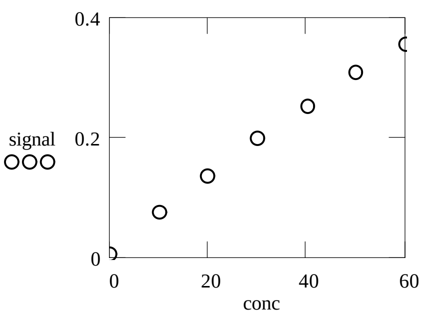
Nonlinear Curve Example

$\text{conc} := (0\ \ 10.0\ \ 20.0\ \ 30.0\ \ 40.0\ \ 50.0\ \ 60.0)^T$ $\quad \text{signal} := (.006\ \ .077\ \ .138\ \ .199\ \ .253\ \ .309\ \ .356)^T$

Let's plot the calibration data



Are these data well described by a line? They look a little curved. Still, let's try a line first, and then examine the residuals.

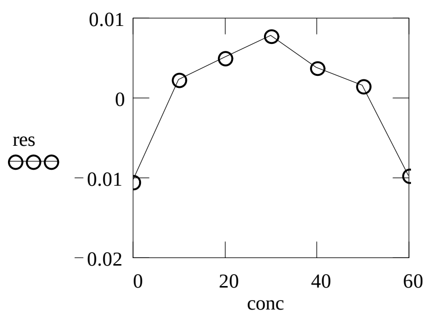
$b_0 := \text{intercept}(\text{conc}, \text{signal}) \qquad b_1 := \text{slope}(\text{conc}, \text{signal})$

$\text{fit} := b_1 \cdot \text{conc} + b_0 \qquad \text{res} := \text{signal} - \text{fit}$

$s_{res} := \sqrt{\frac{1}{5} \cdot \sum \overrightarrow{\text{res}^2}} \qquad MSS_{res} := s_{res}^{\ 2} \qquad MSS_{tot} := \text{Var}(\text{signal})$

Both the "adjusted" R-squared and std dev of the residuals help compare model fit.

$R_{adj} := 1 - \frac{MSS_{res}}{MSS_{tot}} \qquad R_{adj} = 0.9960 \qquad s_{res} = 7.9242 \cdot 10^{-3}$



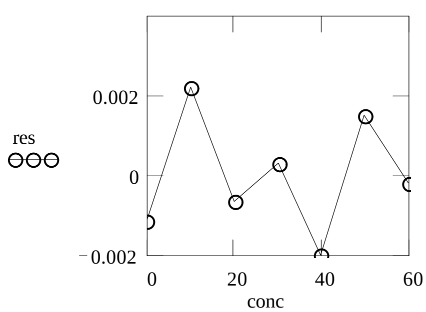
Definitely some structure to the residuals! Let's try a second-order polynomial fit and see if we can do a little better.

$\text{vs} := \text{regress}(\text{conc}, \text{signal}, 2) \qquad b := \text{submatrix}(\text{vs}, 3, 5, 0, 0)$

$b^T = \left[\ 7.1429 \cdot 10^{-3} \quad 6.9536 \cdot 10^{-3} \quad -1.8929 \cdot 10^{-5}\ \right]$

these are the LS estimates of the 2nd-order polynomial coefficients

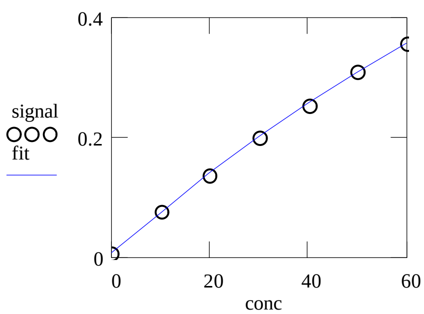
$\text{fit} := \overrightarrow{b_2 \cdot \text{conc}^2 + b_1 \cdot \text{conc} + b_0} \qquad \text{res} := \text{signal} - \text{fit} \qquad s_{res} := \sqrt{\frac{1}{4} \cdot \sum \overrightarrow{\text{res}^2}} \qquad s_{res} = 1.8028 \cdot 10^{-3}$



Residuals look more random now, which is good. Let's calculate the adjusted R-squared value to see if it improved.

$MSS_{res} := s_{res}^{\ 2}$

$R_{adj} := 1 - \frac{MSS_{res}}{MSS_{tot}} \qquad R_{adj} = 0.9998$ Yep, better than before,
as is s_{res} (above)



The fit now looks pretty good.

We will use the second-order polynomial fit as the calibration curve function - it will give better prediction capabitility than a best-fit line.

Now let's obtain a point-estimate for the tap water sample, which requires solving the following quadratic equation:

$$b_2 \cdot x^2 + b_1 \cdot x + b_0 = y_u$$

$y_u := 0.278$ $\quad x_u := \text{polyroots}\left(\left[b_0 - y_u \;\; b_1 \;\; b_2\right]^T\right)^T$

$x_u = [\,44.2926 \quad 323.0659\,]$ $\quad x_u := x_{u_0}$

These are the roots.
The first root looks like our answer.

Now we must find the standard error of the point estimate. If the calibration function is not too curved, the following expression gives a reasonable approximation to the standard error.

$S_{xx} := 6 \cdot \text{Var}(conc)$ $\quad S_{xx} = 2.8000 \bullet 10^3$ $\quad xbar := \text{mean}(conc)$ $\quad xbar = 30.0000$

$$se_u := \frac{s_{res}}{2 \cdot b_2 \cdot x_u + b_1} \cdot \sqrt{1 + \frac{1}{7} + \frac{(x_u - xbar)^2}{S_{xx}}}$$

$se_u = 0.3767$

The value $2b_2x_u+b_1$ is the slope of the curve at $x_u = 44.2926$

$t := qt(.975, 4)$ $\quad t = 2.7764$ $\quad t \cdot se_u = 1.0459$ $\quad$ Width of 95% confidence interval.

Thus, **the concentration of lead in tap water is 44.3 ± 1.0 ppb**.