

“SAMPLING”

**UM SOFTWARE NA ÁREA DE AMOSTRAGEM
INTEGRADO AO SOFTWARE ESTATÍSTICO
“MINITAB FOR WINDOWS”**

MANUAL DO USUÁRIO

APOIO FINANCEIRO DA FAPEMIG

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

**(*) PROFS. ADJUNTO DO DEPARTAMENTO DE ESTATÍSTICA DA UFMG
(**) PROFA. APOSENTADA DO DEPARTAMENTO DE ESTATÍSTICA DA UFMG
(***) BACHAREL E MESTRANDO EM ESTATÍSTICA PELA UFMG**

2001

“SAMPLING”

***UM SOFTWARE NA ÁREA DE AMOSTRAGEM
INTEGRADO AO SOFTWARE ESTATÍSTICO
“MINITAB FOR WINDOWS”***

AGRADECIMENTO

***OS AUTORES AGRADECEM À FAPEMIG PELO APOIO
FINANCEIRO QUE POSSIBILITOU A PRODUÇÃO DO
SOFTWARE “SAMPLING”.***

“SAMPLING”

UM SOFTWARE NA ÁREA DE AMOSTRAGEM INTEGRADO AO SOFTWARE ESTATÍSTICO “MINITAB FOR WINDOWS”

ÍNDICE

O SOFTWARE “SAMPLING”	03
AMOSTRAGEM ALEATÓRIA SIMPLES	07
AMOSTRAGEM SISTEMÁTICA	12
AMOSTRAGEM DO TIPO P.P.S	15
AMOSTRAGEM ESTRATIFICADA	17
AMOSTRAGEM POR CONGLOMERADOS EM 1, 2 OU 3 ESTÁGIOS – SUBAMOSTRAGEM	24
COMO FAZER UMA ANÁLISE	35
REFERÊNCIAS BIBLIOGRÁFICAS	164



Introdução

O software **Sampling** foi desenvolvido como parte de um projeto de pesquisa financiado pela FAPEMIG, uma instituição de fomento à pesquisa do estado de Minas Gerais. Sua estrutura foi concebida de modo a permitir 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. Este programa funciona como um complemento do popular software estatístico "**Minitab For Windows**" (2000). Tal opção pela criação do **Sampling** como um complemento de outro software deve-se principalmente à grande difusão, baixo custo e facilidade de manipulação do "**Minitab For Windows**". **Sampling** é um programa muito simples de ser usado, funcionando de forma totalmente interativa. Seu manuseio requer do usuário conhecimentos básicos sobre o "**Minitab For Windows**" além de, é claro, conhecimento das técnicas de amostragem. O programa funciona 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 `a uma análise estatística, sem precisar sair do software. 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. Cabe salientar que no sistema *Windows 95 (ou superior)*, a capacidade de armazenamento de dados é limitada apenas pela quantidade de memória disponível ("**Minitab**" versão 11.0 ou superior). Configurações inferiores podem ser utilizadas, porém o processamento torna-se mais lento. O programa também pode ser utilizado com versões anteriores do **Minitab** a partir da versão 10.0.

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](#)
- . [Subamostragem Estratificada](#).
- . [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 Estratificada 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 [Subamostragem](#).

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 neste caso 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 %. Para a construção dos intervalos é utilizada a distribuição t-Student para o

caso de amostras pequenas (amostras com tamanho menor que 30 unidades amostrais) e a distribuição Normal para caso de amostras grandes.

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” com a respectiva [alocação ótima ou proporcional](#).

Exemplos de uso do software **Sampling** estão apresentados na seção [Como Fazer uma Análise](#).

Uma breve descrição dos recursos disponíveis em cada um dos planos amostrais existentes no Sampling é apresentada a seguir.

. Amostragem Aleatória Simples com ou sem reposição

No Sampling para executar uma análise estatística de dados quando o usuário dispõe de uma única amostra coletada pelo método da Amostragem Aleatória Simples este deverá proceder como se estivesse executando uma análise pelo procedimento de [Amostragem Estratificada](#) com um único estrato.

Para detalhes sobre a entrada de dados e a forma de saída de resultados veja o a seção [Como Fazer uma Análise-Amostragem Aleatória Simples](#).

. Amostragem Sistemática

Neste módulo o usuário poderá estimar pontualmente os parâmetros populacionais usando os estimadores usuais da Amostragem Sistemática (Thompson,1992) como também fazer a estimação por intervalo. Neste caso, o cálculo da variância dos estimadores é feita através de 3 procedimentos diferentes:

. [Método 1](#): A variância é estimada utilizando-se a aproximação pela Amostragem Aleatória Simples (Cochran, 1977);

. [Método 2](#): A variância é estimada utilizando-se o [método das diferenças](#) (Mingoti e Aguiar,1988).

. [Método 3](#): A variância é estimada pelo [método da amostragem sistemática repetida](#) (Scheaffer, Mendenhall e Ott, 1996) . Neste caso, são considerados partições em que o tamanho da amostra n é um múltiplo de $m=2$ ou $m=3$;

Como uma forma do usuário avaliar se a aproximação pela Amostragem Aleatória Simples é apropriada para a estimação da variância dos estimadores, o software **Sampling** fornece uma [estimativa da correlação intra-classe das unidades amostrais populacionais](#) para avaliação. Valores próximos de zero indicariam que a aproximação seria razoável.

Uma análise descritiva dos dados amostrais (opcional) também está disponível neste módulo. Para maiores detalhes sobre entrada de dados e forma de saída de resultados veja a seção [Como Fazer uma Análise- Amostragem Sistemática](#).

. Amostragem Estratificada

Neste módulo presume-se que as unidades amostrais da população estejam previamente divididas em L estratos disjuntos. Além da estimação de parâmetros populacionais através dos métodos não-viciados usuais ou pelos métodos de Razão e Regressão, o usuário poderá também fazer uma [análise descritiva](#) dos dados para cada estrato individualmente ou para amostra global.

Os planos amostrais permitidos no software Sampling para seleção de unidades amostrais dentro de cada estrato são o da [Amostragem Aleatória Simples](#) ([Amostragem Aleatória Estratificada](#)) e o da [Amostragem Sistemática](#). No caso em que as unidades amostrais de cada estrato estão divididas em conglomerados ou ocorre a [subamostragem](#) dentro de cada estrato, os planos amostrais permitidos são o da [Amostragem Aleatória Simples](#), o da [Amostragem tipo P.P.S.](#) e o de

[Amostragem Sistemática](#) no qual a estimação da variância dos estimadores é feita pela aproximação da Amostragem Aleatória Simples.

Os métodos de estimação considerados em cada estrato são : o [método não-viciado usual](#), [os métodos de Razão e o de Regressão](#).

No caso da estimação por intervalo para amostras pequenas a distribuição t-Student é utilizada com os graus de liberdade dados pela fórmula sugerida por Cochran (1977, pg.96). O software **Sampling** permite também que o usuário proceda ao cálculo do tamanho global da amostra n e a respectiva alocação dos tamanhos amostrais dos estratos populacionais, isto é (n_1, n_2, \dots, n_L) . São consideradas a [Alocação Ótima](#) e a [Alocação Proporcional](#). No caso da Alocação Ótima dois casos são considerados: a determinação do tamanho da amostra levando-se em consideração a [minimização do custo final de coleta de dados](#) ou a [minimização da variabilidade dos estimadores](#) não-viciados usuais dos parâmetros populacionais considerando-se que o plano amostral utilizado em cada estrato é o de Amostragem Aleatória Simples ou a Amostragem Sistemática com variância aproximada pela Amostragem Aleatória Simples. Para maiores detalhes sobre a entrada de dados e a forma de saída de resultados veja a [seção Como Fazer uma Análise-Amostragem Estratificada](#).

. Amostragem do Tipo P.P.S.

A [amostragem do tipo P.P.S.](#) é um procedimento amostral probabilístico no qual a probabilidade de seleção de cada unidade amostral populacional é proporcional ao respectivo tamanho da unidade amostral. Portanto, é um procedimento de amostragem bastante útil nas situações em que as unidades amostrais populacionais diferem de tamanho. A notação P.P.S. significa “probabilidade de seleção proporcional ao tamanho” (Scheaffer, Mendenhall e Ott, 1996).

O software **Sampling** permite que o usuário estime os parâmetros populacionais no caso em que tem uma única amostra selecionada pelo método P.P.S com reposição, ou nos casos em que a Amostragem P.P.S. é utilizada em combinação com a [Amostragem Estratificada](#) ou [Amostragem em Dois ou Três Estágios](#).

Em termos operacionais no software **Sampling** para executar uma análise estatística de uma única amostra proveniente do plano amostral do tipo P.P.S. o usuário deverá proceder como se estivesse executando uma análise pelo procedimento de Amostragem por Conglomerados escolhendo-se a opção P.P.S.

Para maiores detalhes sobre a entrada de dados e a forma de saída de resultados veja a [seção Como Fazer uma Análise-Amostragem do Tipo P.P.S.](#)

. Amostragem por Conglomerados

Neste caso presume-se que as unidades amostrais populacionais tenham sido previamente divididas em grupos ou conglomerados disjuntos. Os métodos de seleção de conglomerados permitidos são o de [Amostragem Aleatória Simples](#) sem reposição, o da Amostragem tipo P.P.S com reposição e o da [Amostragem Sistemática](#) como uma aproximação da Amostragem Aleatória Simples. Os estimadores para o caso em que a [Amostragem Aleatória Simples com reposição](#) tenha sido utilizado na coleta de dados podem ser obtidos por aproximação considerando-se o caso de populações infinitas, ou de tamanhos muito grandes.

São considerados os casos de conglomerados de mesmo tamanho e de tamanhos diferentes. Os métodos de estimação de parâmetros utilizados são o de [estimadores não-viciados usuais](#) e o [método de Razão](#).

Para maiores detalhes sobre a entrada de dados e a forma de saída de resultados veja a [seção Como Fazer uma Análise-Amostragem por Conglomerados](#).

. Amostragem em Dois ou Três Estágios-Subamostragem

Neste módulo presume-se que as unidades amostrais populacionais estão previamente divididas em [unidades primárias e secundárias](#) (no caso de [Amostragem em Dois Estágios](#)) ou em [unidades primárias, secundárias e de terceiro estágio](#) (no caso de [Amostragem em Três Estágios](#)). Os procedimentos de seleção para as unidades primárias permitidos são o de [Amostragem Aleatória Simples](#), o de [Amostragem do tipo P.P.S com reposição](#) e o de [Amostragem Sistemática](#) com estimativas obtidas por aproximação a Amostragem Aleatória Simples sem reposição. Para a seleção das unidades secundárias e de terceiro estágio são considerados apenas os procedimentos da

Amostragem Aleatória Simples sem reposição e o de Amostragem Sistemática com estimativas aproximadas pela Amostragem Aleatória Simples . Entretanto, o caso da Amostragem Aleatória Simples com reposição pode ser obtido através da aproximação para populações com tamanhos grandes (população infinita)

O software **Sampling** executa a estimação de parâmetros para os casos de unidades primárias de tamanhos iguais ou diferentes combinados com os casos de unidades secundárias e de terceiro estágio de tamanhos iguais ou diferentes. Os métodos de estimação de parâmetros utilizado são o dos [estimadores não-viciados usuais](#) e o [método de Razão](#) (Sukhatme e Sukhatme, 1970).

Para maiores detalhes sobre a entrada de dados e a saída de resultados veja a seção de [Como Fazer uma Análise-Amostragem em Dois ou Três Estágios](#).

. Subamostragem Estratificada

A subamostragem estratificada, isto é a Amostragem Estratificada combinada com a Amostragem por Conglomerados ou a Amostragem em Dois ou Três Estágios em cada estrato também está disponível no software **Sampling**. A estimação de parâmetros é feita pelo método usual de [estimadores não-viciados](#) ou pelo [método da Razão](#).

Para maiores detalhes sobre a entrada de dados e a forma de saída dos resultados veja a seção de [Como Fazer uma Análise](#).

. Pós-Estratificação

A Pós-Estratificação (Scheaffer, Mendenhall e Ott, 1996), aparece com grande frequência em situações nas quais a amostra inicial foi coletada sem levar em consideração “à priori” fatores importantes de estratificação (planos como Amostragem Aleatória Simples ou Sistemática). Assim os estratos populacionais são determinados após a coleta da amostra de n unidades amostrais. Neste caso, os tamanhos amostrais dos estratos (n_h) deixam de ser fixos e passam a ser variáveis aleatórias. Devido a este fato correções são necessárias no cálculo das estimativas das variâncias dos estimadores dos parâmetros populacionais. No software **Sampling** estas correções seguem as sugestões indicadas em Scheaffer, Mendenhall e Ott (1996, e Cochran, 1977).

Para maiores detalhes sobre a entrada de dados e a forma de saída de resultados veja a seção [Como Fazer uma Análise](#).

. Estimação de Proporções

No software **Sampling** a [estimação de proporções](#) populacionais é considerada como um caso particular da estimação de médias populacionais. A variável resposta Y só pode assumir valores do tipo "0" ou "1" . Para maiores detalhes veja a seção [Como Fazer uma Análise-Estimação de Proporções](#).



AMOSTRAGEM ALEATÓRIA SIMPLES

Na Amostragem Aleatória Simples uma amostra de n elementos é selecionada através de um mecanismo aleatório de uma população constituída de N unidades amostrais, $n < N$. O mecanismo de selecção das unidades deve ser tal que cada uma das N unidades amostrais populacionais tenha a mesma probabilidade de pertencer a amostra de pertencer a amostra final. A selecção pode ser feita com ou sem reposição. Quando este tipo de procedimento amostral é utilizado na coleta de dados, o software **Sampling** permite que o usuário obtenha as estimativas pontual e por intervalo da média (\bar{y}) e do total (Y) e proporções populacionais usando os estimadores não-viciados usuais, os de Razão e os de Regressão. Além disso, o usuário poderá calcular o tamanho da amostra necessário para obter uma precisão e erro de estimação fixos.

No que segue será utilizada a seguinte notação:

- N : tamanho da população
- n : tamanho da amostra
- \bar{y} : média da amostra
- s^2 : variância da amostra
- \bar{Y} : média populacional
- \hat{Y} : estimador da média populacional
- Y : total populacional
- \hat{Y} : estimador do total populacional
- σ^2 : variância populacional
- R : razão populacional
- \hat{R} : estimador da razão populacional
- \hat{Y}_R : estimador de razão da média populacional
- \hat{Y}_R : estimador de razão do total populacional
- \bar{y}_{lr} : estimador de regressão da média populacional
- \hat{Y}_{lr} : estimador de regressão do total populacional

Amostragem Aleatória Simples Sem Reposição

Suponha que uma amostra de tamanho n tenha sido selecionada de uma população constituída de N elementos através do procedimento de Amostragem Aleatória Simples sem reposição. Neste caso, os estimadores não viciados da média e total populacionais são respectivamente:

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

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

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

Os intervalos de confiança para a média e total populacional são construídos de acordo com o nível de significância escolhido pelo usuário. Para amostras pequenas ($n < 30$) utiliza-se a distribuição t de Student. Para amostras maiores utiliza-se a distribuição normal.

Amostragem Aleatória Simples com Reposição

No caso da amostra ter sido obtida através do procedimento de Amostragem Aleatória Simples Com Reposição, os estimadores não viciados da média e total populacionais são respectivamente:

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

As variâncias estimadas de \widehat{Y} e de \widehat{Y} são respectivamente:

$$\widehat{Var}(\bar{Y}) = \frac{s^2}{n} \quad \text{e} \quad \widehat{Var}(\widehat{Y}) = N^2 \frac{s^2}{n}$$

Estimação de uma Proporção Populacional

A proporção populacional P é estimada por \widehat{p} a fração de elementos na amostra que possui a característica de interesse e pode ser vista como uma média de valores zeros e uns da amostra. Assim, $\widehat{p} = \bar{y}$ e o tratamento é análogo ao da estimação da média populacional. Para maiores detalhes de entrada de dados e a forma de saída de resultados veja [Como Fazer uma Análise](#).

Determinação do Tamanho da Amostra na Amostragem Aleatória Simples

Se o usuário deseja estimar \bar{Y} com um erro de estimação de no máximo d (tolerância), e com uma probabilidade de que o erro de estimação seja maior que d no máximo igual a α , $0 < \alpha < 1$, isto é:

$$P(|\bar{Y} - \widehat{Y}| > d) \leq \alpha$$

o tamanho da amostra no procedimento de Amostragem Aleatória Simples sem Reposição é calculado através da seguinte fórmula:

$$n = \frac{N \sigma^2}{(N - 1) \left(\frac{d}{z_{\alpha/2}} \right)^2 + \sigma^2}$$

e no caso de Amostragem Aleatória Simples com Reposição:

$$n = \left(\frac{z_{\alpha/2} \sigma}{d} \right)^2$$

O **Sampling** solicitará ao usuário os valores das constantes d , α e uma estimativa do desvio padrão populacional σ . Veja exemplos em [Como Fazer uma Análise](#).

Estimadores de Razão na Amostragem Aleatória Simples

Suponha que uma amostra aleatória de tamanho n tenha sido retirada, sem reposição, de uma população finita contendo N elementos. Deseja-se estimar a razão, a média e o total populacionais utilizando-se a informação amostral da variável resposta Y e de uma variável auxiliar X escolhida previamente. Neste caso, o estimador da razão populacional $R = \frac{Y}{X}$ é definido por:

$$\hat{R} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} = \frac{\bar{y}}{\bar{x}}$$

onde $y_i, x_i, i = 1, 2, \dots, n$, denotam respectivamente os valores amostrais das variáveis Y e X para o i -ésimo elemento amostral coletado.

Seja f a fração de amostragem, $f = n / N$.

A variância estimada de \hat{R} é calculada por:

$$\hat{Var}(\hat{R}) \cong \frac{(1-f)}{n \bar{X}^2} \left[\frac{\sum_{i=1}^n (y_i - \hat{R} x_i)^2}{n-1} \right] = \frac{(1-f)}{n \bar{X}} \left(s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R} s_{yx} \right) \text{ onde:}$$

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}, \quad s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} \quad \text{e} \quad s_{yx} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{n-1}.$$

No cálculo da variância no software **Sampling**, é perguntado ao usuário se ele deseja entrar com o valor da média da variável auxiliar \bar{X} . Em caso negativo, a variância é calculada substituindo \bar{X} por \bar{x} , ou seja a média amostral da variável X .

Os estimadores de razão do total populacional \hat{Y}_R e da média populacional $\hat{\bar{Y}}_R$ são respectivamente,

$$\hat{Y}_R = \frac{\bar{y}}{\bar{x}} X \quad \text{e} \quad \hat{\bar{Y}}_R = \frac{\bar{y}}{\bar{x}} \bar{X}$$

e suas variâncias são estimadas respectivamente por:

$$\hat{Var}(\hat{Y}_R) \cong \frac{N^2(1-f)}{n} \left[\frac{\sum_{i=1}^n (y_i - \hat{R} x_i)^2}{n-1} \right] = \frac{N^2(1-f)}{n} \left(s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R} s_{yx} \right)$$

$$\hat{Var}(\hat{Y}_R) \cong \frac{(1-f)}{n} \left[\frac{\sum_{i=1}^n (y_i - \hat{R} x_i)^2}{n-1} \right] = \frac{(1-f)}{n} (s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R} s_{yx})$$

Para utilizar os estimadores de razão da média e total populacionais é necessário que o usuário forneça o valor da média da variável auxiliar \bar{X} . Caso este valor não seja fornecido, o **Sampling** utilizará os estimadores usuais não viciados de média e total populacionais para a variável X .

As fórmulas apresentadas para as variâncias de \hat{R} , \hat{Y}_R e \hat{Y}_R são aproximadas e as aproximações são satisfatórias somente quando o tamanho da amostra é maior que 30 e os coeficientes de variação de \bar{x} e \bar{y} são menores que 10%. O **Sampling** calcula os coeficientes de variação de \bar{x} e \bar{y} e dá um “aviso de alerta” ao usuário quando algum destes coeficientes de variação não satisfaz a condição exigida.

Os intervalos de confiança para R , Y_R e \bar{Y}_R são construídos com base na aproximação normal.

No caso da Amostragem Aleatória Simples com reposição as fórmulas para a variância dos estimadores de razão são obtidas retirando o fator de correção para população finita $(1-f)$.

Na seção [Como Fazer Uma Análise - Amostragem Aleatória Simples](#) o usuário encontrará um exemplo de aplicação dos estimadores de razão.

. Método de Regressão na Amostragem Aleatória Simples

Um outro método de estimação que incorpora a informação de variáveis auxiliares é o método de Regressão (Cochran, 1977; Thompson, 1992). Suponha que se disponha de uma variável auxiliar X linearmente relacionada com a variável resposta de interesse Y , isto é,

$$Y_i = \beta_0 + \beta_1 X_i$$

onde (X_i, Y_i) é o par de valores observados para a i -ésima unidade amostral populacional, $i=1, 2, \dots, n$. Neste caso, uma estimativa da média populacional é dada por:

$$\bar{y}_{lr} = \bar{y} + \hat{\beta}(\bar{X} - \bar{x})$$

onde $\hat{\beta}$ é a estimativa da inclinação da reta obtida pelo método de mínimos quadrados ordinários (Draper & Smith, 1981), isto é,

$$\hat{\beta} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

\bar{X} é a média populacional da variável auxiliar X , \bar{x} é a média amostral da variável X .

