Chemometrics

MA4605

Week 7. Lecture 14. Prediction Intervals for Regression

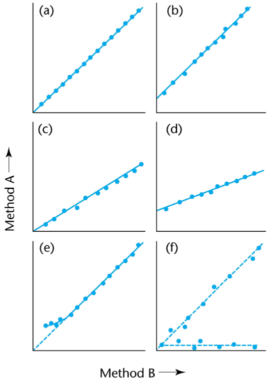
October 18, 2011

Using regression for comparing analytical methods

- New analytical methods can be validated against standard procedures.
- The validation is done by comparing two methods of determination of an analyte: an old reliable one and a new one that is examined for accuracy.
- The aim of the comparison is to identify systematic errors in the new method.
- Does the new method give results higher or lower than the established procedure?

- In cases where the analysis is repeated over a small concentration range we can use methods such as the t-test for paired samples.
- In cases where the analysis is done over large concentration ranges we use a regression line to compare the methods.
- One axis of a regression graph is used for the results obtained by the new method and the other axis for the standard method.
- How to allocate the method to the x and y axis?
- The old(standard) method is assigned to the x-axis and the new one is assigned to the y-axis.

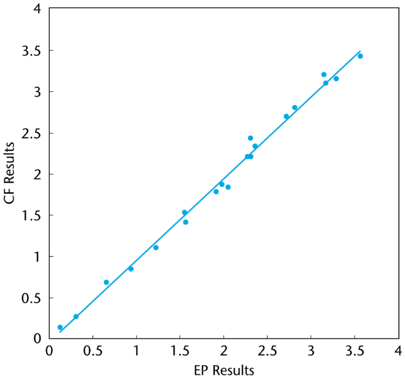
- Linear regression is then applied to calculate the slope (a) and the intercept(b) of the regression line.
- If the two methods yield similar results the the regression line will have a=0 and b=1 and correlation coefficient =1.
- The ideal situation a=0 and b=r=1. In practice this never occurs.
- Regression analysis is performed of y-s on x-s in order to detect any significant deviation from the $y = x$ relation.


(a)
(b)
(c)
(d)
(e)
(f)
Method A
Method B

In practice we want to test if the intercept differs significantly from 0 and if the slope differs significantly from 1. We can test this by calculating the confidence intervals for the slope and the intercept estimates.

Example 5.9.1 The level of phytic acid in 20 urine samples is measured by a new catalytic fluorimetric (CF) method and compared with those obtained using an established extraction photometric (EP) technique.

```
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)
```


CF Results
EP Results
0
0.5
1
1.5
2
2.5
3
3.5
4
0
0.5
1
1.5
2
2.5
3
3.5
4

$>$ model $<$- lm(CF $\sim$ EP)

$>$ summary(model)

Call:

lm(formula = CF$\sim$ EP)

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	***

The output from the *lm* function indicates that the intercept *a* is not significantly different from 0.

Test $H_0 : \beta = 1$ vs $H_a : \beta \neq 1$.

Test statistic $t = \frac{0.98794-1}{0.01903} = -0.6337362$.

The test statistic is higher than the critical value qt(0.025,18)=-2.100922, hence we accept the null hypothesis that the slope is 1.

y = -0.04563 + 0.98794 x

Calculate the 95% confidence intervals for the slope(b) and the intercept(a).

```
> confint(model)
```

	2.5 %	97.5 %
(Intercept)	-0.1352092	0.04395558
EP	0.9479627	1.02791138

The 95% CI for the intercept contains the ideal values of 0.

The 95% CI for the slope contains the ideal values of 1.