

A model for changes in length frequencies

1 Data

Read in length data and modify some column names and variable labels for use below.

```
library(gdata)

setwd("../data")

neph.dat <- read.xls("Celtic Warrior Diamond mesh July 2014 Celtic Sea.xls",
  sheet = "Nephrops Lengths", stringsAsFactors = FALSE)

## Show the first 2 rows
head(neph.dat, 2)

## Change the carapace length name
names(neph.dat)[names(neph.dat) == "Carapace.Length..mm.."] <- "Carapace.Length"

## Make the 'HAUL' variable character
neph.dat$HAUL <- paste("H", neph.dat$HAUL, sep = "")
```

Make one row per length measurement assuming, for example, that a sub-sampling ratio of 0.1 corresponds to 10% of the catch sampled (CHECK).

```
## get a row per length measurement (raise them also)
n <- nrow(neph.dat)

neph.dat2 <- neph.dat[rep(1:n, times = round(neph.dat$COUNT/neph.dat$SUBSRATIO,
  0)), ]
```

Read in the haul weights

```
setwd("../data")

weight.dat <- read.xls("Celtic Warrior Diamond mesh July 2014 Celtic Sea.xls",
  sheet = "Weights", stringsAsFactors = FALSE)
```

```

## Show the first 2 rows
head(weight.dat, 2)

## create a new 'HAUL' variable for the merge
weight.dat$HAUL <- paste("H", weight.dat$Haul., sep = "")

## re-name total weight column
names(weight.dat)[names(weight.dat) == "Total.weight..kg."] <- "Total.Weight"

```

Merge the bulk weights with the length data

```

neph.dat3 <- merge(neph.dat2,
                  subset(weight.dat, Species == "Bulk")[, c("Mesh.Size", "HAUL",
                                                             "Total.Weight")],
                  by = c("Mesh.Size", "HAUL"))

## subset the data by mesh size
neph.70mm <- subset(neph.dat3, Mesh.Size == "70mm")
neph.80mm <- subset(neph.dat3, Mesh.Size == "80mm")
neph.90mm <- subset(neph.dat3, Mesh.Size == "90mm")
neph.100mm <- subset(neph.dat3, Mesh.Size == "100mm")

## convert HAUL to factor
## with levels depending on the haul weight
neph.70mm$HAUL <- factor(neph.70mm$HAUL, levels =
                        unique(neph.70mm$HAUL[order(neph.70mm$Total.Weight)]))
neph.80mm$HAUL <- factor(neph.80mm$HAUL, levels =
                        unique(neph.80mm$HAUL[order(neph.80mm$Total.Weight)]))
neph.90mm$HAUL <- factor(neph.90mm$HAUL, levels =
                        unique(neph.90mm$HAUL[order(neph.90mm$Total.Weight)]))
neph.100mm$HAUL <- factor(neph.100mm$HAUL, levels =
                        unique(neph.100mm$HAUL[order(neph.100mm$Total.Weight)]))

```

Produce a summary plot of the data by length, haul and catch weight.

```

library(ggplot2)
library(gridExtra)

## quick function for plot
plot.lfreq <- function(data, title.string){
  p <- ggplot(data, aes(x = Carapace.Length, group = HAUL)) +
    geom_density(position = "stack",
                aes(fill = Total.Weight),
                colour = 1, lwd = 0.005) +
    xlim(10, 45) +
    scale_fill_gradient2(low = "white", high = "blue",
                        limits = c(0, max(neph.dat3$Total.Weight))) +
    ggtitle(title.string) +
    theme(axis.text.x=element_blank())

  return(p)
}

```

```

}

p70mm <- plot.lfreq(data = neph.70mm, title.string = "70mm")
p80mm <- plot.lfreq(data = neph.80mm, title.string = "80mm")
p90mm <- plot.lfreq(data = neph.90mm, title.string = "90mm")
p100mm <- plot.lfreq(data = neph.100mm, title.string = "100mm")

grid.arrange(p70mm, p80mm, p90mm, p100mm, ncol = 1)