Para grandes amostras uma estimativa da variância de \bar{y}_{lr} é dada por:

$$\hat{Var}(\bar{y}_{lr}) = \frac{1-n/N}{n(n-2)} \sum_{i=1}^n [(y_i - \bar{y}) - \hat{\beta}(x_i - \bar{x})]^2 =$$

O estimador do total populacional é dado por:

$$\hat{Y}_{lr} = N \bar{y}_{lr}$$

e para amostras grandes a variância deste estimador é dada por:

$$\hat{Var}(\hat{Y}_{lr}) = N^2 \hat{Var}(\bar{y}_{lr})$$

Intervalos de confiança para os parâmetros populacionais são construídos da forma usual ou através da distribuição t-Student ou através da distribuição Normal. Os estimadores \bar{y}_{lr} e \hat{Y} são viciados (Cochran, 1977).



O software **Sampling** permite que o usuário proceda as estimações pontual e por intervalo dos parâmetros populacionais de interesse no caso em que o procedimento de amostragem sistemática for utilizado. As estimativas das variâncias dos estimadores da média, total e proporção populacionais, podem ser obtidas por três métodos diferentes: *Método 1*: o da aproximação pela [Amostragem Aleatória Simples](#) *Método 2*: o da Amostragem Sistemática Repetida e o *Método 3*: o das Diferenças Sucessivas. A seguir descrevemos brevemente os procedimentos disponíveis no módulo de amostragem sistemática. A notação utilizada será semelhante àquela do módulo de Amostragem Aleatória Simples.

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. Neste caso, os estimadores não-viciados da média e do total populacionais são dados respectivamente por:

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

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

Uma estimativa das variância de \bar{y}_{sy} pode ser obtidas pelos seguintes métodos:

Método 1 : Aproximação pela Amostragem Aleatória Simples (Cochran,1977);

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

onde s^2 é uma estimativa do quadrado médio populacional descrita no módulo de [Amostragem Aleatória Simples Sem Reposição](#). A qualidade desta aproximação pode ser verificada pelo usuário se ele assim o desejar. Considerando-se a forma alternativa de expressar-se a variância de \bar{y}_{sy} , dada por:

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

onde σ^2 é a 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, conseqüentemente a variância de \bar{y}_{sy} será maior que a variância de \bar{y} da Amostragem Aleatória Simples sem Reposição. 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 forem 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.

A critério do usuário, o software **Sampling** calcula uma estimativa de ρ_w dada por:

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

Método 2: Amostragem Sistemática Repetida (Scheaffer, Mendenhall e Ott,1996)

Neste método, considera-se que a Amostra Sistemática de tamanho n e período de seleção k fornecida pelo usuário, possa ser dividida em duas ou três amostras sistemáticas distintas. No primeiro caso, a amostra de tamanho n é tratada como tendo vindo de c=2 amostras sistemáticas de período de seleção igual a $k^* = 2k$. No segundo caso, a amostra de tamanho n é tratado como vindo de c=3 amostras sistemáticas de período $k^* = 3k$. Neste caso, as estimativas das variâncias de \bar{y}_{sy} e \hat{Y} são dadas por:

$$\hat{Var}(\bar{y}_{sy}) = \frac{\sum_{i=1}^c (\bar{y}_{syi} - \bar{y}_s)^2}{c(c-1)} \quad \text{onde} \quad \bar{y}_s = \frac{\sum_{i=1}^c \bar{y}_{syi}}{c} \quad \text{e} \quad \bar{y}_{syi} \text{ é a média da}$$

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

O software **Sampling** escolhe automaticamente o valor de c conveniente dependendo do tamanho n da amostra fornecida pelo usuário. É importante ressaltar que esta estimativa pode ser bastante imprecisa especialmente quando o tamanho da amostra n for pequeno.

Método 3: Método das Diferenças Sucessivas (Mingoti e Aguiar,1988)

Neste método uma estimativa da variância de \bar{y}_{sy} é calculada através da seguinte fórmula:

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

são as observações sucessivas da amostra sistemática de tamanho n fornecida pelo usuário.

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

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

Nota: Uma outra possibilidade de obter-se uma estimativa da variância de \bar{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 de período de seleção k . Então, um estimador não viciado da média amostral será dado por (Scheaffer, Mendenhall e Ott, 1996):

$$\bar{y}_s = \frac{1}{m} \sum_{i=1}^m \bar{y}_{syi}, \text{ onde } \bar{y}_{syi} \text{ são as respectivas médias amostrais das } m$$

amostras sistemáticas observadas.

Uma estimativa não-viciada da variância deste estimador é dada por:

$$\hat{Var}(\bar{y}_s) = \frac{1}{m(m-1)} \sum_{i=1}^m (\bar{y}_{syi} - \bar{y}_s)^2$$

No software **Sampling** este procedimento de estimação pode ser obtido no módulo que trata de [Amostragem por Conglomerados em Um Estágio - Conglomerados de Tamanhos Iguais](#). Veja o módulo [Como fazer uma análise](#) para exemplos de Amostragem Sistemática no **Sampling**.



Suponha que a população de interesse tenha sido dividida em N unidades amostrais aqui denotadas por U_1, U_2, \dots, U_N . Seja π_i a probabilidade de seleção da unidade amostral

$U_i, i=1, 2, \dots, N, \sum_{i=1}^N \pi_i = 1$. 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:

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

para o total populacional com variância estimada dada por:

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

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:

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

Estes estimadores são não-viciados para qualquer escolha de valores das probabilidades de seleção π_i . Um caso particular é aquele em que as probabilidades de seleção das unidades amostrais são escolhidas proporcionalmente ao tamanho das unidades, onde “tamanho” tem que ser definido por algum critério (porte econômico, número de empregados, área cultivada, etc.). A título de ilustração, se por exemplo U_i tem um certo número M_i de elementos amostrais, 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 elementos entre todas as

unidades amostrais 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:

$$\bar{y}_\pi = \bar{y}_{pps} \quad e \quad \hat{t}_\pi = \hat{t}_{pps} = \hat{Y}_{pps}$$

Os intervalos de Confiança para os parâmetros populacionais são construídos de modo similar ao descrito no módulo de [Amostragem Aleatória Simples](#).

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).

No software **Sampling** a Amostragem do Tipo P.P.S. é executada dentro do módulo de [Amostragem por Conglomerados, Subamostragem e Subamostragem Estratificada](#), pois trata-se de uma opção para estes métodos. Veja [Como fazer uma Análise](#) para maiores detalhes.



Amostragem Estratificada

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.

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](#) e com métodos de estimação de parâmetros como o de [Razão](#) e o de [Regressão](#).

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. A seguir apresentamos os estimadores disponíveis no software **Sampling** para o caso em que a amostra em cada estrato é selecionada através da Amostragem Aleatória Simples com ou sem reposição. Neste caso, os estimadores obtidos são não viciados. No que segue a seguinte notação será adotada:

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

N_h : número de unidades amostrais em cada estrato ($h=1,2,\dots, L$),

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

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

n_h : tamanho amostral no estrato h ($h=1, 2, \dots, L$),

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

Y_h : total populacional do estrato h

Y : total da população

\bar{Y} : média da população,

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

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

A . Amostragem Aleatória Simples sem Reposição em Cada Estrato

O estimador da média populacional é dado por

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

onde \bar{y}_h é a média amostral do estrato h, h=1,2,..., L

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

$$\begin{aligned} \hat{V}(\bar{y}_{st}) &= \sum_{h=1}^L W_h^2 \hat{V}(\bar{y}_h) \\ &= \sum_{h=1}^L W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_h^2}{n_h} \end{aligned}$$

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 $\hat{Y} = N \bar{y}_{st}$. A variância estimada de \hat{Y} é dada por $V(\hat{Y}) = N^2 V(\bar{y}_{st})$

Intervalos de Confiança para a média e total populacionais podem ser construídos de acordo com o nível de significância escolhido pelo usuário. Para amostras grandes utiliza-se a distribuição normal enquanto que para amostras pequenas utiliza-se a distribuição t de Student sendo os graus de liberdade calculados de acordo com a fórmula dada em Cochran (1977, pag. 96).

B. Amostragem Aleatória Simples com Reposição em cada Estrato

Neste caso, a média populacional continua sendo estimada por \bar{y}_{st} como dado anteriormente. Entretanto, a variância estimada deste estimador torna-se igual a :

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

Tamanho da Amostra

A escolha do valor de n e os valores de n_h , ($h=1,2,\dots, L$, $\sum_{h=1}^L n_h = n$), 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 ($h=1,2,\dots, L$).

Seja $V(\bar{y}_{st})$ a variância do estimador da média populacional. Duas soluções são possíveis: a primeira é encontrar-se a alocação $(n, n_1, n_2, \dots, n_L)$ 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:

$$n_h = n \left(\frac{N_h \sigma_h / \sqrt{c_h}}{\sum_{h=1}^L N_h \sigma_h / \sqrt{c_h}} \right) \quad h=1,2, \dots, L$$

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

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

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 substituídos 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 :

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

Estimação de Proporções

No software **Sampling** a estimação de uma proporção 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 a proporção populacional é igual a média populacional, o mesmo ocorrendo para a média e proporção amostral. Para maiores detalhes veja [Como Fazer uma Análise](#).

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.

No software **Sampling** o usuário poderá fazer as estimações de parâmetros considerando a Pós-Estratificação se a amostra inicial de n elementos tiver sido coletada pelos métodos de [Amostragem Aleatória Simples](#) ou por [Amostragem Sistemática](#). Suponha que os pesos W_h dos estratos formados “a posteriori” sejam conhecidos, $h = 1, 2, \dots, L$. Neste caso, um estimador da média populacional é dado por $\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h$, como definido anteriormente. Entretanto, este estimador

não tem a variância como discutido anteriormente devido a aleatoriedade dos valores n_h . 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).

Para fazer a análise de dados pela Pós-Estratificação no software **Sampling**, o usuário deverá proceder a entrada de dados como se estivesse fazendo uma análise pela Amostragem Estratificada. Na série de perguntas iniciais que são feitas pelo software o usuário deverá então informar que se trata de uma Pós-Estratificação para que o **Sampling** possa fazer a correspondente análise procedendo as devidas correções.

A estimação considerando-se a Pós-Estratificação pode a vir 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. Veja [Como fazer uma Análise](#).

Estimadores de Razão e Estratificação

No caso de amostragem estratificada a estimativa do total populacional quando o método de razão é utilizado pode ser obtida por duas formas diferentes descritas a seguir:

. Método Separado

Neste método uma estimativa do total populacional é dada por:

$$\hat{Y}_{Rs} = \sum_{h=1}^L \frac{\bar{y}_h}{\bar{x}_h} X_h$$

onde, \bar{y}_h, \bar{x}_h e X_h são respectivamente as médias amostrais das variáveis Y e X e o total populacional da variável auxiliar X para o h-ésimo estrato. Para grandes amostras a variância estimada deste estimador é dada por:

$$\hat{V}ar(\hat{Y}_{Rs}) = \frac{N_h^2 (1 - n_h/N_h)}{n_h} (s_{yh}^2 + \hat{R}_h^2 s_{xh}^2 - 2 \hat{R}_h \hat{\rho}_h s_{yh} s_{xh})$$

onde

$$\hat{R}_h = \frac{\bar{y}_h}{\bar{x}_h} \quad ; \quad s_{yh}^2 = \frac{1}{n_h - 1} \sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)^2 \quad s_{xh}^2 = \frac{1}{n_h - 1} \sum_{j=1}^{n_h} (x_{hj} - \bar{x}_h)^2$$

$$s_{yhx}^2 = \frac{1}{n_h - 1} \sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)(x_{hj} - \bar{x}_h) \quad e \quad \hat{\rho}_h = \frac{s_{yhx}}{s_{yh} s_{xh}}$$

. Método Combinado

Neste método o total populacional é estimado por:

$$\hat{Y}_{st} = \frac{\bar{y}_{st}}{\bar{x}_{st}} X \quad onde \quad \bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h \quad e \quad \bar{x}_{st} = \sum_{h=1}^L W_h \bar{x}_h$$

e X é o total populacional da variável auxiliar X considerando todos os estratos conjuntamente.

Para grandes amostras a variância estimada deste estimador é dada por:

$$\hat{V}ar(\hat{Y}_{Rs}) = \sum_{h=1}^L \frac{N_h^2 (1 - n_h/N_h)}{n_h} (s_{yh}^2 + \hat{R}^2 s_{xh}^2 - 2 \hat{R} \hat{\rho}_h s_{yh} s_{xh})$$

Em ambos os métodos as estimações para a razão e média populacionais são similares ao apresentado para o total populacional (Cochran, 1977).

Método de Regressão Linear e Estratificação

Para amostras obtidas pelo procedimento de Amostragem Aleatória Estratificada, os estimadores de regressão dos parâmetros populacionais, podem ser obtidos pelos métodos separado e combinado (Cochran, 1977). A seguir descrevemos brevemente estes dois métodos.

. Método Separado

Neste método a média populacional é estimada por:

$$\bar{y}_{lrs} = \sum_{i=1}^n W_h \bar{y}_{lrh} , \text{ onde } \bar{y}_{lrh} = \bar{y}_h + \hat{\beta}_h (\bar{X}_h - \bar{x}_h)$$

onde , \bar{X}_h e \bar{x}_h são respectivamente as médias populacional e amostral da variável auxiliar X , e $\hat{\beta}_h$ é a estimativa de mínimos quadrados da inclinação da reta relativa ao h-ésimo estrato ([Amostragem Aleatória Simples – Método de Regressão](#)).

Para amostras grandes a variância deste estimador pode ser estimada por:

$$\hat{Var}(\bar{y}_{lrs}) = \sum_{h=1}^L \frac{W_h^2 (1 - n_h/N_h)}{n_h} (s_{yh}^2 - 2 \hat{\beta}_h s_{yxh} + \hat{\beta}_h^2 s_{xh}^2)$$

onde s_{yh} , s_{yxh} e s_{xh} são os quadrados médios como definidos anteriormente e são calculados individualmente para cada estrato populacional. Este método é apropriado quando presume-se que a relação linear entre as variáveis Y e X difere de estrato para estrato, ou seja tem-se diferentes valores de β_h .

. Método Combinado

Neste método define-se:

$$\bar{y}_{st} = \sum_{h=1}^L \bar{y}_h \text{ e } \bar{x}_{st} = \sum_{h=1}^L \bar{x}_h$$

A estimativa da média populacional é dada por:

$$\bar{y}_{lrc} = \bar{y}_{st} + \hat{\beta} (\bar{X} - \bar{x}_{st})$$

onde \bar{X} é a média populacional da variável auxiliar considerando todos os estratos conjuntamente, e a inclinação da reta é dada por:

$$\hat{\beta} = [\sum_{h=1}^L \frac{W_h^2 (1 - n_h/N_h)}{n_h (n_h - 1)} s_{xyh}] / [\sum_{h=1}^L \frac{W_h^2 (1 - n_h/N_h)}{n_h (n_h - 1)} s_{xh}^2]$$

Para grandes amostras a variância deste estimador é estimada por:

$$\hat{Var}(\bar{y}_{lrc}) = \sum_{h=1}^L \frac{W_h^2 (1 - n_h/N_h)}{n_h(n_h - 1)} \sum_{i=1}^{n_h} [(y_{hi} - \bar{y}_h) - \hat{\beta}(x_{hi} - \bar{x}_h)]^2$$

Este método é apropriado quando presume-se que a relação linear entre as variáveis Y e X são aproximadamente iguais de estrato para estrato, ou seja tem-se o mesmo valor de β para cada estrato.

Em ambos os métodos as estimações para a razão e média populacionais são similares ao apresentado para o total populacional (Cochran,1977). Para maiores detalhes de aplicação veja seção [Como Fazer uma Análise](#).



AMOSTRAGEM POR CONGLOMERADOS EM 1, 2 OU 3 ESTÁGIOS SUBAMOSTRAGEM

O software **Sampling** permite que o usuário obtenha as estimativas pontual e por intervalo, da média (\bar{Y}), o total (Y) e proporção (P) populacionais, quando o procedimento amostral é o de Amostragem por Conglomerados em 1, 2 ou 3 estágios. Nos casos de 1 ou 2 estágios, as unidades amostrais do segundo estágio podem ter ou não o mesmo tamanho. Se as unidades amostrais têm o mesmo tamanho e todos os conglomerados da amostra são completamente inspecionados, ou seja não existe a subamostragem, então os estimadores utilizados são os não-viciados usuais (Scheaffer, Mendenhall e Ott,1996). Quando os conglomerados têm tamanhos diferentes, a amostra de conglomerados é selecionada pelo procedimento da [Amostragem Aleatória Simples com ou sem reposição](#) e ocorre ou não a [subamostragem](#), o usuário tem a sua disposição dois métodos de estimação diferentes para a estimação de parâmetros: o método de [estimadores não-viciados](#) e o [método de razão](#) (Scheaffer, Mendenhall e Ott,1996). A estimação de parâmetros também é possível de ser executada quando o procedimento utilizado para a escolha dos conglomerados é do tipo [P.P.S. com reposição](#), isto é, quando as probabilidades de seleção dos conglomerados são proporcionais ao tamanho dos mesmos. O software **Sampling** permite também a estimação de média e total populacional no caso em que os dados provêm de esquemas amostrais com [Subamostragem estratificada](#), isto é, esquemas nos quais a população de interesse está dividida em estratos e de cada estrato uma amostra é selecionada através de amostragem por conglomerados em 1, 2 ou 3 estágios.

A [entrada de dados](#) está diretamente relacionada com a escolha do número de estágios do procedimento amostral e com o método de estimação de parâmetros escolhido pelo usuário. A seguir apresentamos os planos amostrais disponíveis no software **Sampling** para Amostragem por Conglomerados.

Para facilitar a exposição a seguinte notação será utilizada :

N : número de conglomerados na população ;

n : número de conglomerados na amostra ;

M_i : número de unidades amostrais secundárias do i -ésimo conglomerado populacional ,
 $i = 1, 2, \dots, N$;

m_i : número de unidades amostrais secundárias selecionadas do i -ésimo conglomerado selecionado na amostra do primeiro estágio , $i = 1, 2, \dots, n$;

P_{ij} : número de unidades amostrais do terceiro estágio que pertencem à j -ésima unidade amostral secundária do i -ésimo conglomerado populacional, $i = 1, 2, \dots, N$;
 $j = 1, 2, \dots, M_i$

p_{ij} : número de unidades amostrais selecionadas no terceiro estágio que pertencem à j -ésima unidade amostral secundária selecionada no segundo estágio e à i -ésimo conglomerado selecionado no primeiro estágio.

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

\bar{y} : estimador da média populacional ;

\hat{Y} : estimador do total populacional ;

(1) Amostragem por Conglomerados em um Estágio

. Conglomerados de Tamanhos iguais

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

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

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

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

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

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

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

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.

. Conglomerados de Tamanhos Diferentes

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

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

Para esta situação o usuário dispõe de dois métodos de estimação diferentes : o de estimadores não-viciados e os estimadores de razão.

Método dos Estimadores Não-Viciados

Os estimadores pontuais para a média e total populacional são dados respectivamente por :

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

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

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

$$\hat{Var}(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{s_b'^2}{n} \text{ onde}$$

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

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

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

Caso o usuário não forneça a informação do valor da constante \bar{M} , o software **Sampling** estima o valor desta constante através da fórmula :

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

ou seja, o tamanho médio amostral dos conglomerados.

Método de Estimação de Razão

De acordo com este procedimento os estimadores da média e total populacional são dados por :

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

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

O estimador da variância de \bar{y} é dada por :

$$\hat{Var}(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{s_b''^2}{n} \text{ onde } s_b''^2 = \frac{1}{(n-1) \bar{M}^2} \sum_{i=1}^n (y_i - M_i \bar{y})^2$$

Caso o valor da constante \bar{M} não seja fornecida pelo usuário esta será estimada por \bar{m} .

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 seja igual a

$$p_i = \frac{M_i}{\sum_{i=1}^N M_i} = \frac{M_i}{N \bar{M}}, \quad i = 1, 2, \dots, N. \quad \text{Suponha ainda que a amostra seja}$$

selecionada com reposição. Então, os estimadores de média e total populacional são dados por :

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

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

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

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

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

$$\hat{Var}(\hat{Y}) = (N \bar{M})^2 \hat{Var}(\bar{y}).$$

Caso o valor da constante \bar{M} não seja fornecida pelo usuário esta será estimada por \bar{m} .

