



TRIBUNAL DE CONTAS DA UNIÃO
Secretaria-Geral de Controle Externo
Secretaria-Adjunta de Fiscalização

TÉCNICAS DE AMOSTRAGEM PARA AUDITORIAS

ADFIS/SEGECEX

2002

TRIBUNAL DE CONTAS DA UNIÃO

Negócio

Controle externo da administração pública e da gestão dos recursos públicos federais.

Missão

Assegurar a efetiva e regular gestão dos recursos públicos, em benefício da sociedade.

Visão

Ser instituição de excelência no controle e contribuir para o aperfeiçoamento da administração pública.

MINISTROS

Humberto Guimarães Souto, Presidente
Antonio Valmir Campelo Bezerra, Vice-Presidente
Marcos Vinicios Rodrigues Vilaça
Iram de Almeida Saraiva
Adylson Motta
Walton Alencar Rodrigues
Guilherme Gracindo Soares Palmeira
Ubiratan Diniz Aguiar
Benjamin Zymler

MINISTROS-SUBSTITUTOS

Lincoln Magalhães da Rocha
Augusto Sherman Cavalcanti
Marcos Bemquerer Costa

MINISTÉRIO PÚBLICO

Lucas Rocha Furtado, Procurador-Geral
Jatir Batista da Cunha, Subprocurador-Geral
Paulo Soares Bugarin, Subprocurador-Geral
Ubaldo Alves Caldas, Subprocurador-Geral
Maria Alzira Ferreira, Procuradora
Marinus Eduardo Vries Marsico, Procurador
Cristina Machado da Costa e Silva, Procuradora



TRIBUNAL DE CONTAS DA UNIÃO
Secretaria-Geral de Controle Externo
Secretaria-Adjunta de Fiscalização

TÉCNICAS DE AMOSTRAGEM PARA AUDITORIAS

Brasília, março de 2002.

Tribunal de Contas da União

Internet: <http://www.tcu.gov.br>

SAFS Q. 04 Lt. 01

70.042-900 – Brasília (DF)

Secretário-Geral de Controle Externo:

Luciano Carlos Batista

Secretário-Adjunto de Fiscalização:

Cláudio Souza Castello Branco

Analistas de Finanças e Controle Externo - Área de Controle Externo:

Carlos Alexandre Amorim Rocha	(Semag	–	Redação)
Fábio Henrique Granja e Barros	(Secob	–	Revisão)
Flávio Marcos Godoy Krecke	(Semag	–	Revisão)
Marcelo Cardoso Soares	(Seprog	–	Revisão)
Márcio Emmanuel Pacheco	(Secex-RJ	–	Revisão)
Paula de Biase Damasceno	(Secex-RJ	–	Revisão)

657.63
B823t

Brasil. Tribunal de Contas da União.
Técnicas de amostragem para
auditorias / Tribunal de Contas da União.
-- Brasília : TCU, Secretaria-Adjunta de
Fiscalização, 2002.
149 p.

1. Auditoria I. Título.

APRESENTAÇÃO

Este documento complementa orientações constantes do Manual de Auditoria de Natureza Operacional desta Corte. As técnicas ora tratadas procuram orientar as equipes de auditoria sobre os conceitos e procedimentos elementares requeridos pelas amostragens probabilísticas, bem como sobre os benefícios que elas podem proporcionar.

É nítida a importância do presente trabalho, pois já se observa uma demanda por técnicas de amostragem nas avaliações de programas públicos, nas fiscalizações de obras e na obtenção de subsídios para os pareceres anuais sobre as contas prestadas pelo Presidente da República.

No intuito de tornar mais claros e acessíveis os temas abordados, este documento contém vários exemplos e, quando possível, as rotinas do *Microsoft Excel* que permitem obter os resultados desejados.

Ademais, convém frisar que este trabalho não apenas descreve as técnicas de amostragem mais tradicionais (i.e., aleatória simples, estratificada e por conglomerados), empregadas em diversos campos do conhecimento, como também resume uma técnica específica para auditorias contábil-financeiras, qual seja: a amostragem por unidade monetária.

Dessa forma, acreditamos que o presente documento contribuirá para tornar mais efetiva a atuação do TCU, aprimorando a sua capacidade para apurar tempestivamente os resultados e as eventuais falhas sistemáticas presentes na administração pública federal.

Naturalmente, é de suma importância a apresentação de críticas e sugestões por todos que utilizarem este documento, pois somente isso permitirá o seu aperfeiçoamento. O tópico “Folha de Sugestões”, incluído no final desta brochura, explica como e a quem enviar quaisquer comentários.

Finalmente, na condição de Secretário-Geral de Controle Externo, parablenizo os dirigentes e servidores cujo esforço resultou na materialização do presente trabalho.

Luciano Carlos Batista
Secretário-Geral de Controle Externo

SUMÁRIO

LISTA DE FIGURAS.....	vii
LISTA DE TABELAS	ix
LISTA DE EXEMPLOS	x
INTRODUÇÃO	11
1. CONCEITOS BÁSICOS	15
1.1. Inferência Estatística.....	15
1.2. Principais Tipos de Amostragem Probabilística	16
1.2.1. Introdução à Amostragem Aleatória Simples.....	16
1.2.2. Introdução à Amostragem Aleatória Estratificada.....	17
1.2.3. Introdução à Amostragem Aleatória por Conglomerados.....	18
1.3. Plano Amostral.....	19
2. PROPRIEDADES DOS DADOS NUMÉRICOS	21
2.1. Medidas de Tendência Central.....	21
2.2. Medidas de Variação	23
2.3. Relação entre a Média e a Mediana	25
3. DISTRIBUIÇÕES DE PROBABILIDADE.....	27
3.1. Modelos de Distribuição Discreta.....	27
3.1.1. Distribuição Binomial	28
3.1.2. Distribuição Hipergeométrica.....	30
3.1.3. Distribuição Binomial Negativa	31
3.1.4. Distribuição de Poisson.....	31
3.2. Modelos de Distribuição Contínua.....	33
3.2.1. Distribuição Normal.....	34
3.2.2. Distribuição Exponencial	39
4. DISTRIBUIÇÕES AMOSTRAIS	41
4.1. Distribuição de Probabilidades das Médias Amostrais	41
4.2. Distribuição de Probabilidades das Proporções Amostrais	47
4.3. Amostragem em Populações Finitas	48
5. ESTIMAÇÃO	49
5.1. Estimativa do Intervalo da Média com σ Conhecido.....	50
5.2. Estimativa do Intervalo da Média com σ Desconhecido.....	52
5.3. Estimativa do tipo “ <i>Bootstrapping</i> ”	55
5.4. Estimativa do Intervalo para Proporções.....	55
5.5. Determinação do Tamanho da Amostra	57
5.5.1. Médias	57
5.5.2. Proporções	58
5.6. Amostragem em Populações Finitas	60
5.7. Estimativa do Intervalo para Totais.....	61
5.8. Estimativa do Intervalo para Diferenças	62

6. VARIANTES DA AMOSTRAGEM ALEATÓRIA SIMPLES.....	65
6.1. Amostragem Aleatória Estratificada	65
6.1.1. Estimação de Médias, Totais e Proporções na <i>AAE</i>	65
6.1.2. Determinação do Tamanho da Amostra na <i>AAE</i>	68
6.2. Amostragem Aleatória por Conglomerados	69
6.2.1. Estimação de Médias, Totais e Proporções na <i>AAC</i>	69
6.2.2. Determinação do Tamanho da Amostra na <i>AAC</i>	72
7. AMOSTRAGEM POR UNIDADE MONETÁRIA.....	73
7.1. Metodologia da Amostragem por Unidade Monetária.....	75
7.2. Aplicação da Amostragem por Unidade Monetária.....	80
8. TIPOS DE ERRO EM AMOSTRAGEM.....	83
8.1. Efeitos da Não-Resposta.....	84
8.2. Métodos de Redução da Não-Resposta	87
8.2.1. <i>Call-Backs</i>	88
8.2.2. Subamostras	89
8.3. Métodos de Compensação da Não-Resposta.....	90
ANEXO I: ROTINAS DO <i>MICROSOFT EXCEL</i>.....	93
ANEXO II: DISTRIBUIÇÕES NORMAL PADRÃO & "t" DE STUDENT.....	107
ANEXO III: BREVIÁRIO DE FÓRMULAS.....	117
ANEXO IV: FORMULÁRIOS PARA AMOSTRAGENS POR UNIDADE MONETÁRIA	127
ANEXO V: APLICAÇÃO DA AMOSTRAGEM POR UNIDADE MONETÁRIA	111
GLOSSÁRIO	137
BIBLIOGRAFIA	141
FOLHA DE SUGESTÕES.....	143

LISTA DE FIGURAS

1.1. Amostragem Aleatória Simples.....	16
1.2. Amostragem Aleatória Estratificada	17
1.3. Amostragem Aleatória por Conglomerados	18
2.1. Propriedades dos Dados Numéricos	21
2.2. Relação entre a Média, a Mediana e a Moda	25
3.1. Exemplo de Distribuição Discreta de Probabilidades	27
3.2. Distribuição Binomial com $n = 5$ e $p = 0,5$	29
3.3. Distribuição Binomial com $n = 5$ e $p = 0,1$	30
3.4. Distribuição de <i>Poisson</i> com $\lambda = 0,5$	32
3.5. Distribuição de <i>Poisson</i> com $\lambda = 6$	33
3.6. Função Densidade de Probabilidades	33
3.7. Probabilidade de Variáveis Aleatórias Contínuas.....	34
3.8. Distribuição Normal	35
3.9. Efeitos de Variações em σ e μ ($\sigma_A > \sigma_B$ e $\mu_C > \mu_B$)	36
3.10. Padronização da Distribuição Normal	37
3.11. Obtenção de Probabilidades [P(Z)] Dado um Valor Normal Padronizado (Z).....	38
3.12. Obtenção de um Valor Normal Padronizado (Z) Dada uma Probabilidade [P(Z)]	38
3.13. Distribuição Exponencial	40
4.1. Medida Não-Viesada	42
4.2. Medida Eficiente	42
4.3. Medida Consistente	43
4.4. Exemplo de Distribuição de Médias Amostrais.....	43
4.5. Exemplo do Impacto de n em Amostras Extraídas de Populações Distribuídas Normalmente.....	44
4.6. Teorema do Limite Central	46
4.7. Aplicação do Teorema do Limite Central.....	46
4.8. Distribuição de Proporções Amostrais	47
5.1. Elementos de uma Estimativa Probabilística.....	49
5.2. Intervalos de Confiança das Médias Amostrais	50

5.3. Distribuição de Probabilidades das Médias Amostrais Conhecendo-se o Desvio-Padrão Populacional	51
5.4. Distribuição de Probabilidades das Médias Amostrais Desconhecendo-se o Desvio-Padrão Populacional.....	53
5.5. Aplicação da Tabela “t” de Student.....	54
8.1. Tipos de Erro em Amostragem	84
A.1. Dados Hipotéticos.....	95
A.2. Geração de Estatísticas Descritivas	95
A.3. Estatísticas Descritivas Geradas	96
A.4. Agrupamento de Dados (Etapa 1 de 4).....	97
A.5. Agrupamento de Dados (Etapa 2 de 4).....	97
A.6. Agrupamento de Dados (Etapa 3 de 4).....	98
A.7. Agrupamento de Dados (Etapa 4 de 4).....	98
A.8. Dados Agrupados.....	98
A.9. Seleção de Amostras Aleatórias Simples com Reposição	100
A.10. Amostra Aleatória Simples Seleccionada.....	100
A.11. Estimação do Intervalo de Confiança para Médias com Desvio-Padrão Conhecido	101
A.12. Intervalo de Confiança Estimado.....	101

LISTA DE TABELAS

3.1. Aplicações da Distribuição Binomial	29
5.1. Tamanho da Amostra para Vários Erros Amostrais e Proporções de “Sucesso” ($Z = 1,96$)	59
6.1. Resumo dos Dados Coletados na Aplicação da Amostragem por Conglomerados	70
7.1. Índices de Confiabilidade para Várias Quantidades Previstas de Erros e Níveis de Confiança	73
7.2. Índice de Confiabilidade para Vários Erros Observados na Amostra.....	75
7.3. Seleção dos Elementos Amostrais.....	78
7.4. Fatores de Ampliação do Intervalo de Precisão (F_{AP}^j)	80
7.5. Diferenças Observadas na Amostra.....	81
8.1. Intervalos de Confiança para Várias Proporções de “Sucesso” e Índices de Não- Resposta ($Z = 1,96$ e $n = 1.000$).....	86
8.2. Tamanho da Amostra para Vários Erros Amostrais e Índices de Não-Resposta ($Z = 1,96$ e $p = 0,5$).....	87
8.3. Eficácia de Três Tentativas de Contato	88
A.1. Obtenção da Amplitude Semi-Interquartílica e da Meia Amplitude Semi- Interquartílica.....	99
A.2. Estimação de Intervalos de Confiança.....	102
A.3. Determinação do Tamanho da Amostra	103
A.4. Estimação do Intervalo de Confiança para Totais.....	104
A.5. Estimação do Intervalo de Confiança para Diferenças.....	105

LISTA DE EXEMPLOS

3.1. Aplicação da Distribuição de Poisson	32
3.2. Aplicação da Distribuição Exponencial.....	39
4.1. Aplicação da Distribuição de Probabilidades das Médias Amostrais	45
4.2. Aplicação da Distribuição de Probabilidades das Proporções Amostrais	48
5.1. Aplicação da Estimativa do Intervalo da Médias com σ Conhecido	52
5.2. Aplicação da Estimativa do Intervalo para Médias com σ Desconhecido	54
5.3. Aplicação da Estimativa do Intervalo para Proporções.....	56
5.5. Cálculo do Tamanho da Amostra para Estimar Duas Proporções	58
5.6. Cálculo do Tamanho da Amostra para Estimar n -Proporções.....	59
5.7. Cálculo Simplificado do Tamanho da Amostra para Estimar n -Proporções.....	60
5.8. Aplicação da Estimativa do Intervalo para Totais	62
5.9. Aplicação da Estimativa do Intervalo para Diferenças	63
6.1. Aplicação da Estimativa de Médias na Amostragem Estratificada	66
6.2. Aplicação da Estimativa de Totais na Amostragem Estratificada	67
6.3. Aplicação da Estimativa de Médias e Proporções na Amostragem por Conglomerados	70
7.1. Cálculo do Índice de Confiabilidade	74
7.2. Aplicação da Tabela de Índices de Confiabilidade	75
7.3. Cálculo do Tamanho da Amostra na <i>AUM</i>	77
7.4. Seleção dos Elementos Amostrais na <i>AUM</i>	77
8.1. Cálculo da Quantidade de Questionários.....	90

INTRODUÇÃO

É cada vez mais comum o uso de técnicas de amostragem pelos órgãos brasileiros de controle das finanças públicas, merecendo destaque os esforços da Secretaria Federal de Controle – SFC. A adoção dessas técnicas pelos órgãos em questão é uma decorrência do processo de reforma do Estado, no qual, sem prejuízo das várias formas de controle jurídico, observa-se uma demanda crescente por controles gerenciais, preocupados com a tempestiva apuração de resultados e com a identificação de falhas sistemáticas que, independentemente de envolverem ou não práticas ilícitas, possam requerer a adoção, em tempo hábil, de medidas corretivas de natureza político-administrativa.

No âmbito do TCU, há uma tendência em favor do uso dessas técnicas de amostragem em avaliações de programas públicos, em fiscalizações de obras e até mesmo na obtenção de subsídios para os pareceres anuais sobre as contas prestadas pelo Presidente da República. Dessa forma, o presente documento deve ser compreendido como uma complementação do Manual de Auditoria de Natureza Operacional.

Este trabalho tem como objetivo permitir que os AFCEs-CE tomem contato com os principais conceitos e métodos empregados pelas modernas técnicas de amostragem, para que possam estruturar melhor os trabalhos de campo cujo propósito seja determinar, p. ex., se os cadastros ou sistemas de controle das unidades jurisdicionadas são confiáveis, se as políticas públicas federais estão alcançando os resultados que lhe deram origem ou se as eventuais insuficiências de desempenho são aleatórias ou sistemáticas.

Isso não significa, entretanto, que este documento bastará para que os membros do corpo técnico do TCU se tornem especialistas em amostragem – uma disciplina por demais vasta e que, por ser eminentemente aplicada, precisa lidar com uma grande variedade de situações concretas. Na verdade, raramente os cadastros apresentam o mesmo grau de confiabilidade. É igualmente pouco comum que as populações apresentem o mesmo nível de dispersão ou o mesmo índice de não-respondentes recalcitrantes. Esses fatores geram diferenças entre os estimadores empregados pelos estudos amostrais, tornando-os, de certa forma, únicos.

Com este trabalho, pretende-se que os AFCEs-CE possam cumprir as tarefas mais simples e, ante situações mais complexas, formular questões que possam ser respondidas com objetividade por especialistas em atividade no próprio TCU ou em outras entidades, públicas ou privadas. Os interessados em se aprofundar um pouco mais nesse tema deverão recorrer a textos especializados, tais como as obras “Curso de Estatística”, “Estatística Aplicada à Administração” e, em especial, “*Elementary Survey Sampling*” (vide o tópico “Bibliografia”).

Ressalte-se que, para que este documento fosse o mais útil possível para os membros do corpo técnico, os cinco primeiros capítulos baseiam-se em rotinas do *Microsoft Excel*, pertencente à plataforma de *softwares* do TCU. A principal fonte para esses capítulos

foi a versão de 1997 da obra “*Statistics for Managers Using Microsoft Excel*”. Cabe registrar o nosso agradecimento a um dos seus autores, Prof.º David M. Levine, da *City University of New York*, por nos ter permitido acessar o material didático derivado desse livro, especialmente as apresentações do *Microsoft PowerPoint* preparadas pelo Prof.º John McGill.

O presente trabalho representa o resultado de seis anos de iniciativas individuais, de treinamentos promovidos pela Assessoria de Relações Internacionais (Arint) e pelo Instituto Serzedello Corrêa (ISC), e de recursos alocados pela extinta Secretaria de Auditoria e Inspeções (Saudi) e pelas atuais Secretaria-Adjunta de Fiscalização (Adfis), Secretaria de Macroavaliação Governamental (Semag), Secretaria de Controle Externo no Estado do Rio de Janeiro (Secex-RJ), Secretaria de Fiscalização de Obras e Patrimônio da União (Secob) e Secretaria de Fiscalização e Avaliação de Programas de Governo (Seprog). As fases do trabalho foram:

- participação no “*TCU Auditing Seminar*”, ministrado, em novembro de 1995, pelos auditores Roger Adams e Richard Regal, da *Association of Certified Chartered Accountants*, com 60 horas de duração;
- manutenção de contatos com auditores da Inspetoria-Geral do Departamento da Defesa dos EUA, como parte do curso “*Executive Development Program in Performance Management*”, ministrado, em junho de 1996, pelo *Virginia Polytechnic Institute and State University* e pela Fundação Getúlio Vargas, com o auxílio da *National Academy of Public Administration*;
- participação no curso “Estatística Aplicada e Noções de Amostragem”, ministrado, em maio de 1997, pelo Prof.º George von Borries, com 20 horas de duração;
- participação na disciplina “Delineamento e Análise de Amostras” do mestrado em Estatística da Universidade de Brasília, ministrada no 2º semestre de 1997;
- elaboração deste documento no período de setembro de 1998 a março de 1999;
- organização do módulo “Técnicas de Amostragem”, ministrado em abril de 1999, com 15 horas de duração, do curso “Introdução à Auditoria” do Programa de Formação de 1999 para AFCEs-CE;
- revisão de uma versão preliminar deste documento pelos participantes do módulo citado acima, o que permitiu a supressão de várias imprecisões;
- participação, em decorrência de convênio firmado pelo TCU e pela Embaixada Britânica em março de 1998, na disciplina “Estatística” do mestrado em Economia da *London School of Economics and Political Science*, ministrada em setembro de 1999;
- complementação deste trabalho em março de 2001;
- revisão final no âmbito da Adfis e por servidores designados pelo Memorando-Circular nº 029 – Adfis, de 15/05/2001, no período de abril a julho de 2001;
- participação no curso “Técnicas de Amostragem com o Auxílio do *Software SAS*”, ministrado, em junho de 1997, pela Prof.ª Édina Shisue Miazaki, com 36 horas de duração;

- finalização deste trabalho em setembro de 2001, incorporando-se quase todas as sugestões apresentadas pelos revisores.¹

Inegavelmente, trata-se de um período de maturação longo. A publicação de um documento como este, entretanto, somente tem sentido quando há uma demanda nítida por esse tipo de trabalho, como é o caso atualmente.

O texto é composto por oito capítulos:

- a) o primeiro discorre sobre alguns conceitos básicos que devem ser fixados desde o primeiro instante;
- b) o segundo revê e exemplifica as propriedades dos dados numéricos;
- c) o terceiro trata dos modelos de distribuição de probabilidade;
- d) o quarto aborda as propriedades das distribuições amostrais;
- e) o quinto discute os problemas e técnicas elementares de estimação, os quais se confundem com a amostragem aleatória simples;
- f) o sexto estende os conceitos do capítulo anterior, discorrendo sobre as versões básicas das amostragens aleatórias estratificada e por conglomerados;
- g) o sétimo discute as características da amostragem por unidade monetária, empregada em auditorias contábil-financeiras;
- h) o oitavo, por fim, trata dos erros em amostragem, em especial do problema da não-resposta.

Em todos os capítulos, recorreu-se a exemplos no intuito de tornar mais claros os temas tratados. A bibliografia, por sua vez, foi dividida em “principal” e “complementar” para que os leitores tenham mais opções para consulta, ainda que somente o primeiro grupo tenha sido ostensivamente consultado.

¹ Naturalmente, compete ao redator a responsabilidade pelas falhas remanescentes.

1. CONCEITOS BÁSICOS

Em auditoria, as **técnicas de amostragem** visam coletar e avaliar evidências numéricas das entidades administrativas no intuito de determinar e relatar o grau de adequação das informações obtidas a critérios previamente definidos. Isso se deve à natureza antieconômica das auditorias que pretendam investigar todo o universo visado.

As técnicas em questão, por se basearem em princípios estatísticos demonstráveis, apresentam as seguintes vantagens:

- a) o tamanho da amostra e o erro amostral podem ser estimados prévia e objetivamente;
- b) as amostragens conduzidas por auditores diferentes podem ser combinadas;
- c) os censos, além de serem demorados, podem conter mais erros não-amostrais do que as amostras;
- d) os resultados amostrais são objetivos e, por extensão, defensáveis;
- e) os resultados da auditoria podem ser avaliados com segurança e extrapolados para toda a população.

1.1. Inferência Estatística

A crescente demanda por dados numéricos observada ao longo da história está estreitamente relacionada com o desenvolvimento da **estatística descritiva**.

UTILIZANDO O *MICROSOFT EXCEL*:

OPÇÃO “FERRAMENTAS/ANÁLISE DE DADOS...” (VIDE FIGURAS A.1 – A.3, ANEXO I) ⇒ gera estatísticas descritivas de um conjunto de observações.

Modernamente, porém, é graças ao desenvolvimento da **inferência estatística**, partindo de extensões da teoria da probabilidade, que a estatística passou a ser amplamente empregada por todo tipo de pesquisa.

A relevância dos métodos de inferência estatística deve-se ao fato de que as técnicas de amostragem tornaram-se indispensáveis, pois o crescimento populacional tornou excessivamente onerosa, demorada e complexa a coleta de dados sobre toda a população. Decisões relacionadas com as características da população devem se basear em informações extraídas de amostras, com a teoria da probabilidade fornecendo o elo entre ambas mediante a definição da probabilidade de que os resultados amostrais espelhem os parâmetros populacionais.

Convém frisar que, para que a análise estatística seja útil ao processo de tomada de decisão, os dados coletados devem ser apropriados, ou seja, livres de vieses, ambigüidades ou outros tipos de erro, pois essas deficiências dificilmente poderão ser compensadas, mesmo pelos mais modernos métodos estatísticos.

1.2. Principais Tipos de Amostragem Probabilística

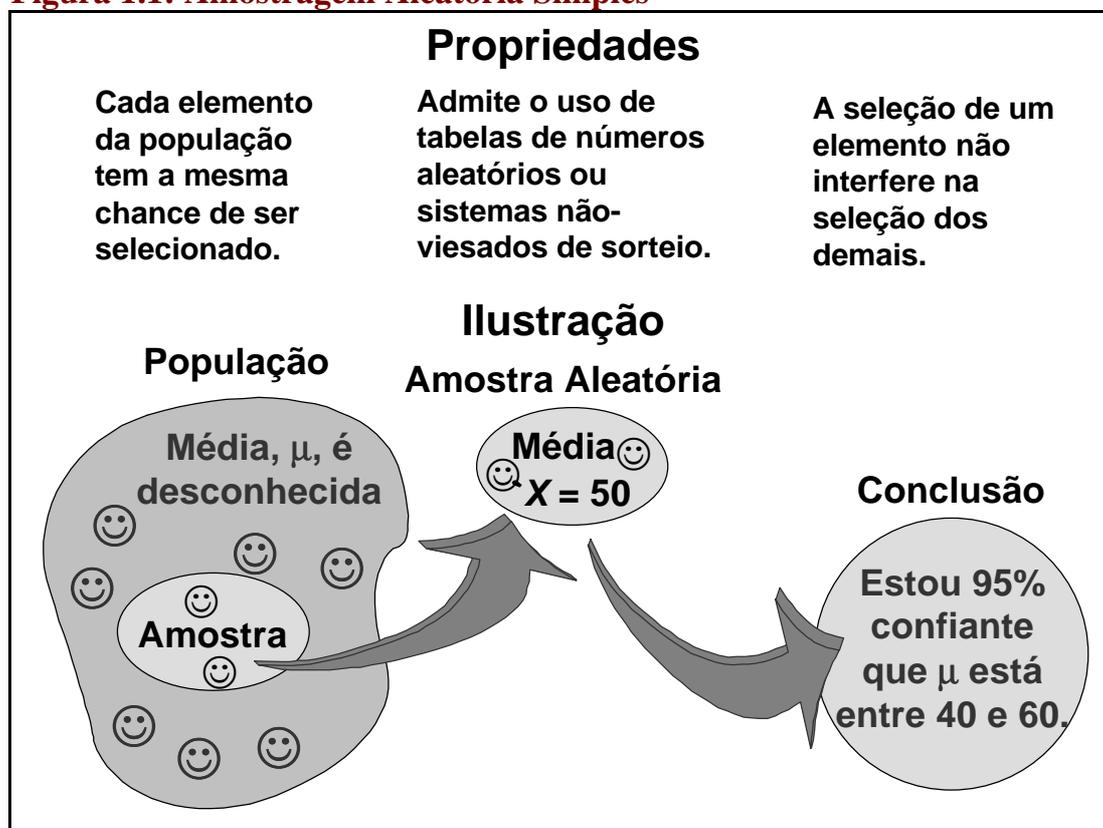
Há dois tipos de amostragens: **não-probabilísticas** e **probabilísticas**. As amostras por quotas e por julgamento do pesquisador são não-probabilísticas, enquanto que as **amostras aleatórias simples, estratificadas e por conglomerados** são probabilísticas, uma vez que seus elementos são selecionados com base em probabilidades conhecidas. Em estudos numéricos, somente as amostragens probabilísticas permitem a correta generalização para a população dos resultados amostrais.

Nas amostras probabilísticas, as inferências podem ser formuladas por intermédio de **estimadores simples, do tipo razão ou do tipo regressão**. Os primeiros se baseiam unicamente nos dados referentes à variável acerca da qual se deseja gerar inferências, enquanto que os outros também recorrem a variáveis auxiliares, supostamente relacionados com o fenômeno estudado. Este documento, contudo, somente tratará de estimadores simples, pois os demais envolvem um nível maior de sofisticação matemática.

1.2.1. Introdução à Amostragem Aleatória Simples

Embora a amostragem aleatória simples não seja, necessariamente, a estratégia amostral mais eficiente e econômica, ela funciona como base para as estratégias mais sofisticadas. As suas principais características constam da Figura 1.1.

Figura 1.1: Amostragem Aleatória Simples



Fonte: McGill, 1997.

O elemento-chave de qualquer estratégia amostral é a obtenção e manutenção de um **cadastro** atualizado (i.e., sistema de referência ou, na língua inglesa, “*frame*”) de todos

os itens ou indivíduos que compõem a população da qual será extraída a amostra (i.e., **população-alvo**). Caso alguns grupos de itens ou indivíduos não sejam adequadamente contemplados pelo cadastro, a população-alvo e a verdadeira população diferirão. Dessa forma, as estimativas geradas pelas amostras aleatórias serão válidas para a população-alvo, mas viesadas para a verdadeira população.

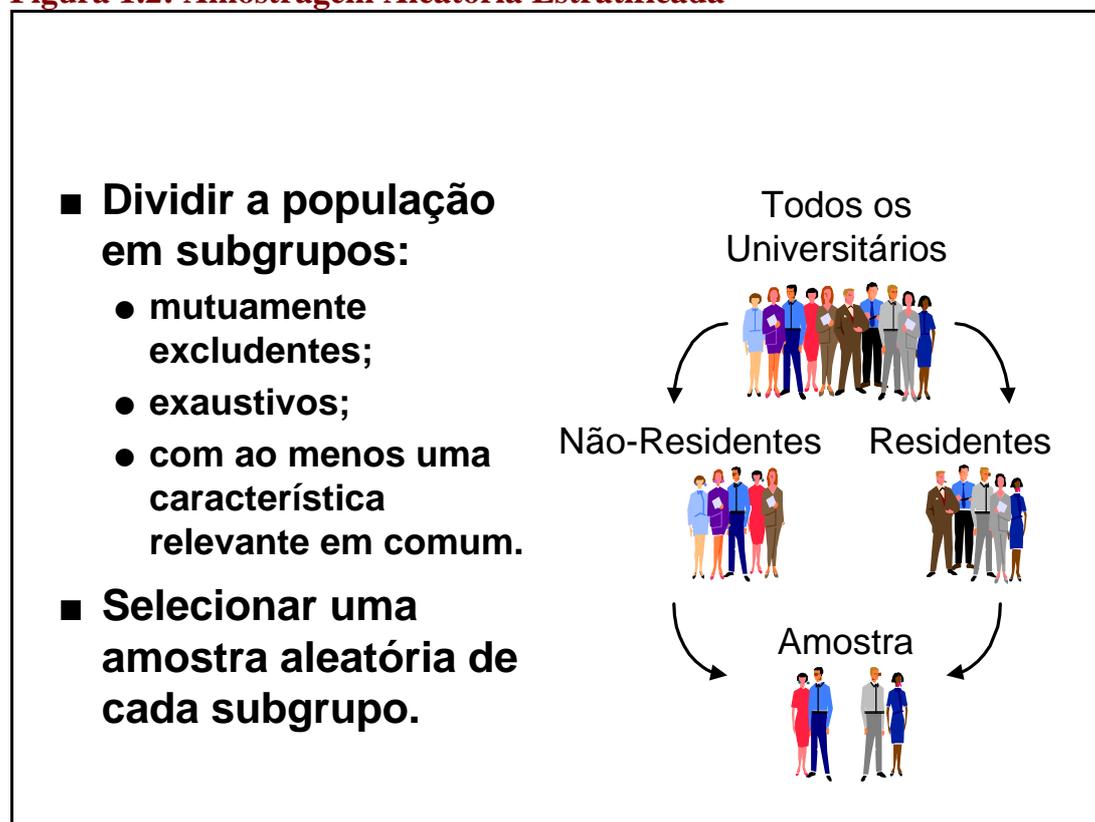
As amostras aleatórias podem ser selecionadas com ou sem reposição de populações finitas ou infinitas. O método empregado deve ser indicado claramente, pois as equações usadas nas inferências estatísticas variam com os métodos. Convém notar que as **amostragens aleatórias simples com reposição de populações finitas e sem reposição de populações infinitas geram estimadores cujas equações são idênticas**.

1.2.2. Introdução à Amostragem Aleatória Estratificada

Quando os elementos da população puderem ser agrupados em conjuntos homogêneos (i.e., **estratos**), pode-se utilizar a amostragem aleatória estratificada, selecionando-se uma amostra para cada estrato. Os tamanhos das amostras e as estatísticas estimadas são independentes. Ao se combinar as estimativas de cada estrato, obtém-se uma estimação para toda a população.

Na estratificação da população, cada elemento deve constar de um único estrato, ou seja, os estratos não podem possuir interseções, enquanto que o conjunto de elementos de um estrato deve ser o mais homogêneo possível em relação à característica que se pretende examinar, como ilustrado pela Figura 1.2.

Figura 1.2: Amostragem Aleatória Estratificada



Fonte: McGill, 1997.

A amostragem estratificada apresenta as seguintes vantagens:

- a) com relação à amostra aleatória simples, supondo-se que haja um acréscimo no volume de informação acerca da população em decorrência da estratificação (p. ex., o desvio-padrão observado em cada um dos estratos é substancialmente inferior ao desvio-padrão do conjunto da população):
 - obter precisão semelhante com uma amostra total menor;
 - obter maior precisão com uma amostra total igual;
- b) obter estimativas para cada estrato, se essas estimativas forem úteis;
- c) dar maior atenção a certos grupos dentro da população que têm uma propensão muito alta ou muito baixa a apresentar a característica examinada.

A estratificação pode ser útil também se os custos de auditoria são diferentes para cada estrato (p. ex., entrevistas de contato direto em alguns casos e questionários enviados pelo correio em outros).

1.2.3. Introdução à Amostragem Aleatória por Conglomerados

Quando os elementos da população podem ser agrupados em conjuntos semelhantes, mas internamente heterogêneos (i.e., **conglomerados**), pode-se usar a amostragem aleatória por conglomerados. Nessa técnica, as amostras são obtidas por meio não da seleção aleatória de alguns de seus elementos, mas sim de alguns dos conglomerados que a compõem. Uma vez selecionados os conglomerados, todos os seus elementos são examinados para que as estatísticas desejadas sejam obtidas, como ilustrado pela Figura 1.3.

Figura 1.3: Amostragem Aleatória por Conglomerados



Fonte: McGill, 1997.

Quando os conglomerados são compostos por uma grande quantidade de elementos, entretanto, pode-se estimar as estatísticas referentes a cada conglomerados por meio da amostragem aleatória simples. Essa técnica é denominada de **amostragem aleatória por conglomerados em dois estágios**, que também não será examinada neste documento em função do nível de sofisticação matemática.

1.3. Plano Amostral

O ponto de partida de toda amostra é o plano amostral, o qual documenta os passos e os procedimentos envolvidos na utilização de técnicas de amostragem, devendo envolver as etapas definidas a seguir:

- a) estipular os objetivos do uso de técnicas de amostragem, explicando-se o porquê da não-utilização de um censo;
- b) definir os elementos da população, ou seja, os indivíduos ou objetos acerca dos quais serão feitas estimativas;
- c) definir o tamanho da população, recorrendo-se, se necessário, a estimativas;
- d) examinar o cadastro da população ou, caso não esteja disponível, a descrição dos itens relevantes para a seleção;
- e) descrever a técnica de amostragem a ser utilizada e justificar a escolha;
- f) descrever os procedimentos seguidos na execução do trabalho;
- g) estabelecer o nível de confiança;
- h) determinar o tamanho da amostra e a precisão desejada, o que freqüentemente requer uma amostra preliminar;
- i) escolher as técnicas de coleta, armazenamento e análise dos dados.

O maior ou menor sucesso dos estudos amostrais, independentemente do seu tipo, costuma guardar relação direta com a melhor ou pior observância dessas etapas.

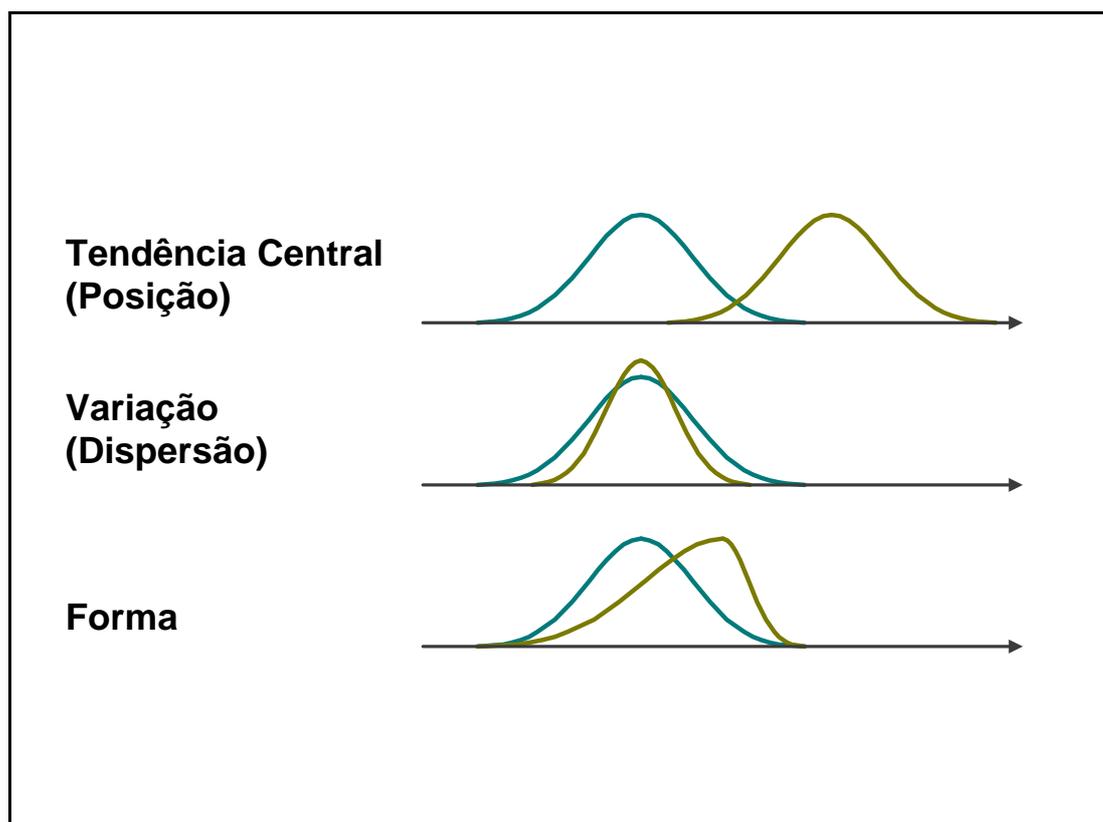
2. PROPRIEDADES DOS DADOS NUMÉRICOS

Em geral, os pesquisadores procuram desenvolver instrumentos destinados a obter respostas objetivas para perguntas acerca de vários tipos de fenômenos ou características (i.e., **variáveis aleatórias**). Os dados coletados representam manifestações das variáveis aleatórias, podendo diferir de resposta para resposta. Essas manifestações podem ser: **categóricas** ou **numéricas**.

UTILIZANDO O *MICROSOFT EXCEL*:
OPÇÃO “DADOS/RELATÓRIO DE TABELA DINÂMICA” (VIDE FIGURAS A.1 & A.4 – A.8, ANEXO I) ⇒ gera estatísticas agrupadas por uma variável categórica.

Com relação aos dados numéricos, as suas principais propriedades, ilustradas pela Figura 2.1, são: **tendência central**, **variação** e **forma**.

Figura 2.1: Propriedades dos Dados Numéricos



Fonte: McGill, 1997.

2.1. Medidas de Tendência Central

As medidas de tendência central são **média aritmética**, **mediana**, **moda**, **meia amplitude** e **meia-amplitude semi-interquartilica**, que podem ser definidas da seguinte maneira:

- a) Média Aritmética Populacional (μ) e Amostral (\bar{X}): dado um conjunto de n observações $X_1, X_2, X_3, \dots, X_n$, μ ou \bar{X} é definida pelo somatório do valor de todas as observações dividido por n , ou seja:²

$$\bullet \mu = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N} = \frac{\sum_{i=1}^N X_i}{N};$$

$$\bullet \bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n}.$$

UTILIZANDO O MICROSOFT EXCEL: =MÉDIA(núm1;núm2;...).

- b) Mediana (Md): dada uma seqüência ordenada de n observações $X_1, X_2, X_3, \dots, X_n$, Md é definida pelo valor que divide a seqüência em dois conjuntos com a mesma quantidade de observações. Caso n seja ímpar, a mediana é dada pelo valor correspondente à $\left(\frac{n+1}{2}\right)$ -ésima posição da seqüência. Caso n seja par, a mediana é dada pela média dos valores correspondentes às $\left(\frac{n}{2}\right)$ e $\left(\frac{n}{2}+1\right)$ -ésimas posições.

