



Instituto Brasileiro de Geografia e Estatística

*II Encontro Nacional de Produtores e
Usuários de Informações Sociais,
Econômicas e Territoriais
21 a 25 de agosto de 2006*

*V Conferência Nacional de Estatística
CONFEST*

*Pesquisas por amostragem:
política de divulgação de estimativas
com baixa precisão amostral*

Sonia Albieri, Diretoria de Pesquisas

Roteiro

- Introdução
- O início das discussões
- A situação atual
- As questões de qualidade na produção de estatísticas oficiais
- Políticas praticadas em outros institutos oficiais de estatística – alguns exemplos
- O debate

Introdução

- O IBGE realiza um conjunto bastante grande de pesquisas por amostragem probabilística, tanto na área sócio-demográfica, investigando unidades domiciliares e seus moradores, como na área econômica, cujas unidades de investigação são as empresas da indústria, do comércio e de serviços.

Introdução

- Todas essas pesquisas passam por um processo de avaliação da precisão de suas estimativas.
- Ou seja, são calculadas estimativas das variâncias, erros padrão ou coeficientes de variação associados às estimativas do plano tabular de divulgação dos resultados.
- Em algumas pesquisas, esse cálculo é feito para todas as estimativas do plano de divulgação.

Formas de Apresentação dos Erros Amostrais

- CV ou erro padrão apresentado para cada estimativa \Rightarrow duplica plano tabular
- Ajuste de modelo de regressão
 - CV x estimativa para variáveis categóricas
 - divulgação dos parâmetros do ajuste
- Indicador do intervalo do CV para cada estimativa

O início das discussões

- Em 1999, a área de metodologia (atual Coordenação de Métodos e Qualidade) sugeriu que fosse definida uma forma padrão de apresentar a precisão de estimativas para todas as estimativas divulgadas.
- Mas, não houve consenso em torno da proposta apresentada e, assim, a variedade de formas de apresentar medidas de precisão permanece.

Proposta para Intervalos de CV preparada em 1999

Indicador	Intervalo de CV(%)	Diretriz
Z	Zero	Sem erro amostral
A	0,0 - 0,5	Publicável
B	0,6 - 1,0	Publicável
C	1,1 - 2,5	Publicável
D	2,6 - 5,0	Publicável
E	5,1 - 7,5	Publicável
F	7,6 - 10,0	Publicável
G	10,1 - 15,0	Publicável
H	15,1 - 25,0	Qualificado
I	25,1 - 35,0	Restrito
J	35,1 - 50,0	Restrito
K	mais de 50,0	Não publicável

Proposta para Intervalos de CV e diretriz

- Indicador **H**: {15,1 a 25,0%}
 - qualificado: indica que a estimativa deve ser divulgada acompanhada de alguma explicação sobre os motivos e o significado de sua precisão.
- Indicadores **I** e **J**: {25,1 a 50,0%}
 - restrito: indica estimativas que devem ser divulgadas de forma cuidadosa dependendo do tipo de uso
 - alerta para o fato de se referirem a estimativas com baixa precisão

Proposta para Intervalos de CV e diretriz

- Indicadores **K**: {mais de 50,0%}
 - não publicável: indica que a estimativa é muito imprecisa, justificando que não seja publicada
 - a proposta era substituir o valor da estimativa pelo seu indicador de precisão (K) ou por algum sinal convencional

Proposta para uma diretriz

- Essa proposta (de supressão do valor da estimativa caso o CV associado apresente um valor alto, indicando que essa é uma estimativa muito imprecisa ou não confiável) causou bastante discussão
- Conseqüência: algumas pesquisas passaram a adotar o quadro simplificado de letras indicadoras de intervalos de CVs, na apresentação da precisão das estimativas.

