Aula inaugural do curso
Análise de Regressão

Profª Silvia Nagib Elian
Sala 215 - Bloco A

Instituto de Matemática e Estatística
Universidade de São Paulo

Agenda

Exemplo

Um psicólogo está investigando a relação entre o tempo que o indivíduo leva para reagir a um certo estímulo e algumas de suas características tais como: sexo, idade e acuidade visual (medida em %).

Dados - Bussab, W.O. Análise de Variância e de Regressão
São Paulo; Atual, 1986

	Y	W	X	Z
Individuo	Tempo de reação	Sexo	Idade	Acuidade visual
1	96	H	20	90
2	92	M	20	100
3	106	H	20	80
4	100	M	20	90
5	98	M	25	100
6	104	H	25	90
7	110	M	25	80
8	101	M	25	90
9	116	H	30	70
10	106	H	30	90
11	109	H	30	90
12	100	M	30	80
13	112	M	35	90
14	105	M	35	80
15	118	H	35	70
16	108	H	35	90
17	113	M	40	90
18	112	M	40	90
19	127	H	40	60
20	117	H	40	80

Objetivo: estudar o tempo de reação.

Possíveis modelos

$$1) \qquad y_i = \mu + \epsilon_i$$

onde

- μ - efeito fixo, comum a todos indivíduos.
- ϵ_i - variável aleatória correspondente ao i-ésimo individuo.
- $\epsilon_i = f(w, x, z, \ldots)$.

$$\hat{\mu} = \bar{y}$$

Possíveis modelos

$$2) \qquad y_{ij} = \mu_i + \epsilon_{ij}$$

onde

- μ_i - efeito comum a todos os elementos do grupo i
  - $i = \begin{cases} 1, & \text{homem} \\ 2, & \text{mulher} \end{cases}$
- ϵ_{ij} - efeito aleatório do j-ésimo individuo do i-ésimo grupo.
- $\epsilon_i = f(x, z, \ldots)$.

$$\hat{\mu}_1 = \bar{y}_1 \qquad \hat{\mu}_2 = \bar{y}_2$$

Se $H_0 : \mu_1 = \mu_2$ é aceita $\rightarrow$ Modelo 1).

Possíveis modelos

$$3) \qquad y_{ij} = \mu_i + \epsilon_{ij}$$

onde

- $i = 1, 2, 3, 4, 5$ (para cada grupo de idade).
- $j = 1, 2, 3, 4$ (individuo).

$$\hat{\mu}_i = \bar{y}_i \quad i = 1, 2, 3, 4, 5$$

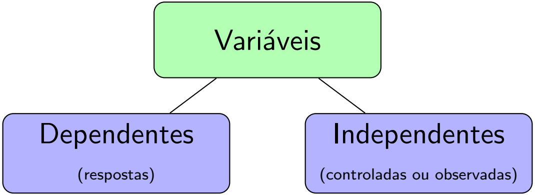
Idade - variável quantitativa

$$\mu = \mu(x)$$

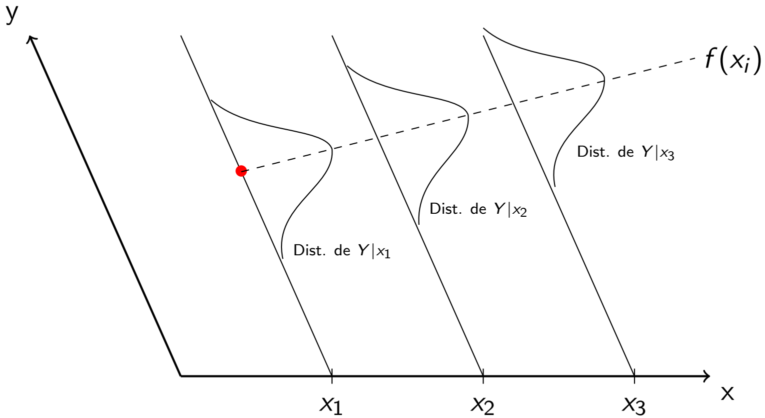
Introdução - Análise de Regressão

Objetivos:

- Na situação em que muitas variáveis quantitativas estão envolvidas, estudar o efeito que algumas variáveis exercem nas outras.
- Construção e análise de uma relação matemática entre as variáveis [No geral, uma variável em função das outras].
- Estudar como alterações nas variáveis independentes influem na variável dependente.
- Interessante obtermos uma relação simples de dependência [Variável dependente em função de uma ou poucas variáveis independentes].



