Module 4 Assignment 3

EKB

2025-12-02

Assignment Details

Purpose

The goal of this assignment is to assess your ability to apply concepts of data visualization, measures of central tendency and variability, fitting logistic growth models, and using inferential statistics to compare populations.

Task

Write R code which produces the correct answers and correctly interpret the results of visualizations and models.

Criteria for Success

- Code is within the provided code chunks
- Code chunks run without errors
- Code produces the correct result
    - Code that produces the correct answer will receive full credit
    - Code attempts with logical direction will receive partial credit
- Written answers address the questions in sufficient detail

Due Date
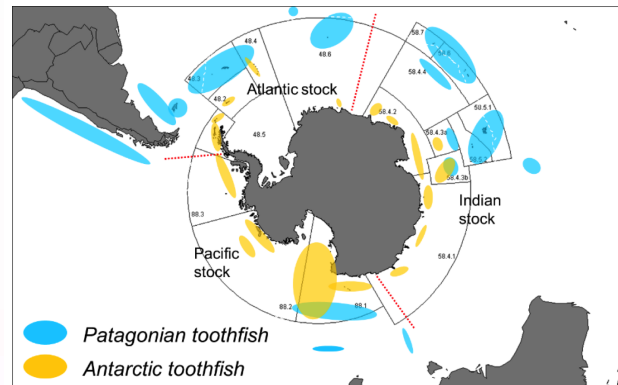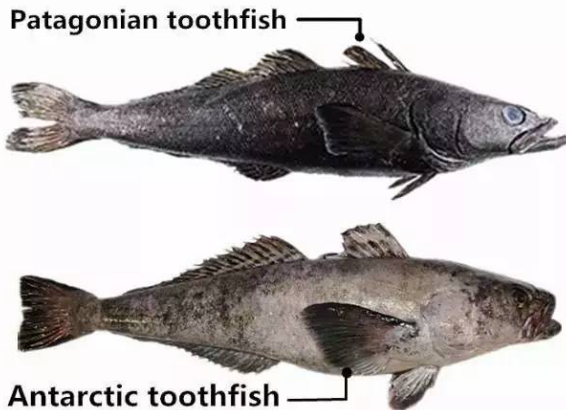
December 9 at 10:45 am MST

Assignment Questions

Some of our colleagues from the "mainland" have heard about our work estimating carrying capacities and fitting logistic growth models to populations and have asked for our help.

In their research, they have been using the mass (weight) of the fish they have caught to estimate populations of fish over the last 50 years. In fisheries, using mass as a proxy for–or representation of–abundance is common practice.

Our colleagues have collected data in an interesting location, where the ranges of both species of toothfish (genus: *Dissostichus*) overlap. Therefore, they have population estimates for both the Patagonian toothfish (AKA Chilean sea bass, *D. eleginoides*) and the Antarctic toothfish (*D. mawsoni*) populations in the area.

Patagonian toothfish

Antarctic toothfish

Patagonian toothfish
Antarctic toothfish

They want to know (a) what the estimated carrying capacity of each population is and (b) if the population sizes are statistically different from each other.

Set-Up

1. Load both of the packages (`tidyverse`, `drc`) that we will need to use to answer the questions.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr    1.5.2
## v ggplot2    4.0.0      v tibble     3.3.0
## v lubridate 1.9.4       v tidyr      1.3.1
## v purrr      1.1.0
## -- Conflicts ------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(drc)
```

```
## Loading required package: MASS
##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##     select
##
##
## 'drc' has been loaded.
##
## Please cite R and 'drc' if used for a publication,
## for references type 'citation()' and 'citation('drc')'.
##
##
## Attaching package: 'drc'
##
## The following objects are masked from 'package:stats':
##
##     gaussian, getInitial
```

2. Read in the dataset sent to us by our colleagues with their data.

