

Aula 14 - Qualidade do Ajuste

Fabricio Murai

Anúncios importantes

- Prova 2
 - Data: 06/05 (sábado)
 - Horário: 08:00 - 10:00
 - Local:
 - Turma TZ: Sala 1020
- Monitoria
 - rafaelbrunosiqueira at gmail dot com
 - Sala: 1015
 - Sexta, 11:10 às 12:50
- Quizz de questões teóricas sobre interpolação polinomial

Aula passada

- Ajuste de curvas (curve fitting):
quero aproximar uma função a partir de m pontos
(pode ser um polinômio de grau $n \leq m-1$ ou outra função qualquer)
 - Relações determinísticas e semideterminísticas
 - Diagramas de dispersão
 - Regressão linear
 - Simples: $y_i = b_0 + b_1 x_i$
 - Múltipla: x_i é um vetor; $y_i = b_0 + b_1 x_{i1} + \dots + b_n x_{in}$
- Dúvidas sobre interpolação polinomial

Relações determinísticas vs. semideterminísticas

- Relação entre duas variáveis
 - X: **preditor**, explicativa ou independente
 - Y: **resposta** ou dependente
- Determinística: relação exata
 - Celsius \Leftrightarrow Fahrenheit
 - Raio \Leftrightarrow Circunferência

Relações determinísticas vs. semideterminísticas

- Semideterminística ou estatística: relação não é perfeita

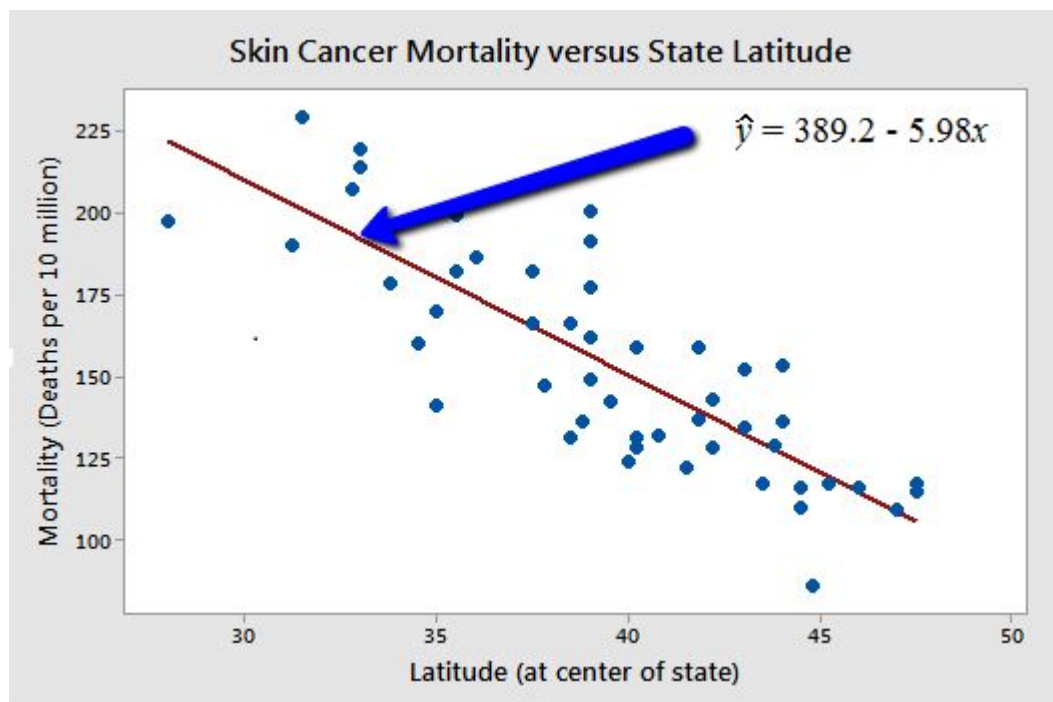


Diagrama de dispersão:

- Existe uma tendência
- Mas existe dispersão

Fonte:

<https://onlinecourses.science.psu.edu/stat501/node/251>

- Exemplos:

- Peso e altura
- Velocidade de condução e consumo de gasolina

Aula de hoje

- Exemplo de regressão simples
 - Interpretação dos coeficientes
- Qualidade do ajuste
 - Coeficiente de Determinação
 - Variância residual
- Transformações não-lineares
- Uso da interpolação vs. uso da regressão

Exemplo de regressão simples

- Formulário:

$$\begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \dots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}x_{i1} & \sum x_{i2}x_{i1} & \dots & \sum x_{ip}x_{i1} \\ \sum x_{i2} & \sum x_{i1}x_{i2} & \sum x_{i2}x_{i2} & \dots & \sum x_{ip}x_{i2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{ip} & \sum x_{i1}x_{ip} & \sum x_{i2}x_{ip} & \dots & \sum x_{ip}x_{ip} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \\ \vdots \\ \sum x_{ip}y_i \end{bmatrix}$$

- Dados

altura	63	64	66	69	69	71	71	72	73	75
peso	127	121	142	157	162	156	159	165	181	208

Interpretação dos coeficientes

- ○ que significa b_0 ?
 - Peso de indivíduo que tem altura zero?
Não. Não pode ser usado para extrapolação.
- ○ que significa b_1 ?
 - Aumento médio no peso por polegada, dentro do intervalo considerado

Qualidade do ajuste

- Vimos desvio como medida de qualidade

$$D(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - u_i)^2$$

- Problemas com essa medida?
 - Medida não é normalizada
 - No caso de regressão múltipla, quanto maior o número de parâmetros, menor o desvio
- Medidas de qualidade do ajuste usadas na prática
 - Coeficiente de determinação
 - Variância residual