Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística

Métodos de Amostragem com Aplicações na Área Empresarial Com Um Enfoque Integrado ao *Software* "Sampling"

S. A. Mingoti, G. S. Atuncar, M. L. G. Nogueira e R. C. Silva

Relatório Técnico
RTE-01/2000

Relatório Técnico
Série Ensino

MÉTODOS DE AMOSTRAGEM
COM APLICAÇÕES NA ÁREA EMPRESARIAL

COM UM ENFOQUE INTEGRADO AO SOFTWARE "SAMPLING"

AUTORES: SUELI APARECIDA MINGOTI (*)
GREGÓRIO SARAVIA ATUNCAR (*)
MARIA LOURDES GRANHA NOGUEIRA(*)
RODRIGO CARAZOLLI DA SILVA (**)

(*) PROFS. ADJUNTO DO DEPARTAMENTO DE ESTATÍSTICA DA UFMG
(**) BACHAREL EM ESTATÍSTICA PELA UFMG

TEXTO ELABORADO COM O APOIO FINANCEIRO DA FAPEMIG
2000

MÉTODOS DE AMOSTRAGEM COM APLICAÇÕES NA ÁREA EMPRESARIAL

COM UM ENFOQUE INTEGRADO AO SOFTWARE "SAMPLING"

AGRADECIMENTO

OS AUTORES AGRADECEM À FAPEMIG PELO APOIO FINANCEIRO QUE POSSIBILITOU A REALIZAÇÃO DESTE TRABALHO.

s

MÉTODOS DE AMOSTRAGEM
COM APLICAÇÕES NA ÁREA EMPRESARIAL

COM UM ENFOQUE INTEGRADO AO SOFTWARE "SAMPLING"

ÍNDICE

CAPÍTULO 1: INTRODUÇÃO AOS PROCEDIMENTOS DE AMOSTRAGEM

1.0 Introdução

Os procedimentos de amostragem vêm sendo utilizados ao longo dos anos em várias áreas diferentes do conhecimento humano. Algumas referências clássicas são Deming (1950), Cochran (1977) e Hansen, Hurwitz e Madow (1953). Se na indústria a seleção de amostras sempre teve sua importância devido principalmente a implementação do controle estatístico de qualidade e a inspeção da qualidade por amostragem (Duncan,1985), na área das ciências humanas a necessidade de obter-se inferências sobre o comportamento social e demográfico de populações humanas determinou a utilização de procedimentos amostrais em grande escala. Nos dias atuais a grande competitividade entre empresas tem incentivado as pesquisas de opinião junto ao mercado consumidor ou para avaliar produtos já existentes ou para obter informações que possam auxiliar nos projetos de desenvolvimento de novos produtos que venham a satisfazer as necessidades deste mercado.

Apesar da grande importância dos procedimentos amostrais, talvez a grande maioria da população tenha ciência da existência das técnicas de amostragem pelo fato destas serem utilizadas nas pesquisas eleitorais realizadas pelos vários institutos de pesquisa na época das eleições.

Embora exista uma grande literatura na área de amostragem, poucos são os textos que apresentam casos reais, principalmente na área industrial. Em muitos livros de Estatística, os conjuntos de dados dos vários exemplos tratados aparecem sem muita referência sobre a forma como foram coletados. Além disso, a grande maioria dos textos está na língua Inglêsa o que introduz uma maior dificuldade no aprendizado e na difusão dos métodos. Daí a importância e necessidade de textos na língua Portuguêsa.

Nosso objetivo com este novo texto sobre o assunto é justamente procurar mostrar ao leitor como os procedimentos de amostragem mais comuns podem ser aplicados em situações práticas, mais propriamente na área empresarial. As técnicas abordadas são aquelas de maior aplicação real, como Amostragem Aleatória Simples, Amostragem Estratificada, Amostragem Sistemática, Amostragem P.P.S., Amostragem por Conglomerados em 1,2,ou 3 estágios e a Subamostragem Estratificação. Em nenhum momento serão apresentadas demonstrações dos resultados colocados no texto pois acreditamos que a literatura de amostragem é bastante vasta sob este aspecto (ver por exemplo Sukthame e Sukhatme, 1970). Nossa maior preocupação é com o aspecto prático das técnicas. Deste modo, ao longo do texto vários exemplos são apresentados e resolvidos em sua grande maioria, com o auxílio do software estatístico ***"Sampling"*** (1998), desenvolvido pelos autores deste texto com o suporte financeiro da FAPEMIG, Instituição de fomento à Pesquisa do Estado de Minas Gerais.

Este texto não tem como sua proposta principal apresentar a utilidade do software ***"Sampling"*** na sua totalidade. A título de ilustração, alguns exemplos de operacionalização prática do software são apresentados no Anexo A.

Finalmente a leitura deste texto pressupõe que o leitor tenha alguns conhecimentos básicos de Inferência Estatística. Deste modo, termos como estimadores não-viciados, intervalos de confiança, etc. são utilizados ao longo do texto sem maiores explicações técnicas .

2.0 O Software "Sampling"

O software ***"Sampling"*** funciona como um complemento do popular software estatístico ***"Minitab For Windows"*** (1998) e permite que o usuário execute facilmente a análise estatística de dados através de métodos estatísticos que levem em consideração o procedimento amostral utilizado na coleta destes dados. O programa ***"Sampling"*** funciona de forma iterativa com o usuário, em um sistema de várias macros interligadas a uma macro *mestre* onde as informações são armazenadas. As perguntas são feitas passo a passo em uma seqüência lógica de análise. Por funcionar a partir do ***"Minitab",*** o usuário, além dos recursos disponíveis no ***"Sampling"***, tem a grande vantagem de desfrutar de todos os outros recursos estatísticos importantes à uma análise estatística, sem precisar sair do software. Uma das vantagens de se operar no sistema Windows 95 (ou 98) é que a capacidade de armazenamento de dados fica limitada apenas pela quantidade de memória disponível *("Minitab" versão 11.0 ou 12.0*) o que torna o ***"Sampling"*** um software capaz de operar com conjuntos de dados bastante extensos. O programa também pode ser utilizado com versões anteriores do ***"Minitab"*** a partir da versão 10.0. Em termos do espaço ocupado no disco rígido, o ***"Sampling"*** é bem econômico. Tecnicamente, para se obter uma boa performance, o programa requer no mínimo um computador com processador Pentium™ 100 e 16 MB de memória RAM.

Os procedimentos amostrais disponíveis no software ***"Sampling"*** são: Amostragem Aleatória Simples com ou sem reposição, Amostragem Sistemática, Amostragem Estratificada, Amostragem do Tipo P.P.S, Amostragem por Conglomerados, Amostragem em Dois ou Três Estágios, Subamostragem Estratificada e Pós- Estratificação.

O software ***"Sampling"*** foi desenhado tendo a Amostragem Estratificada como seu procedimento amostral base. Deste modo, as análises estatísticas para os casos em que se tem apenas uma amostra proveniente de um procedimento amostral como, Amostragem Aleatória Simples, Amostragem Sistemática e Amostragem por Conglomerados são obtidas como casos particulares da Amostragem Estratíficada quando se tem apenas um estrato na população, o que significa dizer que o estrato é a própria população. A Amostragem do tipo P.P.S. é tratada como um caso particular da Sub-Amostragem

De um modo geral o software ***"Sampling"*** permite ao usuário a estimação dos parâmetros populacionais (média, total e proporções) através dos métodos estatísticos não-viciados usuais de estimação pontual e intervalar (Cochran, 1977) ou pelos métodos de Razão e Regressão, sendo que nestes casos a variância dos estimadores poderá ser obtida pelo método separado ou método combinado. A estimação por intervalo é executada para 4 diferentes valores do nível de significância α, isto é, 1, 2,5, 5 e 10 %.

Nos módulos de Amostragem Aleatória Simples, Amostragem Sistemática e Amostragem Estratificada, o usuário também poderá proceder a uma análise descritiva de seus dados (opcional) ou calcular o tamanho *n* da amostra necessário para obter-se uma precisão especificada "à priori" (tolerância e probabilidade de erro). Vários recursos disponíveis neste software não serão tratados neste texto. Para um maior conhecimento de seu funcionamento sugerimos que o leitor veja o manual do usuário (Mingoti,Atuncar, Silva e Granha, 1998).

Outras informações sobre o software ***"Sampling"*** podem ser obtidas com os autores no endereço eletrônico: Sueli@est.ufmg.br .

CAPÍTULO 2: AMOSTRAGEM ALEATÓRIA SIMPLES

1.0 Introdução

O esquema de Amostragem Aleatória Simples é o mais fácil dos procedimentos amostrais existentes. Entretanto, em geral a sua utilização é feita acoplada a outros procedimentos amostrais mais sofisticados. Basicamente, neste método uma amostra de n elementos de uma população constituída de N unidades amostrais, n<N, é selecionada através de um mecanismo aleatório que atribui a cada uma das unidades amostrais populacionais a mesma probabilidade de pertencer a amostra final, ou de outro modo a mesma probabilidade de seleção em cada estágio da coleta. A amostra pode ser selecionada com ou sem reposição. Estimadores não-viciados (Cochran,1977) são então utilizados para a estimação pontual ou por intervalo, dos parâmetros populacionais de interesse, tais como, a média ($\bar{Y}$) ou total (Y) populacionais. Outros métodos de estimação como os de Razão e Regressão também são possíveis de serem utilizados juntamente com a Amostragem Aleatória Simples (Scheaffer,Mendenhall e Ott,1996). Outra questão de interesse relaciona-se com a escolha do tamanho da amostra (n) necessário para obter-se uma precisão e erro de estimação (tolerância) fixados “ à priori”.

2.0 Notação

Para facilitar a exposição, no que segue será utilizada a seguinte notação:

N : tamanho da população
n : tamanho da amostra
y_i : valor da variável resposta para a i-ésima unidade amostral
$\bar{y}$: média da amostra
s^2 : variância da amostra
$\bar{Y}$: média populacional
$\hat{\bar{Y}}$: estimador da média populacional
Y : total populaciomal
$\hat{Y}$: estimador do total populacional
σ^2 : variância populacional

3.0 Amostragem Aleatória Simples Sem Reposição

Conceitualmente, o esquema de Amostragem Aleatória Simples sem reposição pressupõe que a população de interesse esteja dividida em N unidades amostrais (elementos amostrais), e que se disponha de uma lista de todas estas unidades amostrais, lista esta chamada de *Frame*. Portanto, parte-se do princípio de que as N unidades possuem algum tipo de identificação. A forma mais simples de se obter uma amostra aleatória de n unidades desta população seria escolher-se n números aleatórios sem reposição, ou seja n n’’umeros aleatórios distintos. Este números podem ser selecionados ou através de uma tabela de números aleatórios, ou através de um gerador computacional de números aleatórios. As unidades que possuissem a identificação igual aos números aleatórios selecionados seriam então escolhidas para comporem a amostra final. Quando o esquema de amostragem aleatória simples sem reposição é empregado, os estimadores não-viciados da média e total populacionais são dados respectivamente por:

$$\hat{\bar{Y}} = \bar{y} \quad \text{e} \quad \hat{Y} = N\bar{y}$$

sendo as variâncias estimadas de $\bar{y}$ e de $\hat{Y}$ dadas respectivamente por:

$$\hat{V}ar(\bar{y}) = \frac{s^2}{n}\left(\frac{N-n}{N}\right) \quad \text{e} \quad \hat{V}ar(\hat{Y}) = N^2\frac{s^2}{n}\left(\frac{N-n}{N}\right)$$

Para um nível de significância α, $0<\alpha<1$, um intervalo de $(1-\alpha)100\%$ de confiança para a média populacional é dado respectivamente por:

$$\bar{y} \pm t_{n-1,\frac{\alpha}{2}} \sqrt{\hat{V}ar(\bar{y})} \qquad \text{se} \quad n<30$$

$$\bar{y} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{V}ar(\bar{y})} \qquad \text{se} \quad n\geq 30$$

onde t representa a ordenada da distribuição t de Student com n-1 graus de liberdade e z representa a ordenada da distribuição normal.

Os intervalos de confiança para o total populacional são construídos de maneira similar, bastando multiplicar as equações acima pela constante N.

3.1 Exemplos de Aplicação

Exemplo 1. Uma determinada indústria do ramo de eletrônica, avalia continuamente ao longo da linha de produção, a qualidade dos componentes eletrônicos que fabrica. Os componentes fabricados por esta empresa são vendidos para outras companhias que utilizam os componentes na montagem de equipamentos. Apesar da existência deste sistema de monitoramento de qualidade do processo, antes que o produto seja enviado ao cliente a empresa seleciona aleatoriamente por amostragem aleatoria simples sem reposição, um certo número n de componentes dentre aqueles que serão enviados ao cliente. Suponha a remessa a ser enviada para um determinado cliente seja de N=1000 componentes, e suponha que uma amostra de n=77 componentes tenha sido selecionada desta remessa. Para cada componente selecionado é medido o valor de uma determinada característica elétrica. Os dados obtidos estão apresentados na tabela 3.1 a seguir:

Tabela 3.1: Valores da característica elétrica medida nos componentes

16,5	15,7	17,3	16,9	15,5	13,5	16,5	16,5	14,5	16,9	16,5
17,2	17,6	15,8	15,8	16,6	13,5	14,3	16,9	14,9	16,5	16,7
16,6	16,3	16,8	16,9	15,9	16,0	16,9	16,8	15,6	17,1	16,3
15,0	14,6	17,2	16,8	16,5	15,9	14,6	16,1	16,8	15,8	14,0
14,4	14,9	16,2	16,6	16,1	13,7	17,5	16,9	12,9	15,7	14,9
16,5	15,2	16,9	16,0	16,2	15,2	15,5	15,0	16,6	13,0	15,6
15,5	16,1	14,9	16,6	15,7	15,9	16,1	16,1	10,9	15,0	16,8

Com o auxílio do software ***"Sampling"*** obteríamos as seguintes estimativas pontual e por intervalo para o valor médio da característica de qualidade elétrica em consideração:

Estimação Global

. Tipo de Amostragem: Aleatória Simples Sem Reposição
. Estimadores: Usuais

N	Mean	StDev	I.C.
1000	15.8143	0.1315947	(15.5978, 16.0307)

. Z Value : 1.64485 , $\alpha = 0{,}05$

Portanto, podería-se afirmar que o valor médio da característica elétrica dos componentes que estão na remessa (lote) seria igual a 15,8143. Um intervalo de 95% de confiança para o verdadeiro valor da média desta característica de qualidade seria (15,5978 ; 16,0307). Se a especificação do cliente para esta característica fosse um valor médio entre 14 e 18, por exemplo, o lote poderia ser enviado ao cliente sem maiores problemas. Entretanto, se os valores de especificação para a média fossem 15 e 15,8 o lote não poderia ser enviado ao cliente da forma como está. Para tal, seria necessário passar por uma inspeção 100%.

A título de informação sobre o funcionamento e praticidade do software ***"Sampling"*** incluímos no anexo A os detalhes do processamento dos dados deste exemplo.

Exemplo 2. Uma empresa que utiliza a madeira de um determinado tipo como uma das matérias primas para seu processo de produção, está interessada em adquirir uma certa fazenda que contém 150 hectares. Para avaliar a possibilidade de compra a empresa decidiu retirar uma amostra aleatória, por amostragem aleatória simples sem reposição, de 20 pontos de área igual a 1 hectare, da fazenda em consideração e para cada. O mapa da região foi utilizado para esta finalidade. Em cada área selecionada mediu-se o número de árvores que tinham um tamanho maior que um determinado valor especificado. Os valores obtidos estão na tabela a seguir:

Tabela 3.2: Exemplo das árvores

Pontos selecionados	*Número de Árvores*
87	17
58	10
1	19
43	22
55	15
16	23
75	17
97	20
23	15
98	25
17	13
15	15
20	16
37	12
31	14
91	18
20	23
6	21
61	13
66	19

Utilizando o software ***"Sampling"*** obteríamos as seguintes estimativas para os parâmetros populacionais neste caso:

Estimação Global

. Tipo de Amostragem : Aleatória Simples Sem Reposição
. Estimadores : Usuais

N	Mean	StDev	I.C.
150	17.3500	0.8550315	(15.8715, 18.8285)

Total	StDev	I.C.
2602.5000	128.2547321	(2380.73, 2824.27)

. T Value : 1.72914 , $\alpha = 0,05$

Portanto, uma estimativa do número médio de árvores na área considerada com tamanho maior que o especificado seria igual a 17,35, sendo que a empresa esperaria com 95% de confiança que o valor verdadeiro deste número médio estivesse no intervalo (15,8715 ; 18,8285). Para o número total de árvores da área com tamanho maior que o especificado esperaria-se um valor igual a 2602,50 sendo um intevalo de 95% de confiança para o valor verdadeiro deste total igual a (2380,73 ; 2824,27).

4.0 Amostragem Aleatória Simples Com Reposição

Na Amostragem Aleatória Simples com reposição é permitido que a mesma unidade amostral apareça na amostra mais de uma vez o que por sua vez, acarreta perda de informação. Em cada estágio de seleção todas as N unidades amostrais têm a mesma probabilidade de serem selecionadas para compor a amostra final. É um procedimento que pode ser utilizado sem maiores problemas quando o tamanho populacional N é grande, pois assim a probabilidade de se obter a mesma unidade amostral várias vezes na amostra é pequena. A amostra pode ser selecionada usando um mecanismo aleatório similar àquele descrito para o caso da amostragem aleatória simples sem reposição. A diferença é que um mesmo número aleatório poderá aparecer mais de uma vez indicando que a unidade amostral relacionada estaria na amostra mais de uma vez.

Para o procedimento de Amostragem Aleatória Simples com reposição, os estimadores não-viciados da média e total populacionais são dados respectivamente por:

$$\hat{\bar{Y}} = \bar{y} \quad \text{e} \quad \hat{Y} = N\bar{y}$$

As variâncias estimadas de $\bar{y}$ e de $\hat{Y}$ são respectivamente:

$$\hat{V}ar(\bar{y}) = \frac{s^2}{n} \qquad \text{e} \qquad \hat{V}ar(\hat{Y}) = N^2 \frac{s^2}{n}$$

Os intervalos de confiança para os parâmetros populacionais são construídos de modo similar ao descrito para Amostragem Aleatória Simples sem reposição.

