Measure the Error of a Fitted Line

eCornell

7/26/2021

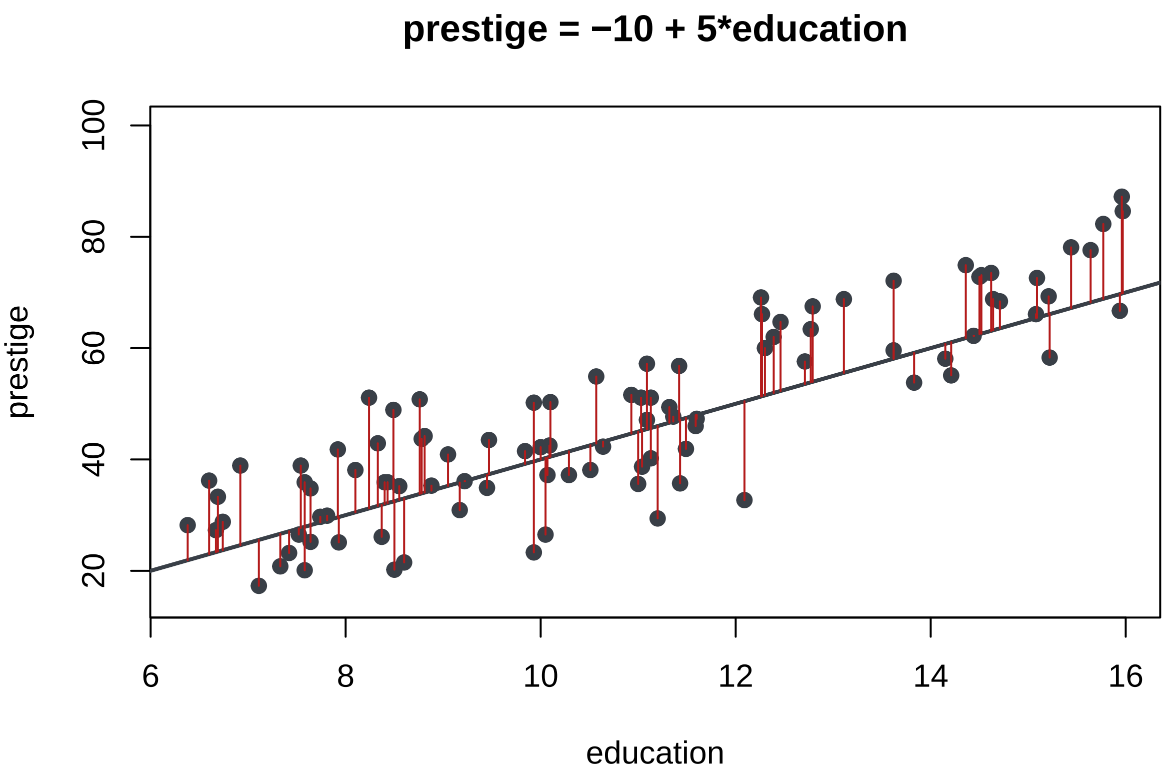
Use this R Markdown file to understand how to draw a fit line through data and use error calculations to measure how well the line fits your data.

Step 1: Load the data and define colors.

Step 2: Plot the data, regression line, and errors.

Create a plot of prestige vs. education with the plot() function. Plot a line through it with the abline() function. Finally, use a for loop to plot the error lines between each point and its predicted value to help you visualize the difference between predicted and observed values.

```
# Create the plot:
plot(prestige ~ education, data = Prestige, pch = 19, col = ecBlack,
     main = 'prestige = -10 + 5*education', ylim = c(15, 100))
# Plot the fit line. the argument a = -10 indicates the intercept is -10, b = 5 indicates the slope is
abline(a = -10, b = 5, col = ecBlack, lwd = 2)
# for loop to draw vertical error lines:
for (i in 1:98){
  lines(rep(Prestige$education[i], 2),
        c(Prestige$prestige[i], -10+5*Prestige$education[i]),
        type = 'l', col = crimson, lwd = 1)
}
```



Step 3: Assess how a line fits the data with MSE.

