Cálculo Diferencial e Integral para Estatísticos

Prof. Wagner H. Bonat

Universidade Federal do Paraná
Departamento de Estatística
Laboratório de Estatística e Geoinformação

Sumário

Funções

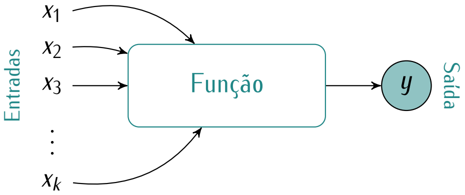
- **Definição 1** - Uma **função** escrita como $y = f(x)$ associa um número y a cada valor de x.
- x é chamada de variável **independente**.
- **Domínio de f(x)** é a faixa de valores que x pode assumir.
- y é chamada de variável **dependente**.
- **Imagem de f(x)** é a faixa de valores que y pode assumir.
- Resumindo temos,

$$\xrightarrow[\text{Independente}]{x \in D} f(x) \xrightarrow[\text{Dependente}]{y \in I}$$

- O **domínio** e **imagem** de uma função são intervalos.
- Tipos de intervalos:
  - Intervalo aberto **não contêm** as extremidades: Notação (a, b).
  - Intervalo fechado **contêm** as extremidades: Notação $[a, b]$.

Ideia intuitiva de função

Ideia intuitiva de função



Exemplo: Função

- Considere a função $y = x^2$.
- Em R temos

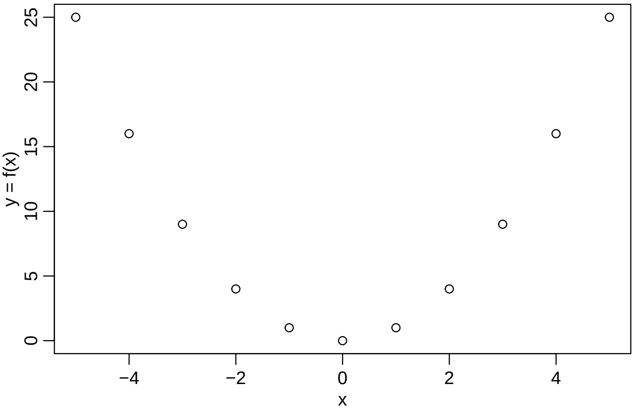
```
fx = function(x) {
  out <- x^2
  return(out)
}
```

- Avaliando a função em alguns pontos.

```
x <- c(-5,-4,-3,-2,-1,0,1,2,3,4,5)
fx(x)
```

```
#  [1] 25 16  9  4  1  0  1  4  9 16 25
```

Gráfico da função



Funções parametrizadas

- **Definição 2** - **Parâmetro** é uma quantidade conhecida que indexa ou parametriza uma determinada função.
- Em geral os parâmetros mudam o comportamento da função e descrevem quantidades/características de interesse.
- Notação: $y = f(x; \theta)$, onde θ denota o parâmetro.
- O conjunto de valores que θ pode assumir é chamado de espaço paramétrico.
- Notação $\theta \in \Theta$.

Exemplo: Função parametrizada

- Considere a seguinte função $y = (x - \theta)^2$.
- Em R temos

```
fx = function(x, theta) {
  out <- (x - theta)^2
  return(out)
}
```

- Avaliando a função em alguns pontos.

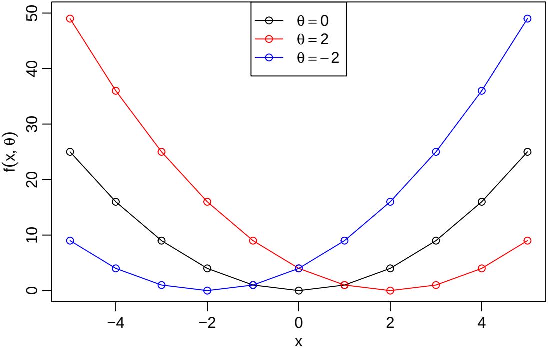
```
x <- c(-5,-4,-3,-2,-1,0,1,2,3,4,5)
fx(x = x, theta = -2)
```

```
#  [1]  9  4  1  0  1  4  9 16 25 36 49
```

```
fx(x = x, theta = 2)
```

```
#  [1] 49 36 25 16  9  4  1  0  1  4  9
```

Gráfico da função



Funções com vários parâmetros

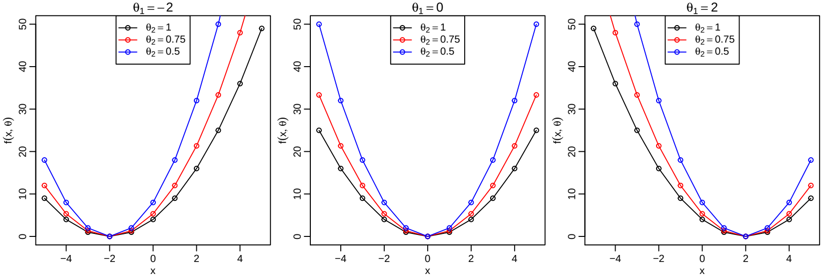
- Em geral uma função pode ter vários parâmetros.
- O ideal é que cada parâmetro controle um aspecto da função.
- Exemplo: $y = f(x; \theta_1, \theta_2)$ ou mais geral $y = f(x; \boldsymbol{\theta})$, onde $\boldsymbol{\theta}$ é um vetor de parâmetros.
- Função com dois parâmetros:

$$y = \frac{(x - \theta_1)^2}{\theta_2}.$$

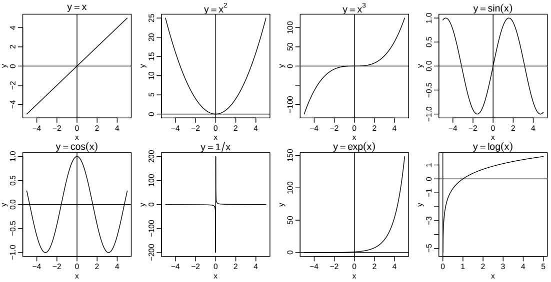
- Função em R.

```
fx = function(x, theta) {
  out <- ((x - theta[1])^2)/theta[2]
  return(out)
}
```

