**Non- Linear Models**

**Problem**. We know that the amount of a radioactive isotope present decays over time exponentially. If A(t) = amount present at time t, then

A(t) = A0ert,

for some values of A0 and r. Suppose at time t = 1, 2 gm are present, and at time t = 2,

1 gm is present. What are the values of r and A0 ? **Note: R uses log for the natural log function, so I use log notation instead of ln notation.**

This approach to determining r and A0 is a bit unrealistic. It presumes that we can measure the amount present perfectly, without error. In the real world, measurements come with errors. So, how do we handle this? Let M be a measured value of the amount at time t. Then, a model has the form

M = A0ert + ,

where  is a random variable that represents the measurement error and A0ert is the “true” value at time t.

There are two variables: time and measured amount. We can think of time as the explanatory variable and measured amount as the response variable. If the relationship between these variables were linear, we already know what to do. We make multiple measurements of the amount at various times and use that data to estimate a simple linear regression model for the relationship. Unfortunately, **the relationship is NOT linear.**

**However, using logarithms, we can transform the non-linear relation into a linear relation.**

**log(M) =**

Here, time is still the explanatory variable, but log(measurment) is the response variable. The intercept parameter is log(A0) and the slope parameter is r. We can take our measurement data (time, measurment), convert them into data (time, log(measurement)) and use the converted data to estimate the linear model with parameters log(A0) and r. From those estimates, we can compute the estimates for the parameters of the original exponential model.

> time<-runif(100,0,3)

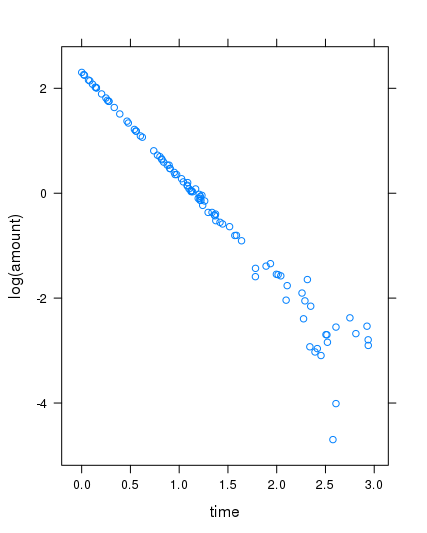
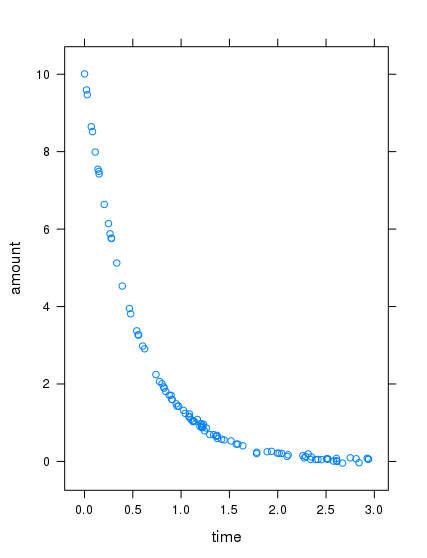
> error<-rnorm(100,0,.04)

> measurement<-10\*exp(-2\*time)+error

> **decay<-data.frame(measurement,time)**

> xyplot(measurement~time,data=decay)

> xyplot(log(measurement)~time,data=decay)



> summary(lm(log(measurement)~time,data=decay))

Call:

lm(formula = log(measurement) ~ time, data = decay)

Residuals:

Min 1Q Median 3Q Max

-1.90712 -0.01915 0.00844 0.04314 0.94134

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(**Intercept) 2.26993** 0.06242 36.37 <2e-16 \*\*\*

**time -1.96400** 0.04061 -48.36 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3241 on 95 degrees of freedom

(3 observations deleted due to missingness)

Multiple R-squared: 0.961, Adjusted R-squared: 0.9606

F-statistic: 2339 on 1 and 95 DF, **p-value: < 2.2e-16**

**So,** the estimated value of **log(A0) = 2.26993** and the estimated value of **r = -1.964.** The resulting estimates for the parameters of the original exponential model are

A0 = e2.26883 = 9.68 and r = -1.96

**Another kind of non-linear relation that can be linearized using logs**

Y = axk

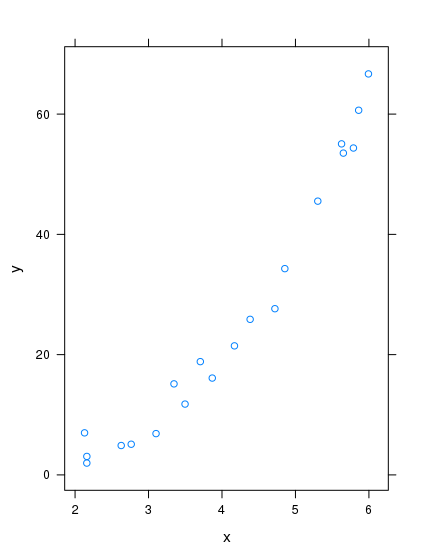
> x<-runif(20,2,6)

> error<-rnorm(20,0,3)

> y<-.3\*x^3+error

> ex<-data.frame(y,x)

> xyplot(y~x,data=ex)



