Combined regulations R-Script

# Introduction

This script calculates the catch reduction when combining a seasonal closure, a bag limit and an increased minimum landing size (MLS) in comparison to the years before 2017. The script works in a similar way than the recreational cod catch estimation R-Script.

The script utilizes data from two data sources: (1) effort data from an off-site population-wide telephone-diary survey and (2) catch data from an annual on-site intercept survey. On-site data are used to calculate CPUE values of the anglers for different strata (federal\_state, SD, year, half\_year, survey\_category). In contrast to the normal assessment, in this script the catches from the closed month are removed from the CPUE calculations. In the other month the catches are reduced to the maximum amount allowed by the bag limit, but only for fishing days where the bag limit is exceeded with cod above the proposed MLS. Thereby the CPUE´s are changed, which in the end will change the total catch in numbers and tons. Off-site data are used to calculate fishing effort for different strata (federal\_state, half\_year, survey\_category). Additionally, to the changed CPUE´s the fishing effort is reduced to zero in the closed month. Catch in numbers (CANUM) is calculated by multiplying the CPUE values with the fishing effort (total fishing days). CANUM is then divided into individual length classes based on the recreational length distribution from onboard length samples. In contrast to the normal assessment, in this script length classes under the proposed MLS would be released and therefore have to be multiplied with the death rate of released cod. With a length-weight relationship from commercial catch sampling CANUM is transferred to catch in tons (CATON) and then transferred to the final CATON (catch in tons at length).

#data work  
library("readr")  
#library("plyr")  
library("dplyr")  
library("tidyr")  
library("reshape2")  
#library("Rmisc")  
  
#plots  
library("ggplot2")  
library("ggrepel")  
#library("cowplot")  
#library("ggpubr")

# Importing data

Before the national recreational catch data can be imported, a copy is made of the "\_original" data frames. This data frames are generated in the “get\_the\_data” script and include all necessary data, they should be saved in the working environment.

