

FORMAÇÃO CIENTISTA DE DADOS

REGRESSÃO LINEAR MÚLTIPLA



SIMPLES E MÚLTIPLA

- **SIMPLES**

- **UMA VARIÁVEL EXPLANATÓRIA PARA PREVER UMA VARIÁVEL DEPENDENTE**
- **$Y \sim X$**

- **MÚLTIPLA**

- **DUAS OU MAIS VARIÁVEIS EXPLANATÓRIAS PARA PREVER UMA VARIÁVEL DEPENDENTE**
- **$Y \sim X_1 + X_2 + X_N$**

ANALISAR CADA X COM Y

- **ANALISAR CADA VARIÁVEL INDEPENDENTE COM Y INDIVIDUALMENTE**
- **GERAR GRÁFICOS DE DISPERSÃO INDIVIDUAIS**
- **BUSCAR REDUNDÂNCIAS (MESMOS EFEITOS DE X SOBRE Y): EXPLICAÇÃO POSTERIOR**

COEFICIENTE DE DETERMINAÇÃO (R^2)

- LEMBRANDO QUE R^2 É O PERCENTUAL DE VARIAÇÃO DA VARIÁVEL DE RESPOSTA QUE É EXPLICADA PELO MODELO
- QUANDO SE COLOCAM MAIS VARIÁVEIS NO MODELO, A TENDÊNCIA É QUE R^2 AUMENTE, MESMO QUE A ADIÇÃO DA VARIÁVEL NÃO AUMENTE A PRECISÃO DO MODELO
- PARA ISSO, UTILIZA-SE R^2 AJUSTADO, QUE AJUSTA A VARIAÇÃO DO MODELO DE ACORDO COM O NÚMERO DE VARIÁVEIS INDEPENDENTES QUE É INCLUÍDA NO MODELO
- R^2 AJUSTADO VAI SER SEMPRE MENOR QUE R^2

COLINEARIDADE E PARCIMÔNIA

- **COLINEARIDADE: DUAS VARIÁVEIS INDEPENDENTES QUE SÃO CORRELACIONADAS**
- **INCLUIR VARIÁVEIS INDEPENDENTES COLINEARES PODE PREJUDICAR O MODELO, CRIANDO PREVISÕES NÃO CONFIÁVEIS**
- **PARCIMÔNIA: NÃO COLOCAR VARIÁVEIS QUE NÃO MELHOREM O MODELO EM NADA: CRIAR MODELOS PARCIMONIOSOS**

REQUISITOS BÁSICOS

- 1. LINEARIDADE ENTRE A VARIÁVEL DEPENDENTE E AS VARIÁVEIS INDEPENDENTES**
- 2. QUE AS VARIÁVEIS SEJAM NORMALMENTE DISTRIBUÍDAS**
- 3. POUCA OU NENHUMA COLINEARIDADE**

RESIDUAIS

- **PRÓXIMOS A DISTRIBUIÇÃO NORMAL**
- **VARIÂNCIA CONSTANTE EM RELAÇÃO A LINHA DE MELHOR AJUSTE**
- **INDEPENDENTES (SEM PADRÃO)**