

# COMPARAÇÃO DE TAMANHOS AMOSTRAIS PARA A SELEÇÃO DE GUARDA DE AUTOS FINDOS ATRAVÉS DA SIMULAÇÃO DE AMOSTRAS UTILIZANDO AMOSTRAGEM ESTRATIFICADA PROPORCIONAL

**Lúcia Adriana dos Santos Gruginskie**

Tribunal Regional Federal da 4ª Região

Otávio Francisco Caruso da Rocha, 300, 4º andar, Porto Alegre - RS

lagruginskie@gmail.com

O objetivo do artigo é simular amostras de treze populações (listagens de autos findos) da Seção Judiciária de Santa Catarina para cinco cenários: um com erro fixado *a priori* e os demais, com tamanhos amostrais fixados em 5, 10, 15 e 20% do tamanho populacional. A partir das amostras simuladas, as estimativas foram comparadas com os valores populacionais, concluindo que o primeiro cenário apresentou as estimativas mais distantes dos valores populacionais. A técnica de amostragem usada foi a estratificada proporcional por classe de processo, conforme a Resolução 23/2008 do Conselho da Justiça Federal.

Palavras-chave: amostragem estratificada, simulação, gestão documental.

The aim of the paper is to simulate samples from 13 populations (processes completed) of the Seção Judiciária de Santa Catarina for 5 scenarios: one with desired precision determined *a priori* and the others, with sample sizes set at 5, 10, 15 and 20% of population size. From the simulated samples, the estimates were compared with the corresponding population, concluding that the first scenario showed the estimates more distant of the population values. The sampling technique used was the stratified sampling with the proportional allocation by class of process, pursuant to Resolution 23/2008 of the Conselho da Justiça Federal.

Key-words: stratified sampling, simulation, documental management.

## 1. INTRODUÇÃO

A guarda da documentação na Justiça Federal é feita, via de regra, nas próprias Seções e Subseções Judiciárias desde a década de 1960. Desta forma, de acordo com o Relatório Justiça em Números de 2009, há 118.093 metros lineares utilizados como arquivos no Judiciário Federal, embora este número desconsidere as 14 Seções Judiciárias da 1ª Região.

Quanto às condições dos arquivos, Mendes (2007) e Axt (2003) destacam como pontos negativos, a partir dos resultados de diagnóstico realizado pelo Conselho da Justiça Federal em 1996, espaços físicos inadequados e insuficientes, problemas de infiltração, ventilação inadequada e a existência de vários depósitos de documentos em diferentes locais de uma mesma cidade sede de Seção Judiciária.

Diante da considerável massa documental guardada e de suas condições surge o paradoxo em relação ao tratamento dos documentos judiciais: por um lado existe a necessidade de redução da massa arquivada e por outro, a seleção dos processos para guarda com a possibilidade de atribuição de valor histórico seletivo aos documentos, em razão de que é impossível saber quais documentos no futuro serão utilizados como fonte prioritária no estudo histórico (MENDES, 2007).

Com o objetivo de estabelecer políticas de gestão das ações judiciais transitadas em julgado e arquivadas, a Resolução do Conselho da Justiça Federal 359/2004, revogada pelas Resoluções 393/2004 e 23/2008, propõe um fluxo para a seleção de autos judiciais, composto por quatro fases: i) seleção e identificação dos processos de guarda permanente; ii) aplicação da tabela de temporalidade aos processos não selecionados no item anterior; iii) aplicação do plano amostral ao conjunto de processos após o transcurso do período determinado a fim de selecionar uma amostra para guarda; iv) análise final casuística dos processos destinados para eliminação.

Especificamente sobre a terceira fase, a Resolução propõe um plano que descreve uma maneira para representar as ações judiciais com base em técnica de amostragem estratificada para a seleção de amostras.

Inicialmente, da primeira listagem de processos analisados é aplicado o plano e guardada uma amostra. Em outro momento, uma segunda listagem é analisada e uma nova amostra é selecionada, e assim por diante. Após várias listagens, há uma reserva de processos guardados composta por todas estas amostras.

Neste contexto, o artigo se propõe a simular amostras de diferentes tamanhos a partir da aplicação do plano sugerido pela Resolução às listagens de processos da Seção Judiciária de Santa Catarina que passaram pelas fases i e ii e compará-las quanto às estimativas produzidas em relação aos valores populacionais.

Em um primeiro momento, apresenta-se o referencial teórico. A seguir, é exposta a metodologia e a discussão dos resultados. Em seguida, apresentam-se as considerações finais.

## 2. REFERENCIAL TEÓRICO

São abordados dois tópicos: amostragem, com ênfase na amostragem estratificada com repartição proporcional e a simulação.