dat <- onsite\_ohneNullbepro\_original  
  
dat\_length <- dat\_length\_original  
   
fishing\_days <- fishing\_days\_original  
  
mass\_coefficient <- mass\_coefficient\_own\_original  
  
SnglIOR <- SnglIOR\_original  
  
all\_strata <- all\_strata\_original

# CPUE

### Season closure

~~~~~~~~~~~~~~~ different from the normal assessment ~~~~~~~~~~~~~~~~

The number of retained and released cod in the proposed closed month are removed from the calculations.

dat$retained[dat$based == "Sea" & dat$month == 2 | dat$based == "Sea" & dat$month == 3] <- NA  
dat$released[dat$based == "Sea" & dat$month == 2 | dat$based == "Sea" & dat$month == 3] <- NA

### Set the bag limit an MLS

#set the value of MLS  
MLS <- 45  
  
#set the bag limit  
bag.limit <- 10

### Calculate how many fish are above MLS

The bag limit is only activated on fishing days when more than 10 cod are above the new MLS, so it is necessary to calculate the percentage of cod caught that must be released due to the increased MLS. This amount of cod is subtracted from the actual catch, and then if the catch is still higher than the bag limit, the catch is reduced.

First, the percentage of cod under the new MLS is calculated using the length samples from the on-site survey (dat\_length data frame). This is done by calculating the total number cod sampled (n.length.total) and then the number of cod in the harvest and release range. The percentages are calculated from these numbers and it is verified that they sum to one.

range <- filter(.data = dat\_length, retained == 1)  
  
range <- range %>%  
 group\_by(SD, year, based) %>%  
 mutate(n.length.total = length(length\_class))  
  
range$range <- "harvest"  
range$range[range$length\_class < MLS] <- "release"  
  
range <- range %>%  
 group\_by(SD, year, based, range) %>%  
 mutate(n.length.harvest = length(length\_class))  
  
range$perc.length.range <- range$n.length.harvest/range$n.length.total  
  
  
range <- range %>%  
 group\_by(SD, year, based, range,   
 n.length.total, n.length.harvest, perc.length.range) %>%  
 summarise()  
  
  
  
#control if percentages are 1 per strata  
range %>%  
 group\_by(SD, year, based, n.length.total) %>%  
 summarise( sum.perc = sum(perc.length.range))

## # A tibble: 38 x 5  
## # Groups: SD, year, based [38]  
## SD year based n.length.total sum.perc  
## <fct> <dbl> <fct> <int> <dbl>  
## 1 22 2005 Sea 435 1  
## 2 22 2005 Land 1026 1  
## 3 22 2006 Sea 352 1  
## 4 22 2006 Land 10 1  
## 5 22 2007 Sea 18 1  
## 6 22 2007 Land 498 1  
## 7 22 2008 Sea 148 1  
## 8 22 2008 Land 340 1  
## 9 22 2009 Sea 572 1  
## 10 22 2010 Sea 1436 1  
## # ... with 28 more rows

Next, only the release range and sea-based fishing methods (no increased MLS for land-based fishing methods) are filtered and joined to the dat data. Where the join generated a NA due to no fitting value, the NA is replaced with a 0. Meaning in this stratum 0% of the fish will be released due to the increased MLS. For example, land-based fishing methods.

range <- filter(.data = range, range == "release" & based == "Sea")  
  
dat <- left\_join(x = dat, y = range)

## Joining, by = c("SD", "year", "based")

dat$perc.length.range[is.na(dat$perc.length.range)] <- 0

With the percentages of the release range, it can be calculated how many cod will be released due to the MLS (mls.released). Now two situations are possible: 1. the angler caught an amount of cod that is smaller than the bag limit (generally or due to the release of to small cod). 2. the angler caught so many cod, that even after the release of cod under the new MLS he exceeds the bag limit.

In the first case, the number of retained cod can stay as they are, because later with the introduced MLS the fish will be released (multiplied with the release mortality). In the second case, the number of retained cod stays as bag limit + the cod released due to the later introduced MLS. All other cod have to be released due to the bag limit.

dat <- mutate(.data = dat, mls.release = round(retained \* perc.length.range))   
  
dat <- mutate(.data = dat, retained.bag10 = if\_else(retained - mls.release > bag.limit, bag.limit + mls.release, retained))  
  
dat <- mutate(.data = dat, released.bag10 = retained - retained.bag10)

Then we continue as in the main script.

CPUE values are calculated for the different strata for retained and released cod from the on-site data (dat data frame).

In the dat data frame every angler is one row (Unit effort = on fishing day). The mean CPUE is calculated for all anglers in a stratum summing the number of retained and released cod and dividing by the total numbers of anglers respectively.

dat <- dat %>%  
 group\_by(federal\_state, SD, year, half\_year, survey\_category)%>%  
 mutate(n.angsamples = length(ID),  
 n.ret = sum(retained.bag10, na.rm = T),  
 n.rel = sum(released.bag10, na.rm = T),  
 mean.CPUE.rel = mean(released, na.rm = T),  
 mean.CPUE = mean(retained.bag10, na.rm = T),  
 mean.CPUE.bag10.rel = mean(released.bag10, na.rm = T))

Here an overview table for the CPUE´s per strata is created.

dat.strata.comb <- dat %>%  
 group\_by(federal\_state, SD, year, half\_year, survey\_category, n.angsamples, n.ret, n.rel,  
 mean.CPUE, mean.CPUE.rel, mean.CPUE.bag10.rel) %>%  
 summarise()

# Catch in numbers

The next step is to calculate catch in numbers. Therefore, CPUE values are multiplied with the effort (fishing days) per strata.

### Ratio between SD22 and SD24

In contrast to the CPUEs, effort is not available per subdivision but for federal states. In one federal state (Schleswig-Holstein, SH) subdivision and federal state overlap in another state (Mecklenburg Western Pomerania. MV) the coastline spans two subdivisions and therefore effort needs to be recalculated. Therefore, effort is split between the MV strata in SD 22 and 24. Here we calculate the ratio between encountered anglers during the on-site survey in SD22 to SD24. In SH the ratio is always 1 (only SD 22). In MV the ratio for the two subdivisions is used to split the effort.

dat <- dat %>%  
 group\_by(federal\_state, year, half\_year, survey\_category)%>%  
 mutate(n.ang.fs = length(ID))  
  
  
dat <- dat %>%  
 group\_by(federal\_state, SD, year, half\_year, survey\_category)%>%  
 mutate(n.ang.SD = length(ID))  
  
  
dat$SD.ratio <- dat$n.ang.SD/dat$n.ang.fs

### Calculate the perc of fishing days per month

~~~~~~~~~~~~~~~ different from the normal assessment ~~~~~~~~~~~~~~~~

With closed seasons not only the catches are reduced, but also the fishing effort.

Since the fishing effort values are per half year, the on-site survey is used as proxy how the fishing effort varies over the month. Therefore, the mean number of anglers sampled per month corrected for the amount of samplings and then the perc to the total number of sampled anglers per half year is calculated.

ang.month <- dat %>%  
 group\_by(federal\_state, survey\_category, year, half\_year, month) %>%  
 summarise(n.ang.month = length(activity\_id),  
 n.bepro.month = length(unique(session\_id)))  
  
ang.month <- mutate(.data = ang.month, ang.pro.bepro = n.ang.month/n.bepro.month)  
  
ang.month <- ang.month %>%  
 group\_by(federal\_state, survey\_category, year, half\_year) %>%  
 mutate(ang.pro.hyear = sum(ang.pro.bepro))  
  
ang.month <- mutate(.data = ang.month, perc.ang = (ang.pro.bepro)/ang.pro.hyear)

In the script two closed months are used, for every closed month a new column is created which indicates which strata are closed. Then the percentages of closed effort from the closed strata are summed.

