

ACH3657

# Métodos Quantitativos para Avaliação de Políticas Públicas

Aula 6

Regressão linear múltipla

Alexandre Ribeiro Leichsenring

[alexandre.leichsenring@usp.br](mailto:alexandre.leichsenring@usp.br)



# Organização

## Exemplo - Determinantes da nota média em curso superior

As variáveis do arquivo `gpa1.RData` incluem a nota média em um curso superior (**colGPA**), a nota média do ensino médio (**hsGPA**) e a nota do teste de avaliação de conhecimentos para ingresso em curso superior (**ACT**), para uma amostra de 141 estudantes de uma grande universidade dos Estados Unidos; tanto **colGPA** como **hsGPA** estão baseados em uma escala de quatro pontos.

- 1 Ajuste no SPSS um modelo de regressão linear para **colGPA** em função de **hsGPA** e **ACT**.
- 2 Interprete a equação estimada. O sinal dos coeficientes está de acordo com o que você esperava? Por quê?
- 3 Se escolhermos dois estudantes, A e B, e esses estudantes tiverem a mesma nota **ACT**, mas **hsGPA** do estudante A é um ponto maior que a **hsGPA** do estudante B, quanto maior, em média, deverá ser a nota **colGPA** do aluno A?
- 4 De acordo com o modelo ajustado, qual das duas variáveis influencia mais a nota média num curso superior (**colGPA**)?
- 5 Ajuste um modelo de regressão linear simples para **colGPA** em função apenas de **hsGPA**. Compare com o modelo de regressão com duas variáveis explicativas.

## Exercício

- 1 Um problema de interesse das autoridades da saúde (e outras) é determinar os efeitos que fumar durante a gravidez exerce sobre a saúde do recém-nascido. Uma medida da saúde do recém-nascido é o peso de nascimento; um peso de nascimento muito baixo pode atribuir à criança o risco de contrair várias doenças. Como outros fatores que afetam o peso de nascimento, além de fumar cigarros, estão provavelmente correlacionados com o fumo, devemos levar em consideração tais fatores. Por exemplo, uma renda maior geralmente permite acesso a pré-natais melhores, bem como a uma nutrição melhor da mulher. Uma equação que reconhece isso é:

$$pesonas = \beta_0 + \beta_1cigs + \beta_2rendfam + u$$

- (i) Qual é o sinal mais provável de  $\beta_2$ ? Discorra.
- (ii) Você acha que *cigs* e *rendfam* devem ser correlacionados? Explique por que a correlação pode ser positiva ou negativa.
- (iii) Agora, estime a equação com e sem *rendfam*, usando os dados em [bwght.RData](#). Relate os resultados na forma de uma equação, incluindo o tamanho da amostra e o *R*-quadrado. Discuta seus resultados, dando ênfase ao fato de acrescentar *rendfam* mudar ou não substancialmente o efeito estimado de *cigs* sobre *pesonas*.