(2) Amostragem por Conglomerados em 2 Estágios

. Unidades Primárias de Mesmo Tamanho

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. A seleção das unidades secundárias é feita por AAS-SR. Neste caso, os estimadores não viciados para a média e total populacional são dados por :

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

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

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

$$\hat{Var}(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{s_b^2}{n} + \frac{1}{N} \left(1 - \frac{m}{M}\right) \frac{s_w^2}{m}$$

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

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

a variância amostral das unidades secundárias observadas do conglomerado i pertencente a amostra do primeiro estágio, $i = 1, 2, \dots, n$.

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

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

Caso o valor da constante \bar{M} não seja fornecida pelo usuário esta será estimada por :

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

. Unidades Primárias de Tamanhos Diferentes

No caso em que os conglomerados têm tamanhos M_i diferentes, $i = 1, 2, \dots, N$ 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.

Para esta situação o usuário dispõe de dois métodos de estimação diferentes: o de de estimadores não-viciados e os estimadores de razão.

Método dos Estimadores Não-Viciados

Os estimadores pontuais para a média e total populacional são dados respectivamente por :

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

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

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

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

onde

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

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

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

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

Caso o usuário não forneça a informação do valor da constante \bar{M} , o software **Sampling** estima o valor desta constante através da fórmula :

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

Método de Estimação de Razão

Neste procedimento a média e o total populacional são estimados por :

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

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

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

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

Caso o valor da constante \bar{M} não seja fornecida pelo usuário esta será estimada por \bar{m} .

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

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 seja igual a

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

Então, os estimadores de média e total populacional são dados por :

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

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

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

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

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

$$\hat{Var}(\hat{Y}) = (N \bar{M})^2 \hat{Var}(\bar{y}).$$

Caso o valor da constante \bar{M} não seja fornecida pelo usuário esta será estimada por \bar{m} .

(3) Amostragem por Conglomerados em Três Estágios

No caso de Amostragem por Conglomerados em 3 estágios dois casos são considerados : Amostragem Aleatória Simples e a Amostragem p.p.s.

. AAS-SR em todos os Estágios de Seleção

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 da média e do total populacional são respectivamente,

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

$$\hat{Y} = NM P \bar{y}$$

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

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

onde :

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

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

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

Suponha que a população em estudo seja tal que N conglomerados populacionais tenham tamanhos diferentes, ou seja $M_i, i = 1, 2, \dots, N$. 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.

$$\text{Seja } p_i = \frac{\sum_{j=1}^{M_i} P_{ij}}{\sum_{i=1}^N \sum_{j=1}^{M_i} P_{ij}} \quad \text{a probabilidade de seleção de cada conglomerado}$$

populacional no primeiro estágio. Sejam m_i e p_{ij} , respectivamente, 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 respectivamente :

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

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

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

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

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

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

Caso o valor da constante $\left(\sum_{i=1}^N \sum_{j=1}^{M_i} P_{ij} \right)$ não seja fornecida pelo usuário esta será estimada por $(N \bar{m} \bar{p})$, onde

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

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

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

Nota: Os casos em que as unidades amostrais são selecionadas por Amostragem Aleatória Simples com reposição em um ou mais estágios dos procedimentos amostrais tratados, as fórmulas de variância simplificam. Nestes casos, bastará eliminar das fórmulas os fatores de correções finitas correspondentes aos estágios em que a Amostragem Aleatória Simples com reposição foi utilizada. No software **Sampling** estas estimativas são obtidas por meio da aproximação para o caso de amostragem em populações infinitas, não havendo portanto, uma opção em separado para tratamento destes planos amostrais.

(4) Subamostragem estratificada

O software **Sampling** permite ao usuário a estimação da média e do total populacional quando a Subamostragem estratificada é utilizada. 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 da média e do total populacional são dados por :

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

onde W_h representa o peso do h-ésimo estrato da população e \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.

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

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

sendo que a $\hat{Var}(\bar{y}_h)$ é calculada de acordo com o esquema amostral utilizado para a coleta de amostra em cada estrato e de acordo com as opções de estimação disponíveis no software "Sampling".

Os valores de W_h são dados respectivamente por :

$$W_h = \frac{\sum_{i=1}^{N_h} M_i}{\sum_{h=1}^L \sum_{i=1}^{N_h} M_i} \quad \text{no caso de Amostragem por Conglomerados em 2 estágios ;}$$

$$W_h = \frac{\sum_{i=1}^{N_h} \sum_{j=1}^{M_i} P_{ij}}{\sum_{h=1}^L \sum_{i=1}^{N_h} \sum_{j=1}^{M_i} P_{ij}} \quad \text{no caso de Amostragem por Conglomerados em 3 estgios.}$$

A constante N_h representa o nmero total de conglomerados da populaoo do h-simo estrato.

O estimador do total populacional  dado por :

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

A varincia estimada de \hat{Y}_{st}  dada por :

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



Como Fazer uma Análise

Este tópico tem como objetivo intensificar a relação entre o **Sampling** e o usuário. Desta forma, perguntas tais como: quais os comandos devem ser utilizados para se fazer uma análise, como os dados devem estar estruturados ou como interpretar os resultados, serão naturalmente respondidas através de exemplos aplicados às mais variadas áreas.

Os exemplos citados ilustrarão de forma clara e objetiva a utilização das diversas técnicas de amostragem implementadas no **Sampling**.

Escolha uma das opções abaixo:

[Amostragem Aleatória Simples](#)

[Amostragem Sistemática](#)

[Amostragem Tipo P.P.S.](#)

[Amostragem por Conglomerados](#)

[Subamostragem](#)

[Amostragem Estratificada](#)

[Cálculo do Tamanho da Amostra](#)

[Trabalhando Com Proporções](#)



Exemplo de Amostragem Aleatória Simples

Considerando o esquema de amostragem aleatória simples, duas situações distintas podem ser avaliadas:

[Amostragem Aleatória Simples Com Reposição](#)

[Amostragem Aleatória Simples Sem Reposição](#)



Exemplo de Amostragem Aleatória Simples

Escolha o tipo de estimador com o qual você deseja trabalhar:

[Estimador Não Viciado Usual](#)

[Estimador Razão](#)

[Estimador Regressão](#)



Amostragem Aleatória Simples Com Reposição

A forma de análise dos dados provenientes de uma [Amostragem Aleatória Simples Com Reposição](#) é muito semelhante à demonstrada para a [Amostragem Aleatória Simples Sem Reposição](#), entretanto, é necessário que se faça a seguinte modificação:

Digite o método de amostragem utilizado dentro dos estratos:

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

Digite a opção 1



Exemplo de Amostragem Aleatória Simples Sem Reposição

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem aleatória simples sem reposição](#) e [estimador não viciado usual](#).

Um técnico está interessado em estimar o número total de árvores, maiores do que um certo tamanho especificado, em uma plantação de 150 hectares. Esta informação será utilizada para determinar o volume total de madeira das árvores da plantação. Uma amostra aleatória simples de 20 pontos, cada um contendo 1 hectare, foi selecionada a partir do mapa da região e o número de árvores maiores do que a especificação foram contadas em cada ponto.

1. Abra o arquivo EX_AASSREU.MTW.

Observe a estrutura dos dados, na primeira coluna, temos o número dos pontos selecionados a partir do mapa, a segunda coluna contém os valores da variável resposta e na terceira coluna são apresentados o tamanho da população e o tamanho da amostra, respectivamente.

Serão utilizadas para a análise a segunda e a terceira coluna.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > **%Sampling;**
 SUBC > **Estrato C2;**
 SUBC > **Tamanho C3.**

4. A seguir o programa deverá fazer uma série de [perguntas](#), entre elas:

Digite o método de amostragem utilizado dentro dos estratos:

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

Digite a opção 2

Escolha o tipo de estimador:

- 1) Usual
- 2) Razão
- 3) Regressão

Digite a opção 1

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

5. Impressão dos Resultados

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

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem e ao estimador selecionado. No exemplo acima, os pontos de coleta dos dados foram selecionados por [amostragem aleatória simples sem reposição](#) sendo o [estimador não viciado usual](#).

N = 150. Representa o tamanho da população. No exemplo, corresponde ao tamanho da plantação.

Mean = 17.35. Representa uma estimativa da média populacional da variável resposta. No exemplo, 17.35 é uma estimativa do número médio de árvores maiores do que o tamanho especificado por ponto (1 hectare).

StDev = 0.8550315. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (15.8715 , 18.8285). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 15.8715 representa o limite inferior e 18.8285 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o número médio populacional de árvores com tamanho superior ao especificado por ponto.

Total = 2602.5000. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do número total de árvores com tamanho maior do que o especificado na plantação de 150 hectares.

StDev = 128.2547321. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (2380.73, 2824.27). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 2380.73 representa o limite inferior e 2824.27 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o número total populacional de árvores com tamanho maior do que o especificado na plantação de 150 hectares.

T Value = 1.72914. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição t-student com 19 graus de liberdade. Neste caso, se o tamanho da amostra fosse maior ou igual à 30 ao invés de **T Value** teríamos **Z Value** que corresponde ao ponto crítico baseado em uma distribuição normal padronizada.



Exemplo de Amostragem Aleatória Simples Sem Reposição

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem aleatória simples sem reposição](#) e [estimador razão](#).

Uma empresa deseja estimar a quantidade média de dinheiro gasto em despesas médicas com empregados durante o primeiro trimestre do ano. Relatórios fiscais do ano anterior mostram que a quantidade média de dinheiro gasto no primeiro quadrimestre do ano passado foi de \$32,50. Uma amostra aleatória simples de 100 registros foi retirada a partir de uma população de 1000 empregados.

1. Abra o arquivo EX_AASSRERZ.MTW.

Observe a estrutura dos dados, na primeira coluna, temos o número dos registros selecionados a partir da população de 1000 empregados, a segunda coluna contém os valores da [variável auxiliar](#), a terceira coluna apresenta os valores da variável resposta e na quarta coluna são apresentados o tamanho da população e o tamanho da amostra, respectivamente.

Serão utilizadas para a análise a segunda, a terceira e a quarta coluna.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %Sampling;
 SUBC > Estrato C3;
 SUBC > Auxiliar C2;
 SUBC > Tamanho C4.

4. A seguir o programa deverá fazer uma série de [perguntas](#), entre elas:

Digite o método de amostragem utilizado dentro dos estratos:

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

Digite a opção 2

Escolha o tipo de estimador:

- 1) Usual
- 2) Razão
- 3) Regressão

Digite a opção 2

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

Deseja entrar com o(s) valor(es) da Média da Variável Auxiliar?

Digite YES

Digite o(s) valor(es) de Mx:

Digite 32.5

5. Impressão dos Resultados

Estimação Global

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

N	Mean	StDev	I.C.
1000	34.0532	0.8273509	(32.6923, 35.4140)
	Total	StDev	I.C.
	34053.1661	827.3509010	(32692.3, 35414.0)
	Ratio	StDev	I.C.
	1.0478	0.0254570	(1.00592, 1.08966)

. Z Value : 1.64485

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem e ao estimador selecionado. No exemplo acima, os registros dos empregados foram selecionados por [amostragem aleatória simples sem reposição](#) sendo o [estimador razão](#).

N = 1000. Representa o tamanho da população. No exemplo, corresponde ao número de empregados da empresa.

Mean = 34.0532. Representa uma estimativa da média populacional da variável resposta. No exemplo, 34.0532 é uma estimativa da quantidade média de dinheiro gasto pela empresa em despesas médicas com empregados no primeiro trimestre do ano.

StDev = 0.8273509. Representa uma estimativa do desvio padrão da média populacional da variável resposta.

I.C. = (32.6923, 35.4140). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 32.6923 representa o limite inferior e 35.4140 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para a quantidade média populacional de dinheiro gasto pela empresa em despesas médicas com os empregados no primeiro trimestre do ano.

Total = 34053.1661. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa da quantidade total de dinheiro gasto pela empresa em despesas médicas com os empregados no primeiro trimestre do ano.

StDev = 827.3509010. Representa uma estimativa do desvio padrão do total populacional da variável resposta.

I.C. = (32692.3, 35414.0). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 32692.3 representa o limite inferior e 35414.0 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para a quantidade total de dinheiro gasto pela empresa em despesas médicas com os empregados no primeiro trimestre do ano.

Ratio = 1.0478. Representa uma estimativa da razão populacional entre os totais populacionais da variável resposta e da [variável auxiliar](#). No exemplo, corresponde a razão entre a quantidade total de dinheiro gasto pela empresa em despesas médicas com empregados no primeiro trimestre deste ano e o primeiro quadrimestre do ano passado.

StDev = 0.0254570. Representa uma estimativa do desvio padrão da razão amostral entre os totais populacionais da variável resposta e da [variável auxiliar](#).

I.C. = (1.00592, 1.08966). Este é um [intervalo de confiança](#) de 95% para a razão populacional entre os totais populacionais da variável resposta e da [variável auxiliar](#) baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 1.00592 representa o limite inferior e 1.08966 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para a razão entre a quantidade total de dinheiro gasto pela empresa em despesas médicas com empregados no primeiro trimestre deste ano e o primeiro quadrimestre do ano passado.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30 ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



Exemplo de Amostragem Aleatória Simples Sem Reposição

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem aleatória simples sem reposição](#) e [estimador regressão](#).

O Ministério da Educação e Cultura interessado em avaliar o reflexo de algumas características do ensino médio sobre o ensino superior, aplica uma prova de matemática nos 486 estudantes aprovados no vestibular para a área de ciências exatas de uma universidade. A nota média obtida na prova foi de 52 pontos. Através do procedimento de amostragem aleatória simples sem reposição, dos 486 estudantes, 10 foram selecionados e tiveram o seu rendimento avaliado na disciplina cálculo I. O objetivo do Ministério é estimar a nota média obtida pelos estudantes aprovados no vestibular para a área de ciências exatas na disciplina cálculo I.

1. Abra o arquivo EX_AASSRERG.MTW.

Observe a estrutura dos dados, na primeira coluna, temos o número dos estudantes selecionados a partir dos 486 aprovados no vestibular, a segunda coluna contém os valores da [variável auxiliar](#), a terceira apresenta os valores da variável resposta e na quarta coluna são apresentados o tamanho da população e o tamanho da amostra, respectivamente.

Serão utilizadas para a análise a segunda, a terceira e a quarta coluna.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

```
3. Session:   MTB > %Sampling;
              SUBC > Estrato C3;
              SUBC > Auxiliar C2;
              SUBC > Tamanho C4.
```

4. A seguir o programa deverá fazer uma série de [perguntas](#), entre elas:

Digite o método de amostragem utilizado dentro dos estratos:

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

Digite a opção 2

Escolha o tipo de estimador:

- 1) Usual
- 2) Razão
- 3) Regressão

Digite a opção 3

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

Digite o(s) valor(es) da Média da Variável Auxiliar:

Digite 52.

5. Impressão dos Resultados

Estimação Global

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

. Coef. Angular : 0.7656

N	Mean	StDev	I.C.
486	80.5934	2.7238672	(75.6002, 85.5865)
	Total	StDev	I.C.
	39168.3783	1323.7994582	(36741.7, 41595.1)

. T Value : 1.83311

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem e ao estimador selecionado. No exemplo acima, os registros dos empregados foram selecionados por [amostragem aleatória simples sem reposição](#) sendo o [estimador regressão](#).

Coef. Angular = 0.7656. Representa uma estimativa da mudança ocasionada na variável resposta ao se acrescentar uma unidade a [variável auxiliar](#).

N = 486. Representa o tamanho da população. No exemplo, corresponde ao número de estudantes aprovados no vestibular para a área de ciências exatas.

Mean = 80.5934. Representa uma estimativa da média populacional da variável resposta. No exemplo, 80.5934 é uma estimativa da nota média obtida pelos alunos aprovados no vestibular para a área de ciências exatas na disciplina cálculo I.

StDev = 2.7238672. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (75.6002, 85.5865). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 75.6002 representa o limite inferior e 85.5865 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para a nota média obtida pelos alunos aprovados no vestibular para a área de ciências exatas na disciplina cálculo I.

Total = 39168.3783. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa da nota total obtida pelos alunos aprovados no vestibular para a área de ciências exatas na disciplina cálculo I.

StDev = 1323.7994582. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (36741.7, 41595.1). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 36741.7 representa o limite inferior e 41595.1 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para a nota total obtida pelos alunos aprovados no vestibular para a área de ciências exatas na disciplina cálculo I.

T Value = 1.83311. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição t-student com 9 graus de liberdade. Neste caso, se o tamanho da amostra fosse maior ou igual à 30 ao invés de **T Value** teríamos **Z Value** que corresponde ao ponto crítico baseado em uma distribuição normal padronizada.



Exemplo de Amostragem Sistemática

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem sistemática](#).

Os administradores de um parque florestal desejam estimar o número médio de pessoas por automóvel que visitam a área em um certo feriado nacional. Experiências passadas mostraram que aproximadamente 400 automóveis entram no parque nesta data. Os administradores decidem tomar uma amostra sistemática de 80 automóveis.

Desta forma, os funcionários irão contar o número de pessoas, a cada cinco automóveis, que passarem pela portaria, a começar pelo segundo automóvel (escolhido aleatoriamente).

1. Abra o arquivo EX_AS.MTW.

Observe a estrutura dos dados, na primeira coluna, temos a ordem de passagem dos automóveis selecionados pela portaria, a segunda coluna contém os valores da variável resposta e na terceira coluna são apresentados o tamanho da população e o tamanho da amostra, respectivamente.

Apenas a segunda e a terceira coluna serão utilizadas para a análise.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %**Sampling**;
 SUBC > **Estrato C2**;
 SUBC > **Tamanho C3**.

4. A seguir o programa deverá fazer uma série de [perguntas](#), entre elas:

Digite o método de amostragem utilizado dentro dos estratos:

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

Digite a opção 3

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

Deseja correlação entre elementos da mesma amostra?

Digite **YES**

5. 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	Total
400	4.1625	1665.0000
StDev Mean		
Method 1	Method 2	Method 3
0.1649271	0.1625000	0.1680115
StDev Total		
Method 1	Method 2	Method 3
65.9708413	65.0000000	67.2045809
IC Mean		IC Total
Method 1		Method 1
(3.89122, 4.43378)		(1556.49, 1773.51)
Method 2		Method 2
(3.89521, 4.42979)		(1558.08, 1771.92)
Method 3		Method 3
(3.88615, 4.43885)		(1554.46, 1775.54)
Estrato		Correlation
-----		-----
1		-0.0125313

. Z Value : 1.64485

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem e ao método de estimação da variância. No exemplo acima, os automóveis foram selecionados por [amostragem sistemática](#) sendo a variância estimada pelos métodos [1](#), [2](#) e [3](#).

N = 400. Representa o tamanho da população. No exemplo, corresponde ao número de automóveis que visitaram o parque em um certo feriado nacional.

Mean = 4.1625. Representa uma estimativa da média populacional da variável resposta. No exemplo, 4.1625 é uma estimativa do número médio de pessoas por automóvel que visitaram o parque em um certo feriado nacional.

StDev Mean

Method 1 = 0.1649271. Representa uma estimativa do desvio padrão da média amostral da variável resposta calculado através do [método 1](#).

Method 2 = 0.1625000. Representa uma estimativa do desvio padrão da média amostral da variável resposta calculado através do [método 2](#).

Method 3 = 0.1680115. Representa uma estimativa do desvio padrão da média amostral da variável resposta calculado através do [método 3](#).

Total = 1665.0000. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde ao número total de pessoas que visitaram o parque em um certo feriado nacional.

StDev Total

Method 1 = 65.9708481. Representa uma estimativa do desvio padrão do total amostral da variável resposta calculado através do [método 1](#).

Method 2 = 65.0000000. Representa uma estimativa do desvio padrão do total amostral da variável resposta calculado através do [método 2](#).

Method 3 = 67.2045822. Representa uma estimativa do desvio padrão do total amostral da variável resposta calculado através do [método 3](#).

I.C. Mean

Method 1 = (3.89122, 4.43378). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 3.89122 representa o limite inferior e 4.43378 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para do número médio de pessoas por automóvel que visitaram o parque em um certo feriado nacional.

Method 2 = (3.89521, 4.42979). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 3.89521 representa o limite inferior e 4.42979 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para do número médio de pessoas por automóvel que visitaram o parque em um certo feriado nacional.

Method 3 = (3.88615, 4.43885). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 3.88615 representa o limite inferior e 4.43885 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para do número médio de pessoas por automóvel que visitaram o parque em um certo feriado nacional.

I.C. Total

Method 1 = (1556.49, 1773.51). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 1556.49 representa o limite inferior e 1773.51 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número total de pessoas que visitaram o parque em um certo feriado nacional.

Method 2 = (1558.08, 1771.92). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 1558.08 representa o limite inferior e 1771.92 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número total de pessoas que visitaram o parque em um certo feriado nacional.

Method 3 = (1554.46, 1775.54). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 1554.46 representa o limite inferior e 1775.54 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número total de pessoas que visitaram o parque em um certo feriado nacional.

Correlacion = -0.0125313. Representa uma estimativa da [correlação intra-classe](#) das unidades populacionais, e neste exemplo indica que a aproximação pela amostragem aleatória simples sem reposição é razoável.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30 ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.

