Lecture 1 – Extending the linear model

Goal: to use mathematical models to describe observed data

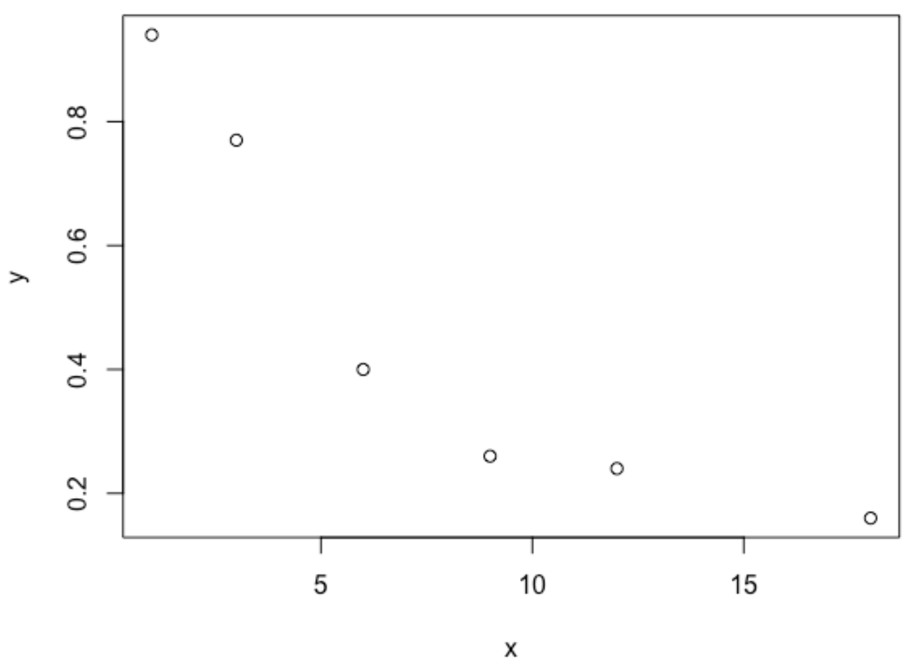
Example: Murdock (1961) – "forgetting curve"

* Ss were presented a "word", counted backward for a short duration, then were asked to recall the word.
  - manipulated the length of the retention interval

Retention interval (x)	0	3	6	9	12	18
Mean prop. recalled (Y)	.94	.77	.40	.26	.24	.16

```
x = c(1,3,6,9,12,18) # retention intervals (in seconds)
y = c(0.94, 0.77, 0.40, 0.26, 0.24, 0.16) # proportion recalled

plot(y~x)
```



Today's goal is to fit some models to these data.

Model 1 - a linear model

Mathematical form: $y = a + bx$

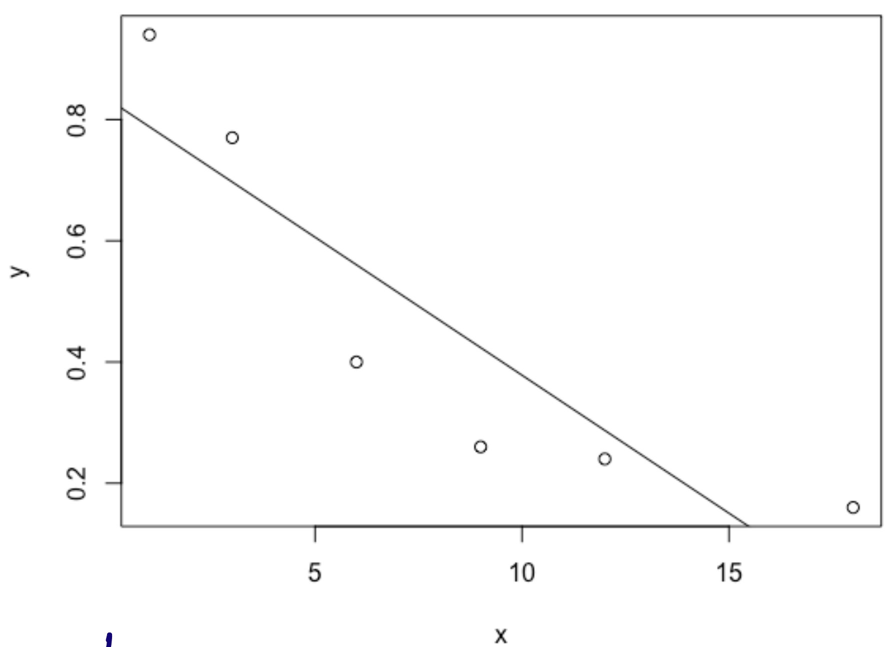
↳ a = "intercept / "initial value"

b = slope / "growth rate"

```r
# linear model
model1 = lm(y~x)
summary(model1)

# extract parameters (a = intercept, b=slope)
a = model1$coefficients[1]
b = model1$coefficients[2]

curve(a+b*x,
      from=0, to=20,
      add=T)
```



Not a very good fit!

