MOODLE ACADÊMICO

Iniciado em	Thursday, 18 Apr 2024, 17:27
Estado	Finalizada
Concluída em	Thursday, 18 Apr 2024, 19:14
Tempo empregado	1 hora 46 minutos
Avaliar	**10,00** de um máximo de 10,00(**100**%)

Questão **1**

Correto

Atingiu 2,00 de 2,00

Faça o download do arquivo T1_q1.rda.

Indique o diretório onde está salvo o arquivo e execute a seguinte rotina:

```
load(file = "T1_q1.rda")
```

Resolva os itens abaixo, referentes aos objetos presentes no arquivo `T1_q1.rda`.

a. Considerando o objeto `lista.q1`, qual é o valor da soma de todos os elementos dessa lista localizados na linha 3 da coluna z? ✔
b. Considerando o objeto `lista.q1`, qual é o valor do elemento localizado na posição 87 dessa lista e na linha 1 da coluna y? ✔
c. Considerando o *data.frame* `data.q1`, qual é o valor da soma de todos os elementos da variável X70? ✔
d. Considerando o *data.frame* `data.q1`, qual é o valor do elemento na posição 2 da variável X148? ✔
e. Considerando o objeto `matriz.q1`, qual é o valor da soma de todos os elementos pares presentes nessa matriz? ✔

Questão **2**

Correto

Atingiu 2,00 de 2,00

Faça o download do arquivo T1_q2_c.rda.

Indique o diretório onde está salvo o arquivo e execute a seguinte rotina:

```
load(file = "T1_q2_c.rda")
```

Resolva esta questão utilizando o objeto `data_q2`, presente no arquivo `T1_q2_c.rda`.

Obs: Arredonde cada resposta para três casas decimais e multiplique o resultado por 1000 (`round(resp, 3) * 1000`). Ou seja, a resposta fornecida por você deverá ser um número inteiro.

Questão: Numa urna estão quatro bolas numeradas de 1 a 4. Duas bolas são retiradas ao acaso e sem reposição. Esse experimento é repetido um grande número de vezes. Uma amostra desse experimento está armazenada no objeto `data_q2`. Com base nesses resultados, estime a probabilidade de que:

a. A média aritmética simples entre os dois valores retirados seja 2 ou 3. ✔

Questão **3**

Correto

Atingiu 2,00 de 2,00

Considere o seguinte modelo de regressão não-paramétrica: $Y_i = m(X_i) + \epsilon_i,\ i = 1, \ldots, n$, onde $m(\cdot)$ é uma função não conhecida de classe C^2.

Seja $(X_1, Y_1), \ldots, (X_n, Y_n)$ uma sequência de pares $iid.$ O estimador não-paramétrico *Nadaraya-Watson* para $m(\cdot)$, no ponto x_0, é dado por:

$\hat{m}(x_0) = \frac{\sum_{i=1}^n k\left(\frac{X_i - x_0}{h}\right) Y_i}{\sum_{i=1}^n k\left(\frac{X_i - x_0}{h}\right)}$, onde $h > 0$ é a *bandwidth* e $k(\cdot)$ é uma densidade simétrica ao redor de zero.

- Crie uma função que implementa o estimador *Nadaraya-Watson* num ponto x_0, com os seguintes argumentos:
  - `x`: vetor com os valores referentes à v.a. X;
  - `y`: vetor com os valores referentes à v.a. Y;
  - `x0`: escalar referente ao ponto x_0;
  - `h`: *bandwidth* (escalar);
  - `k`: função referente à densidade kernel.

  Essa função deve retornar o escalar $\hat{m}(x_0)$.
- Em seguida, crie uma função que implementa o estimador *Nadaraya-Watson* para um *grid* de pontos x_0, com os seguintes argumentos:
  - `x`: vetor com os valores referentes à v.a. X;
  - `y`: vetor com os valores referentes à v.a. Y;
  - `x0`: vetor referente aos pontos x_0;
  - `h`: *bandwidth* (escalar);
  - `k`: função referente à densidade kernel.