OBS: A estimação da variância populacional e por conseqüência a construção dos intervalos de confiança via método 2 considera somente partições em que o tamanho da amostra (n) é um múltiplo de m=2 ou m=3. Portanto, caso isto não ocorra, todas as estatísticas baseadas no método 2 não serão apresentadas.



Exemplo de Cálculo do Tamanho da Amostra

Escolha o tipo de procedimento amostral utilizado para a coleta dos dados:

[Amostragem Aleatória Simples](#)

[Amostragem Sistemática](#)

[Amostragem Estratificada](#)



Exemplo de Amostragem Sistemática

Para calcular o tamanho da amostra quando o planejamento utilizado na coleta dos dados é o de [Amostragem Sistemática](#), o **Sampling** faz uso das mesmas expressões da [Amostragem Aleatória Simples](#). As características desta aproximação podem ser avaliadas na seção [Cálculo do Tamanho da Amostra - Amostragem Sistemática](#).

Portanto, a forma de análise para este caso é exatamente a mesma proposta em [Cálculo do Tamanho da Amostra - Amostragem Aleatória Simples](#).



Exemplo de Amostragem Aleatória Simples

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo [cálculo do tamanho da amostra](#) quando as unidades amostrais são selecionadas pelo procedimento de [amostragem aleatória simples sem reposição](#).

Um pesquisador está interessado em estimar o peso médio ganho, em 4 semanas, por 1000 galinhas alimentadas com um novo tipo de ração. Utilizando dados de estudos anteriores, o pesquisador decide aproximar a variância populacional por 36 gr^2 . Determine o número de galinhas que devem ser selecionadas para que se tenha uma variância global de estimação igual a 0.25 gr^2 .

1. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).
2. Session: MTB > %**Sampling**
3. A seguir o programa deverá fazer uma série de [perguntas](#):

Deseja estimar o tamanho da amostra?

Digite YES

Digite o número de estratos para análise:

Digite 1

Digite o número de elementos na população:

Digite 1000

Digite a variância global:

Digite 0.25

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

Digite 6

4. Impressão dos Resultados

Estimação do Tamanho da Amostra

. Método de Amostragem : Aleatória Simples Sem Reposição
. Variância Global escolhida : 0.25

Tamanho Total da Amostra : 126

OBS: Para o caso em que o planejamento utilizado na coleta dos dados é o de [amostragem aleatória simples com reposição](#), o *Sampling* utiliza uma aproximação pela [amostragem aleatória simples sem reposição](#). Portanto, o procedimento para cálculo do tamanho da amostra para este caso é o mesmo descrito acima. Maiores detalhes podem ser vistos em [Cálculo do Tamanho da Amostra - Amostragem Aleatória Simples Com Reposição](#).



Exemplo de Amostragem Estratificada

Após o cálculo do tamanho total da amostra (n) existem várias formas de se distribuir as observações entre os estratos. Cada configuração resultante têm características particulares que podem ser avaliadas em [Cálculo do Tamanho da Amostra - Amostragem Estratificada](#).

Escolha o método de alocação do tamanho total da amostra:

[Alocação Ótima Minimizando Custo e](#)
[Alocação Ótima Minimizando Variância](#)

[Alocação Proporcional Minimizando Custo e](#)
[Alocação Proporcional Minimizando Variância](#)



Exemplo de Amostragem Estratificada - Alocação Ótima

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo [cálculo do tamanho da amostra](#) e [alocação ótima](#), quando as unidades amostrais são selecionadas pelo procedimento de [amostragem estratificada](#) combinada a [amostragem aleatória simples sem reposição](#).

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 agência acha 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:

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

1. Abra o arquivo EX_CTAEAO.MTW.

Observe a estrutura dos dados, existe apenas uma coluna. Nessa coluna são apresentados o tamanho total da população, o tamanho populacional do primeiro estrato, o tamanho populacional do segundo estrato e o tamanho populacional do terceiro estrato, respectivamente.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %**Sampling**;
 SUBC > **Tamanho C1**.

4. A seguir o programa deverá fazer uma série de [perguntas](#):

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

5. 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

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de alocação, ao critério de minimização, ao método de amostragem e a variância global escolhida. No exemplo acima, a agência selecionará os moradores pelo método de [amostragem estratificada](#) com [amostragem aleatória simples sem reposição](#) em cada área, além disso, foi utilizada a [alocação ótima](#) minimizando custo para distribuir o tamanho total da amostra entre as áreas e uma [variância global](#) de $1h^2$.

Tamanho Total da Amostra = 57. Representa uma estimativa do tamanho total da amostra necessário para se estimar a média populacional da variável resposta com variância global igual a V . No exemplo, corresponde a uma estimativa do número total de moradores que devem ser entrevistados para que a variância global da estimativa do número médio de horas por semana que os moradores assistem televisão seja $1h^2$.

Alocação

Estrato 1 - Tamanho = 18. Representa uma estimativa do número de observações que devem ser coletadas no estrato 1 para que o custo total final seja mínimo. No exemplo, corresponde ao número de moradores que devem ser entrevistados na área A para que o custo total final da pesquisa seja mínimo.

Estrato 2 - Tamanho = 22. Representa uma estimativa do número de observações que devem ser coletadas no estrato 2 para que o custo total final seja mínimo. No exemplo, corresponde ao número de moradores que devem ser entrevistados na área B para que o custo total final da pesquisa seja mínimo.

Estrato 3 - Tamanho = 17. Representa uma estimativa do número de observações que devem ser coletadas no estrato 3 para que o custo total final seja mínimo. No exemplo, corresponde ao número de moradores que devem ser entrevistados na área rural para que o custo total final da pesquisa seja mínimo.

A variância global seja mínima e o custo total final igual a \$1000.

1. Abra o arquivo EX_CTAEAO.MTW.

Observe a estrutura dos dados, existe apenas uma coluna. Nessa coluna são apresentados o tamanho total da população, o tamanho populacional do primeiro estrato, o tamanho populacional do segundo estrato e o tamanho populacional do terceiro estrato, respectivamente.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %**Sampling**;
 SUBC > **Tamanho C1**.

4. A seguir o programa deverá fazer uma série de [perguntas](#):

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

5. Impressão dos Resultados

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

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de alocação, ao critério de minimização, ao método de amostragem e o custo total final escolhido. No exemplo acima, a agência selecionará os moradores pelo método de [amostragem estratificada](#) com [amostragem aleatória simples sem reposição](#) em cada área, além disso, foi utilizada a [alocação ótima](#) minimizando a variância global para distribuir o tamanho total da amostra entre as áreas e um custo total final igual a \$1000.

Tamanho Total da Amostra = 90. Representa uma estimativa do tamanho total da amostra necessário para se estimar a média populacional da variável resposta com custo total final igual a C. No exemplo, corresponde a uma estimativa do número total de moradores que devem ser entrevistados para que o custo total final da pesquisa seja \$1000.

Alocação

Estrato 1 - Tamanho = 29. Representa uma estimativa do número de observações que devem ser coletadas no estrato 1 para que a variância global seja mínima. No exemplo, corresponde ao número de moradores que devem ser entrevistados na área A para que a variância global da estimativa do número médio de horas por semana que os moradores assistem televisão seja mínima.

Estrato 2 - Tamanho = 22. Representa uma estimativa do número de observações que devem ser coletadas no estrato 2 para que a variância global seja mínima. No exemplo, corresponde ao número de moradores que devem ser entrevistados na área B para que a variância global da estimativa do número médio de horas por semana que os moradores assistem televisão seja mínima.

Estrato 3 - Tamanho = 17. Representa uma estimativa do número de observações que devem ser coletadas no estrato 3 para que a variância global seja mínima. No exemplo, corresponde ao número de moradores que devem ser entrevistados na área rural para que a variância global da estimativa do número médio de horas por semana que os moradores assistem televisão seja mínima.

OBS: Para o caso em que o planejamento utilizado na coleta dos dados é o [amostragem estratificada combinada a amostragem aleatória simples com reposição](#), o *Sampling* utiliza uma aproximação pela [amostragem aleatória simples sem reposição](#). Portanto, o procedimento para cálculo do tamanho da amostra para este caso é o mesmo descrito acima. Maiores detalhes podem ser vistos em [Cálculo do Tamanho da Amostra - Amostragem Estratificada](#).



O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo [cálculo do tamanho da amostra](#) e [alocação proporcional](#) quando as unidades amostrais são selecionadas pelo procedimento de [amostragem estratificada](#) combinada a [amostragem aleatória simples sem reposição](#).

Uma indústria siderúrgica 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) $\hat{\sigma}_2 = 1.2\%$, do Sinter Feed 4 $\hat{\sigma}_3 = 1.1\%$. Além disso, a quantidade de minério de ferro normalmente produzida no período é:

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

Qual o número total de incrementos da amostra e o número em cada que estrato devem ser selecionados de forma que:

o custo seja mínimo e a [variância global](#) igual a $0.02\%^2$.

1. Abra o arquivo EX_CTAEAP.MTW.

Observe a estrutura dos dados, existe apenas uma coluna. Nessa coluna são apresentados o tamanho total da população, o tamanho populacional do primeiro estrato, o tamanho populacional do segundo estrato e o tamanho populacional do terceiro estrato, respectivamente.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > **%Sampling;**
 SUBC > **Tamanho C1.**

4. A seguir o programa deverá fazer uma série de [perguntas](#):

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

5. Impressão dos Resultados

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

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de alocação, ao critério de minimização, ao método de amostragem, ao custo da unidade amostral e a variância global escolhida. No exemplo acima, a indústria selecionará os incrementos pelo método de [amostragem estratificada](#) combinada a [amostragem aleatória simples sem reposição](#), além disso, foi utilizada a [alocação proporcional](#) minimizando custo para distribuir o tamanho total da amostra entre os tipos de Sinter Feed. O custo para se amostrar o minério em cada tipo foi considerado constante e a [variância global](#) escolhida foi de 0.02%².

Tamanho Total da Amostra = 73. Representa uma estimativa do tamanho total da amostra necessário para se estimar a média populacional da variável resposta com variância global igual a V. No exemplo, corresponde a uma estimativa do número total de incrementos que devem ser amostrados para que a variância global da estimativa do teor médio de ferro do Sinter Feed B seja 0.02%².

Alocação

Estrato 1 - Tamanho = 18. Representa uma estimativa do número de observações que devem ser coletadas no estrato 1 para que o custo total final seja mínimo. No exemplo, corresponde ao número de incrementos do Sinter Feed 1 que devem ser selecionados para que o custo total final seja mínimo.

Estrato 2 - Tamanho = 22. Representa uma estimativa do número de observações que devem ser coletadas no estrato 2 para que o custo total final seja mínimo. No exemplo, corresponde ao número de incrementos do Sinter Feed (2+3) que devem ser selecionados para que o custo total final seja mínimo.

Estrato 3 - Tamanho = 17. Representa uma estimativa do número de observações que devem ser coletadas no estrato 3 para que o custo total final seja mínimo. No exemplo, corresponde ao número de incrementos do Sinter Feed 4 que devem ser selecionados para que o custo total final seja mínimo.

A variância global seja mínima e o custo total final igual a \$100.

Suponha que o custo para se amostrar um incremento do Sinter Feed 1 é \$0.90, do Sinter Feed (2+3) \$1.80 e do Sinter Feed 4 \$1.30.

1. Abra o arquivo EX_CTAEAP.MTW.

Observe a estrutura dos dados, existe apenas uma coluna. Nessa coluna são apresentados o tamanho total da população, o tamanho populacional do primeiro estrato, o tamanho populacional do segundo estrato e o tamanho populacional do terceiro estrato, respectivamente.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > **%Sampling;**
 SUBC > **Tamanho C1.**

4. A seguir o programa deverá fazer uma série de [perguntas](#):

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 2

Digite o custo total final:

Digite 100

Deseja atribuir valor ao custo da unidade amostral?

Digite YES

Digite o custo por unidade amostral:

Digite 0.90 Pressione a tecla Enter
1.80 Pressione a tecla Enter
1.30 Pressione a tecla Enter

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

5. Impressão dos Resultados

Estimação do Tamanho da Amostra

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

. Valor de custo total escolhido : 100

Tamanho Total da Amostra : 75

Alocação

Estrato	Tamanho
1	17
2	29
3	29

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de alocação, ao critério de minimização, ao método de amostragem e o custo total final escolhido. No exemplo acima, a indústria selecionará o minério de ferro pelo método de [amostragem estratificada](#) com [amostragem aleatória simples sem reposição](#) em cada tipo de Sinter Feed, além disso, foi utilizada a [alocação proporcional](#) minimizando a variância global para distribuir o tamanho total da amostra entre os tipos e um custo total final igual a \$100.

Tamanho Total da Amostra = 90. Representa uma estimativa do tamanho total da amostra necessário para se estimar a média populacional da variável resposta com custo total final igual a C. No exemplo, corresponde a uma estimativa do número total de incrementos que devem ser amostrados para que o custo total final seja \$100.

Alocação

Estrato 1 - Tamanho = 29. Representa uma estimativa do número de observações que devem ser coletadas no estrato 1 para que a variância global seja mínima. No exemplo, corresponde ao número de incrementos do Sinter Feed 1 que devem ser selecionados para que a variância global da estimativa do teor médio de ferro do Sinter Feed B seja mínima.

Estrato 2 - Tamanho = 22. Representa uma estimativa do número de observações que devem ser coletadas no estrato 2 para que a variância global seja mínima. No exemplo, corresponde ao número de incrementos do Sinter Feed (2+3) que devem ser selecionados para que a variância global da estimativa do teor médio de ferro do Sinter Feed B seja mínima.

Estrato 3 - Tamanho = 17. Representa uma estimativa do número de observações que devem ser coletadas no estrato 3 para que a variância global seja mínima. No exemplo, corresponde ao número de incrementos do Sinter Feed 4 que devem ser selecionados para que a variância global da estimativa do teor médio de ferro do Sinter Feed B seja mínima.

OBS: Para o caso em que o planejamento utilizado na coleta dos dados é o [amostragem estratificada](#) combinada a [amostragem aleatória simples com reposição](#), o *Sampling* utiliza uma aproximação pela [amostragem aleatória simples sem reposição](#). Portanto, o procedimento para cálculo do tamanho da amostra para este caso é o mesmo descrito acima. Maiores detalhes podem ser vistos em [Cálculo do Tamanho da Amostra - Amostragem Estratificada](#).



Exemplo de Amostragem por Conglomerados

Considerando o esquema de amostragem por conglomerados, três situações distintas podem ser avaliadas:

[1 Estágio](#)

[2 Estágios](#)

[3 Estágios](#)

(Para 1 e 2 estágios.)



Exemplo de Amostragem por Conglomerados

Escolha o tipo de estimador com o qual você deseja trabalhar:

[Estimador P.P.S.](#)

[Estimador Razão](#)

[Estimador Não Viciado - Unidades de mesmo tamanho em cada estágio](#)

[Estimador Não Viciado - Unidades de tamanho diferente em cada estágio](#)

(Para 3 estágios.)



Exemplo de Amostragem por Conglomerados

Escolha o tipo de estimador com o qual você deseja trabalhar:

[Estimador P.P.S.](#)

[Estimador Não Viciado - Unidades de mesmo tamanho em cada estágio](#)



Exemplo de Amostragem por Conglomerados - 1 Estágio

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em um estágio](#) e [estimador P.P.S.](#)

O gerente de vendas de um jornal está interessado em estimar o número médio de exemplares comprados por morador no último bimestre. Existem 5000 moradores na área e eles estão distribuídos em 40 setores censitários. Sabe-se que o número de jornais comprados é altamente correlacionado com o número de moradores do setor. Uma amostra de 4 setores é selecionada pelo procedimento de [amostragem proporcional ao tamanho](#) e todos os seus moradores são entrevistados.

1. Abra o arquivo EX1_AC1E.MTW.

Observe a estrutura dos dados, a primeira coluna contém os valores da variável resposta, ou seja, o número de jornais comprados no último bimestre por cada morador dos conglomerados selecionados; na segunda coluna são apresentados o número de conglomerados na população e o número populacional de elementos em cada um dos conglomerados selecionados, respectivamente.

Existe uma ordem lógica para entrada dos dados, portanto, as 100 primeiras observações da coluna '*Número de Jornais Comprados*' são referentes ao 1º conglomerado, as 80 subseqüentes ao 2º, as 80 subseqüentes ao 3º e as 90 restantes ao 4º.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %**Sampling**;
 SUBC > **Estrato C1**;
 SUBC > **Conglomerados C2**.

4. A seguir o programa deverá fazer uma série de perguntas:

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 1

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

5. Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Conglomerados
. Número de Estágios : Um
. Estimadores : P.P.S.
. Tamanho Médio Populacional dos Congl. : 125.00

N	Mean	StDev	I.C.
40	5.1242	0.7761477	(3.84752, 6.40082)

n	Total	StDev	I.C.
4	25620.8333	3880.7386942	(19237.6, 32004.1)

. Z Value : 1.64485

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem, ao número de estágios, ao estimador selecionado e ao tamanho médio populacional dos conglomerados. No exemplo acima, os setores foram selecionados pelo procedimento de [amostragem por conglomerados em um estágio](#) sendo o [estimador P.P.S.](#) Além disso, o tamanho médio populacional dos conglomerados é de 125 moradores.

N = 40. Representa o número de conglomerados na população. No exemplo, corresponde ao número de setores censitários na área.

Mean = 5.1242. Representa uma estimativa da média populacional da variável resposta. No exemplo, 5.1242 é uma estimativa do número médio de jornais comprados por morador no último bimestre.

StDev = 0.7761477. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (3.84752, 6.40082). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 3.84752 representa o limite inferior e 6.40082 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número médio de jornais comprados por morador no último bimestre.

n = 4. Representa o número de conglomerados amostrados. No exemplo, corresponde ao número de setores censitários selecionados.

Total = 25620.8333. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do número total de jornais vendidos na área no último bimestre.

StDev = 3880.7386942. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (19237.6, 32004.1). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 19237.6 representa o limite inferior e 32004.1 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número total de jornais vendidos na área no último bimestre.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30 ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



Exemplo de Amostragem por Conglomerados - 1 Estágio

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em um estágio](#) e [estimador razão](#).

O gerente de vendas de um jornal está interessado em estimar o número médio de exemplares comprados por morador no último bimestre. Existem 5000 moradores na área e eles estão distribuídos em 40 setores censitários. Uma amostra de 4 setores é selecionada pelo procedimento de amostragem aleatória simples sem reposição e todos os seus moradores são entrevistados.

1. Abra o arquivo EX1_AC1E.MTW.

Observe a estrutura dos dados, a primeira coluna contém os valores da variável resposta, ou seja, o número de jornais comprados no último bimestre por cada morador dos conglomerados selecionados; na segunda coluna são apresentados o número de conglomerados na população e o número populacional de elementos em cada um dos conglomerados selecionados, respectivamente.

Existe uma ordem lógica para entrada dos dados, portanto, as 100 primeiras observações da coluna '*Número de Jornais Comprados*' são referentes ao 1º conglomerado, as 80 subseqüentes ao 2º, as 80 subseqüentes ao 3º e as 90 restantes ao 4º.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %**Sampling**;
 SUBC > **Estrato C1**;
 SUBC > **Conglomerados C2**.

4. A seguir o programa deverá fazer uma série de perguntas:

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

5. Impressão dos Resultados

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

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem, ao número de estágios, ao estimador selecionado e ao tamanho médio populacional dos conglomerados. No exemplo acima, os setores foram selecionados pelo procedimento de [amostragem por conglomerados em um estágio](#) sendo o [estimador razão](#). Além disso, o tamanho médio populacional dos conglomerados é de 125 moradores.

N = 40. Representa o número de conglomerados na população. No exemplo, corresponde ao número de setores censitários na área.

Mean = 5.1514. Representa uma estimativa da média populacional da variável resposta. No exemplo, 5.1514 é uma estimativa do número médio de jornais comprados por morador no último bimestre.

StDev = 0.5520147. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (4.24345, 6.05941). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 4.24345 representa o limite inferior e 6.05941 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número médio de jornais comprados por morador no último bimestre.

n = 4. Representa o número de conglomerados amostrados. No exemplo, corresponde ao número de setores censitários selecionados.

Total = 25757.1429. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do número total de jornais vendidos na área no último bimestre.

StDev = 2760.0735749. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (21217.2, 30297.1). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 21217.2 representa o limite inferior e 30297.1 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número total de jornais vendidos na área no último bimestre.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30 ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



Exemplo de Amostragem por Conglomerados - 1 Estágio

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em um estágio](#) e [estimador não viciado - unidades de mesmo tamanho](#).

O gerente de vendas de um jornal 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.

1. Abra o arquivo EX2_AC1E.MTW.

Observe a estrutura dos dados, a primeira coluna contém os valores da variável resposta, ou seja, o número de jornais comprados no último bimestre por cada morador dos conglomerados selecionados; na segunda coluna são apresentados o número de conglomerados na população, o número de conglomerados amostrados e seu número populacional de elementos, respectivamente.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %**Sampling**;
 SUBC > **Estrato C1**;
 SUBC > **Conglomerados C2**.

