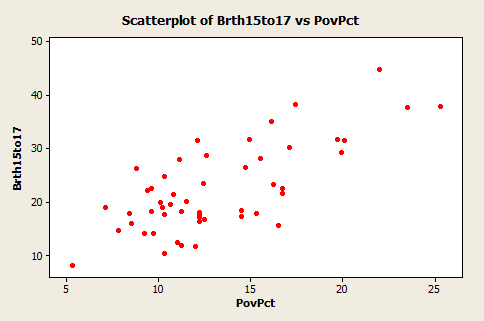
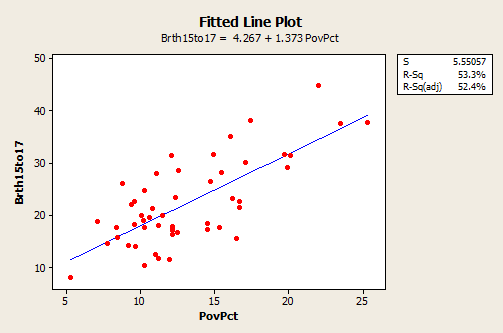
### Simple linear regression

### Example

This dataset of size n = 51 are for the 50 states and the District of Columbia in the United States ([poverty.txt](https://onlinecourses.science.psu.edu/stat462/sites/onlinecourses.science.psu.edu.stat462/files/data/poverty.txt)). The variables are y = year 2002 birth rate per 1000 females 15 to 17 years old and x = poverty rate, which is the percent of the state’s population living in households with incomes below the federally defined poverty level. (Data source: Mind On Statistics, 3rd edition, Utts and Heckard).



The plot of the data (birth rate on the vertical) shows a generally linear relationship, on average, with a positive slope. As the poverty level increases, the birth rate for 15 to 17 year old females tends to increase as well.



Y= 4.267+1.373x

The equation of the fitted regression line is given above. The equation should really state that it is for the “average” birth rate (or “predicted” birth rate would be okay too) because a regression equation describes the average value of y as a function of one or more x-variables. In statistical notation, the equation could be written.

* The interpretation of the slope (value = 1.373) is that the 15 to 17 year old birth rate increases 1.373 units, on average, for each one unit (one percent) increase in the poverty rate.
* The interpretation of the intercept (value=4.267) is that if there were states with poverty rate = 0, the predicted average for the 15 to 17 year old birth rate would be 4.267 for those states. Since there are no states with poverty rate = 0 this interpretation of the intercept is not practically meaningful for this example.

We also note the information that s = 5.55057 and r2 = 53.3%.

* The value of s tells us roughly the standard deviation of the differences between the y-values of individual observations and predictions of y based on the regression line.
* The value of r2 can be interpreted to mean that poverty rates "explain" 53.3% of the observed variation in the 15 to 17 year old average birth rates of the states.

The R2 (adj) value (52.4%) is an adjustment to R2 based on the number of x-variables in the model (only one here) and the sample size. With only one x-variable, the adjusted R2 is not important.