

Introdução à Análise de Dados I

Lupércio F. Bessegato e Marcel T. Vieira - Jan/2013

Estudo de Caso - Análise de Dados do SAEB

Descrição:

A base de dados utilizada nesta atividade é a do Sistema Nacional de Avaliação da Educação Básica (SAEB), coletada pelo INEP no ano de 1999, compreendendo um exame de conhecimentos, a partir dos quais foram calculados escores de proficiência dos alunos, tendo como base a Teoria da Resposta ao Item (TRI). Acrescentou-se a estes exames um levantamento sobre condições sócio-econômico-demográficas de mais de 200.000 alunos, suas escolas, professores e diretores. Estes dados, quando analisados de forma adequada, podem dar respostas a importantes questões educacionais. Por exemplo, uma vez controlando-se pelos determinantes socioeconômicos do aluno, podem ser estudados os fatores ligados à turma e/ou escola, que mais contribuem para um melhor desempenho escolar do aluno.

Consideraremos os dados que estão armazenados no arquivo *Saeb99_Mestrado.xls*, que incluem uma sub-amostra de 722 alunos do 3º Ano do Ensino Médio, avaliados em Física e matriculados em escolas localizadas nos estados de Minas Gerais e São Paulo em 1999.

Base de dados:

Informações sobre as variáveis incluídas na base de dados são apresentadas abaixo:

ufesc_c	Estado onde se localiza a Escola	31 - Minas Gerais 35 - São Paulo
cadalu_c	Código do aluno	
q12_3	Que idade você tinha quando entrou na 1ª. série?	1 - Menos de 5 anos 2 - 5 anos 3 - 6 anos 4 - 7 anos 5 - 8 anos 6 - 9 anos ou mais
q13_3	Qual seu sexo?	1 - Masculino 2 - Feminino
q14_3	Como você se considera?	1 - branco(a) 2 - pardo(a)/mulato(a) 3 - negro(a) 4 - amarelo(a) 5 - indígena
q17_3	Qual o nível de instrução de seu pai?	1 - Nunca estudou 2 - Ensino Fundamental-1ª. a 4ª. série 3 - Ensino Fundamental- 5ª. a 8ª. série 4 - Ensino Médio 5 - Ensino Superior 6 - Pós-Graduação 7 - Não sei
q18_3	Qual o nível de instrução de sua	1 - Nunca estudou

Introdução à Análise de Dados I

Lupércio F. Bessegato e Marcel T. Vieira - Jan/2013

	mãe?	2 - Ensino Fundamental-1ª. a 4ª. série 3 - Ensino Fundamental- 5ª. a 8ª. série 4 - Ensino Médio 5 - Ensino Superior 6 - Pós-Graduação 7 - Não sei
q111_3	Como é sua situação econômica?	1 - Não trabalho e dependo financeiramente da minha família 2 - Trabalho e dependo financeiramente da minha família 3 - Trabalho e não dependo financeiramente da minha família 4 - Trabalho e outras pessoas dependem de meu trabalho para viver
q114_3	Você já deixou de frequentar a escola por algum período?	1 - Não 2 - Sim, por 1 ano 3 - Sim, por 2 anos 4 - Sim, por 3 anos 5 - Sim, por 4 anos 6 - Sim, por mais de 4anos
q116_3	Você já repetiu de ano?	1 - Não 2 - Sim, 1 vez 3 - Sim, 2 vezes 4 - Sim, 3 vezes 5 - Sim, 4 vezes 6 - Sim, mais de 4 vezes
q121_3	Quanto tempo você leva para chegar de sua casa até a escola?	1 - Menos de 15 minutos 2 - De 15 a 30 minutos 3 - De 31 minutos a 1 hora 4 - Mais de 1 hora
q26_3	Você gosta de Física	1 - Não gosto 2 - Gosto mais ou menos 3 - Gosto muito
profic99	Proficiência em Física	

Objetivo:

O objetivo da atividade é a manipulação, análise e interpretação de dados de base de dados do Saeb.

Análise:

Deseja-se efetuar uma análise exploratória dos dados, salientando-se os seguintes pontos:

- Análise e comparação das variáveis relevantes da base de dados, resumindo-as adequadamente por meio de gráficos, tabelas e medida-resumo;

Introdução à Análise de Dados I

Lupércio F. Bessegato e Marcel T. Vieira - Jan/2013

- Identificação de estratificações importantes ao estudo de variáveis relevantes do banco de dados;
- Identificação de padrões de variáveis relevantes do banco de dados;
- Estabelecimento de hipóteses para estudo posterior mais detalhado.

Roteiro:

1. Com relação à proficiência em Física, pede-se:
 - a. Avalie sua distribuição, no que diz respeito à tendência central, dispersão assimetria;
 - b. Considerando os mesmos aspectos considerados no item “a”, compare os alunos dos estados de Minas Gerais e São Paulo. Em qual dos dois estados o desempenho em física é melhor? Onde a homogeneidade entre os alunos é maior?
2. Podemos identificar evidências de diferenças de gênero e de raça no desempenho em Física? Utilize ferramentas gráficas como o boxplot ou histograma.
3. O que parece influenciar mais no desempenho dos alunos em física - o grau de instrução dos pais ou das mães? Considere medidas de tendência central.
4. Alunos que trabalham parece terem um desempenho melhor ou pior do que aqueles que não trabalham?
5. Parece haver alguma relação entre repetência no passado e desempenho atual dos alunos em física?
6. A literatura da educação sugere que alunos que moram mais próximos da escola tendem a ter um desempenho escolar melhor. A base de dados considerada oferece evidências favoráveis a esta hipótese?
7. Alunos que gostam de Física parecem ter um desempenho melhor do que aqueles que não gostam?
8. Considerada a análise exploratória efetuada, pede-se:
 - a. Que hipóteses levantadas por você necessitam de uma verificação formal, por meio de técnicas de inferência estatística?
 - b. Que variáveis você sugere sejam acrescentadas a presente base de dados de maneira a aprimorar suas conclusões (e decisões)?