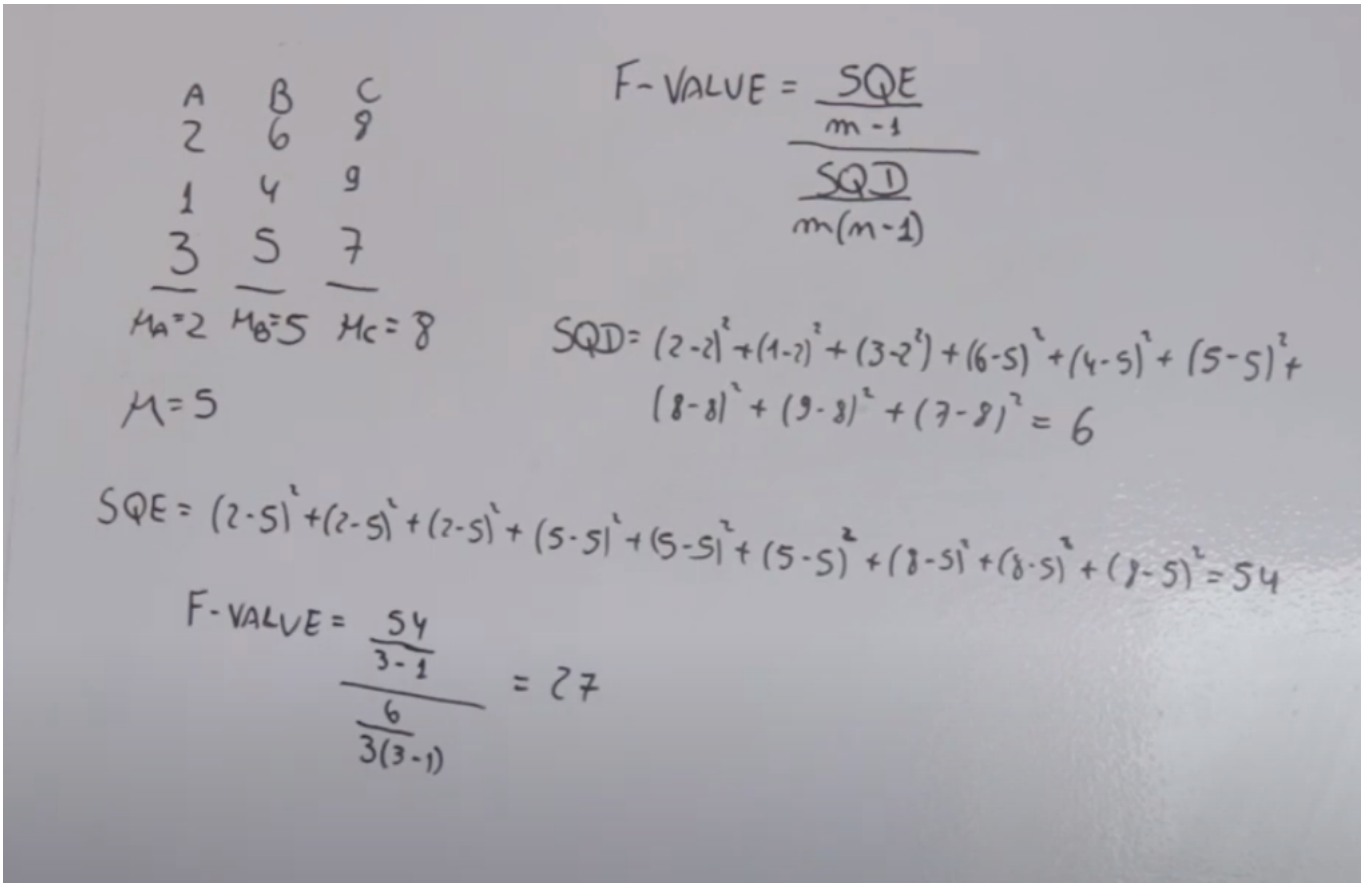
F classif

É o cálculo do F-value da estatística. Basicamente ele verifica o quão distante da média de um grupo está do outro, caso seja distante o F-value é maior, e muito próximo será menor.

Exemplo:

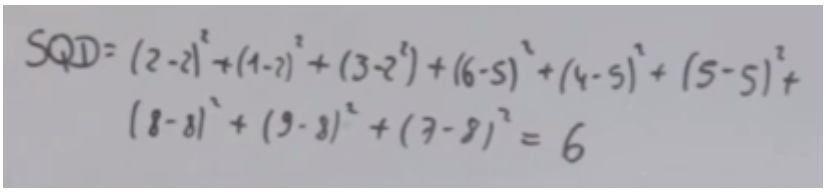


Primeiro é calculado a média de cada grupo A, B e C:

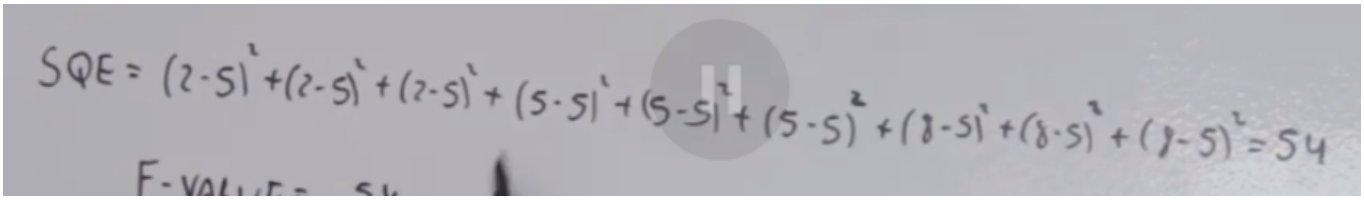
A = 2
B = 5
C = 8

Média geral: 5

SQD = Somatório de cada grupo, sendo o valor da amostra de um grupo X - a média do grupo X elevado ao quadrado

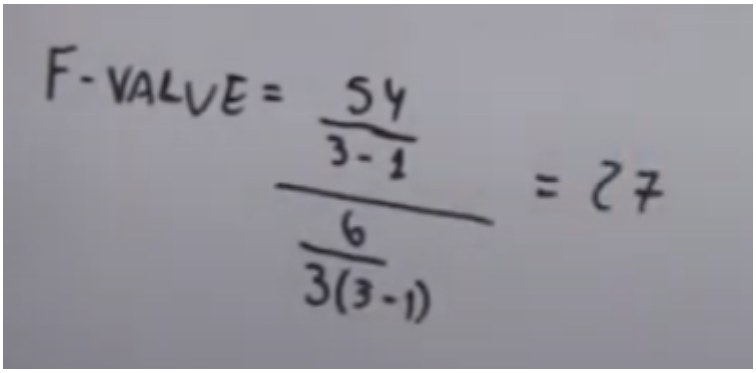


SQD = Vai utilizar a média geral e a média do grupo repetindo por amostras de cada grupo:



Vai se repetindo, pois há 3 valores de cada grupo

Por fim é só calcular o F-value:



Sendo M o número de amostras = 3
e N o número de linhas = 3

Com isso conseguimos utilizar o conceito do F-value para datasets na seleção de features para problemas de classificação. Basta dividir uma coluna do dataframe em N grupos, sendo grupos a quantidade de classes, 0,1 ou N classes. Assim podemos extrair a média de cada grupo e verificar se elas são diferentes, caso sejam, a variável contribui bastante para a classificação do dado, caso sejam próximas, ou seja, um valor baixo de F-value, não é uma variável boa para a classificação. No fim queremos ver se a média dos dados para cada classe é diferente, para haver uma melhor divisão na hora de classificar os dados.