Regression examples

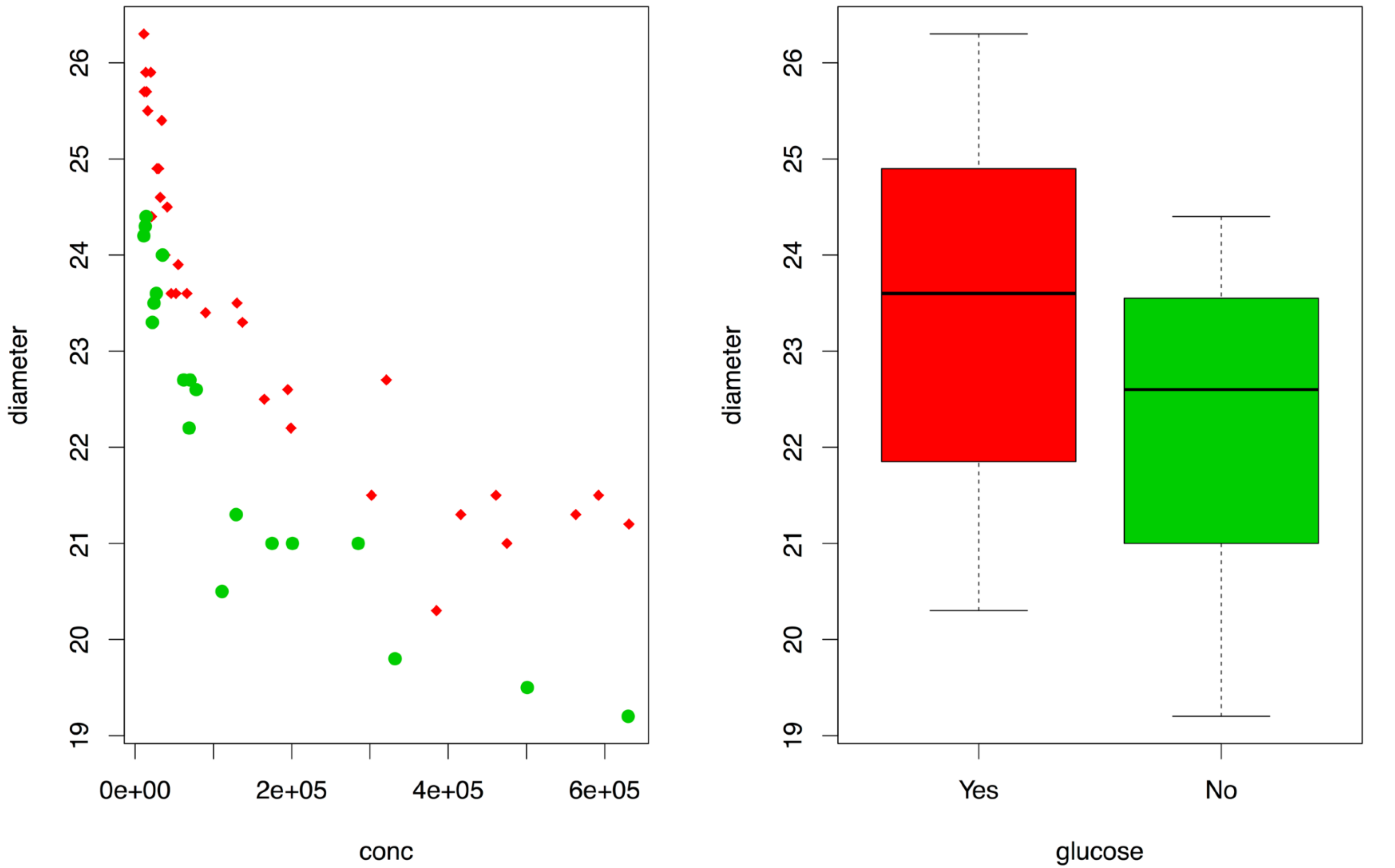
<u>Example, Hellung Data:</u> Two groups of Tetrahymena cell cultures where glucose was either added or not to the growth medium. For each culture the average cell diameter and cell concentration were recorded. It is expected that the cell diameter is affected by the presence of glucose in the medium.

```
> head(hellung)
  glucose   conc diameter
1       1 631000     21.2
2       1 592000     21.5
3       1 563000     21.3
4       1 475000     21.0
5       1 461000     21.5
6       1 416000     21.3
```

```
> is.factor(hellung$glucose)
[1] FALSE
> hellung$glucose=factor(hellung$glucose,labels=c("Yes","No"))
```

