto anteriormente virá:

$$n = \frac{c}{(n_0 - 1)} \cdot Q_1 \sim \frac{c}{(n_0 - 1)} \cdot \chi^2_{(n_0 - 1)}$$

Com isto pode-se obter os valores da média e variância desta variável aleatória, como se segue:

$$E(n) = E \left[\frac{c}{(n_0 - 1)} Q_1 \right] = c$$

$$V(n) = V \left[\frac{c}{(n_0 - 1)} Q_1 \right] = \frac{2 \cdot c^2}{(n_0 - 1)}$$

A fixação do valor da constante K_2 , que especifica a precisão ($d = K_2 \cdot \sigma$), é subjetiva.

Uma sugestão para estimar o desvio padrão σ é a de fazer uso do método prático de dimensionamento de amostra visto em CAMPOS (1985).

Admitindo que a população tenha distribuição normal, aproximadamente 95% dos dados se situam no intervalo $\bar{y} \pm 2\sigma$, o que nos dá uma amplitude $A = 4\sigma$. Então, grosseiramente, pode-se estimar σ tomando os valores extremos dos dados e determinando a amplitude.

Com isto tem-se:

$$\sigma = \frac{A}{4}$$

A partir desta estimativa é possível a obtenção de um valor para constante K_2 com base na precisão d desejada.

Estando fixados o tamanho da amostra inicial e o valor da constante K_2 , ou seja, fixada a precisão desejada para o levantamento, pode-se obter um valor médio para a variável aleatória n (tamanho da amostra). Assim sendo, diferentes pesquisadores deverão obter o mesmo dimensionamento de amostra, partindo da mesma precisão.

Pode-se ainda, construir um intervalo de confiança para a variável aleatória n , conforme se segue:

$$P \{ n_1 \leq n \leq n_2 \} = 1-\alpha \implies$$

$$P \left\{ n_1 \leq \frac{c}{(n_0-1)} Q_1 \leq n_2 \right\} = 1-\alpha \implies$$

$$P \left\{ \frac{(n_0-1)}{c} n_1 \leq Q_1 \leq \frac{(n_0-1)}{c} n_2 \right\} = 1-\alpha$$

onde $1-\alpha$ é o nível de confiança.

Como Q_1 possui distribuição de qui-quadrado, amplamente conhecida e tabelada, torna-se fácil a obtenção dos valores de n_1 e n_2 que serão os extremos do intervalo de confiança.

Quando o tamanho da amostra inicial n_0 for maior que 30, pode-se ainda obter um intervalo de confiança para a variável $\underline{n}$ com o uso da aproximação normal, conforme se segue:

Seja

$$Z = \frac{n - E(n)}{\sqrt{V(n)}}$$

Pelo Teorema do Limite Central, a variável Z tende para uma Normal padronizada, quando o tamanho da amostra cresce. Assim tem-se:

$$Z = \frac{n - c}{\sqrt{\frac{2c^2}{(n_0-1)}}} \sim N(0, 1)$$

deste modo, estribado na probabilidade:

$$P \left\{ -Z \leq \frac{n - c}{\sqrt{\frac{2c^2}{(n_0 - 1)}}} \leq Z \right\} = 1 - \alpha$$

pode-se obter um intervalo de confiança para a variável a leatória n , cujos extremos são dados por:

$$I.C.(n): c \pm Z_{\alpha/2} \cdot \sqrt{\frac{2c^2}{(n_0 - 1)}}$$

onde $Z_{\alpha/2}$ é um limite superior da distribuição normal padronizada, que define uma área $\alpha/2$.

Esta técnica de dimensionamento de uma A mostra Casual Simples ora apresentada, estende-se para as demais técnicas da amostragem, como se verá a seguir.

Um outro procedimento para a obtenção da distribuição da variável n seria o de considerar-se a se mi-amplitude do I.C. como uma fração da média amostral.

Far-se-á, apenas para a técnica da A.C.S., um estudo deste caso com o intuito de incentivar novos trabalhos a esse respeito.

Quando se trabalha com a precisão $d = K_1 \cdot \bar{y}$ o problema da falta de conhecimento dos parâmetros torna-se ainda mais grave, como será visto a seguir.

Como foi visto em (4.2) o valor de n para Amostragem Casual Simples é dado por

$$n = \frac{t^2 \cdot s^2}{d^2}$$

e neste caso

$$n = \frac{t^2 \cdot s^2}{K_1^2 \cdot \bar{y}^2}$$

Fixando o tamanho da amostra inicial e o nível α de probabilidade tem-se que:

$$b = \frac{t^2}{K_1^2}$$

é uma constante e assim

$$n = b \cdot \frac{s^2}{\bar{y}^2}$$

Por analogia ao já visto, pode-se dizer que se $Y \sim N(\mu, \sigma^2)$ então:

$$\frac{n}{n_0} = \frac{b \cdot s^2}{n_0 \cdot \bar{y}^2} = b \cdot \frac{\frac{(n_0-1)s^2}{(n_0-1)\sigma^2}}{\frac{\bar{y}^2}{\sigma^2/n_0}}$$

Mas como se sabe:

$$\frac{(n_0-1)s^2}{\sigma^2} \sim \chi^2_{(n_0-1)}$$

e ainda,

$$\bar{y} \sim N(\mu, \sigma^2/n_0)$$

Assim tem-se:

$$\frac{\bar{y}}{\frac{\sigma}{\sqrt{n_0}}} \sim N(\mu', 1) \quad \text{onde} \quad \mu' = \frac{\mu}{\frac{\sigma}{\sqrt{n_0}}}$$

Deste modo pode-se afirmar que:

$$\frac{\bar{y}^2}{\frac{\sigma^2}{n_0}} \sim \chi^2_{(1, \lambda)}$$

onde $\chi^2_{(1, \lambda)}$ é uma qui-quadrado não central, com parâmetro de não centralidade:

$$\lambda = \frac{1}{2} \cdot \frac{\mu^2}{\frac{\sigma^2}{n_0}} = \frac{n_0 \cdot \mu^2}{2\sigma^2}$$

Tomando-se o inverso da variável aleatória estudada tem-se:

$$\frac{n_0}{n} = \frac{1}{b} \cdot \frac{\frac{\bar{y}^2}{\frac{\sigma^2}{n_0}}}{\frac{(n_0-1)s^2}{(n_0-1)\sigma^2}} \sim \frac{1}{b} \cdot F'_{(1, n_0-1, \lambda)}$$

onde $F'_{(1, n_0-1, \lambda)}$ é uma distribuição F não central, com parâmetro da não centralidade igual a λ .

Deste modo,

$$n \sim n_0 \cdot b \cdot \frac{1}{F'_{(1, n_0-1, \lambda)}}$$

Portanto a variável aleatória n (tamanho de amostra) tem distribuição do inverso de uma F não central, com 1 e n_0-1 graus de liberdade e parâmetro de não centralidade λ , multiplicada por uma constante que só depende do tamanho da amostra inicial e do valor de K_1 desejado.

Analogamente ao caso anterior, quando se trabalha com populações normais padronizadas, deve-se obter a mediana desta variável aleatória.

Como se sabe, se a mediana de uma variável aleatória $Y \sim F_{(n_1, n_2)}$ é igual à $\underline{md}$ então a mediana de uma variável $X = 1/Y$, que é o inverso de Y , é dada pelo inverso da mediana de Y , logo a mediana de X é igual à $1/\underline{md}$.

Assim tem-se que se $Y \sim F'_{(1, n_0-1, \lambda)}$ então

$$P(Y \leq \underline{md}') = \int_0^{\underline{md}'} f(y) dy = \frac{1}{2}$$

onde $\underline{md}'$ é a mediana de Y .

Portanto a mediana da variável aleatória n (tamanho da amostra) é dada por:

$$\text{Mediana } (n) = n_o . b . \frac{1}{md'}$$

Pela escassez de tabelas de F não central, com nível de probabilidade $\alpha = 0,5$, torna-se difícil a obtenção desta mediana.

5.3. Amostragem Estratificada

Conforme já visto em (4.3) a equação de n (tamanho de amostra), numa partilha proporcional é dada por:

$$n = \frac{t^2}{d^2} \cdot \sum_{h=1}^L W_h s_h^2$$

Fixando-se o tamanho da amostra inicial nos estratos e o nível α de probabilidade, então o valor de t , como já visto, será uma constante. Assim ao fixar-se $d = K$ obtém-se a constante

$$a = \frac{1}{V} = \frac{t^2}{d^2} = \frac{t^2}{K^2}$$

Consequentemente

$$n = a \cdot \sum_h W_h s_h^2 ,$$

isto é, n é função da soma ponderada de variáveis aleatórias, multiplicada por uma constante.

Supondo-se que dentro de cada estrato h coleta-se uma Amostra Casual Simples de uma população normalmente distribuída com média zero e variância um ($Y_h \sim N(0,1)$).

Então para cada estrato tem-se:

$$(nh - 1) \cdot s_h^2 \sim \chi^2_{(nh-1)} .$$

Deste modo, baseando-se em SATTERTHWAITTE (1946), para a totalidade dos estratos, virá que:

$$Q = n^* \cdot \sum_{h=1}^L W_h s_h^2$$

tem distribuição aproximada de $\chi^2_{n^*}$, onde o número de graus de liberdade

$$n^* = \sum_{h=1}^L (nh-1) = n_0 - L .$$

Reportando à expressão de n , tem-se:

$$Q = n^* \cdot \sum_h W_h s_h^2 = \frac{n^* \cdot n}{a} \sim \chi^2_{n^*}$$

Daí obtém-se:

$$n = \frac{a \cdot Q}{n^*} \sim \frac{a}{n^*} \chi_{n^*}^2$$

Portanto a variável aleatória n tem distribuição aproximada de $\chi_{n^*}^2$ (qui-quadrado com $n_0 - L$ g.l.), multiplicada por constantes que dependem exclusivamente do tamanho da amostra inicial e do valor de $\underline{d}$ desejado.