ang.month <- mutate(.data = ang.month, clo = if\_else(month == 2 | month == 3, "closed", "open"))  
  
ang.month.clo <- ang.month%>%  
 group\_by(federal\_state, survey\_category, year, half\_year, clo)%>%  
 summarise(perc.clo = sum(perc.ang))  
  
ang.month.clo$clo[ang.month.clo$survey\_category == "Strand"] <- "open"  
ang.month.clo$clo[ang.month.clo$survey\_category == "Brandung"] <- "open"  
ang.month.clo$clo[ang.month.clo$survey\_category == "Watangeln"] <- "open"  
  
ang.month.clo <- filter(.data = ang.month.clo, clo == "closed")

Next, the data frames with the closed percentages is joined to the data frame with the fishing effort. By now the data frame includes month that are not closed, but where the perc.clo row is NA. Therefore, the NA´s are changed to 0. Now the half year fishing effort is multiplied with the percentages of fishing effort from the on-site survey that would be open (1- percentages of fishing effort in closed month).

fishing\_days <- left\_join(x = fishing\_days, y = ang.month.clo)

## Joining, by = c("federal\_state", "half\_year", "survey\_category", "year")

fishing\_days$perc.clo[is.na(fishing\_days$clo)] <- 0  
  
  
  
fishing\_days <- mutate(.data = fishing\_days, fishing\_days.clo = fishing\_days \* (1 - perc.clo),  
 fishing\_days.clo\_lower = fishing\_days\_lower \* (1 - perc.clo),  
 fishing\_days.clo\_upper = fishing\_days\_upper \* (1 - perc.clo))

### Combining data frames.

In order to combine data frames the CPUE data tables (dat data frame) are joined with the effort data (fishing\_days data frame). The data frames should be joined over “survey\_category”, “federal\_state”, “year” and “half\_year”.

dat <- left\_join(x = dat, y = fishing\_days)

## Joining, by = c("year", "survey\_category", "federal\_state", "half\_year")

### Catch and release numbers

With the CPUE´s, the effort and the effort ratio split in SD`s in on data frame, the numbers of retained (n.retained) and released (n.released) cod per strata are calculated. From the off-site telephone survey in 2015, we have upper and lower values of average fishing effort per angler, which are used to calculate upper and lower total fishing days. In the end the attributes are summed , whereby every row represents on strata and not an individual angler.

dat <- mutate(.data = dat, n.catch = mean.CPUE \* SD.ratio \* fishing\_days.clo,  
 n.catch\_lower = mean.CPUE \* SD.ratio \* fishing\_days.clo\_lower,  
 n.catch\_upper = mean.CPUE \* SD.ratio \* fishing\_days.clo\_upper)  
  
dat <- mutate(.data = dat, n.released.bag10 = mean.CPUE.bag10.rel \* SD.ratio \* fishing\_days.clo,  
 n.released.bag10\_lower = mean.CPUE.bag10.rel \* SD.ratio \* fishing\_days.clo\_lower,  
 n.released.bag10\_upper = mean.CPUE.bag10.rel \* SD.ratio \* fishing\_days.clo\_upper)  
  
dat <- mutate(.data = dat, n.released = mean.CPUE.rel \* SD.ratio \* fishing\_days.clo,  
 n.released\_lower = mean.CPUE.rel \* SD.ratio \* fishing\_days.clo\_lower,  
 n.released\_upper = mean.CPUE.rel \* SD.ratio \* fishing\_days.clo\_upper)  
  
  
  
CANUM <- dat %>%  
 group\_by(federal\_state, SD, year, half\_year, based, survey\_category, n.angsamples,  
 mean.CPUE, mean.CPUE.rel, SD.ratio, number\_angler,   
 fishing\_days.clo, fishing\_days.clo\_lower, fishing\_days.clo\_upper,  
 n.catch, n.catch\_lower, n.catch\_upper,   
 n.released.bag10, n.released.bag10\_lower, n.released.bag10\_upper,  
 n.released, n.released\_lower, n.released\_upper)%>%  
 summarise()

