

# Homework8

## Question 1

Bolger et al (1997) investigated the impacts of habitat fragmentation on the occurrence of native rodents. Modeled the presence/absence of any species of native rodent (except *Microtus californicus*) against three predictor variables: distance (meters) to nearest source canyon (DISTX), age (years) since fragment was isolated by urbanization (AGE), and percentage of fragment area covered in shrubs (PERSHRUB)

- RV = RODENTSP
- EVs = DISTX, AGE, PERSHRUB

```
library(rms)

## Loading required package: Hmisc

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##       format.pval, units

library(car)

## Loading required package: carData

##
## Attaching package: 'car'

## The following objects are masked from 'package:rms':
##       Predict, vif

library(MuMIN)
library(ggplot2)

# Data
ds <- readxl::read_excel('/Users/kanoalindiwe/Downloads/Projects/playground/R/Quantitative Ecology/Data')

PERSHRUB <- ds$PERSHRUB
DISTX <- ds$DISTX
AGE <- ds$AGE
RODENTSP <- as.factor(ds$RODENTSP)

# 1) Check collinearity of explanatory variables cor(bolger[,1:3])
cor(ds[,1:3])
```

```

##          PERSHRUB      DISTX       AGE
## PERSHRUB  1.0000000 -0.2745893 -0.7952976
## DISTX     -0.2745893  1.0000000  0.2789722
## AGE       -0.7952976  0.2789722  1.0000000

# Pass, keep all variables

# 2) Specify a full logistic model (note- keep all variables that correlate below 0.80)
glm1 <- glm(RODENTSP ~ PERSHRUB + DISTX + AGE, family = binomial, data = ds)

#use function "summary" to summarize model
summary(glm1)

```

```

##
## Call:
## glm(formula = RODENTSP ~ PERSHRUB + DISTX + AGE, family = binomial,
##      data = ds)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.9099159  3.1125426 -1.899   0.0576 .
## PERSHRUB    0.0958695  0.0406119  2.361   0.0182 *
## DISTX       0.0003087  0.0007741  0.399   0.6900
## AGE         0.0250077  0.0376618  0.664   0.5067
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 34.617 on 24 degrees of freedom
## Residual deviance: 19.358 on 21 degrees of freedom
## AIC: 27.358
##
## Number of Fisher Scoring iterations: 5

```

```

# 3) examine for overdispersion (Residual deviance/residual df)
data.lrm <- lrm(RODENTSP ~ PERSHRUB + DISTX + AGE, data = ds, y = TRUE, x = TRUE)
resid(data.lrm, type = "gof")

