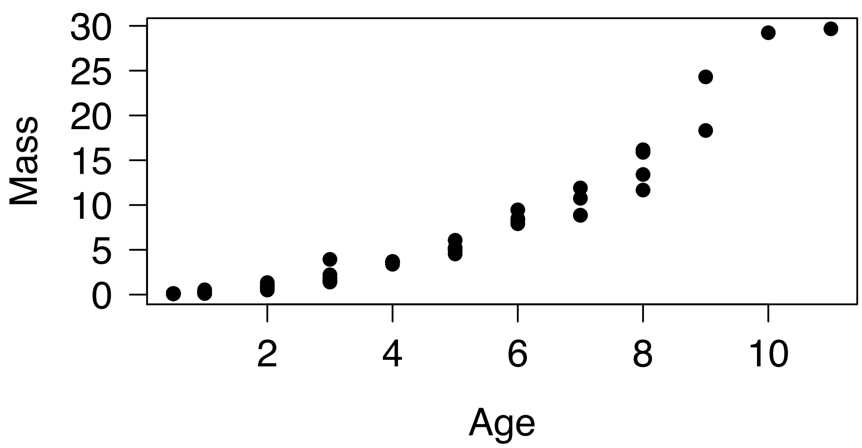
Homework 9

We would like to know how fast dinosaurs grew (obviously). Some clever folks have extracted size-at-age data from fossils using growth rings in the bones. For the species *Psittacosaurus lujiatunensis*:

the data look like this:



(taken from this paper: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0081917)

Age is in units of years and mass is in units of kilograms. There are many different nonlinear models that one could use to model the size of an organism as a function of its age. And there are some complexities in how the data were created that I will gloss over for our purposes. Let's keep things simple, and use model selection to ask:

- How fast does the dinosaur grow when it is young and growing quickly?

- Is there any evidence that this curve saturates at some asymptotic mass? Or is the curve basically exponential?
- If there is a sigmoidal relationship, what is the estimated maximum size?
- How well does a simple linear model fit relative to more realistic models?

To answer these questions, you'll need to fit the following three curves:

1. Linear model: $Mass = a + b*Age$
2. Exponential model: $Mass = a*\exp(r*Age)$
3. Logistic model: $Mass = \frac{M_{max}}{1+\exp\left(-r*(Age-Age_o)\right)}$

The linear model has parameters *a* and *b*; the exponential model has parameters *a* and *r*, and the logistic model has parameters M_{max}, *r*, and Age_0.

I have written the exponential and logistic curves in a different way than what we've used previously in the course for GLMs. These the same shape as what you've seen before, but writing them this way makes it easier to interpret in terms of the growth model. The exponential curve has an intercept (mass at age 0) at *a*, and an exponential growth rate of *r*. The logistic model also has an exponential growth rate of *r* when age ~ 0, but eventually the size of the organism saturates at the asymptote M_{max}, which is the maximum size. The logistic curve has a third parameter Age_o, which is the inflection point.

Fit these three curves to the dataset 'psittacosaurus.csv', using nls() as described in lecture. You'll need to supply start values for the parameters, which may be tricky. If you get error messages when trying to fit the model, use trial-and-error, or try plotting what the curve looks like with different parameter values to come up with reasonable guesses. For each model report the coefficient estimates and confidence intervals, and plot the fitted curves on top of the raw data.

Compare the three models using AICc. Which model is the best? What are the ΔAICc values and the Akaike weights for the three models? How do you interpret these results in terms of the relative support for the three models?

What is the estimated exponential growth rate (*r*) for the exponential and logistic models? What is the confidence interval on this parameter for the two models? For exponential growth, the doubling time is log(2)/r. How long does it take the dinosaur to double in size, based on the two models?

Is there evidence that this dinosaur has a maximum size? If so, what is the estimate for that size, and what is the confidence interval around that estimate? How does the estimated maximum size compare to the largest size in the data? How much stock do you put in the M_{max} estimate, given the data we have? If this estimate of M_{max} is true, about how big does this dinosaur get, relative to a human?

Now compare the three models using leave-one-out cross-validation. Which model is the best at predicting the data, in terms of LOOCV? What is the typical difference between the predicted values and the observed values for the best model? Does cross-validation yield the same ranking of models as AICc?