Curve fitting in real life

Content

Charles's Law - Linear regression

Charles's law is a gas law that describes how gases tend to expand when heated. It describes the relationship between gas temperature (in Kelvin) and the volume.

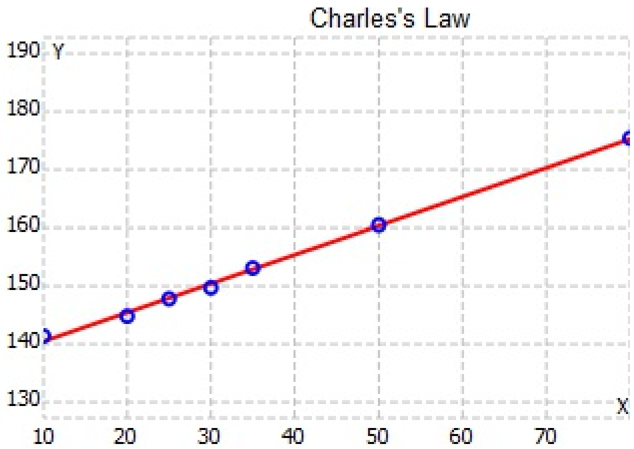
The law was named after Jacques Charles, who formulated the law. He made a series of investigations of the effect of changing temperature on the volume of a sample of air. Specifically, he investigated the change of volume of a fixed amount of air as a function of temperature at a fixed pressure.

When the pressure on a sample of a gas is held constant, the Kelvin temperature and the volume will be directly related.

We will use this sample data to demonstrate how to use curve fitting to present numerically the connection.

Temperature (°C)	Volume (cm^3)
10	141,4
20	144,8
25	147,8
30	149,7
35	153,1
50	160,5
80	175,4

$$xy := \begin{bmatrix} 10 & 20 & 25 & 30 & 35 & 50 & 80 \\ 141.4 & 144.8 & 147.8 & 149.7 & 153.1 & 160.5 & 175.4 \end{bmatrix}$$



Dependency function between temperature and volume

$$a := \text{linfit}(xy)$$

$$a = 0.497\ x + 135.496$$

The volume is directly proportional to the temperature in Kelvin

$$a == 0 = \begin{bmatrix} x & -272.672 \end{bmatrix}$$

$$-272.6723\ ^\circ C = -272.672\ ^\circ C$$

This is the concept behind absolute zero volume and the Kelvin temperature scale: -273 °C = 0 K.

With this example we showed how the volume of the gas is directly proportional to the temperature in Kelvins.

Polynomial regression

Let's take data which gives the age and average number of births per 1000 women at that age. We will later find the function that best fits the data and try to predict the number of births for women with a age of 45.

Age	Births/1000 women
16	34
18,5	86,5
22	111,1
27	113,9
32	84,5
37	35,4
42	6,8
45	We will calculate this value

$$xz := \begin{bmatrix} 16 & 18.5 & 22 & 27 & 32 & 37 & 42 \\ 34 & 86.5 & 111.1 & 113.9 & 84.5 & 35.4 & 6.8 \end{bmatrix}$$

regressiontable(xz) =

Method	R-square	Adj. R-sq	RMSE	SSE	Coeff Num.
"lin"	0.225	0.070	34.155	8166.181	2
"exp"	-0.198	-0.437	42.467	12623.829	2
"log"	0.13	-0.044	36.196	9170.998	2
"pow"	-0.268	-0.521	43.693	13363.935	2
"poly2"	0.869	0.843	14.042	1380.182	3
"poly3"	0.995	0.994	2.843	56.571	4
"poly4"	0.995	0.994	2.839	56.411	5
"poly5"	0.999	0.999	0.988	6.835	6
"poly6"	1	1	3.903e-6	1.066e-10	7

From the regression table above we can determine which regression best fits the inputed data and which type of regression function we will use to describe this model.

As you can see, polynomial regressions of 3rd, 4th and 5th order have the largest R-square values (R-square with the value 1 we will not take into concideration because it tells us that this curve passes through all the data points, so we talk about interpolation and not regression in this case).
On the other hand, Adjusted R-square values for polynomials of 3rd and 4th orders are same, and 5th order value is the largest while it's root mean square error (RMSE) and sum of square errors (SSE) values are the smallest.

So the polynomial regression of the 5th order is the best regression to use, but we will use the 4rd order for this model because we have to use it for prediction (extrapolation) and the 5th order polynomials start to climb too fast,and to predict values with this polynomial would not be credible.

Polynomial regression of the 4th order formula, that we are going to use is:

$$\text{polyfit}(xz, 4) = -5.368e\text{-}5\ x^4 + 0.038\ x^3 - 3.479\ x^2 + 105.811\ x - 916.689$$

$$\text{replace symbols}\left(\text{polyfit}(xz, 4), x, 45\right) = 2.601$$

The average number of births per 1000 women of age 45 is 2.6.