Classificação das Estimativas Quanto à Precisão em uso

Indicador	Intervalo de cv(%)	Conceito
Z	zero	“Exata”
A	até 5	Ótima
B	mais de 5 a 15	Boa
C	mais de 15 a 30	Razoável
D	mais de 30 a 50	Pouco precisa
E	mais de 50	Imprecisa

Situação atual

- A discussão que permanece é a que está relacionada com a decisão de publicar ou não estimativas que possuem baixa precisão.
- **A favor da publicação:** alguns produtores e usuários preferem que seja divulgado o valor de uma estimativa com a indicação de que possui alta variabilidade devido à amostragem e deixar a decisão sobre usar ou não a estimativa para o analista temático.

Situação atual

- **Contra a publicação:** outros preferem suprimir essas estimativas com baixa precisão, substituindo-as por um símbolo que indique tratar-se de uma estimativa não apropriada para uso.
- Esses últimos baseiam seus argumentos na necessidade de garantir a **qualidade** e manter a **credibilidade** da instituição.

As questões de qualidade

- como definir qualidade?
 - ⇒ Juran e Gryna, 1980, propuseram um a definição bem geral e abrangente:
 - ⇒ qualidade pode ser definida como adequação para o uso (ou adequação para os propósitos)
 - ⇒ *fitness for use or fitness for purpose*
- ⇒ variedade de usos implica que a adequação para o uso deve ter múltiplas características, que variam de importância entre os usuários.

As questões de qualidade

- adequação para o uso, traduzido no contexto de uma pesquisa por amostra:
 - ⇒ 1. os dados devem ser tão precisos quanto seja necessário para atender aos usos.
 - ⇒ Ou seja: Acurácia é uma das dimensões da qualidade em pesquisas por amostra

A qualidade em institutos de estatística

- qualidade de um produto estatístico é um conceito multidimensional:
 - ⇒ acurácia, oportunidade, riqueza de detalhes, acessibilidade, nível de proteção da confidencialidade, etc.
 - ⇒ há vários exemplos de marcos de referência para qualidade usados em estatísticas oficiais
 - ⇒ todos com ênfase em acurácia

Acurácia

- A acurácia é definida como a proximidade entre a estimação de um valor e seu verdadeiro valor (não conhecido).
- A acurácia também é considerada como uma medida do erro total, que inclui:
 - Erro de amostragem.
 - Erro de não amostragem [cobertura, medida, elaboração, não resposta (total e parcial) e suposição do modelo].

Políticas praticadas em outros institutos oficiais de estatística

- Statistics Canada
- Statistics New Zealand
- ABS - Australian Bureau of Statistics
- ONS – Office for National Statistics - UK

Statistics Canada (1)

Designação Alfabética do cv

Indicador	Intervalo de CV (%)	Diretriz
A	0 - 0,5	publicável
B	0,6 - 1,0	publicável
C	1,1 - 2,5	publicável
D	2,6 - 5,0	publicável
E	5,1 - 10,0	publicável
F	10,1 - 16,5	publicável
G	16,6 - 25,0	qualificado
H	25,1 - 33,3	restrito
I	33,4 ou mais	não divulgável

Fonte: Survey Skills Development Survey - Statistics Canada

Statistics Canada (2)

- Símbolos usados em publicações do Statistics Canada
 - Além dos tradicionais símbolos para indicar arredondamentos, dados preliminares, dados revisados, dados não disponíveis, dados suprimidos em função de critérios de confidencialidade, dois símbolos relacionados com a questão da precisão das estimativas:
 - E -- usar com cuidado
 - F -- muito impreciso para ser publicado

Statistics New Zealand (1)

- Protocols for Official Statistics - Statistics New Zealand - August 1998
 - É preciso ter cuidado ao divulgar dados sujeitos a erro.
 - Em mais detalhes:
 - dados com vícios sérios, que podem dar impressão errada ou ser mal interpretados pelos usuários, não devem ser publicados.
 - estimativas em células individuais de uma tabela publicada não devem ser suprimidas somente porque estão sujeitas a erros amostrais grandes, desde que os usuários sejam advertidos sobre a falta de confiança dos dados.

