

## ATA “INFRAESTRUTURA CADASTRAL E AMOSTRAGEM”

Elaborada por Marcelo de Moraes Duriez

O Coordenador de Agropecuária, Sr. Flávio Bolliger, convidou o Sr. Marcos Paulo S. de Freitas (IBGE/COMAQ), responsável por estabelecer o plano amostral para o SNPA, a fazer sua explanação a respeito dos estudos e decisões já tomadas nos últimos quatro anos. Atendendo à solicitação da COAGRO, dentre as pesquisas do SNPA, a escolhida para iniciar os trabalhos foi a PNAG (Pesquisa Nacional de Atividade Agropecuária).

O primeiro ponto enfatizado pelo Sr. Marcos Paulo foi a importante decisão de tratar a população de estabelecimentos agropecuários como populações distintas, constituindo dois estratos, e utilizar estratégias de amostragem diferentes para cada estrato. Para uma parte da população, se construirá um cadastro de lista e os estabelecimentos que integrarão este cadastro deverão atender a determinados critérios, previamente estabelecidos. A outra parte da população será abrangida por um cadastro de área, que é o conjunto dos setores censitários do IBGE. O Sr. Marcos esclareceu que no cadastro de lista ficariam os grandes estabelecimentos, pois eles são difíceis de serem captados na área. Já no outro cadastro, ficarão aqueles estabelecimentos que não atenderam aos critérios mencionados. Também salientou-se que um estabelecimento que está em um cadastro, não poderá estar no outro.

A seguir, o técnico da COMAQ relatou que, primeiramente, eles estavam pensando em utilizar o método do cadastro duplo. Neste método, incidem sobre toda a população dois cadastros: um de lista e outro de área. Utiliza-se os dois cadastros para que um supra a deficiência do outro. O principal problema do cadastro de lista é a desatualização. A princípio, o cadastro é uma lista completa e atualizada. Contudo, com o passar do tempo esta lista vai se desatualizando, perdendo sua completude. Conseqüentemente, ela deve passar por frequentes atualizações e, para este processo, é necessário ter fontes cadastrais confiáveis e que disponibilizem facilmente as informações, o que, geralmente, é difícil de se obter. Por outro lado, o cadastro de áreas é completo, cobre todo o território e não se desatualiza. Entretanto, devido à distribuição assimétrica das variáveis de interesse nos estabelecimentos agropecuários, há a possibilidade de se ter problemas ao usar somente o cadastro de áreas para fazer sua amostra. Explicando melhor, o Sr. Marcos Paulo acrescentou que há poucos estabelecimentos grandes, que concentram uma parcela majoritária das informações de interesse e são pouco distribuídos no território. Em contrapartida, há um grande número de estabelecimentos pequenos, com parcela minoritária das variáveis de interesse e bem distribuídos no território. Ao utilizar só o cadastro de áreas para selecionar a amostra, pode ser que

nenhum estabelecimento grande seja selecionado, já que estes são pouco distribuídos no espaço. Além do mais, se for selecionado um estabelecimento grande, este entrará na amostra com o peso do setor no qual ele estava, que poderá ser um peso elevado se no setor houver um número não desprezível de estabelecimentos pequenos. Neste caso, conseqüentemente haverá superestimação para as variáveis de interesse deste estabelecimento.

Ainda explicando a metodologia de cadastro duplo, o Sr. Marcos Paulo salientou que, ao fazer as estimativas, se utiliza um peso para aqueles estabelecimentos advindos do cadastro de lista e outro peso para aqueles estabelecimentos do cadastro de área. E colocou: o que fazer quando o estabelecimento está nos dois cadastros? Segundo ele, quando acontece isto, nos países que adotam esta metodologia, se atribui valor zero ao peso proveniente de um dos cadastros. Com isso, o peso final considera somente um dos pesos, ou o da área ou o da lista. Concluindo este tema, o Sr. Marcos salientou que a COMEQ achou melhor não utilizar a técnica de cadastro duplo e adotar a metodologia descrita no segundo parágrafo deste texto.