UTILIZANDO O MICROSOFT EXCEL: =MED(núm1;núm2;...).

- c) Moda (Mo): dado um conjunto de n observações $X_1, X_2, X_3, \dots, X_n$, Mo é definida pelo(s) valor(es) mais freqüente(s) entre todas as observações.

UTILIZANDO O MICROSOFT EXCEL: =MODO(núm1;núm2;...).

- d) Meia Amplitude (Ma): dado um conjunto de n observações $X_1, X_2, X_3, \dots, X_n$, Ma é definida pela média do menor e maior valores observados, ou seja:

$$Ma = \frac{X_{\text{mínimo}} + X_{\text{máximo}}}{2}.$$

UTILIZANDO O MICROSOFT EXCEL:

=(MÁXIMO(núm1;núm2;...)+MÍNIMO(núm1;núm2;...))/2.

- e) Meia Amplitude Semi-Interquartílica (Mq): dada uma seqüência ordenada de n observações $X_1, X_2, X_3, \dots, X_n$, Mq é definida pela média dos valores correspondentes ao primeiro e terceiro quartis (Q_1 e Q_3)³, ou seja:

$$Mq = \frac{Q_1 + Q_3}{2}.$$

² Caso toda a população seja abrangida pelas observações, n deve ser substituído por N .

³ O segundo quartil (Q_2) corresponde à mediana (Md).

- O primeiro quartil (Q_1) é definido pelo valor correspondente à $\left(\frac{n+1}{4}\right)$ -ésima posição da seqüência, o qual separa os primeiros 25% das observações dos 75% restantes.
- O terceiro quartil (Q_3) é definido pelo valor correspondente à $\left[3 \cdot \left(\frac{n+1}{4}\right)\right]$ -ésima posição da seqüência, o qual separa os primeiros 75% das observações dos 25% restantes.
- Se $(n+1)$ não for múltiplo de 4, as equações $\left(\frac{n+1}{4}\right)$ e $\left[3 \cdot \left(\frac{n+1}{4}\right)\right]$ não gerarão números inteiros. Nesse caso, há três possibilidades:
 - i. caso n seja múltiplo de 4, o quartil desejado será dado pela posição imediatamente inferior ao número obtido;
 - ii. caso $(n+2)$ seja múltiplo de 4, o quartil desejado será dado pela posição imediatamente superior ao número obtido;
 - iii. caso $(n+3)$ seja múltiplo de 4, o quartil desejado será dado pela média dos valores correspondentes às $\left(\frac{n-1}{4}\right)$ e $\left(\frac{n+3}{4}\right)$ -ésimas posições.

UTILIZANDO O *MICROSOFT EXCEL*: VIDE TABELA A.1, ANEXO I.

2.2. Medidas de Variação

Já as medidas de variação são **amplitude**, **amplitude semi-interquartílica**, **variância**, **desvio-padrão** e **coeficiente de variação**, que são definidas como indicado abaixo:

- a) Amplitude (A_t): dado um conjunto de n observações $X_1, X_2, X_3, \dots, X_n$, A_t é definida pela diferença entre o maior e menor valores observados, ou seja:
- $$A_t = X_{\text{máximo}} - X_{\text{mínimo}}.$$

UTILIZANDO O *MICROSOFT EXCEL*:

=(MÁXIMO(núm1;núm2;...)-MÍNIMO(núm1;núm2;...)).

- b) Amplitude Semi-Interquartílica (D_q): dado um conjunto de n observações $X_1, X_2, X_3, \dots, X_n$, D_q é definida pela diferença entre o terceiro e o primeiro quartis, ou seja: $D_q = Q_3 - Q_1$.

UTILIZANDO O *MICROSOFT EXCEL*: VIDE TABELA A.1, ANEXO I.

- c) Variância Populacional (σ^2) e Amostral (S^2): dado um conjunto contendo n observações $X_1, X_2, X_3, \dots, X_n$, σ^2 e S^2 correspondem ao somatório do

quadrado da diferença entre cada observação e a média (μ ou \bar{X}) dividido, respectivamente, por N (**observações abrangem toda a população**) ou por $(n-1)$ (**observações abrangem apenas uma amostra**), ou seja:

$$\bullet \sigma^2 = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + (X_3 - \mu)^2 + \dots + (X_N - \mu)^2}{N} = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N} = \frac{\sum_{i=1}^N X_i^2 - N \cdot \mu^2}{N}.$$

UTILIZANDO O *MICROSOFT EXCEL*: =VARP(núm1;núm2;...).

$$\bullet S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{(n-1)} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)} = \frac{\sum_{i=1}^n X_i^2 - n \cdot \bar{X}^2}{(n-1)}.$$

UTILIZANDO O *MICROSOFT EXCEL*: =VAR(núm1;núm2;...).

d) Desvio-Padrão Populacional (σ) e Amostral (S): dado um conjunto contendo n observações $X_1, X_2, X_3, \dots, X_n$, σ e S correspondem, respectivamente, à raiz quadrada da variância populacional (σ^2) e amostral (S^2), ou seja:⁴

$$\bullet \sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^N X_i^2 - N \cdot \mu^2}{N}}.$$

UTILIZANDO O *MICROSOFT EXCEL*: = DESVPADP(núm1;núm2;...).

$$\bullet S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n X_i^2 - n \cdot \bar{X}^2}{(n-1)}}.$$

UTILIZANDO O *MICROSOFT EXCEL*: = DESVPAD(núm1;núm2;...).

e) Coeficiente de Variação (CV): dado um conjunto de n observações $X_1, X_2, X_3, \dots, X_n$, CV mede a variação relativa das observações em torno da média (μ ou \bar{X}): $CV_{populacional} = \left(\frac{\sigma}{\mu}\right) 100\%$ ou $CV_{amostral} = \left(\frac{S}{\bar{X}}\right) 100\%$.

UTILIZANDO O *MICROSOFT EXCEL*:

=(DESVPADP(núm1;núm2;...)/MÉDIA(núm1;núm2;...)) \Rightarrow $CV_{populacional}$.

=(DESVPAD(núm1;núm2;...)/MÉDIA(núm1;núm2;...)) \Rightarrow $CV_{amostral}$.

⁴ O desvio-padrão e a variância têm como objetivo medir a distribuição das n observações em torno da média. Quanto mais próximas de zero, maior a concentração.

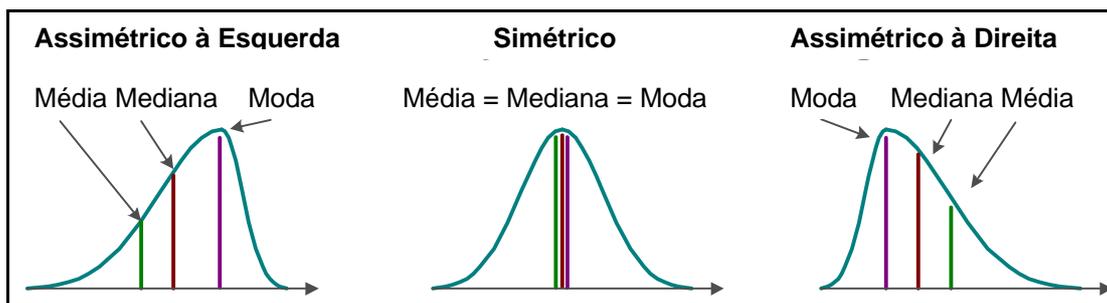
2.3. Relação entre a Média e a Mediana

A curva de frequência de um conjunto de n observações $X_1, X_2, X_3, \dots, X_n$ pode exibir uma forma simétrica, assimétrica à direita ou assimétrica à esquerda. Isso pode ser constatado comparando-se a média (μ ou \bar{X}) e a mediana (Md). Nesse caso, há três possibilidades:

- μ ou $\bar{X} > Md \Rightarrow$ forma assimétrica à direita (mais de 50% das observações concentra-se à esquerda da média);
- μ ou $\bar{X} = Md \Rightarrow$ forma simétrica (a média separa os primeiros 50% das observações dos 50% restantes);
- μ ou $\bar{X} < Md \Rightarrow$ forma assimétrica à esquerda (mais de 50% das observações concentra-se à direita da média).

Convém notar que a análise anterior pode ser estendida às modas, como mostra a Figura 2.2.

Figura 2.2: Relação entre a Média, a Mediana e a Moda



Fonte: McGill, 1997.

3. DISTRIBUIÇÕES DE PROBABILIDADE

As variáveis aleatórias numéricas podem ser descritas por intermédio de **distribuições de probabilidades**, as quais devem contemplar todas as possíveis manifestações do fenômeno examinado e as respectivas probabilidades ou frequências relativas. Esses dados são obtidos por meio de cadastros, os quais se classificam em:

- teóricos, gerados por modelos matemáticos que representem, com graus variados de sucesso, os fenômenos examinados;
- empíricos, elaborados com base nas frequências efetivamente observadas;
- subjetivos, que reflitam as convicções do pesquisador acerca das frequências.

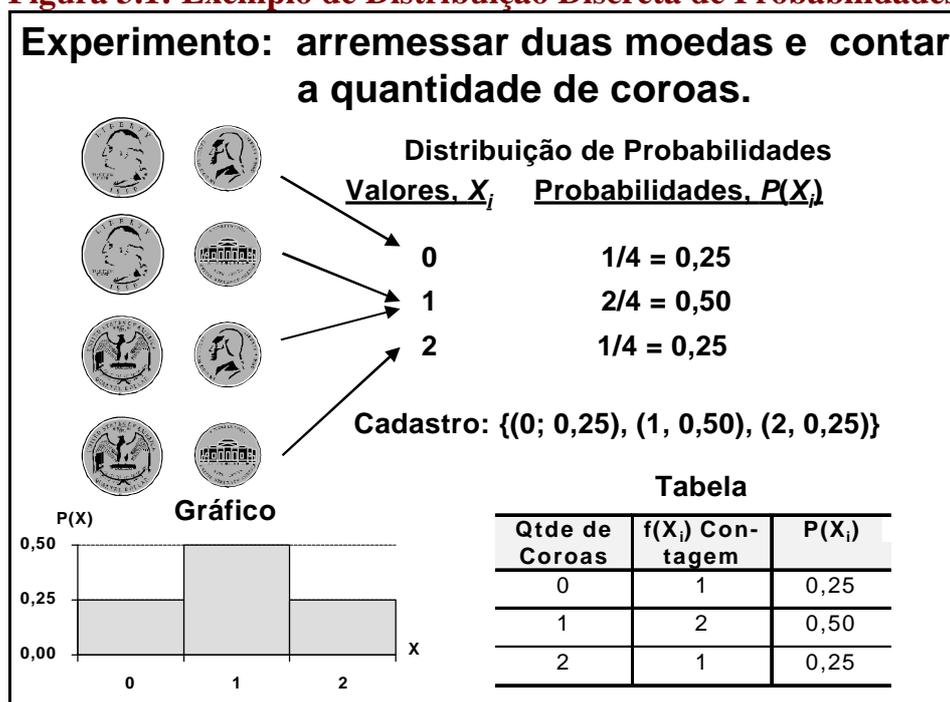
O primeiro tipo de cadastro permite definir **funções de distribuição de probabilidades**, com bases nas quais são calculadas com precisão a probabilidade de ocorrência de qualquer possível manifestação da variável aleatória. Por essa razão, somente esse tipo de cadastro será analisado neste documento.

Os modelos de distribuição de probabilidade apresentam características específicas conforme as variáveis aleatórias numéricas sejam **discretas** ou **contínuas**.

3.1. Modelos de Distribuição Discreta

Um exemplo clássico de distribuição discreta corresponde ao resultado do arremesso de duas moedas, como mostrado na Figura 3.1.

Figura 3.1: Exemplo de Distribuição Discreta de Probabilidades



Fonte: McGill, 1997.

Os modelos de distribuição discreta mais importantes são: a **binomial**, a **hipergeométrica**, a **binomial negativa** e a **de Poisson**.

3.1.1. Distribuição Binomial

As principais propriedades da distribuição binomial são:

- as **amostras** podem ser **selecionadas a partir de populações infinitas sem reposição** ou **finitas com reposição**;
- as observações selecionadas podem ser classificadas em uma de duas categorias mutuamente excludentes e coletivamente exaustivas, usualmente chamadas “**sucesso**” e “**fracasso**”;
- as probabilidades de que uma observação seja classificada um “sucesso” (p) ou um “fracasso” ($1-p$) são constantes e complementares;
- as observações são mutuamente independentes.

A variável aleatória discreta ou o fenômeno estudado que segue uma distribuição binomial é a que trata da quantidade de “sucessos” observados em uma amostra. Conhecendo-se a quantidade n de observações contidas na amostra e a probabilidade p de “sucesso”, a distribuição binomial é representada pela seguinte equação:

$$P(X) = \underbrace{\frac{n!}{X!(n-X)!}}_{\text{quantidade de possíveis seqüências}} \cdot \underbrace{p^X \cdot (1-p)^{(n-X)}}_{\text{probabilidade de uma seqüência em particular}}$$

Onde: $P(X)$ = probabilidade de que sejam observados X “sucessos”, dados n e p ;
 X = quantidade de “sucessos” na amostra $\{X \in \mathbb{1} \mid 0 \leq X \leq n\}$;
 n = tamanho da amostra;
 p = probabilidade de “sucesso”;
 $(1-p)$ = probabilidade de “fracasso”.

USANDO MICROSOFT EXCEL:

=DISTRBINOM(núm_s;tentativas;probabilidade_s;cumulativo).

Onde: núm_s = X ;

tentativas = n ;

probabilidade = p ;

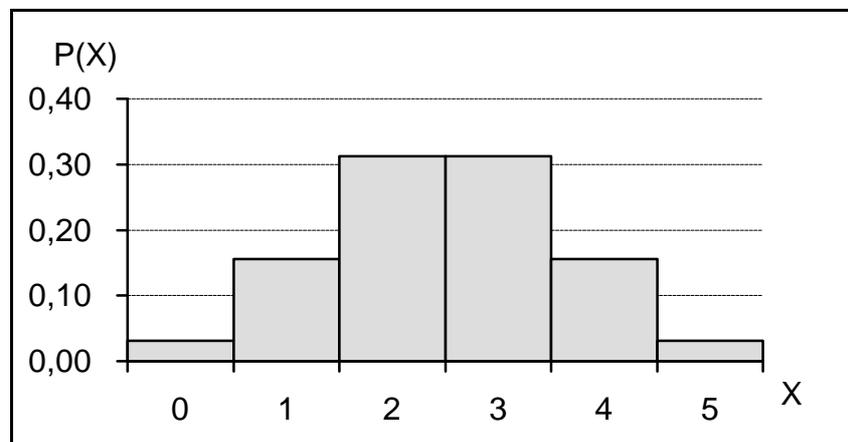
cumulativo = $\begin{cases} \text{VERDADEIRO} = \text{para calcular a prob. de } X \text{ ou menos "sucessos"}; \\ \text{FALSO} = \text{para calcular a prob. de } X \text{ "sucessos"}. \end{cases}$

A distribuição binomial apresenta inúmeras aplicações, como mostra a Tabela 3.1.

Tabela 3.1: Aplicações da Distribuição Binomial

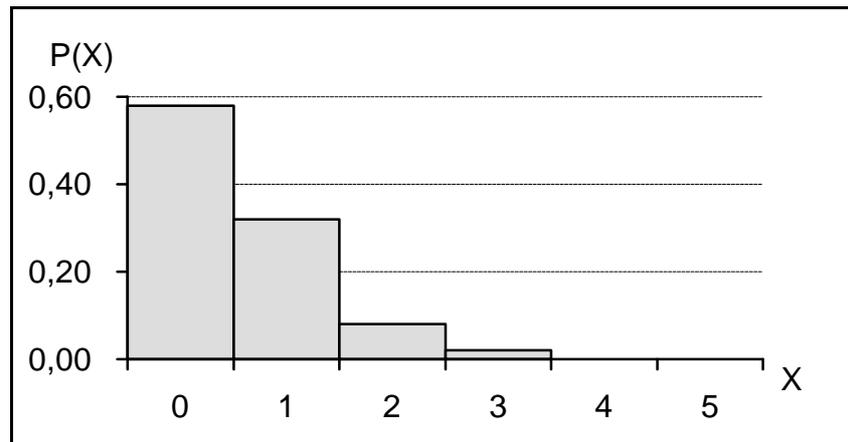
Gênero	Aplicações	X	n	p	$P(X)$
Jogos de Azar	Probabilidade de que, em 20 tentativas, 15 ou mais resultados da roleta sejam vermelhos?	> 14	20	0,50	0,021
Controle de Qualidade	Probabilidade de que, em uma amostra de 20 pneus, nenhum seja defeituoso dado que 8% de todos os pneus sejam defeituosos?	20	20	0,92	0,189
Educação	Probabilidade de que um aluno seja aprovado em um exame com 10 questões, cada questão com quatro opções, caso o aluno tente adivinhar todas as questões?	> 4	10	0,25	0,078
Finanças	Probabilidade de que o valor de uma ação aumente por 10 dias consecutivos dado que os preços das ações efetivamente oscilem aleatoriamente?	10	10	0,50	0,001

A função binomial de distribuição de probabilidades é simétrica quando $p = 0,5$ e assimétrica quando $p \neq 0,5$. Quanto mais próximo p for de 0,5 e quanto maior for n , menor será o viés, como mostrado nas Figuras 3.2 e 3.3.

Figura 3.2: Distribuição Binomial com $n = 5$ e $p = 0,5$ 

Fonte: McGill, 1997.

Figura 3.3: Distribuição Binomial com $n = 5$ e $p = 0,1$



Fonte: McGill, 1997.

A média e o seu desvio-padrão são obtidos, respectivamente, por meio das seguintes equações:

- a) Média: $\mu = E(X) = n.p$;
- b) Desvio-padrão: $\sigma = \sqrt{n.p.(1-p)}$.

3.1.2. Distribuição Hipergeométrica

A distribuição hipergeométrica, a exemplo da distribuição binomial, trata da quantidade de “sucessos” observados em uma amostra com n observações. A diferença entre ambas reside no método de seleção empregado. Enquanto que na binomial as observações são selecionadas com reposição de populações finitas ou sem reposição de populações infinitas, na hipergeométrica as **observações** são **selecionadas sem reposição de populações finitas**. Portanto, a probabilidade p de “sucesso” nesse caso não é constante. A equação da distribuição em questão é:

$$P(X) = \frac{\frac{A!}{X!(A-X)!} \cdot \frac{(N-A)!}{(n-X)![(N-A)-(n-X)]!}}{\frac{N!}{n!(N-n)!}} = \frac{\binom{A}{X} \binom{N-A}{n-X}}{\binom{N}{n}}$$

Onde: $P(X)$ = probabilidade de que sejam observados X “sucessos”, dados n e p ;

A = quantidade de “sucessos” na população $\{X \in \mathbb{1} \mid 0 \leq X \leq N\}$;

X = quantidade de “sucessos” na amostra $\{X \in \mathbb{1} \mid 0 \leq X \leq \text{Min}[n, A]\}$;

n = tamanho da amostra;

N = tamanho da população.

USANDO *MICROSOFT EXCEL*:

=DIST.HIPERGEOM(exemplo_s;exemplo_núm;população_s;núm_população).

Onde: exemplo_s = X ;

exemplo_núm = n ;

população_s = A ;

núm_população = N .

3.1.3. Distribuição Binomial Negativa

A distribuição binomial negativa é usada para determinar a probabilidade de ocorrência de uma certa quantidade de “fracassos” antes que outra quantidade de sucessos ocorra. A sua equação é:

$$P(n) = \frac{(n-1)!}{(X-1)! \cdot [(n-1)-(X-1)]!} \cdot p^X \cdot (1-p)^{(n-X)} = \binom{n-1}{X-1} p^X \cdot (1-p)^{(n-X)}.$$

USANDO *MICROSOFT EXCEL*: =DIST.BIN.NEG(núm_f;núm_s;probabilidade_s).

Onde: núm_f = $(n - X)$;

núm_s = X ;

probabilidade_s = p .

3.1.4. Distribuição de Poisson

Há um processo de Poisson quando eventos discretos podem ser observados em uma **área de oportunidade** (i.e., um intervalo contínuo de tempo, de superfície, de comprimento, etc.) de tal modo que, se essa área for suficientemente reduzida, o seguinte ocorrerá:

- a) a probabilidade de que seja observado um único “sucesso” é estável;
- b) a probabilidade de que mais de um “sucesso” seja observado é nula;
- c) a probabilidade de ocorrência de um “sucesso” em qualquer intervalo é estatisticamente independente da probabilidade observada nos demais intervalos.

A equação correspondente é:

$$P(X) = \frac{e^{-\lambda} \cdot \lambda^X}{X!}.$$

Onde: $P(X)$ = probabilidade de que sejam observados X “sucessos”, dado \mathbf{I} ;

X = quantidade de “sucessos” por unidade da **área de oportunidade**;

\mathbf{I} = quantidade esperada de “sucessos”;

$e = 2,71828$.⁵

⁵ Número de Euler, base do sistema de logaritmos neperianos.

USANDO *MICROSOFT EXCEL*: =POISSON(X;média;cumulativo).

Onde: média = **I**;

cumulativo = $\begin{cases} \text{VERDADEIRO} = \text{para calcular a probabilidade de } X \text{ ou menos "sucessos"}; \\ \text{FALSO} = \text{para calcular a probabilidade de } X \text{ "sucessos"}. \end{cases}$

O Exemplo 3.1 ilustra uma das possíveis aplicações da distribuição de Poisson.

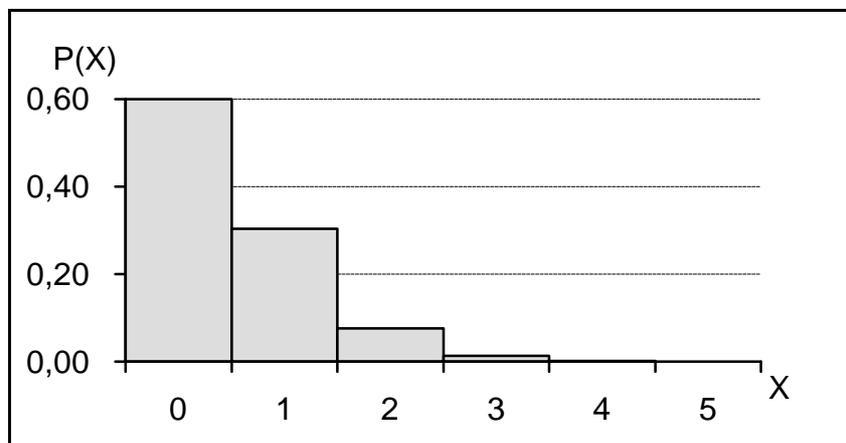
Exemplo 3.1: Aplicação da Distribuição de Poisson

Suponha-se um banco que atende 180 clientes em um período de 1 hora. Toda chegada de um cliente é um evento discreto em um determinado instante de um intervalo contínuo de tempo. Partindo-se o período de 1 hora em 3.600 intervalos de 1 segundo, tem-se:

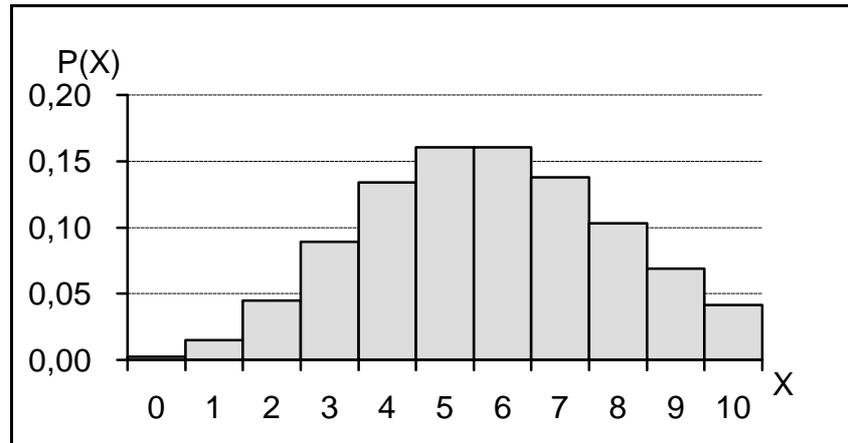
- a) a quantidade esperada (ou média) de clientes atendidos por segundo é 0,05;
- b) a probabilidade de que mais de um cliente seja atendido em um dado segundo é quase nula;
- c) o atendimento de um cliente em um dado segundo não afeta o atendimento dos demais clientes nos demais segundos.

As Figuras 3.4 e 3.5, a seu tempo, mostram os efeitos do parâmetro λ sobre o formato do gráfico da distribuição de probabilidades.

Figura 3.4: Distribuição de Poisson com $\lambda = 0,5$



Fonte: McGill, 1997.

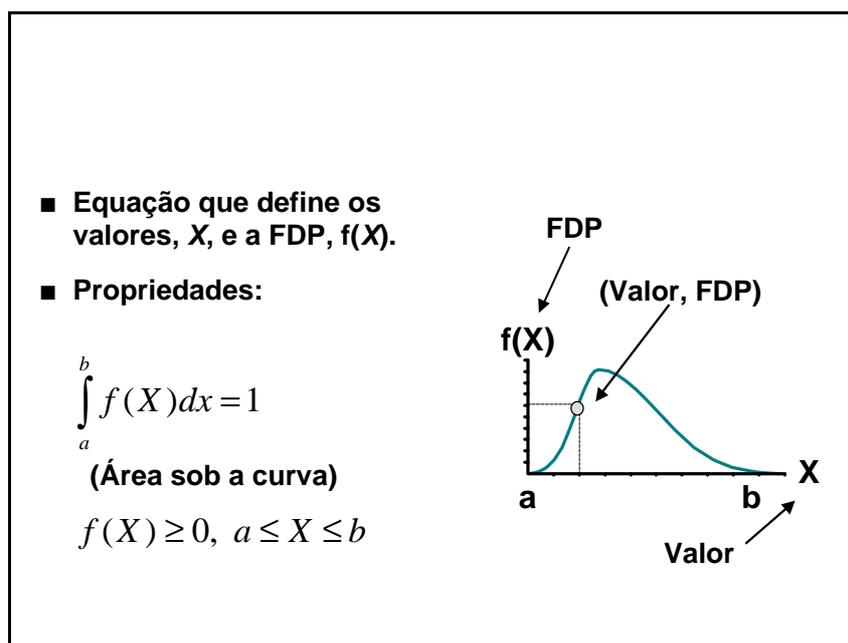
Figura 3.5: Distribuição de Poisson com $\lambda = 6$ 

Fonte: McGill, 1997.

Convém notar que a distribuição de Poisson não apenas representa vários fenômenos (p. ex., telefonemas recebidos em uma manhã, acidentes de trânsito observados em uma semana, greves havidas em um ano, etc.), como também pode funcionar como uma aproximação da distribuição binomial quando n for grande e p pequeno, com $(n.p) \leq 7$. Nesse caso, tem-se: $\lambda = n.p$.

3.2. Modelos de Distribuição Contínua

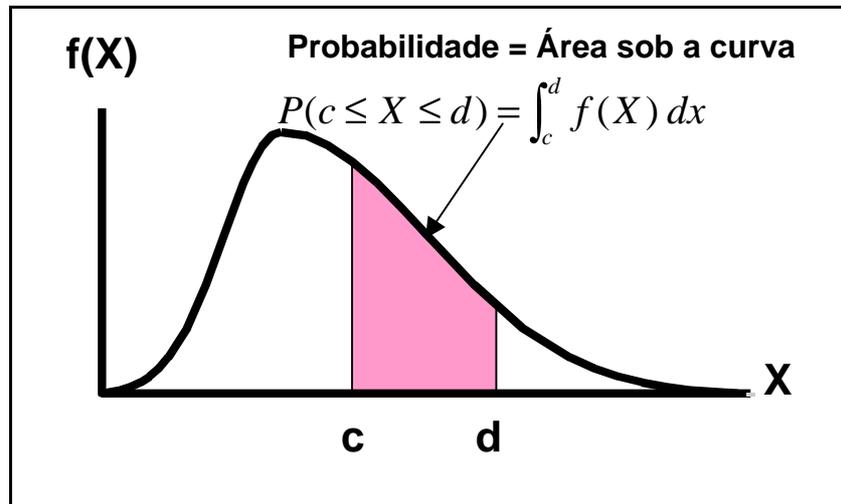
A representação matemática da distribuição de uma variável aleatória contínua é denominada **função densidade de probabilidades (FDP)**, cuja equação-geral e principais características constam da Figura 3.6.

Figura 3.6: Função Densidade de Probabilidades

Fonte: McGill, 1997.

Quando um fenômeno contínuo (p. ex., altura, peso ou tempo) é representado por um modelo matemático, a **probabilidade de que os valores da variável aleatória correspondente pertençam a um dado intervalo pode ser calculada**, como mostra a Figura 3.7, porém **a exata probabilidade de um valor específico é nula**.

Figura 3.7: Probabilidade de Variáveis Aleatórias Contínuas



Fonte: McGill, 1997.

Os modelos de distribuição contínua mais importantes são: a **normal** e a **exponencial**.

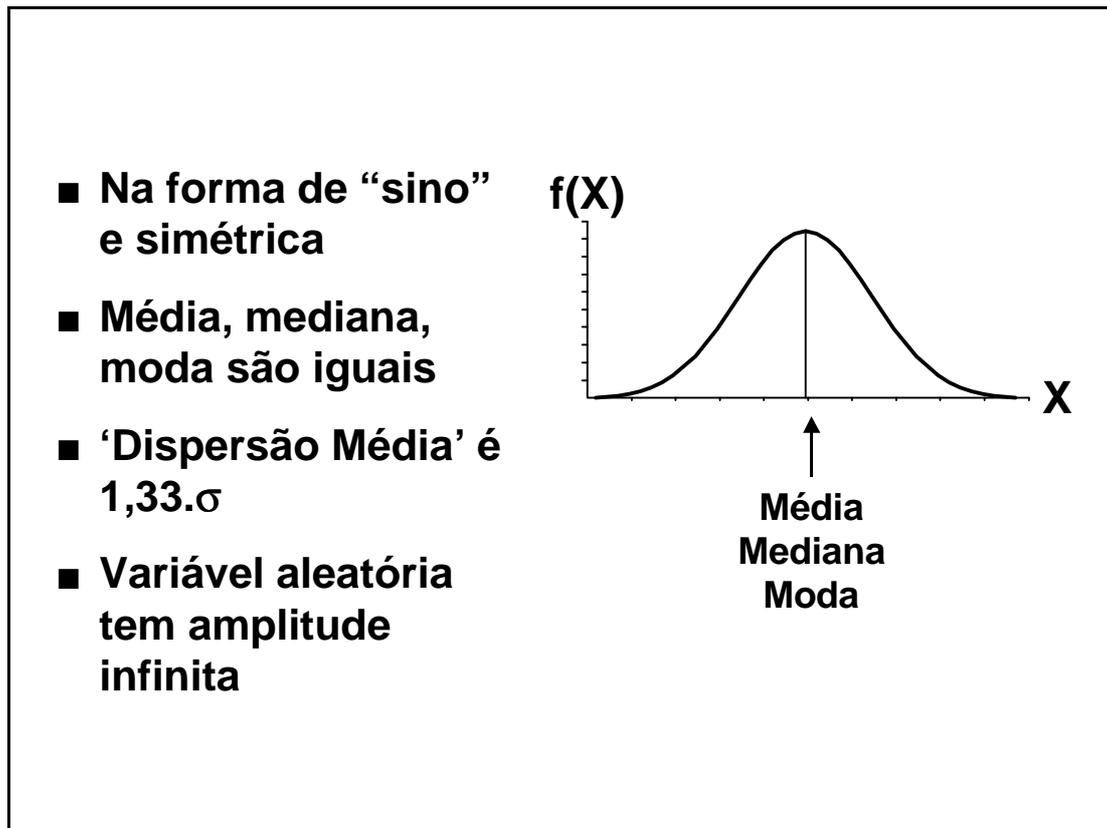
3.2.1. Distribuição Normal

Entre os modelos de distribuição contínua, a mais utilizada é a distribuição normal, uma vez que:

- inúmeros fenômenos contínuos parecem seguir essa distribuição;
- essa distribuição pode funcionar como uma aproximação de várias distribuições discretas de probabilidade;
- essa distribuição funciona como a base da inferência estatística clássica em decorrência da sua relação com o **Teorema do Limite Central**.

Essa distribuição possui as propriedades teóricas indicadas a seguir, ilustradas pela Figura 3.8:

- o formato é simétrico, semelhante a um sino;
- as medidas de tendência central (média, mediana, moda, meia amplitude e meia amplitude semi-interquartílica) são iguais;
- o intervalo semi-interquartílico, compreendendo a 50% das observações, corresponde ao intervalo $\left[\mu - \frac{2}{3} \cdot \sigma; \mu + \frac{2}{3} \cdot \sigma \right]$;
- a amplitude da variável aleatória representada é infinita $\{X \in \square \mid -\infty \leq X \leq +\infty\}$.

Figura 3.8: Distribuição Normal

Fonte: McGill, 1997.

Na prática, a função de distribuição de probabilidades de muitos fenômenos observados apenas se assemelha à distribuição normal. Isso ocorre quando a distribuição de probabilidades da população é aproximadamente normal ou quando a amostra selecionada não contém as características teóricas esperadas. Dessa forma, freqüentemente tem-se:

- a) polígonos quase simétricos;
- b) medidas de tendência central ligeiramente diferentes;
- c) intervalo semi-interquartílico ligeiramente diferente do intervalo $\left[\mu - \frac{2}{3}.\sigma; \mu + \frac{2}{3}.\sigma \right]$;
- d) amplitude finita, geralmente correspondendo ao intervalo $[\mu - 3.\sigma; \mu + 3.\sigma]$.⁶

No presente contexto, a FDP é definida pela seguinte equação:

⁶ Uma restrição comum é a não observação de valores negativos.

$$f(X) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot e^{-\frac{1}{2} \left(\frac{X - \mu}{\sigma} \right)^2}$$

Onde: $\pi = 3,14159$;⁷
 $e =$ n° de Euler;
 $X =$ variável aleatória;
 $\mu =$ média populacional;
 $\sigma =$ desvio-padrão populacional.

USANDO MICROSOFT EXCEL:

=DIST.NORM(x;média;desv_padrão;cumulativo) \Rightarrow calcula a probabilidade de que sejam observados valores menores do que X.

Onde: média = **m**

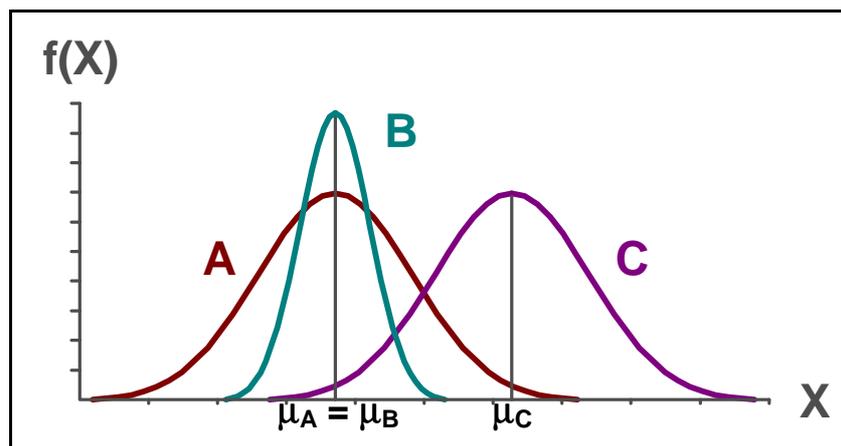
desvio-padrão = σ ;

cumulativo = $\begin{cases} \text{VERDADEIRO} = \text{função de distribuição cumulativa;} \\ \text{FALSO} = \text{função de distribuição de massa.} \end{cases}$

=INV.NORM(probabilidade;média;desv_padrão) \Rightarrow calcula o valor X correspondente a uma dada probabilidade acumulada.

Como e e π são constantes matemáticas, a FDP [$f(X)$] depende unicamente da média e do desvio-padrão populacionais (μ e σ). Cada combinação desses parâmetros gera uma distribuição de probabilidade distinta, como mostrado pela Figura 3.9.

Figura 3.9: Efeitos de Variações em σ e μ ($\sigma_A > \sigma_B$ e $\mu_C > \mu_B$)



Fonte: McGill, 1997.

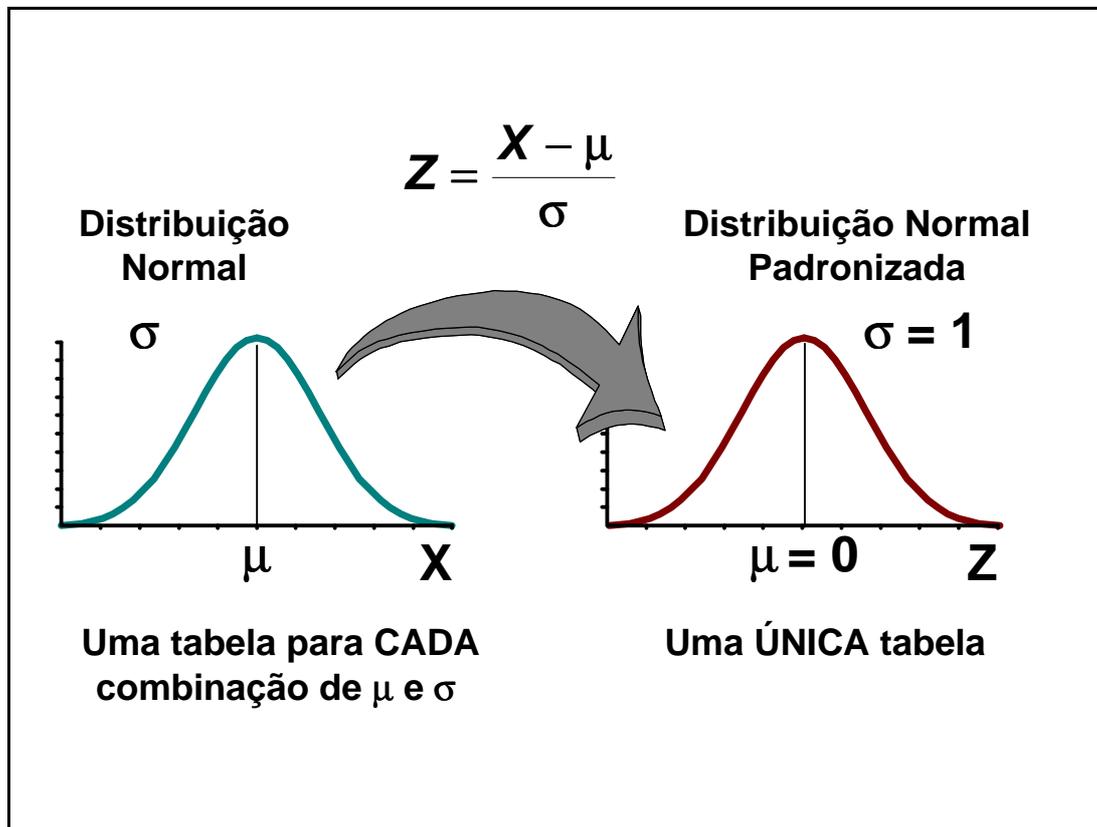
No intuito de tornar desnecessário o uso da equação exposta acima e em face da impossibilidade de que se elaborem infinitas tabelas de probabilidades, a variável aleatória (X) deve ser **normalizada**, ou seja, convertida em uma **variável aleatória normal padronizada** (Z) mediante a seguinte transformação, ilustrada pela Figura 3.10:

⁷ π radianos = 180 graus.

$$Z = \frac{X - m}{s}$$

USANDO MICROSOFT EXCEL: =PADRONIZAR(x;média;desv_padrao).

Figura 3.10: Padronização da Distribuição Normal



Fonte: McGill, 1997.