Gráfico da função



Exemplos de funções



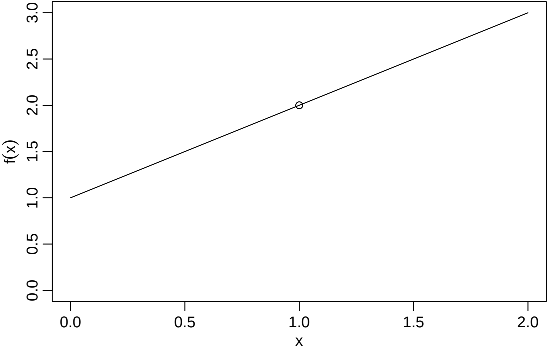
Limite de uma função

- **Definição 3** - Se uma função $f(x)$ se aproxima de um número L conforme x tende a um número a vindo da direita ou da esquerda, dizemos que o limite de $f(x)$ tende a L quando x tende a a.
- Notação
$$\lim_{x \to a} f(x) = f(a) = L.$$
- O limite pode não existir.
- Se o limite de uma função existe ele é único.

Exemplo: Limite de funções

- Considere o limite

$$\lim_{x \to 1} (x + 1) = 2.$$



Exemplo: Limite de funções

- Considere o limite

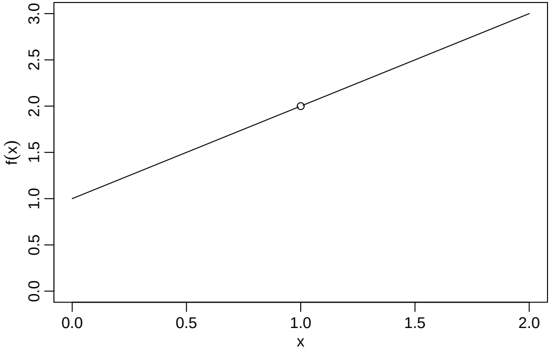
$$\lim_{x \to 1} \frac{x^2 - 1}{x - 1} = ?$$

```r
fx <- function(x) {
  out <- (x^2 - 1)/(x - 1)
  return(out)
}
fx(x = 1)
```

```
# [1] NaN
```

Exemplo: Limite de funções

- Graficamente temos



Exemplo: Limite de funções

- Note que

$$\lim_{x \to 1} \frac{x^2 - 1}{x - 1} = \lim_{x \to 1} \frac{(x+1)(x-1)}{x - 1} = \lim_{x \to 1} (x + 1) = 2.$$

- **Definição intuitiva:** O limite de uma função é o valor que achamos natural para ela em um determinado ponto.

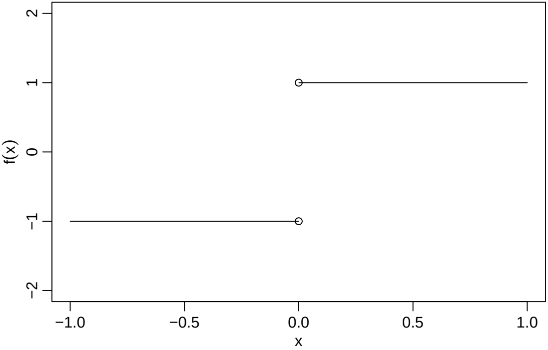
Continuidade de uma função

- **Definição 4** - Dizemos que uma função é **contínua** em $x = a$ se três condições forem satisfeitas: $f(a)$ existe, $\lim_{x \to a} f(x)$ existe e $\lim_{x \to a} f(x) = f(a)$.
- Continuidade significa que pequenas variações na variável independente levam a pequenas variações na variável dependente.
- **Teorema do valor intermediário**: Se a função $f(x)$ é contínua no intervalo fechado $[a, b]$, então existe pelo menos um número c em $[a, b]$ tal que $f(c) = M$.
- Implicação: Se $f(x)$ é contínua seu gráfico não contêm salto vertical.
- Em geral podemos pensar em funções contínuas como sendo funções suaves.

Exemplo: Função não contínua

- Considere a função não continua em 0.

$$\lim_{x \to 0} \frac{|x|}{x} = \begin{cases} -1 & x < 0 \\ 1 & x > 0. \end{cases}$$



Sumário

Derivada de uma função

- **Definição 5** - Derivada ordinária, derivada primeira, ou simplesmente, derivada de uma função $y = f(x)$ em um ponto $x = a$ no domínio de f é representada por $\frac{dy}{dx}$, y', $\frac{df}{dx}$ ou $f'(a)$ é o valor

$$\frac{dy}{dx}|_{x=a} = f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

Exemplo: Derivada de uma função

- Obtenha a derivada de $f(x) = -x^2$.

$$
\begin{aligned}
f'(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\
&= \lim_{h\to 0} \frac{-(x+h)^2 - (-x^2)}{h} \\
&= \lim_{h\to 0} \frac{-(x^2 + 2xh + h^2) + x^2}{h} \\
&= \lim_{h\to 0} \frac{-x^2 - 2xh - h^2}{h} = \frac{-2xh - h^2}{h} \\
&= \lim_{h\to 0} -2x - h = -2x.
\end{aligned}
$$

Interpretação da derivada

- Taxa de mudança instântanea.
- No limite quando $x \to a$ a derivada é a reta tangente ao ponto $(a, f(a))$.
- A reta tangente ao ponto a tem equação dada por $y - f(a) = f'(a)(x - a)$.
- Exemplo: Obtenha a reta tangente a $f(x)$ nos pontos $x = 2$ e $x = -2$.
- Temos $f(x = 2) = -4$ e $f'(x = 2) = -4$, assim

$$
\begin{aligned}
y - f(x = 2) &= f'(x = 2)(x - 2) \\
y - (-4) &= -4(x - 2) \\
y + 4 = &= -4x + 8 \\
y &= 4 - 4x
\end{aligned}
$$

Exemplo: Reta tangente a $f(x)$

- $f(x)$ e $f'(x)$.