Uma vez que o conjunto dos estabelecimentos será tratado como duas populações distintas, a COMEQ pensará em um modelo amostral para cada “subpopulação” separadamente. Na parcela da população que comporá o cadastro de lista, a amostra deverá seguir o mesmo modelo das pesquisas econômicas. Como o IBGE possui conhecimento e tradição em realizar amostragens deste tipo, num primeiro momento, a Coordenação de Métodos e Qualidade se dedicará mais ao plano amostral da parte da população que comporá o cadastro de áreas.

O primeiro passo do plano amostral é definir os cadastros. No caso da PNAG, o cadastro de área será composto pelos estabelecimentos que não estão no cadastro de lista, devendo-se primeiro criar o cadastro de lista e, em seguida, retirar os estabelecimentos que estão na lista dos setores nos quais estão associados, para que eles não façam parte do cadastro de áreas.

A seguir, o Sr. Marcos Paulo comentou que já havia sido construído um cadastro de lista preliminar, a partir das informações do Censo Agropecuário 2006 e de critérios estabelecidos pela COAGRO, para definir os grandes estabelecimentos (área total, empregados permanentes e ter CNPJ). Foram incluídos cerca de 300.00 estabelecimentos, depois de analisar as possíveis fontes de atualização. Chegou-se à conclusão de que, devido ao elevado número de estabelecimentos incluídos, seria muito custoso e difícil atualizá-los. Com isso, foram elaborados novos critérios de inclusão, levando em consideração as diferentes classes de atividade econômica. Para cada classe de atividade econômica foi estipulado um critério específico. Esta nova tentativa resultou em um cadastro de aproximadamente 115.00 estabelecimentos. Agora, o trabalho se encontra numa fase de avaliar e estabelecer efetivamente os critérios para os grandes estabelecimentos; verificar se o cadastro resultante será mais fácil de atualizar; e estabelecer critérios para que os estabelecimentos

que se dedicam a atividades muito específicas, também possam entrar no cadastro de lista. Estes estabelecimentos não são grandes em termos de área ou de produção, mas são concentrados no território e, portanto, difíceis de serem captados no cadastro de área.

Dando seguimento à apresentação, o técnico da COMEQ comentou sobre o que está sendo feito em relação à construção do cadastro de áreas. A primeira observação feita foi quanto à diferença de malhas setoriais. Enquanto os estudos do plano amostral estão sendo realizados usando o Censo Agropecuário 2006, que usou a malha setorial de 2006, o SNPA usará a malha setorial de 2010. A maior diferença entre as bases é a divisão de setores ocorrida entre a malha anterior e a posterior. Então, ele questionou: como compatibilizar as informações de 2006 na base de 2010? Para os estudos iniciais, pensou-se primeiramente em alocar proporcionalmente as informações, levando-se em conta o número de estabelecimentos agropecuários existente nos setores de 2010. Esta opção foi abandonada, devido ao número reduzido de estabelecimentos agropecuários encontrados no Censo Demográfico de 2010. Como o objetivo do Censo Demográfico é localizar domicílios, muitos estabelecimentos agropecuários deixaram de ser contabilizados. Uma outra opção seria utilizar a informação de coordenada geográfica, obtida no Censo Agropecuário 2006. De posse da informação posicional, verificaríamos em qual setor de 2010 o estabelecimento estaria alocado. Contudo, infelizmente, há um número razoável de estabelecimentos aonde não foi possível obter a coordenada em 2006. Para os estudos iniciais, acabou-se optando pela divisão proporcional da seguinte forma: se um setor em 2006 deu origem a três setores em 2010, as informações das variáveis de interesse foram alocadas em 33% para cada novo setor. O Sr. Marcos Paulo disse acreditar que a opção escolhida para a fase de estudos não tenha sido tão prejudicial. Entretanto, para o cadastro final, ele considera mais promissora a alternativa da coordenada geográfica. Além do mais, o técnico acenou com a possibilidade de utilizar uma outra maneira de fazer esta compatibilização. Segundo ele, tão importante quanto adequar as bases é proceder à “limpeza” do cadastro. Esta limpeza consiste em retirar os estabelecimentos grandes do setor onde ele está, ou seja, uma vez identificado um estabelecimento integrante do cadastro de lista, ele é retirado do setor para que suas informações não sejam consideradas, quando houver a seleção dos setores.