### 2.1 Amostragem

De acordo com Good (2006), o melhor jeito de estabelecer os parâmetros da população é medir os seus elementos, mas isso nem sempre é possível. A alternativa é selecionar uma amostra e usá-la para estimar os parâmetros populacionais (por exemplo, média, proporção) que, à medida que o tamanho da amostra cresce, aproxima-se do valor populacional. Mas como a amostra não é a população, se são tomadas diferentes amostras, as estimativas também serão diferentes. Por isso, a melhor maneira de descobrir quão diferentes e quão precisas são as estimativas é tomar repetidas amostras de uma mesma população. Isto pode ser realizado através da simulação de amostras a partir de uma mesma população.

O problema básico de qualquer procedimento de amostragem é a obtenção de estimativas fidedignas de alguma característica da população de interesse, tomando como base somente parte dessa população. Segundo Lhor (1999), uma boa amostra deve reproduzir as características de interesse na população, tão próximas quanto possível. Ou seja, ser uma amostra “representativa” da população.

Sobre a representatividade da amostra, Bussab e Bolfarine (2005) argumentam que frequentemente, em linguagem popular, é dito que uma boa amostra é aquela que é representativa da população, no sentido de ser uma microrrepresentação do universo. Os autores citam que, algumas vezes, o significado de microrrepresentação confunde-se com o de uma amostra estratificada proporcional, porém, este tipo de amostra não conduz, obrigatoriamente, a resultados mais precisos. Porém, diante da dificuldade em definir amostra representativa, pode-se trabalhar com o conceito de amostra probabilística. Neste artigo, a abordagem é dada ênfase à amostragem probabilística estratificada.

De acordo com Cochran (1977), na amostragem estratificada a população de tamanho  $N$  é dividida em subpopulações de tamanho  $N_h$  ( $N_1, \dots, N_h$ ), respectivamente. Essas subpopulações (estratos) não se sobrepõem e juntas abrangem a totalidade da população.

Conforme Lhor (1999) e Cochran (1977), são quatro os principais motivos para a utilização da amostragem estratificada:

- Quando há o desejo de a amostra manter a composição da população segundo suas características básicas;
- Deseja-se uma certa precisão para os subgrupos;
- Uma amostra estratificada pode ser mais conveniente para administrar as diferenças entre os estratos;
- Amostragem estratificada, se bem realizada, pode resultar em estimativas mais precisas para a população.

De acordo com Schwartz (2011), o delineamento da estratificação segue os seguintes passos:

#### 1. Criar os estratos

A estratificação começa pelo agrupamento das unidades dentro de grupos homogêneos (estratos) onde deve haver similaridade no interior dos estratos e diferenças entre eles. A estratificação será benéfica sempre que a variabilidade da variável resposta pode ser antecipada e serem formadas camadas que são mais homogêneas do que o conjunto original das unidades de pesquisa.

Sobre o número de estratos, de acordo com Kish (1992), este pode ser limitado pelo esforço gasto na estratificação e pelo número de unidades amostrais: no mínimo, duas unidades devem ser selecionadas de cada estrato para calcular a estimativa não viciada da variância, além de que muitos estratos podem aumentar a dificuldade de análise das subclasses.

Quanto aos critérios de especificação para a formação dos estratos, de acordo com Cochran (1977), uma vez que a melhor repartição para uma determinada especificação não o será, em geral, para outra, deve-se chegar a uma conciliação nos levantamentos com numerosas especificações. A primeira providência é reduzir as especificações consideradas na repartição ao número relativamente pequeno, das que se admitem sejam as mais importantes.

#### 2. Determinar o tamanho total da amostra

O tamanho da amostra é dado por:

$$n = \frac{z_{\alpha/2}^2 \sum_{h=1}^L W_h S_h^2}{E^2 + \frac{z_{\alpha/2}^2}{N} \sum_{h=1}^L W_h S_h^2}$$

Considerando a amostra sem reposição:

$W_h = \frac{N_h}{N}$  é o peso dos estratos,

$E^2$  é o erro máximo admitido,

$S_h^2$  é a variância populacional por estrato (caso desconhecida, pode ser estimada por  $s^2$ )

$L$  é o número de estratos

$z_{\alpha/2}^2$  é o valor crítico que corresponde ao grau de confiança desejado. Quando a variância é conhecida, caso deste artigo, utiliza-se a distribuição normal para obter o valor crítico.

