

# PLANO AMOSTRAL

## Pesquisa de Emprego e Desemprego

WILTON DE OLIVEIRA BUSSAB

NÁDIA PINHEIRO DINI

SILVIA REGINA MANCINI

---

*Resumo:* Discussão dos aspectos principais do plano amostral de uma pesquisa domiciliar, a PED – Pesquisa de Emprego e Desemprego, realizada mensalmente pela Fundação Seade e pelo Dieese desde 1985. Abordam-se questões metodológicas, teóricas e práticas e as soluções para elas adotadas, bem como procura-se avaliar a contribuição da realização dessa pesquisa para o desenvolvimento da Fundação Seade.

*Palavras-chave:* plano amostral; pesquisa domiciliar; painel rotativo.

*Abstract:* A discussion of the principle aspects of the sampling plan of a household survey, the PED – Survey of Employment and Unemployment, conducted monthly by Fundação Seade and Dieese since 1985. Methodological, theoretical and practical issues, as well as adopted solutions, are addressed. An attempt is made to evaluate the contribution of this survey to the development of Fundação Seade.

*Key words:* sampling plan; household survey; rotating panel.

---

A PED surgiu em 1984, a partir de um convênio firmado entre a Fundação Seade – Sistema Estadual de Análise de Dados – e o Dieese – Departamento Intersindical de Estatística e Estudos Sócio-Econômicos, com o objetivo principal de acompanhar a inserção da população em idade ativa – PIA no mercado de trabalho da Região Metropolitana de São Paulo, caracterizado por grande heterogeneidade e por tênues limites entre as possíveis condições de atividade dos indivíduos (ocupação, desemprego e inatividade).

Desde seu início, a pesquisa buscou ser fonte de informações que subsidiem a formulação de políticas econômicas e sociais, em especial de emprego e de renda, através da geração de indicadores referentes à medição de desemprego, às características dos postos de trabalho e aos rendimentos do trabalho, entre outros (Troiano, 1990).

Para melhor atender a esses objetivos, a pesquisa é realizada em caráter contínuo e através de entrevistas diretas com a população da Região Metropolitana de São Paulo. Selecionam-se, para tanto, domicílios particulares de maneira probabilística, de acordo com um plano amos-

tral predefinido, e entrevistam-se todos os moradores desses domicílios.

### CARACTERÍSTICAS DO PLANO AMOSTRAL

A descrição de um plano amostral probabilístico deve especificar o universo de investigação, as unidades amostrais, os critérios de estratificação, os procedimentos de sorteio das unidades amostrais, as probabilidades de inclusão, os estimadores e os respectivos erros amostrais. Desse modo, saberemos do que e de quem estamos falando e avaliando os desvios esperados para as estimativas (Bolfarine; Bussab, 2000).

Outros aspectos e decisões operacionais também precisam ser considerados a fim de se obter um planejamento amostral eficiente, em relação a custo e precisão, e bem ajustado aos propósitos da investigação a ser realizada. Esses objetivos só serão alcançados com um estudo detalhado das informações e recursos disponíveis para a realização da pesquisa. Após a identificação e conhecimento do cenário à disposição, pode-se escolher o plano amostral e respectivos estimadores que melhor respondam aos interesses do levantamento.

## Universo da Pesquisa

A *população-alvo* ou *universo* inicial de interesse era a população em idade ativa moradora na área urbana da Região Metropolitana de São Paulo, que em 1984 era constituída por 37 municípios. Também tinha-se como acordado que as informações seriam obtidas nos domicílios particulares que, portanto, constituiriam a última unidade amostral, qualquer que fosse o plano amostral adotado.

## Sistema de Referência

O uso de amostras probabilísticas exige uma listagem das unidades amostrais, ou seja, um sistema de referência das unidades amostrais. Em 1984, momento de implantação da pesquisa, consideraram-se, como possíveis fontes de informação, os cadastros telefônicos, de suprimento de água ou ainda de energia elétrica. Além da dificuldade de acesso a esses cadastros, eles excluía parcelas importantes da população urbana da Região Metropolitana de São Paulo, como, por exemplo, aquela moradora em favelas, além de parcelas de domicílios que ainda não estavam cobertos por esses serviços e apresentavam características indesejáveis de identificação, como falta de clareza entre consumidores comerciais e residenciais e outras especificidades. Optou-se, então, por usar as informações fornecidas pelo Instituto Brasileiro de Geografia e Estatística – IBGE – por meio dos Censos Demográficos na Região Metropolitana de São Paulo, detalhadas ao nível menor de Setor Censitário – SC. Tais setores censitários correspondem a regiões geográficas delimitadas com cerca de 300 domicílios cada uma, definidas previamente à realização de cada Censo Demográfico para todo o território nacional. Desse modo, tem-se cobertura total de todos os municípios por setores censitários, com a vantagem adicional de que, em regiões de elevada densidade demográfica, como é o caso da Região Metropolitana de São Paulo, esses setores são geograficamente pequenos, o que facilita muito o trabalho de um entrevistador ao percorrê-los a pé.

A escolha de tal sistema de referência geográfico obriga-nos à utilização de planos amostrais em múltiplos estágios, modelo bastante difundido em pesquisas amostrais domiciliares. Assim, em uma primeira etapa serão selecionados SCs e dentro deles serão sorteados domicílios.