Com isto pode-se obter os valores da média e variância da variável $\underline{n}$, expressos por:

$$E(n) = E \left[\frac{a \cdot Q}{n^*} \right] = \frac{a}{n^*} E(Q) = \frac{a}{n^*} n^* = a$$

$$V(n) = V \left[\frac{a \cdot Q}{n^*} \right] = \frac{a^2}{n^{*2}} V(Q) = \frac{a^2}{n^{*2}} 2n^* = \frac{2 \cdot a^2}{n^*}$$

Por outro lado, adotando-se $d = K_1 \cdot \bar{y}_{est}$ e novamente fixando-se o tamanho da amostra inicial e o nível α de probabilidade, tem-se de maneira análoga:

$$n = \frac{t^2}{d^2} \frac{\sum W h^2 s h^2}{h} = \frac{t^2}{K_1^2} \frac{\sum W h^2 s h^2}{\bar{y}_{est}^2} \quad \text{onde}$$

$$b = \frac{t^2}{K_1^2} \quad \text{é uma constante.}$$

Supondo, mais uma vez, que as subpopulações dos estratos são normais padronizadas, então para cada estrato h , tem-se:

$$\bar{y}_h \sim N(0, 1/nh) \quad \text{e portanto}$$

$$\bar{y}_{est} = \sum_h W_h \bar{y}_h \sim N(0, \sum_h \frac{W_h^2}{nh})$$

Conclui-se então que:

$$\frac{\bar{y}_{est}}{\sqrt{\sum_h \frac{W_h^2}{nh}}} \sim N(0, 1)$$

e deste modo

$$\frac{\bar{y}_{est}^2}{\sum_h \frac{W_h^2}{nh}} \sim \chi^2_{(1)}$$

Como dentro de cada estrato $\bar{y}_h$ e sh^2 são independentes sob normalidade, e por consequência $\bar{y}_{est}$ independe das variâncias dos estratos, então com o uso da aproximação:

$$n^* = \sum_h W_h sh^2 \sim \chi^2_{n^*}$$

pode-se dizer que:

$$F = \frac{\frac{n^* \cdot \sum W_h sh^2}{n^*}}{\frac{\bar{y}_{est}^2}{\sum \frac{Wh^2}{nh}}} \sim F(n^*, 1)$$

Relembrando a expressão de

$$n = b \cdot \frac{\sum W_h^2 sh^2}{\bar{y}_{est}^2}$$

pode-se ver que:

$$F = \frac{1}{b} \cdot \sum \frac{Wh^2}{nh} \cdot n \sim F(n^*, 1) \quad , \quad \text{donde}$$

$$n = b \cdot \frac{1}{\sum \frac{Wh^2}{nh}} \cdot F \sim \frac{b}{\sum \frac{Wh^2}{nh}} \cdot F(n^*, 1)$$

Portanto a variável aleatória n possui distribuição de F-Snedecor com n^* e 1 g.l., multiplicada por uma constante que só depende do tamanho da amostra inicial nos estratos, do peso dos estratos e do valor desejado da proporção K_1 da média.

Por extensão ao já visto na Amostragem Casual Simples, pode-se obter a mediana desta variável, que será dada por:

$$\text{mediana } (n) = \frac{b}{\sum \frac{Wh^2}{nh}} \cdot md$$

onde md é a mediana de uma variável y com distribuição $F_{(n^*, 1)}$, ou seja:

$$P(Y \leq md) = \int_0^{md} f(y) dy = \frac{1}{2}$$

onde $f(y) \sim F_{(n^*, 1)}$

A obtenção da distribuição de probabilidade da variável n , quando os estratos são subpopulações normalmente distribuídas com média e variância desconhecidas, pode ser realizada de maneira análoga ao caso da A.C.S.

Fixando $d = K$ onde K é constante e também o tamanho da amostra piloto nos estratos, tem-se:

$$n = a \cdot \sum_h W_h sh^2$$

onde $a = \frac{t^2}{K^2}$, como já visto.

Uma vez que $Y_h \sim N(\mu_h, \sigma_h^2)$ resulta:

$$\frac{(nh-1)sh^2}{\sigma_h^2} \sim \chi^2_{(nh-1)}$$

Baseando-se novamente na aproximação de SATTERTHWAITTE (1946), a variável

$$Q_1 = \frac{n^* \cdot \sum W_h sh^2}{\sigma^2}$$

possui distribuição aproximada de qui-quadrado com n^* g.l., onde $n^* = n_0 - L$ conforme já visto e σ^2 é a variância populacional.

Deste modo,

$$Q_1 = \frac{n^*}{\sigma^2 \cdot a} (a \cdot \sum W h^2 s h^2) = \frac{n^*}{\sigma^2 \cdot a} n \sim \chi_{n^*}^2$$

e como consequência:

$$n = \sigma^2 \frac{a}{n^*} Q_1 \sim \sigma^2 \frac{a}{n^*} \chi_{n^*}^2$$

Como não se conhece o valor da variância populacional σ^2 , não é possível obter a distribuição de probabilidade da variável n pelo método usual.

Utilizando a técnica já apresentada na A.C.S. pode-se contornar o problema, tomando-se $d = k_2 \sigma$. Assim tem-se:

$$n = \frac{t^2}{d^2} \sum W h^2 s h^2 = \frac{t^2}{K_2^2} \cdot \frac{\sum W h^2 s h^2}{\sigma^2}$$

Fixando n_0 e consequentemente t , resulta ser $c = \frac{t^2}{K_2^2}$ uma constante, e portanto

$$n = c \cdot \frac{\sum W h^2 s h^2}{\sigma^2}$$

Por analogia ao já visto:

$$n = \frac{c}{n^*} Q_1 \sim \frac{c}{n^*} \chi_{n^*}^2 .$$

Com isto, pode-se obter os valores da média e variância da variável n , como se segue:

$$E(n) = E \left[\frac{c \cdot Q_1}{n^*} \right] = c$$

$$V(n) = V \left[\frac{c \cdot Q_1}{n^*} \right] = \frac{2 \cdot c^2}{n^*} .$$

De maneira análoga ao caso da A.C.S., pode-se obter intervalos de confiança para variável aleató-ria n , através da probabilidade:

$$P \left\{ \frac{n^*}{c} n_1 \leq Q_1 \leq \frac{n^*}{c} n_2 \right\} = 1-\alpha .$$

Como Q_1 possui distribuição de qui-quadrado, amplamente conhecida e tabelada, pode-se facilmente obter os extremos n_1 e n_2 do intervalo de confiança, a um nível $1-\alpha$ de confiança.

Pode-se ainda, no caso de grandes amos-tras iniciais, obter uma aproximação normal para este intervalo, ou seja:

$$P \left\{ -Z \leq \frac{n - c}{\sqrt{\frac{2 \cdot c^2}{n^*}}} \leq Z \right\} = 1 - \alpha$$

Assim, a um nível α de probabilidade, os extremos do intervalo de confiança serão dados por:

$$I.C.(n) : c \pm Z_{\alpha/2} \cdot \sqrt{\frac{2 \cdot c^2}{n^*}}$$

onde $Z_{\alpha/2}$ é limite superior da distribuição normal padronizada, a um nível $\alpha/2$ de probabilidade.

5.4. Amostragem Sistemática

Como já visto em 4.4, ao considerar-se a população como aleatória, todos os resultados obtidos na Amostragem Sistemática podem ser estudados através de uma Amostragem Casual Simples e portanto não existe sentido em deduzir-se novamente a distribuição da variável aleatória n (tamanho de amostra) para este caso.

5.5. Amostragem por Conglomerados

Uma vez obtida uma equação que relaciona o valor do tamanho da amostra com o grau de precisão adotado, pode-se estender os resultados obtidos pela técnica de A.C. S. para o caso da Amostragem por Conglomerados em estágio único.

Fixando o tamanho da amostra inicial e o nível α de probabilidade, resulta um valor de t constante, e ainda fixando o desvio $d = K$ tem-se que

$$\frac{t^2}{K^2} = a \quad \text{é uma constante.}$$

Assim a expressão (4.5) fica:

$$n = a \quad s_1^2$$

onde s_1^2 é a única variável aleatória.

Supondo uma população normalmente distribuída, com média zero e variância um ($Y \sim N(0, 1)$) tem-se:

$$Y_{i.} = \sum_{j=1}^M Y_{ij} \sim N(0, M) \quad ,$$

o que é facilmente comprovável com o uso da função geradora de momentos da distribuição normal e propriedades de variáveis aleatórias independentes.

Sabe-se que:

$$\frac{(n_0 - 1) s_1^2}{M} \sim \chi^2_{(n_0 - 1)} \quad \text{(qui-quadrado com } (n_0 - 1) \text{ g.l.)}$$

e conseqüentemente:

$$Q = \frac{(n_0-1)s_1^2}{M} = \frac{(n_0-1)}{M} \cdot \frac{n}{a} \sim \chi^2_{(n_0-1)} .$$

Logo,

$$n = \frac{a \cdot M}{(n_0-1)} Q \sim \frac{a \cdot M}{(n_0-1)} \chi^2_{(n_0-1)} .$$

Deste modo a variável aleatória n tem distribuição de probabilidade de qui-quadrado com (n_0-1) g.l., multiplicada por constantes que dependem do tamanho n_0 da amostra inicial, do número M de unidades secundárias e ainda do valor de d desejado.