4. A seguir o programa deverá fazer uma série de perguntas:

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 3

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

5. Impressão dos Resultados

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

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem, ao número de estágios, ao estimador selecionado e ao tamanho médio populacional dos conglomerados. No exemplo acima, os setores foram selecionados pelo procedimento de [amostragem por conglomerados em um estágio](#) sendo o [estimador não viciado - unidades de mesmo tamanho](#). Neste caso, todos os conglomerados têm o mesmo tamanho populacional.

N = 40. Representa o número de conglomerados na população. No exemplo, corresponde ao número de setores censitários na área.

Mean = 5.0125. Representa uma estimativa da média populacional da variável resposta. No exemplo, 5.0125 é uma estimativa do número médio de jornais comprados por morador no último bimestre.

StDev = 0.4295412. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (4.30597, 5.71903). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 4.30597 representa o limite inferior e 5.71903 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número médio de jornais comprados por morador no último bimestre.

n = 4. Representa o número de conglomerados amostrados. No exemplo, corresponde ao número de setores censitários selecionados.

Total = 20050.0000. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do número total de jornais vendidos na área no último bimestre.

StDev = 1718.1647185. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (17223.9, 22876.1). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 17223.9 representa o limite inferior e 22876.1 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número total de jornais vendidos na área no último bimestre.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30 ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



Exemplo de Amostragem por Conglomerados - 1 Estágio

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em um estágio](#) e [estimador não viciado - unidades de tamanho diferente](#).

O gerente de vendas de um jornal está interessado em estimar o número médio de exemplares comprados por morador no último bimestre. Existem 5000 moradores na área e eles estão distribuídos em 40 setores censitários. Uma amostra de 4 setores é selecionada pelo procedimento de amostragem aleatória simples sem reposição e todos os seus moradores são entrevistados.

1. Abra o arquivo EX1_AC1E.MTW.

Observe a estrutura dos dados, a primeira coluna contém os valores da variável resposta, ou seja, o número de jornais comprados no último bimestre por cada morador dos conglomerados selecionados; na segunda coluna são apresentados o número de conglomerados na população e o número populacional de elementos em cada um dos conglomerados selecionados, respectivamente.

Existe uma ordem lógica para entrada dos dados, portanto, as 100 primeiras observações da coluna '*Número de Jornais Comprados*' são referentes ao 1º conglomerado, as 80 subseqüentes ao 2º, as 80 subseqüentes ao 3º e as 90 restantes ao 4º.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %**Sampling**;
 SUBC > **Estrato C1**;
 SUBC > **Conglomerados C2**.

4. A seguir o programa deverá fazer uma série de perguntas:

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 4

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

5. Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Conglomerados
. Número de Estágios : Um
. Estimadores : Não Viciado
Unidades de Tamanho Diferente
. Tamanho Médio Populacional dos Congl. : 125.00

N	Mean	StDev	I.C.
40	3.6060	0.6090547	(2.60419, 4.60781)

n	Total	StDev	I.C.
4	18030.0000	3045.2733867	(13021.0, 23039.0)

. Z Value : 1.64485

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem, ao número de estágios, ao estimador selecionado e ao tamanho médio populacional dos conglomerados. No exemplo acima, os setores foram selecionados pelo procedimento de [amostragem por conglomerados em um estágio](#) sendo o [estimador não viciado - unidades de tamanho diferente](#). Além disso, o tamanho médio populacional dos conglomerados é de 125 moradores.

N = 40. Representa o número de conglomerados na população. No exemplo, corresponde ao número de setores censitários na área.

Mean = 3.6060. Representa uma estimativa da média populacional da variável resposta. No exemplo, 3.6060 é uma estimativa do número médio de jornais comprados por morador no último bimestre.

StDev = 0.6090547. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (2.60419, 4.60781). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 2.60419 representa o limite inferior e 4.60781 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número médio de jornais comprados por morador no último bimestre.

n = 4. Representa o número de conglomerados amostrados. No exemplo, corresponde ao número de setores censitários selecionados.

Total = 18030.0000. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do número total de jornais vendidos na área no último bimestre.

StDev = 3045.2733867. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (13021.0, 23039.0). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 13021.0 representa o limite inferior e 23039.0 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número total de jornais vendidos na área no último bimestre.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30 ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



Exemplo de Amostragem por Conglomerados - 2 Estágios

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em dois estágios](#) e [estimador P.P.S.](#)

Os técnicos de uma indústria de mineração desejam estimar o teor médio de ferro (%) das 3600 toneladas do Sinter Feed B que será carregado em 60 vagões. Os vagões têm capacidades de armazenamento diferentes variando de 20 a 100 toneladas. Sabe-se que o carregamento é bastante homogêneo e existe uma alta correlação entre a capacidade do vagão e o teor de ferro do minério. Selecionou-se 10 vagões através de uma [amostragem proporcional ao tamanho](#) e de cada um dos vagões selecionados retirou-se, aleatoriamente, uma amostra de incrementos:

Vagão	Número de incrementos retirados
01	3
02	6
03	5
04	3
05	4
06	4
07	4
08	5
09	3
10	5

1. Abra o arquivo EX1_AC2E.MTW.

Observe a estrutura dos dados, a primeira coluna contém os valores da variável resposta, ou seja, o teor de ferro (%) dos incrementos pertencentes aos vagões selecionados; na segunda coluna são apresentados o número de conglomerados na população e o número de elementos amostrados em cada um dos conglomerados selecionados, respectivamente.

Existe uma ordem lógica para entrada dos dados, portanto, as 3 primeiras observações da coluna 'Teor de Ferro (%)' são referentes ao 1º conglomerado, as 6 subseqüentes ao 2º, e assim por diante.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %Sampling;
 SUBC > Estrato C1;
 SUBC > Conglomerados C2.

4. A seguir o programa deverá fazer uma série de perguntas:

Digite o número de estágios:

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

Digite 2

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem, ao número de estágios, ao estimador selecionado e ao tamanho médio populacional dos conglomerados. No exemplo acima, os vagões foram selecionados pelo procedimento de [amostragem por conglomerados em dois estágios](#) sendo o [estimador P.P.S.](#) Além disso, o tamanho médio populacional dos conglomerados é de 100000 incrementos.

N = 60. Representa o número de conglomerados na população. No exemplo, corresponde ao número de vagões do carregamento.

Mean = 64.0683. Representa uma estimativa da média populacional da variável resposta. No exemplo, 64.0683 é uma estimativa do teor médio de ferro (%) do Sinter Feed B.

StDev = 0.0034816. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (64.0625, 64.0740). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 64.0625 representa o limite inferior e 64.0740 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o teor médio de ferro(%) do minério carregado nos vagões.

n = 10. Representa o número de conglomerados amostrados. No exemplo, corresponde ao número de vagões selecionados.

Total = 384409600.0000. Representa uma estimativa do total populacional da variável resposta.

StDev = 20889.5685469. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (3.84E+08, 3.84E+08). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30 ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em dois estágios](#) e [estimador razão](#).

Os técnicos de uma indústria siderúrgica desejam estimar o teor médio de ferro (%) das 3600 toneladas do Sinter Feed B que será carregado em 60 vagões. Para isto, selecionou-se 10 vagões através de uma [amostragem aleatória simples](#) e de cada um dos vagões selecionados retirou-se, aleatoriamente, uma amostra de incrementos:

Vagão	Número de incrementos retirados
01	3
02	6
03	5
04	3
05	4
06	4
07	4
08	5
09	3
10	5

1. Abra o arquivo EX1_AC2E.MTW.

Observe a estrutura dos dados, a primeira coluna contém os valores da variável resposta, ou seja, o teor de ferro (%) dos incrementos pertencentes aos vagões selecionados; na segunda coluna são apresentados o número de conglomerados na população e o número de elementos amostrados em cada um dos conglomerados selecionados, respectivamente.

Existe uma ordem lógica para entrada dos dados, portanto, as 3 primeiras observações da coluna 'Teor de Ferro (%)' são referentes ao 1º conglomerado, as 6 subseqüentes ao 2º, e assim por diante.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %**Sampling**;
 SUBC > **Estrato C1**;
 SUBC > **Conglomerados C2**.

4. A seguir o programa deverá fazer uma série de perguntas:

Digite o número de estágios:

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

Digite 2

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem, ao número de estágios, ao estimador selecionado e ao tamanho médio populacional dos conglomerados. No exemplo acima, os vagões foram selecionados pelo procedimento de [amostragem por conglomerados em dois estágios](#) sendo o [estimador razão](#). Além disso, o tamanho médio populacional dos conglomerados é de 100000 incrementos.

N = 60. Representa o número de conglomerados na população. No exemplo, corresponde ao número de vagões do carregamento.

Mean = 64.0700. Representa uma estimativa da média populacional da variável resposta. No exemplo, 64.0700 é uma estimativa do teor médio de ferro (%) do Sinter Feed B.

StDev = 0.0000001. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (64.0700, 64.0700). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 64.0700 representa o limite inferior e 64.0700 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o teor médio de ferro (%) do minério carregado nos vagões.

n = 10. Representa o número de conglomerados amostrados. No exemplo, corresponde ao número de vagões selecionados.

Total = 384420000.0000. Representa uma estimativa do total populacional da variável resposta.

StDev = 0.8640988. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (3.84E+08, 3.84E+08). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30 ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em dois estágios](#) e [estimador não viciado - unidades de mesmo tamanho](#).

Um sociólogo está interessado em estimar o número total de pessoas aposentadas que residem em uma certa região. Ele decide retirar através do procedimento de [amostragem aleatória simples](#) uma amostra de 4 setores censitários a partir dos 30 que compõem a região e de cada um deles selecionar 6 domicílios, como apresentado na tabela a seguir:

Setor Censitário	Número de Domicílios	Número de domicílios selecionados	Número de pessoas aposentadas em cada domicílio selecionado
6	18	6	2, 0, 1, 3, 0, 1
19	18	6	1, 1, 0, 3, 2, 0
24	18	6	0, 0, 1, 1, 1, 2
29	18	6	3, 0, 2, 0, 2, 2

1. Abra o arquivo EX2_AC2E.MTW.

Observe a estrutura dos dados, a primeira coluna contém os valores da variável resposta, na segunda coluna são apresentados o número de conglomerados populacional e amostral e o número de unidades em cada conglomerado populacional e amostral, respectivamente.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %**Sampling**;
 SUBC > **Estrato C1**;
 SUBC > **Conglomerados C2**.

4. A seguir o programa deverá fazer uma série de perguntas:

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 3

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

5. 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. : 18.00

N	Mean	StDev	I.C.
30	1.1667	0.1431567	(0.921314, 1.41202)
n	Total	StDev	I.C.
4	630.0000	77.3045924	(497.509, 762.491)

. T Value : 1.71388

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem, ao número de estágios, ao estimador selecionado e ao tamanho médio populacional dos conglomerados. No exemplo acima, os setores foram selecionados pelo procedimento de [amostragem por conglomerados em dois estágios](#) sendo o [estimador não viciado - unidades de mesmo tamanho](#). Além disso, o tamanho médio populacional dos conglomerados é de 18 residentes.

N = 30. Representa o número de conglomerados na população. No exemplo, corresponde ao número de setores censitários que compõe a região.

Mean = 1.1667. Representa uma estimativa da média populacional da variável resposta. No exemplo, 1.1667 é uma estimativa do número médio de residentes aposentados por domicílio.

StDev = 0.1431567. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (0.921314, 1.41202). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 0.921314 representa o limite inferior e 1.41202 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o número médio de residentes aposentados por domicílio.

n = 4. Representa o número de conglomerados amostrados. No exemplo, corresponde ao número de setores censitários selecionados.

Total = 630.0000. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do número total de pessoas aposentados que residem na região.

StDev = 77.3045924. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (497.509, 762.491). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 497.509 representa o limite inferior e 762.491 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o número total de pessoas aposentados que residem na região.

T Value = . Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição t-student com 23 graus de liberdade. Neste caso, se o tamanho da amostra fosse maior ou igual à 30 ao invés de **T Value** teríamos **Z Value** que corresponde ao ponto crítico baseado em uma distribuição normal padronizada.



Exemplo de Amostragem por Conglomerados - 2 Estágios

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em dois estágios](#) e [estimador não viciado - unidades de tamanho diferente](#).

Um sociólogo está interessado em estimar o número total de pessoas aposentadas que residem em uma certa região. Ele decide retirar através do procedimento de [amostragem aleatória simples](#) uma amostra de 4 setores censitários a partir dos 30 que compõem a região e de cada um deles selecionar um certo número de domicílios, como apresentado na tabela a seguir:

Setor Censitário	Número de Domicílios	Número de domicílios selecionados	Número de pessoas aposentadas em cada domicílio selecionado
6	18	6	2, 0, 1, 3, 0, 1
19	14	3	1, 1, 0
24	9	3	0, 0, 1
29	12	4	3, 0, 2, 0

Sabe-se que, em média, cada setor censitário da região contém 15 domicílios.

1. Abra o arquivo EX3_AC2E.MTW.

Observe a estrutura dos dados, a primeira coluna contém os valores da variável resposta, na segunda coluna são apresentados o número de conglomerados na população, o número de unidades secundárias amostradas e o número de unidades secundárias populacional para cada conglomerado, respectivamente. Existe uma ordem lógica para entrada dos dados, portanto, as 6 primeiras observações da coluna '*Número de Pessoas Aposentadas*' são referentes ao 1º conglomerado que apresenta tamanho populacional igual a 18, as 3 subseqüentes ao 2º com tamanho populacional igual a 14, e assim por diante.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %**Sampling**;
 SUBC > **Estrato C1**;
 SUBC > **Conglomerados C2**.

4. A seguir o programa deverá fazer uma série de perguntas:

Digite o número de estágios:

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

Digite 2

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem, ao número de estágios, ao estimador selecionado e ao tamanho médio populacional dos conglomerados. No exemplo acima, os setores foram selecionados pelo procedimento de [amostragem por conglomerados em dois estágios](#) sendo o [estimador não viciado - unidades de tamanho diferente](#). Além disso, o tamanho médio populacional dos conglomerados é de 15 domicílios.

N = 30. Representa o número de conglomerados na população. No exemplo, corresponde ao número de setores censitários que compõe a região.

Mean = 0.8056. Representa uma estimativa da média populacional da variável resposta. No exemplo, 0.2417 é uma estimativa do número médio de residentes aposentados por domicílio.

StDev = 0.2486780. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (0.369610, 1.24150). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 0.369610 representa o limite inferior e 1.24150 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o número médio de residentes aposentados por domicílio.

n = 4. Representa o número de conglomerados amostrados. No exemplo, corresponde ao número de setores censitários selecionados.

Total = 362.5000. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do número total de pessoas aposentados que residem na região.

StDev = 111.9050937. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (166.324, 558.676). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 166.324 representa o limite inferior e 558.676 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o número total de pessoas aposentados que residem na região.

T Value = 1.75305. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição t-student com 15 graus de liberdade. Neste caso, se o tamanho da amostra fosse maior ou igual à 30 ao invés de **T Value** teríamos **Z Value** que corresponde ao ponto crítico baseado em uma distribuição normal padronizada.



Exemplo de Amostragem por Conglomerados - 3 Estágios

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em três estágios](#) e [estimador P.P.S.](#)

Um pesquisador está interessado 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, o pesquisador resolveu amostrá-las com probabilidade proporcional ao número de habitantes. Em cada região selecionada será retirada uma amostra de maternidades e de cada uma destas uma amostra de alas e então será verificado o número de cesarianas realizadas no primeiro bimestre do ano. Sabe-se que existem no total 400 alas, considerando-se as 20 maternidades.

Região	Maternidade	Número de alas selecionadas	Número de cesarianas
1	2	8	21, 37, 38, 47, 33, 30, 43, 30
	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

1. Abra o arquivo EX1_AC3E.MTW.

Observe a estrutura dos dados, a primeira coluna contém os valores da variável resposta, na segunda coluna são apresentados o número de conglomerados na população e na amostra, o número de unidades secundárias amostradas em cada conglomerado e o número de unidades do terceiro estágio retiradas de cada unidade secundária, respectivamente.

Existe uma ordem lógica para entrada dos dados, portanto, as 8 primeiras observações da coluna 'Número de Cesarianas' são referentes ao 1º conglomerado e a sua 1ª unidade secundária, as 4 subsequentes ao 1º conglomerado e a sua 2ª unidade secundária, e assim por diante.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %Sampling;
 SUBC > Estrato C1;
 SUBC > Conglomerados C2.

4. A seguir o programa deverá fazer uma série de perguntas:

Digite o número de estágios:

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

Digite 3

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem, ao número de estágios, ao estimador selecionado e ao número de elementos da população. No exemplo acima, as regiões foram selecionados pelo procedimento de [amostragem por conglomerados em três estágios](#) sendo o [estimador P.P.S.](#) Além disso, existem no total 400 alas de hospital na cidade.

N = 4 Representa o número de conglomerados na população. No exemplo, corresponde ao número de regiões em que a cidade está dividida.

Mean = 41.8095. Representa uma estimativa da média populacional da variável resposta. No exemplo, 41.8095 é uma estimativa do número médio de operações cesarianas realizadas por ala de hospital.

StDev = 3.5186980. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (36.0218, 47.5973). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 36.0218 representa o limite inferior e 47.5973 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número médio de operações cesarianas realizadas por ala de hospital.

n = 2. Representa o número de conglomerados amostrados. No exemplo, corresponde ao número de regiões da cidade selecionadas.

Total = 16723.8095. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do número total de operações cesarianas realizadas na cidade no primeiro bimestre do ano.

StDev = 1407.4792121. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (14408.7, 19038.9). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 14408.7 representa o limite inferior e 19038.9 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número total de operações cesarianas realizadas na cidade no primeiro bimestre do ano.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30 ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



Exemplo de Amostragem por Conglomerados - 3 Estágios

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em três estágios](#) e [estimador não viciado - unidades de mesmo tamanho](#).

Com o objetivo de melhorar o serviço telefônico, um dos diretores executivos de uma empresa decide estimar o número total de chamadas telefônicas efetuadas por secretárias durante um dia de trabalho. A empresa é composta por uma matriz e três filiais, cada uma contendo 12 departamentos, e elas efetuam aproximadamente o mesmo número de chamadas telefônicas diárias. Cada departamento emprega 20 secretárias e o número de chamadas efetuadas varia consideravelmente entre elas.

Firma	Departamento	Número de secretárias	Número de secretárias selecionadas	Número de chamadas efetuadas
Matriz	2	20	10	5, 10, 0, 20, 11, 6, 3, 15, 5, 0
	6	20	10	15, 2, 7, 11, 0, 25, 30, 5, 9, 15
	11	20	10	10, 35, 21, 12, 3, 8, 15, 0, 21, 30
	20	20	10	1, 14, 5, 50, 12, 2, 21, 13, 1, 10
Filial 2	5	20	10	15, 0, 0, 29, 10, 6, 3, 15, 5, 1
	6	20	10	25, 2, 7, 1, 0, 35, 30, 5, 0, 9
	13	20	10	40, 15, 21, 2, 3, 14, 15, 0, 11, 38
	18	20	10	21, 4, 59, 5, 21, 12, 1, 13, 1, 10

1. Abra o arquivo EX2_AC3E.MTW.

Observe a estrutura dos dados, a primeira coluna contém os valores da variável resposta, na segunda coluna são apresentados o número de conglomerados na população e na amostra, o número de unidades secundárias na população e na amostra e o número de unidades do terceiro estágio na população e na amostra, respectivamente.

Existe uma ordem lógica para entrada dos dados, portanto, as 8 primeiras observações da coluna 'Número de Cesarianas' são referentes ao 1º conglomerado e a sua 1ª unidade secundária, as 4 subseqüentes ao 1º conglomerado e a sua 2ª unidade secundária, e assim por diante.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %Sampling;
 SUBC > Estrato C1;
 SUBC > Conglomerados C2.

4. A seguir o programa deverá fazer uma série de perguntas:

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

5. 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 : 12.00

. Unidades do terceiro estágio : 20.00

Tamanho Amostral

. Unidades do segundo estágio : 4.00

. Unidades do terceiro estágio : 10.00

N	Mean	StDev	I.C.
4	12.2750	0.8262214	(10.9160, 13.6340)

n	Total	StDev	I.C.
2	11784.0000	793.1725327	(10479.3, 13088.7)

. Z Value : 1.64485

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem, ao número de estágios, ao estimador selecionado e ao tamanho populacional e amostral das unidades do segundo e do terceiro estágios. No exemplo acima, as firmas foram selecionadas pelo procedimento de [amostragem por conglomerados em três estágios](#) sendo o [estimador não viciado - unidades de mesmo tamanho](#). Além disso, existem 12 departamentos em cada firma, destes, 4 foram amostrados. Em cada departamento há 20 secretárias, sendo que 10 foram sorteadas.

N = 4 Representa o número de conglomerados na população. No exemplo, corresponde ao número de firmas que compõe a empresa.

Mean = 12.2750. Representa uma estimativa da média populacional da variável resposta. No exemplo, 12.2750 é uma estimativa do número médio de chamadas telefônicas efetuadas por secretária em um dia de trabalho.

