

Capítulo 14 Estimadores de Calibração

14.1 Motivação: melhorar a precisão com informações auxiliares

Em Amostragem, há várias maneiras de aproveitar *informações auxiliares* disponíveis para melhorar a precisão de estimativas de interesse em uma pesquisa amostral. As duas maneiras mais conhecidas são: *estratificação* e *amostragem com probabilidades proporcionais ao tamanho (PPT)*. Nos dois casos, a informação auxiliar é utilizada durante a etapa de planejamento amostral, portanto deve estar disponível *antes da seleção da amostra* para a realização de uma pesquisa.

Nem sempre se consegue aproveitar toda a informação auxiliar existente nesta etapa da pesquisa. Por exemplo, quando existem muitas variáveis auxiliares, pode não ser possível usar todas na estratificação. Por outro lado, num plano PPT, só podemos usar uma *medida de tamanho* para definir as *probabilidades de inclusão*. Em consequência, pode não ser possível obter ganhos de eficiência para algumas variáveis de pesquisa.

Uma ideia é considerar o uso da informação auxiliar também na *etapa de estimação da pesquisa*, usando *estimação de calibração*.

Este capítulo aborda *quando* usar estimadores de calibração; *estimadores de calibração*; *estimadores de regressão como calibração*; *porque* calibrar; *como* fazer calibração clássica; *quanta* calibração é necessária; *indicadores de qualidade* da calibração. Apresenta uma revisão da literatura sobre métodos de calibração usados na ponderação e estimação de pesquisas por amostra, aponta referências bibliográficas relevantes, discute questões importantes que surgem quando os métodos de calibração são usados em situações reais de pesquisa e aponta critérios que podem ser usados para avaliar se sua utilização foi bem sucedida ou se ocorreram dificuldades que demandem revisão dos resultados.

14.2 O problema

Seja uma *população finita* U de *tamanho* N , que pode ser representada pelo conjunto de rótulos $1, 2, \dots, k, \dots, N$.

Seja $\mathbf{y}_k = (y_{k1}, y_{k2}, \dots, y_{kJ})'$ um vetor com os valores de J *variáveis de pesquisa* para a unidade $k \in U$.

Uma *amostra aleatória* $s \subset U$ é selecionada, e os valores de \mathbf{y}_k para $k \in s$ são observados para fazer inferência sobre alguns *parâmetros populacionais* de interesse.

Suponha por hora que é possível observar as *respostas* \mathbf{y}_k para *todas* as unidades da amostra selecionada (resposta completa).

O *alvo de inferência* que vamos considerar inicialmente é o *total populacional* definido como:

$$\mathbf{T}_y = \sum_{k \in U} \mathbf{y}_k$$

Um estimador ‘padrão’ para este total é o estimador *HT* (Horvitz e Thompson (1952)) definido por:

$$\hat{\mathbf{T}}_y = \sum_{k \in s} d_k \mathbf{y}_k$$

onde d_k são *pesos amostrais do desenho* dados por $d_k = \pi_k^{-1}$, sendo $\pi_k = P(k \in s)$ a *probabilidade de inclusão* da unidade k na amostra s .

Seja $\mathbf{x}_k = (x_{k1}, x_{k2}, \dots, x_{kZ})'$ um vetor com os valores de Z *variáveis auxiliares* para a unidade k .

Quando os *totais populacionais* das variáveis auxiliares \mathbf{T}_x são *conhecidos exatamente*, isto é, quando sabemos o valor de $\mathbf{T}_x = \sum_{k \in U} \mathbf{x}_k$ podemos usar esta informação para obter estimadores mais eficientes que $\hat{\mathbf{T}}_y$ para o total das variáveis de pesquisa \mathbf{T}_y .

Trata-se do uso de informações auxiliares na etapa de estimação da pesquisa.

O estimador *HT* para o total das variáveis auxiliares dado por $\hat{\mathbf{T}}_x = \sum_{k \in s} d_k \mathbf{x}_k$ vai diferir do total populacional conhecido $\mathbf{T}_x = \sum_{k \in U} \mathbf{x}_k$. Esta diferença é o chamado *erro de calibração*.

Na *estimação de calibração* a ideia básica é que as estimativas amostrais do total populacional das variáveis auxiliares podem ser *calibradas* para igualarem o total populacional conhecido \mathbf{T}_x , de tal forma que o *erro de calibração* seja eliminado.

Isto pode ser feito, modificando os pesos amostrais d_k usados no estimador Horvitz-Thompson (*HT*) para ‘expandir’ a amostra, usando um *estimador calibrado* tal que os *novos pesos* (w_k) permitam reproduzir exatamente os totais populacionais conhecidos, definido por:

$$\hat{\mathbf{T}}_x^C = \sum_{k \in s} w_k \mathbf{x}_k = \mathbf{T}_x$$

onde $w_k, k \in s$, são os *pesos calibrados* satisfazendo:

$$\hat{\mathbf{T}}_x^C - \mathbf{T}_x = \sum_{k \in s} w_k \mathbf{x}_k - \mathbf{T}_x = 0$$

As condições da Equação (14.4) são chamadas de *restrições de calibração*. A ideia é que, se os *pesos calibrados* $w_k, k \in s$, tiverem sucesso em reduzir ou evitar erros ao ‘estimar’ os totais x , eles também podem reduzir o erro ao estimar os totais y , usando o estimador de calibração:

$$\hat{\mathbf{T}}_y^C = \sum_{k \in s} w_k \mathbf{y}_k$$

Os pesos calibrados $w_k, k \in s$, podem depender de todas as informações disponíveis sobre as variáveis auxiliares x , mas não das variáveis de pesquisa y . Se este for o caso, então $\hat{\mathbf{T}}_y^C$ é um estimador linear de \mathbf{T}_y .

Neste capítulo, nos concentramos nos estimadores de *calibração para total* da forma da Equação (14.5), ou seja, estimadores lineares definidos por conjuntos de pesos $w_k, k \in s$, satisfazendo as *restrições de calibração para totais* dadas pela Equação (14.4).

Outras formas de restrições de calibração podem ser consideradas, como calibração para momentos de ordem mais alta ou até mesmo para a função de distribuição de população finita das variáveis auxiliares - veja a discussão na seção 10 em Chambers (1997).

Entretanto, essas outras formas de estimadores de calibração não devem ser consideradas aqui e, por simplicidade, seguiremos a denominação simples predominante dos estimadores definidos por (14.5) com pesos que satisfazem a Equação (14.4) como *estimadores de calibração*.

Um grande número de conjuntos de pesos $\{w_k, k \in s\}$, pode satisfazer as restrições de calibração. Uma maneira de selecionar aqueles que levam a conjuntos de pesos ‘razoáveis’ a serem usados para estimar totais para as variáveis y é pensar nos pesos de calibração w_k como modificações nos pesos do desenho d_k que *não diferem muito* dos pesos originais. Isso se justifica porque o uso dos pesos originais de desenho d_k dado pelo estimador *HT* assegura propriedades desejáveis, como *estimação não viciada e consistente* do total (no sentido de que à medida que o tamanho da amostra aumenta, o estimador converge em probabilidade em direção ao alvo correto \mathbf{T}_y).

14.3 Estimadores de regressão como estimadores de calibração

Em Särndal, Swensson, e Wretman (1992) foi definida uma família de estimadores de calibração 'clássica' para T_y , onde os pesos w_k são escolhidos de tal forma que as funções de distância especificadas que medem a distância entre o w_k e a d_k são minimizadas.

A ideia é resolver o problema de encontrar *novos pesos* $\{w_k, k \in s\}$ que *minimizem a distância* para os pesos originais:

$$D(\mathbf{w}; \mathbf{d}) = \sum_{k \in s} G_k(w_k, d_k)$$

tal que os totais estimados das variáveis auxiliares com os novos pesos w_k satisfaçam as *restrições de calibração* definidas em (14.4), isto é, tal que não ocorra *erro de calibração*.

Os *estimadores de regressão como estimadores de calibração* são definidos quando:

$$G_k(w_k, d_k) = \frac{(w_k - d_k)^2}{q_k d_k} \quad \forall k \in s$$

para algumas constantes conhecidas $q_k > 0$, $k \in s$, então a *solução* do problema é dada por:

$$w_k = d_k \times g_k$$

onde:

$$g_k = 1 + (\mathbf{T}_x - \hat{\mathbf{T}}_x)' \left(\sum_{i \in s} d_i q_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} q_k \mathbf{x}_k$$

Os pesos dados por (14.8) e (14.9) são exatamente os obtidos quando se emprega um *estimador de regressão generalizado* (GREG - 'Generalized Regression') motivado pelo seguinte *modelo de superpopulação*:

$$M \left\{ \begin{array}{l} y_{kj} = \mathbf{x}_k' \mathbf{B}_j + E_{kj} \\ V_M(E_{kj}) = \sigma_j^2 / q_k \end{array} \right. \quad k \in U, j = 1, \dots, J$$

com os coeficientes de regressão da população definidos por:

$$\mathbf{B}_j = \left(\sum_{k \in U} q_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \left(\sum_{k \in U} q_k \mathbf{x}_k y_{kj} \right)$$

e E_{kj} como sendo o resíduo do modelo de regressão da população para a variável de pesquisa y_j .

Sob o modelo definido em (14.10), o *estimador GREG* para o total da variável y_j , de acordo com Särndal, Swensson, e Wretman (1992), p. 234, pode ser escrito como:

$$\hat{T}_{y_j}^C = \sum_{k \in s} w_k y_{kj} = \sum_{k \in s} d_k g_k y_{kj} = \hat{T}_{y_j} + (\mathbf{T}_x - \hat{\mathbf{T}}_x)' \hat{\mathbf{B}}_j$$

onde os *coeficientes do modelo* são estimados por:

$$\hat{\mathbf{B}}_j = \left(\sum_{k \in s} d_k q_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \left(\sum_{k \in s} d_k q_k \mathbf{x}_k y_{kj} \right)$$

Uma questão prática importante é a *escolha das constantes* $\{q_k, k \in s\}$ que serão usadas para obter os pesos $\{w_k, k \in s\}$.

Na prática, é comum adotar o *mesmo conjunto de constantes* $\{q_k, k \in s\}$ *para todas as variáveis de pesquisa*, de modo a ter um *único conjunto de pesos* $\{w_k, k \in s\}$ para cada pesquisa.