# CANUM

CANUM refers to all recreational removals (harvest + dead releases). Accordingly, the release component is adjusted with the respective release mortality. Dead releases are calculated separately for sea and land-based fishing methods. The sea-based value for post-release survival is 89,9% based on Weltersbach and Strehlow (2013). Due to lack of data, post-release survival for land-based fishing is assumed to 0%. The inverse mortality rate (death rate) is used for further calculations. After multiplying the released numbers with the mortality rate, harvested and dead released cod are summed to estimate CANUM.

~~~~~~~~~~~~~~~ different from the normal assessment ~~~~~~~~~~~~~~~~

In contrast to the normal assessment we have additional released cod due to the bag limit (e.g. n.released.bag10), which also have to be multiplied with the respective mortality rates.

After multiplying the released numbers with the death rate, we can sum up the retained and dead released cod to get the CANUM.

CANUM$death.rate <- 1  
CANUM$death.rate[CANUM$based == "Sea"] <- 0.112  
  
CANUM <- mutate(.data = CANUM, n.death.released = n.released \* death.rate,  
 n.death.released\_lower = n.released\_lower \* death.rate,  
 n.death.released\_upper = n.released\_upper \* death.rate,  
   
 n.death.released.bag10 = n.released.bag10 \* death.rate,  
 n.death.released.bag10\_lower = n.released.bag10\_lower \* death.rate,  
 n.death.released.bag10\_upper = n.released.bag10\_upper \* death.rate)  
  
  
CANUM <- mutate(.data = CANUM, CANUM = n.catch + n.death.released + n.death.released.bag10,  
 CANUM\_lower = n.catch\_lower + n.death.released\_lower + n.death.released.bag10\_lower,  
 CANUM\_upper = n.catch\_upper + n.death.released\_upper + n.death.released.bag10\_upper)

Here an overview table of the CANUM per year is created.

CANUM.year.comb <- CANUM %>%  
 group\_by(year) %>%  
 summarise(CANUM = sum(CANUM),   
 CANUM\_lower = sum(CANUM\_lower),   
 CANUM\_upper = sum(CANUM\_upper),  
 retained = sum(n.catch),   
 retained\_lower = sum(n.catch\_lower),   
 retained\_upper = sum(n.catch\_upper),  
 released = sum(n.death.released),   
 released\_lower = sum(n.death.released\_lower),   
 released\_upper = sum(n.death.released\_upper),  
 released.bag10 = sum(n.death.released.bag10),   
 released.bag10\_lower = sum(n.death.released.bag10\_lower),   
 released.bag10\_upper = sum(n.death.released.bag10\_upper))

Here an overview table of the CANUM per strata is created.

CANUM.strata.comb <- CANUM %>%  
 group\_by(federal\_state, SD, year, half\_year, survey\_category,) %>%  
 summarise(CANUM = sum(CANUM),   
 CANUM\_lower = sum(CANUM\_lower),   
 CANUM\_upper = sum(CANUM\_upper),  
 retained = sum(n.catch),   
 retained\_lower = sum(n.catch\_lower),   
 retained\_upper = sum(n.catch\_upper),  
 released = sum(n.death.released),   
 released\_lower = sum(n.death.released\_lower),   
 released\_upper = sum(n.death.released\_upper),  
 released.bag10 = sum(n.death.released.bag10),   
 released.bag10\_lower = sum(n.death.released.bag10\_lower),   
 released.bag10\_upper = sum(n.death.released.bag10\_upper))

# CANUM at Length

The next step is to divide the CANUM into length categories. Therefore, a length distribution from the on-site length sampling data (dat\_length) is created.

### Create length distribution

First, the sum of all measured fish per strata (n.length.total) is calculated. Since we want to create length distributions with a minimum amount of samples, land-based samples and sea-based strata under 500 samples (500 is a self-introduced limit and a not tested assumption) are removed. Then the sum of all measured fish per length class for each stratum (n.length.lc) is calculated. With these two numbers we can calculate the percentage of each length class per strata (perc.length.lc). We summarise the dat\_length data frame, so that every length class per strata is represented by on row and not each individual measured fish.

The table output is a control if the percentages per strata sum up to 1. When n.length.total = 0 the sum.perc should be INF.

dat\_length <- dat\_length %>%  
 group\_by(SD, year, based, retained) %>%  
 mutate(n.length.total = length(length\_class))  
  
  
dat\_length$n.length.total[dat\_length$based == "Land"] <- 0  
dat\_length$n.length.total[dat\_length$n.length.total < 500] <- 0  
  
  
dat\_length <- dat\_length %>%  
 group\_by(SD, year, based, retained, length\_class) %>%  
 mutate(n.length.lc = length(length\_class))  
  
  
dat\_length$perc.length.lc <- dat\_length$n.length.lc/dat\_length$n.length.total  
  
  
dat\_length <- dat\_length %>%  
 group\_by(SD, year, based, retained, length\_class,   
 n.length.total, n.length.lc, perc.length.lc) %>%  
 summarise()  
  
  
#control if percentages are 1 per strata  
dat\_length %>%  
 group\_by(SD, year, based, retained, n.length.total) %>%  
 summarise( sum.perc = sum(perc.length.lc))

## # A tibble: 69 x 6  
## # Groups: SD, year, based, retained [69]  
## SD year based retained n.length.total sum.perc  
## <fct> <dbl> <fct> <int> <dbl> <dbl>  
## 1 22 2005 Sea 1 0 Inf  
## 2 22 2005 Land 1 0 Inf  
## 3 22 2006 Sea 1 0 Inf  
## 4 22 2006 Land 1 0 Inf  
## 5 22 2007 Sea 0 0 Inf  
## 6 22 2007 Sea 1 0 Inf  
## 7 22 2007 Land 1 0 Inf  
## 8 22 2008 Sea 0 0 Inf  
## 9 22 2008 Sea 1 0 Inf  
## 10 22 2008 Land 0 0 Inf  
## # ... with 59 more rows

### Fill the length distribution gaps

For some sea-based strata and for land-based fishing methods we do not have enough samples to build a trustworthy length-distribution. To cover these strata summed length distributions are used combining data over several years.

##### Creating length distributions over several years

First, a new column is created with the time spans and the data frame divided into sea- and land-based length distributions, since the calculations differ.