O cadastro de setores censitários urbanos do Censo Demográfico de 1980 do IBGE constituiu o sistema de referência original por ocasião da implantação da pesqui-

sa. Esse sistema é atualizado à medida que as informações dos censos mais recentes são disponibilizadas. Independentemente da atualização do cadastro de setores através do Censo, procede-se regularmente a uma renovação dos setores que compõem a amostra, selecionando-se novos setores para substituí-los. Mensalmente, uma pequena parcela dos setores da amostra é substituída, o que mantém atualizada e ao mesmo tempo evita impactos bruscos às séries de dados.

## Organização do Cadastro

Conhecer bem e trabalhar o sistema de referência é importante para aumentar a precisão dos resultados de pesquisas amostrais. Um dos recursos bastante comuns para isso é a utilização de estratificação, ou seja, a organização do sistema de referência adotado em subgrupos ou estratos de tal forma que esses estratos sejam ao mesmo tempo bastante homogêneos internamente quanto a certas características relacionadas ao que se busca medir, e tão heterogêneos quanto possível entre si, no que se refere a essas mesmas características.

No caso da PED, um dos indicadores cujo erro se pretendia controlar é a taxa de desemprego, que serve de base para todo o planejamento da amostra. As informações disponíveis no Censo Demográfico mais relacionadas a esse indicador são aquelas referentes à inserção da mão-de-obra no mercado de trabalho, tais como a distribuição dos ocupados por setor de atividade econômica e o rendimento dos chefes de domicílio, que foram utilizados para a estratificação do sistema de referência. Por meio de técnicas estatísticas multivariadas como a análise de agrupamentos, produziram-se regiões homogêneas de municípios ou distritos (dez regiões em 1980 e sete regiões em 1991) (Bussab; Dini, 1985). Cada uma dessas regiões pode ser caracterizada pela maior ou menor presença de sua mão-de-obra ocupada em cada um dos principais setores de atividade econômica. Como exemplo, um dos resultados da aplicação dessa técnica aos dados de 1980 foi a obtenção de uma região que incluía os principais municípios do ABC paulista, além de Caieiras e de Cajamar, com forte presença de mão-de-obra inserida no setor industrial, refletindo a realidade dessa região naquele momento.

A análise dos tipos de SC existentes no Censo sugeriu eliminar do cadastro os setores rurais e alguns setores especiais, entre eles os quartéis, as cadeias, os asilos e as aldeias indígenas. Desse modo, a população de referên-

cia da pesquisa passou a ser constituída pelos moradores de domicílios particulares em setores censitários urbanos e suburbanos da RMSP, inclusive aqueles de favelas e de alojamentos. Deve-se salientar que a quase totalidade dos domicílios dessa região é considerada urbana (cerca de 98%).

### Determinação do Tamanho da Amostra

Estudos realizados com o objetivo de balancear o custo da pesquisa com a precisão desejada para seus principais indicadores (dos quais considerou-se principalmente a taxa de desemprego desagregada em seus tipos – total, aberto, oculto pelo trabalho precário e oculto pelo desalento) previam amostras de 9 mil unidades domiciliares a serem selecionadas trimestralmente. Após analisar as dificuldades operacionais de treinamento e aplicação de questionários de forma não-contínua, decidiu-se alterar o procedimento inicial de uma única tomada a cada três meses, e dividi-la em três tomadas a cada mês, com cerca de 3 mil domicílios entrevistados em cada tomada. Esse procedimento implicou propor novos parâmetros e estimadores, bem como a decisão do uso de acúmulo de informações para a produção de estatísticas de tendências.

### Amostras Selecionadas com Probabilidade Proporcional ao Tamanho

O principal cuidado a ser tomado com planos amostrais em múltiplos estágios é o controle do tamanho final da amostra, que pode ser conseguido com sorteio do conglomerado (SC) com probabilidade proporcional ao tamanho (PPT) e, no segundo estágio, selecionar um número fixo de domicílios (Bolfarine; Bussab, 2000). Assim, o plano amostral pode ser resumido como a seleção de “a” setores censitários com probabilidade proporcional ao tamanho, e “b” domicílios com igual probabilidade dentro de cada SC sorteado. Usou-se como medida do tamanho do SC o número de domicílios ocupados no Censo Demográfico à disposição. A probabilidade de seleção do domicílio  $i$  dentro do setor  $j$ , passa a ser:

$$P_{ij} = (aN_j/T)(b/N_j) = ab/T$$

$N_j$  = número de domicílios no setor  $j$   
 $T$  = total de domicílios no Censo

Ou seja, com esse procedimento cada domicílio teria a mesma probabilidade de pertencer à amostra, pelo menos teoricamente.

Entretanto, quanto mais a pesquisa se afasta da data do Censo, mais desatualizadas vão ficando as informações dos SCs, o que recomenda a recontagem do número de domicílios dentro dos setores sorteados (relistagem dos SCs). Como o número de domicílios pode crescer ou decrescer, a recontagem usualmente altera o número existente para um novo número  $N_j^*$ , necessitando-se, então, corrigir a probabilidade de seleção. Em tais casos, costuma-se alterar o número de domicílios sorteados dentro dos setores de tal modo que se mantenha a probabilidade de seleção inicial, ou seja,

$$b/N_j = b_j/N_j^*$$

Com essa alteração, perde-se um pouco o controle sobre o tamanho final da amostra, mas de um modo geral não são modificações dramáticas.