Exemplo de Aplicação (Promoção de Supermercados): Como um exemplo de aplicação poderíamos considerar um supermercado que tem 30 caixas de atendimento ao cliente. O supermercado está fazendo uma promoção com o intuíto de atrair consumidores. De hora em hora, um número aleatório entre 1 e 30 é escolhido através de um mecanismo aleatório. Como prêmio o cliente que estiver com suas compras no caixa sorteado leva todas as suas compras para casa sem ter que pagá-las. Em todos os momentos em que a seleção de números aleatórios é feita, todos os 30 números têm a mesma chance de ocorrência, inclusive aqueles que já foram sorteados anteriormente. Portanto, tem-se aqui um caso típico de amostragem aleatória simples com reposição.Suponha por exemplo que os supervisores do supermercado desejem estimar o gasto médio por dia com esta promoção. No segundo dia da promoção eles observaram os gastos com todos os 14 clientes que foram sorteados naquele dia (o supermercado funciona de 8 as 22 hs. sendo o primeiro sorteio as 8:30 hs. e o último as 21:30 hs.). Suponha que tenha sido observado um gasto médio de 85,40 reais com um desvio padrão de 15,34 reais. Neste caso, uma estimativa para a variabilidade do gasto médio diário que o supermercado tem com esta promoção seria dado por:

$$\hat{V}ar(\,\bar{y}\,) = \frac{(\,15{,}34\,)^2}{14} = 16{,}80$$

e um intervalo de confiança de 95 % para o gasto médio diário seria:

$$85{,}40 \pm (\,2{,}16\,)\sqrt{16{,}808} = 85{,}40 \pm 8{,}855 \Rightarrow (\,76{,}545;\, 94{,}255\,)$$

onde o valor 2,16 vem da distribuição t-Student com 13 graus de liberdade para um nível $\alpha/2$ = 0,025.

5.0 Determinação do Tamanho da Amostra na Amostragem Aleatória Simples

Suponha que se deseje estimar a média populacional $\bar{Y}$ com um erro de estimação (tolerância) de no máximo *d,* e de modo que a probabilidade de que o erro de estimação seja maior que *d* seja no máximo igual a α, $0 < \alpha < 1,$ isto é:

$$P(|\,\bar{Y} - \hat{\bar{Y}}\,| > d\,] \leq \alpha$$

onde d = $\left|\,\bar{Y} - \bar{y}\,\right|$.

Neste caso, o tamanho n da amostra necessário para que os requisitos especificados de estimação sejam satisfeitos é obtido através das seguintes fórmulas:

$$n = \frac{N\sigma^2}{(\,N-1\,)\left(\dfrac{d}{z_{\alpha/2}}\right)^2 + \sigma^2}$$ (se a amostragem for sem reposição)

e

$$n = \left(\frac{z_{\alpha/2}\,\sigma}{d}\right)^2$$ (se a amostragem for com reposição)

onde $z_{\alpha/2}$ refere-se ao valor absoluto da ordenada da distribuição normal padronizada que deixa abaixo dele a área igual $\alpha/2$.

Para utilizar as fórmulas descritas é necessário que se tenha alguma estimativa prévia do desvio padrão populacional. Esta estimativa pode proceder de algum estudo similar feito anteriormente (estudo piloto) ou de considerações práticas sobre a estrutura da população em avaliação. Observe ainda que a fórmula de n para o caso da amostragem aleatória simples com reposição também se aplica àqueles casos em que a amostragem é feita sem reposição, mas o tamanho populacional N pode ser considerado bastante grande. Neste caso, a fração de correção para populações finitas (n/N) aproxima-se de zero e consequentemente tem-se a fórmula similar a amostragem aleatória simples com reposição.

Exemplo de Aplicação: Suponha que uma indústria de mineração esteja interessada em encontrar o número n de incrementos que devem ser extraídos da sua linha de produção diária, com a finalidade de estimar o teor médio de ferro (%) de uma pilha de minério que está sendo formada. Os técnicos da empresa admitem uma tolerância de no máximo 0,01% com 99% de confiança, ou seja considerando-se uma probabilidade de erro de no máximo α = 0,01. Suponha que no momento o processo de produção da empresa esteja controlado para produzir minério com um desvio padrão igual a 0,0413. Supondo que a amostragem aleatória simples sem reposição seja utilizada para a seleção dos incrementos de minério da linha de produção, e que a pilha de minério tem um número N de incrementos bastante grande, então o tamanho da amostra n ideal para satisfazer as exigências dos técnicos da empresa seria:

$$n = \left(\frac{z_{\alpha/2}\sigma}{d} \right)^2 = \left(\frac{3{,}09 \ \ (0{,}0413)}{0{,}01} \right)^2 = 152{,}7696 \approx 153 \text{ incrementos}$$

onde 3,09 = $z_{0,005}$.

Algumas considerações práticas devem ser feitas no que se refere ao uso do esquema de amostragem aleatória simples no caso da indústria de mineração. Primeiramente, teríamos que considerar que a quantidade total de minério (pilha em toneladas) poderia ser decomposta em N partes, cada parte de aproximadamente o mesmo volume de minério (kg). Estas N partes ou os incrementos, seriam as unidades amostrais do procedimento. Como neste caso não dispomos de um *Frame*, pois não podemos listar um a um os incrementos que passam pela correia transportadora, imagina-se que esta enumeração estaria sendo feita naturalmente ao longo do próprio processo de produção, ou seja, se tivessemos um fluxo de minério aproximadamente constante na correia de empilhamento, então as unidades amostrais seriam “identificadas” através do tempo de funcionamento da correia. Assim, se por exemplo cada incremento tem 2,0 kg e o fluxo de minério é de 2000 kg/min então saberíamos que no 10o. segundo de funcionamento da correia já teriam sido transportados aproximadamente 167 incrementos ou unidades amostrais. Neste caso, a unidade amostral de número 167 seria selecionada se fosse retirado da correia transportadora um incremento de minério no 10o. segundo de funcionamento da mesma. Como tem-se a informação do peso total em toneladas da pilha de minério que é formada continuamente tem-se uma noção de qual é o valor do número N de incrementos da pilha, ou seja da população em avaliação. Portanto, a coleta de minério neste caso é feita com base no fluxo e no tempo de funcionamento da correia.

6.0 Estimação de uma Proporção Populacional

A proporção populacional *P* de elementos da população que pertencem a uma certa categoria tal como: “a peça ser defeituosa”, “votar no candidato A”, ou “ter uma determinada doença”, etc. , é estimada por $\hat{p}$ a fração de elementos na amostra que possui a característica de interesse. Se associarmos o valor numérico “1” para aquelas unidades amostrais populacionais que pertencem à categoria de interesse, e o valor numérico “0” para aquelas unidades amostrais populacionais que não pertencem à categoria, a fração amostral $\hat{p}$ pode ser vista como uma média de valores zeros e uns da amostra. Assim, $\hat{p} = \bar{y}$ e o tratamento é análogo ao da estimação da média populacional descrito nas seções 3.0 e 4.0. Neste caso, o valor de s^2 simplifica para:

$$s^2 = \frac{n}{n-1}\,\hat{p}\,(\,1-\hat{p}\,)$$

e consequentemente a variância estimada de $\hat{p}$ simplifica para:

$$\hat{V}ar(\,\hat{p}\,) = \left(1-\frac{n}{N}\right)\frac{\hat{p}\,(\,1-\hat{p}\,)}{n-1}$$

As fórmulas para o cálculo de tamanho de amostra também são análogas às descritas na seção 5.0. A única diferença é que no caso de estimação de proporções a variância populacional σ^2 poderá ser aproximada pela quantidade:

$$\hat{\sigma}^2 = P(\,1-P\,) = 1/4$$

que é a variância populacional máxima. Portanto, por este procedimento o valor obtido de n é o maior dentre todos os conjuntos de valores possíveis que satisfazem as especificações do pesquisador. Caso se disponha de alguma estimativa prévia de *P* podemos utilizar esta estimativa na fórmula de aproximação da variância populacional. Esta alternativa poderá diminuir os custos da pesquisa.

6.1 Exemplos de Aplicação

Exemplo 1. Uma fábrica de tecidos é responsável pela produção de diversos tipos de tecidos, desde a tecelagem, banho e tintura, até a expedição para o cliente. Antes do produto ser submetido ao processo de banho são feitas inspeções em amostras de tecidos, onde são verificadas a presença de falhas, como “bolinhas” e “rasgos”dentre outros defeitos que podem ser considerados nesta etapa do processo. Caso pelo menos um destes defeitos apareça o tecido é classificado como defeituoso. A inspeção dos tecidos é feita por inspetores treinados para tal tarefa. Num conjunto de 20 amostras de tecido selecionadas aleatoriamente da produção constituída de N=500, por amostragem aleatória simples sem reposição, observou-se um total de 8 amostras defeituosas. Neste caso, uma estimatíva para a proporção de tecidos defeituosos que é manufaturado por esta fábrica seria igual a:

$$\hat{p} = \frac{8}{20} = 0{,}40$$

A variância estimada de $\hat{p}$ é dada por:

$$\hat{V}ar(\hat{p}) = \left(1 - \frac{20}{500}\right)\frac{(0{,}40)(0{,}60)}{(19)} = 0{,}012126$$

Um intervalo de 95% de confiança para o verdadeiro valor da proporção de tecidos defeituosos deste processo de produção seria:

$$0{,}40 \pm (2{,}093)(\sqrt{0{,}012126} \Rightarrow 0{,}40 \pm 0{,}2304 \Rightarrow (0{,}169\ ;\ 0{,}6304)$$

Exemplo 2. Suponha que os técnicos de uma indústria de cerâmica desejem estimar a proporção média de defeitos das peças que são fabricadas pela empresa atualmente. Os técnicos decidem utilizar uma tolerância de 0,02 e uma porcentagem máxima de erro igual a 5%. Amostras anteriores do processo sugerem que a proporção de defeitos é aproximadamente igual a 0,15. Qual seria o tamanho de amostra necessário para que a estimação da proporção média seja feita de acordo com o que os técnicos especificaram ?

Vamos supor que a empresa fabrique um volume bastante grande de peças rotineiramente. Neste caso, podemos utilizar a fórmula de amostra aleatória simples com reposição para o cálculo de n. Considerando a estimativa anterior de 0,15 temos uma aproximação para o desvio padrão populacional que é,

$$\hat{\sigma}^2 = 0{,}15(0{,}85) = 0{,}1275$$

Então, o tamanho de amostra seria:

$$n = \left(\frac{1{,}96}{0{,}02}\right)^2 (0{,}1275) = 1224{,}51 \quad peças$$

Os técnicos da empresa resolveram adotar o seguinte esquema amostral: um total de n=1230 peças foram observadas através da seleção de k=10 amostras, cada amostra com m=123 peças cada. As 11 amostras foram selecionadas ao longo da linha de produção em intervalos de tempo pré-determinados pelos técnicos. Portanto, ao final obteve-se 10 estimativas ($\hat{p}_i$) para a proporção de defeitos deste processo de produção que podem ser combinadas do seguinte modo, para estimar a proporção média de defeitos do processo:

$$\overline{p} = \frac{1}{k} \sum_{i=1}^{k} \hat{p}_i$$

estimador muito conhecido na área de controle de qualidade.

É interessante observar que se o valor da tolerância for aumentado para d=0,05 o valor de n será igual a 195,92, ou seja aproximadamente 196. Portanto, uma diminuição bastante considerável. Por outro lado, se optarmos por utilizar o valor de ¼ como estimativa para a variância populacional, teríamos para um d=0,02 um valor n = 2401 e para d = 0,05 um valor n = 385.

CAPÍTULO 3: AMOSTRAGEM SISTEMÁTICA

1.0 Introdução

A Amostragem Sistemática é um procedimento amostral muito utilizado devido a sua facilidade de implementação operacional. Basicamente, as unidades amostrais populacionais são numeradas de 1 a N. Através de algum processo aleatório, seleciona-se uma unidade amostral dentre as N. Se esta unidade ocupar a posição de ordem R, as outras unidades que pertencerão a amostra final serão aquelas que ocuparem as posições:

$$R + Zk \quad onde \quad k \cong \frac{N}{n} \quad e \quad Z = 1,2,\ldots,(n-1).$$

onde n é o tamanho da amostra, e k é denominado de período de seleção (é o espaçamento entre as unidades amostrais pertencentes a amostra final). Quando N=nk as N unidades amostrais ficam divididas em n grupos de tamanho k. Entretanto, problemas aparerem quando N $\neq$ nk. Para ilustrar considere as seguintes situações:

(a) Suponha N=12 , n=4 e k=3. Então teríamos a divisão :

1 2 3 | 4 5 6 |7 8 9 |10 11 12

ou seja 4 grupos de tamanho 3. Seja qual for a unidade amostral escolhida na primeira seleção, as amostras possíveis teríam sempre o tamanho n=4.

(b) Suponha N=14, n=4 então k=3,5 que poderia ser aproximado por 3 ou 4. Se escolhermos a aproximação para k=4 teremos a seguinte divisão:

1 2 3 4 | 5 6 7 8 | 9 10 11 12 | 13 14

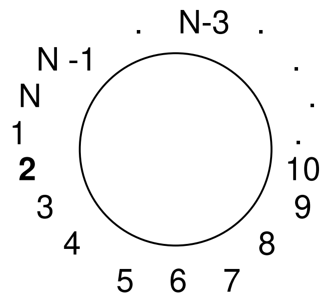
Neste caso, se na primeira seleção fosse escolhida a unidade amostral 1 ou 2 teríamos uma amostra com n=4 unidades como planejado. O mesmo não valeria se fosse escolhido as unidades 3 ou 4, pois neste caso as amostras teríam tamanho n=3.

Caso optemos pela aproximação para k=3 teríamos a divisão

1 2 3 | 4 5 6 | 7 8 9 | 10 11 12 | 13 14

Neste caso, se na primeira seleção fosse escolhida as unidades amostrais 1 ou 2 teríamos uma amostra com tamanho n=5 (maior que o planejado) e se a unidade amostral escolhida fosse a 3 teríamos a amostra com n=4.

Uma forma alternativa que evita problemas em relação ao fato de N ser ou não um múltiplo de k, é o método de Lahiri (1952). Neste método as unidades populacionais são consideradas como se estivessem dispostas num círculo (Lariri,1952). Desta forma, ao chegar-se na unidade amostral N a seleção continuaria a partir da unidade amostral 1 respeitando o valor do período de seleção k. Esquematicamente teríamos:



Por exemplo, se tivermos 14 residências em uma rua e desejarmos selecionar n=4 delas para uma entrevista com k=4 uma amostra possível seria aquela contendo as residências de ordem: 5,9,13 e 3, caso a 5a. residência da lista fosse a primeira a ser selecionada para a amostra.

O espaçamento k pode estar também relacionado com o tempo por exemplo, o que faz com que a amostragem sistemática seja muito utilizada na avaliação da qualidade de produtos cuja linha de produção é contínua, como minério de ferro,manganês, carvão, tubos de pvc, fios de eltricidade, etc. Neste caso, podemos pensar em selecionar incrementos de carvão da linha de produção de 15 em 15 minutos e observar a porcentagem de cinzas de cada incremento ou de um conjunto de incrementos.

Portanto, o esquema da amostragem sistemática é um processo alternativo para a aleatorização e para a escolha aleatória das unidades amostrais podendo ser utilizada em substituição à amostragem alatória simples. Algumas vezes a amostragem sistemática é denominada de “pseudo-aleatória”, pelo fato de ter apenas 1 unidade realmente selecionada por algum mecanismo aleatório, como por exemplo um gerador de números aleatórios computacional. Assim como a Amostragem Aleatória Simples, em geral a Amostragem Sistemática é utilizada em conjunto com a Amostragem Estratificada e Amostragem por Conglomerados, procedimentos que serão tratados nos capítulos 5 e 6 deste texto.

Uma vantagem da amostragem sistemática é que ela pode também ser utilizada em situações que envolvem mais de um dimensão, por exemplo na amostragem de rios, áreas onde a região de amostragem (população) poderia ser vista em duas coordenadas do sistema cartesiano (como Norte e Sul) ou em 4 direções como Norte, Sul, Leste e Oeste (ver Sukthame,1970 ou Thompson,1992).

Na indústria a amostragem sistemática torna-se bastante útil pela simplicidade de implementação e pela flexibilidade, especialmente em situações nas quais é impossível construir um frame como é necessário para a amostragem aleatória simples. Amostragem de líquidos e granulados são casos onde a amostragem sistemática podem ser utilizados. Nestes casos, o período de seleção k seria escolhido de acordo com o tempo ou fluxo de produção.

2.0 Estimação de Parâmetros Populacionais

Suponha que o procedimento de Amostragem Sistemática tenha sido utilizado para obtenção de uma amostra aleatória de tamanho n de uma população com N unidades amostrais. Suponha ainda que o período k satisfaça a igualdade N=nk, ou que o método de Lahiri tenha sido utilizado para a implementação do método sistemático. Neste caso, os estimadores não viciados da média e do total populacional são dados respectivamente por:

$$\hat{\bar{Y}} = \bar{y}_{sy} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad \text{(estimador da média populacional)}$$

$$\hat{Y} = N\ \bar{y}_{sy} \qquad \text{(estimador do total populacional)}$$

Uma estimativa da variância de $\bar{y}_{sy}$ pode ser obtida pelos seguintes métodos:

Método 1 : Aproximação pela Amostragem Aleatória Simples

Neste método é assumido que as unidades amostrais populacionais foram totalmente misturadas antes de terem sido numeradas para efeito da implementação da amostragem sistemática. A amostragem sistemática seria então, semelhante ao esquema da amostragem aleatória simples sem reposição. Portanto, uma estimativa da variância de $\bar{y}_{sy}$ seria,

$$\hat{Var}(\bar{y}_{sy}) = \frac{s^2}{n}\left(\frac{N-n}{N}\right)$$

onde s^2 é uma estimativa do quadrado médio populacional descrita no módulo de amostragem aleatória simples.

Em problemas práticos a qualidade desta aproximação pode ser verificada. Considerando-se a forma alternativa de expressar-se a variância de $\bar{y}_{sy}$, dada por:

$$Var(\bar{y}_{sy}) = \frac{\sigma^2}{n}\left[1+(n-1)\rho_w\right],$$

onde σ^2 é o variância populacional e ρ_w é uma medida da correlação entre os pares de elementos dentro da mesma amostra sistemática (correlação intra-classe), tem-se que se ρ_w for próximo de 1, os elementos da amostra são todos bastante semelhantes com relação à característica que foi medida nas unidades amostrais e, consequentemente a variância de $\bar{y}_{sy}$ será maior que a variância de $\bar{y}$ da amostragem aleatória simples. Se ρ_w for negativo, a amostragem sistemática poderá ser mais precisa que a amostragem aleatória simples. A correlação pode ser negativa se os elementos da amostra são extremamente diferentes. Quando N for muito grande e ρ_w for muito próximo de 0, a amostragem sistemática é aproximadamente equivalente à amostragem aleatória simples.

Uma estimativa amostral de ρ_w é dada por:

$$\hat{\rho}_w = \frac{\sum\limits_{i=1}^{n}\sum\limits_{\substack{j=1\\ j\neq i}}^{n}(y_i-\bar{y}_{sy})(y_j-\bar{y}_{sy})}{\left[\frac{n(n-1)}{2}\right]\left(\frac{N-1}{N}\right)s^2}$$

A presença de tendências ou desvios em relação à aleatoriedade da enumeração que foi realizada para dar início ao seleção da amostra sistemática pode aumentar grandemente a variância e a aproximação pela amostragem aleatória simples não será adequada (ver Cochran,1977).