```

```

## Sum of squared errors      Expected value|H0           SD
##                      3.1538988      3.0291377      0.1382219
##                      Z                  P
##                      0.9026146      0.3667305

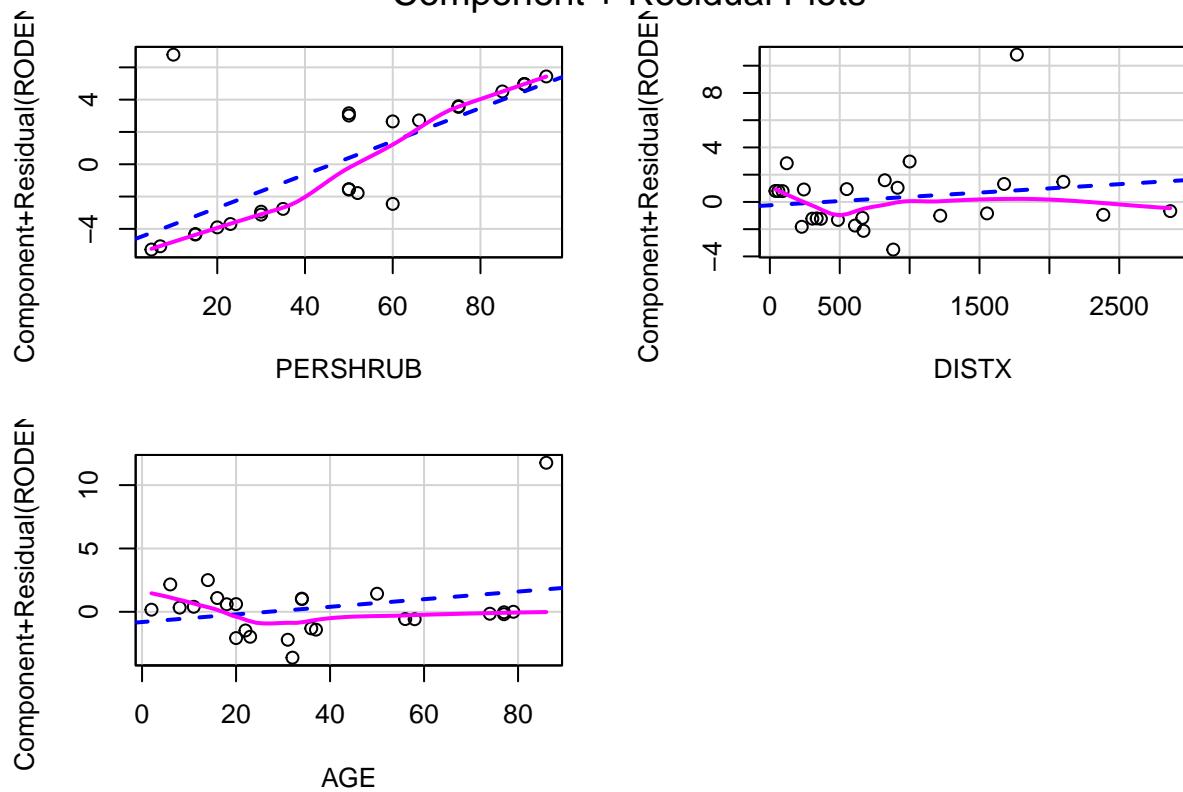
```

```

# 4) Check for linearity
crPlots(glm1, ask = FALSE)

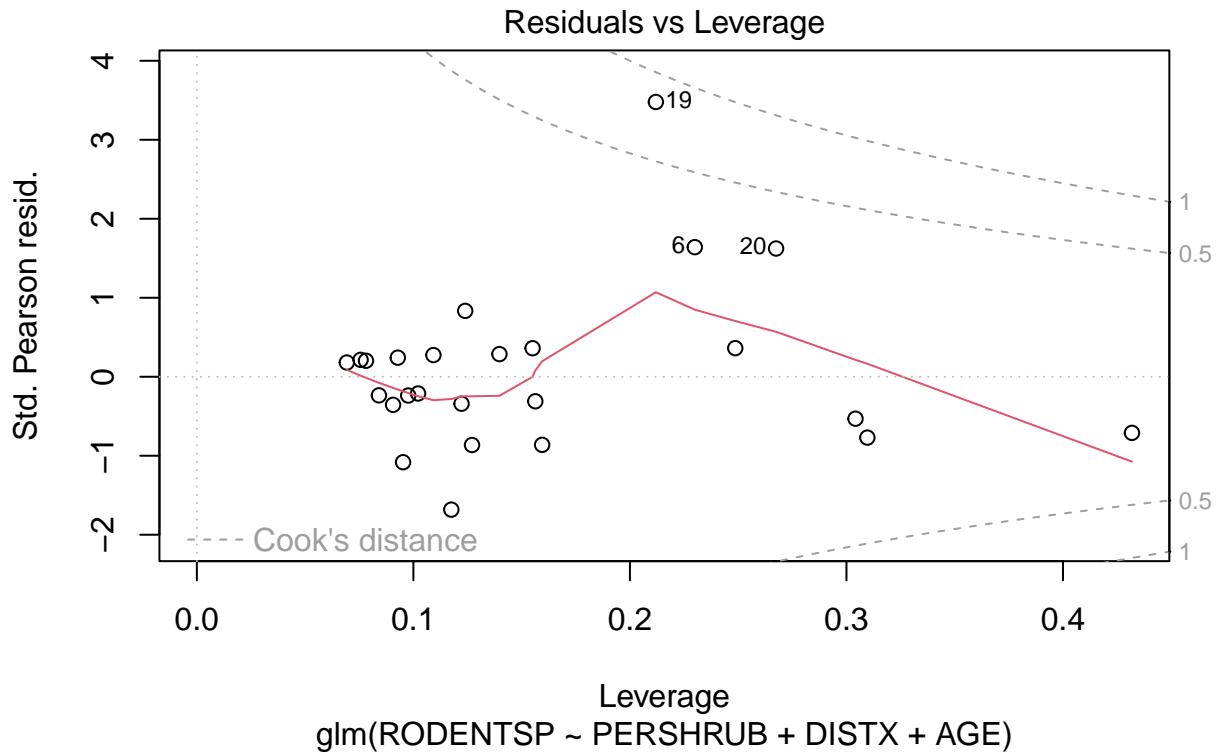
```

### Component + Residual Plots



```
# Pass, data points are generally linear
```

```
# 5) Check for influential observations  
plot(glm1, which = 5)
```



```
# Pass, does not cross cook's lines

# 6) interpret best model and examine odds ratio for presence of rodents (based on the full
# model) *use exp(coef(bolger1)) to get odds ratios
exp(coef(glm1))

## (Intercept)      PERSHRUB        DISTX         AGE
## 0.002712415 1.100615371 1.000308770 1.025322977

# The odds ratios show that shrub cover is the primary factor influencing rodent presence.
# For each 1% increase in shrub cover, the odds of finding native rodents increase by about
# 10%. Distance to the nearest canyon and fragment age show minimal influence, suggesting they
# do not meaningfully alter the likelihood of rodents

# 7) use AIC to determine the best model (use functions "dredge", "model.avg", "summary" and
# "importance" from library MuMin), and using model below: bolger2<-glm(RODENTSP~DISTX+AGE+PERSHRUB, bi
glm2<-glm(RODENTSP~DISTX+AGE+PERSHRUB, binomial, na.action="na.fail")
glm2.all <- dredge(glm2)

## Fixed term is "(Intercept)"

summary(glm2.all)

##      (Intercept)          AGE           DISTX        PERSHRUB
```

```

## Min.   :-5.9099   Min.   :-0.044654   Min.   :-0.0003266   Min.   :0.07662
## 1st Qu.:-4.7303   1st Qu.:-0.044210   1st Qu.:-0.0000421   1st Qu.:0.07963
## Median :-1.9917   Median :-0.009527   Median : 0.0001807   Median :0.08716
## Mean   :-2.1007   Mean   :-0.009370   Mean   : 0.0001007   Mean   :0.08670
## 3rd Qu.: 0.5121   3rd Qu.: 0.025313   3rd Qu.: 0.0003235   3rd Qu.:0.09423
## Max.   : 1.4643   Max.   : 0.026230   Max.   : 0.0003680   Max.   :0.09587
##          NA's    :4          NA's    :4          NA's    :4          NA's    :4
##          df      logLik      AICc      delta
## Min.   :1.0   Min.   :-17.309   Min.   :24.59   Min.   : 0.000
## 1st Qu.:2.0   1st Qu.:-15.127   1st Qu.:26.87   1st Qu.: 2.278
## Median :2.5   Median :-12.241   Median :31.41   Median : 6.818
## Mean   :2.5   Mean   :-12.839   Mean   :31.58   Mean   : 6.988
## 3rd Qu.:3.0   3rd Qu.:-9.865   3rd Qu.:36.24   3rd Qu.:11.647
## Max.   :4.0   Max.   :-9.679   Max.   :38.79   Max.   :14.195
##
##          weight
## Min.   :0.0004656
## 1st Qu.:0.0016852
## Median :0.0293493
## Mean   :0.1250000
## 3rd Qu.:0.1806014
## Max.   :0.5630446
##

```

```

# Model averaging is not recommended as AIC > 2, however, for the sake of homework,
# I will perform it.

```

```

avg_mod <- model.avg(glm2.all, subset = delta < 4)
summary(avg_mod)

```

```

##
## Call:
## model.avg(object = glm2.all, subset = delta < 4)
##
## Component model call:
## glm(formula = RODENTSP ~ <3 unique rhs>, family = binomial, na.action =
##       na.fail)
##
## Component models:
##          df logLik  AICc delta weight
## 3        2 -10.02  24.59  0.00   0.60
## 13      3  -9.76  26.66  2.06   0.21
## 23      3  -9.90  26.95  2.35   0.19
##
## Term codes:
##          AGE     DISTX PERSHRUB
##            1         2         3
##
## Model-averaged coefficients:
## (full average)
##           Estimate Std. Error Adjusted SE z value Pr(>|z|)
## (Intercept) -4.361e+00  2.103e+00  2.211e+00  1.972   0.0486 *
## PERSHRUB    8.102e-02  3.240e-02  3.417e-02  2.371   0.0177 *
## AGE         5.618e-03  1.997e-02  2.080e-02  0.270   0.7871

```

```

## DISTX      6.822e-05  3.486e-04   3.655e-04   0.187   0.8519
##
## (conditional average)
##           Estimate Std. Error Adjusted SE z value Pr(>|z|)
## (Intercept) -4.3608734  2.1026986  2.2112653  1.972   0.0486 *
## PERSHRUB     0.0810198  0.0324030  0.0341650  2.371   0.0177 *
## AGE         0.0262301  0.0363578  0.0384709  0.682   0.4954
## DISTX       0.0003680  0.0007383  0.0007812  0.471   0.6376
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
sw(glm2.all)
```

```

##                  PERSHRUB AGE  DISTX
## Sum of weights:    0.99      0.26  0.23
## N containing models: 4        4     4

```

# *PERSHRUB* is by far the most important variable.

```

#use deletion tests to compare best model with the next best model(s) as a final way to
# determine whether to include specific explanatory variables (see Textbook example for this)
glm_full <- glm(RODENTSP ~ PERSHRUB + DISTX + AGE, family = binomial, data = ds)
glm_no_shrub <- update(glm_full, ~ . - PERSHRUB)
glm_no_age <- update(glm_full, ~ . - AGE)
glm_no_dist <- update(glm_full, ~ . - DISTX)

```

```
anova(glm_full, glm_no_shrub, test = "Chisq")
```

```

## Analysis of Deviance Table
##
## Model 1: RODENTSP ~ PERSHRUB + DISTX + AGE
## Model 2: RODENTSP ~ DISTX + AGE
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       21     19.358
## 2       22     28.915 -1   -9.5577 0.001991 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
anova(glm_full, glm_no_age, test = "Chisq")
```

```

## Analysis of Deviance Table
##
## Model 1: RODENTSP ~ PERSHRUB + DISTX + AGE
## Model 2: RODENTSP ~ PERSHRUB + DISTX
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       21     19.358
## 2       22     19.802 -1  -0.44461   0.5049

```

```
anova(glm_full, glm_no_dist, test = "Chisq")
```

```
## Analysis of Deviance Table
```

```

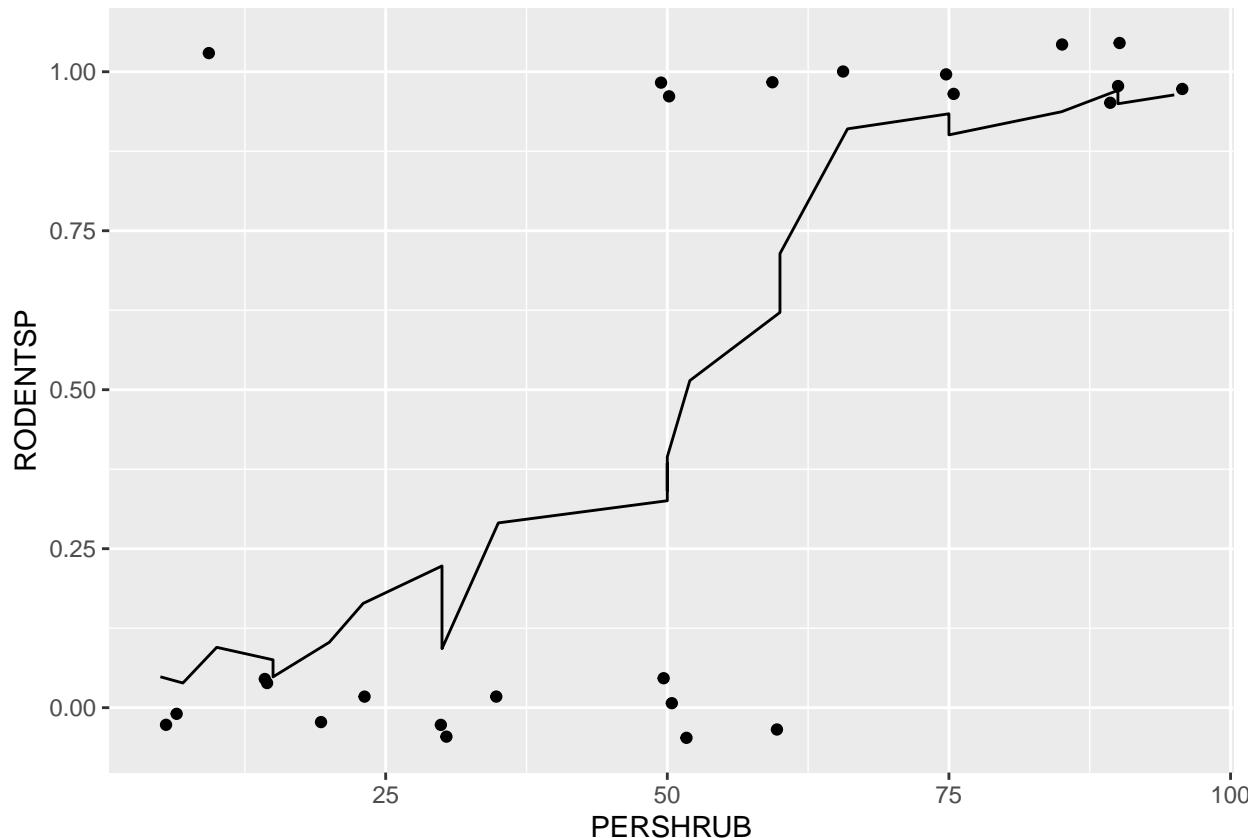
## 
## Model 1: RODENTSP ~ PERSHRUB + DISTX + AGE
## Model 2: RODENTSP ~ PERSHRUB + AGE
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        21     19.358
## 2        22    19.514 -1 -0.15592   0.6929

# The results of the anovas show that AGE and DISTX do not meaningfully reduce the model
# performance, while PERCHRUB is significantly ( $P = 0.001991$ ). This indicates that the
# most parsimonious model only includes PERCHRUB.

# Re-interpret best model based on model averaging results
# No model averaging is required here.

# extra- 2 pts, Produce a graph that illustrates the relationship between percent
# shrub cover and presence/absence of rodents
ds$pred <- predict(glm1, type = "response")
ggplot(ds, aes(PERSHRUB, RODENTSP)) +
  geom_jitter(height = 0.05) +
  geom_line(aes(y = pred),)

