# Box 5.2 Estimating biomass using a length weight relationship

Commonly a subset of fish captured are measured for length and weight and the remaining fish are measured for length only. This is advantageous as the variance for total biomass can be reduced by accounting for variability in weight among fish by exploiting the relationship between length and weight. The length weight relationship predicts the expected (i.e., mean) weight of fish where only length was measured and the estimate of total biomass then estimated by summing the predicted weights. The variance can be estimated as well and 95% confidence intervals calculated using the variance of the predicted weight.

## Setting up the data

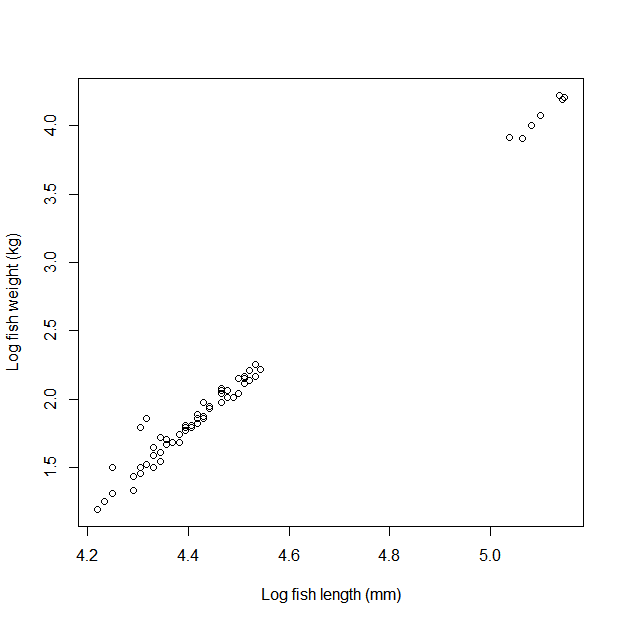
The data is 2 columns, length and weight for common carp in mm and kg respectively. There were xx common carp captured. Length was measured on all captured common carp and weight was measured for xx fish.

ylb\_lw<-read.csv("yellow-bass.csv")  
# look at the first few rows of the data  
head(ylb\_lw)  
# plot the data  
  
plot(weight\_g~length\_mm,data=ylb\_lw,  
 xlab="Length (mm)",  
 ylab="Weight (g)")  
# add a rug plot to see distribution of lengths  
# that do not have lengths to verify adequate coverage  
rug(ylb\_lw[which(!is.na(ylb\_lw$weight\_g)),]$length\_mm)

## Fitting the length weight relationship

The expected weight for a given length is estimated by log linear regression in the R code chunk below. First the length and weight data are log transformed. Then a linear model is fit to the log transformed data. The data is visualized using the plot function.

ylb\_lw$ln\_length<-log(ylb\_lw$length)  
ylb\_lw$ln\_weight<-log(ylb\_lw$weight)  
plot(ln\_weight~ln\_length,data=ylb\_lw,  
 xlab="Log fish length (mm)",  
 ylab="Log fish weight (kg)")



exp(sum(na.omit(ylb\_lw$ln\_weight)) )

## [1] 3.803478e+56

(sum(na.omit(ylb\_lw$weight)) )

## [1] 771.4

sum(ylb\_lw$ln\_weight)

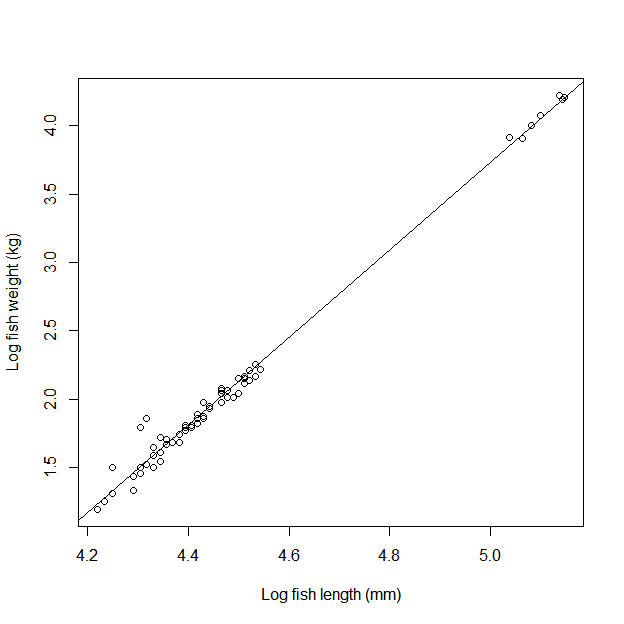
## [1] NA

A linear regression is fit to the log transformed length and weight data using lm. And the model output summarized and model fit added to the plot of raw data to visualize model fit.

