Anexo N – Plano para amostra estatística representativa

I. Introdução

Este documento propõe uma metodologia estatística baseada em técnica de amostragem, com o intuito de uma estratégia de guarda amostral para documentos e processos que serão eliminados. Essa amostra deve permitir ao usuário do acervo preservado reconstruir procedimentos, técnicas e normas utilizadas à época da criação do documento bem como dar condições ao usuário de fazer pesquisas por meio da documentação preservada. Utiliza-se a guarda amostral também para representar as funções e atividades do órgão em determinado período.

Como não é possível determinar com precisão usos potenciais futuros dos documentos, a arquivística utiliza o plano amostral como ferramenta para atender ao futuro usuário desse acervo.

Ao considerar a heterogeneidade dos acervos de todos os segmentos da Justiça, pretende-se obter um plano amostral simples, didático que contemple as especificidades dos tribunais.

II. Metodologia

a) Escolha do Plano Amostral

A proposta do plano amostral que será apresentado neste estudo consiste em elaborar metodologia de fácil aplicação, tanto do ponto de vista matemático quanto do ponto de vista das informações que deverão ser coletadas e utilizadas no cálculo da amostra.

Isso não significa, entretanto, que este documento possa ser utilizado sem análise crítica dos dados disponíveis. Na verdade, como cada tribunal e cada justiça apresenta realidades distintas, é pouco provável que todas as populações contenham o mesmo grau de dispersão. É essencial que a aplicação da amostra seja acompanhada por um estatístico responsável.

Some-se a isso a questão de que, qualquer que seja o plano amostral desenvolvido para armazenamento de um acervo documental, não se sabe ao certo o objeto das pesquisas que serão realizadas no futuro, ou seja, o parâmetro que se quer estimar, um dos elementos mais importantes na seleção da amostra, nesse caso não é conhecido. Por esse motivo, a proposta baseia-se na seleção de amostras para estimar proporções, adotando valores que gerem maior variância possível, o que irá gerar amostra baseada em critério mais conservador.

Há dois tipos de amostragem: as não probabilísticas e as probabilísticas. As amostras feitas por julgamento do pesquisador são não probabilísticas. Nas amostras probabilísticas, a probabilidade de seleção de cada item ou indivíduo da população é conhecida, fazendo que seja possível estimar o nível de erro. O grande diferencial das amostras probabilísticas é que seus resultados podem ser generalizados para toda a população, enquanto nas não probabilísticas o mesmo não ocorre.

Outro ponto relevante é que para uma amostra ser probabilística é essencial que o pesquisador conheça todos os elementos de sua população e a seleção dos itens seja feita utilizando critérios aleatórios, que independem da escolha do pesquisador. Por exemplo, se o pesquisador escolher os processos que serão eliminados segundo sua opinião, a amostra deixará de ser probabilística, já que, no critério de seleção, estão embutidos, mesmo indiretamente, critérios subjetivos que fogem à aleatoriedade.

Nos casos em que há muita heterogeneidade dos dados da população, a amostragem estratificada se apresenta como uma metodologia indicada. A ideia desse plano amostral consiste em dividir a população que, nesse caso, corresponde ao universo de todos os documentos e processos arquivados passíveis de eliminação em grupos homogêneos (parecidos) entre si. Como critério de estratificação, optou-se por considerar o ano de distribuição dos documentos e processos. A opção pela adoção do ano de distribuição na construção dos estratos baseia-se na premissa de que esse critério reflita as questões apresentadas em juízo em determinado momento histórico.

Outro motivo é que, como não há obrigatoriedade de aplicar a amostragem periodicamente, pelo contrário, é, inclusive, preferível que os tribunais aguardem acumular um determinado número de processos — já que, com um universo pequeno, as estimativas amostrais podem perder precisão —, é natural que, com o acúmulo, haja mais processos antigos do que novos. Ao aplicar simplesmente uma amostra aleatória simples, não se garante a seleção dos processos mais recentes, que podem justamente ser objeto de estudo de futuro pesquisador.

Sugere-se que, para evitar número excessivo de estratos, o ano de distribuição seja agrupado em intervalos. Por exemplo, fazendo grupos de dois anos, o primeiro estrato poderia ser formado pelos documentos e processos distribuídos de janeiro de 2008 a dezembro de 2010, o segundo estrato seria formado pelos documentos e processos de janeiro de 2006 a dezembro de 2008 e, assim, sucessivamente. Nos anos mais antigos, caso restem poucos processos, eles podem ser agrupados em intervalos de tempo maiores.

É relevante destacar que o número de estratos deve ser suficiente para separar a heterogeneidade da população, mas não deve ser um número excessivo a fim de não segmentar demais a população e obter muitos estratos com pequenas populações em cada um.