```



## Question 2

A study was conducted to determine if wing molt in Amakihi (N or Y) was related to seasonal period (1-6), elevation(low or high), and sex. Use the amakihi dataset to determine the best fitting model for predicting wing molt in this bird.

```
# a) ensure that wing.molt and sex are being read as categorical variables by R
# (and change them if they aren't (period should be a continuous variable here)
ds <- readxl::read_excel('/Users/kanoalindiwe/Downloads/Projects/playground/R/Quantitative Ecology/Data/Amakihi.xlsx')

period <- ds$period
elevation <- as.factor(ds$elevation)
sex <- as.factor(ds$sex)
molt <- as.factor(ds$`wing molt`)

# b) create a full model for wing.molt, including an interaction term between elevation and sex
amakihi.lrm <- lrm(molt ~ period + elevation + sex + elevation*sex, data = ds, y = TRUE, x = TRUE)

# c) consider checking diagnostics for this model. Based on what we have discussed in lecture, what (if
resid(amakihi.lrm, type = "gof")

## Sum of squared errors      Expected value|H0          SD
##            3.711791e+01      3.581278e+01      1.725835e-01
##            Z                  P
##            7.562303e+00      3.959937e-14

# Lack of good fit as we have a high p value. Dont really need to run
# anything else, we need to fix it.

# d) use functions from library MuMin to assess the best fitting model
# from the global model. Provide comparison of AICc values, model weights, and model averages in your analysis
glm.full <- glm(molt ~ period + elevation + sex + elevation*sex, family = binomial, na.action="na.fail")
dredged <- dredge(glm.full)

## Fixed term is "(Intercept)"

summary(dredged)
```

```
##   (Intercept)    elevation    period      sex  elevation:sex
## Min.   :-18.709   +  :6   Min.   :0.4216   +  :6   +  :2
## 1st Qu.: -4.438   NA's:4   1st Qu.:0.4269   NA's:4   NA's:8
## Median : -4.013           Median :0.4284
## Mean   : -6.397           Mean   :0.4272
## 3rd Qu.: -2.843           3rd Qu.:0.4288
## Max.   : -2.433           Max.   :0.4305
## 
##                   NA's   :5
## 
##      df      logLik      AICc      delta
## Min.   :1.00   Min.   :-142.5   Min.   : 0.000
## 1st Qu.:2.00   1st Qu.:-141.9   1st Qu.: 270.7   1st Qu.: 1.939
## Median :3.00   Median :-136.4   Median :279.2   Median :10.500
## Mean   :2.90   Mean   :-136.7   Mean   :279.3   Mean   :10.606
## 3rd Qu.:3.75   3rd Qu.:-131.9   3rd Qu.:288.4   3rd Qu.:19.656
```

```

##   Max.    :5.00    Max.   :-130.7    Max.   :289.4    Max.   :20.665
##
##      weight
##  Min.   :1.259e-05
##  1st Qu.:2.110e-05
##  Median :4.911e-02
##  Mean   :1.000e-01
##  3rd Qu.:1.489e-01
##  Max.   :3.867e-01
##

sw(dredged)

##                               period sex  elevation elevation:sex
## Sum of weights:       1.00  0.45  0.37      0.10
## N containing models: 5     6     6          2

head(dredged, 5)

## Global model call: glm(formula = molt ~ period + elevation + sex + elevation * sex,
##   family = binomial, na.action = "na.fail")
## ---
## Model selection table
##   (Int) elv    prd sex elv:sex df   logLik  AICc delta weight
## 3 -3.924    0.4305           2 -132.355 268.7  0.00  0.387
## 7 -4.268    0.4288  +        3 -131.801 269.6  0.91  0.245
## 4 -4.103    0.4284  +        3 -132.210 270.5  1.73  0.163
## 8 -4.495    0.4269  +        4 -131.605 271.3  2.56  0.108
## 16 -18.710   0.4216  +       + 5 -130.678 271.5  2.74  0.098
## Models ranked by AICc(x)

# The summary dredge table shows AICc and model weights. It shows that
# period is the best predictor for molt, whereas, elevation, sex, and their
# interaction contribute little to the model performance. The second summary,
# sum of weights, also confirms this.

# e) Assess the assumption of lack of over-dispersion and use a deletion test
# to compare the fit of the best model by AICc to the second best model by AICc
best.model <- get.models(dredged, 1)[[1]]
pp <- sum(resid(best.model, type = "pearson")^2)

1 - pchisq(pp, best.model$df.residual)

## [1] 0.9991569

1 - pchisq(best.model$deviance, best.model$df.residual)

## [1] 1