```
toothfish <- read_csv("../data/toothfish.csv")
```

```
## Rows: 100 Columns: 3
## -- Column specification -------------------------------------------------------
## Delimiter: ","
## chr (1): species
## dbl (2): fish_kg, date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

3. Examine the data frame, either through code or visually. To complete our analyses, we will need data for (a) time, (b) abundance, and (c) population. Below, identify which columns in the data frames represent which data.
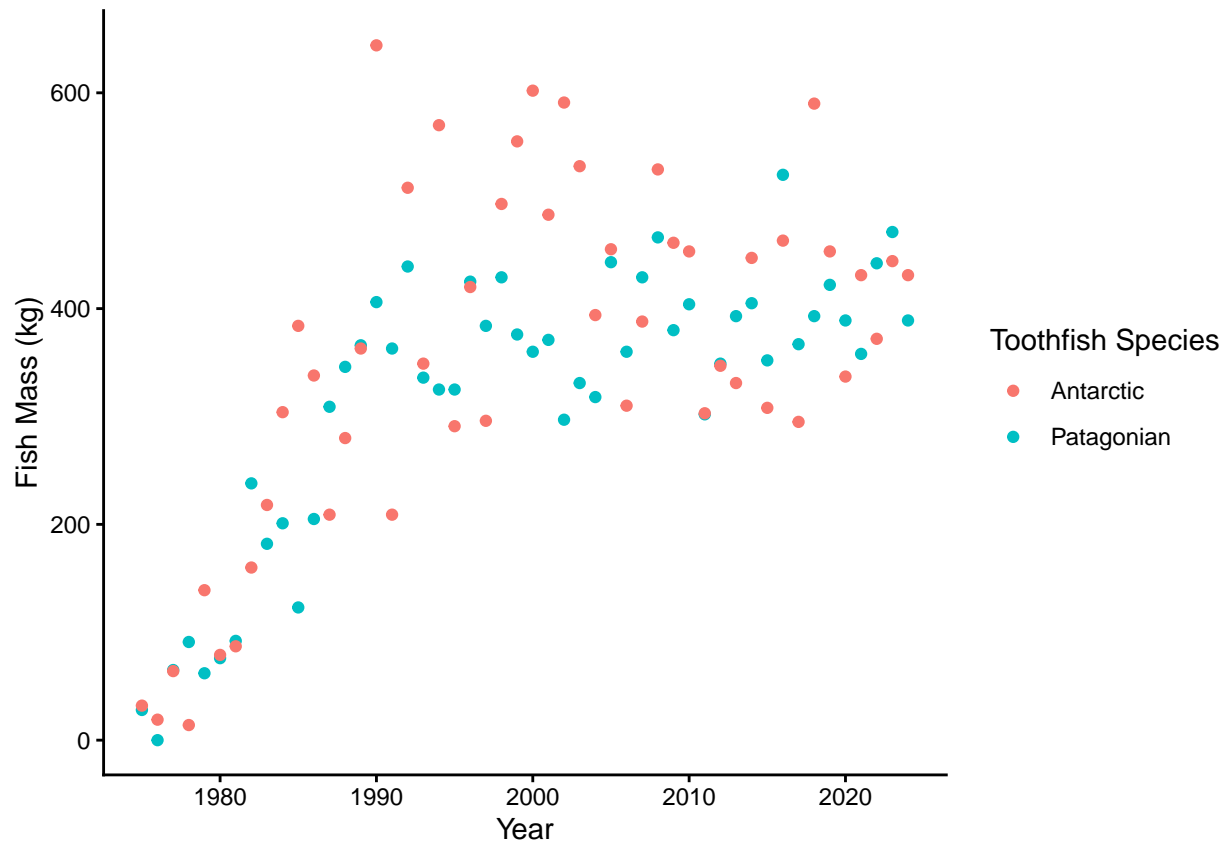
- Time: date
- Abundance: fish_kg
- Population: species

Visual Comparison

Let's start with a visual comparison of the populations.

4. Make a scatter plot of the two populations. Ensure that the dependent variable is on the y-axis and the independent variable is on the x-axis. Use color to differentiate between the fish populations.

