Painted Turtle Survival Rate Analysis

Haoluan Chen

2/13/2021

Statistical Analysis

1. Introduction

In this statistical analysis, we aimed to understand the natural survival rate for painted turtles in the Algonquin Park. Our data was collected using the capture-recapture method, which would cause a high correlation in observations from the same turtle. To take advantage of this type of data, we decided to use the Cormack-Jolly-Seber (CJS) model to fit our data. The CJS model can estimate the survival rate of individuals in a population and identify factors that affected the survival rate using the capture-recapture dataset.

2. Model Selection

Through our collaborator meetings and EDA, I obtained the following information, which helped in my model selection process:

1. Larger turtle have a higher survival rate
2. Female turtle can grow larger than turtles
3. Capture rate for male turtle is lower compared to female turtle
4. Bigger turtle has a metal tag that we can see underwater. This means a higher capture rate.
5. midpl are the most reliable size measurement
6. all the size variables are highly correlated, but the mass has the lowest correlation with other size measurements.

Therefore, I started experimenting and building models with covariate sex and various size measurement.

I had a hard time choosing from the model with sex as static variable for both survival probability and capture probability and the model with sex as static variable and midpl as time-varying variable for both survival probability and capture probability. The main problem is that for the second model, some imputation method produced a non-significant negative estimate for midpl. And some imputation method produced a significant negative estimate for midpl, which indicated that larger turtle has a lower survival probability, which is not correct based on what we already know about the painted turtles. (Full detail in the Appendixes)

After comparing the models, I decided that the model with sex as a static variable is the most appropriate model.

2.1 Model Assuptions Check

1. Capture probability is the same for all individuals on occasion t.
2. Survival probability is the same for all individuals on occasion t.
3. Marks are neither lost nor overlooked and are recorded correctly throughout the study.
4. Sampling periods are instantaneous and recaptured individuals are released immediately.
5. Emigration from the study area is permanent.
6. Individuals are independent from each other.
7. Captures of the same individual on different occasions are independent.

For 1 and 2, I think we can safely assume that same-sex turtles are equally likely to be capture and survival probability because male turtles, female and unknown(juvenile) turtles have very different behaviors. Male move around more, so they are harder to capture and lower survival probability. Females tend to stay close to their nest and are generally larger, so they are relatively easier to capture and higher survival probability. Juvenile is smaller, so they are harder to capture and much lower survival probability.

Since we have a reliable notch system, assumption 3 is satisfied.

Our collaborator said that 90% of the turtles are released within 24 hours, remaining 10% are released right away. So, we may say within 24 hours is acceptable with our assumption.

We may have some problem with the emigration is permanent assumption, because we know that the turtles in our data set immigrate and emigrate to other cite and some of them came back after few years. This assumption may not be satisfied.

Our collaborator said that we could safely assume each turtle and each capture is independent of each other.

2.2 Sensitivity to imputation check

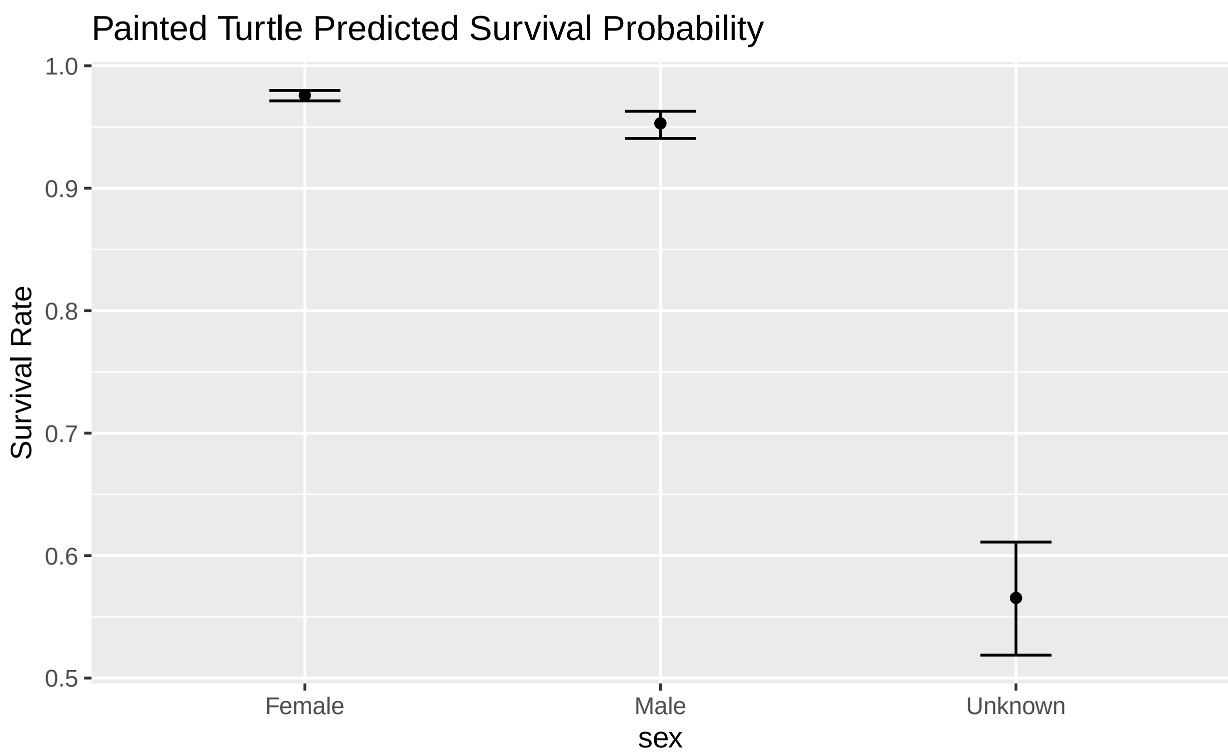
My final model does not use any imputation methods. However, I did sensitivity to imputation check for my second model, which helped my decision on the final model. (Details in the appendixes)