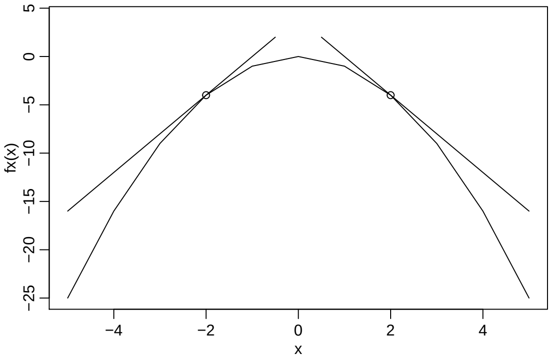
```
fx <- function(x) {
  out <- - x^2
  return(out)
}

f_prime <- function(x) {
  out <- -2*x
  return(out)
}
```

- Equação da reta $y = a + b * x$.

```
intercept = (fx(x = 2) - f_prime(x = 2)*2)
slope <- f_prime(x = 2)
c(intercept, slope)
```

```
# [1]  4 -4
```

Exemplo: Reta tangente a $f(x)$



Regras de derivação

- Seja $n \neq 0$ um natural. São válidas as fórmulas de derivação:
  1. Se $f(x) = c$ então $f'(x) = 0$.
  2. Se $f(x) = x^n$ então $f'(x) = nx^{n-1}$.
  3. Se $f(x) = x^{-n}$ então $f'(x) = -nx^{-n-1}$.
  4. Se $f(x) = x^{1/n}$ então $f'(x) = \frac{1}{n}x^{\frac{1}{n}-1}$.
- Derivada de funções especiais.
  1. Se $f(x) = \exp(x)$ então $f'(x) = \exp(x)$.
  2. Se $f(x) = \ln(x)$ então $f'(x) = \frac{1}{x}, x > 0$.
- Sejam $f(x)$ e $g(x)$ deriváveis em x e seja c uma constante. Então as funções $f(x) + g(x)$, $cf(x)$ e $f(x) \cdot g(x)$ são deriváveis em x e têm-se
  1. $(f + g)' = f'(x) + g'(x)$.
  2. $(cf)'(x) = cf'(x)$.
  3. $(f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x)$.

Regra da cadeia

- Regra da cadeia: Sejam $y = f(x)$ e $x = g(t)$ duas funções deriváveis, com $I \in D_f$. A função composta $h(t) = f(g(t))$ é derivável, sendo

$$h'(t) = f'(g(t))g'(t), t \in D_g.$$

- Existe uma infinidade de fórmulas de derivação.
- Na prática é comum usar um software de matemática simbólica como o wxMaxima.
- Em `R` as funções `deriv()` e `deriv3()`.
- Exemplos.

Por que derivadas são importantes?

- Derivada é a inclinação (slope) da reta tangente à curva $y = f(x)$.
- **Obtenção de máximo ou mínino de uma função (fundamental !!!).**
- O **máximo** de uma função $f(x)$ é o valor x_n tal que, $f(x_n) \geq f(x), \forall x \in D$.
- O **mínimo** de uma função $f(x)$ é o valor x_1 tal que, $f(x_1) \leq f(x), \forall x \in D$.

Problema: Redução de dados

- Suponha que temos um conjunto de observações y_i para $i = 1, \ldots, n$. Queremos resumir a informação contida em y_i em um único número, digamos μ.
- Problema: Como encontrar μ?

Problema: Redução de dados

- Suponha que temos um conjunto de observações y_i para $i = 1, \ldots, n$. Queremos resumir a informação contida em y_i em um único número, digamos μ.
- Problema: Como encontrar μ?
- Solução: Encontrar o valor μ, tal que $f(\mu) = \sum_{i=1}^{n} (y_i - \mu)^2$, seja a menor possível.
- Note que uma vez que temos os números observados y_i a única quantidade desconhecida é μ.
- Note que μ é o parâmetro da nossa função.
- A função $f(\mu)$ mede o quanto **perdemos** em representar y_i apenas usando μ.
- Funções perda muito populares são a **perda quadrática**, **perda absoluta**, **minmax** e a **cross entropia**.

Exemplo: Redução de dados

- Funções em R.

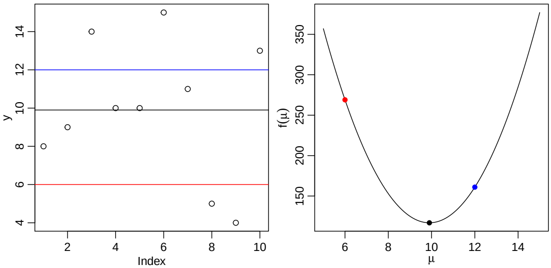
```
y <- c(8,9,14,10,10,15,11,5,4,13)

fmu <- function(mu, y) {
  out <- sum((y - mu)^2)
  return(out)
}

fmu <- Vectorize(fmu, "mu")
fmu(mu = c(10, 12), y = y)
```

```
# [1] 117 161
```

```
f_prime <- function(mu, y) {
  out <- -2*sum(y-mu)
  return(out)
}
```

Exemplo: Redução de dados

- Graficamente, temos

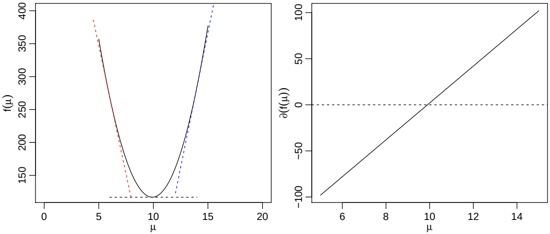


Exemplo: Redução de dados

- Note que o **melhor** resumo dos dados de um número, corresponde ao ponto de mínimo da função $f(\mu) = \sum_{i=1}^{n}(y_i - \mu)^2$.
- Como o mínimo está relacionado com a derivada de $f(\mu)$?

Exemplo: Redução de dados

- Note que o **melhor** resumo dos dados de um número, corresponde ao ponto de mínimo da função $f(\mu) = \sum_{i=1}^{n}(y_i - \mu)^2$.
- Como o mínimo está relacionado com a derivada de $f(\mu)$?