### 3. Alocar a amostra entre os estratos

A respeito de como alocar a população dentro dos estratos, Lhor (1999), Cochran (1977) e Schwarz (2011), colocam quatro formas de repartição: igual, proporcional, ótima com custos envolvidos e de Neyman. Neste artigo, apenas será abordada a alocação proporcional. Ou seja, proporcional no sentido de que o tamanho amostral dentro de cada estrato é proporcional ao seu tamanho na população. No caso da alocação proporcional, o tamanho da amostra em cada estrato é:

$$n_h = W_h n$$

A repartição proporcional, segundo Schwarz (2011) é preferível quando estimativas precisas são necessárias em estratos considerados mais importantes.

O uso de alocação proporcional pode ser justificado, conforme Kish (1992), quando:

- Há ganhos com as variâncias reduzidas;
- A variância pode ser menor que uma amostra não estratificada de mesmo tamanho;
- A realização é simples e fácil;
- O resultado é uma média autoponderada.

### 4. Coletar os dados em cada estrato

O sorteio da amostra pode ser através da amostragem aleatória simples ou sistemática.

### 5. Obtenção de estimativas dos estratos

$$\bar{X}_h = \frac{\sum_{i=1}^{n_h} X_{ih}}{n_h}, \text{ para a média do estrato } h$$

$$s_h^2 = \frac{\sum_{i=1}^{n_h} (X_{ih} - \bar{X}_h)^2}{n_h - 1}, \text{ para a variância}$$

### 6. Combinação das estimativas de cada estrato para a obtenção de um valor geral

$$\bar{X} = \sum_{h=1}^L W_h \bar{X}_h, \text{ para a média do estrato } h \text{ e}$$

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}, \text{ para a variância}$$

## 2.2 Simulação computacional

A simulação computacional tem servido como uma ferramenta importante em várias disciplinas: engenharia, pesquisa operacional, estatística, matemática, física, economia, biologia, medicina, engenharia, química e ciências sociais. Através da simulação, pode-se estudar o comportamento de um sistema real que seria muito difícil analisar de forma analítica (RUBINSTEIN e KROESE, 2007).

Quanto à classificação, de acordo com Law e Kelton (1991), os modelos de simulação podem ser categorizados em: i) discretos ou contínuos; ii) estáticos ou dinâmicos; e iii) probabilísticos ou determinísticos.

A simulação estática é usada para representar um sistema no qual o tempo não desempenha nenhum papel ou representa um sistema em um tempo particular, como por exemplo, simulação de Monte Carlo. Este artigo faz uso da simulação estática e probabilística, com a ênfase na geração de números aleatórios.

No caso probabilístico, a mais simples abordagem, de acordo com Hammersley e Handscomb (1964), é observar números aleatórios, escolhendo-os de tal forma que simulem o processo aleatório do problema original, e então inferir na solução do comportamento destes números aleatórios.

De acordo com Casella e Robert (2010), os métodos de Monte Carlo contam com a possibilidade de produção de um fluxo supostamente interminável de variáveis aleatórias de distribuições conhecidas ou novas, sendo que essa simulação é, por sua vez, baseada na produção de variáveis aleatórias uniformes no intervalo (0, 1).

Neste artigo, a simulação consiste basicamente em gerar aleatoriamente amostras sem reposição a partir da população com a utilização do software R, isto é, através da geração de números pseudoaleatórios.

### **3. ASPECTOS METODOLÓGICOS**

Antes de apresentar o desenvolvimento do estudo é necessário mostrar, ainda que resumidamente, a proposta da Resolução sobre a seleção da amostra.

#### **3.1 Considerações sobre o plano amostral proposto pela Resolução**

Para a realização de um levantamento amostral, Bussab e Bolfarine (2005) apontam oito passos: i) identificação dos objetivos e populações; ii) planejamento da coleta das informações; iii) planejamento e seleção da amostra, quando são tomadas as decisões do tamanho da amostra, os estimadores, entre outros; iv) processo de coleta dos dados, v) processamento dos dados, vi) análise dos resultados, vii) apresentação dos resultados e viii) disponibilidade dos dados.

Assim, há algumas considerações sobre a Resolução no tocante aos itens i a iv.

Quanto à delimitação, a Resolução não sugere critérios para os autos que compõem a população, podendo conter processos de uma ou mais Subseções Judiciárias (cidade da vara), sem a especificação de varas, sem delimitação de ano de entrada ou de saída do processo ou outras classificações, apenas propondo a divisão em estratos de acordo com as classes processuais. Diante da inexistência de critérios, a população é considerada como a listagem de processos de uma Seção Judiciária (Unidade da Federação) analisados para o “edital de eliminação”, já excluídos os autos de guarda permanente (processos anteriores a 1974, ações coletivas, processos criminais, ambientais, de direitos indígenas, entre outros) e findo o período de guarda temporária. Nesta listagem, necessariamente, deve conter o registro da classe e do tempo entre autuação e arquivamento por processo.