3. Result

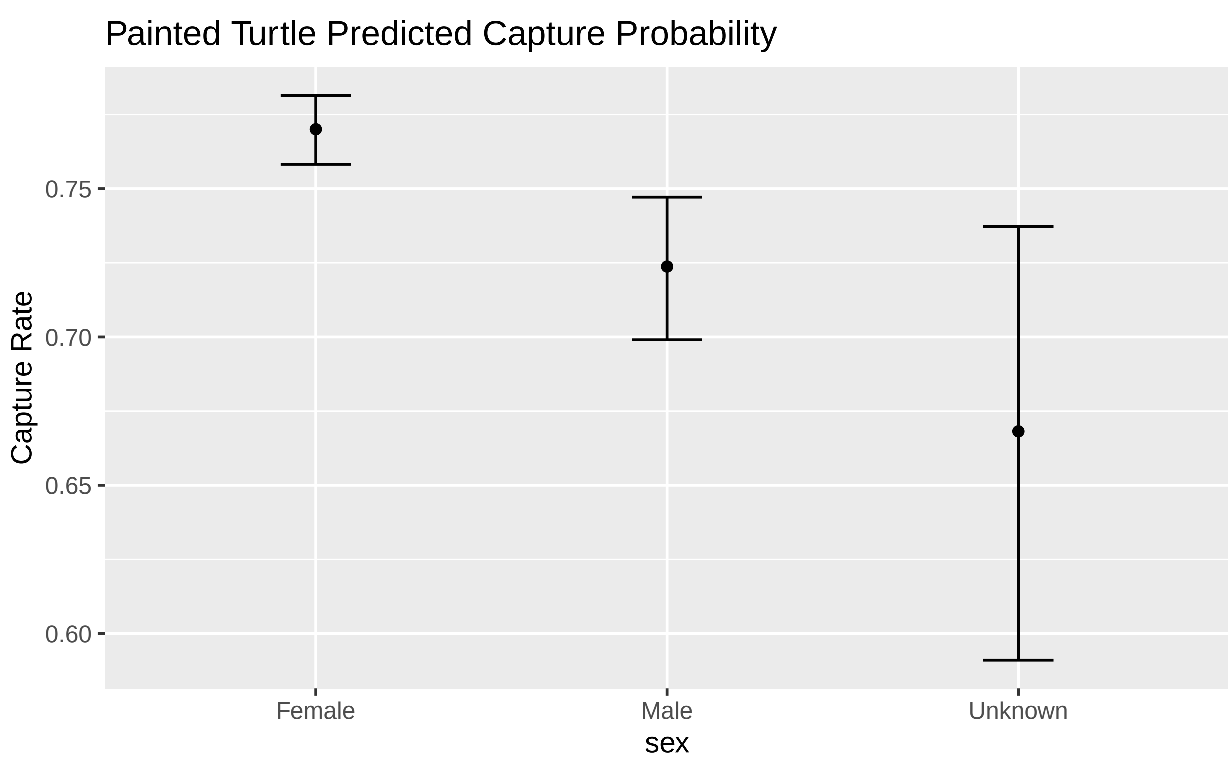
coefficient	Estimate	lcl	ucl
Phi.(Intercept)	40.552	33.760	48.709
Phi.sexMale	0.499	0.366	0.681
Phi.sexUnknown	0.032	0.025	0.042
p.(Intercept)	3.349	3.132	3.581
p.sexMale	0.782	0.680	0.900
p.sexUnknown	0.601	0.426	0.849