A nova variável tem média nula ($\mu = 0$), desvio-padrão unitário ($\sigma = 1$) e FDP definida pela seguinte equação:

$$f(Z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}Z^2}$$

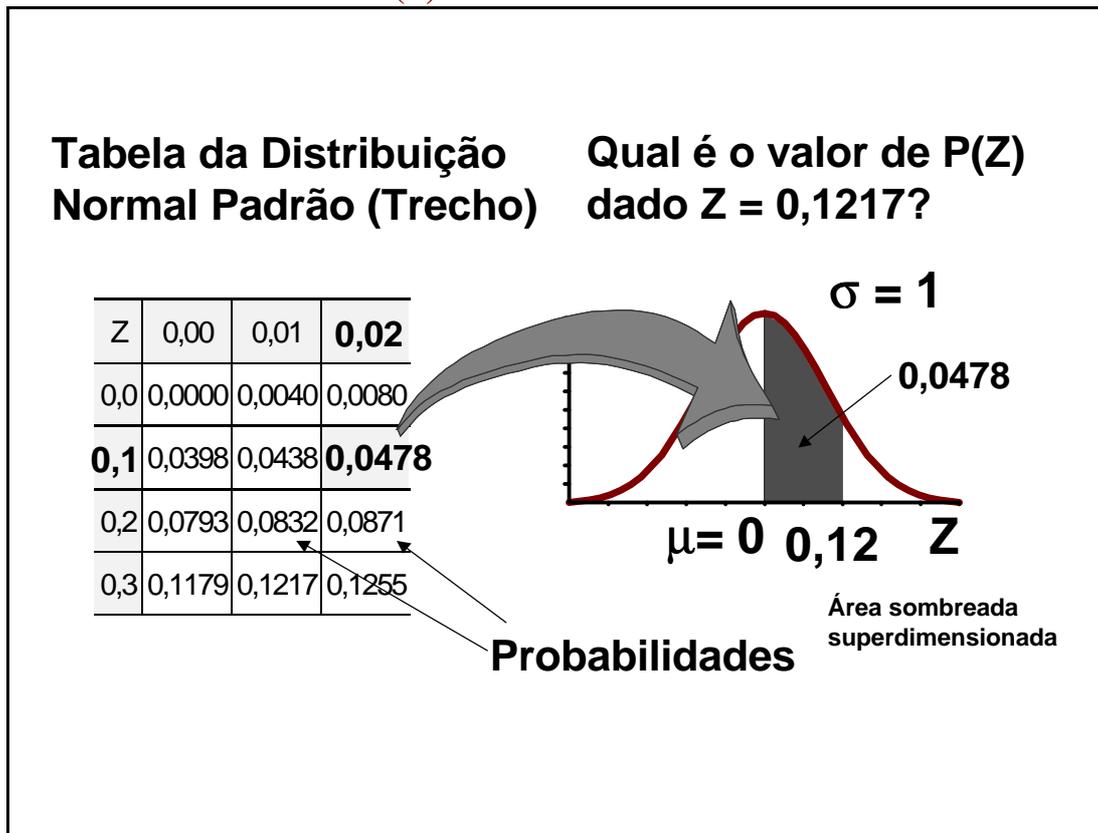
USANDO MICROSOFT EXCEL:

=DIST.NORMP(z) \Rightarrow calcula a probabilidade de que sejam observados valores menores do que Z.

=INV.NORMP(z) \Rightarrow calcula o valor Z correspondente a uma dada probabilidade acumulada.

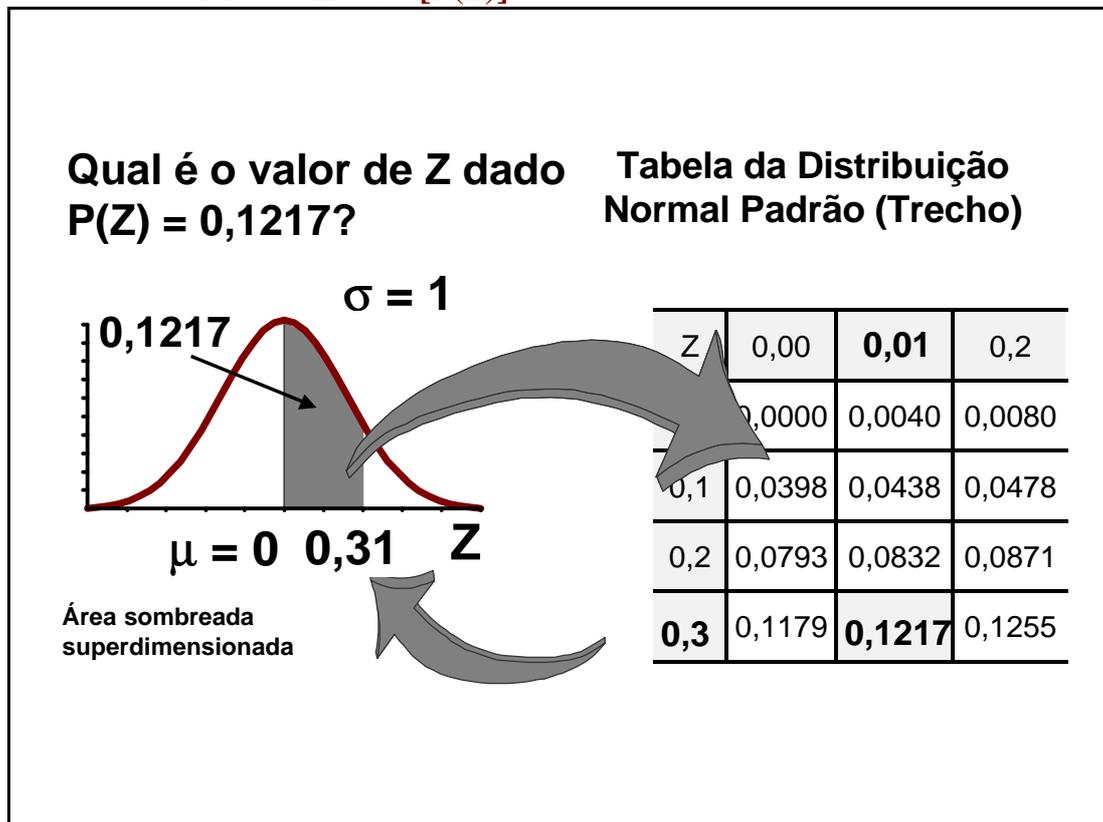
Portanto, todo conjunto de dados distribuídos normalmente pode ser transformado de tal modo que as probabilidades correspondentes podem ser obtidas a partir da tabela da distribuição normal padronizada (*vide* Anexo II), como ilustrado pelas Figuras 3.11 e 3.12.

Figura 3.11: Obtenção de Probabilidades [P(Z)] Dado um Valor Normal Padronizado (Z)



Fonte: McGill, 1997.

Figura 3.12: Obtenção de um Valor Normal Padronizado (Z) Dada uma Probabilidade [P(Z)]



Fonte: McGill, 1997.

Convém notar que o valor de uma observação selecionada aleatoriamente tem, em toda distribuição normal, uma probabilidade de 68,26% de estar contido no intervalo $[\mu - \sigma; \mu + \sigma]$ e de 99,73% de estar contido no intervalo $[\mu - 3\sigma; \mu + 3\sigma]$.

3.2.2. Distribuição Exponencial

A distribuição exponencial é amplamente utilizada em teoria das filas para estimar o **período de tempo entre dois “sucessos”** (p. ex., chegadas, partidas, atendimentos, etc.). O seu comportamento é definido por um único parâmetro: a quantidade média de “sucessos” por unidade de tempo (λ). A probabilidade de que o período de tempo entre dois “sucessos” seja igual ou menor a um valor X é dada pela seguinte equação:

$$P(\text{intervalo entre "sucessos"} \leq X) = 1 - e^{-\lambda \cdot X}.$$

USANDO MICROSOFT EXCEL:

=DISTEXPON(x;lambda;cumulativo).

Onde: lambda = λ ;

cumulativo = $\begin{cases} \text{VERDADEIRO} = \text{função de distribuição cumulativa;} \\ \text{FALSO} = \text{função de densidade de probabilidade.} \end{cases}$

O Exemplo 3.2 ilustra uma típica aplicação da distribuição exponencial.

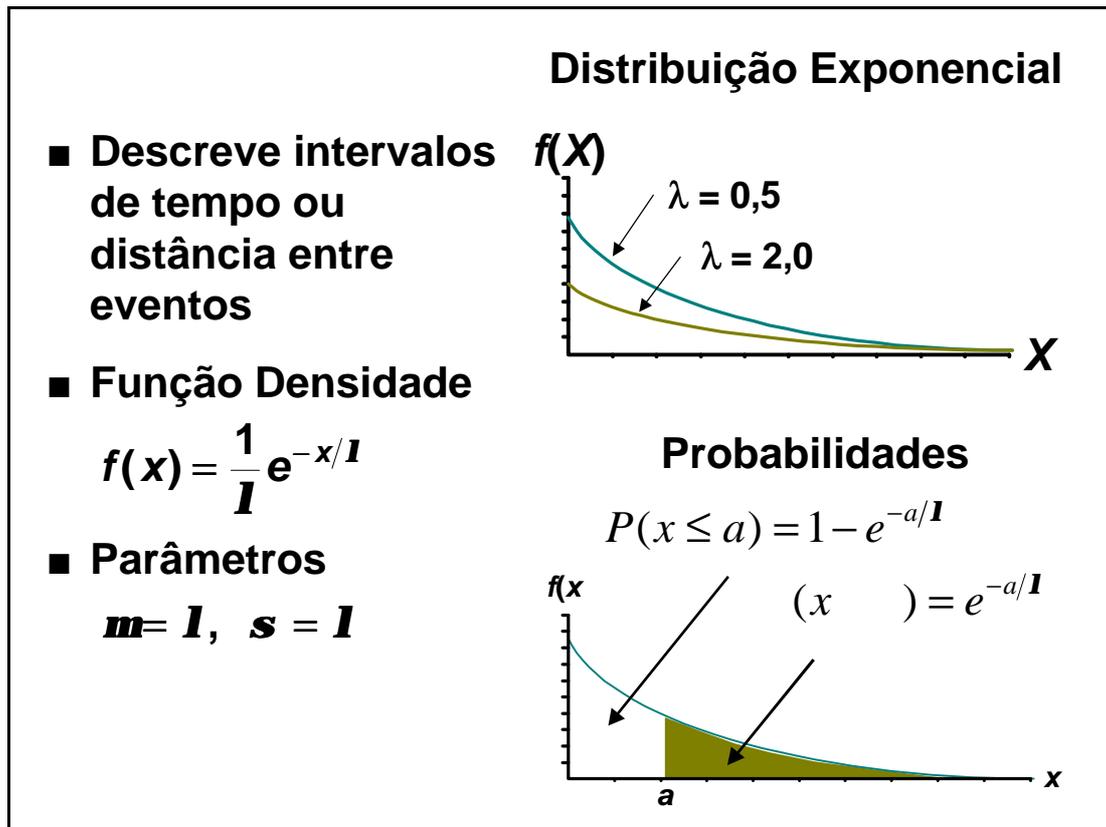
Exemplo 3.2: Aplicação da Distribuição Exponencial

Suponha-se um banco que atende, em média, 20 clientes por hora. Qual é a probabilidade de que o intervalo entre dois atendimentos seja, no máximo, de 6 minutos ou 0,1 hora?

Dados $\lambda = 20$ e $X = 0,1$, tem-se: $P(\text{intervalo entre "sucessos"} \leq 0,1) = 1 - e^{-20 \cdot 0,1} = 0,8647$. Portanto, há uma probabilidade de 86,47% de que o próximo atendimento ocorrerá dentro de 6 minutos.

A Figura 3.13, por sua vez, mostra os efeitos do parâmetro λ sobre o formato do gráfico da distribuição exponencial.

Figura 3.13: Distribuição Exponencial



Fonte: McGill, 1997.

4. DISTRIBUIÇÕES AMOSTRAIS

As análises de dados têm como um de seus principais objetivos o uso de estatísticas como a média e a proporção amostrais para estimar os parâmetros populacionais correspondentes. Desse modo, uma pesquisa sobre intenção de voto tem nos resultados amostrais apenas uma forma de obter estimativas acerca das intenções de todos os eleitores, enquanto que uma auditoria financeira poderá usar a média amostral para estimar o valor total de todas as operações constantes de um cadastro.

Do ponto de vista prático, uma única amostra com uma quantidade predeterminada de observações é extraída da população com o auxílio de um gerador de números aleatórios, que pode ser tanto uma tabela preparada previamente, como uma rotina de computador.

UTILIZANDO O *MICROSOFT EXCEL*:

- OPÇÃO “FERRAMENTAS/ANÁLISE DE DADOS...” (FIGURAS A.1, A.9 & A.10, ANEXO I) \Rightarrow seleciona uma amostra aleatória simples com reposição;
- =ALEATÓRIOENTRE(inferior;superior) \Rightarrow gera um número entre os valores mínimo (inferior) e máximo (superior) especificados, admitindo repetições;
- =ALEATÓRIO \Rightarrow gera um número distribuído uniformemente no intervalo $[0; 1[$.

Do ponto de vista teórico, contudo, a estimação de parâmetros populacionais por meio de estatísticas amostrais deveria envolver o **exame de todas as possíveis amostras**. Se esse exame fosse efetivamente feito, a distribuição dos resultados obtidos corresponderia à **distribuição amostral**. Note-se que a quantidade de possíveis amostras corresponde:

a) no caso de amostras com reposição, ao total de arranjos com repetição de N , n a n : $AR_{N,n} = N^n$;

b) no caso de amostras sem reposição, ao total de combinações de N , n a n :

$$C_{N,n} = \frac{N!}{n!(N-n)!}.$$

USANDO *MICROSOFT EXCEL*: =COMBIN(núm;núm_escolhido).

Onde: núm = N ;

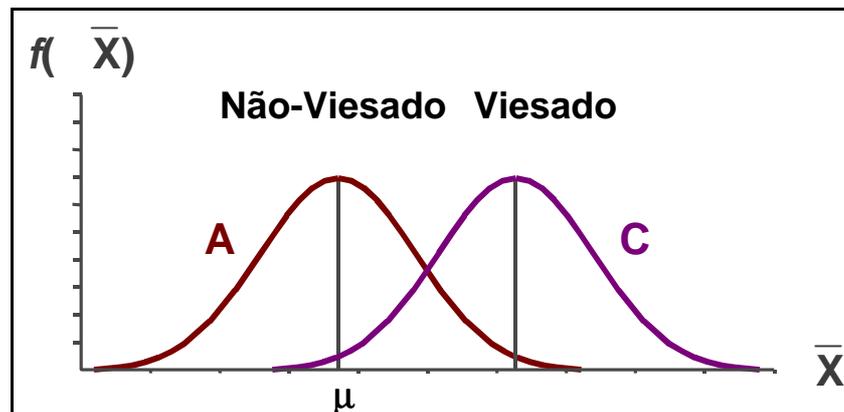
núm_escolhido = n .

4.1. Distribuição de Probabilidades das Médias Amostrais

A medida de tendência central mais amplamente usada é a média aritmética. No caso de populações cuja distribuição de probabilidades seja normal, a média apresenta as seguintes propriedades matemáticas, ilustradas pelas Figuras 4.1, 4.2 e 4.3:

- a) **não-viesada**: a média de todas as possíveis médias amostrais $(\frac{\sum_{i=1}^m \bar{X}_i}{m})$, onde m é dado por $AR_{N,n}$, no caso de amostras com reposição, ou por $C_{N,n}$, no caso de amostras sem reposição) é igual à média populacional (μ);⁸

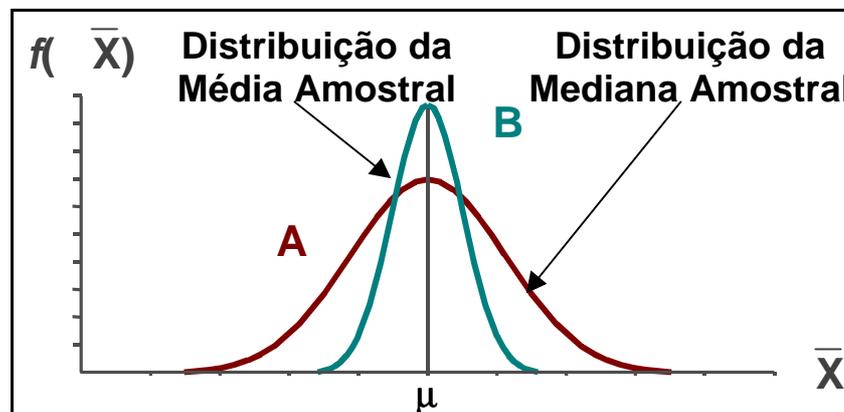
Figura 4.1: Medida Não-Viesada



Fonte: McGill, 1997.

- b) **eficiente**: comparando-se amostras, constata-se que a média é a medida de tendência central mais estável, proporcionando, em geral, as melhores estimativas para a média populacional;

Figura 4.2: Medida Eficiente

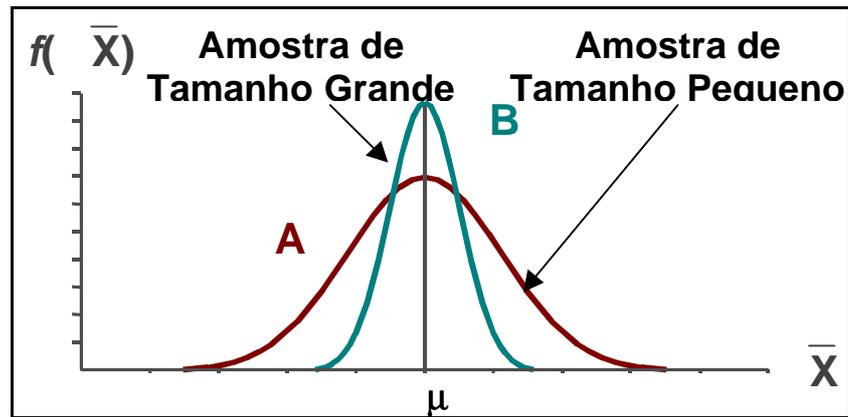


Fonte: McGill, 1997.

- c) **consistente**: aumentando-se o tamanho da amostra, observa-se uma redução na variação das médias amostrais em torno da média populacional.

⁸ Assim, ainda que não se conheça o quanto uma dada média amostral aproxima-se da média populacional, sabe-se, ao menos, que a **média de todas as possíveis médias amostrais é igual à média populacional**.

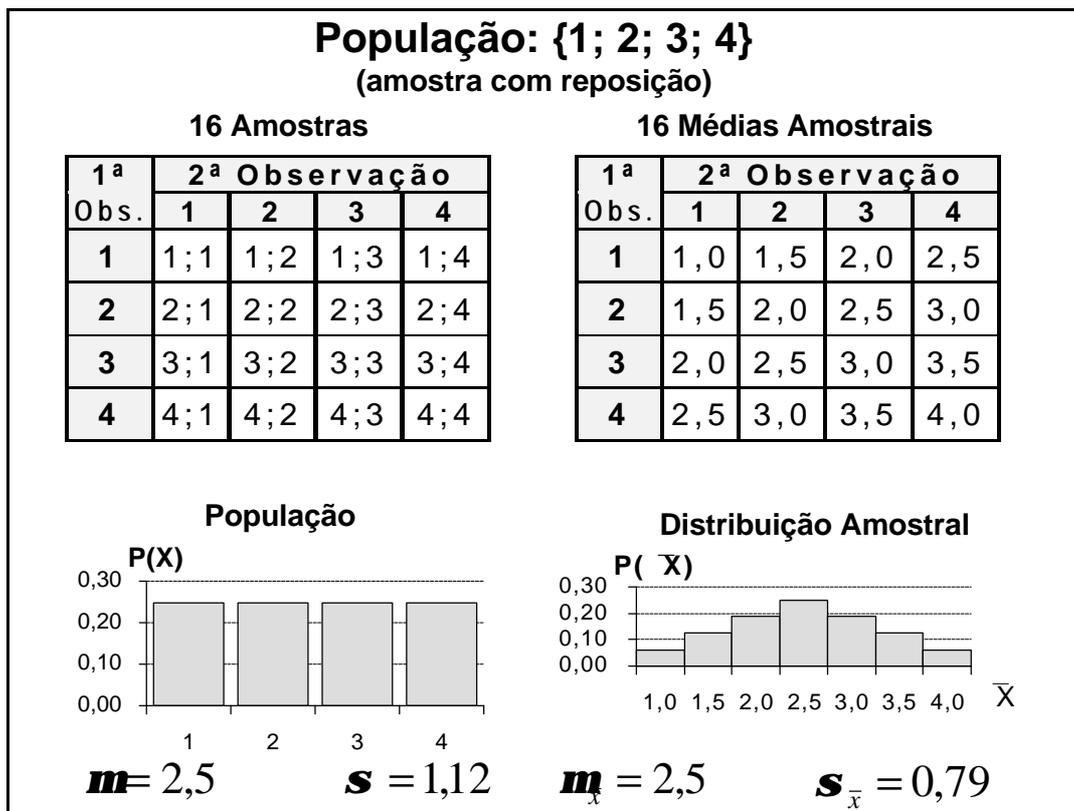
Figura 4.3: Medida Consistente



Fonte: McGill, 1997.

Cabe frisar que a **Lei dos Grandes Números** faz com que as médias amostrais apresentem uma menor variância do que as observações contidas na população. Efetivamente, os efeitos da eventual seleção de valores extremos para as amostras são compensados pelo próprio cálculo da média, como mostra a Figura 4.4.

Figura 4.4: Exemplo de Distribuição de Médias Amostrais



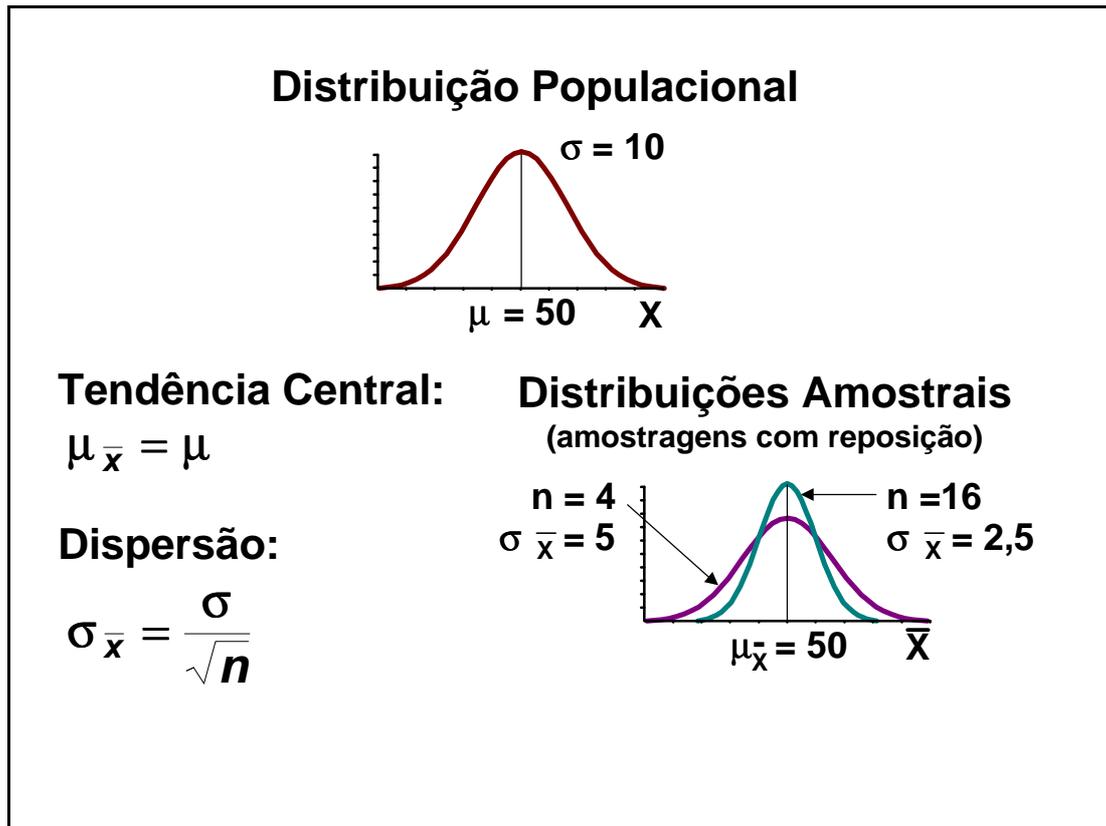
Fonte: McGill, 1997.

Ressalte-se que quanto maior for o tamanho da amostra, menores são os efeitos da seleção de valores extremos. Estatisticamente, esse fenômeno é expresso pelo desvio-padrão da média amostral, denominado de **erro-padrão** da média ($s_{\bar{X}}$). Nas amostras com reposição, o erro-padrão é definido pela seguinte equação:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Portanto, o erro-padrão da média ($s_{\bar{x}}$) diminui à medida que cresce o tamanho da amostra (n), como mostrado pela Figura 4.5.

Figura 4.5: Exemplo do Impacto de n em Amostras Extraídas de Populações Distribuídas Normalmente



Fonte: McGill, 1997.

Assim, nas amostras com reposição extraídas de populações normalmente distribuídas, com média μ e desvio-padrão σ , tem-se que as distribuições de probabilidades das médias amostrais também são, para todo n , normais, com média $\mu_{\bar{x}} = \mu$ e erro-padrão $s_{\bar{x}}$. Além do mais, quanto maior n , maior a concentração das médias amostrais (\bar{X}_i) em torno da média populacional (μ).

O Exemplo 4.1 ilustra uma das possíveis aplicações da distribuição de probabilidades das médias amostrais.

Exemplo 4.1: Aplicação da Distribuição de Probabilidades das Médias Amostrais

Suponha-se que uma máquina empacotadora de caixas de cereais tenha sido configurada de tal modo que as quantidades depositadas nas caixas sejam normalmente distribuídas, com média de 368 g e desvio-padrão de 15 g. Caso seja selecionada uma amostra de 25 das milhares de caixas empacotadas diariamente, qual é a probabilidade de que a média amostral fique entre 365 e 368 g?

Como a população é normalmente distribuída, tem-se que:

$$Z = \frac{\bar{X} - \mathbf{m}_{\bar{x}}}{\mathbf{s}_{\bar{x}}} = \frac{\bar{X} - \mathbf{m}}{\frac{\mathbf{s}}{\sqrt{n}}} = \frac{365 - 368}{\frac{15}{\sqrt{25}}} = -1,00.$$

A área sob a curva normal padronizada no intervalo $[-1; 0]$ é igual a 0,3413. Portanto, 34,13% de todas as possíveis amostras de 25 caixas têm médias amostrais de 365 a 368 g.

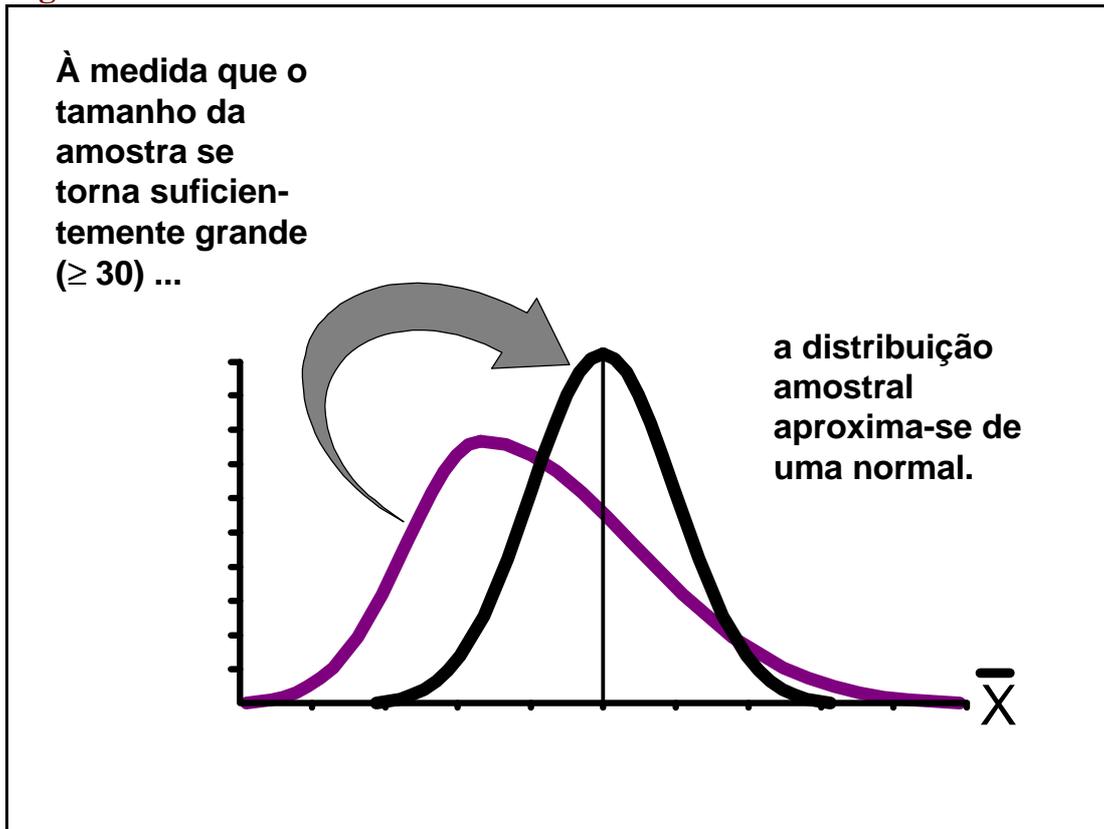
O percentual calculado acima não deve ser confundido com a probabilidade de que um item qualquer da população tenha de 365 a 368 g. Essa probabilidade deve ser obtida da seguinte maneira:

$$Z = \frac{X - \mathbf{m}}{\mathbf{s}} = \frac{365 - 368}{15} = -0,20.$$

A área sob a curva normal padronizada no intervalo $[-0,2; 0]$ é igual a 0,0793. Portanto, apenas 7,93% das caixas têm uma massa esperada de 365 a 368 g. Dessa forma, as médias amostrais estão mais concentradas em torno da média populacional do que as massas de cada caixa. Isso se deve ao próprio método de cálculo da média, que atenua os efeitos dos valores extremos.

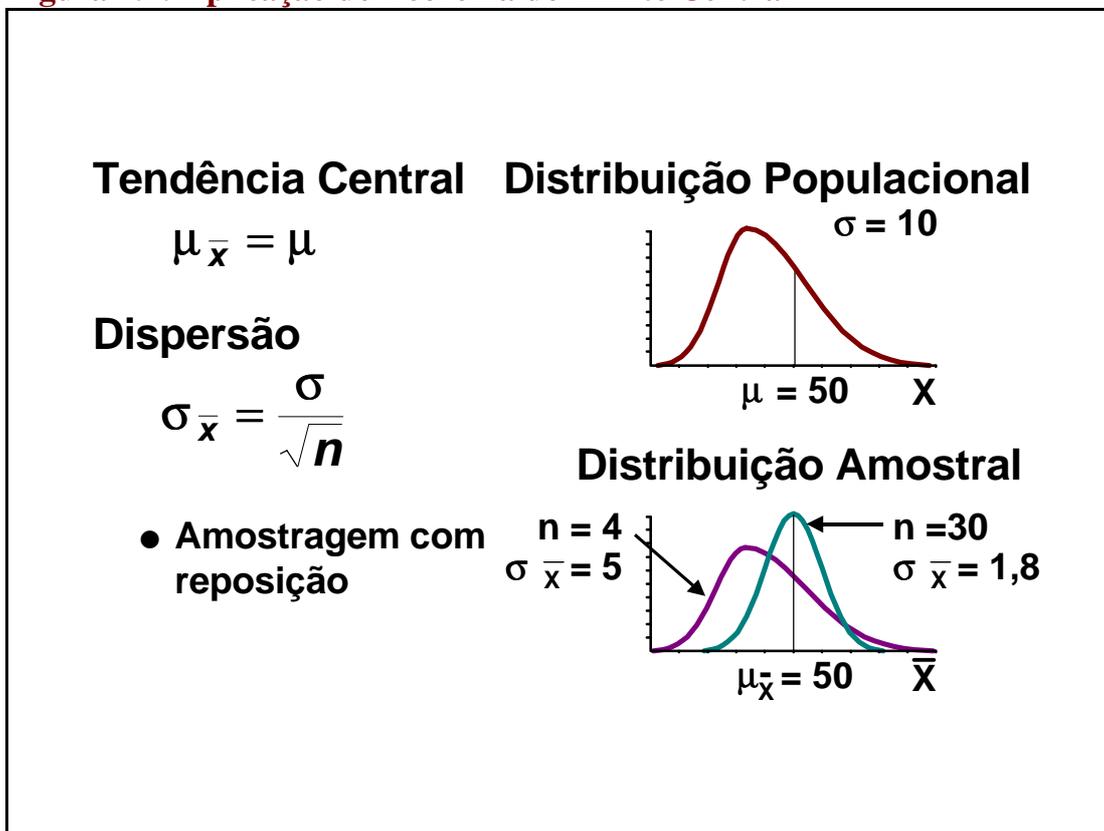
Quando não houver razões para crer que a população examinada é distribuída normalmente, deve-se recorrer ao Teorema do Limite Central. Segundo esse teorema, como mostram as Figuras 4.6 e 4.7, quando o tamanho da amostra é suficientemente grande, a distribuição das probabilidades das médias amostrais aproxima-se de uma distribuição normal. Isso ocorre independentemente da forma da distribuição da população. Como regra geral, tem-se que **amostras com trinta ou mais elementos apresentam distribuições das médias amostrais que assemelham-se a distribuições normais**. O teorema em questão, contudo, pode ser usado em amostras ainda menores caso algumas características da população sejam conhecidas. No caso de populações com distribuições quase simétricas, p. ex., pode-se utilizar amostras com quinze elementos.

Figura 4.6: Teorema do Limite Central



Fonte: McGill, 1997.

Figura 4.7: Aplicação do Teorema do Limite Central



Fonte: McGill, 1997.

4.2. Distribuição de Probabilidades das Proporções Amostrais

Nas variáveis categóricas compostas por itens que podem ser classificados como tendo ou não uma dada característica, os dois possíveis resultados podem ser associados, respectivamente, aos valores “1” e “0”. Nesse caso, a média de uma amostra (\bar{X}) de tamanho n é definida pela razão entre a soma dos valores “1’s” e “0’s” constantes da amostra e o seu tamanho. Isso corresponde à proporção dos elementos da amostra, ou proporção amostral (p_s), que contém a característica examinada, também chamada de “sucesso”, ou seja:

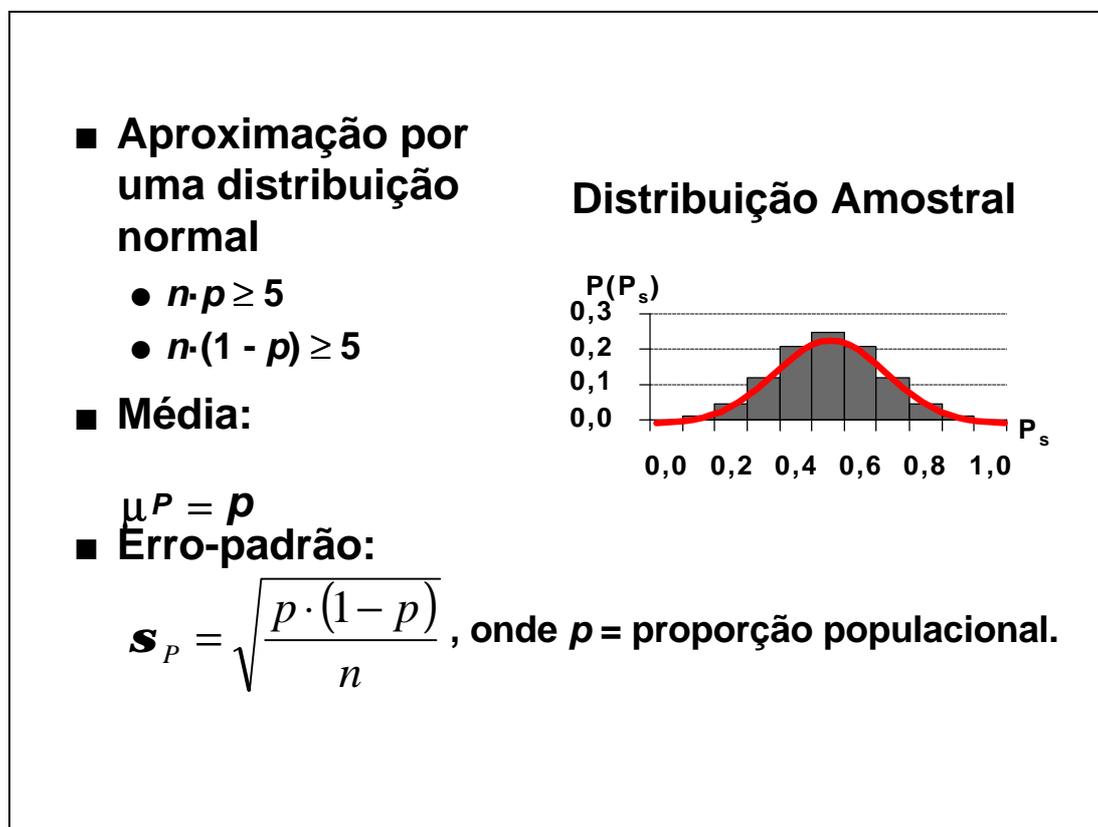
$$\bar{X} \Rightarrow p_s = \frac{\text{qtde de sucessos}}{\text{tamanho da amostra}} = \frac{X}{n}.$$

Dessa forma, a proporção amostral (p_s) está necessariamente contida no intervalo $[0, 1]$. O erro-padrão da proporção (s_{p_s}), por sua vez, é definido pela seguinte equação:

$$s_{p_s} = \sqrt{\frac{p \cdot (1-p)}{n}}.$$

A distribuição de probabilidades das proporções amostrais, por fim, segue a distribuição binomial. **Essa distribuição, entretanto, aproxima-se de uma distribuição normal quando $(n \cdot p)$ e $[n \cdot (1-p)]$ são superiores ou iguais a cinco.**

Figura 4.8: Distribuição de Proporções Amostrais



Fonte: McGill, 1997.

Exemplo 4.2: Aplicação da Distribuição de Probabilidades das Proporções Amostrais

Suponha-se um banco no qual se saiba que 40% de todos os correntistas têm várias contas. Caso seja selecionada uma amostra de 200 correntistas, qual é a probabilidade de que a proporção amostral de correntistas com várias contas fique entre 0,40 e 0,43?

Como $(n.p) = (200.0,4) = 80$ e $[n.(1-p)] = [200.(1-0,40)] = 120$, então a distribuição de probabilidades da proporção amostral é semelhante a uma distribuição normal. Assim, tem-se que:

$$Z = \frac{\bar{X} - m_{\bar{x}}}{s_{\bar{x}}} = \frac{\bar{X} - m}{\frac{s}{\sqrt{n}}} \Rightarrow Z = \frac{p_s - p}{\sqrt{\frac{p.(1-p)}{n}}} = \frac{0,43 - 0,40}{\sqrt{\frac{0,40.(1-0,40)}{200}}} = 0,87.$$

A área sob a curva normal padronizada no intervalo $[0; 0,87]$ é igual a 0,3078. Portanto, é de 30,78% a probabilidade de que, em uma amostra com 200 correntistas, de 40 a 43% deles tenham várias contas.

