Our Lass 70mm versus 80mm analysis

For discussion

1 Data

Read the data in and make a few factor variables

```
library(gdata)

neph.dat <- read.xls("../data/2015 BIM Nephrops quad rig trials/Our Lass 2 70_80_90_100mm c
                     sheet = "All hauls",
                     stringsAsFactors = FALSE)

## subset containing the 70mm and 80mm data
neph.7080 <- subset(neph.dat, Mesh.Size %in% c("70mm", "80mm"))

## make factor haul variable
neph.7080$fHAUL <- factor(paste("H", neph.7080$Haul.No, sep =""))

## make a factor mesh size variable
neph.7080$fMesh.Size <- factor(paste("mesh", neph.7080$Mesh.Size, sep = ""))

## remove trailing dot from raised count column name
names(neph.7080)[names(neph.7080) == "Raised.count."] <- "Raised.count"
```

Re-shape the data to wide format (columns for 70mm, 80mm variables).

```
## get count per length bin per haul by mesh size
## using the reshape package (makes it easier to process data)
library(reshape)

## variables to keep
vars2keep <- c("fHAUL", "fMesh.Size", "Net.position", "Carapace.length",
               "Count", "Raised.count", "Total.catch", "Overall.Sampling.Ratio")

## melt the data frame
neph.7080.melt <- melt(neph.7080[, vars2keep],
                       id = c("fHAUL", "fMesh.Size", "Carapace.length"))

## re-form the dataframe in required format
neph.7080.cast <- cast(neph.7080.melt, Carapace.length + fHAUL ~ fMesh.Size  + variable)
```

```
## first couple of lines
head(neph.7080.cast, 2)
```

```
##   Carapace.length fHAUL mesh70mm_Net.position mesh70mm_Count
## 1              17   H14                     4              1
## 2              17    H3                    NA             NA
##   mesh70mm_Raised.count mesh70mm_Total.catch
## 1                  32.1                  468
## 2                    NA                   NA
##   mesh70mm_Overall.Sampling.Ratio mesh80mm_Net.position mesh80mm_Count
## 1                          0.0311                    NA             NA
## 2                              NA                     3              1
##   mesh80mm_Raised.count mesh80mm_Total.catch
## 1                    NA                   NA
## 2                  32.1                  420
##   mesh80mm_Overall.Sampling.Ratio
## 1                              NA
## 2                          0.0312
```

```
summary(neph.7080.cast) ## note lots of NAs
```

```
##  Carapace.length     fHAUL     mesh70mm_Net.position mesh70mm_Count
##  Min.   :17.0    H11    : 31   Min.   :1.0           Min.   : 1.0
##  1st Qu.:27.0    H14    : 28   1st Qu.:2.0           1st Qu.: 3.0
##  Median :33.0    H4     : 27   Median :3.0           Median :10.0
##  Mean   :33.2    H7     : 27   Mean   :2.7           Mean   :14.8
##  3rd Qu.:40.0    H6     : 26   3rd Qu.:4.0           3rd Qu.:24.0
##  Max.   :54.0    H8     : 26   Max.   :4.0           Max.   :66.0
##                  (Other):172   NA's   :37            NA's   :37
##  mesh70mm_Raised.count mesh70mm_Total.catch
##  Min.   :  11          Min.   :203
##  1st Qu.:  72          1st Qu.:306
##  Median : 206          Median :411
##  Mean   : 375          Mean   :416
##  3rd Qu.: 580          3rd Qu.:490
##  Max.   :2251          Max.   :618
##  NA's   :37            NA's   :37
##  mesh70mm_Overall.Sampling.Ratio mesh80mm_Net.position mesh80mm_Count
##  Min.   :0.0                     Min.   :1.00          Min.   : 1.0
##  1st Qu.:0.0                     1st Qu.:2.00          1st Qu.: 2.2
##  Median :0.0                     Median :2.00          Median :10.0
##  Mean   :0.0                     Mean   :2.44          Mean   :13.8
##  3rd Qu.:0.1                     3rd Qu.:3.00          3rd Qu.:21.8
##  Max.   :0.1                     Max.   :4.00          Max.   :61.0
##  NA's   :37                      NA's   :31            NA's   :31
##  mesh80mm_Raised.count mesh80mm_Total.catch
##  Min.   :   9          Min.   :166
##  1st Qu.:  52          1st Qu.:265
##  Median : 193          Median :407
```