Com isto pode-se obter o valor médio da variável n , que será dado por:

$$E(n) = E \left[\frac{a \cdot M}{(n_0-1)} Q \right] = \frac{a \cdot M}{(n_0-1)} E(Q) = \frac{a \cdot M}{(n_0-1)} (n_0-1) = a \cdot M .$$

Pode-se também obter sua variância, ou seja:

$$\begin{aligned} V(n) &= V \left[\frac{a \cdot M}{(n_0-1)} Q \right] = \frac{a^2 \cdot M^2}{(n_0-1)^2} V(Q) = \frac{a^2 \cdot M^2}{(n_0-1)^2} 2(n_0-1) = \\ &= 2 \frac{a^2 \cdot M^2}{(n_0-1)} . \end{aligned}$$

Por outro lado, adotando o valor da semi-amplitude do intervalo de confiança como fração da média amostral ($d = K_1 \cdot \bar{y}$), e novamente fixando o tamanho da amostra inicial n_0 e o nível α de probabilidade tem-se, de maneira análoga ao já visto, que:

$$n = \frac{t^2 \cdot s_1^2}{d^2} = \frac{t^2 \cdot s_1^2}{K_1^2 \cdot \bar{y}^2}, \quad \text{onde}$$

$$b = \frac{t^2}{K_1^2} \quad \text{é uma constante.}$$

Supondo novamente uma população normal padronizada, virá:

$$\frac{(n_0 - 1) s_1^2}{M} \sim \chi^2_{(n_0 - 1)}$$

e ainda pode-se afirmar que, $\bar{y} \sim N(0, M/n_0)$. Assim,

$$\frac{\bar{y}}{\sqrt{\frac{M}{n_0}}} \sim N(0, 1)$$

e por sua vez,

$$\left[\frac{\bar{y}}{\sqrt{\frac{M}{n_0}}} \right]^2 = \frac{n_0 \bar{y}^2}{M} \sim \chi^2_{(1)}.$$

Como nos casos anteriores $\bar{y}$ e s_1^2 são variáveis aleatórias independentes sob normalidade e deste modo pode-se estruturar:

$$F = \frac{\frac{(n_0-1)s_1^2}{(n_0-1)M}}{\frac{n_0 \bar{y}^2}{M}} \sim F_{(n_0-1, 1)} \quad \text{logo,}$$

$$F = \frac{s_1^2}{n_0 \cdot \bar{y}^2} = \frac{1}{n_0 \cdot b} \left(\frac{b \cdot s_1^2}{\bar{y}^2} \right) = \frac{n}{n_0 \cdot b}$$

portanto,

$$n = b \cdot n_0 \cdot F \sim b \cdot n_0 \cdot F_{(n_0-1, 1)}.$$

Como se verifica, chega-se à mesma equação do caso de Amostragem Casual Simples, e portanto pode-se estender a técnica da obtenção de mediana abordada na A.C.S. para a Amostragem por Conglomerado, ou seja,

$$\text{mediana } (n) = n_0 \cdot b \cdot md$$

onde md é a mediana de uma variável com distribuição

$$F_{(n_0-1, 1)},$$

ou seja:

$$\int_0^{md} f(y) dy = \frac{1}{2} \quad , \quad \text{onde } f(y) \sim F_{(n_0-1, 1)} \quad .$$

A extensão destes resultados, para população com distribuição normal com média e variância desconhecidas, é feita de maneira análoga ao caso da Amostra Casual Simples.

Fixando o tamanho da amostra inicial n_0 e o valor de $d = K$, como no caso anterior, tem-se que $n = a.s_1^2$, mas agora $Y \sim N(\mu, \sigma^2)$ e portanto:

$$y_{i..} \sim N(M\mu, M\sigma^2) \quad ,$$

então como se sabe,

$$Q = \frac{(n_0-1)s_1^2}{M \sigma_1^2} \sim \chi^2_{(n_0-1)}$$

e analogamente aos casos anteriores

$$n \sim \frac{\sigma^2 \cdot M \cdot a}{(n_0-1)} \cdot \chi^2_{(n_0-1)}$$

Como não se conhece o valor da variância populacional σ^2 não se pode obter a distribuição da variável n .

Este problema pode ser resolvido com o uso de técnica já apresentada na A.C.S., ou seja, adotando o valor da semi-amplitude do intervalo de confiança d como parte do desvio padrão populacional $d = K_2 \cdot \sigma$, onde K_2 é uma constante, e assim tem-se:

$$n = \frac{t^2 \cdot s^2}{d^2} = \frac{t^2 \cdot s_1^2}{K_2^2 \cdot \sigma^2} .$$

Como n_0 e conseqüentemente t estão fixados, tem-se:

$$c = \frac{t^2}{K_2^2} \quad \text{é uma constante, e então}$$

$$n = c \cdot \frac{s_1^2}{\sigma^2} .$$

Por analogia ao caso anterior obtém-se:

$$n = \frac{M \cdot c \cdot Q_1}{(n_0 - 1)} \sim \frac{M \cdot c}{(n_0 - 1)} \chi^2_{(n_0 - 1)} .$$

Com isto pode-se obter, os valores da média e variância da variável aleatória n , como se segue:

$$E(n) = E \left[\frac{M \cdot c \cdot Q_1}{(n_0 - 1)} \right] = M \cdot c$$

$$V(n) = V \left[\frac{M \cdot c \cdot Q_1}{(n_0 - 1)} \right] = \frac{2 \cdot M^2 \cdot c^2}{(n_0 - 1)}$$

De maneira análoga ao já visto na A.C.S. pode-se obter intervalos de confiança para a variável n obtida pela técnica de Amostragem por Conglomerados, do seguinte modo:

Toma-se a probabilidade:

$$P \{ n_1 \leq n \leq n_2 \} = 1 - \alpha \implies$$

$$P \left\{ n_1 \leq \frac{M \cdot c \cdot Q_1}{(n_0 - 1)} \leq n_2 \right\} = 1 - \alpha \implies$$

$$P \left\{ \frac{(n_0 - 1)}{M \cdot c} n_1 \leq Q_1 \leq \frac{(n_0 - 1)}{M \cdot c} n_2 \right\} = 1 - \alpha .$$

Como Q_1 tem distribuição de qui-quadrado, pode-se encontrar os valores de n_1 e n_2 que são os extremos do intervalo de confiança com um nível α de probabilidade.

Para amostras grandes é possível obter-se uma aproximação para o intervalo, fornecida pela distribuição normal, com o uso da seguinte probabilidade:

$$P \left\{ -Z \leq \frac{n - M \cdot c}{\sqrt{\frac{2 M^2 c^2}{(n_0 - 1)}}} \leq Z \right\} = 1 - \alpha .$$

Assim, com um nível de confiança $1-\alpha$, os extremos do intervalo serão dados por:

$$I.C.(n) : M.C \pm Z_{\alpha/2} \cdot \sqrt{\frac{2 M^2 C^2}{(n_0-1)}}$$

onde $Z_{\alpha/2}$ é o limite superior da distribuição normal padronizada, com um nível $\alpha/2$ de probabilidade.

6. ILUSTRAÇÃO DAS TÉCNICAS PROPOSTAS

Serão apresentados, para cada uma das técnicas de amostragem abordadas, exemplos dos dois métodos usuais de dimensionamento e também da técnica proposta neste trabalho.

6.1. Amostragem Casual Simples

Dos dados vistos na Tabela 1 obteve-se as seguintes estimativas de parâmetros:

$$\bar{y} = 7,52 \quad ; \quad s^2 = 6,8437$$

Com base nestas estimativas obtém-se os tamanhos de amostras, pelos 3 métodos abordados, como se segue:

6.1.1. Fixando $d = K = 0,70$

Com um nível $\alpha = 0,05$ de probabilidade e como $n_0 = 36$ observações, tem-se que $t_{35,5\%} = 2,03$. Então:

$$n = \frac{t^2 s^2}{d^2} = \frac{(2,03)^2 \cdot 6,8437}{(0,70)^2} = 57,6 \approx 58 \text{ observações}$$

6.1.2. Adotando $d = K_1 \bar{y} = 0,10 \cdot 7,52 = 0,752$

$$n = \frac{t^2 s^2}{(K_1 \bar{y})^2} = \frac{(2,03)^2 \cdot 6,8437}{(0,752)^2} = 49,9 \approx 50 \text{ observações}$$

6.1.3. Fazendo uso da técnica proposta e tomando $d = K_2 \sigma = 0,25\sigma$

Então pode-se obter o valor médio e a variância da variável n que são dadas, conforme já visto no capítulo anterior, por:

$$E(n) = c = \frac{(2,03)^2}{(0,25)^2} = 65,9 \approx 66 \text{ observações}$$

$$V(n) = \frac{2 \cdot c^2}{(n_0 - 1)} = \frac{2(65,9)^2}{(36 - 1)} = 248,16$$

Pode-se também obter um intervalo com nível $1-\alpha = 0,95$ de confiança para n , com o uso da probabilidade:

$$P \left\{ \frac{(n_0-1)}{c} n_1 \leq Q_1 \leq \frac{(n_0-1)}{c} n_2 \right\} = 0,95$$

A partir de uma tabela de qui-quadrado obtêm-se os valores da $\chi^2_{35;2,5\%}$ e $\chi^2_{35;97,5\%}$ e então tem-se:

$$P \{20,57 \leq Q_1 \leq 53,20\} = 0,95, \text{ e assim,}$$

$$\frac{(n_0-1)}{c} n_1 = 20,57 \Rightarrow n_1 = \frac{65,9}{35} \cdot 20,57 = 38,73$$

$$\frac{(n_0-1)}{c} n_2 = 53,20 \Rightarrow n_2 = \frac{65,9}{35} \cdot 53,20 = 100,17$$

Logo um intervalo de confiança para n com nível $\alpha = 0,05$ de probabilidade é dado por:

$$(38 \leq n \leq 101) .$$

Como $(n_0-1) \geq 30$ pode-se ainda obter uma aproximação normal para o intervalo de confiança de n .

Para $\alpha = 0,05$, tem-se que $Z_{\alpha/2} = 1,96$ e assim,

$$\text{I.C.}(n): c \pm z_{\alpha/2} \sqrt{\frac{2 \cdot c^2}{(n_0 - 1)}}, \text{ é dado por:}$$

$$(35 \leq n \leq 97) \quad .$$

Como se vê, os resultados obtidos nos dois intervalos são muito próximos, o que mostra que a aproximação normal é razoável.