```

```

pp / best.model$df.residual

## [1] 0.8141618

# Pearson's  $X^2$  ( $P = 0.99$ ) shows no over dispersion and is surprisingly high.
# Dispersion parameter (0.81) indicates no over dispersion.

# Check if we need to average
head(dredged, 5)

## Global model call: glm(formula = molt ~ period + elevation + sex + elevation * sex,
##   family = binomial, na.action = "na.fail")
## ---
## Model selection table
##   (Int) elv     prd sex elv:sex df   logLik  AICc delta weight
## 3 -3.924    0.4305          2 -132.355 268.7  0.00  0.387
## 7 -4.268    0.4288  +      3 -131.801 269.6  0.91  0.245
## 4 -4.103    0.4284          3 -132.210 270.5  1.73  0.163
## 8 -4.495    0.4269  +      4 -131.605 271.3  2.56  0.108
## 16 -18.710   0.4216  +      5 -130.678 271.5  2.74  0.098
## Models ranked by AICc(x)

avg_mod <- model.avg(dredged, subset = delta < 2)
summary(avg_mod)

## 
## Call:
## model.avg(object = dredged, subset = delta < 2)
## 
## Component model call:
## glm(formula = molt ~ <3 unique rhs>, family = binomial, na.action =
##   na.fail)
## 
## Component models:
##   df  logLik  AICc delta weight
## 2  2 -132.36 268.73  0.00  0.49
## 23 3 -131.80 269.65  0.91  0.31
## 12 3 -132.21 270.47  1.73  0.20
## 
## Term codes:
## elevation     period       sex
##           1           2           3
## 
## Model-averaged coefficients:
## (full average)
##             Estimate Std. Error Adjusted SE z value Pr(>|z|)
## (Intercept) -4.06668   0.51467   0.51580  7.884 < 2e-16 ***
## period       0.42958   0.09774   0.09797  4.385 1.16e-05 ***
## sexM         0.13633   0.31679   0.31724  0.430   0.667
## elevationlow 0.04743   0.21923   0.21966  0.216   0.829
## 
## (conditional average)

```

```

##          Estimate Std. Error Adjusted SE z value Pr(>|z|)
## (Intercept) -4.06668   0.51467    0.51580  7.884 < 2e-16 ***
## period      0.42958   0.09774    0.09797  4.385 1.16e-05 ***
## sexM        0.44229   0.43620    0.43724  1.012   0.312
## elevationlow 0.23167   0.43825    0.43931  0.527   0.598
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# We model average because models have delta less than two.

exp(cbind(OR = coef(avg_mod)))

##          OR
## (Intercept) 0.01713413
## period      1.53660943
## sexM        1.55626067
## elevationlow 1.26070084

# f) summarize your best fitting model, report P values, and explain and
# interpret the results in terms of odds ratios.

# The averaged results indicate that period is the only significant predictor
# of wing molt. The effects of sex and elevation are weak and not significant.
# The best model is logit(p) = -4.07 + 0.43 (period) where both the intercept
# (-4.07 +- 0.51 SE, p < 0.001) and the period (0.43 +- 0.10 SE, z = 4.38, p < 0.001)
# were significant. The odds ratio of molt increased by 54% for each one-unit increase
# in period (OR = 1.54), indicating that amakihi were more likely to molt later
# in the season.

```