A título de comparação, pode ser citado o exemplo do Poder Judiciário do Rio de Janeiro. Vieira (2005) sugere a delimitação por comarca; unidade organizacional; assunto/matéria; item documental/tipo de ação; período. Como exemplo de universo a ser amostrado, o autor cita “número de exemplares de um item documental, produzidos ou recebidos por uma unidade organizacional do PJERJ, num período de tempo previamente estabelecido”.

Em relação ao planejamento da coleta das informações, a variável usada para calcular o tamanho amostral é o tempo entre autuação e arquivamento. A Resolução justifica que, para a obtenção de uma amostra aleatória consistente e com baixa correlação entre os estratos, é necessário trabalhar com uma variável resposta comum a todas as classes de processos e isenta de juízo de valor.

Quanto ao planejamento e a coleta da amostra, a Resolução propõe o uso da técnica de amostragem estratificada proporcional com os processos selecionados através da tabela de

números aleatórios. A Resolução justifica a utilização de amostragem estratificada em decorrência da heterogeneidade do tempo de tramitação entre as classes processuais.

Após a seleção, são calculados os estimadores amostrais e populacionais, com o objetivo de demonstrar a precisão da grandeza amostral encontrada. Quanto à precisão, a Resolução não define ou coloca critérios quanto ao erro amostral ou grau de confiança.

Também para comparação, podem ser citados exemplos do Judiciário Brasileiro e um exemplo mundial quanto ao tamanho amostral ou a precisão estipulada para seu cálculo.

O Tribunal Regional do Trabalho da 7ª Região, através do ato 78 de 2002, que regulamentou os procedimentos referentes ao arquivamento de processos judiciais, preservação da memória institucional e a eliminação de autos, delimitou o número de processos selecionados para fins de reserva restrito a 2% do total a ser eliminado.

Vieira (2005) propõe, para o Tribunal de Justiça do Rio de Janeiro, a técnica de amostragem aleatória simples sem reposição para a seleção de processos, fixando o grau de confiança em 95% e o erro máximo admitido em 10%.

Como exemplo mundial, Axt (2003) cita os exemplos dos Estados Unidos e da França, onde de 5 a 20% dos documentos administrativos são preservados, sendo os demais eliminados.

Se a precisão for estipulada por listagem analisada, em uma listagem qualquer, a fração amostral  $\frac{n}{N}$  pode ser diferente da fração de outras, dependendo da variabilidade.

Desta forma, após sucessivas listagens analisadas, as estimativas de médias e proporções obtidas podem estar “muito” diferentes em relação aos seus valores populacionais. Além disso, processos em populações com maior variabilidade teriam uma chance maior de serem guardados em comparação a processos de listagens com menor variabilidade.

Em geral, a Resolução do CJF tem sido vista como inovadora, como por exemplo, no artigo de Axt (2003):

“o critério da seleção de documentos pelo método de proporcionalidade e amostragem foi a grande novidade introduzida pela política de gestão documental do Judiciário Federal, transparecendo a preocupação em desenvolver um método que garanta o engajamento do acervo, levando em conta a pertinência do argumento administrativo, sem basear-se no critério tradicional de valor.”

### 3.2 Desenvolvimento do estudo

Considerando que, de acordo com a Resolução, não há critérios para a delimitação da população e não estão fixados, para o cálculo do tamanho da amostra, o erro admitido e o grau de confiança, surgem os seguintes questionamentos: i) após sucessivas amostras de tamanho calculado com base na precisão fixada *a priori*, as estimativas da reserva de processos guardada (conjunto das amostras selecionadas) estariam próximas aos valores populacionais (processos guardados somado aos eliminados)? ii) neste sentido, fixar o tamanho da amostra em um percentual da população produziria resultados mais precisos que calcular o tamanho da amostra com um erro fixado *a priori*? iii) que percentual poderia ser usado para estimar não apenas o tempo de tramitação, mas também outras variáveis?

Embora, segundo Lhor (1999), a precisão não seja obtida pela fração que a amostra cobre na população, as questões ii e iii são colocadas por ser uma forma de obter amostras retratando a realidade do acervo ao longo do tempo, considerando que não se sabe os critérios de definição das populações das quais as amostras foram retiradas.

