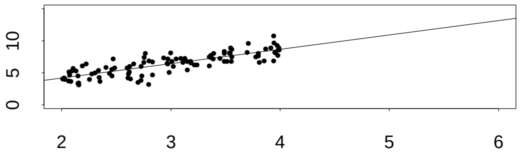
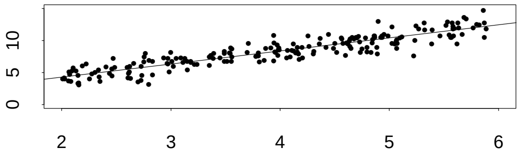
The Meaning of the Residual Standard Deviation and of r^2

Suppose we measure the height of plants grown indoors with light for different numbers of hours a day. We might get data like this:



a=-0.31 b=2.24 res.std.dev.=0.98 r-squared=0.67

Now suppose that we gather more data, for plants exposed to light for longer:

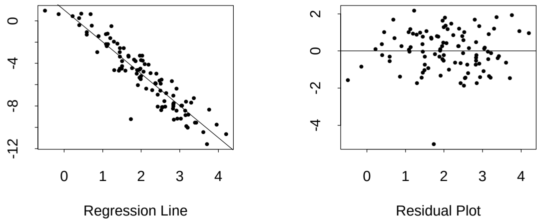


a=0.24 b=2.03 res.std.dev.=1.05 r-squared=0.83

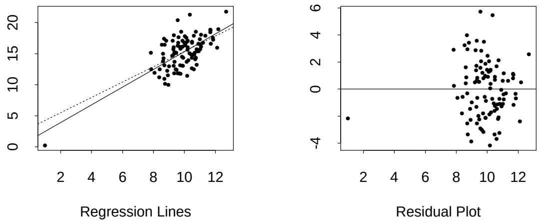
Are the changes here as you would expect?

Outliers and Influential Observations

Here one point doesn't follow the relationship:



Here one point has a big influence on the regression line:



The solid line is with all points; the dotted line is with the influential point omitted.

Factors Affecting Interpretation of a Regression Model

For experimental data:

- What was the range of the explanatory variable?
- How well-controlled were other influences?

For observational data:

- Could the association be due to "lurking' variables?
- Is the association causal?

For all data:

- Might the relationship be different for different sub-groups?