  Essa função deve retornar um vetor, cujo i-ésimo elemento seja igual a $\hat{m}$ avaliado no i-ésimo ponto do vetor x_0, i.e., $\hat{m}(x_0[i])$.

Faça o download do arquivo T1_q3.csv. Indique o diretório onde está salvo o arquivo e execute a seguinte rotina:

```
dados <- read.csv("T1_q3.csv", header = TRUE)

x <- dados$x
y <- dados$y
x0 <- seq(from = 0.1, to = 0.9, by = 0.05)
```

Aqui `x` e `y` representam uma amostra do par (X, Y) e `x0` representa o grid de pontos nos quais se deseja estimar $m(\cdot)$.

Use o kernel Epanechnikov: $k(x) = 0.75(1 - x^2),\ -1 < x < 1.$ Considere $h = 0.05.$

Obs: Arredonde o valor absoluto de cada resposta para três casas decimais e multiplique o resultado por 1000 (`round(abs(resp), 3) * 1000`). Ou seja, a resposta fornecida por você deverá ser um número inteiro não-negativo.

a. Qual é a estimativa de $m(\cdot)$ no ponto $x_0 = 0.25$? ✔
b. Qual é a média das estimativas de $m(\cdot)$ nos pontos do grid fornecido? Ou seja, qual é $\frac{1}{ng}\sum_{i=1}^{ng} \hat{m}(x_{0i})$, onde ng é o tamanho do grid. ✔

Questão **4**

Correto

Atingiu 2,00 de 2,00

Para resolver essa questão utilize a seguinte função que implementa o algoritmo **Newton-Raphson** no caso univariado.

```
f_nr <- function(
  x0, # chute inicial
  f1, # f1 := f(x), onde f(x) = 0
  f2, # f2 := f'(x)
  eps = 1 / 10000, # precisão
  ... # demais argumentos de f1 e f2
) {
  cc <- eps + 1
  conta <- 0
  trajetoria <- x0
  while (cc > eps) {
    x1 <- x0 - f1(x0, ...) / f2(x0, ...)
    cc <- (x1 - x0)^2
    x0 <- x1
    trajetoria <- c(trajetoria, x1)
    conta <- conta + 1
  }
  return(
    list(
      ponto_otimo = x1,
      num_iter = conta,
      trajetoria = trajetoria
    )
  )
}
```

Seja $X_1, X_2, \ldots, X_n$ uma sequência *iid*, onde $X_1 \sim Exp(\lambda)$. Estamos interessados em maximizar, com respeito a λ, a função log-verossimilhança associada a esse problema, usando o método de Newton.

Considere que a densidade $X_1 \sim Exp(\lambda)$ seja dada por:

$$f_{X_1}(\lambda, x) = \lambda \exp(-\lambda x), \;\; x > 0, \;\; \lambda > 0.$$

Considere também que a função log-verossimilhança associada a esse problema seja dada por:

$L(\lambda, \vec{x}) = n^{-1} \sum_{i=1}^{n} \log(f_{X_1}(\lambda, x_i))$.

Usando os dados disponíveis em T1_q4.csv, implemente o algoritmo **Newton-Raphson**, usando a função `f_nr`, para encontrar a estimativa de máxima verossimilhança do parâmetro λ.

Não modifique o valor *default* do argumento `eps`, dado na função `f_nr`.

Dicas:

- Para ler os dados, utilize a seguinte rotina:

```
dados <- read.csv("T1_q4.csv", header = T)

x <- dados$x
```

Obs: Arredonde o valor absoluto de cada resposta para três casas decimais e multiplique o resultado por 1000 (`round(abs(resp), 3) * 1000`). Ou seja, a resposta fornecida por você deverá ser um número inteiro não-negativo.