From the above table, we see that the odds ratio between the male turtle and female turtle survival rate is 0.499 with a 95% confidence interval of [0.366, 0.681]. This means the odds of survival rate for male turtle are about 50% lower than the odds of survival rate for female. Furthermore, the odds ratio between the Unknown sex turtle (juvenile) and the female turtle survival rate is 0.032 with a 95% confidence interval of [0.025, 0.042]. This means the odds of survival rate for Unknown sex turtles is 96.8% lower than the odds of survival rate for the female turtle.

Additionally, looking at the capture probability p, we see that male turtle has lower odds than female turtle, and unknown sex turtle has even lower odds ratio. The odds ratio of capture probability between male and female is 0.782 with a 95% confidence interval of [0.680, 0.9]. And the odds ratio of capture probability between unknown sex and female is 0.601 with a 95% confidence interval of [0.426, 0.849]



The odds of survival rate for female, male and unknown sex turtle are also reflected on the above plot. We predicted that female turtles have a high survival probability with a small 95% confidence interval. Male turtles have a slightly lower survival rate and slightly wider 95% confidence interval. Lastly, the unknown sex turtle has the lowest survival rate with only around 0.57%



Our model also predicted that the female turtle has the highest capture probability, around 77%. Male turtles have a slightly lower capture probability, around 72.5%. Lastly, the unknown sex turtle has the lowest estimated capture probability of 67%, with a very wide 95% confidence interval.

4. Disscussion

4.1 Result implication

From our CJS model result, we see that the female turtle has the highest survival probability. Male turtles have a slightly lower survival probability. Unsurprisingly, unknown sex turtles have a very low survival probability compared to male and female turtles. This is consistent with our knowledge of painted turtles since the female turtle is generally larger than males, so they are much less vulnerable. And the unknown sex turtle are mostly juvenile, which are very vulnerable.

We also observe the same trend for the capture probability that the female turtle has the highest capture probability, followed by male turtle then, unknown sex turtle. This means that we only capture about 77% of female turtles, 72.5% of male turtles, and only 67% of unknown sex turtle. This means we are likely to overestimate the survival probability because we do not have the data for the whole population. There are estimated 1/3 of juveniles are missing from our data, because they are likely to be dead and been eaten.

4.2 Model limitation

Our model has several limitations. First of all, there were many missing values in our dataset, and we have to remove some of the missing values and perform imputation. Removing and imputing the data could have a severe impact on the model result. Secondly, there was some input error in the data, which I chose to remove. This introduces more variation in our model result. Thirdly, some of our assumptions were not satisfied. For example, we cannot ensure that turtle emigrates from the study area is permanent and we are not releasing the turtles right away. Lastly, our model could not take advantage of all the size measurements and other variables in the dataset. Since adding a size measurement as a time-varying covariate leads to a non-significant result, which may because sex and size measurements are highly correlated. And I was not able to add a site random effect to our CJS model.

5. Conclusion

In conclusion, our model shows female turtles have a very high survival rate and capture rate. Male turtles have a slightly lower survival rate and capture compared to females. And the unknown sex (juveniles) turtle has a much lower survival rate probability and capture probability than female and male. However, there is some limitation of the model due to missing value, input error, imputation methods and the assumption that we may need further investigation.

Appendixes

Model comparison

For model 2, I impute the midpl for each turtle with increment of 0.15 every year after first capture, and we see that model 2 has non-significant midpl estimates. I also tried to impute the midpl with previous midpl measurement(with line 245-248) and got a negative estimate for midpl. This means that the larger turtle has lower survival probability, which is not correct based on my knowledge of painted turtle.

For model 3, we see non-siginificant result for both midpl and mass, and a larger AIC compared to model 1.

Therefore, I decided to use modle 1 as my final model.

