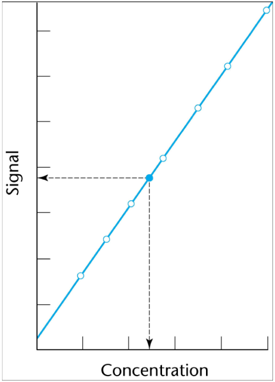
Chemometrics

MA4605

Week 6. Lecture 11. Calibration methods

October 10, 2011

- Most analyses are now performed by instrumental methods.
  Instrumental methods can perform analyses that are difficult or impossible by classical methods.
- The usual procedure: the equipment will take a measured volume of a sample, dilute it appropriately, conduct one or more reactions on it, and determine the concentration of analyte produced in the reactions.
- The concentration of analyte in the sample is *known* and it is used to determine the calibration graph.



- Take a calibration sample with known but different concentrations.
- Based on measurements plot response curve.
- Is it linear?
- Make prediction for concentration between calibrated points.

Questions

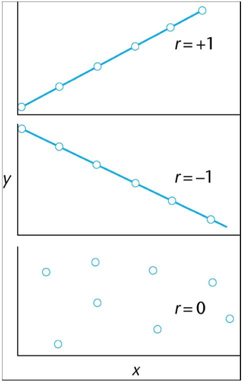
- Is the calibration graph linear?
- What is the best straight line fitting the data?
- What are the errors and confidence limits?

The correlation coefficient

Product-moment correlation coefficient

$$r = \frac{\sum_i\{(x_i - \overline{x})(y_i - \overline{y})\}}{\{[\sum_i(x_i - \overline{x})^2] \cdot [\sum_i(y_i - \overline{y})^2]\}^{\frac{1}{2}}}$$

Measure of linearity



- It can be shown that the correlation coefficient satisfies $-1 < r < +1$ and when $|r| \approx 1$ then the relation is close to linear.

Example 5.3.1

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 the correlation coefficient, r.

$$
\begin{aligned}
r &= \frac{\sum_i\{(x_i - \overline{x})(y_i - \overline{y})\}}{\{[\sum_i (x_i - \overline{x})^2] \cdot [\sum_i (y_i - \overline{y})^2]\}^{\frac{1}{2}}} \\
&= \frac{216.2}{\sqrt{112 \cdot 418.28}} = \frac{216.2}{216.44} = 0.9989
\end{aligned}
$$

Computations in R

Intensities (Y)$<-$ c(2.1,5.0,9.0,12.6,17.3,21.0,24.7)
Concentration (X)$<-$c(0,2,4,6,8,10,12)
The function in R that calculates the correlation coefficient is:
cor(x,y)=cor(y,x)

```
> cor(Concentration ,Intensities)
[1] 0.9988796
```

In analytical practice calibration graphs frequently give r-values greater than 0.99.

Test r for significance

H_0 : correlation is zero.

The test statistics $t = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{|0.9988796|\sqrt{7-2}}{\sqrt{1-0.9988796^2}} = 47.19669$

follows a *t*-dist with *n-2*=5 df.

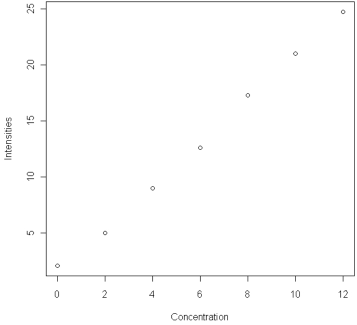
We can test for significance this value using the **cor.test( )**.

```
> cor.test(Concentration ,Intensities)
Pearson's product-moment correlation
data: Conc and Int
t = 47.1967, df = 5, p-value = 8.066e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.992073 0.999842
sample estimates:
cor
0.9988796
```

Regression plot

The regression plot can be obtained in R using:

> plot(Concentration ,Intensities)



We shall assume that the straight-line calibration graph takes the algebraic form:

$y = \alpha + \beta \cdot x$

It can be shown that the estimates for α and β in the above equation, are given by:

Slope of least squares line: b= $\frac{\sum_i[(x_i - \overline{x}) \cdot (y_i - \overline{y})]}{\sum_i (x_i - \overline{x})^2}$

Intercept of least squares line: a= $\overline{y} - b\overline{x}$

Calculate the slope and the intercept of the regression line for the data given in the previous example.

b= $\frac{\sum_i [(x_i - \overline{x}) \cdot (y_i - \overline{y})]}{\sum_i (x_i - \overline{x})^2} = \frac{216.2}{112}$=1.93

a= $\overline{y} - b\overline{x}$=13.1-(1.93)6=13.1-11.58=1.52

The regression estimates can be obtained in R using lm(y $\sim$ x):

```
> summary(lm(Intensities ~ Concentration))
```

Call:
lm(formula = Int $\sim$ Conc)
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	**
Conc	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