Teaching R^2 in Regression

Kenneth Sutrick, Murray State University, Murray, Kentucky, USA

ABSTRACT

In regression the coefficient of determination, R^2, measures how well a regression line fits a set of data. How does R^2 do this? Exactly what does R^2 measure? It is taught that R^2 is the percent of explained variance. These are nice words but what do they really mean? The measure R^2 is perhaps the most confusing topic in regression and is confusing to students. This paper suggests ways to teach what the coefficient of determination and its components and interpretations are about.

Keywords: **r**egression, coefficient of determination, R^2, RMSE, percent of explained variance.

INTRODUCTION

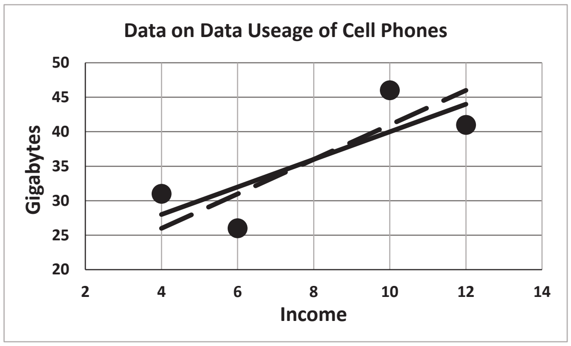
The measure R^2, the coefficient of determination, tells us how well a regression line fits a set of pairs of data. How does R^2 do this? Exactly what do the components of R^2 measure anyway and how does this relate to the fit of a line? Anyone with even just a little more than an introduction to regression also knows that R^2 is the percent of explained variance. Exactly what does percent of explained variance mean? While many people use this phase, there is much confusion about what it is really about. Unfortunately students will likely be confused about all of this when they are taught regression. This paper uses some strategically defined data sets and the visual representations of scatter diagrams to present ideas for the teaching of R^2 and the percent of explained variance.

WHERE THE REGRESSION LINE COMES FROM AND ABOUT R^2

The interpretation of the measure R^2 as percent of explained variance comes from the theoretical formula: $Var(Y) = Var\big(E(Y|X)\big) + E(Var(Y|X))$ with $R^2 = Var(E(Y|X))/Var(Y)$, (Bickel and Doksum, 1977, 36). One must carefully define what all these symbols represent, such as $E(Y|X)$ denoting the regression equation and $Var(Y|X)$ being how far a typical data point tends to be from $E(Y|X)$. Of course using this formula in an introductory statistics class will get you nothing but hate. This formula is far above introductory statistics but can be understood and taught at an elementary level in a simple regression model. We now present methods to explain and teach these concepts.

To put all of this in context it will be necessary to review the subject of regression. For example, a cell phone company will want to predict the data requirements of its customers and will probably use regression to make the predictions. Suppose a hypothetical data set is (X,Y) = (4,31) (6,26) (10,46) (12,41) where X is family household income (in tens of thousands of dollars) and Y is the yearly number of gigabytes of cell phone data used by the household. Figure 1 has the data scatter diagram along with two potential prediction lines (out of infinitely many possible prediction lines), a dashed line: $Y = 16 + 2.5X$, and a black line: $Y = 20 + 2X$.

Figure 1: Data and Potential Prediction Lines



The points in the data set do not fall on a line so that any line that is used to predict with will have some prediction error. Since you will ultimately use only one line to predict with a criteria to distinguish between lines is necessary.

That criteria is the Root Mean Square Prediction Error ($RMSE$). The $RMSE$ calculation for both lines is given below in Table 1. In the calculation 'n' is the number of pairs of data (four here), '$\hat{Y}$' stands for predicted Y, and the symbol 'SSE ' stands for Sum of Squared Errors.

Table 1: Root Mean Square Error Calculations

RMSE Calculations for the **Dashed Line**:

	Actual	Predicted	Prediction	Squared
	Y	Y	Error	Error
X	Y	$\hat{Y} = 16 + 2.5X$	$Y - \hat{Y}$	$(Y - \hat{Y})^2$
4	31	26	5	25
6	26	31	–5	25
10	46	41	5	25
12	41	46	–5	25
			Sum=0	Sum= $SSE = 100$

$RMSE(DashedLine) = \sqrt{\frac{SSE(DashedLine)}{n}} = \sqrt{\frac{100}{4}} = 5,$

RMSE Calculations for the **Black Line**:

