

Conjunto de dados: dados_escolas_catolicas

Dados utilizados no artigo:

Wooldridge Source: Altonji, J.G., T.E. Elder, and C.R. Taber (2005), "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling," Journal of Human Resources 40, 791-821. Professor Elder kindly provided a subset of the data, with some variables stripped away for confidentiality reasons. Data loads lazily.

Variáveis:

id: identificador da pessoa

read12: escore padronizado de leitura (avaliação)

math12: escore padronizado de matemática (avaliação)

female: =1 se sexo feminino, 0 se sexo masculino

race: = 1 se "White", 2 se "asian", 3 se "hispan", 4 se "black"

motheduc: anos de estudo da mãe

fatheduc: anos de estudo do pai

lfaminc: logaritmo da renda familiar

hsgrad: = 1 se conclui o ensino médio em 1994, 0 se conclui em outro ano

cathhs: =1 se frequentou o ensino médio em escola católica, 0 se caso contrário

parcath: =1 se um dos pais se declara católico, 0 se nenhum dos pais se declara católico.

Observe que com exceção das variáveis escolaridade do pai, escolaridade da mãe, desempenho em linguagem, desempenho em matemática e logaritmo da renda, as demais variáveis, embora codificadas com números, são todas qualitativas.

Utilizando esse conjunto de dados, façam:

- a) Definam o tipo de cada uma das variáveis

- b) Façam uma análise descritiva de cada uma das variáveis: medidas descritivas e gráficos. Comente os aspectos principais de cada uma das variáveis.
- c) Crie uma variável chamada desempenho geral calculando a média do desempenho em linguagem com o desempenho em matemática e faça a análise descritiva dessa variável.
- d) Utilizando as medidas e gráficos adequados investigue a associação entre 3 pares de variáveis. (Use as medidas, tabelas e ou gráficos adequados para cada caso) sendo que:
Caso 1: desempenho em linguagem e desempenho em matemática.
Caso 2: etnia e parcat
Caso 3: desempenho em matemática e female (gênero)
Comente os resultados.
- e) Investigue a associação entre desempenho geral e logaritmo da renda. Repita essa análise separadamente para cada sexo e responda se há indicativos de que a relação entre desempenho geral e renda depende da variável gênero. Comente os resultados.

O trabalho será feito em grupo de 5 ou 4 alunos. Foram disponibilizados 11 conjuntos de dados, cada um deles com 2000 observações. Foram disponibilizados 13 conjuntos de dados, 1 para cada grupo.