Transcript of Mick Crawley's R course 2010 Imperial College London, Silwood Park

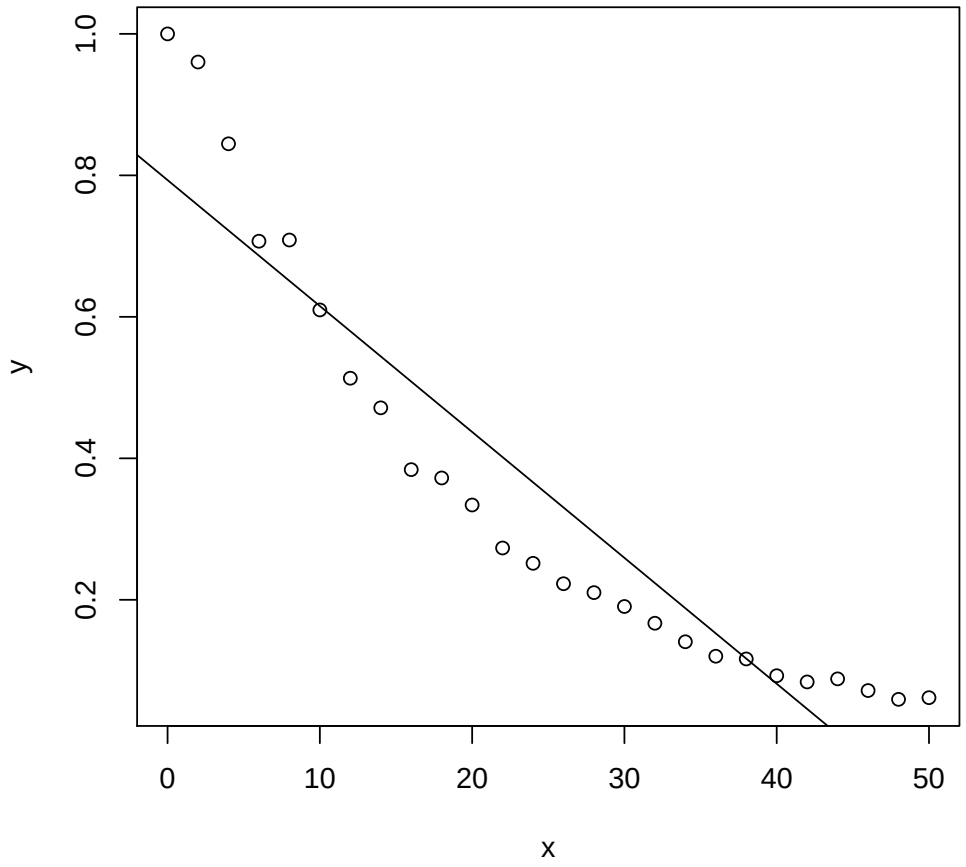
Emanuel G Heitlinger

Disclaimer: The following document is a private transcript of Mick Crawley's R-course. I am a participant in this course and my writeup has in no way been approved by Mick Crawly (from whom the ideas behind the code and teaching concepts are) or any of his staff.

Regression 2: Transformation of response in simple linear models

Non-linear response

```
> decay <- read.table("sapdecay.txt", header = TRUE)
> attach(decay)
> plot(x, y)
> abline(lm(y ~ x))
```