Na comparação dos resultados obtidos pelas 2 técnicas usuais e a técnica proposta, verifica-se que eles não se distanciam muito. A vantagem da técnica proposta é de que uma vez fixada a precisão (tamanho da amostra inicial n_0 , nível α de probabilidade e o valor de K_2) qualquer pesquisador obterá o mesmo valor para tamanho de amostra em qualquer época, sem ser necessária a retirada da amostra piloto.

6.2. Amostragem Estratificada

Dos dados vistos na Tabela 2 obtiveram-se as seguintes estimativas, para os 5 estratos:

ESTRATOS	CLASSES	$\bar{y}_h$	sh^2
1	< 43,9	17,5	132,27
2	43,9 - 126,4	83,3	669,36
3	126,4 - 366,4	238,0	6623,60
4	366,4 - 948,3	587,2	32362,97
5	$\geq$ 948,3	1981,0	1476520,50

$$\bar{y}_{est} = 259,67 \quad ; \quad \sum_{h=1}^5 W_h sh^2 = 103006,609$$

A partir destas estimativas obtiveram-se os dimensionamentos pelos 3 métodos.

6.2.1. Fixando $d = K = 20$

Com um nível $\alpha = 0,05$ de probabilidade, o valor de t com $n_0 - L = 35$ g.l. é dado por $t_{35, 5\%} = 2,030$

$$e \quad n = \frac{t^2}{d^2} (\sum_h W_h sh^2) = \frac{(2,030)^2}{(20)^2} \cdot 103006,3343 = 1061,20.$$

Ajustando o tamanho da amostra para população finita para $N = 300$ tem-se:

$$n' = \frac{n}{1 + \frac{n}{N}} = \frac{1061,20}{1 + \frac{1061,20}{300}} = 233,9 \approx 234 \text{ observações}$$

6.2.2. Adotando $d = K_1 \bar{y}_{est} = 0,10.259,67=25,967$

$$n = \frac{t^2}{(K_1 \bar{y}_{est})^2} (\sum \frac{W_h s_h^2}{h}) = \frac{(2,030)^2}{(25,967)^2} \cdot 103006,3343=629,52$$

Ajustando para população finita tem-se:

$$n' = \frac{629,52}{1 + \frac{629,52}{300}} = 203,2 \approx 204 \text{ observações}$$

6.2.3. Fazendo uso da técnica proposta e tomando
 $d = K_2 \sigma = 0,05\sigma$

Então o valor médio da variável aleatória
 dada por

$$E(n) = c = \frac{(2,030)^2}{(0,05)^2} = 1648,36$$

Ajustando para população finita tem-se:

$$n' = \frac{1648,36}{1 + \frac{1648,36}{300}} = 253,8 \approx 254 \text{ observações}$$

Pode-se agora obter um intervalo de confiança para variável n , com o uso da probabilidade:

$$P \left\{ \frac{n^*}{c} n_1 \leq Q_1 \leq \frac{n^*}{c} n_2 \right\} = 1-\alpha$$

Utilizando-se da tabela de qui-quadrado ob-
têm-se os valores de $\chi^2_{35;2,5\%}$ e $\chi^2_{35;97,5\%}$ e assim tem-
-se:

$$P \{ 20,57 \leq Q_1 \leq 53,20 \} = 0,95 , \text{ e portanto}$$

$$\frac{n^*}{c} n_1 = 20,57 \Rightarrow n_1 = \frac{1648,36}{35} \cdot 20,57 = 968,76$$

$$\frac{n^*}{c} n_2 = 53,20 \Rightarrow n_2 = \frac{1648,36}{35} \cdot 53,20 = 2505,51 .$$

Logo um intervalo com nível de confiança
 $1-\alpha = 0,95$ e dado por:

$$(968 \leq n \leq 2506) .$$

Ajustando este intervalo para população
finita tem-se:

$$n'_1 = \frac{968,76}{1 + \frac{968,76}{300}} = 229,06$$

$$n'_2 = \frac{2505,51}{1 + \frac{2505,51}{300}} = 267,92$$

e o intervalo é dado por:

$$(229 \leq n' \leq 268) .$$

6.3. Amostragem Sistemática

Com os dados da Tabela 3 obtiveram-se as seguintes estimativas de parâmetros:

$$\bar{y} = 188,1 \quad ; \quad s^2 = 117.257,72$$

A partir destas estimativas e considerando a amostra como uma casual simples, dimensiona-se a amostra pelos 3 métodos:

6.3.1. Fixando $d = K = 20$

Com um nível $\alpha = 0,05$ de probabilidade, o valor de t com $n_0 - 1 = 29$ g.l. é dado por $t_{29, 5\%} = 2,045$ e portanto:

$$n = \frac{t^2 s^2}{d^2} = \frac{(2,045)^2}{(20)^2} \cdot 117.257,72 = 1225,94.$$

Ajustando para população finita com $N = 300$ tem-se:

$$n' = \frac{n}{1 + \frac{n}{N}} = \frac{1225,94}{1 + \frac{1225,94}{300}} = 241,0 \text{ observações}$$

6.3.2. Adotando $d = K_1 \bar{y} = 0,10 \cdot 188,1 = 18,81$

$$n = \frac{t^2}{(K_1 \bar{y})^2} s^2 = \frac{(2,045)^2}{(18,81)^2} \cdot 117.257,72 = 1385,96$$

Ajustando para população finita

$$n' = \frac{1385,96}{1 + \frac{1385,96}{300}} = 246,6 \approx 247 \text{ observações}$$

6.3.3. Usando a técnica proposta e tomando $d = K_2 \sigma = 0,05\sigma$

Então o valor médio da variável aleatória

n é dado por:

$$E(n) = c = \frac{(2,045)^2}{(0,05)^2} = 1672,81$$

Ajustando para população finita,

$$n' = \frac{1672,81}{1 + \frac{1672,81}{300}} = 254,4 \approx 255 \text{ observações .}$$

Um intervalo de confiança para variável aleatória n é obtido com o uso da probabilidade

$$P \left\{ \frac{(n_0-1)}{c} n_1 \leq n \leq \frac{(n_0-1)}{c} n_2 \right\} = 1-\alpha$$

A partir de uma tabela de qui-quadrado, com $n_0-1 = 29$ g.l. e nível $\alpha = 0,05$ de probabilidade obtém-se:

$$P \{ 16,05 \leq n \leq 45,72 \} = 0,95$$

e daí tem-se:

$$\frac{(n_0-1)n_1}{c} = 16,05 \rightarrow n_1 = \frac{1672,81}{29} \cdot 16,05 = 925,81$$

$$\frac{(n_0-1)n_2}{c} = 45,72 \Rightarrow n_2 = \frac{1672,81}{29} \cdot 45,72 = 2637,27$$

Deste modo um intervalo de confiança com nível de confiança $1-\alpha = 0,95$ é dado por:

$$(925 \leq n \leq 2638)$$

Este intervalo pode ser ajustado para populações finitas, tomando-se:

$$n'_1 = \frac{925,81}{1 + \frac{925,81}{300}} = 226,58$$

$$n'_2 = \frac{2637,27}{1 + \frac{2637,27}{300}} = 269,36$$

e deste modo o novo intervalo é o seguinte:

$$(226 \leq n' \leq 270) \text{ .}$$

Como se observa os resultados obtidos na Amostragem Sistemática são muito próximos aos vistos na Amostragem Estratificada.

No caso da técnica proposta os resultados são praticamente idênticos, como era de se esperar, pois esta técnica não depende da amostra inicial coletada e sim da precisão adotada para o levantamento.

6.4. Amostragem por Conglomerados

A partir da Tabela 4 obtiveram-se as seguintes estimativas, para os 15 conglomerados:

$$\bar{y} = 603,15 \quad ; \quad s_1^2 = 16.283,2741$$

Com base nestas estimativas obtiveram-se os dimensionamentos pelos 3 métodos.

6.4.1. Fixando $d = K = 60$

Com um nível $\alpha = 0,05$ de significância e com $n_0 - 1 = 14$ g.l., então $t_{14,5\%} = 2,145$, daí:

$$n = \frac{t^2 \cdot s^2}{d^2} = \frac{(2,145)^2}{(60)^2} 16.283,2741 = 20,8 \approx 21 \text{ observações}$$

6.4.2. Adotando $d = K_1 \bar{y} = 0,10 \cdot 603,15 = 60,315$

$$n = \frac{t^2}{(K_1 \bar{y})^2} s^2 = \frac{(2,145)^2}{(60,315)^2} \cdot 16.283,2741 = 20,6 \approx 21 \text{ observações}$$

6.4.3. Fazendo uso da técnica proposta e tomando $d = K_2 \sigma = 1,30\sigma$

Então obteve-se o valor médio da variável de n , por:

$$E(n) = M.c = 8 \cdot \frac{(2,145)^2}{(1,30)^2} = 21,78 \approx 22 \text{ observações}$$

Um intervalo de confiança para a variável n com nível $\alpha = 0,05$ de probabilidade é dado por:

$$P \left\{ \frac{(n_0-1)}{M \cdot c} n_1 \leq Q_1 \leq \frac{(n_0-1)}{M \cdot c} n_2 \right\} = 0,95$$

Com o uso de uma tabela da distribuição de qui-quadrado com $n_0-1 = 14$ g.l. obtém-se:

$$P \{ 5,629 \leq Q_1 \leq 26,119 \} = 0,95$$

e portanto,

$$\frac{(n_0-1)}{M \cdot c} n_1 = 5,629 \Rightarrow n_1 = \frac{21,78}{14} \cdot 5,629 = 8,75$$

$$\frac{(n_0-1)}{M \cdot c} n_2 = 26,119 \Rightarrow n_2 = \frac{21,78}{14} \cdot 26,119 = 40,63$$

Logo um intervalo com nível $1-\alpha = 0,95$ de confiança para a variável n é dado por:

$$(8 \leq n \leq 41)$$

Finalmente, de todos estes exemplos, pode-se notar que a técnica proposta tem vantagem sobre as demais, quando se tem ao menos um indício da variância da população, pois uma vez fixada a precisão, o dimensionamento da amostra perde o caráter aleatório e passa a ser determinístico.