   Be sure to modify your axes labels to be clearer and more descriptive; also, choose a theme for your plot.

```
ggplot(toothfish, aes(x = date, y = fish_kg, color = species)) +
  geom_point() +
  labs(x = "Year", y = "Fish Mass (kg)", color = "Toothfish Species") +
  theme_classic()
```

5. Based on your plot above, make some initial observations. Does each population appear to be showing growth more akin to exponential or logistic growth? Do the populations show a lot or a little variation? Does one population show more variation? Which population do you expect to have a higher carrying capacity, if it has one? (2 points)

Answer: logistic, lot of variation, Antarctic shows more, hard to tell but carrying capacity might be higher in Antarctic

Fitting Models

6. In order to estimate the carrying capacity of each population, we will first separate the data into two dataframes, one for each species of fish.

   Below, create the two new data frames. Be sure to save them as new objects.

```
pat <- toothfish %>%
  filter(species == "Patagonian")
pat
```

```
## # A tibble: 50 x 3
##     species     fish_kg  date
##     <chr>         <dbl> <dbl>
##  1 Patagonian       28  1975
##  2 Patagonian        0  1976
##  3 Patagonian       65  1977
##  4 Patagonian       91  1978
##  5 Patagonian       62  1979
##  6 Patagonian       76  1980
```

```
##  7 Patagonian       92  1981
##  8 Patagonian      238  1982
##  9 Patagonian      182  1983
## 10 Patagonian      201  1984
## # i 40 more rows
```

```
ant <- toothfish %>%
  filter(species == "Antarctic")
ant
```

```
## # A tibble: 50 x 3
##    species    fish_kg  date
##    <chr>        <dbl> <dbl>
##  1 Antarctic       32  1975
##  2 Antarctic       19  1976
##  3 Antarctic       64  1977
##  4 Antarctic       14  1978
##  5 Antarctic      139  1979
##  6 Antarctic       79  1980
##  7 Antarctic       87  1981
##  8 Antarctic      160  1982
##  9 Antarctic      218  1983
## 10 Antarctic      304  1984
## # i 40 more rows
```

Now that we have separate data frames, we can fit the logistic models to each species' population.

7. Let's start with the Patagonian toothfish. Use the `drm` function to fit a logistic model to the population. Remember to use the `fct = LL.4()` argument.

```
pat_model <- drm(fish_kg ~ date, data = pat, fct = LL.4())
pat_model
```

```
##
## A 'drc' model.
##
## Call:
## drm(formula = fish_kg ~ date, data = pat, fct = LL.4())
##
## Coefficients:
## b:(Intercept)  c:(Intercept)  d:(Intercept)  e:(Intercept)
##     -628.9632        -0.6173       391.1327      1983.6795
```

8. Do the same thing for the Antarctic toothfish.

```
ant_model <- drm(fish_kg ~ date, data = ant, fct = LL.4())
ant_model
```

```
##
## A 'drc' model.
##
## Call:
## drm(formula = fish_kg ~ date, data = ant, fct = LL.4())
##
## Coefficients:
## b:(Intercept)  c:(Intercept)  d:(Intercept)  e:(Intercept)
##      -719.317        -1.875       430.739      1983.039
```

9. Report the respective carrying capacities in the table below.

Toothfish Species	Estimate of K
Patagonian	391.1
Antarctic	430.7

Plotting Logistic Models

10. To add the logistic growth curves to our plot, we need to add the "fitted" or "predicted" values from the model to each data frame. We can do this using the `predict()` function within a `mutate()` function.

    In each data frame, create a new column called `predicted_values` that contains the output from the `predict()` function. Be sure to save the resulting data frame to ensure that the new column is included.

```
pat <- pat %>% mutate(predicted_values = predict(pat_model))
pat
```

```
## # A tibble: 50 x 4
##    species    fish_kg  date predicted_values
##    <chr>        <dbl> <dbl>            <dbl>
##  1 Patagonian      28  1975             22.7
##  2 Patagonian       0  1976             30.8
##  3 Patagonian      65  1977             41.3
##  4 Patagonian      91  1978             54.8
##  5 Patagonian      62  1979             71.7
##  6 Patagonian      76  1980             92.3
##  7 Patagonian      92  1981            117.
##  8 Patagonian     238  1982            144.
##  9 Patagonian     182  1983            174.
## 10 Patagonian     201  1984            205.
## # i 40 more rows
```

```
ant <- ant %>% mutate(predicted_values = predict(ant_model))
ant
```

```
## # A tibble: 50 x 4
##    species   fish_kg  date predicted_values
##    <chr>       <dbl> <dbl>            <dbl>
##  1 Antarctic      32  1975             20.2
##  2 Antarctic      19  1976             29.2
##  3 Antarctic      64  1977             41.5
##  4 Antarctic      14  1978             57.9
##  5 Antarctic     139  1979             79.2
##  6 Antarctic      79  1980            106.
##  7 Antarctic      87  1981            138.
##  8 Antarctic     160  1982            174.
##  9 Antarctic     218  1983            213.
## 10 Antarctic     304  1984            252.
## # i 40 more rows
```