Data Cleaning process

```r
# cleaning and transforming midpl variable
data2 <- data2 %>% filter(!is.na(midpl))
midpl <- data2 %>% group_by(notch) %>% spread(Year, midpl)
midpl <- full_join(midpl, datawithcapyear, by = "notch") %>% filter(!is.na(firstcapyear))


# imputation
year = 1991
column = 5
for(c in column:31) {
    for (i in 1:nrow(midpl)){
      if (midpl[i,32] > year){
        midpl[i, c] = 0
      }
      # impute by assumping each year turtles' midpl grow by 0.15
      # 0.15 is estimated by fitting a linear regression (midpl~age)
      else if (is.na(midpl[i,c])){
        if (year > 1991 && midpl[i,32] < year){
        midpl[i,c] = midpl[i, c-1] + 0.15
        # impute by previous observation
        # if(year > 1991 && is.na(midpl[i,c])){
        #   midpl[i,c] = midpl[1, c-1]
        # }
        }
      }

    }
  year = year + 1

}

# renaming the columns
midpl <- rename(midpl, c("midpl1" = "1991", "midpl2" = "1992", "midpl3" = "1993",
                         "midpl4" = "1994", "midpl5" = "1995", "midpl6" = "1996",
                         "midpl7" = "1997", "midpl8" = "1998", "midpl9" = "1999",
                         "midpl10" = "2000", "midpl11" = "2001", "midpl12" = "2002",
                         "midpl13" = "2003", "midpl14" = "2004", "midpl15" = "2005",
                         "midpl16" = "2006",
                         "midpl17" = "2007", "midpl18" = "2008", "midpl19" = "2009",
                         "midpl20" = "2010", "midpl21" = "2011", "midpl22" = "2012",
```

```
                             "midpl23" = "2013", "midpl24" = "2014", "midpl25" = "2015",
                             "midpl26" = "2016",
                             "midpl27" = "2017"))

midpl <- na.omit(midpl)


# cleaning and transforming mass variable

mass <- data1 %>% select(notch, mass, Year) %>% filter(!is.na(mass), Year >= 1991) %>%
  group_by(notch) %>% spread(Year, mass)
mass <- full_join(mass, datawithcapyear) %>% filter(!is.na(firstcapyear))

year = 1991
column = 2
for(c in column:28) {
    for (i in 1:nrow(mass)){
      if (mass[i,29] > year){
        mass[i, c] = 0
      }
      # impute by assumping each year turtles' mass grow by 8.9
      # 8.9 is estimated by fitting a linear regression (mass~age)
      else if (is.na(mass[i,c])){
        if (year > 1991 && mass[i,29] < year){
        mass[i,c] = mass[i, c-1] + 8.9
        }
      }

      # impute by previous observation
      # if(year > 1991 && is.na(mass[i,c])){
      #   mass[i,c] = mass[i, c-1]
      # }
    }
  year = year + 1
}

# renaming the columns
massdata <- rename(mass, c("mass1" = "1991", "mass2" = "1992", "mass3" = "1993",
                           "mass4" = "1994", "mass5" = "1995", "mass6" = "1996",
                           "mass7" = "1997", "mass8" = "1998", "mass9" = "1999",
                           "mass10" = "2000", "mass11" = "2001", "mass12" = "2002",
                           "mass13" = "2003", "mass14" = "2004", "mass15" = "2005",
                           "mass16" = "2006", "mass17" = "2007", "mass18" = "2008",
                           "mass19" = "2009", "mass20" = "2010", "mass21" = "2011",
                           "mass22" = "2012", "mass23" = "2013", "mass24" = "2014",
                           "mass25" = "2015","mass26" = "2016", "mass27" = "2017"))
massdata <- na.omit(massdata)

# Joing cleaned and transformed mass data and midpl data
cjsdata <- inner_join(midpl,massdata, by = "notch")
```

Model 1

Only use sex as static variable for both survival rate and capture rate

```r
 # Process data
turtles.proc = process.data(cjsdata)

 # Create design data with static and time varying covariates
design.Phi=list(static=c("sex"))
design.p=list(static=c("sex"))

 design.parameters=list(Phi=design.Phi,p=design.p)
 ddl=make.design.data(turtles.proc,parameters=design.parameters)

Phi.sfw=list(formula=~sex)
p.ast=list(formula=~sex)
model=crm(turtles.proc,ddl,hessian=TRUE,model.parameters=list(Phi=Phi.sfw,p=p.ast))
```

```
##  Number of evaluations:  100  -2lnl: 9578.847439 Number of evaluations:  200  -2lnl:  9578.23544 Numb
```

```r
model$results
```

```
##
## crm Model Summary
##
## Npar :  6
## -2lnL:  9578.233
## AIC  :  9590.233
##
## Beta
##                   Estimate         se        lcl         ucl
## Phi.(Intercept)  3.7124726 0.09325228  3.5296981  3.89524707
## Phi.sexMale     -0.7018276 0.15836980 -1.0122324 -0.39142277
## Phi.sexUnknown  -3.4654889 0.13420389 -3.7285286 -3.20244931
## p.(Intercept)    1.2067197 0.03381580  1.1404407  1.27299866
## p.sexMale       -0.2335175 0.07135187 -0.3733672 -0.09366785
## p.sexUnknown    -0.4133194 0.17844295 -0.7630675 -0.06357117
```

