# Box 5.1. Estimating biomass from a random sample and a stratified random sample

This box demonstrates how to estimate biomass,associated uncertainties, and 95% confidence intervals for a simple random sample, a stratified random sample, and for length weight data. Common to the 3 approaches is the use of a sampling frame (how many fish are there) and random sample of fish from the sampling frame. In short, the mean weight of fish in the sample is multiplied by the total number of fish to estimate biomass where the mean can be calculated as a simple random sample, from a stratified sample, or accounting for variation among all the fish in the sampling frame with a covariate like fish length.

The data used in this example is from commercial common carp *Cyprinus carpio* harvest in Clear Lake, Iowa (Colvin 2012). Commercial harvesters capture common carp in the spring and fall annually to remove biomass from the system. The biomass removed from the system can be estimated using the 2 approaches. The sampling frame is the number of common carp harvested for each occasion. The sampling unit is individual common carp.

# Estimating biomass for a simple random sample

## Setting up the data

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

# weight data in grams for the  
# simple random sample  
carp\_weights<-data.frame(season=c(rep("spring", 180)),  
 weight=c(0.611, 0.595, 0.601, 0.629, 0.63,   
 0.635, 0.65, 0.65, 0.65, 0.654, 0.63, 0.64, 0.67, 0.63, 0.648, 0.65,   
 0.634, 0.671, 0.759, 0.645, 0.645, 0.701, 0.66, 0.66, 0.666, 0.669,   
 0.674, 0.675, 0.677, 0.678, 0.68, 0.681, 0.71, 0.654, 0.662, 0.665,   
 0.671, 0.68, 0.682, 0.682, 0.633, 0.662, 0.665, 0.675, 0.65, 0.667,   
 0.681, 0.674, 0.674, 0.681, 0.706, 0.71, 0.698, 0.724, 0.663, 0.674,   
 0.694, 0.71, 0.72, 0.682, 0.685, 0.708, 0.692, 0.707, 0.712, 0.685,   
 0.687, 0.69, 0.702, 0.705, 0.73, 0.685, 0.694, 0.724, 0.725, 0.735,   
 0.745, 0.704, 0.715, 0.691, 0.715, 0.731, 0.747, 0.72, 0.725, 0.71,   
 0.713, 0.729, 0.73, 0.73, 0.705, 0.722, 0.728, 0.735, 0.731, 0.749,   
 0.776, 0.705, 0.751, 0.772, 0.754, 0.757, 0.76, 0.76, 0.79, 0.725,   
 0.739, 0.742, 0.743, 0.753, 0.755, 0.76, 0.765, 0.74, 0.745, 0.731,   
 0.745, 0.78, 0.72, 0.745, 0.755, 0.767, 0.78, 0.768, 0.79, 0.715, 0.766,   
 0.77, 0.79, 0.8, 0.823, 0.751, 0.779, 0.78, 0.795, 0.75, 0.766, 0.786,   
 0.684, 0.75, 0.757, 0.765, 0.77, 0.783, 0.79, 0.785, 0.799, 0.8, 0.76,   
 0.78, 0.81, 0.79, 0.77, 0.79, 0.78, 0.79, 0.79, 0.77, 0.797, 0.808,   
 0.819, 0.805, 0.805, 0.83, 0.82, 0.838, 0.846, 0.83, 0.818, 0.83, 0.814,   
 0.82, 0.821, 0.852, 0.796, 0.84, 0.833, 0.862, 0.86, 0.859))   
  
sample\_size<-length(carp\_weights$weight)  
# specify the sampling frame  
# 3367 common carp captured but not weighed  
# therefore the sampling frame is 2877+ plus the number of carped weighed  
sampling\_frame<- 3367+nrow(carp\_weights)

## 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(carp\_weights$weight)  
# estimate the variance of the sample  
variance\_weight<- var(carp\_weights$weight)

## Estimating total biomass

The R code chunk below estimates the total biomass of the common carp captured as the product of the mean weight and the number of fish harvested (i.e., sampling frame).

