Lab week 9

Weighted linear regression

Exercise 1. The florescence of each of a series of acidic solutions of quinine with concentrations of 0, 10, 20, 30, 40 and 50 was determined 5 times 4.0, 21.2, 44.6, 61.8, 78.0, 105.2 .

The standard deviations of these determinations are 0.71, 0.84, 0.89, 1.64, 2.24, 3.03 .

Determine the slopes and the intercepts of the unweighted and the weighted regression lines.

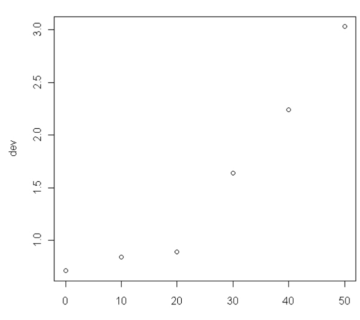
The unweighted linear regression requires that the variability in the response variable (y) is constant across the range of values of the independent variable(x). We first investigate the plot of standard deviations against the concentration values (x) to see whether the homoscedasticity assumption is true.

```
x<- seq(0,50, by=10)

y<- c(4.0, 21.2, 44.6, 61.8, 78.0, 105.2)

stdev<- c(0.71, 0.84, 0.89, 1.64, 2.24, 3.03)

plot(x,stdev)
```



The plot indicates that the standard deviations increase with the concentration values, so the assumption of constant variance is violated. Unweighted linear regression is recommended in this case.

We will analyse the linear relationship between concentration and absorbance using both weighted and unweighed regression. The unweighted linear regression does not require the values of the standard deviation, while the weighted linear regression needs the additional information about the standard deviations of the response.

```
model1<- lm(y ∼ x)
summary(model1)
```

Call:

lm(formula = y ∼ x)

Coefficients:

	Estimate	Std. Error	t value	Pr(> $|t|$)
(Intercept)	2.9238	2.1648	1.351	0.248
x	1.9817	0.0715	27.715	1.01e-05

Residual standard error: 2.991 on 4 degrees of freedom

Multiple R-squared: 0.9948, Adjusted R-squared: 0.9935

F-statistic: 768.1 on 1 and 4 DF, p-value: 1.008e-05

The fitted line obtained for the unweighted linear regression and is y=2.9238 +1.9817 x.

The weighted linear regressions requires calculating the weights according to the formula

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

This formula ensures that higher weights will be given to values with small deviations and less to the ones with larger ones.

```
w<- stdev ˆ(-2)/ mean(stdev ˆ(-2))
model2<- lm(y ∼ x, weights=w)
summary(model2)
```

Call:

lm(formula = y ∼ x, weights = w)

Coefficients:

	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	3.48296	1.16109	3.00	0.0400
x	1.96357	0.06768	29.01	8.4e-06

Residual standard error: 2.036 on 4 degrees of freedom

Multiple R-squared: 0.9953, Adjusted R-squared: 0.9941

F-statistic: 841.7 on 1 and 4 DF, p-value: 8.403e-06

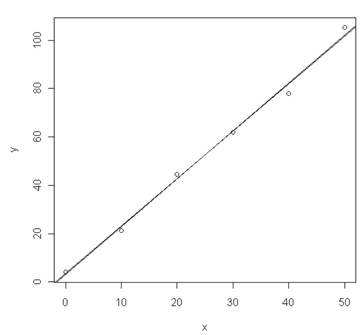
The fitted line obtained for the weighted linear regression is y= 3.48296 + 1.96357 x.

Plot the data and the two fitted lines.

```
plot(x,y)

abline(coef(model1))

abline(coef(model2))
```



Both approaches gives similar linear fits as it can be seen from the plot. The point estimates for the slope and the intercept are very close for the two methods. However, the difference consists in the error of the estimation. This can observed from the confidence intervals for the slope and the intercept that we can extract from the two models with the $confint$ function.

confint(model1)

	2.5 %	97.5 %
(Intercept)	-3.086760	8.934380
x	1.783192	2.180237

confint(model2)

	2.5 %	97.5 %
(Intercept)	0.2592681	6.706662
x	1.7756557	2.151491

The 95% confidence intervals for the slope and the intercept are smaller for model 2 suggesting that the weighted regression is a preferable.