```
##  Mean   : 306          Mean   :388
##  3rd Qu.: 442          3rd Qu.:459
##  Max.   :1621          Max.   :635
##  NA's   :31            NA's   :31
##  mesh80mm_Overall.Sampling.Ratio
##  Min.   :0.02
##  1st Qu.:0.03
##  Median :0.05
##  Mean   :0.05
##  3rd Qu.:0.07
##  Max.   :0.11
##  NA's   :31
```

```
## fill in missing values
## these occur if there is a count for e.g. 20mm CL in 70mm but not in 80mm
neph.7080.cast$mesh70mm_Count[is.na(neph.7080.cast$mesh70mm_Count)] <- 0
neph.7080.cast$mesh70mm_Raised.count[is.na(neph.7080.cast$mesh70mm_Raised.count)] <- 0
neph.7080.cast$mesh80mm_Count[is.na(neph.7080.cast$mesh80mm_Count)] <- 0
neph.7080.cast$mesh80mm_Raised.count[is.na(neph.7080.cast$mesh80mm_Raised.count)] <- 0

for(i in 1:dim(neph.7080.cast)[1]){
  haul.dat <- subset(neph.7080.cast, fHAUL == neph.7080.cast$fHAUL[i])
  ## 70mm net position
  if(is.na(neph.7080.cast$mesh70mm_Net.position[i])){
    neph.7080.cast$mesh70mm_Net.position[i] <-
      unique(na.omit(haul.dat$mesh70mm_Net.position))
  }
  ## 80mm net position
  if(is.na(neph.7080.cast$mesh80mm_Net.position[i])){
    neph.7080.cast$mesh80mm_Net.position[i] <-
      unique(na.omit(haul.dat$mesh80mm_Net.position))
  }
  ## 70mm total catch
  if(is.na(neph.7080.cast$mesh70mm_Total.catch[i])){
    neph.7080.cast$mesh70mm_Total.catch[i] <-
      unique(na.omit(haul.dat$mesh70mm_Total.catch))
  }
  ## 80mm total catch
  if(is.na(neph.7080.cast$mesh80mm_Total.catch[i])){
    neph.7080.cast$mesh80mm_Total.catch[i] <-
      unique(na.omit(haul.dat$mesh80mm_Total.catch))
  }
  ## Sampling ratio
  ## 70mm total catch
  if(is.na(neph.7080.cast$mesh70mm_Overall.Sampling.Ratio[i])){
    neph.7080.cast$mesh70mm_Overall.Sampling.Ratio[i] <-
      unique(na.omit(haul.dat$mesh70mm_Overall.Sampling.Ratio))
  }
  ## 80mm total catch
  if(is.na(neph.7080.cast$mesh80mm_Overall.Sampling.Ratio[i])){
    neph.7080.cast$mesh80mm_Overall.Sampling.Ratio[i] <-
```

```
      unique(na.omit(haul.dat$mesh80mm_Overall.Sampling.Ratio))
  }
}

summary(neph.7080.cast) ## no missing


##  Carapace.length     fHAUL     mesh70mm_Net.position mesh70mm_Count
##  Min.   :17.0    H11    : 31   Min.   :1.00          Min.   : 0.0
##  1st Qu.:27.0    H14    : 28   1st Qu.:2.00          1st Qu.: 2.0
##  Median :33.0    H4     : 27   Median :3.00          Median : 8.0
##  Mean   :33.2    H7     : 27   Mean   :2.66          Mean   :13.1
##  3rd Qu.:40.0    H6     : 26   3rd Qu.:4.00          3rd Qu.:23.0
##  Max.   :54.0    H8     : 26   Max.   :4.00          Max.   :66.0
##                  (Other):172
##  mesh70mm_Raised.count mesh70mm_Total.catch
##  Min.   :   0          Min.   :203
##  1st Qu.:  34          1st Qu.:306
##  Median : 161          Median :411
##  Mean   : 334          Mean   :417
##  3rd Qu.: 521          3rd Qu.:490
##  Max.   :2251          Max.   :618
##
##  mesh70mm_Overall.Sampling.Ratio mesh80mm_Net.position mesh80mm_Count
##  Min.   :0.0231                  Min.   :1.00          Min.   : 0.0
##  1st Qu.:0.0311                  1st Qu.:2.00          1st Qu.: 2.0
##  Median :0.0407                  Median :2.00          Median : 7.0
##  Mean   :0.0496                  Mean   :2.44          Mean   :12.5
##  3rd Qu.:0.0583                  3rd Qu.:3.00          3rd Qu.:20.0
##  Max.   :0.0923                  Max.   :4.00          Max.   :61.0
##
##  mesh80mm_Raised.count mesh80mm_Total.catch
##  Min.   :   0          Min.   :166
##  1st Qu.:  32          1st Qu.:265
##  Median : 158          Median :407
##  Mean   : 278          Mean   :388
##  3rd Qu.: 406          3rd Qu.:459
##  Max.   :1621          Max.   :635
##
##  mesh80mm_Overall.Sampling.Ratio
##  Min.   :0.0231
##  1st Qu.:0.0316
##  Median :0.0482
##  Mean   :0.0545
##  3rd Qu.:0.0687
##  Max.   :0.1059
##
```