Exemplo: Redução de dados

- No ponto de mínimo/máximo a inclinação da reta tangente a $f(\mu)$ é zero.
- Denote por $\hat{\mu}$ o ponto de mínimo/máximo de $f(\mu)$, então $f'(\hat{\mu}) = 0$.
- Assim, temos (regra da cadeia!!)

$$
\begin{aligned}
f'(\mu) &= 2\sum_{i=1}^{n}(y_i - \mu)\frac{d}{d\mu}(y_i - \mu) \\
&= 2\sum_{i=1}^{n}(y_i - \mu)(-1) = -2\sum_{i=1}^{n}(y_i - \mu).
\end{aligned}
$$

Exemplo: Redução de dados

- Agora precisamos achar o ponto $\hat{\mu}$ tal que $f'(\hat{\mu}) = 0$.

$$
\begin{aligned}
f'(\hat{\mu}) &= 0 \\
-2\sum_{i=1}^{n}(y_i - \hat{\mu}) &= 0 \\
-\sum_{i=1}^{n} y_i + n\hat{\mu} &= 0 \\
n\hat{\mu} &= \sum_{i=1}^{n} y_i \\
\hat{\mu} &= \frac{\sum_{i=1}^{n} y_i}{n}.
\end{aligned}
$$

Máximos e mínimos

- Sejam $f(x)$ uma função que admite segunda derivada no intervalo aberto I e $x \in I$.
  1. $f'(x) = 0$ e $f''(x) > 0$ é ponto de mínimo local.
  2. $f'(x) = 0$ e $f''(x) < 0$ é ponto de máximo local.
- $f''(x)$ denota a segunda derivada de $f(x)$, i.e. $\frac{\partial f'(x)}{\partial x}$.
- Em geral em estatística e técnicas padrões de *machine learning* a função objetivo/perda é criada para ter apenas um ponto de mínimo/máximo.
- Em situações patológicas, tais como falta de identificabilidade a função pode ter mais de um mínimo/máximo.

Funções de duas ou mais variáveis indepentes

- Em geral uma função possui apenas uma variável dependente.
- Porém, pode ter duas ou mais variáveis independentes.
- Exemplo 1:

$$z = f(x, y) = \frac{x^2}{2^2} + \frac{y^2}{3^2}.$$

- Exemplo 2: Sejam y_i e x_i quantidades observadas.
- A equação da reta que relaciona x e y é dada por

$$y_i = f(\beta_0, \beta_1; x_i) = \beta_0 + \beta_1 x_i.$$

Derivadas parciais

- Para uma função $z = f(x, y)$, a derivada parcial de f em relação a x é representada por $\frac{\partial f}{\partial x}$, e é definida por

$$\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h},$$

desde que o limite exista.

- De forma similar, a derivada parcial de $f(x, y)$ em relação a y é representada por

$$\frac{\partial f}{\partial y} = \lim_{h \to 0} \frac{f(x, y+h) - f(x, y)}{h}.$$

- De forma geral, $\frac{\partial f(x)}{\partial x}$ é obtida derivando $f(x)$ considerando y fixo e vice-versa.

Gradiente e Hessiano

- O gradiente de uma função $f(x,y)$ é o vetor composto pelas derivadas primeira de $f(x,y)$ em relação a x e y, i.e.

$$\nabla f(x,y) = \left(\frac{\partial f(x,y)}{\partial x}, \frac{\partial f(x,y)}{\partial y} \right).$$

- O Hessiano de uma função $f(x,y)$ é a matrix

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 f(x,y)}{\partial x^2} & \frac{\partial f(x,y)}{\partial x \partial y} \\ \frac{\partial f(x,y)}{\partial y \partial x} & \frac{\partial^2 f(x,y)}{\partial x^2} \end{pmatrix}.$$

- As definições se estendem naturalmente para mais de duas variáveis.
- Vamos revisar as ideias de vetores e matrizes (álgebra linear).

Expansão de funções em Série de Taylor

- Aproximação por Série de Taylor é fundamental em estatística e métodos numéricos.
- É uma forma simples de obter o valor aproximado de uma função perto de um ponto conhecido x_0.
- Dada uma função $f(x)$ derivável $(n+1)$ vezes em um intervalo contendo um ponto $x = x_0$, temos

$$\begin{aligned} f(x) &= f(x_0) + (x - x_0)\frac{df(x)}{dx}|_{x=x_0} + \frac{(x-x_0)^2}{2!}\frac{d^2 f(x)}{dx^2}|_{x=x_0} + \\ \ldots &+ \frac{(x-x_0)^n}{n!}\frac{d^n f(x)}{dx^n}|_{x=x_0} + R_n(x). \end{aligned}$$

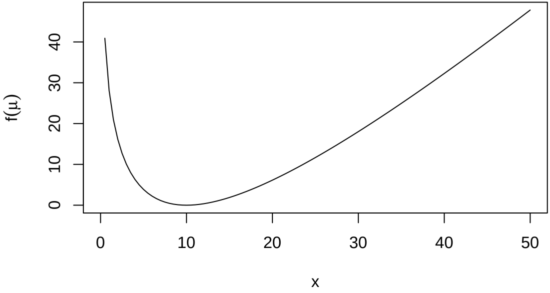
- Aproximação similar é possível para funções com duas ou mais variáveis.

Exemplo: Expansão em Série de Taylor

- Considere a seguinte função

$$f(\mu) = 2\left(10 \log \frac{10}{\mu} - 10 + \mu\right).$$

Faça uma aproximação em série de Taylor de primeira e segunda ordem ao redor do ponto $\mu_0 = 10$.



Exemplo: Expansão em Série de Taylor

- Primeira derivada

$$f'(\mu) = 2\left(1 - \frac{10}{\mu}\right).$$

- Segunda derivada

$$f''(\mu) = \frac{20}{\mu^2}.$$

- Aproximação em série de Taylor (segunda ordem) ao redor de μ_0.

$$f(\mu) \approx f(\mu_0) + (\mu - \mu_0)f'(\mu = \mu_0) + \frac{(\mu - \mu_0)^2}{2!} f''(\mu = \mu_0).$$

Exemplo: Expansão em Série de Taylor

- Aproximação de Taylor (genérica)

```
taylor_ap <- function(mu, mu0, f, f_prime, f_dprime) {
  app <- f(mu = mu0) + (mu - mu0)*f_prime(mu = mu0) +
    (((mu - mu0)^2)/(2))*f_dprime(mu = mu0)
  return(app)
}
```

- Derivadas

```
f_prime <- function(mu) {2*(1- 10/mu)}
f_dprime <- function(mu) {20/mu^2}
```