```

2 Models

The first model we'll try is a model assuming all lengths come from a normal distribution with a single mean μ and standard deviation σ

$$l_i \sim N(\mu, \sigma^2) \quad (1)$$

```

lm0 <- lm(Carapace.Length ~ 1, data = neph.dat3)

summary(lm0)

##
## Call:
## lm(formula = Carapace.Length ~ 1, data = neph.dat3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.698  -2.698  -0.698   3.302  53.302
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  28.698050   0.009518   3015    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.918 on 266990 degrees of freedom

## predict the density
pred.df <- data.frame(Carapace.Length = seq(0, 100))
pred.df$dens <- dnorm(pred.df$Carapace.Length, mean = coef(lm0)[1], sd = summary(lm0)$sigma)

## plot the distributional fit
p <- ggplot(neph.dat3, aes(x = Carapace.Length)) + geom_histogram(aes(y = ..density..),
  fill = "slategrey", color = "white", binwidth = 1) + xlim(10, 50)

p + geom_line(data = pred.df, aes(x = Carapace.Length, y = dens))

```

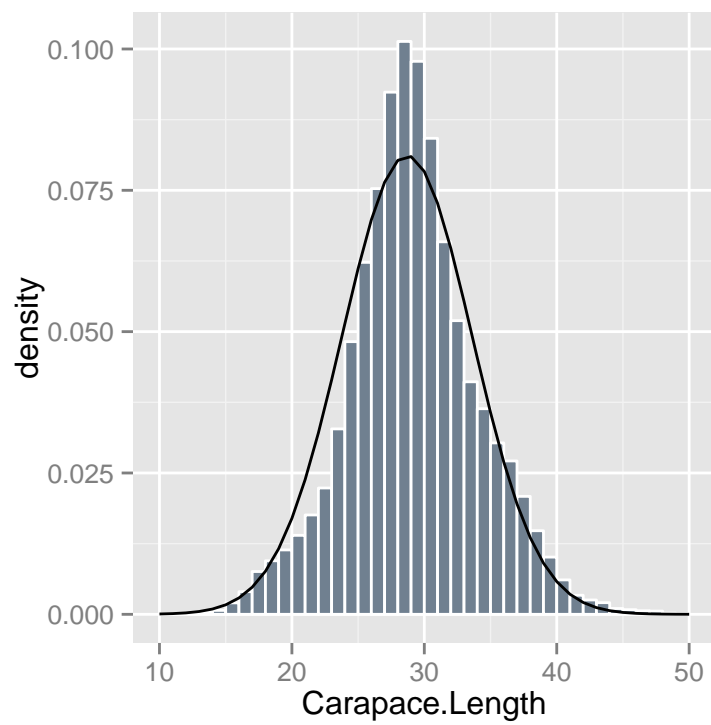


Figure 1: Histogram of carapace length over all hauls and mesh sizes with overlaid fitted single normal curve.

This model does not fit the data very well as there are large parts of the distribution not captured. Next model is to let the mean vary by mesh size

```
lm1 <- lm(Carapace.Length ~ Mesh.Size, data = neph.dat3)

summary(lm1)

##
## Call:
## lm(formula = Carapace.Length ~ Mesh.Size, data = neph.dat3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.753  -3.303  -0.340   2.697  52.660
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   29.33990    0.01985 1478.23  <2e-16 ***
## Mesh.Size70mm -0.86420    0.02690  -32.13  <2e-16 ***
## Mesh.Size80mm -1.03668    0.02741  -37.82  <2e-16 ***
## Mesh.Size90mm -0.58708    0.02756  -21.30  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.903 on 266987 degrees of freedom
## Multiple R-squared:  0.006105, Adjusted R-squared:  0.006094
## F-statistic: 546.7 on 3 and 266987 DF,  p-value: < 2.2e-16

## predict the densities
pred.df <- expand.grid(Carapace.Length = seq(0, 100), Mesh.Size = c("70mm",
  "80mm", "90mm", "100mm"))

pred.df$dens <- dnorm(pred.df$Carapace.Length, mean = predict(lm1, newdata = pred.df),
  sd = summary(lm1)$sigma)

## plot the distributional fit
p <- ggplot(neph.dat3, aes(x = Carapace.Length)) + geom_histogram(aes(y = ..density..),
  fill = "slategrey", color = "white", binwidth = 1) + xlim(10, 50) + facet_wrap(~Mesh.S

p + geom_line(data = pred.df, aes(x = Carapace.Length, y = dens))
```