	Actual	Predicted	Prediction	Squared
	Y	Y	Error	Error
X	Y	$\hat{Y} = 20 + 2X$	$Y - \hat{Y}$	$(Y - \hat{Y})^2$
4	31	28	3	9
6	26	32	–6	36
10	46	40	6	36
12	41	44	–3	9
			Sum=0	Sum= $SSE = 90$

$RMSE(BlackLine) = \sqrt{\frac{SSE(BlackLine)}{n}} = \sqrt{\frac{90}{4}} \approx 4.7434.$

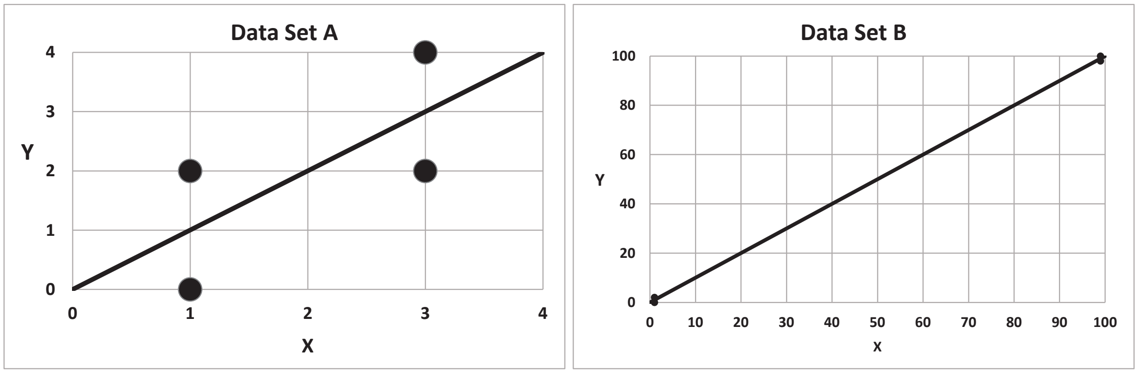
In regression the $RMSE$ measures a typical prediction error. This is evident from the calculation of the $RMSE$ for the dashed line. For the dashed line the prediction error is 5 (ignoring the minus signs) for each pair of points and so a typical prediction error is obviously 5 and the $RMSE$ is therefore 5. The $RMSE$ of 4.74 for the black line is between the actual errors, which are either 3 or 6. The $RMSE$ of 4.74 for black line is smaller than the $RMSE$ of 5 for the dashed line showing that the black line has smaller prediction errors and is the better line for predicting Y from X. The regression line (by definition) is that line that has the smallest $RMSE$ out of all possible lines and for this data set the black line is the regression line. To summarize, the first measure of fit in regression $RMSE$ has two purposes. The first purpose is as a criteria for finding the best fitting line and the second purpose of $RMSE$ is to tell you what a typical prediction error is for that best line.

Calculus (the part for minimizing functions) is used to find the equation of the regression line by finding the line that minimizes the $RMSE$. If we write the equation of a line as: $Y = a + bX$, then as one learns in statistics, calculus proves that the equation for the regression line has slope: $b = r\frac{SD(Y)}{SD(X)}$ and intercept: $a = \bar{Y} - b\bar{X}$, where r is the correlation and the SD's can be either the sample SD's or population SD's (as long as one is consistent between the top and bottom in the formula for b). For the data set in Table 1, it is easy to check that $r = .8$, $\bar{X} = 8$, $\bar{Y} = 36$, and the sample SD's, S_X and S_Y, are $S_X = \sqrt{\frac{SSX}{n-1}} = \sqrt{\frac{(4-8)^2+(6-8)^2+(10-8)^2+(12-8)^2}{4-1}} = \sqrt{40/3}$ and $S_Y = \sqrt{\frac{SSY}{n-1}} = \sqrt{\frac{(31-36)^2+(26-36)^2+(46-36)^2+(44-36)^2}{4-1}} = \sqrt{250/3}$, so that $b = .8\frac{\sqrt{250/3}}{\sqrt{40/3}} = 2$, and $a = 36 - 2(8) = 20$. This gives the regression line as: $Y = 20 + 2X$, which is the black line above.

To look at the next aspect of fitting equations to data, consider another prediction problem. Suppose that we are predicting yearly incomes of people in the United States and the $RMSE$ is \$200,000. (For example, if we predict that every single person in the US makes exactly \$300,000 per year, this would give an $RMSE$ of at least \$200,000.) However this would be a terrible prediction since we could do much better by just predicting that everyone makes the (available) average US income. On the other hand if we were predicting Bill Gates' yearly income and the $RMSE$ were \$200,000 this would be an excellent prediction, given Bill Gates' place on the upper end of the income scale. Both situations have the same $RMSE$, yet in one case the predictions are terrible and in the other case the prediction is excellent. This tells us that we need something else besides $RMSE$ when talking about fitting equations to data. That something else is called R^2.