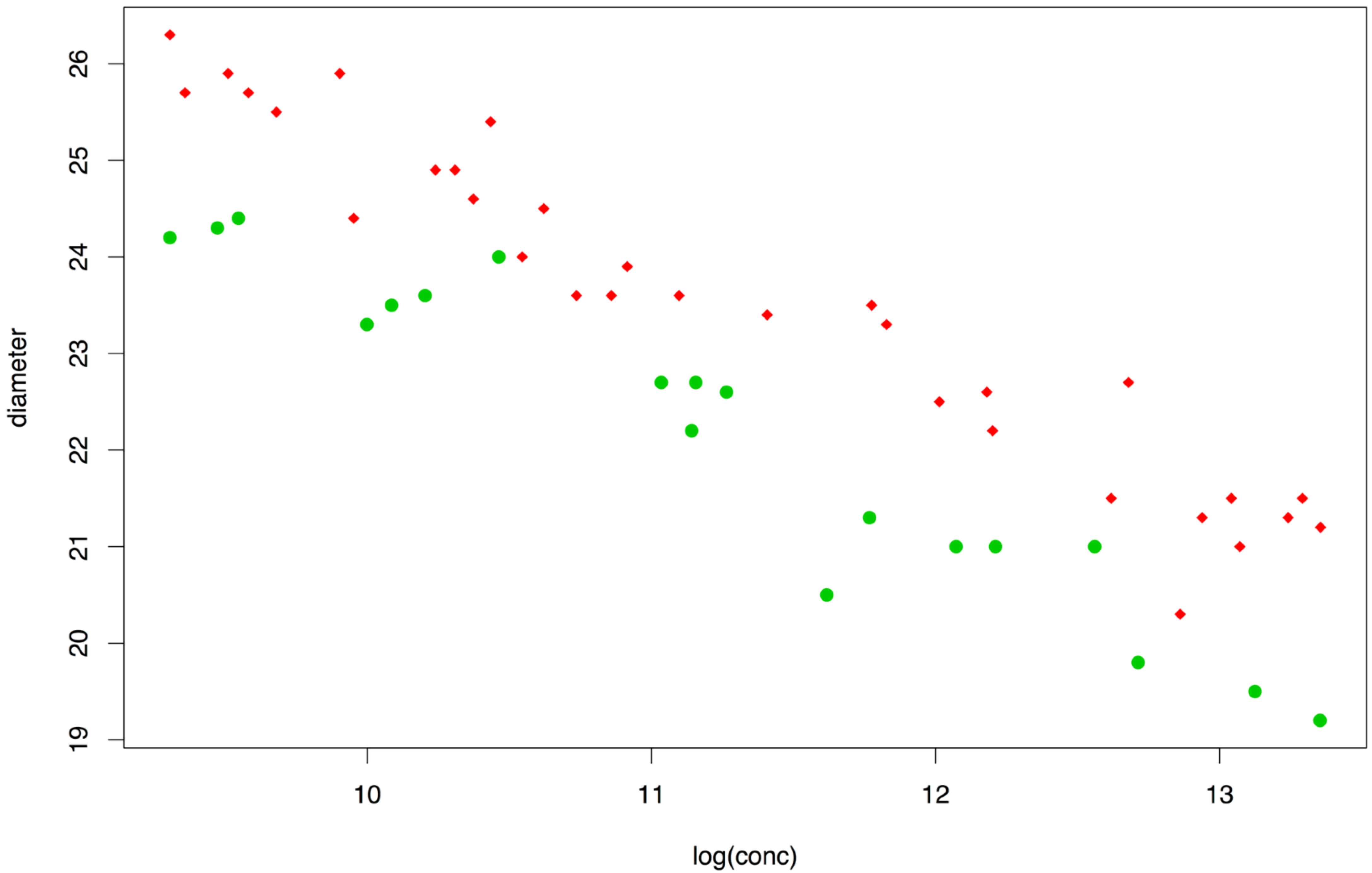
Regression examples

```
> plot(conc,diameter,pch=(as.double(glucose)+17),
+  col=(as.double(glucose)+1),lwd=2)
> boxplot(diameter~glucose,data=hellung,col=c(2,3),ylab="diameter"
+          xlab="glucose")
```



Regression examples

```
>plot(log(conc),log(diameter),pch=(as.double(glucose)+17),
+    col=(as.double(glucose)+1),lwd=2)
```



Regression examples

Possible models?

- Regression diameter ~ concentration with the same intercept and slope regardless of glucose presence

$$y_{i,j} = \alpha + \beta x_{i,j} + \epsilon_{i,j}$$

- Regression diameter ~ concentration with glucose-dependent intercept and slope

$$y_{i,j} = \alpha_i + \beta_i x_{i,j} + \epsilon_{i,j}$$

- Regression diameter ~ concentration with glucose-dependent intercept and common slope. Model with glucose-dependent slope and common intercept also possible but does it make sense?

