

Quotas em pesquisa online: que são e como usá-las?

Obter uma amostra representativa do universo que queremos estudar (o que se conhece como “amostragem”), é uma condição necessária para obter resultados confiáveis em uma pesquisa.

A forma mais direta de obter amostras representativas é a amostragem aleatória simples (M.A.S.). Neste tipo de amostragem, todos os indivíduos que constituem o universo estudado têm igual probabilidade de ser selecionados para fazer parte da amostra. Seria algo bem como pôr a todos os indivíduos que formam o universo numa urna e ir extraindo por casualidade indivíduos até completar o tamanho de mostra desejado. O M.A.S. garante a ausência de riscos no processo de seleção da mostra.

Mas na prática é extremamente difícil usar este método de amostragem. Para poder usar M.A.S. requerem-se várias condições: a primeira é dispor de um marco amostral de referência confiável, isto é, uma listagem completa dos indivíduos que constituem o universo. Também precisamos acesso a ditos indivíduos e a predisposição dos mesmos a participar. Um indivíduo que recuse participar numa enquete seria suficiente para invalidar um processo de amostragem aleatória perfeito.

Amostragem por quotas

As condições anteriores são muito exigentes, fazendo que na prática o M.A.S. seja impossível ou extremamente caro. Por isso, costumam usar-se outras técnicas de amostragem, como a amostragem por quotas.

A amostragem por quotas pretende obter uma amostra representativa a partir de uma seleção de indivíduos não necessariamente aleatória, forçando uma distribuição em relação a umas variáveis concretas idêntica à distribuição do universo estudado.

Para usar amostragem por quotas é necessário ter em conta duas coisas:

- (1) Precisamos selecionar as variáveis relevantes para o objeto de nosso estudo. Por exemplo, se estamos desenhando um estudo eleitoral, variáveis como a “idade”, “localização geográfica” ou “classe social”, são variáveis relevantes porque podem condicionar a orientação do voto. Dito de outro modo, “idade” e “orientação do voto” são variáveis dependentes. Pelo contrário, a “estatura” ou o “nome” do indivíduo são variáveis irrelevantes num estudo eleitoral.
- (2) O segundo fator a considerar é que precisamos conhecer a distribuição de ditas variáveis relevantes no universo estudado. Caso contrário, não poderemos reproduzir dita distribuição na mostra. Por exemplo, se queremos fazer um estudo sobre população geral e desejamos fixar uma quota sobre a variável “idade”, podemos recorrer a estudos do CENSO e outras fontes de informação proporcionados pelos organismos oficiais de cada país. Estes estudos se realizam, ou bem empregando acesso direto a todos os indivíduos do universo estudado, ou bem mediante um

método de amostragem muito próxima ao M.A.S.. Ambas as coisas nos garantem a confiabilidade da informação que trabalhamos para o universo.

Quantas e quais as variáveis devo empregar em minhas quotas?

Fixar quotas em um desenho amostral tem custo. Com independência da metodologia empregada (cara a cara, telefônico ou online), agregar quotas encarece o trabalho de campo já que obriga a descartar possíveis participantes quando excedem um objetivo que nos fixamos. É por isso, que para selecionar variáveis relevantes devemos ter em conta as seguintes considerações:

- (1) Algumas variáveis relevantes podem ser redundantes (isto é, altamente correlacionadas), pelo que controlar só uma delas pode ser suficiente. Por exemplo, se fixamos uma quota sobre a variável “classe social” possivelmente possa prescindir de outras variáveis como “nível de rendimentos”, “metros quadrados da moradia” ou “nível de estudos”.
- (2) (2) Se a forma em que seleciono indivíduos para a amostra, pese a não ser puramente aleatória (M.A.S.), sem que garanta a aleatoriedade com respeito à variável relevante, posso prescindir de incorporá-la às quotas. Por exemplo, poderia estar obtendo uma amostra entre os leitores de uma página web de notícias esportivas, e pese a que claramente não teria uma mostra aleatória (todos os indivíduos são usuários de Internet, aficionados ao esporte e leitores do site em questão), poderia dar-se o caso de que geograficamente os leitores desta página web se distribuíssem de forma idêntica ao universo estudado. Neste caso não precisaria fixar quotas pela variável “região”.

Considerando os dois critérios anteriores, deveríamos selecionar as variáveis realmente essenciais que permitam garantir um bom nível de representatividade da mostra sem encarecer excessivamente o estudo.

Quotas cruzadas ou quotas não cruzadas?

Quando realizamos um controle de quotas sobre duas ou mais variáveis, podemos definir dito controle de forma cruzada ou não cruzada. Vejamos um exemplo: suponhamos que queremos obter uma amostra de 1.000 pessoas para um estudo eleitoral e identificamos duas variáveis relevantes: sexo (50% homens e 50% mulheres) e idade (50% menores de 40 anos e 50% maiores de 40 anos).

Se impomos um controle de quotas não cruzadas, exigiremos que do total de indivíduos da mostra (1.000), 500 sejam homens e 500 mulheres, e que 500 sejam menores de 40 anos e 500 maiores. Dito de outra forma, uma mostra de 500 homens menores de 40 anos e 500 mulheres maiores de 40 anos seria válida.