Deve-se enfrentar agora a questão de como alocar a amostra pelos SCs e domicílios, ou seja, quantos setores e quantos domicílios selecionar para obter a amostra final de cerca de 3 mil domicílios por mês.

Para melhor definir esses dois números, devem ser considerados aspectos como os custos associados à inclusão de cada setor censitário na amostra e à realização das entrevistas em cada unidade domiciliar e o grau de correlação entre as informações de unidades domiciliares pertencentes a um mesmo setor censitário, a chamada correlação intraclasse.

Quanto maior o número de setores censitários na amostra, maiores as despesas com seu arrolamento e com os deslocamentos dos pesquisadores para a realização das entrevistas.

Por outro lado, espera-se uma forte correlação entre as informações de domicílios dentro de um mesmo setor; assim, quanto maior a correlação entre as informações de unidades domiciliares pertencentes a um mesmo setor censitário, menor deverá ser o número de unidades domiciliares selecionadas por setor censitário. Entretanto, como a correlação é desconhecida até que o levantamento seja de fato efetuado, procura-se, nessa fase, utilizar a experiência proveniente de outras pesquisas e o senso comum para se estabelecer a divisão mais conveniente do tamanho total previsto para a amostra entre setores censitários e unidades domiciliares por setor censitário. Experiência anterior e simulações de custos sugeriram que a alocação da amostra deveria ser feita em 600 SCs e 15 domicílios em cada um, perfazendo um total de 9 mil domicílios trimestralmente. Com essa escolha, a fração amostral passou a corresponder a cerca de um domicílio sorteado para cada 500 existentes na RMSP.

Como já mencionado, a fim de realmente garantir que todos os domicílios tenham a mesma probabilidade de seleção, é necessário que cada um dos setores censitários selecionados seja arrolado por completo e o número esperado de unidades domiciliares para cada setor (15) seja ajustado proporcionalmente às alterações encontradas entre o tamanho presumido do setor e o observado em campo no momento do arrolamento. Dessa forma, setores que apresentem, no arrolamento, um número maior de domicílios do que o divulgado no Censo deverão também ter um acréscimo às 15 unidades domiciliares esperadas, e esse acréscimo deverá ser proporcional ao crescimento observado no setor. Inversamente, setores que apresentem menos domicílios no arrolamento do que no Censo deverão ter um número proporcionalmente menor de unidades domiciliares selecionadas. Tais alterações no tamanho do setor são comuns em uma região dinâmica como a metrópole paulistana, onde são freqüentes fenômenos como o brusco aumento do número de residências graças à verticalização ou a sua redução devido ao surgimento de grandes eixos comerciais em áreas anteriormente residenciais. Assim, devem-se realizar esforços constantes para a manutenção de arrolamentos atualizados.

Quando a implantação da pesquisa ocorre em um momento outro que o do Censo Demográfico, convém que o cálculo inicial do tamanho da amostra leve em consideração as estimativas do crescimento populacional anual ocorrido, evitando-se, assim, a indesejada perda de controle do tamanho final.

### Sistema de Ponderação

O sistema de referência (setores censitários) foi inicialmente agrupado de acordo com as regiões homogêneas definidas. Em seguida, dentro de cada região homogênea, foi ordenado por município e distrito e, dentro destes, por rendimento médio dos chefes de domicílio. A partir do sistema assim ordenado, realizou-se o sorteio de setores censitários em pares aleatórios, o que garantiu uma estratificação implícita da amostra. Isso quer dizer que a distribuição da amostra pelos estratos considerados coincide com aquela da população, sem outro esforço adicional para se obter essa característica.

Além disso, a utilização de uma fração amostral constante, ou seja, o fato de que todas as unidades domiciliares têm a mesma chance de ser incluídas na amostra, garante que os dados sejam autoponderados, não havendo, portanto, a necessidade de criação de pesos para o

processamento dos indicadores. Esse aspecto é bastante vantajoso para uma pesquisa de caráter contínuo, pois, se houver necessidade de pesos, o cálculo exato destes não é trivial e exige informações extras sobre a distribuição populacional, podendo estar sujeito a distorções à medida que a pesquisa se distancia de seu sistema de referência, ou seja, do último Censo Demográfico disponível. Além das facilidades computacionais que advêm da autoponderação, existe outra vantagem adicional expressa em menores erros amostrais. A teoria de amostragem determina que quanto maior a diferença entre as frações amostrais utilizadas e, conseqüentemente, entre os pesos posteriormente atribuídos aos dados, maiores os erros amostrais dos indicadores calculados.

Com a autoponderação, a maioria dos indicadores divulgados pode ser calculada diretamente a partir dos dados amostrais; já para os indicadores de contingentes populacionais preferiu-se adotar um sistema misto: utilizam-se os resultados da amostra aplicados a dados externos, ou seja, projeções populacionais produzidas pela Gerência de Demografia da Fundação Seade.

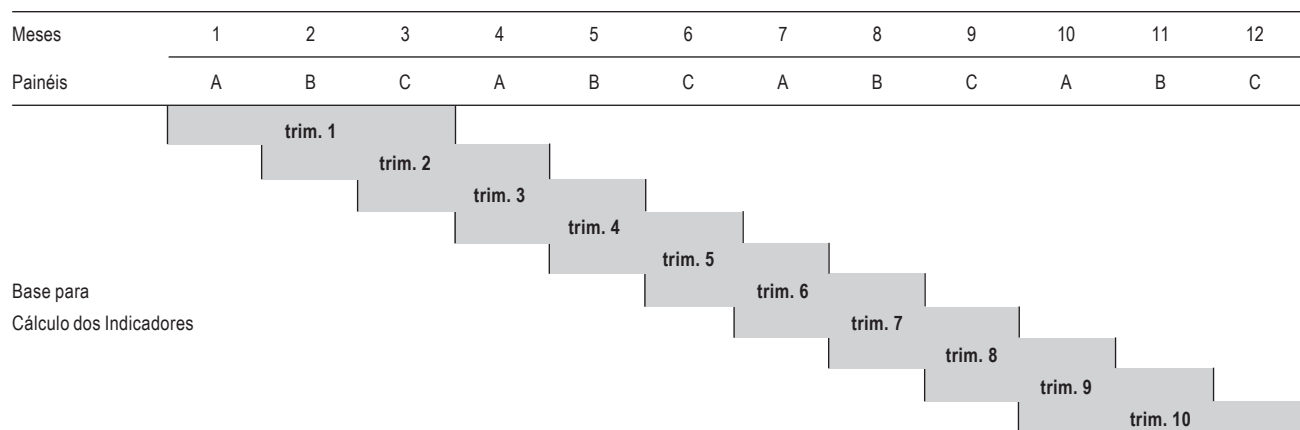
### Periodicidade da Coleta

O tamanho inicial definido para a amostra (9 mil domicílios) por trimestre, por razões já mencionadas de custo, não é levantado em um único mês, mas em três meses, com a coleta de 3 mil domicílios por mês. Para tanto, divide-se o número total de setores censitários sorteados em três painéis distintos rotulados de A, B e C. Nos três primeiros meses da pesquisa, levantam-se as informações dos painéis A, B e C, respectivamente. No quarto, sétimo e décimo mês da pesquisa, utilizam-se novamente os setores censitários do painel A, selecionando-se, entretanto, novas unidades domiciliares nesses setores. No quinto, oitavo e décimo primeiro mês, repetem-se os setores censitários do painel B, ao passo que o painel C serve de base para o sorteio do sexto, nono e décimo segundo mês. A partir do décimo terceiro mês, repete-se esse esquema, sempre com a seleção de novas unidades domiciliares a cada novo mês. Dessa forma, podem-se aproveitar os custos despendidos no processo de arrolamento dos setores censitários e ao mesmo tempo obter amostras mensais independentes (Figura 1).

### Divulgação das Informações

Embora o levantamento seja realizado todos os meses, para garantir a precisão desejada, o cálculo dos indicadores

**FIGURA 1**  
Esquema de Coleta de Dados da PED



Fonte: Fundação Seade. PED.

é efetuado com a amostra acumulada do trimestre. Como será apresentado a seguir, a independência das amostras mensais permite esse acúmulo, bem como a divulgação, todos os meses, de novos resultados obtidos a partir de trimestres móveis, uma vez que incluem sempre os três painéis definidos. Dessa forma, é possível acompanhar mensalmente a tendência dos principais indicadores divulgados, entre os quais se destacam as estimativas dos contingentes de ocupados, desempregados e inativos, as taxas de desemprego total, por tipo e por atributos pessoais, a distribuição dos ocupados por setor de atividade econômica e por posição na ocupação e os rendimentos médios da população ocupada e assalariada.

### Cálculo dos Indicadores

Conforme já mencionado, o plano amostral em dois estágios, utilizado com a seleção, no primeiro estágio, de conglomerados (setores censitários) e, no segundo, de domicílios dentro dos setores previamente selecionados, faz com que o tamanho da amostra a cada mês possa oscilar, dependendo do crescimento ou da diminuição do SC sorteado. Desse modo, a maioria dos indicadores produzidos tais como taxas, distribuições de frequências e médias, é calculada por meio de estimadores do tipo razão, ou seja, pelo quociente de duas variáveis aleatórias ou características. Por exemplo, a taxa de desemprego total divulgada corresponde ao estimador razão combinado no trimestre, expresso pelo quociente entre o número total de desempregados obtido em três meses de pesquisa e o número total de pessoas economicamente ativas obtido no mesmo período.

O cálculo das estimativas é obtido diretamente dos dados acumulados no banco de dados, ao passo que o dos erros amostrais, sendo bastante complexo, requer a utilização de *softwares* específicos, como PCCARP, SAS e Sudaan. Atualmente, a Fundação Seade vem utilizando o *software* SAS para o cálculo dos erros.

### Controle de Qualidade dos Resultados

O tamanho da amostra foi calculado a fim de garantir a precisão desejada para alguns indicadores, considerando-se principalmente a taxa de desemprego total. Para garantir a confiabilidade de todos os resultados divulgados, quer em estudos de caráter conjuntural, quer de caráter estrutural, determinou-se que sejam disponibilizados de forma rotineira apenas os indicadores cujo coeficiente de variação seja no máximo 7,5%. Estudos com erros superiores a esse limite só são divulgados em casos muito especiais.