Ainda sobre o cadastro de áreas, o Sr. Marcos Paulo lembrou a questão do grande número de setores com reduzido número de estabelecimentos agropecuários. Portanto, deverá haver necessariamente agregação de setores, de forma que não se terá um cadastro de setores, mas sim um cadastro de grupo de setores. Esta agregação deverá estabelecer um número mínimo de estabelecimentos por grupo, levando em conta quantos estabelecimentos se terá que ter no grupo para que se possa retirar “n” amostras de “n” estabelecimentos, sem antes esgotá-lo, considerando que os estabelecimentos ficam cinco anos na amostra, e serão retiradas quantas amostras a cada

cinco anos. Para agregar os setores, a princípio, pretende-se utilizar os mesmos critérios que estão sendo utilizados para a divulgação de agregados, por grupos de setores, do Censo Agropecuário 2006. São agregados setores com características semelhantes e resguardando-se o sigilo. Pretende-se aproveitar esta agregação, mas levando em conta o tamanho mínimo da amostra ao invés do sigilo.

A seguir, o Sr. Marcos teceu comentários a respeito do plano amostral a ser empregado no cadastro de áreas, cujo modelo será o de conglomerado em dois estágios. O primeiro estágio é o agregado de setores censitários, e o segundo, os estabelecimentos agropecuários dentro do agregado. Antes de selecionar os agregados de setores será feita uma estratificação, para melhorar a qualidade da estimativa. Ainda se definirá qual estratificação usar: se as divisões existentes na UF, a estratificação usada para o SIPD ou uma outra estratificação mais adequada ao meio agropecuário. Cada agregado de setor será selecionado, com probabilidade proporcional ao tamanho. Ainda será decidida qual variável de tamanho empregar, sendo que já se testaram algumas: o valor da produção, que tem uma variabilidade muito grande; ou o número de estabelecimentos por setor. Contudo, como ainda não se procedeu à agregação dos setores, existe uma quantidade muito grande de setores com poucos estabelecimentos. Este aspecto interferiu muito nos estudos preliminares do tamanho da amostra. Dada a seleção de agrupamento de setores, faz-se a seleção de estabelecimentos no agregado de setores. Será uma amostra aleatória simples, com número fixo de estabelecimentos, algo entre 10 e 20 estabelecimentos, com tendência maior a se adotar 10 estabelecimentos, pois há uma grande homogeneidade no setor. Valeria mais a pena espalhar bem a amostra de agregados de setores. Contudo, ainda estão sendo feitos estudos e tudo dependerá do custo de cada amostra. Uma opção para melhorar a qualidade das estimativas seria fazer uma estratificação de estabelecimentos dentro do setor, de acordo com informações levantadas na PCADE, ou mesmo por atividade econômica. Entretanto, esta ideia surgiu recentemente e ainda carece de mais estudos.

Apesar dos cadastros (lista e área) ainda não estarem prontos, fez-se uma tentativa de dimensionar um tamanho de amostra primordial. Para tal, foi escolhida como variável de dimensionamento o valor total da produção, por ser uma variável importante e por aparentemente ter uma correlação grande com outras variáveis a serem investigadas na pesquisa. O domínio de interesse é a UF e, portanto, teremos um tamanho de amostra para cada uma. Foi estipulado um CV de 10% para o valor total da produção. Contudo, os trabalhos iniciais não deram bons resultados. Utilizando-se o próprio valor total da produção como variável de tamanho, para a seleção de setor com probabilidade proporcional ao tamanho, a amostra obtida foi muito pequena para estimar o valor total da produção. A princípio, um tamanho de amostra pequeno é desejável, contudo, quando

se usa como variável de tamanho a mesma variável que se quer estimar, o tamanho da amostra tende a ficar reduzido. Porém, a precisão das outras variáveis que se quer estimar foi muito ruim. O Sr. Marcos Paulo acrescentou que como o cadastro ainda não foi trabalhado, a variabilidade muito grande do valor total da produção interferiu neste resultado. Utilizando como variável de tamanho para a seleção de setores o número de estabelecimentos por setor, o tamanho de amostra obtido foi muito grande.