Ao contrário, se definimos uma quota cruzada sexo-idade, exigiremos que a mostra se componha de 250 homens menores de 40, 250 homens maiores de 40, 250 mulheres menores de 40 e 250 mulheres maiores de 40 .

Definir quotas cruzadas é mais caro que definir quotas não cruzadas. Para saber se é necessário cruzar quotas, devemos valorizar novamente relevância e método de seleção dos indivíduos:

- (1) Influem as variáveis no objeto do estudo de idêntica forma ao ser consideradas por separado ou conjuntamente? Se duas variáveis relevantes são independentes entre elas, poderemos considerá-las como “não cruzadas”. No exemplo anterior, se a “idade” influi na “orientação do voto” do mesmo modo para homens que para mulheres, poderíamos empregar quotas “não cruzadas”.
- (2) A forma em que seleciono indivíduos para minha amostra pode favorecer a sobre-representação de algum dos grupos definidos pela combinação de variáveis relevantes? Se a forma em que seleciono indivíduos garante por si só, que as proporções de ambas variáveis cruzadas é a correta, poderia renunciar a definir este controle.

Quotas offline e quotas online

As variáveis mais empregadas para definir quotas em qualquer metodologia são as sociodemográficas: “sexo”, “idade”, “região”... Costumam ser variáveis com influência na maior parte de temáticas objeto de estudo, e são facilmente controláveis, ao existir fontes de dados confiáveis com as quais comparar.

No entanto, cada metodologia tem suas particularidades que aconselham adaptar o uso das quotas.

Quando fazemos estudos mediante entrevista pessoal, é frequente simplificar as quotas geográficas pelo enorme custo que teria deslocar entrevistadores a todos os povos e cidades de um país. É por isso que costumam fixar-se umas quotas para as principais cidades, assumindo que ditas cidades representam perfeitamente o comportamento dos indivíduos de uma região ampla. Por exemplo, numa amostra para Brasil, é frequente solicitar um número determinado de indivíduos para São Paulo, Rio de Janeiro e Recife. Para México, usaríamos Distrito Federal e Monterrey. Ou para Espanha, Madri e Barcelona.

Ao fazer estudos telefônicos, é especialmente relevante considerar a variável “ocupação”, já que ao contatar com pessoas que se encontram em seu lar, poderíamos estar sobre-representando indivíduos desocupados e retirados.

O uso de quotas em estudos online está ainda pouco estudado. Uma prática maioritária – e possivelmente errônea – é empregar as mesmas quotas que se definiram quando o estudo se fazia offline.

A seguir, listam-se uma série de considerações a ter em conta ao usar Internet na obtenção de mostras:

- (1) A maioria das pesquisas online empregam painéis: É importante recordar como funcionam os painéis e com que critérios recrutam a seus membros. Por exemplo: os painéis tratam de pesquisar a todos seus membros por igual. Por isso, se alguns perfis concretos são mais demandados nos estudos (mamães, pessoas com altos rendimentos, etc.) podem estar sobre-representados. Um painel online não é representativo da população por si mesmo, é um lugar que pode definir amostras representativas usando as quotas adequadas.
- (2) A dispersão geográfica não é um problema em online, ao invés, é uma oportunidade. Se queremos obter mostras representativas de uma região, não precisamos limitar-nos a uma cidade representante da mesma como fazíamos em offline. Fazê-lo reduz desnecessariamente a representatividade e encarece o estudo, já que os painéis lhes resulta mais natural recrutar membros de regiões amplas do que de uma cidade específica.
- (3) Todos os indivíduos usam Internet: É uma obviedade, mas é necessário tê-lo em conta. Não podemos fixar quotas sobre “não usuários de Internet” ou sobre “analfabetos”: em Internet todo mundo sabe ler. Assim mesmo, o fator Internet pode correlacionar fortemente com outras variáveis como o uso de tecnologias, o qual aconselharia fixar quotas sobre este tipo de variáveis. Por exemplo, se estamos fazendo um estudo sobre telefonia móvel, e o tipo de contrato (contrato/pré-pago) pode influenciar nos resultados que esperamos, em um painel online podemos encontrar sobre-representadas as modalidades de contrato próprias de usuários mais avançados (acesso à Internet móvel).
- (4) A adoção de Internet em cada país é chave para definir boas quotas: Se um grupo de população adotou Internet em maior medida do que outro, sua presença em amostras online estará sobre-representada. Variáveis como “idade” e “classe social” são fundamentais, especialmente em América Latina, uma região que existem diferenças sociais especialmente evidenciadas, que fazem que as classes mais baixas acedam com muita maior dificuldade a Internet.

Definitivamente, quando fazemos uma pesquisa online, as quotas são ferramentas imprescindíveis para obter representatividade, e a seleção das variáveis que empregamos nelas deve realizar-se atendendo ao problema objeto de estudo bem como à forma em que obtemos a amostra.