StDev = 0.8262214. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (10.9160, 13.6340). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 10.9160 representa o limite inferior e 13.6340 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número médio de chamadas telefônicas efetuadas por secretária em um dia de trabalho.

n = 2. Representa o número de conglomerados amostrados. No exemplo, corresponde ao número de firmas da empresa selecionadas.

Total = 11784.0000. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do número total de ligações telefônicas efetuadas pelas secretárias da empresa em um dia de trabalho.

StDev = 793.1725327. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (10479.3, 13088.7). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 10479.3 representa o limite inferior e 13088.7 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número total de ligações telefônicas efetuadas pelas secretárias da empresa em um dia de trabalho.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30 ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



Exemplo de Amostragem Estratificada

Considerando o procedimento de [amostragem estratificada](#) é necessário indicar qual o método utilizado na coleta dos dados dentro de cada estrato:

[Amostragem Aleatória Simples](#)

[Amostragem Sistemática](#)

[Amostragem por Conglomerados](#)



Exemplo de Amostragem Estratificada

Considerando o esquema de amostragem estratificada com amostragem aleatória simples em cada estrato, duas situações distintas podem ser avaliadas:

[Amostragem Estratificada com Amostragem Aleatória Simples Com Reposição](#)

[Amostragem Estratificada com Amostragem Aleatória Simples Sem Reposição](#)



Amostragem Estratificada com Amostragem Aleatória Simples

A forma de análise dos dados provenientes de uma Amostragem Estratificada com [Amostragem Aleatória Simples Com Reposição](#) é muito semelhante à demonstrada para a Amostragem Estratificada com [Amostragem Aleatória Simples Sem Reposição](#), entretanto, é necessário que se faça a seguinte modificação:

Digite o método de amostragem utilizado dentro dos estratos:

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

Digite a opção 1



Exemplo Amostragem Estratificada

Escolha o tipo de estimador com o qual você deseja trabalhar:

[Estimador Não Viciado Usual](#)

[Estimador Razão](#)

[Estimador Regressão](#)



Exemplo de Amostragem Estratificada

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo [a amostragem estratificada](#) com [amostragem aleatória simples sem reposição](#) e [estimador não viciado usual](#).

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.

1. Abra o arquivo EX_AEEU.MTW.

Observe a estrutura dos dados, as três primeiras colunas contêm os valores da variável resposta para cada estrato, ou seja, o tempo (h) em que os moradores entrevistados declararam assistir televisão em uma semana para a cidade A, cidade B e para a área rural; na quarta coluna são apresentados o número de elementos na população e o número populacional de elementos nos estratos 1, 2 e 3; respectivamente.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %Sampling;
 SUBC > Estrato C1 - C3;
 SUBC > Tamanho C4.

4. A seguir o programa deverá fazer uma série de [perguntas](#), entre elas:

Digite o método de amostragem utilizado dentro dos estratos:

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

Digite a opção 2

Escolha os tipos de estimadores:

- 1) Usuais
- 2) Razão
- 3) Regressão

Digite a opção 1

Existe Pós-Estratificação?

Digite NO

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

5. 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	1.4033956	(25.2601, 30.0899)
	Total	StDev	I.C.
	8579.2500	435.0526503	(7830.63, 9327.87)

. T Value : 1.72075

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem global e dentro dos estratos e ao estimador selecionado. No exemplo acima, os moradores foram selecionados por [amostragem estratificada](#) com [amostragem aleatória simples sem reposição](#) sendo o [estimador não viciado usual](#).

N = 310. Representa o tamanho da população. No exemplo, corresponde ao número de moradores no condado.

Mean = 27.6750. Representa uma estimativa da média populacional da variável resposta. No exemplo, 27.6750 é uma estimativa do número médio de horas por semana que os moradores do condado assistem televisão.

StDev = 1.4033956. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (25.2601, 30.0899). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 25.2601 representa o limite inferior e 30.0899 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o número médio de horas por semana que os moradores do condado assistem televisão.

Total = 8579.2500. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do número total de horas por semana que os moradores do condado assistem televisão.

StDev = 435.0526503. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (7830.63, 9327.87). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 7830.63 representa o limite inferior e 9327.87 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o número total de horas por semana que os moradores do condado assistem televisão.

T Value = 1.72075. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição t-student. Neste caso, se o tamanho da amostra fosse maior ou igual à 30, em todos os estratos, ao invés de **T Value** teríamos **Z Value** que corresponde ao ponto crítico baseado em uma distribuição normal padronizada.



Exemplo de Amostragem Estratificada

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo [a amostragem estratificada](#) com [amostragem aleatória simples sem reposição](#) e [estimador razão](#).

Um estudo nutricional está sendo realizado com 180 coelhos, 60 da raça A, 80 da raça B e 40 da raça C. Na 1ª etapa, cada coelho teve seu peso avaliado e a média obtida foi: 3,2 Kg raça A; 3,5 Kg raça B e 2,5 Kg raça C. Após dois meses de tratamento a base de uma nova ração, o pesquisador está interessado em estimar o peso médio dos coelhos. Considerando cada raça como um estrato, ele decide utilizar a amostragem estratificada selecionando pelo procedimento de amostragem aleatória simples sem reposição, 15 coelhos da raça A, 20 da raça B e 10 da raça C. Estime o peso médio dos coelhos observando o procedimento adotado na coleta dos dados e utilizando o estimador razão.

1. Abra o arquivo EX_AEER.MTW.

Observe a estrutura dos dados, as três primeiras colunas contêm os valores da variável auxiliar para cada estrato, as três colunas subseqüentes os valores da variável resposta para cada estrato e a sétima coluna apresenta o número de elementos na população e o número populacional de elementos nos estratos 1, 2 e 3; respectivamente.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

```
3. Session:   MTB > %Sampling;
              SUBC > Estrato C4-C6;
              SUBC > Auxiliares C1-C3;
              SUBC > Tamanho C7.
```

4. A seguir o programa deverá fazer uma série de [perguntas](#), entre elas:

Digite o método de amostragem utilizado dentro dos estratos:

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

Digite a opção 2

Escolha os tipos de estimadores:

- 1) Usuais
- 2) Razão
- 3) Regressão

Digite a opção 2

Existe Pós-Estratificação?

Digite NO

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

Deseja entrar com o(s) valor(es) da Média da Variável Auxiliar?

Digite YES

Digite o(s) valor(es) de Mx:

Digite 3.2 Pressione a tecla Enter (refere-se ao 1º estrato)
3.5 Pressione a tecla Enter (refere-se ao 2º estrato)
2.5 Pressione a tecla Enter (refere-se ao 3º estrato)

5. Impressão dos Resultados

Estimação Global

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

Método Separado

N	Mean	StDev	I.C.
180	4.1392	0.0486880	(4.05723, 4.22120)
	Total	StDev	I.C.
	745.0591	8.7638377	(730.302, 759.816)

Método Combinado

N	Mean	StDev	I.C.
180	4.1375	0.0494995	(4.05416, 4.22086)
	Total	StDev	I.C.
	744.7511	8.9099176	(729.748, 759.754)

. T Value : 1.68385

OBS: Nível de significância igual à 0.05. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem global e dentro dos estratos e ao estimador selecionado. No exemplo acima, os moradores foram selecionados por [amostragem estratificada](#) com [amostragem aleatória simples sem reposição](#) sendo o [estimador razão](#).

Método Separado

N = 180. Representa o tamanho da população. No exemplo, corresponde ao número de coelhos que fizeram parte do estudo nutricional.

Mean = 4.1392. Representa uma estimativa da média populacional da variável resposta. No exemplo, 4.1392 é uma estimativa do peso médio dos coelhos.

StDev = 0.0486880. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (4.05723, 4.22120). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 4.05723 representa o limite inferior e 4.22120 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o peso médio dos coelhos.

Total = 745.0591. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do peso total dos coelhos que fizeram parte do estudo nutricional.

StDev = 8.7638377. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (730.302, 759.816). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 730.302 representa o limite inferior e 759.816 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o peso total dos coelhos que fizeram parte do estudo nutricional.

Método Combinado

N = 180. Representa o tamanho da população. No exemplo, corresponde ao número de coelhos que fizeram parte do estudo nutricional.

Mean = 4.1375. Representa uma estimativa da média populacional da variável resposta. No exemplo, 4.1375 é uma estimativa do peso médio dos coelhos.

StDev = 0.0494995. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (4.05416, 4.22086). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 4.05416 representa o limite inferior e 4.22086 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o peso médio dos coelhos.

Total = 744.7511. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do peso total dos coelhos que fizeram parte do estudo nutricional.

StDev = 8.9099176. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (729.748, 759.754). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 729.748 representa o limite inferior e 759.754 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o peso total dos coelhos que fizeram parte do estudo nutricional.

T Value = 1.68385. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição t-student. Neste caso, se o tamanho da amostra fosse maior ou igual à 30, em todos os estratos, ao invés de **T Value** teríamos **Z Value** que corresponde ao ponto crítico baseado em uma distribuição normal padronizada.



Exemplo de Amostragem Estratificada

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo [a amostragem estratificada](#) com [amostragem aleatória simples sem reposição](#) e [estimador regressão](#).

Um estudo nutricional está sendo realizado com 180 coelhos, 60 da raça A, 80 da raça B e 40 da raça C. Na 1ª etapa, cada coelho teve seu peso avaliado e a média obtida foi: 3,2 Kg raça A; 3,5 Kg raça B e 2,5 Kg raça C. Após dois meses de tratamento a base de uma nova ração, o pesquisador está interessado em estimar o peso médio dos coelhos. Considerando cada raça como um estrato, ele decide utilizar a amostragem estratificada selecionando pelo procedimento de amostragem aleatória simples sem reposição, 15 coelhos da raça A, 20 da raça B e 10 da raça C. Estime o peso médio dos coelhos observando o procedimento adotado na coleta dos dados e utilizando o estimador razão.

1. Abra o arquivo EX_AEER.MTW.

Observe a estrutura dos dados, as três primeiras colunas contêm os valores da variável auxiliar para cada estrato, as três colunas subsequentes os valores da variável resposta para cada estrato e a sétima coluna apresentada o número de elementos na população e o número populacional de elementos nos estratos 1, 2 e 3; respectivamente.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

```
3. Session:  MTB > %Sampling;  
             SUBC > Estrato C4-C6;  
             SUBC > Auxiliares C1-C3;  
             SUBC > Tamanho C7.
```

4. A seguir o programa deverá fazer uma série de [perguntas](#), entre elas:

Digite o método de amostragem utilizado dentro dos estratos:

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

Digite a opção 2

Escolha os tipos de estimadores:

- 1) Usuais
- 2) Razão
- 3) Regressão

Digite a opção 3

Existe Pós-Estratificação?

Digite NO

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

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem global e dentro dos estratos e ao estimador selecionado. No exemplo acima, os moradores foram selecionados por [amostragem estratificada](#) com [amostragem aleatória simples sem reposição](#) sendo o [estimador regressão](#).

Método Separado

Coef. Angular. Representa uma estimativa da mudança ocasionada na variável resposta ao se acrescentar uma unidade a [variável auxiliar](#).

Estrato 1 = 0.7695

Estrato 2 = 1.3529

Estrato 3 = 1.1078

N = 180. Representa o tamanho da população. No exemplo, corresponde ao número de coelhos que fizeram parte do estudo nutricional.

Mean = 4.0951. Representa uma estimativa da média populacional da variável resposta. No exemplo, 4.0951 é uma estimativa do peso médio dos coelhos.

StDev = 0.0527310. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (4.00635, 4.18393). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 4.00635 representa o limite inferior e 4.18393 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o peso médio dos coelhos.

Total = 737.1251. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do peso total dos coelhos que fizeram parte do estudo nutricional.

StDev = 9.4915846. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (721.143, 753.108). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 721.143 representa o limite inferior e 753.108 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o peso total dos coelhos que fizeram parte do estudo nutricional.

Método Combinado

Coef. Angular Combinado = 0.9912. Representa uma estimativa da mudança ocasionada na variável resposta ao se acrescentar uma unidade a [variável auxiliar](#).

N = 180. Representa o tamanho da população. No exemplo, corresponde ao número de coelhos que fizeram parte do estudo nutricional.

Mean = 4.0404. Representa uma estimativa da média populacional da variável resposta. No exemplo, 4.0404 é uma estimativa do peso médio dos coelhos.

StDev = 0.0523715. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (3.95224, 4.12861). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 3.95224 representa o limite inferior e 4.12861 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o peso médio dos coelhos.

Total = 727.2761. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do peso total dos coelhos que fizeram parte do estudo nutricional.

StDev = 9.4268638. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (711.403, 743.150). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 711.403 representa o limite inferior e 743.150 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para o peso total dos coelhos que fizeram parte do estudo nutricional.

T Value = 1.68385. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição t-student. Neste caso, se o tamanho da amostra fosse maior ou igual à 30, em todos os estratos, ao invés de **T Value** teríamos **Z Value** que corresponde ao ponto crítico baseado em uma distribuição normal padronizada.



Exemplo de Amostragem Estratificada

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem estratificada](#) com [amostragem sistemática](#).

Um instituto de pesquisas de mercado foi contratado para estimar a quantidade de dinheiro gasta por empresas da região sudeste com investimentos em cultura no primeiro semestre do ano. Baseando-se na renda declarada no ano passado, 300 empresas foram alocadas em três categorias; 60 na A, 100 na B e 140 na C. Para cada uma delas foi montada uma lista em ordem alfabética e através do procedimento de amostragem sistemática foram selecionadas 20 empresas da categoria A, 24 da categoria B e 50 da C. As empresas sorteadas receberam a visita de um coordenador que coletou as informações necessárias.

1. Abra o arquivo EX_AEAS.MTW.

Observe a estrutura dos dados, as três primeiras colunas contêm os valores da variável auxiliar para cada estrato, as três colunas subseqüentes os valores da variável resposta para cada estrato e a sétima coluna apresentada o número de elementos na população e o número populacional de elementos nos estratos 1, 2 e 3; respectivamente.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

```
1. Session:   MTB > %Sampling;  
             SUBC > Estrato C1-C3;  
             SUBC > Tamanho C4.
```

4. A seguir o programa deverá fazer uma série de [perguntas](#), entre elas:

Digite o método de amostragem utilizado dentro dos estratos:

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

Digite a opção 3

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

Deseja correlação entre elementos da mesma amostra?

Digite YES

5. Impressão dos Resultados

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	Total
	300	26516.1892	7954856.7667
StDev Mean			
	Method 1	Method 2	Method 3
	1358.0726116	1588.5596829	1550.0621247
StDev Total			
	Method 1	Method 2	Method 3
	407421.7834704	476567.9048846	465018.6374227
IC Mean		IC Total	
	Method 1	Method 1	
	(24226.5, 28805.8)	(7267962, 8641752)	
	Method 2	Method 2	
	(23837.9, 29194.4)	(7151384, 8758329)	
	Method 3	Method 3	
	(23902.9, 29129.5)	(7170856, 8738858)	

Estrato	Correlation
-----	-----
1	-0.0508475
2	-0.0420875
3	-0.0201439

. T Value : 1.68596

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem global e dentro dos estratos e ao método de estimação da variância. No exemplo acima, as empresas foram selecionadas pelo procedimento de [amostragem estratificada](#) com [amostragem sistemática](#) sendo a variância estimada pelos métodos [1](#), [2](#) e [3](#).

N = 300. Representa o tamanho da população. No exemplo, corresponde ao número de empresas da região sudeste.

Mean = 26516.1892. Representa uma estimativa da média populacional da variável resposta. No exemplo, 26516.1892 é uma estimativa da quantidade média de dinheiro gasto pelas empresas da região sudeste com investimentos em cultura no primeiro semestre do ano.

StDev Mean

Method 1 = 1358.0726116. Representa uma estimativa do desvio padrão da média amostral da variável resposta calculado através do [método 1](#).

Method 2 = 1588.5596829. Representa uma estimativa do desvio padrão da média amostral da variável resposta calculado através do [método 2](#).

Method 3 = 1550.0621247. Representa uma estimativa do desvio padrão da média amostral da variável resposta calculado através do [método 3](#).

Total = 7954856.7667. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a quantidade total de dinheiro gasto pelas empresas da região sudeste com investimentos em cultura no primeiro semestre do ano.

StDev Total

Method 1 = 407421.7834704. Representa uma estimativa do desvio padrão do total amostral da variável resposta calculado através do [método 1](#).

Method 2 = 476567.9048846. Representa uma estimativa do desvio padrão do total amostral da variável resposta calculado através do [método 2](#).

Method 3 = 465018.6374227. Representa uma estimativa do desvio padrão do total amostral da variável resposta calculado através do [método 3](#).

I.C. Mean

Method 1 = (24226.5, 28805.8). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 24226.5 representa o limite inferior e 28805.8 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para a quantidade média de dinheiro gasto pelas empresas da região sudeste com investimentos em cultura no primeiro semestre do ano.

Method 2 = (23837.9, 29194.4). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 23837.9 representa o limite inferior e 29194.4 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para a quantidade média de dinheiro gasto pelas empresas da região sudeste com investimentos em cultura no primeiro semestre do ano.

Method 3 = (23902.9, 29129.5). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 23902.9 representa o limite inferior e 29129.5 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para a quantidade média de

dinheiro gasto pelas empresas da região sudeste com investimentos em cultura no primeiro semestre do ano.

I.C. Total

Method 1 = (7267962, 8641752). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 7267962 representa o limite inferior e 8641752 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para a quantidade total de dinheiro gasto pelas empresas da região sudeste com investimentos em cultura no primeiro semestre do ano.

Method 2 = (7151384, 8758329). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 7151384 representa o limite inferior e 8758329 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para a quantidade total de dinheiro gasto pelas empresas da região sudeste com investimentos em cultura no primeiro semestre do ano.

Method 3 = (7170856, 8738858). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 7170856 representa o limite inferior e 8738858 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para a quantidade total de dinheiro gasto pelas empresas da região sudeste com investimentos em cultura no primeiro semestre do ano.

Correlacion . Representa uma estimativa da [correlação intra-classe](#) das unidades populacionais.

Estrato 1 = -0.0508475

Estrato 2 = -0.0420875

Estrato 3 = -0.0201439

T Value = 1.68596. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição t-student. Neste caso, se o tamanho da amostra fosse maior ou igual à 30, em todos os estratos, ao invés de **T Value** teríamos **Z Value** que corresponde ao ponto crítico baseado em uma distribuição normal padronizada.

OBS: A estimação da variância populacional e por consequência a construção dos intervalos de confiança via método 2 considera somente partições em que o tamanho da amostra (n) é um múltiplo de $m=2$ ou $m=3$. Portanto, caso isto não ocorra, todas as estatísticas baseadas no método 2 não serão apresentadas.



O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem estratificada](#) com [amostragem por conglomerados em um estágio](#) e [estimador P.P.S.](#)

O gerente de vendas de um jornal está interessado em estimar o número médio de exemplares comprados por morador no último bimestre. A cidade está dividida em três regionais, a primeira contém 20 setores censitários, a segunda 25 e a terceira 10. Sabe-se que o número de jornais comprados é altamente correlacionado com o número de moradores do setor. Através do procedimento de [amostragem proporcional ao tamanho](#) é selecionada uma amostra de setores de cada uma das regionais e todos os seus moradores são entrevistados.

Regional	Número de setores censitários selecionados
1	4
2	6
3	3

1. Abra o arquivo EX1_AEC1E.MTW.

Observe a estrutura dos dados:

1ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do primeiro estrato. As 100 primeiras observações são referentes ao 1º conglomerado, as 80 subseqüentes ao 2º, as 80 subseqüentes ao 3º e as 90 restantes ao 4º.

2ª Coluna - contém o número populacional de conglomerados no primeiro estrato e o número populacional de elementos em cada um dos conglomerados selecionados.

3ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do segundo estrato. As 60 primeiras observações são referentes ao 1º conglomerado, as 110 subseqüentes ao 2º, as 90 subseqüentes ao 3º e assim por diante.

4ª Coluna - contém o número populacional de conglomerados no segundo estrato e o número populacional de elementos em cada um dos conglomerados selecionados.

5ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do terceiro estrato. As 65 primeiras observações são referentes ao 1º conglomerado, as 80 subseqüentes ao 2º e as 75 restantes ao 3º.

6ª Coluna - contém o número populacional de conglomerados no terceiro estrato e o número populacional de elementos em cada um dos conglomerados selecionados.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %**Sampling**;
 SUBC > **Estrato C1 C3 C5**;
 SUBC > **Conglomerados C2 C4 C6**.

4. A seguir o programa deverá fazer uma série de perguntas:

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 1

Deseja entrar com o tamanho médio populacional dos conglomerados?

Digite YES	Pressione a tecla Enter
200	Pressione a tecla Enter (refere-se ao 1º estrato)
230	Pressione a tecla Enter (refere-se ao 2º estrato)
110	Pressione a tecla Enter (refere-se ao 3º estrato)

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

5. Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Estratificada
. Tipo de Amostragem dentro dos Estratos: Conglomerados
. Número de Estágios : Um
. Estimadores : P.P.S.

