# Turma 22

Gabriel Rulka Tardoski

Israel Matias

Jordhan Emmanuel Marciano da Silva

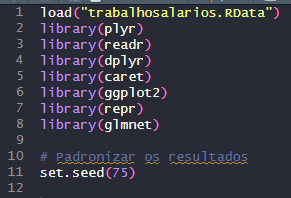
Zaira Mendonça Amorim

# 

# Rotina Genérica

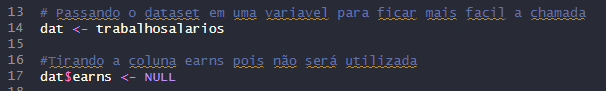
Para os 3 modelos, existem variáveis e dados que são comuns entre eles, por isso, esses dados foram criados primeiro, sendo eles:

Dataset e bibliotecas utilizadas:

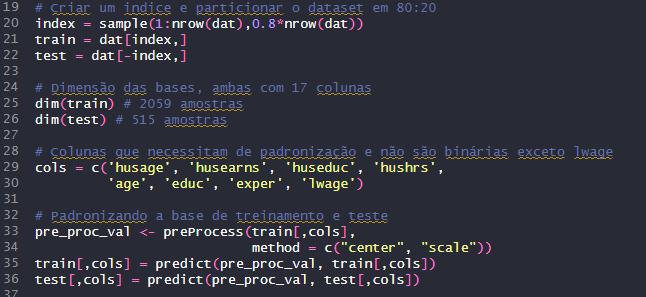


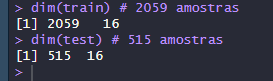
No nosso trabalho, foi escolhido a seed 75 para manter consistência dos resultados.

Também tiramos a coluna ‘earns’ visto que ela não será necessária

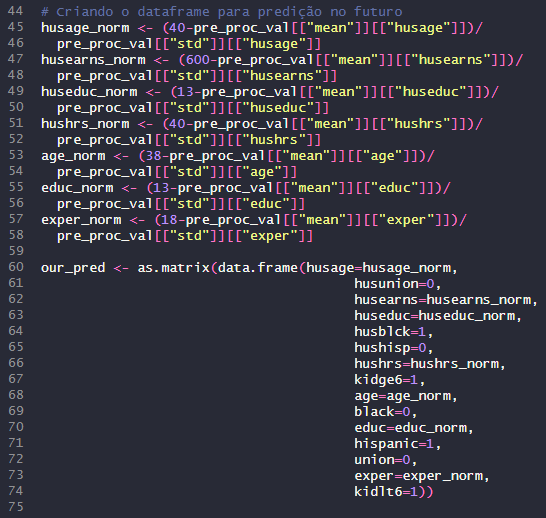


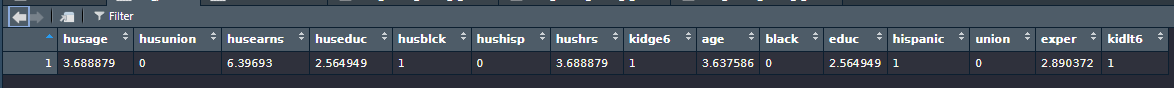
Após isso, foi criado uma variável para guardar o dataset e criado os índices de treinamento além da normalização dos valores não binários:

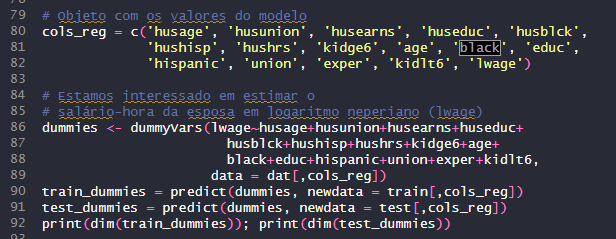




Normalizando os dados e criando uma matriz que será usada para predição (com base nos valores do trabalho e aplicando logaritmo neperiano:

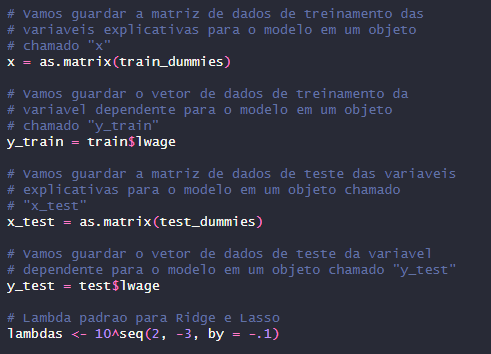




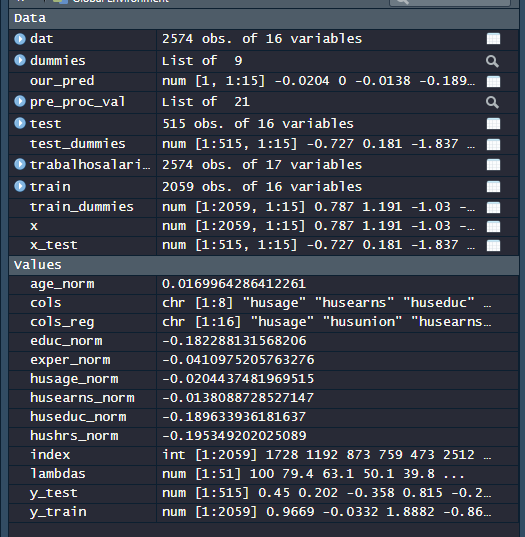




Depois disso, é guardado os valores em variáveis para usarmos posteriormente nos modelos, detalhe que a variável **‘*lambdas’*** só é usada em **Ridge e Lasso.**

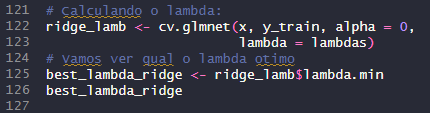
****

**Preview de como ficou os nossos dados genéricos:**

****

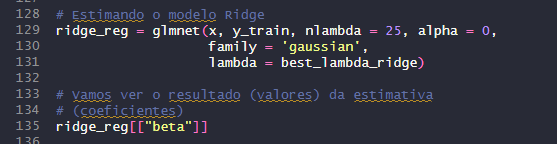
# Regressão Ridge

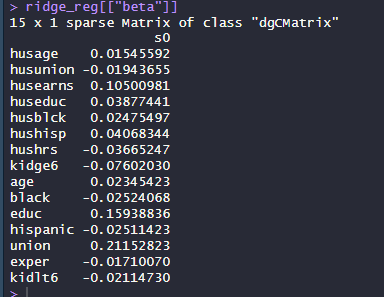
Calculando o lambda



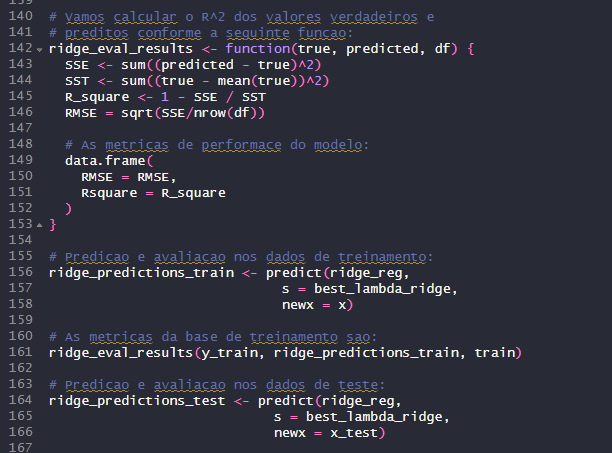


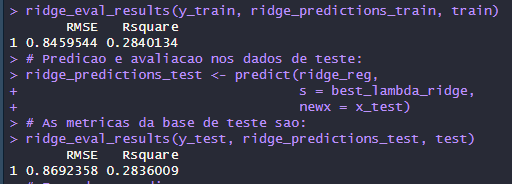
Estimativa de coeficientes:



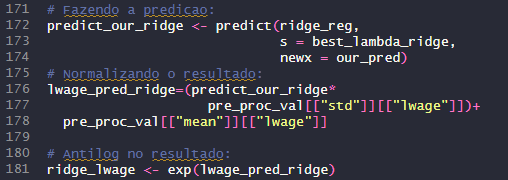


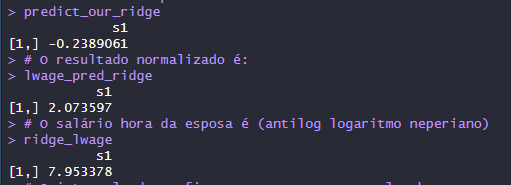
Calculando o RMSE e R² do modelo em Ridge:





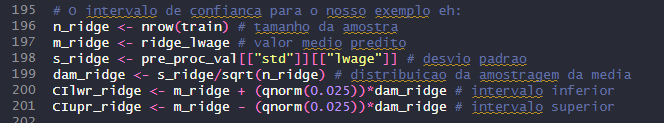
Fazendo a predição em Ridge:

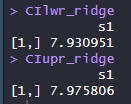




Valor original predito foi -0.238, após a normalização, voltando a ficar em logaritmo neperiano, foi de $3.24, por fim, após a aplicação do antilog, se tornou $7.95.

O intervalo de confiança para nosso modelo ficou em 7.93 o inferior e 7.98 o superior.

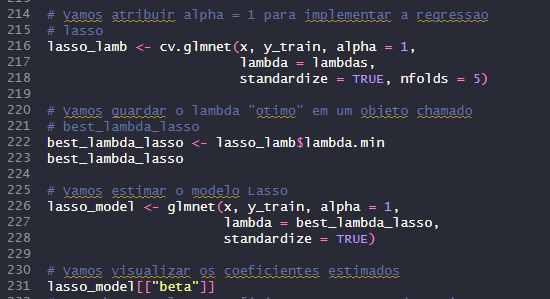


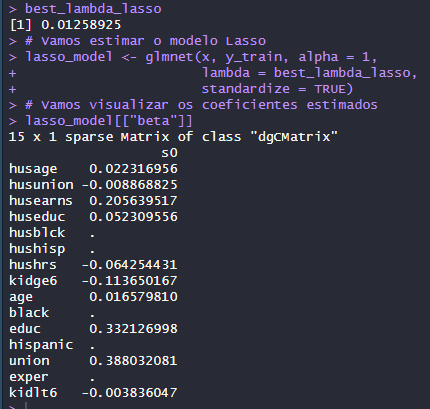


O que significa que o salário pode variar entre 7.93 e 7.97.

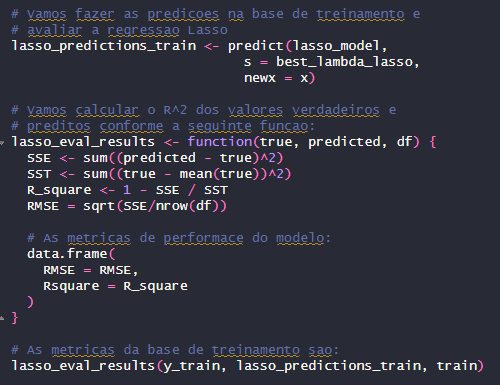
# Regressão Lasso

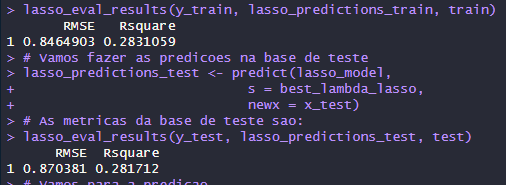
Como feito na Regressão Ridge, não será necessário recriar variáveis para dataset ou para predição, pois são os mesmos valores, além do próprio lambda, porém será feito um lambda otimizado para essa regressão.