Método 2: Amostragem Sistemática Repetida

Neste método, considera-se que a amostra sistemática de tamanho n e período de seleção *k,* possa ser dividida em *c* amostras sistemáticas distintas cada uma de tamanho *m*, ou seja considera-se que *n = cm*. O período de seleção de cada amostra de tamanho *m* seria *k* = c k* Neste caso, as estimativas das variâncias de $\bar{y}_{sy}$ e $\hat{Y}$ são dadas por:

$$\hat{V}ar(\bar{y}_{sy}) = \frac{\sum\limits_{i=1}^{c}(\bar{y}_{sy_i}-\bar{\bar{y}}_s)^2}{c(c-1)} \qquad onde \quad \bar{\bar{y}}_s = \frac{\sum\limits_{i=1}^{c}\bar{y}_{sy_i}}{c} \quad e \quad \bar{y}_{sy_i}$$ é a média da

i-ésima amostra sistemática repetida, i=1,2.

O valor de m é escolhido pelo pesquisador. Não há uma regra geral mas aconselha-se a não utilizar um valor de m muito pequeno, pois as estimativas podem ser muito imprecisas especialmente quando o tamanho da amostra n é pequeno.

Por exemplo, suponha N=18 , n=6 e k=3. Neste caso tem-se que n=2x3 , o que implica que poderíamos dividir a amostra sistemática de 6 unidades amostrais em c=2 amostras de tamanho m=3, ou c=3 amostras de tamanho m=2. No primeiro caso teríamos k*=6 e no segundo k*=9. Suponha que a amostra sistemática fosse constituída das unidades 3, 6, 9, 12, 15 e 18. Então, se considerássemos c=2 teríamos duas amostras sistemáticas repetidas, uma constituída das unidades 3, 09 e 15 e outra com as unidades 6, 12 e 18. Neste caso k* seria igual a 6. Se considerássemos c=3 teríamos três amostras sistemáticas repetidas, uma com as unidades 3 e 12, outra com 6 e 15 e a última com as unidades 9 e 18. Neste caso, k* seria igual a 9.

No software ***"Sampling"*** o método de amostragem sistemática repetida é efetuada para valores de c iguais a 2 ou 3.

Método 3: Método das Diferenças Sucessivas

Neste método uma estimativa da variância de $\bar{y}_{sy}$ é calculada através da seguinte fórmula (Mingoti e Aguiar, 1988):

$$\hat{V}ar(\bar{y}_{sy}) = \left(\frac{N-n}{n}\right)\frac{\sum_{j=1}^{n-1}(y_{j+1}-y_j)^2}{2n(n-1)} \quad , \; onde \;\; y_j \;\; e \;\; y_{j+1}$$

são as observações sucessivas obsvervadas na amostra sistemática de tamanho n. Por exemplo, suponha que uma amostra sistemática de n=4 unidades fosse selecionada de uma população com N=40 unidades amostrais. Se os valores da variável resposta Y observados na amostra fossem iguais a 10,8,6 e 5, então a média amostral seria igual a 7,25 e uma estimativa da variabilidade da média amostral seria estimada por:

$$\hat{V}ar(\bar{y}_{sy}) = \left(\frac{1-\frac{4}{40}}{2(4)(3)}\right)\left[(10-8)^2+(8-6)^2+(6-5)^2\right]$$

$$= 9\,\frac{0,90}{24} = 0,337 \quad \Rightarrow \quad \sqrt{\hat{V}ar(\bar{y}_{sy})} = 0,58$$

Nos três métodos de estimação a variância do estimador do total populacional é dada por:

$$\hat{V}ar(\hat{Y}) = N^2\,\hat{V}ar(\bar{y}_{sy})$$

Método 4: Várias Amostras Sistemáticas

Uma outra possibilidade de obter-se uma estimativa da variância de $\overline{y}_{sy}$ sem a necessidade de utilizar-se algum tipo de aproximação, é aquela em que o usuário dispõe de várias amostras sistemáticas da população que está sendo investigada. Suponha que se disponha de m amostras sistemáticas de tamanho n e todas de período de seleção igual a *k*. Então, um estimador não viciado da média amostral será dado por (Scheaffer, Mendenhall e Ott, 1996):

$$\overline{y}_s = \frac{1}{m} \sum_{i=1}^{m} \overline{y}_{sy_i} \quad , \; onde \quad \overline{y}_{sy_i}$$ são as respectivas médias amostrais

das *m* amostras sistemáticas observadas.

Um estimador não-viciado da variância deste estimador é dada por:

$$\hat{V}ar(\overline{y}_s) = \frac{1}{m(m-1)} \sum_{i=1}^{m} (\overline{y}_{sy_i} - \overline{y}_s)^2$$

3.0 Exemplos de Aplicação

Exemplo 1: Suponha que os técnicos de uma empresa mineradora desejem estimar o teor médio de ferro(%) do minério que será estocado num determinado silo durante o período de 6 horas. Com este objetivo foram selecionados 24 incrementos, um a cada 15 minutos, a partir de um início aleatório a partir do instante que o processo de estocagem do minério foi iniciado. Os resultados estão apresentados na tabela 3.1 a seguir:

Tabela 3.1: Exemplo do minério de ferro

No. da amostra	*Teor de Ferro (%)*
1	61,40
2	61,30
3	61,20
4	61,00
5	61,10
6	61,00
7	61,93
8	61,50
9	61,42
10	61,15
11	61,30
12	61,54
13	61,32
14	61,22
15	61,46
16	61,96
17	61,05
18	61,00
19	61,10
20	61,35
21	61,02
22	61,02
23	61,01
24	61,40

Utilizando se o software ***"Sampling"*** obteríamos as seguintes estimativas para o teor médio de ferro (%) do minério estocado no silo:

Estimação Global

. Tipo de Amostragem : Amostragem Sistemática
. Método 1 : Amostragem Aleatória Simples
. Método 2 : Amostragem Sistemática Repetida
. Método 3 : Diferença

N	Mean
4000000	61.2882

StDev Mean

Method 1	Method 2	Method 3
0.0586301	0.0118160	0.0537267

IC Mean

Method 1	Method 2	Method 3
(61.1873, 61.3891)	(61.2679, 61.3085)	(61.1957, 61.3806)

Estrato	Correlation
1	-0.0000005

. T Value : 1.72075 , $\alpha = 0,05$

Como o minério é misturado ao longo do processo de produção, a aproximação da variância pela amostragem aleatória simples seria um método apropriado, fato que é refletido na estimativa do coeficiente de correlação ρ_w próxima de zero (-0,0000005).

Exemplo 2: Uma empresa está interessada em estimar o número médio de pessoas que visitam a empresa em veículos automotores. Sabe-se que em média a empresa recebe a visita de aproximadamente N= 60 veículos por dia (período de 7 as 18 hs), entretanto nem sempre estes veículos estão totalmente ocupados. Com este objetivo a empresa decidiu realizar o seguinte procedimento amostral: o porteiro da empresa deveria parar 1 a cada 4 veículos que entrassem na empresa em um determinado dia da semana escolhido pela equipe responsável pela amostragem. Deveria então, anotar o número de passageiros constantes do veículo. No final do dia, espera-se ter um total de 15 veículos observados. Suponha que dentre os 15 veículos tenha sido observado um total de 35 passageiros e que a variância amostral dentre o número de passageiros em cada veículo tenha sido igual a

0,2380,ou seja este é o valor de s^2. Então, o número médio de visitantes diários a esta empresa por veículo seria estimado por:

$$\bar{y}_{sy} = \frac{35}{15} = 2,33$$

o que levaria a uma estimativa de (2,33) (60) = 139,80 ≈ 140 visitantes em média por dia. Uma estimativa da variância deste total médio de visitantes, usando a aproximação por amostragem aleatória simples seria,

$$(60)^2 \left(1 - \frac{15}{60}\right) \frac{(0,2380)}{15} = 0,0119048$$

$$(60)^2 \left(1 - \frac{15}{60}\right) \frac{(0,2380)}{15} = 42,48$$

o que indica um desvio padrão igual a 6,51 visitantes.

4.0 Cálculo de Tamanho de Amostra

Para o cálculo de tamanho de amostra deve ser utilizado as fórmulas apresentadas no capítulo 2 referente a Amostragem Aleatória Simples.

Exemplo de Aplicação: Suponha que se saiba que a variabilidade do teor de ferro do minério de um determinado tipo que será estocado num determinado silo seja aproximadamente igual 1,4 $(\%)^2$ (valor estimado por amostragens anteriores). O silo tem capacidade para aproximadamente 4000 toneladas. Neste caso, o tamanho da amostra necessário para que se estimasse o teor médio de ferro do minério que será estocado com uma tolerância igual a 0,3 % e com uma probabilidade de erro igual a 95 %, seria dado por:

$$n = \frac{(1,96)^2 \ (1,4)}{(0,3)^2} = 59,43 \approx 60 \ \ incrementos$$

Como a amostragem é sistemática e o tempo de enchimento do silo é de 12 horas, o intervalo entre as seleções dos incrementos deverá ser igual a k=12 minutos. O primeiro ponto de amostragem deverá ser um tempo escolhido, aleatoriamente, entre no intevalo de 1 a 12 minutos. Suponha que o ponto escolhido seja 3 minutos. Então, os demais tempos de seleção de incrementos serão, em minutos,

3 , 3 + 12 = 15 , 3 + 2(12) = 27, . . . , 3 + 59(12) = 711

CAPÍTULO 4: AMOSTRAGEM DO TIPO P.P.S.

1.0 Introdução

Os esquemas amostrais tratados nos capítulos 2 e 3 assumem que todas as unidades amostrais populacionais têm a mesma probabilidade de pertencer a amostra final de tamanho n, ou seja, em cada estágio de seleção cada unidade amostral tem a mesma chance de ser selecionada. Deste modo, todas as unidades têm o mesmo grau de importância. Entretanto, muitas vezes é interessante utilizar um esquema no qual as unidades amostrais têm diferentes probabilidades de seleção. Por exemplo, numa pesquisa que envolva a amostragem de indústrias é natural imaginarmos que indústrias maiores e mais competitivas tenham maior impacto no mercado de trabalho e no mercado econômico. As taxas de emprego e de desemprego por exemplo, são altamente influenciadas pelas indústrias mais importantes do setor. Deste modo, não seria muito aceitável utilizar-se um esquema amostral que tratasse todas as indústrias como igualmente importantes. Parece ser mais razoável utilizar um esquema no qual indústrias mais representativas do setor tivessem uma maior probabilidade de pertencer a amostra do que aquelas indústrias de menor porte e que não causam grandes impactos no mercado. Numa amostragem que envolvesse fazendas, por exemplo, com o objetivo de estimar-se a produção média ou total de algum tipo de produto cultivado, seria natural aceitarmos que fazendas com maiores áreas de cultivo tivessem uma probabilidade maior de pertencer a amostra final do que aquelas com menores áreas de cultivo. Portanto, nestes casos os esquemas que permitem que unidades amostrais populacionais tenham diferentes probabilidades de seleção seriam preferidos em relação aos procedimentos de Amostragem Aleatória Simples ou Sistemática.

Um caso particular é aquele em que as probabilidades de seleção estão relacionadas com o "tamanho" da unidade amostral, tamanho este medido por algum critério. Se as probabilidades de seleção são escolhidas proporcionalmente ao tamanho de cada unidade amostral tem-se o procedimento amostral conhecido como P.P.S. (esta sigla vem do inglês: sampling with probabilities proportional to size). A seguir descreveremos brevemente os aspectos principais deste tipo de procedimento amostral.

2.0 Estimação dos Parâmetros Populacionais

Suponha que a população de interesse tenha sido dividida em N unidades amostrais aqui denotadas por $U_1, U_2, \ldots, U_N$. Seja π_i a probabilidade de seleção da unidade amostral U_i , $i = 1,2,\ldots,N$, $\sum_{i=1}^{N} \pi_i = 1$. Se uma amostra aleatória de tamanho n for retirada com reposição desta população de acordo com estas probabilidades de seleção, então os estimadores não-viciados para o total e a média populacional seriam dados por:

$$\hat{\tau}_\pi = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

para o total populacional com variância estimada dada por:

$$\hat{Var}(\hat{\tau}_{\pi}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\frac{y_i}{\pi_i} - \hat{\tau}_{\pi} \right)^2$$

e

$$\bar{y}_{\pi} = \frac{1}{nN} \sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

para a média populacional com variância estimada dada por:

$$\hat{Var}(\bar{y}_{\pi}) = \frac{1}{N^2 n(n-1)} \sum_{i=1}^{n} \left(\frac{y_i}{\pi_i} - \hat{\tau}_{\pi} \right)^2$$

Estes estimadores são não-viciados para qualquer escolha de valores π_i . O mais interessante entretanto, é encontrar o conjunto de valores de π_i que resulte em estimadores de maior precisão, ou seja mínima variância. A solução seria escolher-se π_i igual a $\frac{y_i}{Y}$, onde Y é o total populacional e y_i e o valor da variável resposta para a unidade amostral U_i correspondente. Neste caso, a variância dos estimadores $\bar{y}_{\pi}$ e $\hat{\tau}_{\pi}$ seriam iguais a zero. Entretanto, esta escolha é impossível na prática, pois se o fosse não precisariamos realizar a amostragem uma vez que toda a população seria conhecida. A escolha mais natural seria então, escolher alguma variável que fosse altamente correlacionada com y_i e que pudesse ser utilizada para a escolha das probabilidades de seleção. Por exemplo, no caso das taxas de emprego e desemprego, sabe-se que estas variáveis estão altamente correlacionadas com o tamanho (ou porte) das empresas. Naturalmente, as empresas maiores podem oferecer mais oportunidades de emprego. Ao mesmo tempo quando estas empresas decidem pela demissão de empregados o impacto no mercado de trabalho é muito grande e ocorre naturalmente uma elevação na taxa de desemprego. Talvez o exemplo mais próximo da realidade brasileira seja aquele relacionado com as indústrias automobilísticas. Deste modo, uma possibilidade seria a de escolher-se as probabilidades de seleção levando-se a variável “tamanho” em consideração. Por exemplo, se uma indústria U_i tem um certo número M_i de empregados, então uma escolha lógica para π_i seria $\pi_i = \frac{M_i}{M}$ onde $M = \sum_{i=1}^{N} M_i$ seria o número total de empregados entre todas as empresas do *Frame*. Este procedimento amostral é chamado de P.P.S., ou seja as probabilidades de seleção são escolhidas de modo proporcional ao tamanho das unidades amostrais populacionais. Em geral, quando o método P.P.S. é utilizado os estimadores dados anteriormente são denotados por:

$$\overline{y}_{\pi} = \overline{y}_{pps} \quad e \quad \hat{\tau}_{\pi} = \hat{\tau}_{pps} = \hat{Y}_{pps}$$

$$\overline{y}_{\pi} = \overline{y}_{pps} \quad e \quad \hat{\tau}_{\pi} = \hat{\tau}_{pps} = \hat{Y}_{pps}$$

Os intervalos de Confiança para os parâmetros populacioansi são construídos de modo similar ao descrito nos capítulos 2 e 3.

Quando as probabilidades são escolhidas por este procedimento, ou seja P.P.S., a amostragem com probabilidades de seleção diferentes resulta em maior precisão se comparada com a Amostragem Aleatória Simples ou Sistemática. Entretanto, para outras escolhas de π_i os estimadores poderão ter variâncias maiores que aqueles produzidas por estes dois procedimentos amostrais (ver Scheaffer, Mendenhal e Ott,1996, para maiores detalhes.

Exemplo de Aplicação: Os técnicos de uma empresa que produz placas de circuítos eletrônicos desejam estimar o número médio de defeitos por placa. As placas que são fabricadas por esta empresa variam de acordo com o número de componentes. Suponha que se tenha um total de N=10 placas e se deseje obter uma amostra aleatória de tamanho n=4 pelo método P.P.S. com reposição. A tabela a seguir apresenta os números de componentes de cada placa e as respectivas probabilidades de seleção de cada placa.

Tabela 2.1 : Exemplo das placas de circuítos

Placa	Número de Componentes	Probabilidade π_i
01	10	10/150
02	12	12/150
03	22	22/150
04	08	8/150
05	16	16/150
06	24	24/150
07	09	9/150
08	10	10/150
09	08	8/150
10	31	31/150
Total	150	

Suponha que tenham sido selecionadas as placas 2,3,5 e 7 e que o número de defeitos encontrados em cada uma destas placas tenham sido: 1, 3, 2 e 1. Neste caso, uma estimativa não-viciada para o número médio de defeitos por placa será:

$$\overline{y}_{pps} = \frac{1}{10(4)}\left[1(\frac{150}{12}) + 3(\frac{150}{22}) + 2(\frac{150}{16}) + 1(\frac{150}{9})\right] = 1{,}71$$

com desvio padrão estimado igual a :

$$\sqrt{\hat{V}ar(\overline{y}_{pps})} = \sqrt{0{,}0295} = 0{,}1718$$

Um intervalo de 95% de confiança para o valor verdadeiro do número médio de defeitos por placa será:

$$1{,}71 \pm 1{,}96\ (0{,}1718) \Rightarrow (1{,}3733\ ;\ 2{,}0467)$$

Como ilustração, uma forma de selecionar-se as 4 placas para a amostra seria a seguinte: Do conjunto números de 1 a 150, poderíamos relacionar os números de 1 a 10 com a primeira placa , os de 11 a 22 com a segunda placa, os de 23 a 44 com a terceira placa, e assim por diante até termos os números 120 a 150 relacionados com a décima placa. Seriam então gerados 4 números aleatóriamente com reposição entre 1 e 150 através de um mecanismo aleatório qualquer. A placa que tivesse o número correspondente ao sorteado seria aquela que entraria para a amostra. Por exemplo, se fossem gerados os números 9, 25, 95 e 150, as placas 1, 3, 7 e 10 seriam aquelas que iriam compor a amostra de n=4 placas.

Outros exemplos podem ser encontrados em Thompson (1992) e no capítulo 6 deste texto.

CAPÍTULO 5: AMOSTRAGEM ESTRATIFICADA

1.0 Introdução

A amostragem estratificada é um dos esquemas mais utilizados em problemas práticos.Sua finalidade é de melhorar a precisão das estimativas dos parâmetros populacionais através do controle “à priori” de fatores que podem interferir nas respostas dos elementos amostrais e que naturalmente aumentam a variabilidade dos estimadores usuais da amostragem aleatória simples e sistemática. Como um exemplo, suponha que uma empresa esteja investigando as causas relativas a um determinado problema que está ocorrendo no processo de produção. Há vários operadores que trabalham no sistema e várias máquinas. Deste modo, se uma amostra do produto for selecionada do processo de produção sem levar em consideração o fator “humano” e o fator “máquina”, as estimativas da média e variabilidade da característica de qualidade de interesse poderá ter pouca precisão caso haja uma diferença significativa de comportamento entre os operadores e entre as máquinas quando operando no processo. Além disso, um esquema que não levasse estes fatores em consideração não permitiria uma investigação sobre o relacionamento dos operadores e das máquinas com algum problema central que estaria ocorrendo no processo de produção da empresa. Uma alternativa melhor seria um esquema que tivesse uma amostra de produtos manufaturados para cada combinação (operador ; máquina) que atuam no sistema de produção. Deste modo, seria possível avaliar se o problema estaria relacionado com o comportamento dos operadores, ou com o funcionamento das máquinas, ou a interação de ambos, por exemplo.