$$y_{i,j} = \alpha_i + \beta x_{i,j} + \epsilon_{i,j}$$

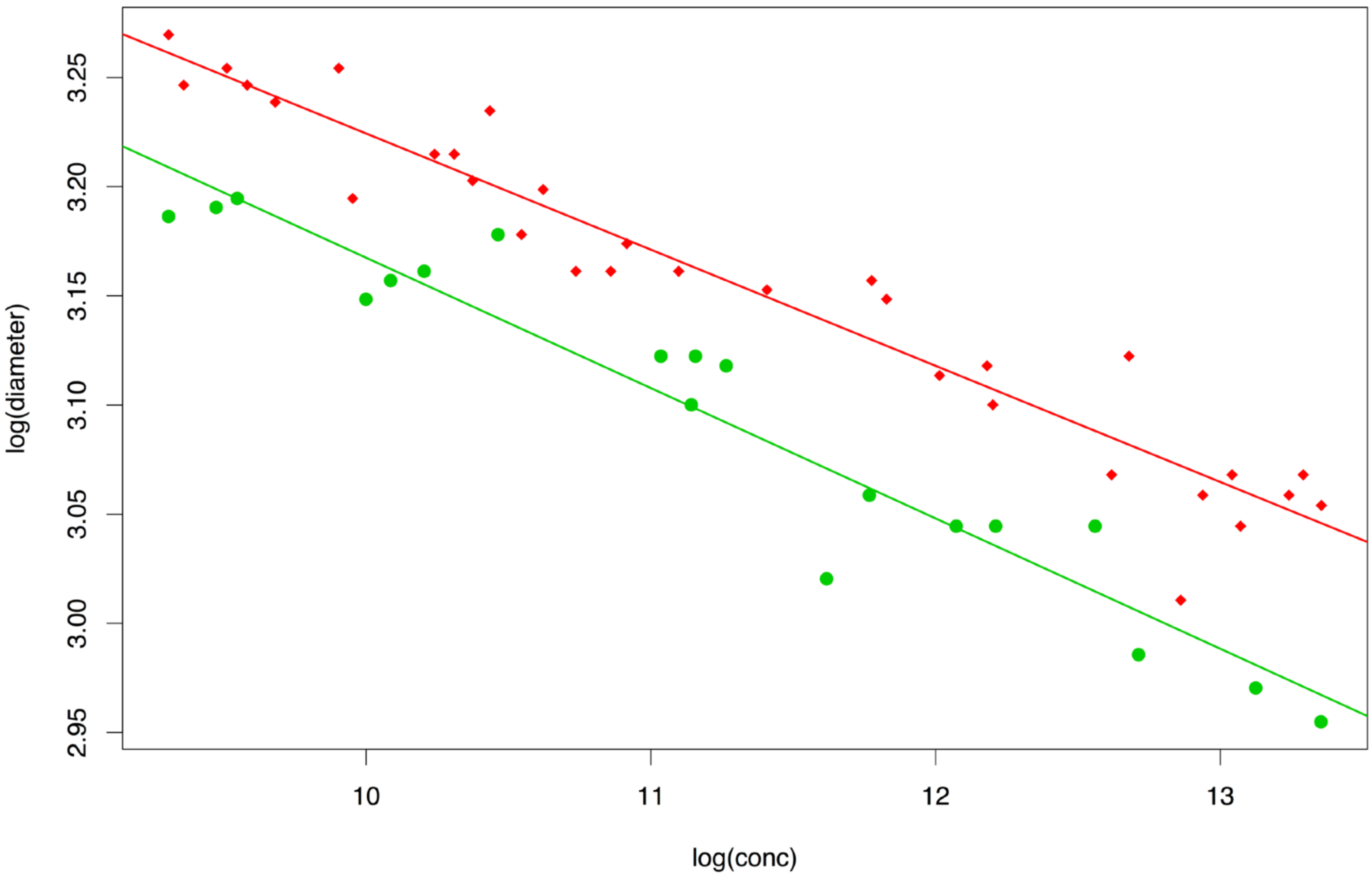
Regression examples

What about variances? σ_i^2 or σ^2

```
> summary(L1)
Call:
lm(formula = log(diameter) ~ log(conc) * glucose)

Residuals:
      Min        1Q    Median        3Q       Max
-0.061530 -0.011254  0.000129  0.008675  0.040543

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          3.756307   0.031957 117.543   <2e-16 ***
log(conc)           -0.053196   0.002807 -18.954   <2e-16 ***
glucoseNo            0.007869   0.054559   0.144    0.886
log(conc):glucoseNo -0.006480   0.004821  -1.344    0.185
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02086 on 47 degrees of freedom
Multiple R-squared:  0.9361,   Adjusted R-squared:  0.9321
F-statistic: 229.6 on 3 and 47 DF,  p-value: < 2.2e-16
```

Regression examples

```
> plot(log(conc),log(diameter),pch=c(18,19),col=2:3,lwd=2)
> abline(a=L1$coeff[1],b=L1$coeff[2],lty=1,lwd=2,col=2)
> abline(a=L1$coeff[1]+L1$coeff[3],b=L1$coeff[2]+L1$coeff[4],
+          lwd=2,col=3)
```



Regression examples

```
> L2=lm(log(diameter)~log(conc)+glucose)
> summary(L2)

Call:
lm(formula = log(diameter) ~ log(conc) + glucose)

Residuals:
      Min        1Q    Median        3Q       Max
-0.058123 -0.013201 -0.000449  0.011270  0.043550

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.781150   0.026290  143.83  < 2e-16 ***
log(conc)   -0.055393   0.002301  -24.07  < 2e-16 ***
glucoseNo   -0.065020   0.006095  -10.67 2.93e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02103 on 48 degrees of freedom
Multiple R-squared:  0.9337, Adjusted R-squared:  0.9309
F-statistic: 337.9 on 2 and 48 DF,  p-value: < 2.2e-16
```

Regression examples