a. Considerando o banco de dados fornecido, qual é a estimativa de λ no ponto de ótimo, se for usado como chute inicial para λ o inverso da mediana da amostra fornecida (`1 / median(x)`)? ✔

b. Qual é o valor da derivada primeira da função log-verossimilhança com respeito a λ, $\left(\frac{\partial\left[n^{-1}\sum_{i=1}^{n}\log(f(x_i;\lambda))\right]}{\partial\lambda}\right)$, considerando o banco de dados fornecido e avaliada no ponto $\lambda = 4$? ✔

c. Qual é o valor da derivada segunda da função log-verossimilhança com respeito a λ, $\left(\frac{\partial^2\left[n^{-1}\sum_{i=1}^{n}\log(f(x_i;\lambda))\right]}{\partial\lambda^2}\right)$, considerando o banco de dados fornecido e avaliada no ponto $\lambda = 2$? ✔

Questão **5**

Correto

Atingiu 2,00 de 2,00

Faça o download do arquivo T1_q5.rda.

Indique o diretório onde está salvo o arquivo e execute a seguinte rotina:

```
load(file = "T1_q5.rda")
```

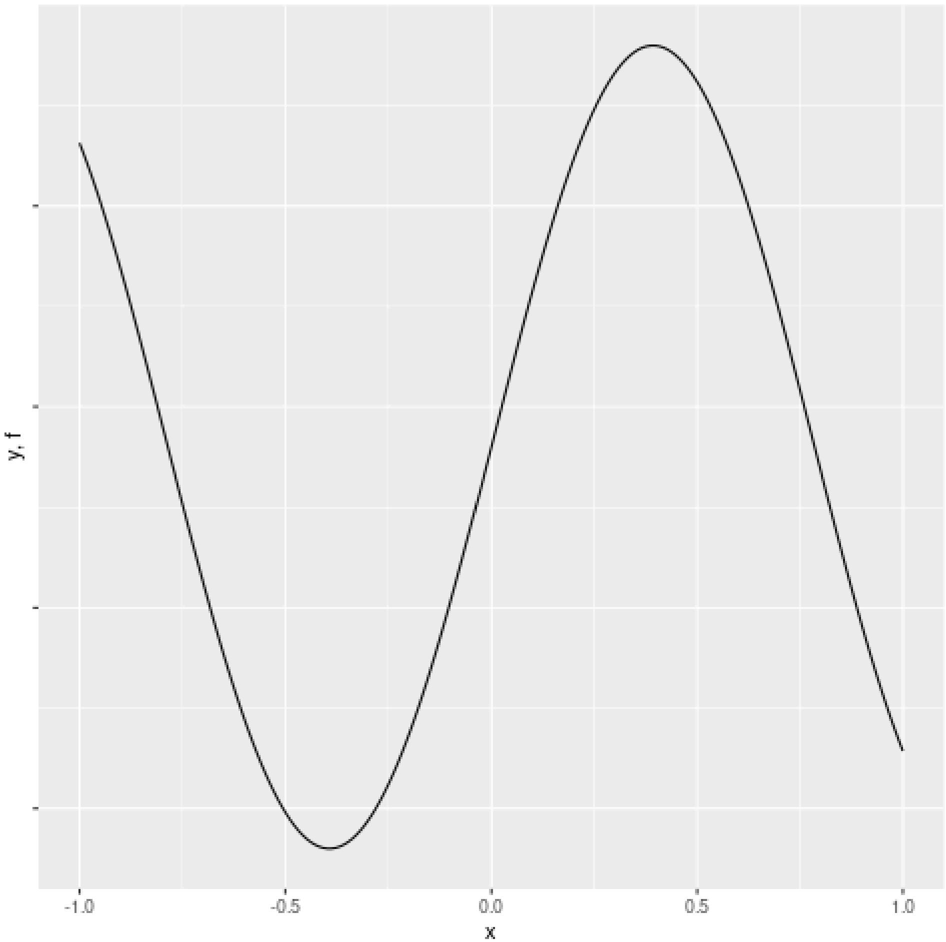
No arquivo `T1_q5.rda`, há um objeto denominado `data_q5`, que é um `data.frame` com duas variáveis: `x` e `y`.