~~~~~~ Check from year to year the labels of the time spans ~~~~~~~~~~

dat\_length$year\_span <- NA  
dat\_length$year\_span[dat\_length$year > 2013] <- "2014-19"  
dat\_length$year\_span[dat\_length$year < 2014] <- "2005-13"  
  
  
dat\_length\_spansea <- filter(.data = dat\_length, based == "Sea")  
dat\_length\_spanland <- filter(.data = dat\_length, based == "Land")

For sea-based fishing methods:

First the sum of all measured fish per year span is calculated (n.length.total). Then the sum of all measured fish per length class in each year span is calculated (n.length.lc). Out of these two numbers the percentage of each length class to the total per year span (perc.length.lc) is calculated. The data frame is summed, so that every length class per year span is represented by on row and not each year.

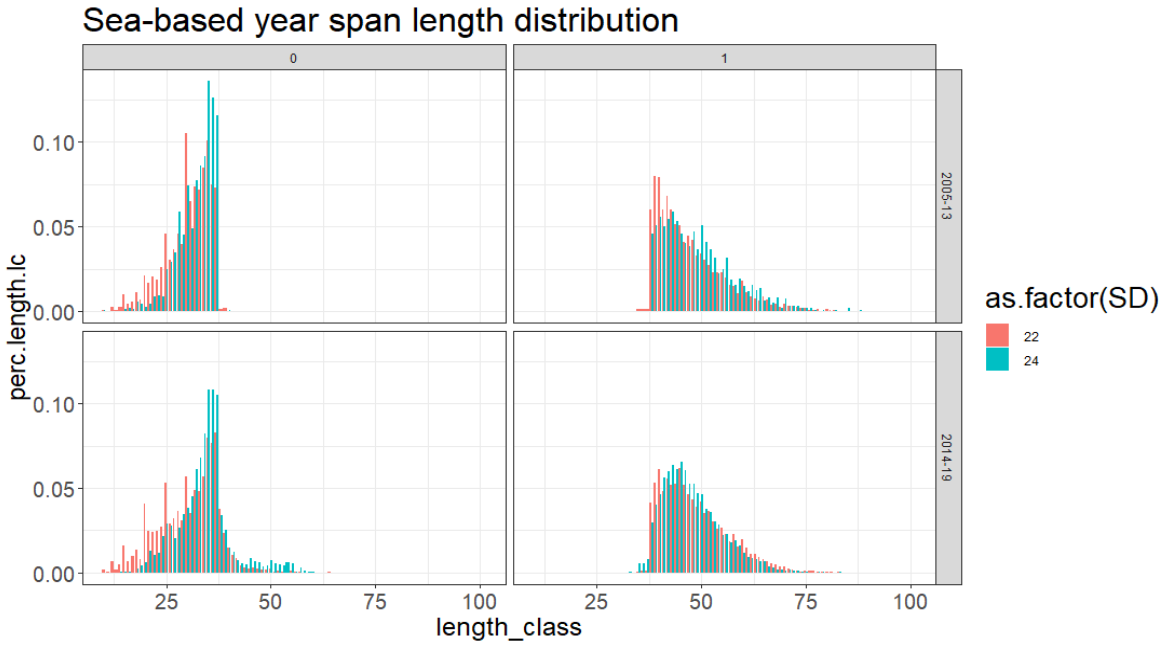
The table output is a control, if the percentages per strata sum up to 1.

dat\_length\_spansea <- dat\_length\_spansea %>%  
 group\_by(based, SD, retained, year\_span) %>%  
 mutate(n.length.total = sum(n.length.lc))  
  
  
dat\_length\_spansea <- dat\_length\_spansea %>%  
 group\_by(based,SD, year\_span, retained, length\_class) %>%  
 mutate(n.length.lc = sum(n.length.lc))  
  
dat\_length\_spansea$perc.length.lc <- dat\_length\_spansea$n.length.lc/dat\_length\_spansea$n.length.total  
  
  
dat\_length\_spansea <- dat\_length\_spansea %>%  
 group\_by(based,SD, year\_span, retained, length\_class,   
 n.length.total, n.length.lc, perc.length.lc) %>%  
 summarise()  
  
  
  
#control if percentages are 1 per strata  
dat\_length\_spansea %>%  
 group\_by(based,SD, year\_span, retained, n.length.total) %>%  
 summarise( sum.perc = sum(perc.length.lc))