Se os pesos $\{w_k, k \in s\}$ propostos forem usados para *todas* as variáveis de pesquisa, então as mesmas constantes $\{q_k, k \in s\}$ serão usadas para todas as variáveis.

Em muitas aplicações, isso não seria um problema, já que uma escolha comum é tornar todas essas constantes iguais, ou seja, $q_k = 1 \forall k \in s$.

Em alguns casos, as diferentes variáveis y podem ter resíduos da regressão linear da população nas variáveis auxiliares que mostram diferentes padrões de heteroscedasticidade. Nesses casos, para os diferentes conjuntos de valores necessários para as constantes q_k representarem adequadamente esses padrões podem levar a diferentes conjuntos de pesos calibrados, para cada conjunto específico de uma ou mais variáveis de pesquisa. Por um lado, usar *pesos distintos para cada variável* pode ser justificado com base em maior eficiência com modelos mais aderentes para estimar o total de cada variável resposta y . Por outro lado, isso levaria a possíveis problemas de coerência na estimação de certos parâmetros populacionais. Por exemplo, para estimar razões de totais. Outra incoerência surge quando estimativas ponderadas de partes de uma soma podem não corresponder à estimativa ponderada do total para a soma das partes.

Portanto, a ideia de usar diferentes conjuntos de pesos para diferentes variáveis y é pouco atraente na prática.

Embora isso não seja um requisito de calibração, assumimos a partir de agora que a derivação dos pesos de calibração é feita com o objetivo de usar um único conjunto de pesos calibrados $\{w_k, k \in s\}$ para estimar todas as variáveis de pesquisa.

14.4 Razões para calibração

Os estimadores de calibração possuem algumas boas propriedades. Primeiro, os pesos de calibração de acordo com a Equação (14.4) fornecem ‘estimativas’ amostrais para os totais das variáveis auxiliares que correspondem exatamente aos totais populacionais conhecidos para essas variáveis. Portanto, se os totais populacionais das variáveis auxiliares tiverem sido publicados antes que os resultados da pesquisa sejam produzidos, a calibração garantirá que as estimativas da pesquisa sejam *coerentes* com as já existentes no domínio público. Essa propriedade, embora não seja essencial, é uma das razões dominantes pelas quais a calibração é usada com frequência na prática de pesquisa.

A segunda propriedade é a sua *simplicidade*, ou seja, o fato de as estimativas de calibração serem lineares. Isso significa que cada registro de pesquisa pode ter um único peso a ser usado para estimar todas as variáveis da pesquisa. O cálculo das estimativas para totais, médias, proporções e muitos outros parâmetros é simples, usando um *software* estatístico padrão. No caso das funções de distância definidas pelas Equações (14.6) e (14.7), os pesos calibrados são dados em uma expressão de forma fechada e são fáceis de calcular usando um *software* estatístico padrão.

A terceira propriedade de tais estimadores de calibração é sua *flexibilidade* para incorporar informações auxiliares que podem incluir variáveis contínuas, categóricas ou de ambos os tipos ao mesmo tempo. Se os totais auxiliares representam contagens dos números de unidades populacionais em certas classes de variáveis categóricas (discretas), então os valores das variáveis x correspondentes são simplesmente indicadores das unidades que são membros das classes correspondentes. A classificação cruzada de duas ou mais variáveis categóricas também pode ser facilmente acomodada pela definição de variáveis indicadoras para as combinações correspondentes de categorias.

Os estimadores de calibração também produzem algum grau de *integração*, no sentido de que alguns estimadores amplamente usados são casos especiais, por ex. estimadores de razão, regressão e pós-estratificação (ver capítulo 7 de Särndal, Swensson, e Wretman (1992)), bem como pós-estratificação incompleta (ver Bethlehem e Keller (1987)).

Os estimadores de calibração, sob certas hipóteses, também podem oferecer alguma proteção contra o *vício de não resposta* e a obtenção de *estimativas mais precisas* de totais.

A pós-estratificação e a estimativa de regressão são amplamente usadas para tentar reduzir o vício de não resposta em pesquisas por amostragem. O estimador de regressão (calibração) será aproximadamente não viciado quando o modelo de regressão definido em (14.10) se mantiver e o mecanismo combinado de amostragem e resposta for ignorável, dado o conjunto de variáveis x para as quais informações de população auxiliar estão disponíveis (por exemplo, Bethlehem (1988), Särndal e Lundström (1999), e também o capítulo 15 de Särndal, Swensson, e Wretman (1992)).

Todas essas razões são argumentos poderosos para usar a calibração. No entanto, ao fazer isso, os usuários também devem estar cientes de alguns problemas ou dificuldades que podem ser encontrados. Primeiro, notamos que os estimadores de calibração não são exatamente não viciados. De fato, o vício do estimador de calibração é dado por:

$$E(\hat{\mathbf{T}}_y^C - \mathbf{T}_y) = E \left[\sum_{k \in s} (w_k - d_k) \mathbf{y}_k \right]$$

A Tabela 14.1 apresenta uma comparação dos estimadores *HT* e de calibração.

Tabela 14.1: Comparação dos estimadores *HT* e de calibração.

Descrição	Estimador HT	Estimador de Calibração
Pesos	$\{d_k, k \in s\}$	$\{w_k, k \in s\}$
Vício	0	$E \left(\sum_{k \in s} (w_k - d_k) y_{kj} \right)$
Variância	$V_p \left(\sum_{k \in s} d_k y_{kj} \right)$	$V_p \left(\sum_{k \in s} w_k E_{kj} \right)$

Se os pesos calibrados estiverem ‘próximos’ dos pesos de desenho para todas as amostras, então o vício será insignificante ou próximo de zero. Isto sustenta o critério usado para definir os pesos de calibração w_k , que requer que sua distância ao d_k seja minimizada. No entanto, para amostras pequenas ou moderadas, é preciso estar ciente da possibilidade de enfrentar uma certa quantidade de vício.

Se os erros E_{kj} do modelo definido em (14.10) forem ‘menos espalhados’ que os valores y_{kj} , então:

$$V_P\left(\sum_{k \in s} w_k E_{kj}\right) < V_P\left(\sum_{k \in s} d_k y_{kj}\right)$$

Podemos então observar que o estimador de regressão (calibração) deverá ter um bom desempenho em termos de precisão quando a variância dos resíduos do modelo de regressão definido pela Equação (14.10) for pequena em comparação com a da variável y original. Este será o caso quando a relação linear for uma boa aproximação para a regressão de y em x e as variáveis x no estimador de regressão tiverem bom poder preditivo para y . Os dois gráficos da Figura 14.1 ilustram essa ideia. Neste exemplo, os resíduos do estimador de regressão para o modelo $y = Bx$ têm variância menor do que a variável y original (modelo $y = B$), levando ao estimador de regressão com menor variância aproximada do que a variância do estimador de HT para amostras do mesmo tamanho.

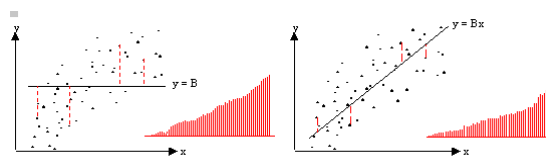


Figura 14.1: Resíduos para o modelo de regressão da população para os estimadores de HT (esquerda) e de calibração (direita)

Exemplo 14.1 Considere o gráfico que ilustra, para uma população de fazendas de cana-de-açúcar, a relação entre a quantidade produzida (y) e a área plantada (x).

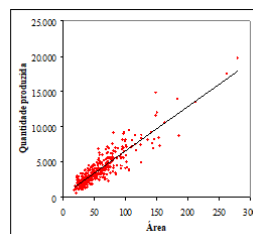


Figura 14.2: Regressão linear para a população de fazendas de cana-de-açúcar

Usando uma amostra aleatória simples com $n = 50$ fazendas, e considerando a *Área* como variável auxiliar, estimamos os totais populacionais de fazendas de cana-de-açúcar por dois métodos: pelo estimador HT e pelo estimador de regressão (GREG). A Tabela 14.2 apresenta as estimativas de totais, bem como as dos respectivos erros padrão e coeficientes de variação (CV).

Tabela 14.2: Estimativas de total de fazendas de cana-de-açúcar e dos erros padrão e coeficientes de variação (CV) usando os estimadores *HT* e de regressão (GREG) para variáveis selecionadas

Variável	Total <i>HT</i>	Total GREG	Erro Padrão <i>HT</i>	Erro Padrão GREG	CV <i>HT</i>	CV GREG
Área	18.015	20.364	1.699	0	9,43	0,00
Quantidade	1.236.046	1.375.931	111.997	55.658	9,06	4,05
Receita	29.957.792	33.915.479	3.123.624	1.439.641	10,43	4,24
Despesa	18.480.035	20.785.241	1.835.615	866.154	9,93	4,17

Os resultados mostram a superioridade do estimador de regressão quanto à redução do erro padrão (ou CV).

Comandos Relevantes do R

```
fazendas.plano = svydesign(data=fazendas.amo, id=~1, fpc=~CPF)
```

```
Totais.HT = svytotal(design=fazendas.plano, ~AREA + QUANT + RECEITA + DESPESA)
```

```
fazendas.calib1 = calibrate(fazendas.plano, formula=~AREA, population=Totais.pop)
```

```
Totais.GREG = svytotal(design=fazendas.calib1, ~AREA + QUANT + RECEITA + DESPESA)
```

Exemplo 14.2 Usando a mesma amostra aleatória simples com $n=50$ fazendas, e considerando *Região* como variável auxiliar, estimamos os totais populacionais de fazendas de cana-de-açúcar usando agora os estimadores *HT* e de pós-estratificação (PS-Região). A Tabela 14.3 apresenta as estimativas de total e dos respectivos erros padrão e coeficientes de variação (CV).

Tabela 14.3: Estimativas de total de fazendas de cana-de-açúcar e dos erros padrão e coeficientes de variação (CV) usando os estimadores *HT* e pós-estratificação (PS-Região) para variáveis selecionadas

Variável	Total <i>HT</i>	Total PS-Região	Erro Padrão <i>HT</i>	Erro Padrão PS-Região	CV <i>HT</i>	CV PS-Região
Área	18.015	18.059	1.699	1.501	9,43	8,31
Quantidade	1.236.046	1.234.739	111.997	104.861	9,06	8,49
Receita	29.957.792	29.719.944	3.123.624	2.788.930	10,43	9,38
Despesa	18.480.035	18.690.411	1.835.615	1.715.898	9,93	9,18

Os resultados mostram pequena redução do CV quando comparado com estimador *HT*. Portanto, a variável auxiliar *Área* leva a estimativas mais precisas.