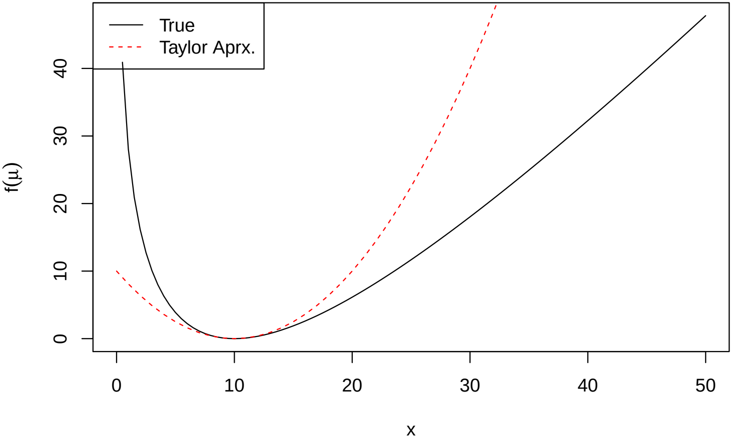
- Aproximando a função

```
taylor_ap(mu = c(9,10,11), mu0 = 10, f = f,
          f_prime = f_prime, f_dprime = f_dprime)
```

```
# [1] 0.1 0.0 0.1
```

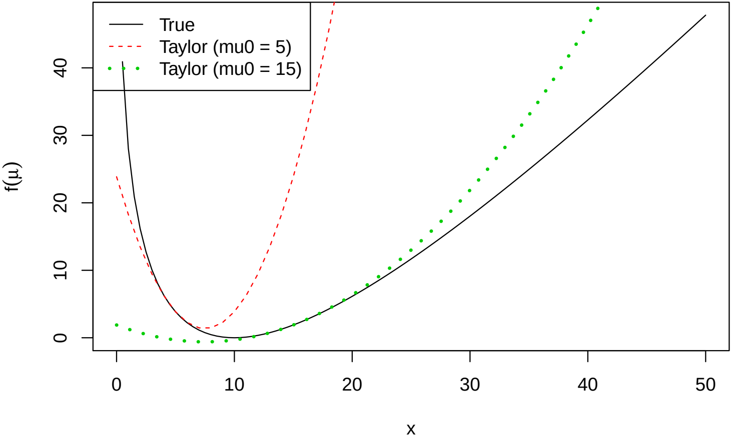
Exemplo: Expansão em Série de Taylor

- Graficamente, temos



Exemplo: Expansão em Série de Taylor

- Graficamente, temos



Exemplo - Regressão linear simples

- Seja $y_i (i = 1, \dots, n)$ observações de alguma variável de interesse.
- Seja x_i uma outra variável que queremos relacionar com y_i através de um reta, i.e.

$$y_i = \beta_0 + \beta_1 x_i.$$

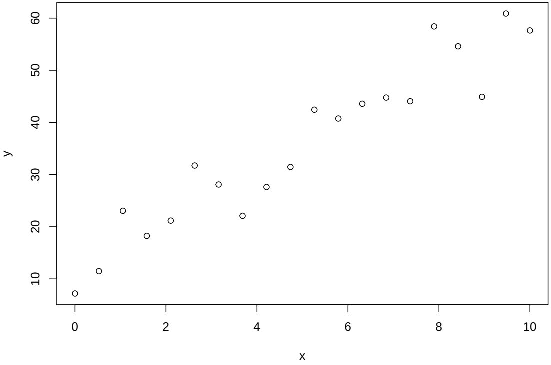
- Problema: Encontrar β_0 e β_1 tal que

$$SQ = f(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2,$$

seja a menor possível.

Exemplo - Regressão linear simples

- Graficamente, temos



Como encontrar β_0 e β_1 que minimizam a SQ?

- Abordagem 1
  1. Obter o vetor gradiente

$$\nabla f(\beta_0, \beta_1) = \left(\frac{\partial f(\beta_0, \beta_1)}{\partial \beta_0}, \frac{\partial f(\beta_0, \beta_1)}{\partial \beta_1} \right).$$

  2. Encontrar $\hat{\beta}_0$ e $\hat{\beta}_1$ tal que

$$\nabla f(\hat{\beta}_0, \hat{\beta}_1) = \mathbf{0}.$$

- Abordagem 2 - Obter a derivada analiticamente, mas resolver o sistema numericamente.
- Abordagem 3 - Obter a derivada e resolver o sistema numericamente.
- Abordagem 4 - Usar um algoritmo de otimização genérico.

Vetor gradiente

- Chame $y_i - (\beta_0 + \beta_1 x_i) = \epsilon_i$.
- Chame $\beta_0 + \beta_1 x_i = \mu_i$.

$$\nabla f(\beta_0, \beta_1) = \left(\frac{\partial f(\beta_0, \beta_1)}{\partial \epsilon_i} \frac{\partial \epsilon_i}{\mu_i} \frac{\partial \mu_i}{\partial \beta_0}, \frac{\partial f(\beta_0, \beta_1)}{\partial \epsilon_i} \frac{\partial \epsilon_i}{\mu_i} \frac{\partial \mu_i}{\partial \beta_1} \right).$$

onde

$$\frac{\partial f(\beta_0, \beta_1)}{\partial \epsilon_i} = \frac{\partial}{\partial \epsilon_i} \sum_{i=1}^{n} \epsilon_i^2 = 2 \sum_{i=1}^{2} \epsilon_i.$$

$$\frac{\partial \epsilon_i}{\partial \mu_i} = \frac{\partial}{\partial \mu_i} (y_i - \mu_i) = -1.$$

$$\frac{\partial \mu_i}{\partial \beta_0} = \frac{\partial}{\partial \beta_0} \beta_0 + \beta_1 x_i = 1.$$

$$\frac{\partial \mu_i}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} \beta_0 + \beta_1 x_i = x_i$$

Vetor gradiente

- Vetor gradiente,

$$\nabla f(\beta_0, \beta_1) = \left(-2 \sum_{i=1}^{n} \epsilon_i (1); -2 \sum_{i=1}^{n} \epsilon_i x_i \right)$$

$$= \left(-2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i); -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i \right).$$

- Resolver o sistema de equações:

$$-2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (1)$$

$$-2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0 \quad (2)$$

Resolvendo o sistema de equações

- Pela Eq. (1) temos,

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \tag{3}$$

- Substituindo Eq.(3) em Eq. (2) e resolvendo em $\hat{\beta}_1$, temos

$$\hat{\beta}_1 = \frac{\bar{y} \sum_{i=1}^{n} - \sum_{i=1}^{n} y_i x_i}{\bar{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2}.$$

Implementação em R

- Simulando um conjunto de dados.

```
set.seed(123)
b0 = 10
b1 = 5
x <- seq(0,10, l = 20)
mu <- b0 + b1*x
y <- rnorm(20, mean = mu, sd = 5)
head(cbind(y, x), 4)
```

```
#              y         x
# [1,]  7.197622 0.0000000
# [2,] 11.480691 0.5263158
# [3,] 23.056699 1.0526316
# [4,] 18.247279 1.5789474
```

Implementação em R

- Fazendo as contas.