Get the empirical proportion 80/(70 + 80) at length. Note that the length-specific CIs do not reflect the non-independence of the observations across lengths at the haul level are therefore not plotted.

```
## vector of carapace lengths
cl.vec <- unique(neph.7080.cast$Carapace.length)
cl.vec <- cl.vec[order(cl.vec)]

count.df <- data.frame(Carapace.length = cl.vec, prop.80 = NA)

for(i in 1:dim(count.df)[1]){
  sub.dat <- subset(neph.7080.cast, Carapace.length == count.df$Carapace.length[i])
  ##
  if(dim(sub.dat)[1] > 1){
    ##
    count.df$prop.80[i] <- with(sub.dat, round(sum(mesh80mm_Raised.count)) /
                                (round(sum(mesh80mm_Raised.count)) + round(sum(mesh70mm_Ra
    rm(list = c("sub.dat"))
  }
}
```

Plot the data (Figure 1)

```
with(count.df, plot(Carapace.length, prop.80, ylim = c(0, 1), pch = 19,
                    xlab = "Carapace length (mm)",
                    ylab = "Proportion (N80mm/(N70mm + N80mm))",
                    bty = "L"))
abline(h = 0.5, lty = 2)
```

2 Models

A catch comparison binomial Generalized Additive/Linear Mixed Model is suitable choice for these count data where we are interested in estimating how the proportion changes with carapace length. We first try a model with only carapace length as an explanatory variable with haul random effects.

```
library(mgcv)

neph.7080.cast$dum <- 1

## no length effect
gamm.null <- gam(cbind(mesh80mm_Count, mesh70mm_Count) ~ 1 +
                   s(fHAUL, bs="re", by = dum),
                   offset =
                   log(mesh80mm_Overall.Sampling.Ratio /
                       mesh70mm_Overall.Sampling.Ratio),
                   family = binomial,
                   data = neph.7080.cast)

gamm.alt <- gam(cbind(mesh80mm_Count, mesh70mm_Count) ~
```