## # A tibble: 8 x 6  
## # Groups: based, SD, year\_span, retained [8]  
## based SD year\_span retained n.length.total sum.perc  
## <fct> <fct> <chr> <int> <int> <dbl>  
## 1 Sea 22 2005-13 0 3900 1  
## 2 Sea 22 2005-13 1 7995 1  
## 3 Sea 22 2014-19 0 4590 1  
## 4 Sea 22 2014-19 1 8065 1  
## 5 Sea 24 2005-13 0 1571 1  
## 6 Sea 24 2005-13 1 4895 1  
## 7 Sea 24 2014-19 0 2624 1  
## 8 Sea 24 2014-19 1 6296 1

The plot shows the sea-based length distributions for each year span and SD. The plot is separated for retained and released cod.

ggplot(data = dat\_length\_spansea, aes(x = length\_class, y = perc.length.lc, fill = as.factor(SD)))+  
 geom\_col(position = "dodge")+  
 facet\_grid(year\_span ~ retained)+  
 labs(title = "Sea-based year span length distribution")+  
 theme\_bw()+  
 theme(axis.text = element\_text(size = 15),  
 axis.title = element\_text(size = 18),  
 title = element\_text(size = 20))



For land-based fishing methods:

First the sum of all measured fish is calculated (n.length.total). Then, the sum of all measured fish per length class is calculated (n.length.lc). From these two numbers the percentage of each length class to the total (perc.length.lc) is calculated. The data frame is summed, so that every length class is represented by on row.

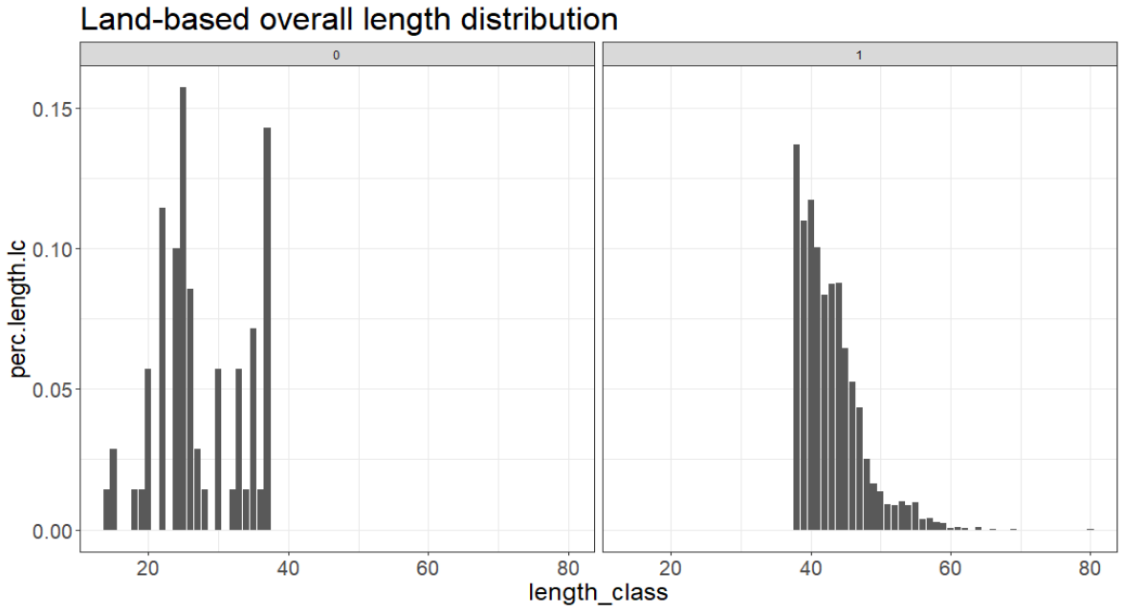
The table output is a control if the percentages per strata sum up to 1.

dat\_length\_spanland <- dat\_length\_spanland %>%  
 group\_by(based, retained)%>%  
 mutate(n.length.total = sum(n.length.lc))  
  
  
dat\_length\_spanland <- dat\_length\_spanland %>%  
 group\_by(based, retained, length\_class) %>%  
 mutate(n.length.lc = sum(n.length.lc))  
  
dat\_length\_spanland$perc.length.lc <- dat\_length\_spanland$n.length.lc/dat\_length\_spanland$n.length.total  
  
  
dat\_length\_spanland <- dat\_length\_spanland %>%  
 group\_by(based, retained, length\_class, n.length.total, n.length.lc, perc.length.lc) %>%  
 summarise()  
  
  
  
#control if percentages are 1 per strata  
dat\_length\_spanland %>%  
 group\_by(based, retained, n.length.total) %>%  
 summarise( sum.perc = sum(perc.length.lc))

## # A tibble: 2 x 4  
## # Groups: based, retained [2]  
## based retained n.length.total sum.perc  
## <fct> <int> <int> <dbl>  
## 1 Land 0 70 1  
## 2 Land 1 3530 1

The plot shows the land-based length distributions separated for retained and released cod.

ggplot(data = dat\_length\_spanland, aes(x = length\_class, y = perc.length.lc))+  
 geom\_col(position = "dodge")+  
 facet\_grid(.~ retained)+  
 labs(title = "Land-based overall length distribution")+  
 theme\_bw()+  
 theme(axis.text = element\_text(size = 15),  
 axis.title = element\_text(size = 18),  
 title = element\_text(size = 20))



##### Fill the strata without samples with the distribution from the year span

To fill strata with insufficient numbers of samples a with the combined length distribution over several years a data frame which includes only strata that had not enough samples or no samples at all is created.