**Linearizing using logs**

**log(Y) =**

Explanatory variable =

Response variable =

Parameters: **intercept =** **slope =**

summary(lm(log(y)~log(x),data=ex))

Call:

lm(formula = log(y) ~ log(x), data = ex)

Residuals:

Min 1Q Median 3Q Max

-0.47923 -0.10416 -0.00088 0.03962 0.83254

Coefficients:

Estimate Std. Error t value Pr(>|t|)

**(Intercept) -1.0989** 0.2371 -4.635 0.000206 \*\*\*

**log(x) 2.9258** 0.1700 17.208 1.27e-12 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2636 on 18 degrees of freedom

Multiple R-squared: 0.9427, Adjusted R-squared: 0.9395

F-statistic: 296.1 on 1 and 18 DF, p-value: 1.269e-12

So, the estimate of **log(a) = -1.0989** and the estimate of **k = 2.9258**. Hence the estimates for the original parameters are

**a = e-1.0989 = .3332. and k = 2.9258**

**Not all non-linear relations can be linearized using logarithms. Substitution is another way of (sometimes) linearizing a model.**

non-linear model y = ax2 + b data : (xi,yi), i = 1,…,n

Let u = x2.

linear model y = au + b data : (ui,yi) where ui = xi2

**Example**

> error<-rnorm(50,0,.5)

> x<-runif(50,0,2)

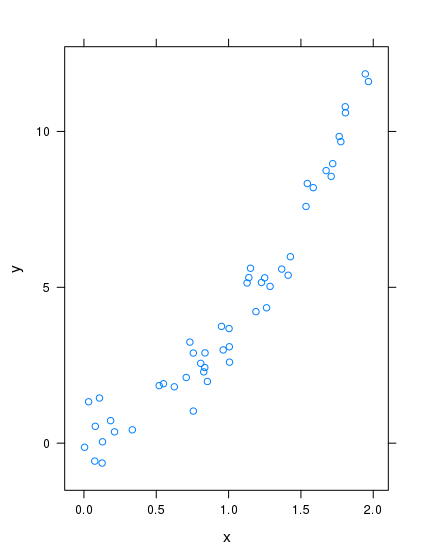
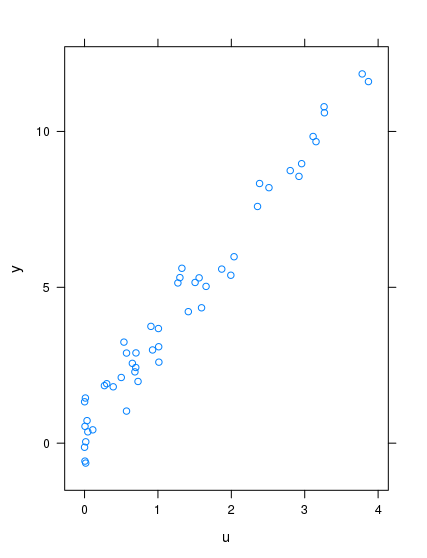
> y<-a\*x\*x+b+error

> u<-x\*x

> example<-data.frame(x,y,u)

> xyplot(y~u,data=example)

> xyplot(y~x,data=example)



> lm(y~u,data=example)

Call:

lm(formula = y ~ u, data = example)

Coefficients:

**(Intercept) u**

* **0.4628 2.9752**

**Estimates of the original coefficients**

**a = 2.975 b = .4628. Y = 2.975x2 + 0.4628**

**Exercises 16**

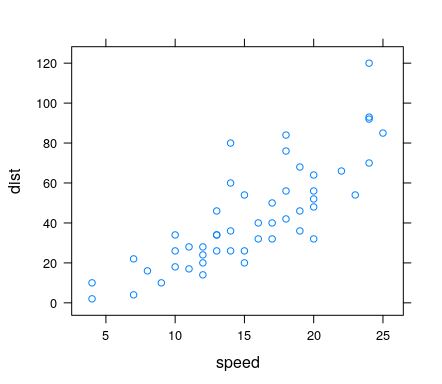
1. There is a connection between a person’s height and that person’s weight. Would a good model be. Weight = a\*Height + b? Since weight is related to volume and height is only one dimensional, it seems like a model of the form

**Weight = a\*Height^r**

might be a better model. The dataframe StudentSurvey (in the package Lock5withR) contains Height (inches) and Weight (in pounds) for 362 college students. Use logarithms to transform this into a linear relation. What are the parameters of the linear relation. Use R and the data in StudentSurvey to get estimates of the original parameters a and r.

1. The data frame cars contains measures of speed (mph) and stopping distance (feet). This data comes from 1020’s. Our problem: find a model that predicts stopping distance from speed. A scatter plot of the data is below.

> xyplot(dist~speed, data = cars)



From the scatter plot, it looks like the relation is not linear. A model of the form

**dist = A\*speed^r** might be correct. Use log to transform this into a linear model. Use R and the data in **cars** to get estimates of the parameters A and r.

1. Try the model **dist = a\*speed^2 + b** for problem 2 above**. Linearize this model using a substitution.** Use R and the data in **cars** to get estimates of the parameters a and b. To get the model of the linearized version with R use:

> f<-function(x) x^2

> lm(dist~f(speed), data = cars)