4.3. Amostragem em Populações Finitas

O Teorema do Limite Central e os erros-padrão da média e da proporção baseiam-se na premissa de que as amostras são selecionadas com reposição, porém, em geral, as pesquisas não apenas são conduzidas sem reposição, como envolvem uma porcentagem significativa da população. Assim, **quando $\frac{n}{N}$ é superior a 0,05, deve-se usar um Fator de Correção para Populações Finitas (CPF) no cálculo dos erros-padrão da média e da proporção.** Esse fator é definido pela equação indicada a seguir:

$$CPF = \sqrt{\frac{N - n}{N - 1}}.$$

Dessa forma, as equações referentes aos erros-padrão da média e da proporção passam a ser, respectivamente:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N - n}{N - 1}} \text{ e } s_{p_s} = \sqrt{\frac{p.(1-p)}{n}} \cdot \sqrt{\frac{N - n}{N - 1}}.$$

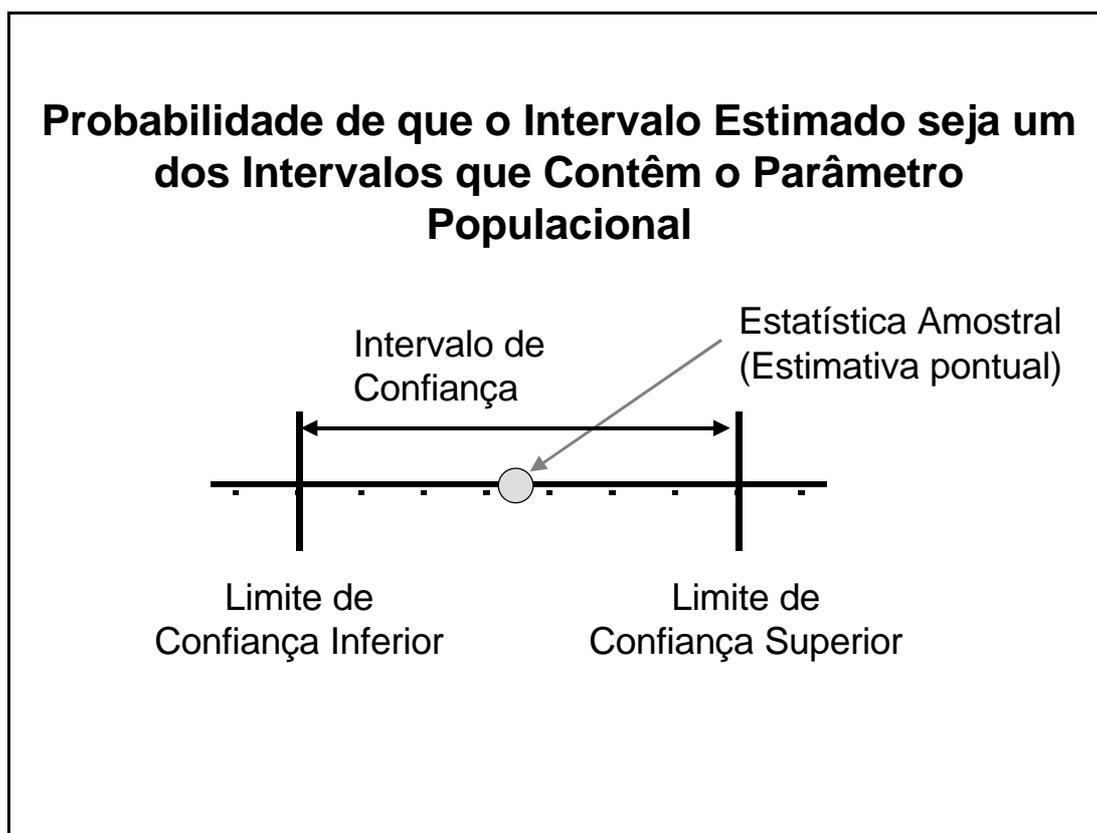
Como o fator em questão é, em geral, inferior a 1, os erros-padrão corrigidos são menores do que os não-corrigidos, o que é coerente com o fato de que a amostra abrange uma fração significativa da população.

5. ESTIMAÇÃO

Como ilustrado pela Figura 5.1, há dois tipos principais de estimativa: a pontual e a do intervalo. A **estimativa pontual** equívale a um único valor, usado como estimativa do parâmetro populacional. A média e a variância amostrais (\bar{X} e S^2) são, respectivamente, estimativas pontuais da média e da variância populacionais (μ e σ^2).

Como, contudo, as estatísticas amostrais variam de amostra para amostra, é importante que se leve em consideração a distribuição amostral de probabilidades, o que é feito mediante a **estimativa do intervalo**, que deve conter o parâmetro populacional. Ademais, associado ao intervalo estimado há uma probabilidade, ou nível de confiança, de que seja um dos intervalos que efetivamente contêm o parâmetro populacional.

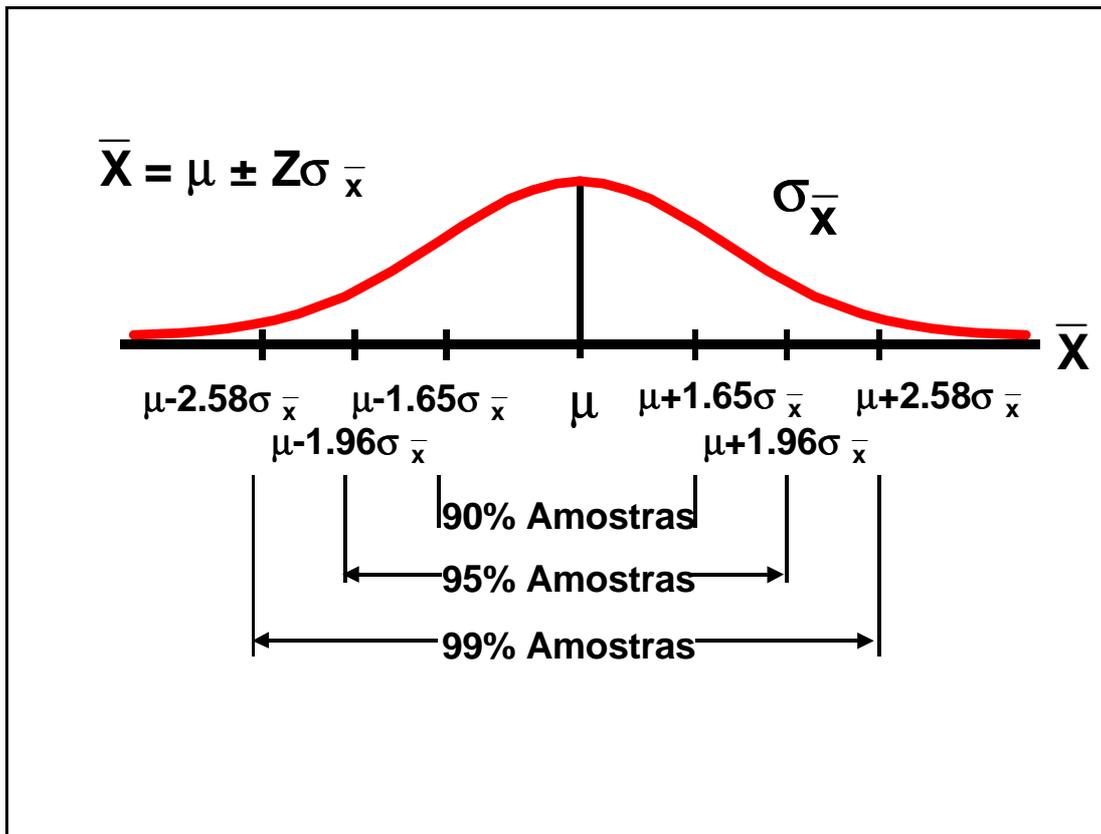
Figura 5.1: Elementos de uma Estimativa Probabilística



Fonte: McGill, 1997.

A Figura 5.2 mostra como as **estimativas dos intervalos de confiança** associadas às médias amostrais variam conforme diferentes níveis de confiança, obtidos a partir da tabela da distribuição normal padronizada.

Figura 5.2: Intervalos de Confiança das Médias Amostrais



Fonte: McGill, 1997.

5.1. Estimativa do Intervalo da Média com σ Conhecido

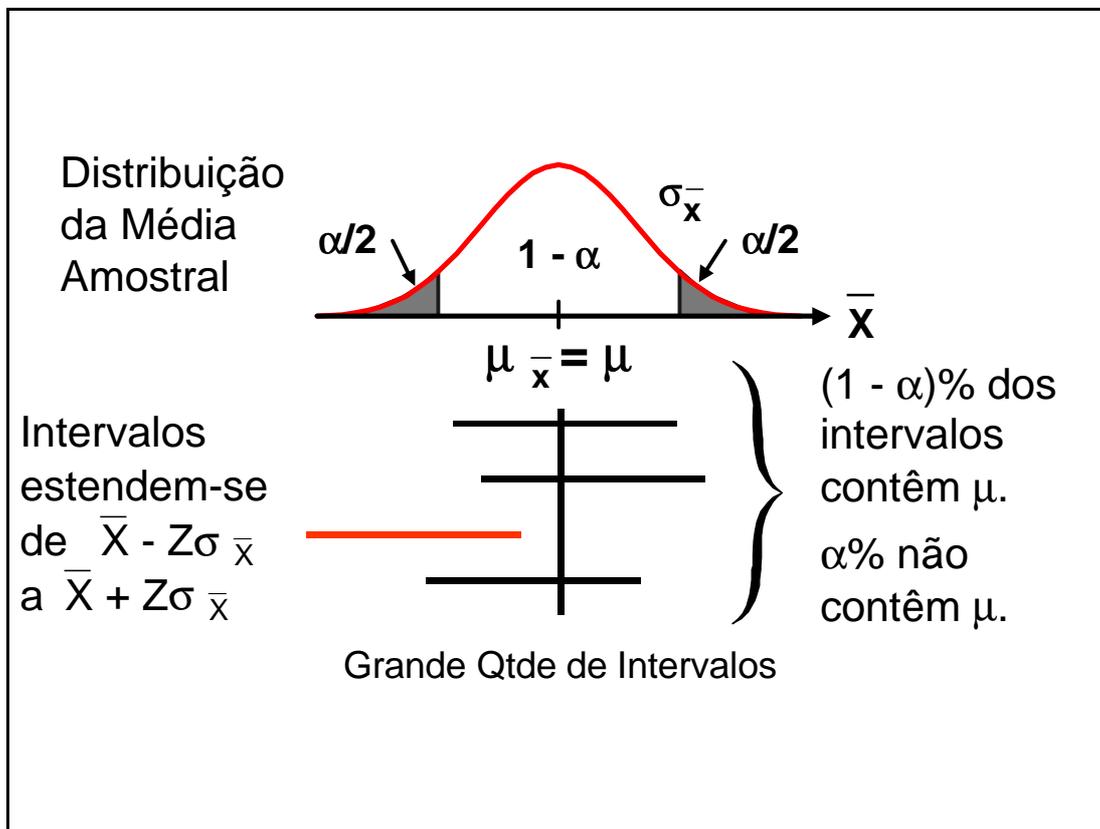
No capítulo anterior, verificou-se que o percentual de médias amostrais contidas em qualquer intervalo centrado na média populacional pode ser obtido a partir do Teorema do Limite Central ou de informações disponíveis acerca da distribuição da população. O presente item fará o contrário: com base nos resultados de uma única amostra, serão extraídas conclusões sobre a população.

Do ponto de vista prático, a média populacional (μ) não é conhecida. Dessa forma, quando o **desvio-padrão** (σ) é **sabido**, em vez do intervalo $\left[\mu - Z \left(\frac{\sigma}{\sqrt{n}} \right); \mu + Z \left(\frac{\sigma}{\sqrt{n}} \right) \right]$, tem-se o intervalo $\left[\bar{X} - Z \left(\frac{\sigma}{\sqrt{n}} \right); \bar{X} + Z \left(\frac{\sigma}{\sqrt{n}} \right) \right]$, com $(1 - \alpha)\%$ das médias amostrais (\bar{X}_m), com o valor $(1 - \alpha)$ correspondendo ao dobro da área sob a curva da distribuição normal padronizada contida no intervalo $[0; Z]^9$, a qual é simétrica em relação à área contida no intervalo $[-Z; 0]$.

⁹ $Z =$ **valor crítico** da distribuição normal padronizada associado à metade do complemento do nível de confiança desejado, ou seja: $\frac{(1 - \alpha)}{2}$.

Considerando-se, para um dado n e um dado α , todas as possíveis médias amostrais e os respectivos intervalos de confiança, tem-se que $(1 - \alpha)\%$ desses intervalos contêm a média populacional, como mostrado pela Figura 5.3.

Figura 5.3: Distribuição de Probabilidades das Médias Amostrais Conhecendo-se o Desvio-Padrão Populacional



Fonte: McGill, 1997.

UTILIZANDO O MICROSOFT EXCEL:

OPÇÃO “FERRAMENTAS/ANÁLISE DE DADOS...” (VIDE FIGURAS A.1, A11 & A.12, ANEXO I) \Rightarrow gera estimativa do intervalo de confiança com σ conhecido.

Assim, um intervalo de confiança de 95% ($\alpha = 5\%$ e $Z = 1,96$) significa que o parâmetro populacional está contido em 95% de todos os possíveis intervalos, para um dado n . Dessa forma, ainda que não se tenha certeza de que um dado intervalo de confiança, obtido a partir de uma amostra específica, contenha o parâmetro desejado, sabe-se que essa condição é satisfeita por 95% de todas as amostras de tamanho n .

Destaque-se que, para um dado n , maior o nível de confiança desejado, maior o intervalo em torno da média amostral, tonando a estimativa obtida menos precisa e, por extensão, menos útil.

Uma possível aplicação das estimativas em questão encontra-se ilustrada no Exemplo 5.1.

Exemplo 5.1: Aplicação da Estimativa do Intervalo da Médias com σ Conhecido

Suponha-se uma fábrica de folhas de papel, cuja produção é normalmente distribuída, com comprimento esperado de 28 cm e desvio-padrão conhecido de 0,05 cm. Caso seja selecionada uma amostra de 100 folhas e a média obtida tenha sido de 27,995 cm, qual é a estimativa do intervalo da média para um nível de confiança de 95%?

Como a população é normalmente distribuída, tem-se que:

$$\frac{\alpha}{2} = \frac{(1 - 0,95)}{2} = 0,025 \Rightarrow Z = 1,96;$$

$$\bar{X} \pm Z \cdot \left(\frac{\sigma}{\sqrt{n}} \right) = 27,995 \pm 1,96 \cdot \left(\frac{0,05}{\sqrt{100}} \right) = 27,995 \pm 0,01 \Rightarrow 27,985 \leq \mu \leq 28,005 .$$

Portanto, a amostra selecionada indica que o intervalo [27,985; 28,005] contém a média populacional, com 95% de nível de confiança de que esse seja um dos intervalos que efetivamente contém o parâmetro desejado.

Concluindo, como o comprimento esperado está contido no intervalo estimado, a hipótese de que a fábrica está operando corretamente é aceita. Esse procedimento corresponde a um **Teste de Hipótese**, no qual a **Hipótese Nula** (H_0) equivale à aceitação do valor esperado, enquanto que a **Hipótese Alternativa** (H_1) equivale à rejeição. Se, neste exemplo, o comprimento esperado estivesse aquém ou além do intervalo estimado, aceitar-se-ia a hipótese alternativa, com risco de 5% de que se estivesse rejeitando um processo de fabricação correto.¹⁰

5.2. Estimativa do Intervalo da Média com σ Desconhecido

Assim como a média populacional costuma ser desconhecida, o **desvio-padrão** também é, em geral, **não-conhecido**. Dessa forma, o intervalo de confiança deve ser estimado com base na média e no desvio-padrão amostrais. Para tanto, deve-se recorrer não à distribuição normal, mas sim à distribuição de “*t*” de Student (*vide* Anexo II).

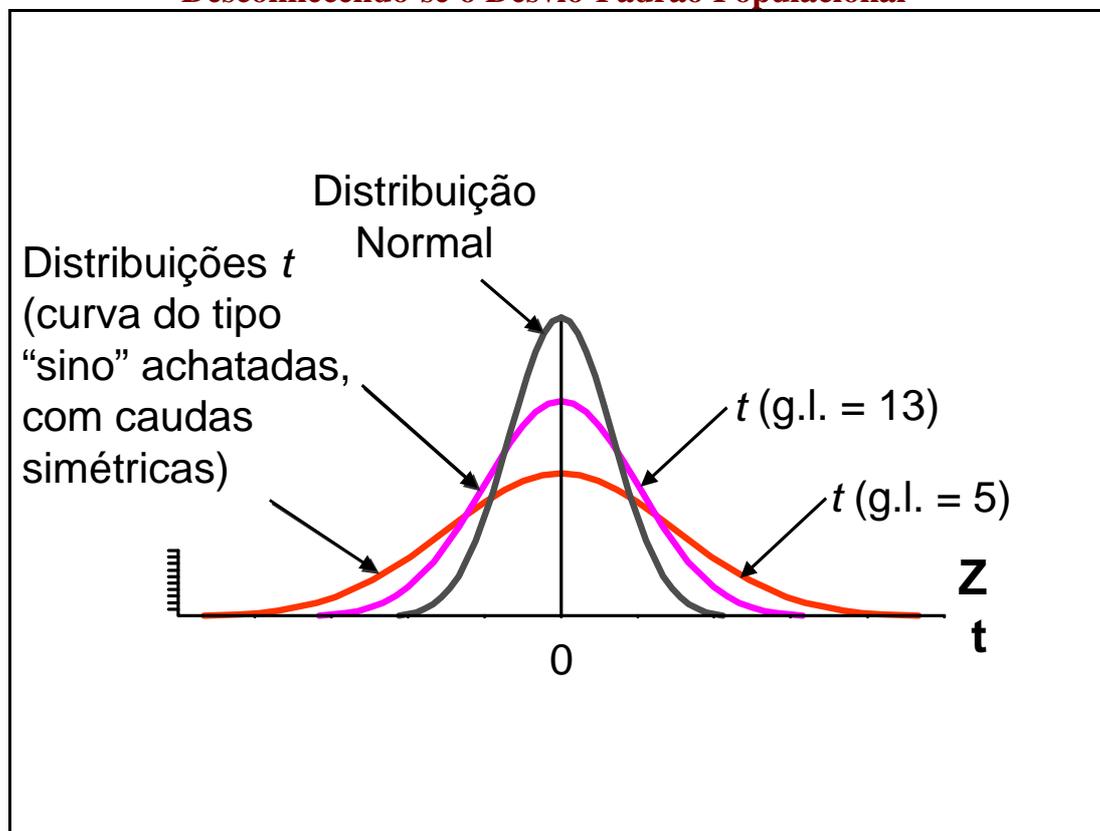
No caso de populações distribuídas normalmente, as probabilidades da estatística $\left[\frac{(\bar{X} - \mu)}{\frac{S}{\sqrt{n}}} \right]$ seguem uma distribuição “*t*” de Student com $(n-1)$ graus de liberdade (g.l.).

Aparentemente, a distribuição “*t*” de Student e a distribuição normal são semelhantes – ambas são simétricas e têm forma de sino. A área definida pela primeira curva, contudo, é menor no centro e maior nas extremidades do que no caso da segunda. À

¹⁰ Há dois tipos de erro inerentes ao processo de **teste de significância**. Há um risco de que H_0 seja julgada falsa ainda que seja verdadeira. A probabilidade de que se cometa esse erro, conhecido como **erro do Tipo I**, é igual ao nível de significância do teste (i.e., α). Também há o risco de que H_0 seja julgada verdadeira embora seja falsa. A probabilidade de que se cometa esse erro, conhecido como **erro do Tipo II**, é definido pelo símbolo β .

medida que aumentam os graus de liberdade (g.l.)¹¹, porém essa diferença diminui, como mostrado pela Figura 5.4.

Figura 5.4: Distribuição de Probabilidades das Médias Amostrais Desconhecendo-se o Desvio-Padrão Populacional



Fonte: McGill, 1997.

Como o grau de liberdade (g.l.) da estatística desejada é obtido a partir do tamanho da amostra, tem-se que, **para n superior ou igual a 120, as distribuições normal e “ t ” de Student são equivalentes.**

Na prática, como mostra a Figura 5.5, desde que a amostra seja numerosa o suficiente e desde que a população não seja muito viesada, a distribuição “ t ” de Student pode ser usada na estimação do intervalo de confiança, bastando substituir a estatística Z pela estatística t_{n-1} , ou seja:

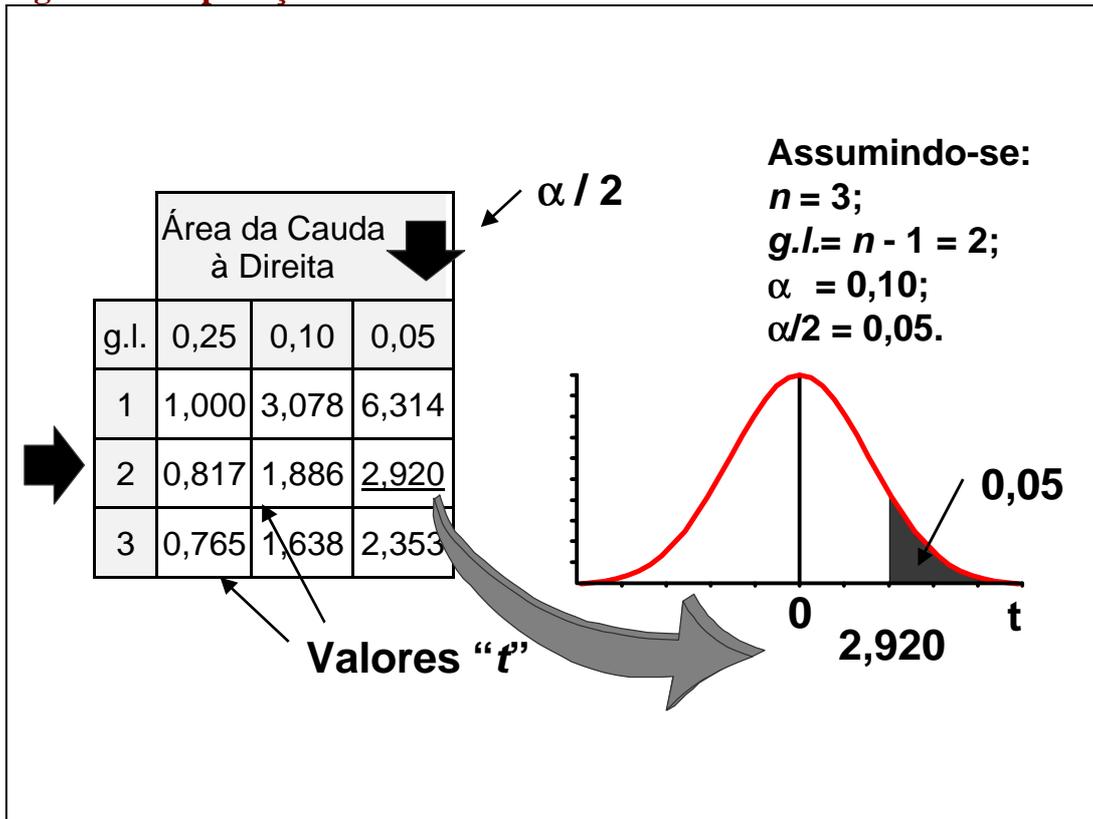
$$\bar{X} - t_{n-1} \cdot \left(\frac{S}{\sqrt{n}} \right) \leq \mu \leq \bar{X} + t_{n-1} \cdot \left(\frac{S}{\sqrt{n}} \right).^{12}$$

UTILIZANDO O MICROSOFT EXCEL: VIDE TABELA A.2, ANEXO I, COLUNAS “C” E “D” ATÉ A LINHA 13.

¹¹ Medida que exprime a quantidade de observações independentes de uma variável. Assim, dada uma amostra com n observações $X_1, X_2, X_3, \dots, X_n$ e média \bar{X} , somente $(n-1)$ observações podem assumir quaisquer valores sem que a restrição definida por \bar{X} seja violada.

¹² t_{n-1} = **valor crítico** da distribuição “ t ” de Student com $(n-1)$ graus de liberdade associado ao complemento do nível de confiança desejado.

Figura 5.5: Aplicação da Tabela “t” de Student



Fonte: McGill, 1997.

Nota: Diferentemente da distribuição “t” de Student exibida no Anexo II, a figura acima refere-se a um teste unicaudal.

Uma possível aplicação das estimativas em questão encontra-se ilustrada no Exemplo 5.2.

Exemplo 5.2: Aplicação da Estimativa do Intervalo para Médias com σ Desconhecido

No exemplo anterior, desconhecendo-se o desvio-padrão populacional, qual é o novo valor crítico?

Dados que a população é normalmente distribuída, o nível de confiança desejado é de 95% e a amostra tem 100 observações, e como o desvio-padrão populacional é desconhecido e $n < 120$, tem-se que:

$$\left. \begin{array}{l} (n - 1) = 100 - 1 = 99 \\ \mathbf{a} = 1 - 0,95 = 0,05 \end{array} \right\} \Rightarrow t_{99;0,05} = 1,984.^{13}$$

O novo valor crítico é um pouco superior ao valor anterior: 1,984 em vez de 1,96. O uso da mesma medida para o desvio-padrão resultaria em um intervalo de confiança um pouco maior no presente exemplo. Ou seja, a nova estimativa é menos precisa que a anterior em decorrência da menor disponibilidade de informações.

¹³ O uso de α no presente exemplo e de $\left(\frac{\mathbf{a}}{2}\right)$ no anterior deve-se ao modo como os percentuais de cada tabela são apresentados.

5.3. Estimativa do tipo “*Bootstrapping*”

Embora as estimativas de intervalos de confiança sejam amplamente utilizados para fazer inferências acerca dos parâmetros populacionais, as hipóteses subjacentes a essas estimativas nem sempre correspondem à realidade. Os intervalos calculados acima, p. ex., baseiam-se na hipótese de que as amostras são extraídas de populações normalmente distribuídas. Ainda que os intervalos de confiança não sejam muito sensíveis a pequenos relaxamentos dessa hipótese, a presença de um substancial nível de não-normalidade na população, especialmente quando o tamanho da amostra é pequeno, pode gerar intervalos imprecisos para a média.

Um modo de contornar o problema citado acima é a estimação do tipo “*bootstrapping*”. Esse método envolve a seleção **sem reposição** de uma amostra inicial de tamanho n , a qual é recombinaada **com reposição** de 100 a 1.000 vezes, dependendo da capacidade de processamento do computador empregado, para que seja obtida uma distribuição das médias amostrais sem que se recorra a hipóteses acerca da distribuição populacional.

Primeiro, forma-se com as médias calculadas uma seqüência ordenada de observações. Em seguida, a seqüência é segmentada conforme o nível de confiança desejado. Caso seja de 95%, p. ex., deve-se identificar as observações que separam os primeiros e os últimos 2,5% das observações dos 95% restantes, concentrados em torno da mediana das médias amostrais. Por fim, os valores associados às observações identificadas definem a estimativa do intervalo de confiança.

UTILIZANDO O *MICROSOFT EXCEL*:

- seleção de uma amostra de tamanho n **sem reposição** \Rightarrow gerar uma amostra de tamanho $j > n$ **com reposição** (*vide* Figuras A.1, A.9 & A.10, Anexo I), aproveitando-se as n -primeiras observações distintas;
- seleção, a partir da amostra inicial, de 100 a 1.000 amostras de tamanho n **com reposição** (*vide* Figuras A.1, A.9 & A.10, Anexo I);
- calcular a média amostral de cada amostra \Rightarrow =MÉDIA(núm1;núm2;...);
- ordenar as médias amostrais \Rightarrow usar a opção “DADOS/CLASSIFICAR”;
- identificar os percentis desejados \Rightarrow ajustar as fórmulas das células “B4” e “B5” e os textos correspondentes da Tabela A.1, Anexo I.

5.4. Estimativa do Intervalo para Proporções

A exemplo da média, também pode-se estimar um intervalo de confiança para a proporção de um conjunto de observações que satisfaz uma dada condição. Como visto anteriormente, quando $(n.p)$ e $[n.(1 - p)]$ são superiores ou iguais a 5, a distribuição binomial assemelha-se à distribuição normal. Nesse caso, o intervalo de confiança é definido da seguinte maneira:

$$p_s - Z \cdot \sqrt{\frac{p_s \cdot (1 - p_s)}{n}} \leq p \leq p_s + Z \cdot \sqrt{\frac{p_s \cdot (1 - p_s)}{n}};$$

Onde: p_s = proporção amostral;

Z = valor crítico da distribuição normal padronizada associado à metade do complemento do nível de confiança;

n = tamanho da amostra;

p = proporção populacional.

UTILIZANDO O *MICROSOFT EXCEL*: VIDE TABELA A.2, ANEXO I, COLUNAS “E” E “F” ATÉ A LINHA 13.

O Exemplo 5.3 ilustra uma das possíveis aplicações da estimativa de intervalos de confiança para proporções.

Exemplo 5.3: Aplicação da Estimativa do Intervalo para Proporções

Suponha-se uma editora que deseje conhecer a proporção de revistas distribuídas com defeito (p. ex., páginas duplicadas ou suprimidas). Caso seja selecionada uma amostra de 200 revistas, 35 das quais defeituosas ($p_s = 17,5\%$), qual é a estimativa do intervalo da proporção para um nível de confiança de 90%?

Como $(n \cdot p_s) = 35$ e $[n \cdot (1 - p_s)] = 165$, tem-se que:

$$\frac{\alpha}{2} = \frac{(1-0,9)}{2} = 0,05 \Rightarrow Z = 1,645;$$

$$p_s \pm Z \cdot \sqrt{\frac{p_s \cdot (1 - p_s)}{n}} = 0,175 \pm 1,645 \cdot \sqrt{\frac{0,175 \cdot 0,825}{200}} = 0,175 \pm 0,0442 \Rightarrow 0,1308 \leq p \leq 0,2192.$$

Portanto, a amostra selecionada indica que de 13,08% a 21,92% das revistas são distribuídas com defeito, com 90% de nível de confiança de que o intervalo $[0,1308; 0,2192]$ seja um dos intervalos que efetivamente contêm o parâmetro desejado. Note-se que quanto menor o nível de confiança, menor o intervalo estimado. Dessa forma, a confiança e a precisão das estimativas variam inversamente.

Para uma dada combinação de n e α , **os intervalos de confiança estimados para proporções, necessariamente envolvendo variáveis categóricas, costumam ser maiores, ou menos precisos, do que os intervalos estimados para variáveis contínuas.** Enquanto que no primeiro tipo de estimativa as observações são agrupadas em uma quantidade finita de categorias, no segundo, as observações são mais variadas, agregando mais informação à estimativa.

5.5. Determinação do Tamanho da Amostra

Nos exemplos examinados anteriormente, o tamanho da amostra foi fixado arbitrariamente, independentemente da extensão do intervalo de confiança. Na prática, a determinação do tamanho apropriado da amostra é um procedimento complexo, envolvendo restrições quanto ao tempo, aos dados e aos recursos financeiros disponíveis e quanto à facilidade de seleção das observações. O **acesso a dados coletados por pesquisas anteriores ou por testes-piloto**, em especial, **permite uma melhor alocação dos recursos disponíveis**.

Independentemente das restrições existentes, todavia, as propriedades matemáticas dos estimadores continuam valendo. Dessa forma, o tamanho da amostra tido como exequível pode não permitir inferências tão confiáveis ou tão precisas quanto se deseja. Isso, porém, não necessariamente significa que a amostragem pretendida é irrelevante, pois os dados coletados podem fornecer tanto conclusões preliminares, como indicações valiosas para futuras pesquisas.

5.5.1. Médias

Do ponto de vista estritamente matemático, o tamanho da amostra para estimar a média é obtido por meio da seguinte transformação:

$$Z = \frac{\bar{X} - \mu}{\frac{S_o}{\sqrt{n}}} \Rightarrow n = \frac{Z^2 \cdot S_o^2}{(\bar{X} - \mu)^2}, \text{ com } S_o = \text{estimativa inicial do desvio-padrão.}$$

Definindo-se o **erro amostral** (e) como a diferença entre as médias amostral e populacional ($\bar{X} - \mu$), tem-se, ainda:

$$n = \frac{Z^2 \cdot S_o^2}{e^2}, \text{ a ser aproximado para o valor inteiro imediatamente superior quando contiver casas decimais.}$$

UTILIZANDO O *MICROSOFT EXCEL*: VIDE TABELA A.3, ANEXO I, COLUNAS “A” E “B” ATÉ A LINHA 08.

Portanto, dados o **nível desejado de confiança**, o **erro amostral admitido**¹⁴ e uma **estimativa inicial do desvio-padrão**, fixada a partir de pesquisas anteriores, de pré-testes ou de consultas a especialistas¹⁵, obtém-se o tamanho apropriado da amostra. Destaque-se que esses valores preliminares não devem ser confundidos com os resultados da pesquisa amostral, pois as estimativas obtidas podem não ser referendadas pelos dados coletados. Assim, as expectativas quanto ao nível de confiança e ao erro amostral podem ser

¹⁴ O erro amostral admitido corresponde a maior diferença entre as médias amostral e populacional tida como insignificante em face do propósito da pesquisa que estiver sendo conduzida.

¹⁵ Alternativamente, caso haja razões para crer que a população examinada é normalmente distribuída, a estimativa inicial do desvio-padrão pode ser dada por $\left(\frac{1}{6}\right)$ da amplitude ($A_r \cong 6 \cdot \sigma$ nas populações normalmente distribuídas).

reformuladas, para melhor ou para pior, à luz do desvio-padrão amostral efetivamente calculado.

5.5.2. Proporções

A exemplo do que foi feito acima, na estimação de proporções o tamanho da amostra é dado pela seguinte transformação:

$$Z = \frac{p_s - p}{\sqrt{\frac{p_o \cdot (1 - p_o)}{n}}} \Rightarrow n = \frac{Z^2 \cdot p_o \cdot (1 - p_o)}{(p_s - p)^2}, \text{ com } p_o = \text{estimativa inicial da proporção de "sucessos"}.$$

Definindo-se o **erro amostral** (e) como a diferença entre as proporções amostral e populacional ($p_s - p$), tem-se, ainda:

$$n = \frac{Z^2 \cdot p_o \cdot (1 - p_o)}{e^2}, \text{ a ser aproximado para o valor inteiro imediatamente superior quando contiver casas decimais.}$$

UTILIZANDO O *MICROSOFT EXCEL*: VIDE TABELA A.3, ANEXO I, COLUNAS “C” E “D” ATÉ A LINHA 08.

Geralmente, a estimativa inicial da proporção de “sucessos” é fixada a partir de dados coletados previamente. Caso, porém, não haja dados disponíveis, pode-se recorrer ao valor de p no qual ocorre o máximo da função $f(p) = p \cdot (1 - p)$.¹⁶ Consequentemente, a estimativa mais conservadora para a proporção de “sucessos” é 0,5. Esse valor define, para cada nível de confiança, a maior amostra exigida por pesquisas amostrais com apenas duas categorias, como pode ser observado no Exemplo 5.5.

Exemplo 5.5: Cálculo do Tamanho da Amostra para Estimar Duas Proporções

Suponha-se uma campanha eleitoral com vários candidatos. Qual deve ser o tamanho da amostra para que se possa obter, com 95% de nível de confiança e 3% de margem de erro, as intenções de votos em um dos candidatos?

Com $(1 - \alpha) = 0,95$ e $e = 0,03$, fazendo-se $p_o = 0,5$ e agrupando-se os votos nulos e em branco e as intenções de voto nos demais candidatos em uma única categoria, tem-se:

$$\left(\frac{\alpha}{2}\right) = \frac{(1 - 0,95)}{2} = 0,025 \Rightarrow Z = 1,96 \therefore n = \frac{Z^2 \cdot p_o \cdot (1 - p_o)}{e^2} = \frac{1,96^2 \cdot 0,5 \cdot (1 - 0,5)}{0,03^2} = 1.067,11 \cong 1.068.$$

Portanto, na ausência de mais informações, são necessárias 1.068 entrevistas para que seja obtida a proporção desejada.

¹⁶ $f(p) = p \cdot (1 - p) = p - p^2 \Rightarrow \begin{cases} \frac{df(p)}{dp} = 1 - 2 \cdot p^* = 0 \Rightarrow p^* = \frac{1}{2}; \\ \frac{d^2 f(p)}{dp^2} = -2 \Rightarrow p^* \text{ é um ponto máximo.} \end{cases}$

A Tabela 5.1 sintetiza vários tamanhos de amostra para estimar proporções, dado um nível de confiança de 95% ($Z = 1,96$):

Tabela 5.1: Tamanho da Amostra para Vários Erros Amostrais e Proporções de “Sucesso” ($Z = 1,96$)

(em %)

Estimativa Inicial de p (p_0)	Erro Amostral Admitido (e)			
	0,15	0,10	0,05	0,03
5	n.d.	n.d.	n.d.	203
10	n.d.	n.d.	139	385
15	n.d.	49	196	545
20	28	62	246	683
30	36	81	323	897
40	41	93	369	1.025
50	43	97	385	1.068

Nota: “n.d.” corresponde a “não definido”, pois (n,p) é inferior a 5.

Nas amostragens com mais de duas categorias, os pesquisadores terão, necessariamente, de definir estimativas iniciais para as k -proporções desejadas, não havendo uma hipótese mais conservadora para cada tipo de “sucesso”. Com base nas estimativas iniciais, os pesquisadores deverão calcular $(k-1)$ tamanhos de amostra conforme mostrado anteriormente, somando-os em seguida para obter o tamanho total. Esse procedimento achase ilustrado no Exemplo 5.6.

Exemplo 5.6: Cálculo do Tamanho da Amostra para Estimar n -Proporções

Suponha-se uma campanha eleitoral com três candidatos. Qual deve ser o tamanho da amostra para que se possa obter, com 95% de nível de confiança e 3% de margem de erro, as intenções de votos em cada candidato, bem como as proporções de votos nulos e em branco? Com base nas pesquisas realizadas anteriormente, as estimativas iniciais para as intenções de voto nos três candidatos são $p_1 = 0,40$, $p_2 = 0,30$ e $p_3 = 0,15$, e que os votos nulos e em branco são, respectivamente, $p_4 = 0,10$ e $p_5 = 0,05$.

Com $(1 - \alpha) = 0,95$, $e = 0,03$ e tratando-se os votos nulos como resíduo, tem-se:

- $p_1 = 0,40 \Rightarrow q_1 = (p_2 + p_3 + p_4 + p_5) = 0,60 \Rightarrow n_1 = 1.025$;
- $p_2 = 0,30 \Rightarrow q_2 = (p_1 + p_3 + p_4 + p_5) = 0,70 \Rightarrow n_2 = 897$;
- $p_3 = 0,15 \Rightarrow q_3 = (p_1 + p_2 + p_4 + p_5) = 0,85 \Rightarrow n_3 = 545$;
- $p_4 = 0,10 \Rightarrow q_4 = (p_1 + p_2 + p_3 + p_5) = 0,95 \Rightarrow n_4 = 203$;
- $n = n_1 + n_2 + n_3 + n_4 = 2.670$.

Portanto, em princípio, são necessárias 2.670 entrevistas para que sejam obtidas as cinco proporções desejadas. **Convém frisar que as proporções obtidas podem ou não coincidir com as estimativas definidas inicialmente. Dessa forma, é fundamental que o nível de confiança e os amostrais sejam recalculados à luz dos novos resultados. Isso poderá ou não gerar estimativas mais precisas e/ou confiáveis do que aquelas pretendidas.**

O método ilustrado acima mais do que duplicou o tamanho original da amostra. Um modo menos oneroso de obter resultados igualmente válidos é o uso da seguinte simplificação (SOUZA, 1990, p. 43):

$n = \frac{Z'^2 \cdot 0,25}{e^2}$, com Z' correspondendo ao valor crítico da distribuição normal padronizada associado à metade do complemento do nível de confiança desejado, ajustado conforme a quantidade de categorias examinadas, ou seja:

$$\left(\frac{\mathbf{y}}{2}\right) = \frac{\left[1 - (1 - \mathbf{a})\left(\frac{1}{k-1}\right)\right]}{2} \Rightarrow Z', \text{ com } k \text{ correspondendo à quantidade de categorias.}^{17}$$

O Exemplo 5.7 ilustra a simplificação descrita acima.

Exemplo 5.7: Cálculo Simplificado do Tamanho da Amostra para Estimar n -Proporções

No exemplo anterior, utilizando-se a simplificação sugerida, qual deve ser o tamanho da amostra para que se possa obter, com 95% de nível de confiança e 3% de margem de erro, as intenções de votos em cada candidato, bem como as proporções de votos nulos e em branco?

Com $(1 - \alpha) = 0,95$, $e = 0,03$ e $k = 5$, tem-se:

$$\left(\frac{\mathbf{y}}{2}\right) = \frac{\left[1 - (1 - \mathbf{a})\left(\frac{1}{k-1}\right)\right]}{2} = \frac{\left[1 - (0,95)\left(\frac{1}{5-1}\right)\right]}{2} = 0,0065 \Rightarrow z' = 2,49 ;$$

$$n = \frac{Z'^2 \cdot 0,25}{e^2} = \frac{2,49^2 \cdot 0,25}{0,03^2} = 1.722,25 \cong 1.723.$$

Portanto, são necessárias 1.723 entrevistas para que sejam obtidas as cinco proporções desejadas. Pode-se, contudo, diminuir o tamanho da amostra por meio de técnicas mais complexas (p. ex., amostragens estratificadas, por conglomerados ou com estimadores do tipo “razão” ou do tipo “regressão”), as quais exigem uma maior quantidade de informações acerca da população.

5.6. Amostragem em Populações Finitas

Nas amostragens sem reposição e com N inferior ou igual a $(20.n)$, os intervalos de confiança e os tamanhos apropriados das amostras devem ser ajustados pelo CPF (vide Item 4.3).