Para responder estas questões foi realizado um estudo a partir da Seção Judiciária de Santa Catarina. Foram 13 listagens de autos findos de 2005 a 2010, de quatro classes (Procedimento do Juizado Especial Federal, Ação Ordinária, Mandado de Segurança e Execução Fiscal). A tabela 1 apresenta as classes e o coeficiente de variação do tempo entre autuação e arquivamento segundo a listagem de processos analisados para compor o edital de eliminação. É apresentado o coeficiente de variação para mostrar a variabilidade das variáveis

em relação à média, sem mostrar a última medida. As classes também são mostradas como A, B, C e D.

Tabela 1 – Coeficiente de variação e quantidade de processos analisados segundo a classe processual e a listagem

Listagem para o edital de eliminação	Classe								Total	
	A		B		C		D		Número	CV
1			5132	136%			3116	110%	8248	136%
2			1050	126%					1050	126%
3	11949	63%	7582	100%					19531	88%
4					6447	41%			6447	41%
5			6000	121%			11317	91%	17317	100%
6					5128	37%			5128	37%
7					5776	37%			5776	37%
8			2706	61%	3525	48%			6231	100%
9	1832	49%			3770	61%			5602	128%
10					3956	53%			3956	53%
11	1535	72%			6550	45%			8085	144%
12					9315	50%	2370	62%	11685	98%
13					4467	33%			4467	33%
<b>Total</b>	<b>15316</b>	<b>62%</b>	<b>22470</b>	<b>133%</b>	<b>48934</b>	<b>53%</b>	<b>16803</b>	<b>91%</b>	<b>103523</b>	<b>145%</b>

Fonte: Listagens de processos selecionados para análise

A tabela 2 apresenta os tamanhos amostrais calculados para os editais de acordo com quatro margens de erro fixadas. Estas dimensões foram calculadas utilizando as variâncias do tempo de tramitação das classes processuais, ao grau de confiança de 95% para a amostragem estratificada com alocação proporcional.

Tabela 2 – Tamanhos amostrais calculados segundo o erro fixado a priori

Listagem para o edital de eliminação	Erro fixado			
	0,04xMédia	0,06xMédia	0,08xMédia	0,10xMédia
1	2920	1616	994	665
2	830	657	509	395
3	1703	796	456	294
4	389	179	102	66
5	2172	1038	600	389
6	329	152	87	56
7	324	149	85	55
8	1104	544	319	208
9	1054	523	307	201
10	601	292	170	111
11	1911	978	581	382
12	1155	543	312	202
13	260	120	68	44
<b>Total</b>	<b>14752</b>	<b>7587</b>	<b>4590</b>	<b>3068</b>

Fonte: Dados calculados com base nas listagens de processos

Como se conhecem os valores das médias populacionais, os erros foram fixados em relação a este parâmetro de tal forma que o erro dividido pela média seja igual a 0,04; 0,06; 0,08 e 0,10.

Devido a diferentes variabilidades, a fração  $n/N$  é distinta entre as listagens. Por exemplo, na primeira listagem para o edital de eliminação, considerando o erro fixado de 0,04 x média, o tamanho amostral é de 2920 processos (fração amostral de 0,35), para a quinta listagem, com as mesmas classes, a amostra é de 2172, ou seja, fração de 0,13.

Para dar o prosseguimento ao estudo, foi possível obter, para estas listagens, além do tempo entre autuação e arquivamento e da classe do processo, a subseção judiciária, a parte, o ano de entrada e de saída, conforme descrito a seguir:

- Parte. Foi agrupada a Parte (ré ou autora) de acordo com a frequência: CEF (Caixa Econômica Federal), Conselhos Regionais e Federais, Fazenda Nacional, Instituto Nacional de Seguridade Social, Receita Federal, União Federal e demais partes.
- Subseção. Consideradas as subseções de Florianópolis, Itajaí, Criciúma, Chapecó, Blumenau, Joinville, Joaçaba, Lages e Tubarão e sem a especificação de subseção.
- Ano de entrada e de saída. Assim como o ano de saída, o ano de entrada foi agrupado nas seguintes faixas: de 1974 a 1978, de 1979 a 1983, de 1984 a 1988, de 1989 a 1993, de 1994 a 1998, de 1999 a 2003 e 2004 em diante. Considerando que, após findo, o auto judicial fica guardado por tempo determinado pela tabela de temporalidade.

O propósito é, a partir das treze listagens, simular amostras possíveis de acordo com cinco cenários. Após a obtenção de amostras “gerais” (de todos as listagens para editais de eliminação), comparar o comportamento destas em relação não só ao tempo médio entre autuação e arquivamento, mas também a outras variáveis.

Foram comparados cinco cenários:

Cenário I – Erro fixado em 0,04 x tamanho da média, conforme a tabela 2. Assim, a amostra da primeira listagem para o edital de eliminação foi igual a 2920, a amostra do segundo, 830 e assim por diante, guardando, ao final, 14.752 processos. Este é o único cenário com erro fixado *a priori*.

