Linear regression as a calibration method

Krzysztof Podgórski
Department of Mathematics and Statistics
University of Limerick

November 10, 2009

Calculation of concentration and its error

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).

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.

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?

ing approximate formula:

$$s_{x_0} = \frac{s_{y/x}}{b}\sqrt{1 + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum_i (x_i - \bar{x})^2}} \quad (5$$

More determinations of the same concentration and its error, example

$$s_{x_0} = \frac{s_{y/x}}{b}\sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum_i (x_i - \bar{x})^2}} \qquad (5.10)$$

As expected, equation (5.10) reduces to equation (5.9) if $m = 1$. As always, confidence limits can be calculated as $x_0 \pm t_{(n-2)} s_{x_0}$, with $(n - 2)$ degrees of freedom. Again, a simple computer program will perform all these calculations, but most calculators will not be adequate.

Example 5.6.1

Using the data from the Section 5.3 example, determine x_0 and s_{x_0} values and x_0 confidence limits for solutions with fluorescence intensities of 2.9, 13.5 and 23.0 units.

The x_0 values are easily calculated by using the regression equation determined in Section 5.4, $y = 1.93x + 1.52$. Substituting the y_0-values 2.9, 13.5 and 23.0, we obtain x_0-values of 0.72, 6.21 and 11.13 pg ml^{-1} respectively.

To obtain the s_{x_0}-values corresponding to these x_0-values we use equation (5.9), recalling from the preceding sections that $n = 7$, $b = 1.93$, $s_{y/x} = 0.4329$, $\bar{y} = 13.1$, and $\sum_i (x_i - \bar{x})^2 = 112$. The y_0 values 2.9, 13.5 and 23.0 then yield s_{x_0}-values of 0.26, 0.24 and 0.26 respectively. The corresponding 95% confidence limits ($t_5 = 2.57$) are 0.72 ± 0.68, 6.21 ± 0.62, and 11.13 ± 0.68 pg ml^{-1} respectively.

Computations of the regression line in **R**

```
Int=c(2.1,5.0,9.0,12.6,17.3,21.0,24.7)
Conc=c(0,2,4,6,8,10,12)
plot(Conc, Int)

#fitting the line
fit=lm(Int~Conc)
coef(fit)
#plotting it
abline(coef(fit))
```

Interpretation of regression results in **R**

```
Call:
lm(formula = Int ~ 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 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1 

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 $y = 1.9304x + 1.5179$, $a = 1.5179$, $b = 1.9304$

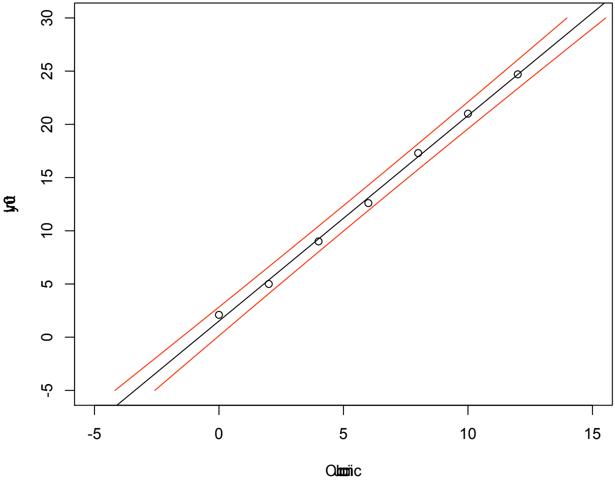
Calculation of *x*-determination error in **R**

```
ra=lm(Int~Conc)  #Regression analysis
bary=mean(Int)   #Means
barx=mean(Conc)
n=length(Int)

coef(ra)         #Coefficients of regression

a=coef(ra)[1]
b=coef(ra)[2]

residuals(ra)    #Residuals

sres=0.4328
 y_0=c(2.9,13.5,23.0)
 x_0=(y_0-a)/b
 ssx=sum((Conc-barx)^2)
 Sx_0=(sres/b)*sqrt(1+1/n+(y_0-bary)^2/(b^2*ssx)) #Standard dev

 tn_2=qt(0.975,n-2)
 x_0-tn_2*Sx_0 #Confidence intervals
 x_0+tn_2*Sx_0
```

Dependence of confidence interval on y_0

