

Pesquisa por Amostragem: Política de Divulgação de Estimativas com Baixa Precisão Amostral

Nádia Pinheiro Dini¹

Resumo: Nesse artigo, apresentam-se vários fatores que podem ser considerados ao se divulgar estimativas com baixa precisão amostral em pesquisas por amostragem. Também, são apresentados alguns critérios de divulgação de estimativas de baixa precisão nas pesquisas amostrais realizadas pela Fundação Seade.

Introdução

As pesquisas amostrais, em geral, têm objetivos amplos e investigam diversos aspectos do fenômeno de interesse. Como reflexo dessa amplitude, tem-se que os questionários aplicados nesse tipo de pesquisa tendem a ser extensos, englobando elevado número de variáveis. Além dos dados coletados diretamente pelo instrumento da pesquisa, é comum que vários outros indicadores sejam gerados a partir deles, perfazendo, assim, um vasto conjunto de informações provenientes de uma única pesquisa.

Contudo, o tamanho da amostra normalmente é determinado fixando-se apenas a precisão desejada para alguns poucos estimadores. Como a gama de indicadores possíveis de ser calculados e divulgados e para os quais não se fixou a priori a precisão desejada costuma ser bastante elevada, os responsáveis pelas pesquisas amostrais deparam-se, então, com o seguinte problema: deve-se divulgar estimativas com baixa precisão amostral?

A última pergunta remete à outra questão – o que é baixa precisão amostral?

Como tratar a questão da precisão amostral nas diversas formas de se divulgar os resultados de uma pesquisa?

¹ Estatística - Gerente de Métodos Quantitativos da Diretoria Adjunta de Produção e Análise de Dados da Fundação Seade.

Precisão Amostral

Na etapa de planejamento da amostra a ser utilizada na pesquisa, quando se trata da questão da precisão amostral desejada, normalmente determina-se que o coeficiente de variação dos principais estimadores a serem divulgados seja no máximo 5%, obtendo-se, assim, o tamanho necessário de amostra. É um valor arbitrário que, portanto, pode ser alterado dependendo dos interesses envolvidos.

Entretanto, qualquer que seja o patamar estipulado, como já foi mencionado anteriormente, quase sempre existem mais indicadores passíveis de divulgação, além dos que se utilizaram para determinação do tamanho da amostra. Ao se calcular o coeficiente de variação para estes indicadores, podem-se obter valores abaixo ou acima do limite estabelecido. Cabe então definir que patamar máximo de coeficiente de variação se deseja para os indicadores principais da pesquisa e que atitude tomar em relação aos demais indicadores, quando sua precisão for inferior à desejada.

Essa decisão precisa ser tomada levando-se em consideração alguns fatores importantes, tais como: a periodicidade do levantamento e a própria natureza do fenômeno em estudo. A seguir, discute-se a influência desses aspectos na política de divulgação das estimativas.

- **Periodicidade do levantamento** – em geral, exige-se uma precisão maior dos indicadores obtidos a partir de levantamentos regulares, principalmente os mensais. Isso porque nesses casos é de grande interesse não apenas o patamar dos indicadores em si, mas também sua comparação com os valores obtidos em outros momentos da pesquisa. Se o coeficiente de variação dos indicadores for muito elevado, fica comprometida a possibilidade de se realizar comparações conclusivas. Já quando a pesquisa é pontual ou realizada em intervalos mais longos de tempo, tende-se a tolerar uma menor precisão das informações.
- **Natureza do fenômeno em estudo** – ao se determinar a precisão desejada para a divulgação de indicadores, não se deve ignorar o importante papel que sua própria natureza desempenha em relação à magnitude dos coeficientes de variação. Frequentemente se observa que alguns fenômenos apresentam tamanha variabilidade que a exigência de elevados patamares de precisão amostral praticamente inviabilizaria a divulgação de quaisquer indicadores sobre eles, a menos que se pudessem levantar amostras excepcionalmente grandes.

Formas de divulgação dos resultados da pesquisa

As pesquisas podem ser divulgadas através de textos analíticos, conjunto de tabelas e disponibilização dos seus microdados. Em todas as formas de divulgação deve-se manter a qualidade das informações no sentido de não serem divulgados indicadores com baixa precisão amostral, além de preservar o sigilo das informações.