14.5 Propriedades dos estimadores para grandes amostras

Para grandes amostras, os estimadores de calibração definidos pelos pesos de regressão (14.8) e (14.9) são *assintoticamente não viciados* sob o plano amostral, ou seja:

$$\text{Vício}\left(\hat{T}_{y_j}^C\right) = E\left[\sum_{k \in s} (w_k - d_k)y_{kj}\right] \rightarrow 0, \text{ quando } n \rightarrow \infty$$

e têm variância assintótica do estimador aproximada (ver Särndal, Swensson, e Wretman (1992), p. 235) dada por:

$$V_{A_p}\left(\hat{T}_{y_j}^C\right) = \sum_{k \in U} \sum_{i \in U} (\pi_{ki} - \pi_k \pi_i) (d_k E_{kj}) (d_i E_{ij})$$

Se o vício é desprezível, podemos comparar essa variância aproximada com a do estimador *HT* padrão, dada por:

$$V_p\left(\hat{T}_{y_j}\right) = \sum_{k \in U} \sum_{i \in U} (\pi_{ki} - \pi_k \pi_i) (d_k y_{kj}) (d_i y_{ij})$$

Os estimadores de calibração são também mais *eficientes* que o estimador *HT* se:

$$V_{A_p}(\hat{T}_{y_j}^C) < V_p(\hat{T}_{y_j})$$

Dois estimadores de variância relativamente diretos estão disponíveis para uso com o estimador de regressão.

Särndal, Swensson, e Wretman (1992), p. 235, recomenda usar o estimador da *variância* do estimador de calibração:

$$\hat{V}_p(\hat{T}_{y_j}^C) = \sum_{k \in s} \sum_{i \in s} (1 - \pi_k \pi_i / \pi_{ki}) (d_k g_k e_{kj}) (d_i g_i e_{ij})$$

A expressão acima é similar à do estimador da variância do estimador *HT*, porém: emprega os *resíduos* $e_{kj} = y_{kj} - \mathbf{x}'_k \hat{\mathbf{B}}_j$ em lugar dos valores da variável y_j ; e também emprega os *pesos ajustados* $d_k g_k = w_k$ como multiplicadores dos resíduos.

Um estimador de variância ainda mais simples que não requer os pesos- g é dado por:

$$\tilde{V}_p(\hat{T}_{y_j}^C) = \sum_{k \in s} \sum_{i \in s} (1 - \pi_k \pi_i / \pi_{ki}) (d_k e_{kj}) (d_i e_{ij})$$

Ambos os estimadores de variância são de primeira ordem assintoticamente não viciados de desenho para a variância aproximada do estimador de regressão, mas (14.20) também é aproximadamente não viciado para o modelo Särndal, Swensson, e Wretman (1989).

Sob *Amostragem Aleatória Simples* tem-se:

$$V_{AAS}(\hat{T}_{y_j}^C) = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \frac{1}{N-1} \sum_{k \in U} (E_{kj})^2$$

e pode ser estimada usando:

$$\hat{V}_{AAS}(\hat{T}_{y_j}^C) = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \frac{1}{n-1} \sum_{k \in s} (d_k g_k e_{kj})^2$$

Silva (1996) mostrou que o estimador em (14.23) tem vício de ordem menor que $n - 2$ para estimar a variância correspondente. Enquanto o vício de (14.21) é da ordem $(n - 2)$. Por isso o estimador da Equação (14.20) deve ser preferido ao da (14.21). Holmes e Skinner (2000) apoiam esta visão com base nos resultados de um estudo empírico realizado para comparar estimadores de variância alternativa para a pesquisa de força de trabalho do Reino Unido (UK Labour Force Survey / UK-LFS).

14.6 Alguns problemas práticos com calibração

Enquanto os estimadores de calibração possuem um número de propriedades atraentes, eles não são livres de problemas quando se trata de aplicações práticas. Nesta seção, revisamos alguns dos problemas que afetam os estimadores de calibração e algumas das abordagens que foram desenvolvidas para resolvê-los. Antes de uma discussão detalhada, no entanto, pode ser útil ter uma lista rápida dos problemas que devem ser motivo de preocupação ao realizar a estimativa de calibração na prática:

1. *Amostras* são finitas, geralmente *pequenas* em certos estratos.
2. *Pesos ajustados* podem ser *negativos* (ou menores que 1) ou *extremos* (muito grandes). O que é uma situação inaceitável em muitas aplicações; e têm implicações sobre a variância do estimador.
3. Pode haver *grande número de variáveis de pesquisa* (y) ou de 'grupos' para ajustar modelos de regressão.
4. Pode haver *grande número de variáveis auxiliares* (x).
5. Pesquisas geralmente afetadas por *não resposta*.
6. *Erros de medida* ou *totais populacionais conhecidos apenas aproximadamente*.

A última questão nesta lista (erros de medida e seus efeitos na calibração), apesar de sua importância, não será discutida aqui. Os leitores podem encontrar alguma discussão em Skinner (1999). Todas as outras questões serão tratadas nas seções a seguir.

14.6.1 Tamanhos Pequenos de Amostra

O problema com os estimadores de calibração quando os tamanhos das amostras são pequenos vem do fato de que seu vício pode se tornar importante, em relação à sua variância. É bem conhecido, por exemplo, que os estimadores de razão são tendenciosos e que o vício é da ordem $(n - 1)$ (ver Cochran (1977), p. 160-162). O estimador de razão é um caso especial do estimador de calibração quando o vetor x inclui uma única variável contínua x , nenhum termo de intercepto é incluído, e as constantes q_k são ajustadas para $q_k = 1/x_k$ (assumindo que $x_k > 0 \forall k$). Recomenda-se que os estimadores de razão sejam usados somente para amostras de tamanhos grandes o suficiente para que o vício seja desprezível. Särndal, Swensson, e Wretman (1992), p. 251, sugerem que amostras de tamanho 20 ou mais devem ser suficientes para que isso aconteça. Cochran (1977), p. 162, sugere que o

coeficiente de variação do estimador de HT do total da variável x , $CV(\hat{T}_x)$, deve ser menor que 0,1 (10%) para que o vício do estimador de razão possa ser ignorado ou considerado pequeno em comparação com o seu erro padrão.

Apesar dessas conhecidas ‘regras práticas’ ou limitações que devem impedir que os estimadores de razão sejam usados com amostras muito pequenas, software moderno torna fácil estimar a razão e outros estimadores de calibração para amostras de qualquer tamanho, muitas vezes sem qualquer aviso de que tamanhos de amostra podem ser insuficientes para garantir a utilização segura. Isso deixa espaço para aplicações em que nem mesmo precauções mínimas sejam tomadas, como essa de verificar se o tamanho da amostra é adequado. Nos casos em que os tamanhos das amostras são muito pequenos, as estimativas de calibração podem estar sujeitas não apenas à grande variação (como esperado devido ao pequeno tamanho da amostra), mas também a um vício perceptível. Os usuários de estimadores de calibração são encorajados a evitar a aplicação da técnica quando os tamanhos das amostras são muito pequenos. Até o momento, não foram desenvolvidas regras simples de segurança em relação ao tamanho mínimo das amostras para a família geral de estimadores de calibração. No entanto, pode-se pelo menos sugerir que as mesmas regras aplicáveis à estimativa de razão simples devem ser satisfeitas antes de aplicar alguma outra forma de estimativa de calibração.

Diante de pequenas amostras, uma ideia é adotar *métodos dependentes de modelos*, e abandonar inferência ‘sob o plano amostral’. As dificuldades a enfrentar neste caso surgem quando: há grande número de variáveis de pesquisa; e há grande número de ‘grupos’ para modelagem. Em consequência, tarefas de *ajuste e validação de modelos* podem se tornar muito onerosas e demoradas, implicando em modelos distintos geralmente requeridos para variáveis distintas, levando a diferentes conjuntos de pesos e as consequências advindas dos problemas de coerência. Portanto, calibração é para amostras moderadas / grandes!

14.6.2 Pesos negativos ou extremos

Outro conjunto de problemas compreende as situações que surgem quando os pesos de calibração são em algum sentido considerados extremos ou não representativos. Um caso importante ocorre quando a calibração resulta em pesos negativos, isto é, em ter alguns pesos $w_k < 0$ (ou $g_k < 0$). Esta situação não representa nenhum problema do ponto de vista estritamente teórico, mas leva a duas dificuldades do ponto de vista prático. Primeiro, a interpretação usual dos pesos dos casos como o número de unidades populacionais representadas pela unidade de amostra correspondente é perdida para esses casos, e a liberação de tais pesos seria uma decisão muito desconfortável para muitos órgãos de estatística. O segundo problema é que os pesos negativos podem eventualmente gerar

estimativas negativas para alguns domínios com tamanhos de amostra pequenos, o que não é um resultado aceitável para a maioria das aplicações práticas de pesquisa quando as variáveis da pesquisa são intrinsecamente não-negativas. Também observamos que pesos negativos podem fornecer uma indicação de algum problema com a tentativa de calibração que precisa de atenção do estatístico responsável.

Para resolver esse problema da possibilidade de pesos negativos, várias abordagens foram desenvolvidas. Uma abordagem que é implementada em pacotes de software desenvolvidos por algumas agências de estatística é calcular os pesos de ajuste de calibração que minimizam:

$$\sum_{k \in s} (w_k - d_k)^2 / q_k d_k = \sum_{k \in s} (d_k g_k - d_k)^2 / q_k d_k = \sum_{k \in s} d_k (g_k - 1)^2 / q_k$$

tal que não haja *erros de calibração*, ou seja, sujeito às restrições de calibração:

$$\hat{\mathbf{T}}_x^C - \mathbf{T}_x = \sum_{k \in s} g_k d_k \mathbf{x}_k - \mathbf{T}_x = 0$$

e que os fatores de ajuste sejam limitados abaixo e acima tal que:

$$LI \leq g_k \leq LS \quad \forall k \in s, \text{ com } 0 < LI < 1 < LS$$

Esta é a abordagem adotada no desenvolvimento do *GES (Generalized Estimation System)* pelo Statistics Canada (ver Estevao e Hidiroglou (1995)).