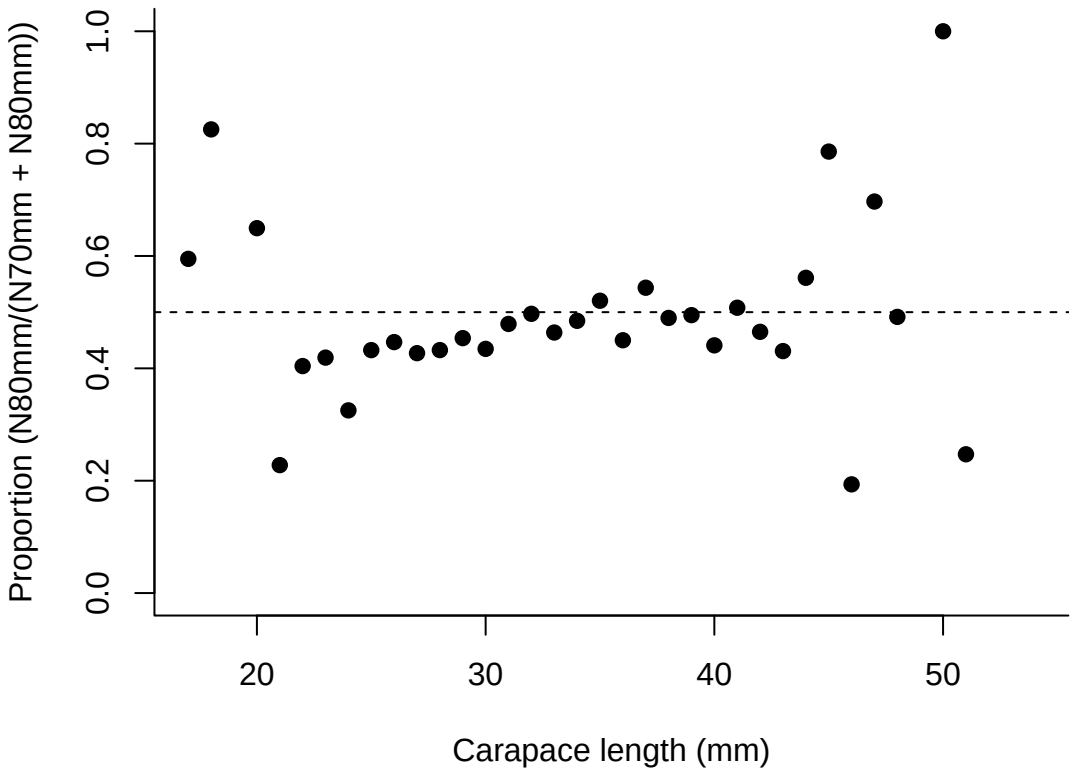

Figure 1: Proportion of Nephrops raised numbers retained in the 80mm over the sum of the 80mm and 70mm meshes.

```
                s(Carapace.length, k = 5) +
                s(fHAUL, bs="re", by = dum),
                offset =
                log(mesh80mm_Overall.Sampling.Ratio /
                    mesh70mm_Overall.Sampling.Ratio),
                family = binomial,
                data = neph.7080.cast)

## likelihood ratio test for the significance of carapace length
anova(gamm.null, gamm.alt, test = "Chisq")


## Analysis of Deviance Table
##
## Model 1: cbind(mesh80mm_Count, mesh70mm_Count) ~ 1 + s(fHAUL, bs = "re",
##     by = dum)
## Model 2: cbind(mesh80mm_Count, mesh70mm_Count) ~ s(Carapace.length, k = 5) +
##     s(fHAUL, bs = "re", by = dum)
##   Resid. Df Resid. Dev    Df Deviance Pr(>Chi)
## 1       324        438
## 2       323        430 0.998     8.41   0.0037 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Plot the predictions from this model.

```
mean.catch <- mean(c(unique(neph.7080.cast$mesh70mm_Total.catch),
                     unique(neph.7080.cast$mesh80mm_Total.catch)))

## data frame to predictfor
pred.df <- data.frame(Carapace.length = cl.vec,
                      fHAUL = "H1",
                      dum = 0,
                      mesh80mm_Overall.Sampling.Ratio = 1,
                      mesh70mm_Overall.Sampling.Ratio = 1,
                      mesh70mm_Total.catch = mean.catch,
                      mesh80mm_Total.catch = mean.catch
                      )

pred.gamm.alt <- predict(gamm.alt, newdata = pred.df, se.fit = TRUE)

## predicted proportions and confidence intervals
pred.df$pred.prop <- plogis(pred.gamm.alt$fit)
pred.df$lwr.prop <- plogis(pred.gamm.alt$fit - qnorm(0.975) * pred.gamm.alt$se.fit)
pred.df$upr.prop <- plogis(pred.gamm.alt$fit + qnorm(0.975) * pred.gamm.alt$se.fit)


with(count.df, plot(Carapace.length, prop.80, ylim = c(0, 1), pch = 19,
                    xlab = "Carapace length (mm)",
                    ylab = "Proportion (N80mm/(N70mm + N80mm))",
                    bty = "L"))
```

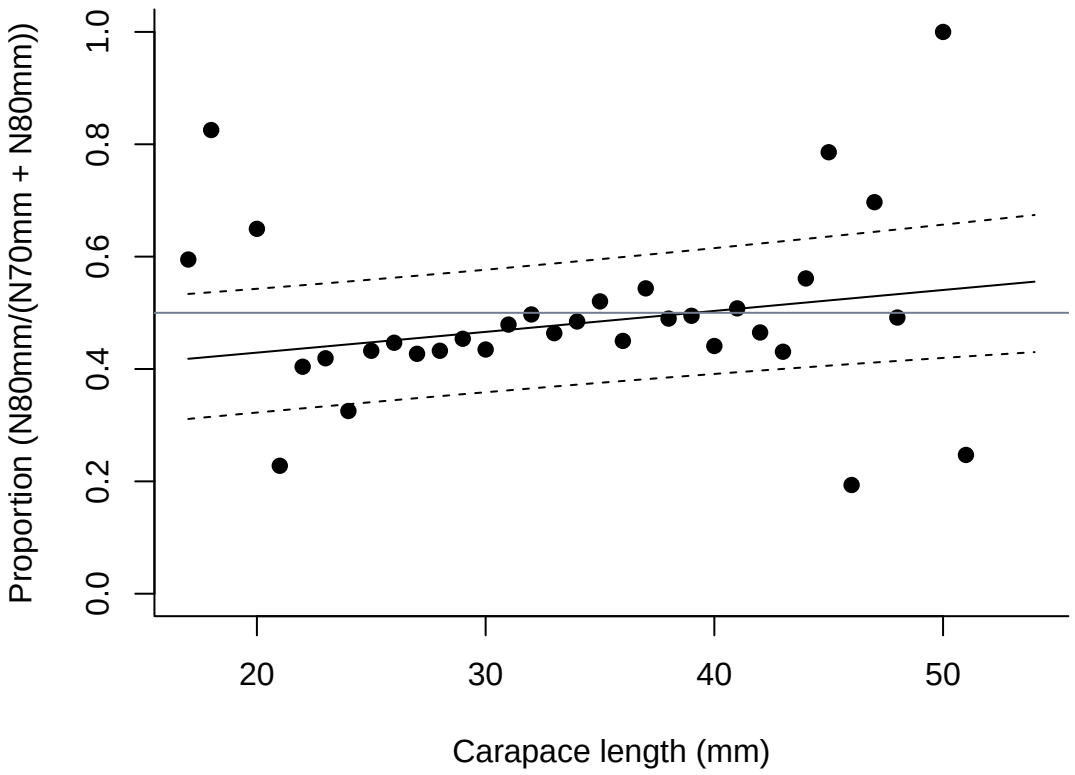

Figure 2: Predicted proportion from binomial GLMM without covariates.