The data seems to curve, but the model does not.

Model 2 - the exponential model

Form: $y = ab^x$

↳ a = intercept / initial value

b = growth factor

How to fit?

* exponential model is based on multiplication
* trick: use logarithm to convert multiplication to addition

$$y = ab^x$$

$$\log(y) = \log(ab^x)$$

$$\log(y) = \log(a) + \log(b^x)$$

$$\log(y) = \underbrace{\log(a)}_{\text{intercept}} + x \cdot \underbrace{\log(b)}_{\text{slope}}$$

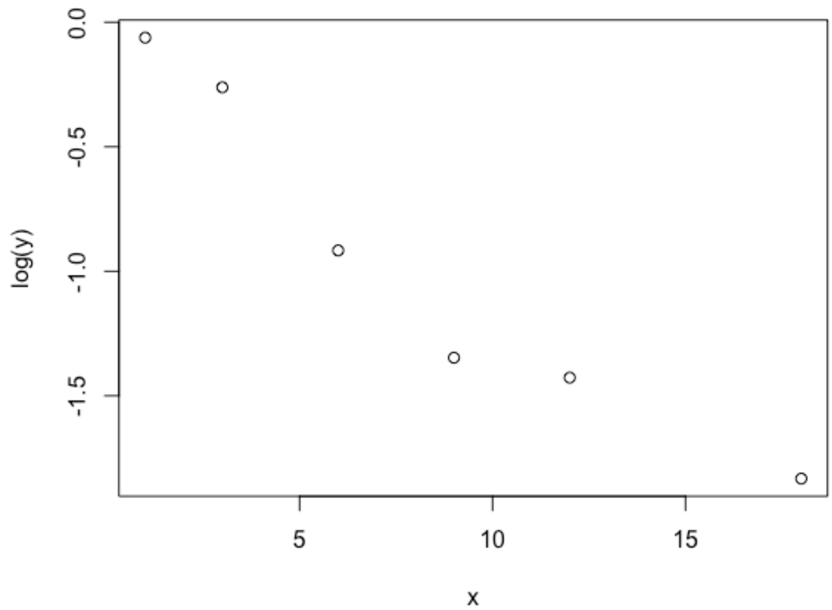
Fact: $\log(y)$ is a linear function of x.

↳ that is, if we plot x versus $\log(y)$, it will be a straight line with

$$\text{intercept} = \log(a)$$
$$\text{slope} = \log(b)$$

```r
# exponential model
plot(log(y)~x)

```

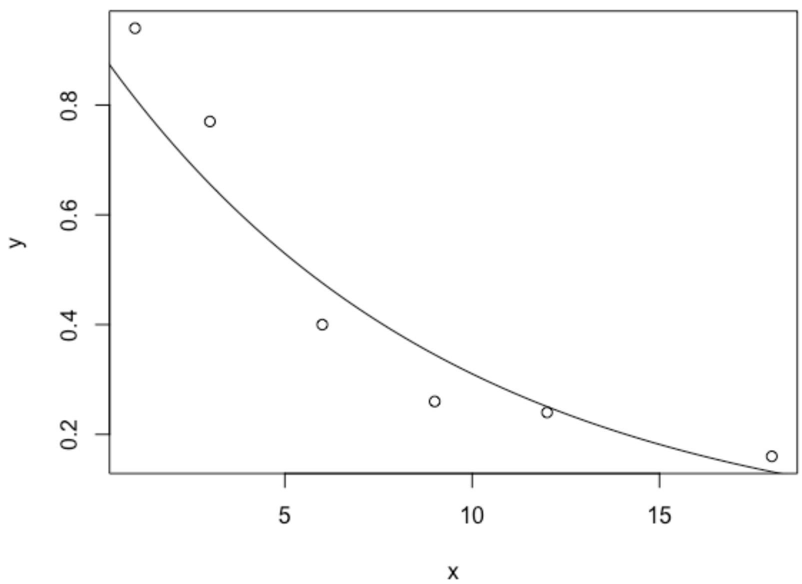


How to fit exponential model:

```r
model2 = lm(log(y)~x)
summary(model2)

int = model2$coefficients[1]
slope = model2$coefficients[2]

# plot exponential curve on data
plot(y~x)
a = exp(int)
b = exp(slope)
curve(a*b^x, 0, 20, add=T)
```



Model 3 – power model

Form: $$y = a x^b$$

↳ a = initial value

b = growth rate

Take logarithm:

$$y = a x^b$$

$$\log(y) = \log(a x^b)$$

$$\log(y) = \log(a) + \log(x^b)$$

$$\log(y) = \underbrace{\log(a)}_{\text{intercept}} + \underbrace{b}_{\text{slope}} \cdot \log(x)$$

This time, $\log(y)$ is a linear function of $\log(x)$.