Before getting to the exact definition, to see what R^2 is about consider the following two data sets: Data Set A is $(X, Y) = (1,0)\ (1,2)\ (3,2)\ (3,4)$, and Data Set B is $(X, Y) = (1,0)\ (1,2)\ (99,98)(99,100)$. After calculation, it is seen that for Data Set A: $r = \frac{\sqrt{2}}{2} \approx .707107$, $\bar{X} = 2$, $\bar{Y} = 2$, $S_x = \sqrt{4/3}$, $SSY = \sum(Y - \bar{Y})^2 = 8$, and $S_Y = \sqrt{8/3}$, giving regression equation coefficients: $b=\frac{\sqrt{2}}{2}\frac{\sqrt{8/3}}{\sqrt{4/3}}=1$ and $a=2-(1)2=0$. Therefore for Data Set A the regression equation is: $Y=0+1X$. After calculation, it is seen that for Data Set B: $r=\frac{\sqrt{9604}}{\sqrt{9608}}\approx .999792$, $\bar{X}=50$, $\bar{Y}=50$, $S_x=\sqrt{9604/3}$, $SSY=\sum(Y-\bar{Y})^2=9608$, and $S_Y=\sqrt{9608/3}$, giving regression equation coefficients: $b=\frac{\sqrt{9604}}{\sqrt{9608}}\frac{\sqrt{9608/3}}{\sqrt{9604/3}}=1$ and $a=50-(1)50=0$, and also giving a regression equation for Data Set B of: $Y=0+1X$. Both data sets have the same regression equation. A graph of both data sets with their regression lines are in Figure 2. Note that Data Set A and Data Set B are on different scales. This will ultimately illustrate what R^2 is about.

Figure 2: Two Data Sets with the Same Regression Equation but Different Fits



The $RMSE$ for both data sets is computed in Table 2.

Table 2: $RMSE$ Calculations for Data Set A and Data Set B

Data Set A

X	Y	$\hat{Y}=0+1X$	$Y-\hat{Y}$	$(Y-\hat{Y})^2$
1	0	1	–1	1
1	2	1	1	1
2	1	2	–1	1
2	3	2	1	1
			Sum=0	$SSE=4$

$$RMSE(DataSetA)=\sqrt{\frac{SSE(DataSetA)}{n}}=\sqrt{\frac{4}{4}}=1,$$

Data Set B

X	Y	$\hat{Y}=0+1X$	$Y-\hat{Y}$	$(Y-\hat{Y})^2$
1	0	1	–1	1
1	2	1	1	1
99	98	99	–1	1
99	100	99	1	1
			Sum=0	$SSE=4$

$$RMSE(DataSetB)=\sqrt{\frac{SSE(DataSetB)}{n}}=\sqrt{\frac{4}{4}}=1.$$

The prediction error, $Y-\hat{Y}$, is 1 for every pair of data in both data sets.