Num caso mais geral, quando a pesquisa envolve o estudo de populações humanas espera-se “a priori” que haja uma certa diferença de comportamento entre os vários grupos humanos devido entre outras razões, pelas diferenças de classe social, faixa etária e nível educacional por exemplo. Deste modo, amostras que tivessem a representação de faixa etárias e classes socais e econômicas diferentes, seriam mais informativas. Quando a amostragem aleatória simples é utilizada em populações humanas como esta, não há garantia que a amostra final terá este tipo de representatividade. A forma de se garantir isto, dividindo-se a população em várias faixas etárias, e em várias classes econônmicas e sociais, e retirando-se uma amostra aleatória de cada combinação do tipo (faixa etária; classe social e econômica).

Situações como estas são resolvidas através da utilização do esquema amostral chamado de Amostragem Estratificada cujas idéias básicas serão descritas a seguir.

2.0 Amostragem Estratificada – Noções Gerais

Uma amostra estratificada é obtida dividindo-se a população de N elementos em L grupos distintos chamados de estratos e escolhendo-se em seguida uma amostra de elementos de cada estrato. As L amostras precisam ser independentes. Quando o procedimento amostral utilizado na seleção das amostras de cada estrato for o de Amostragem Aleatória Simples, o procedimento é chamado de Amostragem Aleatória Estratificada. Outro procedimento muito comum utilizado na seleção da amostra de cada estrato é o de Amostragem Sistemática. A estratificação pode ainda estar combinada com a Amostragem por Conglomerados ou a Amostragem em vários estágios, procedimento conhecido como subamostragem estratificada(ver capítulo 7) e com métodos de estimação de parâmetros como o de Razão e o de Regressão (Sukhatme e Sukthatme,1970, Thompson,1992).

A estratificação é essencial na estimação de parâmetros de populações que apresentam grande heterogeneidade dos elementos no que se refere às variáveis resposta de interesse. Deste modo, uma maior precisão pode ser obtida através da partição dos N elementos populacionais em L grupos, ou estratos, de forma que haja homogeneidade dos elementos em cada estrato e heterogeneidade entre os estratos formados. Os estratos precisam ser disjuntos. Além disso, o custo da coleta dos dados pode ser reduzido.

Para que a partição da população seja feita de forma eficiente é necessário que se tenha conhecimento de variáveis auxiliares que possam estar relacionadas, ou correlacionadas com as variáveis respostas de interesse. As variáveis auxiliares são em geral denominadas de fatores de estratificação. Alguns exemplos comuns são sexo, idade, classe social, linhas e turno de produção diferentes, etc.

No caso da amostragem estratificada, os estimadores da média e total populacionais são obtidos através de combinações lineares dos respectivos estimadores de cada estrato individualmente. Partindo-se do pressuposto de que há independência entre as amostras obtidas nos vários estratos os estimadores são construídos usando-se a informação correspondente ao procedimento amostral usado em cada estrato.

Para facilitar o entendimento do leitor na exposição das seções que seguem, a seguinte notação sera utilizado no decorrer do texto:

L : número de estratos, em que a população foi dividida,

N_h : numero de unidades amostrais em cada estrato (h=1,2,..., L),

N : número de unidades amostrais na população, $\sum_{h=1}^{L} N_h = N$,

$W_h = \dfrac{N_h}{N}$: peso do estrato h (h=1,2,..., L)

n_h : tamanho amostral no estrato h (h=1, 2, ..., L),

y_{hj} : valor da variável resposta para a j-ésima observação no estrato h (h=1,2,..., L : j=1,2,..., N_h),

Y_h : total populacional do estrato h

Y : total da população

$\vec{Y}$: média da população,

$\overline{Y}_h$: média populacional do estrato h (h=1,2,..., . L),

$\overline{y}_h$: média amostral do estrato h (h=1, 2, ..., L).

3.0 Amostragem Aleatória Simples sem Reposição em cada Estrato

Quando as amostras de cada estrato são selecionadas independentemente pelo procedimento de Amostragem Aleatória Simples sem reposição, o estimador não-viciado da média populacional é dado por:

$$\overline{y}_{st} = \sum_{h=1}^{L} W_h \overline{y}_h$$

onde $\bar{y}_h$ é a media amostral no estrato h, h=1,2,..., L. Neste caso, a variância estimada de $\bar{y}_{st}$ é dada por :

$$\hat{V}(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \hat{V}(\bar{y}_h)$$

$$= \sum_{h=1}^{L} W_h^2 (1 - \frac{n_h}{N_h}) \frac{s_h^2}{n_h}$$

onde s_h^2 é a variância amostral do estrato h, definida por

$$s_h^2 = \frac{1}{n_h - 1} \sum_{h=1}^{N_h} (y_{hj} - \bar{y}_h)^2$$

O estimador não viciado do total populacional é dado por $\hat{Y} = N\,\bar{y}_{st}$ e a variância de $\hat{Y}$ é dada por $V(\hat{Y}) = N^2 V(\bar{y}_{st})$.

Os intervalos de confiança para a média e total populacionais podem ser construídos de acordo com o nível de significância escolhido pelo pesquisador. Para amostras grandes utiliza-se a distribuição normal enquanto que para amostras pequenas utiliza-se a distribuição t de Student, isto é:

$\bar{y}_{st} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{V}ar(\bar{y}_{st})}$, para amostras grandes isto é, $n_h \geq 30$, $\forall h$

$\bar{y}_{st} \pm t_{\frac{\alpha}{2},g} \sqrt{\hat{V}ar(\bar{y}_{st})}$, para amostras pequenas

sendo os graus g de liberdade calculados de acordo com a aproximação sugerida por Satterthwaite (1946), isto é,

$$g = \frac{\left(\sum g_h\, s_h^2\right)^2}{\sum \frac{g_h^2\, s_h^4}{n_h - 1}} \text{, onde } \quad g_h = \frac{N_h\,(N_h - n_h)}{n_h}$$

O valor de g está sempre entre o menor valor de $(n_h - 1)$ e a soma destes valores. Se todas os tamanhos populacionais N_h forem iguais e todos os tamanhos amostrais n_h forem iguais g será igual a n – *L*. Algumas possíveis modificações na fórmula sugerida por Satterthwaite podem ser encontradas em Ames e Webster (1991). Para o total populacional o intervalo de confiança pode ser obtido multiplicando-se por N, os limites de confiança calculados para a média populacional.

4.0 Amostragem Estratificada com outros Esquemas de Amostragem dentro dos Estratos

No caso em que a amostragem aleatória simples com reposição, ou amostragem sistemática, for utlilizada em cada estrato a média populacional continuará sendo estimada por $\bar{y}_{st}$ como dado anteriormente. Entretanto, a variância deste estimador torna-se igual a:

$\hat{V}(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \dfrac{s_h^2}{n_h}$ no caso da amostragem aleatória simples com reposição

$\hat{V}(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \; \hat{V}ar(\bar{y}_h)$ no caso da amostragem sistemática

onde $\hat{V}ar(\bar{y}_h)$ é calculada de acordo com os métodos 1,2 e 3 descritos no capítulo 3.

De um modo geral, os esquemas de amostragem podem variar de estrato para estrato. A amostragem aleatória simples com reposição e sem reposição e a amostragem sistemática são alguns dos procedimentos mais utilizados. Entretanto nada impede que o pesquisador opte por um procedimento de amostragem diferente. Neste caso, para que ele possa calcular a estimativa da variância do estimador $\bar{y}_{st}$ correspondente é necessário que ele saiba estimar a variância dos estimadores da média de cada estrato de acordo com o esquema amostral que foi utilizado em cada estrato, isto é, é necessário que se saiba teoricamente a fórmula correspondente a $Var(\hat{\theta}_h)$ onde $\hat{\theta}_h$ é o estimador da média populacional do h=ésimo estrato da população, como também que se tenha uma estimativa para esta quantidade. Deste modo, a estimativas para a média populacional e para a variância deste estimador serão dadas por:

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \, \hat{\theta}_h$$

$$\hat{V}(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \; \hat{V}ar(\hat{\theta}_h)$$

Estes estimadores serão não viciados caso $\hat{\theta}_h$ e $\hat{V}ar(\hat{\theta}_h)$ sejam estimadores não-viciados para a média do estrato h e para a $Var(\hat{\theta}_h)$ respectivamente.

5.0 Exemplos de Aplicação

Exemplo1: Um dos produtos comercializados por uma indústria de mineração é o "Sinter Feed B" que é produzido em 3 tipos diferentes (SF1, SF2, SF3). Uma das características de qualidade deste produto é o teor de ferro. Para estimar-se o teor médio de ferro do "Sinter Feed B" empilhado durante um dia de produção (24 horas) adotou-se o seguinte procedimento amostral (Mingoti & Aguiar, 1988):

(a) os Sinter Feed SF1, SF2 e SF3 produzidos ao longo do dia foram pesados;

(b) de 2 em 2 horas foi retirada um incremento do SF1, sendo que cada incremento foi analisado individualmente no laboratório da empresa.

Portanto, no final do dia os resultados de 12 análises foram obtidos.Os procedimentos amostrais utilizados para avaliação do SF2 e do SF3 são análogos ao do SF1.Neste caso, temos um exemplo de Amostragem Estratificada na qual cada tipo de "Sinter Feed" pode ser considerado como um estrato. O procedimento amostral usado dentro de cada estrato foi o de Amostragem Sistemática no tempo. Os resultados obtidos, nas análises efetuadas no laboratório, como também as pesagens de cada Sinter Feed produzido estão dados na tabela 5.1 a seguir:

Tabela 5.1: Exemplo da empresa de mineração

		Teor de Ferro dos Incrementos (%)											
Tipo	Peso (Kg)	1	2	3	4	5	6	7	8	9	10	11	12
SF1	16000	61,20	61,30	61,15	61,18	61,17	61,21	61,22	61,18	61,20	61,18	61,15	61,17
SF2	27400	64,00	64,10	64,18	64,20	64,15	64,16	64,20	64,24	64,21	64,16	64,25	64,15
SF3	26600	65,10	65,08	65,07	65,06	65,09	65,08	65,11	65,05	65,00	65,02	65,07	65,00
Total	70000												

As estimativas pontual e por intervalo do teor médio do minério de ferro "Sinter Feed B" empilhado, levando-se em consideração o esquema amostral adotado estão apresentadas a seguir usando-se o software ***"Sampling***". No caso da Amostragem Sistemática a estimativa da variância do estimador da média em cada estrato foi calculada pelos 3 métodos descritos no capítulo 3. Para cada estrato é apresentado também o valor do coeficiente de correlação entre as unidades amostrais, para que se possa verificar a qualidade da aproximação pela Amostragem Aleatória Simples.

Estimação Global

. Tipo de Amostragem : Amostragem Estratificada
. Tipo de Amostragem dentro dos estratos : Sistemática
. Método 1 : Amostragem Aleatória Simples
. Método 2 : Amostragem Sistemática Repetida
. Método 3 : Diferença

N	Mean
70000	63.8266

StDev Mean

Method 1	Method 2	Method 3
0.0089847	0.0070417	0.0065004

IC Mean

Method 1	Method 2	Method 3
(63.8114, 63.8418)	(63.8147, 63.8386)	(63.8156, 63.8376)

Estrato	Correlation
1	-0.0833385
2	-0.0833364
3	-0.0833365

. T Value : 1.69236 ; $\alpha = 0{,}05$

Neste caso, como a estimativa do coeficiente de correlação intra-classe para cada estrato é próxima de zero , poderíamos concluir que a aproximação da variância pela Amostragem Aleatória Simples seria razoável. Portanto, uma estimativa para o teor médio de ferro do minério produzido seria 63,8266% com um desvio padrão de 0,0089847 e um intervalo de confiança de 95% para o valor médio real do teor de ferro seria (63,8114 ; 63,8418).

Exemplo 2: Uma agência de publicidade está interessada em determinar o quanto ela deve enfatizar em propagandas de televisão em uma certa cidade. Para isso, ela decidiu conduzir uma pesquisa para estimar o número médio de horas por semana que os moradores assistem televisão. A cidade está dividida em três áreas: área A, área B e área rural. A área A foi construída em torno de uma indústria. A maioria de seus moradores trabalha nela e têm crianças em idade escolar. A área B é subúrbio de uma cidade vizinha. Os moradores são mais velhos e tem poucas crianças em casa. Existem 155 moradores na área A, 62 na área B e 93 na área rural. A tabela 5.2 a seguir mostra alguns resultados desta pesquisa:

Tabela 5.2: Exemplo da Agência de Propaganda

Áreas	n_h	$\overline{y_h}$	s_h	Min	Max
A	20	33,90	5,95	25	45
B	08	25,12	15,25	04	49
Rural	12	19,00	9,36	07	34

Utilizando o software ***"Sampling"*** obteríamos as seguintes estimativas para o número médio e o número total de horas que as pessoas desta cidade assistem televisão:

Estimação Global

. Tipo de Amostragem : Estratificada
. Tipo de Amostragem dentro do estratos : Aleatória Simples Sem Reposição
. Estimadores : Usuais

N	Mean	StDev	I.C.
310	27.6750	1.4033956	(25.2601, 30.0899)

Total	StDev	I.C.
8579.2500	435.0526503	(7830.63, 9327.87)

. T Value : 1.72075 , $\alpha = 0{,}05$

Poderíamos concluir que em média cada morador da cidade assiste 27,675 horas semanais de televisão. Um intervalo de 95 % de confiança para o valor real desta média seria (25,2601 ; 30,0899). O número total de horas semanais que os moradores da cidade assitem televisão seria estimado por 8579,25 com um intervalo de 95% de confiança igual a (7830,63 ; 9327,87).

6.0 Cáculo de Tamanho da Amostra

A escolha do valor de n e os valores de n_h, (h=1,2,..., L, $\sum_{h=1}^{L} n_h$ = n), pode ser feita levando em consideração a precisão especificada para a estimação dos parâmetros populacionais e o custo total da pesquisa. Para o que se segue considere a função custo dada por:

$$C = C_0 + \sum_{h=1}^{L} c_h n_h$$

onde C_0 é um custo inicial para implementação da pesquisa e c_h é o custo unitário para a observação de uma unidade amostral do estrato h (h=1,2,..., L). Seja $V(\bar{y}_{st})$ a variância do estimador da média populacional. Duas soluções são possíveis: a primeira é encontrar-se a alocação $(n, n_1, n_2, ..., n_L)$ que minimiza $V(\bar{y}_{st})$ para um custo fixo; a segunda é encontrar a alocação que minimiza o custo C para um valor de $V(\bar{y}_{st})$ fixo.

Para os dois casos a solução da alocação é a mesma e é dada por:

$$n_h = n\left(\frac{N_h \sigma_h / \sqrt{c_h}}{\sum N_h \sigma_h / \sqrt{c_h}}\right) \qquad \text{h=1,2, ..., L}$$

onde N_h é o tamanho do h-ésimo estrato, σ_h^2 é a variância e c_h é o custo para obter uma unidade de observação no h-ésimo estrato. Esta forma de alocação é denominada de Alocação Ótima.

Se o custo C for fixado, e a variância minimizada, o valor de n é dado por

$$n = \frac{(C - c_0)\sum_{h=1}^{L} N_h \sigma_h / \sqrt{c_h}}{\sum N_h \sigma_h \sqrt{c_h}}$$

Se a variância for fixada isto é, $V(\bar{y}_{st})$=V , e o custo minimizado, o valor de n é dado por :

$$n = \frac{\sum_{h=1}^{L} W_h \sigma_h \sqrt{c_h} \sum_{h=1}^{L} W_h \sigma_h / \sqrt{c_h}}{V + \frac{1}{N}\sum_{h=1}^{L} W_h \sigma_h^2}$$

Na prática nem sempre é possível utilizar-se as fórmulas de alocação ótima de imediato uma vez que estas dependem dos valores populacionais dos desvios padrões dos estratos, que operacionalmente são substituidos por estimativas. Na falta destas estimativas uma outra forma de se proceder a escolha dos tamanhos amostrais dos estratos é a partir da Alocação proporcional. Para um dado valor de n os valores n_h são escolhidos da forma :

$$n_h = W_h n = \frac{N_h}{N} n$$

Outros tipos de funções custo podem ser utilizadas e procedimentos de minimização estão descritos em Cochran (1977, pag. 97).

6.1 Exemplos de Aplicação

Exemplo1. Considere o exemplo da agência de publicidade dado na seção 5. Suponha que a empresa esteja interessada em estimar o número médio de horas por semana que os moradores assistem televisão e deseja calcular o tamanho de amostra n necessário para obter esta estimativa. Os técnicos da agência acham que obter alguma informação referente a um morador da área rural é mais caro do que obter alguma informação referente a um morador da área A ou B. O custo por morador na área A e B foi estimado como sendo de \$9.00 e o custo por morador na área rural como \$16.00. Através de um estudo piloto, os desvios padrão de cada estrato foram aproximados por $\hat{\sigma}_1 = 5h$, $\hat{\sigma}_2 = 15h$ e $\hat{\sigma}_3 = 10h$. Qual o tamanho total da amostra e o tamanho de cada um dos estratos para que:

(a) o custo seja mínimo e a variância global igual a $1h^2$.

Pelo software ***"Sampling"*** teríamos a seguinte solução

Estimação do Tamanho da Amostra

. Tipo de Alocação : Ótima
. Estimativa Minimizando : Custo
. Método de Amostragem : Estratificada
. Amostragem dentro dos estratos : Aleatória Simples Sem Reposição

. Variância Global escolhida : 1

Tamanho Total da Amostra : 57

Alocação

Estrato	Tamanho
1	18
2	22
3	17

Portanto, para que o custo seja mínimo e para que a precisão desejada na estimação seja atendida, seria necessário entrevistar 57 moradores distribuídos da seguinte forma: 18 seriam da área A , 22 da área B e 17 da área rural.

(b) A variância global seja mínima e o custo total final igual a $1000.

Estimação do Tamanho da Amostra

. Tipo de Alocação : Ótima
. Estimativa Minimizando : Variância Global
. Método de Amostragem : Estratificada
. Amostragem dentro dos estratos : Aleatória Simples Sem Reposição

. Valor de custo total escolhido : 1000

Tamanho Total da Amostra: 90

Alocação

Estrato	Tamanho
1	29
2	35
3	26

Portanto, para que o custo final da pesquisa seja $1000,00 e para que a variância do estimdor seja mínima, seria necessário entrevistar um total de 90 moradores sendo 29 da área A, 35 da área B e 26 da área rural.