Cenários II a V – independente da variabilidade de cada listagem de processos analisados para o edital de eliminação, os tamanhos amostrais foram fixados, respectivamente, em 5%, 10%, 15% e 20% do tamanho populacional. Estes percentuais foram propostos seguindo os exemplos de órgãos que eliminam/guardam processos.

Convenções:

$j$  listagens ( $j = 1, \dots, 13$ )

$k$  cenários ( $k = 1, \dots, 5$ )

$i$  amostras geradas para cada cenário  $k$  e listagem  $j$  ( $i = 1, \dots, 10.000$ )

$l$  variáveis  $Y$  ( $l = 1, \dots, 35$ ), para as proporções e para as médias:

- proporções das subseções (A, B, C, D, E, F, G, H, I, J); da faixa de ano de entrada; da faixa de ano de saída; e da parte do processo (A, B, C, D, E, F e G);

- média e variância do tempo entre autuação e arquivamento de acordo com as classes A, B, C, D.  $l = 36$ , refere-se a média de tempo geral

$\theta_l$  como o parâmetro populacional da variável  $l$  (média ou proporção)

$\sigma_l$  como o desvio padrão populacional da variável  $l$

$n_{jk}$  como o tamanho amostral em cada uma das listagens no cenário  $k$

$n_k$  como o tamanho amostral em cada cenário, ou seja,  $n_k = \sum_{j=1}^{13} n_{jk}$

$Y_{ilk}$  como estimativas da variável  $l$  para  $\theta_l$ , na amostra  $i$ , no cenário  $k$ ,

O estudo consistiu nas seguintes etapas:

- Etapa 1: Selecionar, para cada um dos  $k$  cenários e  $j$  listagens, 10.000 amostras sem reposição de tamanho  $n_{jk}$ , obtendo 10.000 amostras de dimensão  $n_k$ . A seleção das



amostras foi feita utilizando o software R através da geração de números pseudoaleatórios;

- Etapa 2: Obter, para cada uma das  $i$  amostras, as estimativas  $Y_{ilk}$  para as  $l$  variáveis no cenário  $k$ ;
- Etapa 3: Obter, para cada amostra  $i$ , a variável  $Z_{ilk}$ , a partir da comparação de  $Y_{ilk}$  com o parâmetro populacional da seguinte forma:

$$Z_{ilk} = 0, \text{ se } \left(\theta_l - \frac{\sigma_l}{20}\right) \leq Y_{ilk} \leq \left(\theta_l + \frac{\sigma_l}{20}\right) \quad \text{ou}$$

$$Z_{ilk} = 1, \text{ caso contrário}$$

Os valores de  $\sigma_l/20$  para as proporções estão na tabela 3, bem como os valores de  $\theta_l$ .

Embora o erro tenha sido arbitrado, ele fornece um referencial comparativo para os cenários. Se um valor comparativo maior ou menor for fixado as conclusões podem ser diferentes.

- Etapa 4: Calcular o valor de  $M_{ik}$ , quantidade de variáveis fora do intervalo por amostra  $i$  cenário  $k$ , podendo assumir valores de 0 a 35.

$$M_{ik} = \sum_{l=1}^{35} Z_{ilk}$$

- Etapa 5: Tabular os valores da variável  $M_{ik}$  para comparação;
- Etapa 6: Calcular o valor  $M_k$  (quantidade de amostras que não contem o valor populacional da média de tempo entre autuação e arquivamento):

$$M_k = \sum_{i=1}^{10000} Z_{i,36,k}$$

A simulação consistiu na extração de amostras das listagens já conhecidas. Contudo, estas não abrangem a totalidade de Subseções Judiciárias nem das classes judiciais. Porém esta é a população referida, ou seja, aquela que, segundo Bussab e Bolfarine (2005), está previamente disponível e descrita pelo sistema de referência e para a qual podem ser construídas e selecionadas as unidades amostrais.

#### 4. RESULTADOS E DISCUSSÕES

A tabela 3 apresenta os erros para cada uma das proporções e o parâmetro populacional. São omitidos os valores das médias. Para as proporções, o maior erro foi observado para a proporção de processos com ano de entrada de 1999 a 2003, porém, quando a proporção é próxima de zero, como por exemplo, para a subseção J, o erro, apesar de baixo, é maior que o valor populacional.