Nos casos de indicadores de baixa precisão, conforme a forma de divulgação, pode-se optar por:

- omitir esses resultados;
- apresentá-los com advertências;
- disponibilizar tabelas com os erros amostrais para o usuário tomar a decisão ou com a apresentação dos erros ou alguns coeficientes de variação calculados através de modelagem.
- divulgar alguns indicadores com o uso de modelagem que em geral reduz os erros amostrais

No caso dos textos analíticos, ao se divulgar informações desta forma, é comum que se realize uma série de afirmações sobre os dados. A fim de se evitar que tais conclusões sejam equivocadas convém que se trabalhe apenas com indicadores de precisão mais elevada, preservando assim a qualidade do produto.

No conjunto de tabelas, pode-se optar por omitir os indicadores de baixa precisão ou apresentar todos os resultados. Se a segunda opção for adotada, deve-se indicar de alguma forma que o coeficiente de variação da variável está acima de um limite estipulado ou apresentar um conjunto de tabelas análogo contendo os coeficientes de variação de todos os resultados divulgados.

Quando se disponibilizam os microdados de uma pesquisa, deve-se também fornecer os erros amostrais associados aos principais indicadores ou pelo menos as informações e a orientação necessárias para que o usuário esteja ciente da questão da precisão amostral e possa calculá-la para qualquer indicador que lhe interessar.

Métodos de Cálculo dos Erros Amostrais

Em pesquisas amostrais de desenho complexo, o cálculo dos erros amostrais não é trivial e até recentemente não estava disponível nos principais *softwares* estatísticos. Devido a essa dificuldade, em muitos levantamentos, principalmente os de caráter contínuo, optou-se por não efetuar o cálculo de todos os erros amostrais, mas sim de se desenvolver modelos em que se estabeleça a relação entre a estimativa obtida e o seu erro amostral, no caso de microdados. Dessa forma, sem a necessidade de se calcular o coeficiente de variação de todos os resultados obtidos, pode-se apresentar quais possuem a precisão desejada e quais estão aquém dela. Todos os usuários da pesquisa devem ser informados sobre esse modelo e ser orientados a não utilizarem as estimativas que apresentarem precisão abaixo da estipulada para a pesquisa ou, caso decidam fazê-lo assim mesmo, de estar cientes da cautela que é necessária na análise dessas informações.

Atualmente, o avanço dos *softwares* mais utilizados no processamento de dados amostrais vem possibilitando que se calculem não apenas as estimativas, mas também o erro amostral a elas associado de maneira relativamente simples. Com isso, passa a ser viável a opção de se divulgar todas as informações acompanhadas de seu coeficiente de variação, mesmo em pesquisas obtidas através de desenhos amostrais complexos.

Política de Divulgação de Estimativas de Baixa Precisão do Seade

A Fundação Seade não adota uma política única para a divulgação dos resultados das pesquisas que realiza, mas orienta sua decisão de acordo com as questões acima mencionadas. A seguir, apresenta-se a política de divulgação de estimativas de algumas das pesquisas amostrais realizadas pelo Seade, bem como os principais motivos que levaram à sua adoção.

PED - Pesquisa de Emprego e Desemprego: esta pesquisa, realizada mensalmente em cerca de 3.000 domicílios da Região Metropolitana de São Paulo desde 1984, teve sua amostra determinada de modo a garantir um coeficiente de variação de no máximo 3,5% para a taxa de desemprego total, seu principal indicador. Para os demais indicadores disponibilizados, fixou-se em 7,5% o coeficiente de variação máximo aceitável. Nas tabulações, os indicadores com precisão inferior ao limite estabelecido não são divulgados, e, em seu lugar, apresenta-se um símbolo com uma nota explicativa ressaltando-se o motivo para a exclusão da informação. Foi possível

modelar o número mínimo de observações que se deve ter na amostra a fim de que a estimativa resultante tenha um coeficiente de variação de até 7,5%. Os microdados da pesquisa são disponibilizados, incluindo-se as informações necessárias para o cálculo dos erros amostrais, mas não os erros em si, procedimento que se pretende adotar em futuro próximo.