7. CONCLUSÕES

Tomando por base os resultados obtidos e supondo que sempre se trabalha com populações normais, pode-se concluir:

7.1. Nos levantamentos por Amostragem Casual Simples a distribuição de probabilidade da variável aleatória n (tamanho da amostra), quando se fixa o tamanho da amostra inicial n_0 , o nível α de probabilidade e o valor K_2 da proporção do desvio padrão, é o seguinte:

$$n \sim \frac{c}{(n_0-1)} \cdot \chi^2_{(n_0-1)}$$

onde

$$c = \left[\frac{t_{\alpha/2; n_0-1}}{K_2} \right]^2 = \text{constante fixada a partir da precisão desejada}$$

Com isto é possível dimensionar uma amostra tomando-se um valor médio para esta variável, que é dado por:

$$E(n) = c .$$

7.2. Considerando a Amostragem Estratificada (Partilha Proporcional) e novamente fixando o tamanho da amostra inicial n_0 , o nível α de probabilidade e valor da proporção K_2 , a distribuição de probabilidade da variável n é dada por:

$$n \sim \frac{c}{(n_0-L)} \cdot \chi^2_{(n_0-L)} , \quad \text{onde}$$

L = número de estratos .

Pode-se então obter o tamanho da amostra por um valor médio de n , como se segue:

$$E(n) = c$$

7.3. Quando se considera a população como aleatória, as conclusões obtidas nos levantamentos por Amostragem Sistemática são idênticas às já vistas na Amostragem Casual Simples.

7.4. Na Amostragem por Conglomerados em estágio único com unidades primárias de mesmo tamanho M , a distribuição da variável aleatória n é a seguinte:

$$n \sim \frac{M \cdot c}{(n_0 - 1)} \cdot \chi^2_{(n_0 - 1)}$$

desde que se fixe o número n_0 de unidades primárias da amostra inicial, o nível α de probabilidade e o valor da proporção K_2 do desvio padrão.

A partir desta distribuição obtém-se o dimensionamento da amostra por um valor médio da variável n , que é dado por:

$$E(n) = M \cdot c$$

7.5. Em todas as técnicas de amostragem probabilística vistas, uma vez que esteja fixada a precisão (tamanho da amostra inicial, nível de probabilidade e o valor da proporção do desvio padrão) adotada para o levantamento, qualquer pesquisador obterá o mesmo tamanho ótimo da amostra, sem que seja necessária a retirada da amostra inicial.

Assim, o dimensionamento da amostra perde a aleatoriedade provocada pela amostra inicial, tornando-se determinístico.

8. BIBLIOGRAFIA

- CÂMPOS, H., 1984. Estatística Aplicada à Experimentação com Cana-de-Açúcar. Piracicaba, FEALQ/USP. 292 p.
- CÂMPOS, H., 1985. Amostragem II (apontamentos de aulas). Piracicaba, ESALQ/USP.
- COCHRAN, W.G., 1965. Técnicas de Amostragem. Tradução. Rio de Janeiro, Fundo de Cultura. 555 p.
- COSTA NETO, P.L.O., 1977. Estatística. São Paulo, Edgard Blücher. 264 p.
- DIXON, W.J. e F.J. MASSEY Jr., 1969. Introduction to Statistical Analysis. Nova York, McGraw-Hill. 638 p.
- GRAYBILL, F.A., 1961. An Introduction to Linear Statistical Models. Nova York, McGraw-Hill. 463 p.

- GRAYBILL, F.A., 1976. Theory and Application of the Linear Model. Belmont, Wadsworth Publishing Company, Inc. 704 p.
- HANSEN, M.H.; W.N. HURWITZ e W.G. MADOW, 1953. Sample Survey Methods and Theory. Vol. 1, Nova York. John Wiley e Sons, Inc. 638 p.
- HOEL, P.G.; S.C. PORT e C.J. STONE, 1978. Introdução à Teoria da Probabilidade. Tradução. Rio de Janeiro. Interciência. 269 p.
- HOEL, P.G., 1980. Estatística Matemática. Tradução. 4ª Edição, Rio de Janeiro. Guanabara Dois. 373 p.
- HOGG, R.V. e A.T. CRAIG, 1978. Introduction to Mathematical Statistics. 4ª Edição, Nova York, Macmillan Publishing. 438 p.
- KISH, L., 1965. Survey Sampling. Nova York, John Wiley. 643 p.
- MACE, A.E., 1964. Sample-Size Determination. Nova York, Reinhold Publishing Corporation. 226 p.
- MEIER, P.L., 1978. Probabilidade: Aplicações à Estatística. Tradução. Rio de Janeiro, Livros Técnicos e Científicos. 391 p.
- MOOD, A.M.; F.A. GRAYBILL e D.C. BOES, 1974. Introduction to the Theory of Statistics. 3ª Edição, Nova York, McGraw-Hill. 564 p.

- MURTEIRA, B.J.F., 1979. Probabilidade e Estatística. Volume I. Lisboa, McGraw-Hill. 387 p.
- RAJ, D., 1968. Sampling Theory. Nova York, McGraw-Hill . 302 p.
- SATTERTHWAITE, F.E., 1946. An Approximate Distribution of Estimates of Variance Components. Biometrics Bulletin, 2: 110-114.
- SEVERO, N.C. e M. ZELEN, 1960. Normal Approximation to the Chi-square and Non-central F Probability Functions. Biometrika, 47: 411-416.
- SUKHATAME, P.V., 1956. Teoria de Encuestas por Muestreo con Aplicaciones. Tradução para o espanhol. Cidade do México, Fundo de Cultura Econômica. 495 p.
- SUKHATAME, P.V. e B.V., 1970. Sampling Theory of Surveys with Applications. 2ª Edição Revisada. Ames, Iowa State University Press. 452 p.
- TIKU, M.L., 1966. A Note on Approximation to the Non-central F Distribution. Biometrika, 53: 606-610.
- YAMANE, T., 1967. Elementary Sampling Theory. Englewood Cliffs, Prentice-Hall, Inc. 405 p.

A P E N D I C E

TABELA I - População de 300 fornecimentos de cana-de-açúcar
(em t.) na região de Piracicaba. Dados ordenados
de forma decrescente pelos fornecimentos.

ALVARO MONTEIRO CRUZ	5790,1
RUBENS AREAS VENANCIO	5249,6
PAULO SILVEIRA COUTINHO	4819,8
ONIZEDI SANTOS FORMOSO	3943,3
ALIATAR MARIANO DE AZEVEDO	3012,3
JOSÉ PAES DE LIMA	2531,7
IRINEU DE AZEVEDO LIMA	2453,9
IZIDORO C. RAPOSO DE ALMEIDA	2299,7
WOLFGANG ECHTERBECK	1879,4
AMORESTINA PESSANHA DA SILVA	1813,7
HENRIQUE DE SOUZA NOGUEIRA	1762,9
VANILDO DA SILVA COSTA	1593,2
NELY MENEZES MUYLAERT	1456,3
GELSON DA CONCEIÇÃO RIBEIRO	1364,1
JÓSE DE OLIVEIRA MEIRELES	1249,0
LAUDELINO BENEVIDES GOMES	1122,9
ALMIR SILVA	1107,8
FRANCISCO DA SILVA VASCONCELOS	1081,1
ANTONIO MIRANDA DE AZEVEDO	976,8
JOÃO LOURENÇO DE SOUZA	948,3
NELSON MOREIRA DE LIMA	904,7
AMARO REIS	878,5
SILVIO DA CUNHA PINTO	831,2
CERLY GOMES DE OLIVEIRA	795,1
NAIR FELIX DE BARROS	774,9
GERALDO PAES	735,6
LUIZ MARQUES SIMÕES	706,1
JOÃO ROSA DA SILVA FILHO	678,5

TABELA I - Continuação.

JOSE RIBEIRO AREAS	675,8
AGUSTINHO MIRANDA DA CONCEIÇÃO	631,1
JOÃO BATISTA DE ABREU RANGEL	628,1
ANTONIO FIDELIS DA SILVA	577,4
FRANCISCO DA C. A. BARBOSA	565,6
JOÃO DE LIMA RANGEL	540,7
JOSÉ DOS SANTOS	533,9
ALFEU NOGUEIRA MACHADO	526,0
WALTER RIBEIRO DA FONSECA	523,9
CARLOS A. MUYLAERT MENEZES	519,7
HELMA DE SOUZA GOMES	505,8
LOPO MEIRELES JUNIOR	476,8
ADERBAL PEREIRA MACHADO	476,3
DORALICIO DE SOUZA	456,9
GERALDO DA CONCEIÇÃO SERAFIM	451,0
JOÃO MATIAS FERREIRA	420,1
ARI DOS SANTOS SIQUEIRA	390,6
BENEDITO RIBEIRO RANGEL	385,4
LUIZ ANDRE GUALDA TERRA	383,8
JULIÃO VIANA DE CARVALHO	379,2
FRANCISCO PEREIRA BARRETO	376,6
SEBASTIAO PANISSET GOUDARD	366,4
DOMINGOS NUNES BARRETO	358,9
WALDOMIRO BENTO DE OLIVEIRA	357,2
RAMIRO RANGEL PAES	336,8
JOSE CARLOS DE AZEVEDO	330,6
JORGE FERREIRA DE SOUZA	328,9
DARIO RIBEIRO BATISTA	326,6
FRANCISCO ARCANJO FERREIRA	318,9
DARCY GOMES LEAL	316,4

TABELA I - Continuação.