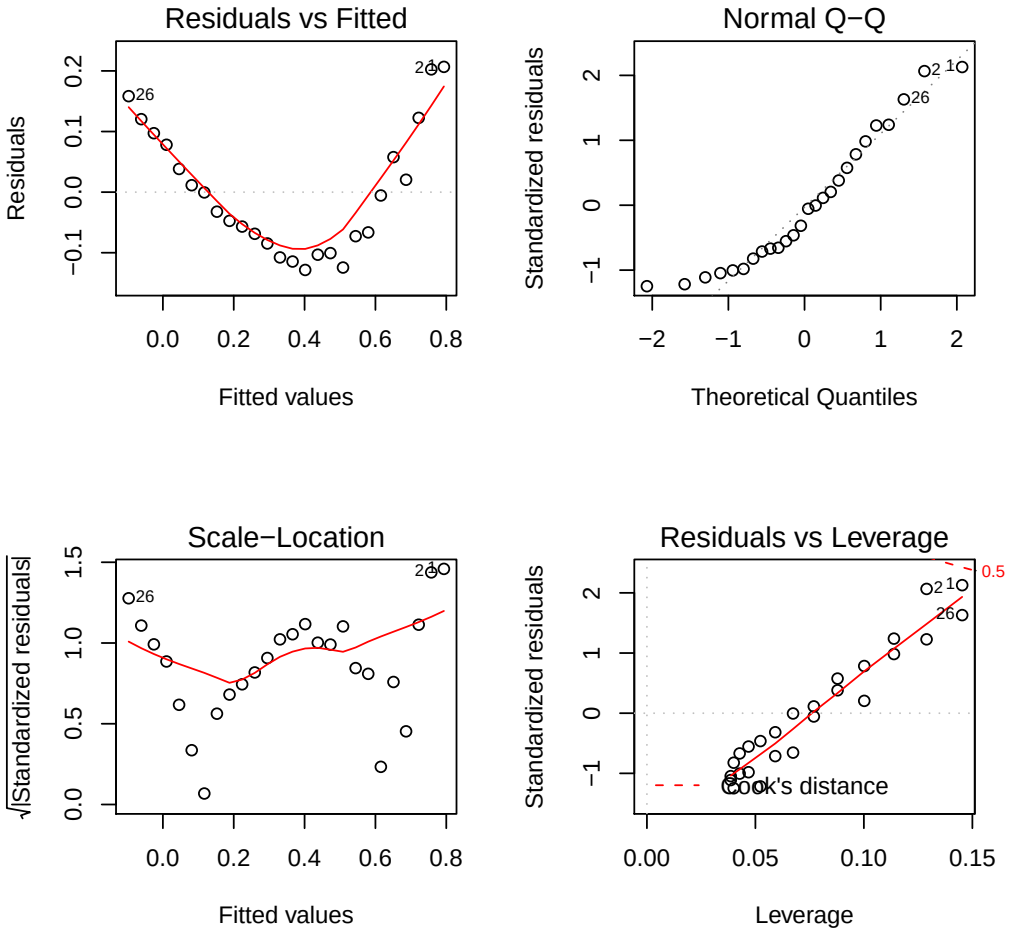
On the first view this relationship doesn't look particularly linear. Let's see how the model behaves:

```
> mod <- lm(y ~ x)
> summary(mod)

Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max
-0.12840 -0.08170 -0.01887  0.07305  0.20672

Coefficients:
```

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.793277   0.040044   19.81 2.22e-16 ***
x           -0.017799   0.001374  -12.96 2.51e-12 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ

Residual standard error: 0.1051 on 24 degrees of freedom
Multiple R-squared: 0.8749,        Adjusted R-squared: 0.8697
F-statistic: 167.9 on 1 and 24 DF,  p-value: 2.506e-12
```

Everything is higly significant, but is the model good? We can use diagnostic plots.

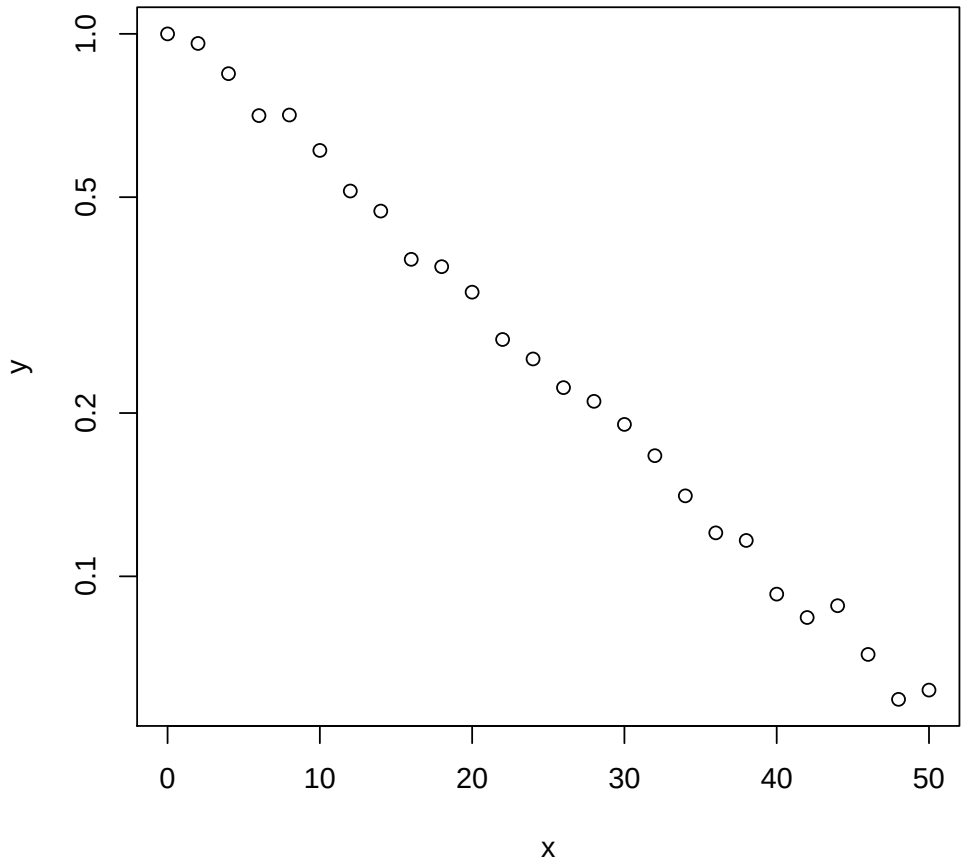
```
> par(mfrow = c(2, 2))
> plot(mod)
```



Most important are the first two plots: We see that the residuals clearly start off big, get small and then big again. The resuduals plotted against theoretical quantiles of the normal distribution are not on a straight line.

Let's try log transformation of the response:

```
> plot(x, y, log = "y")
```



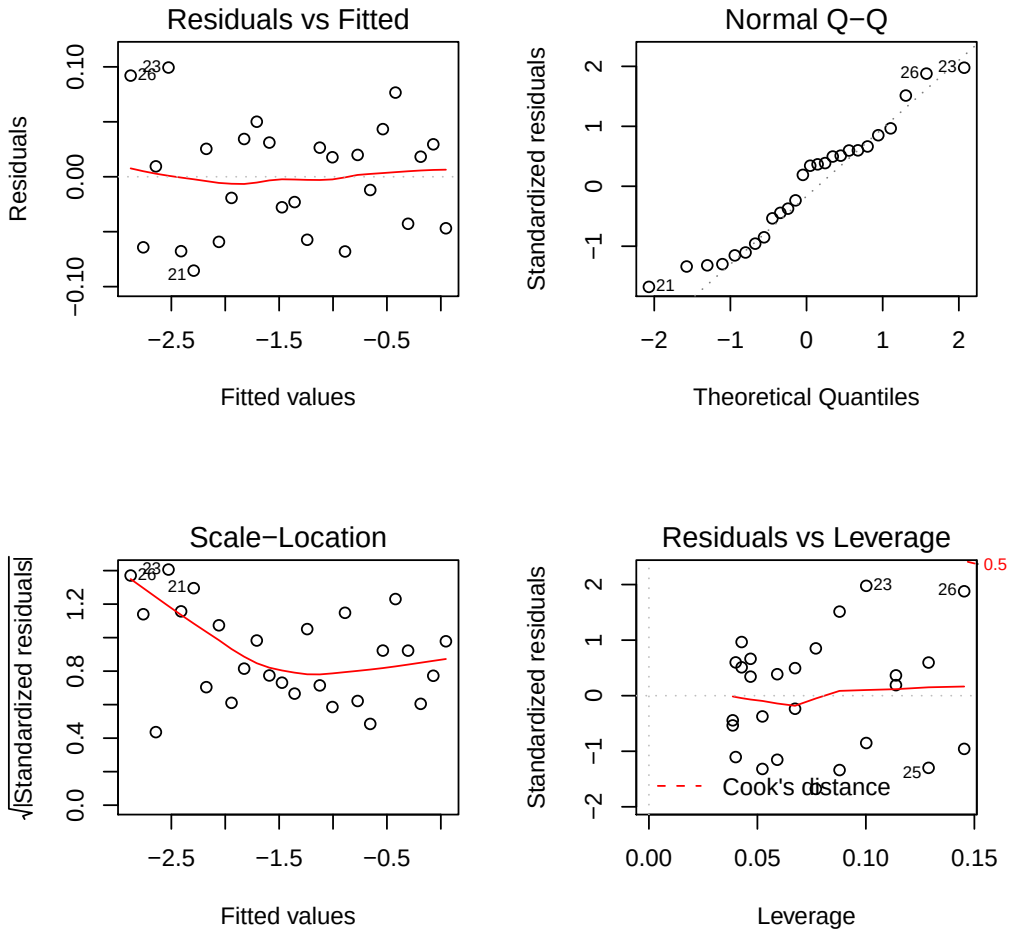
Looks promising, let's model it like this

```
> tra.mod <- lm(log(y) ~ x)
> summary(tra.mod)

Call:
lm(formula = log(y) ~ x)

Residuals:
     Min       1Q   Median       3Q      Max
-0.08542 -0.04586  0.01362  0.03071  0.09940

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)  0.0468837  0.0201983   2.321   0.0291 *
x           -0.0584863  0.0006928 -84.421   <2e-16 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ

Residual standard error: 0.05299 on 24 degrees of freedom
Multiple R-squared: 0.9966,         Adjusted R-squared: 0.9965
F-statistic:  7127 on 1 and 24 DF,  p-value: < 2.2e-16
```

Okay, how do the diagnostics look?

```
> par(mfrow = c(2, 2))
> plot(tra.mod)
```



Much better: The residuals vs. fittet look like a sky at night. The residuals vs. normal Quantiles still show some problems, but we will leave this for now.

How to get this model in a easy interpretable picture? A combined use of predict and back-transformation will help:

```
> plot(x, y)
> smoothx <- seq(0, max(x), 0.1)
> smoothy <- exp(predict(tra.mod, list(x = smoothx)))
> lines(smoothx, smoothy)
```