

Section 4.3

The Coefficient of Determination

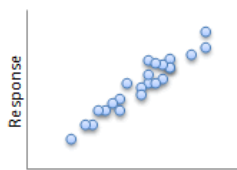
The **coefficient of determination**, R^2 , is the percent of the variation in the response variable (y) that can be explained by the least-squares regression line.

Looking at the definition, we can see that a higher R^2 is better - the LSR line does a *better* job of explaining the variation in the response variable.

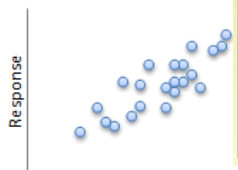
The coefficient of determination measures the percentage of total variation in the response variable that is explained by the least-squares regression line. R^2 is also equal to one minus the unexplained variation divided by the total variation.

$$R^2 = 1 - \frac{\text{unexplained variation}}{\text{total variation}}$$

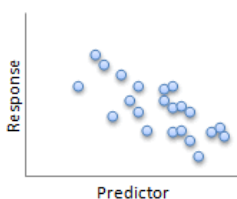
Here are some examples:



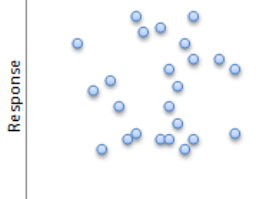
$R^2 = 92.1\%$



$R^2 = 70.0\%$



$R^2 = 53.4\%$



No $R^2 = 13.5\%$

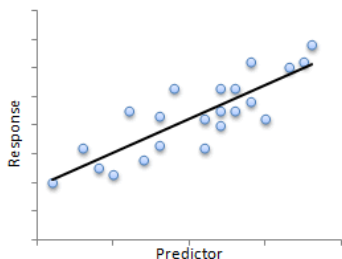
The deviation between the predicted and mean values of the response variable is called the explained deviation, so explained deviation is $\hat{y} - \bar{y}$. The deviation between the observed and predicted values of the response variable is called the unexplained deviation, so unexplained deviation is $y - \hat{y}$. Therefore, total deviation equals unexplained deviation plus explained deviation.

$$y - \bar{y} = (y - \hat{y}) + (\hat{y} - \bar{y})$$

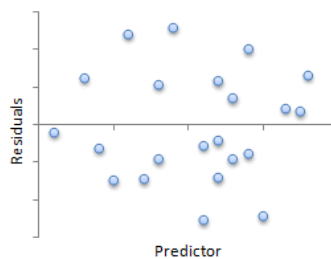
The lower the percent the more scattered the points are on the graph.

When a Linear Model is Appropriate

Sometimes it can be difficult to determine if any of the three above conditions have been violated, but here's a good example of a situation where a linear model does seem appropriate.



The LSR line seems to fit the data.



The residuals are evenly spread above and below zero, there is no discernable pattern, and there are no outliers.

Technology

Here's a quick overview of how to create a residual plot in StatCrunch.

1. Select **Stat > Regression > Simple Linear**
2. Set the X-Variable and Y-Variable and press **Next**.
3. Select **Save residuals** (optional) and press **Next**.
4. Select the options you want - make sure to select "Residuals vs. X-values" is the residual plot.
5. Press **Calculate**.
6. The output will show your regression analysis. On the bottom, press **Next** to see any graphics.

Multiply r-sq by 100 to get % then round