NEIDE DUARTE ALVES	306,0
JOÃO SOARES DE SOUZA	303,1
ALMIR ABDALA SALOMÃO	282,6
EZIO ERNESTO DOS SANTOS	281,8
ABEL INACIO DA SILVEIRA	256,1
JOSE VIANA I	250,5
JOSE GERALDO ALVES QUEIROZ	250,1
FRANCISCA FREITAS SARDINHA	246,0
FRANCISCO FERREIRA MACHADO	244,2
AMARO BARRETO MONTEIRO	244,1
CARLOS AUGUSTO DA SILVA	236,0
ELIZARDO FERREIRA	234,9
LENICIO ROSA	223,0
LUIZ CARLOS AREAS	216,4
ROBERTO DOS SANTOS P. GUIMARÃES	215,9
CLEVES RIBEIRO DE SOUZA	210,5
AMARO ZACARIAS DE ABREU	201,5
JOAQUIM PINTO RIBEIRO	193,1
ANTONIO DE SOUZA PINTO I	193,0
ALDEMIR PEREIRA DA SILVA	186,0
MANOEL TOLEDO DE ALMEIDA	185,4
LUIZ OCTAVIO DA COSTA SILVA	182,9
LUIZ FERNANDO BARROS	181,0
HERVAL TAVARES	177,4
FRANCISCO PAES VIANA	173,0
MANOEL BARRETO SALES	172,4
LAERCIO PAES DA SILVA	172,3
JOÃO FRANCISCO GOMES	171,9
ARIALDE COSTA	170,3

TABELA I - Continuação.

FRANCISCO DE SOUZA PINTO FILHO	157,1
JOSE THOMAZ CASTILHO	150,8
SILVIO CUSTODIO DA SILVA	147,1
MARIETA BATISTA MAGALDI	146,5
ORESTES SALES CAVALINE	145,8
JOÃO RIBEIRO DA FONSECA	143,8
HERMES NOGUEIRA PINTO	141,0
JOÃO DOS SANTOS	139,5
JULIO GOMES SOARES	135,8
JOSE CARLOS DA SILVA IIIII	135,5
JOÃO GOMES PEDRA	135,5
ERCILIO GOMES PAES	131,6
INACIO RIBEIRO MOCO	126,4
MARIA ELMA MANHAES DA SILVA	126,2
LUIZ 'CERQUEIRA GUIMARÃES	121,6
EUCLIDES PESSANHA DE SOUZA	121,4
AIDES VENTURA DO NASCIMENTO	121,3
JOSE DOMINGOS DA SILVA	121,3
IRINEU GONÇALVES NOGUEIRA	116,9
MANOEL DOS SANTIS TRINDADE	110,9
ERALDO DA SILVA SANTOS	109,3
EFIGENIA LEMOS DA SILVA	108,9
LUIZ RIBEIRO GOMES	108,4
FLORIANO MANHAES	106,6
ZELIO NOGUEIRA MARTINS	106,3
BENEDITO DA CONCEIÇÃO	105,0
REINALDO DE SOUZA ROSA	103,6
ELÍAS TAVARES DE AZEVEDO FILHO	103,0
JOÃO E GONÇALVES DOS SANTOS	100,9

TABELA I - Continuação.

REINALDO FERNANDES NASCIMENTO	100,9
ARI GENARIO DA SILVA	99,9
MANOEL PINTO NOGUEIRA	99,7
EUCLIDES HEITOR BARRETO	99,2
GERMANO DIOGO M. DOS SANTOS	98,0
EDMAR AZEVEDO	97,0
VALDEMIRO RODRIGUES	93,0
JOSEMARCIA FATIMA C. GONÇALVES	92,8
MANOEL DIONIZIO BATISTA	92,7
UMBERTO BELATRABLE P. SANTOS	92,3
DIRCON DE SIQUEIRA REIS	88,9
JOÃO BATISTA MACHADO	87,8
JOSE ALVES DE SOUZA	87,8
LAVINA RIBEIRO PINTO	87,8
OZIEL PESSANHA DE OLIVEIRA	87,6
CONCEIÇÃO DE MARIA DA SILVA	83,2
EVERALDO SALES	82,0
CLAUDIO CESAR PEIXOTO DA SILVA	80,0
RENATO GOMES RIBEIRO DA SILVA	79,8
JAIME HENRIQUE NOGUEIRA	79,6
ANA FERREIRA MARTINS	79,3
DENILDA MARCELINA DA PAIXÃO	79,3
ALMIR SOUZA BRAGA	78,4
LUIZ SOARES NOGUEIRA	77,2
JOÃO FIRMO DE ARAUJO	75,5
MARIA PEREIRA GOMES DA SILVA	74,0
ERNANDE GONÇALVES DE DEUS	72,5
NILSON PEREIRA RANGEL	72,0
PEDRO JOAQUIM DOS SANTOS	71,8

TABELA I - Continuação.

ALZEMIRO THOMAZ DE LIMA	70,8
RITA DE CASSIA R. DO ROSARIO	69,7
JOÃO CHAGAS DE CARVALHO	69,1
AMARO RODRIGUES PESSANHA	67,4
JOVIANO FERREIRA NETO	67,1
MARIA ELIZA SALES ANDRADE	66,4
AMARO DE SOUZA	65,6
ANTONIO FRANCISCO DE AZEVEDO 2	65,5
ADELINO GOMES BATISTA	65,1
JORGE RANGEL	64,9
ARTHUR RIBEIRO GOMES	64,7
MANOEL PEREIRA VIANA FILHO	64,4
FRANCISCO PESSANHA DE SOUZA I	61,5
JOSE ALENCAR DA SILVA	59,8
RUTH TRINDADE DE SOUZA	58,9
OTAVIO RIBEIRO	58,5
MANOEL AFONSO	58,4
RENIVALDO JOSE VIANA	58,0
MAURINO ALMEIDA	57,9
MANOEL PESSANHA NOGUEIRA	56,8
UNIGENITO INACIO DA FONSECA	55,4
FRANCISCO INACIO DE OLIVEIRA	52,9
CEMILTON ROCHA	52,6
AMARO ALVES BARRETO II	51,2
MARIA JOSE BARCELOS NETO	50,6
LUIZ CHAGAS	50,5
EDUARDO DA SILVA PESSANHA	50,5
CECI VIANA PEREIRA	49,3

TABELA I - Continuação.

JOILSON PUPE DE ALMEIDA	48,1
SANTILHA MARIA SANTOS EDUARDO	47,7
MARIA RANGEL DE LIMA	46,9
MARIO RIBEIRO DE SOUZA	46,6
MARLI DE SOUZA RIBEIRO	46,4
ANTONIO CARLOS DE OLIVEIRA	44,2
GERALDO JOSE DE ALMEIDA	43,9
MARIA DA C.SILVA RANGEL	42,6
MARIA ANTONIA BRANCO PEREIRA	41,8
AMARO SARDINHA FILHO	41,6
JOSE ANTONIO PAES DE ALMEIDA	41,0
MARIA DOLORES G.DE SIQUEIRA	40,3
LAUDELINA RIBEIRO	39,9
JOSE ANTONIO TOLEDO	39,2
FRANCISCO LEONARDO	38,6
PAULO SALES	38,4
AMARO SIQUEIRA	38,2
ROTCHILDES DE SOUZA RIBEIRO	37,9
BENEDITO TAVARES NETO	37,1
UILSON BAJIENSE DA SILVA	35,3
DOMINGOS MELO RANGEL	34,8
AMARO RIBEIRO II	34,7
JOSE CAETANO ROSA	34,0
MARIA ALVES MANHAES	33,6
MARIA MERCEDES SOARES	32,7
AMARO ALCIDES PESSANHA	32,0
ANTONIO PEREIRA VIANA	31,4
AMARO BARBOSA DA SILVA	31,2
UIZ DE SOUZA LEAL	30,8

TABELA I - Continuação.

LAUDELINA FERREIRA DE FREITAS	30,3
JORGE FRANCISCO DE SOUZA I	29,9
VALDEMAR ROCHA	29,6
FRANCISCO LIMA	29,5
AMARO PINHEIRO	28,3
JOAREZ MAIA GONÇALVES	28,3
SILVIO SIQUEIRA	28,1
ADELINA RANGEL RIBEIRO	28,0
AMARO DE ALMEIDA TRINDADE	27,4
MANOEL DA SILVA SALES	27,2
EDILSA PESSANHA DE ARAUJO	27,2
ANTONIO MARINS	27,0
SILVIO R. TEIXEIRA TAVARES	26,0
FRANCISCO BARROSO DE SIQUEIRA	25,4
COLY FERREIRA DE SIQUEIRA	25,3
CLAUDIO LIZANDRO GODOY	24,4
MARIA IZABEL FERREIRA DA SILVA	24,3
MARIA DA PENHA DA SILVA	24,0
JOSE MANOEL ALVES FIUZA	23,7
JOSE LICIO RIBEIRO MANHAES	23,2
LICIO GONÇALVES DE SOUZA	22,6
FRANCISCO DE SOUZA	22,5
MARIA DE LOURDES MONTEIRO	22,4
DIONICIO RIBEIRO DE CARVALHO	22,0
ANTONIO RATINHA FILHO	21,7
SEBASTIÃO DE O.MACIEL E SILVA	21,3
NILZETE DA ROCHA BATISTA	20,8
AMARO BRAGA DE AZEREDO	20,2
JOSE ROMARIO PEREIRA GOMES	19,7

TABELA I - Continuação.

