Chemometrics

MA4605

Week 7. Lecture 13. Prediction Intervals for Regression

October 17, 2011

Calculation of concentration and its error

- Assume n known concentrations x_i -s have been used to determine the regression line connecting concentrations (x) with instrumentation readings (y).
- A new measurement is made of unknown concentration and it is determined through the regression line that it corresponds to x_0 concentration.
- What is the error of this determination?

$$s_{x_0} = \frac{s_{y/x}}{b}\sqrt{1 + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum_i (x_i - \bar{x})^2}}$$

More determinations of the same concentration and its error:

$$s_{x_0} = \frac{s_{y/x}}{b}\sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \overline{y})^2}{b^2 \sum_i (x_i - \overline{x})^2}}$$

where y_0 is the mean of the m separate readings for y.

More determinations of the same concentration and its error:

$$s_{x_0} = \frac{s_{y/x}}{b}\sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \overline{y})^2}{b^2 \sum_i (x_i - \overline{x})^2}}$$

where y_0 is the mean of the m separate readings for y.
Standard aqueous solutions of fluorescein are examined in a fluorescence spectrometer, and yield the following fluorescence intensities:

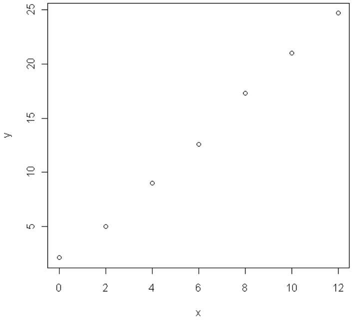
Fluorescence intensities	2.1	5.0	9.0	12.6	17.3	21.0	24.7
Concentration	0	2	4	6	8	10	12

Determine x_0 and s_{x_0} for solutions with fluorescence of 2.9, 13.5 and 23.0 units.

$> y <- c(2.1, 5.0, 9.0, 12.6, 17.3, 21.0, 24.7)$
$> x <- c(0, 2, 4, 6, 8, 10, 12)$

Obtain the calibration plot

$>$ plot(x,y)



Fitting the calibration line:

$>$ model$<-$ lm(y $\sim$ x)

$>$ summary(model) obtains the results

Call:

lm(formula = y $\sim$ x)

Residuals:

1	2	3	4	5	6	7
0.58214	-0.37857	-0.23929	-0.50000	0.33929	0.17857	0.01786

Coefficients:

	Estimate	Std. Error	t value	$\Pr(>\|t\|)$	
(Intercept)	1.5179	0.2949	5.146	0.00363	**
x	1.9304	0.0409	47.197	8.07e-08	***

Residual standard error: 0.4328 on 5 degrees of freedom

Multiple R-squared: 0.9978, Adjusted R-squared: 0.9973

F-statistic: 2228 on 1 and 5 DF, p-value: 8.066e-08

$s_{y/x}$= 0.4328, line **y = 1.5179 + 1.9304 x** , a = 1.5179, b = 1.9304

Calculation of x-determination error in R

$>$ model<-lm(y~ x) Regression analysis
$>$ bary=mean(y) Mean of y
$>$ barx=mean(x) Mean of x
$>$ n=length(y) Sample size
$>$ coef(model) Coefficients of regression
$>$ a=coef(model)[1] Intercept
$>$ b=coef(model)[2] Slope
$>$ sres=0.4328 Random errors in the y-direction
$>$ $y0 = 2.9$
$>$ $x0 = (y0 - a)/b$ Point estimate for $x0$
$>$ ssx= sum((x-barx)*(x-barx))
$>$ Sx0=(sres/b)* sqrt(1+1/n+(y0-bary)*(y0-bary)/(b*b*ssx)) Standard dev
$>$ error <- qt(0.975,n-2)* Sx0 Error of the confidence interval
$>$ x0-error Lower limit of CI
$>$ x0+error Upper limit of CI

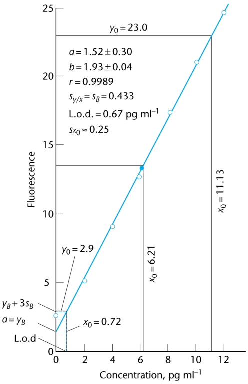
Dependence of confidence interval on y_0

- If $y_0 = 2.9$, the 95% CI for x_0 is $[0.7160037 \pm 0.6800233] = [0.03598042, 1.396027]$.
- If $y_0 = 13.5$, the 95% CI for x_0 is $[6.207216 \pm 0.6162399] = [5.590976, 6.823455]$.
- If $y_0 = 2.9$, the 95% CI for x_0 is $[11.12858 \pm 0.6764852] = [10.4521, 11.80507]$.

Limits of detections

- Set the blank signal, i.e. the concentration x = 0, in the fitted regression model.
- Compute from fitted regression the corresponding response, i.e. y = a.
- Add to *a* three times estimated standard error of y, i.e. a + $3 \cdot s_{y/x}$
- Any measurement below the above value is questionable in representing a positive concentration.
- The corresponding value x0 is the lowest value of concentration that can be trusted as non-zero.

Limits of detections



The limit of detections for the fluorescence example

- Compute the blank signal(obtained for x=0): y = a= 1.517857
- Compute y_0= a + 3$\cdot s_{y/x}$ = 1.517857 + 3(0.4328)= 2.816257
- Any fluorescence intensities below 2.816257 are questionable.
- The corresponding value x_0 is the lowest value of concentration that can be trusted as non-zero.
- $x_0 = \frac{y_0 - a}{b} = \frac{2.816257 - 1.517857}{1.930357}$ =0.6726216
- The limit of detection for the concentration is 0.6726216