

# Models of length distributions

fish5103growth Modelling length at age and length distributions

Gunnar Stefansson and Lorna Taylor

December 19, 2016

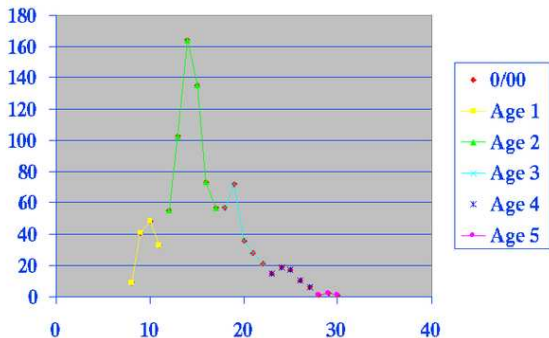
# Statistical and other models of length distributions

A fairly simple statistical model of length distributions is a combination of cohort length distributions, each of which is assumed to be for a specific probability density such as a Gaussian density.

The location of each density is centered on the mean length of the corresponding cohort with some standard deviation.

The multiplicative factors forming the combination reflect the relative strength of each cohort.

# Cohort slicing



**Figure :** A length distribution is sliced by assigning length groups surrounding a mode to an age group.

Slicing is done by grouping length classes into age groups.

# The distribution of length at age

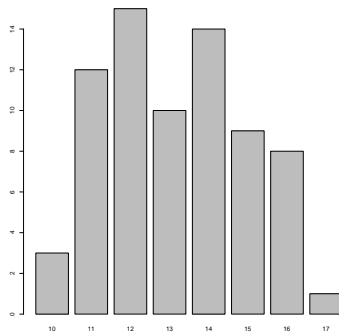
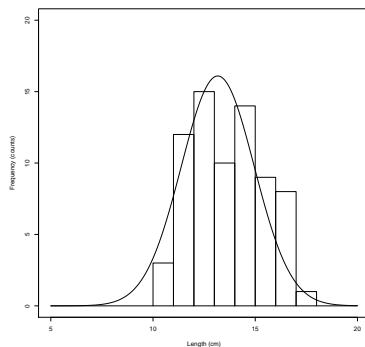


Figure : Some length measurements on age 1 cod in Icelandic waters in 1996.

# The Gaussian density and cdf



**Figure 1:** Length frequency and fitted Gaussian for a single age group.

**Example:** Data on the length of one-group cod and a normal density, with the same mean and standard deviation of length and scaled to the same counts.

# The proportion within a length group

Gaussian density and cumulative distribution function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$$

$$F(x) = \int_{-\infty}^x \phi(t)dt = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Fix  $a$ : Proportion within length category  $l$  of width 1 is

$$\Phi\left(\frac{(l + \frac{1}{2}) - \mu_a}{\sigma_a}\right) - \Phi\left(\frac{(l - \frac{1}{2}) - \mu_a}{\sigma_a}\right)$$

The proportion of fish in length group  $l$ , across all ages becomes:

$$\sum_a \pi_a \left\{ \Phi\left(\frac{(l + \frac{1}{2}) - \mu_a}{\sigma_a}\right) - \Phi\left(\frac{(l - \frac{1}{2}) - \mu_a}{\sigma_a}\right) \right\}.$$

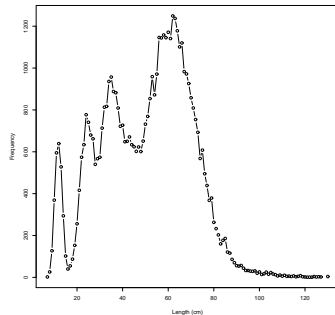
Unknown parameters:  $\pi_a, \mu_a, \sigma_a$ .

# Statistical estimation of proportions at age

Parameters:  $\pi_a$ ,  $\mu_a$ ,  $\sigma_a$

Need some criteria, e.g.  $\sum (y_l - \hat{y}_l)^2$

Where  $y_l$  is the measured proportion in cell  $l$  and  $\hat{y}_l$  is the modeled proportion



**Figure :** Example length distribution. Cod in 1990 survey of Icelandic waters.

If at least some modes can be discerned, it is feasible to use data-fitting methods to estimate normal distributions which add up to give the observed length distribution as the sum of the individual components.

