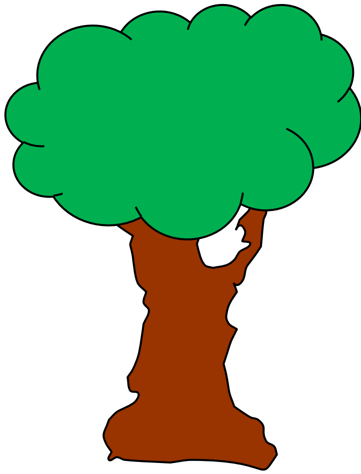
Tree volume as a function of width and height



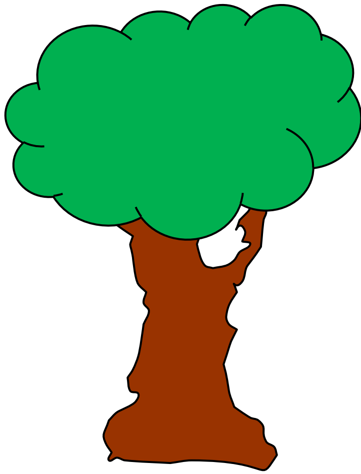
The data consists of measured cherry tree volume, width and height. It can be read in R as follows: trees = read.csv("http://www.finse.uio.no/events/international-workshops/introduction-to-estimation/data/trees.csv")

The point of the exercise is to find a reasonable relationship describing the volume as a function of diameter and height.

1) Try using linear regression on the original model (use lm(Volume~Diameter+Height,data=trees) (Rice - 14.44)
   a) The model behind linear regression says that in expectancy, : V=a+b*H+c*D. From a biological or purely geometrical point of view, would you expect this to be the case?
   b) Fit the model. Discuss the values of the 3 parameters.
   c) For a particularly small tree, H=1, D=0.1, what is the estimated volume?
   d) If the tree was flattened completely (H=0) what would you expect of the volume? What does the linear model say?
   e) The linear model is written as: $V_i=a+b*H_i+c*W_i+\varepsilon_i$, $\varepsilon_i\sim N(0,\sigma^2)$ i.i.d. Discuss the qualities of the error terms and how they compare to the qualities you would expect when comparing a model prediction with the real volume.

Tree volume as a function of width and height

2) Geometrical scaling says that if you change the height or diameter while keeping the shape, then the volume can be written as $V \propto HD^2$.



a) Compare this geometrical relationship with the linear model. What does the geometrical model and the linear model say about the case when the height and diameter of the tree goes to zero?

b) If we log-transform volume, height and diameter and use a linear model on that, log(V)=a+b log(H)+log(D), what relationship does that describe on the original (untransformed) scale? Will the geometrical relationship be a sub-model of this? What would a deviance from the geometrical model signify?

c) Fit log-transformed data. Does the result suggest a significant difference to geometrical scaling?

d) Estimate the volume for the case D=0.1, H=1.

e) The model on the log-transformed data is: $\log(V_i)=a+b*\log(H_i)+c*\log(W_i)+\varepsilon_i$, $\varepsilon_i \sim N(0,\sigma^2)$ i.i.d. What are the qualities of the error terms on the original scale, now? Are they reasonable?