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

## [1] 2589.271

The R code chunk below uses the variance equation provided in table 5.1 to estimate the variance for the estimated total biomass. The standard deviation is calculated as the square root of the variance.

# estimated variance for biomass estimate  
variance\_biomass<- sampling\_frame^2\*(variance\_weight/sample\_size)  
variance\_biomass

## [1] 253.1798

The 95% confidence intervals are calculated by adding the appropriate t-value given the sample size as the degrees of freedom. For a 95% confidence interval the t-value is calculated using the qt function given probabilities of 0.025 and 0.975. 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)  
# calculate the lower and upper 95% confidence interval for  
# the biomass estimate  
lower\_95\_ci<- biomass+t\_value[1]\*sqrt(variance\_biomass)  
upper\_95\_ci<- biomass+t\_value[2]\*sqrt(variance\_biomass)

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

lower\_95\_ci

## [1] 2557.873

biomass

## [1] 2589.271

upper\_95\_ci

## [1] 2620.668

# 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. The stratum weight is calculated because it is need to calculate the weighted mean fish weight. The stratum weight is the number of sampling units in each stratum divided by the total number of sampling units.

# weight data in kilograms for the carp harvested in the autumn  
# 41 fish with weights and another 743 without weights  
carp\_weights<-rbind(carp\_weights,data.frame(season=c(rep("autumn", 41)),  
 weight=c(0.652, 0.632, 0.668, 0.684, 0.69, 0.695, 0.703, 0.698, 0.692,   
 0.714, 0.695, 0.74, 0.715, 0.729, 0.742, 0.713, 0.715, 0.745, 0.759,   
 0.761, 0.768, 0.725, 0.809, 0.775, 0.779, 0.798, 0.813, 0.795, 0.811,   
 0.811, 0.79, 0.801, 0.814, 0.835, 0.834, 0.83, 0.82, 0.853, 0.88, 0.88,   
 0.915)) )  
  
# specify the sampling frame  
sampling\_frame<- data.frame(  
 season=c("spring","autumn"),  
 N=c(3367+180,743+41))  
# calculate the stratum weight, do not confuse with fish weight  
sampling\_frame$stratum\_weight<- sampling\_frame$N/sum(sampling\_frame$N)

## Estimating the mean and variance of the sample

The R code chunk below estimates the mean and variance of the simple random sample taken within each stratum using the aggregate function.

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

## season weight  
## 1 spring 0.7299889  
## 2 autumn 0.7628780

# return the stratum specific variance  
variance\_weight

## season weight  
## 1 spring 0.003622257  
## 2 autumn 0.004470760

## 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$stratum\_weight

The R code chunk below calculates the overall mean fish weight. The biomass is then estimated as the total number of sampling units times the overall mean fish weight.

# overall mean fish weight  
mean\_weight<- sum(weighted\_mean\_weight)  
# biomass estimate  
biomass<- sum(sampling\_frame$N)\*mean\_weight  
biomass

## [1] 3278.24

The estimated variance for the biomass estimate is the sum of the stratum specific variance estimates. The stratum specific variance is the same as the biomass variance for the simple random sample, but in the case of a stratified sample applied to each stratum. The standard deviation is calculated as the square root of the variance and used in the calculation of 95% confidence intervals.

# estimated variance for biomass estimate  
variance\_biomass<- sum(sum(sampling\_frame$N)^2\*variance\_weight$weight/sample\_size$weight)

The 95% confidence intervals are calculated using the biomass estimate, a t-value, and the standard deviation as was done for the simple random sample.

# 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<- biomass+t\_value[1]\*sqrt(variance\_biomass)  
upper\_95\_ci<- biomass+t\_value[2]\*sqrt(variance\_biomass)

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

lower\_95\_ci

## [1] 3181.232

biomass

## [1] 3278.24

upper\_95\_ci

## [1] 3375.248

# 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.