Este problema corresponde à minimização de uma função quadrática (14.24) sob restrições de limite linear (14.25) e (não linear) (14.26). A solução do problema é obtida por *métodos não lineares de otimização*.

O sistema *GES* tenta resolver esse problema usando um algoritmo eficiente, mas *nem sempre existe uma solução* garantida, dependendo dos limites especificados e da configuração dos dados. A probabilidade de encontrar uma solução tende para 1 quando a amostra cresce (ver Deville e Särndal (1992)).

O sistema *GES* inclui, além da determinação dos pesos de calibração, estimação de totais, médias e razões, global e para domínios, juntamente com suas correspondentes variâncias para desenhos de amostragem estratificada ou de conglomeração com um único estágio.

Os estatísticos que procuram uma ferramenta computacional para implementar a calibração devem dar a devida consideração a este pacote. Uma desvantagem é a sua dependência do software estatístico *SAS (Statistical Analysis System)*, o que torna esta uma opção

razoavelmente cara. Se o SAS já estiver disponível, o licenciamento de GES pelo Statistics Canada não é proibitivo para a maioria das agências de estatística, e custaria muito menos do que o desenvolvimento de um software equivalente.

Outra implementação da abordagem acima está disponível no *BASCULA*, que é um sistema desenvolvido pelo Statistics Netherlands (veja Nieuwenbroek e Boonstra (2002)). A principal diferença entre o GES e o *BASCULA* é o algoritmo usado para calcular os pesos de calibração.

O *BASCULA* adota um algoritmo iterativo proposto por Huang e Fuller (1978) para calcular pesos calibrados que satisfazem as restrições de limites. Como é o caso do GES, o *BASCULA* também nem sempre garante encontrar uma solução que satisfaça todas as restrições especificadas.

O sistema trata separadamente casos especiais: pós-estratificação, estimação de razão e estimação tipo 'raking'.

O *BASCULA* é um programa autônomo (standalone), mas funciona bem com o sistema BLAISE para coleta apoiada por computador. Também requer licenciamento junto ao desenvolvedor e pode ser mais barato obter do que o GES se a organização ainda não for usuária do SAS.

Outra abordagem proposta para resolver o problema de pesos negativos é devida a Deville e Särndal (1992), que definiu a família de estimadores de calibração. Na abordagem anterior, a função de distância padrão, levando aos pesos de regressão, foi mantida e as condições de limites foram impostas como restrições adicionais. A abordagem proposta por Deville e Särndal (1992) consiste em modificar a função de distância a ser utilizada no cálculo dos pesos calibrados, de forma a evitar a possibilidade de pesos negativos desde o início. Por isso, a ideia é definir pesos de calibração que *minimizem a distância* $D(w; d)$ definida em (14.6) para cada amostra s , sujeita às restrições de calibração (14.25), mas com a *função de distância* G_k definida como um dos tipos de funções apresentados na Tabela 14.4.

Tabela 14.4: Funções de distância para estimação de calibração propostas em Deville e Särndal (1992)

Tipo	Funções de Distância $q_K G_K(w_k, d_k)$
1	$(w_k - d_k)^2 / 2d_k$
2	$w_k [\log(w_k / d_k) - 1] + d_k$
3	$2(\sqrt{w_k} - \sqrt{d_k})^2$
4	$w_k - d_k [\log(w_k / d_k) + 1]$
5	$(w_k - d_k)^2 / 2w_k$
6	$(g_k - LI) \log\left(\frac{g_k - LI}{1 - LI}\right) + (LS - g_k) \log\left(\frac{LS - g_k}{LS - 1}\right) \quad g_k = w_k / d_k, \quad 0 < LI < 1 < LS$

Todas as funções de distância G consideradas satisfazem algumas condições de regularidade, ou seja, para cada valor fixo $d > 0$:

- $G_k(w, d) \geq 0$ e $G(d, d) = 0$;
- $G_k(w, d)$ é definida num intervalo D_k que contém d ;
- $G_k(w, d)$ é estritamente convexa e duas vezes diferenciável em w ;
- $\partial G_k(w, d) / \partial w$ é contínua e mapeia D_k sobre um intervalo $Im_k(d)$ de forma biunívoca (um para um).

A solução para o problema de *minimizar a distância* $D(w; d)$ pode ser obtida usando o método dos multiplicadores de Lagrange, sob as restrições definidas em (14.25). A solução é encontrada resolvendo:

$$\partial G_k(w_k, d_k) / \partial w_k - \mathbf{x}_k' \lambda = 0 \quad \forall k \in s$$

Se uma solução existir, sob algumas das condições de regularidade ela será única e dada por:

$$w_k = d_k g_k \text{ com } g_k = F(q_k \mathbf{x}_k' \lambda)$$

onde $F(\cdot)$ é a inversa de $\partial G_k(w_k, d_k) / \partial w_k$ e λ é um *multiplicador de Lagrange* que resolve as equações:

$$\sum_{k \in s} d_k \left[F(q_k \mathbf{x}_k' \lambda) - 1 \right] \mathbf{x}_k = \mathbf{T}_x - \hat{\mathbf{T}}_x$$

O estimador de calibração resultante é então dado por:

$$\hat{\mathbf{T}}_y^C = \sum_{k \in s} d_k g_k y_k$$

com os fatores de ajuste de calibração g_k definidos por uma das funções de calibração $F(\cdot)$, correspondentes a alguma função de distância, propostas em Deville e Särndal (1992) e descritas na Tabela 14.5.

Tabela 14.5: Funções de calibração propostas em Deville e Särndal (1992)

Tipo	Funções de Calibração $F(q_k u)$ com $0 < LI < 1 < LS$
1	$1 + q_k u$
2	$\exp(q_k u)$
3	$(1 - q_k u/2)^{-2}$
4	$(1 - q_k u)^{-1}$
5	$(1 - 2q_k u)^{-1/2}$
6	$\frac{LI(LS - 1) + LS(1 - LI)\exp(Aq_k u)}{(LS - 1) + (1 - LI)\exp(Aq_k u)}$ com $A = \frac{LS - LI}{(1 - LI)(LS - 1)}$
7	$\begin{array}{ll} LI & \text{se } u < (LI - 1)/q_k \\ 1 + q_k u & \text{se } (LI - 1)/q_k \leq u \leq (LS - 1)/q_k \\ LS & \text{se } u > (LS - 1)/q_k \end{array}$

Observe que a função de calibração 7 corresponde à função número 1 da Tabela 14.4, mas com limites especificados para os pesos de calibração.

Um algoritmo para calcular os pesos de calibração proposto por Deville e Särndal (1992) pode ser especificado com a seguinte sequência de etapas:

Passo 1: Calcule os erros de calibração para o estimador de HT para as variáveis auxiliares consideradas: $\mathbf{T}_x - \hat{\mathbf{T}}_x$.

Passo 2: Para a função de calibração $F(\cdot)$ escolhida, resolva as Equações (14.29) requeridas para determinar o *multiplicador de Lagrange* λ .

Isto pode ser feito aplicando o método de Newton. Defina:

$$\mathbf{H}_s(\lambda) = \sum_{k \in s} d_k \left[F(q_k \mathbf{x}_k' \lambda) - 1 \right] \mathbf{x}_k$$

Então o passo 2 do algoritmo requer achar o valor de λ tal que:

$$\mathbf{H}_s(\lambda) = \mathbf{T}_x - \hat{\mathbf{T}}_x$$

Primeiro obtenha um valor inicial para λ como:

$$\lambda_1 = \left(\sum_{k \in s} q_k d_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} [\mathbf{T}_x - \hat{\mathbf{T}}_x]$$

Em seguida, realize iterações do método de Newton, calculando a cada iteração $r = 2, 3, \dots$ o valor atualizado para λ como:

$$\lambda_{r+1} = \lambda_r + \left[\mathbf{H}_s'(\lambda_r) \right]^{-1} [\mathbf{T}_x - \hat{\mathbf{T}}_x - \mathbf{H}_s(\lambda_r)]$$

onde:

$$\mathbf{H}_s'(\lambda_r) = \partial \mathbf{H}_s(\lambda) / \partial \lambda \big|_{\lambda = \lambda_r}$$

As iterações devem prosseguir até convergência (até um limite de tolerância especificado ser satisfeito) ou até o número máximo de iterações ser realizado sem haver convergência.

Passo 3: Com a solução obtida para o valor de λ , calcule os pesos de calibração como:

$$w_k = d_k F(q_k \mathbf{x}_k' \lambda)$$

Os pesos de calibração e estimadores correspondentes obtidos como resultado deste algoritmo preservam todas as propriedades desejáveis que discutimos em conexão com os estimadores de regressão (seções 14.2 e 14.3).

- *Simplicidade*: o estimador é linear e calculável de forma bem simples, uma vez calculados os pesos ajustados.
- *Flexibilidade* para incorporar informação auxiliar: tanto para variáveis categóricas, contínuas, ou ambas.
- Os *totais são calibrados* para as variáveis auxiliares: $\mathbf{T}_x - \hat{\mathbf{T}}_x$
- *Integração*: os estimadores de pós-estratificação, razão, GREG e 'raking' (como os usados para ponderar pessoas no UK-LFS) são todos casos particulares de estimadores de calibração.

Deville e Särndal (1992) demonstraram que membros dessa classe de estimadores de calibração possuem propriedades assintóticas idênticas àquelas dos estimadores GREG baseados no mesmo conjunto de variáveis auxiliares.

Portanto, os estimadores de calibração definidos por uma das funções de distância acima são assintoticamente *não viciados*, com *variância* assintótica *igual à do estimador GREG* dada pela Equação (14.17). Além disso, sua *variância* também pode ser *estimada* pelas Expressões (14.20) ou (14.21).

O estimador de calibração resultante do algoritmo *não é mais eficiente* que o estimador GREG, mas oferece *flexibilidade* para lidar com pesos extremos.