Model 2

Use sex as static variable and midpl as time varying covariates for both survival rate and capture rate

```r
 # Process data
turtles.proc = process.data(cjsdata)

 # Create design data with static and time varying covariates
design.Phi=list(static=c("sex"),time.varying=c("midpl"))
design.p=list(static=c("sex"), time.varying=c("midpl"))

 design.parameters=list(Phi=design.Phi,p=design.p)
 ddl=make.design.data(turtles.proc,parameters=design.parameters)

Phi.sfw=list(formula=~sex + midpl)
p.ast=list(formula=~sex + midpl)
model=crm(turtles.proc,ddl,hessian=TRUE,model.parameters=list(Phi=Phi.sfw,p=p.ast))
```

```
##  Number of evaluations:  100  -2lnl:  9562.03937 Number of evaluations:  200  -2lnl: 9559.742257 Numb
```

```r
model
```

```
##
```

```
## crm Model Summary
##
## Npar :  8
## -2lnL:  9559.675
## AIC  :  9575.675
##
## Beta
##                     Estimate         se         lcl          ucl
## Phi.(Intercept)  4.23714365 0.35835620  3.53476549  4.939521808
## Phi.sexMale     -0.77303211 0.16501461 -1.09646076 -0.449603471
## Phi.sexUnknown  -3.74855663 0.24210198 -4.22307650 -3.274036759
## Phi.midpl       -0.04042159 0.02597781 -0.09133810  0.010494923
## p.(Intercept)    0.59823175 0.15023641  0.30376839  0.892695101
## p.sexMale       -0.14866392 0.07424099 -0.29417625 -0.003151585
## p.sexUnknown    -0.16370748 0.18755366 -0.53131266  0.203897700
## p.midpl          0.04623172 0.01118058  0.02431778  0.068145667
```

Model 3

Use sex as static variable and midpl and mass as time varying covariates for both survival rate and capture rate

```
 # Process data
turtles.proc = process.data(cjsdata)


 # Create design data with static and time varying covariates
design.Phi=list(static=c("sex"),time.varying=c("midpl", 'mass'))
design.p=list(static=c("sex"),time.varying=c("midpl", 'mass'))

 design.parameters=list(Phi=design.Phi,p=design.p)
 ddl=make.design.data(turtles.proc,parameters=design.parameters)


Phi.sfw=list(formula=~sex+midpl+mass)
p.ast=list(formula=~sex+midpl+mass)
model=crm(turtles.proc,ddl,hessian=TRUE,model.parameters=list(Phi=Phi.sfw,p=p.ast))
```

```
##  Number of evaluations:  100  -2lnl: 9533.305248 Number of evaluations:  200  -2lnl: 9514.593452 Numb
```

```
model
```

```
##
## crm Model Summary
##
## Npar :  10
## -2lnL:  9510.868
## AIC  :  9530.868
##
## Beta
##                     Estimate           se          lcl           ucl
## Phi.(Intercept)  4.105611602 0.3799420713  3.360925143  4.8502980620
## Phi.sexMale     -0.851328325 0.1737334819 -1.191845950 -0.5108107009
## Phi.sexUnknown  -3.768903665 0.2441615140 -4.247460232 -3.2903470971
## Phi.midpl        0.004222225 0.0445895316 -0.083173257  0.0916177071
## Phi.mass        -0.001129540 0.0008608734 -0.002816851  0.0005577723
```

```
## p.(Intercept)   -0.048326762 0.1923799374 -0.425391439  0.3287379152
## p.sexMale       -0.274893821 0.0770817904 -0.425974130 -0.1238135118
## p.sexUnknown    -0.282622777 0.1860889936 -0.647357205  0.0821116503
## p.midpl          0.175469429 0.0245003096  0.127448822  0.2234900363
## p.mass          -0.002648714 0.0004247474 -0.003481219 -0.0018162092
```