Após a geração das amostras e a obtenção da variável  $M_{ik}$ , através da execução dos passos de 1 a 4, foi obtida a tabela comparando os resultados dos cenários, conforme a etapa 5. São comparados, na tabela 4, os 5 cenários quanto ao número de variáveis fora do intervalo de  $\theta_l \pm \frac{\sigma_l}{20}$ .

Observa-se que, no cenário que fixa o tamanho em 5% da população, 6.835 das 10.000 amostras contêm as estimativas de todas as 35 variáveis dentro do intervalo. Para o cenário fixado em 15%, 9831 amostras contêm todas as variáveis dentro do intervalo de comparação. Quando se toma amostras de tamanho igual a 20% da listagem analisada, este número sobe para 9983 amostras.

Porém, quando o erro máximo admitido é fixado em 0,04 x média, todas as amostras possuem pelo menos 4 das 35 variáveis fora do intervalo de comparação. Ou seja, caso o erro fosse fixado por população analisada, a amostra resultante, após sucessivas listagens estaria com estimativas “distantes” dos valores populacionais, considerando o erro estipulado.

Desta forma, acredita-se que, fixando o tamanho amostral em 10%, o resultado já é satisfatório: 9226 das 10000 amostras apresentam as estimativas das 35 variáveis dentro do intervalo estipulado.

Tabela 3 – Proporções populacionais e valor comparativo

Variável	Proporção populacional	Valor comparativo
Subseção		
A	0,0101	0,0050
B	0,2477	0,0216
C	0,0166	0,0064
D	0,0780	0,0134
E	0,3847	0,0243
F	0,0052	0,0036
F	0,0397	0,0098
H	0,2046	0,0202
I	0,0109	0,0052
J	0,0024	0,0025
Faixa de ano de entrada		
De 1974 a 1978	0,0124	0,0055
De 1979 a 1983	0,0296	0,0085
De 1984 a 1988	0,0548	0,0114
De 1989 a 1993	0,1924	0,0197
De 1994 a 1998	0,1645	0,0185
De 1999 a 2003	0,5402	0,0249
De 2004 em diante	0,0061	0,0039
Faixa de ano de entrada		
De 1974 a 1978	0,0036	0,0030
De 1979 a 1983	0,0116	0,0054
De 1984 a 1988	0,0257	0,0079
De 1989 a 1993	0,0448	0,0103
De 1994 a 1998	0,2376	0,0213
De 1999 a 2003	0,4387	0,0248
De 2004 em diante	0,2380	0,0213
Parte do processo		
A	0,0690	0,0127
B	0,0149	0,0061
C	0,1069	0,0154
D	0,0565	0,0115
E	0,6315	0,0241
F	0,0750	0,0132
G	0,0462	0,0105

Fonte: Dados calculados com base nas listagens de processos

Quanto à média do tempo entre autuação e arquivamento, a quantidade de amostras em que o intervalo  $\theta_i \pm \frac{\sigma_i}{20}$  contém o valor populacional é de 9890 para o cenário de 5%, sendo que para os cenários de 10%, 15% e 20%, a totalidade das amostras localizou-se dentro do intervalo de comparação.

No cenário em que o erro é fixado em 0,04 x média, nenhuma das 10.000 amostras contém a estimativa dentro dos limites.

Tabela 4 – Comparação entre os cenários

Quantidade de variáveis fora do intervalo $\theta_i \pm \frac{\sigma_i}{20}$	Cenário – Tamanho da amostra					Calculado através de erro fixado <i>a priori</i> em 0,04 x média
	5% de N	10% de N	15% de N	20% de N		
0	6835	9226	9831	9983		0
1	2823	757	169	17		0
2	336	17	0	0		0
3	56	0	0	0		0
4 ou mais	0	0	0	0		10000
Total	10000	10000	10000	10000		10000

Fonte: Dados obtidos da simulação de amostras

## 5. CONSIDERAÇÕES FINAIS

As considerações finais tratam da digitalização do acervo em vez da guarda de uma amostra de processos, das restrições do estudo e das conclusões obtidas.

De acordo com Axt (2003), a digitalização tem sido sugerida pelos arquivistas como uma forma não de substituir na íntegra um arquivo em base impressa, mas como uma maneira de facilitar o acesso a uma parte da documentação de um arquivo, normalmente muito consultada. Com o avanço da tecnologia, a digitalização também necessita de atualização das mídias de leitura de tal forma que possibilite seu uso no futuro, o que pode ser caro, como no exemplo “arquivo das Índias”, do departamento do arquivo espanhol, citado por Sordi e Marques *apud* Axt (2003): o departamento gastou US\$ 10 milhões em 1992 para digitalizar 5% do seu acervo e em 1999, US\$ 2,5 milhões para migrar de mídia porque os equipamentos da época não liam os CDs antigos.