Statistics New Zealand (2)

- tabelas não publicadas em um determinado nível de detalhe, porque a maior parte das estimativas teria grande imprecisão, em termos de erros amostrais, se disponibilizadas para usuários precisam ter indicações apropriadas dos erros de amostragem.
- se tratando de dados que não foram publicados, por estarem sujeitos a sérios vícios, a advertência deve ser maior ainda quando disponibilizados para usuários.
- todas as publicações e divulgações (releases) devem incluir um tópico sobre a metodologia e qualidade dos dados. Em muitos casos esse tópico será a orientação sobre o local onde poderá ser encontrada a informação abrangente.

ABS - Australian Bureau of Statistics (1)

- Possui uma Política de Divulgação de Dados Sujeitos a Erros, desde 1988
- Mesmos critérios da Nova Zelândia, porém com as seguintes recomendações:
- se mais de 1/3 das células de uma tabela estiverem sujeitas a erros amostrais de 25% ou mais, as áreas responsáveis são orientadas a agregar linhas ou colunas (combinando classificações) ou a eliminar a tabela da publicação.
- se os usuários invocarem a lei de liberdade de informação, os erros amostrais devem ser disponibilizados junto com os dados.

ABS - Australian Bureau of Statistics (2)

- Em resumo:
- dados sujeitos a erros de amostragem grandes podem ser disponibilizados para os usuários, mas nunca publicados pelo instituto.
- A utilização consciente desses dados é responsabilidade do usuário.
- O instituto não pode ter essa responsabilidade, pois tem a obrigação de colocar um selo de qualidade nas informações que publica.

ONS/UK – Reino Unido

- Protocol on Release Practices – 2002
- QM&R - Quality Measurement and Reporting at the ONS (UK)
 - Indica que ainda não concluíram se será possível adotar um padrão único e ótimo de classes de intervalos de cv pra comunicar a qualidade de estimativas em função da precisão.
 - aponta que o GSS – Government Statistics Service faz uso de um esquema dessa forma: cada categoria de precisão é classificada com um símbolo ou letra para indicar que a estimativa é aceitável, deve ser usada com cautela, ou imprecisa.
 - cita o Statistics Canada como um instituto que tem um padrão.

Debate

- O objetivo aqui é:
- discutir sobre a **utilidade** de uma estimativa com baixa precisão, ou seja, com erro amostral estimado muito grande.
- discutir a questão no contexto da disseminação de estatísticas oficiais, com vistas à definição de uma política de disseminação de informações por parte do IBGE.
- principalmente a questão dos **alertas** junto com estimativas com baixa precisão e da **supressão** de valor por símbolo que indica esse fato, qual seja, estimativa muito imprecisa para ser publicada com a chancela do instituto.

Bibliografia

- Juran, J.M.e Gryna,Jr.F.M. *Quality Planning and Analysis*, 2nd ed., New York, McGraw-Hill, 1980.
- Biemer, Paul.P e Lyberg, Lars. E. *Introduction to Survey Quality*, Wiley, 2003.
- Albieri, Sonia. *Apresentação da precisão de estimativas nas tabelas de pesquisas por amostragem*. IBGE: Rio de Janeiro, 1999. [documento interno]
- Albieri, Sonia e Silva, Ari N. (2001). *Índice de Qualidade de Tabelas, Uma Avaliação do Plano Tabular de Pesquisas por Amostragem baseado na Precisão das Estimativas*. Rio de Janeiro, IBGE. 4p. [apresentado em IBGE/IASI - Seminário Internacional sobre Metodologias para Pesquisas Domiciliares por Amostragem. 2001. Rio de Janeiro - Brasil].
- <http://unstats.um.org/unsd/bpform.asp?>
- <http://www.statcan.ca/english/freepub/63-008-XIE/00905/user.htm>