```
with(pred.df, lines(Carapace.length, pred.prop))
with(pred.df, lines(Carapace.length, lwr.prop, lty = 2))
with(pred.df, lines(Carapace.length, upr.prop, lty = 2))
abline(h = 0.5, col = "slategrey")
```

The cause of the wide confidence intervals (Figure 2) is the large amount of inter-haul variability in the proportion retained in the 80mm (Figure 3)

```
## Get the proportion at length by haul
haul.count.df <- expand.grid(Carapace.length = cl.vec,
                             fHAUL = unique(neph.7080.cast$fHAUL))

## order the levels
haul.count.df$fHAUL <- factor(as.character(haul.count.df$fHAUL),
                              levels = c(paste("H", 1:12, sep = ""), "H14"))
```

```
haul.count.df$prop.80 <- NA
haul.count.df$lwr <- NA
haul.count.df$upr <- NA

for(i in 1:dim(haul.count.df)[1]){
  sub.dat <- subset(neph.7080.cast,
                    Carapace.length == haul.count.df$Carapace.length[i] &
                    fHAUL == haul.count.df$fHAUL[i])
  ##
  ##if((sub.dat$mesh80mm_Raised.count + sub.dat$mesh70mm_Raised.count) > 0){
  if(dim(sub.dat)[1] > 0){
    btest <- with(sub.dat,
                  binom.test(x = round(mesh80mm_Raised.count),
                             n = round(mesh80mm_Raised.count + mesh70mm_Raised.count)))
    ##
    haul.count.df$prop.80[i] <- btest$estimate
    haul.count.df$lwr[i] <- btest$conf.int[1]
    haul.count.df$upr[i] <- btest$conf.int[2]
    ##
    rm(list = c("sub.dat", "btest"))
  }
}

## get predictions at the HAUL level from model
haul.count.df$dum <- 1
haul.count.df$pred.prop <- plogis(predict(gamm.alt, newdata = haul.count.df))


library(ggplot2)

blue2red <- colorRampPalette(c("darkblue", "white", "red"))

ggplot(haul.count.df, aes(x = Carapace.length, y = prop.80)) +
  geom_point(aes(colour = fHAUL)) +
  geom_line(data = haul.count.df, aes(x = Carapace.length, y = pred.prop,  colour = fHAUL))
  scale_colour_manual(values = blue2red(13)) + ylab("Proportion in 80mm")


## Warning:  Removed 118 rows containing missing values (geom_point).
```

We can take a look at additional measured covariates to see if these relate to the haul-level variability (random effects in the model above) (Figure 4).