. Número de Estratos Analisados : 3

N	Mean	StDev	I.C.
10850	12.6970	0.2734650	(12.2472, 13.1468)
n	Total	StDev	I.C.
1150	137762.1013	2967.0952087	(132882, 142643)

. Z Value : 1.64485

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem global e dentro dos estratos, ao número de estágios, ao estimador selecionado e ao número de estratos analisados.

No exemplo acima, os moradores foram selecionados pelo procedimento de [amostragem estratificada](#) com [amostragem por conglomerados em um estágio](#) sendo o [estimador P.P.S.](#) Além disso, três estratos foram analisados.

N = 10850. Representa o número de elementos na população. No exemplo, corresponde ao número de habitantes da cidade.

Mean = 12.6970. Representa uma estimativa da média populacional da variável resposta. No exemplo, 12.6970 é uma estimativa do número médio de jornais comprados por morador no último bimestre.

StDev = 0.2734650. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (12.2472, 13.1468). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 12.2472 representa o limite inferior e 13.1468 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número médio de jornais comprados por morador no último bimestre.

n = 1150. Representa o número total de elementos amostrados. No exemplo, corresponde ao número de habitantes da cidade que foram entrevistados.

Total = 137762.1013. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do número total de jornais vendidos na cidade no último bimestre.

StDev = 2967.0952087. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (132882, 142643). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 132882 representa o limite inferior e 142643 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número total de jornais vendidos na cidade no último bimestre.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30, em todos os estratos, ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



Exemplo de Amostragem Estratificada

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem estratificada](#) com [amostragem por conglomerados em um estágio](#) e [estimador razão](#).

O gerente de vendas de um jornal está interessado em estimar o número médio de exemplares comprados por morador no último bimestre. A cidade está dividida em três regionais, a primeira contém 20 setores censitários, a segunda 25 e a terceira 10. Através do procedimento de amostragem aleatória simples sem reposição é selecionada uma amostra de setores de cada uma das regionais e todos os seus moradores são entrevistados.

Regional	Número de setores censitários selecionados
1	4
2	6
3	3

1. Abra o arquivo EX1_AEC1E.MTW.

Observe a estrutura dos dados:

- 1ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do primeiro estrato. As 100 primeiras observações são referentes ao 1º conglomerado, as 80 subseqüentes ao 2º, as 80 subseqüentes ao 3º e as 90 restantes ao 4º.
- 2ª Coluna - contém o número populacional de conglomerados no primeiro estrato e o número populacional de elementos em cada um dos conglomerados selecionados.
- 3ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do segundo estrato. As 60 primeiras observações são referentes ao 1º conglomerado, as 110 subseqüentes ao 2º, as 90 subseqüentes ao 3º e assim por diante.
- 4ª Coluna - contém o número populacional de conglomerados no segundo estrato e o número populacional de elementos em cada um dos conglomerados selecionados.
- 5ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do terceiro estrato. As 65 primeiras observações são referentes ao 1º conglomerado, as 80 subseqüentes ao 2º e as 75 restantes ao 3º.
- 6ª Coluna - contém o número populacional de conglomerados no terceiro estrato e o número populacional de elementos em cada um dos conglomerados selecionados.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %Sampling;
 SUBC > Estrato C1 C3 C5;
 SUBC > Conglomerados C2 C4 C6.

4. A seguir o programa deverá fazer uma série de perguntas:

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
200	Pressione a tecla Enter (refere-se ao 1º estrato)
230	Pressione a tecla Enter (refere-se ao 2º estrato)
110	Pressione a tecla Enter (refere-se ao 3º estrato)

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

5. Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Estratificada
. Tipo de Amostragem dentro dos Estratos: Conglomerados
. Número de Estágios : Um
. Estimadores : Razão
. Número de Estratos Analisados : 3

N	Mean	StDev	I.C.
10850	12.7044	0.1218289	(12.5040, 12.9048)

n	Total	StDev	I.C.
1150	137843.1281	1321.8430703	(135669, 140017)

. Z Value : 1.64485

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem global e dentro dos estratos, ao número de estágios, ao estimador selecionado e ao número de estratos analisados.

No exemplo acima, os moradores foram selecionados pelo procedimento de [amostragem estratificada](#) com [amostragem por conglomerados em um estágio](#) sendo o [estimador razão](#). Além disso, três estratos foram analisados.

N = 10850. Representa o número de elementos na população. No exemplo, corresponde ao número de habitantes da cidade.

Mean = 12.7044. Representa uma estimativa da média populacional da variável resposta. No exemplo, 12.7044 é uma estimativa do número médio de jornais comprados por morador no último bimestre.

StDev = 0.1218289. Representa uma estimativa do desvio padrão da média amostral variável resposta.

I.C. = (12.5040, 12.9048). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 12.5040 representa o limite inferior e 12.9048 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número médio de jornais comprados por morador no último bimestre.

n = 1150. Representa o número total de elementos amostrados. No exemplo, corresponde ao número de habitantes da cidade que foram entrevistados.

Total = 137843.1281. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do número total de jornais vendidos na cidade no último bimestre.

StDev = 1321.8430703. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (135669, 140017). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 135669 representa o limite inferior e 140017 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número total de jornais vendidos na cidade no último bimestre.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30, em todos os estratos, ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



Exemplo de Amostragem Estratificada

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em um estágio](#) e [estimador não viciado - unidades de mesmo tamanho](#).

O gerente de vendas de um jornal está interessado em estimar o número médio de exemplares comprados por morador no último bimestre. A cidade está dividida em três regionais e estas subdivididas em setores censitários com o mesmo número de moradores. Através do procedimento de amostragem aleatória simples sem reposição é selecionada uma amostra de setores de cada uma das regionais e todos os seus moradores são entrevistados.

Regional	Número de setores censitários	Número de moradores em cada setor censitário	Número de setores censitários selecionados
1	20	200	4
2	25	230	6
3	10	110	3

1. Abra o arquivo EX2_AEC1E.MTW.

Observe a estrutura dos dados:

- 1ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do primeiro estrato.
- 2ª Coluna - contém o número populacional e amostral de conglomerados no primeiro estrato e o número populacional de elementos em cada um dos conglomerados selecionados.
- 3ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do segundo estrato.
- 4ª Coluna - contém o número populacional e amostral de conglomerados no segundo estrato e o número populacional de elementos em cada um dos conglomerados selecionados.
- 5ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do terceiro estrato.
- 6ª Coluna - contém o número populacional e amostral de conglomerados no terceiro estrato e o número populacional de elementos em cada um dos conglomerados selecionados.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %Sampling;
 SUBC > Estrato C1 C3 C5;
 SUBC > Conglomerados C2 C4 C6.

4. A seguir o programa deverá fazer uma série de perguntas:

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 3

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

5. Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Estratificada
. Tipo de Amostragem dentro dos Estratos: Conglomerados
. Número de Estágios : Um
. Estimadores : Não Viciado
Unidades de Mesmo Tamanho

. Número de Estratos Analisados : 3

N	Mean	StDev	I.C.
10850	19.5542	0.2183551	(19.1951, 19.9134)

n	Total	StDev	I.C.
2510	212163.3333	2369.1526849	(208266, 216060)

. Z Value : 1.64485

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem global e dentro dos estratos, ao número de estágios, ao estimador selecionado e ao número de estratos analisados.

No exemplo acima, os moradores foram selecionados pelo procedimento de [amostragem estratificada](#) com [amostragem por conglomerados em um estágio](#) sendo o [estimador não viciado - unidades de mesmo tamanho](#). Além disso, três estratos foram analisados.

N = 10850. Representa o número de elementos na população. No exemplo, corresponde ao número de habitantes da cidade.

Mean = 19.5542. Representa uma estimativa da média populacional da variável resposta. No exemplo, é uma estimativa do número médio de jornais comprados por morador no último bimestre.

StDev = 0.2183551. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (19.1951, 19.9134). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 19.1951 representa o limite inferior e 19.9134 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número médio de jornais comprados por morador no último bimestre.

n = 2510. Representa o número total de elementos amostrados. No exemplo, corresponde ao número de habitantes da cidade que foram entrevistados.

Total = 212163.3333. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do número total de jornais vendidos na cidade no último bimestre.

StDev = 2369.1526849. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (208266, 216060). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 208266 representa o limite inferior e 216060 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número total de jornais vendidos na cidade no último bimestre.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30, em todos os estratos, ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



Exemplo de Amostragem Estratificada

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em um estágio](#) e [estimador não viciado - unidades de tamanho diferente](#).

O gerente de vendas de um jornal está interessado em estimar o número médio de exemplares comprados por morador no último bimestre. A cidade está dividida em três regionais, a primeira contém 20 setores censitários, a segunda 25 e a terceira 10. Através do procedimento de amostragem aleatória simples sem reposição é selecionada uma amostra de setores de cada uma das regionais e todos os seus moradores são entrevistados.

Regional	Número de setores censitários selecionados
1	4
2	6
3	3

1. Abra o arquivo EX1_AEC1E.MTW.

Observe a estrutura dos dados:

- 1ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do primeiro estrato. As 100 primeiras observações são referentes ao 1º conglomerado, as 80 subseqüentes ao 2º, as 80 subseqüentes ao 3º e as 90 restantes ao 4º.
- 2ª Coluna - contém o número populacional de conglomerados no primeiro estrato e o número populacional de elementos em cada um dos conglomerados selecionados.
- 3ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do segundo estrato. As 60 primeiras observações são referentes ao 1º conglomerado, as 110 subseqüentes ao 2º, as 90 subseqüentes ao 3º e assim por diante.
- 4ª Coluna - contém o número populacional de conglomerados no segundo estrato e o número populacional de elementos em cada um dos conglomerados selecionados.
- 5ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do terceiro estrato. As 65 primeiras observações são referentes ao 1º conglomerado, as 80 subseqüentes ao 2º e as 75 restantes ao 3º.
- 6ª Coluna - contém o número populacional de conglomerados no terceiro estrato e o número populacional de elementos em cada um dos conglomerados selecionados.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %**Sampling**;
 SUBC > **Estrato C1 C3 C5**;
 SUBC > **Conglomerados C2 C4 C6**.

4. A seguir o programa deverá fazer uma série de perguntas:

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 4

Deseja entrar com o tamanho médio populacional dos conglomerados?

Digite YES	Pressione a tecla Enter
200	Pressione a tecla Enter (refere-se ao 1º estrato)
230	Pressione a tecla Enter (refere-se ao 2º estrato)
110	Pressione a tecla Enter (refere-se ao 3º estrato)

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

5. Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Estratificada
. Tipo de Amostragem dentro dos Estratos: Conglomerados
. Número de Estágios : Um
. Estimadores : Não Viciado
Unidades de Tamanho Diferente

. Número de Estratos Analisados : 3

N	Mean	StDev	I.C.
10850	5.6121	0.3409291	(5.05136, 6.17291)

n	Total	StDev	I.C.
1150	60891.6667	3699.0808168	(54807.2, 66976.1)

. Z Value : 1.64485

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem global e dentro dos estratos, ao número de estágios, ao estimador selecionado e ao número de estratos analisados.

No exemplo acima, os moradores foram selecionados pelo procedimento de [amostragem estratificada](#) com [amostragem por conglomerados em um estágio](#) sendo o [estimador razão](#). Além disso, três estratos foram analisados.

N = 10850. Representa o número de elementos na população. No exemplo, corresponde ao número de habitantes da cidade.

Mean = 5.6121. Representa uma estimativa da média populacional da variável resposta. No exemplo, é uma estimativa do número médio de jornais comprados por morador no último bimestre.

StDev = 0.3409291. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (5.05136, 6.17291). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 5.05136 representa o limite inferior e 6.17291 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número médio de jornais comprados por morador no último bimestre.

n = 1150. Representa o número total de elementos amostrados. No exemplo, corresponde ao número de habitantes da cidade que foram entrevistados.

Total = 60891.6667. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa do número total de jornais vendidos na cidade no último bimestre.

StDev = 3699.0808168. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (54807.2, 66976.1). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 54807.2 representa o limite inferior e 66976.1 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para o número total de jornais vendidos na cidade no último bimestre.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30, em todos os estratos, ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



Exemplo de Amostragem Estratificada

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem estratificada](#) com [amostragem por conglomerados em dois estágios](#) e [estimador P.P.S.](#)

A prefeitura de uma certa cidade está interessada em avaliar o nível de conhecimento geral dos alunos matriculados nas escolas do Sistema Municipal de Ensino. Com esta finalidade, o estatístico responsável pelo estudo decide dividir a cidade em 3 regiões sócio-econômicas e retirar, através do procedimento de amostragem com probabilidade proporcional ao tamanho, uma amostra de escolas de cada região. A partir das escolas selecionadas sorteou-se um conjunto de salas de aula, sendo aplicado um teste a todos os seus alunos.

Região	Número de escolas	Número de escolas selecionadas
1	30	8
2	10	3
3	50	11

Sabe-se que o número médio de salas de aula nas escolas das regiões 1, 2 e 3 é 15, 20 e 10, respectivamente.

1. Abra o arquivo EX1_AEC2E.MTW.

Observe a estrutura dos dados:

- 1ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do primeiro estrato, ou seja, as notas médias, por turma, dos alunos pertencentes as escolas da região 1. As 6 primeiras observações são referentes ao 1º conglomerado (escola), as 8 subseqüentes ao 2º, as 5 subseqüentes ao 3º e assim por diante.
- 2ª Coluna - contém o número populacional de conglomerados (escolas) no primeiro estrato (região) e o número amostral de elementos (salas) em cada um dos conglomerados selecionados.
- 3ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do segundo estrato, ou seja, as notas médias, por turma, dos alunos pertencentes as escolas da região 2. As 7 primeiras observações são referentes ao 1º conglomerado (escola) e as 8 subseqüentes ao 2º.
- 4ª Coluna - contém o número populacional de conglomerados (escolas) no segundo estrato (região) e o número amostral de elementos (salas) em cada um dos conglomerados selecionados.
- 5ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do terceiro estrato, ou seja, as notas médias, por turma, dos alunos pertencentes as escolas da região 3. As 5 primeiras observações são referentes ao 1º conglomerado (escola), as 6 subseqüentes ao 2º, as 4 subseqüentes ao 3º e assim por diante.
- 6ª Coluna - contém o número populacional de conglomerados (escolas) no terceiro estrato (região) e o número amostral de elementos (salas) em cada um dos conglomerados selecionados.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).
3. Session: MTB > %**Sampling**;
 SUBC > **Estrato C1 C3 C5**;
 SUBC > **Conglomerados C2 C4 C6**.
4. A seguir o programa deverá fazer uma série de perguntas:

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
15	Pressione a tecla Enter (refere-se ao 1º estrato)
20	Pressione a tecla Enter (refere-se ao 2º estrato)
10	Pressione a tecla Enter (refere-se ao 3º estrato)

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

5. Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Estratificada
. Tipo de Amostragem dentro dos Estratos: Conglomerados
. Número de Estágios : Dois
. Estimadores : P.P.S.

. Número de Estratos Analisados : 3

N	Mean	StDev	I.C.
1150	49.5092	3.1246170	(44.3333, 54.6851)
n	Total	StDev	I.C.
135	56935.5833	3593.3095529	(50983.3, 62887.8)

. T Value : 1.65648

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem global e dentro dos estratos, ao número de estágios, ao estimador selecionado e ao número de estratos analisados. No exemplo acima, os alunos foram selecionados pelo procedimento de [amostragem estratificada](#) com [amostragem por conglomerados em dois estágios](#) sendo o [estimador P.P.S.](#) Além disso, três estratos foram analisados.

N = 1150. Representa o número de elementos do segundo estágio na população. No exemplo, corresponde ao número de salas de aula na cidade.

Mean = 49.5092. Representa uma estimativa da média populacional da variável resposta. No exemplo, é uma estimativa da nota média dos alunos por sala de aula.

StDev = 3.1246170. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (44.3333, 54.6851). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 44.3333 representa o limite inferior e 54.6851 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para a nota média dos alunos por sala de aula.

n = 135. Representa o número total de elementos do segundo estágio amostrados. No exemplo, corresponde ao número de salas de aula da cidade que foram amostradas.

Total = 56935.5833. Representa uma estimativa do total populacional da variável resposta.

StDev = 3593.3095529. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (50983.3, 62887.8). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 50983.3 representa o limite inferior e 62887.8 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student.

T Value = 1.65648. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição t-student. Neste caso, se o tamanho da amostra fosse maior ou igual à 30, em todos os estratos, ao invés de **T Value** teríamos **Z Value** que corresponde ao ponto crítico baseado em uma distribuição normal padronizada.



Exemplo de Amostragem Estratificada

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem estratificada](#) com [amostragem por conglomerados em dois estágios](#) e [estimador razão](#).

A prefeitura de uma certa cidade está interessada em avaliar o nível de conhecimento geral dos alunos matriculados nas escolas do Sistema Municipal de Ensino. Com esta finalidade, o estatístico responsável pelo estudo decide dividir a cidade em 3 regiões sócio-econômicas e retirar, através do procedimento de amostragem aleatória simples, uma amostra de escolas de cada região. A partir das escolas selecionadas sorteou-se um conjunto de salas de aula, sendo aplicado um teste a todos os seus alunos.

Região	Número de escolas	Número de escolas selecionadas
1	30	8
2	10	3
3	50	11

Sabe-se que o número médio de salas de aula nas escolas das regiões 1, 2 e 3 é 15, 20 e 10, respectivamente.

1. Abra o arquivo EX1_AEC2E.MTW.

Observe a estrutura dos dados:

- 1ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do primeiro estrato, ou seja, as notas médias, por turma, dos alunos pertencentes as escolas da região 1. As 6 primeiras observações são referentes ao 1º conglomerado (escola), as 8 subseqüentes ao 2º, as 5 subseqüentes ao 3º e assim por diante.
- 2ª Coluna - contém o número populacional de conglomerados (escolas) no primeiro estrato (região) e o número amostral de elementos (salas) em cada um dos conglomerados selecionados.
- 3ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do segundo estrato, ou seja, as notas médias, por turma, dos alunos pertencentes as escolas da região 2. As 7 primeiras observações são referentes ao 1º conglomerado (escola) e as 8 subseqüentes ao 2º.
- 4ª Coluna - contém o número populacional de conglomerados (escolas) no segundo estrato (região) e o número amostral de elementos (salas) em cada um dos conglomerados selecionados.
- 5ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do terceiro estrato, ou seja, as notas médias, por turma, dos alunos pertencentes as escolas da região 3. As 5 primeiras observações são referentes ao 1º conglomerado (escola), as 6 subseqüentes ao 2º, as 4 subseqüentes ao 3º e assim por diante.
- 6ª Coluna - contém o número populacional de conglomerados (escolas) no terceiro estrato (região) e o número amostral de elementos (salas) em cada um dos conglomerados selecionados.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).
3. Session: MTB > **%Sampling;**
SUBC > **Estrato C1 C3 C5;**
SUBC > **Conglomerados C2 C4 C6.**
4. A seguir o programa deverá fazer uma série de perguntas:

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 2

Deseja entrar com o tamanho médio populacional dos conglomerados?

Digite YES	Pressione a tecla Enter
15	Pressione a tecla Enter (refere-se ao 1º estrato)
20	Pressione a tecla Enter (refere-se ao 2º estrato)
10	Pressione a tecla Enter (refere-se ao 3º estrato)

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

5. Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Estratificada
. Tipo de Amostragem dentro dos Estratos: Conglomerados
. Número de Estágios : Dois
. Estimadores : Razão

. Número de Estratos Analisados : 3

N	Mean	StDev	I.C.
1150	48.4510	1.2327308	(46.4090, 50.4930)
n	Total	StDev	I.C.
135	55718.6901	1417.6403954	(53370.4, 58067.0)

. T Value : 1.65648

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem global e dentro dos estratos, ao número de estágios, ao estimador selecionado e ao número de estratos analisados. No exemplo acima, os alunos foram selecionados pelo procedimento de [amostragem estratificada](#) com [amostragem por conglomerados em dois estágios](#) sendo o [estimador razão](#). Além disso, três estratos foram analisados.

N = 1150. Representa o número de elementos do segundo estágio na população. No exemplo, corresponde ao número de salas de aula na cidade.

Mean = 48.4510. Representa uma estimativa da média populacional da variável resposta. No exemplo, é uma estimativa da nota média dos alunos por sala de aula.

StDev = 1.2327308. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (46.4090, 50.4930). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 46.4090 representa o limite inferior e 50.4930 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para a nota média dos alunos por sala de aula.

n = 135. Representa o número total de elementos do segundo estágio amostrados. No exemplo, corresponde ao número de salas de aula da cidade que foram amostradas.

Total = 55718.6901. Representa uma estimativa do total populacional da variável resposta.

StDev = 1417.6403954. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (53370.4, 58067.0). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 53370.4 representa o limite inferior e 58067.0 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student.

T Value = 1.65648. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição t-student. Neste caso, se o tamanho da amostra fosse maior ou igual à 30, em todos os estratos, ao invés de **T Value** teríamos **Z Value** que corresponde ao ponto crítico baseado em uma distribuição normal padronizada.



Exemplo de Amostragem Estratificada

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em dois estágios](#) e [estimador não viciado - unidades de mesmo tamanho](#).