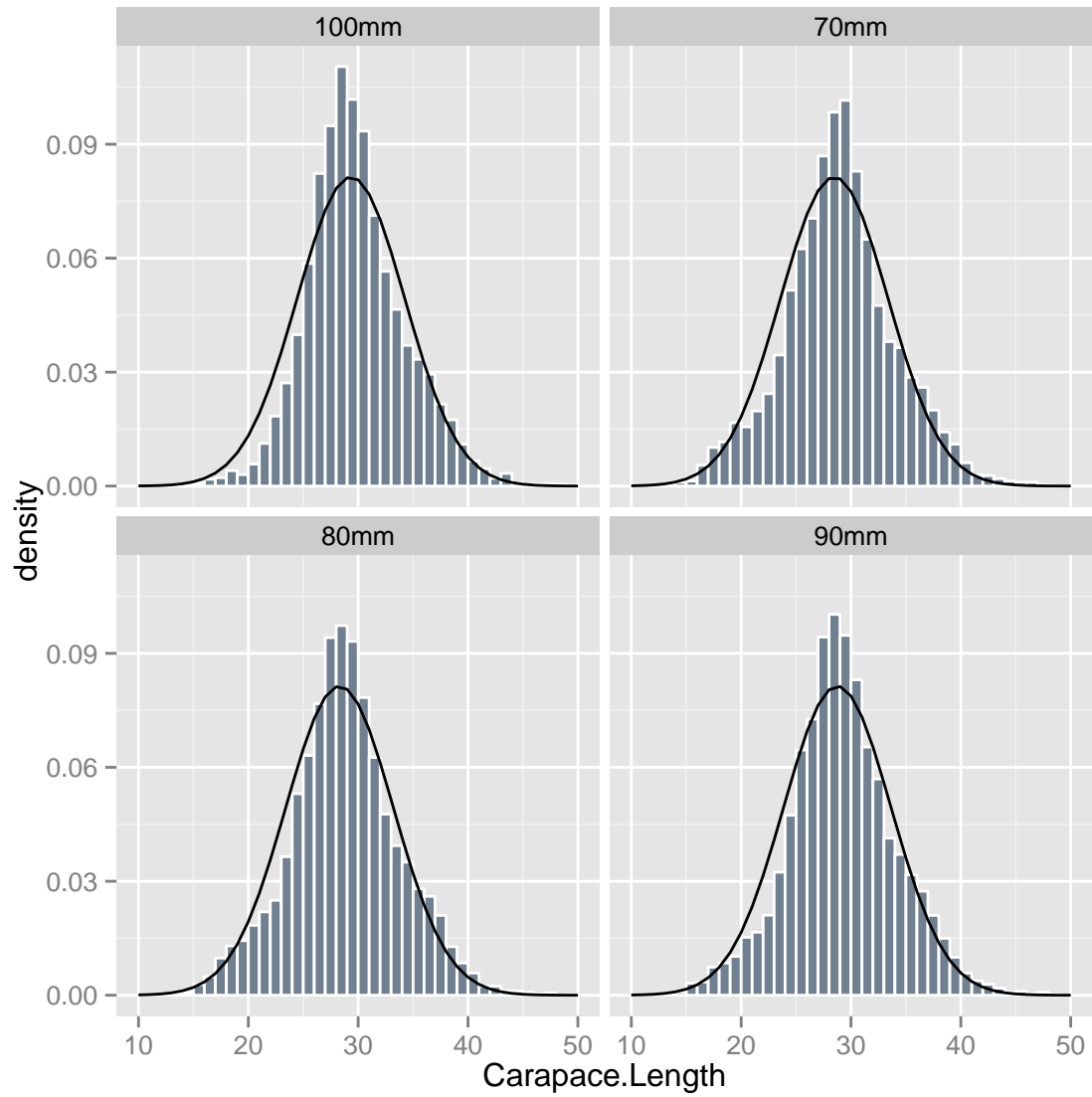


Figure 2: Histogram of carapace length by mesh size with overlaid normal curve with mean varying by mesh size.