Part I:

1) mod1 = glm(Bycatch ~ Season + Area + Gear.Type+Time+offset(log(Tows)), data = byca, family = poisson)

2)

|  |  |  |  |
| --- | --- | --- | --- |
| Model variables | AreaSouth | Gear.TypeMid-Water | TimeNight |
| Estimates | 1.822450 | 1.917687 | 2.176564 |

Mean bycatch ratio is generally consistent across all three variables and is only slightly higher for nighttime. Nighttime mean and SE not high enough to exert a strong-seeming effect.

3) He is not right - the seasonal estimates introduce so much variation (esp. 90-91) in the relationship of bycatch to towing. This obscures understanding any effects area, gear, and time may or may not actually have, and seems to lead to artificially inflated mean ratios that seem higher than they should be.

4) The estimate and error reported for Season 1991-1992 wildly deviates from the other estimates.

A standard error of almost 2500 does not make sense for this data, especially if for all other variables it's between 0-1.

Call:

glm(formula = Bycatch ~ Season + Area + Gear.Type + Time + offset(log(Tows)),

family = poisson, data = byca)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.2576 -0.6964 -0.3018 -0.0002 3.2730

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -7.32782 0.59038 -12.412 < 2e-16 \*\*\*

Season1990-91 -19.66524 2439.00216 -0.008 0.9936

Season1991-92 1.81895 0.28503 6.382 1.75e-10 \*\*\*

Season1992-93 0.07391 0.39557 0.187 0.8518

Season1993-94 -0.93191 0.41132 -2.266 0.0235 \*

Season1994-95 -0.13503 0.30923 -0.437 0.6623

AreaSouth 1.82245 0.41063 4.438 9.07e-06 \*\*\*

Gear.TypeMid-Water 1.91769 0.44308 4.328 1.50e-05 \*\*\*

TimeNight 2.17656 0.45138 4.822 1.42e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 335.096 on 42 degrees of freedom

Residual deviance: 42.078 on 34 degrees of freedom

AIC: 97.461

Number of Fisher Scoring iterations: 18

5) mods2 = glmer(Bycatch~Area+Gear.Type+Time+(1|Season), data=byca, family=poisson)

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']

Family: poisson ( log )

Formula: Bycatch ~ Area + Gear.Type + Time + (1 | Season) + offset(log(Tows))

Data: byca

AIC BIC logLik deviance df.resid

116.3 125.1 -53.2 106.3 38

Scaled residuals:

Min 1Q Median 3Q Max

-1.6007 -0.5736 -0.3236 -0.1313 5.0876

Random effects:

Groups Name Variance Std.Dev.

Season (Intercept) 2.389 1.546

Number of obs: 43, groups: Season, 6

Fixed effects:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -7.7505 0.8397 -9.230 < 2e-16 \*\*\*

AreaSouth 1.8115 0.4052 4.471 7.80e-06 \*\*\*

Gear.TypeMid-Water 1.8872 0.4355 4.334 1.47e-05 \*\*\*

TimeNight 2.2028 0.4425 4.978 6.43e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:

(Intr) AreSth G.TM-W

AreaSouth -0.372

Gr.TypMd-Wt -0.161 -0.215

TimeNight -0.277 0.051 -0.469

6) Worst described observations have little consistently in common save for that they all had Bycatch values of 5 or 6.





Part II:

#Molly Jenkins

#ENEC563

#Assignment 9

#04/24/2017

# These data are Example 3.3 of Manly (2001), pp. 96–100, who obtained the data from Baird (1996).

#Manly describes the data as follows.

#

# The accidental capture of marine mammals and birds in commercial fishing operations

#is of considerable concern in many fisheries around the world.

#This example concerns one such situation, which is of catches of the common dolphin (Delphinus delphis)

#and the bottlenose dolphin (Tursiops truncatus) in the Taranaki Bight trawl fishery for jack mackerel

#(Trachurus declivus, T. novaezalandiae, and T. murphyi) off the west coast of New Zealand.

#The New Zealand Ministry of Fisheries puts official observers on about 10% of fishing vessels

#to monitor dolphin bycatch. Data were collected by these observers for the six fishing seasons

#1989/90 to 1994/95

#The table shows the number of observed trawls (Tows)

#and the number of dolphins accidentally killed (Bycatch)

#categorized by eight conditions for each fishing year:

#the fishing area (the northern or southern Taranaki Bight),

#by the gear type (bottom or midwater),

#by the time (day or night).

#Excluding five cases where there were no observed trawls, this gives 43 observations

#on the bycatch under different conditions, in different years.

####Pre-assignment code####

setwd("C:/git/coursework/ENEC563")

library(ggplot2)

library(dplyr)

library(MASS)

library(gamlss)

library(maps)

library(lme4)

byca = read.csv("https://sakai.unc.edu/access/content/group/7d7a0e1c-4adb-4ee2-ace8-490a89313a59/Data/bycatch.csv", header = TRUE)

# Questions:

#1) Manly (2001) fits a Poisson regression model to these data in which he includes season,

# area, gear type, and time of day as factors in a main effects model.

# He fits the model in such a way that the response variable, Bycatch, can be interpreted

# as a bycatch rate per tow, the number of dolphins killed per tow. Fit this model.

# Recall that rates can be fit by adding an offset with the log of the tow variable.

mod1 = glm(Bycatch ~ Season + Area + Gear.Type+Time+offset(log(Tows)), data = byca, family = poisson)

summary(mod1)

#confint(mod1) #throwing up errors!

# 2) Obtain the estimates of the area, gear type, and time effects.

# Interpret what these estimates represent in practical terms.

# For instance, how do the bycatch rates compare for night versus day,

# south versus north, mid-water versus bottom?

ests = coef(mod1)[7:9]

#Bycatch generally consistent across all three variables; similar means and SE's, slightly higher for nighttime

#but not much higher to exert a strong effect.

# 3) Manly (2001), p. 100, states the following:

# "All effects are highly significant in terms of the reduction in the deviance

# that is obtained by adding them into the model,

# and the final model gives a good fit to the data (chi-squared = 42.08 with 34 degrees of freedom, p = 0.161)."

# Is Manly right? Why or why not?

###He is not right - the seasonal estimates introduce so much variation (esp. 90-91) on bycatch

###in the relationship of bycatch to towing.

###This obscures understanding any effects area, gear, and time may or may not have.

# 4) Examine the summary table.

# Do you see anything unusual in what's reported there

# to suggest that maybe there is a problem with this model? What?

###The estimate and error reported for Season 1991-1992 wildly deviates from the other estimates.

###A standard error of almost 2500 does not make sense for this data, especially if for all other variables

### it's between 0-1.

# 5) In light of the problem you saw in question 4, refit the model

# but this time treat Season as a random effect rather than a fixed effect.

mods2 = glmer(Bycatch~Area+Gear.Type+Time+(1|Season), data=byca, family=poisson)

summary(mods2)

ests2 = coef(mods2)

# 6) Carry out a graphical goodness of fit test (as in the last assignment)

# for the random effects model. What observations appear to be poorly described by the model?

byca$lam <- fitted(mods2)

byca$rat = byca$Bycatch/byca$Tows

upper.p = 1-ppois(byca$Bycatch-1, lambda =byca$lam)

upper.p

sum(upper.p<.025) # = 3

byca$Tows[upper.p<.025]

lower.p <- ppois(byca$Bycatch, lambda =byca$lam)

lower.p

sum(lower.p<.025) #= 0

min(lower.p)

pval.dat <- data.frame(pvalue=c(lower.p, upper.p), Tows=rep(byca$Season,2),

label=rep(c('lower', 'upper'), each=nrow(byca)))

ggplot(pval.dat,aes(x=pvalue,y=Tows))+geom\_point()+facet\_wrap(~label)+

geom\_vline(xintercept=.025,color="red",linetype=2)+

theme(panel.grid.minor.x=element\_blank(),panel.grid.major.x=element\_blank())

byca$pplot <- ifelse(ppois(byca$Bycatch, lambda =byca$lam) < .5,

ppois(byca$Bycatch, lambda =byca$lam),

1-ppois(byca$Bycatch-1, lambda =byca$lam))

byca$lower <- ifelse(ppois(byca$Bycatch, lambda =byca$lam) < .5,

"lower","upper")

plot1 = ggplot(byca,aes(x=pplot,y=Tows,color=lower))+geom\_point()+

geom\_vline(xintercept=.025,linetype=2)+

labs(x = "", color="Tail of distribution")+theme(legend.position = "none")

plot2 = ggplot(byca,aes(x=pplot,y=Season,color=lower))+geom\_point()+

geom\_vline(xintercept=.025,linetype=2)+theme(legend.position = "none")+

labs(x="Tail probability of observation from model",color="Tail of distribution")

library(gridExtra)

library(grid)

library(ggmap)

#create a shared legend

grid\_arrange\_shared\_legend <- function(...) {

plots <- list(...)

g <- ggplotGrob(plots[[1]] + theme(legend.position="right"))$grobs

legend <- g[[which(sapply(g, function(x) x$name) == "guide-box")]]

lw <- sum(legend$width)

gl <- lapply(plots, function(x) x + theme(legend.position="none"))

grid.arrange(arrangeGrob(grobs = gl), legend,

ncol = 2, widths = unit.c(unit(1, "npc") - lw, lw))

}

grid\_arrange\_shared\_legend(plot1, plot2)

ggplot(byca, aes(x=rat, y = Season))+geom\_point(aes(color = "raw data"), size = 1)+

geom\_point(aes(x = lam, color = "Conditional means"), size = 4, shape = "|")+

theme(legend.title = element\_blank())+xlab("Bycatch/Tow Ratio")

#some of our predications are still way high, way off but the rest seem to pair with the raw data really closely