Chemometrics

MA4605

Week 8. Lecture 16. Weighted Linear regression

October 25, 2011

Weighted regression

- **Homoscedasticity** - the standard deviations of y-observations from the straight line are the same independently of the underlying x-observations.
- **Heteroscedasticity** - the standard deviations of y-observations depend on the underlying x-observations.
- In the first case, standard regression analysis should be performed, while in the second the weighted regression is more suitable.

Weights

- The analysis has to be adjusted through weights that gives more emphasize to the values with small deviations and less to the ones with larger ones.
- For this standard deviations s_i has to be given together with (x_i, y_i).
- They can be obtained through multiple measurements at a given x_i .
- Weights are inverse proportional to the variance and are defined as

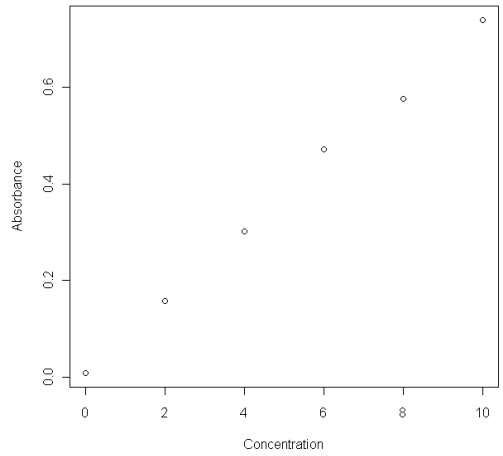
$$wi = \frac{s_i^{-2}}{\frac{\sum_{k=1}^{n} s_k^{-2}}{n}}$$

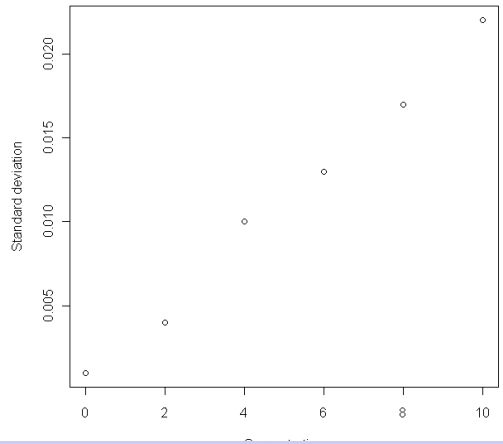
Example 5.10.1

Calculate the unweighted and the weighted regression lines for the following calibration data.

Concentration(X)	0	2	4	6	8	10
standard deviation(s_i)	0.001	0.004	0.010	0.013	0.017	0.022
Absorbance(Y)	0.009	0.158	0.301	0.472	0.577	0.739

Example 5.10.1 plots

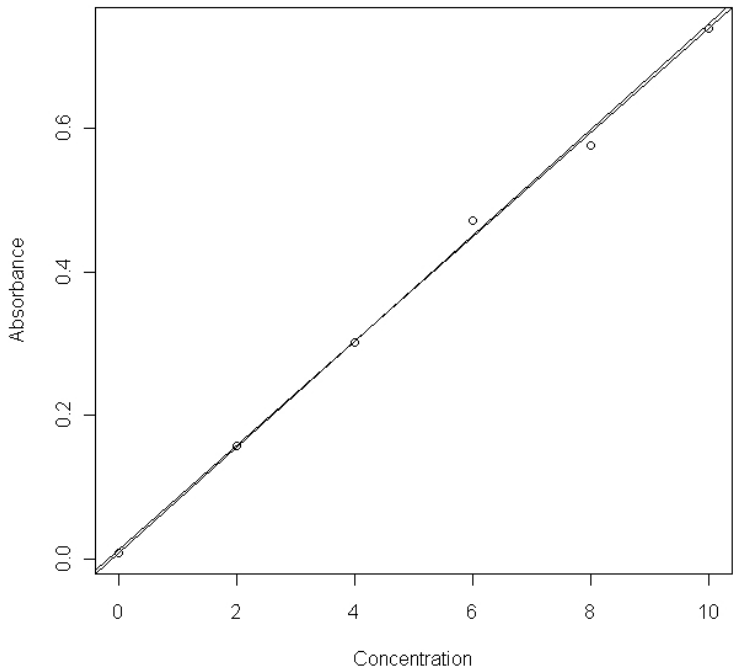




Fitting weighted linear regression in R

```
Conc <- c(0,2,4,6,8,10)
StDev <- c(0.001,0.004,0.010,0.013,0.017,0.022)
Abs <- c(0.009,0.158,0.301,0.472,0.577,0.739)
reg <- lm(Abs ~ Conc)
n <- length(Conc)
w <- StDev^(-2)/ mean(StDev^(-2))
wreg <-lm(Abs ~ Conc, weights=w)
```

Fitted unweighted and weighted linear regression



- Both approaches gives similar linear fits as expressed by slopes and and intercepts.
- They differ in error estimation.

```
> confint(reg)
                 2.5 %       97.5 %
(Intercept) -0.01603033  0.04260176
          x  0.06770147  0.07738424
> confint(wreg)
                 2.5 %       97.5 %
(Intercept)  0.00617518  0.01199264
          x  0.07080612  0.07671381
```