

PERGUNTAS E RESPOSTAS – MBA EM DATA SCIENCE E ANALYTICS

Disciplina: Supervised Machine Learning: Análise de Regressão Simples e Múltipla I

Data: 27/07/2021

Alex De Lima Bassi

Poderia colocar ao invés de colocar o valor da média no gráfico da linha 90. colocar uma função de média?

Olá, Alex. É possível, mas seria necessário ter uma fundamentação teórica neste sentido, considerando que não faz parte do algoritmo padrão da regressão linear.

Alex De Lima Bassi

porque no código da linha 71 os valores mínimos ficaram negativos?

Os valores desses resíduos ficaram abaixo do valor médio, como você pode verificar no gráfico, motivo pelo qual os seus valores ficaram negativos.

Carlos Munhoz Junior

pacote jtools foi instalado e carregado mas com o comando abaixo ocorre o seguinte erro: `export_summs(modelo_tempodist, scale = F, digits = 4)`. Erro apresentado: `> export_summs(modelo_tempodist, scale`

Tente: `install.packages("Jtools")` e depois `library(Jtools)`. Após repita o código. Provavelmente dará certo. Caso o problema persista, tente: `update.packages()`. Caso não consiga encaminhe um email para monitoria tentar ajudar.

Carlos Eduardo Almeida Gomes

Prof boa noite, no caso de São paulo, a distancia ficaria com porcentual pequeno? uma vez que medimos a distancia através de horas

É necessário rodar o modelo com inclusão de São Paulo para ter certeza dessa afirmação.

Maíra Vitoriano Rodrigues de Freitas

Professor, quais são as consequências de impor o alfa igual a zero no modelo?

A constante linear (alfa) é fornecida pela equação, portanto não há como impor seu valor de maneira arbitrária.

Teresa Arlinda De Souza Campos

distância e modo são mais relevantes para determinar o tempo?

Provavelmente, sim. Mas é necessário rodar o modelo com essas variáveis para ter certeza.

Gabriel Cinati Sundfeld

Mesmo comandando as linhas e instalando os pacotes, erro: Error in `summ(modelo_tempodist, confint = T, digits = 4, ci.width = 0.95)` : objeto 'modelo_tempodist' não encontrado

Tente verificar se o banco de dados está escrito da forma correta.

Alessandro Nogueira Corrêa

```
> summ(modelo_tempodist, confint = T, digits = 4, ci.width = .95) MODEL INFO: Observations: 10 Dependent Variable: tempo Type: OLS linear regression MODEL FIT: F(1,8) = 36.3031, p = 0.0003 R2 = 0.81
```

Pelo que verifiquei esse é o padrão da função summ, conforme a documentação do pacote Jtools: <https://jtools.jacob-long.com/reference/summ.lm.html>

Genisson Alves da Fonseca

```
MODEL INFO: Observations: 10 Dependent Variable: tempo Type: OLS linear regression MODEL FIT: F(1,8) = 36.3031, p = 0.0003 R2 = 0.8194 Adj. R2 = 0.7969 Standard errors: OLS
```

Pelo que verifiquei esse é o padrão da função summ, conforme a documentação do pacote Jtools: <https://jtools.jacob-long.com/reference/summ.lm.html>

Tiago Rosas Amaral

Boa noite, documentando o erro da linha 74, após instalar e carregar os pacotes Jtools e Rcpp Espero que tenha conseguido rodar o script. Tente: `install.packages("Rcpp")` e depois `library(Rcpp)` e `install.packages("Jtools")` e depois `library(Jtools)`. Após repita o código. Provavelmente dará certo. Caso o problema persista, tente: `update.packages()`.

Murilo Marin Pechoto

```
> install.packages("rccp") WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding: https://cra
```

Algun erro com pacote Rcpp. Tente: `install.packages("Rcpp")` e depois `library(Rcpp)`. Após repita o código. Provavelmente dará certo. Caso o problema persista, tente: `update.packages()`.

Alex Jornada Queiroz

Na linha 74, o erro foi esse: Error in tableExpand_cpp(cells, cols.width, justify, sep.cols, style) : pacote 'Rcpp_precious_remove' não oferece a função 'Rcpp'

Algun erro com pacote Rcpp. Tente: `install.packages("Rcpp")` e depois `library(Rcpp)`. Após repita o código. Provavelmente dará certo. Caso o problema persista, tente: `update.packages()`.

Miguel Angelo Henley Gomes da Silva Filho

Por que devemos minimizar o quadrado das diferenças? Nao poderia ser minimizar o módulo da diferença?

O professor respondeu na aula. Seria possível, mas seria necessário criar um algoritmo novo e testar o modelo para saber se seria coerente e responderia sua questão de pesquisa de forma satisfatória.

Felipe Francisco Nusda

porque a Análise de Regressão: Cálculo do R² parece um torus?

Olá Felipe. Poderia ser mais específico. Não compreendemos bem sua pergunta.

Ana Jessica Oliveira Batista

Por que o intercept value ou "alpha", é chamado de "bias" em alguns dos outros modelos?

Olá Ana. Trata-se de mera questão de nomenclatura. De toda forma, tente entender o modelo em si para compreender o método independente da nomenclatura adotada.

Sabrina Portela Costa Ferreira

```
> install.packages('rcpp') WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding: https://cra
```

Tente `install.packages("Rcpp")`. A letra "R" do pacote deve ser maiúscula.