Os intervalos de confiança ajustados são:

¹⁷ $\alpha = \psi$ para $k = 2$.

a) estimação da média:

- com desvio-padrão conhecido:

$$\left[\bar{X} - Z \cdot \left(\frac{\sigma}{\sqrt{n}} \right) \left(\sqrt{\frac{N-n}{N-1}} \right); \bar{X} + Z \cdot \left(\frac{\sigma}{\sqrt{n}} \right) \left(\sqrt{\frac{N-n}{N-1}} \right) \right];$$

UTILIZANDO O *MICROSOFT EXCEL*: VIDE TABELA A.2, ANEXO I, COLUNAS “A” E “B” A PARTIR DA LINHA 14.

- com desvio-padrão desconhecido:

$$\left[\bar{X} - t_{n-1} \cdot \left(\frac{S}{\sqrt{n}} \right) \left(\sqrt{\frac{N-n}{N-1}} \right); \bar{X} + t_{n-1} \cdot \left(\frac{S}{\sqrt{n}} \right) \left(\sqrt{\frac{N-n}{N-1}} \right) \right];$$

UTILIZANDO O *MICROSOFT EXCEL*: VIDE TABELA A.2, ANEXO I, COLUNAS “C” E “D” A PARTIR DA LINHA 14.

b) estimação da proporção:

$$\left[p_s - Z \cdot \sqrt{\frac{p_s \cdot (1 - p_s)}{n}} \left(\sqrt{\frac{N-n}{N-1}} \right); p_s + Z \cdot \sqrt{\frac{p_s \cdot (1 - p_s)}{n}} \left(\sqrt{\frac{N-n}{N-1}} \right) \right].$$

UTILIZANDO O *MICROSOFT EXCEL*: VIDE TABELA A.2, ANEXO I, COLUNAS “E” E “F” A PARTIR DA LINHA 14.

Já os tamanhos ajustados das amostras (n_{CPF}), na estimação tanto de médias, como de proporções, são:

$$n_{CPF} = \frac{n \cdot N}{n + (N - 1)}, \text{ com } n \text{ correspondendo ao tamanho não-ajustado.}$$

UTILIZANDO O *MICROSOFT EXCEL*: VIDE TABELA A.3, ANEXO I, A PARTIR DA LINHA 09.

5.7. Estimativa do Intervalo para Totais

O total (T) é obtido multiplicando-se a média amostral pelo tamanho da população, ou seja:

$$T = N \cdot \bar{X}.$$

O respectivo intervalo de confiança, por sua vez, é:

$$\left[N \cdot \bar{X} - N \cdot t_{n-1} \cdot \left(\frac{S}{\sqrt{n}} \right) \left(\sqrt{\frac{N-n}{N-1}} \right); N \cdot \bar{X} + N \cdot t_{n-1} \cdot \left(\frac{S}{\sqrt{n}} \right) \left(\sqrt{\frac{N-n}{N-1}} \right) \right].$$

UTILIZANDO O *MICROSOFT EXCEL*: VIDE TABELA A.4, ANEXO I.

Uma das possíveis aplicações da presente modalidade de estimativa encontra-se ilustrada no Exemplo 5.8.

Exemplo 5.8: Aplicação da Estimativa do Intervalo para Totais

Suponha-se um arquivo contendo 5.000 ordens bancárias, normalmente distribuídas, mas cujos montante e desvio-padrão não são conhecidos. Uma amostra de 100 ordens gerou os seguintes resultados: $\bar{X} = \text{R\$ } 1.076,39$ e $S = \text{R\$ } 273,62$. Qual é o montante estimado das ordens e respectivo intervalo, para um nível de confiança de 95%?

Dado que $N > (20.n)$ e $n < 120$, tem-se que:

$$T = N \cdot \bar{X} = (5.000) \cdot (1.076,39) = 5.381.950.$$

$$\left. \begin{array}{l} (n-1) = 100 - 1 = 99 \\ \mathbf{a} = 1 - 0,95 = 0,05 \end{array} \right\} \Rightarrow t_{99,0,05} = 1,984.$$

$$N \cdot \bar{X} \pm N \cdot t_{n-1} \cdot \left(\frac{S}{\sqrt{n}} \right) = (5.000) \cdot (1.076,39) \pm (5.000) \cdot (1,984) \cdot \frac{273,62}{\sqrt{100}} = (5.381.950) \pm (271.431,04) \Rightarrow$$

$$\Rightarrow 5.110.518,96 \leq T \leq 5.653.381,04.$$

Portanto, a estimativa pontual para o montante é de R\$ 5.381.950,00, enquanto que o intervalo estimado para o nível de confiança desejado, é [R\$ 5.110.518,96; R\$ 5.653.381,04].

5.8. Estimativa do Intervalo para Diferenças

A estimação da diferença é empregada quando se crê que os registros referentes a um conjunto de itens contêm erros, cujo montante deseja-se conhecer. Essa técnica, comum em trabalhos de auditoria, envolve os seguintes procedimentos:

- seleção de uma amostra de tamanho apropriado;
- cálculo da diferença amostral média (\bar{D}) entre o valor observado do item e o valor constante do registro, ou seja:

$$\bar{D} = \frac{\sum_{i=1}^n D_i}{n}, \text{ com } D_i = (X_{i, \text{observado}} - X_{i, \text{registrado}});$$

- cálculo do desvio-padrão (S_D) amostral das diferenças, ou seja:

$$S_D = \sqrt{\frac{\sum_{i=1}^n D_i^2 - n \cdot \bar{D}^2}{n-1}};$$

d) estimativa do intervalo de confiança da diferença total, ou seja:

$$\left[N \cdot \bar{D} - N \cdot t_{n-1} \cdot \left(\frac{S_D}{\sqrt{n}} \right) \left(\sqrt{\frac{N-n}{N-1}} \right); N \cdot \bar{D} + N \cdot t_{n-1} \cdot \left(\frac{S_D}{\sqrt{n}} \right) \left(\sqrt{\frac{N-n}{N-1}} \right) \right].$$

UTILIZANDO O *MICROSOFT EXCEL*: VIDE TABELA A.5, ANEXO I.

Uma aplicação possível da estimativa do intervalo para diferenças encontra-se ilustrada no Exemplo 5.9.

Exemplo 5.9: Aplicação da Estimativa do Intervalo para Diferenças

Suponha-se um cadastro contendo 5.000 ordens bancárias, normalmente distribuídas, mas cujo desvio-padrão não é conhecido. Uma amostra de 100 ordens identificou 14 ordens cujos valores diferiam dos valores constantes do cadastro. Qual é o montante estimado das diferenças e o respectivo intervalo para um nível de confiança de 95%?

A amostragem revelou as seguintes diferenças (D_i): R\$ 75,41, R\$ 38,97, R\$108,54, – R\$ 37,18, R\$ 62,75, R\$ 118,32, – R\$ 88,84, R\$ 127,74, R\$ 55,42, R\$ 39,03, R\$ 29,41, R\$ 47,99, R\$ 28,73 e R\$ 84,05.

$$\bar{D} = \frac{\sum_{i=1}^{100} D_i}{100} = 6,9034 \text{ e } S_D = \sqrt{\frac{\sum_{i=1}^{100} D_i^2 - 100 \cdot 6,9034^2}{100-1}} = 27,23.$$

Dado que $N > (20 \cdot n)$ e $n < 120$, tem-se que:

$$N \cdot \bar{D} = 34.517.$$

$$\left. \begin{array}{l} (n-1) = 100 - 1 = 99 \\ \mathbf{a} = 1 - 0,95 = 0,05 \end{array} \right\} \Rightarrow t_{99;0,05} = 1,984.$$

$$N \cdot \bar{D} \pm N \cdot t_{n-1} \cdot \left(\frac{S_D}{\sqrt{n}} \right) = (5.000) \cdot (6,9034) \pm (5.000) \cdot (1,984) \cdot \frac{27,23}{\sqrt{100}} = (34.517) \pm (27.012,16) \Rightarrow$$

$$\Rightarrow 7.504,84 \leq N \cdot \bar{D} \leq 61.529,16.$$

Portanto, a estimativa pontual para o montante da diferença é de R\$ 34.517,00, enquanto que o intervalo estimado, para o nível de confiança desejado, é [R\$ 7.504,84; R\$ 61.529,16].

6. VARIANTES DA AMOSTRAGEM ALEATÓRIA SIMPLES

Os conceitos enunciados no capítulo anterior referem-se às **amostras aleatórias simples** (*vide* item 1.2.1.). Por conseguinte, essa modalidade de amostragem confunde-se com os próprios fundamentos da inferência estatística. As demais modalidades representam desenvolvimentos da amostragem aleatória simples destinados a capturar, com maior precisão e menor custo, as características relevantes da população. Os próximos itens abordarão dois desses desenvolvimentos: as **amostragens aleatórias estratificada** (AAE) e **por conglomerados** (AAC).¹⁸

6.1. Amostragem Aleatória Estratificada

Caso os membros da equipe de auditoria acreditem que haja uma estreita relação entre o objeto do trabalho e uma dada característica da população ou julguem que um subgrupo específico da população é particularmente interessante para os propósitos da investigação, convém recorrer à amostragem estratificada, na qual a população é dividida em k -estratos mutuamente excludentes e exaustivos, selecionando-se uma amostra aleatória simples de cada estrato (*vide* item 1.2.2).

Note-se que se os k -estratos da população contêm N_1, N_2, \dots, N_k elementos, então:

$$N_1 + N_2 + \dots + N_k = N$$

Da mesma forma, se as k -amostras têm tamanhos n_1, n_2, \dots, n_k , o tamanho total da amostra (n) é:

$$n_1 + n_2 + \dots + n_k = n$$

6.1.1. Estimação de Médias, Totais e Proporções na AAE

As médias populacionais dos k -estratos são denotadas por meio dos símbolos $\mu_1, \mu_2, \dots, \mu_k$, enquanto que os símbolos $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$ correspondem às k -médias amostrais. Considerando-se o j -ésimo estrato, tem-se que \bar{X}_j , obtido a partir de uma amostra aleatória simples, é um estimador não-viesado de μ_j

A variância do estimador da média amostral do estrato em questão ($S_{\bar{x}_j}^2$), por sua vez, é definida pela seguinte equação:

$$S_{\bar{x}_j}^2 = \frac{S_j^2}{n_j} \cdot \frac{N_j - n_j}{N_j}, \text{ com } S_j^2 = \text{estimativa da variância populacional do } j\text{-ésimo estrato.}$$

¹⁸ As equações expostas neste capítulo e no capítulo anterior estão sintetizadas no Anexo III.

Uma vez apuradas as k -médias amostrais, a estimação da média populacional geral (\bar{X}_{AE}) é imediata:

$$\bar{X}_{AE} = \frac{1}{N} \cdot \sum_{j=1}^k N_j \cdot \bar{X}_j, \text{ com } AE = \text{“amostragem estratificada”}.$$

Considerando-se que as amostras selecionadas em cada estrato são mutuamente independentes, tem-se que a estimativa não-viesada da variância do estimador da média populacional geral ($S_{\bar{X}_{AE}}^2$) é:

$$S_{\bar{X}_{AE}}^2 = \frac{1}{N^2} \cdot \sum_{j=1}^k N_j^2 \cdot S_{\bar{X}_j}^2.$$

Para n superior ou igual a 120, o intervalo de confiança da média populacional geral verdadeira (μ), para Z dado pelo α desejado, é: $[\bar{X}_{AE} \pm Z \cdot S_{\bar{X}_{AE}}]$.

O Exemplo 6.1 ilustra algumas das principais propriedades da amostragem estratificada.

Exemplo 6.1: Aplicação da Estimativa de Médias na Amostragem Estratificada

O Ministério do Trabalho e Emprego oferece cursos de requalificação profissional em 155 locais na região sul: 60 no Rio Grande do Sul, 50 no Paraná e 45 em Santa Catarina. O ministério cogita acrescentar mais um curso ao rol já disponível. No intuito de estimar quanto seria a demanda mensal média por local pelo novo curso, procedeu-se ao acréscimo cogitado em 12 locais no Rio Grande do Sul, 10 no Paraná e 9 em Santa Catarina, selecionados aleatoriamente.

Ao longo de um mês, apurou-se, com base nas quantidades de pessoas que se inscreveram no novo curso, os seguintes resultados amostrais (i.e., média e desvio-padrão):

- a) $\bar{X}_{RS} = 21,2$ e $S_{RS} = 12,8$;
- b) $\bar{X}_{PR} = 13,3$ e $S_{PR} = 11,4$;
- c) $\bar{X}_{SC} = 26,1$ e $S_{SC} = 9,2$.

Para que a demanda mensal média por local seja obtida, note-se, primeiramente, que:

- a) $N_{RS} = 60$ e $n_{RS} = 12$;
- b) $N_{PR} = 50$ e $n_{PR} = 10$;
- c) $N_{SC} = 45$ e $n_{SC} = 9$;
- d) $N = 155$ e $n = 31$.

A estimativa da média populacional é:

$$\bar{X}_{AE} = \frac{1}{155} \cdot \sum_{j=1}^3 N_j \cdot \bar{X}_j = \frac{60 \cdot 21,2 + 50 \cdot 13,3 + 45 \cdot 26,1}{155} = 20,1; \text{ com } j = RS, PR \text{ e } SC.$$

Continua

As variâncias dos estimadores da média de cada estrato, por sua vez, são:

$$\begin{aligned} \text{a)} \quad S_{\bar{x}_{RS}}^2 &= \frac{S_{RS}^2}{n_{RS}} \cdot \frac{N_{RS} - n_{RS}}{N_{RS}} = \frac{12,8^2}{12} \cdot \frac{(60-12)}{60} = 10,923; \\ \text{b)} \quad S_{\bar{x}_{PR}}^2 &= \frac{S_{PR}^2}{n_{PR}} \cdot \frac{N_{PR} - n_{PR}}{N_{PR}} = \frac{11,4^2}{10} \cdot \frac{(50-10)}{50} = 10,397; \\ \text{c)} \quad S_{\bar{x}_{SC}}^2 &= \frac{S_{SC}^2}{n_{SC}} \cdot \frac{N_{SC} - n_{SC}}{N_{SC}} = \frac{9,2^2}{9} \cdot \frac{(45-9)}{45} = 7,524. \end{aligned}$$

Com essas variâncias, já se poderia definir os intervalos de confiança das médias dos estratos. No entanto, o intervalo desejado refere-se à média geral. Para isso, é preciso calcular $S_{\bar{X}_{AE}}^2$, ou seja:

$$S_{\bar{X}_{AE}}^2 = \sqrt{S_{\bar{X}_{AE}}^2} = \sqrt{\frac{1}{155^2} \cdot \sum_{j=1}^3 N_j^2 \cdot S_{X_j}^2} = \sqrt{\frac{60^2 \cdot 10,923 + 50^2 \cdot 10,397 + 45^2 \cdot 7,524}{155^2}} = 1,83.$$

Ademais, para $\alpha = 95\%$ e $n = 31$ ($n < 120$), tem-se $t_{30; 0,05} = 2,04$. Assim, o intervalo de confiança da média geral é:

$$[20,10 - 2,04 \cdot 1,83 < \mu < 20,10 + 2,04 \cdot 1,83] = [16,37 < \mu < 23,83].$$

Dessa forma, pode-se afirmar, com um nível de confiança de 95%, que a demanda mensal média por local pelo novo curso está compreendida no intervalo [16,37; 23,83].

Como o valor total de uma população é dado pela multiplicação da quantidade de elementos que a compõem pelo seu valor médio, bastam algumas transformações para que se obtenha o intervalo de confiança correspondente. A estimativa pontual, a variância do estimador e o intervalo desejado, para n superior ou igual a 120, são definidos pelas seguintes equações:

$$\begin{aligned} \text{a)} \quad N \cdot \bar{X}_{AE} &= \sum_{j=1}^k N_j \cdot \bar{X}_j; \\ \text{b)} \quad N^2 \cdot S_{\bar{X}_{AE}}^2 &= \sum_{j=1}^k N_j^2 \cdot S_{X_j}^2; \\ \text{c)} \quad &[N \cdot \bar{X}_{AE} \pm Z \cdot N \cdot S_{\bar{X}_{AE}}]. \end{aligned}$$

No Exemplo 6.2, as novas equações são aplicadas ao exemplo anterior.

Exemplo 6.2: Aplicação da Estimativa de Totais na Amostragem Estratificada

No exemplo anterior, a demanda total pelo novo curso geraria, para $\alpha = 95\%$ e $n = 31$ ($t_{30; 0,05} = 2,04$), o seguinte intervalo de confiança:

$$[155 \cdot 20,10 - 2,04 \cdot 155 \cdot 1,83 < N \cdot \mu < 155 \cdot 20,10 + 2,04 \cdot 155 \cdot 1,83] = [2536,85 < N \cdot \mu < 3694,15].$$

No que se refere à estimação de intervalos para proporções populacionais, a estimativa pontual, a variância do estimador e o intervalo de confiança correspondente, para n superior ou igual a 120, são assim definidos:

- a) $\hat{p}_{AE} = \frac{1}{N} \cdot \sum_{j=1}^k N_j \cdot \hat{p}_j$, com $\hat{p}_j =$ proporção amostral de “sucessos” do j -ésimo estrato;
- b) $S_{\hat{p}_{AE}}^2 = \frac{1}{N^2} \cdot \sum_{j=1}^k N_j^2 \cdot S_{\hat{p}_j}^2$, com $S_{\hat{p}_j}^2 = \frac{\hat{p}_j \cdot (1 - \hat{p}_j)}{n_j - 1} \cdot \frac{(N_j - n_j)}{N_j}$;
- c) $\left[\hat{p}_{AE} \pm Z \cdot S_{\hat{p}_{AE}} \right]$.

6.1.2. Determinação do Tamanho da Amostra na AAE

Uma vez que a amostragem estratificada pode ser entendida como uma combinação de amostras aleatórias simples, **o cálculo do tamanho da amostra de cada estrato deve observar as equações definidas no capítulo anterior**. Convém frisar, porém, que as equipes de auditoria não precisam utilizar o mesmo **nível de confiança** (α) e o mesmo **erro amostral** (e) em todos os estratos. Esses parâmetros **poderão variar conforme a maior ou menor importância de cada estrato para o trabalho proposto**. Assim, uma pesquisa que deseje apurar a opinião de toda população sobre certo tema, mas que esteja particularmente interessada na opinião de um grupo étnico ou profissional específico, poderá atribuir a esse grupo um maior nível de confiança e/ou um menor erro amostral no momento do cálculo do tamanho da amostra.

Importa notar, contudo, que a variância do estimador da média amostral de cada estrato é sensível ao tamanho da amostra. Dessa forma, supondo-se tudo mais constante, quanto menor o tamanho da amostra, maior a variância em questão, o que pode gerar problemas na fase de agregação. Portanto, as equipes de auditoria devem **tomar cuidado para que os estratos menos importantes não fiquem demasiadamente subrepresentados** no estudo amostral, pois isso comprometeria a generalização dos resultados obtidos para toda a população.

Além dessas considerações, o tamanho da amostra de cada estrato também pode ser definido privilegiando-se os estratos com maior variância. Esse procedimento, denominado **“alocação ótima”**, é usado para estimar as médias, totais e proporções gerais, sendo menos útil quando se deseja obter estimativas para cada estrato. Ademais, pressupõe que a variância dos estratos (σ_j) seja conhecida.

Supondo-se uma amostra total com tamanho n , a alocação ótima estipula que as amostras de cada estrato deverão ter os seguintes tamanhos:

- a) estimação de médias: $n_j = \frac{N_j \cdot \sigma_j}{\sum_{j=1}^k N_j \cdot \sigma_j} \cdot n$;

b) estimação de proporções:
$$n_j = \frac{N_j \cdot \sqrt{p_j \cdot (1 - p_j)}}{\sum_{j=1}^k N_j \cdot \sqrt{p_j \cdot (1 - p_j)}} \cdot n.$$

6.2. Amostragem Aleatória por Conglomerados

Confrontadas, p. ex., com a ausência de um cadastro completo ou confiável, ou com os altos custos associados à realização de entrevistas de contato direto, as equipes de auditoria podem optar pela amostragem por conglomerados. Trata-se de um procedimento recomendável quando a população pode ser subdividida em conglomerados menores, que sejam geograficamente compactos (p. ex., bairros ou quarteirões de uma cidade).

Nessa modalidade de amostragem, seleciona-se, primeiramente, uma amostra aleatória simples de conglomerados. Em seguida, todos os seus membros são entrevistados. Em outras palavras, a amostragem por conglomerados equivale a uma amostra aleatória simples seguida de um censo (*vide* item 1.2.3).

6.2.1. Estimação de Médias, Totais e Proporções na AAC

Suponha-se que a população se divida em M conglomerados, que uma amostra aleatória simples de m conglomerados tenha sido selecionada e que seus integrantes (N_1, N_2, \dots, N_m) tenham sido entrevistados. Com base nas médias e proporções amostrais dos conglomerados escolhidos ($\bar{X}_1, \bar{X}_2, \dots, \bar{X}_m$ e $\hat{p}_1, \hat{p}_2, \dots, \hat{p}_m$), tem-se que as estimativas pontuais da média e da proporção gerais (\bar{X}_{AC} e \hat{p}_{AC}), as variâncias dos estimadores e os intervalos de confiança correspondentes, para n superior ou igual a 120, são assim definidos:

a)
$$\bar{X}_{AC} = \frac{\sum_{j=1}^m N_j \cdot \bar{X}_j}{\sum_{j=1}^m N_j} \text{ e } \hat{p}_{AC} = \frac{\sum_{j=1}^m N_j \cdot \hat{p}_j}{\sum_{j=1}^m N_j}, \text{ com } AC = \text{“amostragem por conglomerados”};$$

b)
$$S_{\bar{X}_{AC}}^2 \text{ ou } S_{\hat{p}_{AC}}^2 = \frac{M - m}{M \cdot \left(\sum_{j=1}^m N_j \right)^2} \cdot S_{AC}^2,$$

Onde:
$$S_{AC}^2 = \frac{\sum_{j=1}^m N_j^2 \cdot (\bar{X}_j - \bar{X}_{AC})^2}{m - 1} \text{ ou } \frac{\sum_{j=1}^m N_j^2 \cdot (\hat{p}_j - \hat{p}_{AC})^2}{m - 1} \Rightarrow \text{estimativa da variância populacional};$$

c)
$$\left[\bar{X}_{AC} - Z \cdot S_{\bar{X}_{AC}} < \mu < \bar{X}_{AC} + Z \cdot S_{\bar{X}_{AC}} \right] \text{ e } \left[\hat{p}_{AC} - Z \cdot S_{\hat{p}_{AC}} < p < \hat{p}_{AC} + Z \cdot S_{\hat{p}_{AC}} \right].$$

Observe-se que, no presente contexto, as inferências podem ser feitas com uma quantidade relativamente pequena de informações previamente disponíveis acerca da

população. Basta que a população possa ser claramente dividida em conglomerados. Sequer o total de elementos da população precisa ser conhecido, devendo-se apenas determinar a quantidade de elementos de cada conglomerado selecionado, o que é uma consequência necessária do censo requerido pelas versões mais simples da amostragem por conglomerados.

O Exemplo 6.3 ilustra algumas propriedades da amostragem por conglomerados.

Exemplo 6.3: Aplicação da Estimativa de Médias e Proporções na Amostragem por Conglomerados

Uma amostra aleatória simples com 20 quarteirões é selecionada de uma área residencial com um total de 1.000 quarteirões. Cada uma das residências dos quarteirões selecionados foi visitada no intuito de se apurar a renda familiar. A Tabela 6.1 discrimina a renda anual média (\bar{X}_j) e a proporção de famílias com renda anual inferior a R\$ 15.000,00 (\hat{p}_j) nos quarteirões que compuseram a amostra.

Tabela 6.1: Resumo dos Dados Coletados na Aplicação da Amostragem por Conglomerados

Quarteirão Selecionado (j)	\bar{X}_j (em R\$)	\hat{p}_j	Quantidade de Domicílios (N_j)
1	26.283,00	0,1304	23
2	19.197,00	0,4516	31
3	37.911,00	0,1250	24
4	14.527,00	0,6585	41
5	16.753,00	0,5143	35
6	28.312,00	0,2692	26
7	21.646,00	0,3548	31
8	29.312,00	0,1563	32
9	31.829,00	0,1333	30
10	18.412,00	0,3846	39
11	33.893,00	0,0769	26
12	38.409,00	0,0476	21
13	43.911,00	0,0000	20
14	14.699,00	0,4375	32
15	24.921,00	0,1111	36
16	31.827,00	0,0909	33
17	34.436,00	0,0833	24
18	37.647,00	0,0400	25
19	30.026,00	0,1081	37
20	16.493,00	0,3659	41

Pede-se para que seja calculada a renda anual média e proporção de famílias com renda anual inferior a R\$ 15.000,00 de toda a área.

Primeiramente, tem-se que: $M = 1000$ e $m = 20$.

A quantidade total de domicílios é: $\sum_{j=1}^{20} N_j = (23 + 31 + \dots + 41) = 607$.

Continua

As estimativas pontuais da média e da proporção são:

$$a) \bar{X}_{AC} = \frac{\sum_{j=1}^{20} N_j \cdot \bar{X}_j}{\sum_{j=1}^{20} N_j} = \frac{(23)(26283) + (31)(19197) + \dots + (41)(16493)}{607} = 26109,00;$$

$$b) \hat{p}_{AC} = \frac{\sum_{j=1}^{20} N_j \cdot \hat{p}_j}{\sum_{j=1}^{20} N_j} = \frac{(23)(0,1304) + (31)(0,4516) + \dots + (41)(0,3659)}{607} = 0,2521.$$

Ademais, tem-se que:

$$a) \frac{\left(\sum_{j=1}^{20} N_j\right)^2}{20} = \frac{607^2}{20} = 18422,45;$$

$$b) \frac{\sum_{j=1}^{20} N_j^2 \cdot \left(\hat{p}_j - \hat{p}_{AC}\right)^2}{20 - 1} = \frac{(23)^2 \cdot (26283 - 26109)^2 + \dots + (41)^2 \cdot (16493 - 26109)^2}{19} = 69270551000;$$

$$c) \frac{\sum_{j=1}^{20} N_j^2 \cdot \left(\hat{p}_j - \hat{p}_{AC}\right)^2}{20 - 1} = \frac{(23)^2 \cdot (0,1304 - 0,2521)^2 + \dots + (41)^2 \cdot (0,3659 - 0,2521)^2}{19} = 38,1547.$$

Com isso, os estimadores da média e da proporção gerais apresentam os seguintes desvios-padrão:

$$a) S_{\bar{X}_{AC}} = \sqrt{S_{\bar{X}_{AC}}^2} = \sqrt{\frac{1000 - 20}{1000 \cdot \frac{\left(\sum_{j=1}^{20} N_j\right)^2}{20}} \cdot \frac{\sum_{j=1}^{20} N_j^2 \cdot \left(\bar{X}_j - \bar{X}_{AC}\right)^2}{20 - 1}} = \sqrt{\frac{(980)(69270551000)}{(1000)(18422,45)}} = 1920;$$

$$b) S_{\hat{p}_{AC}} = \sqrt{S_{\hat{p}_{AC}}^2} = \sqrt{\frac{1000 - 20}{1000 \cdot \frac{\left(\sum_{j=1}^{20} N_j\right)^2}{20}} \cdot \frac{\sum_{j=1}^{20} N_j^2 \cdot \left(\hat{p}_j - \hat{p}_{AC}\right)^2}{20 - 1}} = \sqrt{\frac{(980)(38,1547)}{(1000)(18422,45)}} = 0,0451.$$

Portanto, os intervalos de confiança da média e da proporção gerais, para $\alpha = 95\%$ e $\sum_{j=1}^{20} N_j = 607$ ($Z = 1,96$), são:

- a) $[26109 - 1,96 \cdot 1920 < \mu < 26109 + 1,96 \cdot 1920] = [22346 < \mu < 29872];$
 b) $[0,2521 - 1,96 \cdot 0,0451 < \mu < 0,2521 + 1,96 \cdot 0,0451] = [0,164 < \mu < 0,340].$

6.2.2. Determinação do Tamanho da Amostra na AAC

Parece haver alguma semelhança entre as amostragens estratificada e por conglomerado, pois ambas dividem a população em subgrupos. Trata-se, todavia, de uma semelhança superficial.

Na amostragem estratificada, seleciona-se uma amostra aleatória simples de cada estrato para que a pesquisa contenha representantes de todos os segmentos relevantes da população.

Já na amostragem por conglomerados, o objeto da amostra aleatória simples são os próprios subgrupos. Dessa forma, somente alguns conglomerados estarão representados no estudo amostral. **Dependendo de como forem definidos os subgrupos, os seus elementos poderão, no que se refere às características relevantes para o trabalho, ser muito heterogêneos externamente** (i.e., os conglomerados exibem perfis muito distintos) e **homogêneos internamente** (i.e., os elementos de cada conglomerado são muito parecidos). Nesse caso, corre-se o risco de que segmentos importantes da população sejam subrepresentados na pesquisa.

Efetivamente, quanto mais os conglomerados forem parecidos com a população e, por extensão, entre si, menor será a quantidade de subgrupos que precisará ser incluída no estudo amostral. Isso pode ser constatado no próprio **cálculo do tamanho da amostra** (m), o qual é **diretamente proporcional à magnitude da variância populacional dos conglomerados** (σ_{AC}^2) **ou da sua estimativa preliminar** ($S_{AC_0}^2$), como mostra a equação exibida a seguir:

$$m = \frac{M \cdot S_{AC_0}^2}{\frac{e^2}{Z^2} \cdot \frac{\left(\sum_{j=1}^M N_j\right)^2}{M} - S_{AC}^2}.$$

Convém ainda notar que **essa equação requer** não apenas uma estimativa de qual seja a variância populacional, mas **também uma estimativa de qual seja o tamanho da população** $\left(\sum_{j=1}^M N_j\right)$.

Conseqüentemente, enquanto que os trabalhos de estimação expostos no Exemplo 6.3 somente exigiram uma pequena quantidade de informações prévias acerca da população, o cálculo do tamanho da amostra requer muito mais dados ou pressuposições. Isso reforça **a necessidade de que os estudos amostrais sejam precedidos por criteriosos levantamentos de informações, que devem originar planos amostrais detalhados, cujas premissas e conclusões preliminares deverão ser conferidas à luz dos dados efetivamente coletados, devendo-se proceder, quando necessário, as modificações requeridas quanto aos níveis de confiança esperados e aos erros amostrais admitidos.**

7. AMOSTRAGEM POR UNIDADE MONETÁRIA

Os capítulos 5 e 6 discorreram sobre as modalidades clássicas de amostragem de variáveis (i.e., estimação de médias, totais e diferenças) e por atributo (i.e., estimação de proporções). Nas amostragens por atributo, porém, as técnicas examinadas não são aplicáveis aos casos nos quais as probabilidades de seleção variam, como ocorre quando as amostras são selecionadas sem reposição de populações finitas. Nessas situações, a distribuição de probabilidades mais apropriada é, em vez da binomial, a hipergeométrica. Essa distribuição é empregada, p. ex., em controles de qualidade ou de aderência a critérios preestabelecidos, nos quais os desvios em relação ao critério definido são tratados igualmente, ainda que os valores envolvidos difiram entre si.

No controle de qualidade ou de aderência, o tamanho da amostra, para populações com mais de quinhentos elementos, é definido pela equação:

$$n = \frac{C_{i,\alpha}}{p},$$

onde: $C_{i,\alpha}$ = índice de confiabilidade;
 i = quantidade prevista de erros na amostra;
 α = nível de confiança;
 p = proporção máxima de erros admitidos na população.

Geralmente, o índice de confiabilidade ($C_{i,\alpha}$) é obtido a partir de quadros como o da Tabela 7.1.

Tabela 7.1: Índices de Confiabilidade para Várias Quantidades Previstas de Erros e Níveis de Confiança

Nível de Confiança (α)	Quantidade Prevista de Erros na Amostra (i)		
	0 \Rightarrow ($C_{0,\alpha}$) ^a	1 \Rightarrow ($C_{1,\alpha}$) ^b	2 \Rightarrow ($C_{2,\alpha}$) ^c
0,99	4,60	6,61	–
0,95	3,00	4,75	6,30
0,90	2,30	3,89	5,33
0,85	1,90	3,38	4,73
0,80	1,61	3,00	4,28
0,70	1,20	2,44	3,62
0,50	0,69	1,68	2,64

Fonte: Adams & Regal, 1995, p. 19.

Notas: (a) $C_{0,\mathbf{a}} = -\ln(1-\mathbf{a}) \Leftarrow$ aproximação obtida a partir de uma distribuição hipergeométrica de probabilidades;

(b) e (c) para amostras numerosas e populações com pequenas taxas de erros, a distribuição hipergeométrica pode ser aproximada por meio de uma distribuição de Poisson, o que permite obter $C_{i,\alpha}$, mediante interpolações baseadas nas seguintes equações: $\mathbf{a} = e^{-C_{1,\mathbf{a}}} \cdot (C_{1,\mathbf{a}} + 1)$ e

$$\mathbf{a} = e^{-C_{2,\mathbf{a}}} \cdot \left[\frac{(C_{2,\mathbf{a}})^2}{2} + C_{2,\mathbf{a}} + 1 \right].$$

Como mostrado pelo Exemplo 7.1, é imediata a generalização do cálculo de $C_{0,\alpha}$ a partir de uma distribuição hipergeométrica de probabilidades (P_i^H).

Exemplo 7.1: Cálculo do Índice de Confiabilidade

Qual é o índice de confiabilidade ($C_{i,\mathbf{a}}$) associado a uma população (N) com 600 itens, 18 dos quais contendo erros ($p = 0,03$), e a uma amostra (n) com 100, todos perfeitos ($i = 0$)?

A probabilidade (P_i^H) de que semelhante amostra seja extraída de uma população com tais característica é dada pela seguinte equação:

$$P_0^H = \frac{\binom{18}{0} \binom{600-18}{100-0}}{\binom{600}{100}} = \frac{18!}{0!18!} \cdot \frac{582!}{100!482!} \cdot \frac{100!500!}{600!} = \frac{582}{600} \cdot \dots \cdot \frac{483}{501} \cong 0,97^{100} \cong 0,05.$$

Portanto, é de aproximadamente 5% o risco de que uma amostra com 100 elementos perfeitos seja proveniente de uma população com uma taxa de erro de 3%. Em outras palavras, pode-se afirmar, com um nível de confiança de 95% (complemento da taxa de risco), que a população examinada apresenta um erro máximo de 3%.

Generalizando-se:

$$\mathbf{a} = (1 - P_0^H) = [1 - (1 - p)^n] \Rightarrow n = \frac{\ln(1 - \mathbf{a})}{\ln(1 - p)}.$$

Como $n = \frac{C_{0,\mathbf{a}}}{p}$ e $p \cong -\ln(1 - p)$, então $C_{0,\alpha} = -\ln(1 - \alpha)$.

Utilizando-se a distribuição de Poisson (P_i^P), tem-se que o índice de confiabilidade ($C_{i,\alpha}$) corresponde ao produto do tamanho da amostra (n) pela taxa máxima de erros na população (p) que proporciona, para uma dada previsão sobre a quantidade de erros amostrais (i), o nível de confiança desejado (α) na seguinte equação:

$$\mathbf{a} = (1 - P_i^P) = \left\{ 1 - \sum_{j=0}^i \left[\frac{(n.p)^j \cdot e^{-(n.p)}}{j!} \right] \right\} = \left\{ 1 - \sum_{j=0}^i \left[\frac{(C_{i,\mathbf{a}})^j \cdot e^{-C_{i,\mathbf{a}}}}{j!} \right] \right\}.$$

Por interpolação, obtém-se $C_{i,\alpha}$, o qual, juntamente com p , define n .

Como **esses cálculos podem ser exaustivos**, deve-se usar a Tabela 7.1, como ilustrado pelo Exemplo 7.2.

Exemplo 7.2: Aplicação da Tabela de Índices de Confiabilidade

Em uma auditoria, pretende-se realizar um teste de conformidade. A proporção máxima de erros admitidos para que a população examinada seja aprovada pelo teste é 2%, com nível de confiança de 95%. Qual é o tamanho da amostra?

Com $p = 0,02$ e $\alpha = 0,95$, e prevendo-se que:

- a) a amostra não conterá erros ($C_{0;0,95} = 3,0$), tem-se: $n = \frac{C}{p} = \frac{3,0}{0,02} = 150$;
- b) a amostra conterá um único erro ($C_{1;0,95} = 4,75$), tem-se: $n = \frac{C}{p} = \frac{4,75}{0,02} = 235$.

Quando os resultados amostrais observados diferem dos resultados esperados, especialmente quando a quantidade de erros observados na amostra é superior à quantidade prevista, o índice de confiabilidade ($C_{i,\alpha}$) deve ser ajustado com base na Tabela 7.2.

Tabela 7.2: Índice de Confiabilidade para Vários Erros Observados na Amostra

n	Quantidade de Erros Observados na Amostra						
	0	1	2	3	4-5	6-7	> 8
15	Reduzido	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum
25	Reduzido	Reduzido	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum
30	Moderado	Reduzido	Nenhum	Nenhum	Nenhum	Nenhum	Nenhum
35	Moderado	Reduzido	Reduzido	Nenhum	Nenhum	Nenhum	Nenhum
45-50	Elevado	Moderado	Reduzido	Reduzido	Nenhum	Nenhum	Nenhum
70	Elevado	Elevado	Moderado	Reduzido	Reduzido	Nenhum	Nenhum
80	Elevado	Elevado	Moderado	Moderado	Reduzido	Reduzido	Nenhum
105	Elevado	Elevado	Elevado	Moderado	Moderado	Reduzido	Reduzido

Fonte: Adams & Regal, 1995, p. 21.

7.1. Metodologia da Amostragem por Unidade Monetária

A Amostragem por Unidade Monetária (AUM) também é uma amostragem por atributo baseada na distribuição hipergeométrica. Essa técnica permite estimar o montante monetário de uma dada população. Trata-se de uma técnica geralmente utilizada em auditorias financeiras, que têm por objetivo verificar se os registros contábeis são fidedignos.

Diferentemente da amostragem por variáveis, a AUM não exige o prévio conhecimento da quantidade de elementos da população e da variância dos valores que a compõem. Essa técnica baseia-se nas seguintes informações: a proporção máxima de erros admitidos na população (p), previamente definida pela equipe de auditoria, o índice

de confiabilidade ($C_{i,\alpha}$), obtido a partir de testes de conformidade, e **o montante constante dos registros contábeis** (M), o qual define a população da qual a amostra é extraída.

Como os elementos da população correspondem às unidades monetárias que integram o montante, convém notar que **a probabilidade de inclusão na amostra de um dado registro diminui quando o seu valor é subestimado**, sendo nula no caso de omissão. Dessa forma, há na *AUM* um viés em favor dos registros superestimados ou repetidos. Outra característica da *AUM* que merece menção é a **premissa de que não há diferenças superiores a 100% entre os valores registrados e observados**.

O tamanho da amostra na *AUM* é definido pela seguinte equação:

$$n = \frac{C_{0,\alpha}}{p}, \text{ com } p = \frac{M_R}{M} = \frac{M_A - \hat{E}_A - \hat{A}_P - O_A}{M};$$

Onde: $C_{0,\alpha}$ = **Fator de Precisão Básica** (índice de confiabilidade supondo-se que a amostra não contém erros);

M_R = **Materialidade Restrita** ou **Precisão Básica** (diferença máxima entre os montantes registrado e estimado supondo-se que a amostra não contém erros);

M = montante registrado;

M_A = **Materialidade Ampla** (diferença máxima entre os montantes registrado e estimado que não implica a impugnação do conjunto de lançamentos contábeis);

\hat{E}_A = **Erro Amostral Previsto** [estimativa da diferença que será observada na amostra entre os montantes registrado e estimado caso os membros da equipe de auditoria antecipem a ocorrência de erros ($i > 0$)];

\hat{A}_P = **Estimativa da Ampliação do Intervalo de Precisão** (margem de erro suplementar arbitrada quando se prevê a ocorrência de erros na amostra);

O_A = outros ajustes arbitrados pelos membros da equipe de auditoria.

Dessa forma, os membros da equipe de auditoria devem definir, inicialmente, a Materialidade Ampla (M_A). Em seguida, deve-se arbitrar os ajustes julgados necessários, os quais devem ser subtraídos da M_A , definindo a Materialidade Restrita (M_R). Convém notar que M_R e p são diretamente proporcionais. Dessa forma, quanto menor for a primeira, menor será a última, o que aumenta n . Consequentemente, **os ajustes feitos ainda na fase de planejamento devem permitir a obtenção de resultados amostrais robustos ainda que haja pequenos desvios em relação às premissas utilizadas originalmente**.

O Exemplo 7.3 ilustra como o tamanho da amostra deve ser calculado na *AUM*.