```
plot(Conc, Int, xlim=c(-5,15),ylim=c(-5,30))
                %Plotting within specified ranges

fit=lm(Int~Conc)
coef(fit)

abline(coef(fit))

y0=seq(-5,30,by=0.1)
x0=(y0-a)/b
Sx0=(sres/b)*sqrt(1+1/n+(y0-bary)^2/(b^2*ssx))
lci=x0-tn_2*Sx0 %lower bound
uci=x0+tn_2*Sx0 %upper bound

par(new=T) %Preserving previous graph to plot upper and lower limits

plot(lci,y0,xlim=c(-5,15),ylim=c(-5,30),col="red",type="l",axes=F)
par(new=T)
plot(uci,y0,xlim=c(-5,15),ylim=c(-5,30),col="red",type="l",axes=F)
```

Plot of confidence intervals



Limits of detections

Limits of detections

- Set the blank signal, i.e. the concentration $x = 0$, in the fitted regression model.

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. $y = a$.

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. $3s_{y/x}$:

$$a + 3s_{y/x}.$$

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. $3s_{y/x}$:
$$a + 3s_{y/x}.$$
- Any measurement below the above value is questionable in representing a positive concentration.

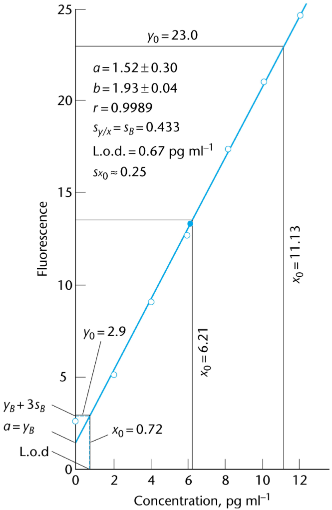
Limits of detections

- 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. $3s_{y/x}$:

$$a + 3s_{y/x}.$$

- Any measurement below the above value is questionable in representing a positive concentration.
- The corresponding value x_0 is the lowest value of concentration that can be trusted as non-zero.

Example 5.7.1



$y_0 = 23.0$
$a = 1.52 \pm 0.30$
$b = 1.93 \pm 0.04$
$r = 0.9989$
$s_{y/x} = s_B = 0.433$
L.o.d. = 0.67 pg ml^{-1}
$s_{x_0} \approx 0.25$
$x_0 = 11.13$
$x_0 = 6.21$
$y_0 = 2.9$
$y_B + 3s_B$
$a = y_B$
$x_0 = 0.72$
L.o.d
Fluorescence
Concentration, pg ml^{-1}
0
5
10
15
20
25
2
4
6
8
10
12

Application of regression: The method of standard additions

Application of regression: The method of standard additions

- It is assumed that the regression line passes through the origin

$$y = bx.$$

Application of regression: The method of standard additions

- It is assumed that the regression line passes through the origin
$$y = bx.$$

- Unknown concentration x_0 is suppose to be detected.

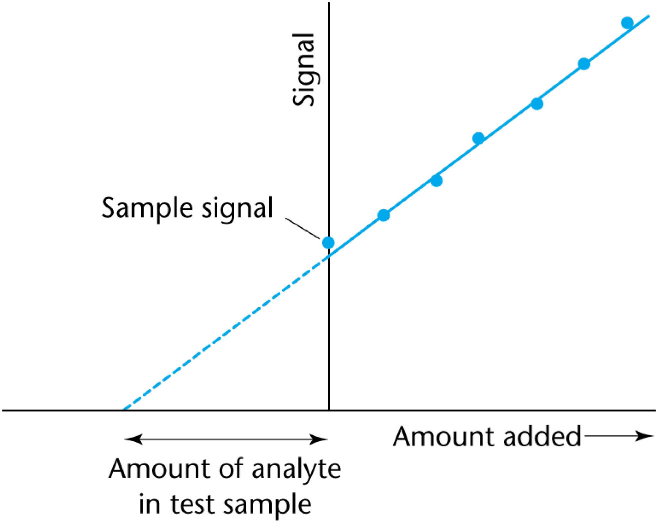
Application of regression: The method of standard additions

- It is assumed that the regression line passes through the origin
$$y = bx.$$
- Unknown concentration x_0 is suppose to be detected.
- Some known concentrations x_i are added to subsamples of the tested sample
$$y_i = bx_i + bx_0$$

Application of regression: The method of standard additions

- It is assumed that the regression line passes through the origin
$$y = bx.$$
- Unknown concentration x_0 is suppose to be detected.
- Some known concentrations x_i are added to subsamples of the tested sample
$$y_i = bx_i + bx_0$$
- The searched value is obtained by taking the intercept of the model (bx_0) and dividing it by the slope b.

Graphical illustration of the method



Standard deviation and error distribution

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

and the error of determination when normalized by s_{x_E} has Student t-distribution with $n-2$ degrees of freedom.

Example 5.8.1

The silver concentration in a sample of photographic waste was determined by atomic-absorption spectrometry with the method of standard additions. The following results were obtained.

Added Ag: µg added per ml of *original* sample solution	0	5	10	15	20	25	30
Absorbance	0.32	0.41	0.52	0.60	0.70	0.77	0.89

Determine the concentration of silver in the sample, and obtain 95% confidence limits for this concentration.

Equations (5.4) and (5.5) yield $a = 0.3218$ and $b = 0.0186$. The ratio of these figures gives the silver concentration in the test sample as 17.3 µg ml^{-1}. The confidence limits for this result can be determined with the aid of equation (5.13). Here $s_{y/x}$ is 0.01094, $\bar{y} = 0.6014$, and $\sum_i (x_i - \bar{x})^2 = 700$. The value of s_{x_E} is thus 0.749 and the confidence limits are $17.3 \pm 2.57 \times 0.749$, i.e. 17.3 ± 1.9 µg ml^{-1}.

Computations in **R**