Visualização



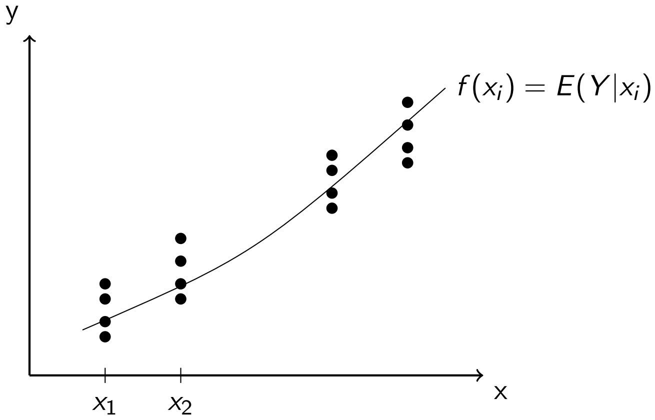
Quem é o ponto vermelho?

$$E[Y|x_i] = f(x_i) = \mu_i$$

Definição

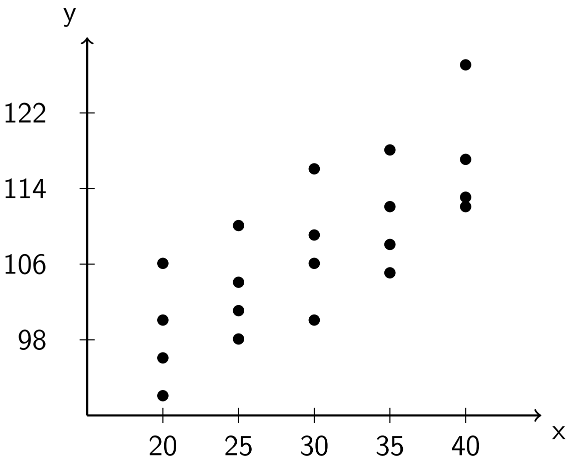
Definição: A função $f(X) = E(Y|X)$ é chamada de regressão de Y em X.

Dado x_i, $Y_i = E(Y|x_i) + \epsilon_i$, $\forall i$, ϵ_i: variável aleatória.



Exemplo

- Y - Tempo de reação.
- X - Idade.



Modelo de regressão linear simples

$$E(Y|x) = \beta_0 + \beta_1 x = f(x)$$

Regressão: Técnica estatística para investigar e modelar a relação entre variáveis quantitativas.

Utilização:

- Análise descritiva de dados:
  Na existência de variáveis relacionadas, medidas no mesmo elemento amostral, procura-se uma equação para resumir ou descrever os dados.
- Busca de uma relação funcional entre a variável resposta e as variáveis independentes (determinação de f)

Modelo de regressão linear simples

Utilização:

- Estimação de parâmetros
  - ◇ Ex: $\beta_1 = 1$ pode ter algum significado prático.
  - ◇ Circuito elétrico com resistência desconhecida R:
    - Várias correntes passam pelo circuito e a correspondente voltagem é medida.
    - Diagrama de dispersão - indicaria relação entre voltagem e corrente do tipo reta pela origem.
    - Pela Lei de Ohm: Voltagem $= R \times$ Corrente.
    - A análise de regressão pode ser usada para ajustar um modelo aos dados, produzindo uma estimativa de R.
- Previsão.
- Determinar quais variáveis independentes são importantes na explicação da variável dependente.

Observação

Como as conclusões da análise de regressão são condicionais aos valores observados da variável X, esforços devem ser feitos para se obter um conjunto de dados com valores representativos.

Comentários finais

Atualmente, métodos avançados de análise de dados estão amplamente disponíveis nos diversos softwares estatísticos. Então, por que devemos utilizar ainda o método de mínimos quadrados em nossa análise?

Comentários finais

Atualmente, métodos avançados de análise de dados estão amplamente disponíveis nos diversos softwares estatísticos. Então, por que devemos utilizar ainda o método de mínimos quadrados em nossa análise?

- Muitos métodos modernos são modificações ou ampliações do modelo de regressão linear. Por exemplo, entender o método de redes neurais é praticamente impossível sem um bom entendimento da metodologia de modelos de regressão linear.

Comentários finais

Atualmente, métodos avançados de análise de dados estão amplamente disponíveis nos diversos softwares estatísticos. Então, por que devemos utilizar ainda o método de mínimos quadrados em nossa análise?

- Muitos métodos modernos são modificações ou ampliações do modelo de regressão linear. Por exemplo, entender o método de redes neurais é praticamente impossível sem um bom entendimento da metodologia de modelos de regressão linear.
- A metodologia de modelos de regressão linear é praticamente transparente. Geralmente, é possível fazer gráficos para verificar se os modelos que estamos utilizando fazem sentido.

Comentários finais

Atualmente, métodos avançados de análise de dados estão amplamente disponíveis nos diversos softwares estatísticos. Então, por que devemos utilizar ainda o método de mínimos quadrados em nossa análise?

- Muitos métodos modernos são modificações ou ampliações do modelo de regressão linear. Por exemplo, entender o método de redes neurais é praticamente impossível sem um bom entendimento da metodologia de modelos de regressão linear.
- A metodologia de modelos de regressão linear é praticamente transparente. Geralmente, é possível fazer gráficos para verificar se os modelos que estamos utilizando fazem sentido.
- Usualmente, outros problemas para os quais o método de mínimos quadrados não é apropriado são obtidos a partir de pequenas modificações na metodologia de modelos de regressão linear.