A segunda consideração é que o artigo esteve restrito às listagens de Santa Catarina em um determinado período de tempo. A generalização do tamanho amostral como procedimento padrão deveria contar com mais alguns estudos.

Por exemplo, poderia ser aplicado na Seção Judiciária do Rio Grande do Sul, onde acredita-se que as variabilidades, pelo menos das variáveis de tempo, sejam maiores.

Além disso, as listagens de processos, até o momento, não representam a totalidade de processos a serem guardados e eliminados da Seção Judiciária de Santa Catarina. Outro estudo poderia ser elaborado considerando não apenas as treze listagens, mas a totalidade de processos através da simulação de populações (listagens) em vez da simulação de amostras. Considerando que a Resolução não delimita a população, este outro estudo poderia comparar diferentes critérios na escolha dos processos que compõem as listagens, como a Parte, o ano de entrada ou de saída, Subseção Judiciária ou outros critérios.

Poderiam ser testadas outras técnicas de amostragem, como a aleatória simples ou sistemática e com a inclusão de outras variáveis às analisadas, como o custo de armazenamento e o volume do processo.

Em relação aos cenários comparados neste estudo, observa-se que calcular o tamanho amostral com o erro fixado por listagem de processos pode levar a uma guarda de processos com estimativas diferentes da população analisada depois de sucessivas listagens.

Quanto ao cálculo do tamanho amostral através da fixação de um percentual, amostras de tamanho 10% da população revelaram-se apropriadas para estimar as variáveis analisadas neste artigo, usando o plano fixado pela Resolução. Dependendo do valor comparativo, amostras menores ou maiores poderiam ser usadas.

Quando da eliminação, esta é feita por meio da reciclagem do material descartado e da destinação do resultado para programas sociais de entidades sem fins lucrativos, ou seja, mediante critérios de responsabilidade social e de preservação ambiental.

## REFERÊNCIAS

- AXT, G.** *Algumas reflexões sobre os critérios para a identificação e guarda dos processos judiciais históricos*, 2003. Disponível em <<http://www.tjrs.jus.br>>. Acesso em mar. 2010.
- BUSSAB, W., BOLFARINE, H.** *Elementos de amostragem*. São Paulo: Blucher, 2005.
- CASELLA, G., ROBERT, C.** *Introducing Monte Carlo method with R*. New York: Springer, 2010.
- COCHRAN, W.** *Sampling Techniques*. 3. ed. New York: John Willey and Sons, 1977.
- CONSELHO DA JUSTIÇA FEDERAL.** *Resolução 23/2008*. Disponível em <<http://www.jf.jus.br>> Acesso em mar. 2011.
- CONSELHO NACIONAL DE JUSTIÇA.** *Justiça em números 2009: Indicadores do Poder Judiciária*. Disponível em <<http://www.cnj.jus.br>>. Acesso em dez. 2010.
- GOOD, P.** *Resampling methods: a practical guide to data analysis*. 3. ed. Boston: Birkhauser, 2006.
- HAMMERSLEY, J.; HANDSCOMB, D.** *Monte Carlo methods*. New York: Jhon Willey & Sons, 1964.
- KISH, L.** *Survey sampling*. New York: Jhon Willey & Sons, 1992.
- LAW, A. M.; KELTON, D.** *Simulation modeling and analysis*. 2. ed. New York: McGraw-Hill, 1991.
- LOHR, S.** *Sampling: designs and analysis*. Belmont, CA: Duxbury Press, 1999.
- MENDES, V.** *Gestão documental e manutenção da informação na Justiça Federal – RS*. Mouseion, V.1, n.2, p. 10-21. 2007. Disponível em <<http://www.unilasalle.edu.br>>. Acesso em mar. 2011.
- RUBINSTEIN, R.; KROESE, D.** *Simulation and the Monte Carlo method*. 2. ed. New York: Jhon Willey & Sons, 2007.
- SCHWARZ, C.** *Sampling, regression, experimental design and analysis for environmental scientists, biologists, and resource managers*. 2011. Disponível em <<http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/PDF/ChapterChangePoint.pdf>>. Acesso em abr. 2011.
- TRIBUNAL REGIONAL DO TRABALHO DA 7ª REGIÃO.** *Ato 78/2002*. Disponível em <<http://www.trt7.gov.br>>. Acesso em abr. 2011.
- VIEIRA, C.** A coleta de amostras representativas de um acervo documental: o caso do poder judiciário do estado do Rio de Janeiro. *Arquivística.net*. v.1, n.2, p. 31-42. 2005. Disponível em <<http://www.arquivistica.net>>. Acesso em abr. 2011.