```
aa=c(0,5,10,15,20,25,30)
ab=c(0.32,0.41,0.52,0.60,0.70,0.77,0.89)
n=length(ab)
barx=mean(aa)
bary=mean(ab)

regr=lm(ab~aa)
a=coef(regr)[1]
b=coef(regr)[2]

summary(regr)

sres=0.01092

ss_x=sum((aa-barx)^2)
sx_e=(sres/b)*sqrt(1/n + bary^2/(b^2*ss_x))
tn_2=qt(0.975,n-2)
 a/b-tn_2*sx_e #Confidence intervals
 a/b+tn_2*sx_e
```

Using regression for comparing analytical methods

Using regression for comparing analytical methods

- Two methods of the determination of an analyte: an old reliable one and a new one that is examined for accuracy.

Using regression for comparing analytical methods

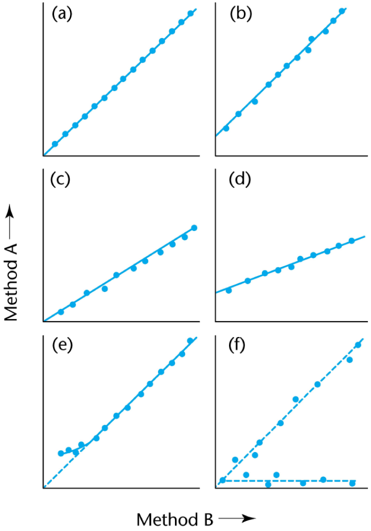
- Two methods of the determination of an analyte: an old reliable one and a new one that is examined for accuracy.
- For several samples each is analysed by the two methods and x-values are assigned to the old method and y-values to the new one.

Using regression for comparing analytical methods

- Two methods of the determination of an analyte: an old reliable one and a new one that is examined for accuracy.
- For several samples each is analysed by the two methods and x-values are assigned to the old method and y-values to the new one.
- Regression analysis is performed of y's on x's in order to detect any significant deviation from the $y = x$ relation.

Using regression for comparing analytical methods

- Two methods of the determination of an analyte: an old reliable one and a new one that is examined for accuracy.
- For several samples each is analysed by the two methods and x-values are assigned to the old method and y-values to the new one.
- Regression analysis is performed of y's on x's in order to detect any significant deviation from the $y = x$ relation.



Example 5.9.1

Example 5.9.1

- The level of phytic acid in 20 urine samples is measured by a new catalytic fluorimetic (CF) method and compared with those obtained using an established extraction photometric (EP) technique.

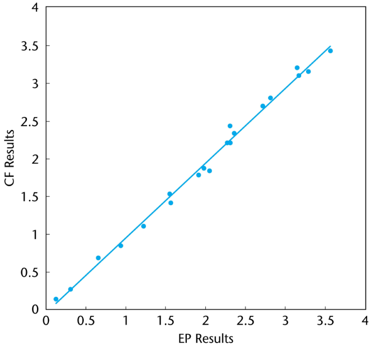
Example 5.9.1

- The level of phytic acid in 20 urine samples is measured by a new catalytic fluorimetic (CF) method and compared with those obtained using an established extraction photometric (EP) technique.
- These are the results