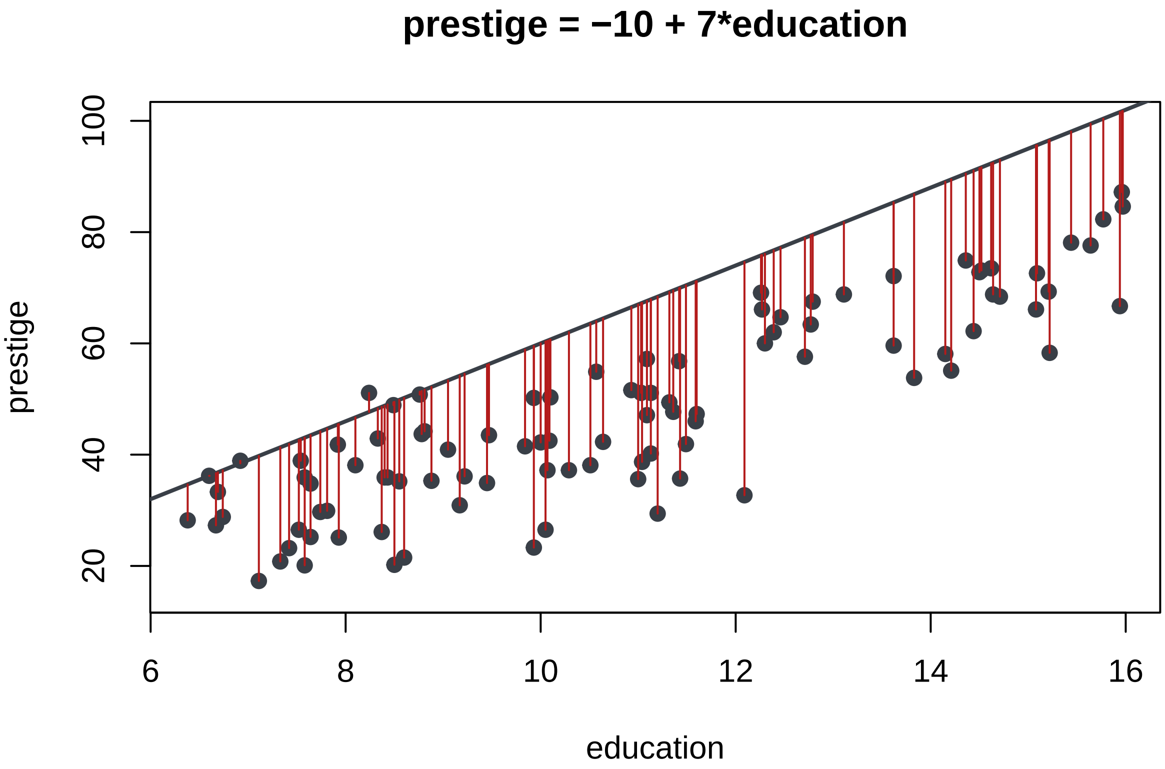
The following code indicates one way to calculate the Mean Square of Errors (MSE) in R. Use MSE to assess whether the line is a good fit to the data. MSE is a measurement that takes the average of the sum of the squared value of each error line.

```
# Add a column to the Prestige data set called fit1 that contains the predicted value of each point bas
Prestige$fit1 = -10 + 5*Prestige$education
# Add a column to the Prestige data set called error1 that contains the difference between the observed
Prestige$error1 = Prestige$prestige - Prestige$fit1
# Calculate MSE by finding the average value of squared error:
MSE1 = mean(Prestige$error1^2)
```

Step 4: Visually compare line fits to data.

Compare the regression line prestige = -10 + 5*education with the fit of a different line: prestige = -10 + 7*education

```
# Plot the line prestige = -10 + 7*education over the prestige vs. education scatterplot
plot(prestige ~ education, data = Prestige, pch = 19, col = ecBlack,
     main = 'prestige = -10 + 7*education', ylim = c(15, 100))
abline(a = -10, b = 7, col = ecBlack, lwd = 2)
# Add vertical error lines to the plot
for (i in 1:98){
  lines(rep(Prestige$education[i], 2),
        c(Prestige$prestige[i], -10+7*Prestige$education[i]),
        type = 'l', col = crimson, lwd = 1)
}
```



Step 5: Use MSE to compare line fits to data.

Use MSE to quantify how bad this fit is, relative to the other line. Higher MSE means higher error and, therefore, a worse fit.

```
# Find the MSE for this line:
Prestige$fit2 = -10 + 7*Prestige$education
Prestige$error2 = Prestige$prestige - Prestige$fit2
MSE2 = mean(Prestige$error2^2)
```