11. Combine the data frames for the Patagonian and Antarctic toothfish back together into one data frame using the `bind_rows()` function.

```
toothfish <- bind_rows(pat, ant)
toothfish
```

```
## # A tibble: 100 x 4
```
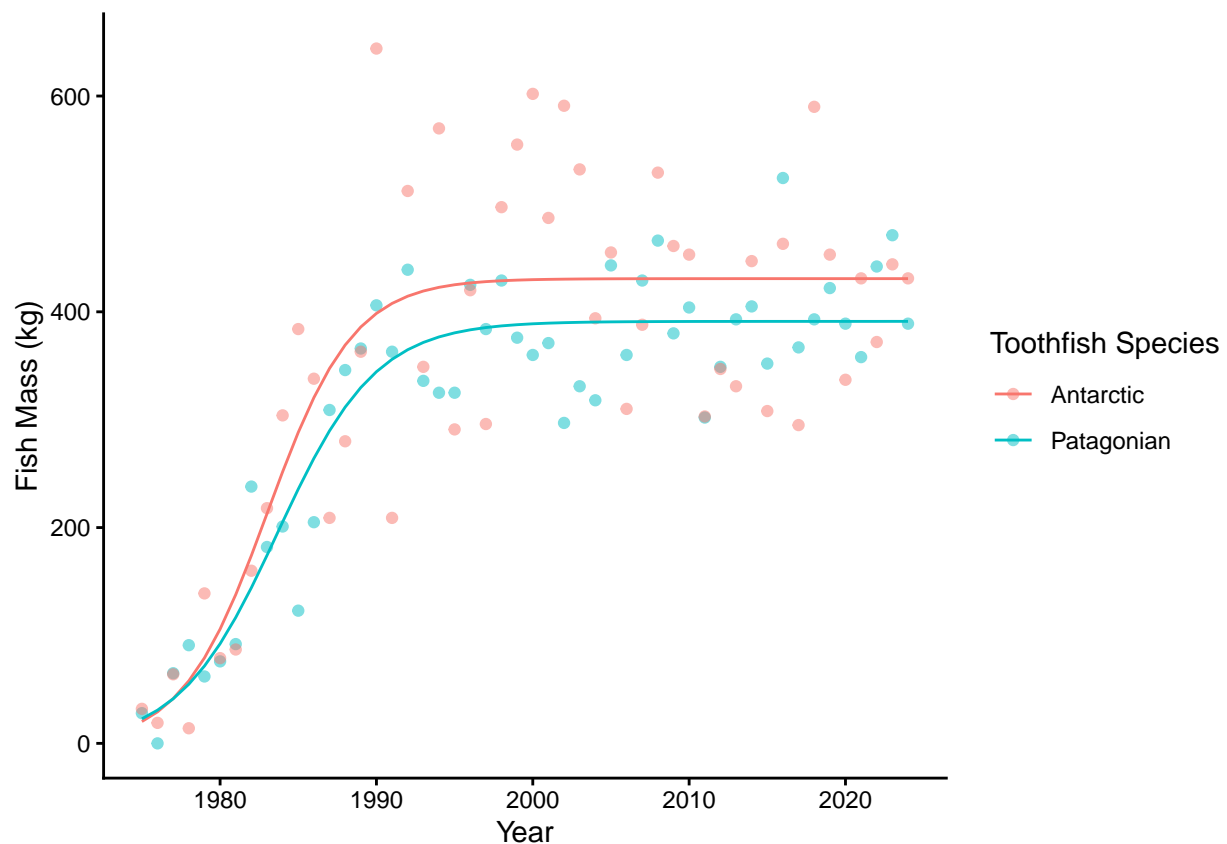
```
##    species   fish_kg  date predicted_values
##    <chr>       <dbl> <dbl>            <dbl>
##  1 Patagonian     28  1975             22.7
##  2 Patagonian      0  1976             30.8
##  3 Patagonian     65  1977             41.3
##  4 Patagonian     91  1978             54.8
##  5 Patagonian     62  1979             71.7
##  6 Patagonian     76  1980             92.3
##  7 Patagonian     92  1981            117.
##  8 Patagonian    238  1982            144.
##  9 Patagonian    182  1983            174.
## 10 Patagonian    201  1984            205.
## # i 90 more rows
```