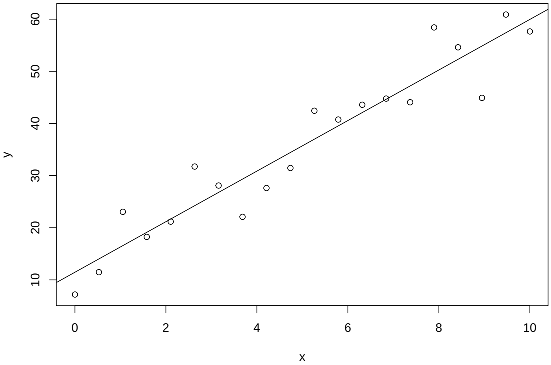
```
b1 <- (mean(y)*sum(x) - sum(y*x))/(mean(x)*sum(x) - sum(x^2))
b0 <- mean(y) - b1*mean(x)
c(b0,b1)
```

```
# [1] 11.470239  4.847576
```

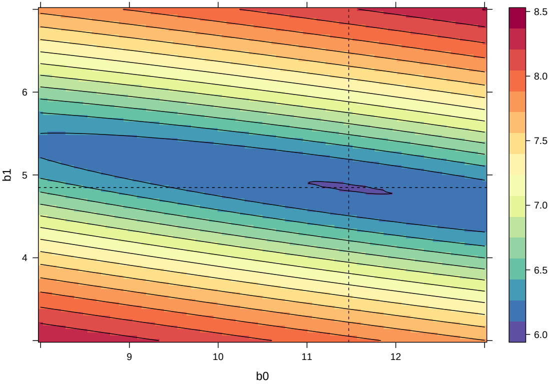
```
# Conferindo
coef(lm(y ~ x))
```

```
# (Intercept)           x
#   11.470239    4.847576
```

Visualizando a reta ajustada



Visualizando a superfície objetivo



Discussão

- Derivadas são essenciais em estatística.
- Maximizar/Minimizar funções perda/objetivo.
- O cálculo é por vezes difícil e tedioso.
- Solução de sistemas lineares é tedioso quando possível.
- Álgebra linear ajuda a generalizar as soluções.
- Em situações mais gerais expressões analíticas não serão possíveis de obter.
- Métodos numéricos para resolução de sistemas lineares.
- Métodos numéricos para resolução de sistemas não-lineares.
- Métodos de otimização numérica.

Sumário

Diferenciação numérica

- Derivada dá uma medida da taxa na qual a variável y muda devido a uma mudança na variável x.
- A função a ser diferenciada pode ser dada por uma função $f(x)$, ou apenas por um conjunto de pontos (y_i, x_i).
- Quando devemos usar derivadas numéricas?
  1. $f'(x)$ é dificil de obter analiticamente.
  2. $f'(x)$ é caro para calcular computacionalmente.
  3. Quando a função é especificada apenas por um conjunto de pontos.
- Abordagens para a diferenciação numérica
  1. Aproximação por diferenças finitas;
  2. Aproximar a função por uma outra função de fácil derivação.

Aproximação da derivada por diferenças finitas

- Derivada $f'(x)$ de uma função $f(x)$ no ponto $x = a$ é definida como:

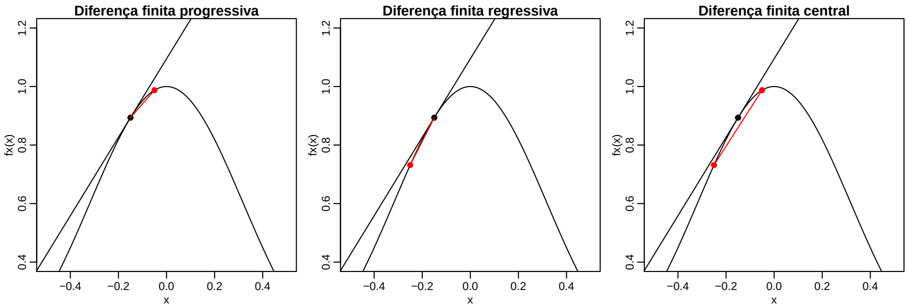
$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

- Derivada é o valor da inclinação da reta tangente à função em $x = a$.
- Escolhe-se um ponto x próximo a a e calcula-se a inclinação da reta que conecta os dois pontos.
- A precisão do cálculo aumenta quando x aproxima de a.
- Aproximação numérica: função será avaliada em diferentes pontos próximos a a para aproximar $f'(a)$.

Aproximação da derivada por diferenças finitas

- Fórmulas para diferenciação numérica:
  1. Diferença progressiva: Inclinação da reta que conecta os pontos $(x_i, f(x_i))$ e $(x_{i+1}, f(x_{i+1}))$:
  $$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}.$$
  2. Diferença regressiva: Inclinação da reta que conecta os pontos $(x_{i-1}, f(x_{i-1}))$ e $(x_i, f(x_i))$:
  $$f'(x_i) = \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}.$$
  3. Diferença central: Inclinação da reta que conecta os pontos $(x_{i-1}, f(x_{i-1}))$ e $(x_{i+1}, f(x_{i+1}))$:
  $$f'(x_i) = \frac{f(x_{i+1}) - f(x_{i-1})}{x_{i+1} - x_{i-1}}.$$

Ilustração: Derivada por diferenças finitas



Aproximação da derivada por diferenças finitas

- Diferença progressiva

```
dif_prog <- function(fx, x, h) {
  df <- (fx(x + h) - fx(x))/( (x + h) - x)
  return(df)
}
```

