

Resolução da lista 6 - Análise de dados - UFPE - 2019.1
Professor: Davi Moreira
Aluna: Ayla Reis de Lima
Endereço do GitHub: https://github.com/AylaReis/AD_resolucao_lista_6

Questões resolvidas: 1-8 e 10.

Exercício 1: Descreva os conceitos abaixo:

a) Variável dependente

A variável dependente, ou o valor de "Y" das equações matemáticas, se refere a uma grandeza em que seu valor irá depender de outra grandeza.

b) Variável independente

A variável independente, ou o valor de "X" das equações matemáticas, se refere a uma grandeza em que seu valor irá não de outra grandeza para existir.

c) Apresente qual a relação existente entre variáveis independentes e dependente

A relação é de que o valor da variável dependente está vulnerável aos valores das variáveis independentes, ou seja, os "X"s influenciará no valor de "Y".

Exercício 2: Em análise de dados, qual o nome dado à equação abaixo?

Modelo de regressão amostral.

Exercício 3: Com suas palavras, apresente uma definição para cada um dos componentes da equação apresentada no exercício 2.

Y_i : é a variável dependente de "X"

$\hat{\alpha}$: é o coeficiente linear e representa o valor de "Y" quando o "X" é zero.

$\hat{\beta}$: é o coeficiente angular da reta de regressão, ou seja, a inclinação da reta.

X_i : é a variável independente.

\hat{u}_i : é o erro amostral, ou a distância do ponto de uma amostra até a reta de regressão.

Exercício 4: Apresente o componente sistemático da equação apresentada no exercício 2. Descreva por quê é sistemático.

O componente sistemático é descrito pela equação $\hat{Y}_i = \alpha + \beta X_i$, sendo ele o " \hat{Y}_i " previsto para " Y_i " sem o erro amostral. Ademais, ele é sistemático porque é composta apenas pelas variáveis explicativas.

Exercício 5: Apresente o componente estocástico da equação apresentada no exercício 2. Descreva por quê é estocástico.

O componente sistemático mais o componente estocástico estão presentes na seguinte equação:

$\hat{u}_i = Y_i - \hat{Y}_i$. Ademais, ele é estocástico pois mostra o erro ou a diferença do valor observado do esperado.

Exercício 6: Descreva a diferença entre Y_i e \hat{Y}_i ? Qual a relação desses dois componentes com o \hat{u}_i ?

O " Y_i " diz respeito ao valor que foi observado na regressão amostral; já o " \hat{Y}_i ", o valor que foi esperado ter na reta de regressão. Ou seja, a diferença entre os dois (\hat{u}_i) será o erro amostral.

Exercício 7: Com suas palavras, apresente o que é o modelo OLS e seu principal uso na análise de dados.

O modelo de regressão bivariada ou o "OLS" serve para minimizar ao máximo a distância entre a reta de regressão e seus pontos, ou seja, é utilizado para minimizar o erro (\hat{u}_i).

Exercício 8: Com base no Google's R Style Guide (<https://google.github.io/styleguide/Rguide.xml#indentation>), apresente exemplos de boas práticas para os seguintes tópicos:

a) File names;

Os nomes dos arquivos devem ser apresentar o ".R" no final e ser auto-didáticos.

Exemplo: lista_6_Ayla_Reis.R CERTO / lista.R ERRADO

b) Identifiers;

Para nomear variáveis, usar pontos finais ou letras maiúsculas para separar palavras

Exemplo: cidades.interior ou cidadesInterior BOM / cidades_interior RUIM

c) Indentation;

```
Exemplo: if ("condição") {  
        código código código  
}
```

O código dentro do if deve ser separado por dois espaços do original.

d) Spacing;

Os espaços devem ser colocados entre operadores binários, depois de uma vírgula, antes do parêntese à esquerda. Ademais, o espaço extra é bom ser colocado para alinhar sinais ou setas.

Exemplos: 1 + 2 CERTO / 1+2 ERRADO

y <- c(sample(1:10, 100, replace = T)) CERTO

instalado ("condição") CERTO

e) Assignment;

Na atribuição de objeto, usar o "<-" e não o "=".

Exemplo: Y <- 3 CERTO / Y = 3 ERRADO

f) Commenting Guidelines;

Para comentários nas linhas deve-se usar o "#" mais um espaço.

Exemplo: # Execício 1 CERTO / Exercício 1# ERRADO

g) Function Definitions and Calls;

As definições de função devem primeiro listar os argumentos sem valores padrão, seguidos daqueles com valores padrão. As quebras entre linhas só são permitidas entre atribuições.

Exemplo: `qplot(movies$votes,
 movies$reviews,
 data = movies,
 geom = c("point", "smooth"),
 method = "lm",
 alpha = I(1 / 5),
 se = F)`

h) Function Documentation;

As funções devem conter uma seção de comentários imediatamente abaixo da linha de definição da função.

Exemplo: `X <- 1`

`y <- 2`

`z <- x + y`

`# Os valores de x e y foram somados para ser criado o objeto z, para que assim`

`# contribua com a pesquisa.`

Exercício 9: Leia o Capítulo 7 do livro R para Ciência de Dados (<http://r4ds.had.co.nz/exploratory-data-analysis.html>) e entregue script no R que

reproduza os exemplos apresentados no capítulo. Comente seu código indicando o que está para ser realizado em cada etapa do seu script.

Exercício 10: Com os dados disponibilizados na plataforma (vote_growth_usa.RData), reproduza os resultados do livro Kellstedt, P. M., & Whitten, G. D. (2013) utilizando o código apresentado nos slides da aula.

```
# Instalando pacote, definindo diretório e carregando base de dados primária

if(require("tidyverse") == F) install.packages("tidyverse") ; require("tidyverse")

setwd("/Users/reis/Desktop/AYLA/UFPE/analise_de_dados_pos_graduacao_davi_moreira/lista_6/AD_resolucao_lista_6")

getwd()

load("vote_growth_usa.RData")

votegrowth <- bd

rm (bd)

summary(votegrowth)

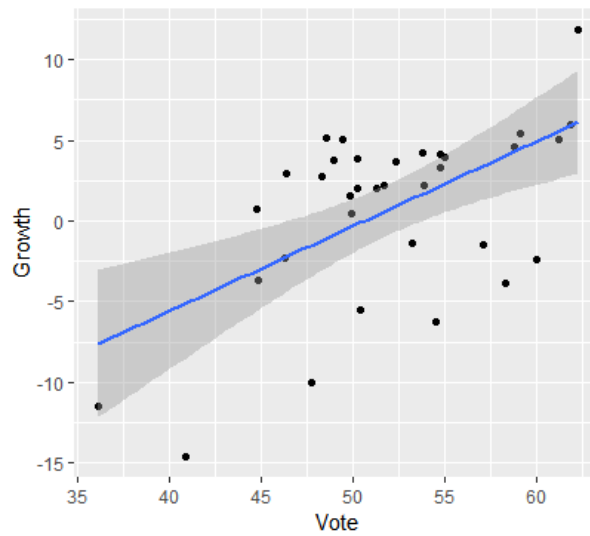
# Modelo de regressão bivariada

regl <- lm(Vote ~ Growth, data = votegrowth)

summary(regl)

ggplot(votegrowth,

       aes(Vote, Growth)) + geom_point() + geom_smooth(method = "lm")
```



Exercício 11: Com os dados e as variáveis do exercício 10, realize uma análise de regressão considerando apenas o período de 1876 a 1932. Apresente os resultados e os compare quanto ao modelo completo (exercício 10) em relação a: