**Equation of the least-squares regression line**

We have data on an explanatory variable *x* and a response variable *y* for *n* individuals. From the data, calculate the means and standard deviations of the two variables and their correlation. The least squares regression line is the line *ŷ* = *a* + *bx* with

**slope**



and ***y* intercept**



Warmup:

A study was done comparing the number of registered automatic weapons (in thousands) along with the murder rate (in murders per 100,000) for 8 randomly selected states. The number of registered weapons had a mean of 4.8 and a standard deviation of 3.74 The murder rate had a mean of 8.1and a standard deviation of 3.6. The correlation r = .885.

1. Find the least square regression line for this relationship?
2. Predict the murder rate if 10 thousand automatic guns are registered

Example #2:

It is believed a student’s Verbal SAT score can be used to predict their Math SAT score. Data for a high school graduating class was compiled and it was found that the mean Verbal score was 596.3 with a standard deviation of 99.5. The Math scores had a mean of 612.2 with a standard deviation of 96.1. If r = .685, find the equation of the regression line. What would you predict the Math score to be if the Verbal score is 615?

**Residuals**

A residual is the difference between an observed value of the response variable and the value predicted by the regression line. That is,

residual = observed *y* – predicted *y*

residual = *y* – *ŷ*

Using the regression line from Example #2:

1. Calculate the predicted Math SAT score if a student scored 610 on the Verbal part
2. John had a Verbal score of 610 and a Math score of 615. What was the residual for John’s result?

A **residual plot** is a scatterplot of the residuals against the explanatory variable. Residual plots help us assess how well a regression line fits the data.



**Interpreting Residual Plots**

A residual plot magnifies the deviations of the points from the line, making it easier to see unusual observations and patterns.

* + - 1. The residual plot should show no obvious patterns
			2. The residuals should be relatively small in size.

  

**The Role of *r*2 in Regression**

There is a numerical quantity that tells us how well the least-squares regression line predicts values of the response *y*.

The **coefficient of determination *r*2** is the fraction of the variation in the values of *y* that is accounted for by the least-squares regression line of *y* on *x*.

Consider the body weight vs backpack weight example. We found that r = .794. Calculate r2.

This tells us what % of the variation in backpack weight is accounted for by the linear model relating pack weight to body weight.