EROTILDE AZEVEDO	19,3
LUIZ FERNANDO MARÇAL PEREIRA	19,2
JUVENAL DE SOUZA PINTO	19,1
JOSE EUZEBIO	19,1
RENATO CEZAR PESSANHA MANHAES	18,8
ALAIDE SALES	18,8
JOEL GOMES DO ESPIRITO SANTO	18,0
JOSE FRANCISCO CHAGAS FILHO	17,2
JOSE CALIFORNIA	17,1
JOSE MANHAES DE SOUZA	17,0
JOSE RANGEL PESSANHA	16,9
FRANCISCO DE ASSIS CORREA	16,8
ANISIO PEDRO SOBRINHO	16,2
MARIA DA CONCEIÇÃO F. BARRETO	16,0
AILTON RIBEIRO VIANA	15,9
ANTONIO RIBEIRO	15,6
APOLINARIO JOAQUIM NETO	15,4
JOSE VITORIO DAS CHAGAS	15,3
SALVADOR DE SOUZA PINTO	14,8
CONSTANTINO MENDONÇA AZEVEDO	14,6
MARIA HELIA SALES DA SILVA	14,5
AMARO NUNES PEIXOTO	13,9
MANOEL FERNANDES SOBRINHO	13,7
ORLANDO GOMES DA SILVA	13,2
SALVADOR FRANCISCO CORDEIRO	13,1
VALDIR ALVES MANHAES	13,1
MANOEL ALBERTO	12,7
EDIVAL MANHAES CARDOSO	12,3
JOSE AMARO FERREIRA	12,3
PAULO DE SOUZA NETO	12,1

TABELA I - Continuação.

ROSEVALDO CONTILIO	12,0
JOSE ANTONIO DE CARVALHO	11,9
MARIA DA CONCEIÇÃO A. DA SILVA	11,2
SELMA RIBEIRO LEAL	10,5
RAUL TOLEDO DE ALMEIDA	9,8
MARIA DE LOURDES RANGEL	9,2
JOÃO FRANCISCO ALVES PEIXOTO	9,2
JAMIL VIEIRA	8,8
JOCELIA DOS SANTOS PINTO	8,6
PORFIRIO DOS SANTOS	8,3
MARIA DE LOURDES R. PESSANHA	8,3
MANOEL THOMAZ PEREIRA	8,2
MANOEL ELICIO DE SOUZA BARRETO	8,1
JOÃO MANOEL PEREIRA	7,9
ISRAEL HENRIQUE	7,8
ELCIMAR FREITAS RIBEIRO	7,8
JOSELINO NOGUEIRA DE SOUZA	7,7
NELSON BARRÊTO DE SOUZA	7,7
FRANCISCO RODRIG DA SILVA III	7,1
MARIA TERESA DA SILVA	6,7
MANOEL ARMINDO FELIPE DA SILVA	6,2
AURELIANO GOMES TERRA	6,0
AGASSIS DA SILVA FERREIRA	5,8
ZENITA TELES	5,5
IRINEA DE ALMEIDA MIRANDA	5,2
NEUSA AZEVEDO HENRIQUE	4,7
MARIA DO CARMO DE SOUZA	4,3
IANTONIA FERREIRA	4,3
ALNITO DOS SANTOS	4,2

TABELA I - Continuação.

MANOEL F. DE OLIVEIRA VIANA	4,2
ANTONIO AZEVEDO VIANA	3,7
POMPEU FRANCISCO DO ROSARIO	3,7
ANOEL ANTONIO MONTEIRO	3,2
ZELIA ALVES BARRETO	3,0
SALVADOR BRAGA DE AZEREDO II	2,8
JUPIRA PESSANHA NIGUEIRA	2,3
MANOEL NEVES	2,0
BENEDITO RANGEL DOS SANTOS	1,9
JOSE MARIA B. DO NASCIMENTO	1,5

TABELA II - População de 300 fornecimentos de cana-de-açúcar (em t.) na região de Piracicaba. Dados ordenados alfabeticamente pelos nomes dos fornecedores.

ABEL INACIO DA SILVEIRA	256,1
ADELINA RANGEL RIBEIRO	28,0
ADELINO GOMES BATISTA	65,1
ADERBAL PEREIRA MACHADO	476,3
AGASSIS DA SILVA FERREIRA	5,8
AGUSTINHO MIRANDA DA CONCEIÇÃO	631,1
AIDES VENTURA DO NASCIMENTO	121,3
AILTON RIBEIRO VIANA	15,9
ALAIDE SALES	18,8
ALDEMIR PEREIRA DA SILVA	186,0
ALFEU NOGUEIRA MACHADO	526,0
ALIATAR MARIANO DE AZEVEDO	3012,3
ALMIR ABDALA SALOMÃO	282,6
ALMIR SILVA	1107,8
ALMIR SOUZA BRAGA	78,4
ALNITO DOS SANTOS	4,2
ALVARO MONTEIRO CRUZ	5790,1
ALZEMIRO THOMAZ DE LIMA	70,8
AMARO ALCIDES PESSANHA	32,0
AMARO ALVES BARRETO II	51,2
AMARO BARBOSA DA SILVA	31,2
AMARO BARRETO MONTEIRO	244,1
AMARO BRAGA DE AZEREDO	20,2
AMARO DE ALMEIDA TRINDADE	27,4
AMARO DE SOUZA	65,6
AMARO NUNES PEIXOTO	13,9

TABELA II - Continuação.

AMARO PINHEIRO	28,3
AMARO REIS	878,5.
AMARO RIBEIRO II	34,7
AMARO RODRIGUES PESSANHA	67,4
AMARO SARDINHA FILHO	41,6
AMARO SIQUEIRA	38,2
AMARO ZACARIAS DE ABREU	201,5
AMORESTINA PESSANHA DA SILVA	1813,7
ANA FERREIRA MARTINS	79,3
ANISIO PEDRO SOBRINHO	16,2
ANOEL ANTONIO MONTEIRO	3,2
ANTONIO AZEVEDO VIANA	3,7.
ANTONIO CARLOS DE OLIVEIRA	44,2
ANTONIO DE SOUZA PINTO I	193,0
ANTONIO FIDELIS DA SILVA	577,4
ANTONIO FRANCISCO DE AZEVEDO 2	65,5
ANTONIO MARINS	27,0
ANTONIO MIRANDA DE AZEVEDO	976,8
ANTONIO PEREIRA VIANA	31,4
ANTONIO RATINHA FILHO	21,7
ANTONIO RIBEIRO	15,6
APOLINARIO JOAQUIM NETO	15,4.
ARIALDE COSTA	170,3
ARI DOS SANTOS SIQUEIRA	390,6
ARI GENARIO DA SILVA	99,9
ARTHUR RIBEIRO GOMES	64,7
AURELIANO GOMES TERRA	6,0
BENEDITO TAVARES NETO	37,1
BENEDITO DA CONCEIÇÃO	105,0

TABELA II - Continuação.

BENEDITO RANGEL DOS SANTOS	1,9
BENEDITO RIBEIRO RANGEL	385,4
CARLOS AUGUSTO SILVA	236,0
CARLOS A. MUYLAERT MENEZES	519,7
CEMILTON ROCHA	52,6
CERLY GOMES DE OLIVEIRA	795,1
CLAUDIO CESAR PEIXOTO DA SILVA	80,0
CLAUDIO LIZANDRO GODOY	24,4
CLEVES RIBEIRO DE SOUZA	210,5
COLY FERREIRA DE SIQUEIRA	25,3
CONCEIÇÃO DE MARIA DA SILVA	83,2
CONSTANTINO MENDONÇA AZEVEDO	14,6
DARCY GOMES LEAL	316,4
DARIO RIBEIRO BATISTA	326,6
DENILDA MARCELINA DA PAIXÃO	79,3
DIONICIO RIBEIRO DE CARVELHO	22,0
DIRCON DE SIQUEIRA REIS	88,9
DOMINGOS MELO RANGEL	34,8
DOMINGOS NUNES BARRETO	358,9
DORALICIO DE SOUZA	456,9
EDILSA PESSANHA DE ARAUJO	27,2
EDMAR AZEVEDO	97,0
EDUARDO DA SILVA PESSANHA	50,5
EDVAL MANHAES CARDOSO	12,3
EFIGENIA LEMOS DA SILVA	108,9
ELCIMAR FREITAS RIBEIRO	7,8
ELIAS TAVARES DE AZEVEDO FILHO	103,0
ELIZARDO FERREIRA	234,9

TABELA II - Continuação.

ERALDO DA SILVA SANTOS	109,3
ERCILIO GOMES PAES	131,6
ERNANDE GONÇALVES DE DEUS	72,5
EROTILDE AZEVEDO	19,3
EUCLIDES HEITOR BARRETO	99,2
EUCLIDES PESSANHA DE SOUZA	121,4
EVERALDO SALES	82,0
EZIO ERNESTO DOS SANTOS	281,8
FLORIANO MANHAES	106,6
FRANCISCA FREITAS SARDINHA	246,0
FRANCISCO ARCANJO FERREIRA	318,9
FRANCISCO BARROSO DE SIQUEIRA	25,4
FRANCISCO DA CONC. A.BARBOSA	565,6
FRANCISCO DA SILVA VASCONCELOS	1081,1
FRANCISCO DE ASSIS CORREA	16,8
FRANCISCO DE SOUZA PINTO FILHO	157,1
FRANCISCO DE SOUZA	22,5
FRANCISCO FERREIRA MACHADO	244,2
FRANCISCO INACIO DE OLIVEIRA	52,9
FRANCISCO LEONARDO	38,6
FRANCISCO LIMA	29,5
FRANCISCO PAES VIANA	173,0
FRANCISCO PEREIRA BARRETO	376,6
FRANCISCO PESSANHA DE SOUZA I	61,5
FRANCISCO RODRIG DA SILVA III	7,1
GECI VIANA PEREIRA	49,3
GELSON DA CONCEIÇÃO RIBEIRO	1364,1
GERALDO DA CONCEIÇÃO SERAFIM	451,0
GERALDO JOSE DE ALMEIDA	43,9

TABELA II - Continuação.

