# Overview

Estimating biomass for a simple random sample requires a sampling frame (i.e., how many possible samples units are there) and a random sample of from the sampling frame. A common example encountered in fisheries is estimating biomass of a large catch from a simple random sample.

This box estimates biomass and associated uncertainties for a simple random sample, a stratified random sample, and a case where a random subset of a sample were measured for weight and length and then lengths were taken on the remaining sample.

# Estimating biomass for a simple random sample

The data used in this example is from commercial common carp *Cyprinus carpio* harvest in Clear Lake Iowa (Colvin xxxx). In short, commercial harvesters capture common carp to remove of the system and the biomass removed is estimated from a simple random sample of the common carp harvested. The sampling frame was determined by counting the common carp removed. A simple random sample of xx fish was taken to estimate the biomass removed.

## Setting up the data

The R code chunk below sets up the sampling frame and the data for the simple random sample.

# specify the sampling frame  
sampling\_frame<- 4000  
# weight data in kilograms for the  
# simple random sample  
data<-data.frame(weight=c(rnorm(50,1.2,0.2)))  
sample\_size<-length(data$weight)

## Estimating the mean and variance of the sample

The R code chunk below estimates the mean and variance of the simple random sample

# estimate the mean weight for the sample  
mean\_weight<-mean(data$weight)  
# estimate the variance of the sample  
variance\_weight<- var(data$weight)

## Estimating total biomass

The R code chunk below estimates the total biomass of the common carp captured.

total\_biomass<- sampling\_frame\*mean\_weight  
# return the biomass estimate  
total\_biomass

## [1] 5058.704

The R code chunk below uses the variance equation provided in table xx to estimate the variance for the estimate of total biomass.

variance\_total\_biomass<- sampling\_frame^2\*(variance\_weight/sample\_size)  
variance\_total\_biomass

## [1] 15705.68

std\_dev\_total\_biomass<- sqrt(variance\_total\_biomass)  
std\_dev\_total\_biomass

## [1] 125.3223

The 95% confidence intervals can be calculated given a t value the estimate of total biomass and the standard deviation for the biomass estimate. The t value is returned using the qt function. The probabilities for the lower and upper are specified and the degrees of freedom is the sample size. The lower and upper 95% confidence interval are calculated by adding the product of the lower and upper t value and the standard deviation for the total biomass estimate.

# t value   
# note the lower value is negative and the upper is positive  
t\_value<-qt(p=c(0.025,0.975), df=sample\_size)  
lower\_95\_ci<- total\_biomass+t\_value[1]\*std\_dev\_total\_biomass  
upper\_95\_ci<- total\_biomass+t\_value[2]\*std\_dev\_total\_biomass

The values below are the lower 95% confidence interval, the total biomass and the upper 95% confidence interval respectively.

lower\_95\_ci

## [1] 4806.986

total\_biomass

## [1] 5058.704

upper\_95\_ci

## [1] 5310.421

# Estimating biomass for a stratified random sample

There are instances where stratification is necessary to estimate biomass. Continuing with the common carp harvest from Clear Lake, Iowa there were harvest events that occurred in the spring and the fall. It makes sense for annual reporting to report the estimated total biomass removed but there are differences in fish weight that should be accounted for in the estimate. Specifically, fish in the spring, after ice out, tend weigh less than fish in the fall after a season of growth. Therefore the time of harvest (i.e., spring, autumn) can be used to account for the weight differences that effect sample. Similar to the simple random sample the data required is a sampling frame and the sample random sample within the frame, but for each strata. For this example there are 2 strata, spring and autumn. In application this type of estimate can be used if there are weight differences between sex and among locations or life stages.

## Setting up the data

The R code chunk below sets up the sampling frame and the data for the stratified random sample.

# specify the sampling frame  
sampling\_frame<- data.frame(season=c("spring","autumn"),  
 N=c(2500,4000))  
sampling\_frame$weight<- sampling\_frame$N/sum(sampling\_frame$N)  
# weight data in kilograms for the  
# stratified sample  
data<-data.frame(season=c(rep("spring",45),rep("autumn", 60)),  
 weight=c(rnorm(45,1.3,0.1),rnorm(60,2.1,0.2)))

## Estimating the mean and variance of the sample

The R code chunk below estimates the mean and variance of the simple random sample

# estimate the mean weight for each stratum  
mean\_weight<-aggregate(weight~season,data,FUN=mean)  
sample\_size<-aggregate(weight~season,data,FUN=length)  
# estimate the variance of the sample  
variance\_weight<- aggregate(weight~season,data,FUN=var)

## Estimating total biomass for the stratified sample

The R code chunk below estimates the weighted mean of the common carp captured in the spring and the autumn. Noe that sampling\_frame is not in the same order as the mean\_weight and variance\_weight. The order function sorts the data.frame so it matches the output of the aggregate function.

sampling\_frame<-sampling\_frame[order(sampling\_frame$season),]  
weighted\_mean\_weight<- mean\_weight$weight\*sampling\_frame$weight

The R code chunk below calculated the overall mean and the variance as the sum of the weighted means and variances.

mean\_weight<- sum(weighted\_mean\_weight)  
total\_biomass<- sum(sampling\_frame$N)\*mean\_weight

variance\_weight<- sum(sampling\_frame$N)^2\*variance\_weight$weight/sample\_size$weight  
std\_dev\_total\_biomass<-sqrt(sum(variance\_weight))

# 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))  
lower\_95\_ci<- total\_biomass+t\_value[1]\*std\_dev\_total\_biomass  
upper\_95\_ci<- total\_biomass+t\_value[2]\*std\_dev\_total\_biomass  
lower\_95\_ci

## [1] 11179.58

total\_biomass

## [1] 11627.54

upper\_95\_ci

## [1] 12075.5

# Estimating biomass using a length weight relationship

A common situation occurring in inland fisheries is the use of a length weight relationship to estimate biomass. In this case, a random sample of fish is used to develop a length weight relationship and the remaining fish are measured for length but not weighed. The length weight relationship is then used to predict the weight of fish where only length was measured and the estimate of total biomass then estimated.

x<-data.frame(W=rnorm(50,1.2,0.1),group=rep(1,50))  
N<-5000  
y<- lm(W~group,x)  
mean(x$W)  
sqrt(var(x$W)/50)# standard error same as se from predict  
pp<-predict(y,newdata=data.frame(group=rep(1,N)),se.fit = TRUE)  
sum(pp$se.fit)^2# variance is the sum of predicted ses squared  
# variance for biomass   
N^2\*(var(x$W)/50)  
sum(pp)

Since the sample biomass is already known the variance for that value is 0 (i.e., it is known with certainty. The remaining uncertainty is due to the estimation of the mean weight for each fish where length was measured but no weight was taken.

Now the biomass is back transformed from log scale to get the estimated biomass and 95% confidence intervals.

# variance for all fish   
total\_variance<- sum(pp$se.fit)^2  
exp(mean)  
exp(lwr)  
exp(upr)

The R code chunk below uses the variance equation provided in table xx to estimate the variance for the estimate of total biomass.

x<-data.frame(W=rnorm(50,1.2,0.1),group=rep(1,50))  
N<-5000  
y<- lm(W~group,x)  
mean(x$W)  
sqrt(var(x$W)/50)# standard error same as se from predict  
pp<-predict(y,newdata=data.frame(group=rep(1,N)),se.fit = TRUE,interval = "confidence")  
sum(pp$se.fit)^2# variance is the sum of predicted ses squared  
# variance for biomass   
N^2\*(var(x$W)/50)  
sum(pp)