Exemplo 2 Uma indústria de mineração está interessada em estimar o teor médio de ferro do Sinter Feed B, que será empilhado em 24 horas, através das estimativas do teor médio de ferro do Sinter Feed 1, Sinter Feed (2+3) e Sinter Feed 4. Experiências anteriores indicam que uma boa estimativa do desvio padrão do teor médio de ferro do Sinter Feed 1 é $\hat{\sigma}_1 = 1{,}4\%$, do Sinter Feed (2+3) $\hat{\sigma}_2 = 1{,}2\%$, e do Sinter Feed 4 é $\hat{\sigma}_3 = 1{,}1\%$. Além disso, a quantidade de minério de ferro normalmente produzida no período é:

Sinter Feed 1	-	13800 ton;
Sinter Feed (2+3)	-	23400 ton;
Sinter Feed 4	-	22800 ton.

Neste caso, o número total de incrementos da amostra e o número em cada que estrato de forma que o custo seja mínimo e a variância global seja igual a 0,02 $(\%)^2$ e a alocação da amostra seja proporcional em cada estrato,será dado por:

Estimação do Tamanho da Amostra

. Tipo de Alocação : Proporcional
. Estimativa Minimizando : Custo
. Método de Amostragem : Estratificada
. Amostragem dentro dos estratos : Aleatória Simples Sem Reposição

. Custo Unid. Amostral Constante : 1

. Variância Global escolhida : 0.02

Tamanho Total da Amostra : 73

Alocação

Estrato	Tamanho
1	17
2	28
3	28

Os resultados podem ser interpretados do seguinte modo: inicialmente para encontrar o valor de n igual a 73 o software ***"Sampling"*** usou o critério de minimização do custo considerando-se que a variância global era igual a 0,02 $\%^2$. A distribuição dos 73 incrementos entre os 3 estratos foi então feita de acordo com a alocação proporcional.

No anexo A pode-se encontrar um exemplo do procedimento operacional do software ***"Sampling"*** para se proceder ao cálculo de tamanhos de amostra.

7.0 Estimação de Proporções

A estimação de uma proporção populacional é considerada como um caso particular de estimação de uma média populacional. A variável resposta ***Y*** discutida assume apenas os valores 0 e 1. Portanto, teoricamente a proporção populacional é igual a média populacional, o mesmo ocorrendo para a média e proporção amostral.

Suponha que *P* seja a proporção populacional que se deseja estimar. Se o procedimento amostral utilizado em cada estrato for o de amostragem aleatória simples ou de amostragem sistemática então, a proporção amostral $\hat{p}_h$ é um estimador não-viciado da proporção populacional do h-ésimo estrato. Portanto um estimador não-viciado para a proporção populacional *P* será dada por:

$$\hat{p}_{st} = \sum_{h=1}^{L} W_h \ \hat{p}_h$$

sendo a respectiva variância estimada por:

$$\hat{V}ar(\ \hat{p}_{st}\ ) = \sum_{h=1}^{L} W_h^2 \quad (\ 1 - \frac{n_h}{N_h}\ ) \ \frac{\hat{p}_h\ (\ 1 - \hat{p}_h\ )}{n_h - 1}$$

Os intervalos de confiança para a proporção *P* e para o total de unidades amostrais populacionais que possuem a categoria em questão, isto é *NP*, são obtidos de modo similar aos discutidos na seção 3.0.

Para o cálculo do tamanho de amostras no caso de estimação de proporções deverá ser seguido os procedimentos apresentados na seção 6.0. Para facilitar o uso das fórmulas de cálculo de tamanho de amostra é importante lembrar que a variância populacional de cada estrato é dada por:

$$\sigma_h^2 = P_h\ (\ 1 - P_h\ )$$

e portanto um estimador para esta quantidade será:

$$\hat{\sigma}_h^2 = \hat{p}_h(\ 1 - \hat{p}_h\ )$$

sendo o valor de $\hat{p}_h$ obtido em alguma amostragem anterior ou por algum estudo similar. Caso não seja possível obter este valor a variância populacional de cada estrato h poderá ser aproximada por: $\hat{\sigma}_h^2 = \frac{1}{4}$. Veja o capítulo 2 seção 6.0.

Exemplo de Aplicação: Os supervisores de uma rede que conta com 4 lojas de departamento que vendem a crédito desejam estimar a proporção inadimplência da rede. Com esta finalidade foi feita uma amostragem estratificada na qual cada loja foi considerada como um estrato. Do arquivo de cada loja selecionou-se um certo número de contas (clientes) para a pesquisa observando-se então, o número de inadimplentes dentre os selecionados. Os dados obtidos estão na tabela 7.1 a seguir:

Tabela 7.1 : Resultados do exemplo da inadinplência da rede de lojas

	Estrato 1	Estrato 2	Estrato 3	Estrato 4
Número total de contas à credito (N_h)	65	42	93	25
Pêso do estrato (W_h)	0,2889	0,1867	0,4133	0,1111
Tamanho da amostra (n_h)	14	09	21	06
Número de contas inadimplentes na amostra	04	02	08	01
Proporção de contas inadimplentes na amostra	0,2857	0,2222	0,3809	0,1667

Para este exemplo a proporção de contas inadimplentes desta rede de lojas será estimada por:

$$\hat{p}_{st} = (0{,}2889)(0{,}2857) + (0{,}1867)(0{,}2222) + (0{,}4133)(0{,}3809) + (0{,}1111)(0{,}1667) = 0{,}30$$

com

$$\hat{V}ar(\hat{p}_{st}) = 0{,}0034396 \quad e \quad \sqrt{\hat{V}ar(\hat{p}_{st})} = 0{,}0586 .$$

8.0 Pós - Estratificação

A Pós-Estratificação aparece com grande frequência em situações nas quais a amostra inicial foi coletada sem levar-se em consideração “a priori” fatores importantes de estratificação. A não estratificação “a priori” pode ocorrer por falha humana quando da escolha do plano amostral para a coleta de dados, ou por desconhecimento ocasionado pela dificuldade de observação inicial de fatores relacionados com o problema, ou seja fatores de estratificação ocultos e que tornam-se visíveis após a análise estatística da amostra coletada sem estratificação. Deste modo, uma alternativa para tentar reparar o erro da não estratificação inicial é realizar a estratificação da população “a posteriori”. Assim os estratos populacionais são determinados após a coleta da amostra de n unidades amostrais. Neste caso, os tamanhos dos estratos, n_h, não são fixados “a priori” e são variáveis aleatórias. Como um exemplo de aplicação, em controle de qualidade uma situação na qual este procedimento de Pós-Estratificação poderia ser utilizado seria aquela em que uma amostra tenha sido coletada através de amostragem aleatória simples ou sistemática sem levar em consideração fatores de estratificação tais como, operadores, máquinas, linhas de produção, matérias primas, etc. Neste caso, estes fatores descobertos “a posteriori” poderiam ser introduzidos na análise estatística através de uma Pós-Estratificação. Entretanto, esta alternativa deve ser encarada com cuidado pois não há garantias de que a Pós-Estratificação venha a produzir estimativas mais precisas que aquelas obtidas sem a Pós-Estratificação. O melhor mesmo, é fazer um "brainstorm" antes da coleta de dados, de modo a planejar corretamente a seleção das informações levando-se em consideração os fatores de estratificação que podem estar relacionados com a(s) variável(is) resposta(s) de interesse.

No caso em que a amostra inicial de n elementos tiver sido coletada pelos métodos de Amostragem Aleatória Simples ou por Amostragem Sistemática as estimações de parâmetros considerando-se a Pós-Estratificação pode ser feita da seguinte forma:

- suponha que os pesos W_h dos estratos formados "a posteriori" sejam conhecidos, h = 1,2,..., L. Neste caso, um estimador da média populacional será dado por $\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$, como definido anteriomente. Entretanto, este estimador não tem a variância como discutido anteriormente devido a aleatoriedade dos valores n_h que somente são conhecidos após a seleção e observação da amostra inicial aleatória de n unidades amostrais. A variância estimada de $\bar{y}_{st}$ neste caso será dada por :

$$\hat{V}(\bar{y}_{st}) = \frac{N-n}{Nn} \sum_{h=1}^{L} W_h S_h^2 + \frac{1}{n^2} \sum_{h=1}^{L} (1 - W_h) S_h^2$$

Esta fórmula é obtida como sugerido em Scheaffer, Mendenhal e Ott (1996) e Cochran (1977).

A estimação considerando-se a Pós-Estratificação pode vir a resultar em estimativas menos precisas que aquelas obtidas através da Amostragem Aleatória Simples e Amostragem Sistemática quando os pesos dos estratos não forem conhecidos ou estiverem muito distantes dos valores populacionais.

CAPÍTULO 6: AMOSTRAGEM POR CONGLOMERADOS EM 1, 2, OU 3 ESTÁGIOS

1.0 Introdução

Neste capílulo vamos introduzir as noções gerais sobre o procedimento amostral conhecido como amostragem por conglomerados que pode ser realizado em 1 ou mais estágios. Primeiramente discutiremos o procedimento em 1 estágio e posteriormente extenderemos para dois e três estágios.

De um modo geral o procedimento de amostragem por conglomerados pressupõe que a população em estudo possa ser dividida em N unidades amostrais distintas e identificáveis, chamadas de conglomerados. Cada conglomerado é constituído por um certo número de elementos amostrais. Uma amostra aleatória constituída de n conglomerados é então selecionada da população em questão e todos os elementos de cada conglomerado selecionado para compor a amostra são observados. Um exemplo simples seria o de uma empresa que produz peças em lotes, cada um contendo 50 peças. Suponha que se tenha em um depósito um total de N=20 lotes. Então, uma amostra de 200 peças poderia ser obtida através da seleção aleatória de n=4 lotes dentre os 20 que estão no estoque. Cada um destes n=4 lotes seriam inspecionados em relação as variáveis de interesse. Para a seleção dos n=4 lotes poderíamos utilizar procedimentos como o de Amostragem Aleatória Simples ou Amostragem Sistemática, por exemplo.

Existem várias populações para as quais não se tem acesso a uma lista de todos os elementos amostrais o que inviabiliza muitas vezes o uso do elemento amostral como unidade amostral básica para efeitos de amostragem da população. Neste caso, o método de amostragem por conglomerados torna-se mais viável. Por exemplo, o caso em que a empresa deseja estimar o teor médio de minério de ferro de um carregamento de minério que vai ser embarcado. Pode-se considerar que o minério estocado nos vários vagões seja composto de um número finito N de incrementos (elementos amostrais), entretanto seria impossível listar estes incrementos ou identificá-los um a um. No entanto, poderíamos considerar que estes incrementos estariam agrupados nos vagões e portanto poderíamos considerar os vagões como as unidades amostrais básicas para se proceder a amostragem. Durante o processo de estocagem do minério nos vagões, amostras poderiam ser obtidas através de um amostrador automático que interviría no fluxo de minério em determinados momentos e coletaria os incrementos para análise laboratorial posterior. As amostras poderiam ser colhidas de todos os vagões do carregamento ou apenas de alguns deles selecionados aleatoriamente. Um outro exemplo muito comum é o de amostragem de indivíduos de uma determinada área para efeitos de alguma pesquisa de mercado. Se não dispomos de uma lista destes indivíduos, poderemos acessá-los através de um procedimento de amostragem no qual a área seja dividida em quarteirões (ou setores censitários), e estes por sua vez em residências. Deste modo, alguns quarteirõres seriam selecionados, dentro de cada quarteirão escolheria-se algumas residências e os indivíduos dentro de cada residência escolhida seriam entrevistados.

Teoricamente, não é necessário que todos os elementos amostrais associados a um conglomerado estejam fisicamente próximos. Por exemplo, num estudo de populações de fazendas, diferentes regiões da mesma fazenda podem constituir um conglomerado, mesmo que elas não estejam situadas próximas umas das outras.

É claro que o tamanho do conglomerado influencia na eficiência da amostragem por conglomerados. Para um dado tamanho de amostra fixo a ser selecionada da população em estudo, tem-se que pequenas unidades amostrais dão resultados mais precisos que grandes unidades amostrais, no que se refere a estimação das características populacionais de interesse. Assim, uma amostra aleatória de 600 residências dá informações de melhor qualidade sobre uma cidade do que uma amostra de 20 quarteirões contendo uma média de 30 residências cada. Entretanto, os custos associados à localização e entrevista das 600 residências são maiores do que aqueles relacionados com os 20 quarteirões e com a entrevista das 30 residências em cada quarteirão. Portanto, quando custo é confrontado com a precisão das estimativas, os esquemas com grandes unidades amostrais levam vantagem. Na prática o que se procura fazer é utilizar uma divisão da população em conglomerados que facilite a operacionalização da amostragem, que mantenha a pesquisa dentro do orçamento previsto e que dê estimativas confiáveis. O que se espera na verdade é que pequenas perdas de eficiência do esquema amostral de conglomerados seja compensada por uma redução nos custos de amostragem e na facilidade de sua implementação.

Se a preocupação na construção de conglomerados for apenas a de aumentar a precisão das estimativas, em relação aquelas que seriam obtidas usando-se a amostragem aleatória simples, então estes deveriam ser formados de modo que a variância da média dentro dos conglomerados fosse maior do que a variabilidade entre conglomerados. Entretanto, este método de escolha nem sempre é o mais prático, o mais conveniente, ou o minimiza custo.

Embora a Amostragem por Conglomerados ou em vários Estágios estabeleça que a população esteja dividida em N grupos disjuntos, a situação é diferente daquela discutida no capítulo 5 (Amostragem Estratificada).

2.0 Notação

Para facilitar a exposição dos resultados contidos neste capítulo, a seguinte notação será empregada:

N : número de conglomerados na população ;

n : número de conglomerados na amostra ;

M_i: número de unidades amostrais do i-ésimo conglomerado populacional , i = 1, 2, . . . , N (tamanho dos conglomerados);

$\bar{y}_i$: média amostral do i-ésimo conglomerado da amostra , i = 1, 2,…, n

y_i : o total amostral do i-ésimo conglomerado da amostra , i = 1,2, … , n

$\bar{\bar{y}}$: estimador da média populacional ;

$\hat{Y}$: estimador do total populacional ;

3.0 Amostragem por Conglomerados em um único Estágio Conglomerados de mesmo Tamanho

Neste procedimento pressupõe-se que a população possa ser dividida em N conglomerados de mesmo tamanho. Uma amostra aleatória de n conglomerados é então selecionada sendo que todos os conglomerados pertencentes a amostra são totalmente observados. Os estimadores pontuais e por intervalo para a média e o total populacional são apresentados a seguir.

Suponha que o procedimento de Amostragem Aleatória Simples sem reposição tenha sido utilizado para a escolha dos n conglomerados pertencentes a amostra. Considerando-se que neste caso $M_i = M$ para todo i = 1, 2, ..., N, tem-se que os estimadores não-viciados para a média e total populacional são dados respectivamente por :

$$\bar{\bar{y}} = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_i \qquad \text{(para a média populacional)}$$

$$\hat{Y} = N M \bar{\bar{y}} \qquad \text{(para o total populacional)}$$

A variância estimada de $\bar{\bar{y}}$ é dada por :

$$\hat{V}ar(\bar{\bar{y}}) = (1 - \frac{n}{N}) \frac{s_b^2}{n} \quad \text{onde} \quad s_b^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\bar{y}_i - \bar{\bar{y}})^2 \ .$$

e a variância estimada de $\hat{Y}$ é respectivamente,

$$\hat{V}ar(\hat{Y}) = (N M)^2 \ \hat{V}ar(\bar{\bar{y}})$$

De um modo geral os intervalos de confiança para a média e total populacionais são construídos através do uso da aproximação pela distribuição normal quando n M $\geq$ 30 e pela distribuição t- Student caso contrário. Portanto, para um nível de significância α, $0 < \alpha < 1$, os intervalos de $(1 - \alpha)100\%$ de confiança para a média populacional e total populacional seriam respectivamente:

$$\bar{\bar{y}} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{V}ar(\bar{\bar{y}})} \quad e \quad \hat{Y} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{V}ar(\hat{Y})} \qquad \text{se n M} \geq 30$$

$$\bar{\bar{y}} \pm t_{\frac{\alpha}{2}, nM-1} \sqrt{\hat{V}ar(\bar{\bar{y}})} \quad e \quad \hat{Y} \pm t_{\frac{\alpha}{2}, nM-1} \sqrt{\hat{V}ar(\hat{Y})} \qquad \text{se n M} < 30$$

Exemplo de Aplicação: O gerente de vendas de uma empresa jornalística está interessado em estimar o número médio de exemplares comprados por morador no último bimestre. Existem 4000 moradores na área e eles estão distribuídos em 40 setores censitários cada um contendo 100 moradores. Uma amostra de 4 setores é selecionada pelo procedimento de amostragem aleatória simples sem reposição e todos os seus moradores são entrevistados. A tabela 3.1 a seguir apresenta um resumo dos dados observados:

Tabela 3.1: Exemplo da empresa jornalística

Setores	No. médio de Exemplares comprados	Desvio Padrão
1	4,38	3,240
2	4,09	2,675
3	5,79	7,841
4	5,79	7,850

As estimativas pontuais e por intervalo, obtidas para o número médio e o total de exemplares adquiridos por cada morador da área de investigação, utilizando-se o software ***"Sampling"*** estão dados a seguir:

Estimação Global

. Tipo de Amostragem : Conglomerados
. Número de Estágios : Um
. Estimadores : Não Viciado
Unidades de Mesmo Tamanho

. Tamanho Médio Populacional dos Congl. : 100.00

N	Mean	StDev	I.C.
40	5.0125	0.4295412	(4.30597, 5.71903)

n	Total	StDev	I.C.
4	20050.0000	1718.1647185	(17223.9, 22876.1)

. Z Value : 1.64485 , $\alpha = 0{,}05$

Neste caso, estima-se que em média cada morador tenha adquirido 5,0125 exemplares (ou seja, aproximadamente 5 exemplares) no último bimestre. O verdadeiro valor do número médio de exemplares comprados por cada morador da área deverá estar dentro do intervalo (4,30597 , 5,71903) com 95% de confiança. O número total de exemplares adquiridos na área considerada é 20050,00 e um intervalo de 95% de confiança para o verdadeiro valor deste total populacional é (17223,9 , 22376,1).

4.0 Amostragem por Conglomerados em um único Estágio Conglomerados de Tamanhos Diferentes

No caso em que os conglomerados têm tamanhos M_i diferentes , i=1,2,..., N, são considerados dois casos: o primeiro no qual a amostra de conglomerados é obtida pelo método de Amostragem Aleatória Simples sem reposição (AAS-SR) e o segundo no qual a amostra é obtida pelo método P.P.S com reposição. A seguir descrevermos estes dois casos.