```
##
ranef.df <- data.frame(fHAUL = levels(neph.7080.cast$fHAUL),
                       ranef = coef(gamm.alt)[-c(1:5)])
##
covar.names <- c("fHAUL", "mesh70mm_Net.position", "mesh70mm_Total.catch",
                 "mesh80mm_Net.position", "mesh80mm_Total.catch")
##
covar.df <- unique(neph.7080.cast[, covar.names])
```

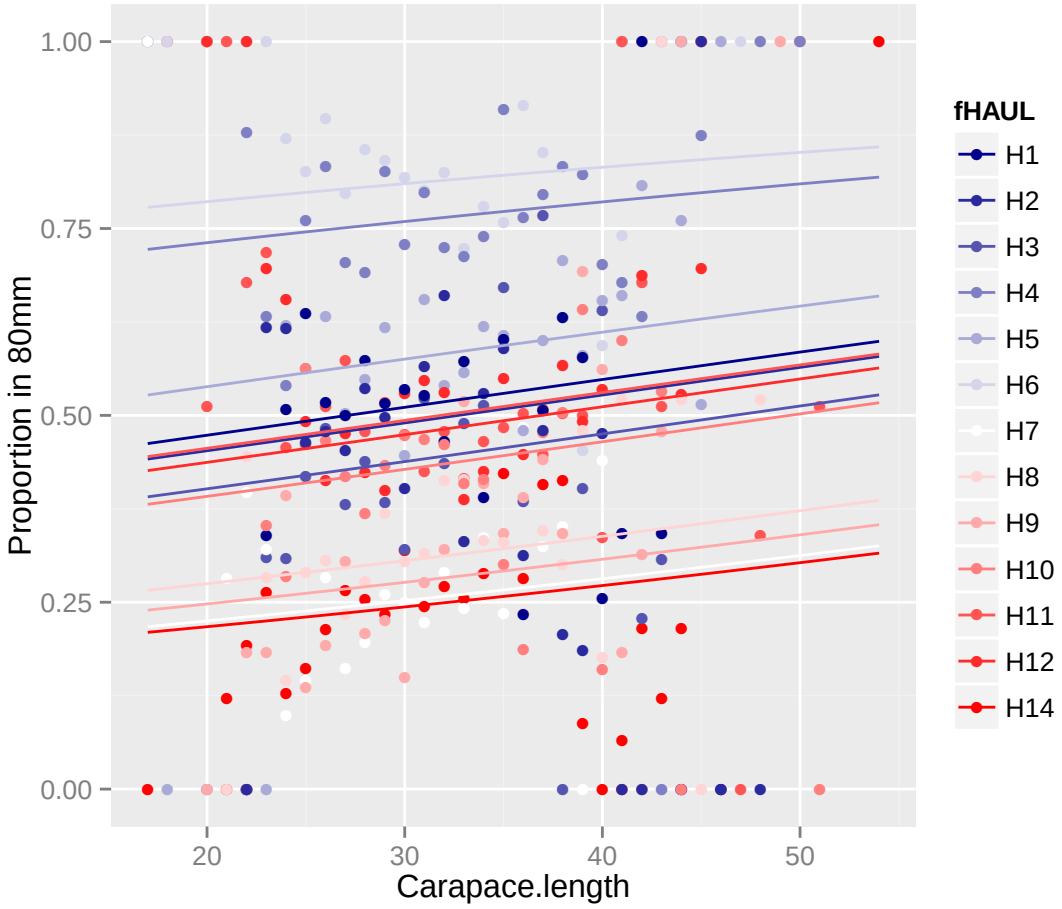

Figure 3: Observed proportion in the 80mm by haul. Note the wide variability of the proportion with some having much higher or lower proportions. Fitted lines come from the GLMM with carapace length only.

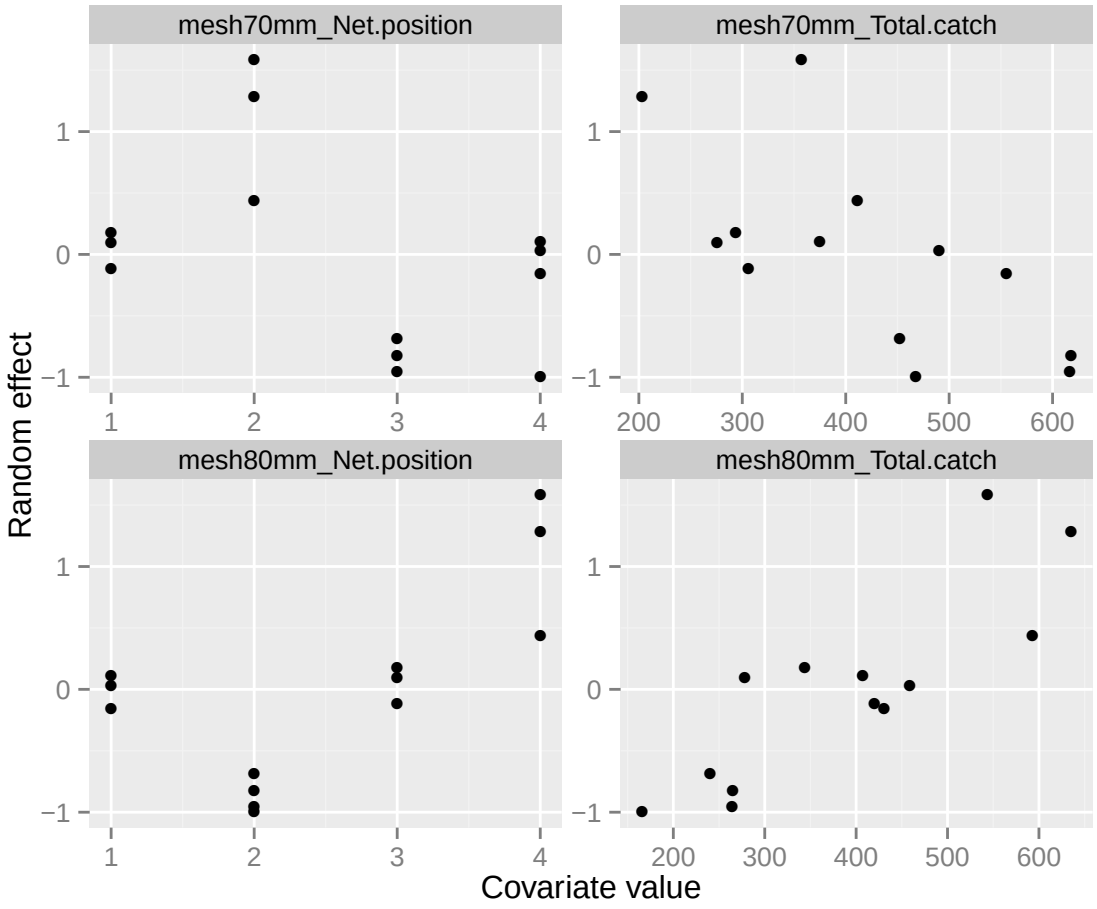

Figure 4: Relationship between the random effects of the carapace length only model and measured covariates .