The all\_strata data frame (Reminder that this data frame needs to be updated with Excel once a year!) is joined with the factor’s sea- & land-based and retained/released.

all\_strata\_sd <- all\_strata %>%  
 group\_by(SD, year, based)%>%  
 summarise()  
  
re <- data.frame(based = c("Sea", "Sea", "Land", "Land"),  
 retained = c(1,0,1,0))  
  
all\_strata\_sd <- left\_join(x = re, y = all\_strata\_sd)

## Joining, by = "based"

In a next step, the resulting all strata data frame is joined with dat\_length. Due to the joining procedure strata that are not in the dat\_length data frame are filled with NA´s. NA´s are changed to 0. Now all the strata that are not sampled at all and the strata with insufficient numbers of samples have a ‘0’ at the column n.length.total. Since we don´t want the strata, that have enough samples for a single length distribution, to be overwritten we only filter for the strata with n.length.total = 0.

dat\_length\_gaps <- right\_join(x = dat\_length, y = all\_strata\_sd)

## Joining, by = c("SD", "year", "based", "retained")

dat\_length\_gaps$n.length.total[is.na(dat\_length\_gaps$n.length.total)] <- 0  
  
  
dat\_length\_gaps <- filter(.data = dat\_length\_gaps, n.length.total == 0)

To fill the gaps with the combined length distributions over several years, only the columns that describe the strata (SD, year, land- & sea-based and retained) are kept and the year span they should be filled with.

Since we have two different data frames for the sea- & land-based length distributions over several years, we also split the data frame into sea- & land-based gaps. Then the sea gap strata are joined over “based”, “SD” , “retained” & “year span” with the year span length distribution and the land gap strata over “based” & “retained” with the combined length distribution over several years. Last, sea- and land-based data frames are brought together again.

dat\_length\_gaps <- dat\_length\_gaps%>%  
 group\_by(SD, year, based, retained, year\_span)%>%  
 summarise()  
  
dat\_length\_gaps$year\_span <- NA  
dat\_length\_gaps$year\_span[dat\_length\_gaps$year > 2013] <- "2014-19"  
dat\_length\_gaps$year\_span[dat\_length\_gaps$year < 2014] <- "2005-13"  
  
  
dat\_length\_gapssea <- filter(.data = dat\_length\_gaps, based == "Sea")  
dat\_length\_gapsland <- filter(.data = dat\_length\_gaps, based == "Land")  
  
  
dat\_length\_gapsfilledsea <- left\_join(x = dat\_length\_gapssea, y = dat\_length\_spansea)

## Joining, by = c("SD", "based", "retained", "year\_span")

dat\_length\_gapsfilledland <- left\_join(x = dat\_length\_gapsland, y = dat\_length\_spanland)

## Joining, by = c("based", "retained")

dat\_length\_gapsfilled <- rbind(dat\_length\_gapsfilledsea, dat\_length\_gapsfilledland)

##### Combine length distributions from sufficient strata and the filled length distributions

We now have a data frame with all length distributions (dat\_length, which include strata with insufficient numbers of samples) and a data frame with the filled length distributions (dat\_length\_gapsfilled). To combine the two sets the strata with insufficient numbers of samples need to be removed from dat\_length. A new column is created to see, if the distribution is filled or from one stratum. Last, the two data frames are combined to on final data frame containing all length distributions.

dat\_length <- filter(.data = dat\_length, n.length.total != 0)  
  
dat\_length$lc.origin <- "one\_strata"  
dat\_length\_gapsfilled$lc.origin <- "filled"  
  
dat\_length <- rbind(dat\_length, dat\_length\_gapsfilled)

### Combine CANUM and length distributions

To distribute CANUM into all length distributions the two data frames are joined together over “SD”, “year” and “based”.

CANUM\_length <- right\_join(x = dat\_length, y = CANUM)

## Joining, by = c("SD", "year", "based")

### Calculate the CANUM per Length class

Since the length distributions are separated for retained and released cod CANUM needs to be separated too and retained cod has to be divided with the retained length distribution and the dead released cod with the released length distribution. As a result, number of removed cod are shown for each length class (lc) separated for harvest and dead releases.