Raphael Fidelis Valadares

Não é este o problema. independentemente disso de instalar e chamar o pacote, o erro persiste.

Olá Fabrício. Nestes casos recomendo reiniciar o Rstudio e tentar rodar o código novamente, verificando se todos os pacotes estão instalados. Pode ser algo específico da sua máquina. Verifique na aba direita e inferior, na aba packages, se todos os pacotes estão marcados e tente rodar o código novamente.

Fabricio Colombo Tezini

Meu R travou no processamento. Acho que foi na linha 74 ou algo assim...o que faço?

Olá Fabrício. Nestes casos recomendo reiniciar o Rstudio e tentar rodar o código novamente, verificando se todos os pacotes estão instalados. Pode ser algo específico da sua máquina.

Kauê Bonato De Araujo

Professor, entendo que pontos outliers podem enviesar a linha de ajuste. Existem técnicas para lidar com esses pontos (que não seja removê-los da base).

Olá Kauê. Existem técnicas para retirar outliers. Uma delas é criar gráficos de box-plot para identificar qual observação é um outlier. Neste site você pode ter alguns exemplos: <https://statsandr.com/blog/outliers-detection-in-r/>

Marcelo Patto

Ymedio é a mesma coisa de Ychapeu?

Ychapéu representa o valor previsto da variável dependente que será obtido por meio do modelo estimado para cada observação. Ymédio representa o fenômeno em estudo (variável dependente quantitativa) em média.

Alessandro Nogueira Corrêa

Qual a versão recomendada do R?

Olá Alessandro. Sempre são lançadas várias versões atualizadas. Neste sentido, é sempre bom manter a versão mais nova do programa no site: <https://www.r-project.org/>. A versão indicada inicialmente pode já ter sido atualizada.

Alfredo Neto

Amanda e monitores. Uma sugestão é instalar o JTools, que pode ser entendido como um complemento do R. E depois tentar reinstalar os pacotes
Obrigado, Alfredo. Podemos tentar para ver se funciona.

Guilherme Piva Magalhaes Da Rocha

Existe algum cálculo métrico para verificar medidas de correlação não lineares? Ex : cálculo de medida exponencial

Se uma relação entre duas variáveis não é linear, a taxa de aumento ou diminuição pode mudar como uma mudança de variáveis, causando um "padrão curvo" nos dados em algumas situações. Esta tendência curva pode ser melhor modelada por uma função não linear, como uma função quadrática ou cúbica ou ser transformada para tornar mais linear. O Gráfico em U pode mostrar uma relação forte entre duas variáveis. Entretanto, como a relação não é linear, o coeficiente de correlação de Pearson é apenas +0,244. Essa relação ilustra o porquê é importante traçar os dados para explorar quaisquer relações que possam existir.

Adriana Melges Quintanilha Weingart

Neste exemplo da "Importância relativa da distância interferindo no tempo", aplica-se "ceteris paribus", a distância afeta diretamente o tempo - aumenta distância, aumenta tempo. É isso? Isso, o parâmetro estimado alfa (coeficiente linear) mostra o ponto da reta de regressão em que $X = 0$, o parâmetro estimado Beta (coeficiente angular) representa a inclinação da reta, ou seja, o incremento (ou decréscimo) de Y para cada unidade adicional de X, em média. Quando o Beta é positivo significa que o aumento do X causa acréscimo no Y.

Vitor Hugo De Marrafon

Professor, quando calculamos uma regressão linear simples calculamos o termo beta (coeficiente linear), esse coeficiente pode ser interpretado como uma tendência linear presente em minha amostra?

Enquanto o parâmetro estimado alfa (coeficiente linear) mostra o ponto da reta de regressão em que $X = 0$, o parâmetro estimado Beta (coeficiente angular) representa a inclinação da reta, ou seja, o incremento (ou decréscimo) de Y para cada unidade adicional de X, em média.

Carla Porto Veiga

professor, como se chega manualmente nos valores de alfa e beta?

Os **parâmetros α e β** são estimados através do Método dos Mínimos Quadrados Ordinários. O objetivo deste método é obter uma reta que minimiza as distâncias entre os valores estimados e os valores observados. O **parâmetro α** representa o intercepto da reta, onde ela cruza o eixo Y, ou seja, o valor de $E(Y)$ para o qual $X = 0$. Este parâmetro só será interpretado na prática quando existir sentido real na variável explicativa assumindo valor zero. Já o **parâmetro β** representa, neste caso, o efeito da variável explicativa sobre a variável resposta. Neste link você pode verificar como esses parâmetros são calculados: <http://dpi.ufv.br/~peternelli/inf162.www.16032004/materiais/CAPITULO9.pdf>

Luiz Monclar

Não entendi qual era o problema de usar ponderação arbitrária para variáveis qualitativas. Pq seria errado fazer isso?

Este tipo de variável não apresenta média e variância sem que seja feita uma incoerente ponderação arbitrária, visto não haver uma equidistância aceita entre as diferentes categorias de uma mesma variável. Isso já foi explicado em algumas aulas.

Cibelle Farias Martins

À partir da linha 89, não consigo rodar mais nada.

Olá, Cibelle. Tente reiniciar o R e reinstalar os pacotes. Rode o código todo de novo, pois ele foi testado e não apresentou nenhum erro. Às vezes pode ser um erro interno que pode ser corrigido com simples reinício do processo.

Luiz Rodriguez Fantini 005.374.61981