### Erros Não Amostrais

Além do esforço para controlar os erros amostrais, através da determinação do tamanho de amostra adequado para a precisão desejada e da divulgação de indicadores com coeficiente de variação de no máximo 7,5%, a PED também realiza diversos procedimentos a fim de controlar os erros de origem não amostral, entre eles os relacionados a cobertura, coleta e processamento de informações.

Uma equipe de checagem investiga aproximadamente 30% do material coletado pela PED, verificando tanto aspectos de cobertura, ou seja, da realização da entrevista

ta no domicílio preestabelecido, como de qualidade das principais informações levantadas. Uma equipe interna de crítica analisa todos os questionários preenchidos, dando atenção especial à coerência das informações e à existência de possíveis vieses devidos à atuação dos entrevistadores. Após a entrada eletrônica dos dados, efetuada com programa específico para minimizar erros, realiza-se o controle final pela execução de um programa de consistência eletrônica de dados, que busca eliminar possíveis erros ocorridos durante a entrada dos mesmos e incoerências que tenham passado despercebidas no processo de crítica.

### **Tipo de Amostra**

Levantamentos amostrais periódicos envolvem decisões metodológicas difíceis e controversas, principalmente aquelas relativas ao uso de painéis fixos, amostras independentes, rotacionadas ou de outras combinações. Maiores informações podem ser encontradas em Duncan e Kalton (1987).

Amostras independentes são aquelas em que, a cada novo momento ou tomada do levantamento, sorteiam-se unidades domiciliares que nunca haviam sido selecionadas. Por outro lado, o uso de painéis fixos implica a utilização das mesmas unidades amostrais em todas as tomadas do levantamento, ao passo que as amostras rotacionadas combinam parcela de unidades amostrais que permanecem na amostra por um número predeterminado de tomadas (sendo substituídas findo esse prazo) e parcela de novas unidades, daí a referência à rotação em seu nome.

Um dos principais guias para orientar a opção por um ou outro plano amostral deve ser o objetivo primário para o levantamento das informações. Frequentemente, as pesquisas contínuas são utilizadas para estimar parâmetros pontuais e ao mesmo tempo avaliar as mudanças ocorridas entre um instante e outro, como nas pesquisas sobre desemprego, por exemplo. Isso poderia sugerir o uso de painéis, rotacionados ou não, devido à menor variabilidade das estimativas entre uma tomada e outra, o que garante a confiabilidade da informação. A variabilidade de um momento a outro tende a ser menor, já que toda a amostra, no caso de painéis fixos, ou parcela dela, no caso de amostras rotacionadas, é composta pelas mesmas unidades domiciliares, onde pode-se esperar que, para intervalos relativamente curtos entre uma tomada e outra, como no caso de levantamentos mensais, os entrevistados não tenham sofrido alterações substanciais em sua condição

de atividade. Por exemplo, é mais provável que um entrevistado que já se encontrava ocupado no mês anterior de pesquisa continue nessa situação no mês atual do que um desempregado se torne ocupado, no mesmo intervalo de tempo. Ou seja, existe correlação, quase sempre positiva, nas partes fixas dos painéis. A utilização desse recurso também pode contribuir para reduzir custos, se houver a possibilidade, para a parte fixa da amostra, de simplificar a coleta das informações, como por exemplo, através da realização da entrevista inicial, pessoalmente, e das subsequentes, por telefone.

As vantagens potenciais desse esquema de coleta de dados, entretanto, são muitas vezes suplantadas pelas sérias dificuldades em sua aplicação, que não devem ser desconsideradas no momento do planejamento. Uma delas reside em se efetuarem análises estatísticas que tirem o justo proveito desses modelos mais complexos de amostragem. Outra, não menos importante, deve-se à já mencionada dinâmica das metrópoles e a conseqüente dificuldade de localização dos domicílios que compõem a parte fixa da amostra – é comum que, embora a mesma unidade domiciliar seja entrevistada em dois momentos diferentes, os moradores já não sejam mais os mesmos ou que simplesmente não seja possível localizar exatamente a mesma unidade domiciliar, em áreas de rápido crescimento e mudanças, como favelas, por exemplo.

Um dos problemas mais sérios desse método de coleta, todavia, encontra-se no aumento da taxa de recusa dos moradores em responder à pesquisa, que tende a ocorrer justamente quando se localiza a unidade domiciliar procurada e ela continua habitada pelos mesmos moradores. Isso se deve a um desgaste natural produzido pela realização de entrevistas consecutivas com os mesmos indivíduos. Outro efeito prejudicial comum é o condicionamento dos entrevistados, ou seja, suas respostas passam a ser influenciadas pelo conhecimento prévio do que lhes será perguntado. Finalmente, pode haver discrepâncias importantes e difíceis de serem contornadas entre as respostas oferecidas pelos mesmos entrevistados em tomadas diferentes.