CANUM\_length <- mutate(.data = CANUM\_length, Canum.lc = if\_else(retained == 1,   
 n.catch \* perc.length.lc,   
 n.death.released \* perc.length.lc),  
   
 Canum.lc\_lower = if\_else(retained == 1,   
 n.catch\_lower \* perc.length.lc,   
 n.death.released\_lower \* perc.length.lc),  
   
 Canum.lc\_upper = if\_else(retained == 1,   
 n.catch\_upper \* perc.length.lc,   
 n.death.released\_upper \* perc.length.lc))  
  
  
  
CANUM\_length <- mutate(.data = CANUM\_length, n.death.released.lc.bag10 = if\_else(retained == 1,   
 n.death.released.bag10 \* perc.length.lc,   
 0),  
   
 n.death.released.lc.bag10\_lower = if\_else(retained == 1,   
 n.death.released.bag10\_lower   
 \* perc.length.lc,   
 0),  
   
 n.death.released.lc.bag10\_upper = if\_else(retained == 1,   
 n.death.released.bag10\_upper  
 \* perc.length.lc,   
 0))

# Introducing a new MLS

~~~~~~~~~~~~~~~ different from the normal assessment ~~~~~~~~~~~~~~~~

Normally released cod are already multiplied with the death rate (see above: n.death.released). But all retained cod catches that are smaller than the new proposed MLS would have been released and therefore have to be multiplied with death rate (land = 1, sea = 0.112) of released cod. Following the values for the proposed MLS are set and all length classes smaller than the MLS that would have been retained are multiplied with the death rate.

CANUM\_length <- mutate(.data = CANUM\_length, Canum.lc.45 = if\_else(based == "Sea" & length\_class < MLS & retained == 1,  
 Canum.lc \* death.rate,  
 Canum.lc),  
   
 Canum.lc.45\_lower = if\_else(based == "Sea" &length\_class < MLS & retained == 1,  
 Canum.lc\_lower \* death.rate,  
 Canum.lc\_lower),  
   
 Canum.lc.45\_upper = if\_else(based == "Sea" &length\_class < MLS & retained == 1,  
 Canum.lc\_upper \* death.rate,  
 Canum.lc\_upper))

Now we can add the dead released by the bag limit to the CANUM from the MLS.

CANUM\_length <- mutate(.data = CANUM\_length, CANUM.lc.45 = Canum.lc.45 + n.death.released.lc.bag10,  
   
 CANUM.lc.45\_lower = Canum.lc.45\_lower +   
 n.death.released.lc.bag10\_lower,  
   
 CANUM.lc.45\_upper = Canum.lc.45\_upper +   
 n.death.released.lc.bag10\_upper)

# CATON

To estimate catch in weight (tons) the mean weight of cod per length class (lc) is calculated using the length-weight relationship from commercial samples.

First, the CANUM per length data frame is joined with the mass coefficient data frame, so every “SD”, “year”, “half\_year” combination will get the parameters of the length-weight relationship.

CATON\_length <- left\_join(x = CANUM\_length, y = mass\_coefficient)

## Joining, by = c("SD", "year", "half\_year")

Second, mean weight [g] is calculated for each length class with the formal: weight = a \* length ^b = exp(ln(a) + b \* ln(length[mm]))

Third, mean weight per length class is multiplied with CANUM of each length class to get the total weight of the length class.

CATON\_length <- mutate(.data = CATON\_length, mean.weight.lc = a \* (length\_class ^ b))  
  
CATON\_length <- mutate(.data = CATON\_length, CATON.g.lc.45 = (CANUM.lc.45 \* mean.weight.lc),  
 CATON.g.lc.45\_lower = (CANUM.lc.45\_lower \* mean.weight.lc),  
 CATON.g.lc.45\_upper = (CANUM.lc.45\_upper \* mean.weight.lc),  
   
 CATON.kg.lc.45 = (CANUM.lc.45 \* mean.weight.lc)/1000,  
 CATON.kg.lc.45\_lower = (CANUM.lc.45\_lower \* mean.weight.lc)/1000,  
 CATON.kg.lc.45\_upper = (CANUM.lc.45\_upper \* mean.weight.lc)/1000,  
   
 CATON.t.lc.45 = (CANUM.lc.45 \* mean.weight.lc)/1000000,  
 CATON.t.lc.45\_lower = (CANUM.lc.45\_lower \* mean.weight.lc)/1000000,  
 CATON.t.lc.45\_upper = (CANUM.lc.45\_upper \* mean.weight.lc)/1000000)

Here an overview table of the CATON per strata is created.

CATON.t.strata.comb <- CATON\_length %>%  
 group\_by(federal\_state, SD, year, half\_year, survey\_category) %>%  
 summarise(comb = sum(CATON.t.lc.45),  
 comb\_lower = sum(CATON.t.lc.45\_lower),  
 comb\_upper = sum(CATON.t.lc.45\_upper))

Here an overview table of the CATON per year is created.

CATON.t.year.comb <- CATON\_length %>%  
 group\_by(year) %>%  
 summarise(comb = sum(CATON.t.lc.45),  
 comb\_lower = sum(CATON.t.lc.45\_lower),  
 comb\_upper = sum(CATON.t.lc.45\_upper))