# Setting initial values

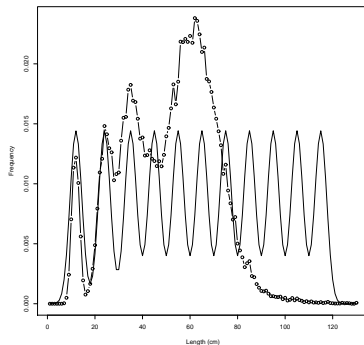


Figure : Example length distribution with initial values for fitted distribution.



# Estimating proportions alone

The proportions are easiest to estimate and can be estimated for given values of the mean and standard deviation.

The results from this first estimation part will typically clear up what needs to be done about standard deviations and means.

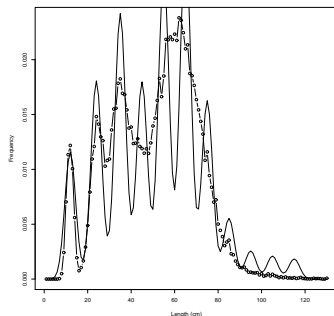


Figure : Example length distribution. Cod in 1990 survey of Icelandic waters.

In this particular case the standard deviations on the older ages are clearly way too low.

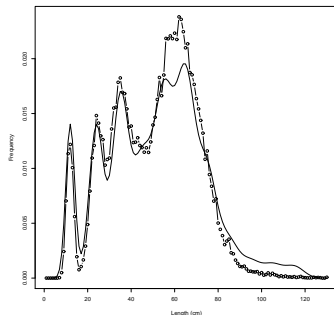
```
# Define a function to evaluate the fit of different vectors of proportions
sseprop<-function(pvec){
  pi<-c(pvec,1-sum(pvec))
  # The pvector is input - the mu and sigma are fixed
  # na-1 values - the last is computed
```

# Typical parameter reduction - sigma

Commonly assume equal standard deviations

Often the standard deviation for the youngest is clearly lower than for the oldest

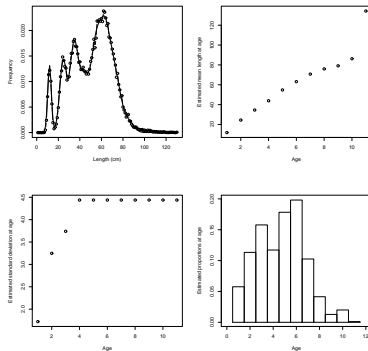
Possibly estimate one or a few for the younger ages



**Figure :** Example length distribution. Cod in 1990 survey of Icelandic waters. Sigma values estimated after proportions.

After the proportions, standard deviations for 4 ages are estimated.  
See `base.r` for the code.

## Full run



**Figure :** Example length distribution. Cod in 1990 survey of Icelandic waters. Parameters estimated in twice-repeated sequence.

# Parameter reduction - using a growth curve

May want to use a von B growth curve in place of  $\mu_a$

**R coding:**

```
ssevonB<-function(beta){
  Linf<-exp(beta[1])
  k<-beta[2]
  muvec<-Linf*(1-exp(-k*(1:11)))
  fit<-rep(0,130)
  for(lgrp in 1:130){
    fit[lgrp]<-sum(pi*(pnorm((lgrp+0.5-muvec)/sigma)-pnorm((lgrp-0.5-muvec)/sigma))
  }
  sse<-sum((dat-fit)^2)
  plot(lgrps,dat,type='b',lwd=2)
  lines(lgrps,fit,type='l',lwd=2,col="red")
  cat("SSE=",sse,"\n")
  return(sse)
}
```

# Caveats

Uncertainty estimation

Time of sampling - fixed-point (survey) or continuous (catches)

Time of spawning - short interval or continuous or bimodal

etc

# The next steps