No caso específico da Justiça Federal, constatou-se que cerca de 47% dos assuntos dos processos distribuídos durante o ano referem-se a direito previdenciário e 52,4% das classes são de execução fiscal, ou seja, grande parte da massa de processo abrange apenas esses dois tipos de matérias. Retirando esses dois tipos de processo, restam, proporcionalmente, poucos, mas são os mais relevantes para fazer a guarda amostral, já que representam todos os demais tipos de processos que tramitam na Justiça Federal.

Assim sendo, cada Justiça elaborará amostras estratificadas por ano de distribuição do processo e outras que versem sobre assuntos repetitivos (ações de massa, como, por exemplo, as existentes sobre execuções fiscais, ações de direito previdenciário revisionais e outras ações versando sobre FGTS, poupança, empréstimo compulsório, acordos trabalhistas, etc.).

b) Cálculo do tamanho da amostra

Para cálculo do tamanho de uma amostra estratificada, aplica-se a seguinte formulação matemática:

$$n = \frac{\sum_{i=1}^{L} N_i^2 p_i(1-p_i)/w_i}{N^2 D + \sum_{i=1}^{L} N_i\, p_i(1-p_i)},$$

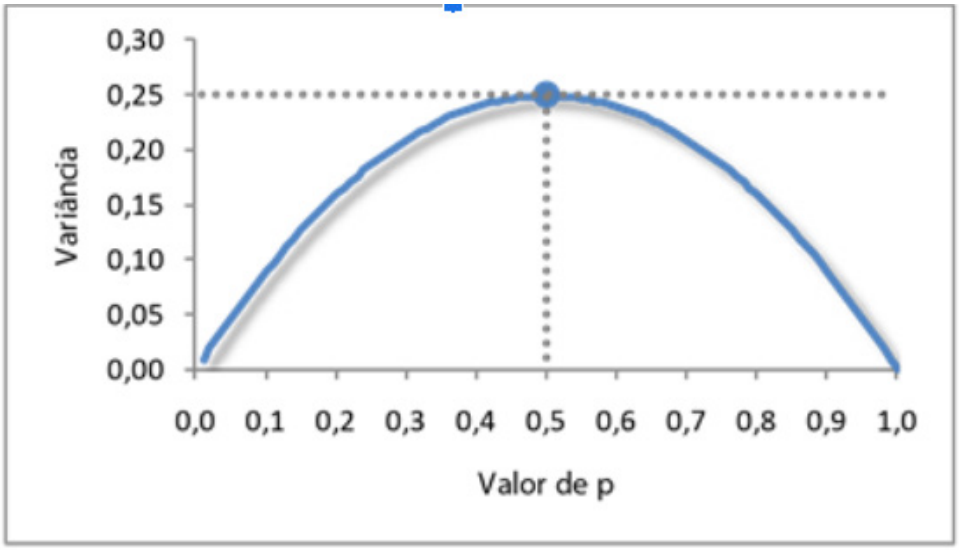
em que é o parâmetro que se quer estimar, L é o número de estratos,

Sendo B o erro máximo desejado e z o grau de confiança, N é o tamanho total da população, Ni é o tamanho da população no i-ésimo estrato.

Considera-se teste esse estudo para o caso em que se deseja estimar a proporção de uma população. Nas estimações por proporções, estima-se que a distribuição de probabilidade da variável analisada segue uma distribuição binomial, cuja média é dada por E(x) = p e variância por Var(x) = p·(1-p). Tendo em vista que, nesse caso, não se sabe o objeto de estudo do pesquisador que, no futuro, venha a utilizar essa amostra, indica-se a atribuição do valor de p igual a meio (p = 0,5), pois dessa forma se garante a maior variância possível.

O Gráfico 1 ilustra como a variância de uma distribuição binomial se altera de acordo com a escolha do parâmetro p e demonstra que, para a escolha p = 0,5, o máximo da variância é encontrada (variância igual a 0,25).

Gráfico 1 – Variância de uma distribuição binominal de acordo com o parâmetro P



Como o parâmetro p é desconhecido, será atribuído p fixo para todos os estratos, ou seja, = 0,5, para i = 1, .., L. Nesse caso, a equação (1) pode ser simplificada para:

$$n = \frac{Np(1-p)}{N \cdot \left(\frac{B^2}{z_\alpha^2}\right) + p(1-p)}.$$

c) Escolha da margem de erro

A margem de erro da amostra vai depender da escolha de dois parâmetros: do erro máximo desejado (B), que está relacionado com a magnitude do parâmetro que se quer estimar (p), e do grau de confiança (z).