As amostras independentes, por outro lado, permitem que se realizem estimativas pontuais e também que se avaliem as alterações ocorridas entre duas tomadas determinadas, sem as dificuldades e desvantagens das amostras de painéis, embora com uma variabilidade maior para as comparações entre tomadas distintas. No entanto, cabe ressaltar que a PED, ao divulgar seus indicadores baseados em trimestres móveis, controla a variabilidade das



comparações entre trimestres consecutivos graças à presença, nesses trimestres, de dois meses com as mesmas informações.

Além disso, as amostras independentes são particularmente interessantes quando se deseja acumular informações de várias tomadas. O acúmulo de dados coletados em momentos diversos constitui uma boa alternativa para o estudo de fenômenos raros pois, sem esse recurso, seria necessário o levantamento de uma única amostra suficientemente grande para tal, cujo custo poderia, muitas vezes, inviabilizar por completo o estudo desses fenômenos. Entretanto, ao utilizar o acúmulo de informações, é importante ter sempre em mente questões relacionadas à estabilidade das populações em estudo. Por exemplo, quando o fenômeno de interesse é bastante mutável com o passar do tempo, não é conveniente que se acumulem dados referentes a longos períodos de coleta, sob pena de se obterem resultados e conclusões distorcidos. Por outro lado, se os fenômenos ou características que se pretende estudar apresentam relativa estabilidade ao longo do tempo, os resultados dos dados acumulados são bastante confiáveis.

Outra importante utilização das amostras independentes encontra-se na construção de bancos de dados que possam servir como uma amostra mestra para o planejamento de amostras futuras, como será visto a seguir.

A PME (Pesquisa Mensal de Emprego) do IBGE e a LFS (*Labor Force Survey* – Pesquisa de Mão-de-Obra) do Canadá constituem exemplos de pesquisas sobre mercado de trabalho realizadas com painéis rotacionados. Além da PED, outro exemplo de pesquisa que se serve de amostras independentes é a *U.K. General Household Survey*, pesquisa geral domiciliar do Reino Unido.

### Utilização da Amostra da PED

Para o planejamento da PED, optou-se pela utilização de amostras independentes graças a duas grandes vantagens que ela apresenta: a possibilidade de acumular informações e a de utilizar a amostra da PED como uma amostra mestra para outras pesquisas.

**Acúmulo de informações** – A PED serve-se do acúmulo de informações oferecido pelas amostras independentes de três formas distintas. A primeira refere-se ao acúmulo dos próprios dados levantados pelo questionário-padrão da pesquisa, seja para análises conjunturais, com a divulgação mensal de indicadores produzidos a partir dos dados de trimestres móveis, como mencionado anteriormente,

seja para análises de caráter mais estrutural, por meio do acúmulo de informações ao longo de períodos mais extensos, em geral um ano ou biênio. Diversos estudos desse tipo foram e continuam sendo realizados periodicamente com os dados acumulados da PED. Destacam-se os perfis traçados de importantes grupos de mão-de-obra, entre eles os jovens, os idosos, os negros e as mulheres e de diversas categorias profissionais como metalúrgicos, trabalhadores nos serviços creditícios e financeiros, na construção civil e nos serviços domésticos. Estudam-se também os desempregados, sua estrutura familiar e seus meios de sobrevivência utilizados, os ocupados e sua inserção setorial, jornada de trabalho e rendimentos recebidos.

A segunda forma de aproveitar a possibilidade de acúmulo de dados oferecida pelas amostras independentes consiste em coletar, além dos dados normalmente levantados pela PED através de seu questionário-padrão, informações adicionais nos mesmos domicílios selecionados para entrevista, por questionários complementares desenhados com o objetivo de colher informações que ajudem a aprofundar as análises normalmente realizadas, possibilitando um melhor entendimento ou detalhamento de questões relacionadas à mão-de-obra da região. Como exemplos, podem ser citados os questionários referentes à mobilidade ocupacional da população em idade ativa e à formação profissional. Tal procedimento mostra-se bastante adequado, uma vez que certos aspectos relacionados ao mercado de trabalho urbano, embora relevantes, não justificam a coleta permanente de informações a seu respeito e sua conseqüente sobrecarga nos custos e prazos da pesquisa.

A terceira forma consiste na obtenção de dados referentes a novas dimensões normalmente não investigadas, por meio de questionários suplementares ao da PED. Constituem exemplos já realizados a pesquisa sobre hábitos de leitura e acesso à Internet e a pesquisa sobre renda e pobreza e acesso a programas sociais.

Esses questionários complementares ou suplementares são geralmente aplicados durante um período predeterminado, como um ano ou um semestre, podendo repetir-se o processo a grandes intervalos de tempo, a fim de se obterem parâmetros de comparação.

**A PED como amostra mestra** – Devido ao tamanho da amostra mensal da PED (aproximadamente 3 mil novos domicílios a cada mês) e também ao fato de vir sendo realizada ininterruptamente desde 1985, tem-se, hoje, um grande conjunto de domicílios já selecionados que pode,

por sua vez, servir como base para a seleção de novas amostras. Esse fim também pode ser alcançado aproveitando-se o arrolamento de domicílios dos 600 setores censitários que totalizam cerca de 300 mil domicílios. Ou seja, a PED pode ser utilizada como uma amostra mestra para outras pesquisas, entre as quais se destacam a de Condições de Vida – PCV na Região Metropolitana de São Paulo e a de Acidentes de Trabalho e Doenças Profissionais.