```
##
ranef.df <- merge(ranef.df, covar.df)
## convert to long format for plotting
ranef.df <- melt(ranef.df, id = c("fHAUL", "ranef"))
##
ggplot(ranef.df, aes(x = value, y = ranef)) +
  geom_point() +
  facet_wrap(~ variable, scales = "free") +
  xlab("Covariate value") +
  ylab("Random effect")
```

There are some strong relationships between the random effects and measured covariates (Figure 4). It is best to include these measured variables in the model as fixed effects.

```
## including additional covariates
## check identifiability
neph.7080.cast$fmesh80mm_Net.position <-
  factor(paste("pos",
               neph.7080.cast$mesh80mm_Net.position, sep = ""))

## using log of total catch weights - return to this
neph.7080.cast$log.mesh80mm_Total.catch <- log(neph.7080.cast$mesh80mm_Total.catch)
neph.7080.cast$log.mesh70mm_Total.catch <- log(neph.7080.cast$mesh70mm_Total.catch)

## Note should return to this warning later
## fits okay in gam but glmm used for effects package
glmm.alt.covar <- glmer(cbind(mesh80mm_Count, mesh70mm_Count) ~
                        ##I(log(mesh80mm_Total.catch / mesh70mm_Total.catch)) * Carapace.le
                        log.mesh80mm_Total.catch + log.mesh70mm_Total.catch +
                        Carapace.length +
                        fmesh80mm_Net.position +
                        (1 | fHAUL),
                        offset =
                        log(mesh80mm_Overall.Sampling.Ratio /
                            mesh70mm_Overall.Sampling.Ratio),
                        family = binomial,
                        data = neph.7080.cast)


## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,
:  Model failed to converge with max|grad| = 0.00492158 (tol = 0.001, component
2)
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,
:  Model is nearly unidentifiable:  large eigenvalue ratio
##  - Rescale variables?

## use effects package to get prediction for model with net position
## set predictor variables
xlevels <- list(Carapace.length = cl.vec
                ##mesh80mm_Total.catch = mean.catch,
                ##mesh70mm_Total.catch = mean.catch
                )
## if we wanted to set the proportions of net positions equivalent
## otherwise set to the proportion observed in the data
##given.values <- c("fmesh80mm_Net.positionpos2" = 1/4,
##                  "fmesh80mm_Net.positionpos3" = 1/4,
##                  "fmesh80mm_Net.positionpos4" = 1/4
##                  )