- Diferença regressiva

```
dif_reg <- function(fx, x, h) {
  df <- (fx(x) - fx(x - h))/( x - (x - h))
  return(df)
}
```

- Diferença central

```
dif_cen <- function(fx, x, h) {
  df <- (fx(x + h) - fx(x - h))/( (x + h) - (x - h))
  return(df)}
```

Exemplo: Aproximação da derivada por diferenças finitas

- Considere $f(x) = x^3$, assim $f'(x) = 3x^2$.

```
fx <- function(x) x^3
# Diferença progressiva
dif_prog(fx, x = 2, h = 0.001)
```

```
# [1] 12.006
```

```
# Diferença regressiva
dif_reg(fx, x = 2, h = 0.001)
```

```
# [1] 11.994
```

```
# Diferença central
dif_cen(fx, x = 2, h = 0.001)
```

```
# [1] 12
```

```
# Exata
3*2^2
```

Diferenças finitas usando expansão em série de Taylor

- As fórmulas anteriores podem ser deduzidas usando expansão em série de Taylor.
- O número de pontos para aproximar a derivada pode mudar.
- Vantagem da dedução por série de Taylor é que ela fornece uma estimativa do erro de truncamento.

Diferença finita progressiva com dois pontos

- Aproximação de Taylor para o ponto x_{i+1}

$$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f'''(x_i)}{3!}h^3 + \ldots,$$

  onde $h = x_{i+1} - x_i$.
- Fixando dois termos e deixando os outros termos como um resíduo, temos

$$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{f''(\xi)}{2!}h^2.$$

- Resolvendo para $f'(x_i)$, temos

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{h} - \frac{f''(\xi)}{2!}h^2.$$

- Erro de truncamento,

$$-\frac{f''(\xi)}{2!}h^2 = O(h).$$

Diferença finita regressiva com dois pontos

- Aproximação de Taylor para o ponto x_{i-1}

$$f(x_{i-1}) = f(x_i) - f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f'''(x_i)}{3!}h^3 + \ldots,$$

  onde $h = x_i - x_{i-1}$.
- Fixando dois termos e deixando os outros termos como um resíduo, temos

$$f(x_{i-1}) = f(x_i) - f'(x_i)h + \frac{f''(\xi)}{2!}h^2.$$

- Resolvendo para $f'(x_i)$, temos

$$f'(x_i) = \frac{f(x_i) - f(x_{i-1})}{h} + \frac{f''(\xi)}{2!}h^2.$$

- Erro de truncamento,

$$\frac{f''(\xi)}{2!}h^2 = O(h).$$

Diferença finita central com dois pontos

- Aproximação de Taylor para o ponto x_{i+1}

$$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f'''(\xi_1)}{3!}h^3,$$

onde ξ_1 está entre x_i e x_{i+1}.

- Aproximação de Taylor para o ponto x_{i-1}

$$f(x_{i-1}) = f(x_i) - f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f'''(\xi_1)}{3!}h^3,$$

onde ξ_2 está entre x_{i-1} e x_i.

- Subtraindo as equações acima, temos

$$f(x_{i+1}) - f(x_{i-1}) = 2f'(x_i)h + \frac{f'''(\xi_1)}{3!}h^3 + \frac{f'''(\xi_1)}{3!}h^3.$$

- Resolvendo para $f'(x_i)$, temos

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_{i-1})}{2h} + O(h^2).$$

Diferença finita progressiva com três pontos

- Aproxima $f'(x_i)$ avaliando a função no ponto e nos dois pontos seguintes x_{i+1} e x_{i+2}.
- Aproximação de Taylor em x_{i+1} e x_{i+2},

$$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f'''(\xi_1)}{3!}h^3, \qquad (4)$$

$$f(x_{i+2}) = f(x_i) + f'(x_i)2h + \frac{f''(x_i)}{2!}(2h)^2 + \frac{f'''(\xi_2)}{3!}(2h)^3. \qquad (5)$$

- Equações 4 e 5 são combinadas de forma que os termos com derivada segunda desapareçam.
- Multiplicando Eq. 4 por 4 e subtraindo Eq. 5, temos

$$4f(x_{i+1}) - f(x_{i+2}) = 3f(x_i) + 2f'(x_i)h + \frac{4f'''(\xi_1)}{3!}h^3 - \frac{f'''(\xi_2)}{3!}(2h)^3.$$

Diferença finita com três pontos

- Resolvendo em $f'(x_i)$, temos

$$f'(x_i) = \frac{-3f(x_i) + 4f(x_{i+1}) - f(x_{i+2})}{2h} + O(h).$$

- Diferença finita regressiva com três pontos

$$f'(x_i) = \frac{f(x_{i-2}) - 4f(x_{i-1}) + 3f(x_i)}{2h} + O(h).$$

Fórmulas de diferenças finitas para a segunda derivada

- Usando as mesmas idéias podemos aproximar a derivada segunda de uma função qualquer por diferenças finitas.
- A derivação das fórmulas são idênticas, porém mais tediosas.
- Fórmula diferença central com três pontos para a derivada segunda

$$f''(x_i) = \frac{f(x_{i-1}) - 2f(x_i) + f(x_{i+1})}{h^2} + O(h^2).$$

- Diferença central com quatro pontos

$$f''(x_i) = \frac{-f(x_{i-2}) + 16f(x_{i-1}) - 30f(x_i) + 16f(x_{i+1}) - f(x_{i+2})}{12h^2} + O(h^4)$$

Fórmulas de diferenças finitas para a segunda derivada

- Diferença progressiva com três pontos

$$f''(x_i) = \frac{f(x_i) - 2f(x_{i+1}) + f(x_{i+2})}{h^2} + O(h).$$

- Diferença regressiva com três pontos

$$f''(x_i) = \frac{f(x_{i-2}) - 2f(x_{i-1}) + f(x_i)}{h^2} + O(h).$$

- Uma infinidade de fórmulas de várias ordens estão disponíveis.
- Fórmulas de diferenciação podem ser obtidas usando polinômios de Lagrange.