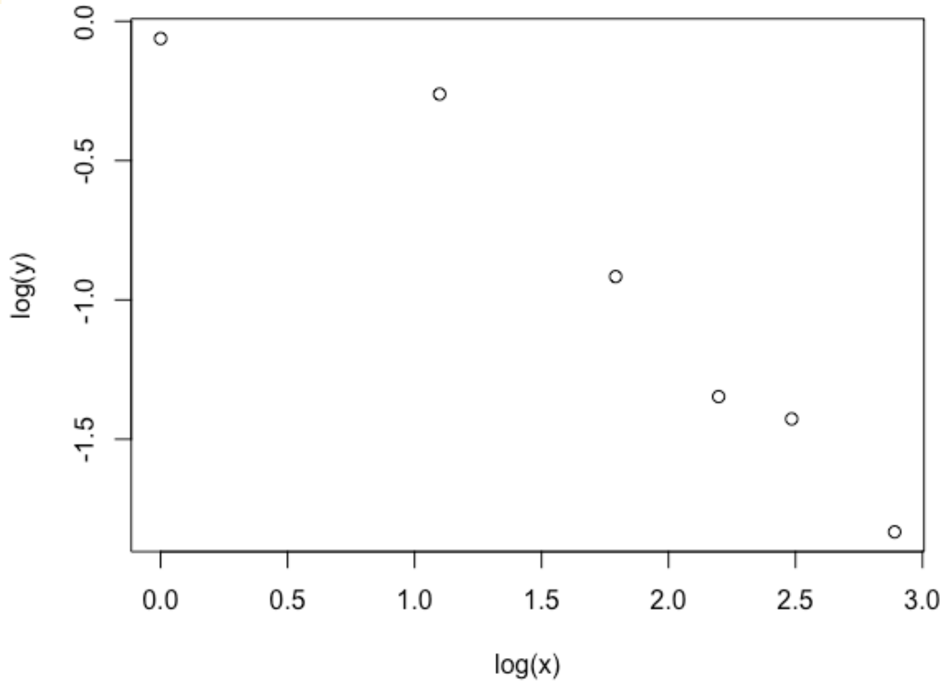
↳ the "log-log" plot will be a straight line.

with

intercept = $\log(a)$

slope = b

```r
# power function model
plot(log(y) ~ log(x))
```

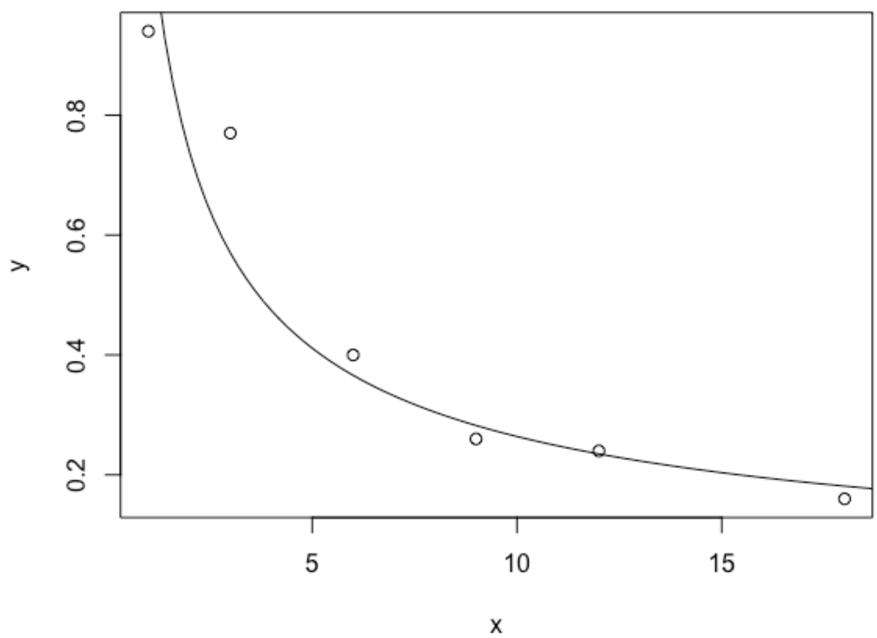



```r
model3 = lm(log(y)~log(x))
summary(model3)

int = model3$coefficients[1]
slope = model3$coefficients[2]

# plot power curve on data
plot(y~x)
a = exp(int)
b = slope
curve(a*x^b, 0, 20, add=T)

```



Wrapping up: which model is best?

* in linear modeling, we usually use R^2 as a measure of fit.

  ↳ R^2 = proportion of variability in y that is explained by the model.

```
# computing model fit
summary(model1)
summary(model2)
summary(model3)
```

Model	R^2
Linear	0.7945
Exponential	0.9188
Power	0.9331