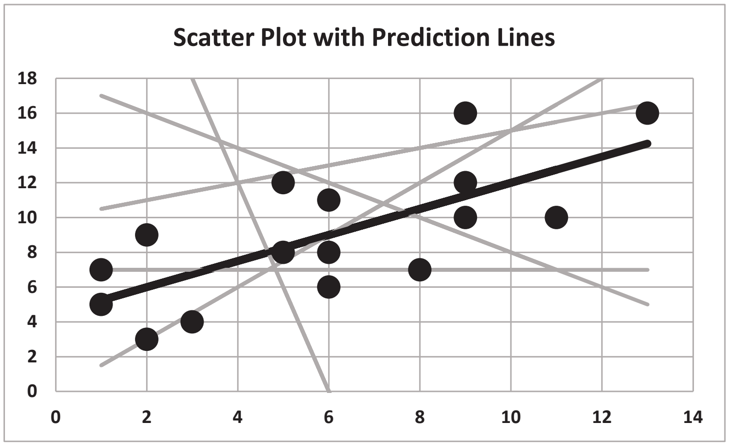
Both data sets have the same regression line and the same $RMSE$. If you are using $RMSE$ as a measure of fit then the line $Y=0+1X$ fits both data sets equally well. However when looking at Figure 2, it appears that the line fits Data Set B better (points look closer to the line). The fact that Data Set B is fit better is also evident in the correlations, since the correlation between X and Y for Data Set B is approximately .9998, while the correlation between X and Y for Data Set A is approximately .7071. So $RMSE$ cannot distinguish between the two data sets. This is where the second measure of fit R^2 comes into play, R^2 *can* tell the difference in fit in these two data sets. Before defining R^2 exactly, let's look closer at why the line fits Data Set B better. In Data Set B the distance of the points to the line (measured by the $RMSE=1$) is small compared to the spread in the Y-values (the range of the Y's is 0 to 100, and from previous the SD of the Y's is $S_Y=\sqrt{9608/3}\approx 56.6$). While in Data Set A the distance of the points to the line (measured by the $RMSE=1$) is similar to the spread in the Y-values (the range of the Y's is 0 to 3, and the SD of the Y's is $S_Y=\sqrt{8/3}\approx 1.6$). In Data Set B we are comparing an error of 1 to a Y-range of 100 (or error 1 to $SD(Y)=56.6$), while in Data Set A we are comparing an error of 1 to a Y-range of 3 (or error 1 to $SD(Y)=1.6$). This is why the line fits Data Set B better. What R^2 does is to measure the extent to which this happens. Since the standard deviation is the more common way to measure spread (compared to the range as a measure of spread), R^2 will be defined by in some way comparing the $RMSE$ to $SD(Y)$. Since $RMSE = \sqrt{SSE/n}$, this gives that SSE is also a measure of the distance of the points to the line (just on a different scale). For example, if SSE is small then $RMSE$ will be small and vice versa. Since the sample standard deviation is $SD(Y) = \sqrt{SSY/(n-1)}$, this shows that SSY is also a measure of the spread in the Y's (just on a different scale). This is in the sense that if SSY is large then $SD(Y)$ will be large also and vice versa. Finally R^2 is defined as

Equation 1: $$R^2 = 1 - \frac{SSE}{SSY} \quad \text{or} \quad R^2 = \left(1 - \frac{SSE}{SSY}\right) \times 100\%.$$

If a line fits the data well then the prediction error, as measured by SSE, will be much smaller than the spread in the Y's, as measured by SSY, and R^2 will be close to 1 (or 100%). For Data Set A: $SSE = 4$, $SSY = 8$, and $R^2 = 1 - 4/8 = .5$, while for Data Set B: $SSE = 4$, but $SSY = 9608$, and $R^2 = 1 - 4/9608 = .999584$. Thus R^2 can distinguish between Data Set A and Data Set B while $RMSE$ cannot. In this sense R^2 can be called the coefficient of determination.

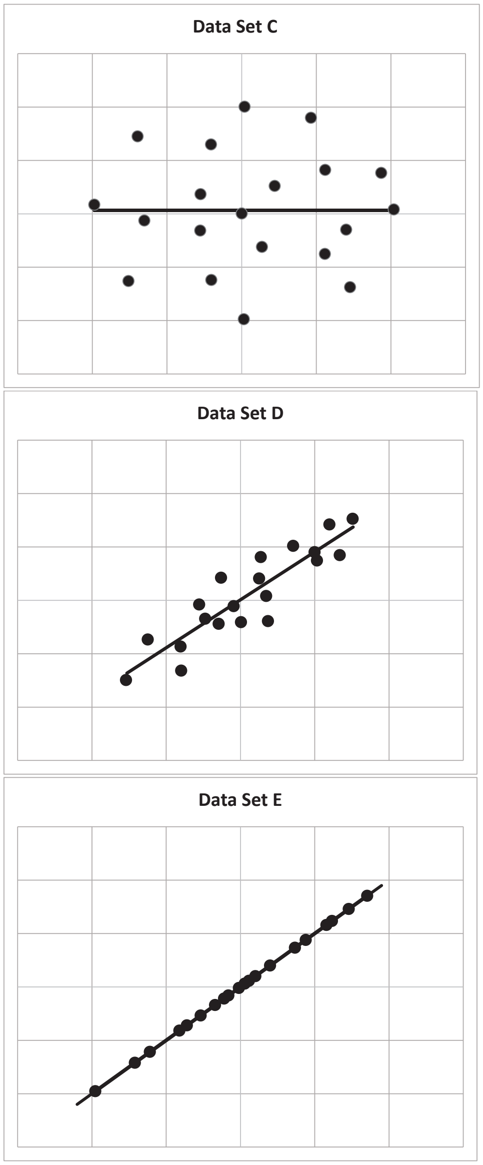
The two measures of fit in regression, $RMSE$ and R^2, can be characterized in the following way. There are infinitely many lines that could be fit to a set of data and $RMSE$ chooses the best one (the regression line) out of that infinite set of lines. This is illustrated in the Figure 3 below. $RMSE$ then measures and tells you what a typical prediction error is for that best line, the regression line. Once you have the best line, R^2 tells if that best line fits or not. This is illustrated in Figure 4 below, where R^2 distinguishes between three different scatter diagrams. Figure 3 and Figure 4 demonstrate the difference between $RMSE$ and R^2. Sometimes even the best line does not do very well, which R^2 will show.

Figure 3: The Purpose of RMSE



$RMSE$ picks out the best line for predicting Y from X. The regression line (in black) is the best prediction line. The $RMSE$ of the regression line, when calculated here, is about 2.5 units. A typical data point is about 2.5 units directly above or directly below the black line. Some points are more than this and some are less.

Figure 4: Data Sets with Different R^2 Values



$RMSE$ says the regression line (in black) is the best line for predicting Y from X. However since this data is basically a ball and balls do not look like lines, no line (including the regression line) fits very well.

Here $R^2 = 0$

$RMSE$ says the regression line (in black) is best here. The regression line fits fairly well.

Here $R^2 = .5$, when calculated.

$RMSE$ says the regression line (in black) is best here also. Here the regression line fits perfectly, there is no prediction error.

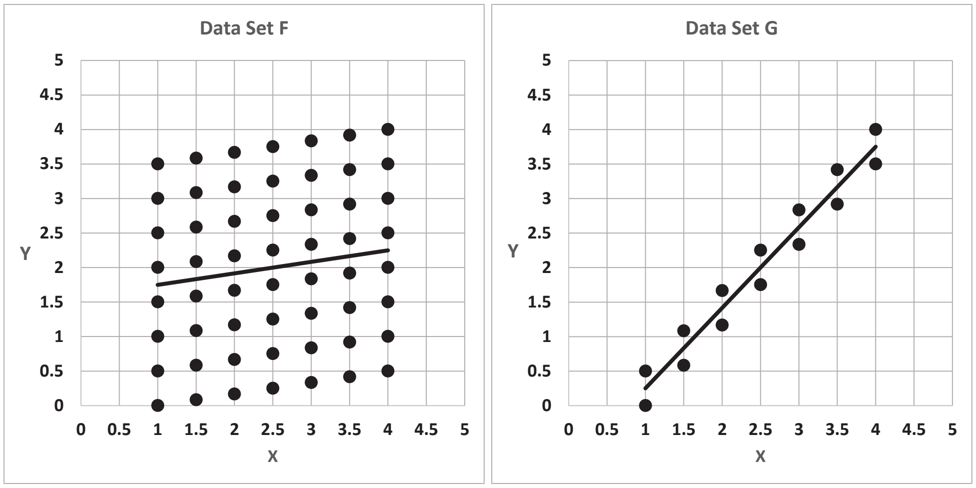
Here $R^2 = 1$.

$RMSE$ is a relative measure of fit while R^2 is a more absolute measure of fit. In a data set if someone tells you that $R^2 = .5$, then you know that the data looks something like Data Set D. If someone tells you that the $RMSE = 108$ and you do not have other any information, then you don't have any idea what the data looks like (except that it could not look like Data Set E, which has $RMSE = 0$). In a relative sense we know that we want $RMSE$ to be small and R^2 to be high. You would need other information to tell if 108 is an acceptable prediction error or not. In one situation 108 might be good and in another situation it might be terrible. The measure R^2 utilizes that other information. For a data set the $RMSE$ can be anything between zero and infinity but with smaller better and zero best. However R^2 has an absolute scale of 0 to 1. In R^2 the absolute best is 1 and the absolute worst is 0. If Data Sets C, D, and E in Figure 4 were all on the same scale (range of X's and Y's the same, which you cannot tell since the axis are not labeled) then the $RMSE$ could distinguish between them and would tell us that the regression line fits E best then D then C. However, ultimately R^2 is a more important measure of fit since it is comparable across all data sets whether they have the same X and Y range or not. [That doesn't mean $RMSE$ is not important since it tells you a typical prediction error. With knowledge of only the $RMSE$, you could still decide if that size of prediction error is acceptable.]