Os estimadores de calibração deste tipo foram implementados na Macro *CALMAR* em SAS do INSEE (ver Sautory (2003)). Este programa executa apenas o cálculo de peso, mas uma variante denominada CALJACK foi desenvolvida no Statistics Canada (ver Bernier e Lavallée (1994)) que inclui a estimativa de variância de Jackknife para totais, médias, razões e diferenças destes. O CALMAR também requer o SAS, mas uma implementação mais recente (mas limitada) do método está disponível: a Macro *g-CALIB-S*, desenvolvida na Statistics Belgium, é executada no SPSS Vanderhoeft (2000). As funções *calibrate*, *postStratify* e *rake* do pacote *survey* do sistema *R* (ver Lumley (2010)) também têm o método implementado.

A estimação por calibração como agora estendida fornece as ferramentas para tentar *resolver o problema de pesos negativos*, o que pode ser evitado escolhendo as funções de calibração 2 a 7 na Tabela #tab:tabcalib5. Também dá algum *controle sobre o problema de pesos extremos ou pesos menores que 1*, o que pode ser evitado escolhendo as funções de calibração 6 ou 7, usando $LI = 1 / \min \{d_k, k \in s\}$ e especificando algum *LS* adequado.

No entanto, vários dos problemas discutidos anteriormente permanecem sem solução.

Primeiro, para amostras pequenas e moderadas, o *vício* pode ser um problema e agora, a *escolha da função de distância* pode se tornar importante nesse aspecto. Segundo, embora o método seja voltado para evitar pesos negativos ou extremos, a computação dos pesos pode ser demorada (requer *método iterativo* de solução) e *pode não haver solução* para as equações de calibração. Deville e Särndal (1992) provaram que a probabilidade de encontrar uma solução para λ no passo 2 do algoritmo tende para 1 quando n aumenta. No entanto, não é 1 com amostras finitas. Portanto, em algumas aplicações, o método pode não convergir dependendo das opções de $F(\cdot)$, *LI* e *LS*. Quando esse for o caso, os usuários do método devem tentar investigar as causas por trás da falha em encontrar uma solução. Pode ser devido a *amostras pequenas* ou 'extremas', no sentido de que os pesos de calibração resultantes podem precisar ser mais extremos do que estamos preparados para permitir quando especificamos as restrições de limite *LI* e *LS*. Isso também pode acontecer porque *grande número de variáveis auxiliares* são consideradas para calibração, o que pode levar a problemas de colinearidade, uma questão que discutiremos na seção 14.6.4.

14.6.3 Grande número de variáveis de pesquisa ou de ‘grupos de modelos’

Outra fonte de dificuldades para o uso de estimadores de regressão (ou calibração) é o fato de que eles são aplicados separadamente para vários ‘grupos de modelos’, definidos como grupos de unidades para os quais informações de participação de amostra e de população auxiliar estão disponíveis. Esses grupos de modelos podem coincidir com estratos de amostragem pré-definidos, ou podem ser formados após a seleção da amostra, caso em que desempenharão o papel de pós-estratos. Quando esses grupos de modelos são numerosos, pequenos tamanhos de amostra podem resultar em alguns (ou na verdade muitos) deles.

O problema muitas vezes é agravado pelo fato de que o número de variáveis de pesquisa também pode ser grande. Neste caso, embora o cálculo dos pesos de calibração seja realizado apenas uma vez (os pesos não dependem das variáveis y), a adequação dos modelos subjacentes que fornecem as condições para que os estimadores de calibração tenham um bom desempenho (no sentido de fornecer resíduos com pequena variação) deve ser verificada. Em alguns casos, essa tarefa pode se tornar grande demais para ser viável dentro de cronogramas apertados de produção que as pesquisas típicas devem adotar. Por esse motivo, os usuários são alertados contra a tentativa de executar a calibragem em níveis detalhados demais no sentido de envolver muitos grupos de modelos. Quanto mais grupos de modelos forem considerados para a calibração, mais recursos deverão ser dedicados à validação do modelo e análise das estimativas de calibração resultantes.

Essa discussão se assemelha à comparação de estimadores de razão separada e combinada. Estimadores de razão separada são estimadores de calibração onde a calibração é realizada para os totais conhecidos no nível do estrato (grupo modelo). Os estimadores de razão combinada envolvem a calibração apenas no nível agregado (para a amostra como um todo ou para alguns grupos mais amplos formados a partir de conjuntos de estratos agrupados). Cochran (1977), p. 167, argumenta que ‘o uso de uma estimativa de razão separada em cada estrato provavelmente será mais preciso se a amostra em cada estrato for grande o suficiente para que a fórmula aproximada para a variância do estimador de razão separada seja válida, e o vício cumulativo que pode afetar o estimador de razão separada é insignificante. Com apenas uma pequena amostra em cada estrato, o estimador de razão combinada deve ser recomendado, a menos que exista uma boa evidência empírica em contrário.’ Sugerimos que esse conselho também seja aplicado a estimadores de calibração em geral.

14.6.4 Grande número de variáveis auxiliares

Um problema que as abordagens discutidas acima não trataram é o que fazer quando um grande número de potenciais variáveis x está disponível para serem consideradas para calibração. Uma opção simplista é considerar cada uma das potenciais variáveis x na calibração. Isso pode parecer desejável do ponto de vista prático, porque o erro de calibração seria zero para todos os totais populacionais conhecidos. No entanto, essa opção também pode causar vários problemas.

Primeiro, pode ser mais difícil resolver o sistema de equações de calibração necessárias para determinar λ no passo 2 do algoritmo, porque seu tamanho aumenta com o número de variáveis x , e o cálculo pode ser demorado.

Segundo, números maiores de variáveis x podem levar a problemas de *dependência linear* que afetem a solução das equações de calibração. Bankier (1989) e Sautory (2003) propuseram o *descarte de variáveis auxiliares* linearmente dependentes antes de tentar a solução das equações de calibração na etapa 2 do algoritmo. Essa solução é bastante fácil de implementar e não leva à perda de calibração para nenhuma variável x , já que as variáveis descartadas são combinações lineares exatas de variáveis retidas no problema de calibração, e os estimadores de calibração resultantes são lineares.

Uma solução alternativa usando *inversa generalizada* de matrizes foi implementada no programa g-CALIB-S (ver Vanderhoeft (2000)).

Bankier (1989) e Bankier, Rathwel, e Majkowski (1992) também propuseram o descarte de variáveis auxiliares *quase linearmente dependentes* para controlar a variação de peso, mantendo a função de distância padrão do tipo 1. Essa solução leva à perda de calibração de variáveis x descartadas, bem como não permite controlar sobre quais variáveis x teriam calibração.

Um problema adicional encontrado quando um número muito grande de variáveis auxiliares são consideradas na calibração é o aumento potencial do erro quadrático médio (MSE) do estimador de calibração resultante.

Silva (1996), cap. 4, e Silva e Skinner (1997) mostraram que, às vezes, um grande número de variáveis auxiliares pode realmente reduzir a eficiência do estimador de calibração (regressão) para tamanhos de amostra pequenos a moderados. Por exemplo, em amostragem aleatória simples sem reposição (AAS) e assumindo o modelo definido em (14.10) com $q_k = 1$ para cada k , Silva (1996), p. 45, mostrou que:

$$EQM_{AAS+M}(\hat{T}_{y_j}^C/N) = \left(\frac{1}{n} - \frac{1}{N}\right)\sigma_j^2 \left(1 + \frac{Z}{n}\right) + O(n^{-5/2})$$

onde:

σ_f^2 é a variância dos resíduos da regressão de y em x ;

Z é o número de variáveis x consideradas.

Essa expressão revela que o EQM de um estimador de regressão pode realmente aumentar à medida que o número de variáveis x aumenta, se o aumento no termo de segunda ordem Z/n compensa a diminuição na variância dos resíduos σ_f^2 . Naturalmente, isso não é um problema se a amostra for grande, mas para amostras pequenas a moderadas, o número de variáveis auxiliares pode ter algum efeito perceptível no EQM do estimador de regressão.

Ou

$$V_M\left(\hat{t}_{y_j}^{C/N} | \mathbf{X}_U\right) = \left(\frac{1}{n} - \frac{1}{N}\right) \sigma_f^2 \left(1 + CV_g\right)$$

onde: CV_g é o coeficiente de variação dos fatores de ajuste g_k .

Como uma ilustração do problema, a Figura 14.3 plota o EQM do estimador de regressão para o aumento de conjuntos de variáveis auxiliares, assumindo amostragem aleatória simples com $n = 100$ de uma população de chefes de domicílios para os quais os dados foram coletados como parte do Censo Demográfico Experimental de Limeira, São Paulo, Brasil, 1988.

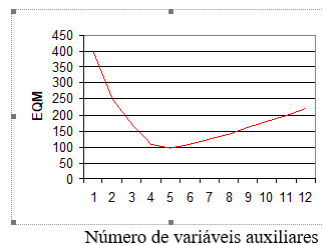


Figura 14.3: EQM do estimador de regressão versus número de variáveis x

Silva e Skinner (1997) mostraram, em um exercício de simulação limitado, que estimadores de regressão *após seleção de subconjuntos de variáveis auxiliares* podem ser mais eficientes do que estimadores de regressão que considera todas as variáveis auxiliares (*modelo saturado*), para amostras de tamanhos moderados ($n = 100$ e $Z = 5; 10$). Achados semelhantes são relatados por Clark (2002) para $n = 100; 250$ e $Z = 24; 40$. Ambas as fontes relataram também que a incidência de pesos negativos foi menor após a seleção de subconjuntos do que quando a calibração usou o conjunto saturado de variáveis x . Isso sugere que parte do problema com pesos negativos vem da calibração ‘excessiva’.

Embora a ideia de aplicar alguma forma de procedimento de *seleção de variáveis* auxiliares x *para calibração* possa levar a estimadores mais eficientes para algumas variáveis y especificadas, essa abordagem não é livre de problemas. Primeiro, leva à *perda de*

calibração para variáveis x descartadas. Segundo, a abordagem é intrinsecamente *univariada*, no sentido de que a seleção de covariáveis é específica para cada variável de pesquisa y , implicando que pesos distintos deveriam ser usados para diferentes variáveis y . Além disso, a *estimação de variância se torna mais difícil* após a seleção de variáveis, como observado em Silva e Skinner (1997).