Caso 1. Os n conglomerados são selecionados através de AAS-SR

Para esta situação dispõe-se de dois métodos de estimação diferentes: o de estimação não-viciada e os estimação de razão. Pelo método da estimação não viciada os estimadores pontuais para a média e total populacional são dados respectivamente por:

$$\bar{\bar{y}} = \frac{1}{n\bar{\bar{M}}} \sum_{i=1}^{n} M_i \, \bar{y}_i \qquad \text{onde} \qquad \overline{M} = \frac{1}{N} \sum_{i=1}^{N} M_i$$

$$\hat{Y} = \left(\sum_{i=1}^{N} M_i \right) \bar{\bar{y}} = N\,\overline{M}\,\bar{\bar{y}}$$

As variâncias estimadas de $\bar{\bar{y}}$ e de $\hat{Y}$ são dadas respectivamente por:

$$\hat{V}ar(\bar{\bar{y}}) = \left(1 - \frac{n}{N}\right) \frac{s_b'^{2}}{n} \quad \text{onde} \quad s_b' = \frac{1}{(n-1)\,\overline{M}^2} \sum_{i=1}^{n} \left(M_i\,\bar{y}_i - \overline{M}\,\bar{\bar{y}} \right)^2$$

$$\hat{V}ar(\hat{Y}) = (N\overline{M})^2 \; \hat{V}ar(\bar{\bar{y}})$$

Pelo método de estimação por razão os estimadores da média e total populacional são dados por :

$$\bar{\bar{y}} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} M_i} = \frac{\sum_{i=1}^{n} M_i\,\bar{y}_i}{\sum_{i=1}^{n} M_i} = \frac{\sum_{i=1}^{n} M_i\,\bar{y}_i}{n\,\bar{m}} \qquad \text{(para a média populacional)}$$

$$\hat{Y} = \left(\sum_{i=1}^{N} M_i \right) \bar{\bar{y}} = N\,\overline{M}\,\bar{\bar{y}} \qquad \text{(para o total populacional)}$$

Os estimadores da variância de $\overline{\overline{y}}$ e de $\hat{Y}$ são dados respectivamente por :

$$\hat{V}ar(\overline{\overline{y}}) = \left(1 - \frac{n}{N}\right)\frac{s_b^{''\,2}}{n}$$

onde
$$s_b^{''\,2} = \frac{1}{(n-1)\overline{M}^2} \sum_{i=1}^{n} (y_i - M_i \overline{\overline{y}})^2$$

e
$$\hat{V}ar(\hat{Y}) = (N\overline{M})^2 \; \hat{V}ar(\overline{\overline{y}})$$

Em casos nos quais não se dispõe da informação do valor da constante $\overline{M}$, este valor pode ser estimado através da fórmula:

$$\overline{m} = \frac{1}{n} \sum_{i=1}^{n} M_i$$

ou seja, a média dos tamanhos de conglomerados da amostra. Esta estimativa entretanto, pode não ser muito boa.

Exemplo de Aplicação: Uma empresa que comercializa serras elétricas está interessada em estimar o custo médio de reparo (manutenção) por mês que é dispendido pela empresa para aquelas indústrias que adquirem as serras elétricas desta companhia. A empresa tem 48 clientes . Para fazer a estimação decidiu pelo seguinte procedimento: selecionou aleatoriamente 10 clientes dentre os 48 e para cada cliente foi observado o custo de manutenção dispendido pela companhia no último mês. A tabela a 4.1 seguir apresenta para cada indústria selecionada na amostra, o número de serras elétricas que a empresa dispõe cuja manutenção é feita por esta empresa, e o custo total de reparos (manutenção) no último mês.

Tabela 4.1 : Exemplo das serras elétricas

Indústria	Número de Serras	Custo Total De Reparo ($)
1	03	48,00
2	07	109,90
3	11	229,90
4	09	139,50
5	02	60,00
6	12	288,00
7	14	252,00
8	03	45,00
9	05	60,00
10	09	229,50
Total	75	1461,80

As estimativas obtidas utilizando-se o software ***"Sampling"*** estão dadas seguir:

Estimação Global

. Tipo de Amostragem : Conglomerados
. Número de Estágios : Um
. Estimadores : Não Viciado
Unidades de Tamanho Diferente

. Estimativa do Tamanho Médio dos Congl.: 7.50

N	Mean	StDev	I.C.
48	19.4907	3.5667861	(13.6238, 25.3575)

n	Total	StDev	I.C.
10	7016.6401	1284.0429688	(4904.58, 9128.7)

. Z Value : 1.64485 , $\alpha = 0{,}05$

Portanto, estima-se que o custo médio mensal de manutenção de cada serra comercializada por esta empresa seja pontulamente igual a 19,4907 dólares. Deste modo, se a empresa vender 5 serras para uma indústria ela esperará gastar aproximadamente 97,4535 dólares mensais na manutenção destas serras. O valor médio verdadeiro deverá estar no intervalo de (13,6238 , 25,3575) com 95% de confiança.

Caso 2. A amostra de n conglomerados é selecionada pelo método p.p.s com reposição.

Suponha que a probabilidade de seleção de cada conglomerado da população seja igual a:

$$p_i = \frac{M_i}{\sum_{i=1}^{N} M_i} = \frac{M_i}{N\,\overline{M}} \quad , \ i = 1, 2, \ldots, N.$$

Suponha ainda que a amostra seja selecionada com reposição. Então, os estimadores de média e total populacional são dados por :

$$\bar{\bar{y}} = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_i \qquad \text{(para a média populacional)}$$

$$\hat{Y} = N \overline{M}\, \bar{\bar{y}} \qquad \text{(para o total populacional)}$$

A variância estimada de $\bar{\bar{y}}$ e de $\hat{Y}$ são dadas respectivamente por :

$$\hat{V}ar(\bar{\bar{y}}) = \left(1 - \frac{n}{N}\right) \frac{s_b^2}{n} \quad \text{onde} \quad s_b^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\bar{y}_i - \bar{\bar{y}})^2$$

e

$$\hat{V}ar(\hat{Y}) = (N\overline{M})^2\, \hat{V}ar(\bar{\bar{y}}).$$

(A constante $\overline{M}$, pode ser estimada por $\overline{m}$ se necessário).

Exemplo de Aplicação: Um supervisor de uma companhia deseja estimar o número médio de dias de trabalho que os empregados perderam no último ano devido a motivos de saúde (doença). A companhia possue 8 departamentos que variam em termos do número de empregados. O supervisor deseja selecionar uma amostra de 3 departamentos com probabilidades de seleção proporcionais ao número de empregados do departamento. A tabela 4.2 a seguir apresenta o número de empregados por Departamento:

Tabela 4.2: Exemplo da empresa com vários departamentos

Departamento	No. de Empregados	No.Acumulado de Empregados
1	1200	1200
2	450	1650
3	2100	3750
4	860	4610
5	2840	7450
6	1910	9360
7	390	9750
8	3200	12950
Total	12950	12950

O supervisor selecionou três números aleatórios com reposição entre 1 e 12950 (números com 4 dígitos). Suponha que os números selecionados tenham sido: 02011, 07972 e 10281. Neste caso, a amostra aleatória seria composta dos departamentos 3, 6 e 8. A tabela 4.3 a seguir apresenta o número total de dias perdidos para os departamentos selecionados.

Tabela 4.3: Resultados amostrais observados
Empresa com vários departamentos

Depto.	Número de dias perdidos	Média ($\bar{y}_i$)
3	4320	2,057
6	4160	2,178
8	5790	1,809
Total	14240	1,975

Neste caso, a estimativa pontual para o número médio de dias perdidos devido a problemas de saúde seria:

$$\bar{y}_{pps} = \frac{1}{3}\left[\frac{4320}{2100} + \frac{4160}{1910} + \frac{5790}{3200}\right] = 2{,}02 \ \ dias$$

A variância estimada de $\bar{\bar{y}}_{pps}$ será dada por:

$$\hat{V}ar(\bar{y}_{pps}) = \frac{s_{bz}^2}{3} = 0{,}011809 \quad onde \quad s_{bz}^2 = \frac{1}{3-1}\sum_{i=1}^{3}(\bar{y}_i - \bar{y}_{pps})^2 = 0{,}03543$$

Portanto, um intervalo de 95% de confiança para o número médio de dias perdidos por empregados desta companhia devido a problemas de saúde seria:

$$2{,}02 \pm (1{,}96)\sqrt{0{,}011809} \Rightarrow (\ 1{,}8062 \ ; \ 2{,}2338\ )$$

Uma estimativa para o total de dias perdidos por problemas de saúde, dentre todos os empregados desta empresa seria:

$$\hat{Y}_{pps} = (12950)(2{,}02) = 26159 \ \ dias$$

com variância estimada igual a:

$$\hat{\sigma}^2_{\hat{Y}_{pps}} = (12950)^2 \ (0{,}011809) = 1980398{,}823$$

Um intervalo de confiança de 95% para o total de dias perdidos por todos os empregados da empresa seria:

$$26159 \pm (1{,}96)(1407{,}2664) \Rightarrow (23400{,}76 \ ; \ 28917{,}24)$$

5.0 Amostragem de Conglomerados em 2 Estágios
Unidades Primárias de Mesmo Tamanho

Neste procedimento a amostra é obtida em dois estágios. No primeiro estágio seleciona-se n conglomerados dentre os N da população. Num segundo estágio de cada conglomerado selecionado, escolhe-se aleatoriamente um certo número de elementos para observação.

Suponha que a amostra de conglomerados no primeiro estágio tenha sido selecionada por AAS-SR, e que para cada conglomerado selecionado sejam observadas *m* unidades secundárias sendo a seleção destas unidades secundárias feita por AAS-SR. Neste caso, os estimadores não viciados para a média e total populacional são dados por:

$$\bar{\bar{y}} = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_i \qquad \text{(para a média populacional)}$$

$$\hat{Y} = N \overline{M} \; \bar{\bar{y}} \qquad \text{(para o total populacional)}$$

A variância estimada de $\bar{\bar{y}}$ é dada por :

$$\hat{V}ar(\bar{\bar{y}}) = \left(1 - \frac{n}{N}\right) \frac{s_b^2}{n} + \frac{1}{N}\left(1 - \frac{m}{M}\right) \frac{\bar{s}_w^2}{m}$$

onde $s_b^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\bar{y}_i - \bar{\bar{y}})^2$; $\bar{s}_w^2 = \frac{1}{n} \sum_{i=1}^{n} s_i^2$

sendo $s_i^2 = \frac{1}{(m-1)} \sum_{i=1}^{m} (y_{ij} - \bar{y}_i)^2$ a variância amostral das unidades secundárias observadas do conglomerado i pertencente a amostra do primeiro estágio , i= 1, 2, . . . , n.

A variância estimada de $\hat{Y}$ é respectivamente,

$$\hat{V}ar(\hat{Y}) = (N \overline{M})^2 \; \hat{V}ar(\bar{\bar{y}})$$

(A constante $\overline{M}$, pode ser estimada por $\overline{m}$ se necessário).

Exemplo de Aplicação: Uma empresa que fabrica chapas de fibra de madeira está interessada em avaliar a qualidade das chapas produzidas em relação a um teste chamado de “teste de absorção da água pela chapa”. Este teste é requerido pelas normas técnicas deste setor de produção. Com esta finalidade os técnicos responsáveis optaram pelo seguinte esquema amostral: de um lote contendo 35 chapas de fibra de madeira de mesma espessura fabricados num determinado período do dia, selecionou-se 4 para realizar o experimento. Cada uma das chapas selecionadas foi recortada em 50 corpos de prova de mesma dimensão. Para cada chapa selecionou-se 10 corpos de prova que foram então utilizados na experimentação. O teste de absorção de água consiste em submergir os corpos de prova em uma banheira com água destilada circulante (mas não corrente) a determinadas temperaturas.Estes corpos devem ficar na água por um período de 24 horas. Depois disso, são retirados sendo então medido a quantia de absorção de água pela chapa. Antes de serem colocados na banheira os corpos de prova passam ficam por um determinado período de tempo numa câmara de refrigeração para um acondicionamento físico. Os dados relativos a este exemplo foram observados a um nível de temperatura igual a 20 C.

Tabela 5.1: Exemplo das chapas de fibra de madeira
Teste de absorção de água

Absorção		de		Água (%)	
Chapa 1		Chapa 2		Chapa 3	Chapa 4
22,9	22,9	20,5	20,6	23,4	22,6
				22,7	24,1
25,3	22,0	23,3	22,9	20,8	23,2
				22,3	24,3
23,2	24,2	20,8	21,7	21,9	22,5
				23,2	20,2
20,8	22,4	22,5	23,2	22,7	23,3
				22,3	20,4
20,3	22,5	23,0	22,7	22,1	23,0
				22,6	22,3

As estimativas da quantia média de absorção de água para chapas de fibra de madeira deste tipo, calculadas usando o software ***"Sampling"*** estão apresentadas a seguir:

Estimação Global

. Tipo de Amostragem : Conglomerados
. Número de Estágios : Dois
. Estimadores : Não Viciado
Unidades de Mesmo Tamanho

. Tamanho Médio Populacional dos Congl. : 50.00

N	Mean	StDev	I.C.
35	22.4400	0.1261221	(22.2325, 22.6475)

. Z Value : 1.64485 , $\alpha = 0{,}05$

Portanto, concluímos que em média cada chapa de fibra de madeira absorve 22,44 % de água. O intervalo de 95% para o valor real da absorção média de água pela chapa é (22,2325 , 22,6475).

6.0 *Amostragem por Conglomerados em 2 Estágios Unidades Primárias de Tamanhos Diferentes*

No caso em que os conglomerados têm tamanhos M_i diferentes, i=1,2,..., N são considerados dois casos :

Caso 1. AAS no Primeiro e Segundo Estágios

Os n conglomerados do primeiro estágio são selecionados através de AAS-SR. Para cada conglomerado i selecionado toma-se uma amostra constituída de m_i unidades secundárias escolhidas através de AAS-SR. Dois métodos de estimação diferentes podem ser utilizados nesta situação: o de estimadores não-viciados e os estimadores de razão. Pelo método dos estimadores não-viciados, os estimadores pontuais para a média e total populacional são dados respectivamente por :

$$\bar{\bar{y}} = \frac{1}{n\bar{M}} \sum_{i=1}^{n} M_i \, \bar{y}_i \quad \text{onde} \quad \bar{M} = \frac{1}{N} \sum_{i=1}^{N} M_i$$ (para a média populacional)

$$\hat{Y} = \left(\sum_{i=1}^{N} M_i \right) \bar{\bar{y}} = N \, \bar{M} \, \bar{\bar{y}}$$ (para o total populacional)

As variâncias estimadas de $\bar{\bar{y}}$ e de $\hat{Y}$ são dadas por:

$$\hat{V}ar(\bar{\bar{y}}) = \left(1 - \frac{n}{N}\right) \frac{s_b^{'2}}{n} + \frac{1}{nN(\bar{M}^2)} \sum_{i=1}^{n} M_i^2 \left(1 - \frac{m_i}{M_i}\right) \frac{s_i^2}{m_i}$$

onde

$$s_b^{'2} = \frac{1}{(n-1)\,\bar{M}^2} \sum_{i=1}^{n} \left(M_i \, \bar{y}_i - \bar{M} \, \bar{\bar{y}} \right)^2 \;;$$

$$s_i^2 = \frac{1}{(m-1)} \sum_{i=1}^{m} \left(y_{ij} - \bar{y}_i \right)^2$$

e

$$\hat{V}ar(\hat{Y}) = (N\bar{M})^2 \; \hat{V}ar(\bar{\bar{y}})$$

Pelo procedimento de estimação de razão a média e o total populacional são estimados respectivamente por :

$$\bar{\bar{y}} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i} = \frac{\sum_{i=1}^{n} m_i \bar{y}_i}{\sum_{i=1}^{n} m_i} = \frac{\sum_{i=1}^{n} m_i \bar{y}_i}{n \bar{m}}$$

$$\hat{Y} = (\sum_{i=1}^{N} M_i) \bar{\bar{y}} = N \bar{M} \bar{\bar{y}}$$

As variâncias estimadas de $\bar{\bar{y}}$ e de $\hat{Y}$ são dadas por :

$$\hat{V}ar(\bar{\bar{y}}) = (1 - \frac{n}{N}) \frac{1}{n \bar{M}^2 (n-1)} \sum_{i=1}^{n} (y_i - m_i \bar{\bar{y}})^2$$

$$\hat{V}ar(\hat{Y}) = (N \bar{M})^2 \hat{V}ar(\bar{\bar{y}})$$

(A constante $\bar{M}$, pode ser estimada por $\bar{m}$ se necessário).

Exemplo de Aplicação: Os técnicos de uma empresa de mineração desejam estimar o teor médio de ferro (%) de 3600 toneladas do Sinter Feed que será carregado em 60 vagões, todos os vagões de mesma capacidade. Para isto, selecionou-se 10 vagões através de uma Amostragem Aleatória Simples sem reposição e de cada um dos vagões selecionados retirou-se, aleatoriamente, uma amostra de incrementos no momento em que estes vagões estavam sendo carregados (corte automático no fluxo de minério). A tabela 6.1 ilustra para cada vagão selecionado o teor de ferro observado dos incrementos que foram retirados:

Tabela 6.1: Exemplo do carregamento de minério em vagões

Vagão	Número de incrementos selecionados	Teor de Ferro Observado (%)
01	3	64,04 64,05 64,06
02	6	64,09 64,08 64,05 64,10 64,07 64,02
03	5	64,10 64,09 64,07 64,06 64,05
04	3	64,06 64,05 64,06
05	4	64,08 64,07 64,08 64,06
06	4	64,05 64,06 64,07 64,05
07	4	64,06 64,05 64,08 64,07
08	5	64,08 64,09 64,10 64,09 64,05
09	3	64,07 64,06 64,07
10	5	64,08 64,10 64,08 64,09 64,10

As estimativas obtidas usando o software "***Sampling***" para o teor médio de ferro do minério carregado nos vagões utilizando o estimador de razão, estão apresentadas a seguir**:**

Estimação Global

. Tipo de Amostragem : Conglomerados
. Número de Estágios : Dois
. Estimadores : Razão

N	Mean	StDev	I.C.
60	64.0700	0.0000001	(64.0700, 64.0700)

. Z Value : 1.64485 , $\alpha = 0{,}05$

Caso 2. Procedimento de Amostragem p.p.s. no Primeiro Estágio e AAS-SR no Segundo Estágio.

Neste procedimento a amostra de n conglomerados é selecionada pelo método P.P.S. com reposição. Para cada conglomerado i na amostra seleciona-se m_i unidades amostrais através de AAS-SR.

Suponha que a probabilidade de seleção de cada conglomerado populacional seja igual a :

$$p_i = \frac{M_i}{\sum_{i=1}^{N} M_i} = \frac{M_i}{N\overline{M}} \quad , \; i = 1, 2, \ldots, N.$$

Então, os estimadores não-viciados de média e total populacional são dados respectivamente por :

$$\bar{\bar{y}} = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_i \qquad \text{(para a média populacional)}$$

$$\hat{Y} = N M \bar{\bar{y}} \qquad \text{(para o total populacional)}$$

As variâncias estimadas de $\bar{\bar{y}}$ e de $\hat{Y}$ são dadas respectivamente por:

$$\hat{V}ar(\bar{\bar{y}}) = (1 - \frac{n}{N}) \frac{s_b^2}{n} \qquad \text{onde} \qquad s_b^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\bar{y}_i - \bar{\bar{y}})^2$$

e

$$\hat{V}ar(\hat{Y}) = (N \overline{M})^2 \; \hat{V}ar(\bar{\bar{y}}).$$

Exemplo de Aplicação: No exemplo anterior suponha que os 60 vagões do carregamento não tivessem a mesma capacidade de armazenamento e que a capacidade variasse de 20 a 100 toneladas e que para estimar-se o teor de ferro (%) do minério que estava sendo carregado os 10 vagões da amostra tenham sido selecionados através do método P.P.S.. De cada um dos vagões selecionados retirou-se, aleatoriamente, uma amostra de incrementos. Considerando-se os mesmos dados apresentados na tabela 6.1 as estimativas pontual e por intervalo para o teor médio de minério de ferro do carregamento, considerando-se este novo esquema amostral, seria:

Estimação Global

. Tipo de Amostragem : Conglomerados
. Número de Estágios : Dois
. Estimadores : P.P.S.