Erros na diferenciação numérica

- Em todas as fórmulas o erro de truncamento é função de h.
- h é o espaçamento entre os pontos, i.e. $h = x_{i+1} - x_i$.
- Fazendo h pequeno o erro de truncamento será pequeno.
- Em geral usa-se a precisão da máquina, algo como $1e^{-16}$.
- O erro de arredondamento depende da precisão finita de cada computador.
- Mesmo que h possa ser tão pequeno quanto desejado o erro de arredondamento pode crescer quando se diminue h.

Extrapolação de Richardson

- Extrapolação de Richardson é usada para obter uma aproximação mais precisa da derivada a partir de duas aproximações menos precisas.
- Considere o valor D de uma derivada (desconhecida) calculada pela fórmula

$$D = D(h) + k_2 h^2 + k_4 h^4, \tag{6}$$

  onde $D(h)$ aproxima D e k_2 e k_4 são termos de erro.
- O uso da mesma fórmula, porém com espaçamento $h/2$ resulta

$$D = D(\frac{h}{2}) + k_2 \left(\frac{h}{2}\right)^2 + k_4 \left(\frac{h}{2}\right)^4. \tag{7}$$

Extrapolação de Richardson

- A Eq. 7 pode ser rescrita (após multiplicar por 4):

$$4D = 4D(\frac{h}{2}) + k_2 h^2 + k_4 \frac{h^4}{4}. \tag{8}$$

- Subtraindo 6 de 8 elimina os termos com h^2 e fornece

$$3D = 4D(\frac{h}{2}) + D(h) - k_4 \frac{3h^4}{4}. \tag{9}$$

- Resolvendo 9, temos

$$D = \frac{1}{3}\left(4D(\frac{h}{2}) + D(h)\right) - k_4 \frac{h^4}{4}. \tag{10}$$

Extrapolação de Richardson

- O erro na Eq. 10 é agora $O(h^4)$. O valor de D é aproximado por

$$D = \frac{1}{3}\left(4D(\frac{h}{2}) + D(h)\right) + O(h^4).$$

- A partir de duas aproximações de ordem inferiores, obtemos uma aproximação de $O(h^4)$ mais precisa.
- Procedimento a partir de duas aproximações com erro $O(h^4)$ mostra que

$$D = \frac{1}{15}\left(16D(\frac{h}{2}) + D(h)\right) + O(h^6).$$

- Aproximação ainda mais precisa.

Exemplo: Extrapolação de Richardson

- Calcule a derivada de $f(x) = \frac{2^x}{x}$ no ponto $x = 2$.
- Solução exata: $\frac{\log(2)2^x}{x} - \frac{2^x}{x^2}$.
- Solução numérica usando diferença central

```
fx <- function(x) (2^x)/x
fpx <- function(x)(log(2)*(2^x))/x - (2^x)/x^2
erro <- fpx(x = 2)/dif_cen(fx = fx, x = 2, h = 0.2)
(erro-1)*100
```

```
# [1] 0.345544
```

- Extrapolação de Richardson

```
D2 <- dif_cen(fx = fx, x = 2, h = 0.2/2)
D <- dif_cen(fx = fx, x = 2, h = 0.2)
der <- (1/3)*( 4*D2 - D)
erro2 <- fpx(x = 2)/der
(erro2-1)*100
```

```
# [1] -0.001585268
```

Sumário

Derivadas parciais

- Para funções com muitas variáveis, a derivada parcial da função em relação a uma das variáveis representa a taxa de variação da função em relação a essa variável, mantendo as demais constantes.
- Assim, as fórmulas de diferenças finitas podem ser usadas no cálculo das derivadas parciais.
- As fórmulas são aplicadas em cada uma das variáveis, mantendo as outras fixas.
- A mesma ideia se aplica para derivadas de mais alta ordem.

Implementação: Derivadas parciais

- Derive $f(\beta_0, \beta_1) = \sum_{i=1}^{n} |y_i - (\beta_0 + \beta_1 x_i)|$.
- Fórmula dois pontos central

```
dif_cen <- function(fx, pt, h, ...) {
  df <- (fx(pt + h, ...) - fx(pt - h, ...))/( (pt + h) - (pt - h))
  return(df)
}
```

- Função a ser diferenciada

```
fx <- function(par, y, x1) {sum ( abs( y - (par[1] + par[2]*x1)) )}
```

- Gradiente usando diferenças finita.

```
grad_fx <- function(fx, par, h, ...) {
  fbeta0 <- function(beta0, beta1, y, x) fx(par = c(beta0, beta1), y = y, x = x)
  fbeta1 <- function(beta1, beta0, y, x) fx(par = c(beta0, beta1), y = y, x = x)
  db0 <- dif_cen(fx = fbeta0, pt = par[1], h = h, beta1 = par[2], y = y, x = x)
  db1 <- dif_cen(fx = fbeta1, pt = par[2], h = h, beta0 = par[1], y = y, x = x)
  return(c(db0, db1))
}
```

Exemplo: Derivadas parciais

- Simulando y_i's e x_i's.

```
set.seed(123)
x <- runif(100)
y <- rnorm(100, mean = 2 + 3*x, sd = 1)
```

- Gradiente numérico

```
grad_fx(fx = fx, par = c(2, 3), h = 0.001, y = y, x1 = x)
```

```
# [1] 6.000000 2.272805
```

- Gradiente analítico

```
c(sum(((y - 2 - 3*x)/abs(y - 2 -3*x))*(-1)),
  sum(((y - 2 - 3*x)/abs(y - 2 -3*x))*(-x)))
```

```
# [1] 6.000000 2.272805
```

Sumário

Uso de funções residentes do R para diferenciação numérica.

- Pacote `numDeriv` implementa derivadas por diferença finita.
- Gradiente

```
require(numDeriv)
args(grad)
```

```
# function (func, x, method = "Richardson", method.args = list(),
#     ...)
# NULL
```

- Hessiano

```
args(hessian)
```

```
# function (func, x, method = "Richardson", method.args = list(),
#     ...)
# NULL
```

Exemplo de aplicação

```r
grad(func = fx, x = c(2, 3), y = y, x1 = x)
```

```
# [1] 6.000000 2.272805
```

```r
hessian(func = fx, x = c(2, 3), y = y, x1 = x)
```

```
#          [,1]     [,2]
# [1,] 58.91271 29.53710
# [2,] 29.53710 48.86648
```