No entanto, a mensagem emergente desses estudos é que realizar a estimativa de calibração ou regressão ‘automaticamente’ com todas as variáveis auxiliares disponíveis pode não ser uma estratégia eficiente, particularmente para amostras com tamanhos pequenos a moderados ou quando o número de variáveis auxiliares é grande em relação a o tamanho da amostra. Nesses casos, recomenda-se dedicar alguma atenção à seleção de subconjuntos adequados das variáveis disponíveis, mesmo que não se esteja preparado para usar procedimentos formais de seleção de modelos, como os descritos por Silva e Skinner (1997). Em pesquisas repetidas, por exemplo, pode-se dedicar esforços nas primeiras rodadas da pesquisa para estabelecer um conjunto adequado de variáveis para calibrar e, em seguida, usar o subconjunto fixo para calibração nas rodadas subsequentes da pesquisa.

Algumas outras abordagens foram propostas para lidar com o caso de pesos negativos ou extremos. Chambers (1997) propôs os chamados estimadores de *ridge calibration*, onde a ideia básica é minimizar a função de distância modificada dada por:

$$\sum_{k \in S} (w_k - d_k)^2 / q_k d_k + \frac{1}{\gamma} (\mathbf{T}_x - \hat{\mathbf{T}}_x^C)' \Delta (\mathbf{T}_x - \hat{\mathbf{T}}_x^C)$$

onde: Δ é uma matriz diagonal de custos de erro de calibração e γ é um *scala ridging parameter* a ser especificado. Nesta abordagem, não há restrições a serem satisfeitas. Os pesos resultantes são dados por:

$$w_k = d_k \left[1 + q_k \left(\mathbf{T}_x - \hat{\mathbf{T}}_x^C \right)' \left(\gamma \Delta^{-1} + \sum_{i \in S} q_i d_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \mathbf{x}_k \right]$$

Observe que, nessa abordagem, uma medida da quantidade de erro de calibração é incorporada como o segundo termo da função de distância. No entanto, como não há restrições de calibração vinculadas, não é mais garantido que os pesos e estimadores resultantes evitem totalmente o erro de calibração. Alguma escolha de γ tal que todos os pesos ‘ridge-calibrated’ definidos em (14.40) sejam positivos é sempre possível. Uma ideia pode ser escolher o menor γ que satisfaz essa condição.

Outras abordagens para a escolha de γ são discutidas por Chambers (1997). A especificação cuidadosa da ‘matriz de custo do erro de calibração’ permite a seleção flexível de subconjuntos de variáveis auxiliares para as quais o erro de calibração deve ser eliminado.

Para isso, basta usar elementos diagonais muito grandes nessa matriz, correspondentes às variáveis auxiliares para as quais o erro de calibração deve ser zero. A abordagem é uma melhoria em relação aos procedimentos que descartam completamente variáveis auxiliares, no sentido de que algum controle sobre a quantidade de erro de calibração pode ser mantido para todas as variáveis x .

Chambers (1997) considerou outras versões de *ridge calibration* que têm como ponto de partida pesos derivados sob uma abordagem baseada em modelo ou abordagem baseada em modelos não paramétricos. Ele também considerou modificações robustas discrepantes dessa abordagem que podem ajudar nos casos em que os valores discrepantes das variáveis y são preocupantes. Estes são específicos da variável y , no entanto, e, portanto, não devem ser considerados mais aqui.

Rao e Singh (2009) propuseram uma outra abordagem em linhas similares, que é chamada de *ridge-shrinkage calibration*. Novamente, a ideia é minimizar uma função de distância modificada, mas desta vez, sob restrições de limite. O procedimento é semelhante à calibração de Chambers (1997) se a solução inicial satisfaz as restrições de limite. Se não, essas restrições de limite são relaxadas e o procedimento itera entre modificar adaptativamente a matriz de custo de erro de calibração e as restrições de intervalo desejadas até a convergência.

Hedlin et al. (2001) também discutiram o problema de pesos de calibração extremos. Este artigo explorou o comportamento dos estimadores de calibração (*GREG*) quando os modelos subjacentes foram especificados de forma incorreta, e propôs algumas medidas de diagnóstico para avaliar a adequação do modelo para uma dada situação de pesquisa. Parte dos diagnósticos considerou a ideia de que os pesos- g definidos pela Equação (14.9) são funções de medidas de influência bem conhecidas de uma unidade de amostra no ajuste de modelos de regressão linear. Os diagnósticos foram usados para localizar os pesos- g mais extremos e para propor soluções que envolviam, por exemplo, a estratificação da amostra e o uso de estimadores de regressão ou calibração somente para aquelas unidades amostrais para as quais os pesos- g não são extremos e usando estimador de expansão simples para o pós-estrato formado com as unidades com pesos- g extremos.

A mensagem, novamente, é que o mero cálculo automático dos pesos de calibração não é uma boa prática, e alguma atenção deve ser dedicada para analisar os pesos resultantes para avaliar se o uso de estimadores de calibração ou regressão é 'seguro' e eficiente. Uma maneira simples de fazer isso é executar a análise de dados dos pesos- g e tentar sinalizar aqueles que são extremos em algum sentido. Os casos mais óbvios são os pesos- g negativos ou pequenos (aqueles que levam a pesos calibrados finais menores que 1 ou os pesos- g

muito grandes (digamos, com $g_k > LS$). O ponto de corte LS pode ser determinado arbitrariamente (digamos, fazer $LS = 5$ ou 10), ou por métodos dependentes de dados ($LS > Q3 + 1,5 (Q3 - Q1)$), onde $Q1$ e $Q3$ denotam os quartis de amostra de g_k , $k \in s$.

14.6.5 Não resposta e calibração

Até agora, discutimos a estimativa de calibração sob a hipótese de observação completa da amostra selecionada. No entanto, a não resposta é um problema comum. A maioria das pesquisas na prática terá alguma quantidade de não resposta, apesar de incorporar métodos e procedimentos para evitar a não resposta.

Quando a não resposta é '*ignorável*' ou '*completamente ao acaso*', o seu efeito sobre estimadores é *aumentar a variância* devido à redução do tamanho da amostra. Neste caso, a *compensação da não resposta* costuma ser feita por ajustamentos simples dos pesos usados para estimação. Porém, na prática, não resposta nunca é '*completamente ao acaso*'.

Uma importante nova questão trazida pela não resposta é a do vício. Os *estimadores de HT usuais* (ponderado pelo inverso de π) *serão viciados* a menos que a não resposta seja completamente aleatória, e mesmo nesta situação improvável, a estimativa dos totais requer pelo menos algum ajuste simples para compensar a perda de unidades da amostra devido a não resposta.

A calibração é uma abordagem útil para tentar *reduzir o vício* causado pela não resposta. Lundström (1997) e Särndal e Lundström (1999) até sugerem a '*calibração como um método padrão para o tratamento da não resposta*'. Os estimadores de calibração são aproximadamente não viciados se houver resposta completa, para qualquer escolha fixa de variáveis auxiliares. Sob a não resposta, no entanto, os estimadores de calibração podem ser viciados, mesmo em grandes amostras.

Novo critério para avaliar eficiência do estimador é necessário quando o *vício* do estimador *HT não é ignorável* quando ocorre não resposta. A ideia básica é considerar:

$$EQM(\hat{T}_{y_j}^C) = V(\hat{T}_{y_j}^C) + B^2(\hat{T}_{y_j}^C)$$

Para avaliar este novo critério, é essencial adotar um *modelo* para o *mecanismo de resposta*.

Suponha que a unidade $k \in U$ responde com probabilidade $r_k > 0$. Sob a *distribuição conjunta* do plano amostral e do modelo de resposta, o estimador de regressão tem vício aproximado dado por:

$$B\left(\hat{T}_{y_j}^R\right) \doteq N^{-1} \sum_{k \in U} \left(y_{kj} - \mathbf{x}_k \gamma_j\right)$$

onde:

$$\gamma_j = \left(\sum_{k \in U} r_k q_k \mathbf{x}_k' \mathbf{x}_k \right)^{-1} \left(\sum_{k \in U} r_k q_k \mathbf{x}_k' y_{kj} \right) \doteq E(\hat{\mathbf{B}}_j)$$

A *redução de vício* pretendida por calibração somente será alcançada, no entanto, se o mecanismo combinado de resposta e amostragem é *ignorável* dadas as variáveis x consideradas para calibração, isto é, a seleção dos respondentes é condicionalmente independente de $y_{kj} | \mathbf{x}_k \forall k \in U$.

Alternativamente, o vício se anula quando $\mathbf{c}' \mathbf{x}_k = 1/r_k \forall k \in U$, sendo \mathbf{c} um vetor de constantes.

Isto sugere que a escolha das variáveis x a serem consideradas para a calibração deve levar em conta os prováveis efeitos da não resposta e, em particular, deve-se incorporar todas as variáveis x para as quais dados populacionais auxiliares estão disponíveis e que contêm informações sobre as probabilidades desconhecidas de responder à pesquisa. Sob um modelo simplificado onde δ_k denota a probabilidade de que uma unidade responderá à pesquisa, dado que ela é selecionada na amostra, e a resposta é independente para unidades distintas, uma condição para o estimador de calibração ser aproximadamente não viciado sob a amostragem conjunta e distribuição de resposta é que $\delta_k^{-1} = 1 + q_k \mathbf{x}_k \boldsymbol{\lambda}$, para todo k e algum vetor de constantes $\boldsymbol{\lambda}$ (ver Lundström (1997), p. 46). No entanto, como os pesos de calibração w_k são sempre da forma $w_k = d_k F(q_k \mathbf{x}_k \boldsymbol{\lambda})$ (ver Equação (14.36)). É fácil ver que a calibração levará a estimativas aproximadamente não viciadas quando $w_k = d_k \delta_k^{-1}$, uma condição que depende tanto da escolha das variáveis x quanto da forma da função de distância (ou calibração) usada para obter os pesos de calibração.

Exemplo 14.2 Um exemplo em que esta questão pode ser bem ilustrada vem da ponderação realizada para a pesquisa de força de trabalho do Reino Unido (UK Labour Force Survey / UK-LFS) (ver Statistics (2001), seção 9). A não resposta foi da ordem de 22% na primavera de 1998, sendo 16% de recusas e 6% de ‘falhas de contacto’. Há evidência de *não resposta diferencial detectada* usando estudo de ligação com dados do Censo de 1991 (Statistics (2001)). A tabela 14.6 ilustra grupos sub-representados e super-representados na UK-LFS.