N	Mean	StDev	I.C.
60	64.0683	0.0034816	(64.0625, 64.0740)

. Z Value : 1.64485 , $\alpha = 0{,}05$

7.0 Amostragem por Conglomerados em Três Estágios

No caso de Amostragem por Conglomerados em 3 estágios duas situações são consideradas: Amostragem Aleatória Simples e a Amostragem P.P.S.

Caso 1. AAS-SR em todos os Estágios de Seleção
Unidades Amostrais de mesmo Tamanho

Suponha que todas as unidades do primeiro estágio tenham o mesmo número de unidades secundárias, isto é, $M_i = M$, e as unidades secundárias tenham todas o mesmo número de unidades amostrais (unidades do terceiro estágio), isto é, $P_{ij} = P \ , \forall\ i, j$. Seja n o número de conglomerados selecionados por AAS-SR no primeiro estágio e seja *m* o número de unidades secundárias selecionadas por AAS-SR de cada conglomerado pertencente a amostra do primeiro estágio. Seja p o número de unidades amostrais selecionadas por AAS-SR de cada unidade secundária escolhida no segundo estágio. Neste caso, os estimadores não-viciados pontuais da média e do total populacional são respectivamente,

(a) para a média populacional

$$\bar{\bar{y}} = \frac{1}{nmp} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{p} y_{ijk} = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_i$$

(b) para o total populacional

$$\hat{Y} = N\ M\ P\ \bar{\bar{y}}$$

As variâncias estimadas de $\bar{\bar{y}}$ e de $\hat{Y}$ são respectivamente dadas por :

$$\hat{Var}(\bar{\bar{y}}) = (1 - \frac{n}{N}) \frac{s_b^2}{n} + \frac{1}{N}(1 - \frac{m}{M}) \frac{\bar{s}_w^2}{m} + \frac{1}{NM}(1 - \frac{p}{P}) \frac{\bar{\bar{s}}_p^2}{p}$$

onde :

$$s_b^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\bar{y}_{i..} - \bar{\bar{y}})^2 \quad ; \quad \bar{s}_w^2 = \frac{1}{n} \sum_{i=1}^{n} s_i^2$$

$$s_i^2 = \frac{1}{n-1} \sum_{j=1}^{m} (\bar{y}_{ij.} - \bar{y}_{i..})^2$$

$$\bar{\bar{s}}_p^2 = \frac{1}{nm} \sum_{j=1}^{m} s_{ij}^2 \quad ; \quad s_{ij}^2 = \frac{1}{p-1} \sum_{k=1}^{p} (y_{ijk} - \bar{y}_{ij.})^2 \quad ;$$

Exemplo de Aplicação: Um outro teste para verificar a qualidade das chapas de fibra de madeira produzidas pela empresa é o “teste de absorção da gota de óleo”. Neste teste, gotas de óleo são colocadas nos corpos de prova . Observa-se o tempo necessário para que a gota seja absorvida pela chapa. Numa experimentação feita pela empresa, esta selecionou 3 dentre 10 lotes que tinham sido fabricados no mesmo período de produção, para inspeção. Todas as chapas tinham a mesma espessura. Cada lote continha 100 chapas, medindo 1,20 x 1,50 cm de largura. De cada lote selecionado para inspeção decidiu-se selecionar 2 chapas de fibra, aleatoriamente. Cada chapa foi dividida em 24 corpos de prova (de 5cm x 40 cm). De cada chapa foi selecionado um total de 4 corpos de prova, que foram então submetidos ao teste da gota de óleo. A massa da gota foi fixada em 12 mg ($\pm$ 1mg). Os corpos de prova restantes de cada chapa foram utilizados para outras experiências. Neste exemplo, tem-se uma amostragem em 3 estágios. Suponha que os dados observados sejam aqueles apresentados no quadro 7.1 a seguir:

Quadro 7.1: Tempo de absorção da gota de óleo pela chapa de fibra de madeira

Amostra	Lote 1	Lote 2	Lote 3
Chapa 1	60,35 61,70 62,12 61,50	62,05 61,06 60,04 61,12	63,04 62,04 63,01 62,02
Média Chapa 1	61,42	61,07	62,52
Chapa 2	60,09 61,10 61,12 60,15	61,01 62,04 62,03 60,01	60,12 60,10 60,15 60,18
Média Chapa 2	60,62	61,27	60,14
Média do Lote	61,02	61,17	61,33

Neste caso as estimativas pontual e por intervalo para o tempo médio de absorção da gota de óleo pela chapa de fibra de madeira, obtidas pelo software “***Sampling***” seriam:

Estimação Global

. Tipo de Amostragem : Conglomerados
. Número de Estágios : Três
. Estimadores : Não Viciado
Unidades de Mesmo Tamanho

Tamanho Populacional

. Unidades do segundo estágio : 100.00

. Unidades do terceiro estágio : 24.00

Tamanho Amostral

. Unidades do segundo estágio : 2.00

. Unidades do terceiro estágio : 4.00

N	Mean	StDev	I.C.
10	61.1729	0.2412176	(60.7595, 61.5863)

. T Value : 1.71388 , $\alpha = 0,05$

Portanto, podemos concluir que em média o tempo para absorção da gota de óleo pela chapa de fibra é de 61,1729 segundos. Um intervalo de 95% de confiança para o valor real deste tempo médio é dado por: (60,7595 , 61,5863).

Caso 2. Amostragem p.p.s com Reposição no Primeiro Estágio e AAS-SR nos outros dois Estágios Unidades Amostrais Populacionais de Tamanhos Diferentes.

Suponha que a população em estudo seja tal que os N conglomerados populacionais tenham tamanhos diferentes, ou seja M_i , i = 1, 2, . . . , N. Para cada uma das M_i unidades amostrais secundárias seja P_{ij} o número de unidades amostrais populacionais do terceiro estágio. Suponha que os n conglomerados do primeiro estágio sejam selecionados através do método p.p.s. com reposição, enquanto que as unidades amostrais do segundo e terceiro estágio sejam selecionadas por AAS-SR. Seja

$$p_i = \frac{\sum_{j=1}^{M_i} P_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{M_i} P_{ij}}$$

a probabilidade de seleção de cada conglomerado populacional no primeiro estágio. Sejam m_i e p_{ij}, repectivamente, o número de unidades secundárias selecionadas de cada unidade primária no segundo estágio, e o número de unidades amostrais selecionadas de cada unidade secundária no terceiro estágio. Então, neste caso os estimadores não-viciados para a média e o total populacional são dados respectivamente :

$$\bar{\bar{y}} = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_i \qquad \text{(para a média populacional)}$$

$$\hat{Y} = \left(\sum_{i=1}^{N} \sum_{j=1}^{M_i} P_{ij} \right) \bar{\bar{y}} \qquad \text{(para o total populacional)}$$

As variâncias estimadas de $\bar{\bar{y}}$ e de $\hat{Y}$ são dadas respectivamente por :

$$\hat{V}ar(\bar{\bar{y}}) = \left(1 - \frac{n}{N} \right) \frac{s_b^2}{n} \quad \text{onde} \quad s_b^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(\bar{y}_{i..} - \bar{\bar{y}} \right)^2 .$$

e

$$\hat{V}ar(\hat{Y}) = \left(\sum_{i=1}^{N} \sum_{j=1}^{M_i} P_{ij} \right)^2 \hat{V}ar(\bar{\bar{y}}).$$

Quando o valor da constante $\left(\sum_{i=1}^{N} \sum_{j=1}^{M_i} P_{ij} \right)$ não é conhecida, ela poderá ser estimada pela quantidade $(N \, \bar{m} \, \bar{p})$, onde:

$$\bar{m} = \frac{1}{n} \sum_{i=1}^{n} m_i \quad e \quad \bar{p} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} p_{ij} .$$

Se o procedimento amostral utilizado nos segundos e terceiros estágios forem do tipo AAS com reposição a variância estimada de $\bar{\bar{y}}$ torna-se :

$$\hat{V}ar(\bar{\bar{y}}) = \frac{s_b^2}{n}$$

Exemplo de Aplicação: Os técnicos de uma companhia de seguros de saúde estão interessados em estimar o número total de operações cesarianas realizadas no primeiro bimestre do ano em uma certa cidade. A cidade está subdividida em 4 regiões contendo no total 20 maternidades. Como as regiões apresentam uma grande variação de tamanho, os técnicos decidiram amostrá-las com probabilidade proporcional ao número de habitantes. Em cada região selecionada foi retirada uma amostra de maternidades e de cada uma destas uma amostra de alas foi observada. Para cada ala selecionada observou-se o número de cesarianas realizadas no primeiro bimestre do ano. Sabe-se que existem no total 400 alas, considerando-se as 20 maternidades. A tabela a seguir apresenta os resultados amostrais observados:

Tabela 7.1: Exemplo da empresa de seguros de saúde

Região	*Maternidade*	*Número de alas selecionadas*	*Número de cesarianas*
	2	8	21, 37, 38, 47, 33, 30, 43, 30
1	3	4	42, 34, 42, 37
	5	6	35, 34, 41, 38, 47, 34
3	1	9	40, 49, 44, 52, 42, 50, 46, 50, 48
	2	5	52, 44, 46, 47, 45

As estimativas pontual e por intervalo do número médio e do número total de cesarianas da cidade estão dadas a seguir:

Estimação Global

. Tipo de Amostragem : Conglomerados
. Número de Estágios : Três
. Estimadores : P.P.S.

. Número de elementos da população : 400.00

N	Mean	StDev	I.C.
4	41.8095	3.5186980	(36.0218, 47.5973)

n	Total	StDev	I.C.
2	16723.8095	1407.4792121	(14408.7, 19038.9)

. Z Value : 1.64485 , $\alpha = 0,05$

Deste modo, a empresa de seguros de saúde estima que nesta cidade sejam realizadas em média 41,8095 cesarianas por bimestre. Um intervalo de 95% de confiança para o verdadeiro valor desta média é (36,0218 , 47,5973). A empresa estima que sejam realizadas um total de 16723,8095 cesarianas por bimestre., sendo um intervalo de 95% de confiança para o verdadeiro total igual a (14408,7 , 19038,9).

NOTA: AMOSTRAGEM POR CONGLOMERADOS E AMOSTRAGEM ALEATÓRIA SIMPLES COM REPOSIÇÃO:

Os casos em que as unidades amostrais são selecionadas por Amostragem Aleatória Simples com reposição em um ou mais dos estágios dos procedimentos amostrais tratados neste capítulo, as fórmulas de variância simplificam. Nestes casos, bastará eliminar das f'órmulas apresentadas os fatores de correções finitas correspondentes aos estágios em que a Amostragem Aleatória Simples com reposição foi utilizada.

CAPÍTULO 7: SUBAMOSTRAGEM ESTRATIFICADA

1.0 Introdução

A Subamostragem estratificada (Sukhatme e Sukhatme,1970) é um método de amostragem muito utilizado nas situções práticas, principalmente em pesquisas que envolvem populações humanas. Como um simples exemplo, suponhamos que uma pesquisa envolva os habitantes da cidade de Belo Horizonte. Claramente, sabemos que o comportamento das pessoas está relacionado com a situação econômica e social, nível educacional e faixa etária, entre outros fatores. Deste modo, um esquema amostral adequado sIria aquele que procurasse ter na amostra elementos que representassem todas as camadas sociais e as várias faixas etárias. Como em geral, há uma certa relação entre o local onde o indivíduo reside e sua classe social e nível econômico, uma sugestão seria divididir a cidade de Belo Horizonte em 5 regiões por exemplo, segundo estes critérios, e então selecionar de cada região uma amostra de setores censitários. De cada setor selecionado seria escolhido aleatoriamente algumas residências para entrevista. A seleção das residências poderia ser feita em um único estágio ou em dois estágios escolhendo-se primeiramente os quarteirões e posteriormente as residências em cada quarteirão escolhido. Portanto, neste esquema teríamos 5 estratos (5 regiões) e em cada estrato a amostragem de residências seria feita por uma Amostragem por Conglomerados em 2 ou 3 estágios.

Um exemplo na área industrial está relacionado com a indústria de mineração. Supondo-se que a indústria tenha várias minas produtoras de minério de ferro, digamos A, B, e C, por exemplo, e que o minério produzido em cada uma delas seja embarcado para um determinado porto, no qual primeiramente forma-se uma pilha de homogeneização sendo este minério posteriormente transportado pelos navios para países distintos, então para se avaliar o teor médio de ferro (%) do minério que estaria sendo embarcado em um navio, o seguinte esquema poderia ser utilizado: para cada mina seria selecionado aleatoriamente um certo número n de vagões para observação. De cada vagão selecionado, seria coletado uma amostra aleatória de m incrementos do minério para ser avaliado no laboratório. Para cada mina terse-ia uma estimativa do teor médio de ferro do minério proveniente daquela mina específica. Estas estimativas seriam então combinadas de modo a fornecer uma estimativa do teor médio de ferro global da pilha de minério. Neste caso, cada carregamento de cada mina, constituiria um estrato e em cada estrato a amostragem do minério seria feita por um esquema de Amostragem por Conglomerados em 2 estágios.

A forma de combinar as estimativas dos parâmetros populacionais obtidas nos vários estratos é similar àquela apresentada no capítulo 5 onde a Amostragem Estratificada foi discutida. Na seção a seguir mostraremos os estimadores mais usuais para estas situações.

2.0 Estimação de Parâmetros Populacionais

Os estimadores utilizados são aqueles usuais de Amostragem Estratificada. Supondo-se que a população esteja dividida em L estratos, e que as amostras sejam selecionadas de cada estrato de forma independente, os estimadores não-viciados da média e do total populacional são dados por :

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \; \bar{\bar{y}}_h$$

onde W_h representa o pêso do h-ésimo estrato da população e $\bar{\bar{y}}_h$ representa o estimador da média populacional do h-ésimo estrato. Este estimador pode ser selecionado de acordo com aqueles disponíveis para os casos de Amostragem por Conglomerados em 1,2 ou 3 estágios descritos anteriormente no capítulo 6.

A variância estimada de $\bar{y}_{st}$ é dada por :

$$\hat{V}ar(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \; \hat{V}ar(\bar{\bar{y}}_h)$$

sendo que a $\hat{V}ar(\bar{\bar{y}}_h)$ é calculada de acordo com o esquema amostral utilizado para a coleta de amostra em cada estrato. Os valores de W_h são dados respectivamente por :

- no caso de Amostragem por Conglomerados em 2 estágios:

$$W_h = \frac{\sum_{i=1}^{N_h} M_i}{\sum_{h=1}^{L} \sum_{i=1}^{N_h} M_i}$$

- no caso de Amostragem por Conglomerados em 3 estágios:

$$W_h = \frac{\sum_{i=1}^{N_h} \sum_{j=1}^{M_i} P_{ij}}{\sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{j=1}^{M_i} P_{ij}}$$

onde a constante N_h representa o número total de conglomerados da população do h-ésimo estrato.

O estimador do total populacional é dado por :

$$\hat{Y}_{st} = \left(\sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{j=1}^{M_i} P_{ij} \right) \overline{y}_{st}$$

com variância estimada igual a:

$$\hat{V}ar(\overline{y}_{st}) = \left(\sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{j=1}^{M_i} P_{ij} \right)^2 \hat{V}ar(\overline{y}_{st})$$

Exemplo de Aplicação: Uma pesquisa sócio-econômica está sendo realizada com o objetivo de estimar o gasto médio com ligações telefônicas, por domicílio, em uma determinada região. Como não há um cadastro centralizado contendo informações referentes a todos os domicílios, o estatístico responsável pelo estudo decidiu adotar o seguinte procedimento

- a região foi dividida em 2 áreas economicamente distintas, A e B;
- selecionou-se, através do procedimento de amostragem com probabilidade proporcional ao tamanho, um certo número de setores censitários em cada área;
- em cada setor sorteou-se, de forma aleatória, um certo número de quarteirões;
- em cada quarteirão sorteou-se, de forma aleatória, um certo número de domicílios para receberem a visita do pesquisador.

Utilizando o software ***"Sampling"*** teríamos as seguintes estimativas para a média e o total populacionais:

Estimação Global

. Tipo de Amostragem	: Estratificada
. Tipo de Amostragem dentro dos Estratos	: Conglomerados
. Número de Estágios	: Três
. Estimadores	: P.P.S.
. Número de Estratos Analisados	: 2

N	Mean	StDev	I.C.
3000	294.8705	5.2180080	(286.288, 303.453)

n	Total	StDev	I.C.
838	884611.4353	15654.0239955	(858863, 910360)

. Z Value : 1.64485 , $\alpha = 0{,}05$

Portanto, poderíamos dizer que em média cada domicílio desta região dispende 294,8705 reais mensais, sendo um intervalo de 95% de confiança igual a (286,288 , 303,453). O total dispendido em ligações telefônicas nesta região é 884611,4353 reais mensais sendo um intervalo de 95% de confiança igual a (858863 , 910360).

ANEXO A: EXEMPLOS DE OPERACIONALIZAÇÃO DO SOFTWARE SAMPLING

Neste anexo apresentamos alguns exemplos de operacionalização do software ***"Sampling"*** .

Estando no ambiente do software ***"Minitab for Windows"*** o usuário deverá chamar o software ***"Sampling"*** e proceder a leitura do arquivo de dados através dos comandos correspondentes (ver manual do usuário). A partir de então, a seguinte tela irá aparecer no monitor:

Aguarde um momento por favor ... programa em processamento

SAMPLING

Produzido por : Sueli Aparecida Mingoti
Gregório Saravia Atuncar
Maria Lourdes G. Nogueira
Rodrigo Carazolli Da Silva

DEPARTAMENTO DE ESTATÍSTICA UFMG

Programa desenvolvido com o apoio financeiro da **FAPEMIG**

1998

Para continuar digite "yes" ou digite "no" para parar o processamento :

Digite Yes

Departamento de Estatística UFMG

A seguir várias perguntas aparecerão sequencialmente e os resultados das análises aparecerão tão logo o processamento é feito, como segue:

Exemplo 1: Considere o exemplo 1 tratado no capítulo 2, seção 3.1 (indústria eletrônica – amostragem aleatória simples)

Deseja estimar o tamanho da amostra ?

Digite No

Deseja estatística descritiva por estrato ?

Digite No

Deseja estatística descritiva global ?

Digite No

Deseja análise gráfica por estrato ?

Digite No

Deseja análise gráfica Global ?