##cl.effect <- effect("Carapace.length", glmm.alt.covar, xlevels = xlevels, offset = 0, gi
cl.effect <- effect("Carapace.length", glmm.alt.covar, xlevels = xlevels, offset = 0)
```

Finally plot the effect of carapace length with the other variables set to their mean in the data (Figure 5).

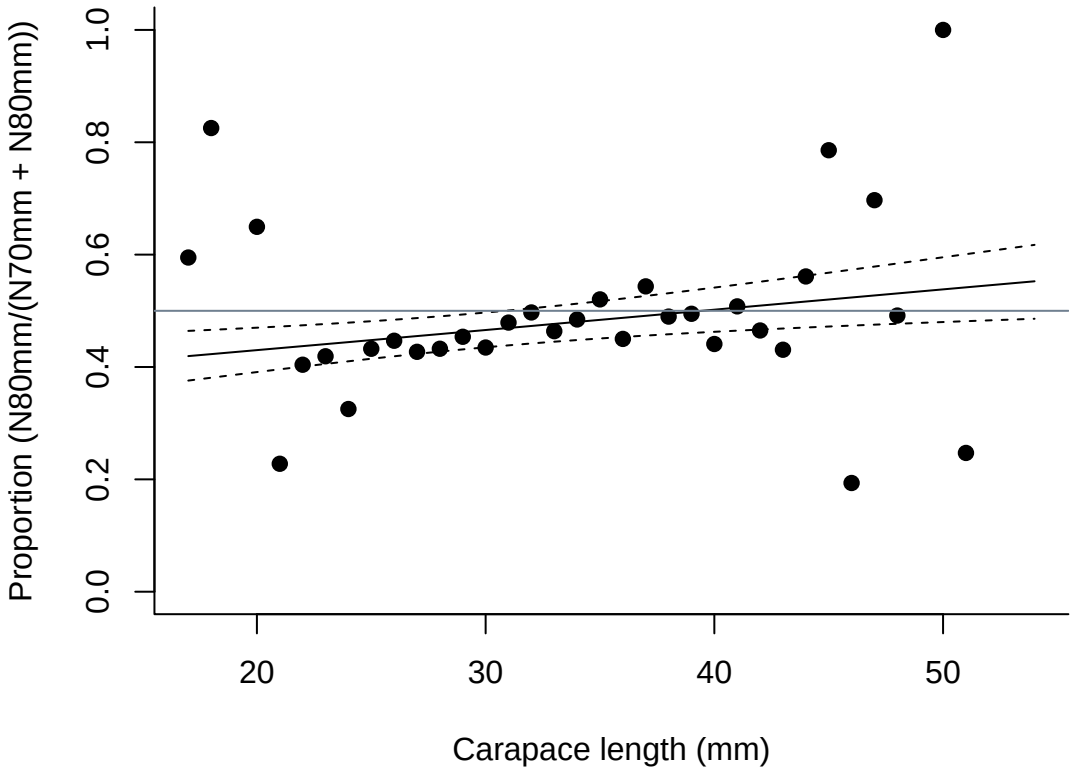

Figure 5: Predicted proportion from binomial GLMM with covariates. Note in the predictions the bulk weights are set to their mean and the net positions to their proportional occurence in the data.

```
with(count.df, plot(Carapace.length, prop.80, ylim = c(0, 1), pch = 19,
                    xlab = "Carapace length (mm)",
                    ylab = "Proportion (N80mm/(N70mm + N80mm))",
                    bty = "L"))

## effects prediction
lines(cl.vec, plogis(cl.effect$fit[, 1]))
lines(cl.vec, plogis(cl.effect$lower[, 1]), lty = 2)
lines(cl.vec, plogis(cl.effect$upper[, 1]), lty = 2)
abline(h = 0.5, col = "slategrey")
```

Finally test length effect in covariate model

```
glmm.alt.covar.nolength <- glmer(cbind(mesh80mm_Count, mesh70mm_Count) ~
                         log.mesh80mm_Total.catch + log.mesh70mm_Total.catch +
```

```
                        fmesh80mm_Net.position +
                        (1 | fHAUL),
                        offset =
                        log(mesh80mm_Overall.Sampling.Ratio /
                            mesh70mm_Overall.Sampling.Ratio),
                        family = binomial,
                        data = neph.7080.cast)

## likelihood ratio test
anova(glmm.alt.covar.nolength, glmm.alt.covar)


## Data: neph.7080.cast
## Models:
## glmm.alt.covar.nolength: cbind(mesh80mm_Count, mesh70mm_Count) ~ log.mesh80mm_Total.cato
## glmm.alt.covar.nolength:     log.mesh70mm_Total.catch + fmesh80mm_Net.position + (1 |
## glmm.alt.covar.nolength:     fHAUL)
## glmm.alt.covar: cbind(mesh80mm_Count, mesh70mm_Count) ~ log.mesh80mm_Total.catch +
## glmm.alt.covar:     log.mesh70mm_Total.catch + Carapace.length + fmesh80mm_Net.position
## glmm.alt.covar:     (1 | fHAUL)
##                         Df  AIC  BIC logLik deviance Chisq Chi Df
## glmm.alt.covar.nolength  7 1421 1448   -703     1407
## glmm.alt.covar           8 1415 1445   -699     1399  8.09      1
##                         Pr(>Chisq)
## glmm.alt.covar.nolength
## glmm.alt.covar              0.0044 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


## significant effect of carapace length
```