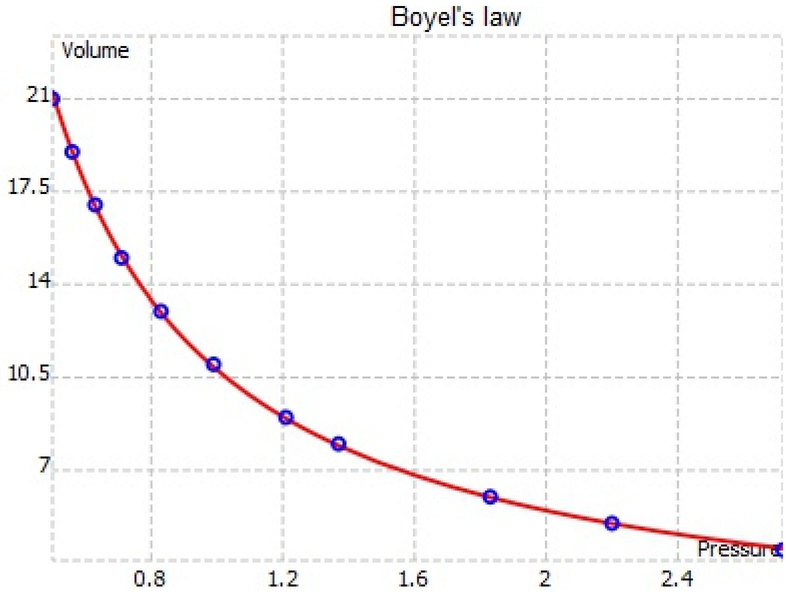
Boyel's Law - Power Regression

To deduce the connection between the pressure and volume of a gas, we will fit a regression function to the data from the following table.

Pressure	Volume
0,5	21
0,56	19
0,63	17
0,71	15
0,83	13
0,99	11
1,21	9
1,37	8
1,83	6
2,2	5
2,72	4

Doing so, we will empirically prove that Boyel's law, which describes how the pressure of a gas tends to increase as the volume of the container decreases.

$$xt := \begin{bmatrix} 0.5 & 0.56 & 0.63 & 0.71 & 0.83 & 0.99 & 1.21 \\ 21 & 19 & 17 & 15 & 13 & 11 & 9 \end{bmatrix}$$



regression(xt , "power") =

	Value
Reg. formula	$10.793\ x^{-0.977197}$
RMSE	0.092
R-sq	1
R-sq(adj)	1

The regression formula can be approximated as

$$y = 10.793\ x^{-1}$$

and this is an inverse variation.

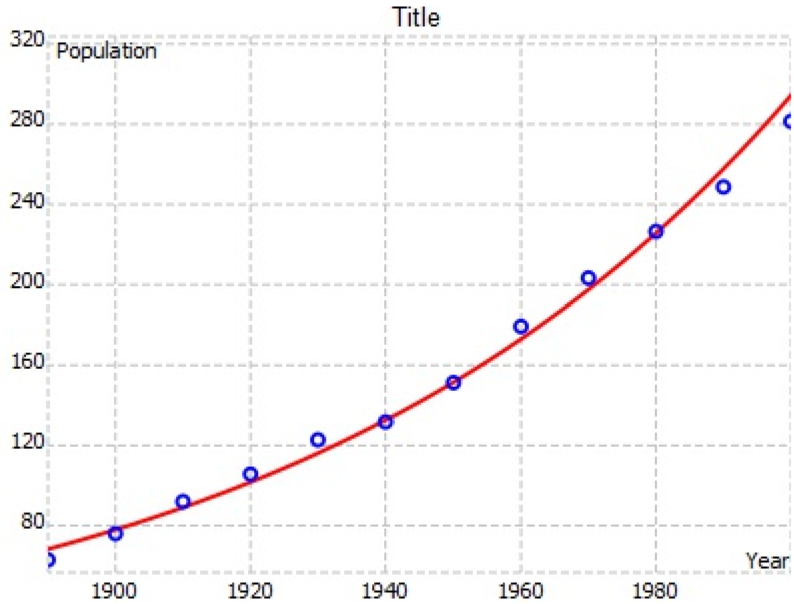
This approximation shows us that the pressure is inversely proportional to the volume of gas at constant pressure, which is the generalization of Boyle's law.

Exponential regression - Prediction

Use data of the population of the US from the table below to predict the population for the year 2010. Compare then with the actual 2010 population of which is approximately 308 million.

Year	Population (millions)
1890	62,9
1900	76
1910	92
1920	105,7
1930	122,8
1940	131,7
1950	151,3
1960	179,3
1970	203,3
1980	226,5
1990	248,7
2000	308,7

$$xu := \begin{bmatrix} 1890 & 1900 & 1910 & 1920 & 1930 & 1940 & 1950 \\ 62.9 & 76 & 92 & 105.7 & 122.8 & 131.7 & 151.3 \end{bmatrix}$$



Exponential regression formula

$$expfit(xu) = 8.466e\text{-}10 \cdot e^{0.013\ x}$$

Exponential model estimates the 2010 population would be 336 million, an overestimate of approximately 28 million people.

To predict the USA population based on the exponential regression from historical data, we will use x = 2010 in the regression formula.

$$\text{Prediction} := \text{replace symbols}\left(\text{expfit}(xu)\ ,\ x\ ,\ 2010\right)$$

$$\text{Prediction} = 336.174$$