PERCENT OF EXPLAINED VARIANCE

The term “percent of explained variance” in regression is short for the “percent of explained variance of Y which is explained by the variability in the regression equation”. Now what does that second phrase mean? The meaning is illustrated by Data Set F and Data Set G in Figure 5 and explained thereafter. The black line in both pictures is their respective regression lines.

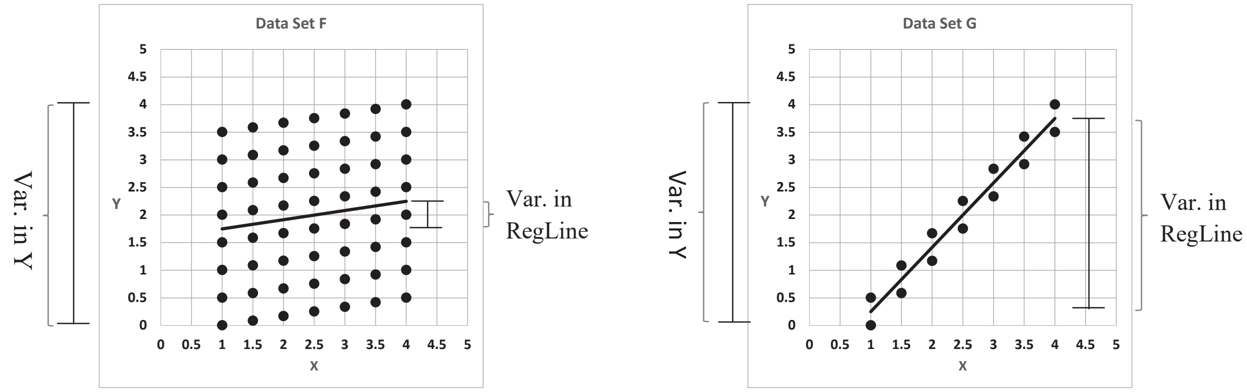
Figure 5: A Graph of the Regression Line for Two Different Data Sets



In Figure 5 you can see that Data Set F has a relatively low but positive correlation, while Data Set G has a relatively high and positive correlation. Since the slope of the regression line is: $b = r\,SD(Y)/SD(X)$, where r is the correlation, if you don’t worry too much about the SD’s then b for Data Set F will be small but positive while b for Data Set G will be high and positive. This tells you that the regression line for Data Set F is relatively flat, while the regression line for Data Set G will be relatively steep. This is evident in Figure 5.

The percent of explained variance compares the variability in the Y-values to the variability in the regression line. For the moment we will use the range (top to bottom) as a measure of variability. In both data sets the lowest Y-value is $Y = 0$, at the point $(X, Y) = (1,0)$, and the highest Y-value is at $Y = 4$ at the point $(4,4)$. Both data sets then have Y-ranges of 4 or said another way a Y-variability of 4. For the range of the regression line for Data Set F, the lowest part of the regression line has $Y = 1.75$ at the point $(X, Y) = (1,1.75)$ and the highest point on the regression line has $Y = 2.25$ at the point $(4,2.25)$, giving a regression line range of $.5 = 2.25 - 1.75$. This regression line range will be called the RegLine-variability. For Data Set F we are comparing a RegLine-variability of 0.5 to a Y-variability of 4.0. Since 0.5 is quite a bit smaller than 4.0, we say that the regression line for Data Set F has a low percent of explained variance. For Data Set G the lowest part of the regression line has $Y = .25$ at the point $(1,.25)$ and the highest point on the regression line has $Y = 3.75$ at the point $(4,3.75)$, giving a regression line range of $3.5 = 3.75 - .25$. For Data Set G we are comparing a RegLine-variability of 3.5 to a Y-variability of 4.0, which are almost the same. We say that the regression line for Data Set G has a high percent of explained variance. The comparative ranges or distances are illustrated in Figure 6, showing what percent of explained variance is about.