GERALDO PAES	735,6
GERMANO DIOGO M. DOS SANTOS	98,0
HELMA DE SOUZA GOMES	505,8
HENRIQUE DE SOUZA NOGUEIRA	1762,9
HERMES NOGUEIRA PINTO	141,0
HERVAL TAVARES	177,4
IAANTONIA FERREIRA	4,3
INACIO RIBEIRO MOCO	126,4
IRINEA DE ALMEIDA MIRANDA	5,2
IRINEU DE AZEVEDO LIMA	2453,9
IRINEU GONÇALVES NOGUEIRA	116,6
ISRAEL HENRIQUE	7,8
IZIDORO C RAPOSO DE ALMEIDA	2299,7
JAIME HENRIQUE NOGUEIRA	79,6
JAMIL VIEIRA	8,8
JOÃO BATISTA DE ABREU RANGEL	628,1
JOÃO BATISTA MACHADO	87,8
JOÃO CHAGAS DE CARVALHO	69,1
JOÃO DE LIMA RANGEL	540,7
JOÃO DOS SANTOS	139,5
JOÃO E GONÇALVES DOS SANTOS	100,9
JOÃO FIRMO DE ARAUJO	75,5
JOÃO FRANCISCO ALVES PEIXOTO	9,2
JOÃO FRANCISCO GOMES	171,9
JOÃO GOMES PEDRA	135,5
JOÃO LOURENÇO DE SOUZA	948,3
JOÃO MANOEL PEREIRA	7,9
JOÃO MATIAS FERREIRA	420,1
JOÃO RIBEIRO DA FONSECA	143,8

TABELA II - Continuação.

JOÃO ROSA DA SILVA FILHO	678,5
JOÃO SOARES DE SOUZA	303,1
JOAQUIM PINTO BARRETO	193,1
JOAREZ MAIA GONÇALVES	28,3
JOCELIA DOS SANTOS PINTO	8,6
JOEL GOMES DO ESPIRITO SANTO	18,0
JOILSON PUPE DE ALMEIDA	48,1
JORGE FERREIRA DE SOUZA	328,9
JORGE FRANCISCO DE SOUZA I	29,9
JORGE RANGEL	64,9
JOSELINO NOGUEIRA DE SOUZA	7,7
JOSEMARCIA FATIMA C. GONÇALVES	92,8
JOSÉ ALENCAR DA SILVA	59,8
JOSÉ ALVES DE SOUZA	87,8
JOSÉ AMARO FERREIRA	12,3
JOSÉ ANTONIO DE CARVALHO	11,9
JOSÉ ANTONIO PAES DE ALMEIDA	41,0
JOSÉ ANTONIO TOLEDO	39,2
JOSÉ CAETANO ROSA	34,0
JOSÉ CALIFORNIA	17,1
JOSÉ CARLOS DA SILVA IIIII	135,5
JOSÉ CARLOS DE AZEVEDO	330,6
JOSÉ DE OLIVEIRA MEIRELES	1249,0
JOSÉ DOMINGOS DA SILVA	121,3
JOSÉ DOS SANTOS	533,9
JOSÉ EUZEBIO	19,1
JOSÉ FRANCISCO CHAGAS FILHO	17,2
JOSÉ GERALDO ALVES QUEIROZ	250,1
JOSÉ LICIO RIBEIRO MANHAES	23,2
JOSÉ MANHAES DE SOUZA	17,0

TABELA II - Continuação.

JOSÉ MANOEL ALVES FIUZA	23,7
JOSÉ MARIA B DO NASCIMENTO	1,5
JOSÉ PAES DE LIMA	2531,7
JOSÉ RANGEL PESSANHA	16,9
JOSÉ RIBEIRO AREAS	675,8
JOSÉ ROMARIO PEREIRA GOMES	19,7
JOSÉ THOMAZ CASTILHO	150,8
JOSÉ VIANA I	250,5
JOSÉ VITORIO DAS CHAGAS	15,3
JOVIANO FERREIRA NETO	67,1
JULIAO VIANA DE CARVALHO	379,2
JULIO GOMES SOARES	135,8
JUPIRA PESSANHA NIGUEIRA	2,3
JUVENAL DE SOUZA PINTO	19,1
LAERCIO PAES DA SILVA	172,3
LAUDELINA FERREIRA MARTINS	30,3
LAUDELINA RIBEIRO	39,9
LAUDELINO BENEVIDES GOMES	1122,9
LAVINA RIBEIRO PINTO	87,8
LENICIO ROSA	223,0
LICIO GONÇALVES DE SOUZA	22,6
LOPO MEIRELES JUNIOR	476,8
LUIZ ANDRE GUALDA TERRA	383,8
LUIZ CARLOS AREAS	216,4
LUIZ CERQUEIRA GUIMARAES	121,6
LUIZ CHAGAS	50,5
LUIZ FERNANDO BARROS	181,0
LUIZ FERNANDO MARÇAL PEREIRA	19,2
LUIZ MARQUES SIMÕES	706,1

TABELA II - Continuação.

LUIZ OCTAVIO DA COSTA SILVA	182,9
LUIZ RIBEIRO GOMES	108,4
LUIZ SOARES NOGUEIRA	77,2
MANOEL AFONSO	58,4
MANOEL ALBERTO	12,7
MANOEL ARLINDO FELIPE DA SILVA	6,2
MANOEL BARRETO SALES	172,4
MANOEL DA SILVA SALES	27,2
MANOEL DIONIZIO BATISTA	92,7
MANOEL DOS SANTOS TRINDADE	110,9
MANOEL ELICIO DE SOUZA BARRETO	8,1
MANOEL FERNANDES SOBRINHO	13,7
MANOEL F DE OLIVEIRA VIANA	4,2
MANOEL NEVES	2,0
MANOEL PEREIRA VIANA FILHO	64,4
MANOEL PESSANHA NOVUEIRA	56,8
MANOEL PINTO NOGUIERA	99,7
MANOEL THOMAS PEREIRA	8,2
MANOEL TOLEDO DE ALMEIDA	185,4
MARIA ALVES MANHAES	33,6
MARIA ANTONIA BRANCO PEREIRA	41,8
MARIA DA CONCEIÇÃO A DA SILVA	11,2
MARIA DA CONCEIÇÃO F BARRETO	16,0
MARIA DA C SILVA RANGEL	42,6
MARIA DA PENHA DA SILVA	24,0
MARIA DE LOURDES MONTEIRO	22,4
MARIA DE LOURDES RANGEL	9,2
MARIA DE LOURDES R PESSANHA	8,3
MARIA DOLORES G. DE SIQUEIRA	40,3

TABELA II - Continuação.

MARIA DO CARMO DE SOUZA	4,3
MARIA ELMA MANHAES DA SILVA	126,2
MARIA ELZA SALES ANDRADE	66,4
MARIA HELIA SALES DA SILVA	14,5
MARIA IZABEL FERREIRA DA SILVA	24,3
MARIA JOSE BARCELOS NETO	50,6
MARIA MERCEDES SOARES	32,7
MARIA PEREIRA GOMES DA SILVA	74,0
MARIA RANGEL DE LIMA	46,9 .
MARIA TEREZA DA SILVA	6,7
MARIETA BATISTA MAGALDI	146,5
MARIO RIBEIRO DE SOUZA	46,6
MARLI DE SOUZA RIBEIRO	46,4
MAURINO ALMEIDA	57,9
NAIR FELIX DE BARROS	774,9
NEIDE DUARTE ALVES	306,0
NELSON BARRETO DE SOUZA	7,7
NELSON MOREIRA DE LIMA	904,7
NELY MENEZES MUYLAERT	1456,3 .
NEUZA AZEVEDO HENRIQUE	4,7
NILSON PEREIRA RANGEL	72,0
NILZETE DA ROCHA BATISTA	20,8
ONIZEDI SANTOS FORMOSO	3943,3
ORESTE SALES CAVALINE	145,8
ORLANDO GOMES DA SILVA	13,2
OTAVIO RIBEIRO	58,5
OZIEL PESSANHA DE OLIVEIRA	87,6
PAULO DE SOUZA NETO	12,1
PAULO SALES	38,4 .

TABELA II - Continuação.

PAULO SILVEIRA COUTINHO	4819,8
PEDRO JOAQUIM DOS SANTOS	71,8
POMPEU FRANCISCO DO ROSARIO	3,7
PORFIRIO DOS SANTOS	8,3
RAMIRO RANGEL PAES	336,8
RAUL TOLEDO DE ALMEIDA	9,8
REINALDO DE SOUZA ROSA	103,6
REINALDO FERNANDES NASCIMENTO	100,9
RENATO CEZAR PESSANHA MANHAES	18,8
RENATO GOMES RIBEIRO DA SILVA	79,8
RENIVALDO JOSE VIANA	58,0
RITA DE CASSIA R DO ROSARIO	69,7
ROBERTO DOS SANTOS P. GUIMARÃES	215,9
ROSEVALDO CONTILIO	12,0
ROTCHILDES DE SOUZA RIBEIRO	37,9
RUBENS AREAS VENANCIO	5249,6
RUTH TRINDADE DE SOUZA	58,9
SALVADOR BRAGA DE ZEREDO II	2,8
SALVADOR DE SOUZA PINTO	14,8
SALVADOR FRANCISCO CORDEIRO	13,1
SANTILHA MARIA SANTOS EDUARDO	47,7
SEBASTIÃO DE O MACIEL E SILVA	21,3
SEBASTIAO PANISSET GOUDARD	366,4
SELMA RIBEIRO LEAL	10,5
SILVIO CUSTODIO DA SILVA	147,1
SILVIO DA CUNHA PINTO	831,2
SILVIO R TEIXEIRA TAVARES	26,0
SILVIO SIQUEIRA	28,1
UILSON NAJIENSE DA SILVA	35,3

TABELA II - Continuação.

UIZ DE SOUZA LEAL	30,8
UMBERTO BELATRABLE PES. SANTOS	92,3
UNIGENITO INACIO DA FONSECA	55,4
VALDEMAR ROCHA	29,6
VALDEMIRO RODRIGUES	93,0
VALDIR ALVES MANHAES	13,1
VANILDO DA SILVA COSTA	1593,2
WALDOMIRO BENTO E OLIVEIRA	357,2
WALTER RIBEIRO DA FONSECA	523,9
WOLFGANG ECHTERBECK	1879,4
ZELIA ALVES BARRETO	3,0
ZELIO NOGUEIRA MARTINS	106,3
ZENITA TELES	5,5