PCV - Pesquisa de Condições de Vida: esta pesquisa é domiciliar e foi realizada em 1990 (na Região Metropolitana de São Paulo), 1994 (nos municípios do Estado de São Paulo de 80.000 habitantes ou mais) e em 1998 (incluindo os municípios com 50.000 habitantes ou mais). Nas tabulações, disponibilizam-se todas as estimativas com coeficiente de variação inferior a 7,5%. Aquelas com coeficiente acima de 7,5%, mas abaixo de 15%, são apresentadas, porém de forma discriminada (entre parênteses) e não se divulgam as estimativas cujo coeficiente de variação ultrapasse 15%. Nesses casos, apresenta-se um símbolo e uma nota explicativa. Os microdados estão disponíveis e, assim como na PED, incluem as informações necessárias para o cálculo dos erros amostrais, mas não os erros em si. Nas próximas edições dessa pesquisa, pretende-se apresentar tabulações em que as estimativas sejam acompanhadas dos respectivos coeficientes de variação. As tabulações cuja maioria das categorias apresentarem precisão inferior à desejada serão descartadas em sua totalidade.

PAEP - Pesquisa da Atividade Econômica Paulista: levantamento realizado em empresas, de forma censitária em alguns estratos e por amostragem em outros, dependendo do porte e do setor de atividade da empresa. Os microdados podem ser acessados por meio do *site* do Seade na Internet, sendo possível solicitar processamentos, para os quais são fornecidos também os coeficientes de variação. Adotou-se uma representação coroplética para estes coeficientes, tomando-se como referência o patamar de 30%, acima do qual utiliza-se o vermelho como sinal de alerta de que se trata de uma informação com baixa precisão. Todos os indicadores são, dessa forma, divulgados, a menos que comprometam o sigilo necessário da pesquisa.

PECOMPE - Pesquisa de Conjuntura das Micro e Pequenas Empresas do Estado de São Paulo: esta pesquisa, realizada para o SEBRAE-SP, levanta mensalmente informações sobre um painel de aproximadamente 2.700 micro e pequenas empresas do Estado de São Paulo. Sua amostra foi determinada a fim de garantir um coeficiente de variação de no máximo 12,5% para o seu principal indicador, o faturamento

mensal, em recortes regionais e setoriais. Esta pesquisa constitui um bom exemplo de como a natureza do indicador em estudo influencia a precisão estipulada, pois o faturamento das micro e pequenas empresas apresenta variabilidade tal que mesmo com uma amostra estratificada de tamanho bastante razoável não é possível garantir um coeficiente de variação de menos de dois dígitos.

Considerações Finais

As políticas de divulgação mencionadas apresentam algumas vantagens e desvantagens, que serão discutidas a seguir.

- Não divulgar os resultados de baixa precisão – essa opção tem a vantagem de garantir de fato a qualidade da pesquisa, uma vez que todos os dados divulgados atendem à exigência de precisão estipulada, não sendo necessárias ressalvas ou cuidados adicionais no uso da informação. Também contribui para um melhor entendimento das pessoas não especializadas na área do que vem a ser estimativas e quais as limitações que apresentam. Por outro lado, é comum que alguns usuários confundam o fato de um resultado não ser apresentado por ter um erro amostral muito elevado com a idéia de que a estimativa em si é insignificante e por isso não foi divulgada.
- Divulgar os resultados de baixa precisão de maneira discriminada – tem a vantagem de possibilitar ao usuário criterioso que ele tenha uma idéia da magnitude de algumas questões, estando ciente, entretanto, de que não deve fazer afirmações categóricas a respeito. Entre usuários menos experientes, é possível que essa distinção entre as estimativas de precisão adequada e aquelas de baixa precisão passe despercebida e que algumas análises e conclusões indevidas sejam realizadas.
- Divulgar todos os resultados sem discriminação, mas acompanhados do erro amostral – tal como na opção anterior, o bom desempenho dessa alternativa depende do grau de consciência dos usuários da pesquisa. Essa política exige que se disponha de métodos ágeis de cálculo dos erros amostrais. O excesso de informações apresentadas pode ter efeitos desagradáveis ou, se a opção for por apresentar os erros amostrais separadamente, pode-se incorrer em alguma dificuldade de consulta e de associação das informações.