R2(ValoresPreditos1, test.data$Fertility)

0.4090483

R2(ValoresPreditos1, test.data$Agriculture)

0.0482434

R2(ValoresPreditos1, test.data$Examination)

0.3689375

R2(ValoresPreditos1, test.data$Education)

0.6637144

R2(ValoresPreditos1, test.data$Catholic)

0.2600994

R2(ValoresPreditos1, test.data$Infant.Mortality)

0.0531055

Escolhendo os 3 maiores R²

modelo2 =lm(Fertility ~Education, Examination, Catholic, data = train.data)

ValoresPreditos2 = predict(modelo2,newdata=data.frame(test.data))

Glane Veveyse Aigle Avenches Oron Payerne

75.09268 76.74319 71.79166 71.79166 80.86946 75.09268

Rolle Vevey Yverdon Monthey Sion La Chauxdfnd

73.44217 66.01488 75.09268 79.21895 70.96641 72.61692

Le Locle Neuchatel ValdeTravers

70.96641 55.28658 75.91794

ModeloAjustado2 = R2(ValoresPreditos2, test.data$Fertility)

0.1838678

ErroAbsoluto2 = MAE(ValoresPreditos2, test.data$Fertility)

7.497056

ErroMedioQuadratico2 = RMSE(ValoresPreditos2, test.data$Fertility)

8.712885

O R² é menor e os erros são maiores, logo o primeiro modelo é melhor

modelo3 =lm(Fertility ~Education, data = train.data)

ValoresPreditos3 = predict(modelo3,newdata=data.frame(test.data))

Glane Veveyse Aigle Avenches Oron Payerne

72.30869 74.15773 68.61062 68.61062 78.78032 72.30869

Rolle Vevey Yverdon Monthey Sion La Chauxdfnd

70.45966 62.13900 72.30869 76.93128 67.68611 69.53514

Le Locle Neuchatel ValdeTravers

67.68611 50.12027 73.23321

ModeloAjustado3 = R2(ValoresPreditos3, test.data$Fertility)

0.1838678

ErroAbsoluto3 = MAE(ValoresPreditos3, test.data$Fertility)

7.30381

ErroMedioQuadratico3 = RMSE(ValoresPreditos3, test.data$Fertility)

8.991637

Como o R² é igual e os erros não alteram muito do 2 para o 3, dependendo da utilização desses modelos, poderia se utilizar o 2 a fins de simplificação.

Contudo, o modelo 1 é o melhor de todos nos 3 quesitos