

## **Tipos de Variáveis**

As variáveis coletadas podem ser classificadas em quantitativas ou qualitativas:

1) Dados quantitativos são números que representam contagens ou medidas (renda e anos de escolaridade, por exemplo). Esses dados se dividem em discretos e contínuos:

1.1) Os dados discretos são aqueles em que o número de valores possíveis são finitos ou “enumeráveis” (tal como o número de cômodos em um domicílio).

1.2) Os dados contínuos resultam de infinitos valores possíveis em uma escala contínua (renda per capita).

2) Os dados qualitativos (ou categóricos ou de atributos) podem ser separados em diferentes categorias que se distinguem por alguma característica não-numérica (sexo e raça, por exemplo).

Os dados podem ainda ser classificados quanto ao seu nível de mensuração (nominal, ordinal, intervalar ou razão):

1) O nível nominal de mensuração utiliza dados que informam nomes, rótulos ou categorias. Os dados não são ordenados e não devem ser usados para cálculos de médias, como é o caso de raça e código do município, por exemplo.

2) O nível ordinal de mensuração engloba dados que podem ser organizados em alguma ordem. Sabemos que há diferenças relativas entre os valores dos dados, mas não sabemos as magnitudes das diferenças. Por exemplo, na escala de frequência “pouco/médio/muito”, é possível ordenar os dados, mas não sabemos se a diferença entre “pouco” e “médio” é a mesma que aquela existente entre “médio” e “muito”.

3) O nível intervalar de mensuração é similar ao ordinal, mas sabemos as magnitudes das diferenças entre dois valores. Os dados não possuem um ponto inicial zero natural. Este é o caso dos anos censitários (1970, 1980, 1991 e 2000), em que sabemos as magnitudes das diferenças entre os anos, mas o tempo não começou em zero.

4) Por fim, o nível de mensuração de razão é similar ao intervalar, mas há um ponto inicial zero natural. Como há um zero que indica nenhuma quantidade, é possível dizer que uma quantidade é maior que outra em X vezes (razões significativas). Por exemplo, 30 anos de idade é 6 vezes maior do que 5 anos de idade.

## **Modelos de Regressão**

Os modelos de regressão estatística são escolhidos dependendo do tipo da variável a ser analisada. Estes modelos procuram verificar a existência e magnitude de relações de causalidade entre diferentes variáveis. Em outras palavras, a regressão estima o impacto que mudanças em uma série de variáveis explicativas (independentes) possuem sobre uma variável dependente de interesse (tal como escolaridade e renda). De uma forma geral, a própria metodologia de avaliação pressupõe que determinadas variáveis oscilam de acordo com mudanças em outras variáveis. Por exemplo, espera-se que com a implementação de políticas de saneamento básico (variável independente), a incidência de doenças infecto-parasitárias (variável dependente) diminua em uma determinada localidade. Ou seja, tais relações implicam que uma mudança de valor de uma determinada variável dependente está associada com uma mudança de valor de uma variável sócio-econômica ou com a implementação de uma política pública. Esse tipo de análise possibilita mensurar de que modo as diferentes políticas públicas implementadas em uma comunidade podem ter impactos em uma série de variáveis demográficas, sócio-econômicas, de saúde e de satisfação, por exemplo. Abaixo são descritos modelos de mínimos quadrados ordinários, modelos logísticos e modelos multinomiais.

Os modelos de mínimos quadrados ordinários são utilizados para analisar tendências de variáveis dependentes quantitativas com nível de mensuração intervalar ou de razão. Em modelos de regressão estatística, é calculado o tamanho do impacto das variáveis independentes sobre a variável dependente. Também é calculado se a relação entre estas variáveis é estatisticamente significativa ao nível de 95%. Para o caso dos modelos de mínimos quadrados ordinários, estas estimativas são realizadas de forma a minimizar o erro amostral na mensuração de efeitos sobre variáveis dependentes quantitativas intervalares e de razão. Um exemplo de variável de razão que pode ser utilizada nestes modelos são as informações de renda ou escolaridade dos indivíduos.

Os modelos logísticos são utilizados para analisar dados qualitativos com nível de mensuração nominal. As regressões logísticas têm a peculiaridade de estimar efeitos em variáveis binárias (dicotômicas), em que a variável dependente assume o valor de zero ou um, com base em oscilações em uma série de variáveis independentes. Assim como o modelo anterior, a intensidade e significância do impacto de variáveis explicativas sobre a variável dependente é estimada pelo programa estatístico. A informação se a pessoa realizou consulta médica (resposta sim ou não), em um determinado intervalo de tempo, é um exemplo de variável binária que pode ser utilizada neste método de estimação. Um conjunto de informações pode ser utilizado para explicar oscilações nesta variável dependente, tais como falta de infra-estrutura na comunidade que influenciariam incidência de doenças, configuração do domicílio do residente e renda domiciliar.

Por fim, os modelos multinomiais são utilizados para avaliar variáveis dependentes qualitativas com nível de mensuração nominal ou ordinal. Mais uma vez, a intenção é de captar a intensidade e significância de impacto de uma série de variáveis explicativas (independentes) em uma variável dependente. No caso deste tipo de modelo, a variável dependente utilizada pode ser a resposta do morador relativa à escala de satisfação

quanto a uma política pública realizada em sua comunidade. Ou seja, a escala de satisfação seria uma variável qualitativa ordinal.

Na estimação desses modelos estatísticos, a validade e confiabilidade das análises são requisitos metodológicos para que uma avaliação tenha legitimidade. A validade pode ser definida como o grau em que as diferenças obtidas no objeto de estudo refletem verdadeiras diferenças no atributo medido. A validade do instrumento de medição se determina pelo grau de congruência entre os resultados de sua aplicação e evidências significativas disponíveis. A validade interna refere-se às mudanças geradas por tratamentos experimentais em um plano determinado. Já a validade externa diz respeito às possíveis generalizações do resultado obtido. A confiabilidade está relacionada com a estabilidade e exatidão da medição. Uma medição é estável quando um instrumento aplicado repetidamente sobre o mesmo objeto de análise apresenta resultados no mínimo semelhantes. Enquanto isso, a exatidão implica em uma relativa ausência de erros em um determinado instrumento. Para existir validade é necessário que exista também confiabilidade e vice-versa.

### **Referências**

- Jannuzzi, Paulo de Martino. 2006. *Indicadores sociais no Brasil: conceitos, fontes de dados e aplicações*. 3ª ed. Campinas: Editora Alínea.
- Triola, Mario F. 2008. *Introdução à Estatística*. 10ª ed., Rio de Janeiro: LTC.