Figure 6: Illustrating Percent of Explained Variance

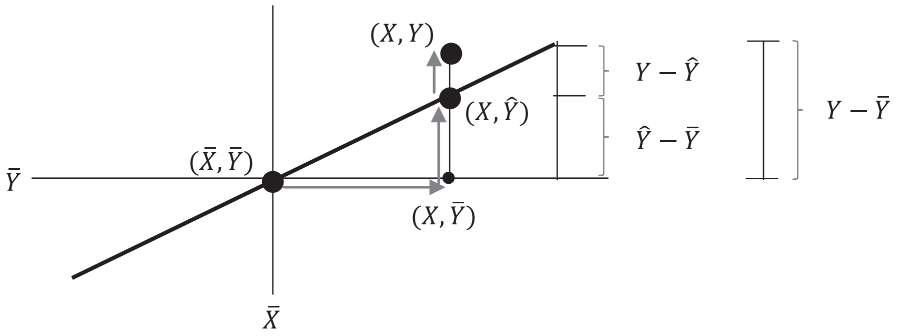


Small % of Explained Variance

High % of Explained Variance

There is a second qualitative and quantitative way to think about the percent of explained variance. Think about how can you go from the point $(\bar{X}, \bar{Y})$ to a data pair (X, Y). This is accomplished as follows. Starting from $(\bar{X}, \bar{Y})$ go along horizontal to the point $(X, \bar{Y})$, then go vertical to the regression line to the point $(X, \hat{Y})$, and then go from the point $(X, \hat{Y})$ to the point (X, Y), as illustrated in Figure 7.

Figure 7: An Alternative Interpretation of Percent of Explained Variance



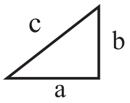
On the journey from $(\bar{X}, \bar{Y})$ to (X, Y), in general the longer the distance $\hat{Y} - \bar{Y}$ compared to the distance $Y - \hat{Y}$ the higher the percent of explained variance. This is evident in Figure 6 where when moving along the regression line you move a greater vertical distance up or down for Data Set G than for Data Set F, while moving off the regression line to the points requires a smaller distance traveled for Data Set G compared to Data Set F. [The point $(X, \hat{Y})$ is on the regression line since you get $\hat{Y}$ by plugging in X into the regression equation. The point $(\bar{X}, \bar{Y})$ is also on the regression line since $a = \bar{Y} - b\bar{X}$, the regression equation $Y = a + bX$ can be rewritten as $Y = \bar{Y} - b\bar{X} + bX$. When plugging in $\bar{X}$ for X into the equation, the predicted Y is $\bar{Y}$.]

Figure 6 showed variability using ranges since they are easy to see on a scatter diagram. The actual percent of explained variance is not defined in terms of ranges, but instead defined with the more commonly used measures of spread such as standard deviations and variances, which are not easy to see on pictures. [We will need to remember that $Var = SD^2$.] For the actual quantitative definition of percent of explained variance, start with Y and then rewrite Y as $Y = \hat{Y} + (Y - \hat{Y})$. Next subtract $\bar{Y}$ from both sides to get: $Y - \bar{Y} = \hat{Y} - \bar{Y} + (Y - \hat{Y})$ or $(Y - \bar{Y}) = (\hat{Y} - \bar{Y}) + (Y - \hat{Y})$. After that perform the calculations in Table 3 (done on the cell phone data since it has fewer numbers). For the percent of explained variance we will quantitatively compare the spread in the regression line to the spread in the Y's.

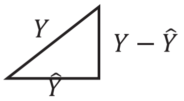
Table 3: Showing SSY = SSR + SSE

X	Y	$\hat{Y}=20+2X$	$Y-\bar{Y}$	$(Y-\bar{Y})^2$	$\hat{Y}-\bar{Y}$	$(\hat{Y}-\bar{Y})^2$	$Y-\hat{Y}$	$(Y-\hat{Y})^2$
4	31	28	–5	25	–8	64	3	9
6	26	32	–10	100	–4	16	–6	36
10	46	40	10	100	4	16	6	36
12	41	44	5	25	8	64	–3	9
	$\bar{Y}=36$			Sum=SSY=250		Sum=SSR=$SS\hat{Y}$=160		Sum=SSE=90
				$=\sum(Y-\bar{Y})^2$		$=\sum(\hat{Y}-\bar{Y})^2$		$=\sum(Y-\hat{Y})^2$.

As learned in introductory statistics and shown previously in the paper, the sum SSY is used to calculate $SD(Y)$. The symbol 'SSR' stands for "sum of squares regression", with the name coming from the fact the predicted Y, the $\hat{Y}$'s, in SSR come from the regression line. In Table 3 you can see that in this case $SSY = SSR + SSE$. It turns out that in the theory of regression it can be proved that this happens for every data set as long as there is an intercept term in the regression equation and the $\hat{Y}$ come from the regression equation (as opposed to predictions from some other line). [This relationship is in fact related to the Pythagorean Theorem, which says that if two legs of a triangle are perpendicular then $a^2 + b^2 = c^2$. If you consider the $\hat{Y}$-values as a vector and the $(Y-\hat{Y})$-values as a vector, the theory of regression shows that $\hat{Y}$ and $Y-\hat{Y}$ are perpendicular vectors. So



$a^2 + b^2 = c^2$



$SSR + SSE = SSY$.]