Para cada valor em `x` está associado um valor em `y`, tal que $y = f(x)$, para uma particular função f.

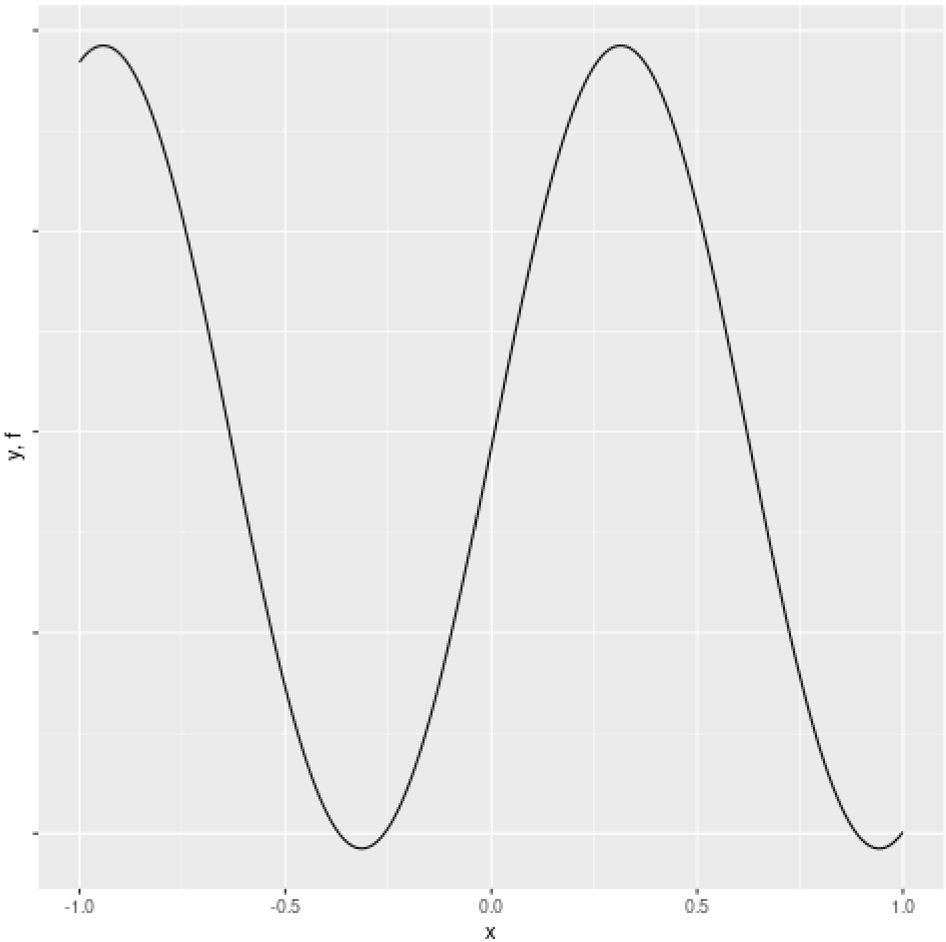
Qual das alternativas abaixo melhor representa o gráfico da função f, no intervalo de ocorrência dos pontos fornecidos na variável `x`?

Escolha uma opção:

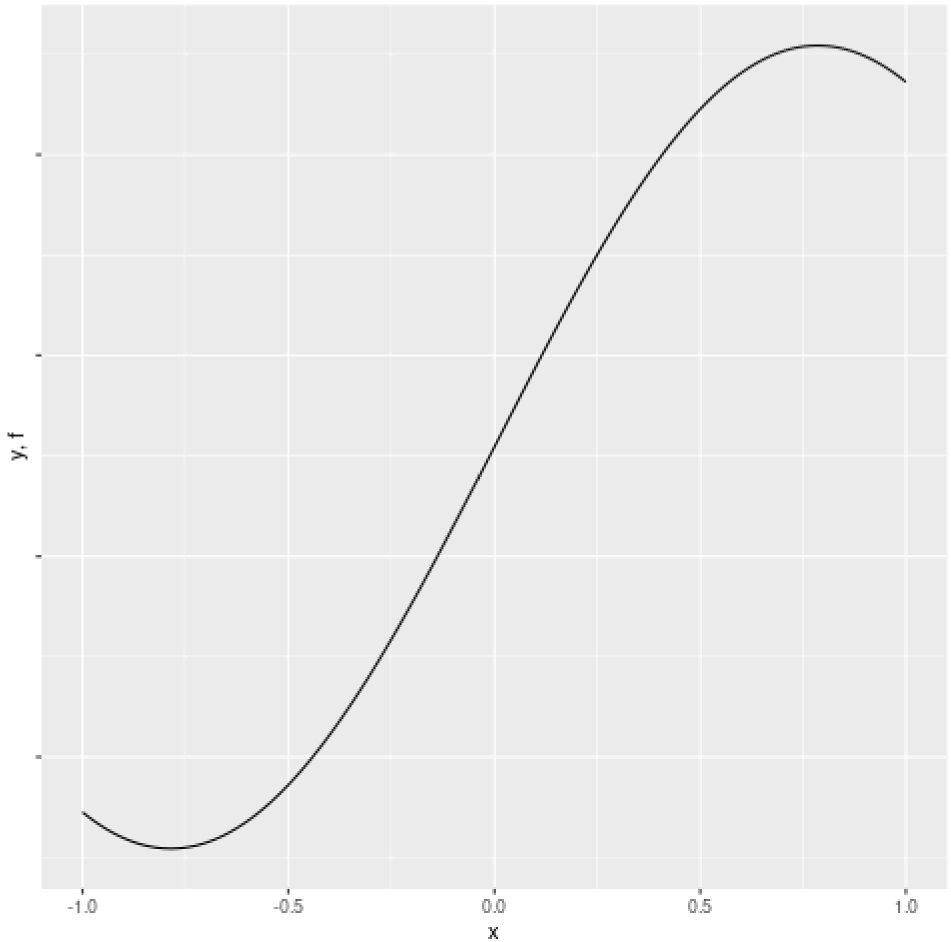
a.



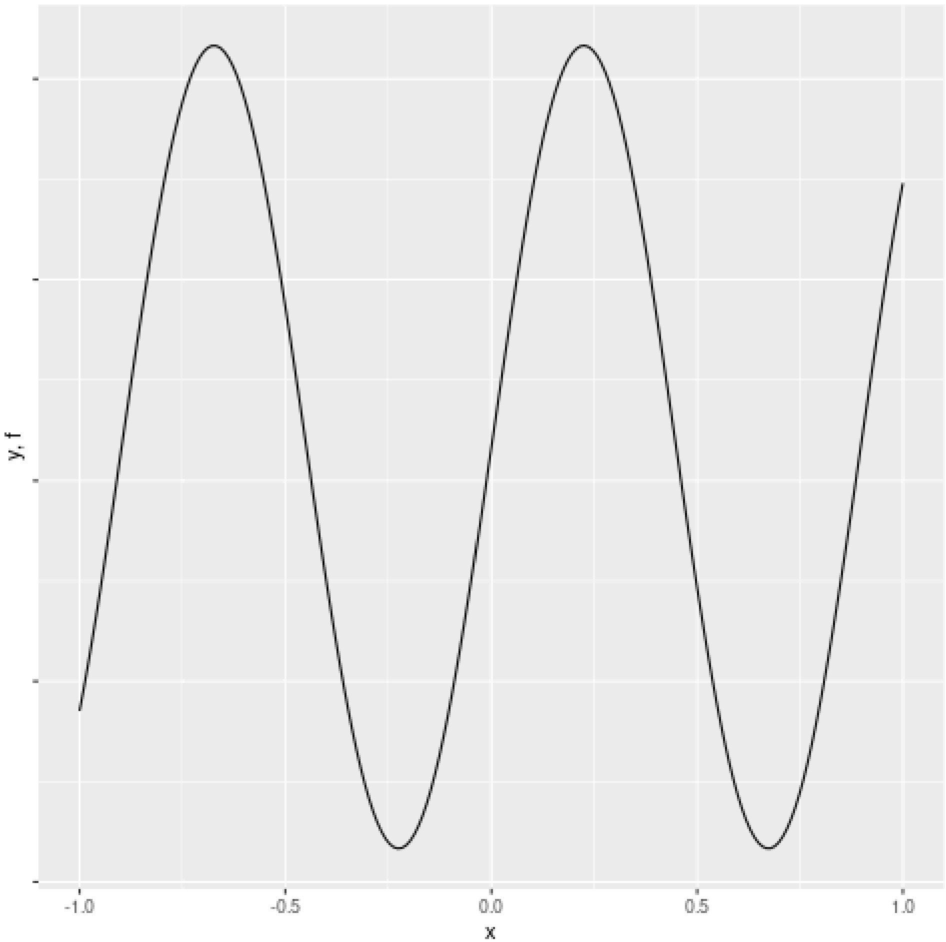
b.