12. Add the logistic growth curves for each toothfish species to the scatterplot you made earlier. You will use `geom_line()` to accomplish this task.

Optional: make the data points partially transparent so we can see the logistic growth curves more clearly.

```
ggplot(toothfish, aes(date, fish_kg, color = species)) +
  geom_point(alpha = 0.5) +
  geom_line(aes(y = predicted_values)) +
  labs(x = "Year", y = "Fish Mass (kg)", color = "Toothfish Species") +
  theme_classic()
```

Comparing Populations

13. Based on where the logistic models "level-off," let's numerically and statistically analyze the two populations starting from the year 1995.

    Create and save a new data frame as an object that contains only rows from the 1995 or later.

```
toothfish <- toothfish %>% filter(date >= 1995)
toothfish
```

```
## # A tibble: 60 x 4
##    species    fish_kg  date predicted_values
##    <chr>        <dbl> <dbl>            <dbl>
##  1 Patagonian     325  1995             381.
##  2 Patagonian     425  1996             383.
##  3 Patagonian     384  1997             385.
##  4 Patagonian     429  1998             387.
##  5 Patagonian     376  1999             388.
##  6 Patagonian     360  2000             389.
##  7 Patagonian     371  2001             389.
##  8 Patagonian     297  2002             390.
##  9 Patagonian     331  2003             390.
## 10 Patagonian     318  2004             390.
## # i 50 more rows
```

14. Before we run any inferential statistics, let's calculate some summary statistics.

    For each species, calculate the average fish mass and the standard deviation of the fish mass.

```
toothfish %>%
  group_by(species) %>%
  summarise(avg_fish_kg = mean(fish_kg),
            stdev_fish_kg = sd(fish_kg))
```

```
## # A tibble: 2 x 3
##    species    avg_fish_kg stdev_fish_kg
##    <chr>            <dbl>         <dbl>
## 1 Antarctic         427.          94.9
## 2 Patagonian        388.          51.8
```

15. Compare the average values to the carrying capacity estimates; they will not be exactly the same, but are they close to what we might expect?

    How about our measure of variability, the standard deviation? Do those values match with your interpretation from Question 5? Why or why not? (2 points)

Answer: yes, both carrying capacities and st deviations match up; higher variability in Antarctic species

16. Our colleagues have asked us to use an inferential statistical test to determine whether or not the Patagonian toothfish and Antarctic toothfish populations in their study area are of significantly different sizes.

    Run the appropriate statistical test to answer their question.

```
t.test(fish_kg ~ species, data = toothfish)
```

```
## 
##  Welch Two Sample t-test
## 
## data:  fish_kg by species
## t = 1.9562, df = 44.877, p-value = 0.05669
```

```
## alternative hypothesis: true difference in means between group Antarctic and group Patagonian is not
## 95 percent confidence interval:
##  -1.147373 78.414040
## sample estimates:
##  mean in group Antarctic mean in group Patagonian
##                  427.1000                 388.4667
```

17. How do we interpret the results of this statistical test? Are the populations different sizes? How do you know? (2 points)

Answer: while close to 0.05, the p-value is above 0.05, meaning we cannot say with certainty that the populations are of statistically significantly different sizes