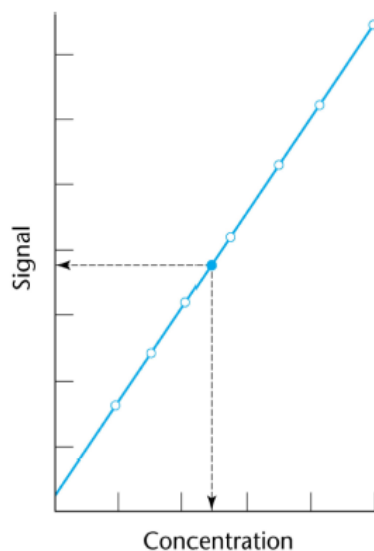## Regression Line

A regression line is a line drawn through the points on a scatterplot to summarise the relationship between the variables being studied. When it slopes down (from top left to bottom right), this indicates a negative or inverse relationship between the variables; when it slopes up (from bottom right to top left), a positive or direct relationship is indicated.

The regression line often represents the regression equation on a scatterplot.

Simple Linear Regression
Simple linear regression aims to find a linear relationship between a response variable and a possible predictor variable by the method of least squares.

Ordinary Least Squares
The method of least squares is a criterion for fitting a specified model to observed data. For example, it is the most commonly used method of defining a straight line through a set of points on a scatterplot.

Regression Equation
A regression equation allows us to express the relationship between two (or more) variables algebraically. It indicates the nature of the relationship between two (or more) variables. In particular, it indicates the extent to which you can predict some variables by knowing others, or the extent to which some are associated with others.

Recall that we class one variable as the response or dependent variable, usually denoted Y, and the other as the predictor, or dependent variable, usually denoted X.
X is said to "cause" changes in Y.

If a linear relationship , the relationship between X and Y is formulated as follows:

$$Y = \beta_o + \beta_1 X + \varepsilon$$

- is the predicted value for the dependent variable
- β_0 is the intercept coefficient
- β_1 is the slope coefficient
- X is the independent variable
- ε is the residual term (i.e random error)

Both β_0 and β_1 are almost always unknown (population) values. However these are the key terms in the model. From a sample of data estimates for the slope and estimate coefficient are derived.

A fitted line to model the data as a linear regression mode (i.e. a regression equation) is usually written as

$$= b_o + b_1 X$$

where
- is the predicted value for the dependent variable
- b_0 is the intercept estimate
- b_1 is the slope estimate (or regression coefficient)
- X is the independent variable

Simple linear regression is from a family of models known as Linear Models. The **R** command used to implement such models is `lm()` .

The regression model is specified in the following form: `lm(Y ~ X)`

The operator "~" (the tilde sign) is taken to mean "is explained by" or "is predicted by".

For our previous example, the simple linear model can be implemented as follows:

```
> lm(Fluo~Conc)

Call:
lm(formula = Fluo ~ Conc)

Coefficients:
(Intercept)          Conc
      1.518         1.930
```
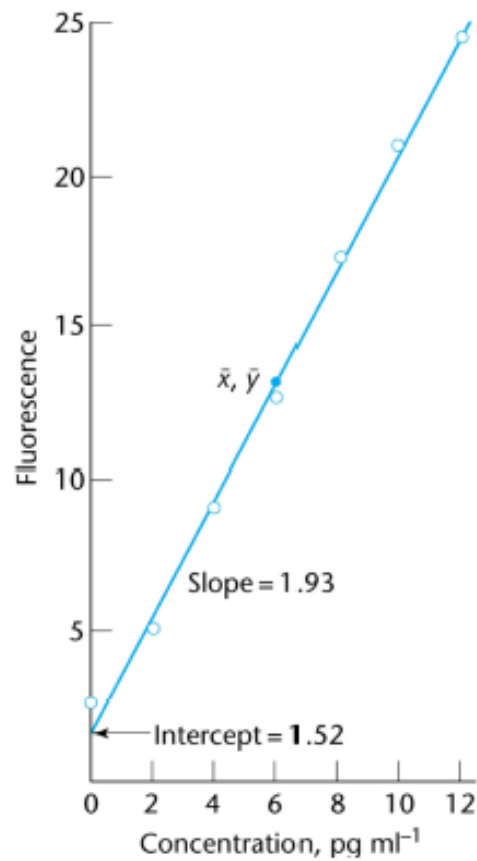
Using the coefficients given in the computer output, the regression equation is therefore
 = 1.52 + 1.93X

Where

= Fluoresence

X = Concentration

The equation will specify the average magnitude of the expected change in Y given a change in X. The regression equation is often represented on a scatterplot by a regression line.

Residual

Residual (or error) represents unexplained (or residual) variation after fitting a regression model. It is the difference (or left over) between the observed value of the variable and the value suggested by the regression model.
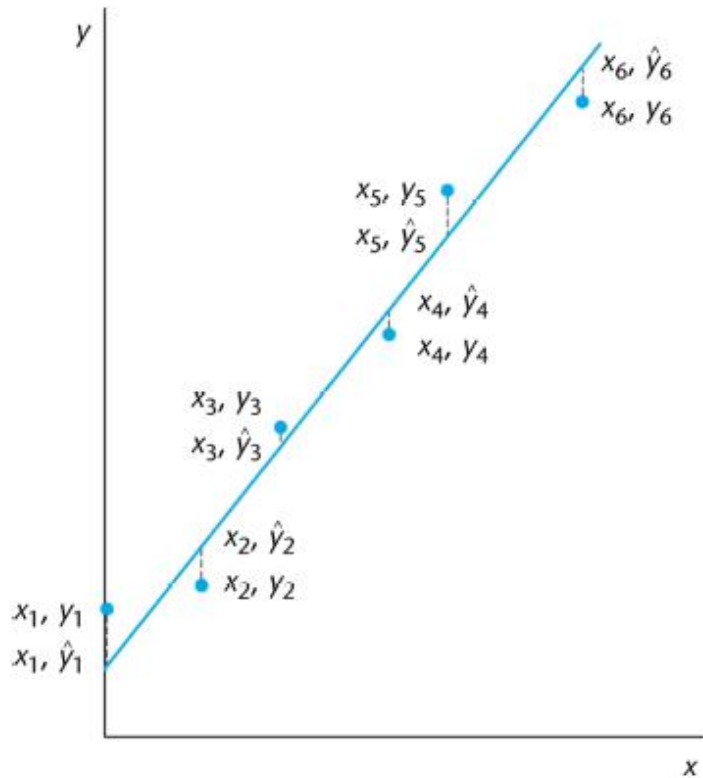
For the example used in this class, the residuals are very small.