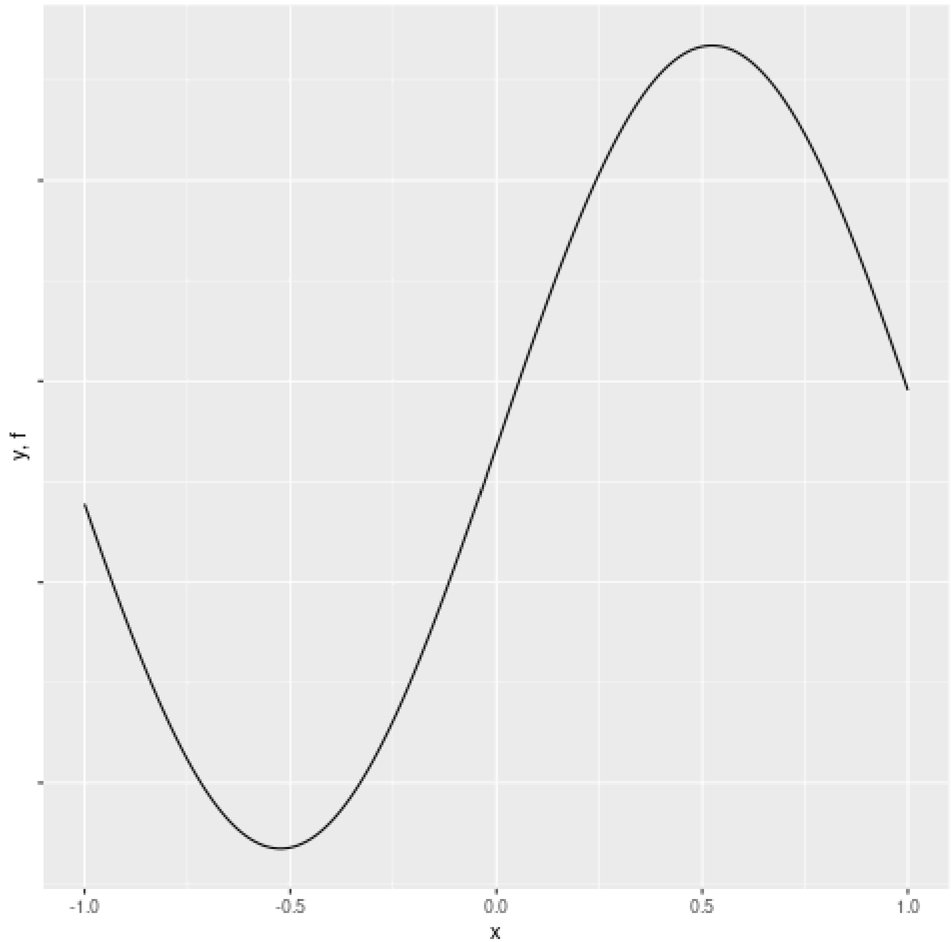
c.



d.



e.



«