Exemplo 7.3: Cálculo do Tamanho da Amostra na AUM

Em uma auditoria, o conjunto de lançamentos contábeis examinados, totalizando R\$ 2.000.000,00, é considerado fidedigno se a diferença entre os montantes estimado e registrado não for superior a R\$ 100.000,00. Para um nível de confiança de 95%, qual é o tamanho da amostra?

Desconsiderando-se a necessidade de ajustes, tem-se:

$$p = \frac{100000}{2000000} = 0,05.$$

Com $i = 0$ e $\alpha = 0,95$ tem-se: $C_{0;0,95} = 3,0$ e $n = \frac{3,0}{0,05} = 60$.

Definido o tamanho da amostra (n), procede-se à seleção dos seus elementos. Primeiro, calcula-se o intervalo amostral (I_A):

$$I_A = \frac{M}{n}.$$

Em seguida, escolhe-se aleatoriamente uma unidade monetária compreendida em cada um dos sucessivos intervalos amostrais. Por fim, são selecionados para análise os registros contábeis correspondentes às unidades monetárias escolhidas,¹⁹ como mostrado no Exemplo 7.4.

Exemplo 7.4: Seleção dos Elementos Amostrais na AUM

No exemplo tratado anteriormente, qual é o intervalo amostral? Dado um conjunto hipotético de lançamentos contábeis, como devem ser selecionados os elementos amostrais?

Supondo-se que o sistema de controle interno é confiável e que, por conseguinte, pode-se prever que a amostra não conterà erros, tem-se $n = 60$. Dessa forma:

$$I_A = \frac{M}{n} = \frac{2000000}{60} = 33333,33.$$

Selecionando-se, aleatoriamente, o valor R\$ 16.000,00 como o primeiro elemento amostral e baseando-se no rol de lançamentos indicados abaixo, a amostra é obtida da seguinte forma:

Continua

¹⁹ A probabilidade de seleção de cada registro é diretamente proporcional ao seu valor.

Tabela 7.3: Seleção dos Elementos Amostrais

(em R\$)

Item ^(a)	Valor Registrado	Valor Acumulado	Unidade Monetária Seleccionada ^(b)
1	10.000,00	10.000,00	–
2	18.500,00	28.500,00	16.000,00
3	20.000,00	48.500,00	–
4	6.000,00	54.500,00	49.333,00
5	6.500,00	61.000,00	–
6	46.000,00	107.000,00	82.666,00
7	9.600,00	116.600,00	115.999,00
8	16.400,00	133.000,00	–
9	126.750,00	259.750,00	149.332,00
			182.665,00
			215.998,00
			249.331,00
10	27.250,00	287.000,00	282.664,00
...

Notas: (a) Itens da população.

(b) Neste exemplo, apenas o primeiro item da amostra foi selecionado aleatoriamente, enquanto que a escolha dos demais ocorreu a intervalos regulares, pois isso permite uma melhor visualização de como a amostra deve se distribuir. Esse procedimento, contudo, pode gerar resultados viesados caso os registros apresentem flutuações cíclicas. Por essa razão, recomenda-se a seleção aleatória de um item em cada intervalo amostral.

Convém notar que os registros com valores iguais ou superiores ao intervalo amostral são necessariamente selecionados para a amostra. O 9º item do rol mostrado acima, p. ex., consta quatro vezes da amostra por ser 3,8 vezes superior ao intervalo. Nesses casos, a equipe de auditoria tem duas alternativas:

- a) diminuir o nível de confiança (α) em decorrência da redução do tamanho efetivo da amostra (no presente exemplo, supondo-se que os demais itens são inferiores a R\$ 33.333,33, o tamanho efetivo da amostra é 57);
- b) examinar separadamente os itens com valores elevados, considerando-os uma única vez na amostra e ajustando-se o intervalo amostral para os demais elementos.

A segunda alternativa é a mais adequada por não implicar redução no nível de confiança.

Uma vez selecionada e examinada a amostra, deve-se estimar a diferença entre os montantes registrado e estimado. Esse cálculo envolve as seguintes etapas:

- 1ª) identificar os itens cujos valores observados diferem dos valores registrados;
- 2ª) calcular a diferença entre os valores observado e registrado;
- 3ª) agrupar as diferenças conforme sejam superiores a zero (registros superestimados) ou inferiores a zero (registros subestimados);

- 4^a) separar as diferenças associadas a itens com valores **superiores** ao intervalo amostral;
- 5^a) extrapolar as diferenças associadas a itens com valores **inferiores** ao intervalo amostral para todo o intervalo;
- 6^a) colocar em ordem decrescente as diferenças associadas a itens com valores **inferiores** ao intervalo amostral e obter a **Ampliação do Intervalo de Precisão** (A_p) quando a quantidade de erros superar as expectativas da equipe de auditoria, de tal modo que as diferenças observadas sejam ajustadas sem que o tamanho da amostra tenha de ser recalculado;
- 7^a) calcular o saldo líquido dos erros ajustados;
- 8^a) somar as diferenças associadas a itens com valores **superiores** ao saldo líquido dos erros ajustados;
- 9^a) comparar as superestimações e as subestimações totais com a materialidade preestabelecida, verificando se os demonstrativos examinados podem ou não ser considerados fidedignos;
- 10^a) estudar que recomendações deverão ser feitas caso a Materialidade Ampla (M_A) preestabelecida não seja respeitada (p. ex.: determinar a correção dos erros observados; reavaliar o nível de confiança e o tamanho da amostra empregados; determinar aumentos nas provisões; apenas atestar a não-fidedignidade dos demonstrativos examinados).

A Ampliação do Intervalo de Precisão (A_p) é obtida multiplicando-se o **Fator de Ampliação do Intervalo de Precisão** (F_{AP}^j) pela extrapolação dos erros amostrais para o intervalo amostral (I_A):

$$A_p^{item} = F_{AP}^j \cdot \left[\frac{(\text{valor registrado}) - (\text{valor observado})}{\text{valor registrado}} \right] I_A;$$

Onde: $(F_{AP}^j) = C_{(j+1),\alpha} - C_{j,\alpha} - 1$;

j = posição relativa dos erros amostrais dispostos em ordem decrescente de grandeza;

$$C_{1^\circ, \mathbf{a}} = C_{0, \mathbf{a}}.$$

Os Fatores de Ampliação do Intervalo de Precisão (F_{AP}^j) constam da Tabela 7.4.

Tabela 7.4: Fatores de Ampliação do Intervalo de Precisão (F_{AP}^j)

Erros Observados (j)	Nível de Confiança (α)						
	50%	60%	70%	80%	85%	90%	95%
1°	0,00	0,11	0,24	0,38	0,48	0,59	0,75
2°	0,00	0,08	0,18	0,28	0,35	0,43	0,55
3°	0,00	0,07	0,15	0,24	0,29	0,36	0,46
4°	0,00	0,06	0,13	0,21	0,25	0,31	0,40
5°	0,00	0,06	0,12	0,19	0,23	0,28	0,36
6°	0,00	0,05	0,11	0,17	0,21	0,26	0,33
7°	0,00	0,04	0,10	0,16	0,19	0,24	0,31
8°	0,00	0,04	0,09	0,15	0,18	0,22	0,29
9°	0,00	0,04	0,09	0,14	0,17	0,21	0,27
10°	0,00	0,04	0,08	0,13	0,16	0,20	0,26

Fonte: *National Audit Office*, s.d., p. 6.

Todas essas etapas estão contempladas nos formulários que compõem o Anexo IV, elaborados no intuito de auxiliar as equipes de auditoria nos muitos cálculos exigidos pela *AUM*

7.2. Aplicação da Amostragem por Unidade Monetária

Verificar se um conjunto hipotético de registros contábeis são fidedignos. As características gerais da pesquisa amostral são (*vide* Anexo V):

- montante registrado (M): R\$ 2.000.000,00;
- materialidade ampla (M_A): R\$ 100.000,00;
- materialidade restrita (M_R): R\$ 95.000,00;
- proporção máxima de erros admitida na população (p): 0,0475;
- quantidade prevista de erros na amostra (i): 0;
- nível de confiança (α): 0,95;
- índice de confiabilidade ($C_{0;0,95}$), dados os itens “c” e “d”: 3,00;
- tamanho da amostra (n): 64, incluindo 5 itens examinados separadamente por serem superiores ao intervalo amostral;
- intervalo amostral (I_A): R\$ 31.250,00.

No decorrer da investigação, constatou-se entre os valores observados e registrados as diferenças indicadas na Tabela 7.5.

Tabela 7.5: Diferenças Observadas na Amostra

Item #	Valor Registrado	Valor Observado	Diferença	Observação
13	43.125,00	42.625,00	500,00	Item superestimado examinado separadamente por ser superior ao intervalo amostral.
29	2.500,00	2.600,00	(100,00)	Item subestimado examinado no contexto da amostra por ser inferior ao intervalo amostral.
35	2.000,00	1.800,00	200,00	Item superestimado examinado no contexto da amostra por ser inferior ao intervalo amostral.
57	4.000,00	4.200,00	(200,00)	Item subestimado examinado no contexto da amostra por ser inferior ao intervalo amostral.

Nota: (#) Itens da amostra.

Agrupando-se os itens superestimados e subestimados, extrapolando-se para o intervalo amostral as diferenças referentes aos itens cujos valores são inferiores ao intervalo, calculando-se os erros globais para maior e para menor e estimando-se a diferença líquida entre os valores registrados e observados, tem-se:

- itens com erros para maior: 13 e 35;
- itens com erros para menor: 29 e 57.
- extrapolação do erro referente ao item 35 (E_{35}):

$$\frac{(\text{valor registrado}) - (\text{valor observado})}{\text{valor registrado}} \cdot I_A = \frac{2000 - 1800}{2000} \cdot 31250 = 3125;$$
- extrapolação do erro referente ao item 29 (E_{29}):

$$\frac{2500 - 2600}{2500} \cdot 31250 = -1250;$$
- extrapolação do erro referente ao item 57 (E_{57}):

$$\frac{4000 - 4200}{4000} \cdot 31250 = -1562,5;$$
- impacto líquido dos erros extrapolados:

$$E_{29} + E_{35} + E_{57} = 312,5$$

Como há mais de um erro para menor referente a itens com valores inferiores ao intervalo amostral, esses erros devem ser dispostos em ordem decrescente, cabendo a primeira posição ao item 57 e a segunda ao item 29. Essa ordenação deve-se à necessidade de se multiplicar os erros por diferentes Fatores de Ampliação do Intervalo de Precisão (F_{AP}^j). Para um nível de confiança de 95% e uma previsão inicial de que a amostra não conteria erros, o que faz i igual a 0 para j igual a 1, tem-se:

- $F_{AP}^{1^\circ} = C_{1;0,95} - C_{0;0,95} - 1 = 4,75 - 3,00 - 1 = 0,75;$
- $F_{AP}^{2^\circ} = C_{2;0,95} - C_{1;0,95} - 1 = 6,30 - 4,75 - 1 = 0,55.$

O Ampliação do Intervalo de Precisão (A_P) é obtido multiplicando-se os F_{AP}^j pelas extrapolações dos erros referentes a itens com valores inferiores ao intervalo amostral²⁰ e somando-se os valores referentes a cada tipo de erro:

- a) $A_P^{35} = E_{35} \cdot F_{AP}^{1^\circ} = 3125 \cdot 0,75 = 2343,75$ (A_P para os itens superestimados);
- b) $A_P^{57} = E_{57} \cdot F_{AP}^{1^\circ} = -1562,5 \cdot 0,75 = -1171,87$;
- c) $A_P^{29} = E_{29} \cdot F_{AP}^{2^\circ} = -1250 \cdot 0,55 = -687,50$;
- d) $A_P^{29} + A_P^{57} = -1859,37$ (A_P para os itens subestimados).

Acumulando-se os erros amostrais, a materialidade restrita e a ampliação do intervalo de precisão para erros superestimados e subestimados, tem-se:

- a) erros para maior:
 $(E_{29} + E_{35} + E_{57}) + E_{I3} + M_R + A_P^{35} = 312,5 + 500 + 95000 + 2343,75 = 98156,25$;
- b) erros para menor:
 $-[(E_{29} + E_{35} + E_{57}) + M_R + (A_P^{29} + A_P^{57})] = -312,5 + 95000 + 1859,37 = 96546,87$.

Como a materialidade ampla é superior aos dois valores calculados acima, conclui-se que os lançamentos contábeis examinados são, com 95% de certeza, fidedignos, não apresentando uma diferença superior a 5%.

Convém notar que o uso de uma materialidade restrita de R\$ 95.000,00, ao aumentar o tamanho da amostra de 60 para 64 elementos, permitiu a obtenção de estimativas robustas embora tenham sido observados 4 erros na amostra.

²⁰ É importante lembrar que os itens superestimados e subestimados devem ser tratadas separadamente. Dessa forma, os erros para maior e para menor mais expressivos são multiplicados pela FAPI.

8. TIPOS DE ERRO EM AMOSTRAGEM

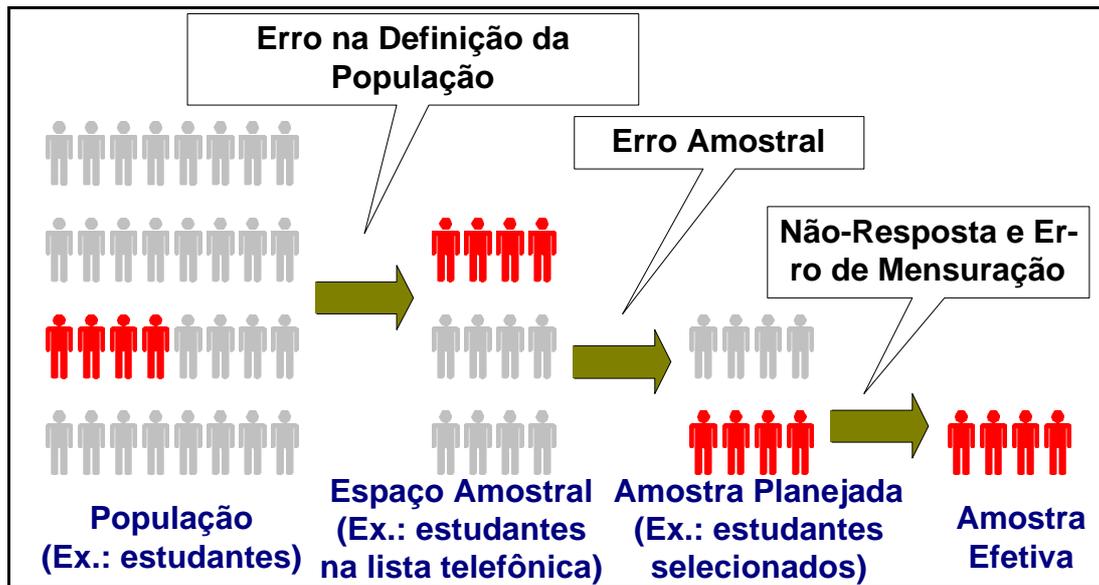
Os métodos de estimação apresentados anteriormente assumem que algum tipo de amostragem probabilística está sendo empregado e que os valores obtidos representam os valores corretos de todos os elementos amostrais selecionados. Nesse caso, os erros de estimação decorrem unicamente da variação randômica presente quando se examina, em vez de toda população N , um subconjunto n desta (i.e., **erros amostrais aleatórios**).

Esses pressupostos são respeitados por pesquisas de campo simples, com instrumentos de medida precisos e execução cuidadosa. As pesquisas complexas, porém, especialmente aquelas que envolvem mensurações de difícil obtenção, estão sujeitas a mais três tipos de erro:

- a) **Erro na definição da população:** Não inclusão no cadastro de alguns grupos de itens ou indivíduos, introduzindo um viés seleção da amostra.
- b) **Erro de mensuração:** Ambigüidades presentes nas respostas coletadas em decorrência da formulação de perguntas imprecisas, da influência do entrevistador sobre o entrevistado ou da falta de rigor do respondente. Essas deficiências podem ser minimizadas tomando-se cuidado na elaboração dos formulários, treinando-se adequadamente os entrevistadores, examinando-se os dados atentamente, conferindo-se as respostas atípicas, e selecionando-se, aleatoriamente, alguns respondente no intuito de atestar a validade das respostas obtidas.
- c) **Não-resposta:** Falha na coleta de dados sobre todos os itens ou indivíduos selecionados para a amostra. Como os pesquisadores não podem assumir, em geral, que os respondentes e os não-respondentes são semelhantes, esses últimos devem, após algum tempo, ser novamente consultados no intuito de obter respostas apropriadas. Após relacionar os novos resultados com os resultados obtidos originalmente, os pesquisadores procurarão gerar estimativas válidas sobre a população.

A Figura 8.1 mostra como os quatro tipos de erro podem afetar a amostragem.

Figura 8.1: Tipos de Erro em Amostragem



Fonte: McGill, 1997.

Os erros na definição da população e de mensuração, e os erros devidos a não-resposta demandam modificações na teoria padrão de amostragem destinadas, que permitam uma melhor alocação dos recursos disponíveis, de modo que:

- haja um equilíbrio entre os esforços empreendidos em favor da redução dos erros amostrais aleatórios e da redução dos demais tipos de erros;
- sejam obtidos métodos de cálculo dos erros-padrão e dos intervalos de confiança que permaneçam válidos.

Os itens subseqüentes tratarão do problema da não-resposta por ser aquele que melhor admite tratamento matemático.

8.1. Efeitos da Não-Resposta

No estudo da não-resposta, é conveniente que a população seja concebida como pertencendo a dois estratos: o primeiro contendo todas os elementos que poderiam ser medidos se fossem selecionados para a amostra; e o segundo contendo os demais elementos, cuja medida não seria obtida.

O tamanho relativo dos dois estratos depende do método de coleta de dados empregado. Uma pesquisa de campo com até três tentativas de contato de cada elemento amostral, supervisionada por pessoas especialmente persuasivas, terá menos não-respostas do que uma pesquisa com uma única tentativa de contato.

Naturalmente, as amostras não proporcionam informações sobre o estrato composto pelas não-respostas. Isso não é um problema quando se pode assumir que os dois estratos apresentam as mesmas características. Frequentemente, porém, não é esse o caso. Dessa forma, as estimativas obtidas podem ser viesadas. Na média amostral, p. ex., o viés é definido pela seguinte equação:

$$\text{Viés} = \bar{X}_r - \bar{X} = \frac{N_{nr}}{N} \cdot (\bar{X}_r - \bar{X}_{nr}),$$

Onde: \bar{X} = média amostral correta mas não-conhecida;

\bar{X}_r = média dos elementos amostrais efetivamente medidos;

\bar{X}_{nr} = média dos elementos amostrais não-medidos;

N = tamanho da população;

N_{nr} = tamanho do estrato contendo elementos que não podem ser medidos.

Como \bar{X}_{nr} não pode ser estimada a partir dos elementos amostrais examinados, a extensão do viés só pode ser definido mediante informações obtidas em outras fontes.

Nas variáveis contínuas, os limites que podem ser imputados com certeza aos vieses são tão amplos que não têm valor prático. Portanto, no presente contexto, quando a proporção de não-respondentes é elevada, os intervalos de confiança não podem ser estimados com precisão, a não ser por meio de aproximações não-corroboradas pelos dados disponíveis.

Já nas amostragens que visam estimar uma proporção, o problema anterior não é tão crítico, pois no estrato composto pelas não-respostas a proporção desejada situa-se, necessariamente, no intervalo [0; 1]. **Conhecendo-se** $\frac{N_{nr}}{N}$ a partir de pesquisas realizadas anteriormente, os limites inferior (p_{inf}) e superior (p_{sup}) do intervalo de confiança são:

$$\text{a) } p_{inf} = \frac{N_r}{N} \cdot \left[p_r - Z \cdot \sqrt{\frac{p_r \cdot (1 - p_r)}{n_r}} \right], \text{ com } p_{nr} = 0;^{21}$$

$$\text{b) } p_{sup} = \frac{N_r}{N} \cdot \left[p_r + Z \cdot \sqrt{\frac{p_r \cdot (1 - p_r)}{n_r}} \right] + \frac{N_{nr}}{N}, \text{ com } p_{nr} = 1.^{22}$$

Quanto maior o índice de não-resposta $\left(\frac{N_{nr}}{N} \right)$, maior a amplitude do intervalo

$[p_{inf}, p_{sup}]$, tornando cada vez mais imprecisa a estimativa. A Tabela 8.1 mostra os intervalos de confiança, com um nível de confiança de 95% ($Z = 1,96$) e uma amostra com mil elementos ($n = 1.000$), para várias proporções de “sucesso” e índices de não-resposta.

$$^{21} \left\{ \frac{N_{nr}}{N} \cdot \left[p_{nr} - Z \cdot \sqrt{\frac{p_{nr} \cdot (1 - p_{nr})}{n_{nr}}} \right] \right\} = 0.$$

$$^{22} \left[p_{nr} + Z \cdot \sqrt{\frac{p_{nr} \cdot (1 - p_{nr})}{n_{nr}}} \right] = 1.$$

Tabela 8.1: Intervalos de Confiança para Várias Proporções de “Sucesso” e Índices de Não-Resposta ($Z = 1,96$ e $n = 1.000$)

(em %)

$\left(\frac{N_{nr}}{N}\right)$	p_r			
	5	10	20	50
0	[3,6; 6,4]	[8,1; 11,9]	[17,5; 22,5]	[46,9; 53,1]
5	[3,4; 11,1]	[7,7; 16,3]	[16,6; 26,4]	[44,5; 55,5]
10	[3,2; 15,8]	[7,2; 20,8]	[15,6; 30,4]	[42,1; 57,9]
15	[3,0; 20,5]	[6,8; 25,2]	[14,7; 34,3]	[39,6; 60,4]
20	[2,9; 25,2]	[6,4; 29,7]	[13,9; 38,2]	[37,4; 62,8]

Fonte: Cochran, 1977, p. 362.

Para tornar mais claro o impacto do índice de não-resposta sobre o intervalo de confiança, convém calcular qual seria o tamanho da amostra para um dado intervalo supondo-se um índice nulo de não-resposta. O intervalo grifado em negrito na Tabela 8.1, p. ex., corresponde a um erro amostral (e) igual a 0,055 (i.e., metade da extensão do intervalo indicado). Com $Z = 1,96$ e $p = 0,5$, tem-se que a amostra poderia ser, na ausência de não-resposta:

$$n = \left(\frac{1,96}{0,055}\right)^2 \cdot 0,5 \cdot 0,5 = 317,5 \cong 320.$$

Portanto, um índice de não-resposta de apenas 5% exigiu, para um dado intervalo de confiança, uma amostra três vezes superior a que seria idealmente necessária (i.e., uma amostra com 320 elementos em vez dos 1.000 elementos retratados na Tabela 8.1). Isso reforça a idéia de que uma parte substancial dos recursos empregados na amostragem deve ter como objetivo a redução do índice de não-resposta.

Quando $\frac{N_{nr}}{N}$ não é conhecido, como ocorre geralmente, um estimativa conservadora do intervalo de confiança pode ser obtido da seguinte maneira:

- a) $p_{inf} = p_r - Z \cdot \sqrt{\frac{p_r \cdot (1 - p_r)}{n}}$, com $p_{nr} = 0$;
- b) $p_{sup} = \left(p_r + \frac{n_{nr}}{n}\right) + Z \cdot \sqrt{\frac{\left(p_r + \frac{n_{nr}}{n}\right) \left[1 - \left(p_r + \frac{n_{nr}}{n}\right)\right]}{n}}$, com $p_{nr} = 1$ e $n_{nr} =$ qtde de não-respostas na amostra.

Sabendo-se que a pesquisa de campo conterà não-respostas, o tamanho apropriado da amostra, dado p igual a 0,5 e $\frac{N_{nr}}{N}$ conhecido, é definido pela seguinte equação:

$$n' = \frac{Z^2 \cdot 0,25}{e \cdot \left(e - \frac{N_{nr}}{N} \right) \cdot \frac{N_r}{N}} - 1. \quad ^{23}$$

A Tabela 8.2 ilustra o efeito da não-resposta sobre o tamanho da amostra, para um nível de confiança de 95% ($Z = 1,96$) e uma proporção de “sucessos” de 50% ($p = 0,5$).

Tabela 8.2: Tamanho da Amostra para Vários Erros Amostrais e Índices de Não-Resposta ($Z = 1,96$ e $p = 0,5$)

(em %)

$\left(\frac{N_{nr}}{N} \right)$	Erro Amostral Admitido (e)			
	20	15	10	5
0	24	42	96	384
2	27	50	122	653
4	31	60	166	2.000
6	36	75	255	n.d.
8	43	99	521	n.d.
10	53	142	n.d.	n.d.
15	112	n.d.	n.d.	n.d.

Fonte: Cochran, 1977, p. 363.

Nota: “n.d.” corresponde a “não definido”, pois $n < 0$ ou $n \rightarrow \pm \infty$.

Percebe-se, mais uma vez, que altos índices de não-resposta tornam inviável ou extremamente onerosa a obtenção de estimativas precisas.

8.2. Métodos de Redução da Não-Resposta

No que tange à não-resposta, há quatro tipos de elementos amostrais:

- 1º) não-incluídos na amostra (p. ex., cadastro incompleto) ou não-visitados (p. ex., dificuldade de acesso);
- 2º) incapazes ou avessos a fornecer a informação desejada;
- 3º) não-encontrados;
- 4º) pertencentes ao “núcleo rígido” de pessoas avessos ou incapazes de fornecer a informação desejada ou, ainda, que não possam ser encontradas, independentemente dos esforços que sejam feitos nesse sentido.²⁴

A identificação e subsequente redução do primeiro grupo de não-respondentes não é simples. Sugere-se a comparação com os cadastros utilizados e os resultados gerados por outras pesquisas, bem como a realização de pesquisas amostrais destinadas a verificar quão completos são os cadastros disponíveis.

²³ $n' = (n - 1)$ para $N_{nr} = 0 \Rightarrow$ a diferença entre ambos deve-se a aproximações feitas durante a obtenção da nova equação. (Cochran, 1977, p. 363)

²⁴ O viés introduzido por essas elementos não pode ser reduzido

A diminuição do segundo grupo, por sua vez, envolve a cuidadosa elaboração do questionário, cujas perguntas não podem ser ambíguas ou conter termos de difícil compreensão. É importante que o questionário seja testado previamente junto à população visada pela pesquisa para que os eventuais problemas seja identificados oportunamente.²⁵

Já a redução do terceiro grupo exige o correto escalonamento das entrevistas, bem como mais de uma tentativa de contato mediante *call-backs* ou subamostras.

8.2.1. Call-Backs

Nas pesquisas domiciliares que envolvam entrevistas com adultos específicos, selecionados aleatoriamente, os solteiros ou chefes de família devem, p. ex., ser contatados em dias ou horários não-comerciais. Naturalmente, as pesquisas domiciliares cujas perguntas podem ser respondidas por qualquer adulto presente são menos afetadas por esse tipo de problema.

Quanto à quantidade de *call-backs*, ela deve ser definida à luz dos resultados desejados, dos meios disponíveis e das dificuldades enfrentadas por pesquisas similares. Como indicação preliminar da eficácia de cada rodada de entrevistas, Cochrane propõe o seguinte quadro:

Tabela 8.3: Eficácia de Três Tentativas de Contato

Tipo de Respondente	Elementos Amostrais Contatados			Não-Resposta	Total
	1ª Tentativa	2ª Tentativa	3ª Tentativa		
Adulto Qualquer	70	17	8	5	100
Adulto Específico	37	32	23	8	100

Fonte: Cochrane, 1977, p. 365.

É nítida a diferença entre as pesquisas cujas questões podem ser respondidas por qualquer adulto e as pesquisas que devem ser respondidas por adultos específicos, selecionados aleatoriamente, cujos altos índices de “sucesso” nas duas últimas tentativas são uma decorrência das informações obtidas na primeira tentativa acerca dos hábitos dos respondentes visados.

Convém notar que, em termos relativos, as rodadas de entrevistas tornam-se, progressivamente, mais onerosas com a queda na quantidade de locais que podem ser visitados pelos entrevistadores por unidade de área (p. ex., km^2) ou de tempo (p. ex., h), o que limita a quantidade de tentativas de contato que serão feitas.

²⁵ Na formulação de perguntas embaraçosas (p. ex., referentes à prática de atos ilegais) pode-se usar a técnica de **respostas randômicas**, na qual o entrevistado responde “sim” ou “não” a uma de duas perguntas, selecionada aleatoriamente, sem que o entrevistador conheça a pergunta que está sendo respondida. Como, contudo, a distribuição de probabilidades das perguntas apresentadas é conhecida, a proporção dos entrevistados que apresentam a característica visada pela pesquisa pode ser estimada pelo entrevistador. (Cochrane, 1977, p. 392-5)

8.2.2. Subamostras

Ainda sobre os elementos amostrais não-encontrados, em vez de tentar contatar todos os não-respondentes, os pesquisadores podem selecionar aleatoriamente uma subamostra, extrapolando os resultados obtidos para todo grupo. O presente procedimento equívale a uma amostragem estratificada em dois estágios, sendo amplamente usado em pesquisas pelo correio, nas quais alguns elementos amostrais que não devolveram o questionário são entrevistados pessoalmente. Em geral, a menor precisão dos resultados amostrais obtidos nesse caso só se justificam quando os custos dos *call-backs* são muito elevados.

A média amostral é não-viesada quando todos os elementos da subamostra respondem às perguntas feitas pelos entrevistadores. A média é definida pela seguinte equação:

$$\bar{X} = \frac{\sum_{i=1}^{n_{r,1}} X_i + n_{nr,1} \cdot \frac{\sum_{j=1}^{n_{r,2}} X_j}{n_{r,2}}}{n},$$

Onde: \bar{X} = média amostral;

n = tamanho da amostra;

$n_{r,1}$ = quantidade de respondentes na primeira tentativa;

X_i = valores obtidos na primeira tentativa;

$n_{nr,1}$ = quantidade de não-respondentes na primeira tentativa;

$n_{r,2}$ = tamanho da subamostra ($n_{r,2} \supset n_{nr,1}$);

X_j = valores obtidos a partir da subamostra.

A variância da média amostral, necessária para que se calcule o intervalo de confiança, é obtida como mostrado abaixo:

$$S_{\bar{X}}^2 = \left(\frac{N-n}{n \cdot N} \right) S^2 + \frac{(n_{nr,1} - n_{r,2}) \cdot N_{nr,1} \cdot S_2^2}{n \cdot n_{r,2} \cdot N}, \quad 26$$

Onde: $S_{\bar{X}}^2$ = variância da média amostral;

N = tamanho da população;

S^2 = variância dos elementos amostrais;

$N_{nr,1}$ = tamanho do estrato contendo elementos que podem ser medidos na primeira tentativa;

S_2^2 = variância dos elementos da subamostra.

²⁶ A razão $\left(\frac{N_{nr,1}}{N} \right)$ pode ser aproximada por meio da razão $\left(\frac{n_{nr,1}}{n} \right)$.

Por fim, a determinação do tamanho apropriado da amostra no presente contexto depende dos valores desejados para a variância média amostral e a razão entre $n_{nr,1}$ e $n_{r,2}$:

$$n = \frac{N \cdot \left[S^2 + S_2^2 \cdot \left(\frac{N_{nr,1}}{N} \right) \left(\frac{n_{nr,1}}{n_{r,2}} - 1 \right) \right]}{[N \cdot S_{\bar{x}}^2 + S_2^2]}.$$

O Exemplo 8.1 ilustra como o tamanho da amostra pode ser calculado na presença de não-respostas.

Exemplo 8.1: Cálculo da Quantidade de Questionários

Uma pesquisa amostral é feita pelo correio, estimando-se em 50% a taxa de resposta. A precisão desejada equivale àquela que seria obtida por uma amostra aleatória simples de 1.000 questionários se não houvesse não-resposta e $\frac{1}{3}$ dos não-respondentes devem ser entrevistados pessoalmente. Pede-se a quantidade de questionários que deve ser enviada.

Inicialmente, tem-se que:

- a) a taxa estimada de não-resposta: $\frac{N_{nr,1}}{N} = 0,5$;
- b) o tamanho da amostra quando não há não-resposta: $n^* = 1.000$;
- c) a razão entre a quantidade de não-respondentes e o tamanho da subamostra: $\frac{n_{nr,1}}{n_{r,2}} = 3$.
- d) a variância desejada: $S_{\bar{x}}^2 = \frac{S^2}{n^*} = \frac{S^2}{1.000}$.

Supondo-se que os elementos amostrais e da subamostra apresentam variâncias idênticas e que a população é muito numerosa, tem-se que:

$$n = \frac{S^2 \cdot \left[1 + \left(\frac{N_{nr,1}}{N} \right) \left(\frac{n_{nr,1}}{n_{r,2}} - 1 \right) \right]}{S_{\bar{x}}^2} = 1000 \cdot [1 + 0,5 \cdot (3 - 1)] = 2000.$$

Consequentemente, 2.000 questionários deverão ser distribuídos, estimando-se que 1.000 dos quais não serão respondidos. Dessa forma, a subamostra envolverá 334 entrevistas pessoais.

8.3. Métodos de Compensação da Não-Resposta

Ainda que os *call-backs* ou as subamostras sejam sucedidos, é improvável que a não-resposta seja suprimida totalmente. Assim, é comum que se recorra a alguns métodos que buscam compensar esse tipo de problema com base nas características conhecidas dos respondentes e dos não-respondentes. Os principais métodos são:

- a) **pós-estratificação:** criação de estratos com base nas informações coletadas na pesquisa amostral, distribuindo-se os não-respondentes pelos estratos que melhor reflitam as suas características conhecidas;
- b) **substituição da não-resposta:** após a pós-estratificação, as não-respostas são substituídas por respostas selecionadas aleatória ou sistematicamente (p. ex., pela resposta imediatamente anterior) no âmbito do estrato correspondente;
- c) **substituição dos não-respondentes:** ao longo do trabalho de campo, os não-respondentes são substituídos por elementos da população tidos como semelhantes;
- d) **utilização de dados complementares:** as não-respostas podem ser ajustadas por meio de informações coletadas em censos ou estudos amostrais realizados anteriormente;
- e) **ponderação ou substituição baseada nos *call-backs*:** as tentativas de *call-back* podem permitir que se façam inferências acerca das características dos não-respondentes recalcitrantes;

Caso haja indícios de que os não-respondentes recalcitrantes e os respondentes apresentam distribuições de frequência similares (p. ex., os indícios proporcionados por uma ou mais tentativas de *call-back*), as equipes de auditoria ainda têm a opção de não fazer qualquer compensação.

É sumamente importante que as equipes sejam bastante criteriosas no uso das opções relacionadas acima, certificando-se de que as **substituições ou ponderações** efetuadas sejam **efetivamente compatíveis com as características conhecidas dos não-respondentes**. A não-observação desse princípio certamente introduzirá vieses nas estimativas que comprometerão as conclusões do trabalho. Dessa forma, **os relatórios de auditoria deverão explicitar porque e como as eventuais compensações (ou não-compensações) foram feitas**.

ANEXO I
ROTINAS DO *MICROSOFT EXCEL*

Figura A.1: Dados Hipotéticos

Microsoft Excel - TDinam.xls

Arquivo Editar Exibir Inserir Formatar Ferramentas Dados Janela Ajuda

Arial 10 N I S

F1 =

	A	B	C	D	E	F	G	H	I	J
1	DADOS SOBRE 20 HIPOTÉTICOS CURSOS DE ESPECIALIZAÇÃO									
2										
3	Curso	Mensalidade (em R\$)	Tipo	Local	Duração (em horas)					
4	A	149,00	Privado	RJ	120					
5	B	164,00	Privado	SP	360					
6	C	93,00	Privado	SP	360					
7	D	84,00	Privado	CE	360					
8	E	123,00	Privado	RJ	360					
9	F	49,00	Público	SP	360					
10	G	171,00	Privado	RJ	360					
11	H	177,00	Privado	SP	360					
12	I	97,00	Privado	RJ	360					
13	J	63,00	Público	SP	360					
14	K	94,00	Privado	CE	360					
15	L	170,00	Privado	CE	360					
16	M	137,00	Privado	RJ	360					
17	N	126,00	Privado	CE	120					
18	O	95,00	Privado	RJ	360					
19	P	27,00	Público	SP	360					
20	Q	44,00	Público	SP	360					
21	R	100,00	Privado	RJ	360					
22	S	116,00	Privado	RJ	360					
23	T	177,00	Privado	CE	360					
24										

Pronto NUM

Figura A.2: Geração de Estatísticas Descritivas

Estatística descritiva

Entrada

Intervalo de entrada: Dados!\$B\$3:\$B\$23

Agrupado por: Colunas Linhas

Rótulos na primeira linha

Opções de saída

Intervalo de saída:

Nova planilha: Descrição

Nova pasta de trabalho

Resumo estatístico

Nível de confiabilidade p/ média: 95 %

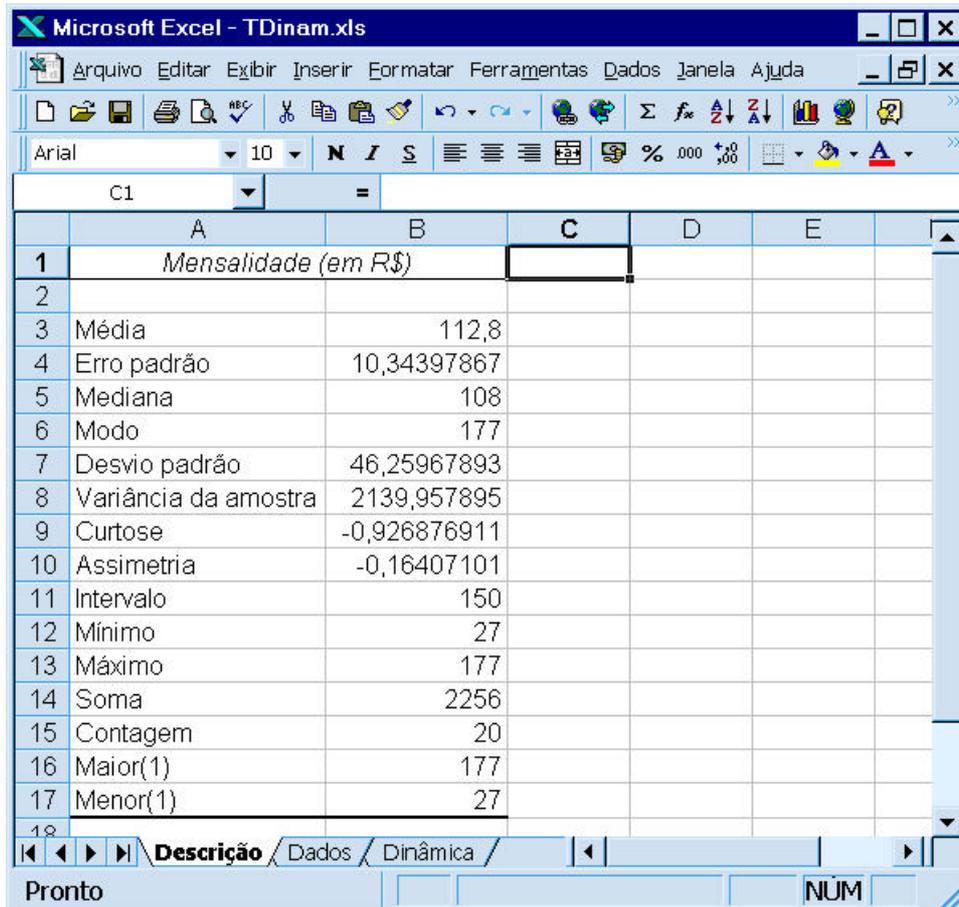
Enésimo maior: 1

Enésimo menor: 1

OK Cancelar Ajuda

Nota: referente aos dados constantes da Figura A.1.

Figura A.3: Estatísticas Descritivas Geradas



	A	B	C	D	E
1	Mensalidade (em R\$)				
2					
3	Média	112,8			
4	Erro padrão	10,34397867			
5	Mediana	108			
6	Modo	177			
7	Desvio padrão	46,25967893			
8	Variância da amostra	2139,957895			
9	Curtose	-0,926876911			
10	Assimetria	-0,16407101			
11	Intervalo	150			
12	Mínimo	27			
13	Máximo	177			
14	Soma	2256			
15	Contagem	20			
16	Maior(1)	177			
17	Menor(1)	27			

Nota: referente aos dados constantes da Figura A.1.

Observações:

a) Erro-Padrão: $S_{\bar{x}} = \frac{S}{\sqrt{n}}$.

