**Regression:**

model the relationship between 2 or more variables from which we can:

* observe the effect of independent variables (inputs) on the dependent variable (output)
* predict the values for new data (eg. forecasting)
* determine the relative importance of variables
* linear regression assumes a straight-line relationship
* many other relationships can be modelled
* fitting a regression model is a form of supervised learning (model is ‘learned’ from data consisting of known inputs and outputs)

**simple least squares regression:**

* assumptions:
  + y=ax+b
  + x & y: numerical
  + a & b: calculated to minimise the squared error between the observed values and the fitted values
  + errors approximately normally distributed
* lm() function

multiple linear regression:

ordinary least squares (OLS) applies to multiple predictors:

y = a1x1+a2x2+…+b(+e)

**qualitative/categorical predictors:**

each factor level is included as a variable in the regression equation

indicator variables (0, 1) show the status of each observation at each factor level