A extensão do banco de dados acumulado da PED também permite que ele seja considerado como universo para estudos de simulação (Figura 2).

### Disponibilização dos Microdados

Todas as informações coletadas pela PED desde seu início estão disponíveis aos usuários que tenham interesse em estudá-las mediante recortes diferentes daqueles normalmente divulgados. Os dados encontram-se em bancos estruturados anualmente e extensivamente documentados, com o objetivo de simplificar sua utilização.

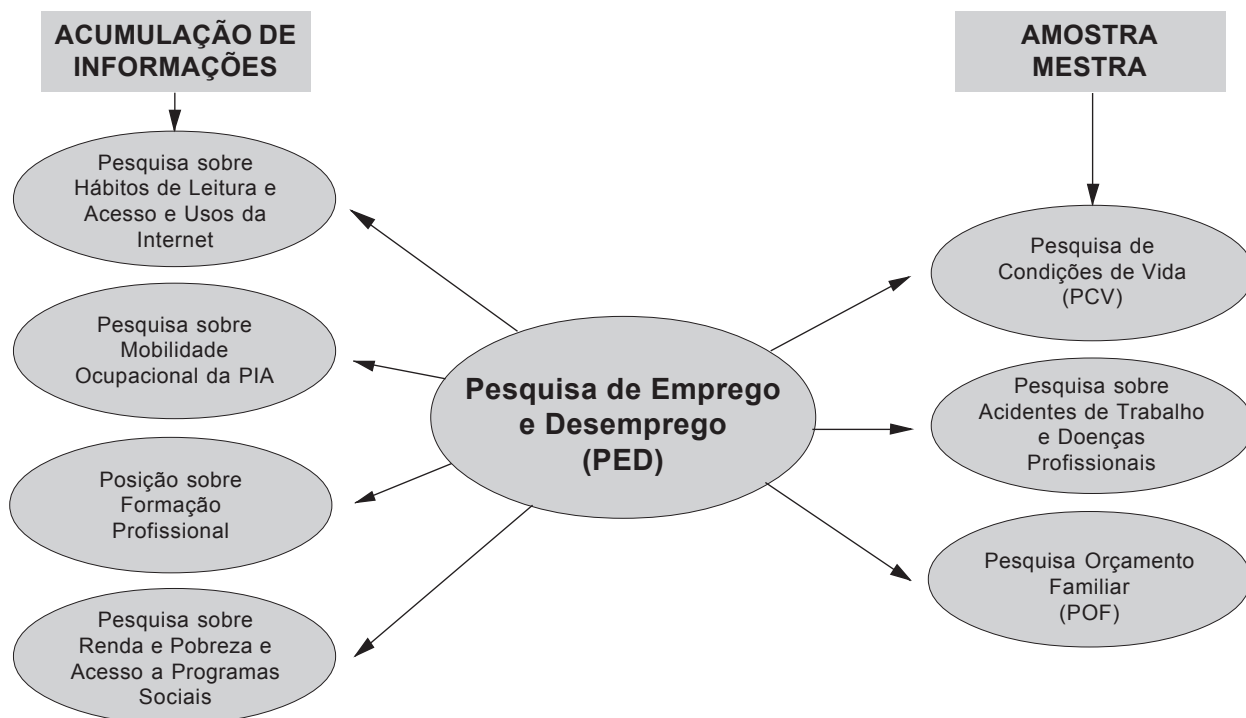
Embora os microdados estejam disponíveis, a privacidade dos entrevistados está completamente protegida graças a diversas características desses bancos de dados: não constam os nomes nem os endereços dos entrevistados e as informações referentes aos setores censitários encontram-se descaracterizadas, servindo apenas para o cálculo de erros amostrais.

### Resumo do Plano Amostral

Conforme o apresentado, podemos resumir que o plano amostral utilizado na PED é:

- uma amostra probabilística estratificada obtida em dois estágios dos moradores da RMSP localizados nos setores censitários urbanos do IBGE listados no último Censo disponível;
- o uso de dois critérios de estratificação: um explícito, resultante de agrupamentos homogêneos segundo as características do emprego, e outro implícito, obtido da ordenação geográfica e por renda dos SCs;

**FIGURA 2**  
Utilização da Amostra da PED



Fonte: Fundação Seade. PED.



- sorteio dos SCs com probabilidade proporcional ao número de domicílios residenciais ocupados na época do Censo, e este número é atualizado antes do sorteio dos domicílios;
- sorteio dos domicílios com igual probabilidade dentro de cada SC;
- sorteio mensal dos SCs, conjugado em três painéis rotativos.

### CONTRIBUIÇÃO AO DESENVOLVIMENTO DA FUNDAÇÃO SEADE

No momento da implantação da PED, a Fundação Seade não dispunha de um corpo técnico especializado na elaboração de planos amostrais de pesquisas domiciliares, o que a levou a buscar profissionais acadêmicos com esse perfil. Além da realização do plano amostral, a necessidade de acompanhamento permanente da pesquisa, devida ao seu caráter de levantamento contínuo, fez com que alguns desses profissionais passassem a integrar o corpo de funcionários da Fundação.