UTILIZANDO O *MICROSOFT EXCEL*:

`=(DESVPAD(núm1;núm2;...)/RAIZ(CONT.NÚM(núm1;núm2;...)))`.

b) Assimetria:

- i. $e > 0 \Rightarrow$ curva de frequência de um conjunto de n observações $X_1, X_2, X_3, \dots, X_n$ é assimétrica à direita;
- ii. $e = 0 \Rightarrow$ curva de frequência de um conjunto de n observações $X_1, X_2, X_3, \dots, X_n$ é simétrica;
- iii. $e < 0 \Rightarrow$ curva de frequência de um conjunto de n observações $X_1, X_2, X_3, \dots, X_n$ é assimétrica à esquerda.

UTILIZANDO O *MICROSOFT EXCEL*: `=DISTRORÇÃO(núm1;núm2;...)`.

c) Curtose:

- i. $k > 0 \Rightarrow$ curva de frequência de um conjunto de n observações $X_1, X_2, X_3, \dots, X_n$ é mais achatada do que a distribuição normal;
- ii. $k = 0 \Rightarrow$ curva de frequência de um conjunto de n observações $X_1, X_2, X_3, \dots, X_n$ é tão achatada quanto a distribuição normal;
- iii. $k < 0 \Rightarrow$ curva de frequência de um conjunto de n observações $X_1, X_2, X_3, \dots, X_n$ é menos achatada quanto a distribuição normal.

UTILIZANDO O MICROSOFT EXCEL: =CURT(núm1;núm2;...).

Figura A.4: Agrupamento de Dados (Etapa 1 de 4)

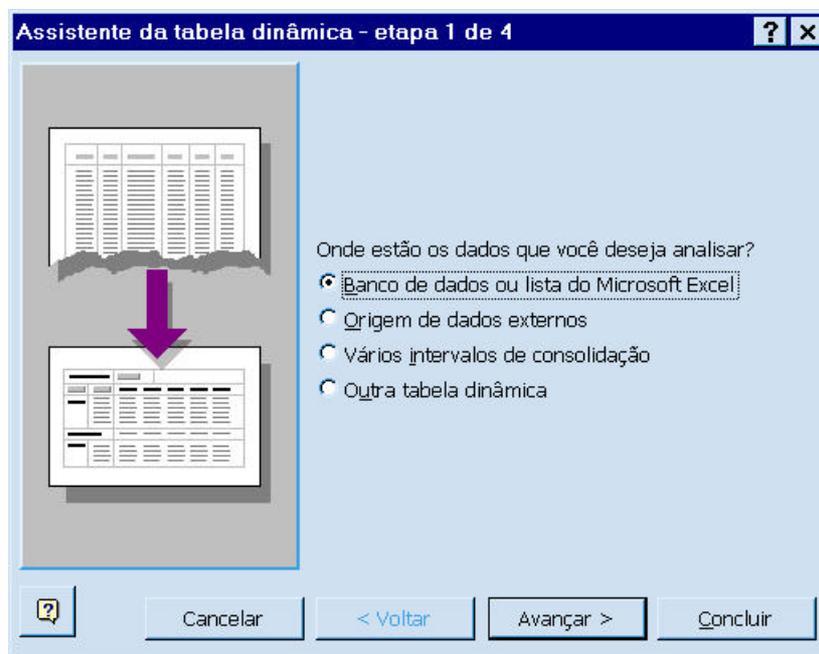
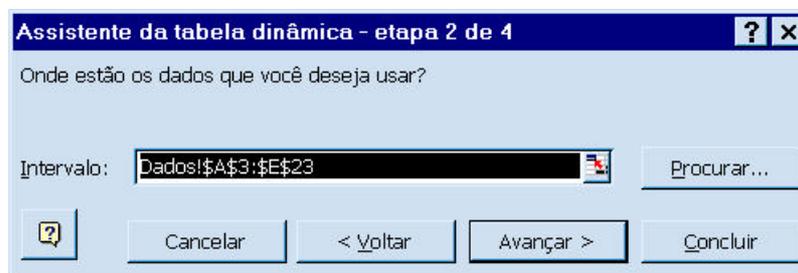


Figura A.5: Agrupamento de Dados (Etapa 2 de 4)



Nota: referente aos dados constantes da Figura A.1.

Figura A.6: Agrupamento de Dados (Etapa 3 de 4)

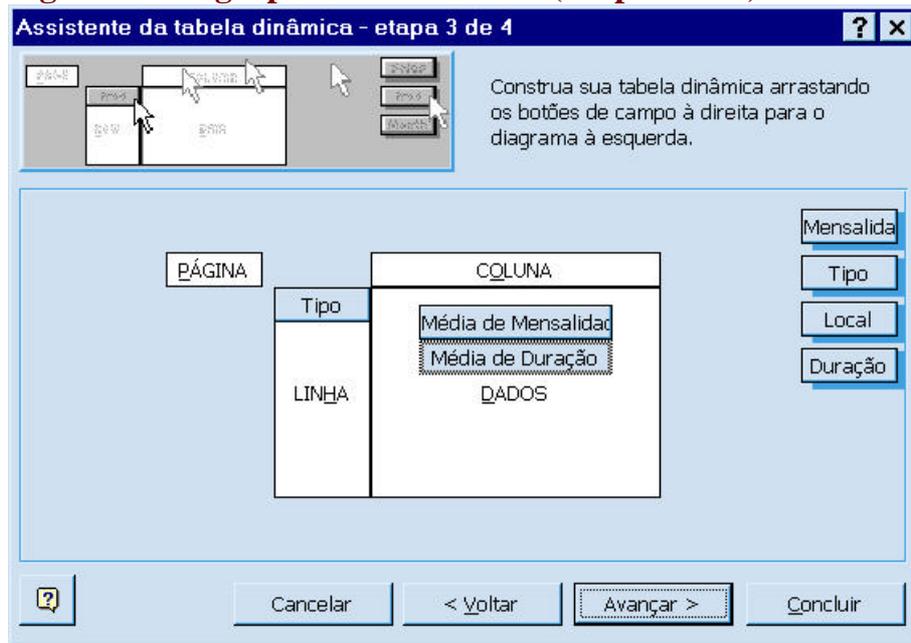
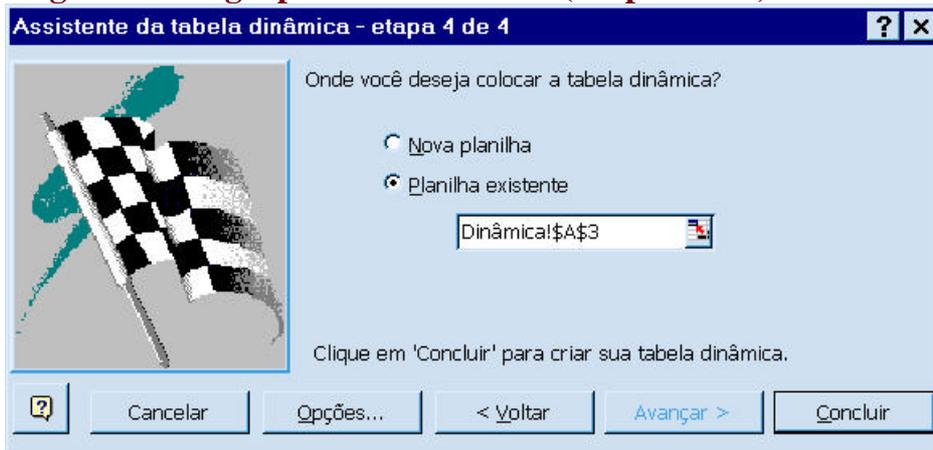


Figura A.7: Agrupamento de Dados (Etapa 4 de 4)



Nota: referente aos dados constantes da Figura A.1.

Figura A.8: Dados Agrupados

	A	B	C	D	E	F	G
1	RELATÓRIO DE TABELA DINÂMICA						
2							
3	Tipo	Dados	Total				
4	Privado	Média de Mensalidade (em R\$)	129,5625				
5		Média de Duração (em horas)	330				
6	Público	Média de Mensalidade (em R\$)	45,75				
7		Média de Duração (em horas)	360				
8	Total	Média de Mensalidade (em R\$)	112,8				
9		Média de Duração (em horas)	336				

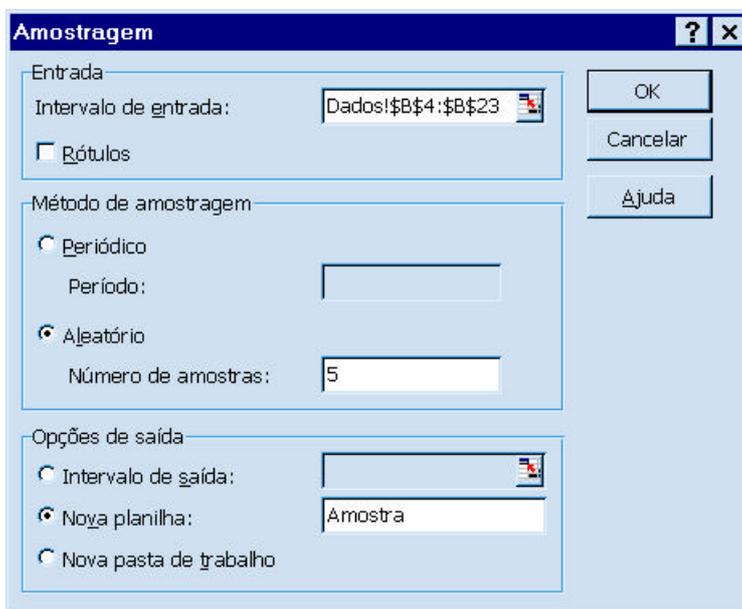
Nota: referente aos dados constantes da Figura A.1.

Tabela A.1: Obtenção da Amplitude Semi-Interquartílica e da Meia Amplitude Semi-Interquartílica

	A	B
1	CÁLCULO AUTOMÁTICO DA AMPLITUDE SEMI-INTERQUARTÍLICA E DA MEIA AMPLITUDE SEMI-INTERQUARTÍLICA #	
2		
3	POSIÇÃO INICIAL	
4	Primeiro Quartil:	= (CONT.NÚM(valor1;valor2;...)+1)/4
5	Terceiro Quartil:	= (3*(CONT.NÚM(valor1;valor2;...)+1)/4
6		
7	PARA O PRIMEIRO QUARTIL	
8	=SE(B4=INT(B4);"utilize a posição inicial:"; SE(B4=TETO(B4;0,5);"utilize as seguintes posições:"; "arredonde para a seguinte posição:"))	=SE(A8="utilize a posição inicial:"; ARREDMULTB(B4;1); ARRED(B4;0))
9		=SE(A8="utilize a posição inicial:"; TETO(B4;1); "")
10	=SE(A8="utilize a posição inicial:"; "valores das posições indicadas:"; "valor da posição indicada:")	=MENOR(valor1;valor2;...;B8)
11		=SE(A8="utilize a posição inicial:"; MENOR(valor1;valor2;...); "")
12		
13	PARA O TERCEIRO QUARTIL	
14	=SE(B5=INT(B5);"utilize a posição inicial:"; SE(B5=TETO(B5;0,5);"utilize as seguintes posições:"; "arredonde para a seguinte posição:"))	=SE(A14="utilize a posição inicial:";ARREDMULTB(B5;1);ARRED(B5;0))
15		=SE(A14="utilize a posição inicial:"; TETO(B5;1); "")
16	=SE(A14="utilize a posição inicial:"; "valores das posições indicadas:"; "valor da posição indicada:")	=MENOR(valor1;valor2;...;B14)
17		=SE(A14="utilize a posição inicial:"; MENOR(valor1;valor2;...); "")
18		
19	CONCLUSÃO	
20	Primeiro Quartil:	=SE(A8="utilize a posição inicial:";(B10+B11)/2;B10)
21	Terceiro Quartil:	=SE(A14="utilize a posição inicial:";(B16+B17)/2;B16)
22	Amplitude Semi-Interquartílica	= (B20+B21)
23	Meia Amplitude Semi-Interquartílica	= (B22)/2

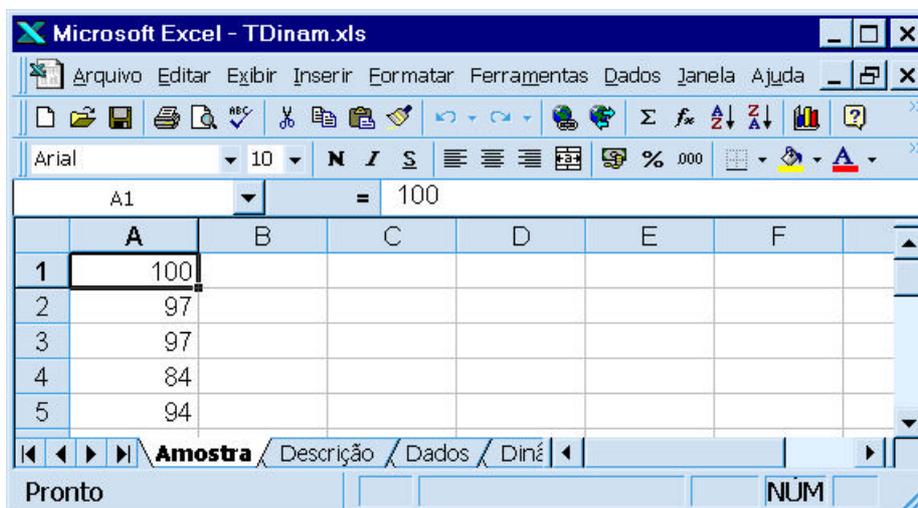
Nota: (#) os comandos do tipo “valor#” correspondem a valores extraídos de outras planilhas ou inseridos pelos pesquisadores.

Figura A.9: Seleção de Amostras Aleatórias Simples com Reposição



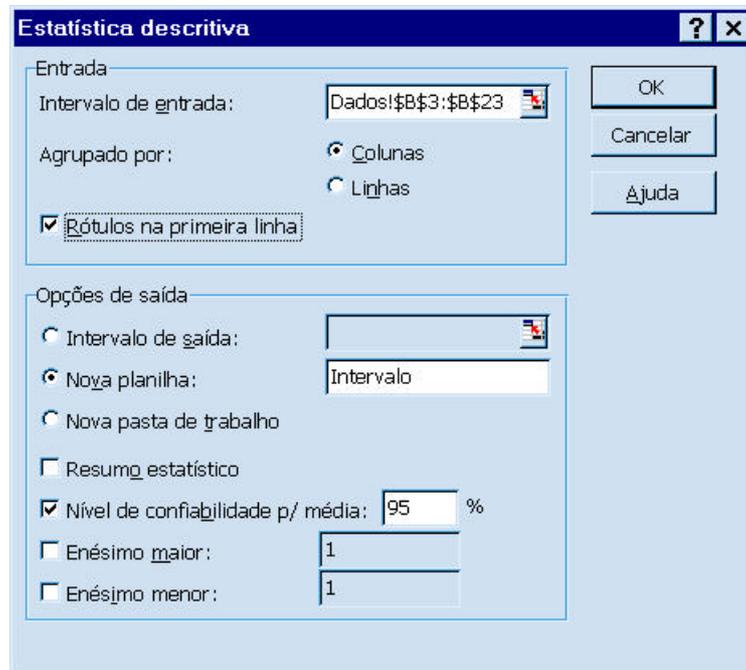
Nota: referente aos dados constantes da Figura A.1.

Figura A.10: Amostra Aleatória Simples Selecionada



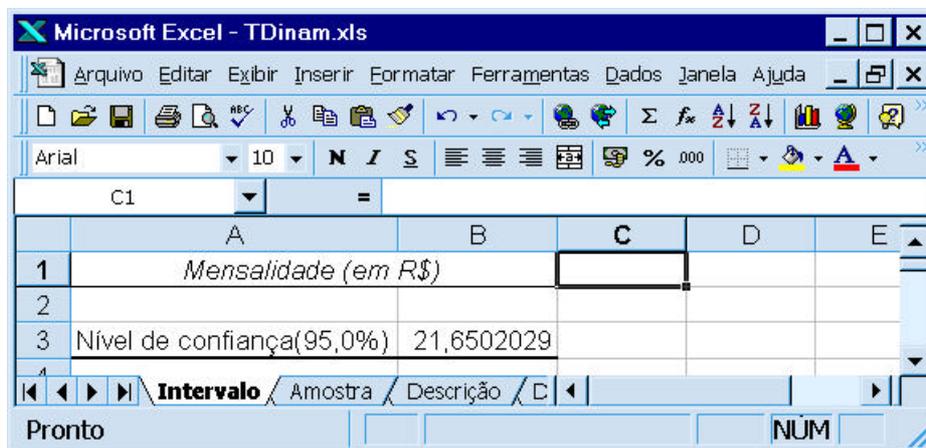
Nota: referente aos dados constantes da Figura A.1.

Figura A.11: Estimação do Intervalo de Confiança para Médias com Desvio-Padrão Conhecido



Nota: referente aos dados constantes da Figura A.1.

Figura A.12: Intervalo de Confiança Estimado



Notas: (a) o valor obtido corresponde à metade do intervalo de confiança estimado em torno da média amostral;

(b) referente aos dados constantes da Figura A.1.

Tabela A.2: Estimação de Intervalos de Confiança

	A	B	C	D	E	F
1	ESTIMAÇÃO DE INTERVALOS DE CONFIANÇA^(a)					
2	Média (com σ conhecido)		Média (com σ desconhecido)		Proporção	
3	n	valor1	n	valor6	n	valor11
4	Média Amostral	valor2	Média Amostral	valor7	“Sucessos”	valor12
5	Desvio-Padrão Populacional	valor3	Desvio-Padrão Amostral	valor8	“Fracassos”	=F3-F4
6	Nível de Confiança ^(b)	valor4	Nível de Confiança	valor9	Nível de Confiança ^(b)	valor13
7						
8	Erro-Padrão	=B5/RAIZ(B3)	Erro-Padrão	=D5/RAIZ(D3)	Proporção Amostral	=F4/F3
9	Z	=INV.NORMP(0,5+(B6/2))	g.l.	=(D3-1)	Z	=INV.NORMP(0,5+(F6/2))
10			t	=INVT(1-D6;D9)	Erro-Padrão	=RAIZ(F8*(1-F8)/F3)
11	Meio Intervalo de Confiança	=INT.CONFIANÇA(1-B6;B5;B3)	Meio Intervalo de Confiança	=D10*D8	Meio Intervalo de Confiança	=F9*F10
12	Limite Inferior	=B4-B11	Limite Inferior	=D4-D11	Limite Inferior	=F8-F11
13	Limite Superior	=B4+B11	Limite Superior	=D4+D11	Limite Superior	=F8+F11
14	Ajuste para Populações Finitas					
15	N	valor5	N	valor10	N	valor14
16	CPF	=RAIZ((B15-B3)/(B15-1))	CPF	=RAIZ((D15-D3)/(D15-1))	CPF	=RAIZ((F15-F3)/(F15-1))
17	Meio Inter. de Conf. Ajustado	=B11*B16	Meio Inter. de Conf. Ajustado	=D11*D16	Meio Inter. de Conf. Ajustado	=F11*F16
18	Limite Inferior Ajustado	=B4-B17	Limite Inferior Ajustado	=D4-D17	Limite Inferior Ajustado	=F8-F17
19	Limite Superior Ajustado	=B4+B17	Limite Superior Ajustado	=D4+D17	Limite Superior Ajustado	=F8+F17

Notas: (a) os comandos do tipo “valor#” correspondem a valores extraídos de outras planilhas ou inseridos pelos pesquisadores;

(b) o nível de confiança deve ser informado na sua forma decimal.

Tabela A.3: Determinação do Tamanho da Amostra

	A	B	C	D
1	DETERMINAÇÃO DO TAMANHO DA AMOSTRA^(a)			
2	Média		Proporção	
3	Nível de Confiança Desejado ^(b)	valor1	Nível de Confiança Desejado ^(b)	valor5
4	Estimativa Inicial do Desvio-Padrão	valor2	Estimativa Inicial do Desvio-Padrão	valor6
5	Erro Amostral Admitido	valor3	Erro Amostral Admitido	valor7
6	Z	=INV.NORMP (0,5+(B3/2))	Z	=INV.NORMP (0,5+(D3/2))
7	Tamanho da Amostra	=((B6*B4)/B5)^2	Tamanho da Amostra	=(D6^2*D4*(1-D4))/D5^2
8	Amostra Requerida	=ARREDONDAR.PARA. CIMA(B7;0)	Amostra Requerida	=ARREDONDAR.PARA. CIMA(D7;0)
9	Ajuste para Populações Finitas			
10	N	valor4	N	valor8
11	n	=(B7*B10)/(B7+B10-1)	n	=(D7*D10)/(D7+D10-1)
12	Amostra Requerida Ajustada	=ARREDONDAR.PARA. CIMA(B11;0)	Amostra Requerida Ajustada	=ARREDONDAR.PARA. CIMA(D11;0)

Notas: (a) os comandos do tipo “valor#” correspondem a valores extraídos de outras planilhas ou inseridos pelos pesquisadores;

(b) o nível de confiança deve ser informado na sua forma decimal.

Tabela A.4: Estimação do Intervalo de Confiança para Totais

	A	B
1	ESTIMAÇÃO DO INTERVALO DE CONFIANÇA PARA TOTAIS^(a)	
2		
3	n	valor1
4	Média Amostral	valor2
5	Desvio-Padrão Amostral	valor3
6	Nível de Confiança ^(b)	valor4
7	N	valor5
8	Total	=B7*B4
9	CPF	=RAIZ((B7-B3)/(B7-1))
10	Erro-Padrão	=(B7*B5*B9)/RAIZ(B3)
11	g.l.	=B3-1
12	t	=INVT(1-B6;B11)
13	Meio Intervalo de Confiança Ajustado	=B12*B10
14	Limite Inferior Ajustado	=B8-B13
15	Limite Superior Ajustado	=B8+B13

Notas: (a) os comandos do tipo “valor#” correspondem a valores extraídos de outras planilhas ou inseridos pelos pesquisadores;

(b) o nível de confiança deve ser informado na sua forma decimal.

Tabela A.5: Estimação do Intervalo de Confiança para Diferenças

	A	B	C	D	E	F
1	ESTIMAÇÃO DO INTERVALO DE CONFIANÇA PARA DIFERENÇAS^(a)					
2	Diferenças ^(b)			Quadrado dos Desvios das Obs. com Diferença		
3	valor1	Somatório das Diferenças	=SOMA (A3:A16)	=(A3-\$C\$6)^2	Som. do Quad. dos Desv. das Obs. com Dif.	=SOMA (D3:D16)
4	valor2	Contagem das Diferenças	=CONT.VALORES (A3:A16)	=(A4-\$C\$6)^2	Produto da Dif. Média pela Qtde. de Obs. sem Dif.	=C7*(-C6)^2
5	valor3	n	valor15	=(A5-\$C\$6)^2	Somatório dos Desvios	=F3+F4
6	valor4	Diferença Média	=C3/C5	=(A6-\$C\$6)^2	g.l.	=C4-1
7	valor5	Qtde de Obs. sem Diferença	=C5-C4	=(A7-\$C\$6)^2	Variância	=F5/F6
8	valor6			=(A8-\$C\$6)^2	Desvio-Padrão	=RAIZ(F7)
9	valor7			=(A9-\$C\$6)^2		
10	valor8			=(A10-\$C\$6)^2	Nível de Conf. ^(c)	valor16
11	valor9			=(A11-\$C\$6)^2	N	valor 17
12	valor10			=(A12-\$C\$6)^2	Diferença Total	=F11*C6
13	valor11			=(A13-\$C\$6)^2	CPF	=RAIZ ((F11-C5)/(F11-1))
14	valor12			=(A14-\$C\$6)^2	Erro-Padrão	=(F11*F8*F13)/RAIZ(C5)
15	valor13			=(A15-\$C\$6)^2	g.l.	=C5-1
16	valor14			=(A16-\$C\$6)^2	t	=INVT (1-F10;F15)
17					Meio Intervalo de Conf. Ajust.	=F16*F14
18					Limite Inferior Ajustado	=F12-F17
19					Limite Superior Ajustado	=F12+F17

Notas: (a) os comandos do tipo “valor#” correspondem a valores extraídos de outras planilhas ou inseridos pelos pesquisadores;

(b) por hipótese, foram observadas catorze diferenças entre os valores constantes do cadastro e os valores observados;

(c) o nível de confiança deve ser informado na sua forma decimal.

ANEXO II
DISTRIBUIÇÕES NORMAL PADRÃO &
“*t*” DE STUDENT

ÁREAS DE UMA DISTRIBUIÇÃO NORMAL PADRÃO

USANDO MICROSOFT EXCEL: =DIST.NORMP(z).

Onde: $\text{DIST.NORMP}(z) = \left(\frac{1 - \alpha}{2} \right)$

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

DISTRIBUIÇÃO “t” DE STUDENT BI-CAUDAL

USANDO *MICROSOFT EXCEL*: =INVT(probabilidade;graus de liberdade).

Onde: probabilidade de que se cometa o erro do Tipo I = α ;
 graus de liberdade = g.l.

g.l.	α							
	0,2000	0,1000	0,0500	0,0200	0,0100	0,0050	0,0020	0,0010
1	3,0777	6,3137	12,7062	31,8210	63,6559	127,3211	318,2888	636,5776
2	1,8856	2,9200	4,3027	6,9645	9,9250	14,0892	22,3285	31,5998
3	1,6377	2,3534	3,1824	4,5407	5,8408	7,4532	10,2143	12,9244
4	1,5332	2,1318	2,7765	3,7469	4,6041	5,5975	7,1729	8,6101
5	1,4759	2,0150	2,5706	3,3649	4,0321	4,7733	5,8935	6,8685
6	1,4398	1,9432	2,4469	3,1427	3,7074	4,3168	5,2075	5,9587
7	1,4149	1,8946	2,3646	2,9979	3,4995	4,0294	4,7853	5,4081
8	1,3968	1,8595	2,3060	2,8965	3,3554	3,8325	4,5008	5,0414
9	1,3830	1,8331	2,2622	2,8214	3,2498	3,6896	4,2969	4,7809
10	1,3722	1,8125	2,2281	2,7638	3,1693	3,5814	4,1437	4,5868
11	1,3634	1,7959	2,2010	2,7181	3,1058	3,4966	4,0248	4,4369
12	1,3562	1,7823	2,1788	2,6810	3,0545	3,4284	3,9296	4,3178
13	1,3502	1,7709	2,1604	2,6503	3,0123	3,3725	3,8520	4,2209
14	1,3450	1,7613	2,1448	2,6245	2,9768	3,3257	3,7874	4,1403
15	1,3406	1,7531	2,1315	2,6025	2,9467	3,2860	3,7329	4,0728
16	1,3368	1,7459	2,1199	2,5835	2,9208	3,2520	3,6861	4,0149
17	1,3334	1,7396	2,1098	2,5669	2,8982	3,2224	3,6458	3,9651
18	1,3304	1,7341	2,1009	2,5524	2,8784	3,1966	3,6105	3,9217
19	1,3277	1,7291	2,0930	2,5395	2,8609	3,1737	3,5793	3,8833
20	1,3253	1,7247	2,0860	2,5280	2,8453	3,1534	3,5518	3,8496
21	1,3232	1,7207	2,0796	2,5176	2,8314	3,1352	3,5271	3,8193
22	1,3212	1,7171	2,0739	2,5083	2,8188	3,1188	3,5050	3,7922
23	1,3195	1,7139	2,0687	2,4999	2,8073	3,1040	3,4850	3,7676
24	1,3178	1,7109	2,0639	2,4922	2,7970	3,0905	3,4668	3,7454
25	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,4502	3,7251
26	1,3150	1,7056	2,0555	2,4786	2,7787	3,0669	3,4350	3,7067
27	1,3137	1,7033	2,0518	2,4727	2,7707	3,0565	3,4210	3,6895
28	1,3125	1,7011	2,0484	2,4671	2,7633	3,0470	3,4082	3,6739
29	1,3114	1,6991	2,0452	2,4620	2,7564	3,0380	3,3963	3,6595
30	1,3104	1,6973	2,0423	2,4573	2,7500	3,0298	3,3852	3,6460
40	1,3031	1,6839	2,0211	2,4233	2,7045	2,9712	3,3069	3,5510
60	1,2958	1,6706	2,0003	2,3901	2,6603	2,9146	3,2317	3,4602
120	1,2886	1,6576	1,9799	2,3578	2,6174	2,8599	3,1595	3,3734
+∞	1,2816	1,6449	1,9600	2,3264	2,5758	2,8070	3,0902	3,2905

Nota: Para $\alpha = 0,05$, as estatísticas “t” de Student, com g.l. = +∞, e Z são iguais a 1,96.

ANEXO III
BREVIÁRIO DE FÓRMULAS

AMOSTRAGEM ALEATÓRIA SIMPLES

I. Utilidade:

A população forma um todo indissociável em relação ao objeto da pesquisa e os custos do trabalho de campo não são elevados \Rightarrow p. ex., no caso de uma pesquisa que pretenda estimar a proporção de beneficiários do crédito educativo que conseguem completar o curso superior, cujos questionários sejam remetidos pelo correio e cujo preenchimento seja compulsório (ou seja, a taxa de não-resposta é assumida baixa), é possível que a estratificação por curso, gênero ou localidade, entre outras possibilidades, não seja relevante, pois os fatores determinantes do sucesso ou insucesso seriam comuns a todos os possíveis subconjuntos.

II. Estimativa do Intervalo:

da Média:

- com \mathbf{s} conhecido e população infinita:

vide item 5.1

$$\left[\bar{X} - Z \cdot \left(\frac{\sigma}{\sqrt{n}} \right); \bar{X} + Z \cdot \left(\frac{\sigma}{\sqrt{n}} \right) \right];$$

- com \mathbf{s} desconhecido e população infinita:

vide item 5.2

$$\left[\bar{X} - t_{n-1} \cdot \left(\frac{S}{\sqrt{n}} \right); \bar{X} + t_{n-1} \cdot \left(\frac{S}{\sqrt{n}} \right) \right];$$

- com \mathbf{s} conhecido e população finita:

vide item 5.6

$$\left[\bar{X} - Z \cdot \left(\frac{\sigma}{\sqrt{n}} \right) \left(\sqrt{\frac{N-n}{N-1}} \right); \bar{X} + Z \cdot \left(\frac{\sigma}{\sqrt{n}} \right) \left(\sqrt{\frac{N-n}{N-1}} \right) \right];$$

- com \mathbf{s} desconhecido e população finita:

vide item 5.6

$$\left[\bar{X} - t_{n-1} \cdot \left(\frac{S}{\sqrt{n}} \right) \left(\sqrt{\frac{N-n}{N-1}} \right); \bar{X} + t_{n-1} \cdot \left(\frac{S}{\sqrt{n}} \right) \left(\sqrt{\frac{N-n}{N-1}} \right) \right];$$

para Proporções:

- população infinita:

vide item 5.4

$$\left[p_s - Z \cdot \sqrt{\frac{p_s \cdot (1-p_s)}{n}}; p_s + Z \cdot \sqrt{\frac{p_s \cdot (1-p_s)}{n}} \right];$$

- população finita:

vide item 5.6

$$\left[p_s - Z \cdot \sqrt{\frac{p_s \cdot (1 - p_s)}{n}} \cdot \left(\sqrt{\frac{N - n}{N - 1}} \right); p_s + Z \cdot \sqrt{\frac{p_s \cdot (1 - p_s)}{n}} \cdot \left(\sqrt{\frac{N - n}{N - 1}} \right) \right];$$

do Total (com **s** desconhecido e população finita):

vide item 5.7

$$\left[N \cdot \bar{X} - N \cdot t_{n-1} \cdot \left(\frac{S}{\sqrt{n}} \right) \cdot \left(\sqrt{\frac{N - n}{N - 1}} \right); N \cdot \bar{X} + N \cdot t_{n-1} \cdot \left(\frac{S}{\sqrt{n}} \right) \cdot \left(\sqrt{\frac{N - n}{N - 1}} \right) \right];$$

- para Diferenças (com **s** desconhecido e população finita):

vide item 5.8

$$\left[N \cdot \bar{D} - N \cdot t_{n-1} \cdot \left(\frac{S_D}{\sqrt{n}} \right) \cdot \left(\sqrt{\frac{N - n}{N - 1}} \right); N \cdot \bar{D} + N \cdot t_{n-1} \cdot \left(\frac{S_D}{\sqrt{n}} \right) \cdot \left(\sqrt{\frac{N - n}{N - 1}} \right) \right], \quad \text{com} \quad \bar{D} = \frac{\sum_{i=1}^n D_i}{n} \quad \text{e}$$

$$S_D = \sqrt{\frac{\sum_{i=1}^n D_i^2 - n \cdot \bar{D}^2}{n - 1}};$$

III. Determinação do Tamanho da Amostra:

com População Infinita:

- para médias:

vide item 5.5.1

$$n = \frac{Z^2 \cdot S_o^2}{e^2};$$

- para duas proporções:

vide item 5.5.2

$$n = \frac{Z^2 \cdot p_o \cdot (1 - p_o)}{e^2};$$

- para *n*-proporções:

vide item 5.5.2

$$n = \frac{Z^2 \cdot 0,25}{e^2}, \quad \text{com } Z \text{ dado por } \left(\frac{\mathbf{y}}{2} \right) = \frac{\left[1 - (1 - \mathbf{a})^{\left(\frac{1}{k-1} \right)} \right]}{2};$$

com População Finita:

vide item 5.6

$$n_{CPF} = \frac{n \cdot N}{n + (N - 1)}.$$

AMOSTRAGEM ALEATÓRIA ESTRATIFICADA

I. Utilidade:

A população contém estratos que têm relação com o objeto da pesquisa \Rightarrow p. ex., no caso de uma pesquisa que pretenda estimar a remuneração média, cinco anos após a formatura, dos graduandos da Universidade de Brasília, é provável que a estratificação por curso seja relevante, uma vez que alguns são mais demandados ou agregam mais valor econômico do que outros; a presente modalidade de amostragem evitaria o risco de que alguns cursos fossem sub ou super-representados, o que geraria estimativas pontuais distantes do parâmetro, ainda que o intervalo de confiança continuasse válido.

II. Estimativa do Intervalo:

vide item 6.1.1

da Média (com t_{n-1} aproximado por Z):

$$\left[\bar{X}_{AE} \pm Z \cdot S_{\bar{X}_{AE}} \right], \text{ com } \bar{X}_{AE} = \frac{1}{N} \cdot \sum_{j=1}^k N_j \cdot \bar{X}_j, \quad S_{\bar{X}_{AE}}^2 = \frac{1}{N^2} \cdot \sum_{j=1}^k N_j^2 \cdot S_{\bar{X}_j}^2 \quad \text{e} \quad S_{\bar{X}_j}^2 = \frac{S_j^2}{n_j} \cdot \frac{N_j - n_j}{N_j};$$

para Proporções:

$$\left[\hat{p}_{AE} \pm Z \cdot S_{\hat{p}_{AE}} \right], \quad \text{com} \quad \hat{p}_{AE} = \frac{1}{N} \cdot \sum_{j=1}^k N_j \cdot \hat{p}_j, \quad S_{\hat{p}_{AE}}^2 = \frac{1}{N^2} \cdot \sum_{j=1}^k N_j^2 \cdot S_{\hat{p}_j}^2 \quad \text{e}$$

$$S_{\hat{p}_j}^2 = \frac{\hat{p}_j \cdot (1 - \hat{p}_j)}{n_j - 1} \cdot \frac{(N_j - n_j)}{N_j};$$

do Total (com t_{n-1} aproximado por Z):

$$\left[N \cdot \bar{X}_{AE} \pm Z \cdot N \cdot S_{\bar{X}_{AE}} \right], \text{ com } N \cdot \bar{X}_{AE} = \sum_{j=1}^k N_j \cdot \bar{X}_j \quad \text{e} \quad N^2 \cdot S_{\bar{X}_{AE}}^2 = \sum_{j=1}^k N_j^2 \cdot S_{\bar{X}_j}^2;$$

III. Alocação Ótima por Estrato do Tamanho da Amostra:

vide item 6.1.2

Para Médias (com σ_j conhecido):

$$n_j = \frac{N_j \cdot \sigma_j}{\sum_{j=1}^k N_j \cdot \sigma_j} \cdot n;$$

Para Duas Proporções (com p_j conhecido):

$$n_j = \frac{N_j \cdot \sqrt{p_j \cdot (1 - p_j)}}{\sum_{j=1}^k N_j \cdot \sqrt{p_j \cdot (1 - p_j)}} \cdot n.$$

AMOSTRAGEM ALEATÓRIA POR CONGLOMERADOS – UM ÚNICO ESTÁGIO

I. Utilidade:

A população é formada por subconjuntos semelhantes e os custos do trabalho de campo são elevados \Rightarrow p. ex., no caso de uma pesquisa que pretenda estimar a produtividade média dos assentamentos criados pelo Ministério do Desenvolvimento Agrário, é possível que os assentamentos possam ser agrupados em conjuntos geograficamente compactos, mas suficientemente diversificados internamente para que os conjuntos examinados possam ser tratados, a um custo menor, como representativos de toda população.

II. Estimativa do Intervalo:

vide item 6.2.1

da Média (com t_{n-1} aproximado por Z):

$$\left[\bar{X}_{AC} - Z.S_{\bar{X}_{AC}} < \mu < \bar{X}_{AC} + Z.S_{\bar{X}_{AC}} \right], \text{ com } \bar{X}_{AC} = \frac{\sum_{j=1}^m N_j \cdot \bar{X}_j}{\sum_{j=1}^m N_j}, \quad S_{\bar{X}_{AC}}^2 = \frac{M - m}{M \cdot \frac{\left(\sum_{j=1}^m N_j \right)^2}{m}} \cdot S_{AC}^2 \quad e$$

$$S_{AC}^2 = \frac{\sum_{j=1}^m N_j^2 \cdot (\bar{X}_j - \bar{X}_{AC})^2}{m - 1};$$

para Proporções:

$$\left[\hat{p}_{AC} - Z.S_{\hat{p}_{AC}} < p < \hat{p}_{AC} + Z.S_{\hat{p}_{AC}} \right], \text{ com } \hat{p}_{AC} = \frac{\sum_{j=1}^m N_j \cdot \hat{p}_j}{\sum_{j=1}^m N_j}, \quad S_{\hat{p}_{AC}}^2 = \frac{M - m}{M \cdot \frac{\left(\sum_{j=1}^m N_j \right)^2}{m}} \cdot S_{AC}^2 \quad e$$

$$S_{AC}^2 = \frac{\sum_{j=1}^m N_j^2 \cdot \left(\hat{p}_j - \hat{p}_{AC} \right)^2}{m - 1};$$

III. Determinação da Quantidade de Conglomerados da Amostra:

vide item 6.2.2

$$m = \frac{M \cdot S_{AC_0}^2}{\frac{e^2}{Z^2} \cdot \frac{\left(\sum_{j=1}^M N_j \right)^2}{M} - S_{AC}^2}.$$

ANEXO IV
FORMULÁRIOS PARA AMOSTRAGENS
POR UNIDADE MONETÁRIA

 <p>Tribunal de Contas da União Secretaria-Geral de Controle Externo</p>	<p>AMOSTRAGEM POR UNIDADE MONETÁRIA FORMULÁRIO 1</p>
--	--

Objeto(s) da Auditoria:	UG:
	TC:

Objeto da Amostragem:	Período de:
	<ul style="list-style-type: none"> • Abrangência: __/__/____ - __/__/____. • Realização: __/__/____ - __/__/____.

Plano Amostral		
Preparado por:	Matrícula:	Data: __/__/____.
Revisto por:	Matrícula:	Data: __/__/____.

PRECISÃO AMOSTRAL E AJUSTES ESTIMADOS

Erro Tolerável/Materialidade Ampla:	<input style="width:90%;" type="text"/>
Materialidade Restrita:	<input style="width:90%;" type="text"/>
Ajustes Arbitrados:	
Erro Amostral Previsto:	<input style="width:40%;" type="text"/>
Estimativa da Ampliação do Intervalo de Precisão:	<input style="width:40%;" type="text"/>
Outros Ajustes:	<input style="width:40%;" type="text"/>
Ajuste Total:	<input style="width:40%;" type="text"/>
Erro Máximo Admitido:	<input style="width:90%;" type="text"/>

Notas (premissas, precedentes ou ressalvas pertinentes):



Tribunal de Contas da União
Secretaria-Geral de Controle Externo

**AMOSTRAGEM POR UNIDADE MONETÁRIA
FORMULÁRIO 3**

Objeto(s) da Auditoria:	UG:
	TC:

Objeto da Amostragem:	Período de:
	<ul style="list-style-type: none"> Abrangência: __/__/____ - __/__/____. Realização: __/__/____ - __/__/____.