fit<- lm(ln\_weight~ln\_length,data=ylb\_lw)  
summary(fit)

##   
## Call:  
## lm(formula = ln\_weight ~ ln\_length, data = ylb\_lw)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.12195 -0.04186 -0.01244 0.01789 0.31261   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -12.31097 0.17976 -68.48 <2e-16 \*\*\*  
## ln\_length 3.20896 0.04006 80.10 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.07431 on 61 degrees of freedom  
## (41 observations deleted due to missingness)  
## Multiple R-squared: 0.9906, Adjusted R-squared: 0.9904   
## F-statistic: 6416 on 1 and 61 DF, p-value: < 2.2e-16

plot(ln\_weight~ln\_length,data=ylb\_lw,  
 xlab="Log fish length (mm)",  
 ylab="Log fish weight (kg)")  
abline(coef(fit)[1],coef(fit)[2])



The fitted model can be used to calculate the expected weight given the length for fish that were not measured for length using the predict function. Using the argument se.fit=TRUE in the predict function returns the standard error for each estimated mean weight in the data set. The standard error associated with each estimated mean will be used to estimate the variance for the biomass estimate.

# calculate the estimated mean weight for a fish of  
# a given length  
mean\_weight<-predict(fit,  
 newdata=data.frame(ln\_length=ylb\_lw$ln\_length),  
 se.fit = TRUE)  
# get the need bits from the list returned  
mean\_weight<-data.frame(mean\_weight=mean\_weight$fit,  
 std\_error=mean\_weight$se.fit)  
  
# take a look at the predicted mean weights and standard errors  
head(mean\_weight)

## mean\_weight std\_error  
## 1 1.750786 0.010165305  
## 2 1.790649 0.009981973  
## 3 1.830023 0.009822309  
## 4 2.128747 0.009393111  
## 5 1.543684 0.011423450  
## 6 2.128747 0.009393111

x<-rlnorm(10,10,0.3)  
sd\_log\_scale= sqrt(var(x))/mean(x)  
  
sd\_log\_scale\*mean=sqrt(var(x))  
(sd\_log\_scale\*mean)^2=var(x) # variance for each predication

The biomass estimate is now the sum of the mean weights given the measured lengths. Recall though weight values remain on log scale.

ln\_biomass<- sum(mean\_weight$mean\_weight)

The variance associated with the biomass estimate is the sum of the standard errors squared and is calculated in the R code chunk below.

ln\_biomass\_variance<- sum(mean\_weight$std\_error)^2

The 95% confidence intervals for the biomass estimate are calculated as previously done for the simple random sample and the stratified random sample, but in this case the values are still on log scale.

# t value   
# note the lower value is negative and the upper is positive  
t\_value<-qt(p=c(0.025,0.975), df=sum(sample\_size$weight)-1)  
lower\_95\_ci<- ln\_biomass+t\_value[1]\*sqrt(ln\_biomass\_variance)  
upper\_95\_ci<- ln\_biomass+t\_value[2]\*sqrt(ln\_biomass\_variance)

The real biomass and 95% confidence interval estimates can be calculated by exponentiating each value to move from log scale back to kilograms.

# biomass estimate  
exp(ln\_biomass)  
# lower 95% confidence interval for biomass  
exp(lower\_95\_ci)  
# upper 95% confidence interval for biomass  
exp(upper\_95\_ci)

# References

Colvin, M. E., C. L. Pierce, T. W. Stewart, and S. Grummer. 2012. Strategies to control a common carp (*Cyprinus carpio*) population by pulsed commercial harvest. North American Journal of Fisheries Management 32:1251-1264.