A prefeitura de uma certa cidade está interessada em avaliar o nível de conhecimento geral dos alunos matriculados nas escolas do Sistema Municipal de Ensino. Com esta finalidade, o estatístico responsável pelo estudo decide dividir a cidade em 3 regiões sócio-econômicas e retirar, através do procedimento de amostragem aleatória simples, uma amostra de escolas de cada região. A partir das escolas selecionadas sorteou-se um conjunto de salas de aula, sendo aplicado um teste a todos os seus alunos.

Região	Número de escolas	No. de escolas selecionadas	No. de salas em cada escola	No. de salas selecionadas em cada escola
1	30	8	10	5
2	10	3	15	7
3	50	11	8	3

1. Abra o arquivo EX2_AEC2E.MTW.

Observe a estrutura dos dados:

- 1ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do primeiro estrato, ou seja, as notas médias, por turma, dos alunos pertencentes as escolas da região 1. As 5 primeiras observações são referentes ao 1º conglomerado (escola), as 5 subseqüentes ao 2º, as 5 subseqüentes ao 3º e assim por diante.
- 2ª Coluna - contém o número populacional e amostral de conglomerados (escolas) no primeiro estrato (região) e o número populacional e amostral de elementos (salas) em cada um dos conglomerados.
- 3ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do segundo estrato, ou seja, as notas médias, por turma, dos alunos pertencentes as escolas da região 2. As 3 primeiras observações são referentes ao 1º conglomerado (escola), as 3 subseqüentes ao 2º, as 3 subseqüentes ao 3º e assim por diante.
- 4ª Coluna - contém o número populacional e amostral de conglomerados (escolas) no segundo estrato (região) e o número populacional e amostral de elementos (salas) em cada um dos conglomerados.
- 5ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do terceiro estrato, ou seja, as notas médias, por turma, dos alunos pertencentes as escolas da região 3. As 5 primeiras observações são referentes ao 1º conglomerado (escola), as 5 subseqüentes ao 2º, as 5 subseqüentes ao 3º e assim por diante.
- 6ª Coluna - contém o número populacional e amostral de conglomerados (escolas) no terceiro estrato (região) e o número populacional e amostral de elementos (salas) em cada um dos conglomerados.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).
3. Session: MTB > %**Sampling**;
SUBC > **Estrato C1 C3 C5**;
SUBC > **Conglomerados C2 C4 C6**.
4. A seguir o programa deverá fazer uma série de perguntas:

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 3

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

5. Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Estratificada
. Tipo de Amostragem dentro dos Estratos: Conglomerados
. Número de Estágios : Dois
. Estimadores : Não Viciado
Unidades de Mesmo Tamanho

. Número de Estratos Analisados : 3

N	Mean	StDev	I.C.
850	48.1536	2.0972890	(44.6795, 51.6277)
n	Total	StDev	I.C.
135	40930.5425	1782.6956137	(37977.5, 43883.5)

. T Value : 1.65648

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem global e dentro dos estratos, ao número de estágios, ao estimador selecionado e ao número de estratos analisados. No exemplo acima, os alunos foram selecionados pelo procedimento de [amostragem estratificada](#) com [amostragem por conglomerados em dois estágios](#) sendo o [estimador não viciado – unidades de mesmo tamanho](#). Além disso, três estratos foram analisados.

N = 850. Representa o número de elementos do segundo estágio na população. No exemplo, corresponde ao número de salas de aula na cidade.

Mean = 48.1536. Representa uma estimativa da média populacional da variável resposta. No exemplo, é uma estimativa da nota média dos alunos por sala de aula.

StDev = 2.0972890. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (44.6795, 51.6277). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 44.6795 representa o limite inferior e 51.6277 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para a nota média dos alunos por sala de aula.

n = 135. Representa o número total de elementos do segundo estágio amostrados. No exemplo, corresponde ao número de salas de aula da cidade que foram amostradas.

Total = 40930.5425. Representa uma estimativa do total populacional da variável resposta.

StDev = 1782.6956137. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (37977.5, 43883.5). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 37977.5 representa o limite inferior e 43883.5 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student.

T Value = 1.65648. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição t-student. Neste caso, se o tamanho da amostra fosse maior ou igual à 30, em todos os estratos, ao invés de **T Value** teríamos **Z Value** que corresponde ao ponto crítico baseado em uma distribuição normal padronizada.



Exemplo de Amostragem Estratificada

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em dois estágios](#) e [estimador não viciado - unidades de tamanho diferente](#).

A prefeitura de uma certa cidade está interessada em avaliar o nível de conhecimento geral dos alunos matriculados nas escolas do Sistema Municipal de Ensino. Com esta finalidade, o estatístico responsável pelo estudo decide dividir a cidade em 3 regiões sócio-econômicas e retirar, através do procedimento de amostragem aleatória simples, uma amostra de escolas de cada região. A partir das escolas selecionadas sorteou-se um conjunto de salas de aula, sendo aplicado um teste a todos os seus alunos.

<i>Região</i>	<i>Número de escolas</i>	<i>Número de escolas selecionadas</i>
1	30	8
2	10	2
3	50	14

Sabe-se que o número médio de salas de aula nas escolas das regiões 1, 2 e 3 é 15, 20 e 10, respectivamente.

1. Abra o arquivo EX3_AEC2E.MTW.

Observe a estrutura dos dados:

- 1ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do primeiro estrato, ou seja, as notas médias, por turma, dos alunos pertencentes as escolas da região 1. As 6 primeiras observações são referentes ao 1º conglomerado (escola), as 8 subseqüentes ao 2º, as 5 subseqüentes ao 3º e assim por diante.
- 2ª Coluna - contém o número populacional de conglomerados (escolas) no primeiro estrato (região) e o número amostral de elementos (salas) em cada um dos conglomerados selecionados.
- 3ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do segundo estrato, ou seja, as notas médias, por turma, dos alunos pertencentes as escolas da região 2. As 7 primeiras observações são referentes ao 1º conglomerado (escola) e as 8 subseqüentes ao 2º.
- 4ª Coluna - contém o número populacional de conglomerados (escolas) no segundo estrato (região) e o número amostral de elementos (salas) em cada um dos conglomerados selecionados.
- 5ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do terceiro estrato, ou seja, as notas médias, por turma, dos alunos pertencentes as escolas da região 3. As 5 primeiras observações são referentes ao 1º conglomerado (escola), as 6 subseqüentes ao 2º, as 4 subseqüentes ao 3º e assim por diante.
- 6ª Coluna - contém o número populacional de conglomerados (escolas) no terceiro estrato (região) e o número amostral de elementos (salas) em cada um dos conglomerados selecionados.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).
3. Session: MTB > %Sampling;
SUBC > **Estrato C1 C3 C5;**
SUBC > **Conglomerados C2 C4 C6.**
4. A seguir o programa deverá fazer uma série de perguntas:

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 4

Deseja entrar com o tamanho médio populacional dos conglomerados?

Digite YES	Pressione a tecla Enter
15	Pressione a tecla Enter (refere-se ao 1º estrato)
20	Pressione a tecla Enter (refere-se ao 2º estrato)
10	Pressione a tecla Enter (refere-se ao 3º estrato)

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

5. Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Estratificada
. Tipo de Amostragem dentro dos Estratos: Conglomerados
. Número de Estágios : Dois
. Estimadores : Não Viciado
Unidades de Tamanho Diferente

. Número de Estratos Analisados : 3

N	Mean	StDev	I.C.
1150	47.1256	4.9141527	(38.9854, 55.2658)

n	Total	StDev	I.C.
135	54194.4452	5651.2755600	(44833.2, 63555.7)

. T Value : 1.65648

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem global e dentro dos estratos, ao número de estágios, ao estimador selecionado e ao número de estratos analisados. No exemplo acima, os alunos foram selecionados pelo procedimento de [amostragem estratificada](#) com [amostragem por conglomerados em dois estágios](#) sendo o [estimador não viciado – unidades de tamanho diferente](#). Além disso, três estratos foram analisados.

N = 1150. Representa o número de elementos do segundo estágio na população. No exemplo, corresponde ao número de salas de aula na cidade.

Mean = 47.1256. Representa uma estimativa da média populacional da variável resposta. No exemplo, é uma estimativa da nota média dos alunos por sala de aula.

StDev = 4.9141527. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (38.9854, 55.2658). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 38.9854 representa o limite inferior e 55.2658 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para a nota média dos alunos por sala de aula.

n = 135. Representa o número total de elementos do segundo estágio amostrados. No exemplo, corresponde ao número de salas de aula da cidade que foram amostradas.

Total = 54194.4452. Representa uma estimativa do total populacional da variável resposta.

StDev = 5651.2755600. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (44833.2, 63555.7). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 44833.2 representa o limite inferior e 63555.7 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student.

T Value = 1.65648. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição t-student. Neste caso, se o tamanho da amostra fosse maior ou igual à 30, em todos os estratos, ao invés de **T Value** teríamos **Z Value** que corresponde ao ponto crítico baseado em uma distribuição normal padronizada.



Exemplo de Amostragem Estratificada

O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em três estágios](#) e [estimador P.P.S.](#)

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 domicílio, 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.

1. Abra o arquivo EX1_AEC3E.MTW.

Observe a estrutura dos dados:

- 1ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do primeiro estrato, ou seja, a quantidade de dinheiro gasto com ligações telefônicas para cada domicílio pertencente a área A. As 4 primeiras observações representam o 1º quarteirão do 1º setor censitário da área A, as observações de ordem 47 a 52 representam o 1º quarteirão do 2º setor censitário da área A e assim por diante.
- 2ª Coluna - contém o número populacional e amostral de conglomerados (setores) no primeiro estrato (área), o número amostral de unidades do segundo estágio (quarteirões) e o número amostral de unidades do terceiro estágio (domicílios) em cada um dos conglomerados.
- 3ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do segundo estrato, ou seja, a quantidade de dinheiro gasto com ligações telefônicas para cada domicílio pertencente a área B. As 3 primeiras observações representam o 1º quarteirão do 1º setor censitário da área B, as observações de ordem 28 a 32 representam o 1º quarteirão do 2º setor censitário da área B e assim por diante.
- 4ª Coluna - contém o número populacional e amostral de conglomerados (setores) no segundo estrato (área), o número amostral de unidades do segundo estágio (quarteirões) e o número amostral de unidades do terceiro estágio (domicílios) em cada um dos conglomerados.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

```
3. Session:  MTB > %Sampling;  
            SUBC > Estrato C1 C3;  
            SUBC > Conglomerados C2 C4.
```

4. A seguir o programa deverá fazer uma série de perguntas:

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

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

Escolha uma das opções abaixo:

- 1) Entrar com os fatores de ponderação dos estratos
- 2) Entrar com o número de elementos da população para cada estrato

Digite 1 ou 2

Opção 2

Digite o número de elementos da população para cada estrato:

- | | |
|--------------------|--|
| Digite 2000 | Pressione a tecla Enter (refere-se ao número de unidades do 3º estágio no 1º estrato) |
| 1000 | Pressione a tecla Enter (refere-se ao número de unidades do 3º estágio no 2º estrato) |

5. Impressão dos Resultados

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

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem global e dentro dos estratos, ao número de estágios, ao estimador selecionado e ao número de estratos analisados. No exemplo acima, os alunos foram selecionados pelo procedimento de [amostragem estratificada](#) com [amostragem por conglomerados em três estágios](#) sendo o [estimador P.P.S.](#) Além disso, dois estratos foram analisados.

N = 3000. Representa o número de elementos do terceiro estágio na população. No exemplo, corresponde ao número de domicílios da região.

Mean = 294.8705. Representa uma estimativa da média populacional da variável resposta. No exemplo, é uma estimativa da quantidade média de dinheiro gasto com ligações telefônicas por domicílio.

StDev = 5.2180080. Representa uma estimativa do desvio padrão da média amostral da variável resposta.

I.C. = (286.288, 303.453). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 286.288 representa o limite inferior e 303.453 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para a quantidade média de dinheiro gasto com ligações telefônicas por domicílio.

n = 838. Representa o número total de elementos do terceiro estágio amostrados. No exemplo, corresponde ao número total de domicílios da região que foram amostrados.

Total = 884611.4353. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa da quantidade total de dinheiro gasto com ligações telefônicas na região.

StDev = 15654.0239955. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (858863, 910360). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 858863 representa o limite inferior e 910360 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para a quantidade total de dinheiro gasto com ligações telefônicas na região.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30 ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



O exemplo a seguir ilustra a utilização do **Sampling** em um problema envolvendo a [amostragem por conglomerados em três estágios](#) e [estimador não viciado - unidades de mesmo tamanho](#).

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 domicílio, o estatístico responsável pelo estudo decidiu adotar o seguinte procedimento:

- a região foi dividida em 2 áreas economicamente distintas, A e B;
- selecionou-se, de forma aleatória, 20 setores censitários na área A e 10 na área B;
- em cada setor, 5 quarteirões foram sorteados na área A e 3 na área B;
- em cada quarteirão, 2 domicílios foram sorteados em ambas as áreas para receberem a visita do pesquisador.

1. Abra o arquivo EX2_AEC3E.MTW.

Observe a estrutura dos dados:

1ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do primeiro estrato, ou seja, a quantidade de dinheiro gasto com ligações telefônicas para cada domicílio pertencente a área A. As 2 primeiras observações representam o 1º quarteirão do 1º setor censitário da área A, as observações de ordem 11 e 12 representam o 1º quarteirão do 2º setor censitário da área A e assim por diante.

2ª Coluna - contém o número populacional e amostral de conglomerados (setores) no primeiro estrato (área), o número populacional e amostral de unidades do segundo estágio (quarteirões) e o número populacional e amostral de unidades do terceiro estágio (domicílios) em cada um dos conglomerados.

3ª Coluna - contém os valores da variável resposta para cada um dos conglomerados selecionados a partir do segundo estrato, ou seja, a quantidade de dinheiro gasto com ligações telefônicas para cada domicílio pertencente a área B. As 2 primeiras observações representam o 1º quarteirão do 1º setor censitário da área B, as observações de ordem 6 e 7 representam o 1º quarteirão do 2º setor censitário da área B e assim por diante.

4ª Coluna - contém o número populacional e amostral de conglomerados (setores) no segundo estrato (área), o número populacional e amostral de unidades do segundo estágio (quarteirões) e o número populacional e amostral de unidades do terceiro estágio (domicílios) em cada um dos conglomerados.

2. Digite o comando **SWAP** (caso ele ainda não tenha sido digitado).

3. Session: MTB > %**Sampling**;
SUBC > **Estrato C1 C3**;
SUBC > **Conglomerados C2 C4**.

4. A seguir o programa deverá fazer uma série de perguntas:

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 2

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

5. Impressão dos Resultados

Estimação Global

. Tipo de Amostragem : Estratificada
. Tipo de Amostragem dentro dos Estratos: Conglomerados
. Número de Estágios : Três
. Estimadores : Não Viciado
Unidades de Mesmo Tamanho

. Número de Estratos Analisados : 2

N	Mean	StDev	I.C.
16750	113.8525	5.0039652	(105.622, 122.083)

n	Total	StDev	I.C.
260	1907030.1046	83816.4172794	(1769164, 2044896)

. Z Value : 1.64485

OBS: Nível de significância igual à 0.05.

6. Interpretação dos Resultados

Inicialmente o programa apresenta algumas informações referentes ao tipo de amostragem global e dentro dos estratos, ao número de estágios, ao estimador selecionado e ao número de estratos analisados. No exemplo acima, os alunos foram selecionados pelo procedimento de [amostragem estratificada](#) com [amostragem por conglomerados em três estágios](#) sendo o [estimador não viciado – unidades de mesmo tamanho](#). Além disso, dois estratos foram analisados.

N = 16750. Representa o número de elementos do terceiro estágio na população. No exemplo, corresponde ao número de domicílios da região.

Mean = 113.8525. Representa uma estimativa da média populacional da variável resposta. No exemplo, é uma estimativa da quantidade média de dinheiro gasto com ligações telefônicas por domicílio.

StDev = 5.0039652. Representa uma estimativa do desvio padrão da média populacional da variável resposta.

I.C. = (105.622, 122.083). Este é um [intervalo de confiança](#) de 95% para a média populacional da variável resposta baseado na distribuição normal ($n \geq 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 105.622 representa o limite inferior e 122.083 o limite superior de um intervalo de 95% de confiança, baseado na distribuição normal, para a quantidade média de dinheiro gasto com ligações telefônicas por domicílio.

n = 260. Representa o número total de elementos do terceiro estágio amostrados. No exemplo, corresponde ao número total de domicílios da região que foram amostrados.

Total = 1907030.1046. Representa uma estimativa do total populacional da variável resposta. No exemplo, corresponde a uma estimativa da quantidade total de dinheiro gasto com ligações telefônicas na região.

StDev = 83816.4172794. Representa uma estimativa do desvio padrão do total amostral da variável resposta.

I.C. = (1769164, 2044896). Este é um [intervalo de confiança](#) de 95% para o total populacional da variável resposta baseado na distribuição t-student ($n < 30$). O nível de significância α pode ser selecionado a partir de uma lista de opções, a qual sempre é fornecida. No exemplo, 1769164 representa o limite inferior e 2044896 o limite superior de um intervalo de 95% de confiança, baseado na distribuição t-student, para a quantidade total de dinheiro gasto com ligações telefônicas na região.

Z Value = 1.64485. Representa o ponto crítico que deixa uma probabilidade de 2,5% em cada uma das caudas de uma distribuição normal. Neste caso, se o tamanho da amostra fosse menor do que 30 ao invés de **Z Value** teríamos **T Value** que corresponde ao ponto crítico baseado em uma distribuição t-student.



Apesar dos exemplos aqui expostos não fazerem referência ao caso em que a variável resposta apresenta uma característica dicotômica, ou seja, uma observação é sempre classificada como pertencente ou não a categoria de interesse, o software **Sampling** está totalmente habilitado a trabalhá-lo. A forma pela qual os dados devem estar estruturados bem como os comandos que devem ser utilizados são exatamente os mesmos ilustrados pelas seções anteriores, não sendo, portanto, necessário maiores discussões.



Inicialmente, ao se executar o Sampling, uma série de perguntas são realizadas. Estas perguntas têm como objetivo abortar o programa caso seja detectado algum erro na especificação dos parâmetros, direcionar a análise e fornecer uma maior exploração dos dados.

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

Se uma ou mais colunas foram incorretamente especificadas na linha de comando, ou por algum outro motivo seja necessário parar a execução do programa, basta digitar no, de modo contrário, digite yes.

Deseja estimar o tamanho da amostra?

Se o objetivo é estimar o tamanho da amostra e/ou calcular alocações digite yes, caso contrário, digite no. Esta pergunta tem a finalidade de desviar o processamento diretamente para o módulo de [cálculo do tamanho da amostra](#) e não estará disponível para a amostragem por conglomerados.

Sendo a resposta afirmativa, as demais perguntas não serão realizadas.

Deseja estatística descritiva por estrato? Deseja análise gráfica por estrato?

Respondendo yes :

Digite (1) para análise de todos os estratos ou (2) para escolha dos estratos:

Se o plano amostral utilizado na coleta dos dados não foi a [amostragem estratificada](#), existe apenas um estrato, portanto, digite 1.

Se o plano amostral utilizado na coleta dos dados foi a [amostragem estratificada](#), digite 1 se todos os estratos devem ser analisados ou 2 para selecionar os estratos de interesse.

No caso da seleção de estratos:

Digite o número de estratos para análise:

Digite o número total de estratos que devem ser analisados.

Digite o número dos estratos para análise:

Digite o número dos estratos que deverão ser analisados (seguido da tecla enter) observe que cada estrato é referenciado pela ordem em que aparece na linha de comando inicial.

**EX: %Sampling;
 Estratos C2 C1 C3;
 Tamanho C4.**

Indica que o estrato número 1 está em C2, o número 2 em C1 e o número 3 em C3.

Deseja estatística descritiva global?

Deseja análise gráfica Global?

Responda **yes** ou **no**.

Se o plano amostral utilizado na coleta dos dados foi a [amostragem estratificada](#) esta opção permite que seja feita uma análise da situação geral, isto é, desconsiderando a existência dos estratos.

Caso o procedimento utilizado não seja a [amostragem estratificada](#) a análise global fornece exatamente os mesmos resultados da análise por estrato.

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

Se algum problema foi detectado digite **yes** para abortar o programa, caso contrário, digite **no** para continuar a análise.

OBS: (1) Todas as perguntas acima não estarão disponíveis quando o procedimento utilizado na coleta dos dados for a [amostragem por conglomerados](#).

(2) As perguntas realizadas com relação a análise descritiva da variável resposta também serão feitas para a variável auxiliar, caso ela exista.

(3) Se não existir variável auxiliar a estatística `corr` será substituída pelo símbolo “*”.



REFERÊNCIAS BIBLIOGRÁFICAS

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

MINITAB Inc. *User's guide 1: Data, Graphics, and Macros*, version 13.0, Pennsylvania, 2000.

MINITAB Inc. *User's guide 2: Data Analysis and Quality Tools*, version 13.0, Pennsylvania, 2000.

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

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.