Commands for the zinc data

The zinc data set is 91 observations with values for Concentration and Peak.Area. There are 11 distinct concentrations, with 7 to 11 replicates per concentration.

If we want to compute the mean peak area for each concentration, we can use

```
> tapply(zinc$Peak.Area,zinc$Concentration,mean)
          0          10          20         100         200         500
   264.7500    316.8571    692.4286   1291.3636   2187.8571   4109.2857
       1000        2000        5000       10000       25000
  7589.4444  14388.5714  35072.3333  70267.2000 181354.4444
```

This results in a vector of length 11 with the 11 means and with labels (names) of the concentration.

To examine the relationship between the concentration and the variance of the peak areas, we need a vector of the 11 concentrations and a vector of the 11 variances of the peak areas. We can do this with

```
> vars <- tapply(zinc$Peak,zinc$Conc,var)
           0           10           20          100          200          500         1000
4.203907e+04 1.714762e+02 7.869524e+02 1.438745e+04 1.431810e+03 8.074238e+03 5.365478e+04
 2000         5000  10000        25000
4.098095e+04 1.849742e+06 1.358443e+07 2.835566e+07
```

To get a vector of the 11 distinct concentrations, we can use

```
> concs <- unique(zinc$Conc)
 [1]     0    10    20   100   200   500  1000  2000  5000 10000 25000
```

According to theory, the variance at a given concentration should be a linear function of the square of the concentration, so we do this using

```
> concs2 <- concs^2
> var.lm <- lm(vars ~ concs2)
> summary(var.lm)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 8.298e+05  9.245e+05   0.898    0.393
concs2      4.611e-02  4.840e-03   9.525 5.36e-06 ***

Residual standard error: 2861000 on 9 degrees of freedom
Multiple R-squared:  0.9098,    Adjusted R-squared:  0.8997
F-statistic: 90.73 on 1 and 9 DF,  p-value: 5.357e-06
```

For weighted least squares, we should use weights that are the reciprocals of the variances. There are two obvious choices for weighting, which are the reciprocals of the observed variances at the given concentration or the reciprocals of the fitted variances from the regression of the variances on the square concentrations. Either choice is defensible.

The vector `vars` is of length 11 as is the vector `predict(var.lm)`. But weights for a regression with 91 observations needs to be of length 91, so we need to repeat each weight as many times as there are replicates at that concentration. Then use the `weights=` parameter of lm.

```
> concnums <- table(zinc$Conc)
    0    10    20   100   200   500  1000  2000  5000 10000 25000
    8     7     7    11     7     7     9     7     9    10     9
> wt1 <- 1/rep(vars,concnums)
> wt2 <- 1/rep(predict(var.lm),concnums)
```

The `rep()` command has many options, so consult the documentation.

```
> ?rep
> rep(1:3,2)
[1] 1 2 3 1 2 3
> rep(1:3,each=2)
[1] 1 1 2 2 3 3
> rep(1:3,c(2,3,2))
[1] 1 1 2 2 2 3 3
```