Tabela 14.6: Identificação dos grupos sub-representados e super-representados na pesquisa de força de trabalho do Reino Unido

Grupos sub-representados	Grupos super-representados
Moradores de Londres	Domicílios com 3 ou mais crianças
Locatários de associações imobiliárias ou empregados	Domicílios com criança mais jovem entre 0 e 4 anos de idade
Moradores de habitações compartilhadas ou convertidas	Chefes com escolaridade superior ou qualificação equivalente
Moradores em domicílios com 3 ou mais automóveis ou vans disponíveis	Domicílios com 2 adultos e crianças
Domicílios unipessoais com 1 adulto de idade entre 16 e 19 anos	Casais com crianças dependentes
Chefes com idades entre 16 e 24 anos	
Chefes solteiros	
Chefes nascidos em países da 'New Commonwealth'	

A não resposta diferencial na UK-LFS depende de características como: composição do domicílio; situação ocupacional; condição de ocupação do domicílio; tipo de acomodação; situação socioeconômica; região de residência; região de origem; estado civil, idade e escolaridade do chefe.

Nesta pesquisa, a ponderação leva em conta a *região de residência* (17 Regiões ou 454 Autoridades Locais), a *idade* (11 faixas etárias para jovens e 17 faixas de 5 anos) e o *sexo* dos indivíduos amostrados. Estas são variáveis para as quais há informações disponíveis de qualidade aceitável para os totais populacionais.

O número de variáveis x usadas para a calibração é razoavelmente grande (1.002) e a função de distância escolhida é do tipo 2 na Tabela #tab:tabcalib4, correspondendo aos pesos obtidos pelo 'raking ratio estimator'. No entanto, um estudo sobre a incidência de não resposta nesta pesquisa mostrou que a não resposta não é completamente aleatória, conforme indicado na tabela da página 43 do UK-LFS User Guide (Statistics (2001)).

Claramente, então, pode-se ver que a calibração apenas na idade, sexo e região, como é atualmente o caso, não pode esperar eliminar todo o vício devido à não resposta. Não é o número de variáveis x que importa, mas sim as variáveis x corretas! Claro que isso é mais

fácil falar do que fazer, e no caso do UK-LFS, claramente há dificuldades. Por exemplo, se a não resposta for dependente da situação ocupacional, pode-se tentar testar e calibrar informações externas fornecidas por fontes baseadas em registros, como a 'claimant count'. Para muitas das outras variáveis, no entanto, informações confiáveis de população auxiliar podem estar indisponíveis, difíceis ou custosas de serem obtidas.

A mensagem aqui é que não é suficiente calibrar 'tudo o que está disponível' para estar livre de vícios. Ainda mais, Gambino (1999) sugere que, em alguns casos, a calibração pode até piorar, e argumenta que 'é bem sabido que, em muitas pesquisas, os jovens do sexo masculino tendem a não ser entrevistados proporcionalmente. Como as estimativas demográficas por idade-sexo são normalmente usadas na calibração, o efeito é aumentar os pesos dos jovens do sexo masculino que respondem à pesquisa. Se, para algumas variáveis de interesse, os jovens do sexo masculino que tendem a se tornar respondentes diferem substancialmente dos jovens do sexo masculino que tendem a não responder ..., então o efeito da calibração é dar mais peso a um componente não representativo da amostra'. Suponhamos que soubéssemos que os jovens do sexo masculino que mais tarde seriam não investigados são aqueles que moram sozinhos, e aqueles que são mais propensos a responder são aqueles que vivem com os pais ou a família. Portanto, a ponderação deve ter como objetivo aumentar os pesos daqueles no primeiro grupo (jovens do sexo masculino vivendo sozinhos), mas não daqueles do segundo grupo (vivendo com a família). A informação crucial que precisaríamos fazer seria o total da população por idade e sexo e composição do domicílio (domicílios de uma única pessoa versus outros domicílios). Se esta informação não estiver disponível, ainda há algum recurso limitado a ser usado. A ponderação pode ser realizada ao nível do domicílio e não ao nível individual. Portanto, os jovens do sexo masculino que moram sozinhos teriam pesos que dependem do tipo de domicílio em que vivem, mas essa não seria a correção de vício que estaríamos almejando, apenas o melhor substituto dada a disponibilidade dos dados.

O exemplo acima ilustra as questões que se deve abordar em relação ao uso da calibração para compensar a não resposta. Se o mecanismo de resposta depende das características dos domicílios (além da região de residência), como o seu tamanho e composição, bem como as do chefe, então talvez o domicílio seja a unidade para a qual os pesos são calculados, com os membros individuais a receber o mesmo peso do domicílio.

Skinner (1999) discute calibração na presença da não resposta. Entre as consequências e problemas, destacam-se:

- pesos negativos ou extremos podem ocorrer com muito mais frequência;
- pesos calibrados (w_k) podem não convergir para os pesos originais (d_k) do desenho;

- a variância do estimador de calibração depende da escolha da função de distância adotada na calibração (função G ou F);
- calibração pode mascarar efeitos da não resposta ou falhas de cobertura.

Veja também Lundström (1997), Särndal e Lundström (1999) e Särndal e Lundström (2005).

Gambino (1999) adverte que, para o ajuste de não resposta, ‘a má escolha de variáveis ou classes de ajuste pode piorar’, e conclui que ‘é nosso dever como estatísticos trabalhar com os usuários para garantir que as ferramentas de calibração sejam usadas com sabedoria’. Uma razão pela qual a calibração pode piorar é porque ela pode mascarar os efeitos da não resposta. Por exemplo, usando pesos amostrais não ajustados para estimar as contagens da população por idade e sexo, pode-se localizar as células para as quais as estimativas estão abaixo do nível esperado por um valor que é muito grande para ser devido ao erro de amostragem. Estas são as células cujos elementos têm maior probabilidade de não responderem a pesquisa. Tais estimativas poderiam então ser usadas para detectar as células para as quais os efeitos prováveis da não resposta são mais altos. Mas se as estimativas são calculadas apenas com pesos calibrados, tais desvios da contagem esperada ou conhecida não aparecerão. Seria necessário esforço extra para calcular as estimativas de pré-calibração necessárias para analisar os efeitos prováveis da não resposta, se os pesos

$d_k = \frac{1}{\pi_k}$ forem disponibilizados juntamente com os dados da pesquisa.

Logo, é importante a escolha das variáveis auxiliares que serão usadas para calibração.

14.7 Seleção de variáveis para calibração com não resposta

Silva e Skinner (1997) propõem procedimento de seleção passo a passo para inclusão de variáveis na calibração, levando em conta a não resposta. Tal critério de seleção é baseado em estimativa do EQM .

Särndal e Lundström (2008) propõem critério para seleção das variáveis de calibração baseado apenas no vício.

Em muitas aplicações práticas, as variáveis auxiliares são categóricas, e são usadas para *pós-estratificação da amostra*.

Nesse caso, quando são conhecidas as contagens populacionais numa tabela cruzada definida pelas variáveis auxiliares, o número de pós-estratos pode crescer muito rapidamente.

Exemplo 14.3 Bethlehem e Keller (1987) consideraram um caso em que cinco variáveis estavam disponíveis: sexo (2 categorias), idade (10 categorias), estado civil (4 categorias), região (11 categorias) e grau de urbanização (6 categorias).

Cruzando todas as 5 variáveis para definir os pós-estratos, o número de celas na tabela resultante seria de 5.280. Em termos do modelo de regressão, isto significaria ter que estimar 5.280 parâmetros!

Neste exemplo, o *modelo saturado* é representado como: sexo x idade x estado civil x região x grau de urbanização

Em casos como este, uma ideia atraente é a de considerar *calibração nas marginais*, o que levará a modelos com menor número de parâmetros.

Como alternativa, os autores propõem considerar a ideia de *pós-estratificação multivariada incompleta*.

A ideia é usar classificações incompletas, similares ao ajuste de modelos log-lineares, para representar as tabelas completas.

Um exemplo citado pelos autores considera o modelo: (sexo x idade x estado civil)+(região x grau de urbanização)

Neste exemplo, o modelo conteria ‘somente’ $80 + 66 = 146$ parâmetros.

Uma consequência natural é que a amostra não será calibrada para as 5.280 contagens conhecidas, mas somente para as 146 contagens consideradas no modelo.

Mas os autores sugerem que, na prática, o modelo incompleto servirá quase tão bem quanto o modelo completo.

Silva e Skinner (1997) se inspiraram neste artigo para desenvolver o procedimento de seleção de variáveis para uso com estimadores de regressão. A Tabela 14.7 mostra uma análise dos autores sobre a aplicação da ideia a uma pesquisa domiciliar holandesa realizada em 1977/1978, onde *LI* e *LS* representam os limites inferior e superior, respectivamente, dos intervalos com 95% de confiança; e a variável “Mar” representa o estado conjugal.

Tabela 14.7: Estimativas do rendimento médio domiciliar por esquema de ponderação da pesquisa domiciliar da Holanda 1977/1978

Esquema de ponderação	Número de parâmetros	Estimativa	Erro Padrão	<i>LI</i>	<i>LS</i>
Nenhum	1	23.494	182	23.137	23.852
Sexo	2	23.613	182	23.137	23.852
Idade	6	23.990	170	23.657	24.323
Mar	2	23.624	161	23.308	23.940
Sexo x Idade	12	24.012	167	23.684	24.340
Sexo + Idade	7	24.065	168	23.736	24.349
Sexo x Mar	4	23.809	160	23.496	24.123
Sexo + Mar	3	23.675	160	23.361	23.990
Idade x Mar	12	23.987	153	23.687	24.287
Idade + Mar	7	24.071	154	23.769	24.374
Sexo + Idade + Mar	8	24.104	154	23.802	24.405
(Sexo x Mar) + Idade	9	24.172	153	23.871	24.472
(Sexo x Idade) + Mar	13	24.078	153	23.777	24.379
(Idade x Mar) + Sexo	13	24.004	152	23.705	24.302
(Sexo x Idade) + (Sexo x Mar)	14	24.149	153	23.849	24.449
(Idade x Mar) + (Sexo x Mar)	14	24.076	152	23.778	24.374
(Sexo x Idade) + (Idade x Mar)	18	23.985	152	23.687	24.283
(Sexo x Idade) + (Idade x Mar) + (Sexo x Mar)	19	24.054	152	23.757	24.352
Sexo x Idade x Mar	24	24.048	152	23.751	24.345

14.8 Indicadores de qualidade da calibração

Um componente importante de qualquer análise estatística ou trabalho de estimação é a avaliação de quão bem os procedimentos adotados foram executados na aplicação em questão. Com as aplicações de estimadores de calibração, além da estimação usual de variância que deve ser realizada rotineiramente, sugerimos que também é importante avaliar vários outros aspectos do resultado. Isso é importante para verificar os objetivos pretendidos da calibração foram atingidos e para verificar os possíveis problemas que podem afetar o resultado.

