Topic: Coefficient of determination and RMSE

Question: If the correlation coefficient of a data set is $r = 0.92$, what percent of the variation in y can be explained by the variation in x-values?

Answer choices:

A $92\,\%$

B $84.64\,\%$

C $8\,\%$

D $0.64\,\%$

Solution: B

The percent of the variation in y that can be explained by the variation in x -values is the coefficient of determination, or the r^2 value. To find this, we square the r-value.

$$r^2 = 0.92^2$$

$$r^2 = 0.8464$$

As a percentage, this is $84.64\,\%$.

Topic: Coefficient of determination and RMSE

Question: When you calculate the RMSE, a smaller RMSE means that ...

Answer choices:

A ... the model is probably a good fit.

B ... the model is probably a bad fit.

C ... the r^2 value will decrease.

D None of these

Solution: A

The root mean square error is the standard deviation of the residuals. If the residuals are closer to the line of best fit, then the standard deviation will be smaller.

This also means that the line of best fit is better correlated, so the r^2 value will increase as the RMSE decreases.

Topic: Coefficient of determination and RMSE

Question: Which RMSE will have the weakest correlation in the data for the line of best fit?

Answer choices:

A RMSE $= 0.6873$

B RMSE $= 0.0871$

C RMSE $= 0.9423$

D RMSE $= 0.0001$

Solution: C

Remember that the larger the RMSE (standard deviation of the residuals from the least squares line), the more scattered the data points are around the line of best fit, and the weaker the linear correlation in the data.