Para esse estudo, sugere-se a utilização do erro B = 0,03, com uma margem de confiança de 97% (ou margem de erro de 3%), que gera um valor de z = 2,17. Para a amostra referente às ações repetitivas é possível usar uma margem de confiança um pouco menor, igual a 95% (z = 1,96), já que toda a diversidade em relação ao tipo de processo está em outra amostra.

Tendo em vista que p = 0,5 será sempre fixo, é possível simplificar ainda um pouco mais a fórmula, conforme a equação seguinte.

$$\mathbf{n = \frac{0,25 \cdot N}{N \cdot \left(\frac{B^2}{z_{\alpha}^2}\right) + 0,25}.}$$

Foi inserido um arquivo Excel anexo que já contém a fórmula e possui um campo para inserir o tamanho da população (número de processos que podem ser eliminados), em que as opções de escolha do grau de confiança do erro máximo desejado são de livre escolha.

d) Alocação dos estratos

Após a determinação do tamanho total da amostra, faz-se necessário verificar o tamanho da amostra que será selecionada em cada estrato. O critério adotado baseia-se na alocação proporcional, cujo procedimento consiste em distribuir proporcionalmente a amostra de tamanho em relação ao tamanho do estrato, isto é:

$$n_h = nW_h = n \cdot \frac{N_h}{N}$$

e) Ponto de corte: quantidade mínima de processos arquivados para aplicar a amostra

Para que a estimativa do tamanho da amostra seja precisa, é importante que a população seja suficientemente grande, ou seja, o número de processos que estão arquivados e são passíveis de eliminação deve ser grande suficiente para garantir uma amostra adequada. Verificou-se, para os parâmetros p = 0,5 e B = 0,03, qual seria o tamanho da amostra de acordo com o tamanho da população. Os resultados estão apresentados no gráfico 2.

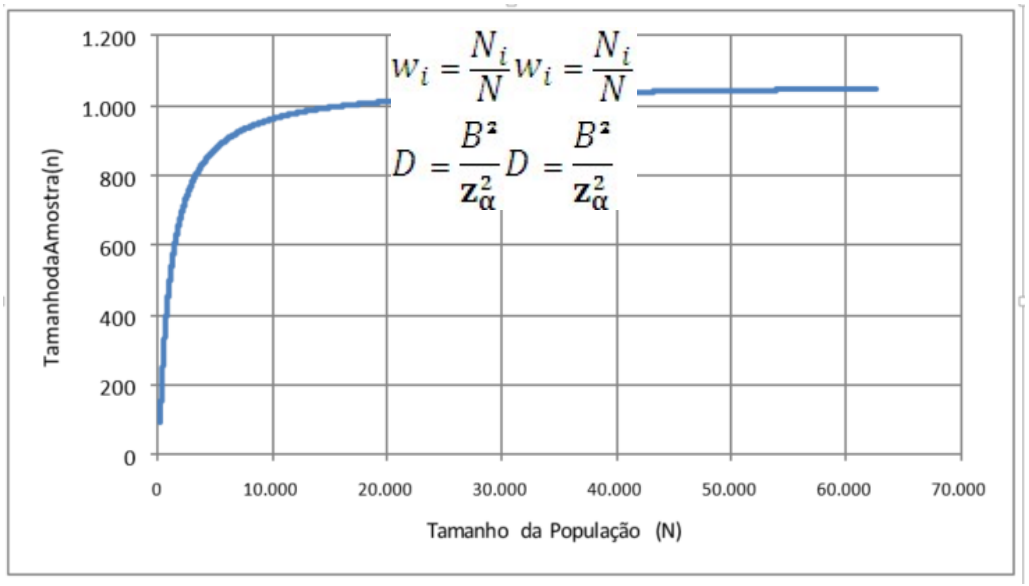
Note-se que, a partir de 10.000 processos arquivados, a amostra atinge o quantitativo 964 e, mesmo aumentando o universo para 20.000 processos arquivados, a amostra calculada fica igual a 1.013, ou seja, o dobro do tamanho da população gera uma amostra apenas 5% maior.

O objetivo de fazer essa análise consiste em determinar um ponto de corte a partir do qual é possível aplicar a amostra. Sendo assim, orienta-se que, para os parâmetros

escolhidos, acumulem-se no mínimo 10.000 processos para calcular a amostra. É relevante destacar que esse ponto de corte foi escolhido para uma margem de erro de B = 0,03 e esse valor muda sensivelmente a qualquer alteração do valor do erro escolhido.

Destaca-se ainda, que o tamanho da população se refere apenas àqueles processos passíveis de eliminação e não incluem os de guarda permanente.

Gráfico 2 – Variação do tamanho da amostra em relação à população



Anexo 0 – Planilha de cálculo

Download da planilha Excel para cálculo do tamanho da amostra e dos estratos.