Digite No

Digite "Yes" para continuar ou "No" para parar a análise:

Digite Yes

Digite o método de amostragem utilizado dentro dos estratos :

1) Amostragem Aleatória Simples com Reposição
2) Amostragem Aleatória Simples sem Reposição
3) Amostragem Sistemática

DATA> **Digite a opção 2**

Escolha os tipos de estimadores :

1) Usuais
2) Razão
3) Regressão

DATA> **Digite a opção 1**

Escolha o nível de significância desejado :

1) 0.010
2) 0.025
3) 0.050
4) 0.100

DATA> **Digite a opção 3**

Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Aleatória Simples Sem Reposição
. Estimadores : Usuais

N	Mean	StDev	I.C.
1000	15.8143	0.1315947	(15.5978, 16.0307)

Total	StDev	I.C.
15814.2861	131.5947418	(15597.8, 16030.7)

. Z Value : 1.64485

Exemplo 2: Considere o exemplo 1 tratado no capítulo 3 seção 3 (indústria de mineração – amostragem sistemática)

Deseja estimar o tamanho da amostra ?

Digite No

Deseja estatística descritiva por estrato ?

Digite No

Deseja estatística descritiva global ?

Digite No

Deseja análise gráfica por estrato ?

Digite No

Deseja análise gráfica Global ?

Digite No

Digite "Yes" para continuar ou "No" para parar a análise:

Digite Yes

Digite o método de amostragem utilizado dentro dos estratos :

1) Amostragem Aleatória Simples com Reposição
2) Amostragem Aleatória Simples sem Reposição
3) Amostragem Sistemática

DATA> **Digite 3**

Escolha o nível de significância desejado :

1) 0.010
2) 0.025
3) 0.050
4) 0.100

DATA> **Digite 3**

Deseja correlação entre elementos da mesma amostra ?

Digite Yes

Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Amostragem Sistemática
. Método 1 : Amostragem Aleatória Simples
. Método 2 : Amostragem Sistemática Repetida
. Método 3 : Diferença

N	Mean
4000000	61.2882

StDev Mean

Method 1	Method 2	Method 3
0.0586301	0.0118160	0.0537267

IC Mean

Method 1
(61.1873, 61.3891))

Method 2
(61.2679, 61.3085)

Method 3
(61.1957, 61.3806)

Estrato	Correlation
1	-0.0000005

. T Value : 1.72075

Exemplo 3: Considere o exemplo 2 tratado no capítulo 5 na seção 5.0 (agência de propaganda – amostragem estratificada)

Deseja estimar o tamanho da amostra?

Digite No

Deseja estatística descritiva por estrato?

Digite No

Deseja estatística descritiva global?

Digite No

Deseja análise gráfica por estrato?

Digite No

Deseja análise gráfica Global?

Digite No

Digite "Yes" para continuar ou "No" para parar a análise:

Digite Yes

Digite o método de amostragem utilizado dentro dos estratos:

1) Amostragem Aleatória Simples com Reposição
2) Amostragem Aleatória Simples sem Reposição
3) Amostragem Sistemática

DATA> **Digite 2**

Escolha os tipos de estimadores:

1) Usuais
2) Razão
3) Regressão

DATA> **Digite 1**

Existe Pós-Estratificação?

Digite No

Escolha o nível de significância desejado:

1) 0.010
2) 0.025
3) 0.050
4) 0.100

DATA> **Digite 3**

Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Estratificada
. Tipo de Amostragem dentro do estratos : Aleatória Simples Sem Reposição
. Estimadores : Usuais

N	Mean	StDev	I.C.
310	27.6750	1.4033957	(25.2601, 30.0899)

Total	StDev	I.C.
8579.2500	435.0526428	(7830.63, 9327.87)

. T Value : 1.72075

Exemplo 4: Considere o exemplo 1 tratado no capítulo 5 seção 6.1 (agência de propaganda – alocação ótima)

Caso (a) Minimização de Custo para uma Variância Fixa igual a $1h^2$

Deseja estimar o tamanho da amostra?

Digite YES

Digite o número de estratos para análise:

Digite 3

Qual o tipo de Alocação:

1) Alocação Ótima
2) Alocação Proporcional

Digite a opção 1

Calcular a estimativa minimizando:

1) Custo
2) Variância Global

Digite a opção 1

Deseja atribuir valor ao custo da unidade amostral?

Digite YES

Digite o custo por unidade amostral:

Digite 9 Pressione a tecla Enter
9 Pressione a tecla Enter
16 Pressione a tecla Enter

Digite a variância global:

Digite 1

Digite os desvios padrão amostrais:

Digite 5 Pressione a tecla Enter
15 Pressione a tecla Enter
10 Pressione a tecla Enter

Deseja calcular o valor do tamanho total da amostra?

Digite YES

Impressão dos Resultados

Estimação do Tamanho da Amostra

. Tipo de Alocação : Ótima
. Estimativa Minimizando : Custo
. Método de Amostragem : Estratificada
. Amostragem dentro dos estratos : Aleatória Simples Sem Reposição

. Variância Global escolhida : 1

Tamanho Total da Amostra : 57

Alocação

Estrato	Tamanho
1	18
2	22
3	17

Caso (b)
A variância global seja mínima e o custo total final igual a $1000.

Deseja estimar o tamanho da amostra?

Digite YES

Digite o número de estratos para análise:

Digite 3

Qual o tipo de Alocação:

1) Alocação Ótima
2) Alocação Proporcional

Digite a opção 1

Calcular a estimativa minimizando:

1) Custo
2) Variância Global

Digite a opção 2

Digite o custo total final:

Digite 1000

Deseja atribuir valor ao custo da unidade amostral?

Digite YES

Digite o custo por unidade amostral:

Digite 9 Pressione a tecla Enter
9 Pressione a tecla Enter
16 Pressione a tecla Enter

Digite os desvios padrão amostrais:

Digite 5 Pressione a tecla Enter
15 Pressione a tecla Enter
10 Pressione a tecla Enter

Deseja calcular o valor do tamanho total da amostra?

Digite YES

Estimação do Tamanho da Amostra

. Tipo de Alocação : Ótima
. Estimativa Minimizando : Variância Global
. Método de Amostragem : Estratificada
. Amostragem dentro dos estratos : Aleatória Simples Sem Reposição

. Valor de custo total escolhido : 1000

Tamanho Total da Amostra : 90

Alocação

Estrato	Tamanho
1	29
2	35
3	26

Exemplo 5 : Considere o exemplo 2 tratado no capítulo 5 seção 6.0 (indústria de mineração- alocação proporcional)

Requisito: **o custo seja mínimo e a variância global igual a 0.02%^{2}.**

Deseja estimar o tamanho da amostra?

Digite YES

Digite o número de estratos para análise:

Digite 3

Qual o tipo de Alocação:

1) Alocação Ótima
2) Alocação Proporcional

Digite a opção 2

Calcular a estimativa minimizando:

1) Custo
2) Variância Global

Digite a opção 1

Deseja atribuir valor ao custo da unidade amostral?

Digite NO

Digite a variância global:

Digite 0.02

Digite os desvios padrão amostrais:

Digite 1.4 Pressione a tecla Enter
1.2 Pressione a tecla Enter
1.1 Pressione a tecla Enter

Deseja calcular o valor do tamanho total da amostra?

Digite YES

Estimação do Tamanho da Amostra

. Tipo de Alocação : Proporcional
. Estimativa Minimizando : Custo
. Método de Amostragem : Estratificada
. Amostragem dentro dos estratos : Aleatória Simples Sem Reposição

. Custo Unid. Amostral Constante : 1

. Variância Global escolhida : 0.02

Tamanho Total da Amostra : 73

Alocação

Estrato	Tamanho
1	17
2	28
3	28

Exemplo 6 : Considere o exemplo tratado no capítulo 6 seção 3 (exemplo da empresa jornalística – Análise de Conglomerados em 1 Estágio)

Digite o número de estágios:

1) Um estágio
2) Dois estágios
3) Três estágios

Digite 1

Escolha o tipo de estimador:

1) Estimador P.P.S.
2) Estimador de Razão
3) Estimador Não Viciado -
   Unidades de mesmo tamanho em cada estágio
4) Estimador Não Viciado -
   Unidades de tamanho diferente em cada estágio

Digite 2

Deseja entrar com o tamanho médio populacional dos conglomerados?

Digite YES   Pressione a tecla Enter
125   Pressione a tecla Enter

Escolha o nível de significância desejado:

1) 0.010
2) 0.025
3) 0.050
4) 0.100

Digite uma das opções: 1, 2, 3 ou 4

Estimação Global

. Tipo de Amostragem : Conglomerados
. Número de Estágios : Um
. Estimadores : Razão

. Tamanho Médio Populacional dos Congl. : 125.00

N	Mean	StDev	I.C.
40	5.1514	0.5520147	(4.24345, 6.05941)

n	Total	StDev	I.C.
4	25757.1429	2760.0735749	(21217.2, 30297.1)

. Z Value : 1.64485

Exemplo 7: Considere o exemplo de aplicação tratado bo capítulo 6 seção 5 (chapas de fibra de madeira – teste de absorção de água – Análise de Conglomerados 2 Estágios)

Digite o número de estágios:

1) Um estágio
2) Dois estágios
3) Três estágios

DATA> **Digite 2**

Escolha os tipos de estimadores :

1) Estimador P.P.S.
2) Estimador de Razão
3) Estimador Não Viciado -
   Unidades de mesmo tamanho em cada estágio
4) Estimador Não Viciado -
   Unidades de tamanho diferente em cada estágio

DATA> **Digite 3**

Escolha o nível de significância desejado :

1) 0.010
2) 0.025
3) 0.050
4) 0.100

DATA> **Digite 3**

Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Conglomerados
. Número de Estágios : Dois
. Estimadores : Não Viciado
Unidades de Mesmo Tamanho

. Tamanho Médio Populacional dos Congl. : 50.00

N	Mean	StDev	I.C.
35	22.4400	0.1261221	(22.2325, 22.6475)

n	Total	StDev	I.C.
4	39270.0000	220.7136536	(38907.0, 39633.0)

. Z Value : 1.64485

---- ....

Exemplo 8: Considere o exemplo tratado no capítulo 6 seção 6 (indústria de mineração - Amostragem por Conglomerados em 2 Estágios – Tipo P.P.S.)

Digite o número de estágios:

1) Um estágio
2) Dois estágios
3) Três estágios

Digite 2

Escolha o tipo de estimador:

1) Estimador P.P.S.
2) Estimador de Razão
3) Estimador Não Viciado -
   Unidades de mesmo tamanho em cada estágio
4) Estimador Não Viciado -
   Unidades de tamanho diferente em cada estágio

Digite 1

Deseja entrar com o tamanho médio populacional dos conglomerados?

Digite YES Pressione a tecla Enter
100000 Pressione a tecla Enter

Escolha o nível de significância desejado:

1) 0.010
2) 0.025
3) 0.050
4) 0.100

Digite uma das opções: 1, 2, 3 ou 4

Estimação Global

. Tipo de Amostragem : Conglomerados
. Número de Estágios : Dois
. Estimadores : P.P.S.

N	Mean	StDev	I.C.
60	64.0683	0.0034816	(64.0625, 64.0740)

n	Total	StDev	I.C.
10	384409600.0000	20889.5685469	(3.84E+08, 3.84E+08)

. Z Value : 1.64485

Exemplo 9 : Considere o exemplo tratado no capítulo 6 seção 7 (chapas de fibra de madeira – teste de absorção de óleo- Amostragem por Conglomerados em 3 Estágios)

Digite o número de estágios:

1) Um estágio
2) Dois estágios
3) Três estágios

DATA> **Digite 3**

Escolha os tipos de estimadores :

1) Estimador P.P.S.
2) Estimador Não Viciado -
   Unidades de mesmo tamanho em cada estágio

DATA> **Digite 2**

Escolha o nível de significância desejado :

1) 0.010
2) 0.025
3) 0.050
4) 0.100

DATA> **Digite 3**

Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Conglomerados
. Número de Estágios : Três
. Estimadores : Não Viciado
Unidades de Mesmo Tamanho

Tamanho Populacional

. Unidades do segundo estágio : 100.00

. Unidades do terceiro estágio : 24.00

Tamanho Amostral

. Unidades do segundo estágio : 2.00

. Unidades do terceiro estágio : 4.00

N	Mean	StDev	I.C.
10	61.1729	0.2412176	(60.7595, 61.5863)

n	Total	StDev	I.C.
3	1468150.0000	5789.2231445	(1458228, 1478072)

. T Value : 1.71388

Exemplo 10: Considere o exemplo disutido na seção 7.0 capítulo 6 (exemplo das maternidades – Amostragem por Conglomerados e 3 Estágios – Tipo P.P.S.)

Digite o número de estágios:

1) Um estágio
2) Dois estágios
3) Três estágios

Digite 3

Escolha o tipo de estimador:

1) Estimador P.P.S.
2) Estimador Não Viciado -
   Unidades de mesmo tamanho em cada estágio

Digite 1

Deseja entrar com o número de elementos da população?

Digite YES Pressione a tecla Enter
400 Pressione a tecla Enter

Escolha o nível de significância desejado:

1) 0.010
2) 0.025
3) 0.050
4) 0.100

Digite uma das opções: 1, 2, 3 ou 4

Estimação Global

. Tipo de Amostragem : Conglomerados
. Número de Estágios : Três
. Estimadores : P.P.S.

. Número de elementos da população : 400.00

N	Mean	StDev	I.C.
4	41.8095	3.5186980	(36.0218, 47.5973)

n	Total	StDev	I.C.
2	16723.8095	1407.4792121	(14408.7, 19038.9)

. Z Value : 1.64485

Outros Exemplos

Exemplo 11. O seguinte exemplo refere-se a notas de estudantes de segundo grau de uma cidade. O estrato 1 corresponde a curso supletivo e os estratos 2, 3 e 4 correspondem aos turnos de manhã, tarde e noite respectivamente. Estimaremos a média e construiremos intervalos de confiança de 95%.

Digite o método de amostragem utilizado dentro dos estratos:

1) Amostragem Aleatória Simples com Reposição
2) Amostragem Aleatória Simples sem Reposição
3) Amostragem Sistemática

DATA> 2

Escolha os tipos de estimadores:

1) Usuais
2) Razão
3) Regressão

DATA> 1

Existe Pós-Estratificação?

n

Escolha o nível de significância desejado:

1) 0.010
2) 0.025
3) 0.050
4) 0.100

DATA> 3

Estimação Global

. Tipo de Amostragem : Estratificada
. Tipo de Amostragem dentro do estratos : Aleatória Simples Sem Reposição
. Estimadores : Usuais

N Mean StDev I.C. 8950
72.5644 0.2516977 (72.1504, 72.9784)

Total StDev I.C.

649451.0881 2252.6948336 (645746, 653156)

. Z Value : 1.64485

Exemplo 12. *No exemplo seguinte, determinaremos os tamanhos amostrais em cada estrato minimizando custo.*

Deseja estimar o tamanho da amostra?

y

Digite o número de estratos para análise:

DATA> 4

Qual o tipo de Alocação:

1) Alocação Ótima
2) Alocação Proporcional

DATA> 1

Calcular a estimativa minimizando:

1) Custo
2) Variância Global

DATA> 1

Deseja atribuir valor ao custo da unidade amostral?

y

Digite o custo por unidade amostral:

DATA> 1
DATA> 4
DATA> 3
DATA> 2

Digite a variância global:

DATA> 40

Digite o(s) desvio(s) padrão amostral(is):

Digite o(s) desvio(s) padrão amostral(is):

DATA> 3.2
DATA> 4.1
DATA> 4.9
DATA> 4.2

Deseja calcular o valor do tamanho total da amostra?

n

Digite o valor do tamanho total da amostra:

DATA> 327

Estimação do Tamanho da Amostra

. Tipo de Alocação : Ótima
. Estimativa Minimizando : Custo
. Método de Amostragem : Estratificada
. Amostragem dentro dos estratos : Aleatória Simples Sem Reposição

. Variância Global escolhida : 40

Tamanho Total da Amostra : 326

Alocação

Estrato	Tamanho
1	82
2	54
3	112
4	78

Exemplo 13 : No exemplo seguinte desejamos encontrar os tamanhos amostrais por estrato minimizando a variância global.

Deseja estimar o tamanho da amostra?

y

Digite o número de estratos para análise:

DATA> 4

Qual o tipo de Alocação:

1) Alocação Ótima
2) Alocação Proporcional

DATA> 1

Calcular a estimativa minimizando:

1) Custo
2) Variância Global

DATA> 2

Digite o custo total final:

DATA> 1000

Deseja atribuir valor ao custo da unidade amostral?

y

Digite o custo por unidade amostral:

DATA> 1
DATA> 4
DATA> 3
DATA> 2

Digite o(s) desvio(s) padrão amostral(is):

DATA> 3.2
DATA> 4.1
DATA> 4.9
DATA> 4.2

Deseja calcular o valor do tamanho total da amostra?

y

Estimação do Tamanho da Amostra

. Tipo de Alocação : Ótima
. Estimativa Minimizando : Variância Global
. Método de Amostragem : Estratificada
. Amostragem dentro dos estratos : Aleatória Simples Sem Reposição

. Valor de custo total escolhido : 1000

Tamanho Total da Amostra : 575

Alocação

Estrato	Tamanho
1	247
2	73
3	150
4	105

REFERÊNCIAS BIBLIOGRÁFICAS

AMES, M. H. e WEBSTER, J. T. On estimating approximate degrees of freedom. *The American Statistician* 45, 45-50, 1991.

COCHRAN, W.G.,*Sampling techniques*, 3a. ed., New York:John Wiley & Sons, 1977.

DEMING, W. E. *Some theory of sampling*. New York: Dover Publications, Inc, 1950.

DUNCAN, A. J. *Quality control and industrial statistics*. New York: Richard D. Irwin, Inc, 1986.

HANSEN, M.M., HURWITZ, W.N. E MADOW, W.G.*Sample survey methods and theory* (in two volumes). New York: John Wiley, 1953.

LAHIRI, D. B. A method for sample selection providing unbiased ratio estimates. *Bullettin of International Statistics Institute* 33, 24-57, 1951.

MINITAB Inc. *User's guide: Minitab for windows*, version 12, Pennsilvania, 1998.

MINGOTI, S. A., ATUNCAR, G. S., SILVA, GRANHA, M. L. E SILVA, R. C. *Manual do software "Sampling",* Belo Horizonte: Departamento de Estatística da UFMG,1998

MINGOTI, S.A. e AGUIAR, S., *Elementos básicos de estatística e amostragem com enfoque em mineração*, São Paulo: ABM, 1988.

SATTERTHWAITE, F. E. An approximate distribution of estimates of variance components. *Biometrics Bulletin* 2, 110-114, 1946.

SCHEAFFER, R.A. , MENDENHALL, W. e OTT, L., *Elementary survey sampling*, 5a. ed., Belmont: Duxbury Press, 1996.

SUKHATME, P. V. e SUKHATME, B. V. *Sampling theory of surveys with applications*, Ames: Iowa State University Press, 1970.

THOMPSON, S. K. *Sampling,* New York: John Wiley & Sons, 1992