A incorporação desses técnicos e a experiência adquirida com a realização da PED possibilitaram à Fundação Seade o desenvolvimento e a realização de diversas outras pesquisas domiciliares e não-domiciliares por amostragem. Além disso, o desafio de obter controle de qualidade praticamente total nos resultados da pesquisa, que só poderia ser superado com o trabalho integrado de todas as equipes envolvidas em sua produção, levou a uma nova consciência e a um novo padrão de trabalho, por sua vez também estendido a futuras pesquisas e estudos realizados.

Um dos aspectos de controle de qualidade das informações que merece destaque é o da opção por divulgar apenas indicadores com confiabilidade fixada (coeficiente de variação de no máximo 7,5%), o que garante a precisão de todos os dados que vêm a público, e permite que especialistas ou leigos utilizem os resultados da pesquisa divulgados sem a necessidade de cuidados adicionais.

Outra contribuição importante oferecida pela PED à instituição foi a introdução e incorporação de técnicas estatísticas até então pouco exploradas, por exemplo, os métodos multivariados, como análise de agrupamentos (*clusters*), discriminante e outras. A familiarização dos analistas com tais técnicas permitiu a geração de indicadores capazes de refletir diversas dimensões de uma realidade. A análise de agrupamentos foi utilizada com sucesso na Pesquisa de Condições de Vida, na qual se procura classificar as famílias com

base em sua condição de vida, avaliada em dimensões distintas, como características da moradia e educação, rendimento, inserção no mercado de trabalho e saúde dos membros da família. Outro exemplo de seu uso encontra-se na geração do IPRS – Índice Paulista de Responsabilidade Social, que tem o objetivo de analisar os municípios do Estado de São Paulo por meio de indicadores sintéticos que abranquem diversas dimensões de riqueza dos municípios e de educação e saúde de seus habitantes.

### CONSIDERAÇÕES FINAIS

Assim como em sua parte conceitual, a PED busca atender a questões específicas da realidade brasileira, como a grande heterogeneidade dos mercados de trabalho urbanos e a dificuldade de se estabelecerem limites específicos entre as possíveis condições de atividade da população em idade ativa, no que se refere ao levantamento da amostra também foi considerada a especificidade regional, bem como a experiência internacional, a fim de se obter um planejamento eficiente e condizente com os recursos disponíveis.

A limitação de recursos financeiros e a necessidade de um tamanho mínimo de amostra para garantir a confiabilidade dos resultados levaram à solução de se realizarem levantamentos trimestrais com divulgação mensal de dados, obtendo-se, assim, a possibilidade de acompanhamento conjuntural do mercado de trabalho com uma amostra mensal relativamente pequena e custos menores.

A adoção de amostras independentes, além de simplificar o procedimento de coleta dos dados, permitiu que se acumulassem informações por períodos mais extensos e, dessa forma, se realizassem análises estruturais ou estudos de fenômenos raros. Tal possibilidade de acúmulo de dados, combinada à longa duração da pesquisa, realizada ininterruptamente desde sua origem em 1985, faz com que a amostra da PED possa, por sua vez, ser considerada uma amostra mestra para futuros estudos.

A organização de todos os dados até hoje coletados em bancos estruturados e bem documentados garante a facilidade de acesso da sociedade às informações que buscam retratar alguns de seus aspectos importantes.

Finalmente, a realização da PED exigiu a formação de uma equipe interna com um novo perfil profissional que pôde servir-se da experiência adquirida nessa pesquisa para o desenvolvimento e a realização de outras investigações, domiciliares ou não, e para a geração de indicadores sociais inovadores.

## REFERÊNCIAS BIBLIOGRÁFICAS

BOLFARINE, H.; BUSSAB, W.O. *Elementos de amostragem*. Versão preliminar. São Paulo: Instituto de Matemática e Estatística da Universidade de São Paulo, 2000.

BUSSAB, W.O.; DINI, N.P. Pesquisa de Emprego e Desemprego Seade/Dieese: regiões homogêneas da Grande São Paulo. *São Paulo em Perspectiva*, São Paulo, Fundação Seade, v.1, n.3, p.5-11, set./dez. 1985.

COCHRAN, W.G. *Sampling techniques*. 3. ed. New York: John Wiley & Sons, Inc., 1977.

DUNCAN, G.J.; KALTON, G. Issues of design and analysis of surveys across time. *International Statistical Review*, Great Britain, v.55, n.1, p.99-117, 1987.

KALTON, G. *Introduction to survey sampling*. 4. ed. Beverly Hills: Sage Publications, Inc., 1983.

KISH, L. *Survey sampling*. 3. ed. New York: John Wiley & Sons, Inc., 1965.

TROYANO, A.A.A. Trajetória de uma pesquisa: avanços e obstáculos. *São Paulo em Perspectiva*, São Paulo, Fundação Seade, v.4, n.3/4, jul./dez. 1990.

---

**WILTON DE OLIVEIRA BUSSAB:** *Professor de Métodos Quantitativos da Escola de Administração de Empresas de São Paulo – FGV.*

**NÁDIA PINHEIRO DINI:** *Estatística, Gerente de Métodos Quantitativos da Fundação Seade.*

**SILVIA REGINA MANCINI:** *Estatística, Analista da Fundação Seade.*