Plano Amostral		
Preparado por:	Matrícula:	Data: __/__/____.
Revisto por:	Matrícula:	Data: __/__/____.

ESTIMATIVA DO ERRO AMOSTRAL MÁXIMO

	Item	Valor		Erro			
		Registrado (A)	Auditado (B)	Valor (C) = [(A)- (B)]	Proporção (D)=[(C)/(A)]	Ordem (1º, 2º, etc.)	Projetado [(Intervalo Amostral).(D)]
Registros superestimados referentes a itens com valores inferiores ao do intervalo amostral							
Registros subestimados referentes a itens com valores inferiores ao do intervalo amostral							

Erro Líquido Observado na Amostra	<input type="text"/>
Erro Líquido Extrapolado para a População	<input type="text"/>

 <p>Tribunal de Contas da União Secretaria-Geral de Controle Externo</p>	<p>AMOSTRAGEM POR UNIDADE MONETÁRIA FORMULÁRIO 4</p>
--	--

Objeto(s) da Auditoria:	UG:
	TC:

Objeto da Amostragem:	Período de:
	<ul style="list-style-type: none"> • Abrangência: __/__/____ - __/__/____. • Realização: __/__/____ - __/__/____.

Plano Amostral		
Preparado por:	Matrícula:	Data: __/__/____.
Revisto por:	Matrícula:	Data: __/__/____.

ITENS CRÍTICOS OU COM VALORES ELEVADOS

	Item	Valor		Erro (C) = [(A)- (B)]
		Registrado (A)	Auditado (B)	
<p>Itens críticos ou com valores superiores ao intervalo amostral</p> <p>A) Superestimados <input type="checkbox"/></p> <p>B) Subestimados <input type="checkbox"/></p> <p><i>(utilizar formulários distintos para valores superestimados e subestimados)</i></p>				
<p>Erro Total</p>				

 <p>Tribunal de Contas da União Secretaria-Geral de Controle Externo</p>	<p>AMOSTRAGEM POR UNIDADE MONETÁRIA FORMULÁRIO 5</p>
--	--

Objeto(s) da Auditoria:	UG: TC:
-------------------------	------------

Objeto da Amostragem:	Período de: <ul style="list-style-type: none"> • Abrangência: __/__/____ - __/__/____. • Realização: __/__/____ - __/__/____.
-----------------------	---

Plano Amostral		
Preparado por:	Matrícula:	Data: __/__/____.
Revisto por:	Matrícula:	Data: __/__/____.

AMPLIAÇÃO DO INTERVALO DE PRECISÃO – AIP

	Ordem (1º, 2º, etc.)	Fator AIP (A)	Proporção (B)	AIP [(Intervalo Amostral.(A).(B))]
<p>Registros</p> <p>C) Superestimados <input type="checkbox"/></p> <p>D) Subestimados <input type="checkbox"/></p> <p style="text-align: center;"><i>(utilizar formulários distintos para valores superestimados e subestimados)</i></p>				
AIP Total				

 <p>Tribunal de Contas da União Secretaria-Geral de Controle Externo</p>	<p>AMOSTRAGEM POR UNIDADE MONETÁRIA FORMULÁRIO 6</p>
--	--

Objeto(s) da Auditoria:	UG:
	TC:

Objeto da Amostragem:	Período de:
	<ul style="list-style-type: none"> • Abrangência: __/__/____ - __/__/____. • Realização: __/__/____ - __/__/____.

Plano Amostral		
Preparado por:	Matrícula:	Data: __/__/____.
Revisto por:	Matrícula:	Data: __/__/____.

ERRO E PRECISÃO TOTAIS

	Superestimados	Subestimados
Erro Provável <i>(Formulários 3 & 4)</i>		
Materialidade Restrita <i>(Formulário 2)</i>		
Ampliação do Intervalo de Precisão <i>(Formulário 5)</i>		
Limite Superior dos Erros (1)		
Materialidade Ampla (2)		
Diferença [(2)-(1)]		
Diferença Aceitável?	<input type="checkbox"/> Sim <input type="checkbox"/> Não	<input type="checkbox"/> Sim <input type="checkbox"/> Não

ANEXO V
APLICAÇÃO DA AMOSTRAGEM
POR UNIDADE MONETÁRIA



Tribunal de Contas da União
Secretaria-Geral de Controle Externo

AMOSTRAGEM POR UNIDADE MONETÁRIA
FORMULÁRIO 1

Objeto(s) da Auditoria: XYZ	UG: ????? TC: ???./???.1999-?
-----------------------------	----------------------------------

Objeto da Amostragem: CADASTRO DE COMPRAS	Período de: <ul style="list-style-type: none"> Abrangência: 01/01/1997 - 31/12/1998. Realização: 25/01/1999 - 01/03/1999.
---	---

Plano Amostral		
Preparado por: UVW	Matrícula: ???-?	Data: 15/12/1998.
Revisto por: RST	Matrícula: ???-?	Data: 20/01/1999.

PRECISÃO AMOSTRAL E AJUSTES ESTIMADOS

Erro Tolerável/Materialidade Ampla:	100.000,00
Materialidade Restrita:	95.000,00
Ajustes Arbitrados:	
Erro Amostral Previsto:	—
Estimativa da Ampliação do Intervalo de Precisão:	—
Outros Ajustes:	5.000,00
Ajuste Total:	5.000,00
Erro Máximo Admitido:	100.000,00

Notas (premissas, precedentes ou ressalvas pertinentes):

A) *Materialidade Ampla*: maior diferença entre os montantes registrado e estimado admitido pela equipe de auditoria ou pelo órgão encarregado da auditoria;

B) *Materialidade Restrita*: diferença entre a *materialidade ampla* e os *ajustes arbitrados* pela equipe de auditoria;

C) *Ajustes Arbitrados*: maior diferença entre os montantes registrado e estimado indicada pelos resultados amostrais; seus componentes (*erro amostral previsto*, *estimativa da ampliação do intervalo de precisão* e *outros ajustes*) são arbitrados pela equipe de auditoria com base em trabalhos anteriores ou de proporções fixadas pelo órgão encarregada da auditoria (no presente exemplo: 0,25% do montante registrado).

 <p>Tribunal de Contas da União Secretaria-Geral de Controle Externo</p>	<p>AMOSTRAGEM POR UNIDADE MONETÁRIA FORMULÁRIO 2</p>
--	--

Objeto(s) da Auditoria: XYZ	UG: ?????
	TC: ???./???.1999-?

Objeto da Amostragem: CADASTRO DE COMPRAS	Período de:
	<ul style="list-style-type: none"> • Abrangência: 01/01/1997 - 31/12/1998. • Realização: 25/01/1999 - 01/03/1999.

Plano Amostral		
Preparado por: UVW	Matrícula: ???-?	Data: 15/12/1998.
Revisto por: RST	Matrícula: ???-?	Data: 20/01/1999.

DETERMINAÇÃO DO TAMANHO DA AMOSTRA

População (compras, salários, estoques, etc.)	COMPRAS
(1) Materialidade Restrita	95.000,00
(2) Montante Registrado	2.000.000,00
(3) Índice de Confiabilidade	3,00
(4) Tamanho da Amostra [(2).(3)/(1)]	64
(5) Intervalo Amostral [(2)/(4)]	31.250,00

Notas (inclusões e expurgos pertinentes):

O montante registrado abrange 500 operações de compra. Os valores de cinco delas são superiores ao intervalo amostral.



Tribunal de Contas da União
Secretaria-Geral de Controle Externo

AMOSTRAGEM POR UNIDADE MONETÁRIA
FORMULÁRIO 3

Objeto(s) da Auditoria: XYZ	UG: ?????
	TC: ???./???/1999-?

Objeto da Amostragem: CADASTRO DE COMPRAS	Período de:
	<ul style="list-style-type: none"> Abrangência: 01/01/1997 - 31/12/1998. Realização: 25/01/1999 - 01/03/1999.

Plano Amostral		
Preparado por: UVW	Matrícula: ???-?	Data: 15/12/1998.
Revisto por: RST	Matrícula: ???-?	Data: 20/01/1999.

ESTIMATIVA DO ERRO AMOSTRAL MÁXIMO

	Item	Valor		Erro			
		Registrado (A)	Auditado (B)	Valor (C) = [(A)- (B)]	Proporção (D)=[(C)/(A)]	Ordem (1º, 2º, etc.)	Extrapolação [(Intervalo Amostral),(D)]
Registros superestimados referentes a itens com valores inferiores ao do intervalo amostral	35	2.000,00	1.800,00	200,00	0,10	1º	3.125,00
Registros subestimados referentes a itens com valores inferiores ao do intervalo amostral	29	2.500,00	2.600,00	(100,00)	0,04	2º	(1.250,00)
	57	4.000,00	4.200,00	(200,00)	0,05	1º	(1.562,50)
Erro Líquido Observado na Amostra				(100,00)			
Erro Líquido Extrapolado para a População							312,50

 <p>Tribunal de Contas da União Secretaria-Geral de Controle Externo</p>	<p>AMOSTRAGEM POR UNIDADE MONETÁRIA FORMULÁRIO 4</p>
--	--

Objeto(s) da Auditoria: XYZ	UG: ????? TC: ???./???./1999-?
-----------------------------	-----------------------------------

Objeto da Amostragem: CADASTRO DE COMPRAS	Período de: <ul style="list-style-type: none"> • Abrangência: 01/01/1997 - 31/12/1998. • Realização: 25/01/1999 - 01/03/1999.
---	---

Plano Amostral		
Preparado por: UVW	Matrícula: ???-?	Data: 15/12/1998.
Revisto por: RST	Matrícula: ???-?	Data: 20/01/1999.

ITENS CRÍTICOS OU COM VALORES ELEVADOS

	Item	Valor		Erro (C) = [(A)- (B)]
		Registrado (A)	Auditado (B)	
<p>Itens críticos ou com valores superiores ao intervalo amostral</p> <p>E) Superestimados <input checked="" type="checkbox"/></p> <p>F) Subestimados <input type="checkbox"/></p> <p><i>(utilizar formulários distintos para valores superestimados e subestimados)</i></p>	13	43.125,00	42.625,00	500,00
Erro Total				500,00

 <p>Tribunal de Contas da União Secretaria-Geral de Controle Externo</p>	<p>AMOSTRAGEM POR UNIDADE MONETÁRIA FORMULÁRIO 5</p>
--	--

Objeto(s) da Auditoria: <i>XYZ</i>	UG: ?????
	TC: ???./???./1999-?

Objeto da Amostragem: CADASTRO DE COMPRAS	Período de:
	<ul style="list-style-type: none"> • Abrangência: 01/01/1997 - 31/12/1998. • Realização: 25/01/1999 - 01/03/1999.

Plano Amostral		
Preparado por: <i>UVW</i>	Matrícula: ???-?	Data: 15/12/1998.
Revisto por: <i>RST</i>	Matrícula: ???-?	Data: 20/01/1999.

AMPLIAÇÃO DO INTERVALO DE PRECISÃO – AIP

	Ordem (1º, 2º, etc.)	Fator AIP (A)	Proporção (B)	AIP [(Intervalo Amostral.(A).(B))]
Registros G) Superestimados <input checked="" type="checkbox"/> H) Subestimados <input type="checkbox"/> <i>(utilizar formulários distintos para valores superestimados e subestimados)</i>	1º	0,75	0,10	2.343,75
AIP Total				2.343,75

 <p>Tribunal de Contas da União Secretaria-Geral de Controle Externo</p>	<p>AMOSTRAGEM POR UNIDADE MONETÁRIA FORMULÁRIO 5</p>
--	--

Objeto(s) da Auditoria: XYZ	UG: ?????
	TC: ???./???./1999-?

Objeto da Amostragem: CADASTRO DE COMPRAS	Período de:
	<ul style="list-style-type: none"> • Abrangência: 01/01/1997 - 31/12/1998. • Realização: 25/01/1999 - 01/03/1999.

Plano Amostral		
Preparado por: UVW	Matrícula: ???-?	Data: 15/12/1998.
Revisto por: RST	Matrícula: ???-?	Data: 20/01/1999.

AMPLIAÇÃO DO INTERVALO DE PRECISÃO – AIP

	Ordem (1°, 2°, etc.)	Fator AIP (A)	Proporção (B)	AIP [(Intervalo Amostral.(A).(B))]
<p>Registros</p> <p>I) Superestimados <input type="checkbox"/></p> <p>J) Subestimados <input checked="" type="checkbox"/></p> <p style="text-align: center;"><i>(utilizar formulários distintos para valores superestimados e subestimados)</i></p>	1°	0,75	0,05	(1.171,87)
	2°	0,55	0,04	(687,50)
AIP Total				(1.859,37)



Tribunal de Contas da União
Secretaria-Geral de Controle Externo

AMOSTRAGEM POR UNIDADE MONETÁRIA
FORMULÁRIO 6

Objeto(s) da Auditoria: XYZ	UG: ????? TC: ???./???.-?
-----------------------------	------------------------------

Objeto da Amostragem: CADASTRO DE COMPRAS	Período de: <ul style="list-style-type: none"> Abrangência: 01/01/1997 - 31/12/1998. Realização: 25/01/1999 - 01/03/1999.
---	---

Plano Amostral		
Preparado por: UVW	Matrícula: ???-?	Data: 15/12/1998.
Revisto por: RST	Matrícula: ???-?	Data: 20/01/1999.

ERRO E PRECISÃO TOTAIS

	Superestimados	Subestimados
Erro Provável (Formulários 3 & 4)	312,50	312,50
	500,00	
Materialidade Restrita (Formulário 2)	95.000,00	(95.000,00)
Ampliação do Intervalo de Precisão (Formulário 5)	2.343,75	(1.859,37)
Limite Superior dos Erros (1)	98.156,25	(96.546,87)
Materialidade Ampla (2)	100.000,00	(100.000,00)
Diferença [(2)-(1)]	1.843,75	(3.453,13)
Diferença Aceitável?	<input checked="" type="checkbox"/> Sim <input type="checkbox"/> Não	<input checked="" type="checkbox"/> Sim <input type="checkbox"/> Não

GLOSSÁRIO

AMOSTRA: subconjunto de elementos da população selecionado para análise; a quantidade de elementos da amostra (i. e., tamanho da amostra) é usualmente denotada como n , enquanto que a quantidade de elementos da população é geralmente denotada como N .

AMOSTRA ALEATÓRIA: amostra cujos elementos são selecionados aleatoriamente, podendo-se computar a probabilidade de que uma amostra em particular seja obtida; a seleção pode ser com ou sem reposição, ou seja, os elementos podem ser selecionados simultaneamente ou um por vez, devolvendo-se à população, após cada seleção, o elemento selecionado; no caso da seleção com reposição, elementos da população podem integrar a amostra mais de uma vez.

APROXIMAÇÃO PELA NORMAL: mensuração aproximada das áreas sob um histograma (i. e., plotagem da distribuição de freqüências relativas de uma variável aleatória) por meio das áreas correspondentes definidas pela curva normal, após a transformação dos valores originais em unidades-padrão; diversas distribuições de probabilidades podem ser aproximadas por intermédio da distribuição normal (p. ex.: no intervalo $[9,5; 17,5]$, a área definida pelo histograma de uma distribuição binomial com $n = 50$ e $p = 0,3$ é igual a 0,742; no caso da distribuição normal, esse intervalo corresponde a $\left[\frac{9,5 - 50 \cdot 0,3}{\sqrt{50 \cdot 0,3 \cdot (1 - 0,3)}}; \frac{17,5 - 50 \cdot 0,3}{\sqrt{50 \cdot 0,3 \cdot (1 - 0,3)}} \right] = [-1,697; 0,772]$, ao qual está associado uma área igual a 0,735 – apenas um pouco inferior ao valor correto, portanto; convém notar que a aproximação será tão melhor quanto maior for n e mais próximo de 0,5 for p).

CORREÇÃO PARA POPULAÇÕES FINITAS: no caso de amostragens sem reposição, os erros-padrão dos totais e das médias amostrais dependem da fração da população incluída na amostra – maior a fração, menor o erro-padrão; o erro-padrão nas amostragens sem reposição é inferior ao erro-padrão nas amostragens com reposição, com a diferença correspondendo ao fator de correção para populações finitas (*CPF*), qual seja, $\left[\sqrt{\frac{N - n}{N - 1}} \right]$ – com $n = 1$, as duas modalidades de amostragem são equivalentes, o que torna os erros-padrão iguais ($CPF = 1$), enquanto que, com $n = N$, a amostra selecionada sem reposição confunde-se com a própria população, de modo que as estimativas são iguais aos parâmetros, não havendo qualquer variação, ou seja, o erro-padrão é igual a zero ($CPF = 0$).

DADOS: informações numéricas relevantes para a tomada de decisão em um contexto específico.

DEFINIÇÃO OPERACIONAL: atribuição de um significado a um conceito ou variável que possa ser entendido por todos.

DISTRIBUIÇÃO: modo como se distribuem as freqüências relativas dos valores associados a um conjunto de observações.

DISTRIBUIÇÃO ACUMULADA DE PROBABILIDADES: modo como se acumulam, para cada possível resultado x_i da variável aleatória X , as observações cujos valores sejam iguais ou inferiores [$P(X \leq x_i)$]; a distribuição é igual a zero para valores suficientemente baixos e igual a um para valores suficientemente altos; o gráfico correspondente cresce monotonicamente – se $x_1 < x_2$, então $P(X \leq x_1) \leq P(X \leq x_2)$.

DISTRIBUIÇÃO AMOSTRAL: distribuição das estimativas geradas por amostras aleatórias.

DISTRIBUIÇÃO DE PROBABILIDADES: especificação da probabilidade de que uma variável aleatória X assumirá cada um dos seus possíveis resultados x_i , no caso de variáveis discretas [$P(X = x_i)$], ou de que esteja compreendida em qualquer intervalo [$x_1; x_2$] do segmento de possíveis resultados, no caso de variáveis contínuas [$P(x_1 < X < x_2)$].

DISTRIBUIÇÃO DE PROBABILIDADES UNIFORME: todos os possíveis resultados da variável aleatória têm a mesma probabilidade de ocorrer.

DISTRIBUIÇÃO NORMAL: uma variável aleatória X , com média m e desvio-padrão s , é normalmente distribuída [$X \sim N(m, s^2)$] se $P[x_1 < (X - \mu)/\sigma < x_2]$, para $x_1 < x_2$, corresponder à área sob a curva normal compreendida no intervalo [$x_1; x_2$]; caso m e s sejam iguais a zero e a um, respectivamente, tem-se que X apresenta uma distribuição normal padronizada, ou seja, $X \sim N(0, 1)$.

ELEMENTO: componente de um conjunto de dados.

ERRO ALEATÓRIO: erro que, ao contrário do viés, afeta as estimativas de forma diferente a cada medição; comporta-se como um número selecionado, aleatoriamente e com reposição, de um conjunto de valores cuja média é igual a zero.

ESPAÇO OU CADASTRO AMOSTRAL: conjunto de dados a partir do qual a amostra é selecionada; idealmente, o cadastro contém toda a população acerca da qual se deseja fazer inferências.

ESPAÇO DOS POSSÍVEIS RESULTADOS: conjunto de todos os possíveis resultados de um experimento aleatório (i.e., experimento cujo resultado, em uma única tentativa, não pode ser previsto com exatidão, mas cujas frequências relativas de longo prazo, para sucessivas repetições, são previsíveis).

ESTATÍSTICA: medida estimada a partir de uma amostra que exprime uma característica da população.

ESTATÍSTICA DESCRITIVA: métodos de coleta, apresentação e caracterização de um conjunto de dados que permitam descrever apropriadamente vários dos seus aspectos.

ESTIMADOR: regra cujo objetivo seja gerar, a partir de amostras aleatórias, estimativas acerca dos parâmetros (p. ex., a média amostral \bar{x} é um estimador da média populacional m); é uma variável aleatória, pois seus resultados dependem das amostras que forem selecionadas; idealmente, não deve ser viesado e a sua variância deve ser pequena, podendo-se avaliá-lo por meio do erro-médio quadrático (EMQ)²⁷, o qual é igual ao valor esperado do quadrado da diferença entre o estimador (p. ex., \bar{x}) e o parâmetro (p. ex., m), ou seja,

²⁷ “Mean Square Error” (MSE), na língua inglesa.

$EMQ(\bar{x}) = \sum_{j=1}^k \left[(\bar{x}_j - \mu)^2 / k \right]$, com k igual à quantidade de estimativas (i.e., o EMQ é uma medida da acurácia do estimador e também pode ser expresso da seguinte maneira: $EMQ(\bar{x}) = [\text{viés}(\bar{x})]^2 + [\text{erro-padrão}(\bar{x})]^2$).

EVENTO: subconjunto do espaço de possíveis resultados que contenha resultados efetivamente observados (p. ex., se A é um subconjunto do espaço de possíveis resultados de uma variável aleatória, então A é um evento caso o resultado efetivamente observado pertença a A).

FUNÇÃO DE DISTRIBUIÇÃO DE PROBABILIDADES: representação matemática da distribuição de uma variável aleatória discreta.

FUNÇÃO DENSIDADE DE PROBABILIDADE: função matemática cuja área sob a curva correspondente define as probabilidades de que os resultados de uma variável aleatória contínua estejam compreendidos em intervalos do segmento de possíveis resultados – para toda variável aleatória X , há uma função densidade de probabilidade $f(X)$ tal que $P(x_1 < X < x_2)$, para $x_1 < x_2$, corresponda à área compreendida no intervalo $[x_1; x_2]$.

INFERÊNCIA ESTATÍSTICA: métodos que permitem estimar características de uma população ou tomar decisões com base em resultados amostrais.

INTERVALO DE CONFIANÇA: intervalo aleatório cuja probabilidade de que contenha o verdadeiro valor do parâmetro é conhecida.

LEI DOS GRANDES NÚMEROS: no caso de repetições sucessivas e independentes de experimentos aleatórios, com igual probabilidade p de sucesso, a probabilidade amostral \hat{p} será tão mais próxima de p quanto maior for a quantidade de repetições; a probabilidade de que a diferença entre \hat{p} e p seja maior do que um valor positivo e converge para zero à medida que a quantidade de repetições tende ao infinito.

MODELO: representação estilizada de um fenômeno.

MODELO MATEMÁTICO: representação matemática de um fenômeno.

NÃO-RESPOSTA: diferença entre a quantidade de consultas feitas (p. ex., questionários enviados, telefonemas feitos, etc.) e a quantidade de consultas respondidas (p. ex., questionários devolvidos, telefonemas atendidos, etc.).

NÍVEL DE CONFIANÇA: probabilidade de que o intervalo de confiança contenha o verdadeiro valor do parâmetro.

PARÂMETRO: propriedade numérica de uma população, como a sua média.

POPULAÇÃO OU UNIVERSO: conjunto de elementos visado por uma pesquisa, os quais podem ser épocas, lugares, objetos, pessoas, procedimentos, etc.

TAXA DE NÃO-RESPOSTA: razão entre as quantidades de consultas não-respondidas e de consultas feitas; caso essa taxa seja alta, a pesquisa conterà um viés de não-resposta elevado.

TIPOS DE AMOSTRAGEM: as técnicas de amostragem podem ser classificadas de várias maneiras, quais sejam:

- a) quanto à divisão em estratos: estratificada ou não-estratificada;
- b) quanto à quantidade de estágios: um estágio ou mais de um;
- c) quanto à unidade amostral: um elemento ou um conglomerado;
- d) quanto ao processo de seleção das unidades amostrais: aleatória ou sistemática;
- e) quanto às probabilidades de seleção das unidades amostrais: probabilidades iguais ou desiguais.

UNIDADE AMOSTRAL: unidade fundamental de uma amostra.

VALOR ESPERADO OU ESPERANÇA MATEMÁTICA: limite de longo prazo da média dos resultados observados em repetições sucessivas e independentes de experimentos aleatórios; no caso de variáveis discretas, esse limite corresponde à média dos possíveis resultados, ponderados pela respectiva frequência relativa; o valor esperado de uma variável aleatória X é denotado como $E(X)$.

VARIÁVEL: valor ou categoria que possa diferir de observação para observação.

VARIÁVEL ALEATÓRIA: atribuição de valores aos possíveis resultados de um experimento aleatório (p. ex., no arremesso de três moedas, a quantidade de “caras” que podem ser observadas é uma variável aleatória, atribuindo-se o valor “0” ao resultado {coroa, coroa, coroa}, o valor “1” aos resultados {coroa, coroa, cara}, {coroa, cara, coroa} e {cara, coroa, coroa}, o valor “2” aos resultados {coroa, cara, cara}, {cara, coroa, cara} e {cara, cara, coroa} e o valor três ao resultado {cara, cara, cara}).

VARIÁVEL ALEATÓRIA CONTÍNUA: variável numérica cujo conjunto de possíveis resultados é incontável (p. ex., altura exata, idade exata, incluindo frações de segundo, temperatura, etc.).

VARIÁVEL ALEATÓRIA DISCRETA: variável numérica cujo conjunto de possíveis resultados é contável (p. ex., variáveis cujos possíveis resultados são um subconjunto dos números inteiros, tais como: idades arredondadas para o valor inteiro mais próximo, números das cédulas de identidade, tamanhos das famílias, etc.).

VARIÁVEL CATEGÓRICA OU ATRIBUTO: variável cujos resultados correspondam a categorias, tais como: {alto, baixo}, {feminino, masculino}, {vermelho, verde, azul}, {Amazonas, Ceará, Goiás e São Paulo}, etc.

VIÉS: diferença sistemática, deliberada ou não, entre as estimativas e o parâmetro (p. ex.: uma pessoa que sempre se pese estando vestida obterá medidas sistematicamente superiores ao seu peso correto, introduzindo um viés “para cima” nas estimativas).

VIÉS DA NÃO-RESPOSTA: viés decorrente de diferenças existentes entre os respondentes e os não-respondentes em aspectos relevantes para a pesquisa (p. ex., uma pesquisa por telefone que deseje estimar a duração média da jornada de trabalho poderia, inadvertidamente, ignorar as pessoas que trabalham até mais tarde, o que introduziria um viés “para baixo” nas estimativas).

BIBLIOGRAFIA

- **Principal**

- ADAMS, Roger; REGAL, Richard. **TCU Auditing Seminar Programme**. Brasília, 1995. (Mimeo.)
- BERNSTEIN, Peter. **Against the Gods: the remarkable story of risk**. Nova Iorque : John Wiley & Sons, 1996.
- BOLFARINE, Heleno; BUSSAB, Wilton O. **Elementos de Amostragem**. Belo Horizonte, 1994. (11º Simpósio Nacional de Probabilidade e Estatística. Mimeo.)
- BORRIES, George F. **Estatística Aplicada e Noções de Amostragem**. Brasília, 1997. (Mimeo.)
- COCHRAN, William. **Sampling Techniques**. 3 ed. Nova Iorque : John Wiley & Sons, 1977. (Wiley Series in Probability and Mathematical Statistics).
- DAMASCENO, Paula de B.; PACHECO, Márcio E. “Técnica de Amostragem”. In: **Auditoria Pública**. Rio de Janeiro, 1998, p. 54-70. (Curso de Pós-Graduação em Contabilidade Pública da Fundação Getúlio Vargas - RJ. Mimeo.)
- FONSECA, Jairo S. da; MARTINS, Gilberto de A. **Curso de Estatística**. 3 ed. São Paulo : Atlas, 1982.
- LEVINE, David M.; BERENSON, Mark L.; STEPHAN, David. **Statistics for Managers Using Microsoft Excel**. New Jersey : Prentice Hall, 1997.
- MCGILL, John J. **Statistics for Managers Using Microsoft Excel: Microsoft PowerPoint slides**. New Jersey : Prentice Hall, 1997. (Meio Magnético.)
- NEWBOLD, Paul. **Statistics for Business and Economics**. 4 ed. New Jersey : Prentice Hall, 1995.
- PLATEK, R.; SINGH, M. P.; TREMBLAY, V. “Adjustment for Nonresponse in Surveys”. In: NAMBOODIRI, N. Krishnan (Org.). **Survey Sampling and Measurement**. Nova Iorque : Academic Press, 1978, p. 157-74.
- SCHEAFFER, Richard; MENDENHALL, William; OTT, Lyman. **Elementary Survey Sampling**. 4 ed. Belmont : Duxbury, 1990.
- SOUZA, Jorge de. **Pesquisa Eleitoral: críticas e técnicas**. Brasília, 1990.
- SPIEGEL, Murray R. **Estatística**. São Paulo : McGraw-Hill, 1976
- STEVENSON, William. **Estatística Aplicada à Administração**. São Paulo : Harbra, 1986.
- TOLEDO, Geraldo L.; OVALLE, Ivo I. **Estatística Básica**. 2 ed. São Paulo : Atlas, 1985.

- **Complementar**

ANDERSON, R. J.; TEITLEBAUM, A. Dollar Unit Sampling. **CA Magazine**, Abril/1973, p. 30-9.

ARKIN, Herbert. **Handbook of Sampling for Auditing and Accounting**. 2 ed. Nova Iorque : McGraw-Hill, 1974. (McGraw-Hill Accounting Series.)

COMMISSION on Audit. **Comprehensive Audit Manual on COA – Sampling-Card Methodology: concepts and application guidelines**. Manila, 1985.

EFRON, Bradley. **The Jackknife, the Bootstrap and Other Resampling Plans**. Philadelphia : Society for Industrial and Applied Mathematics, 1982. (Regional Conference Series in Applied Mathematics.)

GENERAL Accounting Office. **Using Statistical Sampling**. Washington, 1992.

GOVERNMENT Auditing Training Institute. **Statistical Sampling Using Computer Applications**. Washington, 1995.

HORVITZ, D. G. “Some Design Issues in Sample Surveys”. In: NAMBOODIRI, N. Krishnan (Org.). **Survey Sampling and Measurement**. Nova Iorque : Academic Press, 1978, p. 3-11.

LANPHIER, C. Michael; BAILAR, Barbara A. “A Survey of Surveys: some sampling frame problems”. In: NAMBOODIRI, N. Krishnan (Org.). **Survey Sampling and Measurement**. Nova Iorque : Academic Press, 1978, p. 67-85.

LOYNES, R. **A Critical Review of Monetary Sampling and the Stringer Bound**. Sheffield : University of Sheffield, 1994.

MURTHY, M. N. “Use of Sample Surveys in National Planning in Developing Countries”. In: NAMBOODIRI, N. Krishnan (Org.). **Survey Sampling and Measurement**. Nova Iorque : Academic Press, 1978, p. 231-53.

MURTHY, M. N.; RAO, T. J. “Systematic Sampling with Illustrative Examples”. In: KRISHNAIAH, P. R.; RAO, C. R. (Org.). **Handbook of Statistics, v. 6**. Amsterdã : Elsevier Science, 1988, p. 147-85.

NATIONAL Audit Office. **Collecting, Analysing and Presenting Data: how software can help**. Londres, 1996.

_____. **Financial Audit Planning**. Londres, s.d. (Mimeo.)

RAO, Poduri S. R. S. “Ratio and Regression Estimators”. In: KRISHNAIAH, P. R.; RAO, C. R. (Org.). **Handbook of Statistics, v. 6**. Amsterdã : Elsevier Science, 1988, p. 449-68.

SNELLING, J. **Use of Sampling: value for money studies**. Londres : National Audit Office, 1992.

SUDMAN, Seymour. “How Big Should a Sample Be?”. In: **Applied Sampling**. Nova Iorque : Academic Press, 1976, p. 85-105.

TOWS, J.; SNELLING, J. **Designing and Carrying out a Survey**. Londres : National Audit Office, 1991.

FOLHA DE SUGESTÕES

O TCU preocupa-se com o constante aperfeiçoamento da qualidade de seus manuais e orientações, buscando, para isso, ouvir a valiosa opinião do público-alvo dos referidos trabalhos.

O questionário a seguir refere-se especificamente às Técnicas de Amostragem para Auditorias, distribuídas a partir de março de 2002. Será muito útil para o TCU se o leitor deste documento puder dispor de alguns minutos para responder às perguntas constantes no referido questionário e enviá-lo pelos Correios (não é preciso selar, pois o porte será pago pelo TCU).

Sugestões sobre este documento também podem ser enviadas das seguintes formas:

E-mail: *adfis@tcu.gov.br*

Fax: (61) 316-7538

Fone: (61) 316-7311

Endereço: Tribunal de Contas da União – TCU

ADFIS

Setor de Administração Federal Sul – Lote 01

CEP: 70042-900 – Brasília-DF



FINALIDADE

Este questionário de avaliação tem por objetivo obter a opinião dos leitores sobre as **Técnicas de Amostragem para Auditorias**, com vistas ao seu aperfeiçoamento.

Por favor, responda às questões abaixo assinalando com um “X” a alternativa mais adequada. Desde já agradecemos a sua colaboração.

1. Em que esfera do governo você trabalha?

- Federal • Estadual ou DF • Municipal

2. Em que órgão você trabalha?

- Poder Legislativo • Poder Judiciário • Outro [especificar] _____
• Poder Executivo • Controle Interno

3. Que partes das Técnicas de Amostragem para Auditorias você leu?

- Todo • Capítulos I e II [todo ou parte] • Capítulos III e IV [todo ou parte] • Capítulos V e VI [todo ou parte] • Capítulos VII e VIII [todo ou parte]

4. Leia com atenção cada indicador e escolha o ponto da escala que melhor descreve a sua opinião sobre as Técnicas de Amostragem para Auditorias. Marque com um “X” a opção que melhor representa o seu julgamento.

Concorda integralmente 5	Concorda 4	Indiferente 3	Discorda 2	Discorda integralmente 1
-----------------------------	---------------	------------------	---------------	-----------------------------

O manual é:	5	4	3	2	1
Fácil de ser lido	•	•	•	•	•
Fácil de ser entendido	•	•	•	•	•
Lógico	•	•	•	•	•
Sucinto	•	•	•	•	•
Completo	•	•	•	•	•
Útil	•	•	•	•	•

5. Como você tomou conhecimento das Técnicas de Amostragem para Auditorias?

- Quando recebeu • Pela Internet
• Divulgação interna do TCU • Pela imprensa
• Por mensagem do SIAFI • Outros [especificar] _____

6. Como você obteve das Técnicas de Amostragem para Auditorias?

- Solicitou diretamente ao TCU • *Download* pela Internet • Outros [especificar] _____

7. Apresente, a seguir, comentários e sugestões para o aprimoramento da qualidade das Técnicas de Amostragem para Auditorias. No caso de sugestões para alteração/supressão/aditamento de itens de verificação, favor preencher o quadro anexo.



Tribunal de Contas da União
Secretaria-Adjunta de Fiscalização

QUADRO DE SUGESTÕES

FINALIDADE

Este quadro de sugestões tem por objetivo obter a opinião dos leitores sobre as Técnicas de Amostragem para Auditorias, com vistas ao seu aperfeiçoamento.

Nº do item	Proposta de alteração, supressão ou aditamento	Fundamentação

PTR/BSB
880/92
UP-AC/TCU
DR/BSB

CARTA - RESPOSTA
NÃO É NECESSÁRIO SELAR

O SELO SERÁ PAGO POR
TRIBUNAL DE CONTAS DA UNIÃO
BRASÍLIA-DF

70099-999

UNIDADES DA SECRETARIA DO TRIBUNAL DE CONTAS DA UNIÃO

Secretaria da Presidência	<i>Paulo Emílio Lustosa</i>
Consultoria Jurídica	<i>José Moacir Cardoso da Costa</i>
Instituto Serzedello Corrêa	<i>Salvatore Palumbo</i>
Secretaria de Controle Interno	<i>Leila Fonseca dos Santos V. Ferreira</i>
Secretaria de Planejamento e Gestão	<i>Mauro Giacobbo</i>
Secretaria de Tecnologia da Informação	<i>Cláudio Silva da Cruz</i>
Assessoria de Cerimonial e Relações Institucionais	<i>Erivan Carlos de Carvalho</i>
Assessoria de Comunicação Social	<i>Francisco Raul Félix de Souza Ramos</i>
Assessoria de Relações Internacionais	<i>Ana Beatriz Pascal Kraft</i>
Assessoria Parlamentar	<i>Severino Lucena da Nóbrega</i>
Secretaria-Geral das Sessões	<i>Eugênio Lisboa Vilar de Melo</i>
Secretaria do Plenário	<i>Elenir Teodoro Gonçalves dos Santos</i>
Secretaria-Geral de Administração	<i>Antônio José Ferreira da Trindade</i>
Secretaria de Recursos Humanos	<i>Cláudia de Faria Castro</i>
Secretaria de Orçamento, Finanças e Contabilidade	<i>Pedro Martins de Sousa</i>
Secretaria de Material, Pat. e Com. Administrativa	<i>Ary Fernando Beirão</i>
Secretaria de Engenharia e Serviços Gerais	<i>Alfredo Henrique Bauchspiess</i>
Secretaria-Geral de Controle Externo	<i>Luciano Carlos Batista</i>
Secretaria-Adjunta de Contas	<i>Ricardo de Mello Araújo</i>
Secretaria-Adjunta de Fiscalização	<i>Cláudio Souza Castello Branco</i>
Secretaria de Fiscalização de Desestatização	<i>Jorge Pereira de Macedo</i>
Secretaria de Fiscalização de Pessoal	<i>Antonio Júlio Ferreira</i>
Secretaria de Fiscalização de Obras e Pat. da União	<i>Cláudio Sarian Altounian</i>
Secretaria de Fiscalização e Aval. de Prog. de Governo	<i>Marília Zinn Salvucci</i>
Secretaria de Macroavaliação Governamental	<i>Paulo Roberto Pinheiro Dias Pereira</i>
Secretaria de Recursos	<i>Odilon Cavallari de Oliveira</i>
1ª Secretaria de Controle Externo	<i>Rosendo Severo dos Anjos Neto</i>
2ª Secretaria de Controle Externo	<i>Eduardo Duailibe Murici</i>
3ª Secretaria de Controle Externo	<i>Carlos Nivan Maia</i>
4ª Secretaria de Controle Externo	<i>Maria do P. Socorro Teixeira Rosa</i>
5ª Secretaria de Controle Externo	<i>Alexandre Valente Xavier</i>
6ª Secretaria de Controle Externo	<i>Ismar Barbosa Cruz</i>
Secretaria de Controle Externo/AC	<i>João Batista Diniz Capanema</i>
Secretaria de Controle Externo/AL	<i>Edimilson Monteiro Batista</i>
Secretaria de Controle Externo/AP	<i>Jorge Luiz Carvalho Lugão</i>
Secretaria de Controle Externo/AM	<i>Helena Montenegro Valente</i>
Secretaria de Controle Externo/BA	<i>Evilásio Magalhães Vieira</i>
Secretaria de Controle Externo/CE	<i>Paulo Nogueira de Medeiros</i>
Secretaria de Controle Externo/ES	<i>Raimundo Nonato Coutinho</i>
Secretaria de Controle Externo/GO	<i>Rosângela Paniago Curado Fleury</i>
Secretaria de Controle Externo/MA	<i>José Maria Araújo Lima</i>
Secretaria de Controle Externo/MT	<i>Luiz Guilherme da Boamorte Silveira</i>
Secretaria de Controle Externo/MS	<i>Mário Júnior Bertuol</i>
Secretaria de Controle Externo/MG	<i>Élsio Geová dos Santos</i>
Secretaria de Controle Externo/PA	<i>Octávio José Pessoa Ferreira</i>
Secretaria de Controle Externo/PB	<i>Raimundo Nonato Soares Araújo</i>
Secretaria de Controle Externo/PR	<i>Nazaré do Socorro G. Rosário Zuardi</i>
Secretaria de Controle Externo/PE	<i>Ildê Ramos Rodrigues</i>
Secretaria de Controle Externo/PI	<i>José Ulisses Rodrigues Vasconcelos</i>
Secretaria de Controle Externo/RJ	<i>Francisco Carlos Ribeiro de Almeida</i>
Secretaria de Controle Externo/RN	<i>Marcos Valério de Araújo</i>
Secretaria de Controle Externo/RS	<i>Carlos Martins dos Santos</i>
Secretaria de Controle Externo/RO	<i>Fábio Arruda de Lima</i>
Secretaria de Controle Externo/RR	<i>Rainério Rodrigues Leite</i>
Secretaria de Controle Externo/SC	<i>Rafael Blanco Muniz</i>
Secretaria de Controle Externo/SP	<i>Eloi Carnovali</i>
Secretaria de Controle Externo/SE	<i>Maria Salete Fraga Silva Palma</i>
Secretaria de Controle Externo/TO	<i>Dion Carvalho Gomes de Sá</i>