```
CF=c(1.87,2.20,3.15,3.42,1.10,1.41,1.84,0.68,0.27,2.80,0.14,
         3.20,2.70,2.43,1.78,1.53,0.84,2.21,3.10,2.34)
EP=c(1.98,2.31,3.29,3.56,1.23,1.57,2.05,0.66,0.31,2.92,0.13,
         3.15,2.72,2.31,1.92,1.56,0.94,2.27,3.17,2.36)
```

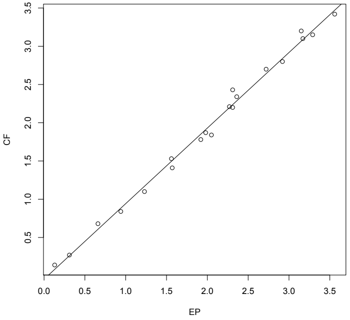
Example 5.9.1

- The level of phytic acid in 20 urine samples is measured by a new catalytic fluorimetic (CF) method and compared with those obtained using an established extraction photometric (EP) technique.
- These are the results

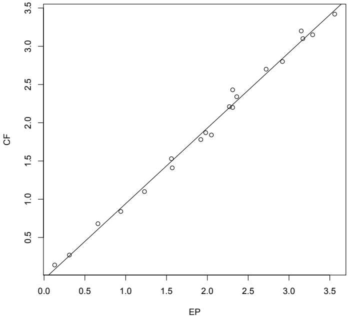


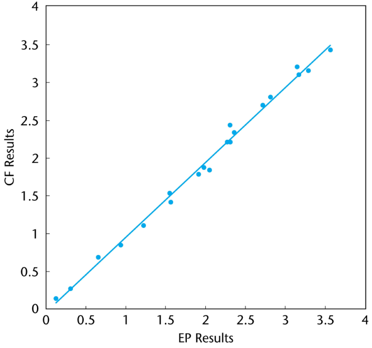

```
CF=c(1.87,2.20,3.15,3.42,1.10,1.41,1.84,0.68,0.27,2.80,0.14,
           3.20,2.70,2.43,1.78,1.53,0.84,2.21,3.10,2.34)
EP=c(1.98,2.31,3.29,3.56,1.23,1.57,2.05,0.66,0.31,2.92,0.13,
           3.15,2.72,2.31,1.92,1.56,0.94,2.27,3.17,2.36)
```

Comparison figures



Comparison figures





Analysis

```
plot(EP,CF)
regr=lm(CF~EP)
abline(coef(regr))
locator(1) #Locates a point that differs
           #from the graph in the textbook
regr
anova(regr)
summary(regr)
```

Comparison of reports

```
regr
Call:
lm(formula = CF ~ EP)
Coefficients:
(Intercept)           EP
   -0.04563      0.98794

anova(regr)
Analysis of Variance Table
Response: CF
          Df  Sum Sq Mean Sq F value    Pr(>F)
EP         1 18.3424 18.3424    2696 < 2.2e-16 ***
Residuals 18  0.1225  0.0068
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Comparison of reports cont

```
 summary(regr)

Call:
lm(formula = CF ~ EP)

Residuals:
     Min       1Q   Median       3Q      Max
-0.13964 -0.05224 -0.01357  0.05487  0.19349

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.04563    0.04264   -1.07    0.299
EP           0.98794    0.01903   51.92   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.08248 on 18 degrees of freedom
Multiple R-squared: 0.9934, Adjusted R-squared: 0.993
F-statistic:  2696 on 1 and 18 DF,  p-value: < 2.2e-16
```

Errors in the slope and intercept

$$s_{y/x} = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n-2}} \qquad (5.6)$$

for the slope (b) and the intercept (a). These are given by:

Standard deviation of slope: $s_b = \dfrac{s_{y/x}}{\sqrt{\sum_i (x_i - \bar{x})^2}}$

Standard deviation of intercept: $s_a = s_{y/x}\sqrt{\dfrac{\sum_i x_i^2}{n\sum_i (x_i - \bar{x})^2}}$

Distribution of the normalized errors

The errors of estimation of the slope and intercept, when normalized by the corresponding standard deviations are distributed as Student t-distribution with $n - 2$ degrees of freedom.

Confidence intervals in the example

```
n=length(CF)

-0.04563+0.04264*qt(0.975,n)
-0.04563-0.04264*qt(0.975,n)
 0.98794+0.01903*qt(0.975,n)
 0.98794-0.01903*qt(0.975,n)
```

```
[-0.147,0.055]
```

```
[0.94,1.03]
```

Thus we can not exclude the intercept equal to zero and the slope equal to one. Two analytical methods are equivalent.