This relationship will give an alternate formula for R^2, compared to Equation 1. You get that by dividing everything in the relationship by SSY, which gives $SSY/SSY = SSR/SSY + SSE/SSY$ or $1 - SSE/SSY = SSR/SSY$. Since the definition of R^2 is $R^2 = 1 - SSE/SSY$, the above relationship now gives $R^2 = SSR/SSY = SS\hat{Y}/SSY$. From this version of R^2 we get the interpretation "percent of explained variance" in the following way. One must first be specific about how to calculate variance. Remember that $SD(Y) = \sqrt{SSY/(n-1)}$, so that $Variance(Y) = Var(Y) = SD(Y)^2 = SSY/(n-1)$. In other words you get $Var(Y)$ by taking the Y's, subtracting their average, squaring them, adding the squares, and then dividing by $n-1$. You would do the same thing to the $\hat{Y}$'s to get $Var(\hat{Y})$. It is necessary to first find the average of the $\hat{Y}$'s. If predicting with a line so that $\hat{Y}_i = a + bX_i$, for constants a and b, then averaging over all predicted values leads to $\bar{\hat{Y}} = a + b\bar{X}$. If as is the case for the regression line, $a = \bar{Y} - b\bar{X}$, then $\bar{\hat{Y}} = (\bar{Y} - b\bar{X}) + b\bar{X} = \bar{Y}$. [For the Black Line, in Table 1: $\bar{\hat{Y}} = (28+32+40+44)/4 = 36 = \bar{Y}$.] So in $SSR = SS\hat{Y}$ you are taking the $\hat{Y}$'s, subtracting their average, squaring them, adding the squares, and then dividing by $n-1$. You must then get $Var(\hat{Y})$. Therefore

$$R^2 = \frac{SSR}{SSY} = \frac{SS\hat{Y}}{SSY} = \frac{SS\hat{Y}/(n-1)}{SSY/(n-1)} = \frac{Var(\hat{Y})}{Var(Y)}.$$

$Var(Y)$ is a representative of the variability of Y, and since $\hat{Y}$ comes from the regression equation $Var(\hat{Y})$ represents the variability in the regression equation. The above ratio is where the meaning of R^2 as the percent of explained variance comes from (a percent is a part divided by the whole times 100%). [Note: some books use SST, "sum of squares total" for SSY.] For the cell phone data: $R^2 = SSR/SSY = 160/250 = .64$.

The multiple correlation R is defined as $R = \sqrt{R^2}$. In the case, where there is only one predictor X, it can be proved that $R = r$, where $r = Corr(Y, X)$. Thus R can be interpreted as a correlation. For the cell phone data: $r = .8$ and $R = \sqrt{.64} = .8$. [When there is one predictor then $R^2 = r^2$ also]. There is also a second interpretation of R. From Table 3, $(Y, \hat{Y}) = (31,28)\ (26,32)\ (46,40)\ (41,44)$ and if you correlate these two sets of numbers it can be proved that $R = Corr(Y, \hat{Y})$ also. [For the cell phone data: $Corr(Y, \hat{Y}) = .8$.] The closer the predictions $\hat{Y}$ are to the actual Y's, the higher the $Corr(Y, \hat{Y})$ will be, and the higher R and R^2 will be.

CONCLUSION

The measure R^2, the coefficient of determination, is the ultimate measure of whether a regression equation fits a data set. Therefore R^2 is quoted all the time. However despite its constant use there is a lot of mystery about what it really means. That results in a lot of difficulty in the teaching of R^2. This paper illustrates how one might teach what the components of R^2 are about and therefore lets students understand how R^2 measures whether a regression equation fits or not. Next the coefficient of determination has an interpretation as the percent of explained variance. This phrase is also confusing and the paper suggests ways of looking at explained variance so that it can be understood by students. Finally the multiple correlation, R, is discussed, completing the teaching of fit in regression.

REFERENCES

Bickel, P. and Doksum, K. (1977). *Mathematical Statistics*. San Francisco: Holden Day.

Kenneth Sutrick, Ph.D., teaches statistics in the Bauernfeind College of Business at Murray State University, Murray, Kentucky. He was born in Green Bay, WI, therefore born a Green Bay Packers Fan. His research interests include portfolio theory, and options.