Portanto, o emprego de estimadores de calibração deve ser acompanhado da obtenção de alguns *indicadores de qualidade*, que ajudem a avaliar o resultado da calibração numa aplicação (ver Silva (2004)).

Como primeira medida, sugerimos examinar a quantidade de erro de calibração restante para o conjunto completo de variáveis x que foram inicialmente selecionadas para calibração. Portanto, um primeiro indicador útil é o *erro absoluto médio* de calibração para os totais estimados das variáveis auxiliares x , *expresso em porcentagem*, dado por:

$$IQ1 = 100 \times \frac{1}{Z} \sum_{j=1}^Z |\hat{T}_{x_j}^C - T_{x_j}| / T_{x_j}$$

Num cenário ideal este indicador terá valor nulo, quando a calibração for bem sucedida em eliminar os erros de calibração para todas as variáveis auxiliares. Mas isto pode não ocorrer se algumas das variáveis x forem descartadas durante o processo de determinação dos pesos de calibração ou se algumas das abordagens que não levam à calibração exata forem adotadas.

Um segundo indicador que pode ajudar a avaliar o potencial de vício do estimador de calibração é a média dos coeficientes de variação dos estimadores HT para totais das variáveis auxiliares x , a saber:

$$IQ2 = 100 \times \frac{1}{Z} \sum_{j=1}^Z \left[\hat{V}_p(\hat{T}_{x_j}) \right]^{1/2} / \hat{T}_{x_j}$$

onde: $\hat{V}_p(\hat{T}_{x_j})$ é uma estimativa da variância do estimador HT para o total da j -ésima variável auxiliar, obtida por:

$$\hat{V}_p(\hat{T}_{x_j}) = \sum_{k \in S} \sum_{i \in S} (1 - \pi_k \pi_i / \pi_{ki}) (d_k x_{kj}) (d_i x_{ij})$$

Três outros indicadores que podem ser de interesse são baseados na distribuição dos fatores de ajuste dos pesos. Dois deles são as proporções de pesos g extremos (pequenos ou grandes), em que é necessário definir valores de limites inferior (LI) e superior (LS) aceitáveis para os fatores de ajuste dos pesos. Valores típicos são 1/5 e 5, respectivamente.

Porcentagem de pesos g *muito pequenos*:

$$IQ3 = 100 \times \frac{1}{n} \sum_{k \in s} I(g_k < LI)$$

Porcentagem de pesos g *muito grandes*:

$$IQ4 = 100 \times \frac{1}{n} \sum_{k \in s} I(g_k > LS)$$

Coeficiente de variação dos pesos g :

$$IQ5 = 100 \times \left[\frac{\sum_{k \in s} (g_k - \bar{g})^2}{(n-1)} \right]^{-1/2} / \bar{g}$$

onde:

$\bar{g} = \sum_{k \in s} g_k / n$ é o fator de ajuste médio (que deve ficar próximo de 1).

Outro indicador útil é a *distância* entre os pesos ajustados e os pesos básicos, dada por:

$$IQ6 = \frac{1}{n-Z} \sum_{k \in s} (w_k - d_k)^2 / q_k d_k = \frac{1}{n-Z} \sum_{k \in s} d_k (g_k - 1)^2 / q_k$$

Observe que normalizamos essa distância dividindo a função de distância de calibração pelo tamanho da amostra, para facilitar a comparação entre amostras de tamanhos diferentes. Outra possibilidade seria dividir pelo tamanho da amostra menos o número de variáveis x consideradas, o que permitiria conjuntos de variáveis auxiliares de tamanhos diferentes. Sugerimos que a função de distância qui-quadrado seja usada mesmo quando a função de distância que foi minimizada para obter os pesos de calibração é uma das outras funções na Tabela #tab:tabcalib4.

Também é de interesse avaliar a *eficiência média* obtida em comparação com o estimador HT para algum conjunto de variáveis de pesquisa. Isso pode ser feito comparando a eficiência média de um conjunto especificado de variáveis y , dadas por:

$$IQ7 = 100 \times \frac{1}{J} \sum_{j=1}^J \hat{V}_g(\hat{T}_{y_j}^C) / \hat{V}_p(\hat{T}_{y_j})$$

onde as variâncias são estimadas usando:

$$\hat{V}_g(\hat{T}_{y_j}^C) = \sum_{k \in s} \sum_{i \in s} (1 - \pi_k \pi_i / \pi_{ki}) (d_k g_{kj}) (d_i g_{ij})$$

$$\hat{V}_p(\hat{T}_{y_j}) = \sum_{k \in s} \sum_{i \in s} (1 - \pi_k \pi_i / \pi_{ki}) (d_k y_{kj}) (d_i y_{ij})$$

Sarndal e Lundström (2010) sugere usar indicadores relacionados com a redução do vício no contexto em que os estimadores de calibração são empregados para compensar vício causado por não resposta:

$$IQ8 = 100 \times (\hat{T}_{y_j}^C - \hat{T}_{y_j}) / \hat{T}_{y_j}^C$$

Usar esse conjunto de indicadores e quaisquer outras medidas que possam fornecer informações sobre os mesmos aspectos é altamente recomendável para usuários de estimadores de calibração.

14.9 Um guia para aplicação de calibração

Comece por identificar claramente que variáveis auxiliares estão disponíveis para calibração, e obtendo todas as informações requeridas sobre estas variáveis.

Os passos seguintes podem ser distintos, dependendo do objetivo principal estabelecido para a calibração. Decida PORQUE calibração é necessária ou desejável.

- Para redução de variância: - Escolha o(s) principal(is) alvo(s) de inferência;
 - Guie a escolha de variáveis auxiliares considerando os alvos de inferência escolhidos;
 - Se houver muitas variáveis auxiliares, considere um procedimento de seleção de variáveis guiado por estimativas da variância do estimador de calibração;
 - Comece considerando o estimador de regressão, sem limites para os fatores de ajuste;
 - Lembre que nem sempre usar o maior número de variáveis auxiliares leva a menor variância;
 - Avalie os resultados da calibração usando indicadores.

- Para redução de vício devido a não resposta:
 - Concentre-se na inclusão de variáveis auxiliares que ajudem a explicar o mecanismo de resposta;
 - Se houver muitas variáveis auxiliares, considere um procedimento de seleção de variáveis guiado por estimativas do *EQM* ou do vício do estimador de calibração;
 - Comece considerando o estimador de regressão, sem limites para os fatores de ajuste;
 - Avalie os resultados da calibração usando indicadores.
- Ideias aplicáveis nos dois casos:
 - Para variáveis auxiliares categóricas, considere subconjuntos baseados em *calibrar nas marginais*, em vez de classificações cruzadas (*pós-estratificação multivariada incompleta*);
 - Para pesquisas que empreguem amostragem conglomerada, considere adicionar restrições de que os pesos sejam iguais para todas as unidades elementares de um conglomerado. Por exemplo: considere o emprego de '*ponderação integrada por domicílios*'.
 - Para pesquisas repetidas, pode ser útil considerar o *tempo na pesquisa* para a calibração.

14.10 Observações finais

Neste capítulo, revisamos a literatura sobre estimação de calibração e tentamos transmitir a mensagem de que é uma abordagem flexível, poderosa e conveniente para ponderação e estimação em pesquisas amostrais. Ao mesmo tempo, apontamos algumas das dificuldades que os usuários da técnica podem enfrentar em aplicações práticas, além de fornecer algumas orientações sobre como elas podem ser detectadas e contornadas.

O valor da estimação de calibração para situações práticas de pesquisa é claramente demonstrado pelo número crescente de pacotes de software desenvolvidos para sua aplicação, bem como pelo número de pesquisas importantes em vários países que dependem de calibração para sua ponderação e estimação. Em países como o Canadá, os Estados Unidos, o Reino Unido e a França, a estimação de calibração é usada para as pesquisas sobre força de trabalho. No Brasil e no Canadá, também é utilizado para as amostras coletadas com o questionário ampliado como parte dos Censos de População.

Embora esse valor seja reconhecido, incentivamos os usuários a criticar o resultado da ponderação de calibração e enfatizamos a importância de uma verificação cuidadosa dos pesos e estimativas resultantes, para verificar se eles atendem aos critérios de desempenho e altos padrões de qualidade e confiabilidade que é esperado da pesquisa.

A calibração não deve ser usada para ‘disfarçar’ resultados de pesquisas com viés, nos quais a cobertura e o vício de não resposta são ‘cobertos’ por meio de calibração simples para os totais de população conhecidos. Os usuários da técnica devem, acima de tudo, procurar ser transparentes sobre a aplicação da metodologia e da avaliação de seu sucesso, a fim de proporcionar aos usuários dos resultados da pesquisa a possibilidade de exercer uma avaliação crítica da adequação à finalidade do resultado das estatísticas produzidas com estimadores de calibração.

Finalmente, encorajariamos os produtores de dados que optassem por adotar a estimação de calibração em pesquisas onde os arquivos de microdados seriam liberados, para fornecer aos usuários as informações necessárias para que eles fizessem o uso adequado dos dados, no sentido de serem capazes de calcular as estimativas da variância corretamente. Isso é muito mais desafiador do que quando os estimadores HT usuais são usados, diante do fato que as informações sobre o conjunto completo de variáveis x usadas para calibração também devem ser liberadas juntamente com os pesos d originais e os pesos w finais para cada registro da pesquisa. Uma investigação cuidadosa ainda é necessária sobre a melhor forma de atingir esse objetivo sem sacrificar as restrições necessárias para a proteção da confidencialidade.