Homework_1

2025-01-15

#1: #The range of the common loon varies in the winter and summer. During the breeding season in the summer months they inhabit the freshwater systems in Canada and the northern U.S. In the winter they inhabit the southern atlantic and pacific coasts of the U.S and extend even as far as northern Mexico. So yes they have seperate ranges for the breeding and nonbreeding seasons and yes they are migratory. They prefer freshwater lakes or ponds with very clear water so that they can effectively hunt for prey during the breeding season. Their diet primaily consists of freshwater fish but will also occasionally contain crustacieans and mollusks.

#2: #For the first covariate, fish abundance, I would initially expect loon abundance to be higher in areas where there is a lot of fish however this may be misleading becasue lots of loons in one area might cause fish abundance to go down affect this relationship. My hypothesis for the second covariate, sechhi depth or water quality, is that loon abundance will go up with higher sechhi depth values since loons like clear water for hunting so much.

```
#3:
library(tidyr)
loon <- read.csv("loons.csv")
print(loon)
```

```
##     y      secchi        fish
## 1   0 0.789636338 0.5607780
## 2   1 0.499180505 2.9089782
## 3   1 0.628640731 3.7522151
## 4   1 0.922113765 2.3733732
## 5   1 0.509853943 3.2624076
## 6   1 0.585470191 3.1601256
## 7   1 0.473474896 2.8113611
## 8   1 0.656063013 2.6429035
## 9   1 0.593938123 3.6244432
## 10  0 0.662965918 2.2274086
## 11  0 0.166833242 0.0000000
## 12  1 0.466973742 2.4751029
## 13  1 0.775394473 2.7494446
## 14  0 0.247514589 1.2813835
## 15  0 0.823274035 2.2894691
## 16  1 0.707509453 0.0000000
## 17  1 0.496472385 2.2566887
## 18  1 1.005157552 2.8302943
## 19  1 0.563863433 2.3606973
## 20  0 0.916519262 1.3433279
## 21  1 0.466455745 2.6205250
## 22  1 0.454869856 2.7478954
## 23  0 0.106078857 1.7252142
## 24  0 0.440935797 3.7297510
## 25  1 0.843602622 2.5484833
## 26  1 0.800789340 2.8634067
```

```
## 27 1 0.538108576 0.5107353
## 28 0 0.094176550 1.1457460
## 29 1 0.813592229 1.8148031
## 30 1 0.826169924 1.2770733
## 31 1 0.667552438 1.2157700
## 32 0 0.025334563 1.3843847
## 33 1 0.774658192 3.6662143
## 34 1 0.788560178 3.6756793
## 35 1 0.709552972 2.4830206
## 36 0 0.676635907 2.6466459
## 37 0 0.072660178 2.2752530
## 38 1 0.768050620 2.2780461
## 39 1 0.951521045 0.9480208
## 40 0 0.484714421 1.3562800
## 41 1 1.016778926 3.4676999
## 42 1 0.684039188 1.3456216
## 43 1 0.256744977 2.4243581
## 44 1 0.871015420 3.3210718
## 45 0 0.892494158 2.6541500
## 46 0 0.009192966 1.8130362
## 47 1 0.751654028 3.1256067
## 48 1 0.542067873 4.4179252
## 49 1 0.636704755 1.5269971
## 50 1 0.840420801 2.2768123
## 51 1 0.686782736 4.1197560
## 52 0 0.829127306 1.6494785
## 53 0 0.708139689 2.5040751
## 54 1 0.688601219 1.0441945
## 55 1 0.490015956 2.4388655
## 56 0 0.817869235 1.5915134
## 57 0 0.061393257 1.7385712
## 58 1 0.777842666 2.9442176
## 59 1 0.630700109 1.6053124
```

```
dim(loon)
```

```
## [1] 59  3
```

```
head(loon)
```

```
##   y    secchi     fish
## 1 0 0.7896363 0.560778
## 2 1 0.4991805 2.908978
## 3 1 0.6286407 3.752215
## 4 1 0.9221138 2.373373
## 5 1 0.5098539 3.262408
## 6 1 0.5854702 3.160126
```

```
#The distripution that best describes the y column of data is a binomial distribution.
```

```
#4:
mod1 <- glm(loon$y ~ 1, family = "binomial")
summary(mod1)
```

```
##
## Call:
## glm(formula = loon$y ~ 1, family = "binomial")
```

```
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -1.5054  -1.5054   0.8817   0.8817   0.8817
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.7444     0.2786   2.672  0.00754 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 74.15  on 58  degrees of freedom
## Residual deviance: 74.15  on 58  degrees of freedom
## AIC: 76.15
##
## Number of Fisher Scoring iterations: 4
```

```r
mod2 <- glm(loon$y ~ loon$fish, family = "binomial")
summary(mod2)
```

```
##
## Call:
## glm(formula = loon$y ~ loon$fish, family = "binomial")
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -2.0808  -1.2031   0.6212   0.8082   1.6279
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.0161     0.7425  -1.369   0.1711
## loon$fish     0.8202     0.3321   2.470   0.0135 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 74.150  on 58  degrees of freedom
## Residual deviance: 67.005  on 57  degrees of freedom
## AIC: 71.005
##
## Number of Fisher Scoring iterations: 3
```

```r
mod3 <- glm(loon$y ~ loon$secchi, family = "binomial")
summary(mod3)
```

```
##
## Call:
## glm(formula = loon$y ~ loon$secchi, family = "binomial")
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -2.0147  -0.8033   0.6491   0.8240   1.3388
```

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.2515     0.7658  -1.634  0.10222
## loon$secchi   3.4263     1.2363   2.771  0.00558 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 74.15  on 58  degrees of freedom
## Residual deviance: 65.08  on 57  degrees of freedom
## AIC: 69.08
##
## Number of Fisher Scoring iterations: 4
```

```
mod4 <- glm(loon$y ~ loon$fish + loon$secchi, family = "binomial")
summary(mod4)
```

```
##
## Call:
## glm(formula = loon$y ~ loon$fish + loon$secchi, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0856  -0.7144   0.5228   0.8394   1.4367
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.6278     1.0721  -2.451   0.0142 *
## loon$fish     0.7221     0.3480   2.075   0.0379 *
## loon$secchi   3.0985     1.2914   2.399   0.0164 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 74.15  on 58  degrees of freedom
## Residual deviance: 60.25  on 56  degrees of freedom
## AIC: 66.25
##
## Number of Fisher Scoring iterations: 4
```

#5a: Since the intercept value for model 1 is .7444, by using the log-odds to probability formula I can calulculate that the propability of any given site having a loon is .678 or 67.8 percent with matches with teh data set since 40/59 sites that were studied had loons.

```
#5b:
AIC(mod1, mod2, mod3,mod4)
```

```
##      df      AIC
## mod1  1 76.15038
## mod2  2 71.00469
## mod3  2 69.08007
## mod4  3 66.25002
```

#5c: With the addition of each coviarite there was a subsequent decrease in AIC. This means that there is a positive correlation between both covariates and the liklihood of seeing a loon at a site. In other words if either covariate increases, whether that is teh water quality getting clearer or there being more fish at teh site, there is a higher liklihood for one to find loons there. this is why the top model turned out to be that one that incorperated both covariates.