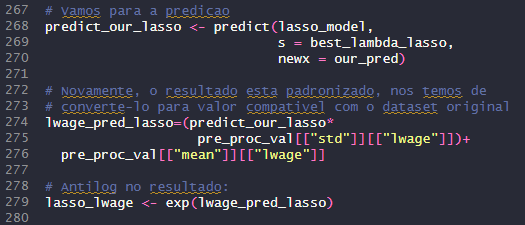
Agora será realizado a predição do modelo Lasso e encontrar o RMSE e R²:

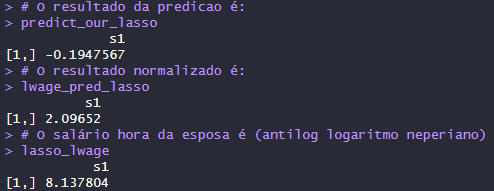




Em comparação a Regressão Ridge, o RMSE e R² também apresentou resultados baixos, explicando que o modelo não está tão bem otimizado.

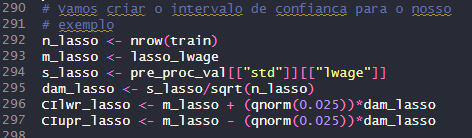
Vamos realizar a predição com os dados montados:

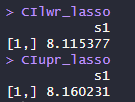




Valor original predito foi -0.19, após a normalização, voltando a ficar em logaritmo neperiano, foi de $2.10, por fim, após a aplicação do antilog, se tornou $8.14.

Agora calculando o intervalo de confiança:

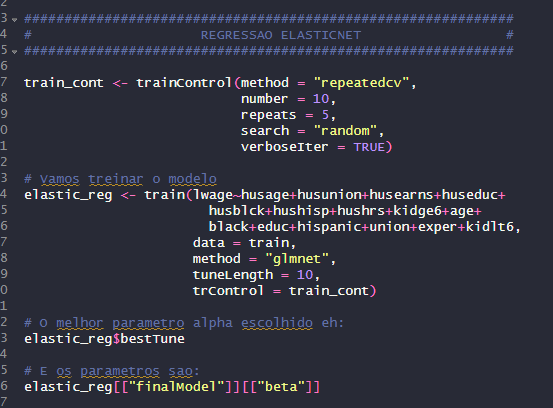




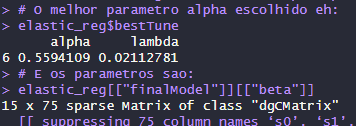
O que significa que o salário pode variar entre 8.12 e 8.16 na Regressão Lasso.

# Regressão ElasticNet

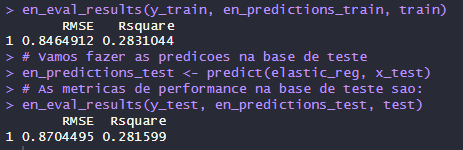
Novamente, usando os mesmos dados genéricos, iremos direto para o treinamento do modelo, não há parâmetro alpha pois a própria regressão ElasticNet irá encontrá-lo.



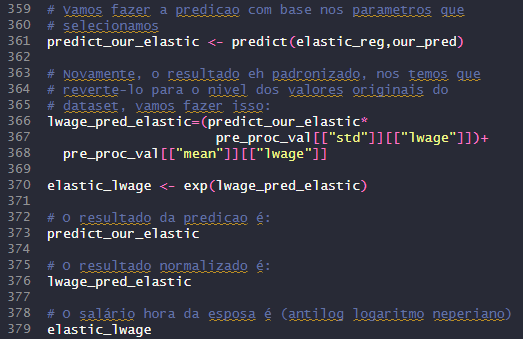
Melhor alpha escolhido:



Fazendo a predição e adquirindo os valores de RMSE e R²:

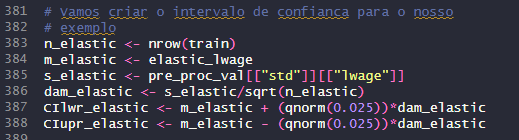


Novamente não saíram tão bons.



Valor original predito foi -0.19, após a normalização, voltando a ficar em logaritmo neperiano, foi de $2.10, por fim, após a aplicação do antilog, se tornou $8.13, bem semelhante ao Lasso.

Para os intervalos de confiança:





Valores entre 8.11 e 8.15.