Por fim, o técnico da COMEQ lamentou não ter apresentado um tamanho de amostra inicial, mas reafirmou que o fato dos cadastros não estarem prontos influenciou muito os resultados. Inclusive, ele comentou que quando forem definidos e estipulados quais são os estabelecimentos que comporão o cadastro de lista, e depois que os setores com poucos estabelecimentos forem agrupados, novos cálculos serão feitos e acredita-se que os resultados serão bem diferentes. Acrescentando um comentário a sua apresentação, o Sr. Marcos citou que uma possibilidade que ainda merece estudo é a coincidência dos setores da amostra do SIPD e do SNPA.

Em seguida, a sessão foi aberta para perguntas/comentários. O Sr. Pedro Andrade (Supervisor do Piauí) reforçou a questão da grande diferença existente entre os estabelecimentos grandes e pequenos. Ilustrando, ele citou o caso do seu Estado. Ele disse que no Piauí existem 243.000 estabelecimentos agropecuários e que aproximadamente 300 estabelecimentos são responsáveis por 97% da produção de grãos (soja, milho, arroz e feijão). Comentando, o Sr. Marcos Paulo salientou que a ideia de tratar diferentemente os estabelecimentos, como se eles constituíssem populações distintas, irá fazer com que os 300 estabelecimentos citados, fiquem no cadastro de lista e sejam amostrados de uma forma. Já os outros 242.700 pequenos estarão no cadastro de área, sendo amostrados de outra maneira. Posteriormente, o Sr. Jonatas Picanço (representante do Amazonas) perguntou como seriam pesquisadas as áreas indígenas. O Sr. Flávio Bolliger (COAGRO) respondeu à indagação, dizendo que o SNPA não pesquisaria muitos tipos de estabelecimentos que foram levantados no Censo Agropecuário. Citou que os estabelecimentos sem área não fariam parte do público-alvo do SNPA. Certamente, segundo ele, as áreas indígenas também serão excluídas. O Coordenador da COAGRO agradeceu a intervenção do Sr. Jonatas, salientando que esta exclusão deverá constar na metodologia do SNPA. A seguir, o próprio Sr. Flávio perguntou o que significava “truncamento”. O Sr. Marcos Paulo explicou que, quando se utilizou o número de estabelecimentos do setor para ser a variável de tamanho, geralmente se define um valor máximo que esta variável deva ter. A isto denomina-se “truncamento”. Isto é feito para que a probabilidade de seleção não tenha amplitude muito grande. Ele ainda disse que normalmente se usa que o limite máximo seja 20 vezes o limite mínimo. Então, todos setores que tem um valor acima do máximo, recebem o valor máximo. Contudo, ele espera que nos setores agropecuários não

seja necessário “truncar” nenhum setor. Em seguida, o Sr. Jorge Elarrat (Supervisor de Rondônia) perguntou como seria a rotação da amostra de setores do SNPA. O Sr. Marcos Paulo disse que mais estudos serão feitos, pois ainda há muitos detalhes a abordar. Colocou que inicialmente se pensava numa rotação anual de 10%, isto é, a cada ano, 10% dos setores seriam substituídos, entrando novos setores na amostra. Assim, a amostra após 10 anos seria totalmente diferente daquela que iniciou o ciclo. Na sequência, o Sr. Pedro Andrade comentou que no Piauí estava havendo uma expansão agrícola muito acentuada. Ele questionou como novas áreas entram na pesquisa. O técnico da COMEQ acrescentou que a ideia de incluir a cada ano novos setores na amostra, bem como fazer rotação de setores, é para possibilitar captar estas mudanças conjunturais. O Sr. Flávio Bolliger externou sua preocupação quanto à agregação dos setores, pois há um contingente muito grande de setores aonde há poucos estabelecimentos. Segundo ele, o quantitativo chega a 25% e, ao serem agregados, se teme que fisicamente a área da agregação fique elevada. Este aspecto dificultaria muito a operação de coleta. Portanto, o Sr. Flávio realçou que devemos ter isto em mente quando formos realizar a agregação. Encerrando a sessão, o Coordenador da COAGRO agradeceu a presença do Sr. Marcos Paulo, parabenizando-o pela excelente explanação.